(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116096873 A
(43)申请公布日 2023.05.09

(21)申请号 202180049194.2
(22)申请日 2021.05.07
(30)优先权数据
63/022,397 2020.05.08 US
63/116,785 2020.11.20 US
(85)PCT国际申请进入国家阶段日
2023.01.09
(86)PCT国际申请的申请数据
PCT/US2021/031439 2021.05.07
(87)PCT国际申请的公布数据
WO2021/226558 EN 2021.11.11
(71)申请人 布罗德研究所股份有限公司
地址 美国马萨诸塞州
申请人 哈佛大学的校长及成员们
(72)发明人 D.R.刘 A.V.安扎隆 J.M.利维
X.高 C.J.波德拉茨基
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 张文辉
(51)Int.Cl.
C12N 9/12(2006.01)

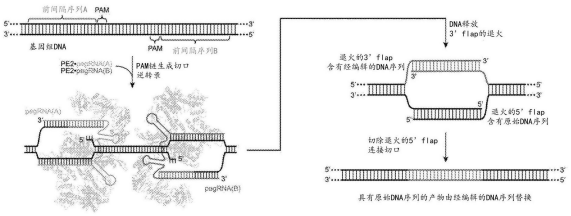
权利要求书31页 说明书612页
序列表（电子公布） 附图232页

(54)发明名称

同时编辑靶标双链核苷酸序列的两条链的方法和组合物

(57)摘要

本公开提供了用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统、组合物和方法。在一些方面，系统包括第一和第二引导编辑器复合物，其中第一和第二引导编辑器复合物中的每一个包含：(1)引导编辑器，其包含：(i)核酸可编程DNA结合蛋白(napDNAbp)，和(ii)具有RNA依赖性DNA聚合酶活性的多肽；和(2)pegRNA，其包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点，其中DNA合成模板编码期望的DNA序列或其互补物，其中期望的DNA序列及其互补物形成包含整合到待编辑的靶位点中的编辑部分的双链体。在一些方面中，系统包括各自包含引导编辑器和PEgRNA的第一、第二、第三和第四引导编辑器复合物。本文还提供了用于在待编辑的靶位点处同时编辑双链DNA的两条链的方法。本文进一步提供了药物组合物、多核苷酸、载体、细胞和试剂盒。



CN 116096873 A

1.用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统,其中所述系统包含第一引导编辑器复合物和第二引导编辑器复合物,其中所述第一和第二引导编辑器复合物中的每一个包含:(1)引导编辑器,其包含:(i)核酸可编程DNA结合蛋白(napDNAbp),和(ii)具有RNA依赖性DNA聚合酶活性的多肽;和(2)pegRNA,其包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点,其中所述第一引导编辑器复合物的所述pegRNA的所述DNA合成模板编码第一单链DNA序列,并且所述第二引导编辑器复合物的所述pegRNA的所述DNA合成模板编码第二单链DNA序列,其中所述第一单链DNA序列和所述第二单链DNA序列各自包含彼此互补的区域,并且其中所述第一单链DNA序列和所述第二单链DNA序列形成双链体,所述双链体包含与所述待编辑的靶位点处的所述DNA序列相比的编辑部分,所述编辑部分整合到所述待编辑的靶位点中。

2.权利要求1的系统,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白,所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

3.权利要求1或2的系统,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含相同的引导编辑器。

4.权利要求1或2的系统,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含不同的引导编辑器。

5.权利要求1至4中任一项的系统,其中所述napDNAbp是Cas9结构域或其变体。

6.权利要求1至5中任一项的系统,其中所述napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域、Cas9切口酶结构域或其变体。

7.权利要求1至4中任一项的系统,其中所述napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c、CasX、CasY和Argonaute。

8.权利要求1至7中任一项的系统,其中所述napDNAbp具有切口酶活性。

9.权利要求1至4中任一项的系统,其中所述napDNAbp包含SEQ ID NO:18-88中任何一个的氨基酸序列,或与SEQ ID NO:18-88中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

10.权利要求1至9中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的多肽是逆转录酶。

11.权利要求1至10中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的多肽包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159和700-766中任何一个的氨基酸序列,或与SEQ ID NO:89-100、105-122、128-129和132中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

12.权利要求1至11中任一项的系统,其中所述引导编辑器进一步包含接头,所述接头接合所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

13.权利要求12的系统,其中所述接头包含SEQ ID NO:166-177中任何一个的氨基酸序列,或与SEQ ID NO:166-177中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

14.权利要求12或13的系统,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、

38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

15.权利要求1至14中任一项的系统，其中所述pegRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列，或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

16.权利要求1至15中任一项的系统，其中所述第一引导编辑器复合物的所述pegRNA的所述间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

17.权利要求1至16中任一项的系统，其中所述第二引导编辑器复合物的所述pegRNA的所述间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

18.权利要求17的系统，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述pegRNA的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被所述第一和第二引导编辑器复合物中每一个的所述napDNAbp生成切口。

19.权利要求18的系统，其中所述第一和第二结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

20.权利要求19的系统，其中所述第一和第二结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

21.权利要求19的系统，其中所述连续区域被包含所述编辑部分的双链体置换。

22.权利要求1至21中任一项的系统，其中所述DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

23.权利要求1至22中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有相同的长度。

24.权利要求1至22中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有不同的长度。

25.权利要求1至24中任一项的系统，其中所述互补区域为所述第一单链DNA序列或所述第二单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

26.权利要求25的系统，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷

酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109核苷酸或至少110个核苷酸，并且小于或等于所述期望的DNA序列及其互补物的长度。

27.权利要求25的系统，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域形成包含所述编辑部分的所述双链体。

28.权利要求1至27中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列各自是具有3'末端的单链DNA链。

29.权利要求21的系统，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

30.用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统，所述系统包含：

a)第一引导编辑器复合物，其包含：

i.第一引导编辑器，其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽；和

ii.第一引导编辑向导RNA(第一PEgRNA)，其结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点；

b)第二引导编辑器复合物，其包含：

iii.第二引导编辑器，其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽；和

iv.第二引导编辑向导RNA(第二PEgRNA)，其结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点；

其中所述第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板，并且所述第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板，并且其中所述双链DNA序列的所述第一链和所述第二链彼此互补。

31.权利要求30的系统，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白，所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

32.权利要求30的系统,其中所述第一和第二单链DNA序列各自包含彼此互补的区域。

33.权利要求30的系统,其中所述第一单链DNA序列和所述第二单链DNA序列形成包含整合到所述待编辑的靶位点中的所述编辑部分的双链体。

34.权利要求30的系统,其中所述第一单链DNA序列和/或所述第二单链序列不包含与所述靶位点处的所述DNA序列相比的序列同源性。

35.权利要求30的系统,其中所述第一PEgRNA进一步包含第一间隔区序列、第一gRNA核心和第一引物结合位点,并且所述第二PEgRNA进一步包含第二间隔区序列、第二gRNA核心和第二引物结合位点。

36.权利要求30至35中任一项的系统,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含相同的引导编辑器。

37.权利要求30至36中任一项的系统,其中所述包含RNA依赖性DNA聚合酶活性的第一多肽和所述包含RNA依赖性DNA聚合酶活性的第二多肽是相同的。

38.权利要求30至37中任一项的系统,其中所述第一和/或第二napDNAbp是Cas9结构域或其变体。

39.权利要求30至38中任一项的系统,其中所述第一和/或第二napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域、Cas9切口酶结构域或其变体。

40.权利要求30至39中任一项的系统,其中所述第一和/或第二napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c、CasX、CasY和Argonaute。

41.权利要求30至40中任一项的系统,其中所述napDNAbp具有切口酶活性。

42.权利要求30至41中任一项的系统,其中所述第一和/或第二napDNAbp包含SEQ ID NO:18-88中任何一个的氨基酸序列,或与SEQ ID NO:18-88中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

43.权利要求30至42中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽是逆转录酶。

44.权利要求30至43中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159和700-766中任何一个的氨基酸序列,或与SEQ ID NO:89-100、105-122、128-129和132中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

45.权利要求30至44中任一项的系统,其中所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽;和/或所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽。

46.权利要求45的系统,其中所述接头包含SEQ ID NO:166-177中任何一个的氨基酸序列,或与SEQ ID NO:166-177中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

47.权利要求45的系统,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

48.权利要求30至47中任一项的系统，其中所述第一和/或第二pegRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列，或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

49.权利要求30至48中任一项的系统，其中所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

50.权利要求35的系统，其中所述第二引导编辑器复合物的所述间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

51.权利要求35的系统，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被所述第一和第二引导编辑器复合物中每一个的所述第一和/或第二napDNAbp生成切口。

52.权利要求30的系统，其中所述第一和第二结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

53.权利要求52的系统，其中所述第一和第二结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

54.权利要求52的系统，其中所述连续区域被包含所述编辑部分的双链体置换。

55.权利要求30至54中任一项的系统，其中所述第一和/或第二DNA合成模板的长度为

至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

56.权利要求30至55中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有相同的长度。

57.权利要求30至56中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有不同的长度。

58.权利要求30至57中任一项的系统，其中所述互补区域为所述第一单链DNA序列或所述第二单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

59.权利要求30至58中任一项的系统，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷

酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

60.权利要求30至59中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域形成双链体。

61.权利要求30至60中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列各自是具有3’末端的单链DNA链。

62.权利要求54的系统，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

63.多核苷酸，其编码权利要求1至62中任一项的系统。

64.多核苷酸，其编码权利要求1至62中任一项的第一和第二引导编辑器。

65.多核苷酸，其编码权利要求1至62中任一项的第一和/或第二pegRNA。

66.多核苷酸，其编码权利要求30至62中任一项的第一和/或第二pegRNA。

67.多核苷酸，其编码权利要求1至29中任一项的第一和第二pegRNA。

68.多核苷酸，其编码权利要求30至62中任一项的第一和第二pegRNA。

69.载体，其包含权利要求63至68中任一项的多核苷酸。

70.细胞，其包含权利要求63至68中任一项的多核苷酸或权利要求69的载体。

71.药物组合物，其包含权利要求63至68中任一项的多核苷酸、权利要求69的载体或权利要求70的细胞以及药物赋形剂。

72.试剂盒，其包含权利要求63至68中任一项的多核苷酸或权利要求69的载体以及用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的使用说明。

73.用于在靶位点处同时编辑双链DNA序列的第一和第二互补链的方法，所述方法包括将所述双链DNA序列与引导编辑器复合物对接触，所述对包含：

a.第一引导编辑器复合物，其包含：

i.第一引导编辑器，其包含第一核酸可编程DNA结合蛋白(napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽；和

ii.第一引导编辑向导RNA(第一PEgRNA)，其结合至所述靶位点上游的基因组DNA序列的第一链上的第一结合位点；

b.第二引导编辑器复合物，其包含：

i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述靶位点下游的基因组DNA序列的第二链上的第二结合位点;

其中所述第一引导编辑器复合物导致与所述第一结合位点互补的序列处的第一切口以及随后从由所述第一切口形成的可用5'末端聚合具有3'末端的第一单链DNA序列;

其中所述第二引导编辑器复合物导致与所述第二结合位点互补的序列处的第二切口以及随后从由所述第二切口形成的可用5'末端聚合具有3'末端的第二单链DNA序列;

其中所述第一单链DNA序列和所述第二单链DNA序列至少在互补区域上是反向互补物并且形成包含编辑的双链体;并且

其中所述双链体置换所述双链DNA序列的带切口的第一和第二互补链。

74.权利要求73的方法,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白,所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

75.权利要求73的方法,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含相同的引导编辑器。

76.权利要求73的方法,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含不同的引导编辑器。

77.权利要求73至76中任一项的方法,其中所述第一和/或第二napDNAbp是Cas9结构域或其变体。

78.权利要求73至77中任一项的方法,其中所述第一和/或第二napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域、Cas9切口酶结构域或其变体。

79.权利要求73至78中任一项的方法,其中所述第一和/或第二napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c、CasX、CasY和Argonaute。

80.权利要求73至79中任一项的方法,其中所述第一和/或第二napDNAbp具有切口酶活性。

81.权利要求73至80中任一项的方法,其中所述第一和/或第二napDNAbp包含SEQ ID NO:18-88中任何一个的氨基酸序列,或与SEQ ID NO:18-88中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

82.权利要求73至81中任一项的方法,其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽是逆转录酶。

83.权利要求73至82中任一项的方法,其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159和700-766中任何一个的氨基酸序列,或与SEQ ID NO:89-100、105-122、128-129和132中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

84.权利要求73至83中任一项的方法,其中所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽,和/或所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性

DNA聚合酶活性的第二多肽。

85.权利要求84的方法，其中所述接头包含SEQ ID NO:166-177中任何一个的氨基酸序列，或与SEQ ID NO:166-177中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

86.权利要求84或85的方法，其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

87.权利要求73至86中任一项的方法，其中所述第一和/或第二pegRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列，或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

88.权利要求73至87中任一项的方法，其中所述第一和/或第二PEgRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列，或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

89.权利要求73至88中任一项的方法，其中所述第一引导编辑器复合物的所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

90.权利要求73至89中任一项的方法，其中所述第二引导编辑器复合物的所述第二间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

91.权利要求90的方法，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述第一和第二间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第一和第二引导编辑器复合物各自的napDNAbp生成切口。

92.权利要求91的方法，其中所述第一和第二结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

93.权利要求92的方法，其中所述第一和第二结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷

酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

94.权利要求92或93的方法，其中所述连续区域被包含所述编辑部分的所述双链体置换。

95.权利要求73至94中任一项的方法，其中所述第一和/或第二DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

96.权利要求73至95中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列具有相同的长度。

97.权利要求73至96中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列具有不同的长度。

98.权利要求73至97中任一项的方法，其中所述互补区域为所述第一单链DNA序列或所述第二单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

99.权利要求73至98中任一项的方法，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸，并且小于或等于所述期望的DNA序列及其互补物的长度。

100.权利要求73至99中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域形成所述双链体。

101.权利要求73至100中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列各自是具有3’末端的单链DNA链。

102.权利要求94的方法，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

103.权利要求73至102中任一项的方法，其中所述双链体包含编辑，所述编辑包含一个或多个重组酶位点。

104.权利要求103的方法，其进一步包括提供重组酶以促进所述待编辑的靶位点处的所述DNA序列的重组酶介导的编辑。

105.pegRNA对，其用于双重引导编辑，其中所述对包含：

a.第一引导编辑向导RNA(第一PEgRNA)，其结合至待编辑的靶位点上游的双链DNA序列

的第一链上的第一结合位点，其中所述第一PEgRNA包含第一间隔区序列、第一gRNA核心、第一DNA合成模板和第一引物结合位点，其中所述第一DNA合成模板编码包含编辑部位的第一单链DNA序列；

b.第二引导编辑向导RNA(第二PEgRNA)，其结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点，其中所述第二PEgRNA包含第二间隔区序列、第二gRNA核心、第二DNA合成模板和第二引物结合位点，其中所述第二DNA合成模板编码包含所述编辑部分的第二单链DNA序列，其中所述第二单链DNA序列包含所述第一单链DNA序列的互补区域。

106.权利要求105的pegRNA对，其中所述第一PEgRNA包含以下结构：

5’-[第一gRNA间隔区序列]-[第一gRNA核心]-[第一DNA合成模板]-[第一引物结合位点]-3’；且

其中所述第二PEgRNA包含以下结构：

5’-[第二gRNA间隔区序列]-[第二gRNA核心]-[第二DNA合成模板]-[第二引物结合位点]-3’。

107.权利要求105或106的pegRNA对，其中所述第一单链DNA序列和所述第二单链DNA序列形成双链体，所述双链体在所述待编辑的靶位点处整合到所述双链DNA序列中。

108.权利要求107的pegRNA对，其中所述双链体的所述整合导致所述靶位点处的DNA的插入、缺失或替代。

109.权利要求108的pegRNA对，其中所述插入的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、

至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

110.权利要求108的pegRNA对，其中所述缺失的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

111.权利要求108的pegRNA对，其中所述替代的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷

酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

112.多核苷酸，其编码权利要求105至111中任一项的pegRNA对。

113.包含权利要求112的多核苷酸的载体，其用于在细胞中表达所述pegRNA对。

114.细胞，其包含权利要求113的载体。

115.药物组合物，其包含权利要求105至111中任一项的pegRNA对、权利要求113的载体或权利要求114的细胞以及药物赋形剂。

116.引导编辑器复合物对，所述对包含：

a.第一引导编辑器复合物，其包含：

i.第一引导编辑器，其包含第一核酸可编程DNA结合蛋白(napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽；和

ii.第一引导编辑向导RNA(第一PEgRNA)，其结合至靶位点上游的基因组DNA序列的链上的第一靶序列；

b.第二引导编辑器复合物，其包含：

i.第二引导编辑器，其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽；和

ii.第二引导编辑向导RNA(第二PEgRNA)，其结合至靶位点下游的基因组DNA序列的第二链上的第二结合位点。

117.权利要求116的引导编辑器复合物对，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白，所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

118.多核苷酸，其编码权利要求116或117的引导编辑器复合物对。

119.包含权利要求118的多核苷酸的载体，其用于在细胞中表达所述引导编辑器复合物对。

120.细胞，其包含权利要求119的载体。

121.药物组合物，其包含权利要求116或117的引导编辑器复合物对、权利要求119的载体或权利要求120的细胞以及药物赋形剂。

122.权利要求1至62中任一项的系统或权利要求73至104中任一项的方法，其中包含所述编辑部分的所述双链体的整合导致所述靶位点处的DNA的插入、缺失或替代。

123.权利要求122的系统或方法，其中所述插入是肽标签。

124.权利要求123的系统或方法，其中所述肽标签是多组氨酸(例如，HHHHHH)(SEQ ID

NO:252-262)、FLAG(例如,DYKDDDDK)(SEQ ID NO:2)、V5(例如,GKPIPNPLLGLDST)(SEQ ID NO:3)、GCN4、HA(例如,YPYDVPDYA)(SEQ ID NO:5)、Myc(例如EQKLISEED)(SEQ ID NO:6)或GST。

125.权利要求123的系统或方法,其中所述肽标签具有选自由以下各项组成的组的氨基酸序列:SEQ ID NO:1-6、245-249、252-262、264-273、275-276、281、278-288和622。

126.权利要求122的系统或方法,其中所述插入是免疫表位标签。

127.权利要求126的系统或方法,其中所述免疫表位标签选自由以下各项组成的组:破伤风类毒素(SEQ ID NO:396);白喉毒素突变体CRM197(SEQ ID NO:630);腮腺炎免疫表位1(SEQ ID NO:400);腮腺炎免疫表位2(SEQ ID NO:402);腮腺炎免疫表位3(SEQ ID NO:404);风疹病毒(SEQ ID NO:406);血凝素(SEQ ID NO:408);神经氨酸酶(SEQ ID NO:410);TAP1(SEQ ID NO:412);TAP2(SEQ ID NO:414);血凝素表位,针对HLAI类(SEQ ID NO:416);神经氨酸酶表位,针对HLAI类(SEQ ID NO:418);血凝素表位,针对HLAII类(SEQ ID NO:420);神经氨酸酶表位,针对HLAII类(SEQ ID NO:422);血凝素表位,H5N1结合的HLAI类和II类(SEQ ID NO:424);和神经氨酸酶表位,H5N1结合的HLAI类和II类(SEQ ID NO:426)。

128.权利要求122的系统或方法,其中所述插入是二聚化结构域。

129.权利要求128的系统或方法,其中所述二聚化结构域是SEQ ID NO:488的FKBP12、SEQ ID NO:489的FKBP12-F36V或SEQ ID NO:490或493-494的亲环蛋白的小分子结合结构域。

130.用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统,所述系统包含:

a)第一引导编辑器复合物,其包含:

i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

ii.第一引导编辑向导RNA(第一PEgRNA),其结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点;

b)第二引导编辑器复合物,其包含:

i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述待编辑的靶位点上游的所述双链DNA序列的第二链上的第二结合位点;

c)第三引导编辑器复合物,其包含:

i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

ii.第三引导编辑向导RNA(第三PEgRNA),其结合至所述待编辑的靶位点下游的所述双链DNA序列的所述第一链上的第三结合位点;

d)第四引导编辑器复合物,其包含:

i.第四引导编辑器,其包含第四核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

ii.第四引导编辑向导RNA(第四PEgRNA),其结合至所述待编辑的靶位点下游的所述双链DNA序列的所述第二链上的第四结合位点;

其中所述第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板,所述第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板,所述第三PEgRNA包含编码第三单链DNA序列的第三DNA合成模板,且所述第四PEgRNA包含编码第四单链DNA序列的第四DNA合成模板;

其中所述第一和所述第三单链DNA序列各自包含彼此互补的区域;并且其中,其中所述第二和第四单链DNA序列各自包含彼此互补的区域。

131.用于编辑一个或多个双链DNA序列的系统,所述系统包含:

a)第一引导编辑器复合物,其包含:

i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

ii.第一引导编辑向导RNA(第一PEgRNA),其结合至第一待编辑的靶位点处的第一双链DNA序列的第一链上的第一结合位点;

b)第二引导编辑器复合物,其包含:

i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述第一待编辑的靶位点处的所述第一双链DNA序列的第二链上的第二结合位点;

c)第三引导编辑器复合物,其包含:

i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

ii.第三引导编辑向导RNA(第三PEgRNA),其结合至第二待编辑的靶位点处的第二双链DNA序列的第一链上的第一结合位点;

d)第四引导编辑器复合物,其包含:

i.第四引导编辑器,其包含第四核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

ii.第四引导编辑向导RNA(第四PEgRNA),其结合至所述第二待编辑的靶位点处的所述第二双链DNA序列的第二链上的第二结合位点;

其中所述第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板,所述第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板,所述第三PEgRNA包含编码第三单链DNA序列的第三DNA合成模板,且所述第四PEgRNA包含编码第四单链DNA序列的第四DNA合成模板;

其中所述第一和所述第三单链DNA序列各自包含彼此互补的区域;并且其中,其中所述第二和第四单链DNA序列各自包含彼此互补的区域。

132.权利要求131的系统,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和/或所述第四引导编辑器复合物的引导编辑器是融合蛋白,所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

133.权利要求130至132中任一项的系统,其中

(i)所述第一PEgRNA进一步包含第一间隔区序列、第一gRNA核心和第一引物结合位点;

(ii)所述第二PEgRNA进一步包含第二间隔区序列、第二gRNA核心和第二引物结合位点;

(iii)所述第三PEgRNA进一步包含第三间隔区序列、第三gRNA核心和第三引物结合位

点;且

(iv)所述第四PEgRNA包含第四间隔区序列、第四gRNA核心、第四DNA合成模板和第四引物结合位点。

134.权利要求131的系统,其中所述第一和第三单链DNA序列的所述互补区域和/或所述第二和第四单链DNA序列的所述互补区域形成双链体。

135.权利要求134的系统,其中所述第一间隔区序列结合至所述第一待编辑的靶位点上游的所述第一双链DNA序列的所述第一结合位点,并且所述第二间隔区序列结合至所述第一待编辑的靶位点下游的所述第一双链DNA序列上的所述第二结合位点。

136.权利要求134或135的系统,其中所述第三间隔区序列结合至所述第二待编辑的靶位点上游的所述第二双链DNA序列的所述第一结合位点,并且所述第四间隔区序列结合至所述第二待编辑的靶位点下游的所述第二双链DNA序列上的所述第二结合位点。

137.权利要求134至136中任一项的系统,其中所述第一和第二结合位点定义了所述第一双链DNA的第一连续区域的两端。

138.权利要求134至137中任一项的系统,其中所述第三和第四结合位点定义了所述第二双链DNA的第二连续区域的两端。

139.权利要求134的系统,其中所述第一和第二结合位点之间或所述第三和第四结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对,至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

140.权利要求134至139中任一项的系统,其中所述第一单链DNA序列和所述第二单链

DNA序列的所述互补区域和/或所述第二单链DNA序列和所述第四单链DNA序列的所述互补区域形成双链体。

141.权利要求140的系统,其中所述双链体掺入所述第一双链DNA和/或所述第二双链DNA中。

142.权利要求134至141中任一项的系统,其中所述系统导致所述第二连续区域置换所述第一连续区域进入所述第一双链DNA中。

143.权利要求134至142中任一项的系统,其中所述系统导致在所述第一双链DNA和所述第二双链DNA之间交换所述第一连续区域和所述第二连续区域。

144.权利要求134至143中任一项的系统,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和所述第四引导编辑器复合物包含相同的引导编辑器。

145.权利要求134至144中任一项的系统,其中所述包含RNA依赖性DNA聚合酶活性的第一、第二、第三和第四多肽是相同的。

146.权利要求134至145中任一项的系统,其中所述第一、第二、第三和/或第四napDNAbp是Cas9结构域或其变体。

147.权利要求134至146中任一项的系统,其中所述第一、第二、第三和/或第四napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域或Cas9切口酶结构域或其变体。

148.权利要求134至147中任一项的系统,其中所述第一、第二、第三和/或第四napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c和Argonaute。

149.权利要求134至148中任一项的系统,其中所述第一、第二、第三和/或第四napDNAbp具有切口酶活性。

150.权利要求134至149中任一项的系统,其中所述第一、第二、第三和/或第四napDNAbp包含SEQ ID NO:2-65中任何一个的氨基酸序列,或与SEQ ID NO:2-65中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

151.权利要求134至150中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽是逆转录酶。

152.权利要求134至151中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽包含SEQ ID NO:37、68-79、82-98、81、98和110中任何一个的氨基酸序列,或者与SEQ ID NO:37、68-79、82-98、81、98和110中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

153.权利要求134至152中任一项的系统,其中

(i)所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽;

(ii)所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽;

(iii)所述第三引导编辑器进一步包含接头,所述接头接合所述第三napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第三多肽;和/或

(iv)所述第四引导编辑器进一步包含接头,所述接头接合所述第四napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第四多肽。

154.权利要求153的系统,其中所述接头包含SEQ ID NO:119-128中任何一个的氨基酸序列,或与SEQ ID NO:119-128中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

155.权利要求153或154的系统,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

156.权利要求134至155中任一项的系统,其中所述第一、第二、第三和/或第四PEgRNA包含SEQ ID NO:192-203中任何一个的核苷酸序列,或与SEQ ID NO:192-203中任何一个具有至少80%、85%、90%、95%或99%序列同一性的核苷酸序列。

157.权利要求134至156中任一项的系统,其中所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

158.权利要求134至157中任一项的系统,其中所述第二引导编辑器复合物的所述第二间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第二链上的第二结合位点。

159.权利要求134至158中任一项的系统,其中所述第三间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第一链上的第三结合位点。

160.权利要求134至159中任一项的系统,其中所述第四引导编辑器复合物的所述第四间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第四结合位点。

161.权利要求160的系统,其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第一和第二引导编辑器复合物各自的所述第一和/或第二napDNAbp生成切口。

162.权利要求161的系统,其中在PAM序列存在下所述第三和第四引导编辑器复合物的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第三和第四引导编辑器复合物各自的所述第三和/或第四napDNAbp生成切口。

163.权利要求133的系统,其中所述第一和第三结合位点或所述第二和第四结合位点定义了双链DNA的连续区域的两端。

164.权利要求163的系统,其中所述第一和第三结合位点之间或所述第二和第四结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对,至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48

个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

165.权利要求130至164中任一项的系统，其中所述第一单链DNA序列和所述第三单链DNA序列的所述互补区域和/或所述第二单链DNA序列和所述第四单链DNA序列的所述互补区域形成双链体。

166.权利要求164的系统，其中所述连续区域被所述双链体置换。

167.权利要求166的系统，其中所述双链体在所述靶位点处掺入所述双链DNA中。

168.权利要求122至167中任一项的系统，其中所述第一、第二、第三和/或第四DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至

少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

169.权利要求122至168中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有相同的长度。

170.权利要求122至169中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有不同的长度。

171.权利要求122至170中任一项的系统，其中所述第一、第二、第三和/或第四单链DNA序列的所述互补区域为所述第一单链DNA序列和所述第三单链DNA序列或所述第二单链DNA序列和所述第四单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

172.权利要求122至171中任一项的系统，其中所述第一、第二、第三和/或第四单链DNA序列的所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸，并且小于或等于所述第一单链DNA序列和所述第三单链DNA序列以及所述第二单链DNA序列和所述第四单链DNA序列的长度。

173.权利要求130至172中任一项的系统，其中所述第一、第二、第三和第四单链DNA序列各自是具有3’末端的单链DNA链。

174.权利要求173的系统，其中包含所述编辑部分的所述双链体的掺入导致所述靶位点处的DNA的插入、缺失或替代。

175.多核苷酸,其编码权利要求130至174中任一项的第一、第二、第三和第四引导编辑器。

176.多核苷酸,其编码权利要求130至175中任一项的第一、第二、第三和第四PEgRNA。

177.载体,其包含权利要求175或176的多核苷酸。

178.细胞,其包含权利要求175或176的多核苷酸或权利要求177的载体。

179.药物组合物,其包含权利要求175或176的多核苷酸、权利要求177的载体或权利要求178的细胞以及药物赋形剂。

180.试剂盒,其包含权利要求175或176的多核苷酸或权利要求177的载体以及用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的使用说明。

181.用于在靶位点处同时编辑双链DNA序列的第一和第二互补链的方法,所述方法包括将所述双链DNA序列与引导编辑器复合物对接触,所述对包含:

(a)第一引导编辑器复合物,其包含:

i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

ii.第一引导编辑向导RNA(第一PEgRNA),其结合至所述靶位点上游的基因组DNA序列的第一链上的第一靶序列;

(b)第二引导编辑器复合物,其包含:

i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述靶位点上游的基因组DNA序列的第二链上的第二结合位点;

(c)第三引导编辑器复合物,其包含:

i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

ii.第三引导编辑向导RNA(第三PEgRNA),其结合至所述靶位点下游的基因组DNA序列的第一链上的第三靶序列;

(d)第四引导编辑器复合物,其包含:

i.第四引导编辑器,其包含第二核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

ii.第四引导编辑向导RNA(第四PEgRNA),其结合至所述靶位点下游的基因组DNA序列的第二链上的第四结合位点;

其中所述第一引导编辑器复合物导致第一靶序列处的第一切口以及随后从由所述第一切口形成的可用5'末端聚合具有3'末端的第一单链DNA序列;

其中所述第二引导编辑器复合物导致第二靶序列处的第二切口以及随后从由所述第二切口形成的可用5'末端聚合具有3'末端的第二单链DNA序列;

其中所述第三引导编辑器复合物导致第三靶序列处的第三切口以及随后从由所述第三切口形成的可用5'末端聚合具有3'末端的第三单链DNA序列;

其中所述第四引导编辑器复合物导致第四靶序列处的第四切口以及随后从由所述第四切口形成的可用5'末端聚合具有3'末端的第四单链DNA序列;

其中所述第一单链DNA序列和所述第二单链DNA序列至少在互补区域上是反向互补物并且形成双链体,其中所述双链体置换所述双链DNA序列的带切口的第一和第二互补链;并且

其中所述第三单链DNA序列和所述第四单链DNA序列至少在互补区域上是反向互补物并且形成双链体,其中所述双链体置换所述双链DNA序列的带切口的第一和第二互补链。

182.权利要求181的方法,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和/或所述第四引导编辑器复合物的引导编辑器是融合蛋白,所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

183.权利要求181或182的方法,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和所述第四引导编辑器复合物包含相同的引导编辑器。

184.权利要求181至183中任一项的方法,其中所述第一、第二、第三和/或第四引导编辑器复合物包含不同的引导编辑器。

185.权利要求181至184中任一项的方法,其中第一、第二、第三和/或第四napDNAbp是Cas9结构域或其变体。

186.权利要求181至185中任一项的方法,其中第一、第二、第三和/或第四napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域或Cas9切口酶结构域其变体。

187.权利要求181至184中任一项的方法,其中所述第一、第二、第三和/或第四napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c。

188.权利要求181至187中任一项的方法,其中所述第一、第二、第三和/或第四napDNAbp具有切口酶活性。

189.权利要求181至188中任一项的方法,其中所述第一、第二、第三和/或第四napDNAbp包含SEQ ID NO:2-65中任何一个的氨基酸序列,或与SEQ ID NO:2-65中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

190.权利要求181至189中任一项的方法,其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽是逆转录酶。

191.权利要求181至190中任一项的方法,其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽包含SEQ ID NO:37、68-79、82-98、81、98和110中任何一个的氨基酸序列,或者与SEQ ID NO:37、68-79、82-98、81、98和110中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

192.权利要求181至191中任一项的方法,其中

(i)所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽;

(ii)所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽;

(iii)所述第三引导编辑器进一步包含接头,所述接头接合所述第三napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第三多肽;和/或

(iv)所述第四引导编辑器进一步包含接头,所述接头接合所述第四napDNAbp和所述具

有所述RNA依赖性DNA聚合酶活性的第四多肽。

193.权利要求192的方法，其中所述接头包含SEQ ID NO:119-128中任何一个的氨基酸序列，或与SEQ ID NO:119-128中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

194.权利要求192或193的方法，其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

195.权利要求192至194中任一项的方法，其中所述第一、第二、第三和/或第四PEgRNA包含SEQ ID NO:192-203中任何一个的核苷酸序列，或与SEQ ID NO:192-203中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

196.权利要求192至195中任一项的方法，其中所述第一引导编辑器复合物的所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

197.权利要求192至196中任一项的方法，其中所述第二引导编辑器复合物的所述第二间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

198.权利要求192至197中任一项的方法，其中所述第三引导编辑器复合物的所述第三间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第三结合位点。

199.权利要求192至198中任一项的方法，其中所述第四引导编辑器复合物的所述第四间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第四结合位点。

200.权利要求199的方法，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述第一和第二间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第一和第二引导编辑器复合物各自的napDNAbp生成切口。

201.权利要求200的方法，其中在PAM序列存在下所述第三和第四引导编辑器复合物的所述第三和第四间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第三和第四引导编辑器复合物各自的所述napDNAbp生成切口。

202.权利要求200或201的方法，其中所述第一和第二结合位点和/或所述第三和第四结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

203.权利要求202的方法，其中所述第一和第二结合位点之间和/或所述第三和第四结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至

少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

204.权利要求202或203的方法，其中所述连续区域被所述双链体置换。

205.权利要求181至204中任一项的方法，其中所述第一、第二、第三和/或第四DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

206.权利要求181至205中任一项的方法，其中所述第一单链DNA序列和所述第二单链

DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有相同的长度。

207.权利要求181至206中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有不同的长度。

208.权利要求181至207中任一项的方法，其中所述互补区域为所述第一单链DNA序列和所述第二单链DNA序列或所述第三单链DNA序列和所述第四单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

209.权利要求181至208中任一项的方法，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

210.权利要求188至209中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列的所述互补区域形成所述双链体。

211.权利要求188至210中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列各自是具有3’末端的单链DNA链。

212.权利要求204的方法，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

213.权利要求181至212中任一项的方法，其中所述第一单链DNA序列和所述第二单链

DNA序列位于靶DNA序列的末端，并且其中所述第三单链DNA序列和所述第四单链DNA序列位于同一靶DNA序列的相反末端。

214.权利要求213的方法，其中所述靶DNA序列是反向的。

215.权利要求181至214中任一项的方法，其进一步包括提供环状DNA供体。

216.权利要求215的方法，其中所述第一单链DNA序列和所述第三单链DNA序列位于所述靶DNA序列的相反末端，并且其中所述第二单链DNA序列和所述第四单链DNA序列位于所述环状DNA供体上。

217.权利要求216的方法，其中所述第二单链DNA序列和所述第四单链DNA序列之间所述环状DNA供体的部分置换所述第一单链DNA序列和所述第三单链DNA序列之间的所述靶序列。

218.权利要求181至217中任一项的方法，其中所述第一单链DNA序列和所述第三单链DNA序列位于第一核酸分子上，并且其中所述第二单链DNA序列和所述第四单链DNA序列位于第二核酸分子上。

219.权利要求218的方法，其中所述第一单链DNA序列和所述第三单链DNA序列之间所述第一核酸分子的部分掺入到所述第二核酸分子中，并且其中所述第二单链DNA序列和所述第四单链DNA序列之间所述第二核酸分子的部分掺入到所述第一核酸分子中。

220.权利要求218或219的方法，其中所述第一核酸分子是第一染色体，并且所述第二核酸分子是第二染色体。

221.PEgRNA组，其用于多flap引导编辑，其中所述PEgRNA包含：

(a)第一引导编辑向导RNA(第一PEgRNA)，其结合至待编辑的靶位点上游的双链DNA序列的第一链上的第一结合位点，其中所述第一PEgRNA包含第一间隔区序列、第一gRNA核心、第一DNA合成模板和第一引物结合位点，其中所述第一DNA合成模板编码第一单链DNA序列；

(b)第二引导编辑向导RNA(第二PEgRNA)，其结合至所述待编辑的靶位点上游的所述双链DNA序列的第二链上的第二结合位点，其中所述第二PEgRNA包含第二间隔区序列、第二gRNA核心、第二DNA合成模板和第二引物结合位点，其中所述第二DNA合成模板编码包含所述编辑部分的第二单链DNA序列，其中所述第二单链DNA序列与所述第一单链DNA序列互补；

(c)第三引导编辑向导RNA(第三PEgRNA)，其结合至待编辑的靶位点下游的双链DNA序列的第一链上的第三结合位点，其中所述第三PEgRNA包含第三间隔区序列、第三gRNA核心、第三DNA合成模板和第三引物结合位点，其中所述第三DNA合成模板编码第三单链DNA序列；

(d)第四引导编辑向导RNA(第四PEgRNA)，其结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第四结合位点，其中所述第四PEgRNA包含第四间隔区序列、第四gRNA核心、第四DNA合成模板和第四引物结合位点，其中所述第四DNA合成模板编码第四单链DNA序列，其中所述第四单链DNA序列包含所述第三单链DNA序列的互补区域。

222.权利要求221的PEgRNA组，其中

所述第一PEgRNA包含以下结构：

5’-[第一gRNA间隔区序列]-[第一gRNA核心]-[第一DNA合成模板]-[第一引物结合位点]-3’；

所述第二PEgRNA包含以下结构：

5’-[第二gRNA间隔区序列]-[第二gRNA核心]-[第二DNA合成模板]-[第二引物结合位

点]-3’；

所述第三PEgRNA包含以下结构：

5’-[第三gRNA间隔区序列]-[第三gRNA核心]-[第三DNA合成模板]-[第三引物结合位点]-3’；且

所述第四PEgRNA包含以下结构：

5’-[第四gRNA间隔区序列]-[第四gRNA核心]-[第四DNA合成模板]-[第四引物结合位点]-3’。

223.权利要求221或222的PEgRNA组，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列形成双链体，所述双链体在所述待编辑的靶位点处整合到所述双链DNA序列中。

224.权利要求221至223中任一项的PEgRNA组，其中所述双链体的所述整合导致所述靶位点处的DNA的插入、缺失或替代。

225.权利要求224的PEgRNA组，其中所述插入的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

226.权利要求224的PEgRNA组，其中所述缺失的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核

苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

227.权利要求224的PEgRNA组，其中所述替代的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷

酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

228.多核苷酸，其编码权利要求221至227中任一项的PEgRNA组。

229.包含权利要求228的多核苷酸的载体，其用于在细胞中表达多个PEgRNA。

230.细胞，其包含权利要求229的载体。

231.药物组合物，其包含多个权利要求221至227中任一项的PEgRNA、权利要求的载体229或权利要求230的细胞以及药物赋形剂。

232.用于修饰细胞中基因组的方法，其包括：

(i)将基因组与双flap引导编辑系统接触，所述系统包含能够在靶位点的第一链上的第一切口位点处安装第一3’核酸flap的第一引导编辑器/pegRNA复合物和能够在靶位点的第二链上的第二切口位置处安装第二3’核酸flap的第二引导编辑器/pegRNA复合物，

其中所述第一3’核酸flap和所述第二3’核酸flap是形成双链体的反向互补序列，并且

其中所述双链体包含第一重组酶位点，并且

其中所述细胞在所述靶位点处安装包含所述第一重组酶位点的双链体，替代了定位于所述第一和第二切口位点之间的内源性双链体；

(ii)将所述第一重组酶位点与包含第二重组酶位点的供体核酸接触；和

(iii)将所述基因组与能够在所述第一和第二重组酶位点之间重组的重组酶接触，从而将所述供体核酸在所述靶位点处整合到所述基因组中。

同时编辑靶标双链核苷酸序列的两条链的方法和组合物

[0001] 政府支持

[0002] 本发明是在获得美国国家卫生研究院授予的资助号U01AI142756、RM1HG009490、R01EB022376和R35GM118062的政府支持下进行的。政府对本发明享有一定的权利。

[0003] 相关申请和引用并入

[0004] 本申请要求于2020年5月8日提交的美国临时专利申请号63/022,397和于2020年11月20日提交的美国临时专利申请号63/116,785的优先权,以上每篇申请的全部内容通过引用全部并入本文。

[0005] 本美国临时申请涉及并通过引用并入以下申请,即,2019年3月19日提交的美国临时申请号62/820,813(代理人案卷号B1195.70074US00)、2019年6月7日提交的美国临时申请号62/858,958(代理人案卷号B1195.70074US01),2019年8月21日提交的美国临时申请号62/889,996(代理人案卷号B1195.70074US02),2019年8月21日提交的美国临时申请号62/922,654(代理人案卷编号B1195.70083US00),美国临时申请号62/889,996,2019年10月10日提交的美国临时申请号62/913,553(代理人案卷编号B1195.70074US03),2019年10月10日提交的美国临时申请号62/973,558(代理人案卷编号B1195.70083US01),2019年11月5日提交的美国临时申请号62/931,195(代理人案卷号B1195.70074US04),2019年12月5日提交的美国临时申请号62/944,231(代理人案卷号B1195.70074US05),2019年12月5日提交的美国临时申请号62/974,537(代理人案卷号B1195.70083US02),2020年3月17日提交的美国临时申请号62/991,069(代理人案卷号B1195.70074US06),以及2020年3月17日提交的美国临时申请号63/100,548(代理人案卷号B1195.70083US03)。此外,本美国临时申请涉及并通过引用并入各项:国际PCT申请号:PCT/US20/23721;PCT/US20/23730;PCT/US20/23713;PCT/US20/23712;PCT/US20/23727;PCT/US20/23724;PCT/US20/23725;PCT/US20/23728;PCT/US20/23732;PCT/US20/23723;PCT/US20/23553;和PCT/US20/23583,每一篇均于2020年3月19日提交。

[0006] 发明背景

[0007] 据某些估计,致病性单核苷酸突变导致约50%的有遗传组分的人类疾病[7]。不幸地,尽管进行了数十年的基因治疗探索,但这些遗传性疾病患者的治疗选择仍然非常有限[8]。也许应对这一治疗挑战的最节俭的解决方案是直接校正患者基因组中的单核苷酸突变,这将解决疾病的根本原因并可能提供持久的益处。尽管这种策略以前是不可想象的,但最近CRISPR/Cas系统[9]的出现带来的基因组编辑能力的改进现在已经使这种治疗方法触手可及。通过直接设计包含与靶DNA序列互补的约20个核苷酸的向导RNA(guide RNA,gRNA)序列,CRISPR相关(Cas)核酸酶可特异性接近几乎任何可想到的基因组位点[1,2]。迄今为止,已鉴定了几种单体细菌Cas核酸酶系统并调整用于基因组编辑应用[10]。Cas核酸酶的这种天然多样性,以及越来越多的工程化变体[11-14],为开发新的基因组编辑技术提供了肥沃的土壤。

[0008] 虽然利用CRISPR进行基因破坏目前是成熟的技术,但精确编辑人类基因组中的单碱基对仍然是主要挑战[3]。同源定向修复(HDR)长期以来一直用于在人类细胞和其他生物体中使用编码期望编辑的供体DNA修复模板在双链断裂(DSB)位置插入、校正或交换DNA序列。

然而,传统HDR在大多数人类细胞类型中的效率非常低,尤其是在非分裂细胞中,并且竞争性非同源末端连接(NHEJ)主要导致插入-缺失(indel)副产物[16]。其他问题与DSB的产生有关,这会导致靶基因座处大的染色体重排和缺失[17],或激活p53轴,导致生长停滞和凋亡[18,19]。

[0009] 已经探索了几种方法来解决HDR的这些缺点。例如,已证明利用寡核苷酸供体修复单链DNA断裂(缺口)减少插入/缺失形成,但期望的修复产物的产率仍然很低[20]。其他策略尝试使用小分子和生物试剂将修复偏向于HDR而非NHEJ[21-23]。然而,这些方法的有效性可能取决于细胞类型,并且扰动正常细胞状态会导致不期望且不可预见的影响。

[0010] 最近,由DavidLiu教授等领导的发明人开发了碱基编辑作为编辑靶核苷酸而不形成DSB或依赖于HDR的技术[4-6,24-27]。通过Cas融合脱氨酶直接修饰DNA碱基可在短的靶窗口(约5-7个碱基)内以非常高的效率将C·G转换为T·A,或A·T转换为G·C。因此,碱基编辑器(editor)已迅速被科学界采用。然而,以下因素限制了它们在精确基因组编辑中的普遍性:(1)观测到靶窗口内的非靶C或A碱基的"旁观者编辑";(2)观测到靶核苷酸产物混合物;(3)靶碱基必须位于PAM序列上游15±2个核苷酸处;以及(5)小的插入和缺失突变的修复是不可能的。

[0011] 因此,开发能够灵活地引入任何期望的单核苷酸变化和/或能够安装(install)碱基对插入或缺失(如,至少1、2、3、4、5、6、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100或更多个碱基对插入或缺失)和/或能够以高特异性和高效率改变或修饰靶位点处的核苷酸序列将大大扩展基于CRISPR的基因组编辑技术的范围和治疗潜力。

[0012] 发明概述

[0013] 本发明描述了用于基因组编辑的新平台,称为"多瓣引导编辑(multi-flap prime editing)"(包括,例如,"双瓣(dual-flap)引导编辑"和"四瓣(quadruple-flap)引导编辑"),并且代表了"引导编辑"或"经典引导编辑"的创新进展,如本发明人在Anzalone,A.V.et al.Search-and-replace genome editing without double-链breaks或donor DNA.Nature 576,149-157(2019)中所述,该文献通过引用并入本文。然而不同的实施方案中的经典引导编辑在切口位点处聚合单个3′瓣,其被整合到同一链上的靶标核酸中,本文所述多瓣引导编辑系统涉及不同的构建体、系统和方法,它们在各个方案中产生位于不同链上的一对或多对3′瓣,后者形成包含期望的编辑的双链体,并且整合到靶标核酸分子中,例如,在基因组中的特定基因座处或编辑位点处。在各个方面,一对或多对3′瓣形成双链体,这是因为它们包含反向互补序列,这些序列一旦由本文所述引导编辑器产生就彼此退火。双链体通过细胞驱动机制整合到靶位点中,自然地替代位于相邻切口位点之间的内源性双链体序列。在某些实施方案中,新的双链体序列可以在一个或多个位置处引入(例如,在相邻的基因组基因座之间或在两个不同的染色体位置上),并且可以包含一个或多个感兴趣的序列,例如,蛋白质编码序列、肽编码序列或RNA编码序列。在一个实施方案中,通过多瓣引导编辑系统建立的新的双链体序列可以包含重组酶位点,例如,Bxb1重组酶attB(38bp)和/或attP(50bp)位点,或者由Hin重组酶、Gin重组酶、Tn3重组酶、β-six重组酶、CinH重组酶、ParA重组酶、γδ重组酶、ΦC31重组酶、TP901重组酶、TG1重组酶、φBT1重组酶、R4重组酶、φRV1重组酶、φFC1重组酶、MR11重组酶、A118重组酶、U153重组酶和gp29

重组酶、Cre重组酶、FLP重组酶、R重组酶、Lambda重组酶、HK101重组酶、HK022重组酶以及pSAM2重组酶识别的重组酶位点。

[0014] 发明人最近开发了能够插入、缺失或替换基因组DNA序列而不需要易错双链DNA断裂的引导编辑。引导编辑使用了与工程化引导编辑向导RNA(pegRNA)配对的工程化Cas9切口酶-逆转录酶融合蛋白(PE1或PE2),其既将Cas9引导至靶基因组位点也编码用于安装期望的编辑的信息。引导编辑通过多步编辑过程进行:1)Cas9结构域结合靶基因组DNA位点并使其产生切口,该位点由pegRNA间隔区序列来指定;2)逆转录酶结构域使用带切口的基因组DNA作为引物,使用pegRNA上的工程化延伸作为逆转录模板,启动编辑的DNA链的合成-这产生了含有编辑的DNA序列的单链3'瓣;3)细胞DNA修复通过以下方式拆分3'瓣中间体:经由编辑的3'瓣侵入进行5'瓣种类的置换,切除含有原始DNA序列的5'瓣,并且连接新的3'瓣以掺入编辑的DNA链,形成一条编辑链和一条未编辑链的异源双链体;和4)细胞DNA修复使用编辑链作为修复模板,替换异源双链体内的未编辑链,完成编辑过程。

[0015] 期望的编辑的有效掺入需要新合成的3'瓣含有与基因组DNA位点同源的一部分序列。这种同源能够使编辑的3'瓣与内源性DNA链(对应的5'瓣)竞争掺入DNA双链体中。由于编辑的3'瓣与内源性5'瓣相比将含有较低的序列同源性,预计竞争有利于5'瓣链。因此,引导编辑效率的潜在限制因素可能是内源性DNA的3'瓣侵入以及随后的5'瓣链置换和替换的效率。此外,成功的3'瓣侵入和5'瓣去除仅将编辑掺入到双链DNA基因组的一条链上。编辑的永久安装需要细胞DNA修复以使用编辑链作为模板来替换未编辑的互补DNA链。虽然可以通过使用二级sgRNA(PE3系统)在与编辑相邻的未编辑链中引入切口,使细胞相对于编辑链有利于未编辑链的置换(上述步骤4),但是该过程仍然依赖于DNA修复的第二阶段。这些DNA修复步骤对于需要长的5'和3'瓣中间体的平衡或者含有长的非同源区域(比如长的插入或长的缺失等)的编辑来说是特别没有效率。引导编辑的进一步开发将推动本领域。

[0016] 在多个方面,本说明书描述了多瓣引导编辑系统(包括,例如,双引导编辑系统和四引导编辑系统),其通过同时编辑两条DNA链解决了与瓣平衡和随后将编辑掺入未编辑的互补基因组DNA链相关的挑战。在双瓣引导编辑系统中,例如,使用两个pegRNA靶向基因组位点的相对链并且指导含有编辑的DNA序列的两个互补3'瓣的合成(图91)。与经典引导编辑不同,这对编辑的DNA链(3'瓣)不需要与5'瓣在内源性基因组DNA中直接竞争,因为互补编辑链可替代用于杂交。由于双链体的两条链都作为编辑的DNA合成,双瓣引导编辑系统省略了经典引导编辑需要的替换未编辑的互补DNA链的需求。相反,细胞DNA修复机制仅需要切除成对的5'瓣(原始基因组DNA)并且将成对的3'瓣(编辑的DNA)连接到基因座。因此,在新合成的DNA链中也不需要包含与基因组DNA同源的序列,允许新链的选择性杂交并且有利于含有最小基因组同源性的编辑。切割DNA的两条链的引导编辑器的核酸酶活性形式也可以用于加快原始DNA序列的去除。使用四个pegRNA的四瓣引导编辑系统提供了相似的优点。

[0017] 与经典引导编辑一样,多瓣引导编辑是通用且精确的基因组编辑方法,其使用与聚合酶联合作用的核酸可编程DNA结合蛋白("napDNAbp")(即,以融合蛋白的形式或在其它情况下与napDNAbp以反式提供)直接将新的遗传信息写入指定的DNA位点,其中引导编辑系统利用引导编辑(PE)向导RNA("PEgRNA")编程,该PEgRNA既指定靶位点,又以通过延伸(DNA或RNA)工程化至向导RNA(如,在向导RNA的5′或3′端处或内部部分中)的置换DNA链的形式为期望编辑的合成提供模板。含有期望编辑(如,单核碱基取代)的置换链与待编辑的靶位

点的内源性链共有相同的序列(除了它包括期望编辑)。通过DNA修复和/或复制机制,靶位点的内源性链被新合成的包含期望编辑的置换链替换。在某些情况下,可认为引导编辑是“搜索和置换”基因组编辑技术,因为本文所述引导编辑不仅搜索和定位待编辑的期望的靶位点,而且同时编码含有期望编辑的置换链,其得到安装代替相应靶位点内源性DNA链。

[0018] 本公开的多瓣引导编辑器部分地涉及以下发现:可利用或调整靶标引发(target-primed)的逆转录(TPRT)或“引导编辑”的机制进行基于CRISPR/Cas的精确基因组编辑,具有高效率和遗传可塑性(如,如图1A至1F的不同实施方案所描绘)。可移动的DNA元件如哺乳动物非LTR逆转录转座子和细菌II组内含子天然使用TPRT[28,29]。本发明人在本文中使用Cas蛋白-逆转录酶融合物或相关系统,利用向导RNA来靶向特定DNA序列,在靶位点处产生单链切口,并使用有切口的DNA作为引物进行与向导RNA一起整合的工程化逆转录酶模板的逆转录。然而,虽然该概念始于使用逆转录酶作为DNA聚合酶组件的引导编辑器,但本文所述的多瓣引导编辑器并不限于逆转录酶,而是可包括使用实际上任何DNA聚合酶。事实上,虽然本申请自始至终都涉及具有“逆转录酶”的多瓣引导编辑器,但在此提出逆转录酶只是可与多瓣引导编辑一起作用的DNA聚合酶的一种类型。因此,无论说明书何处提及“逆转录酶”,本领域普通技术人员都应理解,可使用任何合适的DNA聚合酶代替逆转录酶。因此,一方面,多瓣引导编辑器可包含Cas9(或等同napDNAbp),其被编程为通过将其与特化向导RNA(即,PEgRNA)缔合来靶向DNA序列,该特化向导RNA包含与靶DNA中前间隔序列(protospacer)序列的互补物退火的间隔区序列。特化向导RNA还以延伸形式包含新的遗传信息,该遗传信息编码包含期望的基因变化的DNA置换链,其用于替换靶位点处的相应内源性DNA链。为了将信息从PEgRNA转移到靶DNA,多瓣引导编辑机制包括在DNA的一条链中的靶位点处产生切口以暴露3’-羟基。然后可使用暴露的3’-羟基直接对靶位点中引发PEgRNA上的编码编辑的延伸的DNA聚合。在不同实施方案中,延伸——其为含有编辑的置换链的聚合提供模板——可由RNA或DNA形成。在RNA延伸的情况下,引导编辑器的聚合酶可以是RNA依赖性DNA聚合酶(如,逆转录酶)。在DNA延伸的情况下,引导编辑器的聚合酶可以是DNA依赖性DNA聚合酶。

[0019] 在经典引导编辑中,由本文公开的引导编辑器形成的新合成链(即,含有期望编辑的置换DNA链)除了包含期望的核苷酸变化(如,单核苷酸变化、缺失、或插入、或其组合),还与基因组靶序列同源(即,具有相同的序列)。还可将新合成(或置换)的DNA链称为单链DNA瓣(flap),它可竞争与互补的同源内源性DNA链的杂交,从而置换相应的内源性链。在某些实施方案中,该系统可与易错逆转录酶组合使用(如,作为与Cas9结构域的融合蛋白提供,或与Cas9结构域以反式提供)。易错逆转录酶可在单链DNA瓣的合成过程中引入变化。因此,在某些实施方案中,可利用易错逆转录酶将核苷酸变化引入靶DNA。根据与系统一起使用的易错逆转录酶,变化可以是随机或非随机的。

[0020] 在经典引导编辑中,杂合中间体(包括由与内源性DNA链杂交的逆转录酶合成的单链DNA瓣)的分解可包括去除所产生的内源性DNA的置换瓣(如,利用5’端DNA瓣核酸内切酶,FEN1),合成的单链DNA瓣与靶DNA的连接,以及由于细胞DNA修复和/或复制过程而导致期望的核苷酸变化的同化。由于提供了模板的DNA合成为任何核苷酸的修饰(包括插入和缺失)提供了单核苷酸精确度,因此这种方法的范围非常广泛,可预见可用于在基础科学和治疗学中的无数应用。

[0021] 在一些方面,说明书提供了一对引导编辑器,它们各自包含核酸可编程DNA结合蛋

白(napDNAbp)和DNA聚合酶。在一些实施方案中,每个引导编辑器均能够在延伸的引导RNA存在下通过靶向引导的逆转录来进行基因组编辑。

[0022] 在一些方面,说明书提供了一对引导编辑器,它们各自包含核酸可编程DNA结合蛋白(napDNAbp)和DNA聚合酶,其中DNA聚合酶相对于napDNAbp以反式提供。在不同的实施方案中,每个引导编辑器均能够在延伸的向导RNA存在下通过靶向引导的逆转录进行基因组编辑。

[0023] 在一些方面,说明书提供了一对引导编辑器,它们各自包含核酸可编程DNA结合蛋白(napDNAbp)和逆转录酶。在不同的实施方案中,每个引导编辑器均能够在延伸的向导RNA存在下通过靶向引导的逆转录进行基因组编辑。

[0024] 在一些方面,说明书提供了一对引导编辑器,它们各自包含核酸可编程DNA结合蛋白(napDNAbp)和逆转录酶,其中逆转录酶相对于napDNAbp以反式提供。在不同的实施方案中,每个引导编辑器均能够在延伸的向导RNA存在下通过靶向引导的逆转录进行基因组编辑。

[0025] 在某些实施方案中,napDNAbp具有切口酶活性。napDNAbp也可以是Cas9蛋白或其功能等同物,如核酸酶活性Cas9、无核酸酶活性Cas9(dCas9)或Cas9切口酶(nCas9)。

[0026] 在某些实施方案中,napDNAbp选自:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c和Argonaute蛋白,并且任选地具有切口酶活性。

[0027] 在其他实施方案中,双引导编辑器中的每个引导编辑器在与延伸的向导RNA复合时能够结合靶DNA序列。

[0028] 在其他实施方案中,靶DNA序列包含靶链和互补的非靶链。

[0029] 在其他实施方案中,与延伸的向导RNA复合的引导编辑器的结合形成R-环。R-环可包含(i)包含延伸的向导RNA和靶链的RNA-DNA杂合体,和(ii)互补的非靶链。

[0030] 在其他实施方案中,将互补的非靶链产生切口以形成具有游离3'端的逆转录酶引发序列。

[0031] 在不同实施方案中,延伸的向导RNA包含(a)向导RNA和(b)在向导RNA的5'或3'端处或在向导RNA的分子内位置中的RNA延伸。RNA延伸可包含(i)包含期望的核苷酸变化的逆转录模板序列,(ii)逆转录引物结合位点,和(iii)任选的接头序列。在不同实施方案中,逆转录模板序列可编码与邻近切口位点的内源性DNA序列互补的单链DNA瓣,其中单链DNA瓣包含期望的核苷酸变化。

[0032] 在不同实施方案中,RNA延伸的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、或至少25个核苷酸。

[0033] 在其他实施方案中,单链DNA瓣可与邻近切口位点的内源性DNA序列杂交,从而安装期望的核苷酸变化。在其他实施方案中,单链DNA瓣置换邻近切口位点且具有游离5'端的内源性DNA序列。在某些实施方案中,置换的具有5'端的内源性DNA被细胞切除。

[0034] 在不同实施方案中,单链DNA瓣的细胞修复导致安装期望的核苷酸变化,从而形成期望的产物。

[0035] 在不同的其他实施方案中，期望的核苷酸变化安装在PAM序列的约-4至+10之间的编辑窗口中。

[0036] 在其他实施方案中，期望的核苷酸变化安装在切口位点的约-5至+5之间、或切口位点的约-10至+10之间、或切口位点的约-20至+20之间，或切口位点的约-30至+30之间，或切口位点的约-40至+40之间，或约-50至+50之间切口位点的约-60至+60之间，或切口位点的约-70至+70之间，或切口位点的约-80至+80之间，或切口位点的约-90至+90之间，或切口位点的约-100至+100之间，或切口位点的约-200至+200之间的编辑窗口中。

[0037] 在不同实施方案中，双引导编辑器的napDNAbp各自包含SEQ ID NO:18的氨基酸序列。在不同的其他实施方案中，napDNAbp包含与SEQ ID NO:26-39、42-61、75-76、126、130、137、141、147、153、157、445、460、467和482-487(Cas9)；SEQ ID NO:77-86(CP-Cas9)；SEQ ID NO:18-25和87-88(SpCas9)；和SEQ ID NO:62-72(Cas12)中的任一氨基酸序列至少80％、85％、90％、95％、98％或99％相同的氨基酸序列。

[0038] 在其他实施方案中，所公开的引导编辑器和/或双引导编辑器的组合物的逆转录酶可包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159、235、454、471、516、662、700-716、739-742和766中的任一氨基酸序列。在其他实施方案中，逆转录酶可包含与SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159、235、454、471、516、662、700-716、739-742和766中的任一氨基酸序列至少80％、85％、90％、95％、98％或99％相同的氨基酸序列。这些序列可以是如来自逆转录病毒或逆转录转座子的天然存在的逆转录酶序列，这些序列可以是重组的。

[0039] 在不同的其他实施方案中，本文公开的双引导编辑器的引导编辑器可包含各种结构构造。例如，在其中引导编辑器作为融合蛋白提供的实施方案中，双引导编辑器融合蛋白中的每一个可包含结构NH_2-[napDNAbp]-[逆转录酶]-COOH；或NH_2-[逆转录酶]-[napDNAbp]-COOH，其中每个“]-[”情况表示存在任选的接头序列。

[0040] 在不同实施方案中，接头序列包含SEQ ID NO:127、165-176、446、453和767-769的氨基酸序列，或者与SEQ ID NO:127、165-176、446、453和767-769中的任一接头氨基酸序列至少80％、85％、或90％、或95％、或99％相同的氨基酸序列。

[0041] 在不同实施方案中，掺入靶DNA中的期望的核苷酸变化可以是单核苷酸变化(如，转换或颠换)、一个或多个核苷酸的插入、或者一个或多个核苷酸的缺失。

[0042] 在某些情况下，插入的长度为至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少200、至少300、至少400、或至少500个核苷酸。

[0043] 在某些其他情况下，缺失的长度为至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少200、至少300、至少400、或至少500个核苷酸。

[0044] 另一方面，本公开提供包含向导RNA和至少一个RNA延伸的延伸的向导RNA。RNA延伸可位于向导RNA的3’端。在其他实施方案中，RNA延伸可位于向导RNA的5’端。在其他实施方案中，RNA延伸可位于向导RNA的分子内位置，然而，优选地，延伸部分的分子内定位不破

坏前间隔序列的功能。

[0045] 在不同实施方案中，延伸的向导RNA能够结合napDNAbp并将napDNAbp引导至靶DNA序列。靶DNA序列可包含靶链和互补的非靶链，其中向导RNA与靶链杂交以形成RNA-DNA杂合体和R-环。

[0046] 在延伸的向导RNA的不同实施方案中，至少一个RNA延伸可包含逆转录模板序列。在不同的其他实施方案中，RNA延伸可进一步包含逆转录引物结合位点。在更进一步的实施方案中，RNA延伸可包含将RNA延伸连接至向导RNA的接头或间隔区序列。

[0047] 在不同实施方案中、RNA延伸的长度可为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少150个核苷酸、至少200个核苷酸、至少300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0048] 在其他实施方案中，逆转录模板序列的长度为至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少长度为300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0049] 在其他实施方案中，其中逆转录引物结合位点序列的长度为至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0050] 在其他实施方案中，任选的接头或间隔区序列的长度为至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少为300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0051] 在延伸的向导RNA的不同实施方案中，逆转录模板序列可编码与邻近切口位点的内源性DNA序列互补的单链DNA瓣，其中单链DNA瓣包含期望的核苷酸变化。单链DNA瓣可在切口位点置换内源性单链DNA。在切口位点处置换的内源性单链DNA可具有5'端并形成内源

性瓣，其可被细胞切除。在不同实施方案中，5'端内源性瓣的切除可有助于驱动产物形成，因为去除5'端内源性瓣促进单链3'DNA瓣与相应互补DNA链的杂交，以及掺入或同化由单链3'DNA瓣携带到靶DNA中的期望的核苷酸变化。

[0052] 在延伸的向导RNA的不同实施方案中，单链DNA瓣的细胞修复导致安装期望的核苷酸变化，从而形成期望的产物。

[0053] 在某些实施方案中，PEgRNA包含SEQ ID NO:101-104、181-183、223-234、237-244、277、324-330、332、334、336、338、340、342、344、346、348、350、352、354、356、358、360、362、364、366、368、394、429-442、499-505、641-649、678-692、735-736、757-761、776-777、2997-3103、3113-3121、3305-3455、3479-3493、3522-3540、3549-3556、3628-3698、3755-3810、3874、3890-3901、3905-3911、3913-3929和3972-3989的核苷酸序列，或者与SEQ ID NO:101-104、181-183、223-234、237-244、277、324-330、332、334、336、338、340、342、344、346、348、350、352、354、356、358、360、362、364、366、368、394、429-442、499-505、641-649、678-692、735-736、757-761、776-777、2997-3103、3113-3121、3305-3455、3479-3493、3522-3540、3549-3556、3628-3698、3755-3810、3874、3890-3901、3905-3911、3913-3929和3972-3989中任何一个具有至少85％、或至少90％、或至少95％、或至少98％、或至少99％序列同一性的核苷酸序列。

[0054] 在本发明的又一方面，本说明书提供了包含本文所述引导编辑器和上述任何延伸的向导RNA的复合物。

[0055] 在本发明的其他方面，本说明书提供了包含napDNAbp和延伸的向导RNA的复合物。napDNAbp可以是Cas9切口酶，或可以是SEQ ID NO:42-57(Cas9切口酶)和65(AsCas12a切口酶)的氨基酸序列，或与SEQ ID NO:42-57(Cas9切口酶)和65(AsCas12a切口酶)中的任一氨基酸序列至少80％、85％、90％、95％、98％或99％相同的氨基酸序列。

[0056] 在涉及复合物的不同实施方案中，延伸的向导RNA能够将napDNAbp引导至靶DNA序列。在不同实施方案中，逆转录酶可以反式提供，即由与复合物本身不同的来源提供。例如，通过引入单独编码逆转录酶的独立载体向具有复合物的同一细胞提供逆转录酶。

[0057] 在另一方面，本公开提供了包含第一和第二引导编辑器复合物的系统，每一个复合物均包含引导编辑器和引导编辑向导RNA(PEgRNA)。在一些实施方案中，每个引导编辑器包核酸可编程DNA结合蛋白(napDNAbp)和具有RNA依赖性DNA聚合酶活性的多肽，并且每个PEgRNA包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。在某些实施方案中，每个DNA合成模板编码包含编辑部分的单链DNA序列。编码的两个单链DNA序列可以彼此互补并且形成整合到待编辑的靶位点中的双链体。在一些实施方案中，编码的两个单链DNA序列可以包括彼此互补的区域。在某些实施方案中，编码的两个单链DNA序列可以包括彼此互补的区域，该区域的长度为至少2bp、至少3bp、至少4bp、至少5bp、至少10bp、至少20bp、至少30bp、至少40bp、至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp或至少1000bp。在一些实施方案中，引导编辑器作为融合蛋白提供。在某些实施方案中，引导编辑器的组分(即，napDNAbp和具有RNA依赖性DNA聚合酶活性的多肽)以反式提供。

[0058] 在另一方面，本公开提供了包含第一、第二、第三和第四引导编辑器复合物的系统，每个复合物包含引导编辑器和引导编辑向导RNA(PEgRNA)。在一些实施方案中，每个引

导编辑器包含核酸可编程DNA结合蛋白(napDNAbp)和具有RNA依赖性DNA聚合酶活性的多肽,并且每个PEgRNA包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。在某些实施方案中,每个DNA合成模板编码包含编辑部分的单链DNA序列。编码的两个单链DNA序列可以彼此互补并且形成整合到待编辑的位点中的双链体。在一些实施方案中,编码的两个单链DNA序列可以包括彼此互补的区域。在某些实施方案中,编码的两个单链DNA序列可以包括彼此互补的区域,该区域的长度为至少2bp、至少3bp、至少4bp、至少5bp、至少10bp、至少20bp、至少30bp、至少40bp、至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp或至少1000bp。在一些实施方案中,引导编辑器作为融合蛋白提供。在某些实施方案中,引导编辑器的组分(即,napDNAbp和具有RNA依赖性DNA聚合酶活性的多肽)以反式提供。

[0059] 在一些实施方案中,每个napDNAbp是Cas9结构域或其变体。在一些实施方案中,每个napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域或Cas9切口酶结构域或其变体。在某些实施方案中,每个napDNAbp独立地选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c和Argonaute,并且任选地具有切口酶活性。在不同的实施方案中,每个napDNAbp包含SEQ ID NO:2-65中任何一个的氨基酸序列或与SEQ ID NO:2-65中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[0060] 在一些实施方案中,包含RNA依赖性DNA聚合酶活性的多肽是逆转录酶。在某些实施方案中,包含RNA依赖性DNA聚合酶活性的多肽包含SEQ ID NO:37、68-79、82-98、81、98和110中任何一个的氨基酸序列或与SEQ ID NO:37、68-79、82-98、81、98和110中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[0061] 在一些实施方案中,每个引导编辑器可以包含接合napDNAbp和逆转录酶的接头。在某些实施方案中,接头包含SEQ ID NO:119-128中任何一个的氨基酸序列或与SEQ ID NO:119-128中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。每个接头的长度可以为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

[0062] 在一些实施方案中,每个PEgRNA可以独立地包含SEQ ID NO:192-203中任何一个的核苷酸序列或与SEQ ID NO:192-203中至少一个具有至少80%、85%、90%、95%或99%序列同一性的核苷酸序列。

[0063] 在不同的实施方案中,每个PEgRNA的每个间隔区序列可以结合至双链DNA序列的邻近于待编辑的位点的特定结合位点。在一些实施方案中,一个引导编辑器复合物的间隔区序列与双链DNA序列的一条链的结合以及另一个引导编辑器复合物的间隔区序列与双链DNA序列的相对链的结合导致两条DNA链在每个链上靠近PAM序列的切口位点处的切口。

[0064] 在另一方面,本公开提供了多核苷酸。在一些实施方案中,多核苷酸可以编码本文所述任何复合物。在某些实施方案中,多核苷酸可以编码本文所述任何PEgRNA。

[0065] 另一方面,本说明书提供了多核苷酸。在某些实施方案中,多核苷酸可编码本文公开的任何引导编辑器。在某些其他实施方案中,多核苷酸可编码本文公开的任何napDNAbps。在更进一步的实施方案中,多核苷酸可编码本文公开的任何逆转录酶。在其他实施方案中,多核苷酸可编码本文公开的任何延伸的向导RNA、任何逆转录模板序列、或任

何逆转录引物位点、或任何任选的接头序列。

[0066] 在其他方面，本说明书提供了包含本文所述多核苷酸的载体。因此，在某些实施方案中，载体包含用于编码包含napDNAbp和逆转录酶的引导编辑器的多核苷酸(即，作为融合蛋白，或者以反式表达)。在某些实施方案中，载体包含用于编码本文所述任何复合物的多核苷酸。在其他实施方案中，载体包含分别编码napDNAbp和逆转录酶的多核苷酸。在其他实施方案中，载体可包含编码延伸的向导RNA的多核苷酸。在不同实施方案中，载体可包含在相同或不同载体上的编码napDNAbp、逆转录酶和延伸的向导RNA的一个或多个多核苷酸。在一些实施方案中，载体包含用于编码本文所述任何pegRNA的多核苷酸。

[0067] 在其他方面，本说明书提供了包含如本文所述引导编辑器和延伸的向导RNA的细胞。细胞可利用包含引导编辑器、napDNAbp、逆转录酶和延伸的向导RNA的载体转化。这些遗传元件可包含在同一载体或不同载体上。细胞可以用编码本文所公开的任何系统、复合物和/或pegRNA的多核苷酸或者包含编码本文所公开的任何系统、复合物和/或pegRNA的多核苷酸的载体来转化。

[0068] 另一方面，本说明书提供了药物组合物。在某些实施方案中，药物组合物包含napDNAbp、引导编辑器、逆转录酶和延伸的向导RNA中的一种或多种。在某些实施方案中，药物组合物包含本文所述任何系统和/或复合物。在某些实施方案中，药物组合物包含本文所述任何引导编辑器、系统或复合物以及药学上可接受的赋形剂。在其他实施方案中，药物组合物包含本文所述任何延伸的向导RNA和药学上可接受的赋形剂。在其他实施方案中，药物组合物包含本文所述任何延伸的向导RNA组合本文所述任何引导编辑器和药学上可接受的赋形剂。在其他实施方案中，药物组合物包含编码napDNAbp、引导编辑器、逆转录酶和延伸的向导RNA中的一种或多种的任何多核苷酸序列，或者本文所述的任何载体。在其他实施方案中，本文公开的各种组分可分离成一种或多种药物组合物。例如，第一药物组合物可包含融合蛋白或napDNAbp，第二药物组合物可包含逆转录酶，并且第三药物组合物可包含延伸的向导RNA。

[0069] 又一方面，本公开提供了试剂盒。在一个实施方案中，试剂盒包括编码一种或多种组分的一种或多种多核苷酸，所述组分包括引导编辑器、napDNAbp、逆转录酶和延伸的向导RNA。试剂盒还可包含载体、细胞和分离的多肽制剂，包括本文公开的任何引导编辑器、napDNAbp或逆转录酶。

[0070] 又一方面，本公开提供了使用所公开的物质组合物的方法，包括使用本文所述任何系统在靶位点同时编辑双链DNA序列的两条互补链的方法。在一些实施方案中，该方法包括将双链DNA序列与本文公开的任何系统接触。

[0071] 一方面，本公开提供了以下方法，该方法包括将双链DNA序列在靶位点处与第一和第二引导编辑器复合物接触，每个复合物包含引导编辑器和引导编辑向导RNA(PEgRNA)。在一些实施方案中，每个引导编辑器包含核酸可编程DNA结合蛋白(napDNAbp)和具有RNA依赖性DNA聚合酶活性的多肽，并且每个PEgRNA包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。在一些实施方案中，每个引导编辑器作为融合蛋白提供。在一些实施方案中，引导编辑器的组分以反式提供。在某些实施方案中，每个DNA合成模板编码包含编辑部分的的单链DNA序列。编码的两个单链DNA序列可以彼此互补并且形成双链体，该双链体整合到待编辑的靶位点中。引导编辑器复合物的各元件可以包括本文公开的系统的任何实施方案。

[0072] 另一方面，本公开提供了以下方法，该方法包括将双链DNA序列在靶位点处与第一、第二、第三和第四引导编辑器复合物接触，每个复合物包含引导编辑器和引导编辑向导RNA(PEgRNA)。在一些实施方案中，每个引导编辑器包含核酸可编程DNA结合蛋白(napDNAbp)和具有RNA依赖性DNA聚合酶活性的多肽，并且每个PEgRNA包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。在一些实施方案中，每个引导编辑器作为融合蛋白提供。在一些实施方案中，引导编辑器的组分以反式提供。在某些实施方案中，每个DNA合成模板编码单链DNA序列。编码的两个单链DNA序列可以彼此互补并且形成双链体，该双链体整合到待编辑的靶位点中。引导编辑器复合物的各个元件可以包括本文公开的系统的任何实施方案。

[0073] 在一些实施方案中，本文提供的方法允许靶标DNA序列的倒置。在一些实施方案中，第一DNA合成模板编码的第一单链DNA序列和第二DNA合成模板编码的第二单链DNA序列位于靶标DNA序列的相对端，并且第三DNA合成模板编码的第三单链DNA序列和第四DNA合成模板编码的第四单链DNA序列位于同一靶标DNA序列的相对端。

[0074] 在一些实施方案中，本文提供的方法还包括提供环状DNA供体。在某些实施方案中，第一DNA合成模板编码的第一单链DNA序列和第三DNA合成模板编码的第三单链DNA序列位于靶标DNA序列的相对端，并且第二DNA合成模板编码的第二单链DNA序列和第四DNA合成模板编码的第四单链DNA序列位于环状DNA供体上。在一些实施方案中，环状DNA供体的介于第二单链DNA序列和第四单链DNA序列之间的部分替代了介于第一单链DNA序列和第三单链DNA序列之间的靶标DNA序列。

[0075] 在一些实施方案中，本文提供的方法允许靶标DNA序列从第一核酸分子(例如，第一染色体)易位到第二核酸分子(例如，第二染色体)。在一些实施方案中，第一DNA合成模板编码的第一单链DNA序列和第三单链DNA合成模板编码的第三单链DNA序列位于第一核酸分子上，并且第二DNA合成模板编码的第二单链DNA序列和第四DNA合成模板编码的第四单链DNA序列位于第二核酸分子上。在某些实施方案中，第一单链DNA序列和第三单链DNA序列之间第一核酸分子的部分掺入到第二核酸分子中。在某些实施方案中，第二单链DNA序列和第四单链DNA序列之间第二核酸分子的部分掺入到第一核酸分子中。

[0076] 另一方面，本公开提供了用于多瓣引导编辑的PEgRNA对。在一些实施方案中，该对包含第一PEgRNA和第二PEgRNA，并且每个PEgRNA独立地包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。在某些实施方案中，每个DNA合成模板编码单链DNA序列。在不同的实施方案中，多瓣引导编辑器与PEgRNA对相关使用，后者使单独的引导编辑器白相靶位点的一侧，其中该PEgRNA对各自编码3’核酸瓣，这些瓣包含彼此反向互补的核酸。在不同的实施方案中，包含反向互补序列的3’瓣可以彼此重组以形成双链体，该双链体包含期望的编辑或通过PEgRNA编辑的核酸序列。双链体随后通过替换相邻切口位点之间的相应内源性双链体整合到靶位点中。

[0077] 另一方面，本公开提供了多个PEgRNA，其用于多瓣引导编辑。在一些实施方案中，该多个包含第一、第二、第三和第四PEgRNA。在一些实施方案中，四个PEgRNA中的每个独立地包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。在某些实施方案中，每个DNA合成模板编码单链DNA序列。编码的两个单链DNA序列可以彼此互补。

[0078] 在不同方面，本公开提供了编码所述PEgRNA对或多个PEgRNA的多核苷酸。在某些

方面，本公开提供了编码多核苷酸的载体，该多核苷酸编码本文所述的PEgRNA对或多个PEgRNA中的任何一个。又一方面，本公开提供了包含编码多核苷酸的载体的细胞，多核苷酸编码本文所述的PEgRNA对或多个PEgRNA中的任何一个。在其他方面，本公开提供了药物组合物，其包含本文所述的PEgRNA对或多个PEgRNA中的任何一个、编码本文所述的PEgRNA对或多个PEgRNA中的任何一个的载体或者包含编码本文所述的PEgRNA对或多个PEgRNA中的任何一个的载体的细胞。在某些实施方案中，药物组合物包含药物赋形剂。

[0079] 在一个实施方案中，方法涉及在双链DNA序列中安装期望的核苷酸变化的方法。方法首先包括使双链DNA序列与包含引导编辑器和延伸的向导RNA的复合物接触，其中引导编辑器包含napDNAbp和逆转录酶，其中延伸的向导RNA包含含有期望的核苷酸变化的逆转录模板序列。在一些实施方案中，每个引导编辑器作为融合蛋白提供。在一些实施方案中，引导编辑器的组分以反式提供。接着，方法包括对非靶链上的双链DNA序列产生切口，从而生成具有3'端的游离单链DNA。然后方法包括将游离单链DNA的3'端与逆转录模板序列杂交，从而引发逆转录酶结构域。然后，方法涉及从3'端聚合DNA的链，从而生成包含期望的核苷酸变化的单链DNA瓣。然后，方法涉及用单链DNA瓣置换邻近切割位点的内源性DNA链，从而在双链DNA序列中安装期望的核苷酸变化。

[0080] 在其他实施方案中，本公开提供了用于在靶基因座处的DNA分子核苷酸序列中引入一个或多个变化的方法，包括将DNA分子与核酸可编程DNA结合蛋白(napDNAbp)和将napDNAbp靶向靶基因座的向导RNA接触，其中所述向导RNA包含含有至少一个期望的核苷酸变化的逆转录酶(RT)模板序列。接着，方法涉及在靶基因座处的DNA链中形成暴露的3'端，然后将暴露的3'端与RT模板序列杂交以引发逆转录。接着，通过逆转录酶基于RT模板序列合成或聚合含有至少一个期望的核苷酸变化的单链DNA瓣。最后，将至少一个期望的核苷酸变化掺入相应的内源性DNA，从而在靶基因座处的DNA分子核苷酸序列中引入一个或多个变化。

[0081] 在其他实施方案中，本公开提供了通过靶标引发的逆转录在靶基因座处的DNA分子核苷酸序列中引入一个或多个变化的方法，该方法包括：(a)用以下接触靶基因座处的DNA分子：(i)包含核酸可编程DNA结合蛋白(napDNAbp)和逆转录酶的引导编辑器，和(ii)包含含有期望的核苷酸变化的RT模板的向导RNA；(b)对RT模板进行靶标引发的逆转录，以生成包含期望的核苷酸变化的单链DNA；以及(c)通过DNA修复和/或复制过程将期望的核苷酸变化掺入靶基因座处的DNA分子中。

[0082] 在某些实施方案中，置换内源性DNA链的步骤包括：(i)将单链DNA瓣与邻近切割位点的内源性DNA链杂交以产生序列错配；(ii)切除内源性DNA链；以及(iii)修复错配以在两条DNA链中形成包含期望的核苷酸变化的期望的产物。

[0083] 在不同实施方案中，期望的核苷酸变化可以是单核苷酸取代(如，转换或颠换变化)、缺失或插入。例如，期望的核苷酸变化可以是(1)G至T取代，(2)G至A取代，(3)G至C取代，(4)T至G取代，(5)T至A取代，(6)T至C取代，(7)C至G取代，(8)C至T取代，(9)C至A取代，(10)A至T取代，(11)A至G取代，或(12)A至C取代。

[0084] 在其他实施方案中，期望的核苷酸变化可转换(1)G:C碱基对至T:A碱基对，(2)G:C碱基对至A:T碱基对，(3)G:C碱基对至C:G碱基对，(4)T:A碱基对至G:C碱基对，(5)T:A碱基对至A:T碱基对，(6)T:A碱基对至C:G碱基对，(7)C:G碱基对至G:C碱基对，(8)C:G碱基对至

T:A碱基对,(9)C:G碱基对至A:T碱基对,(10)A:T碱基对至T:A碱基对,(11)A:T碱基对至G:C碱基对,或(12)A:T碱基对至C:G碱基对。

[0085] 在其他实施方案中,方法引入期望的核苷酸变化为插入。在某些情况下,插入的长度为至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少200、至少300、至少400、或至少500个核苷酸。

[0086] 在其他实施方案中,方法引入期望的核苷酸变化为缺失。在某些其他情况下,缺失的长度为至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11,至少12,至少13,至少14,至少15,至少16、至少17、至少18、至少19、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少200、至少300、至少400、或至少500个核苷酸。

[0087] 在不同实施方案中,期望的核苷酸变化校正疾病相关基因。疾病相关基因可与选自下组的单基因病症(monogenetic disorder)相关:腺苷脱氨酶(ADA)缺乏症;α-1抗胰蛋白酶缺乏症;囊性纤维化;杜氏肌营养不良;半乳糖血症;血色素沉积症(hemochromatosis);亨廷顿氏病;枫糖尿病;马凡(Marfan)综合征;1型神经纤维瘤病;先天性厚甲症;苯丙酮尿症;重度联合免疫缺陷;镰状细胞病;史-李-欧综合征;和泰-萨二氏病(Tay-Sachs)病。在其他实施方案中,疾病相关基因可与选自下组的多基因(polygenic)病症相关:心脏病;高血压;阿尔茨海默氏病;关节炎;糖尿病;癌症;以及肥胖。

[0088] 本文公开的方法可涉及具有napDNAbp的融合蛋白,所述napDNAbp为核酸酶死亡Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9。在其他实施方案中,napDNAbp和逆转录酶不作为单一融合蛋白编码,而是可在单独的构建体中提供。因此,在一些实施方案中,逆转录酶可相对于napDNAbp以反式提供(而不是通过融合蛋白的方式)。

[0089] 在涉及方法的不同实施方案中,napDNAbp可包含SEQ ID NO:26-61、75-76、126、130、137、141、147、153、157、445、460、467、和482-487(Cas9);(SpCas9);SEQ ID NO:77-86(CP-Cas9);SEQ ID NO:18-25和87-88(SpCas9);和SEQ ID NO:62-72(Cas12)的氨基酸序列。napDNAbp还可包含与SEQ ID NO:26-61、75-76、126、130、137、141、147、153、157、445、460、467和482-487(Cas9);(SpCas9);SEQ ID NO:77-86(CP-Cas9);SEQ ID NO:18-25和87-88(SpCas9);和SEQ ID NO:62-72(Cas12)中的任一氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

[0090] 在涉及方法的不同实施方案中,逆转录酶可包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159、235、454、471、516、662、700-716、739-742和766的氨基酸序列。逆转录酶还可包含与SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159、235、454、471、6216、132、139、143、149、154、159、235、454,471、516、662,700-716、739-742、和766中的任一氨基酸序列至少80%、85%、90%、95%、98%、或99%相同的氨基酸序列。

[0091] 方法可涉及使用包含以下的核苷酸序列的PEgRNA:SEQ ID NO:101-104、181-183、223-234、237-244、277、324-330、332、334、336、338、340、342、344、346、348、350、352、354、356、358、360、362、364、366、368、394、429-442、499-505、641-649、678-692、735-736、757-

761、776-777、2997-3103、3113-3121、3305-3455、3479-3493、3522-3540、3549-3556、3628-3698、3755-3810、3874、3890-3901、3905-3911、3913-3929和3972-3989或者与其具有至少80％、或至少85％、或至少90％、或至少95％、或至少99％序列同一性的核苷酸序列。方法可包括使用延伸的向导RNA，其在3'端包含RNA延伸，其中RNA延伸包含逆转录模板序列。

[0092] 方法可包括使用延伸的向导RNA，其在5'端包含RNA延伸，其中RNA延伸包含逆转录模板序列。

[0093] 方法可包括使用延伸的向导RNA，其在向导RNA的分子内位置包含RNA延伸，其中RNA延伸包含逆转录模板序列。

[0094] 方法可包括使用延伸的向导RNA，其具有长度为至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少200、至少300、至少400、或至少500个核苷酸的一个或多个RNA延伸。

[0095] 应当理解，前述概念和下面讨论的其他概念可以任何合适的组合进行排列，因为本公开在这方面不受限制。此外，当结合附图考虑时，根据不同非限制性实施方案的以下详细描述，本公开的其他优点和新颖特征将变得显而易见。

[0096] 附图简述

[0097] 以下附图构成本说明书的一部分并被包括在内以进一步阐述本公开的某些方面，通过参考这些附图中的一个或多个并结合本文所呈现的具体实施方案的详细描述，可更好地理解本公开。

[0098] 图1A提供使用包含与Cas9蛋白融合的逆转录酶的融合蛋白复合延伸的向导RNA将单核苷酸变化和/或插入和/或缺失引入DNA分子(如，基因组)的示例性过程的示意图。在该实施方案中，向导RNA在3'端延伸以包含逆转录酶模板序列。该示意图显示与Cas9切口酶融合的逆转录酶(RT)与向导RNA(gRNA)形成的复合物如何结合DNA靶位点并对邻近靶核苷酸的含PAM的DNA链产生切口。RT酶使用有切口的DNA作为由gRNA合成DNA的引物，其用作合成编码期望编辑的新DNA链的模板。所示的编辑过程可被称为靶标引发的逆转录编辑(TRT编辑)或等同地“引导编辑”。

[0099] 图1B提供与图1A相同的图示，除了引导编辑器复合物更普遍地表示为[napDNAbp]-[P]:PEgRNA或[P]-[napDNAbp]:PEgRNA，其中“P”指任何聚合酶(如逆转录酶)，“napDNAbp”指核酸可编程DNA结合蛋白(如，SpCas9)，并且“PEgRNA”指引导编辑向导RNA，“]-[”指可选的接头。如别处所述，例如图3A至3G所示，PEgRNA包含含有引物结合位点和DNA合成模板的5'延伸臂。虽然未显示，但预期PEgRNA的延伸臂(即，其包含引物结合位点和DNA合成模板)可以是DNA或RNA。该构造中考虑的特定聚合酶取决于DNA合成模板的性质。例如，如果DNA合成模板是RNA，则聚合酶可以是RNA依赖性DNA聚合酶(如，逆转录酶)。如果DNA合成模板是DNA，则聚合酶可以是DNA依赖性DNA聚合酶。

[0100] 图1C提供使用包含与Cas9蛋白融合的逆转录酶的融合蛋白复合延伸的向导RNA分子将单核苷酸变化和/或插入和/或缺失引入DNA分子(如，基因组)的示例性过程的示意图。在该实施方案中，向导RNA在5'端延伸以包含逆转录酶模板序列。该示意图显示与Cas9切口酶融合的逆转录酶(RT)与向导RNA(gRNA)形成的复合物如何结合DNA靶位点并对邻近靶核

苷酸的含PAM的DNA链产生切口。RT酶使用有切口的DNA作为由gRNA合成DNA的引物，其用作合成编码期望编辑的新DNA链的模板。所示的编辑过程可被称为靶标引发的逆转录编辑(TRT编辑)或等同地“引导编辑”。

[0101] 图1D提供与图1D相同的图示，除了引导编辑器复合物更普遍地表示为[napDNAbp]-[P]:PEgRNA或[P]-[napDNAbp]:PEgRNA，其中“P”是指任何聚合酶(如，逆转录酶)，“napDNAbp”是指核酸可编程DNA结合蛋白(如，SpCas9)，“PEgRNA”指引导编辑向导RNA，“]-[”指可选的接头。如别处所述，例如图3A至3G所示，PEgRNA包含含有引物结合位点和DNA合成模板的3’延伸臂。虽然未显示，但预期PEgRNA的延伸臂(即，其包含引物结合位点和DNA合成模板)可以是DNA或RNA。该构造中考虑的特定聚合酶取决于DNA合成模板的性质。例如，如果DNA合成模板是RNA，则聚合酶可以是RNA依赖性DNA聚合酶(如，逆转录酶)。如果DNA合成模板是DNA，则聚合酶可以是DNA依赖性DNA聚合酶。在不同实施方案中，PEgRNA可被工程化改造或合成以掺入基于DNA的DNA合成模板。

[0102] 图1E是描绘合成的DNA单链(其包含期望的核苷酸变化)如何被分解从而将期望的核苷酸变化掺入DNA中的示例性过程的示意图。如所示，接着合成编辑链(或“诱变链”)，与内源性链平衡，内源性链的瓣切割和连接导致通过内源性DNA修复和/或复制过程的作用分解错配的DNA双链体后掺入DNA编辑。

[0103] 图1F是显示可将“相反链切口”掺入图1E的分解方法以助于驱动形成期望产物相对于回复产物(reversionproduct)的示意图。在相反链切口产生中，将第二Cas9/gRNA复合物用于在与初始切口链相反的链上引入第二切口。这诱导内源性细胞DNA修复和/或复制过程，以优先置换非编辑链(即，包含第二切口位点的链)。

[0104] 图1G提供使用核酸可编程DNA结合蛋白(napDNAbp)复合延伸的向导RNA将单核苷酸变化和/或插入和/或缺失引入靶基因座的DNA分子(如，基因组)的示例性过程的另一示意图。该过程可被认为是引导编辑的实施方案。延伸的向导RNA包含在向导RNA的3’或5’端处或在向导RNA的分子内位置中的延伸。步骤(a)中，napDNAbp/gRNA复合物接触DNA分子，gRNA引导napDNAbp与靶基因座结合。步骤(b)中，在靶基因座处的DNA链之一(R环链、或含有PAM的链、或非靶DNA链、或前间隔序列链)引入切口(如，通过核酸酶或化学试剂)，从而在靶基因座处的一条链中产生可用的3’端。在某些实施方案中，在对应于R-环链的DNA链中产生切口，即未与向导RNA序列杂交的链。步骤(c)中，3’端DNA链与向导RNA的延伸部分相互作用以引发逆转录。在某些实施方案中，3’端DNA链与向导RNA的延伸部分上的特定RT引发序列杂交。步骤(d)中，引入逆转录酶，该酶从引发位点的3’端向向导RNA的3’端合成单链DNA。这形成包含期望的核苷酸变化(如，单碱基变化、插入、或缺失、或其组合)的单链DNA瓣。步骤(e)中，napDNAbp和向导RNA得到释放。步骤(f)和(g)涉及单链DNA瓣的分解，使得期望的核苷酸变化掺入靶基因座。该过程可向通过去除相应的5’内源性DNA瓣来形成期望的产物驱动，一旦3’单链DNA瓣侵入并与另一条链的互补序列杂交就会形成5’内源性DNA瓣。该过程也可向利用第二链切口产生形成产物驱动，如图1F所示。该过程会引入至少一个或多个以下基因变化：颠换、转换、缺失和插入。

[0105] 图1H是描绘利用本文所述引导编辑过程可能的基因变化类型的示意图。通过引导编辑可实现的核苷酸变化类型包括缺失(包括短和长的缺失)、单核苷酸变化(包括转换和颠换)、倒置、和插入(包括短和长的插入)。

[0106] 图1I是描绘由PE3b(PE3b＝PE2引导编辑器融合蛋白+PEgRNA+第二链切口产生向导RNA)示例的时序(temporal)第二链切口产生的示意图。时序第二链切口产生是第二链切口产生的变体,以促进期望编辑产物的形成。“时序”术语指非编辑链的第二链切口仅在期望编辑安装到编辑链之后才出现的事实。这避免了两条链同时出现切口导致双链DNA断裂。

[0107] 图1J至1K描绘本文预期的引导编辑的变体,其利用任何可编程核酸酶结构域如锌指核酸酶(ZFN)或转录激活因子样效应核酸酶(TALEN)替换napDNAbp(如,SpCas9切口酶)。因此,预期合适的核酸酶不一定需要由核酸靶向分子(如,向导RNA)“编程”,而是可通过定义DNA结合结构域的特异性来编程,例如尤其是核酸酶。正如利用napDNAbp部分的引导编辑,优选对此类替代的可编程核酸酶进行修饰,以便仅切割一条靶DNA链。换言之,可编程核酸酶应优选起切口酶的作用。一旦选择了可编程核酸酶(如,ZFN或TALEN),则可将附加功能工程化改造到系统中以使其按引导编辑样机制作用。例如,可编程核酸酶可通过对其偶联(如,通过化学接头)RNA或DNA延伸臂来修饰,其中延伸臂包含引物结合位点(PBS)和DNA合成模板。可编程核酸酶也可偶联(如,通过化学或氨基酸接头)到聚合酶,所述聚合酶的性质取决于延伸臂是DNA还是RNA。在RNA延伸臂的情况下,聚合酶可以是RNA依赖性DNA聚合酶(如,逆转录酶)。在DNA延伸臂的情况下,聚合酶可以是DNA依赖性DNA聚合酶(如,原核聚合酶,包括PolI、PolII或PolIII,或真核聚合酶,包括Pola、Polb、Polg、Pold、Pole或Polz)。该系统还可包括作为融合物添加至可编程核酸酶或以反式添加的其他功能来促进整体反应(如,(a)解旋酶解开切割位点处的DNA以形成具有可用3’端的切割链用作引物,(b)瓣核酸内切酶(如,FEN1)有助于去除切割链上的内源性链,以驱动反应向着用合成链置换内源性链,或(c)nCas9:gRNA复合物在相反链上形成第二位点切口,这可有助于通过非编辑链的有利细胞修复来驱动合成修复的整合)。以类似于利用napDNAbp的引导编辑的方式,可将这种具有其他情况下可编程核酸酶的复合物用于合成和然后永久地安装携带感兴趣的编辑的新合成的DNA置换链到DNA的靶位点中。

[0108] 图1L描绘在一个实施方案中通过引导编辑可编辑的靶DNA的解剖特征。靶DNA包含“非靶链”和“靶链”。靶链是与识别PAM位点的引导编辑器复合物的PEgRNA的间隔区退火的链(在这种情况下,NGG,其被基于SpCas9的经典引导编辑器识别)。靶链也可被称为“非PAM链”或“非编辑链”。相比之下,非靶链(即,包含NGG的前间隔序列和PAM序列的链)可被称为“PAM链”或“编辑链”。在不同实施方案中,PE复合物的切口位点位于PAM链的前间隔序列中(如,利用基于SpCas9的PE)。切口的定位将是形成PE的特定Cas9的特征。例如,利用基于SpCas9的PE,碱基3(相对于PAM序列的位点1的“-3”位点)和碱基4(相对于PAM序列的位点1的“-4”位点)之间的磷酸二酯键中的切口位点。前间隔序列中的切口位点形成游离的3’羟基,如下图所示,其与PEgRNA延伸臂的引物结合位点复合,并提供底物以通过PEgRNA延伸臂的DNA合成模板开始单链DNA编码的聚合。该聚合反应由PE融合蛋白的聚合酶(如,逆转录酶)在5’至3’方向催化。聚合在到达gRNA核心之前终止(如,通过包含聚合终止信号或二级结构,其起终止PE的聚合活性的作用),产生从有切口的PAM链的原始3’羟基延伸的单链DNA瓣。DNA合成模板编码与内源性5’端DNA单链同源的单链DNA,该内源性5’端DNA单链紧邻PAM链的切口位点并掺入期望的核苷酸变化(如,单碱基替换、插入、缺失、倒置)。期望编辑的位置可在PAM链上切口位点下游之后的任何位点,可包括位点+1、+2、+3、+4(PAM位点的起点)、+5(PAM位点的位点2)、+6(PAM位点的位点3)、+7、+8、+9、+10、+11、+12、+13、+14、+15、+16、+

17、+18、+19、+20、+21、+22、+23、+24、+25、+26、+27、+28、+29、+30、+31、+32、+33、+34、+35、+36、+37、+38、+39、+40、+41、+42、+43、+44、+45、+46、+47、+48、+49、+50、+51、+52、+53、+54、+55、+56、+57、+58、+59、+60、+61、+62、+63、+64、+65、+66、+67、+68、+69、+70、+71、+72、+73、+74、+75、+76、+77、+78、+79、+80、+81、+82、+83、+84、+85、+86、+87、+88、+89、+90、+91、+92、+93、+94、+95、+96、+97、+98、+99、+100、+101、+102、+103、+104、+105、+106、+107、+108、+109、+110、+111、+112、+113、+114、+115、+116、+117、+118、+119、+120、+121、+122、+123、+124、+125、+126、+127、+128、+129、+130、+131、+132、+133、+134、+135、+136、+137、+138、+139、+140、+141、+142、+143、+144、+145、+146、+147、+148、+149、或+150或更多(相对于切口位点的下游位点)。一旦3'端单链DNA(包含感兴趣的编辑)置换内源性5'端单链DNA,DNA修复和复制过程将导致永久安装PAM链上的编辑位置,然后校正存在于编辑位置的非PAM链上的错配。这样,编辑将延伸到靶DNA位置上的两条DNA链。应当理解,提及"编辑链"和"非编辑"仅旨在描绘PE机制中涉及的DNA链。"编辑链"是首先通过利用合成的含有期望编辑的3'端单链DNA置换在切口位点的紧邻下游的5'端单链DNA进行编辑的链。"非编辑"链是与编辑链配对的链,但其本身也通过修复和/或复制被编辑以与编辑链互补,特别是感兴趣的编辑。

[0109] 图1M描述显示靶DNA的解剖特征、引导编辑器复合物以及PEgRNA与靶DNA之间的相互作用的引导编辑机制。首先,包含具有聚合酶(如,逆转录酶)和napDNAbp(如,SpCas9切口酶,例如SpCas9,在HNH核酸酶结构域中具有失活突变(如,H840A)或在RuvC核酸酶结构域中具有失活突变(D10A))的引导编辑器与PEgRNA和具有待编辑的靶DNA的DNA复合。PEgRNA包含间隔区、gRNA核心(又名gRNA支架或gRNA骨架)(其与napDNAbp结合)和延伸臂。延伸臂可位于PEgRNA分子的3'端、5'端或PEgRNA分子内的某处。如所示,延伸臂位于PEgRNA的3'端。延伸臂在3'至5'方向包含引物结合位点和DNA合成模板(包括感兴趣的编辑和与紧邻PAM链的切口位点的5'端单链DNA同源的同源区域(即,同源臂)。如所示,一旦引入切口,从而在紧邻切口位点的上游产生游离的3'羟基,紧邻PAM链的切口位点上游的区域与被称为"引物结合位点"的延伸臂3'端处的互补序列退火,产生具有可用的3'羟基端的短的双链区域,这形成引导编辑器复合物的聚合酶的底物。聚合酶(如,逆转录酶)然后从3'羟基端至延伸臂端聚合DNA链。单链DNA的序列由DNA合成模板编码,其是被聚合酶"读取"以合成新DNA的延伸臂的部分(即,不包括引物结合位点)。这种聚合有效延伸初始切口位点的原始3'羟基端的序列。DNA合成模板编码单链DNA,该单链DNA不仅包含期望编辑,而且还包含与紧邻PAM链的切口位点下游的内源性单链DNA同源的区域。接着,编码的3'端单链DNA(即3'单链DNA瓣)置换紧邻PAM链的切口位点下游的相应同源的内源性5'端DNA单链,形成具有5'端单链DNA瓣的DNA中间体,该5'端单链DNA瓣被细胞移除(如,通过瓣核酸内切酶)。3'端单链DNA瓣(其与内源性5'端单链DNA瓣的互补序列退火)在去除5'DNA瓣后连接到内源性链。3'端单链DNA瓣中期望编辑,现在退火和连接,与互补链形成错配,其经历DNA修复和/或复制轮次,从而在两条链上永久安装期望编辑。

[0110] 图2显示可用于本文所述的引导编辑器的三种Cas复合物(SpCas9、SaCas9和LbCas12a)及其PAM、gRNA和DNA切割特征。该图显示涉及SpCas9、SaCas9和LbCas12a的复合物的设计。

[0111] 图3A至3F显示工程化5'引导编辑器gRNA(图3A)、3'引导编辑器gRNA(图3B)和分子内延伸(图3C)的设计。延伸的向导RNA(或延伸的gRNA)在本文中又被称为PEgRNA或"引导编

辑向导RNA”。图3D和图3E分别提供了3′和5′引导编辑器gRNA(PEgRNA)的其他实施方案。图3F显示3′端引导编辑器向导RNA与靶DNA序列之间的相互作用。图3A至3C的实施方案描绘了逆转录模板序列(即,或更广泛地被称为DNA合成模板,如所示,因为RT仅是可在引导编辑器的背景下使用的一类聚合酶),引物结合位点,和3’、5’和分子内版本的延伸部分中的可选接头序列的示例性排列,以及间隔区和核心区的一般排列。所公开的引导编辑过程不限于延伸的向导RNA的这些构造。图3D的实施方案提供本文预期的示例性PEgRNA的结构。PEgRNA包含以5′至3′方向排序的三个主要组成元件,即:间隔区、gRNA核心和3′端的延伸臂。延伸臂在5′至3′方向可进一步分为以下结构元件,即:引物结合位点(A)、编辑模板(B)和同源臂(C)。此外,PEgRNA可包含任选的3′端修饰区(e1)和任选的5′端修饰区(e2)。更进一步地,PEgRNA可包含在PEgRNA3′端的转录终止信号(未描绘)。这些结构元件在本文中进一步定义。PEgRNA结构的描述并不意味着是限制性的,而是包含元件排列的变化。例如,任选的序列修饰区(modifier)(e1)和(e2)可位于所示的任何其他区域之内或之间,并且并不限于位于3’和5’端。在某些实施方案中,PEgRNA可包含二级RNA结构,例如但不限于发夹、茎/环、趾环(toe loop)、RNA结合蛋白募集结构域(如,募集并结合MS2cp蛋白的MS2适体)。例如,这样的二级结构可位于间隔区、gRNA核心或延伸臂内,特别是位于e1和/或e2修饰区内。除二级RNA结构外,PEgRNA还可包含(如,在e1和/或e2修饰区域内)化学接头或poly(N)接头或尾,其中“N”可以是任何核碱基。在一些实施方案中(如,如图72(c)中所示),化学接头可起到防止sgRNA支架或核心逆转录的作用。此外,在某些实施方案中(如,参见图72(c)),延伸臂(3)可由RNA或DNA组成,和/或可包括一个或多个核碱基类似物(如,其可添加功能性,如温度弹性)。更进一步,延伸臂(3)的定向可以是自然的5′至3′方向,或在3′至5′方向(相对于整个PEgRNA分子的定向)以相反定向合成。还应注意,本领域普通技术人员将能够根据延伸臂的核酸材料(即,DNA或RNA)的性质选择合适的DNA聚合酶,用于引导编辑中,所述引导编辑中可作为与napDNAbp的融合物或作为单独的部分以反式实施,以合成含有期望编辑的期望模板编码的3′单链DNA瓣。例如,如果延伸臂是RNA,DNA聚合酶可以是逆转录酶或任何其他合适的RNA依赖性DNA聚合酶。然而,如果延伸臂是DNA,DNA聚合酶可以是DNA依赖性DNA聚合酶。在不同实施方案中,DNA聚合酶的提供可以反式,例如通过使用RNA-蛋白募集结构域(如,在PEgRNA上安装的MS2发卡(如,在e1或e2区或其他地方,以及与DNA聚合酶融合的MS2cp蛋白中,从而将DNA聚合酶共定位到PEgRNA)。还应注意,引物结合位点通常不形成DNA聚合酶(如,逆转录酶)用来编码所得到的包含期望编辑的3′单链DNA瓣的模板的一部分。因此,“DNA合成模板”的名称是指由DNA聚合酶用作模板来编码包含编辑的期望的3′单链DNA瓣的延伸臂(3)的区域或部分和与5’内源性单链DNA瓣同源的区域,其中5’内源性单链DNA瓣被引导编辑DNA合成的3’单链DNA链产物置换。在一些实施方案中,DNA合成模板包括“编辑模板”和“同源臂”,或一个或多个同源臂,例如在编辑模板之前和之后。编辑模板可以小到单核苷酸取代,也可以是插入或DNA倒置。此外,编辑模板还可包含缺失,这可通过编码包含期望的缺失的同源臂来工程化改造。在其他实施方案中,DNA合成模板还可包括e2区或其一部分。例如,如果e2区包含导致DNA聚合酶活性终止的二级结构,则在e2区的任何部分被实际编码到DNA中之前,DNA聚合酶的功能有可能终止。部分甚至全部e2区域也有可能被编码到DNA中。有多少e2被实际用作模板将取决于其组成以及该组成是否中断DNA聚合酶功能。

[0112] 图3E的实施方案提供本文预期的另一种PEgRNA结构。PEgRNA包含以5′到3′方向排序的三个主要组成元件，即：间隔区、gRNA核心和3′端的延伸臂。延伸臂在5′至3′方向可进一步分为以下结构元件，即：引物结合位点(A)、编辑模板(B)和同源臂(C)。此外，PEgRNA可包含任选的3′端修饰区(e1)和任选的5′端修饰区(e2)。更进一步地，PEgRNA可包含在PEgRNA的3′端的转录终止信号(未描绘)。这些结构元件在本文中进一步定义。PEgRNA结构的描述并不意味着是限制性的，而是包含元件排列的变化。例如，任选的序列修饰区(e1)和(e2)可位于所示的任何其他区域之内或之间，并且不限于位于3’和5’端。在某些实施方案中，PEgRNA可包含二级RNA结构，例如但不限于发夹、茎/环、趾环、RNA结合蛋白募集结构域(如，募集并结合MS2cp蛋白的MS2适体)。这些二级结构可位于PEgRNA分子中的任何位置。例如，这种二级结构可位于间隔区、gRNA核心或延伸臂内，特别是位于e1和/或e2修饰区内。除二级RNA结构外，PEgRNA还可包含(如，在e1和/或e2修饰区域内)化学接头或poly(N)接头或尾，其中“N”可以是任何核碱基。在一些实施方案中(如，如图72(c)所示)，化学接头可起到防止sgRNA支架或核心逆转录的作用。此外，在某些实施方案中(如，参见图72(c))，延伸臂(3)可由RNA或DNA组成，和/或可包括一个或多个核碱基类似物(如，其可添加功能性，如温度弹性)。更进一步，延伸臂(3)的定向可以是自然的5′至3′方向，或在3′至5′方向(相对于整个PEgRNA分子的定向)以相反定向合成。还应注意，本领域普通技术人员将能够根据延伸臂的核酸材料(即，DNA或RNA)的性质选择合适的DNA聚合酶，用于引导编辑，该引导编辑可作为与napDNAbp的融合物或作为单独的部分以反式提供以合成含有期望编辑的期望模板编码的3′单链DNA瓣。例如，如果延伸臂是RNA，DNA聚合酶可以是逆转录酶或任何其他合适的RNA依赖性DNA聚合酶。然而，如果延伸臂是DNA，DNA聚合酶可以是DNA依赖性DNA聚合酶。在不同实施方案中，DNA聚合酶的提供可以为反式，例如通过使用RNA-蛋白募集结构域(如，在PEgRNA上安装的MS2发卡(如，在e1或e2区或其他地方，以及与DNA聚合酶融合的MS2cp蛋白中，从而将DNA聚合酶共定位到PEgRNA)。还应注意，引物结合位点通常不形成DNA聚合酶(如，逆转录酶)用来编码所得到的包含期望编辑的3′单链DNA瓣的模板的一部分。因此，“DNA合成模板”的名称是指由DNA聚合酶用作模板来编码包含编辑的期望的3′单链DNA瓣的延伸臂(3)的区域或部分和与5’内源性单链DNA瓣同源的区域，其中5’内源性单链DNA瓣被引导编辑DNA合成的3’单链DNA链产物置换。在一些实施方案中，DNA合成模板包括“编辑模板”和“同源臂”，或一个或多个同源臂，例如在编辑模板之前和之后。编辑模板可以小到单核苷酸取代，也可以是插入或DNA倒置。此外，编辑模板还可包含缺失，这可通过编码包含期望的缺失的同源臂来工程化改造。在其他实施方案中，DNA合成模板还可包括e2区或其一部分。例如，如果e2区包含导致DNA聚合酶活性终止的二级结构，在e2区的任何部分被实际编码到DNA中之前，DNA聚合酶的功能就有可能终止。部分甚至全部e2区域也有可能被编码到DNA中。有多少e2被实际用作模板将取决于其组成以及该组成是否中断DNA聚合酶功能。

[0113] 图3F的示意图描绘典型的PEgRNA与双链DNA的靶位点的相互作用以及伴随产生的包含感兴趣基因变化的3′单链DNA瓣。双链DNA显示为3’至5’方向的上面(top)链(即，靶链)和5’至3’方向的下面(lower)链(即，PAM链或非靶链)。上面链包含“前间隔序列”的互补序列和PAM序列的互补序列，并且被称为“靶链”，因为它是PEgRNA的间隔区的靶标并与其退火的链。互补的下面链被称为“非靶链”或“PAM链”或“前间隔序列链”，因为它包含PAM序列(如，NGG)和前间隔序列。虽然未显示，但描绘的PEgRNA将与引导编辑器融合蛋白的Cas9或

等同结构域复合。如示意图中所示，PEgRNA的间隔区与靶链的前间隔区序列的互补区域退火。这种相互作用在间隔区RNA和前间隔序列DNA的互补序列之间形成为DNA/RNA杂合体，并诱导在前间隔序列中形成R环。如本文别处所教导，Cas9蛋白(未显示)然后在非靶链中诱导切口，如所示。然后，这会导致紧靠切口位点上游的3′ssDNA瓣区的形成，根据*z*，该3′ssDNA瓣区与引物结合位点的PEgRNA的3′端相互作用。ssDNA瓣的3′端(即，逆转录酶引物序列)与PEgRNA上的引物结合位点(A)退火，从而引发逆转录酶。接着，逆转录酶(如，以反式提供或作为融合蛋白顺式提供，连接到Cas9构建体)聚合由DNA合成模板(包括编辑模板(B)和同源臂(C))编码的单链DNA。继续向延伸臂的5′端聚合。ssDNA的聚合链形成ssDNA 3’端瓣，如别处所述(如，如图1G所示)，侵入内源性DNA，置换相应的内源性链(其作为内源性DNA的5’端DNA瓣被去除)，并通过自然发生的DNA修复/复制循环安装期望的核苷酸编辑(单核苷酸碱基对变化、缺失、插入(包括整个基因)。

[0114] 图3G描绘本文预期的引导编辑的又一个实施方案。具体地，顶部示意图描绘引导编辑器(PE)的一个实施方案，其包含通过接头连接的napDNAbp(如，SpCas9)和聚合酶(如，逆转录酶)的融合蛋白。PE通过与PEgRNA的gRNA核心结合而与PEgRNA形成复合物。在所示实施方案中，PEgRNA配备有3’延伸臂，其从3’端开始包含引物结合位点(PBS)，后接DNA合成模板。底部示意图描绘引导编辑器的变体，被称为“反式引导编辑器(tPE)”。在该实施方案中，DNA合成模板和PBS与PEgRNA分开并呈现在单独的分子上，被称为反式引导编辑器RNA模板(“tPERT”)，其包含RNA-蛋白募集结构域(如，MS2发卡)。PE本身被进一步修饰以包含与rPERT募集蛋白(“RP”)的融合，RP是特异性识别并结合RNA蛋白募集结构域的蛋白。在RNA蛋白募集结构域是MS2发夹的示例中，相应rPERT募集蛋白可以是MS2标签化系统(tagging system)的MS2cp。MS2标签化系统基于MS2噬菌体外壳蛋白(“MCP”或“MS2cp”)与噬菌体基因组中存在的茎环或发夹结构，即“MS2发夹”或“MS2适体”的自然相互作用。在反式引导编辑的情况下，RP-PE:gRNA复合物“募集”具有适当RNA-蛋白募集结构域的tPERT以与PE:gRNA复合物共定位，从而以反式提供在引导编辑中使用的PBS和DNA合成模板，如图3H所示的示例。

[0115] 图3H描绘反式引导编辑的过程。在该实施方案中，反式引导编辑器包含与MS2cp蛋白(即，识别并结合MS2适体的一类募集蛋白)融合并且与sgRNA(即，与PEgRNA相反的标准向导RNA)复合的“PE2”引导编辑器(即，Cas9(H840A)和变体MMLVRT的融合物)。反式引导编辑器与靶DNA结合并对非靶链产生切口。MS2cp蛋白通过与tPERT分子的RNA蛋白募集结构域的特异性相互作用来反式募集tPERT。tPERT与反式引导编辑器共定位，从而以反式提供PBS和DNA合成模板，供逆转录酶聚合酶用来合成单链DNA瓣，其具有3’端并包含通过DNA合成模板编码的期望的遗传信息。

[0116] 图4A至4E显示体外TPRT分析(即，引导编辑分析)。图4A是通过RT酶进行的荧光标记的DNA底物gRNA模板化延伸，PAGE的示意图。图4B示出用预先产生切口的底物、dCas9和不同合成模板长度的5′-延伸的gRNA的TPRT(即，引导编辑)。图4C示出在不存在Cas9的情况下与预先产生切口的DNA底物的RT反应。图4D示出用具有Cas9(H840A)和5′-延伸的gRNA在完整dsDNA底物上的TPRT(即，引导编辑)。图4E示出在预先产生切口的且完整的dsDNA底物的情况下3′延伸的gRNA模板。所有反应均利用M-MLV RT。

[0117] 图5示出使用具有不同长度合成模板的5′延伸的gRNA进行的体外验证。将荧光标记的(Cy5)DNA靶标用作底物，并在这组实验中预先产生切口。这些实验中使用的Cas9是催

化死亡Cas9(dCas9),所用的RT是源自莫洛尼-鼠白血病病毒(M-MLV)的商业RT,SuperscriptIII。dCas9:gRNA复合物由纯化的组件形成。然后,将荧光标记的DNA底物与dNTP和RT酶一起加入。于37℃温育1小时后,通过变性尿素-聚丙烯酰胺凝胶电泳(PAGE)分析反应产物。凝胶图像显示原始DNA链延伸到与逆转录模板的长度一致的长度。

[0118] 图6示出使用具有不同长度合成模板的5'-延伸的gRNA的体外验证,其与图5中所示的那些紧密平行。然而,在这组实验中,DNA底物没有预先产生切口。这些实验中使用的Cas9是Cas9切口酶(SpyCas9 H840A突变体),所用的RT是源自莫洛尼-鼠白血病病毒(M-MLV)的商业RT,SuperscriptIII。通过变性尿素-聚丙烯酰胺凝胶电泳(PAGE)分析反应产物。如凝胶所示,当使用标准gRNA时,切口酶可有效切割DNA链(gRNA_0,泳道3)。

[0119] 图7显示3'延伸支持DNA合成并且不明显影响Cas9切口酶活性。当使用dCas9或Cas9切口酶时,预先产生切口的底物(黑色箭头)几乎定量地转化为RT产物(泳道4和泳道5)。利用完整底物观测到超过50%的RT产物转化率(红色箭头)(泳道3)。使用Cas9切口酶(SpyCas9 H840A突变体)、催化死亡Cas9(dCas9)和源自莫洛尼鼠白血病病毒(M-MLV)的商业RT,SuperscriptIII。

[0120] 图8显示用于确定RT反应是否优先以顺式与gRNA(结合在同一复合物中)一起发生的双色实验。对5′-延伸和3′-延伸gRNA进行两个独立的实验。通过PAGE分析产物。产品比率计算为(Cy3cis/Cy3trans)/(Cy5trans/Cy5cis)。

[0121] 图9A至9D显示瓣模型底物。图9A示出用于瓣定向诱变的双FP报道分子。图9B示出HEK细胞中的终止密码子修复。图9C示出瓣修复后测序的酵母集落。图9D示出人细胞中不同瓣特征的测试。

[0122] 图10显示对质粒底物的引导编辑。双荧光报告质粒是为酵母(酿酒酵母)表达而构建的。该构建体在酵母中的表达仅产生GFP。体外引导编辑反应引入点突变,并将亲本质粒或体外Cas9(H840A)切口质粒转化到酵母中。通过荧光成像使克隆可视化。示出酵母双FP质粒转化体。转化亲本质粒或体外Cas9(H840A)产生切口的质粒只产生绿色GFP表达集落。使用5′-延伸或3′-延伸的gRNA的引导编辑反应产生绿色和黄色集落的混合物。后者同时表达GFP和mCherry两者。利用3′延伸的gRNA观测到更多黄色集落。还显示不含终止密码子的阳性对照。

[0123] 图11显示对与图10中的实验类似的质粒底物的引导编辑,但不是在终止密码子中安装点突变,引导编辑安装修复移码突变并允许下游mCherry合成的单核苷酸插入(左)或缺失(右)。两个实验都使用3′延伸的gRNA。

[0124] 图12示出通过Sanger测序表征的对质粒底物的引导编辑的编辑产物。选择和通过Sanger测序分析来自TRT转化的单个集落。通过测序选择的集落来观测精确的编辑。绿色集落包含具有原始DNA序列的质粒,而黄色集落包含由引导编辑gRNA设计的精确突变。未观测到其他点突变或插入/缺失。

[0125] 图13示出新的引导编辑技术的潜在范围,并与脱氨酶介导的碱基编辑器技术比较。

[0126] 图14示出在人细胞中编辑的示意图。

[0127] 图15显示gRNA中引物结合位点的延伸。

[0128] 图16显示用于邻近靶向的截短的gRNA。

[0129] 图17A至17C是显示在人胚胎肾(HEK)细胞中组件转染之后靶核苷酸处的%T至A转化的图。图17A显示呈现使用野生型MLV逆转录酶与Cas9(H840A)切口酶的N端融合物(32个氨基酸接头)的结果的数据。图17B类似于图17A,除了RT酶的C-端融合外。图17C类似于图17A,但MLVRT和Cas9之间的接头长为60个氨基酸而不是32个氨基酸。

[0130] 图18显示通过高通量扩增子测序在HEK3位置处的高纯度T至A编辑。测序分析的输出显示经编辑的细胞的最丰富的基因型。

[0131] 图19显示靶核苷酸处的编辑效率(蓝色条)以及indel率(橙色条)。WT是指野生型MLV RT酶。突变酶(M1至M4)包含右侧列出的突变。编辑率通过基因组DNA扩增子的高通量测序来定量。

[0132] 图20显示在互补DNA链中邻近靶核苷酸处引入单链切口时靶核苷酸的编辑效率。测试在靶核苷酸不同距离处的切口产生(三角形)。靶碱基对的编辑效率(蓝色条)与indel形成率(橙色条)一起显示。“无”示例不包含互补链切口产生向导RNA。编辑率通过基因组DNA扩增子的高通量测序来定量。

[0133] 图21显示经处理的高通量测序数据,示出期望的T至A颠换突变和其他主要基因组编辑副产物的普遍缺失。

[0134] 图22提供使用与延伸的向导RNA复合的核酸可编程DNA结合蛋白(napDNAbp)在靶基因座上利用易错逆转录酶进行靶向诱变(即,利用易错RT的引导编辑)的示例性过程的示意图。该过程可被称为用于靶向诱变的引导编辑的实施方案。延伸的向导RNA包含在向导RNA的3’或5’端或在向导RNA的分子内位置处的延伸。步骤(a)中,napDNAbp/gRNA复合物与DNA分子接触,gRNA引导napDNAbp与待诱变的靶基因座结合。步骤(b)中,在靶基因座处的一条DNA链中引入切口(如,通过核酸酶或化学试剂),从而在靶基因座处的一条链中产生可用的3’端。在某些实施方案中,切口是在对应于R-环链的DNA链中产生的,即未与向导RNA序列杂交的链。步骤(c)中,3’端DNA链与向导RNA的延伸部分相互作用以引发逆转录。在某些实施方案中,3’端DNA链与向导RNA的延伸部分上的特定RT引发序列杂交。步骤(d)中,引入易错逆转录酶,该酶从引发位点的3’端至向导RNA的3’端合成诱变的单链DNA。示例性突变用星号“*”表示。这形成了包含期望的诱变区的单链DNA瓣。步骤(e)中,napDNAbp和向导RNA被释放。步骤(f)和(g)涉及单链DNA瓣(包含诱变区)的分解,使得期望的诱变区整合到靶基因座中。可通过去除相应的5’内源性DNA瓣来驱动该过程向着期望的产物形成,一旦3’单链DNA瓣侵入另一条链上的互补序列并与其杂交就形成该相应的5’内源性DNA瓣。也可利用第二链切口产生驱动该过程向着产物形成,如图1F所示例。在内源性DNA修复和/或复制过程之后,将诱变区掺入DNA基因座的两条DNA链中。

[0135] 图23是用于缩减三核苷酸重复序列的gRNA设计示意图和利用TPRT基因组编辑(即,引导编辑)缩减三核苷酸重复序列。三核苷酸重复扩增与许多人类疾病有关,包括亨廷顿氏病、脆性X综合征和弗里德赖希氏共济失调。最常见的三核苷酸重复包含CAG三联体,但也存在GAA三联体(弗里德赖希氏共济失调)和CGG三联体(脆性X综合征)。继承扩增的素因或获得已扩增的亲本等位基因增加患病的可能性。假设地,可以使用引导编辑来校正三核苷酸重复的致病性扩增。重复区域上游的区域可被RNA引导的核酸酶产生切口,然后用于引发包含健康重复数目(取决于特定基因和疾病)的新DNA链的合成。在重复序列之后,添加短的同源区段,其匹配邻近重复序列另一端的序列(红色链)的身份。新合成链的侵入,以及随

后用新合成的瓣置换内源性DNA,导致缩减的重复等位基因。

[0136] 图24是显示利用引导编辑得到的精确10个核苷酸的缺失的示意图。靶向HEK3基因座的向导RNA设计有逆转录模板,该模板在切口位点后编码10个核苷酸的缺失。使用扩增子测序评估经转染的HEK细胞的编辑效率。

[0137] 图25是显示内源性基因组基因座处的肽标签化基因的gRNA设计和利用TPRT基因组编辑(即,引导编辑)的肽标签化的示意图。FlAsH和ReAsH标签化系统包括两部分:(1)荧光团-双砷探针,和(2)包含四半胱氨酸基序的基因编码肽,例如序列FLNCCPGCCMEP(SEQ ID NO:1)。在细胞内表达时,含有四半胱氨酸基序的蛋白可用荧光团-砷探针进行荧光标记(参见参考文献:J.Am.Chem.Soc.,2002,124(21),pp6063-6076.DOI:10.1021/ja017687n)。"分选标签化(sortagging)"系统采用细菌分选酶,其将标记的肽探针与含有合适肽底物的蛋白共价缀合(参见参考文献:Nat.Chem.Biol.2007Nov;3(11):707-8.DOI:10.1038/nchembio.2007.31)。FLAG标签(DYKDDDDK(SEQ ID NO:2))、V5标签(GKPIPNPLLGLDST(SEQ ID NO:3))、GCN4标签(EELLSKNYHLENEVARLKK(SEQ ID NO:4))、HA标签(YPYDVPDYA(SEQ ID NO:4))ID NO:5))和Myc-标签(EQKLISEEDL(SEQ ID NO:6))通常用作免疫测定的表位标签。π钳(p-clamp)编码的肽序列(FCPF(SEQ ID NO:622))可用五氟芳香族底物标记(参考:Nat.Chem.2016Feb;8(2):120-8.doi:10.1038/nchem.2413)。

[0138] 图26A示出将His_6-标签和FLAG-标签精确安装到基因组DNA中。靶向HEK3基因座的向导RNA设计有逆转录模板,该模板编码18-nt His标签插入或24-ntFLAG标签插入。使用扩增子测序评估经转染的HEK细胞中的编辑效率。注意,FLAG标签的完整24-nt序列在观测框(viewing frame)外(测序确认完整和精确插入)。图26B示出概述涉及蛋白/肽标签化的各种应用的示意图,包括(a)使蛋白可溶或不可溶,(b)改变或追踪蛋白的细胞定位,(c)延长蛋白半衰期,(d)促进蛋白纯化,和(e)促进蛋白检测。

[0139] 图27示出通过在PRNP中安装保护性突变的引导编辑的概述,所述保护性突变防止或阻止朊病毒病进展。PEgRNA序列对应于左侧的SEQ ID NO:810的残基编号1-20(即,sgRNA支架的5')和右侧的SEQ ID NO:810的残基编号21-43(即,sgRNA支架的3')

[0140] 图28A是编码RNA基序的序列的基于PE的插入的示意图。图28B是可能被插入的一些示例基序及其功能的列表(非详尽的)。

[0141] 图29A是引导编辑器的描绘。图29B示出对PE引导的基因组、质粒或病毒DNA的可能修饰。图29C示出通过PEgRNA文库将肽环文库插入规定蛋白(在这种情况下为GFP)的示例方案。图29D示出使用不同PEgRNA的蛋白的密码子或N-或C-端截短的可能的可编程缺失的实例。预计会以移码突变的最少产生的情况下发生缺失。

[0142] 图30示出在诸如PACE的连续进化系统中重复插入密码子的可能方案。

[0143] 图31是示出gRNA核心、与靶向基因序列匹配的约20nt间隔区、具有免疫原性表位核苷酸序列的逆转录模板和与靶向基因序列匹配的引物结合位点的工程化gRNA的图。

[0144] 图32是示出使用引导编辑作为将已知免疫原性表位插入内源或外源基因组DNA,导致相应蛋白修饰的手段的示意图。

[0145] 图33的示意图示出使用引导编辑确定脱靶编辑用于引物结合序列插入和引物结合插入基因组DNA的PEgRNA设计。在该实施方案中,在活细胞、组织或动物模型内进行引导编辑。第一步,设计合适的PEgRNA。上方示意图示出可用于该方面的示例性PEgRNA。PEgRNA

中的间隔区(标记为“前间隔序列”)与基因组靶标的链之一互补。PE:PEgRNA复合物(即,PE复合物)在切口位点安装单链3′端瓣,其包含编码的引物结合序列和同源区域(由PEgRNA的同源臂编码),该同源区域与正好在切割位点下游的区域(红色)互补。通过瓣侵入和DNA修复/复制过程,合成的链被掺入DNA中,从而安装引物结合位点。这个过程可发生在期望的基因组靶标处,也可发生在可能以脱靶方式与PEgRNA相互作用的其他基因组位点处(即,由于间隔区区域与非期望的基因组位点的其他基因组位点的互补性,PEgRNA将PE复合物引导至其他脱靶位点)。因此,引物结合序列不仅可安装在期望的基因组靶标处,还可能安装在基因组其他地方的脱靶基因组位点。为了检测这些引物结合位点在意图的基因组靶位点和脱靶基因组位点的插入,可以将基因组DNA(PE后)分离、片段化并连接至衔接子核苷酸(以红色显示)。接着,可利用与衔接子和插入的引物结合序列退火的PCR寡核苷酸进行PCR,以扩增由PE插入引物结合位点的中靶(on-target)和脱靶(off-target)基因组DNA区域。然后可以进行高通量测序和序列比对,以鉴定PE插入的引物结合序列在中靶位点或脱靶位点处的插入点。

[0146] 图34是示出利用PE的基因的精确插入的示意图。

[0147] 图35A是示出天然胰岛素信号传导途径的示意图。图35B是示出由FK1012控制的FKBP12标签化的胰岛素受体激活的示意图。

[0148] 图36示出小分子单体。参考文献:bumpedFK506模拟物(2)[107]。

[0149] 图37A至37B示出小分子二聚体。参考文献:FK10124[95,96];FK10125[108];FK10126[107];AP19037[107];环孢霉素A二聚体8[98];FK506-环孢霉素A二聚体(FkCsA)9[100]。

[0150] 图38A至38F提供体外和酵母细胞中引导编辑和可行性研究的概述。图38A示出ClinVar中的75,122个已知的人类致病性遗传变体,按类型分类(2019年7月访问)。图38B示出由引导编辑器(PE)蛋白组成的引导编辑复合物,所述PE蛋白包含融合至工程化逆转录酶结构域并与引导编辑向导RNA(PEgRNA)复合的RNA引导的DNA切口结构域,如Cas9切口酶。PE:PEgRNA复合物结合靶DNA位置,并中靶位点的前间隔序列邻近基序(PAM)之前或之后的各种DNA位点处实现极其多种精确的DNA编辑。图38C示出在DNA靶标结合时,PE:PEgRNA复合物对含有PAM的DNA链产生切口。产生的游离3′端与PEgRNA的引物结合位点杂交。逆转录酶结构域使用PEgRNA的RT模板催化引物延伸,产生包含期望编辑的新合成的DNA链(3’瓣)。编辑的3’瓣和包含原始DNA的非编辑的5’瓣之间平衡,接着细胞5’瓣切割和连接,以及DNA修复或复制以分解异源双链体DNA,导致稳定编辑的DNA。图38D示出利用含有5’-Cy5标记的PAM链的预先产生切口的dsDNA底物、dCas9和商业M-MLV RT变体(RT,SuperscriptIII)的体外5’-延伸的PEgRNA引物延伸测定。dCas9与包含不同长度的RT模板的PEgRNA复合,然后与指定的成分一起添加到DNA底物。反应于37℃下温育1小时,然后通过尿素变性PAGE进行分析并针对Cy5荧光可视化。图38E示出使用与dCas9或Cas9 H840A切口酶预先复合的3’-延伸PEgRNA和预先产生切口或未产生切口的5’-Cy5-标记的dsDNA底物如在图38D进行引物延伸测定。图38F示出利用在体外用PEgRNA、Cas9切口酶和RT编辑的GFP-mCherry融合报告基因质粒转化酵母集落。包含GFP和mCherry之间无义或移码突变的质粒利用5’-延伸或3’-延伸PEgRNA进行编辑,这些PEgRNA通过颠换突变、1-bp插入或1-bp缺失恢复mCherry翻译。GFP和mCherry双阳性细胞(黄色)反映成功的编辑。

[0151] 图39A至39D示出人类细胞中基因组DNA通过PE1和PE2的引导编辑。图39A示出

PEgRNA包含间隔区序列、sgRNA支架和包含引物结合位点(绿色)和逆转录(RT)模板(紫色)的3'延伸,其包含编辑的碱基(红色)。引物结合位点与紧邻切口产生位点上游的含PAM的DNA链杂交。除编码编辑外,RT模板与切口下游的DNA序列同源。图39B示出使用与野生型M-MLV逆转录酶(PE1)融合的Cas9 H840A切口酶和不同引物结合位点长度的PEgRNA在HEK293T细胞中的HEK3位点安装T·A至A·T颠换编辑。图39C示出在PE2中使用工程化五突变M-MLV逆转录酶(D200N、L603W、T306K、W313F、T330P)显著提高HEK293T细胞中5个基因组位点处引导编辑颠换效率,以及HEK3处的小插入和小缺失编辑。图39D是HEK293T细胞中5个基因组位点处用不同RT模板长度的PE2编辑效率的比较。值和误差棒反映3个独立生物学重复的平均值和标准差。

[0152] 图40A至40C示出PE3和PE3b系统对非编辑链产生切口以增加引导编辑效率。图40A是通过PE3的引导编辑的概述。在编辑链的初始合成后,DNA修复将去除包含编辑的新合成链(3'瓣切除)或原始基因组DNA链(5'瓣切除)。5'瓣切除留下包含一条编辑链和一条非编辑链的DNA异源双链体。错配修复机制或DNA复制可分解异源双链体以提供编辑或非编辑的产物。对非编辑链产生切口有利于修复该链,优先生成包含期望编辑的稳定双链体DNA。图40B示出互补链切口产生对PE3介导的引导编辑效率和indel形成的影响。"无"是指PE2对照,其不对互补链产生切口。图40C是利用PE2(无互补链切口)、PE3(一般互补链切口)和PE3b(编辑特异性互补链切口)的编辑效率的比较。所有编辑产率都反映了没有分选的情况下,所有经处理的细胞中包含预期编辑但不包含indel的总测序读段百分比。值和误差棒反映3个独立生物学重复的平均值和标准差。

[0153] 图41A至41K示出HEK293T细胞中7个内源性人基因组基因座处用PE3得到的靶向插入、缺失和全部12种类型的点突变。图41A的图示出使用10-nt RT模板从HEK3位置的位点+1至位点+8(计算PEgRNA诱导的切口为位点+1和位点-1之间的位置)的全部12种类型的单核苷酸转换和颠换编辑。图41B的图示出使用34-nt RT模板在HEK3位置进行远程PE3颠换编辑。图41C至41H的图示出(图41C)RNF2、(图41D)FANCF、(图41E)EMX1、(图41F)RUNX1、(图41G)VEGFA和(图41H)DNMT1的引导编辑窗口中不同位点处全部12种类型的转换和颠换编辑。图41I的图示出利用PE3在7个内源性基因组基因座处靶向的1和3bp插入以及1和3bp缺失。图41J是示出在HEK3靶基因座处5至80bp的靶向精确缺失的图。图41K是示出3个内源性基因组基因座处插入和缺失、插入和点突变、缺失和点突变以及双点突变的组合编辑的图。所有编辑产率都反映没有分选的情况下,经处理细胞中包含预期编辑但不包含indel的总测序读段百分比。值和误差棒反映3个独立生物学重复的平均值和标准差。

[0154] 图42A至42H示出已知Cas9脱靶位点处由Cas9和PE3进行的引导编辑和碱基编辑以及脱靶编辑的比较。图42A示出在HEK293T细胞的内源性HEK3、FANCF和EMX1位置处PE2、PE3、BE2max和BE4max的相同靶核苷酸处的总C·G至T·A编辑效率。图42B示出来自图42中处理的indel频率。图42C示出HEK3、FANCF和EMX1处PE2、PE3、BE2max和BE4max的精确C·G至T·A编辑的编辑效率(没有旁观者编辑或indel)。对于EMX1,还示出在3个靶向核苷酸处C·G至T·A转换的所有可能组合的精确PE组合编辑。图42D示出在HEK3和FANCF处PE2、PE3、ABEdmax和ABEmax的总A·T至G·C编辑效率。图42E示出在HEK3和FANCF处没有旁观者编辑或indel的精确A·T至G·C编辑效率。图42F示出来自图42中处理的indel频率。图42G示出HEK293T细胞中Cas9核酸酶在4个中靶位点和16个已知的脱靶位点处的一式三份平均的编

辑效率(具有indel的测序读段百分比)。所检测的16个脱靶位点是4个中靶位点中每个的前4个之前报告的脱靶位点[118,159]。对于每个中靶位点,Cas9与sgRNA或识别同一前间隔序列的4个PEgRNA中的每一个配对。图42H示出与(图42G)中的每个PEgRNA配对的PE2或PE3在HEK293T细胞中一式三份平均的中靶和脱靶编辑效率和indel效率(以下在括号中)。中靶编辑产率反映没有分选的情况下所有经处理的细胞中包含意图的编辑且不包含indel的总测序读段百分比。脱靶编辑产率反映与引导编辑一致的脱靶基因座修饰。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0155] 图43A至43I示出不同人类细胞系和原代小鼠皮层神经元中的引导编辑,致病性颠换、插入或缺失突变的安装和校正,以及引导编辑和HDR的比较。图43A是示出在HEK293T细胞的HBB中致病性E6V突变的安装(通过T·A至A·T颠换)和校正(通过A·T至T·A颠换)的图。示出对野生型HBB或包含破坏PEgRNAPAM的沉默突变的HBB的校正。图43B是示出HEK293T细胞的致病性HEXA 1278+TATC等位基因的安装(通过4-bp插入)和校正(通过4-bp缺失)的图。示出对野生型HEXA或包含破坏PEgRNA PAM的沉默突变的HEXA的校正。图43C是示出通过G·C至T·A颠换在HEK293T细胞的PRNP中安装保护性G127V变体的图。图43D是示出包括K562(白血病骨髓细胞)、U2OS(骨肉瘤细胞)和HeLa(宫颈癌细胞)在内的其他人类细胞系中的引导编辑的图。图43E是示出使用双断裂内含肽(dual split-intein)PE3慢病毒系统在小鼠原代皮层神经元的DNMT1中安装G·C至T·A颠换突变的图,其中N-端半部是融合至N-内含肽,并通过P2A自切割肽融合至GFP-KASH的Cas9(1-573),并且C-端半部是与PE2其余部分融合的C-内含肽。PE2半部由对成熟神经元高度特异的人类突触蛋白启动子表达。分选的值反映来自GFP阳性细胞核的编辑或indel,而未分选的值来自所有细胞核。图43F是HEK293T细胞中内源性基因组基因座处PE3和Cas9介导的HDR编辑效率的比较。图43G是K562、U2OS和HeLa细胞的内源性基因组基因座处PE3和Cas9介导的HDR编辑效率的比较。图43H是HEK293T、K562、U2OS和HeLa细胞中PE3和Cas9介导的HDR indel副产物生成的比较。图43I示出HEK293T细胞中由PE3靶向插入His6标签(18bp)、FLAG表位标签(24bp)或延伸的LoxP位点(44bp)。所有编辑产率都反映所有经处理的细胞中包含意图的编辑但不包含indel的总测序读段百分比。值和误差棒反映3个独立生物学重复的平均值和标准差。

[0156] 图44A至44G示出利用荧光标记的DNA底物的体外引导编辑验证研究。图44A示出利用dCas9、5′-延伸PEgRNA和5′-Cy5-标记的DNA底物的电泳迁移率分析。PEgRNA 1至5包含间隔区与PBS之间的15-nt接头序列(接头A用于PEgRNA 1,接头B用于PEgRNA2至5)、5-nt的PBS序列和7nt(PEgRNA 1和2)、8nt(PEgRNA3)、15nt(PEgRNA4)和22nt(PEgRNA5)的RT模板。PEgRNA是图44E和44F中使用的那些;完整序列列于表2A至2C中。图44B示出使用5′-延伸和3′-延伸PEgRNA的Cas9 H840A的体外切口产生测定。图44C示出使用5′-延伸和3′-延伸PEgRNA在HEK293T细胞中在HEK3处的Cas9介导的indel形成。图44D示出体外生化分析的引导编辑概述。测试5′-Cy5标记的预先切口和无切口的dsDNA底物。sgRNA、5′-延伸PEgRNA或3′-延伸PEgRNA与dCas9或Cas9 H840A切口酶预先复合,然后与dsDNA底物、M-MLV RT和dNTP组合。允许反应于37℃进行1小时,然后通过尿素变性PAGE分离并通过Cy5荧光可视化。图44E示出使用5′-延伸PEgRNA、预先产生切口的DNA底物和dCas9的引物延伸反应导致显著转化为RT产物。图44F示出使用如图44B中的5′-延伸PEgRNA和未产生切口的DNA底物及Cas9 H840A切口酶的引物延伸反应。与预先产生切口的底物相比,产物产率大大降低。图44G示出

使用3′-PEgRNA的体外引物延伸反应通过尿素变性PAGE产生单一表观产物。RT产物条带被切离，从凝胶上洗脱，然后使用dGTP或dATP进行用末端转移酶(TdT)的同聚物加尾。加尾产物通过poly-T或poly-C引物进行延伸，并对所得到的DNA进行测序。Sanger迹线表明源自gRNA支架的三个核苷酸被逆转录(作为最后的3′核苷酸添加至DNA产物中)。注意，在哺乳动物细胞引导编辑实验中，PEgRNA支架插入比体外罕见得多(图56A-56D)，这可能是由于栓系的逆转录酶无法接近Cas9结合的向导RNA支架和/或含有PEgRNA支架序列的3′瓣的错配3′端的细胞切除。

[0157] 图45A至45G示出酵母中来自体外引导编辑反应的3′DNA瓣的细胞修复。图45A示出双荧光蛋白报告质粒含有由编码框内终止密码子，+1移码或-1移码的靶位点分离的GFP和mCherry开放阅读框。利用Cas9 H840A切口酶、PEgRNA、dNTP和M-MLV逆转录酶在体外进行引导编辑反应，然后转化到酵母。含有非编辑质粒的集落产生GFP但不产生mCherry。含有编辑质粒的酵母集落产生作为融合蛋白的GFP和mCherry。图45B示出利用报告质粒转化的酵母集落的GFP和mCherry荧光叠加，其中报告质粒包含在GFP和mCherry之间的终止密码子(非编辑的阴性对照，上图)，或不包含在GFP和mCherry之间的终止密码子或移码(预先编辑的阳性对照，下图)。图45C至45F示出来自利用体外引导编辑反应产物转化的酵母集落的mCherry和GFP荧光的可视化。图45C示出使用3′-延伸PEgRNA或使用5′-延伸PEgRNA通过T·A至A·T颠换的终止密码子校正，如图45D所示。图45E示出使用3′-延伸PEgRNA通过1-bp缺失的+1移码校正。图45F示出使用3′-延伸PEgRNA通过1-bp插入的-1移码校正。图45G示出来自从图45B中仅GFP集落和图45C中GFP和mCherry双阳性集落分离的质粒的SangerDNA测序迹线。

[0158] 图46A至46F示出利用PE1的正确编辑相对于indel生成。图46A示出使用包含10-nt RT模板和范围为8-17nt的PBS序列的PEgRNA在HEK3的位点+1处由PE1进行的T·A至A·T颠换编辑效率和indel生成。图46B示出使用包含13-nt RT模板和范围为9-17nt的PBS序列的PEgRNA在EMX1的位点+5处由PE1进行的G·C至T·A颠换编辑效率和indel生成。图46C示出使用包含17-nt RT模板和范围为8-17nt的PBS序列的PEgRNA在FANCF的位点+5处由PE1进行的G·C至T·A颠换编辑效率和indel生成。图46D示出使用包含11-ntRT模板和范围为9-17nt的PBS序列的PEgRNA在RNF2的位点+1处由PE1进行的C·G至A·T颠换编辑效率和indel生成。图46E示出使用含有13-nt RT模板和范围为7-15nt的PBS序列的PEgRNA在HEK4的位点+2由PE1进行的G·C至T·A颠换编辑效率和iindel生成。图46F示出使用13-ntPBS和10-nt RT模板在HEK3位置处由PE1介导的+1T缺失、+1A插入和+1CTT插入。PEgRNA的序列是图39C中使用的那些(见表3A-3R)。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0159] 图47A至47S示出用于引导编辑的M-MLV RT变体的评估。图47A示出该图中使用的引导编辑器变体的缩写。图47B示出在HEK3基因座处利用PE1的靶向插入和缺失编辑。图47C至47H示出比较包含M-MLV RT变体的18个引导编辑器构建体在HEK3处安装+2G·C至C·G颠换编辑的能力，如图47C示出，在HEK3处安装24-bp FLAG插入的能力，如图47D示出，在RNF2处安装+1C·G至A·T颠换编辑的能力，如图47E示出，在EMX1处安装+1G·C至C·G颠换编辑的能力，如图47F示出，在HBB处安装+2T·A至A·T颠换编辑的能力，如图47G示出，以及在FANCF处安装+1G·C至C·G颠换编辑的能力，如图47H示出。图47I至47N示出比较包含M-MLV变体的4种引导编辑器构建体在第二轮独立实验中安装图47C至47H中示出的安装编辑的能

力。图47O至47S示出具有不同PBS长度的5个基因组基因座处的PE2编辑效率。图47O示出HEK3处+1T·A至A·T变化。图47P示出EMX1处+5G·C至T·A变化。图47Q示出FANCF处+5G·C至T·A变化。图47R示出RNF2处+1C·G至A·T变化。图47S示出HEK4处+2G·C至T·A变化。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0160] 图48A至48C示出PEgRNAPBS和RT模板序列的设计特征。图48A示出HEK293T细胞中VEGFA处PE2介导的+5G·C至T·A颠换编辑效率(蓝线)作为RT模板长度的函数。绘制indel(灰线)进行比较。图下方的序列示出由PEgRNA合成的最后一个提供模板的核苷酸。G核苷酸(在PEgRNA中以C为模板)突出显示;在PEgRNA设计过程中应避免以C结尾的RT模板,以使引导编辑效率最大化。图48B示出DNMT的+5G·C至T·A颠换编辑和indel,如图48A中一样。图48C示出RUNX1的+5G·C至T·A颠换编辑和indel,如图48A中一样。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0161] 图49A至49B示出PE2、PE2 R110S K103L、Cas9 H840A切口酶和dCas9对细胞活力的影响。利用编码PE2、PE2 R110S K103L、Cas9H840A切口酶或dCas9的质粒以及靶向HEK3的PEgRNA质粒转染HEK293T细胞。使用CellTiter-Glo2.0测定法(Promega)在转染后每24小时测量一次细胞活力,持续3天。图49A示出转染后1、2或3天通过发光测量的活力。值和误差棒反映三个独立生物学重复的平均值和s.e.m,每个重复技术上进行一式三份。图49B示出PE2、PE2 R110S K103L、Cas9 H840A切口酶或dCas9连同编码+5G至A编辑的靶向HEK3的PEgRNA质粒的百分比编辑和indel。转染后第3天测量经处理细胞连同图49A中用于测量活力的那些细胞的编辑效率。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0162] 图50A至50B示出由不同PEgRNA进行的PE3介导的HBB E6V校正和HEXA 1278+TATC校正。图50A示出筛选14种PEgRNA在HEK293T细胞中利用PE3校正HBB E6V等位基因。评估的所有PEgRNA将HBB E6V等位基因转换回野生型HBB,而不会引入任何沉默PAM突变。图50B示出筛选41种PEgRNA在HEK293T细胞中利用PE3或PE3b校正HEXA 1278+TATC等位基因。那些标记为HEXA的PEgRNA通过移动4bp缺失来校正致病性等位基因,该缺失会破坏PAM并留下沉默突变。那些标记为HEXA的PEgRNA将致病性等位基因校正回野生型。以“b”结尾的条目将编辑特异性切口产生性sgRNA与PEgRNA结合使用(PE3b系统)。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0163] 图51A至51G示出人类细胞系中的PE3活性以及比较PE3和Cas9启动的HDR。如图51A所示在HEK293T细胞中,如图51B中所示在K562细胞中,如图51C所示在U2OS细胞中,如图51D所示在HeLa细胞中,PE3和Cas9启动的HDR生成正确编辑(无indel)和indel频率的效率。每个括号内的编辑比较都利用PE3和Cas9启动的HDR安装相同的编辑。非靶向对照是PE3和靶向非靶基因座的PEgRNA。图51E示出利用非靶向PEgRNA+PE3和利用dCas9+sgRNA的对照实验,与野生型Cas9 HDR实验相比,证实ssDNA供体HDR模板(人工提高表观HDR效率的常见污染物)对图51A至51D中的HDR测量没有贡献。图51F至51G示出利用PE3或利用Cas9启动的HDR编辑后从K562细胞中分离的基因组DNA样品的示例HEK3位点等位基因表。等位基因用Illumina MiSeq测序并利用CRISPResso2进行分析[178]。来自该区域的参考HEK3序列位于顶端。示出非靶向PEgRNA阴性对照、使用PE3在HEK3处的+1CTT插入和使用Cas9启动的HDR在HEK3处的+1CTT插入的等位基因表。示出每个等位基因的等位基因频率和相应的Illumina测序读段计数。示出所观测的频率≥0.20%的所有等位基因。值和误差棒反映三个独立生

物学重复的平均值和标准差。

[0164] 图52A至52D示出ClinVar数据库中根据致病性插入、重复、缺失和indel的长度的分布。ClinVar变体摘要于2019年7月15日从NCBI下载。报告的插入、缺失和重复的长度使用参考和替代等位基因、变体开始和终止位点或变体名称中的适当识别信息来计算。分析排除了未报告任何上述信息的变体。报告的indel(包括相对于参考基因组的插入和缺失的单个变体)的长度通过确定参考和替代等位基因之间最佳配对比对中的错配或缺口数目来计算。

[0165] 图53A至53E示出用于GFP阳性细胞分选的FACS门控实例。以下是原始批量分析文件的示例,概述用于生成HEXA 1278+TATC和HBB E6V HEK293T细胞系的分选策略。图像数据在SonyLE-MA900细胞仪上使用Cell Sorter软件v.3.0.5生成。图1示出不表达GFP的细胞的门控图。图2示出用于分离HBB E6V HEK293T细胞系的P2A-GFP表达细胞的示例。HEK293T细胞最初使用FSC-A/BSC-A(门控A)对群体进行门控,然后使用FSC-A/FSC-H(门控B)分选单峰(singlets)。通过门控DAPI阴性细胞(门控C)对活细胞进行分选。使用EGFP作为荧光染料分选具有高于阴性对照细胞的GFP荧光水平的细胞(门控D)。图53A示出HEK293T细胞(GFP阴性)。图53B示出表达PE2-P2A-GFP的细胞的FACS门控的代表性图。图53C示出HEK293 T细胞HEXA 1278+TATC纯合体的基因型。图53D-53E示出HEK293 T细胞系HBB E6V纯合体的等位基因表。

[0166] 图54是概括PEgRNA克隆程序的示意图。

[0167] 图55A至55G是PEgRNA设计的示意图。图55A示出PEgRNA的简化图,其中结构域被标记(左)并在基因组位点与nCas9结合(右)。图55B示出预期增加活性的对PEgRNA的各种类型的修饰。图55C示出对PEgRNA的修饰,以通过启动子选择和5′、3′加工和终止来增加较长RNA的转录。图55D示出P1系统的延长,这是支架修饰的示例。图55E示出在模板区域内或在PEgRNA内的其他地方掺入合成修饰可增加活性。图55F示出设计模板内最小二级结构的掺入可防止形成更长、更具抑制性的二级结构。图55G示出具有由PEgRNA的3′端(左)RNA元件锚定的第二模板序列的断裂PEgRNA。在PEgRNA的5′或3′端掺入元件可增强RT结合。

[0168] 图56A至56D示出将PEgRNA支架序列掺入靶基因座。如图60A-60B所述,分析PEgRNA支架序列插入的HTS数据。图56A示出对于EMX1基因座的分析。显示在邻近RT模板的插入中包含一个或多个PEgRNA支架序列核苷酸的总测序读段百分比(左);包含指定长度的PEgRNA支架序列插入的总测序读段百分比(中间);以及直至并包括X轴上指定长度的PEgRNA插入的累计总百分比。图56B示出除了FANCF外与图56A相同。图56C示出除了HEK3外与图56A相同。图56D示出除了RNF2外与图56A相同。值和误差棒反映三个独立生物学重复的平均值和标准差。

[0169] 图57A至57I示出PE2、PE2-dRT和Cas9 H840A切口酶对全转录组RNA丰度的影响。分析从表达PE2、PE2-dRT或Cas9H840A切口酶和PRNP靶向或HEXA靶向性PEgRNA的HEK293T细胞中分离的细胞RNA,去除了核糖体RNA。分别检测PRNP和HEXA样品中对应于14,410个基因和14,368个基因的RNA。图57A至57F示出显示AeachRNA转录本丰度的-log10 FDR调整的p值相比log2倍变化的火山图,比较(图57A)在靶向PRNP的PEgRNA的情况下PE2相比PE2-dRT,(图57B)在靶向PRNP的PEgRNA的情况下PE2相比Cas9 H840A,(图57C)在靶向PRNP的PEgRNA的情况下PE2-dRT相比Cas9H840A,(图57D)在靶向HEXA的PEgRNA的情况下PE2相比PE2-dRT,(图

57E)在靶向HEXA的PEgRNA的情况下PE2相比Cas9 H840A,(图57F)在靶向HEXA的PEgRNA的情况下PE2-dRT相比Cas9 H840A。红点表示相对丰度变化≥2倍且统计学显著的基因(FDR调整的p<0.05)。图57G至57I是上调和下调转录本(≥2倍变化)的维恩(Venn)图,比较PRNP和HEXA样品,(图57G)PE2相比PE2-dRT、(图57H)PE2相比Cas9 H840A和(图57I)PE2-dRT相比Cas9H840A。

[0170] 图58A至58B示出用于神经元核分选的代表性FACS门控。基于DyeCycle Ruby信号、FSC/SSC比、SSC宽度/SSC高度比和GFP/DyeCycle比依次对核进行门控。

[0171] 图59A至59G示出通过Golden Gate组装将3′-延伸PEgRNA克隆到哺乳动物U6表达载体的方案。图59A示出克隆概述。图59B示出“步骤1:消化pU6-PEgRNA-GG-载体质粒(组件1)”。图59C示出“步骤2和3:对寡核苷酸部分(组件2、3和4)进行排序和退火”。图59D示出“步骤2.b.ii.:sgRNA支架磷酸化(如果购买了磷酸化寡核苷酸则不需要)”。图59E示出“步骤4:PEgRNA组装”。图59F示出“步骤5和6:组装质粒的转化”。图59G示出总结PEgRNA克隆方案的图。

[0172] 图60A至60B示出用于定量PEgRNA支架整合的Python脚本。生成了自定义python脚本来表征和定量靶基因组基因座处的PEgRNA插入。该脚本将从参考序列(向导RNA支架序列)中采集的长度增加的文本字符串与fastq文件中的测序读段迭代匹配,并计算与搜索查询匹配的测序读段数目。每个连续的文本字符串对应于向导RNA支架序列的附加核苷酸。以这种方式计算精确长度积分和累计积分达到指定长度。在参考序列的起始处,包含逆转录酶合成的新DNA链3′端的5至6个碱基,以确保sgRNA短片段的比对和准确计数。

[0173] 图61是示出对于SaCas9(N580A)-MMLV RT HEK3+6C>A具有指定编辑的总测序读段的百分比的图。示出正确编辑和indel的值。

[0174] 图62A至62B示出前间隔序列对于利用引导编辑在精确位置处有效安装期望编辑的重要性。图62A是示出针对各个HEK3基因座具有转换为T·A的靶标A·T碱基对的总测序读段的百分比的图。图62B是示出相同的序列分析。

[0175] 图63是示出PAM编辑中SpCas9 PAM变体(N=3)的图。示出对于SpCas9(H840A)-VRQR-MMLV RT,其中NGA>NTA,以及对于SpCas9(H840A)-VRER-MMLV RT,其中NGCG>NTCG,具有靶向PAM编辑的总测序读段百分比。列出PEgRNA引物结合位点(PBS)长度、RT模板(RT)长度和使用的PE系统。

[0176] 图64A至64F是示出使用PE将各个位点特异性重组酶(SSR)靶标引入基因组的示意图。图64A提供通过引导编辑器插入重组酶靶序列的一般示意图。图64B示出如何将PE插入的单个SSR靶标用作DNA供体模板的基因组整合的位点。图64C示出如何可以使用串联插入SSR靶位点来删除基因组的部分。图64D示出可以如何使用串联插入SSR靶位点来倒置基因组的部分。图64E示出在两个远端染色体区域插入两个SSR靶位点如何可以导致染色体易位。图64F示出如何可以使用基因组中两个不同SSR靶位点的插入来交换来自DNA供体模板的盒。

[0177] 图65示出1)人类细胞基因组中SSR靶位点的PE介导的合成和2)使用该SSR靶位点整合包含GFP表达标志物的DNA供体模板。一旦成功整合,GFP就会使细胞发荧光。

[0178] 图66描绘作为两个PE半部蛋白提供的引导编辑器的一个实施方案,所述蛋白通过位于每个引导编辑器半部蛋白的末端或起点处的断裂内含肽半部的自剪接作用再生为完

整的引导编辑器。

[0179] 图67A至67B描绘从多肽序列中去除内含肽和在N-端和C-端外显肽序列之间重新形成肽键的机制。图67A描述两个半部蛋白的一般机制，每个蛋白都包含内含肽序列的一半，当它们在细胞内接触时产生全功能的内含肽，所述内含肽然后进行自剪接和切除。切除过程导致在N端蛋白半部(或“N外显肽”)和C端蛋白半部(或“C外显肽”)之间形成肽键，以形成完整的包含N外显肽和C外显肽部分的单一多肽。在不同实施方案中，N外显肽可对应于断裂引导编辑器融合蛋白的N-端半部，并且C外显肽可对应于断裂引导编辑器的C-端半部。(b)示出内含肽切除和重新形成连接N外显肽半部(红色半部)和C外显肽半部(蓝色半部)的肽键的的化学机制。断裂内含肽(即，断裂内含肽构造中的N内含肽和C内含肽)的切除也可称为“反式剪接”，因为它涉及以反式提供的两个独立组件的剪接作用。

[0180] 图68A显示共转染到EK293T细胞中时接头处SpPE(SEQ ID NO:3875,3876)的两个断裂内含肽半部的递送在三个测试基因座处保持活性。

[0181] 图68B显示共转染到HEK293T细胞中时SaPE2的两个断裂内含肽(如，SEQ ID NO:443和SEQ ID NO:450)的递送重现全长SaPE2(SEQ ID NO:134)的活性。引号中指示的残基是SaCas9中氨基酸741-743的序列(C-端外显肽的第一个残基)，其对于内含肽反式剪接反应是重要的。“SMP”是天然残基，我们也将其突变为“CFN”共有剪接序列。如通过引导编辑百分比测量，显示共有序列产生最高的重构。

[0182] 图68C提供示出可以这种方式递送各种公开的PE核糖核蛋白复合物(高浓度PE2、高浓度PE3和低浓度PE3)的数据。

[0183] 图69示出确定PANCE中PE有效性的噬菌体噬菌斑测定。斑块(黑圈)表明噬菌体能够成功感染大肠杆菌。L-鼠李糖浓度的增加导致PE表达增加和噬菌斑形成增加。噬菌斑的测序揭示存在PE安装的基因组编辑。

[0184] 图70A至70I提供了编辑靶序列的示例，作为用于设计PEgRNA和切口产生性sgRNA进行引导编辑的分步说明的图解。图70A:步骤1，定义靶序列和编辑。检索围绕期望编辑(点突变、插入、缺失或其组合)位置为中心的靶DNA区域(约200bp)的序列。图70B:步骤2，定位靶PAM。鉴定接近编辑位置的PAM。一定在两条链上寻找PAM。虽然优选接近编辑位置的PAM，但可使用前间隔序列和将切口放置在距编辑位置≥30nt处的PAM安装编辑。图70C:步骤3，定位切口位点。对于正在考虑的各PAM，鉴定相应切口位点。对于Sp Cas9 H840A切口酶，切割发生在含PAM的链中NGG PAM 5′的第3个和第4个碱基之间。所有编辑核苷酸必须存在于切口位点的3′，因此适当的PAM必须将切口置于含PAM的链上的靶编辑的5′处。在下面显示的示例中，有两种可能的PAM。为简单起见，其余步骤将显示仅使用PAM1的PEgRNA设计。图70D:步骤4，设计间隔区序列。Sp Cas9的前间隔序列对应于含PAM的链上NGG PAM的5′的20个核苷酸。有效的PolIII转录起始需要G作为第一个转录的核苷酸。如果前间隔序列的第一个核苷酸是G，则PEgRNA的间隔区序列就是前间隔序列。如果前间隔序列的第一个核苷酸不是G，则PEgRNA的间隔区序列是G，后面为前间隔序列。图70E:步骤5，设计引物结合位点(PBS)。使用起始等位基因序列，鉴定含PAM的链上的DNA引物。DNA引物的3′端正好是切口位点上游的核苷酸(即，Sp Cas9的NGG PAM的5′的第4个碱基)。作为与PE2和PE3一起使用的一般设计原则，包含与DNA引物的互补性的12至13个核苷酸的PEgRNA引物结合位点(PBS)可用于包含约40-60％GC含量的序列。对于具有较低GC含量的序列，应测试较长(14至15nt)的

PBS。对于具有较高GC含量的序列，应测试较短(8至11nt)的PBS。最佳PBS序列应根据经验确定，无论GC含量如何。为设计长度为p的PBS序列，使用起始等位基因序列，取含PAM的链中切口位点5′的前p个核苷酸的反向互补序列。图70F：第6步，设计RT模板。RT模板编码设计的编辑和与邻近编辑的序列的同源性。最佳RT模板长度因靶位点而异。对于短程编辑(位置+1至+6)，建议测试短(9至12nt)、中(13至16nt)和长(17至20nt)RT模板。对于远程编辑(+7位及以上)，建议使用RT模板，该模板在编辑位置后至少延伸5nt(优选是10nt或更多)，以允许足够的3′DNA瓣同源性。对于远程编辑，应筛选几个RT模板以鉴定功能设计。对于较大的插入和缺失(≥5nt)，建议将更大的3′同源性(约20nt或更多)掺入RT模板。当RT模板将G的合成编码为逆转录DNA产物中的最后一个核苷酸(对应于PEgRNA的RT模板中的C)时，通常会损害编辑效率。由于许多RT模板支持有效的引导编辑，因此设计RT模板时建议避免G作为最终合成的核苷酸。为了设计长度为r的RT模板序列，使用期望的等位基因序列，并取原始含PAM的链中切口位点3′的前r个核苷酸的反向互补序列。注意，与SNP编辑相比，使用相同长度的RT模板得到的插入或缺失编辑不会包含相同的同源性。图70G：步骤7，组装完整的PEgRNA序列。按以下顺序(5′至3′)使PEgRNA组件连环化(concatenate)：间隔区、支架、RT模板和PBS。图70H：步骤8，为PE3设计切口产生性sgRNA。鉴定编辑上游和下游的非编辑链上的PAM。最佳切口产生性位点高度依赖于基因座，应根据经验确定。一般地，置于PEgRNA诱导的切口对面的位点5′的40到90个核苷酸的切口导致更高的编辑产率和更少的iindel。切口产生性sgRNA具有与起始等位基因中的20-nt前间隔序列匹配的间隔区序列，如果前间隔序列不以G开头，则添加5′-G。图70I：步骤9，设计PE3b切口产生性sgRNA。如果PAM存在于互补链中，并且其相应的前间隔序列与靶向编辑的序列重叠，则此编辑可以是PE3b系统的候选者。在PE3b系统中，切口产生性sgRNA的间隔区序列匹配期望的经编辑的等位基因的序列，但不匹配起始等位基因的序列。当经编辑的核苷酸落在切口产生性sgRNA前间隔序列的种子区(邻近PAM的约10nt)内时，PE3b系统有效运行。这防止在安装编辑链之前对互补链产生切口，从而防止PEgRNA和sgRNA之间竞争结合靶DNA。PE3b还避免两条链上同时切口的生成，从而在保持高编辑效率的情况下显著减少indel形成。PE3b sgRNA应当具有与期望等位基因中的20-nt前间隔序列匹配的间隔区序列，在需要时添加5′G。

[0185] 图71A示出SpCas9 PEgRNA分子的核苷酸序列(上部)，其在3′端终止于“UUU”且不包含趾环元件。该图的下部描绘相同的SpCas9 PEgRNA分子，但进一步修饰为包含趾环元件，该趾环元件具有正好在“UUU”3′端之前插入的序列5′-“GAAANNNN”-3′。“N”可以是任何核碱基。

[0186] 图71B表明使用含有趾环元件的PEgRNA提高HEK细胞或EMX细胞中的引导编辑效率，而indel形成百分比基本没有变化。

[0187] 图72A至72C描绘可在引导编辑中使用的替代PEgRNA配置。图72A描绘引导编辑的PE2:PEgRNA实施方案。该实施方案涉及与PEgRNA复合的PE2(包含Cas9和逆转录酶的融合蛋白)(也如图1A-1I和/或图3A-3E所述)。在该实施方案中，将用于逆转录的模板掺入sgRNA上的3′延伸臂以制备PEgRNA，并且DNA聚合酶是直接与Cas9融合的逆转录酶(RT)。图72B描绘MS2cp-PE2:sgRNA+tPERT实施方案。该实施方案包括PE2融合物(Cas9+逆转录酶)，其进一步与MS2噬菌体外壳蛋白(MS2cp)融合形成MS2cp-PE2融合蛋白。为实现引导编辑，MS2cp-PE2融合蛋白与sgRNA复合，该sgRNA将复合物靶向到DNA中的特定靶位点。然后，该实施方案涉

及引入反式引导编辑RNA模板(“tPERT”),其通过在分开的分子(即,tPERT)上提供引物结合位点(PBS)和DNA合成模板来代替PEgRNA运行,所述分开的分子还配备了MS2适体(茎环)。MS2cp蛋白通过与分子的MS2适体结合来募集tPERT。图72C描绘可通过核酸分子化学合成的已知方法实现的PEgRNA的替代设计。例如,化学合成可用于合成在引导编辑中使用的杂合RNA/DNAPEgRNA分子,其中杂合PEgRNA的延伸臂是DNA而不是RNA。在这样的实施方案中,可使用DNA依赖性DNA聚合酶代替逆转录酶来合成3′DNA瓣,其包含通过引导编辑形成的期望的基因变化。在另一个实施方案中,可合成延伸臂来包含化学接头,其防止DNA聚合酶(如,逆转录酶)使用sgRNA支架或骨架作为模板。在又一个实施方案中,延伸臂可包含相对于PEgRNA分子的整体定向具有相反定向的DNA合成模板。例如,如以5′至3′定向对PEgRNA所示且利用连接至sgRNA支架3′端的延伸,DNA合成模板以相反反向定向,即3′至5′方向。该实施方案对于具有位于gRNA的3’端处的延伸臂的PEgRNA实施方案可能是有利的。通过反转延伸臂的定向,聚合酶(如,逆转录酶)的DNA合成将在到达延伸臂新定向的5’端时终止,因此没有使用gRNA核心作为模板的风险。

[0188] 图73显示利用tPERT和MS2募集系统(又名MS2标签化技术)的引导编辑。将引导编辑器蛋白(PE2)靶向到靶基因座的sgRNA与tPERT组合表达,所述tPERT包含引物结合位点(13-nt或17-ntPBS)、编码His6标签插入和同源臂的RT模板,以及MS2适体(位于tPERT分子的5′或3′端)。使用引导编辑器蛋白(PE2)或MS2cp与PE2的N端的融合物。编辑是在有或没有互补链切口产生性sgRNA的情况下进行的,如在先前开发的PE3系统中(在x轴上分别指定为标签“PE2+切口”或“PE2”)。这在本文中又称为并定义为“第二链切口产生”。

[0189] 图74表明逆转录酶的MS2适体以反式表达及其利用MS2适体系统募集。PEgRNAPEgRNA包含插入至两个sgRNA支架发夹之一的MS2 RNA适体。野生型M-MLV逆转录酶表达为与MS2外壳蛋白(MCP)的N端或C端融合物。编辑位于HEK293T细胞中的HEK3位置。

[0190] 图75提供比较PE2、PE2-截短、PE3和PE3-截短在不同细胞系的不同靶位点上的效率(即,“具有指定编辑或indel的总测序读段%”)的柱状图。数据表明包含截短的RT变体的引导编辑器与包含非截短的RT蛋白的引导编辑器效率差不多。

[0191] 图76显示内含肽-断裂引导编辑器的编辑效率。利用编码全长PE2或内含肽-断裂PE2、PEgRNA和切口产生性向导RNA的质粒转染HEK239T细胞。显示了共有序列(C端外显肽的大部分氨基端残基)。示出两个位置处的编辑百分比:HEK3+1CTT插入和PRNP+6G至T。重复n=3独立转染。

[0192] 图77显示内含肽断裂引导编辑器的编辑效率。在通过ICV注射向P0小鼠递送5E10vg/SpPE3半部和少量1E10核定位的GFP:KASH后,通过大块皮层和GFP+亚群中的靶向深度测序评估编辑。将编辑器和GFP包装在具有EFS启动子的AAV9中。注射后3周收获小鼠并通过流式细胞术分离GFP+细胞核。显示单个数据点,每个条件分析了1-2只小鼠。

[0193] 图78显示内含肽-断裂引导编辑器的编辑效率。具体地,该图描绘在实施例20中使用的AAV断裂-SpPE3构建体。通过分别表达SpPE3-N和SpPE3-C的AAV颗粒的共转导重演PE3活性。注意N端基因组包含表达切口产生性sgRNA的U6-sgRNA盒,标签C端基因组包含表达PEgRNA的U6-PEgRNA盒。

[0194] 图79示出某些优化接头的编辑效率。特别地,数据显示针对转换、颠换、插入和缺失编辑,对于代表性PEgRNA,在HEK3、EMX1、FANCF、RNF2基因座处,相比于具有用如所示序列

替换的接头的各种版本,具有当前接头的PE2构建体的编辑效率(标注为PE2-白框)。替换接头被称为“1xSGGS”(SEQ ID NO:174)、“2xSGGS”(SEQ ID NO:446)、“3xSGGS”(SEQ ID NO:3889)、“1xXTEN”(SEQ ID NO:171)、“无接头”、“1xGly”、“1xPro”、“1xEAAAK”(SEQ ID NO:3968)、“2xEAAAK”(SEQ ID NO:3969)和“3xEAAAK”(SEQ ID NO:3970)。编辑效率以相对于PE2的“对照”编辑效率的柱状图形式测量。PE2的接头是SGGSSGGSSGSETPGTSESATPESSGGSSGGSS(SEQ ID NO:127)。所有编辑都在PE3系统的背景下完成,即,指PE2编辑构建体加上添加最佳二级sgRNA切口产生性向导物。

[0195] 图80.取相对于PE2的平均倍数功效得到所示图,表明使用1xXTEN(SEQ ID NO:171)接头序列将编辑效率平均提高1.14倍(n=15)。

[0196] 图81描绘来自不同启动子的PEgRNA的转录水平。

[0197] 图82描绘对PEgRNA结构的不同类型的修饰相对于未修饰PEgRNA对编辑效率的影响。

[0198] 图83描绘靶向编辑HEK3基因的PE实验,特别是靶向在相对于切口位点的位点+1处10nt插入的插入并且使用PE3。

[0199] 图84A描绘具有间隔区、gRNA核心和延伸臂(RT模板+引物结合位点)的示例性PEgRNA,其在PEgRNA的3’端用tRNA分子修饰,通过UCU接头偶联。tRNA包括各种转录后修饰。然而,所述修饰不是必需的。

[0200] 图84B描绘可用于修饰PEgRNA结构的tRNA结构。P1的长度可以是可变的。可以延伸P1以有助于防止PEgRNA-tRNA融合物的RNAseP加工。

[0201] 图85描绘靶向编辑FANCF基因的PE实验,特别是靶向在相对于切口位点的位点+5处的G至T转化并且使用PE3构建体。

[0202] 图86描绘靶向编辑HEK3基因的PE实验,特别是靶向在相对于切口位点的位点+1处71ntFLAG标签插入的插入并且使用PE3构建体。

[0203] 图87是来自N2A细胞中的筛选结果,其中pegRNA安装了1412Adel,具有关于引物结合位点(PBS)长度和逆转录酶(RT)模板长度的详细信息(显示有和没有indel)。

[0204] 图88是来自N2A细胞中的筛选结果,其中pegRNA安装了1412Adel,具有关于引物结合位点(PBS)长度和逆转录酶(RT)模板长度的详细信息(显示有和没有indel)。

[0205] 图89描绘在β-珠蛋白基因中的代理(proxy)基因座处和健康HSC中的HEK3处编辑的结果,改变编辑器对pegRNA和切口产生性gRNA的浓度。

[0206] 图90提供了双瓣引导编辑的实施方案的示意图。通过两个引导编辑复合物(由pegRNA-A和pegRNA-B向导)作用于DNA靶序列。两个pegRNA靶向双螺旋的相对链。每个引导编辑器(PE2•pegRNA)在单个DNA链上产生切口,随后使用pegRNA作为模板合成3’DNA瓣。两个引导编辑器复合物的作用导致在DNA的相对链上产生包含两个3’瓣的中间体。两个3’瓣在其3’端彼此互补。3’瓣的3’端的退火导致形成双重双链体结构,一个双链体由含有新DNA序列(红色)的成对3’瓣形成,另一个双链体由含有原始DNA序列(黑色)的成对5’瓣形成。切除插入的原始DNA双链体(黑色成对5’瓣)得到含有期望的新DNA序列(红色)替代原始DNA序列的双切口DNA种类。连接两个切口完成编辑过程。

[0207] 图91A至91B提供了实施例7的结果。图91A示出使用双引导编辑器在HEK293T细胞的HEK3位点用新的22bp序列替换90bp序列的Crispresso2输出等位基因表(针对期望的产

物进行比对)。期望的产物占测序读数的80％以上。为了比较,参考起始等位基因显示在测序等位基因的上方。图91B示出了用于实现图91A所示的序列替换的pegRNA的序列。pegRNA1和pegRNA2靶向DNA双螺旋的不同链,并且产生5’置换切口,如图90中所绘制的。

[0208] 图92示出针对双瓣引导编辑的pegRNA设计的设计实施方案。用于双瓣引导编辑的两个pegRNA在图中显示为pegRNA A和pegRNA B。每个pegRNA含有间隔区序列(深蓝色),其将引导编辑复合物向导到靶标DNA位点。两个pegRNA靶向DNA双螺旋的相对链。与其他pegRNA一样,双瓣引导编辑pegRNA含有3’延伸和逆转录模板(RT模板,浅蓝色),该3’延伸具有与带切口的基因组DNA链退火以启动逆转录的引物结合序列(PBS,绿色),该逆转录模板作为逆转录酶合成新DNA的模板。与用于经典引导编辑的pegRNA(其需要新合成的编辑3’瓣与内源性5’瓣竞争),不需要编码与RT模板内的靶位点的同源性。相反,两个合成的3’瓣的3’端仅需要彼此含有互补性(即3’瓣的3’端是彼此的反向互补序列)。这种互补允许两个3’瓣退火并促进期望的编辑的DNA序列的形成。

[0209] 图93A至93B示出了使用双引导编辑来安装具有双瓣引导编辑的Bxb1attB和attP位点的实施例7的结果。图93A示出了在HEK3上安装38bp Bxb1attB位点。构建了六个pegRNA,三个靶向(+)链(A1,A2和A3),三个靶向(-)链(B1,B2和B3)。它们的区别在于RT模板中编码的attB序列的量,导致两个瓣之间的互补核苷酸数量不同。评估pegRNA的3x 3矩阵,以在HEK293细胞中的靶基因组位置处安装attB序列。图93A和图93B示出了在HEK3上安装50bp Bxb1 attP位点。构建了六个pegRNA,三个靶向(+)链(A1,A2和A3),三个靶向(-)链(B1,B2和B3)。它们的区别在于RT模板中编码的attP序列的量,导致两个瓣之间的互补核苷酸数量不同。评估pegRNA的3x 3矩阵,以在HEK293细胞中的靶基因组位置处安装attP序列。对于图93A和图93B中的两个编辑,attB或attP位点的安装伴随同时删除位于两个切口位点之间的90bp的基因组DNA序列。

[0210] 图94示出了用双瓣引导编辑在人类安全港基因座安装Bxb1 attB和attP位点的结果。在HEK293T细胞中,在AAVS1基因座(左)安装Bxb1 attP位点或在CCR5基因座(右)安装Bxb1 attB位点。正确编辑以蓝色(AAVS1)和红色(CCR5)示出,而插入缺失副产物以灰色显示。

[0211] 图95提供了使用四瓣引导编辑的基因组倒置的示意图。基因组DNA的区域被靶向以进行倒置(绿色和橙色片段)。用PE2引导编辑器将四个pegRNA递送至细胞。一对pegRNA靶向单个基因组DNA链并且作为合成两个互补DNA瓣(A和A’,蓝色)的模板,而第二对靶向另一个基因组DNA链并且作为合成具有正交的DNA序列的两个互补DNA瓣(B和B’,粉色)的模板。互补瓣退火形成3′突出端双链体。通过内源性细胞修复酶切除5′突出端双链体。连接切口以产生产物等位基因,其含有倒置DNA序列(绿色和橙色片段)和倒置接合处的pegRNA模板的序列(蓝色和粉色片段)。

[0212] 图96示出了AAVS1倒置接合处的扩增子测序。使用四瓣引导编辑策略对HEK293T细胞中AAVS1基因座的2.7kb倒置进行CRISPResso2分析输出。预计的倒置接合处的PCR扩增和测序显示了在倒置接合处插入Bxb1 attP或attB序列的期望的产物。

[0213] 图97A-97B示出了使用四瓣引导编辑将环状DNA质粒靶向整合到基因组中的结果。(图97A)靶向基因组DNA的区域和质粒DNA的区域以进行四瓣引导编辑整合。用PE2引导编辑器将四个pegRNA递送至细胞。两个pegRNA模板化互补序列,一个靶向单个基因组DNA链,另

一个靶向单个质粒DNA链(产生蓝色瓣)。另外两个pegRNA靶向与前两个pegRNA相对的基因组DNA和质粒DNA链,并且它们为合成两个互补DNA瓣(粉红色)提供模板,这两个互补DNA瓣与第一对正交。互补瓣退火形成3′突出端双链体。通过内源性细胞修复酶切除5′突出端双链体。连接切口以产生产物等位基因,其含有整合的质粒DNA序列(绿色和橙色片段)和整合接合处的pegRNA模板化序列(蓝色和粉色片段)。(图97B)预期接合处的扩增子测序的CRISPResso2分析显示通过pegRNA模板的attP序列桥接的质粒骨架和基因组DNA序列。

[0214] 图98A至98B示出了用四瓣引导编辑进行靶向染色体易位的结果。(图98A)pegRNA靶向不同染色体上的两个区域。互补的3' DNA瓣桥接了两个染色体序列并指导易位的方向。(图98B)HEK239T细胞中MYC和TIMM44基因座之间的靶向易位。染色体8上的MYC基因座和染色体19上的TIMM44基因座之间易位的预期连接点的扩增子测序的CRISPResso2分析输出。大多数测序读数对应于期望的等位基因序列。

[0215] 图99示出了用双瓣引导编辑在IDS基因座安装Bxb1 attB和attP位点。HEK293T细胞用PE2和不同对的pegRNA(例如,在第一列中,具有用于正向安装attP位点的模板的pegRNAA1_a和pegRNAB2_a)转染。通过HTS测量效率。该数据表明双瓣编辑可以以高达～80%的效率将感兴趣的序列插入到IDS基因座。

[0216] 图100A至100B描述了双瓣介导的复制。图100A是示出在293T细胞的AAVS1基因座处双瓣介导的复制的示意图。图100B示出了使用双瓣pegRNA和PE2在AAVS1处诱导遗传序列复制的结果。

[0217] 图101示出了多瓣诱导的新易位MYC-CCR5。MYC-CCR5易位通过四瓣pegRNA和PE2诱导。MYC-CCR5易位事件由四重pegRNA诱导。四组不同的pegRNA在HEK293T细胞中测试。衍生的chr8和chr3之间的易位连接产物通过连接引物扩增。图表中显示了与预计连接等位基因比对的读数百分比。结果表明,四瓣可以介导MYC和CCR5基因的易位,产物纯度在接合1处接近100%,在接合2处约为50%。代表性等位基因图显示了与接合1处的预计等位基因序列比对的序列。

[0218] 图102A至102B示出了在其他人细胞系中的双瓣和多瓣编辑。图102A示出了在四个不同的人细胞系中的双瓣编辑。HEK293T和HeLa细胞用双重pegRNA和PE2转染以编辑三个不同的基因组基因座(IDS、MYC和TIMM44)。U2OS和K562细胞用相同组分进行核转染。双瓣在所有四种人细胞中的靶基因座处都表现出强大的编辑功能,特别是在HEK293T和K562细胞中。能够进行双瓣编辑的细胞机制在许多人细胞类型中是保守的。图102B显示了HeLa细胞中AAVS1处的多瓣(四重pegRNA)指导的2.7kb倒置。

[0219] 图103A至103B示出了通过HTS在CCR5基因座进行的倒置效率测量的结果。预计倒置编辑等位基因的百分比通过HTS测量。将四个四重pegRNA组PE2分别转染到HEK293T细胞中。图103A示出了HEK293T细胞中CCR5基因座的双瓣介导的序列复制(～100nt)。通过HTS 300循环对端测序分析,编辑效率达到了～1.5%。图103B示出了Hela细胞中CCR5基因座的四瓣介导的序列倒置(～95-117nt)。通过HTS 300循环对端测序分析,编辑效率达到了～1.2%。这一结果表明多瓣能成功地精确介导CCR5基因座处的复制和倒置。当靶标序列被复制时,编辑特异性高(插入缺失百分比<2%)。

[0220] 图104A至104E示出了双瓣编辑的靶向细胞修复途径。在图104A-104D中,将HEK293T细胞用质粒转染,该质粒表达Exo1、Fen1、红色荧光蛋白(对照)、DNA2、Mlh1 neg和

具有pegRNA和PE2的P53抑制剂。通过HTS测量编辑效率。比较候选物和RFP对照之间的编辑效率。在图104E中，将HEK293T细胞用siRNA质粒和针对每个靶基因座的pegRNA和PE2转染。使用非靶向的siRNA(siNT)作为对照。通过HTS测量编辑效率。在每个靶基因座上比较每个siRNA敲低和siNT对照之间的编辑效率。将HEK293T细胞分别用双重pegRNA、PE2和表达Exo1、Fen1、红色荧光蛋白(对照)、DNA2、Mlh1 neg和P53抑制剂转染。通过HTS测量编辑效率。比较候选物和RFP对照之间的编辑效率。使用双尾配对Studentt-检验来测量每个处理和RFP对照之间的统计差异(P<0.05,*;P<0.01,**;P<0.001,***)。FEN1的过表达提高了所有四个靶向位点(MYC、TIMM44、IDS、CCR5)的双活编辑效率。

[0221] 图105A至105B示出了AAVS1基因座处的双瓣介导的序列复制。图105A示出AAVS1基因座处的双瓣介导的序列复制的示意图。图105B示出了通过使用产生两个独特的3'瓣结构的双重pegRNA，在293T细胞中AAVS1基因座处诱导了～300bp的序列复制。用特异性引物扩增预期的等位基因，并进行HTS。～94％的读数与重复的预期等位基因一致。在未处理的样品中没有观察到复制产物。

[0222] 图106示出了使用四瓣引导编辑的靶向IDS基因组序列倒置。～13％的亨特氏综合征患者显示有IDS基因序列的倒置(Bondeson et al.,Human Molecular Genetics,1995)。应用四瓣引导编辑在HEK293T细胞中诱导这种～40kb IDS基因组序列的致病性倒置。通过用pegRNA和PE2转染HEK293T细胞来测试六组四重peg RNA。引物用于特异性扩增连接点“ab”和连接点“cd”处的倒置序列。～95％的预期倒置等位基因序列在IDS_QF1的两个连接处都观察到。其他组的pegRNA也在两个连接处产生高百分比的预期等位基因序列。在未处理的样品中没有观察到倒置连接产物。

[0223] 图107显示PegRNA3′基序修饰改善了IDS基因座处的双瓣编辑效率。为进一步改进双瓣编辑效率，引入了假结evoPreQ1基序来保护pegRNA 3'末端。通过比较未修饰和evoPreQ1修饰的双重pegRNA产生的编辑效率，在靶向IDS基因座处，修饰的peg RNA的编辑效率总体上有所增加。双瓣编辑效率的改善可以高达5.3倍。

[0224] 图108A至108C显示了twinPE和twinPE介导的序列替换的概况；图108A显示了twinPE系统靶向基因组DNA序列，该序列在DNA的相对链上含有两个前间隔序列序列。PE2•pegRNA复合物靶向每个前间隔序列，产生单链切口，并逆转录含有期望的插入序列的pegRNA编码模板。合成并释放3'DNA瓣后，存在一个假设的中间体，它具有包含编辑DNA序列的退火的3'瓣和包含原始DNA序列的退火的5'瓣。切除包含在5'瓣中的原始DNA序列，然后将3'瓣连接到相应的切除位点，产生期望的编辑产物。图108B示出了用38bp的Bxb1 attB序列替换HEK位点3中的90bp序列的twinPE介导的示例。图108C显示了在HEK293T细胞中使用模板化不同长度的插入序列的pegRNA来评估twinPE的图108B所示38bp Bxb1 attB位点或HEK位点3的50bp Bxb1attP位点的安装。数值和误差线反映了三个独立生物重复的平均值和标准差。

[0225] 图109A至109E示出了人细胞中使用twinPE的靶向序列插入、缺失和再编码。图109A显示在HEK293T细胞的HEK位点3插入具有PE3(12bp、36bp、108bp或321bp)或twinPE(108bp)的FKBP编码序列片段。图109B显示使用twinPE对HEK293T细胞中PAH的外显子4和7内的序列进行再编码。使用24、36或59bp的重叠瓣编辑外显子4中的64bp靶序列，使用22或42bp的重叠活编辑外显子7中的46bp靶序列，或者使用24或47bp的重叠瓣编辑外显子7中的

64bp序列。使用含有3’evoPreQ1基序的标准pegRNA或epegRNA比较编辑活性。图109C是三种不同的双瓣缺失策略的示意图，其被研究用于进行靶向缺失。“基本锚定(BA)”twinPE策略允许从一个缺口位点的任意位置3’开始到另一个缺口位点结束的灵活缺失。“杂交锚定(HA)”twinPE策略允许在两个缺口位点之间任意选择的位置灵活缺失序列。此处测试的“引导缺失(PrimeDel,PD)”策略允许从一个缺口位点开始并在另一个缺口位点结束的序列缺失。图109D显示了使用靶向相同前间隔序列对的BA-twinPE、HA-twinPE或PD策略在HEK293T细胞中HEK位点3处的序列缺失。使用含有3’evoPreQ1基序的标准pegRNA或epegRNA比较编辑活性。图109E显示使用BA-twinPE、PD、成对Cas9核酸酶或twinPE介导的attB序列替换，在HEK293T细胞中的DMD基因座缺失外显子51序列。数值和误差线反映了三个独立生物重复的平均值和标准差。在DMD外显子51跳过实验中，至少进行了两次独立的生物重复。

[0226] 图110A至110E显示了人细胞中DNA货物与twinPE和Bxb1重组酶的位点特异性基因组整合。图110A显示了在HEK293T细胞中的AAVS1基因座处安装Bxb1 attP序列的twinPE pegRNA对的筛选。图110B显示了在HEK293T细胞中CCR5基因座处安装Bxb1 attB序列的twinPE pegRNA对的筛选。图110C显示了使用靶向CCR5(最左边的四个柱)或AAVS1(最右边的三个柱)的twinPE pegRNA对的5.6kb DNA供体的单转染敲入。twinPE pegRNA在CCR5处安装attB或在AAVS1处安装attP。Bxb1然后将带有相应附着位点的供体整合到基因组附着位点中。图110D显示了使用531/584双PE pegRNA对CCR5处单转染敲入的优化。模板编辑的同一性(attB与attP)、中央二核苷酸的同一性(野生型GT与正交突变型GA)以及瓣间重叠的长度进行改变以确定支持最高敲入效率的组合。图110E显示了在HEK293T和Huh7细胞系中，在ALB的内含子1中插入Bxb1 attB序列。图110F显示了HEK293T和Huh7细胞系中CCR5和ALB处单转染敲入效率的比较。

[0227] 图111A至111E显示了在人细胞中用twinPE和Bxb1重组酶进行的位点特异性基因组序列倒置。图111A是导致致病性39kb倒置的IDS和IDS2中的重组热点以及用于安装或校正IDS倒置突变的组合twinPE-Bxb1策略的示意图。图111B显示了在IDS和IDS2位点处pegRNA对的筛选，用于在IDS和IDS2位点安装具有特定DNA靶的attP或attB重组位点插入。图111C显示了具有顺序DNA转染的attP或attB插入的DNA测序分析。图111D显示了在倒置连接1和连接2(顺序转染)处的倒置产物纯度，表明在两个连接处成功倒置。图111E显示了连接处转化效率的定量(顺序转染和“一锅”RNA核转染)。

[0228] 图112显示了通过twinPE对HEK293T细胞中PAH的外显子10、11和12内的序列重新编码。使用28bp的重叠瓣编辑外显子10中64bp的靶序列，使用25bp的重叠瓣编辑外显子11中61bp和55bp的靶序列，或者分别使用27bp和24bp的重叠瓣编辑外显子12中68bp和58bp的序列。数值和误差线反映了三个独立生物重复的平均值和标准差。

[0229] 图113显示了用BxBI质粒和含有attP的供体DNA质粒转染含有纯合attB位点插入的HEK293T克隆细胞系。通过ddPCR测量，靶位点处的敲入效率在12-17％之间。

[0230] 图114显示了在twinPE和BxBI介导的一锅敲入后，含有attL和attR重组产物的预期连接序列的HTS测量。数值和误差线反映了三个独立生物重复的平均值和标准差。

[0231] 图115A显示了twinPE介导的attB插入效率，其中双重pegRNA中的瓣重叠长度减少。

[0232] 图115B显示了由特异性引物扩增的PCR产物，用于捕获琼脂糖凝胶上显示的供体

DNA和pegRNA质粒之间的重组。供体DNA和pegRNA质粒之间的重组随着瓣重叠的减少而减少。

[0233] 图116是开发的用于量化IDS转化效率的PCR策略的示意图。

[0234] 定义

[0235] 除非另有定义,本文使用的所有技术和科学术语均具有本发明所属领域的技术人员通常理解的含义。以下参考文献为技术人员提供了本发明中使用的许多术语的一般定义:Singleton et al.,DictionaryofMicrobiology and Molecular Biology(2nd ed.1994);The Cambridge Dictionary of Science and Technology(Walker ed.,1988);The Glossary ofGenetics,5th Ed.,R.Rieger et al.(编),Springer Verlag(1991);以及Hale&Marham,The Harper Collins Dictionary ofBiology(1991)。如本文所用,除非另有说明,否则以下术语具有赋予它们的含义。

[0236] 反义链

[0237] 在遗传学中,双链DNA内区段的“反义”链是模板链,并且认为以3’至5’方向延伸。相比之下,“有义”链是双链DNA内从5’至3’延伸的区段,它与DNA的反义链或模板链(从3’至5’延伸)互补。在编码蛋白质的DNA区段的情况下,有义链是与mRNA具有相同序列的DNA链,它在转录过程中以反义链为其模板,并最终(通常,并非总是)经历翻译成为蛋白。因而,反义链负责随后翻译成蛋白的RNA,而有义链具有与mRNA几乎相同的组成。注意对于dsDNA的每个区段,可能会有两组有义和反义,这取决于读取的方向(因为有义和反义是相对于视角而言的)。最终,规定dsDNA区段的哪条链称为有义或反义是基因产物或mRNA。

[0238] 双特异性配体

[0239] 如本文所用,术语“双特异性配体”或“双特异性部分”是指结合两个不同配体结合结构域的配体。在某些实施方案中,配体是小分子化合物或肽或多肽。在其他实施方案中,配体结合结构域是“二聚化结构域”,其可作为肽标签安装到蛋白上。在不同实施方案中,可以诱导各自包含相同或不同二聚化结构域的两种蛋白通过每个二聚化结构域与双特异性配体的结合二聚化。如本文所用,“双特异性配体”可等同地指“二聚化化学诱导物”或“CID”。

[0240] Cas9

[0241] 术语“Cas9”或“Cas9核酸酶”是指包含Cas9结构域或其片段(如,包含Cas9的活性或非活性DNA切割域,和/或Cas9的gRNA结合结构域的蛋白)。如本文所用,“Cas9结构域”是包含Cas9的活性或非活性切割域和/或Cas9的gRNA结合结构域的蛋白片段。“Cas9蛋白”是全长Cas9蛋白。Cas9核酸酶有时又被称为casn1核酸酶或CRISPR(成簇规律间隔短回文重复序列)相关核酸酶。CRISPR是适应性免疫系统,其可针对移动遗传元件(病毒、可转座元件和接合质粒)提供保护。CRISPR簇包含间隔区、与前面移动元件互补的序列和靶侵入核酸。将CRISPR簇转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,正确加工pre-crRNA需要反式编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9结构域。tracrRNA充当核糖核酸酶3辅助加工pre-crRNA的向导物。随后,Cas9/crRNA/tracrRNA以内切溶核方式切割与间隔区互补的线性或环状dsDNA靶标。首先以内切溶核方式切割不与crRNA互补的靶链,然后以3’-5’以外切溶核方式修剪。事实上,DNA结合和切割通常需要蛋白和这两种RNA。然而,可对单一向导RNA(“sgRNA”,或简称为“gNRA”)进行工程化改造,以便将crRNA和

tracrRNA两者的各个方面掺入单一RNA种类中。参见，例如，JinekM.,Chylinski K.,FonfaraI.,Hauer M.,Doudna J.A.,CharpentierE.Science 337:816-821(2012)，其全部内容通过引用并入本文。Cas9识别CRISPR重复序列(PAM或前间隔序列邻近基序)中的短基序，以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员所熟知的(参见，例如，“Complete genome sequence ofanM1 strain of Streptococcus pyogenes.”Ferretti et al.,J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,QianY.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001)；“CRISPR RNAmaturationby trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011)；以及“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,HauerM.,Doudna J.A.,Charpentier E.Science 337:816-821(2012)，其各自的全部内容通过引用并入本文)。Cas9直系同源物已在各种物种中得到描述，包括但不限于化脓性链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开，其他合适的Cas9核酸酶和序列对本领域技术人员是显而易见的，并且此类Cas9核酸酶和序列包括来自Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9 families oftype II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737(其全部内容通过引用并入本文)中公开的生物体和基因座的Cas9序列。在一些实施方案中，Cas9核酸酶包含一个或多个部分损害或失活DNA切割域的突变。

[0242] 核酸酶失活的Cas9结构域可互换地被称为“dCas9”蛋白(代表核酸酶-“死亡”Cas9)。用于产生具有无活性DNA切割域的Cas9结构域(或其片段)的方法是已知的(参见，例如，Jineketal.,Science.337:816-821(2012)；Qi etal.,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control ofGene Expression”(2013)Cell.28;152(5):1173-83，每篇文献的全部内容通过引用并入本文)。例如，已知Cas9的DNA切割域包括两个亚结构域，HNH核酸酶亚结构域和RuvC1亚结构域。HNH亚结构域切割与gRNA互补的链，而RuvC1亚结构域切割非互补链。这些亚结构域内的突变可使Cas9的核酸酶活性沉默。例如，突变D10A和H840A使化脓性链球菌Cas9的核酸酶活性完全失活(Jinek et al.,Science.337:816-821(2012)；Qi et al.,Cell.28;152(5):1173-83(2013))。在一些实施方案中，提供了包含Cas9片段的蛋白。例如，在一些实施方案中，蛋白包含两个Cas9结构域之一：(1)Cas9的gRNA结合结构域；或(2)Cas9的DNA切割域。在一些实施方案中，包含Cas9或其片段的蛋白被称为“Cas9变体”。Cas9变体与Cas9或其片段具有同源性。例如，Cas9变体与野生型Cas9(如，SEQ ID NO:18的SpCas9)至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、至少约99.8％相同、或至少约99.9％相同。在一些实施方案中，与野生型Cas9(如，SEQ ID NO:18的SpCas9)相比，Cas9变体可具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个氨基酸变化。在一些实施方案中，

Cas9变体包含SEQ ID NO:18Cas9的片段(如,gRNA结合结构域或DNA切割域),使得该片段与野生型Cas9的相应片段(如,SEQ ID NO:18的SpCas9)至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同。在一些实施方案中,该片段为相应野生型Cas9(如,SEQ ID NO:18的SpCas9)的氨基酸长度的至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％、或至少99.5％。

[0243] cDNA

[0244] 术语“cDNA”是指从RNA模板复制的DNA链。cDNA与RNA模板互补。

[0245] 环状排列体(permutant)

[0246] 如本文所用,术语“环状排列体”是指包含环状排列(permutation)的蛋白或多肽(如,Cas9),所述环状排列为蛋白结构构造的变化,涉及出现在蛋白氨基酸序列中的氨基酸顺序变化。换言之,环状排列体是与野生型对应物相比具有改变的N端和C端的蛋白,如蛋白的野生型C端半部变成新的N端半部。环状排列(或CP)本质上是蛋白一级序列的拓扑重排,通常使用肽接头连接其N和C端,同时在不同位置分割其序列以形成新的相邻N和C-端。结果是具有不同连接性的蛋白结构,但通常可具有相同的整体相似的三维(3D)形状,并且可能包括改进或改变的特征,包括降低的蛋白水解易感性、提高的催化活性、改变的底物或配体结合,和/或提高的热稳定性。环状排列的蛋白在自然界中可以存在(如,伴刀豆球蛋白A和凝集素)。此外,环状排列可作为翻译后修饰的结果出现,或可使用重组技术进行工程化改造。

[0247] 环状排列的Cas9

[0248] 术语“环状排列的Cas9”是指已作为环状排列体存在,由此其N-和C-端已局部重排的任何Cas9蛋白或其变体。这种环状排列的Cas9蛋白(“CP-Cas9”)或其变体在与向导RNA(gRNA)复合时保留了结合DNA的能力。参见Oakes et al.,“Protein Engineering ofCas9 for enhanced function,”Methods Enzymol,2014,546:491-511和Oakes et al.,“CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification,”Cell,January 10,2019,176:254-267,以上的每一篇通过引用并入本文。本公开考虑任何先前已知的CP-Cas9或使用新的CP-Cas9,只要所得到的环状排列的蛋白在与向导RNA(gRNA)复合时保留结合DNA的能力。示例性CP-Cas9蛋白是SEQ ID NO:77-86。

[0249] CRISPR

[0250] CRISPR是细菌和古细菌中的DNA序列家族(即,CRISPR簇),其代表已侵入原核生物的病毒在先感染的片段。原核细胞使用DNA片段来检测和破坏DNA免受类似病毒的后续攻击,并与一系列CRISPR相关蛋白(包括Cas9及其同源物)和CRISPR相关RNA一起有效地构成原核免疫防御系统。事实上,CRISPR簇被转录并加工成CRISPRRNA(crRNA)。在某些类型的CRISPR系统(如,II型CRISPR系统)中,pre-crRNA的正确加工需要反式编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当核糖核酸酶3辅助加工pre-crRNA的向导物。随后,Cas9/crRNA/tracrRNA以内切溶核方式切割与RNA互补的线性或环状dsDNA靶标。具体地,首先以内切溶核方式切割不与crRNA互补的靶链,然后以3’-5’以外切溶核方式修剪。事实上,DNA结合和切割通常需要蛋白和这两种RNA。然而,可对单一向

导RNA(“sgRNA”,或简称为“gNRA”)进行工程化改造,以将crRNA和tracrRNA两者的各个方面掺入单个RNA种类——向导RNA中。参见,例如,JinekM.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science337:816-821(2012),其全部内容通过引用并入本文。Cas9识别CRISPR重复序列(PAM或前间隔序列邻近基序)中的短基序,以帮助区分自我与非自我。CRISPR生物学以及Cas9核酸酶序列和结构是本领域技术人员所熟知的(参见,例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti et al.,J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature471:602-607(2011);以及“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,HauerM.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其全部内容通过引用并入本文)。已在不同物种中描述了Cas9直系同源物,包括但不限于化脓性链球菌和嗜热链球菌。基于本公开,其他合适的Cas9核酸酶和序列对本领域技术人员而言是显而易见的,并且这类Cas9核酸酶和序列包括来自Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737(其全部内容通过引用并入本文)中的Cas9序列。

[0251] 在某些类型的CRISPR系统(如,II型CRISPR系统)中,pre-crRNA的正确加工需要反式编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当核糖核酸酶3辅助加工pre-crRNA的向导物。随后,Cas9/crRNA/tracrRNA以内切溶核方式切割与RNA互补的线性或环状核酸靶标。具体地,首先与内切溶核方式切割不与crRNA互补的靶链,然后以3’-5’以外切溶核方式修剪。事实上,DNA结合和切割通常需要蛋白和这两种RNA。然而,可以对单个向导RNA(“sgRNA”,或简称为“gRNA”)进行工程化改造,以将crRNA和tracrRNA两者的实施方案并入单一RNA种类——向导RNA中。

[0252] 通常,“CRISPR系统”统称为参与CRISPR相关(“Cas”)基因的表达或指导其活性的转录本和其他元件,包括编码Cas基因的序列、tracr(反式-激活CRISPR)序列(如,tracrRNA或活性部分tracrRNA)、tracr配对序列(包括在内源性CRISPR系统背景下的“直接重复”和tracrRNA加工的部分直接重复)、向导序列(内源性CRISPR系统背景下又被称为“间隔区”),或来自CRISPR基因座的其他序列和转录本。该系统的tracrRNA与向导RNA上存在的tracr配对序列互补(完全或部分)。

[0253] DNA合成模板

[0254] 如本文所用,术语“DNA合成模板”是指PEgRNA的延伸臂的区域或部分,其被引导编辑器的聚合酶用作模板链来编码3’单链DNA瓣,其含有期望的编辑,然后通过引导编辑机制置换靶位点处相应内源性DNA链。在不同实施方案中,DNA合成模板示于图3A(在包含5’延伸臂的PEgRNA背景下),图3B(在包含3’延伸臂的PEgRNA背景下),图3C(在内部延伸臂背景

下),图3D(在3'延伸臂背景下),以及图3E(在5′延伸臂背景下)。延伸臂(包括DNA合成模板)可由DNA或RNA组成。在RNA的情况下,引导编辑器的聚合酶可以是RNA依赖性DNA聚合酶(如,逆转录酶)。在DNA的情况下,引导编辑器的聚合酶可以是DNA依赖性DNA聚合酶。在不同实施方案中(例如,如图3D-3E中所示),DNA合成模板(4)可包含"编辑模板"和"同源臂",以及可选的5'端修饰区e2的全部或部分。也就是说,取决于e2区的性质(如,它是否包括发夹、趾环或茎/环二级结构),聚合酶也可编码无一、某个或整个的e2区。换言之,在3'延伸臂的情况下,DNA合成模板(3)可包括跨越引物结合位点(PBS)的5'端至gRNA核心的3'端的延伸臂(3)的部分,其可作为聚合酶(如,逆转录酶)合成DNA单链的模板运行。在5′延伸臂的情况下,DNA合成模板(3)可包括跨越PEgRNA分子的5′端至编辑模板的3′端的延伸臂(3)的部分。优选地,DNA合成模板不包括具有3'延伸臂或5'延伸臂的PEgRNA的引物结合位点(PBS)。在此描述的某些实施方案(如,图71A)指"RT模板",其包括编辑模板和同源臂,即在DNA合成过程中实际用作模板的PEgRNA延伸臂的序列。术语"RT模板"等同于术语"DNA合成模板"。

[0255] 在反式引导编辑的情况下(如,图3G和图3H),引物结合位点(PBS)和DNA合成模板可被工程化改造为分开的分子,被称为反式引导编辑器RNA模板(tPERT)。

[0256] 二聚化结构域

[0257] 术语"二聚化结构域"是指与双特异性配体的结合部分结合的配体结合结构域。"第一"二聚化结构域结合双特异性配体的第一结合部分,并且"第二"二聚化结构域结合同一双特异性配体的第二结合部分。当第一二聚化结构域与第一蛋白融合(如,通过PE,如本文所讨论的)并且第二二聚化结构域与第二蛋白融合(如,通过PE,如本文所讨论的)时,第一和第二蛋白在双特异性配体的存在下二聚化,其中所述双特异性配体具有至少一个与第一二聚化结构域结合的部分和至少另一个与第二二聚化结构域结合的部分。

[0258] 下游

[0259] 如本文所用,术语"上游"和"下游"是相对性术语,其定义了位于以5'至3'方向定向的核酸分子(无论是单链还是双链)中的至少两个元件的线性位置。具体地,第一元件在核酸分子中第二元件的上游,其中第一元件位于第二元件5'的某处。例如,如果SNP位于切口位点的5'侧,则SNP位于Cas9诱导的切口位点的上游。相反地,第一元件位于核酸分子中第二元件的下游,其中第一元件位于第二元件3'的某处。例如,如果SNP位于切口位点的3'侧,则SNP位于Cas9诱导的切口位点的下游。核酸分子可以是DNA(双链或单链)。RNA(双链或单链),或DNA和RNA的杂合体。单链核酸分子和双链分子的分析相同,因为上游和下游的术语仅参考核酸分子的单链,除了考虑需要选择双链分子的哪条链。通常,可用于确定至少两个元件的位置相对性的双链DNA的链是"有义"或"编码"链。在遗传学中,"有义"链是双链DNA中从5'至3'延伸,且与DNA从3'至5'延伸的反义链或模板链互补的区段。因此,例如,如果SNP核碱基在有义链或编码链的启动子的3'侧,则SNP核碱基在基因组DNA(双链)中启动子序列的"下游"。

[0260] 编辑模板

[0261] 术语"编辑模板"是指在由聚合酶合成的单链3'DNA瓣中编码期望编辑的延伸臂的一部分,例如DNA依赖性DNA聚合酶、RNA依赖性DNA聚合酶(如,逆转录酶)。在此描述的某些实施方案(如,图71A)是指"RT模板",其指编辑模板和同源臂两者,即,在DNA合成过程中实际用作模板的PEgRNA延伸臂的序列。术语"RT编辑模板"也等同于术语"DNA合成模板",但其

中RT编辑模板反映了使用具有作为逆转录酶的聚合酶的引导编辑器，其中DNA合成模板更广泛地反映了使用具有任何聚合酶的引导编辑器。

[0262] 有效量

[0263] 如本文所用，术语“有效量”是指足以引发期望的生物反应的生物活性剂的量。例如，在一些实施方案中，引导编辑器(PE)的有效量可指足以编辑靶位点核苷酸序列(如，基因组)的编辑器的量。在一些实施方案中，本文提供的引导编辑器(PE)的有效量，例如包含切口酶Cas9结构域和逆转录酶的融合蛋白的有效量，可指足以引起融合蛋白特异性结合和编辑靶位点的编辑的融合蛋白的量。本领域技术人员将理解，有效量的试剂，例如融合蛋白、核酸酶、杂合蛋白、蛋白二聚体、蛋白(或蛋白二聚体)和多核苷酸的复合物，或多核苷酸可能因各种因素而不同，例如期望的生物反应，例如特定的等位基因、基因组或待编辑的靶位点，被靶向的细胞或组织，以及所使用的试剂。

[0264] 易错逆转录酶

[0265] 如本文所用，术语“易错”逆转录酶(或更广泛地，任何聚合酶)是指天然存在或衍生自另一种逆转录酶(如，野生型M-MLV逆转录酶)的逆转录酶(或更广泛地，任何聚合酶)，其错误率小于野生型M-MLV逆转录酶的错误率。据报道，野生型M-MLV逆转录酶的错误率在15,000中的一个错误(较高)至27,000中的一个错误(较低)的范围内。15,000分之一的错误率对应于$6.7x10^{-5}$的错误率。27,000分之一的错误率对应于$3.7x10^{-5}$的错误率。参见Boutabout et al.(2001)“DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposonTy1,”NucleicAcidsRes 29(11):2217-2222，其通过引用并入本文。因此，出于本申请的目的，术语“易错”是指具有错误率大于15,000个核碱基掺入中一个错误($6.7x10^{-5}$或更高)的那些RT，例如，14,000个核碱基中有1个错误($7.14x10^{-5}$或更高)，13,000个或更少的核碱基中有1个错误($7.7x10^{-5}$或更高)，12,000个或更少的核碱基中有1个错误($7.7x10^{-5}$或更高)，11,000个或更少的核碱基中有1个错误($9.1x10^{-5}$或更高)，10,000个或更少的核碱基中有1个错误($1x10^{-4}$或0.0001或更高)，9,000个或更少的核碱基中有1个错误(0.00011或更高)，8,000个或更少的碱基中有1个错误(0.00013或更高)，7,000个或更少的核碱基中有1个错误(0.00014或更高)，6,000个或更少的核碱基中有1个错误(0.00016或更高)，5,000个或更少的核碱基中有1个错误(0.0002或更高)，4,000个或更少的核碱基中有1个错误(0.00025或更高)、3,000个或的更少核碱基中有1个错误(0.00033或更高)、2,000个或更少的核碱基中有1个错误(0.00050或更高)、或1,000个或更少的核碱基中有1个错误(0.001或更高)，或500个或更少的核碱基中有1个错误(0.002或更高)，或250个或更少的核碱基中有1个错误(0.004或更高)。

[0266] 外显肽

[0267] 如本文所用，术语“外显肽”是指侧翼是内含肽并在蛋白剪接形成成熟的剪接蛋白的过程中与另一外显肽连接的多肽序列。通常，内含肽的两侧是两个外显肽序列，当内含肽催化其自身切除时连接在一起。因此，外显肽是mRNA中存在的外显子的蛋白类似物。例如，包含内含肽的多肽可具有外显肽(N)-内含肽-外显肽(C)的结构。切除内含肽并剪接两个外显肽后，得到的结构是外显肽(N)-外显肽(C)和游离内含肽。在不同配置中，外显肽可以是单独的蛋白(如，Cas9或PE融合蛋白的半部)，各自融合到断裂内含肽，其中断裂内含肽的切除导致外显肽序列剪接在一起。

[0268] 延伸臂

[0269] 术语“延伸臂”是指PEgRNA的核苷酸序列组件，其提供多种功能，包括引物结合位点和逆转录酶的编辑模板。在一些实施方案中，例如图3D，延伸臂位于向导RNA的3’端。在其他实施方案中，例如图3E，延伸臂位于向导RNA的5’端。在一些实施方案中，延伸臂还包括同源臂。在不同实施方案中，延伸臂以5’至3’方向包含以下组件：同源臂、编辑模板和引物结合位点。由于逆转录酶的聚合活性为5’至3’方向，同源臂、编辑模板和引物结合位点的优选排列为5’至3’方向，是的逆转录酶一旦被退火引物序列引发，就使用编辑模板作为互补模板链聚合DNA单链。在本文别处描述更多细节，例如延伸臂的长度。

[0270] 延伸臂也可描述为通常包括两个区域：引物结合位点(PBS)和DNA合成模板，例如，如图3G(上图)所示。引物结合位点结合引物序列，所述引物序列是从靶位点的内源性DNA链当它被引导编辑器复合物产生切口，从而暴露内源性切口链上的3’端时形成的。如本文解释，引物序列与PEgRNA延伸臂上的引物结合位点的结合产生具有暴露的3’端(即，引物序列的3’)的双链体区域，该区域然后为聚合酶提供底物以从暴露的3’端沿着DNA合成模板的长度开始聚合DNA单链。单链DNA产物的序列是DNA合成模板的互补序列。继续向着DNA合成模板(或延伸臂)的5’方向聚合，直到聚合终止。因此，DNA合成模板代表延伸臂的一部分，该部分被引导编辑器复合物的聚合酶编码成单链DNA产物(即，含有期望的遗传编辑信息的3’单链DNA瓣)，并最终置换位于紧邻PE诱导的切口位点下游的靶位点的相应内源性DNA链。不受理论的束缚，DNA合成模板继续向着延伸臂的5’端聚合，直到终止事件。聚合可能以多种方式终止，包括但不限于(a)到达PEgRNA的5’端(如，在5’延伸臂的情况下，其中DNA聚合酶简单地耗尽模板)，(b)到达不可逾越的RNA二级结构(如，发夹或茎/环)，或(c)到达复制终止信号，例如阻断或抑制聚合酶的特定核苷酸序列，或核酸拓扑信号，如超螺旋DNA或RNA。

[0271] 瓣核酸内切酶(如，FEN1)

[0272] 如本文所用，术语“瓣核酸内切酶”是指催化去除5′单链DNA瓣的酶。这些是天然存在的酶，其处理细胞过程，包括DNA复制期间形成的5′瓣的除去。本文所述引导编辑方法可利用内源提供的瓣核酸内切酶或以反式提供的那些来去除引导编辑过程中的靶位点处形成的内源性DNA的5’瓣。瓣核酸内切酶是本领域已知的，并且描述于Patel et al.,“Flap endonucleases pass5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends,”Nucleic Acids Research,2012,40(10):4507-4519,Tsutakawa et al.,“Human flap endonuclease structures,DNA double-base flipping,and a unified understanding of the FEN1 superfamily,”Cell,2011,145(2):198-211,以及Balakrishnan et al.,“Flap Endonuclease 1,”Annu RevBiochem,2013,Vol 82:119-138(各自通过引用并入本文)。示例性瓣核酸内切酶是FEN1，其可由以下氨基酸序列表示：

[0273]

描述	序列	SEQ ID NO:
FEN1 野生型	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVDL CILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPN KYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPNE EELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 7

[0274] 功能等同物

[0275] 术语“功能等同物”是指与第一生物分子在功能上等同但在结构上不一定等同的第二生物分子。例如,“Cas9等同物”是指具有与Cas9相同或基本相同功能但不一定具有相同氨基酸序列的蛋白。在本公开的背景中,本说明书自始至终都提到“蛋白X或其功能等同物”。在这种情况下,蛋白X的“功能等同物”包括具有等同功能的蛋白X的任何同系物、旁系同源物、片段、天然存在、工程化、突变或合成版本。

[0276] 融合蛋白

[0277] 如本文所用,术语“融合蛋白”是指包含来自至少两种不同蛋白的蛋白结构域的杂合多肽。一种蛋白可位于融合蛋白的氨基端(N端)部分或羧基端(C端)蛋白,从而分别形成“氨基端融合蛋白”或“羧基端融合蛋白”。蛋白可包含不同的结构域,例如,核酸结合结构域(如,引导蛋白与靶位点结合的Cas9的gRNA结合结构域)和核酸编辑蛋白的核酸切割结构域或催化结构域。另一个示例包括Cas9或其与逆转录酶的等同物。本文提供的任何蛋白可通过本领域已知的任何方法产生。例如,本文提供的蛋白可通过重组蛋白表达和纯化产生,这尤其适用于包含肽接头的融合蛋白。用于重组蛋白表达和纯化的方法是众所周知的,包括Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))中描述的那些,其全部内容通过引用并入本文。

[0278] 感兴趣基因(GOI)

[0279] 术语“感兴趣基因”或“GOI”是指编码感兴趣的生物分子(如,蛋白或RNA分子)的基因。感兴趣的蛋白可包括任何细胞内蛋白、膜蛋白或细胞外蛋白,例如核蛋白、转录因子、核膜转运蛋白、细胞内细胞器相关蛋白、膜受体、催化蛋白和酶、治疗蛋白、膜蛋白、膜转运蛋白、信号转导蛋白或免疫蛋白(如,IgG或其他抗体蛋白)等。感兴趣基因还可编码RNA分子,包括但不限于信使RNA(mRNA)、转运RNA(tRNA)、核糖体RNA(rRNA)、小核RNA(snRNA)、反义RNA、向导RNA、微小RNA(miRNA)、小干扰RNA(siRNA)和无细胞(cell-free)RNA(cfRNA)。

[0280] 向导RNA(“gRNA”)

[0281] 如本文所用,术语“向导RNA”是特定类型的向导核酸,其通常与CRISPR-Cas9的Cas蛋白相关并且与Cas9缔合,将Cas9蛋白引导至DNA分子中包括与向导RNA的前间隔序列的互补性的特定序列。然而,该术语还包括等同的向导核酸分子,其与Cas9等同物、同源物、直系同源物或旁系同源物缔合,无论是天然存在的还是非天然存在的(如,工程化或重组的),并且在其它方面编程Cas9等同物以定位到特定的靶核苷酸序列。Cas9等同物可包括来自任何

类型的CRISPR系统(如,II、V、VI型)的其他napDNAbp,包括Cpf1(V型CRISPR-Cas系统)、C2c1(V型CRISPR-Cas系统)、C2c2(VI型CRISPR-Cas系统)和C2c3(V型CRISPR-Cas系统)。其他Cas等同物描述于Makarova et al.,"C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector,"Science 2016;353(6299)中,其内容通过引用并入本文。本文提供了向导RNA的示例性序列和结构。此外,本文提供了用于设计合适的向导RNA序列的方法。如本文所用,"向导RNA"又被称为"传统向导RNA",以将其与被称为"引导编辑向导RNA"(或"PEgRNA")的修饰形式的向导RNA形成对比,所述修饰形式的向导RNA已经被发明用于本文公开的引导编辑方法和组合物。

[0282] 向导RNA或PEgRNA可包含各种结构元件,包括但不限于:

[0283] 间隔区序列-向导RNA或PEgRNA中的序列(具有约20nt的长度),其与靶DNA中的前间隔序列结合。

[0284] gRNA核心(或gRNA支架或骨架序列)-指gRNA内负责Cas9结合的序列,其不包括用于引导Cas9至靶DNA的20bp间隔区/靶向序列。

[0285] 延伸臂-在PEgRNA的3′端或5′端处的单链延伸,其包含引物结合位点和DNA合成模板序列,所述DNA合成模板序列通过聚合酶(如,逆转录酶)编码含有感兴趣的基因变化的单链DNA瓣,然后通过置换相应内源性链整合至内源性DNA中,从而安装期望的基因变化。

[0286] 转录终止子-向导RNA或PEgRNA可在分子的3′端包含转录终止序列。

[0287] 同源臂

[0288] 术语"同源臂"是指编码所得到的逆转录酶编码的单链DNA瓣的部分的延伸臂的一部分,该单链DNA瓣将通过置换内源性链整合到靶DNA位置。由同源臂编码的单链DNA瓣的部分与靶DNA序列的非编辑链互补,这有利于内源性链的置换和单链DNA瓣就地退火,从而安装编辑。该组件在别处进一步定义。同源臂是DNA合成模板的一部分,因为根据定义,它由本文所述的引导编辑器的聚合酶编码。

[0289] 宿主细胞

[0290] 如本文所用,术语"宿主细胞"是指可以含有、复制和表达本文所述载体的细胞,例如包含编码融合蛋白的核酸分子的载体,所述融合蛋白包含Cas9或Cas9等同物和逆转录酶。

[0291] 内含肽

[0292] 如本文所用,术语"内含肽"是指在来自所有生命领域的生物体中发现的自动加工多肽结构域。内含肽(中间蛋白)执行被称为蛋白剪接的独特自动加工事件,其中它通过两个肽键的裂解从较大的前体多肽中切除自身,并在此过程中通过形成新的肽键连接侧翼外显肽(外部蛋白)序列。这种重排发生在翻译后(或可能是共同翻译),因为发现内含肽基因以符合读码框的方式嵌入其他蛋白编码基因内。此外,内含肽介导的蛋白剪接是自发的;它不需要外部因素或能量源,只需要折叠内含肽结构域。这个过程也被称为顺式蛋白剪接,与具有"断裂内含肽"的反式蛋白剪接的自然过程相反。内含肽是自剪接RNA内含子的蛋白等同物(参见,Perleret al.,NucleicAcids Res.22:1125-1127(1994)),其催化它们自己从前体蛋白中切除,同时伴随被称为外显肽的侧翼蛋白序列的融合(综述参见Perler et al.,Curr.Opin.Chem.Biol.1:292-299(1997);Perler,F.B.Cell 92(1):1-4(1998);Xu et al.,EMBO J.15(19):5146-5153(1996))。

[0293] 如本文所用,术语“蛋白剪接”是指将其中前体蛋白(内含肽)的内部区域切除并且将蛋白的侧翼区域(外显肽)连接以形成成熟蛋白的过程。这种自然过程已在来自原核生物和真核生物的许多蛋白中观测到(Perler,F.B.,Xu,M.Q.,Paulus,H.Current Opinionin Chemical Biology 1997,1,292-299;Perler,F.B.NucleicAcids Research 1999,27,346-347)。内含肽单元包含催化蛋白剪接所需的必要成分,并且通常包含参与内含肽迁移的核酸内切酶结构域(Perler,F.B.,Davis,E.O.,Dean,G.E.,Gimble,F.S.,Jack,W.E.,Neff,N.,Noren,C.J.,Thomer,J.,Belfort,M.NucleicAcids Research 1994,22,1127-1127)。然而,产生的蛋白是连接的,不作为分开的蛋白表达。蛋白剪接也可以反式进行,在分开的多肽上表达的断裂内含肽自发组合形成单个内含肽,其然后经历蛋白剪接过程以连接至分开的蛋白。

[0294] 蛋白剪接机制的阐明导致了许多基于内含肽的应用((Comb,et al.,美国专利号5,496,714;Comb,et al.,美国专利号5,834,247;Camarero and Muir,J.Amer.Chem.Soc.,121:5597-5598(1999);Chong,et al.,Gene,192:271-281(1997),Chong,et al.,NucleicAcids Res.,26:5109-5115(1998);Chong,et al.,J.Biol.Chem.,273:10567-10577(1998);Cotton,et al.J.Am.Chem.Soc.,121:1100-1101(1999);Evans,et al.,J.Biol.Chem.,274:18359-18363(1999);Evans,et al.,J.Biol.Chem.,274:3923-3926(1999);Evans,et al.,Protein Sci.,7:2256-2264(1998);Evans,et al.,J.Biol.Chem.,275:9091-9094(2000);Iwai and Pluckthun,FEBS Lett.459:166-172(1999);Mathys,et al.,Gene,231:1-13(1999);Mills,et al.,Proc.Natl.Acad.Sci.USA 95:3543-3548(1998);Muir,et al.,Proc.Natl.Acad.Sci.USA 95:6705-6710(1998);Otomo,et al.,Biochemistry 38:16040-16044(1999);Otomo,et al.,J.Biolmol.NMR14:105-114(1999);Scott,et al.,Proc.Natl.Acad.Sci.USA 96:13638-13643(1999);Severinov and Muir,J.Biol.Chem.,273:16205-16209(1998);Shingledecker,et al.,Gene,207:187-195(1998);Southworth,et al.,EMBO J.17:918-926(1998);Southworth,et al.,Biotechniques,27:110-120(1999);Wood,et al.,Nat.Biotechnol.,17:889-892(1999);Wu,et al.,Proc.Natl.Acad.Sci.USA 95:9226-9231(1998a);Wu,et al.,Biochim Biophys Acta 1387:422-432(1998b);Xu,et al.,Proc.Natl.Acad.Sci.USA 96:388-393(1999);Yamazaki,et al.,J.Am.Chem.Soc.,120:5591-5592(1998))。各文献通过引用并入本文。

[0295] 配体依赖性内含肽

[0296] 如本文所用,术语“配体依赖性内含肽”是指包含配体结合结构域的内含肽。通常,将配体结合结构域插入内含肽的氨基酸序列中,形成结构:

[0297] 内含肽(N)-配体结合结构域-内含肽(C)。通常,配体依赖性内含肽在不存在合适配体时未显示蛋白剪接活性或仅显示最少的蛋白剪接活性,而在配体存在下蛋白剪接活性显著增加。在一些实施方案中,配体依赖性内含肽在配体不存在时未显示可观测的剪接活性,但在配体存在下确实显示剪接活性。在一些实施方案中,配体依赖性内含肽在配体不存在时显示可观测的蛋白剪接活性,并且在合适配体存在下的蛋白剪接活性比不存在配体时观测到的活性大至少5倍、至少10倍、至少50倍、至少100倍、至少150倍、至少200倍、至少250倍、至少500倍、至少1000倍、至少1500倍、至少2000倍、至少2500倍、至少5000倍、至少10000

倍、至少20000倍、至少25000倍、至少50000倍、至少100000倍、至少500000倍、或至少1000000倍。在一些实施方案中，活性的增加在至少1个数量级、至少2个数量级、至少3个数量级、至少4个数量级、或至少5个数量级上是剂量依赖性的，允许通过调节配体浓度来微调内含肽活性。合适的配体依赖性内含肽是本领域已知的，并且包括以下提供的那些和在公开的美国专利申请U.S.2014/0065711A1中描述的那些；Mootz et al.，“Protein splicing triggeredby a small molecule.”J.Am.Chem.Soc.2002;124,9044-9045;Mootz et al.，“Conditional protein splicing:anewtool to controlprotein structure and function in vitro and in vivo.”J.Am.Chem.Soc.2003;125,10561-10569;Buskirk et al.,Proc.Natl.Acad.Sci.USA.2004;101,10505-10510);Skretas&Wood,“Regulation of protein activitywith small-molecule-controlled inteins.”Protein Sci.2005;14,523-532;Schwartz,et al.,“Post-translational enzyme activation in an animal via optimized conditional protein splicing.”Nat.Chem.Biol.2007;3,50-54;Peck etal.,Chem.Biol.2011;18(5),619-630；各自的全部内容在此通过引用并入。示例性序列如下：

[0298]

名称	配体依赖性内含肽的序列
2-4 内含肽:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 8)

[0299]

3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDR VAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASM MGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEIL MIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS SRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRAL DKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL YSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDK FLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHN C (SEQ ID NO: 9)
30R3-1 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 10)
30R3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 11)
30R3-3 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 12)
37R3-1 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC ((SEQ ID NO: 13)
37R3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD

[0300]

	RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 14)
37R3-3 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 15)

[0301] 接头

[0302] 如本文所用，术语“接头”是指连接两个其他分子或部分的分子。在连接两个融合蛋白的接头的情况下，接头可以是氨基酸序列。例如，Cas9可通过氨基酸接头序列与逆转录酶融合。在将两个核苷酸序列连接在一起的情况下，接头也可以是核苷酸序列。例如，在当前情况下，传统的向导RNA通过间隔区或接头核苷酸序列连接至可包含RT模板序列和RT引物结合位点的引导编辑向导RNA的RNA延伸。在其他实施方案中，接头是有机分子、基团、聚合物或化学部分。在一些实施方案中，接头的长度为5至100个氨基酸，例如，长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。还考虑更长或更短的接头。

[0303] 分离的

[0304] “分离的”是指从自然状态改变或取出。例如，天然存在于活体动物中的核酸或肽不是“分离的”，而与其天然状态的共存材料部分或完全分离的同一核酸或肽是“分离的”。分离的核酸或蛋白可以以基本上纯化的形式存在，或可存在于非天然环境中如例如宿主细胞中。

[0305] 在一些实施方案中，感兴趣基因由分离的核酸编码。如本文所用，术语“分离的”是指如本文提供的材料从其原始或天然环境(如，如果它是天然存在的，则为天然环境)取出的特性。因此，存在于活体动物中的天然存在的多核苷酸或蛋白或多肽不是分离的，而通过人工干预与天然系统中的一些或全部共存材料分开的同一多核苷酸或多肽是分离的。因此，人工或工程化材料，例如非天然存在的核酸构建体，如本文所述表达构建体和载体，又被称为分离的。材料不必为了分离而纯化。因此，材料可以是载体的一部分和/或组合物的一部分，并且仍然是分离的，因为这样的载体或组合物不是在自然界中找到材料的环境的一部分。

[0306] MS2标签化技术

[0307] 在不同实施方案中(例如，如图72-73的实施方案和实施例19中所描绘)，术语“MS2标签化技术”是指“RNA-蛋白相互作用结构域”(又名“RNA-蛋白募集结构域或蛋白”)与特异

性识别并结合RNA-蛋白相互作用结构域(如,特定发夹结构)的RNA结合蛋白配对的组合。可利用这些类型的系统将各种功能性募集到结合至靶位点的引导编辑器复合物。MS2标签化技术基于MS2噬菌体外壳蛋白("MCP"或"MS2cp")与噬菌体基因组中存在的茎环或发夹结构(即"MS2发夹")的自然相互作用。在引导编辑的情况下,MS2标签化技术包括将MS2发夹引入参与引导编辑的期望的RNA分子(如,PEgRNA或tPERT),然后构成识别并结合至该结构的RNA结合蛋白的特异性可相互作用结合靶标。在MS2发夹的情况下,其被MS2噬菌体外壳蛋白(MCP)识别和结合。并且,若MCP与另一个蛋白(如,逆转录酶或其他DNA聚合酶)融合,则可将MS2发夹用于反式"募集"其他蛋白至由引导编辑复合物占据的靶位点。

[0308] 一方面,本文所述的引导编辑器可并入任何已知的RNA-蛋白相互作用结构域以募集或"共定位"感兴趣的特定功能性至引导编辑器复合物。本领域中已描述其他模块化RNA-蛋白相互作用结构域的综述,例如,Johansson et al.,"RNArecognitionby the MS2 phage coatprotein,"Sem Virol.,1997,Vol.8(3):176-185;Delebecque et al.,"Organization of intracellular reactions with rationally designed RNA assemblies,"Science,2011,Vol.333:470-474;Mali et al.,"Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering,"Nat.Biotechnol.,2013,Vol.31:833-838;以及Zalatan et al.,"Engineering complex synthetic transcriptional programs with CRISPR RNAscaffolds,"Cell,2015,Vol.160:339-350,各自通过引用整体并入本文。其他系统包括PP7发夹(其专门募集PCP蛋白)和"com"发夹(其专门募集Com蛋白)。参见Zalatan et al.。

[0309] MS2发夹(或等同地被称为"MS2适体")的核苷酸序列为:

[0310] GCCAACATGAGGATCACCCATGTCTGCAGGGCC(SEQ ID NO:763)。

[0311] MCP或MS2cp的氨基酸序列为:

[0312] GSASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGEELPVAGWRSYLNMELTIPFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY(SEQ ID NO:764)。

[0313] MS2发夹(或"MS2适体")也可被称为一类"RNA效应物募集结构域"(或等同地被称为"RNA结合蛋白募集结构域"或简称为"募集结构域"),因为它是安装到PEgRNA或tPERT中的物理结构(如,发夹),其有效地将其他效应物功能(如,具有不同功能的RNA结合蛋白,例如DNA聚合酶或其他DNA修饰酶)募集到如此修饰的PEgRNA或rPERT,从而将效应物功能以反式共定位到引导编辑机制。本申请无意以任何方式限制于任何特定的RNA效应物募集结构域,并且可包括任何可用的此类结构域,包括MS2发夹。实施例19和图72(b)描绘了使用连接到DNA合成结构域(即,tPERT分子)的MS2适体和引导编辑器,所述引导编辑器包含融合至PE2的MS2cp蛋白以引起与靶DNA位置结合的引导编辑器复合物(MS2cp-PE2:sgRNA复合物)和tPERT分子的DNA合成结构域共定位以完成。

[0314] napDNAbp

[0315] 如本文所用,术语"核酸可编程DNA结合蛋白"或"napDNAbp"(其中Cas9是示例)是指使用RNA:DNA杂交来靶向并结合DNA分子中的特定序列的蛋白。每个napDNAbp与至少一个向导核酸(如,向导RNA)缔合,这将napDNAbp定位于包含与向导核酸或其部分(如,向导RNA

的前间隔序列)互补的DNA链(即,靶链)的DNA序列。换言之,向导核酸“编程”napDNAbp(如,Cas9或等同物)以定位并结合互补序列。

[0316] 不受理论的束缚,napDNAbp-向导RNA复合物的结合机制通常包括形成R环的步骤,由此napDNAbp诱导双链DNA靶标的解旋,从而分离由napDNAbp结合的区域中的链。然后向导RNA前间隔序列与“靶链”杂交。这置换了与靶链互补的“非靶链”,形成了R环的单链区域。在一些实施方案中,napDNAbp包括一种或多种核酸酶活性,其然后切割DNA,留下各种类型的损伤。例如,napDNAbp可包含在第一位置处切割非靶链和/或在第二位置处切割靶链的核酸酶活性。根据核酸酶活性,可切割靶DNA形成“双链断裂”,从而切割两条链。在其他实施方案中,靶DNA可仅在单个位置被切割,即DNA在一条链上有“切口”。具有不同核酸酶活性的示例性napDNAbp包括“Cas9切口酶”(“nCas9”)和没有核酸酶活性的失活Cas9(“死亡Cas9”或“dCas9”)。本文提供了这些和其他napDNAbp的示例性序列。

[0317] 切口酶

[0318] 术语“切口酶”是指两个核酸酶结构域之一失活的Cas9。这种酶只能切割靶DNA的一条链。

[0319] 核定位序列(NLS)

[0320] 术语“核定位序列”或“NLS”是指如通过核转运促进蛋白输入细胞核的氨基酸序列。核定位序列是本领域已知的并且对技术人员而言是显而易见的。例如,lank et al.于2000年11月23日提交的国际PCT申请号PCT/EP2000/011690,于2001年5月31日以WO/2001/038547公开,其公开的示例性核定位序列的内容通过引用并入本文。在一些实施方案中,NLS包含氨基酸序列PKKKRKV(SEQ ID NO:16)或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:17)。

[0321] 核酸分子

[0322] 如本文所用,术语“核酸”是指核苷酸的聚合物。聚合物可包括天然核苷(即,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷)、核苷类似物(如,2-氨基腺苷、2-硫胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷,5-甲基胞苷,C5溴尿苷,C5氟尿苷,C5碘尿苷,C5丙炔基尿苷,C5丙炔基胞苷,C5甲基胞苷,7脱氮腺苷,7脱氮鸟苷,8氧腺苷,8氧鸟苷、O(6)甲基鸟嘌呤、4-乙酰基胞苷、5-(羧基羟甲基)尿苷、二氢尿苷、甲基假尿苷、1-甲基腺苷、1-甲基鸟苷、N6-甲基腺苷和2-硫代胞苷)、化学修饰的碱基、生物修饰的碱基(如,甲基化碱基)、插入的碱基、修饰的糖(如,2'-氟核糖、核糖、2'-脱氧核糖、2'-O-甲基胞苷、阿拉伯糖和己糖),或修饰的磷酸基(如,硫代磷酸酯和5′N亚磷酰胺连接)。

[0323] PEgRNA

[0324] 如本文所用,术语“引导编辑向导RNA”或“PEgRNA”或“延伸的向导RNA”是指已修饰以包括用于实施本文描述的引导编辑方法和组合物的一个或多个额外序列的特定形式的向导RNA。如本文所述,引导编辑向导RNA包含核酸序列的一个或多个“延伸区”。延伸区可包括但不限于单链RNA或DNA。此外,延伸区可能存在于传统向导RNA的3'端。在其他配置中,延伸区可以存在于传统向导RNA的5'端。在其他配置中,延伸区域可能存在于传统向导RNA的分子内区域,例如,在与napDNAbp缔合和/或结合的gRNA核心区域中。延伸区包含“DNA合成模板”,其编码(通过引导编辑器的聚合酶)单链DNA,所述单链DNA继而被设计为(a)与待编辑的内源性靶DNA同源,并且(b)其包含至少一个待引入或整合到内源性靶DNA中的期望的

核苷酸变化(如,转变、颠换、缺失、或插入)。延伸区还可包含其他功能序列元件,例如但不限于“引物结合位点”和“间隔区或接头”序列,或附加结构元件,例如但不限于适体、茎环、发夹、趾环(如,3’趾环)或RNA-蛋白募集结构域(如,MS2发夹)。如本文所用,“引物结合位点”包含与从R环的有切口的DNA产生的具有3’端的单链DNA序列杂交的序列。

[0325] 在某些实施方案中,PEgRNA由图3A所示,显示具有5′延伸臂、间隔区和gRNA核心的PEgRNA。5’延伸在5’至3’方向还包括逆转录酶模板、引物结合位点和接头。如所示,逆转录酶模板也可更广泛地被称为“DNA合成模板”,其中本文所述的引导编辑器的聚合酶不是RT,而是另一类型的聚合酶。

[0326] 在某些其他实施方案中,PEgRNA由图3B所示,显示具有5′延伸臂、间隔区和gRNA核心的PEgRNA。5’延伸在5’至3’方向还包括逆转录酶模板、引物结合位点和接头。如所示,逆转录酶模板也可更广泛地被称为“DNA合成模板”,其中本文所述的引导编辑器的聚合酶不是RT,而是另一类型的聚合酶。

[0327] 在其他实施方案中,PEgRNA由图3D所示,显示在5’至3’方向具有间隔区(1)、gRNA核心(2)和延伸臂(3)的PEgRNA。延伸臂(3)位于PEgRNA的3′端。延伸臂(3)在5’至3’方向还包括“引物结合位点”(A)、“编辑模板”(B)和“同源臂”(C)。延伸臂(3)在3’端和5’端还可包含任选的修饰区,可以是相同序列或不同序列。此外,PEgRNA的3’端可包含转录终止子序列。PEgRNA的这些序列元件在本文中进一步描述和定义。

[0328] 在其他实施方案中,PEgRNA由图3E所示,显示在5’至3’方向具有延伸臂(3)、间隔区(1)和gRNA核心(2)的PEgRNA。延伸臂(3)位于PEgRNA的5’端。延伸臂(3)在5’至3’方向还包括“引物结合位点”(A)、“编辑模板”(B)和“同源臂”(C)。延伸臂(3)在3’端和5’端还可包含任选的修饰区,其可以是相同序列或不同序列。PEgRNA在3’端还可包含转录终止子序列。PEgRNA的这些序列元件在本文中进一步描述和定义。

[0329] PE1

[0330] 如本文所用,“PE1”是指具有以下结构的包含融合蛋白的PE复合物,该融合蛋白包含Cas9(H840A)和野生型MMLV RT:[NLS]-[Cas9(H840A)]-[接头]-[MMLV_RT(wt)]+期望的PEgRNA,其中PE融合物具有SEQ ID NO:123的氨基酸序列,如下所示;

[0331] MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL

VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL
DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLHETSKEP
DVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP
WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA
FEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ
TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMA
APLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFG
PVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWA
KALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALF
LPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV(SEQ
ID NO:123)

[0332] 解释：

[0333] 核定位序列(NLS)开端：(SEQ ID NO:124)，末端：(SEQ ID NO:133)

[0334] CAS9(H840A)(SEQ ID NO:126)

[0335] 33-氨基酸接头(SEQ ID NO:127)

[0336] M-MLV逆转录酶(SEQ ID NO:128)。

[0337] PE2

[0338] 如本文使用，“PE2”是指具有以下结构的包含融合蛋白的PE复合物，该融合蛋白包含Cas9(H840A)和变体MMLV RT：(NLS)-[(Cas9(H840A)]-[接头]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)]+期望的PEgRNA，其中PE融合物具有SEQ ID NO:134的氨基酸序列，如下所示；

[0339] MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV
AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL

VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL
DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLHETSKEP
DVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP
WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA
FEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ
TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMA
APLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFG
PVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWA
KALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALF
LPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV(SEQ
ID NO:134)

[0340] 解释:

[0341] 核定位序列(NLS)开端:(SEQ ID NO:124),末端:(SEQ ID NO:133)

[0342] CAS9(H840A)(SEQ ID NO:137)

[0343] 33-氨基酸接头(SEQ ID NO:127)

[0344] M-MLV逆转录酶(SEQ ID NO:139)。

[0345] PE3

[0346] 如本文所用,“PE3”是指PE2加上第二链切口产生性向导RNA,其与PE2复合并在非编辑的DNA链中引入切口以引起编辑链的优先置换。

[0347] PE3b

[0348] 如本文所用,“PE3b”是指PE3但其中第二链切口产生性向导RNA设计用于时序控制,使得直到安装期望编辑之后才引入第二链切口。这通过设计具有间隔区序列的gRNA来实现,该序列仅与编辑链相匹配,而不与原始等位基因相匹配。使用此策略(以下被称为PE3b),前间隔序列和非编辑等位基因之间的错配应当不利于被sgRNA产生切口,直到PAM链上的编辑事件发生之后。

[0349] PE-短

[0350] 如本文所用,“PE-短”是指与C-端截短的逆转录酶融合的PE构建体,具有以下氨基酸序列:

[0351] MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV
AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY

PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL
VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL
DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLHETSKEP
DVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP
WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA
FEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ
TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMA
APLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFG
PVVALNPATLLPLPEEGLQHNCLDNSRLINSGGSKRTADGSEFEPKKKRKV(SEQ ID NO:765)解释：

[0352] 核定位序列(NLS)开端：(SEQ ID NO:124)，末端：(SEQ ID NO:133)

[0353] CAS9(H840A)(SEQ ID NO:157)

[0354] 33-氨基酸接头1(SEQ ID NO:127)

[0355] M-MLV截短的逆转录酶(SEQ ID NO:766)

[0356] 肽标签

[0357] 术语“肽标签”是指与蛋白序列基因融合以赋予蛋白以一种或多种功能的肽氨基酸序列，这些功能有助于出于不同目的对蛋白进行操作，如可视化、纯化、增溶和分离等。肽标签可包括按目的或功能分类的不同类型的标签，其可包括“亲和标签”(促进蛋白纯化)、“增溶标签”(协助蛋白正确折叠)、“色谱标签”(改变蛋白的色谱特性)、“表位标签”(与高亲和力抗体结合)、“荧光标签”(促进细胞中或体外的蛋白可视化)。

[0358] 聚合酶

[0359] 如本文所用，术语“聚合酶”是指合成核苷酸链且可与本文所述的引导编辑器系统结合使用的酶。聚合酶可以是“模板依赖性”聚合酶(即，基于模板链的核苷酸碱基顺序合成核苷酸链的聚合酶)。聚合酶还可以是“模板非依赖性”聚合酶(即，不需要模板链就合成核苷酸链的聚合酶)。聚合酶还可进一步分类为“DNA聚合酶”或“RNA聚合酶”。在不同实施方案中，引导编辑器系统包含DNA聚合酶。在不同实施方案中，DNA聚合酶可以是“DNA依赖性DNA聚合酶”(即，由此模板分子是DNA链)。在这种情况下，DNA模板分子可以是PEgRNA，其中延伸臂包含DNA链。在这种情况下，PEgRNA可被称为嵌合或杂合PEgRNA，其包含RNA部分(即，向导RNA组件，包括间隔区和gRNA核心)和DNA部分(即，延伸臂)。在不同的其他实施方案中，DNA

聚合酶可以是“RNA依赖性DNA聚合酶”(即,由此模板分子是RNA链)。在这种情况下,PEgRNA是RNA,即包括RNA延伸。术语“聚合酶”还可指催化核苷酸聚合(即,聚合酶活性)的酶。通常,酶将在与多核苷酸模板序列退火的引物(如,与PEgRNA的引物结合位点退火的引物序列)的3’-端开始合成,并将向着模板链的5’端前进。“DNA聚合酶”催化脱氧核苷酸的聚合。如本文关于DNA聚合酶所用,术语DNA聚合酶包括“其功能片段”。“其功能片段”是指野生型或突变型DNA聚合酶的任何部分,其包含少于聚合酶的完整氨基酸序列并在至少一组条件下保留催化多核苷酸聚合的能力。这样的功能片段可作为单独的实体存在,或它可以是较大多肽(如融合蛋白)的组分。

[0360] 引导编辑和多瓣引导编辑

[0361] 如本文所用,术语“引导编辑”或“经典引导编辑”是指使用napDNAbps、聚合酶(如,逆转录酶)和特化向导RNA进行基因编辑的新方法,所述特化向导RNA包括用于编码期望的新遗传信息(或删除遗传信息)的DNA合成模板,所述新遗传信息然后被掺入靶DNA序列中。图1A至1H和图72(a)至72(c)的实施方案中特别描述了引导编辑的某些实施方案。经典引导编辑在发明人的以下公开出版物中描述:Anzalone,A.V.et al.Search-and-replace genome editing without double-strand breaks or donor DNA.Nature 576,149-157(2019),该文献通过引用整体并入本文。

[0362] 引导编辑代表了通用且精确的基因组编辑方法的基因组编辑平台,其使用与聚合酶(即,以融合蛋白的形式或在其它情况下以与napDNAbp反式提供)一起运行的核酸可编程DNA结合蛋白(“napDNAbp”)将新遗传信息写入规定的DNA位点,其中引导编辑系统利用引导编辑(PE)向导RNA(“PEgRNA”)编程,该PEgRNA通过工程化至向导RNA(如,在向导RNA的5′或3′端或内部部分)的延伸(DNA或RNA)既指定靶位点,又为以置换DNA链的形式合成期望编辑提供模板。含有期望编辑(如,单核碱基取代)的置换链与待编辑的靶位点的内源性链(在切口位点的紧邻下游)共有相同(或同源)的序列(只是它包括期望编辑)。通过DNA修复和/或复制机制,切口位点下游的内源性链被新合成的包含期望编辑的置换链替换。在某些情况下,引导编辑可以认为是“搜索和置换(search-and-replace)”基因组编辑技术,因为本文所述的引导编辑器不仅搜索和定位待编辑的期望的靶位点,而且同时编码被安装而替换相应靶位点内源性DNA链的含有期望编辑的置换链。本公开的引导编辑器部分地涉及以下发现:可利用或调整靶标引发的逆转录(TPRT)或“引导编辑”的机制进行基于CRISPR/Cas的精确基因组编辑,具有高效率和遗传可塑性(如,如图1A至1F的不同实施方案所描绘)。TPRT自然被移动DNA元件所使用,如哺乳动物非LTR逆转录转座子和细菌II组内含子[28,29]。本发明人在本文中使用Cas蛋白-逆转录酶融合物或相关系统以利用向导RNA来靶向特定DNA序列,在中靶位点处产生单链切口,并使用有切口的DNA作为引物以进行与向导RNA一起整合的工程化逆转录酶模板的逆转录。然而,虽然该概念始于使用逆转录酶作为DNA聚合酶组件的引导编辑器,但本文所述的引导编辑器不限于逆转录酶,而是可包括使用几乎任何DNA聚合酶。事实上,虽然本申请自始至终都可以涉及具有“逆转录酶”的引导编辑器,但这里提出逆转录酶只是可与引导编辑一起起作用的一种类型DNA聚合酶。因此,无论说明书何处提及“逆转录酶”,本领域普通技术人员都应理解,可使用任何合适的DNA聚合酶代替逆转录酶。因此,一方面,引导编辑器可包含Cas9(或等同的napDNAbp),其被编程为通过将它与特化向导RNA(即,PEgRNA)缔合而靶向DNA序列,所述特化向导RNA包含与靶DNA中的前间隔序列的互

补物退火的间隔区序列。特化向导RNA还包含延伸形式的新遗传信息，该信息编码包含期望的基因变化的DNA置换链，其用于替换靶位点处的相应内源性DNA链。为了将信息从PEgRNA转移到靶DNA，引导编辑机制涉及在DNA的一条链上对靶位点产生切口以暴露3'-羟基。然后，可以使用暴露的3'-羟基对中靶位点中直接引发PEgRNA上编码编辑的延伸的DNA聚合。在不同实施方案中，延伸(其为含有编辑的置换链的聚合提供模板)可由RNA或DNA形成。在RNA延伸的情况下，引导编辑器的聚合酶可以是RNA依赖性DNA聚合酶(如，逆转录酶)。在DNA延伸的情况下，引导编辑器的聚合酶可以是DNA依赖性DNA聚合酶。由本文公开的引导编辑器形成的新合成链(即，含有期望编辑的置换DNA链)除了包含期望的核苷酸变化(如，单核苷酸变化、缺失或插入，或其组合)以外，将与基因组靶序列同源(即，具有相同的序列)。新合成(或置换)的DNA链又被称为单链DNA瓣，它将竞争与互补的同源内源性DNA链的杂交，从而置换相应的内源性链。在某些实施方案中，该系统可与易错逆转录酶的使用组合(如，作为与Cas9结构域的融合蛋白提供，或与Cas9结构域以反式提供)。易错逆转录酶可在单链DNA瓣的合成过程中引入变化。因此，在某些实施方案中，可利用易错逆转录酶将核苷酸变化引入靶DNA。根据与系统一起使用的易错逆转录酶，变化可以是随机的或非随机的。杂交中间体(包括由与内源性DNA链杂交的逆转录酶合成的单链DNA瓣)的分解可包括去除所产生的内源性DNA的置换瓣(如，使用5'端DNA瓣核酸内切酶，FEN1)，连接合成的单链DNA瓣与靶DNA，以及由于细胞DNA修复和/或复制过程而导致的期望的核苷酸变化的同化。由于提供模板的DNA合成为任何核苷酸的修饰(包括插入和缺失)提供了单核苷酸精确度，因此这种方法的范围非常广泛，可预见地可以用于基础科学和治疗学中的无数应用。

[0363] 在不同实施方案中，引导编辑通过将靶DNA分子(期望对其引入核苷酸序列的变化)与同引导编辑向导RNA(PEgRNA)复合的核酸可编程DNA结合蛋白(napDNAbp)接触运行。参考图1G所示，引导编辑向导RNA(PEgRNA)包含在向导RNA的3'或5'端或向导RNA的分子内位置处的延伸，并编码期望的核苷酸变化(如，单核苷酸变化、插入、或缺失)。步骤(a)中，napDNAbp/延伸的gRNA复合物接触DNA分子，并且延伸的gRNA引导napDNAbp与靶因座结合。步骤(b)中，在靶基因座的DNA链之一中引入切口(如，通过核酸酶或化学试剂)，从而在靶基因座的链之一中产生可用的3'端。在某些实施方案中，在对应于R-环链的DNA链(即，未与向导RNA序列杂交的链，即“非靶链”)中产生切口。然而，可在任何一条链中引入切口。换言之，可将切口引入R环“靶链”(即，与延伸的gRNA的前间隔序列杂交的链)或“非靶链”(即，形成R环的单链部分的链，其与靶链互补)。步骤(c)中，DNA链的3'端(由切口形成)与向导RNA的延伸部分相互作用以引发逆转录(即“靶物引发的RT”)。在某些实施方案中，3'端DNA链与向导RNA的延伸部分上的特定RT引发序列，即，PEgRNA上的“逆转录酶引发序列”或“引物结合位点”杂交。步骤(d)中，引入逆转录酶(或其他合适的DNA聚合酶)，该酶从引发位点的3'端向着引导编辑向导RNA的5'端合成单链DNA。DNA聚合酶(如，逆转录酶)可与napDNAbp融合，或者备选地可以与napDNAbp以反式提供。这形成了单链DNA瓣，其包含期望的核苷酸变化(如，单碱基变化、插入或缺失或其组合)，并且在其它方面在切口位点处或切口位点附近与内源性DNA同源。步骤(e)中，将napDNAbp和向导RNA释放。步骤(f)和(g)涉及单链DNA瓣的分解，以便将期望的核苷酸变化掺入靶基因座。这可通过去除相应的5'内源性DNA瓣向期望的产物形成驱动该过程，所述5'内源性DNA瓣在3'单链DNA瓣侵入内源DNA序列并且与内源DNA序列杂交后形成。不受理论的束缚，细胞内源性DNA修复和复制过程分解错配的DNA以掺入核

苷酸变化来形成期望的改变产物。也可通过“第二链切口产生”向产物形成驱动该过程，如图1F所示。该过程可引入至少一个或多个以下基因变化：颠换、转换、缺失和插入。

[0364] 术语“引导编辑器(PE)系统”或“引导编辑器(PE)”或“PE系统”或“PE编辑系统”是指使用本文所述靶标引发的逆转录(TPRT)进行基因组编辑方法中涉及的组合物，包括但不限于napDNAbp、逆转录酶、融合蛋白(如，包含napDNAbp和逆转录酶)、引导编辑向导RNA，以及包含融合蛋白和引导编辑向导RNA的复合物，以及辅助元件，例如第二链切口产生性组件(如，第二链sgRNA)和5’内源性DNA瓣去除核酸内切酶(如，FEN1)，用于帮助向着编辑产物形成驱动引导编辑过程。

[0365] 尽管在迄今为止描述的实施方案中，PEgRNA构成包含向导RNA(其本身包含间隔区序列和gRNA核心或支架)和含有引物结合位点和DNA合成模板的5’或3’延伸臂的单分子(如，参见图3D，PEgRNA也可采用由向导RNA和反式引导编辑器RNA模板(tPERT)组成的两个单独分子的形式，所述反式引导编辑器RNA模板基本上容纳同一分子中的延伸臂(包括特别是引物结合位点和DNA合成结构域)和RNA-蛋白募集结构域(如，MS2适体或发夹)，其共定位或募集到包含tPERT募集蛋白(如，MS2cp蛋白，其与MS2适体结合)的经修饰的引导编辑器复合物。参见图3G和图3H，作为可与引导编辑一起使用的tPERT的示例。

[0366] 在“双瓣引导编辑系统”中，使用两个pegRNA靶向基因组位点的相对链并且指导两个含有编辑DNA序列的互补3’瓣的合成(图91)。与经典引导编辑不同，不需要成对的编辑DNA链(3’瓣)与内源性基因组DNA的5’瓣直接竞争，因为互补编辑链也可用于杂交。由于双链体的两条链作为编辑的DNA合成，双瓣引导编辑系统省略了经典引导编辑所需的替换未编辑的互补DNA链的需求。相反，细胞DNA修复机制仅需要切除成对的5’瓣(原始基因组DNA)并将成对的3’瓣(编辑的DNA)连接到基因座中。因此，在新合成的DNA链中不需要包含与基因组DNA同源的序列，允许新链的选择性杂交并且有利于含有最小基因组同源性的编辑。也可以使用切割DNA的两条链的核酸酶活性形式的引导编辑器以加快原始DNA序列的去除。

[0367] 与经典引导编辑一样，多瓣引导编辑(包括双瓣和四瓣引导编辑)是多用途的、精确的基因组编辑方法，其使用与聚合酶协同工作的核酸可编程DNA结合蛋白(“napDNAbp”)将新的遗传信息直接写入指定的DNA位点(即，以融合蛋白的形式或以反式形式与napDNAbp一起提供)，其中引导编辑系统使用引导编辑(PE)向导RNA(“PEgRNA”)进行编程，该向导RNA既指定靶位点，又通过工程化到向导RNA上(例如，在向导RNA的5’或3’端，或者在其内部)的延伸(DNA或RNA)的方式以替换DNA链的形式为期望的编辑的合成提供模板。含有期望的编辑的替换链(例如，单一核酸酶取代)与待编辑的靶位点的内源性链共享相同序列(不同之处在于其包含期望的编辑)。通过DNA修复和/或复制机制，靶位点的内源性链被替换为新合成的含有期望的编辑的替换链。在一些情况下，引导编辑可以被认为是“检索和替换”基因组编辑技术，因为本文所述的引导编辑器不仅检索和定位待编辑的位点，而且同时还编码含有期望的编辑的替换链，该期望的编辑安装在相应的靶位点内源性DNA链的位置。

[0368] 引导编辑器

[0369] 本文所述的双引导编辑系统包含一对引导编辑器。术语“引导编辑器”是指本文所述融合构建体，其包含napDNAbp(如，Cas9切口酶)和逆转录酶，并且能够在PEgRNA(或“延伸的向导RNA”)存在下对靶核苷酸序列进行引导编辑。术语“引导编辑器”可指融合蛋白或与PEgRNA复合和/或进一步与第二链切口产生性sgRNA复合的融合蛋白。在一些实施方案中，

引导编辑器还可指包含融合蛋白(与napDNAbp融合的逆转录酶)、PEgRNA和能够指导非编辑链的第二位点切口产生步骤的常规向导RNA的复合物,如本文所述。在其他实施方案中,“引导编辑器”的逆转录酶组件可以反式提供。

[0370] 本文所述的双瓣引导编辑系统包含一对引导编辑器。本文所述的四瓣引导编辑系统包含四个引导编辑器。

[0371] 引物结合位点

[0372] 术语“引物结合位点”或“PBS”是指位于PEgRNA上作为延伸臂组件(通常在延伸臂的3’端)的核苷酸序列,并用于结合由引导编辑器对靶序列的Cas9切口产生后形成的引物序列。如别处详述,当引导编辑器的Cas9切口酶组件对靶DNA序列的一条链产生切口时,形成3′端的ssDNA瓣,其充当与PEgRNA上的引物结合位点退火以引发逆转录的引物序列。

[0373] 启动子

[0374] 术语“启动子”是本领域公认的,是指具有被细胞转录机制识别并能够启动下游基因转录的序列的核酸分子。启动子可以是组成型活性,意味着该启动子在给定的细胞环境中总是有活性,或者是条件型活性,意味着该启动子仅在存在特定条件时才具有活性。例如,条件型启动子可以仅在将与启动子中的调控元件相关的蛋白与基本转录机制连接起来的特定蛋白存在时才具有活性,或者仅在不存在抑制性分子的情况下才具有活性。条件型活性启动子的一个亚类是诱导型启动子,其需要存在小分子“诱导剂”才有活性。诱导型启动子的实例包括但不限于阿拉伯糖诱导型启动子、Tet-on启动子和他莫昔芬诱导型启动子。各种组成型、条件型和诱导型启动子是本领域技术人员所熟知的,并且本领域技术人员将能够确定用于实施本发明的多种此类启动子,其在这方面不受限制。

[0375] 前间隔序列

[0376] 如本文所用,术语“前间隔序列(protospacer)”是指DNA中与PAM(前间隔序列邻近基序)序列相邻的序列(约20bp)。前间隔序列与向导RNA的间隔区序列共有相同的序列。向导RNA与靶DNA上的前间隔序列的互补序列退火(具体地,其一条链,即靶DNA序列的“靶链”相对于“非靶链”)。为了使Cas9发挥作用,还需要特定的前间隔序列邻近基序(PAM),该基序随Cas9基因的细菌种类而异。源自化脓性链球菌的最常用的Cas9核酸酶识别非靶链上的NGG的PAM序列,该序列直接位于基因组DNA靶序列下游。技术人员会理解,现有技术中的文献有时将“前间隔序列”称为向导RNA自身上的约20-nt靶标特异性向导序列,而不是将其称为“间隔区”。因此,在一些情况下,如本文使用的术语“前间隔序列”可与术语“间隔区”互换使用。围绕“前间隔序列”或“间隔区”出现的说明书上下文将有助于告知读者该术语是指gRNA还是指DNA靶标。

[0377] 前间隔序列邻近基序(PAM)

[0378] 如本文所用,术语“前间隔序列邻近基序”或“PAM”是指作为Cas9核酸酶的重要靶向组件的约2-6个碱基对的DNA序列。通常,PAM序列位于任一链上,并且位于Cas9切割位点5′至3′方向的下游。经典PAM序列(即,与化脓性链球菌的Cas9核酸酶或SpCas9相关的PAM序列)为5′-NGG-3′,其中“N”是任何核碱基,后面有两个鸟嘌呤(“G”)核碱基。不同的PAM序列可与来自不同生物体的不同Cas9核酸酶或等同蛋白相关联。此外,可修饰任何给定的Cas9核酸酶,如SpCas9,以改变核酸酶的PAM特异性,使得核酸酶识别替代的PAM序列。

[0379] 例如,参考经典SpCas9氨基酸序列为SEQ ID NO:18,PAM序列可通过引入一个或多

个突变来修饰，包括(a)D1135V、R1335Q和T1337R“VQR变体”，其改变了PAM对NGAN或NGNG的特异性，(b)D1135E、R1335Q和T1337R“EQR变体”，其改变了PAM对NGAG的特异性，和(c)D1135V、G1218R、R1335E和T1337R“VRR变体”，其改变了PAM对NGCG的特异性。此外，经典SpCas9的D1135E变体仍可识别NGG，但它相比野生型SpCas9蛋白更具选择性。

[0380] 还应理解来自不同细菌物种的Cas9酶(即，Cas9直系同源物)可具有不同的PAM特异性。例如，来自金黄色葡萄球菌(Staphylococcus aureus)的Cas9(SaCas9)识别NGRRT或NGRRN。此外，来自脑膜炎奈瑟菌(Neisseria meningitis)的Cas9(NmCas)识别NNNNGATT。在另一个示例中，来自嗜热链球菌(Streptococcus thermophilis)的Cas9(StCas9)识别NNAGAAW。在另一个示例中，来自齿状密螺旋体(Treponema denticola)的Cas9(TdCas)识别NAAAAC。这些是示例并不意味着限制。应进一步理解，非SpCas9结合多种PAM序列，这使得它们在期望的靶切割位点不存在合适的SpCas9 PAM序列时有用。此外，非SpCas9可以具有可使它们比SpCas9更有用的其他特征。例如，来自金黄色葡萄球菌的Cas9(SaCas9)比SpCas9小1kb左右，因此它可包装至腺相关病毒(AAV)中。可进一步参考Shah et al.,“Protospacer recognition motifs:mixed identities and functional diversity,”RNA Biology,10(5):891-899(通过引用并入本文)。

[0381] 重组酶

[0382] 如本文所用，术语“重组酶”是指介导重组酶识别序列之间的DNA重组从而导致重组酶识别序列之间DNA片段的切除、整合、倒置或交换(如，易位)的位点特异性酶。重组酶可分为两个不同家族：丝氨酸重组酶(如，解离酶和转化酶)和酪氨酸重组酶(如，整合酶)。丝氨酸重组酶的实例包括但不限于Hin、Gin、Tn3、β-6、CinH、ParA、γδ、Bxb1、φC31、TP901、TG1、φBT1、R4、φRV1、φFC1、MR11、A118、U153和gp29。酪氨酸重组酶的实例包括但不限于Cre、FLP、R、Lambda、HK101、HK022和pSAM2。丝氨酸和酪氨酸重组酶的名称源于重组酶用来攻击DNA并在链交换过程中与DNA共价连接的保守亲核氨基酸残基。重组酶有许多应用，包括创建基因敲除/敲入和基因治疗应用。参见，例如，Brown etal.,“Serine recombinases as tools for genome engineering.”Methods.2011;53(4):372-9;Hirano et al.,“Site-specific recombinases as tools for heterologous gene integration.”Appl.Microbiol.Biotechnol.2011;92(2):227-39;Chavez and Calos,“Therapeutic applications of theΦC31 integrase system.”Curr.Gene Ther.2011;11(5):375-81;Turan and Bode,“Site-specific recombinases:from tag-and-target-to tag-and-exchange-based genomic modifications.”FASEB J.2011;25(12):4088-107;Venken and Bellen,“Genome-wide manipulations of Drosophila melanogaster with transposons,Flp recombinase,andΦC31integrase.”Methods Mol.Biol.2012;859:203-28;Murphy,“Phage recombinases and their applications.”Adv.Virus Res.2012;83:367-414;Zhang et al.,“Conditional gene manipulation:Cre-ating a new biological era.”J.Zhejiang Univ.Sci.B.2012;13(7):511-24;Karpenshif and Bernstein,“From yeast to mammals:recent advances in genetic control of homologous recombination.”DNARepair(Amst).2012;1;11(10):781-8;其各自的全部内容都通过引用以其整体并入本文。本文提供的重组酶并不意味着是可用于本发明实施方案

的重组酶的排他性实例。本发明的方法和组合物可通过挖掘新的正交重组酶的数据库或设计具有限定的DNA特异性的合成重组酶来扩展(参见,例如,Groth et al.,“Phage integrases:biology and applications.”J.Mol.Biol.2004;335,667-678;Gordley et al.,“Synthesis of programmable integrases.”Proc.Natl.Acad.Sci.USA.2009;106,5053-5058;其各自的全部内容通过引用以其整体并入)。可用于本文所述方法和组合物的重组酶的其他实例是本领域技术人员已知的,并且预期发现或产生的任何新重组酶都能够用于本发明的不同实施方案中。在一些实施方案中,重组酶的催化结构域与核酸酶失活的RNA可编程核酸酶(如,dCas9或其片段)融合,使得重组酶结构域不包含核酸结合结构域或不能结合到靶核酸(如,将重组酶结构域工程化改造为使得其不具有特异性DNA结合活性)。缺乏DNA结合活性的重组酶和其工程化改造方法是已知的,包括描述于Klippel et al.,“Isolation and characterisation ofunusual gin mutants.”EMBO J.1988;7:3983-3989:Burke et al.,“Activating mutations of Tn3 resolvase marking interfaces important inrecombination catalysis andits regulation.MolMicrobiol.2004;51:937-948;Olorunniji et al.,“Synapsis and catalysis by activated Tn3 resolvase mutants.”Nucleic Acids Res.2008;36:7181-7191;Rowland et al.,“Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome.”Mol Microbiol.2009;74:282-298;Akopian et al.,“Chimeric recombinases with designed DNA sequence recognition.”Proc Natl Acad Sci USA.2003;100:8688-8691;Gordley et al.,“Evolution of programmable zinc finger-recombinases with activity inhuman cells.JMolBiol.2007;367:802-813;Gordley et al.,“Synthesis of programmable integrases.”ProcNatlAcadSci USA.2009;106:5053-5058;Arnold et al.,“Mutants ofTn3resolvase which do not require accessory binding sites for recombination activity.”EMBO J.1999;18:1407-1414;Gaj et al.,“Structure-guided reprogramming of serine recombinase DNA sequence specificity.”Proc Natl Acad Sci USA.2011;108(2):498-503;以及Proudfoot et al.,“Zinc finger recombinases with adaptable DNA sequence specificity.”PLoS One.2011;6(4):e19537的那些;其各自的全部内容通过引用并入。例如,解离酶-转化酶组(如,Tn3和γδ解离酶与Hin和Gin转化酶)中的丝氨酸重组酶具有含有自主催化和DNA结合结构域的模块结构(参见,例如,Grindley et al.,“Mechanism of site-specific recombination.”Ann Rev Biochem.2006;75:567-605,其全部内容通过引用并入)。因此,如本文所述,这些重组酶的催化结构域易于与核酸酶失活的RNA-可编程核酸酶(如,dCas9或其片段)重组,例如,在分离不需要任何辅助因子(如,DNA结合活性)的“激活”重组酶突变体之后(参见,例如,Klippel etal.,“Isolation and characterisationofunusual gin mutants.”EMBO J.1988;7:3983-3989:Burke et al.,“Activating mutations ofTn3 resolvase marking interfaces important inrecombination catalysis and its regulation.Mol Microbiol.2004;51:937-948;Olorunniji et al.,“Synapsis and catalysis by activated Tn3 resolvase mutants.”Nucleic Acids Res.2008;36:7181-7191;Rowland et al.,“Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome.”Mol Microbiol.2009;74:282-298;

Akopian et al.,“Chimeric recombinases with designed DNA sequence recognition.”ProcNatlAcadSci USA.2003;100:8688-8691)。此外,许多其他具有N端催化结构域和C端DNA结合结构域的天然丝氨酸重组酶是已知的(如,phiC31整合酶、TnpX转座酶、IS607转座酶),并且它们的催化结构域可被共同选择来工程化改造可编程位点特异性重组酶,如本文所述(参见,例如,Smith et al.,“Diversity in the serine recombinases.”Mol Microbiol.2002;44:299-307,其全部内容通过引用并入)。类似地,酪氨酸重组酶(如,Cre、λ整合酶)的核心催化结构域是已知的,并且可类似地被共同选择来工程化改造可编程位点特异性重组酶,如本文所述(参见,例如,Guo et al.,“Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse.”Nature.1997;389:40-46;Hartung et al.,“Cre mutants with altered DNA binding properties.”J Biol Chem 1998;273:22884-22891;Shaikh et al.,“Chimeras ofthe Flp and Cre recombinases:Tests ofthe mode of cleavage by Flp and Cre.JMolBiol.2000;302:27-48;Rongrong et al.,“Effect ofdeletion mutation on the recombination activity ofCre recombinase.”Acta Biochim Pol.2005;52:541-544;Kilbride et al.,“Determinants of product topology in a hybrid Cre-Tn3 resolvase site-specific recombination system.”J MolBiol.2006;355:185-195;Warren et al.,“A chimeric cre recombinase with regulated directionality.”ProcNatlAcadSci USA.2008105:18278-18283;Van Duyne,“Teaching Cre to follow directions.”ProcNatlAcadSci USA.2009Jan6;106(1):4-5;Numrych et al.,“A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophageλ.”Nucleic Acids Res.1990;18:3953-3959;Tirumalai et al.,“The recognition ofcore-type DNA sites byλintegrase.”JMol Biol.1998;279:513-527;Aihara et al.,“A conformational switch controls the DNAcleavage activity ofλ integrase.”Mol Cell.2003;12:187-198;Biswas et al.,“A structural basis for allosteric control of DNA recombination byλintegrase.”Nature.2005;435:1059-1066;以及Warren et al.,“Mutations in the amino-terminal domain ofλ-integrase have differential effects on integrative and excisive recombination.”MolMicrobiol.2005;55:1104-1112;其各自的全部内容通过引用并入)。

[0383] 重组酶识别序列

[0384] 如本文所用,术语“重组酶识别序列”或等同地“RRS”或“重组酶靶序列”或“重组酶位点”是指如下的核苷酸序列靶标,其由重组酶识别,并且与具有RRS的另一DNA分子进行链交换,从而导致重组酶识别序列之间的DNA片段切除、整合、倒置或交换。在不同的实施方案中,多链引导编辑器可以在靶序列中或者在一个以上靶序列中安装一个或多个重组酶位点。当通过多链引导编辑器安装一个以上重组酶位点时,重组酶位点可以被安装在邻近的靶位点处或者不邻近的靶位点处(例如,单独的染色体)。在不同的实施方案中,单个安装的重组酶位点可以用作用于基因组重组酶位点与外源性供应的核酸分子(例如,质粒)内的第二重组酶位点之间的重组酶介导的反应的“登陆位点”。这使得能够靶向整合期望的核酸分子。在其他实施方案中,当将两个重组酶位点插入到DNA的邻近区域中时(例如,相隔25-

50bp、50-100bp、100-200bp、200-300bp、300-400bp、400-500bp、500-600bp、600-700bp、700-800bp、800-900bp、900-1000bp、1000-2000bp、2000-3000bp、3000-4000bp、4000-5000bp或更多)，重组酶位点可以用于重组酶介导的插入序列的切除或倒置，或者用于重组酶介导的与具有相同重组酶位点的外源性DNA的盒交换。当通过多瓣引导编辑器在两个不同染色体上安装一个或多个重组酶位点时，可以发生从第一个染色体位置到第二的插入序列易位。

[0385] 重组

[0386] 在核酸修饰(如，基因组修饰)的背景中，术语“重组”用于指两个或多个核酸分子，或单个核酸分子的两个或多个区域通过重组酶蛋白(如，本文提供的本发明的重组酶融合蛋白)的作用修饰的过程。重组尤其可导致例如在一个或多个核酸分子之中或之间，核酸的插入、倒置、切除或易位。

[0387] 逆转录酶

[0388] 术语“逆转录酶”描述了表征为RNA依赖性DNA聚合酶的一类聚合酶。所有已知的逆转录酶都需要引物才能从RNA模板合成DNA转录本。历史上，逆转录酶主要用于将mRNA转录成cDNA，然后可将其克隆到载体中进行进一步操作。禽类成肌细胞增多(myoblastosis)病毒(AMV)逆转录酶是首次广泛使用的RNA依赖性DNA聚合酶(Verma，Biochim.Biophys.Acta473:1(1977))。该酶具有5′至3′RNA引导的DNA聚合酶活性、5′至3′DNA引导的DNA聚合酶活性和RNase H活性。RNase H是对用于RNA-DNA杂合体的RNA链特异的持续性5′和3′核糖核酸酶(Perbal,A Practical Guide to MolecularCloning,NewYork:Wiley&Sons(1984))。逆转录酶不能校正转录中的错误，因为已知的病毒逆转录酶缺乏校对所必需的3′至5′核酸外切酶活性(Saunders and Saunders,Microbial Genetics Applied to Biotechnology,London:Croom Helm(1987))。AMV逆转录酶活性及其相关RNase H活性的详细研究由Berger et al.,Biochemistry22:2365-2372(1983)提供。另一种广泛用于分子生物学的逆转录酶是源自莫洛尼(Moloney)鼠白血病病毒(M-MLV)的逆转录酶。参见，例如，Gerard,G.R.,DNA 5:271-279(1986)和Kotewicz,M.L.,et al.,Gene 35:249-258(1985)。还描述了基本上缺乏RNase H活性的M-MLV逆转录酶。参见，例如，美国专利号5,244,797。本发明考虑使用任何此类逆转录酶，或其变体或突变体。

[0389] 此外，本发明考虑使用易错的逆转录酶，即，其可称为易错逆转录酶或聚合过程中不支持高保真核苷酸掺入的逆转录酶。在与向导RNA一起整合的基于RT模板的单链DNA瓣合成过程中，易错逆转录酶可引入一个或多个与RT模板序列错配的核苷酸，从而通过单链DNA瓣的错误聚合向核苷酸序列引入变化。这些在单链DNA瓣合成过程中引入的错误然后如下整合至双链分子：与相应的内源性靶链杂交、去除内源性置换链、连接，然后再通过一轮内源性DNA修复和/或测序过程。

[0390] 逆转录

[0391] 如本文所用，术语“逆转录”表示酶使用RNA作为模板合成DNA链(即，互补DNA或cDNA)的能力。在一些实施方案中，逆转录可以是“易错逆转录”，其指某些逆转录酶在其DNA聚合活性中易错的特性。

[0392] PACE

[0393] 如本文所用，术语“噬菌体辅助连续进化(PACE)”是指采用噬菌体作为病毒载体的

连续进化。PACE技术的一般概念已在例如2009年9月8日提交的国际PCT申请PCT/US2009/056194,2010年3月11日以WO2010/028347公布;2011年12月22日提交的国际PCT申请PCT/US2011/066747,2012年6月28日以WO2012/088381公布;美国申请,于2015年5月5日授权的美国专利号9,023,594;于2015年1月20日提交的国际PCT申请PCT/US2015/012022,2015年9月11日以WO2015/134121公布;以及于2015年1月20日提交的国际PCT申请PCT/US2015/012022,2016年10月20日以WO2016/168631公布,其各自的全部内容通过引用并入本文。

[0394] 噬菌体

[0395] 如本文使用,可与术语“细菌噬菌体”互换使用的术语“噬菌体”,是指感染细菌细胞的病毒。通常,噬菌体由包裹遗传物质的外部蛋白衣壳组成。遗传物质可以是线性或环状形式的ssRNA、dsRNA、ssDNA或dsDNA。噬菌体和噬菌体载体是本领域技术人员所熟知的,并且可用于实施本文提供的PACE方法的噬菌体的非限制性示例是λ(溶素原)、T2、T4、T7、T12、R17、M13、MS2、G4、P1、P2、P4、PhiX174、N4、Φ6和Φ29。在某些实施方案中,本发明中使用的噬菌体是M13。其他合适的噬菌体和宿主细胞对本领域技术人员而言是显而易见的,并且本发明在这方面不受限制。有关其他合适的噬菌体和宿主细胞的示例性描述,参见Elizabeth Kutter and Alexander Sulakvelidze:Bacteriophages:Biology and Applications.CRC Press;1st edition(December 2004),ISBN:0849313368;Martha R.J.Clokie and Andrew M.Kropinski:Bacteriophages:Methods and Protocols,Volume 1:Isolation,Characterization,and Interactions(Methods in Molecular Biology)HumanaPress;1st edition(December,2008),ISBN:1588296822;Martha R.J.Clokie and Andrew M.Kropinski:Bacteriophages:Methods and Protocols,Volume 2:Molecular and Applied Aspects(Methods in Molecular Biology)HumanaPress;1st edition(December2008),ISBN:1603275649;为了公开合适的噬菌体和宿主细胞以及用于分离、培养和操作此类噬菌体的方法和方案,所有这些均通过引用以其整体并入本文)。

[0396] 蛋白、肽和多肽

[0397] 术语“蛋白”、“肽”和“多肽”在本文中可互换使用,是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。这些术语是指任何大小、结构或功能的蛋白、肽或多肽。通常,蛋白、肽或多肽的长度为至少3个氨基酸。蛋白、肽或多肽可指单个的蛋白或蛋白的集合。可修饰蛋白、肽或多肽中的一个或多个氨基酸,例如,通过添加化学实体如碳水化合物基团、羟基、磷酸基、法尼基、异法尼基、脂肪酸基团、用于缀合、功能化或其他修饰的接头等。蛋白、肽或多肽也可以是单分子或可以是多分子复合物。蛋白、肽或多肽可以只是天然存在的蛋白或肽的片段。蛋白、肽或多肽可以是天然存在的、重组的或合成的,或其任何组合。本文提供的任何蛋白可通过本领域已知的任何方法产生。例如,本文提供的蛋白可通过重组蛋白表达和纯化产生,这尤其适用于包含肽接头的融合蛋白。重组蛋白表达和纯化的方法是众所周知的,包括Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))中描述的那些,其全部内容其通过引用并入本文。

[0398] 蛋白剪接

[0399] 如本文所用,术语“蛋白剪接”是指从氨基酸序列内切除序列、内含肽(或断裂内含肽,视情况而定),并且氨基酸序列的剩余片段外显肽通过酰胺键连接形成连续的氨基酸序

列的过程。术语“反式”蛋白剪接是指内含肽是断裂内含肽并且它们位于不同蛋白上的特定情况。

[0400] 第二链切口产生

[0401] 作为引导编辑的结果形成的异源双链体DNA(即,包含一条编辑链和一条非编辑链)的分解决定了长期编辑结果。换言之,引导编辑的目标是通过将编辑链永久整合到互补的内源性链中来分解作为PE中间体形成的异源双链体DNA(编辑链与内源性非编辑链配对)。“第二链切口产生”的方法可在本文中用于帮助驱动异源双链体DNA的分解,以有利于编辑链永久整合到DNA分子中。如本文所用,“第二链切口产生”的概念是指在第一切口下游的位置引入第二切口(即,提供游离3’端的初始切口位点,用于在向导RNA的延伸部分上引发逆转录酶),优选在未编辑链上。在某些实施方案中,第一切口和第二切口位于相反链上。在其他实施方案中,第一切口和第二切口位于相反链上。在又一实施方案中,第一切口位于非靶链上(即,形成R环的单链部分的链),并且第二切口位于靶链上。在其他实施方案中,第一切口位于编辑链上,并且第二切口位于非编辑链上。第二切口可位于第一切口下游至少5个核苷酸处,或第一切口下游至少6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40、50、60、70、80、90、100、110、120、130、140或150个或更多个核苷酸处。在某些实施方案中,可在非编辑链上远离PEgRNA诱导的切口位点约5至150个核苷酸之间、或约5至140个之间、或约5至130个之间、或约5至120个之间、或约5至110个之间、或约5至100个之间、或约5至90个之间、或约5至80个之间、或约5至70个之间、或约5至60个之间、或约5至50个之间、或约5至40个之间、或约5至30个之间、或约5至20个之间、或约5至10个之间引入第二切口。在一个实施方案中,在远离PEgRNA诱导的切口14至116个核苷酸之间引入第二切口。不受理论的束缚,第二切口诱导细胞的内源性DNA修复和复制过程向着未编辑链置换或编辑,从而在两条链上永久安装编辑序列并分解由于PE形成的异源双链体。在一些实施方案中,编辑链是非靶链,并且未编辑链是靶链。在其他实施方案中,编辑链是靶链,并且未编辑链是非靶链。

[0402] 有义链

[0403] 在遗传学中,“有义”链是双链DNA中从5’至3’延伸,并且与DNA从3’至5’延伸的反义链或模板链互补的片段。在编码蛋白的DNA片段的情况下,有义链是与mRNA具有相同序列的DNA链,其在转录过程中以反义链为模板,并最终经历(通常,并非总是)翻译成蛋白。因此,反义链负责生成后来翻译成蛋白的RNA,而有义链具有与mRNA几乎相同的组成。注意,对于dsDNA的每个区段,可能会有两组有义和反义,这取决于读取的方向(因为有义和反义是相对于视角而言的)。最终规定dsDNA片段的哪条链被称为有义或反义的是基因产物或mRNA。

[0404] 在PEgRNA的情况下,第一步是离开PEgRNA延伸臂作为模板的以5’至3’方向定向的单链互补DNA(即,被掺入的3′ssDNA瓣)的合成。3′ssDNA瓣应当被视为有义链还是反义链取决于转录的方向,因为公认两条DNA链都可充当转录的模板(但不能同时)。因此,在一些实施方案中,3′ssDNA瓣(其总体上沿5’至3’方向延伸)将充当有义链,因为其是编码链。在其他实施方案中,3’ssDNA瓣(其总体上沿5’至3’方向延伸)将充当反义链,并因此作为转录模板。

[0405] 序列同源性

[0406] “同源性序列”或显示与另一序列的“同源性”的序列表示核酸分子的序列显示出与另一核酸分子有至少约65%、70%、75%、80%、85%或90%序列同一性。在其他实施方案中，核酸的“同源性序列”可以显示出与参考核酸有93%、95%或98%序列同一性。

[0407] 当指定序列同源性或同一性的百分比时，在两个核酸序列或两个多肽序列的情况下，同源性或同一性的百分比通常是指当进行最大对应比较或比对时两个或更多个序列在其一部分长度上的比对。除非另有说明，序列同源性或同一性在核酸、多肽或其部分的指定长度上评估。在一些实施方案中，同源性或同一性在长度的功能性部分或指定部分上评估。

[0408] 用于评估序列同源性的序列比对可以通过本领域已知的算法进行，比如基础局部比对检索工具(Basic Local Alignment Search Tool,BLAST)算法，其在Altschul et al，J.Mol.Biol.215:403-410,1990中描述。公众可用的执行BLAST分析的网络接口可以通过美国国家生物技术信息中心(the National Center for Biotechnology Information)访问。其他已知算法包括在以下文献中公布的那些算法:Smith&Waterman，“Comparison of Bio sequences”,Adv.Appl.Math.2:482,1981;Needleman&Wunsch，“A general method applicable to the search for similarities in the amino acid sequence of two proteins”J.Mol.Biol.48:443,1970;Pearson&Lipman“Improved tools for biological sequence comparison”,Proc.Natl.Acad.Sci.USA 85:2444,1988;或者通过这些或类似算法的自动化实现。也可以使用全局比对来比对大小大致相同的类似序列。全局比对程序的示例包括NEEDLE(在www.ebi.ac.uk/Tools/psa/emboss needle/可用)，其是EMBOSS包(Rice P et al.,Trends Genet.,2000;16:276-277)的一部分，以及GGSEARCH程序fasta.bioch.Virginia.edu/fasta_www2/fasta_www.cgi?rm＝compare&pgm＝gnw)，其是FASTA包(Pearson W and Lipman D,1988,Proc.Natl.Acad.Sci.USA,85:2444-2448)的一部分。这些程序都是基于Needleman-Wunsch算法，用于沿整个长度找到两个序列的最佳比对(包括间隙)。序列分析的详细讨论也可见于Ausubel et al(″Current Protocols in Molecular Biology″John Wiley&Sons Inc,1994-1998,Chapter 15,1998)的Unit 19.3。

[0409] 间隔区序列

[0410] 如本文所用，与向导RNA或PEgRNA相关的术语“间隔区序列”是指向导RNA或PEgRNA的约20个核苷酸的部分，其包含核苷酸序列，该核苷酸序列与靶DNA序列中前间隔序列序列具有相同的序列。间隔区序列与前间隔序列退火以在靶位点处形成ssRNA/ssDNA杂合体结构和前间隔序列内源性DNA链的相应R环ssDNA结构。

[0411] 受试者

[0412] 如本文所用，术语“受试者”是指个体生物体，如个体哺乳动物。在一些实施方案中，受试者是人。在一些实施方案中，受试者是非人类哺乳动物。在一些实施方案中，受试者是非人灵长类动物。在一些实施方案中，受试者是啮齿动物。在一些实施方案中，受试者是绵羊、山羊、牛、猫或狗。在一些实施方案中，受试者是脊椎动物、两栖动物、爬行动物、鱼、昆虫、苍蝇或线虫。在一些实施方案中，受试者是研究动物。在一些实施方案中，受试者是基因工程化的，例如基因工程化的非人类受试者。受试者可以是任何性别，也可以处于任何发育阶段。

[0413] 断裂内含肽

[0414] 尽管最常发现内含肽为连续结构域，但有些以自然断裂形式存在。在这种情况下，

这两个片段作为单独的多肽表达,并且必须在剪接发生之前缔合,即所谓的蛋白反式剪接。

[0415] 示例性断裂内含肽是Ssp DnaE内含肽,其包含两个亚基,即DnaE-N和DnaE-C。这两个不同的亚基由不同的基因编码,即dnaE-n和dnaE-c,分别编码DnaE-N和DnaE-C亚基。DnaE是天然存在于集胞藻属种(Synechocytissp.)PCC6803的断裂内含肽,且能够指导两种不同蛋白的反式剪接,每种蛋白都包含与DnaE-N或DnaE-C的融合物。

[0416] 其他天然存在或工程化的断裂内含肽序列是本领域已知的或可由本文所述完整内含肽序列或本领域可获得的那些制备。断裂内含肽序列的示例可见于Stevens et al.,"A promiscuous split intein with expanded protein engineering applications," PNAS,2017,Vol.114:8538-8543;Iwai et al.,"Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme,FEBS Lett,580:1853-1858,其各自通过引用并入本文。其他断裂内含肽序列可见于例如WO2013/045632、WO2014/055782、WO2016/069774和EP2877490,其各自的内容通过引用并入本文。

[0417] 此外,已在体内和体外描述了反式蛋白剪接(Shingledecker,et al.,Gene207:187(1998),Southworth,et al.,EMBOJ.17:918(1998);Mills,et al.,Proc.Natl.Acad.Sci.USA,95:3543-3548(1998);Lew,et al.,J.Biol.Chem.,273:15887-15890(1998);Wu,et al.,Biochim.Biophys.Acta 35732:1(1998b),Yamazaki,et al.,J.Am.Chem.Soc.120:5591(1998),Evans,et al.,J.Biol.Chem.275:9091(2000);Otomo,et al.,Biochemistry 38:16040-16044(1999);Otomo,et al.,J.Biolmol.NMR 14:105-114(1999);Scott,et al.,Proc.Natl.Acad.Sci.USA 96:13638-13643(1999)),并提供了关于随后经历连接以形成功能性产物的两个无活性片段表达蛋白的机会,例如,如图66和图67关于从两个单独表达的半部形成完整的PE融合蛋白所示。

[0418] 靶位点

[0419] 术语"靶位点"是指由本文公开的引导编辑器(PE)编辑的核酸分子内的序列。靶位点还指核酸分子内与引导编辑器(PE)和gRNA的复合物结合的序列。

[0420] tPERT

[0421] 参见"反式引导编辑器RNA模板(tPERT)"的定义。

[0422] 时序第二链切口产生

[0423] 如本文所用,术语"时序第二链切口产生"是指第二链切口产生的变体,由此在未编辑链中安装第二切口仅在编辑链中安装期望编辑之后发生。这避免了在两条链上同时出现切口,这可导致双链DNA断裂。将第二链切口产生性向导RNA设计用于时序控制,以便在安装期望编辑之后才引入第二链切口。这是通过设计具有间隔区序列的gRNA来实现的,该间隔区序列仅与编辑链相匹配,而不与原始等位基因相匹配。使用这种策略,前间隔序列和未编辑等位基因之间的错配应当不利于被sgRNA产生切口,直到PAM链上的编辑事件发生之后。

[0424] 反式引导编辑

[0425] 如本文所用,术语"反式引导编辑"或类似的"反式双引导编辑"是指利用断裂型PEgRNA的引导编辑或双引导编辑的修饰形式,即,其中将PEgRNA分成两个单独的分子:sgRNA和反式引导编辑RNA模板(tPERT)。sgRNA用于将引导编辑器靶向(或更一般地,将引导编辑器的napDNAbp组件靶向)到期望的基因组靶位点,而一旦通过位于引导编辑器和tPERT

上的结合结构域的相互作用将tPERT以反式募集至引导编辑器，tPERT由聚合酶(如，逆转录酶)用于将新的DNA序列写入靶基因座。在一个实施方案中，结合结构域可包括RNA-蛋白募集部分，例如位于tPERT上的MS2适体和融合至引导编辑器的MS2cp蛋白。反式引导编辑的优点是，通过将DNA合成模板与向导RNA分开，可潜在使用更长长度的模板。

[0426] 反式引导编辑的实施方案显示在图3G和图3H中。图3G左侧示出反式引导编辑器复合物("RP-PE:gRNA复合物")的组分，其包含与聚合酶(如，逆转录酶)和rPERT募集蛋白(如，MS2sc)中的每一个融合的napDNAbp，并与向导RNA复合。图3G还示出单独的tPERT分子，其包含PEgRNA的延伸臂特征，包括DNA合成模板和引物结合序列。tPERT分子还包括RNA-蛋白募集结构域(在这种情况下，其为茎环结构，可以是例如MS2适体)。如图3H中描述的过程所示，RP-PE:gRNA复合物结合靶DNA序列并对靶DNA序列产生切口。然后，募集蛋白(RP)募集tPERT以共定位至与DNA靶位点结合的引导编辑复合物，从而使引物结合位点与切口链上的引物序列结合，随后，允许聚合酶(如，RT)针对DNA合成模板合成DNA单链直到tPERT的5′。

[0427] 虽然图3G和图3H中示出tPERT在RNA-蛋白募集结构域的5'端包含PBS和DNA合成模板，但是其他配置中的tPERT可设计为具有位于RNA-蛋白募集结构域的3'端的PBS和DNA合成模板。然而，具有5'延伸的tPERT的优点是DNA单链的合成将在tPERT的5'端自然终止，因此在引导编辑的DNA合成阶段期间不会有使用RNA-蛋白募集结构域的任何部分作为模板的风险。

[0428] 反式引导编辑器RNA模板(tPERT)

[0429] 如本文所用，"反式引导编辑器RNA模板(tPERT)"是指用于反式引导编辑或反式双引导编辑的组件，所述反式引导编辑或反式双引导编辑是通过将PEgRNA分成两个不同的分子(向导RNA和tPERT分子)来运行的引导编辑的修饰版本。将tPERT分子编程为在靶DNA位点处与引导编辑器复合物共定位，从而将引物结合位点和DNA合成模板以反式带到引导编辑器。例如，参见图3G，用于反式引导编辑器(tPE)的实施方案，其示出双组件系统，其包含(1)RP-PE:gRNA复合物和(2)包含引物结合位点和连接到RNA-蛋白募集结构域的DNA合成模板的tPERT，其中RP-PE:gRNA复合物的RP(募集蛋白)组件将tPERT募集至待编辑的靶位点，从而将PBS和DNA合成模板与引导编辑器反式缔合。换言之，将tPERT工程化改造为包含(全部或部分)PEgRNA的延伸臂，其包括引物结合位点和DNA合成模板。

[0430] 转换

[0431] 如本文所用，"转换"是指嘌呤核碱基的互换$(A\leftrightarrow G)$或嘧啶核碱基的互换$(C\leftrightarrow T)$。这类互换涉及相似形状的核碱基。本文公开的组合物和方法能够诱导靶DNA分子中的一种或多种转换。本文公开的组合物和方法还能够在同一靶DNA分子中诱导转换和颠换。这些变化涉及$A\leftrightarrow G$、$G\leftrightarrow A$、$C\leftrightarrow T$、或$T\leftrightarrow C$。在具有Watson-Crick配对核碱基的双链DNA的情况下，颠换是指以下碱基对交换：$A:T\leftrightarrow G:C$、$G:G\leftrightarrow A:T$、$C:G\leftrightarrow T:A$、或$T:A\leftrightarrow C:G$。本文公开的组合物和方法能够诱导靶DNA分子中的一个或多个转换。本文公开的组合物和方法还能够诱导同一靶DNA分子中的转换和颠换，以及其他核苷酸变化，包括缺失和插入。

[0432] 颠换

[0433] 如本文所用，"颠换"是指嘌呤核碱基与嘧啶核碱基的互换，或相反，因此涉及具有

不相似形状的核碱基的互换。这些变化涉及T↔A、T↔G、C↔G、C↔A、A↔T、A↔C、G↔C、和G↔T。在具有Watson-Crick配对核酸碱基的双链DNA的情况下，颠换是指以下碱基对交换：T:A↔A:T、T:A↔G:C、C:G↔G:C、C:G↔A:T、A:T↔T:A、A:T↔C:G、G:C↔C:G、和G:C↔T:A。本文公开的组合物和方法能够诱导靶DNA分子中的一个或多个颠换。本文公开的组合物和方法还能够诱导同一靶DNA分子中的转换和颠换两者，以及其他核苷酸变化，包括缺失和插入。

[0434] 治疗

[0435] 术语“治疗”是指旨在对如本文所述疾病或病症或其一种或多种症状进行逆转、减轻、延迟发作或抑制进展的临床干预。如本文所用，术语“治疗”是指旨在对如本文所述疾病或病症或其一种或多种症状进行逆转、减轻、延迟发作或抑制进展的临床干预。在一些实施方案中，可在一种或多种症状出现后和/或疾病被诊断后进行治疗。在其他实施方案中，可在没有症状的情况下进行治疗，例如以预防或延迟症状的发作或抑制疾病的发作或进展。例如，可在症状出现之前对易感个体进行治疗(如，根据症状史和/或根据遗传或其他易感因素)。在症状消退后也可继续治疗，例如，以防止或延迟其复发。

[0436] 三核苷酸重复病症

[0437] 如本文所用，“三核苷酸重复病症”(或者，“扩增重复病症”或“重复扩增病症”)是指由“三核苷酸重复扩增”引起的一组遗传病症，其是某些基因或内含子中某些三核苷酸重复的一种突变。曾经认为三核苷酸重复是基因组中常见的重复，但在20世纪90年代澄清了这些病症。这些明显“良性”的DNA段有时会扩增并导致疾病。三核苷酸重复扩增引起的病症共有几个限定的特征。首先，突变重复显示体细胞和种系的不稳定性两者，而且更常见地，它们在连续传播中扩增而不是缩减。其次，较早的发病年龄和后代(预期)中表型的严重程度增加通常与较大的重复长度相关。最后，疾病等位基因的亲本起源通常会影响预期，对于这些病症中的许多种，父系传播具有更大的扩增风险。

[0438] 认为三联体扩增是由DNA复制过程中的滑移引起的。由于这些区域中DNA序列的重复性质，可能在保持母链和正在合成的子链之间的互补碱基配对的情况下在DNA复制过程中形成“环出(loop out)”结构。如果环出结构是由子链上的序列形成的，这将导致重复数增加。但是，如果在母链上形成环出结构，则发生重复数减少。这些重复的扩增似乎比减少更常见。一般地，扩增越多，它们就越有可能引起疾病或增加疾病的严重性。此特性导致在三核苷酸重复病症中看到预期的特征。预期描述了由于这些重复的扩增，经由受影响家庭的连续几代，发病年龄减小和症状严重程度增加的趋势。

[0439] 核苷酸重复病症可包括其中三联体重复出现在非编码区(即，非编码三核苷酸重复病症)或编码区中的那些。

[0440] 本文所述的引导编辑器(PE)系统可用于治疗核苷酸重复病症，其可包括脆性X综合征(FRAXA)、脆性XE MR(FRAXE)、弗里德赖希氏共济失调(FRDA)、肌强直性营养不良(DM)、脊髓小脑共济失调8型(SCA8)和脊髓小脑共济失调12型(SCA12)等。

[0441] 上游

[0442] 如本文所用，术语“上游”和“下游”是相对性术语，其定义了位于以5’至3′方向定

向的核酸分子(无论是单链还是双链)中的至少两个元件的线性位置。特别地,在第一元件位于第二元件5'的某处的核酸分子中,第一元件在第二元件的上游。例如,如果SNP位于切口位点的5'侧,则SNP位于Cas9诱导的切口位点的上游。相反,在第一元件位于第二元件3'的某处的核酸分子中,第一元件位于第二元件的下游。例如,如果SNP位于切口位点的3'侧,则SNP位于Cas9诱导的切口位点的下游。核酸分子可以是DNA(双链或单链),RNA(双链或单链),或DNA和RNA的杂合体。单链核酸分子和双链分子的分析相同,因为术语上游和下游仅参考核酸分子的单链,只是需要选择双链分子的哪条链来考虑。通常,可用于确定至少两个元件的位置相对性的双链DNA的链是"有义"或"编码"链。在遗传学中,"有义"链是双链DNA中从5'至3'延伸的片段,其与DNA的反义链或模板链(其从3'至5'延伸)互补。因此,例如,如果SNP核碱基在有义链或编码链上启动子的3'侧,则SNP核碱基位于基因组DNA(其是双链)中启动子序列的"下游"。

[0443] 变体

[0444] 如本文所用,术语"变体"应被理解指表现出具有与自然界发生的模式偏离的模式的特性,例如,与野生型Cas9氨基酸序列相比,变体Cas9是包含一个或多个氨基酸残基变化的Cas9。术语"变体"包括如下的同源蛋白,其与参考序列具有至少75%、或至少80%、或至少85%、或至少90%、或至少95%、或至少99%百分比同一性,并且具有与参考序列相同或基本相同的功能活性。该术语还包括参考序列的突变体、截短或结构域,并且显示与参考序列相同或基本相同的一个或多个功能活性。

[0445] 载体

[0446] 如本文所用,术语"载体"是指可经修饰以编码感兴趣基因并且能够进入宿主细胞,在宿主细胞内发生突变和复制,然后转移复制形式的载体进入另一宿主细胞的核酸。示例性的合适载体包括病毒载体,例如逆转录病毒载体或噬菌体和丝状噬菌体,以及接合质粒。基于本公开,其他合适的载体对本领域技术人员而言将是显而易见的。

[0447] 野生型

[0448] 如本文所用,术语"野生型"是技术人员理解的本领域术语并且意指如其在自然界中存在的典型形式的生物体、菌株、基因或特征,与突变体或变体形式不同。

[0449] 5'内源性DNA瓣

[0450] 如本文所用,术语"5'内源性DNA瓣"是指位于靶DNA中紧邻PE诱导切口位点下游的DNA链。由PE对靶DNA链产生切口暴露切口位点上游侧的3'羟基和切口位点下游侧的5'羟基。以3'羟基结尾的内源性链用于引发引导编辑器的DNA聚合酶(如,其中DNA聚合酶是逆转录酶)。在切口位点下游侧且以暴露的5'羟基开始的内源性链称为"5'内源性DNA瓣",最终被去除并被由PEgRNA的延伸编码的新合成的置换链(即,"3'置换DNA瓣")置换。

[0451] 5'内源性DNA瓣去除

[0452] 如本文所用,术语"5'内源性DNA瓣去除"或"5'瓣去除"是指当RT合成的单链DNA瓣竞争性侵入内源性DNA并与内源性DNA杂交,在此过程中置换内源性链时形成的5'内源性DNA瓣的去除。去除这种内源性置换链可驱动反应向着形成包含期望的核苷酸变化的期望产物。细胞自身的DNA修复酶可催化5'内源性瓣的去除或切除(如,瓣核酸内切酶,如EXO1或FEN1)。也可转化宿主细胞以表达催化去除所述5'内源性瓣,从而驱动过程向着产物形成的一种或多种酶(如,瓣核酸内切酶)。瓣核酸内切酶是本领域已知的,可见于且描述于Patel

et al.,“Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends,” Nucleic Acids Research,2012,40(10):4507-4519,以及Tsutakawa et al.,“Human flap endonuclease structures,DNA double-base flipping,and a unified understanding of the FEN1 superfamily,”Cell,2011,145(2):198-211(各自都通过引用并入本文)。

[0453] 3’置换DNA瓣

[0454] 如本文所用,术语“3’置换DNA瓣”或简称为“置换DNA瓣”是指由引导编辑器合成并由引导编辑器PEgRNA的延伸臂编码的DNA链。更具体地,3’置换DNA瓣由PEgRNA的聚合酶模板编码。3’置换DNA瓣包含与5’内源性DNA瓣相同的序列,只是其还包含编辑序列(如,单核苷酸变化)。3’置换DNA瓣与靶DNA退火,置换或替换5’内源性DNA瓣(如,可通过5’瓣核酸内切酶如FEN1或EXO1切除),然后连接以联结3’置换DNA瓣的3’端与内源性DNA的暴露的5’羟基端(在切除5’内源性DNA瓣后暴露,从而重新形成磷酸二酯键并安装3’置换DNA瓣以形成异源双链体DNA,其包含一条编辑链和一条未编辑链。DNA修复过程通过将编辑链中的信息复制到互补链来分解异源双链体,从而将编辑永久安装到DNA中。此分解过程可通过对未编辑链产生切口进一步驱动,即,通过“第二链切口产生”,如本文所述。

[0455] 某些实施方案的详细描述

[0456] 采用成簇规律间隔短回文重复序列(CRISPR)系统进行基因组编辑已经彻底改变了生命科学[1-3]。尽管使用CRISPR进行基因破坏现已成为常规,但单核苷酸编辑的精确安装仍然是重大挑战,尽管对于研究或校正大量致病性突变是必要的。同源定向修复(HDR)能够实现此类编辑,但遭受效率低下(通常<5%),供体DNA修复模板的需要以及双链DNA断裂(DSB)形成的有害影响。最近,David Liu教授等的实验室开发了碱基编辑,无需DSB即可实现高效的单核苷酸编辑。碱基编辑器(BE)将CRISPR系统与碱基修饰性脱氨酶结合,将靶C·G或A·T碱基对分别转换为A·T或G·C[4-6]。尽管已被全世界的研究人员广泛使用,但当前的BE仅实现12种可能的碱基对转换中的4种,并且无法校正小的插入或缺失。此外,碱基编辑的靶向范围受到与靶碱基相邻的非靶C或A碱基的编辑(“旁观者编辑”)以及受到PAM序列存在于距靶碱基15±2bp的要求的限制。因此,克服这些限制将大大拓宽基因组编辑的基础研究和治疗应用。

[0457] 已开发了新的精确编辑方法(例如,经典引导编辑),该方法提供了碱基编辑的许多益处——即避免双链断裂和供体DNA修复模板——而克服其主要限制。本文所提出的方法使用靶标引发的逆转录(TPRT)在靶基因组位点处直接安装编辑的DNA链。在本文讨论的设计中,将CRISPR向导RNA(gRNA)工程化改造为携带逆转录酶(RT)模板序列,该序列编码包含期望的核苷酸变化的单链DNA。CRISPR核酸酶(Cas9)产生切口的靶位点DNA将充当经修饰的gRNA上模板序列逆转录的引物,允许直接掺入任何期望的核苷酸编辑。

[0458] 可利用或调整靶标引发的逆转录(TPRT)的机制进行基于CRISPR/Cas的精确基因组编辑,具有高效率和遗传可塑性(例如如图1A-1F的不同实施方案所描绘)。使用Cas蛋白-逆转录酶融合物来靶向具有经修饰的向导RNA(“延伸的向导RNA”)的特定DNA序列,在靶位点处产生单链切口,并使用有切口的DNA作为已整合到延伸的向导RNA中的工程化逆转录酶模板的逆转录引物。除了包含期望的核苷酸变化(如,单核苷酸变化、缺失或插入,或其组合)外,新合成的链将与基因组靶序列同源。新合成的DNA链可被称为单链DNA瓣,其将与互

补的同源内源性DNA链竞争杂交,从而置换相应的内源性链。这种杂合中间体的分解可包括去除由此产生的内源性DNA的置换瓣(如,使用5'端DNA瓣核酸内切酶,FEN1),将合成的单链DNA瓣连接到靶DNA,以及由于细胞DNA修复和/或复制过程导致期望的核苷酸变化的同化。由于提供模板的DNA合成提供单核苷酸精确度,因此这种方法的范围非常广泛,可预见地用于基础科学和治疗学中的无数应用。

[0459] 本发明描述了用于基因组编辑的新平台,称为"多瓣引导编辑"(包括,例如,"双瓣引导编辑"和"四瓣引导编辑"),并且代表了"引导编辑"或"经典引导编辑"的创新进展,如本发明人在Anzalone,A.V.et al.Search-and-replace genome editing without double-strandbreaks ordonorDNA.Nature 576,149-157(2019),该文献通过引用并入本文。然而不同的实施方案中的经典引导编辑在切口位点处具有单个3′瓣,后者被整合到同一链上的靶标核酸中,本文所述多瓣引导编辑系统涉及不同的构建体、系统和方法,它们在各个方案中产生位于不同链上的一对或多对3′瓣,后者形成包含期望的编辑的双链体,并且整合到靶标核酸分子中,例如,在基因组中的特定基因座处或编辑位点处。在各个方面,一对或多对3′瓣形成双链体,这是因为它们包含反向互补序列,这些序列一旦由本文所述引导编辑器产生就彼此退火。双链体通过细胞驱动机制整合到靶位点中,自然地替代位于相邻切口位点之间的内源性双链体序列。在某些实施方案中,新的双链体序列可以在一个或多个位置处引入(例如,在相邻的基因组基因座之间或在两个不同的染色体位置上),并且可以包含一个或多个感兴趣的序列,例如,蛋白质编码序列、肽编码序列或RNA编码序列。在一个实施方案中,通过多瓣引导编辑系统建立的新的双链体序列可以包含重组酶位点,例如,Bxb1重组酶attB(38bp)和/或attP(50bp)位点,或者由Hin重组酶、Gin重组酶、Tn3重组酶、β-six重组酶、CinH重组酶、ParA重组酶、γδ重组酶、ΦC31重组酶、TP901重组酶、TG1重组酶、φBT1重组酶、R4重组酶、φRV1重组酶、φFC1重组酶、MR11重组酶、A118重组酶、U153重组酶和gp29重组酶、Cre重组酶、FLP重组酶、R重组酶、Lambda重组酶、HK101重组酶、HK022重组酶以及pSAM2重组酶识别的重组酶位点。

[0460] 在多个方面,本说明书描述了多瓣引导编辑系统(包括,例如,双引导编辑系统和四引导编辑系统),其通过同时编辑两条DNA链解决了与瓣平衡和随后将编辑加入未编辑的互补基因组DNA链相关的挑战。在双瓣引导编辑系统中,例如,使用两个pegRNA靶向基因组位点的相对链并且指导含有编辑的DNA序列的两个互补3'瓣的合成(图91)。与经典引导编辑不同,这对编辑的DNA链(3'瓣)不需要与5'瓣在内源性基因组DNA中直接竞争,因为互补编辑链可替代用于杂交。由于双链体的两条链都作为编辑的DNA合成,双瓣引导编辑系统省略了经典引导编辑需要的替换未编辑的互补DNA链的需求。相反,细胞DNA修复机制仅需要切除成对的5'瓣(初始基因组DNA)并且将成对的3'瓣(编辑的DNA)连接到基因座。因此,在新合成的DNA链中也不需要包含与基因组DNA同源的序列,允许新链的选择性杂交并且有利于含有最小基因组同源性的编辑。切割DNA的两条链的引导编辑器的核酸酶活性形式也可以用于加快初始DNA序列的去除。使用四pegRNA的四瓣引导编辑系统提供了相似的优点。

[0461] [1]napDNAbp

[0462] 本文所述的多瓣引导编辑器和反式引导编辑器可包含核酸可编程DNA结合蛋白(napDNAbp)。

[0463] 一方面,napDNAbp可与至少一个向导核酸(如,向导RNA或PEgRNA)结合或复合,所

述向导核酸将napDNAbp定位于包含与向导核酸或其一部分(如,与DNA靶的前间隔序列退火的向导RNA的间隔区)互补的DNA链(即,靶链)的DNA序列。换言之,向导核酸“编程”napDNAbp(如,Cas9或等同物)以定位并结合到DNA中的前间隔序列的互补序列。

[0464] 可将任何合适的napDNAbp用于本文所述的多瓣引导编辑器中。在不同实施方案中,napDNAbp可以是任何2类CRISPR-Cas系统,包括任何II型、V型或VI型CRISPR-Cas酶。鉴于CRISPR-Cas作为基因组编辑工具的快速发展,用于描述和/或鉴定CRISPR-Cas酶的命名法一直在不断发展,例如Cas9和Cas9直系同源物。本申请参考可能是旧的和/或新的CRISPR-Cas酶命名法。本领域技术人员将能够基于所使用的命名法(无论它是旧(即“遗留”)的命名法还是新的命名法)来确定在本申请中提及的特定CRISPR-Cas酶。CRISPR-Cas命名法广泛讨论于Makarova et al.,“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?,”The CRISPR Journal,Vol.1.No.5,2018,其全部内容通过引用并入本文。在本申请的任何给定实例中使用的特定CRISPR-Cas命名法不以任何方式限制,并且本领域技术人员将能够确定所引用的是哪种CRISPR-Cas酶。

[0465] 例如,以下II型、V型和VI型2类CRISPR-Cas酶具有以下本领域公认的旧(即,旧有)和新的名称。这些酶中的每一种和/或其变体可与本文所述的多瓣引导编辑器一起使用:

