



(12)发明专利申请

(10)申请公布号 CN 109923211 A

(43)申请公布日 2019.06.21

(21)申请号 201780065641.7

(22)申请日 2017.09.08

(30)优先权数据

62/385,079 2016.09.08 US

62/414,279 2016.10.28 US

(85)PCT国际申请进入国家阶段日

2019.04.23

(86)PCT国际申请的申请数据

PCT/US2017/050774 2017.09.08

(87)PCT国际申请的公布数据

W02018/049226 EN 2018.03.15

(71)申请人 蓝鸟生物公司

地址 美国马萨诸塞州剑桥宾尼街60号

(72)发明人 贾斯迪普·曼恩 乔尔·盖伊

乔丹·贾儒尔 张珏

(74)专利代理机构 北京英赛嘉华知识产权代理有限公司 11204

代理人 王达佐 洪欣

(51)Int.Cl.

C12N 15/66(2006.01)

C12N 15/09(2006.01)

C12N 5/10(2006.01)

C12N 9/22(2006.01)

C12N 15/90(2006.01)

C07K 16/30(2006.01)

C07K 14/54(2006.01)

C07K 16/28(2006.01)

A61K 38/46(2006.01)

A61K 35/17(2006.01)

权利要求书13页 说明书92页

序列表118页 附图44页

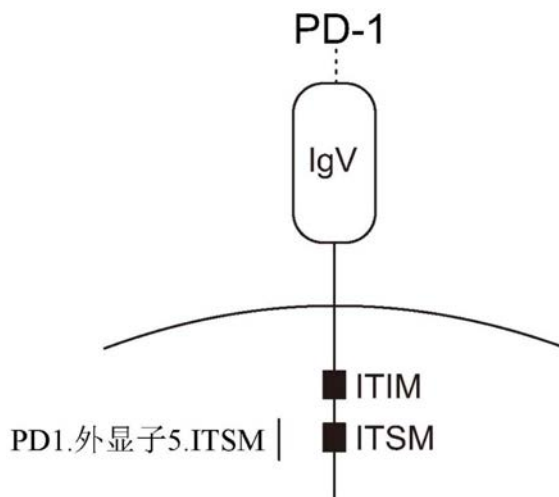
(54)发明名称

PD-1归巢核酸内切酶变体、组合物及使用方法

法

(57)摘要

本公开提供了改良的用于编辑PD-1基因的基因组编辑组合物和方法。本公开还提供了用于预防、治疗或改善癌症、感染性疾病、自身免疫性疾病、炎性疾病或免疫缺陷的至少一种症状的基因组编辑过的细胞。



1. 一种多肽,其包含使人程序性细胞死亡蛋白1 (PD-1) 基因中的靶位点裂解的归巢核酸内切酶 (HE) 变体。

2. 根据权利要求1所述的多肽,其中所述HE变体是LAGLIDADG归巢核酸内切酶 (LHE) 变体。

3. 根据权利要求1或权利要求2所述的多肽,其中所述多肽包含所述HE变体的生物活性片段。

4. 根据权利要求3所述的多肽,其中相较于相应野生型HE,所述生物活性片段缺乏1、2、3、4、5、6、7或8个N末端氨基酸。

5. 根据权利要求4所述的多肽,其中相较于相应野生型HE,所述生物活性片段缺乏4个N末端氨基酸。

6. 根据权利要求4所述的多肽,其中相较于相应野生型HE,所述生物活性片段缺乏8个N末端氨基酸。

7. 根据权利要求3所述的多肽,其中相较于相应野生型HE,所述生物活性片段缺乏1、2、3、4或5个C末端氨基酸。

8. 根据权利要求7所述的多肽,其中相较于相应野生型HE,所述生物活性片段缺乏C末端氨基酸。

9. 根据权利要求7所述的多肽,其中相较于相应野生型HE,所述生物活性片段缺乏2个C末端氨基酸。

10. 根据权利要求1至9中任一项所述的多肽,其中所述HE变体是选自由以下组成的组的LHE变体:I-AabMI、I-AaeMI、I-AniI、I-ApaMI、I-CapIII、I-CapIV、I-CkaMI、I-CpaMI、I-CpaMII、I-CpaMIII、I-CpaMIV、I-CpaMV、I-CpaV、I-CraMI、I-EjeMI、I-GpeMI、I-GpiI、I-GzeMI、I-GzeMII、I-GzeMIII、I-HjeMI、I-LtrII、I-LtrI、I-LtrWI、I-MpeMI、I-MveMI、I-NcrII、I-NcrI、I-NcrMI、I-OheMI、I-OnuI、I-OsoMI、I-OsoMII、I-OsoMIII、I-OsoMIV、I-PanMI、I-PanMII、I-PanMIII、I-PnoMI、I-ScuMI、I-SmaMI、I-SscMI以及I-Vdi141I。

11. 根据权利要求1至10中任一项所述的多肽,其中所述HE变体是选自由以下组成的组的LHE变体:I-CpaMI、I-HjeMI、I-OnuI、I-PanMI以及SmaMI。

12. 根据权利要求1至11中任一项所述的多肽,其中所述HE变体是I-OnuI LHE变体。

13. 根据权利要求1至12中任一项所述的多肽,其中所述HE变体在如SEQ ID NO:1-5中所述的I-OnuI LHE氨基酸序列或其生物活性片段的DNA识别界面中选自由以下组成的组的氨基酸位置处包含一个或多个氨基酸取代:19、24、26、28、30、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76 77、78、80、82、168、180、182、184、186、188、189、190、191、192、193、195、197、199、201、203、223、225、227、229、231、232、234、236、238以及240。

14. 根据权利要求1至13中任一项所述的多肽,其中所述HE变体在如SEQ ID NO:1-5中所述的I-OnuI LHE氨基酸序列或其生物活性片段的DNA识别界面中选自由以下组成的组的氨基酸位置处包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:19、24、26、28、30、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76 77、78、80、82、168、180、182、184、186、188、189、190、191、192、193、195、197、199、201、203、223、225、227、229、231、232、234、236、238以及240。

15. 根据权利要求1至14中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位

点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:26、28、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76、78、80、138、143、159、168、178、180、182、184、186、188、189、190、191、192、193、197、199、201、203、207、223、224、225、227、229、232、236以及238。

16. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40K、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、S72R、N75S、A76Y、S78K、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、S201M、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

17. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40K、E42R、G44R、Q46E、T48D、V68K、A70Y、S72Q、N75S、A76Y、S78K、K80R、L138M、T143N、S159P、F168L、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、K225R、K229I、F232K、D236E以及V238E。

18. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72R、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

19. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

20. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、

K225R、F232K、D236E以及V238E。

21. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201M、T203G、Y223R、K225R、F232K、D236E以及V238E。

22. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个,或甚至更优选至少40个或更多个氨基酸取代:26、28、30、32、34、35、36、37、38、40、42、44、46、48、68、70、72、75、76、78、80、100、132、138、143、155、159、178、180、184、186、189、190、191、192、193、195、201、203、207、223、225、227、232、236、238以及240。

23. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、V37G、G38R、S40H、E42R、G44S、Q46A、Q46T、T48V、T48M、V68I、V68S、A70T、A70Y、A70L、S72D、S72N、N75R、N75H、A76Y、S78R、S78T、K80R、K80C、K80E、K80V、T82F、T82Y、I100V、V132A、L138M、T143N、S155G、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

24. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46A、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80R、I100V、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

25. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、V37G、G38R、S40H、E42R、G44S、Q46T、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80C、I100V、V132A、L138M、T143N、S155G、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

26. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、

R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68I、A70T、S72N、N75H、A76Y、S78T、K80R、I100V、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

27. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70Y、S72N、N75H、A76Y、K80E、T82F、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

28. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70L、S72N、N75H、A76Y、K80V、T82Y、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

29. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37G、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70T、S72N、N75H、A76Y、K80V、T82Y、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

30. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:24、26、28、30、32、34、35、36、37、38、40、41、42、44、46、48、68、70、72、74、75、76、78、80、82、116、138、143、159、168、178、180、182、184、186、188、189、190、191、192、193、195、199、203、207、225、227、229、232、236以及238。

31. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:S24C、L26Q、R28Y、R28H、R30S、N32V、N32L、K34N、K34R、S35N、S35T、S36R、V37S、V37T、G38R、G38K、S40R、T41A、E42R、G44S、G44R、Q46E、Q46A、T48E、V68I、A70N、S72I、D74N、N75T、N75R、A76S、A76R、S78R、K80S、T82G、T82R、V116L、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203A、T203S、K207R、K225N、K225T、K227W、K227S、K229A、K229P、F232R、W234A、

W234D、D236E以及V238R。

32. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28Y、R30S、N32V、K34N、S35N、S36R、V37S、G38R、S40R、T41A、E42R、G44R、Q46A、T48E、A70N、S72I、N75T、A76S、S78R、K80S、T82G、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203A、K207R、K225N、K227W、K229A、F232R、W234A、D236E以及V238R。

33. 根据权利要求1至15中任一项所述的多肽,其中所述HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:S24C、R28H、N32L、K34R、S35T、V37T、G38K、S40R、E42R、G44S、Q46E、T48E、V68I、A70N、S72I、D74N、N75R、A76R、S78R、K80S、T82R、V116L、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203S、K207R、K225T、K227S、K229P、F232R、W234D、D236E以及V238R。

34. 根据权利要求1至33中任一项所述的多肽,其中所述HE变体包含与SEQ ID NO:6-14和60-63中任一个或其生物活性片段中所述的氨基酸序列至少80%、优选至少85%、更优选至少90%、或甚至更优选至少95%一致的氨基酸序列。

35. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:6中或其生物活性片段所述的氨基酸序列。

36. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:7或其生物活性片段中所述的氨基酸序列。

37. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:8或其生物活性片段中所述的氨基酸序列。

38. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:9或其生物活性片段中所述的氨基酸序列。

39. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:10或其生物活性片段中所述的氨基酸序列。

40. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:11或其生物活性片段中所述的氨基酸序列。

41. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:12或其生物活性片段中所述的氨基酸序列。

42. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:13或其生物活性片段中所述的氨基酸序列。

43. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:14或其生物活性片段中所述的氨基酸序列。

44. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:60或其生物活性片段中所述的氨基酸序列。

45. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:61或其

生物活性片段中所述的氨基酸序列。

46. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:62或其生物活性片段中所述的氨基酸序列。

47. 根据权利要求1至34中任一项所述的多肽,其中所述HE变体包含SEQ ID NO:63或其生物活性片段中所述的氨基酸序列。

48. 根据权利要求1至21和30至39中任一项所述的多肽,其中所述多肽结合SEQ ID NO:25中所述的聚核苷酸序列。

49. 根据权利要求1至14、22至29、40至41和44至47中任一项所述的多肽,其中所述多肽结合SEQ ID NO:30中所述的聚核苷酸序列。

50. 根据权利要求1至14、30至33和42至43中任一项所述的多肽,其中所述多肽结合SEQ ID NO:35中所述的聚核苷酸序列。

51. 根据权利要求1至50中任一项所述的多肽,另外包含DNA结合结构域。

52. 根据权利要求51所述的多肽,其中所述DNA结合结构域选自由以下组成的组:TALE DNA结合结构域和锌指DNA结合结构域。

53. 根据权利要求52所述的多肽,其中所述TALE DNA结合结构域包含约9.5个TALE重复单元至约15.5个TALE重复单元。

54. 根据权利要求52或权利要求53所述的多肽,其中所述TALE DNA结合结构域结合所述PD-1基因中的聚核苷酸序列。

55. 根据权利要求52至54中任一项所述的多肽,其中所述TALE DNA结合结构域结合SEQ ID NO:26中所述的聚核苷酸序列。

56. 根据权利要求55所述的多肽,其中所述多肽结合并裂解SEQ ID NO:27中所述的聚核苷酸序列。

57. 根据权利要求52至54中任一项所述的多肽,其中所述TALE DNA结合结构域结合SEQ ID NO:31中所述的聚核苷酸序列。

58. 根据权利要求57所述的多肽,其中所述多肽结合并裂解SEQ ID NO:32中所述的聚核苷酸序列。

59. 根据权利要求52至54中任一项所述的多肽,其中所述TALE DNA结合结构域结合SEQ ID NO:36中所述的聚核苷酸序列。

60. 根据权利要求59所述的多肽,其中所述多肽结合并裂解SEQ ID NO:37中所述的聚核苷酸序列。

61. 根据权利要求52所述的多肽,其中所述锌指DNA结合结构域包含2、3、4、5、6、7或8个锌指基元。

62. 根据权利要求1至61中任一项所述的多肽,另外包含肽连接子和末端加工酶或其生物活性片段。

63. 根据权利要求1至61中任一项所述的多肽,另外包含病毒自裂解2A肽和末端加工酶或其生物活性片段。

64. 根据权利要求62或权利要求63所述的多肽,其中所述末端加工酶或其生物活性片段具有5'-3'核酸外切酶、5'-3'碱性核酸外切酶、3'-5'核酸外切酶、5'瓣状核酸内切酶、解螺旋酶或不依赖模板的DNA聚合酶活性。

65. 根据权利要求62至64中任一项所述的多肽,其中所述末端加工酶包含Trex2或其生物活性片段。

66. 根据权利要求1至65中任一项所述的多肽,其中所述多肽包含SEQ ID NO:15-23和64中任一个或其生物活性片段所述的氨基酸序列。

67. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:15或其生物活性片段中所述的氨基酸序列。

68. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:16或其生物活性片段中所述的氨基酸序列。

69. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:17或其生物活性片段中所述的氨基酸序列。

70. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:18或其生物活性片段中所述的氨基酸序列。

71. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:19或其生物活性片段中所述的氨基酸序列。

72. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:20或其生物活性片段中所述的氨基酸序列。

73. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:21或其生物活性片段中所述的氨基酸序列。

74. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:22或其生物活性片段中所述的氨基酸序列。

75. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:23或其生物活性片段中所述的氨基酸序列。

76. 根据权利要求66所述的多肽,其中所述多肽包含SEQ ID NO:64或其生物活性片段中所述的氨基酸序列。

77. 根据权利要求1至61中任一项所述的多肽,其中所述多肽裂解所述人PD-1基因中SEQ ID NO:25、27、30、32、35或37中所述的聚核苷酸序列。

78. 一种聚核苷酸,其编码根据权利要求1至77中任一项所述的多肽。

79. 一种mRNA,其编码根据权利要求1至77中任一项所述的多肽。

80. 一种cDNA,其编码根据权利要求1至77中任一项所述的多肽。

81. 一种载体,其包含编码根据权利要求1至77中任一项所述的多肽的聚核苷酸。

82. 一种细胞,其编码根据权利要求1至77中任一项所述的多肽。

83. 一种细胞,其包含编码根据权利要求1至77中任一项所述的多肽的聚核苷酸。

84. 一种细胞,其包含根据权利要求81所述的载体。

85. 一种细胞,其包含一个或多个由根据权利要求1至77中任一项所述的多肽引入的基因组修饰。

86. 根据权利要求82至85中任一项所述的细胞,其中所述细胞包含编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体中的一种或多种的聚核苷酸。

87. 根据权利要求86所述的细胞,其中所述聚核苷酸还包含RNA聚合酶II启动子,所述RNA聚合酶II启动子可操作地连接至编码所述免疫效力增强剂、免疫抑制信号阻断剂或工

程改造的抗原受体的聚核苷酸。

88. 根据权利要求87所述的细胞,其中所述RNA聚合酶II启动子选自由以下组成的组:短EF1 α 启动子、长EF1 α 启动子、人ROSA 26基因座、泛素C (UBC) 启动子、磷酸甘油酸激酶-1 (PGK) 启动子、巨细胞病毒增强子/鸡 β -肌动蛋白 (CAG) 启动子、 β -肌动蛋白启动子和骨髓增生肉瘤病毒增强子、负控制区缺失、dl587rev引物结合位点取代 (MND) 的启动子。

89. 根据权利要求86至88中任一项所述的细胞,其中所述聚核苷酸还编码一个或多个自裂解病毒肽,所述自裂解病毒肽可操作地连接至所述免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体;散布于其间和/或与其侧接。

90. 根据权利要求89所述的细胞,其中所述自裂解病毒肽是2A肽。

91. 根据权利要求86至90中任一项所述的细胞,其中所述聚核苷酸还包含异源聚腺苷酸化信号。

92. 根据权利要求86至91中任一项所述的细胞,其中所述免疫抑制信号阻断剂包含抵消免疫抑制因子作用的酶功能。

93. 根据权利要求92所述的细胞,其中所述免疫抑制信号阻断剂包含犬尿氨酸酶活性。

94. 根据权利要求86至91中任一项所述的细胞,其中所述免疫抑制信号阻断剂包含:

(a) 结合免疫抑制因子的胞外结构域,任选地其中所述胞外结构域是抗体或其抗原结合片段;

(b) 结合免疫抑制因子的胞外结构域和跨膜结构域;或

(c) 结合免疫抑制因子的胞外结构域、跨膜结构域和经过修饰的胞内结构域,所述经过修饰的胞内结构域不能够将免疫抑制信号转导至所述细胞中。

95. 根据权利要求86至91中任一项所述的细胞,其中所述免疫抑制信号阻断剂是显性负性TGF β RII受体。

96. 根据权利要求86至91中任一项所述的细胞,其中所述免疫效力增强剂选自由以下组成的组:双特异性T细胞接合分子 (BiTE)、免疫增强因子和翻转受体。

97. 根据权利要求96所述的细胞,其中所述免疫增强因子选自由以下组成的组:细胞因子、趋化因子、细胞毒素、细胞因子受体以及其变体。

98. 根据权利要求97所述的细胞,其中所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。

99. 根据权利要求97所述的细胞,其中所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21并且可操作地连接至内源性PD-1启动子。

100. 根据权利要求97所述的细胞,其中所述细胞因子是IL-7并且可操作地连接至所述内源性PD-1启动子。

101. 根据权利要求97所述的细胞,其中所述细胞因子是IL-12并且可操作地连接至所述内源性PD-1启动子。

102. 根据权利要求97所述的细胞,其中所述细胞因子是IL-15并且可操作地连接至所述内源性PD-1启动子。

103. 根据权利要求96所述的细胞,其中所述翻转受体包含PD-1胞外结构域和跨膜结构域;以及与所述PD-1跨膜结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或CD3 ζ 的胞内结构域。

104. 根据权利要求96所述的细胞,其中所述翻转受体包含PD-1胞外结构域;自CD3多肽、CD4、CD8 α 、CD28、CD134或CD137分离的跨膜结构域;以及与所述PD-1胞外结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或CD3 ζ 的胞内结构域。