[0466]

旧有命名法	当前命名法*
II型CRISPR-Cas酶	
Cas9	相同
V型CRISPR-Cas酶	
Cpf1	Cas12a
CasX	Cas12e
C2c1	Cas12b1
Cas12b2	相同
C2c3	Cas12c
CasY	Cas12d
C2c4	相同
C2c8	相同
C2c5	相同
C2c10	相同
C2c9	相同

[0467]

VI型CRISPR-Cas酶	
C2c2	Cas13a
Cas13d	相同
C2c7	Cas13c
C2c6	Cas13b

[0468] *参见Makarovaetal.,The CRISPRJournal,Vol.1,No.5,2018

[0469] 不受理论的束缚,本文考虑的某些napDNAbp的作用机制包括形成R-环的步骤,由此napDNAbp诱导双链DNA靶的解旋,从而分离由napDNAbp结合的区域中的链。然后,向导RNA间隔区与前间隔序列处的“靶链”杂交。这置换了与靶链互补的“非靶链”,这形成了R环的单链区域。在一些实施方案中,napDNAbp包括一种或多种核酸酶活性,其切割DNA,留下不同类型的损伤。例如,napDNAbp可包含在第一位置处切割非靶链和/或在第二位置处切割靶链的核酸酶活性。根据核酸酶活性,可切割靶DNA以形成“双链断裂”,从而切割两条链。在其他实施方案中,可仅在单个位点处切割靶DNA,即在一条链上对DNA“产生切口”。具有不同核酸酶

活性的示例性napDNAbp包括“Cas9切口酶”(“nCas9”)和没有核酸酶活性的失活Cas9(“死亡Cas9”或“dCas9”)。

[0470] 可与目前公开的多瓣引导编辑器结合使用的各种napDNAbps的以下描述并不意味着以任何方式进行限制。多瓣引导编辑器可包括经典SpCas9,或任何直系同源Cas9蛋白,或任何变体Cas9蛋白——包括任何天然存在的Cas9变体、突变体或其他工程化版本——其是已知的或可通过定向进化或其他诱变过程制备或进化。在不同实施方案中,Cas9或Cas9变体具有切口酶活性,即仅切割靶DNA序列的一条链。在其他实施方案中,Cas9或Cas9变体具有无活性的核酸酶,即“死亡”Cas9蛋白。可使用的其他变体Cas9蛋白是具有比经典SpCas9更小的分子量(如,为了更容易递送)或具有经修饰或重排的一级氨基酸结构(如,环状排列形式)的那些。

[0471] 本文所述的多瓣引导编辑器还可包含Cas9等同物,包括Cas12a(Cpf1)和Cas12b1蛋白,其是趋同进化的结果。本文使用的napDNAbps(如,SpCas9、Cas9变体或Cas9等同物)还可包含改变/增强其PAM特异性的各种修饰。最后,本申请考虑与参考Cas9序列(如,参考SpCas9经典序列或参考Cas9等同物(如,Cas12a(Cpf1)))具有至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.9%序列同一性的任何Cas9、Cas9变体或Cas9等同物。

[0472] napDNAbp可以是CRISPR(成簇规律间隔短回文重复序列)相关的核酸酶。如上所述,CRISPR是针对可动遗传元件(病毒、转座元件和接合质粒)提供保护的适应性免疫系统。CRISPR簇包含间隔区、与前面的可动元件互补的序列和靶侵入核酸。将CRISPR簇转录并加工成CRISPRRNA(crRNA)。在II型CRISPR系统中,正确处理pre-crRNA需要反式编码小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当核糖核酸酶3辅助加工pre-crRNA的向导物。随后,Cas9/crRNA/tracrRNA以内切溶核的方式切割与间隔区互补的线性或环状dsDNA靶。与crRNA不互补的靶链首先以内切溶核方式切割,然后以外切融合方式以3’-5’修剪。事实上,DNA结合和切割通常需要蛋白和这两种RNA。然而,可对单链向导RNA(“sgRNA”,或简称为“gRNA”)进行工程化改造,以便将crRNA和tracrRNA的各个方面都并入单一RNA种类中。参见,例如,JinekM.et al.,Science 337:816-821(2012),其全部内容通过引用并入本文。

[0473] 在一些实施方案中,napDNAbp引导切割靶序列位置(如,靶序列内和/或靶序列的互补序列内)处的一条或两条链。在一些实施方案中,napDNAbp引导从靶序列的第一个或最后一个核苷酸的约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多碱基对内一条或两条链的切割。在一些实施方案中,载体编码napDNAbp,其相对于相应的野生型酶发生突变,使得突变的napDNAbp缺乏切割含有靶序列的靶多核苷酸的一条或两条链的能力。例如,来自化脓性链球菌的Cas9的RuvC I催化域中的天冬氨酸至丙氨酸取代(D10A)将Cas9从切割两条链的核酸酶转化为切口酶(切割单链)。使Cas9成为切口酶的其他突变实例包括但不限于参考经典SpCas9序列或其他Cas9变体或Cas9等同物中的等同氨基酸位点的H840A、N854A和N863A。

[0474] 如本文所用,术语“Cas蛋白”是指从自然界获得的全长Cas蛋白、具有与天然存在的Cas蛋白不同的序列的重组Cas蛋白、或Cas蛋白的任何片段,不过其保留所公开方法所需

的全部或大量必需的基本功能，即，(i)具有Cas蛋白与靶DNA的核酸可编程结合，和(ii)对一条链上靶DNA序列产生切口的能力。本文考虑的Cas蛋白包括CRISPR Cas9蛋白，以及Cas9等同物、变体(如，Cas9切口酶(nCas9)或核酸酶无活性Cas9(dCas9))同源物、直系同源物或旁系同源物，无论是天然存在的还是非天然存在的(如，工程化或重组)，并且可包括来自任何2类CRISPR系统(如，II、V、VI型)的Cas9等同物，包括Cas12a(Cpf1)、Cas12e(CasX)、Cas12b1(C2c1)、Cas12b2、Cas12c(C2c3)、C2c4、C2c8、C2c5、C2c10、C2c9Cas13a(C2c2)、Cas13d、Cas13c(C2c7)、Cas13b(C2c6)和Cas13b。其他Cas等同物描述于Makarova et al.,“C2c2 is a single-componentprogrammable RNA-guided RNA-targeting CRISPR effector,”Science 2016;353(6299)以及Makarova et al.,“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?,”The CRISPR Journal,Vol.1.No.5,2018，其内容通过引用并入本文。

[0475] 术语“Cas9”或“Cas9核酸酶”或“Cas9部分”或“Cas9结构域”包括来自任何生物体的任何天然存在的Cas9、任何天然存在的Cas9等同物或其功能片段、任何Cas9同源物、直系同源物或来自任何生物体的旁系同源物，以及Cas9的天然存在或工程化改造的任何突变体或变体。术语Cas9并不意味着特别限制并且可被称为“Cas9或等同物”。示例性Cas9蛋白在本文中进一步描述和/或在本领域中描述，并且通过引用并入本文。本公开对于在本文所述的多瓣引导编辑器中使用的特定Cas9是不受限制的。

[0476] 如本文所述，Cas9核酸酶序列和结构是本领域技术人员所熟知的(参见，例如，“Complete genome sequence ofanM1 strainofStreptococcuspyogenes.”Ferretti et al.,J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001)；“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011)；and“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,CharpentierE.Science 337:816-821(2012)，其各自的全部内容通过引用并入本文)。

[0477] Cas9和Cas9等同物的实例如下提供；然而，这些具体示例并不意味着限制。本公开的多瓣引导编辑器可使用任何合适的napDNAbp，包括任何合适的Cas9或Cas9等同物。

[0478] A.野生型经典SpCas9

[0479] 在一个实施方案中，本文所述的多瓣引导编辑器构建体可包含来自化脓性链球菌的“经典SpCas9”核酸酶，其已被广泛用作基因组工程的工具并被归类为2类CRISPR-Cas系统的II型亚组酶。这种Cas9蛋白是大型的多结构域蛋白，包含两个不同的核酸酶结构域。可将点突变引入Cas9以消除一种或两种核酸酶活性，分别导致仍保留以sgRNA编程方式结合DNA的能力的切口酶Cas9(nCas9)或死亡Cas9(dCas9)。原则上，当与另一蛋白或结构域融合时，Cas9或其变体(如，nCas9)可通过与适当的sgRNA共表达将蛋白靶向到几乎任何DNA序列。如本文所用，经典SpCas9蛋白是指来自具有以下氨基酸序列的化脓性链球菌的野生型

蛋白：

[0480]

描述	序列	SEQ ID NO:
SpCas9 化脓性链球菌M1 SwissProt 登录号 Q99ZW2 野生型	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV	SEQ ID NO: 18

[0481]

	QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
SpCas9 反向翻译 SwissProt 登录号 Q99ZW2 化脓性链球 菌	ATGGATAAAAAATATAGCATTGGCCTGGATATTGGC ACCAACAGCGTGGGCTGGGCGGTGATTACCGATGAA TATAAAGTGCCGAGCAAAAAATTTAAAGTGCTGGGC AACACCGATCGCCATAGCATTAAAAAAAACCTGATT GGCGCGCTGCTGTTTGATAGCGGCGAAACCGCGGAA GCGACCCGCCTGAAACGCACCGCGCGCCGCCGCTAT ACCCGCCGCAAAAACCGCATTTGCTATCTGCAGGAA ATTTTTAGCAACGAAATGGCGAAAGTGGATGATAGC TTTTTTCATCGCCTGGAAGAAAGCTTTCTGGTGGAAG AAGATAAAAAACATGAACGCCATCCGATTTTTGGCA ACATTGTGGATGAAGTGGCGTATCATGAAAAATATC CGACCATTTATCATCTGCGCAAAAAACTGGTGGATA GCACCGATAAAGCGGATCTGCGCCTGATTTATCTGG CGCTGGCGCATATGATTAAATTTCGCGGCCATTTTCT GATTGAAGGCGATCTGAACCCGGATAACAGCGATGT GGATAAACTGTTTATTCAGCTGGTGCAGACCTATAA CCAGCTGTTTGAAGAAAACCCGATTAACGCGAGCGG CGTGGATGCGAAAGCGATTCTGAGCGCGCGCCTGAG CAAAAGCCGCCGCCTGGAAAACCTGATTGCGCAGCT GCCGGGCGAAAAAAAAAACGGCCTGTTTGGCAACCT GATTGCGCTGAGCCTGGGCCTGACCCCGAACTTTAA AAGCAACTTTGATCTGGCGGAAGATGCGAAACTGCA GCTGAGCAAAGATACCTATGATGATGATCTGGATAA CCTGCTGGCGCAGATTGGCGATCAGTATGCGGATCT GTTTCTGGCGGCGAAAAACCTGAGCGATGCGATTCT GCTGAGCGATATTCTGCGCGTGAACACCGAAATTAC CAAAGCGCCGCTGAGCGCGAGCATGATTAAACGCTA TGATGAACATCATCAGGATCTGACCCTGCTGAAAGC GCTGGTGCGCCAGCAGCTGCCGGAAAAATATAAAG AAATTTTTTTTGATCAGAGCAAAAACGGCTATGCGG GCTATATTGATGGCGGCGCGAGCCAGGAAGAATTTT ATAAATTTATTAAACCGATTCTGGAAAAAATGGATG GCACCGAAGAACTGCTGGTGAAACTGAACCGCGAA GATCTGCTGCGCAAACAGCGCACCTTTGATAACGGC AGCATTCCGCATCAGATTCATCTGGGCGAACTGCAT GCGATTCTGCGCCGCCAGGAAGATTTTTATCCGTTTC TGAAAGATAACCGCGAAAAAATTGAAAAAATTCTG ACCTTTCGCATTCCGTATTATGTGGGCCCGCTGGCGC GCGGCAACAGCCGCTTTGCGTGGATGACCCGCAAAA GCGAAGAAACCATTACCCCGTGGAACTTTGAAGAAG TGGTGGATAAAGGCGCGAGCGCGCAGAGCTTTATTG AACGCATGACCAACTTTGATAAAAACCTGCCGAACG AAAAAGTGCTGCCGAAACATAGCCTGCTGTATGAAT ATTTTACCGTGTATAACGAACTGACCAAAGTGAAAT ATGTGACCGAAGGCATGCGCAAACCGGCGTTTCTGA	SEQ ID NO: 19

[0482]

	GCGGCGAACAGAAAAAAGCGATTGTGGATCTGCTGT TTAAAACCAACCGCAAAGTGACCGTGAAACAGCTGA AAGAAGATTATTTTAAAAAAATTGAATGCTTTGATA GCGTGGAAATTAGCGGCGTGGAAGATCGCTTTAACG CGAGCCTGGGCACCTATCATGATCTGCTGAAAATTA TTAAAGATAAAGATTTTCTGGATAACGAAGAAAACG AAGATATTCTGGAAGATATTGTGCTGACCCTGACCC TGTTTGAAGATCGCGAAATGATTGAAGAACGCCTGA AAACCTATGCGCATCTGTTTGATGATAAAGTGATGA AACAGCTGAAACGCCGCCGCTATACCGGCTGGGGCC GCCTGAGCCGCAAACTGATTAACGGCATTCGCGATA AACAGAGCGGCAAAACCATTCTGGATTTTCTGAAAA GCGATGGCTTTGCGAACCGCAACTTTATGCAGCTGA TTCATGATGATAGCCTGACCTTTAAAGAAGATATTC AGAAAGCGCAGGTGAGCGGCCAGGGCGATAGCCTG CATGAACATATTGCGAACCTGGCGGGCAGCCCGGCG ATTAAAAAAGGCATTCTGCAGACCGTGAAAGTGGTG GATGAACTGGTGAAAGTGATGGGCCGCCATAAACCG GAAAACATTGTGATTGAAATGGCGCGCGAAAACCA GACCACCCAGAAAGGCCAGAAAAACAGCCGCGAAC GCATGAAACGCATTGAAGAAGGCATTAAAGAACTG GGCAGCCAGATTCTGAAAGAACATCCGGTGGAAAA CACCCAGCTGCAGAACGAAAAACTGTATCTGTATTA TCTGCAGAACGGCCGCGATATGTATGTGGATCAGGA ACTGGATATTAACCGCCTGAGCGATTATGATGTGGA TCATATTGTGCCGCAGAGCTTTCTGAAAGATGATAG CATTGATAACAAAGTGCTGACCCGCAGCGATAAAAA CCGCGGCAAAAGCGATAACGTGCCGAGCGAAGAAG TGGTGAAAAAAATGAAAAACTATTGGCGCCAGCTGC TGAACGCGAAACTGATTACCCAGCGCAAATTTGATA ACCTGACCAAAGCGGAACGCGGCGGCCTGAGCGAA CTGGATAAAGCGGGCTTTATTAAACGCCAGCTGGTG GAAACCCGCCAGATTACCAAACATGTGGCGCAGATT CTGGATAGCCGCATGAACACCAAATATGATGAAAAC GATAAACTGATTCGCGAAGTGAAAGTGATTACCCTG AAAAGCAAACTGGTGAGCGATTTTCGCAAAGATTTT CAGTTTTATAAAGTGCGCGAAATTAACAACTATCAT CATGCGCATGATGCGTATCTGAACGCGGTGGTGGGC ACCGCGCTGATTAAAAAATATCCGAAACTGGAAAGC GAATTTGTGTATGGCGATTATAAAGTGTATGATGTG CGCAAAATGATTGCGAAAAGCGAACAGGAAATTGG CAAAGCGACCGCGAAATATTTTTTTTATAGCAACAT TATGAACTTTTTTAAAACCGAAATTACCCTGGCGAA CGGCGAAATTCGCAAACGCCCGCTGATTGAAACCAA CGGCGAAACCGGCGAAATTGTGTGGGATAAAGGCC GCGATTTTGCGACCGTGCGCAAAGTGCTGAGCATGC CGCAGGTGAACATTGTGAAAAAAACCGAAGTGCAG ACCGGCGGCTTTAGCAAAGAAAGCATTCTGCCGAAA CGCAACAGCGATAAACTGATTGCGCGCAAAAAAGA TTGGGATCCGAAAAAATATGGCGGCTTTGATAGCCC GACCGTGGCGTATAGCGTGCTGGTGGTGGCGAAAGT GGAAAAAGGCAAAAGCAAAAAACTGAAAAGCGTGA	

[0483]

	AAGAACTGCTGGGCATTACCATTATGGAACGCAGCA GCTTTGAAAAAAACCCGATTGATTTTCTGGAAGCGA AAGGCTATAAAGAAGTGAAAAAAGATCTGATTATTA AACTGCCGAAATATAGCCTGTTTGAACTGGAAAACG GCCGCAAACGCATGCTGGCGAGCGCGGGCGAACTG CAGAAAGGCAACGAACTGGCGCTGCCGAGCAAATA TGTGAACTTTCTGTATCTGGCGAGCCATTATGAAAA ACTGAAAGGCAGCCCGGAAGATAACGAACAGAAAC AGCTGTTTGTGGAACAGCATAAACATTATCTGGATG AAATTATTGAACAGATTAGCGAATTTAGCAAACGCG TGATTCTGGCGGATGCGAACCTGGATAAAGTGCTGA GCGCGTATAACAAACATCGCGATAAACCGATTCGCG AACAGGCGGAAAACATTATTCATCTGTTTACCCTGA CCAACCTGGGCGCGCCGGCGGCGTTTAAATATTTTG ATACCACCATTGATCGCAAACGCTATACCAGCACCA AAGAAGTGCTGGATGCGACCCTGATTCATCAGAGCA TTACCGGCCTGTATGAAACCCGCATTGATCTGAGCC AGCTGGGCGGCGAT	

[0484] 本文所述的多瓣引导编辑器可包括经典SpCas9，或其与上面提供的野生型Cas9序列具有至少80％、至少85％、至少90％、至少95％、或至少99％序列同一性的任何变体。这些变体可能包括含有一个或多个突变的SpCas9变体，包括SwissProt登录号Q99ZW2(SEQ ID NO:18)条目报告的任何已知突变，其包括：

[0485]

SpCas9 突变(相对于经典 SpCas9 序列的氨基酸序列，SEQ ID NO: 18)	功能/特征(如报告) (参见 UniProtKB – Q99ZW2 (CAS9_STRPT1)条目–通过引用并入本文)
D10A	切割前间隔序列链的切口酶突变体(但不裂解非前间隔序列链)
S15A	降低的 DNA 切割活性
R66A	降低的 DNA 切割活性
R70A	无 DNA 切割
R74A	降低的 DNA 切割
R78A	降低的 DNA 切割
97-150 缺失	无核酸酶活性
R165A	降低的 DNA 切割
175-307 缺失	约 50%降低的 DNA 切割
312-409 缺失	无核酸酶活性
E762A	切口酶
H840A	切割非前间隔序列链但不切割前间隔序列链的切口酶突变体
N854A	切口酶
N863A	切口酶
H982A	降低的 DNA 切割
D986A	切口酶
1099-1368 缺失	无核酸酶活性
R1333A	减少的 DNA 结合

[0486] 可用于本公开的其他野生型SpCas9序列，包括：

[0487]

描述	序列	SEQ ID NO:
SpCas9 化脓性链球菌 MGAS1882 野生型 NC_017053.1	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCA CAAATAGCGTCGGATGGGCGGTGATCACTGATGATTAT AAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATAC AGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCT CTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCG TCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAA TGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGAC TTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCAT GAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT TGCTTATCATGAGAAATATCCAACTATCTATCATCTGC GAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTT GCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGT TTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTG ATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTA CAAATCTACAATCAATTATTTGAAGAAAACCCTATTAA CGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCAC GATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCT CAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGA ATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTA AATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAG CTTTCAAAAGATACTTACGATGATGATTTAGATAATTT ATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTT TGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCA GATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTC CCCTATCAGCTTCAATGATTAAGCGCTACGATGAACAT CATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACA ACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATC AATCAAAAAACGGATATGCAGGTTATATTGATGGGGG AGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAA TTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTG AAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGA CCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTG GGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTT TTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAA AAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCAT TGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGG AAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAG AAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATT GAACGCATGACAAACTTTGATAAAAATCTTCCAAATGA AAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATT TTACGGTTTATAACGAATTGACAAAGGTCAAATATGTT ACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGA ACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAA ATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTA TTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTT CAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCC TACCATGATTTGCTAAAAATTATTAAAGATAAAGATTT TTTGGATAATGAAGAAAATGAAGATATCTTAGAGGAT	SEQ ID NO: 20

[0488]

ATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTG
ATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGG
TATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGC
AGCTGATCCATGATGATAGTTTGACATTTAAAGAAGAT
ATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTT
ACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTA
TTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGAT
GAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAA
AAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAA
TCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCT
TAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATG
AAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGAC
ATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAG
TGATTATGATGTCGATCACATTGTTCCACAAAGTTTCA
TTAAAGACGATTCAATAGACAATAAGGTACTAACGCG
TTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAA
GTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAG
ACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGT
TTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGT
GAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGT
TGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTT
TGGATAGTCGCATGAATACTAAATACGATGAAAATGA
TAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAAT
CTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
ATAAAGTACGTGAGATTAACAATTACCATCATGCCCAT
GATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGAT
TAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATG
GTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCT
AAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAAT
ATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAG
AAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCT
CTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCT
GGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGT
ATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAG
AAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTA
CCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAA
AAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGT
CCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGT
GGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAA
GAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTT
TGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGAT
ATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCT
AAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACG
GATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAAT
GAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATA
TTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAG
AAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCA
TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTG

[0489]

	AATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTA GATAAAGTTCTTAGTGCATATAACAAACATAGAGACA AACCAATACGTGAACAAGCAGAAAATATTATTCATTTA TTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAA ATATTTTGATACAACAATTGATCGTAAACGATATACGT CTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAA TCCATCACTGGTCTTTATGAAACACGCATTGATTTGAG TCAGCTAGGAGGTGACTGA	
SpCas9 化脓性链球菌 MGAS1882 野生型 NC_017053.1	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTD RHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEE NPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLK TYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG QGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKP ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEE VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 21
SpCas9 化脓性链球菌 野生型 SWBC2D7W014	ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCAC TAATTCCGTTGGATGGGCTGTCATAACCGATGAATACA AAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACAC AGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCC TCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCG CCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAA TGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTT GGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACAT	SEQ ID NO: 22

[0490]

GAACGGCACCCCATCTTTGGAAACATAGTAGATGAGG
TGGCATATCATGAAAAGTACCCAACGATTTATCACCTC
AGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACC
TGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAG
TTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCC
GGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAG
TACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATA
AATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGC
CCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCG
CACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGG
TAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATT
TTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTG
CAGCTTAGTAAGGACACGTACGATGACGATCTCGACA
ATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTA
TTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCT
ATCTGACATACTGAGAGTTAATACTGAGATTACCAAGG
CGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAA
CATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCG
TCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTG
ATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGG
CGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAA
CCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGC
TTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCA
GCGGACTTTCGACAACGGTAGCATTCCACATCAAATCC
ACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGA
GGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGA
TTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTG
GGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGAT
GACAAGAAAGTCCGAAGAAACGATTACTCCATGGAAT
TTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATC
GTTCATCGAGAGGATGACCAACTTTGACAAGAATTTAC
CGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTAC
GAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAA
GTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAA
GCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATT
CAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAA
GAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGT
CGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCAC
TTGGTACGTATCATGACCTCCTAAAGATAATTAAAGAT
AAGGACTTCCTGGATAACGAAGAGAATGAAGATATCT
TAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGAT
CGGGAAATGATTGAGGAAAGACTAAAAACATACGCTC
ACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAG
GCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAA
CTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAA
CTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAAT
AGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGA
CAAGGGGACTCATTGCACGAACATATTGCGAATCTTGC
TGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACA
GTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGAC
GTCACAAACCGGAAAACATTGTAATCGAGATGGCACG

[0491]

	CGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGT CGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAG AACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGA AAATACCCAATTGCAGAACGAGAAACTTTACCTCTATT ACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA ACTGGACATAAACCGTTTATCTGATTACGACGTCGATC ACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATC GACAATAAAGTGCTTACACGCTCGGATAAGAACCGAG GGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAA GAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCG AAACTGATAACGCAAAGAAAGTTCGATAACTTAACTA AAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGC CGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAA TCACAAAGCATGTTGCACAGATACTAGATTCCCGAATG AATACGAAATACGACGAGAACGATAAGCTGATTCGGG AAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCG GACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGA GATAAATAACTACCACCATGCGCACGACGCTTATCTTA ATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCG AAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGT TTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAG GAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTC TAACATTATGAATTTCTTTAAGACGGAAATCACTCTGG CAAACGGAGAGATACGCAAACGACCTTTAATTGAAAC CAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGC CGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCC CCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACC GGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGA ATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTG CCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGA AAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGG GGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAAC CCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGT AAAAAAGGATCTCATAATTAAACTACCAAAGTATAGT CTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGC TAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCA CTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCC CATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACG AACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTAT CTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAA GAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTAT TAAGCGCATACAACAAGCACAGGGATAAACCCATACG TGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTA CCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGAC ACAACGATAGATCGCAAACGATACACTTCTACCAAGG AGGTGCTAGACGCGACACTGATTCACCAATCCATCACG GGATTATATGAAACTCGGATAGATTTGTCACAGCTTGG GGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCG AGCGACTACAAAGACCATGACGGTGATTATAAAGATC ATGACATCGATTACAAGGATGACGATGACAAGGCTGC AGGA	

[0492]

SpCas9 化脓性链球菌 野生型 编码的 SWBC2D7W014 的产物	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSPKKKRKVSSDYKDHDGDY KDHDIDYKDDDDKAAG	SEQ ID NO: 23
SpCas9 化脓性链球菌 M1GAS 野生型 NC_002737.2	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCA CAAATAGCGTCGGATGGGCGGTGATCACTGATGAATA TAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATA CAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGC TCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTC GTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCG GAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAA ATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGC ATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAA GTTGCTTATCATGAGAAATATCCAACTATCTATCATCT GCGAAAAAAAATTGGTAGATTCTACTGATAAAGCGGAT TTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAA GTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCC TGATAATAGTGATGTGGACAAACTATTTATCCAGTTGG TACAAACCTACAATCAATTATTTGAAGAAAACCCTATT	SEQ ID NO: 24

[0493]

AACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGC
ACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTG
CTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGG
AATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTT
AAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACA
GCTTTCAAAAGATACTTACGATGATGATTTAGATAATT
TATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTT
TTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCA
GATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCC
CCTATCAGCTTCAATGATTAAACGCTACGATGAACATC
ATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAA
CAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCA
ATCAAAAAACGGATATGCAGGTTATATTGATGGGGGA
GCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAAT
TTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGA
AACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGAC
CTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGG
GTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTT
TATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAA
AATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGA
AGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGA
AGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTG
AACGCATGACAAACTTTGATAAAAATCTTCCAAATGAA
AAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTT
TACGGTTTATAACGAATTGACAAAGGTCAAATATGTTA
CTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAA
CAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAA
TCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTAT
TTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTC
AGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCT
ACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTT
TTGGATAATGAAGAAAATGAAGATATCTTAGAGGATA
TTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATG
ATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGA
TGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATA
CTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGT
ATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTT
TTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGC
AGCTGATCCATGATGATAGTTTGACATTTAAAGAAGAC
ATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTT
TACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGA
TGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAA
AATATCGTTATTGAAATGGCACGTGAAAATCAGACAA
CTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAA
ACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG
ATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCA
AAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAA
GAGACATGTATGTGGACCAAGAATTAGATATTAATCGT
TTAAGTGATTATGATGTCGATCACATTGTTCCACAAAG
TTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAA

[0494]

	CGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTT CCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATT GGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGT AAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTT GAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAAT TGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAA ATTTTGGATAGTCGCATGAATACTAAATACGATGAAAA TGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAA AATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAA TTCTATAAAGTACGTGAGATTAACAATTACCATCATGC CCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTT TGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTC TATGGTGATTATAAAGTTTATGATGTTCGTAAAATGAT TGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCA AAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAA ACAGAAATTACACTTGCAAATGGAGAGATTCGCAAAC GCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAAT TGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCA AAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAA ACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAA TTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGT AAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTG ATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCT AAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCG TTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAG TTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTA AAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAA ACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTC GTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAA AGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAAT TTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGG TAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTG GAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCA AATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATG CCAATTTAGATAAAGTTCTTAGTGCATATAACAAACAT AGAGACAAACCAATACGTGAACAAGCAGAAAATATTA TTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTG CTTTTAAATATTTTGATACAACAATTGATCGTAAACGA TATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTAT CCATCAATCCATCACTGGTCTTTATGAAACACGCATTG ATTTGAGTCAGCTAGGAGGTGACTGA	
SpCas9 化脓性链球菌 M1GAS 野生型 编码的 NC_002737.2 的产物 (与经典 Q99ZW2	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE	SEQ ID NO: 25

[0495]

野生型至少100%相同)	KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	

[0496] 本文所述的多瓣引导编辑器可包括任何上述SpCas9序列，或与其具有至少80％、至少85％、至少90％、至少95％或至少99％序列同一性的任何变体。

[0497] B.野生型Cas9直系同源物

[0498] 在其他实施方案中，Cas9蛋白可以是不同于来自化脓性链球菌的经典Cas9的来自另一细菌物种的野生型Cas9直系同源物。例如，以下Cas9直系同源物可与本说明书所述的多瓣引导编辑器构建体结合使用。此外，与任何以下直系同源物具有至少80％、至少85％、至少90％、至少95％、或至少99％序列同一性的任何变体Cas9直系同源物也可与本多瓣引导编辑器一起使用。

[0499]

描述	序列
LfCas9 发酵乳杆菌 (*Lactobacillus fermentum*) 野生型 GenBank: SNX31424.1 1	MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERR TFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPTK NQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVMNIYKLRE AMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFN VLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLE VKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSS ETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRY WTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKERGFDLEKGL KKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQ AKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQK

	描述	序列
		ATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLL NEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKT VKASDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDK VDDNRYQTDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRY KGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFKEIVDQPVFKEQIDQL NQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVG NAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEK APDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDNSL DNRVLTSRKENNKKSDQVPAKLYAAKMKPYWNQLLKQGLITQRKFEN LTKDVDQNIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTEIIETR AGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSF FVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTR DEVAKSFDRLLNMKYMLVSKEVHDRSDQLYGATIVTAKESGKLTSPIEI KKNRLVDLYGAYTNGTSAFMTIIKFTGNKPKYKVIGIPTTSAASLKRAGK PGSESYNQELHRIIKSNPKVKKGFEIVVPHVSYGQLIVDGDCKFTLASPTV QHPATQLVLSKKSLETISSGYKILKDKPAIANERLIRVFDEVVGQMNRYF TIFDQRSNRQKVADARDKFLSLPTESKYEGAKKVQVGKTEVITNLLMGL HANATQGDLKVLGLATFGFFQSTTGLSLSEDTMIVYQSPTGLFERRICLK DI (SEQ ID NO: 26)
[0500]	SaCas9 金黄色葡萄球菌 (*Staphylococcus aureus*) 野生型 GenBank: AYD60528.1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 27)
	SaCas9 金黄色葡萄	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRS KRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQ KLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEE

[0501]

描述	序列
球菌	KYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKP TLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENA ELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP VVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP FNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY ATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYK HHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDIT DDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEV NSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIE VNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSK KHPQIIKK (SEQ ID NO: 28)
StCas9 嗜热链球菌 (*Streptococcus thermophilus*) UniProtKB/ Swiss-Prot: G3ECR1.2 野生型	MLFNKCIIISINLDFSNKEKCMTKPYSIGLDIGTNSVGWAVITDNYKVPSK KMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRI LYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYH DEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKN NDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLF PGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLL GYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEH KEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLK NLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQA KFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNF EDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVR FIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIE LKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIK QRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLI DDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEVVKSLPGSPAI KKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLK RLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTG DDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDFPSLEVV KKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETR QITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYK VREINDFHHAHDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERK SATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLA TVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNEN LVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITNVLEFQGISI LDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILSTN NKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEEL FYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERKG LFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRID LAKLGEG (SEQ ID NO: 29)
LcCas9	MKIKNYNLALTPSTSAVGHVEVDDDLNILEPVHHQKAIGVAKFGEGETA

[0502]

描述	序列
卷曲乳酸杆菌 (*Lactobacillus crispatus*) NCBI 参考序列: WP_133478044.1 野生型	EARRLARSARRTTKRRANRINHYFNEIMKPEIDKVDPLMFDRIKQAGLSP LDERKEFRTVIFDRPNIASYYHNQFPTIWHLQKYLMITDEKADIRLIYWA LHSLLKHRGHFFNTTPMSQFKPGKLNLKDDMLALDDYNDLEGLSFAVA NSPEIEKVIKDRSMHKKEKIAELKKLIVNDVPDKDLAKRNNKIITQIVNAI MGNSFHLNFIFDMDLDKLTSKAWSFKLDDPELDTKFDAISGSMTDNQIGI FETLQKIYSAISLLDILNGSSNVVDAKNALYDKHKRDLNLYFKFLNTLPD EIAKTLKAGYTLYIGNRKKDLLAARKLLKVNVAKNFSQDDFYKLINKEL KSIDKQGLQTRFSEKVGELVAQNNFLPVQRSSDNVFIPYQLNAITFNKILE NQGKYYDFLVKPNPAKKDRKNAPYELSQLMQFTIPYYVGPLVTPEEQV KSGIPKTSRFAWMVRKDNGAITPWNFYDKVDIEATADKFIKRSIAKDSY LLSELVLPKHSLLYEKYEVFNELSNVSLDGKKLSGGVKQILFNEVFKKTN KVNTSRILKALAKHNIPGSKITGLSNPEEFTSSLQTYNAWKKYFPNQIDNF AYQQDLEKMIEWSTVFEDHKILAKKLDEIEWLDDDQKKFVANTRLRGW GRLSKRLLTGLKDNYGKSIMQRLETTKANFQQIVYKPEFREQIDKISQAA AKNQSLEDILANSYTSPSNRKAIRKTMSVVDEYIKLNHGKEPDKIFLMFQ RSEQEKGKQTEARSKQLNRILSQLKADKSANKLFSKQLADEFSNAIKKS KYKLNDKQYFYFQQLGRDALTGEVIDYDELYKYTVLHIIPRSKLTDDSQ NNKVLTKYKIVDGSVALKFGNSYSDALGMPIKAFWTELNRLKLIPKGKL LNLTTDFSTLNKYQRDGYIARQLVETQQIVKLLATIMQSRFKHTKIIEVR NSQVANIRYQFDYFRIKNLNEYYRGFDAYLAAVVGTYLYKVYPKARRL FVYGQYLKPKKTNQENQDMHLDSEKKSQGFNFLWNLLYGKQDQIFVN GTDVIAFNRKDLITKMNTVYNYKSQKISLAIDYHNGAMFKATLFPRNDR DTAKTRKLIPKKKDYDTDIYGGYTSNVDGYMLLAEIIKRDGNKQYGFYG VPSRLVSELDTLKKTRYTEYEEKLKEIIKPELGVDLKKIKKIKILKNKVPF NQVIIDKGSKFFITSTSYRWNYRQLILSAESQQTLMDLVVDPDFSNHKAR KDARKNADERLIKVYEEILYQVKNYMPMFVELHRCYEKLVDAQKTFKS LKISDKAMVLNQILILLHSNATSPVLEKLGYHTRFTLGKKHNLISENAVL VTQSITGLKENHVSIKQML (SEQ ID NO: 30)
PdCas9 有害片球菌 (*Pedicoccus damnosus*) NCBI 参考序列: WP_062913273.1 野生型	MTNEKYSIGLDIGTSSIGFAVVNDNNRVIRVKGKNAIGVRLFDEGKAAA DRRSFRTTRRSFRTTRRRLSRRRWRLKLLREIFDAYITPVDEAFFIRLKES NLSPKDSKKQYSGDILFNDRSDKDFYEKYPTIYHLRNALMTEHRKFDVR EIYLAIHHIMKFRGHFLNATPANNFKVGRLNLEEKFEELNDIYQRVFPDE SIEFRTDNLEQIKEVLLDNKRSRADRQRTLVSDIYQSSEDKDIEKRNKAV ATEILKASLGNKAKLNVITNVEVDKEAAKEWSITFDSESIDDDLAKIEGQ MTDDGHEIIEVLRSLYSGITLSAIVPENHTLSQSMVAKYDLHKDHLKLFK KLINGMTDTKKAKNLRAAYDGYIDGVKGKVLPQEDFYKQVQVNLDDS AEANEIQTYIDQDIFMPKQRTKANGSIPHQLQQQELDQIIENQKAYYPWL AELNPNPDKKRQQLAKYKLDELVTFRVPYYVGPMITAKDQKNQSGAEF AWMIRKEPGNITPWNFDQKVDRMATANQFIKRMTTTDTYLLGEDVLPA QSLLYQKFEVLNELNKIRIDHKPISIEQKQQIFNDLFKQFKNVTIKHLQDY LVSQGQYSKRPLIEGLADEKRFNSSLSTYSDLCGIFGAKLVEENDRQEDL EKIIEWSTIFEDKKIYRAKLNDLTWLTDDQKEKLATKRYQGWGRLSRKL LVGLKNSEHRNIMDILWITNENFMQIQAEPDFAKLVTDANKGMLEKTDS QDVINDLYTSPQNKKAIRQILLVVHDIQNAMHGQAPAKIHVEFARGEER NPRRSVQRQRQVEAAYEKVSNELVSAKVRQEFKEAINNKRDFKDRLFL YFMQGGIDIYTGKQLNIDQLSSYQIDHILPQAFVKDDSLTNRVLTNENQV KADSVPIDIFGKKMLSVWGRMKDQGLISKGKYRNLTMNPENISAHTENG FINRQLVETRQVIKLAVNILADEYGDSTQIISVKADLSHQMREDFELLKN RDVNDYHHAFDAYLAAFIGNYLLKRYPKLESYFVYGDFKKFTQKETKM RRFNFIYDLKHCDQVVNKETGEILWTKDEDIKYIRHLFAYKKILVSHEVR

[0503]

描述	序列
	EKRGALYNQTIYKAKDDKGSGQESKKLIRIKDDKETKIYGGYSGKSLAY MTIVQITKKNKVSYRVIGIPTLALARLNKLENDSTENNGELYKIIKPQFTH YKVDKKNGEIIETTDDFKIVVSKVRFQQLIDDAGQFFMLASDTYKNNAQ QLVISNNALKAINNTNITDCPRDDLERLDNLRLDSAFDEIVKKMDKYFSA YDANNFREKIRNSNLIFYQLPVEDQWENNKITELGKRTVLTRILQGLHAN ATTTDMSIFKIKTPFGQLRQRSGISLSENAQLIYQSPTGLFERRVQLNKIK (SEQ ID NO: 31)
FnCas9 具核梭杆菌 (*Fusobacterium nucleatum*) NCBI 参考序列: WP_060798984.1	MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFEE AKTAAERRVQRNSRRRLKRRKWRLNLLEEIFSNEILKIDSNFFRRLKESSL WLEDKSSKEKFTLFNDDNYKDYDFYKQYPTIFHLRNELIKNPEKKDIRLV YLAIHSIFKSRGHFLFEGQNLKEIKNFETLYNNLIAFLEDNGINKIIDKNNI EKLEKIVCDSKKGLKDKEKEFKEIFNSDKQLVAIFKLSVGSSVSLNDLFD TDEYKKGEVEKEKISFREQIYEDDKPIYYSILGEKIELLDIAKTFYDFMVL NNILADSQYISEAKVKLYEEHKKDLKNLKYIIRKYNKGNYDKLFKDKNE NNYSAYIGLNKEKSKKEVIEKSRLKIDDLIKNIKGYLPKVEEIEEKDKAIF NKILNKIELKTILPKQRISDNGTLPYQIHEAELEKILENQSKYYDFLNYEE NGIITKDKLLMTFKFRIPYYVGPLNSYHKDKGGNSWIVRKEEGKILPWNF EQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDTFLYSEYVILNELNKVQ VNDEFLNEENKRKIIDELFKENKKVSEKKFKEYLLVKQIVDGTIELKGVK DSFNSNYISYIRFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKIFEKKIK NEYGDILTKDEIKKINTFKFNNWGRLSEKLLTGIEFINLETGECYSSVMDA LRRTNYNLMELLSSKFTLQESINNENKEMNEASYRDLIEESYVSPSLKRAI FQTLKIYEEIRKITGRVPKKVFIEMARGGDESMKNKKIPARQEQLKKLYD SCGNDIANFSIDIKEMKNSLISYDNNSLRQKKLYLYYLQFGKCMYTGREI DLDRLLQNNDTYDIDHIYPRSKVIKDDSFDNLVLVLKNENAEKSNEYPV KKEIQEKMKSFWRFLKEKNFISDEKYKRLTGKDDFELRGFMARQLVNV RQTTKEVGKILQQIEPEIKIVYSKAEIASSFREMFDFIKVRELNDTHHAKD AYLNIVAGNVYNTKFTEKPYRYLQEIKENYDVKKIYNYDIKNAWDKEN SLEIVKKNMEKNTVNITRFIKEKKGQLFDLNPIKKGETSNEIISIKPKVYN GKDDKLNEKYGYYKSLNPAYFLYVEHKEKNKRIKSFERVNLVDVNNIK DEKSLVKYLIENKKLVEPRVIKKVYKRQVILINDYPYSIVTLDSNKLMDF ENLKPLFLENKYEKILKNVIKFLEDNQGKSEENYKFIYLKKKDRYEKNET LESVKDRYNLEFNEMYDKFLEKLDSKDYKNYMNNKKYQELLDVKEKFI KLNLFDKAFTLKSFLDLFNRKTMADFSKVGLTKYLGKIQKISSNVLSKNE LYLLEESVTGLFVKKIKL (SEQ ID NO: 32)
EcCas9 盲肠肠球菌 (*Enterococcus cecorum*) NCBI 参考序列: WP_047338501.1 野生型	RRKQRIQILQELLGEEVLKTDPGFFHRMKESRYVVEDKRTLDGKQVELP YALFVDKDYTDKEYYKQFPTINHLIVYLMTTSDTPDIRLVYLALHYYMK NRGNFLHSGDINNVKDINDILEQLDNVLETFLDGWNLKLKSYVEDIKNIY NRDLGRGERKKAFVNTLGAKTKAEKAFCSLISGGSTNLAELFDDSSLKEI ETPKIEFASSSLEDKIDGIQEALEDRFAVIEAAKRLYDWKTLTDILGDSSS LAEARVNSYQMHHEQLLELKSLVKEYLDRKVFQEVFVSLNVANNYPAY IGHTKINGKKKELEVKRTKRNDFYSYVKKQVIEPIKKKVSDEAVLTKLSE IESLIEVDKYLPLQVNSDNGVIPYQVKLNELTRIFDNLENRIPVLRENRDK IIKTFKFRIPYYVGSLNGVVKNGKCTNWMVRKEEGKIYPWNFEDKVDLE ASAEQFIRRMTNKCTYLVNEDVLPKYSLLYSKYLVLSELNNLRIDGRPLD VKIKQDIYENVFKKNRKVTLKKIKKYLLKEGIITDDDELSGLADDVKSSL TAYRDFKEKLGHLDLSEAQMENIILNITLFGDDKKLLKKRLAALYPFIDD KSLNRIATLNYRDWGRLSERFLSGITSVDQETGELRTIIQCMYETQANLM QLLAEPYHFVEAIEKENPKVDLESISYRIVNDLYVSPAVKRQIWQTLLVIK DIKQVMKHDPERIFIEMAREKQESKKTKSRKQVLSEVYKKAKEYEHLFE

[0504]

描述	序列
	KLNSLTEEQLRSKKIYLYFTQLGKCMYSGEPIDFENLVSANSNYDIDHIYP QSKTIDDSFNNIVLVKKSLNAYKSNHYPIDKNIRDNEKVKTLWNTLVSK GLITKEKYERLIRSTPFSDEELAGFIARQLVETRQSTKAVAEILSNWFPESE IVYSKAKNVSNFRQDFEILKVRELNDCHHAHDAYLNIVVGNAYHTKFTN SPYRFIKNKANQEYNLRKLLQKVNKIESNGVVAWVGQSENNPGTIATVK KVIRRNTVLISRMVKEVDGQLFDLTLMKKGKGQVPIKSSDERLTDISKY GGYNKATGAYFTFVKSKKRGKVVRSFEYVPLHLSKQFENNNELLKEYIE KDRGLTDVEILIPKVLINSLFRYNGSLVRITGRGDTRLLLVHEQPLYVSNS FVQQLKSVSSYKLKKSENDNAKLTKTATEKLSNIDELYDGLLRKLDLPIY SYWFSSIKEYLVESRTKYIKLSIEEKALVIFEILHLFQSDAQVPNLKILGLS TKPSRIRIQKNLKDTDKMSIIHQSPSGIFEHEIELTSL (SEQ ID NO: 33)
AhCas9 *Anaerostipes hadrus* NCBI 参考序列: WP_044924278.1 野生型	MQNGFLGITVSSEQVGWAVTNPKYELERASRKDLWGVRLFDKAETAED RRMFRTNRRLNQRKKNRIHYLRDIFHEEVNQKDPNFFQQLDESNFCEDD RTVEFNFDTNLYKNQFPTVYHLRKYLMETKDKPDIRLVYLAFSKFMKN RGHFLYKGNLGEVMDFENSMKGFCESLEKFNIDFPTLSDEQVKEVRDIL CDHKIAKTVKKKNIITITKVKSKTAKAWIGLFCGCSVPVKVLFQDIDEEIV TDPEKISFEDASYDDYIANIEKGVGIYYEAIVSAKMLFDWSILNEILGDHQ LLSDAMIAEYNKHHDDLKRLQKIIKGTGSRELYQDIFINDVSGNYVCYV GHAKTMSSADQKQFYTFLKNRLKNVNGISSEDAEWIDTEIKNGTLLPKQ TKRDNSVIPHQLQLREFELILDNMQEMYPFLKENREKLLKIFNFVIPYYV GPLKGVVRKGESTNWMVPKKDGVIHPWNFDEMVDKEASAECFISRMT GNCSYLFNEKVLPKNSLLYETFEVLNELNPLKINGEPISVELKQRIYEQLF LTGKKVTKKSLTKYLIKNGYDKDIELSGIDNEFHSNLKSHIDFEDYDNLS DEEVEQIILRITVFEDKQLLKDYLNREFVKLSEDERKQICSLSYKGWGNL SEMLLNGITVTDSNGVEVSVMDMLWNTNLNLMQILSKKYGYKAEIEHY NKEHEKTIYNREDLMDYLNIPPAQRRKVNQLITIVKSLKKTYGVPNKIFF KISREHQDDPKRTSSRKEQLKYLYKSLKSEDEKHLMKELDELNDHELSN DKVYLYFLQKGRCIYSGKKLNLSRLRKSNYQNDIDYIYPLSAVNDRSMN NKVLTGIQENRADKYTYFPVDSEIQKKMKGFWMELVLQGFMTKEKYFR LSRENDFSKSELVSFIEREISDNQQSGRMIASVLQYYFPESKIVFVKEKLIS SFKRDFHLISSYGHNHLQAAKDAYITIVVGNVYHTKFTMDPAIYFKNHK RKDYDLNRLFLENISRDGQIAWESGPYGSIQTVRKEYAQNHIAVTKRVV EVKGGLFKQMPLKKGHGEYPLKTNDPRFGNIAQYGGYTNVTGSYFVLV ESMEKGKKRISLEYVPVYLHERLEDDPGHKLLKEYLVDHRKLNHPKILL AKVRKNSLLKIDGFYYRLNGRSGNALILTNAVELIMDDWQTKTANKISG YMKRRAIDKKARVYQNEFHIQELEQLYDFYLDKLKNGVYKNRKNNQA ELIHNEKEQFMELKTEDQCVLLTEIKKLFVCSPMQADLTLIGGSKHTGMI AMSSNVTKADFAVIAEDPLGLRNKVIYSHKGEK (SEQ ID NO: 34)
KvCas9 *Kandleria vitulina* NCBI 参考序列: WP_031589969.1 野生型	MSQNNNKIYNIGLDIGDASVGWAVVDEHYNLLKRHGKHMWGSRLFTQ ANTAVERRSSRSTRRRYNKRRERIRLLREIMEDMVLDVDPTFFIRLANVS FLDQEDKKDYLKENYHSNYNLFIDKDFNDKTYYDKYPTIYHLRKHLCES KEKEDPRLIYLALHHIVKYRGNFLYEGQKFSMDVSNIEDKMIDVLRQFN EINLFEYVEDRKKIDEVLNVLKEPLSKKHKAEKAFALFDTTKDNKAAYK ELCAALAGNKFNVTKMLKEAELHDEDEKDISFKFSDATFDDAFVEKQPL LGDCVEFIDLLHDIYSWVELQNILGSAHTSEPSISAAMIQRYEDHKNDLK LLKDVIRKYLPKKYFEVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKK LIEKIDDPDVKTILNKIELESFMLKQNSRTNGAVPYQMQLDELNKILENQ SVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKDRQFDWIIKKEGKENERIL PWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVMAKNSLTVSKYEVLN EINKLRINDHLIKRDMKDKMLHTLFMDHKSISANAMKKWLVKNQYFSN

[0505]

描述	序列
	TDDIKIEGFQKENACSTSLTPWIDFTKIFGKINESNYDFIEKIIYDVTVFED KKILRRRLKKEYDLDEEKIKKILKLKYSGWSRLSKKLLSGIKTKYKDSTR TPETVLEVMERTNMNLMQVINDEKLGFKKTIDDANSTSVSGKFSYAEVQ ELAGSPAIKRGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDEKERKDSF VNQMLKLYKDYDFEDETEKEANKHLKGEDAKSKIRSERLKLYYTQMG KCMYTGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLD DLVIPSSIRNKMYGFWEKLFNNKIISPKKFYSLIKTEFNEKDQERFINRQIV ETRQITKHVAQIIDNHYENTKVVTVRADLSHQFRERYHIYKNRDINDFHH AHDAYIATILGTYIGHRFESLDAKYIYGEYKRIFRNQKNKGKEMKKNND GFILNSMRNIYADKDTGEIVWDPNYIDRIKKCFYYKDCFVTKKLEENNG TFFNVTVLPNDTNSDKDNTLATVPVNKYRSNVNKYGGFSGVNSFIVAIK GKKKKGKKVIEVNKLTGIPLMYKNADEEIKINYLKQAEDLEEVQIGKEIL KNQLIEKDGGLYYIVAPTEIINAKQLILNESQTKLVCEIYKAMKYKNYDN LDSEKIIDLYRLLINKMELYYPEYRKQLVKKFEDRYEQLKVISIEEKCNII KQILATLHCNSSIGKIMYSDFKISTTIGRLNGRTISLDDISFIAESPTGMYSK KYKL (SEQ ID NO: 35)
EfCas9 粪肠球菌 (*Enterococcus faecalis)* NCBI 参考序列: WP_016631044.1 野生型	MRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFF ARLQESFLVPEDKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSE QADLRLIYLALAHIVKYRGHFLIEGKLSTENTSVKDQFQQFMVIYNQTFV NGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQF LKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVF LAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIR ENCPDEYDNLFKNEQKDGYAGYIAHAGKVSQLKFYQYVKKIIQDIAGAE YFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQE KIEQLVTFRIPYYVGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQS ATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKA NFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFN ASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFK GQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGV SKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSETVNELAGSPAIKK GIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEK AMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSLHRLS HYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAY WEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNV AGILDQRYNAKSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQD AYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLL RFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFS KESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIK QEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRL LASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEF QEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFN AMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSPTGLYETRRKVVD (SEQ ID NO: 36)
金黄色葡萄球菌 (*Staphylococcus aureus)* Cas9	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKL SEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFI DTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL

[0506]

描述	序列
	KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAEL LDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVK RSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNY EVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETF KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATR GLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHA EDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYK EIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDE KNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKC YEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMI DITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQ IIKKG (SEQ ID NO: 37)
嗜热脱氮土芽孢杆菌 *(Geobacillus thermodenitrificans)* Cas9	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRL ARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRV EALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQ SILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQ REYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPK ATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHD VRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVY GKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLA DKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGY TFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHI ELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIV KFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKV LVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLR LHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVN GRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRR EQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEK LESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQ LDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGE LGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTI DMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIK TAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKY QVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ ID NO: 38)
ScCas9 *S. canis* 1375 AA 159.2 kDa	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLM GALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSF FQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPE KADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTP NFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDA ILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAE IFKDDTKNGYAGYVGIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAEEL LAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKI EKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQS

	描述	序列
[0507]		FIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGF SNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIK ELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDS RMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFAT VRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKY GGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGF LEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQ HLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKV NSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRL RYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD (SEQ ID NO: 39)

[0508] 本文所述的多瓣引导编辑器可包括任何上述Cas9直系同源序列,或与其具有至少80%、至少85%、至少90%、至少95%或至少99%序列同一性的任何变体。

[0509] napDNAbp可包括任何合适的同源物和/或直系同源物或天然存在的酶,例如,Cas9、Cas9同源物和/或直系同源物已在不同物种中得到描述,所述物种包括但不限于化脓性链球菌和嗜热链球菌。优选地,Cas部分被配置(如,诱变、重组工程化或以其他方式从自然界获得)为切口酶,即,能够仅切割靶标双链DNA。其他合适的Cas9核酸酶的单链,且序列基于本公开对技术人员而言是显而易见的,此类Cas9核酸酶和序列包括来自Chylinski,Rhun,and Charpentier,"The tracrRNAand Cas9 families oftype II CRISPR-Cas immunity systems"(2013)RNABiology 10:5,726-737(其全部内容通过引用并入本文)中公开的生物体和基因座的Cas9序列。在一些实施方案中,Cas9核酸酶具有无活性(如,失活)的DNA切割结构域;即Cas9是切口酶。在一些实施方案中,Cas9蛋白包含与由表3的任何一种变体提供的Cas9蛋白的氨基酸序列至少80%相同的氨基酸序列。在一些实施方案中,Cas9蛋白包含与由上表中Cas9直系同源物的任何一种提供的Cas9蛋白的氨基酸序列至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同的氨基酸序列。

[0510] C.死亡Cas9变体

[0511] 在某些实施方案中,本文所述的多瓣引导编辑器可包括死亡Cas9,例如死亡SpCas9,其由于一个或多个使Cas9的两个核酸酶结构域失活的突变而没有核酸酶活性,即RuvC结构域(其切割非前间隔序列DNA链)和HNH结构域(切割前间隔序列DNA链)。核酸酶失活可能是由于一个或多个突变,其导致编码的蛋白或与其具有至少80%、至少85%、至少90%、至少95%或至少99%的序列同一性的任何变体的氨基酸序列中的一个或多个取代和/或缺失。

[0512] 如本文所用,术语"dCas9"是指无核酸酶活性Cas9或核酸酶死亡Cas9,或其功能片段,并且包括来自任何生物体的任何天然存在的dCas9、任何天然存在的dCas9等同物或其功能片段、来自任何生物体的任何dCas9同系物、直系同源物或旁系同源物,以及天然存在

或工程化的dCas9的任何突变体或变体。术语dCas9并不意味着特别限制并且可被称为“dCas9或等同物”。示例性dCas9蛋白和用于制备dCas9蛋白的方法在本文中进一步描述和/或在本领域中描述并且通过引用并入本文。

[0513] 在其他实施方案中，dCas9对应于或包含部分或整个的具有一个或多个使Cas9核酸酶活性失活的突变的Cas9氨基酸序列。在其他实施方案中，提供了具有除D10A和H840A之外的突变的Cas9变体，所述突变可导致内源性Cas9核酸酶(如，分别为nCas9或dCas9)活性的完全或部分失活。例如，参考野生型序列如来自化脓性链球菌的Cas9(NCBI参考序列:NC_017053.1)，此类突变包括在Cas9的D10和H840处的其他氨基酸取代，或核酸酶结构域内的其他取代(如，HNH核酸酶亚结构域和/或RuvC1亚结构域中的取代)。在一些实施方案中，提供了Cas9的变体或同源物(如，来自化脓性链球菌(NCBI参考序列:NC_017053.1(SEQ ID NO:20))的Cas9的变体)，其与NCBI参考序列:NC_017053.1至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同。在一些实施方案中，提供了Cas9的变体(如，NCBI参考序列:NC_017053.1(SEQ ID NO:20)的变体)，其具有比NC_017053.1(SEQ ID NO:20)更短或更长约5个氨基酸，约10个氨基酸，约15个氨基酸，约20个氨基酸，约25个氨基酸，约30个氨基酸，约40个氨基酸，约50个氨基酸，约75个氨基酸，约100个氨基酸或以上的氨基酸序列。

[0514] 在一个实施方案中，死亡Cas9可基于Q99ZW2的经典SpCas9序列，并且可具有以下序列，其包含D10X和H810X，其中X可以是任何氨基酸、取代(下划线和粗体)、或变体为与SEQ ID NO:40具有至少80％、至少85％、至少90％、至少95％、或至少99％序列同一性的变体。

[0515] 在一个实施方案中，死亡Cas9可基于Q99ZW2的经典SpCas9序列并且可具有以下序列，其包含D10A和H810A取代(下划线和粗体)，或是与SEQ ID NO:41具有至少80％、至少85％、至少90％、至少95％、或至少99％序列同一性的变体。

[0516]

描述	序列	SEQ ID NO:
死亡Cas9或dCas9 化脓性链球菌 Q99ZW2 Cas9具有D10**X**和H810**X** 其中“X”为任何氨基酸	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 40
死亡Cas9或dCas9	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI	SEQ ID NO: 41

[0517]

描述	序列	SEQ ID NO:
化脓性链球菌 Q99ZW2 Cas9 具有 D10**A** 和 H810**A**	CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	

[0518] D.Cas9切口酶变体

[0519] 在一个实施方案中,本文所述的多瓣引导编辑器包含Cas9切口酶。术语“nCas9”或“Cas9切口酶”是指能够在双链DNA分子靶中引入单链断裂的Cas9变体。在一些实施方案中,Cas9切口酶仅包含单个功能性核酸酶结构域。野生型Cas9(如,经典SpCas9)包含两个独立的核酸酶结构域,即,RuvC结构域(切割非前间隔序列DNA链)和HNH结构域(切割前间隔序列DNA链)。在一个实施方案中,Cas9切口酶包含在RuvC结构域中使RuvC核酸酶活性失活的突变。例如,天冬氨酸(D)10、组氨酸(H)983、天冬氨酸(D)986或谷氨酸(E)762中的突变已被报道为RuvC核酸酶结构域的功能丧失突变和功能性Cas9切口酶的创建(如,Nishimasu et al.,“Crystal structure ofCas9 in complex with guide RNA andtarget DNA,”Cell 156(5),935-949,通过引用并入本文)。因此,RuvC结构域中的切口酶突变可包括D10X、H983X、D986X、或E762X,其中X是除野生型氨基酸之外的任何氨基酸。在某些实施方案中,切口酶可以是D10A、H983A、D986A、或E762A、或其组合。

[0520] 在不同实施方案中,Cas9切口酶可在RuvC核酸酶结构域中具有突变并且具有以下氨基酸序列中的一种或与其具有至少80%、至少85%、至少90%、至少95%、或至少99%序列同一性的氨基酸序列的变体。

[0521]

描述	序列	SEQ ID NO:
Cas9切口酶 化脓性链球菌Q99ZW2Cas9具有D10X，其中X为任何替代氨基酸	MDKKYSIGLXIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 42
Cas9切口酶 化脓性链球	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL	SEQ ID NO: 43

[0522]

描述	序列	SEQ ID NO:
菌 Q99ZW2 Cas9具有 E762X，其中 X为任何替代氨基酸	AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIXMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 H983X，其中 X为任何替代氨基酸	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK	SEQ ID NO: 44

	描述	序列	SEQ ID NO:
		PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYH**X**AHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	
[0523]	Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有D986X，其中X为任何替代氨基酸	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH**X**AYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH	SEQ ID NO: 45

[0524]

描述	序列	SEQ ID NO:
	QSITGLYETRIDLSQLGGD	
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 D10**A**	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 46
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 E762A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR	SEQ ID NO: 47

	描述	序列	SEQ ID NO:
		KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVI**A**MARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	
[0525]	Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 H983A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYH**A**AHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK	SEQ ID NO: 48

描述	序列	SEQ ID NO:
	VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 D986A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH**A**AYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 49

[0526]

[0527] 在另一个实施方案中,Cas9切口酶包含在HNH结构域中使HNH核酸酶活性失活的突变。例如,组氨酸(H)840或天冬酰胺(R)863中的突变已被报道为HNH核酸酶结构域的功能丧失突变和功能性Cas9切口酶的创建(如,Nishimasu et al.,“Crystal structure of Cas9 in complex with guide RNA and targetDNA,”Cell 156(5),935-949,其通过引用并入本文)。因此,HNH结构域中的切口酶突变可包括H840X和R863X,其中X是除野生型氨基酸之外的任何氨基酸。在某些实施方案中,切口酶可以是H840A或R863A或其组合。

[0528] 在不同实施方案中,Cas9切口酶可在HNH核酸酶结构域中具有突变并且具有以下氨基酸序列中的一种或与其具有至少80%、至少85%、至少90%、至少95%、或至少99%序列同一性的氨基酸序列的变体。

[0529]

描述	序列	SEQ ID NO:
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 H840**X**，其中X为任何替代氨基酸	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 50
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 H840**A**	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN	SEQ ID NO: 51

[0530]

描述	序列	SEQ ID NO:
	LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVD**A**IVPQSFLKDDSIDNKVLTRSDKN**R**GKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有R863X，其中X为任何替代氨基酸	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKN**X**GKSDNVPSE	SEQ ID NO: 52

描述	序列	SEQ ID NO:
	EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	
Cas9切口酶 化脓性链球菌 Q99ZW2 Cas9具有 R863**A**	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKN**A**GKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGD	SEQ ID NO: 53

[0531]

[0532] 在一些实施方案中,N-端甲硫氨酸从Cas9切口酶或从本文公开或考虑的任何Cas9变体、直系同源物或等同物去除。例如,减甲硫氨酸的Cas9切口酶包括以下序列或与其具有至少80%、至少85%、至少90%、至少95%或至少99%序列同一性的氨基酸序列的变体。

[0533]

描述	序列
Cas9 切口酶 (减 Met) 化脓性链球菌 Q99ZW2 Cas9 具有 H840**X**, 其中 X 为任何替代氨基酸	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD (SEQ ID NO: 54)
Cas9 切口酶 (减 Met) 化脓性链球菌 Q99ZW2 Cas9 具有 H840**A**	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK

[0534]

	RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKN**R**GKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD (SEQ ID NO: 55)
Cas9 切口酶 (减 Met) 化脓性链球菌 Q99ZW2 Cas9 具有 R863X, 其中 X 为任何替代氨基酸	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**X**GKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD(SEQ ID NO: 56)
Cas9 切口酶 (减 Met) 化脓性链球菌 Q99ZW2 Cas9 具有 R863**A**	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF

[0535]	IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNAGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD (SEQ ID NO: 57)

[0536] E.其他Cas9变体

[0537] 除了死亡Cas9和Cas9切口酶变体外，本文使用的Cas9蛋白还可包括具有与任何参考Cas9蛋白(包括任何野生型Cas9或突变体Cas9(如，失活Cas9或Cas9切口酶)，或Cas9片段，或环状排列的Cas9，或本文公开或本领域已知的Cas9的其他变体)至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同的其他“Cas9变体”。在一些实施方案中，Cas9变体与参考Cas9相比可具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个氨基酸变化。在一些实施方案中，Cas9变体包括参考Cas9的片段(如，gRNA结合结构域或DNA切割结构域)，使得该片段与野生型Cas9的相应片段至少约70％相同，至少约80％相同，至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同。在一些实施方案中，该片段为相应野生型Cas9(如，SEQ ID NO:18)的氨基酸长度的至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％相同、至少96％、至少97％、至少98％、至少99％，或至少99.5％。

[0538] 在一些实施方案中，本公开还可利用Cas9片段，其保留功能性并且为本文公开的任何Cas9蛋白的片段。在一些实施方案中，Cas9片段的长度为至少100个氨基酸。在一些实施方案中，该片段的长度为至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或至少1300个氨基酸。

[0539] 在不同实施方案中，本文公开的多瓣引导编辑器可包含如下所述的Cas9变体中的一种或其与任何参考Cas9变体具有至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同的Cas9变体。

[0540] F.小型Cas9变体

[0541] 在一些实施方案中，本文考虑的多瓣引导编辑器可包括分子量小于经典SpCas9序列的Cas9蛋白。在一些实施方案中，小型Cas9变体可促进递送至细胞，例如通过表达载体、纳米颗粒或其他递送方式。在某些实施方案中，小型Cas9变体可包括归类为2类CRISPR-Cas系统的II型酶的酶。在一些实施方案中，小型Cas9变体可包括归类为2类CRISPR-Cas系统的V型酶的酶。在其他实施方案中，小型Cas9变体可包括归类为2类CRISPR-Cas系统的VI型酶的酶。

[0542] 经典SpCas9蛋白的长度为1368个氨基酸并且具有158千道尔顿的预测分子量。如本文所用，术语“小型Cas9变体”是指任何Cas9变体——天然存在的、工程化的或以其他方式——少于至少1300个氨基酸，或至少少于1290个氨基酸，或少于1280个氨基酸，或少于1270个氨基酸，或少于1260个氨基酸，或少于1250个氨基酸，或少于1240个氨基酸，或少于1230个氨基酸，或少于1220个氨基酸，或少于1210个氨基酸，或少于1200个氨基酸，或少于1190个氨基酸，或少于1180个氨基酸，或少于1170个氨基酸，或少于1160个氨基酸，或少于1150个氨基酸，或少于1140个氨基酸，或少于1130个氨基酸，或少于1120个氨基酸，或少于1110个氨基酸，或少于1100个氨基酸，或少于1050个氨基酸，或少于1000个氨基酸，或少于950个氨基酸，或少于900个氨基酸，或少于850个氨基酸，或少于800个氨基酸，或少于75个0个氨基酸，或少于700个氨基酸，或少于650个氨基酸，或少于600个氨基酸，或少于550个氨基酸，或少于500个氨基酸，但至少多于约400个氨基酸并保留所需的Cas9蛋白功能。Cas9变体可包括归类为2类CRISPR-Cas系统的II型、V型或VI型酶的那些。

[0543] 在不同实施方案中，本文公开的多瓣引导编辑器可包含如下所述的小型Cas9变体或其与任何参考小型Cas9蛋白具有至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同的Cas9变体。

[0544]

描述	序列	SEQ ID NO:
SaCas9 金黄色葡萄球菌 1053 AA 123 kDa	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEA NVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLT DHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGV HNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLER LKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLM GHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENE KLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYR VTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILT IYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLK AINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTT LVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREK NSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIE KIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSV SFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYET FKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTS FLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIK HIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTL IVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQ KLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVI KKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLK KISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVN MIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILG NLYEVKSKKHPQIIKK	SEQ ID NO: 58
NmeCas9 脑膜炎奈瑟菌 1083 AA 124.5 kDa	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDL GVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLL RTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALD RKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGAL LKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRS DYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIET LLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAE RFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKL TYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAY HAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDED	SEQ ID NO: 59

	描述	序列	SEQ ID NO:
		ITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLME QGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPV VLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDR KEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKL RLYEQQHGKCLYSGKEINLGRLNEKGYVEIDAALPFSRT WDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREW QEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRY VNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFW GLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYK EMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRV FGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTP LFVSRAPNRKMSGQGHMETVKSAKRLDEGVSVLRVPLT QLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAF AEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNG IADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRA VVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMF GYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSF QKYQIDELGKEIRPCRLKKRPPVR	
[0545]	CjCas9 空肠弯曲菌(*C. jejuni)* 984 AA 114.9 kDa	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTG ESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNY EDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFAR VILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQ SVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFL KDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFS HLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKN TEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDY EFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDIT LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKA LKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFN ETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINI ELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLG LKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEI DHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGN DSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKD RNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQK GSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDA VIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKN KRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETF RKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMF RVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKG EIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYN AFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKS IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK	SEQ ID NO: 60
	GeoCas9 嗜热脂肪土芽孢杆菌(*G. stearothermophilus*)	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENP QTGESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILT KEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLH LAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRTV GEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSK QREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFC TFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTD	SEQ ID NO: 61

	描述	序列	SEQ ID NO:
	1087 AA 127 kDa	EERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDR GESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPI DFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLAN KVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVY SSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRALTQ ARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQ DENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQNG RCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNKVL VLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFS KKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIR EHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREES DLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKELAKKT EPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQK LESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQ TVVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRLLEH NNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPL NDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGI LPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPRE KTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRFS LRGVGSRTLKRFEKYQVDVLGNIYKVRGEKRVGLASSAH SKPGKTIRPLQSTRD	
[0546]	LbaCas12a *L. bacterium* 1228 AA 143.9 kDa	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVED EKRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISL FRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFK KDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENM FSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEV QEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGG FVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSD RESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKL FKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNA EYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEY ADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLE KSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNR DESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKF KLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMD KKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVF FSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLI DFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQ GYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHG TPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKE ELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQ YELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGE RNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLL DKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVE KYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLN YMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIF YIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIM YVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIF RNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRAL LCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPV	SEQ ID NO: 62

[0547]

描述	序列	SEQ ID NO:
	KNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLW AIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH	
BhCas12b *B. hisashii* 1108 AA 130.4 kDa	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILK LIRQEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQK CNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEE EKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIV KEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLK VKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQ LLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEK YLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPE YPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFE ERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDES IKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFN MTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKD SKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQ KPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVL RKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTK WISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQL HKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRT RKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKED RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFED LSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQ VGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKN LQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVT THADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTV YIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKK GSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGN VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQ SM	SEQ ID NO: 63

[0548] G.Cas9等同物

[0549] 在一些实施方案中，本文所述的多瓣引导编辑器可包括任何Cas9等同物。如本文所用，术语“Cas9等同物”是广义术语，包括在本多瓣引导编辑器中起到与Cas9相同的功能的任何napDNAbp蛋白，尽管其氨基酸一级序列和/或其三维结构可能不同和/或从进化的角度来看无关。因此，虽然Cas9等同物包括本文所述或包含的进化相关的任何Cas9直系同源物、同源物、突变体或变体，但Cas9等同物还包括可能已通过趋同进化过程进化为具有与Cas9相同或相似功能的蛋白，但是它们不一定在氨基酸序列和/或三维结构方面具有任何相似性。此处描述的多瓣引导编辑器包括将提供与Cas9相同或相似功能的任何Cas9等同物，尽管Cas9等同物可能基于通过趋同进化产生的蛋白。例如，如果Cas9是指CRISPR-Cas系统的II型酶，那么Cas9等同物可指CRISPR-Cas系统的V型或VI型酶。

[0550] 例如，Cas12e(CasX)是据报道具有与Cas9相同功能但通过趋同进化而进化的Cas9等同物。因此，考虑描述于Liu et al.,“CasX enzymes comprises a distinct family of RNA-guided genome editors,”Nature,2019,Vol.566:218-223中的Cas12e(CasX)蛋白与本文所述的多瓣引导编辑器一起使用。此外，Cas12e(CasX)的任何变体或修饰都是可想到的并且在本公开的范围内。

[0551] Cas9是在极其多种物种中进化的细菌酶。然而，本文考虑的Cas9等同物也可从古

细菌获得,所述古细菌构成不同于细菌的单细胞原核微生物的域和界。

[0552] 在一些实施方案中,Cas9等同物可指Cas12e(CasX)或Cas12d(CasY),其已描述于例如Burstein et al.,"New CRISPR-Cas systems from uncultivated microbes." CellRes.2017Feb 21.doi:10.1038/cr.2017.21,其全部内容在此通过引用并入。使用基因组解析的宏基因组学,确定了许多CRISPR-Cas系统,包括首个在古细菌生命领域报道的Cas9。这种不同的Cas9蛋白是在研究很少的纳米古细菌(nanoarchaea)中作为活性CRISPR-Cas系统的一部分发现的。在细菌中,发现了两个以前未知的系统,CRISPR-Cas12e和CRISPR-Cas12d,它们是迄今为止发现的最紧凑的系统之一。在一些实施方案中,Cas9是指Cas12e,或Cas12e的变体。在一些实施方案中,Cas9是指Cas12d或Cas12d的变体。应当理解,其他RNA引导的DNA结合蛋白可用作核酸可编程DNA结合蛋白(napDNAbp),并且在本公开的范围内。另请参见Liu et al.,"CasX enzymes comprises a distinct family of RNA-guided genome editors,"Nature,2019,Vol.566:218-223。任何这些Cas9等同物都考虑在内。

[0553] 在一些实施方案中,Cas9等同物包含与天然存在的Cas12e(CasX)或Cas12d(CasY)蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的Cas12e(CasX)或Cas12d(CasY)蛋白。在一些实施方案中,napDNAbp包含与野生型Cas部分或本文提供的任何Cas部分至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同的氨基酸序列。

[0554] 在不同实施方案中,核酸可编程DNA结合蛋白包括但不限于Cas9(如,dCas9和nCas9)、Cas12e(CasX)、Cas12d(CasY)、Cas12a(Cpf1)、Cas12b1(C2c1)、Cas13a(C2c2)、Cas12c(C2c3)、Argonaute蛋白和Cas12b1。具有与Cas9不同的PAM特异性的核酸可编程DNA结合蛋白的一个示例是来自普雷沃菌属(Prevotella)和弗朗西斯氏菌属(Francisella)1(即,Cas12a(Cpf1))的成簇规律间隔短回文重复序列。与Cas9类似,Cas12a(Cpf1)也是2类CRISPR效应物,但它为V型亚群而不是II型亚群的酶。已表明Cas12a(Cpf1)介导强大的DNA干扰,其特征与Cas9不同。Cas12a(Cpf1)是单链RNA引导的核酸内切酶,缺乏tracrRNA,它利用富含T的前间隔序列邻近基序(TTN、TTTN或YTN)。此外,Cpf1通过交错的DNA双链断裂来切割DNA。在16种Cpf1家族蛋白中,来自氨基酸球菌属(Acidaminococcus)和毛螺菌科(Lachnospiraceae)的2种酶证明在人类细胞中具有有效的基因组编辑活性。Cpf1蛋白是本领域已知的,并且之前已描述于例如Yamano etal.,"Crystal structure ofCpf1 in complexwith guide RNAandtargetDNA."Cell(165)2016,p.949-962;其全部内容在此通过引用并入。

[0555] 在其他实施方案中,Cas蛋白可包括任何CRISPR相关蛋白,包括但不限于Cas12a、Cas12b1、Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(又被称为Csn1和Csx12)、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr、Csb、Csb、Csm、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、其同源物或其修饰版本,并且优选包含切口酶突变(如,对应于SEQ ID NO:18的野生型Cas9多肽的D10A突变的突变)。

[0556] 在不同的其他实施方案中，napDNAbp可以是以下蛋白中的任一种：Cas9、Cas12a(Cpf1)、Cas12e(CasX)、Cas12d(CasY)、Cas12b1(C2c1)、Cas13a(C2c2)、Cas12c(C2c3)、GeoCas9、CjCas9、Cas12g、Cas12h、Cas12i、Cas13b、Cas13c、Cas13d、Cas14、Csn2、xCas9、SpCas9-NG、环状排列的Cas9，或Argonaute蛋白(Ago)结构域，或其变体。

[0557] 示例性的Cas9等同蛋白序列可包括以下：

[0558]

描述	序列
AsCas12a (以前又被称为 Cpf1)	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKEL KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQA TYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVT TTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKF

[0559]

氨基酸球菌属种 (Acidaminococcus sp.)(菌株 BV3L6) UniProtKB U2UMQ6	KENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQT QIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELN SIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKV QRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKK QEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPS LSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFV KNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIP KCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLG EYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKP NLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKM LNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGI DRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQA WSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGI AEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAK MGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFL HYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTP FIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLE NDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQ NPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYI QELRN (SEQ ID NO: 64)
AsCas12a 切口酶(如, R1226A)	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKEL KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQA TYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVT TTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKF KENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQT QIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELN SIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKV QRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKK QEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPS LSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFV KNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIP KCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLG EYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKP NLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKM LNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGI DRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQA WSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGI AEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAK MGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFL HYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTP FIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLE NDDSHAIDTMVALIRSVLQMANSNAATGEDYINSPVRDLNGVCFDSRFQ NPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYI QELRN (SEQ ID NO: 65)
LbCas12a	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQ QELKEIMDDYYRTFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKI

	(以前又被称为 Cpf1) *Lachnospiraceae bacterium GAM79* 参考序列 WP_119623382.1	QNEKRKEICCYFTSDKRFKDLFNAKLITDILPNFIKDNKEYTEEEKAEKEQ TRVLFQRFATAFTNYFNQRRNNFSEDNISTAISFRIVNENSEIHLQNMRAF QRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYSVDFYDRELTQPGIEYY NGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFRFES DQEVYDALNEFIKTMKKKEIIRRCVHLGQECDDYDLGKIYISSNKYEQIS NALYGSWDTIRKCIKEEYMDALPGKGEKKEEKAEAAAKKEEYRSIADID KIISLYGSEMDRTISAKKCITEICDMAGQISIDPLVCNSDIKLLQNKEKTTEI KTILDSFLHVYQWGQTFIVSDIIEKDSYFYSELEDVLEDFEGITTLYNHVR SYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQKFYLGI FNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSRSGQE TYKPSKHILDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPDWKNY DFHFSDTKDYEDISGFYREVEMQGYQIKWTYISADEIQKLDEKGQIFLFQI YNKDFSVHSTGKDNLHTMYLKNLFSEENLKDIVLKLNGEAELFFRKASI KTPIVHKKGSVLVNRSYTQTVGNKEIRVSIPEEYYTEIYNYLNHIGKGKLS SEAQRYLDEGKIKSFTATKDIVKNYRYCCDHYFLHLPITINFKAKSDVAV NERTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFNIVNGYD YQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNA VVAMEDLNYGFKTGRFKVERQVYQKFETMLIEKLHYLVFKDREVCEEG GVLRGYQLTYIPESLKKVGKQCGFIFYVPAGYTSKIDPTTGFVNLFSFKN LTNRESRQDFVGKFDEIRYDRDKKMFEFSFDYNNYIKKGTILASTKWKV YTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIEYHDGKDLKG QIVEKGIEAEIIDIFRLTVQMRNSRSESEDREYDRLISPVLNDKGEFFDTAT ADKTLPQDADANGAYCIALKGLYEVKQIKENWKENEQFPRNKLVQDNK TWFDFMQKKRYL (SEQ ID NO: 66)
[0560]	PcCas12a – 以前又被称为 Cpf1 *Prevotella copri* 参考序列 WP_119227726.1	MAKNFEDFKRLYSLSKTLRFEAKPIGATLDNIVKSGLLDEDEHRAASYV KVKKLIDEYHKVFIDRVLDDGCLPLENKGNNNSLAEYYESYVSRAQDED AKKKFKEIQQNLRSVIAKKLTEDKAYANLFGNKLIESYKDKEDKKKIIDS DLIQFINTAESTQLDSMSQDEAKELVKEFWGFVTYFYGFFDNRKNMYTA EEKSTGIAYRLVNENLPKFIDNIEAFNRAITRPEIQENMGVLYSDFSEYLN VESIQEMFQLDYYNMLLTQKQIDVYNAIIGGKTDDEHDVKIKGINEYINL YNQQHKDDKLPKLKALFKQILSDRNAISWLPEEFNSDQEVLNAIKDCYE RLAENVLGDKVLKSLLGSLADYSLDGIFIRNDLQLTDISQKMFGNWGVIQ NAIMQNIKRVAPARKHKESEEDYEKRIAGIFKKADSFSISYINDCLNEADP NNAYFVENYFATFGAVNTPTMQRENLFALVQNAYTEVAALLHSDYPTV KHLAQDKANVSKIKALLDAIKSLQHFVKPLLGKGDESDKDERFYGELAS LWAELDTVTPLYNMIRNYMTRKPYSQKKIKLNFENPQLLGGWDANKEK DYATIILRRNGLYYLAIMDKDSRKLLGKAMPSDGECYEKMVYKFFKDV TTMIPKCSTQLKDVQAYFKVNTDDYVLNSKAFNKPLTITKEVFDLNNVL YGKYKKFQKGYLTATGDNVGYTHAVNVWIKFCMDFLNSYDSTCIYDFS SLKPESYLSLDAFYQDANLLLYKLSFARASVSYINQLVEEGKMYLFQIYN KDFSEYSKGTPNMHTLYWKALFDERNLADVVYKLNGQAEMFYRKKSIE NTHPTHPANHPILNKNKDNKKKESLFDYDLIKDRRYTVDKFMFHVPITM NFKSVGSENINQDVKAYLRHADDMHIIGIDRGERHLLYLVVIDLQGNIKE QYSLNEIVNEYNGNTYHTNYHDLLDVREEERLKARQSWQTIENIKELKE GYLSQVIHKITQLMVRYHAIVVLEDLSKGFMRSRQKVEKQVYQKFEKM LIDKLNYLVDKKTDVSTPGGLLNAYQLTCKSDSSQKLGKQSGFLFYIPA WNTSKIDPVTGFVNLLDTHSLNSKEKIKAFFSKFDAIRYNKDKKWFEFNL DYDKFGKKAEDTRTKWTLCTRGMRIDTFRNKEKNSQWDNQEVDLTTE MKSLLEHYYIDIHGNLKDAISAQTDKAFFTGLLHILKLTLQMRNSITGTET DYLVSPVADENGIFYDSRSCGNQLPENADANGAYNIARKGLMLIEQIKN AEDLNNVKFDISNKAWLNFAQQKPYKNG (SEQ ID NO: 67)

[0561]

ErCas12a – 以前又被称为 Cpf1 直肠真杆菌 (*Eubacterium rectale*) 参考序列 WP_119223642.1	MFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKDYFKNRA NCFSANDISSSSCHRIVNDNAEIFFSNALVYRRIVKNLSNDDINKISGDMK DSLKEMSLEEIYSYEKYGEFITQEGISFYNDICGKVNLFMNLYCQKNKEN KNLYKLRKLHKQILCIADTSYEVPYKFESDEEVYQSVNGFLDNISSKHIVE RLRKIGENYNGYNLDKIYIVSKFYESVSQKTYRDWETINTALEIHYNNILP GNGKSKADKVKKAVKNDLQKSITEINELVSNYKLCPDDNIKAETYIHEIS HILNNFEAQELKYNPEIHLVESELKASELKNVLDVIMNAFHWCSVFMTEE LVDKDNNFYAELEEIYDEIYPVISLYNLVRNYVTQKPYSTKKIKLNFGIPT LADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPDKKIIEGNTSENKGD YKKMIYNLLPGPNKMIPKVFLSSKTGVETYKPSAYILEGYKQNKHLKSS KDFDITFCHDLIDYFKNCIAIHPEWKNFGFDFSDTSTYEDISGFYREVELQ GYKIDWTYISEKDIDLLQEKGQLYLFQIYNKDFSKKSSGNDNLHTMYLK NLFSEENLKDIVLKLNGEAEIFFRKSSIKNPIIHKKGSILVNRTYEAEEKDQ FGNIQIVRKTIPENIYQELYKYFNDKSDKELSDEAAKLKNVVGHHEAATN IVKDYRYTYDKYFLHMPITINFKANKTSFINDRILQYIAKEKDLHVIGIDR GERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQIKLKQQEGARQIARKEWK EIGKIKEIKEGYLSLVIHEISKMVIKYNAIIAMEDLSYGFKKGRFKVERQV YQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPDKLKNVGHQCGC IFYVPAAYTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRYDSDKNLF CFTFDYNNFITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNESDTIDITK DMEKTLEMTDINWRDGHDLRQDIIDYEIVQHIFEIFKLTVQMRNSLSELE DRDYDRLISPVLNENNIFYDSAKAGDALPKDADANGAYCIALKGLYEIK QITENWKEDGKFSRDKLKISNKDWFDFIQNKRYL (SEQ ID NO: 68)
CsCas12a – 以前又被称为 Cpf1 梭菌属种 (*Clostridium sp.*) *AF34-10BH* 参考序列 WP_118538418.1	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQ QELKEIMDDYYRAFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKI QNEKRKEICCYFTSDKRFKDLFNAKLITDILPNFIKDNKEYTEEEKAEKEQ TRVLFQRFATAFTNYFNQRRNNFSEDNISTAISFRIVNENSEIHLQNMRAF QRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYLVDFYDRVLTQPGIEYY NGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFRFES DQEVYDALNEFIKTMKEKEIICRCVHLGQKCDDYDLGKIYISSNKYEQIS NALYGSWDTIRKCIKEEYMDALPGKGEKKEEKAEAAAKKEEYRSIADID KIISLYGSEMDRTISAKKCITEICDMAGQISTDPLVCNSDIKLLQNKEKTTE IKTILDSFLHVYQWGQTFIVSDIIEKDSYFYSELEDVLEDFEGITTLYNHVR SYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQKFYLGI FNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSRSGQE TYKPSKHILDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPDWKNY DFHFSDTKDYEDISGFYREVEMQGYQIKWTYISADEIQKLDEKGQIFLFQI YNKDFSVHSTGKDNLHTMYLKNLFSEENLKDIVLKLNGEAELFFRKASI KTPVVHKKGSVLVNRSYTQTVGDKEIRVSIPEEYYTEIYNYLNHIGRGKL STEAQRYLEERKIKSFTATKDIVKNYRYCCDHYFLHLPITINFKAKSDIAV NERTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFNIVNGYD YQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNA VVAMEDLNYGFKTGRFKVERQVYQKFETMLIEKLHYLVFKDREVCEEG GVLRGYQLTYIPESLKKVGKQCGFIFYVPAGYTSKIDPTTGFVNLFSFKN LTNRESRQDFVGKFDEIRYDRDKKMFEFSFDYNNYIKKGTMLASTKWK VYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIEYHDGKDLK GQIVEKGIEAEIIDIFRLTVQMRNSRSESEDREYDRLISPVLNDKGEFFDTA TADKTLPQDADANGAYCIALKGLYEVKQIKENWKENEQFPRNKLVQDN KTWFDFMQKKRYL (SEQ ID NO: 69)
BhCas12b	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHH EQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELY EELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLK

[0562]	外村尚芽孢杆菌 *(Bacillus hisashii)* 参考序列 WP_095142515.1	IAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVK EIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKE YKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGW REIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENH FIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERS GSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLL PSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRR YPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKEL TEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIE GKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQ KLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIREL MYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKN IDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLK KMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRF ENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIR CSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLS KDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPE SKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSD ILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISK LTNQYSISTIEDDSSKQSM (SEQ ID NO: 70)
	ThCas12b *Thermomonas hydrothermalis* 参考序列 WP_072754838	MSEKTTQRAYTLRLNRASGECAVCQNNSCDCWHDALWATHKAVNRGA KAFGDWLLTLRGGLCHTLVEMEVPAKGNNPPQRPTDQERRDRRVLLAL SWLSVEDEHGAPKEFIVATGRDSADDRAKKVEEKLREILEKRDFQEHEID AWLQDCGPSLKAHIREDAVWVNRRALFDAAVERIKTLTWEEAWDFLEP FFGTQYFAGIGDGKDKDDAEGPARQGEKAKDLVQKAGQWLSARFGIGT GADFMSMAEAYEKIAKWASQAQNGDNGKATIEKLACALRPSEPPTLDT VLKCISGPGHKSATREYLKTLDKKSTVTQEDLNQLRKLADEDARNCRKK VGKKGKKPWADEVLKDVENSCELTYLQDNSPARHREFSVMLDHAARR VSMAHSWIKKAEQRRRQFESDAQKLKNLQERAPSAVEWLDRFCESRSM TTGANTGSGYRIRKRAIEGWSYVVQAWAEASCDTEDKRIAAARKVQAD PEIEKFGDIQLFEALAADEAICVWRDQEGTQNPSILIDYVTGKTAEHNQK RFKVPAYRHPDELRHPVFCDFGNSRWSIQFAIHKEIRDRDKGAKQDTRQ LQNRHGLKMRLWNGRSMTDVNLHWSSKRLTADLALDQNPNPNPTEVT RADRLGRAASSAFDHVKIKNVFNEKEWNGRLQAPRAELDRIAKLEEQG KTEQAEKLRKRLRWYVSFSPCLSPSGPFIVYAGQHNIQPKRSGQYAPHA QANKGRARLAQLILSRLPDLRILSVDLGHRFAAACAVWETLSSDAFRREI QGLNVLAGGSGEGDLFLHVEMTGDDGKRRTVVYRRIGPDQLLDNTPHP APWARLDRQFLIKLQGEDEGVREASNEELWTVHKLEVEVGRTVPLIDRM VRSGFGKTEKQKERLKKLRELGWISAMPNEPSAETDEKEGEIRSISRSVD ELMSSALGTLRLALKRHGNRARIAFAMTADYKPMPGGQKYYFHEAKEA SKNDDETKRRDNQIEFLQDALSLWHDLFSSPDWEDNEAKKLWQNHIAT LPNYQTPEEISAELKRVERNKKRKENRDKLRTAAKALAENDQLRQHLHD TWKERWESDDQQWKERLRSLKDWIFPRGKAEDNPSIRHVGGLSITRINTI SGLYQILKAFKMRPEPDDLRKNIPQKGDDELENFNRRLLEARDRLREQR VKQLASRIIEAALGVGRIKIPKNGKLPKRPRTTVDTPCHAVVIESLKTYRP DDLRTRRENRQLMQWSSAKVRKYLKEGCELYGLHFLEVPANYTSRQCS RTGLPGIRCDDVPTGDFLKAPWWRRAINTAREKNGGDAKDRFLVDLYD HLNNLQSKGEALPATVRVPRQGGNLFIAGAQLDDTNKERRAIQADLNAA ANIGLRALLDPDWRGRWWYVPCKDGTSEPALDRIEGSTAFNDVRSLPTG DNSSRRAPREIENLWRDPSGDSLESGTWSPTRAYWDTVQSRVIELLRRH AGLPTS (SEQ ID NO: 71)

[0563]	LsCas12b *Laceyella sacchari* WP_132221894.1	MSIRSFKLKLKTKSGVNAEQLRRGLWRTHQLINDGIAYYMNWLVLLRQ EDLFIRNKETNEIEKRSKEEIQAVLLERVHKQQQRNQWSGEVDEQTLLQ ALRQLYEEIVPSVIGKSGNASLKARFFLGPLVDPNNKTTKDVSKSGPTPK WKKMKDAGDPNWVQEYEKYMAERQTLVRLEEMGLIPLFPMYTDEVGD IHWLPQASGYTRTWDRDMFQQAIERLLSWESWNRRVRERRAQFEKKTH DFASRFSESDVQWMNKLREYEAQQEKSLEENAFAPNEPYALTKKALRG WERVYHSWMRLDSAASEEAYWQEVATCQTAMRGEFGDPAIYQFLAQK ENHDIWRGYPERVIDFAELNHLQRELRRAKEDATFTLPDSVDHPLWVRY EAPGGTNIHGYDLVQDTKRNLTLILDKFILPDENGSWHEVKKVPFSLAKS KQFHRQVWLQEEQKQKKREVVFYDYSTNLPHLGTLAGAKLQWDRNFL NKRTQQQIEETGEIGKVFFNISVDVRPAVEVKNGRLQNGLGKALTVLTHP DGTKIVTGWKAEQLEKWVGESGRVSSLGLDSLSEGLRVMSIDLGQRTSA TVSVFEITKEAPDNPYKFFYQLEGTEMFAVHQRSFLLALPGENPPQKIKQ MREIRWKERNRIKQQVDQLSAILRLHKKVNEDERIQAIDKLLQKVASWQ LNEEIATAWNQALSQLYSKAKENDLQWNQAIKNAHHQLEPVVGKQISL WRKDLSTGRQGIAGLSLWSIEELEATKKLLTRWSKRSREPGVVKRIERFE TFAKQIQHHINQVKENRLKQLANLIVMTALGYKYDQEQKKWIEVYPAC QVVLFENLRSYRFSFERSRRENKKLMEWSHRSIPKLVQMQGELFGLQVA DVYAAYSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFIKEEHRP YLQQGDLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQKRFWHPS MWFRVNCESVMEGEIVTYVPKNKTVHKKQGKTFRFVKVEGSDVYEWA KWSKNRNKNTFSSITERKPPSSMILFRDPSGTFFKEQEWVEQKTFWGKV QSMIQAYMKKTIVQRMEE (SEQ ID NO: 72)
	DtCas12b *Dsulfonatronum thiodismutans* *WP_031386437*	MVLGRKDDTAELRRALWTTHEHVNLAVAEVERVLLRCRGRSYWTLDR RGDPVHVPESQVAEDALAMAREAQRRNGWPVVGEDEEILLALRYLYEQ IVPSCLLDDLGKPLKGDAQKIGTNYAGPLFDSDTCRRDEGKDVACCGPF HEVAGKYLGALPEWATPISKQEFDGKDASHLRFKATGGDDAFFRVSIEK ANAWYEDPANQDALKNKAYNKDDWKKEKDKGISSWAVKYIQKQLQL GQDPRTEVRRKLWLELGLLPLFIPVFDKTMVGNLWNRLAVRLALAHLLS WESWNHRAVQDQALARAKRDELAALFLGMEDGFAGLREYELRRNESIK QHAFEPVDRPYVVSGRALRSWTRVREEWLRHGDTQESRKNICNRLQDR LRGKFGDPDVFHWLAEDGQEALWKERDCVTSFSLLNDADGLLEKRKGY ALMTFADARLHPRWAMYEAPGGSNLRTYQIRKTENGLWADVVLLSPRN ESAAVEEKTFNVRLAPSGQLSNVSFDQIQKGSKMVGRCRYQSANQQFEG LLGGAEILFDRKRIANEQHGATDLASKPGHVWFKLTLDVRPQAPQGWL DGKGRPALPPEAKHFKTALSNKSKFADQVRPGLRVLSVDLGVRSFAACS VFELVRGGPDQGTYFPAADGRTVDDPEKLWAKHERSFKITLPGENPSRK EEIARRAAMEELRSLNGDIRRLKAILRLSVLQEDDPRTEHLRLFMEAIVD DPAKSALNAELFKGFGDDRFRSTPDLWKQHCHFFHDKAEKVVAERFSR WRTETRPKSSSWQDWRERRGYAGGKSYWAVTYLEAVRGLILRWNMRG RTYGEVNRQDKKQFGTVASALLHHINQLKEDRIKTGADMIIQAARGFVP RKNGAGWVQVHEPCRLILFEDLARYRFRTDRSRRENSRLMRWSHREIVN EVGMQGELYGLHVDTTEAGFSSRYLASSGAPGVRCRHLVEEDFHDGLP GMHLVGELDWLLPKDKDRTANEARRLLGGMVRPGMLVPWDGGELFAT LNAASQLHVIHADINAAQNLQRRFWGRCGEAIRIVCNQLSVDGSTRYEM AKAPKARLLGALQQLKNGDAPFHLTSIPNSQKPENSYVMTPTNAGKKYR AGPGEKSSGEEDELALDIVEQAEELAQGRKTFFRDPSGVFFAPDRWLPSE IYWSRIRRRIWQVTLERNSSGRQERAEMDEMPY (SEQ ID NO: 73)

[0564] 本文所述的多瓣引导编辑器还可包含Cas12a(Cpf1)(dCpf1)变体,其可用作向导核苷酸序列-可编程DNA结合蛋白结构域。Cas12a(Cpf1)蛋白具有与Cas9的RuvC结构域相似的RuvC样核酸内切酶结构域,但没有HNH核酸内切酶结构域,并且Cas12a(Cpf1)的N端没有Cas9的α-螺旋识别裂片(lobe)。Zetsche et al.,Cell,163,759-771,2015(通过引用并入

本文)表明,Cas12a的RuvC样结构域(Cpf1)负责切割两条DNA链,并且失活RuvC样结构域使Cas12a(Cpf1)核酸酶活性失活。

[0565] 在一些实施方案中,napDNAbp是微生物CRISPR-Cas系统的单一效应物。微生物CRISPR-Cas系统的单一效应物包括但不限于Cas9、Cas12a(Cpf1)、Cas12b1(C2c1)、Cas13a(C2c2)和Cas12c(C2c3)。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应物复合物,而2类系统具有单一蛋白效应物。例如,Cas9和Cas12a(Cpf1)是2类效应物。除Cas9和Cas12a(Cpf1)之外,三种不同的2类CRISPR-Cas系统(Cas12b1、Cas13a和Cas12c)描述于Shmakov et al.,"Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems",Mol.Cell,2015Nov 5;60(3):385-397,其全部内容通过引用在此并入。

[0566] 系统中两种(Cas12b1和Cas12c)的效应物包含与Cas12a相关的RuvC样核酸内切酶结构域。第三种系统Cas13a包含具有两个预测的HEPN RNase结构域的效应物。成熟CRISPR RNA的产生不依赖于tracrRNA,这与Cas12b1产生的CRISPRRNA不同。Cas12b1依赖CRISPR RNA和tracrRNA进行DNA切割。已证明细菌Cas13a对CRISPR RNA成熟具有独特的RNase活性,不同于其RNA激活的单链RNA降解活性。这些RNase功能彼此不同并且与Cas12a的CRISPR RNA加工行为不同。参见,例如East-Seletsky,et al.,"Two distinct RNase activities of CRISPR-Cas13a enable guide-RNAprocessing and RNAdetection",Nature,2016Oct 13;538(7624):270-273,其全部内容在此通过引入并入。Leptotrichiashahii中Cas13a的体外生化分析表明,Cas13a由单链CRISPRRNA引导,并且可编程以切割带有互补前间隔序列的ssRNA靶标。两个保守的HEPN结构域中的催化残基介导切割。催化残基中的突变产生催化死亡的RNA结合蛋白。参见,例如,Abudayyeh et al.,"C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector",Science,2016Aug 5;353(6299),其全部内容在此通过引用并入。

[0567] 已报道Alicyclobaccillus acidoterrastris Cas12b1(AacC2c1)的晶体结构与嵌合单分子向导RNA(sgRNA)复合。参见,例如,Liu et al.,"C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism",Mol.Cell,2017Jan 19;65(2):310-322,其全部内容通过在此引用并入。也报道了在以三元复合物形式与靶DNA结合的嗜酸耐热菌(Alicyclobacillus acidoterrestris)C2c1中的晶体结构。参见,例如,Yang et al.,"PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease",Cell,2016Dec 15;167(7):1814-1828,其全部内容在此通过引用并入。已独立地捕获AacC2c1的催化能力的构象(在靶DNA链和非靶DNA链的这两种情况下),位于单个RuvC催化口袋内,其中C2c1介导的切割导致靶DNA的交错的七核苷酸断裂。C2c1三元复合物与先前确定的Cas9和Cpf1对应物之间的结构比较表明CRISPR-Cas9系统使用的机制的多样性。

[0568] 在一些实施方案中,napDNAbp可以是C2c1、C2c2或C2c3蛋白。在一些实施方案中,napDNAbp是C2c1蛋白。在一些实施方案中,napDNAbp是Cas13a蛋白。在一些实施方案中,napDNAbp是Cas12c蛋白。在一些实施方案中,napDNAbp包含与天然存在的Cas12b1(C2c1)、Cas13a(C2c2)或Cas12c(C2c3)蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同的氨基酸

序列。在一些实施方案中,napDNAbp是天然存在的Cas12b1(C2c1)、Cas13a(C2c2)或Cas12c(C2c3)蛋白。

[0569] H.Cas9环状排列体

[0570] 在不同实施方案中,本文公开的多瓣引导编辑器可包含Cas9的环状排列体。

[0571] 术语“环状排列的Cas9”或Cas9的“环状排列体”或“CP-Cas9”,是指任何Cas9蛋白或其变体,其出现或已修饰以工程化为环状排列体变体,这意味着Cas9蛋白(如,野生型Cas9蛋白)的N端和C端已局部重排。当与向导RNA(gRNA)复合时,这种环状排列的Cas9蛋白或其变体保留了结合DNA的能力。参见Oakes et al.,“Protein Engineering ofCas9 for enhanced function,”Methods Enzymol,2014,546:491-511,以及Oakes et al.,“CRISPR-Cas9Circular Permutants as Programmable Scaffolds for Genome Modification,”Cell,January 10,2019,176:254-267,其各自通过引用并入本文。本公开考虑任何先前已知的CP-Cas9或使用新的CP-Cas9,只要所得到的环状排列的蛋白在与向导RNA(gRNA)复合时保留结合DNA的能力。

[0572] 本文所述任何Cas9蛋白,包括任何变体、直系同源物或天然存在的Cas9或其等同物,可被重新配置为环状排列体变体。

[0573] 在不同实施方案中,Cas9的环状排列体可具有以下结构:

[0574] N-端-[原始C-端]-[可选的接头]-[原始N-端]-C-端。

[0575] 作为示例,本公开考虑了经典化脓性链球菌Cas9(UniProtKB-Q99ZW2(CAS9_STRP1)的1368个氨基酸(编号基于SEQ ID NO:18中的氨基酸位置))的以下环状排列体:

[0576] N-端-[1268-1368]-[可选的接头]-[1-1267]-C-端;

[0577] N-端-[1168-1368]-[可选的接头]-[1-1167]-C-端;

[0578] N-端-[1068-1368]-[可选的接头]-[1-1067]-C-端;

[0579] N-端-[968-1368]-[可选的接头]-[1-967]-C-端;

[0580] N-端-[868-1368]-[可选的接头]-[1-867]-C-端;

[0581] N-端-[768-1368]-[可选的接头]-[1-767]-C-端;

[0582] N-端-[668-1368]-[可选的接头]-[1-667]-C-端;

[0583] N-端-[568-1368]-[可选的接头]-[1-567]-C-端;

[0584] N-端-[468-1368]-[可选的接头]-[1-467]-C-端;

[0585] N-端-[368-1368]-[可选的接头]-[1-367]-C-端;

[0586] N-端-[268-1368]-[可选的接头]-[1-267]-C-端;

[0587] N-端-[168-1368]-[可选的接头]-[1-167]-C-端;

[0588] N-端-[68-1368]-[可选的接头]-[1-67]-C-端;或

[0589] N-端-[10-1368]-[可选的接头]-[1-9]-C-端,或其他Cas9蛋白(包括其他Cas9直系同源物、变体等)的相应环状排列体。

[0590] 在特定实施方案中,环状排列的Cas9具有以下结构(基于化脓性链球菌Cas9(UniProtKB-Q99ZW2(CAS9_STRP1)的1368个氨基酸(编号基于SEQ ID NO:18中的氨基酸位置)):

[0591] N-端-[102-1368]-[可选的接头]-[1-101]-C-端;

[0592] N-端-[1028-1368]-[可选的接头]-[1-1027]-C-端;

[0593] N-端-[1041-1368]-[可选的接头]-[1-1043]-C-端;

[0594] N-端-[1249-1368]-[可选的接头]-[1-1248]-C-端;或

[0595] N-端-[1300-1368]-[可选的接头]-[1-1299]-C-端,或其他Cas9蛋白(包括其他Cas9直系同源物、变体等)的相应环状排列体。

[0596] 在其他实施方案中,环状排列的Cas9具有以下结构(基于化脓性链球菌Cas9(UniProtKB-Q99ZW2(CAS9_STRP1)的1368个氨基酸(编号基于SEQ ID NO:18中的氨基酸位置):

[0597] N-端-[103-1368]-[可选的接头]-[1-102]-C-端;

[0598] N-端-[1029-1368]-[可选的接头]-[1-1028]-C-端;

[0599] N-端-[1042-1368]-[可选的接头]-[1-1041]-C-端;

[0600] N-端-[1250-1368]-[可选的接头]-[1-1249]-C-端;或

[0601] N-端-[1301-1368]-[可选的接头]-[1-1300]-C-端,或其他Cas9蛋白(包括其他Cas9直系同源物、变体等)的相应环状排列体。

[0602] 在一些实施方案中,环状排列体可通过将Cas9的C端片段直接或通过使用接头(如,氨基酸接头)连接至Cas9的N端片段来形成。在一些实施方案中,C-端片段可对应于Cas9的C-端的95%或更多的氨基酸(如,约1300-1368个氨基酸),或Cas9的C-端的90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%或5%或更多的氨基酸(如,SEQ ID NO:77-86中的任一序列)。N-端部分可对应于Cas9的N-端的95%或更多的氨基酸(如,约1-1300个氨基酸),或Cas9的N-端的90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%、或5%或更多的氨基酸(如,SEQ ID NO:18)。

[0603] 在一些实施方案中,环状排列体可通过将Cas9的C端片段直接或通过使用接头(如,氨基酸接头)连接至Cas9的N端片段来形成。在一些实施方案中,重排至N端的C端片段包括或对应于Cas9的C端的30%或更少的氨基酸(如,SEQ ID NO:18的氨基酸1012-1368)。在一些实施方案中,重排至N端的C端片段包括或对应于Cas9(如,SEQ ID NO:18的Cas9)的C端的30%、29%、28%、27%、26%、25%、24%、23%、22%、21%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%或1%的氨基酸。在一些实施方案中,重排至N端的C端片段包括或对应于Cas9(如,SEQ ID NO:18的Cas9)的C端的410个残基或更少。在一些实施方案中,重排至N端的C端部分包括或对应于Cas9(如,SEQ ID NO:18的Cas9)的C端的410、400、390、380、370、360、350、340、330、320、310、300、290、280、270、260、250、240、230、220、210、200、190、180、170、160、150、140、130、120、110、100、90、80、70、60、50、40、30、20、或10个残基。在一些实施方案中,重排至N端的C端部分包括或对应于Cas9(如,SEQ ID NO:18的Cas9)的C端的357、341、328、120或69个残基。

[0604] 在其他实施方案中,环状排列体Cas9变体可定义为基于以下方法的Cas9一级结构的拓扑重排,其基于SEQ ID NO:18的化脓性链球菌Cas9:(a)选择对应于Cas9一级结构的内部氨基酸残基的环状排列体(CP)位点,其将原始蛋白分为两个半部:N端区域和C端区域;(b)通过将原始C端区域(包含CP位点氨基酸)移动至原始N端区域之前来修饰Cas9蛋白序列(如,通过基因工程技术),从而形成现在以CP位点氨基酸残基开头的Cas9蛋白的新N端。CP位点可位于Cas9蛋白的任何结构域中,包括例如螺旋-II结构域、RuvCIII结构域或CTD结构

域。例如,CP位点可位于(相对于SEQ ID NO:18的化脓性链球菌Cas9)原始氨基酸残基181、199、230、270、310、1010、1016、1023、1029、1071、1241249或1282。因此,一旦重新定位到N端,原始氨基酸181、199、230、270、310、1010、1016、1023、1029、1041、1247、1249将成为新的N-端氨基酸。这些CP-Cas9蛋白的命名可分别被认为是Cas9-CP^{181},Cas9-CP^{199}、Cas9-CP^{230}、Cas9-CP^{270}、Cas9-CP^{310}、Cas9-CP^{1010}、Cas9-CP^{1016}、Cas9-CP^{1023}、Cas9-CP^{1029}、Cas9-CP^{1041}、Cas9-CP^{1247}、Cas9-CP^{1249}、和Cas9-CP^{1282}。该描述并不意味着限于由SEQ ID NO:18制备CP变体,而是可实施以制备任何Cas9序列的CP变体,无论是在对应于这些位置的CP位点处,还是全部在其他CP位点处。此描述并非旨在以任何方式限制特定CP位点。几乎任何CP位点都可用于形成CP-Cas9变体。

[0605] 下面提供了基于SEQ ID NO:18的Cas9的示例性CP-Cas9氨基酸序列,其中接头序列加下划线显示并且任选的甲硫氨酸(M)残基以粗体显示。应当理解,本公开提供了不包括接头序列或包括不同接头序列的CP-Cas9序列。应当理解,CP-Cas9序列可基于除SEQ ID NO:18之外的Cas9序列,并且本文提供的任何示例不意味着限制。示例性CP-Cas9序列如下:

[0606]

CP命名	序列	SEQ ID NO:
CP1012	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR	SEQ ID NO: 77

[0607]

CP命名	序列	SEQ ID NO:
	VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGDKKYSIG LAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEI FSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY PKLESEFVYG	
CP1028	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGG SGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVIT DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR	SEQ ID NO: 78

[0608]

CP命名	序列	SEQ ID NO:
	TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF VYGDYKVYDVRKMIAKSEQ	
CP1041	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSG GSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVL GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG	SEQ ID NO: 79

[0609]

CP命名	序列	SEQ ID NO:
	GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV RKMIAKSEQEIGKATAKYFFYS	
CP1249	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL GGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTN SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGS	SEQ ID NO: 80
CP1300	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGS GGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKV PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF	SEQ ID NO: 81

[0610]

CP命名	序列	SEQ ID NO:
	DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSAYNKHRD	

[0611] 本文描述了可用于多瓣引导编辑构建体中的Cas9环状排列体。以下提供了Cas9(基于SEQ ID NO:18的Cas9)的示例性C端片段,其可重排至Cas9的N端。应当理解,Cas9的此类C端片段是示例性的,并不意味着限制。这些示例性CP-Cas9片段具有以下序列:

[0612]

CP命名	序列	SEQ ID NO:
CP1012 C-端片段	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYL ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL YETRIDLSQLGGD	SEQ ID NO: 82
CP1028 C-端片段	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE	SEQ ID NO: 83

[0613]

CP命名	序列	SEQ ID NO:
	LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
CP1041 C-端片段	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 84
CP1249 C-端片段	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL GGD	SEQ ID NO: 85
CP1300 C-端片段	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 86

[0614] I.具有经修饰的PAM特异性的Cas9变体

[0615] 本公开的引导编辑器还可包含具有经修饰的PAM特异性的Cas9变体。本公开的一些方面提供了对靶序列表现出活性的Cas9蛋白，该靶序列在其3'-端不包含经典PAM(5'-NGG-3'，其中N是A、C、G或T)。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NGG-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NNG-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NNA-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NNC-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NNT-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NGT-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NGA-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NGC-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NAA-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NAC-3'PAM序列的靶序列表现出活性。在一些实施方案中，Cas9蛋白对在其3'-端包含5'-NAT-3'PAM序列的靶序列表现出活性。在其他实施方案中，Cas9蛋白对在其3'-端包含5'-NAG-3'PAM序列的靶序列表现出活性。

[0616] 应当理解，本文所述从第一氨基酸残基(如，A)至第二氨基酸残基(如，T)的任何氨基酸突变(如，A262T)也可包括从第一氨基酸残基至与第二氨基酸残基相似(如，保守)的氨基酸残基的突变。例如，具有疏水侧链(如，丙氨酸、缬氨酸、异亮氨酸、亮氨酸、甲硫氨酸、苯丙氨酸、酪氨酸或色氨酸)的氨基酸的突变可为具有不同疏水侧链(如，丙氨酸、缬氨酸、异亮氨酸、亮氨酸、甲硫氨酸、苯丙氨酸、酪氨酸或色氨酸)的第二氨基酸的突变。例如，丙氨酸至苏氨酸的突变(如，A262T突变)也可以是丙氨酸至与苏氨酸在大小和化学性质上相似的

氨基酸(如,丝氨酸)的突变。作为另一个示例,具有带正电荷侧链(如,精氨酸、组氨酸或赖氨酸)的氨基酸突变可为具有不同的带正电荷侧链(如,精氨酸、组氨酸或赖氨酸)的第二氨基酸的突变。作为另一个示例,具有极性侧链(如,丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺)的氨基酸的突变可为具有不同极性侧链(如,丝氨酸、苏氨酸、天冬酰胺、或谷氨酰胺)的第二氨基酸的突变。其他类似的氨基酸对包括但不限于以下:苯丙氨酸和酪氨酸;天冬酰胺和谷氨酰胺;蛋氨酸和半胱氨酸;天冬氨酸和谷氨酸;以及精氨酸和赖氨酸。技术人员会认识到,这种保守的氨基酸取代可能对蛋白结构的影响很小,并且可能被很好地耐受而不损害功能。在一些实施方案中,本文提供的从一个氨基酸至苏氨酸的氨基酸突变的任何氨基可以是氨基酸突变为丝氨酸。在一些实施方案中,本文提供的从一个氨基酸至精氨酸的氨基酸突变中的任何氨基可以是氨基酸突变为赖氨酸。在一些实施方案中,本文提供的从一个氨基酸至异亮氨酸的氨基酸突变的任何氨基可以是氨基酸突变为丙氨酸、缬氨酸、甲硫氨酸或亮氨酸。在一些实施方案中,本文提供的从一个氨基酸至赖氨酸的氨基酸突变中的任何氨基可以氨基酸突变为精氨酸。在一些实施方案中,本文提供的从一个氨基酸至天冬氨酸的氨基酸突变的任何氨基可以是氨基酸突变为谷氨酸或天冬酰胺。在一些实施方案中,本文提供的从一个氨基酸至缬氨酸的氨基酸突变中的任何氨基可以是氨基酸突变为丙氨酸、异亮氨酸、甲硫氨酸或亮氨酸。在一些实施方案中,本文提供的从一个氨基酸至甘氨酸的氨基酸突变的任何氨基可以是氨基酸突变为丙氨酸。然而,应当理解,其他保守氨基酸残基将被本领域技术人员识别且其他保守氨基酸残基的任何氨基酸突变也在本公开的范围内。

[0617] 在一些实施方案中,Cas9蛋白包含对靶序列表现出活性的突变组合,该靶序列在其3'-端包含5'-NAA-3'PAM序列。在一些实施方案中,突变组合存在于表1中所列的任何一个克隆中。在一些实施方案中,突变组合是表1中所列克隆的保守突变。在一些实施方案中,Cas9蛋白包含表1中列出的任何一个Cas9克隆的突变组合。

[0618] 表1:NAAPAM克隆

[0619]

来自野生型 SpCas9 (如，SEQ ID NO: 18)的突变
D177N, K218R, D614N, D1135N, P1137S, E1219V, A1320V, A1323D, R1333K
D177N, K218R, D614N, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, G715C, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A367T, K710E, R1114G, D1135N, P1137S, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D861N, D1135N, K1188R, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, V743I, R753G, E762G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, S1274R, A1320V, R1333K
A10T, I322V, S409I, E427G, A589S, R753G, D1135N, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, E757K, G865G, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, R753G, E757K, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, K599R, M631A, R654L, K673E, V743I, R753G, N758H, E762G, D1135N, D1180G, E1219V, Q1221H, Q1256R, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N869S, N1054D, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, L727I, V743I, R753G, E762G, R859S, N946D, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, N1317T, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, K1151E, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1080S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y,

[0620]

L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, E630K, R654L, K673E, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, G1223S, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1801S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, M673I, N803S, N869S, G1077D, R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, R1114G, D1135N, E1219V, Q1221H, A1320V, R1333K

[0621] 在一些实施方案中,Cas9蛋白包含与表1的任何一个变体提供的Cas9蛋白的氨基酸序列至少80%相同的氨基酸序列。在一些实施方案中,Cas9蛋白包含与表1的任何一个变体提供的Cas9蛋白的氨基酸序列至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同的氨基酸序列。

[0622] 在一些实施方案中,与由SEQ ID NO:18提供的化脓性链球菌Cas9相比,Cas9蛋白对在其3'端不包含经典PAM(5'-NGG-3')的靶序列表现出增加的活性。在一些实施方案中,与由SEQ ID NO:18提供的化脓性链球菌Cas9对相同靶序列的活性相比,Cas9蛋白对具有与经典PAM序列(5'-NGG-3')不直接相邻的3′端的靶序列表现出活性增加至少5倍。在一些实施方案中,与由SEQ ID NO:2提供的化脓性链球菌对相同靶序列的活性相比,Cas9蛋白对与经典PAM序列(5'-NGG-3')不直接相邻的靶序列表现出活性增加至少10倍、至少50倍、至少100倍、至少500倍、至少1,000倍、至少5,000倍、至少10,000倍、至少50,000倍、至少100,000倍、至少500,000倍、或至少1,000,000倍。在一些实施方案中,靶序列的3′端与AAA、GAA、CAA或TAA序列直接相邻。在一些实施方案中,Cas9蛋白包含对在其3'端包含5'-NAC-3'PAM序列的靶序列表现出活性的突变组合。在一些实施方案中,突变组合存在于表2中所列的任何一个克隆中。在一些实施方案中,突变组合是表2中所列的克隆的保守突变。在一些实施方案中,Cas9蛋白包含表2中列出的任何一个Cas9克隆的突变组合。

[0623] 表2:NAC PAM克隆

[0624]

来自野生型 SPCAS9 (如，SEQ ID NO: 18)的突变
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
D1135N, E1219V, D1332N, R1335Q, T1337N
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
T472I, R753G, Q771H, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, K1156E, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, V647I, R753G, N803S, K959N, G1030R, I1055E, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, E630G, T638P, V647A, G687R, N767D, N803S, K959N, R1114G, D1135N, E1219V, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, I1057T, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, M631I, T638P, R753G, N803S, K959N, Y1036H, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, V875I, K959N, Y1016C, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N, I1348V
K608R, E627K, T638P, V647I, R654L, R753G, N803S, T804A, K848N, V922A, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R753G, N803S, V922A, K959N, K1014N, V1015A, R1114G, D1135N, K1156N, E1219V, N1252D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, A711T, R753G, K775R, K789E, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
I670S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, K797N, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, G752R, R753G, K797N, N803S, K948E, K959N, V1015A, Y1016S, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
I570T, A589V, K608R, E627K, T638P, V647I, R654L, Q716R, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1207G, E1219V, N1234D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, R654L, Q740R, R753G, N803S, K959N, N990S, T995S, V1015A, Y1036D, R1114G, D1135N, E1207G, E1219V, N1234D, N1266H, D1332N, R1335Q, T1337N
I562F, V565D, I570T, K608R, L625S, E627K, T638P, V647I, R654I, G752R, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, N1177S, N1234D, D1332N, R1335Q, T1337N
I562F, I570T, K608R, E627K, T638P, V647I, R753G, E790A, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, D1180E, A1184T, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1015A,

[0625]

R1114G, D1127A, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, L625S, E627K, T638P, V647I, R654I, T703P, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1016A, R1114G, D1135N, E1219V, K1246E, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R654L, K673E, R753G, E790A, N803S, K948E, K959N, R1114G, D1127G, D1135N, D1180E, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, L625S, E627K, T638P, V647I, R654I, I670T, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
E627K, M631V, T638P, V647I, K710E, R753G, N803S, N808D, K948E, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T, H1349R

[0626] 在一些实施方案中，Cas9蛋白包含与表2的任何一个变体提供的Cas9蛋白的氨基酸序列至少80％相同的氨基酸序列。在一些实施方案中，Cas9蛋白包含与表2的任何一个变体提供的Cas9蛋白的氨基酸序列至少85％、至少90％、至少92％、至少95％、至少96％、至少97％、至少98％、至少99％、或至少99.5％相同的氨基酸序列。

[0627] 在一些实施方案中，与SEQ ID NO:18提供的化脓性链球菌Cas9相比，Cas9蛋白对在其3'末端不包含经典PAM(5'-NGG-3')的靶序列表现出增加的活性。在一些实施方案中，与SEQ ID NO:18提供的化脓性链球菌Cas9对相同靶序列的活性相比，Cas9蛋白对具有与经典PAM序列(5'-NGG-3')不直接相邻的3′端的靶序列表现出活性增加至少5倍。在一些实施方案中，与SEQ ID NO:18提供的化脓性链球菌对相同靶序列的活性相比，Cas9蛋白对与经典PAM序列(5'-NGG-3')不直接相邻的靶序列表现出的活性增加至少10倍、至少50倍、至少100倍、至少500倍、至少1,000倍、至少5,000倍、至少10,000倍、至少50,000倍、至少100,000倍、至少500,000倍、或至少1,000,000倍。在一些实施方案中，靶序列的3'端与AAC、GAC、CAC或TAC序列直接相邻。

[0628] 在一些实施方案中，Cas9蛋白包含对在其3'端包含5'-NAT-3'PAM序列的靶序列表现出活性的突变组合。在一些实施方案中，突变组合存在于表3中所列的任何一个克隆中。在一些实施方案中，突变组合是表3中所列克隆的保守突变。在一些实施方案中，Cas9蛋白包含表3中列出的任何一个Cas9克隆的突变组合。

[0629] 表3:NAT PAM克隆

[0630]

来自野生型 SPCAS9 (如，SEQ ID NO: 18)的突变
K961E, H985Y, D1135N, K1191N, E1219V, Q1221H, A1320A, P1321S, R1335L

[0631]

D1135N, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
V743I, R753G, E790A, D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, V748I, V743I, R753G, D853E, V922A, R1114G D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1286K, P1321S, D1322G, R1335L
M631L, R654L, R753G, K797E, D853E, V922A, D1012A, R1114G D1135N, G1218S, E1219V, Q1221H, P1249S, N1317K, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, D596Y, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, Q1256R, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, K710E, V750A, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, K649R, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, K1156E, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, I1057G, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1308D, P1321S, D1322G, R1335L
M631L, R654L, R753G, D853E, V922A, R1114G, Y1131C, D1135N, E1150V, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, D853E, I1057V, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, I1057V, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L

[0632] 可与当前公开的多瓣引导编辑器结合使用的各种napDNAbps的以上描述并不意味着以任何方式进行限制。多瓣引导编辑器可包括经典SpCas9，或任何直系同源Cas9蛋白，或任何变体Cas9蛋白——包括任何天然存在的Cas9变体、突变体或其他工程化版本——其是已知的或可通过定向进化或其他诱变过程制备或进化。在不同实施方案中，Cas9或Cas9变体具有切口酶活性，即仅切割靶DNA序列的链。在其他实施方案中，Cas9或Cas9变体具有无活性的核酸酶，即“死亡”Cas9蛋白。可使用的其他变体Cas9蛋白是具有比经典SpCas9更小的分子量(如，为了更容易递送)或具有经修饰或重排的一级氨基酸结构(如，环状排列体形式)的那些。本文所述的多瓣引导编辑器还可包含Cas9等同物，包括Cas12a/Cpf1和Cas12b蛋白，它们是趋同进化的结果。本文使用的napDNAbps(如，SpCas9、Cas9变体或Cas9等同物)还可包含改变/增强其PAM特异性的各种修饰。最后，本申请考虑与参考Cas9序列(如，参考SpCas9经典序列或参考Cas9等同物(如，Cas12a/Cpf1))具有至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％、或至少99.9％序列同一性的任何Cas9、Cas9变体或Cas9等同物。

[0633] 在特定实施方案中，具有扩展的PAM能力的Cas9变体是SpCas9(H840A)VRQR(SEQ ID NO:87)，其具有以下氨基酸序列(其中V、R、Q、R取代相对于SEQ ID NO:51的SpCas9(H840A)以粗下划线显示。此外，SpCas9(H840A)VRQR去除了SpCas9(H840)中的甲硫氨酸残基：

[0634]

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR
KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH
AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 87)

[0635] 在另一个特定实施方案中，具有扩展的PAM能力的Cas9变体是SpCas9(H840A)VRER，其具有以下氨基酸序列(其中V、R、E、R取代相对于SEQ ID NO:51的SpCas9(H840A)以粗下划线显示。此外，SpCas9(H840A)VRER去除了SpCas9(H840)中的甲硫氨酸残基：

[0636]

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK
KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD
NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM

[0637]

TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ
LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY
LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 88)

[0638] 在一些实施方案中,在非经典PAM序列的情况下发挥功能的napDNAbp是Argonaute蛋白。这种核酸可编程DNA结合蛋白的一个示例是来自格氏嗜盐碱杆菌(Natronobacterium gregoryi)的Argonaute蛋白(NgAgo)。NgAgo是ssDNA引导的核酸内切酶。NgAgo结合约24个核苷酸(gDNA)的5'磷酸化ssDNA,以将其引导至靶位点,并在gDNA位点处产生DNA双链断裂。与Cas9相比,NgAgo-gDNA系统不需要前间隔序列邻近基序(PAM)。使用核酸酶无活性的NgAgo(dNgAgo)可极大地扩展可能靶向的碱基。NgAgo的特征和应用描述于Gao et al.,Nat Biotechnol.,2016Jul;34(7):768-73.PubMed PMID:27136078;Swarts et al.,Nature.507(7491)(2014):258-61;以及Swarts etal.,NucleicAcidsRes.43(10)(2015):5120-9,其各自通过引用并入本文。

[0639] 在一些实施方案中,napDNAbp是Argonaute蛋白的原核同源物。Argonaute蛋白的原核同源物是已知的,并且已描述于例如MakarovaK.,et al.,"Prokaryotic homologs ofArgonaute proteins are predicted to function as key components ofa novel system ofdefense against mobile genetic elements",BiolDirect.2009Aug 25;4:29.doi:10.1186/1745-6150-4-29,其全部内容在此通过引用并入。在一些实施方案中,napDNAbp是Marinitoga piezophile Argunaute(MpAgo)蛋白。CRISPR相关的Marinitoga piezophile Argunaute(MpAgo)蛋白使用5′-磷酸化向导物切割单链靶序列。所有已知的Argonaute蛋白都使用5′向导物。MpAgo-RNA复合物的晶体结构显示包含阻止5′磷酸酯相互作用的残基的向导链结合位点。该数据表明对5′-羟基化向导物具有非经典特异性的Argonaute蛋白亚类的进化。参见,例如,Kaya et al.,"A bacterialArgonaute with noncanonical guide RNA specificity",Proc NatlAcad Sci USA.2016Apr 12;113(15):4057-62,其全部内容通过引用在此并入)。应当理解,可使用其他argonaute蛋白,并且在本公开的范围内。

[0640] 本公开的一些方面提供了具有不同PAM特异性的Cas9结构域。通常,Cas9蛋白,例如来自化脓性链球菌(spCas9)的Cas9,需要经典NGG PAM序列来结合特定的核酸区域。这可能会限制编辑基因组内期望的碱基的能力。在一些实施方案中,本文提供的碱基编辑融合蛋白可能需要被放置在精确的位置,例如将靶碱基放置在4碱基区域内(如,"编辑窗口"),其位于PAM上游约15个碱基。参见Komor,A.C.,et al.,"Programmable editing ofa target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016),其全部内容通过引用在此并入。因此,在一些实施方案中,本文提供的任何融合蛋白可包含能够结合不包含经典(如,NGG)PAM序列的核苷酸序列的Cas9结构域。结合非经典PAM序列的Cas9结构域已在本领域中描述并且对技术人员而言是显而易见的。例如,结合非经典PAM序列的Cas9结构域已描述于Kleinstiver,B.P.,et al.,"Engineered CRISPR-Cas9nucleases with altered PAM specificities"Nature 523,481-485(2015);以及Kleinstiver,B.P.,et al.,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 bymodifyingPAM recognition"NatureBiotechnology 33,1293-1298(2015);各自的全部内容在此通过引用并入。

[0641] 例如,具有改变的PAM特异性的napDNAbp结构域,例如与具有以下氨基酸序列的野生型新凶手弗朗西斯菌(Francisella novicida)Cpf1(D917、E1006和D1255)(SEQ ID NO:74)具有至少80%、至少85%、至少90%、至少95%、或至少99%序列同一性的结构域:

[0642]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVEN

[0643]

QGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTE
DKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGN
IIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHE
IAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKT
GGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQ
EFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNW
DTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
TGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQ
EGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 74)

[0644] 具有改变的PAM特异性的其他napDNAbp结构域，例如与具有以下氨基酸序列的嗜热脱氮土芽孢杆菌(Geobacillus thermodenitrificans)Cas9(SEQ ID NO:75)具有至少80％、至少85％、至少90％、至少95％、或至少99％序列同一性的结构域：

[0645]

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRL
RRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARIL
LHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKR
NKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIE
KKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAF
HKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVY
GKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELI
EELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNI
PPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNR
KKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDH
VIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKK
KRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNG
RITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKT
DPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITG
AAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRL
LEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDV
FEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEF
PREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQV
DVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ ID NO: 75)

[0646] 在一些实施方案中，核酸可编程DNA结合蛋白(napDNAbp)是不需要经典(NGG)PAM序列的核酸可编程DNA结合蛋白。在一些实施方案中，napDNAbp是argonaute蛋白。这种核酸可编程DNA结合蛋白的一个示例是来自格氏嗜盐碱杆菌(Natronobacterium gregoryi)的Argonaute蛋白(NgAgo)。NgAgo是ssDNA引导的核酸内切酶。NgAgo结合约24个核苷酸(gDNA)

的5'磷酸化ssDNA,将其引导至靶位点,并在gDNA位点使DNA双链断裂。与Cas9相比,NgAgo-gDNA系统不需要前间隔序列邻近基序(PAM)。使用核酸酶失活的NgAgo(dNgAgo)可极大地扩展可能靶向的碱基。NgAgo的表征和应用描述于Gao et al.,NatBiotechnol.,34(7):768-73(2016),PubMed PMID:27136078;Swarts et al.,Nature,507(7491):258-61(2014);and Swarts et al.,NucleicAcidsRes.43(10)(2015):5120-9,其各自通过引用并入本文。格氏嗜盐碱杆菌Argonaute蛋白的序列在SEQ ID NO:76中提供。

[0647] 所公开的融合蛋白可包含与具有以下氨基酸序列的野生型格氏嗜盐碱杆菌Argonaute蛋白(SEQ ID NO:76)具有至少80%、至少85%、至少90%、至少95%、或至少99%序列同一性的napDNAbp结构域:

[0648]

MTVIDLDSTTTADELTSGHTYDISVTLTGVVYDNTDEQHPRMSLAFEQDNGERRYITLW
KNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDE
TFAGGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATDG
AKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLDCGVRVEA
DETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDLHER
YDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIVWGL
RDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGDDAVS
FPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDPVRLNGST
VEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVL
PEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEVD
AAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYTRNVA
LGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVYKDGTI
LGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATE
FLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFGAPEYLA
TRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTARLPITTAYADQASTHA
TKGYLVQTGAFESNVGFL (SEQ ID NO: 76)

[0649] 此外,可利用任何可获得的方法来获得或构建变体或突变体Cas9蛋白。如本文所用,术语"突变"是指序列(如,核酸或氨基酸序列)内的残基被另一个残基取代,或序列内一个或多个残基的缺失或插入。突变通常通过确定原始残基随后是该残基在序列中的位置以及通过新取代残基的身份来描述。用于进行本文提供的氨基酸取代(突变)的各种方法是本领域公知的,并且提供于例如Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor LaboratoryPress,Cold Spring Harbor,N.Y.(2012))。突变可以包括多种类别,例如单碱基多态性、微重复区域、插入缺失和倒位,并且不意味着以任何方式进行限制。突变可包括"功能丧失性"突变,这是降低或消除蛋白活性的突变的正常结果。大多数功能丧失突变是隐性的,因为在杂合子中,第二个染色体拷贝携带编码全功能蛋白的基因的未突变版本,其存在补偿了突变的影响。突变还包括"功能获得性"突变,这是赋予正常情况下不存在的蛋白或细胞异常活性的突变。许多功能获得性突变位于调节序列中而不是编码区中,因此会产生许多后果。例如,突变可能导致一个或多个基因在错误的组织中表达,这些组织获得了它们通常没有的功能。由于其性质,功能获得性突变通常是显性的。

[0650] 可使用定点诱变将突变引入参考Cas9蛋白。本领域已知的较旧的定点诱变方法依赖于待突变的序列亚克隆至载体中,例如M13噬菌体载体,其允许分离单链DNA模板。在这些方法中,将诱变引物(即,能够与待突变位点退火但在待突变位点具有一个或多个错配核苷酸的引物)与单链模板退火,然后从诱变引物的3′端开始聚合模板的互补序列。然后将得到

的双链体转化到宿主细菌中，并针对期望的突变筛选噬菌斑。最近，定点诱变采用了PCR方法，其优点是不需要单链模板。此外，还开发了不需要亚克隆的方法。在进行基于PCR的定点诱变时，必须考虑几个问题。首先，在这些方法中，需要减少PCR循环的次数，以防止由聚合酶引入非期望的突变扩大。其次，必须进行选择以减少反应中持续存在的非突变亲本分子的数量。第三，优选延伸长度的PCR方法，以便允许使用单个PCR引物组。第四，由于一些热稳定聚合酶的非模板依赖性末端延伸活性，在PCR产生的突变产物的平端连接之前，通常需要在程序中并入末端补齐步骤。

[0651] 突变也可通过定向进化过程引入，例如噬菌体辅助连续进化(PACE)或噬菌体辅助非连续进化(PANCE)。如本文所用，术语“噬菌体辅助连续进化(PACE)”是指采用噬菌体作为病毒载体的连续进化。PACE技术的一般概念已描述于，例如于2009年9月8日提交的国际PCT申请PCT/US2009/056194，2010年3月11日以WO2010/028347公布；于2011年12月22日提交的PCT国际申请PCT/US2011/066747，2012年6月28日以WO2012/088381公布；美国申请，2015年5月5日授权的美国专利号9,023,594；于2015年1月20日提交的国际PCT申请PCT/US2015/012022，2015年9月11日以WO2015/134121公布；于2016年4月15日提交的国际PCT申请PCT/US2016/027795，2016年10月20日以WO2016/168631公布，其各自的全部内容通过引用并入本文。变体Cas9也可通过噬菌体辅助非连续进化(PANCE)获得，如本文所用，是指使用噬菌体作为病毒载体的非连续进化。PANCE是快速体内定向进化的简化技术，使用进化的“选择噬菌体”(SP)的连续瓶转移(serial flasktransfers)，其中包含在新鲜大肠杆菌宿主细胞中待进化的感兴趣基因，从而允许宿主大肠杆菌内的基因保持恒定，而SP中包含的基因不断进化。连续瓶转移一直是广泛使用的微生物实验室进化方法，最近已开发出类似的方法用于噬菌体进化。PANCE系统的特点是比PACE系统的严格性低。

[0652] 如果尚未说明，以上提到的涉及Cas9或Cas9等同物的任何参考文献在此通过引用以其整体并入。

[0653] J.用于断裂PE递送的分开的napDNAbp结构域

[0654] 在不同实施方案中，本文所述的引导编辑器可作为两个或更多个片段递送至细胞，这些片段在细胞内组装(通过被动组装，或通过主动组装，如使用断裂内含肽序列)成为重新组成的引导编辑器。在某些情况下，自组装可能是被动的，通过两个或更多个引导编辑器片段在细胞内共价或非共价结合以重新组成引导编辑器。在其他情况下，自组装可由安装在每个片段上的二聚化结构域催化。二聚化结构域的示例在本文中描述。在其他情况下，自组装可由安装在每个引导编辑器片段上的断裂内含肽序列催化。

[0655] 断裂PE递送可能有利于解决不同递送方法的各种尺寸限制。例如，递送方法可包括基于病毒的递送方法、基于信使RNA的递送方法或基于RNP的递送(基于核糖核蛋白的递送)。并且，通过将引导编辑器分成较小的部分，这些递送方法中的每一种都可能更高效和/或更有效。一旦进入细胞，较小的部分就可组装成功能性引导编辑器。根据断裂的方式，分离的引导编辑器片段可以非共价方式或共价方式重新组装以重新组成引导编辑器。在一个实施方案中，引导编辑器可在一个或多个断裂位点断裂成两个或多个片段。片段可以是未经修饰的(除了被断裂)。一旦将片段递送至细胞(如，通过核糖核蛋白复合物的直接递送或通过核酸递送——例如，mRNA递送或基于病毒载体的递送)，片段可共价或非共价地重新结合以重新组成引导编辑器。在另一个实施方案中，引导编辑器可在一个或多个断割位点断

裂成两个或多个片段。每个片段可经修饰以包含二聚化结构域，由此形成的每个片段都与二聚化结构域结合。一旦在细胞内传递或表达，不同片段的二聚化结构域就会相互关联并结合，将不同的引导编辑器片段聚集在一起重新形成功能性引导编辑器。在又一个实施方案中，引导编辑器片段可经修饰以包含断裂内含肽。一旦在细胞内递送或表达，不同片段的断裂内含肽结构域相互关联并结合，然后进行反式剪接，这导致从各个片段中切除断裂内含肽结构域，并在片段之间相伴形成肽键，从而恢复引导编辑器。

[0656] 在一个实施方案中，使用断裂内含肽方法递送引导编辑器。

[0657] 断裂位点的位置可位于引导编辑器中的任何一对或多对残基之间以及其中的任何结构域中，包括在napDNAbp结构域内、聚合酶结构域(如，RT结构域)内、连接napDNAbp结构域和聚合酶结构域的接头结构域内。

[0658] 在一个实施方案中，如图66所示，将引导编辑器(PE)在napDNAbp内的断裂位点分开。

[0659] 在某些实施方案中，napDNAbp是SEQ ID NO:18的经典SpCas9多肽，如下：

[0660]

SpCas9 化脓性链球菌M1 SwissProt登录号 Q99ZW2 野生型 1368 AA	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 18

[0661] 在某些实施方案中，SpCas9在位于SEQ ID NO:18的经典SpCas9的残基1和2、或2和

3、或3和4、或4和5、或5和6、或6和7、或7和8、或8和9、或9和10之间，或者位于残基1至10、10至20、20至30、30至40、40至50、50至60、60至70、70至80、80至90、90至100、100至200、200至300、300至400、400至500、500至600、600至700、700至800、800至900、1000至1100、1100至1200、1200至1300、或1300至1368之间任意位置的任何两对残基之间的断裂位点处断裂成两个片段。

[0662] 在某些实施方案中，napDNAbp在位于对应于位于SEQ ID NO:18的经典SpCas9的位点1至10、10至20、20至30、30至40、40至50、50至60、60至70、70至80、80至90、90至100、100至200、200至300、300至400、400至500、500至600、600至700、700至800、800至900、1000至1100、1100至1200、1200至1300、或1300至1368之间任意位置的任何两对残基的一对残基处的断裂位点断裂成两个片段。

[0663] 在某些实施方案中，SpCas9在位于SEQ ID NO:18的经典SpCas9的残基1和2、或2和3、或3和4、或4和5、或5和6、或6和7之间，或者位于残基1至10、10至20、20至30、30至40、40至50、50至60、60至70、70至80、80至90、90至100、100至200、200至300、300至400、400至500、500至600、600至700、700至800、800至900、1000至1100、1100至1200、1200至1300、或1300至1368之间任意位置的任何两对残基之间的断裂位点处断裂成两个片段。在某些实施方案中，断裂位点位于一个或多个多肽键位置(即，“断裂位置或断裂内含肽断裂位置”)，与断裂内含肽融合，然后作为单独编码的融合蛋白递送至细胞。一旦断裂内含肽融合蛋白(即，蛋白半部)在细胞内表达，蛋白就会进行反式剪接以形成完整或全部PE，同时去除连接的断裂内含肽序列。

[0664] 例如，如图66所示，N-端外显肽可与第一断裂内含肽(如，N内含肽)融合，并且C-端外显肽可与第二断裂内含肽(如，C内含肽)融合。N端外显肽与C端外显肽融合，以在细胞内的N内含肽和C内含肽自结合时重新形成完整的引导编辑器融合蛋白，包括napDNAbp结构域和聚合酶结构域(如，RT结构域)，然后进行自我切除，并在完整的引导编辑器(PE)的N端外显肽和C端外显肽部分之间同时形成肽键。

[0665] 为了利用使用断裂内含肽的断裂PE递送策略，需要将引导编辑器在一个或多个断裂位点处分开以产生引导编辑器的至少两个分开的半部，如果每个半部都融合到断裂内含肽序列，则其各自在细胞内可重新连接。

[0666] 在某些实施方案中，引导编辑器在单个断裂位点处断裂。在某些其他实施方案中，引导编辑器在两个断裂位点、或三个断裂位点、或四个断裂位点或更多个断裂位点处断裂。

[0667] 在优选实施方案中，将引导编辑器在单个断裂位点处断裂以产生引导编辑器的两个分开的半部，每个半部可融合至断裂内含肽序列。

[0668] 示例性断裂内含肽是Ssp DnaE内含肽，其包含两个亚基，即DnaE-N和DnaE-C。这两个不同亚基由单独的基因编码，即dnaE-n和dnaE-c，它们分别编码DnaE-N和DnaE-C亚基。DnaE是天然存在于集胞藻(Synechocytissp.)PCC6803的断裂内含肽，能够指导两种不同蛋白的反式剪接，每种蛋白包含与DnaE-N或DnaE-C的融合物。

[0669] 其他天然存在或工程化的断裂内含肽序列是已知的或可由本文所述完整内含肽序列或本领域可获得的那些制备。断裂内含肽序列的示例可见于Stevens et al.,“A promiscuous split intein with expanded protein engineering applications,” PNAS,2017,Vol.114:8538-8543;Iwai et al.,“Highly efficient protein trans-

splicing by a naturally split DnaE intein from Nostc punctiforme,FEBS Lett,580:1853-1858,其各自通过引用并入本文。其他断裂内含肽序列可见于例如WO2013/045632、WO2014/055782、WO2016/069774和EP2877490,其各自的内容通过引用并入本文。

[0670] 此外,体内和体外蛋白反式剪接已描述于(Shingledecker,et al.,Gene207:187(1998),Southworth,et al.,EMBOJ.17:918(1998);Mills,et al.,Proc.Natl.Acad.Sci.USA,95:3543-3548(1998);Lew,et al.,J.Biol.Chem.,273:15887-15890(1998);Wu,et al.,Biochim.Biophys.Acta 35732:1(1998b),Yamazaki,et al.,J.Am.Chem.Soc.120:5591(1998),Evans,et al.,J.Biol.Chem.275:9091(2000);Otomo,et al.,Biochemistry 38:16040-16044(1999);Otomo,et al.,J.Biolmol.NMR 14:105-114(1999);Scott,et al.,Proc.Natl.Acad.Sci.USA 96:13638-13643(1999)),并提供将蛋白表达为两个无活性片段随后进行连接以形成功能性产物的机会,例如,如图66和图67关于从两个单独表达的半部形成完整的PE融合蛋白所示。

[0671] 在本文所述的不同实施方案中,连续进化方法(如,PACE)可用于进化碱基编辑器的第一部分。第一部分可包括单个组件或结构域,例如Cas9结构域、脱氨酶结构域或UGI结构域。然后可通过利用断裂内含肽多肽结构域分别表达进化部分和剩余非进化部分,将单独进化的组件或结构域融合到细胞内碱基编辑器的剩余部分。第一部分可更广泛地包括期望使用本文所述连续进化方法进化的碱基编辑器的任何第一氨基酸部分。在该实施方案中,第二部分是指未使用本文方法进化的碱基编辑器的剩余氨基酸部分。碱基编辑器的进化的第一部分和第二部分都可各自在细胞中利用断裂内含肽多肽结构域表达。细胞的天然蛋白剪接机制将重新组装进化的第一部分和非进化的第二部分,以形成单个融合蛋白进化的碱基编辑器。进化的第一部分可包含单个融合蛋白的N-或C-端部分。以类似的方式,使用第二正交反式剪接内含肽对可允许进化的第一部分包含单个融合蛋白的内部部分。

[0672] 因此,本文所述碱基编辑器的进化和非进化组件中的任何一个可用断裂内含肽标签表达,以促进包含细胞内进化和非进化组件的完整碱基编辑器的形成。

[0673] 已非常详细地研究了蛋白剪接加工的机制(Chong,et al.,J.Biol.Chem.1996,271,22159-22168;Xu,M-Q&Perler,F.B.EMBO Journal,1996,15,5146-5153),以及在内含肽和外显肽剪接点发现保守的氨基酸(Xu,et al.,EMBO Journal,1994,135517-522)。本文所述构建体包含与第一基因的5'-端融合的内含肽序列(如,碱基编辑器的进化部分)。合适的内含肽序列可选自已知含有蛋白剪接元件的任何蛋白。可在万维网上找到包含所有已知内含肽的数据库(Perler,F.B.NucleicAcids Research,1999,27,346-347)。内含肽序列在第二基因的3'端与5'端融合。为了将该基因靶向某个细胞器,可将肽信号与基因的编码序列融合。在第二基因之后,内含肽基因序列可根据需要重复多次,以在同一细胞中表达多种蛋白。对于含有多个内含肽的构建体,使用来自不同来源的内含肽元件可能有用。在待表达的最后一个基因序列之后,必须插入转录终止序列。在一个实施方案中,设计了经修饰的内含肽剪接单元,以便其既能催化切除来自内含肽的外显肽,又能防止外显肽的连接。发现火球菌(Pyrococcus)物种GB-DDNA聚合酶中C端外显肽连接的诱变产生改变的剪接元件,其诱导外显肽和内含肽的切割,但阻止随后的外显肽连接(Xu,M-Q&Perler,F.B.EMBO Journal,1996,15,5146-5153)。丝氨酸538突变为丙氨酸或甘氨酸诱导切割但阻止连接。由于在C端外显肽连接到内含肽处的保守氨基酸,其他内含肽剪接单元中等效残基的突变也应防止外

显肽连接。不含核酸内切酶结构域的优选内含肽是蟾分枝杆菌(Mycobacterium xenopi)GyrA蛋白(Telenti,et al.J.Bacteriol.1997,179,6378-6382)。其他已在自然界中发现或通过从含有内含肽的核酸内切酶中去除核酸内切酶结构域而人工产生(Chong,et al.J.Biol.Chem.1997,272,15587-15590)。在优选实施方案中,选择内含肽以使其组成执行剪接功能需要的最低数量的氨基酸,例如,来自蟾分枝杆菌GyrA蛋白的内含肽(Telenti,A.,et al.,J.Bacteriol.1997,179,6378-6382)。在替代实施方案中,选择没有核酸内切酶活性的内含肽,如来自蟾分枝杆菌GyrA蛋白的内含肽或经修饰去除核酸内切酶结构域的酿酒酵母(Saccharaomyces cerevisiae)VMA内含肽(Chong,1997)。内含肽剪接单元的进一步修饰可允许改变切割反应的反应速率,从而允许通过简单修饰剪接单元基因序列来控制蛋白剂量。

[0674] 内含肽也可作为由两个单独转录和翻译的基因编码的两个片段存在。这些所谓的断裂内含肽自结合并催化反式蛋白剪接活性。已在不同的蓝藻细菌和古细菌中鉴定了断裂内含肽(Caspi et al,Mol Microbiol.50:1569-1577(2003);Choi J.et al,J Mol Biol.556:1093-1106(2006.);Dassa B.et al,Biochemistry.46:322-330(2007.);Liu X.and Yang J.,J Biol Chem.275:26315-26318(2003);Wu H.et al.Proc Natl Acad Sci USA.£5:9226-9231(1998.);以及ZettlerJ.et al,FEBS Letters.553:909-914(2009)),但迄今为止尚未在真核生物中发现。最近,环境宏基因组数据的生物信息学分析揭示了26个具有新基因组排列的不同基因座。在各基因座上,保守的酶编码区被断裂内含肽中断,其中在编码内含肽亚结构域的部分之间插入独立的核酸内切酶基因。其中,5个基因座完全组装:DNA解旋酶(gp41-1、gp41-8);肌苷-5'-单磷酸脱氢酶(IMPDH-1);和核糖核苷酸还原酶催化亚基(NrdA-2和NrdJ-1)。这种断裂的基因组织似乎主要存在于噬菌体中(Dassa et al,Nucleic Acids Research.57:2560-2573(2009))。

[0675] 断裂内含肽NpuDnaE的特征在于具有蛋白反式剪接反应的最高报告率。此外,对于不同的外显肽序列的Npu DnaE蛋白剪接反应,6至37℃的温度和存在高达6M尿素被认为是稳健且高产的(Zettler J.et al,FEBS Letters.553:909-914(2009);Iwai I.et al,FEBS Letters 550:1853-1858(2006))。如预期,在这些内含肽的N结构域引入CyslAla突变时,最初的N至S-酰基转移以及因此蛋白剪接被阻断。不幸地,C-端切割反应也几乎被完全抑制。C端剪接连接处的天冬酰胺环化对N端易裂肽键的酰基转移的依赖性似乎是自然断裂DnaE内含肽等位基因共有的独特性质(ZettlerJ.et al.FEBS Letters.555:909-914(2009))。

[0676] 蛋白剪接的机制通常有四个步骤[29-30]:1)内含肽N-端的N-S或N-O酰基转移,这破坏上游肽键并在N-外显肽和内含肽的第一个氨基酸(Cys或Ser)侧链之间形成酯键;2)酯基转移将N-外显肽重新定位至内含肽C-端,形成将N-外显肽连接到C-外显肽的第一个氨基酸(Cys、Ser或Thr)侧链的新酯键;3)Asn环化破坏内含肽和C-外显肽之间的肽键;和4)S-N或O-N酰基转移,其利用N-外显肽和C-外显肽之间的肽键代替酯键。

[0677] 由断裂内含肽催化的蛋白反式剪接提供了用于蛋白连接的完全酶促方法[31]。断裂内含肽本质上是断裂成两部分(分别命名为N-内含肽和C-内含肽)的连续内含肽(如,微型内含肽)。断裂内含肽的N-内含肽和C-内含肽可以非共价结合形成活性内含肽,并以与连续内含肽基本相同的方式催化剪接反应。已在自然界中发现了断裂内含肽,还在实验室中进行了工程化改造[31-35]。如本文所用,术语"断裂内含肽"是指其中在N-端和C-端氨基酸

序列之间存在一个或多个肽键断裂，使得N-端和C-端序列成为单独的分子的任何内含肽，其可非共价地重新结合或重组为对反式剪接反应起作用的内含肽。任何具有催化活性的内含肽或其片段均可用于衍生用于本发明方法的断裂内含肽。例如，一方面，断裂内含肽可源自真核内含肽。另一方面，断裂内含肽可源自细菌内含肽。另一方面，断裂内含肽可源自古细菌内含肽。优选地，如此衍生的断裂内含肽将仅具有催化反式剪接反应所必需的氨基酸序列。

[0678] 如本文所用，“N-端断裂内含肽(In)”是指包含对反式剪接反应起作用的N-端氨基酸序列的任何内含肽序列。因此，In还包括在发生反式剪接时剪出的序列。In可包含天然存在的内含肽序列的N-端部分的修饰的序列。例如，In可包含额外的氨基酸残基和/或突变的残基，只要包含这样的额外和/或突变残基不会使In在反式剪接中失去功能。优选地，包含额外和/或突变残基提高或增强In的反式剪接活性。

[0679] 如本文所用，“C-端断裂内含肽(Ic)”是指包含对反式剪接反应起作用的C-端氨基酸序列的任何内含肽序列。一方面，Ic包含4至7个连续氨基酸残基，其中至少4个氨基酸来自衍生其的内含肽的最后的β-链。因此，Ic还包括在发生反式剪接时剪出的序列。Ic可包含天然存在的内含肽序列的C-端部分的修饰的序列。例如，Ic可包含额外的氨基酸残基和/或突变的残基，只要包含这样的额外和/或突变的残基不会使In在反式剪接中失去功能。优选地，包含额外和/或突变的残基提高或增强Ic的反式剪接活性。

[0680] 在本发明的一些实施方案中，与Ic或In连接的肽可包含额外的化学部分，其中包括荧光基团、生物素、聚乙二醇(PEG)、氨基酸类似物、非天然氨基酸、磷酸盐基、糖基、放射性同位素标记和药物分子。在其他实施方案中，与Ic连接的肽可包含一个或多个化学反应性基团，其中包括酮、醛、Cys残基和Lys残基。存在“内含肽剪接多肽(ISP)”时，断裂内含肽的N-内含肽和C-内含肽可非共价结合形成活性内含肽并催化剪接反应。如本文所用，“内含肽剪接多肽(ISP)”是指从断裂内含肽去除Ic、In或两者时保留的断裂内含肽的氨基酸序列部分。在某些实施方案中，In包括ISP。在另一个实施方案中，Ic包括ISP。在又一个实施方案中，ISP是单独的肽，其既不与In也不与Ic共价连接。

[0681] 断裂内含肽可通过工程化改造非结构化环中的一个或多个断裂位点或干预微型内含肽结构中存在的-12个保守β-链之间的氨基酸序列而由连续内含肽产生[25-28]。β-链之间区域内断裂位点的位置可能存在一定的灵活性，条件是断裂的产生不会破坏内含肽的结构，特别是结构化的β-链，达到足以使蛋白剪接活性丧失的程度。

[0682] 在蛋白反式剪接中，一个前体蛋白由N-外显肽部分和后面的N-内含肽组成，另一个前体蛋白由C-内含肽和后面的C-外显肽部分组成，并且反式剪接反应(由N-和C-内含肽共同催化)切除两个内含肽序列并用肽键连接这两个外显肽序列。蛋白反式剪接是酶促反应，可在非常低(如，微摩尔)浓度的蛋白下进行，并且可在生理条件下进行。

[0683] [2]其他可编程核酸酶

[0684] 在本文所述不同实施方案中，多瓣引导编辑器包含napDNAbp，如Cas9蛋白。这些蛋白通过与向导RNA(或PEgRNA，视情况而定)复合而“可编程”，其引导Cas9蛋白到DNA的靶位点，其具有与gRNA(或PEgRNA)的间隔区部分互补的序列，并且还具有所需的PAM序列。然而，在此设想的某些实施方案中，napDNAbp可用不同类型的可编程蛋白取代，例如锌指核酸酶或转录激活因子样效应物核酸酶(TALEN)。

[0685] 图1J描绘了本文考虑的引导编辑的这种变化,即将napDNAbp(如,SpCas9切口酶)用任何可编程核酸酶结构域(如,锌指核酸酶(ZFN)或转录激活因子样效应物核酸酶(TALEN))置换。因此,考虑合适的核酸酶不一定需要由核酸靶向分子(如,向导RNA)"编程",而是可通过定义DNA结合结构域的特异性来编程,例如尤其是核酸酶。正如利用napDNAbp部分进行引导编辑,最好对此类替代的可编程核酸酶进行修饰,以便仅切割靶DNA的一条链。换言之,可编程核酸酶应当优选起切口酶的作用。一旦选择了可编程核酸酶(如,ZFN或TALEN),就可将额外的功能工程化到系统中,以使其能够按照类似引导编辑机制进行操作。例如,可编程核酸酶可通过结合(如,通过化学接头)RNA或DNA延伸臂来修饰,其中延伸臂包含引物结合位点(PBS)和DNA合成模板。可编程核酸酶也可(如,通过化学或氨基酸接头)结合聚合酶,聚合酶的性质取决于延伸臂是DNA还是RNA。在RNA延伸臂的情况下,聚合酶可以是RNA依赖性DNA聚合酶(如,逆转录酶)。在DNA延伸臂的情况下,聚合酶可以是DNA依赖性DNA聚合酶(如,原核生物聚合酶,包括PolI、PolII或PolIII,或真核生物聚合酶,包括Pola、Polb、Polg、Pold、Pole或Polz)。该系统还可包括作为融合物添加到可编程核酸酶的其他功能或以反式添加来促进整个反应(如,(a)解旋酶在切割位点解开DNA以获得具有可用的3'端的切割链作为引物,(b)FEN1以帮助去除切割链上的内源性链,以驱动反应向着用合成链置换内源性链,或(c)nCas9:gRNA复合物形成相反链上的第二位点切口,这可能有助于通过非编辑链的有利细胞修复来驱动合成修复的整合)。以类似于使用napDNAbp进行引导编辑的方式,这种具有其他可编程核酸酶的复合物可用于合成,然后将携带感兴趣的编辑的新合成的DNA置换链永久地安装到DNA的靶位点中。

[0686] 合适的替代可编程核酸酶是本领域所熟知的,可用于代替napDNAbp:gRNA复合物来构建替代的引导编辑器系统,其可被编程以选择性地结合DNA的靶位点,并且其可进一步以上述方式修饰来将聚合酶和包含引物结合位点和DNA合成模板的RNA或DNA延伸臂共定位到特定切口位点。例如,如图1J所示,转录激活因子样效应物核酸酶(TALEN)可用作本文所述多瓣引导编辑方法和组合物中的可编程核酸酶。TALENS是通过将TAL效应物DNA结合结构域与DNA切割结构域融合而产生的人工限制酶。这些试剂可实现高效、可编程和特异的DNA切割,是原位基因组编辑的强大工具。可快速工程化改造转录激活因子样效应物(TALE)来结合几乎任何DNA序列。如本文所用,术语TALEN是广义的并且包括单体TALEN,其可切割双链DNA而无需另一个TALEN的帮助。术语TALEN也用于指一对TALEN中的一个或两个成员,其被工程化改造为在同一位点合作切割DNA。合作的TALEN可被称为左TALEN和右TALEN,其与DNA的旋向有关。参见美国系列号12/965,590;美国系列号13/426,991(美国专利号8,450,471);美国系列号13/427,040(美国专利号8,440,431);美国系列号13/427,137(美国专利号8,440,432);以及美国系列号13/738,381,其全部通过引用以其整体并入本文。此外,TALENS描述于WO 2015/027134,US 9,181,535,Boch et al.,"Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors",Science,vol.326,pp.1509-1512(2009),Bogdanove et al.,TAL Effectors:Customizable Proteins for DNA Targeting,Science,vol.333,pp.1843-1846(2011),Cade et al.,"Highly efficient generation of heritable zebrafish gene mutations using homo-and heterodimeric TALENs",Nucleic Acids Research,vol.40,pp.8001-8010(2012),以及Cermak et al.,"Efficient design and assembly of custom TALEN and other TAL effector-based

constructs for DNA targeting",NucleicAcids Research,vol.39,No.17,e82(2011),其各自通过引用并入本文。

[0687] 如图1J所示,锌指核酸酶也可用作替代的可编程核酸酶,用于代替napDNAbps(如,Cas9切口酶)进行多瓣引导编辑。与TALENS一样,ZFN蛋白可经修饰,使其起切口酶的作用,即,工程化改造ZFN,使其以类似于与本文所述的多瓣引导编辑器一起使用的napDNAbp的方式仅切割靶DNA的一条链。ZFN蛋白已在本领域中广泛描述,例如,Carroll et al.,"Genome Engineering with Zinc-Finger Nucleases,"Genetics,Aug 2011,Vol.188:773-782;Durai et al.,"Zinc finger nucleases:custom-designed molecular scissors for genome engineering ofplant and mammalian cells,"Nucleic Acids Res,2005,Vol.33:5978-90;and Gaj et al.,"ZFN,TALEN,and CRISPR/Cas-basedmethods forgenome engineering,"TrendsBiotechnol.2013,Vol.31:397-405,其各自通过引用以其整体并入本文。

[0688] [3]聚合酶(如,逆转录酶)

[0689] 在不同实施方案中,本文公开的多瓣引导编辑器系统包括聚合酶(如,DNA依赖性DNA聚合酶或RNA依赖性DNA聚合酶,例如逆转录酶),或其变体,其可与napDNAbp或其他可编程核酸酶作为融合蛋白提供,或以反式提供。

[0690] 任何聚合酶均可用于本文公开的多瓣引导编辑器中。聚合酶可以是野生型聚合酶、功能片段、突变体、变体或截短变体等。聚合酶可包括来自真核生物、原核生物、古细菌或病毒生物体的野生型聚合酶,和/或聚合酶可通过基因工程、诱变、基于定向进化的过程进行修饰。聚合酶可包括T7DNA聚合酶、T5DNA聚合酶、T4DNA聚合酶、Klenow片段DNA聚合酶、DNA聚合酶III等。聚合酶还可以是热稳定的,并且可包括Taq、Tne、Tma、Pfu、Tfl、Tth、Stoffel片段、VENT®和DEEPVENT®DNA聚合酶、KOD、Tgo、JDF3及其突变体、变体和衍生物(参见美国专利号5,436,149;美国专利号4,889,818;美国专利号4,965,185;美国专利号5,079,352;美国专利号5,614,365;美国专利号5,374,553;美国专利号5,270,179;美国专利号5,047,342号;美国专利号5,512,462;WO92/06188;WO92/06200;WO96/10640;Barnes,W.M.,Gene 112:29-35(1992);Lawyer,F.C.,et al.,PCR Meth.Appl.2:275-287(1993);Flaman,J.-M,et al.,Nuc.Acids Res.22(15):3259-3260(1994),其各自通过引用并入)。为了合成更长的核酸分子(如,长度超过约3-5Kb的核酸分子),可使用至少两种DNA聚合酶。在某些实施方案中,一种聚合酶可基本上缺乏3'核酸外切酶活性,而另一种聚合酶可具有3'核酸外切酶活性。这种配对可包括相同或不同的聚合酶。基本上缺乏3'核酸外切酶活性的DNA聚合酶的示例包括但不限于Taq、Tne(exo-)、Tma(exo-)、Pfu(exo-)、Pwo(exo-)、exo-KOD和TthDNA聚合酶及其突变体、变体和衍生物。

[0691] 优选地,可用于本文公开的多瓣引导编辑器中的聚合酶是"模板依赖性"聚合酶(因为聚合酶旨在依赖于DNA合成模板来指定引导编辑期间合成的DNA链的序列。如本文所使用,术语"模板DNA分子"是指例如在PEgRNA的DNA合成模板的引物延伸反应中通过DNA聚合酶合成互补核酸链的核酸链。

[0692] 如本文所用,术语"模板依赖性方式"旨在涉及引物分子的模板依赖性延伸(如,通过DNA聚合酶合成DNA)的过程。术语"模板依赖性方式"是指RNA或DNA的多核苷酸合成,其中新合成的多核苷酸链的序列由众所周知的互补碱基配对规则决定(参见,例如,Watson,

J.D.et al.,In:Molecular Biologyofthe Gene,4th Ed.,W.A.Benjamin,Inc.,Menlo Park,Calif.(1987))。术语“互补”是指两条多核苷酸链的区域之间或两个核苷酸之间通过碱基配对的序列互补的广义概念。众所周知,腺嘌呤核苷酸能够与胸腺嘧啶或尿嘧啶核苷酸形成特定的氢键(“碱基配对”)。类似地,已知胞嘧啶核苷酸能够与鸟嘌呤核苷酸碱基配对。因此,在引导编辑的情况下,可以说引导编辑器的聚合酶针对DNA合成模板合成的DNA单链与DNA合成模板的序列“互补”。

[0693] A.示例性聚合酶

[0694] 在不同实施方案中,本文所述的多瓣引导编辑器包含聚合酶。本公开涵盖从任何天然存在的生物体或病毒获得或从商业或非商业来源获得的任何野生型聚合酶。此外,可用于本公开的多瓣引导编辑器中的聚合酶可包括任何天然存在的突变聚合酶、工程化突变聚合酶或其他变体聚合酶,包括保留功能的截短变体。可用于本文的聚合酶也可被设计为包含特定的氨基酸取代,例如本文具体公开的那些。在某些优选实施方案中,可用于本公开的多瓣引导编辑器的聚合酶是基于模板的聚合酶,即它们以模板依赖性方式合成核苷酸序列。

[0695] 聚合酶是合成核苷酸链,并且可与此处描述的多瓣引导编辑器系统结合使用的酶。聚合酶优选为“模板依赖性”聚合酶(即,基于模板链的核苷酸碱基的顺序合成核苷酸链的聚合酶)。在某些配置中,聚合酶也可以是“非模板依赖性”(即,不需要模板链就合成核苷酸链的聚合酶)。聚合酶还可进一步分类为“DNA聚合酶”或“RNA聚合酶”。在不同实施方案中,多瓣引导编辑器系统包含DNA聚合酶。在不同实施方案中,DNA聚合酶可以是“DNA依赖性DNA聚合酶”(即,模板分子由此是DNA链)。在这种情况下,DNA模板分子可以是PEgRNA,其中延伸臂包含DNA链。在这种情况下,PEgRNA可被称为嵌合或杂合PEgRNA,其包含RNA部分(即,向导RNA组件,包括间隔区和gRNA核心)和DNA部分(即,延伸臂)。在不同的其他实施方案中,DNA聚合酶可以是“RNA依赖性DNA聚合酶”(即,模板分子由此是RNA链)。在这种情况下,PEgRNA是RNA,即包括RNA延伸。术语“聚合酶”还可指催化核苷酸聚合(即,聚合酶活性)的酶。通常,酶将在与多核苷酸模板序列退火的引物(如,与PEgRNA的引物结合位点退火的引物序列)的3’-端处开始合成,并将向着模板链的5’端前进。“DNA聚合酶”催化脱氧核苷酸的聚合。如本文所用提及DNA聚合酶,术语DNA聚合酶包括“其功能片段”。“其功能片段”是指野生型或突变型DNA聚合酶的任何部分,其包含少于聚合酶的完整氨基酸序列并在至少一组条件下保留催化核苷酸聚合的能力。这样的功能片段可作为单独的实体存在,或者它可以是较大多肽如融合蛋白的组分。

[0696] 在一些实施方案中,聚合酶可来自噬菌体。噬菌体DNA聚合酶通常缺乏5’至3’核酸外切酶活性,因为该活性由单独的多肽编码。合适的DNA聚合酶的示例是T4、T7和phi29DNA聚合酶。可商购的酶是:T4(可从许多来源获得,如Epicentre)和T7(可从许多来源获得,如未经修饰来自Epicenter,和用于3’至5’外切T7“测序酶”DNA聚合酶来自USB)。

[0697] 在其他实施方案中,聚合酶是古细菌聚合酶。已在古细菌中鉴定出2种不同类型的DNA聚合酶:1.B/pol家族I型(来自强烈火球菌(Pyrococcus furiosus)的Pfu的同源物)和2.pol II型(强烈火球菌(P.furiosus)DP1/DP22-亚基聚合酶的同源物)。来自这两类的DNA聚合酶已显示天然缺乏相关的5’至3’核酸外切酶活性,并具有3’至5’核酸外切酶(校对)活性。合适的DNA聚合酶(pol I或pol II)可源自具有与期望的测定温度相似的最佳生长温度

的古细菌。

[0698] 热稳定的古细菌DNA聚合酶分离自火球菌属(Pyrococcus)物种(火球菌(furiosus),GB-D物种,woesii,abysii,horikoshii),热球菌属(Thermococcus)物种(kodakaraensis KOD1,litoralis,物种9degrees North-7,物种JDF-3,gorgonarius),隐蔽热网菌(Pyrodictium occultum)和闪烁古生球菌(Archaeoglobusfulgidus)。

[0699] 聚合酶也可来自真细菌物种。有3类真细菌DNA聚合酶pol I、II和III。Pol I DNA聚合酶家族中的酶具有5'至3'核酸外切酶活性,某些成员还表现出3'至5'核酸外切酶活性。Pol II DNA聚合酶天然缺乏5'至3'核酸外切酶活性,但确实表现出3'至5'核酸外切酶活性。Pol III DNA聚合酶代表细胞的主要复制型DNA聚合酶,由多个亚基组成。Pol III催化亚基缺乏5'至3'核酸外切酶活性,但在某些情况下,3'至5'核酸外切酶活性位于同一多肽中。

[0700] 有各种可商购的Pol I DNA聚合酶,其中一些经修饰以降低或消除5'至3'核酸外切酶活性。

[0701] 合适的热稳定性pol I DNA聚合酶可从多种嗜热真细菌中分离,包括栖热菌属(Thermus)物种和海栖热袍菌(Thermotoga maritima),例如栖热水生菌(Thermus aquaticus,Taq)、嗜热栖热菌(Thermus thermophilus,Tth)和海栖热袍菌(Thermotoga maritima,TmaUlTma)。

[0702] 与以上所列那些相关的其他真细菌描述于Thermophilic Bacteria(Kristjansson,J.K.,ed.)CRC Press,Inc.,Boca Raton,Fla.,1992。

[0703] 本发明进一步提供嵌合或非嵌合DNA聚合酶,根据美国专利号5,677,152、6,479,264和6,183,998中公开的方法进行化学修饰,其内容通过引用以其整体在此并入。

[0704] 在以下参考文献中描述了与以上所列那些相关的其他古细菌DNA聚合酶:Archaea:A Laboratory Manual(Robb,F.T.and Place,A.R.,eds.),Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,1995and Thermophilic Bacteria(Kristjansson,J.K.,ed.)CRC Press,Inc.,Boca Raton,Fla.,1992。

[0705] B.示例性逆转录酶

[0706] 在不同实施方案中,本文所述的多瓣引导编辑器包含逆转录酶作为聚合酶。本公开涵盖从任何天然存在的生物体或病毒获得的任何野生型逆转录酶或从商业或非商业来源获得。此外,可用于本公开的多瓣引导编辑器中的逆转录酶可包括任何天然存在的突变RT、工程化突变RT或其他变体RT,包括保留功能的截短变体。RT也可被设计成含有特定的氨基酸取代,例如本文具体公开的那些。

[0707] 逆转录酶是多功能酶,通常具有三种酶活性,包括RNA依赖性和DNA依赖性DNA聚合活性,以及催化RNA-DNA杂合体中RNA裂解的RNaseH活性。一些逆转录酶突变体已禁用RNaseH部分以防止对mRNA的意外损伤。这些使用mRNA作为模板合成互补DNA(cDNA)的酶在RNA病毒中首次鉴定。随后,直接从病毒颗粒、细胞或组织中分离和纯化逆转录酶(例如,参见,Kacian et al.,1971,Biochim.Biophys.Acta 46:365-83;Yang et al.,1972,Biochem.Biophys.Res.Comm.47:505-11;Gerard et al.,1975,J.Virol.15:785-97;Liu et al.,1977,Arch.Virol.55187-200;Kato et al.,1984,J.Virol.Methods 9:325-39;Luke et al.,1990,Biochem.29:1764-69,以及Le Grice et al.,1991,J.Virol.65:7004-

07,其各自通过引用并入)。最近,为了寻求改进的特性,例如热稳定性、保真度和活性,已创建了突变体和融合蛋白。本文考虑本领域已知的或可使用本领域已知的方法制备的逆转录酶的任何野生型、变体和/或突变体形式。

[0708] 逆转录酶(RT)基因(或其中包含的遗传信息)可从许多不同的来源获得。例如,该基因可从逆转录病毒感染的真核细胞中获得,或从包含逆转录病毒基因组的一部分或整个基因组的许多质粒中获得。此外,可从逆转录病毒中获得含有RT基因的信使RNA样RNA。RT来源的示例包括但不限于莫洛尼鼠白血病病毒(M-MLV或MLVRT);人T细胞白血病病毒1型(HTLV-1);牛白血病病毒(BLV);劳斯肉瘤病毒(RSV);人免疫缺陷病毒(HIV);酵母,包括酵母菌、脉孢菌属、果蝇属;灵长类动物;和啮齿动物。参见,例如,Weiss,et al.,美国专利号4,663,290(1987);Gerard,G.R.,DNA:271-79(1986);Kotewicz,M.L.,et al.,Gene 35:249-58(1985);Tanese,N.,et al.,Proc.Natl.Acad.Sci.(USA):4944-48(1985);Roth,M.J.,at al.,J.Biol.Chem.260:9326-35(1985);Michel,F.,et al.,Nature 316:641-43(1985);Akins,R.A.,et al.,Cell 47:505-16(1986),EMBO J.4:1267-75(1985);以及Fawcett,D.F.,Cell 47:1007-15(1986)(其各自通过引用以其整体并入本文)。

[0709] 野生型RT

[0710] 与本文公开的多瓣引导编辑器一起使用的示例性酶可包括但不限于M-MLV逆转录酶和RSV逆转录酶。具有逆转录酶活性的酶是可商购的。在某些实施方案中,逆转录酶以反式提供给多瓣引导编辑器(PE)系统的其他组件。即,逆转录酶作为单独的组分表达或以其他方式提供,即,不是作为与napDNAbp的融合蛋白。

[0711] 本领域普通技术人员将认识到野生型逆转录酶,包括但不限于莫洛尼鼠白血病病毒(M-MLV);人免疫缺陷病毒(HIV)逆转录酶和禽肉瘤-白血病病毒(ASLV)逆转录酶,包括但不限于劳斯肉瘤病毒(RSV)逆转录酶、禽成髓细胞瘤病毒(AMV)逆转录酶、禽成红细胞增多症病毒(AEV))辅助病毒MCAV逆转录酶,禽骨髓细胞瘤病毒MC29辅助病毒MCAV逆转录酶,禽网状内皮增生病毒(REV-T)辅助病毒REV-A逆转录酶,禽肉瘤病毒UR2辅助病毒UR2AV逆转录酶,禽肉瘤病毒Y73AV辅助病毒逆转录酶、劳斯相关病毒(RAV)逆转录酶和成髓细胞血症相关病毒(MAV)逆转录酶可适当地用于本文所述主题方法和组合物。

[0712] 示例性的野生型RT酶如下:

[0713]

描述	序列
逆转录酶(M-MLV RT)野生型	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL

[0714]

描述	序列
莫洛尼鼠白血病病毒 用于 PE1 (本文公开的引导编辑器 1 融合蛋白)	VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP(SEQ ID NO: 700)
逆转录酶 莫洛尼鼠白血病病毒 参考序列 AAA66622.1	AFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNL AKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNVS MSFIWQSAPDIGRKLGRLEDLKSKTLGDLVREAEKIFNKRETP EEREERIRRETEEKEERRRTVDEQKEKERDRRRHREMSKLLAT VVIGQEQDRQEGERKRPQLDKDQCAYCKEKGHWAKDCPKKP RGPRGPRPQTSLLTLGDXGGQGQDPPPEPRITLKVGGQPVTFL VDTGAQHSVLTQNPGPLSDKSAWVQGATGGKRYRWTTDRKV HLATGKVTHSFLHVPDCPYPLLGRDLLTKLKAQIHFEGSGAQV VGPMGQPLQVLTLNIEDEYRLHETSKEPDVSLGFTWLSDFPQA WAESGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREV NKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRL HPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEAL HRDLADFR(SEQ ID NO: 90)
逆转录酶 猫白血病病毒 参考序列 NP955579.1	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAH CQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGIL KPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTV PNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEW RDPEIGLSGQLTWTRLPQGFKNSPTLFDEALHSDLADFRVRYP ALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAK KAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQV REFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLA FEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGP WKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLG QPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVH FGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLP DADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSA QRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRR GLLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQ AKGNRLADDTAKKAATETHSSLTVL SEQ ID NO: 91)
逆转录酶 HIV-1 RT, A 链	PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKE GKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTA FTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQ

[0715]

描述	序列
参考序列 ITL3-A	NPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTP DKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQ KLVGKLNWASQIYPGIKVRQLXKLLRGTKALTEVIPLTEEAEL ELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQ EPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGK TPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLW YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNRGRQKVV TLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPD QSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSA GIRKV(SEQ ID NO: 92) SEE MARTINELLI ET AL., VIROLOGY, 1990, 174(1): 135-144, WHICH IS INCORPORATED BY REFERENCE
逆转录酶 HIV-1 RT, B 链 参考序列 ITL3-B	PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKE GKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTA FTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQ NPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTP DKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQ KLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELE LAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTP KFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQ LEKEPIVGAETF(SEQ ID NO: 93) SEE STAMMERS ET AL., J. MOL. BIOL., 1994, 242(4): 586-588, WHICH IS INCORPORATED BY REFERENCE
逆转录酶 劳斯肉瘤病毒 RT 参考序列 ACL14945	TVALHLAIPLKWKPNHTPVWIDQWPLPEGKLVALTQLVEKEL QLGHIEPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFG AVQQGAPVLSALPRGWPLMVLDLKDCFFSIPLAEQDREAFAF TLPSVNNQAPARRFQWKVLPQGMTCSPTICQLIVGQILEPLRL KHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTIS PDKVQKEPGVQYLGYKLGSTYAAPVGLVAEPRIATLWDVQKL VGSLQWLRPALGIPPRLRGPFYEQLRGSDPNEAREWNLDMK MAWREIVQLSTTAALERWDPALPLEGAVARCEQGAIGVLGQG LSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGK EVDILLLPACFRDELPLPEGILLALRGFAGKIRSSDTPSIFDIARP LHVSLKVRVTDHPVPGPTVFTDASSSTHKGVVVWREGPRWEI KEIADLGASVQQLEARAVAMALLLWPTTPTNVVTDSAFVAKM LLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPG FFTEGNDVADSQATFQAYPLREAKDLHTALHIGPRALSKACNI SMQQAREVVQTCPHCNSAPALEAGVNPRGLGPLQIWQTDFTL EPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAAQHHWATVI AVLGRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQG QAMVERANRLLKDKIRVLAEGDGFMKRIPTSKQGELLAKAM YALNHFERGENTKTPIQKHWRPTVLTEGPPVKIRIETGEWEKG WNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDIAQKDEVT KKDEASPLFA(SEQ ID NO: 94) SEE YASUKAWA ET AL., J. BIOCHEM. 2009, 145(3): 315-324, WHICH IS INCORPORATED BY REFERENCE

[0716]

描述	序列
逆转录酶 花椰菜花叶病毒(CAULIFLOWER MOSAIC VIRUS) RT 参考序列 AGT42196	MMDHLLQKTQIQNQTEQVMNITNPNSIYIKGRLYFKGYKKIEL HCFVDTGASLCIASKFVIPEEHWINAERPIMVKIADGSSITINKV CRDIDLIIAGEIFHIPTVYQQESGIDFIIGNNFCQLYEPFIQFTDRV IFTKDRTYPVHIAKLTRAVRVGTEGFLESMKKRSKTQQPEPVNI STNKIAILSEGRRLSEEKLFITQQRMQKIEELLEKVCSENPLDP NKTKQWMKASIKLSDPSKAIKVKPMKYSPMDREEFDKQIKEL LDLKVIKPSKSPHMAPAFLVNNEAEKRRGKKRMVVNYKAMN KATVGDAYNLPNKDELLTLIRGKKIFSSFDCKSGFWQVLLDQD SRPLTAFTCPQGHYEWNVVPFGLKQAPSIFQRHMDEAFRVFR KFCCVYVDDILVFSNNEEDHLLHVAMILQKCNQHGIILSKKKA QLFKKKINFLGLEIDEGTHKPQGHILEHINKFPDTLEDKKQLQR FLGILTYASDYIPKLAQIRKPLQAKLKENVPWKWTKEDTLYM QKVKKNLQGFPPLHHPLPEEKLIIETDASDDYWGGMLKAIKIN EGTNTELICRYASGSFKAAEKNYHSNDKETLAVINTIKKFSIYL TPVHFLIRTDNTHFKSFVNLNYKGDSKLGRNIRWQAWLSHYS FDVEHIKGTDNHFADFLSREFNRVNS(SEQ ID NO: 95) SEE FARZADFAR ET AL., VIRUS GENES, 2013, 47(2): 347-356, WHICH IS INCORPORATED BY REFERENCE
逆转录酶 肺炎克雷伯菌(KLEBSIELLA PNEUMONIA) 参考序列 RFF81513.1	MKEKISKIDKNFYTDIFIKTSFQNEFEAGGVIPPIAKNQVSTISN KNKTFYSLAHSSPHYSIQTRIEKFLLKNIPLSASSFAFRKERSYL HYLEPHTQNVKYCHLDIVSFFHSIDVNIVRDTFSVYFSDEFLVK EKQSLLDAFMASVTLTAELDGVEKTFIPMGFKSSPSISNIIFRKI DILIQKFCDKNKITYTRYADDLLFSTKKENNILSSTFFINEISSIL SINKFKLNKSKYLYKEGTISLGGYVIENILKDNSSGNIRLSSSKL NPLYKALYEIKKGSSSKHICIKVFNLKLKRFIYKKNKEKFEAKF YSSQLKNKLLGYRSYLLSFVIFHKKYKCINPIFLEKCVFLISEIE SIMNRKF(SEQ ID NO: 96)
逆转录酶 大肠杆菌(ESCERICHIA COLI) RT 参考序列 TGH57013	MKITSNNVTAVINGKGWHSINWKKCHQHVKTIQTRIAKAACN QQWRTVGRLQRLLVRSFSARALAVKRVTENSGRKTPGVDGQI WSTPESKWEAIFKLRRKGYKPLPLKRVFIPKSNGKKRPLGIPV MLDRAMQALHLLGLEPVSETNADHNSYGFRPARCTADAIQQ VCNMYSSRNASKWVLEGDIKGCFEHISHEWLLENIPMDKQIL RNWLKAGIIEKSIFSKTLSGTPQGGIISPVLANMALDGLERLLQ NRFGRNRLI(SEQ ID NO: 97)
逆转录酶 枯草芽孢杆菌(BACILLUS SUBTILIS) RT 参考序列 QBJ66766	MSKIKINYEKYHIKPFPHFDQRIKVNKKVKENLQNPFYIAAHS FYPFIHYKKISYKFKNGTLSSPKERDIFYSGHMDGYIYKHYGEI LNHKYNNTCIGKGIDHVSLAYRNNKMGKSNIHFAAEVINFISE QQQAFIFVSDFSSYFDSLDHAILKEKLIEVLEEQDKLSKDWWN VFKHITRYNWVEKEEVISDLECTKEKIARDKKSRERYYTPAEF REFRKRVNIKSNDTGVGIPQGTAISAVLANVYAIDLDQKLNQY ALKYGGIYRRYSDDIIMVLPMTSDGQDPSNDHVSFIKSVVKRN KVTMGDSKTSVLYYANNNIYEDYQRKRESKMDYLGFSFDGM TVKIREKSLFKYYHRTYKKINSINWASVKKEKKVGRKKLYLL YSHLGRNYKGHGNFISYCKKAHAVFEGNKKIESLINQQIKRH WKKIQKRLVDV(SEQ ID NO: 98)
直肠真杆菌(EUBACTERIUM RECTALE) II 组内含子 RT	DTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKE HLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPT VTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALN IMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIV

	描述	序列
		RKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEM EKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKV NMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKF KKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTL CKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARR VAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCV TC(SEQ ID NO: 99)
[0717]	嗜热脂肪土芽孢杆菌(GEOBACILLUS STEAROTHERMOPHILUS) II 组内含子 RT	ALLERILARDNLITALKRVEANQGAPGIDGVSTDQLRDYIRAH WSTIHAQLLAGTYRPAPVRRVEIPKPGGGTRQLGIPTVVDRLIQ QAILQELTPIFDPDFSSSSFGFRPGRNAHDAVRQAQGYIQEGYR YVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGL KFCRYADDCNIYVKSLRAGQRVKQSIQRFLEKTLKLKVNEEKS AVDRPWKRAFLGFSFTPERKARIRLAPRSIQRLKQRIRQLTNPN WSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTT KTPQLHQALGKTYWTAQGLKSLTQR(SEQ ID NO: 100)

[0718] 变体和易错RT

[0719] 逆转录酶对于从RNA模板合成互补DNA(cDNA)链是必不可少的。逆转录酶是由表现出不同生化活性的不同结构域组成的酶。这些酶催化从RNA模板合成DNA,如下所示:在退火引物存在的情况下,逆转录酶与RNA模板结合并引发聚合反应。RNA依赖性DNA聚合酶活性合成互补DNA(cDNA)链,并结合dNTP。RNaseH活性降解DNA:RNA复合物的RNA模板。因此,逆转录酶包含(a)识别并结合RNA/DNA杂合体的结合活性,(b)RNA依赖性DNA聚合酶活性,和(c)RNaseH活性。此外,通常认为逆转录酶具有不同的属性,包括它们的热稳定性、持续合成能力(dNTP掺入率)和保真度(或错误率)。本文考虑的逆转录酶变体可包括影响或改变这些酶活性(如,RNA依赖性DNA聚合酶活性、RNaseH活性或DNA/RNA杂合结合活性)或酶特性(如,热稳定性、持续合成能力或保真度)中的任何一种或多种的逆转录酶的任何突变。此类变体可在本领域的公共领域中获得,可商购获得,或可使用已知的诱变方法制备,包括定向进化过程(如,PACE或PANCE)。

[0720] 在不同实施方案中,逆转录酶可以是变体逆转录酶。如本文所用,“变体逆转录酶”包括任何天然存在或基因工程改造的变体,其包含一个或多个相对于参考序列(如,参考野生型序列)的突变(包括单一突变、倒置、缺失、插入和重排)。RT自然具有几种活性,包括RNA依赖性DNA聚合酶活性、核糖核酸酶H活性和DNA依赖性DNA聚合酶活性。总的来说,这些活性使酶能够将单链RNA转化为双链cDNA。在逆转录病毒和逆转录转座子中,这种cDNA可整合到宿主基因组中,通过宿主细胞转录可从中产生新的RNA拷贝。变体RT可包含影响这些活性中的一个或多个突变(减少或增加这些活性,或一起消除这些活性)。此外,变体RT可包含一种或多种突变,这些突变使RT或多或少稳定,更不容易聚集,并促进纯化和/或检测,和/或其他性质或特征的修饰。

[0721] 本领域普通技术人员将认识到衍生自其他逆转录酶的变体逆转录酶,包括但不限于莫洛尼鼠白血病病毒(M-MLV);人免疫缺陷病毒(HIV)逆转录酶和禽肉瘤-白血病病毒(ASLV)逆转录酶,包括但不限于劳斯肉瘤病毒(RSV)逆转录酶、禽成髓细胞瘤病毒(AMV)逆转录酶、禽成红细胞增多症病毒(AEV))辅助病毒MCAV逆转录酶,禽骨髓细胞瘤病毒MC29辅助病毒MCAV逆转录酶,禽网状内皮增生病毒(REV-T)辅助病毒REV-A逆转录酶,禽肉瘤病毒

UR2辅助病毒UR2AV逆转录酶，禽肉瘤病毒Y73AV辅助病毒逆转录酶、劳斯相关病毒(RAV)逆转录酶和成髓细胞相关病毒(MAV)逆转录酶可适当地用于本文所述主题方法和组合物。

[0722] 制备变体RT的一种方法是通过遗传修饰(如，通过修饰野生型逆转录酶的DNA序列)。本领域已知许多允许DNA序列随机和靶向突变的方法(参见，例如，Ausubel et.al.ShortProtocols in MolecularBiology(1995)3.sup.rd Ed.JohnWiley&Sons,Inc.)。此外，还有许多用于定点诱变的市售试剂盒，包括常规方法和基于PCR的方法。示例包括QuikChange定点诱变试剂盒(AGILENT®)、Q5®定点诱变试剂盒(NEWENGLAND BIOLABS®)和GeneArt™定点诱变系统(THERMOFISHER SCIENTIFIC®)。

[0723] 此外，可根据本领域技术人员已知的方法通过插入突变或截短(N-末端、内部或C-末端插入或截短)来产生突变逆转录酶。如本文所用，术语“突变”是指序列(如，核酸或氨基酸序列)内的残基被另一残基取代，或缺失或在序列中插入一个或多个残基。突变在本文中通常通过确定原始残基随后是该残基在序列中的位置以及新取代残基的身份来描述。用于进行本文提供的氨基酸取代(突变)的各种方法是本领域公知的，并且由例如Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。突变可包括多种类别，例如单碱基多态性、微重复区域、插入缺失和倒置，并且不意味着以任何方式进行限制。突变可包括“功能丧失性”突变，这是降低或消除蛋白活性的突变的正常结果。大多数功能丧失性突变是隐性的，因为在杂合子中，第二染色体拷贝携带全功能蛋白的基因编码的未突变版本，其存在补偿了突变的影响。突变还包括“功能获得性”突变，这是一种赋予正常情况下不存在的蛋白或细胞异常活性的突变。许多功能获得性突变位于调节序列中而不是编码区中，因此会产生许多后果。例如，突变可能导致一个或多个基因在错误的组织中表达，这些组织获得了它们通常缺乏的功能。由于其性质，功能获得性突变通常是显性的。

[0724] 本领域已知的较早的定点诱变方法依赖于将待突变的序列亚克隆到载体中，例如M13噬菌体载体，其允许分离单链DNA模板。在这些方法中，将诱变引物(即，能够与待突变位点退火但在待突变位点带有一个或多个错配核苷酸的引物)与单链模板退火，然后从诱变引物的3′端开始聚合模板的互补序列。然后将得到的双链体转化到宿主细菌中，并针对期望的突变筛选噬菌斑。

[0725] 最近，定点诱变采用了PCR方法，其优点是不需要单链模板。此外，还开发了不需要亚克隆的方法。在进行基于PCR的定点诱变时，必须考虑几个问题。首先，在这些方法中，需要减少PCR循环的次数，以防止聚合酶引入非期望的突变扩大。其次，必须进行选择以减少反应中持续存在的非突变亲本分子的数量。第三，为了允许使用单个PCR引物组，优选延伸长度的PCR方法。第四，由于一些热稳定聚合酶的非模板依赖性末端延伸活性，在PCR产生的突变产物的平端连接之前，通常需要在程序中加入末端补齐步骤。

[0726] 本领域中存在随机诱变方法，其将导致一组带有一个或多个随机定位突变的突变体。然后可筛选出这样一组表现出期望特性的突变体，例如相对于野生型逆转录酶增加的稳定性。

[0727] 随机诱变方法的一个示例是所谓的“易错PCR方法”。顾名思义，该方法在DNA聚合酶不支持高保真掺入的条件下扩增给定序列。尽管促进不同DNA聚合酶易错掺入的条件各不相同，但本领域技术人员可确定给定酶的此类条件。许多DNA聚合酶在扩增保真度方面的

关键变量是,例如缓冲液中二价金属离子的类型和浓度。因此可应用锰离子的使用和/或镁或锰离子浓度的变化来影响聚合酶的错误率。

[0728] 在不同方面,多瓣引导编辑器的RT可以是“易错”逆转录酶变体。可使用本领域已知和/或可获得的易错逆转录酶。应当理解,逆转录酶天然不具有任何校对功能;因此逆转录酶的错误率通常高于包含校对活性的DNA聚合酶。任何特定逆转录酶的错误率都是酶“保真度”的特性,其代表了DNA针对其RNA模板的模板定向聚合的准确性。具有高保真度的RT具有低错误率。相反,具有低保真度的RT具有高错误率。据报道,基于M-MLV的逆转录酶的保真度在合成的15,000至27,000个核苷酸中有一个错误范围内的错误率。参见Boutabout et al.,“DNAsynthesis fidelitybythe reverse transcriptase ofthe yeast retrotransposon Tyl,”NucleicAcids Res,2001,29:2217-2222,其通过引用并入。因此,出于本申请的目的,被认为“易错”或被认为具有“易错保真度”的那些逆转录酶是在合成的15,000个核苷酸中具有小于一个错误的错误率的那些逆转录酶。

[0729] 还可通过起始RT酶(如,野生型M-MLVRT)的诱变产生易错逆转录酶。诱变方法不受限制,可包括定向进化过程,例如噬菌体辅助连续进化(PACE)或噬菌体辅助非连续进化(PANCE)。如本文所用,术语“噬菌体辅助连续进化(PACE)”是指采用噬菌体作为病毒载体的连续进化。PACE技术的一般概念已描述于,例如2009年9月8日提交的国际PCT申请PCT/US2009/056194,2010年3月11日以WO2010/028347公布;2011年12月22日提交的PCT国际申请PCT/US2011/066747,2012年6月28日以WO2012/088381公布;美国申请,2015年5月5日授权的美国专利号9,023,594;2015年1月20日提交的国际PCT申请PCT/US2015/012022,2015年9月11日以WO2015/134121公布,以及2016年4月15日提交的国际PCT申请PCT/US2016/027795,2016年10月20日以WO2016/168631公开,其各自全部内容通过引用并入本文。

[0730] 还可通过噬菌体辅助非连续进化(PANCE)来获得易错逆转录酶,如本文所用,其是指采用噬菌体作为病毒载体的非连续进化。PANCE是快速体内定向进化的简化技术,使用进化的“选择噬菌体”(SP)的连续瓶转移,其中包含在新鲜大肠杆菌宿主细胞中待进化的感兴趣基因,从而允许宿主大肠杆菌内的基因保持恒定,而SP中包含的基因不断进化。连续瓶转移一直是广泛使用的微生物实验室进化方法,最近已经开发出类似的方法用于噬菌体进化。PANCE系统的特点是比PACE系统的严格性更低。

[0731] 其他易错逆转录酶已在文献中有所描述,其各自都被考虑用于本文的方法和组合物中。例如,易错逆转录酶已描述于Bebenek et al.,“Error-prone Polymerization by HIV-1Reverse Transcriptase,”JBiol Chem,1993,Vol.268:10324-10334,以及Sebastian-Martin et al.,“Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases,”Scientific Reports,2018,Vol.8:627,其各自通过引用并入。更进一步地,逆转录酶,包括易错逆转录酶可获自商业供应商,包括ProtoScript®(II)逆转录酶,AMV逆转录酶,WarmStart®逆转录酶和M-MuLV逆转录酶,均来自NEW ENGLAND BIOLABS®,或AMV逆转录酶XL,SMARTScribe逆转录酶,GPR超纯MMLV逆转录酶,均来自TAKARABIO USA,INC.(以前的CLONTECH)。

[0732] 本公开还考虑在RNaseH结构域中具有突变的逆转录酶。如上所述,逆转录酶的内在特性之一是RNaseH活性,其在聚合的同时切割RNA:cDNA杂合体的RNA模板。RNaseH活性可

能不适合长cDNA的合成，因为RNA模板可能会在全长逆转录完成之前降解。RNaseH活性也可能降低逆转录效率，可能是由于它与酶的聚合酶活性竞争。因此，本公开内容考虑包含经修饰的RNaseH活性的任何逆转录酶变体。

[0733] 本公开还考虑在RNA依赖性DNA聚合酶结构域中具有突变的逆转录酶。如上所述，逆转录酶的内在特性之一是RNA依赖性DNA聚合酶活性，其将核碱基结合到由RNA:cDNA杂合体的模板RNA链编码的新生cDNA链中。可增加或减少RNA依赖性DNA聚合酶活性(即，就其掺入速率而言)以增加或减少酶的持续合成能力。因此，本公开考虑了包含经修饰的RNA依赖性DNA聚合酶活性的任何逆转录酶变体，使得酶的持续合成能力相对于未经修饰的版本增加或减少。

[0734] 本文还考虑了具有改变的热稳定性特征的逆转录酶变体。逆转录酶耐受高温的能力是cDNA合成的重要方面。升高的反应温度有助于使具有强二级结构和/或高GC含量的RNA变性，从而允许逆转录酶读取序列。因此，在较高温度下进行逆转录可实现全长cDNA合成和更高的产率，这可能会导致改善3′瓣ssDNA的生成，这是多瓣引导编辑加工的结果。野生型M-MLV逆转录酶的最佳温度范围通常为37-48℃；然而，可能会引入允许在超过48℃的更高温度下进行逆转录活性的突变，包括49℃、50℃、51℃、52℃、53℃、54℃、55℃、56℃、57℃、58℃、59℃、63℃、63℃、63℃、64℃、65℃、66℃及更高。

[0735] 本文考虑的变体逆转录酶，包括易错RT、热稳定RT、增加持续合成能力的RT，可通过不同的常规策略进行工程化改造，包括诱变或进化过程。在某些情况下，可通过引入单个突变来产生变体。在其他情况下，变体可能需要不止一个突变。对于包含多于一个突变的那些突变体，可通过定点诱变将已鉴定的突变引入野生型基因中并与特定突变体携带的其他突变分离来评估给定突变的影响。这样产生的单一突变体的筛选试验将允许单独确定该突变的影响。

[0736] 本文使用的变体RT酶还可包括与本文公开或考虑的或本领域已知的任何参考RT蛋白(包括任何野生型RT)、或突变体RT、或片段RT、或RT的其他变体具有至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同的其他“RT变体”。

[0737] 在一些实施方案中，RT变体可具有与参考RT相比具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、45、46、47、48、49、50、或多达100、或多达200、或多达300、或多达400、或多达500个或更多的氨基酸变化。在一些实施方案中，RT变体包含参考RT的片段，使得该片段与参考RT的相应片段至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同。在一些实施方案中，该片段为相应野生型RT(M-MLV逆转录酶)(如，SEQ ID NO:89)或SEQ ID NO:90-100的任何逆转录酶的氨基酸长度的至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％，或至少99.5％。

[0738] 在一些实施方案中，本公开还可利用保留其功能性并且是本文公开的任何RT蛋白

的片段的RT片段。在一些实施方案中，RT片段的长度为至少100个氨基酸。在一些实施方案中，该片段的长度为至少100、150、200、250、300、350、400、450、500、550或多达600个或更多氨基酸。

[0739] 在其他实施方案中，本公开还可利用在N-端或C-端或两者处截短一定数量的氨基酸的RT变体，这导致在截短变体中仍然保留足够的聚合酶功能。在一些实施方案中，RT截短变体具有蛋白的N端末端处至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个氨基酸。在其他实施方案中，RT截短变体具有蛋白的C端末端处至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、1090210、220、230、240或250个氨基酸。在其他实施方案中，RT截短变体在N-端和C-端末端具有相同或不同长度的截短。

[0740] 例如，本文公开的多瓣引导编辑器可包括M-MLV逆转录酶的截短版本。在该实施方案中，逆转录酶包含4个突变(D200N、T306K、W313F、T330P；注意PE2中存在的L603W突变由于截短而不再存在)。编码这种截短编辑器的DNA序列比PE2更小522bp，因此它可能适用于DNA序列的递送因其大小而具有挑战性的应用(即，腺相关病毒和慢病毒递送)。该实施方案被称为MMLV-RT(截短)并且具有以下氨基酸序列：

[0741]

MMLV-RT (截短)	*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQ APLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGL PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAA TSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEG QRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMA APLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDNSRLI N* (SEQ ID NO: 766)

[0742] 在不同实施方案中，本文公开的多瓣引导编辑器可包含本文所述RT变体之一，或其与任何参考Cas9变体具有至少约70％相同、至少约80％相同、至少约90％相同、在至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同的RT变体。

[0743] 在其他实施方案中，本方法和组合物可利用已进化成逆转录酶的DNA聚合酶，描述于Effefson et al.,“Synthetic evolutionaryoriginofaproofreading reverse transcriptase,”Science,June 24,2016,Vol.352:1590-1593，其内容通过引用并入本文。

[0744] 在某些其他实施方案中，逆转录酶以还包含napDNAbp的融合蛋白的组件提供。换言之，在一些实施方案中，逆转录酶融合至napDNAbp成为融合蛋白。

[0745] 在不同实施方案中，变体逆转录酶可由如SEQ ID NO:89所示的野生型M-MLV逆转录酶工程化改造。

[0746] 在不同实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括包含一个或多个以下突变的变体RT：SEQ ID NO:89的野生型M-MLV RT或其他野生型RT多肽序列中相应氨基酸位置处的P51L、S67K、E69K、L139P、T197A、D200N、H204R、F209N、E302K、E302R、T306K、F309N、W313F、T330P、L345G、L435G、N454K、D524G、E562Q、D583N、H594Q、L603W、E607K、或D653N。

[0747] 下面提供了一些示例性逆转录酶，其可与napDNAbp蛋白融合或作为根据本公开的各种实施方案的单独蛋白提供。示例性逆转录酶包括与以下野生型酶或部分酶具有至少80％、至少85％、至少90％、至少95％、或至少99％序列同一性的变体：

[0748]

描述	序列 (相对野生型的变体取代)
逆转录酶 (M-MLV RT) 野生型 莫洛尼鼠白血病病毒 用于 PE1 (本文公开的引导编辑器 1 融合蛋白)	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 89)
M-MLV RT	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL

[0749]

描述	序列 (相对野生型的变体取代)
D200N	LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 701)
M-MLV RT D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 702)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 740)
M-MLV RT D200N T330P L603W E69K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQ**K**ARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRA LLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNW

[0750]

描述	序列 (相对野生型的变体取代)
	GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMG QPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDT STLLIENSSP (SEQ ID NO: 703)
M-MLV RT D200N T330P L603W E302R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**R**FLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP(SEQ ID NO: 704)
M-MLV RT D200N T330P L603W E607K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 705)
M-MLV RT D200N T330P L603W L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA

[0751]

描述	序列 (相对野生型的变体取代)
	EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 706)
M-MLV RT D200N T330P L603W L435G	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVI**G**APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 707)
M-MLV RT D200N T330P L603W N454K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 708)
M-MLV RT D200N T330P L603W T306K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLG**K**AGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 709)
M-MLV RT D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL

[0752]

描述	序列 (相对野生型的变体取代)
T330P L603W W313F	LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 710)
M-MLV RT D200N T330P L603W D524G E562Q D583N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRA**Q**LIALTQALKMA EGKKLNVYT**N**SRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 711)
M-MLV RT D200N T330P L603W E302R W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**R**FLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 712)
M-MLV RT D200N T330P L603W E607K L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG

[0753]

描述	序列 (相对野生型的变体取代)
	PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 713)
M-MLV RT P51L S67K T197A H204R E302K F309N W313F T330P L435G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LII**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSP**A**LFDEAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVI**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT **G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDT STLLIENSSP (SEQ ID NO: 714)
M-MLV RT D200N P51L S67K T197A H204R E302K F309N W313F T330P L345G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LII**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSP**A**LF**N**EAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVI**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT **G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDT STLLIENSSP (SEQ ID NO: 715)
M-MLV RT D200N T330P L603W T306K W313F PE2 中	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA

[0754]

描述	序列 (相对野生型的变体取代)
	EGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 716)

[0755] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括包含一个或多个以下突变的变体RT:SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的P51X、S67X、E69X、L139X、T197X、D200X、H204X、F209X、E302X、T306X、F309X、W313X、T330X、L345X、L435X、N454X、D524X、E562X、D583X、H594X、L603X、E607X、或D653X，其中“X”可以是任何氨基酸。

[0756] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的P51X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是L。

[0757] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的S67X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是K。

[0758] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的E69X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是K。

[0759] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的L139X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是P。

[0760] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列中的相应氨基酸位点处的T197X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是A。

[0761] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的D200X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是N。

[0762] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的H204X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是R。

[0763] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的F209X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是N。

[0764] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反

式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的E302X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是K。

[0765] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的E302X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是R。

[0766] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的T306X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是K。

[0767] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的F309X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是N。

[0768] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的W313X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是F。

[0769] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的T330X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是P。

[0770] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的L345X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是G。

[0771] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的L435X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是G。

[0772] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的N454X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是K。

[0773] 在不同的其他实施方案中,本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT,其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的D524X突变,其中“X”可以是任何氨基酸。在某些实施方案中,X是G。

[0774] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的E562X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是Q。

[0775] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLVRT或另外的野生型RT多肽序列的相应氨基酸位点处的D583X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是N。

[0776] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的H594X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是Q。

[0777] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的L603X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是W。

[0778] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的E607X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是K。

[0779] 在不同的其他实施方案中，本文所述的多瓣引导编辑器(RT作为融合伴侣或以反式提供)可包括变体RT，其包含SEQ ID NO:89的野生型M-MLV RT或另外的野生型RT多肽序列的相应氨基酸位点处的D653X突变，其中“X”可以是任何氨基酸。在某些实施方案中，X是N。

[0780] 下面提供了一些示例性逆转录酶，其可与napDNAbp蛋白融合或根据本公开的各种实施方案作为单独的蛋白提供。示例性逆转录酶包括与以下野生型酶或部分酶具有至少80％、至少85％、至少90％、至少95％、或至少99％的序列同一性：

[0781]

描述	序列 (相对于野生型的变体取代)
逆转录酶 (M-MLV RT) 野生型 莫洛尼鼠白血病病毒 用于 PE1 (本文公开的引导编辑器 1 融合蛋白)	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 89)
M-MLV RT D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA

[0782]

描述	序列 (相对于野生型的变体取代)
	PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 106)
M-MLV RT D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 107)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 108)
M-MLV RT E69K D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQ**K**ARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP

[0783]

描述	序列 (相对于野生型的变体取代)
	(SEQ ID NO: 109)
M-MLV RT D200N T330P L603W E302R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLR**R**FLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP(S EQ ID NO: 110)
M-MLV RT D200N T330P L603W E607K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 111)
M-MLV RT D200N T330P L603W L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 112)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP

[0784]

描述	序列 (相对于野生型的变体取代)
L435G	TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI**G**A PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 113)
M-MLV RT D200N T330P L603W N454K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 114)
M-MLV RT D200N T330P L603W T306K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPD QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQ EGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSS P (SEQ ID NO: 115)
M-MLV RT D200N T330P L603W W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA

[0785]

描述	序列 (相对于野生型的变体取代)
	PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 116)
M-MLV RT D200N T330P L603W D524G E562Q D583N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRA**Q**LIALTQALKMAEGKKLN VYT**N**SRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 117)
M-MLV RT D200N T330P L603W E302R W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLR**R**FLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 118)
M-MLV RT D200N T330P L603W E607K L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP

描述	序列 (相对于野生型的变体取代)
	(SEQ ID NO: 119)
M-MLV RT P51L S67K T197A H204R E302K F309N W313F T330P L435G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL II**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SP**A**LFDEAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI **G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPA TLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GSSLL QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILALLKALFL PKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDTSTLLIENS SP (SEQ ID NO: 120)
M-MLV RT P51L S67K T197A D200N H204R E302K F309N W313F T330P L345G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL II**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SP**A**LF**N**EAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI **G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPA TLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GSSLL QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILALLKALFL PKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDTSTLLIENS SP (SEQ ID NO: 121)
M-MLV RT D200N T330P L603W T306K W313F PE2 中	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 122)

[0786]

[0787] 本文所述的多瓣引导编辑器考虑描述或公开于以下美国专利(其各自通过引用以其整体并入)中的任何公开可用的逆转录酶:美国专利号:10,202,658;10,189,831;10,150,955;9,932,567;9,783,791;9,580,698;9,534,201;和9,458,484,以及其使用安装突变的已知方法或进化蛋白的已知方法的任何变体。以下参考文献描述了本领域的逆转录酶。其各自通过引用以其整体并入本文。

[0788] Herzig,E.,Voronin,N.,Kucherenko,N.&Hizi,A.A Novel Leu92 Mutant of

HIV-1 Reverse Transcriptase with a Selective Deficiency in Strand Transfer Causes a Loss of Viral Replication.J.Virol.89,8119-8129(2015).

[0789] Mohr,G.et al.A Reverse Transcriptase-Cas1 Fusion Protein Contains a Cas6 Domain Required for Both CRISPR RNA Biogenesis and RNA Spacer Acquisition.Mol.Cell 72,700-714.e8(2018).

[0790] Zhao,C.,Liu,F.&Pyle,A.M.An ultraprocessive,accurate reverse transcriptase encoded by a metazoan group II intron.RNA 24,183-195(2018).

[0791] Zimmerly,S.&Wu,L.An Unexplored Diversity of Reverse Transcriptases in Bacteria.Microbiol Spectr3,MDNA3-0058-2014(2015).

[0792] Ostertag,E.M.&Kazazian Jr,H.H.Biology of Mammalian L1 Retrotransposons.Annual Review of Genetics 35,501-538(2001).

[0793] Perach,M.&Hizi,A.Catalytic Features of the Recombinant Reverse Transcriptase of Bovine Leukemia Virus Expressed in Bacteria.Virology 259,176-189(1999).

[0794] Lim,D.et al.Crystal structure of the moloney murine leukemia virus RNase H domain.J.Virol.80,8379-8389(2006).

[0795] Zhao,C.&Pyle,A.M.Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution.Nature Structural &Molecular Biology 23,558-565(2016).

[0796] Griffiths,D.J.Endogenous retroviruses in the human genome sequence.Genome Biol.2,REVIEWS1017(2001).

[0797] Baranauskas,A.et al.Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants.Protein EngDes Sel 25,657-668(2012).

[0798] Zimmerly,S.,Guo,H.,Perlman,P.S.&Lambowltz,A.M.Group II intron mobility occurs by target DNA-primed reverse transcription.Cell82,545-554(1995).

[0799] Feng,Q.,Moran,J.V.,Kazazian,H.H.&Boeke,J.D.Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.Cell87,905-916(1996).

[0800] Berkhout,B.,Jebbink,M.&Zsíros,J.Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus.Journal of Virology 73,2365-2375(1999).

[0801] Kotewicz,M.L.,Sampson,C.M.,D'Alessio,J.M.&Gerard,G.F.Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity.Nucleic AcidsRes 16,265-277(1988).

[0802] Arezi,B.&Hogrefe,H.Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer.Nucleic AcidsRes 37,473-481(2009).

[0803] Blain,S.W.&Goff,S.P.Nuclease activities of Moloney murine leukemia virus reverse transcriptase.Mutants with altered substrate specificities.J.Biol.Chem.268,23585-23592(1993).

[0804] Xiong,Y.&Eickbush,T.H.Origin and evolution of retroelements based upon their reverse transcriptase sequences.EMBOJ9,3353-3362(1990).

[0805] Herschhorn,A.&Hizi,A.Retroviral reverse transcriptases.Cell.Mol.Life Sci.67,2717-2747(2010).

[0806] Taube,R.,Loya,S.,Avidan,O.,Perach,M.&Hizi,A.Reverse transcriptase of mouse mammary tumourvirus:expression inbacteria,purification and biochemical characterization.Biochem.J.329(Pt3),579-587(1998).

[0807] Liu,M.et al.Reverse Transcriptase-Mediated Tropism Switching in Bordetella Bacteriophage.Science 295,2091-2094(2002).

[0808] Luan,D.D.,Korman,M.H.,Jakubczak,J.L.&Eickbush,T.H.Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism fornon-LTR retrotransposition.Cell72,595-605(1993).

[0809] Nottingham,R.M.et al.RNA-seq of human reference RNA samples using a thermostable group IIintron reverse transcriptase.RNA 22,597-613(2016).

[0810] Telesnitsky,A.&Goff,S.P.RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer-template.Proc.Natl.Acad.Sci.U.S.A.90,1276-1280(1993).

[0811] Halvas,E.K.,Svarovskaia,E.S.&Pathak,V.K.Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity.Journal of Virology74,10349-10358 (2000).

[0812] Nowak,E.et al.Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid.Nucleic AcidsRes41,3874-3887 (2013).

[0813] Stamos,J.L.,Lentzsch,A.M.&Lambowitz,A.M.Structure ofa Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications.Molecular Cell68,926-939.e4(2017).

[0814] Das,D.&Georgiadis,M.M.The Crystal Structure of the Monomeric Reverse Transcriptase from Moloney Murine Leukemia Virus.Structure 12,819-829(2004).

[0815] Avidan,O.,Meer,M.E.,Oz,I.&Hizi,A.The processivity and fidelity of DNA synthesis exhibited by the reverse transcriptase of bovine leukemia virus.European Journal of Biochemistry269,859-867(2002).

[0816] Gerard,G.F.et al.The role of template-primer in protection of reverse transcriptase from thermal inactivation.Nucleic AcidsRes30,3118-3129(2002).

[0817] Monot,C.et al.The Specificity and Flexibility ofL1 Reverse Transcription Priming at Imperfect T-Tracts.PLOSGenetics9,e1003499(2013).

[0818] Mohr,S.et al.Thermostable group II intron reverse transcriptase fusion proteins and their use incDNA synthesis and next-generation RNA sequencing.RNA 19,958-970(2013).

[0819] 如果尚未说明,以上提及的涉及逆转录酶的任何参考文献通过引用以其整体在此并入。

[0820] [4]引导编辑器

[0821] 本公开提供了包含引导编辑器的系统。一方面,本公开提供了用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统,该系统包含第一引导编辑器复合物和第二引导编辑器复合物的,其中第一和第二引导编辑器复合物中的每个包含:(1)引导编辑器,其包含:(i)核酸可编程DNA结合蛋白(napDNAbp),和(ii)具有RNA依赖性DNA聚合酶活性的多肽;和(2)包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点的pegRNA,其中第一引导编辑器复合物的pegRNA的DNA合成模板编码第一单链DNA序列,并且第二引导编辑器复合物的pegRNA的DNA合成模板编码第二单链DNA序列,其中第一单链DNA序列和第二单链DNA序列各自包含彼此互补的区域,并且其中第一单链DNA序列和第二单链DNA序列形成双链体,该双链体与待编辑的靶位点处的DNA序列相比包含编辑部分,该编辑部分整合到待编辑的靶位点中。

[0822] 另一方面,本公开提供了用于编辑一个或多个双链DNA序列的系统,该系统包含:

[0823] a)第一引导编辑器复合物,其包含:

[0824] i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

[0825] ii.第一引导编辑向导RNA(第一PEgRNA),其在第一待编辑的靶位点处结合至第一双链DNA序列的第一链上的第一结合位点;

[0826] b)第二引导编辑器复合物,其包含:

[0827] i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

[0828] ii.第二引导编辑向导RNA(第二PEgRNA),其在第一待编辑的靶位点处结合至第一双链DNA序列的第二链上的第二结合位点;

[0829] c)第三引导编辑器复合物,其包含:

[0830] i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

[0831] ii.第三引导编辑向导RNA(第三PEgRNA),其在第二待编辑的靶位点处结合至第二双链DNA序列的第一链上的第一结合位点;

[0832] d)第四引导编辑器复合物,其包含:

[0833] i.第四引导编辑器,其包含第四核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

[0834] ii.第四引导编辑向导RNA(第四PEgRNA),其在第二待编辑的靶位点处结合至第二双链DNA序列的第二链上的第二结合位点;

[0835] 其中第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板,第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板,第三PEgRNA包含编码第三单链DNA序列的第三

DNA合成模板，并且第四PEgRNA包含编码第四单链DNA序列的第四DNA合成模板；

[0836] 其中第一和第三单链DNA序列各自包含彼此互补的区域；并且其中，其中第二和第四单链DNA序列各自包含彼此互补的区域

[0837] 在一些实施方案中，本文所述系统中使用的引导编辑器(例如，上面所述系统中的第一引导编辑器、第二引导编辑器、第三引导编辑器和/或第四引导编辑器)作为融合蛋白提供。在一些实施方案中，引导编辑器融合蛋白包含napDNAbp和聚合酶(例如，DNA依赖性DNA聚合酶或RNA依赖性DNA聚合酶，比如，逆转录酶等)。在一些实施方案中，napDNAbp和聚合酶任选地由接头连接形成融合蛋白。不同构造的引导编辑器融合蛋白和以及融合蛋白的其他结构域在本文中进一步描述。在不同实施方案中，本公开提供的系统和方法设想了使用本文所述的任何引导编辑器融合蛋白。

[0838] 在一些实施方案中，本文所述系统中使用引导编辑器复合物包含引导编辑器(例如，上文所述系统中的第一引导编辑器，第二引导编辑器，第三引导编辑器，和/或第四引导编辑器)，其中一个或多个引导编辑器中的组分以反式提供，如同在本说明书通篇中额外详细描述的。在一些实施方案中，引导编辑器包含以反式表达的napDNAbp和聚合酶。在一些实施方案中，napDNAbp和聚合酶从一种或多种载体表达(例如，两种组分从同一种载体表达，或者每种组分从不同载体表达)。在某些实施方案，引导编辑器包如本文所述的其他组分，这些组分以反式表达。在一些实施方案中，本文所述系统中使用的引导编辑器可以包含作为融合蛋白提供的一种或多种引导编辑器和其组分以反式提供的一种或多种引导编辑器。

[0839] 在一些实施方案中，多瓣引导编辑器系统考虑包含napDNAbp和聚合酶(如，DNA依赖性DNA聚合酶或RNA依赖性DNA聚合酶，如逆转录酶)并且任选地通过接头连接的融合蛋白。本申请考虑将任何合适的napDNAbp和聚合酶(如，DNA依赖性DNA聚合酶或RNA依赖性DNA聚合酶，如逆转录酶)组合在单个融合蛋白中。napDNAbps和聚合酶(如，DNA依赖性DNA聚合酶或RNA依赖性DNA聚合酶，如逆转录酶)的实例分别在本文中定义。由于聚合酶是本领域公知的，并且氨基酸序列易于获得，因此本公开并不意味着以任何方式限于本文鉴定的那些特定聚合酶。

[0840] 在不同实施方案中，融合蛋白可包含任何合适的结构构造。例如，融合蛋白可从N-端至C-端方向包含与聚合酶(如，DNA依赖性DNA聚合酶或RNA依赖性DNA聚合酶，如逆转录酶)融合的napDNAbp。在其他实施方案中，融合蛋白可从N-端至C-端方向包含与napDNAbp融合的聚合酶(如，逆转录酶)。融合结构域可任选地通过接头连接，例如氨基酸序列。在其他实施方案中，融合蛋白可包含结构NH_2-[napDNAbp]-[聚合酶]-COOH；或NH_2-[聚合酶]-[napDNAbp]-COOH，其中“]-[”的每个情况表示存在任选的接头序列。在聚合酶是逆转录酶的实施方案中，融合蛋白可包含结构NH_2-[napDNAbp]-[RT]-COOH；或NH_2-[RT]-[napDNAbp]-COOH，其中“]-[”的每个情况表示存在任选的接头序列。

[0841] 示例性融合蛋白描述于图14中，显示包含通过接头序列与切口酶Cas9(“Cas9(H840A)”)融合的MLV逆转录酶(“MLV-RT”)的融合蛋白。本示例不旨在限制可用于本文所述的引导编辑器(PE)系统的融合蛋白的范围。

[0842] 在不同实施方案中，多瓣引导编辑器融合蛋白可具有以下氨基酸序列(本文称为“PE1”)，其包括含有H840A突变的Cas9变体(即，Cas9切口酶)和M-MLV RT野生型，以及N端NLS序列(19个氨基酸)和将Cas9切口酶域的C端连接到RT域的N端的氨基酸接头(32个氨基

酸)。PE1融合蛋白具有以下结构:[NLS]-[Cas9(H840A)]-[接头]-[MMLV_RT(wt)]。PE1及其各个组件的氨基酸序列如下:

[0843]

描述	序列
PE1融合蛋白 CAS9(H840A)-MMLV_RT(WT)	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT

[0844]

描述	序列
	NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD<u>**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**</u>*TLNIEDEYRLH ETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQV KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPG FAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKAL FLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSS P**SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 123) 解释: 核定位序列(NLS) 开端:(SEQ ID NO: 124), 末端: (SEQ ID NO: 133) **CAS9(H840A)** (SEQ ID NO: 126) **33-氨基酸接头** (SEQ ID NO: 127) *M-MLV 逆转录酶*(SEQ ID NO: 128)
PE1 – N-端 NLS	MKRTADGSEFESPKKKRKV (SEQ ID NO: 124)
PE1 – CAS9 (H840A) (减 MET)	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT

[0845]

描述	序列
	NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD (SEQ ID NO: 130)
PE1 –CAS9 结构域与 RT 结构域之间的接头(33 个氨基酸)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 127)
PE1 – M-MLV RT	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 132)
PE1 – C-端 NLS	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 133)

[0846] 在另一个实施方案中，多瓣引导编辑器融合蛋白可具有以下氨基酸序列(本文称为“PE2”)，其包括含有H840A突变的Cas9变体(即，Cas9切口酶)和包含突变D200N、T330P、L603W、T306K和W313F的M-MLV RT，以及N端NLS序列(19个氨基酸)和连接Cas9切口酶结构域C端至RT结构域的N端的氨基酸接头(33个氨基酸)。PE2融合蛋白具有以下结构：[NLS]-[Cas9(H840A)]-[接头]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)]。PE2的氨基酸序列如下：

[0847]

PE2 融合蛋白 CAS9(H840A)-MMLV_RT D200N T330P L603W T306K W313F	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM**

[0848]

	IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLH ETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQV KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPG FAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKA LFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENS SP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 134) 解释: 核定位序列(NLS) 开端:(SEQ ID NO: 124), 末端: (SEQ ID NO: 133) **CAS9(H840A) (SEQ ID NO: 137) **33-氨基酸接头** (SEQ ID NO: 127) *M-MLV 逆转录酶*(SEQ ID NO: 139)
PE2 – N-端 NLS	MKRTADGSEFESPKKKRKV (SEQ ID NO: 124)
PE2 – CAS9 (H840A) (减 MET)	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY

[0849]

	AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD (SEQ ID NO: 141)
PE2 –CAS9 结构域和 RT 结构域之间的接头(33 个氨基酸)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 127)
PE2 – MMLV_RT D200N T330P L603W T306K W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 143)
PE2 – C-端 NLS	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 133)

[0850] 在进一步的其他实施方案中，引导编辑器融合蛋白可具有以下氨基酸序列：

[0851]

PE 融合蛋白 MMLV_RT(WT)-32AA-CAS9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYN LLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELD CQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK ETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE*

[0852]

		*GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAH IHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSP***SGGSSGGSSGSETPGTSES ATPESSGGSSGGSS*****DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL QNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD*****SGGSKRTADGSEFEPKKKRKV*** (SEQ ID NO: 145) 解释: 核定位序列(NLS) 开端:(SEQ ID NO: 124), 末端: (SEQ ID NO: 133) **CAS9(H840A)** (SEQ ID NO: 147) **33-氨基酸接头** (SEQ ID NO: 127) *M-MLV 逆转录酶* (SEQ ID NO: 149)
PE 融合蛋白 MMLV_RT(WT)-60AA-CAS9(H840A)		MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYN LLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELD CQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK ETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAA*

[0853]

		*VTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAH IHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSPS***GGSSGGSSGSETPGTSES ATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSSGGS*****DKKYS IGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE VLDATLIHQSITGLYETRIDLSQLGGD*SGGSKRTADGSEFEPKKKRKV* (SEQ ID NO: 150) 解释: 核定位序列(NLS)开端:(SEQ ID NO: 124), 末端: (SEQ ID NO: 133) **CAS9(H840A)**(SEQ ID NO: 153) ***氨基酸接头***(SEQ ID NO: 175) *M-MLV 逆转录酶*(SEQ ID NO: 149)
PE 融合蛋白 CAS9(H840A)-FEN1-MMLV_RT D200N T330P L603W T306K W313F		MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE**

[0854]

**RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSGIQGLAKLIA DVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQNEEGETT SHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAE KQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDA PSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLP IQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHK SIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPNE EELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDFFKVTGSLSS AKRKEPEPKGSTKKKAKTGAAGKFKRGKSGGSSGGSSGSETPGTSES ATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETG GMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPC QSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPP SHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRA LLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQP TPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALV KQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHN CLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTET EVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGN RMADQAARKAAITETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV
(SEQ ID NO: 154)
解释:
核定位序列(NLS) 开端:(SEQ ID NO: 124), 末端: (SEQ ID NO: 133)
CAS9(H840A) (SEQ ID NO: 157)
33-氨基酸接头 1 (SEQ ID NO: 127)
M-MLV 逆转录酶(SEQ ID NO: 159)
33-氨基酸接头 2 (SEQ ID NO: 127)
FEN1 (SEQ ID NO: 161)

[0855] 在其他实施方案中,多瓣引导编辑器融合蛋白可基于具有改变的PAM特异性的SaCas9或SpCas9切口酶,如以下示例的序列:

[0856]

SACAS9-M-MLV RT 引导编辑器	MKRTADGSEFESPKKKRKVGKRNYILGLDIGITSVGYGIIDYETR DVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKK LLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLA KRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLE RLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFID TYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPE ELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIEN VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYH DIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQE EIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG LPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGK ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII PRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETF KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFK KERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMF EEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKK PNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLIN KSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGN YLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYE EAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNL YEVKSKKHPQIIKKGSGGSSGGSSGSETPGTSESATPESSGGSSG GSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGL AVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR DPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDL ILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQ ICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREF LGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRR PVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAEL IALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSE GKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKR KV (SEQ ID NO: 162)
SPCAS9(H840A)-VRQR-莫洛尼鼠白血病病毒逆转酶引导编辑器	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL

[0857]

	GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK KDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESAT PESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLAD FRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYR ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPD QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKL TMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGT SAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSE FEPKKKRKV (SEQ ID NO: 163)
SPCAS9(H840A)-VRER-莫洛尼鼠白血病病毒逆转酶引导编辑器	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI

[0858]

	LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK KDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVL DATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESAT PESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLAD FRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYR ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPD QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKL TMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGT SAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSE FEPKKKRKV (SEQ ID NO: 164)

[0859] 在其他实施方案中，本文考虑的多瓣引导编辑器融合蛋白可包括与M-MLV逆转录酶的截短版本融合的Cas9切口酶(如，Cas9(H840A))。在该实施方案中，逆转录酶还包含4个突变(D200N、T306K、W313F、T330P；注意PE2中存在的L603W突变由于截短而不再存在)。编码这种截短编辑器的DNA序列比PE2更小522bp，因此它可能适用于DNA序列的递送因其大小而具有挑战性的应用(即，腺相关病毒和慢病毒递送)。该实施方案被称为Cas9(H840A)-MMLV-RT(截短)或“PE2-短”或“PE2-截短”，并具有以下氨基酸序列：

[0860]

CAS9(H840A)-MMLV-RT(截短)或PE2-短	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGD**<u>SGGSSGGSSGSETPGTSE SATPESSGGSSGGSS</u>*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVE DIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA FEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQH PDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQI CQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGK AGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEA LVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLP EEGLQHNCLDNSRLIN**<u>SGGSKRTADGSEFEPKKKRKV</u> (SEQ ID NO: 765) 解释: <u>核定位序列(NLS)</u> 开端:(SEQ ID NO: 124), 末端: (SEQ ID NO: 133)

[0861]

	CAS9(H840A) (SEQ ID NO: 157) **<u>33-氨基酸接头 1</u>** (SEQ ID NO: 127) *M-MLV 截短的逆转录酶*(SEQ ID NO: 766)

[0862] 图75提供了比较PE2、PE2-截短、PE3和PE3-截短在不同细胞系中不同靶位点的效

率(即“具有指定编辑或indel的总测序读段的百分比%”)的柱状图。数据显示包含截短的RT变体的引导编辑器与包含非截短的RT蛋白的引导编辑器的效率差不多。

[0863] 在不同实施方案中,本文考虑的多瓣引导编辑器融合蛋白还可包括以上公开的序列的任何变体,其具有与PE1、PE2或任何以上显示的引导编辑器融合序列至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同、或至少约99.9%相同的氨基酸序列。

[0864] 在某些实施方案中,可将接头用于连接本发明的任何肽或肽结构域或部分(如,与逆转录酶连接或融合的napDNAbp)。

[0865] [5]接头和其他结构域

[0866] 除了napDNAbp(如,Cas9结构域)和聚合酶结构域(如,RT结构域)外,PE融合蛋白可包含各种其他结构域。例如,在napDNAbp是Cas9并且聚合酶是RT的情况下,PE融合蛋白可包含一个或多个连接Cas9结构域和RT结构域的接头。接头还可将其他功能结构域,例如核定位序列(NLS)或FEN1(或其他瓣核酸内切酶)连接到PE融合蛋白或其结构域。

[0867] 此外,在涉及反式引导编辑的实施方案中,接头可用于将tPERT募集蛋白连接到引导编辑器,如在tPERt募集蛋白和napDNAbp之间。参见,例如图3G,反式引导编辑器(tPE)的示例性示意图,其包括将聚合酶结构域和募集蛋白结构域分别融合到napDNAbp的接头。

[0868] A.接头

[0869] 如以上所定义,如本文所用,术语“接头”是指连接两个分子或部分(如,核酸酶的结合结构域和切割结构域)的化学基团或分子。在一些实施方案中,接头连接RNA可编程核酸酶的gRNA结合结构域和聚合酶(如,逆转录酶)的催化结构域。在一些实施方案中,接头连接dCas9和逆转录酶。通常,接头位于两个基团、分子或其他部分之间或两侧,并通过共价键将其彼此连接,从而将二者连接起来。在一些实施方案中,接头是一个氨基酸或多个氨基酸(如,肽或蛋白)。在一些实施方案中,接头是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头的长度为5-100个氨基酸,例如,长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。还考虑更长或更短的接头。

[0870] 接头的长度可以与共价键一样简单,或者可以为许多原子的聚合接头。在某些实施方案中,接头是多肽或基于氨基酸。在其他实施方案中,接头不是肽样的。在某些实施方案中,接头是共价键(如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,接头是酰胺键的碳-氮键。在某些实施方案中,接头是环状或非环状、取代或未取代、支链或无支链的脂族或杂脂族接头。在某些实施方案中,接头是聚合的(如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,接头包含氨基链烷酸的单体、二聚体或聚合物。在某些实施方案中,接头包含氨基链烷酸(如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在某些实施方案中,接头包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,接头基于碳环部分(如,环戊烷、环己烷)。在其他实施方案中,接头包含聚乙二醇部分(PEG)。在其他实施方案中,接头包含氨基酸。在某些实施方案中,接头包含肽。在某些实施方案中,接头包含芳基或杂芳基部分。在某些实施方案中,接头基于苯环。接头可包括功能化部分以促进亲核试剂(如,硫醇、氨基)从肽连接到接头。任何亲电子试剂都可用作接头的一部分。示例性亲电子试剂包括但不限于活化酯、活化酰胺、迈克尔(Michael)受体、卤

代烷、芳基卤、酰卤和异硫氰酸酯。

[0871] 在一些其他实施方案中，接头包含氨基酸序列(GGGGS)n(SEQ ID NO:165)、(G)n(SEQ ID NO:166)、(EAAAK)n(SEQ ID NO:167)、(GGS)n(SEQ ID NO:168)、(SGGS)n(SEQ ID NO:169)、(XP)n(SEQ ID NO:170)或其任何组合，其中n独立地为1至30的整数，X为任何氨基酸。在一些实施方案中，接头包含氨基酸序列(GGS)N(SEQ ID NO:176)，其中n为1、3或7。在一些实施方案中，接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:171)。在一些实施方案中，接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:172)。在一些实施方案中，接头包含氨基酸序列SGGSGGSGGS(SEQ ID NO:173)。在一些实施方案中，接头包含氨基酸序列SGGS(SEQ ID NO:174)。在其他实施方案中，接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESAGSYPYDV PDYAGSAAPAAKKKKLDGSGSGGSSGGS(SEQ ID NO:175,60AA)。

[0872] 在某些实施方案中，接头可用于连接本发明的任何肽或肽结构域或部分(如，与逆转录酶连接或融合的napDNAbp)。

[0873] 如以上所定义，如本文所用，术语“接头”是指连接两个分子或部分(如，核酸酶的结合结构域和切割结构域)的化学基团或分子。在一些实施方案中，接头连接RNA可编程核酸酶的gRNA结合结构域和重组酶的催化域。在一些实施方案中，接头连接dCas9和逆转录酶。通常，接头位于两个基团、分子或其他部分之间或两侧，并通过共价键彼此连接，从而将二者连接起来。在一些实施方案中，接头是一个氨基酸或多个氨基酸(如，肽或蛋白)。在一些实施方案中，接头是有机分子、基团、聚合物或化学部分。在一些实施方案中，接头的长度为5-100个氨基酸，例如，长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。还考虑更长或更短的接头。

[0874] 接头的长度可以与共价键一样简单，或者可以为许多原子的聚合接头。在某些实施方案中，接头是多肽或基于氨基酸。在其他实施方案中，接头不是肽样的。在某些实施方案中，接头是共价键(如，碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中，接头是酰胺键的碳-氮键。在某些实施方案中，接头是环状或非环状、取代或未取代、支链或无支链的脂族或杂脂族接头。在某些实施方案中，接头是聚合的(如，聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中，接头包含氨基链烷酸的单体、二聚体或聚合物。在某些实施方案中，接头包含氨基链烷酸(如，甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在某些实施方案中，接头包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中，接头基于碳环部分(如，环戊烷、环己烷)。在其他实施方案中，接头包含聚乙二醇部分(PEG)。在其他实施方案中，接头包含氨基酸。在某些实施方案中，接头包含肽。在某些实施方案中，接头包含芳基或杂芳基部分。在某些实施方案中，接头基于苯环。接头可包括功能化部分以促进亲核试剂(如，硫醇、氨基)从肽连接到接头。任何亲电子试剂都可用作接头的一部分。示例性的亲电子试剂包括但不限于活化酯、活化酰胺、迈克尔受体、卤代烷、芳基卤、酰基卤和异硫氰酸酯。

[0875] 在一些其他实施方案中，接头包含氨基酸序列(GGGGS)n(SEQ ID NO:165)、(G)n(SEQ ID NO:166)、(EAAAK)n(SEQ ID NO:167)、(GGS)n(SEQ ID NO:168)、(SGGS)n(SEQ ID NO:169)、(XP)n(SEQ ID NO:170)或其任何组合，其中n独立地为1至30的整数，X为任何氨基酸。在一些实施方案中，接头包含氨基酸序列(GGS)N(SEQ ID NO:176)，其中n为1、3或7。在

一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:171)。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:172)。在一些实施方案中,接头包含氨基酸序列SGGSGGSGGS(SEQ ID NO:173)。在一些实施方案中,接头包含氨基酸序列SGGS(SEQ ID NO:174)。

[0876] 特别地,可在不同实施方案中使用以下接头来将引导编辑器结构域彼此连接:

[0877] GGS(SEQ ID NO:767);

[0878] GGSGGS(SEQ ID NO:768);

[0879] GGSGGSGGS(SEQ ID NO:769);

[0880] SGGSSGGSSGSETPGTSESATPESSGGSSGGSS(SEQ ID NO:127)

[0881] SGSETPGTSESATPES(SEQ ID NO:171);

[0882] SGGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSG GSSGGS(SEQ ID NO:175)。

[0883] B.核定位序列(NLS)

[0884] 在不同实施方案中,PE融合蛋白可包含一种或多种核定位序列(NLS),其有助于促进蛋白易位至细胞核中。这类序列是本领域所熟知的并且可包括以下示例:

[0885]

描述	序列	SEQ ID NO:
SV40 较大标签的 NLS	PKKKRKV	SEQ ID NO: 16
NLS	MKRTADGSEFESPKKKRKV	SEQ ID NO: 124
NLS	MDSLLMNRRKFLYQFKNVRWAKGRRETYL C	SEQ ID NO: 17
核质蛋白的 NLS	AVKRPAATKKAGQAKKKKLD	SEQ ID NO: 190
EGL-13 的 NLS	MSRRRKANPTKLSENAKKLAKEVEN	SEQ ID NO: 191
C-MYC 的 NLS	PAAKRVKLD	SEQ ID NO: 192
TUS-蛋白的 NLS	KLKIKRPVK	SEQ ID NO: 193
多瘤的较大标签的 NLS	VSRKRPRP	SEQ ID NO: 194

[0886]

丁型肝炎病毒抗原的 NLS	EGAPPAKRAR	SEQ ID NO: 195
鼠 P53 的 NLS	PPQPKKKPLDGE	SEQ ID NO: 196
PE1 和 PE2 的 NLS	SGGSKRTADGSEFEPKKKRKV	SEQ ID NO: 133

[0887] 以上NLS示例是非限制性的。PE融合蛋白可包含任何已知的NLS序列,包括描述于Cokol et al.,“Finding nuclear localization signals,”EMBO Rep.,2000,1(5):411-415andFreitas et al.,“Mechanisms and Signals forthe Nuclear Import ofProteins,”Current Genomics,2009,10(8):550-7中的那些中的任何一种,其各自通过引用并入本文。

[0888] 在不同实施方案中,本文公开的多瓣引导编辑器和编码引导编辑器的构建体进一步包含一个或多个,优选至少两个核定位信号。在某些实施方案中,多瓣引导编辑器包含至少两个NLS。在具有至少两个NLS的实施方案中,NLS可以是相同的NLS或者可以是不同的NLS。此外,NLS可作为融合蛋白的一部分与引导编辑器的其余部分一起表达。在一些实施方案中,一个或多个NLS是双组分NLS(“bpNLS”)。在某些实施方案中,所公开的融合蛋白包含

两个双组分NLS。在一些实施方案中,所公开的融合蛋白包含两个以上的双组分NLS。

[0889] NLS融合的位置可以在N-端、C-端,或在引导编辑器的序列内(如,在编码的napDNAbp组分(如,Cas9)和聚合酶结构域(如,逆转录酶域)之间插入)。

[0890] NLS可以是本领域中任何已知的NLS序列。NLS也可以是任何未来发现的用于核定位的NLS。NLS还可以是任何天然存在的NLS,或任何非天然存在的NLS(如,具有一个或多个期望突变的NLS)。

[0891] 术语"核定位序列"或"NLS"是指例如通过核转运促进蛋白输入细胞核的氨基酸序列。核定位序列是本领域已知的并且对技术人员而言是显而易见的。例如,NLS序列描述于Plank等于2000年11月23日提交的国际PCT申请PCT/EP2000/011690,2001年5月31日以WO/2001/038547公开,其内容通过引用并入本文。在一些实施方案中,NLS包含氨基酸序列PKKKRKV(SEQ ID NO:16)、MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:17)、KRTADGSEFESPKKKRKV(SEQ ID NO:3864)或KRTADGSEFEPKKKRKV(SEQ ID NO:125)。在其他实施方案中,NLS包含氨基酸序列NLSKRPAAIKK AGQAKKKK(SEQ ID NO:136)、PAAKRVKLD(SEQ ID NO:192)、RQRRNELKRSF(SEQ ID NO:3934)、NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO:3935)。

[0892] 在本公开的一方面,多瓣引导编辑器用一个或多个核定位信号(NLS),优选至少两个NLS修饰。在某些实施方案中,多瓣引导编辑器用两个或更多个NLS修饰。本公开考虑使用在公开时本领域已知的任何核定位信号,或在本申请提交时间之后在现有技术中被确认或以其他方式可用的任何核定位信号。代表性的核定位信号是将蛋白引导至表达该序列的细胞核的肽序列。核定位信号主要是碱性的,几乎可位于蛋白氨基酸序列的任何位置,通常包含4个氨基酸的短序列(Autieri&Agrawal,(1998)J.Biol.Chem.273:14731-37,通过引用并入本文)至8个氨基酸,并且通常富含赖氨酸和精氨酸残基(Magin etal.,(2000)Virology 274:11-16,通过引用并入本文)。核定位信号通常包含脯氨酸残基。已鉴定多种核定位信号并已用于影响生物分子从细胞质到细胞核的转运。参见,例如,Tinland etal.,(1992)Proc.Natl.Acad.Sci.U.S.A.89:7442-46;Moede etal.,(1999)FEBSLett.461:229-34,其通过引用并入。目前认为易位涉及核孔蛋白。

[0893] 大多数NLS可分为三类:(i)单组分NLS,如SV40较大T抗原NLS(PKKKRKV(SEQ ID NO:16));(ii)由被不同数量的间隔区氨基酸分开的两个基本结构域组成的二组分基序,以爪蟾核质蛋白NLS(KRXXXXXXXXXXXKKKL(SEQ ID NO:3936))为例;(iii)非经典序列,如hnRNPA1蛋白的M9、流感病毒核蛋白NLS和酵母Gal4蛋白NLS(Dingwall and Laskey 1991)。

[0894] 核定位信号出现在蛋白的氨基酸序列的不同点。NLS已在N端、C端和蛋白的中心区域处鉴定。因此,本公开提供了可在多瓣引导编辑器的C-端、N-端及内部区域处用一个或多个NLS修饰的多瓣引导编辑器。应选择不作为组分NLS残基发挥作用的较长序列的残基,以免如在张力或空间上干扰核定位信号本身。因此,虽然对包含NLS的序列的组成没有严格限制,但实际上,这样的序列在长度和组成上受功能限制。

[0895] 本公开考虑了任何合适的手段,通过该手段将多瓣引导编辑器修饰为包括一个或多个NLS。一方面,可将多瓣引导编辑器设计成表达在其N端或C端(或两者)翻译融合一个或多个NLS的引导编辑器蛋白,即形成引导编辑器-NLS融合构建体。在其他实施方案中,编码引导编辑器的核苷酸序列可经遗传修饰以在编码的引导编辑器的内部区域并入编码一个

或多个NLS的阅读框。此外，NLS可包括在引导编辑器和N-端、C-端或内部连接的NLS氨基酸序列之间(如，在蛋白的中心区域)编码的各种氨基酸接头或间隔区。因此，本公开还提供了用于表达融合蛋白的核苷酸构建体、载体和宿主细胞，所述融合蛋白包含引导编辑器和一个或多个NLS。

[0896] 本文所述的多瓣引导编辑器还可包含核定位信号，其通过一个或多个接头连接至引导编辑器，例如聚合物、氨基酸、核酸、多糖、化学或核酸接头元件。本公开考虑范围内的接头不意图具有任何限制，可以是任何合适类型的分子(如，聚合物、氨基酸、多糖、核酸、脂质或任何合成的化学接头结构域)，并且通过实现在引导编辑器和一个或多个NLS之间形成键(如，共价键、氢键)的任何合适策略被连接至引导编辑器。

[0897] C.瓣核酸内切酶(如，FEN1)

[0898] 在不同实施方案中，PE融合蛋白可包含一种或多种瓣核酸内切酶(如，FEN1)，其是指催化去除5′单链DNA瓣的酶。这些是天然存在的酶，用于去除细胞过程包括DNA复制中形成的5′瓣。本文所述多瓣引导编辑方法可利用内源性提供的瓣核酸内切酶或以反式提供的那些来去除多瓣引导编辑期间靶位点形成的内源性DNA的5'瓣。瓣核酸内切酶是本领域已知的，可见描述于Patel et al.,“Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends,”Nucleic Acids Research,2012,40(10):4507-4519和Tsutakawa et al.,“Human flap endonuclease structures,DNA double-base flipping,and a unified understanding ofthe FEN1 superfamily,”Cell,2011,145(2):198-211(其各自通过引用并入本文)。示例性瓣核酸内切酶是FEN1，其可由以下氨基酸序列表示：

[0899]

描述	序列	SEQ ID NO:
FEN1 野生型 (wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD	SEQ ID NO: 198

[0900]

	DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	

[0901] 瓣核酸内切酶还可包括任何FEN1变体、突变体或其他瓣核酸内切酶直系同源物、同源物或变体。非限制性FEN1变体示例如下所示：

[0902]

描述	序列	SEQ ID NO:
FEN1 K168R (相对于 FEN1 wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALV**R**AGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 199
FEN1 S187A (相对于 FEN1 wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFG **A**PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFV DLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLD PNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEP NEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRL DDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFK RGK	SEQ ID NO: 200
FEN1 K354R (相对于 FEN1 wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSA**R**RKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 201
GEN1	MGVNDLWQILEPVKQHIPLRNLGGKTIAVDLSLWVCE AQTVKKMMGSVMKPHLRNLFFRISYLTQMDVKLVFV MEGEPPKLKADVISKRNQSRYGSSGKSWSQKTGRSHF KSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGG HVDGCLTNDGDTFLYGAQTVYRNFTMNTKDPHVDCY TMSSIKSKLGLDRDALVGLAILLGCDYLPKGVPGVGKE QALKLIQILKGQSLLQRFNRWNETSCNSSPQLLVTKKL AHCSVCSHPGSPKDHERNGCRLCKSDKYCEPHDYEYC CPCEWHRTEHDRQLSEVENNIKKKACCCEGFPFHEVIQ EFLLNKDKLVKVIRYQRPDLLLFQRFTLEKMEWPNHY ACEKLLVLLTHYDMIERKLGSRNSNQLQPIRIVKTRIRN GVHCFEIEWEKPEHYAMEDKQHGEFALLTIEEESLFEA	SEQ ID NO: 202

		AYPEIVAVYQKQKLEIKGKKQKRIKPKENNLPEPDEVM SFQSHMTLKPTCEIFHKQNSKLNSGISPDPTLPQESISAS LNSLLLPKNTPCLNAQEQFMSSLRPLAIQQIKAVSKSLI SESSQPNTSSHNISVIADLHLSTIDWEGTSFSNSPAIQRN TFSHDLKSEVESELSAIPDGFENIPEQLSCESERYTANIK KVLDEDSDGISPEEHLLSGITDLCLQDLPLKERIFTKLSY PQDNLQPDVNLKTLSILSVKESCIANSGSDCTSHLSKDL PGIPLQNESRDSKILKGDQLLQEDYKVNTSVPYSVSNT VVKTCNVRPPNTALDHSRKVDMQTTRKILMKKSVCLD RHSSDEQSAPVFGKAKYTTQRMKHSSQKHNSSHFKES GHNKLSSPKIHIKETEQCVRSYETAENEESCFPDSTKSS LSSLQCHKKENNSGTCLDSPLPLRQRLKLRFQST	
[0903]	ERCC5	MGVQGLWKLLECSGRQVSPEALEGKILAVDISIWLNQ ALKGVRDRHGNSIENPHLLTLFHRLCKLLFFRIRPIFVF DGDAPLLKKQTLVKRRQRKDLASSDSRKTTEKLLKTF LKRQAIKTAFRSKRDEALPSLTQVRRENDLYVLPPLQE EEKHSSEEEDEKEWQERMNQKQALQEEFFHNPQAIDIE SEDFSSLPPEVKHEILTDMKEFTKRRRTLFEAMPEESDD FSQYQLKGLLKKNYLNQHIEHVQKEMNQQHSGHIRRQ YEDEGGFLKEVESRRVVSEDTSHYILIKGIQAKTVAEV DSESLPSSSKMHGMSFDVKSSPCEKLKTEKEPDATPPSP RTLLAMQAALLGSSSEEELESENRRQARGRNAPAAVD EGSISPRTLSAIKRALDDDEDVKVCAGDDVQTGGPGAE EMRINSSTENSDEGLKVRDGKGIPFTATLASSSVNSAEE HVASTNEGREPTDSVPKEQMSLVHVGTEAFPISDESMI KDRKDRLPLESAVVRHSDAPGLPNGRELTPASPTCTNS VSKNETHAEVLEQQNELCPYESKFDSSLLSSDDETKCK PNSASEVIGPVSLQETSSIVSVPSEAVDNVENVVSFNAK EHENFLETIQEQQTTESAGQDLISIPKAVEPMEIDSEESE SDGSFIEVQSVISDEELQAEFPETSKPPSEQGEEELVGTR EGEAPAESESLLRDNSERDDVDGEPQEAEKDAEDSLHE WQDINLEELETLESNLLAQQNSLKAQKQQQERIAATVT GQMFLESQELLRLFGIPYIQAPMEAEAQCAILDLTDQTS GTITDDSDIWLFGARHVYRNFFNKNKFVEYYQYVDFH NQLGLDRNKLINLAYLLGSDYTEGIPTVGCVTAMEILN EFPGHGLEPLLKFSEWWHEAQKNPKIRPNPHDTKVKK KLRTLQLTPGFPNPAVAEAYLKPVVDDSKGSFLWGKP DLDKIREFCQRYFGWNRTKTDESLFPVLKQLDAQQTQ LRIDSFFRLAQQEKEDAKRIKSQRLNRAVTCMLRKEKE AAASEIEAVSVAMEKEFELLDKAKRKTQKRGITNTLEE SSSLKRKRLSDSKRKNTCGGFLGETCLSESSDGSSSEDA ESSSLMNVQRRTAAKEPKTSASDSQNSVKEAPVKNGG ATTSSSSDSDDDGGKEKMVLVTARSVFGKKRRKLRRA RGRKRKT	SEQ ID NO: 203

[0904] 在不同实施方案中,本文考虑的多瓣引导编辑器融合蛋白可包括与上述任何序列具有至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同、或至少约99.9%相同的氨基酸序列。

[0905] 可由本方法利用以促进去除5'端单链DNA瓣的其他核酸内切酶包括但不限于(1)trex 2,(2)exo1核酸内切酶(如,Keijzers et al.,BiosciRep.2015,35(3):e00206)。

[0906] Trex2

[0907] 3'三原初(prime)修复核酸外切酶2(TREX2)-人

[0908] 登录号NM_080701

[0909] MSEAPRAETFVFLDLEATGLPSVEPEIAELSLFAVHRSSLENPEHDESGALVLPRVLDKLTLCMCPER PFTAKASEITGLSSEGLARCRKAGFDGAVVRTLQAFLSRQAGPICLVAHNGFDYDFPLLCAELRRLGARLPRDTVC LDTLPALRGLDRAHSHGTRARGRQGYSLGSLFHRYFRAEPSAAHSAEGDVHTLLLIFLHRAAELLAWADEQARGWA HIEPMYLPPDDPSLEA(SEQ ID NO:3865)

[0910] 3'三原初修复核酸外切酶2(TREX2)-小鼠

[0911] 登录号NM_011907

[0912] MSEPPRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLPRVLDKLTLCMCPER PFTAKASEITGLSSESLMHCGKAGFNGAVVRTLQGFLSRQEGPICLVAHNGFDYDFPLLCTELQRLGAHLPQDTVC LDTLPALRGLDRAHSHGTRAQGRKSYSLASLFHRYFQAEPSAAHSAEGDVHTLLLIFLHRAPELLAWADEQARSWA HIEPMYVPPDGPSLEA(SEQ ID NO:3866)

[0913] 3'三原初修复核酸外切酶2(TREX2)-大鼠

[0914] 登录号NM_001107580

[0915] MSEPLRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLPRVLDKLTLCMCPER PFTAKASEITGLSSEGLMNCRKAAFNDAVVRTLQGFLSRQEGPICLVAHNGFDYDFPLLCTELQRLGAHLPRDTVC LDTLPALRGLDRVHSHGTRAQGRKSYSLASLFHRYFQAEPSAAHSAEGDVNTLLLIFLHRAPELLAWADEQARSWA HIEPMYVPPDGPSLEA(SEQ ID NO:3867)

[0916] ExoI

[0917] 人核酸外切酶1(EXO1)涉及许多不同的DNA代谢过程,包括DNA错配修复(MMR)、微介导的末端连接、同源重组(HR)和复制。人EXO1属于真核核酸酶Rad2/XPG家族,其还包括FEN1和GEN1。Rad2/XPG家族在从噬菌体到人类的物种的核酸酶结构域中都是保守的。EXO1基因产物表现出5'核酸外切酶和5'瓣活性。此外,EXO1包含固有的5'RNase H活性。人EXO1对加工双链DNA(dsDNA)、缺口、间隙、假Y结构具有高亲和力,并且可使用其遗传的瓣活性分解霍利迪(Holliday)连结体。人EXO1与MMR相关,并包含与MLH1和MSH2直接相互作用的保守结合结构域。PCNA、MutSα(MSH2/MSH6复合物)、14-3-3、MRN和9-1-1复合物可积极刺激EXO1溶核活性。

[0918] 核酸外切酶1(EXO1)登录号NM_003686(智人核酸外切酶1(EXO1),转录本变体3)-同种型A

[0919] MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYVGFCMKFVNMLLS HGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVSEARECFTRSINITHAMAHKVIKAARSQGVDC LVAPYEADAQLAYLNKAGIVQAIITEDSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCI LSGCDYLSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKRKLI PLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRSHSWDDKTCQKSANVSSIWH RNYSPRPESGTVSDAPQLKENPSTVGVERVISTKGLNLPRKSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSE GNKSLSFSEVFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCVSNKVSIQP LDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKFTRTISPP TLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQS QESGEFSLQSSNASKLSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLST TKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKF(SEQ ID NO:3868)

[0920] 核酸外切酶1(EXO1)登录号NM_006027(智人核酸外切酶1(EXO1),转录本变体3)-同种型B

[0921] MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVSEARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKRKLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVGVERVISTKGLNLPRKSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSEVFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCVSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSNASKLSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLSTTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKDSEKLPPCKKPLSPVRDNIQLTPEAEEDIFNKPECGRVQRAIFQ(SEQ ID NO:3869)

[0922] 核酸外切酶1(EXO1)登录号NM_001319224(智人核酸外切酶1(EXO1),转录本变体4)-同种型C

[0923] MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVSEARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKRKLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVGVERVISTKGLNLPRKSSIVKRPRSELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSEVFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCVSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSNASKLSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLSTTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKDSEKLPPCKKPLSPVRDNIQLTPEAEEDIFNKPECGRVQRAIFQ(SEQ ID NO:3870)

[0924] D.内含肽和断裂内含肽

[0925] 应当理解,在一些实施方案中(如,使用AAV颗粒在体内递送多瓣引导编辑器),可能有利的是将多肽(如,脱氨酶或napDNAbp)或融合蛋白(如,多瓣引导编辑器)断裂成N端半部和C端半部,分别递送它们,然后让它们共定位以在细胞内重新形成完整的蛋白(或融合蛋白,视情况而定)。蛋白或融合蛋白分开的半部可各自包含断裂内含肽标签以通过蛋白反式剪接机制促进完整蛋白或融合蛋白的重新形成。

[0926] 由断裂内含肽催化的蛋白反式剪接提供了用于蛋白连接的完全酶促方法。断裂内含肽本质上是断裂成两部分(分别命名为N-i内含肽和C-内含肽)的连续内含肽(如,微型内含肽)。断裂内含肽的N-内含肽和C-内含肽可以非共价结合形成活性内含肽,并以与连续内含肽基本相同的方式催化剪接反应。已在自然界中发现了断裂内含肽,也在实验室中进行了工程化改造。如本文所用,术语“断裂内含肽”是指其中在N-端和C-端氨基酸序列之间存

在一个或多个肽键断裂，使得N-端和C-端序列成为单独的分子的任何内含肽，其可非共价地重新结合或重组为对反式剪接反应起作用的内含肽。任何具有催化活性的内含肽或其片段均可用于衍生用于本发明方法的断裂内含肽。例如，一方面，断裂内含肽可源自真核内含肽。另一方面，断裂内含肽可源自细菌内含肽。另一方面，断裂内含肽可源自古细菌内含肽。优选地，如此衍生的断裂内含肽将仅具有催化反式剪接反应所必需的氨基酸序列。

[0927] 如本文所用，“N-端断裂内含肽(In)”是指包含对反式剪接反应起作用的N-端氨基酸序列的任何内含肽序列。因此，In还包含发生反式剪接时剪出的序列。In可包含作为天然存在的内含肽序列的N-端部分的修饰的序列。例如，In可包含额外的氨基酸残基和/或突变的残基，只要包含这样的额外和/或突变的残基不会使In在反式剪接中失去功能。优选地，包含额外和/或突变的残基提高或增强In的反式剪接活性。

[0928] 如本文所用，“C-端断裂内含肽(Ic)”是指包含对反式剪接反应起作用的C-端氨基酸序列的任何内含肽序列。一方面，Ic包含4至7个连续氨基酸残基，其中至少4个氨基酸来自衍生其的内含肽的最后的β-链。因此，Ic还包含在发生反式剪接时剪出的序列。Ic可包含作为天然存在的内含肽序列的C-端部分的修饰的序列。例如，Ic可包含额外的氨基酸残基和/或突变的残基，只要包含这样的额外和/或突变的残基不会使In在反式剪接中失去功能。优选地，包含额外的和/或突变的残基提高或增强Ic的反式剪接活性。

[0929] 在本发明的一些实施方案中，与Ic或In连接的肽可包含额外的化学部分，尤其包括荧光基团、生物素、聚乙二醇(PEG)、氨基酸类似物、非天然氨基酸、磷酸盐基、糖基、放射性同位素标记和药物分子。在其他实施方案中，与Ic连接的肽可包含一个或多个化学反应性基团，其中包括酮、醛、Cys残基和Lys残基。存在“内含肽剪接多肽(ISP)”时，断裂内含肽的N-内含肽和C-内含肽可非共价结合形成活性内含肽并催化剪接反应。如本文所用，“内含肽剪接多肽(ISP)”是指从断裂内含肽去除Ic、In或两者时保留的断裂内含肽的氨基酸序列部分。在某些实施方案中，In包括ISP。在另一实施方案中，Ic包括ISP。在又一实施方案中，ISP是单独的肽，其既不与In也不与Ic共价连接。

[0930] 断裂内含肽可通过工程化改造非结构化环中的一个或多个断裂位点或干预微型内含肽结构中存在的-12保守β-链之间的氨基酸序列而由连续内含肽产生。β-链之间区域内断裂位点的位置可能存在一定的灵活性，条件是断裂的产生不会破坏内含肽的结构，特别是结构化的β-链，达到足以使蛋白剪接活性丧失的程度。

[0931] 在蛋白反式剪接中，一个前体蛋白由N-外显肽部分和后面的N-内含肽组成，另一个前体蛋白由C-内含肽和后面的C-外显肽部分组成，并且反式剪接反应(由N-和C-内含肽共同催化)切除两个内含肽序列并用肽键连接这两个外显肽序列。蛋白反式剪接是酶促反应，可在非常低(如，微摩尔)浓度的蛋白下进行，并且可在生理条件下进行。

[0932] 示例性序列如下：

[0933]

名称	配体依赖性内含肽的序列
2-4 内含肽:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 8)

[0934]

3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDR VAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASM MGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEIL MIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS SRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRAL DKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL YSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDK FLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHN C (SEQ ID NO: 9)
30R3-1 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 10)
30R3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 11)
30R3-3 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 12)
37R3-1 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC ((SEQ ID NO: 13)
37R3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD

		RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 14)
[0935]	37R3-3 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 15)

[0936] 尽管内含肽最常被发现为连续结构域,但有些以自然断裂的形式存在。在这种情况下,这两个片段作为单独的多肽表达,并且必须在剪接发生之前结合,即所谓的蛋白反式剪接。

[0937] 示例性断裂内含肽是Ssp DnaE内含肽,其包含两个亚基,即DnaE-N和DnaE-C。这两个不同的亚基由单独的基因编码,即dnaE-n和dnaE-c,分别编码DnaE-N和DnaE-C亚基。DnaE是天然存在于集胞藻(Synechocytissp.)的断裂内含肽。PCC6803能够指导两种不同蛋白的反式剪接,每种蛋白都包含与DnaE-N或DnaE-C的融合物。

[0938] 其他天然存在或工程化改造的断裂内含肽序列是已知的或可由本文所述完整内含肽序列或本领域可获得的那些制备。断裂内含肽序列的示例可见于Stevens et al.,"A promiscuous split intein with expanded protein engineering applications," PNAS,2017,Vol.114:8538-8543;Iwai et al.,"Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme,FEBS Lett,580:1853-1858,其各自通过引用并入本文。另外的断裂内含肽序列可见于如WO2013/045632、WO2014/055782、WO2016/069774和EP2877490,其各自的内容通过引用并入本文。

[0939] 此外,已在体内和体外描述了蛋白反式剪接(Shingledecker,et al.,Gene207:187(1998),Southworth,et al.,EMBOJ.17:918(1998);Mills,et al.,Proc.Natl.Acad.Sci.USA,95:3543-3548(1998);Lew,et al.,J.Biol.Chem.,273:15887-15890(1998);Wu,et al.,Biochim.Biophys.Acta 35732:1(1998b),Yamazaki,et al.,J.Am.Chem.Soc.120:5591(1998),Evans,et al.,J.Biol.Chem.275:9091(2000);Otomo,et al.,Biochemistry 38:16040-16044(1999);Otomo,et al.,J.Biolmol.NMR 14:105-114(1999);Scott,et al.,Proc.Natl.Acad.Sci.USA 96:13638-13643(1999)),并提供了表达随后经历连接以形成功能性产物的两个无活性片段的蛋白的机会,例如,如图66和图67关于从两个单独表达的半部形成完整的PE融合蛋白所示。

[0940] E.RNA-蛋白相互作用结构域

[0941] 在不同实施方案中,两个单独的蛋白结构域(如,Cas9结构域和聚合酶结构域)可通过使用"RNA-蛋白募集系统"(如,"MS2标签化技术")彼此共定位以形成功能复合物(类似

于包含两个单独的蛋白结构域的融合蛋白的功能)。这样的系统通常用“RNA-蛋白相互作用结构域”(又名“RNA-蛋白募集结构域”)标记一个蛋白结构域,用特异性识别并结合RNA-蛋白相互作用结构域的“RNA结合蛋白”(如,特定的发夹结构)标记另一个蛋白结构域。可利用这些类型的系统来共同定位多瓣引导编辑器的结构域,以及为多瓣引导编辑器募集附加功能,如UGI结构域。在一个示例中,MS2标签化技术基于MS2噬菌体外壳蛋白(“MCP”或“MS2cp”)与噬菌体基因组中存在的茎环或发夹结构的自然相互作用,即“MS2发卡”。在MS2发夹的情况下,其被MS2噬菌体外壳蛋白(MCP)识别和结合。因此,在一个示例性场景中,脱氨酶-MS2融合物可募集Cas9-MCP融合物。

[0942] 本领域对其他模块化RNA-蛋白相互作用结构域的综述描述于,例如,Johansson et al.,“RNArecognitionby the MS2 phage coatprotein,”Sem Virol.,1997,Vol.8(3):176-185;Delebecque et al.,“Organization of intracellular reactions with rationally designed RNA assemblies,”Science,2011,Vol.333:470-474;Mali et al.,“Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering,”Nat.Biotechnol.,2013,Vol.31:833-838;and Zalatan et al.,“Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds,”Cell,2015,Vol.160:339-350,其各自通过引用以其整体并入本文。其他系统包括专门募集PCP蛋白的PP7发夹和专门募集Com蛋白的“com”发夹。参见Zalatan et al.。

[0943] MS2发夹(或等同地被称为“MS2适体”)的核苷酸序列为:

[0944] GCCAACATGAGGATCACCCATGTCTGCAGGGCC(SEQ ID NO:763)。

[0945] MCP或MS2cp的氨基酸序列为:

[0946] GSASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCS

[0947] VRQSSAQNRKYTIKVEVPKVATQTVGGEELPVAGWRSYLNMELTIPIFAT NSDCELIVKAMQGLLKD GNPIPSAIAANSGIY(SEQ ID NO:764)。

[0948] F.UGI结构域

[0949] 在其他实施方案中,本文所述的多瓣引导编辑器可包含一个或多个尿嘧啶糖苷酶抑制剂结构域。如本文所用,术语“尿嘧啶糖苷酶抑制剂(UGI)”或“UGI结构域”是指能够抑制尿嘧啶-DNA糖苷酶碱基切除修复酶的蛋白。在一些实施方案中,UGI结构域包含野生型UGI或如SEQ ID NO:3873所示的UGI。在一些实施方案中,本文提供的UGI蛋白包含UGI的片段和与UGI或UGI片段同源的蛋白。例如,在一些实施方案中,UGI结构域包含SEQ ID NO:3873所示氨基酸序列的片段。在一些实施方案中,UGI片段包含含有SEQ ID NO:3873所示氨基酸序列的至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%的氨基酸序列。在一些实施方案中,UGI包含与SEQ ID NO:3873所示氨基酸序列同源的氨基酸序列,或与SEQ ID NO:3873所示氨基酸序列的片段同源的氨基酸序列。在一些实施方案中,包含UGI或UGI片段或UGI或UGI片段的同源物的蛋白被称为“UGI变体”。UGI变体与UGI或其片段具有同源性。例如,UGI变体与野生型UGI或SEQ ID NO:3873所示的UGI至少70%相同、至少75%相同、至少80%相同、至少85%相同、至少90%相同、至少95%相同、至少96%相同、至少97%、至少98%、至少99%、至少99.5%、或至少99.9%相同。在一些实施方案中,UGI变体包含UGI的片

段,使得该片段与野生型UGI或SEQ ID NO:3873所示的UGI的相应片段至少70%相同、至少80%相同、至少90%相同、至少95%相同、至少96%相同、至少97%相同、至少98%相同、至少99%相同、至少99.5%相同、或至少99.9%相同。在一些实施方案中,UGI包含以下氨基酸序列:

[0950] 尿嘧啶-DNA糖苷酶抑制剂:

[0951] >sp|P14739|UNGI_BPPB2

[0952] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTD ENVMLLTSDAPEYKP WALVIQDSNGENKIKML(SEQ ID NO:3873)。

[0953] 本文所述的多瓣引导编辑器可包含多于一个UGI结构域,其可被如本文所述一个或多个接头分开。

[0954] G.其他PE元件

[0955] 在某些实施方案中,本文所述的多瓣引导编辑器可包含碱基修复抑制剂。术语"碱基修复抑制剂"或"IBR"是指能够抑制核酸修复酶(如,碱基切除修复酶)活性的蛋白。在一些实施方案中,IBR是OGG碱基切除修复抑制剂。在一些实施方案中,IBR是碱基切除修复抑制剂("iBER")。碱基切除修复的示例性抑制剂包括APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGG1、hNEIL1、T7EndoI、T4PDG、UDG、hSMUG1和hAAG的抑制剂。在一些实施方案中,IBR是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是可以是无催化活性的糖苷酶或无催化活性的双加氧酶或氧化酶的小分子或肽抑制剂或其变体的iBER。在一些实施方案中,IBR是可以是TDG抑制剂、MBD4抑制剂或AlkBH酶抑制剂的iBER。在一些实施方案中,IBR是包含无催化活性的TDG或无催化活性的MBD4的iBER。示例性无催化活性的TDG是SEQ ID NO:3872(人TDG)的N140A突变体。

[0956] 以下提供了一些示例性糖苷酶。这些糖苷酶结构域中的任何无催化活性变体是可融合至本公开提供的多瓣引导编辑器的napDNAbp或聚合酶结构域的iBER。

[0957] OGG(人)

[0958] MPARALLPRRMGHRTLASTPALWASIPCPRSELRLDLVLPSGQSFRWREQSPAHWSGVLADQVWTLTQ TEEQLHCTVYRGDKSQASRPTPDELEAVRKYFQLDVTLAQLYHHWGSVDSHFQEVAQKFQGVRLLRQDPIECLFSF ICSSNNNIARITGMVERLCQAFGPRLIQLDDVTYHGFPSLQALAGPEVEAHLRKLGLGYRARYVSASARAILEEQG GLAWLQQLRESSYEEAHKALCILPGVGTKVADCICLMALDKPQAVPVDVHMWHIAQRDYSWHPTTSQAKGPSPQTN KELGNFFRSLWGPYAGWAQAVLFSADLRQSRHAQEPPAKRRKGSKGPEG(SEQ ID NO:3937)

[0959] MPG(人)

[0960] MVTPALQMKKPKQFCRRMGQKKQRPARAGQPHSSSDAAQAPAEQPHSSSDAAQAPCPRERCLGPPTTP GPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTELRGRIVETEAYLGPEDEAAHSRGGRQTP RNRGMFMKPGTLYVYIIYGMYFCMNISSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSK LCQALAINKSFDQRDLAQDEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDT QA(SEQ ID NO:3938)

[0961] MBD4(人)

[0962] MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSECNPLLQEP IASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFT VLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGV

DDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKS QLDRTVCISDAGA CGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHL HTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNL VQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEY LTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS(SEQ ID NO:3871)

[0963] TDG(人)

[0964] MEAENAGSYSLQQAQAFYTFPFQQLMAEAPNMAVVNEQQMPEEVPAPAPAQEPVQEAPKGRKRKPRTT EPKQPVEPKKPVESKKSGKSAKSKEKQEKITDTFKVKRKVDRFNGVSEAELLTKTLPDILTFNLDIVIIGINPGLM AAYKGHHYPGPGNHFWKCLFMSGLSEVQLNHMDDHTLPGKYGIGFTNMVERTTPGSKDLSSKEFREGGRILVQKLQ KYQPRIAVFNGKCIYEIFSKEVFGVKVKNLEFGLQPHKIPDTETLCYVMPSSSARCAQFPRAQDKVHYYIKLKDLR DQLKGIERNMDVQEVQYTFDLQLAQEDAKKMAVKEEKYDPGYEAAYGGAYGENPCSSEPCGFSSNGLIESVELRGE SAFSGIPNGQWMTQSFTDQIPSFSNHCGTQEQEEESHA(SEQ ID NO:3872)

[0965] 在一些实施方案中,本文所述融合蛋白可包含一个或多个异源蛋白结构域(如,除了引导编辑器组件外,约或多于约1、2、3、4、5、6、7、8、9、10或更多个结构域)。融合蛋白可包含任何额外的蛋白序列,并且任选包含任何两个结构域之间的接头序列。可能存在的其他示例性特征是定位序列,例如细胞质定位序列、输出序列(如,核输出序列或其他定位序列),以及可用于融合蛋白的溶解、纯化或检测的序列标签。

[0966] 可融合至多瓣引导编辑器或其组件(如,napDNAbp结构域、聚合酶结构域或NLS结构域)的蛋白结构域的示例包括但不限于表位标签和报告基因序列。表位标签的非限制性示例包括组氨酸(His)标签、V5标签、FLAG标签、流感血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的示例包括但不限于谷胱甘肽-5-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)、β-半乳糖苷酶、β-葡萄糖醛酸酶、萤光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)和自发荧光蛋白,包括蓝色荧光蛋白(BFP)。多瓣引导编辑器可与编码结合DNA分子或结合其他细胞分子的蛋白或蛋白片段的基因序列融合,包括但不限于麦芽糖结合蛋白(MBP)、S-标签、LexADNA结合结构域(DBD)融合物、GAL4DNA结合结构域融合物,以及纯疱疹病毒(HSV)BP16蛋白融合物。可形成引导编辑器一部分的其他结构域描述于2011年3月10日公布的美国专利公开号2011/0059502,并通过引用以其整体并入本文。

[0967] 在本公开的一方面,报告基因包括但不限于谷胱甘肽-5-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)β-半乳糖苷酶、β-葡糖醛酸苷酶、萤光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)和自发荧光蛋白,包括蓝色荧光蛋白(BFP),其可被引入细胞以编码基因产物,用作测量基因产物表达的改变或修饰的标志物。在本公开的某些实施方案中,基因产物是萤光素酶。在本公开的另一个实施方案中,基因产物的表达降低。

[0968] 本文提供的合适的蛋白标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc-标签、钙调蛋白-标签、FLAG-标签、血凝素(HA)-标签、多组氨酸标签(又被称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softags(如,Softag1、Softag3)、链标签、生物素连接酶标签、Flash标签、V5标签和SBP标签。其他合适的序列对本领域技术人员而言

是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0969] 在本公开的一些实施方案中,可通过调整停留时间、量和/或PE系统的表达组件的活性来时序调节多瓣引导编辑系统的活性。例如,如本文所述,PE可与能够改变PE的细胞内半衰期的蛋白结构域融合。在涉及两个或更多个载体的某些实施方案中(如,其中在两个或更多个单独载体上编码本文所述组件的载体系统),PE系统的活性可通过控制递送载体的时间来时序调节。例如,在一些实施方案中,编码核酸酶系统的载体可在编码模板的载体之前递送PE。在其他实施方案中,编码PEgRNA的载体可在编码PE系统的载体之前递送向导。在一些实施方案中,同时递送编码PE系统和PEgRNA的载体。在某些实施方案中,同时递送的载体时序递送例如PE、PEgRNA和/或第二链向导RNA组件。在进一步的实施方案中,从载体上的编码序列转录的RNA(如,核酸酶转录物)可进一步包含至少一种能够改变RNA的细胞内半衰期和/或调节翻译控制的元件。在一些实施方案中,可增加RNA的半衰期。在一些实施方案中,可减少RNA的半衰期。在一些实施方案中,元件可能能够增加RNA的稳定性。在一些实施方案中,元件可能能够降低RNA的稳定性。在一些实施方案中,元件可在RNA的3'UTR内。在一些实施方案中,元件可包括聚腺苷酸化信号(PA)。在一些实施方案中,元件可包括帽,如上游mRNA或PEgRNA末端。在一些实施方案中,RNA可不包含PA,使得转录后其在细胞中更快地降解。在一些实施方案中,元件可包括至少一种富含AU的元件(ARE)。ARE可以依赖于组织类型、细胞类型、时间、细胞定位和环境的方式被ARE结合蛋白(ARE-BP)结合。在一些实施方案中,去稳定元件可促进RNA衰变、影响RNA稳定性或激活翻译。在一些实施方案中,ARE可包含50至150个核苷酸长度。在一些实施方案中,ARE可包含序列AUUUA的至少一个拷贝。在一些实施方案中,至少一个ARE可被添加到RNA的3'UTR。在一些实施方案中,元件可以是土拨鼠肝炎病毒(WHP)。

[0970] 转录后调控元件(WPRE),其产生三级结构以增强转录本的表达。在进一步的实施方案中,元件是经修饰和/或截短的WPRE序列,其能够增强转录本的表达,描述于例如Zufferey et al.,J Virol,73(4):2886-92(1999)and Flajolet et al.,JVirol,72(7):6175-80(1998)。在一些实施方案中,WPRE或等同物可添加到RNA的3'UTR。在一些实施方案中,元件可选自富含快速或缓慢衰变的转录本的其他RNA序列基序。

[0971] 在一些实施方案中,编码PE或PEgRNA的载体可通过PE系统切割载体上存在的靶序列而自毁。切割可能会阻止PE或PEgRNA从载体继续转录。虽然转录可能在线性化载体上发生一段时间,但进行细胞内降解的表达的转录本或蛋白将有更少的时间产生脱靶效应,而无需由编码载体的表达继续供应。

[0972] [6]PEgRNA

[0973] 本文所述多瓣引导编辑系统考虑使用任何合适的PEgRNA。通过使用包含编码期望的核苷酸变化的逆转录(RT)模板序列的特殊配置的向导RNA,可利用或调整靶标引发的逆转录(TPRT)的机制,以进行精确且通用的基于CRISPR/Cas的基因组编辑。本申请将这种特殊配置的向导RNA称为"延伸的向导RNA"或"PEgRNA",因为RT模板序列可作为标准或传统向导RNA分子的延伸提供。本申请考虑了用于双瓣和四瓣引导编辑的延伸的向导RNA的任何合适的配置或排列。

[0974] 用于双瓣或四瓣引导编辑的pegRNA的一般涉及如图92所示。用于双瓣或四瓣引导编辑的pegRNA具有与用于经典引导编辑的pegRNA相似的设计,但是编码任何与靶基因座的

同源性的RT模板区域不是必需的。相反，在不同的实施方案中，两个pegRNA可以含有编码其3’端为彼此的反向互补序列的3’瓣的合成的RT模板。这种3′瓣之间的互补促进了其退火和内源性DNA序列替换为预期的新DNA序列。这需要两个pegRNA中的RT模板的5’区域是彼此反向互补的序列，并且这种互补性的量可以变化(图92)。

[0975] PEgRNA结构

[0976] 图3A显示可用于本文公开的多瓣引导编辑系统中的延伸的向导RNA的一个实施方案，其中传统向导RNA(绿色部分)包括约20nt前间隔序列和与napDNAbp结合的gRNA核心区。在该实施方案中，向导RNA在5’端包括延伸的RNA区段，即5’延伸。在该实施方案中，5’延伸包括逆转录模板序列、逆转录引物结合位点和任选的5-20个核苷酸接头序列。如图1A-1B所示，RT引物结合位点与在R环的非靶链中形成切口后形成的游离3′端杂交，从而引发逆转录酶以5’至3’方向进行DNA聚合。

[0977] 图3B显示可用于本文公开的引导编辑系统中的延伸的向导RNA的另一个实施方案，其中传统向导RNA(绿色部分)包括约20nt前间隔序列和与napDNAbp结合的gRNA核心。在该实施方案中，向导RNA在3’端包括延伸的RNA片段，即3’延伸。在该实施方案中，3’延伸包括逆转录模板序列和逆转录引物结合位点。如图1C-1D所示，RT引物结合位点与在R环的非靶链中形成切口后形成的游离3’端杂交，从而引发逆转录酶以5’至3’方向进行DNA聚合。

[0978] 图3C显示可用于本文公开的引导编辑系统中的延伸的向导RNA的另一个实施方案，其中传统向导RNA(绿色部分)包括约20nt前间隔序列和与napDNAbp结合的gRNA核心。在该实施方案中，向导RNA包括在gRNA核心内分子间位置处的延伸RNA区段，即，分子内延伸。在该实施方案中，分子内延伸包括逆转录模板序列和逆转录引物结合位点。RT引物结合位点与在R环的非靶链中形成切口后形成的游离3’端杂交，从而引发逆转录酶以5’至3’方向进行DNA聚合。

[0979] 在一个实施方案中，分子间RNA延伸的位置不在向导RNA的前间隔序列中。在另一个实施方案中，gRNA核心中分子间RNA延伸的位置。在又一个实施方案中，分子间RNA延伸的位置是向导RNA分子除了在前间隔序列内的任何位置，或在破坏前间隔序列的位置。

[0980] 在一个实施方案中，分子间RNA延伸在前间隔序列的3’端下游插入。在另一个实施方案中，分子间RNA延伸在前间隔区序列的3’端下游的至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸处插入。

[0981] 在其他实施方案中，分子间RNA延伸插入gRNA中，gRNA是指向导RNA的对应于或包含tracrRNA的部分，其结合Cas9蛋白或其等同物(即，不同的napDNAbp)和/或与其相互作用。优选地，分子间RNA延伸的插入不破坏或最小限度地破坏tracrRNA部分和napDNAbp之间的相互作用。

[0982] RNA延伸(至少包括RT模板和引物结合位点，例如，参见图92)的长度可以是任何有用的长度。在不同实施方案中，RNA延伸的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷

酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0983] RT模板序列还可以是任何合适的长度。例如,RT模板序列的长度可为至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少300个核苷酸、至少400个核苷酸,或至少500个核苷酸。

[0984] 在其他实施方案中,其中逆转录引物结合位点序列的长度为至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0985] 在其他实施方案中,任选的接头或间隔区序列的长度为至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少30个核苷酸、至少40个核苷酸、至少50个核苷酸、至少60个核苷酸、至少70个核苷酸、至少80个核苷酸、至少90个核苷酸、至少100个核苷酸、至少200个核苷酸、至少300个核苷酸、至少400个核苷酸、或至少500个核苷酸。

[0986] 在某些实施方案中,RT模板序列编码与非靶链同源(因此与靶链的相应位点互补)但包括一个或多个核苷酸变化的单链DNA分子。至少一个核苷酸变化可包括一个或多个单碱基核苷酸变化、一个或多个缺失和一个或多个插入。

[0987] 如图1G所示,RT模板序列合成的单链DNA产物与非靶链同源,并含有一个或多个核苷酸变化。RT模板序列的单链DNA产物与互补靶链序列平衡杂交,从而置换同源的内源性靶链序列。在一些实施方案中,置换的内源性链可被称为5'内源性DNA瓣种类(如,参见图1E)。这种5'内源性DNA瓣种类可通过5'瓣核酸内切酶(如,FEN1)去除,可连接现在与内源性靶链杂交的单链DNA产物,从而在内源性序列和新合成的链之间形成错配。错配可通过细胞的先天DNA修复和/或复制过程解决。

[0988] 在不同实施方案中,RT模板序列的核苷酸序列对应于非靶链的核苷酸序列,该非靶链作为5'瓣种类被置换并且与待编辑的位点重叠。

[0989] 在延伸的向导RNA的不同实施方案中,逆转录模板序列可编码与邻近切口位点的

内源性DNA序列互补的单链DNA瓣，其中单链DNA瓣包含期望的核苷酸变化。单链DNA瓣可在切口位点置换内源性单链DNA。切口位点处被置换的内源性单链DNA可具有5’端并形成内源性瓣，其可被细胞切除。在不同实施方案中，5’端内源性瓣的切除可有助于驱动产物形成，因为去除5’端内源性瓣促进单链3’DNA瓣与相应互补DNA链的杂交，以及掺入或同化由单链3’DNA瓣携带到靶DNA中的期望的核苷酸变化。

[0990] 在延伸的向导RNA的不同实施方案中，单链DNA瓣的细胞修复导致期望的核苷酸变化的安装，从而形成期望的产物。

[0991] 在其他实施方案中，将期望的核苷酸变化安装在以下的编辑窗口：切口位点的约-5至+5之间、或切口位点的约-10至+10之间、或切口位点的约-20至+20之间、或切口位点的约-30至+30之间、或切口位点的约-40至+40之间、或切口位点的约-50至+50之间、或切口位点的约-60至+60之间、或切口位点的约-70至+70之间、或切口位点的约-80至+80之间、或切口位点的约-90至+90之间，或切口位点的约-100至+100之间，或切口位点的约-200至+200之间。

[0992] 在其他实施方案中，将期望的核苷酸变化安装在以下的编辑窗口：距切口位点约+1至+2之间，或距切口位点约+1至+3、+1至+4、+1至+5、+1至+6、+1至+7、+1至+8、+1至+9、+1至+10、+1至+11、+1至+12、+1至+13、+1至+14、+1至+15、+1至+16、+1至+17、+1至+18、+1至+19、+1至+20、+1至+21、+1至+22、+1至+23、+1至+24、+1至+25、+1至+26、+1至+27、+1至+28、+1至+29、+1至+30、+1至+31、+1至+32、+1至+33、+1至+34、+1至+35、+1至+36、+1至+37、+1至+38、+1至+39、+1至+40、+1至+41、+1至+42、+1至+43、+1至+44、+1至+45、+1至+46、+1至+47、+1至+48、+1至+49、+1至+50、+1至+51、+1至+52、+1至+53、+1至+54、+1至+55、+1至+56、+1至+57、+1至+58、+1至+59、+1至+60、+1至+61、+1至+62、+1至+63、+1至+64、+1至+65、+1至+66、+1至+67、+1至+68、+1至+69、+1至+70、+1至+71、+1至+72、+1至+73、+1至+74、+1至+75、+1至+76、+1至+77、+1至+78、+1至+79、+1至+80、+1至+81、+1至+82、+1至+83、+1至+84、+1至+85、+1至+86、+1至+87、+1至+88、+1至+89、+1至+90、+1至+90、+1至+91、+1至+92、+1至+93、+1至+94、+1至+95、+1至+96、+1至+97、+1至+98、+1至+99、+1至+100、+1至+101、+1至+102、+1至+103、+1至+104、+1至+105、+1至+106、+1至+107、+1至+108、+1至+109、+1至+110、+1至+111、+1至+112、+1至+113、+1至+114、+1至+115、+1至+116、+1至+117、+1至+118、+1至+119、+1至+120、+1至+121、+1至+122、+1至+123、+1至+124、或+1至+125。

[0993] 在其他实施方案中，将期望的核苷酸变化安装在以下的编辑窗口：距切口位点约+1至+2之间，或距切口位点约+1至+5、+1至+10、+1至+15、+1至+20、+1至+25、+1至+30、+1至+35、+1至+40、+1至+45、+1至+50、+1至+55、+1至+100、+1至+105、+1至+110、+1至+115、+1至+120、+1至+125、+1至+130、+1至+135、+1至+140、+1至+145、+1至+150、+1至+155、+1至+160、+1至+165、+1至+170、+1至+175、+1至+180、+1至+185、+1至+190、+1至+195、或+1至+200。

[0994] 在不同方面，延伸的向导RNA是向导RNA的修饰版本。向导RNA可能是天然存在的、由编码核酸表达的或化学合成的。用于获得或以其他方式合成向导RNA和确定向导RNA的适当序列的方法是本领域所熟知的，包括与感兴趣基因组靶位点的靶链相互作用和杂交的前间隔序列。

[0995] 在不同实施方案中，除其他因素外，向导RNA序列的特定设计方面取决于感兴趣基因组靶位点(即，待编辑的期望位点)的核苷酸序列和存在于本文所述多瓣引导编辑系统中

的napDNAbp的类型(如,Cas9蛋白),例如PAM序列位置、靶序列中的G/C含量百分比、微同源区域的程度、二级结构等。

[0996] 一般地,向导序列是与靶多核苷酸序列具有足够互补性以与靶序列杂交并指导napDNAbp(如,Cas9、Cas9同源物或Cas9变体)与靶序列的序列特异性结合的任何多核苷酸序列。在一些实施方案中,当使用合适的比对算法最佳比对时,向导序列与其对应的靶序列之间的互补程度为约或大于约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或以上。最佳比对可使用用于比对序列的任何合适算法来确定,其非限制性示例包括Smith-Waterman算法、Needleman-Wunsch算法、基于Burrows-Wheeler变换的算法(如,Burrows WheelerAligner)、ClustalW、ClustalX、BLAT、Novoalign(NovocraftTechnologies,ELAND(Illumina,San Diego,Calif.)、SOAP(可获自soap.genomics.org.cn)和Maq(可获自maq.sourceforge.net)。在一些实施方案中,向导序列的长度为约或多于约5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、2728、29、30、35、40、45、50、75个、或更多的核苷酸。

[0997] 在一些实施方案中,向导序列的长度小于约75、50、45、40、35、30、25、20、15、12个或更少的核苷酸。可通过任何合适的测定来评估向导序列引导多瓣引导编辑器与靶序列的序列特异性结合的能力。例如,可向具有相应靶序列的宿主细胞提供多瓣引导编辑器的组件,包括待测试的向导序列,例如通过用编码本文公开的多瓣引导编辑器组件的载体转染,然后评估靶序列内的优先切割,例如通过本文所述Surveyor检测法。类似地,靶多核苷酸序列的切割可通过以下在试管中来评估:提供靶序列、多瓣引导编辑器的组件,包括待测试的向导序列和不同于测试向导序列的对照向导序列,以及比较测试和对照向导序列反应之间靶序列的结合或切割率。其他测定是可能的,并且是本领域技术人员会想到。

[0998] 可选择向导序列以靶向任何靶序列。在一些实施方案中,靶序列是细胞基因组内的序列。示例性靶序列包括靶基因组中独特的那些。例如,对于化脓性链球菌Cas9,基因组中的独特靶序列可包括形式MMMMMMMMNNNNNNNNNNNNXGG(SEQ ID NO:204)的Cas9靶位点,其中NNNNNNNNNNNNXGG(SEQ ID NO:205)(N是A、G、T,或C;X可以是任何碱基)。基因组中的独特靶序列可包括形式MMMMMMMMMNNNNNNNNNNNNXGG(SEQ ID NO:206)的化脓性链球菌Cas9靶位点,其中NNNNNNNNNNNXGG(SEQ ID NO:207)(N是A、G、T或C;X可以是任何碱基)。对于嗜热链球菌CRISPR1Cas9,基因组中的独特靶序列可包括形式MMMMMMMMNNNNNNNNNNNNXXAGAAW(SEQ ID NO:208)的Cas9靶位点,其中NNNNNNNNNNNNXXAGAAW(SEQ ID NO:209)(N是A、G、T或C;X可以是任何碱基;W是A或T)。基因组中的独特靶序列可包括形式MMMMMMMMMNNNNNNNNNNNXXAGAAW(SEQ ID NO:210)的嗜热链球菌CRISPR1Cas9靶位点,其中NNNNNNNNNNNXXAGAAW(SEQ ID NO:211)(N是A、G、T或C;X可以是任何碱基;W是A或T)。对于化脓性链球菌Cas9,基因组中的独特靶序列可包括形式MMMMMMMMNNNNNNNNNNNNXGGXG(SEQ ID NO:212)的Cas9靶位点,其中NNNNNNNNNNNNXGGXG(SEQ ID NO:213)(N是A、G、T或C;X可以是任何碱基)。基因组中的独特靶序列可包括形式MMMMMMMMMNNNNNNNNNNNXGGXG(SEQ ID NO:214)的化脓性链球菌Cas9靶位点,其中NNNNNNNNNNNXGGXG(SEQ ID NO:215)(N是A、G、T或C;和X可以是任何碱基)。在这些序列的各序列中,“M”可以是A、G、T或C,并且在确定序列为独特时无需考虑。

[0999] 除非另外表明,否则如本文在PEgRNA或向导RNA序列中使用的,应意识到字母“T”

或“胸腺嘧啶”表示DNA序列中编码PEgRNA或向导RNA序列的核碱基，并且旨在指PEgRNA或向导RNA序列的尿嘧啶(U)核碱基或本领域已知的任何化学修饰的尿嘧啶，比如，5-甲氧基尿嘧啶。

[1000] 在一些实施方案中，选择向导序列以降低向导序列内的二级结构的程度。二级结构可通过任何合适的多核苷酸折叠算法来确定。一些程序基于计算最小吉布斯(Gibbs)自由能。一种此类算法的示例是mFold，如Zuker and Stiegler(NucleicAcids Res.9(1981)，133-148)。另一个示例折叠算法是维也纳大学(the University ofVienna)理论化学研究所开发的在线网络服务器RNAfold，使用质心结构预测算法(参见，例如A.R.Gruber et al.，2008，Cell106(1):23-24；以及PA Carr and GM Church，2009，Nature Biotechnology27(12):1151-62)。进一步的算法可见于美国申请系列号61/836,080(Broad Reference BI-2013/004A)；通过引用并入本文。

[1001] 一般地，tracr配对序列包括与tracr序列具有足够互补性以促进以下一项或多项的任何序列：(1)在含有相应tracr序列的细胞中切除位于tracr配对序列两侧的向导序列；以及(2)在靶序列处形成复合物，其中该复合物包含与tracr序列杂交的tracr配对序列。一般地，互补性程度参考tracr配对序列和tracr序列沿两个序列中较短序列长度的最佳比对。最佳比对可通过任何合适的比对算法确定，并且可进一步导致二级结构，例如tracr序列或tracr配对序列内的自互补性。在一些实施方案中，最佳比对为约或大于约25%、30%、40%、50%、60%、70%、80%、90%、95%、97.5%、99%或更高时，tracr序列和tracr配对序列之间的互补程度沿着两者中较短者的长度。在一些实施方案中，tracr序列的长度为约或大于约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50个或更多的核苷酸。在一些实施方案中，tracr序列和tracr配对序列包含在单个转录本内，使得两者之间的杂交产生具有二级结构(如，发夹结构)的转录本。用于发夹结构的优选环形成序列的长度为四个核苷酸，最优选具有序列GAAA。然而，可使用更长或更短的环序列，也可使用替代序列。序列优选包括核苷酸三联体(如，AAA)和额外的核苷酸(如，C或G)。环形成序列的示例包括CAAA和AAAG。在本发明的实施方案中，转录本或转录的多核苷酸序列具有至少两个或更多个发夹。在优选实施方案中，转录本具有2个、3个、4个或5个发夹。在本发明的另一个实施方案中，转录本具有至多5个发夹。在一些实施方案中，单个转录本还包括转录终止序列；优选地，这是polyT序列，如6个T核苷酸。包含向导序列、tracr配对序列和tracr序列的单个多核苷酸的进一步非限制性示例如下(从5’至3’列出)，其中“N”代表向导序列的碱基，第一块小写字母代表tracr配对序列，第二块小写字母代表tracr序列，最后的poly-T序列代表转录终止子：

[1002] (1)NNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT(SEQ ID NO:216)；

[1003] (2)NNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT(SEQ ID NO:217)；

[1004] (3)NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgtTTTTT(SEQ ID NO:218)；

[1005] (4)NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcTTTTTT(SEQ ID NO:219)；

[1006] (5)NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT(SEQ ID NO:220);以及

[1007] (6)NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTTTTT(SEQ ID NO:221)。

[1008] 在一些实施方案中,序列(1)至(3)与来自嗜热链球菌CRISPR1的Cas9组合使用。在一些实施方案中,序列(4)至(6)与来自化脓性链球菌的Cas9组合使用。在一些实施方案中,tracr序列是与包含tracr配对序列的转录本分开的转录本。

[1009] 对于本领域技术人员显而易见地,如本文公开,为了将包含Cas9结构域和单链DNA结合蛋白的任何融合蛋白靶向靶位点,如包含待编辑的点突变的位点,通常需要将融合蛋白与向导RNA(如,sgRNA)一起共表达。如本文别处更详细解释的,向导RNA通常包含允许Cas9结合的tracrRNA框架和赋予Cas9:核酸编辑酶/结构域融合蛋白序列特异性的向导序列。

[1010] 在一些实施方案中,向导RNA包含结构5′-[向导序列]-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3′(SEQ ID NO:222),其中向导序列包含与靶序列互补的靶序列。向导序列通常为20个核苷酸长。基于本公开,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向特定基因组靶位点的合适的向导RNA的序列对于本领域技术人员而言是显而易见的。这类合适的向导RNA序列通常包含与待编辑的靶核苷酸上游或下游50个核苷酸内的核酸序列互补的向导序列。本文提供了一些适用于将任何提供的融合蛋白靶向特定靶序列的示例性向导RNA序列。其他向导序列是本领域公知的并且可与本文所述的多瓣引导编辑器一起使用。

[1011] 在其他实施方案中,PEgRNA包括图3D中描绘的那些。

[1012] 在其他实施方案中,PEgRNA可包括图3E中描绘的那些。

[1013] 图3D提供了本文考虑的PEgRNA的实施方案的结构,其可根据实施例2中定义的方法进行设计。PEgRNA包含以5′至3′方向排列的3个主要组分元件,即:间隔区、gRNA核心和3′端处的延伸臂。延伸臂在5′至3′方向可进一步分为以下结构元件,即:引物结合位点(A)、编辑模板(B)和同源臂(C)。此外,PEgRNA可包含可选的3’端修饰区(e1)和可选的5’端修饰区(e2)。更进一步地,PEgRNA可在PEgRNA的3’端包含转录终止信号(未描绘)。这些结构元件在本文中进一步定义。PEgRNA结构的描述并不意味着是限制性的,而是包含元件排列的变化。例如,可选的序列修饰区(e1)和(e2)可位于所示的任何其他区域之内或之间,并且不限于位于3’和5’端。

[1014] 图3E提供了本文考虑的PEgRNA的另一个实施方案的结构,其可根据实施例2中定义的方法进行设计。PEgRNA包含以5’至3’方向排列的3个主要组分元件,即:间隔区、gRNA核心和3′端处的延伸臂。延伸臂在5′至3′方向上可进一步分为以下结构元件,即:引物结合位点(A)、编辑模板(B)和同源臂(C)。此外,PEgRNA可包含可选的3’端修饰区(e1)和可选的5’端修饰区(e2)。更进一步地,PEgRNA可在PEgRNA的3’端包含转录终止信号(未描绘)。这些结构元件在本文中进一步定义。PEgRNA结构的描述并不意味着是限制性的,而是包含元件排列的变化。例如,可选的序列修饰区(e1)和(e2)可位于所示的任何其他区域之内或之间,并且不限于位于3’和5’端。

[1015] PEgRNA改进

[1016] PEgRNA还可包括额外的设计改进,其可改变PEgRNA的性质和/或特征,从而改进多瓣引导编辑的功效。在不同实施方案中,这些改进可能属于许多不同类别中的一个或多个,包括但不限于:(1)设计以能够从非聚合酶III(pol III)启动子有效表达功能性PEgRNA,这能够表达更长的PEgRNA而无需繁琐的序列要求;(2)对核心Cas9结合的PEgRNA支架的改进,其可提高效力;(3)修饰PEgRNA以提高RT持续合成能力,从而能够在靶基因组位点插入更长的序列;(4)在PEgRNA的5′或3′端添加RNA基序,以提高PEgRNA稳定性,增强RT持续合成能力,防止PEgRNA错误折叠,或募集对基因组编辑重要的其他要素。

[1017] 在一个实施方案中,PEgRNA可设计具有pol III启动子以提高具有较大延伸臂的更长长度的PEgRNA的表达。sgRNA通常由U6 snRNA启动子表达。该启动子募集pol III以表达相关的RNA,可用于表达保留在细胞核内的短RNA。然而,pol III的加工能力不强,无法在有效基因组编辑需要的水平上表达长度超过几百个核苷酸的RNA。此外,pol III可在U的延伸处停止或终止,这可能会限制使用PEgRNA插入的序列多样性。已检测募集聚合酶II(如,pCMV)或聚合酶I(如,U1 snRNA启动子)的其他启动子表达更长sgRNA的能力。然而,这些启动子通常是部分转录的,这会导致表达的PEgRNA中间隔区5′的额外序列,这已被证明会导致Cas9:sgRNA活性以位点依赖性方式明显降低。此外,虽然pol III转录的PEgRNA可简单地在6-7U延伸中终止,但从pol II或pol I转录的PEgRNA需要不同的终止信号。通常,此类信号还会导致聚腺苷酸化,从而导致PEgRNA从细胞核中不期望的转运。类似地,从pol II启动子(如,pCMV)表达的RNA通常是5′-加帽的,这也导致它们的核输出。

[1018] 此前,Rinn和同事筛选了多种表达平台,用于生产长的非编码RNA-(lncRNA)标记的sgRNA[183]。这些平台包括从pCMV表达并终止于来自人的MALAT1 ncRNA的ENE元件[184]、来自KSHV的PANENE元件[185]或来自U1 snRNA的3′框[186]。值得注意地,MALAT1 ncRNA和PAN ENE形成保护poly A尾的三螺旋[184,187]。这些构建体还可增强RNA稳定性。还考虑能够表达更长的PEgRNA的这些表达系统。

[1019] 此外,已设计了一系列方法来切割将作为PEgRNA一部分转录的pol II启动子部分,添加自切割核酶(如,锤头型[188]、手枪型[189]、斧头型[189]、发夹型[190]、VS[191]、twister[192]或twister sister[192]核酶)或其他自切割元件来加工转录的向导,或被Csy4[193]识别并且也导致加工向导的发夹。此外,假设并入多个ENE基序可提高PEgRNA表达和稳定性,如先前对KSHV PAN RNA和元件所证明的那样[185]。还预计以环状内含子RNA(ciRNA)形式环化PEgRNA也可能导致增强的RNA表达和稳定性,以及核定位[194]。

[1020] 在不同实施方案中,PEgRNA可包括各种上述元件,如以下序列所示例。

[1021] 非限制性示例1-由pCMV、Csy4发夹、PEgRNA和MALAT1 ENE组成的

[1022] PEgRNA表达平台

[1023] TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA

TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCG
TATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTTAGGGTCATGAAGGTTTT
TCTTTTCCTGAGAAAACAACACGTATTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAA
AAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTTGACT
(SEQ ID NO:223)

[1024] 非限制性示例2-由pCMV、Csy4发卡、PEgRNA和PANENE组成的PEgRNA表达平台

[1025] TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACAT
AACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTA
CATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG
GTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA
TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCG
TATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTTGTTTTGGCTGGGTTTTT
CCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTAT
ACTTTTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAA
AGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAA(SEQ ID NO:224)

[1026] 非限制性示例3-由pCMV、Csy4发卡、PEgRNA和3xPAN ENE组成的PEgRNA表达平台

[1027] TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACAT
AACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTA
CATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG
GTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA
TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCG
TATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTTGTTTTGGCTGGGTTTTT
CCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTAT
ACTTTTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAA
AGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAACACACTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCG
GACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTT
AACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTTAA
TCCATAAAAAAAAAAAAAAAAAAAATCTCTCTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGA
CCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTAGAGC
TCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAA

AAAAAAAAAA(SEQ ID NO:225)

[1028] 非限制性示例4-由pCMV、Csy4发卡、PEgRNA和3′框组成的PEgRNA表达平台

[1029] TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACAT AACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTA CATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG GTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCG TATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC AACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTGTTTCAAAAGTAGACTGT ACGCTAAGGGTCATATCTTTTTTTGTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA(SEQ ID NO:226)

[1030] 非限制性示例5-由pU1、Csy4发卡、PEgRNA和3′框组成的PEgRNA表达平台

[1031] CTAAGGACCAGCTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAGGGAGAGGCAGACGTC ACTTCCCCTTGGCGGCTCTGGCAGCAGATTGGTCGGTTGAGTGGCAGAAAGGCAGACGGGGACTGGGCAAGGCACT GTCGGTGACATCACGGAC AGGGCGACTTCTATGTAGATGAGGCAGCGCAGAGGCTGCTGCTTCGCCACTTGCTGC TTCACCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGTTCAGGACCGCTGATCGGAAGTGAGAATCCCAGCTGTGTG TCAGGGCTGGAAAGGGCTCGGGAGTGCGCGGGGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTATGAGGCTGTGT CGGGGCAGAGGCCCAAGATCTCAGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAG AAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAG CGTGCTCAGTCTGTTTCAGCAAGTTCAGAGAAATCTGAACTTGCTGGATTTTTGGAGCAGGGAGATGGAATAGGAG CTTGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGCACCCACTTTCTGGAGTTTC AAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA(SEQ ID NO:227).