105. 根据权利要求96所述的细胞,其中所述翻转受体包含PD-1胞外结构域;以及与所述PD-1胞外结构域的C末端同框融合的自CD3多肽、CD4、CD8 α 、CD28、CD134或CD137分离的跨膜结构域和胞内结构域。

106. 根据权利要求86至91中任一项所述的细胞,其中所述工程改造的抗原受体选自自由以下组成的组:工程改造的TCR、CAR、DARIC或 ζ 因子。

107. 根据权利要求106所述的细胞,其中所述工程改造的受体未被整合至所述PD-1基因中。

108. 根据权利要求86至91中任一项所述的细胞,其中所述编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体中的一种或多种的聚核苷酸被整合至所述PD-1基因中。

109. 根据权利要求86至91中任一项所述的细胞,其中包含所述编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体中的一种或多种的聚核苷酸的供体修复模板被整合至所述PD-1基因中由根据权利要求1至77中任一项所述的多肽引入的DNA双链断裂位点处。

110. 根据权利要求82至109中任一项所述的细胞,其中所述细胞是造血细胞。

111. 根据权利要求82至110中任一项所述的细胞,其中所述细胞是T细胞。

112. 根据权利要求82至111中任一项所述的细胞,其中所述细胞是CD3⁺、CD4⁺和/或CD8⁺细胞。

113. 根据权利要求82至112中任一项所述的细胞,其中所述细胞是免疫效应细胞。

114. 根据权利要求82至113中任一项所述的细胞,其中所述细胞是细胞毒性T淋巴细胞(CTL)、肿瘤浸润性淋巴细胞(TIL)或辅助T细胞。

115. 根据权利要求82至113中任一项所述的细胞,其中所述细胞是自然杀手(NK)细胞或自然杀手T(NKT)细胞。

116. 根据权利要求82至115中任一项所述的细胞,其中所述细胞的来源是外周血液单核细胞、骨髓、淋巴结组织、脐带血、胸腺组织、来自感染部位的组织、腹水、胸腔积液、脾组织或肿瘤。

117. 根据权利要求82至116中任一项所述的细胞,其中所述细胞包含一个或多个经过修饰的PD-1等位基因。

118. 根据权利要求117所述的细胞,其中所述一个或多个经过修饰的PD-1等位基因是无功能的或具有基本上减弱的PD-1功能和/或细胞内信号传导。

119. 根据权利要求82至118中任一项所述的细胞,其中所述细胞包含被引入所述一个或多个经过修饰的PD-1等位基因中的编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的核酸。

120. 根据权利要求82至118中任一项所述的细胞,其中所述细胞包含被引入所述一个或多个经过修饰的PD-1等位基因中的编码免疫效力增强剂或免疫抑制信号阻断剂的核酸并且所述细胞还包含未被引入所述一个或多个经过修饰的PD-1等位基因中的工程改造的

抗原受体。

121. 多种细胞,其包含一种或多种根据权利要求82至120中任一项所述的细胞。

122. 一种组合物,其包含一种或多种根据权利要求82至121中任一项所述的细胞。

123. 一种组合物,其包含一种或多种根据权利要求82至121中任一项所述的细胞和生理学上可接受的载体。

124. 一种编辑细胞中的人PD-1基因的方法,其包含:将编码根据权利要求1至77中任一项所述的多肽的聚核苷酸引入所述细胞中,其中所述多肽的表达在人PD-1基因中的靶位点处产生双链断裂。

125. 一种编辑细胞中的人PD-1基因的方法,其包含:将编码根据权利要求1至77中任一项所述的多肽的聚核苷酸引入所述细胞中,其中所述多肽的表达在人PD-1基因中的靶位点处产生双链断裂,其中所述断裂是通过非同源末端接合(NHEJ)修复。

126. 一种编辑细胞中的人PD-1基因的方法,其包含:将编码根据权利要求1至77中任一项所述的多肽的聚核苷酸和供体修复模板引入所述细胞中,其中所述多肽的表达在人PD-1基因中的靶位点处产生双链断裂,并且所述供体修复模板是通过同源介导的修复(HDR)并入所述人PD-1基因中的所述双链断裂(DSB)位点处。

127. 根据权利要求124至126中任一项所述的方法,其中所述细胞是造血细胞。

128. 根据权利要求124至127中任一项所述的方法,其中所述细胞是T细胞。

129. 根据权利要求124至128中任一项所述的方法,其中所述细胞是CD3⁺、CD4⁺和/或CD8⁺细胞。

130. 根据权利要求124至129中任一项所述的方法,其中所述细胞是免疫效应细胞。

131. 根据权利要求124至130中任一项所述的方法,其中所述细胞是细胞毒性T淋巴细胞(CTL)、肿瘤浸润性淋巴细胞(TIL)或辅助T细胞。

132. 根据权利要求124至130中任一项所述的方法,其中所述细胞是自然杀手(NK)细胞或自然杀手T(NKT)细胞。

133. 根据权利要求124至132中任一项所述的方法,其中所述细胞的来源是外周血液单核细胞、骨髓、淋巴结组织、脐带血、胸腺组织、来自感染部位的组织、腹水、胸腔积液、脾组织或肿瘤。

134. 根据权利要求124至133中任一项所述的方法,其中编码所述多肽的聚核苷酸是mRNA。

135. 根据权利要求124至134中任一项所述的方法,其中将编码3'-5'核酸外切酶的聚核苷酸引入所述细胞中。

136. 根据权利要求124至134中任一项所述的方法,其中将编码Trex2或其生物活性片段的聚核苷酸引入所述细胞中。

137. 根据权利要求124至136中任一项所述的方法,其中所述供体修复模板编码PD-1基因或其部分,相较于野生型PD-1基因,所述PD-1基因或其部分包含一个或多个突变。

138. 根据权利要求124至136中任一项所述的方法,其中所述供体修复模板编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体中的一种或多种。

139. 根据权利要求138所述的方法,其中所述供体修复模板还包含可操作地连接至所述免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的RNA聚合酶II启动子。

140. 根据权利要求139所述的方法, 其中所述RNA聚合酶II启动子选自由以下组成的组: 短EF1 α 启动子、长EF1 α 启动子、人ROSA 26基因座、泛素C (UBC) 启动子、磷酸甘油酸激酶-1 (PGK) 启动子、巨细胞病毒增强子/鸡 β -肌动蛋白 (CAG) 启动子、 β -肌动蛋白启动子和骨髓增生肉瘤病毒增强子、负控制区缺失、d1587rev引物结合位点取代 (MND) 的启动子。

141. 根据权利要求138至140中任一项所述的方法, 其中所述供体修复模板还编码一种或多种自裂解病毒肽, 所述自裂解病毒肽可操作地连接至所述免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体; 散布于其间和/或与其侧接。

142. 根据权利要求141所述的方法, 其中所述自裂解病毒肽是2A肽。

143. 根据权利要求138至142中任一项所述的方法, 其中所述供体修复模板还包含异源聚腺苷酸化信号。

144. 根据权利要求138至143中任一项所述的方法, 其中所述免疫抑制信号阻断剂包含抵消免疫抑制因子作用的酶功能。

145. 根据权利要求144所述的方法, 其中所述免疫抑制信号阻断剂包含犬尿氨酸酶活性。

146. 根据权利要求138至143中任一项所述的方法, 其中所述免疫抑制信号阻断剂包含:

(a) 结合免疫抑制因子的胞外结构域, 任选地其中所述胞外结构域是抗体或其抗原结合片段;

(b) 结合免疫抑制因子的胞外结构域和跨膜结构域; 或

(c) 结合免疫抑制因子的胞外结构域、跨膜结构域和经过修饰的胞内结构域, 所述经过修饰的胞内结构域不能将免疫抑制信号转导至所述细胞中。

147. 根据权利要求138至143中任一项所述的方法, 其中所述免疫抑制信号阻断剂是显性负性TGF β R2受体。

148. 根据权利要求138至143中任一项所述的方法, 其中所述免疫效力增强剂选自由以下组成的组: 双特异性T细胞接合分子 (BiTE)、免疫增强因子和翻转受体。

149. 根据权利要求148所述的方法, 其中所述免疫增强因子选自由以下组成的组: 细胞因子、趋化因子、细胞毒素、细胞因子受体以及其变体。

150. 根据权利要求149所述的方法, 其中所述细胞因子选自由以下组成的组: IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。

151. 根据权利要求149所述的方法, 其中所述细胞因子选自由以下组成的组: IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21并且可操作地连接至内源性PD-1启动子。

152. 根据权利要求149所述的方法, 其中所述细胞因子是IL-7并且可操作地连接至所述内源性PD-1启动子。

153. 根据权利要求149所述的方法, 其中所述细胞因子是IL-12并且可操作地连接至所述内源性PD-1启动子。

154. 根据权利要求149所述的方法, 其中所述细胞因子是IL-15并且可操作地连接至所述内源性PD-1启动子。

155. 根据权利要求148所述的方法, 其中所述翻转受体包含PD-1胞外结构域和跨膜结构域; 以及与所述PD-1跨膜结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或

CD3 ζ 的胞内结构域。

156. 根据权利要求148所述的方法, 其中所述翻转受体包含PD-1胞外结构域; 自CD3多肽、CD4、CD8 α 、CD28、CD134或CD137分离的跨膜结构域; 以及与所述PD-1胞外结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或CD3 ζ 的胞内结构域。

157. 根据权利要求148所述的方法, 其中所述翻转受体包含PD-1胞外结构域; 以及与所述PD-1胞外结构域的C末端同框融合的自CD3多肽、CD4、CD8 α 、CD28、CD134或CD137分离的跨膜结构域和胞内结构域。

158. 根据权利要求138至143中任一项所述的方法, 其中所述工程改造的抗原受体选自由以下组成的组: 工程改造的TCR、CAR、Daric或 ζ 因子。

159. 根据权利要求138至158中任一项所述的方法, 其中所述供体修复模板包含与在所述DSB 5'端的人PD-1基因序列同源的5'同源臂和与在所述DSB 3'端的人PD-1基因序列同源的3'同源臂。

160. 根据权利要求159所述的方法, 其中所述5'同源臂和所述3'同源臂的长度独立地选自约100bp至约2500bp。

161. 根据权利要求159或权利要求160所述的方法, 其中所述5'同源臂和所述3'同源臂的长度独立地选自约600bp至约1500bp。

162. 根据权利要求159至161中任一项所述的方法, 其中所述5'同源臂是约1500bp并且所述3'同源臂是约1000bp。

163. 根据权利要求159至162中任一项所述的方法, 其中所述5'同源臂是约600bp并且所述3'同源臂是约600bp。

164. 根据权利要求159至163中任一项所述的方法, 其中使用病毒载体将所述供体修复模板引入所述细胞中。

165. 根据权利要求164所述的方法, 其中所述病毒载体是重组腺相关病毒载体(rAAV)或逆转录病毒。

166. 根据权利要求165所述的方法, 其中所述rAAV具有一个或多个来自AAV2的ITR。

167. 根据权利要求165或权利要求166所述的方法, 其中所述rAAV具有选自由以下组成的组的血清型: AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9以及AAV10。

168. 根据权利要求165至167中任一项所述的方法, 其中所述rAAV具有AAV2或AAV6血清型。

169. 根据权利要求165所述的方法, 其中所述逆转录病毒是慢病毒。

170. 根据权利要求169所述的方法, 其中所述慢病毒是整合酶缺陷型慢病毒(IDLV)。

171. 一种治疗、预防或改善癌症、感染性疾病、自身免疫性疾病、炎性疾病和免疫缺陷或其相关病况的至少一种症状的方法, 其包含向受试者施用有效量的根据权利要求122或权利要求123所述的组合物。

172. 一种治疗实体癌的方法, 其包含向受试者施用有效量的根据权利要求122或权利要求123所述的组合物。

173. 根据权利要求172所述的方法, 其中所述实体癌包含肝癌、胰腺癌、肺癌、乳癌、卵巢癌、前列腺癌、睾丸癌、膀胱癌、脑癌、肉瘤、头颈癌、骨癌、甲状腺癌、肾癌或皮肤癌。

174. 一种治疗血液恶性病的方法, 其包含向受试者施用有效量的根据权利要求122或

权利要求123所述的组合物。

175. 根据权利要求174所述的方法, 其中所述血液恶性病是白血病、淋巴瘤或多发性骨髓瘤。

PD-1归巢核酸内切酶变体、组合物及使用方法

[0001] 相关申请的交叉引用

[0002] 本申请根据35 U.S.C.119 (e) 要求2016年10月28日提交的美国临时申请第62/414,279号和2016年9月8日提交的美国临时申请第62/385,079号的权益,以上申请各自以全文引用的方式并入本文中。

[0003] 关于序列表的声明

[0004] 与本申请相关联的序列表是以文本格式提供以代替纸质副本,并以引用的方式并入本说明书中。含有序列表的文本文件的名称是BLBD_076_02W0_ST25.txt。文本文件是266KB,创建于2017年9月8日,并在本说明书提交的同时,通过EFS-Web以电子方式提交。

技术领域

[0005] 本公开涉及改良的基因组编辑组合物。更确切地说,本公开涉及核酸酶变体、组合物和使用其编辑人程序性细胞死亡蛋白1 (PD-1) 基因的方法。

背景技术

[0006] 全球癌症负担在1975年与2000年间增加一倍。癌症是世界范围内导致发病率和死亡率的第二大主导原因,且在2012年间有约一千四百一十例新增病例和八百二十万例癌症相关死亡。最常见的癌症是乳癌、肺和支气管癌症、前列腺癌、结肠和直肠癌、膀胱癌、皮肤黑素瘤、非霍奇金氏淋巴瘤 (non-Hodgkin lymphoma)、甲状腺癌、肾脏和肾盂癌症、子宫内膜癌、白血病以及胰腺癌。预计在接下来的二十年里新增癌症病例的数量将增加至二千二百万例。

[0007] 免疫系统在人癌症检测和防治中具有关键作用。大部分的转化细胞被免疫哨兵 (immune sentinel) 迅速地检测到并且通过克隆地表达的T细胞受体 (TCR) 活化抗原特异性T细胞而破坏。因此,癌症可以被视为一种免疫病症,即免疫系统无法产生必要的抗肿瘤反应以持久地抑制和消除疾病。为了更有效地对抗癌症,在过去数十年里开发出的某些免疫治疗干预专门地集中在增强T细胞免疫。这些治疗仅引起偶发的疾病缓解病例,并没有取得显著的总体成功。新近使用单克隆抗体靶向分子抑制T细胞活化的疗法,如CTLA-4或PD-1显示出较为显著的抗肿瘤作用;不过,这些治疗因全身免疫活化还会伴随显著的毒性。

[0008] 最近,已经在研究基于T细胞分离、修饰、扩增和再输注的过继性细胞免疫疗法策略并在早期临床试验中进行测试。由于T细胞具有选择性识别和强效的效应子机制,故这些细胞通常是选择用于癌症免疫疗法的效应细胞。这些治疗显示出一定的成功率,而且少量患者经历持久的缓解,突出显示了基于T细胞的免疫疗法的尚未实现的潜力。

[0009] 细胞溶解T细胞成功地识别肿瘤细胞相关抗原引起靶肿瘤溶解并且成为任何有效癌症免疫疗法的基础。肿瘤浸润性T细胞 (TIL) 表达特异性针对肿瘤相关抗原的TCR;不过,大量的TIL仅局限于少数人类癌症。工程改造的T细胞受体 (TCR) 和嵌合抗原受体 (CAR) 潜在地增加基于T细胞的免疫疗法在许多癌症和其它免疫病症中的适用性。

[0010] 此外,现有技术工程改造的T细胞还受到复杂免疫抑制肿瘤微环境的调控,所述免

疫抑制肿瘤微环境由癌细胞、炎性细胞、基质细胞和细胞因子组成。在这些组分中,癌细胞、炎性细胞和抑制性细胞因子调控T细胞表型和功能。总体来说,肿瘤微环境驱使T细胞终末分化成耗竭T细胞。

[0011] T细胞耗竭是在持续环境中以抑制性受体表达增加或抑制性受体引起的信号传导增加;效应细胞因子产生减少;以及持续存在和消除癌症的能力降低为标志的T细胞功能障碍状态。耗竭T细胞还以分级方式显示功能损失:在耗竭早期,IL-2产生减少且离体杀灭能力丧失;在中期,无法产生TNF- α ;以及在耗竭晚期,无法产生IFN- γ 和GzmB。大部分T细胞在肿瘤微环境中分化成耗竭T细胞并丧失消除癌症的能力,且最终被清除。

[0012] 程序性细胞死亡蛋白1(PD-1)在T细胞上表达并且通过结合至肿瘤微环境中存在的免疫抑制因子,例如PD-L1和PD-L2来介导免疫抑制作用。PD-L1和PD-L2的表达与一些人恶性病的预后相关。PD-L1/PD-1信号传导路径是T细胞耗竭的一个重要调控路径。PD-L1在癌细胞和基质细胞中大量表达,并且使用单克隆抗体阻断PD-L1/PD-1可增强T细胞的抗肿瘤功能。PD-L2也结合至PD-1并且负面地调控T细胞功能。

发明内容

[0013] 本公开大体上部分涉及包含裂解人PD-1基因中的靶位点的归巢核酸内切酶变体和megaTAL的组合物以及其使用方法。

[0014] 在各种实施例中,本公开部分地涵盖一种多肽,该多肽包含裂解人程序性细胞死亡蛋白1(PD-1)基因中的靶位点的归巢核酸内切酶(HE)变体。

[0015] 在特定实施例中,HE变体是LAGLIDADG归巢核酸内切酶(LHE)变体。

[0016] 在某些实施例中,所述多肽包含HE变体的生物活性片段。

[0017] 在一些实施例中,相较于相应野生型HE,所述生物活性片段缺乏1、2、3、4、5、6、7或8个N末端氨基酸。

[0018] 在另外的实施例中,相较于相应野生型HE,所述生物活性片段缺乏4个N末端氨基酸。

[0019] 在某些实施例中,相较于相应野生型HE,所述生物活性片段缺乏8个N末端氨基酸。

[0020] 在特定实施例中,相较于相应野生型HE,所述生物活性片段缺乏1、2、3、4或5个C末端氨基酸。

[0021] 在特定实施例中,其中相较于相应野生型HE,所述生物活性片段缺乏C末端氨基酸。

[0022] 在一些实施例中,相较于相应野生型HE,所述生物活性片段缺乏2个C末端氨基酸。

[0023] 在另外的实施例中,HE变体是选自由以下组成的组的LHE变体:I-AabMI、I-AaeMI、I-AniI、I-ApaMI、I-CapIII、I-CapIV、I-CkaMI、I-CpaMI、I-CpaMII、I-CpaMIII、I-CpaMIV、I-CpaMV、I-CpaV、I-CraMI、I-EjeMI、I-GpeMI、I-GpiI、I-GzeMI、I-GzeMII、I-GzeMIII、I-HjeMI、I-LtrII、I-LtrI、I-LtrWI、I-MpeMI、I-MveMI、I-NcrII、I-NcrI、I-NcrMI、I-OheMI、I-OnuI、I-OsoMI、I-OsoMII、I-OsoMIII、I-OsoMIV、I-PanMI、I-PanMII、I-PanMIII、I-PnoMI、I-ScuMI、I-SmaMI、I-SscMI以及I-Vdi141I。