[1032] 在不同的其他实施方案中,可通过对支架或核心序列引入改进来改善PEgRNA。这可通过引入已知的完成。

[1033] 可能可以改善核心,结合Cas9的PEgRNA支架以增强PE活性。已证明了几种这样的方法。例如,支架(P1)的第一配对元件包含GTTTT-AAAAC(SEQ ID NO:3939)配对元件。已证明这种T延伸导致pol III暂停和RNA转录本的过早终止。在P1的这一部分,已证明T-A对之一至G-C对的合理突变可增强sgRNA活性,表明这种方法对于PEgRNA也是可行的[195]。此外,还证明增加P1的长度可增强sgRNA折叠并导致提高活性[195],表明这是提高PEgRNA活性的另一条途径。对核心的改进示例可包括:

[1034] 含有对P1的6nt延伸的PEgRNA

[1035] GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGCTCATGAAAATGAGCTAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGCTGTTTTTTT(SE Q ID NO:228)

[1036] P1内含有T-A至G-C突变的PEgRNA

[1037] GGCCCAGACTGAGCACGTGAGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCA

ACTTGAAAAAGTGGGACCGAGTCGGTCCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTTTTT(SEQ ID NO:229)

[1038] 在不同的其他实施方案中,可通过将修饰引入编辑模板区来改进PEgRNA。随着PEgRNA提供模板化的插入物尺寸增加,其更有可能被核酸内切酶降解,发生自发水解,或者折叠成无法被RT逆转录或破坏PEgRNA支架折叠和后续Cas9-RT结合的二级结构。因此,可能需要对PEgRNA模板进行修饰才能影响大的插入,例如整个基因的插入。这样做的一些策略包括在合成或半合成的PEgRNA中插入经修饰的核苷酸,使RNA对降解或水解更具抵抗力,或者不太可能采用抑制性二级结构[196]。这类修饰可包括8-氮杂-7-去氮杂鸟苷,其会减少富含G的序列中的RNA二级结构;锁核酸(LNA),减少降解并增强某些种类的RNA二级结构;增强RNA稳定性的2'-O-甲基、2'-氟或2'-O-甲氧基乙氧基修饰。这些修饰也可包含在PEgRNA的其他地方来增强稳定性和活性。可选择或另外地,可设计PEgRNA的模板,使其既编码期望的蛋白产物,也更可能采用能够通过RT展开的简单二级结构。这类简单结构会充当热力学源,从而不太可能出现阻止逆转录的更复杂的结构。最后,还可将模板分成两个独立的PEgRNA。在这样的设计中,将PE用于启动转录,并通过与Cas9融合的RNA结合蛋白或PEgRNA本身上的RNA识别元件(如,MS2适体)将单独的模板RNA募集到靶向位置。RT可直接结合到这个单独的模板RNA,或者在切换到第二模板之前在原始PEgRNA上启动逆转录。这种方法可通过防止添加长模板后PEgRNA的错误折叠以及不需要为了发生长插入将Cas9从基因组解离来实现长的插入,这种解离可能会抑制基于PE的长插入。

[1039] 在其他实施方案中,PEgRNA可通过在PEgRNA的5'和3'端处或甚至在它们之间的位点处(如,在gRNA核心区域或间隔区中)引入额外的RNA基序来改进。以上讨论了几个这样的基序—例如来自KSHV的PAN ENE和来自MALAT1的ENE作为终止来自非pol III启动子的较长PEgRNA的表达的可能手段。这些元件形成吞没polyA尾的RNA三螺旋,导致它们保留在细胞核内184,187。然而,通过在PEgRNA的3′端形成封闭末端核苷酸的复杂结构,这些结构也可能有助于防止核酸外切酶介导的PEgRNA降解。

[1040] 在3'端插入的附加结构元件也可增强RNA稳定性,尽管不能从非pol III启动子终止。这类基序可包括会封闭3′端的发夹或RNA四链体[197],或自切割核酶(如,HDV),其会导致在3′端形成2'-3′-环状磷酸酯,并且还可能使PEgRNA不太可能被核酸外切酶降解[198]。通过不完全剪接诱导PEgRNA环化以形成ciRNA也可增加PEgRNA稳定性并导致PEgRNA保留在细胞核内[194]。

[1041] 其他RNA基序也可通过增强RT与DNA-RNA双链体的结合来改进RT持续合成能力或增强PEgRNA活性。在其同源逆转录病毒基因组中添加由RT结合的天然序列可增强RT活性[199]。这可能包括天然引物结合位点(PBS)、聚嘌呤区(PPT)或参与逆转录病毒基因组二聚化和转录起始的吻环(kissing loop)[199]。

[1042] 在PEgRNA的5'和3'端添加二聚化基序(如,吻环或GNRA四环/四环受体对[200])也可导致PEgRNA的有效环化,提高稳定性。此外,预计添加这些基序能够物理分离PEgRNA间隔区和引物,防止间隔区闭塞阻碍PE活性。在间隔区或沿引物结合位点形成小立足点(toehold)发夹的PEgRNA的短5′延伸或3'延伸也可有利地竞争沿PEgRNA长度的互补内区域的退火,例如可能存在的间隔区和引物结合位点之间的相互作用。最后,吻环也可用于将其他模板RNA募集到基因组位点,并能够将RT活性从一种RNA交换到另一种。作为不同二级结构的示例性实施方案,图3D和图3E描绘的PEgRNA列出可被工程化改造到PEgRNA的任何区域(包括延伸

臂的末端部分(即,e1和e2))的一些二级RNA结构,如图所示。

[1043] 示例改进包括但不限于:

[1044] PEgRNA-HDV融合物

[1045] GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTTTTTTT(SEQ ID NO:230)

[1046] PEgRNA-MMLV吻环

[1047] GGTGGGAGACGTCCCACCGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCTTTTTTT(SEQ ID NO:231)

[1048] PEgRNA-VS核酶吻环

[1049] GAGCAGCATGGCGTCGCTGCTCACGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTCCATCAGTTGACACCCTGAGGTTTTTTT(SEQ ID NO:232)

[1050] PEgRNA-GNRA四环/四环受体

[1051] GCAGACCTAAGTGGUGACATATGGTCTGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAUACGTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTUACGAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTGCATGCGATTAGAAATAATCGCATGTTTTTTT(SEQ ID NO:233)

[1052] PEgRNA模板切换的二级RNA-HDV融合物

[1053] TCTGCCATCAAAGCTGCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTTTTTTT(SEQ ID NO:234)

[1054] 可通过定向进化以类似于如何改进SpCas9和引导编辑器(PE)的方式进一步改进PEgRNA支架。定向进化可增强Cas9或进化的Cas9变体识别PEgRNA。此外,不同的PEgRNA支架序列在不同基因组基因座处增强相关位点的PE活性,减少脱靶活性,或两者可能是最佳的。最后,添加其他RNA基序的PEgRNA支架的进化几乎肯定会提高融合PEgRNA相对于未进化的融合RNA的活性。例如,由c-di-GMP-I适体和锤头型核酶组成的变构核酶的进化导致显著提高活性[202],表明进化也会提高锤头型-PEgRNA融合物的活性。此外,虽然Cas9目前通常不允许sgRNA的5′延伸,但定向进化可能会产生使这种不耐受性减轻的突变,从而允许利用其他RNA基序。

[1055] 本公开考虑任何这样的方式来进一步提高在此公开的多瓣引导编辑系统的功效。

[1056] 在不同实施方案中,限制来自延伸臂的连续T序列的出现可能是有利的,因为T的连续系列可能限制PEgRNA被转录的能力。例如,设计PEgRNA时应避免或应从最终的设计序列至少去除至少连续三个T、至少连续4个T、至少连续5个T、至少连续6个T、至少连续7个T、至少连续8个T、至少连续9个T、至少连续10个T、至少连续11个T、至少连续12个T、至少连续13个T、至少连续14个T、或至少连续15个T。在一个实施方案中,可避免在PEgRNA延伸臂中包含不需要的连续T串,但避免富含连续A:T核碱基对的靶位点。

[1057] 用于反式引导编辑的断裂PEgRNA设计

[1058] 本公开还考虑了反式引导编辑,其指通过将PEgRNA分成两个不同的分子(向导RNA和tPERT分子)来操作的引导编辑的修饰版本。将tPERT分子编程为与靶DNA位点的引导编辑

器复合物共定位，从而将引物结合位点和DNA合成模板以反式引入引导编辑器。例如，参见图3G，用于反式引导编辑器(tPE)的实施方案，其显示双组件系统，包含(1)募集蛋白(RP)-PE:gRNA复合物和(2)包括引物结合位点和连接到RNA蛋白募集结构域(如，茎环或发夹)的DNA合成模板的tPERT，其中RP-PE:gRNA复合物的募集蛋白组件将tPERT募集到待编辑的靶位点，从而将PBS和DNA合成模板以反式与引导编辑器结合。换言之，将tPERT工程化改造为包含PEgRNA的延伸臂(全部或部分)，其包括引物结合位点和DNA合成模板。这种方法的一个优点是将PEgRNA的延伸臂与向导RNA分开，从而使延伸臂的PBS和向导RNA的间隔区序列之间倾向于发生的退火相互作用最小化。

[1059] 反式引导编辑的关键特征是反式引导编辑器将tPERT募集到DNA编辑位点的能力，从而在引导编辑位点有效地共定位PEgRNA的所有功能。可通过将RNA蛋白募集结构域(如，MS2适体)安装到tPERT中并将相应的募集蛋白融合到引导编辑器(如，通过至napDNAbp的接头或通过至聚合酶的接头)来实现募集，该引导编辑器能够与RNA-蛋白募集结构域特异性结合，从而将tPERT分子募集到引导编辑器复合物。如图3H描述的过程所示，RP-PE:gRNA复合物结合并对靶DNA序列产生切口。然后，募集蛋白(RP)募集tPERT共定位至与DNA靶位点结合的引导编辑复合物，从而允许位于tPERT上的引物结合位点与切口链上的引物序列结合，随后允许聚合酶(如，RT)针对位于tPERT上的DNA合成模板合成单链DNA，直至tPERT的5′端。

[1060] 虽然tPERT在图3G和图3H中以包含RNA蛋白募集结构域5′端的PBS和DNA合成模板示出，其他配置中的tPERT可设计为具有位于RNA蛋白募集结构域3′端的PBS和DNA合成模板。然而，具有5’延伸的tPERT的优点是单链DNA的合成将在tPERT的5’端自然终止，因而在引导编辑的DNA合成过程中不会冒险使用RNA蛋白募集结构域的任何部分作为模板。

[1061] PEgRNA设计方法

[1062] 本公开还涉及用于设计PEgRNA的方法。

[1063] 在设计的一方面，设计方法可考虑待使用的引导编辑的特定应用。例如，如本文所示例和讨论，引导编辑可使用但不限于：(a)对核苷酸序列安装校正突变的变化，(b)安装蛋白和RNA标签，(c)在感兴趣的蛋白上安装免疫表位，(d)在蛋白中安装可诱导的二聚化结构域，(e)安装或去除序列以改变生物分子的活性，(f)安装重组酶靶位点以引导特定的基因变化，以及(g)使用易错RT的靶序列诱变。除了通常在感兴趣的靶位点插入、改变或缺失核苷酸序列的这些方法之外，引导编辑器还可用于构建高度可编程的文库，以及进行细胞数据记录和谱系追踪研究。在这些不同的用途中，如本文所述，可能存在与制备任何给定的这些应用中特别有用的PEgRNA有关的特定设计方面。

[1064] 为任何特定的应用或引导编辑的使用设计PEgRNA时，会考虑许多考虑因素，包括但不限于：

[1065] (a)靶序列，即，期望引导编辑器在其中安装一个或多个核碱基修饰的核苷酸序列；

[1066] (b)靶序列内切割位点的位置，即，引导编辑器将在特定核碱基位点诱导单链切口，以在切口的一侧产生3′端RT引物序列和切口的另一侧产生5′端内源性瓣(最终被FEN1或其等同物去除，并由3′ssDNA瓣置换)。切割位点类似于“编辑位置”，因为这产生了3′端RT引物序列，其在RNA依赖性DNA聚合过程中被RT延伸，以产生包含期望编辑的3′ssDNA瓣，然后该3′ssDNA瓣置换靶序列中的5′内源性DNA瓣。

[1067] (c)可用的PAM序列(包括经典SpCas9 PAM位点,以及Cas9变体识别的非经典PAM位点以及具有扩展或不同PAM特异性的等同物);

[1068] (d)可用的PAM序列与靶序列的切割位点的位置之间的间距;

[1069] (e)所使用的引导编辑器的特定Cas9、Cas9变体或Cas9等同物;

[1070] (f)引物结合位点的序列和长度;

[1071] (g)编辑模板的序列和长度;

[1072] (h)同源臂的序列和长度;

[1073] (i)间隔区序列和长度;和

[1074] (j)核心序列。

[1075] 本公开讨论了以上这些方面。

[1076] 在一个实施方案中,在此提供了设计合适的PEgRNA和任选用于第二位点切口产生的切口产生性sgRNA设计指南的方法。该实施方案提供了用于引导编辑的设计PEgRNA和切口产生性sgRNA的一组分步说明,其考虑了上述考虑中的一个或多个。这些步骤参考图70A至图70I中所示的示例。

[1077] 1.定义靶序列和编辑。检索以期望编辑(点突变、插入、缺失或其组合)位置为中心的靶DNA区域的序列(约200bp)。参见图70A。

[1078] 2.定位靶PAM。确定接近期望编辑位置的PAM。可在邻近期望编辑位置的任一DNA链上确定PAM。虽然优选接近编辑位置的PAM(即,其中切口位点距编辑位点小于30nt,或从编辑位点至切口位点小于29nt、28nt、27nt、26nt、25nt、24nt、23nt、22nt、21nt、20nt、19nt、18nt、17nt、16nt、15nt、14nt、13nt、12nt、11nt、10nt、9nt、8nt、7nt、6nt、5nt、4nt、3nt或2nt),但可使用前间隔序列和PAM安装编辑,将切口放置在距编辑位点≥30nt处。参见图70B。

[1079] 3.定位切口位点。对于正在考虑的各PAM,确定相应切口位点以及在哪条链上。对于Sp Cas9 H840A切口酶,切割发生在含PAM的链中NGG PAM5′的第3个和第4个碱基之间。所有编辑核苷酸必须存在于切口位点的3′,因此适当的PAM必须将切口置于含PAM的链的靶编辑的5′处。在以下所示的示例中,有两种可能的PAM。为简单起见,其余步骤将显示仅使用PAM1的PEgRNA设计。参见图70C。

[1080] 4.设计间隔区序列。SpCas9的前间隔序列对应于含PAM的链中NGG PAM的5′的20个核苷酸。有效的Pol III转录起始需要G作为第一个转录的核苷酸。如果前间隔序列的第一个核苷酸是G,则PEgRNA的间隔区序列就是前间隔序列。如果前间隔序列的第一个核苷酸不是G,则PEgRNA的间隔区序列为G接着前间隔序列。参见图70D。

[1081] 5.设计引物结合位点(PBS)。使用起始等位基因序列,确定含PAM的链的DNA引物。DNA引物的3′端正好是切口位点上游的核苷酸(即,Sp Cas9的NGG PAM的5′的第4个碱基)。作为使用PE2和PE3的一般设计原则,包含与DNA引物互补的12至13个核苷酸的PEgRNA引物结合位点(PBS)可用于包含约40-60%GC含量的序列。对于较低GC含量的序列,应当测试较长(14至15nt)的PBS。对于较高GC含量的序列,应当测试较短(8至11nt)的PBS。最佳PBS序列应根据经验确定,无论GC含量如何。为设计长度为p的PBS序列,使用起始等位基因序列,取含PAM的链中切口位点5′的前p个核苷酸的反向互补序列。参见图70E。

[1082] 6.设计RT模板(或DNA合成模板)。RT模板(或DNA合成模板,其中聚合酶不是逆转录

酶)编码设计的编辑并与邻近编辑的序列同源。在一个实施方案中,这些区域对应于图3D和图3E的DNA合成模板,其中DNA合成模板包括“编辑模板”和“同源臂”。最佳RT模板长度因靶位点而异。对于短程编辑(位置+1至+6),建议测试短(9至12nt)、中(13至16nt)和长(17至20nt)RT模板。对于远程编辑(+7位及以上),建议使用RT模板,该模板在编辑位点后至少延伸5nt(最好是10nt或更多),以允许足够的3′DNA瓣同源性。对于远程编辑,应筛选多个RT模板以识别功能设计。对于较大的插入和缺失(≥5nt),建议将更大的3′同源性(约20nt或更多)掺入RT模板。当RT模板将编码G的合成作为逆转录DNA产物中的最后一个核苷酸(对应于PEgRNA的RT模板中的C)时,通常会损害编辑效率。由于许多RT模板支持有效的引导编辑,因此设计RT模板时建议避免G作为最终合成的核苷酸。为设计长度为r的RT模板序列,使用期望等位基因序列,并取原始含PAM的链中切口位点3′的前r个核苷酸的反向互补序列。注意,与SNP编辑相比,使用相同长度的RT模板进行插入或缺失编辑不包含相同的同源性。参见图70F。

[1083] 7.组装完整的PEgRNA序列。按以下顺序(5′至3′)连接PEgRNA组件:间隔区、支架、RT模板和PBS。参见图70G。

[1084] 8.设计PE3的切口产生性sgRNA。确定编辑上游和下游的非编辑链上的PAM。最佳切口位点高度依赖于基因座,应根据经验确定。一般地,在PEgRNA诱导的切口对面的位点5′放置40至90个核苷酸的切口导致更高的编辑产率和更少的indel。切口产生性sgRNA具有与起始等位基因中的20-nt前间隔序列匹配的间隔区序列,如果前间隔序列不以G开头,则添加5′-G。参见图70H。

[1085] 9.设计PE3b切口产生性sgRNA。如果PAM存在于互补链中,并且其相应的前间隔序列与靶向编辑的序列重叠,则此编辑可能是PE3b系统的候选者。在PE3b系统中,切口产生性sgRNA的间隔区序列匹配期望编辑等位基因的序列,但不匹配起始等位基因的序列。当编辑核苷酸落在切口产生性sgRNA前间隔序列的种子区(邻近PAM的约10nt)内时,PE3b系统有效运行。这可防止在安装编辑链之前对互补链产生切口,从而防止PEgRNA和sgRNA之间竞争结合靶DNA。PE3b还避免同时对两条链产生切口,从而在保持高编辑效率的同时显著减少indel形成。PE3b sgRNA应当具有与期望等位基因中的20nt前间隔序列匹配的间隔区序列,并在需要时添加5′G。参见图70I。

[1086] 上述用于设计合适的PEgRNA和第二位点切口产生性sgRNA的分步方法并不意味着以任何方式进行限制。本公开考虑了本领域普通技术人员可从中推导出的上述分步方法的变化。

[1087] [7]利用多瓣引导编辑的应用

[1088] 双瓣和四瓣引导编辑(即,多瓣引导编辑)具有许多潜在功能,比如安装肽肽标签、RNA标签、免疫表位、二聚结构域和重组酶靶位点。一种双瓣引导编辑的此类应用是在用户指定的基因组中的位置安装重组酶或整合酶序列。图93说明了将Bxb1重组酶attB(38bp)和attP(50bp)位点安装到人基因组的靶标位置(HEK293T位点3或HEK3)中,同时在两个切口位点之间缺失90bp的插入序列。3′瓣之间不同程度的互补允许成功编辑,尽管更长的互补序列产生更有利的期望编辑与插入缺失标记的比率。其他双瓣引导编辑应用包括用肽或蛋白质序列对基因进行内源性标记,或者用新的DNA序列替换外显子,该DNA序列具有替代突变属于外显子序列的多个变体的潜力。

[1089] 双瓣引导编辑可以用于在人基因组的靶标位置引入一个或两个重组酶位点。如果插入单个重组酶位点,这些位点可用作基因组重组酶位点和外源供应的DNA(如质粒)内的第二个重组酶位点之间重组酶介导的反应的登陆位点。这使得DNA货物的靶向整合成为可能。如果将两个重组酶位点插入DNA的相邻区域,根据重组酶位点的方向,这些位点可以用于重组酶介导的切除或插入序列的倒置,或用于重组酶介导的与外源DNA的盒交换以整合货物。不同染色体上相容重组酶位点的整合使得定向染色体易位成为可能。双瓣引导编辑可以用于在人基因组的许多基因座有效引入重组酶位点(图94)。因此,通过双瓣引导编辑的重组酶位点整合与DNA重组酶的配对代表了实现多种类型的SV编辑和DNA货物的靶整合的有力方法。

[1090] 本文描述的多瓣引导编辑器(例如,图90中描绘的实施方案)可以用于精确插入新的DNA序列、精确缺失内源性基因组DNA序列或者用新的DNA序列替换内源性基因组DNA序列。

[1091] 例如,如本文所示例和讨论,双引导编辑可用于(a)对核苷酸序列安装校正突变的变化,(b)安装蛋白和RNA标签,(c)在感兴趣的蛋白上安装免疫表位,(d)在蛋白中安装可诱导的二聚化结构域,(e)安装重组酶靶位点以引导特定的基因变化。除了这些通常在感兴趣的靶位点插入、改变或缺失核苷酸序列的方法外,双引导编辑器还可用于构建高度可编程的文库,以及进行细胞数据记录和谱系追踪研究。

[1092] 双引导编辑的这些特定示例性应用绝不旨在限制。本申请考虑双引导编辑的任何应用,其通常涉及在核苷酸序列如基因组DNA的靶位点处的一个或多个核碱基的某种形式的安装、缺失和/或替换。

[1093] 对于引导编辑的任何示例性应用,可使用本文公开的任何双引导编辑器,包括PE1、PE2、PE3和PE3b,或PE-短。

[1094] A.引导编辑与多瓣引导编辑

[1095] 经典引导编辑

[1096] 在不同实施方案中,引导编辑通过将靶DNA分子(期望对其引入核苷酸序列的变化)与复合延伸的向导RNA的核酸可编程DNA结合蛋白(napDNAbp)接触来进行操作。参考图1G,延伸的向导RNA包含位于向导RNA的3'或5'端或向导RNA的分子内位置的延伸,并编码期望的核苷酸变化(如,单核苷酸变化、插入或缺失)。步骤(a)中,napDNAbp/延伸的gRNA复合物接触DNA分子,延伸的gRNA引导napDNAbp与靶基因座结合。步骤(b)中,在靶基因座处的一条DNA链中引入切口(如,通过核酸酶或化学试剂),从而在靶基因座处的一条链中产生可用的3'端。在某些实施方案中,在对应于R-环链的DNA链(即,未与向导RNA序列杂交的链,即"非靶链")中产生切口。然而,切口可在任何一条链中引入。也就是说,可将切口引入R环"靶链"(即,与延伸的gRNA的前间隔序列杂交的链)或"非靶链"(即,形成R环的单链部分的链,且与靶链互补)。步骤(c)中,DNA链的3'端(由切口形成)与向导RNA的延伸部分相互作用以引发逆转录(即"靶引发的RT")。在某些实施方案中,3'端DNA链与向导RNA的延伸部分上的特定RT引发序列(即,"逆转录酶引发序列")杂交。步骤(d)中,引入逆转录酶(作为与napDNAbp的融合蛋白或以反式),该酶从引发位点的3'端向着延伸的向导RNA的5'端合成单链DNA。这形成包含期望的核苷酸变化(如,单碱基变化、插入或缺失或其组合)的单链DNA瓣,并且其与切口位点或邻近处的内源性DNA同源。步骤(e)中,将napDNAbp和向导RNA释放。

步骤(f)和(g)涉及单链DNA瓣的分解,使得期望的核苷酸变化被掺入靶基因座。可通过去除相应的5’内源性DNA瓣(如,通过FEN1或以反式提供的类似酶,作为与引导编辑器的融合物,或内源性提供)来驱动该过程向着期望的产物形成,该5’内源性DNA瓣在3’单链DNA瓣侵入并杂交内源性DNA序列时形成。不受理论的束缚,细胞内源性DNA修复和复制过程分解了错配的DNA以掺入核苷酸变化来形成期望的改变产物。还可驱动该过程向着具有“第二链切口”(如图1G所示例)或“时序第二链切口”(如图1I所示例并在此讨论)的产物形成。

[1097] 引导编辑过程可引入至少一种或多种以下基因变化:颠换、转换、缺失和插入。此外,可为特定应用实施引导编辑。例如,如本文所示例和讨论,引导编辑可用于(a)对核苷酸序列安装校正突变的变化,(b)安装蛋白和RNA标签,(c)在感兴趣的蛋白上安装免疫表位,(d)在蛋白中安装可诱导的二聚化结构域,(e)安装或去除序列以改变生物分子的活性,(f)安装重组酶靶位点以指导特定的基因变化,以及(g)使用易错RT诱变靶序列。除了这些通常在感兴趣的靶位点插入、改变或缺失核苷酸序列的方法外,引导编辑器还可用于构建高度可编程的文库,以及进行细胞数据记录和谱系追踪研究。本发明人还考虑旨在提高引导编辑效率的PEgRNA的额外设计特征。更进一步地,本发明人已构思了使用载体递送系统成功递送引导编辑器的方法,并且包括使用内含肽结构域断裂napDNAbp。

[1098] 术语“引导编辑系统”或“引导编辑器(PE)”是指本文所述使用靶标引发的逆转录(TPRT)的基因组编辑方法中涉及的组合物,包括但不限于napDNAbp、逆转录转录酶、融合蛋白(如,包含napDNAbp和逆转录酶)、延伸的向导RNA、和包含融合蛋白和延伸的向导RNA的复合物,以及辅助元件,例如第二链切口组件和5’内源性DNA瓣去除核酸内切酶(如,FEN1),用于帮助驱动引导编辑过程向着编辑产物形成。

[1099] 在另一个实施方案中,图3F的示意图描绘了典型的PEgRNA与双链DNA的靶位点的相互作用以及伴随产生的包含感兴趣基因变化的3′单链DNA瓣。双链DNA显示上面链为3′至5′方向,下面链为5′至3′方向。上面链包含“前间隔序列”和PAM序列,被称为“靶链”。互补的下面链被称为“非靶链”。尽管未显示,但所描述的PEgRNA将与Cas9或等同物复合。如所示,PEgRNA的间隔区与靶链的互补区(被称为前间隔序列,位于PAM序列的下游,长度约为20个核苷酸)退火。这种相互作用形成间隔区RNA和前间隔序列DNA之间的DNA/RNA杂合体,并诱导在与前间隔序列相对的区域形成R环。如本文别处所教导的,Cas9蛋白(未显示)然后在非靶链中诱导切口,如所示。然后,这会导致3′ssDNA瓣区的形成,根据*z*,该区域在引物结合位点处与PEgRNA的3′端相互作用。ssDNA瓣的3′端(即,逆转录酶引物序列)与PEgRNA上的引物结合位点(A)退火,从而引发逆转录酶。接着,逆转录酶(如,以反式提供或作为融合蛋白顺式提供,连接到Cas9构建体)聚合由编辑模板(B)和同源臂(C)编码的单链DNA。聚合继续向着延伸臂的5′端延伸。ssDNA的聚合链形成ssDNA3′端瓣,如别处所述(例如,如图1G所示),侵入内源性DNA,置换相应的内源性链(其作为内源性DNA的5′DNA瓣去除),并通过自然发生的DNA修复/复制循环安装期望的核苷酸编辑(单核苷酸碱基对变化、缺失、插入(包括整个基因)。

[1100] 在实施例1中进一步描述了引导编辑的这种应用。

[1101] 双引导编辑

[1102] 本说明书描述了双引导编辑系统(或双瓣引导编辑系统),其解决了与瓣平衡和随后通过同时编辑两条DNA链将编辑加入到未编辑的互补基因组DNA链相关的挑战。在双瓣引

导编辑系统中，两个pegRNA用于靶向基因组位点的相对链并且指导含有编辑DNA序列的两个互补3’瓣的合成(图91)。与经典引导编辑不同，不需要成对的编辑DNA链(3’瓣)直接与内源性基因组DNA中的5’瓣直接竞争，因为互补的编辑链也可以用于杂交。由于双链体的两条链均作为编辑DNA合成，双瓣引导编辑系统省略了经典引导编辑所需的替换未编辑的互补DNA链的需求。相反，细胞DNA修复机制仅需要切除成对的5’瓣(原始基因组DNA)并且将成对的3’瓣(编辑DNA)连接到基因座中。因此，在新合成的DNA链中也不需要包含与基因组DNA同源的序列，允许新链的选择性杂交并且有利于含有最小基因组同源性的编辑。切割两条DNA链的核酸酶活性形式的引导编辑器也可以用用于加快原始DNA序列的去除。因此。在某些实施方案中，双瓣引导编辑系统涉及一对新合成的DNA链(例如3’瓣)，其与待编辑的靶位点处的内源性DNA序列具有同源性。在一些实施方案中，双瓣引导编辑系统涉及一对新合成的DNA链，其中至少一个新合成的DNA链与待编辑的靶位点处的内源性DNA序列不包含同源性。在一些实施方案中，新合成的DNA链中的一条与靶位点处的内源性DNA序列包含同源性并且与另一条新合成的DNA链不包含同源性。在一些实施方案中，新合成的DNA链中的每一条与靶位点处的内源性DNA序列包含同源性并且与另一条新合成的DNA链不包含同源性。例如，双瓣pegRNA之一编码的新合成的3’瓣可以包含与另一个双瓣pegRNA的前间隔序列序列互补的区域。因此，在一些实施方案中，一对双瓣pegRNA各自与另一个pegRNA的间隔区序列具有同源性可能导致定位于该对双瓣pegRNA的前间隔序列序列之间的内源性的DNA序列的缺失。在一些实施方案中，双瓣引导编辑系统涉及一对新合成的DNA链，其中没有一条新合成的DNA链与待编辑的靶位点处的内源性DNA序列具有同源性。相反，两条新合成的DNA链(例如3’瓣)各自包含彼此互补的区域并且可以通过互补形成双链体。随后可以在靶位点处加入与双链体中待编辑的靶位点处的内源性DNA序列相比期望的编辑部分。与经典引导编辑一样，双引导编辑是多功能且精确的基因组编辑方法，其使用与聚合酶一起工作的合适可编程DNA结合(“napDNAbp”)(即，以融合蛋白的形式或者与napDNAbp反式提供)将新的遗传信息直接写入指定的DNA位点中，其中引导编辑系统用引导编辑(PE)向导RNA(“PEgRNA”)编程，该向导RNA既指定靶位点也通过工程化到向导RNA上(例如，在向导RNA的5′或3′端，或在其内部部分)的延伸(DNA或RNA)的方式作为替换DNA链形式的期望的编辑的模板。含有期望的编辑的替换链(例如，单个核碱基取代)与待编辑的靶位点的内源性链具有相同序列(不同之处在于其包含期望的编辑)。通过DNA修复和/或复制机制，靶位点的内源性链被替换为含有期望的编辑的新合成的替换链。在一些情况下，引导编辑可以被认为是“检索和替换”基因组编辑技术，因为本文所述的引导编辑器不仅检索和定位待编辑的靶位点，还同时编码含有期望的编辑的替换链，该替换链被安装在相应的靶位点内源性DNA链的位置。

[1103]

B		

[1104] B.双引导编辑在肽标签化中的应用

[1105] 另一方面，本公开提供了使用本文所述的双引导编辑器通过引导编辑将一个或多个肽标签基因移植到蛋白上的方法(比如图90所示的实施方案)。更具体地，本公开提供了用于从遗传学上安装一个或多个肽标签至蛋白上的方法，包括：使编码蛋白的靶核苷酸序列与配置为在其中插入编码一个或多个肽标签的第二核苷酸序列的引导编辑器接触，以产生编码融合蛋白的重组核苷酸序列，该融合蛋白包含与蛋白标签融合的蛋白。

[1106] 在其他实施方案中，本公开提供了制备包含感兴趣的肽和一个或多个肽标签的融

合蛋白的方法，该方法包括：使编码该蛋白的靶核苷酸序列与配置为在其中插入编码一个或多个肽标签的第二核苷酸序列的引导编辑器接触，以产生编码融合蛋白的重组核苷酸序列，该融合蛋白包含与蛋白标签融合的蛋白。

[1107] 在不同实施方案中，靶核苷酸序列是基因组DNA中的感兴趣的特定基因。感兴趣基因可编码感兴趣的蛋白(如，受体、酶、治疗性蛋白、膜蛋白、转运蛋白、信号转导蛋白、或免疫学蛋白等)。感兴趣基因还可编码RNA分子，包括但不限于信使RNA(mRNA)、转移RNA(tRNA)、核糖体RNA(rRNA)、小核RNA(snRNA)、反义RNA、向导RNA、微RNA(miRNA)、小干扰RNA(siRNA)和无细胞RNA(cfRNA)。

[1108] 肽标签可以是任何肽标签或其变体，其将一种或多种功能赋予蛋白以用于诸如分离、纯化、可视化、增溶或检测的目的。肽标签可包括“亲和标签”(促进蛋白纯化)、“溶解标签”(辅助蛋白正确折叠)、“色谱标签”(改变蛋白的色谱特性)、“表位标签”(用于结合高亲和力抗体)和“荧光标签”(促进细胞或体外蛋白的可视化)。肽标签的示例包括但不限于以下标签：

[1109]

名称	氨基酸序列	SEQ ID NO:
AVITAG™	GLNDIFEAQKIEWHE	SEQ ID NO: 245
C-标签	EPEA	SEQ ID NO: 246
钙调蛋白-标签	KRRWKKNFIAVSAANRFKKISSSGAL	SEQ ID NO: 247
聚谷氨酸-标签	EEEEEE	SEQ ID NO: 248
E-标签	GAPVPYPDPLEPR	SEQ ID NO: 249
FLAG-标签	DYKDDDDK	SEQ ID NO: 250
HA-标签	YPYDVPDYA	SEQ ID NO: 251
HIS-标签	H (HIS_1) HH (HIS_2) HHH (HIS_3) HHHH (HIS_4) HHHHH (HIS_5) HHHHHH (HIS_6) HHHHHHH (HIS_7) HHHHHHHH (HIS_8) HHHHHHHHH (HIS_9) HHHHHHHHHH (HIS_{10}) HHHHHHHHHH…H…(HIS_N, WHEREIN N = 1-25)	SEQ ID NO: 252 SEQ ID NO: 253 SEQ ID NO: 254 SEQ ID NO: 255 SEQ ID NO: 256 SEQ ID NO: 257 SEQ ID NO: 258 SEQ ID NO: 259 SEQ ID NO: 260 SEQ ID NO: 261 SEQ ID NO: 262
MYC-标签	EQKLISEEDL	SEQ ID NO: 263
NE-标签	TKENPRSNQEESYDDNES	SEQ ID NO: 264
RHO1D4-标签	TETSQVAPA	SEQ ID NO: 265
S-标签	KETAAAKFERQHMDS	SEQ ID NO: 266
SBP-标签	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	SEQ ID NO: 267
SOF 标签-1	SLAELLNAGLGGS	SEQ ID NO: 268
SOF 标签-2	TQDPSRVG	SEQ ID NO: 269
SPOT-标签	PDRVRAVSHWSS	SEQ ID NO: 270
STREP-标签	WSHPQFEK	SEQ ID NO: 271
TC 标签	CCPGCC	SEQ ID NO: 272
TY 标签	EVHTNQDPLD	SEQ ID NO: 273
V5 标签	GKPIPNPLLGLDST	SEQ ID NO:274
VSV-标签	YTDIEMNRLGK	SEQ ID NO: 275
XPRESS 标签	DLYDDDDK	SEQ ID NO: 276

[1110] 肽标签还可能是以下亲和标签(用于蛋白的分离和/或纯化)(描述于Kimple et al.,“Overview ofAffinity Tags for Protein Purification,”CurrProtoc Protein Sci,2013,73:Unit-9.9的表9.9.1中,其通过引用并入本文)。

[1111]

名称	氨基酸序列	
AU1 表位	DTYRYI	SEQ ID NO: 278
AU5 表位	TDFYLK	SEQ ID NO: 279
噬菌体 T7 表位 (T7-标签)	MASMTGGQQMG	SEQ ID NO: 280
蓝舌病毒标签 (B-标签)	QYPALT	SEQ ID NO: 281
E2 表位	SSTSSDFRDR	SEQ ID NO: 282

[1112]

组氨酸亲和标签 (HAT)	KDHLIHNVHKEFHAHAHNK	SEQ ID NO: 283
HSV 表位	QPELAPED	SEQ ID NO: 284
聚精氨酸(ARG-标签)	RRRRR	SEQ ID NO: 285
聚天冬氨酸(ASP-标签)	CCCC	SEQ ID NO: 286
聚苯丙氨酸(PHE-标签)	FFFFFFFFFF	SEQ ID NO: 287
S1-标签	NANNPDWDF	SEQ ID NO: 288
S-标签	KETAAAKFERQHMDS	SEQ ID NO: 289
VSV-G	YTDIEMNRLGK	SEQ ID NO: 290

[1113] 在特定实施方案中，肽标签可包括His^6标签、FLAG-标签、V5-标签、GCN4-标签、HA-标签、Myc-标签、FIAsH/ReAsH-标签、分选酶底物、π钳(pi-clamp)。

[1114] 在不同实施方案中，肽标签可用于包括蛋白荧光标记、免疫沉淀、免疫印迹、免疫组织化学、蛋白募集、诱导型蛋白降解和全基因组筛选的应用。

[1115] 在不同的其他实施方案中，肽标签可包括内含肽序列以安装蛋白自剪接功能。如本文所用，术语“内含肽”是指在来自所有生命领域的生物体中发现的自动加工多肽结构域。内含肽(中间蛋白)执行被称为蛋白剪接的独特自动加工事件，其中其通过两个肽键的裂解从较大的前体多肽中切出自身，并在此过程中通过形成新的肽键连接侧翼外显肽(外部蛋白)序列。这种重排发生在翻译后(或可能是共翻译)，因为发现内含肽基因嵌入在其他蛋白编码基因的框架内。此外，内含肽介导的蛋白剪接是自发的。它不需要外部因素或能源，只需折叠内含肽结构域。这个过程又被称为顺式蛋白剪接，与具有“断裂内含肽”的反式蛋白剪接的自然过程相反。内含肽是自剪接RNA内含子的蛋白等同物(参见Perler et al., Nucleic Acids Res.22:1125-1127(1994))，其催化它们自身从前体蛋白中切除，同时伴随融合被称为外显肽的侧翼蛋白序列(综述见Perler et al.,Curr.Opin.Chem.Biol.1:292-299(1997);Perler,F.B.Cell 92(1):1-4(1998);Xu et al.,EMBO J.15(19):5146-5153(1996)))。

[1116] 已非常详细地研究了蛋白剪接过程的机制(Chong,et al.,J.Biol.Chem.1996,271,22159-22168;Xu,M-Q&Perler,F.B.EMBO Journal,1996,15,5146-5153)，并且在内含肽和外显肽剪接点存在保守氨基酸(Xu,et al.,EMBO Journal,1994,135517-522)。

[1117] 内含肽也可作为由两个单独转录和翻译的基因编码的两个片段存在。这些所谓的断裂内含肽自结合并催化反式蛋白剪接活性。已在不同的蓝藻和古细菌中鉴定出断裂内含肽(Caspi et al,Mol Microbiol.50:1569-1577(2003);Choi J.et al,J Mol Biol.556:1093-1106(2006.);Dassa B.et al,Biochemistry.46:322-330(2007.);Liu X.and Yang J.,J Biol Chem.275:26315-26318(2003);Wu H.et al.Proc Natl Acad Sci USA.£5:9226-9231(1998.);以及Zettler J.et al,FEBS Letters.553:909-914(2009))，但迄今为止尚未在真核生物中发现。最近，环境宏基因组数据的生物信息学分析揭示26个具有新基因组排列的不同位点。在每个基因座上，保守的酶编码区被断裂内含肽中断，其中独立的核酸内切酶基因插入编码内含肽亚结构域的部分之间。其中，5个基因座完全组装:DNA解旋酶(gp41-1、gp41-8)；肌苷-5′-单磷酸脱氢酶(IMPDH-1)；和核糖核苷酸还原酶催化亚基(NrdA-2和NrdJ-1)。这种断裂的基因结构似乎主要存在于噬菌体中(Dassa et al,Nucleic Acids Research.57:2560-2573(2009))。

[1118] 在某些实施方案中，本文所述的引导编辑器可用于在两个不同的蛋白中插入断裂

内含肽标签，当共表达形成融合蛋白时导致它们的细胞内连接。在蛋白反式剪接中，一个前体蛋白由N-外显肽部分和后面的N-内含肽组成，另一个前体蛋白由C-内含肽和后面的C-外显肽部分组成，并且反式剪接反应(由N-和C-内含肽共同催化)切除两个内含肽序列并用肽键连接这两个外显肽序列。蛋白反式剪接是酶促反应，可在非常低(如，微摩尔)浓度的的蛋白下工作，并且可在生理条件下进行。

[1119] 断裂内含肽NpuDnaE的特征在于具有蛋白反式剪接反应的最高报告率。此外，对于不同的外显肽序列的Npu DnaE蛋白剪接反应，6至37℃的温度和存在高达6M尿素被认为是稳健且高产的(Zettler J.et al,FEBS Letters.553:909-914(2009)；Iwai I.et al,FEBS Letters 550:1853-1858(2006))。如预期，在这些内含肽的N结构域引入CyslAla突变时，最初的N至S-酰基转移以及因此蛋白剪接被阻断。不幸地，C-端切割反应也几乎被完全抑制。C端剪接连接处的天冬酰胺环化对N端易裂肽键的酰基转移的依赖性似乎是自然断裂DnaE内含肽等位基因共有的独特性质(ZettlerJ.et al.FEBS Letters.555:909-914(2009))。

[1120] 由断裂内含肽催化的蛋白反式剪接提供了用于蛋白连接的完全酶促方法。断裂内含肽本质上是断裂成两部分(分别命名为N-内含肽和C-内含肽)的连续内含肽(如，微型内含肽)。断裂内含肽的N-内含肽和C-内含肽可以非共价结合形成活性内含肽，并以与连续内含肽基本相同的方式催化剪接反应。已在自然界中发现了断裂内含肽，还在实验室中进行了工程化改造。如本文所用，术语“断裂内含肽”是指其中在N-端和C-端氨基酸序列之间存在一个或多个肽键断裂，使得N-端和C-端序列成为单独的分子的任何内含肽，其可非共价地重新结合或重组为对反式剪接反应起作用的内含肽。任何具有催化活性的内含肽或其片段均可用于衍生用于本发明方法的断裂内含肽。例如，一方面，断裂内含肽可源自真核内含肽。另一方面，断裂内含肽可源自细菌内含肽。另一方面，断裂内含肽可源自古细菌内含肽。优选地，如此衍生的断裂内含肽将仅具有催化反式剪接反应所必需的氨基酸序列。

[1121] 断裂内含肽可通过工程化改造非结构化环中的一个或多个断裂位点或干预微型内含肽结构中存在的-12个保守β-链之间的氨基酸序列而由连续的内含肽产生。β-链之间区域内断裂位点的位置可能存在一定的灵活性，条件是断裂的产生不会破坏内含肽的结构，特别是结构化的β-链，达到足以使蛋白剪接活性丧失的程度。

[1122] 本文所述的引导编辑器可将肽标签(包括内含肽)掺入感兴趣的蛋白的C-末端。在其他实施方案中，可将肽标签(包括内含肽)掺入感兴趣的蛋白的N-末端。也可将肽标签掺入感兴趣的蛋白的内部。由本文所述的引导编辑器产生的融合蛋白可具有以下结构：

[1123] [感兴趣的蛋白]-[肽标签]；

[1124] [肽标签]-[感兴趣的蛋白]；或

[1125] [感兴趣的蛋白-N-端区域]-[肽标签]-[感兴趣的蛋白-C-端区域]。

[1126] 可将在肽标签化中使用的向导RNA设计原则应用于肽标签化。例如，在一个实施方案中，用于肽标签化的PEgRNA结构可具有以下结构：5′-[间隔区序列]-[gRNA核心或支架]-[延伸臂]-3′，其中延伸臂在5′至3′方向包括同源臂、编辑模板(包含编码肽标签的序列)和引物结合位点。该配置描绘于图3D和图24中。

[1127] 在另一个实施方案中，用于肽标签化的PEgRNA结构可具有以下结构：5’-[延伸臂]-[间隔区序列]-[gRNA核心或支架]-3’，其中延伸臂在5’至3’方向包括同源臂、编辑模板(包含编码肽标签的序列)和引物结合位点。该配置在图3E中描绘。

[1128] 使用引导编辑的肽标签化的实施方案描绘于图25和图26中，并描述于实施例4中。

[1129] C.双引导编辑在RNA标签化中的应用

[1130] 还可将双引导编辑用于通过RNA标签化操纵、改变和以其他方式修饰编码RNA功能的DNA序列，并以此方式提供间接修饰RNA的结构和功能的手段。例如，可将双PE用于插入在RNA水平上作用的基序(以下称为RNA基序)以标签化或以其他方式操纵非编码RNA或mRNA。这些基序可通过自切割或RNA酶用于增加基因表达、减少基因表达、改变剪接、改变转录后修饰、影响RNA的亚细胞定位、能够分离或确定RNA的细胞内或细胞外位置(使用如荧光RNA适体，如Spinach,Spinach2,Baby Spinach或Broccoli)、募集内源性或外源性蛋白或RNA结合物、引入sgRNA、或诱导RNA加工(更多细节参见图28B和实施例6)。

[1131] 可使用双引导编辑(例如，使用图90的双PE方法)利用适当PEgRNA(使用本文提供的指导设计)将以下RNA标签或基序插入感兴趣基因中来影响RNA的各种特性，包括RNA转运、表达水平、剪接和检测。

[1132]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
多瘤病毒猿猴病毒 40 (SV40) 1 型	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 331)	标签化的基因转录的终止; MRNA 转运至胞液; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAGGGCCATGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC*ACCTGAACCGTATATC* **TAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTT**CTATG*GCCCTGACTGGAA* (SEQ ID NO: 332)

[1133]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
多瘤病毒猿猴病毒 40 (SV40) 2 型	CCATGGCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTC (SEQ ID NO: 333)	标签化的基因转录的终止; MRNA 转运至胞液; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAGGGCCATGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC*ACCTGAACCGTATATC* **GAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTTGTAACCATTATAAGCTGCAATAAACAAGTTGGGCCATGG**CTATG*GCCCTGACTGGAA* (SEQ ID NO: 334)
多瘤病毒猿猴病毒 40 (SV40) 3 型	TGATCATAATCAAGCCATATCACATCTGTAGAGGTTTACTTGCTTTAAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGC (SEQ ID NO: 335)	标签化的基因转录的终止; MRNA 转运至胞液; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAGGGCCATGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC*ACCTGAACCGTATATC* **GCAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGAGGTTTTTTAAAGCAAGTAAACCTCTACAGATGT**

[1134]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
			GATATGGCTTGAT TATGATCACTATG*G CCCTGACTGGAA* (SEQ ID NO: 336)
人生长激素 (HGH)	ACGGGTGGCATCCCTGTG ACCCCTCCCCAGTGCCTC TCCTGGCCCTGGAAGTTG CCACTCCAGTGCCCACCA GCCTTGTCCTAATAAAATT AAGTTGCATCATTTTGTCT GACTAGGTGTCCTTCTATA ATATTATGGGGTGGAGGG GGGTGGTATGGAGCAAGG GGCAAGTTGGGAAGACA ACCTGTAGGGCCTGCGGG GTCTATTGGGAACCAAGC TGGAGTGCAGTGGCACAA TCTTGGCTCACTGCAATCT CCGCCTCCTGGGTTCAAG CGATTCTCCTGCCTCAGCC TCCCGAGTTGTTGGGATT CCAGGCATGCATGACCAG GCTCAGCTAATTTTTGTTT TTTTGGTAGAGACGGGGT TTCACCATATTGGCCAGGC TGGTCTCCAACTCCTAATC TCAGGTGATCTACCCACCT TGGCCTCCCAAATTGCTG GGATTACAGGCGTGAACC ACTGCTCCCTTCCCTGTCC TT (SEQ ID NO: 337)	RNA 转运至细胞质; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **AAGGACAGGGAA GGGAGCAGTGGT TCACGCCTGTAAT CCCAGCAATTTGG GAGGCCAAGGTG GGTAGATCACCTG AGATTAGGAGTTG GAGACCAGCCTG GCCAATATGGTGA AACCCCGTCTCTA CCAAAAAAACAAA AATTAGCTGAGCC TGGTCATGCATGC CTGGAATCCCAAC AACTCGGGAGGCT GAGGCAGGAGAAT CGCTTGAACCCAG GAGGCGGAGATTG CAGTGAGCCAAGA TTGTGCCACTGCA CTCCAGCTTGGTT CCCAATAGACCCC GCAGGCCCTACAG GTTGTCTTCCCAA CTTGCCCCTTGCT CCATACCACCCCC CTCCACCCCATAA TATTATAGAAGGA CACCTAGTCAGAC AAAATGATGCAAC TTAATTTTATTAGG ACAAGGCTGGTG GGCACTGGAGTG GCAACTTCCAGGG CCAGGAGAGGCA

[1135]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
			CTGGGGAGGGGT CACAGGGATGCCA CCCGTCTATG*GCCC TGACTGGAA* (SEQ ID NO: 338)
牛生长激素 (BGH)	CGACTGTGCCTTCTAGTTG CCAGCCATCTGTTGTTTGC CCCTCCCCCGTGCCTTCCT TGACCCTGGAAGGTGCCA CTCCCACTGTCCTTTCCTA ATAAAATGAGGAAATTGC ATCGCATTGTCTGAGTAGG TGTCATTCTATTCTGGGGG GTGGGGTGGGGCAGGAC AGCAAGGGGGAGGATTG GGAAGACAATAGCAGGCA TGCTGGGGATGCGGTGGG CTCTATGG (SEQ ID NO: 339)	RNA 转运至细胞质; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **CCATAGAGCCCAC CGCATCCCCAGCA TGCCTGCTATTGT CTTCCCAATCCTC CCCCTTGCTGTCC TGCCCCACCCCAC CCCCCAGAATAGA ATGACACCTACTC AGACAATGCGATG CAATTTCCTCATTT TATTAGGAAAGGA CAGTGGGAGTGG CACCTTCCAGGGT CAAGGAAGGCAC GGGGGAGGGGCA AACAACAGATGGC TGGCAACTAGAAG GCACAGTCG**CTAT G*GCCCTGACTGGAA* (SEQ ID NO: 340)
兔 BETA-珠蛋白 (RBGLOB)	TTCACTCCTCAGGTGCAG GCTGCCTATCAGAAGGTG GTGGCTGGTGTGGCCAAT GCCCTGGCTCACAAATAC CACTGAGATCTTTTTCCCT CTGCCAAAAATTATGGGG ACATCATGAAGCCCCTTG AGCATCTGACTTCTGGCTA ATAAAGGAAATTTATTTTC ATTGCAATAGTGTGTTGGA ATTTTTTGTGTCTCTCACT CGGAAGGACATATGGGAG GGCAAATCATTTAAAACAT CAGAATGAGTATTTGGTTT	RNA 转运至细胞质; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GATCTCCATAAGA GAAGAGGGACAG CTATGACTGGGAG TAGTCAGGAGAGG AGGAAAAATCTGG CTAGTAAAACATG**

[1136]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
	AGAGTTTGGCAACATATG CCCATATGCTGGCTGCCAT GAACAAAGGTTGGCTATA AAGAGGTCATCAGTATATG AAACAGCCCCCTGCTGTC CATTCCTTATTCCATAGAA AAGCCTTGACTTGAGGTT AGATTTTTTTTATATTTTGT TTTGTGTTATTTTTTTCTTT AACATCCCTAAAATTTTCC TTACATGTTTTACTAGCCA GATTTTTCCTCCTCTCCTG ACTACTCCCAGTCATAGCT GTCCCTCTTCTCTTATGGA GATC (SEQ ID NO: 341)		**TAAGGAAAATTTT AGGGATGTTAAAG AAAAAAATAACAC AAAACAAAATATA AAAAAAATCTAAC CTCAAGTCAAGGC TTTTCTATGGAATA AGGAATGGACAGC AGGGGGCTGTTTC ATATACTGATGAC CTCTTTATAGCCA ACCTTTGTTCATG GCAGCCAGCATAT GGGCATATGTTGC CAAACTCTAAACC AAATACTCATTCT GATGTTTTAAATG ATTTGCCCTCCCA TATGTCCTTCCGA GTGAGAGACACAA AAAATTCCAACAC ACTATTGCAATGA AAATAAATTTCCTT TATTAGCCAGAAG TCAGATGCTCAAG GGGCTTCATGATG TCCCCATAATTTTT GGCAGAGGGAAA AAGATCTCAGTGG TATTTGTGAGCCA GGGCATTGGCCAC ACCAGCCACCACC TTCTGATAGGCAG CCTGCACCTGAGG AGTGAA**CTATG*GCC CTGACTGGAA* (SEQ ID NO: 342)
胸苷激酶(TK)	GGGGGAGGCTAACTGAAA CACGGAAGGAGACAATAC CGGAAGGAACCCGCGCTA TGACGGCAATAAAAAGAC AGAATAAAACGCACGGGT GTTGGGTCGTTTGTTCATA AACGCGGGGTTCGGTCCC AGGGCTGGCACTCTGTCG ATACCCCACCGAGACCCC ATTGGGGCCAATACGCCC	RNA 转运至细胞质; 增强 RNA 稳定性和编码蛋白的表达	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **CTATGGCAGGGCC TGCCGCCCCGACG**

[1137]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
	GCGTTTCTTCCTTTTCCCC ACCCCACCCCCCAAGTTC GGGTGAAGGCCCAGGGCT CGCAGCCAACGTCGGGGC GGCAGGCCCTGCCATAG (SEQ ID NO: 343)		**TTGGCTGCGAGCC CTGGGCCTTCACC CGAACTTGGGGG GTGGGGTGGGGA AAAGGAAGAAAC GCGGGCGTATTGG CCCCAATGGGGTC TCGGTGGGGTATC GACAGAGTGCCA GCCCTGGGACCG AACCCCGCGTTTA TGAACAAACGACC CAACACCCGTGCG TTTTATTCTGTCTT TTTATTGCCGTCAT AGCGCGGGTTCCT TCCGGTATTGTCT CCTTCCGTGTTTC AGTTAGCCTCCCC** CCTATG*GCCCTGAC TGGAA* (SEQ ID NO: 344)
MALAT1 ENE-MASCRN A	TAGGGTCATGAAGGTTTTT CTTTTCCTGAGAAAACAA CACGTATTGTTTTCTCAGG TTTTGCTTTTTGGCCTTTT TCTAGCTTAAAAAAAAAA AAAGCAAAAGATGCTGGT GGTTGGCACTCCTGGTTT CCAGGACGGGGTTCAAAT CCCTGCGGCGTCTTTGCTT TGACT (SEQ ID NO: 345)	导致细胞核中保留 RNA, 转录终止并稳定	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **AGTCAAAGCAAAG ACGCCGCAGGGAT TTGAACCCCGTCC TGGAAACCAGGA GTGCCAACCACCA GCATCTTTTGCTT TTTTTTTTTTTAAG CTAGAAAAAGGCC AAAAAGCAAAACC TGAGAAAACAATA CGTGTTGTTTTCT CAGGAAAAGAAAA ACCTTCATGACCC TA**CTATG*GCCCTGA CTGGAA* (SEQ ID NO: 346)
KSHV PAN	TGTTTTGGCTGGGTTTTTC	导致细胞核中	ATCCTTCCAGTCAG

[1138]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
ENE	CTTGTTCGCACCGGACAC CTCCAGTGACCAGACGGC AAGGTTTTTATCCCAGTGT ATATTGGAAAAACATGTTA TACTTTTGACAATTTAACG TGCCTAGAGCTCAAATTA AACTAATACCATAACGTAA TGCAACTTACAACATAAAT AAAGGTCAATGTTTAATCC ATAAAAAAAAAAAAAAA AAAA (SEQ ID NO: 347)	保留 RNA，转录终止并稳定	GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTTTTTTTTTTT TTTTTATGGATTAA ACATTGACCTTTA TTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACA**CTATG *GCCCTGACTGGAA* (SEQ ID NO: 348)
三个连续的 KSHV PAN ENES，具有短的非保守的 RNA 接头	TGTTTTGGCTGGGTTTTTC CTTGTTCGCACCGGACAC CTCCAGTGACCAGACGGC AAGGTTTTTATCCCAGTGT ATATTGGAAAAACATGTTA TACTTTTGACAATTTAACG TGCCTAGAGCTCAAATTA AACTAATACCATAACGTAA TGCAACTTACAACATAAAT AAAGGTCAATGTTTAATCC ATAAAAAAAAAAAAAAA AAAAACACACTGTTTTGG CTGGGTTTTTCCTTGTTCG CACCGGACACCTCCAGTG ACCAGACGGCAAGGTTTT TATCCCAGTGTATATTGGA AAAACATGTTATACTTTTG ACAATTTAACGTGCCTAG AGCTCAAATTAAACTAATA CCATAACGTAATGCAACTT ACAACATAAATAAAGGTC AATGTTTAATCCATAAAAA	导致细胞核中保留 RNA，转录终止并稳定，预计比单个 PAN ENE 更大	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTTTTTTTTTTT TTTTTATGGATTAA ACATTGACCTTTA TTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA**

[1139]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
	AAAAAAAAAAAAAATCTC TCTGTTTTGGCTGGGTTTT TCCTTGTTCGCACCGGAC ACCTCCAGTGACCAGACG GCAAGGTTTTTATCCCAGT GTATATTGGAAAAACATGT TATACTTTTGACAATTTAA CGTGCCTAGAGCTCAAAT TAAACTAATACCATAACGT AATGCAACTTACAACATA AATAAAGGTCAATGTTTAA TCCATAAAAAAAAAAAAA AAAAAA (SEQ ID NO: 349)		**AGGAAAAACCCAG CCAAAACAGAGAG ATTTTTTTTTTTTT TTTTTTATGGATTA AACATTGACCTTT ATTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACAGTGTG TTTTTTTTTTTTTT TTTTTTATGGATTA AACATTGACCTTT ATTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACA**CTATG *GCCCTGACTGGAA* (SEQ ID NO: 350)
SMBOX/U1 SNRNA 框	CAGCAAGTTCAGAGAAAT CTGAACTTGCTGGATTTTT GGAGCAGGGAGATGGAAT AGGAGCTTGCTCCGTCCA CTCCACGCATCGACCTGG TATTGCAGTACCTCCAGG AACGGTGCACCCACTTTC TGGAGTTTCAAAAGTAGA CTGTACGCTAAGGGTCATA TCTTTTTTTGTTTGGTTTG TGTCTTGGTTGGCGTCTTA	导致细胞核中保留 RNA 和转录终止	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTAAGACGCCAA CCAAGACACAAAC CAAACAAAAAAAG**

[1140]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
	AA (SEQ ID NO: 351)		**ATATGACCCTTAG CGTACAGTCTACT TTTGAAACTCCAG AAAGTGGGTGCAC CGTTCCTGGAGGT ACTGCAATACCAG GTCGATGCGTGGA GTGGACGGAGCA AGCTCCTATTCCA TCTCCCTGCTCCA AAAATCCAGCAAG TTCAGATTTCTCT GAACTTGCTG**CTA TG*GCCCTGACTGGA A* (SEQ ID NO: 352)
U1 SNRNA 3′ 框	GTTTCAAAAGTAGACTGT ACGCTAAGGGTCATATCTT TTTTTGTTTGGTTTGTGTC TTGGTTGGCGTCTTAAA (SEQ ID NO: 353)	导致细胞核中保留 RNA 和转录终止	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTAAGACGCCAA CCAAGACACAAAC CAAACAAAAAAAG ATATGACCCTTAG CGTACAGTCTACT TTTGAAAC**CTATG*G CCCTGACTGGAA* (SEQ ID NO: 354)
TRNA-赖氨酸	GCCCGGCTAGCTCAGTCG GTAGAGCATGAGACTCTT AATCTCAGGGTCGTGGGT TCGAGCCCCACGTTGGGC G (SEQ ID NO: 355)	报告能够将 RNA 转运至线粒体	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **CGCCCAACGTGG GGCTCGAACCCAC GACCCTGAGATTA AGAGTCTCATGCT CTACCGACTGAGC TAGCCGGGC**CTAT G*GCCCTGACTGGAA*

[1141]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
			(SEQ ID NO: 356)
BROCCOLI 适体	GAGACGGTCGGGTCCAGA TATTCGTATCTGTCGAGTA GAGTGTGGGCTC (SEQ ID NO: 357)	可视化(荧光)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GAGCCCACACTCT ACTCGACAGATAC GAATATCTGGACC CGACCGTCTC**CTA TG*GCCCTGACTGGA A* (SEQ ID NO: 358)
SPINACH 适体	GACGCAACTGAATGAAAT GGTGAAGGACGGGTCCA GGTGTGGCTGCTTCGGCA GTGCAGCTTGTTGAGTAG AGTGTGAGCTCCGTAACT AGTCGCGTC (SEQ ID NO: 359)	可视化(荧光)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GACGCGACTAGTT ACGGAGCTCACAC TCTACTCAACAAG CTGCACTGCCGAA GCAGCCACACCTG GACCCGTCCTTCA CCATTTCATTCAG TTGCGTC**CTATG*GC CCTGACTGGAA* (SEQ ID NO: 360)
SPINACH2 适体	GATGTAACTGAATGAAAT GGTGAAGGACGGGTCCA GTAGGCTGCTTCGGCAGC CTACTTGTTGAGTAGAGT GTGAGCTCCGTAACTAGTT ACATC (SEQ ID NO: 361)	可视化(荧光)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GATGTAACTAGTT ACGGAGCTCACAC TCTACTCAACAAG TAGGCTGCCGAAG CAGCCTACTGGAC**

[1142]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
			CCGTCCTTCACCA TTTCATTCAGTTA CATCCTATG*GCCCT GACTGGAA* (SEQ ID NO: 362)
MANGO 适体	GGCACGTACGAAGGGACG GTGCGGAGAGGAGAGTAC GTGC (SEQ ID NO: 363)	可视化(荧光)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GCACGTACTCTCC TCTCCGCACCGTC CCTTCGTACGTGC C**CTATG*GCCCTGAC TGGAA* (SEQ ID NO: 364)
HDV 核酶	GGCCGGCATGGTCCCAGC CTCCTCGCTGGCGCCGGC TGGGCAACATGCTTCGGC ATGGCGAATGGGAC (SEQ ID NO: 365)	3′端 RNA 加工	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GTCCCATTCGCCA TGCCGAAGCATGT TGCCCAGCCGGC GCCAGCGAGGAG GCTGGGACCATGC CGGCC**CTATG*GCCC TGACTGGAA* (SEQ ID NO: 366)
N^6-甲基腺苷标志物(M^6A)	GGACTCTAGGACTGGACT TTGGACT (SEQ ID NO: 367)	靶向甲基化(划线为甲基化的) M6A 甲基化可导致增强 RNA 稳定性和表达，但尚未完全了解	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **AGTCCAAAGTCCA GTCCTAGAGTCC**C TATG*GCCCTGACTG GAA*

[1143]

RNA 基序	RNA 基序的序列	功能/作用	用于引导编辑将 RNA 基序插入示例性 HEXA 基因的示例性 PEGRNA *
			(SEQ ID NO: 368)

[1144] *各PEGRNA以5′至3′方向显示，并具有以下图3F的结构元件，由字体指定，如下：5′-间隔区(正常字体)-gRNA核心(下划线序列)-同源臂(斜体)-RT模板(粗体)-引物结合位点(斜体)-3′

[1145] 将上表的PEgRNA设计为上述基序的位点特异性插入HEXA基因(泰-萨二氏病缺陷)(如，GenBank号KR710351.1(SEQ ID NO:369)的示例，然而，这只是为了说明的目的。引导编辑在RNA标记中的应用不限于HEXA基因，实际上可以是任何基因。HEXA mRNA具有以下核苷酸序列：

[1146] GTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGTTGGCATG
ACAAGTTCCAGGCTTTGGTTTTCGCTGCTGCTGGCGGCAGCGTTCGCAGGACGGGCGACGGCCCTCTGGCCCTGGC
CTCAGAACTTCCAAACCTCCGACCAGCGCTACGTCCTTTACCCGAACAACTTTCAATTCCAGTACGATGTCAGCTC
GGCCGCGCAGCCCGGCTGCTCAGTCCTCGACGAGGCCTTCCAGCGCTATCGTGACCTGCTTTTCGGTTCCGGGTCT
TGGCCCCGTCCTTACCTCACAGGGAAACGGCATACACTGGAGAAGAATGTGTTGGTTGTCTCTGTAGTCACACCTG
GATGTAACCAGCTTCCTACTTTGGAGTCAGTGGAGAATTATACCCTGACCATAAATGATGACCAGTGTTTACTCCT
CTCTGAGACTGTCTGGGGAGCTCTCCGAGGTCTGGAGACTTTTAGCCAGCTTGTTTGGAAATCTGCTGAGGGCACA
TTCTTTATCAACAAGACTGAGATTGAGGACTTTCCCCGCTTTCCTCACCGGGGCTTGCTGTTGGATACATCTCGCC
ATTACCTGCCACTCTCTAGCATCCTGGACACTCTGGATGTCATGGCGTACAATAAATTGAACGTGTTCCACTGGCA
TCTGGTAGATGATCCTTCCTTCCCATATGAGAGCTTCACTTTTCCAGAGCTCATGAGAAAGGGGTCCTACAACCCT
GTCACCCACATCTACACAGCACAGGATGTGAAGGAGGTCATTGAATACGCACGGCTCCGGGGTATCCGTGTGCTTG
CAGAGTTTGACACTCCTGGCCACACTTTGTCCTGGGGACCAGGTATCCCTGGATTACTGACTCCTTGCTACCCTGG
GTCTGAGCCCTCTGGCACCTTTGGACCAGTGAATCCCAGTCTCAATAATACCTATGAGTTCATGAGCACATTCTTC
TTAGAAGTCAGCTCTGTCTTCCCAGATTTTTATCTTCATCTTGGAGGAGATGAGGTTGATTTCACCTGCTGGAAGT
CCAACCCAGAGATCCAGGACTTTATGAGGAAGAAAGGCTTCGGTGAGGACTTCAAGCAGCTGGAGTCCTTCTACAT
CCAGACGCTGCTGGACATCGTCTCTTCTTATGGCAAGGGCTATGTGGTGTGGCAGGAGGTGTTTGATAATAAAGTA
AAGATTCAGCCAGACACAATCATACAGGTGTGGCGAGAGGATATTCCAGTGAACTATATGAAGGAGCTGGAACTGG
TCACCAAGGCCGGCTTCCGGGCCCTTCTCTCTGCCCCCTGGTACCTGAACCGTATATCCTATGGCCCTGACTGGAA
GGATTTCTACGTAGTGGAACCCCTGGCATTTGAAGGTACCCCTGAGCAGAAGGCTCTGGTGATTGGTGGAGAGGCT
TGTATGTGGGGAGAATATGTGGACAACACAAACCTGGTCCCCAGGCTCTGGCCCAGAGCAGGGGCTGTTGCCGAAA
GGCTGTGGAGCAACAAGTTGACATCTGACCTGACATTTGCCTATGAACGTTTGTCACACTTCCGCTGTGAGTTGCT
GAGGCGAGGTGTCCAGGCCCAACCCCTCAATGTAGGCTTCTGTGAGCAGGAGTTTGAACAGACCTGCCCAACTTTC
TTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAAC(SEQ ID NO:369)。

[1147] 相应HEXA蛋白具有以下氨基酸序列:MTSSRLWFSLLLAAAFAGRATALWPWPQNFQTSDQRY
VLYPNNFQFQYDVSSAAQPGCSVLDEAFQRYRDLLFGSGSWPRPYLTGKRHTLEKNVLVVSVVTPGCNQLPTLESV
ENYTLTINDDQCLLLSETVWGALRGLETFSQLVWKSAEGTFFINKTEIEDFPRFPHRGLLLDTSRHYLPLSSILDT
LDVMAYNKLNVFHWHLVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYARLRGIRVLAEFDTPGHTLS
WGPGIPGLLTPCYPGSEPSGTFGPVNPSLNNTYEFMSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSNPEIQDFMRK

KGFGEDFKQLESFYIQTLLDIVSSYGKGYVVWQEVFDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALLS APWYLNRISYGPDWKDFYVVEPLAFEGTPEQKALVIGGEACMWGEYVDNTNLVPRLWPRAGAVAERLWSNKLTSDL TFAYERLSHFRCELLRRGVQAQPLNVGFCEQEFEQT(SEQ ID NO:370)。

[1148] 注意,所得到的RNA基序将包含在HEXA基因的翻译区内,破坏蛋白编码基因的功能。插入的聚腺苷酸化基序会导致转录本过早终止。该位置仅说明潜在的PEgRNA,其可能导致上表所列RNA基序插入被转录的基因组位点内,从而产生RNA产物。

[1149] 与PE一起用于RNA标签化的PEgRNA可由以下表达:U6启动子(在这种情况下,将单个鸟苷添加至PEgRNA的5'端以包括不以G开头的前间隔序列,并将6-7胸腺嘧啶添加到3′端),或pol II启动子,如pCMV(在这种情况下,可能需要通过自切割元件从RNA的5′端去除该启动子的固有转录序列或Csy4基序,并且需要将终止基序添加到RNA的3′端,该终止基序不导致RNA从细胞核中输出,例如以上列出的3′框基序。注意,由于PE该基序不会插入基因组,因为它是退火区的3′)。核心PEgRNA支架加下划线表示,同源和退火区用斜体表示,插入序列为粗体。注意,插入的序列是上述示例的反向互补序列——如下所述,因此这些PEgRNA需要靶向编码链。

[1150] 另外,在一些实施方案中,注意除HDV之外,自切割核酶需要针对给定的靶位点进行调整;换言之,虽然HDV切割紧靠其自身5′端的编码的转录本,但所有其他自切割核酶的切割位点都在核酶本身内。因此,前面和最后约5-10个核苷酸(在某些情况下可能超过10个)实际上是编码序列的一部分。例如,为了切割5N之后的序列5′NNNNNTCATCCTGATAAACTGCAAA3′(SEQ ID NO:371),其中N是任何核苷酸,使用锤头型自切割核酶插入以下序列,其中加下划线的序列形成不完全RNA配对元件。

[1151] 5′NNNNNCAGTTTGTACGGATGACTGATGAGTCCCAAATAGGACGAAACGCGCTTCGGTGCGTCTCAT CCTGATAAACTGCAAA-3′(SEQ ID NO:372)。

[1152] 在该配对元件的长度和性质方面存在明显的灵活性,并且对原始提交中列出的任何非HDV自切割核酶而言都是如此。为安装锤头型核酶以使用具有与上述结构相同的前间隔序列的PEgRN来切割hexAmRNA,可使用以下PEgRNA序列(标签与以上相同):

[1153] 5′ACCTGAACCGTATATCGACGCACCGAAGCGCGTTTCGTCCTATTTGG GACTCATCAGGATATACGGTTCAGGTGATATACGGTTCAGGTGACGCA CCGAAGCGCGTTTCGTCCTATTTGGGACTCATCAGACCTGAACCGTAT ATCATCCTTCCAGTCAGGGCCATGTTTGAGAGCTAGAAATAGCAAGTT TAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCG GTCCACCTGAACCGTATATC**GACGCACCGAAGCGCGTTTCGTCCTA TTTGGGACTCATCAGGATATACGGT**TCAGGTCTATGGCCCTGACTG

[1154] GAA-3′(其中核心PEgRNA支架加下划线,同源和退火区为斜体,插入序列为粗体)(SEQ ID NO:373)。

[1155] 设计用于插入RNA基序的其他PEgRNA可遵循本文所述一般原则。然而,值得注意地,许多RNA基序可能是高度结构化的,这可能使其难以逆转录并插入基因组。尽管对于一些RNA序列,例如简单的发夹,RNA序列本身及其互补序列都是结构化的。然而,对于以上提到的序列不太可能如此。因此,插入这些基序时,可能最好是PEgRNA编码这些序列的反向互补序列,从而将实际编码该基序的DNA序列插入基因组中。类似地,在PEgRNA模板区域中包含自切割核酶会导致加工和低效活性,而包含其反向互补序列则不会。因此,这些PEgRNA可

能必须靶向编码链,而编码其他类型插入物(如,治疗性校正)的PEgRNA理论上能够靶向任一链。

[1156] 另外,注意,对于许多插入的基序,所得到的PEgRNA可能无法由U6启动子转录,因而需要使用其他启动子,如pCMV。同样,较长的PEgRNA也可能不太稳定。较短的基序,如m^6A标志物不会有这个挑战。

[1157] D.双引导编辑在插入免疫表位中的应用

[1158] 双引导编辑器(例如,比如图90所示的实施方案等)也可用作将已知免疫原性表位插入内源或外源基因组DNA中的手段,导致相应蛋白的修饰以用于治疗或生物技术应用(参见图31和图32)。在经典引导编辑或双引导编辑发明之前,这样的插入只能低效地实现,并且来自DSB的indel形成率很高。引导编辑解决了插入编辑导致的高indel形成问题,同时通常提供比HDR更高的效率。这种较低的indel形成率显示引导编辑作为靶向DNA插入方法优于HDR的主要优势,特别是在插入免疫原性表位的所述应用中。表位的长度在几个碱基到几百个碱基的范围内。引导编辑器是在哺乳动物细胞中实现此类靶向插入的有效方法。

[1159] 本发明的关键概念是使用双引导编辑器将包含先前描述的免疫原性表位的核苷酸序列插入内源或外源基因组DNA中进行下调和/或破坏它们的蛋白产物和/或表达细胞类型。用于免疫原性表位插入的核苷酸序列将以产生靶向基因的编码蛋白的融合蛋白和插入的免疫原性表位的相应蛋白翻译的方式靶向基因。作为针对例如破伤风、白喉或麻疹的常规疫苗接种的标准预先免疫的结果,患者的免疫系统之前已接受过训练以识别这些表位。由于融合表位的免疫原性,预期患者的免疫系统会识别和禁用引导编辑的蛋白(不仅仅是插入的表位)以及可能表达它的细胞。

[1160] 最近在广泛应用中探索了使用CRISPR/Cas系统的精确基因组靶向技术,包括将工程化DNA序列插入靶向基因组基因座。以前,同源定向修复(HDR)已用于此应用,需要ssDNA供体模板并通过双链DNA断裂(DSB)启动修复。这种策略提供了在细胞中进行最广泛的可能变化,并且是将大的DNA序列插入哺乳动物细胞的唯一可用方法。然而,HDR受其起始DSB产生的非期望的细胞副作用的阻碍,例如高水平的indel形成、DNA易位、大的缺失和P53激活。除了这些缺点外,HDR还受在许多细胞类型中低效率的限制(T细胞是该观测结果的明显例外)。最近为克服这些缺点所做的努力包括将人Rad51突变体与Cas9D10A切口酶(RDN)融合,从而形成无DSB的HDR系统,该系统具有增加的HDR产物:indel比率和较低的脱靶编辑,但仍受到细胞类型依赖性阻碍且只有适度的HDR编辑效率。

[1161] 最近开发的Cas9与结合PEgRNA的逆转录酶的融合物(“引导编辑器”)代表了新的基因组编辑技术,提供了许多优于现有基因组编辑方法的优势,包括能够安装任何单核苷酸取代,以及以位点特异性方式插入或缺失任何短的核苷酸段(多达至少几十个碱基)。值得注意地,PE编辑通常实现较低的意外indel率。因此,PE基于靶向插入使以前不可能或不切实际的编辑应用成为可能。

[1162] 该特定方面描述了使用引导编辑作为将已知免疫原性表位插入内源或外源基因组DNA的手段的方法,导致用于治疗或生物技术应用的相应蛋白的修饰(参见图31和图32)。在引导编辑发明之前,这样的插入只能低效地实现,并且来自DSB的indel形成率很高。引导编辑解决了插入编辑的高indel形成问题,同时通常提供比HDR更高的效率。这种较低的indel形成率代表作为靶向DNA插入方法的引导编辑优于HDR的主要优势,特别是在所述插

入免疫原性表位的应用中。表位长度在几个碱基到几百个碱基的范围内。引导编辑器是在哺乳动物细胞中实现此类靶向插入的最高效且最清洁的技术。

[1163] 该方面的关键概念是使用引导编辑器将包含先前所述的免疫原性表位的核苷酸序列插入内源或外源基因组DNA中进行下调和/或破坏它们的蛋白产物和/或表达细胞类型。用于免疫原性表位插入的核苷酸序列将以产生靶向基因的编码蛋白和插入的免疫原性表位的相应蛋白翻译的融合蛋白的方式靶向基因。由于针对如破伤风、白喉或麻疹的常规疫苗接种的标准预先免疫，患者的免疫系统之前已接受过训练以识别这些表位。由于融合表位的免疫原性，预计患者的免疫系统会识别和禁用引导编辑的蛋白(不仅仅是插入的表位)以及可能表达它的细胞。

[1164] 与靶向基因的融合将根据需要工程化改造以确保暴露插入的表位蛋白翻译进行免疫系统识别。这可包括导致蛋白翻译的靶向核苷酸插入，产生免疫原性表位与靶向基因的C端融合物，免疫原性表位与靶向基因的N端融合物，或将核苷酸插入基因中，以便在蛋白结构的表面暴露区域内编码免疫原性表位。

[1165] 在靶基因序列和插入的免疫原性表位核苷酸序列之间插入的核苷酸编码的蛋白接头，可能需要经工程化改造为本发明的一部分以促进靶基因的免疫系统识别、细胞运输、蛋白功能或蛋白折叠。这些插入的核苷酸编码的蛋白接头可包括(但不限于)可变长度和序列的XTEN接头或可变长度和序列的甘氨酸-丝氨酸接头。这些工程化接头以前曾被用于成功促进蛋白融合。示例性接头可包括本文所述任何接头，包括氨基酸序列(GGGGS)n(SEQ ID NO:165)、(G)n(SEQ ID NO:166)、(EAAAK)n(SEQ ID NO:167)、(GGS)n(SEQ ID NO:168)、(SGGS)n(SEQ ID NO:169)、(XP)n(SEQ ID NO:170)或其任何组合，其中n独立地为1至30的整数，并且其中X是任何氨基酸。在一些实施方案中，接头包含氨基酸序列(GGS)n(SEQ ID NO:176)，其中n为1、3或7。在一些实施方案中，接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:171)。在一些实施方案中，接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:172)。在一些实施方案中，接头包含氨基酸序列SGGSGGSGGS(SEQ ID NO:173)。在一些实施方案中，接头包含氨基酸序列SGGS(SEQ ID NO:174)。

[1166] 该方面的区别特征包括使用对特定氨基酸序列的先前获得性免疫反应作为诱导针对其他非免疫原性蛋白的免疫反应的手段的能力。另一个区别特征是能够以靶标方式插入这些免疫原性表位的核苷酸序列，该方式不会诱导高水平的不需要的indel作为副产物编辑，并且其插入有效。PE的这种特定应用能够结合细胞类型特异性递送方法(如，AAV血清型)在感兴趣的细胞类型中插入表位以触发免疫反应。

[1167] 可将引导编辑作为将免疫原性表位插入致病基因的手段用于对患者的免疫系统进行编程以对抗多种疾病(不限于癌症，正如免疫肿瘤学策略)。这项技术的直接相关应用将是作为癌症治疗剂，因为它可通过引起对相关致癌基因(如，HER2)或生长因子(如，EGFR)的免疫反应来破坏肿瘤的免疫逃逸机制。这种方法似乎类似于T细胞工程，但这种方法的一个新进展是它可用于许多细胞类型和癌症以外的疾病，而无需生成工程化T细胞并将其引入患者。

[1168] 使用PE将大多数人已针对(破伤风、百日咳、白喉、麻疹、腮腺炎、风疹等)接种疫苗的免疫原性表位插入驱动疾病的外源或内源基因中，因此患者的免疫系统学会了禁用该蛋白。

[1169] 从上述策略中获得潜在治疗益处的疾病包括由有毒蛋白聚集引起的疾病，例如致命性家族性失眠症。其他可能受益的疾病包括由原本无毒的内源性蛋白的致病性过表达引起的疾病，以及由外源病原体引起的疾病。

[1170] 主要治疗的适应症包括上述那些，例如治疗癌症、朊病毒和其他神经退行性疾病、传染病和预防医学。次要治疗的适应症可包括对迟发性遗传病患者的预防性护理。预期当前的护理药物标准可与某些疾病的引导编辑结合使用，例如特别具有侵袭性的癌症，或在药物有助于缓解疾病症状直到疾病完全治愈的情况下。以下是可由本文公开的引导编辑器插入基因中的免疫表位的示例：

[1171]

疫苗	疾病	表位氨基酸序列	示例核酸序列 (8)
1	破伤风类毒素	QYIKANSKFI	CATGATATAAAAGCAAATTCTAAATTTA

[1172]

疫苗	疾病	表位氨基酸序列	示例核酸序列 (8)
		GITEL (SEQ ID NO: 396)	TAGGTATAACTGAACTA (SEQ ID NO: 397)
2	白喉毒素突变体 CRM197	GADDVVDSS KSFVMENFSS YHGTKPGYV DSIQKGIQKP KSGTQGNYD DDWKEFYST DNKYDAAGY SVDNENPLSG KAGGVVKVT YPGLTKVLAL KVDNAETIKK ELGLSLTEPL MEQVGTEEFI KRFGDGASRV VLSLPFAEGSS SVEYINNWEQ AKALSVELEI NFETRGKRGQ DAMYEYMAQ ACAGNRVRRS VGSSLSCINL DWDVIRDKT KTKIESLKEH GPIKNKMSES PNKTVSEEKA KQYLEEFHQT ALEHPELSEL KTVTGTNPVF AGANYAAWA VNVAQVIDSE TADNLEKTTA ALSILPGIGSV MGIADGAVH HNTEEIVAQSI ALSSLMVAQA IPLVGELVDIG FAAYNFVESII NLFQVVHNS YNRPAYSPGH KTQPFLHDGY AVSWNTVEDS IIRTGFQGESG HDIKITAENTP LPIAGVLLPTI PGKLDVNKS	GGCGCCGACGACGTGGTGGACAGCAG CAAGAGCTTCGTGATGGAGAACTTCA GCAGCTACCACGGCACCAAGCCCGGC TACGTGGACAGCATCCAGAAGGGCAT CCAGAAGCCCAAGAGCGGCACCCAGG GCAACTACGACGACGACTGGAAGGAG TTCTACAGCACCGACAACAAGTACGA CGCCGCCGGCTACAGCGTGGACAACG AGAACCCCCTGAGCGGCAAGGCCGGC GGCGTGGTGAAGGTGACCTACCCCGG CCTGACCAAGGTGCTGGCCCTGAAGG TGGACAACGCCGAGACCATCAAGAAG GAGCTGGGCCTGAGCCTGACCGAGCC CCTGATGGAGCAGGTGGGCACCGAGG AGTTCATCAAGAGGTTCGGCGACGGC GCCAGCAGGGTGGTGCTGAGCCTGCC CTTCGCCGAGGGCAGCAGCAGCGTGG AGTACATCAACAACTGGGAGCAGGCC AAGGCCCTGAGCGTGGAGCTGGAGAT CAACTTCGAGACCAGGGGCAAGAGGG GCCAGGACGCCATGTACGAGTACATGG CCCAGGCCTGCGCCGGCAACAGGGTG AGGAGGAGCGTGGGCAGCAGCCTGAG CTGCATCAACCTGGACTGGGACGTGAT CAGGGACAAGACCAAGACCAAGATCG AGAGCCTGAAGGAGCACGGCCCCATC AAGAACAAGATGAGCGAGAGCCCCAA CAAGACCGTGAGCGAGGAGAAGGCC AAGCAGTACCTGGAGGAGTTCCACCA GACCGCCCTGGAGCACCCCGAGCTGA GCGAGCTGAAGACCGTGACCGGCACC AACCCCGTGTTCGCCGGCGCCAACTA CGCCGCCTGGGCCGTGAACGTGGCCC AGGTGATCGACAGCGAGACCGCCGAC AACCTGGAGAAGACCACCGCCGCCCT GAGCATCCTGCCCGGCATCGGCAGCGT GATGGGCATCGCCGACGGCGCCGTGC ACCACAACACCGAGGAGATCGTGGCC CAGAGCATCGCCCTGAGCAGCCTGAT GGTGGCCCAGGCCATCCCCCTGGTGG GCGAGCTGGTGGACATCGGCTTCGCC GCCTACAACTTCGTGGAGAGCATCATC AACCTGTTCCAGGTGGTGCACAACAG CTACAACAGGCCCGCCTACAGCCCCG GCCACAAGACCCAGCCCTTCCTGCAC

疫苗	疾病	表位氨基酸序列	示例核酸序列 (8)
		KTHISVNGRK IRMRCRAIDG DVTFCRPKSP VYVGNGVHA NLHVAFHRSS SEKIHSNEISS DSIGVLGYQK TVDHTKVNS KLSLFFEIKS (SEQ ID NO: 398)	GACGGCTACGCCGTGAGCTGGAACAC CGTGGAGGACAGCATCATCAGGACCG GCTTCCAGGGCGAGAGCGGCCACGAC ATCAAGATCACCGCCGAGAACACCCC CCTGCCCATCGCCGGCGTGCTGCTGCC CACCATCCCCGGCAAGCTGGACGTGA ACAAGAGCAAGACCCACATCAGCGTG AACGGCAGGAAGATCAGGATGAGGTG CAGGGCCATCGACGGCGACGTGACCT TCTGCAGGCCCAAGAGCCCCGTGTAC GTGGGCAACGGCGTGCACGCCAACCT GCACGTGGCCTTCCACAGGAGCAGCA GCGAGAAGATCCACAGCAACGAGATC AGCAGCGACAGCATCGGCGTGCTGGG CTACCAGAAGACCGTGGACCACACCA AGGTGAACAGCAAGCTGAGCCTGTTC TTCGAGATCAAGAGC (SEQ ID NO: 399)
3	流行性腮腺 (MUMPS)炎免疫表位 1	GTYRLIPNAR ANLTA (SEQ ID NO: 400)	GGCACCTACAGGCTGATCCCCAACGC CAGGGCCAACCTGACCGCC (SEQ ID NO: 401)
[1173] 4	流行性腮腺炎免疫表位 2 流行性腮腺炎免疫表位 1	PSKFFTISDSA TFAPGPVSNA (SEQ ID NO: 402) PSKLFIMLDN ATFAPGPVVN A (SEQ ID NO: 404)	CCGAGCAAATTTTTTACCATTAGCGAT AGCGCGACCTTTGCGCCGGGCCCGGT GAGCAACGCG (SEQ ID NO: 403) CCGAGCAAACTGTTTATTATGCTGGAT AACGCGACCTTTGCGCCGGGCCCGGT GGTGAACGCG (SEQ ID NO: 405) 选择的示例来自血凝素–神经氨酸苷酶 (HN) 爆发株系中(表 1) 的多样化 疫苗株 JL5 和爆发株系之间的差异 (表 2)
5	风疹病毒 (RV)	TPPPYQVSCG GESDRASARV IDPAAQS (SEQ ID NO: 406)	ACCCCCCCCCCCTACCAGGTGAGCTGC GGCGGCGAGAGCGACAGGGCCAGCG CCAGGGTGATCGACCCCGCCGCCCAG AGC (SEQ ID NO: 407)
6	血凝素	PEYAYKIVKN KKMEDGFLQ GMVDGWYG HHSNEQGSGL MENERTLDK ANPNNDLCS WSDHEASSN NTNQEDLLQR	CCCGAGTACGCCTACAAGATCGTGAA GAACAAGAAGATGGAGGACGGCTTCC TGCAGGGCATGGTGGACGGCTGGTAC GGCCACCACAGCAACGAGCAGGGCAG CGGCCTGATGGAGAACGAGAGGACCC TGGACAAGGCCAACCCCAACAACGAC CTGTGCAGCTGGAGCGACCACGAGGC CAGCAGCAACAACACCAACCAGGAGG

[1174]

疫苗	疾病	表位氨基酸序列	示例核酸序列 (8)
		ESRRKKRIGT STLNQRGNCN TKCQTEEARL KREEVSLVKS DQCSNGSLQC RANNSTEQV D (SEQ ID NO: 408)	ACCTGCTGCAGAGGGAGAGCAGGAGG AAGAAGAGGATCGGCACCAGCACCCT GAACCAGAGGGGCAACTGCAACACCA AGTGCCAGACCGAGGAGGCCAGGCTG AAGAGGGAGGAGGTGAGCCTGGTGA AGAGCGACCAGTGCAGCAACGGCAGC CTGCAGTGCAGGGCCAACAACAGCAC CGAGCAGGTGGAC (SEQ ID NO: 409)
7	神经氨酸苷酶	TKSTNSRSGG ISGPDNEAPV GEAPSPYGDN PRPNDGNNIRI GSKGYNGIIT DTIEESCSCYP DAKVVKSVE LDSTIWTSGS SPNQKIITIGW DPNGWTGTP MSPNGAYGT DGPSNGQAN QHQAESISAG NSSLCPIRDN WHGSNRSWS WPDGAE (SEQ ID NO: 410)	ACCAAGAGCACCAACAGCAGGAGCG GCGGCATCAGCGGCCCCGACAACGAG GCCCCCGTGGGCGAGGCCCCCAGCCC CTACGGCGACAACCCCAGGCCCAACG ACGGCAACAACATCAGGATCGGCAGC AAGGGCTACAACGGCATCATCACCGA CACCATCGAGGAGAGCTGCAGCTGCT ACCCCGACGCCAAGGTGGTGAAGAGC GTGGAGCTGGACAGCACCATCTGGAC CAGCGGCAGCAGCCCCAACCAGAAGA TCATCACCATCGGCTGGGACCCCAACG GCTGGACCGGCACCCCCATGAGCCCC AACGGCGCCTACGGCACCGACGGCCC CAGCAACGGCCAGGCCAACCAGCACC AGGCCGAGAGCATCAGCGCCGGCAAC AGCAGCCTGTGCCCCATCAGGGACAA CTGGCACGGCAGCAACAGGAGCTGGA GCTGGCCCGACGGCGCCGAG (SEQ ID NO: 411)
8	H5N1 病毒血凝素上的 TAP1 (转运抗原呈递)	EKIVLLLAMM EKIVLLLAKC QTPMGAIKAV DGVTNKCPY LGSPSF (SEQ ID NO: 412)	GAGAAGATCGTGCTGCTGCTGGCCATG ATGGAGAAGATCGTGCTGCTGCTGGCC AAGTGCCAGACCCCCATGGGCGCCAT CAAGGCCGTGGACGGCGTGACCAACA AGTGCCCCTACCTGGGCAGCCCCAGC TTC (SEQ ID NO: 413)
9	H5N1 病毒神经氨酸苷酶上的 TAP2 (转运抗原呈递)	IRPCFWVELN PNQKIITIRPC FWVELICYPD AGEIT (SEQ ID NO: 414)	ATCAGGCCCTGCTTCTGGGTGGAGCTG AACCCCAACCAGAAGATCATCACCATC AGGCCCTGCTTCTGGGTGGAGCTGATC TGCTACCCCGACGCCGGCGAGATCAC C (SEQ ID NO: 415)
10	血凝素 表位 针对 HLA I 类	MEKIVLLLAE KIVLLLAMCP YLGSPSFKCQ TPMGAIKAVD GVTNK (SEQ ID NO: 416)	ATGGAGAAGATCGTGCTGCTGCTGGCC GAGAAGATCGTGCTGCTGCTGGCCATG TGCCCCTACCTGGGCAGCCCCAGCTTC AAGTGCCAGACCCCCATGGGCGCCAT CAAGGCCGTGGACGGCGTGACCAACA AG (SEQ ID NO: 417)

[1175]

疫苗	疾病	表位氨基酸序列	示例核酸序列 (8)
11	神经氨酸苷酶表位 针对 HLA I 类	NPNQKIITICYPDAGEITIRPCFWVELRPCFWVELI (SEQ ID NO: 418)	AACCCCAACCAGAAGATCATCACCATCTGCTACCCCGACGCCGGCGAGATCACCATCAGGCCCTGCTTCTGGGTGGAGCTGAGGCCCTGCTTCTGGGTGGAGCTGATC (SEQ ID NO: 419)
12	血凝素表位 针对 HLA II 类	MVSLVKSDQIGTSTLNQR (SEQ ID NO: 420)	ATGGTGAGCCTGGTGAAGAGCGACCAGATCGGCACCAGCACCCTGAACCAGAGG (SEQ ID NO: 421)
13	神经氨酸苷酶表位 针对 HLA II 类	YNGIITDTI (SEQ ID NO: 422)	TACAACGGCATCATCACCGACACCATC (SEQ ID NO: 423)
14	H5N1 血凝素表位-结合 HLA I 类和 II 类	MEKIVLLLAEKIVLLLAMMVSLVKSDQCPYLGSPSFIGTSTLNQRKCQTPMGAIKAVDGVTNK (SEQ ID NO: 424)	ATGGAGAAGATCGTGCTGCTGCTGGCCGAGAAGATCGTGCTGCTGCTGGCCATGATGGTGAGCCTGGTGAAGAGCGACCAGTGCCCCTACCTGGGCAGCCCCAGCTTCATCGGCACCAGCACCCTGAACCAGAGG (SEQ ID NO: 425)
15	H5N1 神经氨酸苷酶表位-结合 HLA I 类和 II 类	NPNQKIITIYNGIITDTICYPDAGEITIRPCFWVELRPCFWVELI (SEQ ID NO: 426)	AACCCCAACCAGAAGATCATCACCATCTACAACGGCATCATCACCGACACCATCTGCTACCCCGACGCCGGCGAGATCACCATCAGGCCCTGCTTCTGGGTGGAGCTGAGGCCCTGCTTCTGGGTGGAGCTGATC (SEQ ID NO: 427)

[1176] 还可安装本领域已知的其他免疫表位。可通过本文公开的引导编辑器安装可从免疫表位数据库和分析资源(iedb.org/epitopedetails_v3.php)(其内容通过引用并入本文)获得的任何免疫表位。

[1177] 在一些实施方案中,可由本文公开的引导编辑器安装的免疫表位可包括以下表位中的任一种:

[1178]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1	破伤风类毒素	QYIKANSKFIGITEL	396	NA

[1179]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
2	白喉毒素突变体 CRM197	GADDVVDSSKSFVMENFSSYHGT KPGYVDSIQKGIQKPKSGTQGNY DDDWKEFYSTDNKYDAAGYSVD NENPLSGKAGGVVKVTYPGLTKV LALKVDNAETIKKELGLSLTEPLM EQVGTEEFIKRFGDGASRVVLSLP FAEGSSSVEYINNWEQAKALSVEL EINFETRGKRGQDAMYEYMAQA CAGNRVRRSVGSSLSCINLDWDVI RDKTKTKIESLKEHGPIKNKMSES PNKTVSEEKAKQYLEEFHQTALE HPELSELKTVTGTNPVFAGANYAA WAVNVAQVIDSETADNLEKTTAA LSILPGIGSVMGIADGAVHHNTEEI VAQSIALSSLMVAQAIPLVGELVDI GFAAYNFVESIINLFQVVHNSYNR PAYSPGHKTQPFLHDGYAVSWNT VEDSIIRTGFQGESGHDIKITAENT PLPIAGVLLPTIPGKLDVNKSKTHI SVNGRKIRMRCRAIDGDVTFCRP KSPVYVGNGVHANLHVAFHRSSS EKIHSNEISSDSIGVLGYQKTVDH TKVNSKLSLFFEIKS	428	NA
3	流行性腮腺炎 (MUMPS)	GTYRLIPNARANLTA	400	NA
4	流行性腮腺炎	PSKFFTISDSATFAPGPVSNA; PSKLFIMLDNATFAPGPVVNA	402; 404	NA
5	风疹病毒(RV)	TPPPYQVSCGGESDRASARVIDPA AQS	406	NA
6	血凝素	PEYAYKIVKNKKMEDGFLQGMV DGWYGHHSNEQGSGLMENERTL DKANPNNDLCSWSDHEASSNNTN QEDLLQRESRRKKRIGTSTLNQRG NCNTKCQTEEARLKREEVSLVKS DQCSNGSLQCRANNSTEQVD	408	NA
7	神经氨酸苷酶	TKSTNSRSGGISGPDNEAPVGEAP SPYGDNPRPNDGNNIRIGSKGYNG IITDTIEESCSCYPDAKVVKSVELD STIWTSGSSPNQKIITIGWDPNGW TGTPMSPNGAYGTDGPSNGQANQ HQAESISAGNSSLCPIRDNWHGSN RSWSWPDGAE	410	NA
8	H5N1 病毒 血凝素上的 TAP (转运抗原呈递)	EKIVLLLAMMEKIVLLLAKCQTP MGAIKAVDGVTNKCPYLGSPSF	412	NA
9	H5N1 病毒神经氨酸苷酶上的 TAP (转	IRPCFWVELNPNQKIITIRPCFWVE LICYPDAGEIT	414	NA

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	运抗原呈递)			
10	血凝素表位 针对 HLA I 类	MEKIVLLLAEKIVLLLAMCPYLGSPSFKCQTPMGAIKAVDGVTNK	416	NA
11	神经氨酸苷酶表位针对 HLA I 类	NPNQKIITICYPDAGEITIRPCFWVELRPCFWVELI	418	NA
12	血凝素表位 针对 HLA II 类	MVSLVKSDQIGTSTLNQR	420	NA
13	神经氨酸苷酶表位针对 HLA II 类	YNGIITDTI	422	NA
14	H5N1 血凝素表位-结合 HLA I 类和 II 类	MEKIVLLLAEKIVLLLAMMVSLVKSDQCPYLGSPSFIGTSTLNQRKCQTPMGAIKAVDGVTNK	424	NA
15	H5N1 神经氨酸苷酶表位-结合 HLA I 类和 II 类	NPNQKIITIYNGIITDTICYPDAGEITIRPCFWVELRPCFWVELI	426	NA
16	白喉杆菌 (CORYNEBACTERIUM DIPHTHERIAE)	AACAGNRVRRSVGSSLKC	899	SRC280292
[1180] 17	麻疹病毒(麻疹 VIRUS)EDMONSTON 株	AADHCPVVEVNGVTI	900	P69353.1
18	麻疹病毒 EDMONSTON 株	AAHLPTGTPLDID	901	P04851.1
19	百日咳博德特氏菌 (BORDETELLA PERTUSSIS)	AALAVWAGLAVQ	902	Q00879.1
20	麻疹病毒 EDMONSTON 株	AALGVATAAQITAGI	903	P69353.1
21	风疹病毒 (RUBELLA VIRUS)THERIEN	AALLNTPPPYQVSCGGESDRATAR	904	P07566.1
22	风疹病毒	AAQSFTGVVYGTHTT	905	BAA28178.1
23	风疹病毒	ACEVEPAFGHSDAAC	906	BAA28178.1
24	风疹病毒	ACTFWAVNAYSSGGY	907	BAA28178.1
25	风疹病毒	ADDPLLR	908	BAA19893.1
26	风疹病毒	ADDPLLRT	909	CAJ88851.1
27	麻疹病毒 EDMONSTON 株	AEMICDIDTYIVEAG	910	P04851.1

[1181]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
28	麻疹病毒 EDMONSTON 株	AEMICDIDTYIVEAGLASFI	911	P04851.1
29	麻疹病毒 EDMONSTON 株	AEPLLSC	912	P04851.1
30	百日咳博德特氏菌	AFVSTSSSRRYTEVY	913	CAD44970.1
31	百日咳博德特氏菌	AGFIYRETFCITTIYKTGQPAADHYYSKVTA	914	P04979.1
32	风疹病毒	AGLLACCAKCLYYLR	915	BAA28178.1
33	风疹病毒	AHTTSDPWHPPG	916	BAA19893.1
34	麻疹病毒 EDMONSTON 株	AIAKLEDAKELLESS	917	P69353.1
35	麻疹病毒 EDMONSTON 株	AIDNLRASLETTNQA	918	P69353.1
36	百日咳博德特氏菌	AKGVEFR	919	ACI16088.1
37	麻疹病毒 EDMONSTON 株	AKWAVPTTRTDDKLR	920	P08362.1
38	麻疹病毒 EDMONSTON 株	ALAEVLKKPV	921	ABO69699.1
39	麻疹病毒 EDMONSTON 株	ALGVINTLEWIPRFK	922	P08362.1
40	麻疹病毒 EDMONSTON 株	ALHQSMLNSQAIDNL	923	P69353.1
41	麻疹病毒 EDMONSTON 株	ALIGILSLFV	924	ABI54110.1
42	风疹病毒	ALLNTPPPYQVSCGGESDRA	925	CAJ88851.1
43	风疹病毒 M33 株	ALLNTPPPYQVSCGGESDRASARV	926	CAJ88851.1
44	风疹病毒	ALVEGLAPGGGNCHL	927	BAA28178.1
45	百日咳博德特氏菌	AMAAWSERAGEA	928	P04977.1
46	麻疹病毒 EDMONSTON 株	ANCASILCKCYTTGT	929	P69353.1
47	百日咳博德特氏菌	ANPNPYTSRRSV	930	P04977.1
48	风疹病毒	APGPGEVW	931	CAJ88851.1
49	风疹病毒	APLPPHTTERIETRSARHPWRIR	932	ABD64214.1
50	风疹病毒疫苗株 RA27/3	APPMPPQPPRAHGQHYGHHHHQLPFLG	933	CAA33016.1
51	风疹病毒 THERIEN	APPTLPQPPCAHGQHYGHHHHQL	934	P07566.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	株	PFLG		
52	风疹病毒	APPTLPQPPRAHGQHYGHHHHQLPFLG	935	ABD64214.1
53	百日咳博德特氏菌	APQPGPQPPQPPQPQPEAPAPQ	936	P14283.3
54	风疹病毒	AQLASYFNPGGSYYK	937	BAA28178.1
55	麻疹病毒 EDMONSTON 株	ARAAHLPTGTPLD	938	P04851.1
56	麻疹病毒 EDMONSTON 株	ARLVSEIAMHTTEDK	939	P04851.1
57	麻疹病毒 EDMONSTON 株	ARLVSEIAMHTTEDKISRAV	940	P04851.1
58	麻疹病毒 EDMONSTON-B 株	ASDVETAEGGEIHELLR	941	P03422.1
59	麻疹 麻疹病毒属 (MEASLES MORBILLIVIRUS)	ASDVETAEGGEIHELLRLQ	942	ABO69699.1
60	麻疹病毒 EDMONSTON-B 株	ASDVETAEGGEIHELLRLQSR	943	P03422.1
[1182] 61	麻疹 麻疹病毒属	ASDVETAEGGEIHKLLRLQ	944	AAA63285.1
62	风疹病毒 M33 株	ASDVLPGHWLQG	945	NP_740663.1
63	麻疹病毒 EDMONSTON 株	ASELGITAEDARLVS	946	P04851.1
64	麻疹病毒 EDMONSTON 株	ASELGITAEDARLVSEIAMH	947	P04851.1
65	麻疹病毒 EDMONSTON 株	ASESSQDPQDSRR	948	P04851.1
66	麻疹病毒 EDMONSTON 株	ASILCKCYTTGTIIN	949	P69353.1
67	风疹病毒	ASPVCQRHSPDCSRL	950	BAA28178.1
68	百日咳博德特氏菌	ASQARWTGATRA	951	BAF35031.1
69	风疹病毒 THERIEN 株	ASYFNPGGSYYKQYHPTACEVEPAFGHS	952	P07566.1
70	麻疹病毒 EDMONSTON 株	ASYKVMTRSSHQSLV	953	P69353.1
71	百日咳博德特氏菌	ASYVKKPKEDVD	954	ACI16088.1
72	麻疹病毒 EDMONSTON 株	ATAAQITAGIALHQS	955	P69353.1

[1183]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
73	风疹病毒	ATPERPRL	956	CAJ88851.1
74	麻疹病毒 EDMONSTON 株	AVCLGGLIGIPALIC	957	P69353.1
75	麻疹病毒 EDMONSTON 株	AVGPRQAQVSF	958	P04851.1
76	风疹病毒	AVNAYSSGGYAQLAS	959	BAA28178.1
77	风疹病毒	AVSETRQTWAEWAAA	960	BAA28178.1
78	麻疹病毒 EDMONSTON 株	AVTAPDTAADSELRR	961	P04851.1
79	麻疹病毒 EDMONSTON 株	AVTAPDTAADSELRRWIKYT	962	P04851.1
80	百日咳博德特氏菌	AYGGIIKDAPPGAGFIYRETFC	963	P04979.1
81	风疹病毒	CALPLAGLLACCAKC	964	BAA28178.1
82	风疹病毒	CARIWNGTQRACTFW	965	BAA28178.1
83	麻疹病毒 EDMONSTON 株	CARTLVSGSFGNRFI	966	P69353.1
84	百日咳博德特氏菌	CASPYEGRYRDMYDALRBRLLY	967	SRC280066
85	百日咳博德特氏菌	CAVFVRSGQPVIGA	968	AAA83981.1
86	风疹病毒	CCAKCLYYLRGAIAPR	969	BAA28178.1
87	麻疹病毒 EDMONSTON 株	CCRGRCNKKGEQVGM	970	P69353.1
88	风疹病毒	CEIPTDVSCEGLGAW	971	BAA28178.1
89	百日咳博德特氏菌	CFGKDLKRPGSSPMEV	972	P0A3R5.1
90	麻疹 麻疹病毒属	CFQQACKGKIQALCE	973	P06830.1
91	麻疹 麻疹病毒属	CFQQACKGKIQALCENPEWAPLKDNRIPS	974	AAR89413.1
92	风疹病毒	CGGESDRASARVIDP	975	BAA28178.1
93	百日咳博德特氏菌	CITTIYKTGQPAADHYYSKVTA	976	P04979.1
94	麻疹 麻疹病毒属	CKGKIQALCENPEWA	977	AAR89413.1
95	麻疹病毒 EDMONSTON 株	CKPWQESRKNKAQ	978	P04851.1
96	麻疹病毒 EDMONSTON 株	CNKKGEQVGMSRPGL	979	P69353.1

[1184]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
97	风疹病毒	CNVTTEHPFCNTPHG	980	BAA28178.1
98	百日咳博德特氏菌	CQVGSSNSAF	981	P04977.1
99	麻疹 麻疹病毒属	CSGPTTIRGQFS	982	P08362.1
100	百日咳博德特氏菌	CTSPYDGKYWSMYSRL	983	AAA83981.1
101	麻疹病毒 EDMONSTON 株	CVLADSESGGHITHS	984	P08362.1
102	麻疹 麻疹病毒属	CYTTGTIINQDPDKILTYIAADHC	985	AAF02706.1
103	风疹病毒	DADDPLLR	986	CAJ88851.1
104	麻疹病毒 EDMONSTON 株	DARAAHLPTGTPLDI	987	P04851.1
105	麻疹病毒 EDMONSTON 株	DARAAHLPTGTPLDIDTASE	988	P04851.1
106	麻疹病毒 EDMONSTON 株	DCHAPTYLPAEVDGD	989	P08362.1
107	风疹病毒	DCSRLVGATPERPRL	990	BAA28178.1
108	麻疹病毒 EDMONSTON 株	DDKLRMETCFQQACK	991	P08362.1
109	风疹病毒	DDPLLRTA	992	CAJ88851.1
110	风疹病毒	DDPLLRTAPGPGEVW	993	BAA28178.1
111	百日咳博德特氏菌	DDPPATVYRYD	994	P04977.1
112	百日咳博德特氏菌	DDPPATVYRYDSRPPED	995	CAD44970.1
113	百日咳博德特氏菌	DDPPATVYRYDSRPPEDV	996	ACI04548.1
114	麻疹 麻疹病毒属	DEVGLRTPQRFTDLV	997	P06830.1
115	麻疹病毒 EDMONSTON 株	DHCPVVEVNGVTIQV	998	P69353.1
116	麻疹病毒 EDMONSTON 株	DIDTASESSQDPQ	999	P04851.1
117	麻疹病毒 EDMONSTON 株	DINKVLEKLGYSGGD	1000	P69353.1
118	百日咳博德特氏菌	DLIAYKQ	1001	ACI16088.1
119	麻疹病毒 EDMONSTON 株	DLIGQKLGLKLLRYY	1002	P69353.1

[1185]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
120	风疹病毒	DLQKALEAQSRALRAELAA	1003	P07566.1
121	麻疹 麻疹病毒属	DLQYVLATYDTSRVE	1004	P06830.1
122	麻疹病毒 EDMONSTON 株	DLSLRRFMV	1005	P04851.1
123	麻疹病毒 EDMONSTON 株	DLSNCMVALGELKLA	1006	P08362.1
124	风疹病毒	DLVEYIMNYTGNQQSRWGLGSPNC	1007	CAJ88851.1
125	麻疹 麻疹病毒属	DLVKFISDKIKFLNP	1008	AAR89413.1
126	麻疹 麻疹病毒属	DLVKFISTKIKFLNP	1009	SRC280117
127	麻疹病毒 EDMONSTON 株	DLYKSNHNNV	1010	P08362.1
128	百日咳博德特氏菌	DNVLDHLTGR	1011	ACI04548.1
129	百日咳博德特氏菌	DNVLDHLTGRSC	1012	P04977.1
130	百日咳博德特氏菌	DNVLDHLTGRSCQ	1013	P04977.1
131	麻疹 麻疹病毒属	DPDKILTYIAA	1014	AAF02706.1
132	风疹病毒 THERIEN 株	DPGDLVEYIMNYTGNQQSR	1015	P07566.1
133	风疹病毒	DPLLRTAP	1016	CAJ88851.1
134	风疹病毒	DPLLRTAPGPGEVWVTPVIGSQ	1017	CAJ88851.1
135	麻疹病毒 EDMONSTON 株	DPQDSRRSAEPLL	1018	P04851.1
136	麻疹病毒 EDMONSTON 株	DPVIDRLYLSSHRGV	1019	P08362.1
137	麻疹病毒 EDMONSTON 株	DQILRSMKGLSSTSI	1020	P69353.1
138	麻疹 麻疹病毒属	DQYCADVAAEELMNA	1021	P06830.1
139	麻疹病毒 EDMONSTON 株	DSESGGHITH	1022	P08362.1
140	麻疹病毒 EDMONSTON 株	DTASESSQDPQDS	1023	P04851.1
141	风疹病毒	DTVMSVFALASYVQH	1024	BAA28178.1
142	百日咳博德特氏菌	DVFQNGFTAWGNND	1025	P04977.1
143	风疹病毒	DVGAVPPGKFVTAAL	1026	BAA28178.1

[1186]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
144	白喉杆菌	DVNKSKTHISVNGRKI	1027	CAE11230.1
145	风疹病毒	DVSCEGLGAWVPAAP	1028	BAA28178.1
146	风疹病毒 THERIEN 株	DWASPVCQRHSPDCSRLVGATPERPRL	1029	P07566.1
147	麻疹病毒 EDMONSTON 株	EARESYRETGPSR	1030	P04851.1
148	百日咳博德特氏菌	EAVEAERAGRGTG	1031	ACI04548.1
149	麻疹病毒 EDMONSTON 株	EDAKELLESSDQILR	1032	P69353.1
150	麻疹病毒 EDMONSTON 株	EDRRVKQSRGEAR	1033	P04851.1
151	麻疹病毒 EDMONSTON 株	EDSITIPYQGSGKGV	1034	P08362.1
152	风疹病毒	EEAFTYLCTAPGCAT	1035	BAA28178.1
153	麻疹病毒 EDMONSTON 株	EGFNMILGTILAQIW	1036	P04851.1
154	麻疹病毒 EDMONSTON 株	EGFNMILGTILAQIWVLLAK	1037	P04851.1
155	风疹病毒	EHPFCNTPHGQLEVQ	1038	BAA28178.1
156	麻疹病毒 EDMONSTON 株	EISDIEVQDPEGFNM	1039	P04851.1
157	麻疹病毒 EDMONSTON 株	EISDIEVQDPEGFNMILGTI	1040	P04851.1
158	麻疹病毒 EDMONSTON 株	EKPNLSSKRSE	1041	P08362.1
159	麻疹病毒 EDMONSTON 株	ELKLAALCHGEDSIT	1042	P08362.1
160	麻疹 麻疹病毒属	ELMNALVNSTLLETR	1043	P06830.1
161	麻疹病毒 EDMONSTON 株	ELPRL	1044	P04851.1
162	麻疹 麻疹病毒属	ENPEWAPLKDNRIPSYGVLSVDL	1045	AAR89413.1
163	麻疹病毒 EDMONSTON 株	EPIRDALNAMTQNIR	1046	P69353.1
164	麻疹病毒 EDMONSTON 株	EQVGMSRPGLKPDLT	1047	P69353.1
165	风疹病毒	ERPRLRLV	1048	CAJ88851.1

[1187]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
166	风疹病毒	ERPRLRLVDADDPLL	1049	BAA28178.1
167	麻疹病毒 CAM/RB	ESPGQLIQRITDDPDVS	1050	P04851.1
168	麻疹病毒 EDMONSTON 株	ESRGIKARITHVDTE	1051	P69353.1
169	麻疹病毒 EDMONSTON 株	ESSCTFMPEGTVCSQ	1052	P69353.1
170	麻疹病毒 EDMONSTON 株	ESSQDPQDSRRSA	1053	P04851.1
171	麻疹病毒 EDMONSTON 株	ETRTTNQFLAVSKGN	1054	P08362.1
172	麻疹病毒 EDMONSTON 株	EVDGDVKLSSNLVIL	1055	P08362.1
173	麻疹病毒 EDMONSTON-B 株	EVNGVTIQV	1056	P26031.1
174	风疹病毒	EVWVTPVI	1057	CAJ88851.1
175	风疹病毒	EVWVTPVIGSQA	1058	BAA19893.1
176	风疹病毒	EWAAAHWWQLTLGAT	1059	BAA28178.1
177	麻疹 麻疹病毒属	EWIPRFKVSPYLFTV	1060	P06830.1
178	百日咳博德特氏菌	FEYVDTYGDNAG	1061	P04977.1
179	麻疹 麻疹病毒属	FGPLITHGSGMDLYK	1062	P06830.1
180	麻疹病毒 EDMONSTON 株	FIFDALAEV	1063	ABK40531.1
181	麻疹病毒 EDMONSTON 株	FISDKIKFL	1064	P08362.1
182	麻疹病毒 EDMONSTON 株	FKRNKDKPPITSGSG	1065	P04851.1
183	麻疹病毒 EDMONSTON 株	FKRNKDKPPITSGSGGAIRG	1066	P04851.1
184	风疹病毒	FKTVRPVALPRTLAP	1067	BAA28178.1
185	麻疹病毒 EDMONSTON 株	FLMDRHIIV	1068	ABK40531.1
186	麻疹病毒 EDMONSTON 株	FMAVLLTLQTPTGQI	1069	P69353.1
187	麻疹病毒 EDMONSTON 株	FMPEGTVCSQNALYP	1070	P69353.1
188	麻疹	FMYMSLLGV	1071	AAN09804.1

[1188]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	麻疹病毒属			
189	麻疹病毒 EDMONSTON 株	FNVPIKEAGEDCHAP	1072	P08362.1
190	麻疹 麻疹病毒属	FRDLTWCINPPERIK	1073	AAC35876.2
191	麻疹病毒 EDMONSTON 株	FSHDDPISSDQSRFG	1074	P04851.1
192	麻疹病毒 EDMONSTON 株	FSHDDPISSDQSRFGWFENK	1075	P04851.1
193	麻疹 麻疹病毒属	FTDLVKFISDKIKFL	1076	P06830.1
194	麻疹 麻疹病毒属	FTWDQKLWCRHFCVL	1077	P06830.1
195	百日咳博德特氏菌	FVRDGQSVIGACASPYEGRYRDLYDALRRLLY	1078	SRC280066
196	百日咳博德特氏菌	FVRSGQPVIGACTSPYDGKYWSILYSRLRKMLY	1079	SRC280066
197	麻疹 麻疹病毒属	FYKDNPHPKGSRIVI	1080	P06830.1
198	百日咳博德特氏菌	GAASSYFEYVDTYG	1081	ACI04548.1
199	百日咳博德特氏菌	GAFDLKTTFCIMTTRNTGQPA	1082	AAA83981.1
200	麻疹病毒 EDMONSTON 株	GALIGILSLFVESPG	1083	P04851.1
201	麻疹病毒 EDMONSTON 株	GALIGILSLFVESPGQLIQR	1084	P04851.1
202	风疹病毒	GATPERPR	1085	CAJ88851.1
203	百日咳博德特氏菌	GAYGRCPNGTRALTVAELRGNAEL	1086	P04979.1
204	风疹病毒	GCFAPWDLEATGACI	1087	BAA28178.1
205	麻疹 麻疹病毒属	GDINKVLEKLGYS	1088	BAB60865.1
206	麻疹 麻疹病毒属	GDINKVLEKLGYSGGDLLG	1089	AAL29688.1
207	风疹病毒	GDLRAVHHRPVPA	1090	CAA28880.1
208	风疹病毒	GDLVEYIMNYTGNQQ	1091	BAA28178.1
209	麻疹病毒 EDMONSTON 株	GDSSITTRSRLLDRL	1092	P04851.1
210	麻疹病毒 EDMONSTON 株	GDSSITTRSRLLDRLVRLIG	1093	P04851.1

[1189]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
211	麻疹 麻疹病毒属	GEDCHAPTYLPAEVD	1094	P06830.1
212	麻疹病毒 EDMONSTON 株	GELSTLESLMNLYQQ	1095	P04851.1
213	麻疹病毒 EDMONSTON 株	GELSTLESLMNLYQQMGKPA	1096	P04851.1
214	流行性腮腺炎 腮腺炎病毒属 (RUBULAVIRUS)	GEQARYLALLEA	1097	P21186.1
215	风疹病毒	GEVWVT	1098	BAA19893.1
216	风疹病毒	GEVWVTPV	1099	CAJ88851.1
217	风疹病毒	GEVWVTPVIGSQAR	1100	BAA19893.1
218	百日咳博德特氏菌	GEYGGVIKDGTPGGA	1101	AAA83981.1
219	风疹病毒	GFLSGVGPMRLRHGADT	1102	SRC265968
220	麻疹病毒 EDMONSTON-B 株	GFRASDVETAEGGEIHELLRLQ	1103	P03422.1
221	百日咳博德特氏菌	GGAVPGGAVPGGAVPGGFGPGGFGP	1104	P14283.3
222	百日咳博德特氏菌	GGAVPGGAVPGGFGPGGFGPGGFGP	1105	CAA09475.1
223	百日咳博德特氏菌	GGAVPGGAVPGGFGPGGFGPGGFGPGGFGP	1106	CAA09474.1
224	麻疹 麻疹病毒属	GGHITHSGMVGMGVS	1107	P06830.1
225	麻疹 麻疹病毒属	GILESRGIKARITHVDTESY	1108	P26032.1
226	百日咳博德特氏菌	GITGETTTTEYSNARYV	1109	CAD44970.1
227	麻疹病毒 EDMONSTON 株	GKEDRRVKQSRGE	1110	P04851.1
228	百日咳博德特氏菌	GKVTNGS	1111	ACI16088.1
229	风疹病毒	GLGAWVPAAPCARIW	1112	BAA28178.1
230	麻疹病毒 EDMONSTON 株	GLIGIPALICCCRGR	1113	P69353.1
231	风疹病毒 THERIEN 株	GLLACCAKCLYYLRGAIAPR	1114	P07566.1
232	麻疹 麻疹病毒属	GLLAIAGIRLHRAAI	1115	P06830.1

[1190]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
233	风疹病毒	GLQPRADMAAPPTLPQ	1116	NP_740663.1
234	麻疹 麻疹病毒属	GMGVSCTVTREDGTNRR	1117	AAR89413.1
235	麻疹 麻疹病毒属	GMYGGTYLVEKP	1118	AAR89413.1
236	百日咳博德特氏菌	GNAELQTYLRQITPGWSIYGLYDGTYLG	1119	P04979.1
237	风疹病毒	GNCHLTVNGEDVGAV	1120	BAA28178.1
238	百日咳博德特氏菌	GNNDNVLDHLTGR	1121	P04977.1
239	百日咳博德特氏菌	GNNDNVLDHLTGRSC	1122	P04977.1
240	麻疹病毒 EDMONSTON 株	GNRFILSQGNLIANC	1123	P69353.1
241	风疹病毒	GNRGRGQRRDWSRAPPPPEERQETRSQTPAPKPS	1124	P07566.1
242	风疹病毒	GPGEVWVT	1125	CAJ88851.1
243	风疹病毒 THERIEN 株	GPMRLRHGADTRCGRLI	1126	P07566.1
244	百日咳博德特氏菌	GPNHTKV	1127	ACI16083.1
245	麻疹病毒HALLE株	GPRQAQVSF	1128	P10050.1
246	麻疹病毒 EDMONSTON 株	GPRQAQVSFLQGDQS	1129	P04851.1
247	麻疹病毒 EDMONSTON 株	GPRQAQVSFLQGDQSENELP	1130	P04851.1
248	麻疹 麻疹病毒属	GRGYNVSSIVTMTSQ	1131	P06830.1
249	百日咳博德特氏菌	GRTPFII	1132	ACI16083.1
250	风疹病毒	GSPNCHGPDWASPVC	1133	BAA28178.1
251	风疹病毒	GSQARKCGLHIRAGP	1134	BAA28178.1
252	百日咳博德特氏菌	GSSNSAFVSTSSSRR	1135	P04977.1
253	麻疹病毒 EDMONSTON 株	GSTKSCARTLVSGSF	1136	P69353.1
254	风疹病毒	GSYYKQYHPTACEVE	1137	BAA28178.1
255	风疹病毒	GTHTTAVSETRQTWA	1138	BAA28178.1
256	麻疹病毒 EDMONSTON 株	GTIINQDPDKILTYI	1139	P69353.1
257	百日咳博德特氏菌	GTLVRMAPVIG	1140	ADA8512

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				4.1
258	麻疹病毒 EDMONSTON 株	GTPLDIDTASESS	1141	P04851.1
259	百日咳博德特氏菌	GTYLGQAYGGIIKDAPPGAGFIYR ETFC	1142	P04979.1
260	百日咳博德特氏菌	GVATKGLGVHAKSSDWG	1143	P15318.2
261	白喉杆菌	GVLLPTIPGKLDVNKSKTHI	1144	AAV7048 6.1
262	麻疹 麻疹病毒属	GVLSVDLSLTVELKI	1145	P06830.1
263	麻疹 麻疹病毒属	GVPIELQVECFTWDQ	1146	P06830.1
264	麻疹病毒 EDMONSTON 株	GVSCTVTREDGTNRR	1147	P08362.1
265	麻疹病毒 EDMONSTON 株	GVSYNIGSQEWYTTV	1148	P69353.1
266	麻疹病毒 EDMONSTON 株	GYNVSSIVTMTSQGM	1149	P08362.1
267	麻疹 麻疹病毒属	HFCVLADSESGGHIT	1150	P06830.1
[1191] 268	麻疹 麻疹病毒属	HGEDSITIPYQGSGK	1151	P06830.1
269	风疹病毒	HGPDWASP	1152	BAA1989 3.1
270	风疹病毒	HGPDWASPVCQRHSP	1153	BAA2817 8.1
271	风疹病毒	HGPDWASPVCQRHSPDCSRLVG	1154	CAJ88851 .1
272	风疹病毒 M33 株	HGPDWASPVCQRHSPDCSRLVGA TPERPRLRLV	1155	CAJ88851 .1
273	麻疹病毒 EDMONSTON 株	HITHSGMEGMGVSCT	1156	P08362.1
274	麻疹 麻疹病毒属	HKSLSTNLDVTNSIE	1157	P06830.1
275	麻疹病毒 EDMONSTON 株	HLMIDRPYV	1158	P08362.1
276	麻疹病毒 EDMONSTON 株	HLPTGTPLDIDTA	1159	P04851.1
277	麻疹病毒 EDMONSTON-B 株	HLPTGTPLDIDTATESSQDPQDSR	1160	Q77M43. 1
278	麻疹 麻疹病毒属	HMTNYLEQPVSNDLS	1161	P06830.1
279	麻疹病毒 EDMONSTON 株	HQSLVIKLMPNITLL	1162	P69353.1
280	麻疹	HRAAIYTAEIHKSLS	1163	P06830.1

[1192]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	麻疹病毒属			
281	百日咳博德特氏菌	HRMQEAVEAERAGRGTGH	1164	P04977.1
282	麻疹病毒 EDMONSTON 株	HVDTESYFIVLSIAY	1165	P69353.1
283	麻疹病毒 EDMONSTON 株	HWGNLSKIGVVGIGS	1166	P69353.1
284	风疹病毒	HWWQLTLGATCALPL	1167	BAA28178.1
285	风疹病毒	HYRNASDVLPGHWLQGGWGCYNL	1168	NP_740663.1
286	麻疹病毒 EDMONSTON 株	IDLGPPISLERLDVG	1169	P69353.1
287	麻疹病毒 EDMONSTON 株	IEAIRQAGQEMILAV	1170	P69353.1
288	风疹病毒 M33 株	IETRSARHP	1171	CAA28880.1
289	麻疹病毒 EDMONSTON 株	IGSQEWYTTVPKYVA	1172	P69353.1
290	麻疹病毒 EDMONSTON 株	IKGVIVHRLEGVSYN	1173	P69353.1
291	麻疹病毒 EDMONSTON 株	IKHIIIVPIPGDSSI	1174	P04851.1
292	麻疹病毒 EDMONSTON 株	IKHIIIVPIPGDSSITTRSR	1175	P04851.1
293	百日咳博德特氏菌	IKLKDCP	1176	ACI16083.1
294	麻疹病毒 EDMONSTON 株	IKLMPNITLLNNCTR	1177	P69353.1
295	麻疹 麻疹病毒属	ILLERLDVGT	1178	AAF85664.1
296	麻疹 麻疹病毒属	ILPGQDLQYV	1179	P08362.1
297	麻疹病毒 EDMONSTON 株	ILTYIAADHCPVVEV	1180	P69353.1
298	麻疹 麻疹病毒属	INQDPDKILTY	1181	AAL29688.1
299	麻疹病毒 EDMONSTON 株	IPRFKVSPYLFNVPI	1182	P08362.1
300	麻疹病毒 EDMONSTON 株	IQALSYALGGDINKV	1183	P69353.1
301	风疹病毒	IRAGPYGHATVEMPE	1184	BAA28178.1
302	百日咳博德特氏菌	IRMGTDK	1185	ACI16088.1

[1193]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
303	麻疹病毒 EDMONSTON 株	ISNFDESSCTFMPEG	1186	P69353.1
304	麻疹病毒 EDMONSTON 株	ITAGIALHQSMLNSQ	1187	P69353.1
305	麻疹病毒 EDMONSTON 株	ITDDPDVSIRLLEVV	1188	P04851.1
306	麻疹病毒 EDMONSTON 株	ITDDPDVSIRLLEVVQSDQS	1189	P04851.1
307	百日咳博德特氏菌	ITTYV	1190	ACI16083.1
308	麻疹病毒 EDMONSTON 株	IVEAGLASFILTIKF	1191	P04851.1
309	麻疹病毒 EDMONSTON 株	IVEAGLASFILTIKFGIETM	1192	P04851.1
310	麻疹病毒 EDMONSTON 株	IVEAGLASFILTIKFGIETMYPALG	1193	P04851.1
311	风疹病毒	KALEAQSRALRAELAA	1194	P07566.1
312	麻疹病毒 EDMONSTON 株	KARITHVDTESYFIV	1195	P69353.1
313	风疹病毒	KCGLHIRAGPYGHAT	1196	BAA28178.1
314	麻疹病毒 EDMONSTON 株	KCYTTGTIINQDPDK	1197	P69353.1
315	麻疹病毒 EDMONSTON 株	KDNPHPKGSR	1198	P08362.1
316	麻疹病毒 EDMONSTON 株	KDNRIPSYGVLSVDL	1199	P08362.1
317	麻疹 麻疹病毒属	KFLNPDREYDFRDLT	1200	AAC35876.2
318	风疹病毒	KFVTAALLN	1201	BAA28178.1
319	麻疹病毒 EDMONSTON 株	KGNCSGPTTIR	1202	P08362.1
320	麻疹 麻疹病毒属	KIKFLNPDREYDFRD	1203	P06830.1
321	风疹病毒	KIVDGGCFAPWDLEA	1204	BAA28178.1
322	麻疹病毒 EDMONSTON 株	KLGLKLLRYYTEILS	1205	P69353.1
323	麻疹病毒 EDMONSTON 株	KLGVWKSPTDMQSWV	1206	P08362.1
324	百日咳博德特氏菌	KLKECPQ	1207	ACI16088.1

[1194]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
325	麻疹 麻疹病毒属	KLLRYYTEI	1208	P26031.1
326	麻疹病毒 EDMONSTON 株	KLMPFSGDFV	1209	ABK40531.1
327	麻疹 麻疹病毒属	KLMPNITLL	1210	P26031.1
328	麻疹 麻疹病毒属	KLRMETCFQQACKGKIQALCENPEWA	1211	AAR89413.1
329	麻疹 麻疹病毒属	KLWCRHFCV	1212	P08362.1
330	麻疹病毒 EDMONSTON 株	KLWCRHFCVL	1213	P08362.1
331	麻疹 麻疹病毒属	KLWESPQEI	1214	BAB60863.1
332	麻疹病毒 EDMONSTON 株	KMSSAVGFV	1215	ABO69699.1
333	麻疹病毒 EDMONSTON-B 株	KMSSAVGFVPDTGPASR	1216	P03422.1
334	百日咳博德特氏菌	KMVYATN	1217	ACI16083.1
335	麻疹病毒 EDMONSTON 株	KPDLTGTSKSYVRSL	1218	P69353.1
336	麻疹 麻疹病毒属	KPNLSSKRSELSQLS	1219	P08362.1
337	麻疹病毒 EDMONSTON 株	KPNLSSKRSELSQLSMYRVF	1220	P08362.1
338	麻疹病毒 EDMONSTON 株	KQSRGEARESYRETG	1221	P04851.1
339	麻疹病毒 EDMONSTON 株	KQSRGEARESYRETGPSRAS	1222	P04851.1
340	麻疹病毒 EDMONSTON 株	KRFAGVVLAGAALGV	1223	P69353.1
341	麻疹病毒 EDMONSTON 株	KRTPGNKPRIAEMIC	1224	P04851.1
342	麻疹病毒 EDMONSTON 株	KRTPGNKPRIAEMICDIDTY	1225	P04851.1
343	麻疹 麻疹病毒属	KSNHNNVYWLTIPPMKNLALGVINTL	1226	AAR89413.1
344	麻疹 麻疹病毒属	KVSPYLFNV	1227	P08362.1
345	百日咳博德特氏菌	KVVQLPKISKNALKANG	1228	ACI16083.1
346	百日咳博德特氏菌	KVVQLPKISKNALRNDG	1229	ACI16087

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				.1
347	百日咳博德特氏菌	LAHRRIPPENIR	1230	P04977.1
348	百日咳博德特氏菌	LALALWAGFALS	1231	P11092.1
349	风疹病毒	LAPGGGNCHLTVNGE	1232	BAA28178.1
350	麻疹病毒 EDMONSTON 株	LAQIWVLLAKAVTAP	1233	P04851.1
351	麻疹病毒 EDMONSTON 株	LAQIWVLLAKAVTAPDTAAD	1234	P04851.1
352	风疹病毒	LASYFNPGGSYYKQYHPTACEVEPAFGHS	1235	BAA19893.1
353	麻疹 麻疹病毒属	LAVSKGNCSGPTTIR	1236	P06830.1
354	麻疹病毒 EDMONSTON 株	LCENPEWAPLKDNRI	1237	P08362.1
355	麻疹 麻疹病毒属	LDRLVRLIG	1238	ABI54110.1
356	白喉杆菌	LEEEGVTPL	1239	P33120.2
357	百日咳博德特氏菌	LEHRMQEAVEAERAGRGTGHFI	1240	CAD44970.1
358	麻疹病毒 EDMONSTON 株	LEKLGYSGGDLLGIL	1241	P69353.1
359	麻疹病毒 EDMONSTON 株	LEQPVSNDLS	1242	P08362.1
360	麻疹病毒 EDMONSTON 株	LERKWLDVVRNIIAE	1243	P04851.1
361	麻疹病毒 EDMONSTON 株	LERKWLDVVRNIIAEDLSLR	1244	P04851.1
362	麻疹病毒 EDMONSTON 株	LFGPSLRDPISAEIS	1245	P69353.1
363	麻疹 麻疹病毒属	LGELKLAALCHGEDS	1246	P06830.1
364	麻疹病毒 EDMONSTON 株	LGGKEDRRVKQSR	1247	P04851.1
365	风疹病毒	LGHDGHHGGTLRVGQHHRNASDVL	1248	ABD64214.1
366	风疹病毒 THERIEN 株	LGSPNCHGPDWASPVCQRHS	1249	P07566.1
367	风疹病毒 THERIEN 株	LGSPNCHGPDWASPVCQRHSPDCSRLV	1250	P07566.1
368	风疹病毒	LHDPDTEAPTEACVTSWL	1251	ABD64214.1
369	麻疹病毒 EDMONSTON 株	LIANCASILCKCYTT	1252	P69353.1
370	麻疹	LIGLLAIAGIRLHRAAIYTAEIHK	1253	AAR8941

[1195]

[1196]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	麻疹病毒属			3.1
371	麻疹病毒 EDMONSTON 株	LIPSMNQLSCDLIGQ	1254	P69353.1
372	麻疹病毒 EDMONSTON 株	LKIKIASGFGPLITH	1255	P08362.1
373	麻疹 麻疹病毒属	LKIKIASGFGPLITHGSGMDLYK	1256	AAR89413.1
374	百日咳博德特氏菌	LKLYFEP	1257	ACI16088.1
375	麻疹病毒 EDMONSTON 株	LLAVLFVMFL	1258	P08362.1
376	麻疹病毒 EDMONSTON 株	LLDRLVRLIGNPDVS	1259	P04851.1
377	麻疹病毒 EDMONSTON 株	LLDRLVRLIGNPDVSGPKLT	1260	P04851.1
378	麻疹病毒 EDMONSTON 株	LLESSDQILRSMKGL	1261	P69353.1
379	麻疹 麻疹病毒属	LLETRTTNQFLAVSK	1262	P06830.1
380	麻疹病毒 EDMONSTON 株	LLEVVQSDQSQSGLT	1263	P04851.1
381	麻疹病毒 EDMONSTON 株	LLEVVQSDQSQSGLTFASR	1264	P04851.1
382	麻疹病毒 EDMONSTON 株	LLEVVQSDQSQSGLTFASRG	1265	P04851.1
383	麻疹 麻疹病毒属	LLGILESRGIKARIT	1266	AAL29688.1
384	风疹病毒	LLRTAPGP	1267	CAJ88851.1
385	麻疹病毒 EDMONSTON 株	LLRYYTEILSLFGPS	1268	P69353.1
386	风疹病毒 THERIEN 株	LLVPWVLIFMVCRRACRRRG	1269	P07566.1
387	麻疹病毒 EDMONSTON 株	LLWRSRCKIV	1270	ABK40528.1
388	麻疹 麻疹病毒属	LLWSYAMGV	1271	P04851.1
389	麻疹病毒 EDMONSTON 株	LLWSYAMGVGVELEN	1272	P04851.1
390	麻疹病毒 EDMONSTON 株	LLWSYAMGVGVELENSMGGL	1273	P04851.1
391	麻疹 麻疹病毒属	LMIDRPYVL	1274	P08362.1
392	麻疹病毒 EDMONSTON 株	LNAMTQNIRPVQSVA	1275	P69353.1
393	风疹病毒	LNTPPPYQVSCGGES	1276	BAA2817

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				8.1
394	麻疹病毒 EDMONSTON 株	LRDPISAEISIQALS	1277	P69353.1
395	百日咳博德特氏菌	LRGSGDLQEYLRHVTR	1278	AAA83981.1
396	风疹病毒	LRLVDADD	1279	CAJ88851.1
397	风疹病毒	LRLVDADDPLLR	1280	BAA19893.1
398	风疹病毒	LRLVDADDPLLRTAPGPGEVWVTPVIGSQAR	1281	BAA19893.1
399	百日咳博德特氏菌	LRRLLYMIYMSGLAVRVHVSKEEQYYDY	1282	P04979.1
400	风疹病毒	LRTAPGPG	1283	CAJ88851.1
401	风疹病毒	LRVGQHYRNASDVLPGHWLQ	1284	NP_740663.1
402	百日咳博德特氏菌	LRYLA	1285	ACI16088.1
403	麻疹病毒 EDMONSTON 株	LSCKPWQESRKNK	1286	P04851.1
[1197] 404	麻疹 麻疹病毒属	LSEIKGVIVHRLEGV	1287	AAL29688.1
405	麻疹病毒 EDMONSTON 株	LSIAYPTLSEIKGVI	1288	P69353.1
406	麻疹病毒 EDMONSTON 株	LSLLDLYLGRGYNVS	1289	P08362.1
407	麻疹病毒 EDMONSTON 株	LSQGNLIANCASILC	1290	P69353.1
408	麻疹 麻疹病毒属	LSSHRGVIADNQAKW	1291	P06830.1
409	麻疹病毒 EDMONSTON 株	LSVDLSLTVELKIKI	1292	P08362.1
410	百日咳博德特氏菌	LTGISICNPGSSLC	1293	AAA83981.1
411	麻疹病毒 EDMONSTON 株	LTIKFGIETMYPALG	1294	P04851.1
412	麻疹病毒 EDMONSTON 株	LTIKFGIETMYPALGLHEFA	1295	P04851.1
413	麻疹病毒 EDMONSTON 株	LTLQTPTGQIHWGNL	1296	P69353.1
414	风疹病毒	LVDADDPL	1297	CAJ88851.1
415	风疹病毒	LVDADDPLLR	1298	BAA19893.1

[1198]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
416	麻疹病毒 EDMONSTON 株	LVEKPNLSSKRSELS	1299	P08362.1
417	风疹病毒	LVGATPE	1300	BAA19893.1
418	风疹病毒	LVGATPER	1301	CAJ88851.1
419	麻疹 麻疹病毒属	LVKLGVWKSPTGMQS	1302	P06830.1
420	麻疹病毒 EDMONSTON-B 株	LVSGSFGNRFILSQGNLI	1303	P26031.1
421	麻疹病毒 EDMONSTON 株	LYKSNHNNVYWLTIP	1304	P08362.1
422	麻疹病毒 EDMONSTON 株	LYPMSPLLQECLRGSTKSCARTLVS	1305	P69353.1
423	风疹病毒	MASTTPITMEDLQKALEA	1306	P07566.1
424	风疹病毒	MASTTPITMEDLQKALEAQSR	1307	ABD64200.1
425	风疹病毒 THERIEN 株	MASTTPITMEDLQKALEAQSRALRAELAA	1308	P07566.1
426	风疹病毒	MASTTPITMEDLQKALEAQSRALRAGLAA	1309	ABD64200.1
427	风疹病毒疫苗株 RA27/3	MASTTPITMEDLQKALETQSRVLRAGLTA	1310	CAA33016.1
428	麻疹病毒 EDMONSTON 株	MATLLRSLALFKRNK	1311	P04851.1
429	麻疹病毒 EDMONSTON 株	MATLLRSLALFKRNKDKPPI	1312	P04851.1
430	麻疹 麻疹病毒属	MDLYKSNHNNVYWLT	1313	P06830.1
431	风疹病毒	MEDLQKALEAQSRA	1314	P07566.1
432	风疹病毒	MEDLQKALEAQSRALRAELAA	1315	P07566.1
433	麻疹病毒 EDMONSTON 株	MGLKVNVSAIFMAVL	1316	P69353.1
434	麻疹 麻疹病毒属	MIDRPYVLLAVLFVM	1317	P06830.1
435	麻疹病毒 EDMONSTON 株	MILAVQGVQDYINNE	1318	P69353.1
436	麻疹病毒 EDMONSTON 株	MLNSQAIDNLRASLE	1319	P69353.1
437	麻疹病毒 EDMONSTON 株	MNALVNSTLLETRTT	1320	P08362.1
438	风疹病毒	MNYTGNQQSRWGLGSPNCHGPDWASPVCQRHS	1321	BAA19893.1
439	麻疹病毒	MQSWVPLSTDDPVID	1322	P08362.1

[1199]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	EDMONSTON 株			
440	麻疹 麻疹病毒属	MSLSLLDLYLGRGYN	1323	P06830.1
441	麻疹病毒 EDMONSTON 株	MSPLLQECLRGSTKS	1324	P69353.1
442	麻疹 麻疹病毒属	MSPQRDRINAFYKDN	1325	P06830.1
443	麻疹 麻疹病毒属	MYRVFEVSVIRNPGL	1326	P06830.1
444	麻疹病毒 EDMONSTON 株	NALYPMSPLLQECLR	1327	P69353.1
445	风疹病毒 THERIEN 株	NCHGPDWASPVCQRHSPDCSRLVGAT	1328	P07566.1
446	麻疹病毒 EDMONSTON 株	NFGRSYFDPAYFRLG	1329	P04851.1
447	麻疹病毒 EDMONSTON 株	NFGRSYFDPAYFRLGQEMVR	1330	P04851.1
448	风疹病毒	NGTQRACTFWAVNAY	1331	BAA28178.1
449	麻疹病毒 EDMONSTON 株	NGVTIQVGSRRYPDA	1332	P69353.1
450	麻疹病毒 EDMONSTON 株	NIIAEDLSLRRFMVA	1333	P04851.1
451	麻疹病毒 EDMONSTON 株	NIIAEDLSLRRFMVALILDI	1334	P04851.1
452	麻疹病毒 EDMONSTON 株	NITLLNNCTRVEIAE	1335	P69353.1
453	麻疹 麻疹病毒属	NLALGVINTLEWIPR	1336	P06830.1
454	白喉杆菌	NLFQVVHWSYNRPAYSPG	1337	SRC280292
455	麻疹病毒 EDMONSTON 株	NLVILPGQDLQYVLA	1338	P08362.1
456	麻疹病毒 EDMONSTON 株	NLYQQMGKPAPYMVN	1339	P04851.1
457	麻疹病毒 EDMONSTON 株	NLYQQMGKPAPYMVNLENSI	1340	P04851.1
458	麻疹病毒 EDMONSTON 株	NNCTRVEIAEYRRLL	1341	P69353.1
459	麻疹病毒 EDMONSTON 株	NPDREYDFRD	1342	P08362.1
460	麻疹病毒 HALLE 株	NPDVSGPKL	1343	P10050.1
461	麻疹病毒	NPDVSGPKLTGALIG	1344	P04851.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	EDMONSTON 株			
462	麻疹病毒 EDMONSTON 株	NPDVSGPKLTGALIGILSLF	1345	P04851.1
463	麻疹 麻疹病毒属	NPPERIKLDYDQYCA	1346	P06830.1
464	麻疹 麻疹病毒属	NQAKWAVPTTRTDDK	1347	P06830.1
465	麻疹 麻疹病毒属	NQDPDKILTYIAADH	1348	AAF02706.1
466	麻疹病毒 EDMONSTON 株	NQLSCDLIGQKLGLK	1349	P69353.1
467	风疹病毒	NQQSRWGLGSPNCHGPDWASPVCQRHS	1350	ABD64214.1
468	流行性腮腺炎 腮腺炎病毒属	NSTLGVKSAREF	1351	ABP48111.1
469	风疹病毒	NTPHGQLEVQVPPDP	1352	BAA28178.1
470	麻疹病毒 EDMONSTON 株	NVSAIFMAVLLTLQT	1353	P69353.1
471	麻疹 麻疹病毒属	PAEVDGDVKLSSNLV	1354	P06830.1
472	风疹病毒	PAFGHSDAACWGFPT	1355	BAA28178.1
473	麻疹病毒 EDMONSTON 株	PALICCCRGRCNKKG	1356	P69353.1
474	麻疹 麻疹病毒属	PDKILTYIAADHC	1357	AAF02706.1
475	麻疹病毒 EDMONSTON 株	PERIKLDYDQYCADV	1358	P08362.1
476	风疹病毒	PERPRLRL	1359	CAJ88851.1
477	风疹病毒	PGCATQAPVPVRLAG	1360	BAA28178.1
478	风疹病毒疫苗株 RA27/3	PGCATQAPVPVRLAGVRFESKIVDGGCFA	1361	CAJ88851.1
479	风疹病毒	PGEVWVTP	1362	CAJ88851.1
480	风疹病毒	PGEVWVTPVIGSQAR	1363	BAA28178.1
481	白喉杆菌	PGKLDVNKSKTHISVN	1364	CAE11230.1
482	麻疹病毒 EDMONSTON 株	PGLGAPVFHMTNYLE	1365	P08362.1
483	风疹病毒	PGPGEVWV	1366	CAJ88851.1

[1200]

[1201]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
484	风疹病毒	PHKTVRVKFHTETRT	1367	BAA28178.1
485	麻疹 麻疹病毒属	PIELQVECFTWDQKL	1368	AAR89413.1
486	麻疹病毒 EDMONSTON-B 株	PISLERLDVG	1369	P26031.1
487	麻疹病毒 EDMONSTON 株	PISLERLDVGTNLGN	1370	P69353.1
488	百日咳博德特氏菌	PKALFTQQGGAYGRC	1371	P04979.1
489	麻疹病毒 EDMONSTON 株	PKYVATQGYLISNFD	1372	P69353.1
490	麻疹病毒 EDMONSTON 株	PLDIDTASESSQD	1373	P04851.1
491	风疹病毒	PLGLKFKTVRPVALP	1374	BAA28178.1
492	麻疹病毒 EDMONSTON 株	PLITHGSGMDLYKSN	1375	P08362.1
493	麻疹 麻疹病毒属	PLKDNRIPSYGVLSV	1376	P06830.1
494	风疹病毒	PLLRTAPG	1377	CAJ88851.1
495	麻疹病毒 EDMONSTON 株	PLLSCKPWQESRK	1378	P04851.1
496	百日咳博德特氏菌	PPATVYRYDSRPPE	1379	P04977.1
497	麻疹 麻疹病毒属	PPISLERLDVGT	1380	AAL29688.1
498	风疹病毒	PPPPEERQETRSQTPAPKPS	1381	P07566.1
499	百日咳博德特氏菌	PQEQITQHGSPYGRC	1382	AAA83981.1
500	百日咳博德特氏菌	PQEQITQHGSPYGRCANK	1383	AAA83981.1
501	百日咳博德特氏菌	PQPGPQPPQPPQPQPEAPAPQPPAG	1384	P14283.3
502	麻疹病毒 EDMONSTON 株	PRLGGKEDRRVKQ	1385	P04851.1
503	风疹病毒	PRLRLVDA	1386	CAJ88851.1
504	风疹病毒	PRNVRVTGCYQCGTP	1387	BAA28178.1
505	麻疹病毒 EDMONSTON 株	PSRASDARAAHLP	1388	P04851.1
506	麻疹病毒 EDMONSTON 株	PTGQIHWGNLSKIGV	1389	P69353.1
507	麻疹病毒 EDMONSTON 株	PTGTPLDIDTASE	1390	P04851.1
508	麻疹病毒	PTLSEIKGVIVHRLE	1391	P69353.1

[1202]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	EDMONSTON 株			
509	麻疹 麻疹病毒属	PTTIRGQFSNMSLSL	1392	P06830.1
510	麻疹 麻疹病毒属	PTTRTDDKLR	1393	AAR89413.1
511	麻疹 麻疹病毒属	PTTRTDDKLRMETCFQQACKG	1394	AAR89413.1
512	风疹病毒	PVALPRTLAPPRNVR	1395	BAA28178.1
513	白喉杆菌	PVFAGANYAAWAVNVAQVI	1396	AAV70486.1
514	风疹病毒	PVIGSQAR	1397	CAJ88851.1
515	麻疹 麻疹病毒属	PVVEVNGVTIQVGSR	1398	AAL29688.1
516	风疹病毒 THERIEN 株	PWELVVLTARPEDGWTCRGV	1399	P07566.1
517	麻疹病毒 EDMONSTON 株	PWQESRKNKAQTR	1400	P04851.1
518	麻疹病毒 EDMONSTON 株	PYMVNLENSIQNKFS	1401	P04851.1
519	麻疹病毒 EDMONSTON 株	PYMVNLENSIQNKFSAGSYP	1402	P04851.1
520	麻疹病毒 EDMONSTON 株	PYQGSGKGVS	1403	P08362.1
521	风疹病毒	PYQVSCGGESDRASA	1404	BAA28178.1
522	麻疹 麻疹病毒属	QACKGKIQALCEN	1405	P08362.1
523	麻疹病毒 EDMONSTON 株	QAGQEMILAVQGVQD	1406	P69353.1
524	麻疹 麻疹病毒属	QALCENPECVPLKDN	1407	P06830.1
525	百日咳博德特氏菌	QALGALK	1408	ACI16088.1
526	风疹病毒	QAPVPVRLAGVRFES	1409	BAA28178.1
527	风疹病毒	QCGTPALVEGLAPGG	1410	BAA28178.1
528	麻疹病毒 EDMONSTON 株	QDPDKILTYIAADHC	1411	P69353.1
529	麻疹病毒 EDMONSTON 株	QECLRGSTKSCARTL	1412	P69353.1
530	百日咳博德特氏菌	QEQITQHGSPYGRC	1413	AAA8398

[1203]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				1.1
531	麻疹病毒 EDMONSTON 株	QESRKNKAQTRTP	1414	P04851.1
532	麻疹病毒 EDMONSTON 株	QGDQSENELPRLGGK	1415	P04851.1
533	麻疹病毒 EDMONSTON 株	QGDQSENELPRLGGKEDRRV	1416	P04851.1
534	白喉杆菌	QGESGHDIKITAENTPLPIA	1417	AAV70486.1
535	麻疹 麻疹病毒属	QGSGKGVSFQLVKLG	1418	P06830.1
536	麻疹病毒 EDMONSTON 株	QGVQDYINNELIPSM	1419	P69353.1
537	风疹病毒	QLEVQVPPDPGDLVE	1420	BAA28178.1
538	麻疹 麻疹病毒属	QLPEATFMV	1421	ABK40528.1
539	风疹病毒	QLPFLGHDGHHGGTLRVGQHYRNAS	1422	NP_740663.1
540	麻疹病毒 EDMONSTON 株	QLSMYRVFEV	1423	P08362.1
541	百日咳博德特氏菌	QLSNIT	1424	ACI16083.1
542	麻疹病毒 EDMONSTON 株	QNIRPVQSVASSRRH	1425	P69353.1
543	麻疹病毒 EDMONSTON 株	QNKFSAGSYPLLWSY	1426	P04851.1
544	麻疹病毒 EDMONSTON 株	QNKFSAGSYPLLWSYAMGVG	1427	P04851.1
545	麻疹病毒 EDMONSTON 株	QNKFSAGSYPLLWSYAMGVGVELEN	1428	P04851.1
546	麻疹病毒 EDMONSTON 株	QQACKGKIQALCENP	1429	P08362.1
547	麻疹病毒 EDMONSTON 株	QQRRVVGEFRLERKW	1430	P04851.1
548	麻疹病毒 EDMONSTON 株	QQRRVVGEFRLERKWLDVVR	1431	P04851.1
549	百日咳博德特氏菌	QQTRANPNPYTSRRSVAS	1432	P04977.1
550	风疹病毒	QRHSPDCSRLVGATP	1433	BAA28178.1
551	麻疹病毒 EDMONSTON 株	QSGLTFASRGTNMED	1434	P04851.1
552	麻疹病毒 EDMONSTON 株	QSGLTFASRGTNMEDEADQY	1435	P04851.1

[1204]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
553	白喉杆菌	QSIALSSLMVAQAIPLVGEL	1436	AAV70486.1
554	麻疹病毒 EDMONSTON 株	QSRFGWFENKEISDI	1437	P04851.1
555	麻疹病毒 EDMONSTON 株	QSRFGWFENKEISDIEVQDP	1438	P04851.1
556	麻疹病毒 EDMONSTON 株	QSRGEAR	1439	P04851.1
557	麻疹病毒 EDMONSTON 株	QSRGEARESYRETGPSRA	1440	P04851.1
558	风疹病毒 THERIEN 株	QTGRGGSAPRPELGPPTN	1441	P07566.1
559	风疹病毒 THERIEN 株	QTPAPKPSRAPPQQPQPPRMQTGRG	1442	P07566.1
560	麻疹病毒 EDMONSTON 株	QVGSRRYPDAVYLHR	1443	P69353.1
561	麻疹病毒 EDMONSTON 株	QVSFLQGDQSENE	1444	P04851.1
562	风疹病毒	QYHPTACEVEPAFGH	1445	BAA28178.1
563	麻疹病毒 EDMONSTON 株	QYVLATYDTSRVEHA	1446	P08362.1
564	百日咳博德特氏菌	RANPNPYTSRRSV	1447	ACI04548.1
565	麻疹病毒 EDMONSTON 株	RASDARAAHLPTG	1448	P04851.1
566	麻疹病毒 EDMONSTON 株	RASLETTNQAIEAIR	1449	P69353.1
567	风疹病毒 THERIEN 株	RCGRLICGLSTTAQYPPTRF	1450	P07566.1
568	百日咳博德特氏菌	RDGQSVIGACASPYEGRYR	1451	P04979.1
569	麻疹病毒 EDMONSTON 株	RESYRETGPSRAS	1452	P04851.1
570	麻疹病毒 EDMONSTON 株	RETGPSRASDARA	1453	P04851.1
571	流行性腮腺炎腮腺炎病毒属	RFAKYQQQGRLEAR	1454	P21186.1
572	风疹病毒 THERIEN 株	RFGAPQAFLAGLLLATVAVGTARA	1455	P07566.1
573	麻疹病毒 EDMONSTON 株	RFMVALILDIKRTPG	1456	P04851.1
574	麻疹病毒 EDMONSTON 株	RFMVALILDIKRTPGNKPRI	1457	P04851.1

[1205]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
575	麻疹病毒 EDMONSTON 株	RGEARESYRETGP	1458	P04851.1
576	风疹病毒	RGTTPPAYG	1459	CAA28880.1
577	风疹病毒 M33 株	RIETRSARH	1460	ABD64214.1
578	百日咳博德特氏菌	RILAGALATYQ	1461	P04977.1
579	百日咳博德特氏菌	RIPPENIRRVT	1462	ACI04548.1
580	百日咳博德特氏菌	RISNLND	1463	ACI16083.1
581	百日咳博德特氏菌	RLANLNG	1464	ACI16088.1
582	麻疹病毒 EDMONSTON 株	RLDVGTNLGNAIAKL	1465	P69353.1
583	麻疹病毒 EDMONSTON 株	RLERKWLDV	1466	P04851.1
584	麻疹病毒 EDMONSTON 株	RLGGKEDRRVKQSRG	1467	P04851.1
585	麻疹病毒 EDMONSTON 株	RLGGKEDRRVKQSRGEARES	1468	P04851.1
586	麻疹病毒 EDMONSTON 株	RLLDRLVRL	1469	ABI54110.1
587	风疹病毒	RLRLVDAD	1470	CAJ88851.1
588	风疹病毒	RLRLVDADDPLLR	1471	BAA19893.1
589	风疹病毒	RLRLVDADDPLLRTAPGPGEVWVTPVIGSQA	1472	BAA19893.1
590	风疹病毒	RLRLVQDADDPLLRIAPGPGEVWVTPVIGSQA	1473	SRC265968
591	麻疹病毒 EDMONSTON 株	RLSDNGYYTV	1474	ABK40528.1
592	风疹病毒	RLVDADDP	1475	CAJ88851.1
593	风疹病毒	RLVDADDPLLRTAPG	1476	BAA28178.1
594	风疹病毒	RLVGATPE	1477	CAJ88851.1
595	风疹病毒	RMQTGRGGSAPRPELGPPTNPFQAAVA	1478	ABD64216.1
596	麻疹 麻疹病毒属	RMSKGVFKV	1479	ABY21184.1
597	麻疹 麻疹病毒属	RNPGLGAPVFHMTNY	1480	P06830.1

[1206]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
598	风疹病毒	RPRLRLVD	1481	CAJ88851.1
599	百日咳博德特氏菌	RQAESSEAMAAWSERAGEA	1482	P04977.1
600	风疹病毒	RQTWAEWAAAHWWQL	1483	BAA28178.1
601	麻疹病毒 EDMONSTON 株	RRVKQSRGEARES	1484	P04851.1
602	麻疹 麻疹病毒属	RRYPDAVYL	1485	ACA09725.1
603	麻疹病毒 EDMONSTON 株	RRYPDAVYLHRIDLG	1486	P69353.1
604	麻疹病毒 EDMONSTON 株	RSAGKVSSTLASELG	1487	P04851.1
605	麻疹病毒 EDMONSTON 株	RSAGKVSSTLASELGITAED	1488	P04851.1
606	麻疹病毒 EDMONSTON 株	RSELSQLS	1489	P08362.1
607	麻疹病毒 EDMONSTON 株	RSELSQLSMYRVFEV	1490	P08362.1
608	风疹病毒	RSQTPAPKPSRAPPQQPQPPRMQT	1491	ABD64214.1
609	风疹病毒	RTAPGPGE	1492	CAJ88851.1
610	风疹病毒	RTAPGPGEVWVTPVI	1493	BAA28178.1
611	麻疹 麻疹病毒属	RTDDKLRMETCFQQA	1494	P06830.1
612	风疹病毒	RTLAPPRNVRVTGCY	1495	BAA28178.1
613	麻疹病毒 EDMONSTON 株	RTVLEPIRDALNAMT	1496	P69353.1
614	麻疹病毒 EDMONSTON 株	RVEHAVVYYVYSPSR	1497	P08362.1
615	麻疹病毒 EDMONSTON 株	RVFEVGVIRNPGLGA	1498	P08362.1
616	百日咳博德特氏菌	RVHVSKEEQYYDYEDATFE	1499	P04978.2
617	风疹病毒	RVIDPAAQ	1500	BAA28178.1
618	风疹病毒	RVKFHTETRTVWQLS	1501	BAA28178.1
619	百日咳博德特氏菌	RVYHNGITGET	1502	ACI04548.1
620	百日咳博德特氏菌	RYDSRPPEDVF	1503	ACI04548.1
621	麻疹病毒	RYPDAVYLHRIDLGP	1504	P69353.1

[1207]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	EDMONSTON 株			
622	麻疹病毒 EDMONSTON 株	SAEISIQALSYALGG	1505	P69353.1
623	麻疹病毒 EDMONSTON 株	SAEPLLSCKPWQESR	1506	P04851.1
624	麻疹病毒 EDMONSTON 株	SAEPLLSCKPWQESRKNKAQ	1507	P04851.1
625	麻疹 麻疹病毒属	SAGKVSSTLASELG	1508	P04851.1
626	麻疹病毒 EDMONSTON 株	SAGKVSSTLASELGITAEDARLVS	1509	P04851.1
627	麻疹病毒 EDMONSTON 株	SCTVTREDGT	1510	P08362.1
628	风疹病毒	SDAACWGFPTDTVMS	1511	BAA28178.1
629	麻疹病毒 EDMONSTON 株	SDARAAHLPTGTP	1512	P04851.1
630	风疹病毒 THERIEN 株	SDWHQGTHVCHTKHMDFWCVEHD	1513	P07566.1
631	麻疹病毒 EDMONSTON 株	SELRRWIKYTQQRRV	1514	P04851.1
632	麻疹病毒 EDMONSTON 株	SELRRWIKYTQQRRVVGEFR	1515	P04851.1
633	麻疹病毒 EDMONSTON 株	SELSQL	1516	P08362.1
634	麻疹病毒 EDMONSTON 株	SELSQLS	1517	P08362.1
635	百日咳博德特氏菌	SEYLAHRRIPPENIRRVTRV	1518	CAD44970.1
636	麻疹病毒 EDMONSTON 株	SFLQGDQSENELP	1519	P04851.1
637	麻疹病毒 EDMONSTON 株	SGKGVSFQLVKLGVW	1520	P08362.1
638	麻疹病毒 EDMONSTON 株	SHRGVIADNQAKWAV	1521	P08362.1
639	麻疹病毒 EDMONSTON 株	SIEHQVKDVLTPLFK	1522	P08362.1
640	麻疹病毒 EDMONSTON 株	SKIGVVGIGSASYKV	1523	P69353.1
641	麻疹 麻疹病毒属	SKRSELSQLSMYRVF	1524	P06830.1
642	麻疹 麻疹病毒属	SLFVESPGQLIQRITDDPDVS	1525	ABI54110.1
643	麻疹病毒	SLSTNLDVTNSIEHQ	1526	P08362.1

[1208]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	EDMONSTON 株			
644	麻疹病毒 EDMONSTON 株	SLSTNLDVTNSIEHQVKDVLTPLFK	1527	P08362.1
645	麻疹病毒 EDMONSTON 株	SLWGSGLLML	1528	BAE98296.1
646	麻疹病毒 EDMONSTON 株	SMKGLSSTSIVYILI	1529	P69353.1
647	麻疹 麻疹病毒属	SMYRVFEVGV	1530	P08362.1
648	麻疹 麻疹病毒属	SNDLSNCMVALGELK	1531	P06830.1
649	麻疹病毒 HALLE 株	SPGQLIQR	1532	P10050.1
650	麻疹病毒 EDMONSTON 株	SQDPQDSRRSAEP	1533	P04851.1
651	麻疹 麻疹病毒属	SRIVINREHLMIDRP	1534	P06830.1
652	麻疹病毒 EDMONSTON 株	SRKNKAQTRTPLQ	1535	P04851.1
653	风疹病毒	SRLVGATP	1536	CAJ88851.1
654	风疹病毒	SRLVGATPERPRLRLVDADDPLLR	1537	CAJ88851.1
655	麻疹病毒 EDMONSTON 株	SRPGLKPDLTGTSKS	1538	P69353.1
656	百日咳博德特氏菌	SRRSVASIVGTLVRM	1539	CAD44970.1
657	风疹病毒	SRWGLGSPNCHGPDW	1540	BAA28178.1
658	百日咳博德特氏菌	SSATK	1541	ACI16088.1
659	风疹病毒	SSGGYAQLASYFNPG	1542	BAA28178.1
660	百日咳博德特氏菌	SSLGNGV	1543	ACI16083.1
661	麻疹 麻疹病毒属	SSNLVILPGQDLQYV	1544	P06830.1
662	麻疹病毒 EDMONSTON 株	SSQDPQDSRRSAEPL	1545	P04851.1
663	麻疹病毒 EDMONSTON 株	SSQDPQDSRRSAEPLLSCKP	1546	P04851.1
664	麻疹病毒 EDMONSTON 株	SSRRHKRFAGVVLAG	1547	P69353.1
665	麻疹病毒 EDMONSTON 株	SSTSIVYILIAVCLG	1548	P69353.1
666	百日咳博德特氏菌	STPGIVI	1549	AAA83981.1

[1209]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
667	百日咳博德特氏菌	STPGIVIPPQEQITQHGSPYGRC	1550	AAA83981.1
668	百日咳博德特氏菌	STSSSRRYTEVY	1551	P04977.1
669	麻疹病毒 EDMONSTON 株	SYFIVLSIAYPTLSE	1552	P69353.1
670	麻疹病毒 EDMONSTON 株	SYRETGPSRASDA	1553	P04851.1
671	百日咳博德特氏菌	SYVK	1554	ACI16083.1
672	风疹病毒	SYVQHPHKTVRVKFH	1555	BAA28178.1
673	风疹病毒	TAPGPGEV	1556	CAJ88851.1
674	百日咳博德特氏菌	TATRLLSSTNSRLC	1557	AAA83981.1
675	麻疹 麻疹病毒属	TDDPVIDRLYLSSHR	1558	P06830.1
676	麻疹病毒 EDMONSTON 株	TEILSLFGPSLRDPI	1559	P69353.1
677	风疹病毒	TETRTVWQLSVAGVS	1560	BAA28178.1
678	百日咳博德特氏菌	TEVYLEHRMQEAVE	1561	P04977.1
679	麻疹病毒 EDMONSTON 株	TFMPEGTVCSQNALY	1562	P69353.1
680	风疹病毒	TGACICEIPTDVSCE	1563	BAA28178.1
681	百日咳博德特氏菌	TGDLRAY	1564	ACI16083.1
682	麻疹 麻疹病毒属	TGMQSWVPLSTDDPV	1565	P06830.1
683	风疹病毒	TGNQQSRWGLGSPNC	1566	BAA28178.1
684	麻疹病毒 EDMONSTON 株	TGPSRASDARAAH	1567	P04851.1
685	麻疹 麻疹病毒属	TGTIINQDPDKILTY	1568	AAF02706.1
686	风疹病毒	TGVVYGTHTTAVSET	1569	BAA28178.1
687	麻疹病毒 EDMONSTON 株	TIRGQFSNMSLSLLD	1570	P08362.1
688	风疹病毒	TLGATCALPLAGLLA	1571	BAA28178.1
689	麻疹 麻疹病毒属	TLLNNCTRV	1572	P26031.1
690	麻疹病毒 EDMONSTON-B 株	TLNVPPPPDPGR	1573	P03422.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
691	麻疹病毒 EDMONSTON-B 株	TLNVPPPPDPGRASTSGTPIKK	1574	P03422.1
692	麻疹 麻疹病毒属	TMTSQGMYGGTYPVE	1575	P06830.1
693	麻疹病毒 EDMONSTON 株	TNLGNAIAKLEDAKE	1576	P69353.1
694	麻疹病毒 EDMONSTON 株	TNMEDEADQYFSHDD	1577	P04851.1
695	麻疹病毒 EDMONSTON 株	TNMEDEADQYFSHDDPISSD	1578	P04851.1
696	风疹病毒 M33 株	TNPFQAAVARGLRPP	1579	CAA28880.1
697	麻疹 麻疹病毒属	TNSIEHQVKDVLTPL	1580	P06830.1
698	麻疹病毒 EDMONSTON 株	TNYLEQPVSNDLSNC	1581	P08362.1
699	麻疹 麻疹病毒属	TNYLEQPVSNDLSNCMVALGELKLAAL	1582	AAR89413.1
700	风疹病毒	TPERPRLR	1583	CAJ88851.1
[1210] 701	风疹病毒 THERIEN 株	TPERPRLRLVDADDPLLRTA	1584	P07566.1
702	麻疹病毒 HALLE 株	TPGNKPRIA	1585	P10050.1
703	麻疹病毒 EDMONSTON 株	TPLDIDTASESSQDP	1586	P04851.1
704	麻疹病毒 EDMONSTON 株	TPLDIDTASESSQDPQDSRR	1587	P04851.1
705	麻疹病毒 EDMONSTON 株	TPLFKIIGDEVGLRT	1588	P08362.1
706	麻疹病毒 EDMONSTON 株	TPLQCTM	1589	P04851.1
707	风疹病毒	TPVIGSQA	1590	CAJ88851.1
708	风疹病毒	TPVIGSQARK	1591	BAA19893.1
709	麻疹病毒 EDMONSTON 株	TQGYLISNFDESSCT	1592	P69353.1
710	百日咳博德特氏菌	TRANPNPYTSRRSVASIVGTLVRM	1593	P04977.1
711	风疹病毒	TRFGCAMRWGLPP	1594	NP_740663.1
712	百日咳博德特氏菌	TRNTGQPATDHYYSNV	1595	AAA83981.1
713	麻疹病毒 EDMONSTON 株	TRTPLQCTMTEIF	1596	P04851.1

[1211]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
714	风疹病毒	TRWHRLLRMPVR	1597	ABD64216.1
715	麻疹病毒 EDMONSTON 株	TSGSGGAIRGIKHII	1598	P04851.1
716	麻疹病毒 EDMONSTON 株	TSGSGGAIRGIKHIIIVPIP	1599	P04851.1
717	麻疹病毒 EDMONSTON 株	TSQGMYGGTYLVEKP	1600	P08362.1
718	麻疹 麻疹病毒属	TSRVEHAVVYYVYSP	1601	P06830.1
719	百日咳博德特氏菌	TSSSRRYTEVYL	1602	ACI04548.1
720	百日咳博德特氏菌	TSYVG	1603	ACI16088.1
721	麻疹病毒 EDMONSTON 株	TTEDKISRAVGPRQA	1604	P04851.1
722	麻疹病毒 EDMONSTON 株	TTEDKISRAVGPRQAQVSFL	1605	P04851.1
723	风疹病毒 M33 株	TTERIETRSARHP	1606	ABD64214.1
724	麻疹病毒 EDMONSTON 株	TTNQAIEAIRQAGQE	1607	P69353.1
725	风疹病毒	TTSDPWHPPGPLGLK	1608	BAA28178.1
726	麻疹病毒 EDMONSTON 株	TVCSQNALYPMSPLL	1609	P69353.1
727	风疹病毒	TVNGEDVGAVPPGKF	1610	BAA28178.1
728	麻疹 麻疹病毒属	TYPVEKPNLSSKRSE	1611	P06830.1
729	风疹病毒	VAGVSCNVTTEHPFC	1612	BAA28178.1
730	百日咳博德特氏菌	VAPGIVIPPKALFTQQGGAYGRC	1613	P04979.1
731	风疹病毒	VCHTKHMDFWCVEHDRPPPATPTPL	1614	NP_740663.1
732	风疹病毒	VCQRHSPDCSRLVGATPER	1615	BAA19893.1
733	风疹病毒	VDADDPLL	1616	CAJ88851.1
734	风疹病毒	VDADDPLLRTAPGPGEVWVT	1617	BAA19893.1
735	白喉杆菌	VDIGFAAYNFVESIINLFQV	1618	AAV70486.1
736	麻疹病毒 EDMONSTON 株	VEIAEYRRLLRTVLE	1619	P69353.1

[1212]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
737	麻疹病毒 EDMONSTON 株	VELENSMGGLNFGRS	1620	P04851.1
738	麻疹病毒 EDMONSTON 株	VELENSMGGLNFGRSYFDPA	1621	P04851.1
739	麻疹 麻疹病毒属	VELKIKIASGFGPLI	1622	P06830.1
740	风疹病毒	VEMDEWIHAHTTSD	1623	SRC265968
741	风疹病毒	VEMPEWIHAHTTSDP	1624	BAA28178.1
742	白喉杆菌	VERRLVKVL	1625	P33120.2
743	麻疹 麻疹病毒属	VESPGQLI	1626	ABI54110.1
744	麻疹病毒 EDMONSTON 株	VESPGQLIQRITDDP	1627	P04851.1
745	麻疹病毒 EDMONSTON 株	VESPGQLIQRITDDPDVSIR	1628	P04851.1
746	风疹病毒	VFALASYVQHPHKTV	1629	BAA28178.1
747	风疹病毒	VGATPERP	1630	CAJ88851.1
748	风疹病毒	VGATPERPRL	1631	BAA19893.1
749	风疹病毒	VGATPERPRLRLVDA	1632	BAA28178.1
750	麻疹病毒 EDMONSTON 株	VGIGSASYKVMTRSS	1633	P69353.1
751	麻疹病毒 EDMONSTON 株	VGLRTPQRFTDLVKF	1634	P08362.1
752	白喉杆菌	VHHNTEEIVAQSIALSSLMV	1635	AAV70486.1
753	麻疹病毒 EDMONSTON 株	VHRLEGVSYNIGSQE	1636	P69353.1
754	风疹病毒	VIGSQARK	1637	CAJ88851.1
755	百日咳博德特氏菌	VITGSI	1638	ACI16088.1
756	百日咳博德特氏菌	VITGTI	1639	ACI16083.1
757	麻疹病毒 EDMONSTON 株	VKQSRGEA	1640	P04851.1
758	麻疹病毒 EDMONSTON 株	VLFVMFLSLI	1641	P08362.1
759	麻疹 麻疹病毒属	VLFVMFLSLIGLLAI	1642	P06830.1

[1213]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
760	麻疹 麻疹病毒属	VLTPLFKIIGDEVGL	1643	P06830.1
761	麻疹病毒 EDMONSTON 株	VMTRSSHQSLVIKLM	1644	P69353.1
762	风疹病毒	VPAAPCARIWNGTQR	1645	BAA28178.1
763	风疹病毒	VPPDPGDLVEYIMNY	1646	BAA28178.1
764	麻疹病毒 EDMONSTON 株	VQSVASSRRHKRFAG	1647	P69353.1
765	百日咳博德特氏菌	VQTGGTSRTVTMRYLAS	1648	ACI16083.1
766	百日咳博德特氏菌	VQVRI	1649	ACI16083.1
767	风疹病毒	VRAYNQPAGDV	1650	NP_740662.1
768	风疹病毒	VRFESKIVDGGCFAP	1651	BAA28178.1
769	风疹病毒	VRLAGVRFESKIVDG	1652	BAA28178.1
770	麻疹病毒 EDMONSTON 株	VSGSFGNRFILSQGN	1653	P69353.1
771	百日咳博德特氏菌	VSKEEQYYDYEDAT	1654	AAA83981.1
772	麻疹病毒 EDMONSTON 株	VSKGNCSGPTTIRGQ	1655	P08362.1
773	风疹病毒	VTAALLNTPPPYQVS	1656	BAA28178.1
774	风疹病毒	VTGCYQCGTPALVEG	1657	BAA28178.1
775	风疹病毒	VTPVIGSQ	1658	CAJ88851.1
776	风疹病毒 THERIEN 株	VTTEHPFCNTPHGQLEVQVPPD	1659	P07566.1
777	麻疹病毒 EDMONSTON 株	VVLAGAALGVATAAQ	1660	P69353.1
778	风疹病毒	VWQLSVAGVSCNVTT	1661	BAA28178.1
779	风疹病毒	VWVTPVIG	1662	CAJ88851.1
780	风疹病毒	VWVTPVIGSQAR	1663	BAA19893.1
781	麻疹病毒 EDMONSTON 株	VYILIAVCLGGLIGI	1664	P69353.1
782	麻疹病毒 EDMONSTON 株	VYLHRIDLGPPISLE	1665	P69353.1

[1214]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
783	百日咳博德特氏菌	VYRYDSRP	1666	P04977.1
784	百日咳博德特氏菌	VYRYDSRPPEDV	1667	P04977.1
785	麻疹 麻疹病毒属	VYWLTIPPMKNLALG	1668	P06830.1
786	风疹病毒	WDLEATGACICEIPT	1669	BAA28178.1
787	麻疹 麻疹病毒属	WDQKLWCRHFCVLAD	1670	AAR89413.1
788	风疹病毒	WGFPTDTVMSVFALA	1671	BAA28178.1
789	风疹病毒	WHPPGPLGLKFKTVR	1672	BAA28178.1
790	风疹病毒	WIHAHTTSDPWHPPG	1673	BAA28178.1
791	麻疹病毒 EDMONSTON 株	WLTIPPMKNLALGVI	1674	P08362.1
792	麻疹病毒 EDMONSTON 株	WQESRKNKAQTRTPLQCTMT	1675	P04851.1
793	风疹病毒	WVCIFMVCRRACR	1676	SRC265968
794	风疹病毒	WVTPVIGS	1677	CAJ88851.1
795	麻疹病毒 EDMONSTON 株	WYTTVPKYVATQGYL	1678	P69353.1
796	麻疹病毒 EDMONSTON 株	YALGGDINKVLEKLG	1679	P69353.1
797	麻疹病毒 EDMONSTON 株	YAMGVGVELE	1680	P04851.1
798	麻疹 麻疹病毒属	YAMGVGVELEN	1681	ABI54110.1
799	麻疹 麻疹病毒属	YCADVAAEELMNALV	1682	AAR89413.1
800	麻疹 麻疹病毒属	YDFRDLTWCINPPER	1683	P06830.1
801	百日咳博德特氏菌	YFEPGPT	1684	ACI16083.1
802	风疹病毒	YFNPGGSYYKQYHPT	1685	BAA28178.1
803	麻疹病毒 EDMONSTON 株	YFRLGQEMVRRSAGK	1686	P04851.1
804	麻疹病毒 EDMONSTON 株	YFRLGQEMVRRSAGKVSSTL	1687	P04851.1
805	麻疹 麻疹病毒属	YFYPFRLPIKGVPIE	1688	P06830.1

[1215]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
806	百日咳博德特氏菌	YGDNAGRILAGALAT	1689	P04977.1
807	风疹病毒	YGHATVEMPEWIHAH	1690	BAA28178.1
808	风疹病毒	YIMNYTGNQQSRWGL	1691	BAA28178.1
809	麻疹病毒 EDMONSTON 株	YINNELIPSMNQLSC	1692	P69353.1
810	风疹病毒	YLCTAPGCATQAPVP	1693	BAA28178.1
811	麻疹 麻疹病毒属	YLFTVPIKEAGEDCH	1694	P06830.1
812	麻疹病毒 EDMONSTON 株	YLHDPEFNL	1695	ABK40531.1
813	麻疹病毒 EDMONSTON 株	YLNMSRLFV	1696	ABK40531.1
814	麻疹病毒 HALLE 株	YPALGLHEF	1697	P10050.1
815	麻疹 麻疹病毒属	YPALGLHEFA	1698	ABI54110.1
816	麻疹病毒 EDMONSTON 株	YPALGLHEFAGELST	1699	P04851.1
817	麻疹病毒 EDMONSTON 株	YPALGLHEFAGELSTLESLM	1700	P04851.1
818	麻疹病毒 EDMONSTON 株	YPFRLPIKGVPIELQ	1701	P08362.1
819	麻疹病毒 HALLE 株	YPLLWSYAM	1702	P10050.1
820	流行性腮腺炎 腮腺炎病毒属	YQQQGRL	1703	P21186.1
821	百日咳博德特氏菌	YQSEYLAHRR	1704	P04977.1
822	麻疹病毒 EDMONSTON 株	YRETGPSRASDARAA	1705	P04851.1
823	麻疹病毒 EDMONSTON 株	YRETGPSRASDARAAHLPTG	1706	P04851.1
824	风疹病毒	YRNASDVLPGHWLQGGWGCYNLSDW	1707	NP_740663.1
825	麻疹病毒 EDMONSTON 株	YRRLLRTVLEPIRDA	1708	P69353.1
826	百日咳博德特氏菌	YRYDSRPP	1709	P04977.1
827	麻疹病毒 EDMONSTON 株	YSGGDLLGILESRGI	1710	P69353.1
828	百日咳博德特氏菌	YSKVTATBLLASTNSRLCAVFVRDG	1711	SRC280066
829	麻疹病毒 EDMONSTON 株	YSPSRSFSYFYPFRL	1712	P08362.1
830	风疹病毒 THERIEN	YTGNQQSRWGLGSPNCHGPDWA	1713	P07566.1

[1216]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
	株	SPV		
831	麻疹 麻疹病毒属	YVLLAVLFV	1714	P08362.1
832	麻疹 麻疹病毒属	YVYSPGRSFSYFYPF	1715	P06830.1
833	百日咳博德特氏菌	YYDYEDATFQTYALTGISLCNPAASIC	1716	P04979.1
834	麻疹 麻疹病毒属	GDLLGILESRGIKAR	1717	AAF02706.1
835	麻疹 麻疹病毒属	TVPKYVATQGYLISN	1718	AAL29688.1
836	麻疹 麻疹病毒属	KPWDSPQEI	1719	P26035.1
837	麻疹 麻疹病毒属	KPWESPQEI	1720	CAA34579.1
838	百日咳博德特氏菌	ATYQSEYLAHRRIPP	1721	ACI04548.1
839	百日咳博德特氏菌	CMARQAESSEAMAAWSERAGEAMVLVYYESIAYSF	1722	ACI04548.1
840	百日咳博德特氏菌	CQVGSSNSAFVSTSSSRRYTEVYL	1723	ACI04548.1
841	百日咳博德特氏菌	DDPPATVYRYDSRPP	1724	ACI04548.1
842	百日咳博德特氏菌	GALATYQSEYLAHRRIPP	1725	ACI04548.1
843	百日咳博德特氏菌	MAAWSERAGEAMVLVYYESIAYSF	1726	ACI04548.1
844	百日咳博德特氏菌	MVLVYYESIAYSF	1727	ACI04548.1
845	百日咳博德特氏菌	PATVYRYDSRPPEDV	1728	ACI04548.1
846	百日咳博德特氏菌	YDSRPPEDV	1729	ACI04548.1
847	百日咳博德特氏菌	EPGITTNYDT	1730	ACI16087.1
848	百日咳博德特氏菌	GDLRAYKMVYATNPQTQLSN	1731	ACI16083.1
849	百日咳博德特氏菌	KNGDVEASAITTYVGFSVVYP	1732	ACI16083.1
850	百日咳博德特氏菌	KVTNGSKSYTLRYLASYVK	1733	ACI16088.1
851	百日咳博德特氏菌	QALGALKLYFEPGITTNYDTGDLIAYKQTYNASGN	1734	ACI16088.1
852	百日咳博德特氏菌	YATNPQTQLS	1735	ACI16083.1
853	白喉杆菌	DNENPLSGKAGGVVKVTYPGLTK	1736	AAV7048

[1217]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
		V		6.1
854	白喉杆菌	ENFSSYHGTKPGYVDSI	1737	AAV70486.1
855	白喉杆菌	KVDNAETIKKELGLSLTEP	1738	AAV70486.1
856	风疹病毒 M33 株	MEDLQKALEAQSRALRAGLAA	1739	CAA28880.1
857	白喉杆菌	QKGIQKPKSGTQGNYDDDWKGFY	1740	AAV70486.1
858	风疹病毒 M33 株	RTGAWQRKDWSRAPPPPEERQESRSQTPAPKPSR	1741	CAA28880.1
859	风疹病毒	AAGASQSRRPRPPRHARAQHLPEMTPAVT	1742	SRC265968
860	风疹病毒	CVTSWLWSEGEGAVFYRVDLHFINLGTP	1743	CAA28880.1
861	风疹病毒	FRVGGTRWHRLLRMPVRGLDGDTAPLP	1744	CAA28880.1
862	白喉杆菌	GRKIRMRCRAIDGDVTFCRPKSPVYVGN	1745	1007216A
863	风疹病毒	GTPPLDEDGRWDPALMYNPCGPEPPAHV	1746	CAA28880.1
864	白喉杆菌	GVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTKVNSKLSLFFEIKS	1747	AAV70486.1
865	风疹病毒	MASTTPITMEDLQKALEAQSRALRAGLAAG	1748	ABD64200.1
866	风疹病毒	PELGPPTNPFQAAVARGLRPPLHDPDTEAPTEAC	1749	CAA28880.1
867	风疹病毒	PLPPHTTERIETRSARHPWRIRFGAP	1750	CAA28880.1
868	风疹病毒	SRAPPPPEERQESRSQTPAPKPSRAPP	1751	CAA28880.1
869	风疹病毒	SRAPPQQPQPPRMQTGRGGSAPRPELGP	1752	CAA28880.1
870	风疹病毒	TPAVTPEGPAPPRTGAWQRKDWSRAPP	1753	CAA28880.1
871	风疹病毒	VRAYNQPAGDVRGVWGKGERTYAEQDFRV	1754	CAA28880.1
872	风疹病毒	AFGHSDAACWGFPTDTVMSV	1755	CAA28880.1
873	风疹病毒	CARIWNGTQRACTFWAVNAYS	1756	CAA28880.1
874	风疹病毒	EEAFTYLCTAPGCATQTPVPVR	1757	CAA28880.1
875	风疹病毒	FAPWDLEATGACICEIPTDV	1758	CAA28880.1
876	风疹病毒	GEDVGAFPPGKFVTAAL	1759	CAA28880.1

[1218]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
877	风疹病毒	GEVWVTPVIGSQARKCGLHI	1760	CAA28880.1
878	风疹病毒	GQLEVQVPPDPGDLVEYIMN	1761	CAA28880.1
879	风疹病毒	GSYYKQYHPTACEVEPAFGH	1762	CAA28880.1
880	风疹病毒	IHAHTTSDPWHPPGPLGLKF	1763	CAA28880.1
881	风疹病毒	IMNYTGNQQSRWGLGSPNCH	1764	CAA28880.1
882	风疹病毒	LHIRAGPYGHATVEMPEWIH	1765	CAA28880.1
883	风疹病毒	LKFKTVRPVALPRALAPPRN	1766	CAA28880.1
884	风疹病毒	LNTPPPYQVSCGGESDRASAGH	1767	CAA28880.1
885	风疹病毒	NCHGPDWASPVCQRHSPDCS	1768	CAA28880.1
886	风疹病毒	PDCSRLVGATPERPRLRLVD	1769	CAA28880.1
887	风疹病毒	PRNVRVTGCYQCGTPALVEG	1770	CAA28880.1
888	风疹病毒	PTDVSCEGLGAWVPTAPCARI	1771	CAA28880.1
889	风疹病毒	RLVDADDPLLRTAPGPGEVW	1772	CAA28880.1
890	风疹病毒	SVFALASYVQHPHKTVRVKF	1773	CAA28880.1
891	风疹病毒	VEGLAPGGGNCHLTVNGEDV	1774	CAA28880.1
892	风疹病毒	VKFHTETRTVWQLSVAGVSC	1775	CAA28880.1
893	风疹病毒	VPVRLAGVGFESKIVDGGCF	1776	CAA28880.1
894	风疹病毒	VSCNVTTEHPFCNTPHGQLE	1777	CAA28880.1
895	百日咳博德特氏菌	AAASSPDAHVPF	1778	AAA22983.1
896	百日咳博德特氏菌	AASSPDA	1779	AAA22983.1
897	百日咳博德特氏菌	AKLGAAASSPDA	1780	AAA22983.1
898	百日咳博德特氏菌	AMKPYEVTPTRM	1781	AAA22983.1
899	百日咳博德特氏菌	AMTHLSPALADVPYVLVKTNMVVTS	1782	AAA22983.1
900	百日咳博德特氏菌	ASSPDAHVPFCF	1783	AAA22983.1

[1219]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
901	百日咳博德特氏菌	ASSPDAHVPFCFGKDLKRPGSSPME	1784	AAA22983.1
902	百日咳博德特氏菌	CFGKDLKRPGSS	1785	AAA22983.1
903	百日咳博德特氏菌	CFGKDLKRPGSSPMEVMLRAVFMQQ	1786	AAA22983.1
904	百日咳博德特氏菌	CGIAAKLGAAAS	1787	AAA22983.1
905	百日咳博德特氏菌	CGIAAKLGAAASSPDAHVPFCFGKD	1788	AAA22983.1
906	百日咳博德特氏菌	DAHVPFCFGKDL	1789	AAA22983.1
907	百日咳博德特氏菌	DLKRPGSSPMEV	1790	AAA22983.1
908	百日咳博德特氏菌	DVPYVLVKTNMV	1791	AAA22983.1
909	百日咳博德特氏菌	DVPYVLVKTNMVVTSVAMKPYEVTPT	1792	AAA22983.1
910	百日咳博德特氏菌	EVMLRAVFMQQR	1793	AAA22983.1
911	百日咳博德特氏菌	FEGKPALELIRM	1794	AAA22983.1
912	百日咳博德特氏菌	FLGPKQLTFEGK	1795	AAA22983.1
913	百日咳博德特氏菌	FLGPKQLTFEGKPALELIRMVECSG	1796	AAA22983.1
914	百日咳博德特氏菌	FMQQRPLRMFLGPKQLT	1797	AAA22983.1
915	百日咳博德特氏菌	GKDLKRPGSSPM	1798	AAA22983.1
916	百日咳博德特氏菌	GKDLKRPGSSPME	1799	AAA22983.1
917	百日咳博德特氏菌	GKPALELIRMVE	1800	AAA22983.1
918	百日咳博德特氏菌	GPKQLTFEGKPA	1801	AAA22983.1
919	百日咳博德特氏菌	HVPFCFGKDLKR	1802	AAA22983.1
920	百日咳博德特氏菌	IAAKLGAAASSP	1803	AAA22983.1
921	百日咳博德特氏菌	KPYEVTPTRMLV	1804	AAA22983.1
922	百日咳博德特氏菌	KQLTFEGKPALE	1805	AAA22983.1
923	百日咳博德特氏菌	KRPGSSPMEVML	1806	AAA22983.1
924	百日咳博德特氏菌	LELIRMVECSGK	1807	AAA22983.1

[1220]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
925	百日咳博德特氏菌	LGAAASSPDAHV	1808	AAA22983.1
926	百日咳博德特氏菌	LIRMVECSGKQD	1809	AAA22983.1
927	百日咳博德特氏菌	LVCGIAAKLGAA	1810	AAA22983.1
928	百日咳博德特氏菌	MKPYEVTPTRM	1811	AAA22983.1
929	百日咳博德特氏菌	MLRAVFMQQRPL	1812	AAA22983.1
930	百日咳博德特氏菌	MQQRPLRM	1813	AAA22983.1
931	百日咳博德特氏菌	MQQRPLRMFLGP	1814	AAA22983.1
932	百日咳博德特氏菌	MVVTSVAMKPYE	1815	AAA22983.1
933	百日咳博德特氏菌	MVVTSVAMKPYEVTPTRMLVCGIAA	1816	AAA22983.1
934	百日咳博德特氏菌	PALELIRMVECS	1817	AAA22983.1
935	百日咳博德特氏菌	PALELIRMVECSGK	1818	AAA22983.1
936	百日咳博德特氏菌	PFCFGKDLKRPG	1819	AAA22983.1
937	百日咳博德特氏菌	PGSSPMEVMLRA	1820	AAA22983.1
938	百日咳博德特氏菌	PGSSPMEVMLRAVF	1821	AAA22983.1
939	百日咳博德特氏菌	PKQLTFEGK	1822	AAA22983.1
940	百日咳博德特氏菌	PLRMFLGPKQLT	1823	AAA22983.1
941	百日咳博德特氏菌	PTRMLVCGIAAK	1824	AAA22983.1
942	百日咳博德特氏菌	PYVLVKTNMVVT	1825	AAA22983.1
943	百日咳博德特氏菌	QLTFEGKPALELIRMVECSGKQDCP	1826	AAA22983.1
944	百日咳博德特氏菌	QRPLRMFLGPKQ	1827	AAA22983.1
945	百日咳博德特氏菌	RAVFMQQRPLRM	1828	AAA22983.1
946	百日咳博德特氏菌	RMFLGPKQLTFE	1829	AAA22983.1
947	百日咳博德特氏菌	RMLVCGIAAKLG	1830	AAA22983.1
948	百日咳博德特氏菌	RMVECSGKQDCP	1831	AAA22983.1

[1221]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
949	百日咳博德特氏菌	SPDAHVPFCFGK	1832	AAA22983.1
950	百日咳博德特氏菌	SSPMEVMLRAVF	1833	AAA22983.1
951	百日咳博德特氏菌	SSPMEVMLRAVFMQQRPLRMFLGPK	1834	AAA22983.1
952	百日咳博德特氏菌	SVAMKPYEVTPT	1835	AAA22983.1
953	百日咳博德特氏菌	VECSGKQDCP	1836	AAA22983.1
954	百日咳博德特氏菌	VFMQQRPLRMFL	1837	AAA22983.1
955	百日咳博德特氏菌	VFMQQRPLRMFLGPKQLTFEGKPAL	1838	AAA22983.1
956	百日咳博德特氏菌	VKTNMVVTSVAM	1839	AAA22983.1
957	百日咳博德特氏菌	VLVKTNMVVTSV	1840	AAA22983.1
958	百日咳博德特氏菌	VTPTRMLVCGIA	1841	AAA22983.1
959	百日咳博德特氏菌	VTSVAMKPYEVT	1842	AAA22983.1
960	百日咳博德特氏菌	YEVTPTRMLVCG	1843	AAA22983.1
961	百日咳博德特氏菌	YEVTPTRMLVCGIAAKLGAAASSPD	1844	AAA22983.1
962	百日咳博德特氏菌	CASPYEGRYRDMYDALR	1845	P04979.1
963	百日咳博德特氏菌	CAVFVRDGQSV	1846	P04979.1
964	百日咳博德特氏菌	CITTIYKTG	1847	P04979.1
965	百日咳博德特氏菌	CPNGTRALTV	1848	P04979.1
966	百日咳博德特氏菌	DALRRLLYMIYMSG	1849	P04979.1
967	风疹病毒 THERIEN 株	GNRGRGQRRDWSRAPPPPEERQETRS	1850	P07566.1
968	百日咳博德特氏菌	GQPAADHYYSKVT	1851	P04979.1
969	风疹病毒	GSPNCHGPDWASPVCQRHS	1852	ABD64214.1
970	麻疹病毒 EDMONSTON 株	HKSLSTNLDVTNSIEHQ	1853	P08362.1
971	百日咳博德特氏菌	LFTQQGGAYGRC	1854	P04979.1
972	麻疹病毒 EDMONSTON 株	LIGLLAIAGIRLHRAAIYTAEI	1855	P08362.1
973	麻疹 麻疹病毒属	PDTAADSELRRWIKY	1856	ABI54110.1
974	风疹病毒 THERIEN 株	PNCHGPDWASPVCQRHS	1857	P07566.1

[1222]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
975	风疹病毒	QTPAPKPSRAPPQQPQPPRMQTGR	1858	ABD64216.1
976	风疹病毒疫苗株RA27/3	RAGLTAGASQSRRPRPPR	1859	CAA33016.1
977	风疹病毒 疫苗株RA27/3	RFGAPQAFLAGLLLAAVAVGTARA	1860	ABD64214.1
978	百日咳博德特氏菌	RGNAELQTYLRQITPG	1861	P04979.1
979	百日咳博德特氏菌	RVHVSKEEQYYDYED	1862	P04979.1
980	百日咳博德特氏菌	SIYGLYDGTYL	1863	P04979.1
981	百日咳博德特氏菌	SKVTATRLLASTNS	1864	P04979.1
982	百日咳博德特氏菌	TQQGGAYGRCPNGTRA	1865	P04979.1
983	百日咳博德特氏菌	VAPGIVIPPKAL	1866	P04979.1
984	百日咳博德特氏菌	DSRPPEDVFQNGFTAWG	1867	ACI04548.1
985	百日咳博德特氏菌	EHRMQEAVEAERAGR	1868	ACI04548.1
986	麻疹病毒EDMONSTON株	ETCFQQACKGKIQALCENPEWA	1869	P08362.1
987	百日咳博德特氏菌	EYVDTYGDNAGRILAGALATYQ	1870	ACI04548.1
988	百日咳博德特氏菌	HRRIPPENIRRVTR	1871	ACI04548.1
989	百日咳博德特氏菌	MARQAESSE	1872	ACI04548.1
990	百日咳博德特氏菌	MQEAVEAERAGR	1873	ACI04548.1
991	百日咳博德特氏菌	SQQTRANPNPYTSRR	1874	ACI04548.1
992	百日咳博德特氏菌	TRANPNPYTSRRSVASIVGTLVHG	1875	SRC280066
993	百日咳博德特氏菌	TVYRYDSRPPED	1876	ACI04548.1
994	麻疹 麻疹病毒属	NDRNLLD	1877	P10050.1
995	麻疹 麻疹病毒属	NMEDEADQYFSHDDPISSDQSRFGWFENK	1878	P04851.1
996	麻疹 麻疹病毒属	SRASDARAAHLPTGTPLDID	1879	P04851.1
997	百日咳博德特氏菌	EDVFQNGFTAW	1880	ACI04548.1
998	白喉杆菌	AEGSSSVEYINNWEQAK	1881	AAV70486.1
999	白喉杆菌	GPIKNKMSESPNKT	1882	AAV70486.1
1000	麻疹病毒HALLE株	GPKLTGALIGILSLFVESPGQLIQRITDDPDV	1883	P10050.1
1001	白喉杆菌	GYQKTVDHTKVNSK	1884	AAV70486.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1002	白喉杆菌	KTVDH	1885	AAV70486.1
1003	白喉杆菌	SESPNK	1886	AAV70486.1
1004	白喉杆菌	AEGSSSVEYINNWEQAKALS	1887	AAV70486.1
1005	白喉杆菌	AQAIPLVGELVDIGFAAYNF	1888	AAV70486.1
1006	白喉杆菌	ASRVVLSLPFAEGSSSVEYI	1889	AAV70486.1
1007	白喉杆菌	CINLDWDVIRDKTKTKIESL	1890	AAV70486.1
1008	白喉杆菌	CRAIDGDVTFCRPKSPVYVG	1891	AAV70486.1
1009	白喉杆菌	CRPKSPVYVGNGVHANLHVA	1892	AAV70486.1
1010	白喉杆菌	DAAGYSVDNENPLSGKAGGV	1893	AAV70486.1
1011	白喉杆菌	DKTKTKIESLKEHGPIKNKM	1894	AAV70486.1
1012	白喉杆菌	EEFIKRFGDGASRVVLSLPF	1895	AAV70486.1
[1223] 1013	白喉杆菌	EKAKQYLEEFHQTALEHPEL	1896	AAV70486.1
1014	白喉杆菌	FHRSSSEKIHSNEISSDSIG	1897	AAV70486.1
1015	白喉杆菌	GADDVVDSSKSFVMENFSSY	1898	CAE11230.1
1016	白喉杆菌	GKRGQDAMYEYMAQACAGNR	1899	AAV70486.1
1017	白喉杆菌	GSVMGIADGAVHHNTEEIVA	1900	AAV70486.1
1018	白喉杆菌	GTQGNYDDDWKGFYSTDNKY	1901	AAV70486.1
1019	白喉杆菌	HDGYAVSWNTVEDSIIRTGF	1902	AAV70486.1
1020	白喉杆菌	HGTKPGYVDSIQKGIQKPKS	1903	AAV70486.1
1021	白喉杆菌	HQTALEHPELSELKTVTGTN	1904	AAV70486.1
1022	白喉杆菌	IQKGIQKPKSGTQGNYDDDW	1905	AAV70486.1
1023	白喉杆菌	KEHGPIKNKMSESPNKTVSE	1906	AAV70486.1
1024	白喉杆菌	KGFYSTDNKYDAAGYSVDNE	1907	AAV70486.1
1025	白喉杆菌	LDVNKSKTHISVNGRKIRMR	1908	AAV70486.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1026	风疹病毒	MASTIPITMEDLQKALEA	1909	SRC265968
1027	白喉杆菌	NGVHANLHVAFHRSSSEKIH	1910	AAV70486.1
1028	白喉杆菌	NNWEQAKALSVELEINFETR	1911	AAV70486.1
1029	白喉杆菌	NPLSGKAGGVVKVTYPGLTK	1912	AAV70486.1
1030	白喉杆菌	PVFAGANYAAWAVNVAQVID	1913	AAV70486.1
1031	白喉杆菌	SELKTVTGTNPVFAGANYAA	1914	AAV70486.1
1032	白喉杆菌	SESPNKTVSEEKAKQYLEEF	1915	AAV70486.1
1033	白喉杆菌	SETADNLEKTTAALSILPGI	1916	AAV70486.1
1034	白喉杆菌	SFVMENFSSYHGTKPGYVDS	1917	AAV70486.1
1035	白喉杆菌	SNEISSDSIGVLGYQKTVDH	1918	AAV70486.1
1036	白喉杆菌	SPGHKTQPFLHDGYAVSWNT	1919	AAV70486.1
1037	白喉杆菌	SVNGRKIRMRCRAIDGDVTF	1920	AAV70486.1
1038	白喉杆菌	TAALSILPGIGSVMGIADGA	1921	AAV70486.1
1039	白喉杆菌	TAENTPLPIAGVLLPTIPGK	1922	AAV70486.1
1040	白喉杆菌	TEPLMEQVGTEEFIKRFGDG	1923	AAV70486.1
1041	白喉杆菌	TIKKELGLSLTEPLMEQVGT	1924	AAV70486.1
1042	白喉杆菌	TKVNSKLSLFFEIKS	1925	AAV70486.1
1043	白喉杆菌	VEDSIIRTGFQGESGHDIKI	1926	AAV70486.1
1044	白喉杆菌	VELEINFETRGKRGQDAMYE	1927	AAV70486.1
1045	白喉杆菌	VESIINLFQVVHNSYNRPAY	1928	AAV70486.1
1046	白喉杆菌	VHNSYNRPAYSPGHKTQPFL	1929	AAV70486.1
1047	白喉杆菌	VKVTYPGLTKVLALKVDNAE	1930	AAV70486.1
1048	白喉杆菌	VLALKVDNAETIKKELGLSL	1931	AAV70486.1
1049	白喉杆菌	VLGYQKTVDHTKVNSKLSLF	1932	AAV70486.1

[1224]

[1225]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1050	白喉杆菌	VRRSVGSSLSCINLDWDVIR	1933	AAV70486.1
1051	白喉杆菌	WAVNVAQVIDSETADNLEKT	1934	AAV70486.1
1052	白喉杆菌	YMAQACAGNRVRRSVGSSLS	1935	AAV70486.1
1053	百日咳博德特氏菌	AKAPPAPKPAPQPGP	1936	ABO77783.1
1054	百日咳博德特氏菌	APKPAPQPGP	1937	ABO77783.1
1055	百日咳博德特氏菌	APKPAPQPGPQPPQP	1938	ABO77783.1
1056	麻疹病毒 EDMONSTON 株	AQTRTPLQCTMTEIF	1939	P04851.1
1057	麻疹病毒 EDMONSTON 株	ASRGTNMEDEADQYFSHDD	1940	P04851.1
1058	百日咳博德特氏菌	ATIRR	1941	ABO77783.1
1059	百日咳博德特氏菌	DNRAG	1942	ABO77783.1
1060	百日咳博德特氏菌	EAPAPQPPAGRELSA	1943	ABO77783.1
1061	麻疹病毒 EDMONSTON 株	EMVRRSAGKVSSTLASELGI	1944	P04851.1
1062	百日咳博德特氏菌	GASEL	1945	ABO77783.1
1063	百日咳博德特氏菌	GDALAGGAVP	1946	AAA22980.1
1064	百日咳博德特氏菌	GDAPAGGAVP	1947	ABO77783.1
1065	百日咳博德特氏菌	GDTWDDD	1948	ABO77783.1
1066	百日咳博德特氏菌	GERQH	1949	ABO77783.1
1067	百日咳博德特氏菌	GGAVP	1950	ABO77783.1
1068	百日咳博德特氏菌	GGFGP	1951	P14283.3
1069	百日咳博德特氏菌	GGFGPGGFGP	1952	BAF35031.1
1070	百日咳博德特氏菌	GGFGPVLDGW	1953	ABO77783.1
1071	麻疹病毒 EDMONSTON 株	GGKEDRRVKQSRGEARESYR	1954	P04851.1
1072	百日咳博德特氏菌	GILLEN	1955	ABO77783.1
1073	百日咳博德特氏菌	GIRRFL	1956	ABO77783.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1074	麻疹病毒 EDMONSTON 株	HDDPISSDQSRFGWFENKEI	1957	P04851.1
1075	麻疹病毒 EDMONSTON 株	HEFAGELSTLESLMNLY	1958	P04851.1
1076	百日咳博德特氏菌	HLGGLAGY	1959	ABO77783.1
1077	麻疹病毒 EDMONSTON 株	HTTEDKISRAVGPRQAQVSFL	1960	P04851.1
1078	麻疹病毒 EDMONSTON 株	ICDIDTYIVEAGLASFILTI	1961	P04851.1
1079	麻疹病毒 EDMONSTON 株	IIIVPIPGDSSITTRSRLLD	1962	P04851.1
1080	麻疹病毒 EDMONSTON 株	ILDIKRTPGNKPRIAEMICD	1963	P04851.1
1081	百日咳博德特氏菌	KALLYR	1964	ABO77783.1
1082	麻疹病毒 EDMONSTON 株	KPPITSGSGGAIRGIKHIII	1965	P04851.1
1083	百日咳博德特氏菌	LAGSGL	1966	ABO77783.1
1084	麻疹病毒 EDMONSTON 株	LGITAEDARLVSEIAMHTTE	1967	P04851.1
1085	麻疹病毒 EDMONSTON 株	LGTILAQIWVLLAKAVTA	1968	P04851.1
1086	麻疹病毒 EDMONSTON 株	LPTGTPLDIDTASESSQD	1969	P04851.1
1087	麻疹病毒 EDMONSTON 株	MATLLRSLALFKRNKDK	1970	P04851.1
1088	百日咳博德特氏菌	PAPQPP	1971	ABO77783.1
1089	麻疹病毒 EDMONSTON 株	PDTAADSELRRWIKYTQQRR	1972	P04851.1
1090	麻疹病毒 EDMONSTON 株	PKLTGALIGILSLFVESPGQ	1973	P04851.1
1091	百日咳博德特氏菌	PQP	1974	ABO77783.1
1092	百日咳博德特氏菌	PQPGP	1975	ABO77783.1
1093	百日咳博德特氏菌	PQPGPQPPQPPQPQP	1976	ABO77783.1
1094	麻疹病毒 EDMONSTON 株	QDPQDSRRSAEPLLSCKPWQ	1977	P04851.1
1095	麻疹病毒 EDMONSTON 株	QRRVVGEFRLERKWLDVVR	1978	P04851.1

[1226]

[1227]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1096	百日咳博德特氏菌	RELSA	1979	ABO77783.1
1097	百日咳博德特氏菌	RFAPQ	1980	ABO77783.1
1098	麻疹病毒EDMONSTON株	SIQNKFSAGSYPLLWSYAMG	1981	P04851.1
1099	百日咳博德特氏菌	SITLQAGAH	1982	ABO77783.1
1100	百日咳博德特氏菌	SLQPED	1983	ABO77783.1
1101	百日咳博德特氏菌	SNALSKRL	1984	ABO77783.1
1102	麻疹病毒EDMONSTON株	SPGQLIQRITDDPDVSIRLL	1985	P04851.1
1103	百日咳博德特氏菌	TELPSIPG	1986	ABO77783.1
1104	百日咳博德特氏菌	TFTLANK	1987	ABO77783.1
1105	百日咳博德特氏菌	TWDDD	1988	ABO77783.1
1106	麻疹病毒EDMONSTON株	VSFLQGDQSENELPRLGGKE	1989	P04851.1
1107	麻疹病毒EDMONSTON株	WQESRKNKAQTRTPLQC	1990	P04851.1
1108	麻疹病毒EDMONSTON株	YAMGVGVELENSMGGLNFGR	1991	P04851.1
1109	麻疹病毒EDMONSTON株	YQQMGKPAPYMVNLENSI	1992	P04851.1
1110	麻疹病毒HALLE株	MTRSSHQSLVIKLMP	1993	P69355.1
1111	麻疹病毒HALLE株	PIRDALNAMTQNIRP	1994	P69355.1
1112	风疹病毒M33株	ALLNTPPPYQVSCGGESDRASAGH	1995	CAA28880.1
1113	风疹病毒THERIEN株	GLGSPNCHGPDWASPVCQRHS	1996	P07566.1
1114	风疹病毒THERIEN株	GLGSPNCHGPDWASPVCQRHSPDCSRLV	1997	P07566.1
1115	风疹病毒THERIEN株	NYTGNQQSRWGLGSPNCHGPDWASPV	1998	P07566.1
1116	风疹病毒THERIEN株	TLPQPPCAHGQHYGHHHHQL	1999	P07566.1
1117	风疹病毒THERIEN株	TVRVKFHTETRTVWQLSVAGVSCNVT	2000	P07566.1
1118	风疹病毒	ASYFNPGGSYYKQYH	2001	BAA28178.1
1119	风疹病毒	FALASYVQHPHKTVR	2002	BAA2817

[1228]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				8.1
1120	风疹病毒	GGESDRASARVIDPAAQSFTG	2003	BAA28178.1
1121	风疹病毒	GPGEVWVTPVIGSQARKC	2004	BAA28178.1
1122	风疹病毒	GSQARKCGLHIRAG	2005	BAA28178.1
1123	风疹病毒	LVGATPERPRLRLVDADDPLLRTAP	2006	BAA28178.1
1124	风疹病毒	TAPGPGEVWVTPVI	2007	BAA28178.1
1125	麻疹 麻疹病毒属	FIVLSIAYPTLSEIK	2008	AAL29688.1
1126	麻疹病毒 EDMONSTON 株-ZAGREB	ETCFQQACKGKIQALCENPEWAPLKDNRIPSY	2009	P08362.1
1127	麻疹 麻疹病毒属	LPRLGGKEDR	2010	P04851.1
1128	麻疹 麻疹病毒属	MSKTEWNASQ	2011	SRC280080
1129	麻疹 麻疹病毒属	SRFGWFENKE	2012	P04851.1
1130	风疹病毒	EACVTSWLWSEGEGAVFYRVDLHFINLGT	2013	CAA28880.1
1131	风疹病毒	MDFWCVEHDRPPPATPTSLTT	2014	CAA33016.1
1132	风疹病毒	PFLGHDGHHGGTLRVGQHHRNASDV	2015	CAA33016.1
1133	风疹病毒	RVKFHTETRTVWQLSVAGVSC	2016	BAA19893.1
1134	麻疹 麻疹病毒属	AEEQARHVKNGLE	2017	ABO69699.1
1135	麻疹 麻疹病毒属	ESPQEISKHQALG	2018	SRC280080
1136	麻疹 麻疹病毒属	GVGVELENSMGGLNF	2019	ABI54110.1
1137	麻疹 麻疹病毒属	IKGANDLAKFHQMLMKIIMK	2020	ABO69699.1
1138	麻疹 麻疹病毒属	MSKTLHAQLGFKKT	2021	ABK40528.1
1139	麻疹 麻疹病毒属	NASGLSRPSPSAH	2022	BAE98296.1
1140	麻疹 麻疹病毒属	VRVIDPSLGDRKDE	2023	SRC280148
1141	百日咳博德特氏菌	DLSDGDLLV	2024	AAC31207.1
1142	百日咳博德特氏菌	EAERAGRGTG	2025	ACI04548.1
1143	百日咳博德特氏菌	YRYDSRPPEDV	2026	ACI04548

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				.1
1144	百日咳博德特氏菌	YVDTYGDNAG	2027	ACI04548.1
1145	麻疹 麻疹病毒属	SFSYFYPFR	2028	CAB43772.1
1146	流行性腮腺炎腮腺炎病毒属	DIFIVSPR	2029	ADF49557.1
1147	麻疹 麻疹病毒属	QDSRRSADALLRLQAMAGISEEQGSDTDTPIVYNDRN	2030	CAA59302.1
1148	麻疹 麻疹病毒属	SAEALLRLQA	2031	BAH22350.1
1149	麻疹 麻疹病毒属	RIVINREHL	2032	BAB39835.1
1150	麻疹病毒GENOTYPE A	IPRFK	2033	BAB39848.1
1151	百日咳博德特氏菌	AAALSPMEI	2034	P15318.2
1152	百日咳博德特氏菌	AAASVVGAPV	2035	P15318.2
1153	百日咳博德特氏菌	AALGRQDSI	2036	P15318.2
1154	百日咳博德特氏菌	AAQRLVHAIA	2037	P15318.2
1155	百日咳博德特氏菌	AAVEAAEL	2038	P15318.2
1156	百日咳博德特氏菌	AGANVLNGL	2039	P15318.2
1157	百日咳博德特氏菌	AGYANAAD	2040	P15318.2
1158	百日咳博德特氏菌	AGYEQFEFRV	2041	P15318.2
1159	百日咳博德特氏菌	AITGNADNL	2042	P15318.2
1160	百日咳博德特氏菌	AKEKNATLM	2043	P15318.2
1161	百日咳博德特氏菌	AKGVFLSL	2044	P15318.2
1162	百日咳博德特氏菌	APHEYGFGI	2045	P15318.2
1163	百日咳博德特氏菌	ARQGNDLEI	2046	P15318.2
1164	百日咳博德特氏菌	ASVVGAPV	2047	P15318.2
1165	百日咳博德特氏菌	ATLMFRLV	2048	P15318.2
1166	百日咳博德特氏菌	AVAAAQRL	2049	P15318.2
1167	百日咳博德特氏菌	DAGANVLNGL	2050	P15318.2
1168	百日咳博德特氏菌	DALLAQLYR	2051	P15318.2
1169	百日咳博德特氏菌	DANGVLKHSI	2052	P15318.2
1170	百日咳博德特氏菌	DGDMNIGVI	2053	P15318.2
1171	百日咳博德特氏菌	DHVKNIENL	2054	P15318.2
1172	百日咳博德特氏菌	DIDMFAIM	2055	P15318.2
1173	百日咳博德特氏菌	DMFAIMPHL	2056	P15318.2
1174	百日咳博德特氏菌	DNVRNVENV	2057	P15318.2
1175	百日咳博德特氏菌	DNVRNVENVI	2058	P15318.2
1176	百日咳博德特氏菌	DTVDYSAM	2059	P15318.2
1177	百日咳博德特氏菌	DTVDYSAMI	2060	P15318.2
1178	百日咳博德特氏菌	DYLRQAGL	2061	P15318.2

[1229]

[1230]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1179	百日咳博德特氏菌	DYYDNVRNV	2062	P15318.2
1180	百日咳博德特氏菌	EFTTFVEI	2063	P15318.2
1181	百日咳博德特氏菌	EFTTFVEIV	2064	P15318.2
1182	百日咳博德特氏菌	EGYVFYEN	2065	P15318.2
1183	百日咳博德特氏菌	ENVQYRHV	2066	P15318.2
1184	百日咳博德特氏菌	EQLANSDGL	2067	P15318.2
1185	百日咳博德特氏菌	FGVGYGHDTI	2068	P15318.2
1186	百日咳博德特氏菌	FSPDVLETVP	2069	P15318.2
1187	百日咳博德特氏菌	FSVDHVKNI	2070	P15318.2
1188	百日咳博德特氏菌	GDDTYLFGV	2071	P15318.2
1189	百日咳博德特氏菌	GDDVFLQDL	2072	P15318.2
1190	百日咳博德特氏菌	GEDGNDIFL	2073	P15318.2
1191	百日咳博德特氏菌	GERFNVRKQL	2074	P15318.2
1192	百日咳博德特氏菌	GGAGNDTLV	2075	P15318.2
1193	百日咳博德特氏菌	GGDDFEAV	2076	P15318.2
1194	百日咳博德特氏菌	GKSEFTTFV	2077	P15318.2
1195	百日咳博德特氏菌	GKSLFDDGL	2078	P15318.2
1196	百日咳博德特氏菌	GNADNLKSV	2079	P15318.2
1197	百日咳博德特氏菌	GQLVEVDTL	2080	P15318.2
1198	百日咳博德特氏菌	GRSKFSPDV	2081	P15318.2
1199	百日咳博德特氏菌	GSSAYDTV	2082	P15318.2
1200	百日咳博德特氏菌	GTVEKWPAL	2083	P15318.2
1201	百日咳博德特氏菌	GVDYYDNV	2084	P15318.2
1202	百日咳博德特氏菌	GYEQFEFRV	2085	P15318.2
1203	百日咳博德特氏菌	HAVGAQDVV	2086	P15318.2
1204	百日咳博德特氏菌	IAAGRIGLGI	2087	P15318.2
1205	百日咳博德特氏菌	IGDAQANTL	2088	P15318.2
1206	百日咳博德特氏菌	IGLGILADL	2089	P15318.2
1207	百日咳博德特氏菌	IGNAAGIPL	2090	P15318.2
1208	百日咳博德特氏菌	IGTSMKDVL	2091	P15318.2
1209	百日咳博德特氏菌	IGVITDFEL	2092	P15318.2
1210	百日咳博德特氏菌	IPLTADIDM	2093	P15318.2
1211	百日咳博德特氏菌	ISKSALEL	2094	P15318.2
1212	百日咳博德特氏菌	ITGNADNL	2095	P15318.2
1213	百日咳博德特氏菌	KIFVVSAT	2096	P15318.2
1214	百日咳博德特氏菌	KQLNNANVYR	2097	P15318.2
1215	百日咳博德特氏菌	KVIGNAAGI	2098	P15318.2
1216	百日咳博德特氏菌	LAKVVSQL	2099	P15318.2
1217	百日咳博德特氏菌	LANDYARKI	2100	P15318.2
1218	百日咳博德特氏菌	LDYLRQAGL	2101	P15318.2
1219	百日咳博德特氏菌	LGKGFASL	2102	P15318.2

[1231]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1220	百日咳博德特氏菌	LGKGFASLM	2103	P15318.2
1221	百日咳博德特氏菌	LGVDYYDN	2104	P15318.2
1222	百日咳博德特氏菌	LGVDYYDNV	2105	P15318.2
1223	百日咳博德特氏菌	LKHSIKLDVI	2106	P15318.2
1224	百日咳博德特氏菌	LQAGYIPV	2107	P15318.2
1225	百日咳博德特氏菌	LQLTGGTVE	2108	P15318.2
1226	百日咳博德特氏菌	LSAAVFGL	2109	P15318.2
1227	百日咳博德特氏菌	LSLGKGFASL	2110	P15318.2
1228	百日咳博德特氏菌	LSPMEIYGL	2111	P15318.2
1229	百日咳博德特氏菌	NAHDNFLAGG	2112	P15318.2
1230	百日咳博德特氏菌	NANVYREGV	2113	P15318.2
1231	百日咳博德特氏菌	NDTLYGGL	2114	P15318.2
1232	百日咳博德特氏菌	NGLAGNDVL	2115	P15318.2
1233	百日咳博德特氏菌	NNANVYREGV	2116	P15318.2
1234	百日咳博德特氏菌	NTVSYAAL	2117	P15318.2
1235	百日咳博德特氏菌	NVLRNIENAV	2118	P15318.2
1236	百日咳博德特氏菌	PALTFITPL	2119	P15318.2
1237	百日咳博德特氏菌	PETSNVLRNI	2120	P15318.2
1238	百日咳博德特氏菌	PMEIYGLV	2121	P15318.2
1239	百日咳博德特氏菌	PQAYFEKNL	2122	P15318.2
1240	百日咳博德特氏菌	PVNPNLSKL	2123	P15318.2
1241	百日咳博德特氏菌	QAGWNASSV	2124	P15318.2
1242	百日咳博德特氏菌	QAGWNASSVI	2125	P15318.2
1243	百日咳博德特氏菌	QDAANAGNL	2126	P15318.2
1244	百日咳博德特氏菌	QDAANAGNLL	2127	P15318.2
1245	百日咳博德特氏菌	QDSGYDSL	2128	P15318.2
1246	百日咳博德特氏菌	QQSHYADQL	2129	P15318.2
1247	百日咳博德特氏菌	RALQGAQAV	2130	P15318.2
1248	百日咳博德特氏菌	RGGLGLDTL	2131	P15318.2
1249	百日咳博德特氏菌	RKQLNNANV	2132	P15318.2
1250	百日咳博德特氏菌	RQDSGYDSL	2133	P15318.2
1251	百日咳博德特氏菌	RQFRYDGDM	2134	P15318.2
1252	百日咳博德特氏菌	RSKFSPDVL	2135	P15318.2
1253	百日咳博德特氏菌	SAGAAAGAL	2136	P15318.2
1254	百日咳博德特氏菌	SAHWGQRAL	2137	P15318.2
1255	百日咳博德特氏菌	SAMIHPGRI	2138	P15318.2
1256	百日咳博德特氏菌	SAMIHPGRIV	2139	P15318.2
1257	百日咳博德特氏菌	SAYDTVSGI	2140	P15318.2
1258	百日咳博德特氏菌	SAYGYEGD	2141	P15318.2
1259	百日咳博德特氏菌	SGGAGDDVL	2142	P15318.2
1260	百日咳博德特氏菌	SGLQVAGA	2143	P15318.2

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1261	百日咳博德特氏菌	SGYDSLDGV	2144	P15318.2
1262	百日咳博德特氏菌	SLLTGALNGI	2145	P15318.2
1263	百日咳博德特氏菌	SPMEIYGL	2146	P15318.2
1264	百日咳博德特氏菌	SQMLTRGQL	2147	P15318.2
1265	百日咳博德特氏菌	SSAYDTVSGI	2148	P15318.2
1266	百日咳博德特氏菌	SSLAHGHTA	2149	P15318.2
1267	百日咳博德特氏菌	SSVTSGDSV	2150	P15318.2
1268	百日咳博德特氏菌	SVIGVQTTEI	2151	P15318.2
1269	百日咳博德特氏菌	TNTVSYAAL	2152	P15318.2
1270	百日咳博德特氏菌	TSLIAEGV	2153	P15318.2
1271	百日咳博德特氏菌	TSLLTGAL	2154	P15318.2
1272	百日咳博德特氏菌	TVPASPGL	2155	P15318.2
1273	百日咳博德特氏菌	VAKEKNATL	2156	P15318.2
1274	百日咳博德特氏菌	VAPHEYGFGI	2157	P15318.2
1275	百日咳博德特氏菌	VAVVTSLL	2158	P15318.2
1276	百日咳博德特氏菌	VFYENRAYG	2159	P15318.2
1277	百日咳博德特氏菌	VFYENRAYGV	2160	P15318.2
1278	百日咳博德特氏菌	VITDFELEV	2161	P15318.2
1279	百日咳博德特氏菌	VNPHSTSL	2162	P15318.2
[1232] 1280	百日咳博德特氏菌	VNPHSTSLI	2163	P15318.2
1281	百日咳博德特氏菌	VNPNLSKL	2164	P15318.2
1282	百日咳博德特氏菌	VNPNLSKLF	2165	P15318.2
1283	百日咳博德特氏菌	VQQPIIEKL	2166	P15318.2
1284	百日咳博德特氏菌	VQYRHVEL	2167	P15318.2
1285	百日咳博德特氏菌	VSIAAAASV	2168	P15318.2
1286	百日咳博德特氏菌	VSIAAAASVV	2169	P15318.2
1287	百日咳博德特氏菌	VTSLLTGAL	2170	P15318.2
1288	百日咳博德特氏菌	VVLANASRI	2171	P15318.2
1289	百日咳博德特氏菌	WPALNLFSV	2172	P15318.2
1290	百日咳博德特氏菌	WVRKASAL	2173	P15318.2
1291	百日咳博德特氏菌	YAVQYRRKGG	2174	P15318.2
1292	百日咳博德特氏菌	YGGLGDDTL	2175	P15318.2
1293	百日咳博德特氏菌	YGLVQQSHYA	2176	P15318.2
1294	百日咳博德特氏菌	YGYEGDALL	2177	P15318.2
1295	百日咳博德特氏菌	YIPVNPNL	2178	P15318.2
1296	百日咳博德特氏菌	YSAMIHPGRI	2179	P15318.2
1297	百日咳博德特氏菌	YSQTGAHAGI	2180	P15318.2
1298	白喉杆菌	AYNFVESIINLFQVVHNSYN	2181	CAE11230.1
1299	百日咳博德特氏菌	SGTTIK	2182	BAF35031.1
1300	百日咳博德特氏菌	RGHTLESAEGRKIFG	2183	AAA2297

[1233]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				4.1
1301	百日咳博德特氏菌	AGAMTVRDVAAAADLALQAGDA	2184	AAA22974.1
1302	百日咳博德特氏菌	AGAMTVRDVAAAADLALQAGDAL	2185	AAA22974.1
1303	百日咳博德特氏菌	ALAAVLVNPHIFTRIGAAQTSLADGAAGPA	2186	AAA22974.1
1304	百日咳博德特氏菌	ALSIDSMTALGA	2187	AAA22974.1
1305	百日咳博德特氏菌	DLSAARGADISGEGR	2188	AAA22974.1
1306	百日咳博德特氏菌	DQNRYEYIWGLY	2189	AAA22974.1
1307	百日咳博德特氏菌	DYTVSADAIALA	2190	AAA22974.1
1308	百日咳博德特氏菌	GPIVVEAGELVSHAGG	2191	AAA22974.1
1309	百日咳博德特氏菌	GRPEGLKIGAHSATSVSGSFDAL	2192	AAA22974.1
1310	百日咳博德特氏菌	ITVTSRGGFDNEGKMESNK	2193	AAA22974.1
1311	百日咳博德特氏菌	LDQNRYEYIWGLYP	2194	AAA22974.1
1312	百日咳博德特氏菌	LSAARGADISG	2195	AAA22974.1
1313	百日咳博德特氏菌	NKIRLMGPLQ	2196	AAA22974.1
1314	百日咳博德特氏菌	NKLGRIRAGEDM	2197	AAA22974.1
1315	百日咳博德特氏菌	NKLGRIRAGEDMHLDAPRIE	2198	AAA22974.1
1316	百日咳博德特氏菌	PHLRNTGQVVAG	2199	AAA22974.1
1317	百日咳博德特氏菌	QVDLHDLSAARGADISG	2200	AAA22974.1
1318	百日咳博德特氏菌	RDVAAAADLALQ	2201	AAA22974.1
1319	百日咳博德特氏菌	SAARGADISGEG	2202	AAA22974.1
1320	百日咳博德特氏菌	TKGEMQIAGKGGGSP	2203	AAA22974.1
1321	百日咳博德特氏菌	TVSADAIALAAQ	2204	AAA22974.1
1322	百日咳博德特氏菌	VVAGHDIHI	2205	AAA22974.1
1323	百日咳博德特氏菌	PSGPNHTKVVQLPKISKNALKANG	2206	CAD12823.1
1324	风疹病毒	LVGATPERPRLRLVDADDPLLRTA	2207	BAA1990

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
		PGPGEVWVTPVIGSQAR		2.1
1325	风疹病毒	QQSRWGLGSPNCHGPDWASPVCQRHSP	2208	BAA19902.1
1326	百日咳博德特氏菌	AGEAMVLVYYESIAYSF	2209	ACI04548.1
1327	百日咳博德特氏菌	GGVGLASTLWYAESNALSKRLGEL	2210	AAZ74322.1
1328	百日咳博德特氏菌	GTLVRIAPVIGACMARQA	2211	ACI04548.1
1329	百日咳博德特氏菌	IRRVTRVYHNGITGETTT	2212	ACI04548.1
1330	百日咳博德特氏菌	IVKTGERQHGIHIQGSDP	2213	AAZ74322.1
1331	百日咳博德特氏菌	IVKTGERQHGIHIQGSDPGGVRTA	2214	AAZ74338.1
1332	百日咳博德特氏菌	LRDTNVTAVPASGAPAAVSVLGAS	2215	AAZ74338.1
1333	百日咳博德特氏菌	PEAPAPQPPAGRELSAAANAAVNT	2216	AAZ74322.1
1334	百日咳博德特氏菌	AAADFAHAE	2217	WP_019247158.1
1335	百日咳博德特氏菌	AAAEVAGAL	2218	WP_019249248.1
1336	百日咳博德特氏菌	AAESTFESY	2219	WP_019247158.1
1337	百日咳博德特氏菌	AAGFDPEVQ	2220	WP_019248145.1
1338	百日咳博德特氏菌	AALGRGHSL	2221	AGS56996.1
1339	百日咳博德特氏菌	AAMQGAVVH	2222	AGT50936.1
1340	百日咳博德特氏菌	AAPAAHADW	2223	AGS56996.1
1341	百日咳博德特氏菌	AAQATVVQR	2224	WP_019247158.1
1342	百日咳博德特氏菌	AARVAGDNY	2225	WP_019249248.1
1343	百日咳博德特氏菌	AAVALLNKL	2226	WP_019249248.1
1344	百日咳博德特氏菌	ADDPPATVY	2227	AAW72734.1
1345	百日咳博德特氏菌	AEAGRFKVL	2228	AGS56996.1
1346	百日咳博德特氏菌	AEATQLVTA	2229	WP_019247158.1
1347	百日咳博德特氏菌	AEGGATLGA	2230	WP_019249248.1
1348	百日咳博德特氏菌	AEHGEVSIQ	2231	WP_0192

[1234]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				49248.1
1349	百日咳博德特氏菌	AEIAFYPKE	2232	WP_019249248.1
1350	百日咳博德特氏菌	AEKVTTPAV	2233	WP_019247158.1
1351	百日咳博德特氏菌	AELQTYLRQ	2234	1BCP_C
1352	百日咳博德特氏菌	AEQSLIEVG	2235	WP_019249248.1
1353	百日咳博德特氏菌	AESNALSKR	2236	AGS56996.1
1354	百日咳博德特氏菌	AESSEAMAA	2237	AFK26302.1
1355	百日咳博德特氏菌	AEVKVGYRA	2238	WP_019247158.1
1356	百日咳博德特氏菌	AEVTDTSPS	2239	WP_019249248.1
1357	百日咳博德特氏菌	AGKSLKKKN	2240	WP_019247158.1
1358	百日咳博德特氏菌	AGLAGPSAV	2241	WP_019249248.1
1359	百日咳博德特氏菌	AIRVGRGAR	2242	AGT50936.1
1360	百日咳博德特氏菌	ALAAIASAA	2243	WP_019248145.1
1361	百日咳博德特氏菌	ALADVPYVL	2244	AAA22983.1
1362	百日咳博德特氏菌	ALANDGTIV	2245	WP_019248658.1
1363	百日咳博德特氏菌	ALGRGHSLY	2246	AGS56996.1
1364	百日咳博德特氏菌	ALILAASPV	2247	WP_019248658.1
1365	百日咳博德特氏菌	ALMLACTGL	2248	AAA22974.1
1366	百日咳博德特氏菌	AMQGAVVHL	2249	AGT50936.1
1367	百日咳博德特氏菌	AMTHLSPAL	2250	AAA22983.1
1368	百日咳博德特氏菌	AMYGKHITL	2251	WP_019249248.1
1369	百日咳博德特氏菌	ANEANALLW	2252	WP_019249248.1
1370	百日咳博德特氏菌	APLSITLQA	2253	AGT50936.1
1371	百日咳博德特氏菌	APNALAWAL	2254	AAA22974.1
1372	百日咳博德特氏菌	APPAPKPAP	2255	AGS56996.1

[1235]

[1236]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1373	百日咳博德特氏菌	APPGAGFIY	2256	1BCP_C
1374	百日咳博德特氏菌	APQAAPLSI	2257	AGT50936.1
1375	百日咳博德特氏菌	APRIENTAK	2258	WP_019249248.1
1376	百日咳博德特氏菌	AQGKALLYR	2259	AGT50936.1
1377	百日咳博德特氏菌	AQITSYVGF	2260	WP_019248658.1
1378	百日咳博德特氏菌	AQLEVRGQR	2261	WP_019249248.1
1379	百日咳博德特氏菌	AQQLKQADR	2262	WP_019247699.1
1380	百日咳博德特氏菌	AQVTVAGRY	2263	WP_019249248.1
1381	百日咳博德特氏菌	ARRSRVRAL	2264	NP_882284.1
1382	百日咳博德特氏菌	ASPRRARRA	2265	WP_019249248.1
1383	百日咳博德特氏菌	ASSPDAHVP	2266	AAA22983.1
1384	百日咳博德特氏菌	ASVSNPGTF	2267	WP_019249248.1
1385	百日咳博德特氏菌	ATWNFQSTY	2268	WP_019249248.1
1386	百日咳博德特氏菌	ATYIADSGF	2269	AGS56996.1
1387	百日咳博德特氏菌	AVAAPAVGA	2270	WP_019249248.1
1388	百日咳博德特氏菌	AVFMQQRPL	2271	AAA22983.1
1389	百日咳博德特氏菌	AVLVNPHIF	2272	WP_019249248.1
1390	百日咳博德特氏菌	CFGKDLKRP	2273	AAA22983.1
1391	百日咳博德特氏菌	CPSSLGNGV	2274	WP_019248145.1
1392	百日咳博德特氏菌	DAGHEHDTW	2275	AAA22984.1
1393	百日咳博德特氏菌	DAKHDLTVT	2276	WP_019249248.1
1394	百日咳博德特氏菌	DASGQHRLW	2277	AGS56996.1
1395	百日咳博德特氏菌	DATFETYAL	2278	YP_006628018.1
1396	百日咳博德特氏菌	DATFQTYAL	2279	1BCP_C
1397	百日咳博德特氏菌	DATLVGAKF	2280	WP_019247158.1

[1237]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1398	百日咳博德特氏菌	DDEVDVSGR	2281	AAA22974.1
1399	百日咳博德特氏菌	DENGKPQTY	2282	WP_019247158.1
1400	百日咳博德特氏菌	DGPPSRPTT	2283	WP_019247158.1
1401	百日咳博德特氏菌	DHLTGRSCQ	2284	AAW72734.1
1402	百日咳博德特氏菌	DNEGKMESN	2285	WP_019249248.1
1403	百日咳博德特氏菌	DPPATVYRY	2286	AAW72734.1
1404	百日咳博德特氏菌	EATEGDATL	2287	WP_019247158.1
1405	百日咳博德特氏菌	EATQQAAGF	2288	WP_019248145.1
1406	百日咳博德特氏菌	ECSGKQDCP	2289	AAA22983.1
1407	百日咳博德特氏菌	EGGKLRGKD	2290	AAA22974.1
1408	百日咳博德特氏菌	EGKMESNKD	2291	WP_019249248.1
1409	百日咳博德特氏菌	EHRMQEAVE	2292	AAW72734.1
1410	百日咳博德特氏菌	EKRLDIDDA	2293	WP_019249248.1
1411	百日咳博德特氏菌	EPASANTLL	2294	AGS56996.1
1412	百日咳博德特氏菌	EPQAELAVF	2295	AGS56996.1
1413	百日咳博德特氏菌	EPVKLTLTG	2296	AGT50936.1
1414	百日咳博德特氏菌	ESAEGRKIF	2297	WP_019249248.1
1415	百日咳博德特氏菌	ESYSESHNF	2298	WP_019247158.1
1416	百日咳博德特氏菌	ETFCITTIY	2299	1BCP_C
1417	百日咳博德特氏菌	EVAGALELS	2300	WP_019249248.1
1418	百日咳博德特氏菌	EVAKVEVVP	2301	WP_019247158.1
1419	百日咳博德特氏菌	EVDGIIQEF	2302	WP_019249248.1
1420	百日咳博德特氏菌	EVRADNNFY	2303	WP_019248344.1
1421	百日咳博德特氏菌	FAILSSTTE	2304	WP_019247158.1
1422	百日咳博德特氏菌	FAISAYALK	2305	AAA2298

[1238]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				4.1
1423	百日咳博德特氏菌	FALYDGTYL	2306	AFK26303.1
1424	百日咳博德特氏菌	FDTMLGFAI	2307	AAA22984.1
1425	百日咳博德特氏菌	FEGKPALEL	2308	AAA22983.1
1426	百日咳博德特氏菌	FELGADHAV	2309	AGS56996.1
1427	百日咳博德特氏菌	FEPGITTNY	2310	WP_019248658.1
1428	百日咳博德特氏菌	FETYALTGI	2311	YP_006628018.1
1429	百日咳博德特氏菌	FIYRETFCI	2312	1BCP_C
1430	百日咳博德特氏菌	FPTRTTAPG	2313	NP_882284.1
1431	百日咳博德特氏菌	FQTYALTGI	2314	1BCP_C
1432	百日咳博德特氏菌	FTHADGWFL	2315	AGS56996.1
1433	百日咳博德特氏菌	FVRDGQSVI	2316	1BCP_C
1434	百日咳博德特氏菌	FVRSGQPVI	2317	YP_006628018.1
1435	百日咳博德特氏菌	FVWYVDTVI	2318	WP_019248866.1
1436	百日咳博德特氏菌	GAASSRQAL	2319	WP_019249248.1
1437	百日咳博德特氏菌	GAASSYFEY	2320	AAW72734.1
1438	百日咳博德特氏菌	GAFDLKTTF	2321	AFK26303.1
1439	百日咳博德特氏菌	GAPAAVSVL	2322	AGS56996.1
1440	百日咳博德特氏菌	GATRAVDSL	2323	AGS56996.1
1441	百日咳博德特氏菌	GAVPGGAVP	2324	AGT50936.1
1442	百日咳博德特氏菌	GEAMVLVYY	2325	AFK26302.1
1443	百日咳博德特氏菌	GEIALGDAT	2326	WP_019249248.1
1444	百日咳博德特氏菌	GELMAAQVA	2327	WP_019247158.1
1445	百日咳博德特氏菌	GGVPGGAVP	2328	AAZ74338.1
1446	百日咳博德特氏菌	GHEHDTWFD	2329	AAA22984.1
1447	百日咳博德特氏菌	GIGALKAGA	2330	WP_0192

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				49248.1
1448	百日咳博德特氏菌	GIVIPPKAL	2331	1BCP_C
1449	百日咳博德特氏菌	GKDLKRPGS	2332	AAA22983.1
1450	百日咳博德特氏菌	GKLPKPVTV	2333	WP_019247158.1
1451	百日咳博德特氏菌	GKSLKKKNQ	2334	WP_019247158.1
1452	百日咳博德特氏菌	GLDVQQGTV	2335	WP_019249248.1
1453	百日咳博德特氏菌	GLTDGVSRI	2336	WP_019249248.1
1454	百日咳博德特氏菌	GLYPTYTEW	2337	WP_019249248.1
1455	百日咳博德特氏菌	GLYQTYTEW	2338	YP_006626470.1
1456	百日咳博德特氏菌	GPPSRPTTP	2339	WP_019247158.1
1457	百日咳博德特氏菌	GPSAVAAPA	2340	WP_019249248.1
1458	百日咳博德特氏菌	GVAPTAQQL	2341	WP_019248866.1
1459	百日咳博德特氏菌	GVGLASTLW	2342	AGS56996.1
1460	百日咳博德特氏菌	GYEAGFSLG	2343	WP_019247158.1
1461	百日咳博德特氏菌	HADDGTIVI	2344	WP_019248145.1
1462	百日咳博德特氏菌	HADWNNQSI	2345	AGS56996.1
1463	百日咳博德特氏菌	HAEHEKDVR	2346	WP_019247158.1
1464	百日咳博德特氏菌	HANHYGTRI	2347	WP_019247158.1
1465	百日咳博德特氏菌	HAQGKALLY	2348	AGT50936.1
1466	百日咳博德特氏菌	HFIGYIYEV	2349	AAW72734.1
1467	百日咳博德特氏菌	HLSPALADV	2350	AAA22983.1
1468	百日咳博德特氏菌	HSLYASYEY	2351	AGS56996.1
1469	百日咳博德特氏菌	HVRGMLVPV	2352	AAA22974.1
1470	百日咳博德特氏菌	HVSKEEQYY	2353	YP_006628018.1
1471	百日咳博德特氏菌	HVTRGWSIF	2354	AFK26303.1

[1239]

[1240]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1472	百日咳博德特氏菌	IADSGFYLD	2355	AGS56996.1
1473	百日咳博德特氏菌	IAHRTELRG	2356	AGS56996.1
1474	百日咳博德特氏菌	IENTAKLSG	2357	WP_019249248.1
1475	百日咳博德特氏菌	IESKISQSV	2358	WP_019249248.1
1476	百日咳博德特氏菌	IETGGARRF	2359	AGT50936.1
1477	百日咳博德特氏菌	IIKDAPPGA	2360	1BCP_C
1478	百日咳博德特氏菌	IIQEFAADL	2361	WP_019249248.1
1479	百日咳博德特氏菌	ILAGALATY	2362	AAW72734.1
1480	百日咳博德特氏菌	ILLENPAAE	2363	AGS56996.1
1481	百日咳博德特氏菌	ILPILVLAL	2364	NP_882286.1
1482	百日咳博德特氏菌	IPFQRALRL	2365	WP_019248145.1
1483	百日咳博德特氏菌	ISVRVHVSK	2366	YP_006628018.1
1484	百日咳博德特氏菌	ITNETGKTY	2367	WP_019247158.1
1485	百日咳博德特氏菌	ITNKRAALI	2368	WP_019249248.1
1486	百日咳博德特氏菌	ITSYVGFSV	2369	WP_019248658.1
1487	百日咳博德特氏菌	ITVTSRGGF	2370	WP_019249248.1
1488	百日咳博德特氏菌	IVIPPKALF	2371	1BCP_C
1489	百日咳博德特氏菌	IVVEAGELV	2372	WP_019249248.1
1490	百日咳博德特氏菌	KAAKSVNLM	2373	WP_019247158.1
1491	百日咳博德特氏菌	KAAPLRRTT	2374	AGS56996.1
1492	百日咳博德特氏菌	KAGKLSATG	2375	WP_019249248.1
1493	百日咳博德特氏菌	KAGTIAAPW	2376	WP_019249248.1
1494	百日咳博德特氏菌	KAKSLTTEI	2377	WP_019249248.1
1495	百日咳博德特氏菌	KATVTTVQV	2378	WP_019247158.1
1496	百日咳博德特氏菌	KDYRDKDGG	2379	WP_019247158.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1497	百日咳博德特氏菌	KEAATIVAA	2380	WP_019249248.1
1498	百日咳博德特氏菌	KEDVDAAQI	2381	WP_019248658.1
1499	百日咳博德特氏菌	KEVDGIIQE	2382	WP_019249248.1
1500	百日咳博德特氏菌	KGPKLAMPW	2383	AGS56996.1
1501	百日咳博德特氏菌	KLASGGGAV	2384	WP_019249248.1
1502	百日咳博德特氏菌	KLKGKNQEF	2385	AAA22984.1
1503	百日咳博德特氏菌	KLLHHILPI	2386	NP_882286.1
1504	百日咳博德特氏菌	KPAPQPGPQ	2387	AGS56996.1
1505	百日咳博德特氏菌	KPAPTAPPM	2388	WP_019249248.1
1506	百日咳博德特氏菌	KPAVSVKVA	2389	WP_019249248.1
1507	百日咳博德特氏菌	KPDRAARVA	2390	WP_019249248.1
[1241] 1508	百日咳博德特氏菌	KPLADIAVI	2391	YP_006626470.1
1509	百日咳博德特氏菌	KPLADIAVV	2392	WP_019249248.1
1510	百日咳博德特氏菌	KPLPKPLPV	2393	WP_019247158.1
1511	百日咳博德特氏菌	KQADRDFVW	2394	WP_019247699.1
1512	百日咳博德特氏菌	KSLPGGKLP	2395	WP_019247158.1
1513	百日咳博德特氏菌	KSYTLRYLA	2396	WP_019248658.1
1514	百日咳博德特氏菌	KTNMVVTSV	2397	AAA22983.1
1515	百日咳博德特氏菌	KVLAPRLYL	2398	AAA22974.1
1516	百日咳博德特氏菌	KVLSTKTTL	2399	WP_019247158.1
1517	百日咳博德特氏菌	LAAGAGLTL	2400	WP_019249248.1
1518	百日咳博德特氏菌	LAAIASAAH	2401	WP_019248145.1
1519	百日咳博德特氏菌	LAANGNGQW	2402	AGS56996.1
1520	百日咳博德特氏菌	LAAQVTQRG	2403	WP_019249248.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1521	百日咳博德特氏菌	LAARGDGAL	2404	AAA22974.1
1522	百日咳博德特氏菌	LAAVLVNPH	2405	WP_019249248.1
1523	百日咳博德特氏菌	LAGSGLFRM	2406	AGS56996.1
1524	百日咳博德特氏菌	LAKALSAAL	2407	WP_019249248.1
1525	百日咳博德特氏菌	LALQAGDAL	2408	WP_019249248.1
1526	百日咳博德特氏菌	LAMPWTFHA	2409	AGS56996.1
1527	百日咳博德特氏菌	LANDGTIVI	2410	WP_019248658.1
1528	百日咳博德特氏菌	LAPTVGVAF	2411	WP_019247158.1
1529	百日咳博德特氏菌	LASDGSVDF	2412	AGS56996.1
1530	百日咳博德特氏菌	LATYQSEYL	2413	AAW72734.1
1531	百日咳博德特氏菌	LAWALMLAC	2414	AAA22974.1
1532	百日咳博德特氏菌	LEAGRRFTH	2415	AGS56996.1
1533	百日咳博德特氏菌	LFTQQGGAY	2416	1BCP_C
1534	百日咳博德特氏菌	LKRPGSSPM	2417	AAA22983.1
1535	百日咳博德特氏菌	LLGSHVARA	2418	AFK26303.1
1536	百日咳博德特氏菌	LLHHILPIL	2419	NP_882286.1
1537	百日咳博德特氏菌	LLNAGGTLL	2420	WP_019249248.1
1538	百日咳博德特氏菌	LMGPLQVNA	2421	WP_019249248.1
1539	百日咳博德特氏菌	LNDSKITMG	2422	WP_019248145.1
1540	百日咳博德特氏菌	LPEPVKLTL	2423	AGT50936.1
1541	百日咳博德特氏菌	LPILVLALL	2424	NP_882286.1
1542	百日咳博德特氏菌	LPKISKNAL	2425	WP_019248145.1
1543	百日咳博德特氏菌	LPKPVTVKL	2426	WP_019247158.1
1544	百日咳博德特氏菌	LPLKANPMH	2427	NP_882285.1
1545	百日咳博德特氏菌	LPPRPVVAE	2428	WP_0192

[1242]

[1243]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				47158.1
1546	百日咳博德特氏菌	LPSIPGTSI	2429	AGS56996.1
1547	百日咳博德特氏菌	LPTHLYKNF	2430	AAA22984.1
1548	百日咳博德特氏菌	LPVRGVALV	2431	WP_019249248.1
1549	百日咳博德特氏菌	LPVSLTALD	2432	WP_019249248.1
1550	百日咳博德特氏菌	LQGGNKVPV	2433	WP_019249248.1
1551	百日咳博德特氏菌	LSAALGADW	2434	WP_019249248.1
1552	百日咳博德特氏菌	LSDAGHEHD	2435	AAA22984.1
1553	百日咳博德特氏菌	LSGEVQRKG	2436	WP_019249248.1
1554	百日咳博德特氏菌	LSSPSAITV	2437	WP_019249248.1
1555	百日咳博德特氏菌	LTWLAILAV	2438	AAW72734.1
1556	百日咳博德特氏菌	LVFSHVRGM	2439	AAA22974.1
1557	百日咳博德特氏菌	LVSDAGADL	2440	WP_019249248.1
1558	百日咳博德特氏菌	LVYYESIAY	2441	AFK263022.1
1559	百日咳博德特氏菌	MAAESTFES	2442	WP_019247158.1
1560	百日咳博德特氏菌	MAAGHDATL	2443	WP_019249248.1
1561	百日咳博德特氏菌	MAAWSERAG	2444	AFK263022.1
1562	百日咳博德特氏菌	MALGALGAA	2445	AGS56996.1
1563	百日咳博德特氏菌	MAPVMGACM	2446	ADA85123.1
1564	百日咳博德特氏菌	MATKGEMQI	2447	WP_019249248.1
1565	百日咳博德特氏菌	MDAKGGTLL	2448	WP_019249248.1
1566	百日咳博德特氏菌	MESNKDIVI	2449	WP_019249248.1
1567	百日咳博德特氏菌	MEVMLRAVF	2450	AAA22983.1
1568	百日咳博德特氏菌	MEYFKTPLP	2451	WP_019249248.1
1569	百日咳博德特氏菌	MHTIASILL	2452	AAA2298

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				4.1
1570	百日咳博德特氏菌	MIYMSGLAV	2453	1BCP_C
1571	百日咳博德特氏菌	MLACTGLPL	2454	AAA22974.1
1572	百日咳博德特氏菌	MPIDRKTLC	2455	AFK26303.1
1573	百日咳博德特氏菌	MPKAPELDL	2456	WP_019249248.1
1574	百日咳博德特氏菌	MQRQAGLPL	2457	NP_882285.1
1575	百日咳博德特氏菌	NALAWALML	2458	AAA22974.1
1576	百日咳博德特氏菌	NITNKRAAL	2459	WP_019249248.1
1577	百日咳博德特氏菌	NLMAAESTF	2460	WP_019247158.1
1578	百日咳博德特氏菌	NNETMSGRQ	2461	WP_019249248.1
1579	百日咳博德特氏菌	NPGSLIAEV	2462	WP_019249248.1
1580	百日咳博德特氏菌	NPMHTIASI	2463	NP_882285.1
[1244] 1581	百日咳博德特氏菌	NPQTQLSNI	2464	WP_019248145.1
1582	百日咳博德特氏菌	NPYTSRRSV	2465	AFK26302.1
1583	百日咳博德特氏菌	PAPTAPPMP	2466	WP_019249248.1
1584	百日咳博德特氏菌	PASANTLLL	2467	AGS56996.1
1585	百日咳博德特氏菌	PAVALPRPL	2468	WP_019249248.1
1586	百日咳博德特氏菌	PDAHVPFCF	2469	AAA22983.1
1587	百日咳博德特氏菌	PELGAAIRV	2470	AGT50936.1
1588	百日咳博德特氏菌	PFIIKLKDC	2471	WP_019248145.1
1589	百日咳博德特氏菌	PGPQPPQPP	2472	AGS56996.1
1590	百日咳博德特氏菌	PGPQPPQPQ	2473	AAZ74338.1
1591	百日咳博德特氏菌	PGPTTDYST	2474	WP_019248145.1
1592	百日咳博德特氏菌	PGTFTAGKD	2475	WP_019249248.1
1593	百日咳博德特氏菌	PGTPGDLLE	2476	AAA22984.1

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1594	百日咳博德特氏菌	PKPKPKAER	2477	WP_019247158.1
1595	百日咳博德特氏菌	PKPKPKPKA	2478	WP_019247158.1
1596	百日咳博德特氏菌	PKPKPKPKP	2479	WP_019247158.1
1597	百日咳博德特氏菌	PLPPRPVVA	2480	WP_019247158.1
1598	百日咳博德特氏菌	PPAPKPAPQ	2481	AGS56996.1
1599	百日咳博德特氏菌	PPKPAPVAK	2482	WP_019247158.1
1600	百日咳博德特氏菌	PPRPVAAQV	2483	WP_019247158.1
1601	百日咳博德特氏菌	PPRPVVAEK	2484	WP_019247158.1
1602	百日咳博德特氏菌	PPSRPTTPP	2485	WP_019247158.1
1603	百日咳博德特氏菌	PRRARRALR	2486	WP_019249248.1
1604	百日咳博德特氏菌	QAAPLSITL	2487	AGT50936.1
1605	百日咳博德特氏菌	QADRDFVWY	2488	WP_019247699.1
1606	百日咳博德特氏菌	QAIVVGKDL	2489	WP_019249248.1
1607	百日咳博德特氏菌	QALGALKLY	2490	ACI16088.1
1608	百日咳博德特氏菌	QELALKLKG	2491	AAA22984.1
1609	百日咳博德特氏菌	QITQHGGPY	2492	NP_882283.1
1610	百日咳博德特氏菌	QITQHGSPY	2493	AFK26303.1
1611	百日咳博德特氏菌	QPLPPRPVA	2494	WP_019247158.1
1612	百日咳博德特氏菌	QPPAGRELS	2495	AGS56996.1
1613	百日咳博德特氏菌	QQLKQADRD	2496	WP_019247699.1
1614	百日咳博德特氏菌	QQVQVLQRQ	2497	WP_019247158.1
1615	百日咳博德特氏菌	QSIVEAPEL	2498	AGT50936.1
1616	百日咳博德特氏菌	QVGSSNSAF	2499	AAW72734.1
1617	百日咳博德特氏菌	RAGLSPATW	2500	WP_019249248.1

[1245]

[1246]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1618	百日咳博德特氏菌	RARRALRQD	2501	WP_019249248.1
1619	百日咳博德特氏菌	RASASRARI	2502	WP_019249248.1
1620	百日咳博德特氏菌	RELSAAANA	2503	AGS56996.1
1621	百日咳博德特氏菌	RETFCITTI	2504	1BCP_C
1622	百日咳博德特氏菌	RGFAQRQQL	2505	AGS56996.1
1623	百日咳博德特氏菌	RGSAATFTL	2506	WP_019248295.1
1624	百日咳博德特氏菌	RGWSIFALY	2507	AFK26303.1
1625	百日咳博德特氏菌	RKMLYLIYV	2508	YP_006628018.1
1626	百日咳博德特氏菌	RLRKMLYLI	2509	YP_006628018.1
1627	百日咳博德特氏菌	RPQITDAVT	2510	WP_019249248.1
1628	百日咳博德特氏菌	RPSVNGGRI	2511	WP_019249248.1
1629	百日咳博德特氏菌	RRFTHADGW	2512	AGS56996.1
1630	百日咳博德特氏菌	RSGARATSL	2513	AAA22974.1
1631	百日咳博德特氏菌	RSRVRALAW	2514	NP_882284.1
1632	百日咳博德特氏菌	RSRVRALSW	2515	YP_006628019.1
1633	百日咳博德特氏菌	RTHGVGASL	2516	AGS56996.1
1634	百日咳博德特氏菌	RTRGQARSG	2517	AAA22974.1
1635	百日咳博德特氏菌	RVAPPAVAL	2518	WP_019249248.1
1636	百日咳博德特氏菌	RVLPEPVKL	2519	AGT50936.1
1637	百日咳博德特氏菌	RVRALAWLL	2520	NP_882284.1
1638	百日咳博德特氏菌	RVRALSWLL	2521	YP_006628019.1
1639	百日咳博德特氏菌	RVTVSGGSL	2522	AGT50936.1
1640	百日咳博德特氏菌	SEAMAAWSE	2523	AFK26302.1
1641	百日咳博德特氏菌	SESHNFHAS	2524	WP_019247158.1
1642	百日咳博德特氏菌	SGEGRVNIG	2525	WP_0192

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				49248.1
1643	百日咳博德特氏菌	SGLAVRVHV	2526	1BCP_C
1644	百日咳博德特氏菌	SLADISLGA	2527	WP_019249248.1
1645	百日咳博德特氏菌	SLFAILSST	2528	WP_019247158.1
1646	百日咳博德特氏菌	SLFAPHGNV	2529	AAZ74338.1
1647	百日咳博德特氏菌	SLSIDNATW	2530	AGS56996.1
1648	百日咳博德特氏菌	SPMEVMLRA	2531	AAA22983.1
1649	百日咳博德特氏菌	SPQPIRATV	2532	WP_019247158.1
1650	百日咳博德特氏菌	SPRRARRAL	2533	WP_019249248.1
1651	百日咳博德特氏菌	SPSRLAGTL	2534	WP_019249248.1
1652	百日咳博德特氏菌	SSTPLGSLF	2535	WP_019247158.1
1653	百日咳博德特氏菌	STYELLDYL	2536	WP_019249248.1
1654	百日咳博德特氏菌	SVAMKPYEV	2537	AAA22983.1
1655	百日咳博德特氏菌	SVAPNALAW	2538	AAA22974.1
1656	百日咳博德特氏菌	SVKVAKKLF	2539	WP_019249248.1
1657	百日咳博德特氏菌	TAFMSGRSL	2540	AAA22984.1
1658	百日咳博德特氏菌	TAGATPFDI	2541	WP_019248658.1
1659	百日咳博德特氏菌	TAPVTSPAW	2542	AAW72734.1
1660	百日咳博德特氏菌	TARTGWLTW	2543	AAW72734.1
1661	百日咳博德特氏菌	TEAQGVQVR	2544	WP_019248145.1
1662	百日咳博德特氏菌	TEVYLEHRM	2545	AAW72734.1
1663	百日咳博德特氏菌	TFEGKPALE	2546	AAA22983.1
1664	百日咳博德特氏菌	TFTGKVTNG	2547	WP_019248658.1
1665	百日咳博德特氏菌	TGDGGGHTD	2548	AGS56996.1
1666	百日咳博德特氏菌	TLAKALSAA	2549	WP_019249248.1

[1247]

[1248]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1667	百日咳博德特氏菌	TLANVGDTW	2550	AGS56996.1
1668	百日咳博德特氏菌	TLNASNLTL	2551	WP_019249248.1
1669	百日咳博德特氏菌	TLSSAHGNV	2552	WP_019249248.1
1670	百日咳博德特氏菌	TPFDIKLKE	2553	WP_019248658.1
1671	百日咳博德特氏菌	TPFIIKLKD	2554	WP_019248145.1
1672	百日咳博德特氏菌	TPGWSIYGL	2555	1BCP_C
1673	百日咳博德特氏菌	TPLGSAATF	2556	AGS56996.1
1674	百日咳博德特氏菌	TPLGSLFAI	2557	WP_019247158.1
1675	百日咳博德特氏菌	TPLPVSLTA	2558	WP_019249248.1
1676	百日咳博德特氏菌	TRQGIMDQY	2559	YP_006626873.1
1677	百日咳博德特氏菌	TSKQDERNY	2560	WP_019247158.1
1678	百日咳博德特氏菌	TSPYDGKYW	2561	YP_006628018.1
1679	百日咳博德特氏菌	TSRRSVASI	2562	AFK26302.1
1680	百日咳博德特氏菌	TSRTVTMRY	2563	NP_879898.1
1681	百日咳博德特氏菌	TTEYPNARY	2564	ADA85123.1
1682	百日咳博德特氏菌	TTEYSNARY	2565	AFK26302.1
1683	百日咳博德特氏菌	TTLGLEQTF	2566	WP_019247158.1
1684	百日咳博德特氏菌	TVLAAGAGL	2567	WP_019249248.1
1685	百日咳博德特氏菌	TVQELALKL	2568	AAA22984.1
1686	百日咳博德特氏菌	TVVQRNKHW	2569	WP_019247158.1
1687	百日咳博德特氏菌	VAAAADLAL	2570	WP_019249248.1
1688	百日咳博德特氏菌	VALASQARW	2571	AGS56996.1
1689	百日咳博德特氏菌	VAMKPYEVT	2572	AAA22983.1
1690	百日咳博德特氏菌	VARLVKLQG	2573	WP_019249248.1
1691	百日咳博德特氏菌	VAVAGGRWH	2574	AGS5699

[1249]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				6.1
1692	百日咳博德特氏菌	VEASAITTY	2575	WP_019248145.1
1693	百日咳博德特氏菌	VEDIGGKNY	2576	WP_019247158.1
1694	百日咳博德特氏菌	VEVSSPPPV	2577	WP_019247158.1
1695	百日咳博德特氏菌	VGAGGYEAG	2578	WP_019247158.1
1696	百日咳博德特氏菌	VGGGEHGRW	2579	WP_019249248.1
1697	百日咳博德特氏菌	VHVSKEEQY	2580	YP_006628018.1
1698	百日咳博德特氏菌	VIDGQKVLA	2581	AAA22974.1
1699	百日咳博德特氏菌	VIGACTSPY	2582	YP_006628018.1
1700	百日咳博德特氏菌	VKLGGVYEA	2583	AAA22974.1
1701	百日咳博德特氏菌	VLAPRLYLT	2584	AAA22974.1
1702	百日咳博德特氏菌	VLVKTNMVV	2585	AAA22983.1
1703	百日咳博德特氏菌	VPASGAPAA	2586	AGS56996.1
1704	百日咳博德特氏菌	VPFCFGKDL	2587	AAA22983.1
1705	百日咳博德特氏菌	VPVSEHCTV	2588	AAA22974.1
1706	百日咳博德特氏菌	VPVTPPKVE	2589	WP_019247158.1
1707	百日咳博德特氏菌	VRTVSAMEY	2590	WP_019249248.1
1708	百日咳博德特氏菌	VSGGSLSAP	2591	AGT50936.1
1709	百日咳博德特氏菌	VSSATKAKG	2592	WP_019248658.1
1710	百日咳博德特氏菌	VSSPPPVSV	2593	WP_019247158.1
1711	百日咳博德特氏菌	VSVKVAKKL	2594	WP_019249248.1
1712	百日咳博德特氏菌	VTMRYLASY	2595	WP_019248145.1
1713	百日咳博德特氏菌	VTSVAMKPY	2596	AAA22983.1
1714	百日咳博德特氏菌	VVAEKVTTP	2597	WP_019247158.1
1715	百日咳博德特氏菌	VVDGPPSRP	2598	WP_0192

[1250]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				47158.1
1716	百日咳博德特氏菌	VVETAQPLP	2599	WP_019247158.1
1717	百日咳博德特氏菌	WLTWLAILA	2600	AAW72734.1
1718	百日咳博德特氏菌	WTFHAGYRY	2601	AGS56996.1
1719	百日咳博德特氏菌	WVMTDNSNV	2602	AGS56996.1
1720	百日咳博德特氏菌	YAEHGEVSI	2603	WP_019249248.1
1721	百日咳博德特氏菌	YAIDGTAAG	2604	WP_019249248.1
1722	百日咳博德特氏菌	YALKSRIAL	2605	AAA22984.1
1723	百日咳博德特氏菌	YATNPQTQL	2606	WP_019248145.1
1724	百日咳博德特氏菌	YDTGDLIAY	2607	WP_019248658.1
1725	百日咳博德特氏菌	YEAGFSLGS	2608	WP_019247158.1
1726	百日咳博德特氏菌	YEDATFETY	2609	YP_006628018.1
1727	百日咳博德特氏菌	YENKSSTPL	2610	WP_019247158.1
1728	百日咳博德特氏菌	YEVTPTRML	2611	AAA22983.1
1729	百日咳博德特氏菌	YEYIWGLYP	2612	WP_019249248.1
1730	百日咳博德特氏菌	YEYIWGLYQ	2613	YP_006626470.1
1731	百日咳博德特氏菌	YEYSKGPKL	2614	AGS56996.1
1732	百日咳博德特氏菌	YFEPGPTTD	2615	WP_019248145.1
1733	百日咳博德特氏菌	YIWGLYPTY	2616	WP_019249248.1
1734	百日咳博德特氏菌	YIWGLYQTY	2617	YP_006626470.1
1735	百日咳博德特氏菌	YLRQITPGW	2618	1BCP_C
1736	百日咳博德特氏菌	YMIYMSGLA	2619	1BCP_C
1737	百日咳博德特氏菌	YPALRAALI	2620	WP_019248658.1
1738	百日咳博德特氏菌	YPGTPGDLL	2621	AAA22984.1
1739	百日咳博德特氏菌	YPTYTEWSV	2622	WP_019249248.1

[1251]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1740	百日咳博德特氏菌	YQTYTEWSV	2623	YP_006626470.1
1741	百日咳博德特氏菌	YSTGDLRAY	2624	WP_019248145.1
1742	百日咳博德特氏菌	YTLRYLASY	2625	WP_019248658.1
1743	百日咳博德特氏菌	YVLVKTNMV	2626	AAA22983.1
1744	百日咳博德特氏菌	YYDYEDATF	2627	YP_006628018.1
1745	百日咳博德特氏菌509	AAFIALYPNSQLAPT	2628	Q7VU05
1746	百日咳博德特氏菌509	GGAEYNLALGQRRA	2629	Q7VU04
1747	百日咳博德特氏菌509	GGAEYNLALGQRRADA	2630	Q7VU04
1748	百日咳博德特氏菌509	IALYPNSQLAPT	2631	Q7VU05
1749	百日咳博德特氏菌509	KPDQGEVVAVGPGKKTED	2632	P0A339.1
1750	百日咳博德特氏菌509	KPDQGEVVAVGPGKKTEDG	2633	P0A339.1
1751	百日咳博德特氏菌509	LAEVLDYHNFVLTQ	2634	Q7VWM1.1
1752	白喉杆菌	QSIALSSLMVAQAIP	2635	AAV70486.1
1753	白喉杆菌	SIGVLGYQKTVDHTKVNSKLSLF	2636	AAV70486.1
1754	百日咳博德特氏菌	AAHADWNNQSIVKT	2637	ABO77783.1
1755	百日咳博德特氏菌	AALGRG	2638	ABO77783.1
1756	百日咳博德特氏菌	AALGRGHSLYASYE	2639	ABO77783.1
1757	百日咳博德特氏菌	AAPLRRTTLAMALG	2640	ABO77783.1
1758	百日咳博德特氏菌	AAPLSITLQAGAHA	2641	ABO77783.1
1759	百日咳博德特氏菌	ADAQGDIVATELPS	2642	ABO77783.1
1760	百日咳博德特氏菌	ADSGFYLDATLRAS	2643	ABO77783.1
1761	百日咳博德特氏菌	AELA	2644	ABO77783.1
1762	百日咳博德特氏菌	AELAVFRAGGGAYR	2645	ABO77783.1
1763	百日咳博德特氏菌	AELQFRNGSVTSSG	2646	ABO7778

[1252]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1764	百日咳博德特氏菌	AGGRWHLGGLAGYT	2647	ABO77783.1
1765	百日咳博德特氏菌	AGVAAMQGAVVHLQ	2648	ABO77783.1
1766	百日咳博德特氏菌	AGYTRGDRGFTGDG	2649	ABO77783.1
1767	百日咳博德特氏菌	ALASQARWTGATRA	2650	ABO77783.1
1768	百日咳博德特氏菌	AMPWTFHAGYRYSW	2651	ABO77783.1
1769	百日咳博德特氏菌	AMQGAVVHLQRATIRRGDAP	2652	ABO77783.1
1770	百日咳博德特氏菌	ANGLRVRDE	2653	ABO77783.1
1771	百日咳博德特氏菌	ANGLRVRDEGGSSV	2654	ABO77783.1
1772	百日咳博德特氏菌	ANKDGKVDIGTYRY	2655	ABO77783.1
1773	百日咳博德特氏菌	APAAVSVLGASELT	2656	ABO77783.1
1774	百日咳博德特氏菌	APPAPKPAPQPGPQ	2657	ABO77783.1
1775	百日咳博德特氏菌	AQGILLENPAAELQ	2658	ABO77783.1
1776	百日咳博德特氏菌	ARWTGATRAVDSLS	2659	ABO77783.1
1777	百日咳博德特氏菌	ASLEAGRRFTHADG	2660	ABO77783.1
1778	百日咳博德特氏菌	ASYEYSKGPKLAMP	2661	ABO77783.1
1779	百日咳博德特氏菌	ATFTLANKD	2662	ABO77783.1
1780	百日咳博德特氏菌	ATFTLANKDGKVDI	2663	ABO77783.1
1781	百日咳博德特氏菌	ATRAVDSLSIDNAT	2664	ABO77783.1
1782	百日咳博德特氏菌	DDDGIALYVAGEQAQ	2665	ABO77783.1
1783	百日咳博德特氏菌	DDGIALYVAGEQAQ	2666	ABO77783.1
1784	百日咳博德特氏菌	DGGHITGGRAAGVA	2667	ABO77783.1
1785	百日咳博德特氏菌	DGIRRFLGTVTVKAGK	2668	ABO77783.1
1786	百日咳博德特氏菌	DGSVDFQQPAEAGR	2669	ABO77783.1
1787	百日咳博德特氏菌	DGYAVKGKYRTHGV	2670	ABO7778

[1253]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1788	百日咳博德特氏菌	DIVATELPSIPGTS	2671	ABO77783.1
1789	百日咳博德特氏菌	DKLVVMQDASGQHR	2672	ABO77783.1
1790	百日咳博德特氏菌	DLGLSDKLVVMQDA	2673	ABO77783.1
1791	百日咳博德特氏菌	DNATWVMTDNSNVGA	2674	ABO77783.1
1792	百日咳博德特氏菌	DNATWVMTDNSNVGALRLA	2675	ABO77783.1
1793	百日咳博德特氏菌	DNRAGRRFDQKVAG	2676	ABO77783.1
1794	百日咳博德特氏菌	EAGRFKVLTVNTLA	2677	ABO77783.1
1795	百日咳博德特氏菌	ELAQSIVEAPELGA	2678	ABO77783.1
1796	百日咳博德特氏菌	ELGAAIRVGRGARV	2679	ABO77783.1
1797	百日咳博德特氏菌	ELGADHAVAVAGGR	2680	ABO77783.1
1798	百日咳博德特氏菌	ELPSIPGTSIGPLD	2681	ABO77783.1
1799	百日咳博德特氏菌	EPVKLTLTGGADAQ	2682	ABO77783.1
1800	百日咳博德特氏菌	EQAQASIADSTLQG	2683	ABO77783.1
1801	百日咳博德特氏菌	ERGANVTVQRSAIV	2684	ABO77783.1
1802	百日咳博德特氏菌	ERQHGIHIQGSDPG	2685	ABO77783.1
1803	百日咳博德特氏菌	EVGKRIELAGGRQV	2686	ABO77783.1
1804	百日咳博德特氏菌	FDGAGTVHTNGIAH	2687	ABO77783.1
1805	百日咳博德特氏菌	FQQPAEAGRFKVLT	2688	ABO77783.1
1806	百日咳博德特氏菌	FRAGGGAYRAANGL	2689	ABO77783.1
1807	百日咳博德特氏菌	GAHAQGKALLYRVL	2690	ABO77783.1
1808	百日咳博德特氏菌	GARVTVSGGSLSAP	2691	ABO77783.1
1809	百日咳博德特氏菌	GAYRAANGLRVRDE	2692	ABO77783.1
1810	百日咳博德特氏菌	GDAPAGGAVPGGAV	2693	ABO77783.1
1811	百日咳博德特氏菌	GGAVPGGAVPGGFG	2694	ABO7778

[1254]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1812	百日咳博德特氏菌	GGAVPGGFGPVLDG	2695	ABO77783.1
1813	百日咳博德特氏菌	GGFGPVLDGWYGVD	2696	ABO77783.1
1814	百日咳博德特氏菌	GGLHIGALQSLQPE	2697	ABO77783.1
1815	百日咳博德特氏菌	GGVQIERGANVTVQ	2698	ABO77783.1
1816	百日咳博德特氏菌	GHSLYASYEYSKGP	2699	ABO77783.1
1817	百日咳博德特氏菌	GHTDSVHVGGYATY	2700	ABO77783.1
1818	百日咳博德特氏菌	GIAHRTELRGTRAE	2701	ABO77783.1
1819	百日咳博德特氏菌	GKALLYRVLPEPVK	2702	ABO77783.1
1820	百日咳博德特氏菌	GLGMAAALGRGHSL	2703	ABO77783.1
1821	百日咳博德特氏菌	GNVIETGGARRFAP	2704	ABO77783.1
1822	百日咳博德特氏菌	GPLDVALASQARWT	2705	ABO77783.1
1823	百日咳博德特氏菌	GQHRLWVRN	2706	ABO77783.1
1824	百日咳博德特氏菌	GRGFAQRQQLDNRA	2707	ABO77783.1
1825	百日咳博德特氏菌	GRLGLEVGKRIELA	2708	ABO77783.1
1826	百日咳博德特氏菌	GRQVQPYIKASVLQ	2709	ABO77783.1
1827	百日咳博德特氏菌	GRRFTHADGWFLEPQAELA	2710	ABO77783.1
1828	百日咳博德特氏菌	GSEPASANTLLLVQ	2711	ABO77783.1
1829	百日咳博德特氏菌	GSSVLGRLGLEVGK	2712	ABO77783.1
1830	百日咳博德特氏菌	GTTIKVSGRQAQGI	2713	ABO77783.1
1831	百日咳博德特氏菌	GTVTVKAGKLVADH	2714	ABO77783.1
1832	百日咳博德特氏菌	HAVAVAGGRWHLGG	2715	ABO77783.1
1833	百日咳博德特氏菌	IELAGGRQVQPYIK	2716	ABO77783.1
1834	百日咳博德特氏菌	IHIQGSDPGGVRTA	2717	ABO77783.1
1835	百日咳博德特氏菌	IRRFLGTVTVKAGK	2718	ABO7778

[1255]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1836	百日咳博德特氏菌	IRVGRGARVTVSGG	2719	ABO77783.1
1837	百日咳博德特氏菌	ITLQAGAHA	2720	ABO77783.1
1838	百日咳博德特氏菌	ITLQAGAHAQGKAL	2721	ABO77783.1
1839	百日咳博德特氏菌	IVEAPELGAAIRVG	2722	ABO77783.1
1840	百日咳博德特氏菌	IVKTGERQHGIHIQ	2723	ABO77783.1
1841	百日咳博德特氏菌	KAGKLVADHATLAN	2724	ABO77783.1
1842	百日咳博德特氏菌	KGKYRTHGVGASLE	2725	ABO77783.1
1843	百日咳博德特氏菌	KPAPQPGPQPPQPP	2726	ABO77783.1
1844	百日咳博德特氏菌	KVAGFELGADHAVA	2727	ABO77783.1
1845	百日咳博德特氏菌	KVAGSDGYAVKGKY	2728	ABO77783.1
1846	百日咳博德特氏菌	KVDIGTYRYRLAAN	2729	ABO77783.1
1847	百日咳博德特氏菌	KVLTVNTLAGSGLF	2730	ABO77783.1
1848	百日咳博德特氏菌	LAANGNGQWSLVGA	2731	ABO77783.1
1849	百日咳博德特氏菌	LAMPWTFHAGYRYS	2732	ABO77783.1
1850	百日咳博德特氏菌	LASTLWYAESNALS	2733	ABO77783.1
1851	百日咳博德特氏菌	LENDFKVAGSDGYA	2734	ABO77783.1
1852	百日咳博德特氏菌	LENPAAELQFRNGS	2735	ABO77783.1
1853	百日咳博德特氏菌	LGAAPAAHADWNNQ	2736	ABO77783.1
1854	百日咳博德特氏菌	LGGLAGYTRGDRGFTGDG	2737	ABO77783.1
1855	百日咳博德特氏菌	LLENP	2738	ABO77783.1
1856	百日咳博德特氏菌	LLVQTPLGSAATFT	2739	ABO77783.1
1857	百日咳博德特氏菌	LPPSRVVLRDTNVT	2740	ABO77783.1
1858	百日咳博德特氏菌	LQPEDLPPS	2741	ABO77783.1
1859	百日咳博德特氏菌	LQPEDLPPSRVVLR	2742	ABO7778

[1256]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1860	百日咳博德特氏菌	LRASRLENDFKVAG	2743	ABO77783.1
1861	百日咳博德特氏菌	LRLASDGSVDFQQP	2744	ABO77783.1
1862	百日咳博德特氏菌	LSAAANAAVNTGGV	2745	ABO77783.1
1863	百日咳博德特氏菌	LSAPHGNVIETGGA	2746	ABO77783.1
1864	百日咳博德特氏菌	LSDDGIRRFLGTVT	2747	ABO77783.1
1865	百日咳博德特氏菌	LVGAKAPPAPKPAP	2748	ABO77783.1
1866	百日咳博德特氏菌	LYVAGEQAQASIAD	2749	ABO77783.1
1867	百日咳博德特氏菌	MALGALGAAPAAHA	2750	ABO77783.1
1868	百日咳博德特氏菌	MNMSLSRIVKAAPL	2751	ABO77783.1
1869	百日咳博德特氏菌	MQDASGQHR	2752	ABO77783.1
1870	百日咳博德特氏菌	MQGAVVHLQRATIR	2753	ABO77783.1
1871	百日咳博德特氏菌	NAAVNTGGVGLAST	2754	ABO77783.1
1872	百日咳博德特氏菌	NALSKRLGELRLNP	2755	ABO77783.1
1873	百日咳博德特氏菌	NGQWSLVGAKAPPA	2756	ABO77783.1
1874	百日咳博德特氏菌	NTLAGSGLFRMNVF	2757	ABO77783.1
1875	百日咳博德特氏菌	PAGRELSAAANAAV	2758	ABO77783.1
1876	百日咳博德特氏菌	PAPQPPAGRELSAA	2759	ABO77783.1
1877	百日咳博德特氏菌	PGPQPPQPPQPQPE	2760	ABO77783.1
1878	百日咳博德特氏菌	PGTSIGPLDVALAS	2761	ABO77783.1
1879	百日咳博德特氏菌	PLGSAATFTLANKD	2762	ABO77783.1
1880	百日咳博德特氏菌	PQPEAPAPQPPAGR	2763	ABO77783.1
1881	百日咳博德特氏菌	PQPPQPQPEAPAPQ	2764	ABO77783.1
1882	百日咳博德特氏菌	PYIKASVLQEFDGA	2765	ABO77783.1
1883	百日咳博德特氏菌	RFAPQAAPLSITLQ	2766	ABO7778

[1257]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1884	百日咳博德特氏菌	RLGELRLNPDAGGA	2767	ABO77783.1
1885	百日咳博德特氏菌	RLNPDAGGAWGRGF	2768	ABO77783.1
1886	百日咳博德特氏菌	RNGSVTSSGQLSDD	2769	ABO77783.1
1887	百日咳博德特氏菌	RRFDQKVAGFELGA	2770	ABO77783.1
1888	百日咳博德特氏菌	RTTLAMALGALGAA	2771	ABO77783.1
1889	百日咳博德特氏菌	SAIVDGGLHIGALQ	2772	ABO77783.1
1890	百日咳博德特氏菌	SANTLLLVQTPLGS	2773	ABO77783.1
1891	百日咳博德特氏菌	SDPGGVRTASGTTI	2774	ABO77783.1
1892	百日咳博德特氏菌	SELTLDGGHITGGR	2775	ABO77783.1
1893	百日咳博德特氏菌	SGLFRMNVF	2776	ABO77783.1
1894	百日咳博德特氏菌	SGLFRMNVFADLGL	2777	ABO77783.1
1895	百日咳博德特氏菌	SGSSVELAQSIVEA	2778	ABO77783.1
1896	百日咳博德特氏菌	SIADSTLQGAGGVQ	2779	ABO77783.1
1897	百日咳博德特氏菌	SKGPKLAMPWTFHA	2780	ABO77783.1
1898	百日咳博德特氏菌	SNVGALRLASDGSV	2781	ABO77783.1
1899	百日咳博德特氏菌	SRIVKAAPLRRTTL	2782	ABO77783.1
1900	百日咳博德特氏菌	SVLGASELTLDGGH	2783	ABO77783.1
1901	百日咳博德特氏菌	SVLQEFDGA	2784	ABO77783.1
1902	百日咳博德特氏菌	SVLQEFDGAGTVHT	2785	ABO77783.1
1903	百日咳博德特氏菌	TELR	2786	ABO77783.1
1904	百日咳博德特氏菌	TELRGTRAELGLGM	2787	ABO77783.1
1905	百日咳博德特氏菌	TGDGGGHTDSVHVG	2788	ABO77783.1
1906	百日咳博德特氏菌	TGGARRFAPQAAPL	2789	ABO77783.1
1907	百日咳博德特氏菌	TGGRAAGVAAMQGA	2790	ABO7778

[1258]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1908	百日咳博德特氏菌	TGGVGLASTLWYAE	2791	ABO77783.1
1909	百日咳博德特氏菌	THGVGASLEAGRRF	2792	ABO77783.1
1910	百日咳博德特氏菌	TIRRGDAPA	2793	ABO77783.1
1911	百日咳博德特氏菌	TLANVGDTWDDDGI	2794	ABO77783.1
1912	百日咳博德特氏菌	TLQGAGGVQIERGA	2795	ABO77783.1
1913	百日咳博德特氏菌	TLTGGADAQGDIVA	2796	ABO77783.1
1914	百日咳博德特氏菌	TNVTAVPASGAPAA	2797	ABO77783.1
1915	百日咳博德特氏菌	TRAELGLGMAAALG	2798	ABO77783.1
1916	百日咳博德特氏菌	TSSGQLSDDGIRRF	2799	ABO77783.1
1917	百日咳博德特氏菌	TVHTNGIAHRTELR	2800	ABO77783.1
1918	百日咳博德特氏菌	TYRYRLAANGNGQW	2801	ABO77783.1
1919	百日咳博德特氏菌	VADHATLANVGDTW	2802	ABO77783.1
1920	百日咳博德特氏菌	VHVGGYATYIADSG	2803	ABO77783.1
1921	百日咳博德特氏菌	VLDGWYGVD	2804	ABO77783.1
1922	百日咳博德特氏菌	VPASGAPAAVSVLG	2805	ABO77783.1
1923	百日咳博德特氏菌	VRDEGGSSVLGRLG	2806	ABO77783.1
1924	百日咳博德特氏菌	VRTASGTTIKVSGR	2807	ABO77783.1
1925	百日咳博德特氏菌	VSGGSLSAPHGNVI	2808	ABO77783.1
1926	百日咳博德特氏菌	VSGRQAQGILLENP	2809	ABO77783.1
1927	百日咳博德特氏菌	VTVQRSAIVDGGLH	2810	ABO77783.1
1928	百日咳博德特氏菌	VVLRDTNVTAVPAS	2811	ABO77783.1
1929	百日咳博德特氏菌	WNNQSIVKTGERQH	2812	ABO77783.1
1930	百日咳博德特氏菌	WVRNSGSEPASANT	2813	ABO77783.1
1931	百日咳博德特氏菌	WYAESNALSKRLGE	2814	ABO7778

[1259]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
				3.1
1932	百日咳博德特氏菌	YATYIADSGFYLDA	2815	ABO77783.1
1933	百日咳博德特氏菌	YGVDVSGSS	2816	ABO77783.1
1934	百日咳博德特氏菌	YGVDVSGSSVELAQ	2817	ABO77783.1
1935	百日咳博德特氏菌	YLDATLRASRLEND	2818	ABO77783.1
1936	百日咳博德特氏菌	YRVLPEPVKLTLTG	2819	ABO77783.1
1937	百日咳博德特氏菌	VKAQNITNKRAALIEA	2820	AAA22974.1
1938	百日咳博德特氏菌	YYSNVTATRLLSSTNS	2821	AAA83981.1
1939	百日咳博德特氏菌	SPNLTDERAAQAGVT	2822	CPP72976.1
1940	麻疹 麻疹病毒属	SSRASDERAAHLPTS	2823	BAA33867.1
1941	白喉杆菌	QVVHNSYNRPAYSPG	2824	1007216A
1942	麻疹病毒 EDMONSTON-B 株	AEGGEIHEL	2825	AAF85692.1
1943	麻疹病毒 EDMONSTON-B 株	AENLISNGIGKY	2826	AAF85698.1
1944	麻疹病毒 EDMONSTON-B 株	AEVDGDVKL	2827	CAB43772.1
1945	麻疹病毒 EDMONSTON-B 株	AIYTAEIHK	2828	AAF85697.1
1946	麻疹病毒 EDMONSTON-B 株	APVFHMTNY	2829	CAB43772.1
1947	麻疹病毒 EDMONSTON-B 株	APVFHMTNYLEQPVSN	2830	AAR89413.1
1948	麻疹病毒 EDMONSTON-B 株	AQRLNEIY	2831	AAF85698.1
1949	麻疹病毒 EDMONSTON-B 株	ARVPHAYSL	2832	AAF85698.1
1950	麻疹病毒 EDMONSTON-B 株	AVRDLERAM	2833	P03424.1
1951	麻疹病毒 EDMONSTON-B 株	AVRDLERAMTTLK	2834	P03424.1
1952	麻疹病毒 EDMONSTON-B 株	DALLRLQAM	2835	Q89933.1
1953	麻疹病毒 EDMONSTON-B 株	DIKEKVINL	2836	AAF85698.1

[1260]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1954	麻疹病毒 EDMONSTON-B 株	DQGLFKVL	2837	AAF85695.1
1955	麻疹病毒 EDMONSTON-B 株	DTGVDTRIW	2838	Q9EMA9.1
1956	麻疹病毒 EDMONSTON-B 株	EPIGSLAIEEAM	2839	AAF85692.1
1957	麻疹病毒 EDMONSTON-B 株	EPIRDALNAM	2840	P69354.1
1958	麻疹病毒 EDMONSTON-B 株	FPKLGKTL	2841	AAF85692.1
1959	麻疹病毒 EDMONSTON-B 株	FRSVNAVAF	2842	AAF85695.1
1960	麻疹病毒 EDMONSTON-B 株	GKIIDNTEQL	2843	AAF85695.1
1961	麻疹病毒 EDMONSTON-B 株	GLNEKLVFY	2844	AAF85695.1
1962	麻疹病毒 EDMONSTON-B 株	GMYGGTYLVEK	2845	AAC35876.2
1963	麻疹病毒 EDMONSTON-B 株	GPPISLERLDVGTN	2846	P69354.1
1964	麻疹病毒 EDMONSTON-B 株	GPRQAQVSFL	2847	Q89933.1
1965	麻疹病毒 EDMONSTON-B 株	GRLVPQVRVID	2848	AAF85695.1
1966	麻疹病毒 EDMONSTON-B 株	GSAPISMGFR	2849	AAF85692.1
1967	麻疹病毒 EDMONSTON-B 株	HILAKSTAL	2850	AAF85698.1
1968	麻疹病毒 EDMONSTON-B 株	HYREVNLVY	2851	AAF85698.1
1969	麻疹病毒 EDMONSTON-B 株	IPPMKNLAL	2852	AAC35876.2
1970	麻疹病毒 EDMONSTON-B 株	IPYQGSGKGVSF	2853	CAB43772.1
1971	麻疹病毒 EDMONSTON-B 株	ISKESQHVY	2854	AAF85698.1
1972	麻疹病毒 EDMONSTON-B 株	IVSSHFFVY	2855	AAF85698.1
1973	麻疹病毒 EDMONSTON-B 株	KEIKETGRLF	2856	AAF85698.1
1974	麻疹病毒 EDMONSTON-B 株	KESQHVYYL	2857	AAF85698.1

[1261]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1975	麻疹病毒 EDMONSTON-B 株	KIIDNTEQL	2858	AAF85695.1
1976	麻疹病毒 EDMONSTON-B 株	KKQINRQN	2859	AAA63285.1
1977	麻疹病毒 EDMONSTON-B 株	KKVDTNFIYQ	2860	AAF85698.1
1978	麻疹病毒 EDMONSTON-B 株	KLIDGFFPA	2861	AAF85698.1
1979	麻疹病毒 EDMONSTON-B 株	KPNLSSKRSEL	2862	BAB39848.1
1980	麻疹病毒 EDMONSTON-B 株	KVSPYLFTV	2863	AAR89413.1
1981	麻疹病毒 EDMONSTON-B 株	LETRTTNQFL	2864	CAB43772.1
1982	麻疹病毒 EDMONSTON-B 株	LLKEATEL	2865	AAF85695.1
1983	麻疹病毒 EDMONSTON-B 株	LLKKGNSLY	2866	AAF85698.1
1984	麻疹病毒 EDMONSTON-B 株	LPAPIGGMNY	2867	AAF85698.1
1985	麻疹病毒 EDMONSTON-B 株	MPEETLHQVM	2868	AAF85698.1
1986	麻疹病毒 EDMONSTON-B 株	PTTIRGQFS	2869	CAB43772.1
1987	麻疹病毒 EDMONSTON-B 株	QEISRHQALGY	2870	P03424.1
1988	麻疹病毒 EDMONSTON-B 株	RITHVDTESY	2871	P69354.1
1989	麻疹病毒 EDMONSTON-B 株	RPGLKPDL	2872	P69354.1
1990	麻疹病毒 EDMONSTON-B 株	RPIYGLEV	2873	AAF85698.1
1991	麻疹病毒 EDMONSTON-B 株	RQAGQEMILAV	2874	P69354.1
1992	麻疹病毒 EDMONSTON-B 株	SAVRIATVY	2875	AAF85698.1
1993	麻疹病毒 EDMONSTON-B 株	SLMPEETLHQV	2876	AAF85698.1
1994	麻疹病毒 EDMONSTON-B 株	SMIDLVTKF	2877	AAF85698.1
1995	麻疹病毒 EDMONSTON-B 株	SMLNSQAIDNLRA	2878	P69354.1

[1262]

编号	疾病	表位氨基酸序列	SEQ ID NO	登录号
1996	麻疹病毒 EDMONSTON-B 株	SMYRVFEV	2879	CAB43772.1
1997	麻疹病毒 EDMONSTON-B 株	SQQGMFHAY	2880	AAF85698.1
1998	麻疹病毒 EDMONSTON-B 株	TDTPIVYNDRNLLD	2881	Q89933.1
1999	麻疹病毒 EDMONSTON-B 株	VIINDDQGLFKV	2882	AAF85695.1
2000	麻疹病毒 EDMONSTON-B 株	YESGVRIASL	2883	AAF85698.1
2001	麻疹病毒 EDMONSTON-B 株	YLKDKALA	2884	AAF85698.1
2002	麻疹病毒 EDMONSTON-B 株	YVYDHSGEAVK	2885	AAF85692.1
2003	风疹病毒	ARVIDPAAQSFTGVV	2886	BAA28178.1
2004	风疹病毒	SDRASARVIDPAAQS	2887	BAA28178.1
2005	风疹病毒	VPPGKFVTAALLNTP	2888	BAA28178.1
2006	风疹病毒	WVTPVIGSQARKCGL	2889	BAA28178.1
2007	流行性腮腺炎腮腺炎病毒属	GTYRLIPNARANLTA	400	AGC97176.1

[1263] E.多瓣引导编辑器的递送

[1264] 另一方面，本公开提供了使用不同策略在体外和体内递送多瓣引导编辑器，包括在单独的载体上使用断裂内含肽，使用诸如电穿孔的技术直接递送核糖核蛋白复合物(即，引导编辑器复合PEgRNA和/或第二位点gRNA)的策略，利用阳离子脂质介导的制剂，以及使用与核糖核蛋白复合物融合的受体配体诱导内吞作用的方法。本文考虑了任何此类方法。

[1265] 递送选择概述

[1266] 一些方面，本发明提供了包括向宿主细胞递送一种或多种编码多瓣引导编辑器多核苷酸的方法，例如本文所述一种或多种载体，所述载体编码本文所述多瓣引导编辑系统的一种或多种组件，其一种或多种转录本和/或由其转录的一种或多种蛋白。一些方面，本发明进一步提供了通过此类方法产生的细胞，以及包含此类细胞或由此类细胞产生的生物体(如，动物、植物或真菌)。在一些实施方案中，本文所述的多瓣引导编辑器与向导序列组合(并且任选地与向导序列复合)递送至细胞。可将常规的基于病毒和非病毒的基因转移方法用于在哺乳动物细胞或靶组织中引入核酸。此类方法可用于向培养细胞或宿主生物体中施用编码多瓣引导编辑器系统组分的核酸。非病毒载体递送系统包括DNA质粒、RNA(如，本文所述载体的转录本)、裸核酸和与递送载体(如，脂质体)复合的核酸。病毒载体递送系统包括DNA和RNA病毒，其在递送至细胞后具有游离基因组或整合基因组。关于基因治疗程序的综述，参见Anderson,Science 256:808-813(1992);Nabel&Felgner,TIBTECH 11:211-

217(1993);Mitani&Caskey,TIBTECH 11:162-166(1993);Dillon,TIBTECH 11:167-175(1993);Miller,Nature 357:455-460(1992);Van Brunt,Biotechnology 6(10):1149-1154(1988);Vigne,Restorative Neurology and Neuroscience 8:35-36(1995);Kremer&Perricaudet,British Medical Bulletin 51(1):31-44(1995);Haddada et al.,in Current Topics in Microbiology and Immunology Doerfler and Bihm(编)(1995);以及Yu et al.,Gene Therapy 1:13-26(1994)。

[1267] 核酸的非病毒递送方法包括脂质转染、核转染、显微注射、基因枪、病毒体、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸DNA、人工病毒体和试剂增强的DNA摄取。脂质转染描述于例如美国专利号5,049,386,4,946,787;和4,897,355中,脂质转染试剂可商购获得(如,Transfectam™和Lipofectin™)。适用于多核苷酸的有效受体识别脂质转染的阳离子和中性脂质包括Feigner,WO91/17424;WO91/16024的那些。递送可以是细胞(如,体外或离体施用)或靶组织(如,体内施用)。

[1268] 制备脂质:核酸复合物,包括靶向脂质体如免疫脂质复合物是本领域技术人员所熟知的(参见,例如,Crystal,Science 270:404-410(1995);Blaese et al.,Cancer Gene Ther.2:291-297(1995);Behr et al.,Bioconjugate Chem.5:382-389(1994);Remy et al.,Bioconjugate Chem.5:647-654(1994);Gao et al.,Gene Therapy 2:710-722(1995);Ahmad et al.,Cancer Res.52:4817-4820(1992);美国专利号4,186,183,4,217,344,4,235,871,4,261,975,4,485,054,4,501,728,4,774,085,4,837,028,和4,946,787)。

[1269] 使用基于RNA或DNA病毒的系统来递送核酸利用了高度进化的过程,用于将病毒靶向体内的特定细胞并将病毒有效载荷运输到细胞核。病毒载体可直接施用于患者(体内)或它们可用于体外处理细胞,并且任选地将经修饰的细胞施用于患者(离体)。常规基于病毒的系统可包括用于基因转移的逆转录病毒、慢病毒、腺病毒、腺相关病毒和单纯疱疹病毒载体。利用逆转录病毒、慢病毒和腺相关病毒基因转移方法可能在宿主基因组中整合,通常导致插入转基因的长期表达。此外,在许多不同细胞类型和靶组织中都观测到高转导效率。

[1270] 病毒的趋向性可通过并入外源包膜蛋白来改变,扩大靶细胞的潜在目标群体。慢病毒载体是能够转导或感染非分裂细胞并通常产生高病毒滴度的逆转录病毒载体。因此,逆转录病毒基因转移系统的选择取决于靶组织。逆转录病毒载体由顺式作用的长末端重复序列组成,其包装容量高达6-10kb的外源序列。最小的顺式作用LTR足以复制和包装载体,用于将治疗基因整合到靶细胞中以提供永久的转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)及其组合的那些(参见,例如Buchscher et al.,J.Virol.66:2731-2739(1992);Johann et al.,J.Virol.66:1635-1640(1992);Sommnerfelt et al.,Virol.176:58-59(1990);Wilson et al.,J.Virol.63:2374-2378(1989);Miller et al.,J.Virol.65:2220-2224(1991);PCT/US94/05700)。在优选瞬时表达的应用中,可使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有非常高的转导效率,并且不需要细胞分裂。使用这样的载体,已获得了高滴度和表达水平。该载体可在相对简单的系统中大量生产。腺相关病毒("AAV")载体也可用于用靶核酸转导细胞,例如在核酸和肽的体外生产中,以及用于体内和离体基因治疗程序(参见,例如West et al.,Virology 160:38-47(1987);美国专利号4,797,368;WO 93/24641;Kotin,Human Gene Therapy 5:793-801

(1994);Muzyczka,J.Clin.Invest.94:1351(1994)。重组AAV载体的构建描述于许多出版物中,包括美国专利号5,173,414;Tratschin et al.,Mol.Cell.Biol.5:3251-3260(1985);Tratschin,et al.,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);以及Samulski et al.,J.Virol.63:03822-3828(1989)。

[1271] 包装细胞通常用于形成能够感染宿主细胞的病毒颗粒。这些细胞包括包装腺病毒的293细胞和包装逆转录病毒的ψ2细胞或PA317细胞。用于基因治疗的病毒载体通常是通过产生将核酸载体包装成病毒颗粒的细胞系来产生的。载体通常包含包装和随后整合到宿主中所需的最少病毒序列,将其他病毒序列用于替换待表达的多核苷酸的表达盒。缺失的病毒功能通常由包装细胞系以反式提供。例如,用于基因治疗的AAV载体通常仅具有来自AAV基因组的ITR序列,这些序列是包装和整合到宿主基因组中所必需的。将病毒DNA包装在细胞系中,该细胞系包含编码其他AAV基因的辅助质粒,即rep和cap,但缺少ITR序列。细胞系也可用腺病毒作为辅助者进行感染。辅助病毒促进从辅助质粒中复制AAV载体和表达AAV基因。由于缺乏ITR序列,辅助质粒没有大量包装。腺病毒的污染可通过例如热处理来减少,腺病毒比AAV对热处理更敏感。将核酸递送至细胞的其他方法是本领域技术人员已知的。参见,例如US20030087817,通过引用并入本文。

[1272] 在不同实施方案中,可将双PE构建体(包括断裂构建体)工程化改造用于在一种或多种rAAV载体中递送。与本文提供的任何方法和组合物相关的rAAV可以是任何血清型,包括任何衍生物或假型(如,1、2、3、4、5、6、7、8、9、10、11、12、13、2/1、2/5、2/8、2/9、3/1、3/5、3/8或3/9)。rAAV可包含待递送至细胞的遗传负荷(即,表达感兴趣基因的重组核酸载体,如由rAAV携带至细胞的完整或断裂PE融合蛋白)。rAAV可以是嵌合的。

[1273] 如本文所用,rAAV的血清型是指重组病毒衣壳蛋白的血清型。衍生物和假型的非限制性实例包括rAAV2/1、rAAV2/5、rAAV2/8、rAAV2/9、AAV2-AAV3杂合体、AAVrh.10、AAVhu.14、AAV3a/3b、AAVrh32.33、AAV-HSC15、AAV-HSC17、AAVhu.37、AAVrh.8、CHt-P6、AAV2.5、AAV6.2、AAV2i8、AAV-HSC15/17、AAVM41、AAV9.45、AAV6(Y445F/Y731F)、AAV2.5T、AAV-HAE1/2、AAV克隆32/83、AAVShH10、AAV2(Y->F)、AAV8(Y733F)、AAV2.15、AAV2.4、AAVM41和AAVr3.45。具有嵌合VP1蛋白的衍生物和假型的非限制性示例是rAAV2/5-1VP1u,其具有AAV2的基因组、AAV5的衣壳骨架和AAV1的VP1u。具有嵌合VP1蛋白的衍生物和假型的其他非限制性示例是rAAV2/5-8VP1u、rAAV2/9-1VP1u和rAAV2/9-8VP1u。

[1274] AAV衍生物/假型,以及产生此类衍生物/假型的方法是本领域已知的(参见,例如,Mol Ther.2012Apr;20(4):699-708.doi:10.1038/mt.2011.287.Epub 2012Jan 24.The AAV vector toolkit:poised at the clinical crossroads.AsokanA1,SchafferDV,Samulski RJ.)。产生和使用假型rAAV载体的方法是本领域已知的(参见,例如,Duan et al.,J.Virol.,75:7662-7671,2001;Halbert et al.,J.Virol.,74:1524-1532,2000;Zolotukhin et al.,Methods,28:158-167,2002;以及Auricchio et al.,Hum.Molec.Genet.,10:3075-3081,2001)。

[1275] 制备或包装rAAV颗粒的方法是本领域已知的,并且试剂是可商购的(参见,例如,Zolotukhin et al.Production andpurification ofserotype 1,2,and 5recombinant adeno-associatedviral vectors.Methods 28(2002)158-167;以及美国专利公开号US20070015238和US20120322861,其通过引用并入本文;以及可从ATCC和CellBiolabs,

Inc.获得的质粒和试剂盒)。例如,包含感兴趣基因的质粒可与一个或多个辅助质粒组合,例如包含rep基因(如,编码Rep78、Rep68、Rep52和Rep40)和cap基因(编码VP1、VP2和VP3,包括如本文所述经修饰的VP2区),并转染到重组细胞中,使得rAAV颗粒可被包装并随后纯化。

[1276] 重组AAV可包含核酸载体,其可最低限度包括:(a)一个或多个异源核酸区域,其包含编码感兴趣的蛋白或多肽或者感兴趣的RNA(如,siRNA或microRNA),和(b)包含在一个或多个核酸区域(如,异源核酸区域)侧翼的反向末端重复(ITR)序列(如,野生型ITR序列或工程化ITR序列)的一个或多个区域。在此,包含编码感兴趣的蛋白或感兴趣的RNA的序列的异源核酸区域被称为感兴趣基因。

[1277] 本文提供的任何一种rAAV颗粒可具有衣壳蛋白,该衣壳蛋白在VP1u区之外具有不同血清型的氨基酸。在一些实施方案中,VP1蛋白骨架的血清型不同于ITR和/或Rep基因的血清型。在一些实施方案中,颗粒的VP1衣壳蛋白骨架的血清型与ITR的血清型相同。在一些实施方案中,颗粒的VP1衣壳蛋白骨架的血清型与Rep基因的血清型相同。在一些实施方案中,rAAV颗粒的衣壳蛋白包含导致转导效率提高的氨基酸突变。

[1278] 在一些实施方案中,核酸载体包含含有促进核酸(如,异源核酸)表达的序列(如,与核酸可操作地连接的表达控制序列)的一个或多个区域。许多这样的序列是本领域已知的。表达控制序列的非限制性示例包括启动子、隔离子、沉默子、响应元件、内含子、增强子、起始位点、终止信号和poly(A)尾。本文考虑了此类控制序列的任何组合(如,启动子和增强子)。

[1279] 最终的AAV构建体可并入编码PEgRNA的序列。在其他实施方案中,AAV构建体可并入编码第二位点切口向导RNA的序列。在其他实施方案中,AAV构建体可并入编码第二位点切口向导RNA的序列和编码PEgRNA的序列。

[1280] 在不同实施方案中,PEgRNA和第二位点切口向导RNA可由合适的启动子表达,例如人U6(hU6)启动子、小鼠U6(mU6)启动子或其他合适的启动子。PEgRNA和第二位点切口向导RNA可由相同的启动子或不同的启动子驱动。

[1281] 在一些实施方案中,rAAV构建体或本文的组合物经肠内向受试者施用。在一些实施方案中,rAAV构建体或本文的组合物经肠胃外向受试者施用。在一些实施方案中,将rAAV颗粒或本文的组合物经皮下、眼内、玻璃体内、视网膜下、静脉内(IV)、脑室内、肌内、鞘内(IT)、脑池内、腹膜内、经由吸入、局部或通过直接注射到一种或多种细胞、组织或器官。在一些实施方案中,rAAV颗粒或本文的组合物通过注射到肝动脉或门静脉中向受试者施用。

[1282] 基于断裂多瓣PE载体的策略

[1283] 在这方面,多瓣引导编辑器可在断裂位点分开并作为整个/完整引导编辑器的两个半部提供。可将这两个半部递送至细胞(如,作为表达的蛋白或在单独的表达载体上),并且一旦在细胞内接触,这两个半部通过在每个引导编辑器半部上的内含肽的自剪接作用形成完整的引导编辑器。可将断裂内含肽序列工程化改造到编码的引导编辑器的每一个半部中,以促进它们在细胞内的转剪接和完整的功能性PE的伴随恢复。

[1284] 这些基于断裂内含肽的方法克服了体内递送的几个障碍。例如,编码引导编辑器的DNA大于rAAV包装限制,因此需要特殊的解决方案。一种这样的解决方案是配制与断裂内含肽对融合的编辑器,将这些内含肽对包装到两个单独的rAAV颗粒,当它们共同递送到细胞时,重组功能性编辑器蛋白。描述了一些其他特殊考虑因素以解释多瓣引导编辑的独特

功能，包括优化第二位点切口靶标和将多瓣引导编辑器正确包装到病毒载体中，包括慢病毒和rAAV。

[1285] 在这方面，多瓣引导编辑器可在断裂位点分开并作为完整/全部引导编辑器的两个半部提供。可将这两个半部递送至细胞(如，作为表达的蛋白或在单独的表达载体上)，并且一旦在细胞内接触，这两个半部通过每个引导编辑器半部的内含肽的自剪接作用形成完整的引导编辑器。可将断裂内含肽序列工程化改造到编码的引导编辑器的每一个半部中，以促进它们在细胞内的转剪接和完整的功能性PE的伴随恢复。

[1286] 图66描绘了作为两个PE半部蛋白提供的引导编辑器的一个实施方案，其通过位于每个引导编辑器半部蛋白的末端或起点处的断裂内含肽的自剪接作用再生为完整的引导编辑器。本文所述的多瓣引导编辑器也可以以相同方式提供。如本文所用，术语“PEN-端半部”是指完整的引导编辑器的N-端半部，且其包含在完整的引导编辑器的PE N-端半部(即，N端外显肽)的C-末端处的“N内含肽”。“N内含肽”是指完整、完全形成的断裂内含肽部分的N端半部。如本文所用，术语“PE C-端半部”是指完整引导编辑器的C-端半部，且其包含在完整的引导编辑器的C-端半部(即，C-端外显肽)的N-末端处的“C内含肽”。当这两个半部蛋白，即，PEN-端半部和PE C-端半部相互接触时，例如在细胞内，N内含肽和C内含肽同时经历自切除过程并且在PE N-端半部的C-末端和PE C-端半部的N-末端之间形成肽键，以重新形成包含完整napDNAbp结构域(如，Cas9切口酶)和RT结构域的完整引导编辑器蛋白。尽管图中未显示，引导编辑器还可包含额外的序列，包括N端和/或C端处的NLS，以及连接各结构域的氨基酸接头序列。

[1287] 在不同实施方案中，可通过将整个多瓣引导编辑器在“断裂位点”处“断裂”来将引导编辑器工程化改造为两个半部蛋白(即，PE N-端半部和PE C-端半部)。“断裂位点”是指引导编辑器中两个相邻氨基酸残基之间插入断裂内含肽序列(即，N内含肽和C内含肽)的位置。更具体地，“断裂位点”是指将整个引导编辑器分成两个独立的半部的位置，其中每一个半部在断裂位点与N内含肽或C内含肽基序融合。断裂位点可位于引导编辑器融合蛋白中的任何合适的位置，但优选断裂位点位于允许形成大小适合(如，通过表达载体)递送的两个半部蛋白的位置，并且当一个半部蛋白与细胞内的另一个半部蛋白接触时，在断裂位点末端处与每一个半部蛋白融合的内含肽可用于彼此充分相互作用。

[1288] 在一些实施方案中，断裂位点位于napDNAbp结构域中。在其他实施方案中，断裂位点位于RT结构域中。在其他实施方案中，断裂位点位于连接napDNAbp结构域和RT结构域的接头中。

[1289] 在不同实施方案中，断裂位点设计需要寻找断裂和插入N端和C端内含肽的位置，这些位置在结构上都允许将两个半部引导编辑器结构域包装到两个不同的AAV基因组中。此外，反式剪接所需的内含肽残基可通过突变C端外显肽的N端处的残基或插入会留下内含肽“痕”的残基来掺入。

[1290] 示例性包含SpCas9切口酶或SaCas9切口酶的断裂多瓣(例如，双瓣或四瓣)引导编辑器的断裂配置如下。

[1291]

化脓性链球菌(S. PYOGENES)PE，在接头处断裂，N 端部分 结构：[N 外显肽]-[N 内含肽]
MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN** **TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK** **VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD** **KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN** **ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD** **LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI** **TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG** **ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA** **ILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW** **NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV** **TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE** **DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA** **HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF** **MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV** **KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV** **ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK** **VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL** **SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV** **SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD** **VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV** **WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW** **DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF** **LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF** **LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV** **LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT** **LIHQSITGLYETRIDLSQLGGD**<u>**SGGSSGGS**</u>CLSYETEILTVEYGLLPIGKIVEKRIECT VYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPID EIFERELDLMRVDNLPNSGGS<u>KRTADGSEFEPKKKRKV</u>(SEQ ID NO: 443) 解释： <u>NLS</u> (SEQ ID NO: 124, 125) **CAS9** (SEQ ID NO: 445) **<u>接头</u>**(SEQ ID NO: 446) NPUN 内含肽(SEQ ID NO: 447)

[1292]

化脓性链球菌(S. PYOGENES) PE，在接头处断裂，C 端部分 结构：[C 内含肽]-[C 外显肽]
MKRTADGSEFESPKKKRKV*IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN***<u>CFNSGS</u>** **<u>ETPGTSESATPESSGGSSGGSS</u>***TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGG* *MGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP* *VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRL* *HPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILL*

[1293]	*QYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWL TEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDP VAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQ ALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTW YTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYT DSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSP**SGGS**KRTADGSEFEPKKKRKV*(SEQ ID NO: 450) 解释: NLS (SEQ ID NO: 124, 125) **接头 1** (SEQ ID NO: 453) **接头 2** (SEQ ID NO: 174) *NPUC 内含肽*(SEQ ID NO: 452) *RT* (SEQ ID NO: 454)
	金黄色葡萄球菌(S. AUREUS)PE，在残基 740/741 之间断裂，N 端部分 结构：[N 外显肽]-[N 内含肽]
	MKRTADGSEFESPKKKRKV**GKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFK EANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEA RVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKAL EEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFID TYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYN ADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNE EDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDI QEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEG KCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTP FQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINR NLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAE**CLSYETEILTVE YGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATK DHKFMTVDGQMLPIDEIFERELDLMRVDNLPN**SGGS**KRTADGSEFEPKKKRKV*(SEQ ID NO: 458) 解释: NLS (SEQ ID NO: 124, 125) **CAS9** (SEQ ID NO: 460) **接头** (SEQ ID NO: 174) NPUN 内含肽(SEQ ID NO: 462)
	金黄色葡萄球菌(S. AUREUS)PE，在残基 740/741 之间断裂，C 端部分 结构：[C 内含肽]-[C 外显肽]
	MKRTADGSEFESPKKKRKV*IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN***CFN**EIE TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTL IVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHII KTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG**SGGSSGGSSGSETPGTSESA TPESSGGSSGGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDY

[1294]

*RPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAF EWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAA TSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQ ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPC LRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLL QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATA HIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMAD QAARKAAITETPDTSTLLIENSSP**SGGS**KRTADGSEFEPKKKRKV*(SEQ ID NO: 465)

解释:

NLS (SEQ ID NO: 124, 125)

CAS9 (SEQ ID NO: 467)

接头 1 (SEQ ID NO: 127)

接头 2 (SEQ ID NO: 174)

NPUC 内含肽(SEQ ID NO: 452)

RT (SEQ ID NO: 471)

[1295] 在不同实施方案中,使用SpCas9切口酶(SEQ ID NO:18,1368个氨基酸)作为示例,断裂可在1至1368之间的任何两个氨基酸之间。然而,优选的断裂将位于蛋白的中央区域,例如,从SEQ ID NO:18的氨基酸50至1250、或从100至1200、或从150至1150、或从200至1100、或从250至1050、或从300至1000、或从350至950、或从400至900、或从450至850、或从500至800、或从550至750、或从600至700。在具体的示例性实施方案中、断裂位点可位于740/741、或801/802、或1010/1011、或1041/1042。在其他实施方案中,断裂位点可相对于SEQ ID NO:18的SpCas9位于1/2、2/3、3/4、4/5、5/6、6/7、7/8、8/9、9/10、10/11、12/13、14/15、15/16、17/18、19/20、20/21、21/22、22/23、23/24、24/25、25/26、26/27、27/28、28/29、29/30、30/31、31/32、32/33、33/34、34/35、35/36、36/37、38/39、39/40、41/42、42/43、43/44、44/45、45/46、46/47、47/48、48/49、49/50、51/52、52/53、53/54、54/55、55/56、56/57、57/58、58/59、59/60、61/62、62/63、63/64、64/65、65/66、66/67、67/68、68/69、69/70、71/72、72/73、73/74、74/75、75/76、76/77、77/78、78/79、79/80、81/82、82/83、83/84、84/85、85/86、86/87、87/88、88/89、89/90、或90至100、100至150、150至200、200至250、250至300、300至350、350至400、450至500、500至550、550至600、600至650、650至700、700至750、750至800、800至850、850至900、900至950、950至1000、1000至1050、1050至1100、1100至1150、1150至1200、1200至1250、1250至1300、1300至1350、1300至1350、1350至1368之间的任何两对相邻残基之间,或与SEQ ID NO:18具有至少80%、85%、90%、95%、98%、99%或99.9%序列同一性的氨基酸序列的任意两个相应残基之间,或SEQ ID NO:19-88的SpCas9的任一氨基酸序列的变体或等同物或与SEQ ID NO:19-88的任一序列具有至少80%、85%、90%、95%、98%、99%、或99.9%序列同一性的氨基酸序列的任何两个相应残基之间。

[1296] 在不同实施方案中,断裂内含肽序列可由以下内含肽序列工程化改造。

[1297]

名称	配体依赖性内含肽的序列
2-4 内含肽:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 472)
3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDR VAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASM MGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEIL MIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS SRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRAL DKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL YSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDK FLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHN C (SEQ ID NO: 473)
30R3-1 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 474)
30R3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 475)

[1298]

30R3-3 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 476)
37R3-1 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC ((SEQ ID NO: 477)
37R3-2 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 478)
37R3-3 内含肽	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 479)

[1299] 在不同实施方案中，断裂内含肽序列可如下使用：

[1300]

内含肽-N	内含肽-C
NPU-N CLSYETEILTVEYGLLPIGKIVEKRIE CTVYSVDNNGNIYTQPVAQWHDRG EQEVFEYCLEDGSLIRATKDHKFMT VDGQMLPIDEIFERELDLMRVDNLP NSGGS (SEQ ID NO: 447)	NPU-C IKIATRKYLGKQNVYDIGVERDHNFALKNG FIASN (SEQ ID NO: 452)

[1301] 在不同实施方案中，断裂内含肽可用于将完整PE融合蛋白的单独部分分别递送至细胞，其在细胞中表达后，通过反式剪接重新组成完整PE融合蛋白。

[1302] 在一些实施方案中，本公开提供了将PE融合蛋白递送至细胞的方法，其包括：

[1303] (a)构建第一表达载体,其编码与第一断裂内含肽序列融合的PE融合蛋白的N-端片段;

[1304] (b)构建第二表达载体,其编码与第二断裂内含肽序列融合的PE融合蛋白的C端片段;

[1305] (c)将第一和第二表达载体递送至细胞,

[1306] 其中N-端和C-端片段由于反式剪接活性导致第一和第二断裂内含肽序列的自切除而在细胞中重新组成PE融合蛋白。

[1307] 在一些实施方案中,断裂位点可在引导编辑器融合中的任何位置,包括napDNAbp结构域、接头或逆转录酶结构域。

[1308] 在其他实施方案中,断裂位点在napDNAbp结构域中。

[1309] 在其他实施方案中,断裂位点在逆转录酶或聚合酶结构域中。

[1310] 在其他实施方案中,断裂位点在接头中。

[1311] 在不同实施方案中,本公开提供包含napDNAbp(如,Cas9结构域)和逆转录酶的多瓣引导编辑器,其中napDNAbp和/或逆转录酶之一或两者包含内含肽,如配体依赖性内含肽。通常,内含肽是配体依赖性内含肽,其在没有配体(如,小分子如4-羟基三苯氧胺、肽、蛋白、多核苷酸、氨基酸和核苷酸)的情况下不表现蛋白剪接活性或表现出最小的蛋白剪接活性。配体依赖性内含肽是已知的,包括描述于美国专利申请U.S.S.N.14/004,280,以US2014/0065711A1公布中的那些,其全部内容通过引用并入本文。此外,利用断裂Cas9结构。在一些实施方案中,内含肽包含选自SEQ ID NO:8-15、447、452、462和472-479的氨基酸序列。

[1312] 在不同实施方案中,与SEQ ID NO:18的经典SpCas9结构域相比,napDNAbp结构域是较小尺寸的napDNAbp结构域。

[1313] 经典SpCas9蛋白的长度为1368个氨基酸,并且具有158千道尔顿的预测分子量。如本文所用,术语“小尺寸Cas9变体”是指任何Cas9变体——天然存在的、工程化改造的或以其他方式——少于至少1300个氨基酸,或至少少于1290个氨基酸,或少于1280个氨基酸,或少于1270个氨基酸,或少于1260个氨基酸,或少于1250个氨基酸,或少于1240个氨基酸,或少于1230个氨基酸,或少于1220个氨基酸,或少于1210个氨基酸,或少于1200个氨基酸,或少于1190个氨基酸,或少于1180个氨基酸,或少于1170个氨基酸,或少于1160个氨基酸,或少于1150个氨基酸,或少于1140个氨基酸,或少于1130个氨基酸,或少于1120个氨基酸,或少于1110个氨基酸,或少于1100个氨基酸,或少于1050个氨基酸,或少于1000个氨基酸,或少于950个氨基酸,或少于900个氨基酸,或少于850个氨基酸,或少于800个氨基酸,或少于750个氨基酸,或少于700个氨基酸,或少于650个氨基酸,或少于600个氨基酸,或少于550个氨基酸,或少于500个氨基酸,但至少大于约400个氨基酸并保留Cas9蛋白所需的功能。

[1314] 在一个实施方案中,如实施例20所述,本说明书包括以下断裂内含肽PE构建体,其在经典SpCas9(SEQ ID NO:18)的残基1024和1025之间断裂(或相对于Met减SEQ ID NO:18,其可被分别称为残基1023和1024)。

[1315] 首先,SEQ ID NO:18的氨基酸序列如下所示,表明1024("K")和1025("S")残基之间断裂位点的位置:

[1316]

描述	序列	SEQ ID NO:
SpCas9 化脓性链球菌 *(Streptococcus pyogenes)* M1 SwissProt 登录号 Q99ZW2 野生型	**M**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD	SEQ ID NO: 18, 断裂位点 1024/1025 以粗体指示 在某些实施方案中，位点 1 的 M 不一定存在于 PE 融合蛋白中。因此，在氨基酸序列不包括位点 1 的 Met 的情况下，断裂位点的编号为 1023/1024。
	AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA **KS**EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	

[1317]

[1318] 在此配置中,N端半部的氨基酸序列(氨基酸1-1024)如下:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI
EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR
FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF
[1319] TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA**K** (SEQ ID
NO: 3877).

[1320] 在此配置中,N端半部(其中蛋白在位点1减Met)的氨基酸序列(氨基酸1-1023)如下:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
[1321] WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV

[1322] REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA**K** (SEQ ID NO: 3878).

[1323] 在此配置中,C端半部的氨基酸序列(氨基酸1024-1368(或在减Met Cas9中计数为氨基酸1023-1367)如下:

SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLI
[1324] IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 3879)

[1325] 如实施例20所示,PE2(基于SEQ ID NO:18的SpCas9)构建体在位点1023/1024(相

对于减Met SEQ ID NO:18)处断裂成两个独立的构建体,如下:

[1326] SpPE2在1023/1024N端半部处断裂

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVL
GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
[1327] KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQ
SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITL
KSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAK*CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRG*
*EQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN***SGGS***KRTA*
DGSEFEPKKKRKV (SEQ ID NO: 3875)

[1328] 解释:NLS,Cas9,突变的残基,**接头**,*NpuN内含肽*,NpuC内含肽,RT

[1329] SpPE2在1023/1024C端部分处断裂

MKRTADGSEFESPKKKRKV*IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN*CF
NEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP
KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
[1330] QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**SGGSS**
GGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSD
FPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLD
QGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLS
GLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQ

GFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALL
QTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQ
ALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPV
AAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN
[1331] **ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRP**
DLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQR
AELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNK
DEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST
LLIENSSPSGGS*KRTADGSEFEPKKKRK* (SEQ ID NO: 3876)

[1332] 解释:NLS,Cas9,突变的残基,**接头**,*NpuN内含肽*,NpuC内含肽,RT

[1333] 本公开还考虑将断裂内含肽多瓣引导编辑器递送至细胞和/或利用断裂内含肽多瓣引导编辑器处理细胞的方法。

[1334] 在一些实施方案中,本公开提供了将PE融合蛋白递送至细胞的方法,其包括:

[1335] (a)构建第一表达载体,其编码与第一断裂内含肽序列融合的PE融合蛋白的N-端片段;

[1336] (b)构建第二表达载体,其编码与第二断裂内含肽序列融合的PE融合蛋白的C端片段;

[1337] (c)将第一和第二表达载体递送至细胞,

[1338] 其中N-端和C-端片段由于反式剪接活性导致第一和第二断裂内含肽序列的自我切除而在细胞中重新组成PE融合蛋白。

[1339] 在某些实施方案中,融合至第一断裂内含肽序列的PE融合蛋白的N-端片段为SEQ ID NO:3875,或与SEQ ID NO:3875具有至少80%、至少85%、至少90%、至少95%、至少98%、或至少99.9%序列同一性的氨基酸序列。

[1340] 在其他实施方案中,融合至第一断裂内含肽序列的PE融合蛋白的C-端片段为SEQ ID NO:3876,或与SEQ ID NO:3876具有至少80%、至少85%、至少90%、至少95%、至少98%、或至少99.9%序列同一性的氨基酸序列。

[1341] 在其他实施方案中,本公开提供了在细胞内编辑靶DNA序列的方法,其包括:

[1342] (a)构建第一表达载体,其编码与第一断裂内含肽序列融合的PE融合蛋白的N-端片段;

[1343] (b)构建第二表达载体,其编码与第二断裂内含肽序列融合的PE融合蛋白的C-端片段;

[1344] (c)将第一和第二表达载体递送至细胞,

[1345] 其中N-端和C-端片段由于反式剪接活性导致第一和第二断裂内含肽序列的自我切除而在细胞中重新组成PE融合蛋白。

[1346] 在某些实施方案中,融合至第一断裂内含肽序列的PE融合蛋白的N-端片段为SEQ ID NO:3875,或与SEQ ID NO:3875具有至少80%、至少85%、至少90%、至少95%、至少98%、或至少99.9%序列同一性的氨基酸序列。

[1347] 在其他实施方案中,融合至第一断裂内含肽序列的PE融合蛋白的C-端片段为SEQ ID NO:3876,或与SEQ ID NO:3876具有至少80%、至少85%、至少90%、至少95%、至少98%、或至少99.9%序列同一性的氨基酸序列。

[1348] PE核糖核蛋白复合物的递送

[1349] 在这方面,多瓣引导编辑器可通过非病毒递送策略递送,包括通过各种方法递送与PEgRNA复合的多瓣引导编辑器(即,PE核糖核蛋白复合物),包括电穿孔和脂质纳米颗粒。核酸的非病毒递送方法包括脂质转染、核转染、显微注射、基因枪、病毒体、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸DNA、人工病毒体和试剂增强的DNA摄取。脂质转染描述于如美国专利号5,049,386、4,946,787和4,897,355中),脂质转染试剂是可商购获得的(如,Transfectam™和Lipofectin™)。适用于多核苷酸的有效受体识别脂质转染的阳离子和中性脂质包括Feigner,WO91/17424;WO91/16024中的那些。可递送至细胞(如,体外或离体施用)或靶组织(如,体内施用)。

[1350] 制备脂质:核酸复合物,包括靶向脂质体如免疫脂质复合物是本领域技术人员所熟知的(参见,例如,Crystal,Science 270:404-410(1995);Blaese et al.,Cancer Gene Ther.2:291-297(1995);Behr et al.,Bioconjugate Chem.5:382-389(1994);Remy et

al.,Bioconjugate Chem.5:647-654(1994);Gao et al.,Gene Therapy 2:710-722(1995);Ahmad et al.,Cancer Res.52:4817-4820(1992);美国专利号4,186,183,4,217,344,4,235,871,4,261,975,4,485,054,4,501,728,4,774,085,4,837,028和4,946,787)。

[1351] 可参考以下讨论非病毒递送核糖核蛋白复合物的方法的参考文献,其各自通过引用并入本文。

[1352] Chen,Sean,et al."Highly efficient mouse genome editing by CRISPR ribonucleoprotein electroporation of zygotes."Journal of Biological Chemistry (2016):jbc-M116.PubMed

[1353] Zuris,John A.,et al."Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo."Nature biotechnology 33.1(2015):73.PubMed

[1354] Rouet,Romain,et al."受体-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing."Journal of the American Chemical Society 140.21(2018):6596-6603.PubMed.

[1355] 图68C提供的数据显示可以这种方式递送各种公开的PE核糖核蛋白复合物(PE2在高浓度下、PE3在高浓度下和PE3在低浓度下)。

[1356] 通过mRNA递送PE

[1357] 可用于将多瓣引导编辑器和/或PEgRNA递送至需要基于多瓣引导编辑的基因组编辑的细胞的另一种方法是通过使用信使RNA(mRNA)递送方法和技术。可在本公开中利用的mRNA递送方法和组合物的示例包括,例如PCT/US2014/028330、US8822663B2、NZ700688A、ES2740248T3、EP2755693A4、EP2755986A4、WO2014152940A1、EP3450553B1、BR112016030852A2和EP3362461A1,其各自通过引用以其整体并入本文。通过引用在此并入的其他公开见于Kowalski et al.,"Delivering the Messenger:Advances in Technologies for Therapeutic mRNA Delivery,"Mol Therap.,2019;27(4):710-728。

[1358] 与编码多瓣引导编辑器的DNA载体相反,使用RNA作为多瓣引导编辑器的递送剂具有遗传物质不必进入细胞核来执行其功能的优点。递送的mRNA可在细胞质中直接翻译成期望的蛋白(如,引导编辑器融合蛋白)和核酸产物(如,PEgRNA)。然而,为了更稳定(如,抵抗细胞质中的RNA降解酶),在一些实施方案中使mRNA稳定以提高递送效率是必需的。某些递送载体(如,阳离子脂质或聚合物递送载体)还可帮助保护转染的mRNA免受内源性RNase酶的影响,否则内源性RNase酶可能会降解编码期望的引导编辑器融合蛋白的治疗性mRNA。此外,尽管经修饰的mRNA的稳定性增加,但将mRNA(特别是编码全长蛋白的mRNA)在体内以允许治疗水平的蛋白产生的方式递送至细胞仍然是挑战。

[1359] 除了一些例外,mRNA的细胞内递送通常比小的寡核苷酸的细胞内递送更具挑战性,并且它需要封装到递送纳米颗粒中,部分原因是与其他类型的RNA(小干扰RNA[siRNA],约14kDa;反义寡核苷酸[ASO],4-10kDa)相比,mRNA分子的尺寸明显更大(300-5,000kDa,约1-15kb)。

[1360] mRNA必须穿过细胞膜才能到达细胞质。细胞膜是细胞内递送的动态且强大的屏障。它主要由两性离子和带负电荷的磷脂的脂质双层组成,其中磷脂的极性头部指向水性环境,疏水尾部形成疏水核。

[1361] 在一些实施方案中，本公开的mRNA组合物包含mRNA(编码引导编辑器和/或PEgRNA)、转运载体和任选的促进与靶细胞接触和随后转染的试剂。

[1362] 在一些实施方案中，mRNA可包括一种或多种赋予mRNA稳定性的修饰(如，与mRNA的野生型或天然版本相比)，并且涉及相关的蛋白异常表达。还可包括对野生型进行校正缺陷的一种或多种修饰。例如，本发明的核酸可包括对5'非翻译区或3'非翻译区之一或两者的修饰。此类修饰可包括包含编码巨细胞病毒(CMV)立即早期1(IE1)基因、polyA尾、Cap1结构或人生长激素(hGH)的部分序列的序列。在一些实施方案中，mRNA经修饰以降低mRNA免疫原性。

[1363] 在一个实施方案中，本发明的组合物中的多瓣引导编辑器mRNA可在脂质体转移载体中配制以促进递送至靶细胞。考虑的转移载体可包括一种或多种阳离子脂质、非阳离子脂质和/或PEG修饰的脂质。例如，转移载体可包括以下阳离子脂质中的至少一种：C12-200、DLin-KC2-DMA、DODAP、HGT4003、ICE、HGT5000或HGT5001。在实施方案中，转移载体包括胆固醇(chol)和/或PEG修饰的脂质。在一些实施方案中，转移载体包括DMG-PEG2K。在某些实施方案中，转移载体具有以下脂质制剂：C12-200、DOPE、chol、DMG-PEG2K；DODAP、DOPE、胆固醇、DMG-PEG2K；HGT5000、DOPE、chol、DMG-PEG2K、HGT5001、DOPE、chol、DMG-PEG2K之一。

[1364] 本公开还提供了用于促进利用一种或多种编码PE的mRNA分子转染靶细胞的组合物和方法。例如，本发明的组合物和方法考虑使用可增加组合物对一种或多种靶细胞的亲和力的靶向配体。在一个实施方案中，靶向配体为载脂蛋白B或载脂蛋白E，对应的靶细胞表达低密度脂蛋白受体，从而促进靶向配体的识别。使用本公开的方法和组合物可优先靶向大量靶细胞。例如，考虑的靶细胞包括肝细胞、上皮细胞、造血细胞、上皮细胞、内皮细胞、肺细胞、骨细胞、干细胞、间充质细胞、神经细胞、心脏细胞、脂肪细胞、血管平滑肌包括细胞、心肌细胞、骨骼细胞肌细胞、β细胞、垂体细胞、滑膜衬里细胞、卵巢细胞、睾丸细胞、成纤维细胞、B细胞、T细胞、网织红细胞、白细胞、粒细胞和肿瘤细胞。然而，不限于这些。

[1365] 在一些实施方案中，编码PE的mRNA可任选地具有化学或生物修饰，例如，提高此类mRNA的稳定性和/或半衰期，或者提高或以其他方式促进蛋白生产。转染后，本发明组合物中的天然mRNA可以30分钟至数天的半衰期衰变。本公开的组合物中的mRNA可保留至少一些被翻译的能力，从而产生功能性蛋白或酶。因此，本发明提供了包含稳定化mRNA的组合物及其施用方法。在一些实施方案中，mRNA的活性在延长的时间段内延长。例如，可延长mRNA的活性，使得本公开的组合物每半周或每两周，或更优选地每月、每两月、每季度或以年为基础向受试者施用。本发明的mRNA扩展或延长的活性与由这种mRNA产生的蛋白或酶的量直接相关。类似地，本公开组合物的活性可通过进行改善或增强mRNA翻译的修饰而进一步扩展或延长。此外，靶细胞产生的功能性蛋白或酶的数量是递送至靶细胞的mRNA数量和此类mRNA的稳定性的函数。本发明的mRNA的稳定性在一定程度上可得到改善或增强，可进一步扩展半衰期、产生的蛋白或酶的活性以及组合物的给药频率。

[1366] 因此，在一些实施方案中，本公开组合物中的mRNA包含至少一种修饰，其赋予核酸增加或增强的稳定性，包括例如改善对体内核酸酶消化的抗性。如本文所用，术语“修饰”和“经修饰的”作为涉及本文提供的核酸的这类术语，包括优选增强稳定性并使mRNA比野生型或天然存在的mRNA版本更稳定的至少一种改变(如，对核酸酶消化的抗性)。如本文所用，术语“稳定的”和“稳定性”作为涉及本发明的核酸，特别是mRNA的这类术语，是指增加或增强

对如通常能够降解这类mRNA的核酸酶(即,核酸内切酶或核酸外切酶)的降解的抗性。增加的稳定性可包括,例如,对内源性酶(如,核酸内切酶或核酸外切酶)或者靶细胞或组织内的条件的水解或其他破坏的敏感性降低,从而增加或增强这类mRNA在靶细胞、组织、受试者和/或细胞质中的驻留。本文提供的稳定的mRNA分子表现出相对于它们天然存在的未经修饰的对应物(如,mRNA的野生型版本)更长的半衰期。术语“修饰”和“经修饰的”作为涉及本发明的mRNA的这类术语,还涵盖改善或增强mRNA核酸翻译的改变,包括例如包含在蛋白翻译起始中作用的序列(例如,Kozak共有序列)(Kozak,M.,NucleicAcids Res 15(20):8125-48(1987))。

[1367] 在一些实施方案中,用于本公开组合物中的mRNA经化学或生物修饰以使它们更稳定。对mRNA的示例性修饰包括碱基的耗尽(如,通过缺失或通过一个核苷酸取代另一个核苷酸)或碱基修饰,例如碱基的化学修饰。本文使用的短语“化学修饰”包括引入不同于天然存在的mRNA中看到的那些的化学性质的修饰,例如,共价修饰,如引入经修饰的核苷酸(如,核苷酸类似物,或包含在这类mRNA分子中并非天然存在的侧基)。

[1368] 可并入本公开组合物中使用的编码PE的mRNA中的其他合适的多核苷酸修饰包括但不限于4’-硫代修饰的碱基:4’-硫代-腺苷、4’-硫代-鸟苷、4’-硫代-胞苷、4’-硫代-尿苷、4’-硫代-5-甲基-胞苷、4’-硫代-假尿苷和4’-硫代-2-硫代尿苷、吡啶-4-核糖核苷、5-氮杂-尿苷、2-硫代-5-氮杂-尿苷、2-硫代尿苷、4-硫代-假尿苷、2-硫代-假尿苷、5-羟基尿苷、3-甲基尿苷、5-羧甲基-尿苷、1-羧甲基-伪尿苷、5-丙炔基-尿苷、1-丙炔基-伪尿苷、5-牛磺酸甲基尿苷、1-牛磺酸甲基-伪尿苷、5-牛磺酸甲基-2-硫代尿苷、1-牛磺酸甲基-4-硫代尿苷、5-甲基-尿苷、1-甲基-假尿苷、4-硫代-1-甲基-假尿苷、2-硫代-1-甲基-假尿苷、1-甲基-1-脱氮-假尿苷、2-硫代-1-甲基-1-脱氮-假尿苷、二氢尿苷、二氢假尿苷、2-硫代二氢尿苷、2-硫代二氢假尿苷、2-甲氧基尿苷、2-甲氧基-4-硫代尿苷、4-甲氧基-假尿苷、4-甲氧基-2-硫代-假尿苷、5-氮杂-胞苷、假异胞苷、3-甲基-胞苷、N4-乙酰胞苷、5-甲酰胞苷、N4-甲基胞苷、5-羟甲基胞苷、1-甲基-假异胞苷、吡咯并-胞苷、吡咯--假异胞苷、2-硫代-胞苷、2-硫代-5-甲基-胞苷、4-硫代-假异胞苷、4-硫代-1-甲基-假异胞苷、4-硫代-1-甲基-1-de氮杂-假异胞苷、1-甲基-1-去氮杂-假异胞苷、泽布拉林(zebularine)、5-氮杂-泽布拉林、5-甲基-泽布拉林、5-氮杂-2-硫代-泽布拉林、2-硫代-泽布拉林、2-甲氧基-胞苷、2-甲氧基-5-甲基-胞苷、4-甲氧基-假异胞苷、4-甲氧基-1-甲基-假异胞苷、2-氨基嘌呤、2,6-二氨基嘌呤、7-去氮杂-腺嘌呤、7-去氮杂-8-氮杂-腺嘌呤、7-去氮杂-2-氨基嘌呤、7-去氮杂-8-氮杂-2-氨基嘌呤、7-去氮杂-2,6-二氨基嘌呤、7-去氮杂-8-氮杂-2,6-二氨基嘌呤、1-甲基腺苷、N6-甲基腺苷、N6-异戊烯基腺苷、N6-(顺式-羟基异戊烯基)腺苷、2-甲基硫代-N6-(顺式-羟基异戊烯基)腺苷、N6-甘氨酰氨基甲酰腺苷、N6-苏氨酰氨基甲酰腺苷、2-甲硫基-N6-苏氨酰氨基甲酰腺苷、N6,N6-二甲基腺苷、7-甲基腺嘌呤、2-甲硫腺嘌呤、2-甲硫腺嘌呤、肌苷、1-甲基肌苷、怀俄苷(wyosine)、怀丁苷(wybutosine)、7-去氮杂-鸟苷、7-去氮杂-8-氮杂-鸟苷、6-硫代-鸟苷、6-硫代-7-去氮杂-鸟苷、6-硫代-7-去氮杂-8-氮杂-鸟苷、7-甲基-鸟苷、6-硫代-7-甲基-鸟苷、7-甲基肌苷、6-甲氧基-鸟苷、1-甲基鸟苷、N2-甲基鸟苷、N2,N2-甲基鸟苷、8-氧代鸟苷、7-甲基-8-氧代鸟苷、1-甲基-6-硫代鸟苷、N2-甲基-6-硫代鸟苷和N2,N2-二甲基-6-硫代鸟苷、以及其组合。术语修饰还包括,例如,将非核苷酸键或经修饰的核苷酸掺入本发明的mRNA序列中(如,对编码功能性蛋白或酶的mRNA分子的3’和5’端之

一或两者进行修饰)。此类修饰包括向mRNA序列添加碱基(如,包含polyA尾或更长的polyA尾)、3'UTR或5'UTR的改变、将mRNA与试剂(如、蛋白或互补核酸分子)复合,以及包含改变mRNA分子结构的元件(如,形成二级结构)。

[1369] 在一些实施方案中,编码PE的mRNA包括5'帽结构。通常按如下添加5'帽:首先,RNA端磷酸酶从5'核苷酸中去除一个末端磷酸基团,留下两个末端磷酸基团;然后通过鸟苷酸转移酶将三磷酸鸟苷(GTP)添加到末端磷酸,产生5'5'5三磷酸键;然后鸟嘌呤的7-氮被甲基转移酶甲基化。帽结构的示例包括但不限于m7G(5')ppp(5'(A,G(5')ppp(5')A和G(5')ppp(5')G。天然存在的帽结构包括7-甲基鸟苷,它通过三磷酸桥接到首个转录的核苷酸的5'-端,产生m7G(5')ppp(5')N的二核苷酸帽,其中N是任何核苷。在体内,酶促加帽。加帽在细胞核内,由酶(鸟苷酸转移酶)催化。在转录起始后立即加帽到RNA的5'末端。末端核苷是通常是鸟苷,并且与所有其他核苷酸的方向相反,即G(5')ppp(5')GpNpNp。

[1370] 其他帽类似物包括但不限于选自下组的化学结构:m7GpppG、m7GpppA、m7GpppC;未甲基化的帽类似物(如,GpppG);二甲基化帽类似物(如,m2,7GpppG)、三甲基化帽类似物(如,m2,2,7GpppG)、二甲基化对称帽类似物(如,m7Gpppm7G)、或抗反向帽类似物(如,ARCA;m7,2'OmeGpppG,m72'dGpppG、m7,3'OmeGpppG、m7,3'dGpppG和它们的四磷酸衍生物)(参见,例如,Jemielity,J.et al.,"Novel'anti-reverse'cap analogs with superiortranslationalproperties",RNA,9:1108-1122(2003))。

[1371] 通常,"尾"的存在用于保护mRNA免于核酸外切酶降解。认为PolyA或polyU尾可稳定天然信使和合成有义RNA。因此,在某些实施方案中,可向mRNA分子添加长polyA或polyU尾,从而使RNA更稳定。可使用本领域认可的技术添加PolyA或PolyU尾。例如,可使用PolyA聚合酶将长PolyA尾添加到合成或体外转录的RNA中(Yokoe,et al.Nature Biotechnology.1996;14:1252-1256)。转录载体也可编码长polyA尾。此外,可通过直接从PCR产物转录来添加polyA尾。PolyA也可用RNA连接酶连接到有义RNA的3'端(参见,例如,Molecular Cloning A Laboratory Manual,2nd Ed.,ed.by Sambrook,Fritsch and Maniatis(Cold Spring Harbor Laboratory Press:1991版))。

[1372] 通常,polyA或polyU尾的长度可为至少约10、50、100、200、300、400个、至少500个核苷酸。在一些实施方案中,mRNA3'端的polyA尾通常包括约10至300个腺苷核苷酸(如,约10至200个腺苷核苷酸、约10至150个腺苷核苷酸、约10至100个腺苷核苷酸、约20至70个腺苷核苷酸、或约20至60个腺苷核苷酸)。在一些实施方案中,mRNA包括3'poly(C)尾结构。mRNA的3'端上合适的poly-C尾通常包括约10至200个胞嘧啶核苷酸(如,约10至150个胞嘧啶核苷酸、约10至100个胞嘧啶核苷酸、约20至70个胞嘧啶核苷酸、约20至60个胞嘧啶核苷酸、或约10至40个胞嘧啶核苷酸)。poly-C尾可添加到poly-A或polyU尾或可替换poly-A或poly-U尾。

[1373] 根据本公开的编码PE的mRNA可根据多种已知方法中的任一种合成。例如,本发明的mRNA可通过体外转录(IVT)合成。简言之,IVT通常使用包含启动子的线性或环状DNA模板、三磷酸核糖核苷酸池、可包含DTT和镁离子的缓冲系统以及合适的RNA聚合酶(如,T3、T7或SP6RNA聚合酶)、DNAse I、焦磷酸酶和/或RNAse抑制剂进行。确切的条件将根据具体应用而不同。

[1374] 在涉及mRNA递送的实施方案中,编码PE融合蛋白的mRNA与PEgRNA的比率对于有效

编辑可能是重要的。在某些实施方案中，mRNA(编码PE融合蛋白)与PEgRNA的重量比为1:1。在某些其他实施方案中，mRNA(编码PE融合蛋白)与PEgRNA的重量比为2:1。在其他实施方案中，mRNA(编码PE融合蛋白)与PEgRNA的重量比为1:2。在更进一步的实施方案中，mRNA(编码PE融合蛋白)与PEgRNA的重量比选自：约1:1000、1:900、1:800、1:700、1:600、1:500、1:400、1:300、1:200、1:100、1:90、1:80、1:70、1:60、1:50、1:40、1:30、1:20、1:10、和1:1。在其他实施方案中，mRNA(编码PE融合蛋白)与PEgRNA的重量比选自：约1:1000、1:900、800:1、700:1、600:1、500:1、400:1、300:1、200:1、100:1、90:1、80:1、70:1、60:1、50:1、40:1、30:1、20:1、10:1、和1:1。

[1375] 前间隔序列F.多瓣引导编辑在插入可诱导二聚化结构域中的应用

[1376] 本文所述的多瓣引导编辑器(如图90、95、97A、98A所示)还可用于将二聚化结构域安装到一个或多个蛋白靶标中。二聚化结构域可通过以双特异性方式结合的连接部分(如，小分子、肽或蛋白)促进与一个或多个蛋白靶标的二聚化相关的活性的诱导调节。在不同方面，安装在不同蛋白(如，相同类型或不同蛋白)上时，二聚化结构域各自结合相同的双特异性部分(如，具有至少两个区域分别与二聚化结构域结合的双特异性小分子、肽或多肽)，从而通过蛋白与双特异性配体的共同相互作用引起蛋白的二聚化。以这种方式，双特异性配体充当两个蛋白二聚化的“诱导剂”。在一些情况下，双特异性配体或化合物将具有两个相同的靶向部分。在其他实施方案中，双特异性配体或化合物将具有彼此不同的靶向部分。具有相同的两个靶向部分的双特异性配体或化合物将能够靶向安装在不同蛋白靶标上的相同二聚化结构域。具有不同靶向部分的双特异性配体或化合物将能够靶向安装在不同蛋白靶标上的不同二聚化结构域。

[1377] 如本文所用，术语“二聚化结构域”是指与双特异性配体的结合部分结合的配体结合结构域。“第一”二聚化结构域结合双特异性配体的第一结合部分，“第二”二聚化结构域结合同一双特异性配体的第二结合部分。当第一二聚化结构域与第一蛋白融合(如，通过PE，如本文所讨论)并且第二二聚化结构域与第二蛋白融合(如，通过PE，如本文所讨论)时，第一和第二蛋白在双特异性配体存在下二聚化，其中所述双特异性配体具有与第一二聚化结构域结合的至少一个部分和与第二二聚化结构域结合的至少另一个部分。

[1378] 如本文所用，术语“双特异性配体”或“双特异性部分”是指结合两个不同配体结合结构域的配体。在不同实施方案中，双特异性部分本身是二聚体两个相同或两个不同的化学部分，其中每个部分与二聚化结构域特异性且紧密结合。在某些实施方案中，配体是小分子化合物、肽或多肽。在其他实施方案中，配体结合结构域是“二聚化结构域”，其可作为肽标签安装到蛋白上。在不同实施方案中，可通过每个二聚化结构域与双特异性配体的结合来诱导各自包含相同或不同二聚化结构域的两个蛋白二聚化。这些分子也可被称为“二聚化化学诱导物”或CID。此外，双特异性配体可通过将两个相同部分或两个不同部分结合(如，通过标准化的化学键)来制备，其中每个部分紧密且特异性地结合到二聚化结构域。

[1379] 在不同方面，多瓣PE安装的二聚化结构域可以相同或不同。

[1380] 例如，二聚化结构域可以是FKBP12，其具有以下氨基酸序列：

[1381]

FKBP12	MGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO: 488)

[1382] 在另一个示例中，二聚化结构域可为FKBP12的突变体，被称为FKBP12-F36V，FKBP12的突变体具有结合合成凸起的FK506模拟物的工程化孔(2，图3)[107]：

[1383]

FKBP12-F36V	MGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKVDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO: 489)

[1384] 在另一个示例中，二聚化结构域可为亲环蛋白，如下：

[1385]

亲环蛋白	MVNPTVFFDIAVDGEPLGRVSFELFADKVPKTAENFRALSTGEKGFGYKGSCFHRIIPGFMCQGGDFTRHNGTGGKSIYGEKFEDENFILKHTGPGILSMANAGPNTNGSQFFICTAKTEWLDGKHVVFGKVKEGMNIVEAMERFGSRNGKTSKKITIADCGQLE (SEQ ID NO: 490)

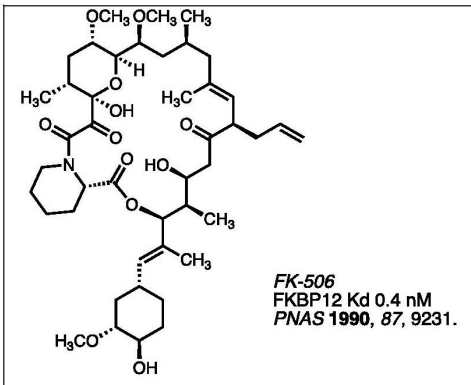
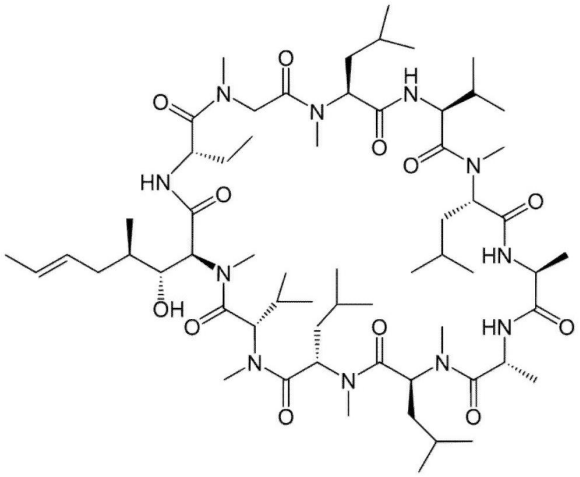
[1386] 在不同实施方案中，可改变这些二聚化结构域的氨基酸序列以优化结合或改善与天然靶标的结合正交性。可改变编码小分子结合蛋白的基因的核酸序列以优化PE加工的效率，例如通过减少PEgRNA二级结构。

[1387] 合适的二聚化结构域和与其结合的同源小分子化合物的其他示例提供如下。注意，同源小分子化合物可结合(如，通过化学接头)至第二小分子化合物(相同的化合物或不同的化合物)，以形成可结合两个二聚化结构域的双特异性配体。在一些情况下，例如FK506和环孢菌素A，各自的二聚化(如，FK506-FK506或环孢菌素A-环孢菌素A)降低或消除单体化合物的免疫抑制活性。

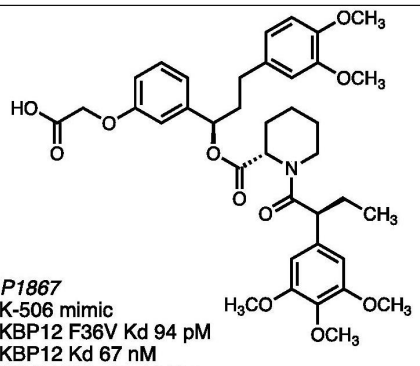
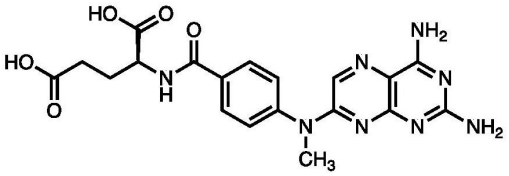
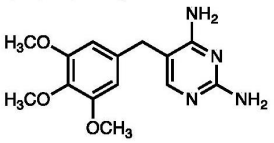
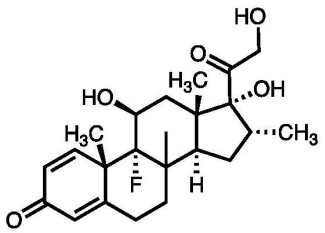
[1388]

小分子–结合至二聚化结构域–这些分子的二聚体构成结合两个二聚化结构域的双特异性配体	二聚化结构域
FK506	FKBP12 的氨基酸序列: MGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO: 491) 钙神经素:

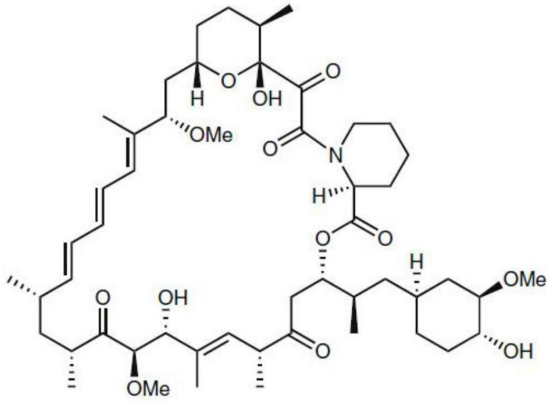
[1389]

 靶标: FKBP12 +钙神经素(CALCINEURIN)	
环孢菌素 A  靶标: 亲环蛋白 + 钙神经素	人亲环蛋白 A 的氨基酸序列: MVNPTVFFDIAVDGEPLGRVSFELF ADKVPKTAENFRALSTGEKGFGYK GSCFHRIIPGFMCQGGDFTRHNGTG GKSIYGEKFEDENFILKHTGPGILSM ANAGPNTNGSQFFICTAKTEWLDG KHVVFGKVKEGMNIVEAMERFGS RNGKTSKKITIADCGQLE (SEQ ID NO: 490) 人亲环蛋白 B 的氨基酸序列: MLRLSERNMKVLLAAALIAGSVFF LLLPGPSAADEKKKGPKVTVKVYF DLRIGDEDVGRVIFGLFGKTVPKTV DNFVALATGEKGFGYKNSKFHRVIK DFMIQGGDFTRGDGTGGKSIYGER FPDENFKLKHYGPGWVSMANAGK DTNGSQFFITTVKTAWLDGKHVVF GKVLEGMEVVRKVESTKTDSRDKP LKDVIIADCGKIEVEKPFAIAKE (SEQ ID NO: 493) 鼠亲环蛋白 C 的氨基酸序列: MSPGPRLLLPAVLCLGLGALVSSSG SSGVRKRGPSVTDKVFFDVRIGDK DVGRIVIGLFGNVVPKTVENFVALA TGEKGYGYKGSIFHRVIKDFMIQGG DFTARDGTGGMSIYGETFPDENFKL KHYGIGWVSMANAGPDTNGSQFFI TLTKPTWLDGKHVVFGKVLDGMT VVHSIELQATDGHDRPLTDCTIVNS GKIDVKTPFVVEVPDW (SEQ ID NO: 494)
AP1867	FKBP12-F36V 的氨基酸序列 MGVQVETISPGDGRTFPKRGQTCV VHYTGMLEDGKKVDSSRDRNKPF KFMLGKQEVIRGWEEGVAQMSVG QRAKLT ISPDYAYGATGHPGIIPPHATLVFDV ELLKLE (SEQ ID NO: 495)

[1390]

 *AP1867* FK-506 mimic FKBP12 F36V Kd 94 pM FKBP12 Kd 67 nM *PNAS* **1998**, *95*, 10437. 靶标: FKBP12	
甲氨蝶呤(METHOTREXATE)  *methotrexate* Human DHFR Kd < 10 nM *J. Biol. Chem.* **1988**, *263*, 10304. *E. coli* DHFR Kd 9.5 nM *PNAS* **2002**, *99*, 13481. 靶标: 二氢叶酸还原酶	人二氢叶酸还原酶的氨基酸序列 MVGSLNCIVAVSQNMGIGKNGDLP WPPLRNEFRYFQRMTTTSSVEGKQ NLVIMGKKTWFSIPEKNRPLKGRIN LVLSRELKEPPQGAHFLSRSLDDAL KLTEQPELANKVDMVWIVGGSSVY KEAMNHPGHLKLFVTRIMQDFESD TFFPEIDLEKYKLLPEYPGVLSDVQ EEKGIKYKFEVYEKND (SEQ ID NO: 496)
甲氧苄啶(TRIMETHOPRIM)  *trimethoprim* *E. coli* DHFR K_I 1.3 nM *Biochemistry* **1982**, *21*, 5068. 靶标: 二氢叶酸还原酶	大肠杆菌二氢叶酸还原酶的氨基酸序列 MISLIAALAVDRVIGMENAMPWNL PADLAWFKRNTLNKPVIMGRHTWE SIGRPLPGRKNIILSSQPGTDDRVTW VKSVDEAIAACGDVPEIMVIGGGR VYEQFLPKAQKLYLTHIDAEVEGDT HFPDYEPDDWESVFSEFHDADAQN SHSYCFEILERR (SEQ ID NO: 497)
地塞米松(DEXAMETHOSONE)  *dexamethasone* Human GR Kd 4.6 nM *Mol. Endocrin.* **1999**, *13*, 1855.	人糖皮质激素受体的氨基酸序列 MDSKESLTPGREENPSSVLAQERGD VMDFYKTLRGGATVKVSASSPSLA VASQSDSKQRRLLVDFPKGSVSNA QQPDLSKAVSLSMGLYMGETETKV MGNDLGFPQQGQISLSSGETDLKLL EESIANLNRSTSVPENPKSSASTAVS AAPTEKEF PKTHSDVSSEQQHLKGQTGTNGGN VKLYTTDQSTFDILQDLEFSSGSPG KETNESPWRSDLLIDENCLLSPLAG EDDSFLLEGNSNEDCKPLILPDTKP KIKDNGDLVLSSPSNVTLPQVKTEK EDFIELCTPGVIKQEKLGTVYCQAS FPGANIIGNKMSAISVHGVSTSGGQ

[1391]

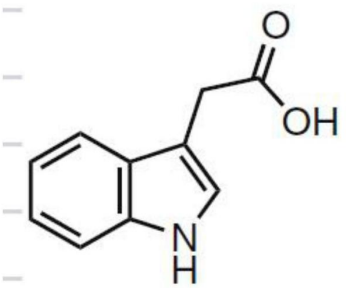
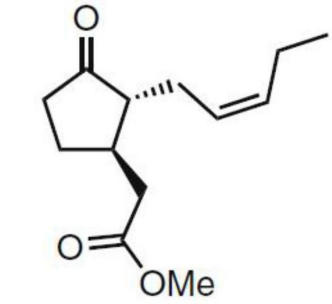
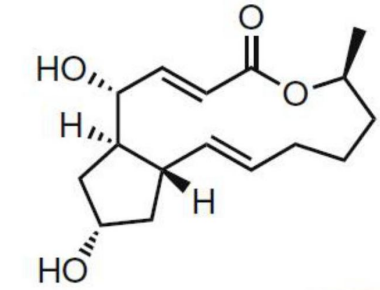
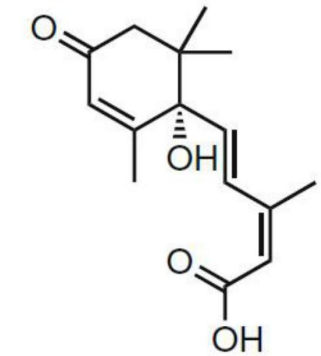
	MYHYDMNTASLSQQQDQKPIFNVI PPIPVGSENWNRCQGSGDDNLTSLG TLNFPGRTVFSNGYSSPSMRPDVSS PPSSSSTATTGPPPKLCLVCSDEASG CHYGVLTCGSCKVFFKRAVEGQHN YLCAGRNDCIIDKIRRKNCPACRYR KCLQAGMNLEARKTKKKIKGIQQA TTGVSQETSENPGNKTIVPATLPQLT PTLVSLLEVIEPEVLYAGYDSSVPDS TWRIMTTLNMLGGRQVIAAVKWA KAIPGFRNLHLDDQMTLLQYSWMF LMAFALGWRSYRQSSANLLCFAPD LIINEQRMTLPCMYDQCKHMLYVS SELHRLQVSYEEYLCMKTLLLLSSV PKDGLKSQELFDEIRMTYIKELGKA IVKREGNSSQNWQRFYQLTKLLDS MHEVVENLLNYCFQTFLDKTMSIE FPEMLAEIITNQIPKYSNGNIKKLLF HQK (SEQ ID NO: 498)
 雷帕霉素(RAPAMYCIN)	FKBP12 的氨基酸序列: MGVQVETISPGDGRTFPKRGQTCV VHYTGMLEDGKKFDSSRDRNKPFK FMLGKQEVIRGWEEGVAQMSVGQ RAKLTISPDYAYGATGHPGIIPPHAT LVFDVELLKLE (SEQ ID NO: 491)

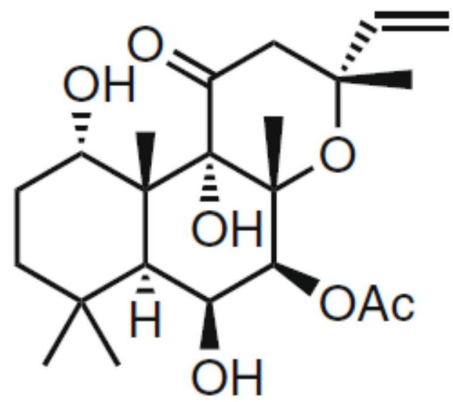
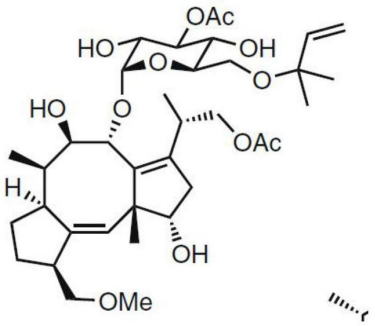
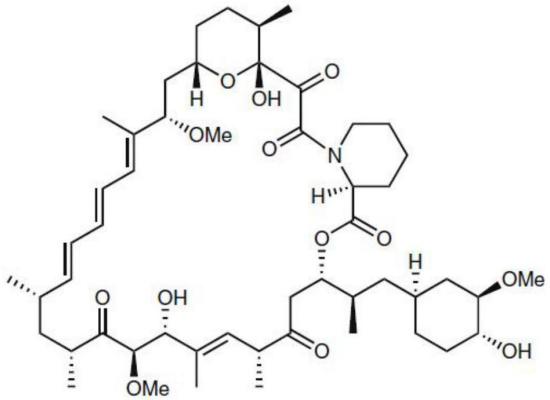
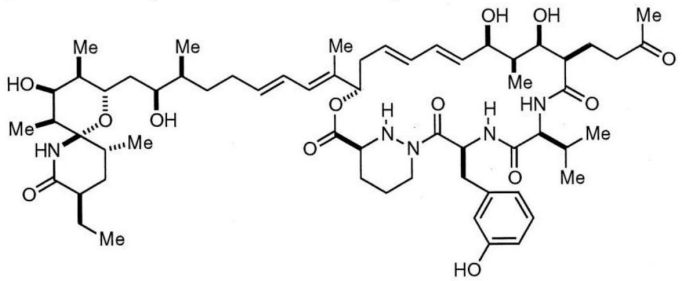
[1392] 其他天然存在的双功能分子及其双靶标受体的示例如下。引导编辑可用于将双靶受体安装到不同的蛋白中。一旦不同的蛋白由PE修饰成含有双功能分子受体，该双功能分子就可能被引入，从而导致蛋白二聚化修饰为包含不同的二聚化结构域。(1)生物功能分子和(2)它们的双靶受体的配对示例：

[1393]

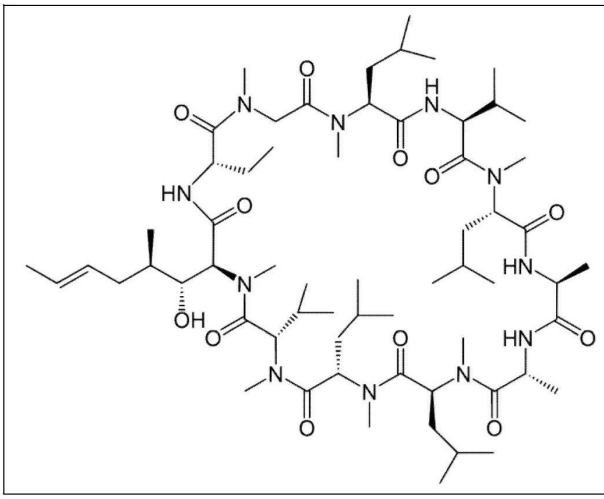
天然存在的生物功能分子	生物功能分子的靶受体

[1394]

 生长素(auxin)	靶受体 1: 生长素受体 靶受体 2: TIR1 E3 连接酶
 茉莉酮酸甲酯(methyl jasmonate)	靶受体 1: JAZ 受体 靶受体 2: Col1 E3 连接酶
布雷菲德菌素 A (brefeldin A) 	靶受体 1: GBF1 靶受体 2: GTPase Arf1p
脱落酸(abscisic acid) 	靶受体 1: PYR 受体 靶受体 2: 磷蛋白磷酸酶 2C
毛喉素(Forskolin)	靶受体 1: 腺苷酸环化酶单体 靶受体 2: 腺苷酸环化酶单体

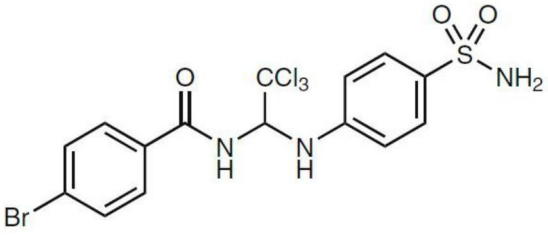
		
	壳梭孢菌素 A(fusicoccin A) 	靶受体 1: 14-3-3 蛋白 靶受体 2: H+-ATPase
[1395]	雷帕霉素(RAPAMYCIN) 	靶受体 1: FKBP12 靶受体 2: MTOR
	萨菲菌素 A(SANGLIFEHRIN A)  Sanglifehrin A (SFA)	靶受体 1: 亲环蛋白 靶受体 2: IMP 脱氢酶 2
	环孢菌素 A(cyclosporin A)	靶受体 1: 亲环蛋白 靶受体 2: 钙神经素

[1396]



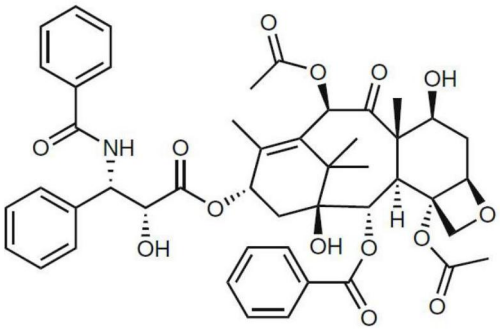
[1397] 可与这方面引导编辑使用的其他双功能分子的示例如下:Synstab A:

[1398]



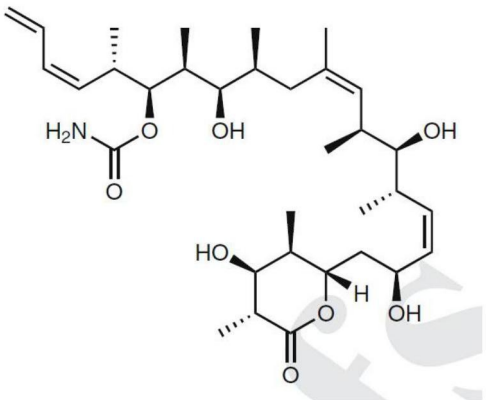
[1399] 紫杉醇:

[1400]

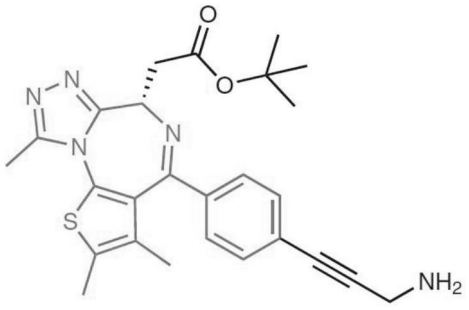


[1401] Discodermolide:

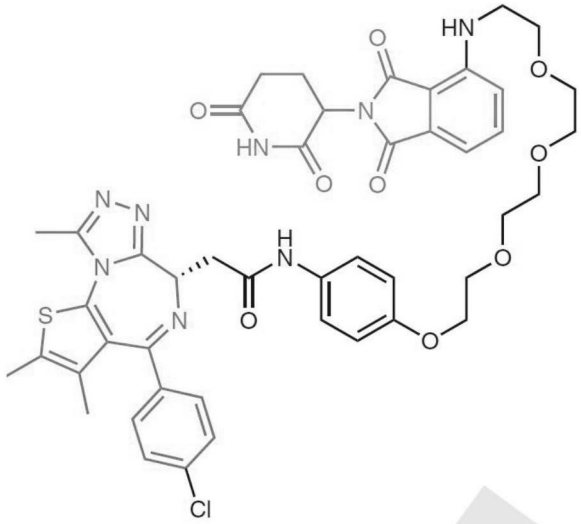
[1402]



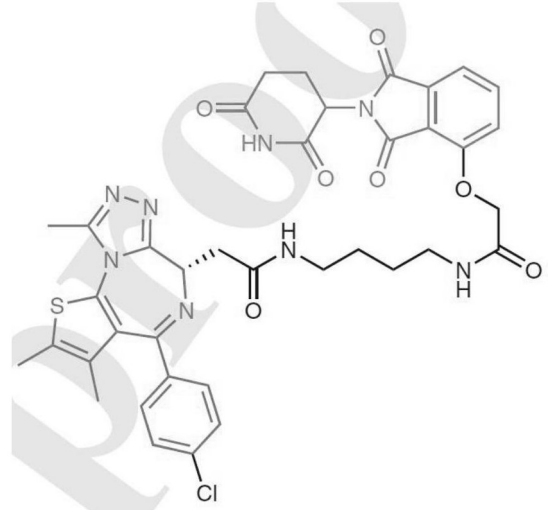
[1403] GNE-0011

[1404] 

[1405] ARV-825,以及

[1406] 

[1407] dBET1

[1408] 

[1409] Synstab A、紫杉醇和discodermolide是微管稳定剂。因此,这些化合物可用于使经PE修饰的蛋白二聚化以包含微管蛋白。GNE-0011、ARV-825和dBET1包含BRD4结合基序和CRBN结合基序。因此,这些化合物可用于使经PE修饰的蛋白二聚化以包含这些靶向结构域。

[1410] 用于安装二聚化结构域的PEgRNA可包含以下结构(参考图3D):

[1411] 5′-[间隔区]-[gRNA核心]-[延伸臂]-3′,其中延伸臂包含5′-[同源臂]-[编辑模板]-[引物结合位点]-3′;或

[1412] 5′-[延伸臂]-[间隔区]-[gRNA核心]-3′,其中延伸臂包含5′-[同源臂]-[编辑模

板]-[引物结合位点]-3′，并且其中任一构造“编辑模板”包含二聚化结构域的核苷酸序列。

[1413] 在一个示例中，用于在人胰岛素受体的C端插入FKBP12二聚化结构域的PEgRNA(间隔区加下划线，gRNA核心为正常文本，瓣同源性为粗体，FKBP12插入为斜体，退火区为粗斜体)：

[1414]

用于在人胰岛素受体中安装FKBP12的PEGRNA	CACGGUAGGCACUGUUAGGAGUUUUAG AGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGC**UUGCCUCGGU CCAAUCCUUCC***GGAGUGCAGGUGGAAA CCAUCUCCCCAGGAGACGGGCGCACCUU CCCCAAGCGCGGCCAGACCUGCGUGGUG CACUACACCGGGAUGCUUGAAGAUGGAA AGAAAUUUGAUUCCUCCCGGGACAGAAA CAAGCCCUUUAAGUUUAUGCUAGGCAAG CAGGAGGUGAUCCGAGGCUGGGAAGAAG GGGUUGCCCAGAUGAGUGUGGGUCAGA GAGCCAAACUGACUAUAUCUCCAGAUUA UGCCUAUGGUGCCACUGGGCACCCAGGC AUCAUCCCACCACAUGCCACUCUCGUCU UCGAUGUGGAGCUUCUAAAACUGGAA**UA ACAGUGCCUACC*** (SEQ ID NO: 499)

[1415] 在另一个示例中，用于在HEK3基因座插入FKBP12二聚化结构域的PEgRNA(用于优化)：

[1416]

用于在HEK3中安装FKBP12的PEGRNA	GGCCCAGACTGAGCACGTGAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGCTGGAGGAAGCAGGG CTTCCTTTCCTCTGCCATCATTCCAGTTT TAGAAGCTCCACATCGAAGACGAGAGT GGCATGTGGTGGGATGATGCCTGGGTGC CCAGTGGCACCATAGGCATAATCTGGAG ATATAGTCAGTTTGGCTCTCTGACCCAC ACTCATCTGGGCAACCCCTTCTTCCCAG CCTCGGATCACCTCCTGCTTGCCTAGCA TAAACTTAAAGGGCTTGTTTCTGTCCCG GGAGGAATCAAATTTCTTTCCATCTTCA AGCATCCCGGTGTAGTGCACCACGCAGG TCTGGCCGCGCTTGGGGAAGGTGCGCCC

[1417]

	GTCTCCTGGGGAGATGGTTTCCACCTGC ACTCCCGTGCTCAGTCTG (SEQ ID NO: 500)

[1418] 用于安装二聚化结构域的靶蛋白没有特别限制；然而，在双特异性配体存在下，它们的二聚化(一旦经PE修饰)产生一些有利的生物学效应是有利的，例如，信号通路、降低的免疫反应性等。在不同方面，待通过PE依赖性安装二聚化结构域进行二聚化的靶蛋白可以是相同的蛋白或不同的蛋白。优选地，蛋白在二聚化时触发一个或多个下游生物级联，例如信号传导级联、磷酸化等。可将PE用于安装二聚化结构域的示例性靶蛋白包括但不限于：

[1419]

膜结合受体	融合至的激酶结构域	使用的 CID	信号级联反应	参考文献
T 细胞受体	FKBP12	FK1012 (FK506 二聚物)	T-细胞受体信号传导	*SCIENCE* 262, 1019–24 (1993) *CHEM. BIOL.* 1, 163–172 (1994).
FAS 受体	MURINE CYCLOPHILIN C	环孢菌素 A 二聚物	细胞凋亡的 FAS 通路	*CHEM. BIOL.* 3, 731–738 (1996).
胰岛素受体	FKBP12	FK1012 (FK506 二聚物)	胰岛素信号传导	*CURR. BIOL.* 8, 11–18 (1998).
血小板衍生的生长因子 (PDGF) BETA	FKBP12	FK1012 (FK506 二聚物)	PDGF 中胚层形成信号传导	*CURR. BIOL.* 8, 11–18 (1998).
红细胞生成素受体(EPOR)	FKBP12	FK1012 (FK506 二聚物)	EPOR-介导的增殖信号传导	*PROC. NATL. ACAD. SCI.* 94, 3076–3081 (2002).

[1420] 一方面，本文所述的多瓣引导编辑器可用于将编码二聚化结构域的序列安装到活细胞或患者的一个或多个编码感兴趣的靶蛋白的基因中。这可被称为“多瓣引导编辑-CID系统”，其中CID是诱导靶蛋白二聚化的双特异性配体，各自与由多瓣PE安装的二聚化结构域融合。单独的这种编辑应当没有生理影响。施用双特异性配体后，双特异性配体通常是可同时结合两个二聚化结构域的二聚体小分子，每个二聚化结构域都与靶蛋白的拷贝融合，双特异性配体引起靶向蛋白的二聚化。该靶蛋白二聚化事件后诱导生物信号传导事件，例如红细胞生成或胰岛素信号。本文描述了在常规小分子药物(即，双特异性配体)的控制下，利用多瓣引导编辑通过基因组整合编码小分子结合蛋白(即，二聚化结构域)的基因来放置诱导二聚化生物过程(如，受体信号传导)的新方法。

[1421] 蛋白二聚化是普遍存在的生物过程。注意，已知许多膜结合受体的同源二聚化会引发信号级联反应，通常会产生深远的生物学后果。许多经批准用作药物的小分子天然产物因其作用机制的一部分充当蛋白二聚化的化学诱导物[92]。例如，FK506与FKBP12紧密结合，产生的小分子-蛋白复合物与磷酸酶钙神经素结合，从而抑制T细胞受体信号传导的一个步骤[93]。同样，环孢菌素A诱导亲环蛋白和钙神经素的二聚化，雷帕霉素诱导FKBP和mTOR的二聚化[93,94]。

[1422] 在一个实施方案中，利用FK506:FKBP12和环孢菌素A:亲环蛋白小分子:蛋白结合相互作用的选择性高亲和力结合，还开发了合成的二聚化化学诱导物。在示例中，当信号传导受体的胞质结构域用FKBP12标记时，由两个FK506单元组成的小分子(被称为FK1012)显示影响信号转导[95]。已将二聚化化学诱导物(CID)用于控制许多信号传导通路[96-103]。

[1423] 虽然是研究生物过程的有用工具，但用于治疗应用的合成CID面临的一个挑战是将FKBP12-或亲环蛋白-靶蛋白嵌合体引入患者具有挑战性。

[1424] 本公开将多个概念结合在一起形成以前无法达到的治疗过程。第一个概念是本文所述多瓣引导编辑(例如，双瓣或多瓣引导编辑)，其允许在活细胞中进行精确的基因组编辑，包括靶向插入。第二个概念是化学诱导的二聚化，这是使小分子能够控制细胞培养中的信号传导和寡聚化过程的强大工具。

[1425] 已确定对蛋白二聚化的化学控制可能具有有益的治疗效果的特定情况。

[1426] 胰岛素受体是异四聚体跨膜蛋白，其通过胞质激酶结构域的磷酸化响应于胰岛素与胞外结构域的结合[104]。工程化嵌合蛋白由膜定位组件、胰岛素受体的C端激酶结构域和三个拷贝的FKBP12组成，在细胞培养中响应于FK1012并启动胰岛素反应[99]。同样，预期FKBP12与患者细胞中天然胰岛素受体激酶结构域的C末端融合应当允许用于FK1012依赖性磷酸化和启动胰岛素信号传导级联。该系统可在不能制造胰岛素的患者(如，1型糖尿病患者)或对胰岛素反应较弱的患者(如，2型糖尿病患者)中替代或补充胰岛素使用。

[1427] 此外，促红细胞生成素通过与促红细胞生成素受体(EpoR)结合来刺激红细胞增殖，诱导预先形成的受体二聚体的二聚化或构象变化，从而导致激活Jak/STAT信号级联[105]。已证明FK1012诱导的用FKBP12标记的EpoR膜锚定胞质结构域的寡聚化足以启动Jak/STAT信号级联并促进细胞增殖[106]。预计通过在患者细胞中进行引导编辑将FKBP12与天然EpoR融合允许FK1012诱导的控制红细胞增殖(红细胞生成)。可将该系统用于触发贫血患者的红细胞生长。FK1012诱导型EpoR也可用作已经历离体工程的血细胞的体内可选择标志物。

[1428] 原则上，任何受体酪氨酸激酶都可能是多瓣引导编辑CID治疗的可行靶标。下表包括人类基因组中所有受体酪氨酸激酶的列表[110]。

家族 受体 别名 NT登录号 PROT登录号 染色体
家族

家族家族	受体	别名	NT登录号	PROT登录号	染色体
ALK家族 家族	ALK	Ki1	NM_004304	NP_004295	2p23
家族	LTK	TYK1	NM_002344	NP_002335	15q15.1-q21.1
AXL家族	AXL	UFO, Tyro7(r) Ark(m)	NM_001699	NP_001690	19q13.1
家族	MER	MERTK, NYK, Eyk(ch)	NM_006343	NP_006334	2q14.1
	TYRO3	RSE, SKY, BRT, DTK, TIF	NM_006293	NP_006284	15q15.1-q21.1
DDR家族	DDR1	CAK, TRKE, NEP NTRK4, EDDR1, PTK3	NM_013993	NP_001945	6p21.3
家族	DDR2	TKT, TYRO10, NTRKR3	NM_006182	NP_006173	1q21-q22
EGFR家族 家族	EGFR	ERBB, ERBB1	NM_005228	NP_005219	7p12
家族	ERBB2	HER2, Neu(r), NGL	NM_004448	NP_004439	17q11.2-q12
家族 家族	ERBB3	HER3	NM_001982	NP_001973	12q13
家族	ERBB4	HER4	NM_005235	NP_005226	2q33.3-q34
EPH家族	EPHA1	EPH, EPHT	NM_005232	NP_005223	7q32-q36
家族 家族	EPHA2	ECK, Sek(m), Myk2(m)	NM_004431	NP_004422	1p34
家族 家族 家族	EPHA3	HEK, ETK1, Tyro4(r), Mek4(m), Cek4(ch)	NM_005233	NP_005224	3p11.2
家族 家族	EPHA4	HEK8, Tyro1(r), Sek1(m), Cek8(ch)	NM_004438	NP_004429	2q36qter

家族

家族

[1429]

[1430]

		EPHA5	HEK7, Ehk(r), Bsk(r), Cek7(ch)	L36644	P54756	
		EPHA6	DKFZp434C1418, Ehk2(r)	AL133666		
		EPHA7	HEK11, Mdk1(m), Ebk(m), Ehk3(r), Cek11(ch)	NM_004440	NP_004431	6q21
		EPHA8	HEK3, KIAA1459, Eek(r), Cek10(ch)	AB040892	CAB81612	1q23-q24
		EPHB1	NET, EPHT2, HEK6, Elk(r), Cek6(ch)	NM_004441	NP_004432	3q21-q23
		EPHB2	HEK5, ERK, DRT, EPHT3, Tyro5(r), Nuk(m), Sek3(m), Cek5(ch)	AF025304	AAB94602	1p36.1-p35
		EPHB3	HEK2, Tyro6, Mdk5(m), Sek4(m)	NM_004443	NP_004434	3q21-qter
		EPHB4	HTK, Tyro11(r), Mdk2(m), Myk1(m)	NM_004444	NP_004435	
		EPHB6	HEP, Mep(m), Cek1(ch)	NM_004445	NP_004436	7q33-q35
	FGFR家族	FGFR1	FLT2, bFGFR, FLG, N-SAM	M34641	AAA35835	8p11.2
		FGFR2	KGFR, K-SAM, Bek(m), CFD1, JWS, Cek3(ch)	NM_000141	NP_000132	10q26
		FGFR3	HBGFR, ACH, Cek2(ch)	NM_000142	NP_000133	4p16.3
		FGFR4		NM_002011	NP_002002	5q35.1-qter
	INSR家族	IGF1R	JTK13	NM_000875	NP_000866	15q25-q26
		INSR	IR	NM_000208	NP_000199	19p13.3-p13.2
		INSRR	IRR	J05046	AAC31759	1q21-q23
	MET家族	MET	HGFR	NM_000245	NP_000236	7q31

[1431]

	RON	MST1R, CDw136, Fv2(m), STK(m), SEA(ch)	NM_002447	NP_002438	3p21.3
MUSK 家族	MUSK	Nsk2(m), Mlk1(m), Mk2(m)	NM_005592	NP_005583	9q31.3-q32
PDGFR 家族	CSF1R	FMS, C-FMS, CD115	NM_005211	NP_005202	5q31-q32
	FLT3	FLK2, STK1, CD135	NM_004119	NP_0041110	13q12
	KIT	Sfr(m), CKIT	NM_000222	NP_000213	4q11-q12
	PDGFRA		NM_006206	NP_006197	4q11-q13
	PDGFRB	PDGFR, JTK12	NM_002609	NP_002600	5q31-q32
PTK7家族	PTK7	CCK4, KLG(ch)	NM_002821	NP_002812	6p21.2-p12.2
RET家族	RET	MEN2A/B, HSCR1, MTC1	X12949	P07949	10q11.2
ROR家族	ROR1	NTRKR1	NM_005012	NP_005003	1p32-p31
	ROR2	NTRKR2	NM_004560	NP_004551	
ROS家族	ROS1	MCF3	NM_002944	NP_002935	6q22
RYK家族	RYK	Vlk(m), Mrk(m)	S59184	AAB263411	3q22
TIE家族	TEK	TIE2	NM_000459	NP_000450	9p21
	TIE	TIE1, JTK14	NM_005424	NP_005415	1p34-p33
TRK家族	NTRK1	TRK, TRKA	NM_002529	NP_002520	1q21-q22
	NTRK2	TRKB	NM_006180	NP_006171	9q22.1
	NTRK3	TRKC	NM_002530	NP_002521	15q25
VEGFR 家族	VEGFR1	FLT1	NM_002019	NP_002010	13q12
	VEGFR2	KDR, FLK1		AAB88005	4q11-q12
	VEGFR3	FLT4, PCL	NM_002020	NP_002011	5q34-q35
AATYK家族	AATYK	AATK, KIAA0641	NM_004920	NP_004911	17q25.3
	AATYK2	KIAA1079	NM_014916	NP_055731	7q21-q22
	AATYK3				19q13.2-q13.3

[1432] 多瓣引导编辑-CID系统有许多优点。一个这样的优点是它可用药物样小分子代替内源性配体，这些内源性配体通常是在制造、成本、递送、生产或储存时并发的蛋白，药物样小分子可口服施用，而不是IV或注射施用，很容易用FDA批准的药物制备(或本身已经是药物)，并且不会产生通常与蛋白药物相关的特殊生产或储存成本。另一个优点是单独的编辑应当没有生理影响。可通过给药小分子CID来控制靶蛋白二聚化的量。此外，通过添加单体

形式的CID,靶蛋白二聚化易于快速地逆转。另一个优点是,在单个配体靶向多个受体的情况下,可通过仅对一个受体进行引导编辑来实现选择性。最后,根据用于引导编辑的递送方法,还可将编辑限制在局部组织或器官,只允许在特定区域进行诱导受体激活。

[1433] 如果利用多瓣引导编辑的编辑效率足够高,两个单独的编辑事件可在高水平发生,那么也有可能利用不同的小分子结合结构域(如,FKBP和亲环蛋白)标记两种感兴趣的蛋白并诱导与小分子异源二聚体的异源二聚化(如,FK506-环孢菌素A二聚物)。

[1434] FKBP12或其他小分子结合蛋白与天然蛋白的融合通常是通过组织培养中的质粒过表达来完成的。已证明随后的化学诱导二聚化会诱导产生融合蛋白的细胞发生表型变化。

[1435] 以下参考文献在在上文部分中引用并且通过引用并入本文。

[1436] 1.Crabtree,G.R.&Schreiber,S.L.Three-part inventions:intracellular signaling and induced proximity.Trends Biochem.Sci.21,418-22(1996).

[1437] 2.Liu,J.et al.Calcineurin Is a Common Target of A and FKBP-FK506 Complexes.Cell 66,807-815(1991).

[1438] 3.Keith,C.T.etal.A mammalian protein targeted by G1-arresting rapamycin-receptor complex.Nature369,756-758(2003).

[1439] 4.Spencer,D.M.,Wandless,T.J.,Schreiber,S.L.S.&Crabtree,G.R.Controlling signal transduction with synthetic ligands.Science262,1019-24(1993).

[1440] 5.Pruschy,M.N.et al.Mechanistic studies of a signaling pathway activated by the organic dimerizer FK1012.Chem.Biol.1,163-172(1994).

[1441] 6.Spencer,D.M.et al.Functional analysis of Fas signaling in vivo using synthetic inducers of dimerization.Curr.Biol.6,839-847(1996).

[1442] 7.Belshaw,P.J.,Spencer,D.M.,Crabtree,G.R.&Schreiber,S.L.Controlling programmed cell death with a cyclophilin-cyclosporin-based chemical inducer of dimerization.Chem.Biol.3,731-738(1996).

[1443] 8.Yang,J.X.,Symes,K.,Mercola,M.&Schreiber,S.L.Small-molecule control of insulin and PDGF receptor signaling and the role of membrane attachment.Curr.Biol.8,11-18(1998).

[1444] 9.Belshaw,P.J.,Ho,S.N.,Crabtree,G.R.&Schreiber,S.L.Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins.Proc.Natl.Acad.Sci.93,4604-4607(2002).

[1445] 10.Stockwell,B.R.&Schreiber,S.L.Probing the role of homomeric and heteromeric receptor interactions in TGF-βsignaling using small molecule dimerizers.Curr.Biol.8,761-773(2004).

[1446] 11.Spencer,D.M.,Graef,I.,Austin,D.J.,Schreiber,S.L.&Crabtree,G.R.A general strategy for producing conditional alleles of Src-like tyrosine kinases.Proc.Natl.Acad.Sci.92,9805-9809(2006).

[1447] 12.Holsinger,L.J.,Spencer,D.M.,Austin,D.J.,Schreiber,S.L.&Crabtree,

G.R.Signal transduction in T lymphocytes usinga conditional allele of Sos.Proc.Natl.Acad.Sci.92,9810-9814(2006).

[1448] 13.Myers,M.G.Insulin Signal Transduction and the IRS Proteins.Annu.Rev.Pharmacol.Toxicol.36,615-658(1996).

[1449] 14.Watowich,S.S.The erythropoietin receptor:Molecular structure and hematopoietic signaling pathways.J.Investig.Med.59,1067-1072(2011).

[1450] 15.Blau,C.A.,Peterson,K.R.,Drachman,J.G.&Spencer,D.M.A proliferation switch for genetically modified cells.Proc.Natl.Acad.Sci.94,3076-3081(2002).

[1451] 16.Clackson,T.et al.Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity.Proc.Natl.Acad.Sci.95,10437-10442(1998).

[1452] 17.Diver,S.T.&Schreiber,S.L.Single-step synthesis of cell-permeable protein dimerizers that activate signal transduction and gene expression.J.Am.Chem.Soc.119,5106-5109(1997).

[1453] 18.Guo,Z.F.,Zhang,R.&Liang,F.Sen.Facile functionalization of FK506 for biological studies by the thiol-ene‘click’reaction.RSCAdv.4,11400-11403(2014).

[1454] 19.Robinson,D.R.,Wu,Y.-M.&Lin,S.-F.The protein tyrosine kinase family ofthe human genome.Oncogene 19,5548-5557(2000).

[1455] 前间隔序列前间隔序列前间隔序列前间隔序列G.多瓣引导编辑在插入重组酶靶位点中的应用

[1456] 另一方面,可使用多瓣引导编辑(例如,如图90、95、97A、98A所示)将重组酶位点(或“重组酶识别序列”)插入期望的基因组位点。重组酶位点的插入提供了用于影响基因组中位点特异性基因变化的编程位置。这类基因变化可包括,例如质粒的基因组整合、基因组缺失或插入、染色体易位和盒交换,以及其他基因变化。这些示例性类型的基因变化示于图64(b)-(f)。然后可使用安装的重组酶识别序列在该位点进行位点特异性重组来产生多种重组结果,例如DNA片段的切除、整合、倒置或交换。例如,图65显示可用于整合包含GFP表达标志物的DNA供体模板的重组酶位点的安装。含有整合到重组酶位点的GFP表达系统的细胞会发荧光。

[1457] 在基因组中安装重组酶位点的机制类似于在基因组中安装其他序列,如肽/蛋白和RNA标签。示例重组酶靶序列的安装的示意图示于图64(a)。该过程从选择将引入重组酶靶序列的期望的靶基因座开始。接着,提供了引导编辑器融合物(“RT-Cas9:gRNA”)。在此,“gRNA”是指PEgRNA,其可使用本文所述原理设计。不同实施方案中的PEgRNA将包括对应于图3D中(5′-[约20-nt间隔区]-[gRNA核心]-[延伸臂]-3′的结构,其中延伸臂在3′至5′方向包含引物结合位点(“A”)、编辑模板(“B”)和同源臂(“C”)。编辑模板(“B”)将包含对应于重组酶位点的序列,即编码互补单链DNA的PEgRNA的单链RNA,其为重组酶位点的有义链或反义链,通过多瓣引导编辑过程并入基因组DNA靶基因座。

[1458] 在不同方面,本公开提供了使用多瓣PE在人或其他基因组中的高价值基因座处引入重组酶识别序列,其在暴露于位点特异性重组酶之后将指导精确且有效的基因组修饰

(图64)。在图64所示的不同实施方案中,可将PE用于(b)插入单个SSR靶标,用作DNA供体模板基因组整合的位点。(c)显示如何使用串联插入SSR靶位点来删除部分基因组。(d)显示如何使用串联插入SSR靶位点来倒置部分基因组。(e)显示如何在两个远端染色体区域插入两个SSR靶位点来导致染色体易位。(f)显示如何使用基因组中两个不同SSR靶位点的插入来交换来自DNA供体模板的盒。每种类型的基因组修饰都是通过使用PE插入SSR靶标来设想的,但该列表也不意味着是限制性的。

[1459] 多瓣PE介导的重组酶识别序列的引入对于治疗由大规模基因组缺陷引起的遗传性疾病特别有用,如基因丢失、倒置或重复,或染色体易位[1-7](表6)。例如,Williams-Beuren综合征是由染色体721中的24个缺失引起的发育病症。目前尚无技术可用于将多个完整基因高效且靶向插入活细胞中(目前正在探索PE进行这种全长基因插入的潜力,但尚未确定);然而,在PE插入的靶标处重组酶介导的整合提供了一种永久治愈这种疾病和其他疾病的方法。此外,靶向引入重组酶识别序列可为包括转基因植物、动物研究模型、生物生产细胞系或其他定制真核细胞系在内的应用提供高度支持。例如,在转基因植物中多瓣PE特异性靶标处重组酶介导的基因组重排可克服产生具有改进特性的农作物的瓶颈之一[8,9]。

[1460] 表6.通过重组酶识别序列的基于多瓣PE的安装可修复与大规模基因组修饰相关的遗传性疾病示例。

[1461]

疾病	起因
三体性17P	基因重复
夏-马-图三氏病I型	基因重复
SMITH-MAGENIS综合征	基因缺失
WILLIAMS-BEUREN综合征	基因缺失
DELACHAPELLE综合征	染色体易位
唐氏(DOWN)综合征(某些类型)	染色体易位
血友病A	基因倒置
亨特氏(HUNTER)综合征	基因倒置

[1462] 已表征了许多SSR家族成员并描述了它们的重组酶识别序列,包括天然和工程化酪氨酸重组酶(表7)、大型丝氨酸整合酶(表8)、丝氨酸解离酶(表9)和酪氨酸整合酶(表10)。针对几个SSR还描述了显示提高基因组整合率的经修饰的靶序列[22-30]。除了天然重组酶之外,还开发了具有不同特异性的可编程重组酶[31-40]。使用PE,可将这些识别序列中的一个或多个引入基因组的指定位置,例如安全港基因座[41-43],取决于期望的应用。

[1463] 例如,通过引导编辑在基因组中引入单个重组酶识别序列将导致与DNA供体模板的整合重组(图64b)。在人细胞中稳健作用的丝氨酸整合酶,可能特别适合于基因整合[44,45]。

[1464] 此外,取决于靶标的身份和方向,引入两个重组酶识别序列可导致间插序列的缺失、间插序列的倒置、染色体易位或盒交换(图64c至f)。通过选择已与重组酶靶标非常相似的内源性序列,引入完整重组酶靶标所需的编辑范围会减少。

[1465] 最后,已证明几种重组酶可在天然存在的假位点(pseudosite)处整合到人或真核生物基因组中[46-64]。可将多瓣引导编辑用于修饰这些基因座,以提高这些天然假位点的整合率,或者可选择地,消除可能用作不需要的脱靶序列的假位点。

[1466] 本公开描述了使用多瓣PE在真核基因组中引入重组酶靶序列的一般方法,其应用

几乎是无限制的。基因组编辑反应旨在与“引导编辑器”一起使用，“引导编辑器”是CRISPR/Cas9蛋白和逆转录酶结构域的嵌合融合物，其利用定制的引导编辑向导RNA(PEgRNA)。相关地，还可利用Cas9工具和同源定向修复(HDR)途径，通过使用几种技术降低indel率，通过DNA模板来引入重组酶识别序列[65-67]。人细胞培养中的概念验证实验示于图65。在一些实施方案中，引导编辑器是多瓣引导编辑器。

[1467] 在以上关于重组酶识别序列的PE引导安装的描述中引用了以下几个表，并提供了可使用的示例性重组酶的列表，以及它们可由多瓣PE安装的同源重组酶识别序列。

[1468] 表7.酪氨酸重组酶和SSR靶序列

[1469]

重组酶	重组酶识别序列	命名
Cre	ATAACTTCGTATAGCATACATTATACGAAGTTAT (SEQ ID NO: 517)	*loxP*
Dre	TAACTTTAAATAATGCCAATTATTTAAAGTTA (SEQ	*rox*

重组酶	重组酶识别序列	命名
	ID NO: 518)	
VCre	TCAATTTCTGAGAACTGTCATTCTCGGAAATTGA (SEQ ID NO: 519)	*loxV*
SCre	CTCGTGTCCGATAACTGTAATTATCGGACATGAT (SEQ ID NO: 520)	*loxS*
Flp	GAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC (SEQ ID NO: 521)	*FRT*
B2	GAGTTTCATTAAGGAATAACTAATTCCCTAATGAA ACTC (SEQ ID NO: 522)	*loxB*
B3	GGTTGCTTAAGAATAAGTAATTCTTAAGCAACC (SEQ ID NO: 523)	*loxB3*
Kw	ACGAAAAATGGTAAGGAATAGACCATTCCTTACC ATTTTTGGT (SEQ ID NO: 524)	
R	TTGATGAAAGAATAACGTATTCTTTCATCAA (SEQ ID NO: 525)	*RS*
TD1-40	GTGCGTCAAATAATAACGTATTATTTGACACTT (SEQ ID NO: 526)	*TDRS*
Vika	AATAGGTCTGAGAACGCCCATTCTCAGACGTATT (SEQ ID NO: 527)	*vox*
Nigri	TGAATGTCCTATAATTACACTTATAGGACATTCA (SEQ ID NO: 528)	*nox*
Panto	GAAACTTTAAATAATAAGTCTTATTTAAAGTTTC (SEQ ID NO: 529)	*pox*
Kd	AAACGATATCAGACATTTGTCTGATAATGCTTCA TTATCAGACAAATGTCTGATATCGTTT (SEQ ID NO: 530)	*loxK*
Fre	ATATATACGTATATAGACATATATACGTATATAT (SEQ ID NO: 531)	*loxH*
CreALSHG	ATAACTCTATATAATGTATGCTATATAGAGTTAT (SEQ ID NO: 532)	*loxM7*
Tre	ACAACATCCTATTACACCCTATATGCCAACATGG (SEQ ID NO: 533)	*loxLTR*
Brec1	AACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT (SEQ ID NO: 534)	*loxBTR*
Cre-R3M3	GATACAACGTATATACCTTTCTATACGTTGTTTA (SEQ ID NO: 535)	*loxK2*

[1470]

[1471] 表8.大型丝氨酸整合酶和SSR靶序列

重组酶	重组酶识别序列 左侧	重组酶识别序列 右侧
Bxb1	GGTTTGTCTGGTCAACCACCGC GGTCTCAGTGGTGTACGGTACA AACC (SEQ ID NO: 536)	GGCTTGTCGACGACGGCGGTCTCC GTCGTCAGGATCAT (SEQ ID NO: 537)

[1472]

[1473]

重组酶	重组酶识别序列 左侧	重组酶识别序列 右侧
phiC31	GTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG (SEQ ID NO: 538)	TGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC (SEQ ID NO: 539)
R4	TGTTCCCCAAAGCGATACCACTTGAAGCAGTGGTACTGCTTGTGGGTACA (SEQ ID NO: 540)	GCATGTTCCCCAAAGCGATACCACTTGAAGCAGTGGTACTGCTTGTGGGTACACTCTGCGGGTG (SEQ ID NO: 541)
phiBT1	GGTGCTGGGTTGTTGTCTCTGGACAGTGATCCATGGGAAACTACTCAGCACC (SEQ ID NO: 542)	CAGGTTTTTGACGAAAGTGATCCAGATGATCCAG (SEQ ID NO: 543)
MJ1 (phiFC1)	ATTTTAGGTATATGATTTTGTTTATTAGTGTAAATAACACTATGTACCTAAAAT (SEQ ID NO: 544)	CAAAGGATCACTGAATCAAAAGTATTGCTCATCCACGCGAAA (SEQ ID NO: 545)
MR11	TTTGTGCGGAACTACGAACAGTTCATTAATACGAAGTGTACAAACTTCCATACAA (SEQ ID NO: 546)	CGAAAATGTATGGAGGCACTTGTATCAATATAGGATGTATACCTTCGAAGACACTT (SEQ ID NO: 547)
TP901-1	GAGTTTTTATTTCGTTTATTTCAATTAAGGTAACTAAAAAACTCCTTTTAAGG (SEQ ID NO: 548)	ATGCCAACACAATTAACATCTCAATCAAGGTAAATGCTTTTTGCTTTTTTTGC (SEQ ID NO: 549)
A118	TTCCTCGTTTTCTCTCGTTGGAAGAAGAAGAAACGAGAAA (SEQ ID NO: 550)	TTTCGGATCAAGCTATGAAGGACGCAAAGAGGGAACTAAA(SEQ ID NO: 551)
U153	TTCCTCGTTTTCTCTCGTTGGACGGAAACGAATCGAGAAA (SEQ ID NO: 552)	TTTCGGATCAAGCTATGAAGGACGCAAAGAGGGAACTAAA (SEQ ID NO: 553)
phiRV1	GTAGTGTATCTCACAGGTCCACGGTTGGCCGTGGACTGCTGAAGAACATTCC (SEQ ID NO: 554)	GAAGGTGTTGGTGCGGGGTTGGCCGTGGTCGAGGTGGGGT (SEQ ID NO: 555)
phi370.1	AAAAAAATACAGCGTTTTTCATGTACAACTATACTAGTTGTAGTGCCTAAAA (SEQ ID NO: 556)	TTGTAAAGGAGACTGATAATGGCATGTACAACTATACTCGTCGGTAAAAAGGCA (SEQ ID NO: 557)
TG1	TCCAGCCCAACAGTGTTAGTCTTTGCTCTTACCCAGTTGGGCGGGA (SEQ ID NO: 558)	GATCAGCTCCGCGGGCAAGACCTTTCTCCTTCACGGGGTGGAAGGTC (SEQ ID NO: 559)
WB	CTAGTTTTAAAGTTGGTTATTAGTTACTGTGATATTTATCACGGTACCCAATAACCAATGAAT (SEQ ID NO: 560)	CGGAAGGTAGCGTCAACGATAGGTGTAACTGTCGTGTTTGTAACGGTACTTCCAACAGCTGGCGCCGCCAC (SEQ ID NO: 561)
BL3	CAATGAAAAACTAGGCATGTAGAAGTTGTTTGT (SEQ ID NO: 562)	TTTCCACAGACAACTCACGTGGAGGTAGTCAC (SEQ ID NO: 563)
SprA	TGTAGTAAGTATCTTAATATACAGCTTTATCTGTTTTTTAAGATACTTACTACTTT (SEQ ID NO:	CACCCATTGTGTTCACAGGAGATACAGCTTTATCTGTACTGATATTAATGACATGCTG (SEQ ID NO: 565)

[1474]

重组酶	重组酶识别序列 左侧	重组酶识别序列 右侧
	564)	
phiJoe	AGTTGTGGCCATGTGTCCATCTGGGGGCAGATGGAGACGGGGTCACA (SEQ ID NO: 566)	ATCTGGATGTGGGTGTCCATCTGCGGGCAGACGCCGCAGTCGAAGCACGG (SEQ ID NO: 567)
--	--	ACCTTGATCTCGGTGTCCATCGCCGGGCAGACGCCGCAGTCGAAGCACGG (SEQ ID NO: 568)
phiK38	CCCTAATACGCAAGTCGATAACTCTCCTGGGAGCGTTGACAACTTGCGCACCCTGATCTG (SEQ ID NO: 569)	GAGCGCCGGATCAGGGAGTGGACGGCCTGGGAGCGCTACACGCTGTGGCTGCGGTCGGTGC (SEQ ID NO: 570)
Int2	GCTCATGTATGTGTCTACGCGAGATTCTCGCCCGAGAACTTCTGCAAGGCACTGCTCTTGGCT (SEQ ID NO: 571)	GGACGGCGCAGAAGGGGAGTAGCTCTTCGCCGGACCGTCGACATACTGCTCAGCTCGTC (SEQ ID NO: 572)
Int3	ATGGATAAAAAAATACAGCGTTTTTCATGTACAACTATACTAGTTGTAGTGCCTAAATAATGCTT (SEQ ID NO: 573)	GTTTGTAAAGGAGACTGATAATGGCATGTACAACTATACTCGTCGGTAAAAAGGCATCTTAT (SEQ ID NO: 574)
Int4	AAAAATTACAAAGTTTTCAACCCTTGATTTGAATTAGCGGTCAAATAATTTGTAATTCGTTT (SEQ ID NO: 575)	TTCCAAAGAGCGCCCAACGCGACCTGAAATTTGAATAAGACTGCTGCTTGTGTAAAGGCGATGATT (SEQ ID NO: 576)
Int7	GTGTTATAAACCTGTGTGAGAGTTAAGTTTACATGCCTAACCTTAACTTTTACGCAGGTTCAGCTT (SEQ ID NO: 577)	AGACGAGAAACGTTCCGTCCGTCTGGGTCAGTTGGGCAAAGTTGATGACCGGGTCGTCCGTT (SEQ ID NO: 578)
Int8	TTAATAAACTATGGAAGTATGTACAGTCTTGCAATGTTGAGTGAACAAACTTCCATAATAAAAT (SEQ ID NO: 579)	CAATCATCAGATAACTATGGCGGCACGTGCATTAACCACGGTTGTATCCCGTCTAAAGTACTCGT (SEQ ID NO: 580)
Int9	GTGGTTGTTTTTGTTGGAAGTGTGTATCAGGTATCTGCATAGTTATTCCGAACTTCCAATTA (SEQ ID NO: 581)	TTTATATTGCGAAAAATAATTGGCGAACGAGGTAACTGGATACCTCATCCGCCAATTAAAATTTG (SEQ ID NO: 582)
Int10	GGAAAATATAAATAATTTTAGTAACCTACATCTCAATCAAGGATAGTAAAACTCTCACTCTT (SEQ ID NO: 583)	AGCACGCTGATAATCAGCAAGACCACCAACATTTCCACCAATGTAAAAGCTTTAACCTTAGC (SEQ ID NO: 584)
Int11	GTTTATATGTTTACTAATAAGACGCTCTCAACCCATAAAGTCTTATTAGTAAACATATTTCAACT (SEQ ID NO: 585)	ATGGATTTTGCAGATTCCCAGATGCCCCTACAGAAAGAGGTACAAAACATTTATTGGAATTAATT (SEQ ID NO: 586)
Int12	TTTTTGTATGTTAGTTGTGTCACTGGGTAGACCTAAATAGTGA	GTTCGTGGTAACTATGGGTGGTACAGGTGCCACATTAGTTGTACCATTT

[1475]

重组酶	重组酶识别序列 左侧	重组酶识别序列 右侧
	CACAACTGCTATTAAAATTTAA (SEQ ID NO: 587)	ATGTTTATGTGGTTAAC (SEQ ID NO: 588)
Int13	CAATAACGGTTGTATTTGTAGA ACTTGACCAGTTGTTTTAGTAA CATAAATACAACTCCGAATA (SEQ ID NO: 589)	GCATACATTGTTGTTGTTTTTCCAG ATCCAGTTGGTCCTGTAAATATAA GCAATCCATGTGAGT (SEQ ID NO: 590)
LI	GTTTAGTATCTCGTTATCTCTC GTTGGAGGGAGAAGAAACGGG ATACCAAAA (SEQ ID NO: 591)	TAACTTTTTCGGATCGAGTTATGAT GGACGTAAAGAGGGAACAAAGCA TCTA (SEQ ID NO: 592)
Peaches	TAGTTTCCAATGTTACAGGAAC TGCTGGCAGAATCCAACACATT GGAAGTCG (SEQ ID NO: 593)	CGGTCTCCATCGGGATCTGCTGATC GAGCAGCATGCCGACCA (SEQ ID NO: 594)
Bxz2	TAACCGCAAGTGTACATCCCTC GGCTGGCCGAGACAAGTACAG TTGCGACAG (SEQ ID NO: 595)	CGGTCTCCATCGGGATCTGCTGATC GAGCAGCATGCCGACCA (SEQ ID NO: 596)
SV1	ATGTGGTCCTTTAGATCCACTG ACGTGGGTCAGTGTCTCTAAAG GACTCGCG (SEQ ID NO: 597)	CATCAGGGCGGTCAGGCCGTAGAT GTGGAAGAAACGGCAGCACGGCG AGGACG (SEQ ID NO: 598)

[1476] 表9.丝氨酸解离酶和SSR靶序列

[1477]

解离酶	重组酶识别序列 左侧	重组酶识别序列 右侧
Gin	CGTTTCCTGTAAACCGAGGTTTTG GATAAACA (SEQ ID NO: 599)	CGTTTCCTGTAAACCGAGGTTTTG GATAATGG (SEQ ID NO: 600)
Cin	GAGTTCTCTTAAACCAAGGTTTA GGATTGAAA (SEQ ID NO: 601)	GAGTTCTCTTAAACCAAGGTATT GGATAACAG (SEQ ID NO: 602)
Hin	TGGTTCTTGAAAACCAAGGTTTTT GATAAAGC (SEQ ID NO: 603)	AAATTTTCCTTTTTGGAAGGTTTT TGATAACCA (SEQ ID NO: 604)
Min	GCCTTCCCCTAAACCAACGTTTTT ATGCCGCC (SEQ ID NO: 605)	GCCTTCCCCCAAACCAAGGTAAT CAAGAACGC (SEQ ID NO: 606)
Sin	TTGTGAAATTTGGGTACACCCTA ATCATACAA (SEQ ID NO: 607)	CGTATGATTAGGGTGTATATTAA TTT (SEQ ID NO: 608)

[1478] 表10.酪氨酸整合酶和靶序列

[1479]

整合酶	*attP*	*attB*
HK022	CAAATGATTTTATTTTGACTAATAA TGACCTACTTACATTAATTTACTGAT AATTAAAGAGATTTTAAATATACAA CTTATTCACCTAAAGGATGACAAAA (SEQ ID NO: 609)	GCACTTTAGGTGAAAAAGGT T (SEQ ID NO: 610)
	TAACATTAATCACTTAAAAATCATC GCATTACACTAATCTGTGGTTAAAT GATAGACTACATAATGCGACAAAA	

[1480]

	CGCAACATATCCAGTCACTATGAAT CAACTACTTAGATAGTATTAGTGAC CT (SEQ ID NO: 611)	
P22	CTAAGTGGTTTGGGACAAAAATGGG ACATACAAATCTTTGCATCGGTTTG CAAGGCTTTGCATGTCTTTCGAAGA TGGGACGTGTGAGCGCAGGTATGAC GTGGTATGTGTTGACTTAAAAGGTA GTTCTTATAATTCGTAATGCGAAGG TCGTAGGTTCGACTCCTATTATCGG CACCAGTTAAATCAAATACTTACGT ATTATTCGTGCCTTCCTTATTTTTAC TGTGGGACATATTTGGGACAGAAGT ACCAAAAA (SEQ ID NO: 612)	GCAGCGCATTCGTAATGCGA AGGTCGT (SEQ ID NO: 613)
L5	GCGATCCCCATCCGCGACGTGCCAA CTAGGTCTCCTCTCGTCGTGAACAA GGCTACCGGGTTGCAACTCCTGTGC AACTCTCAGGCTTCAACGCGCTTCT ACGACCTGCAATTTCTTTCCACTTA GAGGATGCAGCCGAGAGGGGTAAA AACCTATCTTGACCGGCCCATATGT GGTCGGCAGACACCCATTCTTCCAA ACTAGCTACGCGGGTTCGATTCCCG TCGCCCGCTCCGCTGGTCAGAGGGT GTTTTCGCCCTCTGGCCATTTTTCTT TCCAGGGGTCTGCAACTCTTGTGCG ACTCTTCTGACCTGGGCATACGCGG TTGCAACGCATCCCTGATCTGGCTA CTTTCGATGCTGACAAACGAATAGA GCCCCCCGCCTGCGCGAACAGACG AGGGGCATTCACA (SEQ ID NO: 614)	GAGCGGGCGACGGGAATCG AACCCGCGTAGCTAGTTTGG AAGA (SEQ ID NO: 615)

[1481] 在不同的其他方面,本公开涉及使用多瓣PE来安装一个或多个重组酶识别序列的方法及其在位点特异性重组中的应用。

[1482] 在一些实施方案中,位点特异性重组可产生多种重组结果,例如DNA片段的切除、整合、倒置或交换。

[1483] 在一些实施方案中,该方法可用于诱导两个或多个核酸(如,DNA)分子的两个或多个区域之间的重组。在其他实施方案中,该方法可用于诱导单个核酸分子(如,DNA)中的两个或更多个区域的重组或在两个或更多个区域之间的重组。

[1484] 在一些实施方案中,本公开提供了通过位点特异性重组整合供体DNA模板的方法,其包括:(a)通过多瓣引导编辑在基因组基因座安装重组酶识别序列;(b)在重组酶存在下,使基因组基因座与还包含重组酶识别序列的DNA供体模板接触。

[1485] 在其他实施方案中,本发明提供了通过位点特异性重组缺失基因组区域的方法,其包括:(a)通过多瓣引导编辑在基因组基因座安装一对重组酶识别序列;(b)使基因组基因座与重组酶接触,从而催化重组酶识别序列对之间的基因组区域的缺失。

[1486] 在其他实施方案中,本公开提供了通过位点特异性重组倒置基因组区域的方法,其包括:(a)通过多瓣引导编辑在基因组基因座安装一对重组酶识别序列;(b)使基因组基因座与重组酶接触,从而催化重组酶识别序列对之间的基因组区域的倒置。

[1487] 在其他实施方案中，本公开提供了用于诱导第一基因组位点和第二基因组位点之间的染色体易位的方法，其包括：(a)通过引导编辑在第一基因组位基因座安装第一重组酶识别序列；(b)通过多瓣引导编辑在第二基因组基因座安装第二重组酶识别序列；(c)使第一和第二基因组基因座与重组酶接触，从而催化第一和第二基因组基因座的染色体易位。

[1488] 在其他实施方案中，本公开提供了用于诱导基因组位点和包含盒的供体DNA之间的盒交换的方法，其包括：(a)通过多瓣引导编辑在第一基因组位点安装第一重组酶识别序列；(b)通过多瓣phiBT•引导编辑在第二基因组位点安装第二重组酶识别序列；(c)使第一和第二基因组座位与供体DNA接触，该供体DNA包含侧接第一和第二重组酶识别序列和重组酶的盒，从而催化DNA供体中侧接的基因组基因座和盒的交换。

[1489] 在涉及在基因组中插入多于一种重组酶识别序列的不同实施方案中，重组酶识别序列可以相同或不同。在一些实施方案中，重组酶识别序列是相同的。在其他实施方案中，重组酶识别序列是不同的。

[1490] 在不同实施方案中，重组酶可以是酪氨酸重组酶，如Cre、Dre、Vcre、Sre、Flp、B2、B3、Kw、R、TD1-40、Vika、Nigri、Panto、Kd、Fre、Cre(ALSHG)、Tre、Brec1、或Cre-R3M3，如表7所示。在这些实施方案中，重组酶识别序列可以是表7中对应于使用的重组酶的RRS。

[1491] 在不同的其他实施方案中，重组酶可以是大型丝氨酸重组酶，例如Bxb1、PhiC31、R4、phiBT1、MJ1、MR11、TP901-1、A118、V153、phiRV1、phi370.1、TG1、WB、BL3、SprA、phiJoe、phiK38、Int2、Int3、Int4、Int7、Int8、Int9、Int10、Int11、Int12、Int13、L1、peaches、Bxz2、或SV1，如表8所示。在这些实施方案中，重组酶识别序列可以是表8中对应于使用的重组酶的RRS。

[1492] 在其他实施方案中，重组酶可以是丝氨酸重组酶，例如Bxb1、PhiC31、R4、phiBT1、MJ1、MR11、TP901-1、A118、V153、phiRV1、phi370.1、TG1、WB、BL3SprA、phiJoe、phiK38、Int2、Int3、Int4、Int7、Int8、Int9、Int10、Int11、Int12、Int13、L1、peaches、Bxz2、或SV1，如表8所示。在这些实施方案中，重组酶识别序列可以是表8中对应于使用的重组酶的RRS。

[1493] 在其他实施方案中，重组酶可以是丝氨酸解离酶，例如Gin、Cin、Hin、Min、或Sin，如表9所示。在这些实施方案中，重组酶识别序列可以是表9中对应于使用的重组酶的RRS。

[1494] 在不同的其他实施方案中，重组酶可以是酪氨酸整合酶，如HK022、P22、或L5，如表10所示。在这些实施方案中，重组酶识别序列可以是表10中对应于使用的重组酶的RRS。

[1495] 在一些实施方案中，利用多瓣PE进行位点特异性重组的任何方法可在体内或体外进行。在一些实施方案中，用于位点特异性重组的任何方法在细胞中进行(如，细胞中的重组基因组DNA)。细胞可以是原核的或真核的。细胞，如真核细胞，可在个体如受试者中，如本文所述(如，人类受试者)。本文所述方法可用于体外和体内细胞的基因修饰，例如，在产生转基因细胞、细胞系或动物的情况下，或在基因组序列的改变中，例如，校正受试者细胞中的遗传缺陷。

[1496] [8]本文所述的多瓣引导编辑器的实施方案

[1497] 以下非限制性实施方案描述了本申请所述的双引导编辑器及其用途的多个方面。

[1498] 1.用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统，其中所述系统包含第一引导编辑器复合物和第二引导编辑器复合物，其中所述第一和第二引导编辑器复合物中的每一个包含：(1)引导编辑器，其包含：(i)核酸可编程DNA结合蛋白(napDNAbp)，

和(ii)具有RNA依赖性DNA聚合酶活性的多肽;和(2)pegRNA,其包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点,其中所述第一引导编辑器复合物的所述pegRNA的所述DNA合成模板编码第一单链DNA序列,并且所述第二引导编辑器复合物的所述pegRNA的所述DNA合成模板编码第二单链DNA序列,其中所述第一单链DNA序列和所述第二单链DNA序列各自包含彼此互补的区域,并且其中所述第一单链DNA序列和所述第二单链DNA序列形成双链体,所述双链体包含与所述待编辑的靶位点处的所述DNA序列相比的编辑部分,所述编辑部分整合到所述待编辑的靶位点中。

[1499] 2.实施方案1的系统,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白,所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

[1500] 3.实施方案1或2的系统,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含相同的引导编辑器。

[1501] 4.实施方案1或2的系统,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含不同的引导编辑器。

[1502] 5.实施方案1至4中任一项的系统,其中所述napDNAbp是Cas9结构域或其变体。

[1503] 6.实施方案1至5中任一项的系统,其中所述napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域、Cas9切口酶结构域或其变体。

[1504] 7.实施方案1至4中任一项的系统,其中所述napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c、CasX、CasY和Argonaute。

[1505] 8.实施方案1至7中任一项的系统,其中所述napDNAbp具有切口酶活性。

[1506] 9.实施方案1至4中任一项的系统,其中所述napDNAbp包含SEQ ID NO:18-88中任何一个的氨基酸序列,或与SEQ ID NO:18-88中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1507] 10.实施方案1至9中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的多肽是逆转录酶。

[1508] 11.实施方案1至10中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的多肽包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159和700-766中任何一个的氨基酸序列,或与SEQ ID NO:89-100、105-122、128-129和132中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1509] 12.实施方案1至11中任一项的系统,其中所述引导编辑器进一步包含接头,所述接头接合所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

[1510] 13.实施方案12的系统,其中所述接头包含SEQ ID NO:166-177中任何一个的氨基酸序列,或与SEQ ID NO:166-177中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1511] 14.实施方案12或13的系统,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

[1512] 15.实施方案1至14中任一项的系统,其中所述pegRNA包含SEQ ID NO:325-330、

499-505、101-104、181-183和223-244中任何一个的核苷酸序列，或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

[1513] 16.实施方案1至15中任一项的系统，其中所述第一引导编辑器复合物的所述pegRNA的所述间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

[1514] 17.实施方案1至16中任一项的系统，其中所述第二引导编辑器复合物的所述pegRNA的所述间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

[1515] 18.实施方案17的系统，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述pegRNA的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被所述第一和第二引导编辑器复合物中每一个的所述napDNAbp生成切口。

[1516] 19.实施方案18的系统，其中所述第一和第二结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

[1517] 20.实施方案19的系统，其中所述第一和第二结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1518] 21.实施方案19的系统，其中所述连续区域被包含所述编辑部分的双链体置换。

[1519] 22.实施方案1至21中任一项的系统，其中所述DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1520] 23.实施方案1至22中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有相同的长度。

[1521] 24.实施方案1至22中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有不同的长度。

[1522] 25.实施方案1至24中任一项的系统，其中所述互补区域为所述第一单链DNA序列或所述第二单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

[1523] 26.实施方案25的系统，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷

酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109核苷酸或至少110个核苷酸，并且小于或等于所述期望的DNA序列及其互补物的长度。

[1524] 27.实施方案25的系统，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域形成包含所述编辑部分的所述双链体。

[1525] 28.实施方案1至27中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列各自是具有3’末端的单链DNA链。

[1526] 29.实施方案21的系统，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

[1527] 30.用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统，所述系统包含：

[1528] a)第一引导编辑器复合物，其包含：

[1529] i.第一引导编辑器，其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽；和

[1530] ii.第一引导编辑向导RNA(第一PEgRNA)，其结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点；

[1531] b)第二引导编辑器复合物，其包含：

[1532] iii.第二引导编辑器，其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽；和

[1533] iv.第二引导编辑向导RNA(第二PEgRNA)，其结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点；

[1534] 其中所述第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板，并且所述第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板，并且其中所述双链DNA序列的所述第一链和所述第二链彼此互补。

[1535] 31.实施方案30的系统，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白，所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多

肽。

[1536] 32.实施方案30的系统,其中所述第一和第二单链DNA序列各自包含彼此互补的区域。

[1537] 33.实施方案30的系统,其中所述第一单链DNA序列和所述第二单链DNA序列形成包含整合到所述待编辑的靶位点中的所述编辑部分的双链体。

[1538] 34.实施方案30的系统,其中所述第一单链DNA序列和/或所述第二单链序列不包含与所述靶位点处的所述DNA序列相比的序列同源性。

[1539] 35.实施方案30的系统,其中所述第一PEgRNA进一步包含第一间隔区序列、第一gRNA核心和第一引物结合位点,并且所述第二PEgRNA进一步包含第二间隔区序列、第二gRNA核心和第二引物结合位点。

[1540] 36.实施方案30至35中任一项的系统,其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含相同的引导编辑器。

[1541] 37.实施方案30至36中任一项的系统,其中所述包含RNA依赖性DNA聚合酶活性的第一多肽和所述包含RNA依赖性DNA聚合酶活性的第二多肽是相同的。

[1542] 38.实施方案30至37中任一项的系统,其中所述第一和/或第二napDNAbp是Cas9结构域或其变体。

[1543] 39.实施方案30至38中任一项的系统,其中所述第一和/或第二napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域、Cas9切口酶结构域或其变体。

[1544] 40.实施方案30至39中任一项的系统,其中所述第一和/或第二napDNAbp选自由以下各项组成的组:Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c、CasX、CasY和Argonaute。

[1545] 41.实施方案30至40中任一项的系统,其中所述napDNAbp具有切口酶活性。

[1546] 42.实施方案30至41中任一项的系统,其中所述第一和/或第二napDNAbp包含SEQ ID NO:18-88中任何一个的氨基酸序列,或与SEQ ID NO:18-88中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1547] 43.实施方案30至42中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽是逆转录酶。

[1548] 44.实施方案30至43中任一项的系统,其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159和700-766中任何一个的氨基酸序列,或与SEQ ID NO:89-100、105-122、128-129和132中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1549] 45.实施方案30至44中任一项的系统,其中所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽;和/或所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽。

[1550] 46.实施方案45的系统,其中所述接头包含SEQ ID NO:166-177中任何一个的氨基酸序列,或与SEQ ID NO:166-177中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1551] 47.实施方案45的系统,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、

13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

[1552] 48.实施方案30至47中任一项的系统，其中所述第一和/或第二pegRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列，或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

[1553] 49.实施方案30至48中任一项的系统，其中所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

[1554] 50.实施方案35的系统，其中所述第二引导编辑器复合物的所述间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

[1555] 51.实施方案35的系统，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被所述第一和第二引导编辑器复合物中每一个的所述第一和/或第二napDNAbp生成切口。

[1556] 52.实施方案30的系统，其中所述第一和第二结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

[1557] 53.实施方案52的系统，其中所述第一和第二结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1558] 54.实施方案52的系统，其中所述连续区域被包含所述编辑部分的双链体置换。

[1559] 55.实施方案30至54中任一项的系统，其中所述第一和/或第二DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1560] 56.实施方案30至55中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有相同的长度。

[1561] 57.实施方案30至56中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列具有不同的长度。

[1562] 58.实施方案30至57中任一项的系统，其中所述互补区域为所述第一单链DNA序列或所述第二单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

[1563] 59.实施方案30至58中任一项的系统，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷

酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1564] 60.实施方案30至59中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域形成双链体。

[1565] 61.实施方案30至60中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列各自是具有3’末端的单链DNA链。

[1566] 62.实施方案54的系统，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

[1567] 63.多核苷酸，其编码实施方案1至62中任一项的系统。

[1568] 64.多核苷酸，其编码实施方案1至62中任一项的第一和第二引导编辑器。

[1569] 65.多核苷酸，其编码实施方案1至62中任一项的第一和/或第二pegRNA。

[1570] 66.多核苷酸，其编码实施方案30至62中任一项的第一和/或第二pegRNA。

[1571] 67.多核苷酸，其编码实施方案1至29中任一项的第一和第二pegRNA。

[1572] 68.多核苷酸，其编码实施方案30至62中任一项的第一和第二pegRNA。

[1573] 69.载体，其包含实施方案63至68中任一项的多核苷酸。

[1574] 70.细胞，其包含实施方案63至68中任一项的多核苷酸或实施方案69的载体。

[1575] 71.药物组合物，其包含实施方案63至68中任一项的多核苷酸、实施方案69的载体或实施方案70的细胞以及药物赋形剂。

[1576] 72.试剂盒，其包含实施方案63至68中任一项的多核苷酸或实施方案69的载体以及用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的使用说明。

[1577] 73.用于在靶位点处同时编辑双链DNA序列的第一和第二互补链的方法，所述方法包括将所述双链DNA序列与引导编辑器复合物对接触，所述对包含：

[1578] a.第一引导编辑器复合物，其包含：

[1579] i.第一引导编辑器，其包含第一核酸可编程DNA结合蛋白(napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽；和

[1580] ii.第一引导编辑向导RNA(第一PEgRNA)，其结合至所述靶位点上游的基因组DNA

序列的第一链上的第一结合位点；

[1581] b.第二引导编辑器复合物，其包含：

[1582] i.第二引导编辑器，其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽；和

[1583] ii.第二引导编辑向导RNA(第二PEgRNA)，其结合至所述靶位点下游的基因组DNA序列的第二链上的第二结合位点；

[1584] 其中所述第一引导编辑器复合物导致与所述第一结合位点互补的序列处的第一切口以及随后从由所述第一切口形成的可用5'末端聚合具有3'末端的第一单链DNA序列；

[1585] 其中所述第二引导编辑器复合物导致与所述第二结合位点互补的序列处的第二切口以及随后从由所述第二切口形成的可用5'末端聚合具有3'末端的第二单链DNA序列；

[1586] 其中所述第一单链DNA序列和所述第二单链DNA序列至少在互补区域上是反向互补物并且形成包含编辑的双链体；并且

[1587] 其中所述双链体置换所述双链DNA序列的带切口的第一和第二互补链。

[1588] 74.实施方案73的方法，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白，所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

[1589] 75.实施方案73的方法，其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含相同的引导编辑器。

[1590] 76.实施方案73的方法，其中所述第一引导编辑器复合物和所述第二引导编辑器复合物包含不同的引导编辑器。

[1591] 77.实施方案73至76中任一项的方法，其中所述第一和/或第二napDNAbp是Cas9结构域或其变体。

[1592] 78.实施方案73至77中任一项的方法，其中所述第一和/或第二napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域、Cas9切口酶结构域或其变体。

[1593] 79.实施方案73至78中任一项的方法，其中所述第一和/或第二napDNAbp选自由以下各项组成的组：Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c、CasX、CasY和Argonaute。

[1594] 80.实施方案73至79中任一项的方法，其中所述第一和/或第二napDNAbp具有切口酶活性。

[1595] 81.实施方案73至80中任一项的方法，其中所述第一和/或第二napDNAbp包含SEQ ID NO:18-88中任何一个的氨基酸序列，或与SEQ ID NO:18-88中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1596] 82.实施方案73至81中任一项的方法，其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽是逆转录酶。

[1597] 83.实施方案73至82中任一项的方法，其中所述具有RNA依赖性DNA聚合酶活性的第一和/或第二多肽包含SEQ ID NO:89-100、105-122、128-129、132、139、143、149、154、159和700-766中任何一个的氨基酸序列，或与SEQ ID NO:89-100、105-122、128-129和132中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1598] 84.实施方案73至83中任一项的方法,其中所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽,和/或所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽。

[1599] 85.实施方案84的方法,其中所述接头包含SEQ ID NO:166-177中任何一个的氨基酸序列,或与SEQ ID NO:166-177中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1600] 86.实施方案84或85的方法,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

[1601] 87.实施方案73至86中任一项的方法,其中所述第一和/或第二pegRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列,或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80%、85%、90%、95%或99%序列同一性的核苷酸序列。

[1602] 88.实施方案73至87中任一项的方法,其中所述第一和/或第二PEgRNA包含SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个的核苷酸序列,或者与SEQ ID NO:325-330、499-505、101-104、181-183和223-244中任何一个具有至少80%、85%、90%、95%或99%序列同一性的核苷酸序列。

[1603] 89.实施方案73至88中任一项的方法,其中所述第一引导编辑器复合物的所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

[1604] 90.实施方案73至89中任一项的方法,其中所述第二引导编辑器复合物的所述第二间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

[1605] 91.实施方案90的方法,其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述第一和第二间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第一和第二引导编辑器复合物各自的napDNAbp生成切口。

[1606] 92.实施方案91的方法,其中所述第一和第二结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

[1607] 93.实施方案92的方法,其中所述第一和第二结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对,至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至

少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1608] 94.实施方案92或93的方法，其中所述连续区域被包含所述编辑部分的所述双链体置换。

[1609] 95.实施方案73至94中任一项的方法，其中所述第一和/或第二DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1610] 96.实施方案73至95中任一项的方法，其中所述第一单链DNA序列和所述第二单链

DNA序列具有相同的长度。

[1611] 97.实施方案73至96中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列具有不同的长度。

[1612] 98.实施方案73至97中任一项的方法，其中所述互补区域为所述第一单链DNA序列或所述第二单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

[1613] 99.实施方案73至98中任一项的方法，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸，并且小于或等于所述期望的DNA序列及其互补物的长度。

[1614] 100.实施方案73至99中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域形成所述双链体。

[1615] 101.实施方案73至100中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列各自是具有3’末端的单链DNA链。

[1616] 102.实施方案94的方法，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

[1617] 103.实施方案73至102中任一项的方法，其中所述双链体包含编辑，所述编辑包含一个或多个重组酶位点。

[1618] 104.实施方案103的方法，其进一步包括提供重组酶以促进所述待编辑的靶位点

处的所述DNA序列的重组酶介导的编辑。

[1619] 105.pegRNA对,其用于双重引导编辑,其中所述对包含:

[1620] a.第一引导编辑向导RNA(第一PEgRNA),其结合至待编辑的靶位点上游的双链DNA序列的第一链上的第一结合位点,其中所述第一PEgRNA包含第一间隔区序列、第一gRNA核心、第一DNA合成模板和第一引物结合位点,其中所述第一DNA合成模板编码包含编辑部位的第一单链DNA序列;

[1621] b.第二引导编辑向导RNA(第二PEgRNA),其结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点,其中所述第二PEgRNA包含第二间隔区序列、第二gRNA核心、第二DNA合成模板和第二引物结合位点,其中所述第二DNA合成模板编码包含所述编辑部分的第二单链DNA序列,其中所述第二单链DNA序列包含所述第一单链DNA序列的互补区域。

[1622] 106.实施方案105的pegRNA对,其中所述第一PEgRNA包含以下结构:

[1623] 5'-[第一gRNA间隔区序列]-[第一gRNA核心]-[第一DNA合成模板]-[第一引物结合位点]-3';且

[1624] 其中所述第二PEgRNA包含以下结构:

[1625] 5'-[第二gRNA间隔区序列]-[第二gRNA核心]-[第二DNA合成模板]-[第二引物结合位点]-3'。

[1626] 107.实施方案105或106的pegRNA对,其中所述第一单链DNA序列和所述第二单链DNA序列形成双链体,所述双链体在所述待编辑的靶位点处整合到所述双链DNA序列中。

[1627] 108.实施方案107的pegRNA对,其中所述双链体的所述整合导致所述靶位点处的DNA的插入、缺失或替代。

[1628] 109.实施方案108的pegRNA对,其中所述插入的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷

酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1629] 110.实施方案108的pegRNA对，其中所述缺失的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

[1630] 111.实施方案108的pegRNA对，其中所述替代的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷

酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

[1631] 112.多核苷酸,其编码实施方案105至111中任一项的pegRNA对。

[1632] 113.包含实施方案112的多核苷酸的载体,其用于在细胞中表达所述pegRNA对。

[1633] 114.细胞,其包含实施方案113的载体。

[1634] 115.药物组合物,其包含实施方案105至111中任一项的pegRNA对、实施方案113的载体或实施方案114的细胞以及药物赋形剂。

[1635] 116.引导编辑器复合物对,所述对包含:

[1636] a.第一引导编辑器复合物,其包含:

[1637] i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

[1638] ii.第一引导编辑向导RNA(第一PEgRNA),其结合至靶位点上游的基因组DNA序列的链上的第一靶序列;

[1639] b.第二引导编辑器复合物,其包含:

[1640] i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

[1641] ii.第二引导编辑向导RNA(第二PEgRNA),其结合至靶位点下游的基因组DNA序列的第二链上的第二结合位点。

[1642] 117.实施方案116的引导编辑器复合物对,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物或者所述第一引导编辑器复合物和所述第二引导编辑器复合物二者的所述引导编辑器是融合蛋白,所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

[1643] 118.多核苷酸,其编码实施方案116或117的引导编辑器复合物对。

[1644] 119.包含实施方案118的多核苷酸的载体,其用于在细胞中表达所述引导编辑器复合物对。

[1645] 120.细胞,其包含实施方案119的载体。

[1646] 121.药物组合物,其包含实施方案116或117的引导编辑器复合物对、实施方案119的载体或实施方案120的细胞以及药物赋形剂。

[1647] 122.实施方案1至62中任一项的系统或实施方案73至104中任一项的方法,其中包

含所述编辑部分的所述双链体的整合导致所述靶位点处的DNA的插入、缺失或替代。

[1648] 123.实施方案122的系统或方法,其中所述插入是肽标签。

[1649] 124.实施方案123的系统或方法,其中所述肽标签是多组氨酸(例如,HHHHHH)(SEQ ID NO:252-262)、FLAG(例如,DYKDDDDK)(SEQ ID NO:2)、V5(例如,GKPIPNPLLGLDST)(SEQ ID NO:3)、GCN4、HA(例如,YPYDVPDYA)(SEQ ID NO:5)、Myc(例如EQKLISEED)(SEQ ID NO:6)或GST。

[1650] 125.实施方案123的系统或方法,其中所述肽标签具有选自由以下各项组成的组的氨基酸序列:SEQ ID NO:1-6、245-249、252-262、264-273、275-276、281、278-288和622。

[1651] 126.实施方案122的系统或方法,其中所述插入是免疫表位标签。

[1652] 127.实施方案126的系统或方法,其中所述免疫表位标签选自由以下各项组成的组:破伤风类毒素(SEQ ID NO:396);白喉毒素突变体CRM197(SEQ ID NO:630);腮腺炎免疫表位1(SEQ ID NO:400);腮腺炎免疫表位2(SEQ ID NO:402);腮腺炎免疫表位3(SEQ ID NO:404);风疹病毒(SEQ ID NO:406);血凝素(SEQ ID NO:408);神经氨酸酶(SEQ ID NO:410);TAP1(SEQ ID NO:412);TAP2(SEQ ID NO:414);血凝素表位,针对HLAI类(SEQ ID NO:416);神经氨酸酶表位,针对HLAI类(SEQ ID NO:418);血凝素表位,针对HLAII类(SEQ ID NO:420);神经氨酸酶表位,针对HLAII类(SEQ ID NO:422);血凝素表位,H5N1结合的HLAI类和II类(SEQ ID NO:424);和神经氨酸酶表位,H5N1结合的HLAI类和II类(SEQ ID NO:426)。

[1653] 128.实施方案122的系统或方法,其中所述插入是二聚化结构域。

[1654] 129.实施方案128的系统或方法,其中所述二聚化结构域是SEQ ID NO:488的FKBP12、SEQ ID NO:489的FKBP12-F36V或SEQ ID NO:490或493-494的亲环蛋白的小分子结合结构域。

[1655] 130.用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的系统,所述系统包含:

[1656] a)第一引导编辑器复合物,其包含:

[1657] i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

[1658] ii.第一引导编辑向导RNA(第一PEgRNA),其结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点;

[1659] b)第二引导编辑器复合物,其包含:

[1660] i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

[1661] ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述待编辑的靶位点上游的所述双链DNA序列的第二链上的第二结合位点;

[1662] c)第三引导编辑器复合物,其包含:

[1663] i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

[1664] ii.第三引导编辑向导RNA(第三PEgRNA),其结合至所述待编辑的靶位点下游的所述双链DNA序列的所述第一链上的第三结合位点;

[1665] d)第四引导编辑器复合物,其包含:

[1666] i.第四引导编辑器,其包含第四核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

[1667] ii.第四引导编辑向导RNA(第四PEgRNA),其结合至所述待编辑的靶位点下游的所述双链DNA序列的所述第二链上的第四结合位点;

[1668] 其中所述第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板,所述第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板,所述第三PEgRNA包含编码第三单链DNA序列的第三DNA合成模板,且所述第四PEgRNA包含编码第四单链DNA序列的第四DNA合成模板;

[1669] 其中所述第一和所述第三单链DNA序列各自包含彼此互补的区域;并且其中,其中所述第二和第四单链DNA序列各自包含彼此互补的区域。

[1670] 131.用于编辑一个或多个双链DNA序列的系统,所述系统包含:

[1671] a)第一引导编辑器复合物,其包含:

[1672] i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

[1673] ii.第一引导编辑向导RNA(第一PEgRNA),其结合至第一待编辑的靶位点处的第一双链DNA序列的第一链上的第一结合位点;

[1674] b)第二引导编辑器复合物,其包含:

[1675] i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

[1676] ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述第一待编辑的靶位点处的所述第一双链DNA序列的第二链上的第二结合位点;

[1677] c)第三引导编辑器复合物,其包含:

[1678] i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

[1679] ii.第三引导编辑向导RNA(第三PEgRNA),其结合至第二待编辑的靶位点处的第二双链DNA序列的第一链上的第一结合位点;

[1680] d)第四引导编辑器复合物,其包含:

[1681] i.第四引导编辑器,其包含第四核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

[1682] ii.第四引导编辑向导RNA(第四PEgRNA),其结合至所述第二待编辑的靶位点处的所述第二双链DNA序列的第二链上的第二结合位点;

[1683] 其中所述第一PEgRNA包含编码第一单链DNA序列的第一DNA合成模板,所述第二PEgRNA包含编码第二单链DNA序列的第二DNA合成模板,所述第三PEgRNA包含编码第三单链DNA序列的第三DNA合成模板,且所述第四PEgRNA包含编码第四单链DNA序列的第四DNA合成模板;

[1684] 其中所述第一和所述第三单链DNA序列各自包含彼此互补的区域;并且其中,其中所述第二和第四单链DNA序列各自包含彼此互补的区域。

[1685] 132.实施方案131的系统,其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和/或所述第四引导编辑器复合物的引导编辑器是融

合蛋白，所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

[1686] 133.实施方案130至132中任一项的系统，其中

[1687] (i)所述第一PEgRNA进一步包含第一间隔区序列、第一gRNA核心和第一引物结合位点；

[1688] (ii)所述第二PEgRNA进一步包含第二间隔区序列、第二gRNA核心和第二引物结合位点；

[1689] (iii)所述第三PEgRNA进一步包含第三间隔区序列、第三gRNA核心和第三引物结合位点；且

[1690] (iv)所述第四PEgRNA包含第四间隔区序列、第四gRNA核心、第四DNA合成模板和第四引物结合位点。

[1691] 134.实施方案131的系统，其中所述第一和第三单链DNA序列的所述互补区域和/或所述第二和第四单链DNA序列的所述互补区域形成双链体。

[1692] 135.实施方案134的系统，其中所述第一间隔区序列结合至所述第一待编辑的靶位点上游的所述第一双链DNA序列的所述第一结合位点，并且所述第二间隔区序列结合至所述第一待编辑的靶位点下游的所述第一双链DNA序列上的所述第二结合位点。

[1693] 136.实施方案134或135的系统，其中所述第三间隔区序列结合至所述第二待编辑的靶位点上游的所述第二双链DNA序列的所述第一结合位点，并且所述第四间隔区序列结合至所述第二待编辑的靶位点下游的所述第二双链DNA序列上的所述第二结合位点。

[1694] 137.实施方案134至136中任一项的系统，其中所述第一和第二结合位点定义了所述第一双链DNA的第一连续区域的两端。

[1695] 138.实施方案134至137中任一项的系统，其中所述第三和第四结合位点定义了所述第二双链DNA的第二连续区域的两端。

[1696] 139.实施方案134的系统，其中所述第一和第二结合位点之间或所述第三和第四结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷

酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1697] 140.实施方案134至139中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列的所述互补区域和/或所述第二单链DNA序列和所述第四单链DNA序列的所述互补区域形成双链体。

[1698] 141.实施方案140的系统，其中所述双链体掺入所述第一双链DNA和/或所述第二双链DNA中。

[1699] 142.实施方案134至141中任一项的系统，其中所述系统导致所述第二连续区域置换所述第一连续区域进入所述第一双链DNA中。

[1700] 143.实施方案134至142中任一项的系统，其中所述系统导致在所述第一双链DNA和所述第二双链DNA之间交换所述第一连续区域和所述第二连续区域。

[1701] 144.实施方案134至143中任一项的系统，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和所述第四引导编辑器复合物包含相同的引导编辑器。

[1702] 145.实施方案134至144中任一项的系统，其中所述包含RNA依赖性DNA聚合酶活性的第一、第二、第三和第四多肽是相同的。

[1703] 146.实施方案134至145中任一项的系统，其中所述第一、第二、第三和/或第四napDNAbp是Cas9结构域或其变体。

[1704] 147.实施方案134至146中任一项的系统，其中所述第一、第二、第三和/或第四napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域或Cas9切口酶结构域或其变体。

[1705] 148.实施方案134至147中任一项的系统，其中所述第一、第二、第三和/或第四napDNAbp选自由以下各项组成的组：Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c和Argonaute。

[1706] 149.实施方案134至148中任一项的系统，其中所述第一、第二、第三和/或第四napDNAbp具有切口酶活性。

[1707] 150.实施方案134至149中任一项的系统，其中所述第一、第二、第三和/或第四napDNAbp包含SEQ ID NO:2-65中任何一个的氨基酸序列，或与SEQ ID NO:2-65中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1708] 151.实施方案134至150中任一项的系统，其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽是逆转录酶。

[1709] 152.实施方案134至151中任一项的系统，其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽包含SEQ ID NO:37、68-79、82-98、81、98和110中任何一个的氨基酸序列，或者与SEQ ID NO:37、68-79、82-98、81、98和110中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1710] 153.实施方案134至152中任一项的系统,其中

[1711] (i)所述第一引导编辑器进一步包含接头,所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽;

[1712] (ii)所述第二引导编辑器进一步包含接头,所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽;

[1713] (iii)所述第三引导编辑器进一步包含接头,所述接头接合所述第三napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第三多肽;和/或

[1714] (iv)所述第四引导编辑器进一步包含接头,所述接头接合所述第四napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第四多肽。

[1715] 154.实施方案153的系统,其中所述接头包含SEQ ID NO:119-128中任何一个的氨基酸序列,或与SEQ ID NO:119-128中任何一个具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。

[1716] 155.实施方案153或154的系统,其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

[1717] 156.实施方案134至155中任一项的系统,其中所述第一、第二、第三和/或第四PEgRNA包含SEQ ID NO:192-203中任何一个的核苷酸序列,或与SEQ ID NO:192-203中任何一个具有至少80%、85%、90%、95%或99%序列同一性的核苷酸序列。

[1718] 157.实施方案134至156中任一项的系统,其中所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

[1719] 158.实施方案134至157中任一项的系统,其中所述第二引导编辑器复合物的所述第二间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第二链上的第二结合位点。

[1720] 159.实施方案134至158中任一项的系统,其中所述第三间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第一链上的第三结合位点。

[1721] 160.实施方案134至159中任一项的系统,其中所述第四引导编辑器复合物的所述第四间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第四结合位点。

[1722] 161.实施方案160的系统,其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第一和第二引导编辑器复合物各自的所述第一和/或第二napDNAbp生成切口。

[1723] 162.实施方案161的系统,其中在PAM序列存在下所述第三和第四引导编辑器复合物的所述间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第三和第四引导编辑器复合物各自的所述第三和/或第四napDNAbp生成切口。

[1724] 163.实施方案133的系统,其中所述第一和第三结合位点或所述第二和第四结合位点定义了双链DNA的连续区域的两端。

[1725] 164.实施方案163的系统,其中所述第一和第三结合位点之间或所述第二和第四结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、

至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1726] 165.实施方案130至164中任一项的系统，其中所述第一单链DNA序列和所述第三单链DNA序列的所述互补区域和/或所述第二单链DNA序列和所述第四单链DNA序列的所述互补区域形成双链体。

[1727] 166.实施方案164的系统，其中所述连续区域被所述双链体置换。

[1728] 167.实施方案166的系统，其中所述双链体在所述靶位点处掺入所述双链DNA中。

[1729] 168.实施方案122至167中任一项的系统，其中所述第一、第二、第三和/或第四DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷

酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1730] 169.实施方案122至168中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有相同的长度。

[1731] 170.实施方案122至169中任一项的系统，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有不同的长度。

[1732] 171.实施方案122至170中任一项的系统，其中所述第一、第二、第三和/或第四单链DNA序列的所述互补区域为所述第一单链DNA序列和所述第三单链DNA序列或所述第二单链DNA序列和所述第四单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

[1733] 172.实施方案122至171中任一项的系统，其中所述第一、第二、第三和/或第四单链DNA序列的所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷

酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸，并且小于或等于所述第一单链DNA序列和所述第三单链DNA序列以及所述第二单链DNA序列和所述第四单链DNA序列的长度。

[1734] 173.实施方案130至172中任一项的系统，其中所述第一、第二、第三和第四单链DNA序列各自是具有3’末端的单链DNA链。

[1735] 174.实施方案173的系统，其中包含所述编辑部分的所述双链体的掺入导致所述靶位点处的DNA的插入、缺失或替代。

[1736] 175.多核苷酸，其编码实施方案130至174中任一项的第一、第二、第三和第四引导编辑器。

[1737] 176.多核苷酸，其编码实施方案130至175中任一项的第一、第二、第三和第四PEgRNA。

[1738] 177.载体，其包含实施方案175或176的多核苷酸。

[1739] 178.细胞，其包含实施方案175或176的多核苷酸或实施方案177的载体。

[1740] 179.药物组合物，其包含实施方案175或176的多核苷酸、实施方案177的载体或实施方案178的细胞以及药物赋形剂。

[1741] 180.试剂盒，其包含实施方案175或176的多核苷酸或实施方案177的载体以及用于在待编辑的靶位点处同时编辑双链DNA序列的两条链的使用说明。

[1742] 181.用于在靶位点处同时编辑双链DNA序列的第一和第二互补链的方法，所述方法包括将所述双链DNA序列与引导编辑器复合物对接触，所述对包含：

[1743] (a)第一引导编辑器复合物，其包含：

[1744] i.第一引导编辑器，其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽；和

[1745] ii.第一引导编辑向导RNA(第一PEgRNA)，其结合至所述靶位点上游的基因组DNA序列的第一链上的第一靶序列；

[1746] (b)第二引导编辑器复合物，其包含：

[1747] i.第二引导编辑器，其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽；和

[1748] ii.第二引导编辑向导RNA(第二PEgRNA)，其结合至所述靶位点上游的基因组DNA序列的第二链上的第二结合位点；

[1749] (c)第三引导编辑器复合物，其包含：

[1750] i.第三引导编辑器，其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽；和

[1751] ii.第三引导编辑向导RNA(第三PEgRNA)，其结合至所述靶位点下游的基因组DNA序列的第一链上的第三靶序列；

[1752] (d)第四引导编辑器复合物，其包含：

[1753] i.第四引导编辑器，其包含第二核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽；和

[1754] ii.第四引导编辑向导RNA(第四PEgRNA)，其结合至所述靶位点下游的基因组DNA

序列的第二链上的第四结合位点；

[1755] 其中所述第一引导编辑器复合物导致第一靶序列处的第一切口以及随后从由所述第一切口形成的可用5’末端聚合具有3’末端的第一单链DNA序列；

[1756] 其中所述第二引导编辑器复合物导致第二靶序列处的第二切口以及随后从由所述第二切口形成的可用5’末端聚合具有3’末端的第二单链DNA序列；

[1757] 其中所述第三引导编辑器复合物导致第三靶序列处的第三切口以及随后从由所述第三切口形成的可用5’末端聚合具有3’末端的第三单链DNA序列；

[1758] 其中所述第四引导编辑器复合物导致第四靶序列处的第四切口以及随后从由所述第四切口形成的可用5’末端聚合具有3’末端的第四单链DNA序列；

[1759] 其中所述第一单链DNA序列和所述第二单链DNA序列至少在互补区域上是反向互补物并且形成双链体，其中所述双链体置换所述双链DNA序列的带切口的第一和第二互补链；并且

[1760] 其中所述第三单链DNA序列和所述第四单链DNA序列至少在互补区域上是反向互补物并且形成双链体，其中所述双链体置换所述双链DNA序列的带切口的第一和第二互补链。

[1761] 182.实施方案181的方法，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和/或所述第四引导编辑器复合物的引导编辑器是融合蛋白，所述融合蛋白包含所述napDNAbp和所述具有RNA依赖性DNA聚合酶活性的多肽。

[1762] 183.实施方案181或182的方法，其中所述第一引导编辑器复合物、所述第二引导编辑器复合物、所述第三引导编辑器复合物和所述第四引导编辑器复合物包含相同的引导编辑器。

[1763] 184.实施方案181至183中任一项的方法，其中所述第一、第二、第三和/或第四引导编辑器复合物包含不同的引导编辑器。

[1764] 185.实施方案181至184中任一项的方法，其中第一、第二、第三和/或第四napDNAbp是Cas9结构域或其变体。

[1765] 186.实施方案181至185中任一项的方法，其中第一、第二、第三和/或第四napDNAbp是核酸酶活性Cas9结构域、无核酸酶活性Cas9结构域或Cas9切口酶结构域其变体。

[1766] 187.实施方案181至184中任一项的方法，其中所述第一、第二、第三和/或第四napDNAbp选自由以下各项组成的组：Cas9、Cas12e、Cas12d、Cas12a、Cas12b1、Cas13a、Cas12c。

[1767] 188.实施方案181至187中任一项的方法，其中所述第一、第二、第三和/或第四napDNAbp具有切口酶活性。

[1768] 189.实施方案181至188中任一项的方法，其中所述第一、第二、第三和/或第四napDNAbp包含SEQ ID NO:2-65中任何一个的氨基酸序列，或与SEQ ID NO:2-65中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1769] 190.实施方案181至189中任一项的方法，其中所述具有RNA依赖性DNA聚合酶活性的第一、第二、第三和/或第四多肽是逆转录酶。

[1770] 191.实施方案181至190中任一项的方法，其中所述具有RNA依赖性DNA聚合酶活性

的第一、第二、第三和/或第四多肽包含SEQ ID NO:37、68-79、82-98、81、98和110中任何一个的氨基酸序列，或者与SEQ ID NO:37、68-79、82-98、81、98和110中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1771] 192.实施方案181至191中任一项的方法，其中

[1772] (i)所述第一引导编辑器进一步包含接头，所述接头接合所述第一napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第一多肽；

[1773] (ii)所述第二引导编辑器进一步包含接头，所述接头接合所述第二napDNAbp和所述具有RNA依赖性DNA聚合酶活性的第二多肽；

[1774] (iii)所述第三引导编辑器进一步包含接头，所述接头接合所述第三napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第三多肽；和/或

[1775] (iv)所述第四引导编辑器进一步包含接头，所述接头接合所述第四napDNAbp和所述具有所述RNA依赖性DNA聚合酶活性的第四多肽。

[1776] 193.实施方案192的方法，其中所述接头包含SEQ ID NO:119-128中任何一个的氨基酸序列，或与SEQ ID NO:119-128中任何一个具有至少80％、85％、90％、95％或99％序列同一性的氨基酸序列。

[1777] 194.实施方案192或193的方法，其中所述接头的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、38、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。

[1778] 195.实施方案192至194中任一项的方法，其中所述第一、第二、第三和/或第四PEgRNA包含SEQ ID NO:192-203中任何一个的核苷酸序列，或与SEQ ID NO:192-203中任何一个具有至少80％、85％、90％、95％或99％序列同一性的核苷酸序列。

[1779] 196.实施方案192至195中任一项的方法，其中所述第一引导编辑器复合物的所述第一间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第一结合位点。

[1780] 197.实施方案192至196中任一项的方法，其中所述第二引导编辑器复合物的所述第二间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第二结合位点。

[1781] 198.实施方案192至197中任一项的方法，其中所述第三引导编辑器复合物的所述第三间隔区序列结合至所述待编辑的靶位点上游的所述双链DNA序列的第一链上的第三结合位点。

[1782] 199.实施方案192至198中任一项的方法，其中所述第四引导编辑器复合物的所述第四间隔区序列结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第四结合位点。

[1783] 200.实施方案199的方法，其中在PAM序列存在下所述第一和第二引导编辑器复合物的所述第一和第二间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第一和第二引导编辑器复合物各自的napDNAbp生成切口。

[1784] 201.实施方案200的方法，其中在PAM序列存在下所述第三和第四引导编辑器复合物的所述第三和第四间隔区序列的结合导致所述第一和第二链分别在每条链上靠近所述PAM序列的切口位点处被第三和第四引导编辑器复合物各自的所述napDNAbp生成切口。

[1785] 202.实施方案200或201的方法，其中所述第一和第二结合位点和/或所述第三和第四结合位点定义了包含所述待编辑的靶位点的双链DNA的连续区域的两端。

[1786] 203.实施方案202的方法，其中所述第一和第二结合位点之间和/或所述第三和第四结合位点之间的所述连续区域的长度为至少10个核碱基对、至少11个核碱基对、至少12个核碱基对、至少13个核碱基对、至少14个核碱基对、至少15个核碱基对、至少16个核碱基对、至少17个核碱基对、至少18个核碱基对、至少19个核碱基对、至少20个核碱基对、至少21个核碱基对、至少22个核碱基对、至少23个核碱基对、至少24个核碱基对、至少25个核碱基对、至少26个核碱基对、至少27个核碱基对、至少28个核碱基对、至少29个核碱基对、至少30个核碱基对、至少31个核碱基对、至少32个核碱基对、至少33个核碱基对、至少34个核碱基对、至少35个核碱基对、至少36个核碱基对、至少37个核碱基对、至少37个核碱基对、至少38个核碱基对、至少39个核碱基对或者至少40个核碱基对，至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1787] 204.实施方案202或203的方法，其中所述连续区域被所述双链体置换。

[1788] 205.实施方案181至204中任一项的方法，其中所述第一、第二、第三和/或第四DNA合成模板的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷

酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94核苷酸至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1789] 206.实施方案181至205中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有相同的长度。

[1790] 207.实施方案181至206中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列具有不同的长度。

[1791] 208.实施方案181至207中任一项的方法，其中所述互补区域为所述第一单链DNA序列和所述第二单链DNA序列或所述第三单链DNA序列和所述第四单链DNA序列的总长度的至少99％、至少95％、至少90％、至少85％、至少80％、至少75％、至少70％、至少65％、至少60％、至少55％、至少50％、至少45％、至少40％、至少35％、至少30％、至少25％、至少20％、至少15％或至少10％。

[1792] 209.实施方案181至208中任一项的方法，其中所述互补区域的长度为至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷

酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1793] 210.实施方案188至209中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列的所述互补区域形成所述双链体。

[1794] 211.实施方案188至210中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列各自是具有3’末端的单链DNA链。

[1795] 212.实施方案204的方法，其中所述连续区域被包含所述编辑部分的所述双链体置换导致所述靶位点处的DNA的插入、缺失或替代。

[1796] 213.实施方案181至212中任一项的方法，其中所述第一单链DNA序列和所述第二单链DNA序列位于靶DNA序列的末端，并且其中所述第三单链DNA序列和所述第四单链DNA序列位于同一靶DNA序列的相反末端。

[1797] 214.实施方案213的方法，其中所述靶DNA序列是反向的。

[1798] 215.实施方案181至214中任一项的方法，其进一步包括提供环状DNA供体。

[1799] 216.实施方案215的方法，其中所述第一单链DNA序列和所述第三单链DNA序列位于所述靶DNA序列的相反末端，并且其中所述第二单链DNA序列和所述第四单链DNA序列位于所述环状DNA供体上。

[1800] 217.实施方案216的方法，其中所述第二单链DNA序列和所述第四单链DNA序列之间所述环状DNA供体的部分置换所述第一单链DNA序列和所述第三单链DNA序列之间的所述靶序列。

[1801] 218.实施方案181至217中任一项的方法，其中所述第一单链DNA序列和所述第三单链DNA序列位于第一核酸分子上，并且其中所述第二单链DNA序列和所述第四单链DNA序列位于第二核酸分子上。

[1802] 219.实施方案218的方法，其中所述第一单链DNA序列和所述第三单链DNA序列之间所述第一核酸分子的部分掺入到所述第二核酸分子中，并且其中所述第二单链DNA序列和所述第四单链DNA序列之间所述第二核酸分子的部分掺入到所述第一核酸分子中。

[1803] 220.实施方案218或219的方法，其中所述第一核酸分子是第一染色体，并且所述第二核酸分子是第二染色体。

[1804] 221.PEgRNA组，其用于多flap引导编辑，其中所述PEgRNA包含：

[1805] (a)第一引导编辑向导RNA(第一PEgRNA)，其结合至待编辑的靶位点上游的双链DNA序列的第一链上的第一结合位点，其中所述第一PEgRNA包含第一间隔区序列、第一gRNA核心、第一DNA合成模板和第一引物结合位点，其中所述第一DNA合成模板编码第一单链DNA序列；

[1806] (b)第二引导编辑向导RNA(第二PEgRNA)，其结合至所述待编辑的靶位点上游的所述双链DNA序列的第二链上的第二结合位点，其中所述第二PEgRNA包含第二间隔区序列、第二gRNA核心、第二DNA合成模板和第二引物结合位点，其中所述第二DNA合成模板编码包含所述编辑部分的第二单链DNA序列，其中所述第二单链DNA序列与所述第一单链DNA序列互

补；

[1807] (c)第三引导编辑向导RNA(第三PEgRNA)，其结合至待编辑的靶位点下游的双链DNA序列的第一链上的第三结合位点，其中所述第三PEgRNA包含第三间隔区序列、第三gRNA核心、第三DNA合成模板和第三引物结合位点，其中所述第三DNA合成模板编码第三单链DNA序列；

[1808] (d)第四引导编辑向导RNA(第四PEgRNA)，其结合至所述待编辑的靶位点下游的所述双链DNA序列的第二链上的第四结合位点，其中所述第四PEgRNA包含第四间隔区序列、第四gRNA核心、第四DNA合成模板和第四引物结合位点，其中所述第四DNA合成模板编码第四单链DNA序列，其中所述第四单链DNA序列包含所述第三单链DNA序列的互补区域。

[1809] 222.实施方案221的PEgRNA组，其中

[1810] 所述第一PEgRNA包含以下结构：

[1811] 5’-[第一gRNA间隔区序列]-[第一gRNA核心]-[第一DNA合成模板]-[第一引物结合位点]-3’；

[1812] 所述第二PEgRNA包含以下结构：

[1813] 5’-[第二gRNA间隔区序列]-[第二gRNA核心]-[第二DNA合成模板]-[第二引物结合位点]-3’；

[1814] 所述第三PEgRNA包含以下结构：

[1815] 5’-[第三gRNA间隔区序列]-[第三gRNA核心]-[第三DNA合成模板]-[第三引物结合位点]-3’；且

[1816] 所述第四PEgRNA包含以下结构：

[1817] 5’-[第四gRNA间隔区序列]-[第四gRNA核心]-[第四DNA合成模板]-[第四引物结合位点]-3’。

[1818] 223.实施方案221或222的PEgRNA组，其中所述第一单链DNA序列和所述第二单链DNA序列和/或所述第三单链DNA序列和所述第四单链DNA序列形成双链体，所述双链体在所述待编辑的靶位点处整合到所述双链DNA序列中。

[1819] 224.实施方案221至223中任一项的PEgRNA组，其中所述双链体的所述整合导致所述靶位点处的DNA的插入、缺失或替代。

[1820] 225.实施方案224的PEgRNA组，其中所述插入的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷

酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸。

[1821] 226.实施方案224的PEgRNA组，其中所述缺失的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

[1822] 227.实施方案224的PEgRNA组，其中所述替代的长度为至少1个核苷酸、至少2个核苷酸、至少3个核苷酸、至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核

苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸、至少27个核苷酸、至少28个核苷酸、至少29个核苷酸、至少30个核苷酸、至少31个核苷酸、至少32个核苷酸、至少33个核苷酸、至少34个核苷酸、至少35个核苷酸、至少36个核苷酸、至少37个核苷酸、至少38个核苷酸、至少39个核苷酸、至少40个核苷酸、至少41个核苷酸、至少42个核苷酸、至少43个核苷酸、至少44个核苷酸、至少45个核苷酸、至少46个核苷酸、至少47个核苷酸、至少48个核苷酸、至少49个核苷酸、至少50个核苷酸、至少51个核苷酸、至少52个核苷酸、至少53个核苷酸、至少54个核苷酸、至少55个核苷酸、至少56个核苷酸、至少57个核苷酸、至少58个核苷酸、至少59个核苷酸、至少60个核苷酸、至少61个核苷酸、至少62个核苷酸、至少63个核苷酸、至少64个核苷酸、至少65个核苷酸、至少66个核苷酸、至少67个核苷酸、至少68个核苷酸、至少69个核苷酸、至少70个核苷酸、至少71个核苷酸、至少72个核苷酸、至少73个核苷酸、至少74个核苷酸、至少75个核苷酸、至少76个核苷酸、至少77个核苷酸、至少78个核苷酸、至少79个核苷酸、至少80个核苷酸、至少81个核苷酸、至少82个核苷酸、至少83个核苷酸、至少84个核苷酸、至少85个核苷酸、至少86个核苷酸、至少87个核苷酸、至少88个核苷酸、至少89个核苷酸、至少90个核苷酸、至少91个核苷酸、至少92个核苷酸、至少93个核苷酸、至少94个核苷酸、至少95个核苷酸、至少96个核苷酸、至少97个核苷酸、至少98个核苷酸、至少99个核苷酸、至少100个核苷酸、至少101个核苷酸、至少102个核苷酸、至少103个核苷酸、至少104个核苷酸、至少105个核苷酸、至少106个核苷酸、至少107个核苷酸、至少108个核苷酸、至少109个核苷酸或至少110个核苷酸或更长。

[1823] 228.多核苷酸，其编码实施方案221至227中任一项的PEgRNA组。

[1824] 229.包含实施方案228的多核苷酸的载体，其用于在细胞中表达多个PEgRNA。

[1825] 230.细胞，其包含实施方案229的载体。

[1826] 231.药物组合物，其包含多个实施方案221至227中任一项的PEgRNA、实施方案的载体229或实施方案230的细胞以及药物赋形剂。

[1827] 232.用于修饰细胞中基因组的方法，其包括：

[1828] (i)将基因组与双flap引导编辑系统接触，所述系统包含能够在靶位点的第一链上的第一切口位点处安装第一3’核酸flap的第一引导编辑器/pegRNA复合物和能够在靶位点的第二链上的第二切口位置处安装第二3’核酸flap的第二引导编辑器/pegRNA复合物，

[1829] 其中所述第一3’核酸flap和所述第二3’核酸flap是形成双链体的反向互补序列，并且

[1830] 其中所述双链体包含第一重组酶位点，并且

[1831] 其中所述细胞在所述靶位点处安装包含所述第一重组酶位点的双链体，替代了定位于所述第一和第二切口位点之间的内源性双链体；

[1832] (ii)将所述第一重组酶位点与包含第二重组酶位点的供体核酸接触；和

[1833] (iii)将所述基因组与能够在所述第一和第二重组酶位点之间重组的重组酶接触，从而将所述供体核酸在所述靶位点处整合到所述基因组中。

[1834] [9]方法

[1835] 多瓣引导编辑方法

[1836] 一方面，本公开提供了用于在靶位点处同时编辑双链DNA序列的第一和第二互补链的方法，所述方法包括将所述双链DNA序列与引导编辑器复合物对接触，所述对包含：

[1837] a.第一引导编辑器复合物,其包含:

[1838] i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

[1839] ii.第一引导编辑向导RNA(第一PEgRNA),其结合至所述靶位点上游的基因组DNA序列的第一链上的第一结合位点;

[1840] b.第二引导编辑器复合物,其包含:

[1841] i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

[1842] ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述靶位点上游的基因组DNA序列的第二链上的第二结合位点;

[1843] 其中所述第一引导编辑器复合物导致与所述第一结合位点互补的序列的第一切口以及随后的第一单链DNA序列的聚合,所述第一单链DNA序列具有来自可用的5′端的3′端,所述可用的5′端由所述第一切口形成;

[1844] 其中所述第二引导编辑器复合物导致与所述第二结合位点互补的序列的第二切口以及随后的第二单链DNA序列的聚合,所述第二单链DNA序列具有来自可用的5′端的3′端,所述可用的5′端由所述第二切口形成;

[1845] 其中所述第一单链DNA序列和所述第二单链DNA序列至少在互补区域上是反向互补的并且形成包含编辑的双链体;和

[1846] 其中所述双链体替代了所述双链DNA序列的被切口的第一和第二互补链。

[1847] 一方面,本公开提供了用于在靶位点处同时编辑双链DNA序列的第一和第二互补链的方法,所述方法包括将所述双链DNA序列与一对引导编辑器复合物接触,所述对包含:

[1848] (a)第一引导编辑器复合物,其包含:

[1849] i.第一引导编辑器,其包含第一核酸可编程DNA结合蛋白(第一napDNAbp)和包含RNA依赖性DNA聚合酶活性的第一多肽;和

[1850] ii.第一引导编辑向导RNA(第一PEgRNA),其结合至所述靶位点上游的所述基因组DNA序列的所述第一链上的第一靶序列;

[1851] (b)第二引导编辑器复合物,其包含:

[1852] i.第二引导编辑器,其包含第二核酸可编程DNA结合蛋白(第二napDNAbp)和包含RNA依赖性DNA聚合酶活性的第二多肽;和

[1853] ii.第二引导编辑向导RNA(第二PEgRNA),其结合至所述靶位点上游的基因组DNA序列的所述第二链上的第二结合位点;

[1854] (c)第三引导编辑器复合物,其包含:

[1855] i.第三引导编辑器,其包含第三核酸可编程DNA结合蛋白(第三napDNAbp)和包含RNA依赖性DNA聚合酶活性的第三多肽;和

[1856] ii.第三引导编辑向导RNA(第三PEgRNA),其结合至所述靶位点下游的所述基因组DNA序列的所述第一链上的第三靶序列;

[1857] (d)第四引导编辑器复合物,其包含:

[1858] i.第四引导编辑器,其包含第二核酸可编程DNA结合蛋白(第四napDNAbp)和包含RNA依赖性DNA聚合酶活性的第四多肽;和

[1859] ii.第四引导编辑向导RNA(第四PEgRNA),其结合至所述靶位点下游的基因组DNA序列的所述第二链上的第四结合位点;

[1860] 其中所述第一引导编辑器复合物导致所述第一靶序列的第一切口以及随后的第一单链DNA序列的聚合,所述第一单链DNA序列具有来自可用的5′端的3′端,所述可用的5′端由所述第一切口形成;

[1861] 其中所述第二引导编辑器复合物导致第二靶序列的第二切口以及随后的第二单链DNA序列的聚合,所述第二单链DNA序列具有来自可用的5’端的3’端,所述可用的5’端由所述第二切口形成;

[1862] 其中所述第三引导编辑器复合物导致所述第三靶序列的第三切口以及随后的第三单链DNA序列的聚合,所述第三单链DNA序列具有来自可用的5’端的3’端,所述可用的5’端由所述第三切口形成;

[1863] 其中所述第四引导编辑器复合物导致所述第四靶序列的第四切口以及随后的第四单链DNA序列的聚合,所述第四单链DNA序列具有来自可用的5’端的3’端,所述可用的5’端由所述第四切口形成;

[1864] 其中所述第一单链DNA序列和所述第二单链DNA序列至少在互补区域上是反向互补的并且形成双链体,其中所述双链体替代了所切口的所述双链DNA序列的第一和第二互补链;和

[1865] 其中所述第三单链DNA序列和所述第四单链DNA序列至少在互补区域上是反向互补的并且形成双链体,其中所述双链体替代了所切口的所述双链DNA序列的第一和第二互补链。

[1866] 本说明书描述了多瓣引导编辑系统(包括,例如,双瓣引导编辑系统和识别引导编辑系统),其解决了与瓣平衡和随后通过同时编辑两条DNA链将编辑插入到未编辑的互补基因组DNA链相关的挑战。在双瓣引导编辑系统中,使用两个PEgRNA靶向基因组位点的相对链并且指导含有编辑DNA序列的两个互补3′瓣的合成(图90)。在四瓣引导编辑系统中,使用四个PEgRNA并且指导四个3′瓣的合成,这些3’瓣中的两个彼此互补,并且另外两个彼此互补(图95、图96A和图98A)。与经典引导编辑不同,不需要成对的编辑DNA链(3′瓣)与内源性基因组DNA中的5’瓣直接竞争,因为互补编辑链也可以用于杂交。由于双链体的两条链均作为编辑DNA合成,多瓣引导编辑系统省略了经典引导编辑所需的替换未编辑的互补DNA链的需要。相反,细胞DNA修复机制仅需要切除成对的5′瓣(原始基因组DNA)并且将成对的3′瓣(编辑DNA)连接到基因座中。因此,也不需要在新合成的DNA链中包含与基因组DNA同源的序列,允许新链的选择性杂交,并且有利于含有最小基因组同源性的编辑。还可以使用切割DNA的两条链的核酸酶活性形式的多瓣引导编辑器来促进原始DNA序列的去除。

[1867] 与经典引导编辑一样,多瓣引导编辑是多功能和精确的基因组编辑方法,其使用与核酸酶一起工作的核酸可编程DNA结合蛋白(“napDNAbp”)(即,以融合蛋白的形式或者与napDNAbp反式提供)将新的遗传信息直接写入指定DNA位点,其中引导编辑系统用引导编辑(PE)向导RNA(“PEgRNA”)编程,该向导RNA既指定靶位点,又通过工程化到向导RNA上(例如,在向导RNA的5’或3’端,或者在其内部)的延伸(DNA或RNA)的方式以替换DNA链的形式为期望的编辑的合成提供模板。含有期望的编辑的替换链(例如,单一核酸酶取代)与待编辑的靶位点的内源性链共享相同序列(不同之处在于其包含期望的编辑)。通过DNA修复和/或复

制机制，靶位点的内源性链被替换为新合成的含有期望的编辑的替换链。在一些情况下，引导编辑可以被认为是“检索和替换”基因组编辑技术，因为本文所述的引导编辑器不仅检索和定位待编辑的位点，而且同时还编码含有期望的编辑的替换链，该期望的编辑安装在相应的靶位点内源性DNA链的位置。

[1868] 在双瓣引导编辑中，双链DNA序列在靶位点处与第一和第二引导编辑器复合物接触。每个复合物包含融合蛋白和PEgRNA。在一些实施方案中，每个融合蛋白包含核酸可编程DNA结合蛋白(napDNAbp)和具有RNA依赖性DNA聚合酶活性的多肽(例如，逆转录酶)，并且每个PEgRNA包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。每个DNA合成模板可以编码单链DNA序列，其可以包含编辑一个或多个核苷酸的一部分。编码的两个单链DNA序列可以彼此互补并且形成双链体，该双链体可以整合到待编辑的靶位点中。引导编辑器复合物的不同元件(例如，融合蛋白、napDNAbp、聚合酶、PEgRNA等)可以包含本文公开的系统的任何实施方案。

[1869] 在四瓣引导编辑中，双链DNA序列在靶位点处与第一、第二、第三和第四引导编辑器复合物接触。每个复合物包含融合蛋白和PEgRNA。在一些实施方案中，每个融合蛋白包含核酸可编程DNA结合蛋白(napDNAbp)和具有RNA依赖性DNA聚合酶活性的多肽(例如，逆转录酶)，并且每个PEgRNA包含间隔区序列、gRNA核心、DNA合成模板和引物结合位点。每个DNA合成模板编码单链DNA序列。编码的两个单链DNA序列可以彼此互补并且形成双链体，该双链体可以整合到待编辑的靶位点中。引导编辑器复合物的不同元件(例如，融合蛋白、napDNAbp、聚合酶、PEgRNA等)可以包含本文公开的系统的任何实施方案。

[1870] 本文提供的多瓣引导编辑方法可以用于许多应用。例如，它们可以用于促进靶标DNA序列的倒置。在本申请中，第一PEgRNA的DNA合成模板编码的第一单链DNA序列和第二PEgRNA的DNA合成模板编码的第二单链DNA序列位于靶标DNA序列的相对端，并且第三PEgRNA的DNA合成模板编码的第三单链DNA序列和第四PEgRNA的DNA合成模板编码的第四单链DNA序列位于同一靶标DNA序列的相对端。

[1871] 在一些实施方案中，本文提供的用于多瓣引导编辑的方法还包括提供环状DNA供体，该供体的一部分可以在靶位点处整合到双链核酸中。在本申请中，第一PEgRNA的DNA合成模板编码的第一单链DNA序列和第三PEgRNADNA合成模板编码的第三单链DNA序列位于靶标DNA序列的相对端，并且第二PEgRNA的DNA合成模板编码的第二单链DNA序列和第四PEgRNA的DNA合成模板编码的第四单链DNA序列位于环状DNA供体上。环状DNA供体的介于第二单链DNA序列和第四单链DNA序列之间的部分可以形成双链体，该双链体替换了第一单链DNA序列和第三单链DNA序列之间的靶标序列。

[1872] 在另一应用中，本文提供的用于多瓣引导编辑的方法允许靶标DNA序列从第一核酸分子(例如，第一染色体)易位到第二核酸分子(例如，第二染色体)中。在本申请中，第一PEgRNA的DNA合成模板编码的第一单链DNA序列和第三PEgRNADNA合成模板编码的第三单链DNA序列位于第一和Susan分子行，并且第二PEgRNA的DNA合成模板编码的第二单链DNA序列和第四PEgRNA的DNA合成模板编码的第四单链DNA序列位于第二核酸分子上。第一核酸分子的介于第一单链DNA序列和第三单链DNA序列之间的部分可以整合到第二核酸分子中，并且其中第二核酸分子的介于第二单链DNA序列和第四单链DNA序列之间的部分整合到第一核酸分子中。

[1873] 其他方法

[1874] 本公开提供了用于治疗被诊断患有与点突变或其他突变(如,缺失、插入、倒位、重复等)相关或由其引起的疾病的受试者的方法,这些突变可通过本文提供的多瓣引导编辑系统校正,例如但不限于朊病毒病(如,本文的实施例5)、三核苷酸重复扩增病症(如,本文的实施例3)或CDKL5缺乏病症(CDD)(如,本文的实施例23)。

[1875] 实际上,任何引起疾病的遗传缺陷都可通过使用多瓣引导编辑来修复,包括选择合适的引导编辑器融合蛋白(包括napDNAbp和聚合酶(如,逆转录酶),以及设计合适的PEgRNA以(a)靶向包含编辑位点的适当靶DNA,和(b)提供用于从切口位点的3'端合成单链DNA的模板,该模板包括置换并取代紧靠切口位点下游的内源性链的期望编辑。多瓣引导编辑可用于但不限于(a)对核苷酸序列安装校正突变变化,(b)安装蛋白和RNA标签,(c)在感兴趣的蛋白上安装免疫表位,(d)在蛋白中安装诱导型二聚化结构域,(e)安装或去除序列以改变生物分子的活性,(f)安装重组酶靶位点以引导特定的基因变化,以及(g)通过使用易错RT的靶序列的诱变。

[1876] 治疗病症的方法可包括根据本文所述方法设计合适的PEgRNA和引导编辑器融合蛋白作为前期步骤,其包括可以考虑的许多考虑因素,例如:

[1877] (a)靶序列,即,期望引导编辑器在其中安装一个或多个核碱基修饰的核苷酸序列;

[1878] (b)靶序列中切割位点的位置,即,引导编辑器将诱导单链切口以在切口一侧形成3′端RT引物序列和在切口另一侧形成5′端内源性瓣(其最终被FEN1或其等同物去除并被3'ssDNA瓣取代)。切割位点形成3'端引物序列,在RNA依赖性DNA聚合过程中,其被PE融合蛋白的聚合酶(如,RT酶)延伸,以形成包含期望编辑的3′ssDNA瓣,其接着取代靶序列中的5′内源性DNA瓣。

[1879] (c)可用的PAM序列(包括经典SpCas9 PAM位点,以及由具有扩展或不同PAM特异性的Cas9变体和等同物识别的非经典PAM位点);

[1880] (d)可用的PAM序列之间的间距和PAM链中切割位点的位置;

[1881] (e)待使用的可用的引导编辑器的特定Cas9、Cas9变体或Cas9等同物(部分由可用的PAM支配);

[1882] (f)引物结合位点的序列和长度;

[1883] (g)编辑模板的序列和长度;

[1884] (h)同源臂的序列和长度;

[1885] (i)间隔区序列和长度;以及

[1886] (j)gRNA核心序列。

[1887] 合适的PEgRNA和任选地用于产生第二位点切口的切口产生性sgRNA设计指南,可通过以下示例性分步指令组的方式设计,其考虑了的一个或多个上述考虑。这些步骤参考图70A至图70I中所示的示例。

[1888] 1.定义靶序列和编辑。检索以期望编辑位置(点突变、插入、缺失或其组合)为中心的靶DNA区域(约200bp)的序列。参见图70A。

[1889] 2.定位靶PAM。确定邻近期望编辑位置的PAM。可在邻近期望编辑位置的任一DNA链上确定PAM。虽然编辑位置附近的PAM是优选的(即,其中切口位点距编辑位置小于30nt,或

从编辑位点至切口位点小于29nt、28nt、27nt、26nt、25nt、24nt、23nt、22nt、21nt、20nt、19nt、18nt、17nt、16nt、15nt、14nt、13nt、12nt、11nt、10nt、9nt、8nt、7nt、6nt、5nt、4nt、3nt、或2nt)，则可使用前间隔序列和PAM安装编辑，将切口放置在距编辑位置≥30nt处。参见图70B。

[1890] 3.定位切口位点。对于正在考虑的各PAM，确定相应切口位点以及在哪条链上。对于Sp Cas9 H840A切口酶，切割发生在含PAM的链中NGG PAM的5′的第3个和第4个碱基之间。所有编辑核苷酸必须存在于切口位点的3′，因此适当的PAM必须将切口置于含PAM的链的靶编辑的5′。在下面显示的示例中，有两种可能的PAM。为简单起见，其余步骤将显示仅使用PAM1的PEgRNA设计。参见图70C。

[1891] 4.设计间隔区序列。SpCas9的前间隔序列对应于含PAM的链中NGG PAM的5′的20个核苷酸。有效的Pol III转录起始需要G作为第一个转录的核苷酸。如果前间隔序列的第一个核苷酸是G，则PEgRNA的间隔区序列就是前间隔序列。如果前间隔序列的第一个核苷酸不是G，则PEgRNA的间隔区序列是G，后面为前间隔序列。参见图70D。

[1892] 5.设计引物结合位点(PBS)。使用起始等位基因序列，确定含PAM的链的DNA引物。DNA引物的3′端正好是切口位点上游的核苷酸(即，SpCas9的NGG PAM的5′的第4个碱基)。作为使用PE2和PE3的一般设计原则，包含与DNA引物互补的12至13个核苷酸的PEgRNA引物结合位点(PBS)可用作包含约40-60％GC含量的序列。对于低GC含量的序列，应当测试较长(14至15nt)的PBS。对于较高GC含量的序列，应当测试较短(8至11nt)的PBS。最佳PBS序列应根据经验确定，无论GC含量如何。为设计长度为p的PBS序列，使用起始等位基因序列，取含PAM的链中切口位点的5′前p个核苷酸的反向互补序列。参见图70E。

[1893] 6.设计RT模板(或DNA合成模板)。RT模板(或DNA合成模板，其中聚合酶不是逆转录酶)编码设计的编辑和邻近编辑序列的同源性。在一个实施方案中，这些区域对应于图3D和图3E的DNA合成模板，其中DNA合成模板包括“编辑模板”和“同源臂”。最佳RT模板长度因靶位点而异。对于短程编辑(位置+1至+6)，建议测试短(9至12nt)、中(13至16nt)和长(17至20nt)RT模板。对于远程编辑(+7位及以上)，建议使用RT模板，该模板在编辑位置后至少延伸5nt(最好是10nt或更多)，以允许足够的3′DNA瓣同源性。对于远程编辑，应筛选多个RT模板以识别功能设计。对于较大的插入和缺失(≥5nt)，建议将更大的3′同源性(约20nt或更多)掺入RT模板。当RT模板将G的合成编码为逆转录DNA产物中的最后一个核苷酸(对应于PEgRNA的RT模板中的C)时，通常会降低编辑效率。由于许多RT模板支持有效的引导编辑，因此设计RT模板时建议避免G作为最终合成的核苷酸。为设计长度为r的RT模板序列，使用期望等位基因序列并取原始含PAM的链中切口位点3′前r个核苷酸的反向互补序列。注意，与SNP编辑相比，使用相同长度的RT模板进行插入或缺失编辑不包含相同的同源性。参见图70F。

[1894] 7.组装完整的PEgRNA序列。按以下顺序(5′至3′)连接PEgRNA组件：间隔区、支架、RT模板和PBS。参见图70G。

[1895] 8.为PE3设计切口产生性sgRNA。确定编辑上游和下游的非编辑链的PAM。最佳切口位点高度依赖于基因座，应根据经验确定。一般地，在PEgRNA诱导的切口对面的位点5′放置40至90个核苷酸的切口导致更高的编辑产率和更少的indel。切口产生性sgRNA具有与起始等位基因中的20-nt前间隔序列匹配的间隔区序列，如果前间隔序列不以G开头，则添加5′-

G。参见图70H。

[1896] 9.设计PE3b切口产生性sgRNA。如果PAM存在于互补链中,并且其相应的前间隔序列与靶向编辑的序列重叠,则此编辑可能是PE3b系统的候选者。在PE3b系统中,切口产生性sgRNA的间隔区序列匹配期望编辑等位基因的序列,但不匹配起始等位基因的序列。当编辑核苷酸落在切口产生性sgRNA前间隔序列的种子区(邻近PAM的约10nt)内时,PE3b系统有效运行。这可防止在安装编辑链之前对互补链产生切口,从而防止PEgRNA和sgRNA之间竞争结合靶DNA。PE3b还避免同时对两条链产生切口,从而在保持高编辑效率的同时显著减少indel形成。PE3b sgRNA应当具有与期望等位基因中的20-nt前间隔序列匹配的间隔区序列,并在需要时添加5′G。参见图70I。

[1897] 上述用于设计合适的PEgRNA和第二位点切口产生性sgRNA的分步方法并不意味着以任何方式进行限制。本公开考虑了本领域普通技术人员可从中推导出的上述分步方法的变化。

[1898] 一旦选择/设计了合适的PEgRNA和PE融合蛋白,它们可通过合适的方法施用,例如通过基于载体的转染(其中一个或多个载体包含编码PEgRNA和PE融合蛋白的DNA,且利用载体转染后在细胞内表达)、以递送形式(如,脂质颗粒、纳米颗粒)直接递送与PEgRNA复合的PE融合蛋白(如,RNP递送),或通过基于mRNA的递送系统。本公开在此描述了这类方法并且可利用任何已知方法。

[1899] PEgRNA和PE融合蛋白(或一起被称为PE复合物)可以治疗有效量递送至细胞,使得在接触感兴趣的靶DNA时,将期望编辑安装在其中。

[1900] 可以想象,任何疾病都可通过这类方法治疗,只要递送至合适的细胞是可行的。本领域普通技术人员将能够选择和/或挑选适合预期目的和预期靶细胞的PE递送方法。

[1901] 例如,在一些实施方案中,提供的方法包括向患有这种疾病(如,与上述的点突变相关的癌症)的受试者施用有效量的本文所述多瓣引导编辑系统,在包含期望的基因变化的供体DNA分子的存在下,通过同源定向修复介导,该引导编辑系统校正疾病相关基因中的点突变或向其引入失活突变。在一些实施方案中,提供的方法包括向患有这种疾病(如,与上述点突变相关的癌症)的受试者施用有效量的本文所述多瓣引导编辑系统,该引导编辑系统校正疾病相关基因中的点突变或向其引入失活突变。在一些实施方案中,疾病是增殖性疾病。在一些实施方案中,疾病是遗传性疾病。在一些实施方案中,疾病是肿瘤性疾病。在一些实施方案中,疾病是代谢性疾病。在一些实施方案中,疾病是溶酶体贮积病。可通过校正疾病相关基因中的点突变或向其中引入失活突变来治疗的其他疾病是本领域技术人员已知的,并且本公开在这方面不受限制。

[1902] 本公开提供了用于治疗其他疾病或病症的方法,例如,与可通过TPRT介导的基因编辑校正的点突变相关或由其引起的疾病或病症。本文描述了一些这样的疾病,并且基于本公开可用本文提供的策略和融合蛋白治疗的其他合适的疾病对于本领域技术人员将是显而易见的。下面列出示例性的合适的疾病和病症。应当理解,各个序列中特定位置或残基的编号取决于所使用的具体蛋白和编号方案。例如,成熟蛋白的前体和成熟蛋白本身中的编号可能不同,物种之间的序列差异可能会影响编号。本领域技术人员将能够通过本领域所熟知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白和相应编码核酸中的相应残基。示例性的合适疾病和病症包括但不限于:2-甲基-3-羟基丁酸尿症(2-

methyl-3-hydroxybutyric aciduria);3β-羟基类固醇脱氢酶缺乏症(3beta-Hydroxysteroid dehydrogenase deficiency);3-甲基戊二酸尿症(3-Methylglutaconic aciduria);3-氧-5α-类固醇δ4-脱氢酶缺乏症(3-Oxo-5alpha-steroid delta4-dehydrogenase deficiency);46,XY性反转1、3、5型(46,XYsex reversal,type 1,3,and 5);5-羟脯氨酸酶缺乏症(5-Oxoprolinase deficiency);6-丙酮酰-四氢蝶呤合酶缺乏症(6-pyruvoyl-tetrahydropterin synthase deficiency);Aarskog综合征(Aarskog syndrome);Aase综合征(Aase syndrome);软骨成长不全2型(Achondrogenesis type 2);色盲2和7(Achromatopsia2 and 7);获得性长QT综合征(Acquiredlong QT syndrome);Acrocallosal综合征Schinzel型(Acrocallosal syndrome,Schinzeltype);Acrocapitofemoral发育不良(Acrocapitofemoral dysplasia);肢端骨发育不全2,有或无激素抗性(Acrodysostosis 2,with or withouthormone resistance);肢端红斑角化症(Acroerythrokeratoderma);肩峰发育不良(Acromicric dysplasia);Acth非依赖性肾上腺大结节性增生症2(Acth-independent macronodular adrenal hyperplasia 2);激活PI3K-δ综合征(Activated PI3K-delta syndrome);急性间歇性卟啉症(Acute intermittent porphyria);酰基辅酶A脱氢酶家族成员9缺乏症(deficiency of Acyl-CoA dehydrogenase family,member 9);Adams-Oliver综合征5和6;腺嘌呤磷酸核糖转移酶缺乏症(Adenine phosphoribosyltransferase deficiency);腺苷酸激酶缺乏症(Adenylate kinase deficiency);溶血性贫血,腺苷酸琥珀酸裂解酶缺乏所致(hemolytic anemia due toAdenylosuccinate lyase deficiency);青少年肾消耗病(Adolescent nephronophthisis);肾-肝-胰腺发育不良(Renal-hepatic-pancreatic dysplasia);梅克尔(Meckel)综合征7型;肾上腺脑白质营养不良(Adrenoleukodystrophy);成人交界性大疱性表皮松解症(Adult junctional epidermolysis bullosa);交界性大疱性表皮松解症,局部变异(Epidermolysis bullosa,junctional,localisatavariant);成人神经元蜡样质脂褐质沉积症(Adult neuronal ceroid lipofuscinosis);成人神经元蜡样质脂褐质沉积症(Adultneuronal ceroid lipofuscinosis);成人发作性共济失调伴眼动失用症(Adult onset ataxiawith oculomotor apraxia);ADULT综合征;无纤维蛋白原血症和先天性无纤维蛋白原血症(Afibrinogenemia and congenital Afibrinogenemia);常染色体隐性无丙种球蛋白血症2(autosomal recessive Agammaglobulinemia 2);年龄相关性黄斑变性3、6、11和12(Age-related macular degeneration 3,6,11,and 12);Aicardi Goutieres综合征1、4和5;Chilbain狼疮1;Alagille综合征1和2;亚历山大(Alexander)症;黑尿酸尿症(Alkaptonuria);Allan-Herndon-Dudley综合征;广泛先天性脱发症(Alopecia universalis congenital);阿尔珀斯氏脑病(Alpers encephalopathy);α-1-抗胰蛋白酶缺乏症(Alpha-1-antitrypsin deficiency);常染色体显性、常染色体隐性和X连锁隐性Alport综合征(autosomal dominant,autosomal recessive,and X-linkedrecessiveAlport syndromes);家族性阿尔茨海默病3,伴痉挛性下肢轻瘫和失用症(Alzheimer disease,familial,3,with spastic paraparesis and apraxia);阿尔茨海默病1、3和4型(Alzheimer disease,types,1,3,and4);低钙化型和低成熟型,IIA1釉质发育不全(hypocalcification type and hypomaturation type,IIA1 Amelogenesis imperfecta);氨酰化酶1缺乏症(Aminoacylase 1deficiency);阿米什婴儿癫痫综合征

(Amishinfantile epilepsy syndrome);淀粉样蛋白转甲状腺素蛋白淀粉样变性(Amyloidogenic transthyretin amyloidosis);淀粉样心肌病,转甲状腺素蛋白相关(Amyloid Cardiomyopathy,Transthyretin-related);心肌病(Cardiomyopathy);肌萎缩侧索硬化症1型、6型、15型(有或无额颞叶痴呆)、22型(有或无额颞叶痴呆)和10型(Amyotrophic lateral sclerosis types 1,6,15(with or without frontotemporal dementia),22(with or without frontotemporal dementia),and 10);额颞叶痴呆伴TDP43内含物,TARDBP相关(Frontotemporal dementia withTDP43 inclusions,TARDBP-related);Andermann综合征;AndersenTawil综合征;先天性长QT综合征(Congenital long QT syndrome);非血球型溶血性贫血,G6PD缺乏所致(Anemia,nonspherocytic hemolytic,due to G6PD deficiency);Angelman综合征;严重新生儿脑病伴小头畸形(Severe neonatal-onset encephalopathy with microcephaly);易感性自闭症,X-连锁3(susceptibility to Autism,X-linked 3);遗传性血管病伴肾病、动脉瘤和肌肉痉挛(Angiopathy,hereditary,with nephropathy,aneurysms,and muscle cramps);血管紧张素i-转换酶,良性血清升高(Angiotensin i-converting enzyme,benign serum increase);无虹膜、小脑性共济失调和智力低下(Aniridia,cerebellar ataxia,and mental retardation);无甲症(Anonychia);抗凝血酶III缺乏症(Antithrombin III deficiency);伴有生殖器异常和类固醇生成障碍的Antley-Bixler综合征(Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis);家族性胸主动脉瘤4、6和9(Aortic aneurysm,familial thoracic 4,6,and 9);胸主动脉瘤和主动脉夹层(Thoracic aortic aneurysms and aortic dissections);多系统平滑肌功能障碍综合征(Multisystemic smooth muscle dysfunction syndrome);烟雾病5(Moyamoya disease 5);再生障碍性贫血(Aplastic anemia);表观盐皮质激素过量(Apparent mineralocorticoid excess);精氨酸酶缺乏症(Arginase deficiency);精氨基琥珀酸裂解酶缺乏症(Argininosuccinate lyase deficiency);芳香酶缺乏症(Aromatase deficiency);致心律失常性右心室心肌病5型、8型和10型(Arrhythmogenic right ventricular cardiomyopathy types 5,8,and 10);原发性家族性肥厚型心肌病(Primary familial hypertrophic cardiomyopathy);先天性远端多发性关节弯曲症,X连锁(Arthrogryposis multiplex congenita,distal,X-linked);关节弯曲肾功能不全胆汁淤积综合征(Arthrogryposis renal dysfunction cholestasis syndrome);关节弯曲、肾功能障碍和胆汁淤积2(Arthrogryposis,renal dysfunction,and cholestasis 2);天冬酰胺合成酶缺乏症(Asparagine synthetase deficiency);神经元迁移异常(Abnormality of neuronal migration);共济失调伴维生素E缺乏(Ataxia with vitamin E deficiency);常染色体显性感觉性共济失调(Ataxia,sensory,autosomal dominant);共济失调-毛细血管扩张综合征(Ataxia-telangiectasia syndrome);遗传性癌症易感综合征(Hereditary cancer-predisposing syndrome);转铁蛋白缺乏症(Atransferrinemia);家族性心房颤动11、12、13和16(Atrial fibrillation,familial,11,12,13,and 16);心房间隔缺损2、4和7(有或无房室传导缺陷)(Atrial septal defects 2,4,and 7(with orwithout atrioventricular conduction defects));心房阻滞2(Atrial standstill 2);心房间隔缺损4(Atrioventricular septal defect 4);遗传性眼球萎缩(Atrophia

bulborum hereditaria);ATR-X综合征(ATR-X syndrome);耳髁突综合征2(Auriculocondylar syndrome 2);多系统自身免疫性疾病,婴儿期发病(Autoimmune disease,multisystem,infantile-onset);自身免疫性淋巴增生综合征1a型(Autoimmune lymphoproliferative syndrome,type 1a);常染色体显性遗传性少汗性外胚层发育不良(Autosomal dominant hypohidrotic ectodermal dysplasia);常染色体显性进行性外眼肌麻痹伴线粒体DNA缺失1和3(Autosomal dominant progressive external ophthalmoplegia with mitochondrial DNA deletions 1and3);常染色体显性扭转肌张力障碍4(Autosomal dominant torsion dystonia4);常染色体隐性中央核肌病(Autosomal recessive centronuclear myopathy);常染色体隐性先天性鱼鳞癣1、2、3、4A和4B(Autosomal recessive congenital ichthyosis 1,2,3,4A,and 4B);常染色体隐性皮肤松弛IA型和1B型(Autosomal recessive cutis laxatype IAand 1B);常染色体隐性少汗性外胚层发育不良综合征(Autosomal recessive hypohidrotic ectodermal dysplasia syndrome);外胚层发育不良11b(Ectodermal dysplasia 11b);少汗/毛发/牙齿型,常染色体隐性(hypohidrotic/hair/tooth type,autosomal recessive);常染色体隐性低磷酸血症骨病(Autosomal recessive hypophosphatemic bone disease);Axenfeld-Rieger综合征3型;Bainbridge-Ropers综合征;Bannayan-Riley-Ruvalcaba综合征;PTEN错构瘤肿瘤综合征(PTEN hamartoma tumor syndrome);Baraitser-Winter综合征1和2;Barakat综合征;Bardet-Biedl综合征1、11、16和19;裸淋巴细胞综合征2型,互补群E(Bare lymphocyte syndrome type 2,complementation group E);产前Bartter综合征2型(Bartter syndrome antenatal type 2);Bartter综合征3型、3型伴低钙尿和4型(Bartter syndrome types 3,3with hypocalciuria,and 4);特发性基底核钙化4(Basal ganglia calcification,idiopathic,4);念珠状毛(Beadedhair);良性家族性血尿(Benign familial hematuria);良性家族性新生儿癫痫1和2(Benign familial neonatal seizures 1and 2);良性家族性新生儿癫痫1和/或肌纤维颤搐(Seizures,benign familial neonatal,1,and/or myokymia);癫痫,早期婴儿癫痫性脑病7(Seizures,Early infantile epileptic encephalopathy 7);良性家族性新生儿-婴儿癫痫(Benign familial neonatal-infantile seizures);良性遗传性舞蹈病(Benign hereditary chorea);良性肩胛腓骨肌营养不良伴心肌病(Benign scapuloperoneal muscular dystrophy with cardiomyopathy);Bernard-Soulier综合征A1和A2型(常染色体显性);Bestrophinopathy,常染色体隐性;β地中海贫血(beta Thalassemia);Bethlem肌病和Bethlem肌病2(Bethlem myopathy and Bethlem myopathy 2);Bietti结晶样角膜视网膜营养不良(Bietti crystalline corneoretinal dystrophy);先天性胆汁酸合成障碍2(Bile acid synthesis defect,congenital,2);生物素酶缺乏症(Biotinidase deficiency);BirkBarel智力低下畸形综合征(Birk Barel mental retardation dysmorphism syndrome);睑裂狭小、下垂和倒转型内眦赘皮(Blepharophimosis,ptosis,and epicanthus inversus);Bloom综合征;Borjeson-Forssman-Lehmann综合征;BoucherNeuhauser综合征;短指症A1和A2型(Brachydactyly types A1 andA2);短指症伴高血压(Brachydactyly with hypertension);脑小血管病伴出血(Brain small vessel disease with hemorrhage);支链酮酸脱氢酶激酶缺乏症(Branched-chain ketoacid

dehydrogenase kinase deficiency);Branchiootic综合征2和3;早发型乳腺癌(Breast cancer,early-onset);家族性乳腺癌-卵巢癌1、2和4(Breast-ovarian cancer,familial 1,2,and4);角膜脆弱综合征2(Brittle cornea syndrome 2);Brody肌病;支气管扩张伴有或不伴有升高的汗液氯化物3(Bronchiectasis with or without elevated sweat chloride 3);Brown-Vialetto-Van Laere综合征和Brown-Vialetto-Van Laere综合征2;Brugada综合征;Brugada综合征1;心室颤动(Ventricular fibrillation);阵发性家族性心室颤动(Paroxysmal familial ventricular fibrillation);Brugada综合征和Brugada综合征4;长QT综合征(Long QT syndrome);心脏性猝死(Sudden cardiac death);牛眼样黄斑营养不良(Bull eye macular dystrophy);Stargardt病4;视锥视杆营养不良12(Cone-rod dystrophy 12);大疱性鱼鳞癣样红皮病(Bullous ichthyosiform erythroderma);Burn-Mckeown综合征;家族性念珠菌病2、5、6和8(Candidiasis,familial,2,5,6,and 8);碳水化合物缺乏糖蛋白综合征I型和II型(Carbohydrate-deficient glycoprotein syndrome type I and II);碳酸酐酶VA缺乏症,高氨血症所致(Carbonic anhydrase VA deficiency,hyperammonemia due to);结肠癌(Carcinoma ofcolon);心律失常(Cardiac arrhythmia);长QT综合征LQT1亚型(Long QT syndrome,LQT1 subtype);婴儿致命性心脑肌病,细胞色素c氧化酶缺乏所致(Cardioencephalomyopathy,fatal infantile,due to cytochrome c oxidase deficiency);Cardiofaciocutaneous综合征;心肌病(Cardiomyopathy);Danon病;肥厚型心肌病(Hypertrophic cardiomyopathy);左心室致密化不全心肌病(Left ventricularnoncompaction cardiomyopathy);Carnevale综合征;Carney综合征1型;肉碱酰基肉碱转位酶缺乏症(Carnitine acylcarnitine translocase deficiency);肉碱棕榈酰基转移酶I、II、II(迟发性)和II(婴儿)缺乏症(Carnitine palmitoyltransferase I,II,II(late onset),andII(infantile) deficiency);白内障1、4,常染色体显性,常染色体显性,多种类型,伴小角膜,coppock样,幼年,伴小角膜和糖尿,以及弥漫性非进行性核(Cataract 1,4,autosomal dominant,autosomal dominant,multiple types,with microcornea,coppock-like,juvenile,with microcornea and glucosuria,andnuclear diffuse nonprogressive);儿茶酚胺敏感性多形性室性心动过速(Catecholaminergic polymorphic ventricular tachycardia);尾部退化综合征(Caudal regression syndrome);家族性Cd8缺乏症(Cd8 deficiency,familial);中央轴突症(Central core disease);1、9和16号染色体着丝粒不稳定和免疫缺陷(Centromeric instability of chromosomes 1,9and 16and immunodeficiency);小儿小脑共济失调伴进行性眼外麻痹及小脑共济失调、智力低下和平衡失调综合征2(Cerebellar ataxia infantile with progressive external ophthalmoplegi and Cerebellar ataxia,mental retardation,and dysequilibrium syndrome 2);APP相关脑淀粉样血管病(Cerebral amyloid angiopathy,APP-related);常染色体显性和隐性脑动脉病变伴皮质下梗死和脑白质病(Cerebral autosomal dominant and recessive arteriopathy with subcortical infarcts and leukoencephalopathy);脑海绵状血管瘤2(Cerebral cavernous malformations 2);脑眼面骨骼综合征2(Cerebrooculofacioskeletal syndrome 2);脑-眼-面-骨骼综合征(Cerebro-oculo-facio-skeletal syndrome);脑视网膜微血管病伴钙化和囊肿(Cerebroretinal

microangiopathy with calcifications and cysts);神经元蜡样质脂褐质沉积症2、6、7和10();Ch\xc3\xa9diak-Higashi综合征,成人型Chediak-Higashi综合征(Ceroid lipofuscinosis neuronal 2,6,7,and 10;Ch\xc3\xa9diak-Higashi syndrome,Chediak-Higashi syndrome,adult type);夏-马-图三氏(Charcot-Marie-Tooth)病1B、2B2、2C、2F、2I、2U(轴突)、1C(脱髓鞘)、显性中间体C、隐性中间体A、2A2、4C、4D、4H、IF、IVF和X型;肩腓型脊髓性肌萎缩(Scapuloperoneal spinal muscular atrophy);先天非进行性远端脊髓性肌萎缩(Distal spinal muscular atrophy,congenital nonprogressive);常染色体隐性远端脊髓性肌萎缩5(Spinal muscular atrophy,distal,autosomal recessive,5);CHARGE联合征;儿童磷酸酶过低症(Childhoodhypophosphatasia);成人磷酸酶过低症(Adulthypophosphatasia);胆囊炎(Cholecystitis);进行性家族性肝内胆汁淤积3(Progressive familial intrahepatic cholestasis 3);妊娠期肝内胆汁淤积3(Cholestasis,intrahepatic,of pregnancy 3);胆固烷醇贮积病(Cholestanol storage disease);胆固醇单加氧酶(侧链切割)缺乏症(Cholesterol monooxygenase(side-chain cleaving)deficiency);软骨发育异常Blomstrand型(Chondrodysplasia Blomstrand type);点状软骨发育异常1,X连锁隐性和2X连锁显性(Chondrodysplasia punctata 1,X-linked recessive and 2X-linked dominant);CHOPS综合征;慢性肉芽肿病,常染色体隐性细胞色素b阳性,1型和2型(Chronic granulomatous disease,autosomal recessive cytochrome b-positive,types 1and 2);Chudley-McCullough综合征;原发性纤毛运动障碍7、11、15、20和22(Ciliarydyskinesia,primary,7,11,15,20and22);瓜氨酸血症I型(Citrullinemiatype I);瓜氨酸血症I型和II型(Citrullinemia type I and II);颅锁骨发育不全(Cleidocranial dysostosis);C样综合征(C-like syndrome);Cockayne综合征A型;原发性辅酶Q10缺乏症1、4和7(Coenzyme Q10deficiency,primary 1,4,and 7);CoffinSiris/智力低下(Coffin Siris/Intellectual Disability);Coffin-Lowry综合征;Cohen综合征;冷诱发出汗综合征1(Cold-induced sweating syndrome 1);COLE-CARPENTER综合征2;细胞和体液免疫联合缺陷伴肉芽肿(Combined cellular and humoral immune defects with granulomas);联合d-2-和l-2-羟基戊二酸尿症();结合性丙二酸和甲基丙二酸尿症(Combined d-2-and l-2-hydroxyglutaric aciduria;Combined malonic and methylmalonic aciduria);联合氧化磷酸化缺陷1、3、4、12、15和25(Combined oxidative phosphorylation deficiencies 1,3,4,12,15,and 25);联合部分和完全17-α-羟化酶/17,20-裂解酶缺乏症(Combined partial and complete 17-alpha-hydroxylase/17,20-lyase deficiency);常见变异免疫缺陷9(Common variable immunodeficiency9);c1抑制剂补体成分4部分缺乏症,功能失调所致(Complement component 4,partial deficiency of,due to dysfunctional c1 inhibitor);补体因子B缺乏症(Complement factor B deficiency);锥体细胞单色型色觉(Cone monochromatism);锥杆营养不良2和6(Cone-rod dystrophy 2and 6);锥杆营养不良性釉质发育不全(Cone-rod dystrophy amelogenesis imperfecta);X-连锁先天性肾上腺增生症和先天性肾上腺发育不全(Congenital adrenal hyperplasia and Congenital adrenal hypoplasia,X-linked);先天性无巨核细胞血小板减少症(Congenital amegakaryocytic thrombocytopenia);先天性无虹膜(Congenital aniridia);先天性中

枢性换气不足(Congenital central hypoventilation);先天性巨结肠病3(Hirschsprung disease 3);先天性挛缩性蜘蛛样指综合征(Congenital contractural arachnodactyly);先天性四肢和面部挛缩、肌张力减退和发育迟缓(Congenital contractures ofthe limbs and face,hypotonia,and developmental delay);先天性糖基化障碍1B、1D、1G、1H、1J、1K、1N、1P、2C、2J、2K、IIm型(Congenital disorder ofglycosylation types 1B,1D,1G,1H,1J,1K,1N,1P,2C,2J,2K,IIm);先天性红细胞生成异常性贫血I型和II型(Congenital dyserythropoietic anemia,type I and II);先天性面部外胚层发育不良(Congenital ectodermal dysplasia of face);先天性红细胞生成性卟啉症(Congenital erythropoietic porphyria);先天性泛发性脂肪代谢障碍2型(Congenital generalized lipodystrophy type 2);先天性心脏病多类型2(Congenital heart disease,multiple types,2);先天性心脏病(Congenital heart disease);主动脉弓离断(Interrupted aortic arch);先天性脂肪瘤过度生长、血管畸形和表皮痣(Congenital lipomatous overgrowth,vascular malformations,and epidermal nevi);非小细胞肺癌(Non-small cell lung cancer);卵巢肿瘤(Neoplasm of ovary);非特异性心脏传导缺陷(Cardiac conduction defect,nonspecific);先天性微绒毛萎缩(Congenital microvillous atrophy);先天性肌营养不良症(Congenital muscular dystrophy);先天性肌营养不良症,部分LAMA2缺乏所致(Congenital muscular dystrophy due to partial LAMA2deficiency);先天性肌营养不良-抗肌萎缩相关糖蛋白病伴脑和眼异常A2、A7、A8、A11和A14型(Congenital muscular dystrophy-dystroglycanopathy with brain and eye anomalies,typesA2,A7,A8,A11,andA14);先天性肌营养不良-抗肌萎缩相关糖蛋白病伴智力低下B2、B3、B5和B15型(Congenital muscular dystrophy-dystroglycanopathy with mental retardation,types B2,B3,B5,and B15);先天性肌营养不良-抗肌萎缩相关糖蛋白病无智力低下B5型(Congenital muscular dystrophy-dystroglycanopathy without mental retardation,type B5);先天性肌肥大-脑综合征(Congenital muscular hypertrophy-cerebral syndrome);先天性肌无力综合征,乙酰唑胺反应性(Congenital myasthenic syndrome,acetazolamide-responsive);先天性肌病伴纤维型不均衡(Congenital myopathy with fiber type disproportion);先天性眼球缺损(Congenital ocular coloboma);先天性静止性夜盲症1A、1B、1C、1E、1F、2A型(Congenital stationary night blindness,type 1A,1B,1C,1E,1F,and 2A);粪卟啉症(Coproporphyria);扁平角膜2(Cornea plana2);Fuchs角膜内皮营养不良4(Corneal dystrophy,Fuchs endothelial,4);角膜内皮营养不良2型(Corneal endothelial dystrophytype 2);角膜脆性球形角膜、蓝色巩膜和关节运动过度(Corneal fragility keratoglobus,blue sclerae and joint hypermobility);Cornelia de Lange综合征1和5;常染色体显性冠状动脉病变2(Coronary artery disease,autosomal dominant 2);冠状动脉性心脏病(Coronary heart disease);高α脂蛋白血症2(Hyperalphalipoproteinemia2);复杂型皮质发育不良伴其他脑畸形5和6(Cortical dysplasia,complex,with other brain malformations 5and 6);枕骨皮质畸形(Cortical malformations,occipital);皮质醇结合球蛋白缺陷症(Corticosteroid-binding globulin deficiency);皮质酮甲基氧化酶2型缺乏症(Corticosterone

methyloxidase type 2deficiency);Costello综合征;Cowden综合征1;扁平髋(Coxa plana);常染色体显性颅骨骨干发育不良(Craniodiaphyseal dysplasia,autosomal dominant);颅缝早闭1和4(Craniosynostosis 1and4);颅缝早闭和牙齿异常(Craniosynostosis anddental anomalies);肌酸缺乏症,X连锁(Creatine deficiency, X-linked);Crouzon综合征;隐眼综合征(Cryptophthalmos syndrome);单侧或双侧隐睾症(Cryptorchidism,unilateral or bilateral);Cushing指(趾)关节粘连(Cushing symphalangism);皮肤恶性黑素瘤1(Cutaneous malignantmelanoma 1);皮肤松弛症伴骨营养不良和严重肺、胃肠道和泌尿系统异常(Cutis laxa with osteodystrophy and with severe pulmonary,gastrointestinal,and urinary abnormalities);新生儿暂时性紫绀和非典型肾病(Cyanosis,transient neonatal and atypical nephropathic);囊性纤维化病(Cystic fibrosis);胱氨酸尿症(Cystinuria);细胞色素c氧化酶i缺乏症(Cytochrome c oxidase i deficiency;);细胞色素-c氧化酶缺乏症(Cytochrome-c oxidase deficiency);D-2-羟基戊二酸尿症2(D-2-hydroxyglutaric aciduria 2);节段性Darier病(Darier disease,segmental);耳聋伴复杂发育不全小耳症和小牙症(LAMM)(Deafness with labyrinthine aplasia microtia and microdontia(LAMM));耳聋,常染色体显性3a、4、12、13、15,常染色体显性非综合征型感觉神经17、20和65(Deafness, autosomal dominant 3a,4,12,13,15,autosomal dominant nonsyndromic sensorineural 17,20,and 65);耳聋,常染色体隐性1A、2、3、6、8、9、12、15、16、18b、22、28、31、44、49、63、77、86和89(Deafness,autosomal recessive 1A,2,3,6,8,9,12,15,16,18b, 22,28,31,44,49,63,77,86,and 89);耳蜗性耳聋伴近视和智力低下,无前庭受累,常染色体显性,X-连锁2(Deafness,cochlear,with myopia and intellectual impairment, without vestibular involvement,autosomal dominant,X-linked2);2-甲基丁酰辅酶A脱氢酶缺乏症(Deficiency of 2-methylbutyryl-CoAdehydrogenase);3-羟酰基辅酶A脱氢酶缺乏症(Deficiency of 3-hydroxyacyl-CoA dehydrogenase);α-甘露糖苷酶缺乏症(Deficiency of alpha-mannosidase);芳香族-L-氨基酸脱羧酶缺乏症(Deficiency of aromatic-L-amino-acid decarboxylase);双磷酸甘油酸变位酶缺乏症(Deficiency of bisphosphoglycerate mutase);丁酰辅酶A脱氢酶缺乏症(Deficiency of butyryl-CoA dehydrogenase);铁氧化酶缺乏症(Deficiency of ferroxidase);半乳糖激酶缺乏症(Deficiency of ferroxidase;Deficiency of galactokinase);胍基乙酸甲基转移酶缺乏症(Deficiency ofguanidinoacetate methyltransferase);透明质酸氨基葡糖苷酶缺乏症(Deficiency of hyaluronoglucosaminidase);5-磷酸核糖异构酶缺乏症(Deficiency of ribose-5-phosphate isomerase);类固醇11-β-单加氧酶缺乏症(Deficiency of steroid 11-beta-monooxygenase);UDP葡糖-己糖-1-磷酸尿苷酰转移酶缺乏症(Deficiency ofUDPglucose-hexose-1-phosphate uridylyltransferase);黄嘌呤氧化酶缺乏症(Deficiency ofxanthine oxidase);Dejerine-Sottas病;夏-马-图三氏病ID和IVF型;常染色体显性Dejerine-Sottas综合征;树突状细胞、单核细胞、B淋巴细胞和自然杀伤淋巴细胞缺乏症(Dendritic cell,monocyte,B lymphocyte,and natural killerlymphocyte deficiency);Desbuquois发育不良2;Desbuquois综合征;DFNA2非综合征性听觉损失(DFNA 2Nonsyndromic Hearing Loss);糖尿病和尿崩症伴视神经萎缩和耳

聋(Diabetes mellitus and insipidus with optic atrophy and deafness);糖尿病2型和胰岛素依赖型20(Diabetes mellitus,type 2,and insulin-dependent,20);Diamond-Blackfan贫血1、5、8和10;腹泻3(先天性分泌钠综合征型)和5(伴有先天性簇绒肠病)(Diarrhea 3(secretory sodium,congenital,syndromic)and 5(with tufting enteropathy,congenital));二羧基氨基酸尿症(Dicarboxylic aminoaciduria);弥漫性掌跖角化病Bothnian型(Diffuse palmoplantar keratoderma,Bothnian type);Digitorenocerebral综合征;二氢蝶呤还原酶缺乏症(Dihydropteridine reductase deficiency);扩张型心肌病1A、1AA、1C、1G、1BB、1DD、1FF、1HH、1I、1KK、1N、1S、1Y和3B(Dilated cardiomyopathy 1A,1AA,1C,1G,1BB,1DD,1FF,1HH,1I,1KK,1N,1S,1Y,and 3B);左心室致密化不全3(Left ventricular noncompaction 3);类固醇生成紊乱,细胞色素p450氧化还原酶缺乏所致(Disordered steroidogenesis due to cytochrome p450 oxidoreductase deficiency);远端关节弯曲2B型(Distal arthrogryposis type 2B);远端遗传性运动神经元病2B型(Distal hereditary motor neuronopathy type 2B);远端肌病Markesbery-Griggs型(Distal myopathy Markesbery-Griggs type);远端脊肌萎缩症,X连锁3(Distal spinal muscular atrophy,X-linked 3);双行睫-淋巴水肿综合征(Distichiasis-lymphedema syndrome);显性营养不良型大疱性表皮松解症伴皮肤缺损(Dominant dystrophic epidermolysis bullosa with absence of skin);显性遗传性视神经萎缩(Dominant hereditary optic atrophy);Donnai Barrow综合征;多巴胺β羟化酶缺乏症(Dopamine beta hydroxylase deficiency);多巴胺受体d2,脑密度降低(Dopamine receptord2,reduced brain density of);Dowling-degos病4;Doyne蜂窝状视网膜营养不良(Doyne honeycomb retinal dystrophy);Malattia leventinese;Duane综合征2型;Dubin-Johnson综合征;杜氏肌营养不良症(Duchenne muscular dystrophy);贝克肌营养不良(Becker muscular dystrophy);血纤维蛋白原异常(Dysfibrinogenemia);常染色体显性和常染色体显性先天性角化不良3(Dyskeratosis congenita autosomal dominant and autosomal dominant,3);先天性角化不良,常染色体隐性1、3、4和5(Dyskeratosis congenita,autosomal recessive,1,3,4,and 5);先天性角化不良,X连锁(Dyskeratosis congenita X-linked);家族性运动障碍伴面肌纤维颤搐(Dyskinesia,familial,with facial myokymia);异常纤溶酶原血症(Dysplasminogenemia);肌张力障碍2(扭转,常染色体隐性),3(扭转,X连锁),5(多巴反应型),10、12、16、25、26(肌阵挛)(Dystonia 2(torsion,autosomal recessive),3(torsion,X-linked),5(Dopa-responsive type),10,12,16,25,26(Myoclonic));良性家族性婴儿癫痫2(Seizures,benign familial infantile,2);婴儿早期癫痫性脑病2、4、7、9、10、11、13和14(Early infantile epileptic encephalopathy 2,4,7,9,10,11,13,and14);非典型Rett综合征;早期T细胞祖细胞急性淋巴细胞白血病(EarlyT cell progenitor acute lymphoblastic leukemia);外胚层发育不良皮肤脆性综合征(Ectodermal dysplasia skin fragility syndrome);外胚层发育不良-并指(趾)综合征1(Ectodermal dysplasia-syndactyly syndrome 1);晶状体异位,孤立型常染色体隐性和显性(Ectopia lentis,isolated autosomalrecessive and dominant);缺指(趾)畸形、外胚层发育不良和唇裂/腭裂综合征3(Ectrodactyly,ectodermal dysplasia,and cleft lip/palate syndrome 3);Ehlers-Danlos综合征7型

(常染色体隐性),经典型,2型(早衰),羟赖氨酸缺乏,4型,4型变异,以及腱生蛋白-X缺乏所致(Ehlers-Danlos syndrome type 7(autosomal recessive),classic type,type 2(progeroid),hydroxylysine-deficient,type 4,type 4variant,and due to tenascin-X deficiency);Eichsfeld型先天性肌肉萎缩症(Eichsfeld type congenital muscular dystrophy);内分泌-脑骨发育不良(Endocrine-cerebroosteodysplasia);增强型s-锥体综合征(Enhanced s-cone syndrome);大前庭导水管综合征(Enlargedvestibular aqueduct syndrome);肠激酶缺乏症(Enterokinase deficiency);疣状表皮发育不良(Epidermodysplasia verruciformis);单纯性大疱性表皮松解症和肢带型肌营养不良症,单纯性伴斑状色素沉着,单纯性伴幽门闭锁,单纯性,常染色体隐性,以及伴幽门闭锁(Epidermolysa bullosa simplex and limb girdle muscular dystrophy,simplex with mottled pigmentation,simplex with pyloric atresia,simplex,autosomal recessive,and with pyloric atresia);表皮松解性掌跖角化病(Epidermolytic palmoplantarkeratoderma);家族性热性惊厥8(Familial febrile seizures 8);儿童失神癫痫2、12(特发性泛发性,易感)5(夜间额叶),夜间额叶1型,部分性,伴可变病灶,进行性肌阵挛3和X连锁,伴可变学习障碍和行为障碍(Epilepsy,childhood absence 2,12(idiopathic generalized,susceptibility to)5(nocturnal frontal lobe),nocturnal frontal lobe type 1,partial,with variable foci,progressive myoclonic 3,and X-linked,with variable learning disabilities and behavior disorders);癫痫性脑病,儿童期发病,早期婴儿型,1、19、23、25、30和32(Epileptic encephalopathy,childhood-onset,early infantile,1,19,23,25,30,and 32);多发性骨骺发育不良伴近视和传导性耳聋(Epiphyseal dysplasia,multiple,with myopia and conductive deafness);发作性共济失调2型(Episodic ataxiatype 2);家族性阵发性疼痛综合征3(Episodic pain syndrome,familial,3);Epstein综合征;Fechtner综合征;红细胞生成性原卟啉症(Erythropoietic protoporphyria);雌激素抗性(Estrogen resistance);渗出性玻璃体视网膜病变6(Exudative vitreoretinopathy 6);法布里病和法布里病心脏变异(Fabry disease and Fabry disease,cardiac variant);因子H、VII、X、v和因子viii,2种联合缺乏,xiii,亚基,缺乏(FactorH,VII,X,v and factor viii,combined deficiency of2,xiii,a subunit,deficiency);家族性腺瘤性息肉病1和3(Familial adenomatous polyposis 1and 3);家族性淀粉样蛋白肾病伴荨麻疹和耳聋(Familial amyloid nephropathy with urticaria and deafness);家族性寒冷性荨麻疹(Familial cold urticarial);小脑吲体家族性发育不全(Familial aplasia of the vermis);家族性良性天疱疮(Familial benign pemphigus);家族性乳腺癌(Familial cancer of breast);易感性乳腺癌(Breast cancer,susceptibility to);骨肉瘤(Osteosarcoma);胰腺癌3(Pancreatic cancer 3);家族性心肌病(Familial cardiomyopathy);家族性寒冷性自身炎症综合征2(Familial cold autoinflammatory syndrome 2);家族性结直肠癌(Familial colorectal cancer);家族性渗出性玻璃体视网膜病变,X连锁(Familial exudative vitreoretinopathy,X-linked);家族性偏瘫偏头痛1型和2型(Familial hemiplegic migraine types 1and 2);家族性高胆固醇血症(Familial hypercholesterolemia);家族性肥厚型心肌病1、2、3、4、7、10、23和24(Familial

hypertrophic cardiomyopathy 1,2,3,4,7,10,23and24);家族性低钾血症-低镁血症(Familial hypokalemia-hypomagnesemia);家族性肾小球囊肿性肾病(Familial hypoplastic,glomerulocystic kidney);家族性婴儿肌无力(Familial infantile myasthenia);家族性青少年痛风(Familial juvenile gout);家族性地中海热和家族性地中海热,常染色体显性(Familial Mediterranean fever and Familial mediterranean fever,autosomal dominant);家族性孔洞脑畸形(Familial porencephaly);家族性迟发性皮肤卟啉症(Familial porphyria cutanea tarda);家族性肺毛细血管瘤病(Familial pulmonary capillary hemangiomatosis);家族性肾糖尿(Familial renal glucosuria);家族性肾低尿酸血症(Familial renal hypouricemia);家族性限制性心肌病1(Familial restrictive cardiomyopathy 1);家族性高脂蛋白血症1型和3型(Familialtype 1and 3hyperlipoproteinemia);范可尼贫血,互补群E、I、N和O(Fanconi anemia,complementationgroup E,I,N,and O);Fanconi-Bickel综合征;易感性蚕豆病(Favism,susceptibility to);家族性高热惊厥11(Febrile seizures,familial,11);Feingold综合征1;胎儿血红蛋白数量性状位点1(Fetal hemoglobin quantitative traitlocus 1);FG综合征和FG综合征4;先天性眼外肌纤维化1、2、3a(有或没有眼外肌受累)、3b(Fibrosis of extraocular muscles,congenital,1,2,3a(with or without extraocularinvolvement),3b);鱼眼病(Fish-eye disease);斑点状角膜营养不良(Fleck corneal dystrophy);Floating-Harbor综合征;语言障碍局灶性癫痫伴有或不伴有智力低下(Floating-Harbor syndrome;Focal epilepsy with speech disorder with or without mental retardation);局灶节段性肾小球硬化症5(Focal segmental glomerulosclerosis 5);前脑缺陷(Forebrain defects);Frank Ter Haar综合征;Borrone Di Rocco Crovato综合征;Frasier综合征;维尔姆氏瘤1(Wilms tumor 1);Freeman-Sheldon综合征;额干骺端发育不良1和3(Frontometaphyseal dysplasia 1and 3);额颞叶痴呆(Frontotemporal dementia);额颞叶痴呆和/或肌萎缩侧索硬化症3和4(Frontotemporal dementia and/or amyotrophic lateral sclerosis 3and 4);额颞叶痴呆染色体3连锁和额颞叶痴呆泛素阳性(Frontotemporal Dementia Chromosome 3-Linked and Frontotemporal dementia ubiquitin-positive);果糖-双磷酸酶缺乏症(Fructose-biphosphatase deficiency);富尔曼综合征(Fuhrmann syndrome);γ-氨基丁酸转氨酶缺乏症(Gamma-aminobutyric acid transaminase deficiency);Gamstorp-Wohlfart综合征;戈谢病1型和亚急性神经病(Gaucherdisease type 1and Subacute neuronopathic);家族性水平凝视麻痹伴进行性脊柱侧凸(Gaze palsy,familial horizontal,with progressive scoliosis);泛发性显性营养不良型大疱性表皮松解症(Generalized dominant dystrophic epidermolysis bullosa);泛发性癫痫伴发热性惊厥加3、1型、2型(Generalized epilepsy with febrile seizures plus 3,type 1,type 2);癫痫性脑病Lennox-Gastaut型(Epileptic encephalopathy Lennox-Gastaut type);巨轴索神经病(Giant axonal neuropathy);Glanzmann血小板无力症(Glanzmann thrombasthenia);开角型青光眼1,e、F和G(Glaucoma 1,open angle,e,F,and G);原发性先天性青光眼3,d(Glaucoma 3,primary congenital,d);先天性青光眼和先天性青光眼缺损(Glaucoma,congenital and Glaucoma,congenital,Coloboma);青少年型原发性开角型

青光眼(Glaucoma,primary open angle,juvenile-onset);胶质瘤易感性1(;Glioma susceptibility 1);葡萄糖转运蛋白1型缺乏综合征(Glucose transportertype 1deficiency syndrome);葡萄糖-6-磷酸转运缺陷(Glucose-6-phosphate transport defect);GLUT1缺乏综合征2(GLUT1 deficiency syndrome 2);特发性泛发性易感癫痫12(Epilepsy,idiopathic generalized,susceptibility to,12);谷氨酸亚胺甲基转移酶缺乏症(Glutamate formiminotransferase deficiency);戊二酸血症IIA和IIB(Glutaric acidemia IIAand IIB);戊二酸尿症1型(Glutaric aciduria,type 1);谷胱甘肽合成酶缺乏症(Gluthathione synthetase deficiency);糖原贮积病0(肌肉),II(成人型),IXa2,IXc,1A型;II型、IV型、IV型(肝病和肌病联合)、V型和VI型(Glycogen storage disease 0(muscle),II(adult form),IXa2,IXc,type 1A;type II,type IV,IV(combined hepatic and myopathic),type V,and type VI);Goldmann-Favre综合征;戈登综合征(Gordon syndrome);Gorlin综合征;前脑无裂畸形序列征(Holoprosencephaly sequence);前脑无裂畸形7(Holoprosencephaly 7);慢性肉芽肿病,X连锁,变异(Granulomatous disease,chronic,X-linked,variant);卵巢颗粒细胞瘤(Granulosa cell tumor of the ovary);灰色血小板综合征(Gray platelet syndrome);Griscelli综合征3型;Groenouw角膜营养不良I型(Groenouw corneal dystrophytype I);生长和智力低下、下颌面骨发育不全、小头畸形和腭裂(Growth andmental retardation,mandibulofacial dysostosis,microcephaly,and cleftpalate);生长激素缺乏症伴垂体异常(Growth hormone deficiency with pituitary anomalies);生长激素不敏感伴免疫缺陷(Growth hormone insensitivity with immunodeficiency);GTP环化水解酶I缺乏症(GTP cyclohydrolase I deficiency);Hajdu-Cheney综合征;手足子宫综合征(Hand foot uterus syndrome);听觉受损(Hearing impairment);婴儿毛细血管血管瘤(Hemangioma,capillary infantile);血液肿瘤(Hematologic neoplasm);血色病1、2B和3型(Hemochromatosis type 1,2B,and3);糖尿病微血管并发症7(Microvascular complications of diabetes 7);转铁蛋白血清水平定量性状位点2(Transferrin serum level quantitative trait locus 2);非缺失性血红蛋白H病(Hemoglobin H disease,nondeletional);非血球型溶血性贫血,葡萄糖磷酸异构酶缺乏所致(Hemolytic anemia,nonspherocytic,due to glucose phosphate isomerase deficiency);家族性噬血细胞性淋巴组织细胞增生症2(Hemophagocytic lymphohistiocytosis,familial,2);家族性噬血细胞性淋巴组织细胞增生症3(Hemophagocytic lymphohistiocytosis,familial,3);肝素辅因子II缺乏症(Heparin cofactorII deficiency);遗传性肠病性肢端皮炎(Hereditary acrodermatitis enteropathica);遗传性乳腺癌和卵巢癌综合征(Hereditary breast and ovarian cancer syndrome);共济失调-毛细血管扩张样障碍(Ataxia-telangiectasia-like disorder);遗传性弥漫性胃癌(Hereditary diffuse gastric cancer);遗传性弥漫性球状体白质脑病()Hereditary diffuse leukoencephalopathywith spheroids;遗传因素II、IX、VIII缺乏症(Hereditary factors II,IX,VIII deficiency disease);遗传性出血性毛细血管扩张症2型(Hereditary hemorrhagic telangiectasia type 2);遗传性无汗性疼痛不敏感(Hereditary insensitivity to pain with anhidrosis);遗传性淋巴水肿I型(Hereditary

lymphedema type I);遗传性运动和感觉神经病伴视神经萎缩(Hereditary motor and sensory neuropathy with optic atrophy);遗传性肌病伴早期呼吸衰竭(Hereditary myopathy with early respiratory failure);遗传性神经痛性肌萎缩症(Hereditary neuralgic amyotrophy);遗传性非息肉病性结直肠肿瘤(Hereditary Nonpolyposis Colorectal Neoplasms);Lynch综合征I和II();遗传性胰腺炎(Hereditary pancreatitis);慢性易感性胰腺炎(Pancreatitis,chronic,susceptibility to);IIB型和IIA型遗传性感觉和自主神经病变(Hereditary sensory and autonomic neuropathy type IIB amd IIA);遗传性铁粒幼细胞性贫血(Hereditary sideroblastic anemia);Hermansky-Pudlak综合征1、3、4和6;内脏异位2、4和6,常染色体(Heterotaxy,visceral,2,4,and 6,autosomal);内脏异位,X-连锁(Heterotaxy,visceral,X-linked);异位(Heterotopia);组织细胞性髓性网状细胞增生症(Histiocytic medullary reticulosis);组织细胞增生症-淋巴结病加综合征(Histiocytosis-lymphadenopathy plus syndrome);全羧化酶合成酶缺乏症(Holocarboxylase synthetase deficiency);前脑无裂畸形2、3、7和9(Holoprosencephaly 2,3,7,and 9);Holt-Oram综合征;同型半胱氨酸血症,吡哆醇反应性,MTHFR缺乏、CBS缺乏和高半胱氨酸尿所致(Homocysteinemia due to MTHFR deficiency,CBS deficiency,and Homocystinuria,pyridoxine-responsive);高胱氨酸尿症-巨幼红细胞性贫血,钴胺素代谢缺陷所致,cblE互补型(Homocystinuria-Megaloblastic anemia due to defect in cobalamin metabolism,cblE complementation type);Howel-Evans综合征;Hurler综合征;Hutchinson-Gilford综合征;脑积水(Hydrocephalus);高氨血症III型(Hyperammonemia,type III);高胆固醇血症和高胆固醇血症,常染色体隐性(Hypercholesterolaemia and Hypercholesterolemia,autosomal recessive);过度惊骇2和遗传性过度惊骇(Hyperekplexia2 andHyperekplexia hereditary);高铁蛋白血症白内障综合征(Hyperferritinemia cataract syndrome);高甘氨酸尿症(Hyperglycinuria);高免疫球蛋白D伴周期性发热(Hyperimmunoglobulin D with periodic fever);甲羟戊酸尿症(Mevalonic aciduria);高免疫球蛋白E综合征(Hyperimmunoglobulin E syndrome);家族性高胰岛素性低血糖症3、4和5(Hyperinsulinemic hypoglycemia familial 3,4,and 5);高胰岛素血症-高氨血症综合征(Hyperinsulinism-hyperammonemia syndrome);高赖氨酸血症(Hyperlysinemia);高锰血症伴肌张力障碍、红细胞增多症和肝硬化症(Hypermanganesemia with dystonia,polycythemia and cirrhosis);高鸟氨酸血症-高氨血症-高瓜氨酸尿综合征(Hyperornithinemia-hyperammonemia-homocitrullinuriasyndrome);甲状旁腺功能亢进1和2(Hyperparathyroidism 1and2);新生儿严重性甲状旁腺功能亢进(Hyperparathyroidism,neonatal severe);高苯丙氨酸血症,bh4缺乏,a,部分pts缺乏所致,BH4缺乏,D和非pku(Hyperphenylalaninemia,bh4-deficient,a,due to partialpts deficiency,BH4-deficient,D,andnon-pku);高磷酸酯酶症伴精神发育迟滞综合征2、3和4(Hyperphosphatasia with mental retardation syndrome 2,3,and 4);多发性骨软骨发育不良(Hypertrichotic osteochondrodysplasia);与apob32相关的家族性低β脂蛋白血症(Hypobetalipoproteinemia,familial,associatedwith apob32);低钙血症,常染色体显

性1(Hypocalcemia,autosomal dominant 1);家族性低钙尿性高钙血症1型和3型(Hypocalciuric hypercalcemia,familial,types 1and 3);软骨症(Hypochondrogenesis);铁过载的低色素性小细胞性贫血(Hypochromic microcytic anemia with iron overload);低血糖症,肝糖原合成酶缺乏所致(Hypoglycemia with deficiency of glycogen synthetase in the liver);低促性腺激素性性腺机能减退11伴有或不伴有嗅觉障碍(Hypogonadotropic hypogonadism 11with or without anosmia);少汗性外胚层发育不良伴免疫缺陷(Hypohidrotic ectodermal dysplasia with immune deficiency);少汗性X连锁外胚层发育不良(Hypohidrotic X-linked ectodermal dysplasia);低钾血症碘麻痹1和2(Hypokalemic periodic paralysis 1and 2);肠道低镁血症1(Hypomagnesemia 1,intestinal);低镁血症、癫痫发作和智力低下(Hypomagnesemia,seizures,and mental retardation);低髓鞘性脑白质营养不良7(Hypomyelinating leukodystrophy 7);左心发育不良综合征(Hypoplastic left heart syndrome);房室间隔缺损和共同房室交界(Atrioventricular septal defect and common atrioventricular junction);尿道下裂1和2,X连锁(Hypospadias 1and 2,X-linked);先天性甲状腺功能减退症无甲状腺肿1(Hypothyroidism,congenital,nongoitrous,1);少毛症8和12(Hypotrichosis 8and 12);少毛症-淋巴水肿-毛细血管扩张综合征(Hypotrichosis-lymphedema-telangiectasia syndrome);I血型系统(I blood group system);Siemens大疱性鱼鳞癣(Ichthyosis bullosa of Siemens);剥脱性鱼鳞癣(Ichthyosis exfoliativa);早产儿鱼鳞癣综合征(Ichthyosis prematurity syndrome);特发性基底神经节钙化5(Idiopathic basal ganglia calcification 5);特发性纤维化肺泡炎,慢性形式(Idiopathic fibrosing alveolitis,chronic form);先天性角化不良,常染色体显性,2和5(Dyskeratosis congenita,autosomal dominant,2and 5);婴儿特发性高钙血症(Idiopathic hypercalcemia of infancy);免疫功能障碍伴T细胞失活,钙进入缺陷所致,2(Immune dysfunction with T-cell inactivation due to calcium entry defect 2);免疫缺陷15、16、19、30、31C、38、40、8,cd3-ζ缺陷所致,具有高IgM1型和2型,且X连锁,伴镁缺陷、Epstein-Barr病毒感染和瘤形成(Immunodeficiency 15,16,19,30,31C,38,40,8,due to defect in cd3-zeta,withhyperIgM type 1and 2,and X-Linked,with magnesium defect,Epstein-Barr virus infection,and neoplasia);免疫缺陷-着丝粒不稳定-面部异常综合征2(Immunodeficiency-centromeric instability-facial anomalies syndrome 2);包涵体肌病2和3(Inclusion body myopathy 2and 3);Nonaka肌病;家族性婴儿惊厥和阵发性舞蹈手足徐动症(Infantile convulsions and paroxysmal choreoathetosis,familial);婴儿皮质骨质增生症(Infantile cortical hyperostosis);婴儿GM1神经节苷脂沉积症(Infantile GM1 gangliosidosis);婴儿低磷酸酯酶症(Infantile hypophosphatasia);婴儿肾炎(Infantile nephronophthisis);婴儿眼球震颤,X连锁(Infantile nystagmus);婴儿帕金森病-肌张力障碍(Infantile Parkinsonism-dystonia);与多尾精子和DNA过多相关的不育症(Infertility associated with multi-tailed spermatozoa and excessive DNA);胰岛素抵抗(Insulin resistance);胰岛素抵抗糖尿病和黑棘皮病(Insulin-resistant diabetes mellitus and acanthosis nigricans);胰岛素依赖型糖尿病分泌性腹泻综合征(Insulin-

dependent diabetes mellitus secretory diarrhea syndrome);巨核间质性肾炎(Interstitial nephritis,karyomegalic);宫内发育迟缓、干骺端发育不良、先天性肾上腺发育不良、生殖器异常(Intrauterine growth retardation,metaphyseal dysplasia,adrenal hypoplasia congenita,and genital anomalies);碘酪氨酰偶联缺陷(Iodotyrosyl coupling defect);IRAK4缺乏症;虹膜发育不全显性型和1型(Iridogoniodysgenesis dominant type and type 1);脑组织铁沉积(Iron accumulation in brain);坐骨髌骨发育不良(Ischiopatellardysplasia);胰岛细胞增生症(Islet cell hyperplasia);孤立型17,20-裂解酶缺乏症(Isolated 17,20-lyase deficiency);孤立型促黄体素缺乏症(Isolated lutropin deficiency);异戊酰辅酶A脱氢酶缺乏症(Isovaleryl-CoA dehydrogenase deficiency);Jankovic Rivera综合征;Jervell-Lange Nielsen综合征2;Joubert综合征1、6、7、9/15(双基因)、14、16和17,以及Orofaciodigital综合征xiv(Jankovic Rivera syndrome;Jervell and Lange-Nielsen syndrome 2;Joubert syndrome 1,6,7,9/15(digenic),14,16,and 17,and Orofaciodigital syndrome xiv);Herlitz交界性大疱性表皮松解症(Junctional epidermolysis bullosa gravis ofHerlitz);幼年GM>1<神经节苷脂沉积症(Juvenile GM>1<gangliosidosis);幼年性息肉病综合征(Juvenile polyposis syndrome);幼年性息肉病/遗传性出血性毛细血管扩张综合征(Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome);青少年视网膜劈裂症(Juvenile retinoschisis);Kabuki make-up综合征;卡尔曼综合征1、2和6(Kallmann syndrome 1,2,and 6);青春期延迟(Delayed puberty);Kanzaki病;卡拉克综合征(Karak syndrome);Kartagener综合征;Kenny-Caffey综合征2型;Keppen-Lubinsky综合征;圆锥形角膜1(Keratoconus 1);毛囊角化病(Keratosis follicularis);掌跖角化病1(Keratosis palmoplantaris striata 1);Kindler综合征;L-2-羟基戊二酸尿症(L-2-hydroxyglutaric aciduria);Larsen综合征,显性型;III型晶格角膜营养不良(Lattice corneal dystrophy Type III);莱伯氏黑蒙症(Leber amaurosis);Zellweger综合征;过氧化物酶体生物发生障碍(Peroxisome biogenesis disorders);Zellweger综合征谱(Zellweger syndrome spectrum);莱伯氏先天性黑蒙症11、12、13、16、4、7和9(Leber congenital amaurosis 11,12,13,16,4,7,and 9);莱伯氏视神经萎缩(Leberoptic atrophy);氨基糖苷类诱发的耳聋(Aminoglycoside-induced deafness);非综合征型感觉神经性耳聋,线粒体(Deafness,nonsyndromic sensorineural,mitochondrial);左心室致密化不全5(Left ventricular noncompaction 5);左右轴畸形(Left-right axis malformations);Leigh病;线粒体短链烯酰辅酶A水合酶1缺乏症(Mitochondrial short-chain Enoyl-CoA Hydratase 1deficiency);Leigh综合征,线粒体复合物I缺乏所致;Leiner病;Leri Weill软骨生成障碍(Leri Weill dyschondrosteosis);致死性先天性挛缩综合征6(Lethal congenital contracture syndrome 6);白细胞粘附缺陷症I型和III型(Leukocyte adhesion deficiency type I and III);脑白质营养不良,髓鞘形成不足11和6(Leukodystrophy,Hypomyelinating,11and 6);脑白质病,伴共济失调、脑干和脊髓受累且乳酸升高,伴白质消失且进展性,伴卵巢功能衰竭(Leukoencephalopathy with ataxia,with Brainstem and Spinal Cord Involvement and Lactate Elevation,with vanishing white matter,and

progressive,with ovarian failure);全白甲病(Leukonychia totalis);路易体痴呆症(Lewy body dementia);Lichtenstein-Knorr综合征;Li-Fraumeni综合征1;Lig4综合征;肢带型肌营养不良症1B、2A、2B、2D、C1、C5、C9、C14型(Lig4 syndrome;Limb-girdle muscular dystrophy,type 1B,2A,2B,2D,C1,C5,C9,C14);先天性肌营养不良-肌营养不良症伴脑和眼异常A14和B14型(Congenital muscular dystrophy-dystroglycanopathy with brain and eye anomalies,typeA14 and B14);联合脂肪酶缺乏症(Lipase deficiency combined);脂质蛋白沉积症(Lipid proteinosis);家族性部分脂肪代谢障碍2型和3型(Lipodystrophy,familial partial,type 2and 3);无脑回畸形1、2(X连锁)、3、6(小头畸形),X连锁(Lissencephaly 1,2(X-linked),3,6(with microcephaly),X-linked);皮质下层状异位,X-连锁(Subcortical laminar heterotopia,X-linked);急性婴儿肝功能衰竭(Liver failure acute infantile);Loeys-Dietz综合征1、2、3;长QT综合征1、2、2/9、2/5、(二基因型)、3、5和5,获得性,易感性;肺癌(Lung cancer);遗传性淋巴水肿,id(Lymphedema,hereditary,id);原发性淋巴水肿伴脊髓发育不良(Lymphedema,primary,with myelodysplasia);淋巴增生综合征1、1(X连锁)和2(Lymphoproliferative syndrome 1,1(X-linked),and 2);溶酶体酸性脂肪酶缺乏症(Lysosomal acid lipase deficiency);巨头畸形、巨大儿、面部畸形综合征(Macrocephaly,macrosomia,facial dysmorphism syndrome);成年型卵黄样黄斑营养不良(Macular dystrophy,vitelliform,adult-onset;);恶性高热易感性1型(Malignant hyperthermia susceptibility type 1);非霍奇金恶性淋巴瘤(Malignant lymphoma,non-Hodgkin);恶性黑素瘤(Malignant melanoma);前列腺恶性肿瘤(Malignant tumor of prostate);下颌骨骨发育不良(Mandibuloacral dysostosis);下颌骨发育不良伴A型或B型脂肪营养不良,非典型(Mandibuloacral dysplasia with type A or B lipodystrophy,atypical);下颌面骨发育不良,TreacherCollins型,常染色体隐性(Mandibulofacial dysostosis,Treacher Collins type,autosomal recessive);甘露糖结合蛋白缺乏症(Mannose-binding protein deficiency);枫糖尿病1A型和3型(Maple syrup urine disease type 1A and type 3);Marden Walker样综合征;马凡综合征(Marfan syndrome);Marinesco-Sj\xc3\xb6gren综合征;Martsolf综合征;青少年发病的成年型糖尿病1型、2型、11型、3型和9型(Maturity-onset diabetes of the young,type 1,type 2,type 11,type 3,andtype 9);May-Hegglin异常;MYH9相关疾病(MYH9 related disorders);Sebastian综合征;McCune-Albright综合征;生长激素细胞腺瘤(Somatotroph adenoma);性索间质瘤(Sex cord-stromal tumor);Cushing综合征;McKusick Kaufman综合征;McLeod神经棘红细胞增多症综合征(McLeod neuroacanthocytosis syndrome);Meckel-Gruber综合征;中链酰基辅酶A脱氢酶缺乏症(Medium-chain acyl-coenzyme A dehydrogenase deficiency);髓母细胞瘤(Medulloblastoma);巨脑性脑白质病伴皮质下囊肿1和2a(Megalencephalic leukoencephalopathy with subcortical cysts 1and 2a);先天性巨脑毛细血管扩张性大理石样皮肤(Megalencephaly cutis marmorata telangiectatica congenital);PIK3CA相关过度生长谱(PIK3CA Related Overgrowth Spectrum);巨脑-多小脑回-多指(趾)畸形-脑积水综合征2(Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 2);硫胺素反应性巨幼红细胞性贫血伴糖尿病和感觉神经性耳

聋(Megaloblastic anemia,thiamine-responsive,with diabetes mellitus and sensorineural deafness);Meier-Gorlin综合征1和4;Melnick-Needles综合征;脑膜瘤(Meningioma);智力低下,X-连锁,3、21、30和72(Mental retardation,X-linked,3,21,30,and 72);智力低下和小头畸形伴脑桥和小脑发育不全(Mental retardation and microcephaly with pontine and cerebellar hypoplasia);智力低下X连锁综合征5(Mental retardation X-linked syndromic 5);智力低下、上颌前突和斜视(Mental retardation,anterior maxillary protrusion,and strabismus);智力低下,常染色体显性12、13、15、24、3、30、4、5、6和9(Mental retardation,autosomal dominant 12,13,15,24,3,30,4,5,6,and 9);智力低下,常染色体隐性15、44、46和5(Mental retardation,autosomal recessive 15,44,46,and 5);智力低下、刻板运动、癫痫和/或脑畸形(Mental retardation,stereotypic movements,epilepsy,and/or cerebral malformations);智力低下,综合征性,Claes-Jensen型,X连锁(Mental retardation,syndromic,Claes-Jensentype,X-linked);X-连锁非特异性智力低下,综合征性Hedera型和综合征性wu型(Mental retardation,X-linked,nonspecific,syndromic,Hedera type,and syndromic,wu type);Merosin缺乏先天性肌营养不良症(Merosin deficient congenital muscular dystrophy);异染性脑白质营养不良,青少年型、婴儿晚期型和成人型(Metachromatic leukodystrophy juvenile,late infantile,and adult types);异染性脑白质营养不良(Metachromatic leukodystrophy);后生营养性发育不良(Metatrophic dysplasia);高铁血红蛋白血症I型和2型(Methemoglobinemia types I and 2);蛋氨酸腺苷转移酶缺乏症,常染色体显性(Methionine adenosyltransferase deficiency,autosomal dominant);甲基丙二酸血症伴高胱氨酸尿症(Methylmalonic acidemia with homocystinuria);甲基丙二酸尿症cblB型(Methylmalonic aciduria cblB type);甲基丙二酸尿症,甲基丙二酰辅酶A变位酶缺乏所致(Methylmalonic aciduria due to methylmalonyl-CoA mutase deficiency);甲基丙二酸尿症,mut(0)型(METHYLMALONIC ACIDURIA,mut(0)TYPE);小头畸形骨发育不良原始侏儒症2型(Microcephalic osteodysplastic primordial dwarfism type 2);小头畸形伴有或不伴有脉络膜视网膜病变、淋巴水肿或智力低下(Microcephaly with or without chorioretinopathy,lymphedema,or mental retardation);小头畸形、食管裂孔疝和肾病综合征(Microcephaly,hiatal hernia and nephrotic syndrome);小头畸形(Microcephaly);胼胝体发育不全(Hypoplasia of the corpus callosum);痉挛性截瘫50,常染色体隐性(Spastic paraplegia 50,autosomal recessive);整体发育迟缓症(Global developmental delay);CNS髓鞘形成不足(CNS hypomyelination);脑萎缩(Brain atrophy);小头畸形、正常智力和免疫缺陷(Microcephaly,normal intelligence and immunodeficiency);小头畸形-毛细血管畸形综合征(Microcephaly-capillary malformation syndrome);小细胞性贫血(Microcytic anemia);小眼球综合征5、7和9(Microphthalmia syndromic 5,7,and 9);小眼球,孤立型3、5、6、8和伴眼缺损6(Microphthalmia,isolated 3,5,6,8,andwith coloboma 6);球形晶状体(Microspherophakia;);家族性基底型偏头痛(Migraine,familial basilar);Miller综合征;微核肌病伴眼外肌麻痹(Minicore myopathy with external ophthalmoplegia);先天性中央核肌病(Myopathy,congenital with cores);Mitchell-Riley综合征;线粒体3-羟

基-3-甲基戊二酰辅酶A合酶缺乏症(mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency);线粒体复合物I、II、III、III(核型2、4或8)缺乏症(Mitochondrial complex I,II,III,III(nucleartype 2,4,or 8)deficiency);线粒体DNA耗竭综合征11、12(心肌病型)、2、4B(MNGIE型)、8B(MNGIE型)(Mitochondrial DNA depletion syndrome 11,12(cardiomyopathic type),2,4B(MNGIE type),8B(MNGIE type));线粒体DNA耗竭综合征3和7、肝脑型和13(脑肌病型)(Mitochondrial DNA-depletion syndrome 3and 7, hepatocerebraltypes,and 13(encephalomyopathic type));线粒体磷酸盐载体和丙酮酸载体缺乏症(Mitochondrial phosphate carrier and pyruvate carrier deficiency);线粒体三功能蛋白缺乏症(Mitochondrial trifunctional protein deficiency);长链3-羟烷基辅酶A脱氢酶缺乏症(Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency);Miyoshi肌营养不良1(Miyoshi musculardystrophy 1);远端肌病,伴有胫骨前肌发病(Myopathy,distal,with anterior tibial onset);Mohr-Tranebjaerg综合征;钼辅因子缺乏症,互补群A(Molybdenum cofactor deficiency,complementation group A);Mowat-Wilson综合征;粘多糖症IIIγ(Mucolipidosis III Gamma);黏多糖贮积症VI型、VI型(重度)和VII型(Mucopolysaccharidosis type VI,type VI(severe),andtype VII);黏多糖贮积症,MPS-I-H/S、MPS-II、MPS-III-A、MPS-III-B、MPS-III-C、MPS-IV-A、MPS-IV-B(Mucopolysaccharidosis,MPS-I-H/S,MPS-II,MPS-III-A,MPS-III-B,MPS-III-C,MPS-IV-A,MPS-IV-B);色素性视网膜炎73(Retinitis Pigmentosa 73);神经节苷脂沉积症GM1型1(伴心脏受累)3(Gangliosidosis GM1 type1(with cardiac involvement)3);多中心骨质溶解肾病(Multicentric osteolysis nephropathy);多中心骨质溶解、结节病和关节病(Multicentric osteolysis,nodulosis and arthropathy);多发性先天性畸形(Multiple congenital anomalies);心房间隔缺损2(Atrial septal defect 2);多发性先天性畸形-张力减退-癫痫综合征3(Multiple congenital anomalies-hypotonia-seizures syndrome 3);多发性皮肤和粘膜静脉畸形(Multiple Cutaneous andMucosal Venous Malformations);多发性内分泌肿瘤1型和4型(Multiple endocrine neoplasia, types 1and 4);多发性骨骺发育不良5或显性(Multiple epiphyseal dysplasia 5orDominant);多发性胃肠道闭锁(Multiple gastrointestinal atresias);多发性翼状胬肉综合征Escobar型(Multiple pterygium syndrome Escobartype);多发性硫酸酯酶缺乏症(Multiple sulfatase deficiency);多发性骨性连接综合征3(Multiple synostoses syndrome 3);肌肉AMP鸟嘌呤氧化酶缺乏症(MuscleAMP guanine oxidase deficiency);肌眼脑病(Muscle eye braindisease);先天性肌肉萎缩症,大锥状颗粒型(Musculardystrophy,congenital,megaconialtype);家族性婴儿肌无力1(Myasthenia, familial infantile,1);先天性肌无力综合征11,与乙酰胆碱受体缺乏症有关(Myasthenic Syndrome,Congenital,11,associated with acetylcholine receptor deficiency);先天性肌无力综合征17、2A(慢通道),4B(快通道),且无管状聚集体(Myasthenic Syndrome,Congenital,17,2A(slow-channel),4B(fast-channel),and without tubular aggregates);髓过氧化物酶缺乏症(Myeloperoxidase deficiency);MYH相关息肉病(MYH-associated polyposis);子宫内膜癌(Endometrial carcinoma);心肌梗塞1(Myocardial infarction 1);肌阵挛性肌张力障碍(Myoclonic dystonia);肌阵

挛-失张力癫痫(Myoclonic-Atonic Epilepsy);肌阵挛性癫痫伴破碎红纤维(Myoclonus with epilepsy with ragged red fibers);肌原纤维肌病1和ZASP相关(Myofibrillar myopathy 1and ZASP-related);急性复发性肌红蛋白尿,常染色体隐性(Myoglobinuria, acute recurrent,autosomal recessive);肌神经性胃肠道脑病综合征(Myoneural gastrointestinal encephalopathy syndrome);婴儿小脑性共济失调伴进行性外眼肌麻痹(Cerebellar ataxia infantile with progressive external ophthalmoplegia);线粒体DNA耗竭综合征4B,MNGIE型(Mitochondrial DNA depletion syndrome 4B,MNGIE type);先天性中央核肌病1,伴过远端多肌梭1,乳酸酸中毒和铁粒幼红细胞性贫血1,线粒体进行性伴先天性白内障,听力损失和发育迟缓,且有管状聚集体2(Myopathy, centronuclear,1,congenital,with excess of muscle spindles,distal,1,lactic acidosis,and sideroblastic anemia 1,mitochondrial progressive with congenital cataract,hearing loss,and developmental delay,and tubular aggregate,2);近视6(Myopia6);肌硬化症,常染色体隐性(Myosclerosis,autosomal recessive);先天性肌强直(Myotonia congenital);先天性肌强直,常染色体显性和隐性形式(Congenital myotonia,autosomal dominant andrecessive forms);指甲-髌骨综合征(Nail-patella syndrome);Nance-Horan综合征;真性小眼球2(Nanophthalmos 2);Navajo神经肝病(Navajo neurohepatopathy);线状体肌病3和9(Nemaline myopathy 3and 9);新生儿肌张力低下(Neonatal hypotonia);智能障碍(Intellectual disability);癫痫(Seizures);言语和语言发育迟缓(Delayed speech and language development);智力低下,常染色体显性31(Mental retardation,autosomal dominant 31);新生儿肝内胆汁淤积,希特林缺乏所致(Neonatal intrahepatic cholestasis caused by citrin deficiency);肾源性尿崩症,X连锁肾源性尿崩症(Nephrogenic diabetes insipidus,Nephrogenic diabetes insipidus,X-linked);肾结石/骨质疏松症,低磷血症2(Nephrolithiasis/osteoporosis, hypophosphatemic,2);肾消耗病13、15和4(Nephronophthisis 13,15and 4);不孕症(Infertility);小脑-眼-肾综合征(肾消耗病、动眼神经失用和小脑异常)(Cerebello-oculo-renal syndrome(nephronophthisis,oculomotor apraxia and cerebellar abnormalities));肾病综合征3型,5型,有或无眼部异常,7型和9型(Nephrotic syndrome, type 3,type 5,with or without ocular abnormalities,type 7,andtype 9);Nestor-Guillermo早衰综合征(Nestor-Guillermo progeria syndrome);Neu-Laxova综合征1;神经退行性伴脑铁沉积4和6(Neurodegeneration with brain iron accumulation 4and 6);神经铁蛋白病变(Neuroferritinopathy);神经纤维瘤病1型和2型(Neurofibromatosis,type 1andtype 2);神经纤维肉瘤(Neurofibrosarcoma);神经垂体尿崩症(Neurohypophyseal diabetes insipidus);遗传性感觉神经病IC型(Neuropathy, Hereditary Sensory,Type IC);中性1氨基酸转运缺陷(Neutral 1amino acid transport defect);中性脂质贮积病伴肌病(Neutral lipid storage disease with myopathy);中性粒细胞免疫缺陷综合征(Neutrophil immunodeficiency syndrome);Nicolaides-Baraitser综合征;成人型尼曼病C1型、C2型、A型和C1型(Niemann-Pick disease type C1, C2,typeA,and type C1,adult form);非酮症高甘氨酸血症(Non-ketotic hyperglycinemia);努南综合征1和4,LEOPARD综合征1(Noonan syndrome 1and4,LEOPARD

syndrome1)；努南综合征样疾病伴有或不伴有幼年型粒单核细胞白血病(Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia)；正常血钾型周期性麻痹，钾敏感(Normokalemic periodic paralysis,potassium-sensitive)；诺鲁姆病(Norum disease)；癫痫、听力损失和智力低下综合征(Epilepsy,Hearing Loss, AndMental Retardation Syndrome)；智力低下，X连锁102和综合征性13(Mental Retardation,X-Linked 102and syndromic 13)；肥胖症(Obesity)；眼白化病I型(Ocular albinism,type I)；眼皮肤白化病1B型、3型和4型(Oculocutaneous albinism type 1B, type 3,and type4)；眼齿指发育不良(Oculodentodigital dysplasia)；牙齿型低碱性磷酸酯酶症(Odontohypophosphatasia)；Odontotrichomelic综合征；小口氏病(Oguchi disease)；少牙畸形-结直肠癌综合征(Oligodontia-colorectal cancersyndrome)；OpitzG/BBB综合征；视神经萎缩9(Optic atrophy 9)；口-面-指综合征(Oral-facial-digital syndrome)；鸟氨酸转氨酶缺乏症(Ornithine aminotransferase deficiency)；口面裂畸形11和7，唇裂/腭裂-外胚层发育不良综合征(Orofacial cleft 11and 7,Cleft lip/palate-ectodermal dysplasia syndrome)；Orstavik Lindemann Solberg综合征；骨关节炎伴轻度软骨发育异常(Osteoarthritis with mild chondrodysplasia)；剥脱性骨软骨炎(Osteochondritis dissecans)；成骨不全症12型、5型、7型、8型、I型、III型，伴有巩膜正常，显性型，隐性围产期致死(Osteogenesis imperfecta type 12,type 5,type 7, type 8,type I,type III,with normal sclerae,dominant form,recessive perinatal lethal)；纹状骨病伴患有颅骨硬化(Osteopathia striatawith cranial sclerosis)；骨硬化症，常染色体显性1型和2型，隐性4，隐性1，隐性6(Osteopetrosis autosomal dominant type 1and 2,recessive 4,recessive 1,recessive 6)；骨质疏松症伴假神经胶质瘤(Osteoporosis with pseudoglioma)；耳-腭-指综合征，I型和II型(Oto-palato-digital syndrome,types I andII)；卵巢发育不全1(Ovarian dysgenesis 1)；卵巢脑白质营养不良(Ovarioleukodystrophy)；先天性厚甲症4和2型(Pachyonychia congenita4 and type 2)；家族性骨佩吉特病(Paget disease of bone,familial)；Pallister-Hall综合征；掌跖角化病，非表皮松懈性、局灶性或弥漫性(Palmoplantar keratoderma, nonepidermolytic,focal or diffuse)；胰腺发育不全和先天性心脏病(Pancreatic agenesis and congenital heart disease)；Papillon-Lef\xc3\xa8vre综合征；副神经节瘤3(Paragangliomas 3)；von Eulenburg先天性副肌强直(Paramyotonia congenita of von Eulenburg)；甲状旁腺癌(Parathyroid carcinoma)；帕金森病14、15、19(青少年发病)、2、20(早发)、6、常染色体隐性早发和9(Parkinson disease 14,15,19(juvenile-onset),2,20(early-onset),6,(autosomal recessive early-onset,and 9)；部分白化病(Partial albinism)；部分次黄嘌呤-鸟嘌呤磷酸核糖转移酶缺乏症(Partial hypoxanthine-guanine phosphoribosyltransferase deficiency)；视网膜色素上皮图形状营养不良(Patterned dystrophy of retinal pigment epithelium)；PC-K6a；Pelizaeus-Merzbacher病；Pendred综合征；外周脱髓鞘神经病，中枢性髓鞘形成障碍(Peripheral demyelinating neuropathy,central dysmyelination)；Hirschsprung病；永久性新生儿糖尿病(Permanent neonatal diabetes mellitus)；永久性新生儿糖尿病伴神经病学特征(Diabetes mellitus,permanent neonatal,with neurologic features)；

新生儿胰岛素依赖型糖尿病(Neonatal insulin-dependent diabetes mellitus);青少年发病的成年型糖尿病2型(Maturity-onset diabetes of the young,type 2);过氧化物酶体生物发生障碍14B、2A、4A、5B、6A、7A和7B(Peroxisome biogenesis disorder 14B,2A,4A,5B,6A,7A,and 7B;Perrault syndrome 4);Perrault综合征4;Perry综合征;婴儿持续高胰岛素性低血糖症(Persistent hyperinsulinemic hypoglycemia of infancy);家族性高胰岛素血症(familial hyperinsulinism);表现型(Phenotypes);苯丙酮尿症(Phenylketonuria);嗜铬细胞瘤(Pheochromocytoma);遗传性副神经节瘤-嗜铬细胞瘤综合征(Hereditary Paraganglioma-Pheochromocytoma Syndromes);副神经节瘤1(Paragangliomas 1);肠类癌瘤(Carcinoid tumor of intestine);Cowden综合征3;磷酸甘油酸脱氢酶缺乏症(Phosphoglycerate dehydrogenase deficiency);磷酸甘油酸激酶1缺乏症(Phosphoglycerate kinase 1 deficiency);光敏毛发硫营养不良(Photosensitive trichothiodystrophy);植烷酸贮积病(Phytanic acid storage disease);皮克病(Pick disease);皮尔森综合征(Pierson syndrome);色素性视网膜营养不良(Pigmentary retinal dystrophy);原发性色素性结节性肾上腺皮质病1(Pigmented nodular adrenocortical disease,primary,1);毛母质瘤(Pilomatrixoma);Pitt-Hopkins综合征;垂体依赖性皮质醇增多症(Pituitary dependent hypercortisolism);联合性垂体激素缺乏症1、2、3和4(Pituitary hormone deficiency,combined 1,2,3,and 4);纤溶酶原激活物抑制剂1型缺乏症(Plasminogen activator inhibitor type 1deficiency);纤溶酶原缺乏症I型(Plasminogen deficiency,type I);血小板型出血障碍15和8(Platelet-type bleeding disorder 15and 8);遗传性纤维化皮肤异色病伴肌腱挛缩、肌病和肺纤维化(Poikiloderma,hereditary fibrosing,with tendon contractures,myopathy,and pulmonary fibrosis);多囊性肾病2,成人型、婴儿型(Polycystic kidney disease 2,adult type,and infantile type);多囊性脂膜样骨发育不良伴硬化性白质脑病(Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy);聚葡糖体肌病1伴有或不伴有免疫缺陷(Polyglucosan body myopathy 1with or without immunodeficiency);不对称双侧额顶多小脑回畸形(Polymicrogyria,asymmetric,bilateral frontoparietal);多发性神经病、听力损失、共济失调、色素性视网膜炎和白内障(Polyneuropathy,hearing loss,ataxia,retinitis pigmentosa,and cataract);脑桥小脑发育不全4型(Pontocerebellar hypoplasia type 4);腘窝翼状胬肉综合征(Popliteal pterygium syndrome);孔洞脑畸形2(Porencephaly 2);弥漫性浅表性光敏性汗孔角化症8(Porokeratosis 8,disseminated superficial actinic type);胆色素原合酶缺乏症(Porphobilinogen synthase deficiency);迟发性皮肤卟啉症(Porphyria cutanea tarda);后柱共济失调伴视网膜色素变性(Posterior column ataxia with retinitis pigmentosa);后极白内障2型(Posterior polar cataracttype 2);Prader-Willi样综合征;卵巢功能早衰4、5、7和9(Premature ovarian failure 4,5,7,and 9);原发性常染色体隐性小头畸形10、2、3和5(Primary autosomal recessive microcephaly 10,2,3,and 5);原发性纤毛运动障碍24(Primary ciliary dyskinesia 24);原发性扩张型心肌病(Primary dilated cardiomyopathy);左心室致密化不全6(Left ventricular noncompaction 6);4,左心室

致密化不全10(Left ventricular noncompaction 6);阵发性心房颤动(Paroxysmal atrial fibrillation);原发性高草酸尿症I型、III型和III型(Primary hyperoxaluria, type I,type,and type III);原发性肥厚性骨关节病,常染色体隐性2(Primary hypertrophic osteoarthropathy,autosomal recessive 2);原发性低镁血症(Primary hypomagnesemia);原发性开角型青少年青光眼1(Primary open angle glaucoma juvenile onset 1);原发性肺动脉高压症(Primary pulmonary hypertension);Primrose综合征;进行性家族性心脏传导阻滞1B型(Progressive familial heart block type 1B);进行性家族性肝内胆汁淤积2和3(Progressive familial intrahepatic cholestasis 2and 3);进行性肝内胆汁淤积(Progressive intrahepatic cholestasis);进行性肌阵挛性癫痫伴共济失调(Progressive myoclonus epilepsy with ataxia);进行性假类风湿发育不良(Progressive pseudorheumatoid dysplasia);进行性硬化性灰质营养不良(Progressive sclerosing poliodystrophy);脯肽酶缺乏症(Prolidase deficiency);脯氨酸脱氢酶缺乏症(Proline dehydrogenase deficiency);精神分裂症4(Schizophrenia4);备解素缺乏症,X连锁(Properdin deficiency,X-linked);丙酸血症(Propionic academia);前蛋白转化酶1/3缺乏症(Proprotein convertase 1/3deficiency);遗传性前列腺癌2(Prostate cancer,hereditary,2);红色素缺陷(Protan defect);蛋白尿(Proteinuria);芬兰型先天性肾病综合征(Finnish congenital nephrotic syndrome);Proteus综合征;乳腺癌(Breast adenocarcinoma);假性软骨发育不全性脊椎骨骺发育不良综合征(Pseudoachondroplastic spondyloepiphyseal dysplasia syndrome);假性醛固酮减少症1型常染色体显性和隐性和2型(Pseudohypoaldosteronismtype 1autosomal dominant and recessive and type 2);假性甲状旁腺功能减退症1A型,假性假甲状旁腺功能减退症
(Pseudohypoparathyroidismtype 1A,Pseudopseudohypoparathyroidism);假性新生儿肾上腺脑白质营养不良(Pseudoneonatal adrenoleukodystrophy);假原发性醛固酮增多症(Pseudoprimary hyperaldosteronism);弹性假黄瘤(Pseudoxanthoma elasticum);婴儿泛发性动脉钙化2(Generalized arterial calcification of infancy 2);弹性假黄瘤样障碍伴多发性凝血因子缺乏症(Pseudoxanthoma elasticum-like disorder with multiple coagulation factor deficiency);银屑病易感性2(Psoriasis susceptibility 2);PTEN错构瘤肿瘤综合征(PTEN hamartoma tumor syndrome);肺动脉高血压,与遗传性出血性毛细血管扩张相关(Pulmonary arterial hypertension related to hereditary hemorrhagic telangiectasia);端粒相关肺纤维化和/或骨髓衰竭1和3(Pulmonary FibrosisAnd/OrBone Marrow Failure,Telomere-Related,1and 3);原发性肺动脉高血压1伴遗传性出血性毛细血管扩张(Pulmonary hypertension,primary,1,with hereditary hemorrhagic telangiectasia);嘌呤核苷磷酸化酶缺乏症(Purine-nucleoside phosphorylase deficiency);丙酮酸羧化酶缺乏症(Pyruvate carboxylase deficiency);丙酮酸脱氢酶E1-α缺乏症(Pyruvate dehydrogenase E1-alpha deficiency);红细胞丙酮酸激酶缺乏症(Pyruvate kinase deficiency of red cells);雷恩综合征(Raine syndrome);Rasopathy;隐性营养不良性大疱性表皮松解症(Recessive dystrophic epidermolysis bullosa);非综合征先天性指甲疾病8(Nail disorder,

nonsyndromic congenital,8);雷凡斯坦综合征(Reifenstein syndrome);肾发育不良(Renal adysplasia);肾肉碱转运缺陷(Renal carnitine transport defect);肾缺损综合征(Renal coloboma syndrome);肾发育不良(Renal dysplasia);肾发育不良、视网膜色素营养不良、小脑共济失调和骨骼发育不良(Renal dysplasia,retinal pigmentary dystrophy,cerebellar ataxia and skeletal dysplasia);远端肾小管酸中毒,常染色体隐性,伴迟发性感觉神经性听力损失或伴溶血性贫血(Renal tubular acidosis,distal,autosomal recessive,with late-onset sensorineural hearing loss,or with hemolytic anemia);近端肾小管酸中毒伴眼部异常和智力低下(Renal tubular acidosis,proximal,with ocular abnormalities and mental retardation);视锥营养不良3B(Retinal cone dystrophy 3B);视网膜色素变性(Retinitis pigmentosa);色素性视网膜炎10、11、12、14、15、17和19(Retinitis pigmentosa10,11,12,14,15,17,and 19);色素性视网膜炎2、20、25、35、36、38、39、4、40、43、45、48、66、7、70、72(Retinitis pigmentosa2,20,25,35,36,38,39,4,40,43,45,48,66,7,70,72);视网膜母细胞瘤(Retinoblastoma);雷特氏症(Rett disorder);横纹肌样肿瘤易感综合征2(Rhabdoid tumor predisposition syndrome 2);孔源性视网膜脱离,常染色体显性(Rhegmatogenous retinal detachment,autosomal dominant);肢根型点状软骨发育不良2型和3型(Rhizomelic chondrodysplasiapunctata type 2and type 3);Roberts-SC短肢畸形综合征(Roberts-SC phocomelia syndrome);Robinow Sorauf综合征;Robinow综合征,常染色体隐性,常染色体隐性,伴短同时并指(趾)多指(趾)(Robinow syndrome,autosomal recessive,autosomal recessive,with brachy-syn-polydactyly);Rothmund-Thomson综合征;Rapadilino综合征;RRM2B相关线粒体疾病(RRM2B-related mitochondrial disease);Rubinstein-Taybi综合征;Salla病;Sandhoff病,成人型和婴儿型;早发结节病(Sarcoidosis,early-onset);Blau综合征;辛德勒(Schindler)病1型(Schindler disease,type 1);脑裂畸形(Schizencephaly);精神分裂症15(Schizophrenia 15);Schneckenbecken发育不良;神经鞘瘤病2(Schwannomatosis 2);Schwartz Jampel综合征1型;硬化性角膜,常染色体隐性(Sclerocornea,autosomal recessive);硬化性骨化病(Sclerosteosis);继发性甲状腺功能减退(Secondary hypothyroidism);Segawa综合征,常染色体隐性;Senior-Loken综合征4和5;感觉共济失调神经病、构音障碍和眼肌麻痹(Sensory ataxic neuropathy,dysarthria,and ophthalmoparesis);Sepiapterin还原酶缺乏症;SeSAME综合征;严重联合免疫缺陷,ADA缺乏所致,伴小头畸形、生长迟缓、对电离辐射敏感,非典型、常染色体隐性遗传、T细胞阴性、B细胞阳性、NK细胞阴性或NK阳性(Severe combined immunodeficiency due to ADA deficiency,with microcephaly,growth retardation,and sensitivity to ionizing radiation,atypical,autosomal recessive,T cell-negative,B cell-positive,NK cell-negative of NK-positive);严重先天性中性粒细胞减少症(Severe congenital neutropenia);严重先天性中性粒细胞减少症3,常染色体隐性或显性(Severe congenital neutropenia 3,autosomal recessive or dominant);严重先天性中性粒细胞减少症6,常染色体隐性(Severe congenital neutropenia and 6,autosomal recessive);婴儿严重肌阵挛性癫痫(Severe myoclonic epilepsy in infancy);泛发性癫痫伴高热惊厥加1型和2型(Generalized

epilepsy with febrile seizures plus,types 1and 2);严重X连锁肌管肌病(Severe X-linked myotubularmyopathy);短QT综合征3(Short QT syndrome 3);身材矮小伴非特异性骨骼异常(Short stature withnonspecific skeletal abnormalities);身材矮小、耳道闭锁、下颌发育不全、骨骼异常(Short stature,auditory canal atresia,mandibularhypoplasia,skeletal abnormalities);身材矮小、指(趾)甲发育不良、面部畸形和少毛症(Short stature,onychodysplasia,facial dysmorphism,and hypotrichosis);原始侏儒症(Primordial dwarfism);胸短肋发育不良11或3伴有或不伴有多指(趾)畸形(Short-rib thoracic dysplasia 11or 3with or without polydactyly);唾液酸沉积症I型和II型(Sialidosis type I andII);银色痉挛性截瘫综合征(Silver spasticparaplegia syndrome);神经传导速度减慢,常染色体显性(Slowed nerve conduction velocity,autosomal dominant);Smith-Lemli-Opitz综合征;Snyder Robinson综合征;生长激素细胞腺瘤(Somatotroph adenoma);泌乳素瘤(Prolactinoma);家族性垂体腺瘤易感性(familial,Pituitary adenomapredisposition);Sotos综合征1或2(Sotos syndrome 1or2);痉挛性共济失调5,常染色体隐性,Charlevoix-Saguenay型、1、10或11,常染色体隐性(Spastic ataxia 5,autosomal recessive,Charlevoix-Saguenaytype,1,10,or 11,autosomal recessive);肌萎缩性侧索硬化症5型(Amyotrophic lateral sclerosis type 5);痉挛性截瘫15、2、3、35、39、4,常染色体显性,55,常染色体隐性,和5A(Spastic paraplegia 15,2,3,35,39,4,autosomal dominant,55,autosomal recessive,and 5A);先天性胆汁酸合成缺陷3(Bile acid synthesis defect,congenital,3);精子生成障碍11、3和8(Spermatogenic failure 11,3,and 8);球形红细胞症4型和5型(Spherocytosis types 4and 5);球状体肌病(Spheroid body myopathy);脊髓性肌萎缩,下肢显性2,常染色体显性(Spinal muscular atrophy,lower extremity predominant2,autosomal dominant);脊髓性肌萎缩症II型(Spinal muscular atrophy,type II);脊髓小脑性共济失调14、21、35、40和6(Spinocerebellar ataxia 14,21,35,40,and 6);脊髓小脑性共济失调常染色体隐性1和16(Spinocerebellar ataxia autosomal recessive 1and 16);脾发育不全(Splenic hypoplasia);脊椎腕骨跗骨融合综合征(Spondylocarpotarsal synostosis syndrome);脊椎手发育不良,Ehlers-Danlos综合征样,伴免疫失调,Aggrecan型,伴先天性关节脱位,短肢手型,Sedaghatian型,伴锥杆营养不良,Kozlowski型(Spondylocheirodysplasia,Ehlers-Danlos syndrome-like,with immune dysregulation,Aggrecan type,with congenital joint dislocations,short limb-hand type,Sedaghatian type,with cone-rod dystrophy,and Kozlowski type);类扭伤性侏儒症(Parastremmatic dwarfism);Stargardt病1;锥杆营养不良3(Cone-rod dystrophy 3);Stickler综合征1型;Kniest发育不良(Kniest dysplasia);Stickler综合征1型(非综合征性眼疾)和4型(Stickler syndrome,types 1(nonsyndromic ocular)and 4);Sting相关血管病变,婴儿期发病(Sting-associated vasculopathy,infantile-onset);Stormorken综合征;Sturge-Weber综合征,先天性毛细血管畸形1(Sturge-Weber syndrome,Capillary malformations,congenital,1);琥珀酰辅酶A乙酰乙酸转移酶缺乏症(Succinyl-CoA acetoacetate transferase deficiency);蔗糖酶-异麦芽糖酶缺乏症(Sucrase-isomaltase deficiency);婴儿猝死综合征(Sudden infant death syndrome);

孤立型亚硫酸盐氧化酶缺乏症(Sulfite oxidase deficiency,isolated);瓣上主动脉瓣狭窄(Supravalvar aortic stenosis);肺表面活性物质代谢功能障碍2和3(Surfactant metabolism dysfunction,pulmonary,2and 3);近端指关节粘连1b(Symphalangism,proximal,1b);并指(趾)Cenani Lenz型(Syndactyly Cenani Lenz type);并指(趾)3型(Syndactylytype 3);综合征性X连锁智力低下16(Syndromic X-linked mental retardation 16);马蹄内翻足畸形(Talipes equinovarus);Tangier病;TARP综合征;泰萨二氏病,B1变异,Gm2-神经节苷脂沉积症(成人),Gm2-神经节苷脂沉积症(成人发病)(Tay-Sachs disease,B1 variant,Gm2-gangliosidosis(adult),Gm2-gangliosidosis(adult-onset));Temtamy综合征;Tenorio综合征;终末骨发育不良(Terminal osseous dysplasia);睾酮17-β-脱氢酶缺乏症(Testosterone17-beta-dehydrogenase deficiency);先天性四肢切断症,常染色体隐性(Tetraamelia,autosomal recessive);法洛四联症(Tetralogy ofFallot);左心发育不全综合征2(Hypoplastic left heart syndrome 2);动脉干(Truncus arteriosus);心脏和大血管畸形(Malformationofthe heart and greatvessels);室间隔缺损1(Ventricular septal defect 1);Thiel-Behnke角膜营养不良(Thiel-Behnke corneal dystrophy);胸主动脉瘤和主动脉夹层(Thoracic aortic aneurysms and aortic dissections);类马凡氏体态(Marfanoidhabitus);三M综合征2(Three M syndrome 2);血小板减少症、血小板功能障碍、溶血和球蛋白合成不平衡(Thrombocytopenia,platelet dysfunction,hemolysis,and imbalanced globin synthesis);血小板减少症,X连锁(Thrombocytopenia,X-linked);遗传性血栓形成倾向,蛋白C缺乏所致,常染色体显性和隐性(Thrombophilia,hereditary,due to protein C deficiency,autosomal dominant and recessive);甲状腺发育不全(Thyroid agenesis);滤泡性甲状腺癌(Thyroid cancer,follicular);甲状腺激素代谢异常(Thyroid hormone metabolism,abnormal);泛发性甲状腺激素抵抗,常染色体显性(Thyroid hormone resistance,generalized,autosomal dominant);甲亢性周期性麻痹和甲亢性周期性麻痹2(Thyrotoxicperiodicparalysis andThyrotoxic periodicparalysis 2);泛发性促甲状腺激素释放激素抵抗(Thyrotropin-releasing hormone resistance,generalized);Timothy综合征;TNF受体相关周期性发热综合征(TNF receptor-associated periodic feversyndrome(TRAPS));选择性牙齿发育不全3和4(Tooth agenesis,selective,3and 4);扭转型室性心动过速(Torsades de pointes);Townes-Brocks-branchiootorenal样综合征;新生儿暂时性大疱性皮肤溶解症(Transient bullous dermolysis of the newborn);Treacher collins综合征1;毛发肿大伴智力低下、侏儒症和视网膜色素变性(Trichomegaly with mental retardation,dwarfism and pigmentary degeneration of retina);毛发鼻指骨发育不良Ⅰ型(Trichorhinophalangeal dysplasia type I);毛发鼻指骨综合征3型(Trichorhinophalangeal syndrome type 3);三甲基胺尿症(Trimethylaminuria);结节性硬化综合征(Tuberous sclerosis syndrome);淋巴管肌瘤病(Lymphangiomyomatosis);结节性硬化症1和2(Tuberous sclerosis 1and 2);酪氨酸酶阴性眼皮肤白化病(Tyrosinase-negative oculocutaneous albinism);酪氨酸酶阳性眼皮肤白化病(Tyrosinase-positive oculocutaneous albinism);酪氨酸血症I型(Tyrosinemia type

I);UDP葡糖-4-差向异构酶缺乏症(UDPglucose-4-epimerase deficiency);Ullrich先天性肌营养不良(Ullrich congenital muscular dystrophy);尺骨和腓骨缺如伴严重肢体缺陷(Ulna and fibula absence ofwith severe limb deficiency);Upshaw-Schulman综合征;尿苷酸合酶缺乏症(Urocanate hydratase deficiency);Usher综合征1、1B、1D、1G、2A、2C和2D型;视网膜色素变性39(Retinitis pigmentosa 39);UV敏感综合征(UV-sensitive syndrome);Van derWoude综合征;VanMaldergem综合征2;Hennekam淋巴管扩张-淋巴水肿综合征2(Hennekam lymphangiectasia-lymphedema syndrome 2);杂色卟啉症(Variegate porphyria);脑室扩大伴囊性肾病(Ventriculomegaly with cystic kidney disease);Verheij综合征;极长链酰基辅酶A脱氢酶缺乏症(Very long chain acyl-CoA dehydrogenase deficiency);膀胱输尿管反流8(Vesicoureteral reflux 8);内脏异位5,常染色体(Visceral heterotaxy 5,autosomal);内脏肌病(Visceral myopathy);维生素D依赖性佝偻病1型和2型(Vitamin D-dependent rickets,types 1and2);卵黄样营养不良(Vitelliform dystrophy);vonWillebrand病2M型和3型;Waardenburg综合征1、4C和2E型(有神经系统受累);Klein-Waardenberg综合征;Walker-Warburg先天性肌营养不良(Walker-Warburg congenital muscular dystrophy);Warburgmicro综合征2和4;疣、低丙球蛋白血症、感染和骨髓粒细胞缺乏症(Warts,hypogammaglobulinemia,infections,and myelokathexis);Weaver综合征;Weill-Marchesani综合征1和3;Weill-Marchesani样综合征;Weissenbacher-Zweymuller综合征;Werdnig-Hoffmann病;夏-马-图三氏病;Werner综合征;WFS1相关疾病;Wiedemann-Steiner综合征;威尔逊病(Wilson disease);Wolfram样综合征,常染色体显性;Worth病;VanBuchem病2型;着色性干皮病,互补群b、群D、群E和群G;X-连锁无丙种球蛋白血症(Xeroderma pigmentosum,complementation group b,group D,group E,and group G);X连锁遗传性运动和感觉神经病(X-linked agammaglobulinemia;X-linked hereditary motor and sensory neuropathy);X连锁鱼鳞癣伴甾醇硫酸酯酶缺乏症(X-linked ichthyosis with steryl-sulfatase deficiency);X连锁脑室周围异位(X-linked periventricular heterotopia);耳-腭-指综合征综合征I型(Oto-palato-digital syndrome,type I);X连锁严重联合免疫缺陷(X-linked severe combined immunodeficiency);Zimmermann-Laband综合征以及Zimmermann-Laband综合征2;以及板层粉尘状白内障3(Zonularpulverulent cataract 3)。

[1903] 靶核苷酸序列可包含与疾病、障碍或病症相关的靶序列(如,点突变)。靶序列可包含与疾病、障碍或病症相关的T至C(或A至G)点突变,并且其中突变C碱基的脱氨基作用导致错配修复介导的校正与疾病、障碍或病症不相关的序列。靶序列可包含与疾病、障碍或病症相关的G至A(或C至T)点突变,并且其中突变A碱基的脱氨基作用导致错配修复介导的校正与疾病、障碍或病症不相关的序列。靶序列可编码蛋白,并且其中点突变在密码子中并导致突变密码子编码的氨基酸与野生型密码子相比发生变化。靶序列还可位于剪接位点,并且点突变导致mRNA转录本的剪接与野生型转录本相比发生变化。此外,靶标可能位于基因的非编码序列,如启动子,点突变导致基因表达的增加或减少。

[1904] 因此,一些方面,突变C的脱氨基作用导致突变密码子编码的氨基酸变化,这在一些情况下可导致野生型氨基酸的表达。在其他方面,突变A的脱氨基作用导致突变密码子编

码的氨基酸变化,这在一些情况下可导致野生型氨基酸的表达。

[1905] 本文所述涉及使细胞与组合物或rAAV颗粒接触的方法可在体外、离体或体内发生。在某些实施方案中,接触步骤发生在受试者中。在某些实施方案中,受试者已被诊断患有疾病、障碍或病症。

[1906] 在一些实施方案中,本文公开的方法包括使哺乳动物细胞与组合物或rAAV颗粒接触。在特定实施方案中,该方法涉及接触视网膜细胞、皮质细胞或小脑细胞。

[1907] 使用本文所述方法递送的断裂Cas9蛋白或断裂引导编辑器优选具有与原始Cas9蛋白或引导编辑器(即,递送至细胞或在细胞中作为整体表达的未断裂蛋白)相当的活性。例如,断裂Cas9蛋白或断裂引导编辑器保留原始Cas9蛋白或引导编辑器活性的至少50%(如,至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%、或100%)。在一些实施方案中,断裂Cas9蛋白或断裂引导编辑器比原始Cas9蛋白或引导编辑器的活性更强(如,2倍、5倍、10倍、100倍、1000倍或更多)。

[1908] 本文所述组合物可以治疗有效量施用于有需要的受试者来治疗和/或预防受试者患有的疾病或病症。可使用基于CRISPR/Cas9的基因组编辑技术治疗和/或预防的任何疾病或病症都可通过本文所述断裂Cas9蛋白或断裂多瓣引导编辑器治疗。应理解,如果编码断裂Cas9蛋白或多瓣引导编辑器的核苷酸序列不进一步编码gRNA,则可将编码gRNA的单独核酸载体与本文所述组合物一起施用。

[1909] 示例性的合适疾病、病症或病症包括但不限于选自下组的疾病或病症:囊性纤维化、苯丙酮尿症、表皮松解性角化过度症(EHK)、慢性阻塞性肺病(COPD)、Charcot-Marie-Toot病4J型、神经母细胞瘤(NB)、冯·维勒布兰德氏(vonWillebrand)病(vWD)、先天性肌强直、遗传性肾淀粉样变性、扩张型心肌病、遗传性淋巴水肿、家族性阿尔茨海默病、朊病毒病、慢性婴儿神经性皮肤关节综合征(CINCA)、先天性耳聋、尼曼-匹克氏(Niemann-Pick)病C型(NPC)病和结蛋白(desmin)相关肌病(DRM)。在特定实施方案中,疾病或病症是尼曼-匹克氏病C型(NPC)病。

[1910] 在一些实施方案中,疾病、障碍或病症与NPC基因、DNMT1基因、PCSK9基因或TMC1基因中的点突变相关。在某些实施方案中,点突变是NPC中的T3182C突变,其导致I1061T氨基酸替换。

[1911] 在某些实施方案中,点突变是TMC1中的A545G突变,其导致Y182C氨基酸替换。TMC1编码在内耳感觉毛细胞中形成机械敏感离子通道的蛋白,是正常听觉功能所必需的。Y182C氨基酸置换与先天性耳聋有关。

[1912] 在一些实施方案中,疾病、障碍或病症与产生终止密码子的点突变相关,例如,基因编码区内的提前终止密码子。

[1913] 其他示例性疾病、障碍和病症包括囊性纤维化(参见,例如,Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids ofcystic fibrosis patients.Cellstem cell.2013;13:653-658;以及Wu et.al., Correction of a genetic disease in mouse via use of CRISPR-Cas9.Cell stem cell.2013;13:659-662,两者都没有使用脱氨酶融合蛋白来校正遗传缺陷);苯丙酮尿症——例如,苯丙氨酸羟化酶基因中位点835(小鼠)或位点240(人)或同源残基处苯丙氨酸至丝氨酸突变(T>C突变)——参见,例如,McDonald et al.,Genomics.1997;39:402-405;

Bernard-Soulier综合征(BSS)——例如,血小板膜糖蛋白IX中位点55或同源残基处苯丙氨酸至丝氨酸突变,或残基24或同源残基处半胱氨酸至精氨酸突变(T>C突变)——参见,例如,Noris etal.,British Journal of Haematology.1997;97:312-320,and Ali et al.,Hematol.2014;93:381-384;表皮松解性角化过度(EHK)——例如,角蛋白1中位点160或161(如果计算起始蛋氨酸)或同源残基处亮氨酸至脯氨酸突变(T>C突变)——参见,例如,Chipev etal.,Cell.1992;70:821-828,另见UNIPROT数据库(www.uniprot.org)登录号P04264;慢性阻塞性肺疾病(COPD)——例如,α1-抗胰蛋白酶的加工形式位点54或55或同源残基(如果计算起始蛋氨酸)或未加工形式残基78或同源残基处亮氨酸至脯氨酸突变(T>C突变)——参见,例如,Poller et al.,Genomics.1993;17:740-743,另见UNIPROT数据库登录号P01011;Charcot-Marie-Toot病4J型——例如,FIG4中位点41或同源残基处异亮氨酸至苏氨酸突变(T>C突变)——参见,例如Lenk et al.,PLoS Genetics.2011;7:e1002104;成神经细胞瘤(NB)——例如,Caspase-9中位点197或同源残基处亮氨酸至脯氨酸突变(T>C突变)——参见,例如,Kundu et al.,3Biotech.2013,3:225-234;冯·维勒布兰德氏病(vWD)-例如,冯·维勒布兰德因子的加工形式位点509或同源残基处或冯·维勒布兰德因子的未加工形式位点1272或同源残基处半胱氨酸至精氨酸突变(T>C突变)——参见,例如,Lavergne et al.,Br.J.Haematol.1992,82:66-72;另见UNIPROT数据库登录号P04275;先天性肌强直——例如,肌肉氯离子通道基因CLCN1中位点277或同源残基处半胱氨酸至精氨酸突变(T>C突变)——参见,例如,Weinberger et al.,The J.of Physiology.2012,590:3449-3464;遗传性肾淀粉样变性——例如,载脂蛋白AII的加工形式位点78或同源残基或未加工形式位点101或同源残基处终止密码子至精氨酸突变(T>C突变)——参见,例如,Yazaki et al.,Kidney Int.2003;64:11-16;扩张型心肌病(DCM)——例如,FOXD4基因中位点148或同源残基色氨酸至精氨酸突变(T>C突变),参见,例如,Minoretti et.al.,Int.J.of Mol.Med.2007,19:369-372;遗传性淋巴水肿——例如,VEGFR3酪氨酸激酶中位点1035或同源残基处组氨酸至精氨酸突变(A>G突变),参见,例如,Irrthum et al.,Am.J.Hum.Genet.2000,67:295-301;家族性阿尔茨海默病——例如,早老素1中位点143或同源残基处异亮氨酸至缬氨酸突变(A>G突变),参见,例如,Gallo et.al.,J.Alzheimer's disease.2011,25:425-431;朊病毒病——例如,朊病毒蛋白中位点129同源残基处甲硫氨酸至缬氨酸突变(A>G突变)——参见,例如,Lewis et.al.,J.of General Virology.2006,87:2443-2449;慢性婴儿神经性皮肤关节综合征(CINCA)——例如,cryopyrin蛋白中位点位点570或同源残基处酪氨酸至半胱氨酸突变(A>G突变)——参见,例如,Fujisawa et.al.Blood.2007,109:2903-2911;以及结蛋白相关肌病(DRM)——例如,αβ晶状体蛋白中位点120或同源残基处精氨酸至甘氨酸突变(A>G突变)——参见,例如,Kumar et al.,J.Biol.Chem.1999;274:24137-24141。所有参考文献和数据库条目的全部内容通过引用并入本文。

[1914] 三核苷酸重复扩增病

[1915] 三核苷酸重复扩增与许多人类疾病有关,包括亨廷顿病、脆性X综合征和弗里德赖希氏共济失调。最常见的三核苷酸重复包含CAG三联体,但也存在GAA三联体(弗里德赖希氏共济失调)和CGG三联体(脆性X综合征)。继承扩增的倾向或获得已经扩增的亲本等位基因增加患病的可能性。三核苷酸重复的致病性扩增可假设使用多瓣引导编辑来校正。

[1916] 重复区域上游的区域可被RNA引导的核酸酶产生切口，然后用于引发合成包含健康数量重复的新DNA链(取决于特定基因和疾病)，根据图1G或图22中概述的一般机制。在重复序列之后，添加匹配邻近重复序列的另一端的序列(红色链)的同一性的同源短序列。TPRT系统侵入新合成的链，随后用新合成的瓣置换内源性DNA，导致缩减的重复等位基因。术语“缩减”是指核苷酸重复区长度的缩短，从而导致三核苷酸重复区的修复。

[1917] 本文所述多瓣引导编辑系统(例如，双瓣和四瓣引导编辑系统)可用于缩减三核苷酸重复突变(或“三联体扩增疾病”)，以治疗诸如亨廷顿病和其他三核苷酸重复病症的病症。三核苷酸重复扩增病是复杂的进行性障碍，涉及发育神经生物学，通常影响认知和感觉运动功能。这些障碍显示出遗传早现(即，每一代都增加严重程度)。DNA扩增或缩减通常以减数分裂方式发生(即，在配子发生期间或胚胎发育早期)，并且通常具有性别偏倚，这意味着某些基因仅在通过女性遗传时才会扩增，而其他基因则仅通过男性遗传。在人类中，三核苷酸重复扩增病会导致转录或翻译水平的基因沉默，这本质上破坏基因功能。或者，三核苷酸重复扩增病会导致改变的蛋白产生大的重复氨基酸序列，通常以显性失活的方式(如，多聚谷氨酰胺疾病)消除或改变蛋白功能。

[1918] 不希望受理论的束缚，三联体扩增由DNA复制期间或DNA修复合成期间的滑移引起。由于串联重复序列彼此具有相同的序列，两条DNA链之间的碱基配对可在序列的多个点发生。这可能会导致在DNA复制或DNA修复合成过程中形成“环出”结构。这可能导致重复序列的重复复制，扩大重复数目。已提出了涉及杂交RNA:DNA中间体的其他机制。多瓣引导编辑可用于通过删除一个或多个或有问题的重复密码子三联体来减少或消除这些三联体扩增区。在这种应用的实施方案中，图23提供了用引导编辑缩减或减少三核苷酸重复序列的PEgRNA设计的示意图。

[1919] 可实施多瓣引导编辑以通过用包含适当靶向切割位点的PEgRNA的引导编辑器对三联体重复区域上游的区域产生切口来缩减三联体扩增区。然后，引导编辑器基于PEgRNA作为模板(即，其编辑模板)合成新的DNA链(ssDNA瓣)，该模板链编码健康数量的三联体重复(取决于特定的基因和疾病)。还合成包含健康三联体重复序列的ssDNA链的新合成的ssDNA链，以包括匹配邻近重复序列的另一端的序列(红色链)的同源短序列(即，同源臂)。新合成链的侵入，以及随后用新合成的ssDNA瓣置换内源性DNA，导致缩减重复等位基因。

[1920] 取决于特定的三核苷酸扩增病，缺陷诱导三联体扩增可发生在“三核苷酸重复扩增蛋白”中。三核苷酸重复扩增蛋白是与发生三核苷酸重复扩增病的易感性、三核苷酸重复扩增病的存在、三核苷酸重复扩增病的严重程度或其任何组合相关的一组不同的蛋白。三核苷酸重复扩增病根据重复序列的类型分为两类。最常见的重复是三联体CAG，当其存在于基因的编码区时，编码氨基酸谷氨酰胺(Q)。因此，这些疾病被称为聚谷氨酰胺(poly Q)疾病且包括以下疾病：亨廷顿病(HD)；脊髓延髓肌萎缩症(SBMA)；脊髓小脑性共济失调(SCA1、2、3、6、7和17型)；和齿状核红核苍白球路易体萎缩(DRPLA)。其余的三核苷酸重复扩增病不涉及CAG三联体或CAG三联体不在基因的编码区，因此被称为非聚谷氨酰胺疾病。非聚谷氨酰胺疾病包括脆性X综合征(FRAXA)；脆性XE智力低下(FRAXE)；弗里德赖希氏共济失调(FRDA)；肌强直性营养不良(DM)；以及脊髓小脑性共济失调(SCA 8型和12型)。

[1921] 可基于与三核苷酸重复扩增病相关的蛋白与三核苷酸重复扩增病的实验关联来选择与三核苷酸重复扩增病相关的蛋白。例如，相对于没有三核苷酸重复扩增病的群体，与

三核苷酸重复扩增病相关的蛋白的生产率或循环浓度在具有三核苷酸重复扩增病的群体中可能升高或降低。可使用蛋白组学技术评估蛋白水平的差异，包括但不限于Western印迹、免疫组织化学染色、酶联免疫吸附测定(ELISA)和质谱法。或者，可通过使用基因组技术获得编码蛋白的基因的基因表达谱来鉴定与三核苷酸重复扩增病相关的蛋白，包括但不限于DNA微阵列分析、基因表达系列分析(SAGE)和实时定量聚合酶链反应(Q-PCR)。

[1922] 与三核苷酸重复扩增病相关的蛋白的非限制性示例可通过多瓣引导编辑来校正，包括AR(雄激素受体)、FMR1(脆性X智力低下1)、HTT(亨廷顿蛋白)、DMPK(肌强直性肌营养不良症-蛋白激酶)、FXN(frataxin)、ATXN2(ataxin2)、ATN1(atrophin1)、FEN1(瓣结构特异性核酸内切酶1)、TNRC6A(含有6A的三核苷酸重复序列)、PABPN1(poly(A)结合蛋白、核1)、JPH3(亲联蛋白3)、MED15(中介复合体亚基15)、ATXN1(ataxin1)、ATXN3(ataxin3)、TBP(TATA盒结合蛋白)、CACNA1A(钙通道、电压依赖性P/Q型，α1A亚基)、ATXN8OS(ATXN8反链(非蛋白编码))、PPP2R2B(蛋白磷酸酶2、调节亚基B、β)、ATXN7(ataxin7)、TNRC6B(含6B的三核苷酸重复)、TNRC6C(含6C的三核苷酸重复)、CELF3(CUGBP，Elav-样家族成员3)、MAB21L1(mab-21-样1(秀丽隐杆线虫))、MSH2(mutS同源物2，结肠癌，非息肉病1型(大肠杆菌))、TMEM185A(跨膜蛋白185A)、SIX5(SIX同源框5)、CNPY3(冠层3同源物(斑马鱼))、FRAXE(脆弱部位、叶酸型、罕见、fra(X)(q28)E)、GNB2(鸟嘌呤核苷酸结合蛋白(G蛋白)、β多肽2)、RPL14(核糖体蛋白L14)、ATXN8(ataxin8)、INSR(胰岛素受体)、TTR(转甲状腺素蛋白)、EP400(E1A结合蛋白p400)、GIGYF2(GRB10相互作用的GYF蛋白2)、OGG1(8-氧鸟嘌呤DNA糖苷酶)、STC1(斯钙素1)、CNDP1(肌肽二肽酶1(金属肽酶M20家族))、C10orf2(染色体10开放阅读框2)、MAML3(mastermind-样3(果蝇))、DKC1(先天性角化不良1，角化不良蛋白)、PAXIP1(PAX相互作用(与转录激活域)蛋白1)、CASK(钙/钙调蛋白依赖性丝氨酸蛋白激酶(MAGUK家族))、MAPT(微管相关蛋白tau)、SP1(Sp1转录因子)、POLG(聚合酶(DNA定向)、γ)、AFF2(AF4/FMR2家族，成员2)、THBS1(血小板反应蛋白1)、TP53(肿瘤蛋白p53)、ESR1(雌激素受体1)、CGGBP1(CGG三联体重复结合蛋白1)、ABT1(基础转录激活因子1)、KLK3(激肽释放酶相关肽酶3)、PRNP(朊病毒蛋白)、JUN(jun致癌基因))，KCNN3(钾中/小电导钙激活通道，亚家族N，成员3)，BAX(BCL2相关X蛋白)，FRAXA(脆弱部位，叶酸型，罕见，fra(X)(q27.3)A(巨睾丸病，智力低下)、KBTBD10(kelch重复和BTB(POZ)域包含10个)、MBNL1(盲肌样(果蝇))、RAD51(RAD51同源物(RecA同源物，大肠杆菌)(酿酒酵母))、NCOA3(核受体共激活因子3)、ERDA1(扩增重复结构域，CAG/CTG1)、TSC1(结节性硬化症1)、COMP(软骨寡聚基质蛋白)、GCLC(谷氨酸-半胱氨酸连接酶，催化亚基)，RRAD(与糖尿病相关的Ras)、MSH3(mutS同源物3(大肠杆菌))、DRD2(多巴胺受体D2)、CD44(CD44分子(印度血型))、CTCF(CCCTC结合因子(锌指蛋白)))、CCND1(细胞周期蛋白D1)、CLSPN(claspin同源物(非洲爪蟾)、MEF2A(肌细胞增强因子2A)、PTPRU(蛋白酪氨酸磷酸酶，受体类型，U)、GAPDH(3-磷酸甘油醛脱氢酶)、TRIM22(三重基序含22)、WT1(威尔姆氏(Wilms)肿瘤1)、AHR(芳烃受体)、GPX1(谷胱甘肽过氧化物酶1)、TPMT(硫嘌呤S-甲基转移酶)、NDP(诺里病(假神经胶质瘤))、ARX(无芒相关同源框)、MUS81(MUS81核酸内切酶同源物(酿酒酵母))、TYR(酪氨酸酶(眼皮肤白化病IA))、EGR1(早期生长反应蛋白1)、UNG(尿嘧啶-DNA糖基化酶)、NUMBL(numb同源物(果蝇)样)、FABP2(脂肪酸结合蛋白2，肠道)、EN2(engrailed同源框2)、CRYGC(晶状体蛋白、γC)、SRP14(信号识别颗粒14kDa(同源Alu RNA结合蛋白)、CRYGB(晶状体蛋白、γB)、PDCD1(程序性细胞死亡1)、HOXA1

(同源框A1)、ATXN2L(ataxin2样)、PMS2(PMS2减数分裂后分离增加2(酿酒酵母))、GLA(半乳糖苷酶,α)、CBL(Cas-Br-M(鼠)亲嗜性逆转录病毒转化序列)、FTH1(铁蛋白,重多肽1)、IL12RB2(白细胞介素12受体,β2)、OTX2(orthodenticle同源框2)、HOXA5(同源框A5)、POLG2(聚合酶(DNA定向)、γ2、辅助亚基)、DLX2(无远端同源框2)、SIRPA(信号调节蛋白)α)、OTX1(orthodenticle同源框1)、AHRR(芳烃受体阻遏物)、MANF(中脑星形胶质细胞衍生的神经营养因子)、TMEM158(跨膜蛋白158(基因/假基因))和ENSG00000078687。

[1923] 在特定方面,本公开提供了基于TPRT的方法,其用于治疗被诊断患有扩增重复病(又被称为重复扩增病症或三核苷酸重复病)的受试者。当微卫星重复扩增超过阈值长度时发生扩增重复病。目前,至少有30种遗传性疾病被认为是由重复扩增引起的。20世纪90年代初期,随着发现三核苷酸重复是几种主要遗传性疾病的基础,包括脆性X、脊髓和延髓肌萎缩症、肌强直性营养不良和亨廷顿病(Nelson et al,“The unstable repeats-three evolving faces of neurological disease,”Neuron,March 6,2013,Vol.77;825-843,其通过引用并入本文),以及HawRiver综合征、Jacobsen综合征、齿状核红核苍白球路易体萎缩(DRPLA)、Machado-Joseph病、并指(趾)多指(趾)畸形(SPDII)、手足生殖器综合征(HFGS)、锁骨颅骨发育不全(CCD)、前脑无裂畸形(HPE)、先天性中枢性低通气综合征(CCHS)、ARX非综合征性X连锁智力低下(XLMR)和眼咽肌营养不良(OPMD)(参见,发现微卫星重复不稳定性是这些病症的标志,正如预期,每一代连续发生重复扩增的现象,导致后代出现更严重的表型和更早的发病年龄。认为重复扩增通过几种不同的机制引起疾病。即,扩增可能在基因、mRNA转录本和/或编码的蛋白水平上干扰细胞功能。在某些情况下,突变经由通过沉默包含重复基因的功能丧失机制作用。在其他情况下,疾病是由功能获得机制引起的,即mRNA转录本或蛋白具有新的异常功能。

[1924] 在一个实施方案中,治疗三核苷酸重复病的方法描述于图23中。一般地,该方法涉及使用TPRT基因组编辑(即,多瓣引导编辑)与延伸的gRNA相组合,该gRNA包含编码期望且健康的置换三核苷酸重复序列的区域,该区域旨在通过引导编辑加工机制置换内源性患病三核苷酸重复序列。用于缩减三核苷酸重复序列和利用TPRT基因组编辑(即,多瓣引导编辑)的三核苷酸重复缩减的示例性gRNA设计的示意图示于图23。

[1925] 朊病毒病

[1926] 多瓣引导编辑还可通过将一个或多个保护性突变安装到疾病过程中错误折叠的朊病毒蛋白(PRNP)来预防或阻止朊病毒病的进展。朊病毒病或传染性海绵状脑病(TSE)是影响人类和动物的罕见进行性神经退行性疾病家族。它们的特点是潜伏期长、与神经元丢失相关的特征性海绵状变化以及不能诱导炎症反应。

[1927] 在人类中,朊病毒病包括克雅氏病(CJD)、变异型克雅氏病(vCJD)、杰茨曼-斯脱司勒-史茵克综合征、致命性家族性失眠症和库鲁病。在动物中,朊病毒病包括牛海绵状脑病(BSE或“疯牛病”)、慢性消耗性疾病(CWD)、羊瘙痒症、传染性水貂脑病、猫海绵状脑病和有蹄类海绵状脑病。多瓣引导编辑可用于将保护性点突变安装到朊病毒蛋白,以预防或阻止这些朊病毒病中任何一种疾病的进展。

[1928] 经典CJD是一种人朊病毒病。它是神经退行性疾病,具有典型的临床和诊断特征。这种疾病进展迅速并且总是致命的。感染这种疾病通常会在发病后1年内导致死亡。CJD是快速进展且总是致命的神经退行性疾病,据信是由被称为朊病毒蛋白的细胞糖蛋白的异常

同种型引起的。CJD在世界范围内发生，据报道包括美国在内的许多国家的估计年发病率约为每百万人口一例。绝大多数CJD患者通常在发病1年内死亡。将CJD与人类和动物中发生的其他朊病毒病一起归类为传染性海绵状脑病(TSE)。在约85％的患者中，CJD作为散发性疾病发生，没有可识别的传播模式。由于朊病毒蛋白基因的遗传突变，一小部分患者(5％至15％)会发展为CJD。这些遗传形式包括杰茨曼-斯脱司勒-史茵克综合征和致命性家族性失眠症。目前对CJD尚无已知的治疗方法。

[1929] 变异型克雅氏病(vCJD)是1996年首次在英国被描述的一种朊病毒病。现在有强有力的科学证据表明，导致奶牛朊病毒病，牛海绵状脑病(BSE或“疯牛病”)爆发的病原体与导致人vCJD爆发的病原体相同。变异型CJD(vCJD)与经典CJD(通常简称为CJD)不同。它具有不同于经典CJD的临床和病理特征。每种疾病还具有朊病毒蛋白基因的特定遗传特征。这两种疾病都是致命的脑部疾病，潜伏期非常长，以年为单位，并且由被称为朊病毒的非常规传染性因子引起。目前对vCJD尚无已知的治疗方法。

[1930] BSE(牛海绵状脑病或“疯牛病”)是牛的进行性神经系统疾病，其由被称为朊病毒的罕见传染性因子感染引起。传染性因子的性质还不是很清楚。目前，最公认的理论是该因子是被称为朊病毒蛋白的正常蛋白的修饰形式。由于尚不清楚的原因，正常朊病毒蛋白会转变为致病(有害)形式，然后损害牛的中枢神经系统。越来越多的证据表明存在不同的BSE株系：导致英国爆发的典型或经典BSE株系和两种非典型株系(H和L株)。目前对BSE尚无已知的治疗方法。

[1931] 慢性消耗性疾病(CWD)是影响鹿、麋鹿、驯鹿、梅花鹿和驼鹿的朊病毒病。已在北美一些地区包括加拿大和美国、挪威和韩国发现。受感染的动物可能需要一年多的时间才会出现症状，包括体重急剧下降(消瘦)、跌跌撞撞、无精打采和其他神经系统症状。CWD可影响所有年龄的动物，一些受感染的动物可能在没有患病的情况下死亡。CWD对动物是致命的，并且没有治疗方法或疫苗。

[1932] 认为TSE的病原体是朊病毒。术语“朊病毒”是指异常的病原体，它们是可传播的，并且能够诱导在大脑中含量最高的被称为朊病毒蛋白的特定正常细胞蛋白的异常折叠。这些正常朊病毒蛋白的功能仍未完全了解。朊病毒蛋白的异常折叠导致脑损伤和疾病的特征性体征和症状。朊病毒病通常进展迅速且总是致命的。

[1933] 如本文所用，术语“朊病毒”是指已知会引起人类和动物疾病(海绵状脑病)的传染性颗粒。术语“朊病毒”是“蛋白”和“感染”这两个词的缩写，并且颗粒大部分(如果不是全部)由表达$PRNP^{C}$的PRNP基因编码的$PRNP^{Sc}$分子组成，$PRNP^{C}$构象改变成为$PRNP^{Sc}$。朊病毒不同于细菌、病毒和类病毒。已知的朊病毒包括感染动物引起羊瘙痒症(绵羊和山羊神经系统的可传染的退行性疾病)以及牛海绵状脑病(BSE)或疯牛病和猫海绵状脑病的那些。如上所述，已知影响人类的四种朊病毒病是(1)库鲁病，(2)克雅氏病(CJD)，(3)杰茨曼-斯脱司勒-史茵克病(GSS)，以及(4)致命性家族性失眠症(FFI)。如本文所用，朊病毒包括在所用的任何动物，特别是在人和家养农场动物中引起所有或任何这些疾病或其他疾病的所有形式的朊病毒。

[1934] 一般地且不希望受理论的束缚，先前疾病是由朊病毒蛋白的错误折叠引起的。此类疾病——通常被称为沉积病——朊病毒蛋白的错误折叠可解释如下。如果A是正常合成的基因产物，在单体或寡聚状态下发挥预期的生理作用，A*是A的构象激活形式，能够发生

明显的构象变化,B是倾向于多聚体组装的构象改变状态(即,形成沉积物的错误折叠形式),Bn是具有致病性且相对难以回收的多聚体材料。对于朊病毒病,$PRNP^{C}$和$PRNP^{Sc}$对应于状态A和Bn,其中A主要是螺旋和单体,而Bn是富含β的多聚体。

[1935] 已知朊病毒蛋白中的某些突变可能与先前疾病的风险增加有关。相反地,朊病毒蛋白中的某些突变在本质上具有保护作用。参见,Bagynszky et al.,"Characterization of mutations inPRNP(prion)gene and their possible roles in neurodegenerative diseases,"Neuropsychiatr Dis Treat.,2018;14:2067-2085,其内容通过引用并入本文。

[1936] PRNP(NCBI参考序列号NP_000302.1(SEQ ID NO:291))——人朊病毒蛋白——由位于20号染色体16kb长的基因编码(4686151-4701588)。它包含两个外显子,外显子2携带编码253个氨基酸(AA)长的PrP蛋白的开放阅读框。外显子1是非编码外显子,可用作转录起始位点。翻译后修饰导致前22个AA N端片段(NTF)和最后23个AAC端片段(CTF)的去除。NTF在PrP转运到内质网(ER)后裂解,而CTF(糖基磷脂酰肌醇[GPI]信号肽[GPI-SP])被GPI锚裂解。GPI锚可能参与PrP蛋白转运。它还可能发挥将朊病毒蛋白附着至细胞膜外表面的作用。正常的PrP由长N端环(包含八肽重复区域)、两个短β折叠、三个α螺旋和C端区域(包含GPI锚)组成。PrP的裂解产生208个AA长的糖蛋白,锚定于细胞膜。

[1937] PRNP(NP_000302.1)的253个氨基酸序列如下:

[1938] MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV(SEQ ID NO:291).