[0024] 在特定实施例中,HE变体是选自由以下组成的组的LHE变体:I-CpaMI、I-HjeMI、I-OnuI、I-PanMI以及SmaMI。

[0025] 在另外的实施例中,HE变体是I-OnuI LHE变体。

[0026] 在另外的实施例中,HE变体在如SEQ ID NO:1-5中所述I-OnuI LHE氨基酸序列或其生物活性片段的DNA识别界面中选自由以下组成的组的氨基酸位置处包含一个或多个氨基酸取代:19、24、26、28、30、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76、77、78、80、82、168、180、182、184、186、188、189、190、191、192、193、195、197、199、201、203、223、225、227、229、231、232、234、236、238以及240。

[0027] 在特定实施例中,HE变体在SEQ ID NO:1-5中所述I-OnuI LHE氨基酸序列或其生物活性片段的DNA识别界面中选自由以下组成的组的氨基酸位置处包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:19、24、26、28、30、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76、77、78、80、82、168、180、182、184、186、188、189、190、191、192、193、195、197、199、201、203、223、225、227、229、231、232、234、236、238以及240。

[0028] 在特定实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:26、28、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76、78、80、138、143、159、168、178、180、182、184、186、188、189、190、191、192、193、197、199、201、203、207、223、224、225、227、229、232、236以及238。

[0029] 在某些实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40K、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、S72R、N75S、A76Y、S78K、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、S201M、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

[0030] 在一些实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40K、E42R、G44R、Q46E、T48D、V68K、A70Y、S72Q、N75S、A76Y、S78K、K80R、L138M、T143N、S159P、F168L、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、K225R、K229I、F232K、D236E以及V238E。

[0031] 在特定实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72R、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

[0032] 在另外的实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选

至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

[0033] 在另外的实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、K225R、F232K、D236E以及V238E。

[0034] 在特定实施例中,HE变体裂解PD-1外显子5靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201M、T203G、Y223R、K225R、F232K、D236E以及V238E。

[0035] 在特定实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:26、28、30、32、34、35、36、37、38、40、42、44、46、48、68、70、72、75、76、78、80、100、132、138、143、155、159、178、180、184、186、189、190、191、192、193、195、201、203、207、223、225、227、232、236、238以及240。

[0036] 在一些实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、V37G、G38R、S40H、E42R、G44S、Q46A、Q46T、T48V、T48M、V68I、V68S、A70T、A70Y、A70L、S72D、S72N、N75R、N75H、A76Y、S78R、S78T、K80R、K80C、K80E、K80V、T82F、T82Y、I100V、V132A、L138M、T143N、S155G、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0037] 在另外的实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46A、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80R、I100V、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0038] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至

少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、V37G、G38R、S40H、E42R、G44S、Q46T、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80C、I100V、V132A、L138M、T143N、S155G、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0039] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68I、A70T、S72N、N75H、A76Y、S78T、K80R、I100V、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0040] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70Y、S72N、N75H、A76Y、K80E、T82F、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0041] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70L、S72N、N75H、A76Y、K80V、T82Y、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0042] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37G、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70T、S72N、N75H、A76Y、K80V、T82Y、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0043] 在特定实施例中,HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:24、26、28、30、32、34、35、36、37、38、40、41、42、44、46、48、68、70、72、74、75、76、78、80、82、116、138、143、159、168、178、180、182、184、186、188、189、190、191、192、193、195、199、203、207、225、227、229、232、236以及238。

[0044] 在另外的实施例中,HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段的选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:S24C、L26Q、R28Y、R28H、R30S、N32V、N32L、K34N、K34R、S35N、S35T、S36R、V37S、V37T、G38R、G38K、

S40R、T41A、E42R、G44S、G44R、Q46E、Q46A、T48E、V68I、A70N、S72I、D74N、N75T、N75R、A76S、A76R、S78R、K80S、T82G、T82R、V116L、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203A、T203S、K207R、K225N、K225T、K227W、K227S、K229A、K229P、F232R、W234A、W234D、D236E以及V238R。

[0045] 在特定实施例中,HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:L26Q、R28Y、R30S、N32V、K34N、S35N、S36R、V37S、G38R、S40R、T41A、E42R、G44R、Q46A、T48E、A70N、S72I、N75T、A76S、S78R、K80S、T82G、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203A、K207R、K225N、K227W、K229A、F232R、W234A、D236E以及V238R。

[0046] 在一些实施例中,HE变体裂解PD-1外显子2靶位点并在SEQ ID NO:1-5中任一个或其生物活性片段中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个或全部:S24C、R28H、N32L、K34R、S35T、V37T、G38K、S40R、E42R、G44S、Q46E、T48E、V68I、A70N、S72I、D74N、N75R、A76R、S78R、K80S、T82R、V116L、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203S、K207R、K225T、K227S、K229P、F232R、W234D、D236E以及V238R。

[0047] 在另外的实施例中,HE变体包含与SEQ ID NO:6-14、60-63中任一个或其生物活性片段中所述的氨基酸序列至少80%、优选至少85%、更优选至少90%或甚至更优选至少95%一致的氨基酸序列。

[0048] 在特定实施例中,HE变体包含SEQ ID NO:6或其生物活性片段中所述的氨基酸序列。

[0049] 在特定实施例中,HE变体包含SEQ ID NO:7或其生物活性片段中所述的氨基酸序列。

[0050] 在一些实施例中,HE变体包含SEQ ID NO:8或其生物活性片段中所述的氨基酸序列。

[0051] 在特定实施例中,HE变体包含SEQ ID NO:9或其生物活性片段中所述的氨基酸序列。

[0052] 在特定实施例中,HE变体包含SEQ ID NO:10或其生物活性片段中所述的氨基酸序列。

[0053] 在特定实施例中,HE变体包含SEQ ID NO:11或其生物活性片段中所述的氨基酸序列。

[0054] 在一些实施例中,HE变体包含SEQ ID NO:12或其生物活性片段中所述的氨基酸序列。

[0055] 在特定实施例中,HE变体包含SEQ ID NO:13或其生物活性片段中所述的氨基酸序列。

[0056] 在一些实施例中,HE变体包含SEQ ID NO:14或其生物活性片段中所述的氨基酸序列。

- [0057] 在一些实施例中,HE变体包含SEQ ID NO:60或其生物活性片段中所述的氨基酸序列。
- [0058] 在一些实施例中,HE变体包含SEQ ID NO:61或其生物活性片段中所述的氨基酸序列。
- [0059] 在一些实施例中,HE变体包含SEQ ID NO:62或其生物活性片段中所述的氨基酸序列。
- [0060] 在一些实施例中,HE变体包含SEQ ID NO:63或其生物活性片段中所述的氨基酸序列。
- [0061] 在特定实施例中,所述多肽结合SEQ ID NO:25中所述的聚核苷酸序列。
- [0062] 在特定实施例中,所述多肽结合SEQ ID NO:30中所述的聚核苷酸序列。
- [0063] 在特定实施例中,所述多肽结合SEQ ID NO:35中所述的聚核苷酸序列。
- [0064] 在另外的实施例中,所述多肽还包含DNA结合结构域。
- [0065] 在一些实施例中,所述DNA结合结构域选自由以下组成的组:TALE DNA结合结构域和锌指DNA结合结构域。
- [0066] 在某些实施例中,所述TALE DNA结合结构域包含约9.5个TALE重复单元至约15.5个TALE重复单元。
- [0067] 在另外的实施例中,所述TALE DNA结合结构域结合PD-1基因中的聚核苷酸序列。
- [0068] 在特定实施例中,所述TALE DNA结合结构域结合SEQ ID NO:26中所述的聚核苷酸序列。
- [0069] 在某些实施例中,所述多肽结合并裂解SEQ ID NO:27中所述的聚核苷酸序列。
- [0070] 在特定实施例中,所述TALE DNA结合结构域结合SEQ ID NO:31中所述的聚核苷酸序列。
- [0071] 在某些实施例中,所述多肽结合并裂解SEQ ID NO:32中所述的聚核苷酸序列。
- [0072] 在特定实施例中,所述TALE DNA结合结构域结合SEQ ID NO:36中所述的聚核苷酸序列。
- [0073] 在某些实施例中,所述多肽结合并裂解SEQ ID NO:37中所述的聚核苷酸序列。
- [0074] 在某些实施例中,所述锌指DNA结合结构域包含2、3、4、5、6、7或8个锌指基元。
- [0075] 在另外的实施例中,所述多肽还包含肽连接子和末端加工酶或其生物活性片段。
- [0076] 在特定实施例中,所述多肽还包含病毒自裂解2A肽和末端加工酶或其生物活性片段。
- [0077] 在另外的实施例中,所述末端加工酶或其生物活性片段具有5'-3'核酸外切酶、5'-3'碱性核酸外切酶、3'-5'核酸外切酶、5'瓣状核酸内切酶、解螺旋酶或不依赖于模板的DNA聚合酶活性。
- [0078] 在特定实施例中,所述末端加工酶包含Trex2或其生物活性片段。
- [0079] 在某些实施例中,所述多肽包含SEQ ID NO:15-23和64中任一个或其生物活性片段中所述的氨基酸序列。
- [0080] 在另外的实施例中,所述多肽包含SEQ ID NO:15或其生物活性片段中所述的氨基酸序列。
- [0081] 在特定实施例中,所述多肽包含SEQ ID NO:16或其生物活性片段中所述的氨基酸

序列。

[0082] 在各种实施例中,所述多肽包含SEQ ID NO:17或其生物活性片段中所述的氨基酸序列。

[0083] 在特定实施例中,所述多肽包含SEQ ID NO:18或其生物活性片段中所述的氨基酸序列。

[0084] 在特定实施例中,所述多肽包含SEQ ID NO:19或其生物活性片段中所述的氨基酸序列。

[0085] 在一些实施例中,所述多肽包含SEQ ID NO:20或其生物活性片段中所述的氨基酸序列。

[0086] 在一些实施例中,所述多肽包含SEQ ID NO:21或其生物活性片段中所述的氨基酸序列。

[0087] 在另外的实施例中,所述多肽包含SEQ ID NO:22中所述的氨基酸序列,或其生物活性片段。

[0088] 在另外的实施例中,所述多肽包含SEQ ID NO:23或其生物活性片段中所述的氨基酸序列。

[0089] 在另外的实施例中,所述多肽包含SEQ ID NO:64或其生物活性片段中所述的氨基酸序列。

[0090] 在另外的实施例中,本文所涵盖的多肽包含SEQ ID NO:24或其生物活性片段中所述的氨基酸序列。

[0091] 在另外的实施例中,所述多肽裂解人PD-1基因中SEQ ID NO:25、27、30、32、35或37中所述的聚核苷酸序列。

[0092] 在各种实施例中,本公开部分地涵盖一种编码本文所涵盖的多肽的聚核苷酸。

[0093] 在特定实施例中,本公开部分地涵盖一种编码本文所涵盖的多肽的mRNA。

[0094] 在特定实施例中,mRNA包含SEQ ID NO:40-42和65-68中所述的序列。

[0095] 在各种实施例中,本公开部分地涵盖一种编码本文所涵盖的多肽的cDNA。

[0096] 在某些实施例中,本公开部分地涵盖一种载体,该载体包含编码本文所涵盖的多肽的聚核苷酸。

[0097] 在各种实施例中,本公开部分地涵盖一种细胞,该细胞包含本文所涵盖的多肽。

[0098] 在一些实施例中,本公开部分地涵盖一种细胞,该细胞包含编码本文所涵盖的多肽的聚核苷酸。

[0099] 在各种实施例中,本公开部分地涵盖一种细胞,该细胞包含本文所涵盖的载体。

[0100] 在另外的实施例中,本公开部分地涵盖一种细胞,该细胞包含一个或多个由本文所涵盖的多肽引入的基因组修饰。

[0101] 在特定实施例中,所述细胞包含编码以下一种或多种的聚核苷酸:免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体。

[0102] 在某些实施例中,所述聚核苷酸还包含RNA聚合酶II启动子,该RNA聚合酶II启动子可操作地连接至编码所述免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的聚核苷酸。

[0103] 在特定实施例中,所述RNA聚合酶II启动子选自由以下组成的组:短EF1 α 启动子、

长EF1 α 启动子、人ROSA 26基因座、泛素C (UBC) 启动子、磷酸甘油酸激酶-1 (PGK) 启动子、巨细胞病毒增强子/鸡 β -肌动蛋白 (CAG) 启动子、 β -肌动蛋白启动子和骨髓增生肉瘤病毒增强子、负控制区缺失、d1587rev引物结合位点取代 (MND) 的启动子。

[0104] 在一些实施例中,所述聚核苷酸还编码一个或多个自裂解病毒肽,所述自裂解病毒肽可操作地连接至所述免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体;散布于其间;和/或与其侧接。

[0105] 在一些实施例中,所述自裂解病毒肽是2A肽。

[0106] 在某些实施例中,所述聚核苷酸还包含异源聚腺苷酸化信号。

[0107] 在一些实施例中,所述免疫抑制信号阻断剂包含抵消免疫抑制因子作用的酶功能。

[0108] 在一些实施例中,所述免疫抑制信号阻断剂包含犬尿氨酸酶活性。

[0109] 在特定实施例中,所述免疫抑制信号阻断剂包含:结合免疫抑制因子的胞外结构域,任选地其中所述胞外结构域是抗体或其抗原结合片段;结合免疫抑制因子的胞外结构域和跨膜结构域;或结合免疫抑制因子的胞外结构域、跨膜结构域和经过修饰的胞内结构域,所述经过修饰的胞内结构域不能将免疫抑制信号转导至细胞中。

[0110] 在某些实施例中,所述免疫抑制信号阻断剂是显性负性TGF β RII受体。

[0111] 在一些实施例中,所述免疫效力增强剂选自由以下组成的组:双特异性T细胞接合分子 (BiTE)、免疫增强因子和翻转受体 (flip receptor)。

[0112] 在特定实施例中,所述免疫增强因子选自由以下组成的组:细胞因子、趋化因子、细胞毒素、细胞因子受体以及其变体。

[0113] 在特定实施例中,所述细胞因子受体选自由以下组成的组:IL-2受体、IL-7受体、IL-12受体、IL-15受体、IL-18受体以及IL-21受体。

[0114] 在一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的细胞因子受体的聚核苷酸,所述细胞因子受体选自由以下组成的组:IL-2受体、IL-7受体、IL-12受体、IL-15受体、IL-18受体以及IL-21受体。

[0115] 在另一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的IL-12细胞因子受体的聚核苷酸。

[0116] 在特定实施例中,所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。

[0117] 在一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的细胞因子的聚核苷酸,所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。

[0118] 在另一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的IL-12的聚核苷酸。

[0119] 在另外的实施例中,所述翻转受体包含PD-1胞外结构域和跨膜结构域;以及与所述PD-1跨膜结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或CD3 ζ 的胞内结构域。

[0120] 在某些实施例中,所述翻转受体包含PD-1胞外结构域;自CD3多肽、CD4、CD8 α 、CD28、CD134或CD137分离的跨膜结构域;以及与所述PD-1胞外结构域的C末端同框融合的来

自CD28、CD134、CD137、CD278和/或CD3 ζ 的胞内结构域。

[0121] 在特定实施例中,所述翻转受体包含PD-1胞外结构域;以及与所述PD-1胞外结构域的C末端同框融合的自CD3多肽、CD4、CD8 α 、CD28、CD134或CD137分离的跨膜结构域和胞内结构域。

[0122] 在另外的实施例中,工程改造的抗原受体选自由以下组成的组:工程改造的TCR、CAR、Daric或 ζ 因子(zetakine)。

[0123] 在特定实施例中,工程改造的受体未被整合至PD-1基因中。

[0124] 在某些实施例中,编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体中的一种或多种的聚核苷酸被整合至PD-1基因中。

[0125] 在另外的实施例中,包含编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体中的一种或多种的聚核苷酸的供体修复模板被整合至PD-1基因中由本文所涵盖的多肽引入的DNA双链断裂位点处。

[0126] 在特定实施例中,包含编码细胞因子的聚核苷酸的供体修复模板被整合至PD-1基因中由本文所涵盖的多肽引入的DNA双链断裂位点处,所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。在优选实施例中,所述细胞因子被整合至PD-1基因中与内源性PD-1启动子可操作地连接。

[0127] 在特定实施例中,包含编码IL-12细胞因子的聚核苷酸的供体修复模板被整合至PD-1基因中由本文所涵盖的多肽引入的DNA双链断裂位点处。在优选实施例中,IL-12细胞因子被整合至PD-1基因中与内源性PD-1启动子可操作地连接。

[0128] 在特定实施例中,包含编码细胞因子受体的聚核苷酸的供体修复模板被整合至PD-1基因中由本文所涵盖的多肽引入的DNA双链断裂位点处,所述细胞因子受体选自由以下组成的组:IL-2受体、IL-7受体、IL-12受体、IL-15受体、IL-18受体以及IL-21受体。在优选实施例中,所述细胞因子受体被整合至PD-1基因中与内源性PD-1启动子可操作地连接。

[0129] 在特定实施例中,包含编码IL-12细胞因子受体的聚核苷酸的供体修复模板被整合至PD-1基因中由本文所涵盖的多肽引入的DNA双链断裂位点处。在优选实施例中,IL-12细胞因子受体被整合至PD-1基因中与内源性PD-1启动子可操作地连接。