[1939] PRNP(NP_000302.1)的253个氨基酸序列由以下核苷酸序列(NCBI参考序列号NM_000311.5,"智人朊病毒蛋白(PRNP),转录本变体1,mRNA)编码,如下:

[1940] GCGAACCTTGGCTGCTGGATGCTGGTTCTCTTTGTGGCCACATGGAGTGACCTGGGCCTCTGCAAGAAGCGCCCGAAGCCTGGAGGATGGAACACTGGGGGCAGCCGATACCCGGGGCAGGGCAGCCCTGGAGGCAACCGCTACCCACCTCAGGGCGGTGGTGGCTGGGGGCAGCCTCATGGTGGTGGCTGGGGGCAGCCTCATGGTGGTGGCTGGGGGCAGCCCCATGGTGGTGGCTGGGGACAGCCTCATGGTGGTGGCTGGGGTCAAGGAGGTGGCACCCACAGTCAGTGGAACAAGCCGAGTAAGCCAAAAACCAACATGAAGCACATGGCTGGTGCTGCAGCAGCTGGGGCAGTGGTGGGGGGCCTTGGCGGCTACATGCTGGGAAGTGCCATGAGCAGGCCCATCATACATTTCGGCAGTGACTATGAGGACCGTTACTATCGTGAAAACATGCACCGTTACCCCAACCAAGTGTACTACAGGCCCATGGATGAGTACAGCAACCAGAACAACTTTGTGCACGACTGCGTCAATATCACAATCAAGCAGCACACGGTCACCACAACCACCAAGGGGGAGAACTTCACCGAGACCGACGTTAAGATGATGGAGCGCGTGGTTGAGCAGATGTGTATCACCCAGTACGAGAGGGAATCTCAGGCCTATTACCAGAGAGGATCGAGCATGGTCCTCTTCTCCTCTCCACCTGTGATCCTCCTGATCTCTTTCCTCATCTTCCTGATAGTGGGATGAGGAAGGTCTTCCTGTTTTCACCATCTTTCTAATCTTTTTCCAGCTTGAGGGAGGCGGTATCCACCTGCAGCCCTTTTAGTGGTGGTGTCTCACTCTTTCTTCTCTCTTTGTCCCGGATAGGCTAATCAATACCCTTGGCACTGATGGGCACTGGAAAACATAGAGTAGACCTGAGATGCTGGTCAAGCCCCCTTTGATTGAGTTCATCATGAGCCGTTGCTAATGCCAGGCCAGTAAAAGTATAACAGCAAATAACCATTGGTTAATCTGGACTTATTTTTGGACTTAGTGCAACAGGTTGAGGCTAAAACAAATCTCAGAACAGTCTGAAATACCTTTGCCTGGATACCTCTGGCTCCTTCAGCAGCTAGAGCTCAGTATACTAATGCCCTATCTTAGTAGAGATTTCATAGCTATTTAGAGATATTTTCCATTTTAAGAAAACCCGACAACATTTCTGCCAGGTTTGTTAGGAGGCCACATGATACTTATTCAAAAAAATCCTAGAGATTCTTAGCTCTTGGG

ATGCAGGCTCAGCCCGCTGGAGCATGAGCTCTGTGTGTACCGAGAACTGGGGTGATGTTTTACTTTTCACAGTATG
GGCTACACAGCAGCTGTTCAACAAGAGTAAATATTGTCACAACACTGAACCTCTGGCTAGAGGACATATTCACAGT
GAACATAACTGTAACATATATGAAAGGCTTCTGGGACTTGAAATCAAATGTTTGGGAATGGTGCCCTTGGAGGCAA
CCTCCCATTTTAGATGTTTAAAGGACCCTATATGTGGCATTCCTTTCTTTAAACTATAGGTAATTAAGGCAGCTGA
AAAGTAAATTGCCTTCTAGACACTGAAGGCAAATCTCCTTTGTCCATTTACCTGGAAACCAGAATGATTTTGACAT
ACAGGAGAGCTGCAGTTGTGAAAGCACCATCATCATAGAGGATGATGTAATTAAAAAATGGTCAGTGTGCAAAGAA
AAGAACTGCTTGCATTTCTTTATTTCTGTCTCATAATTGTCAAAAACCAGAATTAGGTCAAGTTCATAGTTTCTGT
AATTGGCTTTTGAATCAAAGAATAGGGAGACAATCTAAAAAATATCTTAGGTTGGAGATGACAGAAATATGATTGA
TTTGAAGTGGAAAAAGAAATTCTGTTAATGTTAATTAAAGTAAAATTATTCCCTGAATTGTTTGATATTGTCACCT
AGCAGATATGTATTACTTTTCTGCAATGTTATTATTGGCTTGCACTTTGTGAGTATTCTATGTAAAAATATATATG
TATATAAAATATATATTGCATAGGACAGACTTAGGAGTTTTGTTTAGAGCAGTTAACATCTGAAGTGTCTAATGCA
TTAACTTTTGTAAGGTACTGAATACTTAATATGTGGGAAACCCTTTTGCGTGGTCCTTAGGCTTACAATGTGCACT
GAATCGTTTCATGTAAGAATCCAAAGTGGACACCATTAACAGGTCTTTGAAATATGCATGTACTTTATATTTTCTA
TATTTGTAACTTTGCATGTTCTTGTTTTGTTATATAAAAAAATTGTAAATGTTTAATATCTGACTGAAATTAAACG
AGCGAAGATGAGCACCA(SEQ ID NO:292)

[1941] 目前报道的与CJD和FFI相关的相对于PRNP(NP_000302.1)的突变位点如下。这些突变可使用本文公开的多瓣引导编辑器去除或安装。

[1942]

突变	与 CJD 朊病毒病相关的突变 PRNP 的氨基酸序列(参见 BAGYNSZKY ET AL., 2018 的表 1) (相对于 PRNP NP_000302.1 的 SEQ ID NO: 291)
D178N	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF

[1943]

	GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHNCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 293)
T188K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHKVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 294)
E196K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGKNFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 295)
E196A	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGANFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 296)
E200K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTKTDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 297)
E200G	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTGTDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 298)
V203I	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDIKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPVI LLISFLIFLIVG (SEQ ID NO: 299)
R208H	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMEHVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 300)
V210I	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVIEQMCITQYERESQAYYQRGSSMVLFSSPPVI LLISFLIFLIVG (SEQ ID NO: 301)

[1944]

E211Q	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVQQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 302)
M232R	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSRVLFSSPPVI LLISFLIFLIVG (SEQ ID NO: 303)

[1945] 报道了与GSS相关的相对于PRNP(NP_000302.1)(SEQ ID NO:291)的突变位点,如下:

[1946]

突变	与 GSS 朊病毒病相关的突变 PRNP 的氨基酸序列(参见 BAGYNSZKY ET AL., 2018 的表 2) (相对于 PRNP NP_000302.1 的 SEQ ID NO: 291)
P102L	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKLSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 304)
P105L	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKLKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 305)
A117V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAVAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 306)
G131V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLVSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 307)
V176G	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFGHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 308)
H187R	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQRTVT

[1947]

	TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 309)
	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 291)
F198S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENSTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 311)
D202N	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETNVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 312)
Q212P	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEPMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 313)
Q217R	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITRYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 314)
M232T	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSTVLFSSPPVI LLISFLIFLIVG (SEQ ID NO: 315)

[1948] 与可能对朊病毒病的保护性质相关的相对于PRNP(NP_000302.1)(SEQ ID NO: 291)的突变位点,如下:

[1949]

突变	与对朊病毒病的保护性质相关的突变 PRNP 的氨基酸序列(参见 BAGYNSZKY ET AL., 2018 的表 4)(相对于 PRNP NP_000302.1 的 SEQ ID NO: 291)
G127S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGSYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 316)
G127V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR

[1950]

	YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGVYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 317)
M129V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYVLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 318)
D167G	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMGEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 319)
D167N	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMNEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 320)
N171S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSSQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 321)
E219K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYKRESQAYYQRGSSMVLFSSPPV ILLISFLIFLIVG (SEQ ID NO: 322)
P238S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNR YPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGG GTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHF GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSSPV ILLISFLIFLIVG (SEQ ID NO: 323)

[1951] 因此，在不同实施方案中，可使用多瓣引导编辑来去除与朊病毒病相关的PRNP中的突变或在PRNP中安装被认为对朊病毒病具有保护作用的突变。例如，多瓣引导编辑可用于去除或恢复D178N、V180I、T188K、E196K、E196A、E200K、E200G、V203I、R208H、V210I、E211Q、I215V或M232R突变(相对于NP_000302.1的PRNP)(SEQ ID NO:291)。在其他实施方案中，可使用多瓣引导编辑来去除或恢复PRNP蛋白中的P102L、P105L、A117V、G131V、V176G、H187R、F198S、D202N、Q212P、Q217R或M232T突变(相对于NP_000302.1的PRNP)(SEQ ID NO:291)。通过使用引导编辑去除或校正PRNP中此类突变的存在，可降低或消除朊病毒病的风险。

[1952] 在其他实施方案中，可使用多瓣引导编辑在PRNP中安装保护性突变，该突变与对一种或多种朊病毒病的保护作用相关。例如，可将引导编辑用于在PRNP中安装G127S、

G127V、M129V、D167G、D167N、N171S、E219K或P238S保护性突变(相对于NP_000302.1的PRNP)(SEQ ID NO:291)。在其他实施方案中,保护性突变可以是安装在PRNP中G127、G127、M129、D167、D167、N171、E219或P238处的任何替代氨基酸(相对于NP_000302.1的PRNP)(SEQ ID NO:291)。

[1953] 在特定实施方案中,可将多瓣引导编辑用于在PRNP中安装G127V保护性突变,如图27所示,并在实施例5中讨论。

[1954] 在另一个实施方案中,可将引导编辑用于在PRNP中安装E219K保护性突变。

[1955] PRNP蛋白和保护性突变位点在哺乳动物中是保守的,因此除了治疗人类疾病外,它还可用于产生对朊病毒病免疫的牛和羊,甚至帮助治愈患有朊病毒病的野生动物种群。已将引导编辑用于在人细胞中实现约25%的天然保护性等位基因的安装,先前小鼠实验表明,这种安装水平足以引起对大多数朊病毒病的免疫力。这种方法是首次且目前可能唯一的在大多数细胞类型中以如此高的效率安装该等位基因的方法。另一种可能的治疗策略是使用引导编辑或多瓣引导编辑通过在基因中安装早期终止密码子来减少或消除PRNP的表达。

[1956] 使用本文所述PEgRNA设计原理,可设计合适的PEgRNA来安装期望的保护性突变,或从PRNP中去除朊病毒病相关突变。例如,可将下面的PEgRNA列表用于安装人PRNP中的G127V保护性等位基因和E219K保护性等位基因,以及各种动物PRNP中的G127V保护性等位基因。

[1957] [9]药物组合物

[1958] 本公开的其他方面涉及包含本文所述多瓣引导编辑系统的各种组件(如,包括但不限于napDNAbps、逆转录酶、融合蛋白(如,包含napDNAbps和逆转录酶)、延伸的向导RNA和包含融合蛋白和延伸的向导RNA的复合物,以及辅助元件(如,第二链切口组件和5'内源性DNA瓣去除核酸内切酶,用于帮助驱动多瓣引导编辑过程向着编辑产物形成)中任一种的药物组合物。

[1959] 如本文所用,术语"药物组合物"是指配制用于制药用途的组合物。在一些实施方案中,药物组合物进一步包含药学上可接受的载体。在一些实施方案中,药物组合物包含其他药剂(如,用于特异性递送、增加半衰期或其他治疗化合物)。

[1960] 如本文所用,术语"药学上可接受的载体"是指药学上可接受的材料、组合物或赋形剂,例如液体或固体填充剂、稀释剂、赋形剂、制造助剂(如,润滑剂、滑石镁、硬脂酸钙或锌、或硬脂酸)或参与将化合物从身体的一个部位(如,递送部位)携带或运输到另一个部位(如,器官、组织或身体的一部分)的溶剂包封材料。药学上可接受的载体在与制剂的其他成分相容并且对受试者的组织没有伤害的意义上是"可接受的"(如,生理相容的、无菌的、生理pH等)。可用作药学上可接受的载体的材料的一些示例包括:(1)糖类,如乳糖、葡萄糖和蔗糖;(2)淀粉,如玉米淀粉、马铃薯淀粉;(3)纤维素及其衍生物,如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素、醋酸纤维素等;(4)粉状黄芪胶;(5)麦芽;(6)明胶;(7)润滑剂,如硬脂酸镁、十二烷基硫酸钠、滑石粉等;(8)可可脂、栓剂蜡等赋形剂;(9)油类,如花生油、棉籽油、红花油、香油、橄榄油、玉米油、豆油等;(10)二醇类,如丙二醇;(11)多元醇,如甘油、山梨糖醇、甘露糖醇和聚乙二醇(PEG);(12)油酸乙酯、月桂酸乙酯等酯类;(13)琼脂;(14)缓冲剂,如氢氧化镁、氢氧化铝等;(15)海藻酸;(16)无热原水;(17)等渗盐水;(18)

林格氏(Ringer's)溶液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯、聚碳酸酯和/或聚酐;(22)填充剂,如多肽和氨基酸;(23)血清成分,如血清白蛋白、高密度脂蛋白和低密度脂蛋白;(22)C2-C12醇类,如乙醇;(23)用于药物制剂的其他无毒相容物质。润湿剂、着色剂、脱模剂、涂层剂、甜味剂、调味剂、加香剂、防腐剂和抗氧化剂也可存在于制剂中。诸如"赋形剂"、"载体"、"药学上可接受的载体"等术语在本文中可互换使用。

[1961] 在一些实施方案中,药物组合物被配制用于递送至受试者,例如用于基因编辑。施用本文所述药物组合物的合适途径包括但不限于:局部、皮下、经皮、皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、牙内、耳蜗内、经鼓膜、器官内、硬膜外、鞘内、肌肉内、静脉内、血管内、骨内、眼周、瘤内、脑内和脑室内施用。

[1962] 在一些实施方案中,将本文所述药物组合物局部施用于患病部位(如,肿瘤部位)。在一些实施方案中,本文所述药物组合物通过注射、通过导管、通过栓剂或通过植入物向受试者施用,植入物为多孔、无孔或凝胶状材料,包括膜如唾液酸膜,或纤维。

[1963] 在其他实施方案中,本文所述药物组合物在控释系统中递送。在一个实施方案中,可使用泵(参见,例如,Langer,1990,Science 249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald et al.,1980,Surgery 88:507;Saudek et al.,1989,N.Engl.J.Med.321:574)。在另一个实施方案中,可使用聚合材料(参见,例如,Medical Applications of Controlled Release(Langer and Wise eds.,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolen and Ball eds.,Wiley,New York,1984);Ranger and Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61.还参见Levy et al.,1985,Science 228:190;During et al.,1989,Ann.Neurol.25:351;Howard et al.,1989,J.Neurosurg.71:105)。其他控释系统在例如Langer(同上)中讨论。

[1964] 在一些实施方案中,根据常规程序将药物组合物配制成适合于静脉内或皮下向受试者如人施用的组合物。在一些实施方案中,用于通过注射施用的药物组合物是无菌等渗水性缓冲液中的溶液。必要时,药物还可包括增溶剂和局部麻醉剂(如,利多卡因),以缓解注射部位的疼痛。通常,成分以单位剂型单独提供或混合在一起提供,例如,作为显示活性剂量的密封容器如安瓿或小袋中的干燥冻干粉或无水浓缩物。当药物通过输液施用时,可用装有无菌药用级水或盐水的输液瓶进行配药。当药物组合物通过注射施用时,可提供无菌注射用水或盐水的安瓿,以便在施用前可混合成分。

[1965] 用于全身施用的药物组合物可以是液体,如无菌盐水、乳酸林格氏溶液或汉克氏(Hank's)溶液。此外,药物组合物可以是固体形式并在使用前立即重新溶解或悬浮。还考虑了冻干形式。

[1966] 药物组合物可包含在脂质颗粒或囊泡中,例如脂质体或微晶,其也适用于肠胃外施用。颗粒可具有任何合适的结构,如单层或多层,只要其中包含组合物。化合物可被包裹在含有融合脂质二油酰磷脂酰乙醇胺(DOPE)、低水平(5-10mol%)阳离子脂质核并通过聚乙二醇(PEG)涂层稳定的"稳定型质粒-脂质颗粒"(SPLP)中(ZhangY.P.et al.,Gene Ther.1999,6:1438-47)。带正电荷的脂质如N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵或"DOTAP"特别优选用于此类颗粒和囊泡。这些脂质颗粒的制备是众所周知的。参见,例如美国专利号4,880,635;4,906,477;4,911,928;4,917,951;4,920,016;和4,921,

757;其各自通过引用并入本文。

[1967] 例如,本文所述药物组合物可以单位剂量施用或包装。当用于本公开的药物组合物时,术语“单位剂量”是指适合用于受试者的单位剂量的物理上离散的单位,每个单位包含经计算以产生期望的治疗效果的预定量的活性物质结合所需的稀释剂;即,载体或媒介物。

[1968] 此外,药物组合物可作为药物试剂盒提供,其包括(a)含有冻干形式的本发明化合物的容器,和(b)含有药学上可接受的注射用稀释剂(如,无菌水)的第二容器。药学上可接受的稀释剂可用于复溶或稀释本发明的冻干化合物。任选与这种容器相关联的可以是由管理药物或生物产品的制造、使用或销售的政府机构规定形式的通知,该通知反映了用于人类施用的制造、使用或销售机构的批准。

[1969] 另一方面,包括含有可用于治疗上述疾病的材料的制品。在一些实施方案中,制品包括容器和标签。合适的容器包括如瓶子、小瓶、注射器和试管。容器可由多种材料形成,例如玻璃或塑料。在一些实施方案中,容器容纳有效治疗本文所述疾病的组合物并且可具有无菌进入孔。例如,容器可以是静脉内溶液袋或具有可被皮下注射针刺穿的塞子的小瓶。组合物中的活性剂是本发明的化合物。在一些实施方案中,容器上或与容器相关的标签表明组合物用于治疗疾病的选择。制品还可包括第二容器,其包含药学上可接受的缓冲液,例如磷酸盐缓冲盐水、林格氏溶液或葡萄糖溶液。从商业和用户的角度来看,它还可包括其他所需的材料,包括其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明的包装插页。

[1970] [11]试剂盒、细胞、载体和递送

[1971] 试剂盒

[1972] 本公开的组合物可组装成试剂盒。在一些实施方案中,试剂盒包含用于表达本文所述的多瓣引导编辑器的核酸载体。在其他实施方案中,试剂盒还包含合适的向导核苷酸序列(如,PEgRNA和第二位点gRNA)或用于表达此类向导核苷酸序列的核酸载体,以将Cas9蛋白或引导编辑器靶向期望的靶序列。

[1973] 本文所述试剂盒可包括一个或多个容器,其容纳用于执行本文所述方法的组分和任选的使用说明。本文所述任何试剂盒还可包括执行测定方法所需的组分。在适用的情况下,试剂盒的各组分可以液体形式(如,溶液)或固体形式(如,干粉)提供。在某些情况下,一些组分可以是可复溶的或以其他方式可加工的(如,成为活性形式),例如,通过添加合适的溶剂或其他物质(如,水),这些物质可与试剂盒一起提供或不与试剂盒一起提供提供。

[1974] 在一些实施方案中,试剂盒可任选包括使用所提供组分的说明书和/或宣传。如本文所用,“说明”可标明说明和/或宣传的组成部分,并且通常涉及关于本公开的包装或与本公开的包装相关联的书面说明。说明还可包括以任何方式提供的任何口头或电子说明,使得用户将清楚地认识到说明将与试剂盒相关联,例如,视听(如,录像带、DVD等)、互联网和/或基于网络的通讯等。书面说明可采用监管药品或生物制品制造、使用或销售的政府机构规定的形式,这也可反映制造、使用或销售机构对动物施用的批准。如本文所用,“宣传”包括所有开展业务的方法,包括教育方法、医院和其他临床指导、科学探究、药物发现或开发、学术研究、制药行业活动(包括药品销售)以及任何广告或其他促销活动,包括与本公开相关的任何形式的书面、口头和电子通讯。此外,如本文所述,试剂盒可根据具体应用包括其他组分。

[1975] 试剂盒可在一个或多个容器中包含本文所述任何一种或多种组分。组分可以无菌制备，包装在注射器中并冷藏运输。或者，其可装在小瓶或其他容器中储存。第二容器可具有无菌制备的其他组分。或者，试剂盒可包括在小瓶、管或其他容器中预先混合和运输的活性剂。

[1976] 试剂盒可具有多种形式，例如泡罩袋、收缩包装袋、真空密封袋、密封热成型托盘或类似的袋或托盘形式，其中配件松散地包装在袋，一个或多个管子、容器、盒子或袋子内。在添加附件之后可对试剂盒进行灭菌，从而允许容器中的各个附件以其他方式打开。可使用任何合适的灭菌技术对试剂盒进行灭菌，例如辐射灭菌、热灭菌或本领域已知的其他灭菌方法。试剂盒还可根据具体应用包括其他组分，例如容器、细胞培养基、盐、缓冲液、试剂、注射器、针头、用于施加或去除消毒剂的织物(如，纱布)、一次性手套、施用前用于试剂的支撑物等。本公开的一些方面提供了包含核酸构建体的试剂盒，所述核酸构建体包含编码本文所述多瓣引导编辑系统(例如，双引导编辑或四引导编辑)的各种组件的核苷酸序列(如，包括但不限于napDNAbp、逆转录酶、聚合酶、融合蛋白(如，包含napDNAbps和逆转录酶(或更广泛地，聚合酶)、延伸的向导RNA和包含融合蛋白和延伸的向导RNA的复合物，以及辅助元件，如第二链切口组分(如，第二链切口产生性gRNA)和5’内源性DNA瓣去除核酸内切酶，用于帮助驱动多瓣引导编辑过程向着编辑产物形成)。在一些实施方案中，核苷酸序列包含驱动多瓣引导编辑系统组件表达的异源启动子(或多于一个启动子)。

[1977] 本公开的其他方面提供了包含一种或多种核酸构建体的试剂盒，所述核酸构建体编码本文所述多瓣引导编辑系统的各种组件，例如，包含编码能够修饰靶DNA序列的多瓣引导编辑系统的组件的核苷酸序列。在一些实施方案中，核苷酸序列包含驱动多瓣引导系统组件表达的异源启动子。

[1978] 本公开的一些方面提供了包含核酸构建体的试剂盒，所述核酸构建体包含(a)编码融合到逆转录酶的napDNAbp(如，Cas9结构域)的核苷酸序列，和(b)驱动(a)的序列表达的异源启动子。

[1979] 细胞

[1980] 可包含本文所述任何组合物的细胞包括原核细胞和真核细胞。本文所述方法用于将Cas9蛋白或多瓣引导编辑器递送至真核细胞(如，哺乳动物细胞如人细胞)。在一些实施方案中，细胞是体外的(如，培养的细胞)。在一些实施方案中，细胞是体内的(如，在受试者如人受试者中)。在一些实施方案中，细胞是离体的(如，分离自受试者且可被施用回同一或不同受试者)。

[1981] 本公开的哺乳动物细胞包括人细胞、灵长类细胞(如，vero细胞)、大鼠细胞(如，GH3细胞、OC23细胞)或小鼠细胞(如，MC3T3细胞)。存在多种人细胞系，包括但不限于人胚胎肾(HEK)细胞、HeLa细胞、来自美国国家癌症研究所(National Cancer Institute)60种癌细胞系(NCI60)的癌细胞、DU145(前列腺癌)细胞、Lncap(前列腺癌)细胞、MCF-7(乳腺癌)细胞、MDA-MB-438(乳腺癌)细胞、PC3(前列腺癌)细胞、T47D(乳腺癌)细胞、THP-1(急性髓系白血病)细胞、U87(胶质母细胞瘤)细胞、SHSY5Y人神经母细胞瘤细胞(克隆自骨髓瘤)和Saos-2(骨癌)细胞。在一些实施方案中，将rAAV载体递送至人胚胎肾(HEK)细胞(如，HEK 293或HEK 293T细胞)中。在一些实施方案中，将rAAV载体递送至干细胞(如，人干细胞)中，例如多能干细胞(如，人多能干细胞，包括人诱导多能干细胞(hiPSC))。干细胞是指能够在培养中

无限期分裂并产生特化细胞的细胞。多能干细胞是指能够分化为生物体所有组织的一类干细胞，但不能单独维持完整的生物体发育。人诱导多能干细胞是指通过被迫表达对维持胚胎干细胞的定义特性重要的基因和因子而被重编程为胚胎干细胞样状态的体细胞(如，成熟或成体)细胞(参见，例如，Takahashi and Yamanaka,Cell 126(4):663-76,2006，通过引用并入本文)。人诱导多能干细胞表达干细胞标志物，并且能够产生具有所有三个胚层(外胚层、内胚层、中胚层)特征的细胞。

[1982] 可根据本公开使用的细胞系的其他非限制性示例包括293-T、293-T、3T3、4T1、721、9L、A-549、A172、A20、A253、A2780、A2780ADR、A2780cis、A431、ALC、B16、B35、BCP-1、BEAS-2B、bEnd.3、BHK-21、BR293、BxPC3、C2C12、C3H-10T1/2、C6、C6/36、Cal-27、CGR8、CHO、CMLT1、CMT、COR-L23、COR-L23/5010、COR-L23/CPR、COR-L23/R23、COS-7、COV-434、CT26、D17、DH82、DU145、DuCaP、E14Tg2a、EL4、EM2、EM3、EMT6/AR1、EMT6/AR10.0、FM3、H1299、H69、HB54、HB55、HCA2、Hepa1c1c7、HighFive细胞、HL-60、HMEC、HT-29、HUVEC、JUVEC、Jurkat、JY细胞、K562细胞、KCL22、KG1、Ku812、KYO1、LNCap、Ma-Mel 1、2、3....48、MC-38、MCF-10A、MCF-7、MDA-MB-231、MDA-MB-435、MDA-MB-468、MDCKII、MG63、MONO-MAC 6、MOR/0.2R、MRC5、MTD-1A、MyEnd、NALM-1、NCI-H69/CPR、NCI-H69/LX10、NCI-H69/LX20、NCI-H69/LX4、NIH-3T3、NW-145、OPCN/OPCT Peer、PNT-1A/PNT 2、PTK2、Raji、RBL细胞、RenCa、RIN-5F、RMA/RMAS、S2、Saos-2细胞、Sf21、Sf9、SiHa、SKBR3、SKOV-3、T-47D、T2、T84、THP1、U373、U87、U937、VCaP、WM39、WT-49、X63、YAC-1和YAR细胞。

[1983] 本公开的一些方面提供了包含本文公开的任何构建体的细胞。在一些实施方案中，宿主细胞用本文所述一种或多种载体瞬时或非瞬时转染。在一些实施方案中，细胞在受试者中自然发生转染。在一些实施方案中，被转染的细胞取自受试者。在一些实施方案中，细胞来源于取自受试者的细胞，如细胞系。用于组织培养的多种细胞系是本领域已知的。细胞系的示例包括但不限于C8161、CCRF-CEM、MOLT、mIMCD-3、NHDF、HeLa-S3、Huh1、Huh4、Huh7、HUVEC、HASMC、HEKn、HEKa、MiaPaCell、Panc1、PC-3、TF1、CTLL-2、C1R、Rat6、CV1、RPTE、A10、T24、J82、A375、ARH-77、Calu1、SW480、SW620、SKOV3、SK-UT、CaCo2、P388D1、SEM-K2、WEHI-231、HB56、TIB55、Jurkat、J45.01、LRMB、Bcl-1、BC-3、IC21、DLD2、Raw264.7、NRK、NRK-52E、MRC5、MEF、Hep G2、HeLa B、HeLaT4、COS、COS-1、COS-6、COS-M6A、BS-C-1猴肾上皮、BALB/3T3小鼠胚胎成纤维细胞、3T3 Swiss、3T3-L1、132-d5人胎成纤维细胞;10.1小鼠成纤维细胞、293-T、3T3、721、9L、A2780、A2780ADR、A2780cis、A 172、A20、A253、A431、A-549、ALC、B16、B35、BCP-12B、BCP-12B、3、BHK-21、BR 293。BxPC3。C3H-10T1/2、C6/36、Cal-27、CHO、CHO-7、CHO-IR、CHO-K1、CHO-K2、CHO-T、CHO Dhfr-/-、COR-L23、COR-L23/CPR、COR-L23/5010、COR-L23/R23、COS-7、COV-434、CMLT1、CMT、CT26、D17、DH82、DU145、DuCaP、EL4、EM2、EM3、EMT6/AR1、EMT6/AR10.0、FM3、H1299、H69、HB54、HB55、HCA2、HEK-293、HeLa、Hepa1c1c7、HL-60、HMEC、HT-29、Jurkat、JY细胞、K562细胞、Ku812、KCL22、KG1、KYO1、Ma-Mel 1-48、MC-38、MCF-7、MCF-10A、MDA-MB-231、MDA-MB-468、MDA-MB-435、MDCK II、MDCK 11、MOR/0.2R、MONO-MAC 6、MTD-1A、MyEnd、NCI-H69/CPR、NCI-H69/LX10、NCI-H69/LX20、NCI-H69/LX4、NIH-3T3、NALM-1、NW-145、OPCN/OPCT细胞系、Peer、PNT-1A/PNT 2、RenCa、RIN-5F、RMA/RMAS、Saos-2细胞、Sf-9、SkBr3、T2、T-47D、T84、THP1细胞系、U373、U87、U937、VCaP、Vero细胞、WM39、WT-49、X63、YAC-1、YAR及其转基因变体。

[1984] 细胞系可从本领域技术人员已知的多种来源获得(参见,例如美国典型培养物保藏中心(ATCC)(Manassus,VA))。在一些实施方案中,用本文所述一种或多种载体转染的细胞用于建立包含一种或多种载体衍生序列的新细胞系。在一些实施方案中,将用如本文所述CRISPR系统组件瞬时转染(如通过一种或多种载体的瞬时转染,或用RNA转染)并通过CRISPR复合物的活性修饰的细胞用于建立新的细胞系,其中新的细胞系包含含有修饰但缺乏任何其他外源序列的细胞。在一些实施方案中,将用一种或多种本文所述载体瞬时或非瞬时转染的细胞,或源自此类细胞的细胞系用于评估一种或多种测试化合物。

[1985] 载体

[1986] 本公开的一些方面涉及使用重组病毒载体(如,腺相关病毒载体、腺病毒载体或单纯疱疹病毒载体)将本文所述的多瓣引导编辑器或其组件(如,断裂Cas9蛋白或断裂核碱基多瓣引导编辑器)递送至细胞。在断裂PE方法的情况下,PE融合蛋白的N端部分和PE融合蛋白的C端部分由单独的重组病毒载体(如,腺相关病毒载体、腺病毒载体或单纯疱疹病毒载体)递送至同一细胞,因为全长Cas9蛋白或多瓣引导编辑器超过各种病毒载体的包装限制,如rAAV(约4.9kb)。

[1987] 因此,在一个实施方案中,本公开考虑能够递送断裂多瓣引导编辑器融合蛋白或其断裂组件的载体。在一些实施方案中,提供了用于将断裂Cas9蛋白或断裂引导编辑器递送至细胞(如,哺乳动物细胞、人细胞)的组合物。在一些实施方案中,本公开的组合物包含:(i)第一重组腺相关病毒(rAAV)颗粒,其包含编码Cas9蛋白的N端部分或在其C端与内含肽N融合的引导编辑器的第一核苷酸序列;和(ii)第二重组腺相关病毒(rAAV)颗粒,其包含编码内含肽-C融合至Cas9蛋白或引导编辑器的C端部分的N端的第二核苷酸序列。本公开的rAAV颗粒包含包封在病毒衣壳蛋白中的rAAV载体(即,rAAV的重组基因组)。

[1988] 在一些实施方案中,rAAV载体包含:(1)异源核酸区域,其包含编码本文所述任何形式的断裂Cas9蛋白或断裂多瓣引导编辑器的N-端部分或C-端部分的第一或第二核苷酸序列;(2)一个或多个核苷酸序列,其包含促进异源核酸区域(如,启动子)表达的序列,和(3)一个或多个核酸区域,其包含促进异源核酸区域(任选具有包含促进表达的序列的一个或多个核酸区域)整合至细胞基因组的序列。在一些实施方案中,促进整合的病毒序列包括反向末端重复(ITR)序列。在一些实施方案中,编码断裂Cas9蛋白或断裂多瓣引导编辑器的N-端部分或C-端部分的第一或第二核苷酸序列在每一侧的侧翼有ITR序列。在一些实施方案中,核酸载体还包含编码如本文所述AAV Rep蛋白的区域,其包含在侧翼有ITR的区域内或该区域外。ITR序列可源自任何AAV血清型(如,1、2、3、4、5、6、7、8、9或10)或可源自多于一种血清型。在一些实施方案中,ITR序列源自AAV2或AAV6。

[1989] 因此,在一些实施方案中,本文公开的rAAV颗粒包括至少一种rAAV2颗粒、rAAV6颗粒、rAAV8颗粒、rPHP.B颗粒、rPHP.eB颗粒、或rAAV9颗粒、或其变体。在特定实施方案中,所公开的rAAV颗粒是rPHP.B颗粒、rPHP.eB颗粒、rAAV9颗粒。

[1990] ITR序列和含有ITR序列的质粒是本领域已知的并且可商购获得(参见,例如可获自以下的产品或服务:Vector Biolabs,Philadelphia,PA;Cellbiolabs,San Diego,CA;Agilent Technologies,Santa Clara,Ca;and Addgene,Cambridge,MA;以及Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein.Kessler PD,PodsakoffGM,Chen X,McQuiston SA,

Colosi PC,Matelis LA,Kurtzman GJ,Byrne BJ.Proc Natl Acad Sci USA.1996Nov 26;93(24):14082-7;以及Curtis A.Machida.Methods in Molecular Medicine™.ViralVectors for Gene Therapy Methods and Protocols.10.1385/1-59259-304-6:201©Humana Press Inc.2003.Chapter 10.Targeted Integration by Adeno-Associated Virus.Matthew D.Weitzman,Samuel M.Young Jr.,Toni Cathomen and Richard Jude Samulski;美国专利号5,139,941和5,962,313,所有这些都通过引用并入本文)。

[1991] 在一些实施方案中,本公开的rAAV载体包含一种或多种调控元件来控制异源核酸区域(如,启动子、转录终止子和/或其他调控元件)的表达。在一些实施方案中,第一和/或第二核苷酸序列与一个或多个(如,1、2、3、4、5或更多个)转录终止子可操作地连接。可根据本公开使用的转录终止子的非限制性示例包括牛长激素基因(bGH)、人生长激素基因(hGH)、SV40、CW3、Φ或其组合的转录终止子。已测试了几种转录终止子的效率,以确定它们各自对断裂Cas9蛋白或断裂多瓣引导编辑器的表达水平的影响。在一些实施方案中,本公开使用的转录终止子是bGH转录终止子。在一些实施方案中,rAAV载体还包含土拨鼠肝炎病毒转录后调控元件(WPRE)。在某些实施方案中,WPRE是截短的WPRE序列,如"W3"。在一些实施方案中,WPRE插入转录终止子的5'。这些序列在转录时产生三级结构,该结构增强了表达,特别是来自病毒载体的表达。

[1992] 在一些实施方案中,本文使用的载体可编码PE融合蛋白或其任何组件(如,napDNAbp、接头或聚合酶)。此外,本文使用的载体可编码PEgRNA和/或用于第二链切口的辅助gRNA。载体可能能够驱动细胞中一种或多种编码序列的表达。在一些实施方案中,细胞可以是原核细胞,例如细菌细胞。在一些实施方案中,细胞可以是真核细胞,例如酵母、植物、昆虫或哺乳动物细胞。在一些实施方案中,真核细胞可以是哺乳动物细胞。在一些实施方案中,真核细胞可以是啮齿动物细胞。在一些实施方案中,真核细胞可以是人细胞。在不同类型的细胞中驱动表达的合适启动子是本领域已知的。在一些实施方案中,启动子可以是野生型的。在其他实施方案中,可对启动子进行修饰以进行更有效的表达。在其他实施方案中,启动子可被截短但保留其功能。例如,启动子可具有适合将载体正确包装到病毒中的正常尺寸或减小的尺寸。

[1993] 在一些实施方案中,可用于引导编辑器载体的启动子可为组成型、诱导型或组织特异性的。在一些实施方案中,启动子可为组成型启动子。非限制的示例性组成型启动子包括巨细胞病毒立即早期启动子(CMV)、猿猴病毒(SV40)启动子、腺病毒主要晚期(MLP)启动子、劳斯肉瘤病毒(RSV)启动子、小鼠乳腺肿瘤病毒(MMTV)启动子、磷酸甘油酸激酶(PGK)启动子、延伸因子-α(EFla)启动子、泛素启动子、肌动蛋白启动子、微管蛋白启动子、免疫球蛋白启动子、其功能片段,或任何前述的组合。在一些实施方案中,启动子可为CMV启动子。在一些实施方案中,启动子可为截短的CMV启动子。在其他实施方案中,启动子可为EFla启动子。在一些实施方案中,启动子可为诱导型启动子。非限制的示例性诱导型启动子包括可通过热休克、光、化学品、肽、金属、类固醇、抗生素或酒精诱导的那些启动子。在一些实施方案中,诱导型启动子可为具有低基础(非诱导)表达水平的启动子,如Tet-On®启动子(Clontech)。在一些实施方案中,启动子可为组织特异性启动子。在一些实施方案中,组织特异性启动子在肝脏组织中排他地或主要地表达。非限制的示例性组织特异性启动子包括

B29启动子、CD14启动子、CD43启动子、CD45启动子、CD68启动子、desmin启动子、弹性蛋白酶-1启动子、内皮糖蛋白启动子、纤连蛋白启动子、Flt-1启动子、GFAP启动子、GPIIb启动子、ICAM-2启动子、INF-β启动子、Mb启动子、Nphs1启动子、OG-2启动子、SP-B启动子、SYN1启动子和WASP启动子。

[1994] 在一些实施方案中，引导编辑器载体(如，包括编码引导编辑器融合蛋白和/或PEgRNA和/或辅助第二链切口产生性gRNA的任何载体)可包含诱导型启动子以仅在其被递送至靶细胞之后开始表达。非限制的示例性诱导型启动子包括可通过热休克、光、化学品、肽、金属、类固醇、抗生素或酒精诱导的那些启动子。在一些实施方案中，诱导型启动子可为具有低基础(非诱导)表达水平的启动子，如Tet-On®启动子(Clontech)。

[1995] 在其他实施方案中，引导编辑器载体(如，包括编码引导编辑器融合蛋白和/或PEgRNA和/或辅助第二链切口产生性gRNA的任何载体)可包含组织特异性启动子以仅在其被递送至特定组织后开始表达。非限制的示例性组织特异性启动子包括B29启动子、CD14启动子、CD43启动子、CD45启动子、CD68启动子、结蛋白启动子、弹性蛋白酶-1启动子、内皮糖蛋白启动子、纤连蛋白启动子、Flt-1启动子、GFAPpro启动子、GPIIb启动子、ICAM-2启动子、INF-β启动子、Mb启动子、Nphs1启动子、OG-2启动子、SP-B启动子、SYN1启动子和WASP启动子。

[1996] 在一些实施方案中，编码PEgRNA(或与多瓣引导编辑相关使用的任何向导RNA)的核苷酸序列可与至少一个转录或翻译控制序列可操作地连接。在一些实施方案中，编码向导RNA的核苷酸序列可与至少一个启动子可操作地连接。在一些实施方案中，启动子可被RNA聚合酶III(Pol III)识别。Pol III启动子的非限制性示例包括U6、HI和tRNA启动子。在一些实施方案中，编码向导RNA的核苷酸序列可与小鼠或人U6启动子可操作地连接。在其他实施方案中，编码向导RNA的核苷酸序列可与小鼠或人HI启动子可操作地连接。在一些实施方案中，编码向导RNA的核苷酸序列可与小鼠或人tRNA启动子可操作地连接。在具有多于一种向导RNA的实施方案中，用于驱动表达的启动子可以相同或不同。在一些实施方案中，编码向导RNA的crRNA的核苷酸和编码向导RNA的tracrRNA的核苷酸可在同一载体上提供。在一些实施方案中，编码crRNA的核苷酸和编码tracrRNA的核苷酸可由相同的启动子驱动。在一些实施方案中，crRNA和tracrRNA可被转录成单一转录本。例如，crRNA和tracrRNA可由单一转录本加工形成双分子向导RNA。或者，crRNA和tracrRNA可被转录成单分子向导RNA。

[1997] 在一些实施方案中，编码向导RNA的核苷酸序列可位于包含编码PE融合蛋白的核苷酸序列的同一载体上。在一些实施方案中，向导RNA和PE融合蛋白的表达可由它们相应的启动子驱动。在一些实施方案中，向导RNA的表达可由驱动PE融合蛋白表达的同一启动子驱动。在一些实施方案中，向导RNA和PE融合蛋白转录本可包含在单一转录本内。例如，向导RNA可在Cas9蛋白转录物的非翻译区(UTR)内。在一些实施方案中，向导RNA可在PE融合蛋白转录本的5'UTR内。在其他实施方案中，向导RNA可在PE融合蛋白转录本的3'UTR内。在一些实施方案中，PE融合蛋白转录物的细胞内半衰期可通过在其3'UTR内包含向导RNA并由此缩短其3'UTR的长度而减少。在其他实施方案中，向导RNA可在PE融合蛋白转录本的内含子内。在一些实施方案中，可在向导RNA所在的内含子处添加合适的剪接位点，使得向导RNA从转录本正确剪接出来。在一些实施方案中，Cas9蛋白和在同一载体上紧密接近的向导RNA的表达可促进CRISPR复合物的更有效形成。

[1998] 多瓣引导编辑器载体系统可包括一个载体、或两个载体、或三个载体、或四个载体、或五个载体、或更多。在一些实施方案中，载体系统可包括一种单一载体，其编码PE融合蛋白和PEgRNA。在其他实施方案中，载体系统可包括两种载体，其中一种载体编码PE融合蛋白，另一种编码PEgRNA。在其他实施方案中，载体系统可包括三个载体，其中第三种载体编码本文方法中使用的第二链切口产生性gRNA。

[1999] 在一些实施方案中，包含rAAV颗粒(以本文中考虑的任何形式)的组合物还包含药学上可接受的载体。在一些实施方案中，将组合物配制在合适的药物载体中来施用于人或动物受试者。

[2000] 可用作药学上可接受的载体的材料的一些示例包括：(1)糖类，如乳糖、葡萄糖和蔗糖；(2)淀粉，如玉米淀粉、马铃薯淀粉；(3)纤维素及其衍生物，如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素、醋酸纤维素等；(4)粉状黄芪胶；(5)麦芽；(6)明胶；(7)润滑剂，如硬脂酸镁、十二烷基硫酸钠、滑石粉等；(8)可可脂、栓剂蜡等赋形剂；(9)油类，如花生油、棉籽油、红花油、香油、橄榄油、玉米油、豆油等；(10)二醇类，如丙二醇；(11)多元醇，如甘油、山梨糖醇、甘露糖醇和聚乙二醇(PEG)；(12)油酸乙酯、月桂酸乙酯等酯类；(13)琼脂；(14)缓冲剂，如氢氧化镁、氢氧化铝等；(15)海藻酸；(16)无热原水；(17)等渗盐水；(18)林格氏(Ringer's)溶液；(19)乙醇；(20)pH缓冲溶液；(21)聚酯、聚碳酸酯和/或聚酐；(22)填充剂，如多肽和氨基酸；(23)血清成分，如血清白蛋白、高密度脂蛋白和低密度脂蛋白；(22)C2-C12醇类，如乙醇；(23)用于药物制剂的其他无毒相容物质。润湿剂、着色剂、脱模剂、涂层剂、甜味剂、调味剂、加香剂、防腐剂和抗氧化剂也可存在于制剂中。诸如“赋形剂”、“载体”、“药学上可接受的载体”等术语在本文中可互换使用。

[2001] 递送方法

[2002] 一些方面，本发明提供了包括将一种或多种编码本文所述多瓣引导编辑器的不同组分的多核苷酸，例如本文所述一种或多种载体、其一种或多种转录本和/或由其转录的一种或多种蛋白递送至宿主细胞的方法。在一些方面，本发明还提供了通过此类方法产生的细胞，以及包含此类细胞或由此类细胞产生的生物体(如，动物、植物或真菌)。在一些实施方案中，如本文所述碱基编辑器与向导序列组合(并且任选与向导序列复合)递送至细胞。

[2003] 示例性递送策略在本文其他地方描述，其包括基于载体的策略、PE核糖核蛋白复合物递送和通过mRNA方法递送PE。

[2004] 在一些实施方案中，提供的递送方法包括核转染、显微注射、基因枪、病毒体、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸DNA、人工病毒体和试剂增强的DNA摄取。

[2005] 示例性的核酸递送方法包括脂质转染、核转染、电穿孔、稳定基因组整合(如，piggybac)、显微注射、基因枪、病毒体、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸DNA、人工病毒粒体和试剂增强的DNA摄取。脂质转染描述于例如美国专利号5,049,386,4,946,787;和4,897,355中，脂质转染试剂可商购获得(如，Transfectam™、Lipofectin™和SFCellLine4D-NucleofectorX Kit™(Lonza)))。适用于多核苷酸的有效受体识别脂质转染的阳离子和中性脂质包括Feigner在WO91/17424;WO91/16024中的那些。递送可以是细胞(如，体外或离体施用)或靶组织(如，体内施用)。递送可通过使用RNP复合物来实现。

[2006] 脂质:核酸复合物的制备，包括靶向脂质体，如免疫脂质复合物，是本领域技术人员所熟知的(参见，例如，Crystal,Science 270:404-410(1995);Blaese et al.,Cancer

Gene Ther.2:291-297(1995);Behr et al.,Bioconjugate Chem.5:382-389(1994);Remy et al.,Bioconjugate Chem.5:647-654(1994);Gao et al.,Gene Therapy 2:710-722(1995);Ahmad et al.,Cancer Res.52:4817-4820(1992);美国专利号4,186,183、4,217,344、4,235,871、4,261,975、4,485,054,4,501,728、4,774,085、4,837,028和4,946,787)。

[2007] 在其他实施方案中,本文提供的递送方法和载体是RNP复合物。融合蛋白的RNP递送显著增强了碱基编辑的DNA特异性。融合蛋白的RNP递送导致中靶和脱靶DNA编辑的分离。RNP递送消除了非重复位点的脱靶编辑,同时保持了与质粒递送相当的中靶编辑,并且即使在高度重复的VEGFA位点2上也大大减少了脱靶DNA编辑。参见,Rees,H.A.et al.,Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery,Nat.Commun.8,15790(2017),2016年12月27日授权的美国专利号9,526,784和2017年8月22日授权的美国专利号9,737,604,其各自通过引用并入本文。

[2008] 将核酸递送至细胞的其他方法是本领域技术人员已知的。参见,例如US2003/0087817,通过引用并入本文。

[2009] 本公开的其他方面提供了将多瓣引导编辑器构建体递送至细胞以在细胞内形成完整的功能性引导编辑器的方法。例如,在一些实施方案中,使细胞与本文所述组合物接触(如,包含编码断裂Cas9或断裂引导编辑器的核苷酸序列的组合物或包含含有此类核苷酸序列的核酸载体的AAV颗粒)。在一些实施方案中,接触导致此类核苷酸序列递送到细胞中,其中Cas9蛋白或引导编辑器的N端部分和Cas9蛋白或引导编辑器的C端部分在细胞内表达并连接形成完整的Cas9蛋白或完整的引导编辑器。

[2010] 应当理解,本文提供的任何rAAV颗粒、核酸分子或组合物可以任何合适的方式稳定或瞬时地引入细胞。在一些实施方案中,所公开的蛋白可被转染至细胞中。在一些实施方案中,细胞可用核酸分子转导或转染细胞。例如,可用编码断裂蛋白的核酸分子或含有编码一种或多种核酸分子的病毒基因组的rAAV颗粒转导(如,用编码断裂蛋白的病毒)或转染(如,用编码断裂蛋白的质粒)。这种转导可以是稳定或瞬时转导。在一些实施方案中,表达断裂蛋白或含有断裂蛋白的细胞可用一种或多种向导RNA序列转导或转染,例如递送断裂Cas9(如,nCas9)蛋白时。在一些实施方案中,可通过电穿孔、瞬时(如,脂质转染)和稳定基因组整合(如,piggybac)和病毒转导或本领域技术人员已知的其他方法将表达断裂蛋白的质粒引入细胞。

[2011] 在某些实施方案中,本文提供的组合物包含脂质和/或聚合物。在某些实施方案中,脂质和/或聚合物是阳离子的。此类脂质颗粒的制备是众所周知的。见,例如美国专利号4,880,635;4,906,477;4,911,928;4,917,951;4,920,016;4,921,757;和9,737,604,其各自通过引用并入本文。

[2012] 向导RNA序列的长度可以是15-100个核苷酸并且包含与靶核苷酸序列互补的至少10、至少15或至少20个连续核苷酸的序列。向导RNA可包含15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个与靶核苷酸序列互补的连续核苷酸。向导RNA的长度可为15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸。

[2013] 在一些实施方案中,靶核苷酸序列是基因组中的DNA序列,如真核基因组。在某些

实施方案中,靶核苷酸序列在哺乳动物(如,人)基因组中。

[2014] 例如,本公开的组合物可以单位剂量施用或包装。当用于本公开的药物组合物时,术语“单位剂量”是指适合作为受试者的单位剂量的物理离散单位,每个单位含有与所需的稀释剂(即,载体或媒介物)结合的预定量的活性物质,经计算以产生期望需的治疗效果。

[2015] 疾病或病症的治疗包括延迟疾病的发展或进展,或降低疾病的严重性。治疗疾病不一定需要治愈结果。

[2016] 如本文所用,“延迟”疾病的发展是指推迟、阻碍、减缓、阻止、稳定和/或推迟疾病的进展。这种延迟可以是不同的时间长度,这取决于疾病的历史和/或被治疗的个体。“延迟”或减轻疾病发展或延迟疾病发作的方法是与不使用该方法相比,降低在给定时间范围内出现疾病的一种或多种症状的可能性和/或减轻在给定时间范围内症状程度的方法。此类比较通常基于临床研究,使用的受试者数量足以得出统计学上有意义的结果。

[2017] 疾病的“发展”或“进展”是指疾病的初始表现和/或随后的进展。可使用本领域公知的标准临床技术检测和评估疾病的发展。然而,发展也指可能无法检测到的进展。出于本公开的目的,发展或进展是指症状的生物学过程。“发展”包括发生、复发和发作。

[2018] 如本文所用,疾病的“发作”或“发生”包括初始发作和/或复发。根据待治疗的疾病类型或疾病部位,可使用医学领域普通技术人员已知的常规方法向受试者施用分离的多肽或药物组合物。

[2019] 无需进一步阐述,相信本领域技术人员可基于以上描述充分利用本公开。因此,以下具体实施方案应理解为仅以任何方式说明而非限制本公开的其余部分。出于本文引用的目的或主题,本文引用的所有出版物均通过引用并入。

实施例

[2020] 实施例1.用于在基因组中安装精确核苷酸变化的引导编辑(PE)

[2021] 目标是开发用于在哺乳动物基因组中精确且通用安装单核苷酸变化的转化基因组编辑技术。这项技术将允许研究人员研究几乎任何哺乳动物基因中单核苷酸变异的影响,并潜在实现校正人类患者中的致病性点突变的治疗干预。

[2022] 采用规律成簇间隔短回文重复序列(CRISPR)系统进行基因组编辑已彻底改变了生命科学[1-3]。尽管使用CRISPR进行基因破坏现已成为常规,但单核苷酸编辑的精确安装仍然是重大挑战,尽管对于研究或校正大量致病性突变是必要的。同源定向修复(HDR)能够实现此类编辑,但效率低下(通常<5%)、需要供体DNA修复模板以及双链DNA断裂(DSB)形成的有害影响。最近,Liu的实验室开发了碱基编辑,无需DSB即可实现高效的单核苷酸编辑。碱基编辑器(BE)将CRISPR系统与碱基修饰脱氨酶结合,将靶C·G或A·T碱基对分别转化为A·T或G·C[4-6]。尽管已被全世界的研究人员广泛使用(由Addgene分发了>5,000个Liu实验室BE构建体),但当前的BE仅实现十二种可能的碱基对转换中的四种,并且无法校正小的插入或缺失。此外,碱基编辑的靶向范围受到邻近靶碱基的非靶C或A碱基编辑(“旁观者(bystander)编辑”)以及PAM序列距靶碱基15±2bp存在的要求限制。因此,克服这些限制将大大拓宽基因组编辑的基础研究和治疗应用。

[2023] 在此,建议开发新的精确编辑方法,该方法提供碱基编辑的许多益处——即避免双链断裂和供体DNA修复模板——同时克服其主要限制。为实现这个宏伟目标,旨在使用靶

标引发的逆转录(TPRT)在靶基因组位点处直接安装编辑的DNA链。在本文讨论的设计中，CRISPR向导RNA(gRNA)将被工程化改造为携带编码诱变DNA链合成的模板，由相关的逆转录酶(RT)酶执行。CRISPR核酸酶(Cas9)产生切口的靶位点DNA将充当逆转录的引物，允许直接掺入任何期望的核苷酸编辑。

[2024] 第1部分

[2025] 建立诱变DNA链的向导RNA模板化逆转录。先前研究表明，在DNA切割之后但在复合物解离之前，Cas9释放非靶DNA链以暴露游离的3′端。据推测，该DNA链可进行通过聚合酶的延伸，并且该gRNA可通过延伸其5′或3′端来工程化改造，以充当DNA合成的模板。在初步体外研究中，已确定Cas9:gRNA结合复合物中的切口DNA链确实可使用结合的gRNA作为模板引发逆转录(反式RT酶)。接着，探索不同的gRNA接头、引物结合位点和合成模板，以确定体外优化设计规则。然后，在体外评估以反式作用或作为与Cas9的融合物的不同RT酶。最后，鉴定在细胞中保留有效结合和切割活性的工程化gRNA设计。该目标的成功证明为在细胞中携带我们的诱变链合成提供基础。

[2026] 第2部分

[2027] 在人细胞中建立引导编辑。基于DNA加工和修复机制，假设诱变DNA链(单链瓣)可用于指导靶核苷酸的特异性和高效编辑。在令人鼓舞的初步研究中，通过显示利用含有诱变瓣的模型质粒底物的编辑，建立了该策略的可行性。在目标1同时，通过系统地改变诱变瓣的长度、序列组成、靶核苷酸身份和3′端来进一步评估修复结果。还将测试1至3个小的核苷酸插入和缺失。并行地，且从目标1构建，将评估Cas9-RT结构，包括融合蛋白和非共价募集策略。将对Cas9-RT结构和延伸的gRNA测定在人类基因组中的多个靶位点处的细胞编辑，然后针对高效率进行优化。如果成功，则该目标会立即为基础科学应用建立TPRT基因组编辑(即，引导编辑)

[2028] 第3部分

[2029] 在培养的人细胞中实现致病性突变的位点特异性编辑。这项技术的潜在通用性实现编辑目前不能被BE校正的颠换突变和插入缺失。在目标1和目标2的结果的指导下，在培养的人细胞中将靶向致病性颠换突变，包括β珠蛋白中镰状细胞病创始者突变(需要A•T至T•A颠换来校正)和ATP7B中最普遍的威尔逊氏(Wilson's)病突变(需要G•C至T•A颠换来校正)。还可检测小的插入和缺失突变的校正，包括CFTR中导致囊性纤维化的3核苷酸ΔF508缺失。如果成功，则这将为开发解决这些重要人类疾病的强大治疗方法奠定基础。

[2030] 方法

[2031] 目标是开发基因组编辑策略，其直接在靶向基因组位点处安装点突变。在技术开发阶段，工作重点集中在蛋白质和RNA工程，以将TPRT功能并入CRISPR/Cas系统中。体外测定将用于仔细探索TPRT各个步骤的功能，从头开始构建(目标1)。第二个焦点领域将使用模型底物和工程化CRISPR/Cas系统的组合来评估哺乳动物细胞中的编辑结果(目标2)。最后，应用阶段使用该技术来校正通过其他方法无法进行基因组编辑的突变(目标3)。

[2032] 一般的编辑设计示于图1A-1B中。Cas9切口酶包含对HNH核酸酶结构域的失活突变(Spy Cas9 H840A或N863A)，将DNA切割限制为含有PAM的链(非靶链)。将向导RNA(gRNA)工程化改造为包含用于逆转录的模板(幻灯片5中详述了设计)。显示了gRNA的5′延伸，但也可实施3′延伸。Cas9切口酶通过C端或N端与逆转录酶(RT)酶融合。gRNA:Cas9-RT复合物靶向

感兴趣的DNA区域并在置换非靶链后形成R环。Cas9对非靶DNA链产生切口。切口链的释放暴露了游离的3′-OH端，其能够使用延伸的gRNA作为模板引发逆转录。这种DNA合成反应通过融合的RT酶进行。gRNA模板编码除了靶向编辑的核苷酸外与原始DNA双链体同源的DNA序列。逆转录的产物是编码期望编辑的单链DNA瓣。包含游离的3′端的该瓣可与邻近DNA链平衡，产生5′瓣种类。后面的种类被假设充当FEN1(瓣核酸内切酶1)的有效底物，FEN1是如下的酶，该酶在滞后链DNA合成过程中从冈崎(Okazaki)片段自然切除5’瓣，并在长补丁(long-patch)碱基切除修复过程中发生的链置换合成后去除5’瓣。连接切口的DNA产生错配的碱基对。该中间体可通过错配修复(MMR)过程经历回复到原始碱基对或转换为期望编辑碱基对。或者，半保守的DNA复制可产生回复和编辑的各一个拷贝。

[2033] 实施例2-引导编辑:在人细胞中进行高度通用且精确的搜索和置换基因组编辑而无需双链DNA断裂

[2034] 当前的基因组编辑方法可使用可编程核酸酶破坏、缺失或插入靶基因，伴随着双链DNA断裂的副产物，并使用碱基编辑器在靶基因座处安装四个转换点突变。然而，小的插入、小的缺失和八个颠换点突变共同代表了大多数致病性遗传变体，但在大多数细胞类型中不能有效且在没有过量的副产物的情况下校正。本文描述了引导编辑，是高度通用且精确的基因组编辑方法，其使用与工程化逆转录酶融合的催化受损Cas9(利用既指定靶位点又编码期望编辑的工程化引导编辑向导RNA(PEgRNA)编程)，将新的遗传信息直接写入指定的DNA位点。在人细胞中进行了超过175种不同的编辑，以证实引导编辑可有效地(通常为20-60％，在未分选细胞中高达77％)且在副产物低(通常为1-10％)的情况下进行靶向插入、缺失、所有12种可能类型的点突变及其组合，无需双链断裂或供体DNA模板。在人细胞中应用引导编辑来校正镰状细胞病(需要HBB中的A·T至T·A颠换)和泰-萨二氏病(需要HEXA中的4-碱基缺失)的主要遗传原因，在这两种情况中在最少副产物的情况下有效地将致病基因组等位基因回复为野生型。还可将引导编辑用于创建具有这些致病性HBB颠换和HEXA插入突变的人细胞系，在PRNP中安装赋予对朊病毒病的抗性(需要G·C至T·A颠换)的G127V突变，以及有效地将His6标签、FLAG表位标签和延伸的LoxP位点插入人细胞的靶基因座。引导编辑在效率和产物纯度方面提供了优于HDR的优势，并且与碱基编辑相比提供互补的优势和劣势。与其搜索和替换机制(其需要三个不同的碱基配对事件)一致，相比于Cas9，引导编辑在已知Cas9脱靶位点处更不容易发生脱靶DNA修饰。引导编辑极大地扩展了基因组编辑的范围和能力，原则上可校正约89％的已知致病性人类遗传变体。

[2035] 在任何活细胞或生物体的基因组中进行几乎任何靶向改变的能力是生命科学的长期追求。尽管基因组编辑技术取得了快速进展，但与疾病相关的>75,000种已知人类遗传变体[111]中的大部分无法在大多数治疗相关细胞中校正或安装(图38A)。可编程核酸酶如CRISPR-Cas9产生双链DNA断裂(DSB)，其可通过靶位点处诱导插入和缺失(插入缺失)的混合来破坏基因112[-114]。核酸酶也可用于通过不依赖于同源性的加工缺失靶基因[115,116]或插入外源基因[117-119]。然而，双链DNA断裂也与非期望的结果相关，包括复杂的产物混合物、易位[120]和p53激活[121,122]。此外，绝大多数致病性等位基因与它们的非致病性对应物的不同之处在于需要精确得多的编辑技术来校正的小的插入、缺失或碱基替换(图38A)。由核酸酶诱导的DSB刺激的同源定向修复(HDR)[123]已广泛用于安装各种精确的DNA变化。然而，HDR依赖于外源性供体DNA修复模板，通常从DSB的末端连接修复中产生过量的插入缺失副产物，并

且在大多数治疗相关的细胞类型中效率低下(T细胞和一些干细胞是重要的例外)[124,125]。虽然提高DSB介导的基因组编辑的效率和精确度仍然是有希望的努力的重点[126-130],但这些挑战需要探索替代的精确基因组编辑策略。

[2036] 碱基编辑在极其多种细胞类型和生物体(包括哺乳动物)中可有效地安装或校正四种类型的转换突变(C至T、G至A、A至G和T至C)而无需使用DSB[128-131],但目前无法实现八种颠换突变(C至A、C至G、G至C、G至T、A至C、A至T、T至A和T至G)中的任何一个,例如直接校正镰状细胞病(HBB E6V)的最常见原因需要T·A至A·T突变[132]。此外,未报道无DSB的方法进行靶缺失,例如去除导致泰-萨二氏病(HEXA 1278+TATC)的4-碱基重复[133],或靶向插入,例如直接校正囊性纤维化最常见原因(CFTRΔF508)需要精确的3-碱基插入[134]。因此,靶向颠换点突变、插入和缺失在大多数细胞类型中难以有效地且在没有过量副产物的情况下安装或校正,即使它们共同构成大多数已知的致病性等位基因(图38A)。

[2037] 本文描述了引导编辑的开发,这是新的"搜索和替换"基因组编辑技术,其介导人细胞中靶向基因座处靶向插入、缺失和所有12种可能的碱基至碱基转换,而无需双链DNA断裂或供体DNA模板。引导编辑器,最初以PE1为例,使用与可编程切口酶融合的逆转录酶和引导编辑延伸的向导RNA(PEgRNA)将遗传信息从PEgRNA的延伸直接拷贝至靶基因组基因座。第二代引导编辑器(PE2)使用工程化逆转录酶在插入缺失形成最小(通常<2%)的情况下显著提高编辑效率,而第三代PE3系统添加第二向导RNA来对非编辑链产生切口,从而有利于置换非编辑链并在插入缺失形成约1-10%的情况下进一步提高编辑效率,通常在人细胞中达到约20-50%。与优化的Cas9核酸酶启动的HDR相比,PE3提供少得多的副产物和更高或相似的效率,并且与当代碱基编辑器相比提供互补的优势和劣势。

[2038] PE3应用于人HEK293T细胞的基因组基因座,以实现HBB E6V至野生型HBB的有效转换,删除插入的TATC以使HEXA 1278+TATC恢复为野生型HEXA,在PRNP中安装赋予对朊病毒病的抗性[135](需要G·C至T·A颠换)的G127V突变,以及靶向插入His_6标签(18bp)、FLAG表位标签(24bp)和延伸的LoxP位点进行Cre介导的重组(44bp)。引导编辑在其他三种人细胞系以及有丝分裂后原代小鼠皮层神经元中也取得了成功,效率各不相同。由于初始切口和编辑位置之间的距离的高度灵活性,引导编辑基本上不受Cas9的PAM要求的限制,原则上可靶向绝大多数基因组基因座。在已知的Cas9脱靶基因座处的脱靶引导编辑比脱靶Cas9编辑少得多,这可能是由于需要三个不同的DNA碱基配对事件才能发生有效的引导编辑。通过在极其多种基因组基因座处实现精确的靶向插入、缺失和所有12类可能的点突变而无需DSB或供体DNA模板,引导编辑具有推进许多基因变体的研究和校正的潜力。

[2039] 结果

[2040] 从延伸的向导RNA转移信息至靶DNA位点的策略

[2041] Cas9使用包含与靶DNA位点杂交的间隔区序列的向导RNA靶向DNA[112-114,136,137]。目标是工程化改造向导RNA,既能像在天然CRISPR系统中一样指定DNA靶标[138,139],也能包含置换靶基因座处相应DNA核苷酸的新的遗传信息。将遗传信息从延伸的向导RNA直接转移至指定的DNA位点,然后置换原始非编辑DNA,原则上可提供在活细胞中安装靶向DNA序列变化而不依赖于DSB或供体DNA模板的通用方法。为了实现这种直接的信息转移,目标是使用在靶位点处产生切口以暴露3′-羟基的基因组DNA,以引发遗传信息从工程化向导RNA(以下被称为引导编辑向导RNA或PEgRNA)上的延伸部分直接逆转录至靶位点(图38A)。

[2042] 切口产生和逆转录的这些初始步骤(其类似于某些天然可移动的遗传元件所使用的机制[140])导致在一条链上具有两个冗余的单链DNA瓣的分支中间体:包含未编辑DNA序列的5'瓣和包含从PEgRNA拷贝的编辑序列的3′瓣(图38B)。为实现成功编辑,必须分解该分支中间体,以便经编辑的3′瓣取代未编辑的5′瓣。虽然5′瓣与未编辑链的杂交在热力学上可能是有利的,因为经编辑的3′瓣可与未编辑链形成较少的碱基对,但5′瓣是结构特异性核酸内切酶(如,FEN1)的优选底物[141],其切除在滞后链DNA合成和长补丁碱基切除修复过程中产生的5′瓣。推断优先5'瓣切除和3'瓣连接可驱动编辑DNA链的掺入,形成含有一条编辑链和一条未编辑链的异源双链体DNA(图38B)。

[2043] 编辑的永久安装可源于随后的DNA修复,该修复以将编辑链中的信息拷贝到互补DNA链的方式分解两条DNA链之间的错配(图38C)。基于开发用于最大化DNA碱基编辑效率的类似策略[131-133],预计距初始切口位点足够远对非编辑DNA链产生切口来使双链断裂形成最小化,可以使DNA修复偏向于优先置换非编辑链。

[2044] 在体外和酵母细胞中验证引导编辑步骤

[2045] 在通过Cas9的RuvC核酸酶结构域切割含PAM的DNA链后,该链的PAM远端片段可与其他情况下稳定的Cas9:sgRNA:DNA复合物分离[143]。推测这条游离链的3′端可能足够接近以引发DNA聚合。向导RNA工程化的努力[144-146]和Cas9:sgRNA:DNA复合物的晶体结构[147-149]表明,可延伸sgRNA的5′和3′端而无需消除Cas9:sgRNA活性。通过将sgRNA延伸为包括两个关键组件而设计PEgRNA:允许有切口的DNA链的3′端与PEgRNA杂交的引物结合位点(PBS)和包含期望编辑的逆转录酶(RT)模板,所述期望编辑当有切口的DNA链的3'端通过聚合酶沿着RNA模板延伸时将直接拷贝至基因组DNA位点(图38C)。

[2046] 这些假设使用纯化的化脓性链球菌Cas9蛋白在体外进行了测试。通过利用PBS序列(5至6个核苷酸,nt)和RT模板(7至22nt)在任一末端延伸sgRNA构建了一系列候选PEgRNA。已证实,5′-延伸PEgRNA引导Cas9与靶DNA结合,并且5′-延伸PEgRNA和3′-延伸PEgRNA均在体外支持Cas9介导的靶标切口产生和哺乳动物细胞中的DNA切割活性(图44A至44C)。这些候选PEgRNA设计使用预先产生切口的5′-Cy5标记的dsDNA底物、催化死亡Cas9(dCas9)和莫洛尼鼠白血病病毒(M-MLV)逆转录酶的商业变体进行测试(图44D)。当所有组分都存在时,观测到荧光标记的DNA链有效转化为更长的DNA产物,具有与沿RT模板的逆转录一致的凝胶迁移(图38D,图44D至44E)。利用5′-延伸或3′-延伸PEgRNA形成期望长度的产物(图38D至38E)。省略dCas9导致DNA模板上衍生自逆转录酶介导的DNA聚合的切口平移产物而无PEgRNA信息转移(图38D)。当PEgRNA用常规sgRNA替代时未观测到DNA聚合产物,证实PEgRNA的PBS和RT模板组件的必要性(图38D)。这些结果表明Cas9介导的DNA熔解暴露了单链R环,如果产生切口,该环能够从5′-延伸或3′-延伸PEgRNA引发逆转录。

[2047] 接着,利用专门对含PAM的链产生切口的Cas9切口酶(H840A突变体)[112]测试非产生切口的dsDNA底物。在这些反应中,5′-延伸PEgRNA低效产生逆转录产物,可能是由于Cas9切口酶活性受损(图44F)。然而,3′-延伸PEgRNA允许稳健的Cas9切口产生和有效的逆转录(图38E)。尽管原则上逆转录可能在PEgRNA的其余部分内的任何位置终止,但使用3′延伸PEgRNA仅产生单个表观产物。与Cas9切口酶、RT和3′-延伸PEgRNA反应的产物的DNA测序显示完整的RT模板序列被逆转录至DNA底物中(图44G)。这些实验证实3′-延伸PEgRNA可在保留引导Cas9切口酶活性的能力的情况下为新的DNA链的逆转录提供模板。

[2048] 为了在体外评估由PEgRNA编程逆转录产生的3′瓣的真核细胞DNA修复结果，在体外使用PEgRNA、Cas9切口酶和RT对报告质粒底物进行DNA切口产生和逆转录，并将反应产物转化至酵母(酿酒酵母)细胞中(图45A)。令人鼓舞地，当在体外利用编码校正过早终止密码子的T·A至A·T颠换的3′-延伸PEgRNA编辑质粒时，37％的酵母转化株表达GFP和mCherry蛋白(图38F，图45C)。与图38E和图44F中的结果一致，在体外利用5′-延伸PEgRNA进行的编辑反应产生的GFP和mCherry双阳性克隆(9％)比利用3′-延伸PEgRNA的那些更少(图38F和图45D)。使用插入单核苷酸(15％双阳性转化体)或缺失单核苷酸(29％双阳性转化体)校正移码突变的3′-延伸PEgRNA也观测到高效的编辑(图38F和图45E至45F)。从双阳性酵母集落中回收的编辑质粒的DNA测序证实在期望序列位点处发生经编码的颠换编辑(图45G)。这些结果表明，真核细胞中的DNA修复可分解由引导编辑引起的3′DNA瓣，以掺入精确的DNA编辑，包括颠换、插入和缺失。

[2049] 引导编辑器1(PE1)的设计

[2050] 受体外和酵母中结果的鼓舞，寻求开发能够编辑哺乳动物细胞中基因组DNA的具有最少数量组件的引导编辑系统。假设3′-延伸PEgRNA(以下简称为PEgRNA，图39A)和Cas9 H840A与逆转录酶的直接融合物通过柔性接头可构成功能性双组分引导编辑系统。用编码野生型M-MLV逆转录酶与Cas9 H840A切口酶的任一端的融合物的质粒以及编码PEgRNA的第二质粒转染HEK293T(永生化人胚胎肾)细胞。初步尝试导致在HEK3靶基因座处检测不到T·A至A·T转换。

[2051] 然而，PEgRNA中的PBS延伸至8-15个碱基(图39A)导致在HEK3靶位点处可检测到T·A至A·T编辑(图39B)，将RT与Cas9切口酶C端融合的引导编辑器构建体(3.7％最大T·A至A·T转换，其中PBS长度范围为8-15nt)相比N端RT-Cas9切口酶融合物(1.3％最大T·A至A·T转换)具有更高的效率(图39B；除非另有说明，本文报告的全部哺乳动物细胞数据是整个处理细胞群体的值，未选择或分选)。这些结果表明，与使用反式提供的M-MLVRT商业变体在体外的需求相比，与Cas9融合的野生型M-MLVRT需要更长的PBS序列进行人细胞的基因组编辑。将这种与Cas9 H840A切口酶C端融合的第一代野生型M-MLV逆转录酶命名为PE1。

[2052] 测试了PE1在PEgRNA指定的四个额外基因组靶位点处精确引入颠换点突变的能力(图39C)。类似于HEK3基因座处的编辑，这些基因组位点处的效率取决于PBS长度，最大编辑效率范围为0.7-5.5％(图39C)。来自PE1的插入缺失很低，在使每个位点的编辑效率最大化的条件下，五个位点的平均值为0.2±0.1％(图46A)。PE1还能安装靶向插入和缺失，例如在HEK3基因座处的单核苷酸缺失(4.0％效率)、单核苷酸插入(9.7％)和三核苷酸插入(17％)(图39C)。这些结果证实PE1能够直接安装靶向颠换、插入和缺失，无需双链DNA断裂或DNA模板。

[2053] 引导编辑器2(PE2)的设计

[2054] 虽然PE1可在HEK293T细胞的多个基因座安装多种编辑，但编辑效率通常较低(通常≤5％)(图39C)。推测工程化改造PE1中的逆转录酶可在引导编辑复合物的独特构象限制内提高DNA合成的效率，从而导致更高的基因组编辑产率。先前已报道M-MLV RT突变增加酶的热稳定性[150,151]、持续合成能力[150]和DNA:RNA异源双链体底物亲和力[152]，并使RNaseH活性失活[153]。构建了19种包含各种逆转录酶突变的PE1变体，以评估它们在人细胞中的引导编辑效率。

[2055] 首先,研究了一系列M-MLVRT变体,这些变体先前由于具有在升高温度下支持逆转录的能力而从实验室进化而来[150]。将这些氨基酸替换中的三个(D200N、L603W和T330P)连续引入M-MLV RT(以下称为M3),导致HEK293T细胞中5个基因组基因座的颠换和插入编辑效率相比PE1平均提高6.8倍(图47A至47S)。

[2056] 接着,与M3结合,测试了先前显示增强模板:PBS复合物结合、酶持续合成能力和热稳定性的其他逆转录酶突变[152]。在分析的14种其他突变体中,除了M3突变外,还具有T306K和W313F取代的变体在人细胞的5个基因组位点进行6个颠换或插入编辑的编辑效率相比M3再提高了1.3倍至3.0倍(图47A至47S)。这种M-MLV逆转录酶的五突变体掺入PE1结构(Cas9 H840A-M-MLVRT(D200N L603WT330PT306KW313F))在下文中被称为PE2。

[2057] PE2以基本上高于PE1的效率安装单核苷酸颠换、插入和缺失突变(图39C),并且与更短的PBS PEgRNA序列(图39C)兼容,与提高的有效接合瞬时基因组DNA:PBS复合物的能力一致。平均地,PE2导致引导编辑点突变效率比PE1提高1.6到5.1倍(图39C),并且在某些情况下显著提高编辑产率高达46倍(图47F和图47I)。PE2还比PE1更有效地实现靶向插入和缺失,在HEK3基因座处以4.5%的效率实现24-bp FLAG表位标签的靶向插入,比用PE1安装该插入的效率提高15倍(图47D),在HEK3中以8.6%的效率介导1-bp缺失,比PE1的效率更高2.1倍(图39C)。这些结果证实PE2是比PE1更有效的引导编辑器。

[2058] PEgRNA特征的优化

[2059] 利用PE2在HEK293T细胞的5个基因组位点处系统地探究PEgRNA结构与引导编辑效率之间的关系(图39C)。通常,GC含量越低的引发位点需要越长的PBS序列(EMX1和RNF2,在缺口上游的前10nt分别含有40%和30%GC含量),而那些GC含量越高的位点支持具有更短的PBS序列(HEK4和FANCF,在切口上游的前10nt中分别含有80%和60%GC含量)(图39C)引发编辑,这与有切口的DNA链与PEgRNAPBS杂交的能量要求一致。PBS长度或GC含量水平不能严格预测引导编辑效率,其他因素如DNA引物或PEgRNA延伸的二级结构也可能影响编辑活性。对于典型的靶序列,建议从约13nt的PBS长度开始,如果序列偏离约40-60%GC含量,则探索不同的PBS长度。必要时,应根据经验确定最佳PBS序列。

[2060] 接着,研究PEgRNA的RT模板部分的性能决定因素。使用PE2在五个基因组靶位点处系统评估具有长度范围为10-20nt的RT模板(图39D),并在三个基因组位点处评估具有长达31nt的更长RT模板的PEgRNA(图48A至48C)。与PBS长度一样,也可改变RT模板长度以最大化引导编辑效率,尽管通常长≥10nt的许多RT模板长度支持更有效的引导编辑(图39D)。由于一些靶位点优选较长的RT模板(>15nt)来实现更高的编辑效率(FANCF,EMX1),而其他基因座优选短的RT模板(HEK3,HEK4)(图39D),因此优化PEgRNA时建议使用短和长RT模板两者进行测试,从约10-16nt开始。

[2061] 重要地,与具有相似长度的RT模板的其他PEgRNA相比,将C放置作为与sgRNA支架的末端发夹相邻的核苷酸的RT模板通常导致较低的编辑效率(图48A至48C)。基于与Cas9结合的sgRNA的结构[148,149],推测C作为经典sgRNA的3’延伸的第一个核苷酸的存在可通过与G81配对破坏sgRNA支架折叠,所述G81是与Cas9中的Tyr1356天然形成π堆积且与sgRNAA68非经典碱基配对的核苷酸。由于许多RT模板长度支持引导编辑,建议选择3′延伸的第一个碱基(RT模板的最后一个逆转录碱基)不是C的PEgRNA。

[2062] 引导编辑器3系统设计(PE3和PE3b)

[2063] 虽然PE2可比PE1更有效地将遗传信息从PEgRNA转移到靶基因座,但细胞分解由一条编辑链和一条非编辑链形成的异源双链体DNA的方式决定了编辑是否持久。先前的碱基编辑发展面临着类似的挑战,因为胞嘧啶或腺嘌呤脱氨基的初始产物是包含一条编辑链和一条非编辑链的异源双链体DNA。为了提高碱基编辑的效率,使用编辑链作为模板,使用Cas9 D10A切口酶将切口引入非编辑链,并将DNA修复引导至该链[129,130,142]。为了利用该原理来提高引导编辑效率,测试使用已存在于PE2中的Cas9 H840A切口酶和简单sgRNA对非编辑链产生切口来诱导细胞优先置换非编辑链的类似策略(图40A)。由于编辑DNA链也有切口以启动引导编辑,在非编辑链上测试各种sgRNA编程的切口位置,以最小化导致插入缺失的双链DNA断裂的产生。

[2064] 首先通过筛选诱导位于PEgRNA诱导的切口位点(PAM的5′或3′)14至116个碱基处的切口的sgRNA,在HEK293T细胞的5个基因组位点测试了这种PE3策略。在测试的5个位点中的4个中,对非编辑链产生切口使无插入缺失的引导编辑产物的数量相比PE2系统增加了1.5至4.2倍,高达55%(图40B)。虽然最佳切口位点因基因组位点而异,但位于PAM的3′(图40B中的正距离)的距离PEgRNA诱导的切口约40至90bp的切口通常产生引导编辑效率的有利增加(平均41%),而没有过多的插入缺失形成(sgRNA平均插入缺失为6.8%,导致所测试5个位点中每一个的最高编辑效率)(图40B)。如预期,在某些位点,将非编辑链切口放置在PEgRNA诱导的切口的40bp内导致插入缺失形成大幅增加高达22%(图40B),这可能是由于对两条紧密的链产生切口形成了双链断裂。然而,在其他位点,对距离PEgRNA诱导的切口近达14bp产生切口仅产生5%的插入缺失(图40B),表明基因座依赖性因素控制近端双切口转化为双链DNA断裂。在一个测试位点(HEK4)处,互补链切口未提供任何益处,或导致插入缺失水平超过编辑效率(高达26%),即使放置在距离PEgRNA诱导的切口>70bp处,这与编辑链在该位点处被细胞产生切口或低效连接的不寻常倾向一致。如果插入缺失频率超过可接受的水平,则建议从PEgRNA介导的切口约50bp的非编辑链产生切口开始并测试替代切口位置。

[2065] 这种关于互补链切口产生如何提高引导编辑效率的模型(图40A)预测,仅在编辑链瓣分解后才对非编辑链产生切口可最小化并发切口的存在,从而降低继续形成插入缺失的双链断裂的频率。为了实现时序控制非编辑链切口产生,设计了具有与编辑链匹配但与原始等位基因不匹配的间隔区序列的sgRNA。使用此策略(以下被称为PE3b),间隔区和非编辑等位基因之间的错配应当不利于由sgRNA产生切口,直到PAM链发生编辑事件之后。这种PE3b方法用HEK293T细胞的三个基因组位点处的5种不同的编辑测试,并将结果与用PE2和PE3系统实现的结果进行了比较。在所有情况下,PE3b相比PE3与显著更低水平的插入缺失相关(插入缺失低3.5至30倍,平均低12倍,或0.85%),整体编辑效率相比PE3没有任何明显下降(图40C)。因此,当编辑位于第二原间隔区内时,PE3b系统可减少插入缺失,相比于PE2仍然提高编辑效率,通常达到与PE3相似的水平(图40C)。

[2066] 总之,这些发现证实PE3系统(Cas9切口酶优化的逆转录酶+PEgRNA+sgRNA)相比于PE2提高了编辑效率约3倍(图40B至40C)。如预期考虑到PE3的额外切口产生活性,PE3伴随着比PE2更宽范围的插入缺失。在优先考虑引导编辑效率时,建议使用PE3。当最小化插入缺失至关重要时,PE2提供低约10倍的插入缺失频率。当可使用识别已安装编辑的sgRNA来对非编辑链产生切口时,PE3b系统可在大大减少插入缺失形成的情况下实现类似PE3的编辑

水平。

[2067] 为了证明用PE3进行引导编辑的靶向范围和多功能性，使用PE3和具有10个核苷酸RT模板的PEgRNA，探索了在HEK3靶位点的+1至+8位点(将PEgRNA诱导的切口3′的第一个碱基计为位点+1)间安装所有可能的单核苷酸取代(图41A)。总体地，这24个不同编辑涵盖了所有4个转换突变和所有8个颠换突变，以平均为33±7.9％(范围为14％至48％)的编辑效率(不含插入缺失)进行，平均值为7.5±1.8％插入缺失。

[2068] 重要地，利用PE3，长距离RT模板也可引起高效引导编辑。例如，使用具有34-nt RT模板的PE3，以平均36±8.7％的效率和8.6±2.0％的插入缺失将点突变安装在HEK3基因座中位点+12、+14、+17、+20、+23、+24、+26、+30和+33(PEgRNA诱导的切口12至33个碱基)处(图41B)。尽管未尝试其他基因座处超过+10位点的编辑，但在三个替代位点处≥30nt的其他RT模板也支持高效编辑(图48A至C)。长RT模板的活性能够对来自初始切口位点的数十个核苷酸进行高效的引导编辑。由于任一DNA链上的NGG PAM平均每约8bp出现一次，远小于支持高效引导编辑的编辑和PAM之间的最大距离，与其他精确基因组编辑方法相反[125,142,154]，引导编辑基本不受附近PAM序列可用性的限制。鉴于RNA二级结构与引导编辑效率之间的假定的关系，在设计用于远程编辑的PEgRNA时，谨慎的做法是测试不同长度的RT模板，并在必要时测试序列组成(如，同义密码子)以优化编辑效率。

[2069] 为了进一步测试PE3系统用于引入转换和颠换点突变的范围和局限性，测试了涵盖跨越6个其他基因组靶位点的所有12种可能类型点突变的72种其他编辑(图41C至41H)。总之，无插入缺失编辑效率平均为25±14％，而插入缺失形成平均为8.3±7.5％。由于PEgRNART模板包含PAM序列，引导编辑可诱导PAM序列变化。在这些情况下，观测到更高的编辑效率(平均为39±9.7％)和更低的插入缺失形成(平均为25.0±2.9％)(图41A至41K，位点+5或+6处的点突变)。PAM编辑的效率的此提高和插入缺失形成减少可能是由于Cas9切口酶无法在互补链修复之前重新结合和对编辑链产生切口。由于引导编辑支持组合编辑而不会明显损失编辑效率，因此建议除了其他期望的变化之外，在可能的情况下编辑PAM。

[2070] 接着，使用PE3在7个基因组位点处进行14个靶向小插入和14个靶向小缺失(图41I)。靶向1-bp插入的平均效率为32±9.8％，而3-bp插入的平均效率为39±16％。靶向的1-bp和3-bp缺失也是有效的，分别以29±14％和32±11％的平均产率进行。插入缺失形成(超出靶向的插入或缺失)平均为6.8±5.4％。由于在位点+1和+6之间引入的插入和缺失改变PAM的位点或结构，推测该范围内的插入和缺失编辑通常更高效，因为修复互补链之前Cas9切口酶无法重新结合和对编辑DNA链产生切口，类似于编辑PAM的点突变。

[2071] 还测试了PE3在HEK3位点处介导5bp至80bp的较大精确缺失的能力(图41J)。使用13-ntPBS和包含分别与靶基因座同源的29、24或19bp的RT模板时，对于5-、10-和15-bp缺失观测到非常高的编辑效率(52％至78％)。使用26-nt RT模板支持25bp的更大缺失，效率为72±4.2％，而20-nt RT模板实现80-bp缺失，效率为52±3.8％。这些靶向缺失伴随着平均为11±4.8％的插入缺失频率(图41J)。

[2072] 最后，测试了PE3在同一靶基因座处介导由跨越三个基因组位点的插入和缺失、插入和点突变、缺失和点突变或两个点突变组成的12种多重编辑组合的能力。这些组合编辑是非常高效的，在6.4％的插入缺失的情况下平均为55％的靶编辑(图41K)，并且表明引导编辑能够在单个靶位点处在具有高效率和低插入缺失频率的情况下进行精确插入、缺失和

点突变的组合。

[2073] 总之,图41A至41K的示例代表了跨越7个人类基因组基因座的156种不同的转换、颠换、插入、缺失和组合编辑。这些发现证实引导编辑的多功能性、精确性和靶向灵活性。

[2074] 引导编辑与碱基编辑的比较

[2075] 当前世代的胞苷碱基编辑器(CBE)和腺嘌呤碱基编辑器(ABE)可在具有高效率和低插入缺失的情况下安装C·G至T·A转换突变和A·T至G·C转换突变[129,130,142]。碱基编辑的应用可受限于碱基编辑活性窗口内存在的多个胞苷或腺嘌呤碱基(通常约为5bp宽),这导致不需要的旁观者编辑[129,130,142,155],或者受限于距靶核苷酸约15±2nt定位的PAM的缺乏[142,156]。预计引导编辑对于在无旁观者编辑的情况下精确安装转换突变,或在缺乏适当定位的PAM阻止CBE或ABE活性窗口内有利地定位靶核苷酸时特别有用。

[2076] 通过使用无切口酶活性(BE2max)或具有切口酶活性(BE4max)的优化CBE[157],或使用类似PE2和PE3引导编辑系统,编辑在经典碱基编辑窗口(原间隔区位点4-8,将PAM计数为位点21-23)中包含多个靶胞苷的3个基因组基因座来比较引导编辑和胞嘧啶碱基编辑。在3个位点的碱基编辑窗口内的9个总靶胞嘧啶中,对于碱基编辑窗口中心的碱基(原间隔区位点5-7,图42A),BE4max产生的平均总C·G至T·A转换比PE3高2.2倍。同样,在这些良好定位的碱基处,非切口产生性BE2max平均比PE2表现优异1.4倍(图42A)。然而,对于超出碱基编辑窗口中心的胞嘧啶,PE3比BE4max表现优异2.7倍,PE2比BE2max表现优异2.0倍(PE3平均编辑为40±17%相比于BE4max为15±18%,PE2为22±11%相比于BE2max为11±13%)。总之,PE2的插入缺失频率非常低(平均0.86±0.47%),而PE3的插入缺失频率与BE4max相似或略高(BE4max范围:2.5%至14%;PE3范围:2.5%至21%)(图42B)。

[2077] 比较碱基编辑与引导编辑安装精确C·G至T·A编辑(无任何旁观者编辑)的效率时,上述位点处,引导编辑的效率大大超过碱基编辑的效率,与大多数基因组DNA位点类似,这些位点在约5-bp碱基编辑窗口内包含多个胞嘧啶(图42C)。在这些位点(如,EMX1,其在原间隔区位点C5、C6和C7处包含胞嘧啶)处,BE4max生成很少的产物,该产物只包含单个靶碱基对转换,没有旁观者编辑。相反地,该位点处的引导编辑可用于在任何位点或位点组合(C5、C6、C7、C5+C6、C6+C7、C5+C7,或C5+C6+C7)处选择性安装C·G至T·A编辑(图42C)。所有精确的单碱基或双碱基编辑(即,不修饰任何其他附近碱基的编辑)在利用PE3或PE2的情况下分别比在利用BE4max或BE2的情况下有效得多,而三碱基C·G至T·A编辑在利用BE4max的情况下更高效(图42C),反映了碱基编辑器在活性窗口内编辑所有靶碱基的倾向。总之,这些结果表明,胞嘧啶碱基编辑器可在最佳定位的靶碱基处进行比PE2或PE3更高水平的编辑,但引导编辑可在非最佳定位的靶碱基处优于碱基编辑,并且可利用多个可编辑的碱基以高得多的精确度进行编辑。

[2078] 通过优化的非切口产生性ABE(具有dCas9而非Cas9切口酶的ABEmax[152],以下被称为ABEdmax)相比于PE2,以及通过优化的切口产生性腺嘌呤碱基编辑器ABEmax相比于PE3,在两个基因组基因座处比较A·T至G·C编辑。在碱基编辑窗口(HEK3)中包含两个靶腺嘌呤的位置处,对于A5转换,ABE比PE2或PE3更高效,但对于位于ABEmax编辑窗口边缘的A8转换,PE3更高效(图42D)。当比较其中仅转换单个腺嘌呤的精确编辑的效率时,PE3在A5和A8处均优于ABEmax(图42E)。总之,ABE在HEK3处产生的插入缺失比引导编辑器少得多(ABEdmax为0.19±0.02%相比于PE2为1.5±0.46%,ABEmax为0.53±0.16%相比于PE3为11±2.3%,

图42F)。在碱基编辑窗口内仅存在单个A的FANCF处,ABE2和ABEmax在总靶碱基对转换方面优于它们的引导编辑对应物1.8至2.9倍,几乎所有编辑产物都来自两种碱基编辑,引导编辑仅包含精确编辑(图42D至42E)。与HEK3位点一样,ABE在FANCF位点处产生的插入缺失少得多(图42F)。

[2079] 总之,这些结果表明碱基编辑和引导编辑为进行靶向转换突变提供了互补的优势和劣势。对于碱基编辑窗口内存在单个靶核苷酸的情况,或当旁观者编辑是可接受的,当前的碱基编辑器通常比引导编辑器更高效且产生更少的插入缺失。当存在多个胞嘧啶或腺嘌呤并且不期望旁观者编辑时,或者当用于相对于可用PAM的碱基编辑的靶碱基定位不佳时,引导编辑器提供了巨大的优势。

[2080] 脱靶引导编辑

[2081] 为了导致高效编辑,引导编辑需要与Cas9结构域互补以结合的靶基因座:PEgRNA间隔区,启动PEgRNA引发的逆转录的靶基因座:PEgRNAPBS互补性,以及用于进行瓣分解的靶基因座:逆转录酶产物互补性。推测与其他基因组编辑方法相比,这三种不同的DNA杂交要求可最小化脱靶引导编辑。为了测试这种可能性,利用Cas9和4种靶向相同原间隔区的相应sgRNA,或利用Cas9和相同的16种PEgRNA,用PE3或PE2和总共16种设计用于靶向四个中靶基因组基因座的PEgRNA处理HEK293T细胞。选择这4个靶基因座,因为每个都具有至少4个良好表征的脱靶位点,对于所述脱靶位点,已知在HEK293T细胞中Cas9和相应中靶sgRNA导致大量脱靶DNA修饰[118,159]。处理后,对每个中靶间隔区的4个中靶基因座和前4个已知的Cas9脱靶位点进行测序,总共16个脱靶位点(表1)。

[2082] 与先前研究[118]一致,Cas9和4种靶sgRNA修饰所有16个先前报告的脱靶基因座(图42G)。HEK3靶基因座的4个脱靶位点中Cas9脱靶修饰效率平均为16%。Cas9和靶向HEK4的sgRNA导致4个测试的已知脱靶位点的平均60%的修饰。类似地,EMX1和FANCF的脱靶位点由Cas9:sgRNA修饰,平均频率分别为48%和4.3%(图42G)。值得注意地,相比于sgRNA,平均地,PEgRNA与Cas9核酸酶以相似的效率修饰中靶位点(低1至1.5倍),而PEgRNA与Cas9核酸酶修饰脱靶位点的平均效率比sgRNA低约4倍。

[2083] 引人注目地,具有包含这四个靶间隔区的相同16种测试PEgRNA的PE3或PE2导致脱靶编辑低得多(图42H)。在已知通过Cas9+sgRNA、PE3+PEgRNA或PE2+PEgRNA进行脱靶编辑的16个位点中,16个脱靶位点只有3个发生可检测的脱靶引导编辑,16个中只有1个显示脱靶编辑效率≥1%(图42H)。在这16个已知Cas9脱靶位点处,靶向HEK3、HEK4、EMX1和FANCF的PEgRNA的平均脱靶引导编辑分别为<0.1%、<2.2±5.2%、<0.1%和<0.13±0.11%(图42H)。值得注意地,在Cas9+PEgRNA1以97%的效率编辑的HEK4脱靶3位点处,尽管共有相同的间隔区序列,PE2+PEgRNA1导致仅0.7%的脱靶编辑,表明与Cas9编辑相比引导编辑必需的两个额外DNA杂交事件如何可以大大减少脱靶编辑。综上所述,这些结果表明PE3和PEgRNA在人细胞中诱导的脱靶DNA编辑比靶向相同原间隔区的Cas9和sgRNA低得多。

[2084] 原则上,3′-延伸PEgRNA的逆转录可进行到向导RNA支架。如果所产生的3′瓣(尽管其3′端与未编辑DNA链缺乏互补性)掺入靶基因座中,则结果是插入有助于插入缺失频率的PEgRNA支架核苷酸。我们分析了来自HEK293T细胞中4个基因座处66个PE3介导的编辑实验的测序数据,观测到低频率的PEgRNA支架插入,平均为任意数目的PEgRNA支架核苷酸的1.7±1.5%总插入(图56A-56D)。推测由于Cas9结构域结合,以及在由PEgRNA支架逆转录造成

的3′瓣的错配3′端的瓣分解过程中的细胞切除，向导RNA支架无法接近逆转录酶，从而最小化掺入PEgRNA支架核苷酸的产物。虽然此类事件很少见，但未来致力于最小化PEgRNA支架掺入的工程化PEgRNA或引导编辑器蛋白可能会进一步降低插入缺失频率。

[2085] 一些碱基编辑器中的脱氨酶可以以Cas9非依赖性方式作用，产生第一代CBE(但不是ABE)中低水平但广泛的脱靶DNA编辑[160-162]和第一代CBE和ABE中的脱靶RNA编辑[163-165]，但是具有工程化脱氨酶的较新CBE和ABE变体大大减少了Cas9非依赖性脱靶DNA和RNA编辑[163-165]。引导编辑器缺乏碱基修饰酶如脱氨酶，因此没有以Cas9非依赖性方式修饰DNA或RNA碱基的固有能力。

[2086] 虽然原则上引导编辑器中的逆转录酶结构域可加工细胞中正确引发的RNA或DNA模板，但注意到逆转录转座子如LINE-1家族[166]、内源性逆转录病毒[167、168]和人端粒酶中的那些都提供了活性内源性人逆转录酶。它们在人细胞中的天然存在表明逆转录酶活性本身基本上无毒性。实际上，与表达dCas9、Cas9 H840A切口酶的对照或具有使逆转录酶失活的R110S+K103L(PE2-dRT)突变的PE2[169]相比，HEK293T细胞活力中未观测到PE3依赖性差异(图49A至49B)。

[2087] 尽管有上述数据和分析，仍需要其他研究以无偏倚的全基因组方式评估脱靶引导编辑，以及表征引导编辑器或引导编辑中间体中逆转录酶变体可能影响细胞的程度。

[2088] 在人细胞中引导编辑致病性颠换、插入和缺失突变

[2089] 在人细胞中测试PE3直接安装或校正导致遗传性疾病的颠换、小插入和小缺失突变的能力。镰状细胞病最常由HBB中的A•T至T•A颠换突变导致β-珠蛋白中Glu6→Val突变引起。利用Cas9核酸酶和供体DNA模板进行HDR，然后富集经编辑的细胞、移植和植入来离体治疗造血干细胞，是治疗镰状细胞病有前景的潜在策略[170]。然而，除了正确编辑的HBB等位基因[170-171]之外，这种方法仍然会产生许多含有插入缺失的副产物。虽然碱基编辑器通常产生少得多的插入缺失，但它们目前不能进行直接恢复HBB正常序列所需的T•A至A•T颠换突变。

[2090] 将PE3用于在HEK293T细胞中以44％的效率和4.8％的插入缺失安装HBB E6V突变(图43A)。从PE3处理的细胞混合物中，我们分离出HBB E6V等位基因纯合(三倍体)的6个HEK293T细胞系(图53A至53D)，表明引导编辑产生具有致病性突变的人细胞系的能力。为将HBB E6V等位基因校正为野生型HBB，我们利用PE3和编程为直接将HBB E6V突变回复为野生型HBB的PEgRNA处理纯合的HBB E6V HEK293T细胞。总共测试了14种PEgRNA设计。3天后，DNA测序显示所有14种PEgRNA在与PE3结合时都能高效地将HBB E6V校正为野生型HBB(≥26％野生型HBB，无插入缺失)，并且插入缺失水平平均为2.8±0.70％(图50A)。最佳PEgRNA以2.4％插入缺失导致52％的HBB E6V至野生型的校正(图43A)。引入修饰由PEgRNA识别的PAM的沉默突变适度提高编辑效率和产物纯度至1.4％插入缺失情况下的58％的校正(图43A)。这些结果证实引导编辑可在人细胞系中以高效率和最少的副产物的情况下安装和校正致病性颠换点突变。

[2091] 泰-萨二氏病最常由4-bp插入HEXA基因(HEXA1278+TATC)引起[136]。PE3用于以31％的效率和0.8％插入缺失将该4-bp插入安装到HEK293T细胞中(图43B)，并分离出2个HEXA1278+TATC等位基因纯合的HEK293T细胞系(图53A-53D)。在利用PE3或PE3b系统的情况下使用这些细胞对43种PEgRNA和3种切口产生性sgRNA测试HEXA中的致病性插入的校正(图

50B)，通过完全回复为野生型等位基因或通过破坏PAM并安装沉默突变的转移的4-bp缺失进行。测试的43种PEgRNA中的19种导致≥20％的编辑。利用PE3或PE3b和最佳PEgRNA向野生型HEXA的完全校正以相似的平均效率进行(PE3为30％相比于PE3b为33％)，但PE3b系统伴随着少5.3倍的插入缺失产物(PE3为1.7％相比于PE3b为0.32％)(图43B和图50B)。这些发现表明引导编辑能够进行高效且在副产物最少的情况下安装或校正哺乳动物细胞中的致病性等位基因的精确小插入和缺失。

[2092] 最后，测试了将保护性SNP安装到编码人朊病毒蛋白(PrP)的基因PRNP中。PrP错误折叠导致进行性和致命的神经退行性朊病毒病，该病可通过PRNP基因中的遗传显性突变或通过暴露于错误折叠的PrP而自发出现[172]。天然存在的PRNP G127V突变体等位基因赋予人[138]和小鼠[173]对朊病毒病的抗性。PE3用于将G127V安装到HEK293T细胞的人PRNP等位基因中，这需要G·C至T·A颠换。利用PE3系统评估4种PEgRNA和3种切口产生性sgRNA。在暴露于最有效的PE3和PEgRNA的3天后，DNA测序显示安装G127V突变的效率为53±11％，插入缺失水平为1.7±0.7％(图43C)。总之，这些结果证实引导编辑在人细胞中高效且在副产物最少的情况下安装或校正导致或赋予对疾病抗性的颠换、插入或缺失突变的能力。

[2093] 不同人细胞系和原代小鼠神经元中的引导编辑

[2094] 接着，测试引导编辑在3种其他人细胞系中编辑内源性位点的能力。在K562(白血病骨髓)细胞中，PE3用于在HEK3、EMX1和FANCF位点执行颠换编辑，以及在HEK3中插入18-bp的6xHis标签。对于这四种PE3介导的编辑中的每一种，观测到平均编辑效率为15-30％，插入缺失平均为0.85-2.2％(图43A)。在U2OS(骨肉瘤)细胞中，安装了HEK3和FANCF中的颠换突变，以及对HEK3中的3-bp插入和6xHis标签插入，具有7.9-22％的编辑效率，超过插入缺失形成10至76倍(图43A)。最后，在HeLa(宫颈癌)细胞中，3-bp插入HEK3以12％的平均效率和1.3％的插入缺失进行(图43A)。总之，这些数据表明HEK293T细胞以外的多种细胞系支持引导编辑，但是编辑效率因细胞类型而异，而且效率通常低于HEK293T细胞。在所有测试的人细胞系中编辑:插入缺失比率保持较高。

[2095] 为确定有丝分裂后是否可能进行引导编辑，用双断裂PE3慢病毒递送系统转导从E18.5小鼠收获的终末分化原代细胞，原代皮层神经元，其中断裂内含肽剪接[203]重新组成来自N端半部和C端半部的PE2蛋白，每个半部都由单独的病毒递送。为了限制对有丝分裂后神经元的编辑，将对成熟神经元高度特异的人突触蛋白启动子用于驱动两种PE2蛋白组件的表达。GFP通过自切割P2A肽[205]融合至PE2的N端半部。在双病毒转导后两周从神经元中分离细胞核并直接测序，或在测序前对GFP表达进行分选。观察到在分选的细胞核中以0.58±0.14％的平均插入缺失在DNMT1基因座处7.1±1.2％的平均引导编辑安装颠换(图43D)。同一断裂内含肽双慢病毒系统中的Cas9核酸酶导致分选的皮质神经元核中31±5.5％的插入缺失(图43D)。这些数据表明有丝分裂后终末分化的原代细胞可支持引导编辑，因此证实引导编辑不需要细胞复制。

[2096] 比较引导编辑与Cas9启动的HDR

[2097] 在支持HDR[128]的有丝分裂细胞系中比较PE3的性能与优化Cas9启动的HDR[128,125]的性能。利用Cas9核酸酶、sgRNA和设计用于安装各种颠换和插入编辑的ssDNA供体寡核苷酸模板处理HEK293T、HeLa、K562和U2OS细胞(图43E至43G和图51A至51F)。在所有情况下，Cas9启动的HDR成功安装了期望编辑，但副产物水平高得多(主要是插入缺失)，如导致双链断裂

的处理所预期的。在HEK293T细胞中使用PE3,HBB E6V安装和校正以42％和58％的平均编辑效率进行,分别具有2.6％和1.4％的平均插入缺失(图43E和图43G)。相反地,利用Cas9核酸酶和HDR模板得到的相同编辑导致5.2％和6.7％的平均编辑效率,具有79％和51％的平均插入缺失频率(图43E和图43G)。类似地,PE3以53％的效率和1.7％插入缺失安装PRNP G127V,而Cas9启动的HDR以6.9％的效率和53％插入缺失安装该突变(图43E和图43G)。因此,HBB E6V安装、HBB E6V校正和PRNP G127V安装的编辑:插入缺失比率平均而言对于PE5比对于Cas9启动的HDR高270倍。

[2098] 在除HEK293T之外的人细胞系中,PE3和HDR之间的比较显示类似的结果,尽管PE3编辑效率较低。例如,在K562细胞中,与Cas9启动的HDR的17％编辑和72％插入缺失相比,PE3介导的对HEK3的3-bp插入以25％的效率和2.8％插入缺失进行,有利于PE3的40倍编辑:插入缺失比率优势(图43F至43G)。在U2OS细胞中,PE3以22％的效率和2.2％插入缺失进行此3-bp插入,而Cas9启动的HDR导致15％的编辑与74％插入缺失,低49倍的编辑:插入缺失比率(图43F至3G)。在HeLa细胞中,相比于Cas9启动的HDR的3.0％的编辑和69％插入缺失,PE3以12％的效率和1.3％插入缺失进行插入,210倍的编辑:插入缺失比率差异(图43F至43G)。总之,这些数据表明,在所测试的四种细胞系中,HDR通常导致与PE3相似或更低的编辑效率和远高于PE3的插入缺失(图51A-51F)。

[2099] 讨论和未来方向

[2100] 以单核苷酸精确度插入DNA序列的能力是特别可行的引导编辑能力。例如,PE3用于将His_6标签(18bp,65％的平均效率)、FLAG表位标签(24bp,18％的平均效率)和延伸LoxP位点(其是Cre重组酶的天然底物,44bp,23％的平均效率)精确插入HEK293T细胞的HEK3基因座中。这些示例的平均插入缺失范围为3.0％至5.9％(图43H)。许多生物技术、合成生物学和治疗应用预计源自有效且精确地将新的DNA序列引入活细胞中感兴趣的靶位点的能力。

[2101] 总之,本文所述引导编辑实验在人和小鼠基因组从PAM起始的上游3bp至下游29bp范围的位置处跨越12个内源性基因座安装18个高达44bp的插入、22个高达80bp的缺失、113个点突变(包括77个颠换)和18个组合编辑,而不会造成明显的双链DNA断裂。这些结果证实引导编辑为非常通用的基因组编辑方法。由于ClinVar中绝大多数(85-99％)的插入、缺失、插入缺失和重复≤30bp(图52A至52D),原则上引导编辑可校正ClinVar中目前已知的75,122个致病性人遗传变体中高达约89％(图38A中的转换、颠换、插入、缺失、插入缺失和重复),具有改善由拷贝数增加或减少引起的疾病的额外潜力。

[2102] 重要地,对于任何期望编辑,引导编辑的灵活性提供了PEgRNA诱导的切口位点、sgRNA诱导的第二切口位点、PBS长度、RT模板长度以及首先编辑的链的许多可能的选择,如本文广泛证明的。这种灵活性与通常可用于其他精确基因组编辑方法[125,142,154]的更有限的选择形成对比,允许优化编辑效率、产物纯度、DNA特异性或其他参数来适应给定应用的需要,如图50A至50B所示,其中测试14和43个PEgRNA,分别涵盖优化致病性HBB和HEXA等位基因校正的一系列引导编辑策略。

[2103] 需要更多的研究来进一步了解和改进引导编辑。可能需要额外修饰引导编辑器系统扩展其兼容性来包括其他细胞类型,例如有丝分裂后的细胞。需要将引导编辑与病毒和非病毒体外和体内递送策略相互联系以充分探索引导编辑实现广泛应用的潜力,包括遗传

性疾病的研究和治疗。然而，通过在哺乳动物细胞基因组中实现高精确度的靶向转换、颠换、小插入和小缺失而无需双链断裂或HDR，引导编辑提供了新的“搜索和置换”能力，大大扩展了基因组编辑的范围。

[2104] 方法

[2105] 一般方法

[2106] 除非另有说明，否则使用Phusion U Green Multiplex PCR Master Mix (ThermoFisher Scientific)或Q5 Hot Start High-Fidelity 2x Master Mix(New England BioLabs)通过PCR进行DNA扩增。DNA寡核苷酸，包括Cy5标记的DNA寡核苷酸、dCas9蛋白和Cas9H840A蛋白获自Integrated DNA Technologies。酵母报告质粒衍生自先前描述的质粒[64]并通过Gibson组装方法克隆。本文使用的所有哺乳动物编辑器质粒都使用先前描述的USER克隆方法[175]组装。通过将退火的寡核苷酸连接到BsmBI消化的受体载体中来构建表达sgRNA的质粒。使用定制受体质粒通过Gibson组装或Golden Gate组装构建表达PEgRNA的质粒(参见补充的“Golden Gate组装”概述)。本文使用的sgRNA和PEgRNA构建体的序列列于表2A至2C和表3A至3R中。用于哺乳动物细胞实验的所有载体均使用Plasmid Plus Midiprep试剂盒(Qiagen)或PureYield plasmid miniprep试剂盒(Promega)纯化，其中包括内毒素去除步骤。所有使用活动物的实验都经Broad研究所学会和动物护理和使用委员会(Broad Institute Institutional and Animal Care and Use Committees)批准。野生型C57BL/6小鼠获自CharlesRiver(#027)。

[2107] 体外生化分析

[2108] 使用HiScribe T7体外转录试剂盒(New England Biolabs)从含有T7启动子序列的PCR扩增模板中体外转录PEgRNA和sgRNA。RNA通过尿素变性PAGE纯化，并在使用前通过分析凝胶确认质量。通过加热至95℃3分钟，然后缓慢冷却至室温，使用用于未产生切口的底物的两种寡核苷酸(Cy5-AVA024和AVA025;1:1.1比例)或用于预先产生切口的底物的三种寡核苷酸(Cy5-AVA023、AVA025和AVA026;1:1.1:1.1)退火5′-Cy5标记的DNA双链体底物(表2A至2C)。在补充有dNTP的1x切割缓冲液(20mM HEPES-K,pH7.5;100mM KCl;5％glycerol;0.2mM EDTA,pH 8.0;3mM MgCl2;0.5mM dNTP mix;5mM DTT)[205]中进行Cas9切割和逆转录反应。dCas9或Cas9H840A(最终5μM)和sgRNA或PEgRNA(最终5μM)在5μL反应混合物中于室温预先温育10分钟，然后加入双链体DNA底物(最终400nM)，接着添加Superscript III逆转录酶(ThermoFisher Scientific)，合适时未公开的M-MLV RT变体。反应于37℃进行1小时，然后用水稀释至10μL体积，用0.2μL蛋白酶K溶液(20mg/mL,Thermo Fisher Scientific)处理，并于室温温育30分钟。于95℃加热灭活10分钟后，将反应产物与2x甲酰胺凝胶上样缓冲液(90％甲酰胺;10％甘油;0.01％溴酚蓝)混合，于95℃变性5分钟，然后通过尿素变性-PAGE凝胶(15％TBE-尿素，55℃，200V)分离。使用Typhoon FLA 7000生物分子成像仪通过Cy5荧光信号可视化DNA产物。

[2109] 使用预先温育的dCas9:sgRNA或dCas9:PEgRNA复合物(最终浓度范围为5nM至1μM)和Cy5标记的双链DNA(Cy5-AVA024和AVA025;最终浓度为20nM)，在1x结合缓冲液(1x裂解缓冲液+10μg/mL肝素)中进行电泳迁移率变化分析。于37℃温育15分钟后，样品通过天然PAGE凝胶(10％TBE)分析并进行Cy5荧光成像。

[2110] 对于逆转录产物的DNA测序，从尿素-PAGE凝胶中切下和纯化荧光带，然后根据制

造商方案在dGTP或dATP存在下利用末端转移酶(TdT;New EnglandBiolabs)进行3'加尾。加尾DNA产物用结合缓冲液(40%饱和氯化胍水溶液+60%异丙醇)稀释10倍,并通过QIAquick离心柱(Qiagen)纯化,然后用作模板,使用引物AVA134(A加尾产物)或AVA135(G加尾产物)通过Klenow片段(New EnglandBiolabs)进行引物延伸(表2A至2C)。使用引物AVA110和AVA122通过PCR扩增延伸10个循环,然后使用Sanger方法利用AVA037测序(表2A至2C)。

[2111] 酵母荧光报告基因测定

[2112] 如上所述,所有包含框内终止密码子、+1移码或-1移码的双荧光报告质粒在体外进行5′-延伸PEgRNA或3′-延伸PEgRNA引导编辑反应。于37℃温育1小时后,用水稀释反应物,并用0.3M乙酸钠和70%乙醇沉淀质粒DNA。如前所述[67],通过电穿孔将重悬的DNA转化至酿酒酵母中,并接种于不含亮氨酸(SC(葡萄糖),L-)的合成完全培养基。利用TyphoonFLA7000生物分子成像仪从克隆中可视化GFP和mCherry荧光信号。

[2113] 一般哺乳动物细胞培养条件

[2114] HEK293T(ATCC CRL-3216)、U2OS(ATTC HTB-96)、K562(CCL-243)和HeLa(CCL-2)细胞购自ATCC并在以下培养基中培养和传代:Dulbecco改良Eagle培养基(DMEM)加上GlutaMAX(ThermoFisher Scientific)、McCoy's5A培养基(Gibco)、RPMI培养基1640加上GlutaMAX(Gibco)或Eagle最低基础培养基(EMEM、ATCC),分别各自添加10%(v/v)胎牛血清(Gibco,合格)和1x青霉素链霉素(Coring)。所有细胞类型均于37℃和5%CO_2下温育、维持和培养。细胞系由各自供应商鉴定,并且支原体检测呈阴性。

[2115] HEK293T组织培养转染方案和基因组DNA制备

[2116] 将生长的HEK293T细胞接种于48孔聚-D-赖氨酸包被的板(Corning)。接种后16至24小时,根据制造商方案,用1μL Lipofectamine 2000(Thermo Fisher Scientific)和750ng PE质粒、250ng PEgRNA质粒和83ng sgRNA以约60%的汇合度转染细胞质粒(用于PE3和PE3b)。除非另有说明,细胞在转染后培养3天,然后去除培养基,用1xPBS溶液(Thermo Fisher Scientific)洗涤细胞,并通过直接向组织培养板的每个孔添加150μL新鲜制备的裂解缓冲液(10mM Tris-HCl,pH 7.5;0.05%SDS;25μg/mL蛋白酶K(Thermo Fisher Scientific))提取基因组DNA。基因组DNA混合物于37℃温育1至2小时,然后于80℃进行30分钟的酶灭活步骤。用于哺乳动物细胞基因组DNA扩增的引物列于表4。对于HEK293T细胞中的HDR实验,使用每孔1.4μL Lipofectamine 2000(Thermo Fisher)脂质转染231ng核酸酶表达质粒,69ng sgRNA表达质粒,50ng(1.51pmol)100-nt ssDNA供体模板(PAGE-纯化;IntegratedDNATechnologies)。根据制造商方案,使用AgencourtDNAdvance试剂盒(BeckmanCoulter)纯化来自所有HDR实验的基因组DNA。

[2117] 基因组DNA样品的高通量DNA测序

[2118] 从基因组DNA样品中扩增感兴趣的基因组位点,并用IlluminaMiSeq进行测序,如前所述并进行以下修改[129,130]。简言之,包含Illumina正向和反向适体的扩增引物(表4)用于第一轮PCR(PCR1)扩增感兴趣的基因组区域。利用0.5μM正向和反向引物、1μL基因组DNA提取物和12.5μL Phusion U Green Multiplex PCRMaster Mix进行25μLPCR1反应。PCR反应如下进行:98℃2分钟,接着[98℃10秒,61℃20秒,72℃30秒]的30个循环,最后72℃延伸2分钟。在第二轮PCR反应(PCR2)中,将独特的Illumina条形码引物对添加到每个样品中。具体地,25μL给定的PCR2反应包含各0.5μM独特的正向和反向Illumina条形码引物对、1μL未

纯化的PCR1反应混合物和12.5μL Phusion U Green Multiplex PCR 2x MasterMix。条形码PCR2反应如下进行:98℃2分钟,然后[98℃10秒、61℃20秒和72℃30秒]的12个循环,最后72℃延伸2分钟。PCR产物通过在1.5%琼脂糖凝胶中电泳进行分析评估。PCR2产物(由共同扩增子合并)使用QIAquick凝胶提取试剂盒(Qiagen)通过1.5%琼脂糖凝胶电泳纯化,用40μL水洗脱。根据制造商的方案,DNA浓度通过荧光定量(Qubit,Thermo Fisher Scientific)或qPCR(KAPA文库定量试剂盒-Illumina,KAPA Biosystems)测量,并用Illumina MiSeq仪器测序。

[2119] 使用MiSeq Reporter(Illumina)对测序读段进行多路分解。使用CRISPResso2[178]进行扩增子序列与参考序列的比对。为了定量点突变编辑,CRISPResso2在标准模式下运行,并启用"丢弃_插入缺失_读段"("discard_插入缺失_reads")。编辑效率计算为:(非丢弃读段中指定点突变的频率)x(非丢弃读段的数目#)÷总读段。对于插入或缺失编辑,CRISPResso2在HDR模式下运行,使用期望的等位基因作为预期的等位基因(e flag),并启用"丢弃_插入缺失_读段"。编辑产率计算为HDR比对读段的数目除以总读段。对于所有编辑,插入缺失产率计算为丢弃读段的数目除以总读段。

[2120] U2OS、K562和HeLa细胞的核转染

[2121] 在所有实验中使用K562、HeLa和U2OS细胞进行核转染。对于这些细胞类型的PE条件,在16孔nucleocuvette strip(Lonza)中800ng引导编辑器表达质粒、200ngPEgRNA表达质粒和83ng切口质粒以20μL的最终体积进行核转染。对于这三种细胞类型的HDR条件,在16孔Nucleocuvette strip(Lonza)中350ng核酸酶表达质粒、150ng sgRNA表达质粒和200pmol(6.6μg)100-nt ssDNA供体模板(PAGE纯化;Integrated DNA Technologies)以每个样品20μL的最终体积进行核转染。根据制造商的方案,使用SF细胞系4D-Nucleofector X试剂盒(Lonza)对K562细胞进行核转染,每个样品5×10^5个细胞(程序FF-120)。根据制造商的方案,使用SE细胞系4D-Nucleofector X试剂盒(Lonza)对U2OS细胞进行核转染,每个样品3-4×10^5个细胞(程序DN-100)。根据制造商的方案,使用SE细胞系4D-NucleofectorX试剂盒(Lonza)对HeLa细胞进行核转染,每个样品2×10^5个细胞(程序CN-114)。在核转染后72小时收获细胞用于基因组DNA提取。

[2122] 用于HDR实验的基因组DNA提取

[2123] 根据制造商的方案,使用Agencourt DNAdvance试剂盒(BeckmanCoulter)纯化来自HEK293T、HEK293THBB E6V、K562、U2OS和HeLa细胞中所有HDR比较实验的基因组DNA。

[2124] PE2、PE3、BE2、BE4max、ABEdmax和ABEmax之间的比较

[2125] HEK293T细胞接种于48孔聚-D-赖氨酸包被的板(Corning)。16至24小时后,转染约60%汇合度的细胞。对于利用CBE或ABE构建体的碱基编辑,用750ng碱基编辑器质粒、250ng sgRNA表达质粒和1μL Lipofectamine 2000(ThermoFisher Scientific)转染细胞。如上所述进行PE转染。如上所述进行PE和BE的基因组DNA提取。

[2126] 确定已知Cas9脱靶位点处的PE3活性

[2127] 为评估已知Cas9脱靶位点处的PE3脱靶编辑活性,将用PE3转染3天后从HEK293T细胞提取的基因组DNA用作模板,对16个先前报告的Cas9脱靶基因组位点[118,159]进行PCR扩增(HEK3、EMX1、FANCF和HEK4间隔区各自的前四个脱靶位点;引物序列列于表4)。这些基因组DNA样品与用于定量图41A至41K所示的在靶PE3编辑活性的那些相同;PEgRNA和切口产生性

sgRNA序列列于表3A至3R。在对脱靶位点进行PCR扩增后，如上所述用Illumina MiSeq平台对扩增子进行测序(HTS分析)。为了确定Cas9核酸酶、Cas9H840A切口酶、dCas9和PE2-dRT的中靶和脱靶编辑活性，用750ng编辑器质粒(Cas9核酸酶、Cas9H840A切口酶、dCas9或PE2-dRT)、250ng PEgRNA或sgRNA质粒和1μL Lipofectamine 2000转染HEK293T细胞。如上所述，在转染后3天从细胞中分离基因组DNA。使用表4中的引物序列通过PCR扩增中靶和脱靶基因组位点，并用IlluminaMiSeq测序。

[2128] 使用CRISPResso2[178]进行HTS数据分析。将Cas9核酸酶、Cas9 H840A切口酶和dCas9的编辑效率定量为包含插入缺失的总测序读段的百分比。为定量PE3和PE3-dRT脱靶，检查比对的测序读段是否存在与Cas9切口位点处启动的PEgRNA逆转录的预期产物一致的点突变、插入或缺失。从分析中排除在样品内总读段中以<0.1％总体频率发生的单核苷酸变异。对于包含发生频率≥0.1％且与PEgRNA编码编辑部分一致的单核苷酸变异的读段，用t检验(未配对，单尾，α＝0.5)确定与用含有相同间隔区但编码不同编辑的PEgRNA处理的样品相比，变体是否以显著更高的水平存在。为避免测序错误的差异，在同一MiSeq运行内同时测序的样品之间进行比较。排除不符合p值>0.05标准的变体。然后将脱靶PE3编辑活性计算为满足上述标准的总测序读段的百分比。

[2129] 使用Cas9启动的HDR生成含有HBB E6V突变的HEK293T细胞系

[2130] 将HEK293T细胞接种于48孔板，并以约60％的汇合度转染1.5μL Lipofectamine 2000、300ng Cas9 D10A切口酶质粒、100ng sgRNA质粒和200ng 100-mer ssDNA供体模板(表5)。转染后3天，将培养基更换为新鲜培养基。转染后4天，细胞使用30μLTrypLE溶液分离并悬浮于1.5mL培养基中。通过荧光激活细胞分选(FACS)(Beckman-CoulterAstrios)，将单个细胞分离到两个96孔板的各个孔中。参见代表性FACS分选示例的图53A至53B。如上所述，在基因组DNA测序之前扩增细胞14天。在分离的克隆群体中，HBB E6V突变未发现纯合的，因此在部分编辑的细胞系中重复通过脂质转染、分选和生长的第二轮编辑，以产生E6V等位基因纯合的细胞系。

[2131] 使用PE3产生含有HBB E6V突变的HEK293T细胞系

[2132] 将在不存在抗生素的情况下生长的$2.5x10^4$个HEK293T细胞接种于48孔聚D-赖氨酸包被的板(Corning)。接种后16至24小时，用1μL Lipofectamine2000(Thermo Fisher Scientific，根据制造商的方案)以及750ngPE2-P2A-GFP质粒、250ngPEgRNA质粒和83ng sgRNA质粒转染约70％汇合度的细胞。转染后3天，用磷酸盐缓冲盐水(Gibco)洗涤细胞并使用TrypLE Express(Gibco)解离。然后用补充10％(v/v)FBS(Gibco)的DMEM加上GlutaMax(Thermo Fisher Scientific)稀释细胞，并在分选前通过35-μm细胞过滤器(Corning)。用LE-MA900细胞分选仪(Sony)进行流式细胞术。分选前，细胞用3nM DAPI(BioLegend)处理15分钟。在双重排除门控后，将具有高于GFP阴性对照细胞群的GFP荧光的单个DAPI阴性细胞分选到96孔平底细胞培养板(Corning)，其中填充有补充10％FBS的预冷DMEM和GlutaMax。参见代表性FACS分选示例的图53A至53B。如上所述，在基因组DNA提取和HTS表征之前，培养细胞10天。总共6个克隆细胞系鉴定为HBB中E6V突变纯合的。

[2133] 使用PE3生成含有HEXA1278+TATC插入的HEK293T细胞系

[2134] 按照上述用于创建HBB E6V细胞系的方案生成包含HEXA 1278+TATC等位基因的HEK293T细胞；PEgRNA和sgRNA序列列于表2A至2C，图43A至43H小标题。转染和分选后，在基

因组DNA提取和HTS表征之前，培养细胞10天，如上所述。分离出含有50％HEXA 1278+TATC等位基因的两个杂合细胞系，回收含有100％HEXA 1278+TATC等位基因的两个纯合细胞系。

[2135] 细胞活力测定

[2136] 将HEK293T细胞接种于48孔板，并如上所述用750ng编辑器质粒(PE3、PE3R110SK103L、Cas9H840A切口酶或dCas9)、250ng HEK3-靶向PEgRNA质粒和1μL Lipofectamine 2000转染约70％汇合度的细胞。根据制造商的方案，使用CellTiter-Glo2.0测定(Promega)在转染后每隔24小时测量一次细胞活力，持续3天。使用M1000 Pro酶标仪(Tecan)在96孔平底聚苯乙烯微孔板(Corning)中测量发光，积分时间为1秒。

[2137] 慢病毒生产

[2138] 如先前所述[206]生产慢病毒。根据制造商的说明，使用FuGENE HD(Promega，Madison,WI,USA)用慢病毒生产辅助质粒pVSV-G和psPAX2结合携带内含肽断裂PE2编辑器的经修饰lentiCRISPR_v2基因组转染快速分裂的HEK293T细胞(ATCC;Manassas,VA,USA)的T75细胞培养瓶。设计四种断裂内含肽编辑器构建体:1)编码U6-PEgRNA表达盒和与NpuN-内含肽、自切割P2A肽和GFP-KASH融合的Cas9 H840A切口酶的N端部分(1-573)的病毒基因组;2)编码与PE2的C端其余部分融合的Npu C-内含肽的病毒基因组;3)编码与Cas9的C端其余部分融合的Npu C-内含肽的病毒基因组，用作Cas9对照;以及4)DNMT1的切口产生性sgRNA。断裂内含肽介导反式剪接以连接PE2或Cas9的两个半部，而P2AGFP-KASH能够共同翻译产生核膜定位的GFP。48小时后，收集上清液，以500g离心5分钟去除细胞碎片，并使用0.45μm过滤器过滤。根据制造商的说明，使用PEG-it病毒沉淀溶液(System Biosciences,PaloAlto,CA,USA)浓缩过滤的上清液。使用原始培养基量的1％将所得沉淀重新悬浮于Opti-MEM(Thermo Fisher Scientific,Waltham,MA,USA)。将重悬的沉淀快速冷冻并在使用前于-80℃储存。

[2139] 小鼠原代皮质神经元解剖和培养

[2140] 从定时怀孕的C57BL/6小鼠(Charles River)收获E18.5分离的皮质培养物。在通过CO2实施安乐死后进行断头术从怀孕小鼠收获胚胎。在补充有青霉素/链霉素(Life Technologies)的冰冷Hibernate-E中解剖皮质帽(cap)。用冰冷Hibernate-E冲洗后，将组织用木瓜蛋白酶/DNase(Worthington/Sigma)于37℃消化8分钟。在补充有DNase的NBActiv4(BrainBits)中研磨组织。对细胞进行计数并以每孔100,000个细胞接种于24孔板。每周更换两次一半的培养基。

[2141] 原代神经元和核分离中的引导编辑

[2142] 在DIV1，15μL慢病毒添加10:10:1比例的N端:C端:切口产生性sgRNA。在DIV14，按照制造商的方案，使用EZ-PREP缓冲液(Sigma D8938)分离神经元核。所有步骤均在冰上或于4℃进行。从分离的培养物中去除培养基，并用冰冷PBS洗涤培养物。吸出PBS并用200μLEZ-PREP溶液更换。在冰上温育5分钟后，吸取越过孔表面的EZ-PREP以取出剩余细胞。将样品以500g离心5分钟，去除上清液。样品用200μLEZ-PREP洗涤，并再次以500g离心5分钟。将样品轻轻移入200μL冰冷的Nuclei Suspension缓冲液(NSB)中重悬，该缓冲液由1xPBS中的100μg/mLBSA和3.33μM Vybrant DyeCycle Ruby(Thermo Fisher)组成，然后以500g离心5分钟。去除上清液，将核重悬于100μL NSB中，并在Broad研究所流式细胞仪设备中使用MoFlo Astrios(Beckman Coulter)分选到100μL Agencourt DNAdvance裂解缓冲液中。根

据制造商的Agencourt DNAdvance说明,纯化基因组DNA。

[2143] RNA测序和数据分析

[2144] 用PRNP靶向或HEXA靶向性PEgRNA和PE2、PE2-dRT或Cas9 H840A切口酶共转染HEK293T细胞。转染后72小时,使用TRIzol试剂(Thermo Fisher)从细胞中收获总RNA,并利用RNeasy Mini试剂盒(Qiagen)纯化,包括柱上DNaseI处理。使用TruSeq链式总RNA文库制备试剂盒(Illumina)的rRNA去除方案从总RNA中去除核糖体,然后用RNACleanXP珠(Beckman Coulter)洗涤。按照制造商的方案,用SMARTer PrepXApollo NGS文库制备系统(Takara)使用核糖耗尽的RNA制备测序文库。所得到的文库用2200TapeStation(Agilent Technologies)可视化,使用Qubit dsDNA HS分析(Thermo Fisher)标准化,并用NextSeq550使用高输出v2流动槽(Illumina)测序为75-bp配对末端读段。Fastq文件使用bcl2fastq2版本2.20生成,并使用TrimGalore版本0.6.2(github.com/FelixKrueger/TrimGalore)修剪以去除低质量碱基、未配对序列和适体序列。使用RSEM版本1.3.1[207]将修剪后的读段与具有定制Cas9 H840A基因条目的智人基因组组装GRCh[148]进行比对。limma-voom[208]包用于标准化基因表达水平并利用批次效应校正进行差异表达分析。差异表达的基因以FDR校正的p值<0.05和倍数变化>2截止值命名,结果用R显示。

[2145] ClinVar分析

[2146] 从NCBI下载ClinVar变体摘要(2019年7月15日访问),其中包含的信息用于所有下游分析。所有报告的变体列表都通过等位基因ID过滤,以去除重复,并通过临床意义来限制对致病性变体的分析。致病性变体列表按变体类型顺序过滤,以计算插入、缺失等致病性变体的分数。基于报告的参考和替代等位基因,将单核苷酸变体(SNV)分为两类(转换和颠换)。未报告的参考或替代等位基因的SNV被排除在分析之外。

[2147] 使用参考/替代等位基因、变体起始/终止位点或变体名称中的适当识别信息计算报告的插入、缺失和重复的长度。未报告任何上述信息的变体被排除在分析之外。通过确定参考和替代等位基因之间的最佳双序列比对中错配或缺口的数目,计算报告的插入缺失(包括相对于参考基因组的插入和缺失的单个变体)的长度。使用GraphPad Prism8计算变体长度的频率分布。

[2148] 数据可用性

[2149] 高通量测序数据存入NCBI Sequence ReadArchive数据库。编码PE1、PE2/PE3和PEgRNA表达载体的质粒可从Addgene获得。

[2150] 代码可用性

[2151] 用于定量PEgRNA支架插入的脚本在图60A至60B中提供。

[2152] 补充信息:表格和序列

[2153] 表1:HEK3、HEK4、EMX1和FANCF中靶和脱靶位点处引导编辑器、Cas9核酸酶、Cas9 H840A切口酶和PE2-dRT的活性。PE2/PE3编辑显示为%引导编辑和%插入缺失(在括号中)。显示Cas9、Cas9 H840A切口酶(nCas9)和PE2-dRT在先前表征的前四个脱靶位点[179,180]处的%插入缺失。sgRNA和PEgRNA序列可见于表3A至3R,图42A至42H标题下。所有值都是三个独立生物学重复的平均值。

[2154]

PE

pegRNA 位点	HEK3 (PE3) ---	1	2	3	4	HEK4 (PE2) ---	1	2	3	4	EMX1 (PE3) ---	1	2	3	4	FANCF (PE3) ---	1	2	3	4
中靶		44.2 (11.9)	61.2 (8.8)	40.4 (16.5)	48.4 (3.3)		18.2 (0.9)	14.4 (1.8)	9.8 (2.0)	7.9 (2.2)		28.5 (3.5)	14.1 (2.4)	35.7 (3.3)	15.4 (2.9)		56.8 (9.3)	32.4 (16.7)	42.8 (13.5)	47.6 (12.0)
脱靶1		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)		<0.1 (<0.1)	0.4 (<0.1)	<0.1 (<0.1)	0.4 (<0.1)		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)		<0.1 (0.1)	0.6 (0.1	<0.1 (0.1)	<0.1 (0.1)
脱靶2		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)		<0.1 (<0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (0.1)	<0.1 (<0.1)
脱靶3		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)		0.2 (0.5)	6.8 (1.9)	19.2 (0.5)	7.9 (3.5)		<0.1 (0.3)	<0.1 (0.3)	<0.1 (0.3)	<0.1 (0.3)		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)
脱靶4		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (<0.1		<0.1 (<0.1)	<0.1 (0.1)	<0.1 (0.2)	<0.1 (<0.1		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.2)	<0.1 (0.1)		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)

Cas9

pegRNA 位点	HEK3 sgRNA	1	2	3	4	HEK4 sgRNA	1	2	3	4	EMX1 sgRNA	1	2	3	4	FANCF sgRNA	1	2	3	4
中靶	91.8	87.5	89.2	89.1	86.8	71.8	68.6	72.8	72.8	70.9	85.6	79.7	70.6	76.6	76.0	78.7	55.9	58.3	51.8	52.0
脱靶1	17.2	1.9	5.5	5.2	1.8	54.2	39.5	48.4	49.7	49.2	81.1	63.5	48.1	53.0	59.6	12.6	1.9	1.9	1.7	1.7
脱靶2	38.0	6.5	12.6	11.8	4.7	42.5	19.5	29.4	27.3	30.3	58.3	12.0	6.0	8.2	12.9	1.1	0.2	0.2	0.2	0.1
脱靶3	8.8	0.6	1.7	1.5	0.5	98.1	96.9	97.3	97.6	97.5	14.8	4.2	3.1	3.6	4.8	2.4	0.2	<0.1	0.2	0.2
脱靶4	0.3	<0.1	<0.1	0.1	<0.1	45.3	16.9	28.0	27.5	29.7	39.5	1.3	0.9	0.6	1.3	1.0	0.2	0.2	0.2	0.2

nCas9

pegRNA 位点	HEK3 ---	1	2	3	4	HEK4 ---	1	2	3	4	EMX1 ---	1	2	3	4	FANCF ---	1	2	3	4
脱靶1		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1		0.1	0.1	<0.1	0.1		<0.1	<0.1	<0.1	<0.1
脱靶2		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	< 0.1	<0.1		0.3	0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
脱靶3		<0.1	<0.1	<0.1	<0.1		0.3	0.5	0.7	0.7		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
脱靶4		<0.1	<0.1	<0.1	<0.1		<0.1	0.1	<0.1	<0.1		1.5	0.1	0.1	0.4		<0.1	<0.1	<0.1	<0.1

pegRNA 位点	HEK3 ---	1	2	3	4	HEK4 ---	1	2	3	4	EMX1 ---	1	2	3	4	FANCF ---	1	2	3	4
脱靶1		<0.1	<0.1	<0.1	<0.1		0.1	0.2	<0.1	<0.1		0.1	0.1	0.1	0.1		<0.1	<0.1	<0.1	<0.1
脱靶2		<0.1	<0.1	<0.1	<0.1		<0.1	0.1	<0.1	<0.1		<0.1	0.3	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
脱靶3		<0.1	<0.1	<0.1	<0.1		<0.1	0.1	1.4	0.9		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
脱靶4		<0.1	<0.1	<0.1	<0.1		0.1	0.1	0.1	0.2		0.9	0.1	0.1	0.2		<0.1	<0.1	<0.1	<0.1

[2155] 表2A至2C:在体外实验使用的DNA寡核苷酸、PEgRNA和sgRNA序列。

[2156] 表2A:DNA寡核苷酸

[2157]

寡核苷酸	序列
AVA023	5Cy5-CCTGGGTCAATCCTTGGGGCCCAGACTGAGCACG (SEQ ID NO: 374)
AVA024	5Cy5-CCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGG (SEQ ID NO: 375)
AVA025	5Phos-CCTTTCCTCTGCCATCACGTGCTCAGTCTGGGCCCCAAGGATTGACCCAGG (SEQ ID NO: 376)
AVA026	5Phos-TGATGGCAGAGGAAAGG (SEQ ID NO: 377)

AVA037	GCAGGCTTTAAAGGAACCAATTC (SEQ ID NO: 378)
AVA110	GCAGGCTTTAAAGGAACCAATTCCCTGGGTCAATCCTTGGGGC (SEQ ID NO: 379)
AVA122	CTCTGGAGGATCTAGCGGAG (SEQ ID NO: 380)
AVA134	CTCTGGAGGATCTAGCGGAGTTTTTTTTTTTTTTTTT (SEQ ID NO: 381)
AVA135	CTCTGGAGGATCTAGCGGAGCCCCCCCCCCCCC (SEQ ID NO: 382)

[2158]

[2159] 表2B:5′-延伸PEgRNA

PEGRNA	间隔区序列	5′ 延伸序列	接头长度(NT)	PBS长度(NT)	RT 模板长度(NT)
PEGRNA 1	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 383)	GGCTAACCGTGCCATTTGATCAGGTCA (SEQ ID NO: 429)	15	5	7
PEGRNA 2	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 384)	GGCTAACCGTGCAAATTAACAAACTAA(SEQ ID NO: 430)	15	5	7
PEGRNA 3	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 385)	GGCCATCTCGTGCAAATTAACAAACTAA(SEQ ID NO: 431)	15	5	8
PEGRNA 4	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 386)	GGTCCTCTGCCATCTCGTGCAAATTAACAAACTAA(SEQ ID NO: 432)	15	5	15
PEGRNA 5	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 387)	GGCTTCCTTTCCTCTGCCATCTCGTGCAAATTAACAAACTAA(SEQ ID NO: 433)	15	5	22
5′-PEGRNA_RT_7_A	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 388)	GGCTAACCGTGCCATTTGATCAGGTCA(SEQ ID NO: 434)	15	5	7
5′-PEGRNA_RT_7_B	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 389)	GGCTAACCGTGCAAATTAACAAACTAA(SEQ ID NO: 435)	15	5	7
5′-PEGRNA_RT_8	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 390)	GGCCATCTCGTGCAAATTAACAAACTAA(SEQ ID NO: 436)	15	5	8
5′-PEGRNA_RT_15	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 391)	GGTCCTCTGCCATCTCGTGCAAATTAACAAACTAA(SEQ ID NO: 437)	15	5	15
5′-PEGRNA_RT_22	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 392)	GGCTTCCTTTCCTCTGCCATCTCGTGCAAATTAACAAACTAA(SEQ ID NO: 438)	15	5	22

[2160]

[2161] 表2C:3′-延伸PEgRNA

PEGRNA	间隔区序列	3′延伸序列	PBS长度(NT)	RT 模板长度(NT)
3′-PEGRNA_10	GGCCCAGACTGAGCACGTGA(SEQ ID	TCTGCCATCTCGTGCTC (SEQ ID	7	10

[2162]

[2163]

	NO: 506)	NO: 439)		
3′-PEGRNA_YEAST_TTOA	GGCCCAGACTGAGCACGTGA(SEQ ID NO: 507)	TCTGCCATCTCGTGCTC(SEQ ID NO: 440)	7	10
3′-PEGRNA_YEAST_+1AINS	GGCCCAGACTGAGCACGTGA(SEQ ID NO: 508)	TCTGCCATCATCGTGCTC(SEQ ID NO: 441)	7	11
3′-PEGRNA_YEAST_+1TDEL	GGCCCAGACTGAGCACGTGA(SEQ ID NO: 509)	TCTGCCATCCGTGCTC(SEQ ID NO: 442)	7	9

[2164] 表3A至3R:哺乳动物细胞实验中使用的PEgRNA和sgRNA序列。所有序列以5′至3′方向显示。为了构建PEgRNA,在sgRNA支架的5′端添加如下所列的间隔区序列,并在sgRNA支架的3′端添加如下所列的包含引物结合位点和RT模板的3′延伸。sgRNA支架序列为

[2165] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC(SEQ ID NO:131)

[2166] 表3A:图39A至39D PEgRNA

[2167]

PEGRNA	间隔区序列(SEQ ID NO: 2890-2996)	3′ 延伸(SEQ ID NO: 2997-3103)	PBS 长度 (NT)	RT 模板 长度 (NT)
HEK3_2B-C_8	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCA	8	10
HEK3_2B-C_9	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAG	9	10
HEK3_2B-C_10	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGT	10	10
HEK3_2B-C_11	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTC	11	10
HEK3_2B-C_12	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCT	12	10
HEK3_2B-C_13	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_2B-C_14	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGG	14	10
HEK3_2B-C_15	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGGG	15	10
HEK3_2C_16	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGGGC	16	10
HEK3_2C_17	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGGGCC	17	10
EMX1_2C_9	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGC	9	13
EMX1_2C_10	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCT	10	13
EMX1_2C_11	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTC	11	13

[2168]

EMX1_2C_12	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCG	12	13
EMX1_2C_13	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGG	13	13
EMX1_2C_14	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGA	14	13
EMX1_2C_15	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGAC	15	13
EMX1_2C_16	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGACT	16	13
EMX1_2C_17	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGACTC	17	13
FANCF_2C_8	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGA	8	17
FANCF_2C_9	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAA	9	17
FANCF_2C_10	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAG	10	17
FANCF_2C_11	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGG	11	17
FANCF_2C_12	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGG	12	17
FANCF_2C_13	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	17
FANCF_2C_14	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGAT	14	17
FANCF_2C_15	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATT	15	17
FANCF_2C_16	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATTC	16	17
FANCF_2C_17	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATTCC	17	17
RNF2_2C_9	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACT	9	11
RNF2_2C_10	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTA	10	11
RNF2_2C_11	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAA	11	11
RNF2_2C_12	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAG	12	11
RNF2_2C_13	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGA	13	11
RNF2_2C_14	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGAT	14	11
RNF2_2C_15	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATG	15	11
RNF2_2C_16	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATGA	16	11
RNF2_2C_17	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATGAC	17	11
HEK4_2C_7	GGCACTGCGGCTG	GCTTTAACCCCAACCTC	7	13

	GAGGTGG	CAG		
HEK4_2C_8	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGC	8	13
HEK4_2C_9	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCC	9	13
HEK4_2C_10	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCG	10	13
HEK4_2C_11	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGC	11	13
HEK4_2C_12	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCA	12	13
HEK4_2C_13	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCAG	13	13
HEK4_2C_14	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCAGT	14	13
HEK4_2C_15	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCAGTG	15	13
HEK3_2C_1TDEL	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	10
HEK3_2C_1AINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCATCGTGCTCAGTCTG	13	10
HEK3_2C_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
HEK3_2D_10	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_2D_11	GGCCCAGACTGAGCACGTGA	CTCTGCCATCTCGTGCTCAGTCTG	13	11
HEK3_2D_12	GGCCCAGACTGAGCACGTGA	CCTCTGCCATCTCGTGCTCAGTCTG	13	12
HEK3_2D_13	GGCCCAGACTGAGCACGTGA	TCCTCTGCCATCTCGTGCTCAGTCTG	13	13
HEK3_2D_14	GGCCCAGACTGAGCACGTGA	TTCCTCTGCCATCTCGTGCTCAGTCTG	13	14
HEK3_2D_15	GGCCCAGACTGAGCACGTGA	TTTCCTCTGCCATCTCGTGCTCAGTCTG	13	15
HEK3_2D_16	GGCCCAGACTGAGCACGTGA	CTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	16
HEK3_2D_17	GGCCCAGACTGAGCACGTGA	CCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	17
HEK3_2D_18	GGCCCAGACTGAGCACGTGA	TCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	18
HEK3_2D_19	GGCCCAGACTGAGCACGTGA	TTCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	19
HEK3_2D_20	GGCCCAGACTGAGCACGTGA	CTTCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	20
EMX1_2D_10	GAGTCCGAGCAGAAGAAGAA	GGAGCCCTTGTTCTTCTGCTCGG	13	10
EMX1_2D_11	GAGTCCGAGCAGAAGAAGAA	GGGAGCCCTTGTTCTTCTGCTCGG	13	11
EMX1_2D_12	GAGTCCGAGCAGAAGAAGAA	TGGGAGCCCTTGTTCTTCTGCTCGG	13	12

[2169]

[2170]

EMX1_2D_13	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCCCTTGTTCTTCTGCTCGG	13	13
EMX1_2D_14	GAGTCCGAGCAGAAGAAGAA	GATGGGAGCCCTTGTTCTTCTGCTCGG	13	14
EMX1_2D_15	GAGTCCGAGCAGAAGAAGAA	TGATGGGAGCCCTTGTTCTTCTGCTCGG	13	15
EMX1_2D_16	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	16
EMX1_2D_17	GAGTCCGAGCAGAAGAAGAA	TGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	17
EMX1_2D_18	GAGTCCGAGCAGAAGAAGAA	ATGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	18
EMX1_2D_19	GAGTCCGAGCAGAAGAAGAA	GATGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	19
EMX1_2D_20	GAGTCCGAGCAGAAGAAGAA	TGATGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	20
FANCF_2D_10	GGAATCCCTTCTGCAGCACC	CGATCAAGGTGCTGCAGAAGGGA	13	10
FANCF_2D_11	GGAATCCCTTCTGCAGCACC	GCGATCAAGGTGCTGCAGAAGGGA	13	11
FANCF_2D_12	GGAATCCCTTCTGCAGCACC	AGCGATCAAGGTGCTGCAGAAGGGA	13	12
FANCF_2D_13	GGAATCCCTTCTGCAGCACC	AAGCGATCAAGGTGCTGCAGAAGGGA	13	13
FANCF_2D_14	GGAATCCCTTCTGCAGCACC	AAAGCGATCAAGGTGCTGCAGAAGGGA	13	14
FANCF_2D_15	GGAATCCCTTCTGCAGCACC	AAAAGCGATCAAGGTGCTGCAGAAGGGA	13	15
FANCF_2D_16	GGAATCCCTTCTGCAGCACC	GAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	16
FANCF_2D_17	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	17
FANCF_2D_18	GGAATCCCTTCTGCAGCACC	CGGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	18
FANCF_2D_19	GGAATCCCTTCTGCAGCACC	TCGGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	19
FANCF_2D_20	GGAATCCCTTCTGCAGCACC	CTCGGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	20
RNF2_2D_10	GTCATCTTAGTCATTACCTG	AACACCTCATGTAATGACTAAGATG	15	10
RNF2_2D_11	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATG	15	11
RNF2_2D_12	GTCATCTTAGTCATTACCTG	CGAACACCTCATGTAATGACTAAGATG	15	12
RNF2_2D_13	GTCATCTTAGTCATTACCTG	ACGAACACCTCATGTAATGACTAAGATG	15	13
RNF2_2D_14	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
RNF2_2D_15	GTCATCTTAGTCATTACCTG	CAACGAACACCTCATGTAATGACTAAGATG	15	15
RNF2_2D_1	GTCATCTTAGTCAT	ACAACGAACACCTCATG	15	16

	6	TACCTG	TAATGACTAAGATG		
	RNF2_2D_17	GTCATCTTAGTCATTACCTG	TACAACGAACACCTCATGTAATGACTAAGATG	15	17
	RNF2_2D_18	GTCATCTTAGTCATTACCTG	TTACAACGAACACCTCATGTAATGACTAAGATG	15	18
	RNF2_2D_19	GTCATCTTAGTCATTACCTG	GTTACAACGAACACCTCATGTAATGACTAAGATG	15	19
	RNF2_2D_20	GTCATCTTAGTCATTACCTG	AGTTACAACGAACACCTCATGTAATGACTAAGATG	15	20
	HEK4_2D_7	GGCACTGCGGCTGGAGGTGG	ACCCCAACCTCCAGCCGC	11	7
	HEK4_2D_8	GGCACTGCGGCTGGAGGTGG	AACCCCAACCTCCAGCCGC	11	8
	HEK4_2D_9	GGCACTGCGGCTGGAGGTGG	TAACCCCAACCTCCAGCCGC	11	9
	HEK4_2D_10	GGCACTGCGGCTGGAGGTGG	TTAACCCCAACCTCCAGCCGC	11	10
[2171]	HEK4_2D_11	GGCACTGCGGCTGGAGGTGG	TTTAACCCCAACCTCCAGCCGC	11	11
	HEK4_2D_12	GGCACTGCGGCTGGAGGTGG	CTTTAACCCCAACCTCCAGCCGC	11	12
	HEK4_2D_13	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGC	11	13
	HEK4_2D_14	GGCACTGCGGCTGGAGGTGG	CGCTTTAACCCCAACCTCCAGCCGC	11	14
	HEK4_2D_15	GGCACTGCGGCTGGAGGTGG	CCGCTTTAACCCCAACCTCCAGCCGC	11	15
	HEK4_2D_16	GGCACTGCGGCTGGAGGTGG	TCCGCTTTAACCCCAACCTCCAGCCGC	11	16
	HEK4_2D_17	GGCACTGCGGCTGGAGGTGG	CTCCGCTTTAACCCCAACCTCCAGCCGC	11	17
	HEK4_2D_18	GGCACTGCGGCTGGAGGTGG	CTCCGCTTTAACCCCAACCTCCAGCCGC	11	18
	HEK4_2D_19	GGCACTGCGGCTGGAGGTGG	CTCCGCTTTAACCCCAACCTCCAGCCGC	11	19

[2172] 表3B:图40A至40C PEgRNA

PEGRNA	间隔区序列(SEQ ID NO: 3104-3112)	3′延伸(SEQ ID NO: 3113-3121)	PBS 长度 (NT)	RT 模板长度 (NT)
RNF2_3B	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
EMX1_3B	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGAC	15	13
FANCF_3B	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATT	15	17
HE3_3B	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK4_3B	GGCACTGCGGCTGGAGGTGG	TTAACCCCAACCTCCAGCC	9	10
RNF2_3C_4	GTCATCTTAGTCAT	AACGAACACCGCAGGTA	15	14

[2173]

[2174]

ATOC	TACCTG	ATGACTAAGATG		
RNF2_3C_4ATOG	GTCATCTTAGTCATTACCTG	AACGAACACCCCAGGTAATGACTAAGATG	15	14
FANCF_3C_5GTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	17
FANCF_3C_7ATOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGAGCCAGGTGCTGCAGAAGGGAT	14	17

[2175] 表3C:图40A至40C切口产生性sgRNA序列

[2176]

切口产生性SGRNA	间隔区序列	SEQIDNO:
RNF2_2B_+41	GTCAACCATTAAGCAAAACAT	3122
RNF2_2B_+67	GTCTCAGGCTGTGCAGACAAA	3123
EMX1_2B_-116	GGGGCACAGATGAGAAACTC	3124
EMX1_2B_-57	GCCGTTTGTACTTTGTCCTC	3125
EMX1_2B_+14	GCGCCACCGGTTGATGTGAT	3126
EMX1_2B_+27	GCTTCGTGGCAATGCGCCAC	3127
EMX1_2B_+53	GACATCGATGTCCTCCCCAT	3128
EMX1_2B_+80	GTGGTTGCCCACCCTAGTCAT	3129
FANCF_2B_-78	GCGACTCTCTGCGTACTGAT	3130
FANCF_2B_-50	GCCCTACTTCCGCTTTCACCT	3131
FANCF_2B_-27	GGATTCCATGAGGTGCGCGA	3132
FANCF_2B_-17	GCTGCAGAAGGGATTCCATG	3133
FANCF_2B_+21	GCTTGAGACCGCCAGAAGCT	3134
FANCF_2B_+48	GGGGTCCCAGGTGCTGACGT	3135
HEK3_2B_-108	GCAGAAATAGACTAATTGCA	3136
HEK3_2B_-38	GGATTGACCCAGGCCAGGGC	3137
HEK3_2B_+26	GACGCCCTCTGGAGGAAGCA	3138
HEK3_2B_+37	GCTGTCCTGCGACGCCCTC	3139
HEK3_2B_+63	GCACATACTAGCCCCTGTCT	3140
HEK3_2B_+90	GTCAACCAGTATCCCGGTGC	3141
HEK4_2B_-95	TCCCTTCCTTCCACCCAGCC	3142
HEK4_2B_-51	CCCTGCCTGTCATCCTGCTT	3143
HEK4_2B_-26	GCAGTGCCACCGGGGCGCCG	3144
HEK4_2B_+52	GCGGGGGCTCAGAGAGGGCA	3145
HEK4_2B_+74	GAGACACACACACAGGCCTGG	3146
RNF2_2C_+41	GTCAACCATTAAGCAAAACAT	3147
RNF2_2C_4ATOC_+5	GTGAGTTACAACGAACACCGC	3148
RNF2_2C_4ATOG_+5	GTGAGTTACAACGAACACCCC	3149
FANCF_2C_+48	GGGGTCCCAGGTGCTGACGT	3150
FANCF_2C_5GTOT_+7	GAAGCTCGGAAAAGCGATCA	3151
FANCF_2C_7ATOC_+7	GAAGCTCGGAAAAGCGAGCC	3152

HEK3_2C_+90	GTCAACCAGTATCCCGGTGC	3153

[2177] 表3D:图41A至41KPEgRNA

[2178]

PEGRNA	间隔区序列(SEQ ID NO: 3154-3304)	3′延伸(SEQ ID NO: 3305-3455)	PBS 长度 (NT)	RT 模板长度 (NT)
HEK3_4A_1TTOA	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_4A_	GGCCCAGACTGAG	TCTGCCATCGCGTGCTCA	13	10

[2179]

1TTOC	CACGTGA	GTCTG		
HEK3_4A_1TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCATCCCGTGCTCAGTCTG	13	10
HEK3_4A_2GTOA	GGCCCAGACTGAGCACGTGA	TCTGCCATTACGTGCTCAGTCTG	13	10
HEK3_4A_2GTOC	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK3_4A_2GTOT	GGCCCAGACTGAGCACGTGA	TCTGCCATAACGTGCTCAGTCTG	13	10
HEK3_4A_3ATOC	GGCCCAGACTGAGCACGTGA	TCTGCCAGCACGTGCTCAGTCTG	13	10
HEK3_4A_3ATOG	GGCCCAGACTGAGCACGTGA	TCTGCCACCACGTGCTCAGTCTG	13	10
HEK3_4A_3ATOT	GGCCCAGACTGAGCACGTGA	TCTGCCAACACGTGCTCAGTCTG	13	10
HEK3_4A_4TTOA	GGCCCAGACTGAGCACGTGA	TCTGCCTTCACGTGCTCAGTCTG	13	10
HEK3_4A_4TTOC	GGCCCAGACTGAGCACGTGA	TCTGCCGTCACGTGCTCAGTCTG	13	10
HEK3_4A_4TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCCTCACGTGCTCAGTCTG	13	10
HEK3_4A_5GTOA	GGCCCAGACTGAGCACGTGA	TCTGCTATCACGTGCTCAGTCTG	13	10
HEK3_4A_5GTOC	GGCCCAGACTGAGCACGTGA	TCTGCGATCACGTGCTCAGTCTG	13	10
HEK3_4A_5GTOT	GGCCCAGACTGAGCACGTGA	TCTGCAATCACGTGCTCAGTCTG	13	10
HEK3_4A_6GTOA	GGCCCAGACTGAGCACGTGA	TCTGTCATCACGTGCTCAGTCTG	13	10
HEK3_4A_6GTOC	GGCCCAGACTGAGCACGTGA	TCTGGCATCACGTGCTCAGTCTG	13	10
HEK3_4A_6GTOT	GGCCCAGACTGAGCACGTGA	TCTGACATCACGTGCTCAGTCTG	13	10
HEK3_4A_7CTOA	GGCCCAGACTGAGCACGTGA	TCTTCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_7CTOG	GGCCCAGACTGAGCACGTGA	TCTCCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_7CTOT	GGCCCAGACTGAGCACGTGA	TCTACCATCACGTGCTCAGTCTG	13	10
HEK3_4A_8ATOC	GGCCCAGACTGAGCACGTGA	TCGGCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_8ATOG	GGCCCAGACTGAGCACGTGA	TCCGCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_8ATOT	GGCCCAGACTGAGCACGTGA	TCAGCCATCACGTGCTCAGTCTG	13	10
HEK3_4B_1TTOA	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	34
HEK3_4B_12GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTGCTCTGCCATCACGTGCTCAGTCTG	13	34

	HEK3_4B_14ATOT	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTATCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_17GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTGCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_20GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGGTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_23CTOG	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCACGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_24TTOA	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCTGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_26CTOG	GGCCCAGACTGAGCACGTGA	TGGAGGAACCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_30CTOG	GGCCCAGACTGAGCACGTGA	TGGACGAAGCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
	HEK3_4B_33CTOG	GGCCCAGACTGAGCACGTGA	TCGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
[2180]	RNF2_4C_1CTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
	RNF2_4C_1CTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCACGTAATGACTAAGATG	15	14
	RNF2_4C_1CTOT	GTCATCTTAGTCATTACCTG	AACGAACACCTCAAGTAATGACTAAGATG	15	14
	RNF2_4C_2TTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCTGGTAATGACTAAGATG	15	14
	RNF2_4C_2TTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCCGGTAATGACTAAGATG	15	14
	RNF2_4C_3GTOC	GTCATCTTAGTCATTACCTG	AACGAACACCTGAGGTAATGACTAAGATG	15	14
	RNF2_4C_4ATOC	GTCATCTTAGTCATTACCTG	AACGAACACCGCAGGTAATGACTAAGATG	15	14
	RNF2_4C_4ATOT	GTCATCTTAGTCATTACCTG	AACGAACACCACAGGTAATGACTAAGATG	15	14
	RNF2_4C_4ATOG	GTCATCTTAGTCATTACCTG	AACGAACACCCCAGGTAATGACTAAGATG	15	14
	RNF2_4C_5GTOT	GTCATCTTAGTCATTACCTG	AACGAACACATCAGGTAATGACTAAGATG	15	14
	RNF2_4C_6GTOA	GTCATCTTAGTCATTACCTG	AACGAACATCTCAGGTAATGACTAAGATG	15	14
	RNF2_4C_7TTOC	GTCATCTTAGTCATTACCTG	AACGAACGCCTCAGGTAATGACTAAGATG	15	14
	FANCF_4D_1ATOG	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGCGCTGCAGAAGGGAT	14	17
	FANCF_4D	GGAATCCCTTCTGC	GGAAAAGCGATCCAGGA	14	17

[2181]

_1ATOT	AGCACC	GCTGCAGAAGGGAT		
FANCF_4D_2CTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGTTGCTGCAGAAGGGAT	14	17
FANCF_4D_3CTOG	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCACGTGCTGCAGAAGGGAT	14	17
FANCF_4D_3CTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAAGTGCTGCAGAAGGGAT	14	17
FANCF_4D_4TTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCTGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_4TTOG	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCCGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_5GTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCTAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_6GTOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATGCAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_7ATOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGAGCCAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_8TTOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGGTCCAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_10GTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGAGATCCAGGTGCTGCAGAAGGGAT	14	17
EMX1_4E_2ATOC	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTGCTTCTTCTGCTCGGA	14	16
EMX1_4E_2ATOT	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTACTTCTTCTGCTCGGA	14	16
EMX1_4E_3ATOG	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCCTCTTCTTCTGCTCGGA	14	16
EMX1_4E_4GTOC	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCGTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_5GTOA	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCTCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_5GTOT	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCACTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_7CTOA	GAGTCCGAGCAGAAGAAGAA	GTGATGGGATCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_8TTOA	GAGTCCGAGCAGAAGAAGAA	GTGATGGGTGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_8TTOC	GAGTCCGAGCAGAAGAAGAA	GTGATGGGGGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_8TTOG	GAGTCCGAGCAGAAGAAGAA	GTGATGGGCGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_9CTOG	GAGTCCGAGCAGAAGAAGAA	GTGATGGCAGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_9CTOT	GAGTCCGAGCAGAAGAAGAA	GTGATGGAAGCCCTTCTTCTTCTGCTCGGA	14	16
RUNX1_4F_1CTOA	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCTCTTCCTCCTGAAAAT	15	15
RUNX1_4F_1CTOG	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCCCTTCCTCCTGAAAAT	15	15
RUNX1_4F_1CTOT	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCACTTCCTCCTGAAAAT	15	15

[2182]

RUNX1_4F_2GTOA	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATTGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_3ATOC	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCAGCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_3ATOG	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCACCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_3ATOT	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCAACGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_4TTOA	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCTTCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_4TTOC	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCGTCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_4TTOG	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCCTCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_5GTOT	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_6GTOC	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGGCATCGCTTCCTCCTGAAAAT	15	15
VEGFA_4G_1TTOA	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTCTTCTGGCCTGCAGA	13	22
VEGFA_4G_1TTOC	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTCGTCTGGCCTGCAGA	13	22
VEGFA_4G_1TTOG	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTCCTCTGGCCTGCAGA	13	22
VEGFA_4G_2GTOA	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTTATCTGGCCTGCAGA	13	22
VEGFA_4G_3ATOC	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCGCATCTGGCCTGCAGA	13	22
VEGFA_4G_3ATOG	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCCCATCTGGCCTGCAGA	13	22
VEGFA_4G_3ATOT	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCACATCTGGCCTGCAGA	13	22
VEGFA_4G_5GTOT	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	22
VEGFA_4G_6GTOC	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGGCCTCATCTGGCCTGCAGA	13	22
VEGFA_4G_7CTOA	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGATCCCTCATCTGGCCTGCAGA	13	22
VEGFA_4G_7CTOT	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAACCCTCATCTGGCCTGCAGA	13	22
VEGFA_4G_9CTOG	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGCAGCCCTCATCTGGCCTGCAGA	13	22
DNMT1_4H_1ATOC	GATTCCTGGTGCCAGAAACA	GTCACCCCTGGTTCTGGCACCAGG	13	11
DNMT1_4H_1ATOG	GATTCCTGGTGCCAGAAACA	GTCACCCCTGCTTCTGGCACCAGG	13	11
DNMT1_4H_2CTOA	GATTCCTGGTGCCAGAAACA	GTCACCCCTTTTTCTGGCACCAGG	13	11
DNMT1_4H_2CTOG	GATTCCTGGTGCCAGAAACA	GTCACCCCTCTTTCTGGCACCAGG	13	11
DNMT1_4H	GATTCCTGGTGCCA	GTCACCCCTATTTCTGGC	13	11

[2183]

_2CTOT	GAAACA	ACCAGG		
DNMT1_4H_3ATOT	GATTCCTGGTGCCAGAAACA	GTCACCCCAGTTTCTGGCACCAGG	13	11
DNMT1_4H_4GTOA	GATTCCTGGTGCCAGAAACA	GTCACCCTTGTTTCTGGCACCAGG	13	11
DNMT1_4H_5GTOT	GATTCCTGGTGCCAGAAACA	GTCACCACTGTTTCTGGCACCAGG	13	11
DNMT1_4H_6GTOC	GATTCCTGGTGCCAGAAACA	GTCACGCCTGTTTCTGGCACCAGG	13	11
DNMT1_4H_8TTOA	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCTCCCCTGTTTCTGGCACCAGG	13	19
DNMT1_4H_8TTOC	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCGCCCCTGTTTCTGGCACCAGG	13	19
DNMT1_4H_8TTOG	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCCCCCCTGTTTCTGGCACCAGG	13	19
HEK3_4J_DEL1-5	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCGTGCTCAGTCTG	13	29
HEK3_4J_DEL1-10	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCCGTGCTCAGTCTG	13	24
HEK3_4J_DEL1-15	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCCGTGCTCAGTCTG	13	19
HEK3_4J_DEL1-25	GGCCCAGACTGAGCACGTGA	TGTCCTGCGACGCCCTCTGGAGGAAGCGTGCTCAGTCTG	13	26
HEK3_4J_DEL1-30	GGCCCAGACTGAGCACGTGA	TGTCCTGCGACGCCCTCTGGACGTGCTCAGTCTG	13	21
HEK3_4J_DEL1-80	GGCCCAGACTGAGCACGTGA	AGTATCCCGGTGCAGGAGCTCGTGCTCAGTCTG	13	20
HEK3_4I_1AINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCATCGTGCTCAGTCTG	13	11
HEK3_4I_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	13
HEK3_4I_1TDEL	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	9
HEK3_4I_1-3TGADEL	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCACGTGCTCAGTCTG	13	31
RNF2_4I_1TINS	GTCATCTTAGTCATTACCTG	AACGAACACCTCAGAGTAATGACTAAGATG	15	15
RNF2_4I_1GTAINS	GTCATCTTAGTCATTACCTG	AACGAACACCTCAGTACGTAATGACTAAGATG	15	17
RNF2_4I_4ADEL	GTCATCTTAGTCATTACCTG	AACGAACACCCAGGTAATGACTAAGATG	15	13
RNF2_4I_3-5GAGDEL	GTCATCTTAGTCATTACCTG	AACGAACACAGGTAATGACTAAGATG	15	11
FANCF_4I_3CINS	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGGTGCTGCAGAAGGGAT	14	18
FANCF_4I_4GATINS	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAATCGGTGCTGCAGAAGGGAT	14	20

	FANCF_4I_6GDEL	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAGGTGCTGCAGAAGGGAT	14	16
	FANCF_4I_5-7GGADEL	GGAATCCCTTCTGCAGCACC	GGAAAAGCGAAGGTGCTGCAGAAGGGAT	14	14
	EMX1_4I_6TINS	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCACCTTCTTCTTCTGCTCGGA	14	17
	EMX1_4I_1TGCINS	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTCGCATTCTTCTGCTCGGA	14	19
	EMX1_4I_5GDEL	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCTTCTTCTTCTGCTCGGA	14	15
	EMX1_4I_4-6GGGDEL	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTCTTCTTCTGCTCGGA	14	13
	RUNX1_4I_1CINS	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCGGCTTCCTCCTGAAAAT	15	16
	RUNX1_4I_1ATGINS	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCCATGCTTCCTCCTGAAAAT	15	18
	RUNX1_4I_2GDEL	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATGCTTCCTCCTGAAAAT	15	14
	RUNX1_4I_2-4GATDEL	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCGCTTCCTCCTGAAAAT	15	12
	VEGFA_4I_4CINS	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCGCTCATCTGGCCTGCAGA	13	23
[2184]	VEGFA_4I_2ACAINS	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTTGTCATCTGGCCTGCAGA	13	25
	VEGFA_4I_3ADEL	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCCATCTGGCCTGCAGA	13	21
	VEGFA_4I_2-4GAGDEL	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCATCTGGCCTGCAGA	13	19
	DNMT1_4I_4CINS	GATTCCTGGTGCCAGAAACA	TCCCGTCACCCGCTGTTTCTGGCACCAGG	13	16
	DNMT1_4I_1TCAINS	GATTCCTGGTGCCAGAAACA	TCCCGTCACCCCTGTGATTTCTGGCACCAGG	13	18
	DNMT1_4I_3ADEL	GATTCCTGGTGCCAGAAACA	TCCCGTCACCCCGTTTCTGGCACCAGG	13	14
	DNMT1_4I_3-5AGGDEL	GATTCCTGGTGCCAGAAACA	TCCCGTCACCGTTTCTGGCACCAGG	13	12
	HEK3_4K_1CTTINS_5GDEL	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCATCAAAGCGTGCTCAGTCTG	13	36
	HEK3_4K_1CTTINS_2GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATGAAAGCGTGCTCAGTCTG	13	37
	HEK3_4K_3TDEL_5GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCGATCCGTGCTCAGTCTG	13	33
	HEK3_4K_	GGCCCAGACTGAG	TGGAGGAAGCAGGGCTT	13	34

3GTOC_6GTOT	CACGTGA	CCTTTCCTCTGACATGACGTGCTCAGTCTG		
RNF2_4K_2AAINS_3-4GADEL	GTCATCTTAGTCATTACCTG	AACGAACACCATTGGTAATGACTAAGATG	15	14
RNF2_4K_1AINS_5GTOC	GTCATCTTAGTCATTACCTG	AACGAACACGTCAGTGTAATGACTAAGATG	15	15
RNF2_4K_1-2CTDEL_6GTOT	GTCATCTTAGTCATTACCTG	AACGAACAACTCGTAATGACTAAGATG	15	12
RNF2_4K_1CTOA_5GTOT	GTCATCTTAGTCATTACCTG	AACGAACACATCATGTAATGACTAAGATG	15	14
FANCF_4K_1TINS_3-4TGDEL	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCGGTAGCTGCAGAAGGGAT	14	16
FANCF_4K_1TINS_6GTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATTCAGGTAGCTGCAGAAGGGAT	14	18
FANCF_4K_2CDEL_5GTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGTGCTGCAGAAGGGAT	14	16

[2185]

[2186] 图3E:图41A至41K切口产生性sgRNA

切口产生性SGRNA	间隔区序列(SEQ ID NOS: 3456-3463)
HEK3_4A_+90	GTCAACCAGTATCCCGGTGC
HEK3_4B_+90	GTCAACCAGTATCCCGGTGC
RNF2_4C_+41	GTCAACCATTAAGCAAAACAT
FANCF_4D_+48	GGGGTCCCAGGTGCTGACGT
EMX1_4E_+53	GACATCGATGTCCTCCCCAT
RUNX1_4F_+38	GATGAAGCACTGTGGGTACGA
VEGFA_4G_+57	GATGTACAGAGAGCCCAGGGC
DNMT1_4H_+49	GCCCTTCAGCTAAAATAAAGG

[2187]

[2188] 表3F:图42A至42H PEgRNA

PEGRNA	间隔区序列(SEQ ID NOS: 3464-3478)	3′延伸(SEQ ID NOS: 3479-3493)	PBS长度(NT)	RT模板长度(NT)
HEK3_5A_C3	GGCCCAGACTGAGCACGTGA	TCTGTCATCACGTGCTCAGTCTG	13	10
HEK3_5A_C4	GGCCCAGACTGAGCACGTGA	TCTGCTATCACGTGCTCAGTCTG	13	10
HEK3_5A_C7	GGCCCAGACTGAGCACGTGA	TCTGCCATTACGTGCTCAGTCTG	13	10
FANCF_5A_C3	GGAATCCCTTCTGCAGCACC	GGAAAAGTGATCCAGGTGCTGCAGAAGGGAT	14	17
FANCF_5A_C7	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATTCAGGTGCTGCAGAAGGGAT	14	17
FANCF_5	GGAATCCCTTCT	GGAAAAGCGATCTAGGTGC	14	17

[2189]

A_C8	GCAGCACC	TGCAGAAGGGAT		
EMX1_5A_C5	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTCCTTCTTCTTCTGCTCGGA	14	16
EMX1_5A_C6	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCTCTTCTTCTTCTGCTCGGA	14	16
EMX1_5A_C7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCTTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C5_6	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTCTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C5_7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTCTTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C6_7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCTTTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C5_6_7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTTTTCTTCTTCTGCTCGGA	14	16
HEK3_5D_A5	GGCCCAGACTGAGCACGTGA	TCTGCCGTCACGTGCTCAGTCTG	13	10
HEK3_5D_A8	GGCCCAGACTGAGCACGTGA	TCTGCCATCGCGTGCTCAGTCTG	13	10

[2190]

[2191] 表3G:图42A至42H切口产生性sgRNA

[2192]

切口产生性SGRNA	可能的间隔区序列(SEQ ID NOS: 616-618)	可能的间隔区序列(SEQ ID NO: 619-621)
HEK3_5A-F_+90	GTCAACCAGTATCCCGGTGC	GTCAACCAGTATCCCGGTGC
FANCF_5A-F_+48	GGGGTCCCAGGTGCTGACGT	GATGTACAGAGAGCCCAGGGC
EMX1_5A-F_+57	GATGTACAGAGAGCCCAGGGC	GGGGTCCCAGGTGCTGACGT

[2193] 表3H:图42A至42H碱基编辑sgRNA

[2194]

碱基编辑SGRNA	间隔区序列SEQUENCE
HEK3_5A-F_BE	GTGCCATCACGTGCTCAGTCT(SEQ ID NO:455)
FANCF_5A-F_BE	GAGCGATCCAGGTGCTGCAGA(SEQ ID NO:456)
EMX1_5A-F_BE	GGAGCCCTTCTTCTTCTGCT(SEQ ID NO:457)

[2195] 表3I:图42A至42H中靶sgRNA

[2196]

中靶SGRNA	间隔区序列
HEK3_5G	GGCCCAGACTGAGCACGTGA(SEQ ID NO:510)
HEK4_5G	GGCACTGCGGCTGGAGGTGG(SEQ ID NO:511)
EMX1_5G	GAGTCCGAGCAGAAGAAGAA(SEQ ID NO:512)
FANCF_5G	GGAATCCCTTCTGCAGCACC(SEQ ID NO:513)

[2197] 表3J:图42A至42H中靶PEgRNA

[2198]

中靶 PEGRNA	间隔区序列(SEQ ID NO: 663-677)	3′延伸(SEQ ID NO: 678-692)	PBS 长度 (NT)	RT 模板长度 (NT)
HEK3_5G-H_PEGRNA_1	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_5G-H_PEGRNA_2	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	13
HEK3_5G-H_PEGRNA_3	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	9
HEK3_5G-H_PEGRNA_4	GGCCCAGACTGAGCACGTGA	TCTGCGATCACGTGCTCAGTCTG	13	10
HEK4_5G-H_PEGRNA_1	GGCACTGCGGCTGGAGGTGG	TTAACGCCCACCTCCAGCC	9	10
HEK4_5G-H_PEGRNA_2	GGCACTGCGGCTGGAGGTGG	TTAACCCCCCCCTCCAGCC	9	10
HEK4_5G-H_PEGRNA_3	GGCACTGCGGCTGGAGGTGG	TTAACCCCTTACACCTCCAGCC	9	13
HEK4_5G-H_PEGRNA_4	GGCACTGCGGCTGGAGGTGG	TTAACCCCCCCTCCAGCC	9	9
EMX1_5G-H_PEGRNA_1	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCACTTCTTCTTCTGCTCGGA	14	16
EMX1_5G-H_PEGRNA_2	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTGCTTCTTCTGCTCGGA	14	16
EMX1_5G-H_PEGRNA_3	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTCGCATTCTTCTGCTCGGA	14	19
EMX1_5G-H_PEGRNA_4	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTCTTCTTCTGCTCGGA	14	13
FANCF_5G-H_PEGRNA_1	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATGCAGGTGCTGCAGAAGGGAT	14	17
FANCF_5G-H_PEGRNA_2	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGCGCTGCAGAAGGGAT	14	17
FANCF_5G-H_PEGRNA_3	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAATCGGTGCTGCAGAAGGGAT	14	20

[2199]

[2200] 表3K:图49A至49B PEgRNA

[2201]

PEGRNA	间隔区序列(SEQ ID NO: 3494-3521)	3′延伸(SEQ ID NO: 3522-3540)	PBS 长度 (NT)	RT 模板长度 (NT)
HEK3_6A_2GTOC	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK3_6A_2GTOC	GGCCCAGACTGAGCACGTGA			
EMX1_6A_3GTOC	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCCCTTGTTCTTCTGCTCGG	13	13
EMX1_6A_3GTOC	GAGTCCGAGCAGAAGAAGAA			
FANCF_6A_5GTOT	GGAATCCCTTCTGCAGCACC	AAAAGCGATCAAGGTGCTGCAGAAGGGA	13	15
FANCF_6A_5GTOT	GGAATCCCTTCTGCAGCACC			
HEK3_6A_1HIS6INS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCAATGATGGTGATGATGGTGCGTGCTCAGTCTG	13	52
HEK3_6A_1HIS6INS	GGCCCAGACTGAGCACGTGA			
HEK3_6A_5GTOT	GGCCCAGACTGAGCACGTGA	TCTGCAATCACGTGCTCAGTCTG	13	10
HEK3_6A_5GTOT	GGCCCAGACTGAGCACGTGA			
HEK3_6A_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
HEK3_6A_1C	GGCCCAGACTG			

	TTINS	AGCACGTGA			
	*HBB*_6B_INSALL	GCATGGTGCACCTGACTCCTG	AGACTTCTCCACAGGAGTCAGGTGCAC	13	14
	*HBB*_6B_INSALL	GCATGGTGCACCTGACTCCTG			
	*HBB*_6B_CORRECT	GCATGGTGCACCTGACTCCTG	AGACTTCTCCTCAGGAGTCAGGTGCAC	13	14
	*HBB*_6B_CORRECT	GCATGGTGCACCTGACTCCTG			
	*HBB*_6B_CORRECT_W_SILENT	GCATGGTGCACCTGACTCCTG	AGACTTCTCTTCAGGAGTCAGGTGCAC	13	14
	*HBB*_6B_CORRECT_W_SILENT	GCATGGTGCACCTGACTCCTG			
	*HEXA*_6B_INSTALL	GTACCTGAACCGTATATCCTA	AGTCAGGGCCATAGGATAGATATACGGTTC	12	14
	*HEXA*_6B_CORRECT	GATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTG	10	21
[2202]	*HEXA*_6B_CORRECT_W_SILENT	GATCCTTCCAGTCAGGGCCAT	GTACCTGAACCGTATATCTTATGGCCCTGACT	9	27
	*PRNP*_6C	GCAGTGGTGGGGGGCCTTGG	ATGTAGACGCCAAGGCCCCCCACC	12	12
	HEK3_6E-G_1TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCATCCCGTGCTCAGTCTG	13	10
	HEK3_6E-G_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
	RNF2_6E-G_1CTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCACGTAATGACTAAGATG	15	14
	*HBB*_6E-G_4ATOT	GCATGGTGCACCTGACTCCTG	AGACTTCTCCACAGGAGTCAGGTGCAC	13	14
	HEK3_6H_1HIS6INS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCAATGATGGTGATGATGGTGCGTGCTCAGTCTG	13	52
	HEK3_6H_1FLAGINS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACTTATCGTCGTCATCCTTGTAATCCGTGCTCAGTCTG	13	58

[2203] 表3L:图47A至74D PEgRNA

[2204]

PEGRNA	间隔区序列(SEQ ID NO: 3541-3547)	3′延伸序列(SEQ ID NO: 3549-3556)	PBS长度(NT)	RT 模板长度(NT)
HEK3_ED4B_1TDEL	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	9
HEK3_ED4B_1AINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCATCGTGCTCAGTCTG	13	11
HEK3_ED4B_1	GGCCCAGACTGAGC	TCTGCCATCAAAGCGTGC	13	13

CTTINS	ACGTGA	TCAGTCTG		
HEK3_ED4C_2GTOC	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK3_ED4D_1FLAGINS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACTTATCGTCGTCATCCTTGTAATCCGTGCTCAGTCTG	13	58
RNF2_ED4E_1CTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
EMX1_ED4F_1GTOC	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCCCTTGTTCTTCTGCTCGG	13	13
*HBB*_ED4G_2TTOA	GTAACGGCAGACTTCTCCTC	ATCTGACTCCTGTGGAGAAGTCTGCC	12	14

[2205]

[2206] 表3M:图48A至48C PEgRNA

PEGRNA	间隔区序列(SEQ ID NO: 3557-3627)	3′延伸序列(SEQ ID NO: 3628-3698)	PBS 长度(NT)	RT 模板长度(NT)
VEGFA_ED5A_31	GATGTCTGCAGGCCAGATGA	CCCTCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	31
VEGFA_ED5A_30	GATGTCTGCAGGCCAGATGA	CCTCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	30
VEGFA_ED5A_29	GATGTCTGCAGGCCAGATGA	CTCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	29
VEGFA_ED5A_28	GATGTCTGCAGGCCAGATGA	TCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	28
VEGFA_ED5A_27	GATGTCTGCAGGCCAGATGA	CTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	27
VEGFA_ED5A_26	GATGTCTGCAGGCCAGATGA	TGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	26
VEGFA_ED5A_25	GATGTCTGCAGGCCAGATGA	GACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	25
VEGFA_ED5A_24	GATGTCTGCAGGCCAGATGA	ACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	24
VEGFA_ED5A_23	GATGTCTGCAGGCCAGATGA	CAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	23
VEGFA_ED5A_22	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	22
VEGFA_ED5A_21	GATGTCTGCAGGCCAGATGA	ATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	21
VEGFA_ED	GATGTCTGCAGGC	TGTGCCATCTGGAGCACTCA	13	20

[2207]

[2208]

5A_20	CAGATGA	TCTGGCCTGCAGA		
VEGFA_ED 5A_19	GATGTCTGCAGGC CAGATGA	GTGCCATCTGGAGCACTCAT CTGGCCTGCAGA	13	19
VEGFA_ED 5A_18	GATGTCTGCAGGC CAGATGA	TGCCATCTGGAGCACTCATC TGGCCTGCAGA	13	18
VEGFA_ED 5A_17	GATGTCTGCAGGC CAGATGA	GCCATCTGGAGCACTCATCT GGCCTGCAGA	13	17
VEGFA_ED 5A_16	GATGTCTGCAGGC CAGATGA	CCATCTGGAGCACTCATCTG GCCTGCAGA	13	16
VEGFA_ED 5A_15	GATGTCTGCAGGC CAGATGA	CATCTGGAGCACTCATCTGG CCTGCAGA	13	15
VEGFA_ED 5A_14	GATGTCTGCAGGC CAGATGA	ATCTGGAGCACTCATCTGGC CTGCAGA	13	14
VEGFA_ED 5A_13	GATGTCTGCAGGC CAGATGA	TCTGGAGCACTCATCTGGCC TGCAGA	13	13
VEGFA_ED 5A_12	GATGTCTGCAGGC CAGATGA	CTGGAGCACTCATCTGGCCT GCAGA	13	12
VEGFA_ED 5A_11	GATGTCTGCAGGC CAGATGA	TGGAGCACTCATCTGGCCTG CAGA	13	11
VEGFA_ED 5A_10	GATGTCTGCAGGC CAGATGA	GGAGCACTCATCTGGCCTGC AGA	13	10
VEGFA_ED 5A_9	GATGTCTGCAGGC CAGATGA	GAGCACTCATCTGGCCTGCA GA	13	9
VEGFA_ED 5A_8	GATGTCTGCAGGC CAGATGA	AGCACTCATCTGGCCTGCAG A	13	8
DNMT1_ED 5B_31	GATTCCTGGTGCC AGAAACA	AGGACTAGTTCTGCCCTCCC GTCACCACTGTTTCTGGCAC CAGG	13	31
DNMT1_ED 5B_30	GATTCCTGGTGCC AGAAACA	GGACTAGTTCTGCCCTCCCG TCACCACTGTTTCTGGCACC AGG	13	30
DNMT1_ED 5B_29	GATTCCTGGTGCC AGAAACA	GACTAGTTCTGCCCTCCCGT CACCACTGTTTCTGGCACCA GG	13	29
DNMT1_ED 5B_28	GATTCCTGGTGCC AGAAACA	ACTAGTTCTGCCCTCCCGTC ACCACTGTTTCTGGCACCAG G	13	28
DNMT1_ED 5B_27	GATTCCTGGTGCC AGAAACA	CTAGTTCTGCCCTCCCGTCA CCACTGTTTCTGGCACCAGG	13	27
DNMT1_ED 5B_26	GATTCCTGGTGCC AGAAACA	TAGTTCTGCCCTCCCGTCAC CACTGTTTCTGGCACCAGG	13	26

[2209]

DNMT1_ED5B_25	GATTCCTGGTGCCAGAAACA	AGTTCTGCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	25
DNMT1_ED5B_24	GATTCCTGGTGCCAGAAACA	GTTCTGCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	24
DNMT1_ED5B_23	GATTCCTGGTGCCAGAAACA	TTCTGCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	23
DNMT1_ED5B_22	GATTCCTGGTGCCAGAAACA	TCTGCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	22
DNMT1_ED5B_21	GATTCCTGGTGCCAGAAACA	CTGCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	21
DNMT1_ED5B_20	GATTCCTGGTGCCAGAAACA	TGCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	20
DNMT1_ED5B_19	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	19
DNMT1_ED5B_18	GATTCCTGGTGCCAGAAACA	CCCTCCCGTCACCACTGTTTCTGGCACCAGG	13	18
DNMT1_ED5B_17	GATTCCTGGTGCCAGAAACA	CCTCCCGTCACCACTGTTTCTGGCACCAGG	13	17
DNMT1_ED5B_16	GATTCCTGGTGCCAGAAACA	CTCCCGTCACCACTGTTTCTGGCACCAGG	13	16
DNMT1_ED5B_15	GATTCCTGGTGCCAGAAACA	TCCCGTCACCACTGTTTCTGGCACCAGG	13	15
DNMT1_ED5B_14	GATTCCTGGTGCCAGAAACA	CCCGTCACCACTGTTTCTGGCACCAGG	13	14
DNMT1_ED5B_13	GATTCCTGGTGCCAGAAACA	CCGTCACCACTGTTTCTGGCACCAGG	13	13
DNMT1_ED5B_12	GATTCCTGGTGCCAGAAACA	CGTCACCACTGTTTCTGGCACCAGG	13	12
DNMT1_ED5B_11	GATTCCTGGTGCCAGAAACA	GTCACCACTGTTTCTGGCACCAGG	13	11
DNMT1_ED5B_10	GATTCCTGGTGCCAGAAACA	TCACCACTGTTTCTGGCACCAGG	13	10
DNMT1_ED5B_9	GATTCCTGGTGCCAGAAACA	CACCACTGTTTCTGGCACCAGG	13	9
DNMT1_ED5B_8	GATTCCTGGTGCCAGAAACA	ACCACTGTTTCTGGCACCAGG	13	8
RUNX1_ED5C_31	GCATTTTCAGGAGGAAGCGA	AATGACTCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	31
RUNX1_ED5C_30	GCATTTTCAGGAGGAAGCGA	ATGACTCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTG	15	30

[2210]

		AAAAT		
RUNX1_ED5C_29	GCATTTTCAGGAGGAAGCGA	TGACTCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	29
RUNX1_ED5C_28	GCATTTTCAGGAGGAAGCGA	GACTCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	28
RUNX1_ED5C_27	GCATTTTCAGGAGGAAGCGA	ACTCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	27
RUNX1_ED5C_26	GCATTTTCAGGAGGAAGCGA	CTCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	26
RUNX1_ED5C_25	GCATTTTCAGGAGGAAGCGA	TCAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	25
RUNX1_ED5C_24	GCATTTTCAGGAGGAAGCGA	CAAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	24
RUNX1_ED5C_23	GCATTTTCAGGAGGAAGCGA	AAATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	23
RUNX1_ED5C_22	GCATTTTCAGGAGGAAGCGA	AATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	22
RUNX1_ED5C_21	GCATTTTCAGGAGGAAGCGA	ATATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	21
RUNX1_ED5C_20	GCATTTTCAGGAGGAAGCGA	TATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	20
RUNX1_ED5C_19	GCATTTTCAGGAGGAAGCGA	ATGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	19
RUNX1_ED5C_18	GCATTTTCAGGAGGAAGCGA	TGCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	18
RUNX1_ED5C_17	GCATTTTCAGGAGGAAGCGA	GCTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	17
RUNX1_ED5C_16	GCATTTTCAGGAGGAAGCGA	CTGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	16
RUNX1_ED5C_15	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	15
RUNX1_ED5C_14	GCATTTTCAGGAGGAAGCGA	GTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	14
RUNX1_ED5C_13	GCATTTTCAGGAGGAAGCGA	TCTGAAGCAATCGCTTCCTCCTGAAAAT	15	13
RUNX1_ED5C_12	GCATTTTCAGGAGGAAGCGA	CTGAAGCAATCGCTTCCTCCTGAAAAT	15	12

[2211]

RUNX1_ED5C_11	GCATTTTCAGGAGGAAGCGA	TGAAGCAATCGCTTCCTCCTGAAAAT	15	11
RUNX1_ED5C_10	GCATTTTCAGGAGGAAGCGA	GAAGCAATCGCTTCCTCCTGAAAAT	15	10
RUNX1_ED5C_9	GCATTTTCAGGAGGAAGCGA	AAGCAATCGCTTCCTCCTGAAAAT	15	9

[2212] 表3N:图48A至48C PEgRNA

[2213]

PEGRNA	间隔区序列	3′延伸序列	PBS长度(NT)	RT模板长度(NT)
HEK3_ED6_5GTOA	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 393)	TCTGCTATCACGTGCTCAGTCTG (SEQ ID NO: 394)	13	10

[2214] 表30:图48A至48C切口产生性sgRNA

[2215]

切口产生性SGRNA	间隔区序列
HEK3_ED6_+63	GCACATACTAGCCCCTGTCT (SEQ ID NO: 395)

[2216] 表3P:图50A至50B PEgRNA

[2217]

PEGRNA	间隔区(SEQ ID NO: 699-3754)	3′延伸(5′至3′)(SEQ ID NO: 3755-3810)	PBS长度(NT)	RT模板长度(NT)
HBB 3.5	GTAACGGCAGACTTCTCCAC	AGACTTCTCCTCAGGAGTCAGGTGCAC	12	14
HBB 3.7	GCATGGTGCACCTGACTCCTG	AGACTTCTCTTCAGGAGTCAGGTGCAC	13	14
HBB 5.2	GCATGGTGCACCTGACTCCTG	TAACGGCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	19
HBB 5.3	GCATGGTGCACCTGACTCCTG	ACGGCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	17
HBB 5.4	GCATGGTGCACCTGACTCCTG	GGCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	16
HBB 5.5	GCATGGTGCACCTGACTCCTG	GCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	13
HBB 5.6	GCATGGTGCACCTGACTCCTG	GACTTCTCCTCAGGAGTCAGGTGCAC	13	12
HBB 5.7	GCATGGTGCACCTGACTCCTG	ACTTCTCCTCAGGAGTCAGGTGCAC	13	21
HBB 5.8	GCATGGTGCACCTGACTCCTG	TAACGGCAGACTTCTCCTCAGGAGTCAGGTGCA	12	19
HBB 5.9	GCATGGTGCACCTGACTCCTG	ACGGCAGACTTCTCCTCAGGAGTCAGGTGCA	12	17
HBB 5.10	GCATGGTGCACCTGACTCCTG	GGCAGACTTCTCCTCAGGAGTCAGGTGCA	12	16
HBB 5.11	GCATGGTGCACCTGACTCCTG	GCAGACTTCTCCTCAGGAGTCAGGTGCA	12	13
HBB 5.12	GCATGGTGCACCTGACTCCTG	GACTTCTCCTCAGGAGTCAGGTGCA	12	12

[2218]

HBB 5.13	GCATGGTGCACC TGACTCCTG	ACTTCTCCTCAGGAGTCAGGTGCA	12	14
HEXA S 1	ATCCTTCCAGTC AGGGCCAT	ATATCTTATGGCCCTGACTGGAA	13	14
HEXA S 2	ATCCTTCCAGTC AGGGCCAT	TATATCTTATGGCCCTGACTGGAA	13	15
HEXA S 3	ATCCTTCCAGTC AGGGCCAT	GTATATCTTATGGCCCTGACTGGAA	13	16
HEXA S 4	ATCCTTCCAGTC AGGGCCAT	ACCGTATATCTTATGGCCCTGACTG GAA	13	19
HEXA S 5	ATCCTTCCAGTC AGGGCCAT	AACCGTATATCTTATGGCCCTGACT GGAA	13	20
HEXA S 6	ATCCTTCCAGTC AGGGCCAT	GAACCGTATATCTTATGGCCCTGAC TGGAA	13	21
HEXA S 7	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTGGAA	13	22
HEXA S 8	ATCCTTCCAGTC AGGGCCAT	ATATCTTATGGCCCTGACT	9	14
HEXA S 9	ATCCTTCCAGTC AGGGCCAT	TATATCTTATGGCCCTGACT	9	15
HEXA S 10	ATCCTTCCAGTC AGGGCCAT	GTATATCTTATGGCCCTGACT	9	16
HEXA S 11	ATCCTTCCAGTC AGGGCCAT	ACCGTATATCTTATGGCCCTGACT	9	19
HEXA S 12	ATCCTTCCAGTC AGGGCCAT	AACCGTATATCTTATGGCCCTGACT	9	20
HEXA S 13	ATCCTTCCAGTC AGGGCCAT	GAACCGTATATCTTATGGCCCTGAC T	9	21
HEXA S 14	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CT	9	22
HEXA S 15	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA C	8	22
HEXA S 16	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTG	10	22
HEXA S 17	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTGG	11	22
HEXA S 18	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTGGA	12	22
HEXA S 19	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTGGAA	13	22
HEXA S 20	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTGGAAG	14	22
HEXA S 21	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCTTATGGCCCTGA CTGGAAGG	15	22
HEXA S 22	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCTTATGGCCC TGACT	9	25
HEXA S 23	ATCCTTCCAGTC AGGGCCAT	TACCTGAACCGTATATCTTATGGCCC TGACT	9	26
HEXA S 24	ATCCTTCCAGTC AGGGCCAT	GTACCTGAACCGTATATCTTATGGC CCTGACT	9	27
HEXA	ATCCTTCCAGTC	GGTACCTGAACCGTATATCTTATGG	9	28

S 25	AGGGCCAT	CCCTGACT		
HEXA S 26	ATCCTTCCAGTC AGGGCCAT	TGGTACCTGAACCGTATATCTTATG GCCCTGACT	9	29
HEXA 5	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCCTATGGCCC TGACTGGAA	13	21
HEXA 6	ATCCTTCCAGTC AGGGCCAT	ACCGTATATCCTATGGCCCTGACTG GAA	13	15
HEXA 7	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCCTATGGCCC TGACTGGAAGG	15	21
HEXA 8	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCCTATGGCCC TGACTGGAAG	14	21
HEXA 9	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCCTATGGCCC TGACTGGA	12	21
HEXA 10	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCCTATGGCCC TGACTGG	11	21
HEXA 11	ATCCTTCCAGTC AGGGCCAT	ACCTGAACCGTATATCCTATGGCCC TGACTG	10	21
HEXA 12	ATCCTTCCAGTC AGGGCCAT	AACCGTATATCCTATGGCCCTGACT GGAA	13	16
HEXA 13	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCCTATGGCCCTGA CTGGAA	13	18
HEXA 14	ATCCTTCCAGTC AGGGCCAT	TACCTGAACCGTATATCCTATGGCC CTGACTGGAA	13	22
HEXA 15	ATCCTTCCAGTC AGGGCCAT	TGGTACCTGAACCGTATATCCTATG GCCCTGACTGGAA	13	25
HEXA 16	ATCCTTCCAGTC AGGGCCAT	GTACCTGAACCGTATATCCTATGGC CCTGACTGGAA	13	23
HEXA 17	ATCCTTCCAGTC AGGGCCAT	AACCGTATATCCTATGGCCCTGACT G	10	16
HEXA 18	ATCCTTCCAGTC AGGGCCAT	TGAACCGTATATCCTATGGCCCTGA CTG	10	18
HEXA 19	ATCCTTCCAGTC AGGGCCAT	TACCTGAACCGTATATCCTATGGCC CTGACTG	10	22
HEXA 20	ATCCTTCCAGTC AGGGCCAT	TGGTACCTGAACCGTATATCCTATG GCCCTGACTG	10	25

[2219]

[2220] 表3Q:图50A至50B切口产生性sgRNA

切口产生性 SGRNA	间隔区序列
*HBB*_ED7A_+72	GCCTTGATACCAACCTGCCCA (SEQ ID NO: 626)
*HEXA*_ED7B_+60	GCTGGAACTGGTCACCAAGGC(SEQ ID NO: 627)
*HEXA*_ED7B_CORRECT_WT_PE3B	GTACCTGAACCGTATATCCTA(SEQ ID NO: 628)
*HEXA*_ED7B_CORRECT_SILENT_PE3B	GTACCTGAACCGTATATCTTA(SEQ ID NO: 629)

[2221]

[2222] 表3R:图51A至51G PEgRNA

[2223]

PEGRNA	间隔区序列 (SEQ ID NO: 632-640)	3′延伸(SEQ ID NO: 641-649)	PBS 长度 (NT)	RT 模板长度 (NT)
HEK3_ED8_1	GGCCCAGACTGAGC	TCTGCCATCCCGTGCTCA	13	10

[2224]

TTOG	ACGTGA	GTCTG		
HEK3_ED8_3ATOC	GGCCCAGACTGAGCACGTGA	TCTGCCAGCACGTGCTCAGTCTG	13	10
HEK3_ED8_3ATOT	GGCCCAGACTGAGCACGTGA	TCTGCCAACACGTGCTCAGTCTG	13	10
HEK3_ED8_3ATOT_5-6GGTOTT	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGAAAACACGTGCTCAGTCTG	13	34
HEK3_ED8_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
RNF2_ED8_1CTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
RNF2_ED8_1CTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCACGTAATGACTAAGATG	15	14
RNF2_ED8_1GTAINS	GTCATCTTAGTCATTACCTG	AACGAACACCTCAGTACGTAATGACTAAGATG	15	17
*HBB*_ED8_4ATOT	GCATGGTGCACCTGACTCCTG	AGACTTCTCCACAGGAGTCAGGTGCAC	13	14

[2225] 表4:哺乳动物细胞基因组DNA扩增和HTS中使用的引物序列[181]

[2226]

描述	序列(SEQ ID NO: 3811-3863)
HEK3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
HEK3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
RNF2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGTCTCATATGCCCCTTGG
RNF2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGTAGGAATTTTGGTGGGACA
HEK4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
HEK4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
EMX1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
EMX1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
FANCF FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
FANCF REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
HBB FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGGGTTGGCCAATCTACTCCC
HBB REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTTCTCTGTCTCCACATGCC
PRNP FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGTGGAACAAGCCGAGT
PRNP REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTACTTGGTTGGGGTAACGGTG
HEXA FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATACAGGTGTGGCGAGAGG

[2227]

HEXA REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCTCCTTTGGTTAGCA
RUNX1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCACAAACAAGACAGGGAACTG
RUNX1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGATGTAGGGCTAGAGGGGTG
VEGFA FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACTTGGTGCCAAATTCTTCTCC
VEGFA REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGAGGGAATGGGCTTTGGA
DNMT FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAACAGCTTCATGTCAGCC
DNMT REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGTTAATGTTTCCTGATGGTCC
HEK3 脱靶位点 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
HEK3 脱靶位点 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
HEK3 脱靶位点 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
HEK3 脱靶位点 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
HEK3 脱靶位点 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
HEK3 脱靶位点 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT
HEK3 脱靶位点 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
HEK3 脱靶位点 4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
HEK4 脱靶位点 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCATGGCTTCTGAGACTCA
HEK4 脱靶位点 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
HEK4 脱靶位点 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGGCAATGGAGGCATTGG
HEK4 脱靶位点 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCCCATGAGAG
HEK4 脱靶位点 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCTGAGGCTCGAATCCTG
HEK4 脱靶位点 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
HEK4 脱靶位点 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTCCACCAGAACTCAGCCC
HEK4 脱靶位点 4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCCACAACAC
EMX1 脱靶位点 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGGGGAGATTTGCATCTGTGGAGG
EMX1 脱靶位点 1	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTTATACCATC

[2228]

REV	TTGGGGTTACAG
EMX1 脱靶位点 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAATGTGCTTCAACCCATCACGGC
EMX1 脱靶位点 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCATGAATTTGTGATGGATGCAGTCTG
EMX1 脱靶位点 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGAAGGAGGTGCAGGAGCTAGAC
EMX1 脱靶位点 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCATCCCGACCTTCATCCCTCCTGG
EMX1 脱靶位点 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAGTTCTGACATTCCTCCTGAGGG
EMX1 脱靶位点 4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCAAACAAGGTGCAGATACAGCA
FANCF 脱靶位点 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGGGCAGTGGCGTCTTAGTCG
FANCF 脱靶位点 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCTGGGTTTGGTTGGCTGCTC
FANCF 脱靶位点 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCCTTGCCGCCCAGCCGGTC
FANCF 脱靶位点 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGGGGAAGAGGCGAGGACAC
FANCF 脱靶位点 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGTGTTTCCCATCCCCAACAC
FANCF 脱靶位点 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAATGGATCCCCCCCTAGAGCTC
FANCF 脱靶位点 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGGCCCACAGGTCCTTCTGGA

[2229] 表5:HDR实验和HBB E6VHEK293T细胞系创建中使用的100-mer单链DNA寡核苷酸供体模板序列。寡核苷酸长度为100-103nt,具有围绕编辑位点中心的同源臂。寡核苷酸来自IntegratedDNATechnologies,经PAGE纯化。

[2230]

HEK3 +3 A 至 T	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGTTGGCAGAGGAAAGGAAGCCCTGCTTCCTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 717)
HEK3 +3 A 至 T, +5, 6 GG 至 TT	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGTTTTCAGAGGAAAGGAAGCCCTGCTTCCTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 718)
HEK3 +1 T 至 G	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGGGATGGCAGAGGAAAGGAAGCCCTGCTTCCTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 719)
HEK3 +3 A 至 C	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGCTGGCAGAGGAAAGGAAGCCCTGCTTCCTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 720)
HEK3 +1 CTT 插入	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGGAAGCCCTGCTTCCTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 721)
RNF2 +1 C 至 A	CCCAGTTTACACGTCTCATATGCCCCTTGGCAGTCATCTTAGTCATTACATGAGGTGTTCGTTGTAACTCATATAAACTGAG

	TTCCCATGTTTTGCTTAA (SEQ ID NO: 722)
RNF2 +1 C 至 G	CCCAGTTTACACGTCTCATATGCCCCTTGGCAGTCATCTTA GTCATTACGTGAGGTGTTCGTTGTAACTCATATAAACTGAG TTCCCATGTTTTGCTTAA (SEQ ID NO: 723)
RNF2 +1 GTA 插入	CAGTTTACACGTCTCATATGCCCCTTGGCAGTCATCTTAGT CATTACGTACTGAGGTGTTCGTTGTAACTCATATAAACTGA GTTCCCATGTTTTGCTTA (SEQ ID NO: 724)
HBB E6V 安装(又用于形成 *HBB* E6V HEK293T 细胞系)	ACTTCATCCACGTTCACCTTGCCCCACAGGGCAGTAACGG CAGACTTCTCCACAGGAGTCAGATGCACCATGGTGTCTGT TTGAGGTTGCTAGTGAACAC (SEQ ID NO: 725)
HBB E6V 校正原间隔区 A	ACTTCATCCACGTTCACCTTGCCCCACAGGGCAGTAACGG CAGACTTCTCCTCAGGAGTCAGGTGCACCATGGTGTCTGT TTGAGGTTGCTAGTGAACAC (SEQ ID NO: 726)
HBB E6V 校正原间隔区 B	GTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACCT GACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGC AAGGTGAACGTGGATGAAGT (SEQ ID NO: 727)
HBB E6V 校正原间隔区 B, 沉默 PAM 突变	GTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACCT GACTCCTGATGAGAAGTCTGCCGTTACTGCCCTGTGGGGC AAGGTGAACGTGGATGAAGT (SEQ ID NO: 728)
PRNP G127V	CACATGGCTGGTGCTGCAGCAGCTGGGGCAGTGGTGGGG GGCCTTGGCGTCTACATGCTGGGAAGTGCCATGAGCAGGC CCATCATACATTTCGGCAGTG (SEQ ID NO: 729)

[2231]

[2232] 其他序列本文使用的酵母双荧光报告质粒序列p425-GFP_终止_mCherry：

[2233]

ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGAT
GGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTAC
TTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATG
GCCAACCTTAGTCACTACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGA
TCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAG
AAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAA
GTTTGAAGGTGATACCTTAGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGA
AGATGGTAACATTTTAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTA
CATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACAC
AACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATT
GGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATCTGCCTTA
TCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACTG
CTGCTGGTATTACCCATGGTATGGATGAATTGTACAAAGCTAGCAACCTGGGTCAA
TCCTTGG**GGCCCAGACTGAGCACGTGA**TGGCAGAGCACAGGAGACGTC*ATGGT*
TTCAAAAGGTGAAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAAGTTCATA
TGGAAGGTTCAGTTAATGGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACCAT
ATGAAGGTACTCAAACTGCTAAATTGAAAGTTACTAAAGGTGGTCCATTACCATTTGCTTG
GGATATTTTGTCACCACAATTTATGTATGGTTCAAAAGCTTATGTTAAACATCCAGCTGATA
TTCCAGATTATTTAAAATTGTCATTTCCAGAAGGTTTTAAATGGGAAAGAGTTATGAATTTT
GAAGATGGTGGTGTTGTTACTGTTACTCAAGATTCATCATTACAAGATGGTGAATTTATTT
ATAAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTCCAGTTATGCAAAAAAAAAC
TATGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGTGCTTTAAAAGGTG
AAATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGTTAAAACTAC
TTATAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATTAAATTGGATAT

TACTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTGAAGGTAGACATT
*CAACTGGTGGTATGGATGAATTATATAAAGGTACCGCTCGAGCAGCTGTGATTGATTGA*G
TCGACTTGGTTGAACACGTTGCCAAGGCTTAAGTGAATTTACTTTAAATCTTGCAT
TTAAATAAATTTTCTTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAAT
GACATTTTCGATTCGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTC
CAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATA
GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAGGAGCCG
GAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGGTAACTCACATTAATT
GCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTA
ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCT
TCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG
CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAA
GAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCC
TGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTAT
CTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTA
AGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA
GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACT
AGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
GAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT
GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG
[2234] GTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTA
ATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACC
AGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTT
GCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTT
GGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGT
GCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAA
AACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA
TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC
ACATTTCCCCGAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTTGTAGAAC
AAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGCATTTTT
ACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACGAAGAATCTGTGCTT
CATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAAGAAT
CTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACCAAC
AAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATTTTT
CTAACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTC
TTGATAACTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTTTGGTGT

[2235]

CTATTTTCTCTTCCATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTACTA
GCGAAGCTGCGGGTGCATTTTTTCAAGATAAAGGCATCCCCGATTATATTCTATACC
GATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGCGTTGATGATTCTTCATT
GGTCAGAAAATTATGAACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAGGAAA
TGTTTACATTTTCGTATTGTTTTCGATTCACTCTATGAATAGTTCTTACTACAATTTTT
TTGTCTAAAGAGTAATACTAGAGATAAACATAAAAAATGTAGAGGTCGAGTTTAGA
TGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATAGGGATATAGCACAG
AGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTCGCAATA
TTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTCTTCAG
AGCGCTTTTGGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGA
ACTTCGGAATAGGAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAATGCA
ACGCGAGCTGCGCACATACAGCTCACTGTTCACGTCGCACCTATATCTGCGTGTTG
CCTGTATATATATATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAATGCGT
ACTTATATGCGTCTATTTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTA
TCCCATTCCATGCGGGGTATCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTAT
ATGCTGCCACTCCTCAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCCTTTGAT
ATTGGATCATACTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTA
TCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACA
CATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGA
CAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACT
ATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATCGACTACGTCGTAAGG
CCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCTT
CAAGAAGGTATTGACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATT
TGTTGTATTTTTTTTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCGTAGTT
TCATGATTTTCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAATATTA
ATGTTAAAGTGCAATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTA
CCTGTATTCCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGT
ATATAGTTTCGTCTACCCTATGAACATATTCCATTTTGTAATTTCGTGTCGTTTCTATT
ATGAATTTCATTTATAAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTTTTTAA
GCAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTACTGTT
GGAACCACCTAAATCACCAGTTCTGATACCTGCATCCAAAACCTTTTTAACTGCAT
CTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTTCAACATCATTGCA
GCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAGCAG
AACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGC
CAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATC
GGAGATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGGGTTGG
GTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCAATGT
AGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGC
CAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGG
CCATGAAAGCGGCCATTCTTGTGATTCTTTGCACTTCTGGAACGGTGTATTGTTCA
CTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACCAAAGTAAATACCT
CCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCAAATTGTGGCTTGAT
TGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTGGCG
TACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAGGTCTAACACTACCG
GTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAACCTTCTTGG
AGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGCACC
ACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAATAG
CTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACGTGGTCACCTGGC
AAAACGACGATCTTCTTAGGGGCAGACATAGGGGCAGACATTAGAATGGTATATCC
TTGAAATATATATATATATTGCTGAAATGTAAAAGGTAAGAAAAGTTAGAAAGTAAG
ACGATTGCTAACCACCTATTGGAAAAAACAATAGGTCCTTAAATAATATTGTCAACT
TCAAGTATTGTGATGCAAGCATTTAGTCATGAACGCTTCTCTATTCTATATGAAAAG

CCGGTTCCGGCCTCTCACCTTTCCTTTTTCTCCCAATTTTTCAGTTGAAAAAGGTAT
ATGCGTCAGGCGACCTCTGAAATTAACAAAAAATTTCCAGTCATCGAATTTGATTC
TGTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGAGGAAAAAAATAATGGTTGC
TAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTCCCACAGTTAACTGCG
GTCAAGATATTTCTTGAATCAGGCGCCTTAGACCGCTCGGCCAAACAACCAATTAC
TTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTTGAACACACATGA
ACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGATTTTGATGTAATTGT
TGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACTAGAAGTTCT
CCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA
CCGCATCAGGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTT
AAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAA
AAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACT
ATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAA
AGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAG
CCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAG
GGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTT
AATGCGCCGCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCGCAACTGTT
[2236] GGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGG
ATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGT
TGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGG
GTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTG
CAGCCCGGGGGATCCGTTAGAATCATTTTGAATAAAAAACACGCTTTTTCAGTTCG
AGTTTATCATTATCAATACTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGA
TTTTCCTAACTTTATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACA
TGCCCAAAATAGGGGGCGGGTTACACAGAATATATAACATCGTAGGTGTCTGGGTG
AACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCGCTTTTTAAGCTGGCAT
CCAGAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAG
TTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAA
AAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACA
CAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTATTACCTTC
TGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAA
ATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCT
GTAAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTT
TTAAAACACCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAGAATTC
(SEQ ID NO: 730)

[2237] p425-GFP_+1fs_mCherry:

ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGAT
GGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTAC
TTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATG
GCCAACCTTAGTCACTACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGA
TCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAG
AAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAA
[2238] GTTTGAAGGTGATACCTTAGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGA
AGATGGTAACATTTTAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTA
CATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACAC
AACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATT
GGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATCTGCCTTA
TCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACTG
CTGCTGGTATTACCCATGGTATGGATGAATTGTACAAAGCTAGCAAACCTGGGTCA

[2239]

ATCCTTGG**GGCCCAGACTGAGCACGTGATGG**CAGAGCACAGGAGACGTC*ATGG*
TTTCAAAAGGTGAAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAAGTTCAT
ATGGAAGGTTCAGTTAATGGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACC
ATATGAAGGTACTCAAACTGCTAAATTGAAAGTTACTAAAGGTGGTCCATTACCATTTGCT
TGGGATATTTTGTCACCACAATTTATGTATGGTTCAAAAGCTTATGTTAAACATCCAGCTG
ATATTCCAGATTATTTAAAATTGTCATTTCCAGAAGGTTTTAAATGGGAAAGAGTTATGAAT
TTTGAAGATGGTGGTGTTGTTACTGTTACTCAAGATTCATCATTACAAGATGGTGAATTTA
TTTATAAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTCCAGTTATGCAAAAAAAA
ACTATGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGTGCTTTAAAAGGT
GAAATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGTTAAAACTA
CTTATAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATTAAATTGGATA
TTACTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTGAAGGTAGACAT
TCAACTGGTGGTATGGATGAATTATATAAAGGTACCGCTCGAGCAGCTGTGATTGATTGA
GTCGACTTGGTTGAACACGTTGCCAAGGCTTAAGTGAATTTACTTTAAATCTTGCA
TTTAAATAAATTTTCTTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAA
TGACATTTTCGATTCGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCT
CCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCAT
AGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAGGAGCC
GGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGGTAACTCACATTAA
TTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCG
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA
GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACG
CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCC
CCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT
GTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGT
ATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
CGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC
GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
ACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTT
TTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATT
TTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC
TTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCC
TGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAA
ACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTC
CATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTT
GGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC
CATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTA
AGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTG
TCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT
ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGG
GGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACT
CGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC

AAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTG
TCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC
GCGCACATTTCCCCGAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTTGTA
GAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGCA
TTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACGAAGAATCTG
TGCTTCATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAA
GAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACC
AACAAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATT
TTTCTAACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGT
CTCTTGATAACTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTTTGG
TGTCTATTTTCTCTTCCATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTA
CTAGCGAAGCTGCGGGTGCATTTTTTCAAGATAAAGGCATCCCCGATTATATTCTAT
ACCGATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGCGTTGATGATTCTT
CATTGGTCAGAAAATTATGAACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAG
GAAATGTTTACATTTTCGTATTGTTTTCGATTCACTCTATGAATAGTTCTTACTACAA
TTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAAAAAATGTAGAGGTCGAGT
TTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATAGGGATATAGC
ACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTCG
CAATATTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTC
TTCAGAGCGCTTTTGGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAAT
AGGAACTTCGGAATAGGAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAA
TGCAACGCGAGCTGCGCACATACAGCTCACTGTTCACGTCGCACCTATATCTGCGT
GTTGCCTGTATATATATATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAAT
GCGTACTTATATGCGTCTATTTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGAT
[2240] ATTATCCCATTCCATGCGGGGTATCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTT
CTATATGCTGCCACTCCTCAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCCTT
TGATATTGGATCATACTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG
CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTG
ACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGC
AGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTT
AACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATCGACTACGTCGT
AAGGCCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAAT
GCTTCAAGAAGGTATTGACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTT
TTATTTGTTGTATTTTTTTTTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCG
TAGTTTCATGATTTTCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAA
TATTAATGTTAAAGTGCAATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTT
GCTTACCTGTATTCCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGAT
TGCGTATATAGTTTCGTCTACCCTATGAACATATTCCATTTTGTAATTTCGTGTCGTTT
CTATTATGAATTTCATTTATAAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTT
TTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTA
CTGTTGGAACCACCTAAATCACCAGTTCTGATACCTGCATCCAAAACCTTTTTAAC
TGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTTCAACATCAT
TGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAG
CAGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGA
GGCCAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTC
ATCGGAGATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGGGT
TGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCA
ATGTAGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAG
AGGCCAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGT
AGGGCCATGAAAGCGGCCATTCTTGTGATTCTTTGCACTTCTGGAACGGTGTATTG
TTCACTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACCAAAGTAAAT

ACCTCCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCAAATTGTGGCT
TGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTG
GCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAGGTCTAACACTA
CCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAACCTTCT
TGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGC
ACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAA
TAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACGTGGTCACCT
GGCAAAACGACGATCTTCTTAGGGGCAGACATAGGGGCAGACATTAGAATGGTATA
TCCTTGAAATATATATATATATTGCTGAAATGTAAAAGGTAAGAAAAGTTAGAAAGT
AAGACGATTGCTAACCACCTATTGGAAAAAACAATAGGTCCTTAAATAATATTGTC
AACTTCAAGTATTGTGATGCAAGCATTTAGTCATGAACGCTTCTCTATTCTATATGA
AAAGCCGGTTCCGGCCTCTCACCTTTCCTTTTTCTCCCAATTTTTCAGTTGAAAAA
GGTATATGCGTCAGGCGACCTCTGAAATTAACAAAAAATTTCCAGTCATCGAATTT
GATTCTGTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGAGGAAAAAAATAATG
GTTGCTAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTCCCACAGTTAA
CTGCGGTCAAGATATTTCTTGAATCAGGCGCCTTAGACCGCTCGGCCAAACAACCA
ATTACTTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTTGAACACAC
ATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGATTTTGATGTAA
TTGTTGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACTAGAAG
TTCTCCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAA
AATACCGCATCAGGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTT
TTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAA
ATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGT

[2241] CCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGG
GCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGC
CGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGG
GAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGG
CGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCC
GCGCTTAATGCGCCGCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCGCA
ACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAA
AGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCA
CGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCG
AATTGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAA
TTCCTGCAGCCCGGGGGATCCGTTAGAATCATTTTGAATAAAAAACACGCTTTTTC
AGTTCGAGTTTATCATTATCAATACTGCCATTTCAAAGAATACGTAAATAATTAATAG
TAGTGATTTTCCTAACTTTATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACC
CGTACATGCCCAAAATAGGGGGCGGGTTACACAGAATATATAACATCGTAGGTGTC
TGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCGCTTTTTAAGC
TGGCATCCAGAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAAC
CATCAGTTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAAC
AGGCAAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAAT
GATGACACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTA
TTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGT
TCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATT
GTAATTCTGTAAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTT
TTTTAGTTTTAAAACACCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAA
AGAATTC (SEQ ID NO: 731)

[2242] p425-GFP_-1fs_mCherry:

[2243] ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGAT
GGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTAC

TTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATG
GCCAACCTTAGTCACTACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGA
TCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAG
AAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAA
GTTTGAAGGTGATACCTTAGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGA
AGATGGTAACATTTTAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTA
CATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACAC
AACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATT
GGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATCTGCCTTA
TCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACTG
CTGCTGGTATTACCCATGGTATGGATGAATTGTACAAAGCTAGCAAACCTGGGTCA
ATCCTTGG**<u>GGCCCAGACTGAGCACGTGA</u>ATGG**CAGAGCACAGGACGTC*ATGGTTT*
CAAAAGGTGAAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAAGTTCATATG
GAAGGTTCAGTTAATGGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACCATAT
GAAGGTACTCAAACTGCTAAATTGAAAGTTACTAAAGGTGGTCCATTACCATTTGCTTGG
GATATTTTGTCACCACAATTTATGTATGGTTCAAAAGCTTATGTTAAACATCCAGCTGATAT
TCCAGATTATTTAAAATTGTCATTTCCAGAAGGTTTTAAATGGGAAAGAGTTATGAATTTTG
AAGATGGTGGTGTTGTTACTGTTACTCAAGATTCATCATTACAAGATGGTGAATTTATTTAT
AAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTCCAGTTATGCAAAAAAAAACTA
TGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGTGCTTTAAAAGGTGAA
ATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGTTAAAACTACTTA
TAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATTAAATTGGATATTAC
TTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTGAAGGTAGACATTCA
*ACTGGTGGTATGGATGAATTATATAAAGGTACCGCTCGAGCAGCTGTGATTGATTGA*GTC
GACTTGGTTGAACACGTTGCCAAGGCTTAAGTGAATTTACTTTAAATCTTGCATTTA
[2244] AATAAATTTTCTTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAATGA
CATTTTCGATTCGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCA
GCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGC
TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAGGAGCCGGA
AGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGGTAACTCACATTAATTGC
GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTC
CTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCT
CACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGA
ACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC
TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCA
AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
GCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTT
CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA
GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT
TTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATC
TTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG
TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT
TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAA
TCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCT
GCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCA

[2245]

GCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATC
CAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTT
GCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTT
GGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGT
GCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAA
AACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA
TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC
ACATTTCCCCGAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTTGTAGAAC
AAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGCATTTTT
ACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACGAAGAATCTGTGCTT
CATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAAGAAT
CTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACCAAC
AAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATTTTT
CTAACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTC
TTGATAACTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTTTGGTGT
CTATTTTCTCTTCCATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTACTA
GCGAAGCTGCGGGTGCATTTTTTCAAGATAAAGGCATCCCCGATTATATTCTATACC
GATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGCGTTGATGATTCTTCATT
GGTCAGAAAATTATGAACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAGGAAA
TGTTTACATTTTCGTATTGTTTTCGATTCACTCTATGAATAGTTCTTACTACAATTTTT
TTGTCTAAAGAGTAATACTAGAGATAAACATAAAAAATGTAGAGGTCGAGTTTAGA
TGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATAGGGATATAGCACAG
AGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTCGCAATA
TTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTCTTCAG
AGCGCTTTTGGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGA
ACTTCGGAATAGGAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAATGCA
ACGCGAGCTGCGCACATACAGCTCACTGTTCACGTCGCACCTATATCTGCGTGTTG
CCTGTATATATATATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAATGCGT
ACTTATATGCGTCTATTTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTA
TCCCATTCCATGCGGGGTATCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTAT
ATGCTGCCACTCCTCAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCCTTTGAT
ATTGGATCATACTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTA
TCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACA
CATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGA
CAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACT
ATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATCGACTACGTCGTAAGG
CCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCTT
CAAGAAGGTATTGACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATT
TGTTGTATTTTTTTTTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCGTAGTT
TCATGATTTTCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAATATTA
ATGTTAAAGTGCAATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTA
CCTGTATTCCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGT
ATATAGTTTCGTCTACCCTATGAACATATTCCATTTTGTAATTTCGTGTCGTTTCTATT
ATGAATTTCATTTATAAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTTTTTAA
GCAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTACTGTT

GGAACCACCTAAATCACCAGTTCTGATACCTGCATCCAAAACCTTTTTAACTGCAT
CTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTTCAACATCATTGCA
GCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAGCAG
AACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGC
CAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATC
GGAGATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGGGTTGG
GTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCAATGT
AGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGC
CAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGG
CCATGAAAGCGGCCATTCTTGTGATTCTTTGCACTTCTGGAACGGTGTATTGTTCA
CTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACCAAAGTAAATACCT
CCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCAAATTGTGGCTTGAT
TGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTGGCG
TACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAGGTCTAACACTACCG
GTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAACCTTCTTGG
AGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGCACC
ACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAATAG
CTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACGTGGTCACCTGGC
AAAACGACGATCTTCTTAGGGGCAGACATAGGGGCAGACATTAGAATGGTATATCC
TTGAAATATATATATATATTGCTGAAATGTAAAAGGTAAGAAAAGTTAGAAAGTAAG
ACGATTGCTAACCACCTATTGGAAAAAACAATAGGTCCTTAAATAATATTGTCAACT
TCAAGTATTGTGATGCAAGCATTTAGTCATGAACGCTTCTCTATTCTATATGAAAAG
CCGGTTCCGGCCTCTCACCTTTCCTTTTTCTCCCAATTTTTCAGTTGAAAAAGGTAT
ATGCGTCAGGCGACCTCTGAAATTAACAAAAAATTTCCAGTCATCGAATTTGATTC
TGTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGAGGAAAAAAATAATGGTTGC
[2246] TAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTCCCACAGTTAACTGCG
GTCAAGATATTTCTTGAATCAGGCGCCTTAGACCGCTCGGCCAAACAACCAATTAC
TTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTTGAACACACATGA
ACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGATTTTGATGTAATTGT
TGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACTAGAAGTTCT
CCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA
CCGCATCAGGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTT
AAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAA
AAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACT
ATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAA
AGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAG
CCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAG
GGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTT
AATGCGCCGCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCGCAACTGTT
GGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGG
ATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGT
TGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGG
GTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTG
CAGCCCGGGGGATCCGTTAGAATCATTTTGAATAAAAAACACGCTTTTTCAGTTCG
AGTTTATCATTATCAATACTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGA
TTTTCCTAACTTTATTTAGTCAAAAAATTAGCCTTTTAATTCTGCTGTAACCCGTACA
TGCCCAAAATAGGGGGCGGGTTACACAGAATATATAACATCGTAGGTGTCTGGGTG
AACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCGCTTTTTAAGCTGGCAT
CCAGAAAAAAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAG
TTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAA
AAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGACA

[2247] CAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTATTACCTTC TGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAA ATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCT GTAAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTT TTAAAACACCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAGAATTC (SEQ ID NO: 732)

[2248] 解释:

[2249] GFP开放阅读框

[2250] 接头包含终止密码子+1移码或-1移码

[2251] mCherry开放阅读框

[2252] 质粒骨架(包含GPD启动子,Leu2标志物和AmpR)

[2253] 原间隔区(加下划线)

[2254] PAM(粗体)

[2255] 哺乳动物引导编辑器质粒和示例PEgRNA质粒的DNA序列pCMV-PE1:

[2256] ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAA AGTCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGG GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATG GCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAA GAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTG GCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACA GCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAA GTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTG GACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACC CCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAA GAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGG CCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGC AACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGAC GACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTC TGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAA CACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAG CACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAG AAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTG ACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAA AGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGC GGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAG AGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAA CCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCT CTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACC ATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCT TCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCC CAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTG AAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAA AAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAG CTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCG

[2257]

GCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATA
TCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAA
AACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGA
TACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAG
CAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAA
ACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAA
AGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCC
GGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAG
CTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCA
GAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAG
CGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATG
GGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGA
TGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAG
GTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAA
GAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTG
ATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGC
GAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATC
ACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAG
AATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT
CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCAC
CACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGT
ACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCG
GAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTC
TTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA
GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTG
GGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTG
AATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
TGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTA
AGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGC
CAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGG
GGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA
AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTA
CTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTAC
CTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAG
CGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCC
GCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCC
ACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACC
ACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTG
ATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGG
GAGGTGAC*TCTGGAGGATCTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAA*
*GCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGCAGC***ACCCT**
AAATATAGAAGATGAGTATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTT
CTCTAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGG
GGGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCA
ACCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACT
GGGGATCAAGCCCCACATACAGAGACTGTTGGACCAGGGAATACTGGTACCC
TGCCAGTCCCCCTGGAACACGCCCCTGCTACCCGTTAAGAAACCAGGGACTA
ATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGCGGGTGGAAGA

[2258] **CATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCACCG
TCCCACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAG
ACTCCACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAG
ATGGGAATCTCAGGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAA
ACAGTCCCACCCTGTTTGATGAGGCACTGCACAGAGACCTAGCAGACTTCCG
GATCCAGCACCCAGACTTGATCCTGCTACAGTACGTGGATGACTTACTGCTG
GCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAA
CCCTAGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCA
GAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG
ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCCT
CGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGATCC
CTGGGTTTGCAGAAATGGCAGCCCCCCTGTACCCTCTCACCAAAACGGGGAC
TCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAATCAAGCAA
GCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTAAGCCCTTTG
AACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAA
ACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAAAAAGCTAGACCC
AGTAGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGCAGCCATTGCCGTA
CTGACAAAGGATGCAGGCAAGCTAACCATGGGACAGCCACTAGTCATTCTGG
CCCCCCATGCAGTAGAGGCACTAGTCAAACAACCCCCCGACCGCTGGCTTTC
CAACGCCCGGATGACTCACTATCAGGCCTTGCTTTTGGACACGGACCGGGTC
CAGTTCGGACCGGTGGTAGCCCTGAACCCGGCTACGCTGCTCCCACTGCCT
GAGGAAGGGCTGCAACACAACTGCCTTGATATCCTGGCCGAAGCCCACGGA
ACCCGACCCGACCTAACGGACCAGCCGCTCCCAGACGCCGACCACACCTGG
TACACGGATGGAAGCAGTCTCTTACAAGAGGGACAGCGTAAGGCGGGAGCT
GCGGTGACCACCGAGACCGAGGTAATCTGGGCTAAAGCCCTGCCAGCCGGG
ACATCCGCTCAGCGGGCTGAACTGATAGCACTCACCCAGGCCCTAAAGATGG
CAGAAGGTAAGAAGCTAAATGTTTATACTGATAGCCGTTATGCTTTTGCTACT
GCCCATATCCATGGAGAAATATACAGAAGGCGTGGGTTGCTCACATCAGAAG
GCAAAGAGATCAAAAATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTT
TCTGCCCAAAAGACTTAGCATAATCCATTGTCCAGGACATCAAAAGGGACAC
AGCGCCGAGGCTAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCA
GCCATCACAGAGACTCCAGACACCTCTACCCTCCTCATAGAAAATTCATCACC
CTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAA
GAAGAGGAAAGTCTAA**CCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCT
GATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCC
GTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA
GAAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGG
GGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTC
GACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG
TTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT
AGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
GGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG
CGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATAC
GGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC**

TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT
GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA
GAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCC
GGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA
CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGT
ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATC
TCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA
TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCC
GGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGG
TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGC
TTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCG
ACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGAT
[2259] CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA
TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTC
CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATAT
TTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAA
GTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGA
CTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTT
GTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCA
AGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGC
TTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATA
GTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATA
ACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACG
TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT
GCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATG
CCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGG
TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTT
TTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTG
ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTT
TAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGG
GAGAGCCGCCACC (SEQ ID NO: 733)

[2260] pCMV-PE2:

[2261] ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCGACAAGAAGTA
CAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAG
AAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCG
GCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTA
TCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTG
GTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGT
ACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGC

CCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGAC
AAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACG
CCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAA
GAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTG
GCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCG
ACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAA
CACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTG
CTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACG
CCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG
CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATC
CCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACA
ACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAG
ATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCT
TCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACA
GCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCC
CGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCA
ACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGA
GGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACC
TATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCC
GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGC
CAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCC
GGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGA
CAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAG
AGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAG
CTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACC
TGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCTAT
CGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAG
AGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGA
TTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCAT
CAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAG
TACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGA
AGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGT
GGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTG
CGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAAC
CGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATC
GTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGA
TCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGT

GGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA
AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCA
TCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCA
GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAG
GGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCA
CCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCC
GCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGA
TCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCTGGAGGATCTAG
CGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGC
GGCAGCAGCACCCTAAATATAGAAGATGAGTATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTCTAGGGT
CCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCCTCT
GATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGG
ATCAAGCCCCACATACAGAGACTGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGC
TACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGCGGGTGGAAGA
CATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTT
GATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATC
CAGAGATGGGAATCTCAGGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTAA
TGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGATCCTGCTACAGTACGTGGATGAC
TTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCTAGGGAACCTCG
GGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGG
TCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCCTCGACAACTAAGG
GAGTTCCTAGGGAAGGCAGGCTTCTGTCGCCTCTTCATCCCTGGGTTTGCAGAAATGGCAGCCCCCCTGTACCCTC
TCACCAAACCGGGGACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAATCAAGCAAGCTCTTCT
AACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTAAGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCC
AAAGGTGTCCTAACGCAAAAACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAAAAAGCTAGACCCAGTAG
CAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGCAGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCAT
GGGACAGCCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTCAAACAACCCCCCGACCGCTGGCTTTCC
AACGCCCGGATGACTCACTATCAGGCCTTGCTTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGA
ACCCGGCTACGCTGCTCCCACTGCCTGAGGAAGGGCTGCAACACAACTGCCTTGATATCCTGGCCGAAGCCCACGG
AACCCGACCCGACCTAACGGACCAGCCGCTCCCAGACGCCGACCACACCTGGTACACGGATGGAAGCAGTCTCTTA
CAAGAGGGACAGCGTAAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGTAATCTGGGCTAAAGCCCTGCCAGCCG
GGACATCCGCTCAGCGGGCTGAACTGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAATGT
TTATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGAAATATACAGAAGGCGTGGGTGGCTCACA
TCAGAAGGCAAAGAGATCAAAAATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTTCTGCCCAAAAGACTTA
GCATAATCCATTGTCCAGGACATCAAAAGGGACACAGCGCCGAGGCTAGAGGCAACCGGATGGCTGACCAAGCGGC
CCGAAAGGCAGCCATCACAGAGACTCCAGACACCTCTACCCTCCTCATAGAAAATTCATCACCCTCTGGCGGCTCA
AAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCA
TTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG

CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGAAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCA
TGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGC
TAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACT
GCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTG
CGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG
CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCA
GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA
CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTAT
CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTG
GTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAA
AGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA
CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTC
ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAG
CGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTC
GTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGG
CAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTC
ATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGC
AGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGAT
CCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC
AAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATA
AACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGA
TCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTG
TTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGA
ATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGAC
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAA
CGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGT
GTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATG

ACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG
TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGT
AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCG
CTAATACGACTCACTATAGGGAGAGCCGCCACC(SEQ ID NO:734)

[2262] N-端NLS+Cas9H840A

[2263] 柔性接头

[2264] M-MLV逆转录酶+C-端NLS

[2265] 质粒骨架(包含CMV启动子和AmpR)

[2266] pU6-HEK3_PEgRNA_CTTins:

[2267] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
GAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATC
TTGTGGAAAGGACGAAACACC*GGCCCAGACTGAGCACGTGA*GTTTTAGAGCTAGA
AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAG
TCGGTCC**TCTGCCATCAAAGCGTGCTCAGTCTG**TTTTTTTAAGCTTGGGCCGCT
CGAGGTACCTCTCTACATATGACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAG
CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTG
CCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAT
AGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
CCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCC
AGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAA
CTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCT
TTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCG
TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT
ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTA
GTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG

AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGA
TCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTAC
TCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGC
GTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
[2268] TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAG
GGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCAT
TTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAA
ACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGCTAGC
TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCA
AGGTCGGGCAGGAA (SEQ ID NO: 735)

[2269] U6启动子序列

[2270] 间隔区序列

[2271] sgRNA支架

[2272] 3′延伸(包含PBS和RT模板)

[2273] 骨架(包含AmpR)

[2274] pLenti-hSyn-N-PE2-NpuN-P2A-GFP-KASH_U6-DNMT1-PEgRNA:

GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGC
TGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAAT
TGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGC
CAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
TCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGC
CCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGT
TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC
GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCT
ATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTA
TGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT
TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC
GGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAG
[2275] GCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTTTGCCTGTACTGGGTCTCTC
TGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGC
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTG
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTC
TAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCT
CTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCG
GCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGA
TGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGCGATGGGAAAAA
ATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGC
AAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAA
GGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAG
AACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAG
AGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAA
GTAAGACCACCGCACAGCAAGCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATA
TGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCA
TTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAG

AGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACT
ATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTAT
AGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTG
CAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAA
GATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATT
TGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGAT
TTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGC
TTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAG
AATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATT
GGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAA
TAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATC
GTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA
AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATC
GGCACTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTAAA
AGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATA
GCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATT
TTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGTTAATTAAGGTACCAAGG
TCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA
AGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTA
CAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTA
TGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGG
CTTTATATATCTTGTGGAAAGGACGAAACACC**GCGGGCTGGAGCTGTTCGCGCG**
TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA
AAAAGTGGCACCGAGTCGGTGCAAGATGCAAGCGCGAACAGCTCCAGTTTT
TTTGAATTC*AGTGCAAGTGGGTTTTAGGACCAGGATGAGGCGGGGTGGGGGTGCCT*
[2276] *ACCTGACGACCGACCCCGACCCACTGGACAAGCACCCAACCCCCATTCCCCAAATTGC*
GCATCCCCTATCAGAGAGGGGGAGGGGAAACAGGATGCGGCGAGGCGCGTGCGCAC
TGCCAGCTTCAGCACCGCGGACAGTGCCTTCGCCCCCGCCTGGCGGCGCGCGCCAC
CGCCGCCTCAGCACTGAAGGCGCGCTGACGTCACTCGCCGGTCCCCCGCAAACTCCC
CTTCCCGGCCACCTTGGTCGCGTCCGCGCCGCCGCCGGCCCAGCCGGACCGCACCA
CGCGAGGCGCGAGATAGGGGGGCACGGGCGCGACCATCTGCGCTGCGGCGCCGGC
GACTCAGCGCTGCCTCAGTCTGCGGTGGGCAGCGGAGGAGTCGTGTCGTGCCTGAGA
*GCGCAG**TCGAGAATCTAGAGCGCTGCCACC***ATGAAACGGACAGCCGACGGAAGCG
AGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCGACAAGAAGTACAGCATCGGCC
TGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGT
GCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAA
GAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCG
GCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTA
TCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCAC
AGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCC
ATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCT
ACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGA
TCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGC
GACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAG
ACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCC
AAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCG
CCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAG
CCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAA
CTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAG
ATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCA
TCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAA

GCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGA
GCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCT
ACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGT
GAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAG
CATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAA
GATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCT
TCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTG
GATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGT
GGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCC
CGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACC
AACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAG
TGCCTGTCCTACGAGACAGAGATCCTGACAGTGGAGTATGGCCTGCTGCCAATCGGCA
AGATCGTGGAGAAGAGGATCGAGTGTACCGTGTACTCTGTGGATAACAATGGCAACATC
TATACACAGCCCGTGGCACAGTGGCACGATAGGGGAGAGCAGGAGGTGTTCGAGTATT
GCCTGGAGGACGGCAGCCTGATCAGGGCAACCAAGGACCACAAGTTCATGACAGTGG
ATGGCCAGATGCTGCCCATCGACGAGATTTTCGAGCGGGAGCTGGACCTGATGAGAGT
*GGATAACCTGCCTAAT**AGCGGAGGCAGTAAAAGAACAGCAGACGGGAGTGAGTTTG***
AGCCCAAGAAAAAGAGAAAGGTGGGATCCGGCGCAACAAACTTCTCTCTGCTGAAA
CAAGCCGGAGATGTCGAAGAGAATCCTGGACCGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
AAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCAC
CCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACG
ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC
TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCA
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGT
GCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA
ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT
CTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGAGGAGGAGGAGG
AGACAGACAGCAGGATGCCCCACCTCGACAGCCCCGGCAGCTCCCAGCCGAGACG
CTCCTTCCTCTCAAGGGTGATCAGGGCAGCGCTACCGTTGCAGCTGCTTCTGCTGC
TGCTGCTGCTCCTGGCCTGCCTGCTACCTGCCTCTGAAGATGACTACAGCTGCACC
CAGGCCAACAACTTTGCCCGATCCTTCTACCCCATGCTGCGGTACACCAACGGGCC
ACCTCCCACCTAAACGCGTTAAGTCGACAATCAACCTCTGGATTACAAAATTTGTG
AAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTG
CTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG
TATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACG
TGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCC
ACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGC
GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGC
ACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGC
CTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCC
TCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCG
CGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCG
TCGACTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAA
AGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCT
TTTTGCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTG

[2277]

[2278]

GCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAA
GTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTT
TTAGTCAGTGTGGAAAATCTCTAGCAGGGCCCGTTTAAACCCGCTGATCAGCCTCG
ACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG
ACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGGTGGGGTGGGGCAGGACAGC
AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTA
TGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCC
CTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCT
ACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCC
ACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCG
ATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCAC
GTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACG
TTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGT
CTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAG
CTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGG
TGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCA
ATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTT
TTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAG
TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTAT
ATCCATTTTCGGATCTGATCAGCACGTGTTGACAATTAATCATCGGCATAGTATATCG
GCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGC
CGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGAC
CGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGG
ACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACA
CCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGG
AGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGG
CGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGT
GCACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTCGATTCCACC
GCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGAT
GATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTA
TTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA
GCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATC
ATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGT
TTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC
ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT
GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA
TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC
GCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCAC
TCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGG
CGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA
GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAG
CCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT

GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG
CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTT
TCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCA
TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTA
AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA
GTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCC
CCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCA
ATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCC
AGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAG
TCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGC
AACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGC
[2279] TTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGT
GCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGC
CGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAA
AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCA
CCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAC
AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG
AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCAC
ATTTCCCCGAAAAGTGCCACCTGAC (SEQ ID NO: 736)

[2280] U6启动子

[2281] PEgRNA

[2282] hSyn启动子

[2283] N-端PE2

[2284] N-端Npu

[2285] P2A-GFP-KASH

[2286] pLenti-hSyn-C-PE2-NpuC:

GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGC
TGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAAT
TGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGC
CAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
TCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGC
CCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGT
TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC
GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCT
[2287] ATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTA
TGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT
TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC
GGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAG
GCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTTTGCCTGTACTGGGTCTCTC
TGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGC
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTG
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTC
TAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCT
CTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCG

GCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGA
TGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGCGATGGGAAAAA
ATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGC
AAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAA
GGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAG
AACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAG
AGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAA
GTAAGACCACCGCACAGCAAGCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATA
TGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCA
TTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAG
AGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACT
ATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTAT
AGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTG
CAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAA
GATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATT
TGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGAT
TTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGC
TTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAG
AATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATT
GGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAA
TAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATC
GTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA
AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATC
GGCACTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTAAA
AGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATA
[2288] GCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATT
TTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGTTAATTAAGGTACCAGTG
CAAGTGGGTTTTAGGACCAGGATGAGGCGGGGTGGGGGTGCCTACCTGACGACC
GACCCCGACCCACTGGACAAGCACCCAACCCCCATTCCCCAAATTGCGCATCCCC
TATCAGAGAGGGGGAGGGGAAACAGGATGCGGCGAGGCGCGTGCGCACTGCCAG
CTTCAGCACCGCGGACAGTGCCTTCGCCCCCGCCTGGCGGCGCGCGCCACCGCCG
CCTCAGCACTGAAGGCGCGCTGACGTCACTCGCCGGTCCCCCGCAAACTCCCCTT
CCCGGCCACCTTGGTCGCGTCCGCGCCGCCGCCGGCCCAGCCGGACCGCACCACG
CGAGGCGCGAGATAGGGGGGCACGGGCGCGACCATCTGCGCTGCGGCGCCGGCG
ACTCAGCGCTGCCTCAGTCTGCGGTGGGCAGCGGAGGAGTCGTGTCGTGCCTGAG
AGCGCAG*TCGAGAATCTAGAGCGCTGCCACCATGAAACGGACAGCCGACGGAAGCGA*
*GTTCGAGTCACCAAAGAAGAAGCGGAAAGTC*ATCAAGATTGCTACACGGAAATACCTGG
GAAAGCAGAACGTGTACGACATCGGCGTGGAGCGGGATCACAACTTCGCCCTGAAGAA
*TGGCTTTATCGCCAGCAAT***TGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGA**
TCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAG
GACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCG
TGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGA
AAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCG
GAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCG
GGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTC
GCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGC
ACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGAC
AGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAA

[2289]

CGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGAC
CAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGC
CTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAA
GCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGA
AGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCA
GAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACT
GGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCAC
AAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAG
AATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGG
TGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAA
CTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCT
GATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAG
GTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAG
GCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGA
GATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAA
CGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGT
GCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGT
GCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGA
TAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTT
CGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAA
GGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCAT
CATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAG
GGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCC
TGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAAC
TGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCA
GAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAG
CAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACA
AAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGG
CCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGC
CTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCTGGCGGCTCAAAAAGA
ACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAAACG
CGTTAAGTCGACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTAT
TCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTAT
CATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGC
TGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACT
GTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCT
TTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCT
GCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGT
GTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGA
TTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTT
CCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCC
TCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCGTCGACTTTAAGACCAAT
GACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTG
GAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGGGTC
TCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCA
CTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAA
TCTCTAGCAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCA

[2290]

CTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT
GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGG
AAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGA
AAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTA
AGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCC
TAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG
CACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC
CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGA
CTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTA
TAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAA
ATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCC
AGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
AGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATC
TCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT
CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCA
GAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTT
TTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGA
TCTGATCAGCACGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATA
CGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTC
ACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCT
CCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCT
GTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTG
TGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACG
AACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGG
GGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCC
GAGGAGCAGGACTGACACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATG
AAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCG
CGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAA
TGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
GCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCG
TCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAAT
TGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC
CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCG
CTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGC
TGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT
ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCG
CTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGC
CTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCAC
TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAC
AGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA
TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCC
GGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA
CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGT

ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATC
TCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA
TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCC
GGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGG
TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
[2291] CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGC
TTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCG
ACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGAT
CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA
TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTC
CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATAT
TTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAA
GTGCCACCTGAC (SEQ ID NO: 737)

[2292] hSyn启动子

[2293] C-端Npu

[2294] C-端wtCas9

[2295] pLenti-U6-DNMT1_切口产生_sgRNA:

TAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAG
GTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATT
CACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCAGA*GAGGGCCTA*
TTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAAT
TAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTC
TTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACT
*TGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC**GCCGCG***
CGCGCGAAAAAGCCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC
***GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC**TTTTTTAAGCTTGGCGTAACT
AGATCTTGAGACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGATT
GGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAA
CTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGG
GACAGCAGAGATCCACTTTGGCGCCGGCTCGAGGGGGCCCGGGTGCAAAGATGG
[2296] ATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGACCTTCTAGGTCTTGAA
AGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCAC
AGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGATCCGGTGCCTAGAGAAG
GTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCC
GAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTC
GCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCC
TGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACCTGGCTGCA
GTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGG
CCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGC
GCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTC
GATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGG
CAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGG
GGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCG
GGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCG

[2297]

GCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCA
AGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGC
CCTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGT
GAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTG
ACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTG
GAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACA
CTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTG
GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTT
CAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGACGTACGGCCACCATGACCGAGTA
CAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCAC
CCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGAC
CGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGC
TCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCCGTGGCGGTCTGGA
CCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCA
TGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCT
GGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGAGTCTC
GCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGA
GGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAA
CCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCCC
GAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGAACGCGTTAA
GTCGACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAA
CTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTC
TTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTT
GCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCG
GGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGT
CGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTG
CGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTC
CCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGA
CGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCGTCGACTTTAAGACCAATGACTT
ACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAG
GGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGGGTCTCTC
TGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGC
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTG
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTC
TAGCAGTACGTATAGTAGTTCATGTCATCTTATTATTCAGTATTTATAACTTGCAAAG
AAATGAATATCAGAGAGTGAGAGGAACTTGTTTATTGCAGCTTATAATGGTTACAA
ATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAG
TTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGCTCTAGCTATCCCGC
CCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCC
CATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGA
GCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGGACGTACCCAATTC
GCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTG
ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTC
GCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
GCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGG
TGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCT
CCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCT
CTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCC
CAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACG
GTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA

[2298]

ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTG
CCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAA
TTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCG
CGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGA
CAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCA
ACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCT
CACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGA
GTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC
CGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTAT
TATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAG
AATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA
CAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAAC
TTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACA
TGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATA
CCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGC
AAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTG
GATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGC
TGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGC
AGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACT
GATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTA
AAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATG
ACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAA
AGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA
CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTC
TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATAC
CTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG
AGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTG
ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAAC
GCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG
AGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGA
GGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATT
CATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCA
ACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGC
TTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAAC
AGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAA
AGCTGGAGCTGCAAGCTTAATGTAGTCTTATGCAATACTCTTGTAGTCTTGCAACAT
GGTAACGATGAGTTAGCAACATGCCTTACAAGGAGAGAAAAAGCACCGTGCATGC
CGATTGGTGGAAGTAAGGTGGTACGATCGTGCCTTATTAGGAAGGCAACAGACGG
GTCTGACATGGATTGGACGAACCACTGAATTGCCGCATTGCAGAGATATTGTATTT
AAGTGCCTAGCTCGATACATAAACGGGTCTCTCTGGTTAGACCAGATCTGAGCCTG
GGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
GAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCC
CTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGA
CTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGC
TGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAAT

TTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAG
CGGGGGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAG
AAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGC
AGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAG
CTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTA
GCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTT
TAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCG
GCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAA
TTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGC

[2299] AAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTT
CCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTG
ACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCT
GAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAG
CAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCC
TGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAAT
GCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGA
GTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAAT
CGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGC
AAGTTTGTGGAATTGGTTTAACA (SEQ ID NO: 738)

[2300] U6启动子

[2301] sgRNA

[2302] 本文使用的莫洛尼小鼠白血病病毒逆转录酶(M-MLV RT)变体的氨基酸序列

[2303] PE1M-MLVRT:

[2304] TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQ
EARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQ
WYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILL
QYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDE
KQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPP
DRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD
GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR
RGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP
(SEQ ID NO:739)

[2305] M3M-MLVRT(D200N,T330P,L603W)(参见Baranauskasetal.[182]):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPV
SIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
VNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
EMGISGQLTWTRLPQGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSE

[2306] LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV
MGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQ
EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVA
AGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH
YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDA

DHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA
[2307] EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRLSI
IHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 740)

[2308] PE2 M-MLVRT(D200N,T306K,W313F,T330P,L603W):

[2309] TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPV SIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE VNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRLPQGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSE LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVA AGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDA DHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRLSI IHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 741)

[2310] M3-deadRTM-MLVRT:

[2311] TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPV SIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVK**L**PGTNDY**S**PVQDLRE VNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRLPQGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSE LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVA AGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDA DHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRLSI IHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 742)

[2312] 实施例2的参考文献

[2313] 实施例12中引用下列各参考文献,其各自通过引用并入本文。

[2314] 1.Landrum,M.J.et al.ClinVar:public archive of interpretations of clinically relevant variants.Nucleic Acids Res.44,D862-D868(2016).

[2315] 2.Jinek,M.et al.A Programmable Dual-RNA-GuidedDNA Endonuclease in Adaptive Bacterial Immunity.Science337,816-821(2012).

[2316] 3.Cong,L.et al.Multiplex Genome Engineering Using CRISPR/Cas Systems.Science339,819-823(2013).

[2317] 4.Mali,P.et al.RNA-Guided Human Genome Engineering via Cas9.Science 339,823-826(2013).

[2318] 5.Yang,H.et al.One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering.Cell 154,1370-1379(2013).

[2319] 6.Kim,S.,Kim,D.,Cho,S.W.,Kim,J.&Kim,J.-S.Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.Genome Res.24,1012-1019(2014).

[2320] 7.Orlando,S.J.et al.Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology.Nucleic Acids Res.38,e152-e152(2010).

[2321] 8.Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-

target cleavage by CRISPRCas nucleases.Nat.Biotechnol.33,187-197(2015).

[2322] 9.Suzuki,K.et al.In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration.Nature 540,144-149(2016).

[2323] 10.Kosicki,M.,Tomberg,K.&Bradley,A.Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements.Nat.Biotechnol.36,765-771(2018).

[2324] 11.Haapaniemi,E.,Botla,S.,Persson,J.,Schmierer,B.&Taipale,J.CRISPR-Cas9genome editing induces a p53-mediated DNA damage response.Nat.Med.24,927-930(2018).

[2325] 12.Ihry,R.J.et al.p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells.Nat.Med.24,939-946(2018).

[2326] 13.Rouet,P.,Smih,F.&Jasin,M.Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.Proc.Natl.Acad.Sci.91,6064-6068(1994).

[2327] 14.Chapman,J.R.,Taylor,M.R.G.&Boulton,S.J.Playing the end game:DNA double-strand break repair pathway choice.Mol.Cell 47,497-510(2012).

[2328] 15.Cox,D.B.T.,Platt,R.J.&Zhang,F.Therapeutic genome editing:prospects and challenges.Nat.Med.21,121-131(2015).

[2329] 16.Paquet,D.et al.Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.Nature 533,125-129(2016).

[2330] 17.Chu,V.T.et al.Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.Nat.Biotechnol.33,543-548(2015).

[2331] 18.Maruyama,T.et al.Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.Nat.Biotechnol.33,538-542(2015).

[2332] 19.Rees,H.A.,Yeh,W.-H.&Liu,D.R.Development of hRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks.Nat.Commun.10,1-12(2019).

[2333] 20.Shen,M.W.et al.Predictable and precise template-free CRISPR editing of pathogenic variants.Nature 563,646-651(2018).

[2334] 21.Rees,H.A.&Liu,D.R.Base editing:precision chemistry on the genome and transcriptome of living cells.Nat.Rev.Genet.19,770(2018).

[2335] 22.Komor,A.C.,Kim,Y.B.,Packer,M.S.,Zuris,J.A.&Liu,D.R.Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424(2016).

[2336] 23.Gaudelli,N.M.et al.Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage.Nature 551,464-471(2017).

[2337] 24.Gao,X.et al.Treatment of autosomal dominant hearing loss by in

vivo delivery of genome editing agents.Nature 553,217-221(2018).

[2338] 25.Ingram,V.M.A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin.Nature 178,792-794(1956).

[2339] 26.Myerowitz,R.&Costigan,F.C.The major defect in Ashkenazi Jews with Tay-Sachs disease is an insertion in the gene for the alpha-chain of beta-hexosaminidase.J.Biol.Chem.263,18587-18589(1988).

[2340] 27.Zielenski,J.Genotype and Phenotype in Cystic Fibrosis.Respiration 67,117-133(2000).

[2341] 28.Mead,S.et al.A Novel Protective Prion Protein Variant that Colocalizes with Kuru Exposure.N.Engl.J.Med.361,2056-2065(2009).

[2342] 29.Marraffini,L.A.&Sontheimer,E.J.CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.Science 322,1843-1845(2008).

[2343] 30.Barrangou,R.et al.CRISPR provides acquired resistance against viruses in prokaryotes.Science 315,1709-1712(2007).

[2344] 31.Jiang,F.&Doudna,J.A.CRISPR-Cas9 Structures and Mechanisms.Annu.Rev.Biophys.46,505-529(2017).

[2345] 32.Hille,F.et al.The Biology of CRISPR-Cas:Backward and Forward.Cell 172,1239-1259(2018).

[2346] 33.Luan,D.D.,Korman,M.H.,Jakubczak,J.L.&Eickbush,T.H.Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.Cell 72,595-605(1993).

[2347] 34.Liu,Y.,Kao,H.-I.&Bambara,R.A.Flap endonuclease 1:a central component of DNA metabolism.Annu.Rev.Biochem.73,589-615(2004).

[2348] 35.Richardson,C.D.,Ray,G.J.,DeWitt,M.A.,Curie,G.L.&Corn,J.E.Enhancing homology directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.Nat.Biotechnol.34,339-344(2016).

[2349] 36.Qi,L.S.et al.Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.Cell 152,1173-1183(2013).

[2350] 37.Shechner,D.M.,Hacisuleyman,E.,Younger,S.T.&Rinn,J.L.Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display.Nat.Methods 12,664-670(2015).

[2351] 38.Tang,W.,Hu,J.H.&Liu,D.R.Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation.Nat.Commun.8, 15939(2017).

[2352] 39.Jinek,M.et al.Structures ofCas9 Endonucleases Reveal RNA-Mediated Conformational Activation.Science 343,1247997(2014).

[2353] 40.Nishimasu,H.et al.Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA.Cell 156,935-949(2014).

[2354] 41.Jiang,F.,Zhou,K.,Ma,L.,Gressel,S.&Doudna,J.A.A Cas9-guide RNA complex preorganized for target DNA recognition.Science 348,1477-1481(2015).

[2355] 42.Baranauskas,A.et al.Generation and characterization of new highly thermostable andprocessive M-MuLV reverse transcriptase variants.Protein Eng.Des.Sel.25,657-668(2012).

[2356] 43.Gerard,G.F.et al.The role of template-primer in protection of reverse transcriptase from thermal inactivation.Nucleic Acids Res.30,3118-3129(2002).

[2357] 44.Arezi,B.&Hogrefe,H.Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer.Nucleic Acids Res.37,473-481(2009).

[2358] 45.Kotewicz,M.L.,Sampson,C.M.,D'Alessio,J.M.&Gerard,G.F.Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity.Nucleic Acids Res.16,265-277(1988).

[2359] 46.Thuronyi,B.W.et al.Continuous evolution of base editors with expanded target compatibility and improved activity.Nat.Biotechnol.(2019).

[2360] doi:10.1038/s41587-019-0193-0

[2361] 47.Kim,Y.B.et al.Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions.Nat.Biotechnol.35,371-376(2017).

[2362] 48.Koblan,L.W.et al.Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction.Nat.Biotechnol.(2018).doi:10.1038/nbt.4172

[2363] 49.Kleinstiver,B.P.et al.High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off target effects.Nature 529,490-495(2016).

[2364] 50.Zuo,E.et al.Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos.Science 364,289-292(2019).

[2365] 51.Jin,S.et al.Cytosine,but not adenine,base editors induce genome-wide off-target mutations in rice.Science 364,292-295(2019).

[2366] 52.Kim,D.,Kim,D.,Lee,G.,Cho,S.-I.&Kim,J.-S.Genome-wide target specificity of CRISPR RNA guided adenine base editors.Nat.Biotechnol.37,430-435(2019).

[2367] 53.Grunewald,J.et al.Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors.Nature569,433-437(2019).

[2368] 54.Zhou,C.et al.Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis.Nature571,275-278(2019).

[2369] 55.Rees,H.A.,Wilson,C.,Doman,J.L.&Liu,D.R.Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci.Adv.5,eaax5717(2019).

[2370] 56.Ostertag,E.M.&KazazianJr,H.H.Biology of MammalianL1 Retrotranspos

ons.Annu.Rev.Genet.35,501-538(2001).

[2371] 57.Griffiths,D.J.Endogenous retroviruses in the human genome sequence.Genome Biol.2,REVIEWS1017(2001).

[2372] 58.Berkhout,B.,Jebbink,M.&Zsiros,J.Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus.J.Virol.73,2365-2375(1999).

[2373] 59.Halvas,E.K.,Svarovskaia,E.S.&Pathak,V.K.Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity.J.Virol.74,10349-10358(2000).

[2374] 60.Dever,D.P.etal.CRISPR/Cas9Beta-globin Gene Targeting in Human Hematopoietic Stem Cells.Nature 539,384-389(2016).

[2375] 61.Park,S.H.et al.Highly efficient editing of the β-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease.Nucleic Acids Res.doi:10.1093/nar/gkz475

[2376] 62.Collinge,J.Prion diseases of humans and animals:their causes and molecular basis.Annu.Rev.Neurosci.24,519-550(2001).

[2377] 63.Asante,E.A.et al.A naturally occurring variant of the human prion protein completely prevents prion disease.Nature522,478-481(2015).

[2378] 64.Zettler,J.,Schutz,V.&Mootz,H.D.The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction.FEBS Lett.583,909-914(2009).

[2379] 65.Kugler,S.,Kilic,E.&Bahr,M.Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area.Gene Ther.10,337-347(2003).

[2380] 66.de Felipe,P.,Hughes,L.E.,Ryan,M.D.&Brown,J.D.Co-translational,intraribosomal cleavageof polypeptides by the foot-and-mouth disease virus 2Apeptide.J.Biol.Chem.278,11441-11448(2003).

[2381] 1.建立诱变DNA链的向导RNA模板化逆转录

[2382] 背景和基本原理

[2383] 在提议的基因组编辑策略中,Cas9产生切口的非靶DNA链(形成R环的含有PAM的链)充当DNA合成的引物。假设这根据一些生化和结构数据是可能的。核酸酶保护实验[32]、晶体学研究[33]和碱基编辑窗口[4,24]已显示,在Cas9结合复合物的所谓R环内的-20至-10(编号表示距第一个PAM核苷酸5′的距离)的非靶链核苷酸的较大程度的灵活性和无序性。此外,以反式添加互补ssDNA时,切割的非靶链的PAM远端部分可从紧密结合的三元复合物中置换[20]。这些研究支持非靶链具有高度灵活性,酶可及性,并且在产生切口后PAM远端片段的3′端在Cas9解离之前释放。此外,假设gRNA可延伸至模板DNA合成。先前研究表明,SpCas9、SaCas9和LbCas12a(以前称为Cpf1)的gRNA耐受用RNA适体[34]、配体诱导型自切割核酶[35]和长的非编码RNA[36]的gRNA延伸。该文献为将要利用的两个主要特征建立了先例。在评估该策略时,使用

体外和细胞测定的组合与5’和3’延伸的gRNA设计结合来评估多个CRISPR-Cas系统(图2A至2C)。

[2384] 用于TRT编辑的工程化gRNA的设计示于图3A至3B。DNA合成从5′至3′进行,因此在3′至5′方向复制RNA模板。5′延部分的设计包含接头区域,切口DNA链退火的引物结合位点,以及通过逆转录进行DNA合成的模板。3′延伸的gRNA包含引物结合位点和逆转录模板。在某些情况下,gRNA核心的3′RNA发夹经修饰以匹配DNA靶序列,因为体外实验表明,逆转录将约3个核苷酸延伸至3′延伸的gRNA构建体的gRNA核心(只要进行维持发夹RNA结构的补偿性变化,发夹序列的序列似乎耐受良好)。DNA合成从5′至3′进行,其中核苷酸添加到正在生长的DNA链的3′OH。

[2385] 初步结果

[2386] Cas9产生切口的DNA引发gRNA模板的逆转录。为评估有切口的非靶DNA链的可及性,使用来自化脓性链球菌的Cas9核酸酶(SpCas9)和Cy5荧光标记的双链体DNA底物(51个碱基对)进行体外生化测定。首先,通过体外转录制备一系列包含具有不同合成模板长度的5′延伸的gRNA(总体设计如图2B所示)。利用核酸酶死亡Cas9(dCas9)进行电泳迁移率变化测定(EMSA)建立了5′延伸的gRNA保持靶结合亲和力(数据未显示)。接着,使用dCas9、5′-延伸的gRNA和莫洛尼-鼠白血病病毒(M-MLV)逆转录酶(Superscript III)在预先产生切口的Cy5标记的双链体DNA底物上测试TPRT活性。于37℃温育1小时后,通过变性聚丙烯酰胺凝胶电泳(PAGE)评估产物并使用Cy5荧光成像(图4A)。每个5′-延伸的gRNA变体导致明显的产物形成,观测到的DNA产物大小与延伸模板的长度一致(图4B)。重要地,在没有dCas9的情况下,将预先产生切口的底物延伸至DNA底物的全长51bp,这强烈表明dCas9不存在时,将互补的DNA链而不是gRNA用作DNA合成的模板(图4C)。值得注意地,设计该系统使得新合成的DNA链反映靶位点编辑所需的产物(具有单核苷酸变化的同源链)。该结果建立了Cas9:gRNA结合暴露了有切口的非靶链的3’末端,并且非靶链可进行逆转录。

[2387] 接着,使用Cas9(H840A)突变体评估非产生切口的dsDNA底物,该突变体对非靶DNA链产生切口。首先,为了用5′延伸的gRNA测试Cas9(H840A)切口酶活性,如先前所述[37]进行体外切割测定。尽管与标准gRNA相比切口产生受损,但形成可察觉的切割产物(图4D)。重要地,当用5′-延伸的gRNA和Cas9(H840A)进行TPRT反应时,也观测到RT产物,尽管产率较低,这可能通过切口产生活性降低解释(图4D)。该结果建立了5′-延伸的gRNA:Cas9(H840A)复合物可对DNA产生切口和为逆转录提供模板。

[2388] 最后,针对Cas9(H840A)切口产生和TPRT评估3′gRNA延伸。与5′延伸的gRNA相比,3′延伸的gRNA对DNA的切割与标准gRNA相比没有受到任何可检测程度的损害。重要地,M-MLVRT以反式提供时,3′延伸的gRNA模板也支持利用预先产生切口且完整的双链体DNA底物的有效逆转录(图4E)。令人惊讶地,对于3′延伸的模板仅观测到单一产物,表明逆转录沿着gRNA支架的特定位置终止。用末端转移酶对产物的同聚物加尾,随后是Klenow延伸和Sanger测序揭示,除了gRNA核心的末端3个核苷酸外,还复制了完整的gRNA合成模板。将来,瓣末端将通过修饰末端gRNA序列重新编程[38,39]。该结果表明,3′延伸的gRNA可充当有效的核酸酶靶向向导物,并且可为逆转录提供模板。

[2389] Cas9-TPRT以顺式使用有切口的DNA和gRNA。双色实验用于确定RT反应是否优先在顺式(结合在同一复合物中)的gRNA的情况下发生(参见图8)。对5′-延伸和3′-延伸的gRNA

进行两个独立实验。对于给定实验，dCas9、gRNA和DNA底物的三元复合物在不同管中形成。在一个管中，gRNA编码长的RT产物，并且DNA底物用Cy3(红色)标记；另一方面，gRNA编码短的RT产物，DNA底物用Cy5(蓝色)标记。短暂温育后，将复合物混合，然后用RT酶和dNTP处理。通过尿素变性PAGE分离产物，并通过Cy3和Cy5通道中的荧光可视化。发现使用与DNA底物预先复合的gRNA模板优先形成反应产物，表明RT反应可能以顺式发生。该结果支持单个Cas9:gRNA复合物可靶向DNA位点并为逆转录诱变DNA链提供模板。

[2390] 使用其他Cas系统测试TPRT

[2391] 使用其他Cas系统，包括来自金黄色葡萄球菌的Cas9和来自L.bacterium的Cas12a进行与前面部分中呈现的那些实验类似的实验(参见图2A至2C)。如果对于这些Cas变体也可以证明TRPT，则潜在的编辑范围和在细胞中整体成功的可能性会增加。

[2392] 利用RT-Cas9融合蛋白测试TPRT

[2393] 首先以反式评估一系列市售或可纯化的RT酶的TPRT活性。除了已测试的来自M-MLV的RT外，评估了来自禽成肌细胞增多病毒(AMV)、嗜热脂肪土芽孢杆菌II组内含子(GsI-IIC)[41,42]和直肠真杆菌II组内含子(Eu.re.I2)[43,44]的RT。显著地，后两个RT在其自然生物环境中执行TPRT。在相关情况下，测试RNAse失活突变和其他潜在有益的RT酶修饰。一旦以反式提供时鉴定功能性RT，每个RT将作为与Cas9变体的融合蛋白进行评估。测试N端和C端融合方向，以及各种的接头长度和结构。将动力学时程实验用于确定是否可以顺式使用RT酶发生TPRT。如果可构建允许有效TPRT化学的RT-Cas9融合结构，则这将大大增加在细胞环境中进行功能编辑的可能性。

[2394] 细胞中利用工程化gRNA的Cas9靶向

[2395] 在人类细胞培养实验(HEK293)中评估在前面子目标中开发的候选工程化gRNA，以确认Cas9靶向效率。使用已建立的使用野生型SpCas9的插入缺失形成测定[45]，跨越人基因组中5个或更多位点，并行比较工程化gRNA与标准gRNA。基因组编辑效率通过使用实验室内安装的IlluminaMiSeq平台以多重方式进行扩增子测序来表征。预计此部分和前面部分的结果将产生通告设计-构建-测试循环的后续迭代的见解，其中gRNA可针对在细胞中为逆转录提供模板和有效Cas9靶向进行优化。

[2396] 体外验证的结果显示在图5至图7中。体外实验表明，有切口的非靶DNA链是柔性的，并且可用于引发DNA合成，并且gRNA延伸可充当逆转录的模板(参见图5)。这组实验使用5′延伸的gRNA(设计如图3A至3B所示)与不同长度的合成模板(列在左侧)。将荧光标记的(Cy5)DNA靶标用作底物，并在这组实验中预先产生切口。这些实验中使用的Cas9是催化死亡Cas9(dCas9)，因此不能切割DNA但仍可有效结合。Superscript III(衍生自莫洛尼鼠白血病病毒(M-MLV)的商业RT)以反式提供。首先，dCas9:gRNA复合物由纯化的组件形成。然后，将荧光标记的DNA底物与dNTP和RT酶一起加入。于37℃温育1小时后，通过变性尿素-聚丙烯酰胺凝胶电泳(PAGE)分析反应产物。凝胶图像显示原始DNA链的延伸长度与逆转录模板的长度一致。值得注意地，在不存在dCas9的情况下进行的反应产生长度为51个核苷酸的DNA产物，而与所使用的gRNA无关。该产物对应于使用互补DNA链作为DNA合成的模板，而不是RNA(数据未显示)。因此，需要Cas9结合才能将DNA合成引导至RNA模板。这组体外实验与图5所示的那些非常相似，只是DNA底物没有预先产生切口，并且使用Cas9切口酶(SpyCas9 H840A突变体)。如凝胶所示，当使用标准gRNA时，切口酶有效切割DNA链(gRNA_0，泳道3)。观

测到多种切割产物,与之前SpyCas9的生化研究一致。5′延伸损害切口产生活性(泳道4-8),但仍观测到一些RT产物。图7显示3′延伸支持DNA合成,并且不会显著实现Cas9切口酶活性。当使用dCas9或Cas9切口酶(泳道4和泳道5)时,预先产生切口的底物(黑色箭头)几乎定量地转化为RT产物。使用完整底物(泳道3)观测到超过50%的RT产物转化率(红色箭头)。为了确定RT产物的长度和序列,从凝胶上切下产物条带,提取并测序。这显示RT将3个核苷酸延伸至gRNA核心的3′端发夹中。随后的实验(未显示)表明只要进行保持发夹RNA结构的互补改变,可将这三个核苷酸改变以匹配靶DNA序列。

[2397] 潜在的困难和替代方案

[2398] (1)RT不作为融合物发挥功能:分子拥挤和/或不利的几何构造可阻碍Cas9融合的RT酶的聚合酶延伸。首先,可测试接头优化。评估Cas9的环状排列变体,其可重新定向DNA引物、gRNA和RT酶之间的空间关系。可测试目标2中详述的非共价RT募集策略。(2)延伸的gRNA变体的Cas靶向效率降低:这最可能是5′-延伸的gRNA的问题。基于结构数据[24],可设计和筛选Cas9突变体,以鉴定对gRNA延伸具有更大耐受性的变体。此外,可在细胞中筛选gRNA文库以寻找提高靶向活性的接头。

[2399] 重要性

[2400] 这些初步结果建立了,使用反式提供的逆转录酶,Cas9切口酶和延伸的gRNA可在结合的DNA靶标上启动靶标引发的逆转录。重要地,发现Cas9结合对于产物形成至关重要。虽然可能不是细胞中基因组编辑的绝对要求,但进一步开发以顺式并入RT酶功能的系统将显著增加基于细胞的应用中成功的可能性。该目标的其余方面的实现将为在人类基因组的背景下进行精确基因组编辑提供分子基础。

[2401] 2.在人细胞中建立引导编辑

[2402] 背景和基本原理

[2403] 在提议的策略中,工程化RT-Cas9:gRNA复合物在基因组靶位点处引入诱变的3′DNA瓣。假设含有单个错配的诱变3′瓣通过DNA修复机制经由与邻近5′瓣的能量可达平衡而掺入,所述邻近5′瓣是优先去除的(图1C至1D)。DNA复制和修复机制在加工冈崎片段[46]时和长补丁碱基切除修复(LP-BER)期间[47]遇到5′ssDNA瓣。5′瓣是广泛表达的瓣内切核酸酶FEN1的优选底物,该FEN1被同源三聚体滑动钳复合体PCNA募集至DNA修复位置[48]。PCNA还可充当同时募集其他修复因子(包括DNA连接酶Lig1)的支架[49]。作为“工具带(toolbelt)”起作用,PCNA加速连续瓣切割和连接,这对于加工每次细胞分裂过程中产生的数百万冈崎片段至关重要[50,51]。基于与这些天然DNA中间体的相似性,假设诱变链通过与5′瓣平衡,随后进行协调的5′瓣切除和连接掺入。然后,错配修复(MMR)应当以相等的概率发生在任一条链上,导致编辑或回复(图1C至1D)。或者,可先发生DNA复制,并直接导致编辑掺入新合成的子链中。虽然来自该过程的最高预期产率为50%,但由于编辑修复的不可逆性,多个底物编辑尝试可驱动反应完成。

[2404] 初步结果

[2405] DNA瓣在酵母和HEK细胞中的质粒模型底物中诱导位点特异性诱变。为测试所提议的编辑策略,开始使用模型质粒底物进行研究,该底物含有类似于TPRT产物的诱变3′瓣。创建编码GFP和mCherry之间的终止密码子的双荧光蛋白报告基因。诱变瓣编码对终止密码子的校正(图9A),实现mCherry合成。因此,诱变效率可通过GFP:mCherry比率来定量。体外制

备质粒底物并将其引入酵母(酿酒酵母)或人细胞(HEK293)。在两个系统中都观测到高频率诱变(图9B),并且分离的酵母集落含有回复碱基、突变碱基或两种产物的混合物(图9C)。后者的检测表明在这些情况下质粒复制发生在MMR之前,并进一步表明瓣切除和连接先于MMR。这一结果建立了使用3′诱变链进行DNA编辑的可行性。

[2406] • 用模型瓣底物的系统研究

[2407] 基于上述初步结果,在HEK细胞中评估更广泛的瓣底物,以推断有效编辑的原理。将3′ssDNA瓣系统地改变以确定错配配对的影响、沿瓣的诱变核苷酸的位置、以及末端核苷酸的身份(图9D)。还测试了单核苷酸插入和缺失。将扩增子测序用于分析编辑精确度。这些结果将有助于为gRNA逆转录模板的设计提供信息。

[2408] 质粒底物上的体外TPRT导致有效的编辑结果。将目标1中开发的TPRT反应用于在质粒底物中诱导诱变。反应在环状DNA质粒底物上进行(参见图10)。这排除了DNA链解离作为先前体外实验中RT延伸机制的可能性。其还允许测试细胞中瓣底物的DNA修复。构建了用于酵母(酿酒酵母)表达的双荧光报告质粒。该质粒编码带有中间终止密码子(TGA)的GFP(绿色荧光蛋白)和mCherry(红色荧光蛋白)。该构建体在酵母中的表达仅产生GFP。将该质粒用作体外TRT的底物[Cas9(H840A)切口酶、工程化gRNA、MLVRT酶、dNTPS]。gRNA延伸编码对终止密码子的突变。将瓣链用于修复终止密码子,预计产生作为融合蛋白表达GFP和mCherry两者的质粒。酵母双-FP质粒转化体示于图10。转化亲本质粒或体外Cas9(H840A)产生切口的质粒仅产生绿色GFP表达集落。用5′-延伸或3′-延伸的gRNA进行的TRT反应产生绿色和黄色集落的混合物。后者表达GFP和mCherry两者。使用3′延伸的gRNA观测到更多黄色集落。还显示不含终止密码子的阳性对照。

[2409] 该结果建立了,长的双链底物可经历TPRT,并且TPRT产物在真核细胞中诱导编辑。

[2410] 进行了另一个与上述引导编辑实验类似的实验,但替换在终止密码子中安装点突变,TRT编辑安装单核苷酸插入(左)或缺失(右),其修复移码突变并允许下游mCherry的合成(参见图11)。这两个实验都使用3′延伸的gRNA。选择并且通过Sanger测序分析来自TRT转化的个别集落(参见图12)。绿色集落包含具有原始DNA序列的质粒,而黄色集落包含由TRT编辑gRNA设计的精确突变。未观测到其他点突变或插入缺失。

[2411] 在HEK细胞中使用RT-Cas9结构建立引导编辑

[2412] 来自前面目标的优化构建体将适用于哺乳动物表达和在人基因组中靶向位点处的编辑。除了使用二级gRNA(截短以防止切口产生)的邻近靶向之外,还测试了多种RT酶和融合结构。还使用Sun-Tag系统[52]和MS2适体系统[53]评估非共价RT募集。将插入缺失形成测定用于评估用标准gRNA和RT-Cas9融合物的靶向效率(如上)。然后,对于每个基因组位点,对延伸的gRNA和RT-Cas9对测定单核苷酸编辑。使用MiSeq评估编辑结果。

[2413] 在HEK细胞中使用Cas9-RT融合物进行初步实验。通过细胞内表达的组件进行编辑需要Cas9(H840A)切口酶、逆转录酶(以融合物表达或以反式形式提供)和具有3’延伸的工程化gRNA(参见图14)。初步研究表明,gRNA延伸内的引物结合位点的长度对于提高人细胞中的编辑效率是重要的(参见图15)。

[2414] 优化HEK细胞中的引导编辑参数

[2415] 在鉴定出可在细胞中进行引导编辑的Cas9-RT结构后,优化组件和设计以实现高效编辑。改变编码点突变的位置和核苷酸身份,以及新合成的DNA链的总长度,以评估编辑

范围和潜在限制。还评估短的插入和缺失突变。对蛋白表达构建体进行密码子优化。如果成功,则这将在哺乳动物细胞中建立有效的引导编辑。

[2416] 初步结果

[2417] 在融合的RT酶不可能进行分子内逆转录的情况下,设计其他gRNA以使RT酶在编辑基因座处达到更高的局部浓度。这些辅助向导物在5′端处截短(14-15nt间隔区),这在之前已证明可防止Cas9切割但保留结合(参见图16)。选择HEK3基因座来探索该策略。

[2418] 潜在的困难和替代方案

[2419] (1)细胞中的gRNA降解:如果延伸的gRNA末端在细胞中截短,则可安装稳定化的二级结构,或可测试具有稳定化修饰的合成gRNA。(2)在人细胞中无观测到的编辑:将探索其他策略,包括将RT-Cas9融合物二次靶向到邻近的基因组位点[54]。此外,可探索大肠杆菌或酿酒酵母中潜在的定向进化策略。

[2420] 重要性

[2421] 如果可在实验细胞系中建立引导编辑,则这将通过实现快速生成和表征人基因中的大量点突变,对基础生物医学研究产生直接影响。该方法的通用性及其相对于碱基编辑器的正交编辑窗口将提供安装许多当前难以接近的突变的方法。此外,如果可针对高效率和产物纯度优化引导编辑,那么其校正其他人细胞类型中的疾病突变的潜在适用性将具有重要意义。

[2422] 3.实现对培养人细胞致病性突变的位点特异性编辑

[2423] 背景和基本原理

[2424] 由于PAM限制,或需要颠换或插入缺失突变校正,当前的碱基编辑器无法校正大量致病性突变。通过引导编辑,理论上所有的转换和颠换都是可能的,小的插入和缺失也是如此。此外,关于PAM,引导编辑窗口(预期-3至+4)不同于碱基编辑器的窗口(-18至-12)(图13)。碱基编辑器目前无法校正的孟德尔(Mendelian)病况包括:(1)血红蛋白β中的镰状细胞病Glu6Val创始者突变(需要A·T至T·A的颠换);(2)ATP7B中最常见的威尔逊氏病变体His1069Gln(需要G·C至T·A的颠换);以及(3)导致囊性纤维化的CFTR中的ΔPhe508突变(需要3个核苷酸插入)。这些靶标中的每一个都包含适当定位的PAM,用于SpCas9靶向和引导编辑。

[2425] 初步结果

[2426] ·HEK3细胞中的T至A编辑不能通过当前的碱基编辑实现,但可通过TRPT编辑实现(参见图17A至17C)

[2427] 图17A示出显示在人胚胎肾(HEK)细胞中组件转染后靶核苷酸处%T至A转换的图。该数据展示使用野生型MLV逆转录酶与Cas9(H840A)切口酶的N端融合物(32个氨基酸接头)的结果。当引物结合位点的长度从7个核苷酸延伸至11或12个核苷酸时,编辑效率显著提高。此外,正好位于编辑基因座上游的辅助向导物A(参见图16)显著提高编辑活性,特别是对于较短长度的引物结合位点。使用IlluminaMiSeq平台通过扩增子测序定量编辑效率。图17B还示出在人胚胎肾(HEK)细胞中组件转染后靶核苷酸处%T至A转换,但该数据展示使用RT酶的C-端融合物的结果。在此,辅助向导物A没有那么大的作用,并且编辑效率整体较高。图17C示出的数据展示使用野生型MLV逆转录酶与Cas9(H840A)切口酶的N端融合物的结果,与图17C中的结果类似;然而,MLVRT和Cas9之间的接头长60个氨基酸,而不是32个氨基酸。

[2428] • 通过TRPT编辑结果得到的HEK3位置处T至A编辑显示高纯度

[2429] 图18示出通过高通量扩增子测序进行的测序分析的输出。输出显示编辑细胞的最丰富的基因型。值得注意地，未获得主要的插入缺失产物，并且清洁地安装期望的点突变(T至A)，而无旁观者编辑。第一序列显示参考基因型。上方两个产物是包含内源性多态性(G或A)的起始基因型。下方两个产物代表经正确编辑的基因型。

[2430] • MLVRT突变体改进编辑

[2431] 以与Cas9(H840A)切口酶的C端融合物测试Baranauskas等(doi:10.1093/protein/gzs034)所述的突变体逆转录酶在人胚胎肾(HEK)细胞中的靶核苷酸编辑。Cas9-RT编辑器质粒与为逆转录提供模板的编码3′-引导编辑向导RNA的质粒共转染。图19中在插入缺失比率(橙色柱)旁边示出靶核苷酸处的编辑效率(蓝色柱)。WT是指野生型MLVRT酶。突变酶(M1至M4)包含右侧列出的突变。编辑率通过基因组DNA扩增子的高通量测序来定量。

[2432] 利用第二gRNA进行的互补链切口产生改进编辑

[2433] 本实验评估了在互补DNA链中靠近靶核苷酸处引入单链切口时靶核苷酸的编辑效率，假设这将引导错配修复优先去除原始核苷酸并转换碱基对至期望编辑。Cas9(H840A)-RT编辑构建体与两个向导RNA编码质粒共转染，所述向导RNA编码质粒之一为逆转录反应提供模板，而另一个靶向互补DNA链产生切口。测试距靶核苷酸不同距离处的产生切口(橙色三角形)(参见图20)。在插入缺失形成率(橙色柱)旁边示出靶碱基对的编辑效率(蓝色柱)。“无”示例不包含互补链切口产生性向导RNA。编辑率通过基因组DNA扩增子的高通量测序来定量。

[2434] 图21示出经处理的高通量测序数据，显示期望的T至A颠换突变和其他主要基因组编辑副产物的普遍缺失。

[2435] 范围。新编辑技术的潜在范围如图13所示，并与脱氨酶介导的碱基编辑器技术进行比较。先前开发的碱基编辑器靶向PAM上游约15±2bp的区域。通过将靶标C或A核苷酸分别转换为T或G，先前开发的碱基编辑器可实现所有转换突变(A:T至G:C的转换)。但是，先前开发的碱基编辑器无法安装颠换突变(A至T、A至C、G至T、G至C、T至A、T至G、C至A、C至G)。此外，如果编辑窗口中有多个靶核苷酸，则可导致额外的非期望编辑。

[2436] 新的引导编辑技术理论上可安装任何核苷酸和碱基对转换，以及潜在的小的插入和缺失编辑。对于PAM，引导编辑窗口从DNA切口产生的位点(PAM上游3个碱基)开始，并在PAM下游尚未确定的位点结束。值得注意地，此编辑窗口不同于脱氨酶碱基编辑器的编辑窗口。由于TPRT系统使用DNA聚合酶进行编辑，因此它可能具有所有优点，包括通用性、精确度和保真度。

[2437] 校正患者来源细胞系中的致病性突变

[2438] 从ATCC、Coriell Biobank或合作的Harvard/Broad附属实验室获得携带相关突变的细胞系(镰状细胞病:CD34+造血干细胞;威尔逊氏病:培养的成纤维细胞;囊性纤维化:培养的支气管上皮细胞)。通过高通量测序评估编辑效率，并且使用表型分析(血红蛋白HPLC、ATP7B免疫染色和CFTR膜电位分析)测试经校正的基因型的功效。

[2439] 表征脱靶编辑活性

[2440] 使用与野生型Cas9配对的靶gRNA，利用已建立的方法(如GUIDE-seq[55]和CIRCLE-seq[56])筛选潜在的脱靶编辑。如果鉴定了潜在的脱靶，则这些基因座将在TPRT编辑的细胞中

进行探测,以鉴定真正的脱靶编辑事件。

[2441] 潜在的困难和替代方案

[2442] (1)编辑效率低:引导编辑器(PE)可能需要针对每个靶标进行优化。在这种情况下,可测试gRNA文库,以鉴定针对特定应用的最高功能变体。可优化RT-Cas融合表达和核定位。可将脂质体RNP递送用于限制脱靶编辑。

[2443] 即将进行的实验

[2444] gRNA设计的优化可通过进一步探索引物结合位点长度和延伸合成模板来实现。测试范围和通用性包括不同的核苷酸转换、小的插入和缺失,以及关于PAM的不同编辑位点,以及人基因组中的多个位点。RT组件的优化包括探索MLVRT中的突变以增强活性(Rnase H失活、增加引物-模板结合亲和力、调整持续合成能力)和新的RT酶(II组intro RT、其他逆转录病毒RT)。

[2445] 重要性

[2446] 无数的遗传疾病是由单个基因的单核苷酸变化引起的。开发在此描述的基因组编辑技术,并将其应用于与疾病相关的细胞类型,会为临床转化奠定基础。对于某些疾病,例如镰状细胞病,单点突变代表整个群体的显性基因型。然而,对于许多其他遗传疾病,在整个患者群体中观测到单个基因内不同点突变的巨大异质性,每一种都会导致相似的疾病表型。因此,作为理论上可靶向大量此类突变的通用基因组编辑方法,这项技术可为这些患者及其家庭提供巨大的潜在利益。如果可在细胞中建立这些应用原理的证明,则将为疾病动物模型的研究奠定基础。

[2447] 优点

[2448] 精确度:编码核酸序列中定向的期望编辑。通用性:理论上,可进行任何碱基对转换,包括颠换编辑,以及小的插入或缺失。关于Cas9原间隔区邻近基序(PAM)序列,存在与碱基编辑器不同的编辑窗口。此方法实现了同源定向修复(HDR)的许多编辑能力,但没有HDR的主要限制(在大多数细胞类型中效率低下,并且通常伴随着过量的非期望的副产物,如插入缺失)。此外,其不会造成双链DNA断裂(DSB),因此很少插入缺失、易位、大的缺失、p53激活等。

[2449] 实施例3-使用PE的肽标签化

[2450] 本文所述引导编辑系统(即,PE系统)还可用于将各种肽标签引入蛋白编码基因中。这类标签可包括6x组氨酸标签、FLAG-标签、V5-标签、GCN4-标签、HA-标签、Myc-标签等。这种方法可用于诸如蛋白荧光标记、免疫沉淀、免疫印迹、免疫组织化学、蛋白募集、诱导型蛋白降解决定子(degron)和全基因组筛选的应用中。图25和图26描绘了实施方案。

[2451] 图25是显示内源性基因组基因座处的肽标签化基因和利用TPRT基因组编辑(即,引导编辑)进行肽标签化的gRNA设计的示意图。FlAsH和ReAsH标签化系统包括两部分:(1)荧光团-双砷探针,和(2)含有四半胱氨酸基序的基因编码肽,例如序列FLNCCPGCCMEP(SEQ ID NO:1)。在细胞内表达时,含有四半胱氨酸基序的蛋白可用荧光团-砷探针进行荧光标记(参见文献:J.Am.Chem.Soc.,2002,124(21),pp 6063-6076.DOI:10.1021/ja017687n)。“分选标签化”系统采用细菌分选酶,其将标记的肽探针与含有合适肽底物的蛋白共价缀合(参见文献:Nat.Chem.Biol.2007Nov;3(11):707-8.DOI:10.1038/nchembio.2007.31)。FLAG-标签(DYKDDDDK(SEQ ID NO:2))、V5-标签(GKPIPNPLLGLDST(SEQ ID NO:3))、GCN4-标签

(EELLSKNYHLENEVARLKK(SEQ ID NO:4))、HA-标签(YPYDVPDYA(SEQ ID NO:5))和Myc-标签(EQKLISEEDL(SEQ ID NO:6))通常用作免疫测定的表位标签。π钳(pi-clamp)编码肽序列(FCPF)(SEQ ID NO:622),其可用五氟芳香族底物标记(参考文献:Nat.Chem.2016Feb;8(2):120-8.doi:10.1038/nchem.2413)。

[2452] 图26示出His6-标签和FLAG-标签精确安装至基因组DNA。靶向HEK3基因座的向导RNA设计为具有编码18-nt His标签插入或24-ntFLAG标签插入的逆转录模板。使用扩增子测序评估转染的HEK细胞中的编辑效率。注意,FLAG标签的完整24-nt序列在观测框外(测序确认完整和精确插入)。

[2453] 实施例4-使用PE进行RNA标签化和操作

[2454] 本文描述了将基序插入遗传序列中的新方法,所述基序标签化或者以其它方式操作哺乳动物、真核和细菌细胞内的RNA。虽然据估计只有1%的人类基因组编码蛋白,但实际上所有基因组都在某种程度上转录。所产生的非编码RNA(ncRNA)在多大程度上发挥功能性作用是开放的问题,更不用说大多数这些假定的RNA的作用如何。通过将具有有用特性的新的RNA编码序列插入感兴趣基因来"标签化"这些RNA分子是研究细胞中RNA分子生物学功能的有用方法。将标签安装到编码蛋白的mRNA上作为干扰的手段也是有用的,从而更好地了解mRNA修饰如何影响细胞功能。例如,细胞使用普遍存在的天然RNA标签——聚腺苷酸化来影响mRNA向细胞质的转运。不同类型的聚腺苷酸化信号导致不同的转运速率和不同的mRNA寿命,从而导致编码蛋白表达水平的差异。

[2455] 细胞内表达标签化的RNA的常用方法是外源引入合成构建体,其使用以下进行(i)瞬时质粒转染,其通常以超生理水平产生短期基因表达爆发;或(ii)使用慢病毒整合或转座子将标签化的RNA基因永久整合至基因组(随机位点处),从而实现延长的表达。这两种方法都受到产生改变的表达水平以及缺乏调节基因表达或活性的天然机制的限制。另一种策略是使用由Cas9或其他靶向DNA核酸酶诱导的双链DNA断裂的同源定向修复(HDR),在其内源性基因座处直接对感兴趣基因进行标签化。虽然这种方法能够产生广泛的内源性标签化基因,但HDR的效率明显低下,因此需要进行大量筛选以识别已成功标签化的期望的细胞克隆群体。此外,HDR在大量细胞类型中通常非常低效或完全无活性,尤其是在有丝分裂期后的细胞中。HDR的低效率因产生非期望的插入缺失产物而进一步复杂化,在RNA基因的情况下可能尤其成问题,因为它们可能导致产生RNA,其活性干扰正常等位基因功能。最后,研究人员通常需要筛选RNA分子内的各种标签化位点,以实现最佳性能。总之,这些缺点使HDR成为RNA中安装标签不太理想的方法。

[2456] 引导编辑是新的基因组编辑技术,其可通过将遗传信息从RNA转移到DNA来对基因组基因座进行靶向编辑。使用引导编辑,RNA基因可用各种组件,例如RNA适体、核酶或其他RNA基序标签化。与HDR策略相比,引导编辑具有在更多种类的细胞类型中更快、更便宜和更有效的潜力。因此,所描述的发明代表了用于研究健康和疾病中RNA基因的生物学中新的、有用的且非显而易见的工具。本文描述了使用引导编辑器(PE)将RNA基序插入遗传序列的新方法,所述RNA基序标签化或以其它方式操作RNA。PE能够在由CRISPR/Cas系统可靶向的期望的基因组位点处进行位点特异性插入、突变和/或缺失多个核苷酸。PE由Cas9核酸酶结构域和逆转录酶结构域之间的融合物组成。通过工程化PEgRNA(引导编辑向导RNA)将它们引导至基因组靶标,所述工程化PEgRNA包含用于DNA靶向的向导间隔区部分,以及用于编码

期望的基因组编辑的逆转录的模板(参见图28A)。设想可将PE用于插入在RNA水平上功能性的基序(以下称为RNA基序)以标签化或以其他方式操纵非编码RNA或mRNA。这些基序可用于增加基因表达、减少基因表达、改变剪接、改变转录后修饰、影响RNA的亚细胞定位、实现分离或确定RNA的细胞内或细胞外位置(使用如荧光RNA适体如Spinach、Spinach2、Baby Spinach或Broccoli),募集内源性或外源性蛋白或RNA结合剂,引入sgRNA,或诱导通过自切割或RNA酶的RNA加工(参见图28B)。由于引导编辑的灵活性,不可能提供可安装在基因组内的RNA基序的完整列表。在此显示了一系列示例,广泛地说明可用于对RNA基因进行标签化的PE安装的RNA基序的预测范围。目前不可能使用除PE之外的任何其他报告的基因组编辑方法在大多数类型的细胞(包括许多不支持HDR的细胞)中有效且相当清洁地进行这些改变。

[2457] 基因表达可能受编码3′非翻译区(UTR)的影响,所述3′非翻译区导致核转运或保留或mRNA寿命发生变化。例如,来自多瘤病毒猿猴病毒40(SV40)的polyA尾具有额外的辅助序列,其实现有效的转录终止,并可相对于其他3′UTR增加基因表达[57,58]。SV40 polyA尾的示例序列:

[2458]

SV40 POLY A尾	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGC ATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT TGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 331)

[2459] 除了聚腺苷酸化信号外,翻译后修饰信号也可由PE编码。这些包括掺入N6-甲基腺苷、N1-甲基腺苷、5-甲基胞嘧啶和假尿苷修饰的信号[59]。通过使用PE来包含在RNA转录本中写入或去除这些修饰的由酶结合的序列,有可能诱导它们的写入或清除。这可用作研究这些标志物的影响、诱导细胞分化、影响应激反应的工具,或者鉴于这些标志物的功能尚未得到充分探索,以其他方式影响靶向细胞。

[2460] PE可编码影响亚细胞定位的突变。例如,在mRNA中掺入tRNA-Lys理论上可以导致转运至线粒体[60],而各种3′UTR可导致核保留或转运[61]。

[2461] 示例:

[2462] SV40 polyA信号导致转运。

[2463]

SV40 POLY A尾	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGC ATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT TGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 331)

[2464] U1 snRNA3′框导致保留

[2465]

U1 SNRNA 3′框	TTCATTCAGCAAGTTCAGAGAAATCTGAACTTGCTGGATTTTTG GAGCAGGGAGATGGAATAGGAGCTTGCTCCGTCCACTCCACGC ATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGCACCCACTT TCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTT TTTGTTT GGTTTGTGTCTTGGTTGGCGTCTTAA (SEQ ID NO: 625)

[2466] 确定内源性RNA的亚细胞定位可以具有挑战性,标签需要添加外源性荧光标签化的核苷酸探针,如在FISH的情况下,或者耗时且可能不准确的细胞分级以及然后的RNA检测。在内源性RNA内编码探针将避免许多这些问题。一个示例是在内源性RNA中编码荧光RNA适体,例如Spinach[62]或Broccoli,从而通过添加小分子原(proto-)荧光团来可视化该RNA的存在。

[2467] Broccoli适体

[2468]

BROCCOLI 适体	GAGACGGTCGGGTCCAGATATTCGTATCTGTCGAGTAGAGTGTGGGCTC (SEQ ID NO: 357)

[2469] PE可插入或去除编码由RNA结合蛋白识别的RNA，影响RNA的稳定性、表达、定位或修饰(例如，参见列出的蛋白[63])的序列。

[2470] PE可在基因组中插入编码sgRNA的序列，作为病毒或癌症防御机制。类似地，它可用于插入microRNA(如，pre-microRNA)以指导靶向基因的沉默。

[2471] PE可插入导致RNA加工的序列，无论是通过其自身还是外部因素，作为研究RNA各个部分功能的疗法或工具。例如，HDV核酶[64]包含在RNA序列中时，导致加工紧靠核酶的5′的RNA，而锤头型核酶在核酶内的第三个茎之前切割[65]。其他自切割核酶包括手枪型[66]、斧头型[66]、发夹型[67]、Neuropora Varkud卫星型[68]、glmS[69]、twister[70]和twiner sister[66]。这些序列可能包括核酶的野生型或工程化或进化版本。这些核酶中的大多数可具有不同的序列，取决于它们与RNA相结合的区域，取决于核酶切割位点的位置。还可实现通过外部因素指导RNA加工的序列，例如序列特异性RNAse[71]、识别特定结构的RNAse[72]——如Dicer[73]和Drosha[74]。

[2472] HDV核酶

[2473]

HDV 核酶	GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC (SEQ ID NO: 365)

[2474] 实施例4的参考文献：

[2475] 以下参考文献通过引用整体并入本文。

[2476] 1.SchekN,Cooke C,Alwine JC.Molecular and Cellular Biology.1992.

[2477] 2.Gil A,Proudfoot NJ.Cell.1987.

[2478] 3.Zhao,B.S.,Roundtree,I.A.,He,C.NatRev Mol Cell Biol.2017.

[2479] 4.Rubio,M.A.T.,Hopper,A.K.Wiley Interdiscip Rev RNA 2011.

[2480] 5.Shechner,D.M.,HacisuleymanE.,Younger,S.T.,Rinn,J.L.NatMethods.2015.

[2481] 6.Paige,J.S.,Wu,K.Y.,Jaffrey,S.R.Science 2011.

[2482] 7.Ray D.,…Hughes TR.Nature 2013.

[2483] 8.Chadalavada,D.M.,Cerrone-Szakal,A.L.,Bevilacqua,P.C.RNA2007.

[2484] 9.Forster AC,Symons RH.Cell.1987.

[2485] 10.Weinberg Z,Kim PB,Chen TH,Li S,Harris KA,Lünse CE,Breaker RR.Nat.Chem.

[2486] Biol.2015.

[2487] 11.Feldstein PA,Buzayan JM,Bruening G.Gene 1989.

[2488] 12.Saville BJ,Collins RA.Cell.1990.

[2489] 13.Winkler WC,Nahvi A,Roth A,Collins JA,Breaker RR.Nature 2004.

[2490] 14.Roth A,Weinberg Z,Chen AG,Kim PG,Ames TD,Breaker RR.Nat Chem Biol.2013.

[2491] 15.Choudhury R,Tsai YS,DominguezD,Wang Y,Wang Z.Nat Commun.2012.

[2492] 16.MacRae IJ,Doudna JA.Curr Opin Struct Biol.2007.

[2493] 17.Bernstein E,Caudy AA,Hammond SM,Hannon GJNature 2001.

[2494] 18.Filippov V,Solovyev V,Filippova M,Gill SS.Gene 2000.

[2495] 实施例5-通过PE插入免疫表位

[2496] 最近在极其多种应用中探索了使用CRISPR/Cas系统的精确基因组靶向技术,包括将工程化DNA序列插入靶向基因组位点。先前,同源定向修复(HDR)已用于此应用,需要ssDNA供体模板并通过双链DNA断裂(DSB)启动修复。这种策略提供了在细胞中进行的最广泛的可能的改变,并且是将大的DNA序列插入哺乳动物细胞的唯一可用方法。然而,HDR受其起始DSB产生的非期望的细胞副作用的阻碍,例如高水平的插入缺失形成、DNA易位、大的缺失和P53激活。除了这些缺点之外,HDR还受到许多细胞类型低效率的限制(T细胞是该观测结果的明显例外)。最近为克服这些缺点所做的努力包括将人Rad51突变体与Cas9D10A切口酶(RDN)融合,从而形成无DSB的HDR系统,该系统的特征在于增加的HDR产物:插入缺失比率和较低的脱靶编辑,但仍受到细胞类型依赖性阻碍且只有适度的HDR编辑效率。

[2497] 与PEgRNA偶联的Cas9与逆转录酶的最近开发的融合物(“引导编辑器”)的融合物代表了新的基因组编辑技术,其提供了许多优于现有基因组编辑方法的优势,包括具有安装任何单核苷酸取代的能力,以及以位点特异性方式插入或缺失任何短的核苷酸段(多达至少几十个碱基)。值得注意地,PE编辑通常以较低的意外插入缺失率实现。因此,PE实现基于插入的靶向编辑应用,其以前是不可能或不能实现的。

[2498] 该特定发明描述了使用引导编辑作为将已知免疫原性表位插入内源或外源基因组DNA的手段的方法,导致用于治疗或生物技术应用的相应蛋白的修饰(参见图31和图32)。在引导编辑发明之前,此类插入只能低效地且以较高的来自DSB的插入缺失形成率实现。引导编辑在通常提供比HDR更高的效率的情况下解决了来自插入编辑的高插入缺失形成问题。这种较低的插入缺失形成率代表引导编辑作为靶向DNA插入方法优于HDR的主要优势,特别是在所述插入免疫原性表位的应用中。表位长度在几个碱基到几百个碱基的范围内。引导编辑器是在哺乳动物细胞中实现此类靶向插入的最高效且最清洁的技术。

[2499] 本发明的关键概念是使用引导编辑器将包含先前所述的免疫原性表位的核苷酸序列插入内源或外源基因组DNA中,以下调和/或破坏它们的蛋白产物和/或表达细胞类型。以产生靶向基因的编码蛋白和插入的免疫原性表位的相应蛋白翻译的融合蛋白的方式将用于免疫原性表位插入的核苷酸序列靶向到基因。由于针对如破伤风、白喉或麻疹的常规疫苗接种的标准预先免疫,患者的免疫系统之前已接受过训练以识别这些表位。由于融合表位的免疫原性,预计患者的免疫系统会识别和禁用引导编辑的蛋白(不仅仅是插入的表位)以及可能表达它的细胞。

[2500] 与靶向基因的融合物将根据需要工程化改造以确保暴露插入的表位蛋白翻译进行免疫系统识别。这可包括导致蛋白翻译的靶向核苷酸插入,产生免疫原性表位与靶向基因的C端融合物,免疫原性表位与靶向基因的N端融合物,或核苷酸插入基因中以便在蛋白结构的表面暴露区域内编码免疫原性表位。

[2501] 作为在靶基因序列和插入的免疫原性表位核苷酸序列之间插入的核苷酸编码的蛋白接头可需要经工程化改造为本发明的一部分以促进靶向基因的免疫系统识别、细胞运输、蛋白功能或蛋白折叠。这些插入的核苷酸编码的蛋白接头可包括(但不限于)XTEN接头的可变长度和序列或甘氨酸-丝氨酸接头的可变长度和序列。这些工程化接头以前曾被用于成功促进蛋白融合。

[2502] 本发明的区别特征包括使用对特定氨基酸序列的先前获得性免疫反应作为诱导对其他情况下非免疫原性蛋白的免疫反应的手段的能力。另一个区别特征是能够以靶标方式插入这些免疫原性表位的核苷酸序列，该靶标方式不诱导高水平的作为副产物编辑的不需要的插入缺失，并且在其插入中有效。在此所讨论的本发明还能够结合细胞类型特异性递送方法(如，AAV血清型)以插入在感兴趣的细胞类型中触发免疫反应的表位。

[2503] 可将引导编辑用作将免疫原性表位插入致病基因的手段用于对患者的免疫系统进行编程以对抗极其多种疾病(不限于癌症，如用免疫肿瘤学策略一样)。该技术的直接相关应用将是作为癌症治疗剂，因为它可通过引起对相关致癌基因如HER2或生长因子如EGFR的免疫反应来破坏肿瘤的免疫逃逸机制。这种方法似乎类似于T细胞工程，但这种方法的一个新进展是它可用于许多细胞类型和癌症以外的疾病，而无需生成工程化T细胞并将其引入患者。

[2504] 使用PE将大多数人已针对性(破伤风、百日咳、白喉、麻疹、腮腺炎、风疹等)接种疫苗的免疫原性表位插入驱动疾病的外源或内源基因中，因此患者的免疫系统掌握禁用该蛋白。

[2505] 从上述策略中获得潜在治疗益处的疾病包括由有毒蛋白聚集引起的疾病，例如致命性家族性失眠症。其他可能受益的疾病包括由其它情况下无毒的内源性蛋白的致病性过表达引起的疾病，以及由外源病原体引起的疾病。

[2506] 主要治疗的适应症包括上述那些，例如癌症、朊病毒和其他神经退行性疾病、传染病的治疗剂和预防药物。次要治疗的适应症可包括对迟发性遗传病患者的预防性护理。预期当前的护理药物标准可与某些疾病的引导编辑结合使用，例如特别具有侵袭性的癌症，或在药物有助于缓解疾病症状直到疾病完全治愈的情况下。

[2507] 以下是可使用引导编辑器插入来实现的免疫原性表位的示例：

[2508]

疫苗	疾病	表位氨基酸序列	示例核酸序列 (8)
1	破伤风类毒素	QYIKANSKFIGITEL (SEQ ID NO: 396)	CATGATATAAAAGCAAATTCTAAATTTATAGGTATAACTGAACTA (SEQ ID NO: 397)

[2509]

2	白喉毒素突变体 CRM197	GADDVVDSSKSF VMENFSSYHGTK PGYVDSIQKGIQK PKSGTQGNYDDD WKEFYSTDNKYD AAGYSVDNENPL SGKAGGVVKVTY PGLTKVLALKVD NAETIKKELGLSL TEPLMEQVGTEEF IKRFGDGASRVVL SLPFAEGSSSVEYI NNWEQAKALSVE LEINFETRGKRGQ DAMYEYMAQAC AGNRVRRSVGSS LSCINLDWDVIRD KTKTKIESLKEHG PIKNKMSESPNKT VSEEKAKQYLEEF HQTALEHPELSEL KTVTGTNPVFAG ANYAAWAVNVA QVIDSETADNLEK TTAALSILPGIGSV MGIADGAVHHNT EEIVAQSIALSSL MVAQAIPLVGEL VDIGFAAYNFVES IINLFQVVHNSYN RPAYSPGHKTQPF LHDGYAVSWNTV EDSIIRTGFQGESG HDIKITAENTPLPI AGVLLPTIPGKLD VNKSKTHISVNGR KIRMRCRAIDGD VTFCRPKSPVYVG NGVHANLHVAFH RSSSEKIHSNEISS DSIGVLGYQKTV DHTKVNSKLSLFF EIKS (SEQ ID NO: 630)	GGCGCCGACGACGTGGTGGACAGCA GCAAGAGCTTCGTGATGGAGAACTT CAGCAGCTACCACGGCACCAAGCCC GGCTACGTGGACAGCATCCAGAAGG GCATCCAGAAGCCCAAGAGCGGCAC CCAGGGCAACTACGACGACGACTGG AAGGAGTTCTACAGCACCGACAACA AGTACGACGCCGCCGGCTACAGCGT GGACAACGAGAACCCCCTGAGCGGC AAGGCCGGCGGCGTGGTGAAGGTGA CCTACCCCGGCCTGACCAAGGTGCT GGCCCTGAAGGTGGACAACGCCGAG ACCATCAAGAAGGAGCTGGGCCTGA GCCTGACCGAGCCCCTGATGGAGCA GGTGGGCACCGAGGAGTTCATCAAG AGGTTCGGCGACGGCGCCAGCAGGG TGGTGCTGAGCCTGCCCTTCGCCGA GGGCAGCAGCAGCGTGGAGTACATC AACAACTGGGAGCAGGCCAAGGCCC TGAGCGTGGAGCTGGAGATCAACTT CGAGACCAGGGGCAAGAGGGGCCA GGACGCCATGTACGAGTACATGGCC CAGGCCTGCGCCGGCAACAGGGTGA GGAGGAGCGTGGGCAGCAGCCTGA GCTGCATCAACCTGGACTGGGACGT GATCAGGGACAAGACCAAGACCAA GATCGAGAGCCTGAAGGAGCACGGC CCCATCAAGAACAAGATGAGCGAGA GCCCCAACAAGACCGTGAGCGAGGA GAAGGCCAAGCAGTACCTGGAGGA GTTCCACCAGACCGCCCTGGAGCAC CCCGAGCTGAGCGAGCTGAAGACCG TGACCGGCACCAACCCCGTGTTCGC CGGCGCCAACTACGCCGCCTGGGCC GTGAACGTGGCCCAGGTGATCGACA GCGAGACCGCCGACAACCTGGAGAA GACCACCGCCGCCCTGAGCATCCTG CCCGGCATCGGCAGCGTGATGGGCA TCGCCGACGGCGCCGTGCACCACAA CACCGAGGAGATCGTGGCCCAGAGC ATCGCCCTGAGCAGCCTGATGGTGG CCCAGGCCATCCCCCTGGTGGGCGA GCTGGTGGACATCGGCTTCGCCGCC TACAACTTCGTGGAGAGCATCATCA ACCTGTTCCAGGTGGTGCACAACAG CTACAACAGGCCCGCCTACAGCCCC GGCCACAAGACCCAGCCCTTCCTGC ACGACGGCTACGCCGTGAGCTGGAA CACCGTGGAGGACAGCATCATCAGG ACCGGCTTCCAGGGCGAGAGCGGCC ACGACATCAAGATCACCGCCGAGAA

[2510]

			CACCCCCCTGCCCATCGCCGGCGTG CTGCTGCCCACCATCCCCGGCAAGC TGGACGTGAACAAGAGCAAGACCCA CATCAGCGTGAACGGCAGGAAGATC AGGATGAGGTGCAGGGCCATCGACG GCGACGTGACCTTCTGCAGGCCCAA GAGCCCCGTGTACGTGGGCAACGGC GTGCACGCCAACCTGCACGTGGCCT TCCACAGGAGCAGCAGCGAGAAGAT CCACAGCAACGAGATCAGCAGCGAC AGCATCGGCGTGCTGGGCTACCAGA AGACCGTGGACCACACCAAGGTGAA CAGCAAGCTGAGCCTGTTCTTCGAG ATCAAGAGC (SEQ ID NO: 399)
3	流行性腮腺炎	GTYRLIPNARANLTA (SEQ ID NO: 400)	GGCACCTACAGGCTGATCCCCAACG CCAGGGCCAACCTGACCGCC (SEQ ID NO: 401)
4	流行性腮腺炎	PSKFFTISDSATFAPGPVSNA (SEQ ID NO: 402) PSKLFIMLDNATFAPGPVVNA (SEQ ID NO: 404)	Ccgagcaaatttttaccattagcgatagcgcgacctttgcg ccgggcccggtgagcaacgcg (SEQ ID NO: 403) Ccgagcaaactgtttattatgctggataacgcgacctttgcg ccgggcccggtggtgaacgcg (SEQ ID NO: 405) Selected examples from Hemagglutinin–neuraminidase (HN) diversity among outbreak strains (table1) Divergence between vaccine strain JL5 and outbreak strains (table2)
5	风疹病毒(RV)	TPPPYQVSCGGESDRASARVIDPAAQS (SEQ ID NO: 406)	ACCCCCCCCCCCTACCAGGTGAGCT GCGGCGGCGAGAGCGACAGGGCCA GCGCCAGGGTGATCGACCCCGCCGC CCAGAGC (SEQ ID NO: 407)

[2511]

6	血凝素	PEYAYKIVKNKK MEDGFLQGMVD GWYGHHSNEQGS GLMENERTLDKA NPNNDLCSWSDH EASSNNTNQEDLL QRESRRKKRIGTS TLNQRGNCNTKC QTEEARLKREEVS LVKSDQCSNGSL QCRANNSTEQVD (SEQ ID NO: 408)	CCCGAGTACGCCTACAAGATCGTGA AGAACAAGAAGATGGAGGACGGCT TCCTGCAGGGCATGGTGGACGGCTG GTACGGCCACCACAGCAACGAGCAG GGCAGCGGCCTGATGGAGAACGAG AGGACCCTGGACAAGGCCAACCCCA ACAACGACCTGTGCAGCTGGAGCGA CCACGAGGCCAGCAGCAACAACACC AACCAGGAGGACCTGCTGCAGAGGG AGAGCAGGAGGAAGAAGAGGATCG GCACCAGCACCCTGAACCAGAGGGG CAACTGCAACACCAAGTGCCAGACC GAGGAGGCCAGGCTGAAGAGGGAG GAGGTGAGCCTGGTGAAGAGCGACC AGTGCAGCAACGGCAGCCTGCAGTG CAGGGCCAACAACAGCACCGAGCA GGTGGAC (SEQ ID NO: 409)
7	神经氨酸苷酶	TKSTNSRSGGISG PDNEAPVGEAPSP YGDNPRPNDGNN IRIGSKGYNGIITD TIEESCSCYPDAK VVKSVELDSTIWT SGSSPNQKIITIGW DPNGWTGTPMSP NGAYGTDGPSNG QANQHQAESISA GNSSLCPIRDNWH GSNRSWSWPDGA E (SEQ ID NO: 410)	ACCAAGAGCACCAACAGCAGGAGC GGCGGCATCAGCGGCCCCGACAACG AGGCCCCCGTGGGCGAGGCCCCCAG CCCCTACGGCGACAACCCCAGGCCC AACGACGGCAACAACATCAGGATCG GCAGCAAGGGCTACAACGGCATCAT CACCGACACCATCGAGGAGAGCTGC AGCTGCTACCCCGACGCCAAGGTGG TGAAGAGCGTGGAGCTGGACAGCAC CATCTGGACCAGCGGCAGCAGCCCC AACCAGAAGATCATCACCATCGGCT GGGACCCCAACGGCTGGACCGGCAC CCCCATGAGCCCCAACGGCGCCTAC GGCACCGACGGCCCCAGCAACGGCC AGGCCAACCAGCACCAGGCCGAGA GCATCAGCGCCGGCAACAGCAGCCT GTGCCCCATCAGGGACAACTGGCAC GGCAGCAACAGGAGCTGGAGCTGGC CCGACGGCGCCGAG (SEQ ID NO: 411)
8	H5N1 病毒血凝素上的 TAP (转运抗原呈递)	EKIVLLLAMMEKI VLLLAKCQTPMG AIKAVDGVTNKC PYLGSPSF (SEQ ID NO: 412)	GAGAAGATCGTGCTGCTGCTGGCCA TGATGGAGAAGATCGTGCTGCTGCT GGCCAAGTGCCAGACCCCCATGGGC GCCATCAAGGCCGTGGACGGCGTGA CCAACAAGTGCCCCTACCTGGGCAG CCCCAGCTTC (SEQ ID NO: 413)
9	h5n1 病毒神经氨酸苷酶上的 TAP (转运抗原呈递)	IRPCFWVELNPNQ KIITIRPCFWVELI CYPDAGEIT (SEQ ID NO: 414)	ATCAGGCCCTGCTTCTGGGTGGAGC TGAACCCCAACCAGAAGATCATCAC CATCAGGCCCTGCTTCTGGGTGGAG CTGATCTGCTACCCCGACGCCGGCG AGATCACC (SEQ ID NO: 415)

[2512]

10	血凝素表位针对 HLA I 类	MEKIVLLLAEKIV LLLAMCPYLGSPS FKCQTPMGAIKA VDGVTNK (SEQ ID NO: 416)	ATGGAGAAGATCGTGCTGCTGCTGG CCGAGAAGATCGTGCTGCTGCTGGC CATGTGCCCCTACCTGGGCAGCCCC AGCTTCAAGTGCCAGACCCCCATGG GCGCCATCAAGGCCGTGGACGGCGT GACCAACAAG (SEQ ID NO: 417)
11	神经氨酸苷酶表位针对 HLA I 类	NPNQKIITICYPDAG EITIRPCFWVELRPC FWVELI (SEQ ID NO: 418)	AACCCCAACCAGAAGATCATCACCAT CTGCTACCCCGACGCCGGCGAGATCA CCATCAGGCCCTGCTTCTGGGTGGAG CTGAGGCCCTGCTTCTGGGTGGAGCT GATC (SEQ ID NO: 419)
12	血凝素表位针对 HLA II 类	MVSLVKSDQIGTST LNQR (SEQ ID NO: 420)	ATGGTGAGCCTGGTGAAGAGCGACCA GATCGGCACCAGCACCCTGAACCAGA GG (SEQ ID NO: 421)
13	神经氨酸苷酶表位针对 HLA II 类	YNGIITDTI (SEQ ID NO: 422)	TACAACGGCATCATCACCGACACCAT C (SEQ ID NO: 423)
14	血凝素表位 H5N1 结合的 HLA I 类和 II 类	MEKIVLLLAEKIVL LLAMMVSLVKSDQ CPYLGSPSFIGTSTL NQRKCQTPMGAIK AVDGVTNK (SEQ ID NO: 424)	ATGGAGAAGATCGTGCTGCTGCTGGC CGAGAAGATCGTGCTGCTGCTGGCCA TGATGGTGAGCCTGGTGAAGAGCGAC CAGTGCCCCTACCTGGGCAGCCCCAG CTTCATCGGCACCAGCACCCTGAACC AGAGG (SEQ ID NO: 425)
15	神经氨酸苷酶表位 H5N1 结合的 HLA I 类和 II 类	NPNQKIITIYNGIIT DTICYPDAGEITIRP CFWVELRPCFWVE LI (SEQ ID NO: 426)	AACCCCAACCAGAAGATCATCACCAT CTACAACGGCATCATCACCGACACCA TCTGCTACCCCGACGCCGGCGAGATC ACCATCAGGCCCTGCTTCTGGGTGGA GCTGAGGCCCTGCTTCTGGGTGGAGC TGATC (SEQ ID NO: 427)

[2513] 以下是可整合入靶基因进行免疫表位标签化的表位的其他示例：

[2514] 实施例5中引用的参考文献

[2515] 以下参考文献通过引用以其整体并入：

[2516] 1.X.Wen,K.Wen,D.Cao,G.Li,R.W.Jones,J.Li,S.Szu,Y.Hoshino,L.Yuan, Inclusion of a universal tetanus toxoid CD4(+)T cell epitope P2 significantly enhanced the immunogenicity of recombinant rotavirus ΔVP8* subunit parenteral vaccines.Vaccine 32,4420-4427(2014).

[2517] 2.G.Ada,D.Isaacs,Carbohydrate-protein conjugate vaccines.Clin Microbiol Infect 9,79-85(2003).

[2518] 3.E.Malito,B.Bursulaya,C.Chen,P.L.Surdo,M.Picchianti,E.Balducci, M.Biancucci,A.Brock,F.Berti,M.J.Bottomley,M.Nissum,P.Costantino,R.Rappuoli,

G.Spraggon,Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197.Proceedings of the National Academy of Sciences 109,5229(2012).

[2519] 4.J.de Wit,M.E.Emmelot,M.C.M.Poelen,J.Lanfermeijer,W.G.H.Han,C.van Els,P.Kaaijk,The Human CD4(+)T Cell Response against Mumps Virus Targets a Broadly Recognized Nucleoprotein Epitope.J Virol 93,(2019).

[2520] 5.M.May,C.A.Rieder,R.J.Rowe,Emergent lineages of mumps virus suggest the need for apolyvalent vaccine.Int JInfect Dis 66,1-4(2018).

[2521] 6.M.Ramamurthy,P.Rajendiran,N.Saravanan,S.Sankar,S.Gopalan,B.Nandagopal,Identification of immunogenic B-cell epitope peptides of rubella virus E1 glycoprotein towards development of highly specific immunoassays and/or vaccine.Conference Abstract,(2019).

[2522] 7.U.S.F.Tambunan,F.R.P.Sipahutar,A.A.Parikesit,D.Kerami,Vaccine Design for H5N1 Based on B-and T-cell Epitope Predictions.Bioinform Biol Insights 10,27-35(2016).

[2523] 实施例6-使用PE在体内实现化学剂诱导的靶蛋白二聚化

[2524] 本文所述的引导编辑器还可用于将二聚化诱导的生物过程如受体信号传导置于方便的小分子药物的控制下,通过用本文描述的引导编辑对编码小分子结合蛋白进行基因组整合进行。使用本文所述的引导编辑器,在活细胞或患者中可将编码小分子结合蛋白的基因序列插入编码感兴趣的靶蛋白的基因内。单独的此编辑应当没有生理影响。在施用小分子药物(通常是二聚的小分子,其可同时结合两个药物结合蛋白结构域,每个结构域都与靶蛋白的一个拷贝融合)后,小分子诱导靶向蛋白的二聚化。然后,该靶蛋白二聚化事件诱导生物信号传导事件,如红细胞生成或胰岛素信号传导。

[2525] 实施例7-用双重引导编辑进行的DNA序列插入、缺失和替换

[2526] 本实施例描述了使用多瓣引导编辑,特别是双重引导编辑,用于新DNA序列的精确插入、内源性基因组DNA序列的精确缺失,或用新DNA序列替换内源性基因组DNA序列。参见图90。本发明解决了基因组工程中长期存在的目标:在活细胞的基因组中的DNA序列的可编程插入、缺失或替换。这项技术可以应用于修饰或编辑任何活细胞的基因组,用于研究、治疗或工业应用。

[2527] 先前,核酸酶诱导的双链DNA断裂(DSB)的同源定向修复(HDR)已被用于插入、删除或替换活细胞基因组中的DNA序列。虽然用途广泛,但HDR的低效率限制了它在许多哺乳动物细胞类型中的应用,特别是治疗相关的人细胞类型[1-4]。此外,DSB的生成导致易错末端连接机制的优先修复,该机制产生插入和缺失(插入缺失)修复产物,其不能随后转化为期望的经编辑DNA序列[5-10]。因此,当核酸酶在哺乳动物细胞中使用时,缩合体通常代表了大多数修饰的基因组DNA产物。此外,由核酸酶产生的DSBs可导致不希望发生的基因组变化,如大面积缺失和染色体易位[11],以及p53的激活[12,13]。结构变异体(SV)也可以使用可编程核酸酶产生或校正,而不依赖同源重组,尽管基于核酸酶的策略在这些应用中也存在类似的缺点。例如,双切割策略可用于逆转靶基因组DNA序列,但大多数产物的等位基因在切割点之间含有大量的缺失,而不是所需的逆转序列,或在各个切割点上含有缩进。

[2528] 引导编辑是最近报道的基因组编辑技术,其能够进行基因组DNA序列的插入、缺失

或替换，而无需易错双链DNA断裂[14]。引导编辑使用与工程化引导编辑向导RNA(pegRNA)配对的工程化Cas9切口酶-逆转录酶融合蛋白(PE1或PE2)，该pegRNA将Cas9引导到靶基因组位点并且编码用于安装期望编辑的信息。引导编辑通过多步骤的编辑过程进行：1)Cas9结构域结合并使靶基因组DNA位点生成切口，该位点由pegRNA的间隔区序列指定；2)逆转录酶结构域使用带切口的基因组DNA作为引物，以使用pegRNA上的工程化延伸作为逆转录的模板启动编辑DNA链的合成——这生成含有经编辑的DNA序列的单链3’瓣；3)细胞DNA修复通过5’瓣种类的置换来分解3’瓣中间体，5’瓣种类的置换经由经编辑的3’瓣的侵入，含有原始DNA序列的5’瓣的切除，以及连接新3’瓣以掺入经编辑的DNA链而发生，形成一条经编辑的链和一条未经编辑的链的异源双链体；以及4)细胞DNA修复使用经编辑的链作为修复模板来替换异源双链体中的未经编辑的链，完成编辑过程。

[2529] 期望编辑的高效结合要求新合成的3’瓣含有与基因组DNA位点同源的序列的一部分。这种同源性使经编辑的3’瓣能够与内源性DNA链(相应的5’瓣)竞争，以掺入DNA双链中。由于经编辑的3’瓣将比内源性5’瓣含有更少序列同源性，预计这种竞争将有利于5’瓣链。因此，含有编辑的3’瓣可能无法有效地侵入和置换5’瓣链。此外，成功的3’瓣侵入和5’瓣的去除仅包含双链DNA基因组的一条链上的编辑。编辑的永久安装需要细胞DNA修复，以使用经编辑的链作为模板替换未经编辑的互补DNA链。虽然通过使用次级sgRNA(PE3系统)可以在与编辑相邻的未经编辑链上引入切口，使细胞倾向于替换未经编辑的链而不是经编辑的链(上述步骤4)，这个过程仍然依赖于DNA修复的第二阶段。这些DNA修复步骤对于需要平衡长5’和3’瓣中间体或含有长的非同源区，诸如长的插入或长的缺失的编辑可能特别低效。

[2530] 本实施例描述了双重引导编辑系统(或双重瓣引导编辑系统)，该系统通过同时编辑两条DNA链来解决与瓣平衡和随后将编辑掺入未经编辑的互补基因组DNA链相关联的挑战。在双重瓣引导编辑系统中，使用两个pegRNA来靶向基因组位点的相反链，并指导合成两个互补的含有经编辑的DNA序列的3’瓣(图91)。与经典的引导编辑不同，不要求这对经编辑的DNA链(3’瓣)与内源性基因组DNA中的5’瓣直接竞争，因为互补的经编辑的链反而可用于杂交。由于双联体的两条链作为经编辑的DNA合成，双重瓣引导编辑系统避免了经典引导编辑所需的未经编辑的互补DNA链的替换。相反，细胞DNA修复机制仅需要切除成对的5’瓣(原始基因组DNA)，并将成对的3’瓣(经编辑的DNA)连接到基因座上。因此，也不需要在新合成的DNA链中包括与基因组DNA同源的序列，允许新链的选择性杂交并促进含有最小基因组同源性的编辑。还可以使用切割DNA的两条链的引导编辑的核酸酶活性型式以加速去除原始DNA序列。

[2531] 根据双重引导编辑生成的交错的DNA切口的方向，系统可以用新的期望序列替换两个切口之间的序列(5’突出端交错方向)，或者插入新DNA序列并伴随靶位点复制(3’突出端交错方向)。图90显示了其中切口之间的序列经替换的方向(5’置换的切口)。值得注意的是，新的3’瓣不需要沿着新的DNA序列的长度完全重叠，而是可以在新序列的仅一部分上互补。用这种方式，每个瓣可以比新DNA序列的全长更短。

[2532] HEK293位点3基因座处的双重瓣引导编辑

[2533] 使用人细胞(HEK293T)中HEK293位点3基因座处的双重瓣引导编辑来实现异源DNA序列的高效引入。例如，用编码6xHis标签的新的22bp序列对该基因座中的90bp序列的替换以高于80％的效率发生(图92等位基因表)。类似地，用GFP-11肽序列(52bp)替换这个相同

的90bp序列实现了大于60％的产率。GFP-11标签插入也在MYC基因的C端以>20％的效率进行，并伴随Myc3'UTR序列的一部分的程序化删除发生。

[2534] 双重瓣引导编辑pegRNA

[2535] 用于双重瓣引导编辑的pegRNA的一般设计如图93所示，用于双重瓣引导编辑的pegRNA具有与用于经典引导编辑的pegRNA类似的设计，然而不需要RT模板区来编码与靶基因座的任何同源性。相反，两个pegRNA可以含有编码3'瓣合成的RT模板，其3'末端是彼此的反向互补序列。3'瓣之间的这种互补性促进了它们的退火和用预期的新DNA序列替换内源性DNA序列。这要求两个pegRNA中的RT模板的5'区是彼此的反向互补序列，并且这种互补性的量可以不同(图93)。

[2536] 使用双重瓣引导编辑来安装重组位点

[2537] 双重瓣引导编辑具有许多潜在的应用，诸如安装肽标签、RNA标签、免疫表位、二聚化结构域和重组酶靶位点。

[2538] 这种应用之一是在用户指定的基因组位置安装重组酶或整合酶序列。图94例示了将Bxb1重组酶attB(38bp)和attP(50bp)位点安装到人基因组的靶向区(HEK293T位点3，或HEK3)，同时发生两个切口位点之间90bp的间插序列的缺失。3'瓣之间不同程度的互补性允许成功的编辑，尽管较长的互补性序列产生更有利的期望编辑和插入缺失的比率。其他双重瓣引导编辑的应用包括用肽或蛋白质序列对基因进行内源性标签化，或用新的DNA序列替换外显子，这些新的DNA序列具有取代其突变落在外显子序列内的多种变体的潜力。

[2539] 可以使用双重瓣引导编辑以在人基因组的靶向位置引入一个或两个重组酶位点。如果插入单个重组酶位点，这些位点可以用作基因组重组酶位点与外源提供的DNA(诸如质粒)内的第二重组酶位点之间的重组酶介导的反应的着陆点。这使得DNA货物能够靶向整合。如果两个重组酶位点插入DNA的相邻区域，根据重组酶位点的方向，这些位点可以用于重组酶介导的间插序列的切除或倒位，或用于重组酶介导的与外源DNA进行盒交换用于货物整合。在不同的染色体上整合相容的重组酶位点能够实现靶向且定向的染色体易位。可以使用双重瓣引导编辑以在人基因组的一些基因座处高效引入重组酶位点(图94)。因此，通过双重瓣引导编辑的重组酶位点整合与DNA重组酶的配对，代表了实现多种类型的SV编辑和DNA货物的靶向整合的强大方法。

[2540] 实施例8-使用四重瓣引导编辑进行的DNA序列插入、缺失、倒位、易位和整合

[2541] 本实施例描述了使用多瓣引导编辑，特别是四重瓣引导编辑，用于内源性基因组DNA序列的靶向插入、缺失、复制和替换，以及用于靶向基因组DNA序列倒位、靶向染色体易位和外源性DNA的靶向整合。这项技术有潜力解决基因组工程中长期存在的目标：结构变体的可编程安装或校正(一般定义为>50bp的基因组序列变化，通常表现为大的缺失、插入/复制、倒位和易位)，以及在人基因组的指定位置处的DNA货物的靶向整合。后者能够实现基因扩增(augmentation)治疗策略和其他生物技术应用。这项技术可能潜在用于修饰或编辑任何活细胞的基因组，用于研究、治疗或工业应用。

[2542] 本发明描述了多瓣引导编辑系统，该系统通过同时编辑两条DNA链来解决与瓣平衡和随后将编辑掺入未经编辑的互补基因组DNA链相关联的挑战。其还通过允许新合成的3'瓣指导更大的序列重排，扩展了传统引导编辑的能力。

[2543] 开发了四重瓣引导编辑方法，其可以直接进行期望的序列改变，而不需要重组酶

(如在双重瓣引导编辑中),但具有类似的多功能性。在四重瓣引导编辑中,四个不同的pegRNA序列与PE2编辑器一起递送到细胞中。一对pegRNA是合成互补DNA瓣的模板,与双重瓣引导编辑一样,而另一对pegRNA是合成互补DNA瓣的正交对的模板。这些瓣的位置、它们的相对方向和互补性的配对指示哪一类重排将发生。重排的DNA的接合含有由pegRNA模板化3'瓣编码的序列,其指导重排的方向并可以充当下游过程的有用的DNA序列。靶向基因组序列倒位是用四重瓣引导编辑策略进行的(图95)。在使用将Bxb1 attB和attP序列置于倒位接合处的四个pegRNA,实现了HEK293T细胞中的AAVS1基因座处2.7kb序列的倒位。使用扩增子测序对预期接合进行验证,这证明在≥90%的测序读段中含有attB或attP位点序列的正确接合(图96)。

[2544] 外源DNA质粒的靶向整合也用四重瓣引导编辑实现。在这种方法中,两个pegRNA靶向内源性基因组DNA位点,同时两个pegRNA靶向货物质粒上的区域。在基因组DNA和质粒DNA两者上合成3'DNA瓣,使得质粒和基因组DNA将由杂交的3'瓣桥接(图97A)。5'突出端的切除和切口的连接导致整合的质粒产物。整合的方向可以通过pegRNA模板序列的配对来控制。

[2545] 使用四重瓣引导编辑,在HEK293T细胞中的AAVS1基因座上整合编码表达H2B-GFP融合蛋白的质粒。这导致1.6%的细胞获得持久的、核定位的GFP荧光,如通过流式细胞术所评估。预测接合的PCR显示预期的产物含有pegRNA模板化BxB1 attP序列(图97B)。

[2546] 还使用四重瓣引导编辑(图98A)进行染色体易位,以指导染色体8上的MYC基因座和染色体19上的TIMM44基因座之间的染色体易位(图98B)。

[2547] 用12对pegRNA和PE2进行在IDS基因座处的双重瓣引导编辑。HEK293T细胞用PE2和不同的pegRNA对(例如,第一列中的pegRNAA1_a和pegRNAB2_a与用于正向安装attP位点的模板)进行转染。通过HTS测量效率。该数据显示,双重瓣引导编辑可以成功地以高达约80%的效率将感兴趣的序列插入IDS基因座(图99)。

[2548] 还在293T细胞中的AAVS1基因座处进行双重瓣介导的复制(图100A)。使用双重瓣pegRNA与PE2来诱导AAVS1处的遗传序列的复制(图100B)。

[2549] 还进行多瓣引导编辑诱导的MYC-CCR5的易位(图101)。MYC-CCR5的易位由四瓣pegRNA和PE2诱导。MYC-CCR5易位事件由四重pegRNA诱导。在HEK293T细胞中测试四组不同的pegRNA。衍生的chr8和chr3之间的易位接合产物由接合引物扩增。图中显示了与预期接合等位基因比对的读段的百分比。这表明四重瓣引导编辑可以介导MYC和CCR5基因的易位,其产物纯度在接合1处接近100%并在接合2处为约50%。代表性等位基因图显示了在接合1处与预期等位基因序列比对的序列。

[2550] 还在其他人细胞系中进行双重瓣和四重瓣引导编辑(图102A-102B)。在四个不同的人细胞系中进行双重瓣引导编辑(图102A)。HEK293T和HeLa细胞用双重pegRNA和PE2进行转染,用于编辑三个不同的基因组基因座(IDS、MYC和TIMM44)。U2OS和K562细胞用同样的组分进行核转染。双重瓣引导编辑靶向基因座处显示在所有四种人细胞间的稳健编辑效率,特别是在HEK293T和K562细胞。中,实现双重瓣引导编辑的细胞机制在许多人细胞类型间保守。在HeLa细胞中的AAVS1基因座处也进行2.7kb序列的多瓣(四重pegRNA)定向倒位(图102B)。

[2551] 倒位效率也通过在CCR5基因座处的HTS来测量(图103A-103B)。通过HTS测量预期倒位编辑等位基因的百分比。在HEK293T细胞中分别转染PE2的四组四重pegRNA。进行在

HEK293T细胞中CCR5基因座处的双重瓣介导的序列复制(约100nt)(图103A)。经由HTS 300周期对末端测序分析,编辑效率达到约1.5%。也在HeLa细胞中的CCR5基因座处进行四重瓣介导的序列倒位(约95-117nt)(图103B)。经由HTS 300周期对末端测序分析,编辑效率达到约1.2%。这一结果表明,多瓣引导编辑可以精确地在CCR5基因座处成功地介导复制和倒位。当复制靶向序列时,编辑特异性很高(插入缺失百分比<2%)。

[2552] 接下来,用表达Exo1、Fen1、红色荧光蛋白(对照)、DNA2、Mlh1 neg和P53抑制剂的质粒与pegRNA和PE2转染HEK293T细胞(图104A-104D)。编辑效率通过HTS测量。在候选物和RFP对照之间比较编辑效率。也用针对每个靶基因座的siRNA质粒和pegRNA及PE2转染HEK293T细胞(图104E)。非靶向性siRNA(siNT)用作对照。通过HTS测量编辑效率。每个siRNA敲除和每个靶基因座处的siNT ctrl之间比较编辑效率。用双重egRNA、PE2和分别表达Exo1、Fen1、红色荧光蛋白(ctrl)、DNA2、Mlh1 neg和P53抑制剂的质粒转染HEK293T细胞。通过HTS测量编辑效率。在候选物和RFP对照之间比较编辑效率。发现FEN1的过量表达改善所有四个靶向基因座(MYC、TIMM44、IDS、CCR5)中的双重瓣引导编辑效率。

[2553] 双重瓣介导的序列复制也在AAVS1基因座处进行(图105A)。通过使用生成两个独特的3'瓣结构的双重pegRNA,在293T细胞中的AAVS1基因座处诱导约300bp的序列复制(图105B)。预期的等位基因用特定的引物扩增并经受HTS。约94%的读段与具有复制的预期等位基因比对。在未经处理的样品中没有观察到重复产物。

[2554] 靶向IDS基因组序列倒位也用四重瓣引导编辑进行(图106)。约13%亨特氏综合征患者已显示具有IDS基因序列的倒位(Bondeson et al.,Human Molecular Genetics,1995)。应用四重瓣引导编辑来诱导HEK293T细胞中的约40kb IDS基因组序列的这种致病性倒位。通过用pegRNA和PE2转染HEK293T细胞,测试六组四重pegRNA。使用引物来特异性地扩增接合"ab"和接合"cd"处的倒位序列。在具有IDS_QF1的两个接合处观察到约95%的预期倒位等位基因序列。其他组的pegRNA也在两个接合处产生高百分比的预期等位基因序列。在未经处理的样品中没有观察到倒位接合产物。

[2555] 还发现PegRNA 3'基序修饰改善IDS基因座处的双重瓣引导编辑效率(图107)。为了进一步改善双重瓣引导编辑效率,引入假结(pseudoknot)evoPreQ1基序来保护pegRNA3'末端。在比较由未修饰和evoPreQ1修饰的双重pegRNA生成的编辑效率时,在靶IDS基因座处,修饰的pegRNA的编辑效率有总体改善。双重瓣引导编辑效率的改善可以达到5.3倍。

[2556] 参考文献(实施例7和实施例8):

[2557] 1.Rouet,P.,Smih,F.&Jasin,M.Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.PNAS 91,6064-6068(1994).