[0130] 在一些实施例中,细胞是造血细胞。

[0131] 在另外的实施例中,细胞是T细胞。

[0132] 在特定实施例中,细胞是CD3 $^{+}$ 、CD4 $^{+}$ 和/或CD8 $^{+}$ 细胞。

[0133] 在特定实施例中,细胞是免疫效应细胞。

[0134] 在另外的实施例中,细胞是细胞毒性T淋巴细胞(CTL)、肿瘤浸润性淋巴细胞(TIL)或辅助T细胞。

[0135] 在某些实施例中,细胞是自然杀手(NK)细胞或自然杀手T(NKT)细胞。

[0136] 在特定实施例中,细胞的来源是外周血液单核细胞、骨髓、淋巴结组织、脐带血、胸腺组织、来自感染部位的组织、腹水、胸腔积液、脾组织或肿瘤。

[0137] 在特定实施例中,本公开部分地涵盖包含一种或多种本文所涵盖的细胞的多种细胞。

[0138] 在各种实施例中,本公开部分地涵盖一种组合物,该组合物包含一种或多种本文所涵盖的细胞。

[0139] 在某些实施例中,本公开部分地涵盖一种组合物,该组合物包含一种或多种本文所涵盖的细胞和生理学上可接受的载体。

[0140] 在各种实施例中,本公开部分地涵盖一种编辑细胞中的人PD-1基因的方法,该方法包含:将编码本文所涵盖的多肽的聚核苷酸引入细胞中,其中所述多肽的表达在人PD-1基因中的靶位点处产生双链断裂。

[0141] 在一些实施例中,本公开部分地涵盖一种编辑细胞中的人PD-1基因的方法,该方法包含:将编码本文所涵盖的多肽的聚核苷酸引入细胞中,其中所述多肽的表达在人PD-1基因中的靶位点处产生双链断裂,其中所述断裂是通过非同源末端接合(non-homologous end joining,NHEJ)修复。

[0142] 在各种实施例中,本公开部分地涵盖一种编辑细胞中的人PD-1基因的方法,该方法包含:将编码本文所涵盖的多肽的聚核苷酸和供体修复模板引入细胞中,其中所述多肽的表达在人PD-1基因中的靶位点处产生双链断裂并且所述供体修复模板通过同源介导修复(homology directed repair,HDR)并入人PD-1基因中的双链断裂(DSB)位点处。

[0143] 在另外的实施例中,细胞是造血细胞。

[0144] 在特定实施例中,细胞是T细胞。

[0145] 在特定实施例中,细胞是CD3+、CD4+和/或CD8+细胞。

[0146] 在某些实施例中,细胞是免疫效应细胞。

[0147] 在一些实施例中,细胞是细胞毒性T淋巴细胞(CTL)、肿瘤浸润性淋巴细胞(TIL)或辅助T细胞。

[0148] 在特定实施例中,细胞是自然杀手(NK)细胞或自然杀手T(NKT)细胞。

[0149] 在某些实施例中,细胞的来源是外周血液单核细胞、骨髓、淋巴结组织、脐带血、胸腺组织、来自感染部位的组织、腹水、胸腔积液、脾组织或肿瘤。

[0150] 在特定实施例中,编码所述多肽的聚核苷酸是mRNA。

[0151] 在另外的实施例中,编码3'-5'核酸外切酶的聚核苷酸被引入所述细胞中。

[0152] 在一些实施例中,编码Trex2或其生物活性片段的聚核苷酸被引入所述细胞中。

[0153] 在另外的实施例中,供体修复模板编码PD-1基因或其部分,相较于野生型PD-1基因,所述PD-1基因或其部分包含一个或多个突变。

[0154] 在特定实施例中,供体修复模板编码以下一种或多种:免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体。

[0155] 在另外的实施例中,供体修复模板还包含可操作地连接至所述免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的RNA聚合酶II启动子。

[0156] 在另外的实施例中,RNA聚合酶II启动子选自由以下组成的组:短EF1 α 启动子、长EF1 α 启动子、人ROSA 26基因座、泛素C(UBC)启动子、磷酸甘油酸激酶-1(PGK)启动子、巨细胞病毒增强子/鸡 β -肌动蛋白(CAG)启动子、 β -肌动蛋白启动子和骨髓增生肉瘤病毒增强子、负控制区缺失、d1587rev引物结合位点取代(MND)的启动子。

[0157] 在某些实施例中,供体修复模板还编码一个或多个自裂解病毒肽,所述自裂解病毒肽可操作地连接至免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体;散布于其间;和/或与其侧接。

[0158] 在另外的实施例中,所述自裂解病毒肽是2A肽。

- [0159] 在一些实施例中,供体修复模板还包含异源聚腺苷酸化信号。
- [0160] 在某些实施例中,所述免疫抑制信号阻断剂包含抵消免疫抑制因子作用的酶功能。
- [0161] 在另外的实施例中,所述免疫抑制信号阻断剂包含犬尿氨酸酶活性。
- [0162] 在特定实施例中,所述免疫抑制信号阻断剂包含:结合免疫抑制因子的胞外结构域,任选地其中所述胞外结构域是抗体或其抗原结合片段;结合免疫抑制因子的胞外结构域和跨膜结构域;或结合免疫抑制因子的胞外结构域、跨膜结构域和经过修饰的胞内结构域,所述经过修饰的胞内结构域不能将免疫抑制信号转导至细胞中。
- [0163] 在另外的实施例中,所述免疫抑制信号阻断剂是显性负性TGFβRII受体。
- [0164] 在某些实施例中,所述免疫效力增强剂选自由以下组成的组:双特异性T细胞接合分子(BiTE)、免疫增强因子和翻转受体。
- [0165] 在另外的实施例中,所述免疫增强因子选自由以下组成的组:细胞因子、趋化因子、细胞毒素、细胞因子受体以及其变体。
- [0166] 在特定实施例中,所述免疫增强因子选自由以下组成的组:细胞因子、趋化因子、细胞毒素、细胞因子受体以及其变体。
- [0167] 在特定实施例中,所述细胞因子受体选自由以下组成的组:IL-2受体、IL-7受体、IL-12受体、IL-15受体、IL-18受体以及IL-21受体。
- [0168] 在一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的细胞因子受体的聚核苷酸,所述细胞因子受体选自由以下组成的组:IL-2受体、IL-7受体、IL-12受体、IL-15受体、IL-18受体以及IL-21受体。
- [0169] 在另一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的IL-12受体的聚核苷酸。
- [0170] 在特定实施例中,所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。
- [0171] 在一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的细胞因子的聚核苷酸,所述细胞因子选自由以下组成的组:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21。
- [0172] 在另一个优选实施例中,所述细胞包含编码可操作地连接至内源性PD-1启动子的IL-12的聚核苷酸。
- [0173] 在特定实施例中,所述翻转受体包含PD-1胞外结构域和跨膜结构域;以及与所述PD-1跨膜结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或CD3ζ的胞内结构域。
- [0174] 在另外的实施例中,所述翻转受体包含PD-1胞外结构域;自CD3多肽、CD4、CD8α、CD28、CD134或CD137分离的跨膜结构域;以及与所述PD-1胞外结构域的C末端同框融合的来自CD28、CD134、CD137、CD278和/或CD3ζ的胞内结构域。
- [0175] 在另外的实施例中,所述翻转受体包含PD-1胞外结构域;以及与所述PD-1胞外结构域的C末端同框融合的自CD3多肽、CD4、CD8α、CD28、CD134或CD137分离的跨膜结构域和胞内结构域。
- [0176] 在另外的实施例中,工程改造的抗原受体选自由以下组成的组:工程改造的TCR、

CAR、Daric或 ζ 因子。

[0177] 在另外的实施例中,供体修复模板包含与在DSB 5'端的人PD-1基因序列同源的5'同源臂和与在DSB 3'端的人PD-1基因序列同源的3'同源臂。

[0178] 在特定实施例中,5'同源臂和3'同源臂的长度独立地选自约100bp至约2500bp。

[0179] 在一些实施例中,5'同源臂和3'同源臂的长度独立地选自约600bp至约1500bp。

[0180] 在一些实施例中,5'同源臂是约1500bp并且3'同源臂是约1000bp。

[0181] 在某些实施例中,5'同源臂是约600bp并且3'同源臂是约600bp。

[0182] 在特定实施例中,使用了病毒载体将供体修复模板引入细胞中。

[0183] 在另外的实施例中,病毒载体是重组腺相关病毒载体(rAAV)或逆转录病毒。

[0184] 在另外的实施例中,rAAV具有来自AAV2的一个或多个ITR。

[0185] 在某些实施例中,rAAV具有选自以下组成的组的血清型:AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9以及AAV10。

[0186] 在另外的实施例中,rAAV具有AAV2或AAV6血清型。

[0187] 在一些实施例中,所述逆转录病毒是慢病毒。

[0188] 在特定实施例中,所述慢病毒是整合酶缺陷型慢病毒(IDLV)。

[0189] 在各种实施例中,本公开部分地涵盖一种治疗、预防或改善癌症、感染性疾病、自身免疫性疾病、炎性疾病和免疫缺陷或其相关病况的至少一种症状的方法,该方法包含向受试者施用有效量的本文所涵盖的组合物。

[0190] 在各种实施例中,本公开部分地涵盖一种治疗实体癌的方法,该方法包含向受试者施用有效量的本文所涵盖的组合物。

[0191] 在另外的实施例中,所述实体癌包含肝癌、胰腺癌、肺癌、乳癌、卵巢癌、前列腺癌、睾丸癌、膀胱癌、脑癌、肉瘤、头颈癌、骨癌、甲状腺癌、肾癌或皮肤癌。

[0192] 在各种实施例中,本公开部分地涵盖一种治疗血液恶性病的方法,该方法包含向受试者施用有效量的本文所涵盖的组合物。

[0193] 在另外的实施例中,所述血液恶性病是白血病、淋巴瘤或多发性骨髓瘤。

附图说明

[0194] 图1A显示的草图示出了PD-1的IgV结构域以及ITIM和ITSM基元的位置与外显子中megaTAL靶位点的位置的关系。

[0195] 图1B显示PD-1基因和由外显子5所编码的ITSM基元中靶位点的序列(SEQ ID NO: 106-108),突出显示了位于248位ITSM磷酸酪氨酸残基的密码子处的HE中心-4基元。

[0196] 图2显示如何对I-OnuI进行重编程,这是通过二轮分选,针对嵌合“半位点”(SEQ ID NO:109和110)工程改造NTD和CTD,随后使重编程的结构域(SEQ ID NO:111)融合并针对完整PD-1外显子5靶位点筛选以分离出完全重编程的HE实现。

[0197] 图3A显示在染色体报告子分析中针对活性初步筛选PD-1外显子5HE变体。

[0198] 图3B显示当通过平衡底物滴定法测量时,PD-1HE变体(PD-1.ITSM.ex5_RD1_CV3-08)具有适中的DNA结合亲和力特性。

[0199] 图4显示在分离出增强的变体之后,在染色体报告子分析中针对活性对PD-1HE进行二次筛选,所述增强的变体的分离是通过在较严格的裂解和亲和力条件下对PD-

1.ITSM.ex5_RD1_CV3-08变体的随机诱变文库进行基于展示的流式分选以分离活性改善的变体实现。

[0200] 图5显示对活化的原代人T细胞中PD-1外显子5 (SEQ ID NO:112) 进行亚硫酸氢盐测序分析以测定PD-1外显子5HE内存在的CpG基元 (SEQ ID NO:113) 的甲基化状态的结果。

[0201] 图6显示针对部分和完全甲基化的PD-1外显子5底物,对PD-1.ITSM.ex5_RD5_CV23MK HE变体进行DNA结合亲和力和裂解分析的结果。

[0202] 图7A显示如由TIDE分析所测量,将PD-1.ITSM.ex5_RD5_CV23MK megaTAL连同TREX2共递送至T细胞中以约60%速率编辑靶基因座。

[0203] 图7B显示当在Trex2存在下被PD-1.ITSM.ex5_RD5_CV23MK megaTAL裂解时PD-1ITSM共同基元处的插入缺失(indel)分布。

[0204] 图8显示对PD-1外显子5HE的中心-4特异性进行特征分析的结果。

[0205] 图9A显示的草图示出了PD-1的IgV、ITIM和ITSM结构域的位置与外显子1、2和5的位置关系。

[0206] 图9B显示PD-1基因以及外显子1 (SEQ ID NO:114-116)、外显子2 (SEQ ID NO:117-119) 和外显子5 (SEQ ID NO:106-108) 中靶位点的位置。

[0207] 图10显示在对以下各物的催化活性进行基于展示的流式分选之后,由染色体报告子分析得到的结果:靶向PD-1外显子1的最初重编程的I-OnuI (PD-1.ile3.外显子1_RD1_B1,上图);通过诱变并在较严格的催化条件下筛选以鉴别有助于靶裂解的突变,由此鉴别活性较高的变体而改进的PD-1.ile3.外显子1_RD1_B1 (PD-1.ile3.外显子1_RD2_B1H8和PD-1.ile3.外显子1_RD2_B1G2,中图);以及靶向PD-1外显子2的最初重编程的I-OnuI (PD-1.IgV.外显子2_RD1_G5,下图)。示出了由核酸酶在Trex2存在下(左图)和设计为megaTAL形式(右图)得到的结果。

[0208] 图11A显示当使用相应靶序列,通过平衡底物滴定法测量时PD-1.ile3.外显子1_RD2_B1H8和PD-1.IgV.外显子2_RD1_G5的DNA结合亲和力。

[0209] 图11B显示了针对在碱基对-8、碱基对-7和碱基对-6处有变化的64个DNA靶的特异性改进的核酸酶PD-1.ile3.外显子1 (RD2_B1G2、RD3_B1G2C4、RD3_B1G2C11、RD3_B1G2C5) 间DNA裂解活性的比较。热图表示自点图获得的非裂解:裂解中值的比率。

[0210] 图12显示对活化的原代人T细胞中的PD-1外显子1 (SEQ ID NO:120) 和外显子2 (SEQ ID NO:122) 进行亚硫酸氢盐测序分析的结果(左图),展示PD-1外显子1CpG基元 (SEQ ID NO:121) 保持未甲基化,而PD-1外显子2CpG基元 (SEQ ID NO:123) 被甲基化。图12还显示,PD-1.IgV.外显子2_RD1_G5高效地裂解未甲基化和甲基化的靶位点(分别是右上图和右下图)。

[0211] 图13A显示在电穿孔加入媒介剂、CFP、CCR5megaTAL、PD-1.ITSM.ex5_RD5_CV23MK megaTAL或PD-1.ile3.外显子1_RD2_B1H8megaTAL且随后用媒介剂或佛波醇12-肉豆蔻酸酯13-乙酸酯(PMA)/离子霉素(Ionomycin) (P/I) 刺激的CAR T细胞中的PD-1表面表达。

[0212] 图13B显示在电穿孔加入改进形式的PD-1.ile3.外显子1 (RD2_B1G2、RD3_B1G2C4、RD3_B1G2C11、RD3_B1G2C5) megaTAL的T细胞中的PD-1表面表达。

[0213] 图14显示在Trex2存在或不存在下同时递送PD-1.ITSM.ex5_RD5_CV23MK megaTAL和PD-1.ile3.外显子1_RD2_B1H8megaTAL使PD-1细胞表面表达明显减少。

[0214] 图15A显示电穿孔加入编码PD-1.ITSM.ex5_RD5_CV23MK或PD-1.ile3.外显子1_RD2_B1H8megaTAL的mRNA的抗BCMA CAR T细胞显示的PD-L1介导的细胞因子抑制作用相较于电穿孔加入媒介剂或者编码CFP或CCR5megaTAL的mRNA的抗BCMA CAR T细胞有所降低。

[0215] 图15B显示电穿孔加入编码PD-1.ITSM.ex5_RD5_CV23MK或PD-1.ile3.外显子1_RD2_B1H8megaTAL的mRNA的抗BCMA CAR T细胞显示的PD-L1介导的细胞因子抑制作用相较于电穿孔加入媒介剂或者编码无催化活性TCR α megaTAL的mRNA的抗BCMA CAR T细胞(未处理的实验)有所降低。将PD-1抗体添加至培养物中使得电穿孔加入媒介剂或者编码无催化活性TCR α megaTAL的mRNA的抗BCMA CAR T细胞中的细胞因子抑制作用消除。

[0216] 图16A显示通过同源重组将各种表达盒(GFP,上图;抗CD19CAR,中图;以及抗BCMA CAR,下图)引入PD-1外显子1中的策略。

[0217] 图16B显示了测定在用PD-1.ile3.外显子1_RD2_B1H8megaTAL和含有GFP、抗CD19CAR或抗BCMA CAR表达盒的rAAV靶向载体处理的T细胞中染色体整合的盒的长期表达的代表性流式细胞分析。

[0218] 图17显示了用于在外显子1中的PD-1起始密码子处引入mCherry报告基因的策略,以及有关电穿孔加入媒介剂或PD-1.ile3.外显子1_RD2_B1H8megaTAL并在PMA/离子霉素处理存在和不存在下用媒介剂或编码mCherry的rAAV靶向载体转导24小时的T细胞中mCherry表达的流式细胞分析。

[0219] 图18显示了用于在PD-1外显子5中的ITSM基元处引入MND启动子-BFP表达盒的策略,以及有关电穿孔加入媒介剂或PD-1.ITSM.ex5_RD5_CV23MK megaTAL并用媒介剂或含有pMND-BFP表达盒的rAAV靶向载体转导的T细胞中BFP表达的流式细胞分析。

[0220] 图19显示了用于在PD-1外显子5中的ITSM基元处引入MND启动子-BFP表达盒的策略,以及有关在电穿孔加入媒介剂或PD-1.ITSM.ex5_RD5_CV23MK megaTAL并用媒介剂或含有pMND-BFP表达盒的rAAV靶向载体转导的T细胞中PD-1和BFP表达的流式细胞分析。

[0221] 图20显示了用于在PD-1外显子5中的ITSM基元处引入MND启动子-BFP表达盒的策略,其中同源臂含有单核苷酸多态性(SNP)(上图)。下图显示了有关电穿孔加入PD-1.ITSM.ex5_RD5_CV23MK megaTAL并用含有pMND-BFP表达盒的rAAV靶向载体转导的T细胞的流式细胞分析,所述表达盒具有野生型同源臂、含SNP的5'同源臂或含SNP的3'同源臂。

[0222] 图21显示了用于在PD-1外显子5中的ITSM基元处引入MND启动子-PD-1.CD28翻转受体表达盒的策略,以及有关电穿孔加入媒介剂或PD-1.ITSM.ex5_RD5_CV23MK megaTAL并用媒介剂或含有pMND-PD-1.CD28翻转受体表达盒的rAAV靶向载体转导的T细胞中PD-1和BFP表达的流式细胞分析。