[2558] 2.Heyer,W.-D.,Ehmsen,K.T.& Liu,J.Regulation of homologous recombination in eukaryotes.Annu.Rev.Genet.44,113-139(2010).

[2559] 3.Chapman,J.R.,Taylor,M.R.G.&Boulton,S.J.Playing the end game:DNA double-strand break repair pathway choice.Mol.Cell47,497-510(2012).

[2560] 4.Cox,D.B.T.,Platt,R.J.& Zhang,F.Therapeutic genome editing:prospects and challenges.Nature Medicine 21,121-131(2015).

[2561] 5.Rouet,P.,Smih,F.&Jasin,M.Introduction of double-strand breaks into

the genome of mouse cells by expression of a rare-cutting endonuclease.Mol Cell Biol 14,8096-8106(1994).

[2562] 6.Bibikova,M.,Golic,M.,Golic,K.G.&Carroll,D.Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases.Genetics 161,1169-1175(2002).

[2563] 7.Urnov,F.D.,Rebar,E.J.,Holmes,M.C.,Zhang,H.S.&Gregory,P.D.Genome editing with engineered zinc finger nucleases.Nat Rev Genet 11,636-646(2010).

[2564] 8.Mali,P.et al.RNA-guided human genome engineering via Cas9.Science339,823-826(2013).

[2565] 9.Cong,L.etal.Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823(2013).

[2566] 10.Jinek,M.et al.RNA-programmed genome editing in human cells.eLife2, e00471(2013).

[2567] 11.Kosicki,M.,Tomberg,K.&Bradley,A.Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complexrearrangements.Nature Biotechnology 36,765-771(2018).

[2568] 12.Ihry,R.J.et al.p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells.Nature Medicine24,939-946(2018).

[2569] 13.Haapaniemi,E.,Botla,S.,Persson,J.,Schmierer,B.&Taipale,J.CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response.Nature Medicine 24,927-930(2018).

[2570] 14.Anzalone,A.V.et al.Search-and-replace genome editing without double-strand breaks or donor DNA.Nature 576,149-157(2019).

[2571] 实施例9-使用双生引导编辑和位点特异性重组酶进行的可编程大DNA缺失、替换、整合和倒位。

[2572] 原则上,可以使用在基因组中的特定位置处的DNA序列的靶向缺失、替换、整合或倒位来研究或治疗许多人遗传疾病。然而,在哺乳动物细胞基因组的背景下精确和高效地进行这些DNA序列转化仍然是挑战。本文提供了双生引导编辑(双生PE),这是用于在人基因组内源性位点处可编程替换或切除DNA序列而不需要双链DNA断裂的方法。双生PE采用引导编辑(PE)蛋白和两种引导编辑向导RNA(pegRNA),其作为在基因组DNA的相反链上合成互补DNA瓣的模板,从而用pegRNA编码的序列替换PE诱导的切口位点之间的内源DNA序列。本文显示双生PE可以执行高达780bp的精确删除,以及用高达108bp的新序列精确替换基因组DNA序列。通过将单个或多路复用双生PE与位点特异性丝氨酸重组酶组合,证明了在包括AAVS1、CCR5和ALB的安全港位点处的基因大小的DNA质粒的靶向整合,并且进一步地,还证明了可以用于校正常见亨特氏综合征等位基因的IDS处的39kb倒位。双生PE显著扩展了基因组编辑能力而不需要双链DNA断裂,并且与其他基因组编辑工具组合,能够校正或补充人细胞中的复杂致病等位基因变体。

[2573] 与疾病相关的人基因变体通过各种序列变化产生,从单个碱基对取代到巨型碱基规模的缺失和重排。能够在人细胞中安装、校正或补充这些致病变体的基因组编辑方法具

有促进对人遗传疾病的理解的潜力,并可能导致新的治疗方法的开发。最近,已经开发若干基于CRISPR-Cas系统的基因组编辑方法,包括核酸酶、碱基编辑器和引导编辑。CRISPR-Cas核酸酶,当与一个向导RNA一起使用时,可以用于通过生成双链DNA断裂(DSB)导致随机性插入缺失来破坏基因。碱基编辑器可以用于精确安装C·G到T·A或A·T到G·C转换碱基对编辑,或C·G到G·C碱基对编辑。碱基编辑器可以精确地安装十二个碱基对取代编辑中的任一者,以及小到中等规模的插入和缺失。然而,单独的这些策略并不直接适用于进行更大量的基因组编辑,诸如大的缺失、插入、替换或倒位。

[2574] 此前,已经开发成对Cas9核酸酶策略用于长度范围为约50到>100,000个碱基对的基因组DNA序列的靶向缺失。此外,通过提供线性供体DNA序列,可以在单个切割位点处或成对切割位点之间通过末端连接或同源定向DNA修复过程进行新DNA序列的靶向插入。虽然在许多方面具有多功能性,但单个核酸酶和成对核酸酶的方法均伴随若干缺点。首先,使用成对核酸酶进行缺失可以生成多种中靶副产物,包括具有插入缺失的期望缺失、不具有期望缺失的个体DSB位点处的插入缺失,DSB位点之间DNA序列的不期望倒位,以及在DSB位点处外源性DNA序列的意外整合。此外,缺失的精确位置受PAM可用性和核酸酶生成的相应切割位点限制。类似的限制和副产物存在于DNA供体敲入,当使用同源性非依赖性的方法时,经常在没有控制序列方向的情况下发生。最后,在中靶和脱靶位点同时生成多个DSB,这在使用额外的向导RNA的情况下变得更有可能,可能导致脱靶基因组修饰、染色体重排或染色体碎裂(chromothripsis)。因此,当应用基于核酸酶的编辑策略来切除、替换或插入较大的DNA序列时,在产物控制和纯度方面存在相当大的挑战。

[2575] 事实证明,引导编辑能够在人细胞中进行高达约40bp的精确插入和高达约80bp的缺失,并且具有较高的期望产物与副产物比率。原则上说,可以使用引导编辑以通过避免DSB的生成来规避与核酸酶相关的许多挑战。然而,尚未显示引导编辑介导典型基因编码序列大小的较大插入和缺失,并且简单引导编辑反应的假定机制由于需要长pegRNA逆转录模板而使这些较大的DNA变化变得困难。位点特异性DNA重组酶具有进行若干DNA转化,包括大DNA序列的切除、倒位、整合或交换的能力。然而,重编程位点特异性重组酶的挑战使得其目前在应用于基因组工程和编辑应用方面具有挑战性。

[2576] 本文提出了双生引导编辑(双生PE)的开发,它能够在人细胞中以高效率在内源性基因组位点进行DNA序列的缺失、取代或插入。除了这些变化外,还显示双生PE可以用于在人基因组的靶位点处高效率地整合一个或多个重组酶识别位点,这些位点随后可以用作位点特异性丝氨酸重组酶的底物,能够实现遗传货物的靶向整合或基因组DNA序列的靶向倒位。

[2577] 结果

[2578] 引导编辑使用引导编辑蛋白,包括催化受损的Cas9切口酶和野生型(PE1)或工程化(PE2)MMLV逆转录酶的融合物,以及指定靶基因组位点并且编码期望编辑的引导编辑向导RNA(pegRNA)。在靶位点识别后,PE·pegRNA复合物使含有PAM的DNA链生成切口,并使用带切口的链作为引物将pegRNA的RT模板直接逆转录为基因组DNA。逆转录后,新合成的3'瓣侵入在很大程度上与其同源的相邻DNA,以替换多余的5'瓣序列。然后使用经编辑的DNA链作为模板,修复相反的未经编辑的链。因此,这种提出的编辑途径提供了两个机会,细胞DNA修复可以拒绝期望的编辑并将DNA序列回复到其原始形式。

[2579] 据预测,绕过DNA修复中潜在不受欢迎的步骤可以允许引导编辑以更高的效率发生,并能够实现不同类别的基因组编辑。此外,设想了使用一对pegRNA的双生引导编辑(双生PE)策略,每个pegRNA靶向一条或另一条DNA链并作为与由其他pegRNA模板化的序列互补的3'瓣DNA序列合成的模板(图108A)。据推测,如果新合成的DNA链与内源性靶位点高度不同,3'瓣将优先杂交以形成具有新DNA序列的退火3'突出端和原DNA序列的退火5'突出端的中间体(图108A)。由于两条经编辑的链均由引导编辑复合物合成,因此不需要通过含有编辑的瓣链入侵靶位点,也不需要将编辑拷贝到互补DNA链中。原始DNA序列(退火的5'突出端)的切除,通过聚合酶的空位填充,以及一对切口的连接,将导致切口位点之间的内源性序列用成对的3'瓣序列替换(图108A)。由于模板设计的灵活性,编辑原则上可以引入新的DNA序列,替换DNA序列的一部分,或删除DNA序列的一部分。

[2580] 为了评估双生PE策略,靶向HEK293T中的HEK位点3(以前称为HEK3)基因座以用38bp的Bxb1 attB附接序列替换90bp的内源性序列(图108B)。对于每个前间隔序列,用RT模板设计三种pegRNA,其中含有38bp attB序列的30、34或38nt(图108C)。这些pegRNA的成对组合预测为生成具有范围为22至38bp的重叠互补性的3'瓣(图108C)。令人兴奋的是,当这两种pegRNA连同质粒DNA转染中的PE2递送至细胞时,观察到attB位点插入的高效率,根据扩增子测序,pegRNA的一些组合产生>80%到期望产物的转换(图108C)。用于插入50bp的Bxb1 attP附着序列的类似策略以高达58%的效率实现编辑(图108C)。值得注意的是,每个pegRNA没有必要编码完整的插入序列,因为部分重叠的互补性瓣能够实现全长attB或attP序列的整合。然而,具有更高重叠的3'瓣导致略高的编辑效率以及用于在HEK位点3处插入Bxb1 attB和attP序列的更少插入缺失。

[2581] 接下来,测试了双生PE以确定它是否支持比使用PE2或PE3系统已经证明的更大的DNA序列的插入。首先,研究了PE3介导的范围为12到321bp的FKBP12编码序列片段的插入。用PE3靶向HEK位点3,较短的12bp和36bp插入实现适度效率(分别为31.8%和17.3%),但观察到108bp插入的期望产物很少(0.8%),并且321bp插入没有全长产物(图109A)。相比之下,双生PE能够实现108bp片段的16%插入效率,相比PE3总计20倍改善。此外,在CCR5基因座上,含有成对Bxb1重组酶位点的113bp和103bp插入以相似的效率(分别为10.7%和9.7%)实现(数据未显示)。这些结果说明,与PE3相比,使用双生PE有可能实现更大的序列插入。

[2582] 双生PE策略的一种潜在应用是用维持蛋白质序列的重编码的DNA序列替换外显子编码序列,并具有校正靶双生PE诱导的切口位点之间的任何突变的潜力。为了测试这种方法,靶向PAH基因,其中的突变导致遗传性代谢病症苯丙酮尿症(PKU)。在野生型HEK293T细胞的PAH中测试了外显子4和外显子7的部分的重编码,其中校正性编辑是可检测的。通过测试不同的瓣重叠长度,并在pegRNA的3'末端添加evoPreQ1基序,以对于外显子4中的64bp重编码高达9.4%的平均效率、对于在外显子7的46bp重编码高达22.7%的平均效率、对于在外显子7的64bp重编码高达27.4%的平均效率的适度效率,实现了期望序列重编码(图109B)。根据bioPKU数据库中注释的突变,这些区域的双生PE编辑共同允许校正PKU中涉及的不同突变。经检查的额外外显子也可以重编码,尽管效率较低(图112)。这些结果表明,原则上可以使用双生PE以用单个pegRNA对来校正多个突变。

[2583] 除了插入和替换DNA序列外,双生PE还应该能够进行精确的缺失编辑。成对核酸酶

缺失策略生成跨越两个DSB位点的缺失并因此受PAM可用性限制，并且期望的缺失通常伴随着不期望的插入缺失副产物。相比之下，由于DSB生成的缺乏以及在双生PE诱导的切口位置处写入额外的DNA序列的能力，双生PE具有以更大的灵活性和精确制成缺失的潜力。为了评估用于精确缺失的双生PE，比较了使用成对pegRNA的三种策略："单个-锚定"双生PE策略、"杂交体-锚定"双生PE策略和最近报道的"PrimeDel"策略(图109C)。每种策略在互补瓣中编码的序列上有所不同，允许使缺失相对于切口位点的定位具有灵活性。

[2584] 使用单个-锚定策略，使用13nt互补瓣以14.8％的效率删除与pegRNA诱导的切口位点之一相邻的77bp的序列，并且使用34nt互补瓣以18.8％的效率精确切除56bp的序列(图109D)。使用杂交体-锚定策略，在pegRNA诱导的切口位点之间删除64bp，使得产物在每个切口的3'处保留13bp的序列，效率为11.7％(图109D)。最后，PrimeDel策略测试了在pegRNA诱导的切口位点之间删除90bp，其以40.4％的效率发生(图109D)。值得注意的是，PrimeDel策略破坏了两条链上的PAM序列，这可能增加了效率，同时也减少了插入缺失。通过在pegRNA的3'末端添加evoPreQ1基序，编辑效率可以提高1.5倍到2.5倍，尽管在某些情况下伴随插入缺失的增加(图109D)。这些数据表明，双生PE提供了用于执行具有高度灵活性的靶向删除的策略，它不依赖于完全定位的核酸酶切割位点的可用性。

[2585] 最后，靶向与治疗有关的基因座DMD以将双生PE应用于较大的缺失。致病性DMD等位基因是杜氏肌营养不良的原因，通常包含含有外显子的区域的大缺失，导致移码的mRNA转录物。由于产生全长的肌营养不良蛋白而不替换缺失的外显子可能导致蛋白功能的部分挽救，因此破坏恢复阅读框架的第二外显子已认为是潜在的治疗策略。研究了三种双生PE缺失策略，以及先前报道的用于切除DMD的外显子51的Cas9核酸酶缺失策略。使用单个-锚定双生PE缺失方法，观察到含有DMD中的外显子51的780bp序列的缺失的效率为14.4％至24％(图109E)。虽然成对Cas9核酸酶策略与双生PE策略(平均14％至24％)相比实现更高的缺失效率(平均49.8％)，但成对Cas9核酸酶介导的缺失与双生PE(1％至12％的总插入缺失)相比也伴随着更高的插入缺失水平(33％的期望缺失具有插入缺失，11％的插入缺失不具有期望缺失)(图109E)。也可以使用替代的间隔区对，用PrimeDel策略生成627bp的缺失(平均效率28.8％)或用单个-锚定双生PE策略生成590bp的缺失(平均效率26.4％)，或通过使用双生PE以用38bp的Bxb1 attB序列替换589bp的序列(平均效率高达40.4％)切除DMD的外显子51(图109E)。这些实验表明，与成对Cas9核酸酶策略相比，双生PE和PrimeDel能够在人细胞中的治疗相关基因座生成大的缺失，具有更高的产物纯度，尽管平均效率较低。

[2586] 用双生PE和Bxb1整合酶在安全港位点处进行靶向DNA整合

[2587] 尽管双生PE能够进行比PE3更大的插入编辑，但序列插入大小的上限仍然不足以支持基因大小的DNA片段的整合。因此，在使用双生PE成功地将Bxb1附着序列高效率地插入到内源性人基因组位点后，将双生PE与丝氨酸重组酶组合用于DNA货物的位点特异性整合。为了鉴定DNA货物整合的位置，双生PE介导的Bxb1 attB和attP附着序列插入在HEK293T细胞中既定的安全港位点处。筛选靶向AAVS1基因座的32个间隔区对以插入50bp的Bxb1 attP序列(数据未显示)。在评估的间隔区对中，32个间隔区对组合中的18个的最佳pegRNA通过扩增子测序实现>50％的正确插入效率(图110A)。此外，筛选了靶向CCR5基因座的19个间隔区对以插入38bp的Bxb1 attB序列(数据未显示)，其中6个间隔区对通过扩增子测序实现>50％的理想编辑效率(图110B)。这些结果表明，双生PE可以用于在人细胞的安全港位点处

高效率地插入重组酶附着序列。

[2588] 接下来,对双生PE掺入的Bxb1 attB和attP序列进行检查,以确定它们是否可以充当携带伙伴Bxb1附着位点的质粒DNA货物整合的靶位点。首先,使用双生PE生成在CCR5基因座上携带同源的attB位点插入的单个细胞克隆(数据未显示)。用表达Bxb1重组酶的质粒和5.6kb的含attP的供体DNA质粒转染该克隆细胞系,产生如通过ddPCR测量的在靶位点处12-17%的敲入(图113),与先前报道的Bxb1整合效率一致。在这些结果的鼓舞下,探索了双生PE介导的附着位点插入和BxB1介导的DNA供体整合,以确定它们是否可以在单个转染步骤中实现。令人兴奋的是,用编码PE2、两种pegRNA、BxB1和供体DNA的质粒转染HEK293T细胞,导致如通过ddPCR测量的1.4-6.8%的敲入效率(图110C)。通过扩增子测序确认了含有预期attL和attR重组产物的预期接合序列,产物纯度的范围为71-95%(图114)。

[2589] 为了提高“一锅”敲入的效率,测试了双生PE介导的attB或attP附着位点、典型(GT)Bxb1核心序列、替代(GA)Bxb1核心序列以及不同的瓣重叠长度。结果发现,对于一锅敲入转染,用双生PE掺入attB位点比掺入attP位点表现更好(3.3%对0.5%敲入,图110D)。此外,据观察,插入具有典型GT核心序列的附着位点通常比插入具有替代GA序列的附着位点导致更高的敲入效率(在attB背景下为3.3%与0%,图110D)。然而,值得注意的是,典型和替代核心序列允许与相应的attP附着点进行正交重组。减少pegRNA编码的重组酶序列之间的重叠也提高了敲入效率,当插入attB时,从38个碱基对重叠的3.3%提高到20个碱基对重叠的5.5%(图110D),这可能是由于表达pegRNA的质粒DNA序列内的附着序列范围减少。虽然在不同的重叠长度间观察到类似的双生PE效率,但供体DNA和pegRNA质粒之间的重组减少(图115A-115B)。

[2590] 接下来,在ALB基因座处评估供体整合,该基因座可以用于治疗性蛋白质从肝细胞中分泌。白蛋白由肝脏活性分泌,约占人血浆蛋白的60%。因此,在相对较小百分比的肝细胞中进行ALB基因座的治疗性转基因整合可以实现许多疾病的蛋白表达的治疗性水平。设计了靶向ALB的内含子1的策略,其中Bxb1将介导含有剪接接受体序列的环状DNA,然后是感兴趣的蛋白质的cDNA的整合。原则上,整合将允许将在外显子1中编码的白蛋白分泌信号剪接到治疗性转基因的编码序列。首先,筛选用于插入ALB的内含子1中的Bxb1 attB序列的pegRNA对。在筛选pegRNA对后,鉴定通过扩增子测序实现attB序列43%的正确插入的间隔区组合(图110E)。在HEK293T中以1.3%的效率实现该基因座处的单次转染敲入(图110F)。在Huh7细胞中,观察到attB序列34%的正确插入,单次转染敲入效率为2.6%。有趣的是,尽管在HEK293T细胞中CCR5敲入比ALB敲入更高效,但在Huh7中ALB处的敲入比CCR5处的敲入更高效(2.6%与1.1%)。

[2591] 表14.本实施例中使用的DNA供体序列

[2592]

DNA供体	序列	SEQ ID NO:
attB_Puro_GA供体DNA:	gatgccagctcattcctcccactcatgatctatagatccccgggctgcaggaattctacccactctgtcgataccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccaccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggcttgtcgacgacggcggactccgtcgtcaggatcatccgtgagcaagggcgagga gctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcga gggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgacc accctgacctacggcgtgcagtgcttcagccgctacccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaag gctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccc tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaaca gccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca gcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagca cccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactct cggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgccaccatgggcga tcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccgccaagcttatg catgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggggaggggtcggcaattgaaccggtgcct agagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgccttttcccgagggtgggggagaaccgtatata agtgcagtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcccgcgggcct ggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagcttcgggttgga agtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgggcgctggggcc gccgcgtgcgaatctggtggcacctcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatgacctgctgcga cgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggttttggggccgcgggcggcgacgggg gcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggggtagtctcaag ctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccggtcggcaccag ttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggagagcgggc gggtgagtcacccacacaaaggaaaagggcctttccgtcctcagccgtcgcttcatgtgactccacggagtaccgggcgccgtcc aggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggggaggggttttatgcgatggagtttccccacactga gtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttggttcattctcaa gcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagcccacggtgcgcct cgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcg acccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtggg tcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcc cgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccgggcccaagga gcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagt ggaggcggccgagcgcgccgggggtgcccgccttcctggagacctccgcgccccgcaacctcccccttctacgagcggctcggctt caccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccggatcgggagag ggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctgattaaggagaac atgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagccctacgag ggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccctteгcttcgacatcctggctactagcttcctctacggc agcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggagagagtcac cacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtcaagatcag aggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgtaccccgct gacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatcaagaccacatat agatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaaggaggccaaca acgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaattaaga attctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtg aaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcag ggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcctcgtcgcgcgg tttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatgggatcggccattg	4282

	aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacgatcggctg ctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtccggtgccctgaatgaact gcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggg aagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggc tgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgta ctcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggct caaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgc ttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagc ttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgac gagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaattcggtaccaataa aagagctttatttttcatgatctgtgtgttggtttttggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaa aatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctc tcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggta tctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaac tatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgta ggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccag ttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagatt acgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	
[2593] attB_Puro供体DNA:	gatgccagctcattcctcccactcatgatctatagatcccccgggctgcaggaattctacccactctgtcgatacccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccacccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggcttgtcgacgacggcggtctccgtcgtcaggatcatccgtgagcaagggcgaggga gctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcga gggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgacc accctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaag gctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccc tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaaca gccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca gcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagca cccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactct cggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgccaccatgggcga tcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccgccaagcttatg catgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggggagggtcggcaattgaaccggtgcct agagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgccttttcccgagggtgggggagaaccgtatata agtgcagtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcccgcgggcct ggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagcttcgggttgga agtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgggcgctggggcc gccgcgtgcgaatctggtggcaccttcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatgacctgctgcga cgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggttttttgggccgcgggcggcgacgggg gcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggggtagtctcaag ctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccggtcggcaccag ttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggagagcgggc gggtgagtcacccacacaaaggaaaagggccttccgtcctcagccgtcgcttcatgtgactccacggagtaccgggcgccgtcc aggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggagggttttatgcgatggagtttccccacactga gtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttggttcattctcaa gcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccacatgaccgagtacaagcccacggtgcgcct cgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcg acccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtggg tcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcc cgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaagga gcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagt ggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggctt caccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccggatcgggagag ggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctgattaaggagaac atgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagccctacgag ggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccctttcgccttcgacatcctggctactagcttcctctacggc agcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggagagagtcac cacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtcaagatcag aggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgtaccccgct gacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatcaagaccacatat agatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaaggaggccaaca acgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaattaaga attctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtg	4283

		aaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcag ggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcctcgtcgccgcgg tttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatgggatcggccattg aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacgatcggctg ctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtccggtgccctgaatgaact gcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggg aagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggc tgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgta ctcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggct caaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgc ttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagc ttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgac gagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaattcggtaccaataa aagagctttatttttcatgatctgtgtgttggtttttggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaa aatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctc tcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggta tctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaac tatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgta ggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccag ttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagatt acgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	
[2594]	attP_Puro_GA供体DNA:	gatgccagctcattcctcccactcatgatctatagatcccccgggctgcaggaattctacccactctgtcgatacccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccacccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggtttgtctggtcaaccaccgcggactcagtggtgtacggtacaaaccccgtgagcaa gggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccgg cgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccac cctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgcca tgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagg gcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtac aactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcg aggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccac tacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgcc gggatcactctcggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgcca ccatgggcgatcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccg ccaagcttatgcatgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggggaggggtcggcaattg aaccggtgcctagagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgccttttcccgagggtggggga gaaccgtatataagtgcagtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggtt cccgcgggcctggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgag cttcgggttggaagtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgg gcgctggggccgccgcgtgcgaatctggtggcacctcgcgccgtgtctcgctgctttcgataagtctctagccatttaaaatttttgatg acctgctgcgacgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggtttttggggccgcggg cggcgacggggcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggg gtagtctcaagctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccg gtcggcaccagttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgg gagagcgggcgggtgagtcacccacacaaaggaaaagggccttccgtcctcagccgtcgcttcatgtgactccacggagtaccg ggcgccgtccaggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggggagggttttatgcgatggagtttc cccacactgagtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttg gttcattctcaagcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagccca cggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgc cacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggc aaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgcc gagatcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccg gcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgc tccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgag cggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccgg atcgggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctga ttaaggagaacatgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaa gccctacgagggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccctcgccttcgacatcctggctactagct tcctctacggcagcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatggg agagagtcaccacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacg tcaagatcagaggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgct gtaccccgctgacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatca	4284

		agaccacatatagatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaag gaggccaacaacgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaa gcttaattaagaattctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaa atgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttat gtttcaggttcaggggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcc tcgtcgccgcggtttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatg ggatcggccattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaaca gacgatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtccggtgc cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtc actgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagt atccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcga gcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaact gttcgccaggctcaaggcgcgcatgcccgaccggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtg gaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgat attgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttct atcgccttcttgacgagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaat tcggtaccaataaaagagctttattttcatgatctgtgtgttggtttttggccgcgttgctggcgtttttccataggctccgcccccctgac gagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagct ccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagct cacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcg ccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca gagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcg ctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgttt gcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	
[2595]	attP_Puro供体DNA:	gatgccagctcattcctcccactcatgatctatagatcccccgggctgcaggaattctacccactctgtcgatacccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccacccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggtttgtctggtcaaccaccgcggtctcagtggtgtacggtacaaaccccgtgagcaag ggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggc gagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccacc ctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccat gcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgaggg cgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaa ctacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgag gacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccacta cctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgg gatcactctcggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgccacc atgggcgatcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccgcc aagcttatgcatgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggggagggggtcggcaattgaa ccggtgcctagagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgccttttcccgagggtgggggagaa accgtatataagtgcagtagtcgccgtgaacgttcttttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcc cgcgggcctggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagctt cgggttggaagtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgggcg ctggggccgccgcgtgcgaatctggtggcacctttcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatgacc tgctgcgacgcttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggttttttgggccgcgggcg gcgacggggcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggggt agtctcaagctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgcccgccctgggcggcaaggctggcccggt cggcaccagttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggga gagcgggcgggtgagtcacccacacaaaggaaaagggccttccgtcctcagccgtcgcttcatgtgactccacggagtaccggg cgccgtccaggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggggagggggttttatgcgatggagtttccc cacactgagtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttggtt cattctcaagcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagcccacg gtgcgcctcgccacccgcgacgacgtccccctgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgcca caccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaa ggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccga gatcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggc ccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctcc ccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctcccctttctacgagcgg ctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccggatc gggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctgatta aggagaacatgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagc cctacgagggcacccagaccatgagaatcaaggtggtcgagggcggccctctccccttcgccttcgacatcctggctactagcttc ctctacggcagcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggag	4285

[2596]

agagtcaccacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtc
aagatcagaggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgt
accccgctgacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatcaag
accacatatagatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaagga
ggccaacaacgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagc
ttaattaagaattctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatg
ctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtt
tcaggttcaggggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcctcg
tcgccgcggtttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatgggga
tcggccattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagac
gatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccct
gaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcact
gaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatc
catcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagc
gagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgtt
cgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaa
aatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattg
ctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgccccgattcgcagcgcatcgccttctatc
gccttcttgacgagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaattcg
gtaccaataaaagagctttattttcatgatctgtgtgttggtttttggccgcgttgctggcgtttttccataggctccgcccccctgacgag
catcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccct
cgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcac
gctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgcctt
atccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagag
cgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctct
gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgca
agcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg

[2597] 多路复用编辑和重组酶介导的倒位

[2598] 在许多人致病等位基因中发现了大的结构变体，诸如F8基因座上的0.6Mb的大倒位，其导致约50％的严重血友病A病例。受使用双生PE的重组酶附着位点插入的高效率启发，推断attB和attPBxb1附着位点的多路复用插入可以用于通过间插DNA序列的单向性缺失或倒位来校正更复杂的遗传变体。为了在治疗相关基因座上测试这种方法，尝试在IDS和其假基因IDS2之间进行39kb的倒位。在约13％的亨特氏综合征病例中观察到这些位点之间的倒位，并且对致病等位基因中断点的表征揭示，倒位通常发生在同时存在于IDS和IDS2中的内含子7的重组热点内。因此，靶向侧翼为这些重组热点的区域以插入Bxb1 attP和attB附着序列，当其以相对方向进行定向时，可以用作Bxb1单向性倒位的底物，用于校正患者细胞中的致病等位基因序列(图111A)。在筛选12个pegRNA间隔区对并优化pegRNA序列特征后，发现能够以分别为73.7％和69.8％的效率在左侧和右侧靶向位点插入attP或attB附着位点的pegRNA对(图111B)。

[2599] 接下来，用Bxb1重组酶介导的IDS和IDS2中39kb序列的倒位，探索了多路复用双生PE介导的attP和attB附着位点的插入。最初，用双生PE组分然后是Bxb1重组酶探索序贯DNA转染。测试了第一组pegRNA(第一组)，它在IDS2的内含子3中安装正向attP序列，在IDS的内含子7中安装反向attB序列。此外，使用第二组pegRNA(第二组)以在第一次转染时在IDS2中安装反向attP序列和在IDS中安装正向attB序列。对于用pegRNA组1和组2进行的多路复用编辑，分别观察到52.3％或54.5％的attP序列插入和27.6％或30.9％的attB序列插入(图111C)。当经编辑的细胞随后用Bxb1重组酶转染时，与模拟转染对照相比，观察到含有attP和attB序列的扩增等位基因的百分比显著降低(p值<0.05)，这表明一些经编辑的等位基因用作Bxb1介导的重组的底物(图111C)。扩增预期的倒位接合，然后进行扩增子测序，显示出Bxb1介导的重组的attL和attR序列签名，两个接合处的产物纯度达或超过80％(图111D)。

[2600] 为了进行一锅双生PE和Bxb1介导的倒位并规避pegRNA质粒DNA之间不需要的重

组，核转染包含PE2 mRNA、来自组1的合成pegRNA和Bxb1mRNA的所有RNA组分。使用扩增子测序，捕获了含有attR和attL序列的预期倒位等位基因接合。此外，为了量化倒位效率，设计了可以与非倒位和倒位等位基因中的相同序列结合并且因此使用同一引物对扩增这两种编辑的反向引物(图116)。观察到“序贯”和“一锅”策略的倒位效率分别为约4.6-7.2％和约1.5-2.3％(图111E)。总之，这些数据支持将双生PE与位点特异性丝氨酸重组酶组合，可以用于校正亨特氏综合征等位基因中发现的39kb倒位，并且可以充当靶向其他复杂致病变体的有前途的治疗平台。

[2601] 讨论

[2602] 如本文所述，开发了可以用于删除、替换或插入人细胞的基因组中靶向位置处的DNA序列的双生引导编辑方法。当与位点特异性丝氨酸重组酶组合时，双生PE也可以支持DNA货物的整合，或DNA序列的倒位。

[2603] 方法

[2604] 一般方法

[2605] 除非另有说明，否则使用Phusion U Green Multiplex PCR Master Mix (ThermoFisher Scientific)或Q5 Hot Start High-Fidelity 2x Master Mix(New England BioLabs)通过PCR进行DNA扩增。DNA寡核苷酸从Integrated DNA Technologies获得。通过将退火的寡核苷酸连接到BsmBI消化的接受体载体上构建表达sgRNA的质粒。如前所述通过Gibson组装或Golden Gate组装构建表达pegRNA的质粒。本实施例中使用的sgRNA和pegRNA构建体的序列列于表11-12中。

[2606] 表11.本实施例中使用的sgRNA的序列

[2607]

sgRNA	间隔区序列	SEQ ID NO:
HEK3_3b_+90_切口生成	GTCAACCAGTATCCCGGTGC	616
pU6-Sp-sgRNA-DMDExo51-A1	GATTGGCTTTGATTTCCCTA	3971
pU6-Sp-sgRNA-DMDExo51-A3	GTATATGATTGTTACTGAGA	3972
pU6-Sp-sgRNA-DMDExo51-B1	Gcagttgcctaagaactggt	3973

[2608] 表12本实施例中使用的pegRNA的序列

[2609]

pegRNA	间隔区序列	间隔区 SEQ ID NO:	3'延伸	延伸 SEQ ID NO:
HEK3_attB_A_38	ggcccagactgagcacgtga	383	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCGTGCTCAGTCTG	4029
HEK3_attB_A_34	ggcccagactgagcacgtga	383	TCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCGTGCTCAGTCTG	4030
HEK3_attB_A_30	ggcccagactgagcacgtga	383	GACGACGGAGACCGCCGTCGTCGACAAGCCCGTGCTCAGTCTG	4031
HEK3_attB_B_38	gtcaaccagtatccggtgc	616	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGG	4032
HEK3_attB_B_34	gtcaaccagtatccggtgc	616	TGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGG	4033
HEK3_attB_B_30	gtcaaccagtatccggtgc	616	GACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGG	4034
HEK3_attP_A_43	ggcccagactgagcacgtga	383	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGCTCAGTCTG	4035
HEK3_attP_A_39	ggcccagactgagcacgtga	383	GTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGCTCAGTCTG	4036
HEK3_attP_A_35	ggcccagactgagcacgtga	383	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGCTCAGTCTG	4037
HEK3_attP_B_44	gtcaaccagtatccggtgc	616	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCGGGATACTGG	4038
HEK3_attP_B_40	gtcaaccagtatccggtgc	616	GGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCGGGATACTGG	4039
HEK3_attP_B_36	gtcaaccagtatccggtgc	616	AACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCGGGA	4040

[2610]

			TACTGG	
PE3_HEK3_FKBP_ins12bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcacacctgcactcccgtgctcagtctg	4041
PE3_HEK3_FKBP_ins36bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcaccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4042
PE3_HEK3_FKBP_ins108bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcaaaatttctttccatcttcaagcatcccggtgtagtgcaccacgcaggtctggccgcgcttggggaaggtgcgcccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4043
PE3_HEK3_FKBP_ins321bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcattccagttttagaagctccacatcgaagacgagagtggcatgtggtgggatgatgcctgggtgcccagtggcaccataggcataatctggagatatagtcagtttggctctctgacccacactcatctgggcaacccctttcttccagcctcggatcacctcctgcttgcctagcataaacttaaagggcttgtttctgtcccgggaggaatcaaatttctttccatcttcaagcatcccggtgtagtgcaccacgcaggtctggccgcgcttggggaaggtgcgccccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4044
双生PE_HEK3_FKBP_ins108bp_32bp_突出端_A	ggcccagactgagcacgtga	383	accacgcaggtctggccgcgcttggggaaggtgcgcccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4045
双生PE_HEK3_FKBP_ins108bp_32bp_突出端_B	gtcaaccagtatccggtgc	616	accttccccaagcgcggccagacctgcgtggtgcactacaccgggatgcttgaagatggaaagaaatttccgggatactgg	4046
pU6-Sp-gRNA-HEK3_del_DF_A_13nt TEMP_基础	ggcccagactgagcacgtga	383	tcctctgccatcacgtgctcagtctg	4047
pU6-Sp-gRNA-HEK3_del_DF_B_13nt TEMP_基础	gtcaaccagtatccggtgc	616	tgatggcagaggaccgggatactgg	4048
pU6-Sp-gRNA-HEK3_del_DF_A_34nt TEMP_基础	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcacgtgctcagtctg	4049
pU6-Sp-gRNA-HEK3_del_DF_B_34nt TEMP_基础	gtcaaccagtatccggtgc	616	tgatggcagaggaaaggaagccctgcttcctccaccgggatactgg	4050
pU6-Sp-gRNA-HEK3_del_DF_A_13-13TEMP_杂交体	ggcccagactgagcacgtga	383	tgcaggagctgcatcctctgccatcacgtgctcagtctg	4051
pU6-Sp-gRNA-HEK3_del_DF_B_13-13TEMP_杂交体	gtcaaccagtatccggtgc	616	tgatggcagaggatgcagctcctgcaccgggatactgg	4052
pU6-Sp-gRNA-HE	ggcccagactga	383	gcccagccaaacttgtcaaccagtatcccggcgtg	4053

[2611]

K3_del_DF_A_31nt TEMP_Shendure	gcacgtga		ctcagtctg	
pU6-Sp-gRNA-HE K3_del_DF_B_31nt TEMP_Shendure	gtcaaccagtatc ccggtgc	616	gggtcaatccttggggcccagactgagcacgccg ggatactgg	4054
pU6-Sp-gRNA-HE K3_del_DF_A_13nt TEMP_基础 _EvoPreQ1	ggcccagactga gcacgtga	383	tcctctgccatcacgtgctcagtctgTTAAAT AACGCGGTTCTATCTAGTTACG CGTTAAACCAACTAGAA	4055
pU6-Sp-gRNA-HE K3_del_DF_B_13nt TEMP_基础 _EvoPreQ1	gtcaaccagtatc ccggtgc	616	tgatggcagaggaccgggatactggAAAAT ATACGCGGTTCTATCTAGTTAC GCGTTAAACCAACTAGAA	4056
pU6-Sp-gRNA-HE K3_del_DF_A_34nt TEMP_基础 _EvoPreQ1	ggcccagactga gcacgtga	383	tggaggaagcagggcttcctttcctctgccatcac gtgctcagtctgTTAATAATCGCGGTT CTATCTAGTTACGCGTTAAACC AACTAGAA	4057
pU6-Sp-gRNA-HE K3_del_DF_B_34nt TEMP_基础 _EvoPreQ1	gtcaaccagtatc ccggtgc	616	tgatggcagaggaaaggaagccctgcttcctcca ccgggatactggAAAAAAAACGCGG TTCTATCTAGTTACGCGTTAAA CCAACTAGAA	4058
pU6-Sp-gRNA-HE K3_del_DF_A_13-1 3TEMP_杂交体 _EvoPreQ1	ggcccagactga gcacgtga	383	tgcaggagctgcatcctctgccatcacgtgctcagt ctgATAAATAACGCGGTTCTATC TAGTTACGCGTTAAACCAACTA GAA	4059
pU6-Sp-gRNA-HE K3_del_DF_B_13-1 3TEMP_杂交体 _EvoPreQ1	gtcaaccagtatc ccggtgc	616	tgatggcagaggatgcagctcctgcaccgggata ctggAAAAAAGGCGCGGTTCTAT CTAGTTACGCGTTAAACCAACT AGAA	4060
pU6-Sp-gRNA-HE K3_del_DF_A_31nt TEMP_Shendure_E voPreQ1	ggcccagactga gcacgtga	383	gcccagccaaacttgtcaaccagtatcccggcgtg ctcagtctgTTAAATACCGCGGTTCT ATCTAGTTACGCGTTAAACCAA CTAGAA	4061
pU6-Sp-gRNA-HE K3_del_DF_B_31nt TEMP_Shendure_E voPreQ1	gtcaaccagtatc ccggtgc	616	gggtcaatccttggggcccagactgagcacgccg ggatactggAAAATAATCGCGGTTC TATCTAGTTACGCGTTAAACCA ACTAGAA	4062
pU6-Sp-pegRNA-D MDExo51-A1_a_基 础	GATTGGCT TTGATTTC CCTA	3971	ttgaataggaagtaaattaatttgaagctggaccct agGGAAATCAA	4063
pU6-Sp-pegRNA-D MDExo51-A1_b_基 础	GATTGGCT TTGATTTC CCTA	3971	ttgaataggaagtaaattaatttgaagctggaccct agGGAAATCAAAG	4064
pU6-Sp-pegRNA-D MDExo51-A1_c_基 础	GATTGGCT TTGATTTC CCTA	3971	ttgaataggaagtaaattaatttgaagctggaccct agGGAAATCAAAGCC	4065
pU6-Sp-pegRNA-D MDExo51-B1_A1_ b_基础	Gcagttgcctaa gaactggt	3973	tagggtccagcttcaaattaatttacttcctattcaaa gttcttaggc	4066

[2612]

pU6-Sp-pegRNA-DMDExo51-B1_A1_c_基础	Gcagttgcctaagaactggt	3973	tagggtccagcttcaaattaatttacttcctattcaaagttcttaggcaa	4067
pU6-Sp-pegRNA-DMDExo51-A3_c_基础	GTATATGATTGTTACTGAGA	3972	atgcctggacaagtaacttaagttaaataagccttctCAGTAACAATCAT	4068
pU6-Sp-pegRNA-DMDExo51-B1_A3_a_基础	Gcagttgcctaagaactggt	3973	agaaggcttatttaacttaagttacttgtccaggcatagttcttag	4069
pu6-sp-pegrna-dmdexo51-a3-b1_a_PrimeDel	GTATATGATTGTTACTGAGA	3972	tgctgagagagaaacagttgcctaagaactCAGTAACAAT	4070
pu6-sp-pegrna-dmdexo51-b1-a3_a-PrimeDel	Gcagttgcctaagaactggt	3973	accacttccacaatgtatatgattgttactgagttcttagg	4071
pu6-sp-pegrna-dmdexo51-b1-a3_b-PrimeDel	Gcagttgcctaagaactggt	3973	accacttccacaatgtatatgattgttactgagttcttaggcaa	4072
pU6-Sp-pegRNA-DMDExo51-A3_b_attB_fwd	GTATATGATTGTTACTGAGA	3972	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCAGTAACAATC	4073
pU6-Sp-pegRNA-DMDExo51-A3_c_attB_fwd	GTATATGATTGTTACTGAGA	3972	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCAGTAACAATCATA	4074
pU6-Sp-pegRNA-DMDExo51-B1_b_attB_fwd	Gcagttgcctaagaactggt	3973	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATagttcttaggc	4075
pU6-Sp-pegRNA-DMDExo51-B1_c_attB_fwd	Gcagttgcctaagaactggt	3973	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATagttcttaggcaac	4076
pU6-Sp-pegRNA-DMDExo51-B2_b_attB_fwd	Gaggagagtaaagtgattgg	3974	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATatcactttactc	4077
pU6-Sp-pegRNA-DMDExo51-B2_c_attB_fwd	Gaggagagtaaagtgattgg	3974	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATatcactttactctcc	4078
A_1077a	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCT	4079
A_1077b	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCTGG	4080
A_1077c	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCTGGCT	4081
A_1098a	GGGAAGGGGCAGGAGAGCCA	3976	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCTCCTGC	4082
A_1098b	GGGAAGGGGCAGGA	3976	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACC	4083

[2613]

	GAGCCA		TCTCTCCTGCCC	
A_1098c	GGGAAGGGGCAGGAGAGCCA	3976	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCTCCTGCCCCT	4084
A_1246a	GAATATGTCCCAGATAGCAC	3977	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTATCTGGG	4085
A_1246b	GAATATGTCCCAGATAGCAC	3977	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTATCTGGGAC	4086
A_1246c	GAATATGTCCCAGATAGCAC	3977	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTATCTGGGACAT	4087
A_1267a	GGGGACTCTTTAAGGAAAGA	3978	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTTCCTTAAA	4088
A_1267b	GGGGACTCTTTAAGGAAAGA	3978	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTTCCTTAAAGA	4089
A_1267c	GGGGACTCTTTAAGGAAAGA	3978	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTTCCTTAAAGAGT	4090
A_1293a	GAGAAAGAGAAAGGGAGTAG	3979	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCCCTTTC	4091
A_1293b	GAGAAAGAGAAAGGGAGTAG	3979	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCCCTTTCTC	4092
A_1293c	GAGAAAGAGAAAGGGAGTAG	3979	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCCCTTTCTCTT	4093
A_1307a	GAGTAGAGGCGGCCACGACC	3980	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGGCCGC	4094
A_1307b	GAGTAGAGGCGGCCACGACC	3980	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGGCCGCCT	4095
A_1307c	GAGTAGAGGCGGCCACGACC	3980	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGGCCGCCTCT	4096
A_1582a	GATCAGTGAAACGCACCAGA	3981	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGTGCGTTT	4097
A_1582b	GATCAGTGAAACGCACCAGA	3981	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGTGCGTTTCA	4098
A_1582c	GATCAGTGAAACGCACCAGA	3981	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGTGCGTTTCACT	4099
A_1615a	GCAGCTCAGGTTCTGG	3982	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACC	4100

	GAGA		TCCCAGAACC	
A_1615b	GCAGCTCAGGTTCTGGGAGA	3982	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCAGAACCTG	4101
A_1615c	GCAGCTCAGGTTCTGGGAGA	3982	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCAGAACCTGAG	4102
A_1647a	GTGGCCACTGAGAACCGGGC	3983	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGGTTCTCA	4103
A_1647b	GTGGCCACTGAGAACCGGGC	3983	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGGTTCTCAGT	4104
A_1647c	GTGGCCACTGAGAACCGGGC	3983	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGGTTCTCAGTGG	4105
A_1810a	GAATCTGCCTAACAGGAGGT	3984	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTCCTGTTAG	4106
A_1810b	GAATCTGCCTAACAGGAGGT	3984	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTCCTGTTAGGC	4107
A_1810c	GAATCTGCCTAACAGGAGGT	3984	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTCCTGTTAGGCAG	4108
A_1890a	GTCACCAATCCTGTCCCTAG	3985	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGACAGGA	4109
A_1890b	GTCACCAATCCTGTCCCTAG	3985	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGACAGGATT	4110
A_1890c	GTCACCAATCCTGTCCCTAG	3985	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGACAGGATTGG	4111
A_3786a	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGG	4112
A_3786b	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGGGG	4113
A_3786c	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGGGGGC	4114
A_3835a	GACGTCACGGCGCTGCCCCA	3987	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCAGCGCC	4115
A_3835b	GACGTCACGGCGCTGCCCCA	3987	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCAGCGCCGT	4116
A_3835c	GACGTCACGGCGCTGC	3987	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACC	4117

[2614]

[2615]

	CCCA		TGGCAGCGCCGTGA	
A_3856a	GGTGTGCTGGGCAGGTCGCG	3988	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGACCTGCCC	4118
A_3856b	GGTGTGCTGGGCAGGTCGCG	3988	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGACCTGCCCAG	4119
A_3856c	GGTGTGCTGGGCAGGTCGCG	3988	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGACCTGCCCAGCA	4120
B_1154a	GTCCTTGGCAAGCCCAGGAG	3989	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTG	4121
B_1154b	GTCCTTGGCAAGCCCAGGAG	3989	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTGCC	4122
B_1154c	GTCCTTGGCAAGCCCAGGAG	3989	cTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTGCCAA	4123
B_1314a	GTGCGTCCTAGGTGTTCACC	3990	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAACACCTA	4124
B_1314b	GTGCGTCCTAGGTGTTCACC	3990	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAACACCTAGG	4125
B_1314c	GTGCGTCCTAGGTGTTCACC	3990	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAACACCTAGGAC	4126
B_1376a	GTCCTGGCAGGGCTGTGGTG	3991	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACAGCCCT	4127
B_1376b	GTCCTGGCAGGGCTGTGGTG	3991	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACAGCCCTGC	4128
B_1376c	GTCCTGGCAGGGCTGTGGTG	3991	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACAGCCCTGCCA	4129
B_1640a	GTGACCTGCCCGGTTCTCAG	3992	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGAACCGGG	4130
B_1640b	GTGACCTGCCCGGTTCTCAG	3992	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGAACCGGGCA	4131
B_1640c	GTGACCTGCCCGGTTCTCAG	3992	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGAACCGGGCAGG	4132
B_1676a	GAGCTTGGCAGGGGGTGGGA	3993	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACCCCCTG	4133
B_1676b	GAGCTTGGCAGGGGGT	3993	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACC	4134

[2616]

	GGGA		TCACCCCCTGCC	
B_1676c	GAGCTTGGCAGGGGGTGGGA	3993	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACCCCCTGCCAA	4135
B_1701a	GAGCCAGAGAGGATCCTGGG	3994	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGGATCCTC	4136
B_1701b	GAGCCAGAGAGGATCCTGGG	3994	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGGATCCTCTC	4137
B_1701c	GAGCCAGAGAGGATCCTGGG	3994	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGGATCCTCTCTG	4138
B_1705a	GATGGAGCCAGAGAGGATCC	3995	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTCCTCTCTG	4139
B_1705b	GATGGAGCCAGAGAGGATCC	3995	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTCCTCTCTGGC	4140
B_1705c	GATGGAGCCAGAGAGGATCC	3995	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTCCTCTCTGGCTC	4141
B_1883a	GGGGCCACTAGGGACAGGAT	3996	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGTCCCTA	4142
B_1883b	GGGGCCACTAGGGACAGGAT	3996	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGTCCCTAGT	4143
B_1883c	GGGGCCACTAGGGACAGGAT	3996	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGTCCCTAGTGG	4144
B_1902a	GTCCCCTCCACCCCACAGTG	3997	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGTGGGGTG	4145
B_1902b	GTCCCCTCCACCCCACAGTG	3997	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGTGGGGTGGA	4146
B_1902c	GTCCCCTCCACCCCACAGTG	3997	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGTGGGGTGGAGG	4147
B_1962a	GGGACCACCTTATATTCCCA	3998	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAATATAAG	4148
B_1962b	GGGACCACCTTATATTCCCA	3998	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAATATAAGGT	4149
B_1962c	GGGACCACCTTATATTCCCA	3998	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAATATAAGGTGG	4150
B_3839a	GACCTGCCCAGCACAC	3999	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACC	4151

	CCTG		TGGTGTGCTG	
B_3839b	GACCTGCCCAGCACACCCTG	3999	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGTGTGCTGGG	4152
B_3839c	GACCTGCCCAGCACACCCTG	3999	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGTGTGCTGGGCA	4153
B_3903a	GCGACTCCTGGAAGTGGCCA	4000	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCA	4154
B_3903b	GCGACTCCTGGAAGTGGCCA	4000	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCAGG	4155
B_3903c	GCGACTCCTGGAAGTGGCCA	4000	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCAGGAG	4156
B_3930a	GGACTTCCCAGTGTGCATCG	4001	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTG	4157
B_3930b	GGACTTCCCAGTGTGCATCG	4001	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTGGG	4158
B_3930c	GGACTTCCCAGTGTGCATCG	4001	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTGGGAA	4159
[2617] A_1_223_a	TTTGCAGTTTATCAGGATGA	4002	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGCAGTTTATC	4160
A_1_223_b	TTTGCAGTTTATCAGGATGA	4002	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGCAGTTTATCAG	4161
A_1_223_c	TTTGCAGTTTATCAGGATGA	4002	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGCAGTTTATCAGG	4162
A_2_260_a	GGTTGAGCAGGTAGATGTCA	4003	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCTGAGCAGGT	4163
A_2_260_b	GGTTGAGCAGGTAGATGTCA	4003	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCTGAGCAGGTAG	4164
A_2_260_c	GGTTGAGCAGGTAGATGTCA	4003	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCTGAGCAGGTAGAT	4165
A_3_325_a	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGC	4166
A_3_325_b	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGCAT	4167
A_3_325_c	CTGGGCGGCAGCATAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCG	4168

[2618]

	TGAG		GCAGCATAG	
A_4_360_a	GTCAAGAG TTGACACA TTGT	4005	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAAGA GTTGAC	4169
A_4_360_b	GTCAAGAG TTGACACA TTGT	4005	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAAGA GTTGACAC	4170
A_4_360_c	GTCAAGAG TTGACACA TTGT	4005	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAAGA GTTGACACA	4171
A_5_507_a	CCCAAGTG ATCACACT TGTC	4006	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAAGT GATCAC	4172
A_5_507_b	CCCAAGTG ATCACACT TGTC	4006	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAAGT GATCACAC	4173
A_5_507_c	CCCAAGTG ATCACACT TGTC	4006	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAAGT GATCACACTT	4174
A_6_510_a	CCACCCAA GTGATCAC ACTT	4007	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCCCCA AGTGA	4175
A_6_510_b	CCACCCAA GTGATCAC ACTT	4007	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCCCCA AGTGATC	4176
A_6_510_c	CCACCCAA GTGATCAC ACTT	4007	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCCCCA AGTGATCAC	4177
A_7_532_a	GGGAGAG ACGCAAAC ACAGC	4008	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAGAG ACGCA	4178
A_7_532_b	GGGAGAG ACGCAAAC ACAGC	4008	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAGAG ACGCAAA	4179
A_7_532_c	GGGAGAG ACGCAAAC ACAGC	4008	ATGATCCTGACGACGGAGACC GCCGTCGTCGACAAGCCAGAG ACGCAAACA	4180
B_1_272_a	TCTCTGAC CTGTTTTT CCTT	4009	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATCTGACC TG	4181
B_1_272_b	TCTCTGAC CTGTTTTT CCTT	4009	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATCTGACC TGT	4182
B_1_272_c	TCTCTGAC CTGTTTTT CCTT	4009	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATCTGACC TGTTT	4183
B_2_291_a	TCTTACTG TCCCCTTC TGGG	4010	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATTACTGT CCC	4184
B_2_291_b	TCTTACTG TCCCCTTC	4010	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATTACTGT	4185

[2619]

	TGGG		CCCCT	
B_2_291_c	TCTTACTG TCCCCTTC TGGG	4010	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATTACTGT CCCCTTC	4186
B_3_305_a	TCTGGGCT CACTATGC TGCC	4011	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGGGCTC ACT	4187
B_3_305_b	TCTGGGCT CACTATGC TGCC	4011	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGGGCTC ACTAT	4188
B_3_305_c	TCTGGGCT CACTATGC TGCC	4011	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGGGCTC ACTATGC	4189
B_4_326_a	CCCAGTGG GACTTTGG AAAT	4012	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATAGTGG GACT	4190
B_4_326_b	CCCAGTGG GACTTTGG AAAT	4012	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATAGTGG GACTTTG	4191
B_4_326_c	CCCAGTGG GACTTTGG AAAT	4012	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATAGTGG GACTTTGG	4192
B_5_330_a	GTGGGACT TTGGAAAT ACAA	4013	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGGACTT TGG	4193
B_5_330_b	GTGGGACT TTGGAAAT ACAA	4013	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGGACTT TGGAA	4194
B_5_330_c	GTGGGACT TTGGAAAT ACAA	4013	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGGACTT TGGAAAT	4195
B_6_414_a	CCTGACAA TCGATAGG TACC	4014	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGACAAT CGA	4196
B_6_414_b	CCTGACAA TCGATAGG TACC	4014	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGACAAT CGATA	4197
B_6_414_c	CCTGACAA TCGATAGG TACC	4014	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGACAAT CGATAGG	4198
B_7_538_a	GGAATCAT CTTTACCA GATC	4015	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATATCATC TTTA	4199
B_7_538_b	GGAATCAT CTTTACCA GATC	4015	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATATCATC TTTACC	4200
B_7_538_c	GGAATCAT CTTTACCA GATC	4015	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATATCATC TTTACCAG	4201
B_8_584_a	GCAGCTCT CATTTTCC	4016	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATGCTCTC	4202

	ATAC		ATT	
B_8_584_b	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4203
B_8_584_c	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTCCA	4204
B_9_601_a	TACAGTCAGTATCAATTCTG	4017	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTCAGTAT	4205
B_9_601_b	TACAGTCAGTATCAATTCTG	4017	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTCAGTATCA	4206
B_9_601_c	TACAGTCAGTATCAATTCTG	4017	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTCAGTATCAAT	4207
A_3_325_a	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGC	4166
B_6_414_b	CCTGACAATCGATAGGTACC	4014	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGACAATCGATA	4197
A_5_507_c	CCCAAGTGATCACACTTGTC	4006	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGTGATCACACTT	4174
A_6_510_b	CCACCCAAGTGATCACACTT	4007	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCCCAAGTGATC	4176
A_7_532_b	GGGAGAGACGCAAACACAGC	4008	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCAAA	4179
B_8_584_b	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4203
A_1077c	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCTGGCT	4081
B_1154c	GTCCTTGGCAAGCCCAGGAG	3989	cTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTGCCAA	4123
A_3786c	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGGGGGC	4114
B_3903c	GCGACTCCTGGAAGTGGCCA	4000	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCAGGAG	4156
B_3930c	GGACTTCCCAGTGTGCATCG	4001	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTGGGAA	4159
A_7_532_b	GGGAGAGACGCAAAC	4008	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAG	4179

[2620]

[2621]

	ACAGC		ACGCAAA	
B_8_584_b	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4203
A7_attB_30	GGGAGAGACGCAAACACAGC	4008	TCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCAAA	4208
B8_attB_30	GCAGCTCTCATTTTCCATAC	4016	TGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4209
A7_attB_20	GGGAGAGACGCAAACACAGC	4008	ACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCAAA	4210
B8_attB_20	GCAGCTCTCATTTTCCATAC	4016	ACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4211
A7_attB_GA_38	GGGAGAGACGCAAACACAGC	4008	ATGATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCCAGAGACGCAAA	4212
B8_attB_GA_38	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGACTCCGTCGTCAGGATCATGCTCTCATTTTC	4213
A7_attB_GA_30	GGGAGAGACGCAAACACAGC	4008	TCCTGACGACGGAGTCCGCCGTCGTCGACAAGCCAGAGACGCAAA	4214
B8_attB_GA_30	GCAGCTCTCATTTTCCATAC	4016	TGTCGACGACGGCGGACTCCGTCGTCAGGATCATGCTCTCATTTTC	4215
A7_attB_GA_20	GGGAGAGACGCAAACACAGC	4008	ACGACGGAGTCCGCCGTCGTCGACAAGCCAGAGACGCAAA	4216
B8_attB_GA_20	GCAGCTCTCATTTTCCATAC	4016	ACGACGGCGGACTCCGTCGTCAGGATCATGCTCTCATTTTC	4217
A7_attP_50	GGGAGAGACGCAAACACAGC	4008	GGGTTTGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCAAGAGACGCAAA	4218
B8_attP_50	GCAGCTCTCATTTTCCATAC	4016	TGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCGCTCTCATTTTC	4219
A7_attP_40	GGGAGAGACGCAAACACAGC	4008	TGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCAAGAGACGCAAA	4220
B8_attP_40	GCAGCTCTCATTTTCCATAC	4016	TGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCGCTCTCATTTTC	4221
A7_attP_30	GGGAGAGACGCAAACACAGC	4008	CGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCAAGAGACGCAAA	4222
B8_attP_30	GCAGCTCTCATTTTCC	4016	GGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCGCT	4223

[2622]

	ATAC		CTCATTTTC	
A7_attP_GA_50	GGGAGAG ACGCAAAC ACAGC	4008	GGGTTTGTACCGTACACCACTG AGTCCGCGGTGGTTGACCAGA CAAACCAAGAGACGCAAA	4224
B8_attP_GA_50	GCAGCTCT CATTTTCC ATAC	4016	TGGTTTGTCTGGTCAACCACCG CGGACTCAGTGGTGTACGGTA CAAACCCGCTCTCATTTTC	4225
A7_attP_GA_40	GGGAGAG ACGCAAAC ACAGC	4008	TGTACCGTACACCACTGAGTCC GCGGTGGTTGACCAGACAAAC CAAGAGACGCAAA	4226
B8_attP_GA_40	GCAGCTCT CATTTTCC ATAC	4016	TGTCTGGTCAACCACCGCGGAC TCAGTGGTGTACGGTACAAAC CCGCTCTCATTTTC	4227
A7_attP_GA_30	GGGAGAG ACGCAAAC ACAGC	4008	CGTACACCACTGAGTCCGCGGT GGTTGACCAGACAAACCAAGA GACGCAAA	4228
B8_attP_GA_30	GCAGCTCT CATTTTCC ATAC	4016	GGTCAACCACCGCGGACTCAG TGGTGTACGGTACAAACCCGCT CTCATTTTC	4229
A277_b	GACTGAAA CTTCACAG AATA	4018	AACTTCACAGAGGCTTGTCGAC GACGGCGGTCTCCGTCGTCAG GATCAT	4230
A277_c	GACTGAAA CTTCACAG AATA	4018	GAAACTTCACAGAGGCTTGTC GACGACGGCGGTCTCCGTCGTC AGGATCAT	4231
B358_b	GATTTATG AGATCAAC AGCAC	4019	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATCTGTTG ATCTC	4232
B358_c	GATTTATG AGATCAAC AGCAC	4019	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATCTGTTG ATCTCAT	4233
A7_attB_20	GGGAGAG ACGCAAAC ACAGC	4008	ACGACGGAGACCGCCGTCGTC GACAAGCCAGAGACGCAAA	4210
B8_attB_20	GCAGCTCT CATTTTCC ATAC	4016	ACGACGGCGGTCTCCGTCGTCA GGATCATGCTCTCATTTTC	4211
A277_c	GACTGAAA CTTCACAG AATA	4018	GAAACTTCACAGAGGCTTGTC GACGACGGCGGTCTCCGTCGTC AGGATCAT	4231
B358_c	GATTTATG AGATCAAC AGCAC	4019	GGCTTGTCGACGACGGCGGTCT CCGTCGTCAGGATCATCTGTTG ATCTCAT	4233
pU6-Sp-gRNA-IDS _DF_A1_a_attP_fw d.dna	GACACCAA AAACTGCC ACAGG	4020	TACCGTACACCACTGAGACCG CGGTGGTTGACCAGACAAACC TGTGGCAGTTA	4234
pU6-Sp-gRNA-IDS _DF_A1_a_attP_rev .dna	GACACCAA AAACTGCC ACAGG	4020	GTCTGGTCAACCACCGCGGTCT CAGTGGTGTACGGTACAAACC TGTGGCAGTTA	4235
pU6-Sp-gRNA-IDS _DF_A1_c_attP_fw	GACACCAA AAACTGCC	4020	TACCGTACACCACTGAGACCG CGGTGGTTGACCAGACAAACC	4236

	d.dna	ACAGG		TGTGGCAGTTTTA	
	pU6-Sp-gRNA-IDS_DF_A1_c_attP_rev.dna	GACACCAAAAACTGCCACAGG	4020	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGTGGCAGTTTTA	4237
	pU6-Sp-gRNA-IDS_DF_A4_b_attP_fwd.dna	GCACTCATTTCCTCCAAGCTC	4021	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTTGGAGGAAA	4238
	pU6-Sp-gRNA-IDS_DF_A4_b_attP_rev.dna	GCACTCATTTCCTCCAAGCTC	4021	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTTGGAGGAAA	4239
	pU6-Sp-gRNA-IDS_DF_B2_a_attP_fwd.dna	GTAGGTACAGGACAGGGCAG	4022	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCCTGTCC	4240
	pU6-Sp-gRNA-IDS_DF_B2_a_attP_rev.dna	GTAGGTACAGGACAGGGCAG	4022	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCTGTCC	4241
	pU6-Sp-gRNA-IDS_DF_B3_a_attP_fwd.dna	GAGATAGGTAGGTACAGGACA	4023	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCTGTACCT	4242
	pU6-Sp-gRNA-IDS_DF_B3_a_attP_rev.dna	GAGATAGGTAGGTACAGGACA	4023	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCTGTACCT	4243
	pU6-Sp-gRNA-IDS_DF_B4_b_attP_fwd.dna	GTGAAAAGATAGGTAGGTAC	4024	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCTACCTATCTA	4244
[2623]	pU6-Sp-gRNA-IDS_DF_B4_b_attP_rev.dna	GTGAAAAGATAGGTAGGTAC	4024	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCTACCTATCTA	4245
	pU6-Sp-gRNA-IDS_DF_B7_b_attP_fwd.dna	GTTATGGTTTACTCCATCTA	4025	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTATGGAGTAAACCT	4246
	pU6-Sp-gRNA-IDS_DF_B7_b_attP_rev.dna	GTTATGGTTTACTCCATCTA	4025	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTATGGAGTAAACC	4247
	pU6-Sp-gRNA-IDS_DF_C2_c_attB_fwd.dna	GTTTTGGTTTACCCTATCTA	4026	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCATAGGGTAAACCA	4248
	pU6-Sp-gRNA-IDS_DF_C2_c_attB_rev.dna	GTTTTGGTTTACCCTATCTA	4026	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATATAGGGTAAACCA	4249
	pU6-Sp-gRNA-IDS_DF_D1_b_attB_fwd.dna	GCTGTGGAACTGCAACACACT	4027	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGTGTTGCAGT	4250
	pU6-Sp-gRNA-IDS_DF_D1_b_attB_rev.dna	GCTGTGGAACTGCAACACACT	4027	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGTGTTGCAGT	4251
	pU6-Sp-gRNA-IDS_DF_D2_c_attB_fwd.dna	GTGCCACCTAACAGTGAGCTG	4028	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCTCACTGTTAGGT	4252
	pU6-Sp-gRNA-IDS_DF_D2_c_attB_rev	GTGCCACCTAACAGTG	4028	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCTCAC	4253

[2624]	.dna	AGCTG		TGTTAGGT	

[2625] 用于哺乳动物细胞实验的所有载体均使用PlasmidPlus Midiprep试剂盒

(Qiagen)或PureYield plasmid miniprep试剂盒(Promega),或QIAprep Spin Miniprep试剂盒进行纯化。

[2626] 一般哺乳动物细胞培养条件

[2627] HEK293T(ATCC CRL-3216)、U2OS(ATTC HTB-96)、K562(CCL-243)和HeLa(CCL-2)细胞购自ATCC并在以下培养基中培养和传代:Dulbecco改良Eagle培养基(DMEM)加GlutaMAX(ThermoFisher Scientific)、McCoy's 5A培养基(Gibco)、RPMI培养基1640加GlutaMAX(Gibco),或Eagle最低基础培养基(EMEM,ATCC),分别各自补充有10%(v/v)胎牛血清(Gibco,合格)和1x青霉素链霉素(Coring)。所有细胞类型均在37℃和5%CO_2下温育、维持和培养。细胞系由其各自的供应商鉴定,并且支原体检测呈阴性。

[2628] HEK293T、HeLa和Huh7组织培养转染方案和基因组DNA制备

[2629] 将生长的HEK293T细胞接种于48孔聚D-赖氨酸包被的板(Corning)上。接种后16-24小时,根据制造商的方案,用1μL Lipofectamine 2000(Thermo Fisher Scientific)和750ng PE2质粒、125ngpegRNA 1和125ngpegRNA2(用于双生PE转染);750ngPE2质粒、250ngpegRNA质粒和83ng sgRNA质粒(用于PE3转染);或者,750ng Cas9和125ng sgRNA 1和125ng sgRNA2(用于成对Cas9核酸酶转染)以约60%的汇合度转染细胞。除非另有说明,细胞在转染后培养3天,然后去除培养基,用1xPBS溶液(Thermo Fisher Scientific)清洗细胞,并通过直接向组织培养板的每个孔添加150μL新鲜制备的裂解缓冲液(10mM Tris-HCl,pH 7.5;0.05%SDS;25μg/mL蛋白酶K(ThermoFisher Scientific))提取基因组DNA。基因组DNA混合物在37℃下温育1-2小时,然后在80℃下进行30分钟的酶灭活步骤。用于哺乳动物细胞基因组DNA扩增的引物列于表13。

[2630] 表13.本实施例中使用的引物的序列

[2631]

引物	序列	SEQ ID NO:
HEK3_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG	3811
HEK3_rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC	3812
HEK3_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG	3811

	HEK3_rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC	3812
	DMD_UMI_fwd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNtgctggccagtttactaacaat	4254
	DMD_UMI_fwd2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNcagaaagaagatcttatcccatcttg	4255
	DMD_rev0	TGGAGTTCAGACGTGTGCTCTTCCGATCTggctacttttgttatttgcatt	4256
	AAVS1_AVA1713	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGAGCAGGGTCCCGCTTC	4257
	AAVS1_AVA1717	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGGGGCTCAGTCTGAAGAG	4258
	AAVS1_AVA1651	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCAAGGACTCAAACCCAGA	4259
	AAVS1_AVA1652	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCGTGCGTCAGTTTTACCT	4260
	AAVS1_AVA1653	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACTGCTTCTCCTCTTGGGAA	4261
	AAVS1_AVA1715	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTGCCAGAACCTCTAAGGT	4262
	AAVS1_AVA1655	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATCCTCTCTGGCTCCATCGTA	4263
	AAVS1_AVA1656	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCACTTCAGGACAGCATGTTT	4264
	AAVS1_AVA1707	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCCGGGAACTGCCGCTGGC	4265
	AAVS1_AVA1710	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAGGAGGCCCTCATCTGGCG	4266
[2632]	CCR5_AVA1678	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATCAATGTGAAGCAAATCGCAGC	4267
	CCR5_AVA1679	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCGATTGTCAGGAGGATGATGAA	4268
	CCR5_AVA1680	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTTCTTACTGTCCCCTTCTGGGC	4269
	CCR5_AVA1681	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAACACAGCATGGACGAC	4270
	CCR5_AVA1682	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACAATCGATAGGTACCTGGCTGTC	4271
	CCR5_AVA1683	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCAGCCCCAAGATGACTAT	4272
	IDS_AVA1763	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGAAAACCTGAGCTTGGAGG	4273
	IDS_AVA1764	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTACTCCAGCTTAATGGAAGTGG	4274
	IDS_AVA1765	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGAAGATGTGGAAATGCCTCAC	4275
	IDS_AVA1766	TGGAGTTCAGACGTGTGCTCTTCCGATCTAATCAACATGAAGGGTTGTGTTGT	4276
	IDS_AVA1769	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCCCACACATGCGTTCCTC	4277
	IDS_AVA1770	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCATGAAGGGTTGTTTTTAATTGA	4278
	IDS_UMI_junc1_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNGCCCTGTCCTGTACCTACCT	4279
	IDS_junc2_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCAGTTTGGGGTATGTGCCA	4280
	IDS_universal_rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAAATTTACCCGTGGCAGC	4281

[2633] 对于HeLa细胞，它们以类似于HEK293T细胞的密度生长和接种，并且每孔细胞用190ng PE2_P2A_Blast(杀稻瘟素抗性基因)、31.5ngpegRNA1和31.5ngpegRNA2(用于双生PE转染)与0.75uL HeLamonster试剂(Mirus)进行转染。然后在转染后24小时用10ug/mL处理细胞用于选择。其余的基因组DNA收集步骤与HEK293T细胞相同。将Huh7细胞以150k/孔接种

于24孔聚D-赖氨酸包被的板(Corning)上。接种后16-24小时,根据制造商的方案,用2μL Lipofectamine 2000(Thermo Fisher Scientific)和多达800ng质粒DNA在高汇合度下转染细胞。基因组DNA收集如上述用于HEK293T的进行。

[2634] 基因组DNA样品的高通量DNA测序

[2635] 从基因组DNA样品中扩增感兴趣的基因组位点,并在IlluminaMiSeq上进行测序,如前所述并进行以下修改。简而言之,含有Illumina正向和反向适体的扩增引物(表13)用于第一轮PCR(PCR 1)扩增感兴趣的基因组区域。用各0.5μM正向和反向引物、1μL基因组DNA提取物和12.5μL Phusion U GreenMultiplex PCRMasterMix进行25μL的PCR 1反应。PCR反应如下进行:98℃2分钟,然后[98℃10秒,61℃20秒,72℃30秒]的30个循环,最后72℃延伸2分钟。在第二轮PCR反应(PCR2)中,将独特的Illumina条形码引物对添加到每个样品中。具体地,25μL给定的PCR 2反应包含0.5μM独特的正向和反向Illumina条形码引物对,1μL未纯化的PCR 1反应混合物和12.5μLPhusionU GreenMultiplex PCR 2x MasterMix。条形码PCR 2反应如下进行:98℃2分钟,然后[98℃10秒,61℃20秒,72℃30秒]的12个循环,最后72℃延伸2分钟。PCR产物通过在1.5%琼脂糖凝胶中电泳进行分析评估。PCR 2产物(由共同扩增子合并)使用QIAquick凝胶提取试剂盒(Qiagen)通过1.5%琼脂糖凝胶电泳纯化,用40μL水洗脱。根据制造商的方案,DNA浓度通过荧光定量法(Qubit,ThermoFisher Scientific)或qPCR(KAPA文库定量试剂盒-Illumina,KAPA Biosystems)测量,并在Illumina MiSeq仪器上进行测序。

[2636] 使用MiSeq Reporter(Illumina)对测序读段进行多路分解。使用CRISPResso2进行扩增子序列与参考序列的比对。对于所有引导编辑产率定量,引导编辑效率计算为:[具有期望编辑且不含有插入缺失的读段的数目]%÷[总读段的数目]的%。对于点突变编辑的定量,CRISPResso2在标准模式下运行,并启用“丢弃_插入缺失_读段”。然后用于安装点突变的引导编辑明确计算为:[非丢弃读段中指定点突变的频率]x[非丢弃读段的数目]÷[总读段]。对于编辑定量,CRISPResso2在HDR模式下运行,使用期望的等位基因作为预期的等位基因(e flag),并启用“丢弃_插入缺失_读段”。编辑产率计算为:[非丢弃HDR比对读段的数目]÷[总读段]。插入缺失产率计算为:[含有插入缺失的丢弃读段的数目]÷[总读段]。独特的分子标识符(UMI)用于量化缺失效率和评估这个三步PCR方案的PCR偏倚。简而言之,在线性扩增的第一步中,1uL基因组DNA提取物由含有15bp或16bp的独特分子标识符的0.1uM的单个正向UMI引物和在25uL反应物的12.5uL Phusion U Green Multiplex PCR Master Mix线性扩增。然后用1.6X AmPure珠(Beckman Coulter)纯化PCR产物,并在20uL QIAGEN洗脱缓冲液中洗脱。在第二步中,然后将1或2uL线性扩增的PCR洗脱物在25uL反应物中用各0.5uM正向和反向引物、Phusion U Green Multiplex PCR混合物扩增30个循环。在这种情况下,正向引物与位于UMI引物5’和UMI上游的P5 Illumina适体序列退火。PCR产物由1X ampure珠纯化并在25uL洗脱缓冲液中洗脱。在第三步中,如上所述将1uL洗脱物扩增12个循环,用于添加独特的Illumina条形码。为了评估DMD基因座的大缺失,从1.5%琼脂糖中分别切除顶部条带(未经编辑的大扩增子)和底部条带(具有缺失的经编辑的扩增子),并且将最终文库(60pM顶部条带与20pM底部条带)加载至两个独立的Miseq试剂盒中,以避免大扩增子(约1kb)与短扩增子(约300bp)相比的弱聚类。

[2637] 使用AmpUMI对原始测序读段进行UMI重复删除。对于成对末端的读段,使用SeqKit

将R1与R2的反向互补物连环化(无重叠合并)。连环化R1+R2使用R1的5'端UMI进行UMI重复删除。使用CRISPResso2分析UMI重复删除的R1或连环化R1+R2。对于分析连环化R1+R2,提供了适当的连环化参考扩增子序列以最小化由于连环化引起的测序比对伪影。

[2638] U2OS和K562的核转染

[2639] 在所有实验中使用K562和U2OS细胞进行核转染。每次核转染使用20万个细胞。将计数的细胞旋转下来并用PBS清洗,然后按照Lonza SE Cell Line 4D-NucleofectorKit的建议重新悬浮在核转染溶液中。在装置中核转染细胞后,允许细胞在比色皿中在室温下温育10分钟。将比色皿中的内容物转移到充满预热培养基的48孔板中。如上所述,提取基因组DNA并准备用于Illumina Miseq制备。

[2640] 一锅双生PE和BxbI介导的货物敲入和大倒位:

[2641] 对于单锅介导的大倒位实验,进行了序贯质粒转染和一锅mRNA核转染。在序贯质粒转染实验中,如上所述使用Lipofectamine 2000用750ng PE2、62.5ng各pegRNA(共四种)转染HEK293T细胞。三天后,将细胞胰蛋白酶化并铺板到4孔板中,并且传代约七天。然后再次接种20K细胞,并用500ng BxBI质粒进行转染。如上所述提取基因组DNA并准备用于Illumina Miseq制备。在一锅mRNA核转染实验中,用1ug PE2 mRNA、30pmol各合成pegRNA(共4种pegRNA)和750ngBxBI mRNA核转染200,000个细胞。合成pegRNA是从IDT订购的并且mRNA转录物按照发表的方案进行体外转录。20uLLonza缓冲液和细胞的混合物用程序CM-130进行电穿孔,并用80uL温热的培养基恢复5分钟。然后将比色皿中的25uL样品添加到48孔板的每个孔中并在37℃下温育72小时。

[2642] 本公开中提及的其他参考文献

[2643] 以下参考文献各自通过引用以其整体并入本文。

[2644] 1.Jinek,M.et al.AProgrammable Dual-RNA-Guided DNA Endonuclease in AdaptiveBacterial Immunity.Science337,816-821(2012).

[2645] 2.Cong,L.et al.Multiplex Genome Engineering Using CRISPR/Cas Systems.Science339,819-823(2013).

[2646] 3.Komor,A.C.,Badran,A.H.&Liu,D.R.CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes.Cell168,20-36(2017).

[2647] 4.Komor,A.C.,Kim,Y.B.,Packer,M.S.,Zuris,J.A.&Liu,D.R.Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature533,420-424(2016).

[2648] 5.Nishida,K.et al.Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.Science353,aaf8729(2016).

[2649] 6.Gaudelli,N.M.et al.Programmable base editing of A•Tto G•C in genomic DNA without DNA cleavage.Nature 551,464-471(2017).

[2650] 7.ClinVar,July2019.

[2651] 8.Dunbar,C.E.et al.Gene therapy comes of age.Science359,eaan4672(2018).

[2652] 9.Cox,D.B.T.,Platt,R.J.&Zhang,F.Therapeutic genome editing:prospects andchallenges.Nat.Med.21,121-131(2015).

[2653] 10.Adli,M.The CRISPR tool kit for genome editing and beyond.Nat.Commun.9,1911(2018).

[2654] 11.Kleinstiver,B.P.et al.Engineered CRISPR-Cas9 nucleases with altered PAM specificities.Nature 523,481-485(2015).

[2655] 12.Kleinstiver,B.P.et al.High-fidelity CRISPR-Cas9nucleases with no detectable genome-wide off-target effects.Nature529,490-495(2016).

[2656] 13.Hu,J.H.et al.Evolved Cas9 variants with broad PAM compatibility and high DNA specificity.Nature556,57-63(2018).

[2657] 14.Nishimasu,H.et al.Engineered CRISPR-Cas9 nuclease with expanded targeting space.Science 361,1259-1262(2018).

[2658] 15.Jasin,M.&Rothstein,R.Repair of strand breaks by homologous recombination.Cold Spring Harb.Perspect.Biol.5,a012740(2013).

[2659] 16.Paquet,D.et al.Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.Nature 533,125-129(2016).

[2660] 17.Kosicki,M.,Tomberg,K.&Bradley,A.Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements.Nat.Biotechnol.36,765-771(2018).

[2661] 18.Haapaniemi,E.,Botla,S.,Persson,J.,Schmierer,B.&Taipale,J.CRISPR-Cas9genome editing induces ap53-mediated DNA damage response.Nat.Med.24,927-930(2018).

[2662] 19.Ihry,R.J.et al.p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells.Nat.Med.24,939-946(2018).

[2663] 20.Richardson,C.D.,Ray,G.J.,DeWitt,M.A.,Curie,G.L.&Corn,J.E.Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.Nat.Biotechnol.34,339-344(2016).

[2664] 21.Srivastava,M.et al.An Inhibitor of Nonhomologous End-Joining AbrogatesDouble-Strand Break Repair and Impedes Cancer Progression.Cell 151,1474-1487(2012).

[2665] 22.Chu,V.T.et al.Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.Nat.Biotechnol.33,543-548(2015).

[2666] 23.Maruyama,T.et al.Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.Nat.Biotechnol.33,538-542(2015).

[2667] 24.Kim,Y.B.et al.Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions.Nat.Biotechnol.35,371-376(2017).

[2668] 25.Li,X.et al.Base editing with a Cpf1-cytidine deaminase fusion.Nat.Biotechnol.36,324-327(2018).

[2669] 26.Gehrke,J.M.et al.An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities.Nat.Biotechnol.(2018).doi:10.1038/nbt.4199

[2670] 27.Rees,H.A.&Liu,D.R.Base editing:precision chemistry on the genome and transcriptome of living cells.Nat.Rev.Genet.1(2018).doi:10.1038/s41576-018-0059-1.

[2671] 28.Ostertag,E.M.&Kazazian Jr,H.H.Biology of Mammalian L1 Retrotransposons.Annu.Rev.Genet.35,501-538(2001).

[2672] 29.Zimmerly,S.,Guo,H.,Perlman,P.S.&Lambowltz,A.M.Group II intron mobility occurs by target DNA-primed reverse transcription.Cell 82,545-554(1995).

[2673] 30.Luan,D.D.,Korman,M.H.,Jakubczak,J.L.&Eickbush,T.H.Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site:a mechanism for non-LTR retrotransposition.Cell 72,595-605(1993).

[2674] 31.Feng,Q.,Moran,J.V.,Kazazian,H.H.&Boeke,J.D.Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.Cell87,905-916(1996).

[2675] 32.Jinek,M.et al.Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation.Science 343,1247997(2014).

[2676] 33.Jiang,F.et al.Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.Science aad8282(2016).doi:10.1126/science.aad8282

[2677] 34.Qi,L.S.et al.Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.Cell 152,1173-1183(2013).

[2678] 35.Tang,W.,Hu,J.H.&Liu,D.R.Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation.Nat.Commun.8,15939(2017).

[2679] 36.Shechner,D.M.,Hacisuleyman,E.,Younger,S.T.&Rinn,J.L.Multiplexable,locus-specific targeting of long RNAs with CRISPR-Display.Nat.Methods 12,664-670(2015).

[2680] 37.Anders,C.&Jinek,M.Chapter One-In vitro Enzymology of Cas9.in Methods in Enzymology(eds.Doudna,J.A.&Sontheimer,E.J.)546,1-20(Academic Press,2014).

[2681] 38.Briner,A.E.et al.Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality.Mol.Cell 56,333-339(2014).

[2682] 39.Nowak,C.M.,Lawson,S.,Zerez,M.&Bleris,L.Guide RNA engineering for versatile Cas9functionality.Nucleic Acids Res.44,9555-9564(2016).

[2683] 40.Sternberg,S.H.,Redding,S.,Jinek,M.,Greene,E.C.&Doudna,J.A.DNAinterrogation by the CRISPR RNA-guided endonuclease Cas9.Nature 507,62-67(2014).

[2684] 41.Mohr,S.et al.Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing.RNA 19,958-970(2013).

[2685] 42.Stamos,J.L.,Lentzsch,A.M.&Lambowitz,A.M.Structure ofa Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications.Mol.Cell 68,926-939.e4(2017).

[2686] 43.Zhao,C.&Pyle,A.M.Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution.Nat.Struct.Mol.Biol.23,558-565(2016).

[2687] 44.Zhao,C.,Liu,F.&Pyle,A.M.An ultraprocessive,accurate reverse transcriptase encoded by a metazoan group II intron.RNA 24,183-195(2018).

[2688] 45.Ran,F.A.et al.Genome engineering using the CRISPR-Cas9 system.Nat.Protoc.8,2281-2308(2013).

[2689] 46.Liu,Y.,Kao,H.-I.&Bambara,R.A.Flap endonuclease 1:a central component of DNA metabolism.Annu.Rev.Biochem.73,589-615(2004).

[2690] 47.Krokan,H.E.& Bjørås,M.Base Excision Repair.ColdSpring Harb.Perspect.Biol.5,(2013).

[2691] 48.Kelman,Z.PCNA:structure,functions and interactions.Oncogene 14,629-640(1997).

[2692] 49.Choe,K.N.&Moldovan,G.-L.Forging Ahead through Darkness:PCNA,Still the Principal Conductor at the Replication Fork.Mol.Cell65,380-392(2017).

[2693] 50.Li,X.,Li,J.,Harrington,J.,Lieber,M.R.&Burgers,P.M.Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation ofFEN-1 by proliferating cell nuclear antigen.J.Biol.Chem.270,22109-22112(1995).

[2694] 51.Tom,S.,Henricksen,L.A.&Bambara,R.A.Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1.J.Biol.Chem.275,10498-10505(2000).

[2695] 52.Tanenbaum,M.E.,Gilbert,L.A.,Qi,L.S.,Weissman,J.S.&Vale,R.D.Aprotein-tagging system for signal amplification in gene expression and fluorescence imaging.Cell 159,635-646(2014).

[2696] 53.Bertrand,E.et al.Localization ofASH1 mRNA particles in living yeast.Mol.Cell 2,437-445(1998).

[2697] 54.Dahlman,J.E.et al.Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease.Nat.Biotechnol.33,1159-1161(2015).

[2698] 55.Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nat.Biotechnol.33,187-197(2015).

[2699] 56.Tsai,S.Q.et al.CIRCLE-seq:a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets.Nat.Methods 14,607-614(2017).

[2700] 57.SchekN,Cooke C,Alwine JC.Molecular and Cellular Biology.(1992).

[2701] 58.Gil A,Proudfoot NJ.Cell.(1987).

[2702] 59.Zhao,B.S.,Roundtree,I.A.,He,C.Nat Rev Mol Cell Biol.(2017).

[2703] 60.Rubio,M.A.T.,Hopper,A.K.Wiley Interdiscip Rev RNA(2011).

[2704] 61.Shechner,D.M.,Hacisuleyman E.,Younger,S.T.,Rinn,J.L.Nat Methods.(2015).

[2705] 62.Paige,J.S.,Wu,K.Y.,Jaffrey,S.R.Science(2011).

[2706] 63.Ray D.,…Hughes TR.Nature(2013).

[2707] 64.Chadalavada,D.M.,Cerrone-Szakal,A.L.,Bevilacqua,P.C.RNA(2007).

[2708] 65.Forster AC,Symons RH.Cell.(1987).

[2709] 66.Weinberg Z,Kim PB,Chen TH,Li S,Harris KA,Lünse CE,Breaker RR.Nat.Chem.Biol.(2015).

[2710] 67.Feldstein PA,Buzayan JM,Bruening G.Gene(1989).

[2711] 68.Saville BJ,Collins RA.Cell.(1990).

[2712] 69.Winkler WC,Nahvi A,Roth A,Collins JA,Breaker RR.Nature(2004).

[2713] 70.Roth A,Weinberg Z,Chen AG,Kim PG,Ames TD,Breaker RR.Nat Chem Biol.(2013).

[2714] 71.Choudhury R,Tsai YS,Dominguez D,WangY,WangZ.Nat Commun.(2012).

[2715] 72.MacRae IJ,Doudna JA.CurrOpin Struct Biol.(2007).

[2716] 73.Bernstein E,Caudy AA,Hammond SM,Hannon GJ Nature(2001).

[2717] 74.Filippov V,Solovyev V,Filippova M,Gill SS.Gene(2000).

[2718] 75.Cadwell RC and Joyce GF.PCR Methods Appl.(1992).

[2719] 76.McInerney P,Adams P,andHadi MZ.Mol Biol Int.(2014).

[2720] 77.Esvelt KM,Carlson JC,and Liu DR.Nature.(2011).

[2721] 78.Naorem SS,Hin J,Wang S,Lee WR,Heng X,Miller JF,Guo H.ProcNatl Acad Sci USA(2017).

[2722] 79.Martinez MA,Vartanian JP,Wain-Hobson S.Proc Natl Acad Sci USA(1994).

[2723] 80.Meyer AJ,Ellefson JW,Ellington AD.CurrProtoc Mol Biol.(2014).

[2724] 81.Wang HH,Isaacs FJ,Carr PA,Sun ZZ,Xu G,Forest CR,Church GM.Nature.(2009).

[2725] 82.NyergesÁet al.Proc Natl Acad Sci USA.(2016).

[2726] 83.Mascola JR,Haynes BF.Immunol Rev.(2013).

[2727] 84.X.Wen,K.Wen,D.Cao,G.Li,R.W.Jones,J.Li,S.Szu,Y.Hoshino,L.Yuan,Inclusion of a universal tetanus toxoidCD4(+)T cell epitope P2 significantly enhanced the immunogenicity of recombinant rotavirusΔVP8*subunit parenteral vaccines.Vaccine 32,4420-4427(2014).

[2728] 85.G.Ada,D.Isaacs,Carbohydrate-protein conjugate vaccines.Clin Microbiol Infect 9,79-85(2003).

[2729] 86.E.Malito,B.Bursulaya,C.Chen,P.L.Surdo,M.Picchianti,E.Balducci,M.Biancucci,A.Brock,F.Berti,M.J.Bottomley,M.Nissum,P.Costantino,R.Rappuoli,G.Spraggon,Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197.Proceedings of the National Academy of Sciences 109,5229(2012).

[2730] 87.J.de Wit,M.E.Emmelot,M.C.M.Poelen,J.Lanfermeijer,W.G.H.Han,C.van Els,P.Kaaijk,The Human CD4(+)T Cell Response against Mumps Virus Targets aBroadly Recognized Nucleoprotein Epitope.J Virol 93,(2019).

[2731] 88.M.May,C.A.Rieder,R.J.Rowe,Emergent lineages of mumps virus suggest the need fora polyvalent vaccine.Int J Infect Dis 66,1-4(2018).

[2732] 89.M.Ramamurthy,P.Rajendiran,N.Saravanan,S.Sankar,S.Gopalan,B.Nandagopal,Identification of immunogenic B-cell epitope peptides of rubella virus E1 glycoprotein towards development of highly specific immunoassays and/or vaccine.Conference Abstract,(2019).

[2733] 90.U.S.F.Tambunan,F.R.P.Sipahutar,A.A.Parikesit,D.Kerami,Vaccine Design for H5N1 Based on B- and T-cell Epitope Predictions.Bioinform Biol Insights 10,27-35(2016).

[2734] 91.Asante,EA.et.al."A naturally occurring variant of the human prion protein completely prevents prion disease".Nature.(2015).

[2735] 92.Crabtree,G.R.&Schreiber,S.L.Three-part inventions:intracellular signaling and induced proximity.Trends Biochem.Sci.21,418-22(1996).

[2736] 93.Liu,J.et al.Calcineurin Is aCommon Target of A and FKBP-FK506 Complexes.Cell66,807-815(1991).

[2737] 94.Keith,C.T.etal.A mammalian protein targeted by G1-arresting rapamycin-receptor complex.Nature 369,756-758(2003).

[2738] 95.Spencer,D.M.,Wandless,T.J.,Schreiber,S.L.S.&Crabtree,G.R.Controlling signal transduction with synthetic ligands.Science 262,1019-24(1993).

[2739] 96.Pruschy,M.N.et al.Mechanistic studies of a signaling pathway activated by the organic dimerizer FK1012.Chem.Biol.1,163-172(1994).

[2740] 97.Spencer,D.M.et al.Functional analysis of Fas signaling in vivo using synthetic inducers of dimerization.Curr.Biol.6,839-847(1996).

[2741] 98.Belshaw,P.J.,Spencer,D.M.,Crabtree,G.R.&Schreiber,S.L.Controlling programmed cell deathwith acyclophilin-cyclosporin-based chemical inducer of dimerization.Chem.Biol.3,731-738(1996).

[2742] 99.Yang,J.X.,Symes,K.,Mercola,M.&Schreiber,S.L.Small-molecule control of insulin and PDGF receptor signaling and the role of membrane attachment.Curr.Biol.8,11-18(1998).

[2743] 100.Belshaw,P.J.,Ho,S.N.,Crabtree,G.R.&Schreiber,S.L.Controlling protein association and subcellular localization with a synthetic ligand that

induces heterodimerization of proteins.Proc.Natl.Acad.Sci.93,4604-4607(2002).

[2744] 101.Stockwell,B.R.&Schreiber,S.L.Probing the role of homomeric and heteromeric receptor interactions in TGF-βsignaling using small molecule dimerizers.Curr.Biol.8,761-773(2004).

[2745] 102.Spencer,D.M.,Graef,I.,Austin,D.J.,Schreiber,S.L.&Crabtree,G.R.A general strategy for producing conditional alleles of Src-like tyrosine kinases.Proc.Natl.Acad.Sci.92,9805-9809(2006).

[2746] 103.Holsinger,L.J.,Spencer,D.M.,Austin,D.J.,Schreiber,S.L.&Crabtree,G.R.Signal transduction in T lymphocytes using a conditional allele of Sos.Proc.Natl.Acad.Sci.92,9810-9814(2006).

[2747] 104.Myers,M.G.Insulin Signal Transduction and the IRS Proteins.Annu.Rev.Pharmacol.Toxicol.36,615-658(1996).

[2748] 105.Watowich,S.S.The erythropoietin receptor:Molecular structure and hematopoietic signaling pathways.J.Investig.Med.59,1067-1072(2011).

[2749] 106.Blau,C.A.,Peterson,K.R.,Drachman,J.G.&Spencer,D.M.A proliferation switch for genetically modified cells.Proc.Natl.Acad.Sci.94,3076-3081(2002).

[2750] 107.Clackson,T.et al.Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity.Proc.Natl.Acad.Sci.95,10437-10442(1998).

[2751] 108.Diver,S.T.&Schreiber,S.L.Single-step synthesis of cell-permeable protein dimerizers that activate signal transduction and gene expression.J.Am.Chem.Soc.119,5106-5109(1997).

[2752] 109.Guo,Z.F.,Zhang,R.&Liang,F.Sen.Facile functionalization of FK506 for biological studies by the thiol-ene‘click’reaction.RSC Adv.4,11400-11403(2014).

[2753] 110.Robinson,D.R.,Wu,Y.-M.&Lin,S.-F.The protein tyrosine kinase family ofthe human genome.Oncogene 19,5548-5557(2000).

[2754] 111.Landrum,M.J.et al.ClinVar:public archive of interpretations of clinically relevant variants.Nucleic Acids Res.44,D862-D868(2016).

[2755] 112.Jinek,M.et al.A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.Science 337,816-821(2012).

[2756] 113.Cong,L.et al.Multiplex Genome Engineering Using CRISPR/Cas Systems.Science 339,819-823(2013).

[2757] 114.Mali,P.et al.RNA-Guided Human Genome Engineering via Cas9.Science 339,823-826(2013).

[2758] 115.Yang,H.et al.One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering.Cell 154,1370-1379(2013).

[2759] 116.Kim,S.,Kim,D.,Cho,S.W.,Kim,J.&Kim,J.-S.Highly efficient RNA-

guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.Genome Res.24,1012-1019(2014).

[2760] 117.Orlando,S.J.et al.Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology.Nucleic Acids Res.38,e152-e152(2010).

[2761] 118.Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nat.Biotechnol.33,187-197(2015).

[2762] 119.Suzuki,K.et al.In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration.Nature 540,144-149(2016).

[2763] 120.Kosicki,M.,Tomberg,K.&Bradley,A.Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements.Nat.Biotechnol.36,765-771(2018).

[2764] 121.Haapaniemi,E.,Botla,S.,Persson,J.,Schmierer,B.&Taipale,J.CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response.Nat.Med.24,927-930(2018).

[2765] 122.Ihry,R.J.et al.p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells.Nat.Med.24,939-946(2018).

[2766] 123.Chapman,J.R.,Taylor,M.R.G.&Boulton,S.J.Playing the end game:DNA double-strand break repair pathway choice.Mol.Cell 47,497-510(2012).

[2767] 124.Cox,D.B.T.,Platt,R.J.&Zhang,F.Therapeutic genome editing:prospects and challenges.Nat.Med.21,121-131(2015).

[2768] 125.Paquet,D.et al.Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.Nature 533,125-129(2016).

[2769] 126.Chu,V.T.et al.Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.Nat.Biotechnol.33,543-548(2015).

[2770] 127.Maruyama,T.et al.Increasing the efficiency ofprecise genome editing with CRISPR-Cas9by inhibition of nonhomologous endjoining.Nat.Biotechnol.33,538-542(2015).

[2771] 128.Rees,H.A.,Yeh,W.-H.&Liu,D.R.Development ofhRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks.Nat.Commun.10,1-12(2019).

[2772] 129.Komor,A.C.,Kim,Y.B.,Packer,M.S.,Zuris,J.A.&Liu,D.R.Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424(2016).

[2773] 130.Gaudelli,N.M.et al.Programmable base editing ofA•T to G•C in genomic DNA without DNA cleavage.Nature 551,464-471(2017).

[2774] 131.Gao,X.et al.Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents.Nature 553,217-221(2018).

[2775] 132.Ingram,V.M.A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin.Nature 178,792-794(1956).

[2776] 133.Myerowitz,R.&Costigan,F.C.The major defect in Ashkenazi Jews with Tay-Sachs disease is an insertion in the gene for the alpha-chain of beta-hexosaminidase.J.Biol.Chem.263,18587-18589(1988).

[2777] 134.Zielenski,J.Genotype and Phenotype in Cystic Fibrosis.Respiration 67,117-133(2000).

[2778] 135.Mead,S.et al.A Novel Protective Prion Protein Variant that Colocalizes with Kuru Exposure.N.Engl.J.Med.361,2056-2065(2009).

[2779] 136.Marraffini,L.A.&Sontheimer,E.J.CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.Science 322,1843-1845(2008).

[2780] 137.Barrangou,R.et al.CRISPR provides acquired resistance against viruses in prokaryotes.Science 315,1709-1712(2007).

[2781] 138.Jiang,F.&Doudna,J.A.CRISPR-Cas9 Structures and Mechanisms.Annu.Rev.Biophys.46,505-529(2017).

[2782] 139.Hille,F.et al.The Biology ofCRISPR-Cas:Backward and Forward.Cell 172,1239-1259(2018).

[2783] 140.Luan,D.D.,Korman,M.H.,Jakubczak,J.L.&Eickbush,T.H.Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTRretrotransposition.Cell 72,595-605(1993).

[2784] 141.Liu,Y.,Kao,H.-I.&Bambara,R.A.Flap endonuclease 1:a central component of DNA metabolism.Annu.Rev.Biochem.73,589-615(2004).

[2785] 142.Rees,H.A.&Liu,D.R.Base editing:precision chemistry on the genome and transcriptome of living cells.Nat.Rev.Genet.19,770(2018).

[2786] 143.Richardson,C.D.,Ray,G.J.,DeWitt,M.A.,Curie,G.L.&Corn,J.E.Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.Nat.Biotechnol.34,339-344(2016).

[2787] 144.Qi,L.S.et al.Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.Cell 152,1173-1183(2013).

[2788] 145.Shechner,D.M.,Hacisuleyman,E.,Younger,S.T.&Rinn,J.L.Multiplexable,locus-specific targeting of long RNAs with CRISPR-Display.Nat.Methods 12,664-670(2015).

[2789] 146.Tang,W.,Hu,J.H.&Liu,D.R.Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation.Nat.Commun.8,15939(2017).

[2790] 147.Jinek,M.et al.Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation.Science 343,1247997(2014).

[2791] 148.Nishimasu,H.et al.Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA.Cell 156,935-949(2014).

[2792] 149.Jiang,F.,Zhou,K.,Ma,L.,Gressel,S.&Doudna,J.A.A Cas9-guide RNA complex preorganized for target DNA recognition.Science 348,1477-1481(2015).

[2793] 150.Baranauskas,A.et al.Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants.Protein Eng.Des.Sel.25,657-668(2012).

[2794] 151.Gerard,G.F.et al.The role of template-primer in protection of reverse transcriptase from thermal inactivation.Nucleic Acids Res.30,3118-3129(2002).

[2795] 152.Arezi,B.&Hogrefe,H.Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer.Nucleic Acids Res.37,473-481(2009).

[2796] 153.Kotewicz,M.L.,Sampson,C.M.,D'Alessio,J.M.&Gerard,G.F.Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity.Nucleic Acids Res.16,265-277(1988).

[2797] 154.Shen,M.W.et al.Predictable and precise template-free CRISPR editing of pathogenic variants.Nature 563,646-651(2018).

[2798] 155.Thuronyi,B.W.et al.Continuous evolution of base editors with expanded target compatibility and improved activity.Nat.Biotechnol.(2019).doi:10.1038/s41587-019-0193-0

[2799] 156.Kim,Y.B.et al.Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions.Nat.Biotechnol.35,371-376(2017).

[2800] 157.Koblan,L.W.et al.Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction.Nat.Biotechnol.(2018).doi:10.1038/nbt.4172

[2801] 158.Komor,A.C.et al.Improved base excision repair inhibition and bacteriophage Mu Gamprotein yields C:G-to-T:A base editors with higher efficiency and product purity.Sci.Adv.3,eaao4774(2017).

[2802] 159.Kleinstiver,B.P.et al.High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.Nature 529,490-495(2016).

[2803] 160.Zuo,E.et al.Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos.Science 364,289-292(2019).

[2804] 161.Jin,S.et al.Cytosine,but not adenine,base editors induce genome-wide off-target mutations in rice.Science 364,292-295(2019).

[2805] 162.Kim,D.,Kim,D.,Lee,G.,Cho,S.-I.&Kim,J.-S.Genome-wide target specificity ofCRISPR RNA-guided adenine base editors.Nat.Biotechnol.37,430-435(2019).

[2806] 163.Grünewald,J.et al.Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors.Nature 569,433-437(2019).

[2807] 164.Zhou,C.et al.Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis.Nature 571,275-278(2019).

[2808] 165.Rees,H.A.,Wilson,C.,Doman,J.L.&Liu,D.R.Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci.Adv.5,eaax5717(2019).

[2809] 166.Ostertag,E.M.&Kazazian Jr,H.H.Biology of Mammalian L1 Retrotrans posons.Annu.Rev.Genet.35,501-538(2001).

[2810] 167.Griffiths,D.J.Endogenous retroviruses in the human genome sequence.Genome Biol.2,REVIEWS1017(2001).

[2811] 168.Berkhout,B.,Jebbink,M.&Zsíros,J.Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus.J.Virol.73,2365-2375(1999).

[2812] 169.Halvas,E.K.,Svarovskaia,E.S.&Pathak,V.K.Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity.J.Virol.74,10349-10358(2000).

[2813] 170.Dever,D.P.et al.CRISPR/Cas9 Beta-globin Gene Targeting in Human Hematopoietic Stem Cells.Nature 539,384-389(2016).

[2814] 171.Park,S.H.et al.Highly efficient editing of theβ-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease.Nucleic Acids Res.doi:10.1093/nar/gkz475

[2815] 172.Collinge,J.Prion diseases of humans and animals:their causes and molecular basis.Annu.Rev.Neurosci.24,519-550(2001).

[2816] 173.Asante,E.A.et al.A naturally occurring variant of the human prion protein completely prevents prion disease.Nature 522,478-481(2015).

[2817] 174.Anzalone,A.V.,Lin,A.J.,Zairis,S.,Rabadan,R.&Cornish,V.W.Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches.Nat.Methods 13,453-458(2016).

[2818] 175.Badran,A.H.et al.Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance.Nature 533,58-63(2016).

[2819] 176.Anders,C.&Jinek,M.Chapter One-In Vitro Enzymology of Cas9.in Methods inEnzymology(eds.Doudna,J.A.&Sontheimer,E.J.)546,1-20(Academic Press,2014).

[2820] 177.Pirakitikulr,N.,Ostrov,N.,Peralta-Yahya,P.&Cornish,V.W.PCRless library mutagenesis via oligonucleotide recombination in yeast.Protein Sci.Publ.Protein Soc.19,2336-2346(2010).

[2821] 178.Clement,K.et al.CRISPResso2 provides accurate and rapid genome editing sequence analysis.Nat.Biotechnol.37,224-226(2019).

[2822] 179.Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-

target cleavage by CRISPR-Cas nucleases.Nat.Biotechnol.33,187-197(2015).

[2823] 180.Kleinstiver,B.P.et al.High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.Nature 529,490-495(2016).

[2824] 181.Koblan,L.W.et al.Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction.Nat.Biotechnol.(2018).doi:10.1038/nbt.4172

[2825] 182.Baranauskas,A.et al.Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants.Protein Eng.Des.Sel.25,657-668(2012).

[2826] 183.Schechner,DM,Hacisuleyman E.,Younger ST,Rinn JL.Nat Methods 664-70(2015).

[2827] 184.Brown JA,et al.Nat Struct Mol Biol 633-40(2014).

[2828] 185.Conrad NA and Steitz JA.EMBO J 1831-41(2005).

[2829] 186.Bartlett JS,et al.Proc Natl Acad Sci USA 8852-7(1996).

[2830] 187.Mitton-Fry RM,DeGregorio SJ,Wang J,Steitz TA,Steitz JA.Science 1244-7(2010).

[2831] 188.Forster AC,Symons RH.Cell.1987.

[2832] 189.Weinberg Z,Kim PB,Chen TH,Li S,Harris KA,Lünse CE,Breaker RR.Nat.Chem.Biol.2015.

[2833] 190.Feldstein PA,Buzayan JM,Bruening G.Gene 1989.

[2834] 191.Saville BJ,Collins RA.Cell.1990.

[2835] 192.Roth A,Weinberg Z,Chen AG,Kim PG,Ames TD,Breaker RR.Nat Chem Biol.2013.

[2836] 193.Borchardt EK,et al.RNA 1921-30(2015).

[2837] 194.Zhang Y,et al.Mol Cell 792-806(2013).

[2838] 195.Dang Y,et al.Genome Biol 280(2015).

[2839] 196.Schaefer M,Kapoor U,and Jantsch MF.Open Biol 170077(2017).

[2840] 197.Nahar S,et al.Chem Comm 2377-80(2018).

[2841] 198.Gao Y and Zhao Y.J Integr Plant Biol 343-9(2014).

[2842] 199.Dubois N,Marquet R,Paillart J,Bernacchi S.Front Microbiol 527(2018).

[2843] 200.Costa M and Michel F.EMBO J 1276-85(1995).

[2844] 201.Hu JH,et al.Nature 57-63(2018).

[2845] 202.Furukawa K,Gu H,Breaker RR.Methods Mol Biol 209-20(2014).

[2846] 203.Zettler,J.,Schütz,V.&Mootz,H.D.The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction.FEBS Lett.583,909-914(2009).

[2847] 204.Kügler,S.,Kilic,E.&Bähr,M.Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral

vector in the adult rat brain depending on the transduced area.Gene Ther.10,337-347(2003).

[2848] 205.de Felipe,P.,Hughes,L.E.,Ryan,M.D.&Brown,J.D.Co-translational,intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide.J.Biol.Chem.278,11441-11448(2003).

[2849] 206.Levy,J.M.&Nicoll,R.A.Membrane-associated guanylate kinase dynamics reveal regional and developmental specificity of synapse stability.J.Physiol.595,1699-1709(2017).

[2850] 207.Li,B.&Dewey,C.N.RSEM:accurate transcript quantification from RNA-Seq data with or without a reference genome.BMC Bioinformatics 12,323(2011).

[2851] 208.Ritchie,M.E.etal.limma powers differential expressionanalyses for RNA-sequencing and microarray studies.Nucleic Acids Res.43,e47-e47(2015).

[2852] 实施方案

[2853] 除非上下文中有相反说明或以其他方式明显看出,冠词如“一个”、“一种”和“该”可表示一个或多于一个。除非上下文中有相反说明或以其他方式明显看出,如果一个、多于一个或所有的组成员存在于、被用于或以其他方式与给定的产品或方法相关,则认为在组的一个或多个成员之间包括“或”的实施方案或描述是满足的。本发明包括这样的实施方案,其中该组中的一个成员正好存在于、被用于或以其他方式与给定的产品或方法相关。本发明包括这样的实施方案,其中多于一个或所有的组成员存在于、被用于或以其他方式与给定的产品或方法相关。

[2854] 此外,本公开涵盖所有变化、组合和排列,其中将来自一个或多个所列权利要求的一个或多个限制、元素、条款和描述性术语引入另一个权利要求中。例如,从属于另一个权利要求的任何权利要求可修改为包括在从属于同一基本权利要求的任何其他权利要求中存在的一个或多个限制。在元素以列表形式呈现的情况下,例如,以马库什组形式,还公开了元素的每个亚组,并且可从组中去除任何元素。应当理解,一般而言,在本发明或本发明的方面被称为包括特定要素和/或特征的情况下,本公开的某些实施方案或本公开的方面包括或基本上包括这样的元素和/或特征。为简单起见,这些实施方案并未在本文中具体阐述。还应当注意,术语“包括”和“包含”旨在是开放的并且允许包括额外的元素或步骤。在给出范围的地方包括端点。此外,除非根据上下文和本领域普通技术人员的理解另有说明或以其他方式明显看出,表示为范围的值可推定本发明的不同实施方案中所述范围内的任何特定值或子范围,至范围下限单位的十分之一,除非上下文另有明确规定。

[2855] 本申请涉及各种已授权的专利、公开的专利申请、期刊文章和其他出版物,所有这些都通过引用并入本文。如果任何并入的参考文献与本说明书之间存在冲突,则以本说明书为准。此外,属于现有技术的本发明的任何特定实施方案可以明确地排除在任何一个或多个实施方案之外。因为这样的实施方案被认为是本领域普通技术人员已知的,所以即使在本文中没有明确阐述排除,它们也可被排除。可出于任何原因将本发明的任何特定实施方案排除在任何实施方案之外,无论是否与现有技术的存在有关。

[2856] 本领域技术人员将认识到或能够仅使用常规实验来确定本文描述的特定实施方案的许多等效物。本文描述的本实施方案的范围不旨在限于上述描述,而是如所附实施方

案中所阐述。本领域普通技术人员将理解,在不脱离本发明的精神或范围的情况下,可对本说明进行各种改变和修改,如以下实施方案中所定义。

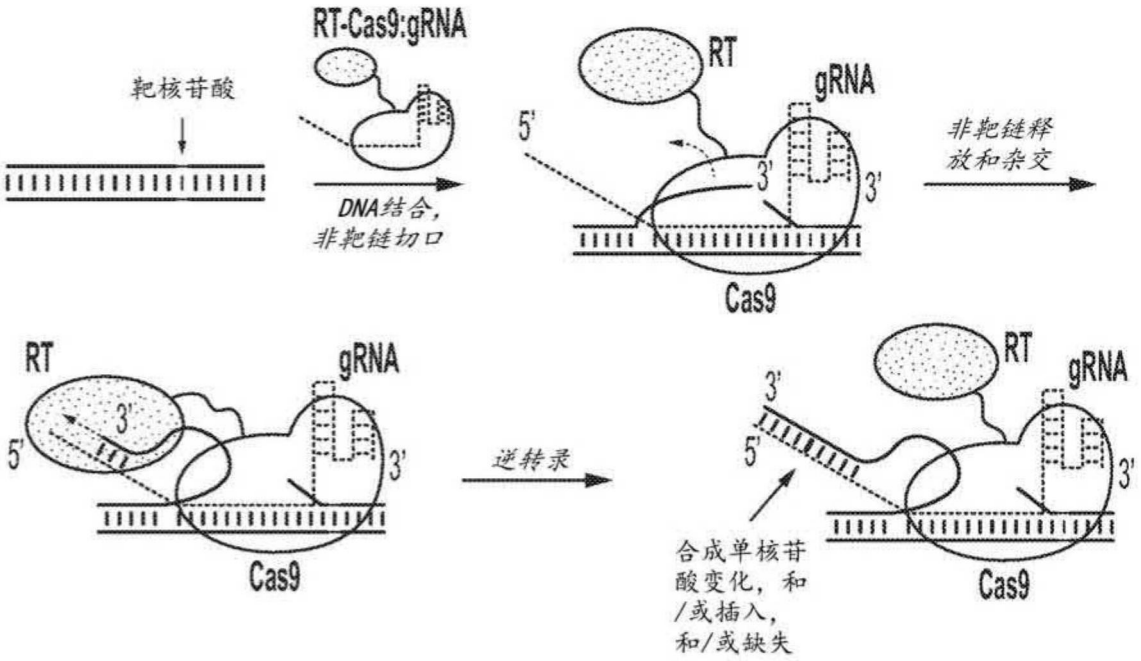

RT-Cas9:gRNA
靶核苷酸
DNA结合，
非靶链切口
RT
gRNA
5'
3'
3'
Cas9
非靶链释
放和杂交
RT
gRNA
5'
3'
3'
Cas9
逆转录
RT
gRNA
3'
5'
3'
Cas9
合成单核苷
酸变化，和
/或插入，
和/或缺失

图1A

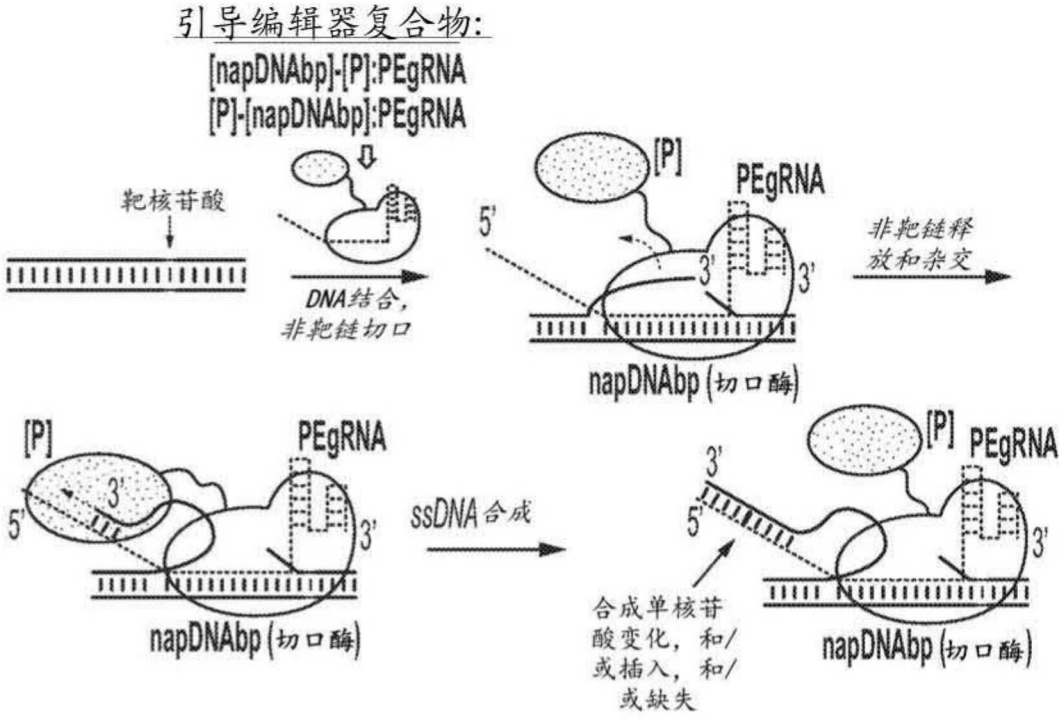

引导编辑器复合物:
[napDNAbp]-[P]:PEgRNA
[P]-[napDNAbp]:PEgRNA
靶核苷酸
DNA结合，
非靶链切口
[P]
PEgRNA
5'
3'
3'
napDNAbp (切口酶)
非靶链释
放和杂交
[P]
PEgRNA
5'
3'
3'
napDNAbp (切口酶)
ssDNA合成
[P]
PEgRNA
3'
5'
3'
napDNAbp (切口酶)
合成单核苷
酸变化，和/
或插入，和/
或缺失

图1B

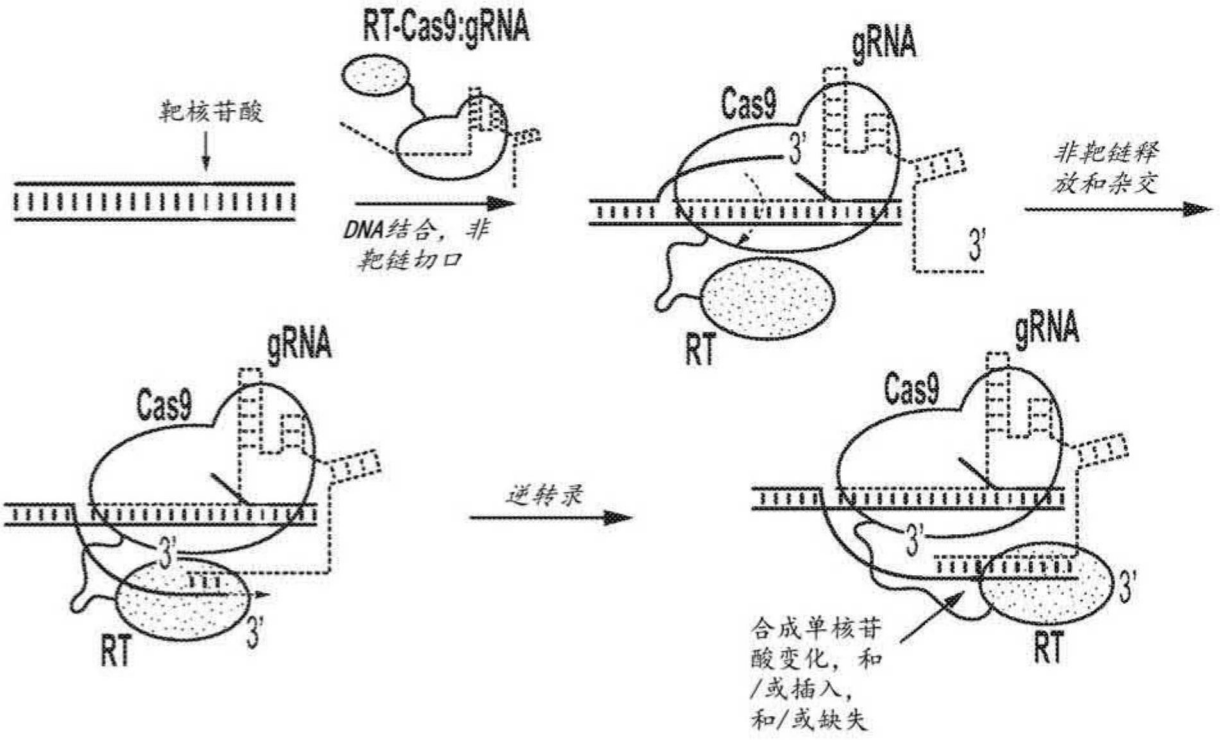

RT-Cas9:gRNA
靶核苷酸
DNA结合，非靶链切口
gRNA
Cas9
3'
非靶链释放和杂交
RT
逆转录
合成单核苷酸变化，和/或插入，和/或缺失

图1C

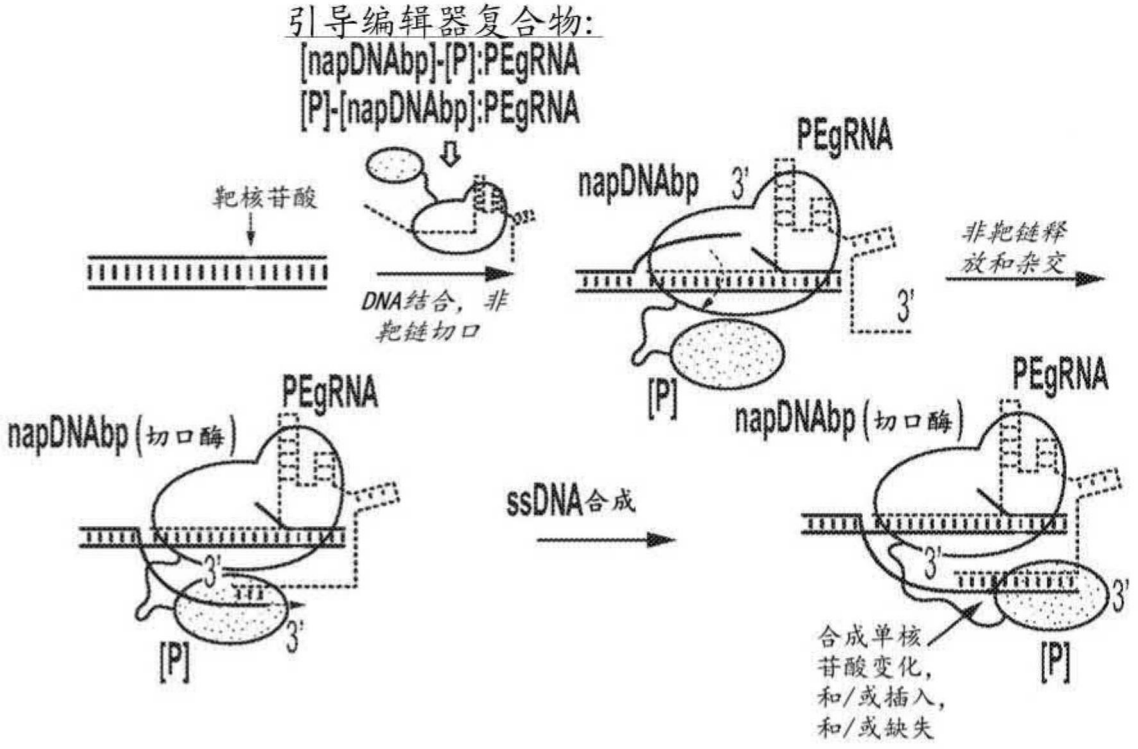

引导编辑器复合物:
[napDNAbp]-[P]:PEgRNA
[P]-[napDNAbp]:PEgRNA
靶核苷酸
DNA结合，非靶链切口
PEgRNA
napDNAbp
3'
非靶链释放和杂交
[P]
napDNAbp(切口酶)
ssDNA合成
合成单核苷酸变化，和/或插入，和/或缺失

图1D

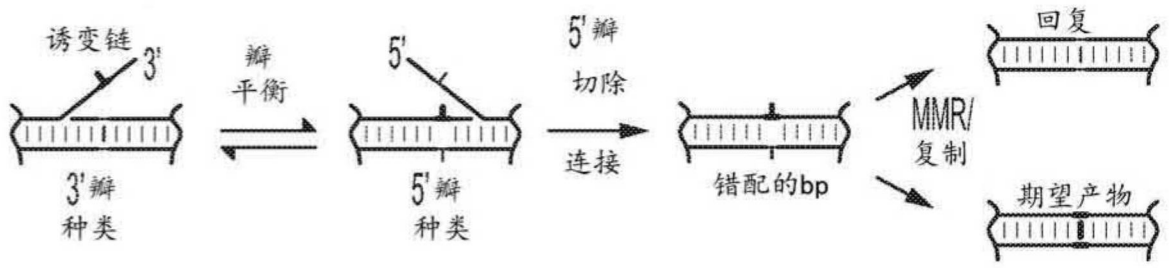

诱变链
3'
3'瓣
种类
瓣
平衡
5'
5'瓣
种类
5'瓣
切除
连接
错配的bp
MMR/
复制
回复
期望产物

图1E

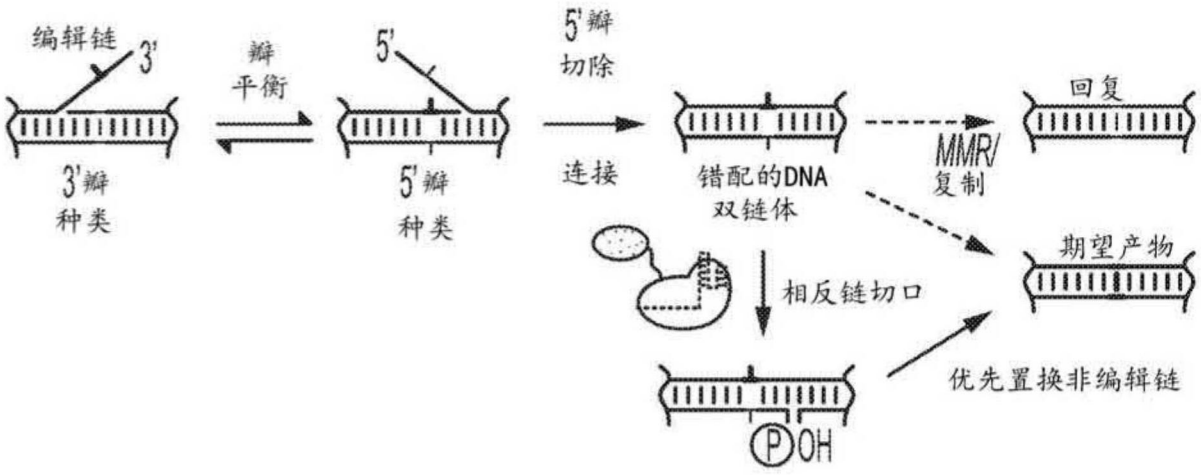

编辑链
3'
3'瓣
种类
瓣
平衡
5'
5'瓣
种类
5'瓣
切除
连接
错配的DNA
双链体
MMR/
复制
回复
期望产物
相反链切口
优先置换非编辑链
POH

图1F

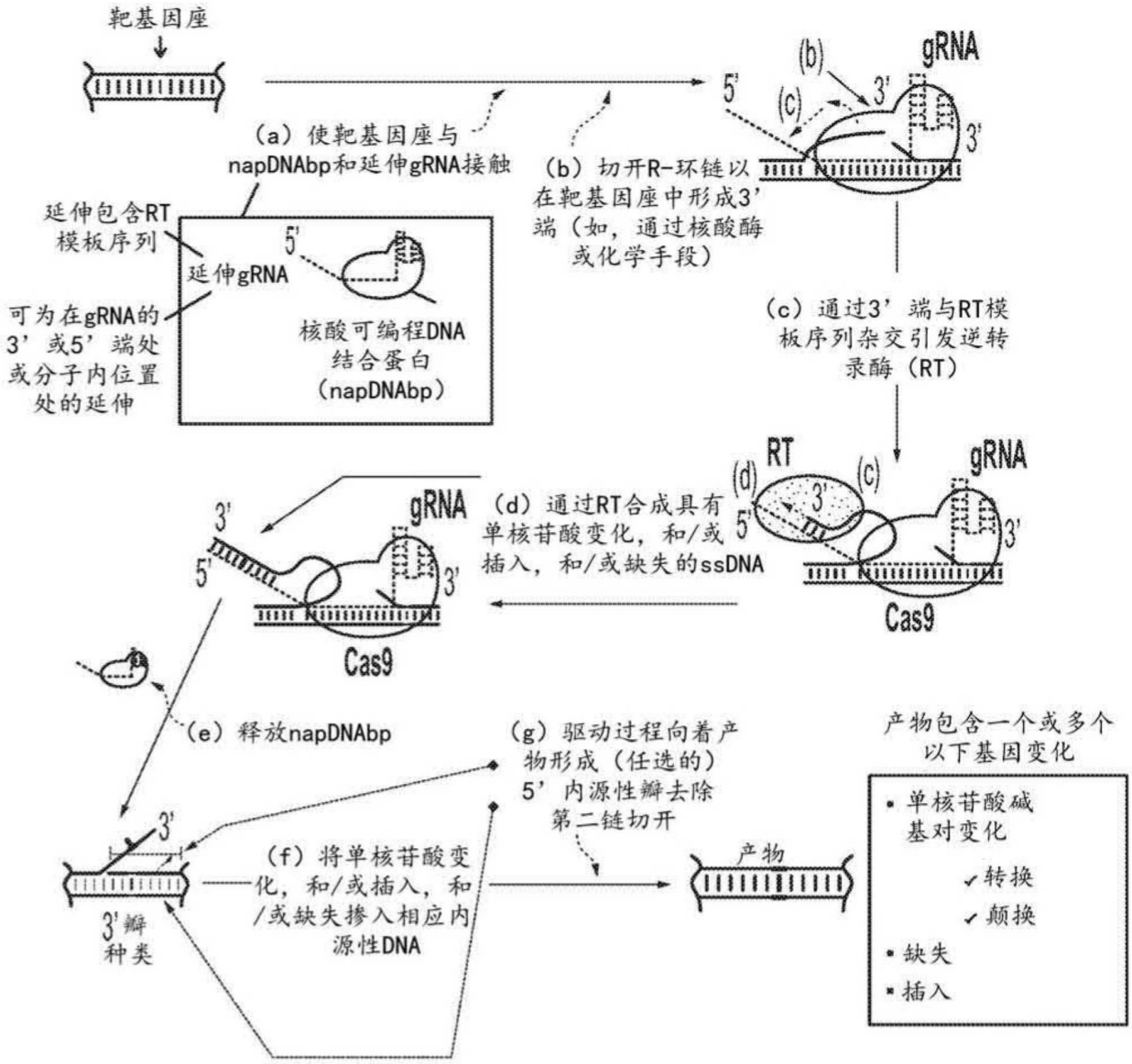

靶基因座
(a) 使靶基因座与napDNAbp和延伸gRNA接触
(b) 切开R-环链以在靶基因座中形成3'端（如，通过核酸酶或化学手段）
gRNA
(b)
(c)
5'
3'
3'
延伸包含RT模板序列
延伸gRNA
5'
可为在gRNA的3'或5'端处或分子内位置处的延伸
核酸可编程DNA结合蛋白（napDNAbp）
(c) 通过3'端与RT模板序列杂交引发逆转录酶（RT）
RT
(d)
(c)
gRNA
5'
3'
3'
Cas9
(d) 通过RT合成具有单核苷酸变化，和/或插入，和/或缺失的ssDNA
3'
5'
gRNA
3'
Cas9
(e) 释放napDNAbp
(g) 驱动过程向着产物形成（任选的）5'内源性瓣去除第二链切开
产物包含一个或多个以下基因变化
• 单核苷酸碱基对变化
✓转换
✓颠换
• 缺失
• 插入
3'
3'瓣种类
(f) 将单核苷酸变化，和/或插入，和/或缺失掺入相应内源性DNA
产物

图1G

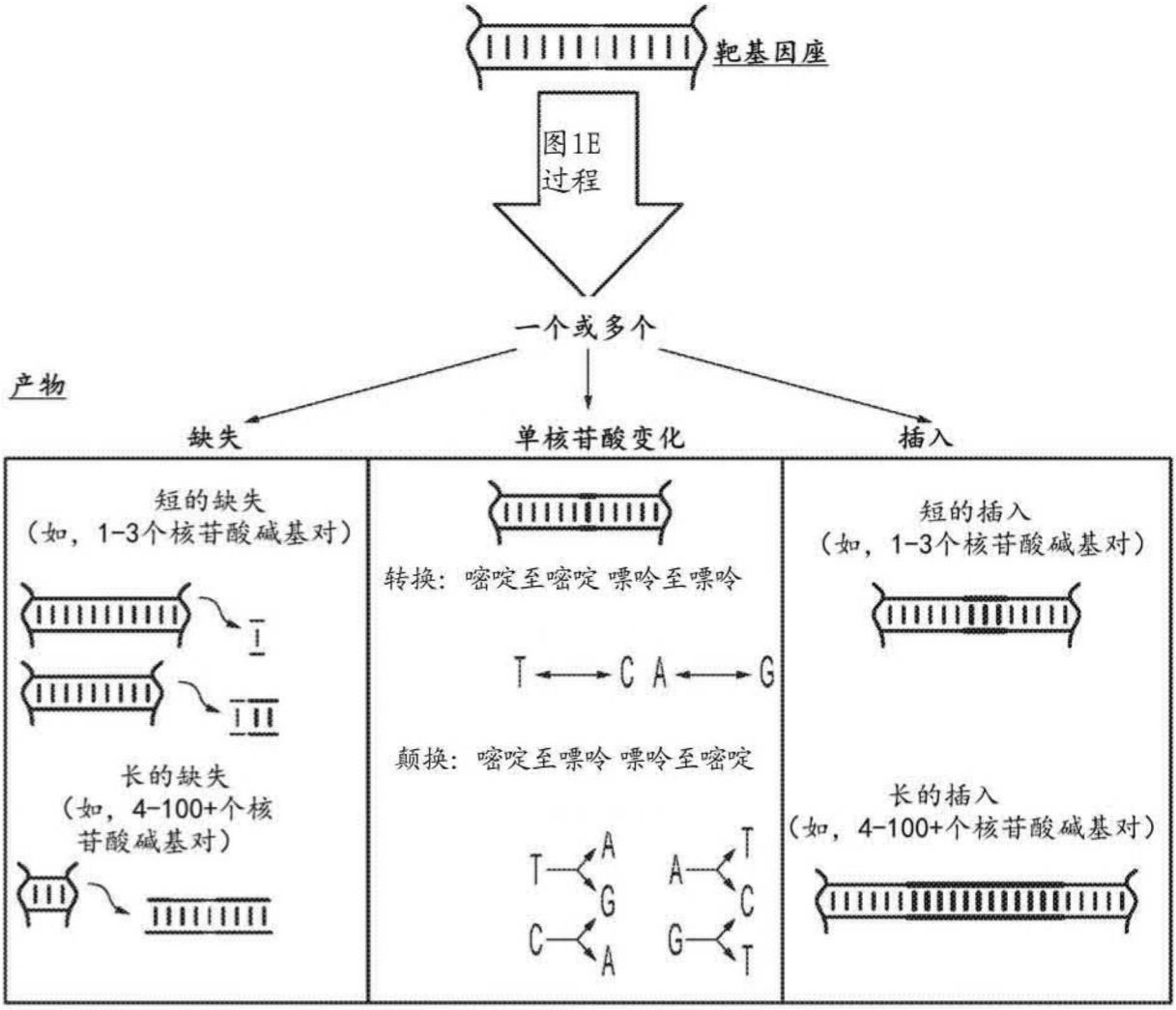

靶基因座
图1E
过程
一个或多个
产物
缺失
单核苷酸变化
插入
短的缺失
（如，1-3个核苷酸碱基对）
长的缺失
（如，4-100+个核
苷酸碱基对）
转换：嘧啶至嘧啶 嘌呤至嘌呤
T ⟷ C A ⟷ G
颠换：嘧啶至嘌呤 嘌呤至嘧啶
T A G C A
A T C G T
短的插入
（如，1-3个核苷酸碱基对）
长的插入
（如，4-100+个核苷酸碱基对）

图1H

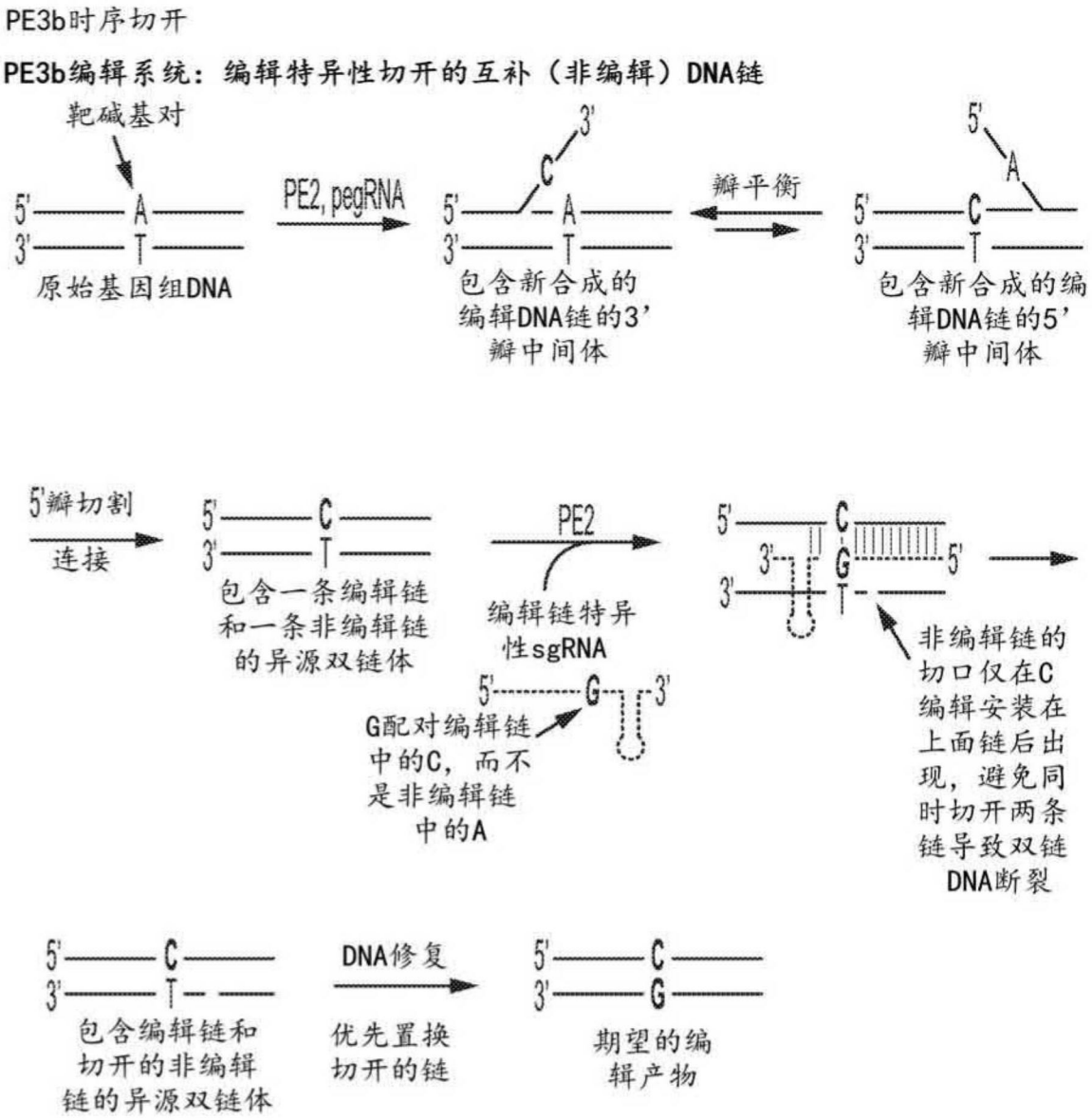

PE3b时序切开
PE3b编辑系统：编辑特异性切开的互补（非编辑）DNA链
靶碱基对
PE2, pegRNA
瓣平衡
原始基因组DNA
包含新合成的编辑DNA链的3’瓣中间体
包含新合成的编辑DNA链的5’瓣中间体
5’瓣切割
连接
包含一条编辑链和一条非编辑链的异源双链体
PE2
编辑链特异性sgRNA
G配对编辑链中的C，而不是非编辑链中的A
非编辑链的切口仅在C编辑安装在上面链后出现，避免同时切开两条链导致双链DNA断裂
DNA修复
包含编辑链和切开的非编辑链的异源双链体
优先置换切开的链
期望的编辑产物

图1I

利用任何可编程核酸酶的引导编辑

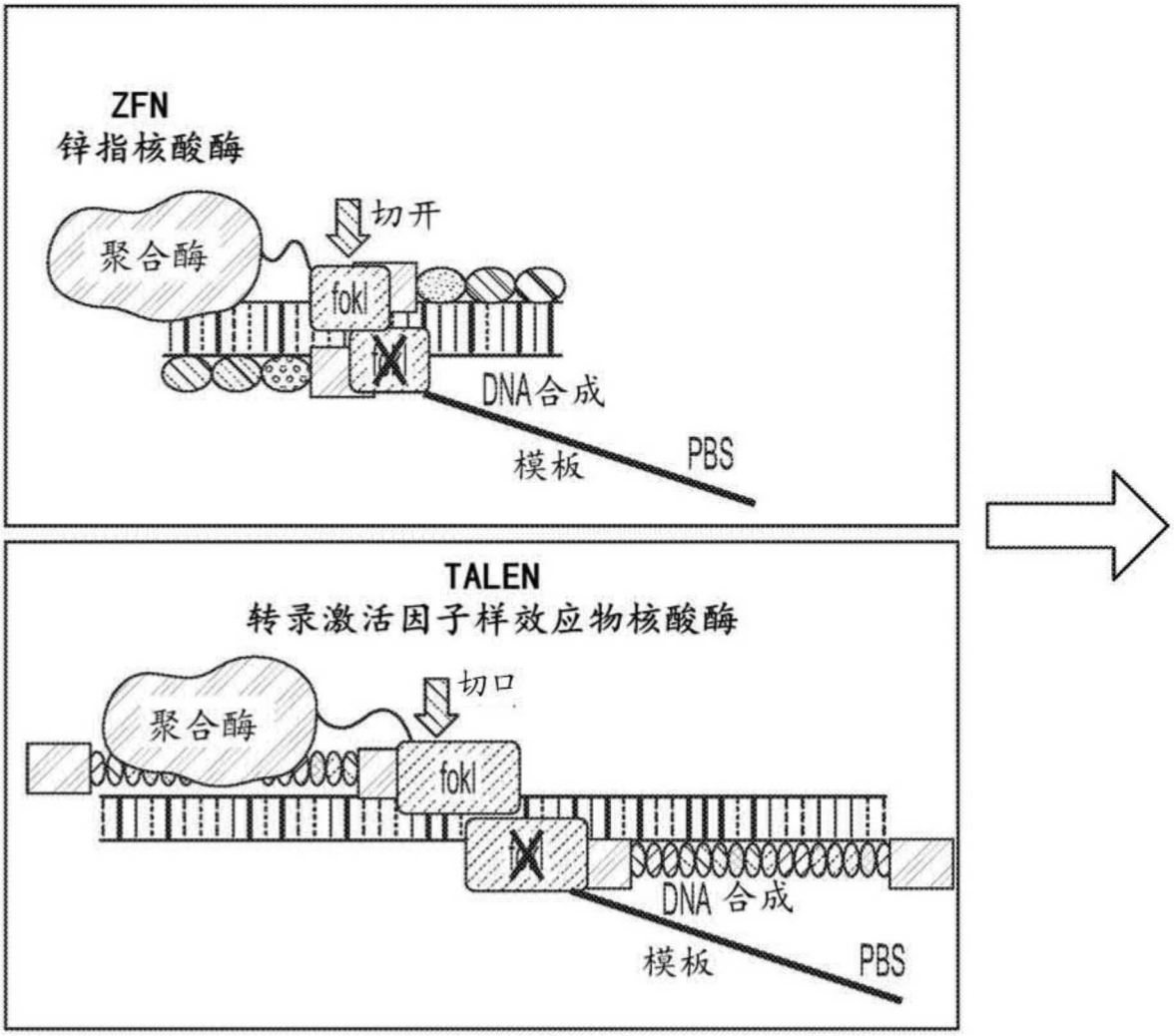


图1J

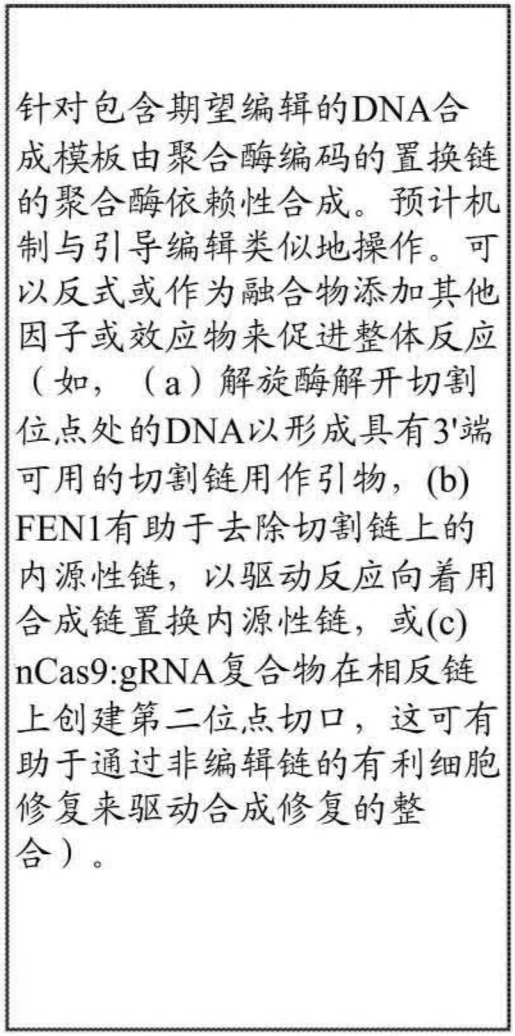

针对包含期望编辑的DNA合成模板由聚合酶编码的置换链的聚合酶依赖性合成。预计机制与引导编辑类似地操作。可以反式或作为融合物添加其他因子或效应物来促进整体反应（如，（a）解旋酶解开切割位点处的DNA以形成具有3'端可用的切割链用作引物，(b) FEN1有助于去除切割链上的内源性链，以驱动反应向着用合成链置换内源性链，或(c) nCas9:gRNA复合物在相反链上创建第二位点切口，这可有助于通过非编辑链的有利细胞修复来驱动合成修复的整合）。

图1K

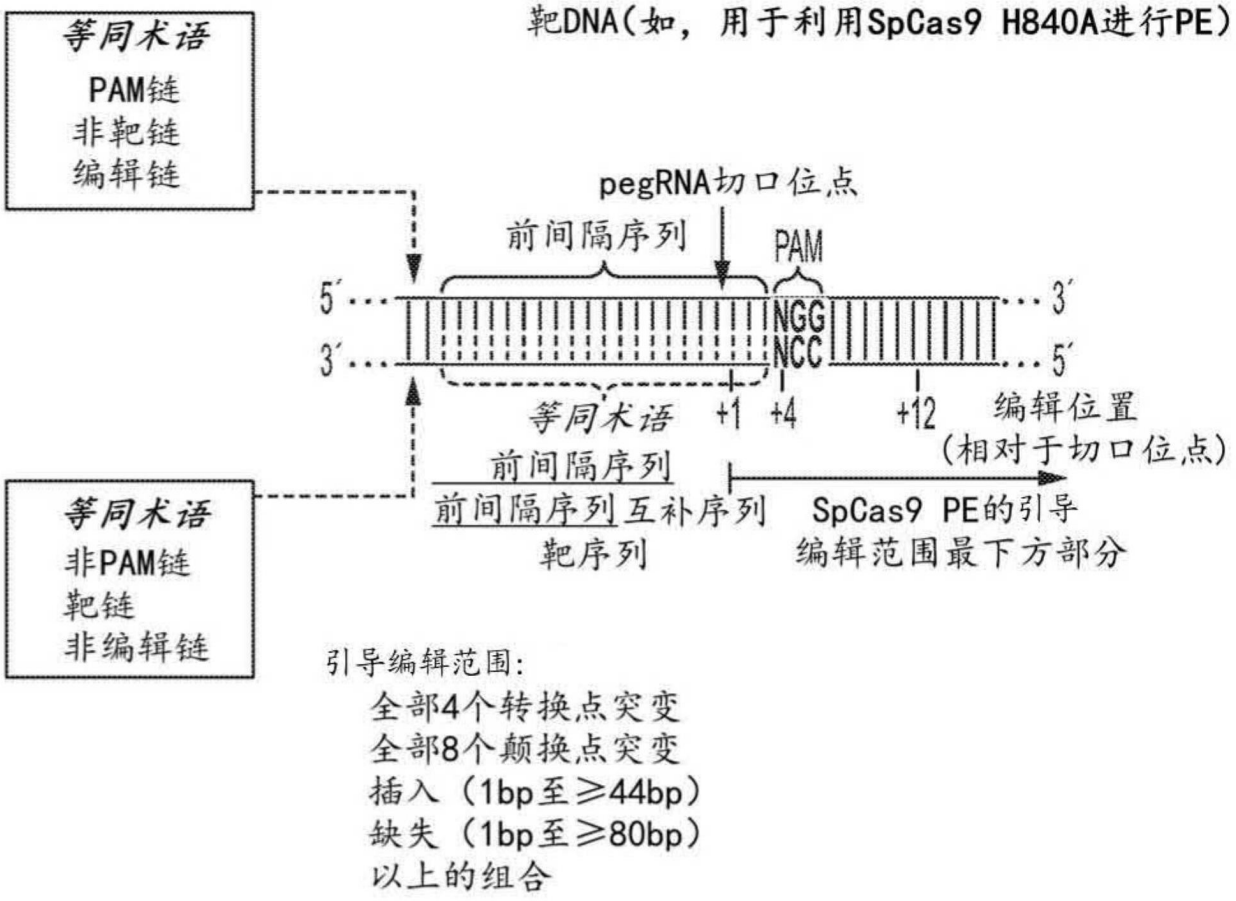

靶DNA(如，用于利用SpCas9 H840A进行PE)
等同术语
PAM链
非靶链
编辑链
pegRNA切口位点
前间隔序列
PAM
5′
3′
NGG
NCC
3′
5′
等同术语
+1
+4
+12
编辑位置
(相对于切口位点)
前间隔序列
前间隔序列互补序列
靶序列
SpCas9 PE的引导
编辑范围最下方部分
等同术语
非PAM链
靶链
非编辑链
引导编辑范围:
全部4个转换点突变
全部8个颠换点突变
插入（1bp至≥44bp）
缺失（1bp至≥80bp）
以上的组合

图1L

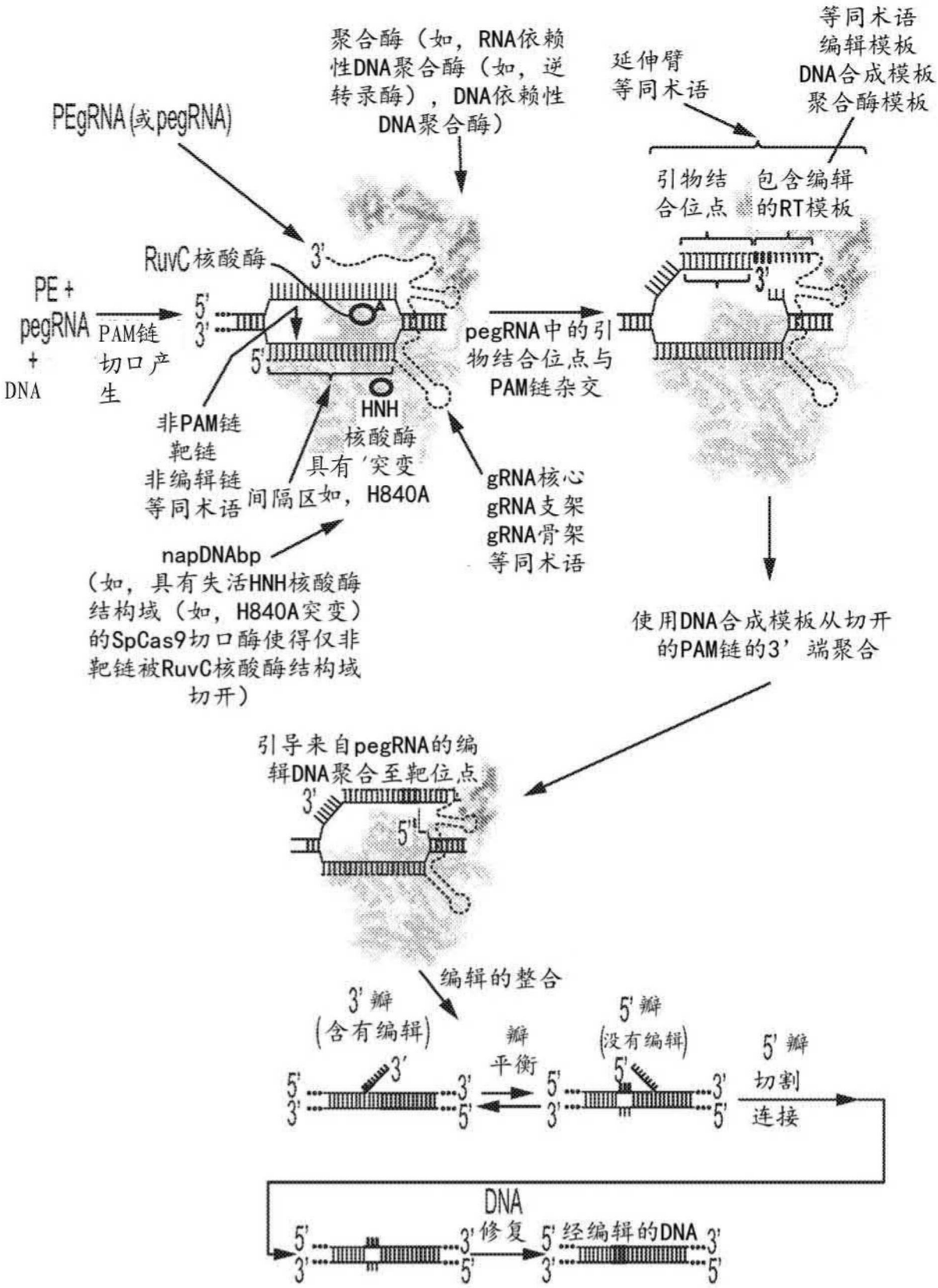

聚合酶（如，RNA依赖性DNA聚合酶（如，逆转录酶），DNA依赖性DNA聚合酶）
延伸臂
等同术语
等同术语
编辑模板
DNA合成模板
聚合酶模板
PEgRNA(或pegRNA)
引物结合位点
包含编辑的RT模板
RuvC核酸酶
PE + pegRNA + DNA
PAM链切口产生
pegRNA中的引物结合位点与PAM链杂交
非PAM链
靶链
非编辑链
等同术语
HNH核酸酶
具有'突变
间隔区如，H840A
gRNA核心
gRNA支架
gRNA骨架
等同术语
napDNAbp
（如，具有失活HNH核酸酶结构域（如，H840A突变）的SpCas9切口酶使得仅非靶链被RuvC核酸酶结构域切开）
使用DNA合成模板从切开的PAM链的3’端聚合
引导来自pegRNA的编辑DNA聚合至靶位点
编辑的整合
3’瓣（含有编辑）
瓣平衡
5’瓣（没有编辑）
5’瓣切割连接
DNA修复
经编辑的DNA

图1M

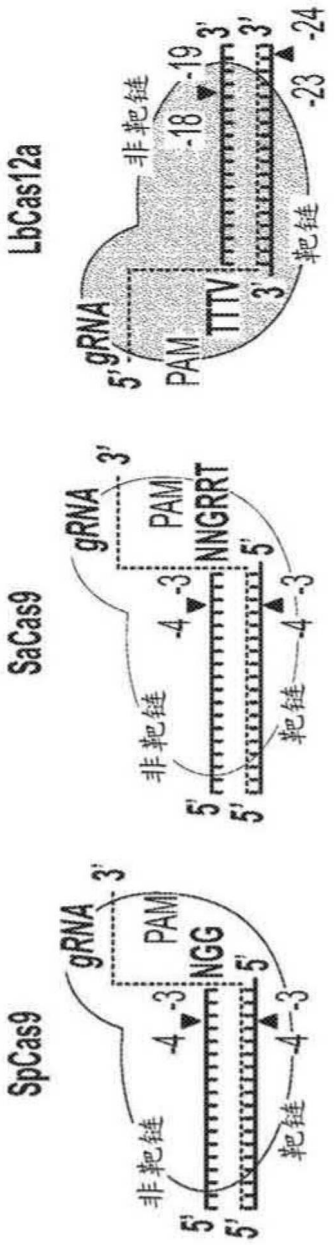

LbCas12a
gRNA
PAM
TTTV
非靶链
靶链
-18
-19
-23
-24
SaCas9
gRNA
PAM
NNGRRT
非靶链
靶链
-4
-3
SpCas9
gRNA
PAM
NGG
非靶链
靶链
-4
-3

图2

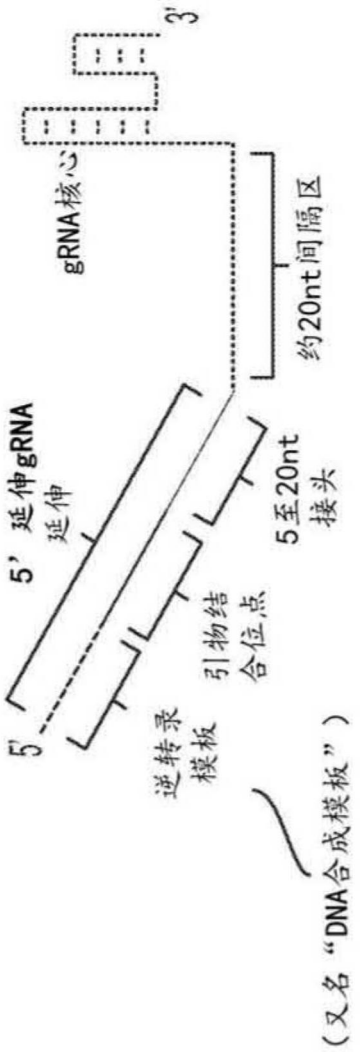

3'
gRNA核心
约20nt间隔区
5'延伸gRNA
延伸
5至20nt
接头
引物结
合位点
5'
逆转录
模板
(又名"DNA合成模板")

图3A

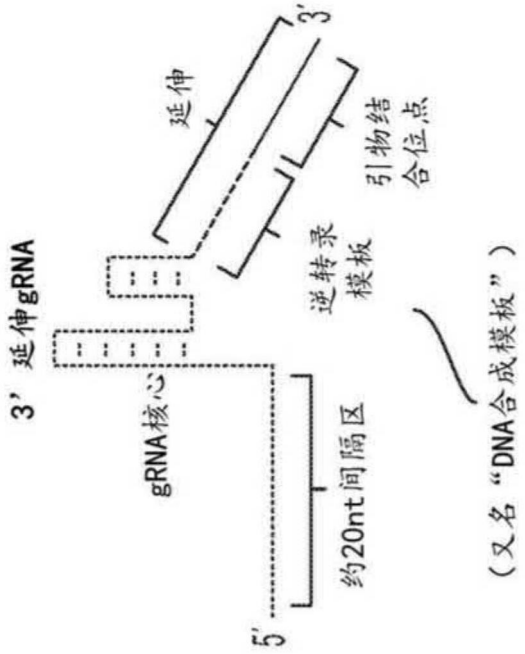

3’ 延伸gRNA
gRNA核心
延伸
引物结合位点
逆转录模板
约20nt间隔区
（又名“DNA合成模板”）
5’
3’

图3B

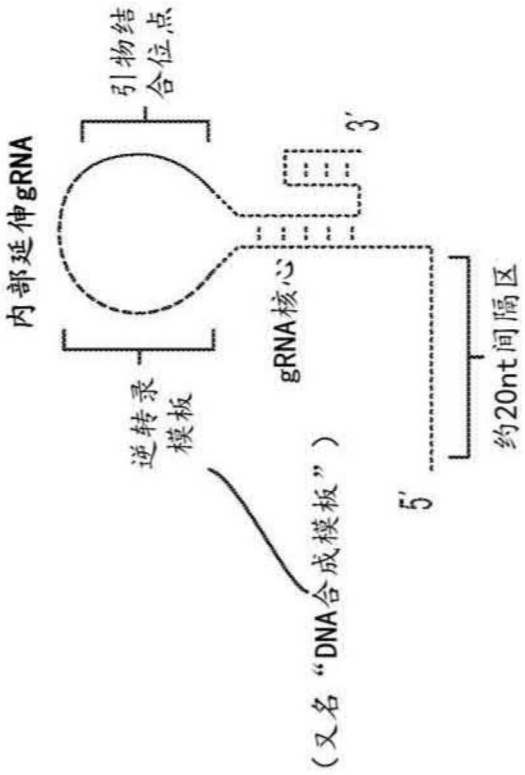

内部延伸gRNA
引物结合位点
逆转录模板
gRNA核心
3’
（又名“DNA合成模板”）
5’
约20nt间隔区

图3C

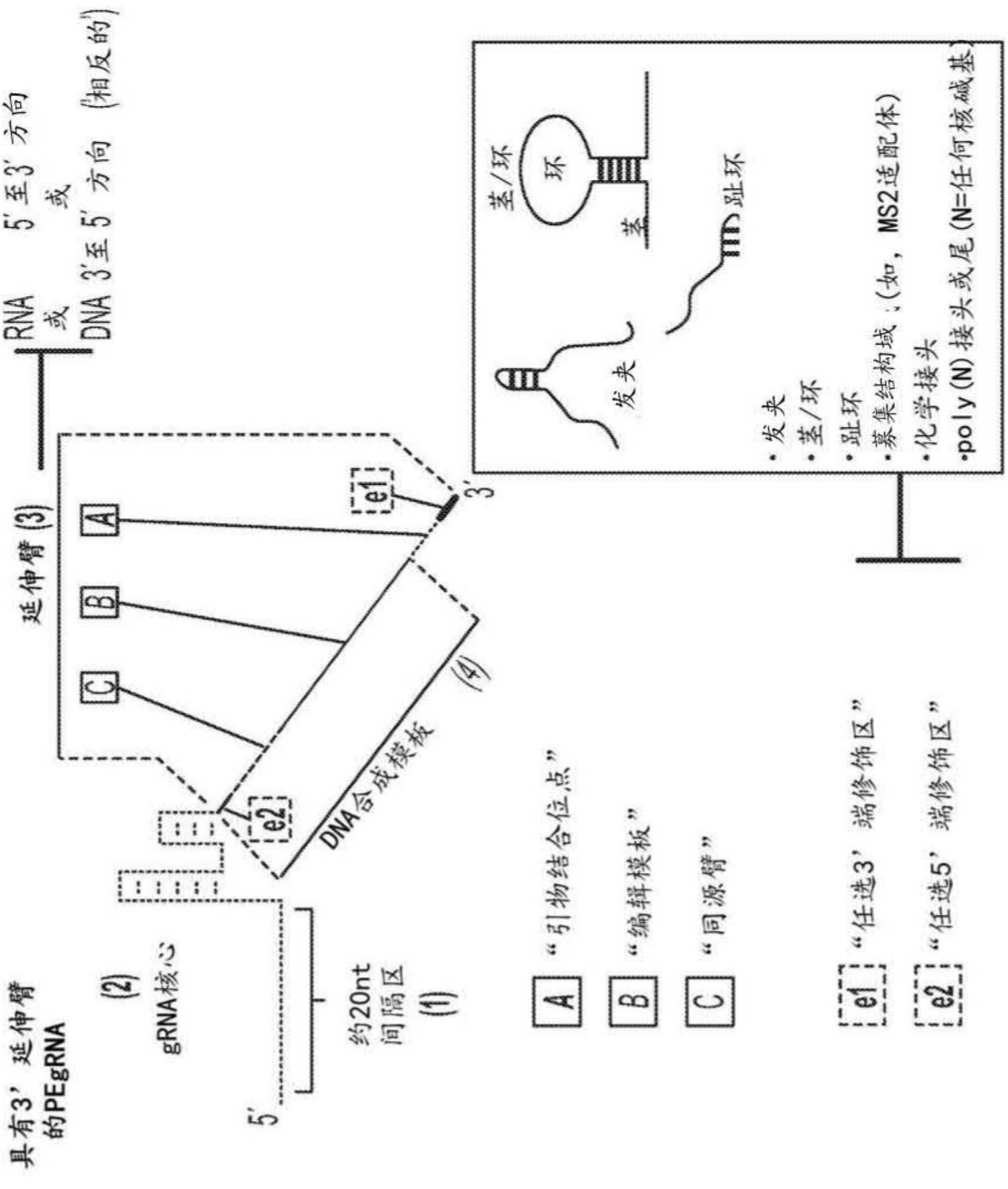

具有3’延伸臂的PEgRNA
(2)
gRNA核心
5’
约20nt间隔区
(1)
延伸臂 (3)
RNA 5’至3’方向
或
DNA 3’至5’方向 (相反的)
A
B
C
e1
e2
3’
DNA合成模板
(4)
茎/环
环
茎
发夹
ITD趾环
• 发夹
• 茎/环
• 趾环
• 募集结构域（如，MS2适配体）
• 化学接头
• poly(N)接头或尾 (N=任何核碱基)
A “引物结合位点”
B “编辑模板”
C “同源臂”
e1 “任选3’端修饰区”
e2 “任选5’端修饰区”

图3D

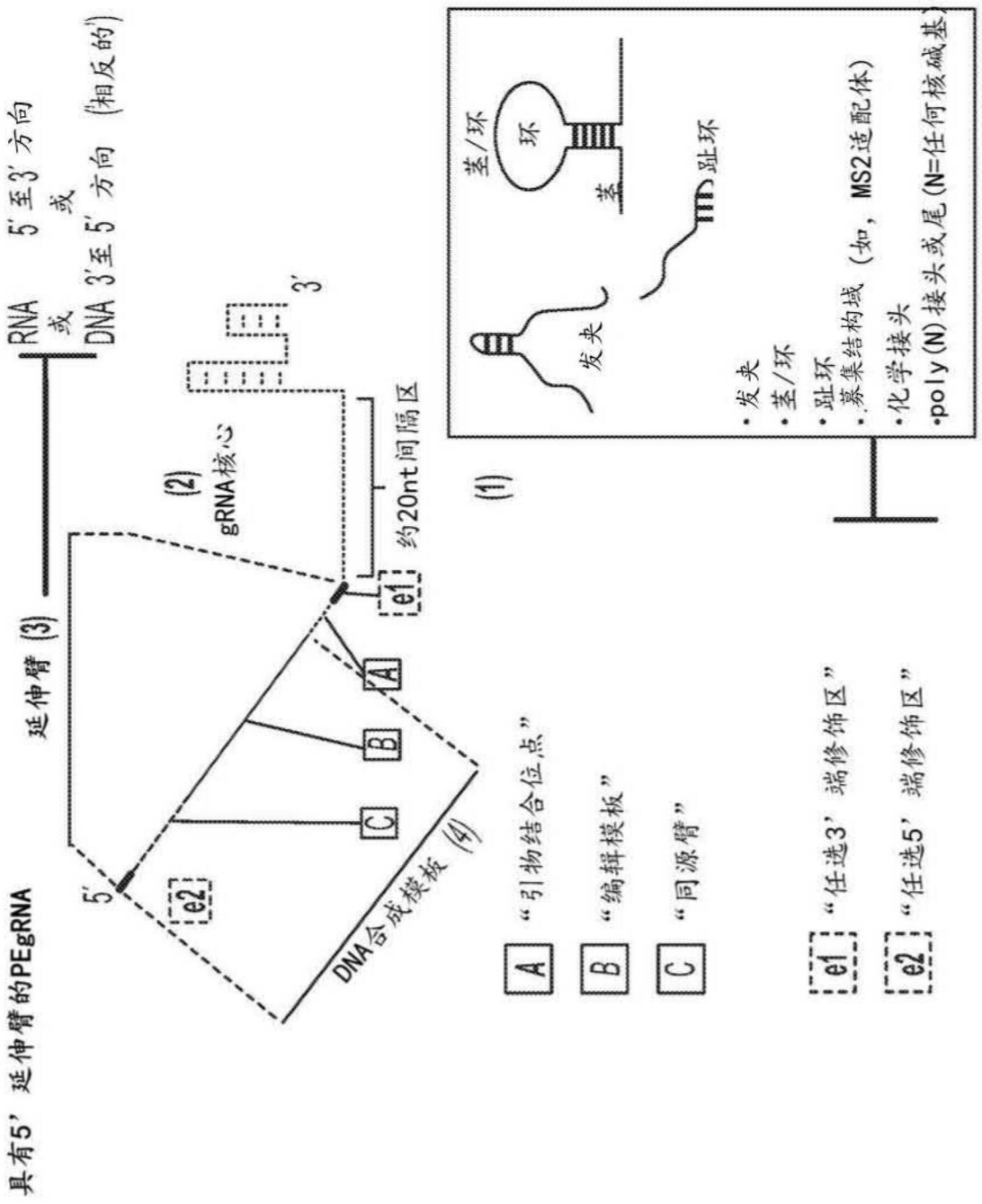

具有5’延伸臂的PEgRNA
延伸臂 (3)
RNA 5’至3’方向
或
DNA 3’至5’方向 (相反的)
5’
3’
(2)
gRNA核心
e1 约20nt间隔区
e2
DNA合成模板 (4)
A
B
C
A “引物结合位点”
B “编辑模板”
C “同源臂”
e1 “任选3’端修饰区”
e2 “任选5’端修饰区”
(1)
茎/环
环
发夹
茎
ITD 趾环
• 发夹
• 茎/环
• 趾环
• 募集结构域 (如，MS2适配体)
• 化学接头
• poly (N) 接头或尾 (N=任何核碱基)

图3E

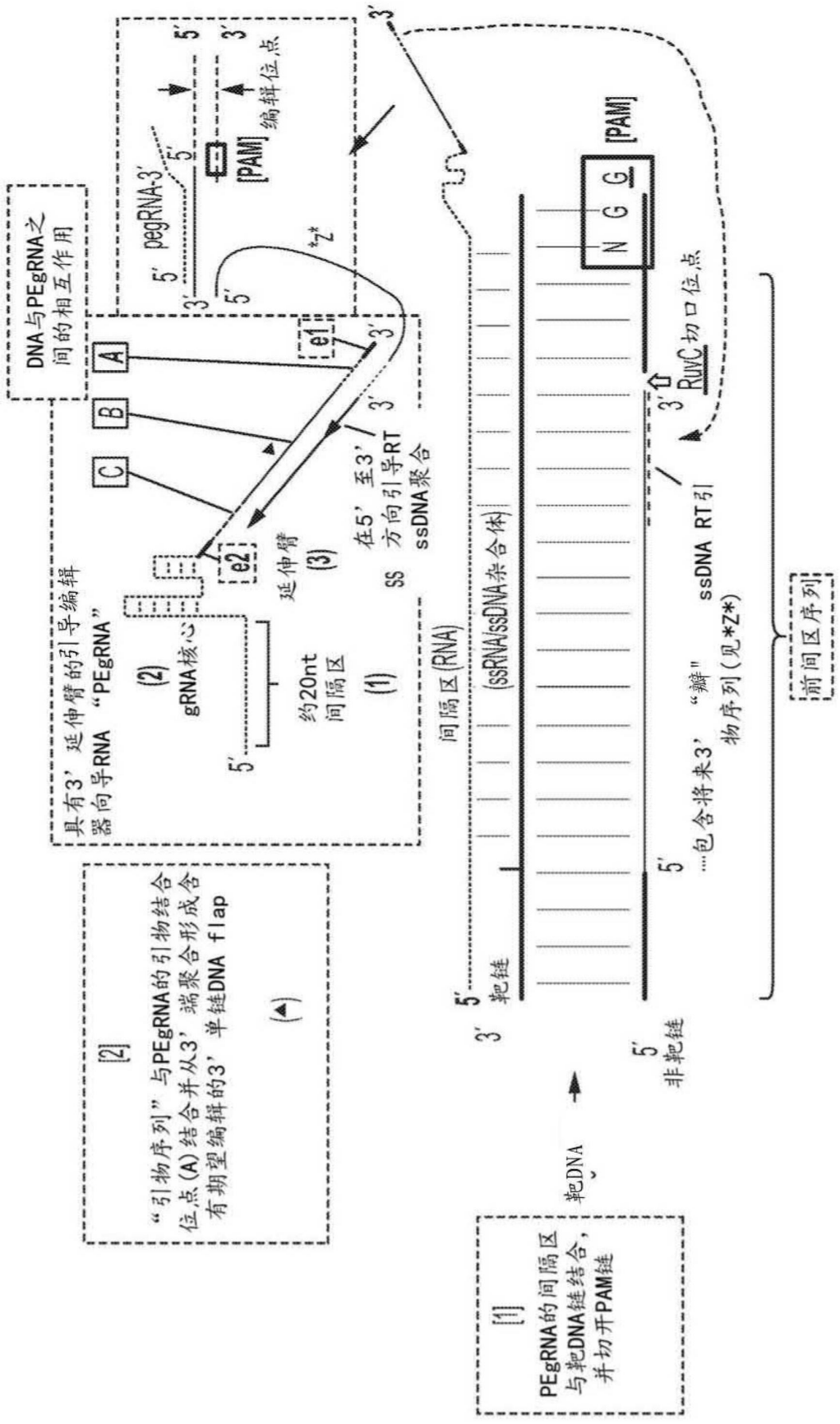

DNA与PEgRNA之
间的相互作用
具有3’延伸臂的引导编辑
器向导RNA “PEgRNA”
(2)
gRNA核心
约20nt
间隔区
(1)
延伸臂
(3)
在5’至3’
方向引导RT
ssDNA聚合
SS
间隔区(RNA)
(ssRNA/ssDNA杂合体)
pegRNA-3’
[PAM]
编辑位点
N G G
[PAM]
RuvC切口位点
ssDNA RT引
物序列(见*Z*)
“瓣”
....包含将来3’
前间区序列
[2]
“引物序列”与PEgRNA的引物结合
位点(A)结合并从3’端聚合形成含
有期望编辑的3’单链DNA flap
(A)
[1]
PEgRNA的间隔区
与靶DNA链结合，
并切开PAM链
靶DNA
靶链
非靶链

图3F

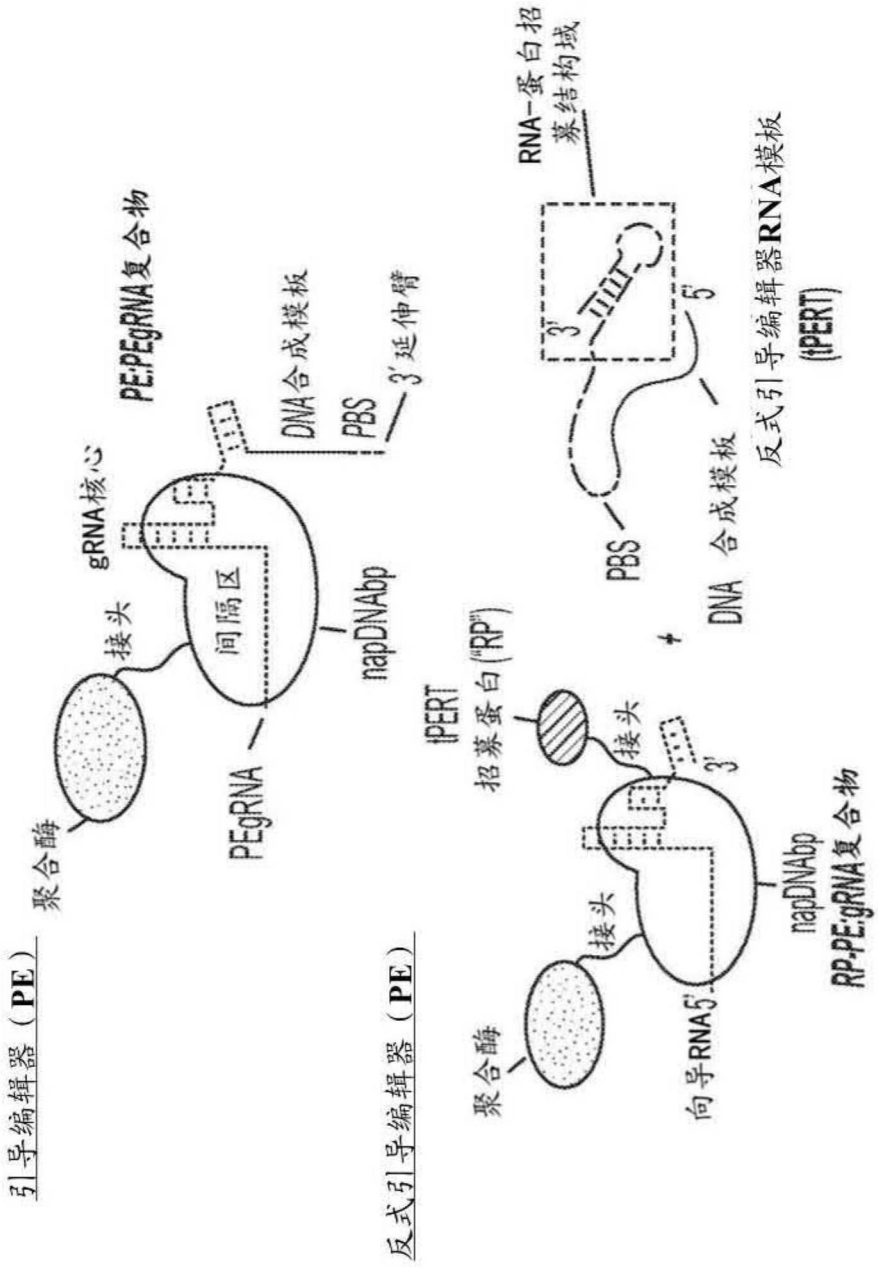

引导编辑器（PE）
聚合酶
接头
gRNA核心
间隔区
PE:PEgRNA复合物
DNA合成模板
PBS
3'延伸臂
PEgRNA
napDNAbp
反式引导编辑器（PE）
tPERT
招募蛋白("RP")
接头
聚合酶
接头
napDNAbp
向导RNA5'
RP-PE:gRNA复合物
RNA-蛋白招募结构域
3'
5'
PBS
DNA 合成模板
反式引导编辑器RNA模板
(tPERT)

图3G

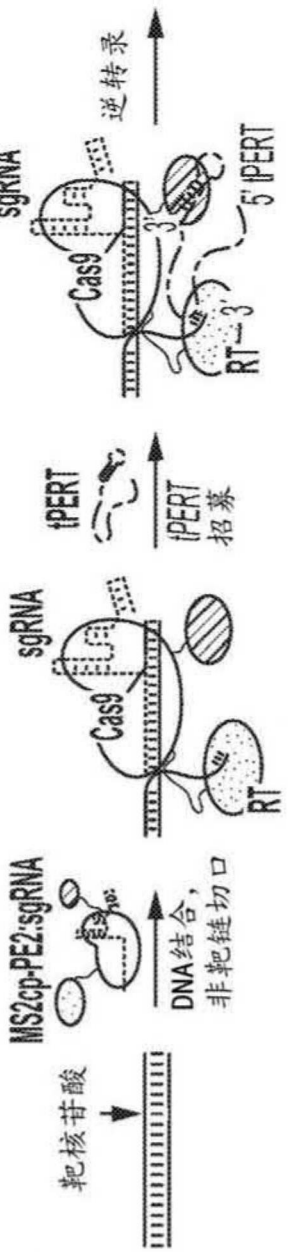

靶核苷酸
MS2cp-PE2:sgRNA
DNA结合，
非靶链切口
sgRNA
Cas9
RT
tPERT
tPERT
招募
sgRNA
Cas9
RT
3'
5' tPERT
逆转录

图3H

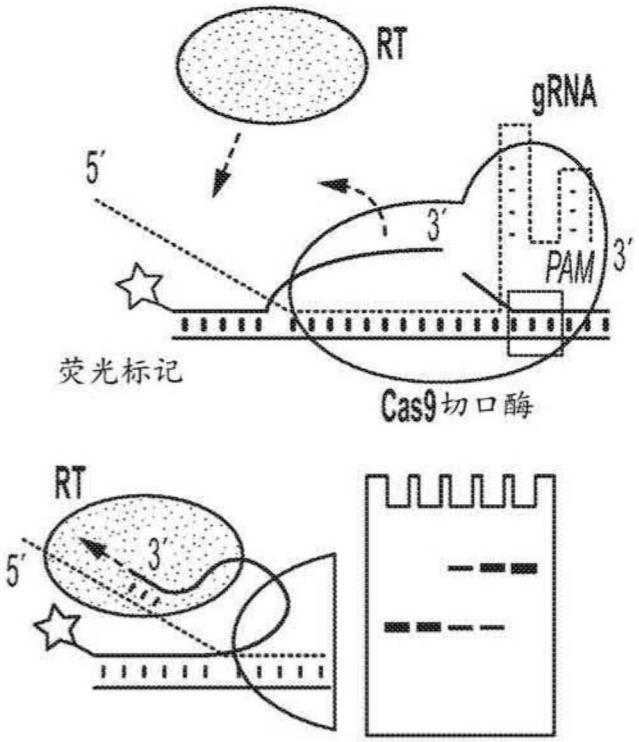

RT
gRNA
5′
3′
PAM
3′
荧光标记
Cas9切口酶
RT
5′
3′

图4A

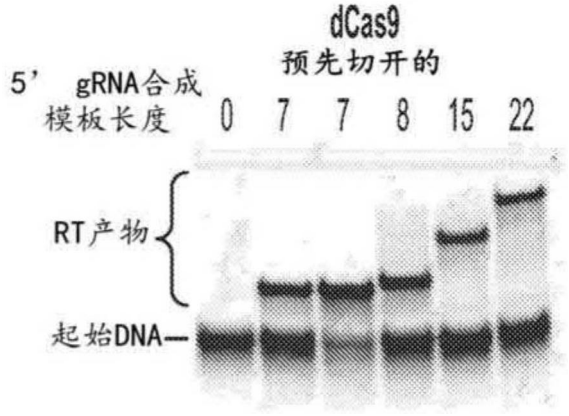

dCas9
预先切开的
5’ gRNA合成
模板长度 0 7 7 8 15 22
RT产物
起始DNA

图4B

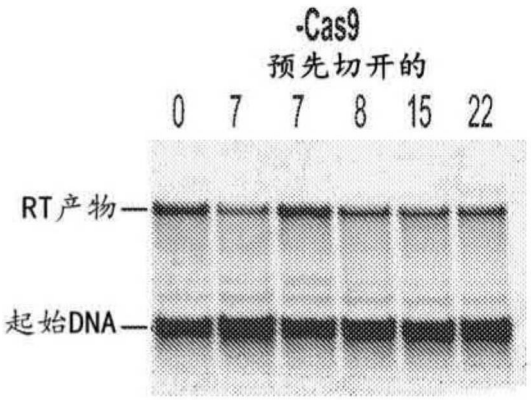

-Cas9
预先切开的
0 7 7 8 15 22
RT产物
起始DNA

图4C

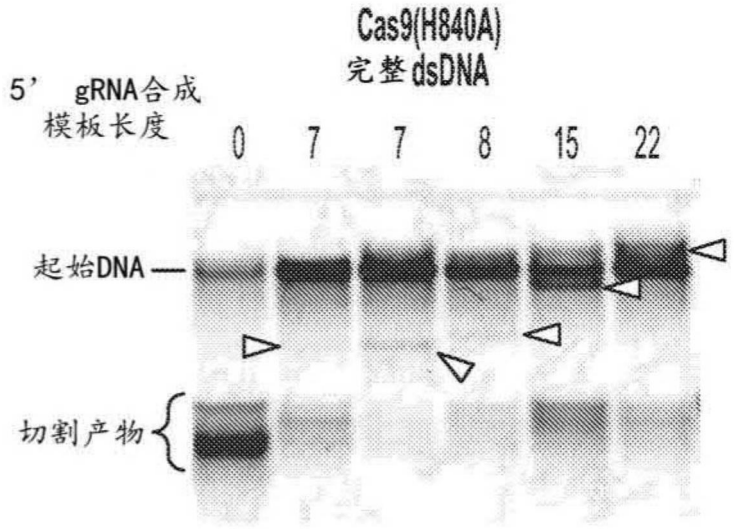

Cas9(H840A)
完整dsDNA
5’ gRNA合成
模板长度
0 7 7 8 15 22
起始DNA
切割产物

图4D

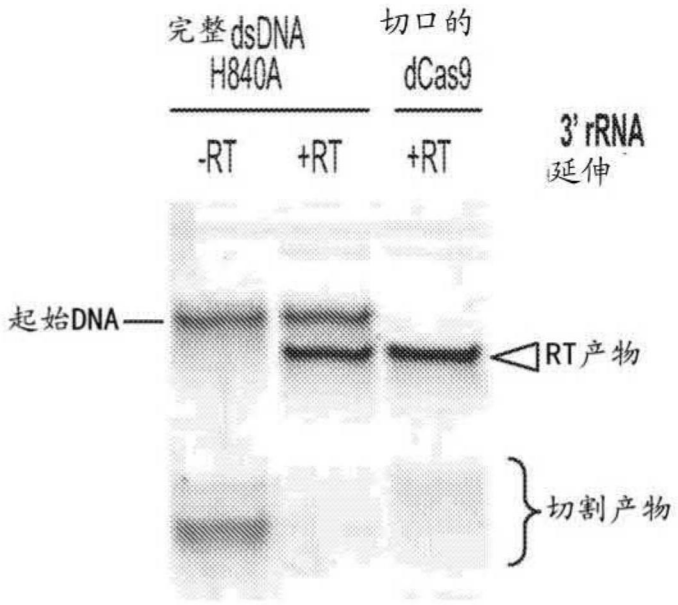


图4E

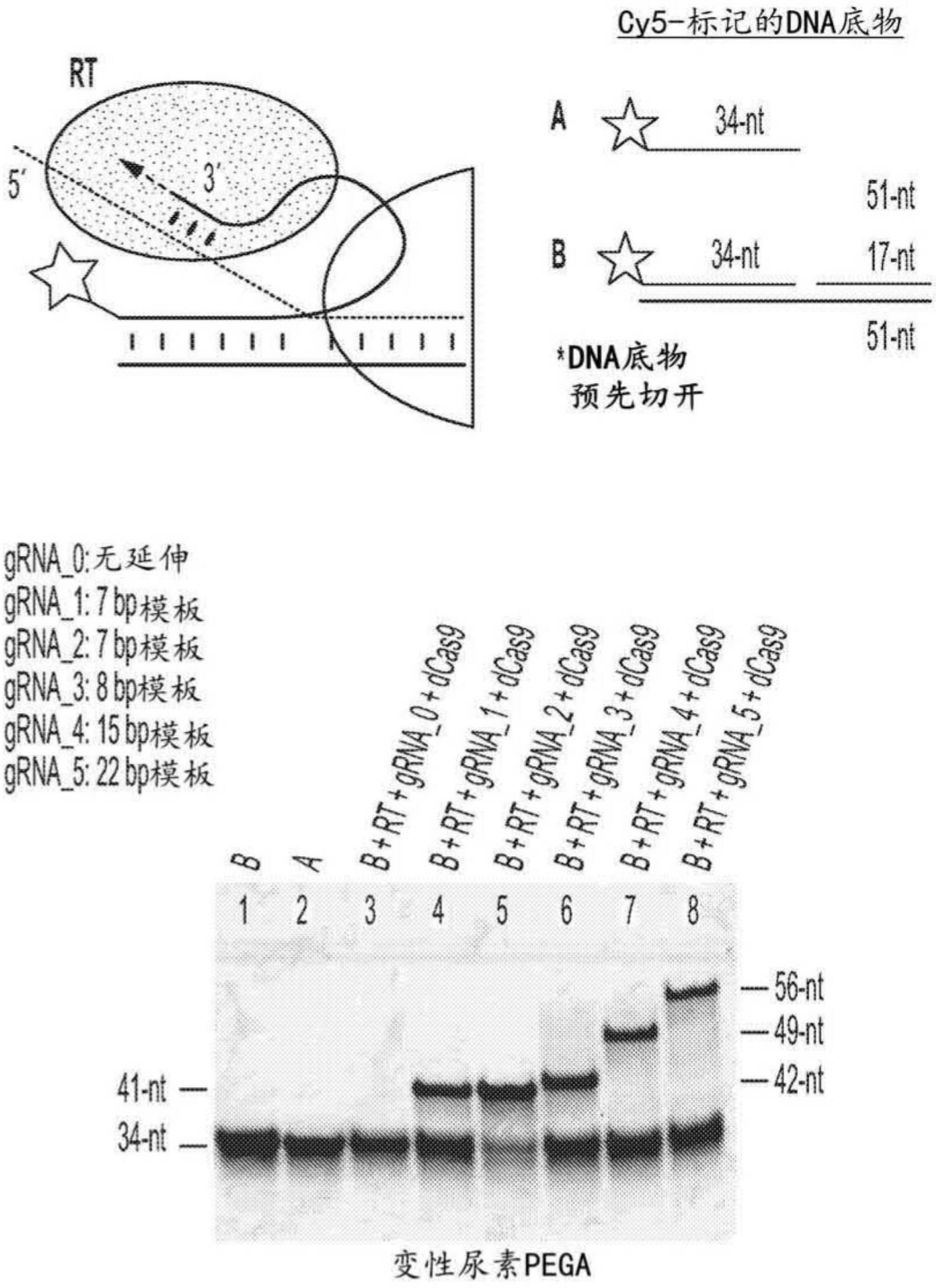

Cy5-标记的DNA底物
RT
5′
3′
A
34-nt
51-nt
B
34-nt
17-nt
51-nt
*DNA底物
预先切开
gRNA_0:无延伸
gRNA_1:7 bp模板
gRNA_2:7 bp模板
gRNA_3:8 bp模板
gRNA_4:15 bp模板
gRNA_5:22 bp模板
B
A
B+RT+gRNA_0+dCas9
B+RT+gRNA_1+dCas9
B+RT+gRNA_2+dCas9
B+RT+gRNA_3+dCas9
B+RT+gRNA_4+dCas9
B+RT+gRNA_5+dCas9
1 2 3 4 5 6 7 8
56-nt
49-nt
42-nt
41-nt
34-nt
变性尿素PEGA

图5

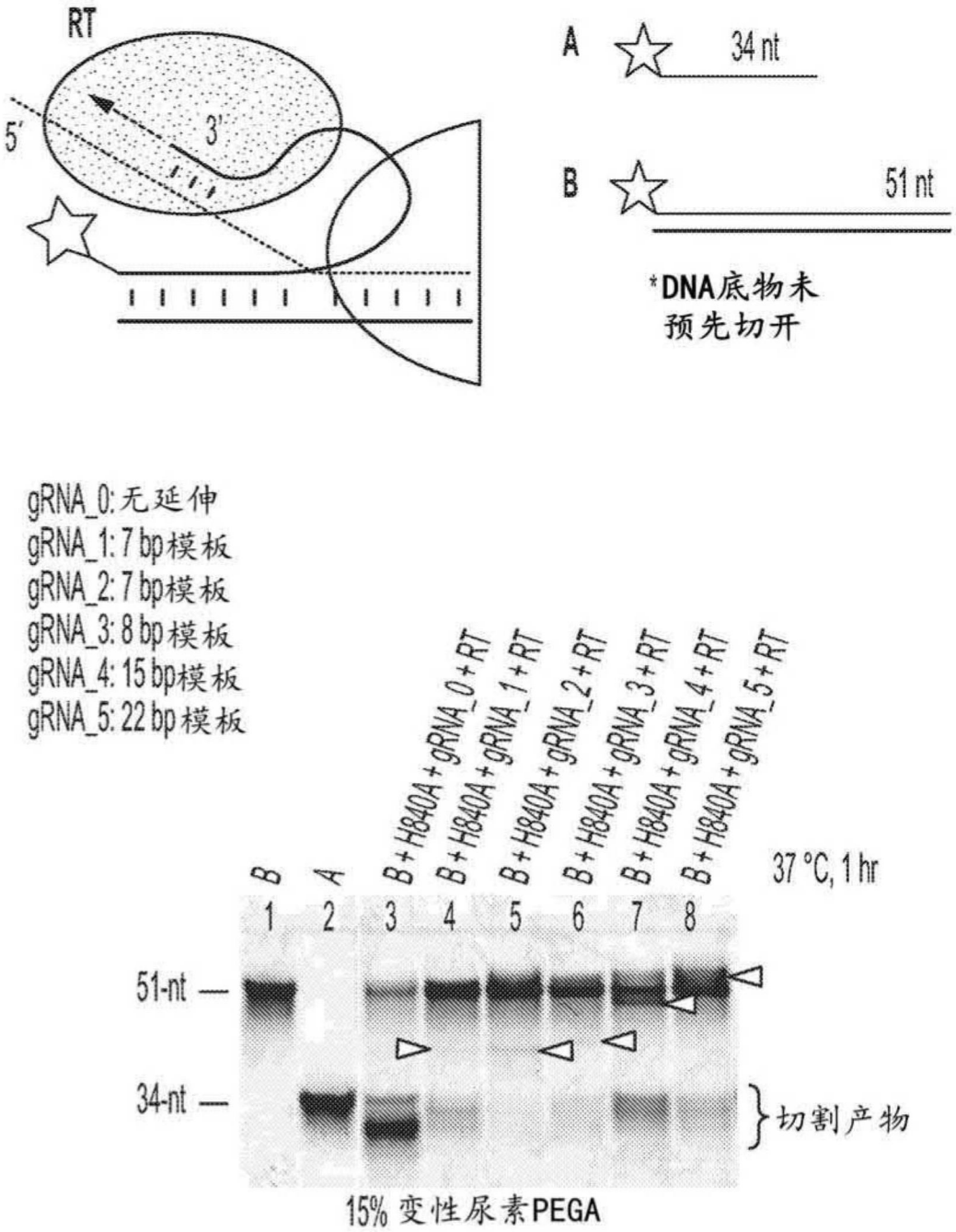

RT
5′
3′
A
34 nt
B
51 nt
*DNA底物未
预先切开
gRNA_0:无延伸
gRNA_1:7 bp模板
gRNA_2:7 bp模板
gRNA_3:8 bp模板
gRNA_4:15 bp模板
gRNA_5:22 bp模板
B
A
B+H840A+gRNA_0+RT
B+H840A+gRNA_1+RT
B+H840A+gRNA_2+RT
B+H840A+gRNA_3+RT
B+H840A+gRNA_4+RT
B+H840A+gRNA_5+RT
37 °C, 1 hr
1 2 3 4 5 6 7 8
51-nt
34-nt
切割产物
15% 变性尿素PEGA

图6

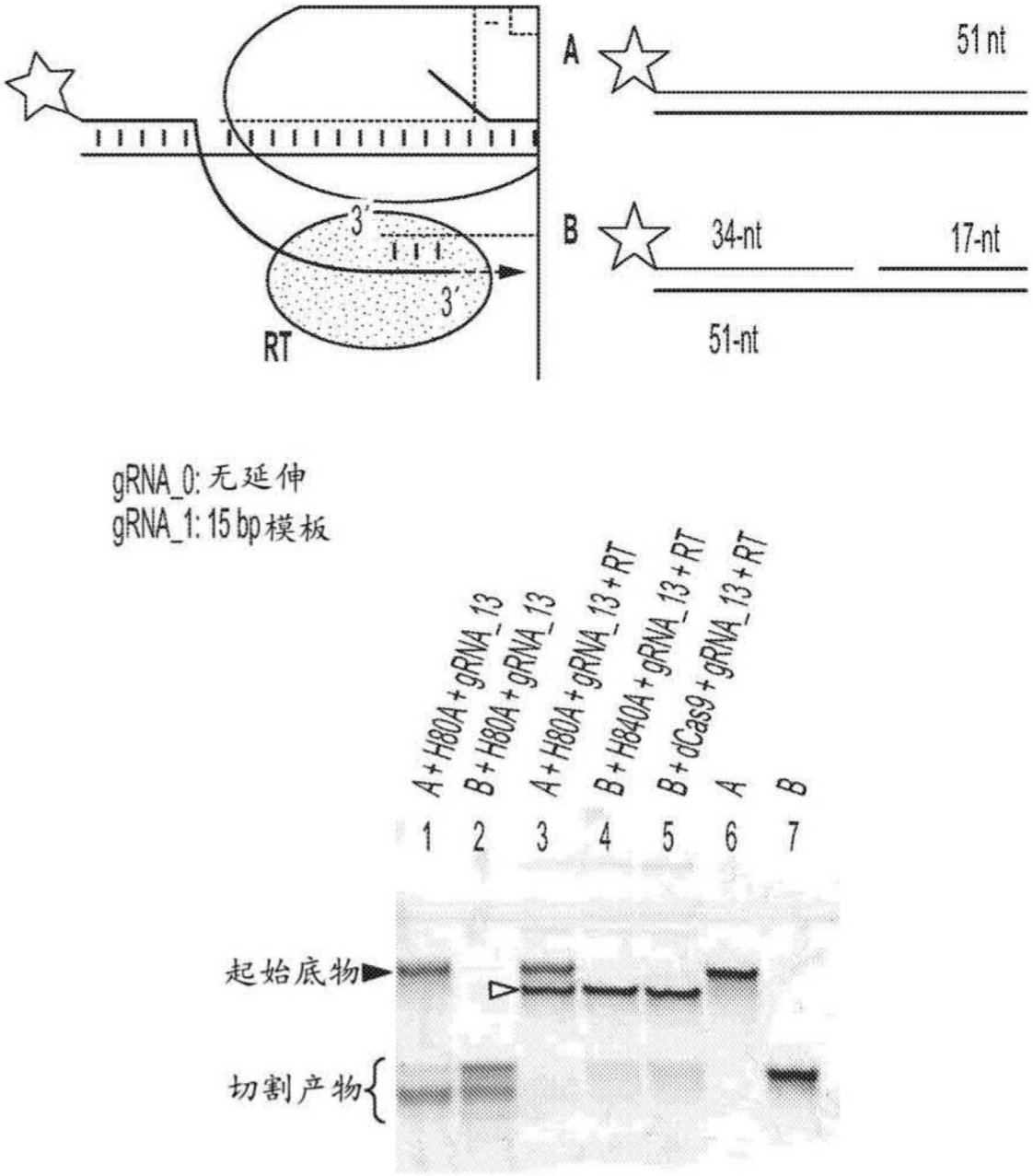
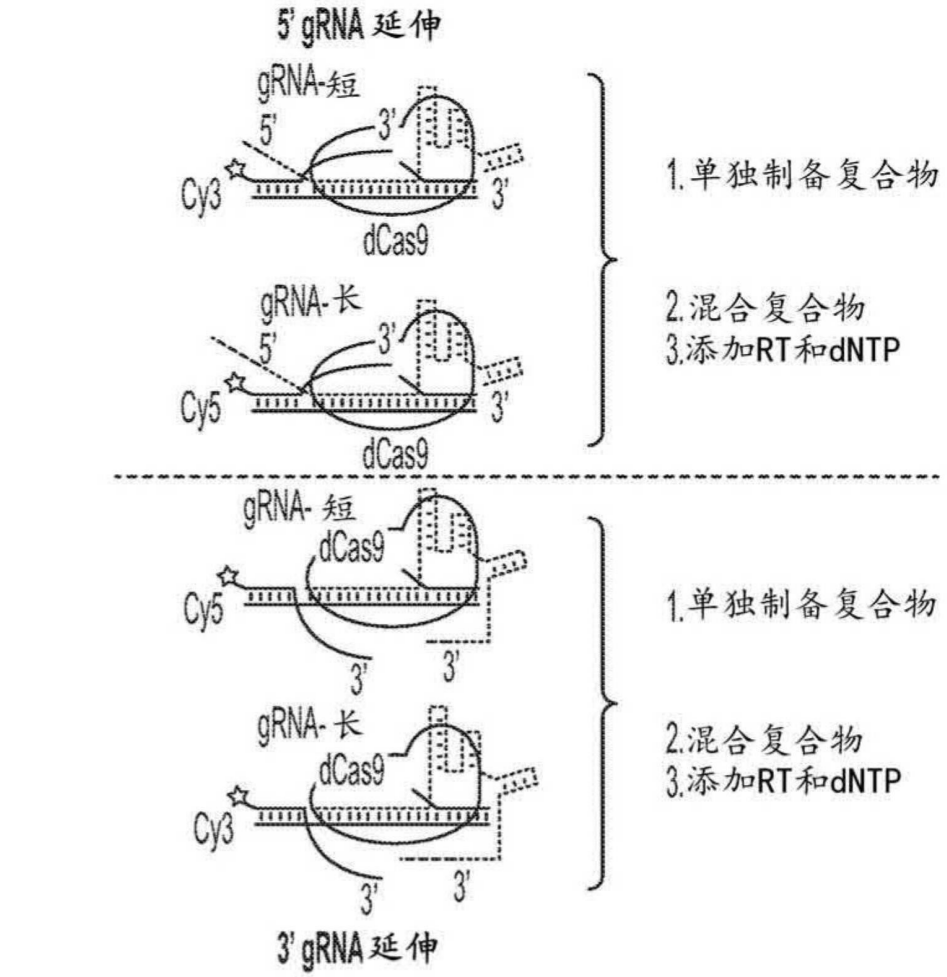

A
51 nt
B
34-nt
17-nt
51-nt
3'
3'
RT
gRNA_0: 无延伸
gRNA_1: 15 bp模板
A+H80A+gRNA_13
B+H80A+gRNA_13
A+H80A+gRNA_13+RT
B+H840A+gRNA_13+RT
B+dCas9+gRNA_13+RT
A
B
1 2 3 4 5 6 7
起始底物
切割产物

图7


5' gRNA 延伸
gRNA-短
5'
3'
Cy3
3'
dCas9
1.单独制备复合物
gRNA-长
5'
3'
Cy5
3'
dCas9
2.混合复合物
3.添加RT和dNTP
gRNA- 短
dCas9
Cy5
3'
3'
1.单独制备复合物
gRNA- 长
dCas9
Cy3
3'
3'
2.混合复合物
3.添加RT和dNTP
3' gRNA 延伸

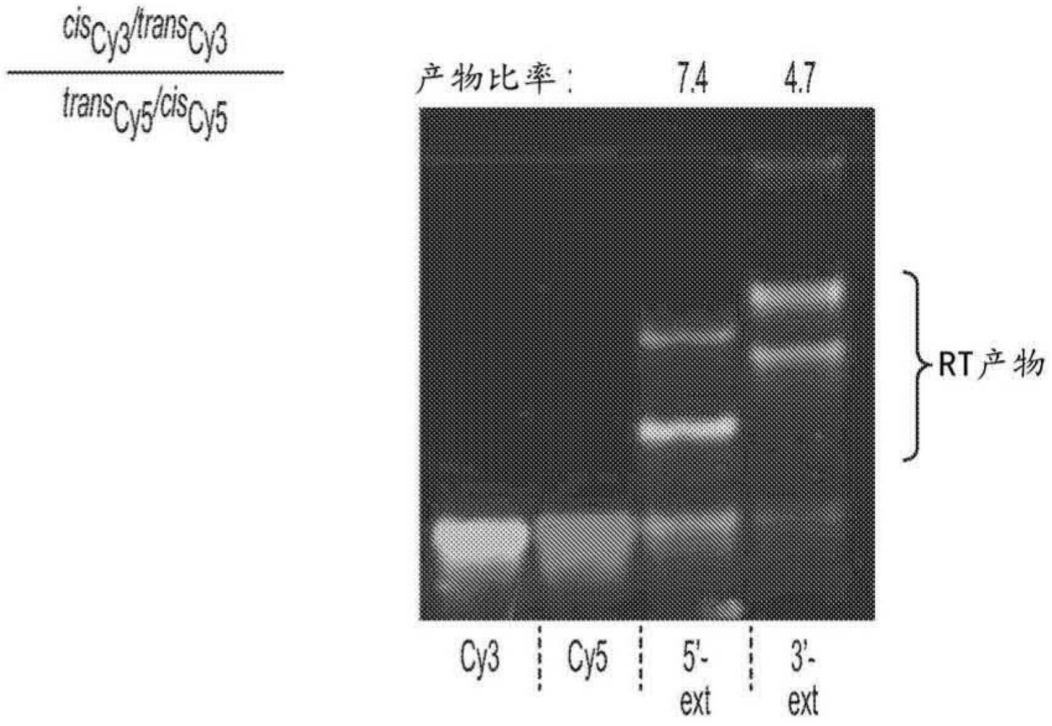

cis Cy3/trans Cy3
trans Cy5/cis Cy5
产物比率：
7.4
4.7
RT产物
Cy3
Cy5
5'-ext
3'-ext

图8

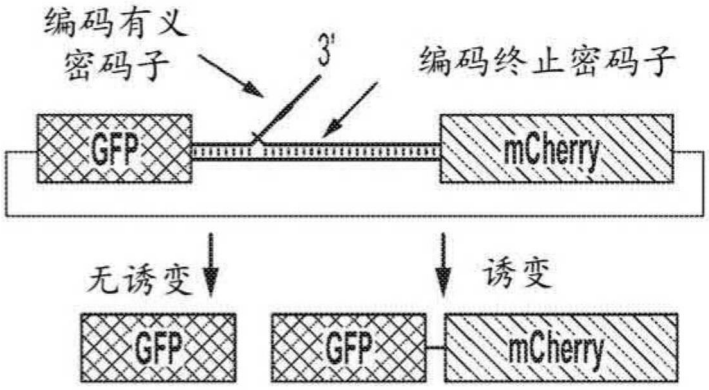

编码有义
密码子
3'
编码终止密码子
GFP
mCherry
无诱变
诱变
GFP
GFP
mCherry

图9A

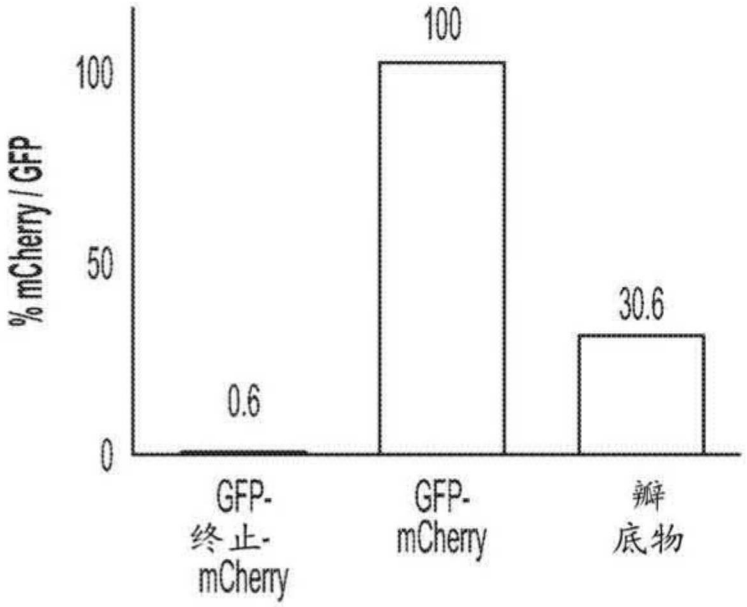

% mCherry / GFP
100
50
0
0.6
100
30.6
GFP-
终止-
mCherry
GFP-
mCherry
瓣辞
底物

图9B

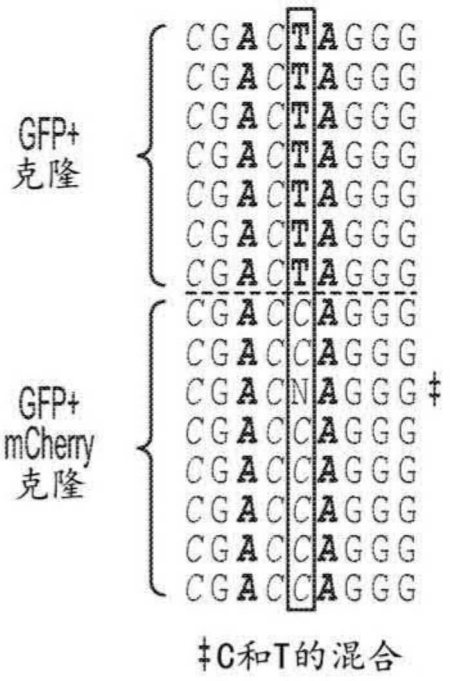

GFP+
克隆
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
GFP+
mCherry
克隆
CGACCAGGG
CGACCAGGG
CGACNAGGG‡
CGACCAGGG
CGACCAGGG
CGACCAGGG
CGACCAGGG
CGACCAGGG
‡C和T的混合

图9C

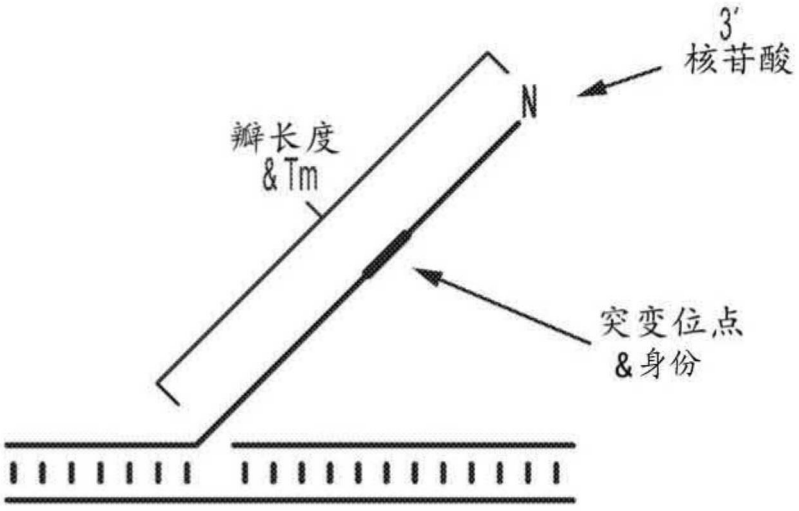

3′
核苷酸
N
瓣长度
&Tm
突变位点
&身份

图9D

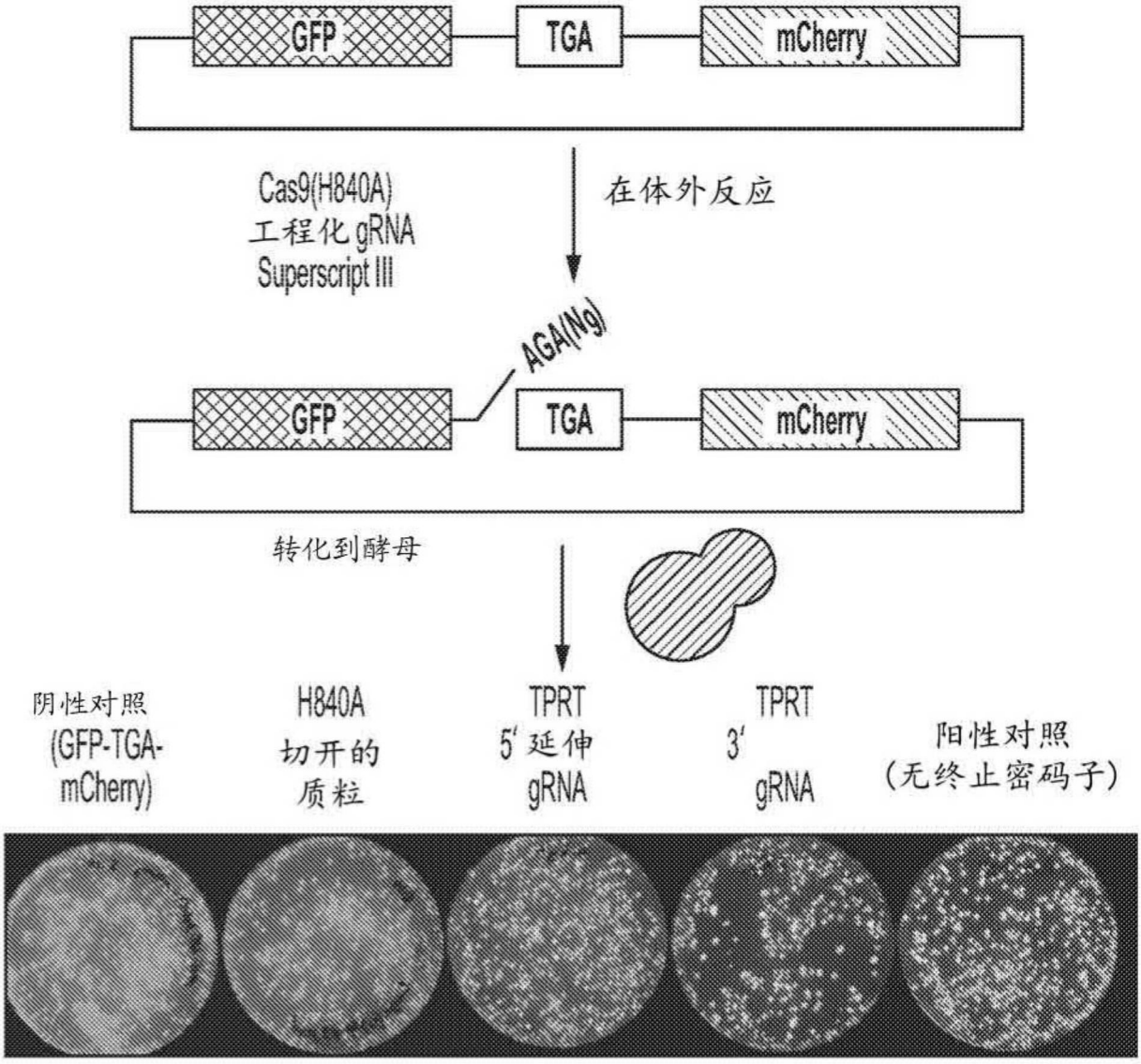

GFP
TGA
mCherry
Cas9(H840A)
工程化 gRNA
Superscript III
在体外反应
AGA(Ng)
GFP
TGA
mCherry
转化到酵母
阴性对照
(GFP-TGA-
mCherry)
H840A
切开的
质粒
TPRT
5' 延伸
gRNA
TPRT
3'
gRNA
阳性对照
(无终止密码子)

图10

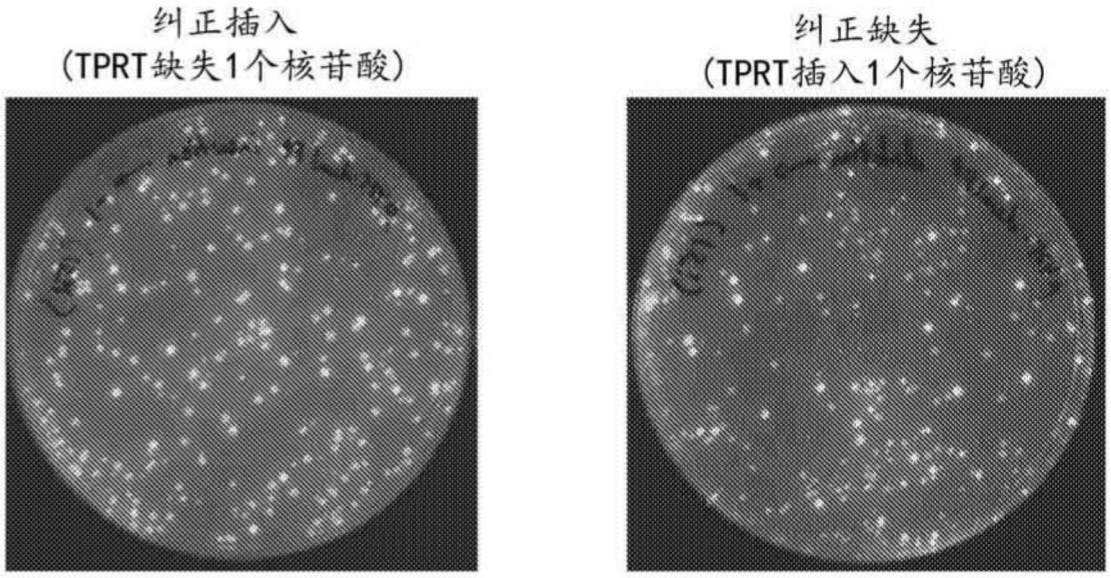


图11

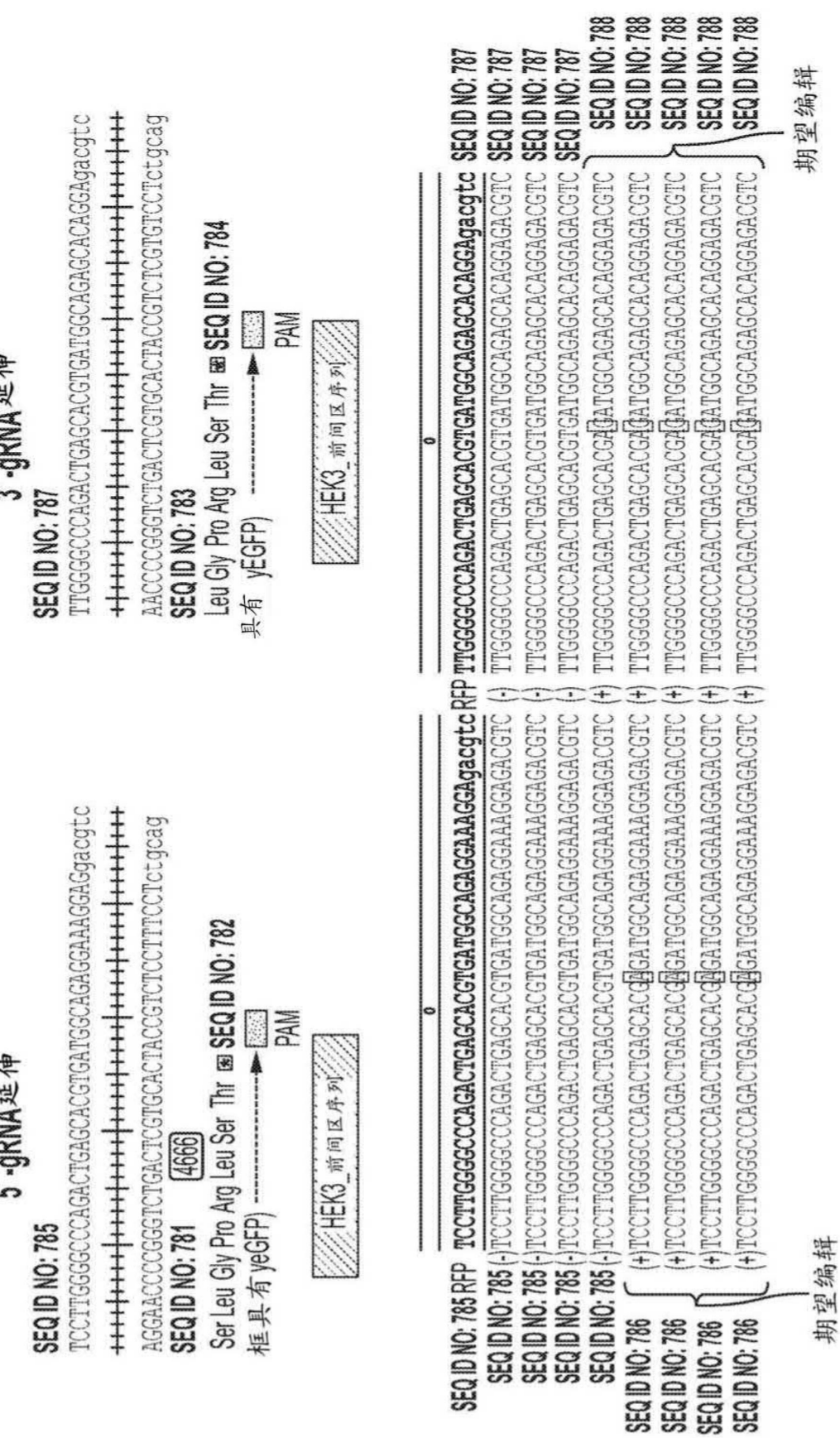

5′-gRNA延伸
SEQ ID NO: 785
TCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAGgacgtc
AGGAACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTctgcag
SEQ ID NO: 781
4666
Ser Leu Gly Pro Arg Leu Ser Thr
SEQ ID NO: 782
框具有 yeGFP)
PAM
HEK3_前间区序列
3′-gRNA延伸
SEQ ID NO: 787
TTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAGgacgtc
AACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTctgcag
SEQ ID NO: 783
Leu Gly Pro Arg Leu Ser Thr
SEQ ID NO: 784
具有 yEGFP)
PAM
HEK3_前间区序列
SEQ ID NO: 785 RFP
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 786 (+)
SEQ ID NO: 786 (+)
SEQ ID NO: 786 (+)
SEQ ID NO: 786 (+)
SEQ ID NO: 786 (+)
期望编辑
TTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAGgacgtc RFP
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 788
SEQ ID NO: 788
SEQ ID NO: 788
SEQ ID NO: 788
SEQ ID NO: 788
期望编辑

图12

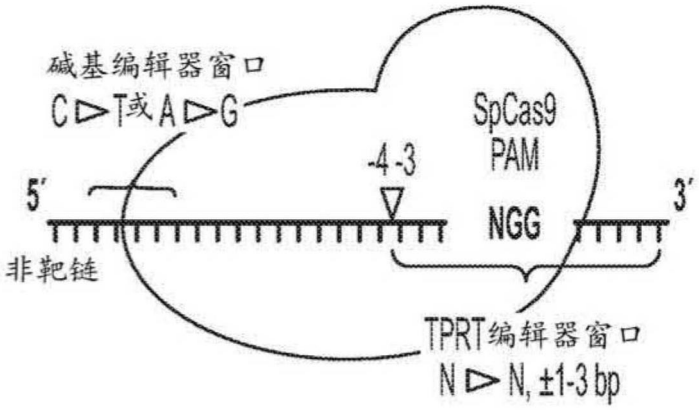

碱基编辑器窗口
C▷T或A▷G
SpCas9
PAM
-4 -3
5′
3′
NGG
非靶链
TPRT编辑器窗口
N▷N, ±1-3 bp

图13

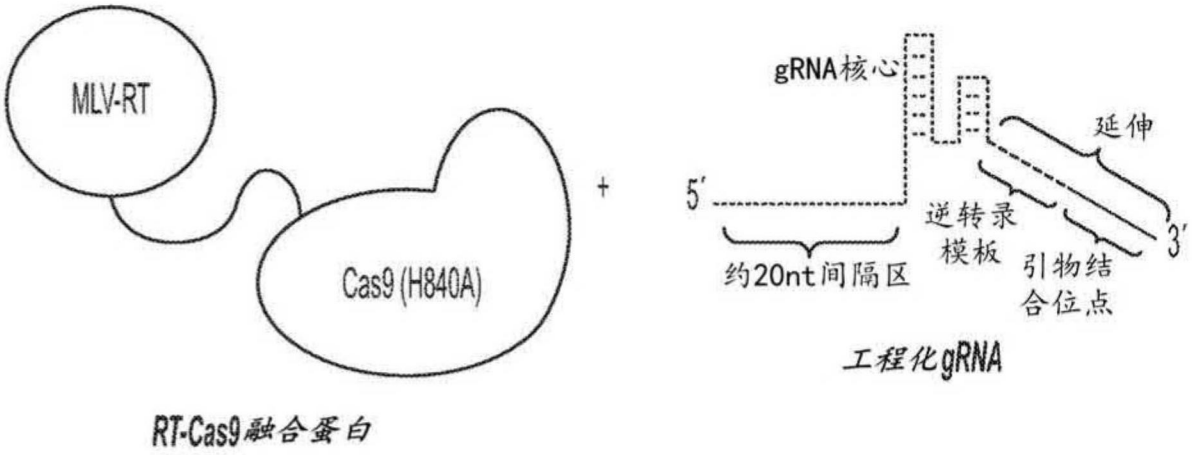

MLV-RT
+
Cas9 (H840A)
RT-Cas9融合蛋白
gRNA核心
延伸
5′
3′
逆转录
模板
引物结
合位点
约20nt间隔区
工程化gRNA

图14

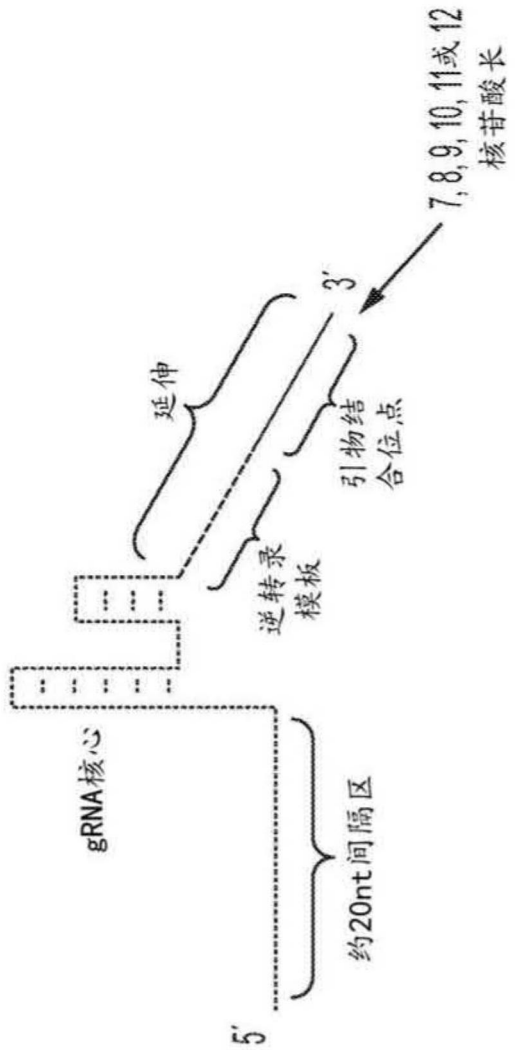


图15

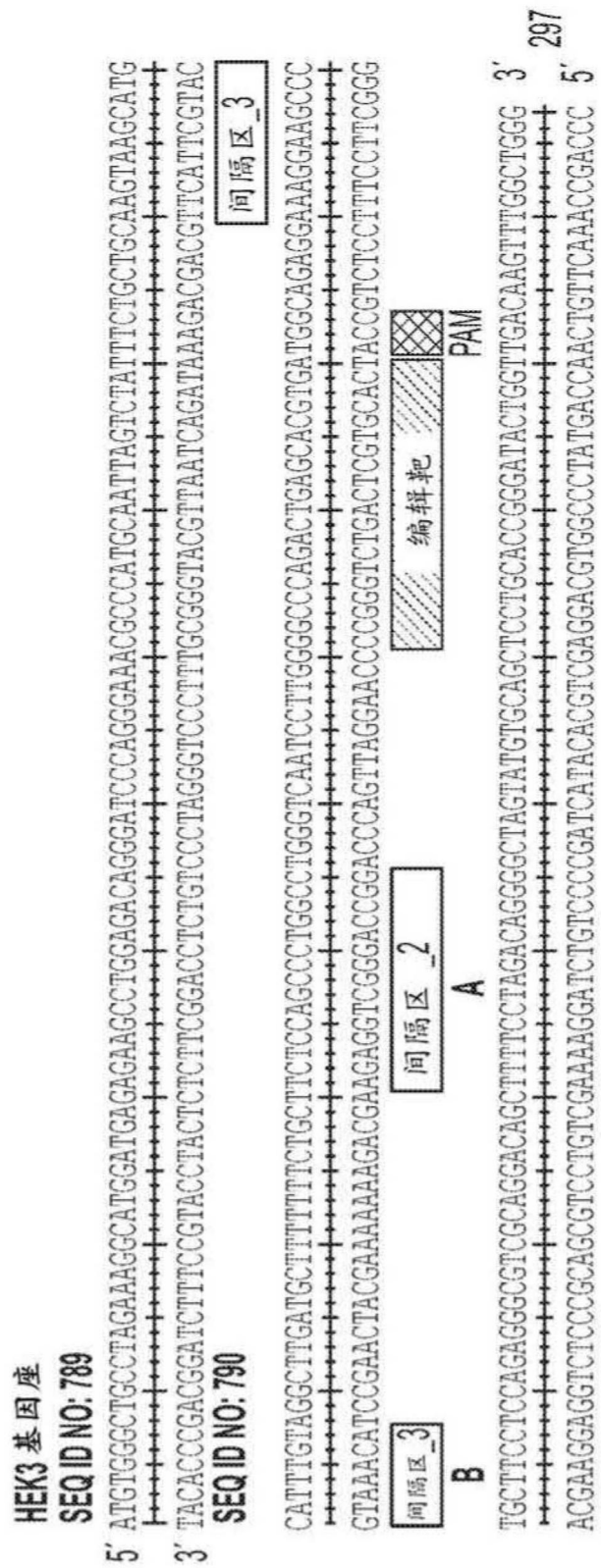

HEK3 基因座
SEQ ID NO: 789
5′ ATGTGGGCTGCCTAGAAAGGCATGATGAGAGAAGCCTGGAGACAGGGATCCCAGGGAAACGCCCATGCAATTAGTCTATTTCTGCTGCAAGTAAGCATG
3′ TACACCCGACGGATCTTTCCGTACCTACTCTCTTCGGACCTCTGTCCCTAGGGTCCCTTTGCGGGTACGTTAATCAGATAAAGACGACGTTCATTCGTAC
SEQ ID NO: 790
间隔区_3
CATTTGTAGGCTTGATGCTTTTTTTTCTGCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAGGAAGCCC
GTAAACATCCGAACTACGAAAAAAAGACGAAGAGGTCGGGACCGGACCCAGTTAGGAACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTTCGGG
间隔区_3
B
间隔区_2
A
编辑靶
PAM
TGCTTCCTCCAGAGGGCGTCGGCAGGACAGCTTTTTCCTAGACAGGGGCTAGTATGTGCAGCTCCTGCACCGGGATACTGGTTGACAAGTTTGGCTGGG 3′
ACGAAGGAGGTCTCCCCGCAGCCGTCCTGTCGAAAAGGATCTGTCCCCCGATCATACACGTCGAGGACGGTGGCCCTATGACCAACTGTTCAAACCGACCC 5′ 297
间隔区_4_rev
C
间隔区_5_rev
D
间隔区_6
E

图16

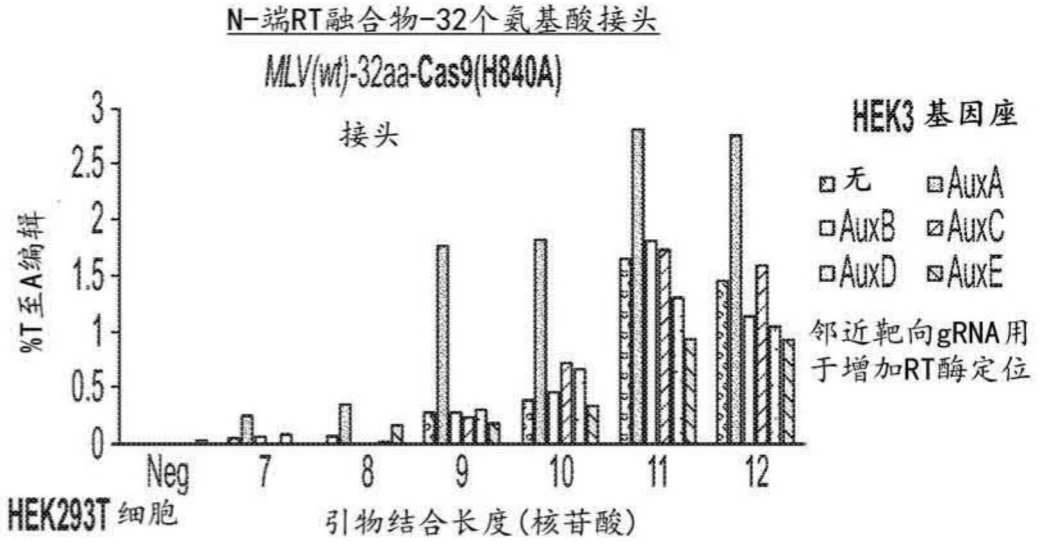

N-端RT融合物-32个氨基酸接头
MLV(wt)-32aa-Cas9(H840A)
接头
HEK3 基因座
无
AuxA
AuxB
AuxC
AuxD
AuxE
邻近靶向gRNA用于增加RT酶定位
%T至A编辑
3
2.5
2
1.5
1
0.5
0
Neg
7
8
9
10
11
12
HEK293T 细胞
引物结合长度(核苷酸)

图17A

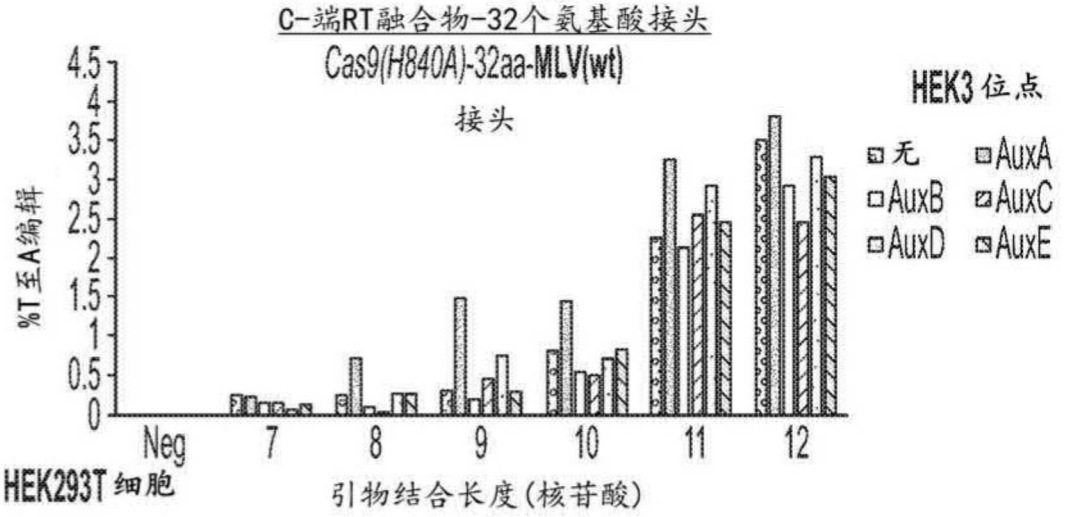

C-端RT融合物-32个氨基酸接头
Cas9(H840A)-32aa-MLV(wt)
接头
HEK3 位点
无
AuxA
AuxB
AuxC
AuxD
AuxE
%T至A编辑
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Neg
7
8
9
10
11
12
HEK293T 细胞
引物结合长度(核苷酸)

图17B

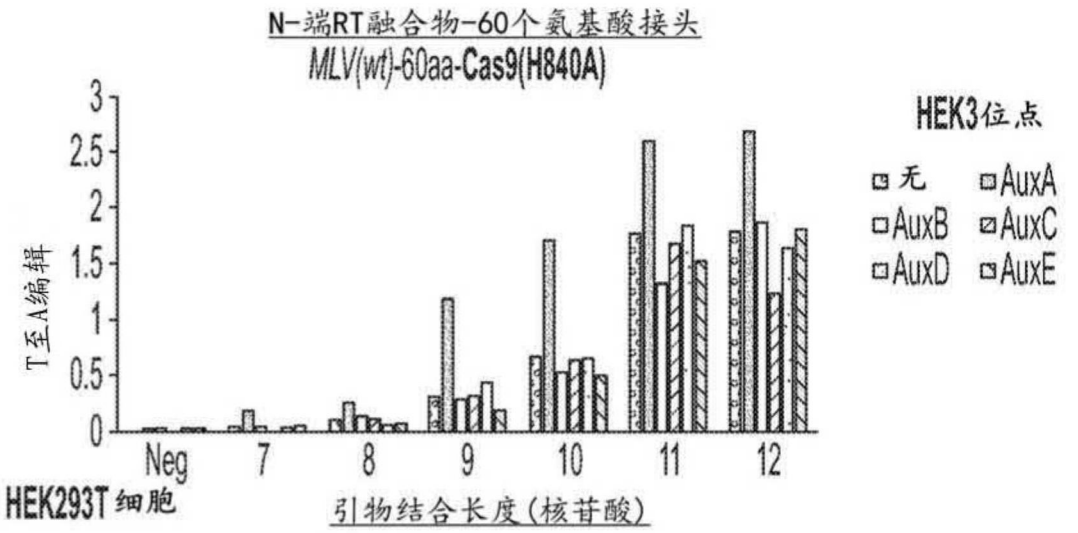

N-端RT融合物-60个氨基酸接头
MLV(wt)-60aa-Cas9(H840A)
HEK3位点
无
AuxA
AuxB
AuxC
AuxD
AuxE
T至A编辑
3
2.5
2
1.5
1
0.5
0
Neg
7
8
9
10
11
12
HEK293T细胞
引物结合长度(核苷酸)

图17C

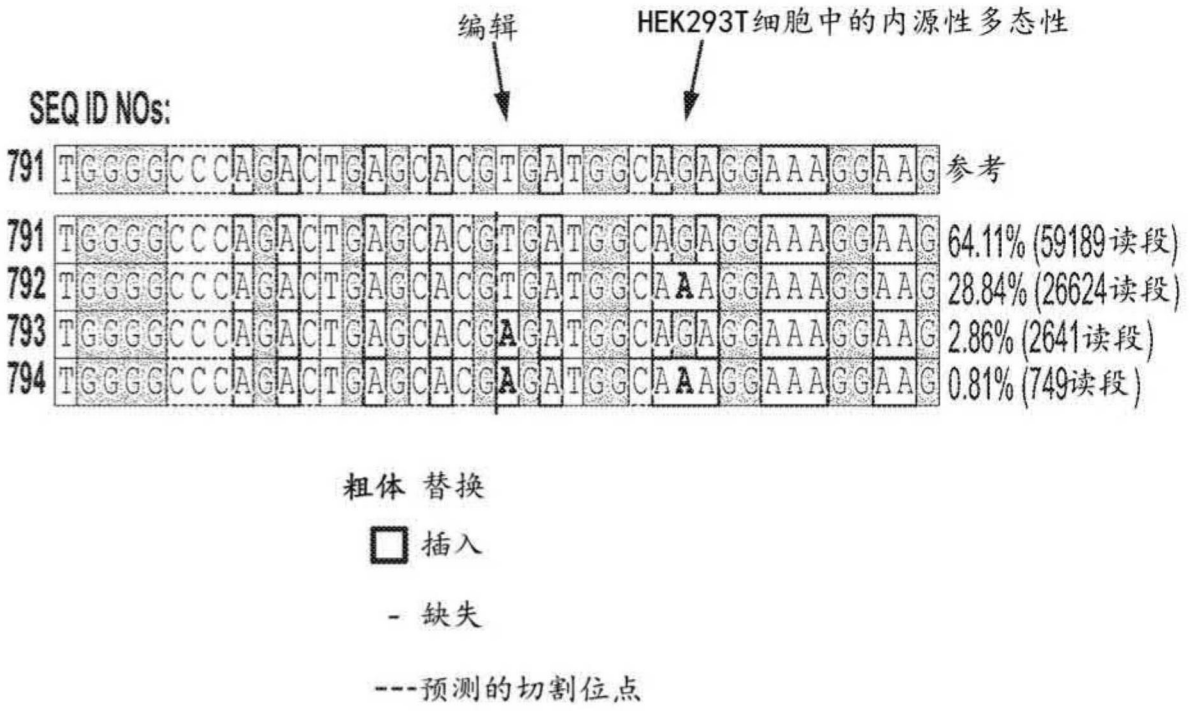

编辑
HEK293T细胞中的内源性多态性
SEQ ID NOs:
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG 参考
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG 64.11% (59189读段)
792 TGGGGCCCAGACTGAGCACGTGATGGCAAAGGAAAGGAAG 28.84% (26624读段)
793 TGGGGCCCAGACTGAGCACGAGATGGCAGAGGAAAGGAAG 2.86% (2641读段)
794 TGGGGCCCAGACTGAGCACGAGATGGCAAAGGAAAGGAAG 0.81% (749读段)
粗体 替换
插入
- 缺失
---预测的切割位点

图18

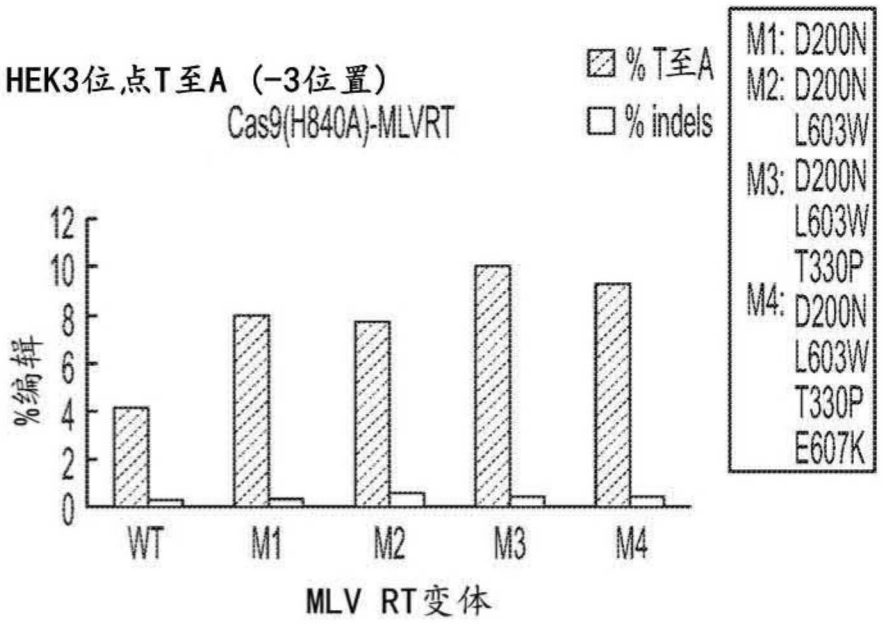

HEK3位点T至A（-3位置）
Cas9(H840A)-MLVRT
% T至A
% indels
M1: D200N
M2: D200N
L603W
M3: D200N
L603W
T330P
M4: D200N
L603W
T330P
E607K
%编辑
0
2
4
6
8
10
12
WT
M1
M2
M3
M4
MLV RT变体

图19

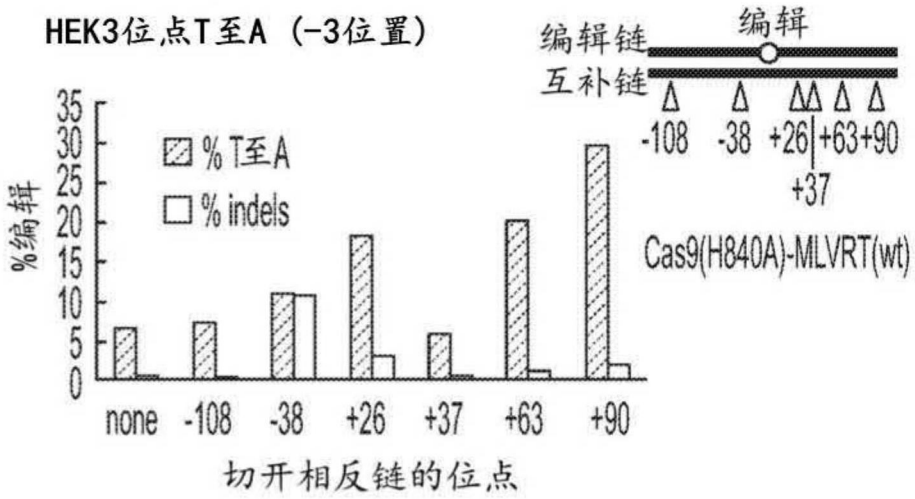

HEK3位点T至A（-3位置）
编辑链
编辑
互补链
-108
-38
+26
+37
+63
+90
Cas9(H840A)-MLVRT(wt)
% T至A
% indels
%编辑
0
5
10
15
20
25
30
35
none
-108
-38
+26
+37
+63
+90
切开相反链的位点

图20

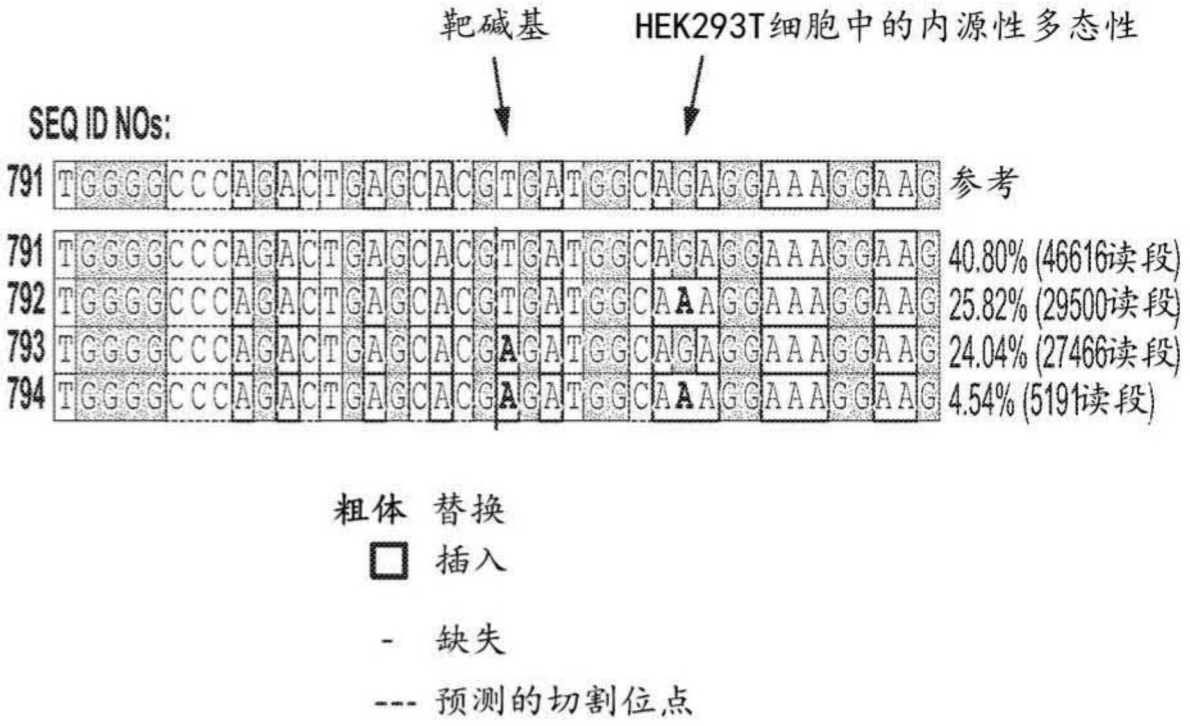

靶碱基
HEK293T细胞中的内源性多态性
SEQ ID NOs:
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG 参考
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG 40.80% (46616读段)
792 TGGGGCCCAGACTGAGCACGTGATGGCAAAGGAAAGGAAG 25.82% (29500读段)
793 TGGGGCCCAGACTGAGCACGAGATGGCAGAGGAAAGGAAG 24.04% (27466读段)
794 TGGGGCCCAGACTGAGCACGAGATGGCAAAGGAAAGGAAG 4.54% (5191读段)
粗体 替换
插入
- 缺失
--- 预测的切割位点

图21

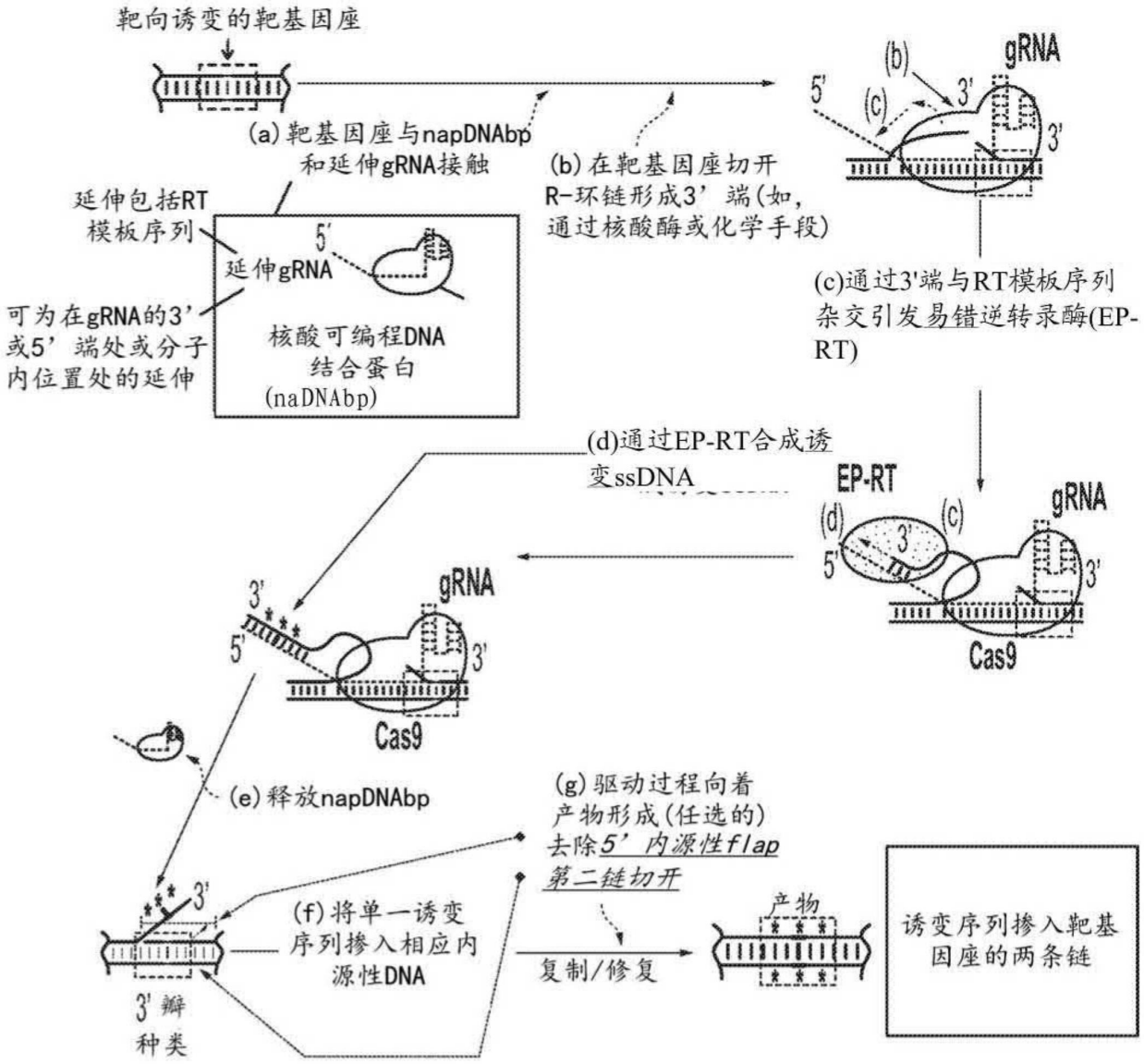

靶向诱变的靶基因座
(a)靶基因座与napDNAbp和延伸gRNA接触
(b)在靶基因座切开R-环链形成3’端(如，通过核酸酶或化学手段)
gRNA
(b)
(c)
5’
3’
3’
延伸包括RT模板序列
延伸gRNA
5’
可为在gRNA的3’或5’端处或分子内位置处的延伸
核酸可编程DNA结合蛋白(naDNAbp)
(c)通过3'端与RT模板序列杂交引发易错逆转录酶(EP-RT)
(d)通过EP-RT合成诱变ssDNA
EP-RT
(d)
(c)
gRNA
5’
3’
3’
Cas9
3’
5’
gRNA
3’
Cas9
(e)释放napDNAbp
(g)驱动过程向着产物形成(任选的)去除5’内源性flap第二链切开
3’
(f)将单一诱变序列掺入相应内源性DNA
3’瓣种类
复制/修复
产物
诱变序列掺入靶基因座的两条链

图22

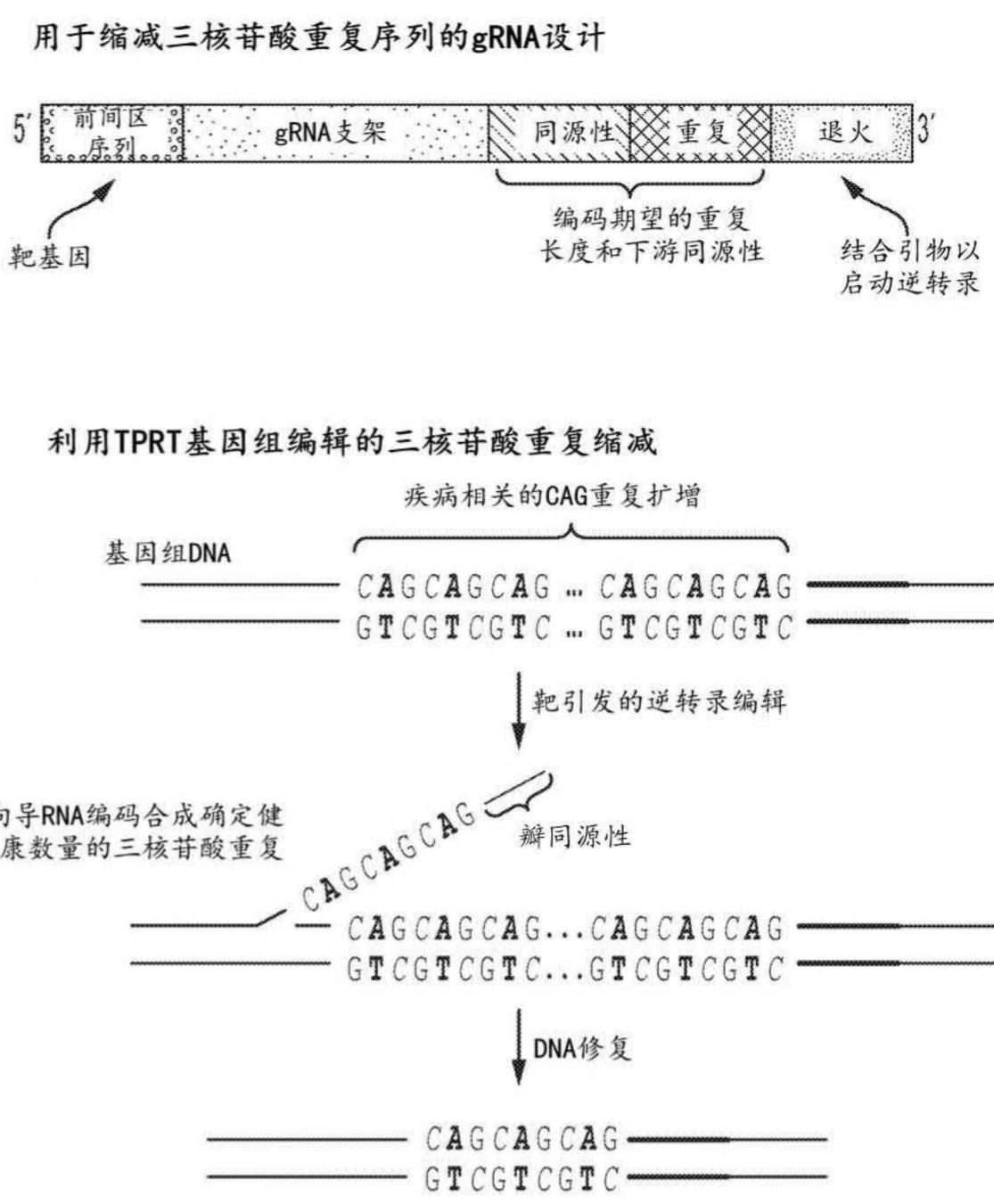

用于缩减三核苷酸重复序列的gRNA设计
5′
前间区序列
gRNA支架
同源性
重复
退火
3′
靶基因
编码期望的重复长度和下游同源性
结合引物以启动逆转录
利用TPRT基因组编辑的三核苷酸重复缩减
疾病相关的CAG重复扩增
基因组DNA
CAGCAGCAG … CAGCAGCAG
GTCGTCGTC … GTCGTCGTC
靶引发的逆转录编辑
向导RNA编码合成确定健康数量的三核苷酸重复
CAGCAGCAG
瓣同源性
CAGCAGCAG...CAGCAGCAG
GTCGTCGTC...GTCGTCGTC
DNA修复
CAGCAGCAG
GTCGTCGTC

图23

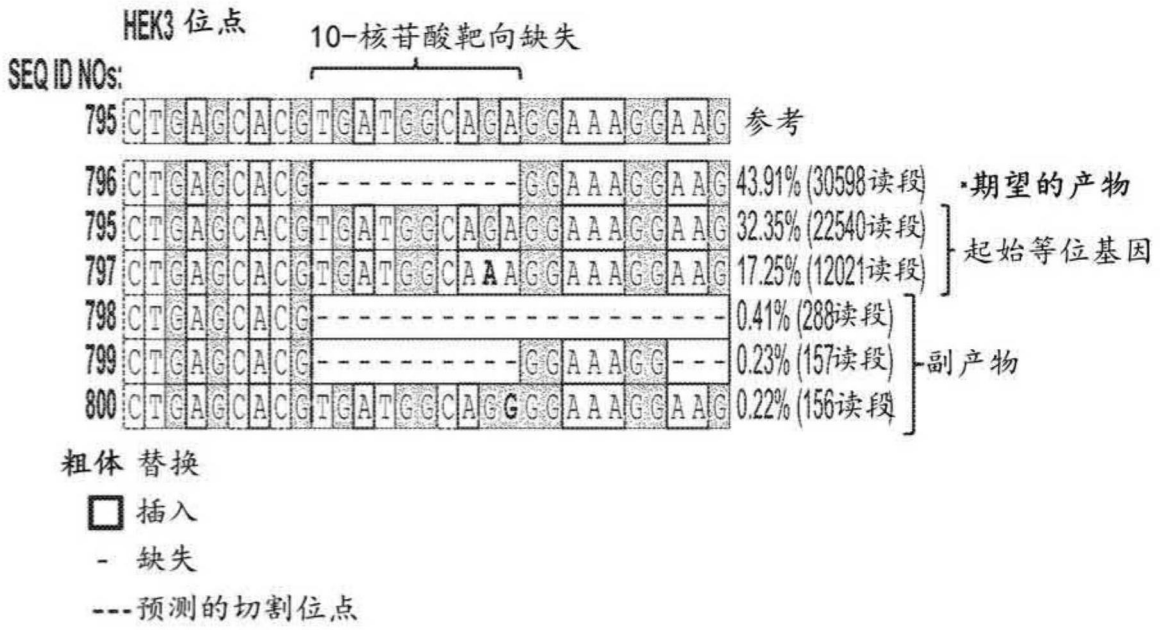

HEK3 位点
10-核苷酸靶向缺失
SEQ ID NOs:
795 CTGAGCACGTGATGGCAGAGGAAAGGAAG 参考
796 CTGAGCACG----------GGAAAGGAAG 43.91% (30598读段)
795 CTGAGCACGTGATGGCAGAGGAAAGGAAG 32.35% (22540读段)
797 CTGAGCACGTGATGGCAAAGGAAAGGAAG 17.25% (12021读段)
798 CTGAGCACG-------------------- 0.41% (288读段)
799 CTGAGCACG----------GGAAAGG--- 0.23% (157读段)
800 CTGAGCACGTGATGGCAGGGGAAAGGAAG 0.22% (156读段)
·期望的产物
起始等位基因
副产物
粗体 替换
插入
- 缺失
---预测的切割位点

图24

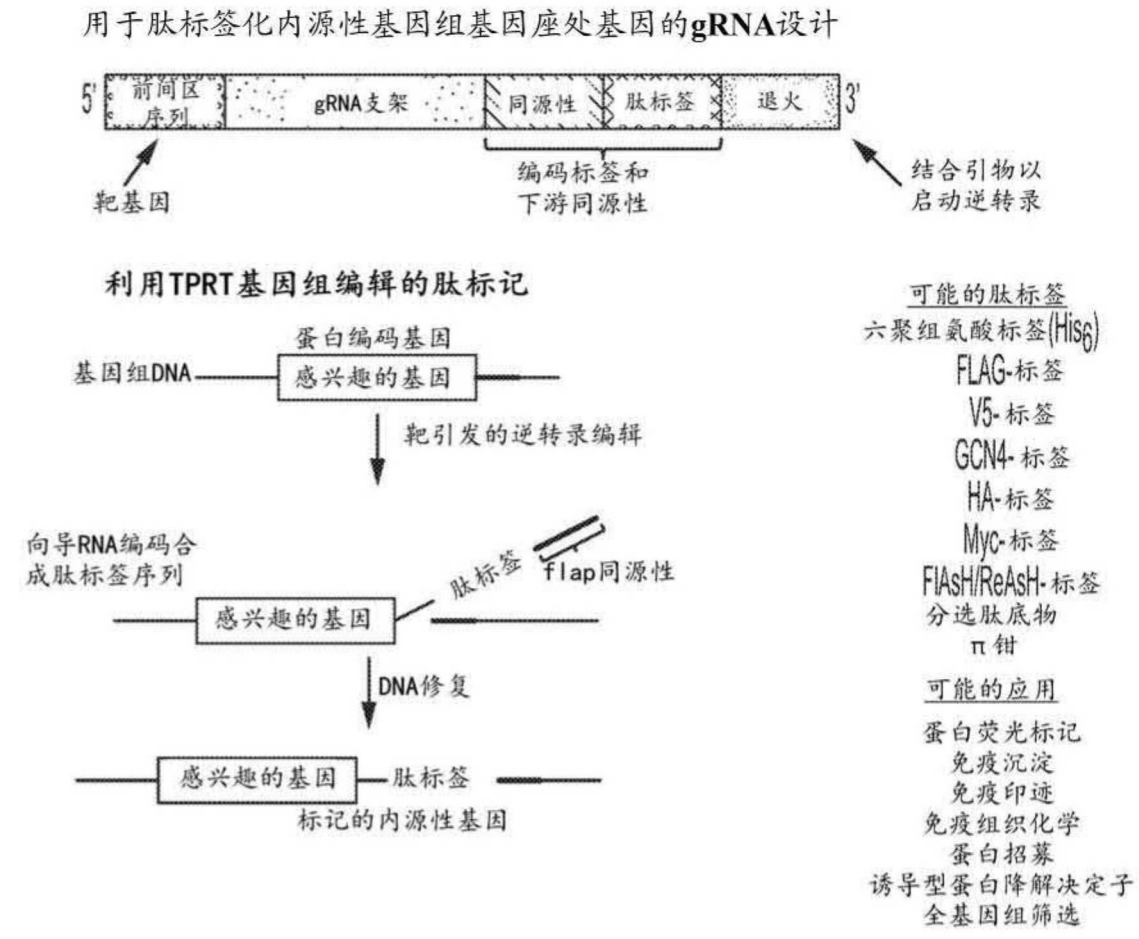

用于肽标签化内源性基因组基因座处基因的gRNA设计
5'
前间区序列
gRNA支架
同源性
肽标签
退火
3'
靶基因
编码标签和下游同源性
结合引物以启动逆转录
利用TPRT基因组编辑的肽标记
蛋白编码基因
基因组DNA
感兴趣的基因
靶引发的逆转录编辑
向导RNA编码合成肽标签序列
肽标签
flap同源性
感兴趣的基因
DNA修复
感兴趣的基因
肽标签
标记的内源性基因
可能的肽标签
六聚组氨酸标签(His6)
FLAG-标签
V5-标签
GCN4-标签
HA-标签
Myc-标签
FlAsH/ReAsH-标签
分选肽底物
π钳
可能的应用
蛋白荧光标记
免疫沉淀
免疫印迹
免疫组织化学
蛋白招募
诱导型蛋白降解决定子
全基因组筛选

图25

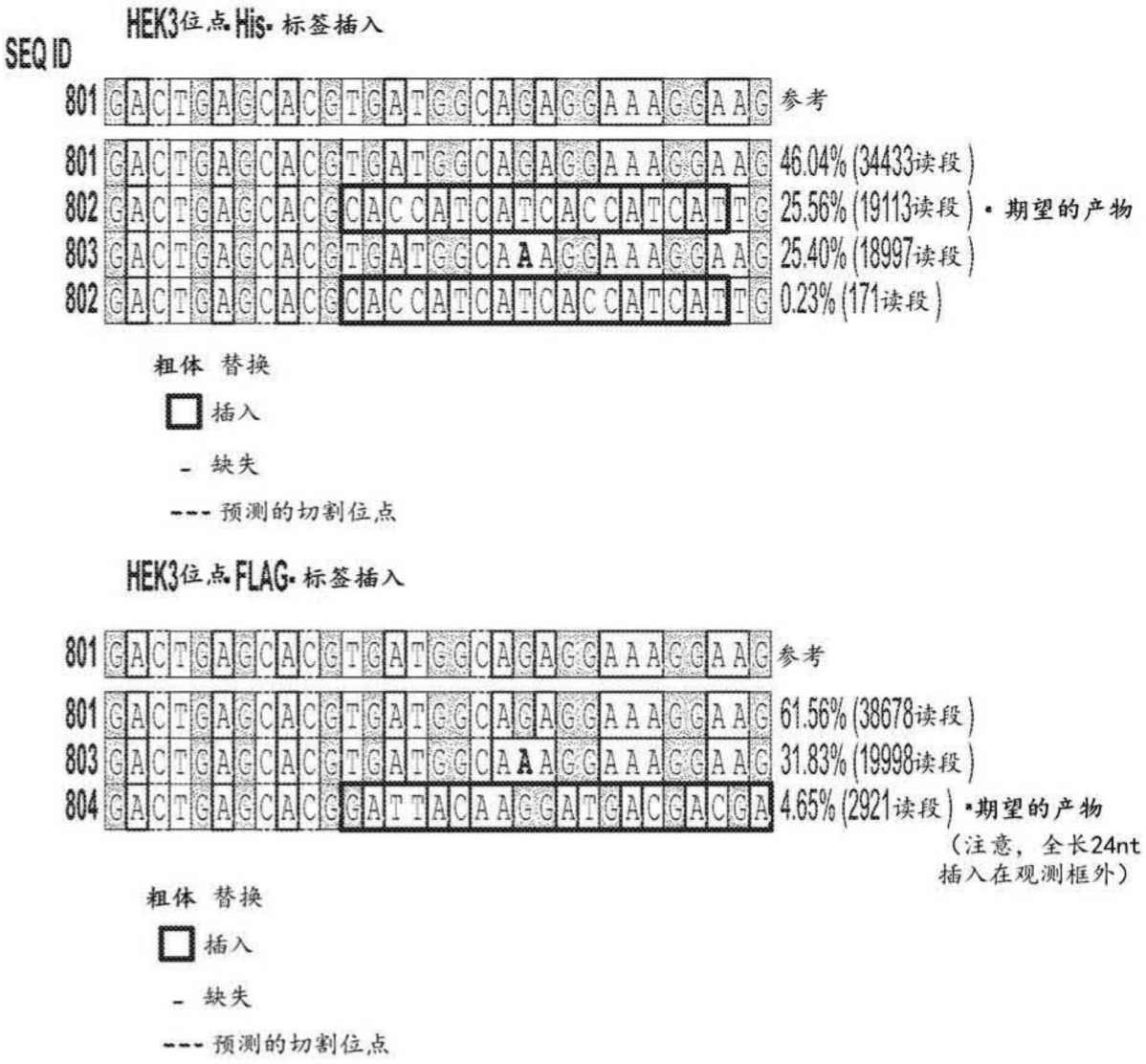

HEK3位点His-标签插入
SEQ ID
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 参考
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 46.04% (34433读段)
802 GACTGAGCACGCACCATCATCACCATCATTG 25.56% (19113读段) • 期望的产物
803 GACTGAGCACGTGATGGCAAAGGAAAGGAAG 25.40% (18997读段)
802 GACTGAGCACGCACCATCATCACCATCATTG 0.23% (171读段)
粗体 替换
插入
- 缺失
--- 预测的切割位点
HEK3位点FLAG-标签插入
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 参考
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 61.56% (38678读段)
803 GACTGAGCACGTGATGGCAAAGGAAAGGAAG 31.83% (19998读段)
804 GACTGAGCACGGATTACAAGGATGACGACGA 4.65% (2921读段) •期望的产物
（注意，全长24nt
插入在观测框外）
粗体 替换
插入
- 缺失
--- 预测的切割位点

图26A

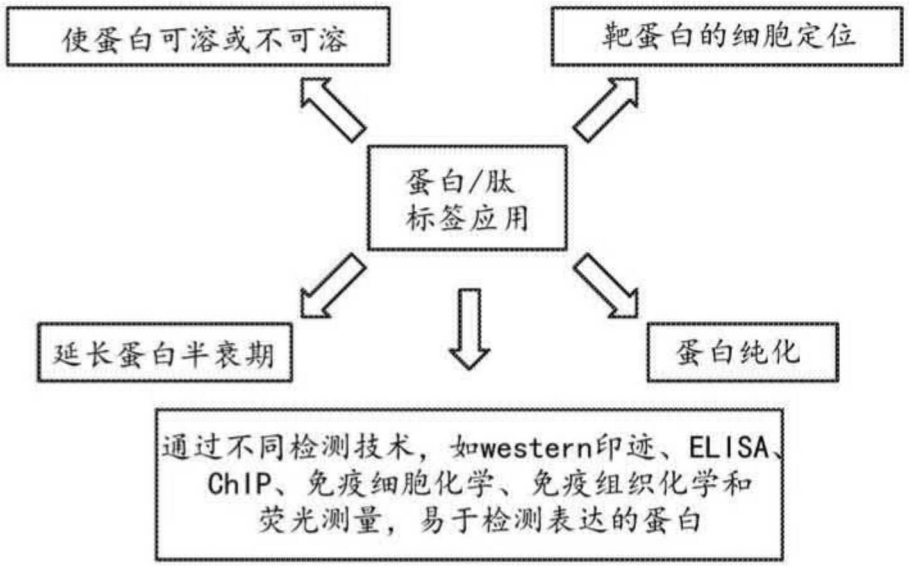
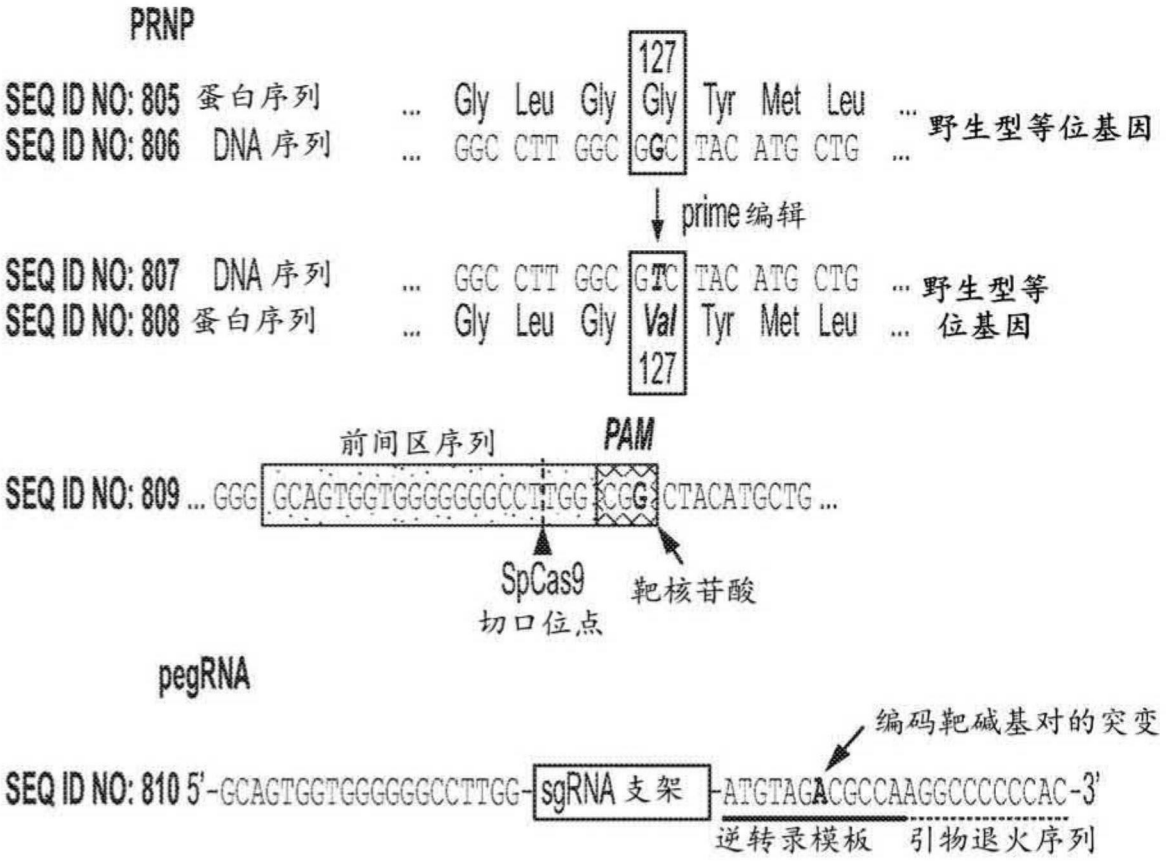

使蛋白可溶或不可溶
靶蛋白的细胞定位
蛋白/肽
标签应用
延长蛋白半衰期
蛋白纯化
通过不同检测技术，如western印迹、ELISA、
ChIP、免疫细胞化学、免疫组织化学和
荧光测量，易于检测表达的蛋白

图26B


PRNP
SEQ ID NO: 805 蛋白序列 ... Gly Leu Gly 127 Gly Tyr Met Leu ...
SEQ ID NO: 806 DNA 序列 ... GGC CTT GGC GGC TAC ATG CTG ...
野生型等位基因
prime编辑
SEQ ID NO: 807 DNA 序列 ... GGC CTT GGC GTC TAC ATG CTG ...
SEQ ID NO: 808 蛋白序列 ... Gly Leu Gly Val Tyr Met Leu ...
野生型等位基因
127
前间区序列
PAM
SEQ ID NO: 809 ... GGG GCAGTGGTGGGGGGGCCTTGG CGG CTACATGCTG ...
SpCas9
切口位点
靶核苷酸
pegRNA
编码靶碱基对的突变
SEQ ID NO: 810 5'-GCAGTGGTGGGGGGGCCTTGG-sgRNA 支架-ATGTAGACGCCAAGGCCCCCCAC-3'
逆转录模板
引物退火序列

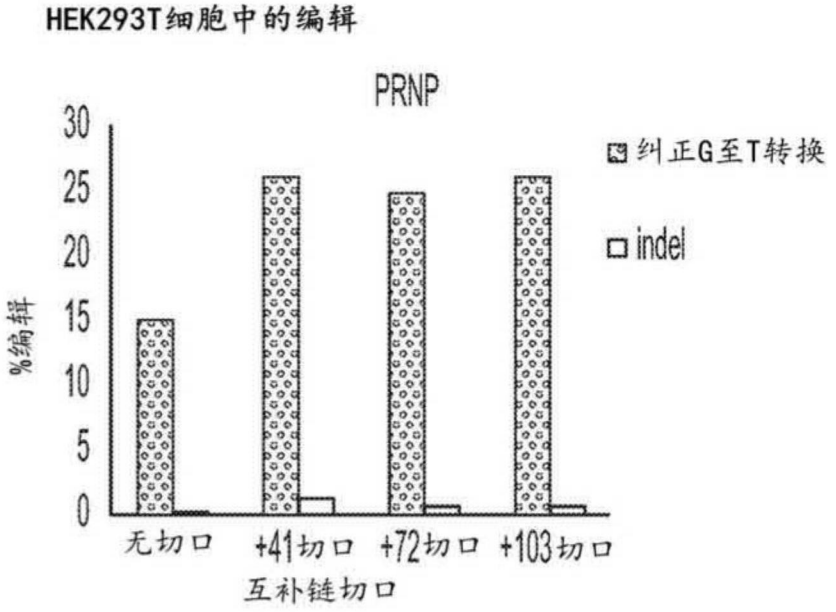

HEK293T细胞中的编辑
PRNP
%编辑
30
25
20
15
10
5
0
纠正G至T转换
indel
无切口
+41切口
+72切口
+103切口
互补链切口

图27

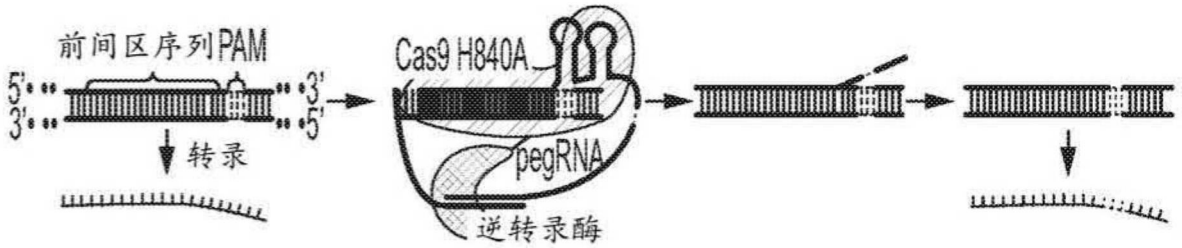

前间区序列PAM
5'
3'
3'
5'
转录
Cas9 H840A
pegRNA
逆转录酶

图28A

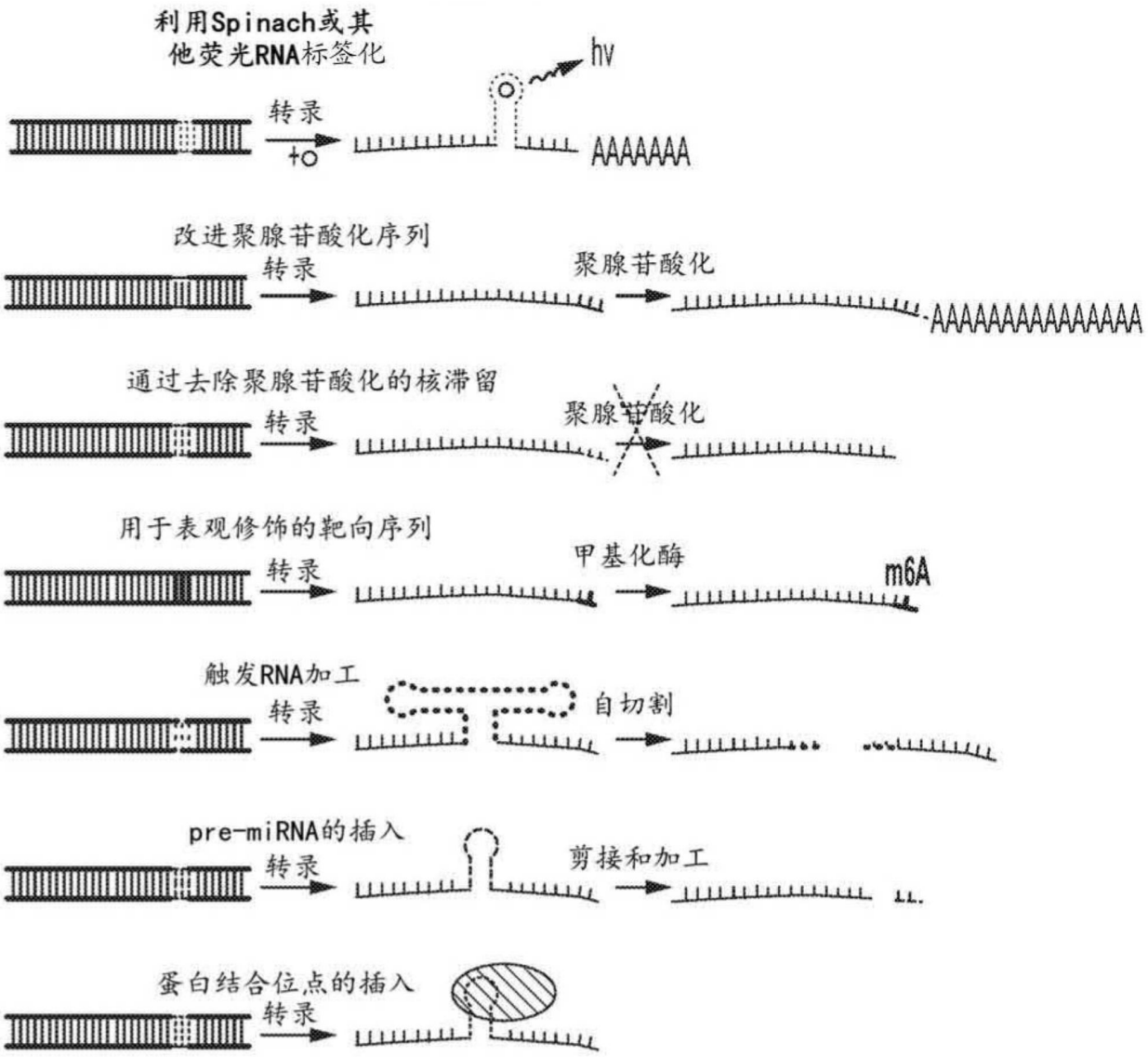

利用Spinach或其
他荧光RNA标签化
hv
转录
AAAAAAA
改进聚腺苷酸化序列
转录
聚腺苷酸化
AAAAAAAAAAAAAAA
通过去除聚腺苷酸化的核滞留
转录
聚腺苷酸化
用于表观修饰的靶向序列
转录
甲基化酶
m6A
触发RNA加工
转录
自切割
pre-miRNA的插入
转录
剪接和加工
蛋白结合位点的插入
转录

图28B

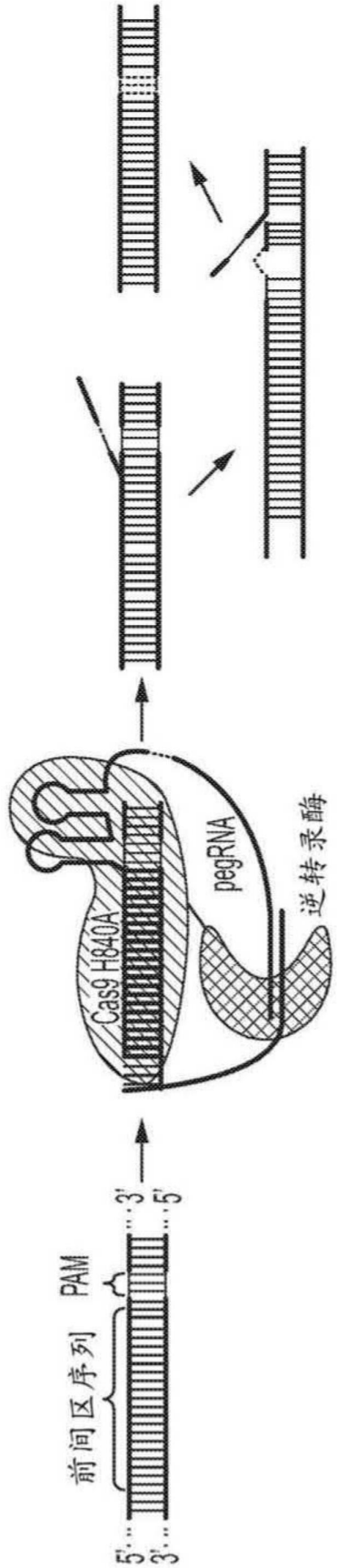


图29A

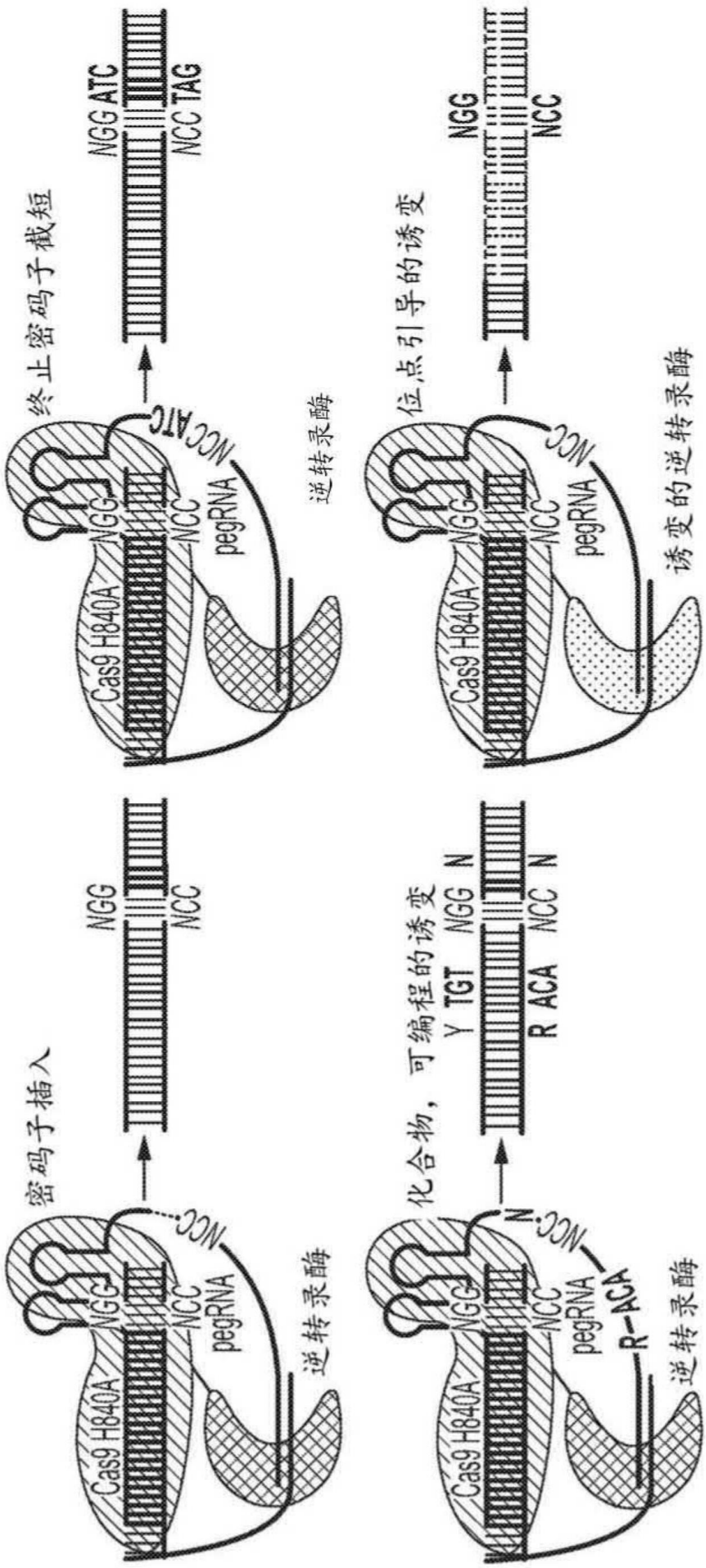

终止密码子截短
NGG ATC
NCC TAG
Cas9 H840A
NGG
NCC
pegRNA
逆转录酶
位点引导的诱变
NGG
NCC
诱变的逆转录酶
密码子插入
NGG
NCC
化合物，可编程的诱变
Y TGT NGG N
R ACA NCC N
R-ACA

图29B

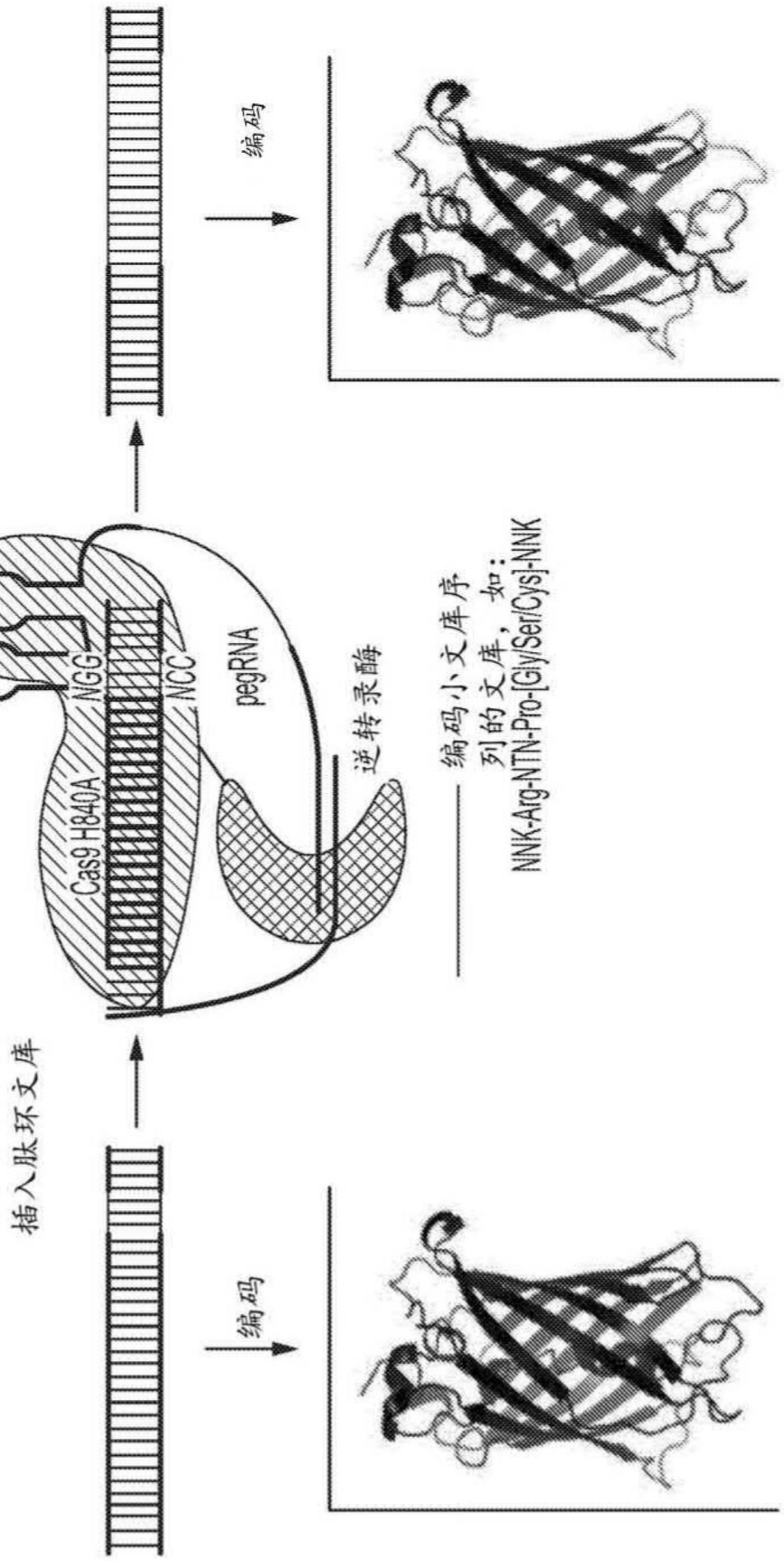

插入肽环文库
编码
编码
Cas9 H840A
NGG
NCC
pegRNA
逆转录酶
编码小文库序列的文库，如：
NNK-Arg-NTN-Pro-[Gly/Ser/Cys]-NNK

图29C

可编程密码子缺失和截短

SEQ ID NO:811	GFP	NFKIRHNIEDGSVQLADHYQQNTPI
SEQ ID NO:812	del5	NFKI-HNIEDGSVQLADHYQQNTPI
SEQ ID NO:813	del6	NFKIR-NIEDGSVQLADHYQQNTPI
SEQ ID NO:814	del8	NFKIRHN-EDGSVQLADHYQQNTPI
SEQ ID NO:815	del 8-13	NFKIRHN-----QLADHYQQNTPI
SEQ ID NO:816	del 1-5	-----HNIEDGSVQLADHYQQNTPI
SEQ ID NO:817	del 15-25	NFKIRHNIEDGSVQ----------

图29D

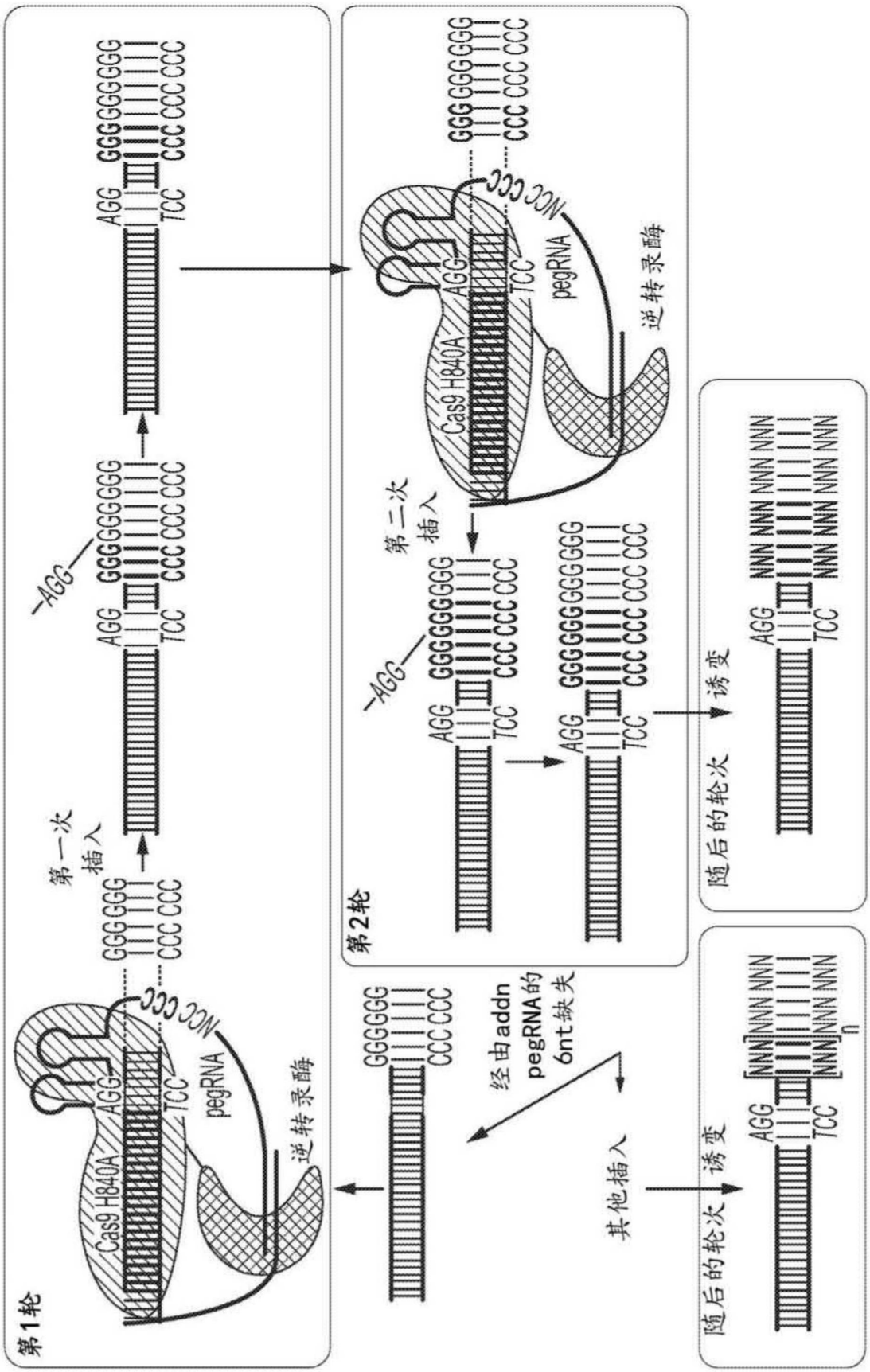

第1轮
Cas9 H840A
pegRNA
逆转录酶
第一次插入
第2轮
第二次插入
经由addn pegRNA的6nt缺失
其他插入
诱变
随后的轮次
诱变
随后的轮次

图30

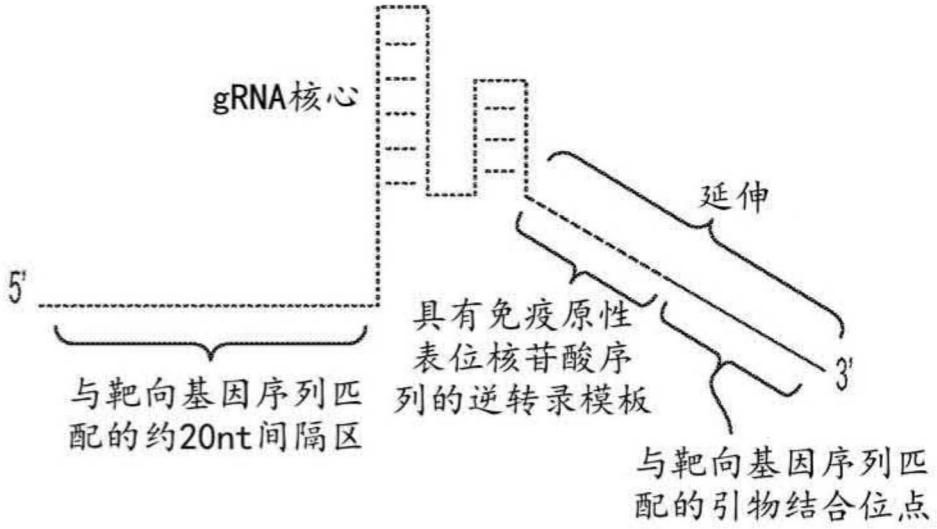

gRNA核心
延伸
5'
具有免疫原性
表位核苷酸序
列的逆转录模板
3'
与靶向基因序列匹
配的约20nt间隔区
与靶向基因序列匹
配的引物结合位点

图31

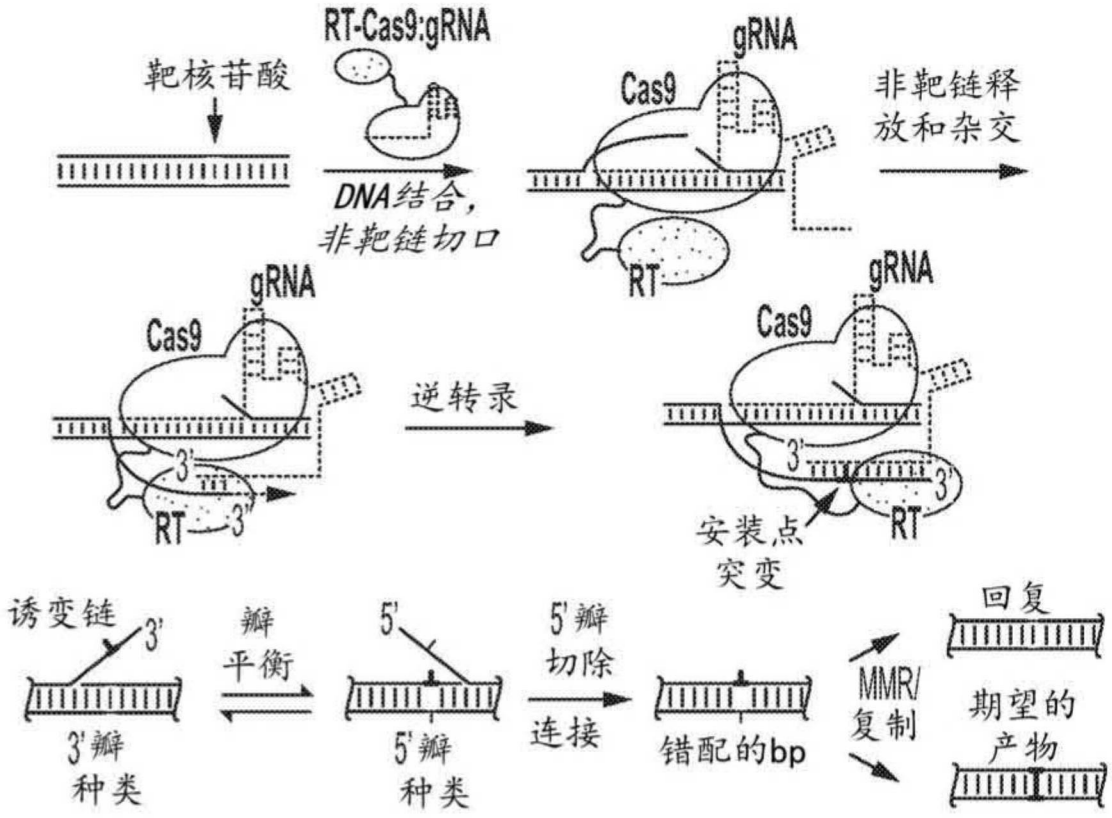
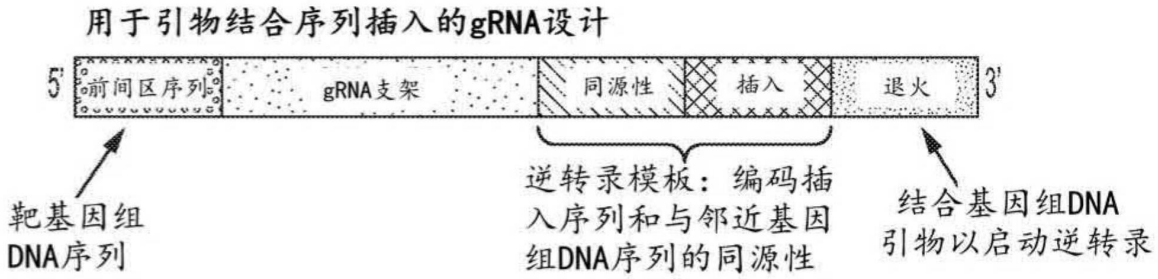

RT-Cas9:gRNA
靶核苷酸
DNA结合,
非靶链切口
gRNA
Cas9
非靶链释
放和杂交
RT
gRNA
Cas9
逆转录
3'
RT
3''
gRNA
Cas9
3'
3'
RT
安装点
突变
诱变链
3'
3'瓣
种类
瓣
平衡
5'
5'瓣
种类
5'瓣
切除
连接
错配的bp
MMR/
复制
回复
期望的
产物

图32



使用引导编辑将引物结合序列插入基因组**DNA**

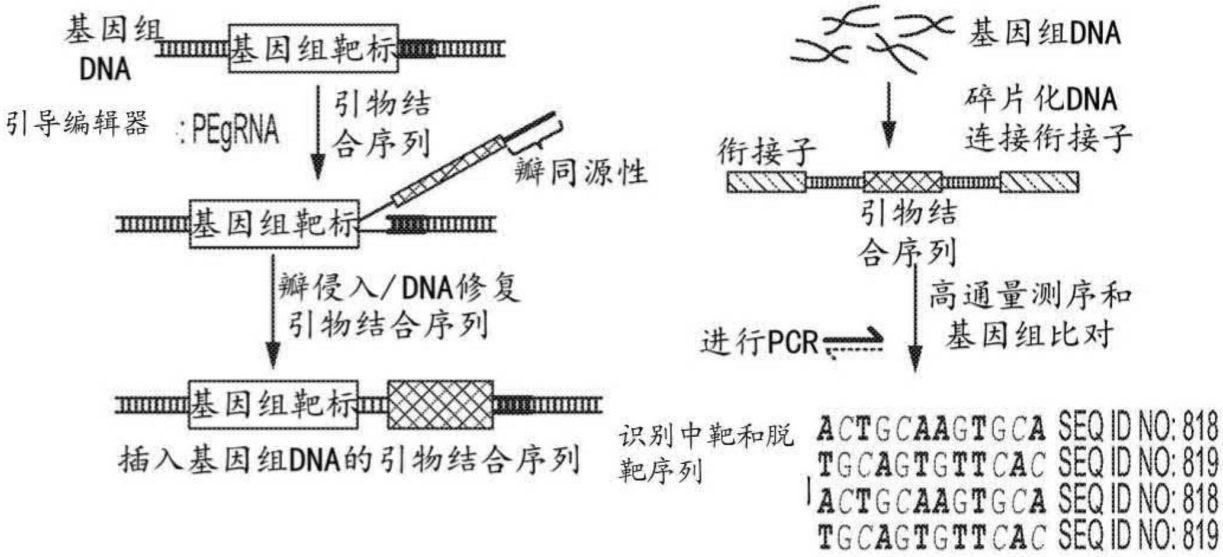


图33

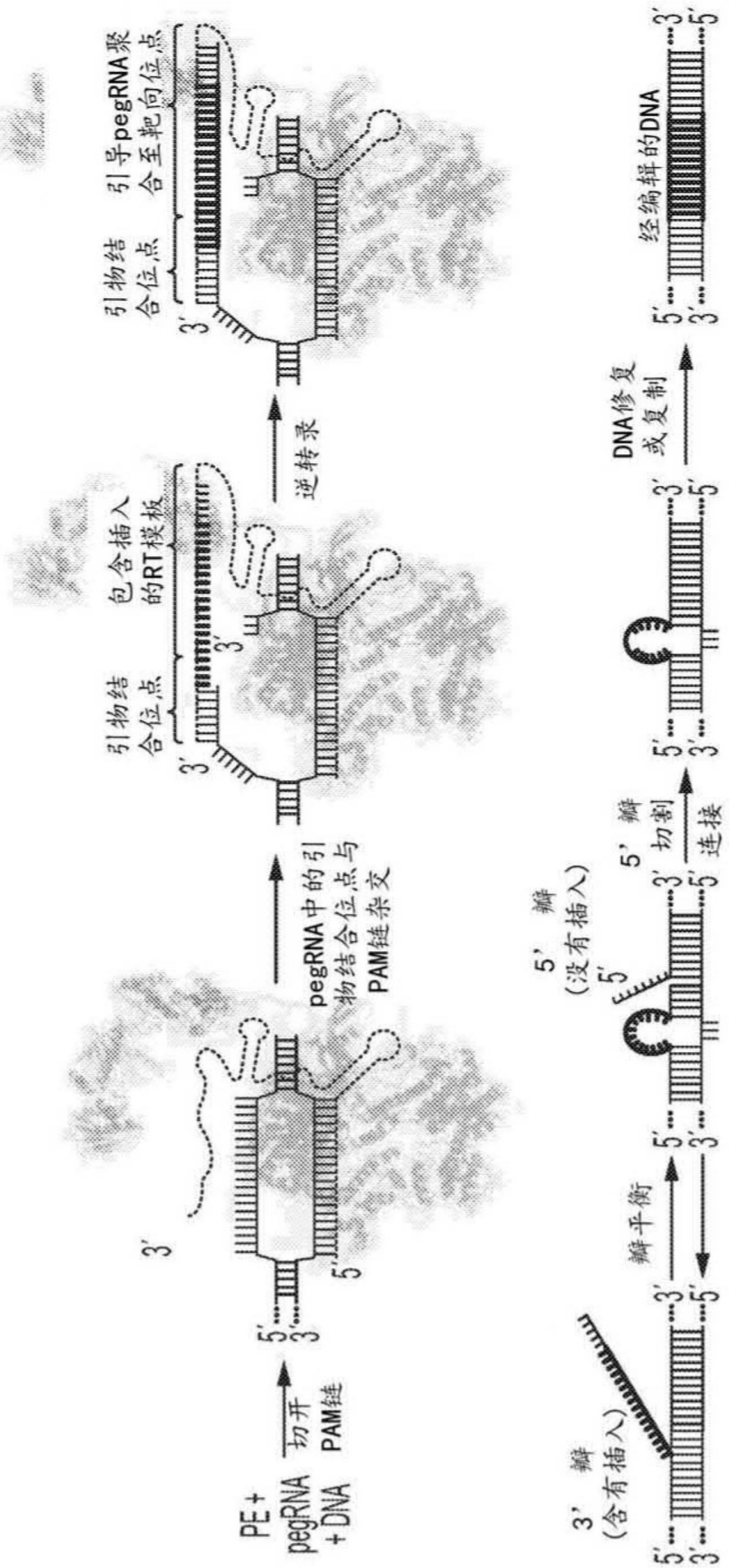

PE+
pegRNA
+DNA
切开
PAM链
pegRNA中的引
物结合位点与
PAM链杂交
引物结
合位点
包含插入
的RT模板
逆转录
引导pegRNA聚
合至靶向位点
3′瓣
(含有插入)
瓣平衡
5′瓣
(没有插入)
5′瓣
切割
连接
DNA修复
或复制
经编辑的DNA

图34

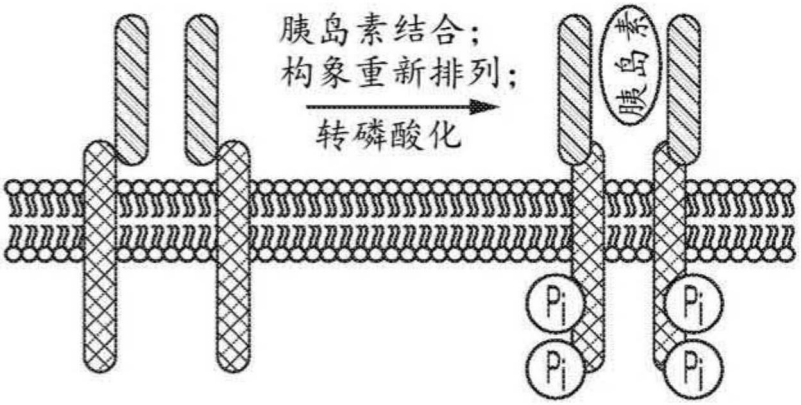

胰岛素结合;
构象重新排列;
转磷酸化
胰岛素
Pi
Pi
Pi
Pi

图35A

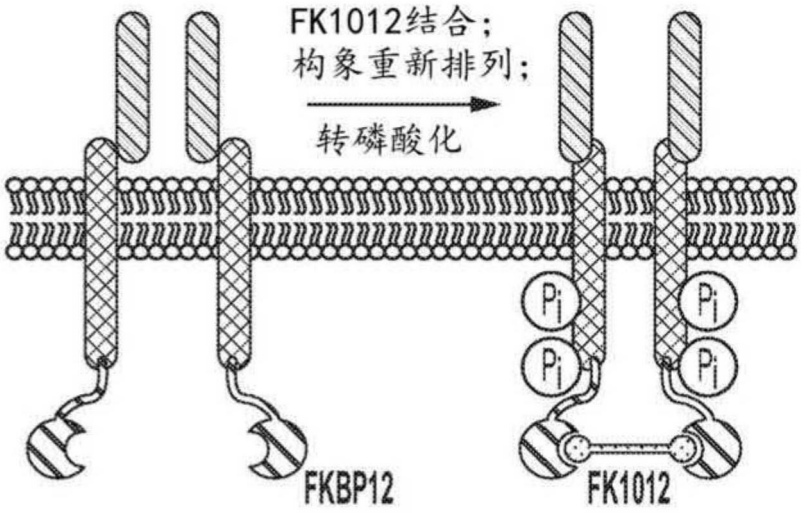

FK1012结合;
构象重新排列;
转磷酸化
Pi
Pi
Pi
Pi
FKBP12
FK1012

图35B

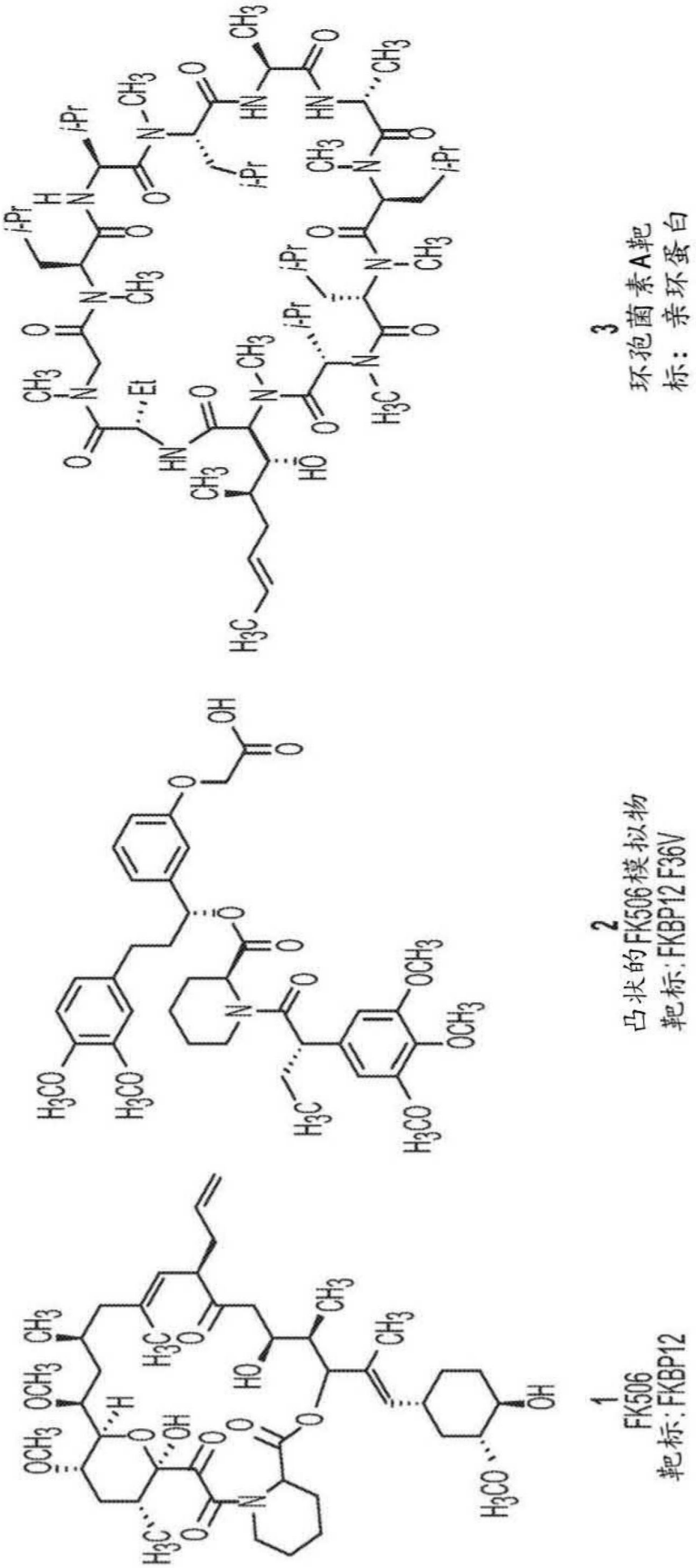

1
FK506
靶标:FKBP12
2
凸状的FK506模拟物
靶标:FKBP12 F36V
3
环孢菌素A靶
标:亲环蛋白

图36

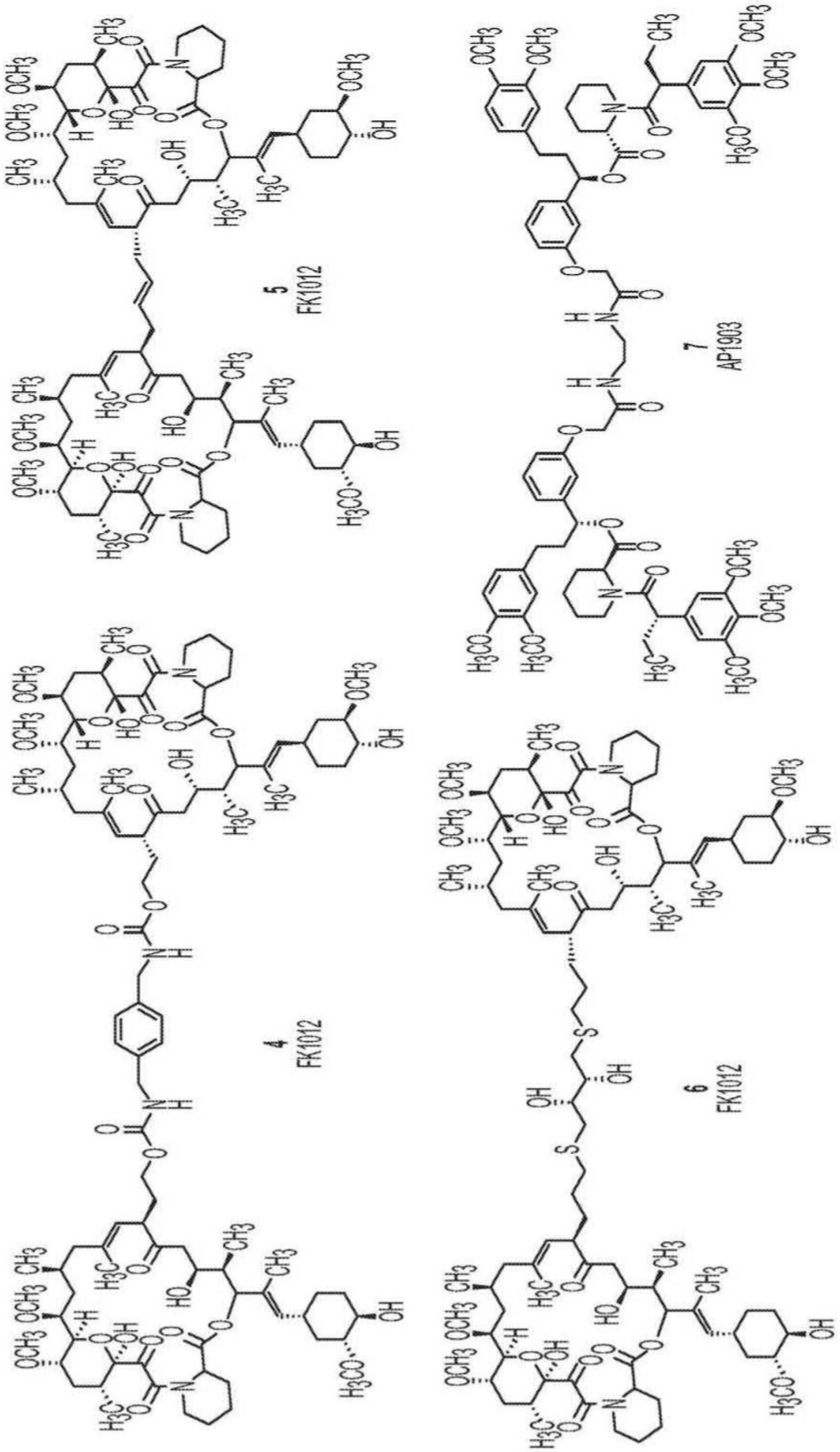


图37A

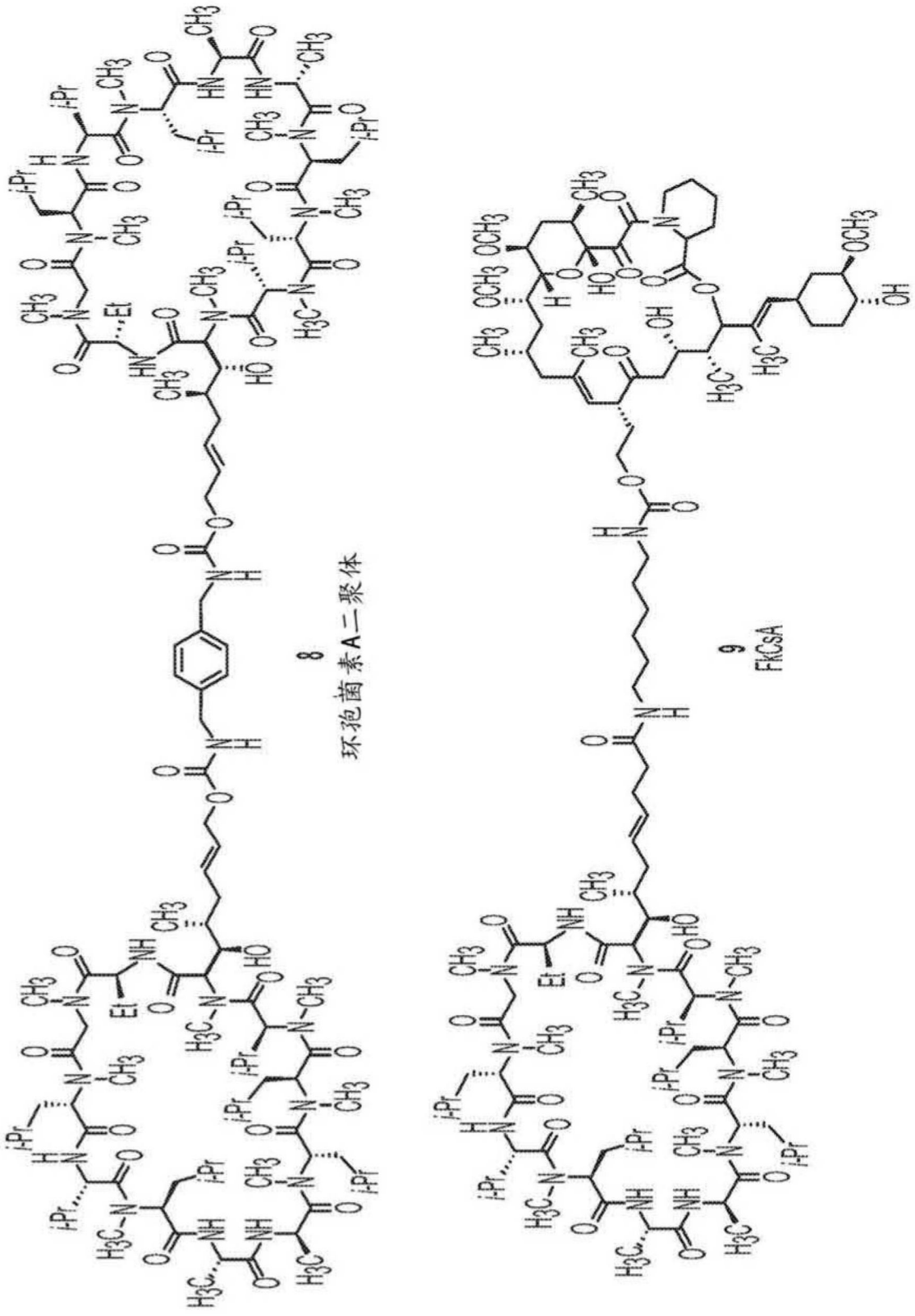


图37B

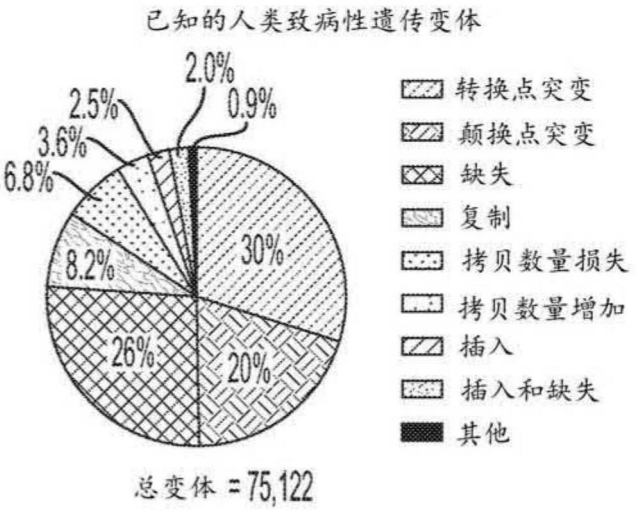

已知的人类致病性遗传变体
2.0%
2.5%
0.9%
3.6%
6.8%
8.2%
30%
26%
20%
转换点突变
颠换点突变
缺失
复制
拷贝数量损失
拷贝数量增加
插入
插入和缺失
其他
总变体 = 75,122

图38A

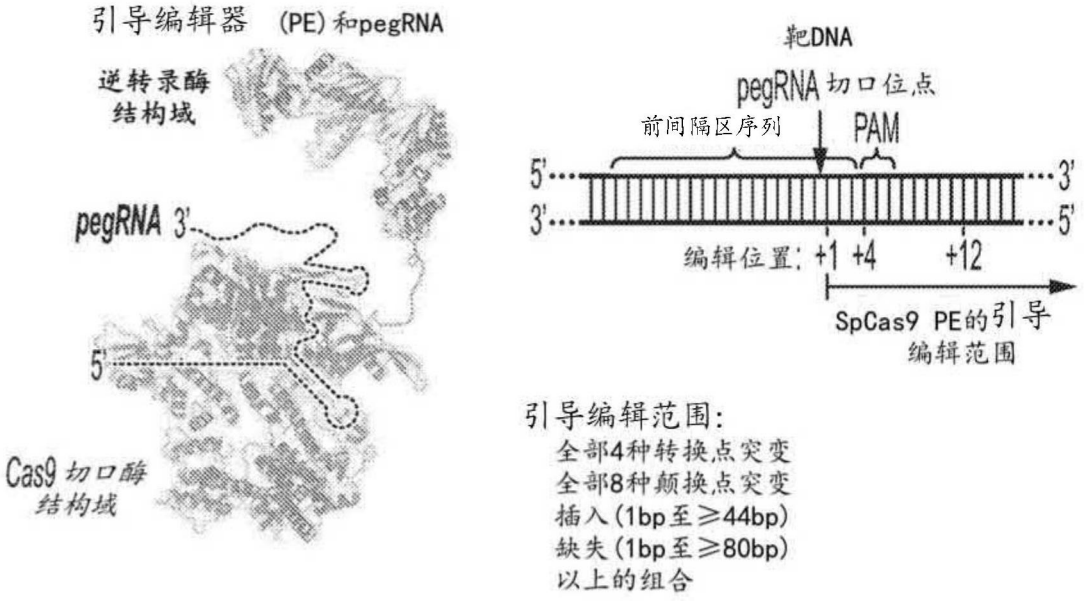

引导编辑器 (PE) 和pegRNA
逆转录酶
结构域
pegRNA 3'
5'
Cas9 切口酶
结构域
靶DNA
pegRNA 切口位点
前间隔区序列
PAM
5'
3'
3'
5'
编辑位置: +1 +4 +12
SpCas9 PE的引导
编辑范围
引导编辑范围:
全部4种转换点突变
全部8种颠换点突变
插入(1bp至≥44bp)
缺失(1bp至≥80bp)
以上的组合

图38B

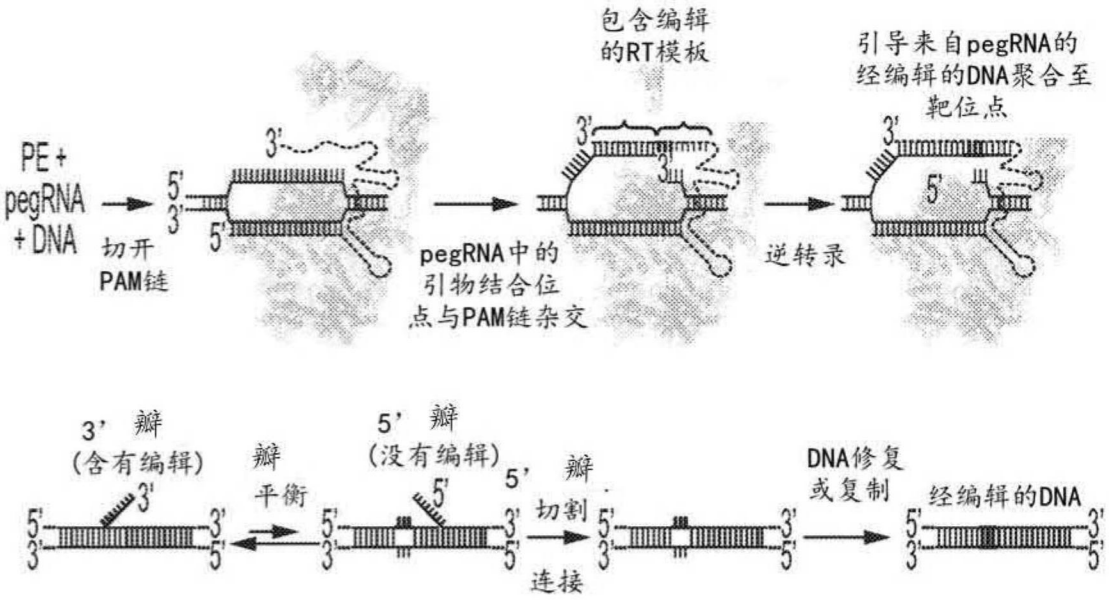

PE+
pegRNA
+DNA
切开
PAM链
pegRNA中的
引物结合位
点与PAM链杂交
包含编辑
的RT模板
逆转录
引导来自pegRNA的
经编辑的DNA聚合至
靶位点
3’ 瓣
(含有编辑)
瓣
平衡
5’ 瓣
(没有编辑)
5’ 瓣
切割
连接
DNA修复
或复制
经编辑的DNA

图38C

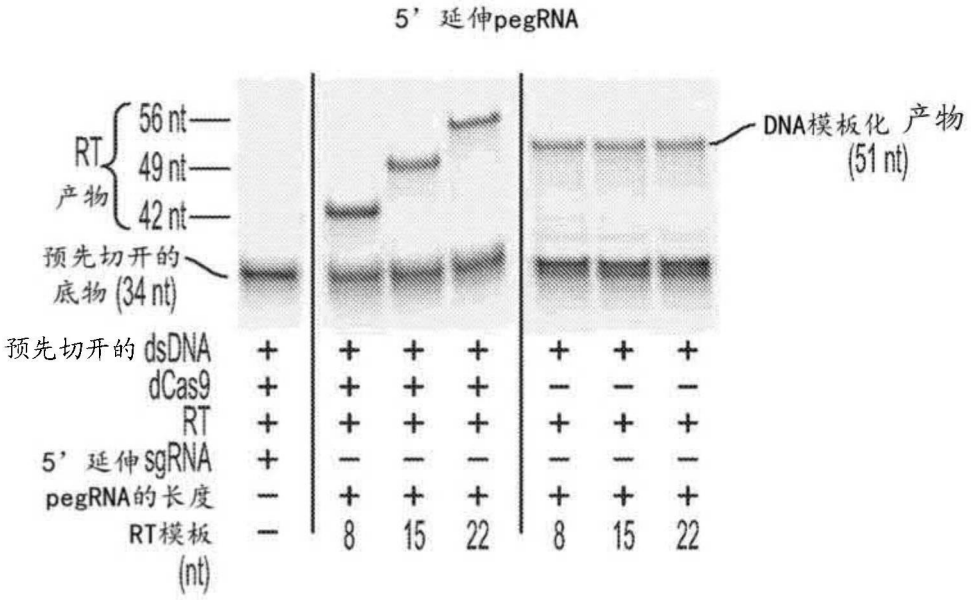

5’ 延伸pegRNA
RT
产物
56 nt
49 nt
42 nt
预先切开的
底物 (34 nt)
DNA模板化 产物
(51 nt)
预先切开的 dsDNA + + + + + + +
dCas9 + + + + − − −
RT + + + + + + +
5’ 延伸sgRNA + − − − − − −
pegRNA的长度 − + + + + + +
RT模板 (nt) − 8 15 22 8 15 22

图38D

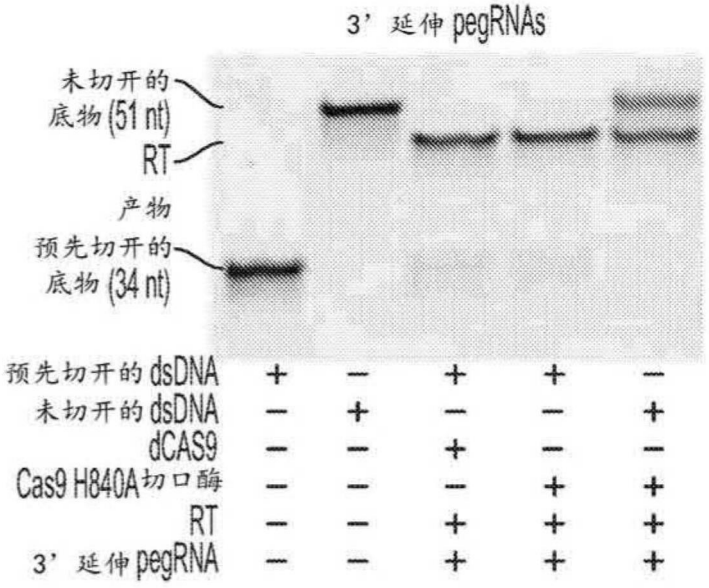

3’ 延伸 pegRNAs
未切开的
底物 (51 nt)
RT
产物
预先切开的
底物 (34 nt)
预先切开的 dsDNA + − + + −
未切开的 dsDNA − + − − +
dCAS9 − − + − −
Cas9 H840A切口酶 − − − + +
RT − − + + +
3’ 延伸 pegRNA − − + + +

图38E

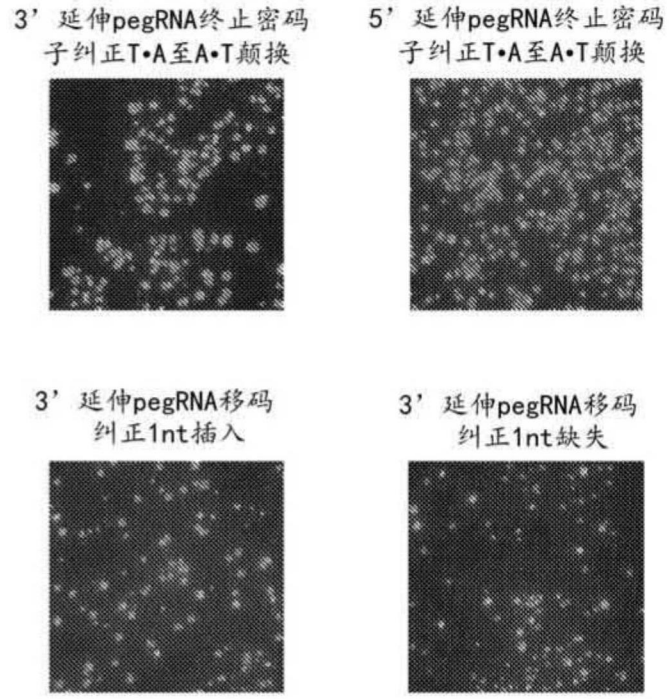

3’ 延伸pegRNA终止密码
子纠正T•A至A•T颠换
5’ 延伸pegRNA终止密码
子纠正T•A至A•T颠换
3’ 延伸pegRNA移码
纠正1nt插入
3’ 延伸pegRNA移码
纠正1nt缺失

图38F

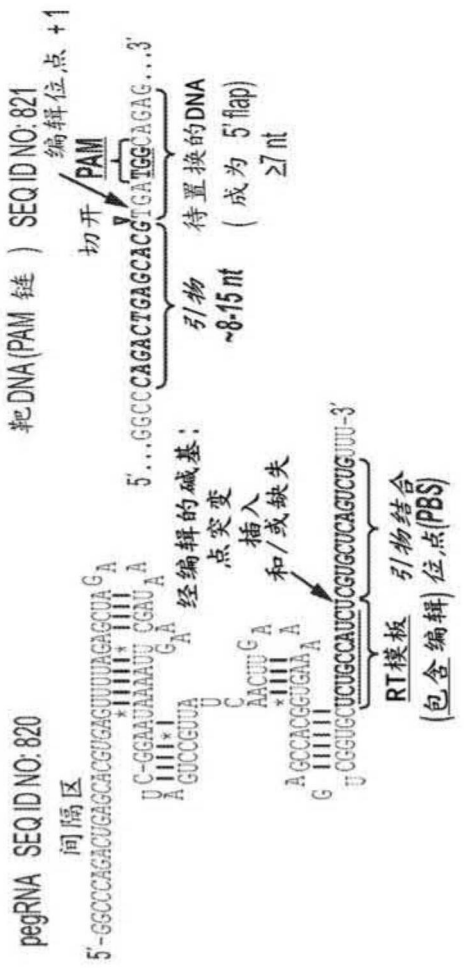

pegRNA SEQ ID NO: 820
间隔区
靶DNA(PAM 链) SEQ ID NO: 821
编辑位点 +1
切开
PAM
引物
~8-15 nt
待置换的DNA
(成为 5' flap)
≥7 nt
经编辑的碱基;
点突变
插入
和/或缺失
RT模板
引物结合
位点(PBS)
(包含编辑)

图39A

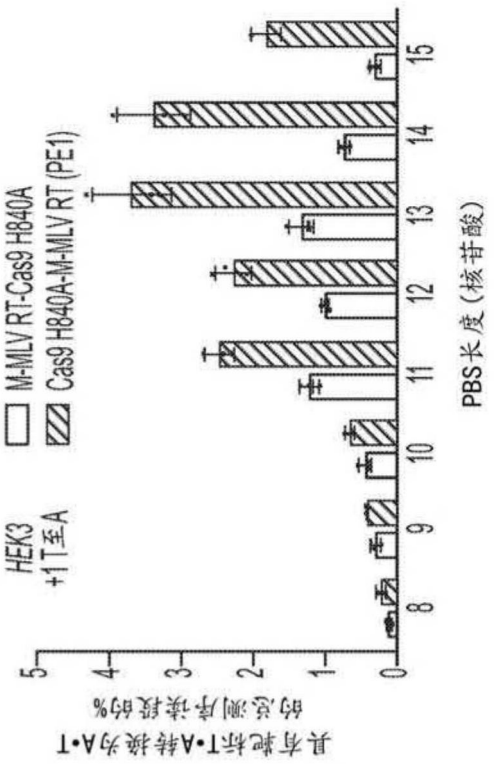

具有靶标T•A转换为A•T的总测序读段的%
HEK3 +1T至A
M-MLV RT-Cas9 H840A
Cas9 H840A-M-MLV RT (PE1)
PBS长度(核苷酸)
0
1
2
3
4
5
8
9
10
11
12
13
14
15

图39B

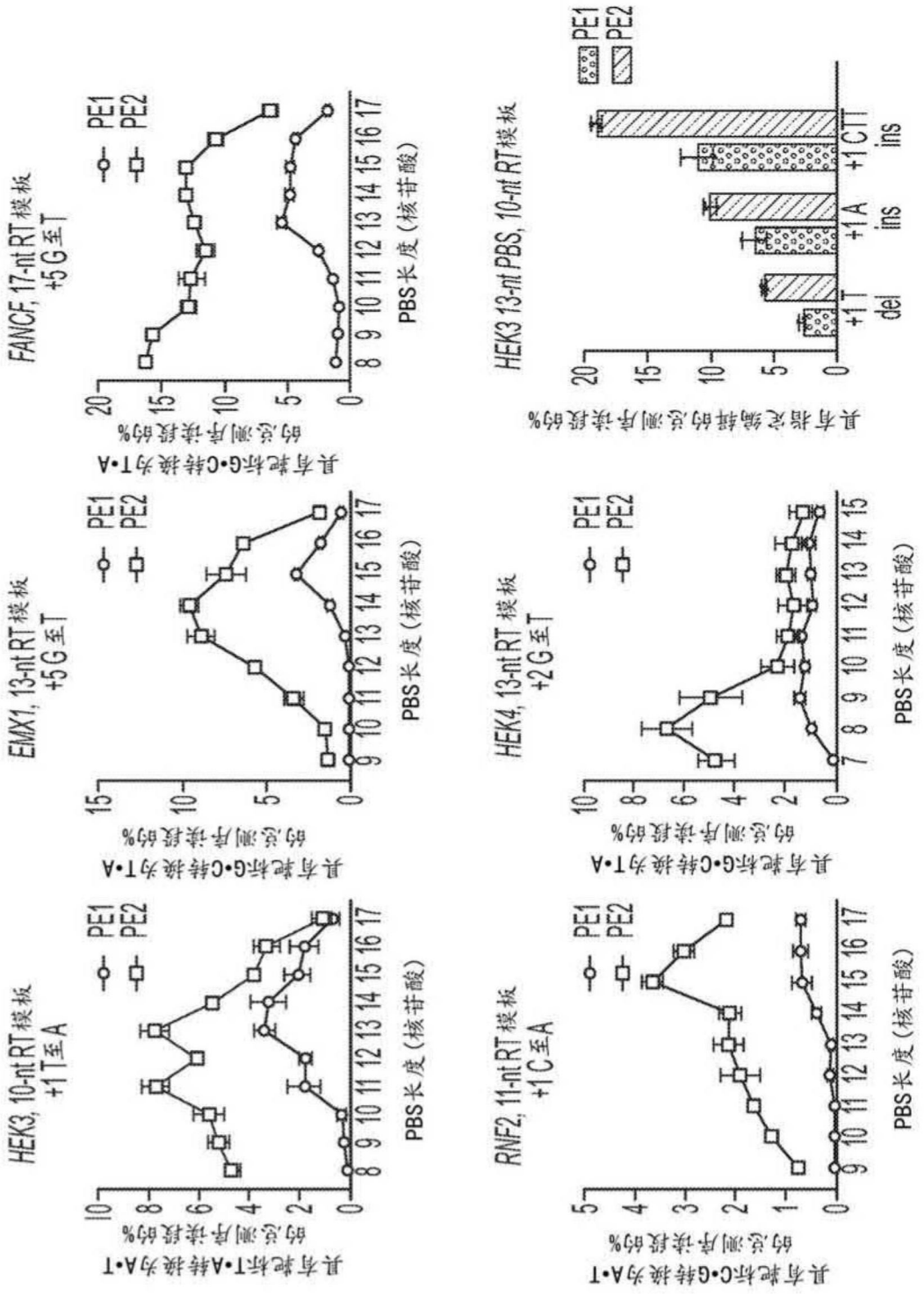

HEK3, 10-nt RT模板
+1 T至A
EMX1, 13-nt RT模板
+5 G至T
FANCF, 17-nt RT模板
+5 G至T
RNF2, 11-nt RT模板
+1 C至A
HEK4, 13-nt RT模板
+2 G至T
HEK3 13-nt PBS, 10-nt RT模板
PE1
PE2
具有靶标T•A转换为A•T的总测序读段的%
具有靶标G•C转换为T•A的总测序读段的%
具有靶标C•G转换为A•T的总测序读段的%
具有指定编辑的总测序读段的%
PBS长度（核苷酸）
del
ins
+1 T
+1 A
+1 CTT

图39C

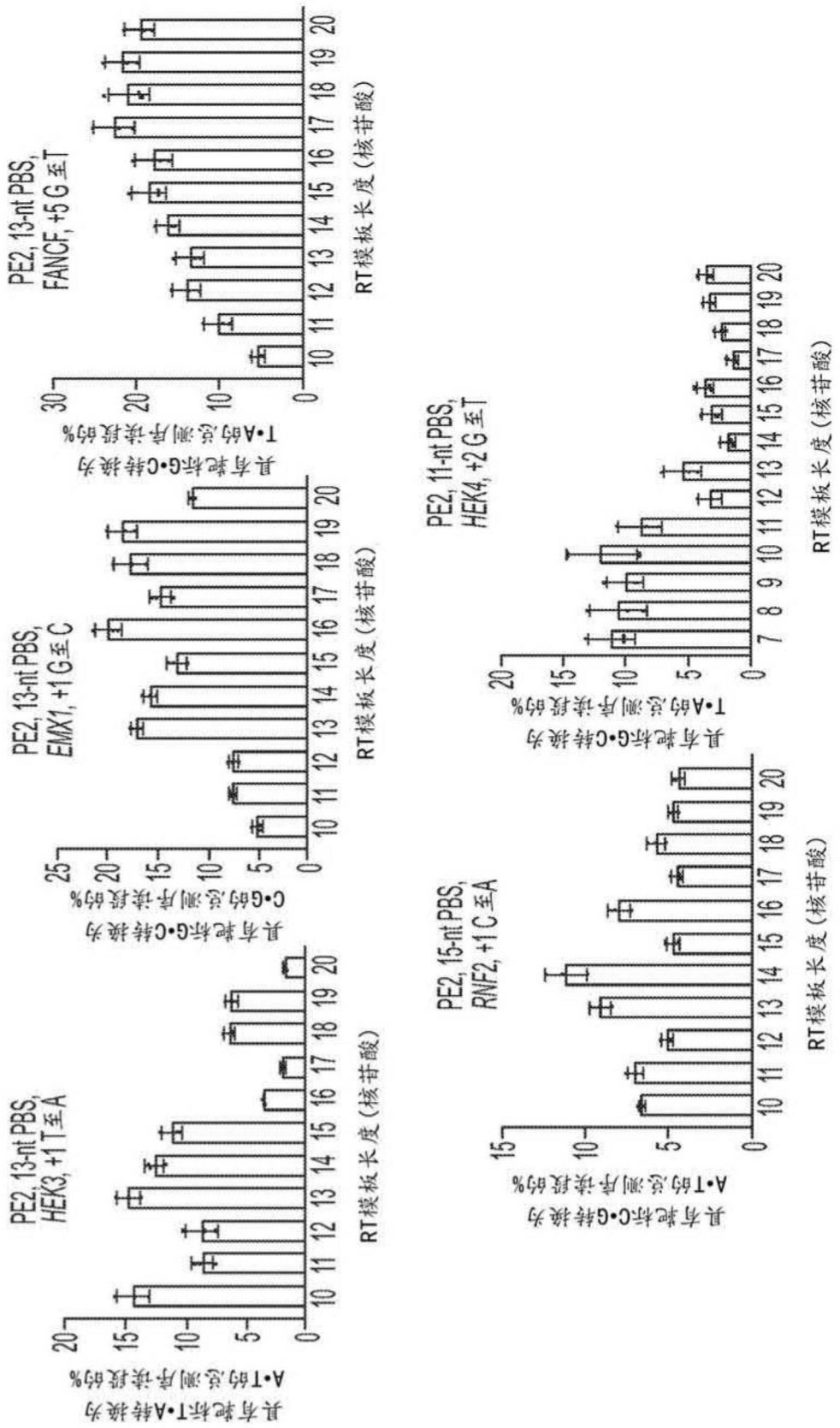

PE2, 13-nt PBS,
HEK3, +1 T至A
具有靶标T•A转换为
A•T的总测序读段的%
RT模板长度(核苷酸)
PE2, 13-nt PBS,
EMX1, +1 G至C
具有靶标G•C转换为
C•G的总测序读段的%
RT模板长度(核苷酸)
PE2, 13-nt PBS,
FANCF, +5 G至T
具有靶标G•C转换为
T•A的总测序读段的%
RT模板长度(核苷酸)
PE2, 15-nt PBS,
RNF2, +1 C至A
具有靶标C•G转换为
A•T的总测序读段的%
RT模板长度(核苷酸)
PE2, 11-nt PBS,
HEK4, +2 G至T
具有靶标G•C转换为
T•A的总测序读段的%
RT模板长度(核苷酸)

图39D

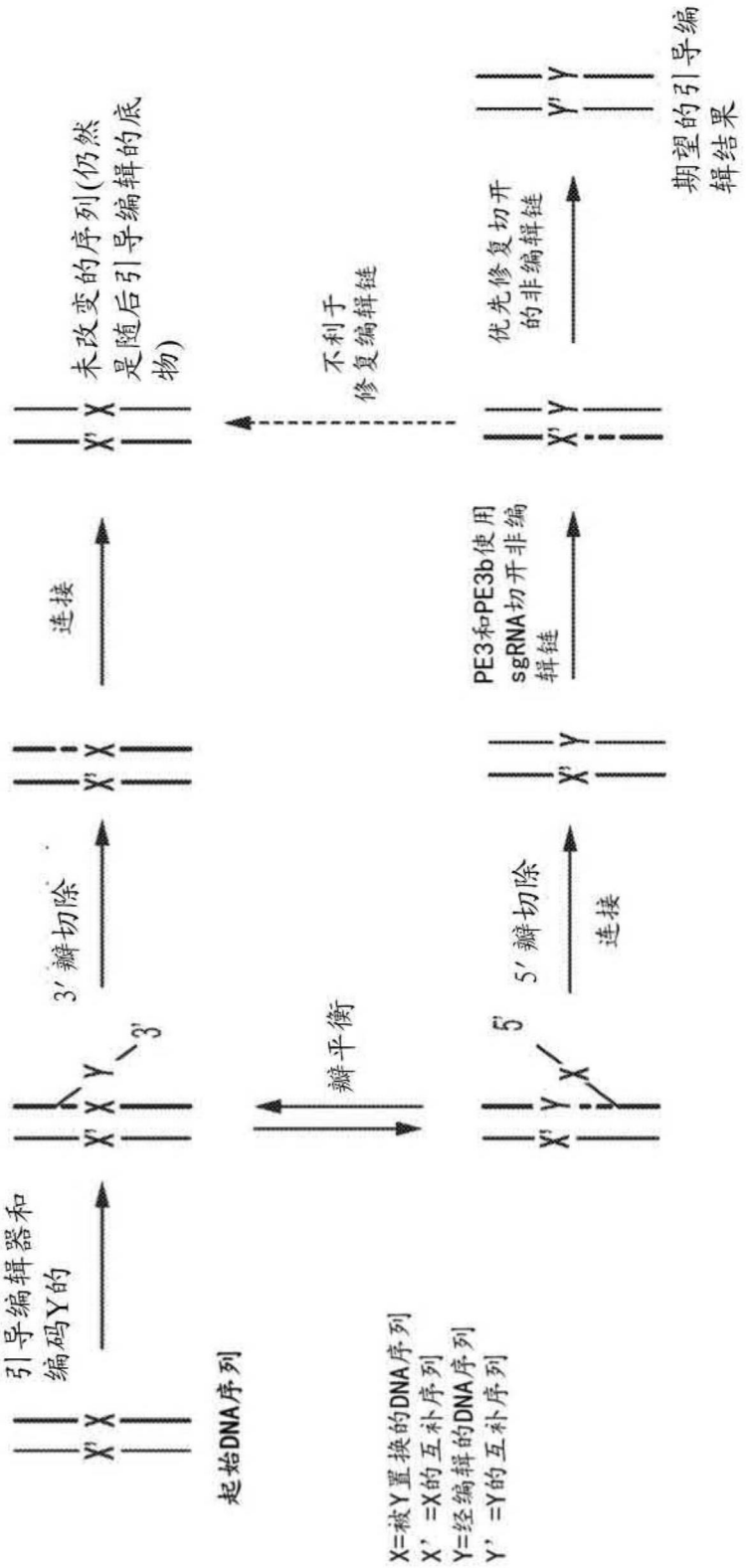

引导编辑器和
编码Y的
起始DNA序列
3′瓣切除
3′
瓣平衡
连接
未改变的序列(仍然
是随后引导编辑的底
物)
不利于
修复编辑链
PE3和PE3b使用
sgRNA切开非编
辑链
优先修复切开
的非编辑链
5′瓣切除
5′
连接
期望的引导编
辑结果
X=被Y置换的DNA序列
X′=X的互补序列
Y=经编辑的DNA序列
Y′=Y的互补序列

图40A

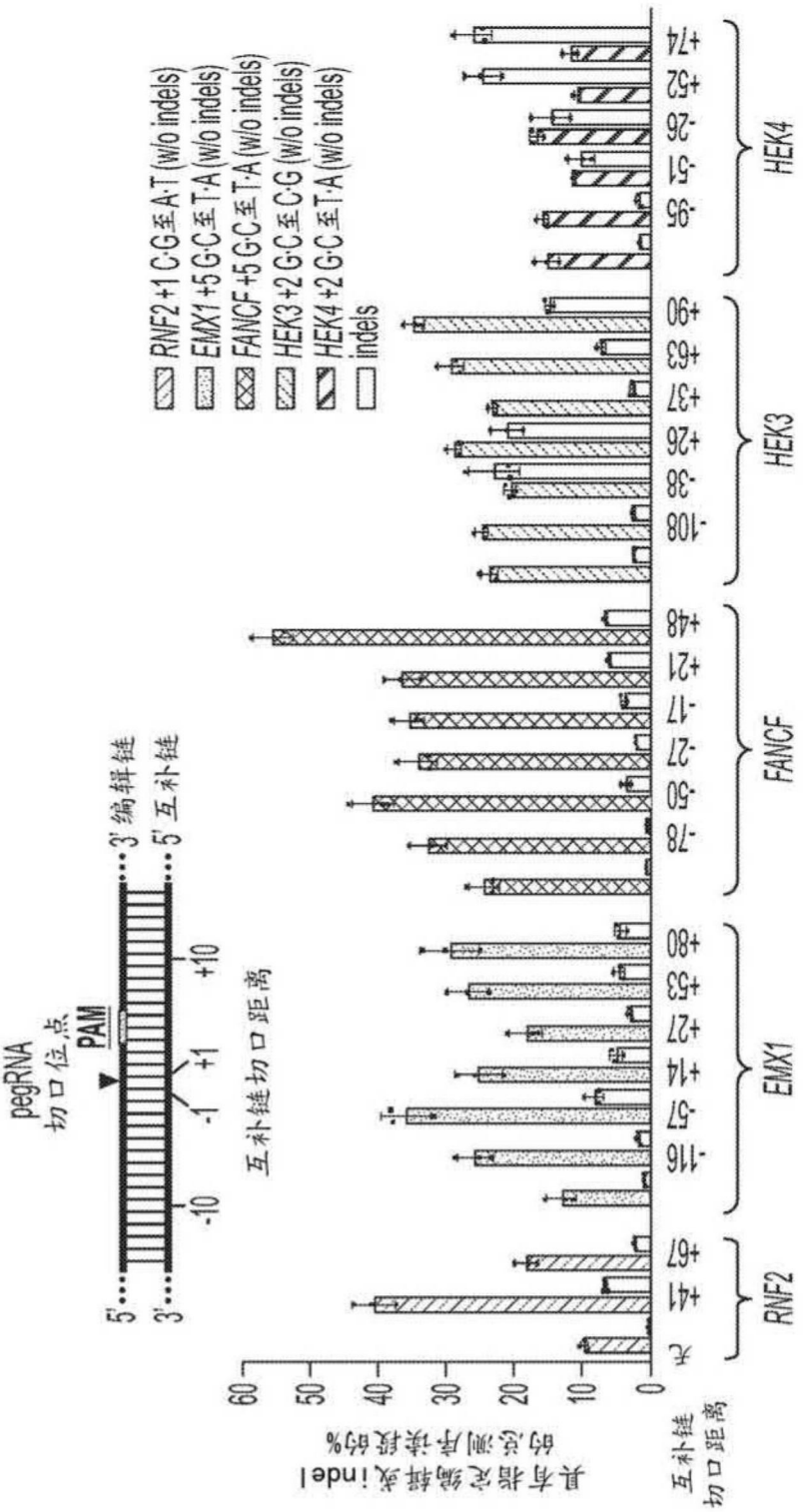

RNF2 +1 C·G至A·T (w/o indels)
EMX1 +5 G·C至T·A (w/o indels)
FANCF +5 G·C至T·A (w/o indels)
HEK3 +2 G·C至C·G (w/o indels)
HEK4 +2 G·C至T·A (w/o indels)
indels
具有指示编辑类indel的总测序读段的%
互补链切口距离
pegRNA
切口位点
PAM
3'编辑链
5'互补链
互补链切口距离
RNF2
EMX1
FANCF
HEK3
HEK4

图40B

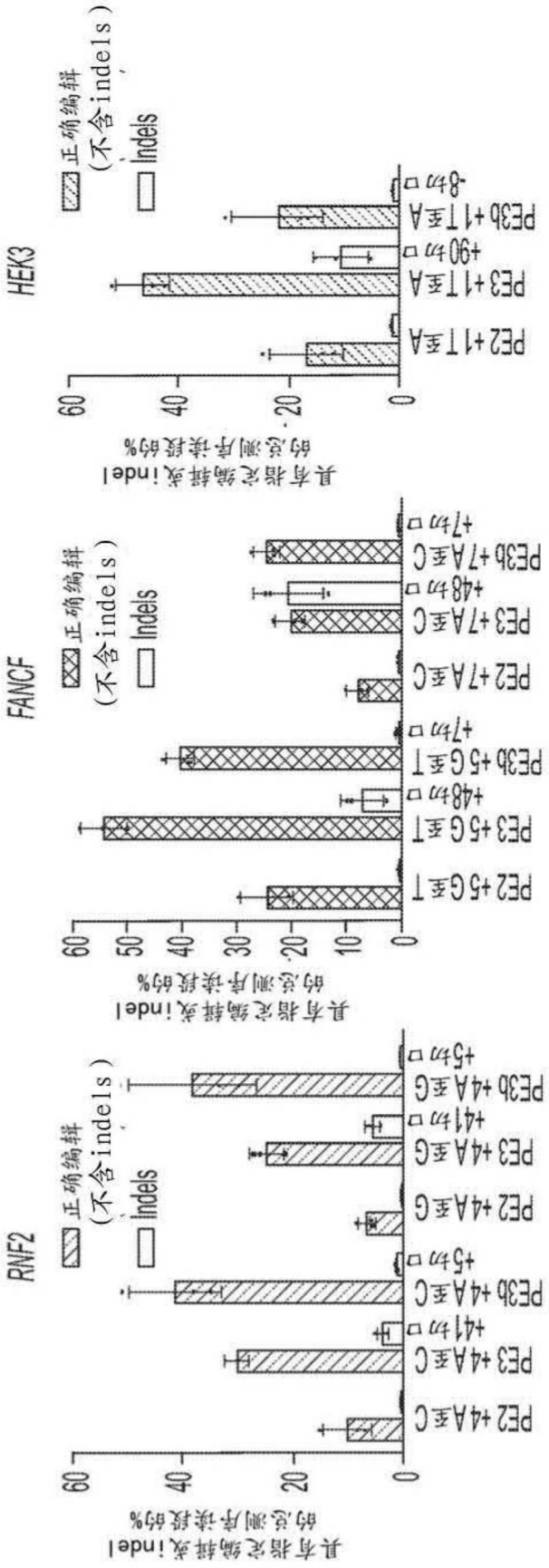

RNF2
FANCF
HEK3
正确编辑
(不含indels)
Indels
具有指定编辑类indel的总测序读段的%
0
20
40
60
0
10
20
30
40
50
60
0
20
40
60
PE2 +4A至C
PE3 +4A至C
+41bp
PE3b +4A至C
+5bp
PE2 +4A至G
PE3 +4A至G
+41bp
PE3b +4A至G
+5bp
PE2 +5G至T
PE3 +5G至T
+48bp
PE3b +5G至T
+7bp
PE2 +7A至C
PE3 +7A至C
+48bp
PE3b +7A至C
+7bp
PE2 +11T至A
PE3 +11T至A
+90bp
PE3b +11T至A
-8bp

图40C

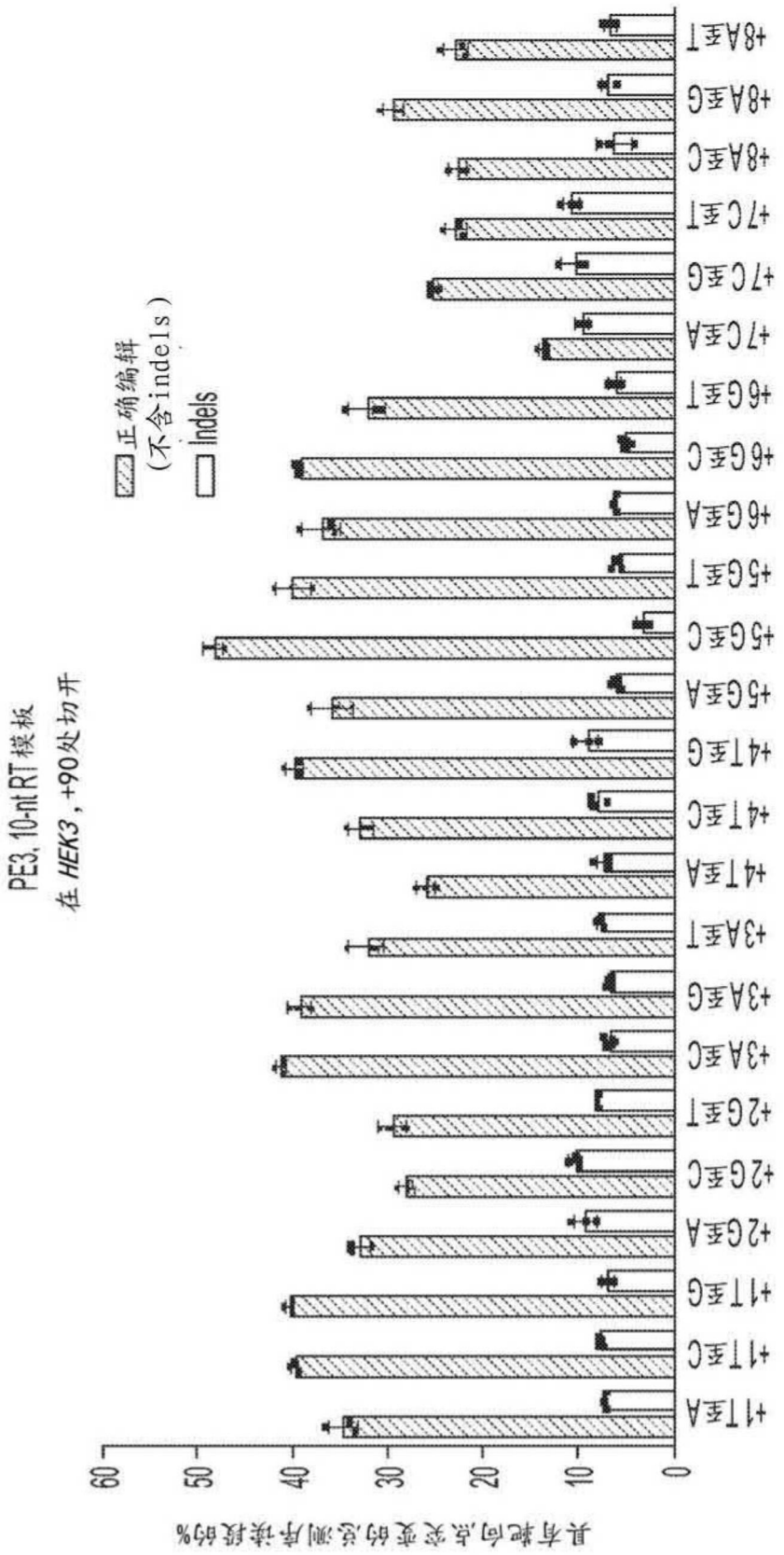

PE3, 10-nt RT模板
在 HEK3，+90处切开
正确编辑
(不含indels)
Indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
60
+1T至A
+1T至C
+1T至G
+2G至A
+2G至C
+2G至T
+3A至C
+3A至G
+3A至T
+4T至A
+4T至C
+4T至G
+5G至A
+5G至C
+5G至T
+6G至A
+6G至C
+6G至T
+7C至A
+7C至G
+7C至T
+8A至C
+8A至G
+8A至T

图41A

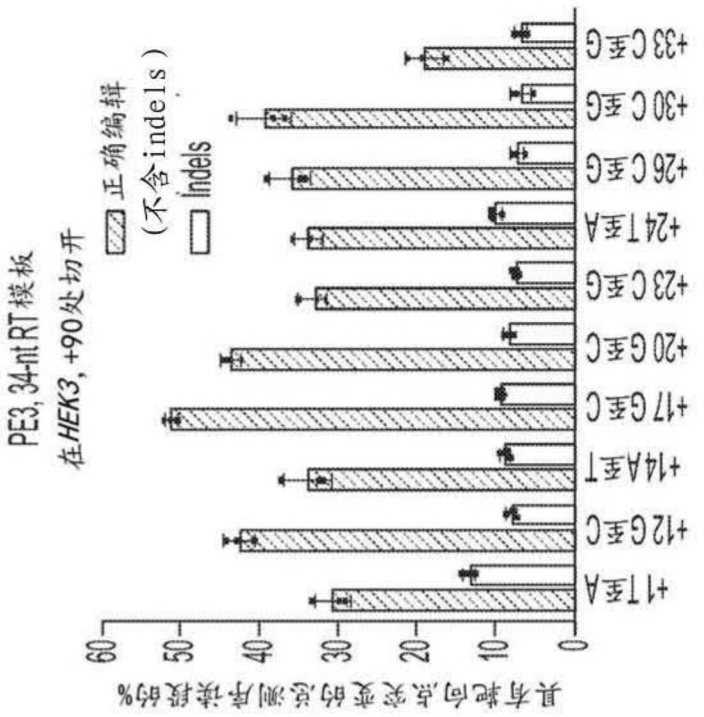

PE3, 34-nt RT模板
在HEK3, +90处切开
正确编辑
(不含indels)
Indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
60
+11T至A
+12G至C
+14A至T
+17G至C
+20G至C
+23C至G
+24T至A
+26C至G
+30C至G
+33C至G

图41B

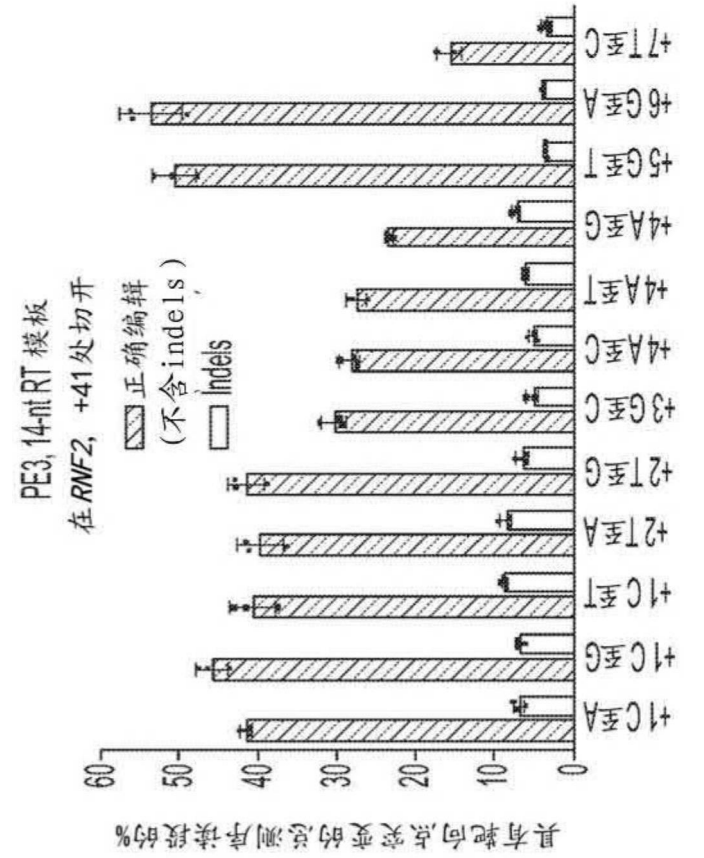

PE3, 14-nt RT 模板
在RNF2, +41处切开
正确编辑
(不含indels)
Indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
60
+1C至A
+1C至G
+1C至T
+2T至A
+2T至G
+3G至C
+4A至C
+4A至T
+4A至G
+5G至T
+6G至A
+7T至C

图41C

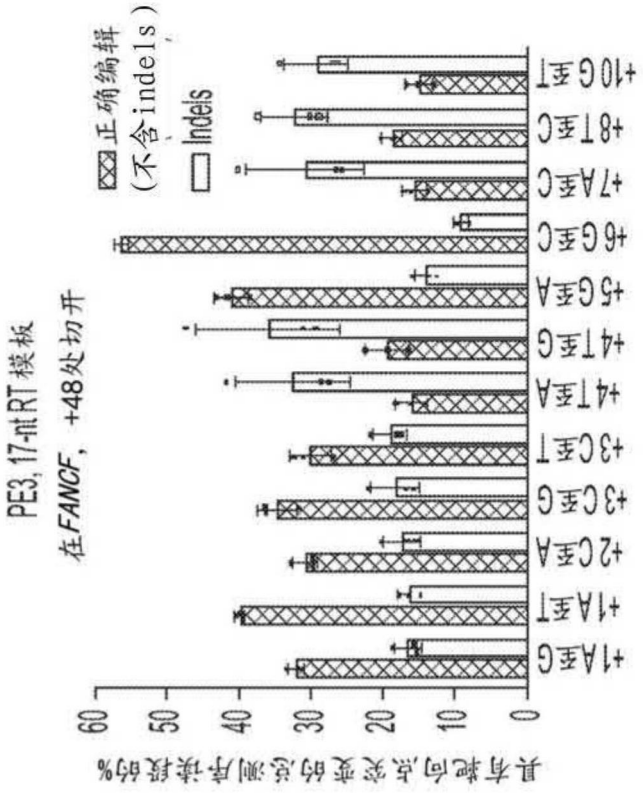

PE3, 17-nt RT模板
在FANCF，+48处切开
正确编辑
(不含indels)
Indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
60
+1A至G
+1A至T
+2C至A
+3C至G
+3C至T
+4T至A
+4T至G
+5G至A
+6G至C
+7A至C
+8T至C
+10G至T

图41D

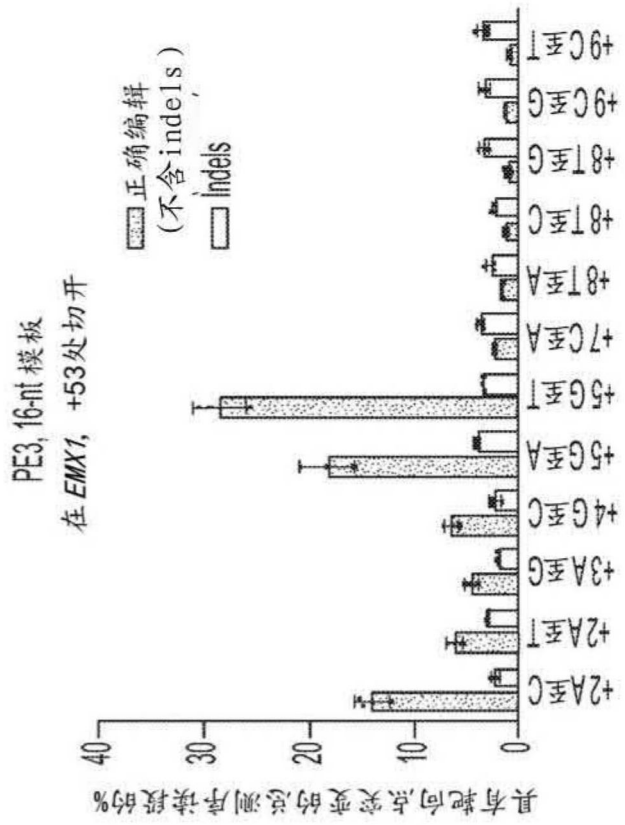

PE3, 16-nt 模板
在 EMX1, +53处切开
正确编辑
(不含indels)
indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
+2A至C
+2A至T
+3A至G
+4G至C
+5G至A
+5G至T
+7C至A
+8T至A
+8T至C
+8T至G
+9C至G
+9C至T

图41E

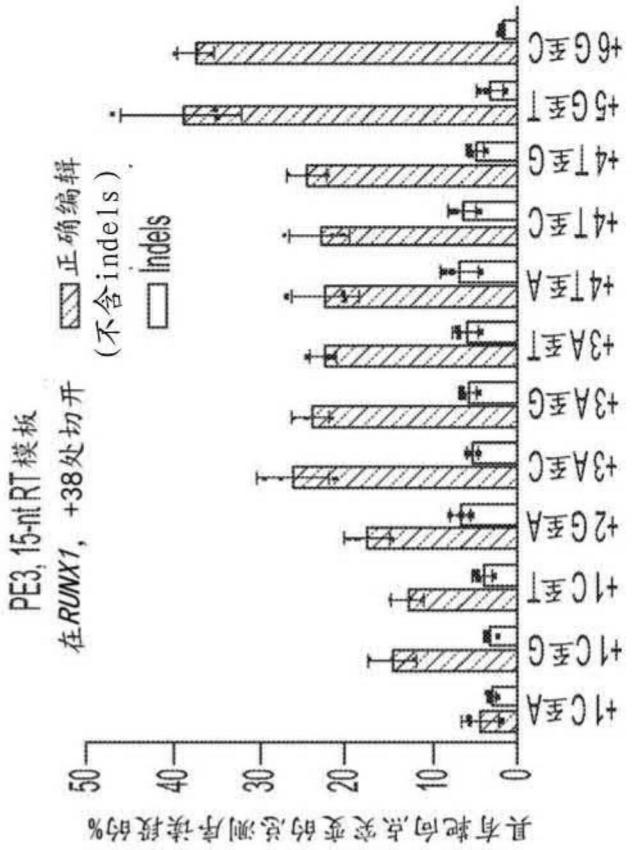

PE3, 15-nt RT模板
在RUNX1，+38处切开
正确编辑
(不含indels)
Indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
+1C至A
+1C至G
+1C至T
+2G至A
+3A至C
+3A至G
+3A至T
+4T至A
+4T至C
+4T至G
+5G至T
+6G至C

图41F

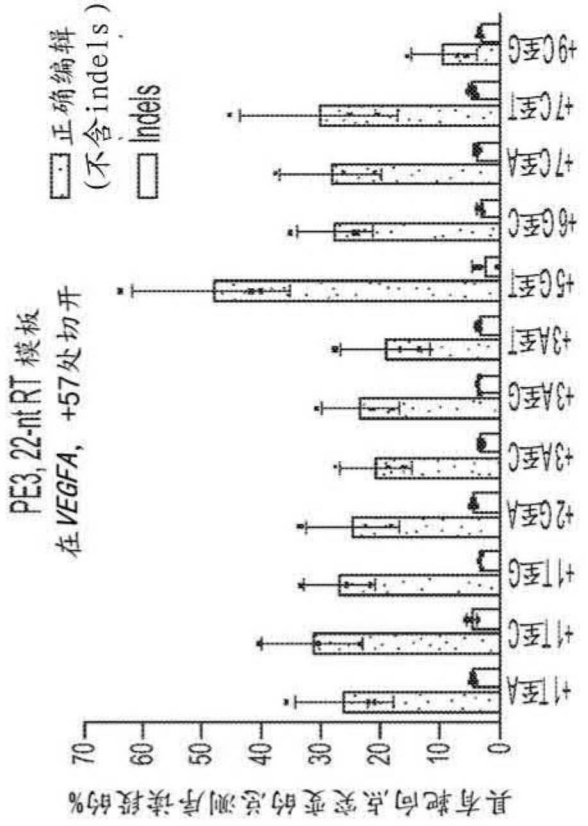

PE3, 22-nt RT 模板
在 VEGFA，+57 处切开
正确编辑
(不含indels)
indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
60
70
+1 T至A
+1 T至C
+1 T至G
+2 G至A
+3 A至C
+3 A至G
+3 A至T
+5 G至T
+6 G至C
+7 C至A
+7 C至T
+9 C至G

图41G

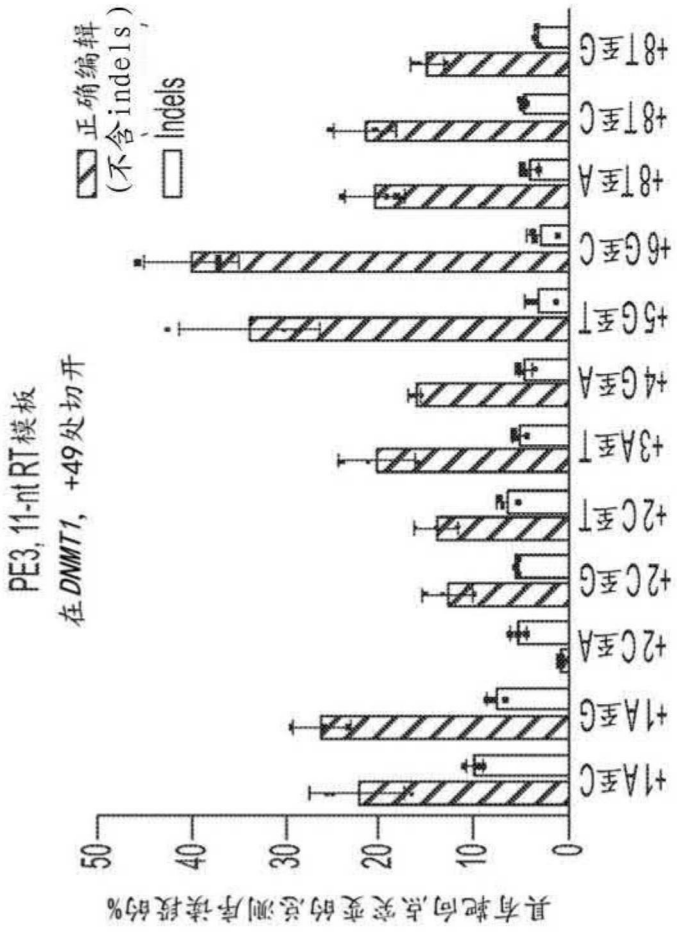

PE3, 11-nt RT模板
在 DNMT1, +49处切开
正确编辑
(不含indels)
indels
具有靶向点突变的总测序读段的%
0
10
20
30
40
50
+1A至C
+1A至G
+2C至A
+2C至G
+2C至T
+3A至T
+4G至A
+5G至T
+6G至C
+8T至A
+8T至C
+8T至G

图41H

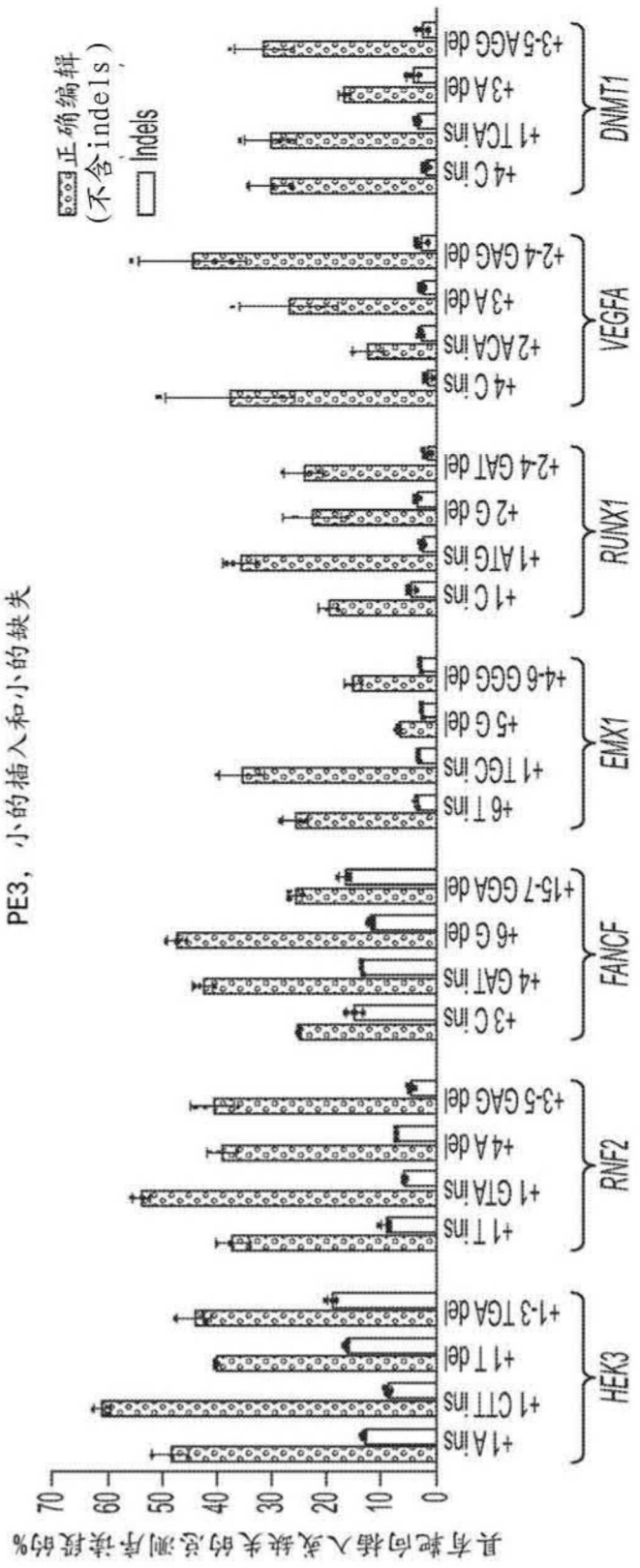

具有靶向插入或缺失的总测序读段的%
70
60
50
40
30
20
10
0
PE3，小的插入和小的缺失
正确编辑
(不含indels)
indels
+1 A ins
+1 CTT ins
+1 T del
+1-3 TGA del
HEK3
+1 T ins
+1 GTA ins
+4 A del
+3-5 GAG del
RNF2
+3 C ins
+4 GAT ins
+6 G del
+15-7 GGA del
FANCF
+6 T ins
+1 TGC ins
+5 G del
+4-6 GGG del
EMX1
+1 C ins
+1 ATG ins
+2 G del
+2-4 GAT del
RUNX1
+4 C ins
+2 ACA ins
+3 A del
+2-4 GAG del
VEGFA
+4 C ins
+1 TCA ins
+3 A del
+3-5 AGG del
DNMT1

图41I

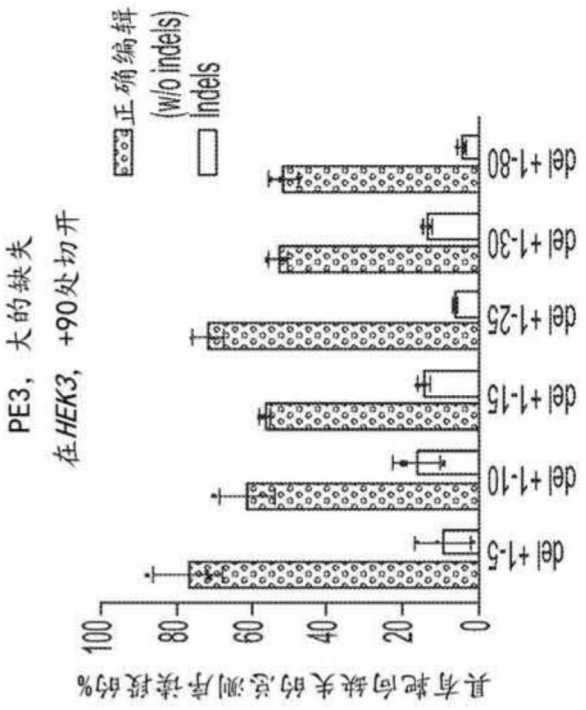

PE3，大的缺失
在HEK3，+90处切开
正确编辑
(w/o indels)
indels
具有靶向缺失的测序读段的%
100
80
60
40
20
0
del +1-5
del +1-10
del +1-15
del +1-25
del +1-30
del +1-80

图41J

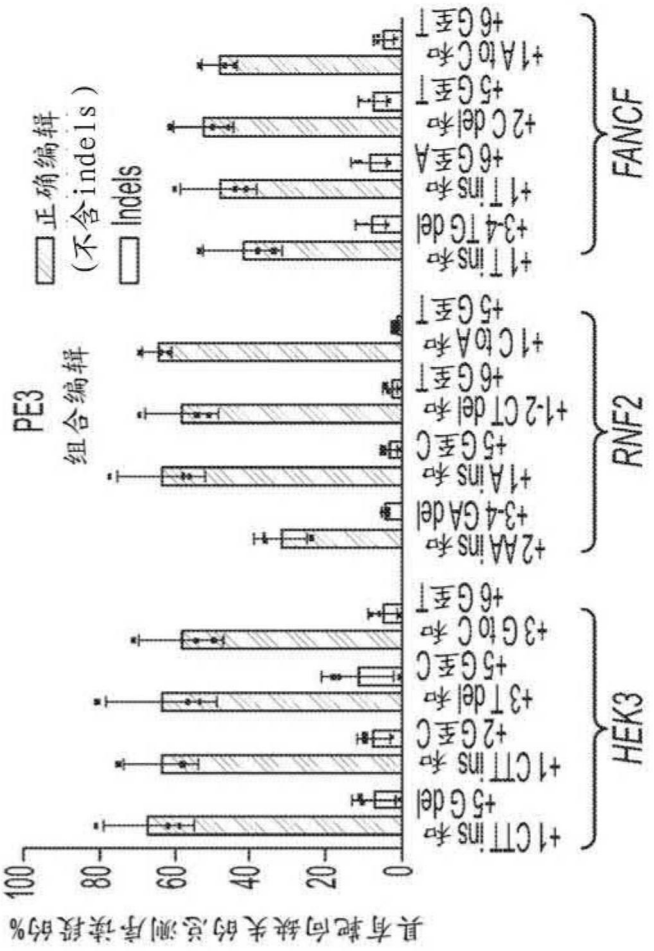

具有靶向缺失的总测序读段的%
100
80
60
40
20
0
PE3
组合编辑
正确编辑
(不含indels)
Indels
+1 CTT ins 和
+5 G del
+1 CTT ins 和
+2 G至C
+3 T del 和
+5 G至C
+3 G to C 和
+6 G至T
+2AA ins 和
+3-4 GA del
+1A ins 和
+5 G至C
+1-2 CT del 和
+6 G至T
+1C to A 和
+5 G至T
+1T ins 和
+3-4 TG del
+1T ins 和
+6 G至A
+2 C del 和
+5 G至T
+1A to C 和
+6 G至T
HEK3
RNF2
FANCF

图41K

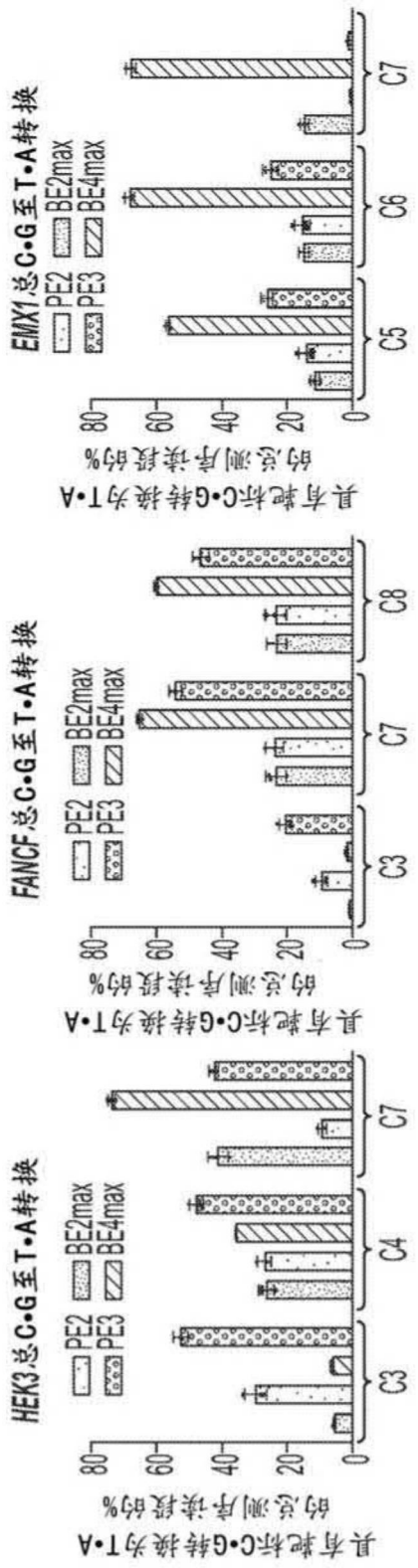

HEK3,总C•G至T•A转换
EMX1,总C•G至T•A转换
FANCF,总C•G至T•A转换
具有靶标C•G转换为T•A的总测序读段的%
PE2
PE3
BE2max
BE4max
C3
C4
C5
C6
C7
C8
0
20
40
60
80

图42A

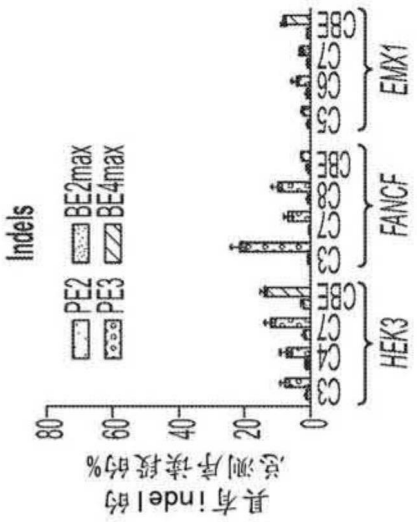

具有indel的
总测序读段的%
0
20
40
60
80
Indels
PE2
PE3
BE2max
BE4max
HEK3
FANCF
EMX1

图42B

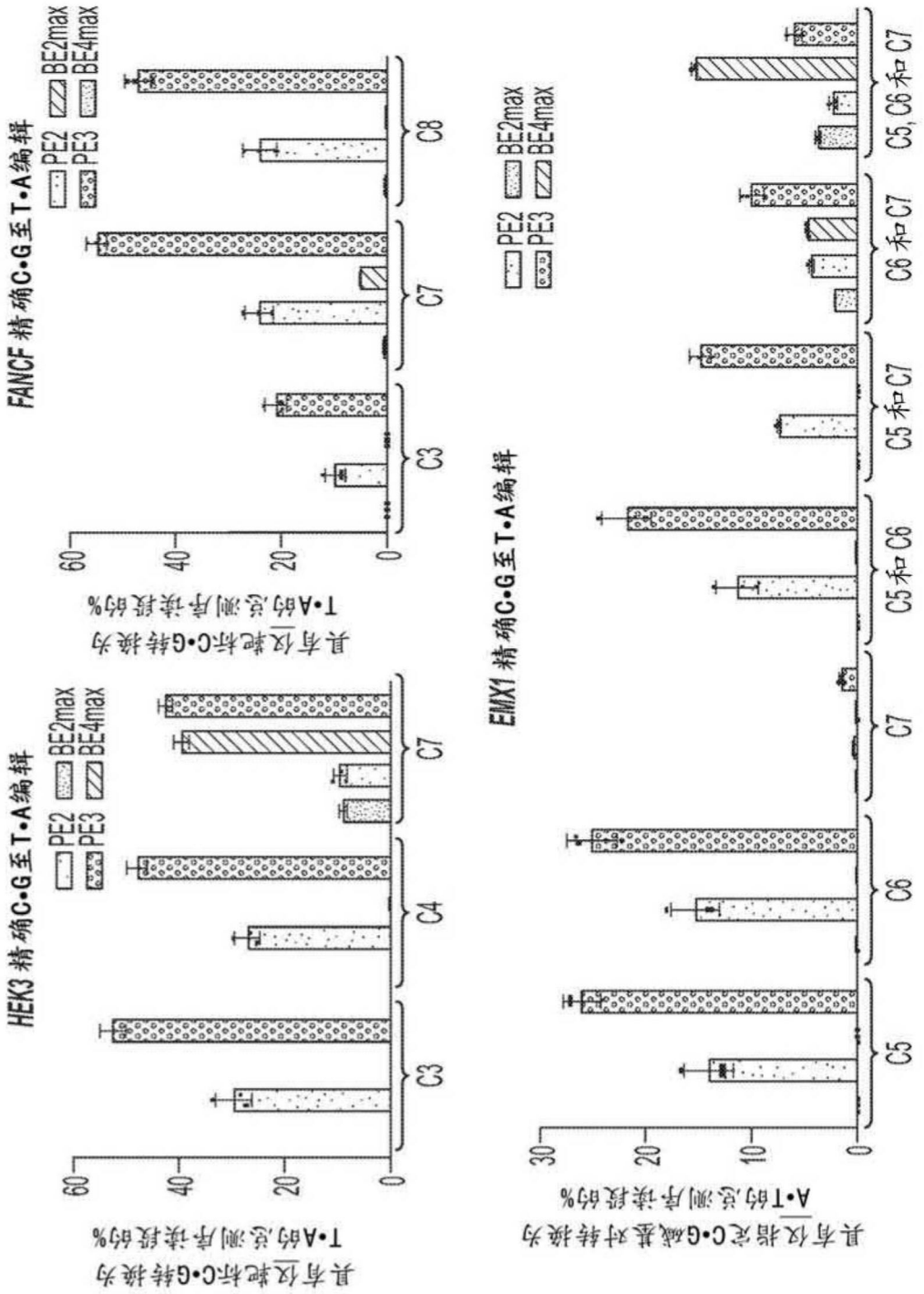

HEK3 精确C•G至T•A编辑
PE2
PE3
BE2max
BE4max
具有仅靶标C•G转换为T•A的总测序读段的%
C3
C4
C7
FANCF 精确C•G至T•A编辑
C8
EMX1 精确C•G至T•A编辑
具有仅指定C•G碱基对转换为A•T的总测序读段的%
C5
C6
C5和C6
C5和C7
C6和C7
C5, C6和C7

图42C

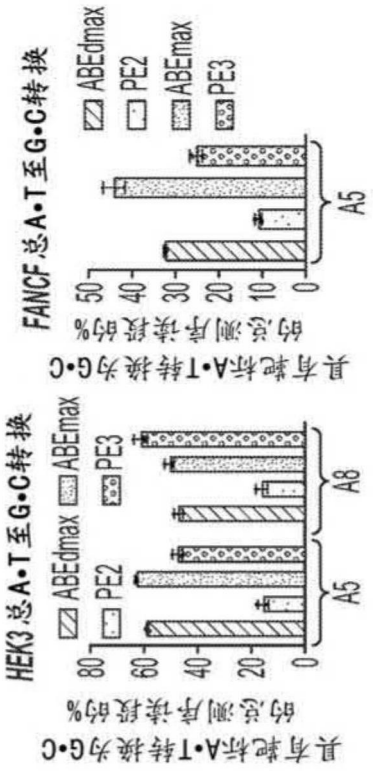

HEK3 总A•T至G•C转换
ABEdmax
ABEmax
PE2
PE3
具有靶标A•T转换为G•C的总测序读段的%
0
20
40
60
80
A5
A8
FANCF 总A•T至G•C转换
ABEdmax
PE2
ABEmax
PE3
具有靶标A•T转换为G•C的总测序读段的%
0
10
20
30
40
50
A5

图42D

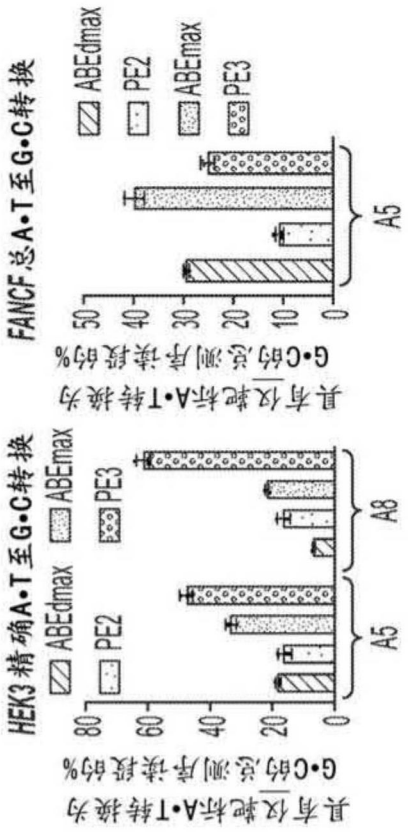

具有仅靶标A•T转换为G•C的总测序读段的%
HEK3 精确A•T至G•C转换
ABEdmax
PE2
ABEmax
PE3
A5
A8
具有仅靶标A•T转换为G•C的总测序读段的%
FANCF 总A•T至G•C转换
ABEdmax
PE2
ABEmax
PE3
A5

图42E

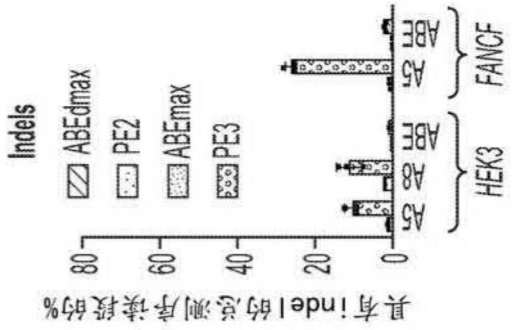

具有indel的总测序读段的%
Indels
ABEdmax
PE2
ABEmax
PE3
A5
ABE8
ABE7
A5
ABE8
ABE7
HEK3
FANCF

图42F

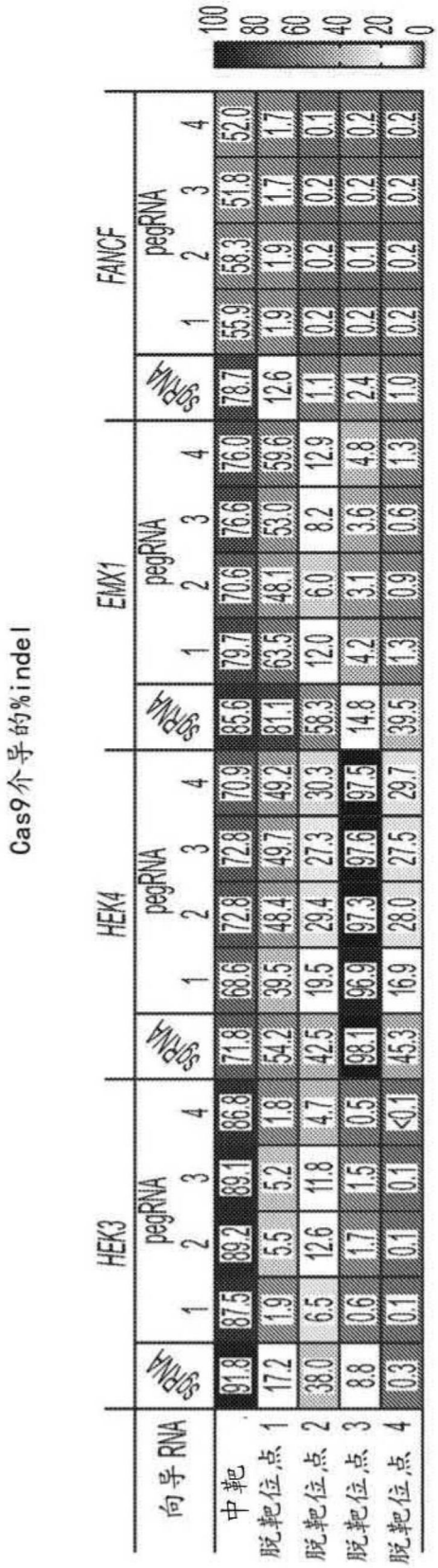

Cas9介导的%indel
向导 RNA
HEK3
HEK4
EMX1
FANCF
sgRNA
pegRNA
1
2
3
4
中靶
脱靶位点 1
脱靶位点 2
脱靶位点 3
脱靶位点 4
91.8 87.5 89.2 89.1 86.8 71.8 68.6 72.8 72.8 70.9 85.6 79.7 70.6 76.6 76.0 78.7 55.9 58.3 51.8 52.0
17.2 1.9 5.5 5.2 1.8 54.2 39.5 48.4 49.7 49.2 81.1 63.5 48.1 53.0 59.6 12.6 1.9 1.9 1.7 1.7
38.0 6.5 12.6 11.8 4.7 42.5 19.5 29.4 27.3 30.3 58.3 12.0 6.0 8.2 12.9 1.1 0.2 0.2 0.2 0.1
8.8 0.6 1.7 1.5 0.5 98.1 96.9 97.3 97.6 97.5 14.8 4.2 3.1 3.6 4.8 2.4 0.2 0.1 0.2 0.2
0.3 0.1 0.1 0.1 <0.1 45.3 16.9 28.0 27.5 29.7 39.5 1.3 0.9 0.6 1.3 1.0 0.2 0.2 0.2 0.2
100
80
60
40
20
0

图42G

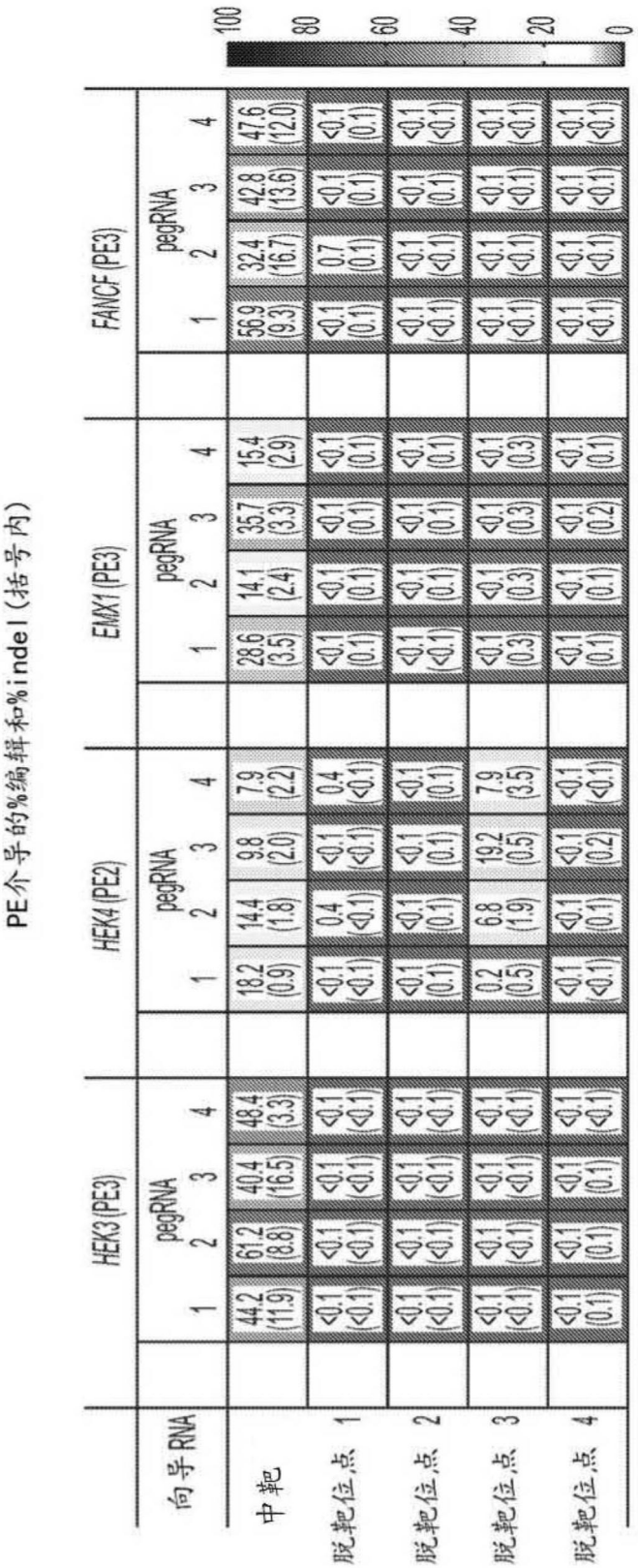

PE个导的%编辑和%indel(括号内)
100
80
60
40
20
0
向导RNA
中靶
脱靶位点 1
脱靶位点 2
脱靶位点 3
脱靶位点 4
HEK3(PE3)
HEK4(PE2)
EMX1(PE3)
FANCF(PE3)
pegRNA
1 2 3 4
44.2 (11.9) 61.2 (8.8) 40.4 (16.5) 48.4 (3.3)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1)
18.2 (0.9) 14.4 (1.8) 9.8 (2.0) 7.9 (2.2)
<0.1 (<0.1) 0.4 (<0.1) <0.1 (<0.1) 0.4 (<0.1)
<0.1 (0.1) 0.4 (<0.1) <0.1 (0.1) <0.1 (0.1)
0.2 (0.5) 6.8 (1.9) 19.2 (0.5) 7.9 (3.5)
<0.1 (<0.1) <0.1 (0.1) <0.1 (0.2) <0.1 (<0.1)
28.6 (3.5) 14.1 (2.4) 35.7 (3.3) 15.4 (2.9)
<0.1 (0.1) <0.1 (0.1) <0.1 (0.1) <0.1 (0.1)
<0.1 (<0.1) <0.1 (0.1) <0.1 (0.1) <0.1 (0.1)
<0.1 (0.3) <0.1 (0.3) <0.1 (0.3) <0.1 (0.3)
<0.1 (0.1) <0.1 (0.1) <0.1 (0.2) <0.1 (0.1)
56.9 (9.3) 32.4 (16.7) 42.8 (13.6) 47.6 (12.0)
<0.1 (0.1) 0.7 (0.1) <0.1 (0.1) <0.1 (0.1)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (0.1) <0.1 (<0.1)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1)
<0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1) <0.1 (<0.1)

图42H

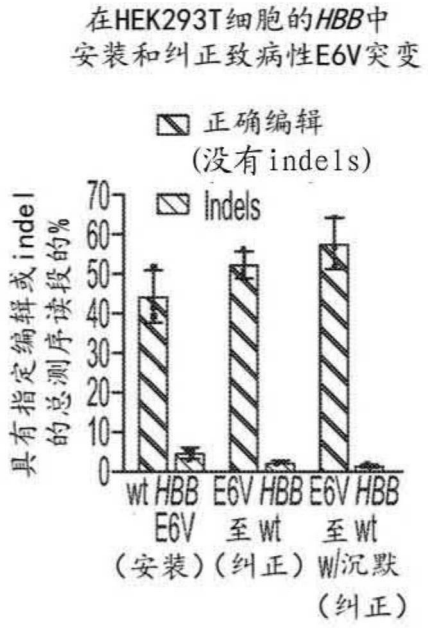

在HEK293T细胞的HBB中
安装和纠正致病性E6V突变
正确编辑
(没有indels)
Indels
具有指定编辑或indel
的总测序读段的%
70
60
50
40
30
20
10
0
wt HBB E6V
E6V 至wt
HBB E6V 至wt
(安装)(纠正)
w/沉默
(纠正)

图43A

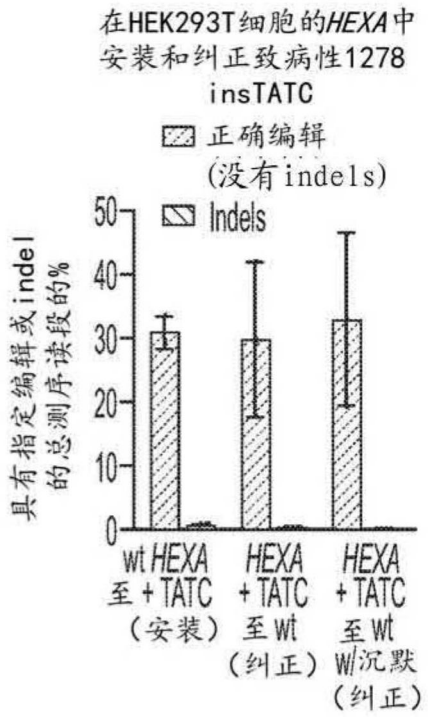

在HEK293T细胞的HEXA中
安装和纠正致病性1278
insTATC
正确编辑
(没有indels)
Indels
具有指定编辑或indel
的总测序读段的%
50
40
30
20
10
0
wt HEXA
至+TATC
(安装)
HEXA
+TATC
至wt
(纠正)
HEXA
+TATC
至wt
w/沉默
(纠正)

图43B

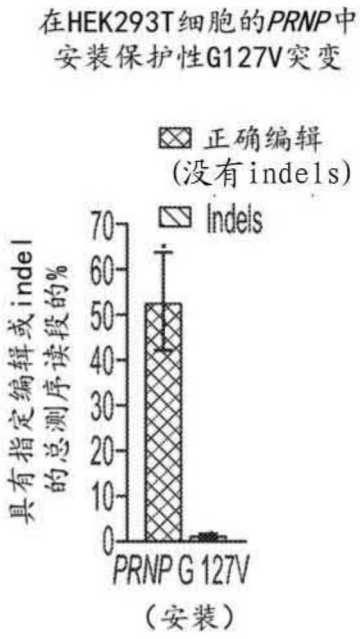

在HEK293T细胞的PRNP中
安装保护性G127V突变
正确编辑
(没有indels)
Indels
具有指定编辑或indel
的总测序读段的%
70
60
50
40
30
20
10
0
PRNP G 127V
(安装)

图43C

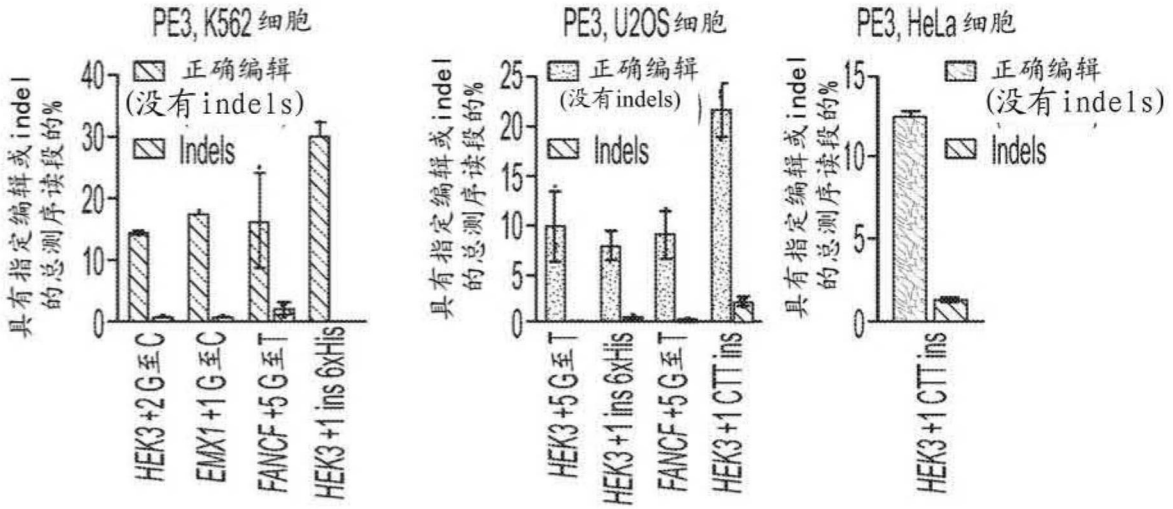

PE3, K562细胞
PE3, U2OS细胞
PE3, HeLa细胞
具有指定编辑或indel
的总测序读段的%
正确编辑
(没有indels)
Indels
HEK3 +2 G至C
EMX1 +1 G至C
FANCF +5 G至T
HEK3 +1 ins 6xHis
HEK3 +5 G至T
HEK3 +1 ins 6xHis
FANCF +5 G至T
HEK3 +1 CTT ins
HEK3 +1 CTT ins

图43D

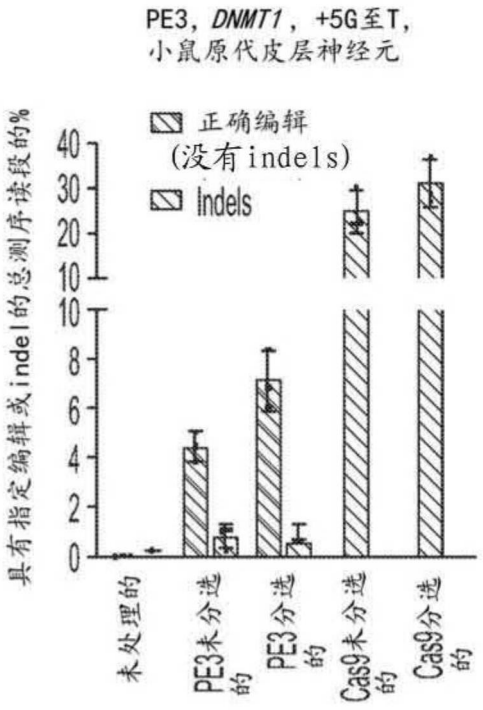

PE3，DNMT1，+5G至T，
小鼠原代皮层神经元
正确编辑
(没有indels)
indels
具有指定编辑或indel的总测序读段的%
0
2
4
6
8
10
20
30
40
未处理的
PE3未分选的
PE3分选的
Cas9未分选的
Cas9分选的

图43E

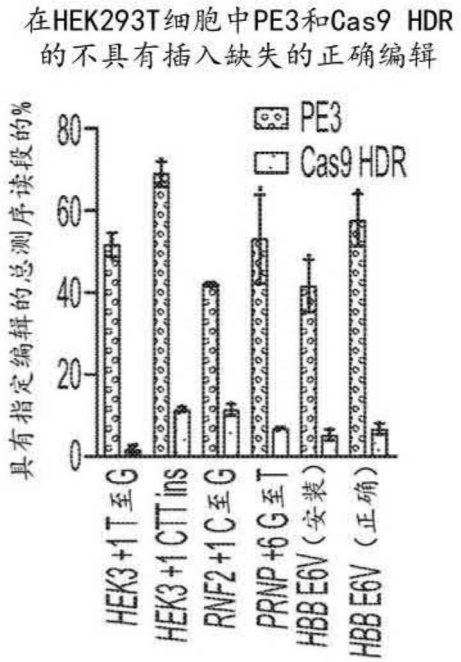

在HEK293T细胞中PE3和Cas9 HDR
的不具有插入缺失的正确编辑
PE3
Cas9 HDR
具有指定编辑的总测序读段的%
0
20
40
60
80
HEK3 +1 T至G
HEK3 +1 CTT ins
RNF2 +1 C至G
PRNP +6 G至T
HBB E6V（安装）
HBB E6V（正确）

图43F

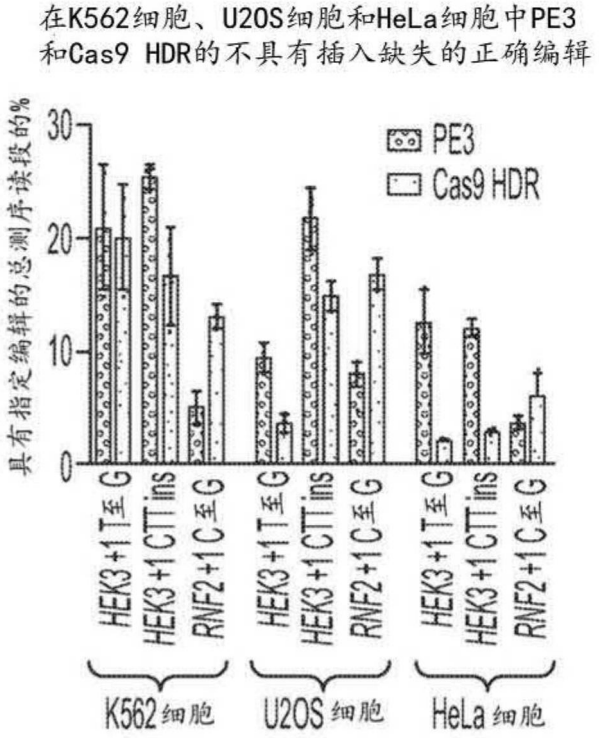

在K562细胞、U2OS细胞和HeLa细胞中PE3
和Cas9 HDR的不具有插入缺失的正确编辑
具有指定编辑的总测序读段的%
30
20
10
0
PE3
Cas9 HDR
HEK3 +1 T至G
HEK3 +1 CTT ins
RNF2 +1 C至G
K562细胞
U2OS细胞
HeLa细胞

图43G

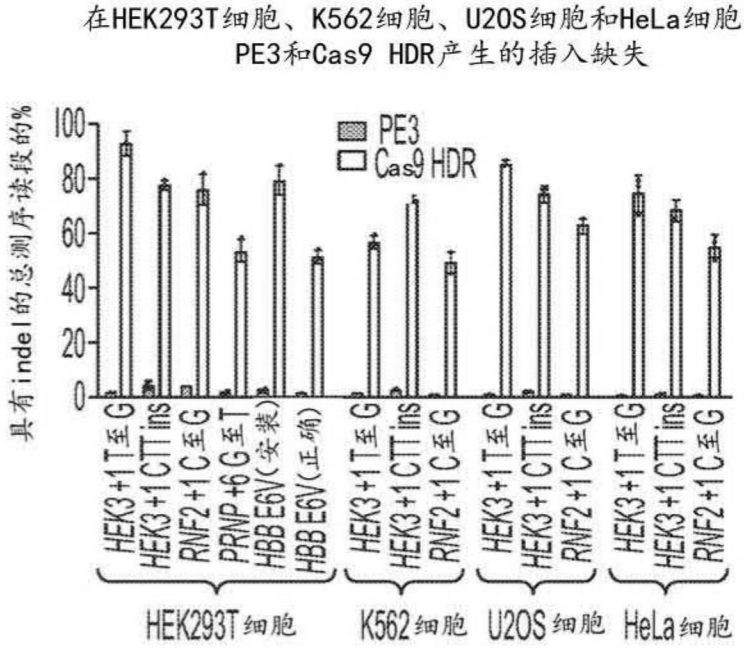

在HEK293T细胞、K562细胞、U2OS细胞和HeLa细胞中
PE3和Cas9 HDR产生的插入缺失
具有indel的总测序读段的%
100
80
60
40
20
0
PE3
Cas9 HDR
HEK3 +1 T至G
HEK3 +1 CTT ins
RNF2 +1 C至G
PRNP +6 G至T
HBB E6V(安装)
HBB E6V(正确)
HEK293T细胞
K562细胞
U2OS细胞
HeLa细胞

图43H

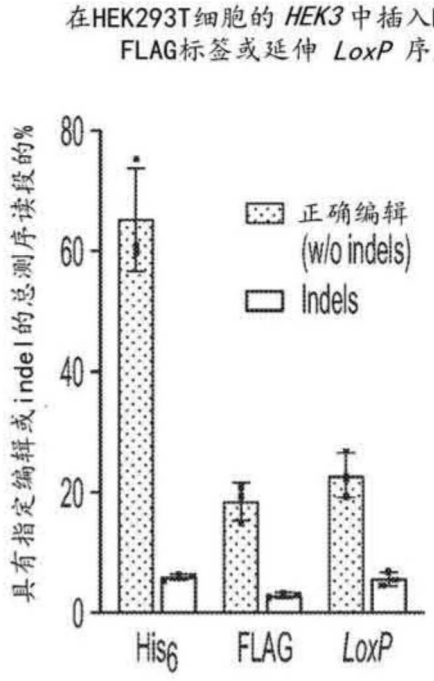

在HEK293T细胞的 HEK3 中插入His6标签、
FLAG标签或延伸 LoxP 序列
具有指定编辑或indel的总测序读段的%
80
60
40
20
0
正确编辑
(w/o indels)
Indels
His6
FLAG
LoxP

图43I

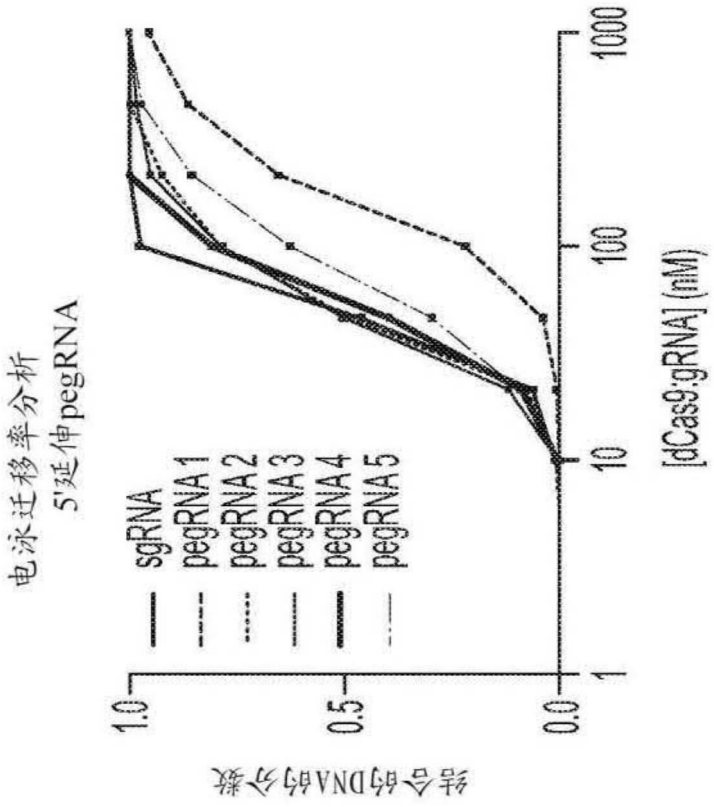

电泳迁移率分析
5'延伸pegRNA
sgRNA
pegRNA 1
pegRNA 2
pegRNA 3
pegRNA 4
pegRNA 5
1.0
0.5
0.0
1000
100
10
1
[dCas9:gRNA] (nM)
结合的DNA的分数

图44A

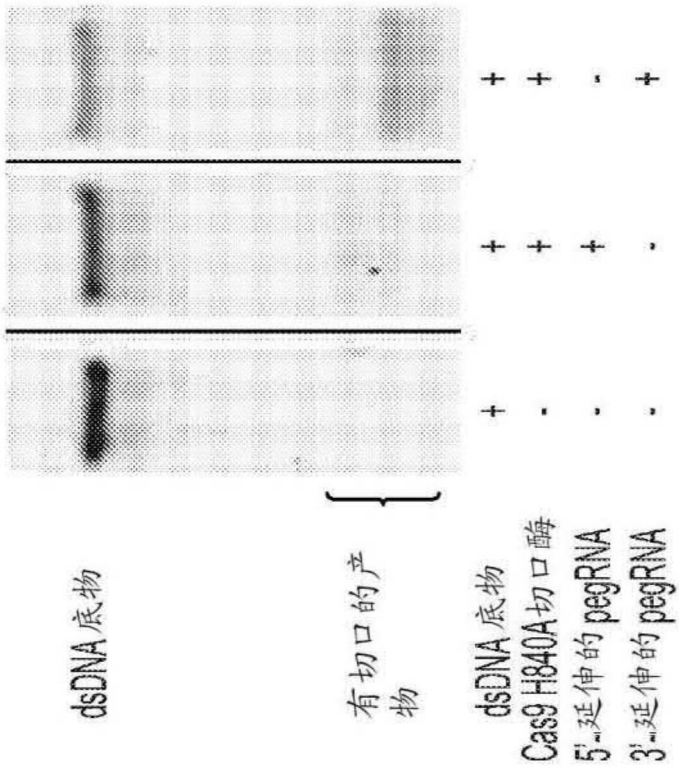

dsDNA 底物
有切口的产物
dsDNA 底物
Cas9 H840A切口酶
5'-延伸的 pegRNA
3'-延伸的 pegRNA
+ - - +
+ + + -
+ + - +

图44B

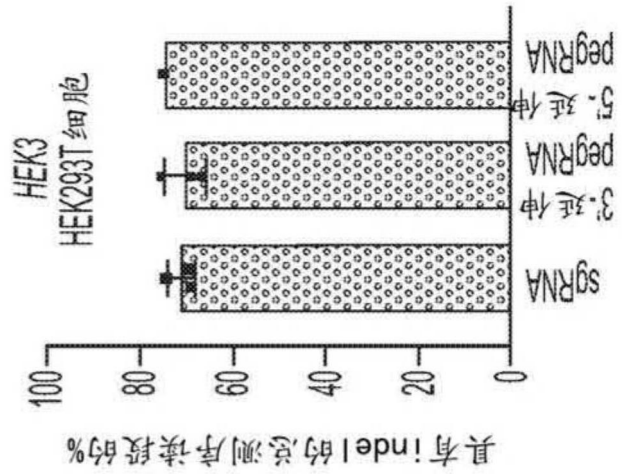

具有 indel 的经测序读段的%
100
80
60
40
20
0
HEK3
HEK293T 细胞
sgRNA
3'-延伸
pegRNA
5'-延伸
pegRNA

图44C

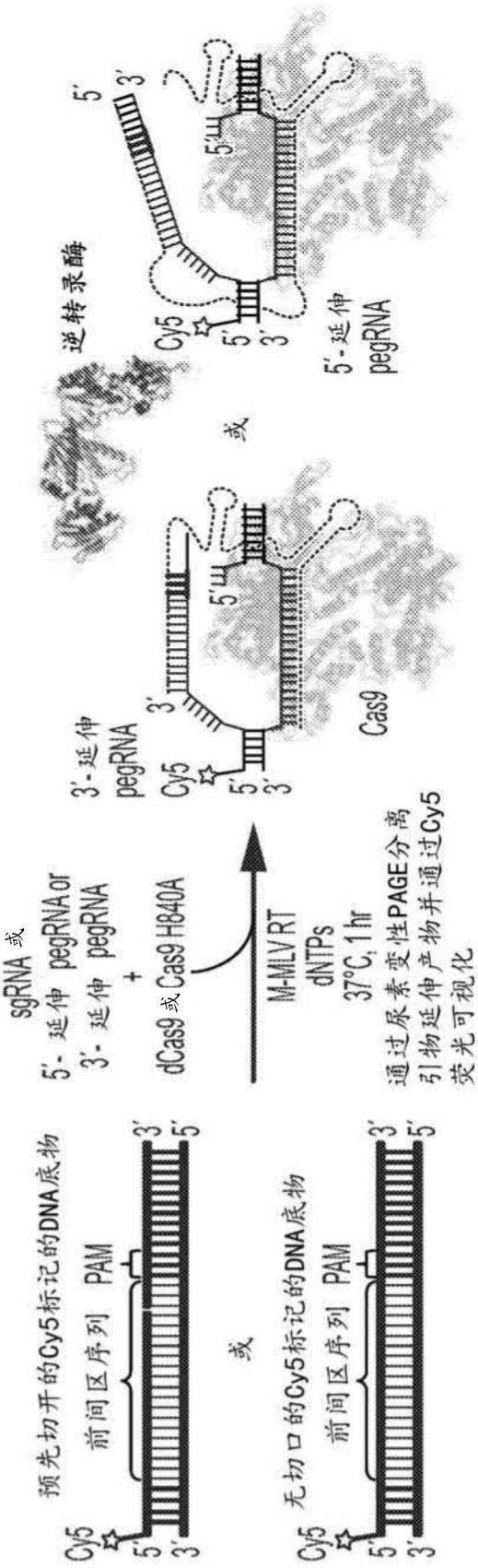

预先切开的Cy5标记的DNA底物
前间区序列 PAM
Cy5
5′
3′
或
无切口的Cy5标记的DNA底物
前间区序列 PAM
Cy5
5′
3′
sgRNA 或
5′- 延伸 pegRNA or
3′- 延伸 pegRNA
+
dCas9或Cas9 H840A
M-MLV RT
dNTPs
37°C, 1 hr
通过尿素变性PAGE分离
引物延伸产物并通过Cy5
荧光可视化
3′- 延伸
pegRNA
Cy5
5′
3′
Cas9
或
逆转录酶
Cy5
5′
3′
5′- 延伸
pegRNA

图44D

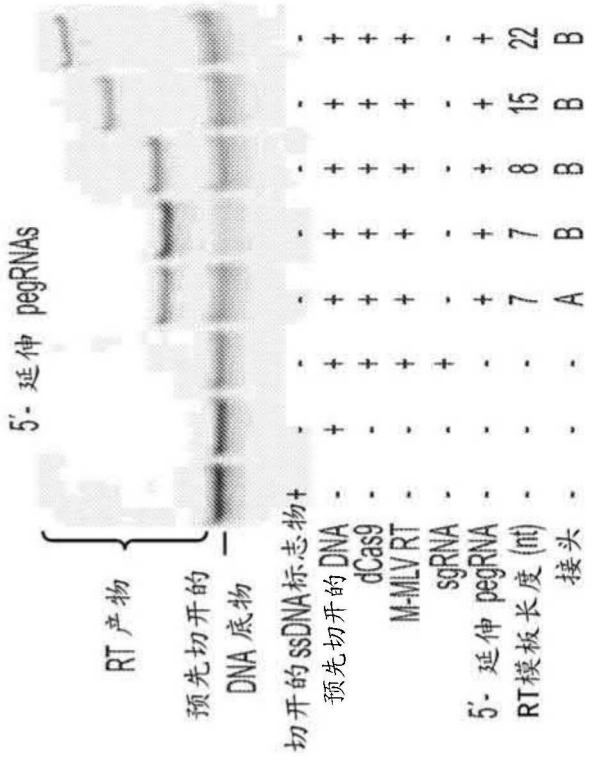

5'- 延伸 pegRNAs
RT 产物
预先切开的
DNA 底物
切开的 ssDNA 标志物 + - - - - - - - -
预先切开的 DNA - + + + + + + +
dCas9 - - + + + + + +
M-MLV RT - - + + + + + +
sgRNA - - + - - - - -
5'- 延伸 pegRNA - - - + + + + +
RT模板长度 (nt) - - - 7 7 8 15 22
接头 - - - A B B B B

图44E

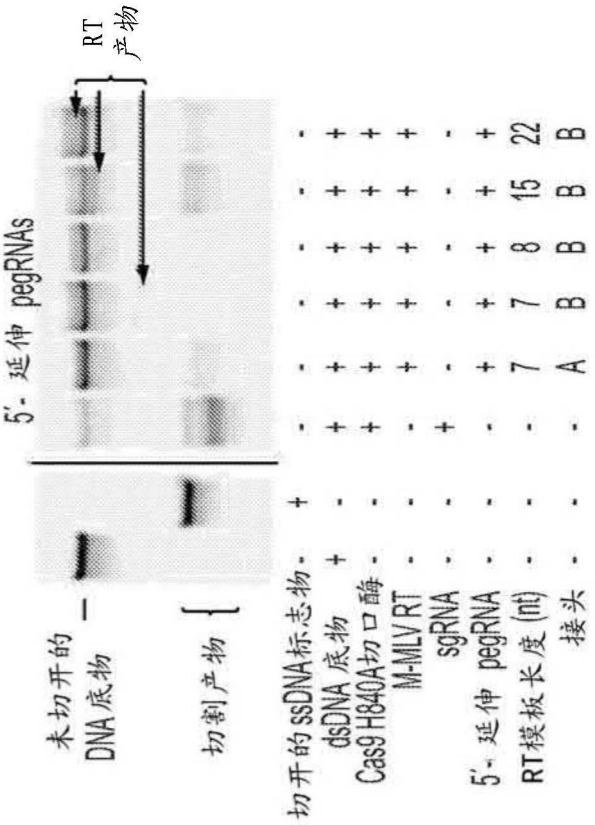

5'- 延伸 pegRNAs
RT
产物
未切开的
DNA 底物
切割产物
切开的 ssDNA 标志物 - + - - - - - - -
dsDNA 底物 + - + + + + + +
Cas9 H840A 切口酶 - - + + + + + +
M-MLV RT - - - + + + + +
sgRNA - - + - - - - -
5'- 延伸 pegRNA - - - + + + + +
RT模板长度 (nt) - - - 7 7 8 15 22
接头 - - - A B B B B

图44F

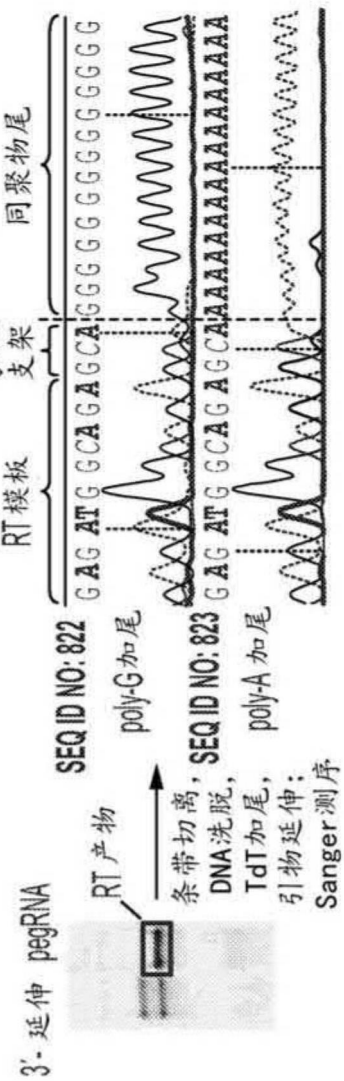

3'-延伸 pegRNA
RT 产物
条带切离，
DNA洗脱，
TdT加尾，
引物延伸；
Sanger测序
SEQ ID NO: 822
poly-G加尾
SEQ ID NO: 823
poly-A 加尾
RT 模板
gRNA 支架
同聚物尾

图44G

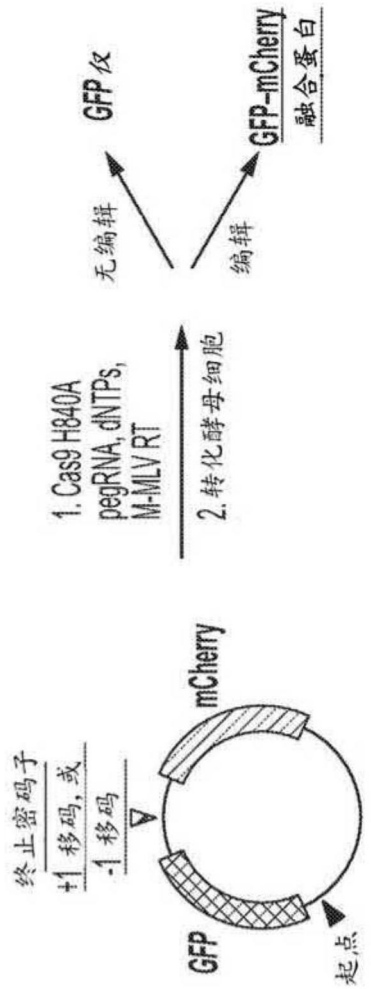

终止密码子
+1 移码, 或
-1 移码
mCherry
GFP
起点
1. Cas9 H840A
pegRNA, dNTPs,
M-MLV RT
2. 转化酵母细胞
无编辑
编辑
GFP 仅
GFP-mCherry
融合蛋白

图45A

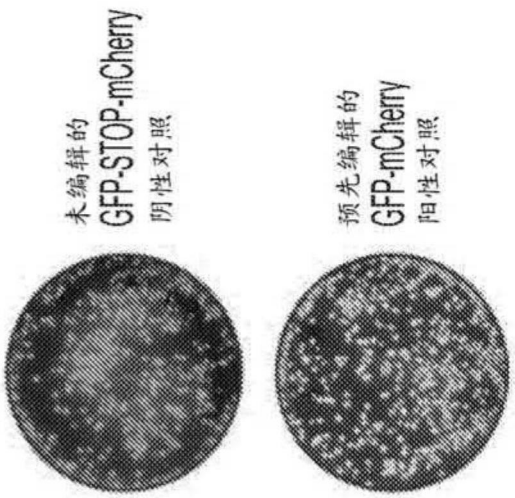

未编辑的
GFP-STOP-mCherry
阴性对照
预先编辑的
GFP-mCherry
阳性对照

图45B

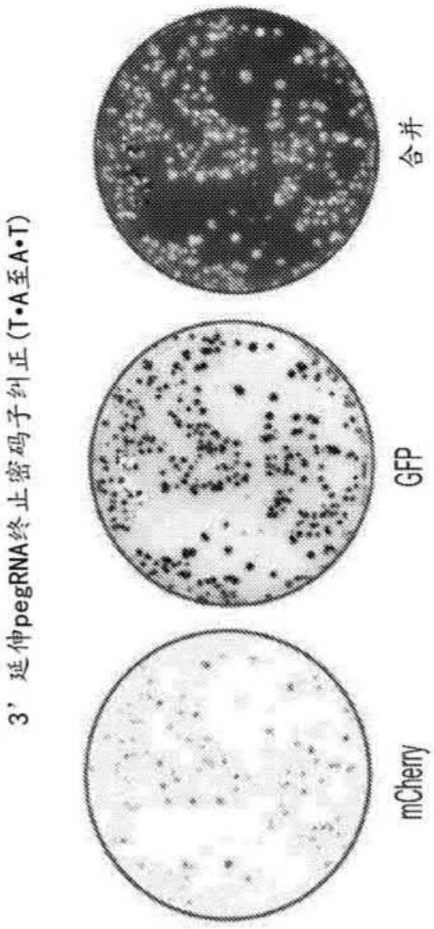


图45C

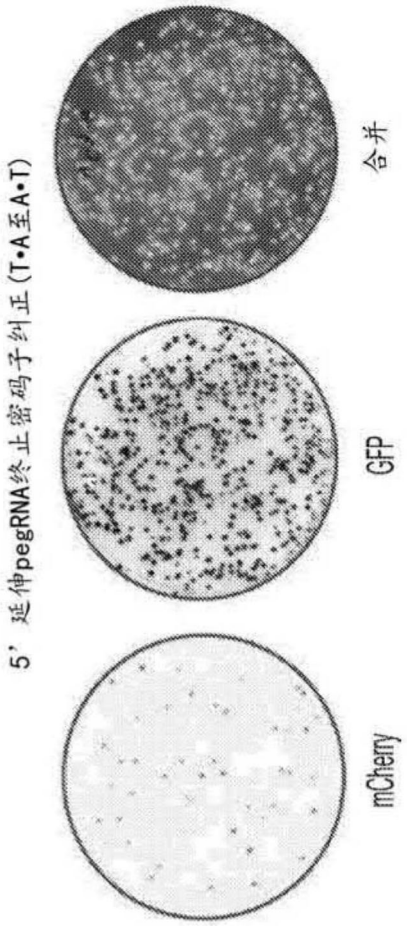


图45D

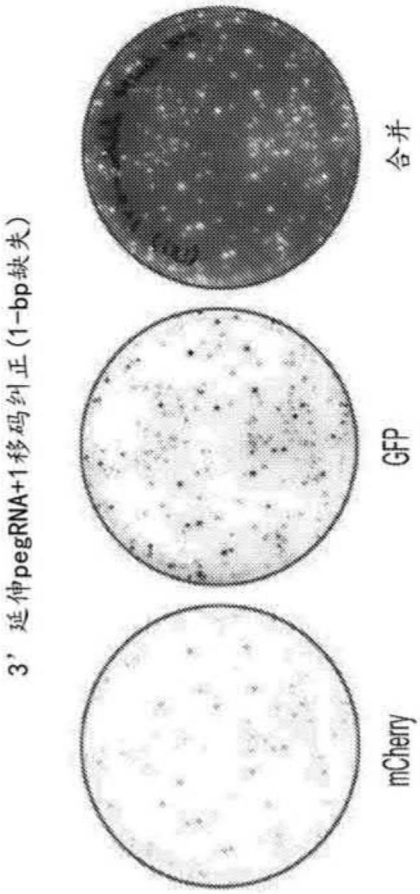


图45E

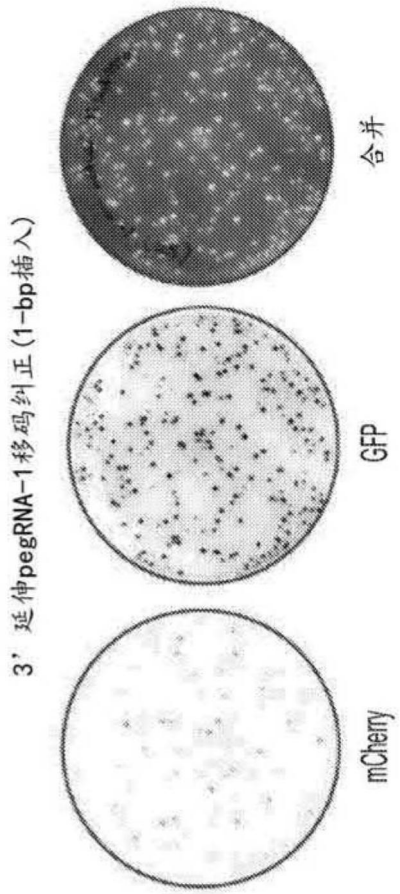


图45F

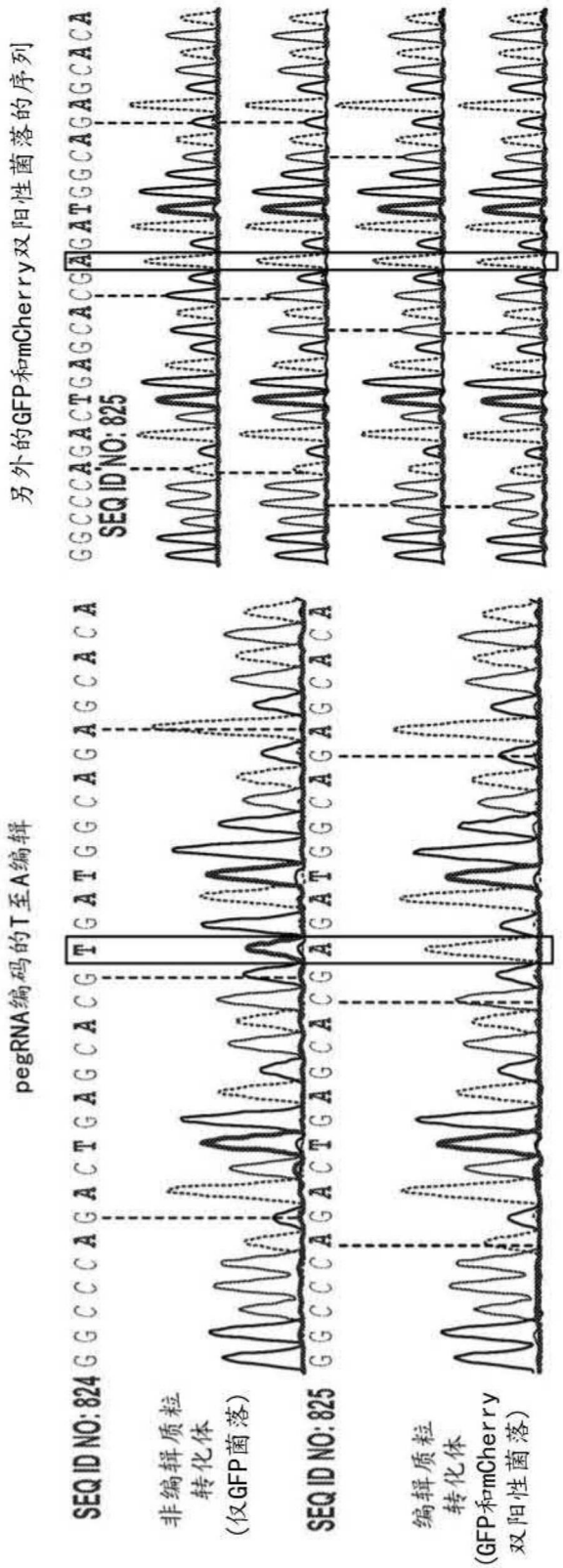

pegRNA编码的T至A编辑
另外的GFP和mCherry双阳性菌落的序列
SEQ ID NO: 824
SEQ ID NO: 825
SEQ ID NO: 825
非编辑质粒
转化体
(仅GFP菌落)
编辑质粒
转化体
(GFP和mCherry
双阳性菌落)

图45G

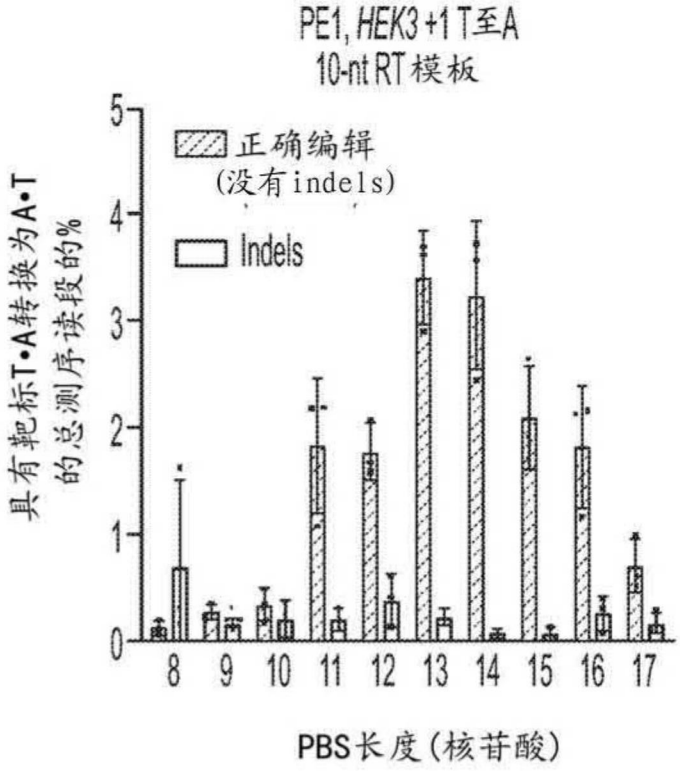

PE1, HEK3 +1 T至A
10-nt RT模板
正确编辑
(没有indels)
Indels
具有靶标T•A转换为A•T
的总测序读段的%
0
1
2
3
4
5
8
9
10
11
12
13
14
15
16
17
PBS长度(核苷酸)

图46A

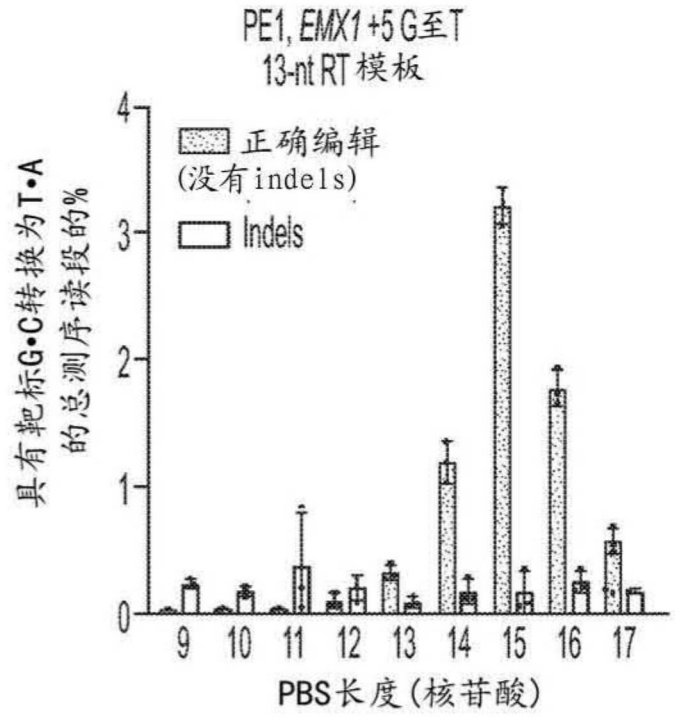

PE1, EMX1 +5 G至T
13-nt RT模板
正确编辑
(没有indels)
Indels
具有靶标G•C转换为T•A
的总测序读段的%
0
1
2
3
4
9
10
11
12
13
14
15
16
17
PBS长度(核苷酸)

图46B

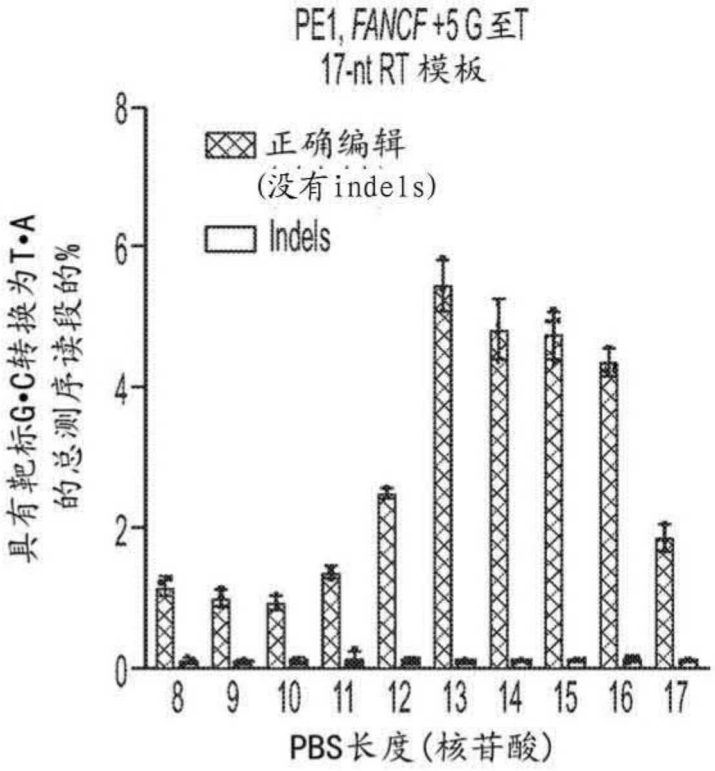

PE1, FANCF +5 G至T
17-nt RT 模板
正确编辑
(没有indels)
Indels
具有靶标G•C转换为T•A
的总测序读段的%
0
2
4
6
8
8 9 10 11 12 13 14 15 16 17
PBS长度(核苷酸)

图46C

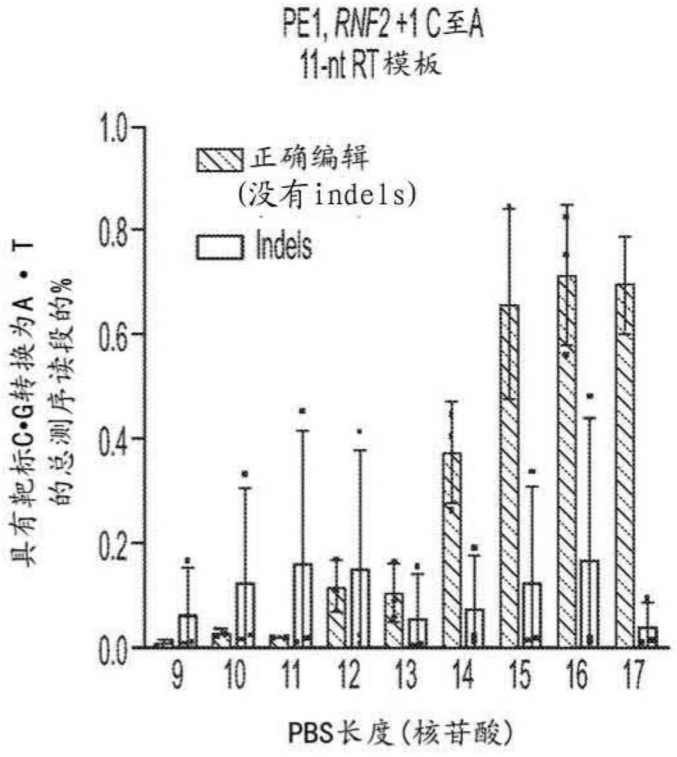

PE1, RNF2 +1 C至A
11-nt RT模板
正确编辑
(没有indels)
Indels
具有靶标C•G转换为A • T
的总测序读段的%
0.0
0.2
0.4
0.6
0.8
1.0
9 10 11 12 13 14 15 16 17
PBS长度(核苷酸)

图46D

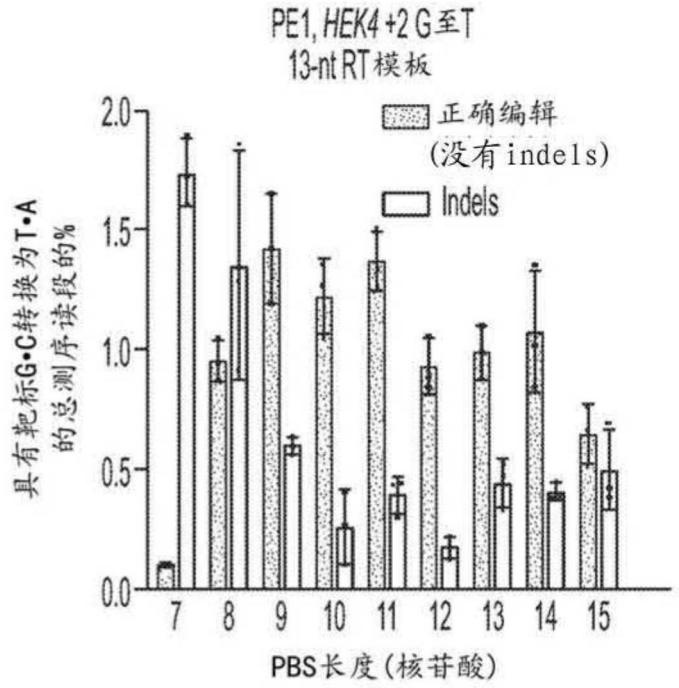

PE1, HEK4 +2 G至T
13-nt RT模板
正确编辑
(没有indels)
Indels
具有靶标G•C转换为T•A
的总测序读段的%
2.0
1.5
1.0
0.5
0.0
7
8
9
10
11
12
13
14
15
PBS长度(核苷酸)

图46E

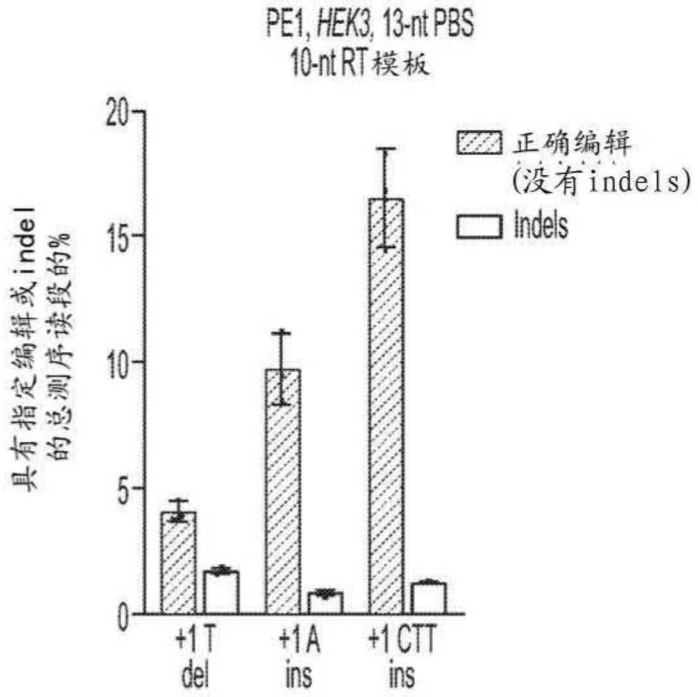

PE1, HEK3, 13-nt PBS
10-nt RT模板
正确编辑
(没有indels)
Indels
具有指定编辑或indel
的总测序读段的%
20
15
10
5
0
+1 T
del
+1 A
ins
+1 CTT
ins

图46F

PE1	Cas9 H840A-M-MLV RT (wt)
PE1-M1	Cas9 H840A-M-MLV RT (D200N)
PE1-M2	Cas9 H840A-M-MLV RT (D200N+L603W)
PE1-M3	Cas9 H840A-M-MLV RT (D200N+L603W+T330P)
PE1-M6	Cas9 H840A-M-MLV RT (D200N+L603W+T330P+D524G+E562Q+D583N)
PE1-M15	Cas9 H840A-M-MLV RT (P51L+S67R+E67K+T197A+H204R+E302K+F309N+W1331F+T330P+L345G+N454K+D524G+D583N+H594Q+D653N)
PE1-M3inv	M-MLV RT (D200N+L603W+T330P)-Cas9 H840A
PE2	Cas9 H840A-M-MLV RT (D200N+L603W+T330P+T306K+W313F)

图47A

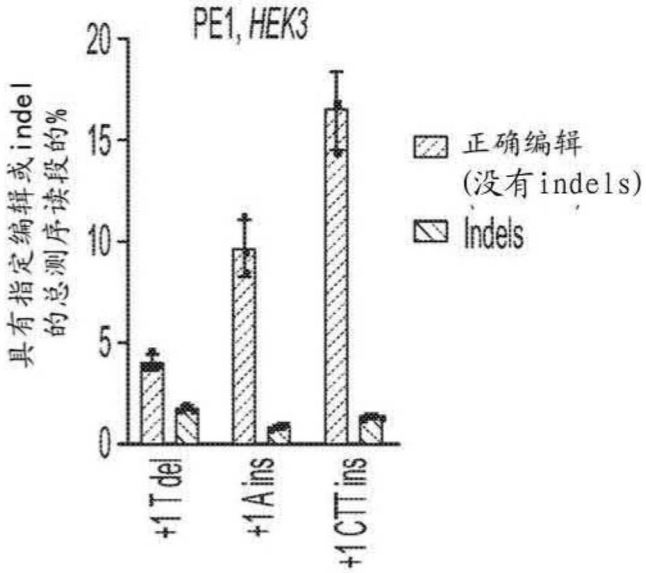


图47B

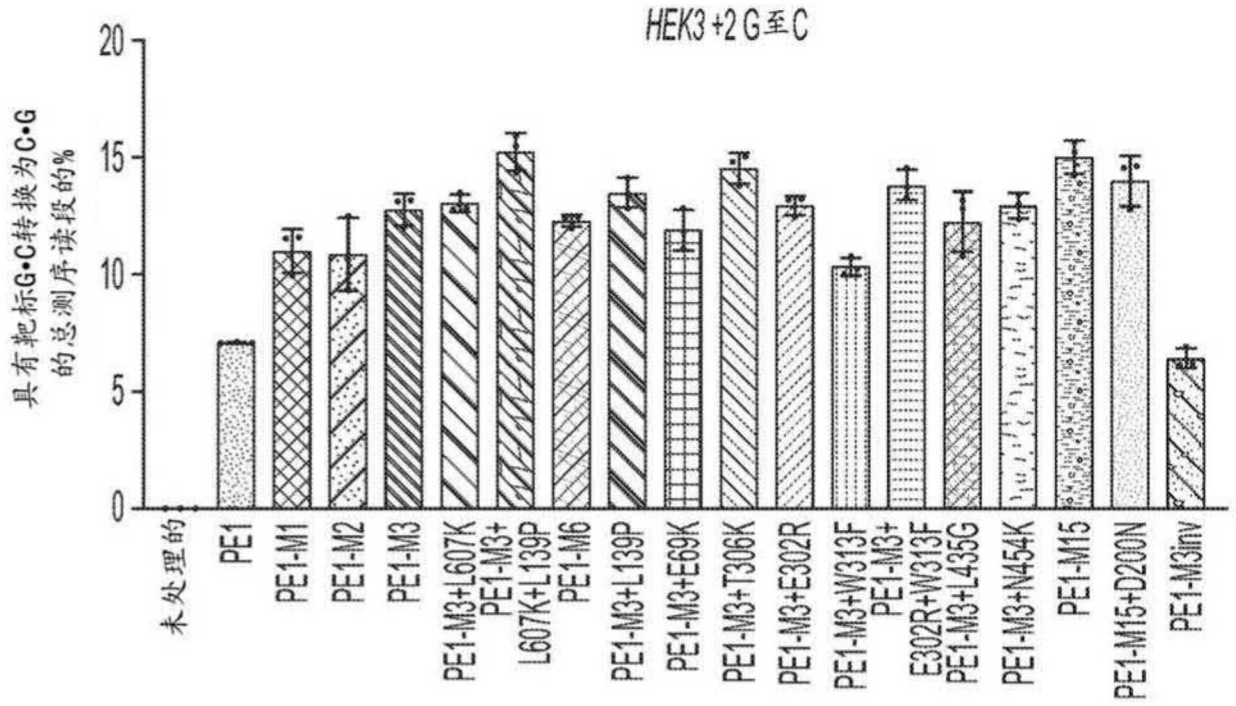

HEK3 +2 G至C
具有靶标G•C转换为C•G的总测序读段的%
20
15
10
5
0
未处理的
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

图47C

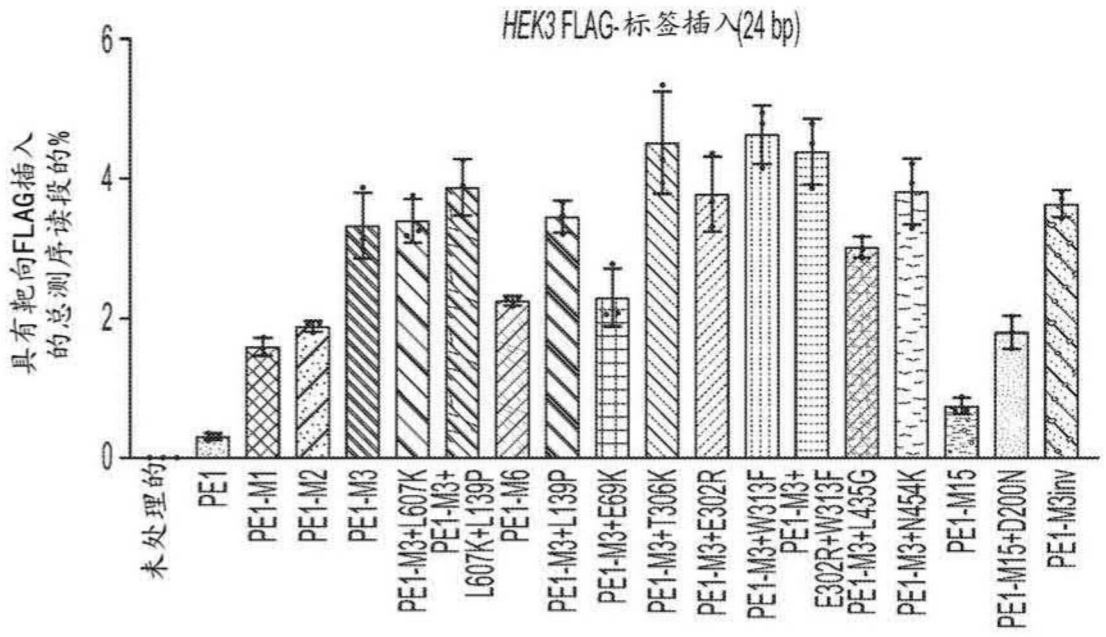

HEK3 FLAG-标签插入(24 bp)
具有靶向FLAG插入的总测序读段的%
6
4
2
0
未处理的
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

图47D

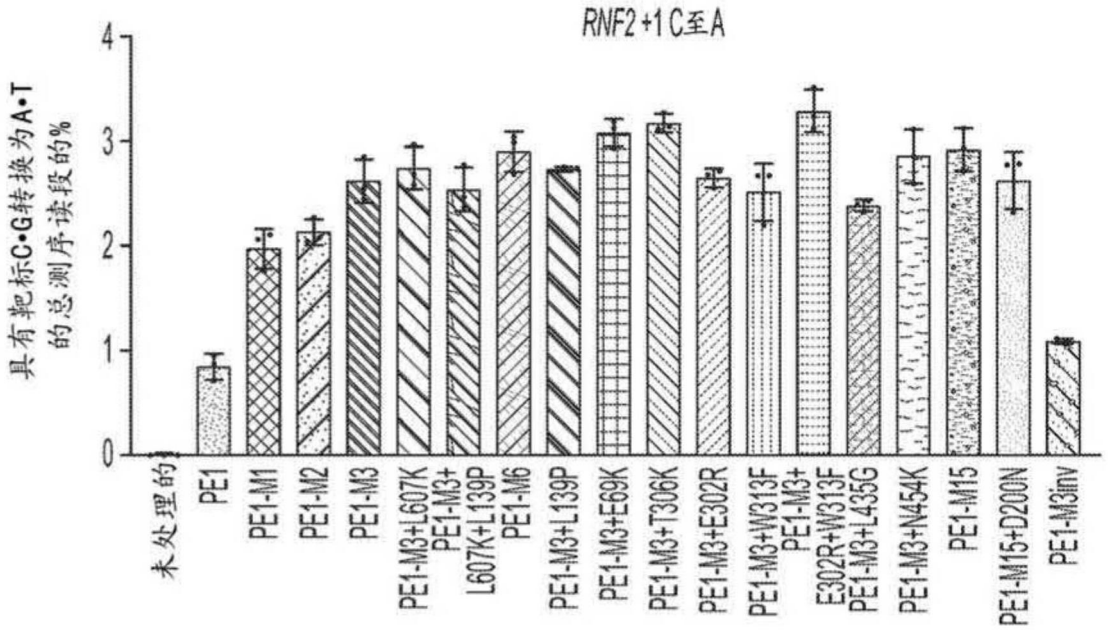

RNF2 +1 C至A
具有靶标C•G转换为A•T的总测序读段的%
0
1
2
3
4
未处理的
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

图47E

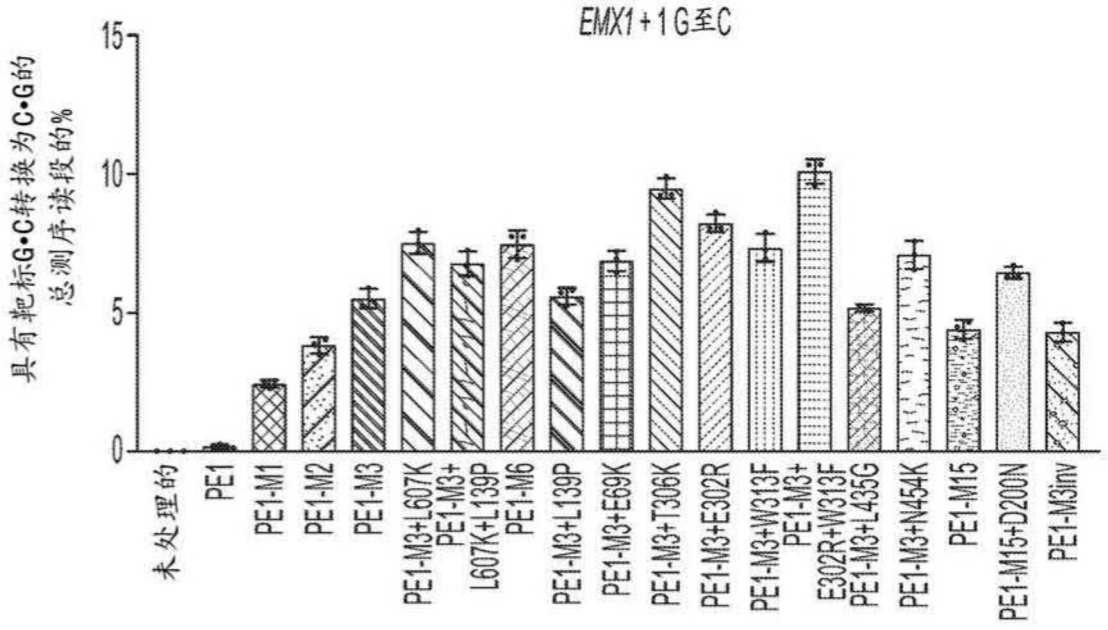

EMX1 +1 G至C
具有靶标G•C转换为C•G的总测序读段的%
0
5
10
15
未处理的
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

图47F

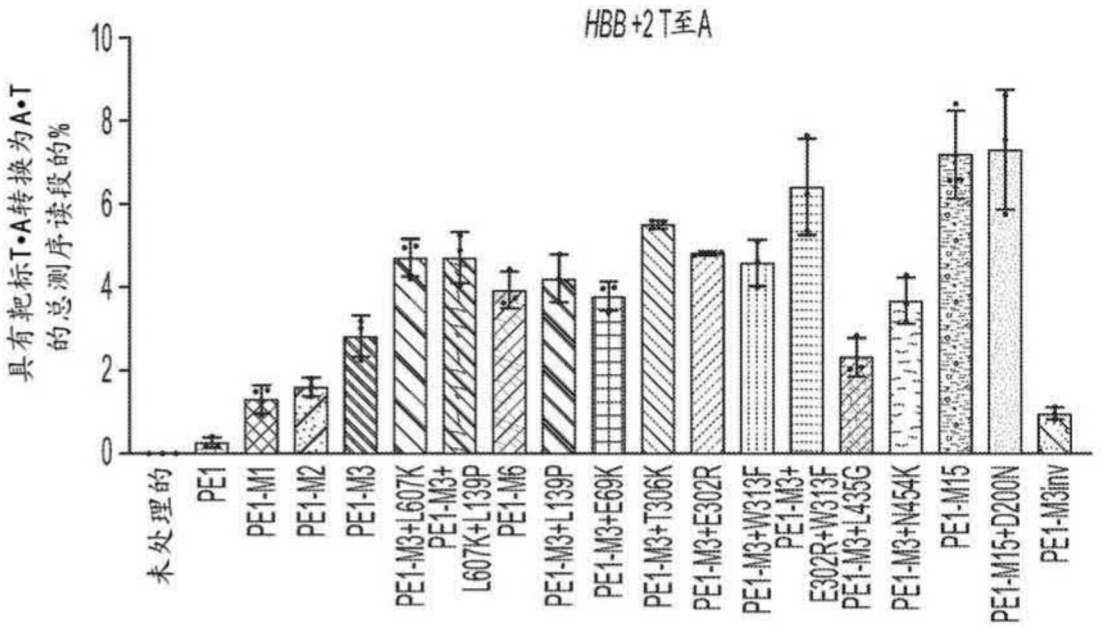

HBB+2T至A
具有靶标T•A转换为A•T的总测序读段的%
0
2
4
6
8
10
未处理的
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

图47G

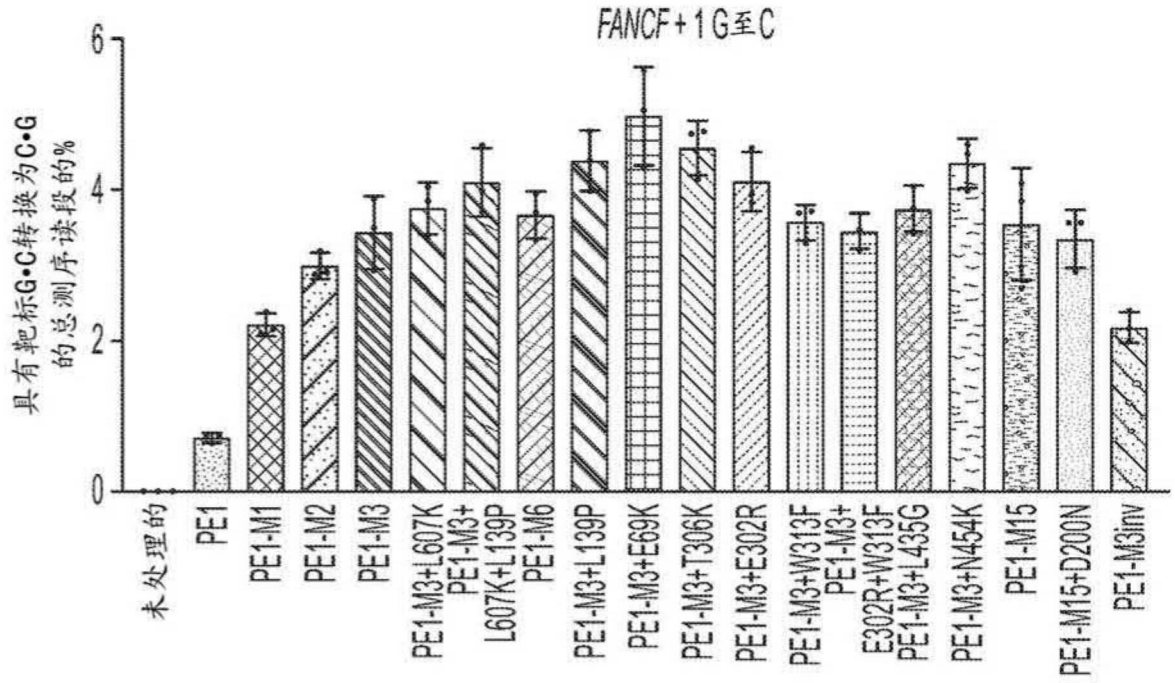

FANCF+1G至C
具有靶标G•C转换为C•G的总测序读段的%
0
2
4
6
未处理的
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

图47H

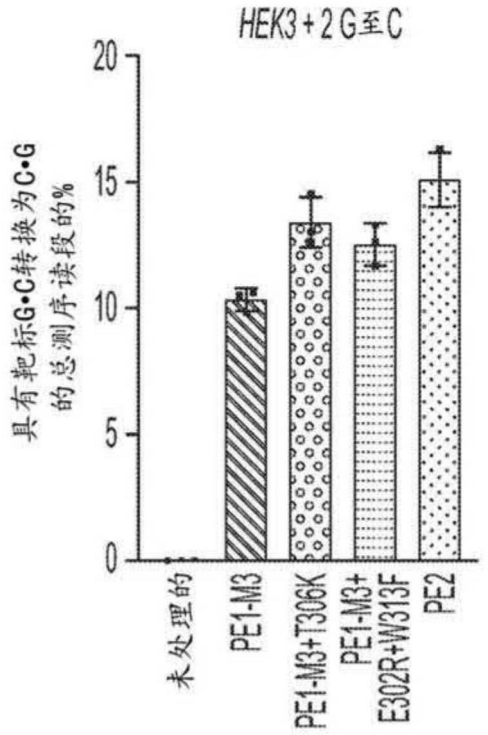

HEK3+2G至C
具有靶标G•C转换为C•G的总测序读段的%
20
15
10
5
0
未处理的
PE1-M3
PE1-M3+T306K
PE1-M3+E302R+W313F
PE2

图47I

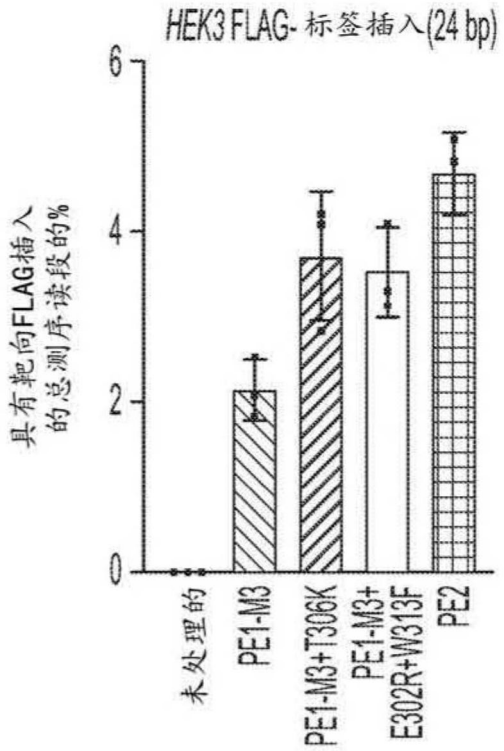

HEK3 FLAG-标签插入(24 bp)
具有靶向FLAG插入的总测序读段的%
6
4
2
0
未处理的
PE1-M3
PE1-M3+T306K
PE1-M3+E302R+W313F
PE2

图47J

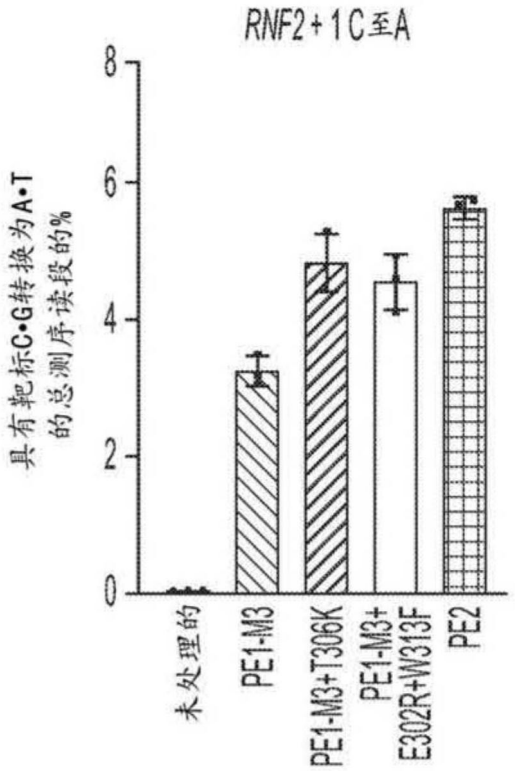

RNF2+1C至A
具有靶标C•G转换为A•T
的总测序读段的%
0
2
4
6
8
未处理的
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

图47K

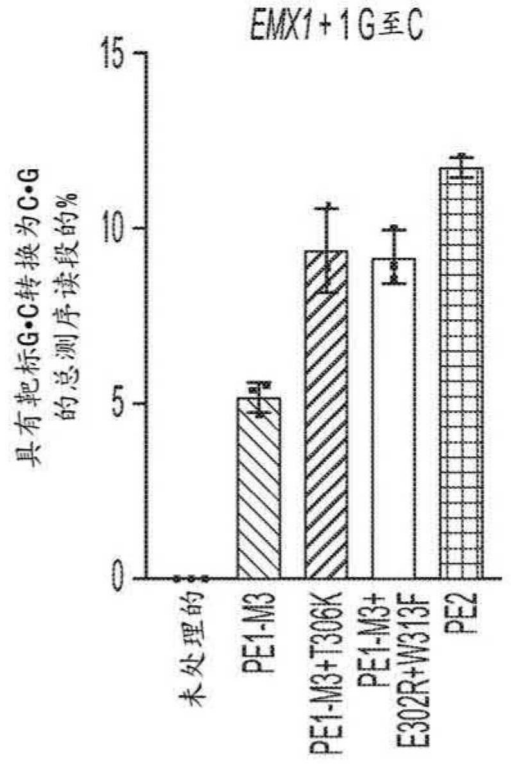

EMX1+1G至C
具有靶标G•C转换为C•G
的总测序读段的%
0
5
10
15
未处理的
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

图47L

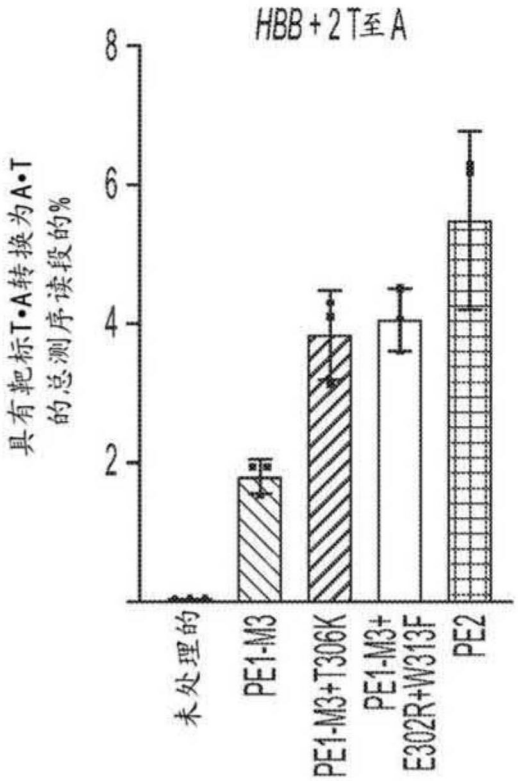

HBB+2T至A
具有靶标T•A转换为A•T的总测序读段的%
8
6
4
2
未处理的
PE1-M3
PE1-M3+T306K
PE1-M3+E302R+W313F
PE2

图47M

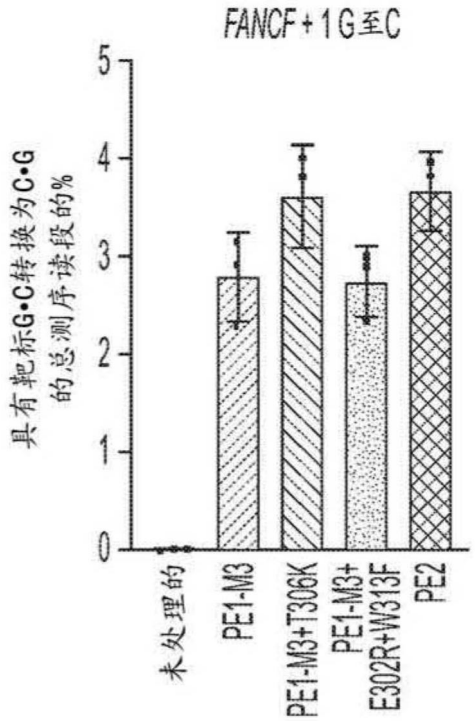

FANCF+1G至C
具有靶标G•C转换为C•G的总测序读段的%
5
4
3
2
1
0
未处理的
PE1-M3
PE1-M3+T306K
PE1-M3+E302R+W313F
PE2

图47N

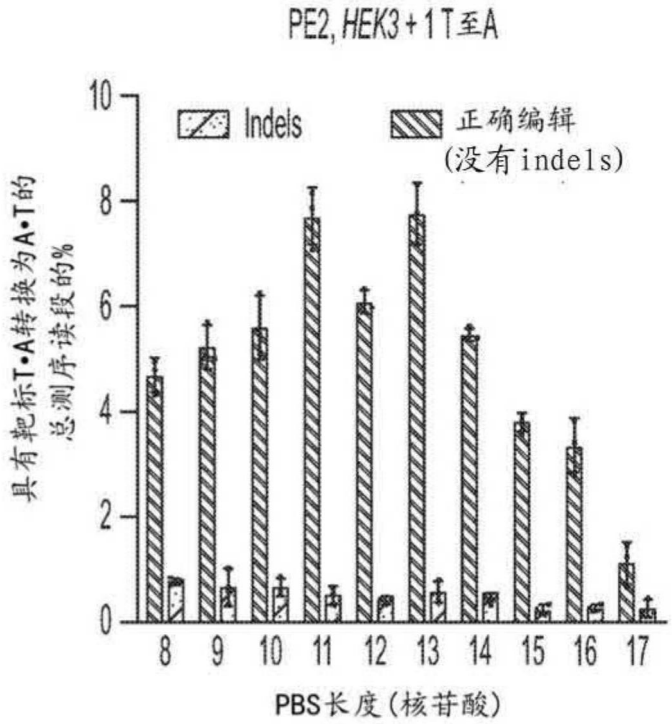

PE2, HEK3 + 1 T至A
Indels
正确编辑
(没有indels)
具有靶标T•A转换为A•T的
总测序读段的%
0
2
4
6
8
10
8 9 10 11 12 13 14 15 16 17
PBS长度(核苷酸)

图470

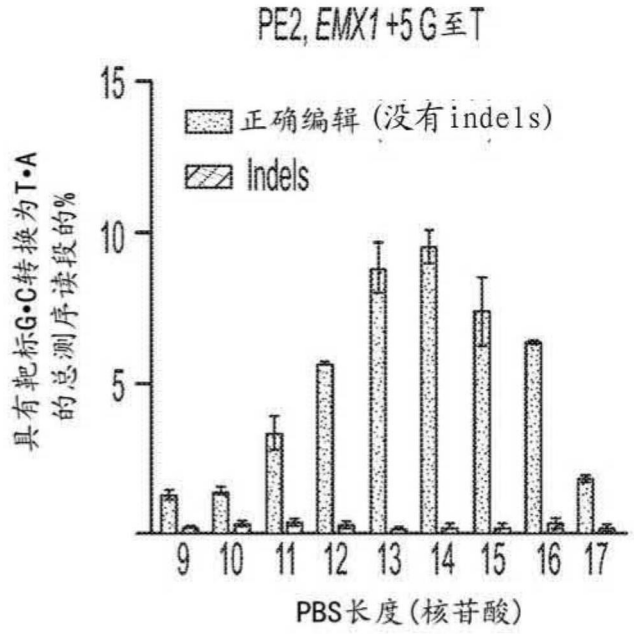

PE2, EMX1 +5 G至T
正确编辑 (没有indels)
Indels
具有靶标G•C转换为T•A
的总测序读段的%
5
10
15
9 10 11 12 13 14 15 16 17
PBS长度(核苷酸)

图47P

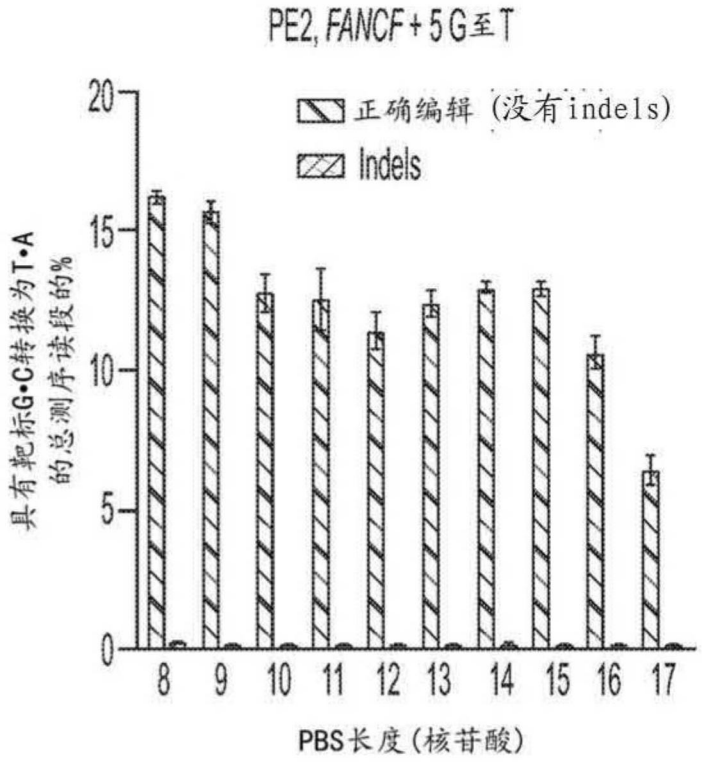

PE2, FANCF + 5 G至T
正确编辑 (没有indels)
Indels
具有靶标G•C转换为T•A的总测序读段的%
20
15
10
5
0
8 9 10 11 12 13 14 15 16 17
PBS长度(核苷酸)

图47Q

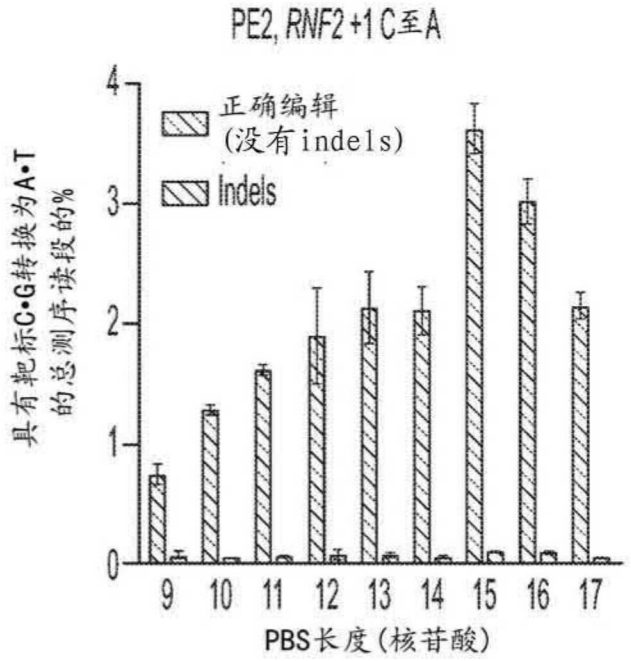

PE2, RNF2 +1 C至A
正确编辑
(没有indels)
Indels
具有靶标C•G转换为A•T的总测序读段的%
4
3
2
1
0
9 10 11 12 13 14 15 16 17
PBS长度(核苷酸)

图47R

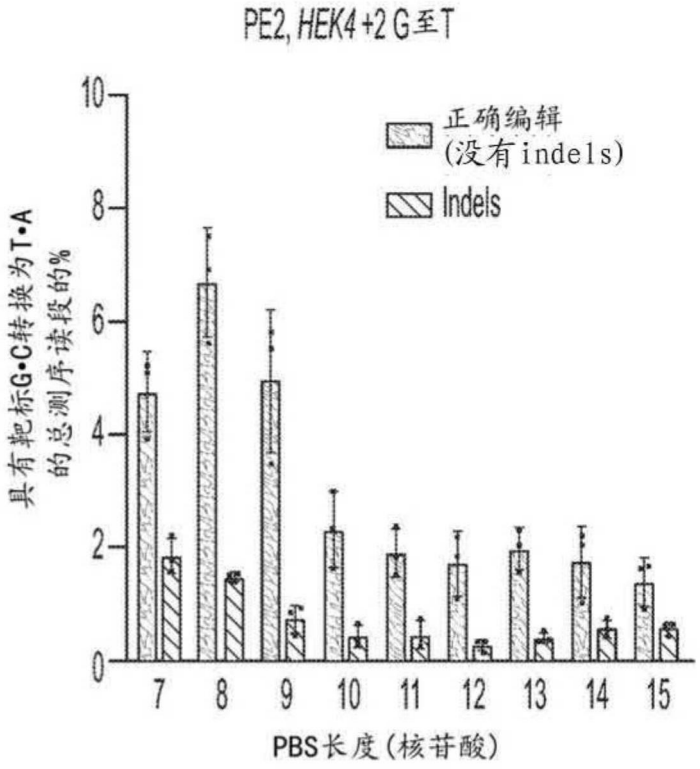

PE2, HEK4 +2 G至T
正确编辑
(没有indels)
Indels
具有靶标G•C转换为T•A
的总测序读段的%
0
2
4
6
8
10
7
8
9
10
11
12
13
14
15
PBS长度(核苷酸)

图47S

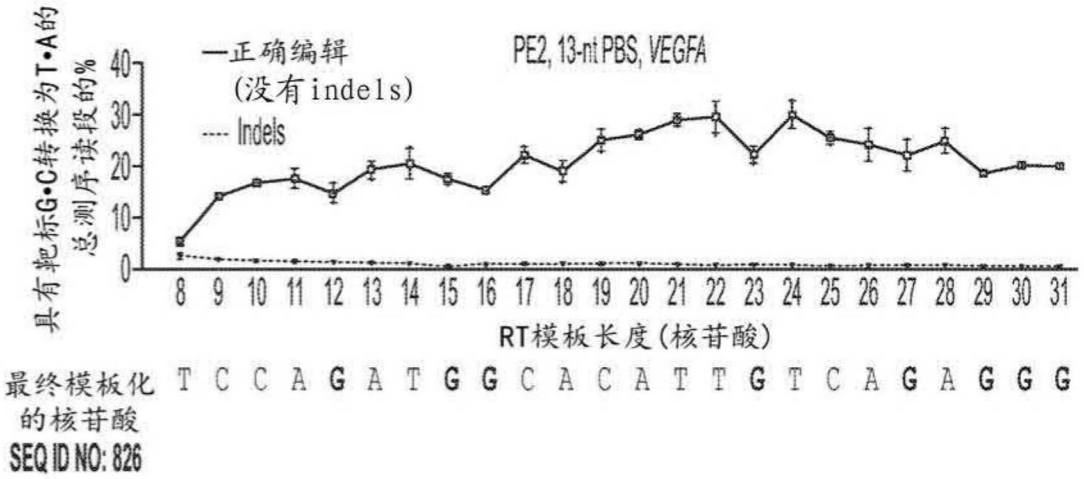

具有靶标G•C转换为T•A的
总测序读段的%
正确编辑
(没有indels)
Indels
PE2, 13-nt PBS, VEGFA
0
10
20
30
40
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
RT模板长度(核苷酸)
最终模板化
的核苷酸
SEQ ID NO: 826
T C C A G A T G G C A C A T T G T C A G A G G G

图48A

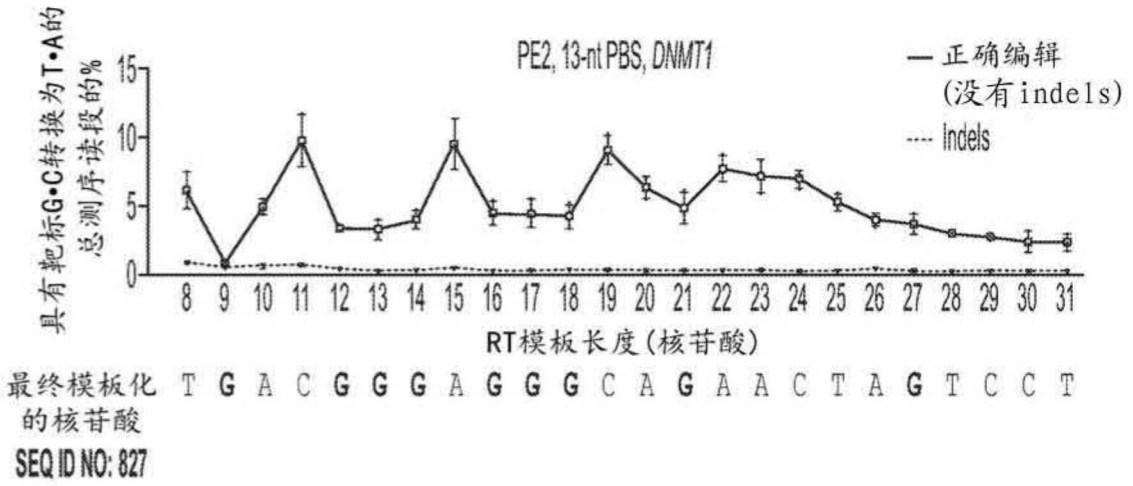

PE2, 13-nt PBS, DNMT1
— 正确编辑
(没有indels)
---- Indels
具有靶标G•C转换为T•A的总测序读段的%
0
5
10
15
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
RT模板长度(核苷酸)
最终模板化的核苷酸 T G A C G G G A G G G C A G A A C T A G T C C T
SEQ ID NO: 827

图48B

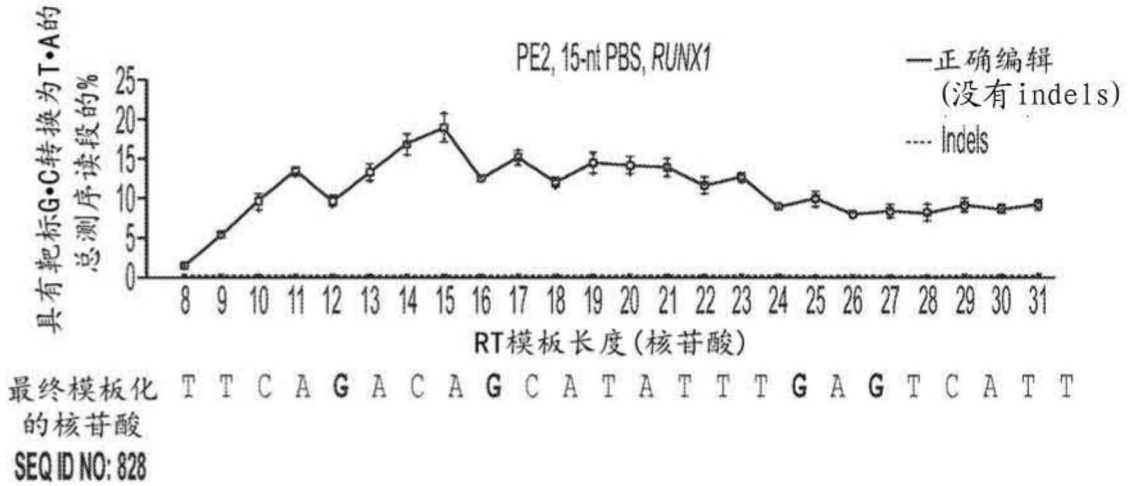

PE2, 15-nt PBS, RUNX1
—正确编辑
(没有indels)
---- Indels
具有靶标G•C转换为T•A的总测序读段的%
0
5
10
15
20
25
8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
RT模板长度(核苷酸)
最终模板化的核苷酸 T T C A G A C A G C A T A T T T G A G T C A T T
SEQ ID NO: 828

图48C

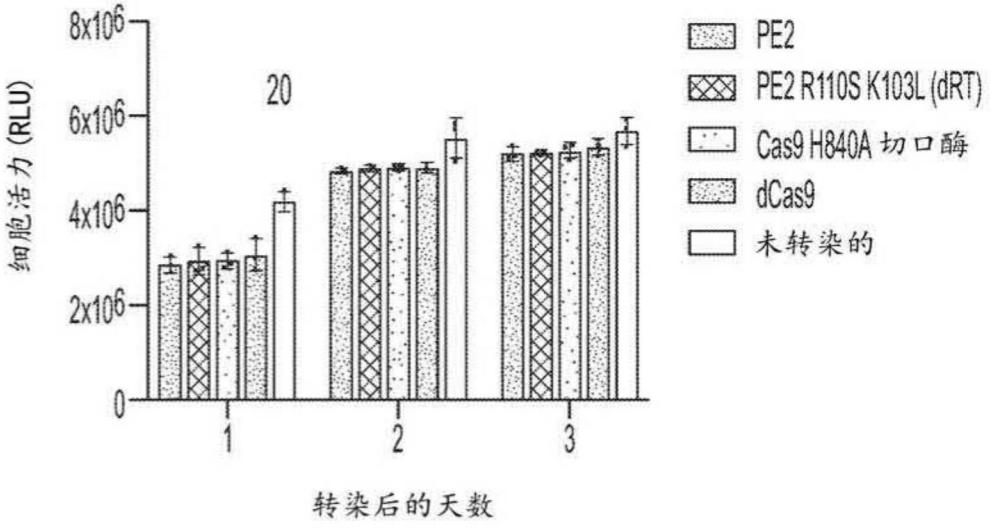

8x10^6
6x10^6
4x10^6
2x10^6
0
细胞活力 (RLU)
20
1
2
3
转染后的天数
PE2
PE2 R110S K103L (dRT)
Cas9 H840A 切口酶
dCas9
未转染的

图49A

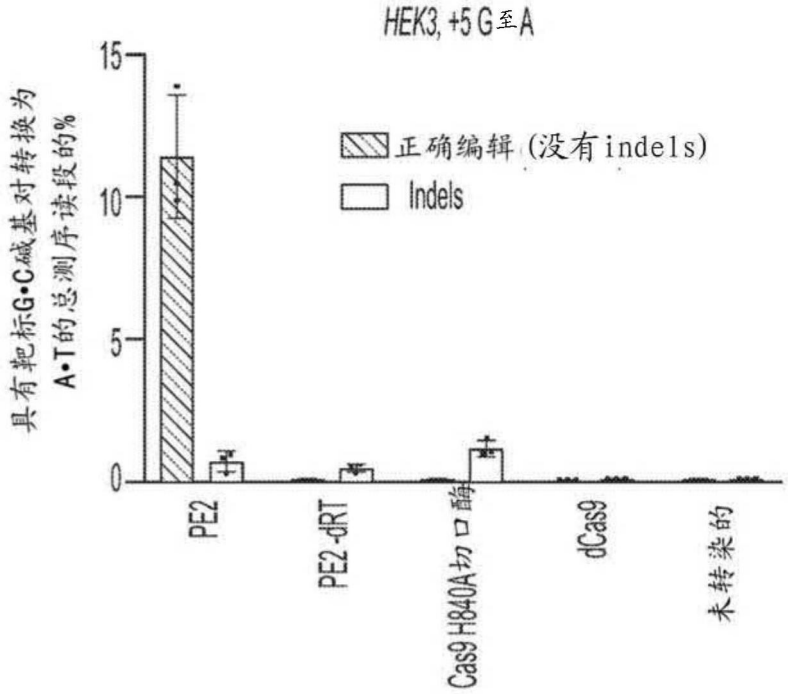

HEK3, +5 G至A
15
10
5
0
具有靶标G•C碱基对转换为A•T的总测序读段的%
正确编辑 (没有indels)
Indels
PE2
PE2-dRT
Cas9 H840A切口酶
dCas9
未转染的

图49B

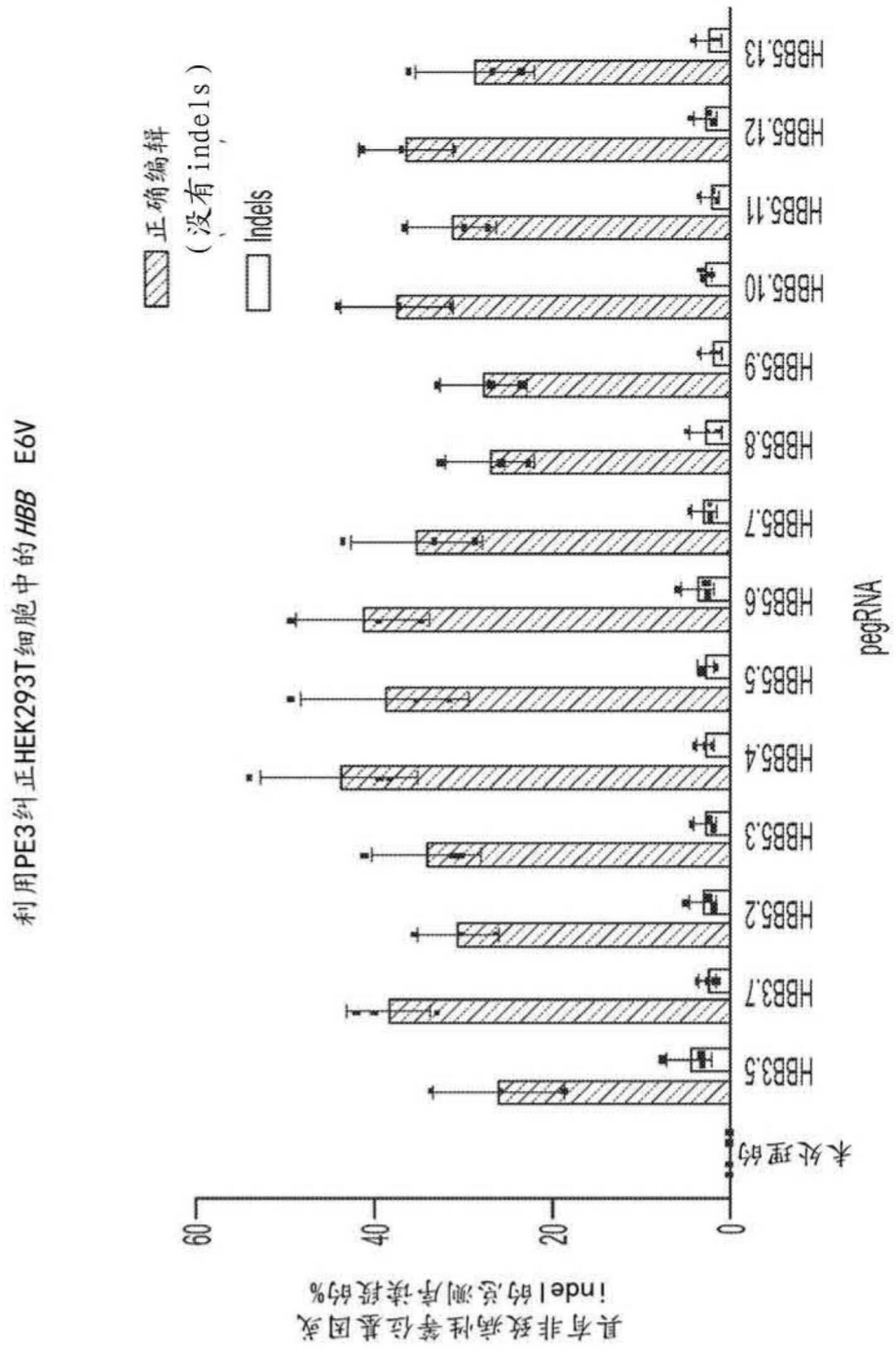

利用PE3纠正HEK293T细胞中的HBB E6V
正确编辑
(没有indels)
Indels
具有非致病性等位基因或
indel的总测序读段的%
0
20
40
60
未处理的
HBB3.5
HBB3.7
HBB5.2
HBB5.3
HBB5.4
HBB5.5
HBB5.6
HBB5.7
HBB5.8
HBB5.9
HBB5.10
HBB5.11
HBB5.12
HBB5.13
pegRNA

图50A

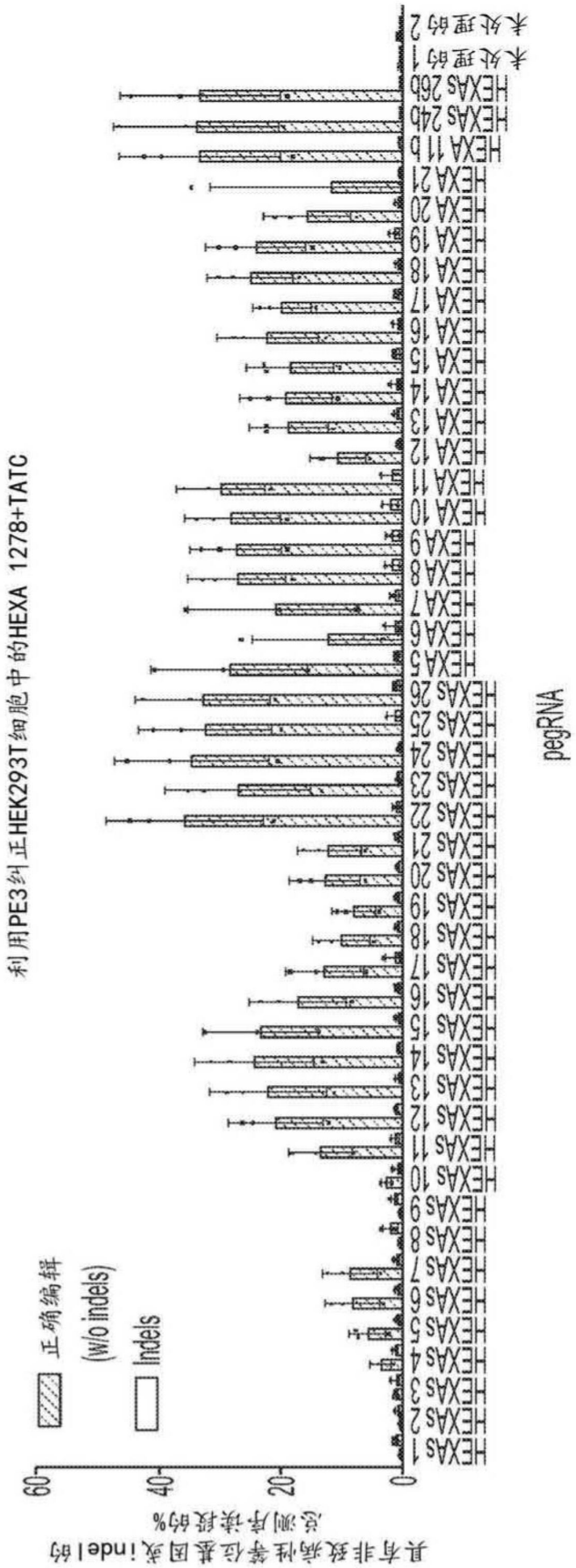

利用PE3纠正HEK293T细胞中的HEXA 1278+TATC
具有非致病性等位基因或indel的总测序读段的%
0
20
40
60
正确编辑 (w/o indels)
Indels
pegRNA
未处理的1
未处理的2
HEXA 11b
HEXAS 24b
HEXAS 26b

图50B

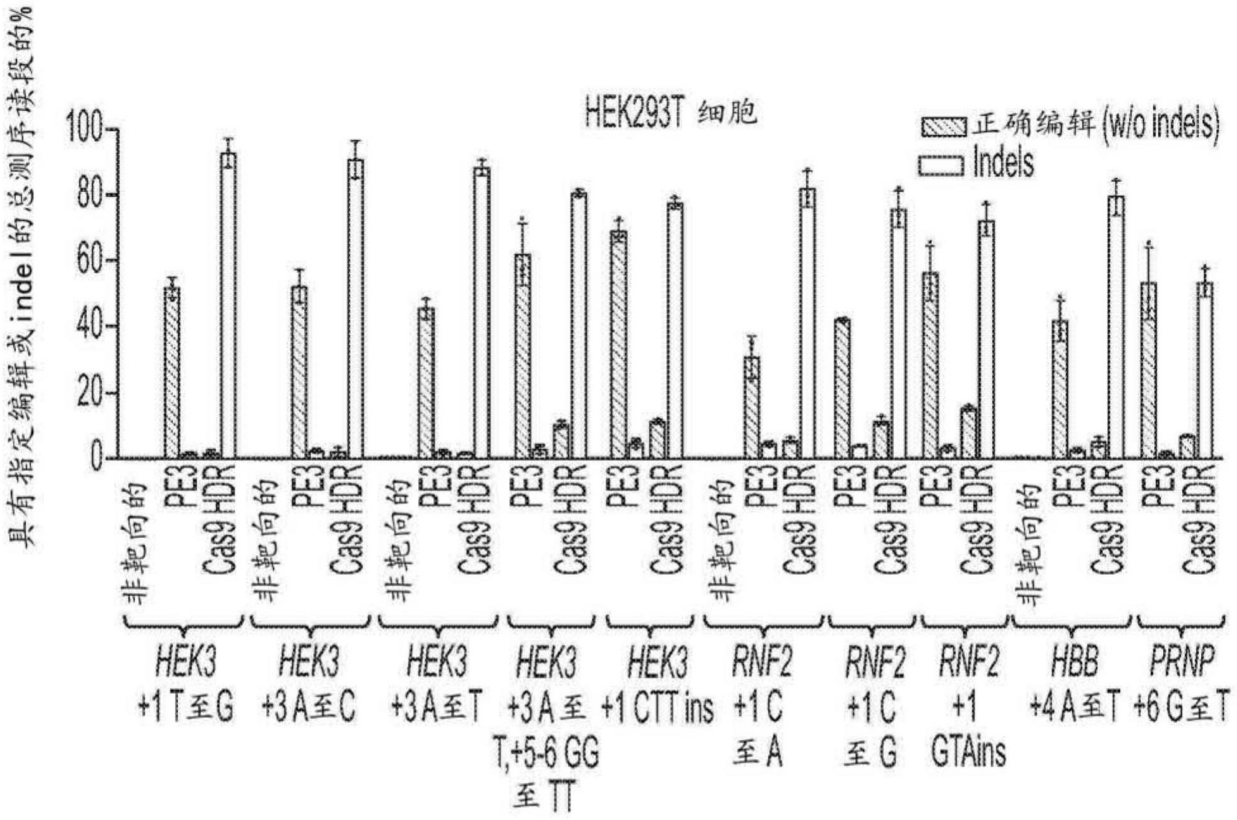

具有指定编辑或indel的总测序读段的%
HEK293T 细胞
正确编辑(w/o indels)
Indels
100
80
60
40
20
0
非靶向的
PE3
Cas9 HDR
HEK3 +1 T至G
HEK3 +3 A至C
HEK3 +3 A至T
HEK3 +3 A至T,+5-6 GG至TT
HEK3 +1 CTT ins
RNF2 +1 C至A
RNF2 +1 C至G
RNF2 +1 GTA ins
HBB +4 A至T
PRNP +6 G至T

图51A

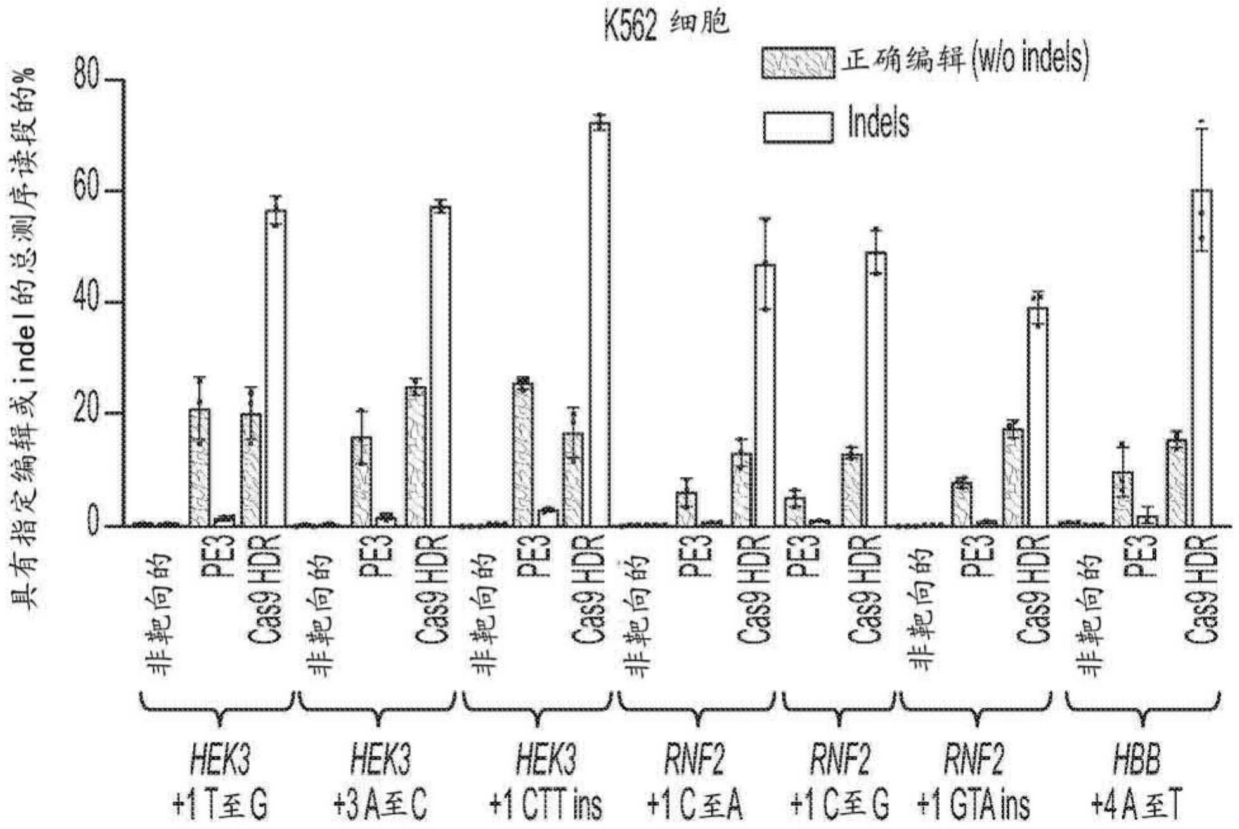

K562 细胞
正确编辑(w/o indels)
Indels
具有指定编辑或indel的总测序读段的%
80
60
40
20
0
非靶向的
PE3
Cas9 HDR
HEK3 +1 T至G
HEK3 +3 A至C
HEK3 +1 CTT ins
RNF2 +1 C至A
RNF2 +1 C至G
RNF2 +1 GTA ins
HBB +4 A至T

图51B

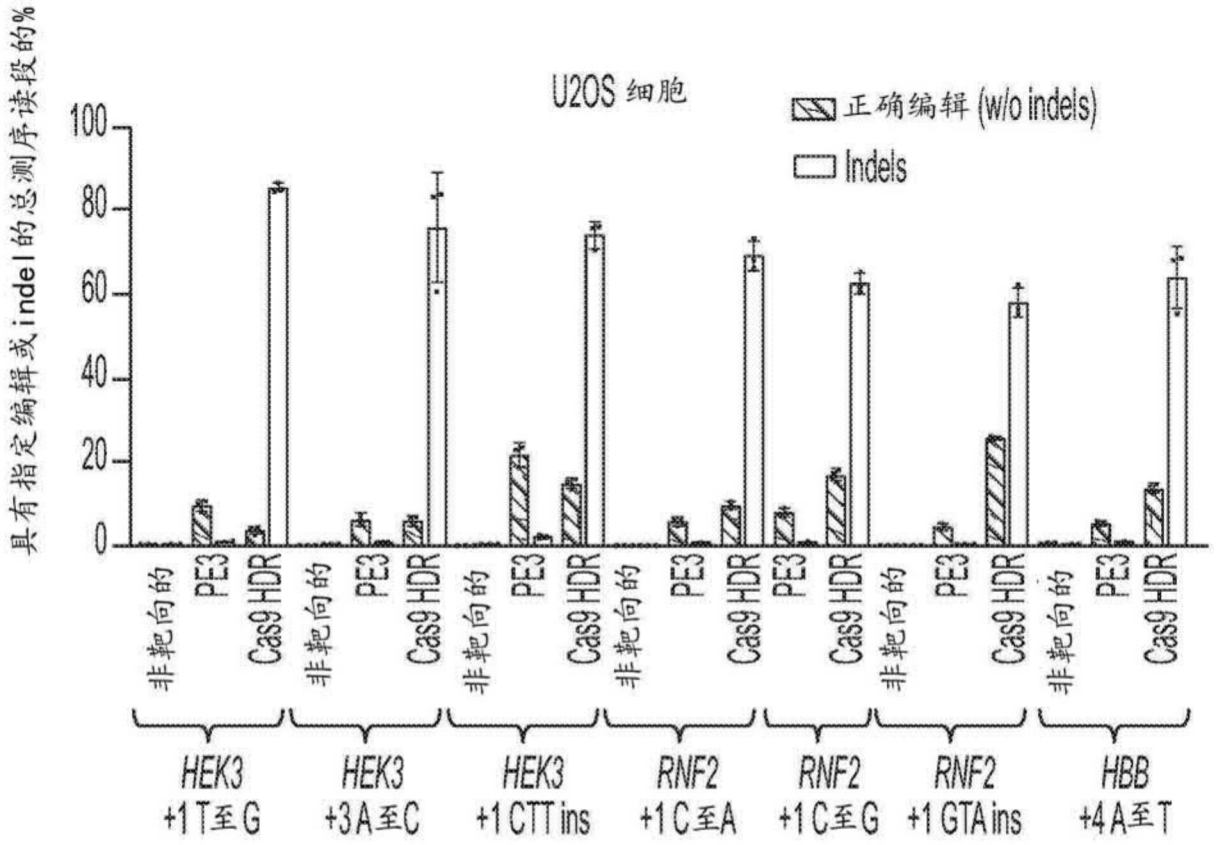

具有指定编辑或indel的总测序读段的%
U2OS 细胞
正确编辑 (w/o indels)
Indels
100
80
60
40
20
0
非靶向的
PE3
Cas9 HDR
HEK3 +1 T至G
HEK3 +3 A至C
HEK3 +1 CTT ins
RNF2 +1 C至A
RNF2 +1 C至G
RNF2 +1 GTA ins
HBB +4 A至T

图51C

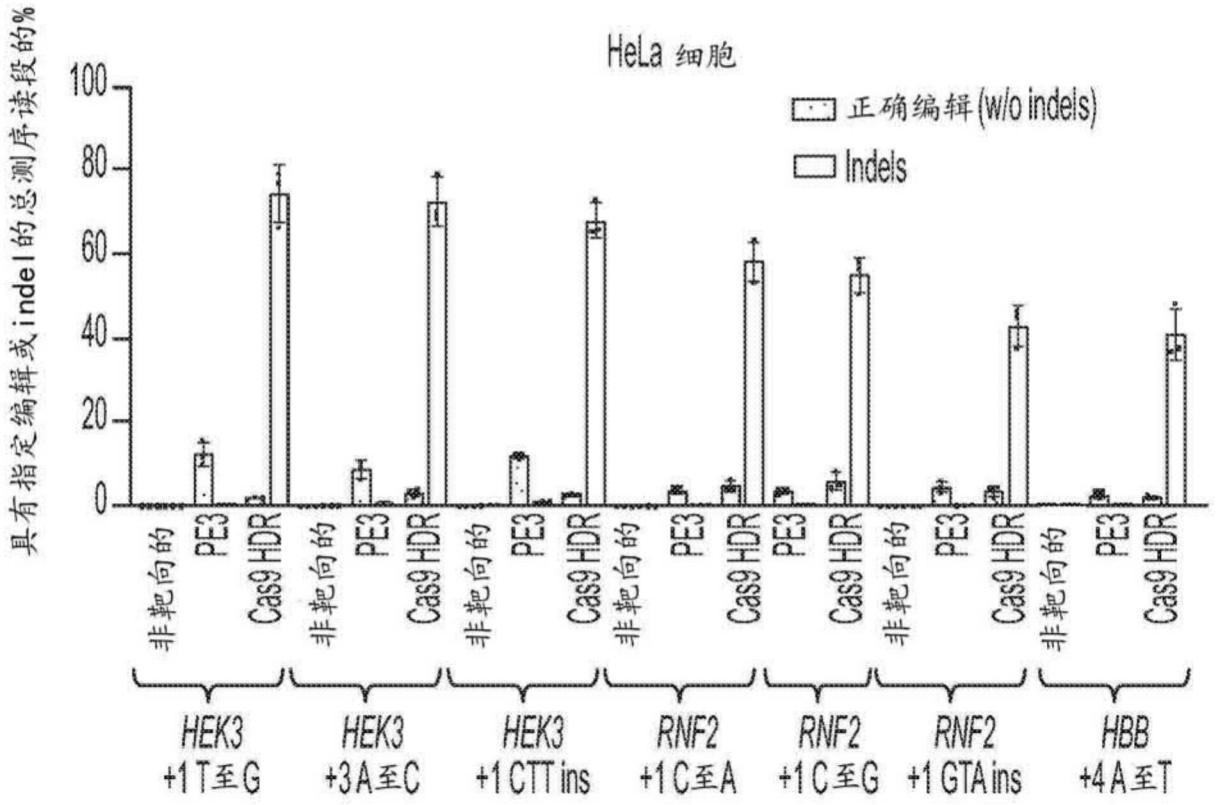

HeLa 细胞
正确编辑(w/o indels)
Indels
具有指定编辑或indel的总测序读段的%
100
80
60
40
20
0
非靶向的
PE3
Cas9 HDR
HEK3 +1 T至G
HEK3 +3 A至C
HEK3 +1 CTT ins
RNF2 +1 C至A
RNF2 +1 C至G
RNF2 +1 GTA ins
HBB +4 A至T

图51D

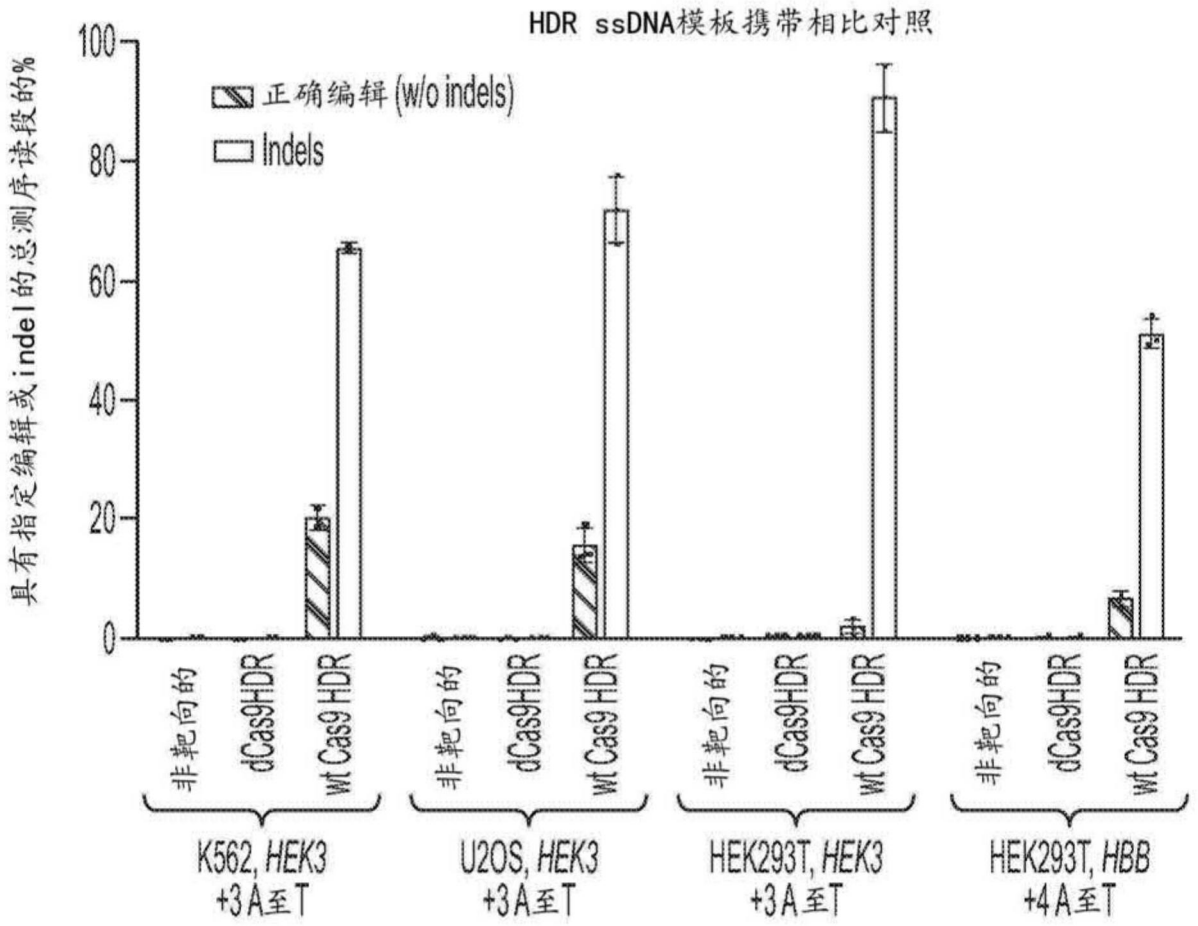

HDR ssDNA模板携带相比对照
具有指定编辑或indel的总测序读段的%
正确编辑 (w/o indels)
Indels
100
80
60
40
20
0
非靶向的
dCas9HDR
Wt Cas9 HDR
K562, HEK3 +3A至T
U2OS, HEK3 +3A至T
HEK293T, HEK3 +3A至T
HEK293T, HBB +4A至T

图51E

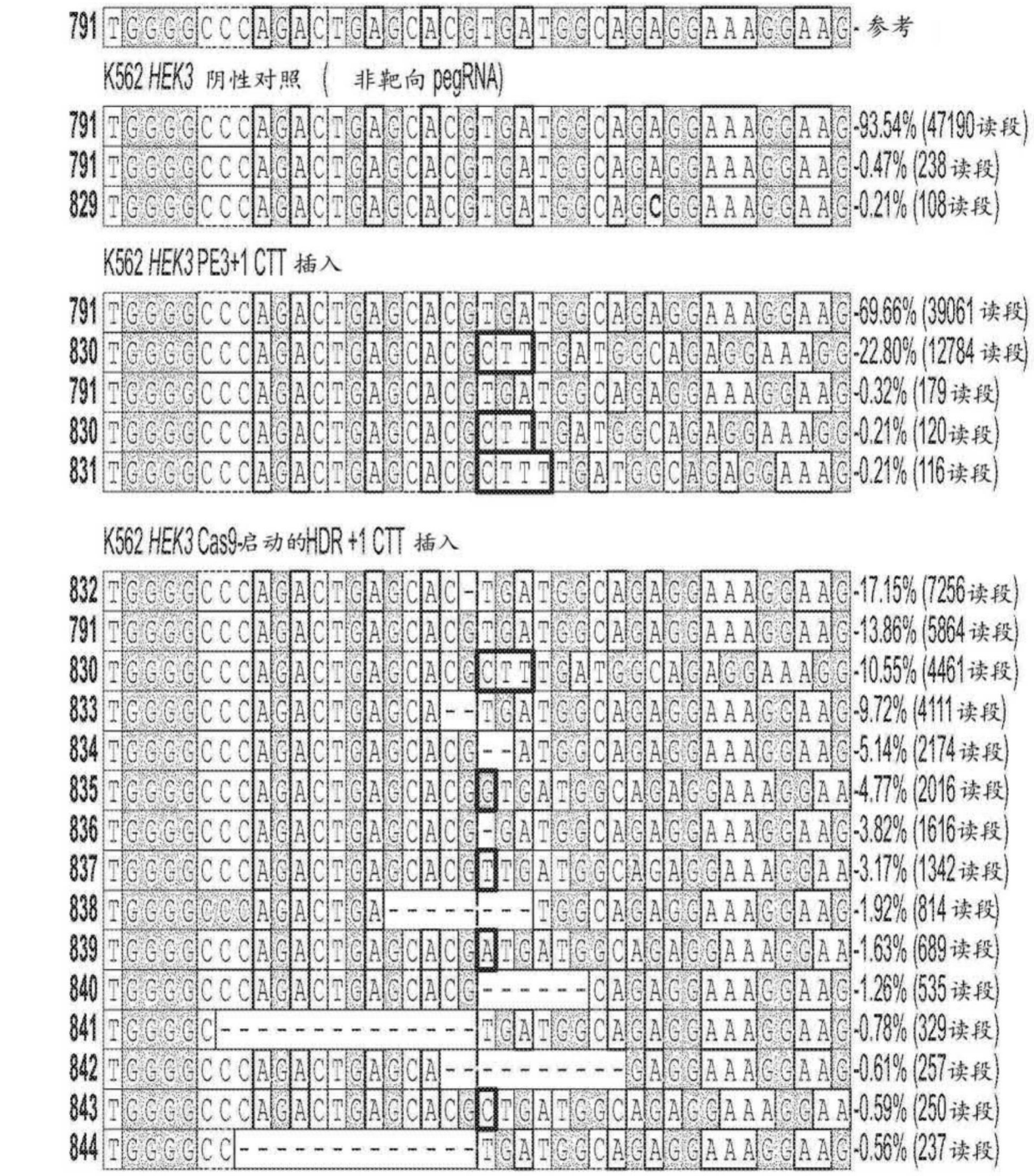

SEQ ID NO:
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG - 参考
K562 HEK3 阴性对照 (非靶向 pegRNA)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -93.54% (47190读段)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -0.47% (238读段)
829 TGGGGCCCAGACTGAGCACGTGATGGCAGCGGAAAGGAAG -0.21% (108读段)
K562 HEK3 PE3+1 CTT 插入
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -69.66% (39061读段)
830 TGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGG -22.80% (12784读段)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -0.32% (179读段)
830 TGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGG -0.21% (120读段)
831 TGGGGCCCAGACTGAGCACGCTTTTGATGGCAGAGGAAAG -0.21% (116读段)
K562 HEK3 Cas9-启动的HDR +1 CTT 插入
832 TGGGGCCCAGACTGAGCAC-TGATGGCAGAGGAAAGGAAG -17.15% (7256读段)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -13.86% (5864读段)
830 TGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGG -10.55% (4461读段)
833 TGGGGCCCAGACTGAGCA--TGATGGCAGAGGAAAGGAAG -9.72% (4111读段)
834 TGGGGCCCAGACTGAGCACG--ATGGCAGAGGAAAGGAAG -5.14% (2174读段)
835 TGGGGCCCAGACTGAGCACGGTGATGGCAGAGGAAAGGAA -4.77% (2016读段)
836 TGGGGCCCAGACTGAGCACG-GATGGCAGAGGAAAGGAAG -3.82% (1616读段)
837 TGGGGCCCAGACTGAGCACGTTGATGGCAGAGGAAAGGAA -3.17% (1342读段)
838 TGGGGCCCAGACTGA-----TGGCAGAGGAAAGGAAG -1.92% (814读段)
839 TGGGGCCCAGACTGAGCACGATGATGGCAGAGGAAAGGAA -1.63% (689读段)
840 TGGGGCCCAGACTGAGCACG------CAGAGGAAAGGAAG -1.26% (535读段)
841 TGGGGC----------TGATGGCAGAGGAAAGGAAG -0.78% (329读段)
842 TGGGGCCCAGACTGAGCA--------GAGGAAAGGAAG -0.61% (257读段)
843 TGGGGCCCAGACTGAGCACGCTGATGGCAGAGGAAA -0.59% (250读段)
844 TGGGGCC----------TGATGGCAGAGGAAAGGAAG -0.56% (237读段)

图51F

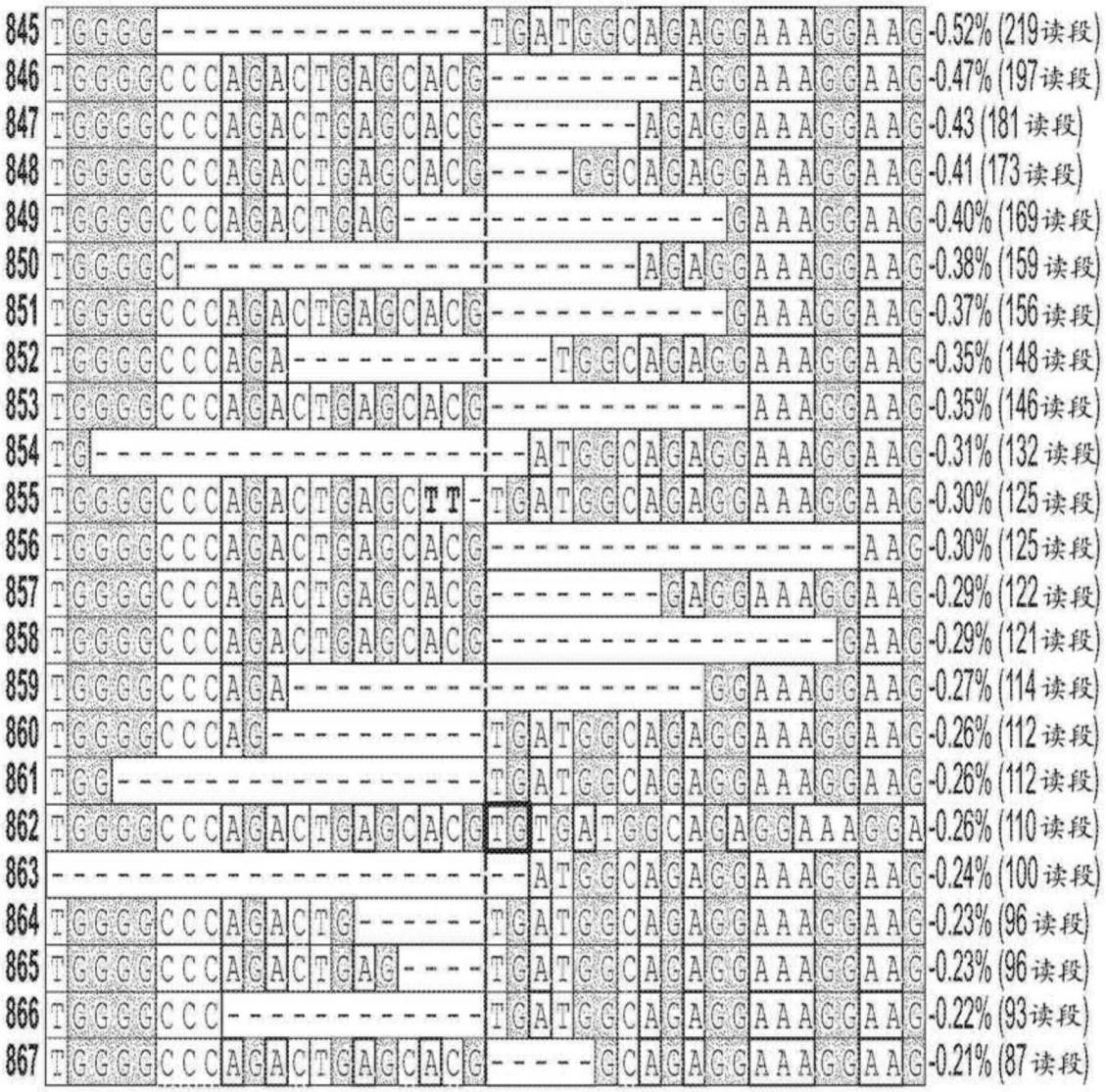

845 -0.52% (219读段)
846 -0.47% (197读段)
847 -0.43 (181 读段)
848 -0.41 (173读段)
849 -0.40% (169读段)
850 -0.38% (159 读段)
851 -0.37% (156读段)
852 -0.35% (148读段)
853 -0.35% (146读段)
854 -0.31% (132 读段)
855 -0.30% (125读段)
856 -0.30% (125读段)
857 -0.29% (122读段)
858 -0.29% (121读段)
859 -0.27% (114 读段)
860 -0.26% (112 读段)
861 -0.26% (112读段)
862 -0.26% (110读段)
863 -0.24% (100 读段)
864 -0.23% (96 读段)
865 -0.23% (96读段)
866 -0.22% (93读段)
867 -0.21% (87 读段)
粗体 替换
插入
- 缺失
---- 预测的
切割位点

图51G

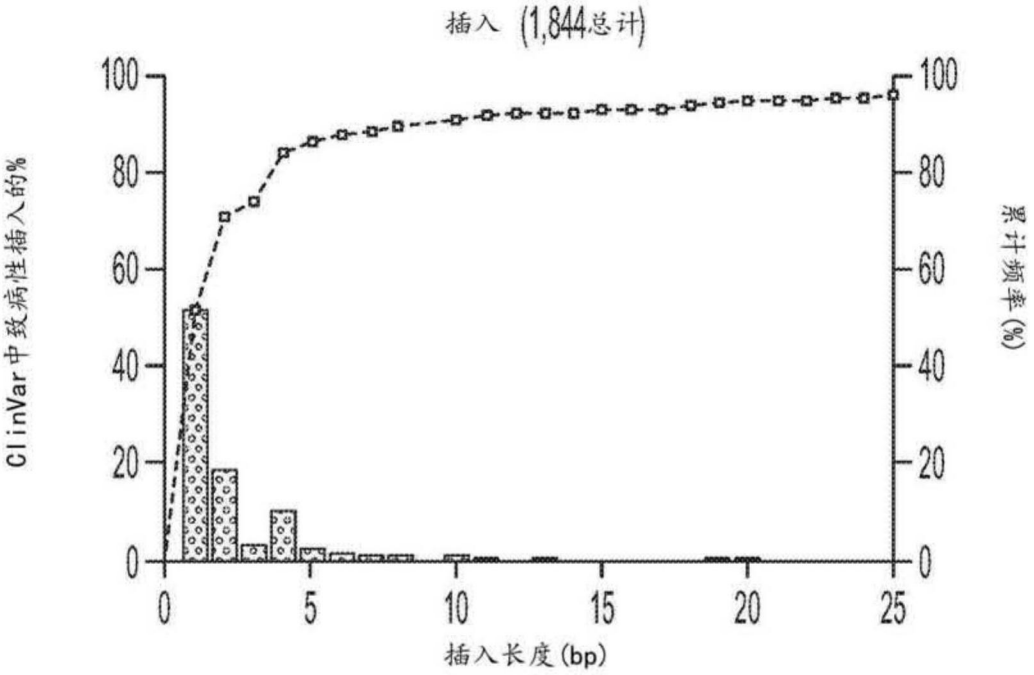

插入 (1,844总计)
ClinVar中致病性插入的%
累计频率(%)
插入长度(bp)
0
20
40
60
80
100
0
5
10
15
20
25

图52A

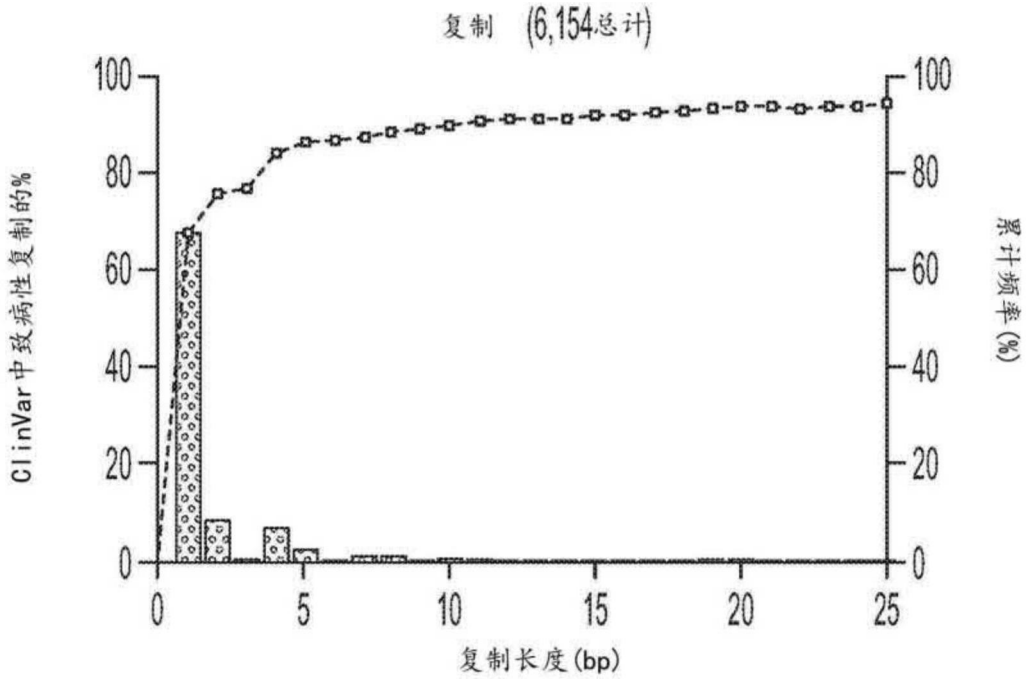

复制 (6,154总计)
ClinVar中致病性复制的%
累计频率(%)
复制长度(bp)
0
20
40
60
80
100
0
5
10
15
20
25

图52B

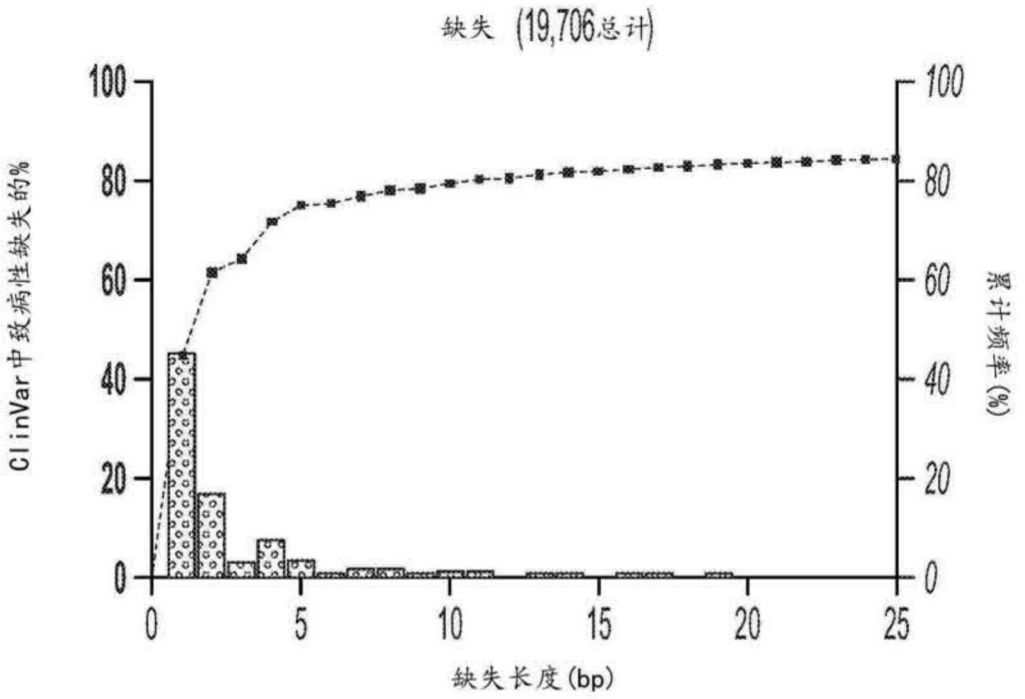

缺失 (19,706总计)
ClinVar中致病性缺失的%
累计频率(%)
缺失长度(bp)
0
5
10
15
20
25
0
20
40
60
80
100

图52C

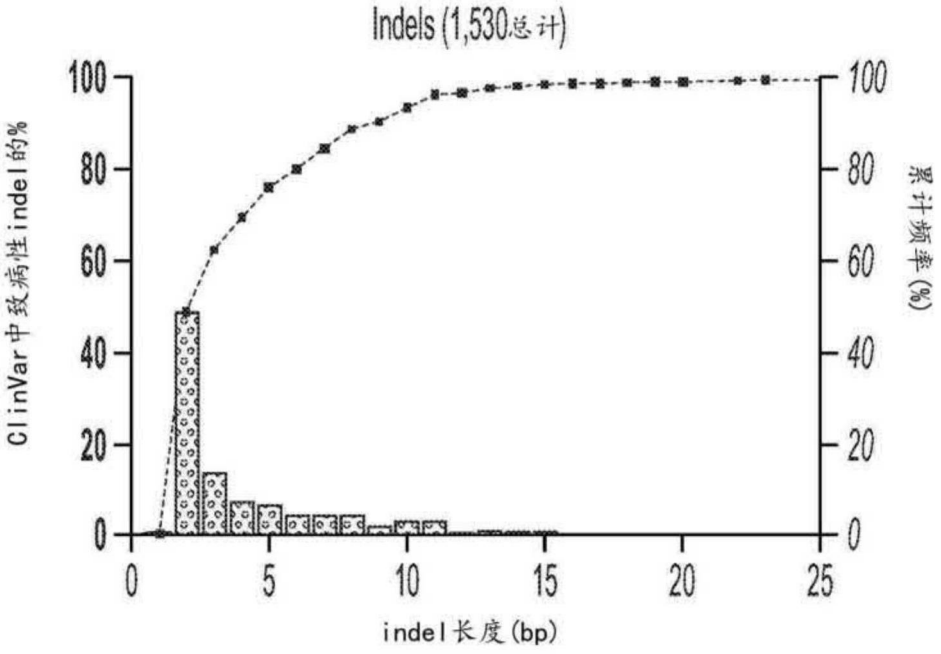

Indels (1,530总计)
ClinVar中致病性indel的%
累计频率(%)
indel长度(bp)
0
5
10
15
20
25
0
20
40
60
80
100

图52D

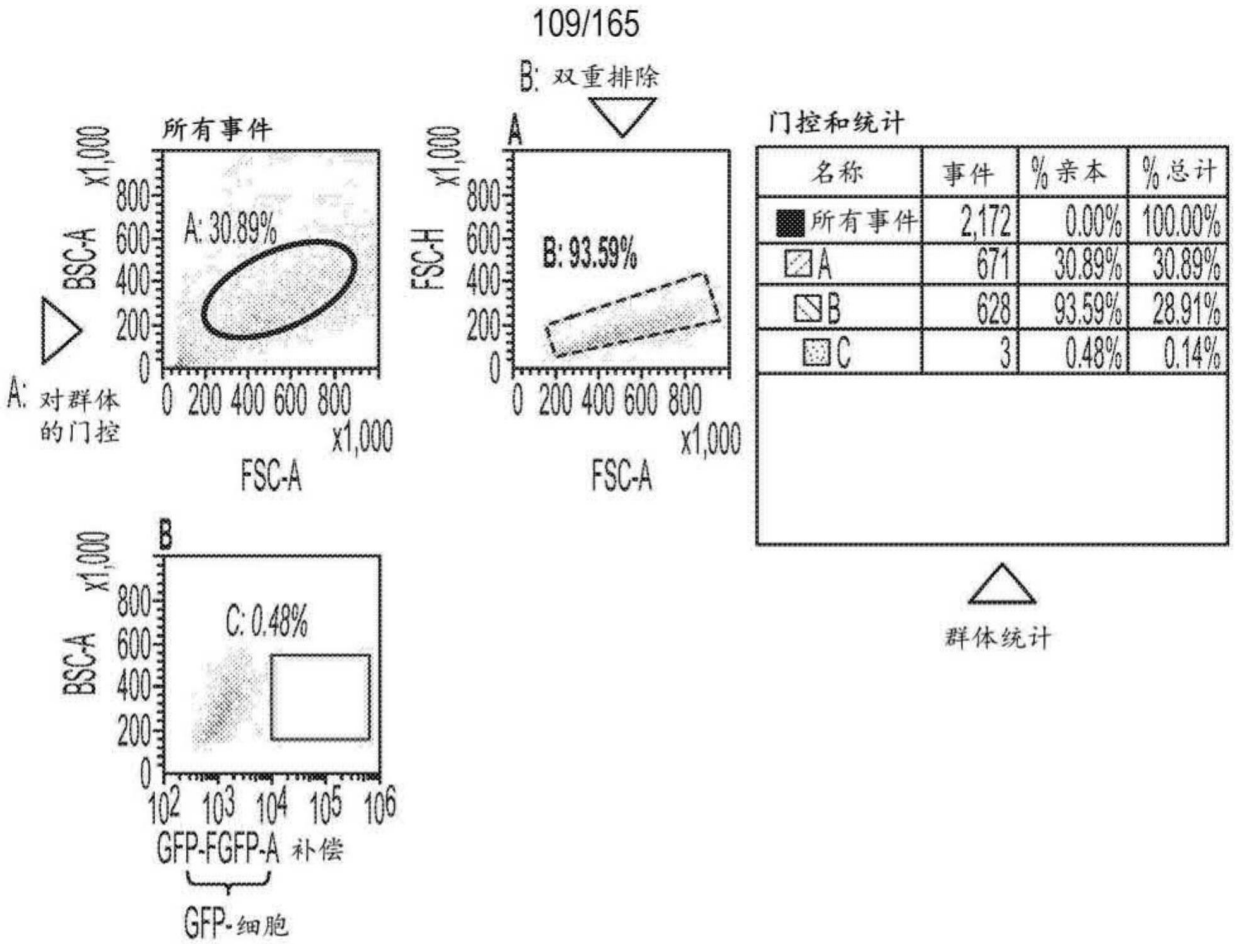

109/165
所有事件
BSC-A
x1,000
A: 30.89%
A: 对群体的门控
FSC-A
x1,000
B: 双重排除
A
FSC-H
B: 93.59%
门控和统计
名称 事件 %亲本 %总计
所有事件 2,172 0.00% 100.00%
A 671 30.89% 30.89%
B 628 93.59% 28.91%
C 3 0.48% 0.14%
群体统计
B
C: 0.48%
GFP-FGFP-A 补偿
GFP-细胞

图53A

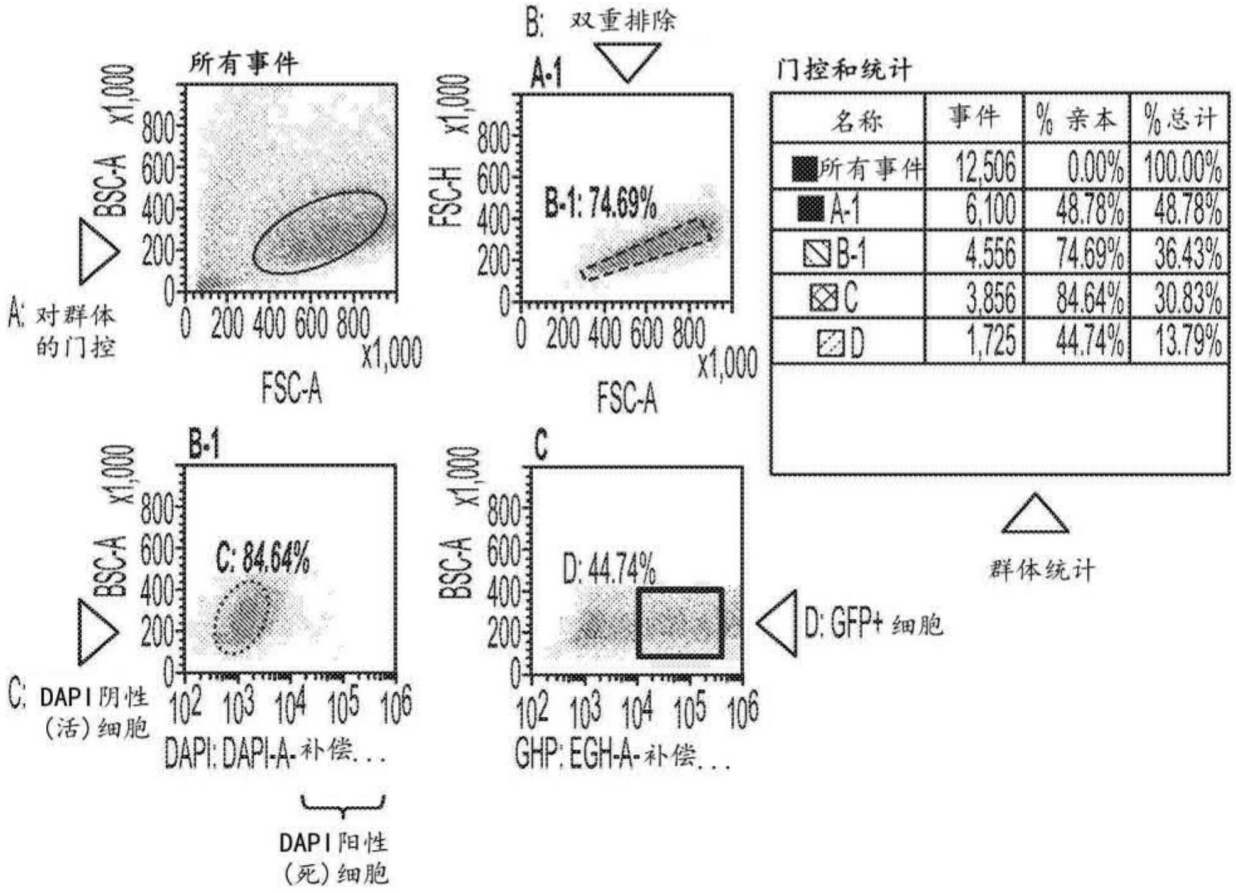

所有事件
BSC-A x1,000
FSC-A x1,000
A: 对群体的门控
B: 双重排除
A-1
FSC-H x1,000
B-1: 74.69%
门控和统计
名称 事件 % 亲本 % 总计
所有事件 12,506 0.00% 100.00%
A-1 6,100 48.78% 48.78%
B-1 4,556 74.69% 36.43%
C 3,856 84.64% 30.83%
D 1,725 44.74% 13.79%
群体统计
B-1
C: 84.64%
C: DAPI 阴性（活）细胞
DAPI: DAPI-A- 补偿...
DAPI 阳性（死）细胞
C
D: 44.74%
D: GFP+ 细胞
GHP: EGH-A- 补偿...

图53B

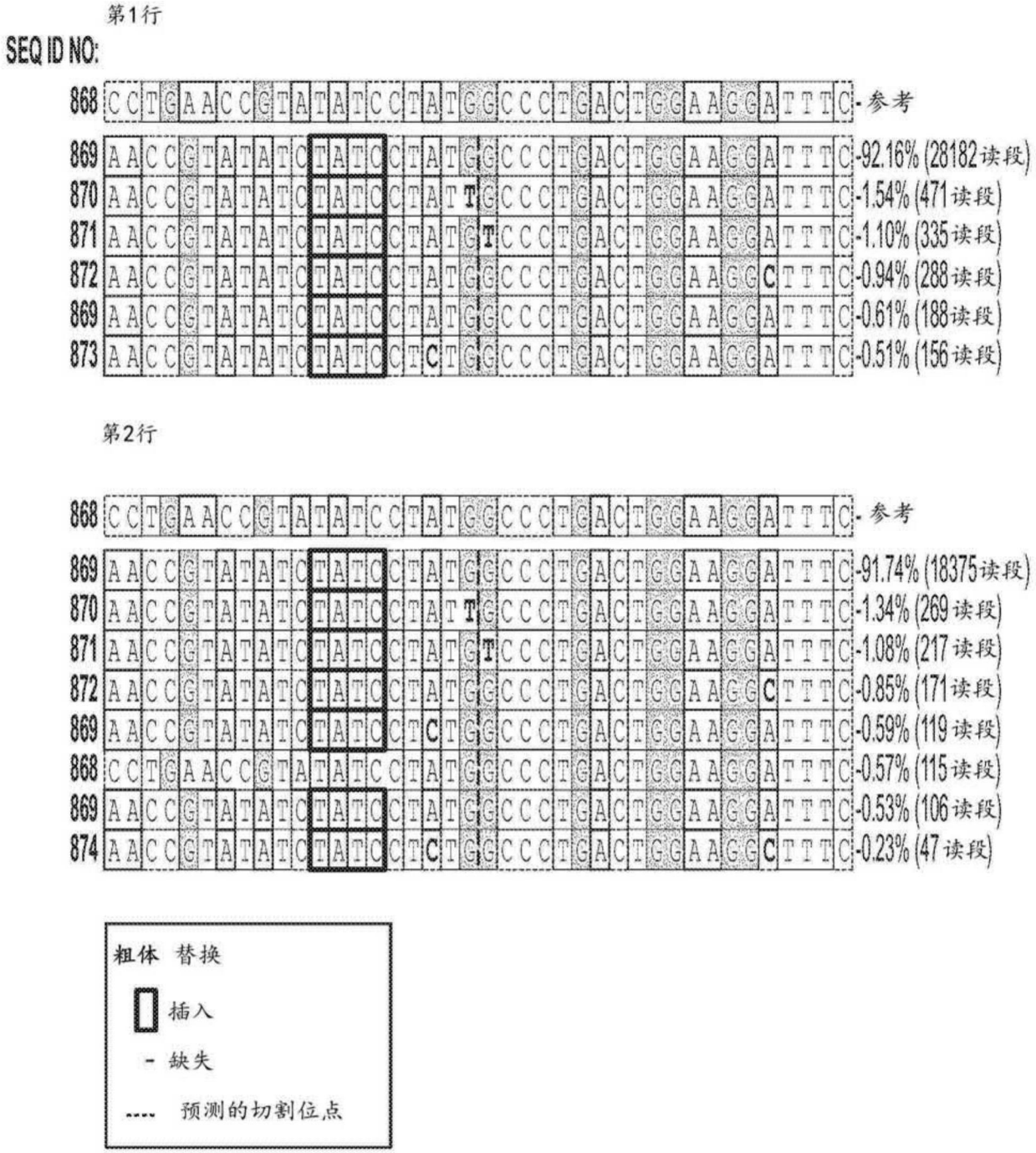

第1行
SEQ ID NO:
868 CCTGAACCGTATATCCTATGGCCCTGACTGGAAGGATTTC - 参考
869 AACCGTATATCTATICCTATGGCCCTGACTGGAAGGATTTC -92.16% (28182读段)
870 AACCGTATATCTATICCTATTGCCCTGACTGGAAGGATTTC -1.54% (471读段)
871 AACCGTATATCTATICCTATGTCCCTGACTGGAAGGATTTC -1.10% (335读段)
872 AACCGTATATCTATICCTATGGCCCTGACTGGAAGGCTTTC -0.94% (288读段)
869 AACCGTATATCTATICCTATGGCCCTGACTGGAAGGATTTC -0.61% (188读段)
873 AACCGTATATCTATICCTCTGGCCCTGACTGGAAGGATTTC -0.51% (156读段)
第2行
868 CCTGAACCGTATATCCTATGGCCCTGACTGGAAGGATTTC - 参考
869 AACCGTATATCTATICCTATGGCCCTGACTGGAAGGATTTC -91.74% (18375读段)
870 AACCGTATATCTATICCTATTGCCCTGACTGGAAGGATTTC -1.34% (269读段)
871 AACCGTATATCTATICCTATGTCCCTGACTGGAAGGATTTC -1.08% (217读段)
872 AACCGTATATCTATICCTATGGCCCTGACTGGAAGGCTTTC -0.85% (171读段)
869 AACCGTATATCTATICCTCTGGCCCTGACTGGAAGGATTTC -0.59% (119读段)
868 CCTGAACCGTATATCCTATGGCCCTGACTGGAAGGATTTC -0.57% (115读段)
869 AACCGTATATCTATICCTATGGCCCTGACTGGAAGGATTTC -0.53% (106读段)
874 AACCGTATATCTATICCTCTGGCCCTGACTGGAAGGCTTTC -0.23% (47读段)
粗体 替换
插入
- 缺失
---- 预测的切割位点

图53C

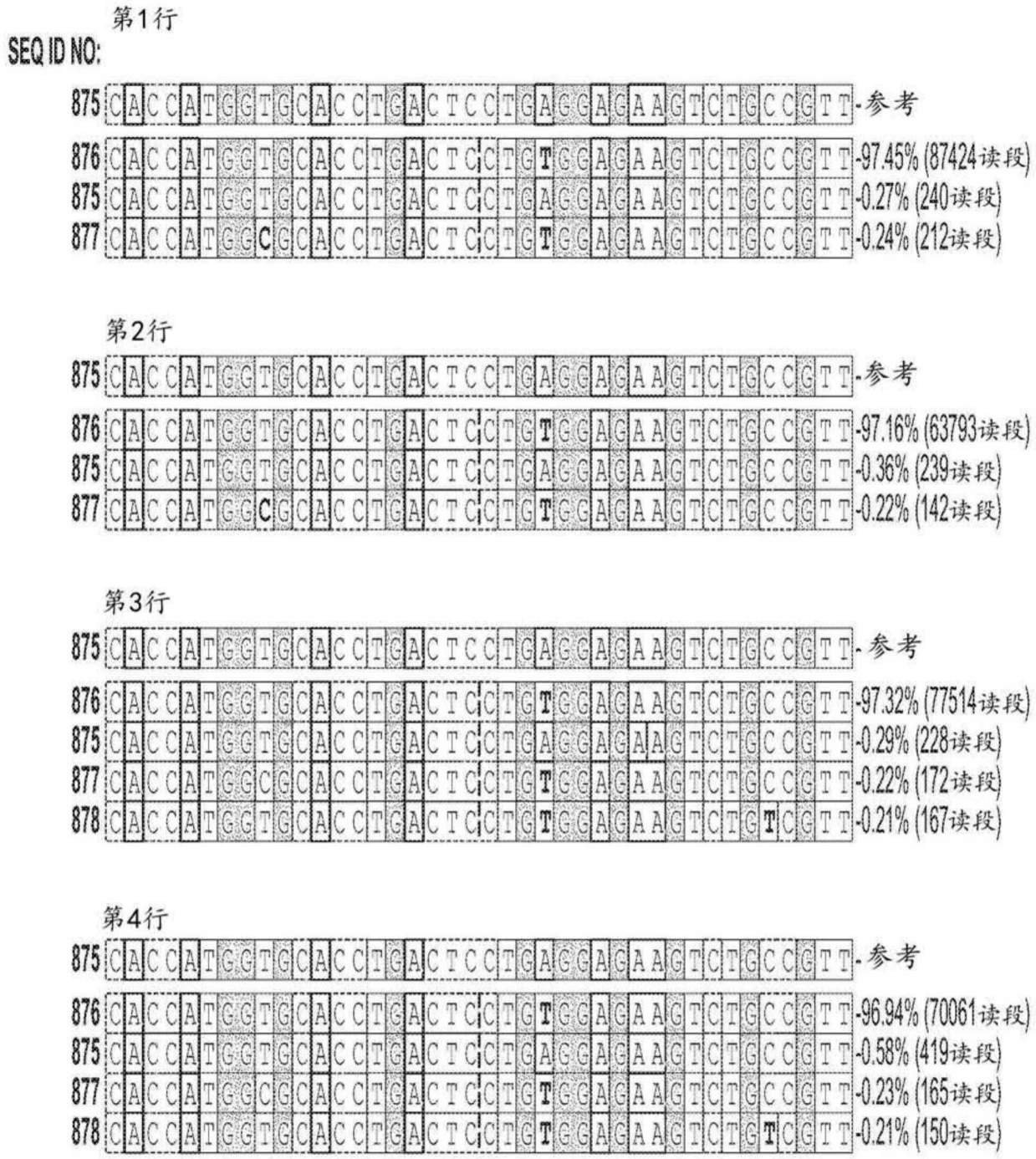

第1行
SEQ ID NO:
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -参考
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.45% (87424读段)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.27% (240读段)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.24% (212读段)
第2行
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -参考
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.16% (63793读段)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.36% (239读段)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.22% (142读段)
第3行
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -参考
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.32% (77514读段)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.29% (228读段)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.22% (172读段)
878 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGTCGTT -0.21% (167读段)
第4行
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -参考
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -96.94% (70061读段)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.58% (419读段)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.23% (165读段)
878 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGTCGTT -0.21% (150读段)

图53D

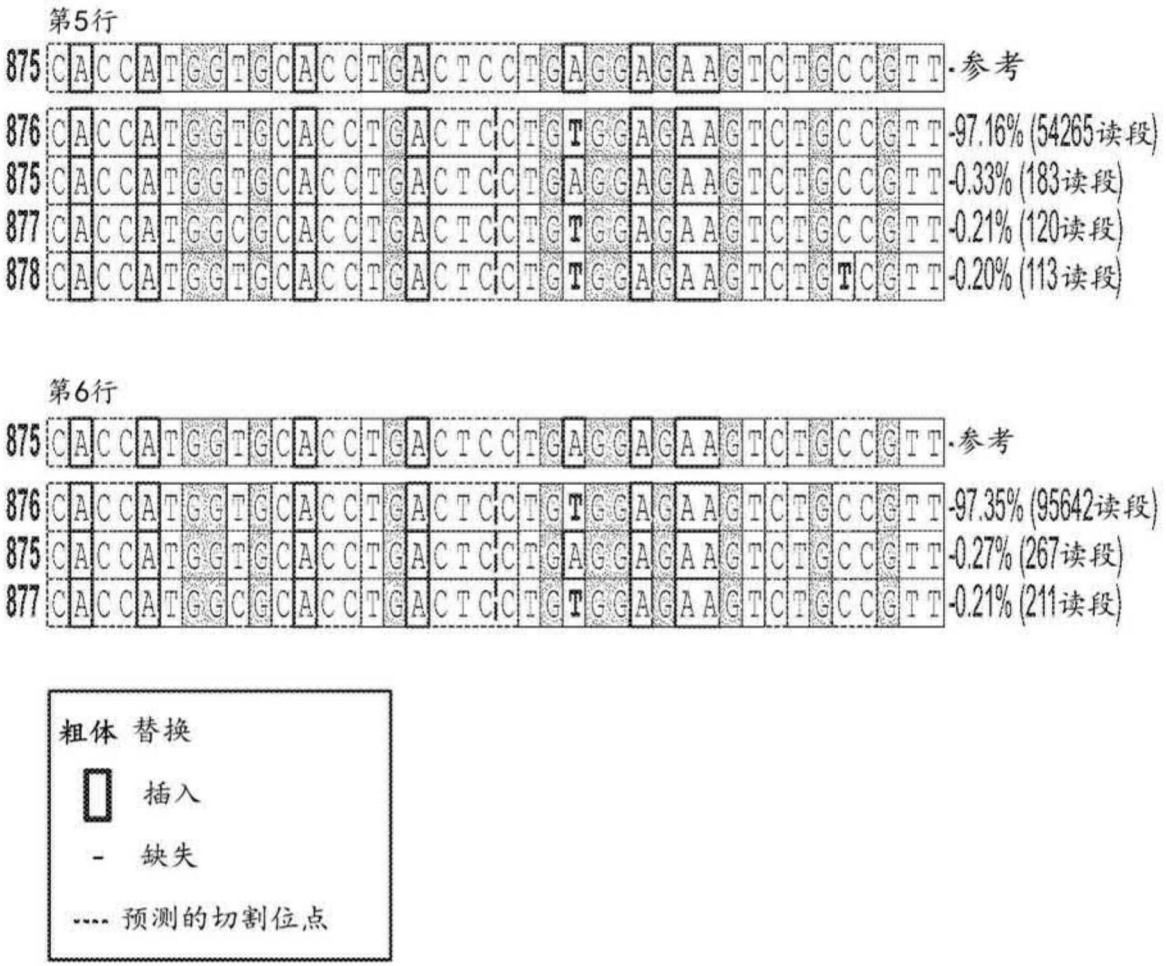
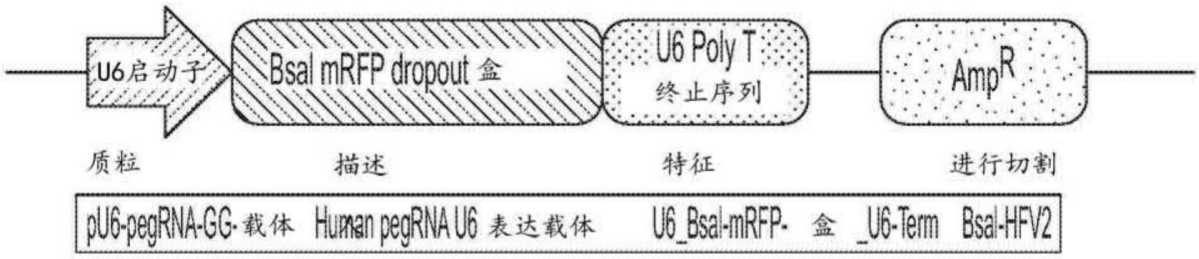

第5行
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -参考
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.16% (54265读段)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.33% (183读段)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.21% (120读段)
878 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGTCGTT -0.20% (113读段)
第6行
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -参考
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.35% (95642读段)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.27% (267读段)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.21% (211读段)
粗体 替换
插入
- 缺失
---- 预测的切割位点

图53E

同时克隆间隔区序列(靶位点)和含有PBS和RT模板的
pegRNA 3’延伸至人U6启动子pegRNA表达载体的方案：

[组件1]未消化的pU6-pegRNA-GG-载体示意图：



[组件2]排序寡核苷酸产生人pegRNA载体的间隔区(靶)序列

上面和下面寡核苷酸退火并组装至消化的pU6-pegRNA-GG-载体

SEQ ID NO: 879 5′-CACCNNNNNNNNNNNNNNNNNNNNGTTTT -3′

SEQ ID NO: 880 3′- NNNNNNNNNNNNNNNNNNNNCAAAATCTC-5′

(N20模拟前间区序列显示可具有按需要改变的长度)

[组件3]排序寡核苷酸产生人pegRNA载体的spCas9 sgRNA支架

上面和下面寡核苷酸退火并组装至消化的pU6-pegRNA-GG-载体

{Phos}5′-AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG -3′ SEQ ID NO: 881

SEQ ID NO: 882 3′- GATCTTTATCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACG-5′{Phos}

(注意，这些寡核苷酸必须5’磷酸化以允许pegRNA载体组装)

[组件4]排序寡核苷酸产生人pegRNA载体的含有PBS和RT
模板序列的pegRNA 3’延伸：

上面和下面核苷酸退火并组装至消化的pU6-pegRNA-GG-载体

SEQ ID NO: 883 5′-GTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN -3′

SEQ ID NO: 884 3′- NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAA-5′

(N30模拟逆转录酶模板序列显示可具有按需要改变的长度)

由组件1,2,3和4最终组装的U6pegRNA表达载体的示意图

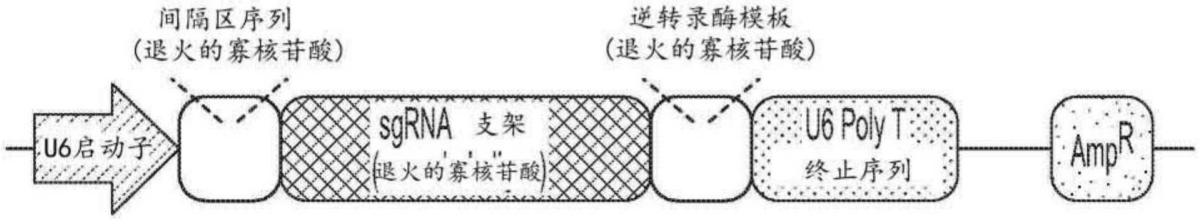


图54

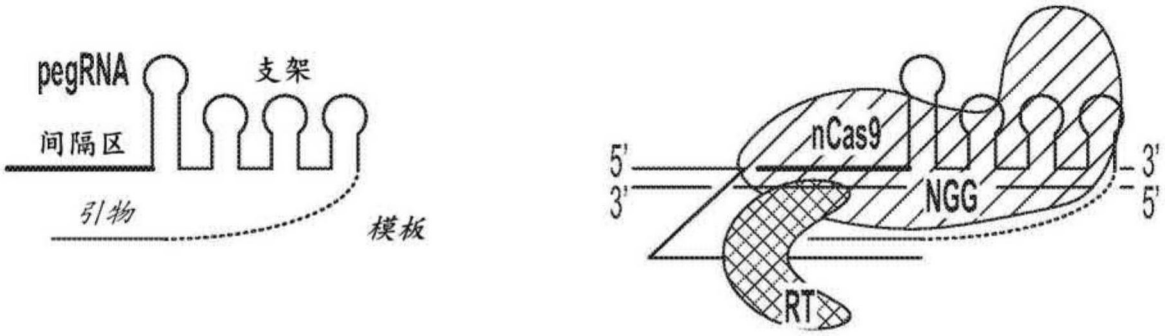


图55A

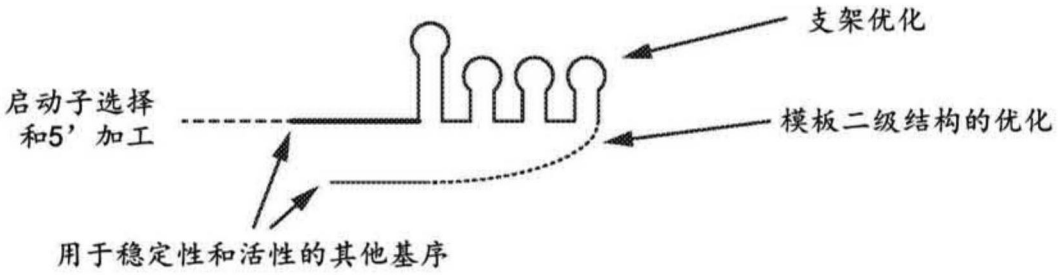


图55B

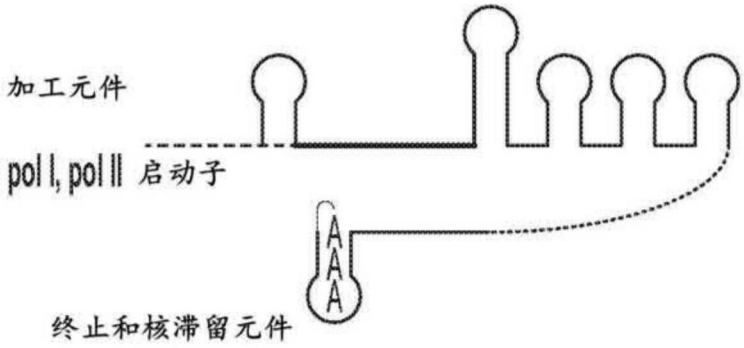


图55C

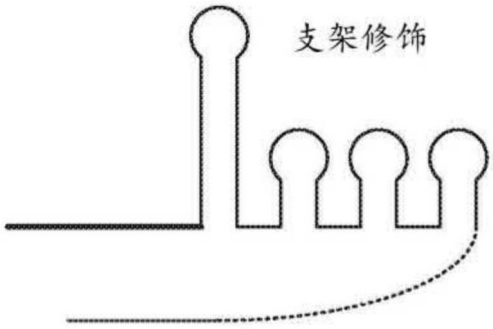


图55D

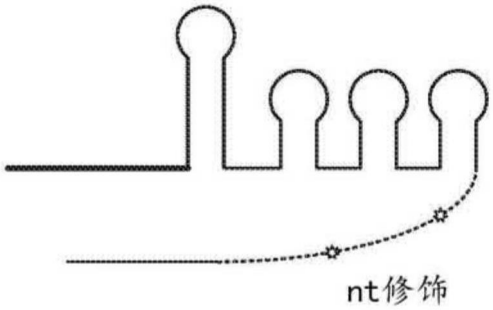


图55E

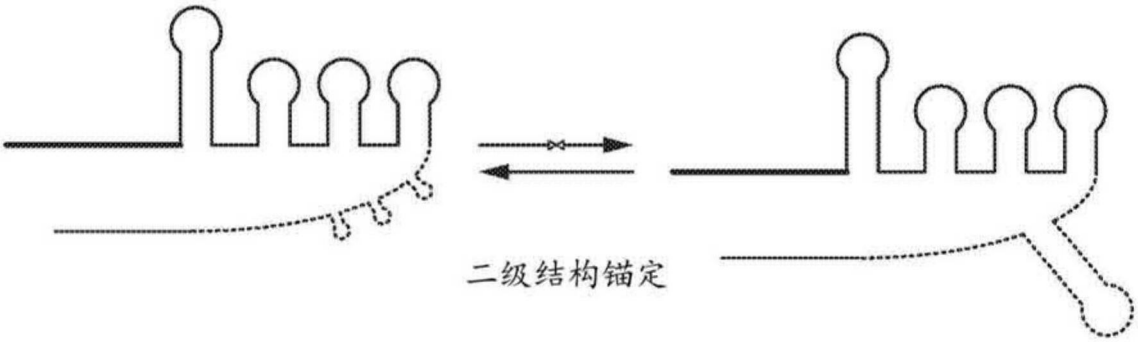


图55F

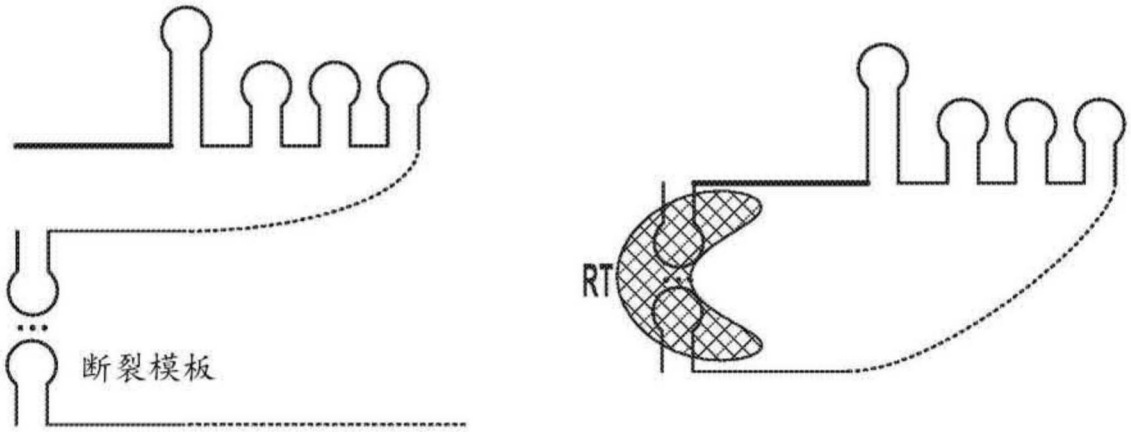
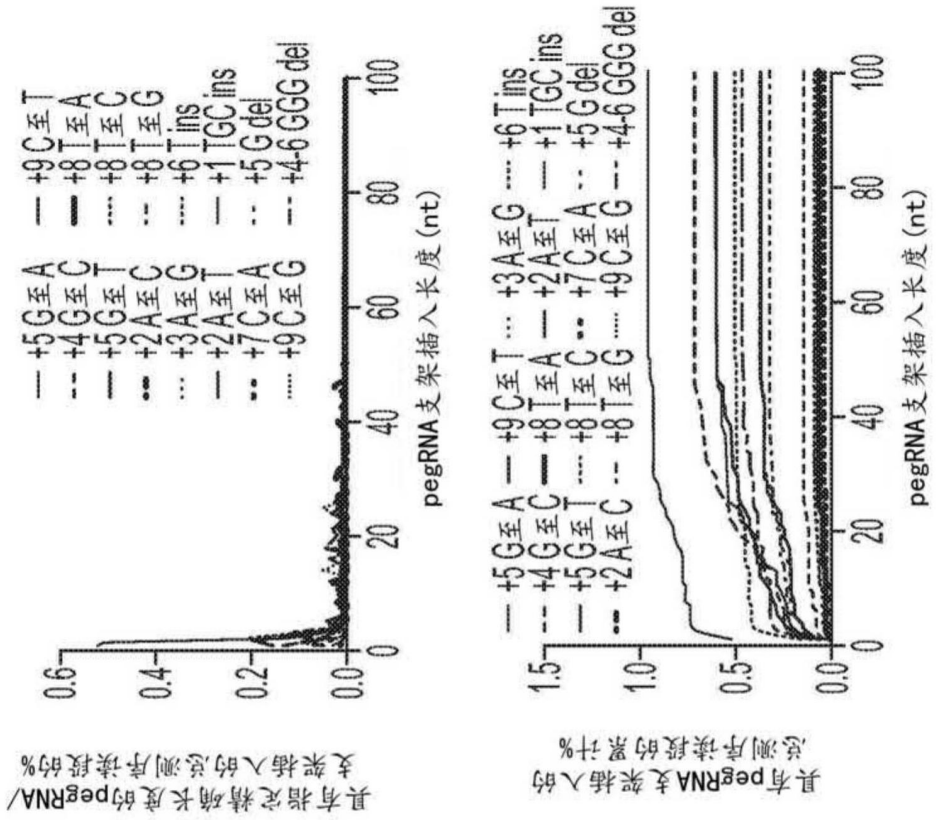


图55G


具有指定精确长度的pegRNA/支架插入的,总测序读段的%
0.6
0.4
0.2
0.0
0
20
40
60
80
100
pegRNA支架插入长度(nt)
+5 G至A
+4 G至C
+5 G至T
+2 A至C
+3 A至G
+2 A至T
+7 C至A
+9 C至G
+9 C至T
+8 T至A
+8 T至C
+8 T至G
+6 T ins
+1 TGC ins
+5 G del
+4-6 GGG del
具有pegRNA支架插入的,总测序读段的累计%
1.5
1.0
0.5
0.0
+5 G至A
+4 G至C
+5 G至T
+2 A至C
+9 C至T
+8 T至A
+8 T至C
+8 T至G
+3 A至G
+6 T ins
+1 TGC ins
+5 G del
+4-6 GGG del
+7 C至A
pegRNA支架插入长度(nt)

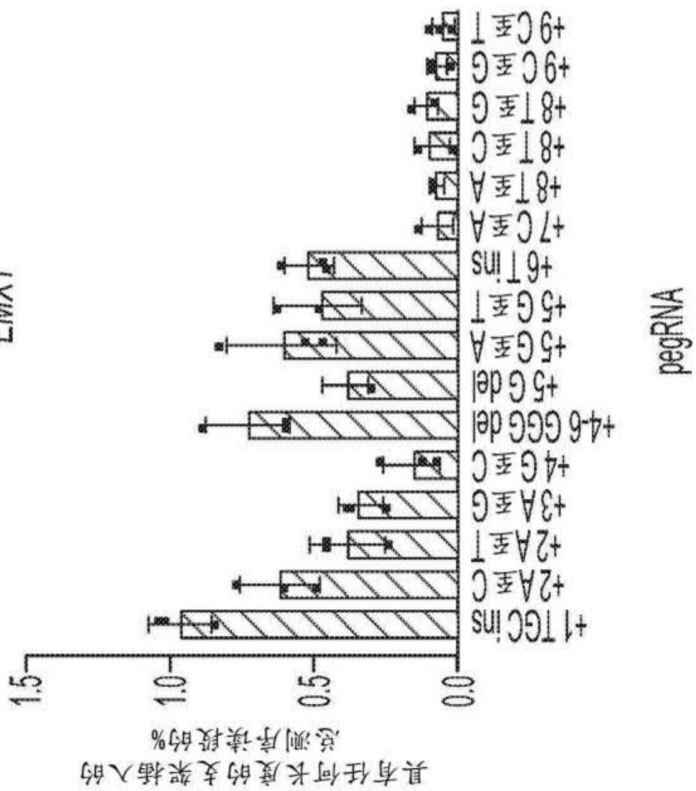
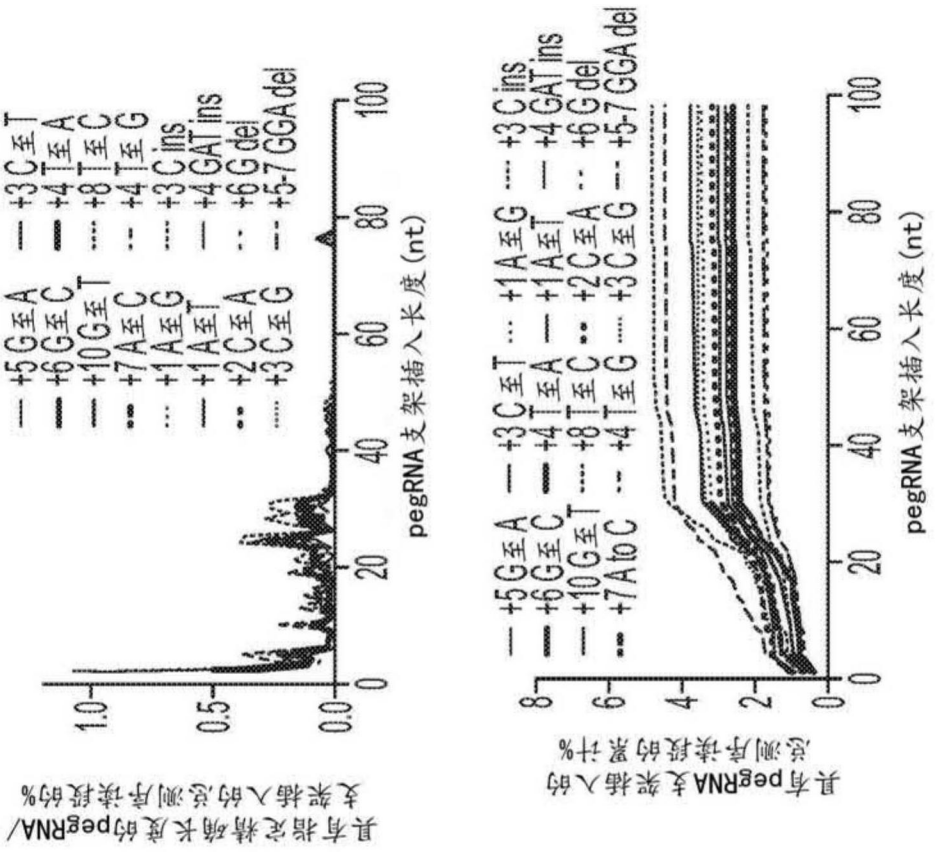

EMX1
具有任何长度的支架插入的,总测序读段的%
1.5
1.0
0.5
0.0
+1 TGC ins
+2 A至C
+2 A至T
+3 A至G
+4 G至C
+4-6 GGG del
+5 G del
+5 G至A
+5 G至T
+6 T ins
+7 C至A
+8 T至A
+8 T至C
+8 T至G
+9 C至G
+9 C至T
pegRNA

图56A


具有指定精确长度的pegRNA/支架插入的,总测序读段的%
pegRNA支架插入长度(nt)
+5 G至A
+6 G至C
+10 G至T
+7 A至C
+1 A至G
+1 A至T
+2 C至A
+3 C至G
+3 C至T
+4 T至A
+8 T至C
+4 T至G
+3 C ins
+4 GAT ins
+6 G del
+5-7 GGA del
具有pegRNA支架插入的总测序读段的累计%
pegRNA支架插入长度(nt)
+1 A至G
+3 C至T
+4 T至A
+8 T至C
+4 T至G
+3 C ins
+1 A至T
+4 GAT ins
+2 C至A
+6 G del
+3 C至G
+5-7 GGA del

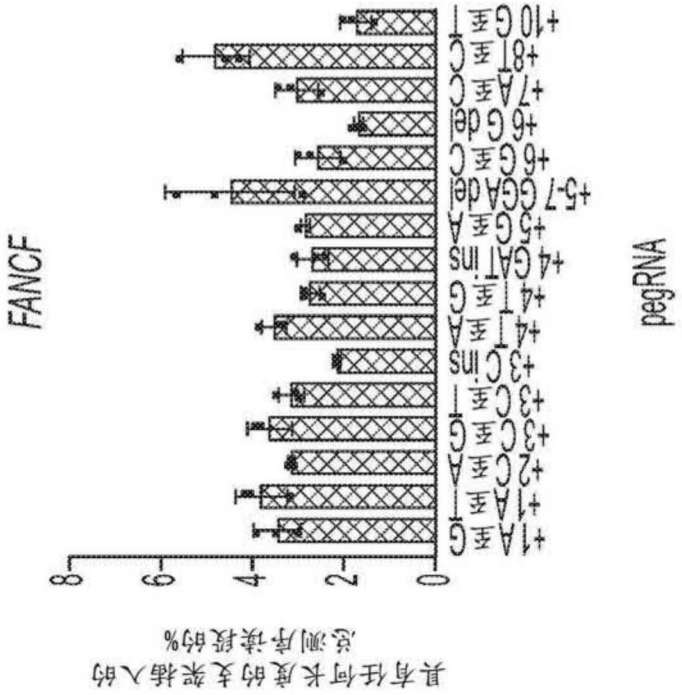

具有任何长度的支架插入的,总测序读段的%
FANCF
pegRNA
+1A至G
+1A至T
+2C至A
+3C至G
+3C至T
+3C ins
+4T至A
+4T至C
+4T至G
+4GAT ins
+5G至A
+5-7 GGA del
+6G至C
+6G del
+7A至C
+8T至C
+10G至T

图56B

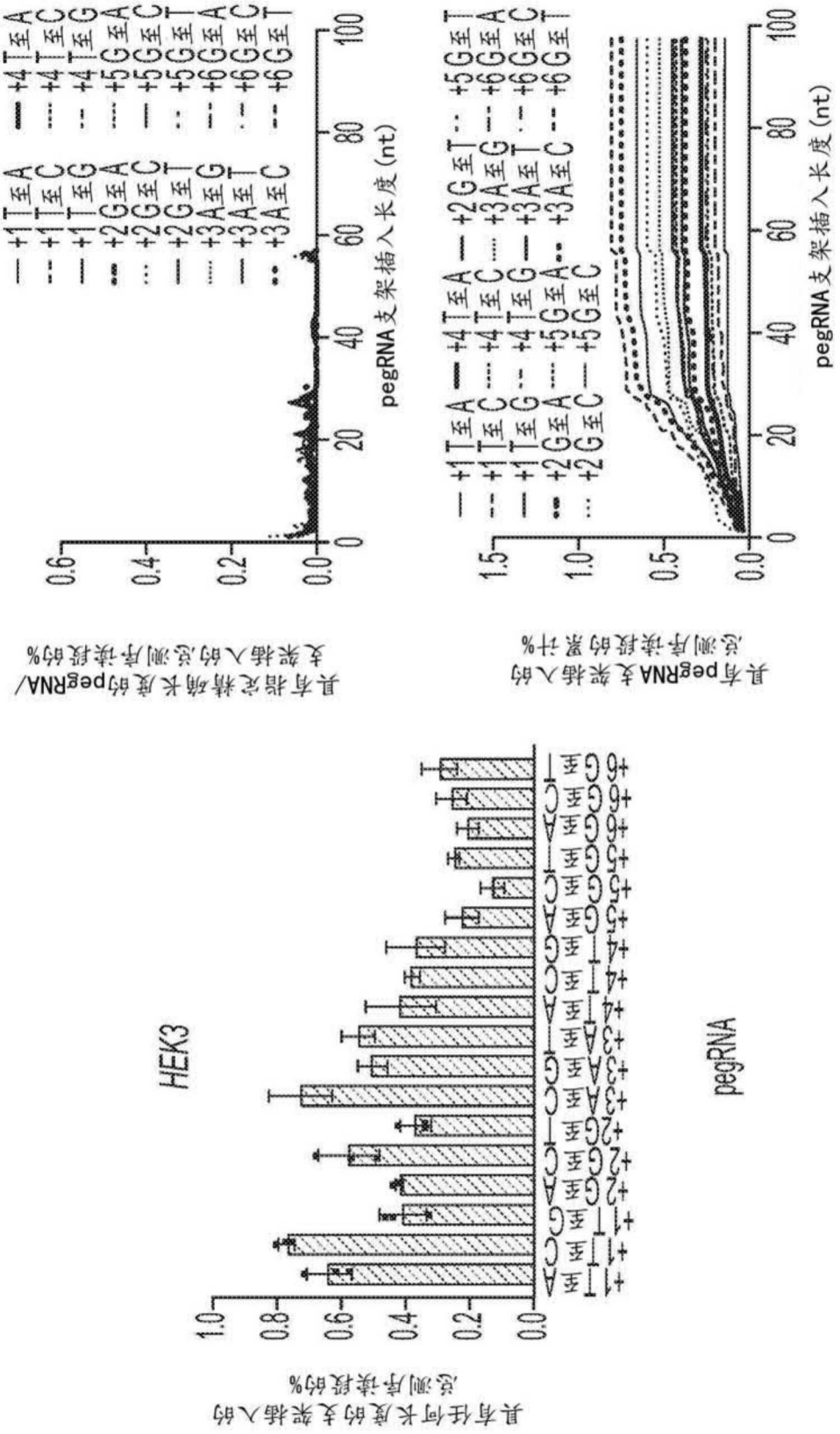
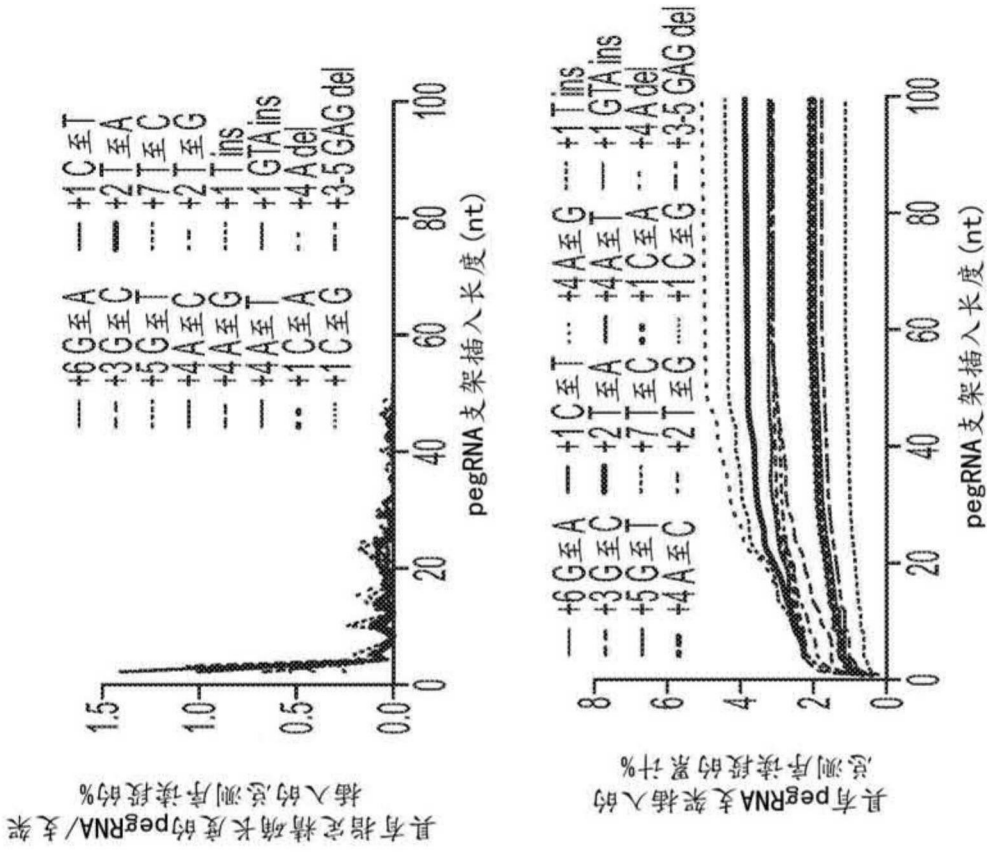

具有任何长度的支架插入的
总测序读段的%
HEK3
pegRNA
具有指定精确长度的pegRNA/
支架插入的总测序读段的%
pegRNA支架插入长度(nt)
具有pegRNA支架插入的
总测序读段的累计%
pegRNA支架插入长度(nt)

图56C


具有指定精确长度的pegRNA/支架插入的总测序读段的%
pegRNA支架插入长度(nt)
+6 G至A
+3 G至C
+5 G至T
+4 A至C
+4 A至G
+4 A至T
+1 C至A
+1 C至G
+1 C至T
+2 T至A
+7 T至C
+2 T至G
+1 T ins
+1 GTA ins
+4 A del
+3-5 GAG del
具有pegRNA支架插入的总测序读段的累计%
pegRNA支架插入长度(nt)

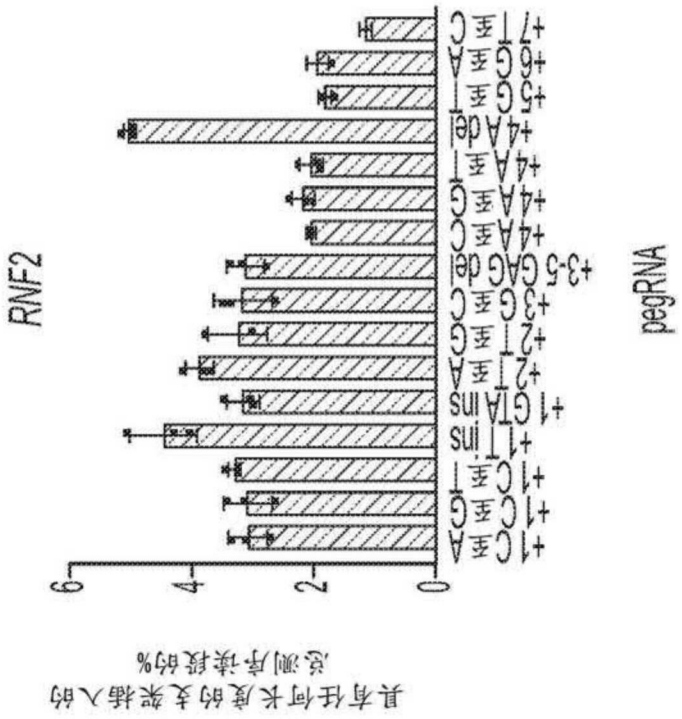

具有任何长度的支架插入的总测序读段的%
RNF2
pegRNA

图56D

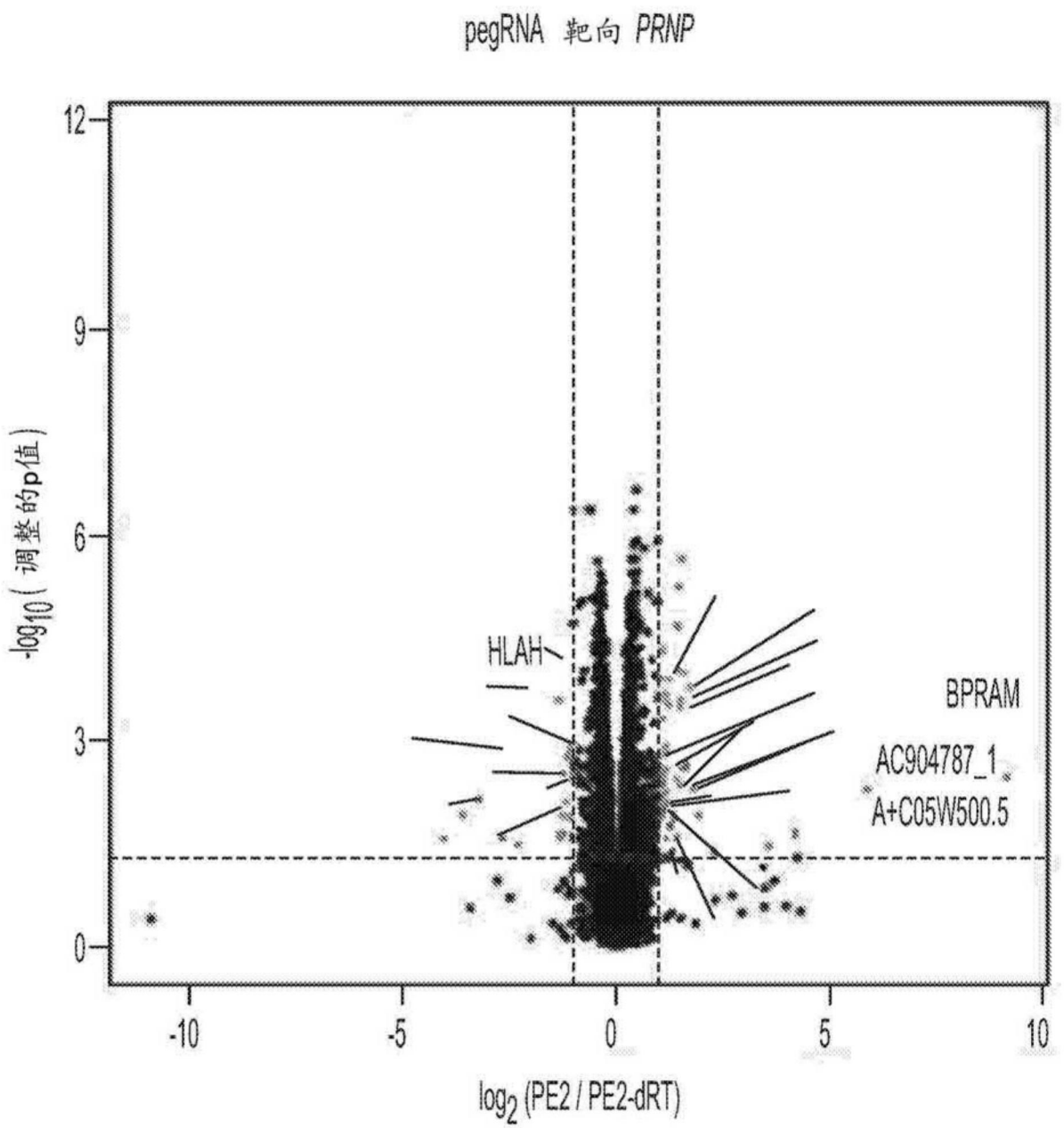

pegRNA 靶向 PRNP
12
9
6
3
0
-log10(调整的p值)
HLAH
BPRAM
AC904787_1
A+C05W500.5
-10
-5
0
5
10
log2 (PE2 / PE2-dRT)

图57A

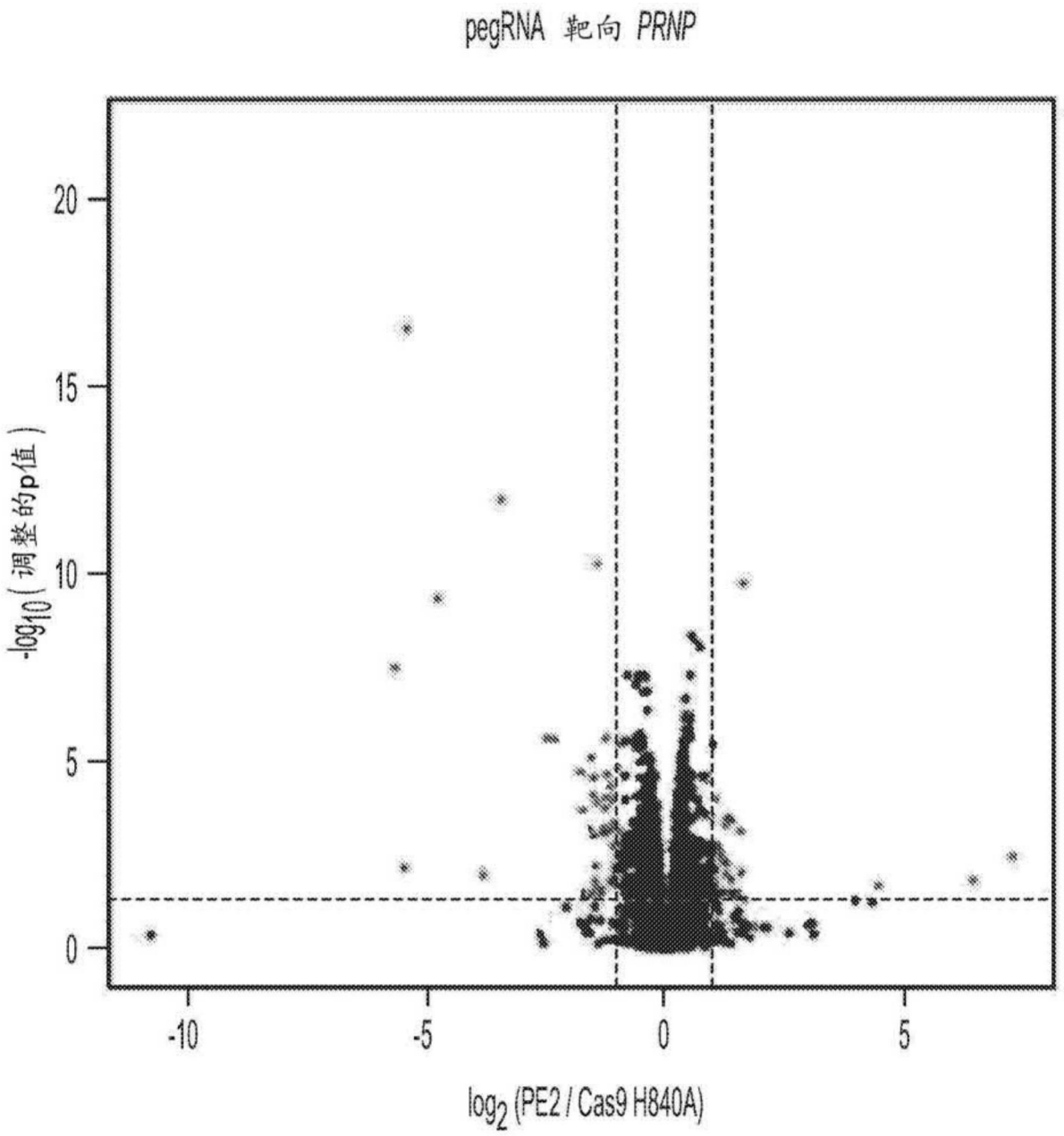

pegRNA 靶向 PRNP
20
15
10
5
0
-10
-5
0
5
-log10(调整的p值)
log2 (PE2 / Cas9 H840A)

图57B

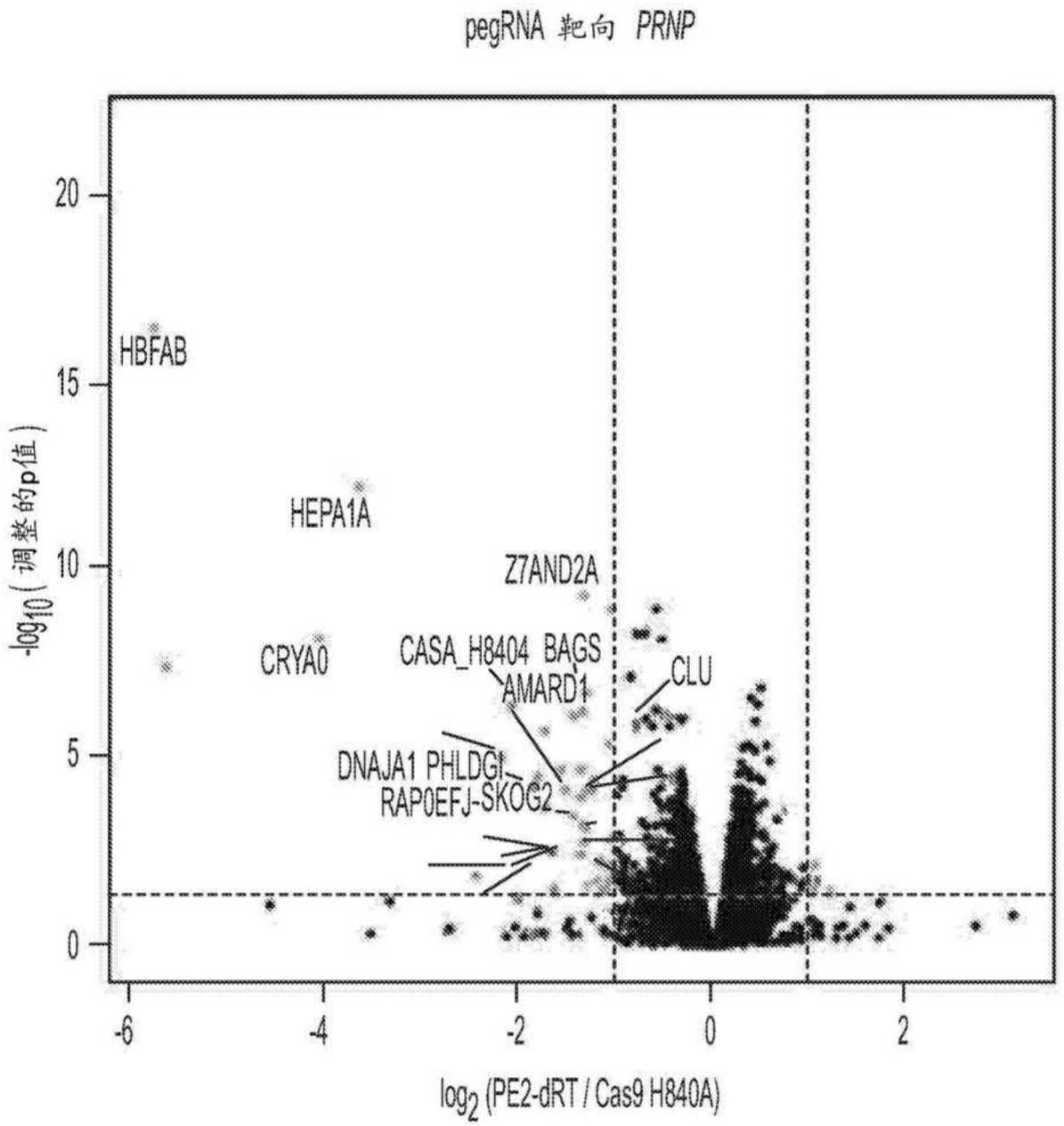

pegRNA 靶向 PRNP
HBFAB
HEPA1A
Z7AND2A
CRYA0
CASA_H8404 BAGS
AMARD1
CLU
DNAJA1 PHLDGI
RAP0EFJ-SKOG2
20
15
10
5
0
-6
-4
-2
0
2
-log10(调整的p值)
log2 (PE2-dRT / Cas9 H840A)

图57C

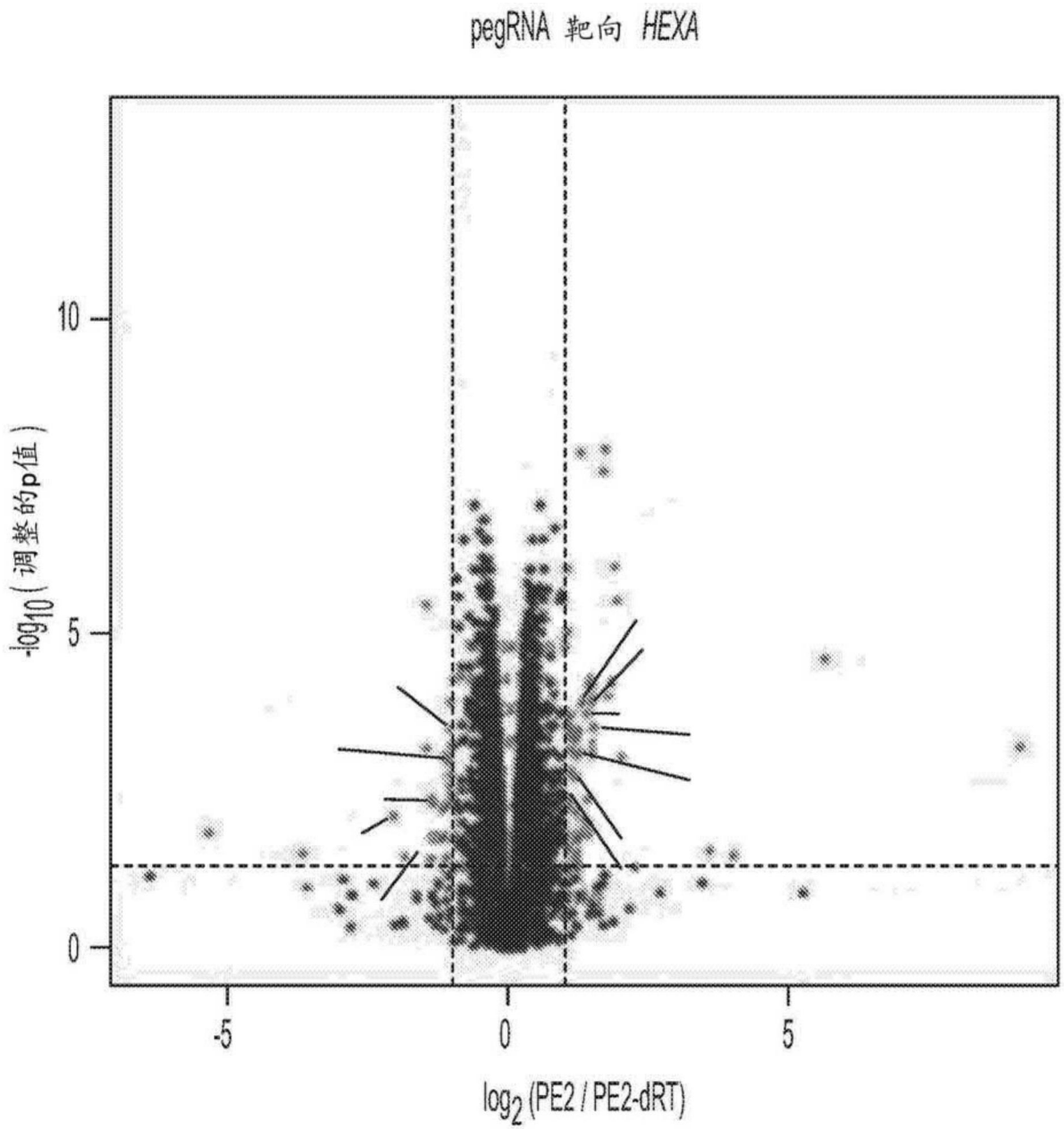


图57D

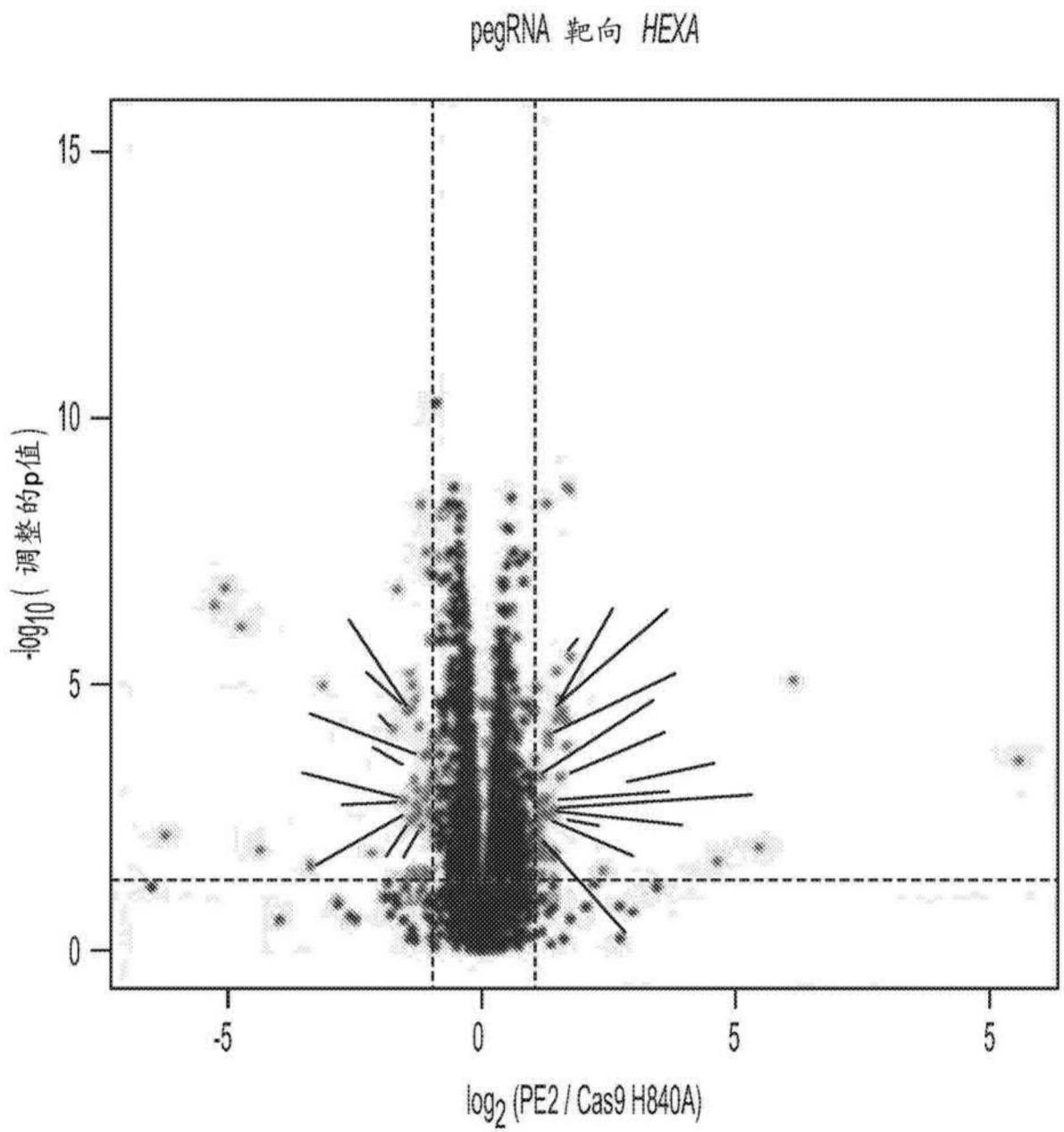


图57E

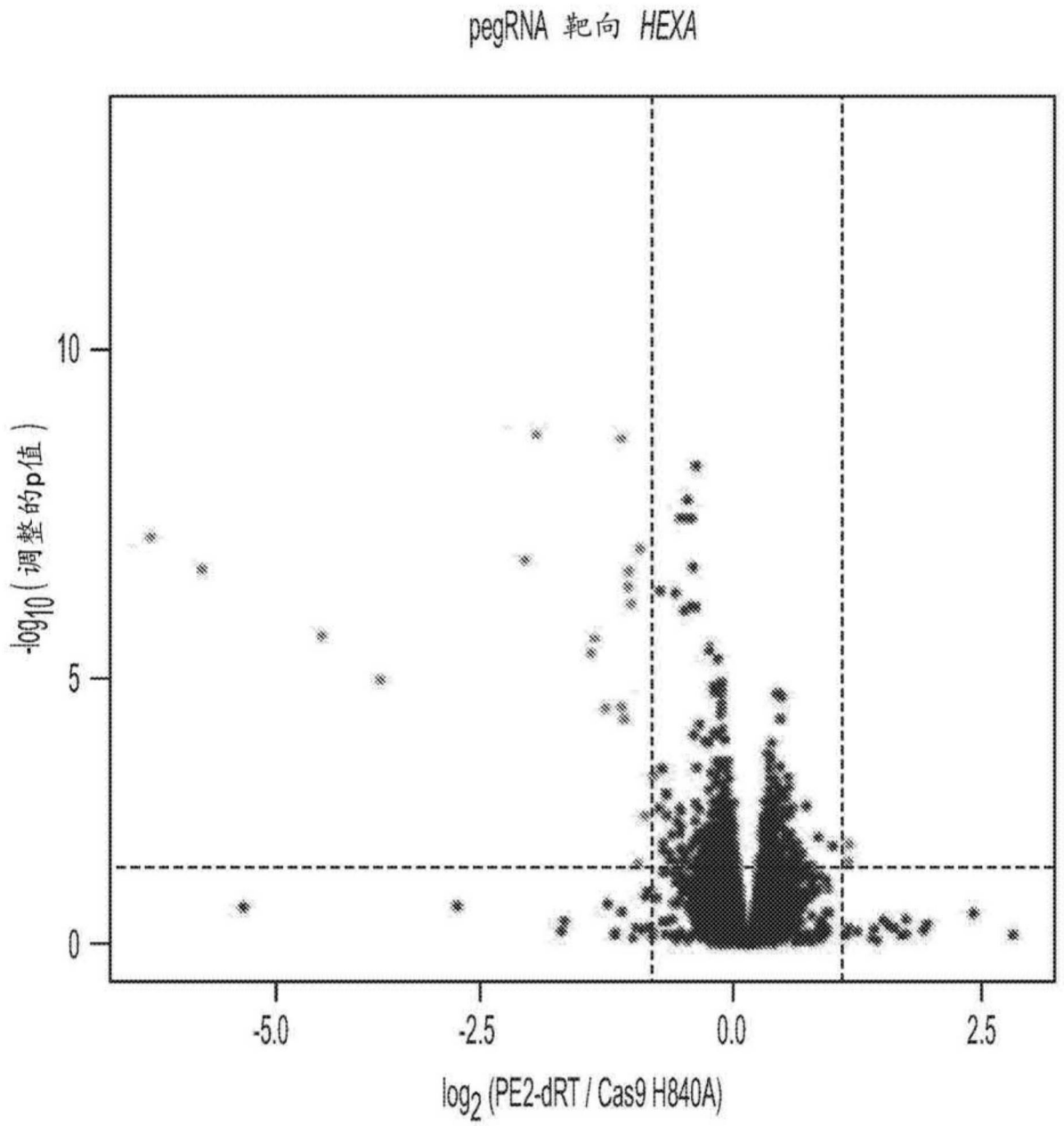

pegRNA 靶向 HEXA
10
5
0
-log10(调整的p值)
-5.0
-2.5
0.0
2.5
log2 (PE2-dRT / Cas9 H840A)