[0223] 图22显示了通过将PD-1.ITSM.ex5_RD5_CV23MK megaTAL和含有同源臂的rAAV盒递送至外显子5靶位点上游超过一千碱基处,在PD-1基因中产生大规模缺失的策略。

[0224] 图23显示了用于将细胞因子插入PD-1基因中的策略,其中表达处于内源性PD-1启动子的控制下(上图)。在PMA/离子霉素处理24小时之后,电穿孔加入PD-1megaTAL mRNA并用编码IL-12的rAAV转导的抗BCMA CAR T细胞显示PD-1表达相较于对照处理的细胞有所减少(下图)。

[0225] 图24A显示在PMA/离子霉素处理24小时之后,电穿孔加入PD-1megaTAL mRNA并用编码IL-12的rAAV转导的抗BCMA CAR T细胞显示IL-12产生相较于对照处理的细胞有所增

加(左图)。电穿孔加入PD-1megaTAL mRNA并用编码IL-12的rAAV转导且在K562-BCMA靶细胞存在下培养的抗BCMA CAR T细胞显示IL-12产生相较于对照处理的细胞有所增加(右图)。

[0226] 图24B显示由连续刺激分析得到的结果。将K52-BCMA细胞与抗BCMA-CAR T细胞混合,培养7天并且再添加K562-BCMA靶细胞以模拟重复刺激。再刺激之后,用重组IL-12处理或用PD-1megaTAL和IL-12HDR模板处理的抗BCMA-CAR T细胞显示IFN γ 产生和细胞毒性相较于对照处理的细胞有所增加。

[0227] 序列标识符的简要说明

[0228] SEQ ID NO:1是野生型I-0nuI LAGLIDADG归巢核酸内切酶(LHE)的氨基酸序列。

[0229] SEQ ID NO:2是野生型I-0nuI LHE的氨基酸序列。

[0230] SEQ ID NO:3是野生型I-0nuI LHE的生物活性片段的氨基酸序列。

[0231] SEQ ID NO:4是野生型I-0nuI LHE的生物活性片段的氨基酸序列。

[0232] SEQ ID NO:5是野生型I-0nuI LHE的生物活性片段的氨基酸序列。

[0233] SEQ ID NO:6是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0234] SEQ ID NO:7是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0235] SEQ ID NO:8是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0236] SEQ ID NO:9是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0237] SEQ ID NO:10是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0238] SEQ ID NO:11是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0239] SEQ ID NO:12是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0240] SEQ ID NO:13是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0241] SEQ ID NO:14是重编程用于结合并裂解人PD-1基因中的靶位点的I-0nuI LHE变体的氨基酸序列。

[0242] SEQ ID NO:15是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0243] SEQ ID NO:16是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0244] SEQ ID NO:17是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0245] SEQ ID NO:18是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0246] SEQ ID NO:19是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0247] SEQ ID NO:20是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0248] SEQ ID NO:21是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0249] SEQ ID NO:22是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0250] SEQ ID NO:23是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

- [0251] SEQ ID NO:24是编码鼠类Trex2的氨基酸序列。
- [0252] SEQ ID NO:25是人PD-1基因外显子5中的I-0nuI LHE变体靶位点。
- [0253] SEQ ID NO:26是人PD-1基因外显子5中的TALE DNA结合结构域靶位点。
- [0254] SEQ ID NO:27是人PD-1基因外显子5中的megaTAL靶位点。
- [0255] SEQ ID NO:28是人PD-1基因外显子5中的I-0nuI LHE变体N末端结构域靶位点。
- [0256] SEQ ID NO:29是人PD-1基因外显子5中的I-0nuI LHE变体C末端结构域靶位点。
- [0257] SEQ ID NO:30是人PD-1基因外显子1中的I-0nuI LHE变体靶位点。
- [0258] SEQ ID NO:31是人PD-1基因外显子1中的TALE DNA结合结构域靶位点。
- [0259] SEQ ID NO:32是人PD-1基因外显子1中的megaTAL靶位点。
- [0260] SEQ ID NO:33是人PD-1基因外显子1中的I-0nuI LHE变体N末端结构域靶位点。
- [0261] SEQ ID NO:34是人PD-1基因外显子1中的I-0nuI LHE变体C末端结构域靶位点。
- [0262] SEQ ID NO:35是人PD-1基因外显子2中的I-0nuI LHE变体靶位点。
- [0263] SEQ ID NO:36是人PD-1基因外显子2中的TALE DNA结合结构域靶位点。
- [0264] SEQ ID NO:37是人PD-1基因外显子2中的megaTAL靶位点。
- [0265] SEQ ID NO:38是人PD-1基因外显子2中的I-0nuI LHE变体N末端结构域靶位点。
- [0266] SEQ ID NO:39是人PD-1基因外显子2中的I-0nuI LHE变体C末端结构域靶位点。
- [0267] SEQ ID NO:40是编码PD-1megaTAL的mRNA。
- [0268] SEQ ID NO:41是编码PD-1megaTAL的mRNA。
- [0269] SEQ ID NO:42是编码PD-1megaTAL的mRNA。
- [0270] SEQ ID NO:43是编码鼠类Trex2蛋白质的mRNA。
- [0271] SEQ ID NO:44是编码含PD-1同源臂和pMND-BFP-SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列。
- [0272] SEQ ID NO:45是编码SEQ ID NO:44的5'同源臂的聚核苷酸。
- [0273] SEQ ID NO:46是编码SEQ ID NO:44的3'同源臂的聚核苷酸。
- [0274] SEQ ID NO:47是编码含PD-1同源臂和pMND-BFP-SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列。相对于野生型基因组序列,3'含有单核苷酸多态性(SNP)。
- [0275] SEQ ID NO:48是编码含PD-1同源臂和pMND-BFP-SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列。相对于野生型基因组序列,5'含有单核苷酸多态性(SNP)。
- [0276] SEQ ID NO:49是编码SEQ ID NO:48的5'同源臂的聚核苷酸。
- [0277] SEQ ID NO:50是编码SEQ ID NO:47的3'同源臂的聚核苷酸。
- [0278] SEQ ID NO:51是编码含pMND-PD-1.CD28转换受体.SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列。
- [0279] SEQ ID NO:52是编码含PD-1同源臂和pMND-BFP.SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列,其中5'同源臂在PD-1外显子5中ITSM基元上游约1.3kb处。
- [0280] SEQ ID NO:53是编码含PD-1同源臂和pMND-GFP-SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列。
- [0281] SEQ ID NO:54是编码SEQ ID NO:53的5'同源臂的聚核苷酸。
- [0282] SEQ ID NO:55是编码SEQ ID NO:53的3'同源臂的聚核苷酸。
- [0283] SEQ ID NO:56是编码含PD-1同源臂和pMND-抗CD19CAR-SV40polyA表达盒的rAAV

靶向载体的聚核苷酸序列。

[0284] SEQ ID NO:57是编码含PD-1同源臂和pMND-抗BCMA CAR-SV40polyA表达盒的rAAV靶向载体的聚核苷酸序列。

[0285] SEQ ID NO:58是编码含PD-1同源臂和编码mCherry的cDNA的rAAV靶向载体的聚核苷酸序列。

[0286] SEQ ID NO:60是重编程用于结合并裂解人PD-1基因中的靶位点的I-OnuI LHE变体的氨基酸序列。

[0287] SEQ ID NO:61是重编程用于结合并裂解人PD-1基因中的靶位点的I-OnuI LHE变体的氨基酸序列。

[0288] SEQ ID NO:62是重编程用于结合并裂解人PD-1基因中的靶位点的I-OnuI LHE变体的氨基酸序列。

[0289] SEQ ID NO:63是重编程用于结合并裂解人PD-1基因中的靶位点的I-OnuI LHE变体的氨基酸序列。

[0290] SEQ ID NO:64是结合并裂解人PD-1基因中的靶位点的megaTAL的氨基酸序列。

[0291] SEQ ID NO:65是编码PD-1megaTAL的mRNA。

[0292] SEQ ID NO:66是编码PD-1megaTAL的mRNA。

[0293] SEQ ID NO:67是编码PD-1megaTAL的mRNA。

[0294] SEQ ID NO:68是编码PD-1megaTAL的mRNA。

[0295] SEQ ID NO:69-79阐述各种连接子的氨基酸序列。

[0296] SEQ ID NO:80-104阐述蛋白酶裂解位点和自裂解多肽裂解位点的氨基酸序列。

[0297] 在前述序列中,如果存在X,则是指任何氨基酸或不存在某一氨基酸。

具体实施方式

[0298] A. 概述

[0299] 本公开大体上部分地涉及改良的基因组编辑组合物和其使用方法。不希望受任何特定理论束缚,各种实施例中所涵盖的基因组编辑组合物可以用于预防或治疗癌症、感染性疾病、自身免疫性疾病、炎症疾病和免疫缺陷或其相关病况,或改善其至少一种症状。困扰现有过继性细胞疗法的一个局限或问题是由肿瘤微环境介导的耗竭引起免疫效应细胞反应性不足。耗竭T细胞具有明显不同于原始T细胞、效应T细胞或记忆T细胞的特有分子特征。其被定义为细胞因子表达和效应功能降低的T细胞。PD-1是T细胞耗竭标记物;PD-1表达增加与T细胞增殖减少以及IL-2、TNF和IFN- γ 产生减少有关。

[0300] 在特定实施例中,通过消除、减少或阻断PD-1表达和/或信号传导使本文所涵盖的基因组编辑过的免疫效应细胞对耗竭更具抗性。

[0301] 各种实施例中所涵盖的基因组编辑组合物和方法包含被设计用于结合并裂解人程序性细胞死亡蛋白1 (PD-1) 基因中的靶位点的核酸酶变体。特定实施例中所涵盖的核酸酶变体可以用于在靶聚核苷酸序列中引入双链断裂,所述双链断裂可以通过在无聚核苷酸模板,例如供体修复模板存在下进行非同源末端接合 (NHEJ),或通过在供体修复模板存在下进行同源介导修复 (homology repair, HDR),即同源重组来修复。某些实施例中所涵盖的核酸酶变体也可以被设计为切口酶形式,所述切口酶产生单链DNA断裂,这些单链DNA断裂

可以使用细胞的碱基切除修复 (BER) 机构或在供体修复模板存在下进行的同源重组加以修复。NHEJ 是一种易错程序, 常常会形成破坏基因功能的小插入和缺失。同源重组需要同源 DNA 作为模板进行修复并且可以通过在靶位点处引入含有所希望序列的供体 DNA, 侧接于与侧接所述靶位点的区域同源的序列的任一侧上, 用于产生无限多种指定修饰。

[0302] 在一个优选实施例中, 本文所涵盖的基因组编辑组合物包含靶向人 PD-1 基因的归巢核酸内切酶变体或 megaTAL。

[0303] 在一个优选实施例中, 本文所涵盖的基因组编辑组合物包含归巢核酸内切酶变体或 megaTAL 和末端加工酶, 例如 Trex2。

[0304] 在各种实施例中, 涵盖基因组编辑过的细胞。基因组编辑过的细胞包含编辑过的 PD-1 基因, 其中编辑策略被设计用于减少或消除 PD-1 表达, 和/或选用 PD-1, 通过表达 PD-1 的细胞外配体结合结构域, 但破坏其转导免疫抑制细胞内信号的能力, 使其以显性负性方式作用。

[0305] 在各种实施例中, 在 T 细胞, 例如免疫效应细胞中 PD-1 基因的靶位点中产生 DNA 断裂, 并且在裂解的基因组序列的两端进行 NHEJ 可以产生具有极少或无 PD-1 表达的细胞, 并且优选地, 产生缺乏或基本上缺乏功能性 PD-1 表达和/或信号传导, 例如缺乏增进 T 细胞耗竭的能力的 T 细胞。不希望受任何特定理论束缚, 缺乏功能性 PD-1 表达的 T 细胞对免疫抑制和 T 细胞耗竭具有较高抗性, 并因此, 具有更持久且更有效的治疗效用。

[0306] 在各种其它实施例中, 提供了用于修复裂解的 PD-1 基因组序列的供体模板。利用所述模板的序列, 通过在 DNA 断裂位点处进行同源重组来修复 PD-1 基因。在特定实施例中, 修复模板包含破坏, 并且优选地基本上降低或消除功能性 PD-1 表达的聚核苷酸序列。

[0307] 在特定实施例中, PD-1 基因是利用编码对配体的亲和力增加的 PD-1 胞外结构域的模板修复。

[0308] 在特定实施例中, PD-1 基因是用编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的聚核苷酸修复。

[0309] 在特定实施例中, PD-1 基因是用编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的聚核苷酸修复并且被引入 PD-1 基因中, 选用内源性 PD-1 启动子以转录控制免疫效力增强剂、免疫抑制信号或工程改造的抗原受体的表达。

[0310] 在优选实施例中, 使用了本文所涵盖的基因组编辑组合物和方法编辑人 PD-1 基因。

[0311] 因此, 本文所涵盖的方法和组合物表示相较于现有过继性细胞疗法的重大改良。

[0312] 除非明确地作相反指示, 否则特定实施例的实践将采用在本领域技术范围内的化学、生物化学、有机化学、分子生物学、微生物学、重组 DNA 技术、遗传学、免疫学以及细胞生物学的常规方法, 其中有许多方法出于说明的目的而在下文有描述。在文献中全面解释了这些技术。参见例如 Sambrook 等人, 《分子克隆实验指南 (Molecular Cloning: A Laboratory Manual)》(第 3 版, 2001); Sambrook 等人, 《分子克隆实验指南》(第 2 版, 1989); Maniatis 等人, 《分子克隆实验指南》(1982); Ausubel 等人, 《现代分子生物学实验技术 (Current Protocols in Molecular Biology)》(John Wiley and Sons, 2008 年 7 月更新); 《精编分子生物学实验指南: 现代分子生物学实验技术方法概要 (Short Protocols in Molecular Biology: A of Methods from Current Protocols in Molecular Biology)》,

Greene Pub.Associates and Wiley-Interscience;Glover,《DNA克隆实用方法 (DNA Cloning:A Practical Approach)》,第I卷和第II卷 (IRL Press,牛津(Oxford),1985);Anand,《复杂基因组分析技术 (Techniques for the Analysis of Complex Genomes)》, (Academic Press,纽约(New York),1992);《转录与翻译 (Transcription and Translation)》(B.Hames和S.Higgins编,1984);Perbal,《分子克隆实践指南 (A Practical Guide to Molecular Cloning)》(1984);Harlow和Lane,《抗体 (Antibodies)》, (Cold Spring Harbor Laboratory Press,纽约冷泉港(Cold Spring Harbor,N.Y.),1998);《现代免疫学方法 (Current Protocols in Immunology)》,Q.E.Coligan,A.M.Kruisbeek,D.H.Margulies,E.M.Shevach和W.Strober编,1991);《免疫学年评 (Annual Review of Immunology)》;以及如《免疫学进展 (Advances in Immunology)》之类杂志中的专论。

[0313] B. 定义

[0314] 除非另外定义,否则本文所使用的所有技术和科学术语都具有与本发明所属领域的一般技术人员通常所了解相同的含义。尽管可以使用与本文所描述的方法和材料类似或相当的任何方法和材料来实践或测试特定实施例,但本文中描述组合物、方法和材料的优选实施例。出于本公开的目的,下文定义以下术语。

[0315] 本文使用的冠词“一(a/an)”是指一个(种)或超过一个(种)(即至少一个(种)或一个(种)或多个(种))该冠词的语法对象。作为实例,“一个元件”意思指一个元件或一个或多个元件。

[0316] 替代选择(例如“或”)的使用应理解为意思指替代选择中的一种、两种或其任何组合。

[0317] 术语“和/或”应理解为意思指替代选择中的一种或两种。

[0318] 如本文所使用,术语“约”或“大约”是指相对于参考数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度变化高达15%、10%、9%、8%、7%、6%、5%、4%、3%、2%或1%的数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度。在一个实施例中,术语“约”或“大约”是指数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度在大致参考数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 9\%$ 、 $\pm 8\%$ 、 $\pm 7\%$ 、 $\pm 6\%$ 、 $\pm 5\%$ 、 $\pm 4\%$ 、 $\pm 3\%$ 、 $\pm 2\%$ 或 $\pm 1\%$ 的范围内。

[0319] 在一个实施例中,范围,例如1至5、约1至5或约1至约5是指所述范围所涵盖的每个数字值。例如,在一个非限制性且仅说明性实施例中,范围“1至5”相当于表述1、2、3、4、5;或1.0、1.5、2.0、2.5、3.0、3.5、4.0、4.5或5.0;或1.0、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4.0、4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9或5.0。

[0320] 如本文所使用,术语“基本上”是指数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度是参考数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度的80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高百分比。在一个实施例中,“基本上相同”是指数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度产生的作用,例如生理作用与参考数量、水平、值、数目、频率、百分比、尺寸、大小、量、重量或长度大致相同。

[0321] 在本说明书通篇,除非上下文另外要求,否则“包含(comprise/comprises/

comprising)”一词应理解为暗指包括所述步骤或要素或一组步骤或要素,但不排除任何其它步骤或要素或任何其它组步骤或要素。“由……组成”意图包括且限于短语“由……组成”之后的事物。因此,短语“由……组成”指示所列要素是必不可少或必需的,并且不能存在其它要素。“基本上由……组成”意图包括短语后所列的任何要素,并且限于不干扰或影响关于所列要素的公开内容中所说明的活性或作用的其它要素。因此,短语“基本上由……组成”指示所列要素是必不可少的或必需的,而且不存在实质上影响所列要素的活性或作用的其它要素。

[0322] 本说明书通篇提到的“一个实施例(one embodiment/an embodiment)”、“一个特定实施例”、“一个相关实施例”、“某一实施例”、“一个另外的实施例”或“另一实施例”或其组合意味着,结合所述实施例描述的特定特征、结构或特性被包括在至少一个实施例中。因此,在本说明书通篇各处出现的前述短语未必都参考同一实施例。此外,特定特征、结构或特性可以通过任何适合方式组合于一个或多个实施例中。还应理解,在一个实施例中关于一个特征的正面叙述用作一个特定实施例中排除所述特征的基础。

[0323] 术语“离体”一般是指在生物体外部发生的活动,如在生物体外部的人造环境,优选在具有最少天然条件变化的人造环境中,在活组织中或对活组织进行的实验或测量。在特定实施例中,“离体”程序涉及从生物体获得活细胞或组织并在实验室设备中,通常在无菌条件下培养或调节,且取决于情况,典型地持续数小时或长达约24小时,但包括长达48或72小时。在某些实施例中,可收集这些组织或细胞并冷冻,且稍后解冻以进行离体处理。使用活细胞或组织进行的持续超过数天的组织培养实验或程序典型地被视为“体外”的,但在某些实施例中,这一术语可与离体互换使用。

[0324] 术语“体内”一般是指在生物体内部进行的活动。在一个实施例中,细胞基因组在体内经历工程改造、编辑或修饰。

[0325] “增强”或“促进”或“增加”或“扩增”或“加强”一般是指本文所涵盖的核酸酶变体、基因组编辑组合物或基因组编辑过的细胞产生、引发或引起比由媒剂或对照所引起的反应要强的反应(即,生理反应)的能力。从本领域中的理解和本文中的说明显而易见,可测量的反应可以包括催化活性、结合亲和力、持久性、细胞溶解活性增加和/或促炎性细胞因子增加等。“增加”或“增强”的量典型地是“统计显著”的量,并且可以包括达到媒剂或对照所产生的反应的1.1、1.2、1.5、2、3、4、5、6、7、8、9、10、15、20、30或更高倍数(例如500、1000倍)(包括在其间并高于1的所有整数和小数点,例如1.5、1.6、1.7、1.8等)的增加。

[0326] “减少”或“减小”或“减低”或“降低”或“减轻”或“除去”或“抑制”或“减弱”一般是指本文所涵盖的核酸酶变体、基因组编辑组合物或基因组编辑过的细胞产生、引发或引起比媒剂或对照所引起的反应要弱的反应(即,生理反应)的能力。可测量的反应可以包括脱靶结合亲和力、脱靶裂解特异性、T细胞耗竭等的减小。“减少”或“降低”的量典型地是“统计显著”的量,并且可以包括达到媒剂或对照所产生的反应(参考反应)1.1、1.2、1.5、2、3、4、5、6、7、8、9、10、15、20、30或更高倍数(例如500、1000倍)(包括在其间和高于1的所有整数和小数点,例如1.5、1.6、1.7、1.8)的减少。

[0327] “维持(maintain)”或“保持(preserve)”或“维持(maintenance)”或“无变化”或“无显著变化”或“无显著减少”一般是指本文所涵盖的核酸酶变体、基因组编辑组合物或基因组编辑过的细胞产生、引发或引起与媒剂或对照所引起的反应基本上类似或相当的生理

反应(即,下游效应)的能力。相当的反应是与参考反应并无显著不同或可测量不同的反应。

[0328] 如本文所使用,术语“特异性结合亲和力”或“特异性结合(specifically binds)”或“特异性结合(specifically bound)”或“特异性结合(specific binding)”或“特异性靶向”描述一个分子与另一个分子结合,例如多肽的DNA结合结构域与DNA结合的结合亲和力高于背景结合。如果结合结构域以例如大于或等于约 10^5M^{-1} 的亲和力或 K_a (即,特定结合相互作用的平衡缔合常数,以 $1/\text{M}$ 为单位)与靶位点结合或缔合,则该结合结构域“特异性结合”至靶位点。在某些实施例中,结合结构域与靶位点结合的 K_a 大于或等于约 10^6M^{-1} 、 10^7M^{-1} 、 10^8M^{-1} 、 10^9M^{-1} 、 10^{10}M^{-1} 、 10^{11}M^{-1} 、 10^{12}M^{-1} 或 10^{13}M^{-1} 。“高亲和力”结合结构域是指 K_a 是至少 10^7M^{-1} 、至少 10^8M^{-1} 、至少 10^9M^{-1} 、至少 10^{10}M^{-1} 、至少 10^{11}M^{-1} 、至少 10^{12}M^{-1} 、至少 10^{13}M^{-1} 或更高的结合结构域。

[0329] 或者,亲和力可以定义为特定结合相互作用的平衡解离常数(K_d)且以 M 为单位(例如 10^{-5}M 至 10^{-13}M ,或更低)。特定实施例中涵盖的包含一个或多个DNA结合结构域的核酸酶变体针对DNA靶位点的亲和力可以使用常规技术,例如酵母细胞表面展示,或通过结合缔合或使用标记过的配体进行的置换分析容易地测定。

[0330] 在一个实施例中,特异性结合的亲和力比背景结合高约2倍、比背景结合高约5倍、比背景结合高约10倍、比背景结合高约20倍、比背景结合高约50倍、比背景结合高约100倍、或比背景结合高约1000倍或更高。

[0331] 术语“选择性结合(selectively binds)”或“选择性结合(selectively bound)”或“选择性结合(selectively binding)”或“选择性靶向”描述在多个脱靶分子存在下,一个分子优先结合至靶分子(在靶结合(on-target binding))。在特定实施例中,HE或megaTAL选择性结合在靶DNA结合位点的频率比HE或megaTAL结合脱靶DNA靶结合位点的频率高约5、10、15、20、25、50、100或1000倍。

[0332] “在靶”是指靶位点序列。

[0333] “脱靶”是指与靶位点序列类似但不一致的序列。

[0334] “靶位点”或“靶序列”是在存在足以实现结合和/或裂解的条件下,界定核酸中结合分子将结合和/或裂解的一部分的染色体或染色体外核酸序列。当提到仅涉及靶位点或靶序列一条链的聚核苷酸序列或SEQ ID NO.时,应理解,核酸酶变体所结合和/或裂解的靶位点或靶序列是双链的并且包含参考序列和其补体。在一个优选实施例中,靶位点是人PD-1基因中的序列。

[0335] “重组”是指二个聚核苷酸之间遗传信息交换的程序,包括但不限于通过非同源末端接合(NHEJ)和同源重组进行的供体捕获。出于本公开的目的,“同源重组(HR)”是指在例如通过同源介导修复(HDR)机制修复细胞中的双链断裂期间发生的此类交换的特殊形式。这一程序需要核苷酸序列同源性,使用“供体”分子作为模板来修复“靶”分子(即,经历双链断裂的分子),并且不同地称为“不产生交换的基因转换(non-crossover gene conversion)”或“短链基因转化(short tract gene conversion)”,因为其导致遗传信息从供体转移至靶。不希望受任何特定理论束缚,此类转移可涉及在断裂的靶与供体之间形成的异双螺旋DNA的错配校正;和/或“合成依赖性链退火(synthesis-dependent strand annealing)”,其中使用供体再合成遗传信息,所述遗传信息将变成靶的一部分;和/或相关程序。此类特殊HR通常引起靶分子序列的改变,使得供体聚核苷酸序列的一部分或全部被

并入靶聚核苷酸中。

[0336] “NHEJ”或“非同源末端接合”是指在没有供体修复模板或同源序列存在下双链断裂的消除。NHEJ可以在断裂位点处引起插入和缺失。NHEJ是由若干子路径介导的,这些子路径各自具有不同的突变结果。经典NHEJ路径(cNHEJ)需要KU/DNA-PKcs/Lig4/XRCC4复合物,通过最低限度的加工,将两端重新连接在一起,并且通常引起断裂的精确修复。替代性NHEJ路径(altnHEJ)在消除dsDNA断裂方面也具有活性,但这些路径明显更易于诱变且通常无法精确修复以插入和缺失为标志的断裂。尽管不希望受任何特定理论束缚,但预期利用末端加工酶,如核酸外切酶,例如Trex2修饰dsDNA断裂可能增加不精确修复的可能性。

[0337] “裂解”是指DNA分子共价主链的断裂。裂解可以通过多种方法引发,包括但不限于磷酸二酯键的酶或化学水解。单链裂解和双链裂解都是可能的。双链裂解可作为二个不同的单链裂解事件的结果发生。DNA裂解可导致产生钝端或交错端。在某些实施例中,将本文所涵盖的多肽和核酸酶变体,例如归巢核酸内切酶变体、megaTAL等用于靶向双链DNA裂解。核酸内切酶裂解识别位点可以在任一DNA链上。

[0338] “外源”分子是通常不存在于细胞中,而是通过一种或多种基因、生物化学或其它方法引入细胞中的分子。示例性外源分子包括但不限于小有机分子、蛋白质、核酸、碳水化合物、脂质、糖蛋白、脂蛋白、多糖、前述分子的任何经过修饰的衍生物或包含一个或多个前述分子的任何复合物。用于将外源分子引入细胞中的方法是本领域技术人员已知的并且包括但不限于脂质介导的转移(即,脂质体,包括中性和阳离子性脂质)、电穿孔、直接注射、细胞融合、粒子轰击、生物聚合物纳米粒子、磷酸钙共沉淀、DEAE-葡聚糖介导的转移以及病毒载体介导的转移。

[0339] “内源”分子是在特定环境条件下通常存在于处于特定发育阶段的特定细胞中的分子。另外的内源分子可以包括蛋白质。

[0340] “基因”是指编码基因产物的DNA区域,以及调控基因产物的产生的所有DNA区域,无论这些调控序列是否邻近于编码序列和/或转录的序列。基因包括但不限于启动子序列、增强子、沉默子、隔离子、边界元件、终止子、聚腺苷酸化序列、转录后反应元件、翻译调控序列如核糖体结合位点和内部核糖体进入位点、复制起点、基质附着位点以及基因座控制区。

[0341] “基因表达”是指基因中所含信息转化成基因产物。基因产物可以是基因(例如mRNA、tRNA、rRNA、反义RNA、核酶、结构RNA或任何其它类型的RNA)的直接转录产物或由mRNA翻译产生的蛋白质。基因产物还包括通过如戴帽、聚腺苷酸化、甲基化和编辑之类方法修饰的RNA,以及通过例如甲基化、乙酰化、磷酸化、泛素化、ADP-核糖基化、肉豆蔻酸化和糖基化修饰的蛋白质。

[0342] 如本文所使用,术语“基因工程改造”或“基因修饰”是指将呈DNA或RNA形式的额外遗传物质染色体或染色体外添加至细胞中的总遗传物质中。基因修饰可以被靶向或不靶向细胞基因组中的特定位点。在一个实施例中,基因修饰具有位点特异性。在一个实施例中,基因修饰不具有位点特异性。

[0343] 如本文所使用,术语“基因组编辑”是指在细胞基因组中靶位点处遗传物质的取代、缺失和/或引入,其恢复、校正、破坏和/或改变基因或基因产物的表达和/或功能。特定实施例中涵盖的基因组编辑包含将一种或多种核酸酶变体引入细胞中以任选地在供体修复模板存在下,在细胞基因组中的靶位点处或附近产生DNA损伤。

[0344] 如本文所使用,术语“基因疗法”是指将额外遗传物质引入细胞中的总遗传物质中以恢复、校正或改变基因或基因产物的表达或达到表达治疗性多肽的目的。在特定实施例中,通过基因组编辑将遗传物质引入细胞基因组中以恢复、校正、破坏或改变基因或基因产物的表达或达到表达治疗性多肽的目的被认为是基因疗法。

[0345] “免疫病症”是指引起免疫系统反应的疾病。在特定实施例中,术语“免疫病症”是指癌症、移植物抗宿主疾病、自身免疫性疾病或免疫缺陷。在一个实施例中,免疫病症包含感染性疾病。

[0346] 如本文所使用,术语“癌症”大体上涉及异常细胞不受控制地分裂且可侵袭邻近组织的一类疾病或病况。

[0347] 如本文所使用,术语“恶性”是指一组肿瘤细胞展示以下一种或多种特性的癌症:不受控制的生长(即,分裂超过正常限值)、侵袭(即,入侵并破坏邻近组织)和转移(即,通过淋巴或血液扩散至身体其它位置)。

[0348] 如本文所使用,术语“转移”是指癌症从身体的一个部分扩散至另一部分。由扩散的细胞形成的肿瘤称为“转移瘤”或“转移”。转移瘤含有的细胞与原始肿瘤(原发瘤)中的细胞相似。

[0349] 如本文所使用,术语“良性”或“非恶性”是指肿瘤可生长至较大但不会扩散至身体的其它部分。良性肿瘤是自限性的且典型地不会侵袭或转移。

[0350] “癌细胞”或“肿瘤细胞”是指癌变生长或组织的各个细胞。肿瘤一般是指细胞异常生长形成的肿胀或病变,它可以是良性、癌前或恶性的。大多数癌症形成肿瘤,但有一些,例如白血病,未必形成肿瘤。对于形成肿瘤的癌症,术语癌症(细胞)和肿瘤(细胞)可互换使用。个体中肿瘤的量是“肿瘤负荷”,它可以通过肿瘤的数目、体积或重量度量。

[0351] “移植物抗宿主疾病”或“GVHD”是指在细胞、组织或实体器官移植之后发生的并发症。GVHD可在干细胞或骨髓移植之后发生,在此情况下移植的供体细胞攻击移植物接受者的身体。人急性GVHD是在移植后约60天内发生并且在细胞溶解性淋巴细胞的作用下引起皮肤、肝和肠道的损伤。慢性GVHD是在稍后发生并且是主要影响皮肤,导致B细胞多克隆活化以及Ig和自身抗体过量产生的一种全身自身免疫性疾病。实体器官移植的移植物抗宿主疾病(SOT-GVHD)是以二种形式发生。较常见类型是抗体介导的,其中来自有O型血液的供体的抗体攻击有A型、B型或AB型血液的接受者体内的红细胞,导致轻度短暂性、溶血性贫血。SOT-GVHD的第二种形式是伴随高死亡率的一种细胞类型,其中供体来源的T细胞针对免疫不同的宿主组织,最通常在皮肤、肝、胃肠道和骨髓中产生免疫攻击,在这些器官中引起并发症。

[0352] “移植物抗白血病”或“GVL”是指供体的移植组织,如骨髓或外周血液中存在的免疫细胞针对个人的白血病细胞的免疫反应。

[0353] “自身免疫性疾病”是指身体针对其自身组织的一些成分产生免疫原性(即,免疫系统)反应的一种疾病。换句话说,免疫系统丧失其将身体内的某一组织或系统识别为“自身的”能力,并且靶向并攻击该组织或系统,就好像它是外来的。自身免疫性疾病的示例性实例包括但不限于:关节炎、炎性肠病、桥本氏甲状腺炎(Hashimoto's thyroiditis)、格雷夫斯氏病(Grave's disease)、狼疮、多发性硬化、风湿性关节炎、溶血性贫血、自身免疫性甲状腺炎、全身性红斑狼疮、乳糜泻、克罗恩氏病(Crohn's disease)、结肠炎、糖尿病、硬皮

病、牛皮癣等。

[0354] “免疫缺陷”意思指患者免疫系统因疾病或因施用化学品而受损的状态。这一病况使系统缺乏防御外来物质所需的血细胞数目和类型。免疫缺陷病况或疾病是本领域中已知的并且包括例如获得性免疫缺陷综合症 (AIDS)、重度联合免疫缺陷疾病 (SCID)、选择性IgA缺乏症、常见变异型免疫缺陷病、X连锁无丙种球蛋白血症、慢性肉芽肿病、高IgM综合症、韦斯考特-奥德里奇综合症 (Wiskott-Aldrich Syndrome, WAS) 以及糖尿病。

[0355] “感染性疾病”是指可以在人与人之间或在生物体与生物体之间传播的一种疾病，并且是由微生物或病毒因子引起 (例如普通感冒)。感染性疾病是本领域中已知的并且包括例如肝炎、性传播疾病 (例如衣原体病、淋病)、肺结核、HIV/AIDS、白喉、乙型肝炎、丙型肝炎、霍乱以及流感。

[0356] 如本文所使用，术语“个体”和“受试者”通常可互换使用并且是指展现可以用本文别处所涵盖的核酸酶变体、基因组编辑组合物、基因疗法载体、基因组编辑载体、基因组编辑过的细胞以及方法治疗的免疫病症的症状的任何动物。适合受试者 (例如患者) 包括实验动物 (如小鼠、大鼠、兔或天竺鼠)、农畜和家畜或宠物 (如猫或狗)。非人类灵长类动物并且优选地人类受试者也包括在内。典型的受试者包括患有免疫病症、被诊断患有免疫病症或有患免疫病症的风险的人类患者。

[0357] 如本文所使用，术语“患者”是指被诊断患有可以用本文别处所涵盖的核酸酶变体、基因组编辑组合物、基因疗法载体、基因组编辑载体、基因组编辑过的细胞以及方法治疗的免疫病症的受试者。

[0358] 如本文所使用，“治疗 (treatment/treating)”包括针对疾病或病理病况的症状或病理的任何有益或所希望的作用，并且可以包括所治疗疾病或病况，例如癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病和免疫缺陷的一种或多种可测量标记物的甚至最低程度的降低。治疗可以任选地涉及疾病或病况进展的延迟。“治疗”未必指示疾病或病况或其相关症状的完全根除或治愈。

[0359] 如本文所使用，“预防 (prevent)”和类似词语，如“预防 (prevention/prevented/preventing)”等指示用于预防、抑制疾病或病况，例如癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病和免疫缺陷的发生或复发，或降低其发生或复发的可能性的方法。这些词语还指延迟疾病或病况的发作或复发，或延迟疾病或病况症状的发生或复发。如本文所使用，“预防”和类似词语还包括在疾病或病况发作或复发之前，降低疾病或病况的强度、影响、症状和/或负荷。

[0360] 如本文中所使用，短语“改善……的至少一种症状”是指减少所治疗受试者所患疾病或病况，例如癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病和免疫缺陷的一种或多种症状。在特定实施例中，所治疗的疾病或病况是癌症，其中被改善的一种或多种症状包括但不限于虚弱、疲劳、呼吸急促、容易瘀伤和出血、频繁感染、淋巴结肿大、腹胀或腹痛 (由腹部器官增大引起)、骨头或关节疼痛、骨折、计划外的体重减轻、食欲不振、盗汗、持续性低热和排尿减少 (由肾功能减退引起)。

[0361] 如本文所使用，术语“量”是指足以实现有益或希望的预防性或治疗性结果，包括临床结果的核酸酶变体、基因组编辑组合物或基因组编辑过的细胞的“有效的量”或“有效量”。

[0362] “预防有效量”是指足以实现所希望的预防结果的核酸酶变体、基因组编辑组合物或基因组编辑过的细胞的量。典型地但非必需地，由于预防剂量是在患病之前或在疾病早期用于受试者，故预防有效量小于治疗有效量。