图57F

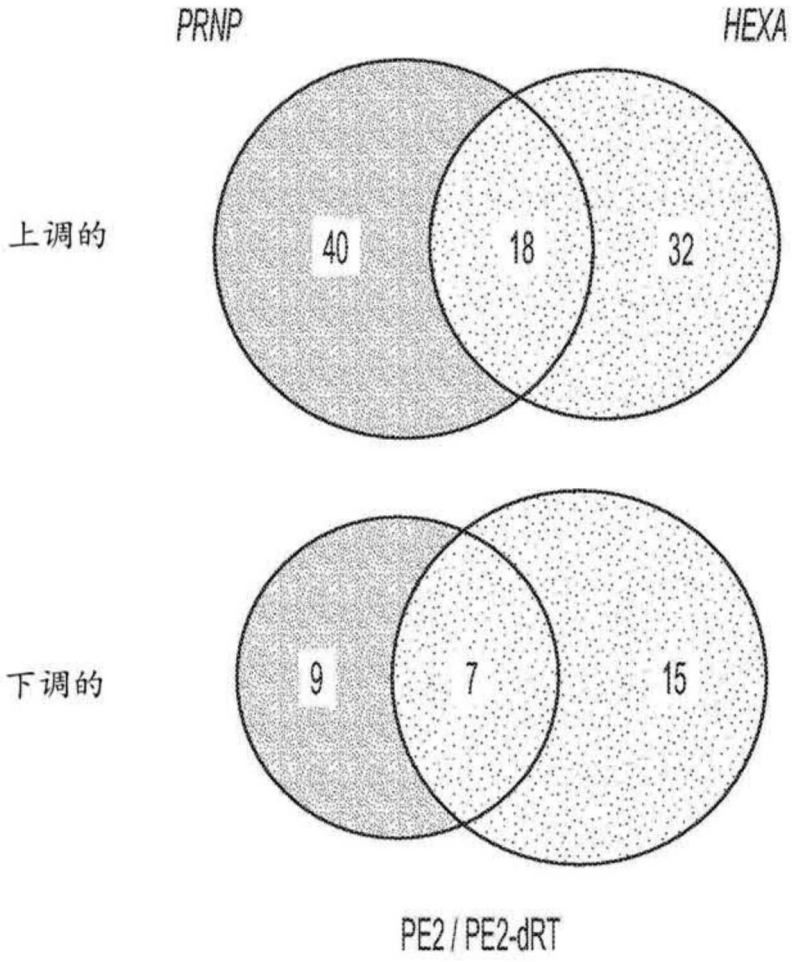


图57G

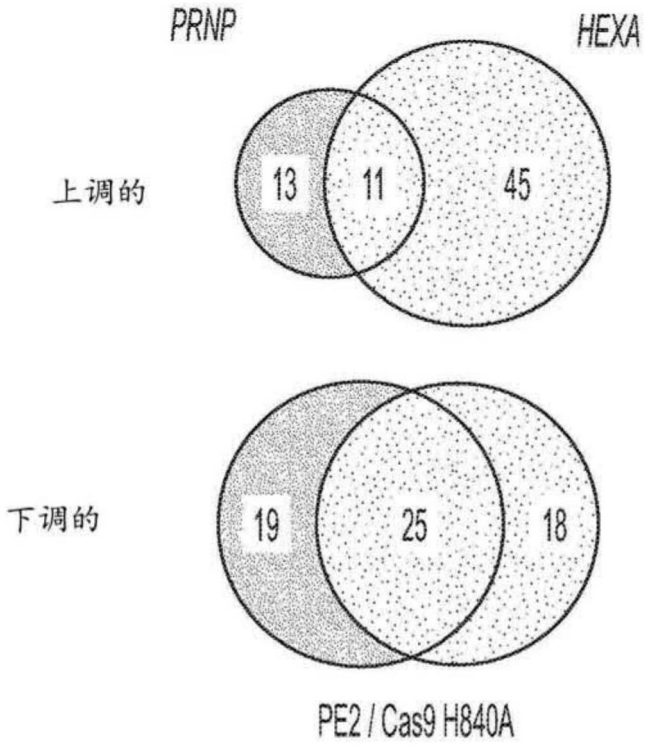

PRNP
HEXA
上调的
13
11
45
下调的
19
25
18
PE2 / Cas9 H840A

图57H

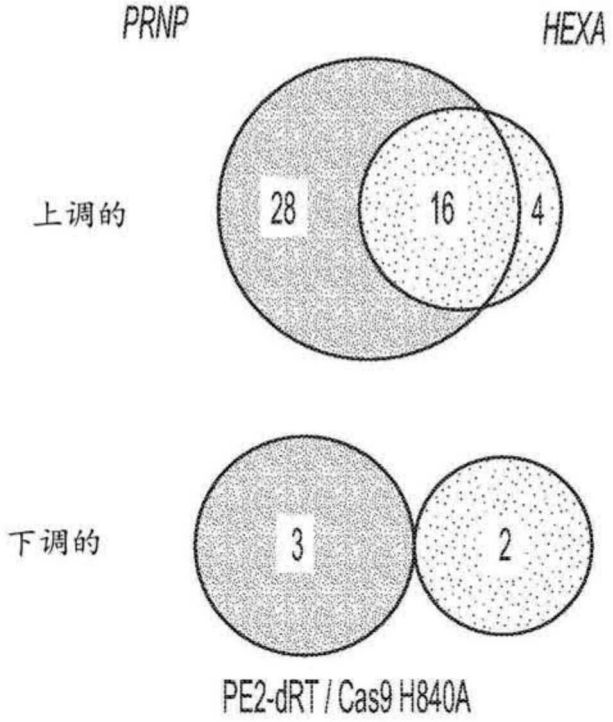

PRNP
HEXA
上调的
28
16
4
下调的
3
2
PE2-dRT / Cas9 H840A

图57I

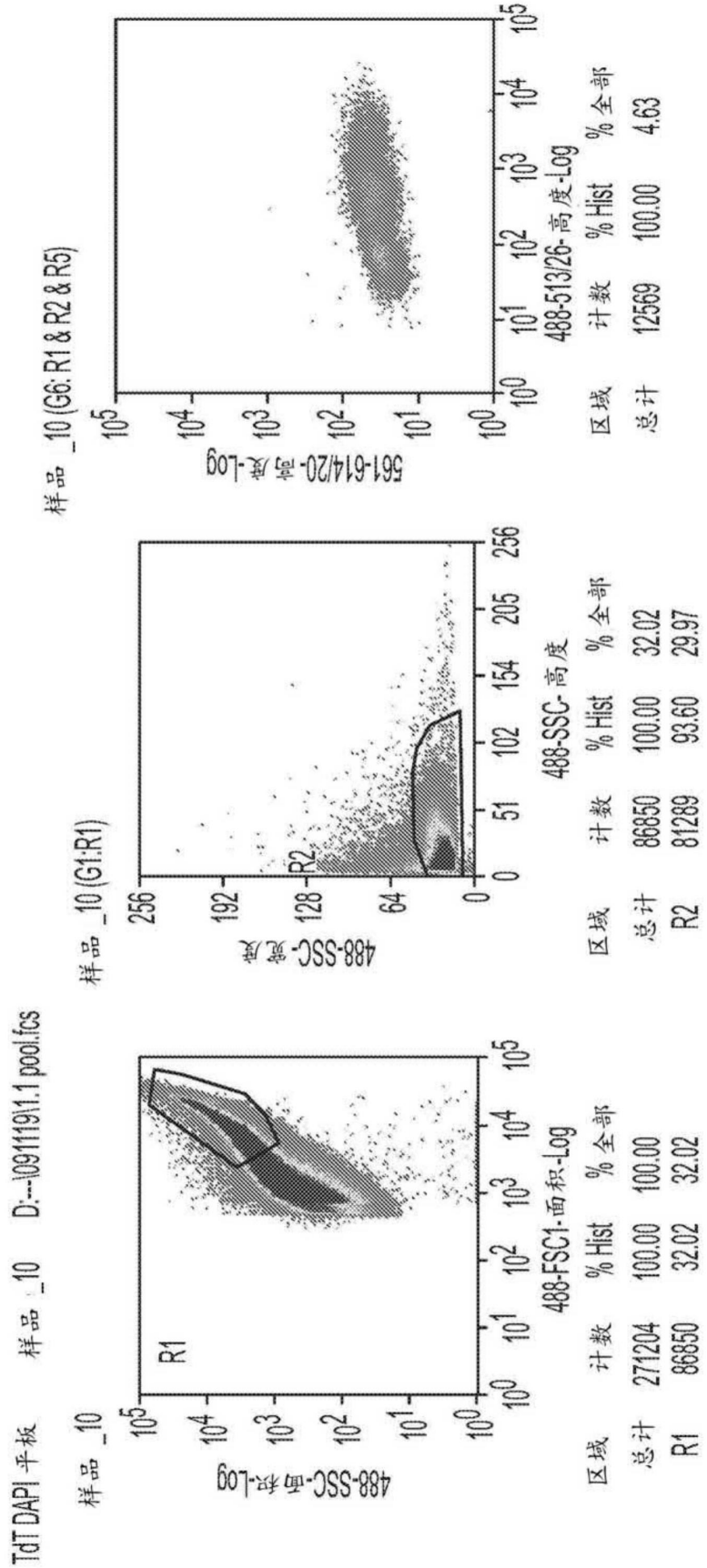

TdT DAPI 平板
样品 _10
D:---\091119\1.1 pool.fcs
样品 _10
R1
488-SSC-高度-Log
488-FSC1-面积-Log
区域 计数 % Hist % 全部
总计 271204 100.00 100.00
R1 86850 32.02 32.02
样品 _10 (G1:R1)
R2
488-SSC-高度
488-SSC-高度
区域 计数 % Hist % 全部
总计 86850 100.00 32.02
R2 81289 93.60 29.97
样品 _10 (G6: R1 & R2 & R5)
561-614/20-高度-Log
488-513/26-高度-Log
区域 计数 % Hist % 全部
总计 12569 100.00 4.63

图58A

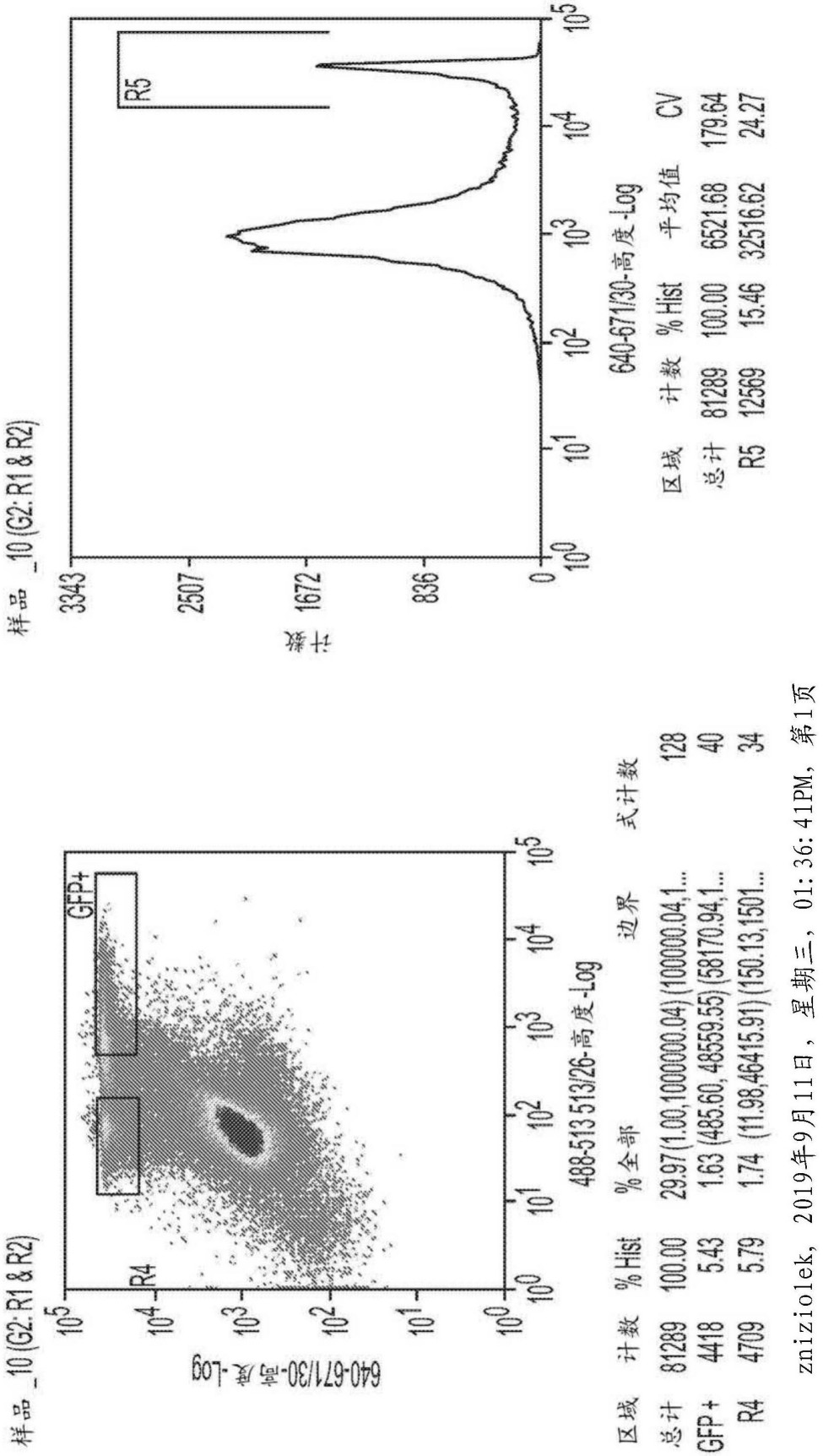

样品 _10 (G2: R1 & R2)
R5
计数
640-671/30-高度-Log
区域 计数 % Hist 平均值 CV
总计 81289 100.00 6521.68 179.64
R5 12569 15.46 32516.62 24.27
样品 _10 (G2: R1 & R2)
GFP+
R4
640-671/30-高度-Log
488-513 513/26-高度-Log
区域 计数 % Hist % 全部 边界 式计数
总计 81289 100.00 29.97(1.00,1000000.04) (100000.04,1... 128
GFP+ 4418 5.43 1.63 (485.60, 48559.55) (58170.94,1... 40
R4 4709 5.79 1.74 (11.98,46415.91) (150.13,1501... 34
zniziolek, 2019年9月11日，星期三，01: 36: 41PM，第1页

图58B

克隆概述

1. 用Bsa消化pU6-pegRNA-GG-载体质粒(组件1)并分离含有复制起点、U6启动子、U6 poly-T终止序列和AmpR基因的质粒片段(约2.2kb)

2. 排序寡核苷酸以:

a. 两侧有所示突出端的期望的间隔区(靶)序列(组件2)

i. 使用期望靶标的5’至3’序列作为上面链寡核苷酸(包含5’ CACC和3’ GTTTT突出端)和使用靶序列的反向互补序列作为下面链寡核苷酸(包含5’ CTCTAAAAC突出端)。间隔区序列必须由G核苷酸开始以有效的转录起始。

b. 两侧有所示突出端的期望的pegRNA 3’延伸模板(组件3)

i. 使用RNA有义序列作为上面链寡核苷酸(以5’ GTGC突出端为特征)和使用该序列的反向互补序列作为下面链寡核苷酸(以5’ AAAA突出端为特征)

c. 以兼容golden gate突出端为特征的SpCas9 sgRNA支架序列(组件4)

i. 这些寡核苷酸不是完整的支架序列,因为来自其余组件的突出端贡献了几个缺少的核苷酸

ii. 注意:这些寡核苷酸必须5’磷酸化。寡核苷酸可由制造商5’磷酸化或使用T4 PNK酶促5’磷酸化(参见以下方案)

3. 根据以下方案在单独的退火反应中退火上面和下面寡核苷酸作为组件2、3和4。如果SpCas9 sgRNA支架序列(组件4)未磷酸化,用T4 PNK磷酸化。

4. Golden Gate组装分离自组件1的2.2kb片段与组件2、3和4

5. 转化连接产物至大肠杆菌。由来自pU6-pegRNA-GG-载体质粒的组件1赋予的抗生素抗性为氨苄青霉素和羧苄青霉素抗性。

6. 分离和测序来自所产生克隆转化体的质粒

图59A

pegRNA克隆方案

步骤1：消化pU6-pegRNA-GG-载体质粒(组件1)

在PCR管中混合以下：

2000 ng pU6-pegRNA-GG-载体(组件 1)	X μL
Bsa1-HFv2 (NEB)	1.0 μL
10x Cutsmart缓冲液	3.0 μL
H_2O	至 30.0 μL
总反应体积	30.0 μL

于37°C温育4至16小时

从切割的质粒分离约2.2kb片段

图59B

步骤2和3：排序和退火寡核苷酸部分(组件2、3和4)

材料

退火缓冲液：补充有10mM Tris-Cl pH8.5和50mM NaCl的H_2O

互补寡核苷酸对

方案

在PCR管中混合以下：

上面寡核苷酸,100 μM	1.0 μL
下面寡核苷酸 ,100 μM	1.0 μL
退火缓冲液(组件 2,3,和 4)	23.0 μL
总反应体积	25.0 μL

在热循环仪中，于95°C加热3分钟，然后逐渐冷却至22°C(0.1°C/s)

通过添加75 μL H20以1:4稀释退火的寡核苷酸。该稀释之后每种寡核苷酸的最终浓度为1μM。如果步骤2.5中通过PNK磷酸化，则不稀释sgRNA支架(组件4)。

图59C

步骤2.b.ii.：sgRNA支架磷酸化(如果购买磷酸化的寡核苷酸则不需要)

方案

在PCR管中混合以下：

4 μM来自步骤1的寡核苷酸双链	6.25 μL
10x T4 DNA连接酶缓冲液(NEB)	2.50 μL
T4 PNK (NEB)	0.50 μL
H_2O	15.75 μL
总反应体积	20.0 μL

在热循环仪中，于37°C温育60分钟

该磷酸化之后，退火支架寡核苷酸目前浓度为1 μM。

继续步骤3

图59D

步骤4：pegRNA组装

pegRNA，Golden Gate组装反应

消化的pU6-pegRNA-GG质粒载体	1.00 μL @ 30 ng/μL
-预先切割，分离的2.2kb片段	
退火的前间区序列寡核苷酸(组件2)	1.00 μL @ 1 μM
退火的pegRNA 3’延伸寡核苷酸(组件4)	1.00 μL @ 1 μM
sgRNA支架退火的寡核苷酸(组件3)	1.00 μL @ 1 μM
-寡核苷酸必须磷酸化	
BsaI-HFv2 (NEB)	0.25 μL
T4 DNA连接酶(NEB)	0.50 μL
10x T4 DNA连接酶缓冲液(NEB)	1.00 μL
H_2O	4.25 μL
总反应体积:	10.0 μL

于室温温育10分钟

温育之后，在热循环仪中进行以下程序：

于37°C，5min；于80°C，5min；于12°C保持

图59E

步骤5和6：组装质粒的转化

将10 μL组装反应中的1 μL转化至10 μL感受态细胞。期望的转化体将对氨苄青霉素和羧苄青霉素具有抗性。

以下图总结了pegRNA克隆方案。

图59F

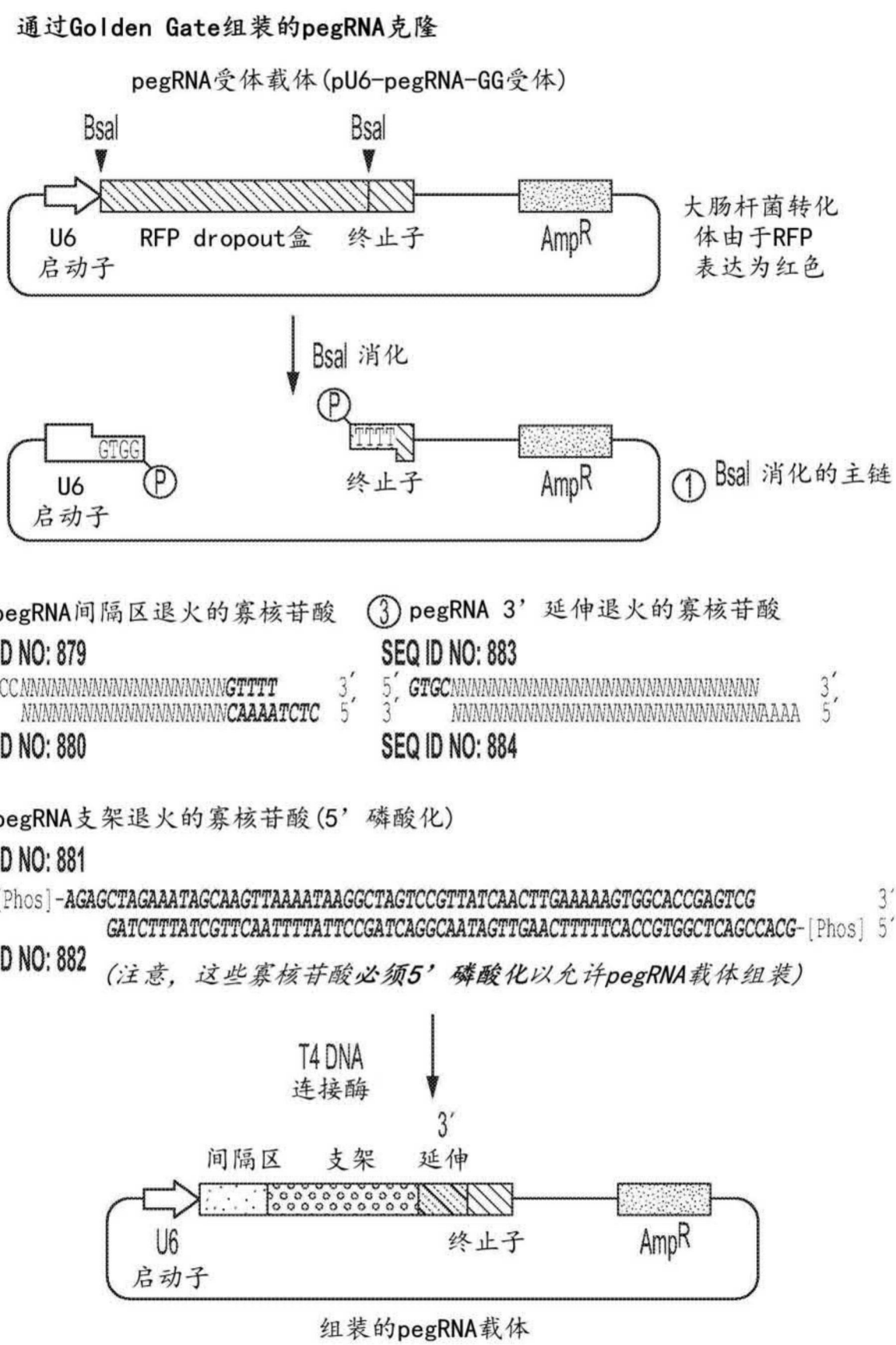

通过Golden Gate组装的pegRNA克隆
pegRNA受体载体(pU6-pegRNA-GG受体)
BsaI
BsaI
U6
启动子
RFP dropout盒
终止子
AmpR
大肠杆菌转化体由于RFP表达为红色
BsaI 消化
GTGG
P
TTTT
① BsaI 消化的主链
② pegRNA间隔区退火的寡核苷酸
SEQ ID NO: 879
5′ CACCNNNNNNNNNNNNNNNNNNGTTTT 3′
3′ NNNNNNNNNNNNNNNNNNCAAAATCTC 5′
SEQ ID NO: 880
③ pegRNA 3’延伸退火的寡核苷酸
SEQ ID NO: 883
5′ GTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3′
3′ NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 5′
SEQ ID NO: 884
④ pegRNA支架退火的寡核苷酸(5’磷酸化)
SEQ ID NO: 881
5′ [Phos]-AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG 3′
3′ GATCTTTATCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACG-[Phos] 5′
SEQ ID NO: 882
(注意，这些寡核苷酸必须5’磷酸化以允许pegRNA载体组装)
T4 DNA
连接酶
3′
间隔区
支架
延伸
U6
启动子
终止子
AmpR
组装的pegRNA载体

图59G

```
##sgRNA支架序列搜索##

输入pandas为pd
输入Bio为bio
从Bio输入Seql()
输入glob

#产生fastq文件列表来分析
来源=glob.glob(‘*.fastq’)

#读取fastq文件至具有关键词作为文件名的词典
fastqdict={}
对于i范围内(len(来源)):
        temp=列表(Seql().解析(来源[i], “fastq”))
        fastqdict[来源[i]]=[str(temp[k].seq)对于k范围内(len(temp))]

#搜索的参考序列进入以下词典具有
#合适的关键词
scaffdict =
{'HEK3':'CAGAGGACCGACTCGGTCCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT
GCTATTTCTAGCTCTAAAACTCACGTGCTCAGTCTGGGCCGGTG',(SEQ ID NO:885)
'EMX1':'ATCACGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT
TGCTATTTCTAGCTCTAAAACTTCTTCTTCTGCTCGGACTCGGTG',(SEQ ID NO:886)
'FANCF':'TTTCCGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC
TTGCTATTTCTAGCTCTAAAACGGTGCTGCAGAAGGGATTCCGGTG',(SEQ ID NO:887)
'RNF2':'TCGTTGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT
GCTATTTCTAGCTCTAAAACCAGGTAATGACTAAGATGACGGTG'}(SEQ ID NO:888)

#匹配和计数参考串的迭代片段至合适的fastq文件
#参考关键词必须包含在fastq文件名中
#产生的值代表最低程度sgRNA整合的累计计数
#即, 给定值x意指x读段含有支架的y个或更多碱基
resultdic=来自关键词的dic.(来源)
对于fastqdict中的关键词:
  对于scaffdict中的支架:
    如果str(关键词)中的支架
      结果列表=[]
      对于j范围内(len(scaffdict[支架])):
```

图60A

```
            范围=scaffdict[支架][():(j+1)]
            计数器=()
            对于i范围内(len(fastqdict[关键词])):
              如果范围在fastqdict[关键词][i]内:
                计数器=计数器+1
              结果列表.附加(计数器)
          resultdic[关键词]= 结果列表

#结果写入由1索引的数据框
resultdf=来自_dict(resultdic)的pd.数据框
resultdf= resultdf.重建索引(分选的(Resultdf.列), 轴=1)
resultdf.索引=范围(1, len(resultdf)+1)

#转化累计计数值为具体计数
#如, 给定值x意指x读段含有支架的精确y个碱基
resultdf2= resultdf.拷贝()
对于进入resultdf:
  对于i范围内(1, len(resultdf)[进入]+1):
    尝试:
      resultdf2[进入][i]= resultdf[进入][i]- resultdf [进入][i+1]
    除了:
      resultdf2[进入][i]= resultdf[进入][i]

#转化具体计数值为频率
resultdf3= resultdf2.拷贝()
对于进入resultdf3:
   resultdf3[进入]= resultdf2[进入].div(resultdf[进入][1])*100

#读取结果至excel文件
resultdf.至_excel('累计计数.xlsx')
resultdf2.至_excel('具体计数.xlsx')
resultdf3.至_excel('具体频率.xlsx')
```

图60B

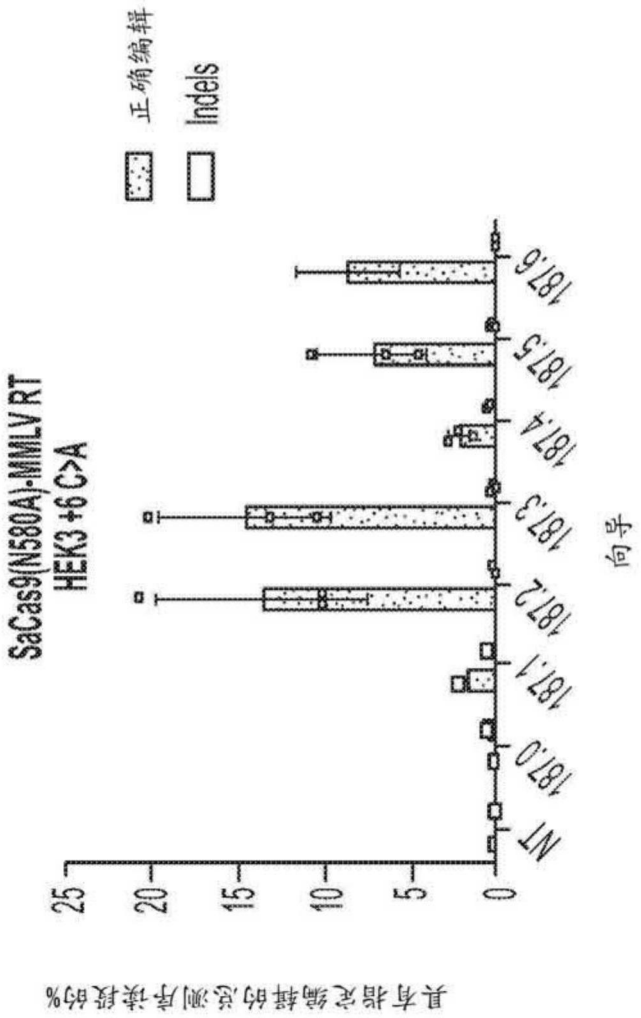

正确编辑
Indels
SaCas9(N580A)-MMLV RT
HEK3 +6 C>A
NT
187.0
187.1
187.2
187.3
187.4
187.5
187.6
向导
具有指定编辑的总测序读段的%
0
5
10
15
20
25

图61

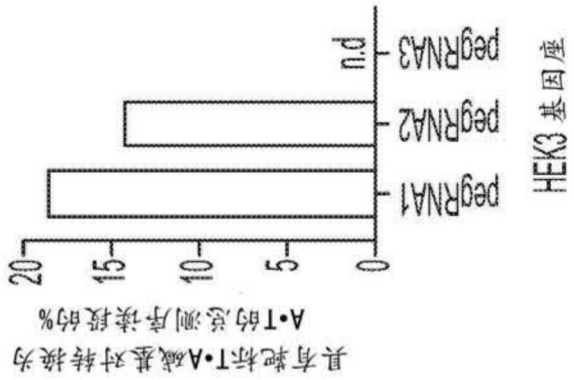

pegRNA1
pegRNA2
pegRNA3
n.d
HEK3 基因座
具有靶标T•A碱基对转换为
A•T的总测序读段的%
0
5
10
15
20

图62A

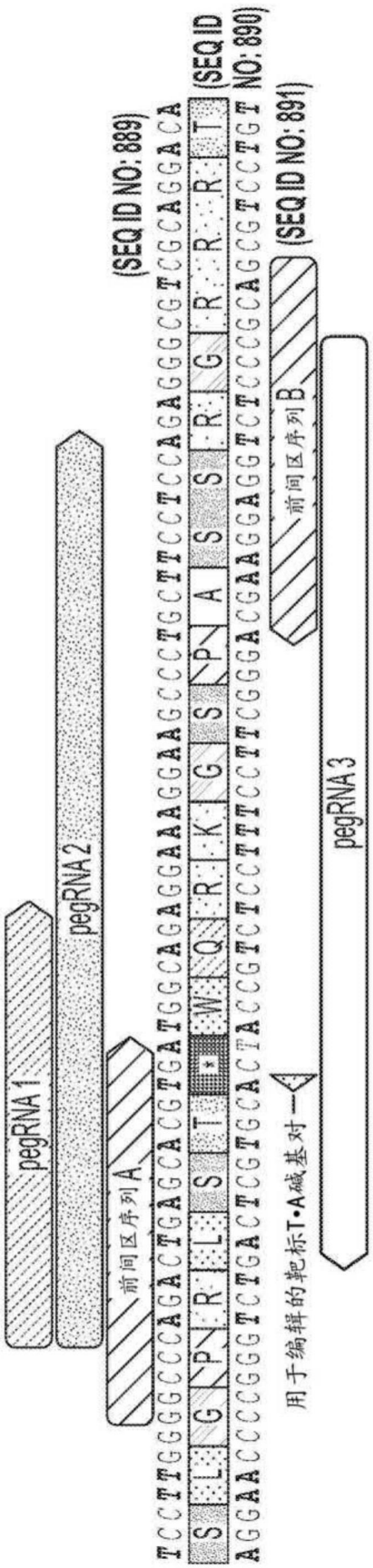

pegRNA1
pegRNA2
前间区序列A
pegRNA3
前间区序列B
用于编辑的靶标T•A碱基对
(SEQ ID NO: 889)
(SEQ ID NO: 890)
(SEQ ID NO: 891)

图62B

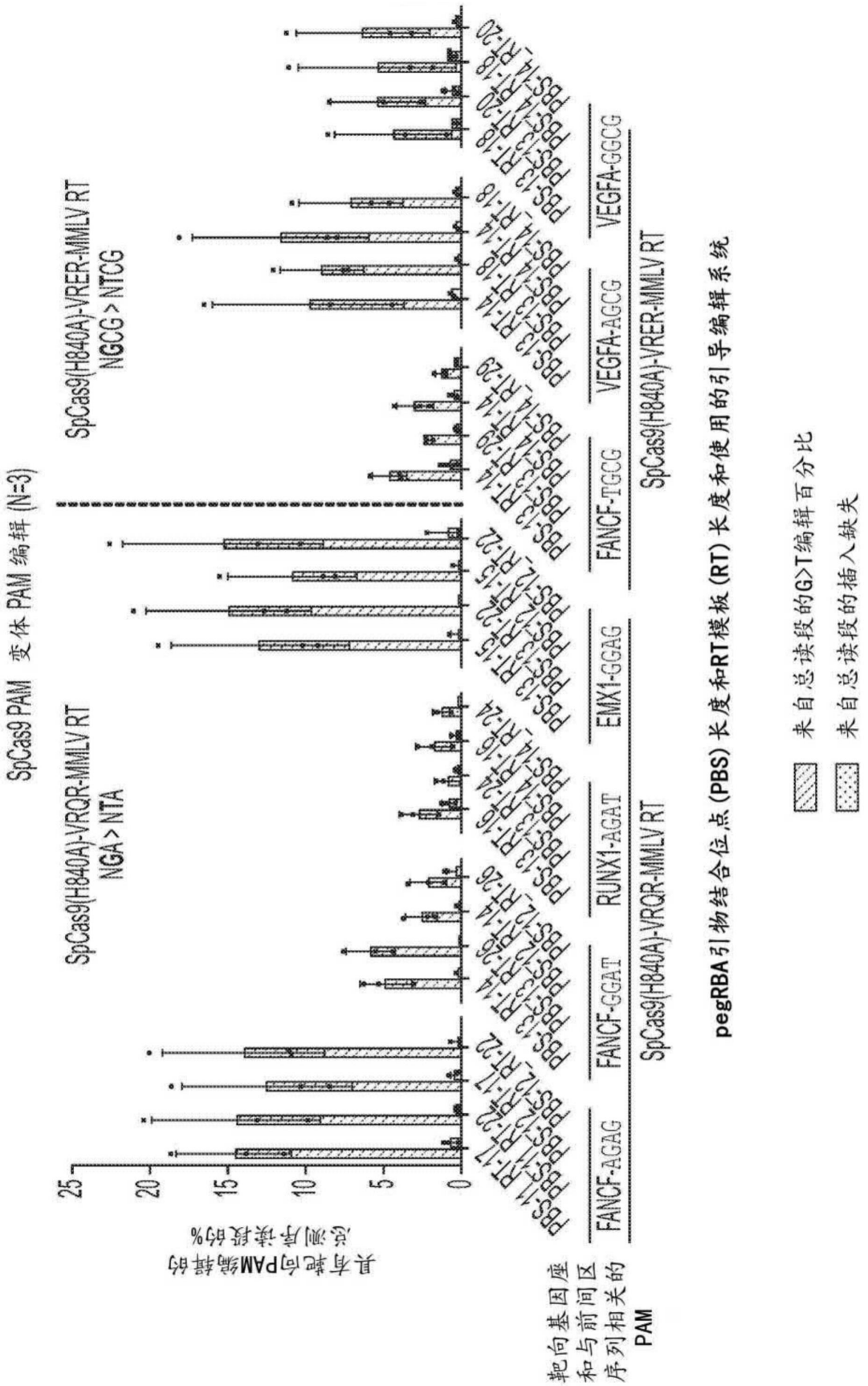

SpCas9 PAM 变体 PAM 编辑 (N=3)
SpCas9(H840A)-VRQR-MMLV RT
NGA > NTA
SpCas9(H840A)-VRER-MMLV RT
NGCG > NTCG
具有靶向PAM编辑的经测序读段的%
25
20
15
10
5
0
靶向基因座和与前间区序列相关的PAM
FANCF-AGAG
FANCF-GGAT
RUNX1-AGAT
EMX1-GGAG
FANCF-TGCG
VEGFA-AGCG
VEGFA-GGCG
SpCas9(H840A)-VRQR-MMLV RT
SpCas9(H840A)-VRER-MMLV RT
pegRBA引物结合位点(PBS)长度和RT模板(RT)长度和使用的引导编辑系统
来自总读段的G>T编辑百分比
来自总读段的插入缺失

图63

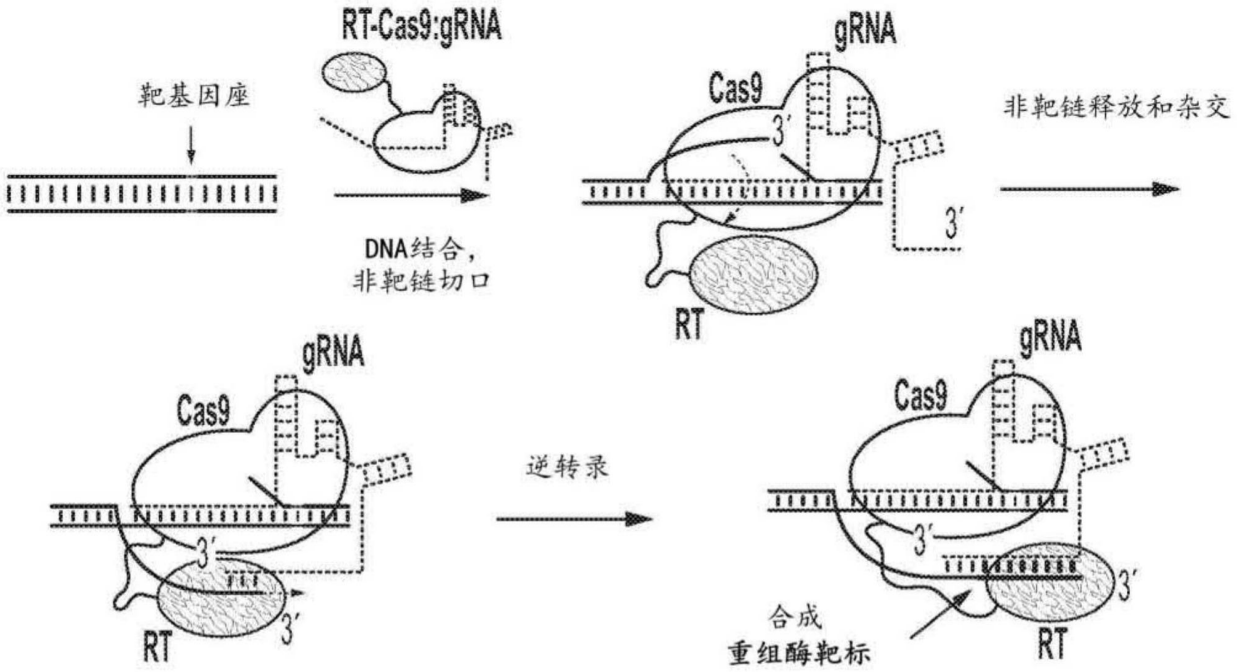


图64A

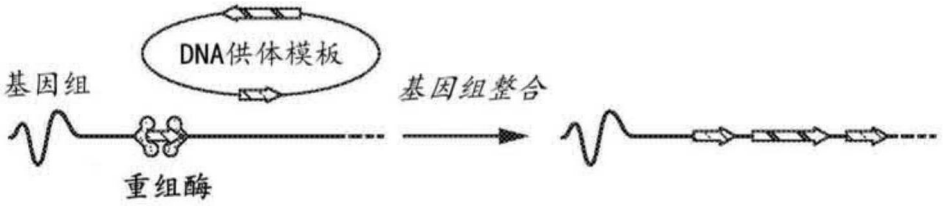


图64B

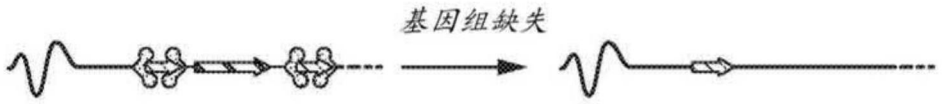


图64C

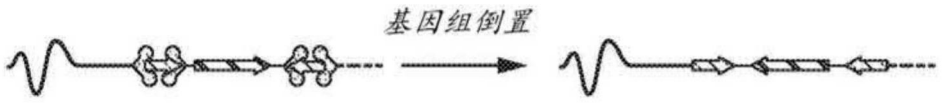


图64D

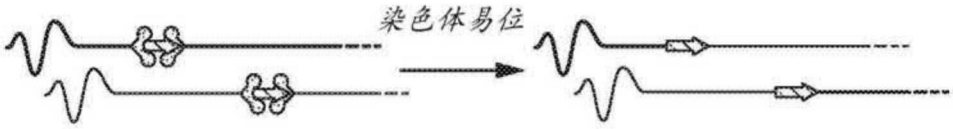


图64E



图64F

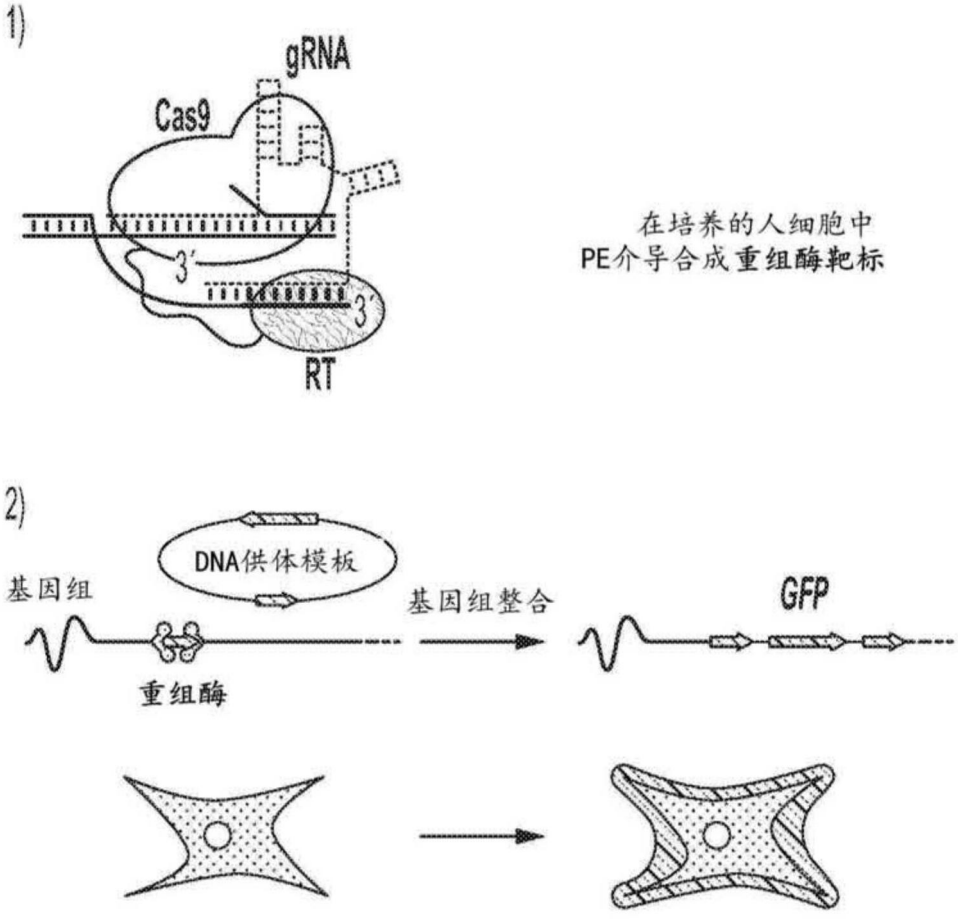
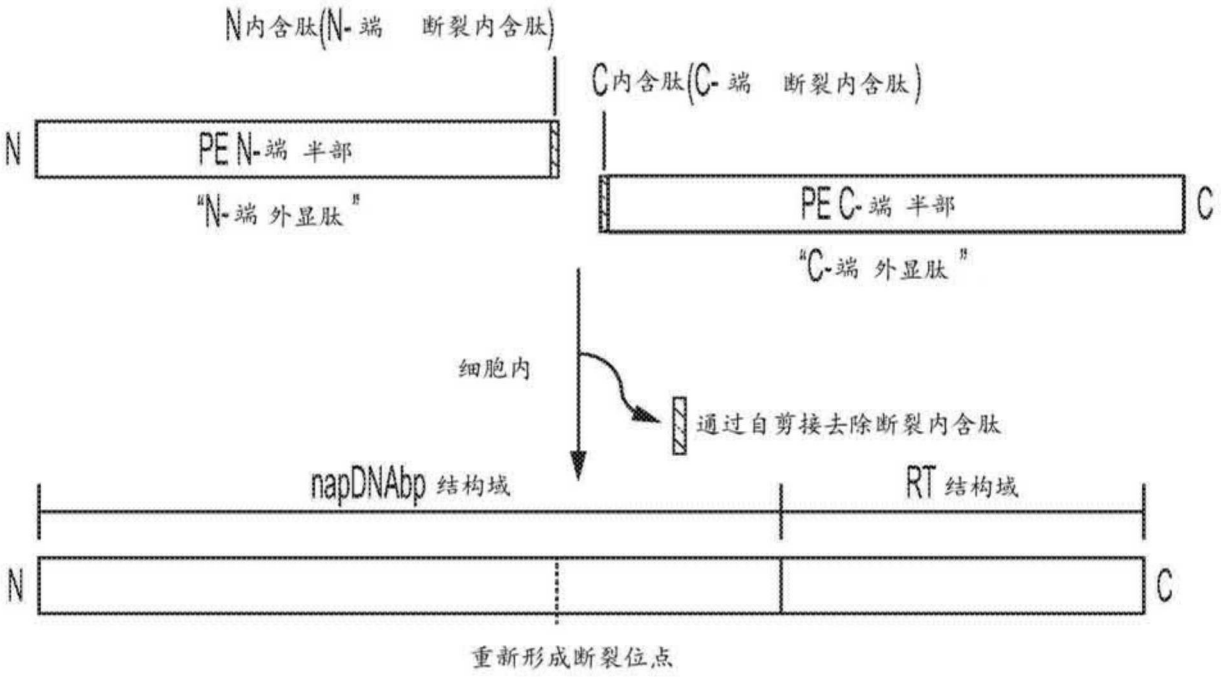


图65


N内含肽(N-端 断裂内含肽)
C内含肽(C-端 断裂内含肽)
N
PE N-端 半部
"N-端 外显肽"
PE C-端 半部
C
"C-端 外显肽"
细胞内
通过自剪接去除断裂内含肽
napDNAbp 结构域
RT 结构域
N
C
重新形成断裂位点

引导编辑器

图66

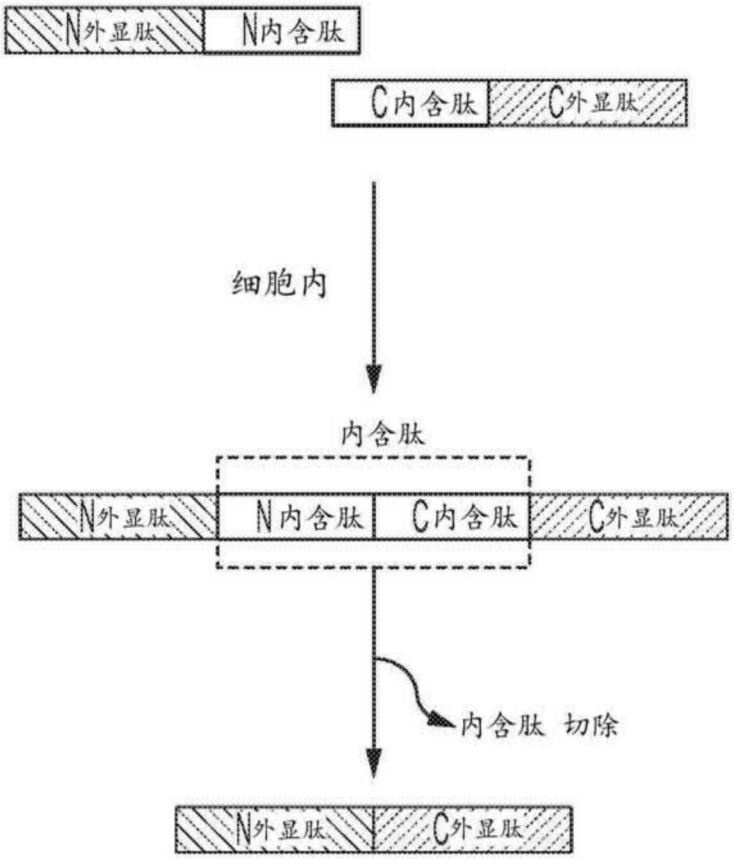

N外显肽
N内含肽
C内含肽
C外显肽
细胞内
内含肽
N外显肽
N内含肽
C内含肽
C外显肽
内含肽 切除
N外显肽
C外显肽

图67A

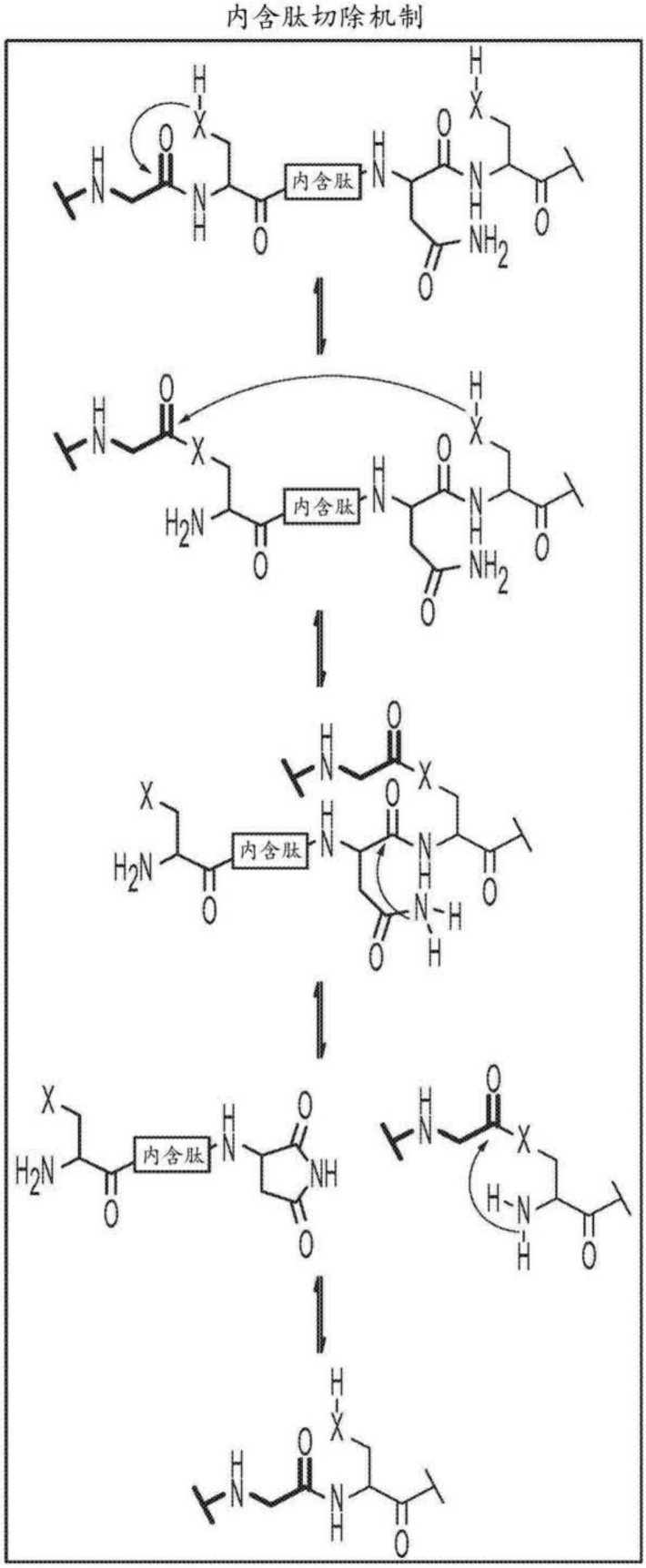


图67B

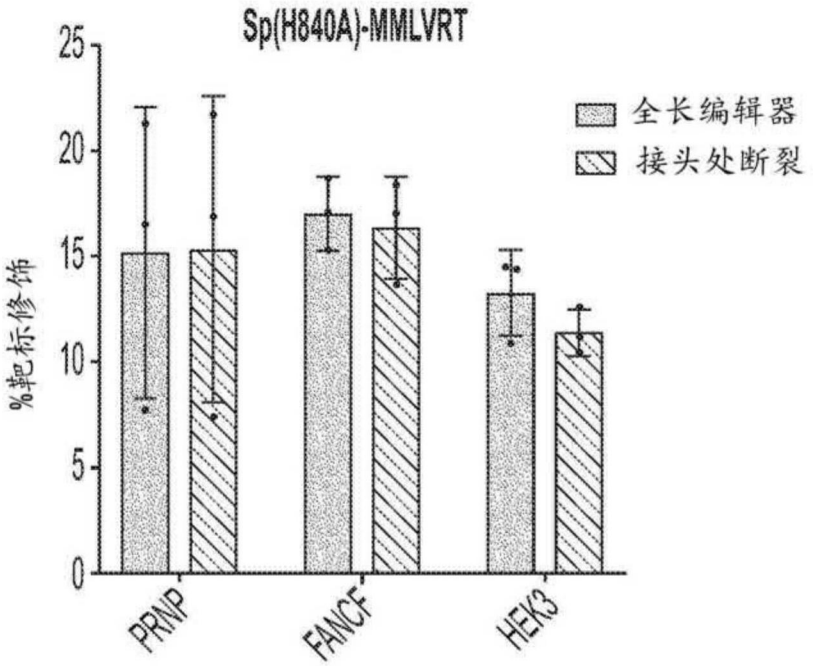

Sp(H840A)-MMLVRT
全长编辑器
接头处断裂
%靶标修饰
25
20
15
10
5
0
PRNP
FANCF
HEK3

图68A

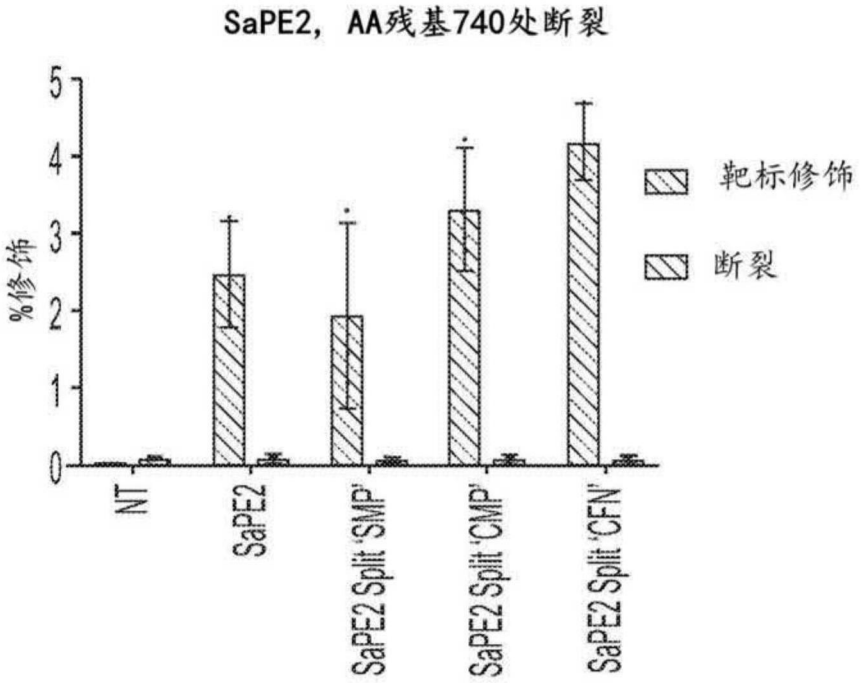

SaPE2，AA残基740处断裂
靶标修饰
断裂
%修饰
5
4
3
2
1
0
NT
SaPE2
SaPE2 Split 'SMP'
SaPE2 Split 'CMP'
SaPE2 Split 'CFN'

图68B

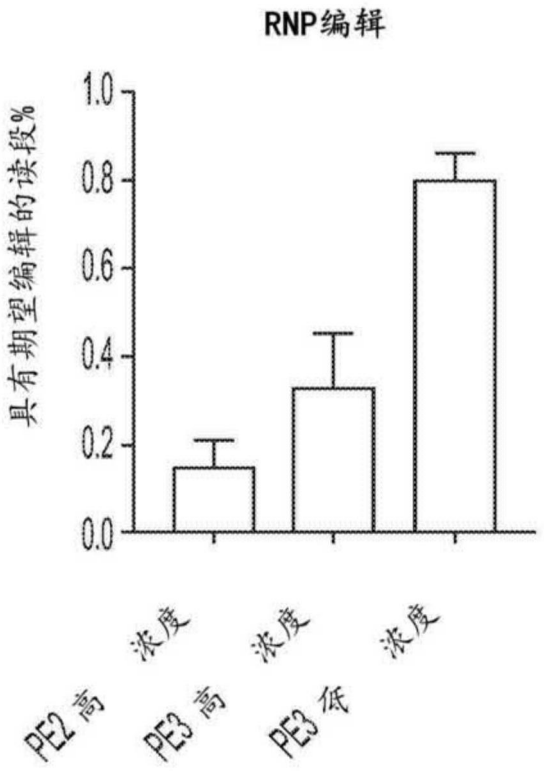

RNP编辑
具有期望编辑的读段%
1.0
0.8
0.6
0.4
0.2
0.0
PE2 高 浓度
PE3 高 浓度
PE3 低 浓度

图68C

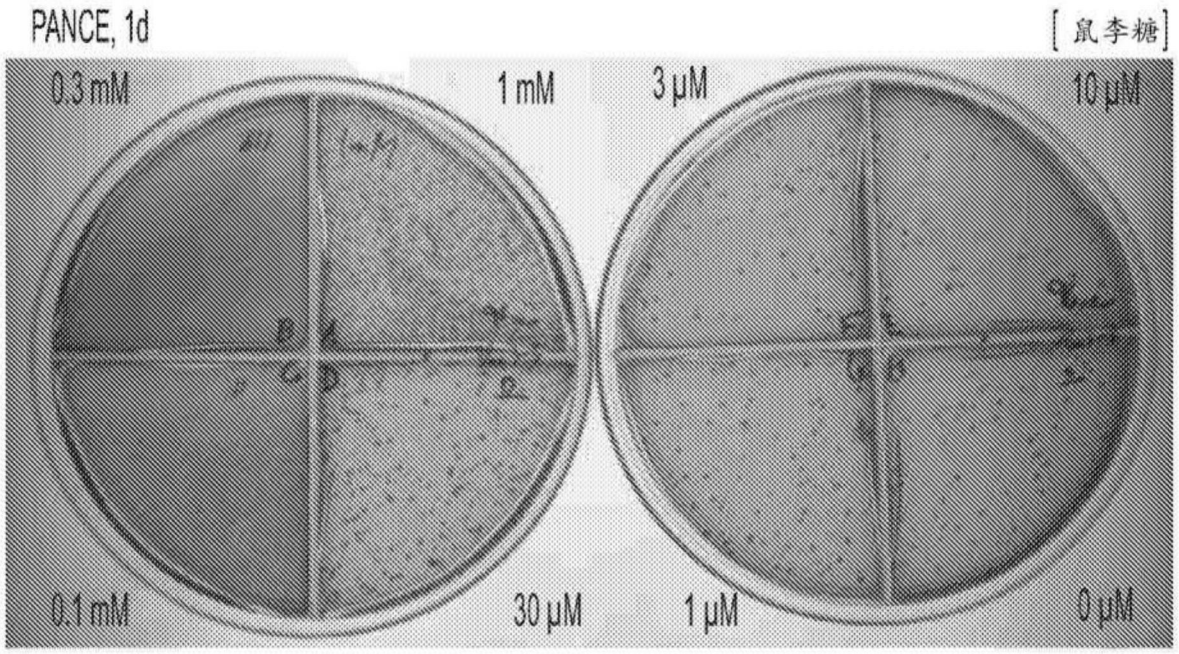

PANCE, 1d
[鼠李糖]
0.3 mM
1 mM
3 μM
10 μM
0.1 mM
30 μM
1 μM
0 μM

图69

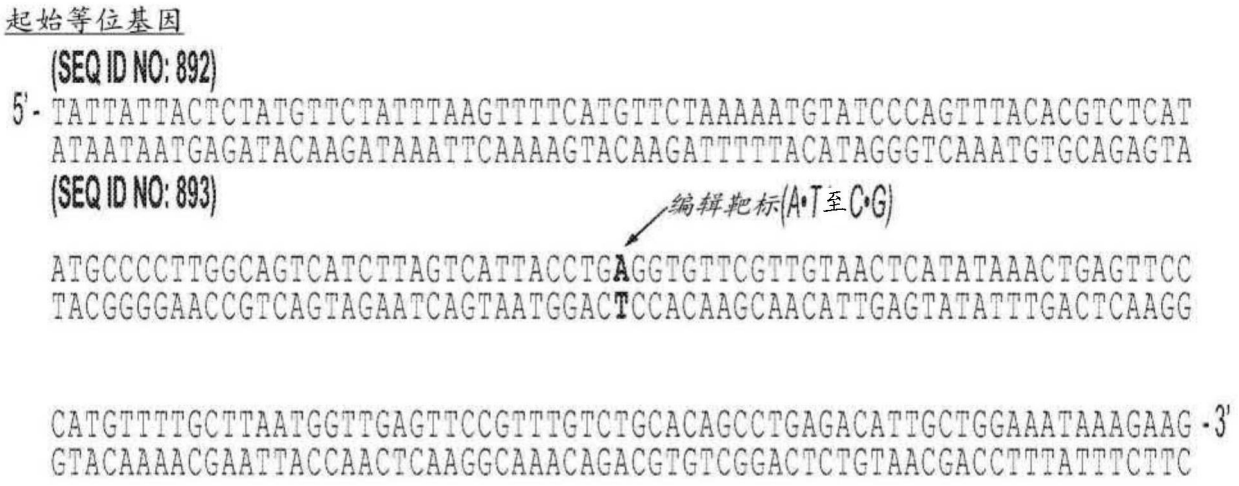

起始等位基因
(SEQ ID NO: 892)
5' - TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
编辑靶标(A•T至C•G)
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC

图70A


起始等位基因
(SEQ ID NO: 892)
5' - TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
NGG PAM 1
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG
NGG PAM 2
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC

图70B

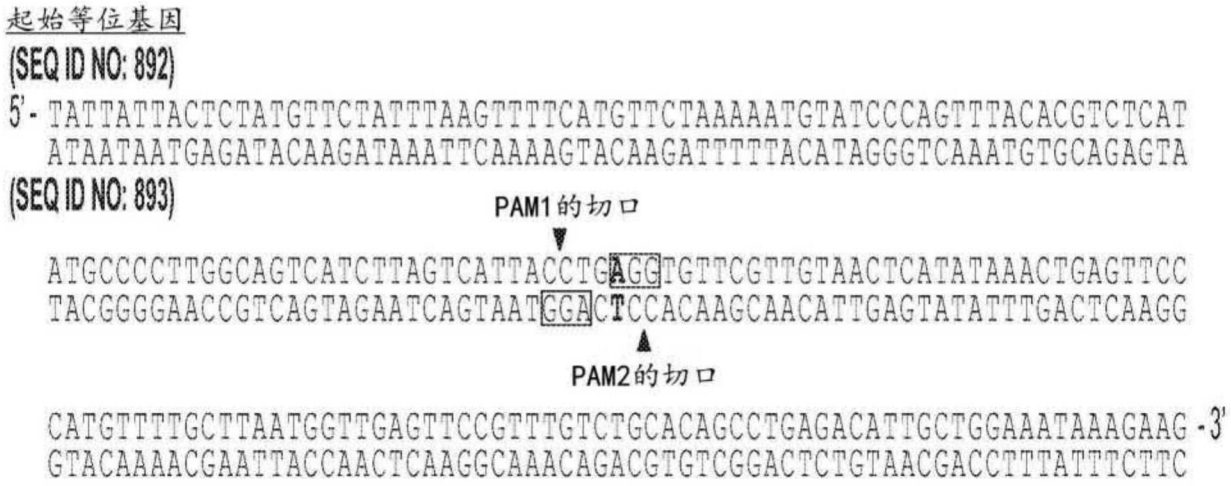

起始等位基因
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
PAM1的切口
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG
PAM2的切口
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC

图70C

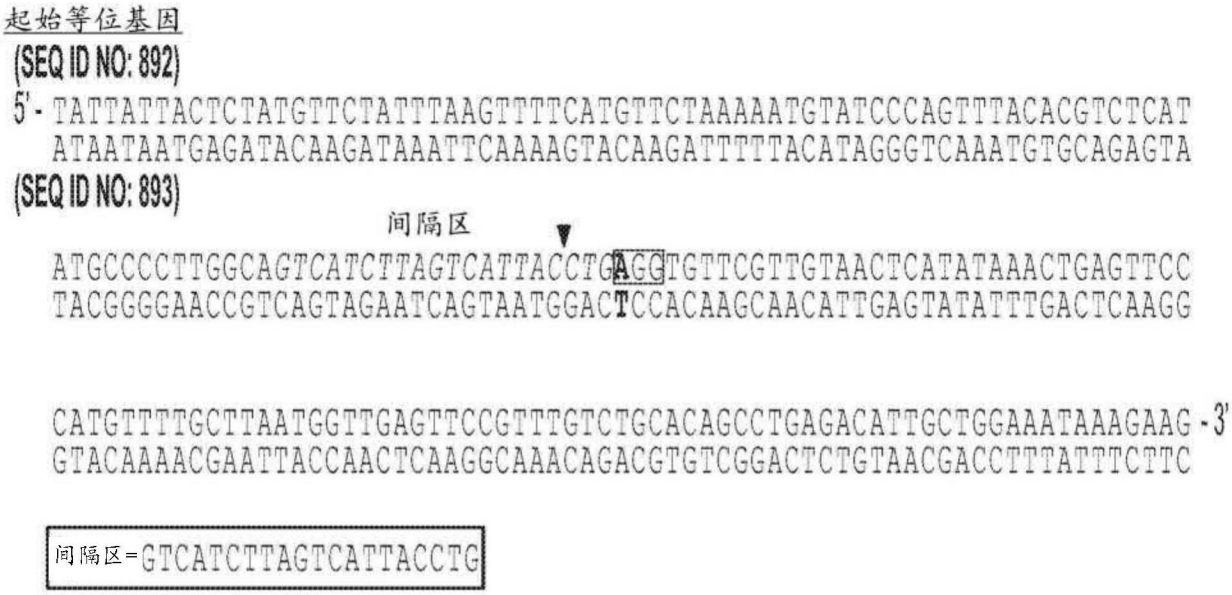

起始等位基因
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
间隔区
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC
间隔区=GTCATCTTAGTCATTACCTG

图70D

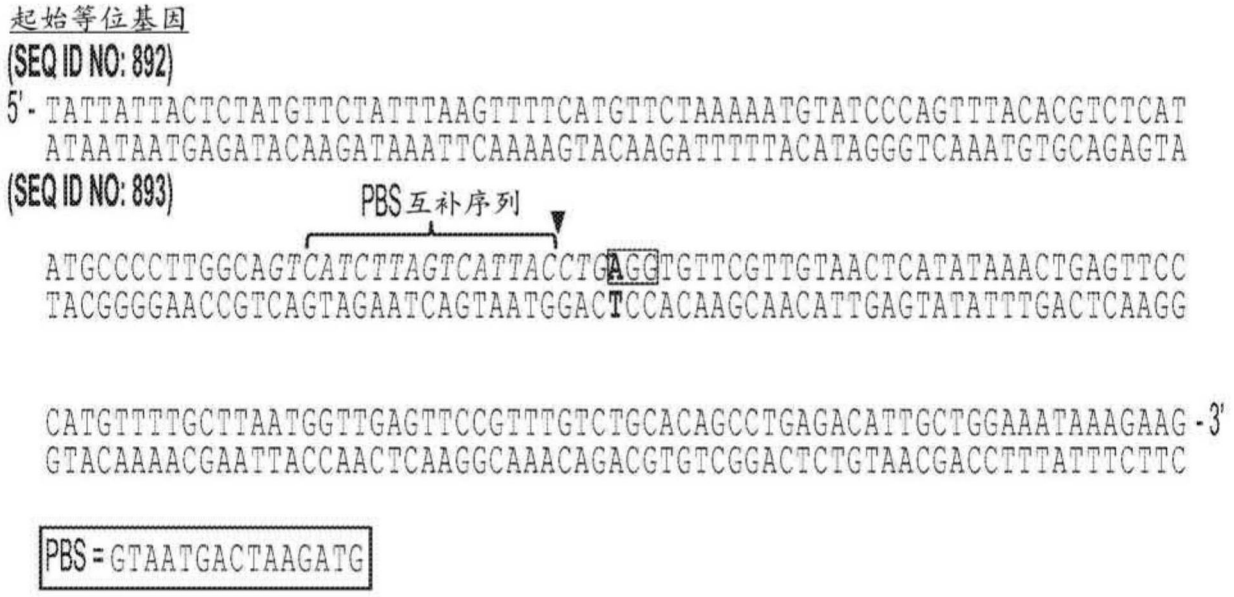

起始等位基因
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
PBS互补序列
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGCTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC
PBS = GTAATGACTAAGATG

图70E

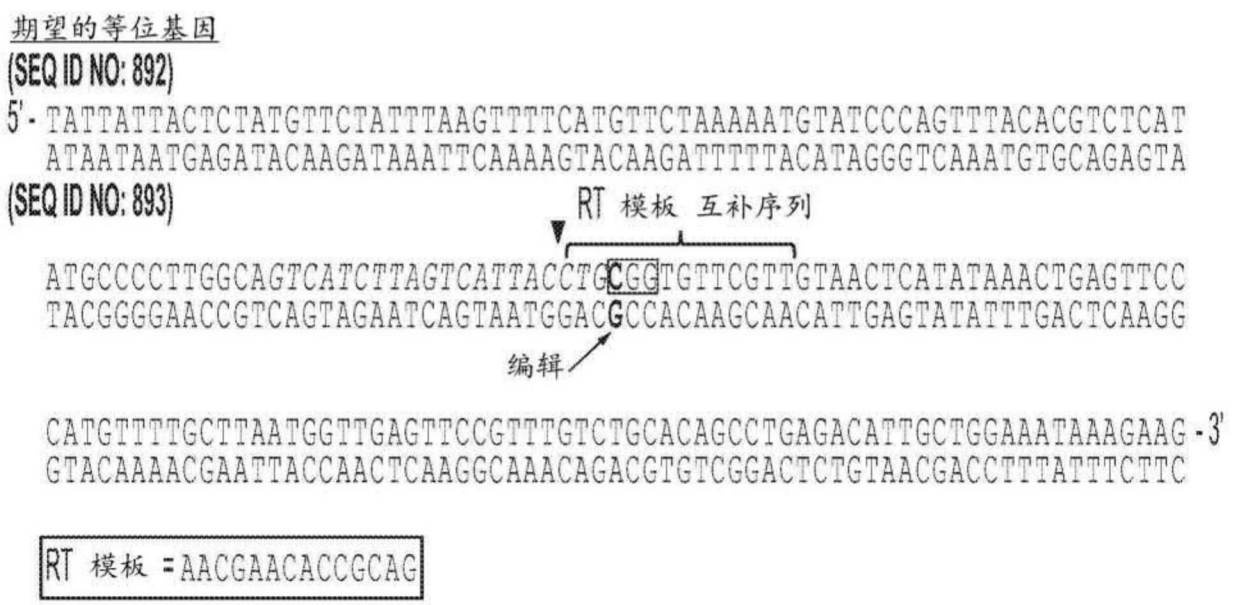

期望的等位基因
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
RT 模板 互补序列
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGCCGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACGCCACAAGCAACATTGAGTATATTTGACTCAAGG
编辑
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC
RT 模板 = AACGAACACCGCAG

图70F

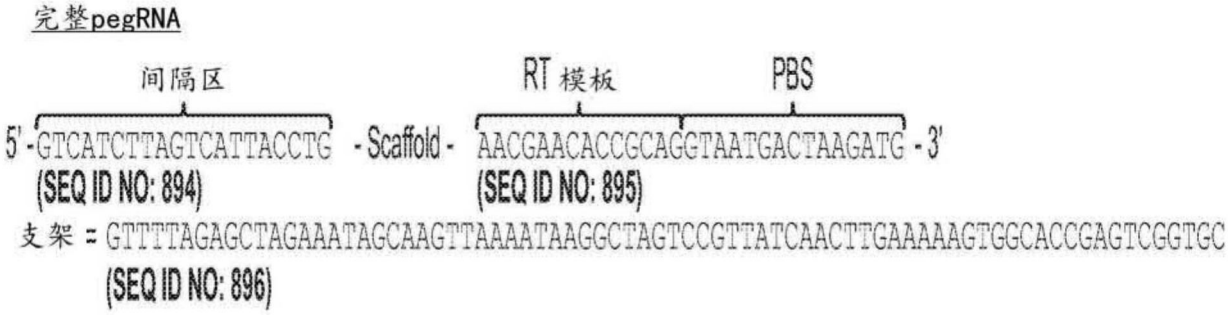

完整pegRNA
间隔区
RT 模板
PBS
5'-GTCATCTTAGTCATTACCTG -Scaffold- AACGAACACCGCAGGTAATGACTAAGATG-3'
(SEQ ID NO: 894)
(SEQ ID NO: 895)
支架 = GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
(SEQ ID NO: 896)

图70G

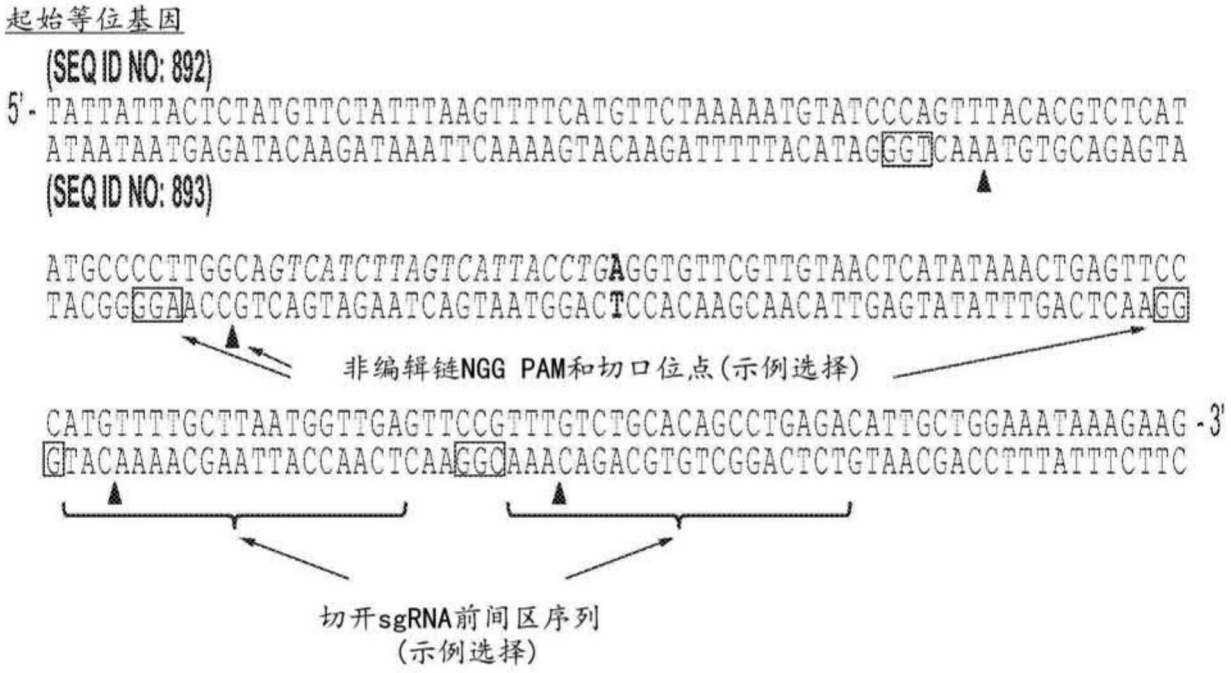

起始等位基因
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG
非编辑链NGG PAM和切口位点(示例选择)
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG-3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC
切开sgRNA前间区序列
(示例选择)

图70H

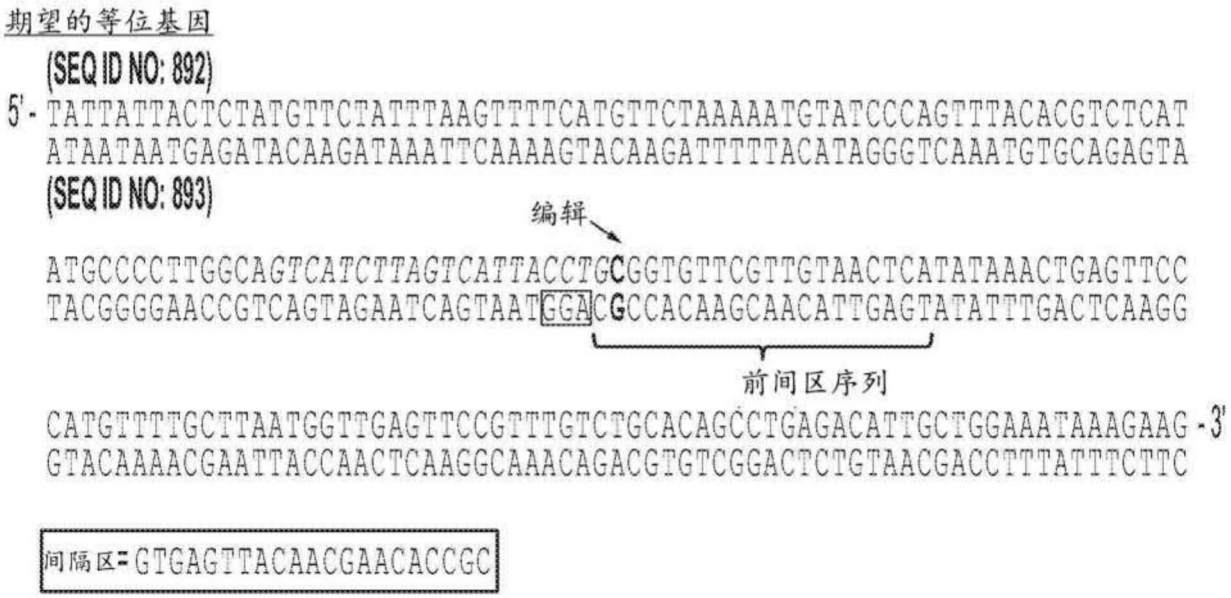

期望的等位基因
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
编辑
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGCGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACGCCACAAGCAACATTGAGTATATTTGACTCAAGG
前间区序列
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC
间隔区= GTGAGTTACAACGAACACCGC

图70I

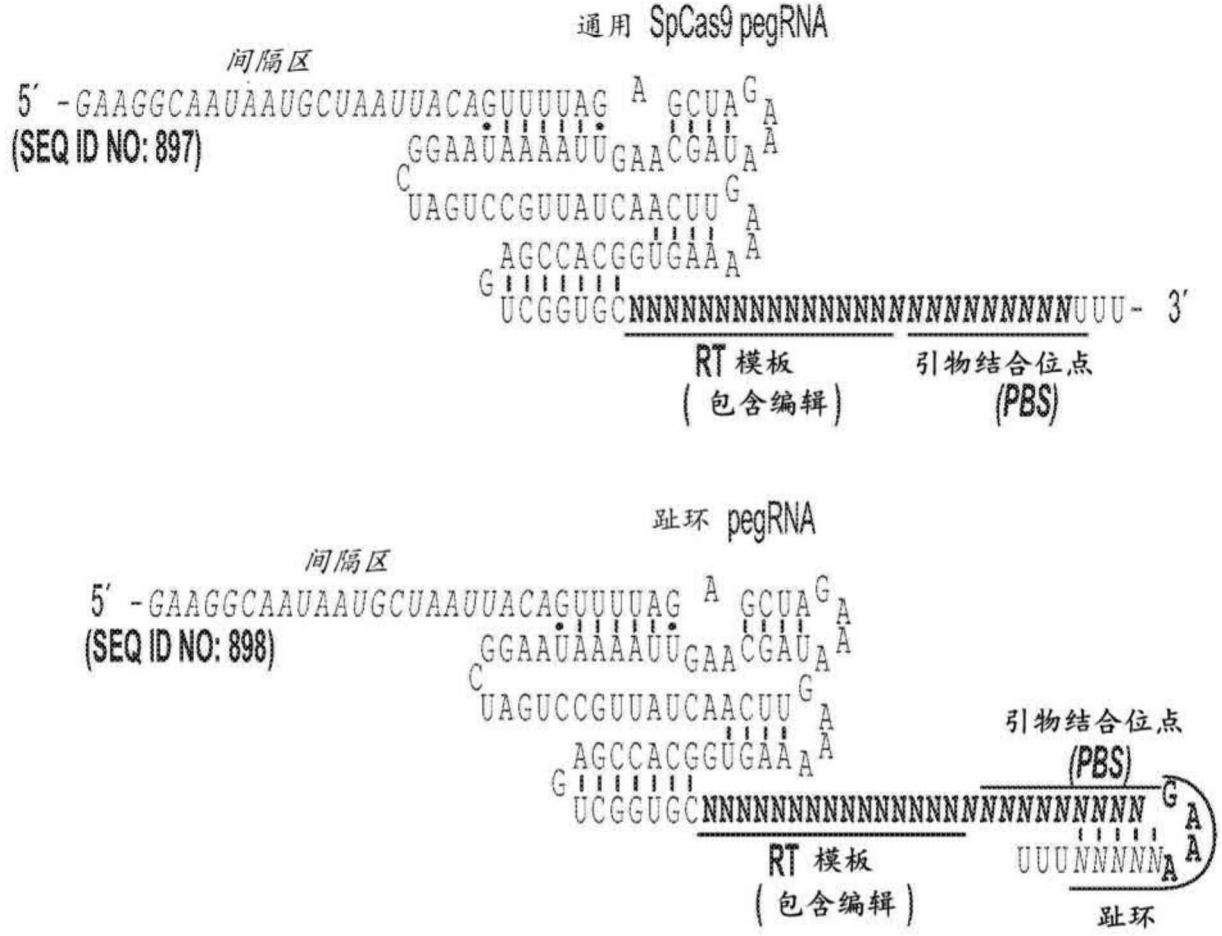

通用 SpCas9 pegRNA
间隔区
5′ -GAAGGCAAUAAUGCUAAUUACAGUUUUAG A GCUA G A
(SEQ ID NO: 897) GGAAUAAAAUU GAA CGAU A A
C UAGUCCGUUAUCAACUU G A
AGCCACGGUGAA A A
G UCGGUGCNNNNNNNNNNNNNNNNNNNNUUU- 3′
RT 模板
(包含编辑)
引物结合位点
(PBS)
趾环 pegRNA
间隔区
5′ -GAAGGCAAUAAUGCUAAUUACAGUUUUAG A GCUA G A
(SEQ ID NO: 898) GGAAUAAAAUU GAA CGAU A A
C UAGUCCGUUAUCAACUU G A
AGCCACGGUGAA A A
G UCGGUGCNNNNNNNNNNNNNNNNNNNN G A
UUUNNNNN A A
RT 模板
(包含编辑)
引物结合位点
(PBS)
趾环

图71A

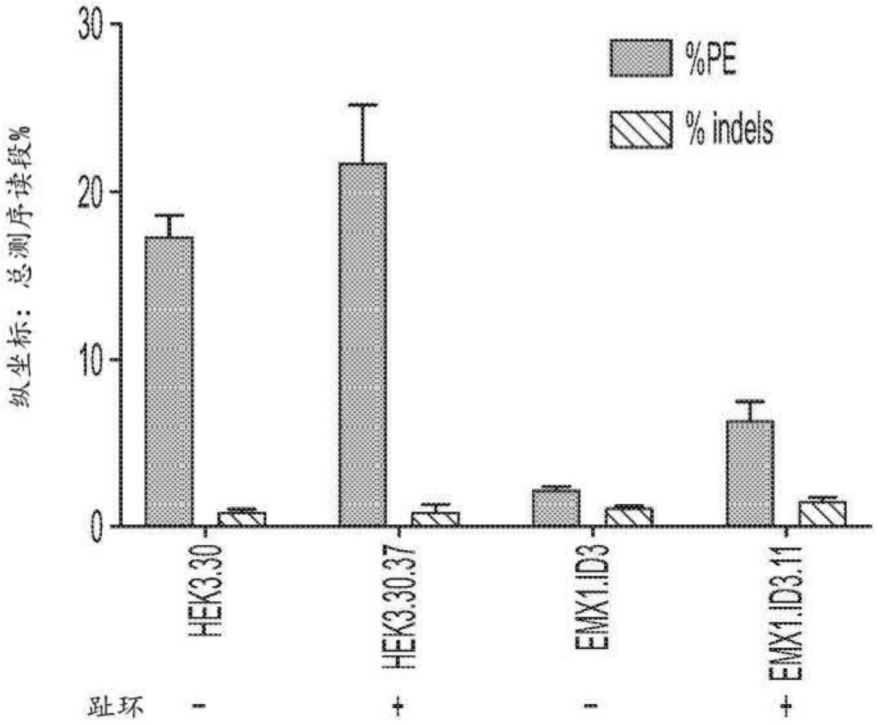

30
20
10
0
%PE
% indels
纵坐标：总测序读段%
HEK3.30
HEK3.30.37
EMX1.ID3
EMX1.ID3.11
趾环
-
+
-
+

图71B

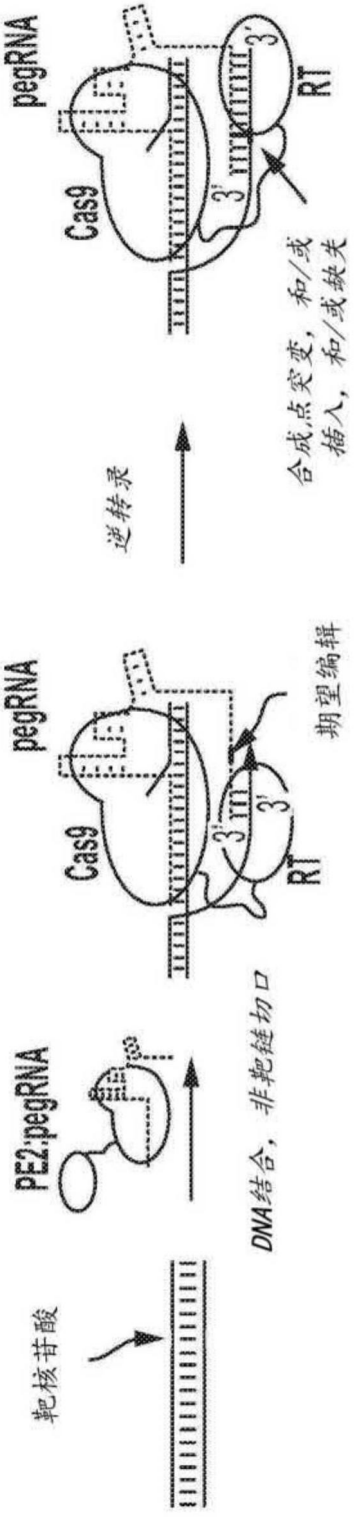

靶核苷酸
PE2:pegRNA
DNA结合，非靶链切口
pegRNA
Cas9
RT
3'
期望编辑
逆转录
合成点突变，和/或
插入，和/或缺失

图72A

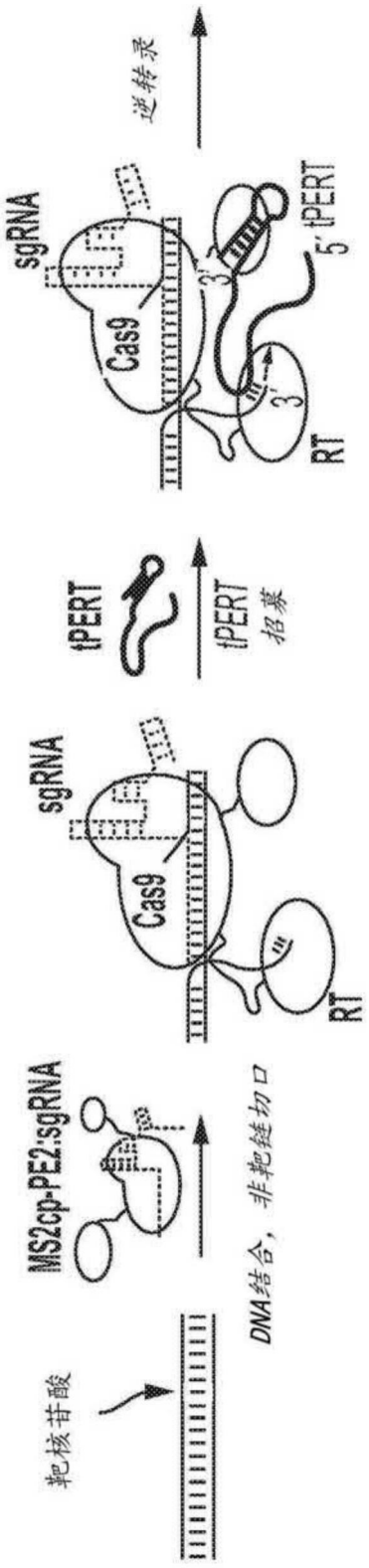

逆转录
sgRNA
Cas9
3'
5' tPERT
3'
RT
tPERT
tPERT
招募
sgRNA
Cas9
RT
MS2cp-PE2:sgRNA
DNA结合，非靶链切口
靶核苷酸

图72B

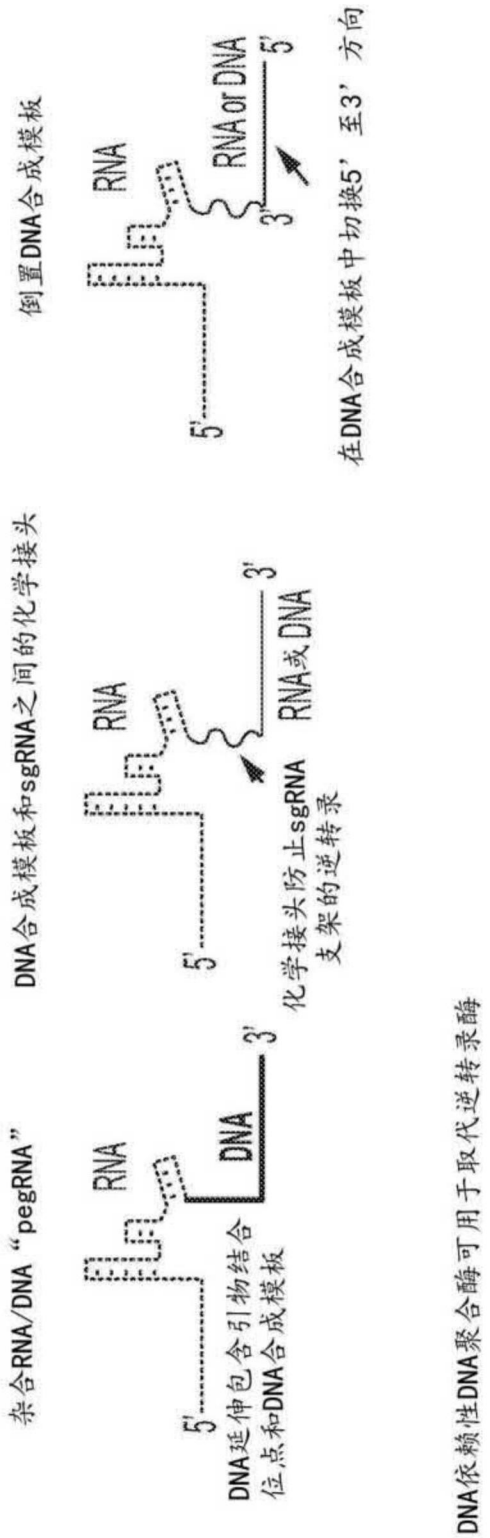

杂合RNA/DNA “pegRNA”
RNA
DNA
5'
3'
DNA延伸包含引物结合位点和DNA合成模板
DNA合成模板和sgRNA之间的化学接头
RNA
5'
3'
RNA或DNA
化学接头防止sgRNA支架的逆转录
倒置DNA合成模板
RNA
5'
3'
RNA or DNA
5'
在DNA合成模板中切换5’至3’方向
DNA依赖性DNA聚合酶可用于取代逆转录酶

图72C

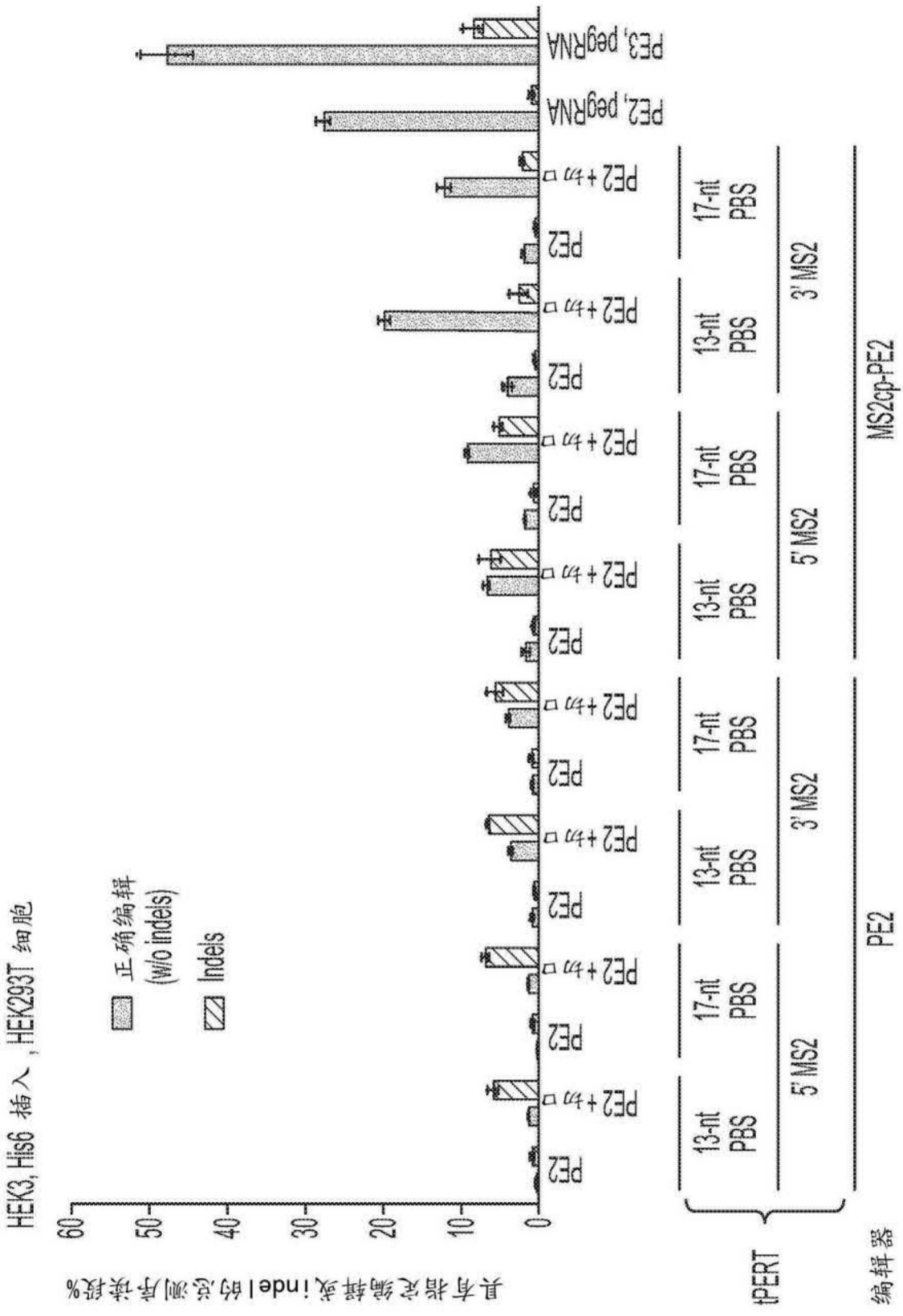

HEK3, His6 插入，HEK293T 细胞
正确编辑 (w/o indels)
Indels
具有指定编辑或 indel 的总测序读段%
0
10
20
30
40
50
60
PE2
PE2 + hMLH1dn
PE3, pegRNA
PE2, pegRNA
13-nt PBS
17-nt PBS
5' MS2
3' MS2
PE2
MS2cp-PE2
tPERT
编辑器

图73

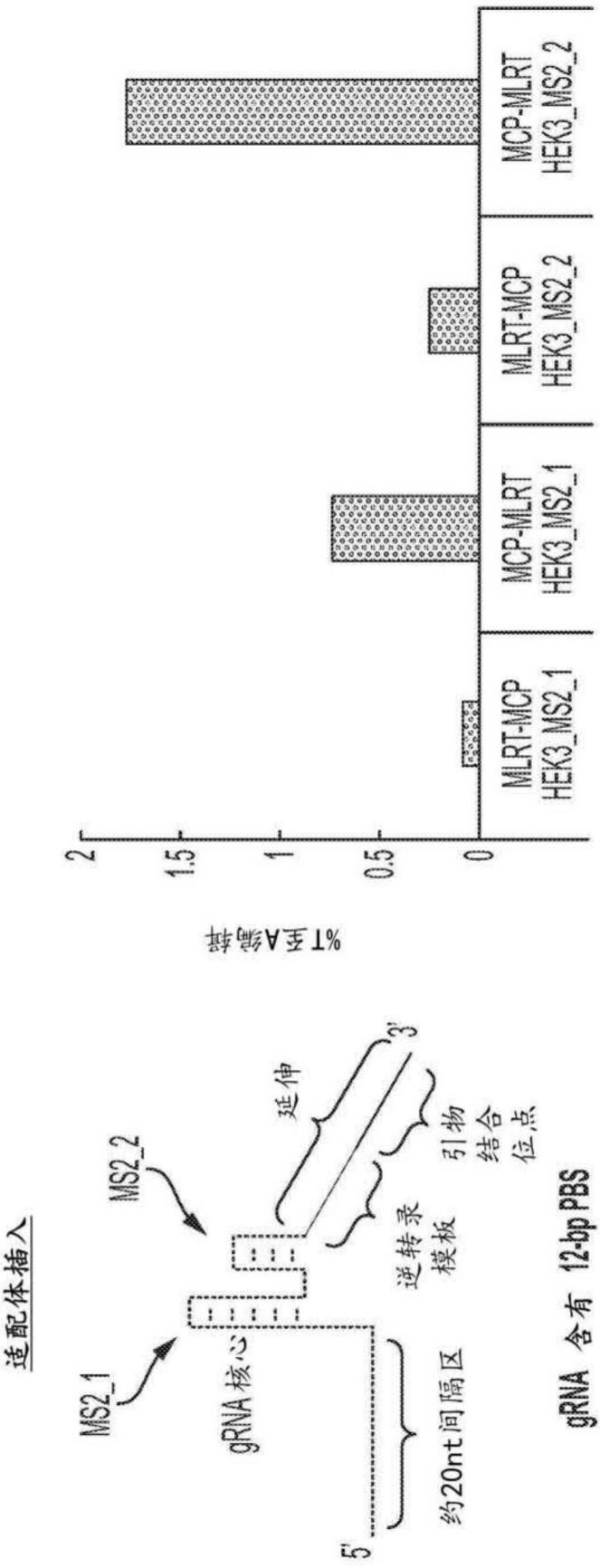

%T至A编辑
2
1.5
1
0.5
0
MLRT-MCP HEK3_MS2_1
MCP-MLRT HEK3_MS2_1
MLRT-MCP HEK3_MS2_2
MCP-MLRT HEK3_MS2_2
适配体插入
MS2_1
MS2_2
gRNA核心
约20nt间隔区
逆转录模板
延伸
引物结合位点
5'
3'
gRNA 含有 12-bp PBS

图74

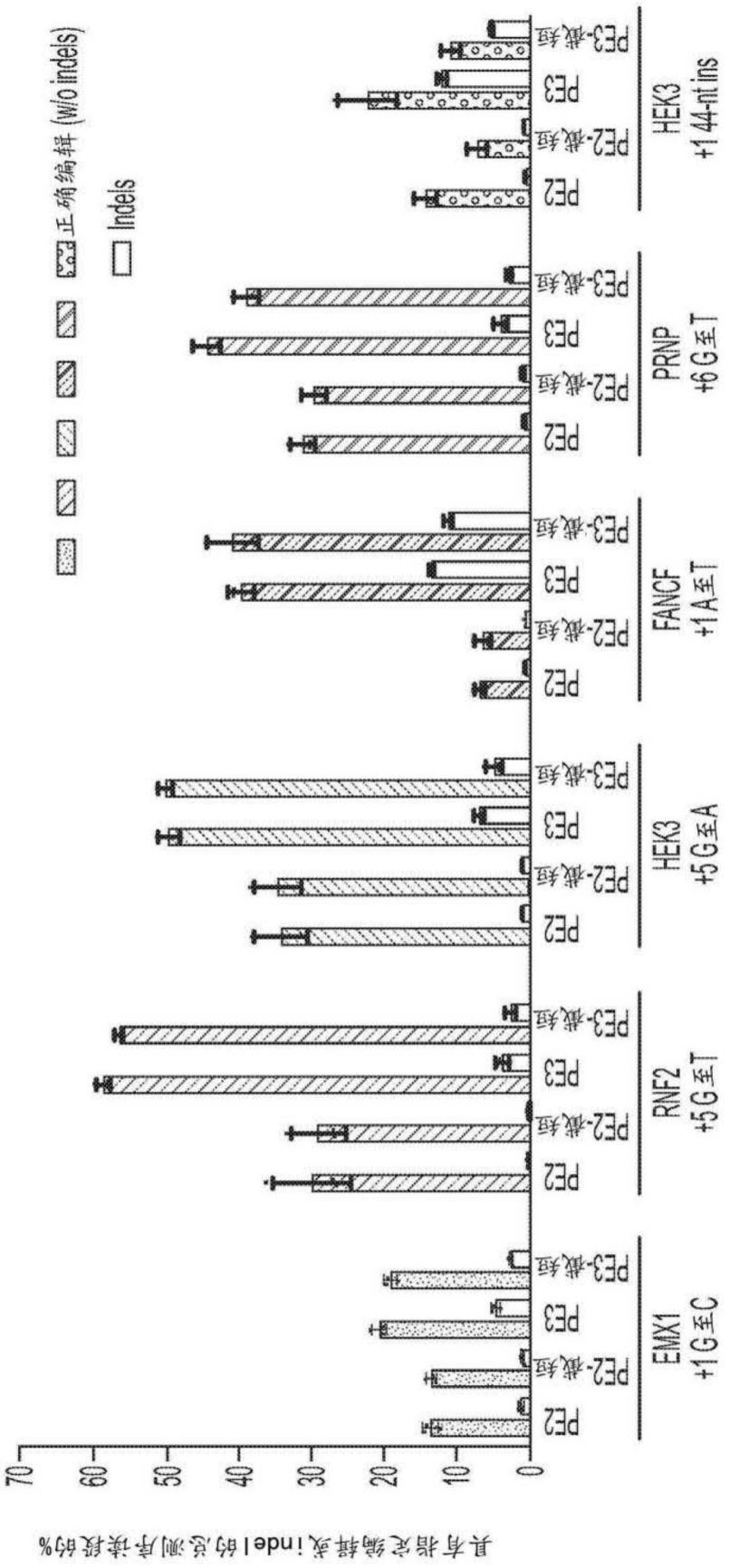

具有指定编辑或indel的总测序读段的%
70
60
50
40
30
20
10
0
Indels
正确编辑 (w/o indels)
EMX1 +1G至C
RNF2 +5G至T
HEK3 +5G至A
FANCF +1A至T
PRNP +6G至T
HEK3 +1 44-nt ins
PE2
PE2-截短
PE3
PE3-截短

图75

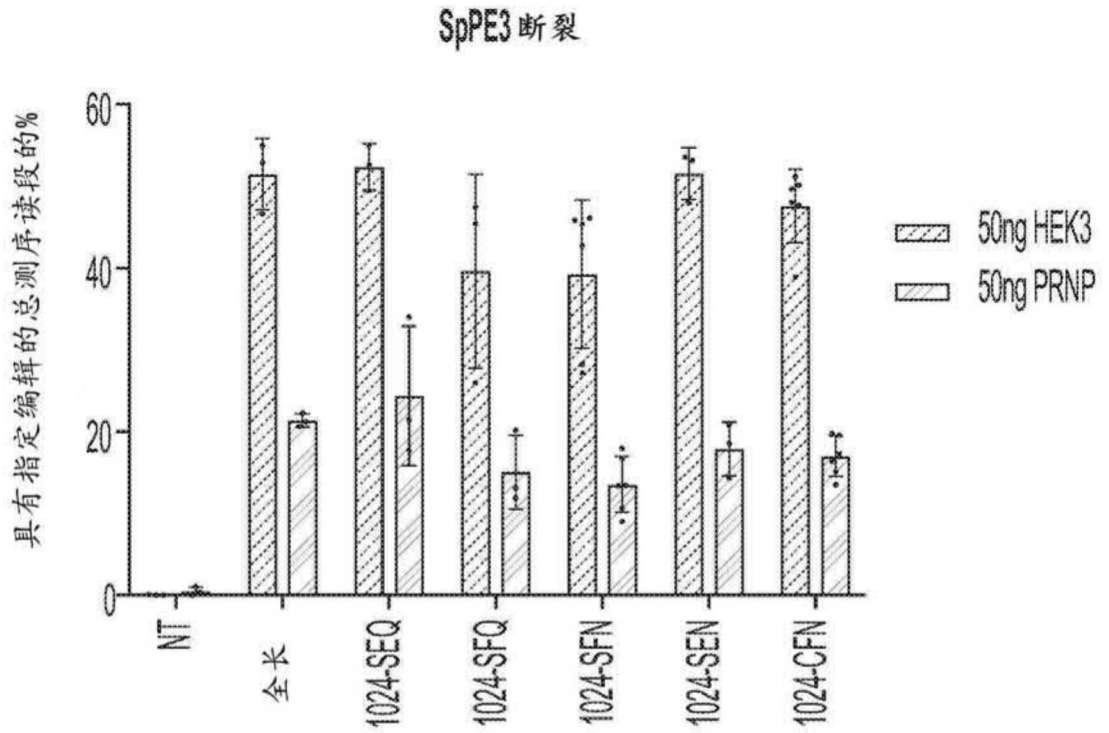

SpPE3断裂
具有指定编辑的总测序读段的%
60
40
20
0
NT
全长
1024-SEQ
1024-SFQ
1024-SFN
1024-SEN
1024-CFN
50ng HEK3
50ng PRNP

图76

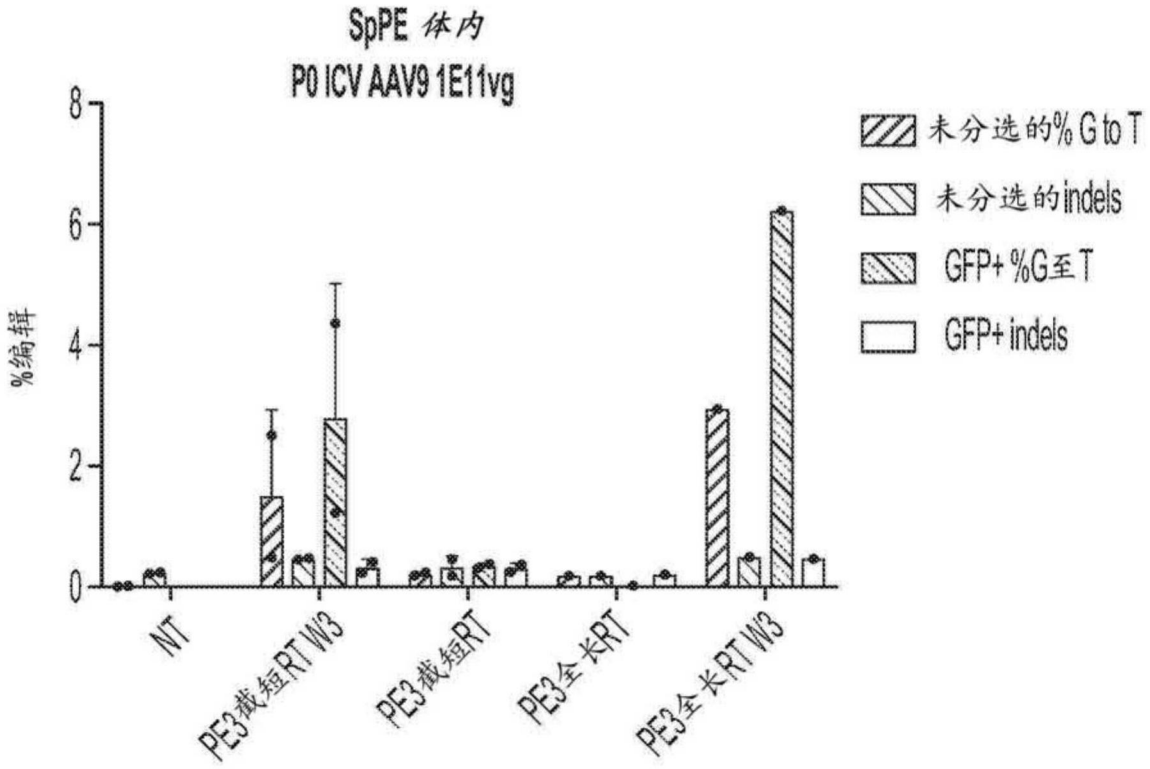

SpPE 体内
P0 ICV AAV9 1E11vg
%编辑
8
6
4
2
0
NT
PE3截短RT W3
PE3截短RT
PE3全长RT
PE3全长RT W3
未分选的% G to T
未分选的indels
GFP+ %G至T
GFP+ indels

图77

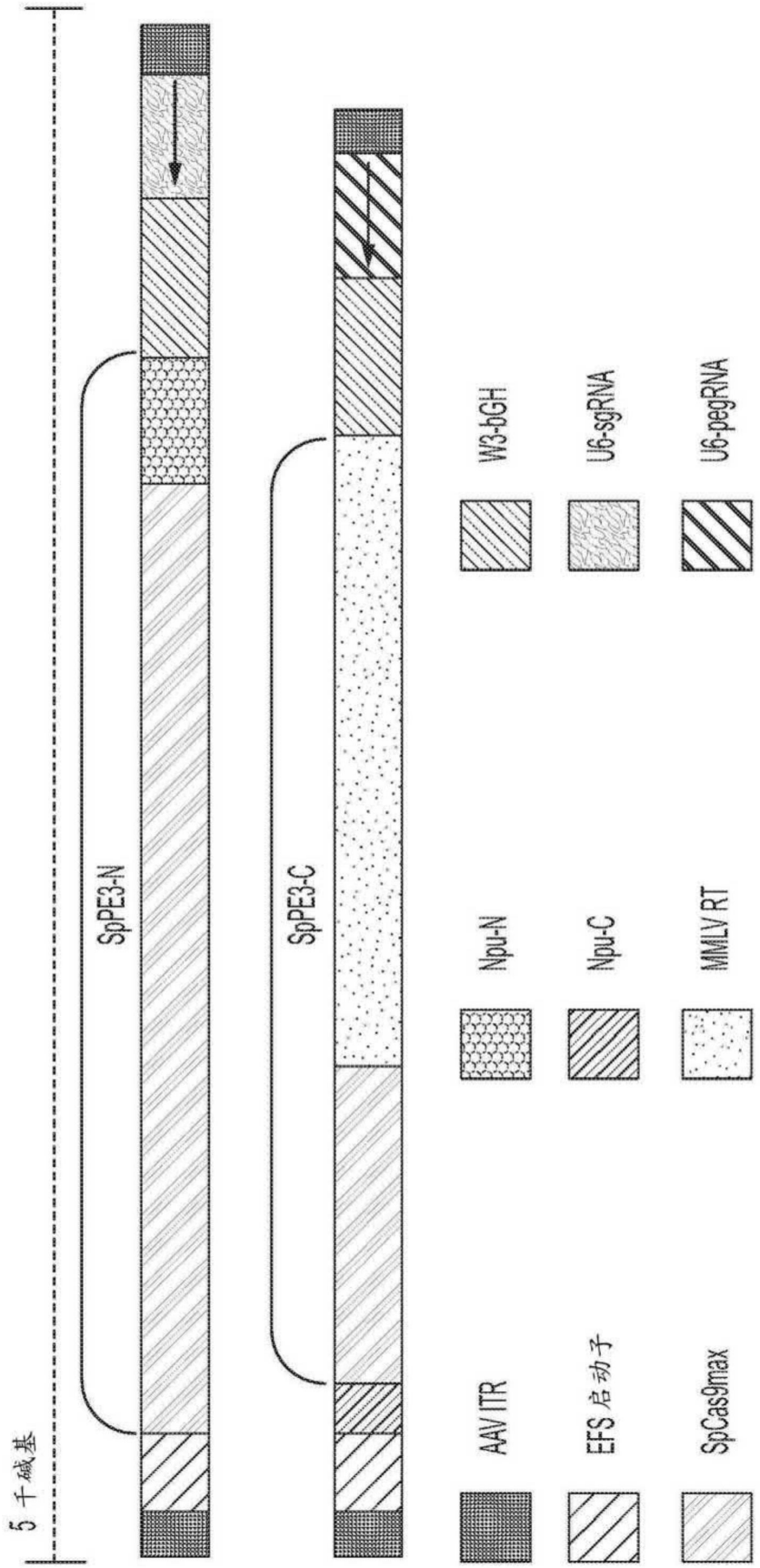

5 千碱基
SpPE3-N
SpPE3-C
AAV ITR
EFS 启动子
SpCas9max
Npu-N
Npu-C
MMLV RT
W3-bGH
U6-sgRNA
U6-pegRNA

图78

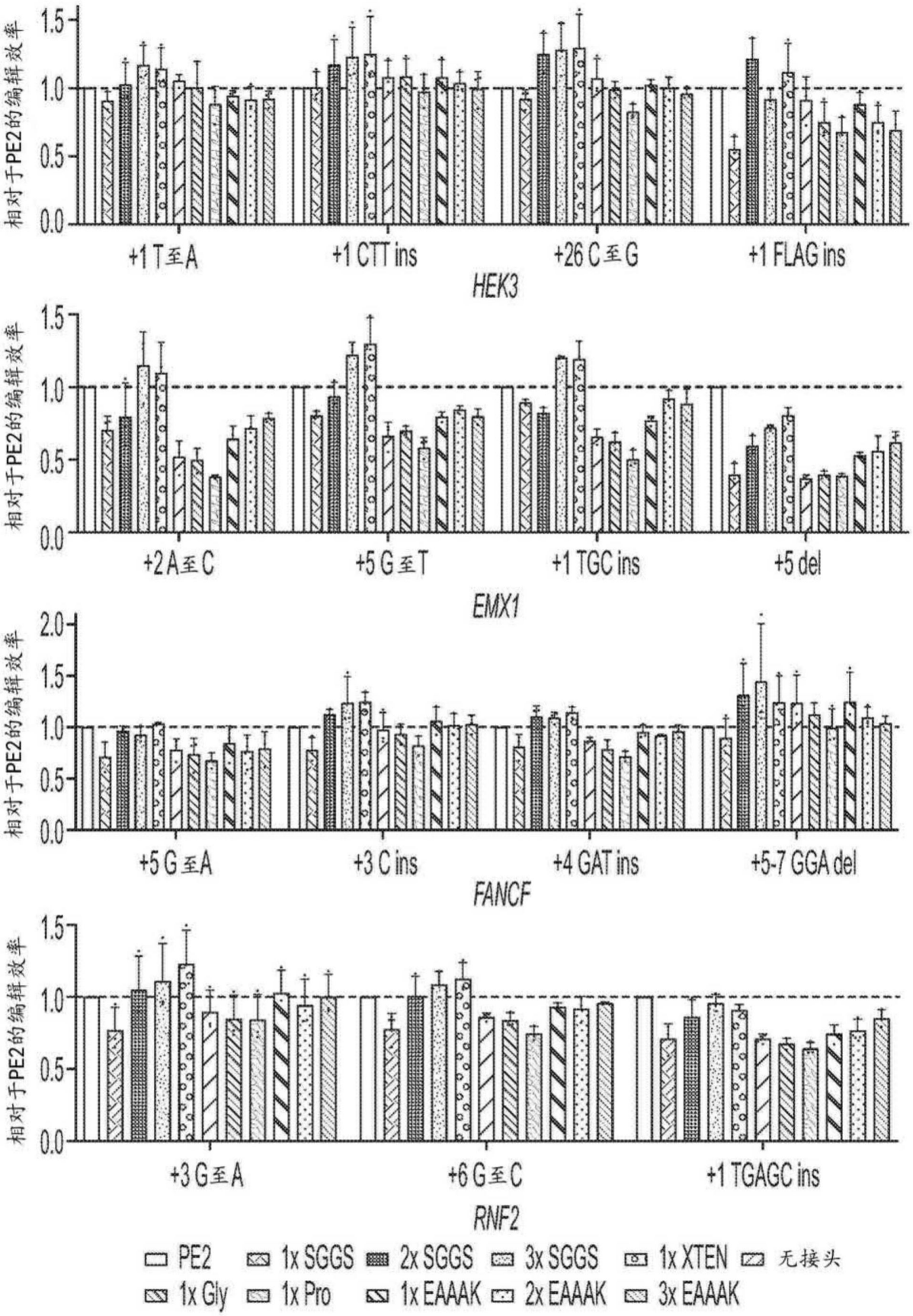

相对于PE2的编辑效率
1.5
1.0
0.5
0.0
+1 T至A
+1 CTT ins
+26 C至G
+1 FLAG ins
HEK3
+2 A至C
+5 G至T
+1 TGC ins
+5 del
EMX1
2.0
+5 G至A
+3 C ins
+4 GAT ins
+5-7 GGA del
FANCF
+3 G至A
+6 G至C
+1 TGAGC ins
RNF2
PE2
1x SGGS
2x SGGS
3x SGGS
1x XTEN
无接头
1x Gly
1x Pro
1x EAAAK
2x EAAAK
3x EAAAK

图79

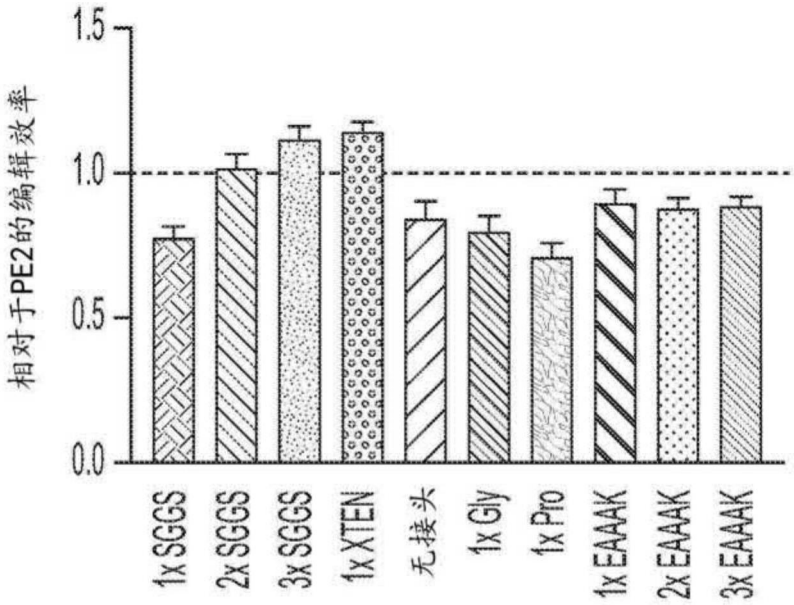

相对于PE2的编辑效率
1.5
1.0
0.5
0.0
1x SGGS
2x SGGS
3x SGGS
1x XTEN
无接头
1x Gly
1x Pro
1x EAAAK
2x EAAAK
3x EAAAK

图80

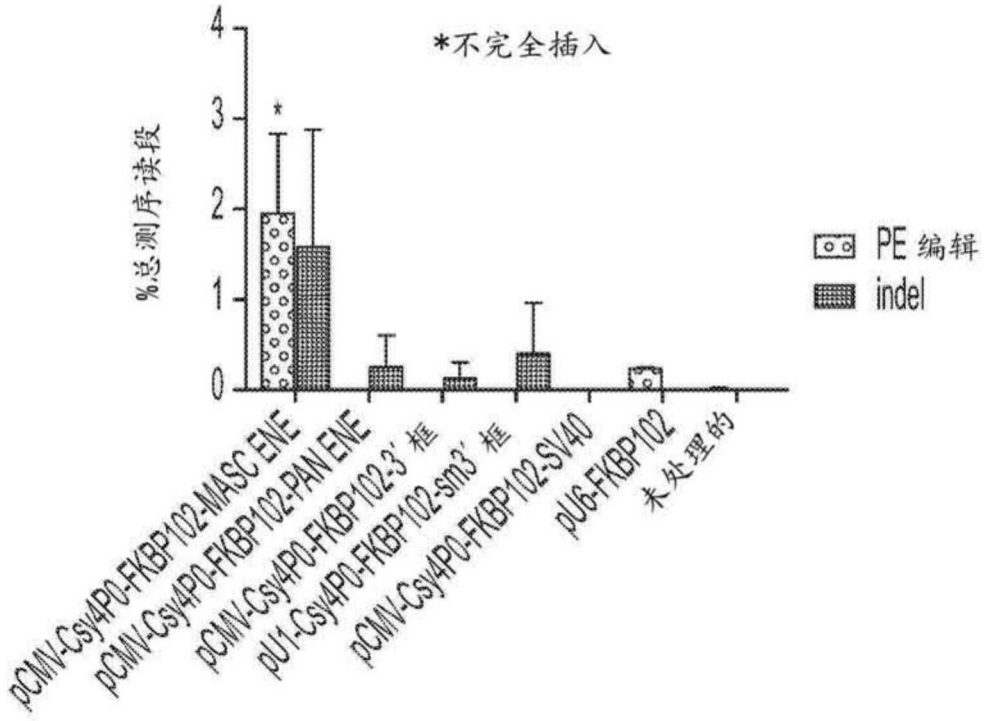

*不完全插入
%总测序读段
4
3
2
1
0
*
PE 编辑
indel
pCMV-Csy4P0-FKBP102-MASC ENE
pCMV-Csy4P0-FKBP102-PAN ENE
pCMV-Csy4P0-FKBP102-3' 框
pU1-Csy4P0-FKBP102-sm3' 框
pCMV-Csy4P0-FKBP102-SV40
pU6-FKBP102
未处理的

图81

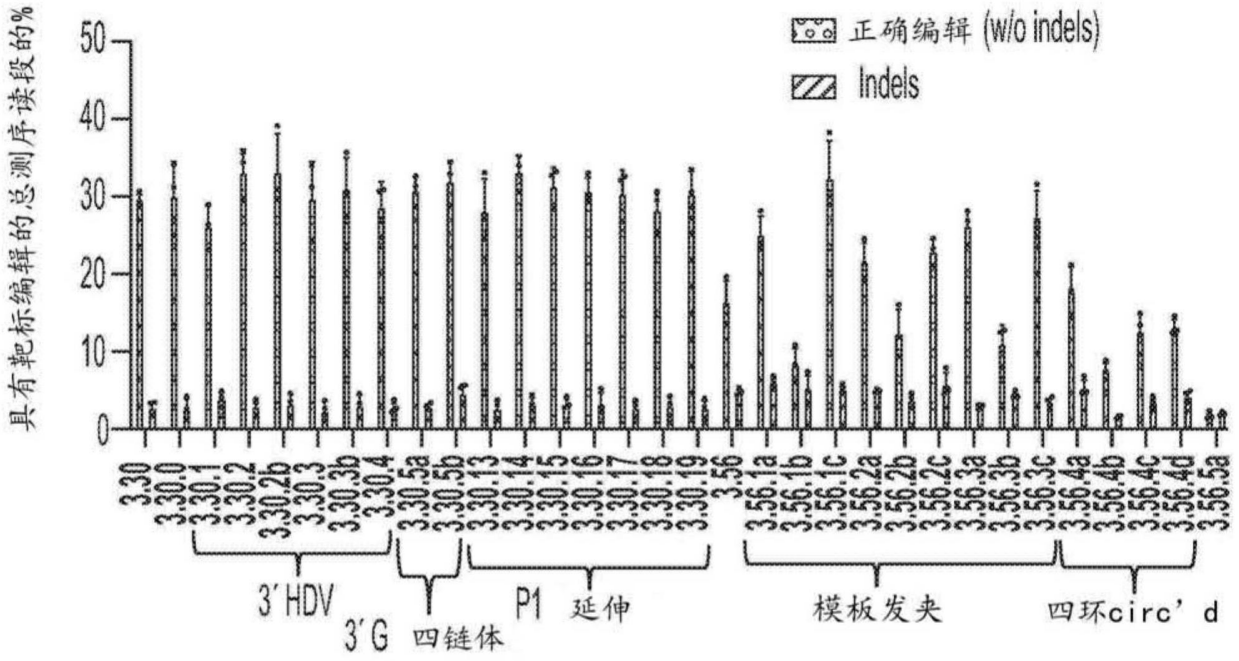

具有靶标编辑的总测序读段的%
正确编辑 (w/o indels)
Indels
3′HDV
3′G 四链体
P1 延伸
模板发夹
四环circ' d

图82

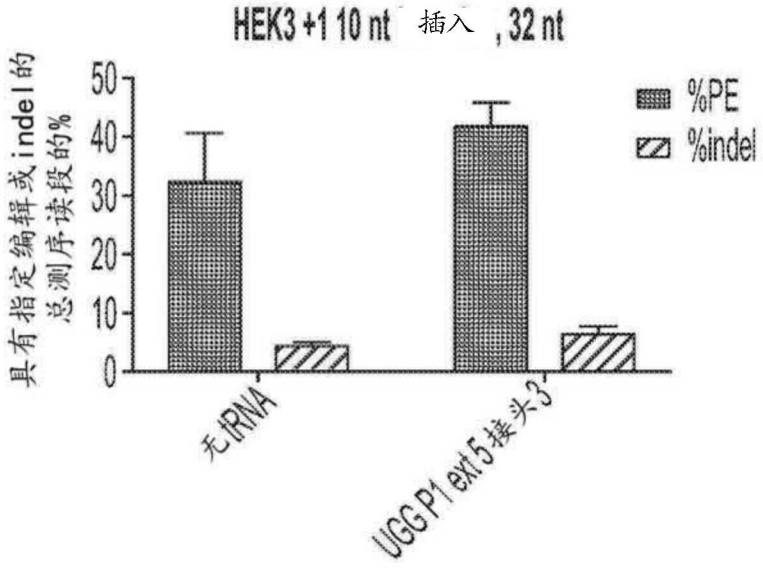

HEK3 +1 10 nt 插入 , 32 nt
具有指定编辑或indel的总测序读段的%
%PE
%indel
无tRNA
UGG P1 ext 5接头3'

图83

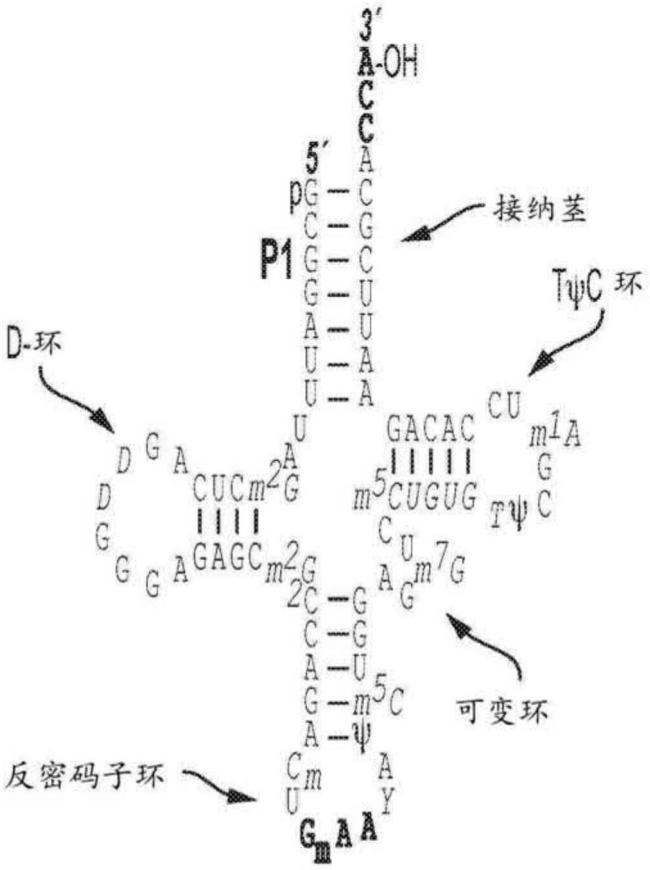

3′
A-OH
5′
P1
接纳茎
TψC 环
D-环
可变环
反密码子环

图84A

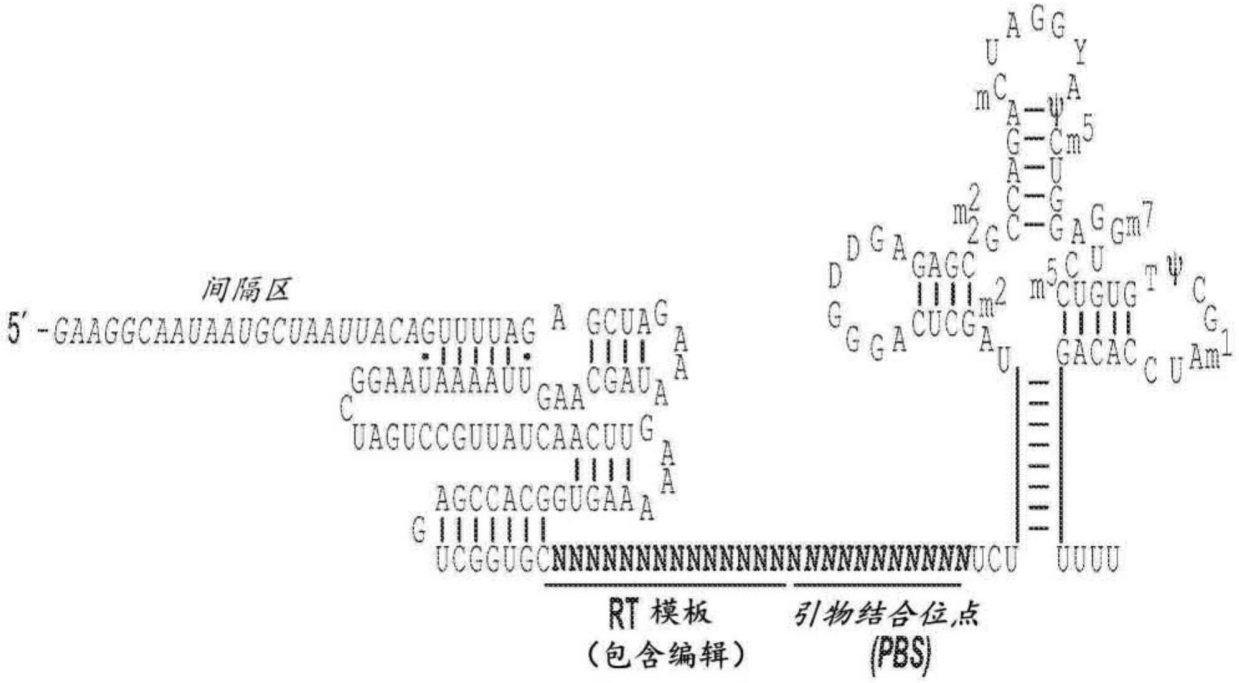

间隔区
5′-GAAGGCAAUAAUGCUAAUUACAGUUUUAG
RT 模板
（包含编辑）
引物结合位点
(PBS)

图84B

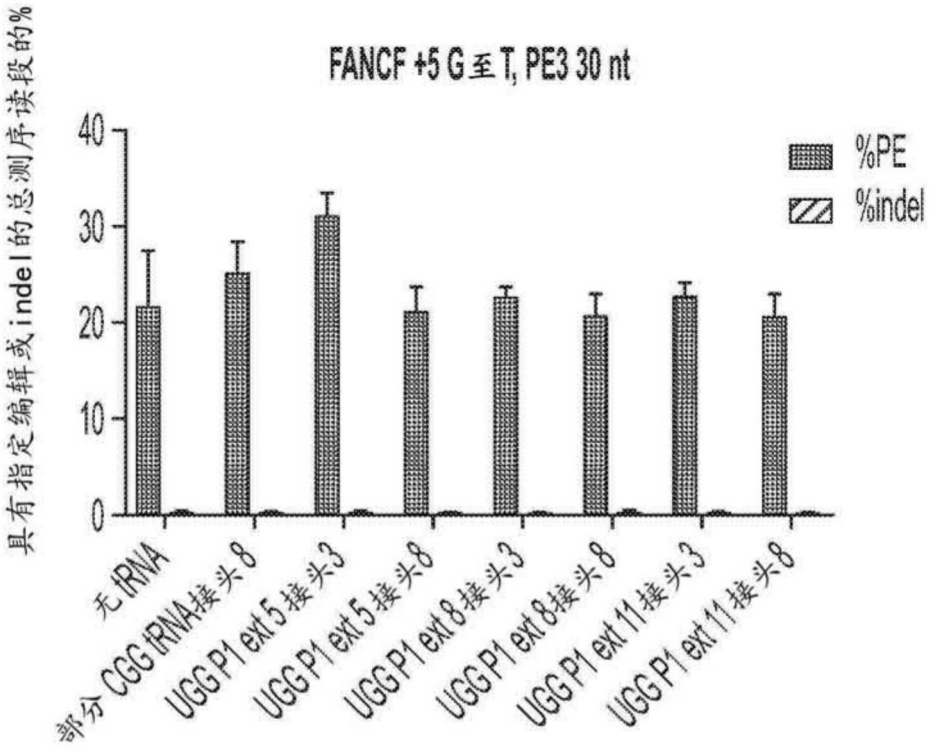

FANCF +5 G至T, PE3 30 nt
具有指定编辑或indel的总测序读段的%
%PE
%indel
0
10
20
30
40
无tRNA
部分CGG tRNA接头8
UGG P1 ext 5接头3
UGG P1 ext 5接头8
UGG P1 ext 8接头3
UGG P1 ext 8接头8
UGG P1 ext 11接头3
UGG P1 ext 11接头8

图85

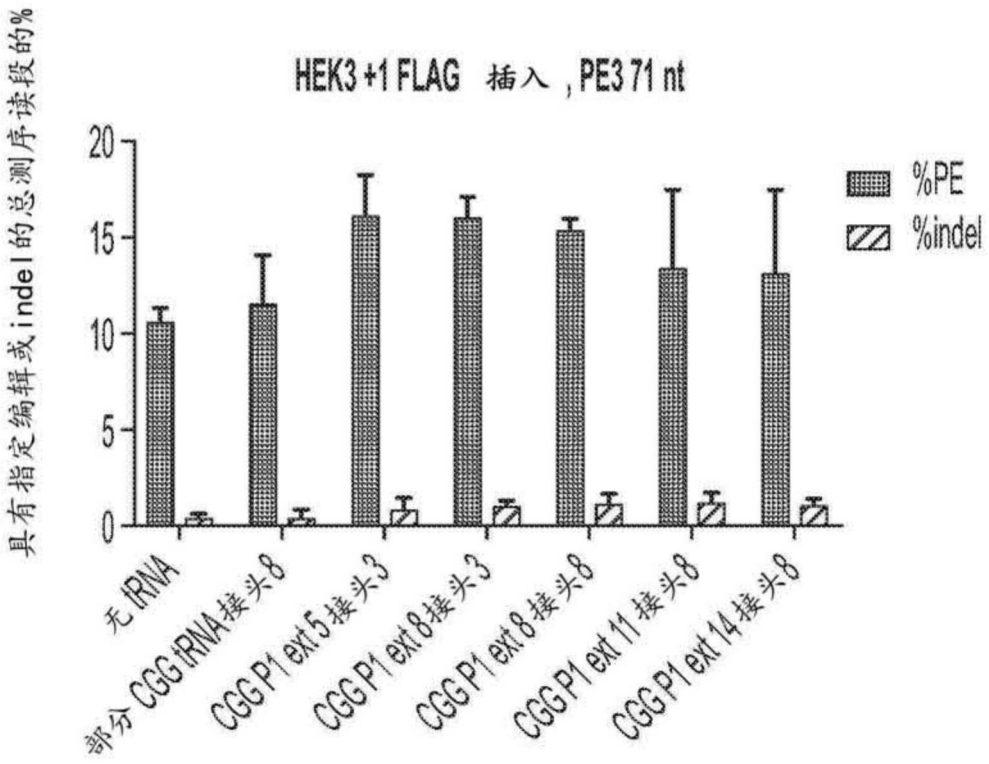

HEK3 +1 FLAG 插入, PE3 71 nt
具有指定编辑或indel的总测序读段的%
%PE
%indel
0
5
10
15
20
无tRNA
部分CGG tRNA接头8
CGG P1 ext 5接头3
CGG P1 ext 8接头3
CGG P1 ext 8接头8
CGG P1 ext 11接头8
CGG P1 ext 14接头8

图86

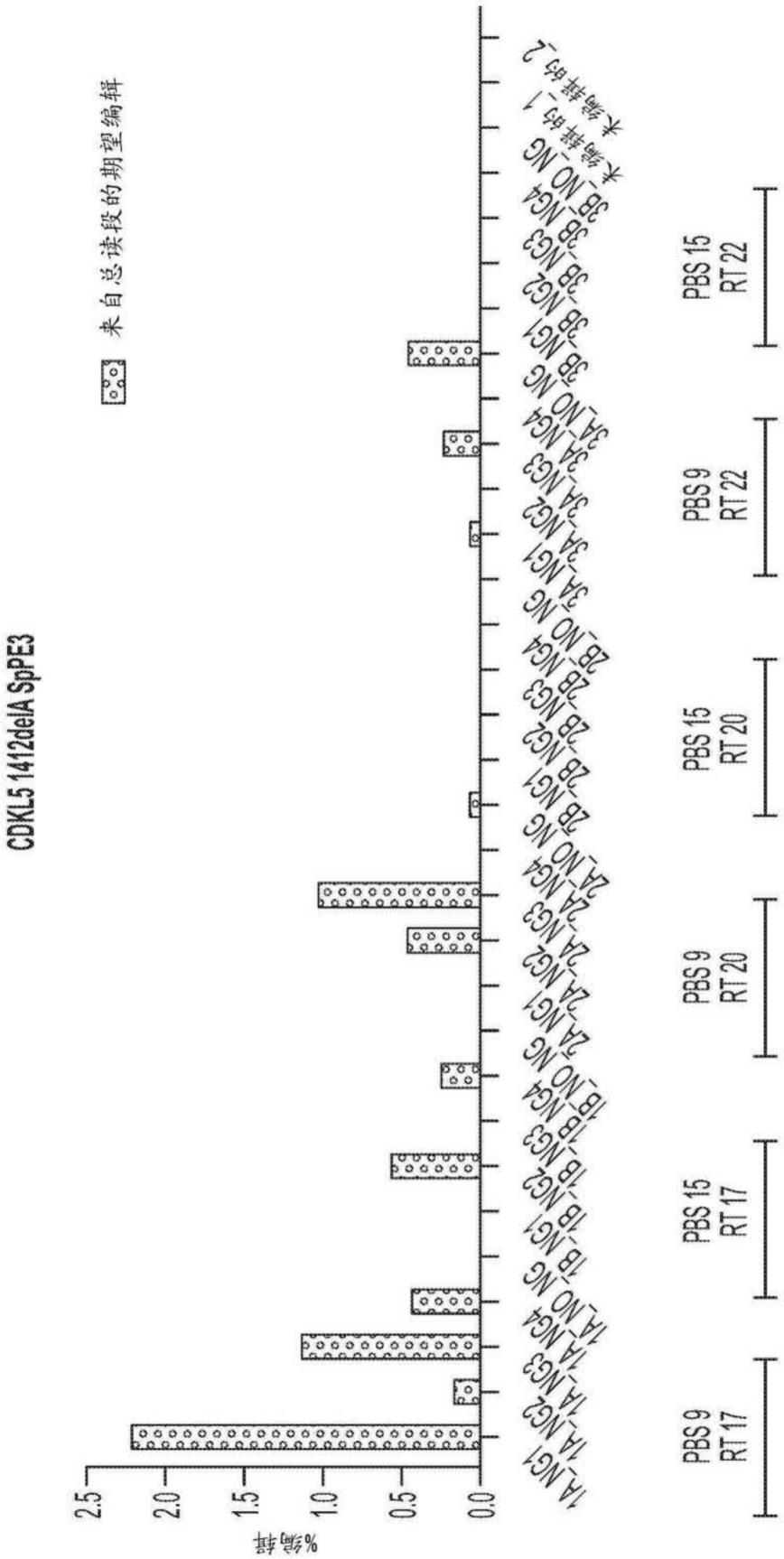

CDKL5 1412delA SpPE3
%编辑
2.5
2.0
1.5
1.0
0.5
0.0
来自总读段的期望编辑
未编辑的_1
未编辑的_2
38_NO_NG
38_NG4
38_NG3
38_NG2
38_NG1
34_NO_NG
34_NG4
34_NG3
34_NG2
34_NG1
28_NO_NG
28_NG4
28_NG3
28_NG2
28_NG1
24_NO_NG
24_NG4
24_NG3
24_NG2
24_NG1
18_NO_NG
18_NG4
18_NG3
18_NG2
18_NG1
14_NO_NG
14_NG4
14_NG3
14_NG2
14_NG1
PBS 9
RT 17
PBS 15
RT 17
PBS 9
RT 20
PBS 15
RT 20
PBS 9
RT 22
PBS 15
RT 22

图87

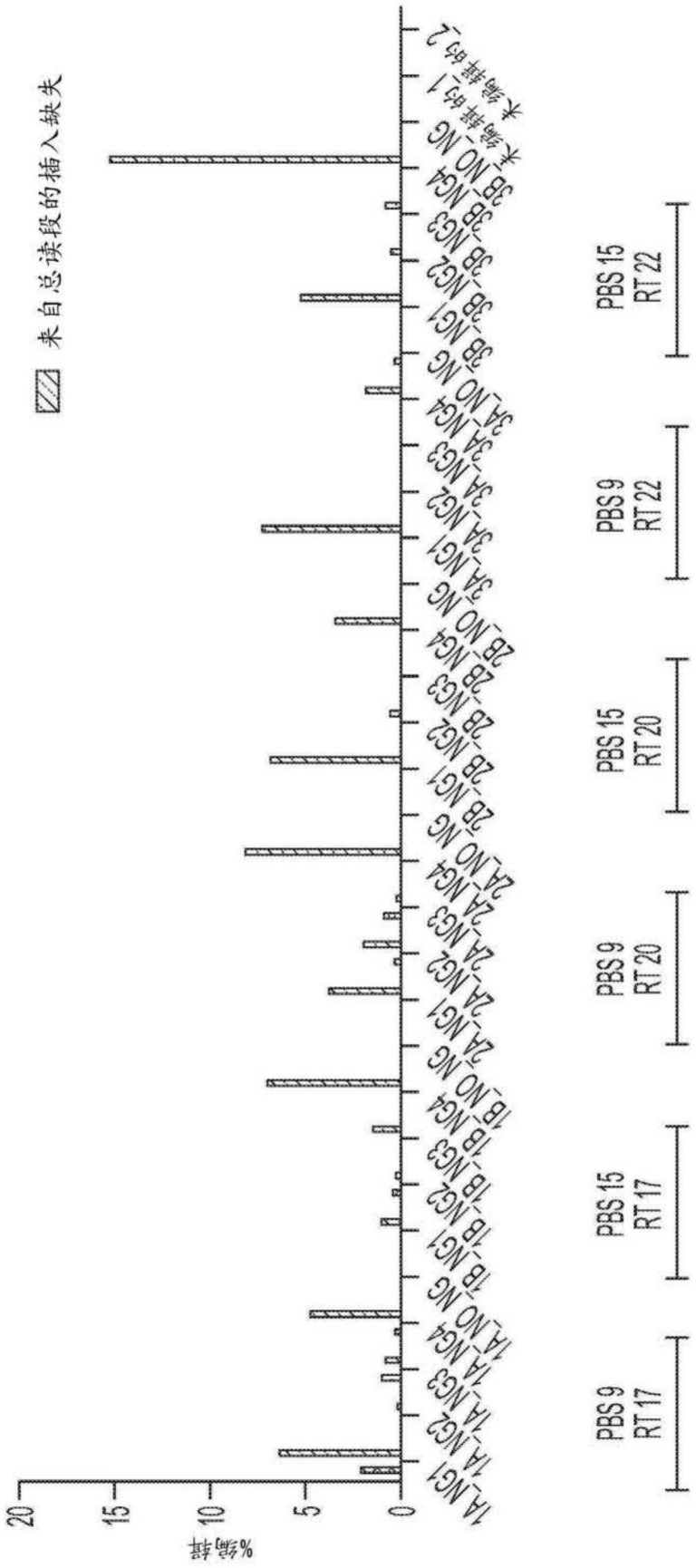

CDKL5 1412delA SpPE3
来自总读段的期望编辑
来自总读段的插入缺失
%编辑
20
15
10
5
0
未编辑的_1
未编辑的_2
1A_NO_NG
1A_1A_NG1
1A_1A_NG2
1A_1A_NG3
1A_1A_NG4
1B_NO_NG
1B_1B_NG1
1B_1B_NG2
1B_1B_NG3
1B_1B_NG4
2A_NO_NG
2A_2A_NG1
2A_2A_NG2
2A_2A_NG3
2A_2A_NG4
2B_NO_NG
2B_2B_NG1
2B_2B_NG2
2B_2B_NG3
2B_2B_NG4
3A_NO_NG
3A_3A_NG1
3A_3A_NG2
3A_3A_NG3
3A_3A_NG4
3B_NO_NG
3B_3B_NG1
3B_3B_NG2
3B_3B_NG3
3B_3B_NG4
PBS 9
RT 17
PBS 15
RT 17
PBS 9
RT 20
PBS 15
RT 20
PBS 9
RT 22
PBS 15
RT 22

图88

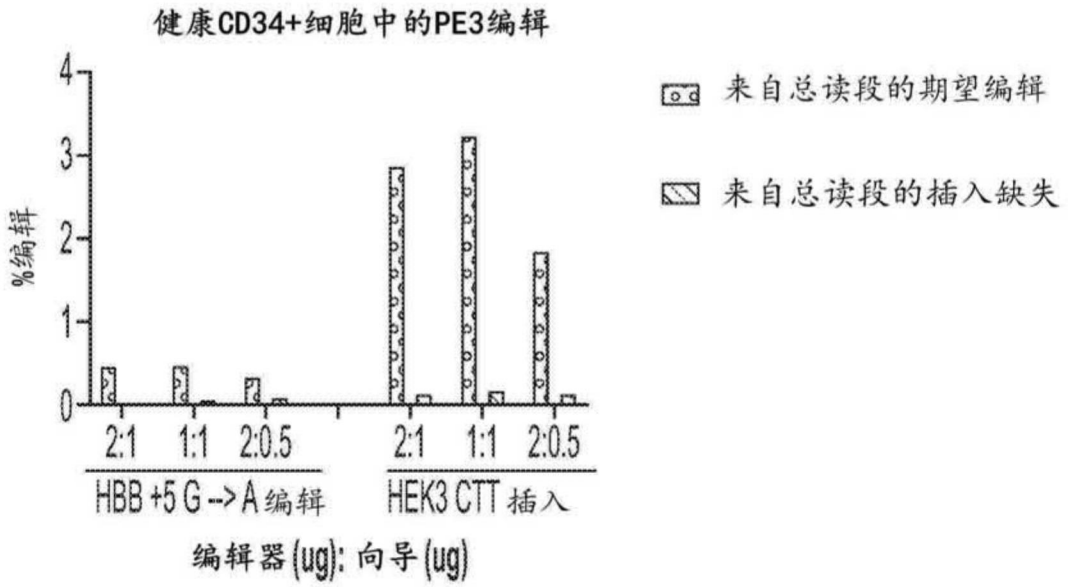

健康CD34+细胞中的PE3编辑
%编辑
4
3
2
1
0
2:1
1:1
2:0.5
2:1
1:1
2:0.5
HBB +5 G --> A 编辑
HEK3 CTT 插入
编辑器(ug): 向导(ug)
来自总读段的期望编辑
来自总读段的插入缺失

图89

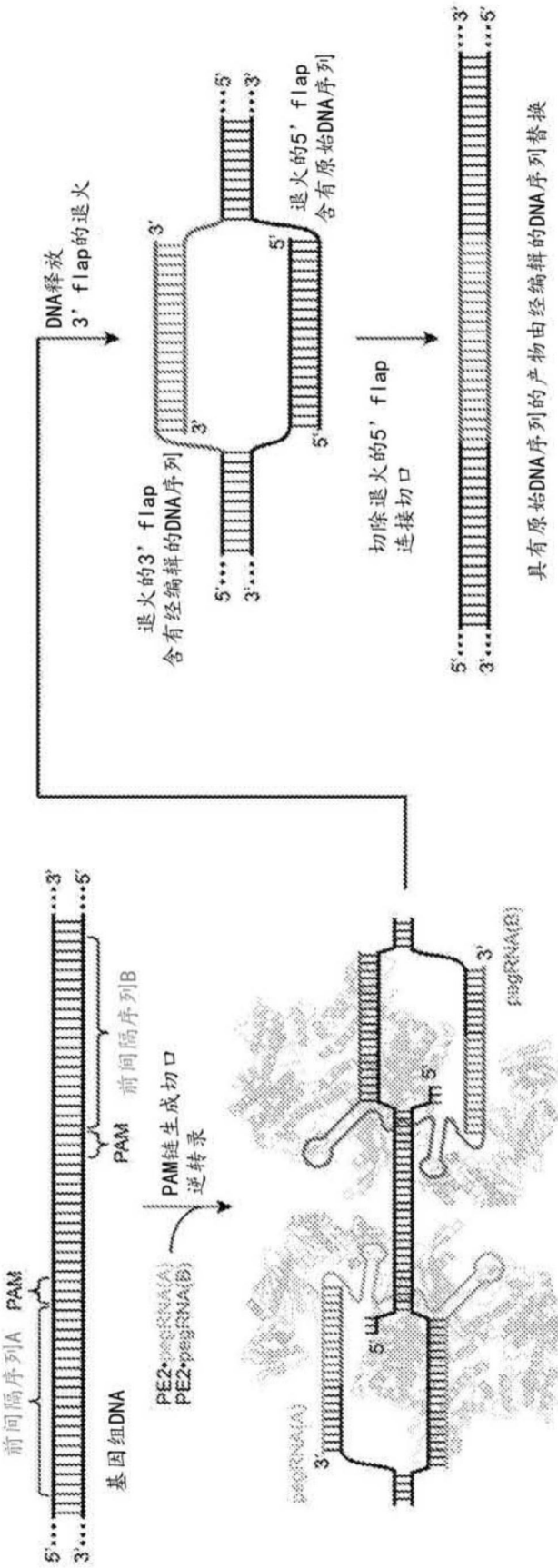

前间隔序列A
PAM
基因组DNA
PE2•pegRNA(A)
PE2•pegRNA(B)
PAM链生成切口
逆转录
前间隔序列B
PAM
DNA释放
3' flap的退火
退火的3' flap
含有经编辑的DNA序列
退火的5' flap
含有原始DNA序列
切除退火的5' flap
连接切口
具有原始DNA序列的产物由经编辑的DNA序列替换

图90

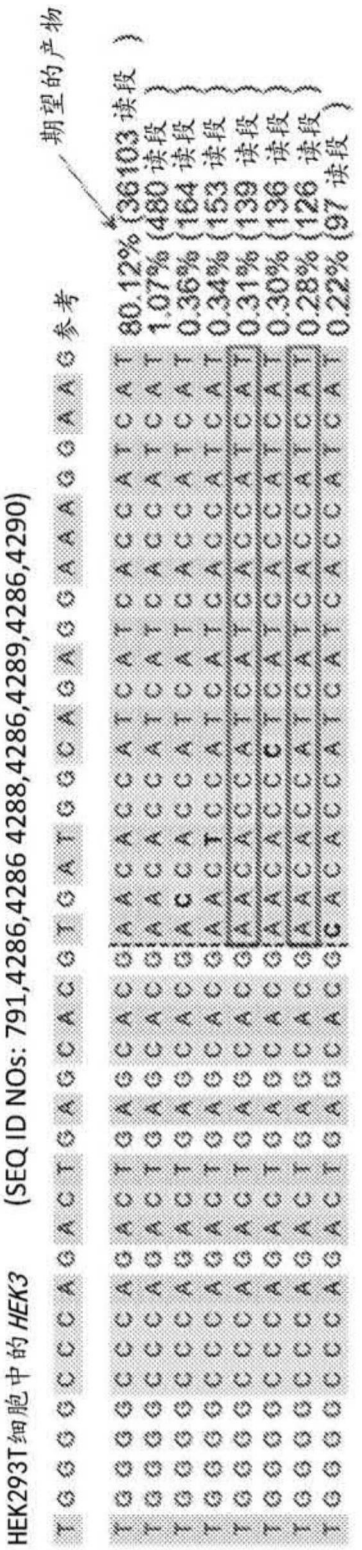

HEK293T细胞中的 HEK3
(SEQ ID NOs: 791,4286,4286 4288,4286,4289,4286,4290)
期望的产物
参考
80.12% {36103 读段 }
1.07% {480 读段 }
0.36% {164 读段 }
0.34% {153 读段 }
0.31% {139 读段 }
0.30% {136 读段 }
0.28% {126 读段 }
0.22% {97 读段 }
加粗 取代
插入
缺失
预测的切割位置

图91A

pegRNA1: (SEQ ID NO: 4291)

5' – GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC

CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC**AAATGATGGTGATGGTGTTCGTGCTCAGTCTG – 3'**

pegRNA2: (SEQ ID NO: 4292)

5' – GTCAACCAGTATCCCGGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC

CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC**AACACCATCATCACCATCATTTCCGGGATACTGG – 3'**

图91B

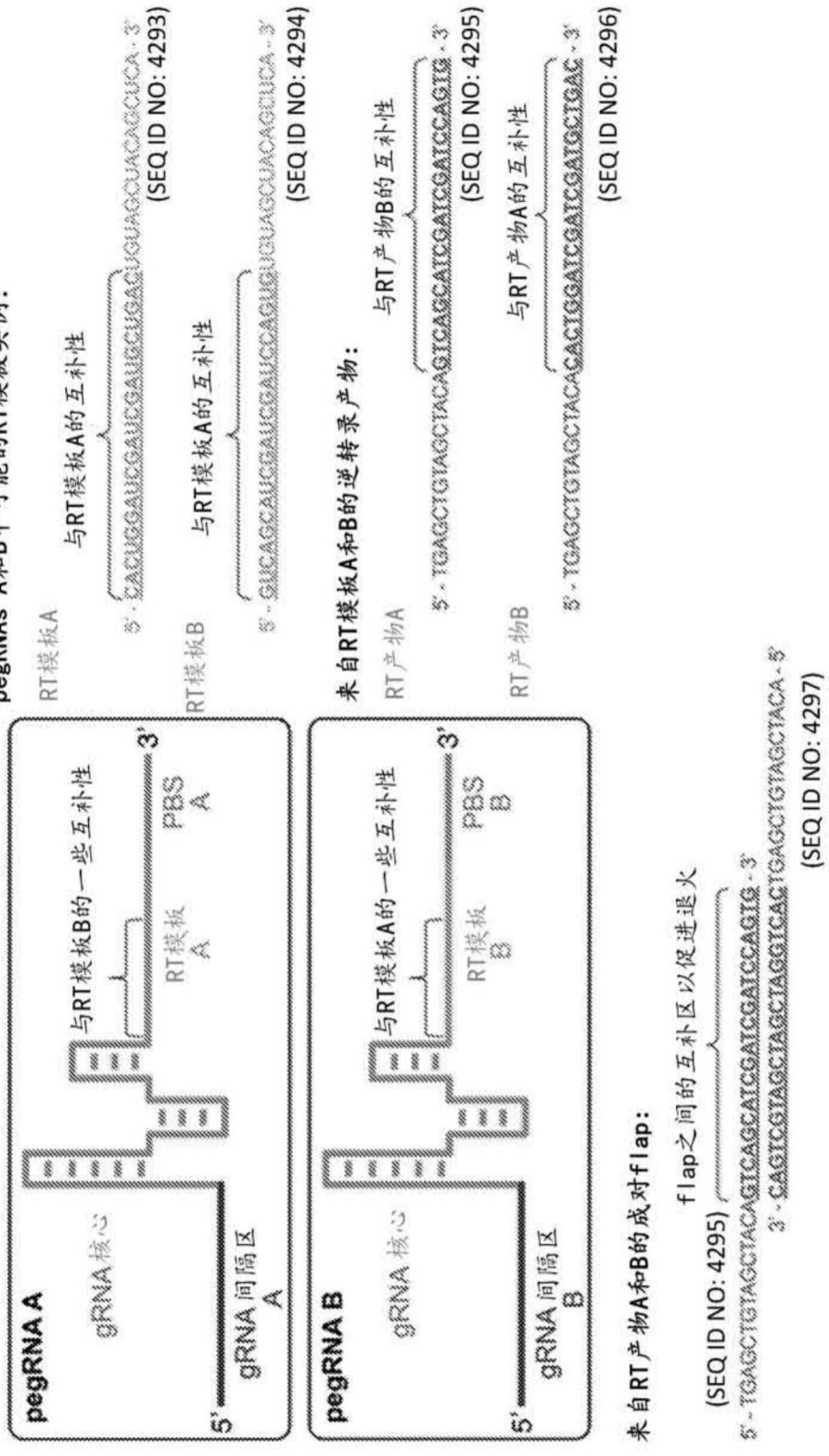

pegRNAs A和B中可能的RT模板实例：
RT模板A
与RT模板A的互补性
5'-CACUGGAUCGAUCGAUCUGACUGUAGCUACAGCUUCA-3'
(SEQ ID NO: 4293)
RT模板B
与RT模板A的互补性
5'-GUCAGCAAUCGAUCGAUCCAGUGCUAGCUACAGCUUCA-3'
(SEQ ID NO: 4294)
pegRNA A
gRNA核心
与RT模板B的一些互补性
3'
RT模板 A
PBS A
5'
gRNA 间隔区 A
pegRNA B
gRNA核心
与RT模板A的一些互补性
3'
RT模板 B
PBS B
5'
gRNA 间隔区 B
来自RT模板A和B的逆转录产物：
RT产物A
与RT产物B的互补性
5'-TGAGCTGTAGCTACAGTCAGCAATCGATCGATCCAGTG-3'
(SEQ ID NO: 4295)
RT产物B
与RT产物A的互补性
5'-TGAGCTGTAGCTACACACTGGATCGATCGATTGCTGAC-3'
(SEQ ID NO: 4296)
来自RT产物A和B的成对flap：
flap之间的互补区以促进退火
(SEQ ID NO: 4295)
5'-TGAGCTGTAGCTACAGTCAGCAATCGATCGATCCAGTG-3'
3'-CAGTCGTAGCTAGCTAGGTCAGCTGAGCTGTAGCTACA-5'
(SEQ ID NO: 4297)

图92

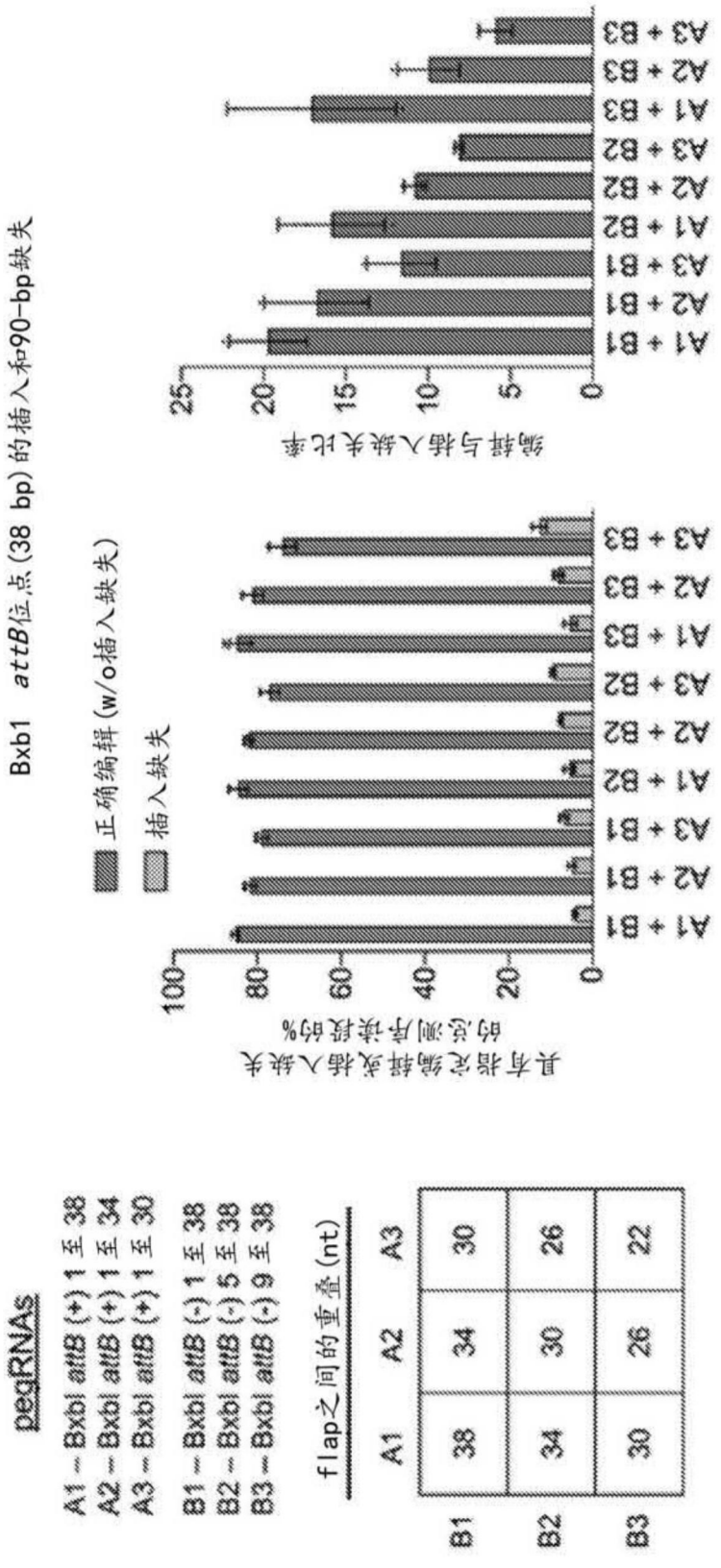

Bxb1 attB位点(38 bp)的插入和90-bp缺失
正确编辑(w/o插入缺失)
插入缺失
具有指定编辑或插入缺失的总测序读段的%
编辑与插入缺失比率
A1 + B1
A2 + B1
A3 + B1
A1 + B2
A2 + B2
A3 + B2
A1 + B3
A2 + B3
A3 + B3
pegRNAs
A1 – Bxb1 attB (+) 1 至 38
A2 – Bxb1 attB (+) 1 至 34
A3 – Bxb1 attB (+) 1 至 30
B1 – Bxb1 attB (-) 1 至 38
B2 – Bxb1 attB (-) 5 至 38
B3 – Bxb1 attB (-) 9 至 38
flap之间的重叠(nt)
A1 A2 A3
B1 38 34 30
B2 34 30 26
B3 30 26 22

图93A

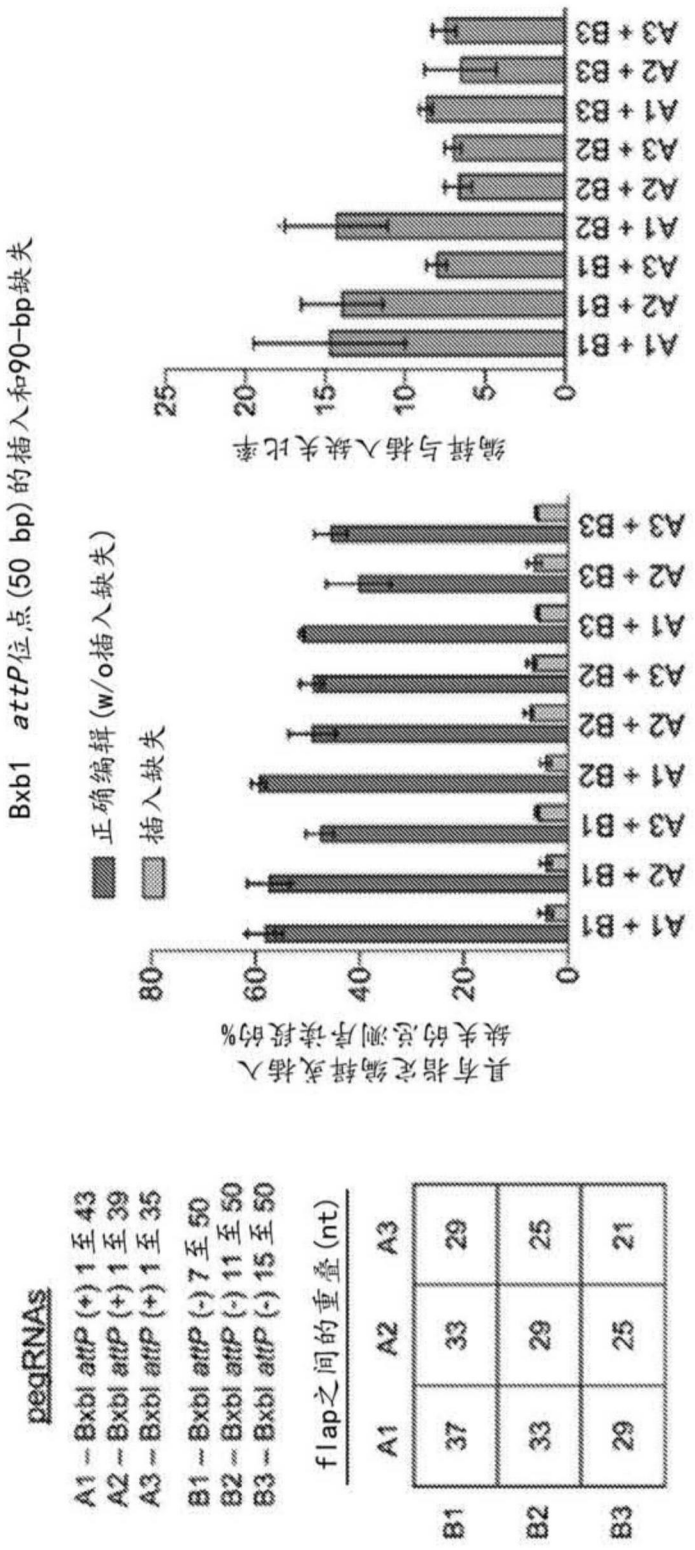

Bxb1 attP位点(50 bp)的插入和90-bp缺失
正确编辑(w/o插入缺失)
插入缺失
具有指定编辑或插入缺失的总测序读段的%
编辑与插入缺失比率
A1 + B1
A2 + B1
A3 + B1
A1 + B2
A2 + B2
A3 + B2
A1 + B3
A2 + B3
A3 + B3
pegRNAs
A1 – Bxb1 attP (+) 1 至 43
A2 – Bxb1 attP (+) 1 至 39
A3 – Bxb1 attP (+) 1 至 35
B1 – Bxb1 attP (-) 7 至 50
B2 – Bxb1 attP (-) 11 至 50
B3 – Bxb1 attP (-) 15 至 50
flap之间的重叠(nt)
A1 A2 A3
B1 37 33 29
B2 33 29 25
B3 29 25 21

图93B

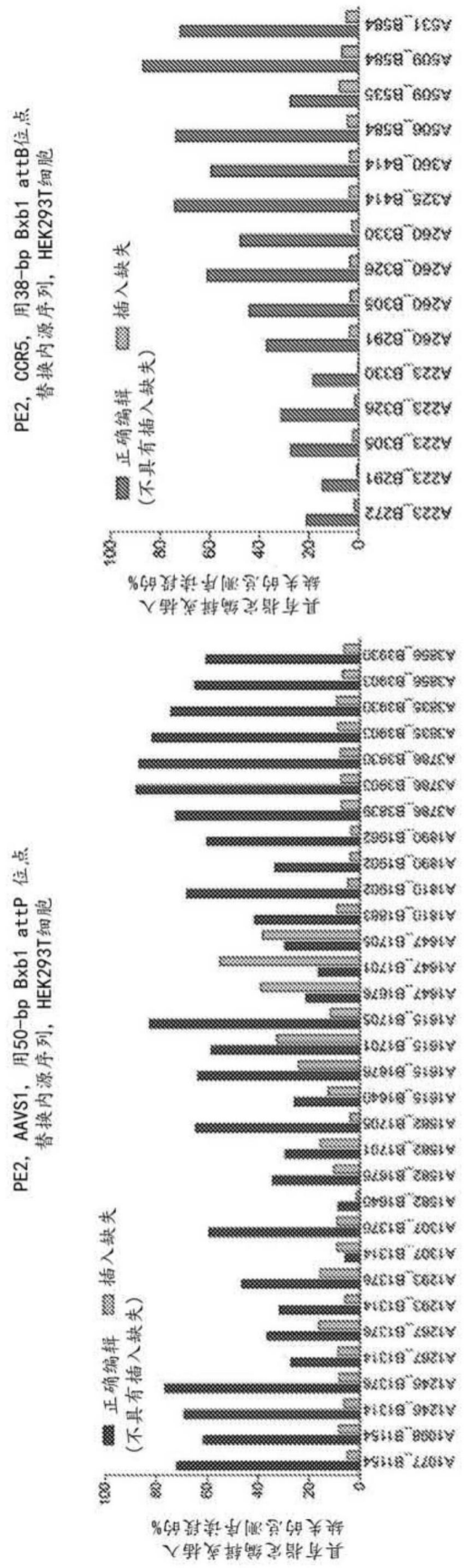

PE2，AAVS1，用50-bp Bxb1 attP 位点
替换内源序列，HEK293T细胞
具有指定编辑或插入缺失的总测序读段的%
正确编辑（不具有插入缺失）
插入缺失
PE2，CCR5，用38-bp Bxb1 attB位点
替换内源序列，HEK293T细胞
具有指定编辑或插入缺失的总测序读段的%
正确编辑（不具有插入缺失）
插入缺失

图94

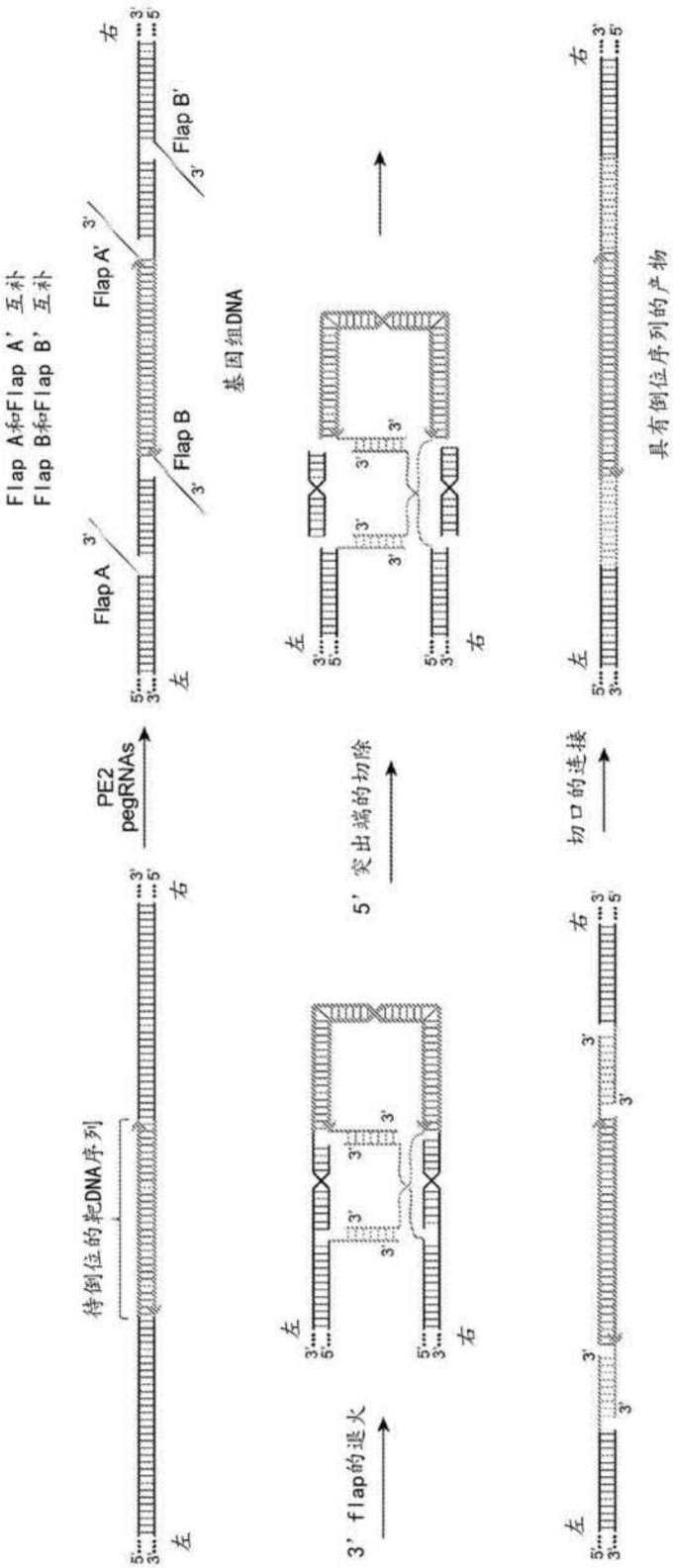

Flap A和Flap A' 互补
Flap B和Flap B' 互补
Flap A
Flap B
Flap A'
Flap B'
基因组DNA
左
右
PE2
pegRNAs
待倒位的靶DNA序列
3' flap的退火
5' 突出端的切除
切口的连接
具有倒位序列的产物

图95

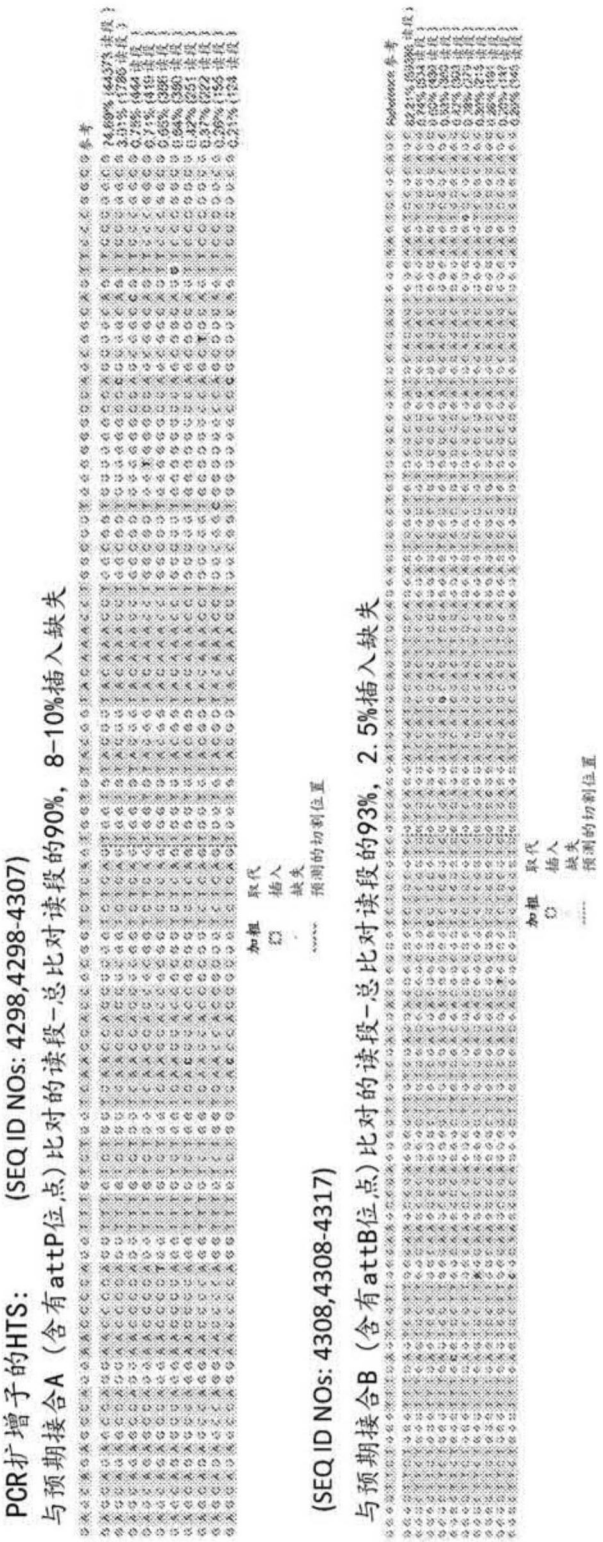

PCR扩增子的HTS: (SEQ ID NOs: 4298,4298-4307)
与预期接合A（含有attP位点）比对的读段-总比对读段的90%，8-10%插入缺失
加框 取代
凸 插入
缺失
..... 预测的切割位置
(SEQ ID NOs: 4308,4308-4317)
与预期接合B（含有attB位点）比对的读段-总比对读段的93%，2.5%插入缺失
加框 取代
凸 插入
缺失
..... 预测的切割位置

图96

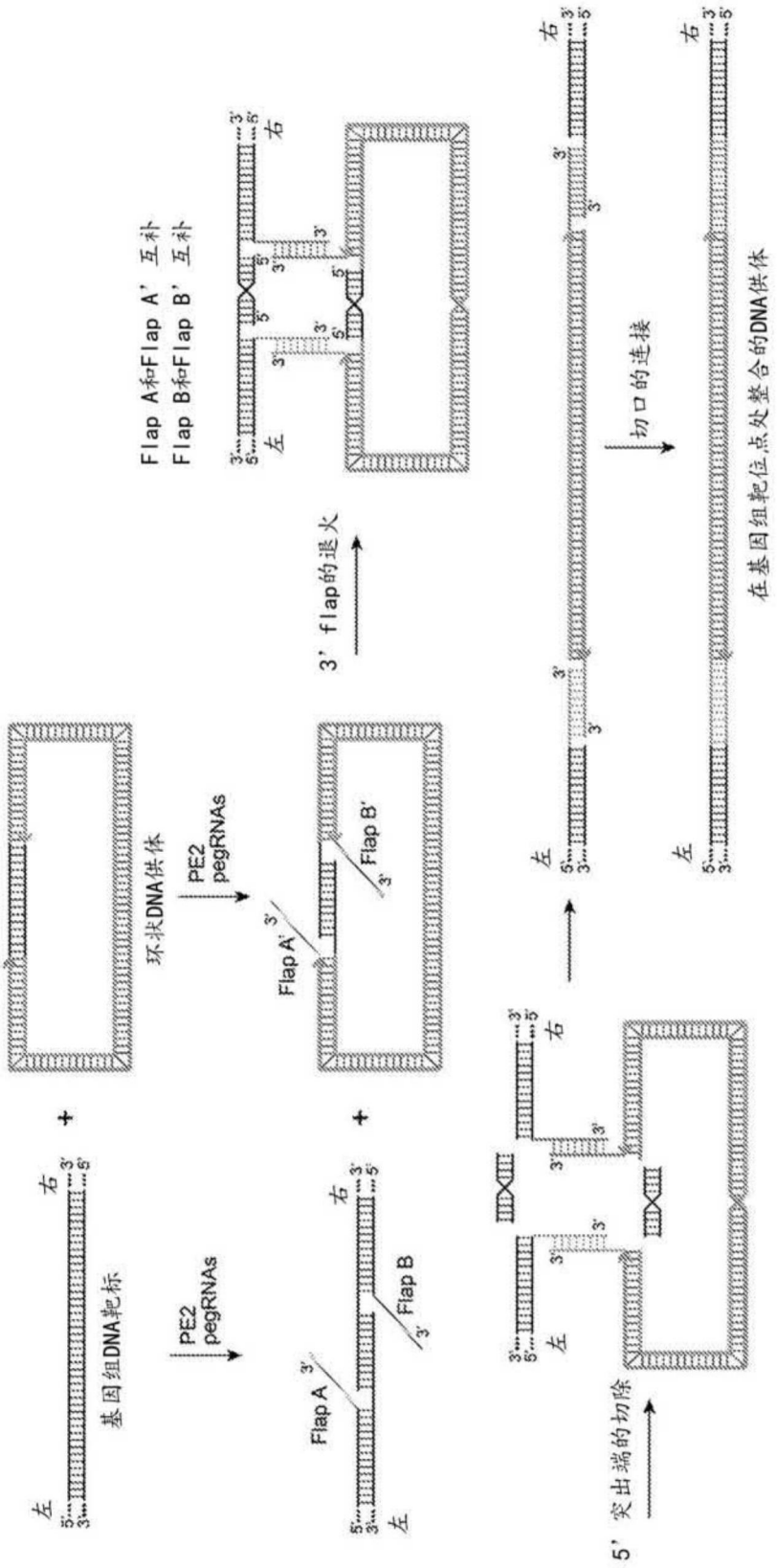

左
右
基因组DNA靶标
PE2
pegRNAs
环状DNA供体
Flap A'
Flap B'
Flap A
Flap B
Flap A和Flap A' 互补
Flap B和Flap B' 互补
3' flap的退火
切口的连接
在基因组靶位点处整合的DNA供体
5' 突出端的切除

图97A

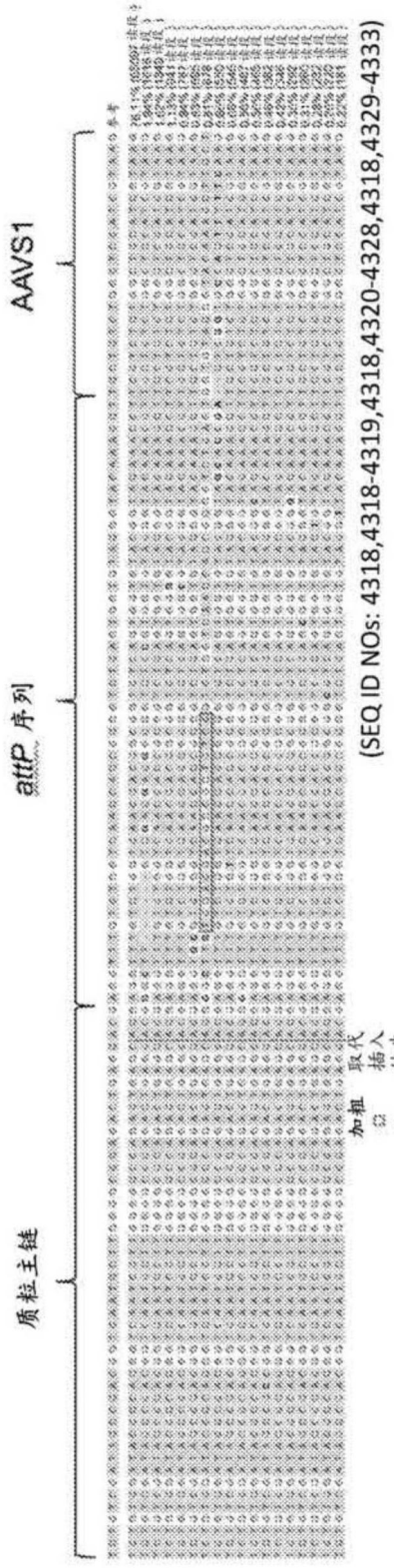

质粒主链
attP 序列
AAVS1
(SEQ ID NOs: 4318,4318-4319,4318,4320-4328,4318,4329-4333)
加粗
取代
插入
缺失
预测的切割位置

图97B

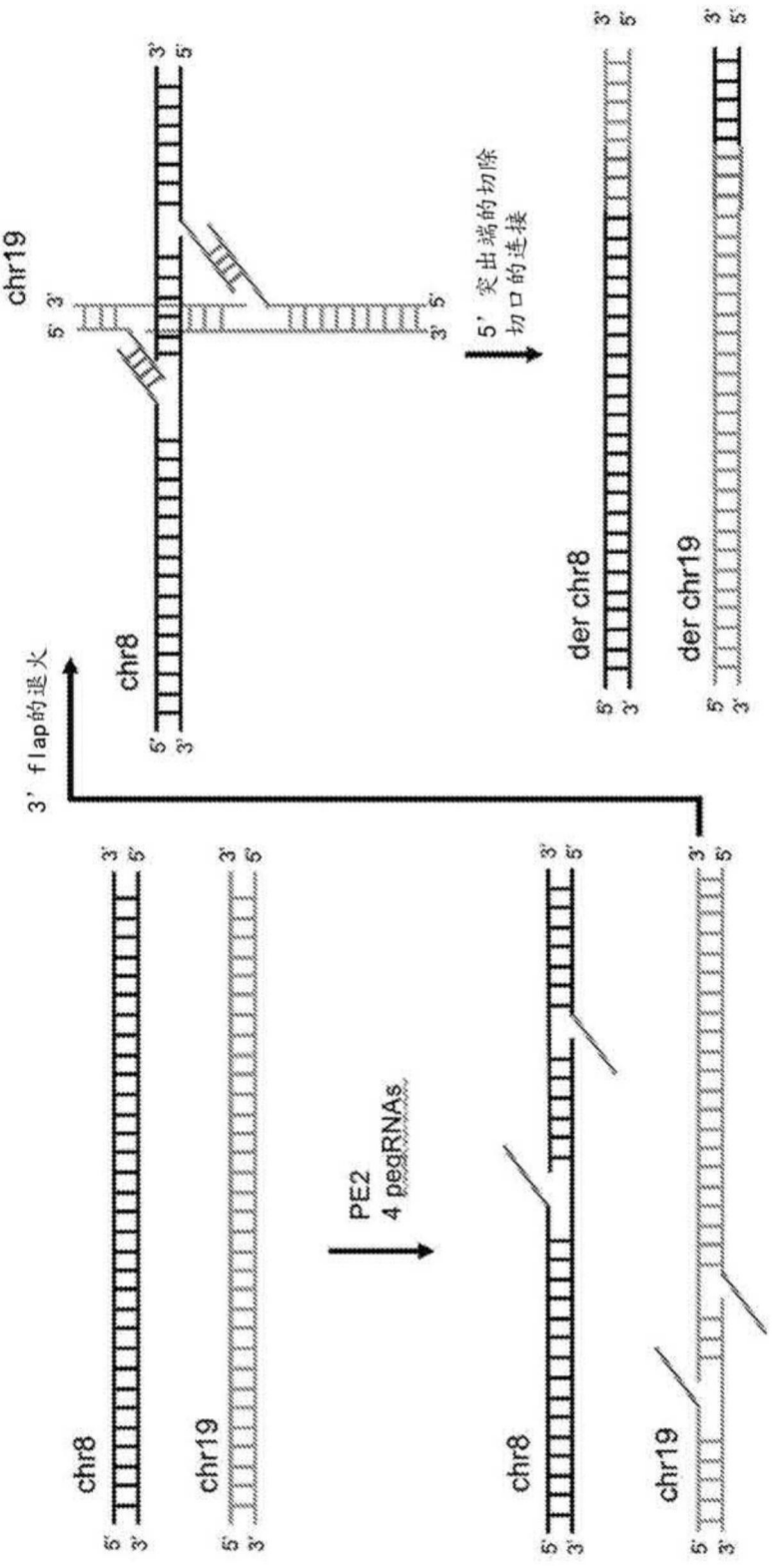

chr19
chr8
3' flap的退火
5' 突出端的切除
切口的连接
der chr8
der chr19
PE2
4 pegRNAs
chr8
chr19
chr8
chr19

图98A

PCR扩增子的HTS (pegRNA Set #5 rep2):
与预期接合1 （含有His6标签）比对的读段-总比对读段的99.5%
(SEQ ID NOs: 4334,4334-4336)
sgRNA
取代
插入
缺失
预测的切割位置
PCR扩增子的HTS (pegRNA Set #5 rep2):
与预期接合2 （含有22nt随机序列）比对的读段-总比对读段的97.9%
sgRNA
取代
插入
缺失
预测的切割位置
(SEQ ID NOs: 4337,4337-4341)

图98B

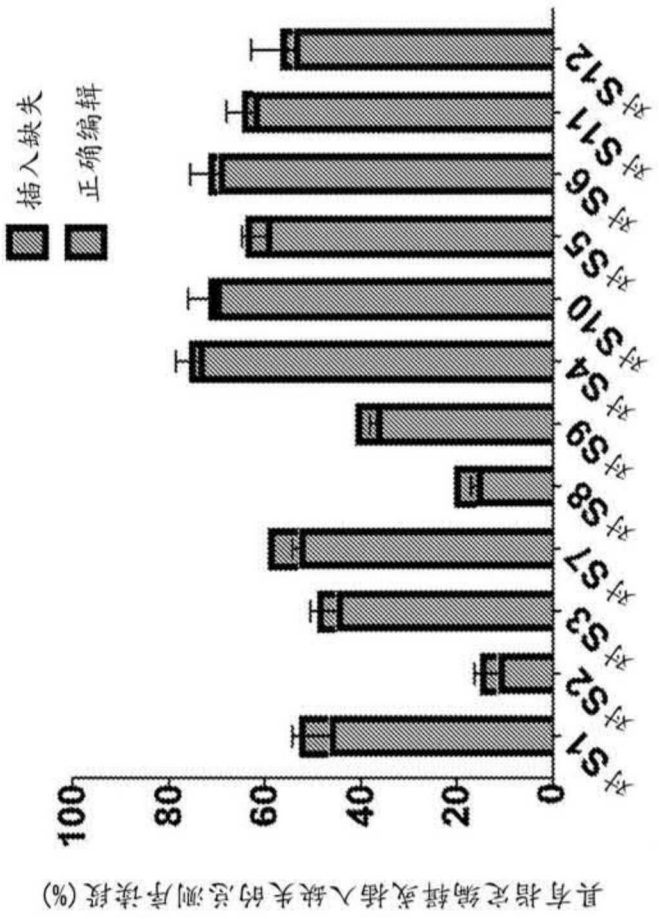

插入缺失
正确编辑
100
80
60
40
20
0
对S1
对S2
对S3
对S7
对S8
对S9
对S4
对S10
对S5
对S6
对S11
对S12
具有指定编辑或插入缺失的总测序读段(%)

图99

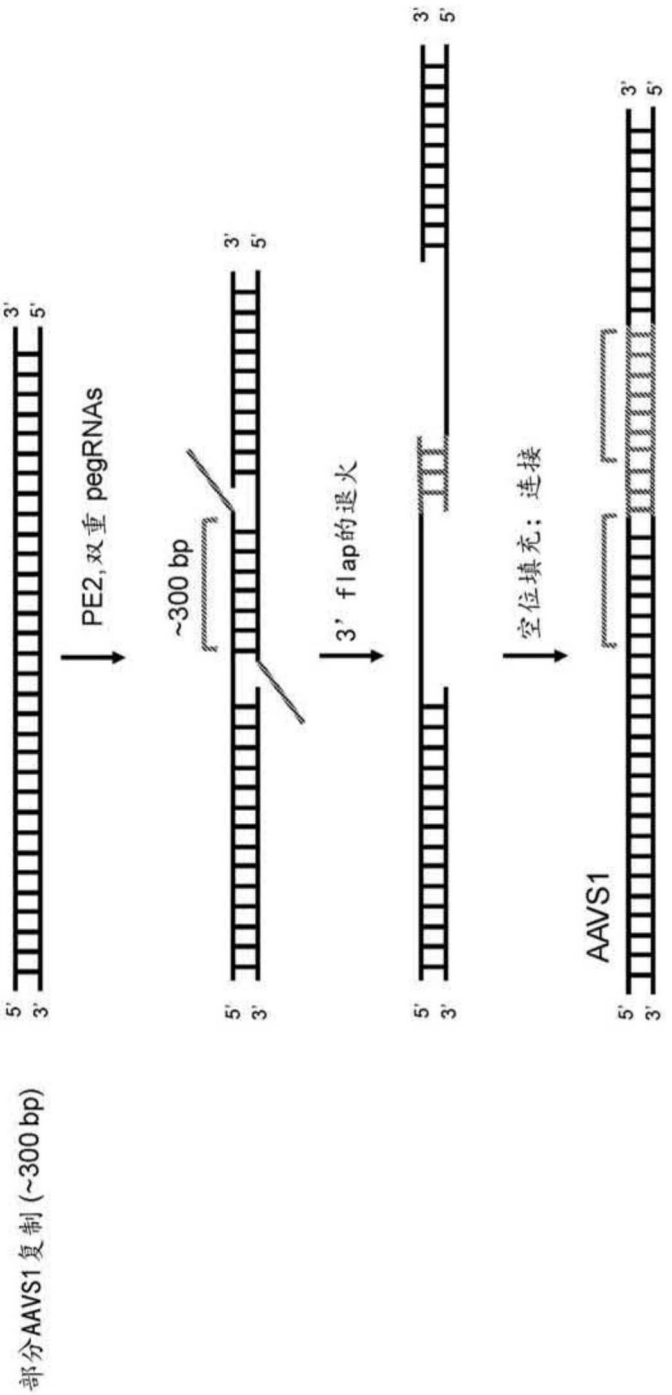

AAVS1
5'
3'
3'
5'
PE2,双重 pegRNAs
~300 bp
3' flap的退火
空位填充；连接
部分AAVS1 复制 (~300 bp)

图100A

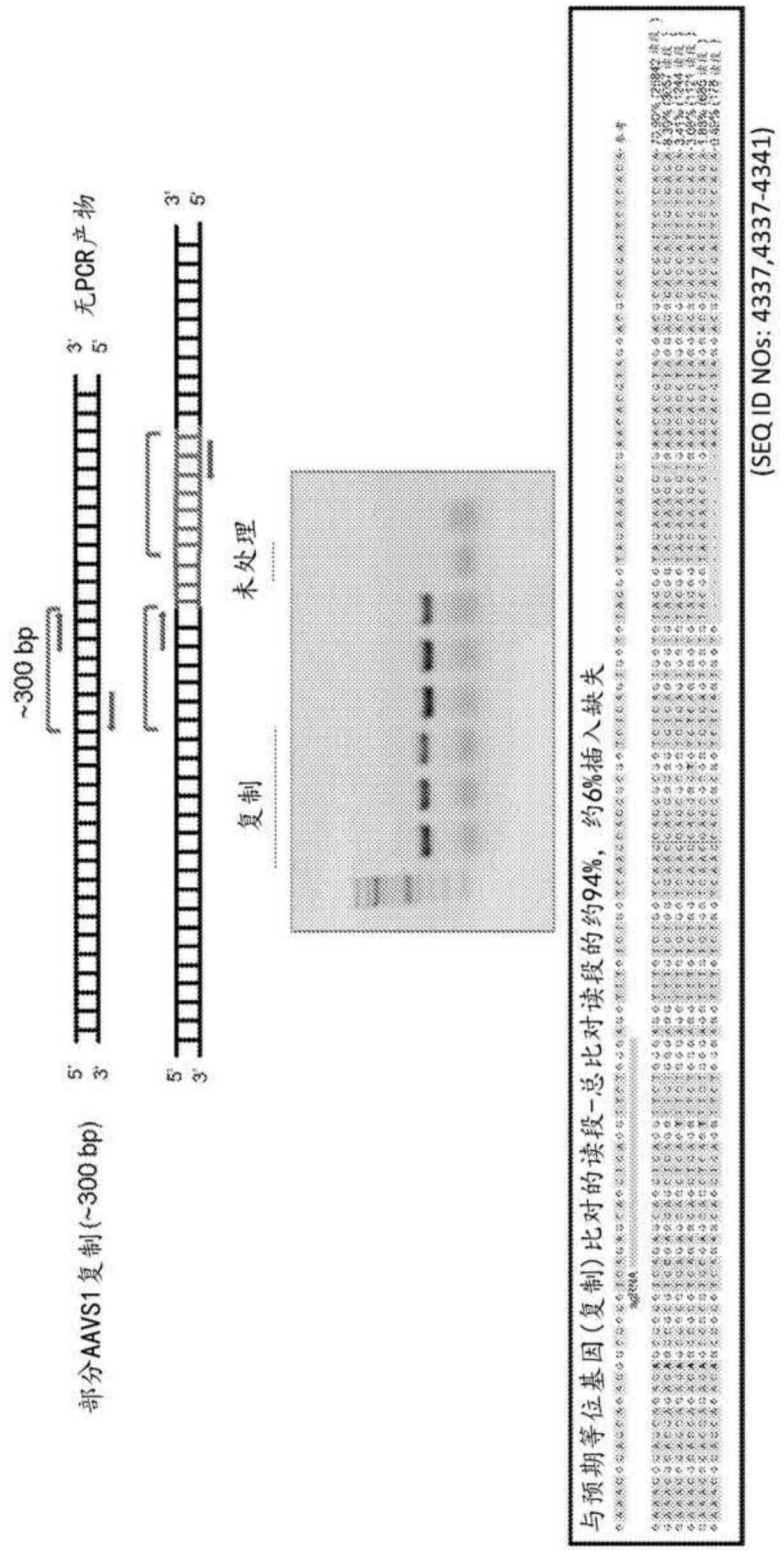

~300 bp
5'
3'
部分AAVS1复制(~300 bp)
3'
5'
无PCR产物
5'
3'
3'
5'
复制
未处理
与预期等位基因(复制)比对的读段-总比对读段的约94%，约6%插入缺失
(SEQ ID NOs: 4337,4337-4341)

图100B

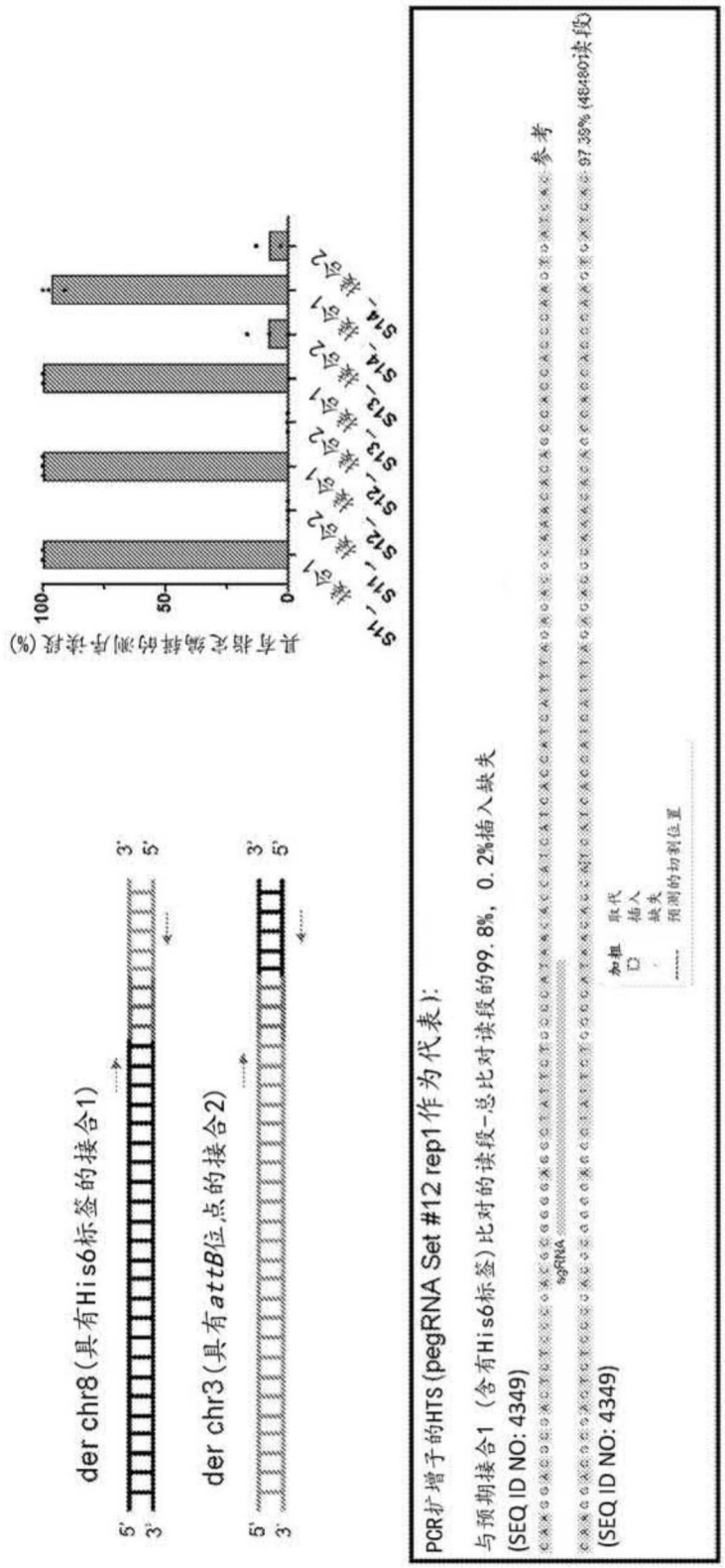

der chr8(具有His6标签的接合1)
5'
3'
3'
5'
der chr3(具有attB位点的接合2)
5'
3'
3'
5'
具有指定编辑的测序读段(%)
100
50
0
S11_接合1
S11_接合2
S12_接合1
S12_接合2
S13_接合1
S13_接合2
S14_接合1
S14_接合2
PCR扩增子的HTS(pegRNA Set #12 rep1作为代表):
与预期接合1（含有His6标签）比对的读段-总比对读段的99.8%，0.2%插入缺失
(SEQ ID NO: 4349)
sgRNA
参考
97.38% (48480读段)
(SEQ ID NO: 4349)
加粗
取代
插入
缺失
预测的切割位置

图101

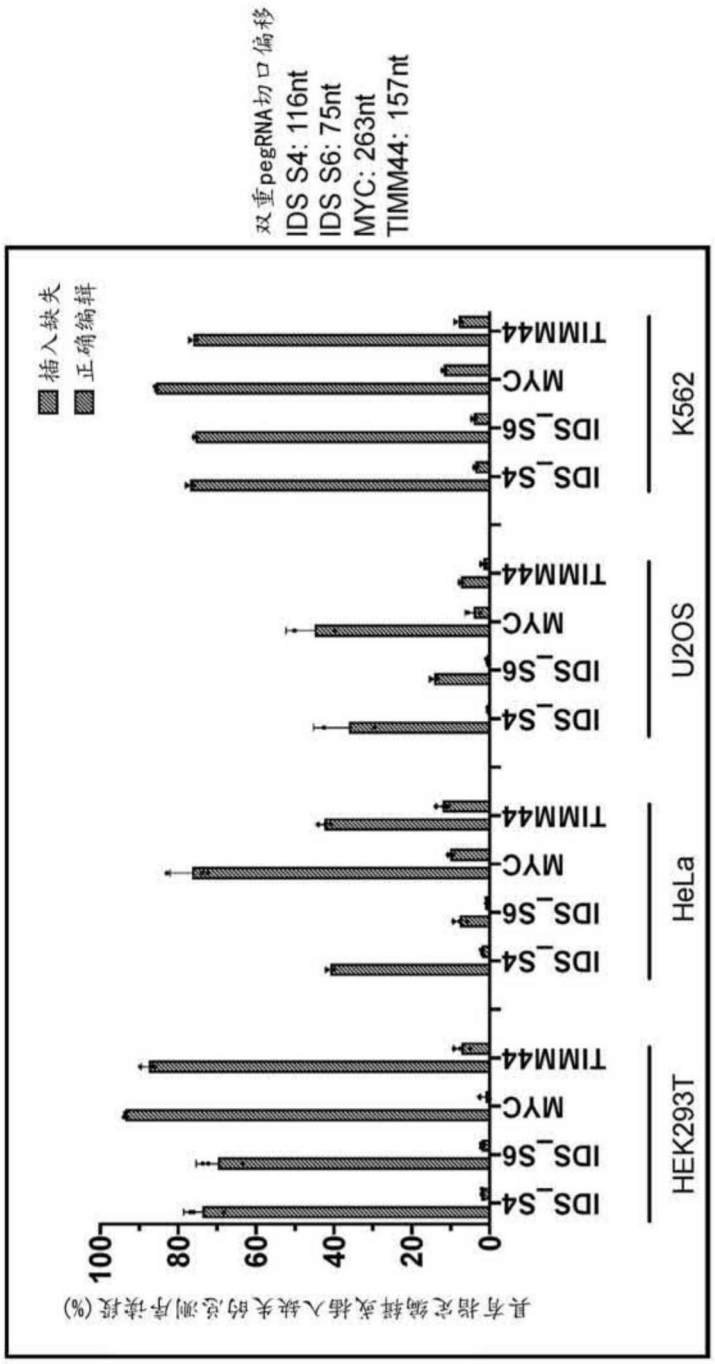

双重pegRNA切口偏移
IDS S4: 116nt
IDS S6: 75nt
MYC: 263nt
TIMM44: 157nt
插入缺失
正确编辑
具有指定编辑或插入缺失的总测序读段(%)
100
80
60
40
20
0
HEK293T
HeLa
U2OS
K562
IDS_S4
IDS_S6
MYC
TIMM44

图102A

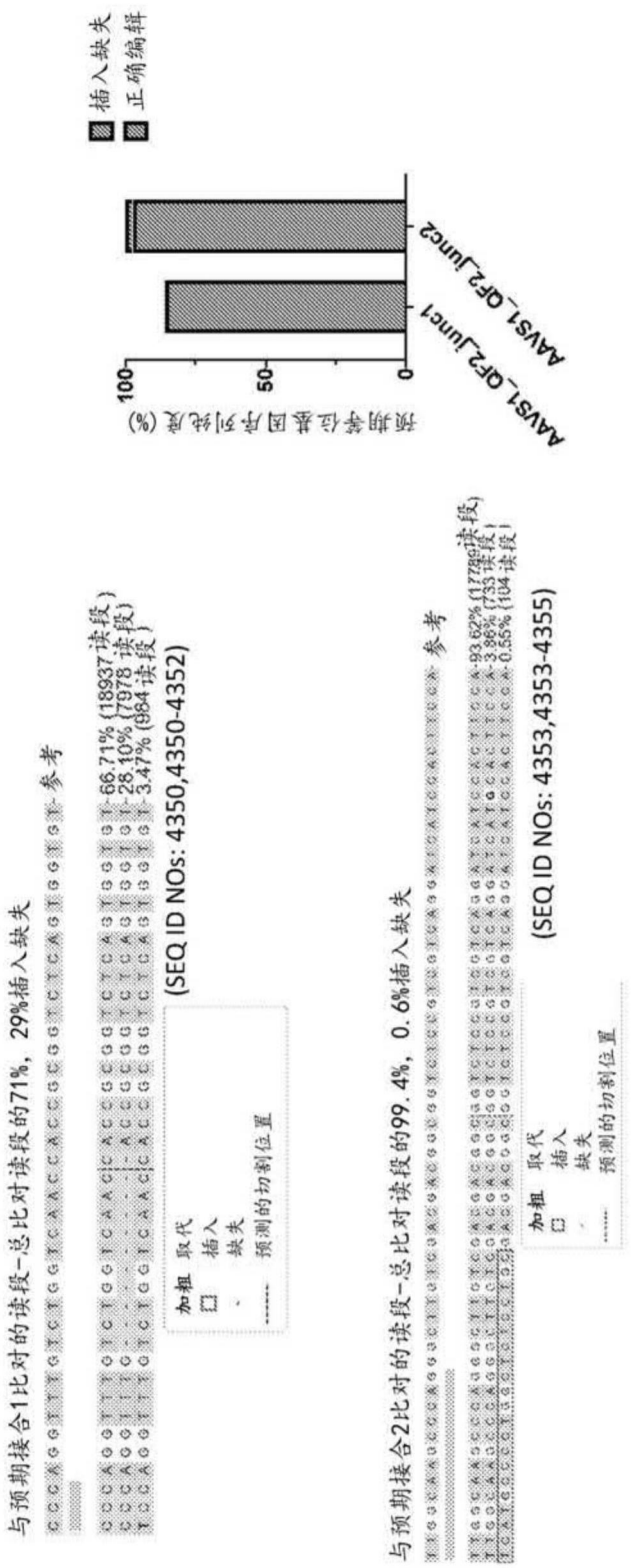

与预期接合1比对的读段-总比对读段的71%，29%插入缺失
参考
加粗 取代
插入
缺失
预测的切割位置
(SEQ ID NOs: 4350,4350-4352)
与预期接合2比对的读段-总比对读段的99.4%，0.6%插入缺失
参考
加粗 取代
插入
缺失
预测的切割位置
(SEQ ID NOs: 4353,4353-4355)
插入缺失
正确编辑
预期等位基因序列纯度(%)
100
50
0
AAVS1_QF2_junc1
AAVS1_QF2_junc2

图102B

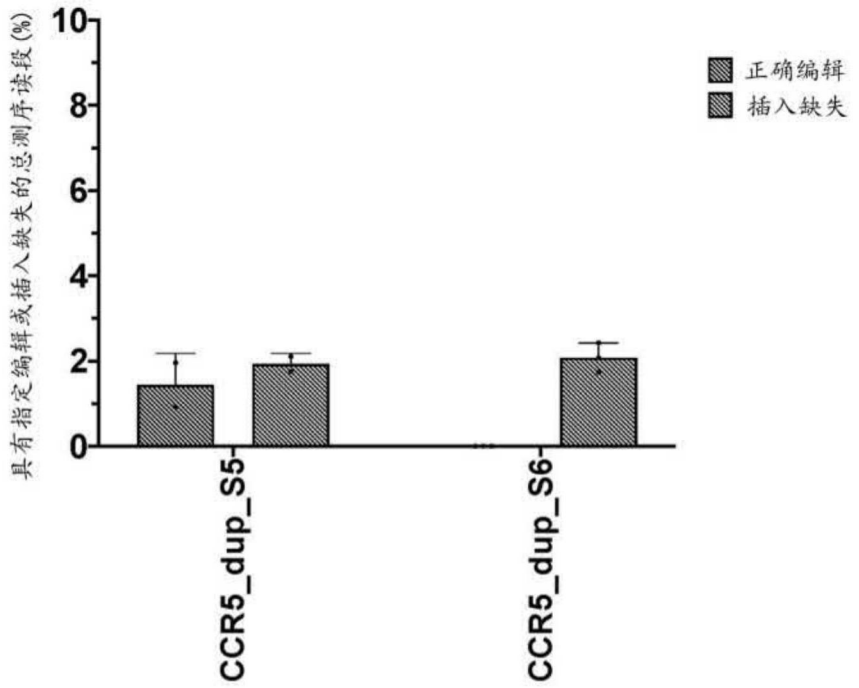

具有指定编辑或插入缺失的总测序读段(%)
10
8
6
4
2
0
正确编辑
插入缺失
CCR5_dup_S5
CCR5_dup_S6

图103A

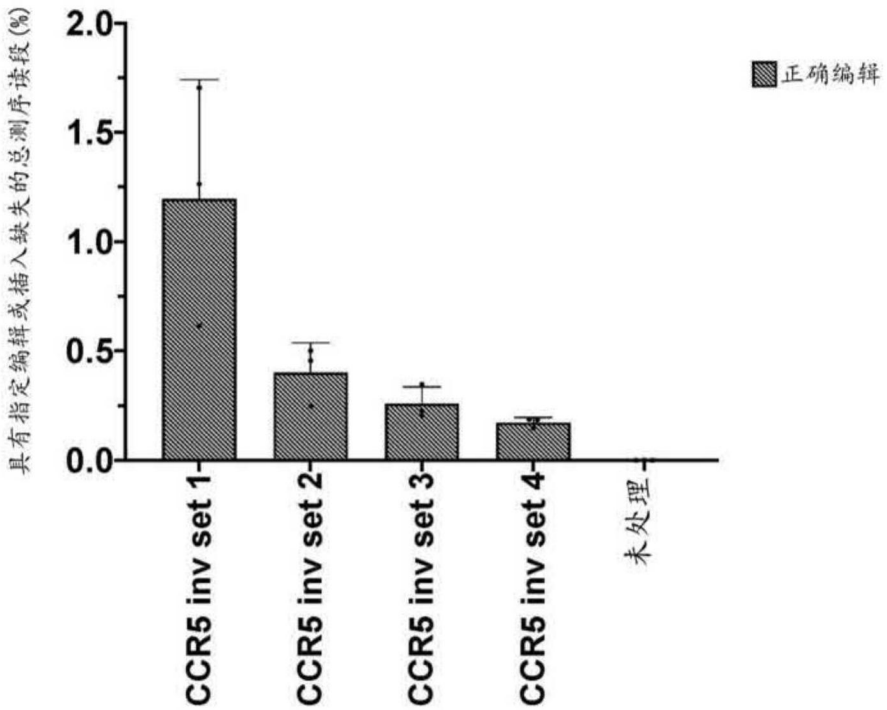

具有指定编辑或插入缺失的总测序读段(%)
2.0
1.5
1.0
0.5
0.0
正确编辑
CCR5 inv set 1
CCR5 inv set 2
CCR5 inv set 3
CCR5 inv set 4
未处理

图103B

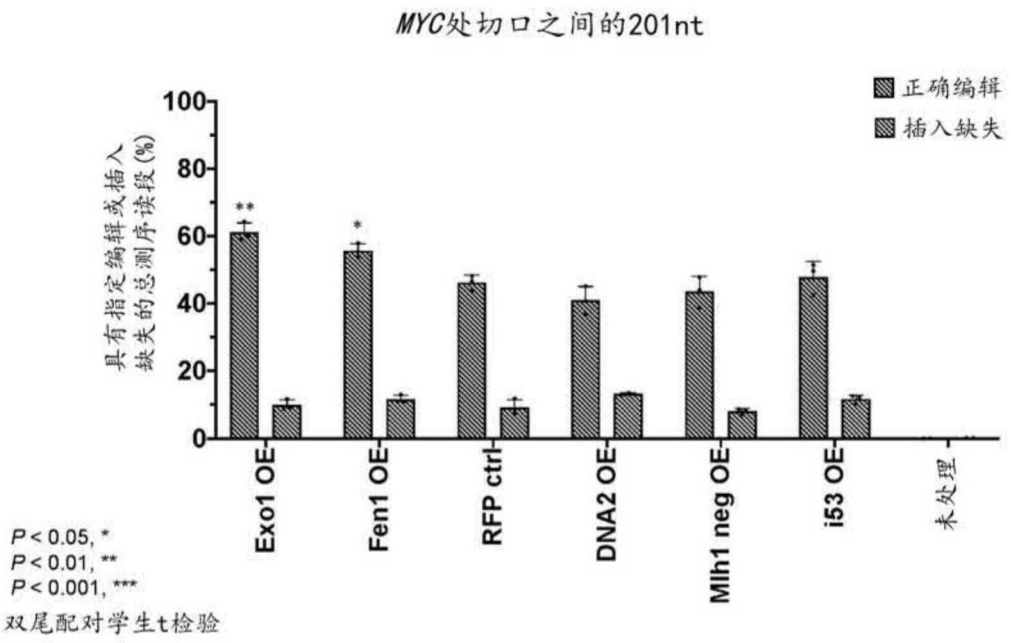

MYC处切口之间的201nt
正确编辑
插入缺失
具有指定编辑或插入缺失的总测序读段(%)
100
80
60
40
20
0
Exo1 OE
Fen1 OE
RFP ctrl
DNA2 OE
Mlh1 neg OE
i53 OE
未处理
P < 0.05, *
P < 0.01, **
P < 0.001, ***
双尾配对学生t检验

图104A

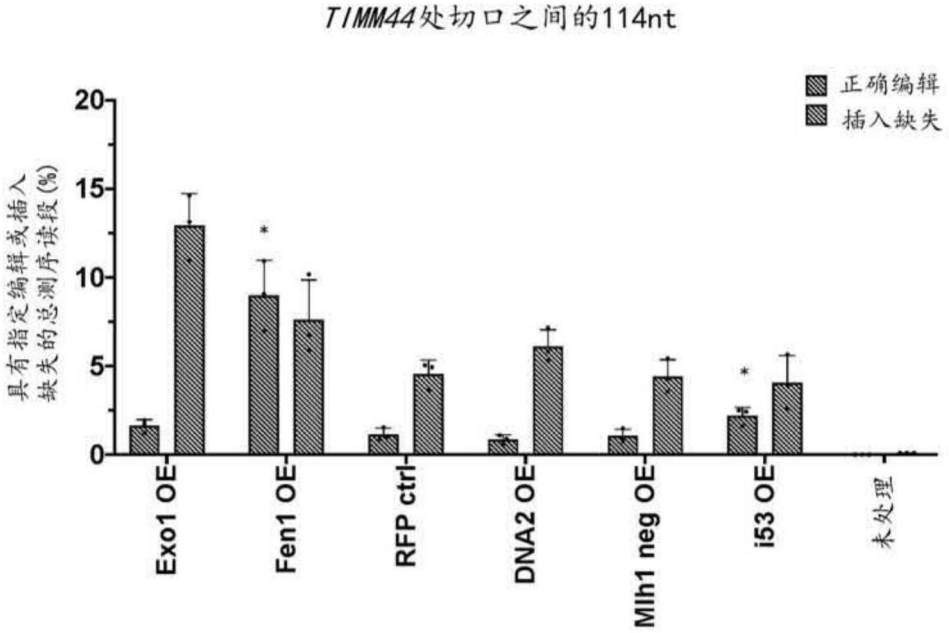

TIMM44处切口之间的114nt
正确编辑
插入缺失
具有指定编辑或插入缺失的总测序读段(%)
20
15
10
5
0
Exo1 OE
Fen1 OE
RFP ctrl
DNA2 OE
Mlh1 neg OE
i53 OE
未处理

图104B

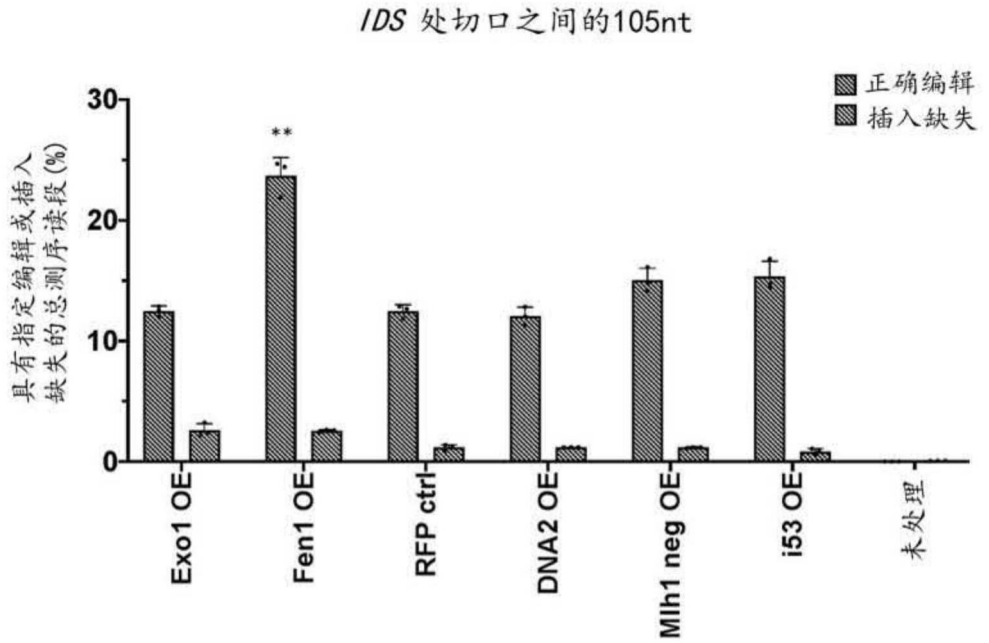

IDS 处切口之间的105nt
正确编辑
插入缺失
具有指定编辑或插入缺失的总测序读段(%)
30
20
10
0
**
Exo1 OE
Fen1 OE
RFP ctrl
DNA2 OE
Mlh1 neg OE
i53 OE
未处理

图104C

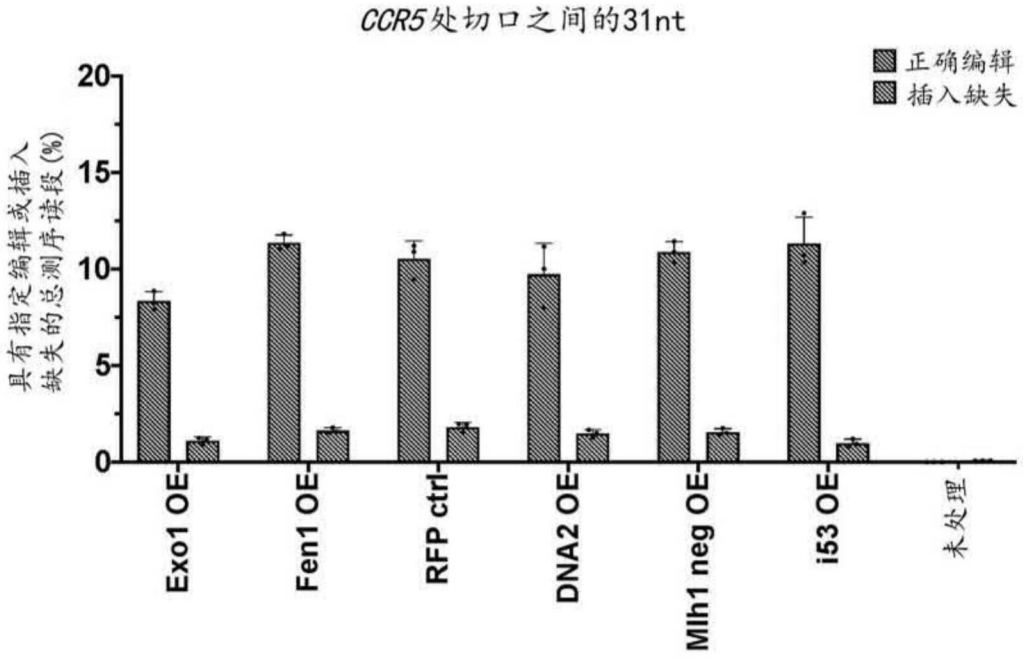

CCR5处切口之间的31nt
正确编辑
插入缺失
具有指定编辑或插入缺失的总测序读段(%)
20
15
10
5
0
Exo1 OE
Fen1 OE
RFP ctrl
DNA2 OE
Mlh1 neg OE
i53 OE
未处理

图104D

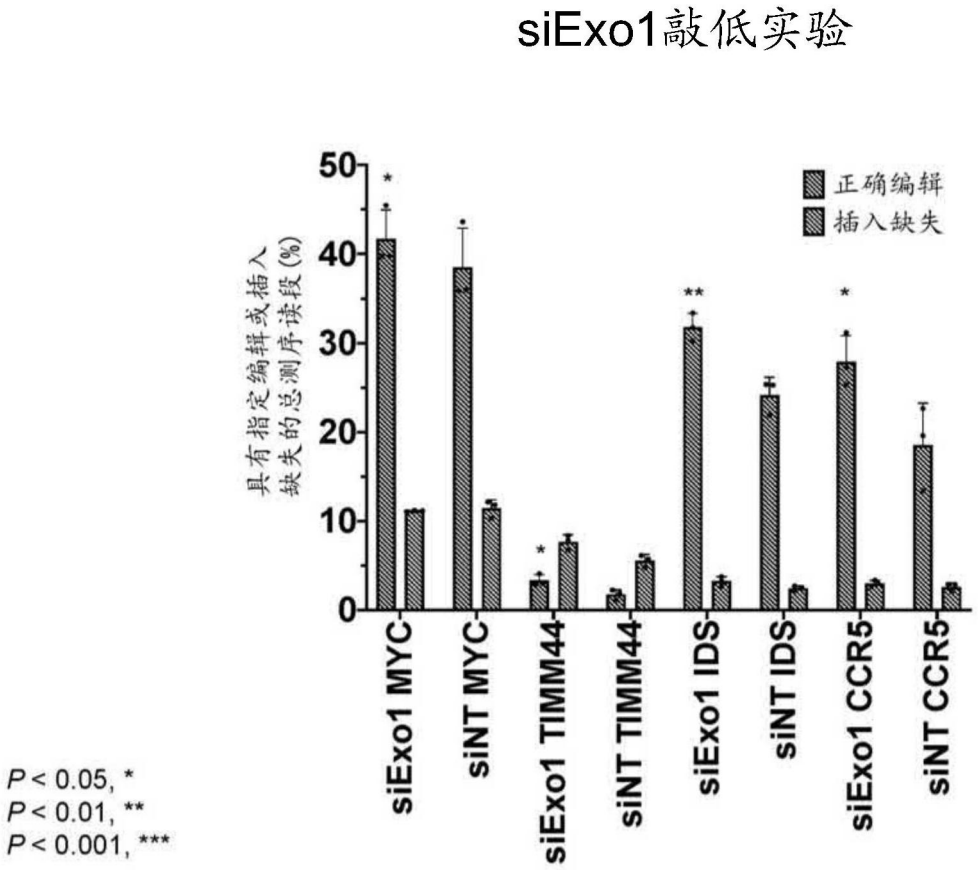

siExo1敲低实验
具有指定编辑或插入缺失的总测序读段(%)
50
40
30
20
10
0
正确编辑
插入缺失
*
**
*
*
siExo1 MYC
siNT MYC
siExo1 TIMM44
siNT TIMM44
siExo1 IDS
siNT IDS
siExo1 CCR5
siNT CCR5
P < 0.05, *
P < 0.01, **
P < 0.001, ***

图104E

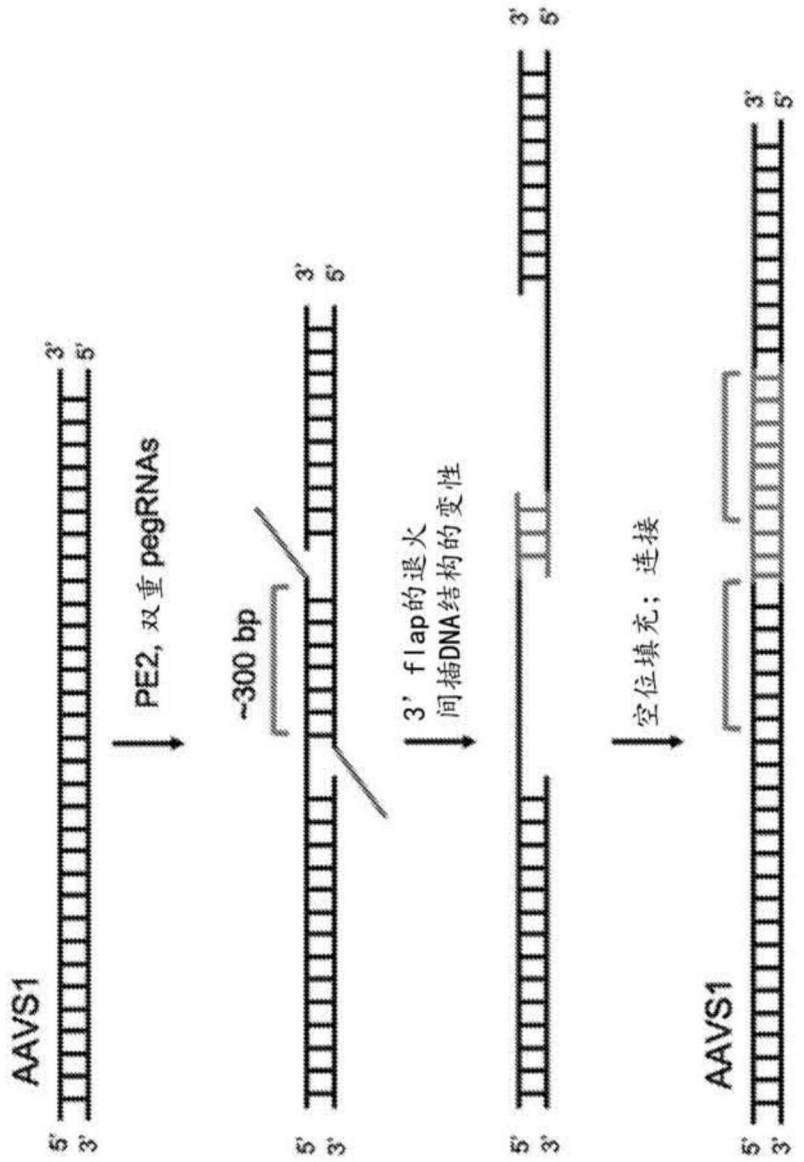


图105A

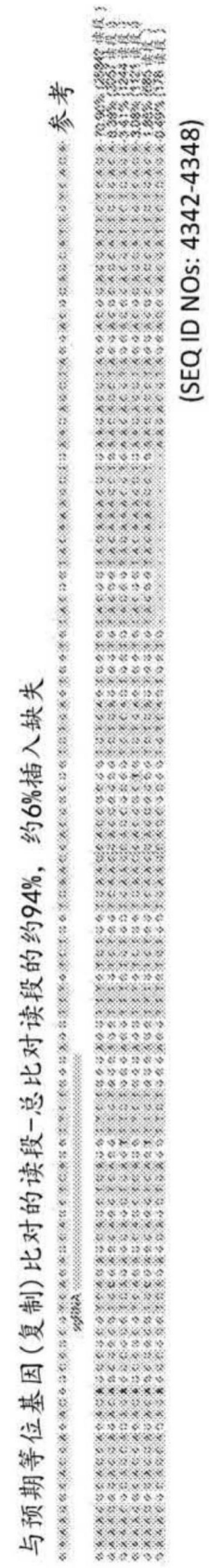


图105B

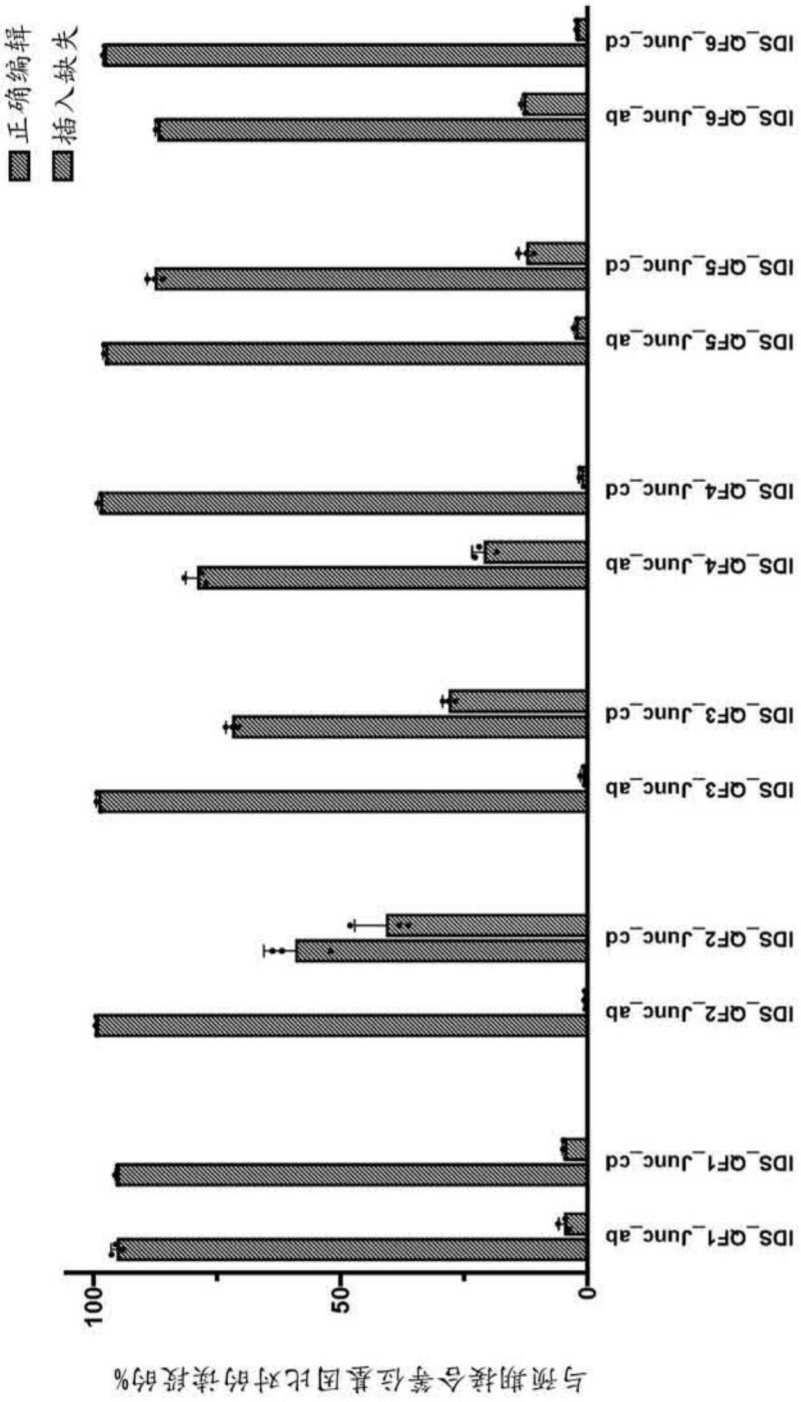

正确编辑
插入缺失
IDS_QF1_Junc_ab
IDS_QF1_Junc_cd
IDS_QF2_Junc_ab
IDS_QF2_Junc_cd
IDS_QF3_Junc_ab
IDS_QF3_Junc_cd
IDS_QF4_Junc_ab
IDS_QF4_Junc_cd
IDS_QF5_Junc_ab
IDS_QF5_Junc_cd
IDS_QF6_Junc_ab
IDS_QF6_Junc_cd
0
50
100
与预期接合等位基因比对的读段的%

图106

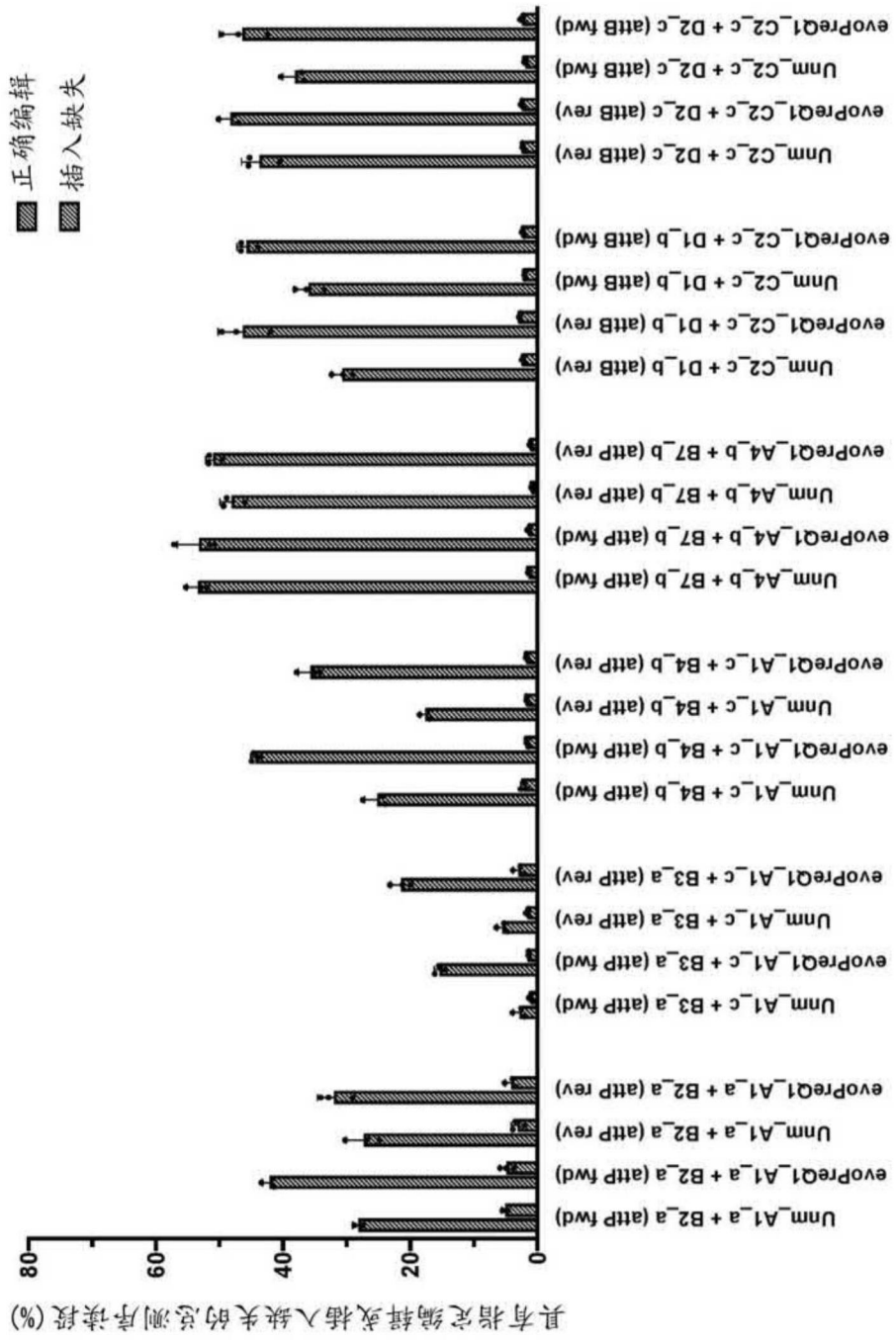

具有指定编辑或插入缺失的总测序读段 (%)
0
20
40
60
80
插入缺失
正确编辑
Unm_A1_a + B2_a (attP fwd)
evoPreQ1_A1_a + B2_a (attP fwd)
Unm_A1_a + B2_a (attP rev)
evoPreQ1_A1_a + B2_a (attP rev)
Unm_A1_c + B3_a (attP fwd)
evoPreQ1_A1_c + B3_a (attP fwd)
Unm_A1_c + B3_a (attP rev)
evoPreQ1_A1_c + B3_a (attP rev)
Unm_A1_c + B4_b (attP fwd)
evoPreQ1_A1_c + B4_b (attP fwd)
Unm_A1_c + B4_b (attP rev)
evoPreQ1_A1_c + B4_b (attP rev)
Unm_A4_b + B7_b (attP fwd)
evoPreQ1_A4_b + B7_b (attP fwd)
Unm_A4_b + B7_b (attP rev)
evoPreQ1_A4_b + B7_b (attP rev)
Unm_C2_c + D1_b (attB rev)
evoPreQ1_C2_c + D1_b (attB rev)
Unm_C2_c + D1_b (attB fwd)
evoPreQ1_C2_c + D1_b (attB fwd)
Unm_C2_c + D2_c (attB rev)
evoPreQ1_C2_c + D2_c (attB rev)
Unm_C2_c + D2_c (attB fwd)
evoPreQ1_C2_c + D2_c (attB fwd)

图107

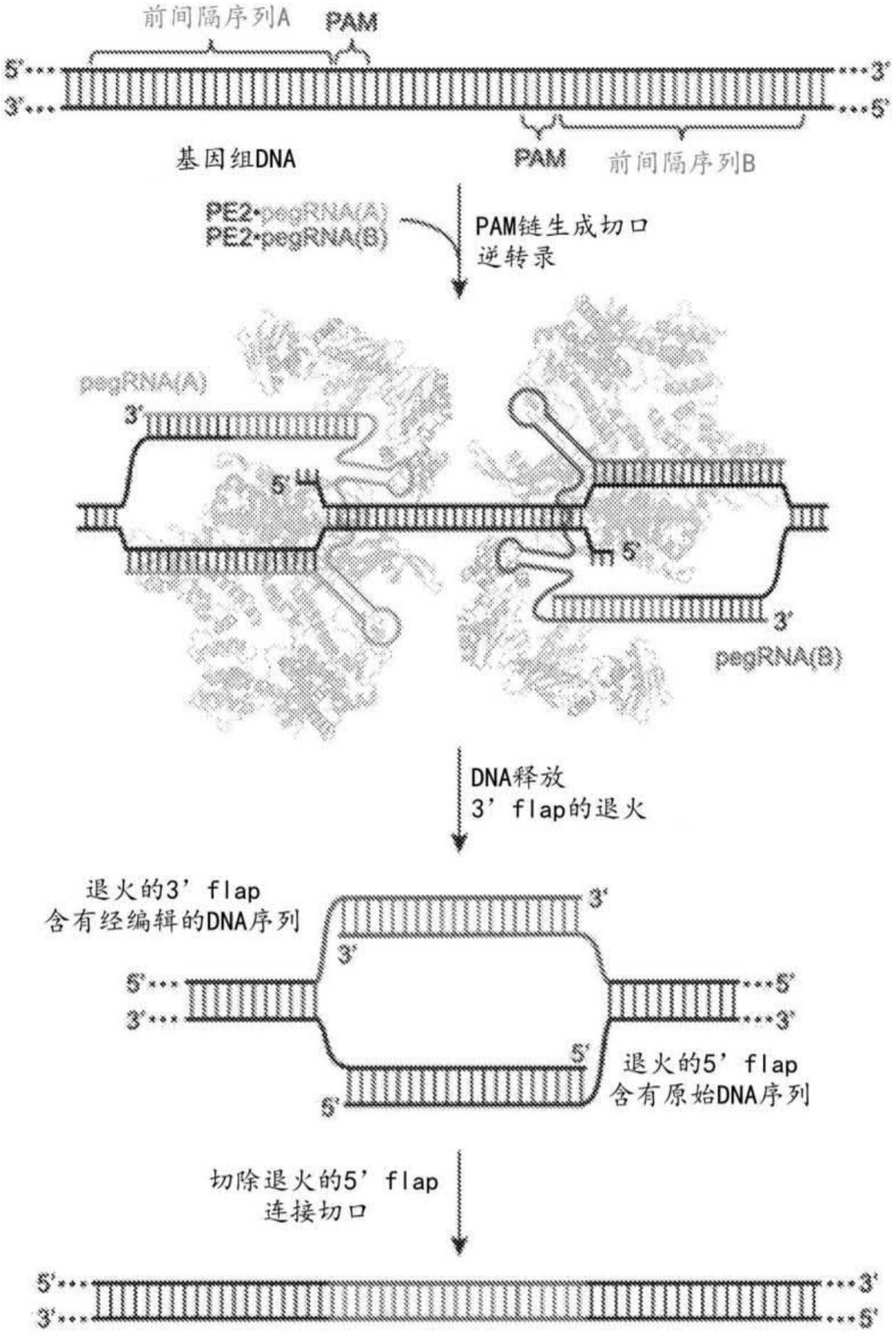

前间隔序列A
PAM
5'
3'
3'
5'
基因组DNA
PAM
前间隔序列B
PE2•pegRNA(A)
PE2•pegRNA(B)
PAM链生成切口
逆转录
pegRNA(A)
3'
5'
5'
3'
pegRNA(B)
DNA释放
3' flap的退火
退火的3' flap
含有经编辑的DNA序列
3'
3'
5'
3'
5'
3'
5'
退火的5' flap
含有原始DNA序列
切除退火的5' flap
连接切口
5'
3'
3'
5'
具有原始DNA序列的产物由经编辑的DNA序列替换

图108A

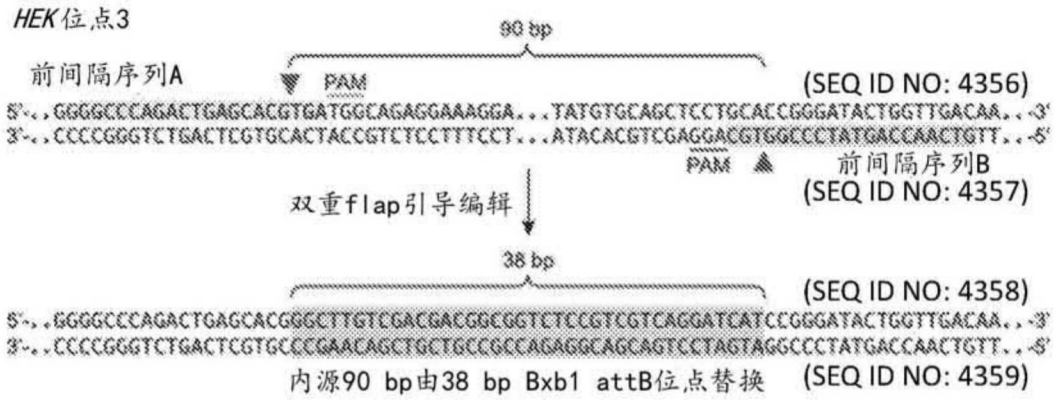

HEK位点3
90 bp
前间隔序列A
PAM
(SEQ ID NO: 4356)
5'-..GGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGA...TATGTGCAGCTCCTGCACCGGGATACTGGTTGACAA..-3'
3'-..CCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCT...ATACACGTCGAGGACGTGGCCCTATGACCAACTGTT..-5'
PAM
前间隔序列B
(SEQ ID NO: 4357)
双重flap引导编辑
38 bp
(SEQ ID NO: 4358)
5'-..GGGGCCCAGACTGAGCACGGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGGTTGACAA..-3'
3'-..CCCCGGGTCTGACTCGTGCCCGAACAGCTGCTGCCGCCAGAGGCAGCAGTCCTAGTAGGCCCTATGACCAACTGTT..-5'
内源90 bp由38 bp Bxb1 attB位点替换
(SEQ ID NO: 4359)

图108B

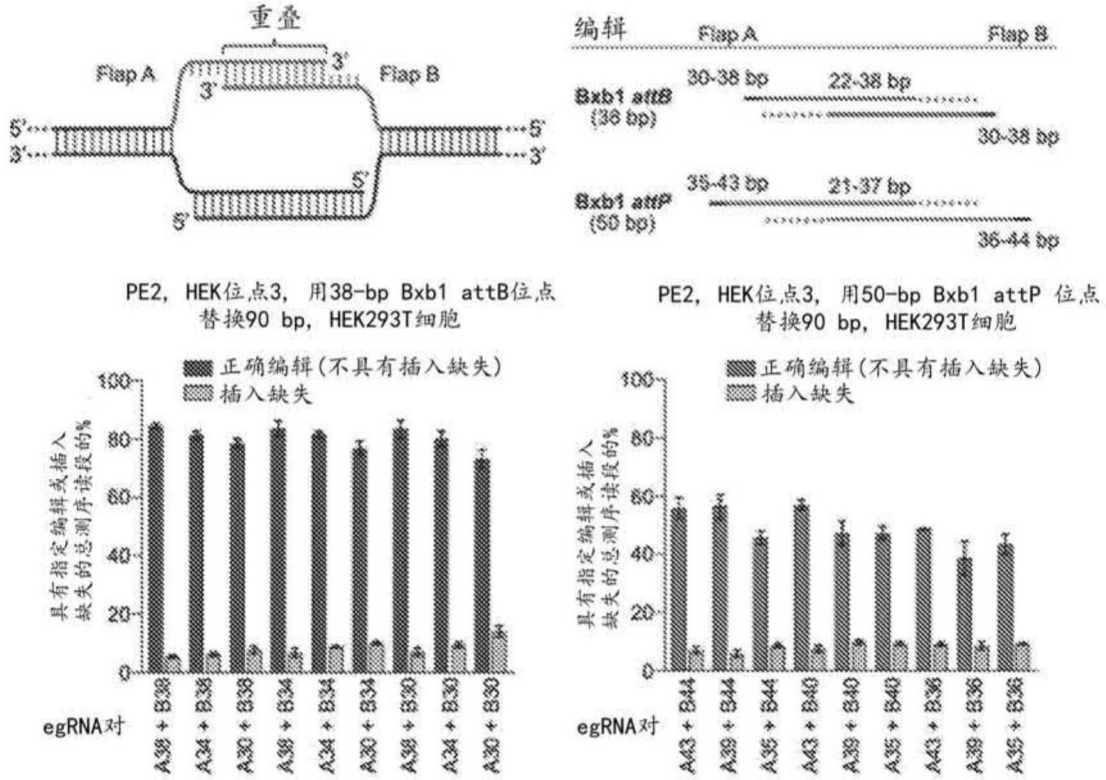

重叠
Flap A
Flap B
3'
5'
编辑
Flap A
Flap B
Bxb1 attB
(38 bp)
30-38 bp
22-38 bp
30-38 bp
Bxb1 attP
(50 bp)
35-43 bp
21-37 bp
36-44 bp
PE2, HEK位点3, 用38-bp Bxb1 attB位点替换90 bp, HEK293T细胞
PE2, HEK位点3, 用50-bp Bxb1 attP 位点替换90 bp, HEK293T细胞
正确编辑(不具有插入缺失)
插入缺失
具有指定编辑或插入缺失的总测序读段的%
0
20
40
60
80
100
egRNA对
A38 + B38
A34 + B38
A30 + B38
A38 + B34
A34 + B34
A30 + B34
A38 + B30
A34 + B30
A30 + B30
A43 + B44
A39 + B44
A35 + B44
A43 + B40
A39 + B40
A35 + B40
A43 + B36
A39 + B36
A35 + B36

图108C

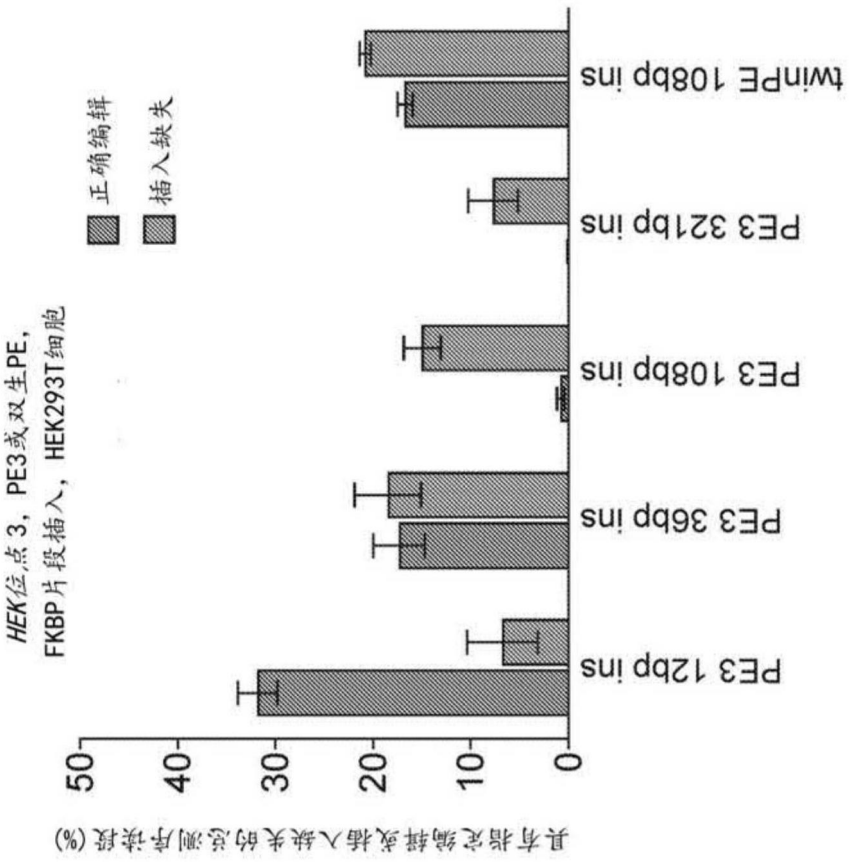

具有指定编辑或插入缺失的总测序读段(%)
0
10
20
30
40
50
HEK位点3，PE3或双生PE，
FKBP片段插入，HEK293T细胞
正确编辑
插入缺失
PE3 12bp ins
PE3 36bp ins
PE3 108bp ins
PE3 321bp ins
twinPE 108bp ins

图109A

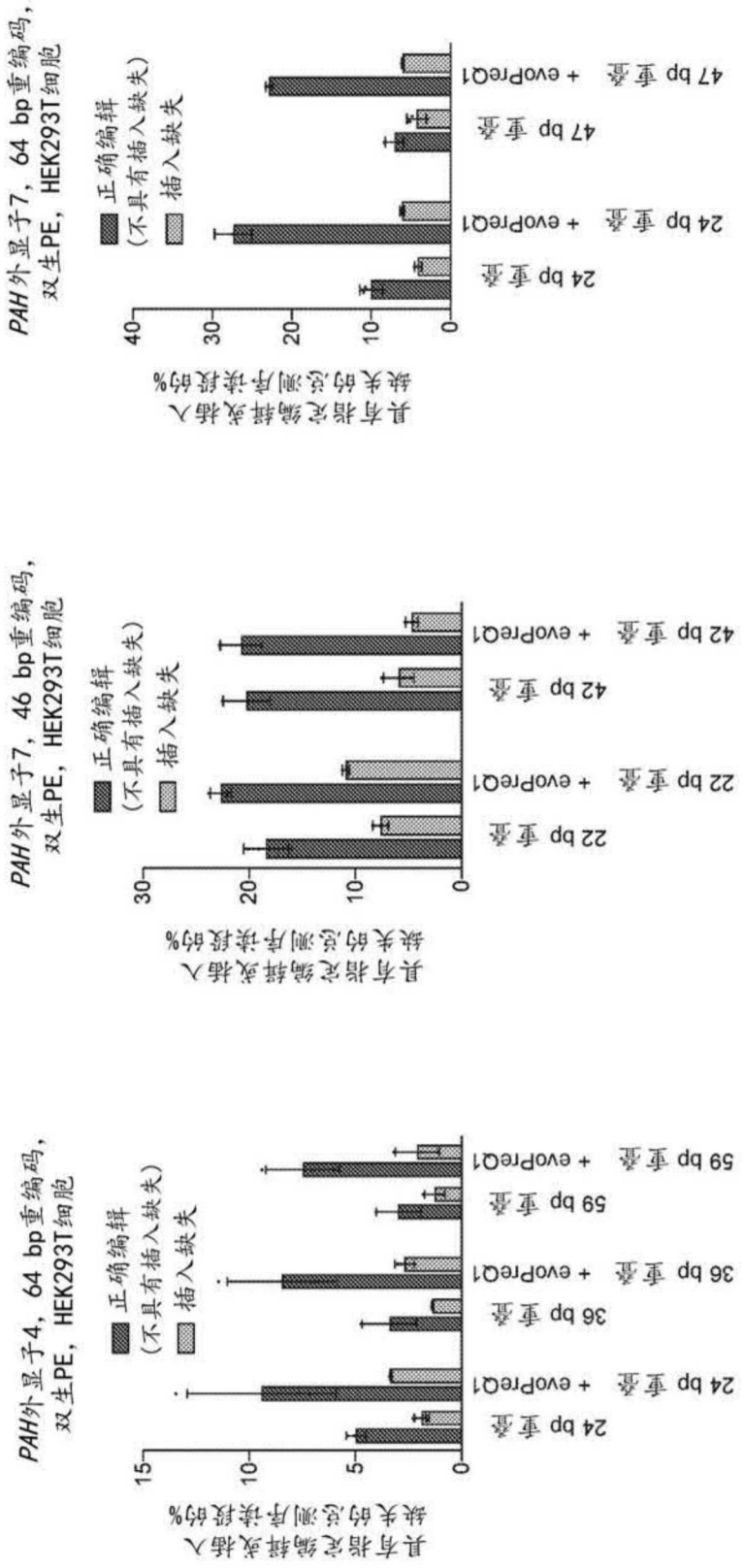

PAH外显子4，64 bp重编码，
双生PE，HEK293T细胞
正确编辑
(不具有插入缺失)
插入缺失
具有指定编辑类插入
缺失的总测序读段的%
0
5
10
15
24 bp 重叠
24 bp 重叠 + evoPreQ1
36 bp 重叠
36 bp 重叠 + evoPreQ1
59 bp 重叠
59 bp 重叠 + evoPreQ1
PAH外显子7，46 bp重编码，
双生PE，HEK293T细胞
正确编辑
(不具有插入缺失)
插入缺失
具有指定编辑类插入
缺失的总测序读段的%
0
10
20
30
22 bp 重叠
22 bp 重叠 + evoPreQ1
42 bp 重叠
42 bp 重叠 + evoPreQ1
PAH外显子7，64 bp重编码，
双生PE，HEK293T细胞
正确编辑
(不具有插入缺失)
插入缺失
具有指定编辑类插入
缺失的总测序读段的%
0
10
20
30
40
24 bp 重叠
24 bp 重叠 + evoPreQ1
47 bp 重叠
47 bp 重叠 + evoPreQ1

图109B

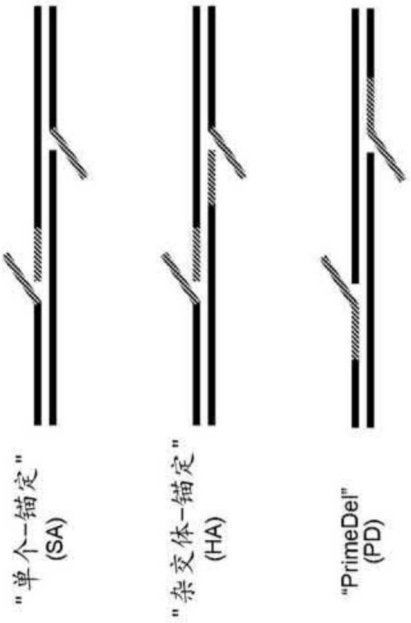

"单个锚定"
(SA)
"杂交体锚定"
(HA)
"PrimeDel"
(PD)

图109C

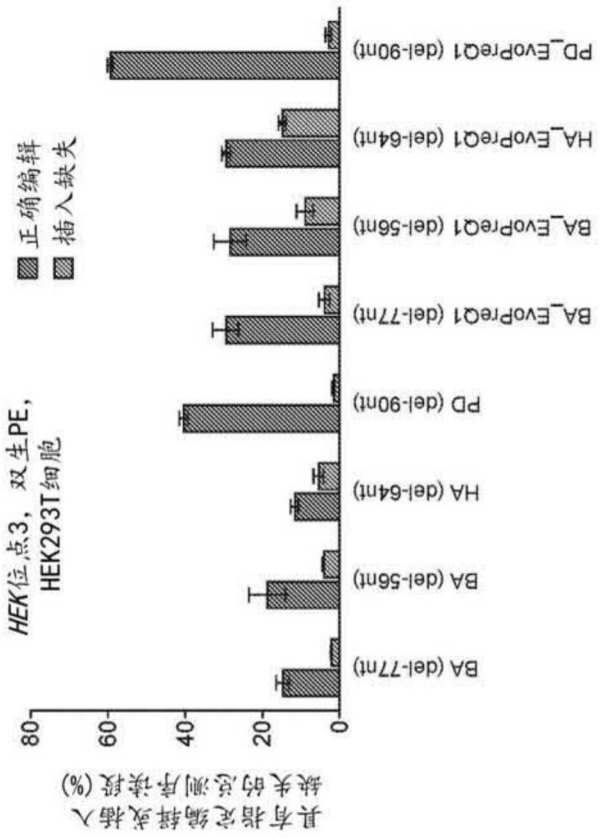

具有指定编辑或插入
缺失的总测序读段(%)
0
20
40
60
80
HEK位点3，双生PE，
HEK293T细胞
正确编辑
插入缺失
BA (del-77nt)
BA (del-56nt)
HA (del-64nt)
PD (del-90nt)
BA_EvoPreQ1 (del-77nt)
BA_EvoPreQ1 (del-56nt)
HA_EvoPreQ1 (del-64nt)
PD_EvoPreQ1 (del-90nt)

图109D

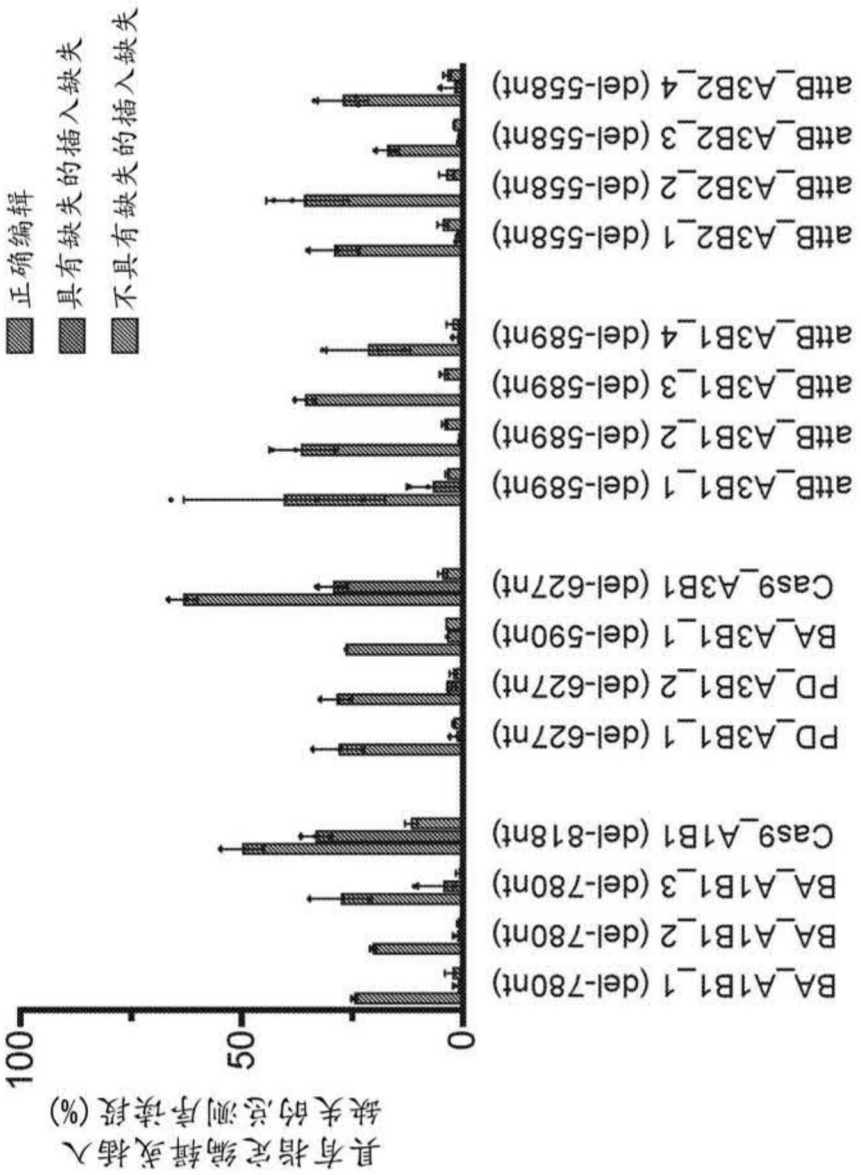

具有指定编辑或插入缺失的总测序读段(%)
100
50
0
正确编辑
具有缺失的插入缺失
不具有缺失的插入缺失
BA_A1B1_1 (del-780nt)
BA_A1B1_2 (del-780nt)
BA_A1B1_3 (del-780nt)
Cas9_A1B1 (del-818nt)
PD_A3B1_1 (del-627nt)
PD_A3B1_2 (del-627nt)
BA_A3B1_1 (del-590nt)
Cas9_A3B1 (del-627nt)
atB_A3B1_1 (del-589nt)
atB_A3B1_2 (del-589nt)
atB_A3B1_3 (del-589nt)
atB_A3B1_4 (del-589nt)
atB_A3B2_1 (del-558nt)
atB_A3B2_2 (del-558nt)
atB_A3B2_3 (del-558nt)
atB_A3B2_4 (del-558nt)

图109E

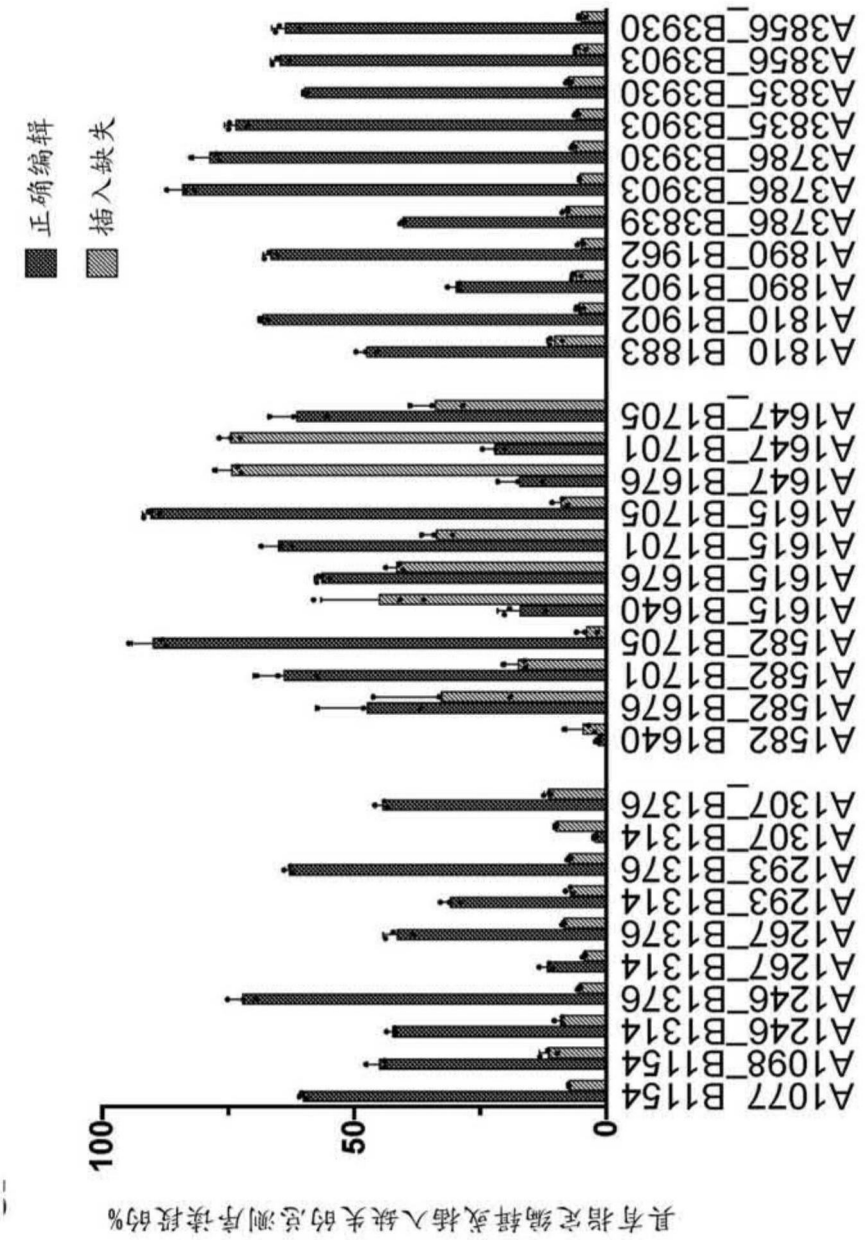

正确编辑
插入缺失
具有指定编辑或插入缺失的总测序读段的%
100
50
0
A10077_B1154
A1098_B1154
A1246_B1314
A1246_B1376
A1267_B1314
A1267_B1376
A1293_B1314
A1293_B1376
A1307_B1314
A1307_B1376
A1582_B1640
A1582_B1676
A1582_B1701
A1582_B1705
A1615_B1640
A1615_B1676
A1615_B1701
A1615_B1705
A1647_B1676
A1647_B1701
A1647_B1705
A1810_B1883
A1810_B1902
A1890_B1902
A1890_B1962
A3786_B3839
A3786_B3903
A3786_B3930
A3835_B3903
A3835_B3930
A3856_B3903
A3856_B3930

图110A

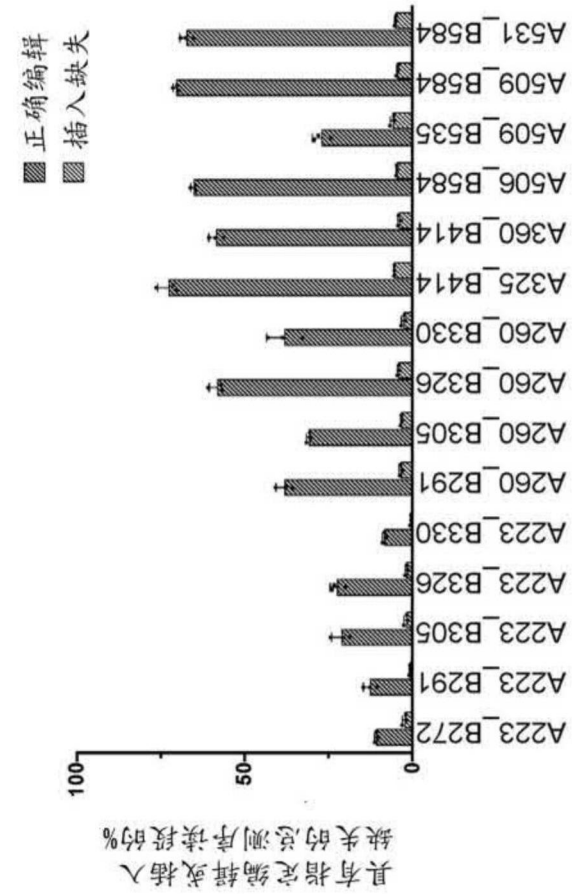

正确编辑
插入缺失
A531_B584
A509_B584
A509_B535
A506_B584
A360_B414
A325_B414
A260_B330
A260_B326
A260_B305
A260_B291
A223_B330
A223_B326
A223_B305
A223_B291
A223_B272
0
50
100
具有指定编辑或插入
缺失的总测序读段的%

图110B

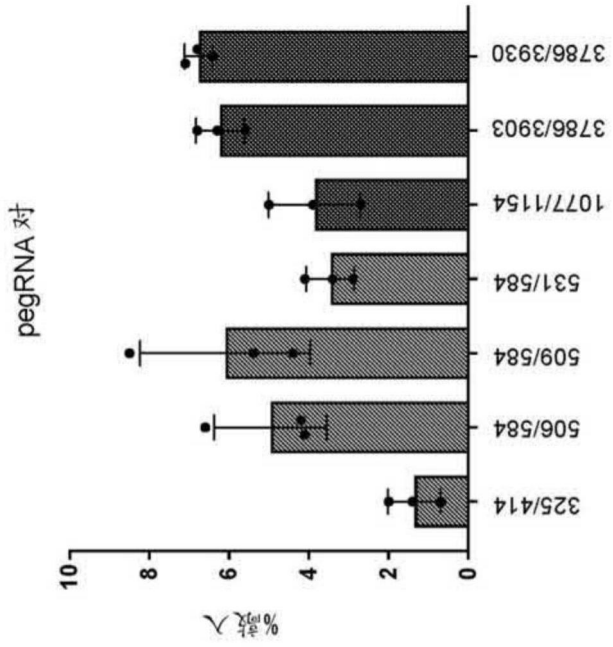

pegRNA 对
325/414
506/584
509/584
531/584
1077/1154
3786/3903
3786/3930
0
2
4
6
8
10
%插入

图110C

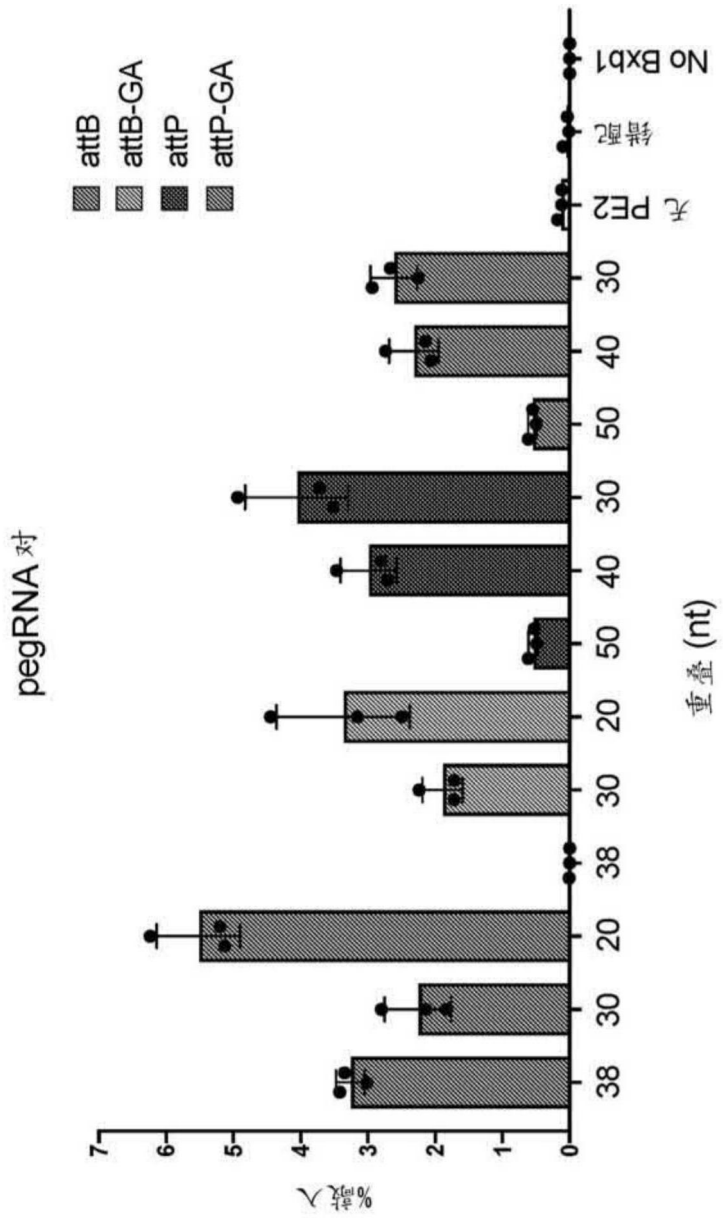

attB
attB-GA
attP
attP-GA
pegRNA对
重叠 (nt)
无 PE2
错配
No Bxb1
0
1
2
3
4
5
6
7
38
30
20
30
38
20
30
40
50
30
40
50
30
40
50

图110D

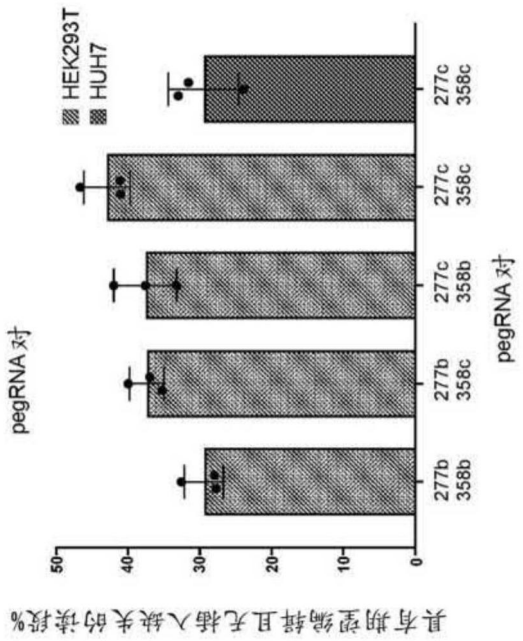

HEK293T
HUH7
277c 358c
277c 358c
277c 358b
277b 358c
277b 358b
pegRNA对
pegRNA对
0
10
20
30
40
50
具有期望编辑且无插入缺失的读段%

图110E

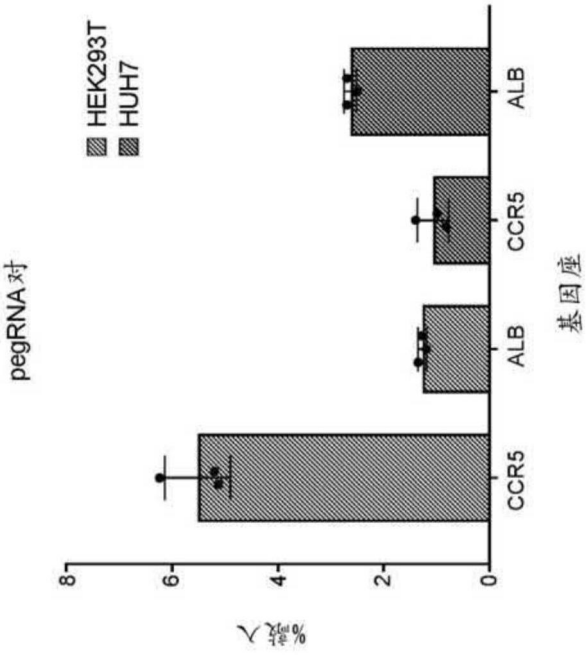

HEK293T
HUH7
pegRNA对
CCR5
ALB
CCR5
ALB
基因座
0
2
4
6
8
%插入

图110F

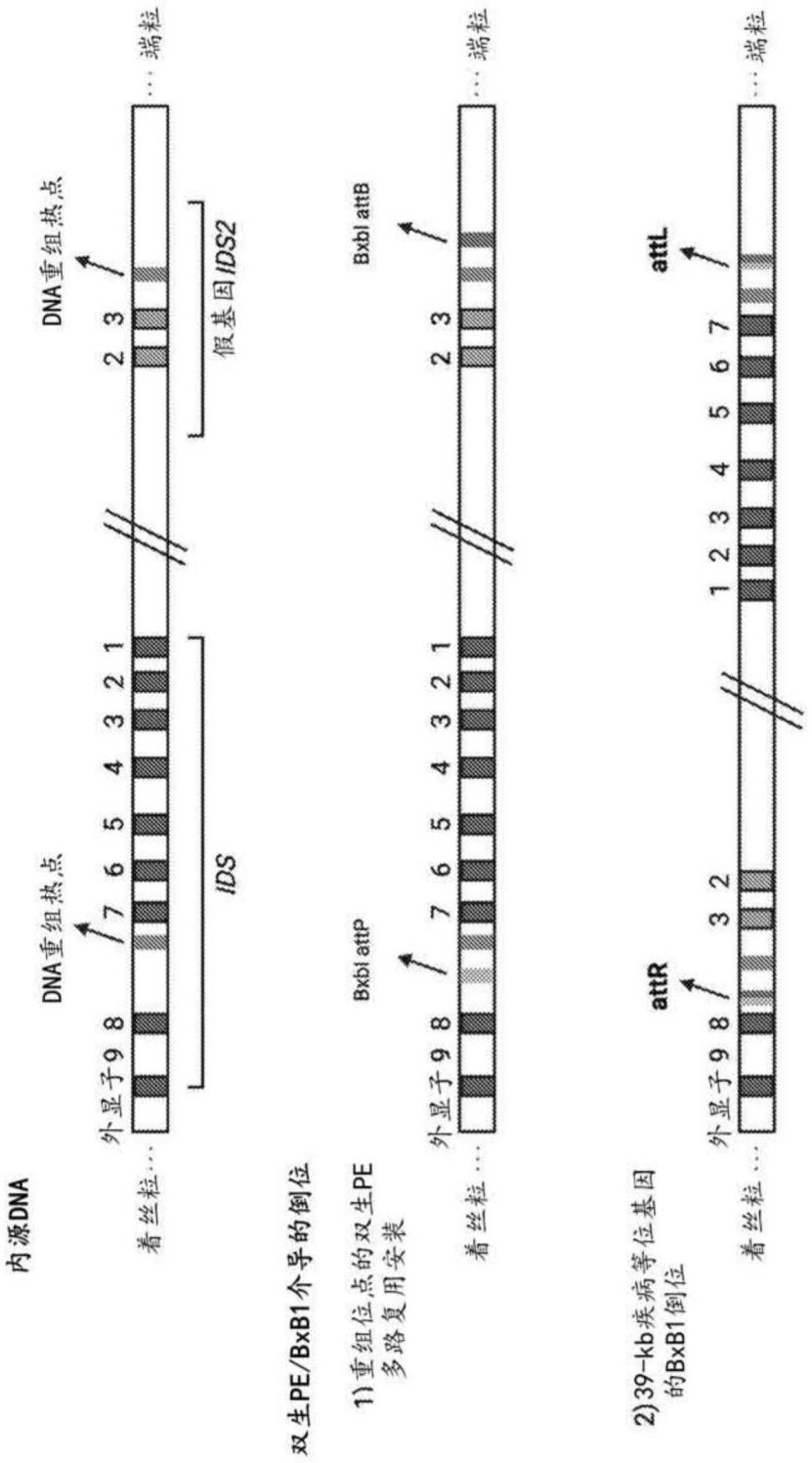

内源DNA
DNA重组热点
外显子 9 8 7 6 5 4 3 2 1
着丝粒 ···
··· 端粒
IDS
假基因IDS2
双生PE/BxB1介导的倒位
1)重组位点的双生PE
多路复用安装
Bxbl attP
Bxbl attB
2)39-kb疾病等位基因
的BxB1倒位
attR
attL
1 2 3 4 5 6 7

图111A

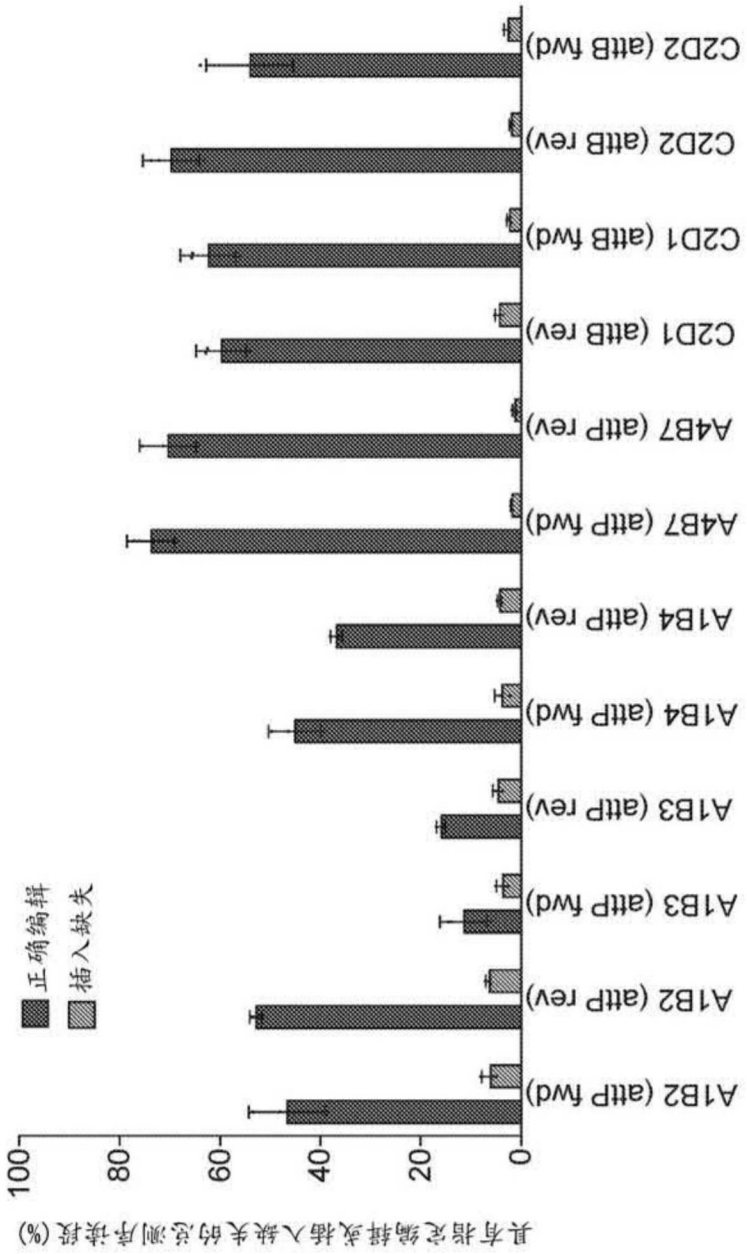

具有指定编辑或插入缺失的总测序读段 (%)
100
80
60
40
20
0
正确编辑
插入缺失
A1B2 (attP fwd)
A1B2 (attP rev)
A1B3 (attP fwd)
A1B3 (attP rev)
A1B4 (attP fwd)
A1B4 (attP rev)
A4B7 (attP fwd)
A4B7 (attP rev)
C2D1 (attB fwd)
C2D1 (attB rev)
C2D2 (attB fwd)
C2D2 (attB rev)

图111B

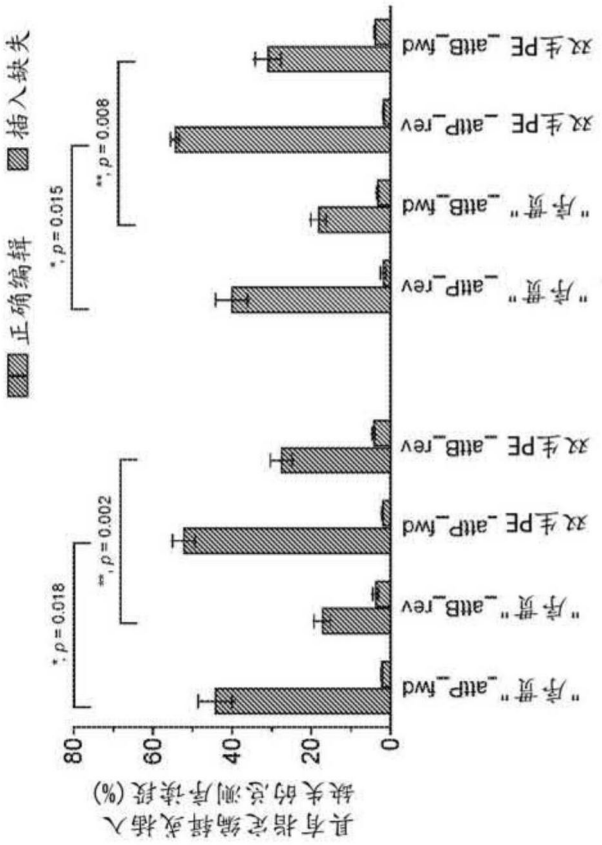

具有指定编辑或插入
缺失的总测序读段(%)
0
20
40
60
80
*, p = 0.018
**, p = 0.002
*, p = 0.015
**, p = 0.008
正确编辑
插入缺失
"序贯"_attP_fwd
"序贯"_attB_rev
双生PE_attP_fwd
双生PE_attB_rev
"序贯"_attP_rev
"序贯"_attB_fwd
双生PE_attP_rev
双生PE_attB_fwd

图111C

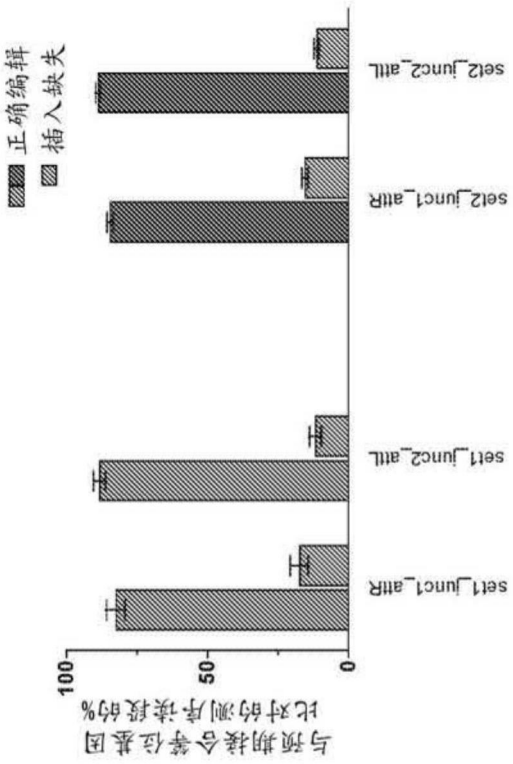

正确编辑
插入缺失
与预期接合等位基因
比对的测序读段的%
100
50
0
set1_Junc1_attR
set1_Junc2_attL
set2_Junc1_attR
set2_Junc2_attL

图111D

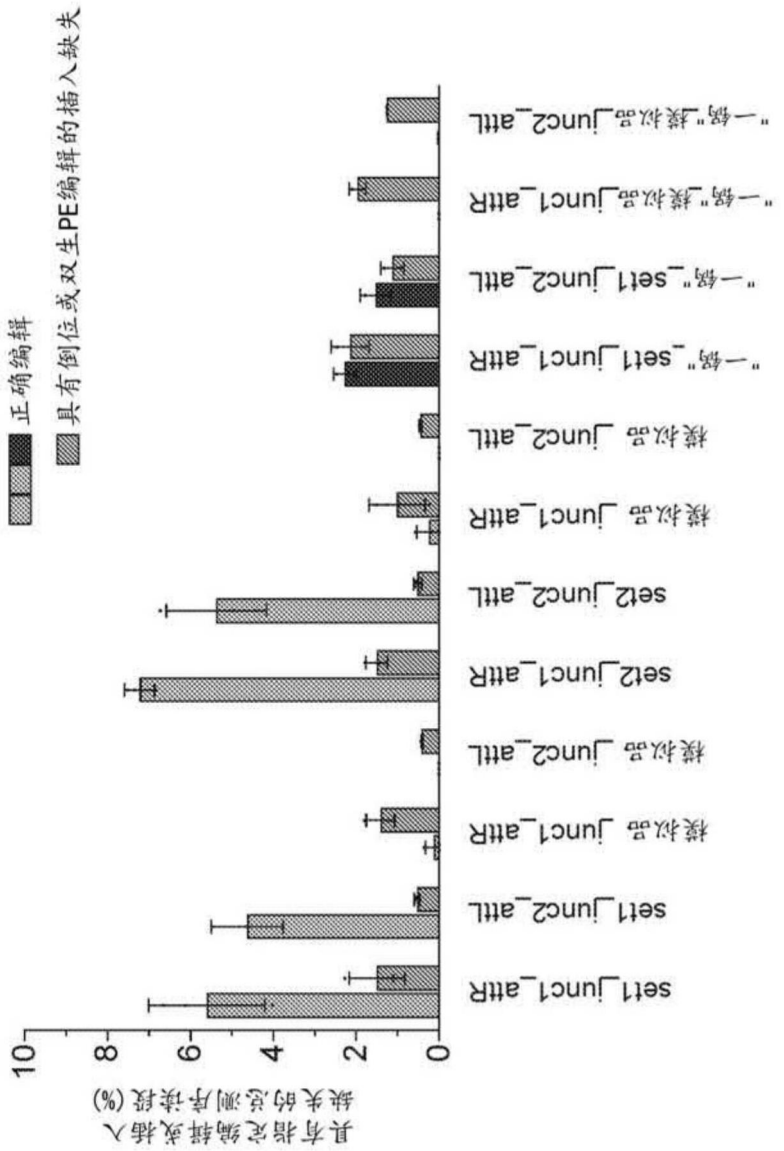

正确编辑
具有倒位或双生PE编辑的插入缺失
具有指定编辑或插入缺失的总测序读段(%)
0
2
4
6
8
10
set1_Junc1_attR
set1_Junc2_attL
模拟_Junc1_attR
模拟_Junc2_attL
set2_Junc1_attR
set2_Junc2_attL
模拟_Junc1_attR
模拟_Junc2_attL
"一般"_set1_Junc1_attR
"一般"_set1_Junc2_attL
"一般"_模拟_Junc1_attR
"一般"_模拟_Junc2_attL

图111E

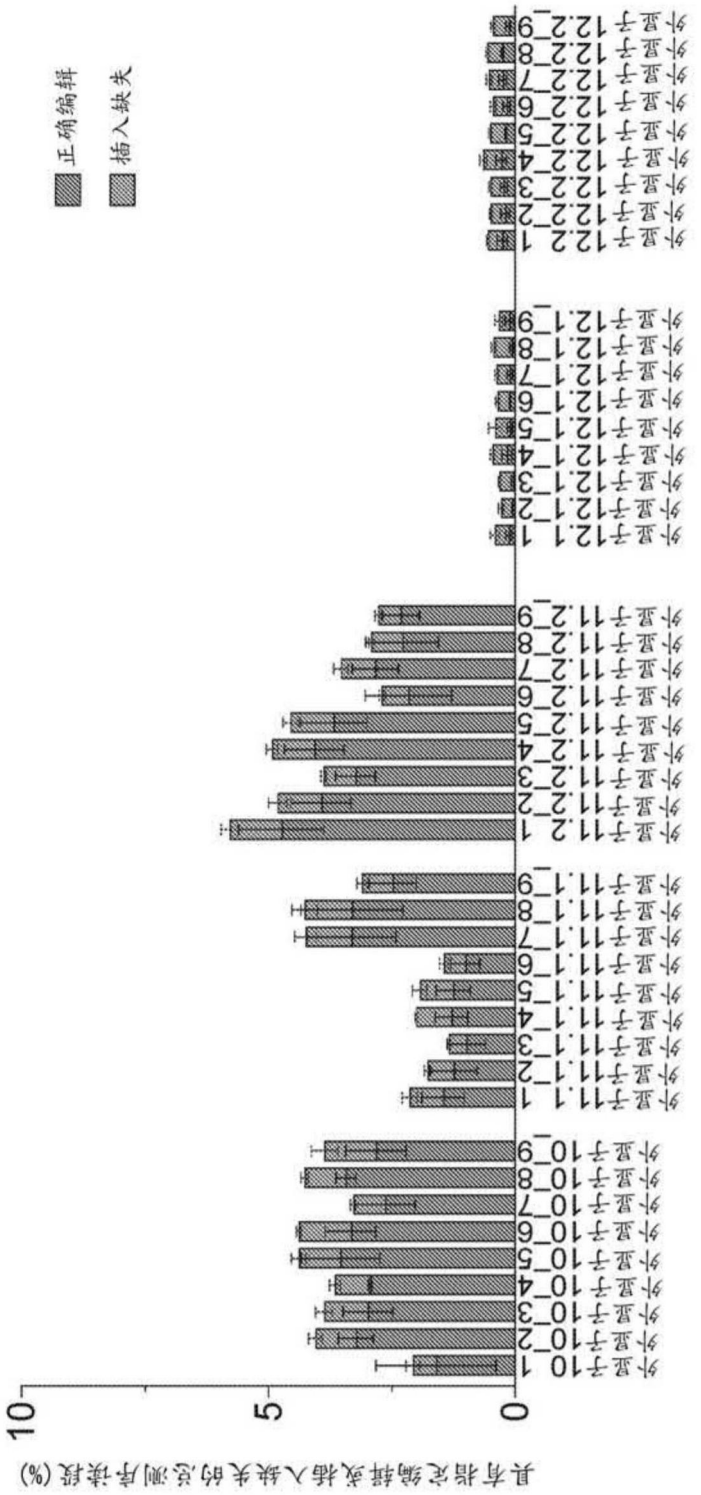


图112

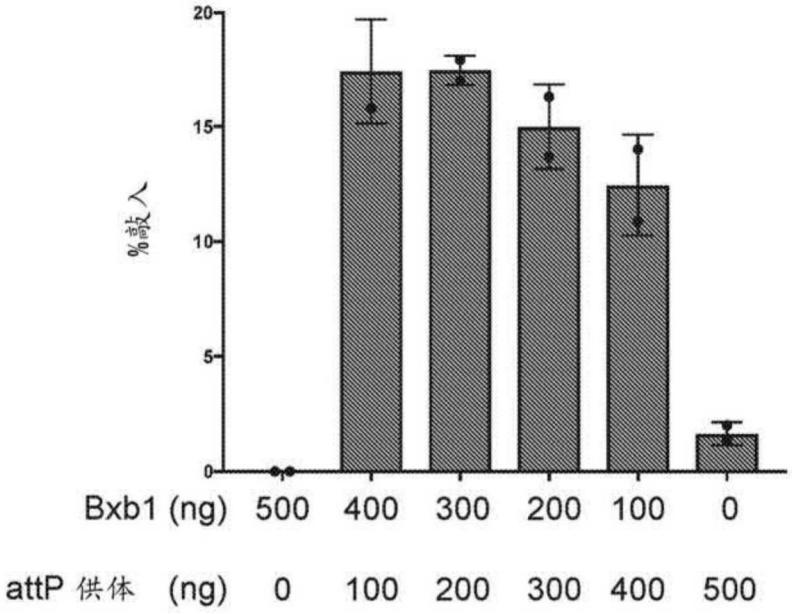

20
15
10
5
0
%敲入
Bxb1 (ng) 500 400 300 200 100 0
attP 供体 (ng) 0 100 200 300 400 500

图113

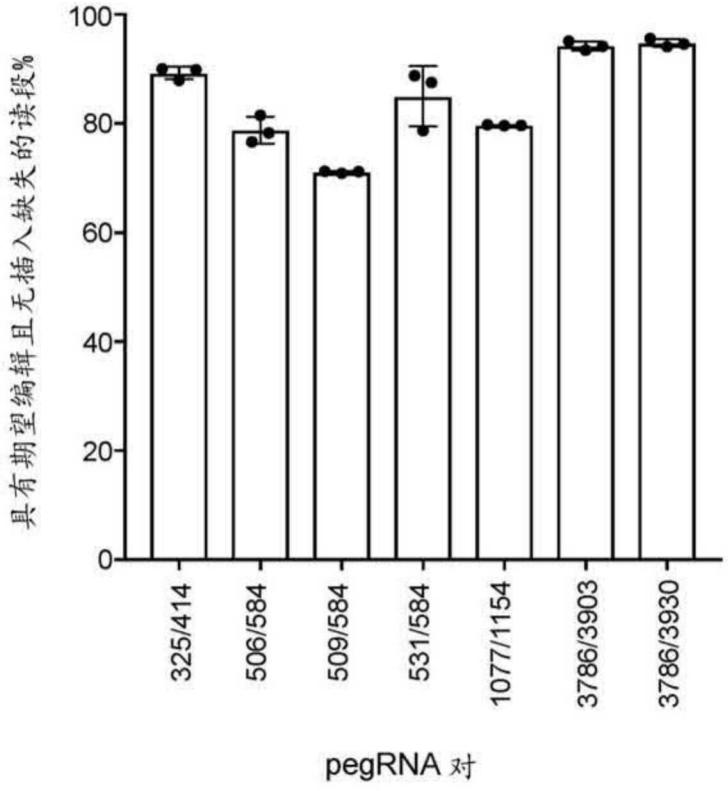

100
80
60
40
20
0
具有期望编辑且无插入缺失的读段%
325/414
506/584
509/584
531/584
1077/1154
3786/3903
3786/3930
pegRNA 对

图114

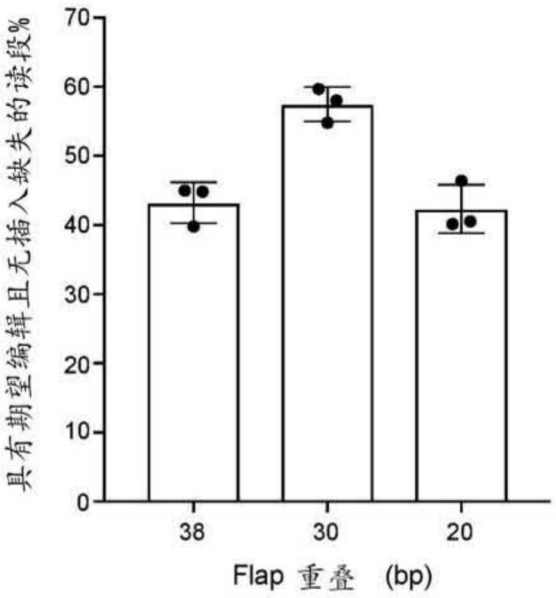

具有期望编辑且无插入缺失的读段%
70
60
50
40
30
20
10
0
38
30
20
Flap 重叠 (bp)

图115A

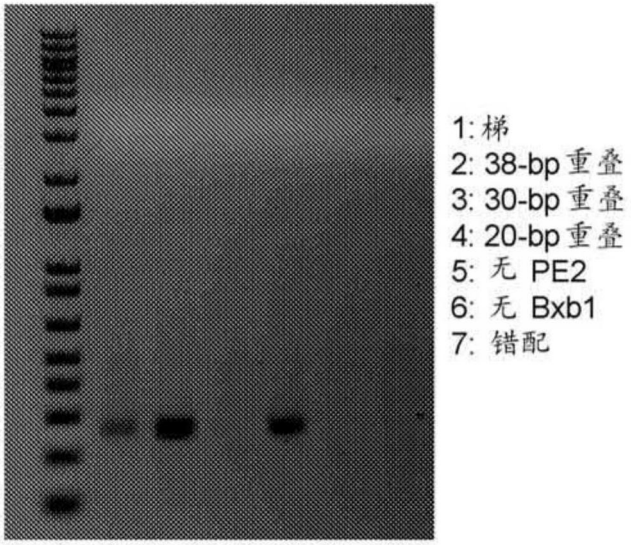

1: 梯
2: 38-bp 重叠
3: 30-bp 重叠
4: 20-bp 重叠
5: 无 PE2
6: 无 Bxb1
7: 错配

图115B

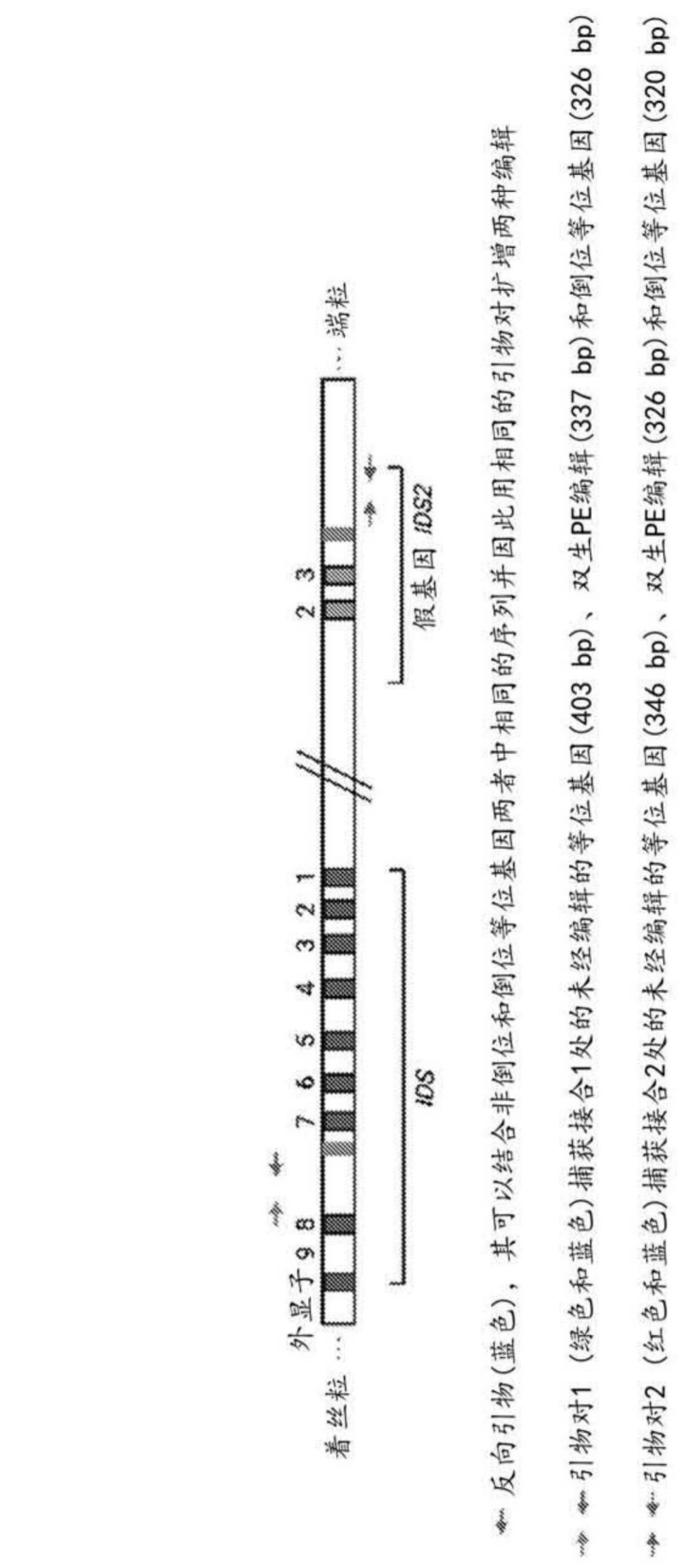

着丝粒
外显子9 8 7 6 5 4 3 2 1
2 3
端粒
IDS
假基因 IDS2
反向引物(蓝色)，其可以结合非倒位和倒位等位基因两者中相同的序列并因此用相同的引物对扩增两种编辑
引物对1 (绿色和蓝色)捕获接合1处的未经编辑的等位基因(403 bp)、双生PE编辑(337 bp)和倒位等位基因(326 bp)
引物对2 (红色和蓝色)捕获接合2处的未经编辑的等位基因(346 bp)、双生PE编辑(326 bp)和倒位等位基因(320 bp)

图116