[0363] 核酸酶变体、基因组编辑组合物或基因组编辑过的细胞的“治疗有效量”可以根据如疾病状态、个体的年龄、性别和体重以及在个体体内引起所希望的反应的能力等因素而变化。治疗有效量还是治疗有益作用超过任何有毒或有害作用的量。术语“治疗有效量”包括有效地“治疗”受试者（例如患者）的量。当指示治疗量时，打算施用的特定实施例中所涵盖的组合物的确切量可由医师根据本说明书并且考虑个体在年龄、体重、肿瘤大小、感染或转移的程度以及患者（受试者）的状况确定。

[0364] C. 核酸酶变体

[0365] 本文在特定实施例中所涵盖的核酸酶变体适于基因组编辑PD-1基因中的靶位点并且包含一个或多个DNA结合结构域和一个或多个DNA裂解结构域（例如一种或多种核酸内切酶和/或一个或多个核酸外切酶结构域），以及任选地一个或多个本文所涵盖的连接子。术语“重编程的核酸酶”、“工程改造的核酸酶”或“核酸酶变体”可互换使用并且是指包含一个或多个DNA结合结构域和一个或多个DNA裂解结构域的核酸酶，其中所述核酸酶自亲本或天然存在的核酸酶设计和/或修饰成结合并裂解PD-1基因中的双链DNA靶序列。

[0366] 在特定实施例中，核酸酶变体结合并裂解PD-1基因外显子5中的靶序列，优选地PD-1基因外显子5中的SEQ ID NO:25，并且更优选地PD-1基因外显子5中SEQ ID NO:25中的序列“ATAC”。

[0367] 在特定实施例中，核酸酶变体结合并裂解PD-1基因外显子1中的靶序列，优选地PD-1基因外显子1中的SEQ ID NO:30，并且更优选地PD-1基因外显子1中SEQ ID NO:30中的序列“ATCC”。

[0368] 在特定实施例中，核酸酶变体结合并裂解PD-1基因外显子2中的靶序列，优选地PD-1基因外显子2中的SEQ ID NO:35，并且更优选地PD-1基因外显子2中SEQ ID NO:35中的序列“ACTT”。

[0369] 核酸酶变体可以自天然存在的核酸酶或先前的核酸酶变体设计和/或修饰得到。特定实施例中所涵盖的核酸酶变体还可包含一个或多个另外的功能结构域，例如末端加工酶中展现5’-3’核酸外切酶、5’-3’碱性核酸外切酶、3’-5’核酸外切酶（例如Trex2）、5’瓣状核酸内切酶、解螺旋酶、模板依赖性DNA聚合酶或不依赖模板的DNA聚合酶活性的末端加工酶结构域。

[0370] 结合并裂解PD-1基因中的靶序列的核酸酶变体的示例性实例包括但不限于归巢核酸内切酶（大范围核酸酶）变体和megaTAL。

[0371] 1. 归巢核酸内切酶（大范围核酸酶）变体

[0372] 在各种实施例中，归巢核酸内切酶或大范围核酸酶被重编程以在PD-1基因中的靶位点中引入双链断裂（DSB）。在特定实施例中，归巢核酸内切酶变体在PD-1基因外显子5中，优选地在PD-1基因外显子5中的SEQ ID NO:25处，并且更优选地在PD-1基因外显子5中SEQ ID NO:25中的序列“ATAC”处引入双链断裂。在特定实施例中，归巢核酸内切酶变体在PD-1基因外显子1中，优选地在PD-1基因外显子1中的SEQ ID NO:30处，并且更优选地在PD-1基因外显子1中SEQ ID NO:30的序列“ATCC”处引入双链断裂。在特定实施例中，归巢核酸内切

酶变体在PD-1基因外显子2中,优选地在PD-1基因外显子2中的SEQ ID NO:35处,并且更优选地在PD-1基因外显子2中的SEQ ID NO:35中的序列“ACTT”处引入双链断裂。

[0373] “归巢核酸内切酶”与“大范围核酸酶”可互换使用并且是指识别具有12-45个碱基对的裂解位点的天然存在的归巢核酸内切酶,并且这些归巢核酸内切酶通常基于序列和结构基元分成五个家族:LAGLIDADG、GIY-YIG、HNH、His-Cys盒以及PD- (D/E) XK。

[0374] “参考归巢核酸内切酶”或“参考大范围核酸酶”是指野生型归巢核酸内切酶或自然界中所发现的归巢核酸内切酶。在一个实施例中,“参考归巢核酸内切酶”是指被修饰成增加基础活性的野生型归巢核酸内切酶。

[0375] “工程改造的归巢核酸内切酶”、“重编程的归巢核酸内切酶”、“归巢核酸内切酶变体”、“工程改造的大范围核酸酶”、“重编程的大范围核酸酶”或“大范围核酸酶变体”是指包含一个或多个DNA结合结构域和一个或多个DNA裂解结构域的归巢核酸内切酶,其中所述归巢核酸内切酶自亲本或天然存在的归巢核酸内切酶设计和/或修饰成结合并裂解PD-1基因中的DNA靶序列。归巢核酸内切酶变体可以自天然存在的归巢核酸内切酶或自另一归巢核酸内切酶变体设计和/或修饰得到。特定实施例中所涵盖的归巢核酸内切酶变体还可包含一个或多个另外的功能结构域,例如末端加工酶中展现5'-3'核酸外切酶、5'-3'碱性核酸外切酶、3'-5'核酸外切酶(例如Trex2)、5'瓣状核酸内切酶、解螺旋酶、模板依赖性DNA聚合酶或不依赖模板的DNA聚合酶活性的末端加工酶结构域。

[0376] 归巢核酸内切酶(HE)变体不存在于自然界中并且可以通过重组DNA技术或通过随机诱变获得。HE变体可以通过使天然存在的HE或HE变体中的一个或多个氨基酸改变,例如突变、取代、添加或缺失一个或多个氨基酸获得。在特定实施例中,HE变体包含DNA识别界面的一个或多个氨基酸变化。

[0377] 特定实施例中所涵盖的HE变体还可包含一个或多个另外的功能结构域,例如末端加工酶中展现5'-3'核酸外切酶、5'-3'碱性核酸外切酶、3'-5'核酸外切酶(例如Trex2)、5'瓣状核酸内切酶、解螺旋酶、模板依赖性DNA聚合酶或不依赖模板的DNA聚合酶活性的末端加工酶结构域。在特定实施例中,利用展现5'-3'核酸外切酶、5'-3'碱性核酸外切酶、3'-5'核酸外切酶(例如Trex2)、5'瓣状核酸内切酶、解螺旋酶、模板依赖性DNA聚合酶或不依赖模板的DNA聚合酶活性的末端加工酶将HE变体引入T细胞中。HE变体和3'加工酶可以分开地引入,例如利用不同载体或独立mRNA引入;或一起引入,例如以融合蛋白形式,或以通过病毒自裂解肽或IRES元件隔开的多顺反子构建体引入。

[0378] “DNA识别界面”是指与核酸靶碱基以及邻近残基相互作用的HE氨基酸残基。对于每个HE,DNA识别界面包含侧链与侧链和侧链与DNA接触构成的广泛网状结构,所述接触中大部分必然是识别特定核酸靶序列特有的。因此,对应于特定核酸序列的DNA识别界面的氨基酸序列变化极大并且是任何天然HE或HE变体的特征。作为非限制性实例,特定实施例中所涵盖的HE变体可以通过构造HE变体文库得到,在所述HE变体文库中,位于天然HE(或先前产生的HE变体)的DNA识别界面中的一个或多个氨基酸残基是变化的。可以使用裂解分析,针对每个预测的PD-1靶位点的靶裂解活性筛选这些文库(参见例如Jarjour等人,2009.《核酸研究(Nuc.Acids Res.)》37(20):6871-6880)。

[0379] LAGLIDADG归巢核酸内切酶(LHE)是研究最充分的归巢核酸内切酶家族,主要在古细菌中以及绿藻和真菌中的细胞器DNA中编码,并且展示最高的总体DNA识别特异性。LHE包

含每条蛋白质链一个或二个LAGLIDADG催化基元并且分别以同二聚体或单链单体形式起作用。关于LAGLIDADG蛋白质的结构研究鉴别出高度保守的核心结构(Stoddard 2005),以 $\alpha\beta\alpha$ 折叠为特征,并且LAGLIDADG基元属于这一折叠的第一个螺旋。LHE的高效特异性裂解表示得到新颖、高度特异性核酸内切酶的蛋白质骨架。然而,将LHE工程改造成结合并裂解非天然或非典型靶位点需要选择适当LHE骨架、检查靶基因座、选择推定的靶位点以及在靶位点中多达三分之二的碱基对位置处广泛地改变LHE以改变其DNA接触点和裂解特异性。

[0380] 在一个实施例中,可以被设计成重编程的LHE或LHE变体的LHE包括但不限于I-CreI和I-SceI。

[0381] 可以被设计成重编程的LHE或LHE变体的LHE的示例性实例包括但不限于I-AabMI、I-AaeMI、I-AniI、I-ApaMI、I-CapIII、I-CapIV、I-CkaMI、I-CpaMI、I-CpaMII、I-CpaMIII、I-CpaMIV、I-CpaMV、I-CpaV、I-CraMI、I-EjeMI、I-GpeMI、I-GpiI、I-GzeMI、I-GzeMII、I-GzeMIII、I-HjeMI、I-LtrII、I-LtrI、I-LtrWI、I-MpeMI、I-MveMI、I-NcrII、I-NcrI、I-NcrMI、I-OheMI、I-OnuI、I-OsoMI、I-OsoMII、I-OsoMIII、I-OsoMIV、I-PanMI、I-PanMII、I-PanMIII、I-PnoMI、I-ScuMI、I-SmaMI、I-SscMI以及I-Vdi141I。

[0382] 在一个实施例中,重编程的LHE或LHE变体选自由以下组成的组:I-CpaMI变体、I-HjeMI变体、I-OnuI变体、I-PanMI变体和I-SmaMI变体。

[0383] 在一个实施例中,重编程的LHE或LHE变体是I-OnuI变体。参见例如SEQ ID NO:6-14和60-63。

[0384] 在一个实施例中,靶向PD-1基因的重编程的I-OnuI LHE或I-OnuI变体是由天然I-OnuI或其生物活性片段(SEQ ID NO:1-5)产生。在一个优选实施例中,靶向人PD-1基因的重编程的I-OnuI LHE或I-OnuI变体是由现有I-OnuI变体产生。在一个实施例中,产生的重编程的I-OnuI LHE是针对SEQ ID NO:25中所述的人PD-1基因靶位点。在一个实施例中,产生的重编程的I-OnuI LHE是针对SEQ ID NO:30中所述的人PD-1基因靶位点。在一个实施例中,产生的重编程的I-OnuI LHE是针对SEQ ID NO:35中所述的人PD-1基因靶位点。

[0385] 在一个特定实施例中,结合并裂解人PD-1基因的重编程的I-OnuI LHE或I-OnuI变体在DNA识别界面中包含一个或多个氨基酸取代。在特定实施例中,结合并裂解人PD-1基因的I-OnuI LHE与I-OnuI(Taekuchi等人,2011.《美国国家科学院院刊(Proc Natl Acad Sci U.S.A.)》2011年8月9日;108(32):13077-13082)或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI LHE变体、其生物活性片段和/或其另外地变体的DNA识别界面具有至少70%、至少71%、至少72%、至少73%、至少74%、至少75%、至少76%、至少77%、至少78%、至少79%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、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%或至少99%序列一致性。

[0386] 在一个实施例中,结合并裂解人PD-1基因的I-OnuI LHE与I-OnuI(Taekuchi等人,2011.《美国国家科学院院刊》2011年8月9日;108(32):13077-13082)或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI LHE变体、其生物活性片段和/或其另外的变体的DNA识别界面具有至少70%、更优选至少80%、更优选至少85%、更优选至少90%、更优选至少95%、更优选至少97%、更优选至少99%序列一致性。

[0387] 在一个特定实施例中,结合并裂解人PD-1基因的I-OnuI LHE变体在SEQ ID NO:1-

14和60-63中任一个中所述的I-OnuI、其生物活性片段和/或其另外的变体的DNA识别界面中包含一个或多个氨基酸取代或修饰。

[0388] 在一个特定实施例中,结合并裂解人PD-1基因的I-OnuI LHE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体的DNA识别界面中,特别是在位于位置24至50、68至82、180至203以及223至240的子基元中包含一个或多个氨基酸取代或修饰。

[0389] 在一个特定实施例中,结合并裂解人PD-1基因的I-OnuI LHE在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体的DNA识别界面中选自由以下组成的组的氨基酸位置处包含一个或多个氨基酸取代或修饰:19、24、26、28、30、32、34、35、36、37、38、40、42、44、46、48、68、70、72、75、76、77、78、80、82、168、180、182、184、186、188、189、190、191、192、193、195、197、199、201、203、223、225、227、229、231、232、234、236、238以及240。

[0390] 在一个特定实施例中,结合并裂解人PD-1基因的I-OnuI LHE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体的DNA识别界面中,特别是位于位置24至50、68至82、180至203以及223至240的子基元中包含5、10、15、20、25、30、35或40个或更多个氨基酸取代或修饰。

[0391] 在一个特定实施例中,结合并裂解人PD-1基因的I-OnuI LHE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体的DNA识别界面中选自由以下组成的组的氨基酸位置处包含5、10、15、20、25、30、35或40个或更多个氨基酸取代或修饰:19、24、26、28、30、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76、77、78、80、82、168、180、182、184、186、188、189、190、191、192、193、195、197、199、201、203、223、225、227、229、231、232、234、236、238以及240。

[0392] 在一个实施例中,结合并裂解人PD-1基因的I-OnuI LHE变体在位于整个I-OnuI序列内任何地方的另外的位置处包含一个或多个氨基酸取代或修饰。可以被取代和/或修饰的残基包括但不限于接触核酸靶或者直接地或通过水分子与核酸主链或核苷酸碱基相互作用的氨基酸。在一个非限制性实例中,本文所涵盖的结合并裂解人PD-1基因的I-OnuI LHE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-14和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体的选自由以下位置组成的位置组的至少一个位置中包含一个或多个取代和/或修饰,优选至少5个、优选至少10个、优选至少15个、优选至少20个、更优选至少25个、更优选至少30个、甚至更优选地至少35个或甚至更优选至少40个或更多个氨基酸取代:26、28、32、34、35、36、37、38、40、42、44、46、48、59、68、70、72、75、76、78、80、138、143、159、168、178、180、182、184、186、188、189、190、191、192、193、197、199、201、203、207、223、224、225、227、229、232、236以及238。

[0393] 在某些实施例中,HE变体裂解PD-1外显子5靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-10中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40K、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、S72R、N75S、A76Y、S78K、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、

G193P、Q197R、V199R、S201A、S201M、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

[0394] 在一些实施例中,HE变体裂解PD-1外显子5靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-10中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40K、E42R、G44R、Q46E、T48D、V68K、A70Y、S72Q、N75S、A76Y、S78K、K80R、L138M、T143N、S159P、F168L、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、K225R、K229I、F232K、D236E以及V238E。

[0395] 在特定实施例中,HE变体裂解PD-1外显子5靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-10中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72R、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

[0396] 在另外的实施例中,HE变体裂解PD-1外显子5靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-10中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、I224T、K225R、K229I、F232K、D236E以及V238E。

[0397] 在另外的实施例中,HE变体裂解PD-1外显子5靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-10中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201A、T203G、K207R、Y223R、K225R、F232K、D236E以及V238E。

[0398] 在特定实施例中,HE变体裂解PD-1外显子5靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:6-10中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28E、N32V、K34P、S35N、S36I、V37P、G38R、S40R、E42R、G44R、Q46E、T48D、N59D、V68K、A70Y、S72Q、N75S、A76Y、K80R、L138M、T143N、S159P、F168L、E178D、C180H、F182G、N184I、I186A、S188R、K189R、S190P、K191A、L192S、G193P、Q197R、V199R、S201M、T203G、Y223R、K225R、F232K、D236E以及V238E。

[0399] 在特定实施例中,HE变体裂解PD-1外显子1靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另

外的变体中选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代：26、28、30、32、34、35、36、37、38、40、42、44、46、48、68、70、72、75、76、78、80、100、132、138、143、155、159、178、180、184、186、189、190、191、192、193、195、201、203、207、223、225、227、232、236、238以及240。

[0400] 在一些实施例中，HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个：L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46A、Q46T、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80R、K80C、I100V、V132A、L138M、T143N、S155G、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0401] 在另外的实施例中，HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个：L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46A、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80R、I100V、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0402] 在某些实施例中，HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个：L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48V、V68I、A70T、S72D、N75R、A76Y、S78R、K80C、I100V、V132A、L138M、T143N、S155G、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0403] 在某些实施例中，HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个：L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68I、A70T、S72N、N75H、A76Y、S78T、K80R、I100V、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0404] 在某些实施例中，HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个：L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70L、S72N、N75H、A76Y、K80V、T82Y、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、

K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0405] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37G、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70T、S72N、N75H、A76Y、K80V、T82Y、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0406] 在某些实施例中,HE变体裂解PD-1外显子1靶位点并在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:11-12和60-63中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26G、R28S、R30L、N32R、K34R、S35G、S36T、V37A、G38R、S40H、E42R、G44S、Q46T、T48M、V68S、A70Y、S72N、N75H、A76Y、K80E、T82F、L138M、T143N、S159P、E178D、C180S、N184R、I186R、K189N、S190V、K191N、L192A、G193R、Q195R、S201E、T203S、K207R、Y223H、K225Y、K227G、F232R、D236Q、V238R以及T240E。

[0407] 在特定实施例中,HE变体裂解PD-1外显子2靶位点并且在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:13-14中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中选自由以下位置组成的位置组的至少一个位置中包含至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个氨基酸取代:24、26、28、30、32、34、35、36、37、38、40、41、42、44、46、48、68、70、72、74、75、76、78、80、82、116、138、143、159、168、178、180、182、184、186、188、189、190、191、192、193、195、199、203、207、225、227、229、232、236以及238。

[0408] 在另外的实施例中,HE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:13-14中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:S24C、L26Q、R28Y、R28H、R30S、N32V、N32L、K34N、K34R、S35N、S35T、S36R、V37S、V37T、G38R、G38K、S40R、T41A、E42R、G44S、G44R、Q46E、Q46A、T48E、V68I、A70N、S72I、D74N、N75T、N75R、A76S、A76R、S78R、K80S、T82G、T82R、V116L、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203A、T203S、K207R、K225N、K225T、K227W、K227S、K229A、K229P、F232R、W234A、W234D、D236E以及V238R。

[0409] 在特定实施例中,HE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:13-14中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个:L26Q、R28Y、R30S、N32V、K34N、S35N、S36R、V37S、G38R、S40R、T41A、E42R、G44R、Q46A、T48E、A70N、S72I、N75T、A76S、S78R、K80S、T82G、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203A、K207R、K225N、K227W、K229A、F232R、W234A、D236E以及V238R。

[0410] 在一些实施例中,HE变体在I-OnuI (SEQ ID NO:1-5) 或如SEQ ID NO:13-14中任一个中所述的I-OnuI变体、其生物活性片段和/或其另外的变体中包含以下氨基酸取代中的

至少5个、至少15个、优选至少25个、更优选至少35个或甚至更优选至少40个或更多个：S24C、R28H、N32L、K34R、S35T、V37T、G38K、S40R、E42R、G44S、Q46E、T48E、V68I、A70N、S72I、D74N、N75R、A76R、S78R、K80S、T82R、V116L、L138M、T143N、S159P、F168L、E178D、C180N、F182Y、N184H、I186K、K189G、S190R、K191T、L192T、G193R、Q195Y、V199R、T203S、K207R、K225T、K227S、K229P、F232R、W234D、D236E以及V238R。

[0411] 在特定实施例中，结合并裂解人PD-1基因的I-0nuI LHE变体包含与SEQ ID NO:6-14和60-63中任一个或其生物活性片段中所述的氨基酸序列至少80%、优选至少85%、更优选至少90%或甚至更优选至少95%一致的氨基酸序列。

[0412] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:6-14和60-63中任一个或其生物活性片段中所述的氨基酸序列。

[0413] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:6或其生物活性片段中所述的氨基酸序列。

[0414] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:7或其生物活性片段中所述的氨基酸序列。

[0415] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:8或其生物活性片段中所述的氨基酸序列。

[0416] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:9或其生物活性片段中所述的氨基酸序列。

[0417] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:10或其生物活性片段中所述的氨基酸序列。

[0418] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:11或其生物活性片段中所述的氨基酸序列。

[0419] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:12或其生物活性片段中所述的氨基酸序列。

[0420] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:13或其生物活性片段中所述的氨基酸序列。

[0421] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:14或其生物活性片段中所述的氨基酸序列。

[0422] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:60或其生物活性片段中所述的氨基酸序列。

[0423] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:61或其生物活性片段中所述的氨基酸序列。

[0424] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:62或其生物活性片段中所述的氨基酸序列。

[0425] 在特定实施例中，I-0nuI LHE变体包含SEQ ID NO:63或其生物活性片段中所述的氨基酸序列。

[0426] 2.MegaTAL

[0427] 在各种实施例中，包含归巢核酸内切酶变体的megaTAL被重编程以在PD-1基因中的靶位点中引入双链断裂(DSB)。在特定实施例中，megaTAL在PD-1基因外显子5中，优选地

在PD-1基因外显子5中的SEQ ID NO:27处,并且更优选地在PD-1基因外显子5中SEQ ID NO:27中的序列“ATAC”处引入DSB。在特定实施例中, megaTAL在PD-1基因外显子1中, 优选地在PD-1基因外显子1中的SEQ ID NO:32处, 并且更优选地在PD-1基因外显子1中SEQ ID NO:32中的序列“ATCC”处引入双链断裂。在特定实施例中, megaTAL在PD-1基因外显子2中, 优选地在PD-1基因外显子2中的SEQ ID NO:37处, 并且更优选地在PD-1基因外显子2中SEQ ID NO:37中的序列“ACTT”处引入双链断裂。

[0428] “megaTAL”是指包含TALE DNA结合结构域以及结合并裂解PD-1基因中的DNA靶序列的归巢核酸内切酶变体的多肽, 并且任选地包含一个或多个连接子和/或另外的功能结构域, 例如末端加工酶中展现5′-3′核酸外切酶、5′-3′碱性核酸外切酶、3′-5′核酸外切酶(例如Trex2)、5′瓣状核酸内切酶、解螺旋酶或不依赖模板的DNA聚合酶活性的末端加工酶结构域。

[0429] 在特定实施例中, megaTAL可以与末端加工酶一起引入细胞中, 所述末端加工酶展现5′-3′核酸外切酶、5′-3′碱性核酸外切酶、3′-5′核酸外切酶(例如Trex2)、5′瓣状核酸内切酶、解螺旋酶、模板依赖性DNA聚合酶或不依赖模板的DNA聚合酶活性。megaTAL和3′加工酶可以分开地引入, 例如利用不同载体或独立mRNA引入; 或一起引入, 例如以融合蛋白形式, 或以通过病毒自裂解肽或IRES元件隔开的多顺反子构建体引入。

[0430] “TALE DNA结合结构域”是转录活化因子样效应子(TALE或TAL效应子)的DNA结合部分, 其模拟植物转录活化因子以操作植物转录物组(参见例如Kay等人, 2007.《科学(Science)》318:648-651)。特定实施例中所涵盖的TALE DNA结合结构域是从头工程改造得到或由天然存在的TALE, 例如来自野油菜黄单胞菌野油菜致病变种(*Xanthomonas campestris* pv.vesicatoria)、大豆黄单胞菌(*Xanthomonas gardneri*)、半透明黄单胞菌(*Xanthomonas translucens*)、地毯草黄单胞菌(*Xanthomonas axonopodis*)、番茄黄单胞菌(*Xanthomonas perforans*)、苜蓿黄单胞菌(*Xanthomonas alfalfa*)、柑橘黄单胞菌(*Xanthomonas citri*)、辣椒-番茄黄单胞菌(*Xanthomonas euvesicatoria*)和水稻黄单胞菌(*Xanthomonas oryzae*)的AvrBs3, 以及来自茄科雷尔氏菌(*Ralstonia solanacearum*)的brg11和hpx17工程改造得到。用于得到并设计DNA结合结构域的TALE蛋白质的示例性实例公开于美国专利第9,017,967号和其中引用的参考文献中, 其全部以引用的方式整体并入本文中。

[0431] 在特定实施例中, megaTAL包含TALE DNA结合结构域, 该TALE DNA结合结构域包含TALE DNA结合结构域与其相应靶DNA序列结合所涉及的一个或多个重复单元。单一“重复单元”(又称为“重复”)典型地是33-35个氨基酸长度。每个TALE DNA结合结构域重复单元典型地在该重复的12位和/或13位包括1或2个DNA结合残基, 构成重复可变双残基(Repeat Variable Di-Residue, RVD)。已测定这些TALE DNA结合结构域的DNA识别序列的天然(典型)编码序列, 由此在12位和13位的HD序列使得结合至胞嘧啶(C), NG结合至T, NI结合至A, NN结合至G或A且NG结合至T。在某些实施例中, 涵盖非典型(非典型性)RVD。

[0432] 适用于特定实施例中所涵盖的特定megaTAL中的非典型RVD的示例性实例包括但不限于用于识别鸟嘌呤(G)的HH、KH、NH、NK、NQ、RH、RN、SS、NN、SN、KN; 用于识别腺嘌呤(A)的NI、KI、RI、HI、SI; 用于识别胸腺嘧啶(T)的NG、HG、KG、RG; 用于识别胞嘧啶(C)的RD、SD、HD、ND、KD、YG; 用于识别A或G的NV、HN; 以及用于识别A或T或G或C的H*、HA、KA、N*、NA、NC、NS、RA、

S*,其中(*)意思指13位的氨基酸不存在。适用于特定实施例中所涵盖的特定megaTAL中的RVD的其它示例性实例还包括以全文引用的方式并入本文中的美国专利第8,614,092号中所公开的那些。

[0433] 在特定实施例中,本文所涵盖的megaTAL包括含3至30个重复单元的TALE DNA结合结构域。在某些实施例中, megaTAL包含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或30个TALE DNA结合结构域重复单元。在一个优选实施例中,本文所涵盖的megaTAL包括含5-15个重复单元、更优选7-15个重复单元、更优选9-15个重复单元并且更优选地9、10、11、12、13、14或15个重复单元的TALE DNA结合结构域。

[0434] 在特定实施例中,本文所涵盖的megaTAL包括含3至30个重复单元和另外的单个截短的TALE重复单元的TALE DNA结合结构域,所述截短的TALE重复单元包含位于一组TALE重复单元的C末端处的20个氨基酸,即,另外的在C末端的一半TALE DNA结合结构域重复单元(本文别处所公开的C-帽的氨基酸-20至-1,见下文)。在特定实施例中,本文所涵盖的megaTAL包括含3.5至30.5个重复单元的TALE DNA结合结构域。在某些实施例中, megaTAL包含3.5、4.5、5.5、6.5、7.5、8.5、9.5、10.5、11.5、12.5、13.5、14.5、15.5、16.5、17.5、18.5、19.5、20.5、21.5、22.5、23.5、24.5、25.5、26.5、27.5、28.5、29.5或30.5个TALE DNA结合结构域重复单元。在一个优选实施例中,本文所涵盖的megaTAL包括含5.5-15.5个重复单元、更优选7.5-15.5个重复单元、更优选9.5-15.5个重复单元并且更优选地9.5、10.5、11.5、12.5、13.5、14.5或15.5个重复单元的TALE DNA结合结构域。

[0435] 在特定实施例中, megaTAL包含TAL效应子架构,所述TAL效应子架构包含“N末端结构域(NTD)”多肽、一个或多个TALE重复结构域/单元、“C末端结构域(CTD)”多肽和归巢核酸内切酶变体。在一些实施例中,所述NTD、TALE重复和/或CTD结构域是来自相同物种。在其它实施例中,所述NTD、TALE重复和/或CTD结构域中的一个或多个是来自不同物种。

[0436] 如本文所使用,术语“N末端结构域(NTD)”多肽是指侧接天然存在的TALE DNA结合结构域的N末端部分或片段的序列。NTD序列如果存在的话,可具有任何长度,只要TALE DNA结合结构域重复单元保持结合DNA的能力即可。在特定实施例中,NTD多肽包含在TALE DNA结合结构域N末端的至少120个到至少140个或更多个氨基酸(0是在最N末端重复单元的氨基酸1)。在特定实施例中,NTD多肽包含在TALE DNA结合结构域N末端的至少约120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139或至少140个氨基酸。在一个实施例中,本文所涵盖的megaTAL包含在黄单胞菌属TALE蛋白质的至少约氨基酸+1至+122到至少约+1至+137的NTD多肽(0是在最N末端重复单元的氨基酸1)。在特定实施例中,NTD多肽包含在黄单胞菌属TALE蛋白质TALE DNA结合结构域N末端的至少约122、123、124、125、126、127、128、129、130、131、132、133、134、135、136或137个氨基酸。在一个实施例中,本文所涵盖的megaTAL包含在罗尔斯通氏菌属TALE蛋白质的至少氨基酸+1至+121的NTD多肽(0是在最N末端重复单元的氨基酸1)。在特定实施例中,NTD多肽包含在罗尔斯通氏菌属TALE蛋白质TALE DNA结合结构域N末端的至少约121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136或137个氨基酸。

[0437] 如本文所使用,术语“C末端结构域(CTD)”多肽是指侧接天然存在的TALE DNA结合结构域的C末端部分或片段的序列。CTD序列如果存在的话,可具有任何长度,只要TALE DNA结合结构域重复单元保持结合DNA的能力即可。在特定实施例中,CTD多肽包含在TALE DNA

结合结构域最后一个完整重复的C末端的至少20个到至少85个或更多个氨基酸(前20个氨基酸是在最后一个C末端完整重复单元C末端的一半重复单元)。在特定实施例中,CTD多肽包含在TALE DNA结合结构域最后一个完整重复的C末端的至少约20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、443、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或至少85个氨基酸。在一个实施例中,本文所涵盖的megaTAL包含在黄单胞菌属TALE蛋白质的至少约氨基酸-20至-1的CTD多肽(-20是在最后一个C末端完整重复单元的C末端的一半重复单元的氨基酸1)。在特定实施例中,CTD多肽包含在黄单胞菌属TALE蛋白质TALE DNA结合结构域的最后完整重复的C末端的至少约20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2或1个氨基酸。在一个实施例中,本文所涵盖的megaTAL包含在罗尔斯通氏菌属TALE蛋白质的至少约氨基酸-20至-1的CTD多肽(-20是在最后一个C末端完整重复单元的C末端的一半重复单元的氨基酸1)。在特定实施例中,CTD多肽包含在罗尔斯通氏菌属TALE蛋白质TALE DNA结合结构域的最后完整重复的C末端的至少约20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2或1个氨基酸。

[0438] 在特定实施例中,本文所涵盖的megaTAL包含融合多肽,所述融合多肽包含被工程改造成结合靶序列的TALE DNA结合结构域、重编程用于结合并裂解靶序列的归巢核酸内切酶以及任选地NTD和/或CTD多肽任选地通过本文别处涵盖的一个或多个连接子多肽彼此接合在一起。不希望受任何特定理论束缚,预期包含TALE DNA结合结构域和任选地NTD和/或CTD多肽的megaTAL与连接子多肽融合,所述连接子多肽又与归巢核酸内切酶变体融合。因此,TALE DNA结合结构域结合在距归巢核酸内切酶变体DNA结合结构域所结合的靶序列约1、2、3、4、5、6、7、8、9、10、11、12、13、14或15个核苷酸内的DNA靶序列。通过这种方式,本文所涵盖的megaTAL增加基因组编辑的特异性和效率。

[0439] 在一个实施例中,megaTAL包含归巢核酸内切酶变体和TALE DNA结合结构域,所述TALE DNA结合结构域结合在重编程的归巢核酸内切酶的结合位点上游约2、3、4、5或6个核苷酸,优选2或4个核苷酸内的核苷酸序列。

[0440] 在一个实施例中,megaTAL包含归巢核酸内切酶变体和结合SEQ ID NO:26中所述的核苷酸序列的TALE DNA结合结构域,所述核苷酸序列是在归巢核酸内切酶变体(SEQ ID NO:25)所结合和裂解的核苷酸序列上游的5个核苷酸。在优选实施例中,megaTAL靶序列是SEQ ID NO:27。

[0441] 在一个实施例中,megaTAL包含归巢核酸内切酶变体和结合SEQ ID NO:31中所述的核苷酸序列的TALE DNA结合结构域,所述核苷酸序列是在归巢核酸内切酶变体(SEQ ID NO:30)所结合和裂解的核苷酸序列上游的2个核苷酸。在优选实施例中,megaTAL靶序列是SEQ ID NO:32。

[0442] 在一个实施例中,megaTAL包含归巢核酸内切酶变体和结合SEQ ID NO:36中所述的核苷酸序列的TALE DNA结合结构域,所述核苷酸序列是在归巢核酸内切酶变体(SEQ ID NO:35)所结合和裂解的核苷酸序列上游的5个核苷酸。在优选实施例中,megaTAL靶序列是SEQ ID NO:37。

[0443] 在特定实施例中,本文所涵盖的megaTAL包含一个或多个TALE DNA结合重复单元和LHE变体,所述LHE变体是由选自以下组成的组的LHE设计或重编程得到:I-AabMI、I-

AaeMI、I-AniI、I-ApaMI、I-CapIII、I-CapIV、I-CkaMI、I-CpaMI、I-CpaMII、I-CpaMIII、I-CpaMIV、I-CpaMV、I-CpaV、I-CraMI、I-EjeMI、I-GpeMI、I-GpiI、I-GzeMI、I-GzeMII、I-GzeMIII、I-HjeMI、I-LtrII、I-LtrI、I-LtrWI、I-MpeMI、I-MveMI、I-NcrII、I-NcrI、I-NcrMI、I-OheMI、I-OnuI、I-OsoMI、I-OsoMII、I-OsoMIII、I-OsoMIV、I-PanMI、I-PanMII、I-PanMIII、I-PnoMI、I-ScuMI、I-SmaMI、I-SscMI、I-Vdi141I 以及其变体,或优选地 I-CpaMI、I-HjeMI、I-OnuI、I-PanMI、SmaMI 和其变体,或更优选地 I-OnuI 和其变体。

[0444] 在特定实施例中,本文所涵盖的 megaTAL 包含 NTD、一个或多个 TALE DNA 结合重复单元、CTD 和 LHE 变体,所述 LHE 变体选自由以下组成的组: I-AabMI、I-AaeMI、I-AniI、I-ApaMI、I-CapIII、I-CapIV、I-CkaMI、I-CpaMI、I-CpaMII、I-CpaMIII、I-CpaMIV、I-CpaMV、I-CpaV、I-CraMI、I-EjeMI、I-GpeMI、I-GpiI、I-GzeMI、I-GzeMII、I-GzeMIII、I-HjeMI、I-LtrII、I-LtrI、I-LtrWI、I-MpeMI、I-MveMI、I-NcrII、I-NcrI、I-NcrMI、I-OheMI、I-OnuI、I-OsoMI、I-OsoMII、I-OsoMIII、I-OsoMIV、I-PanMI、I-PanMII、I-PanMIII、I-PnoMI、I-ScuMI、I-SmaMI、I-SscMI、I-Vdi141I 以及其变体,或优选地 I-CpaMI、I-HjeMI、I-OnuI、I-PanMI、SmaMI 和其变体,或更优选地 I-OnuI 和其变体。

[0445] 在特定实施例中,本文所涵盖的 megaTAL 包含 NTD、约 9.5 至约 15.5 个 TALE DNA 结合重复单元和 LHE 变体,所述 LHE 变体选自由以下组成的组: I-AabMI、I-AaeMI、I-AniI、I-ApaMI、I-CapIII、I-CapIV、I-CkaMI、I-CpaMI、I-CpaMII、I-CpaMIII、I-CpaMIV、I-CpaMV、I-CpaV、I-CraMI、I-EjeMI、I-GpeMI、I-GpiI、I-GzeMI、I-GzeMII、I-GzeMIII、I-HjeMI、I-LtrII、I-LtrI、I-LtrWI、I-MpeMI、I-MveMI、I-NcrII、I-NcrI、I-NcrMI、I-OheMI、I-OnuI、I-OsoMI、I-OsoMII、I-OsoMIII、I-OsoMIV、I-PanMI、I-PanMII、I-PanMIII、I-PnoMI、I-ScuMI、I-SmaMI、I-SscMI、I-Vdi141I 以及其变体,或优选地 I-CpaMI、I-HjeMI、I-OnuI、I-PanMI、SmaMI 和其变体,或更优选地 I-OnuI 和其变体。

[0446] 在特定实施例中,本文所涵盖的 megaTAL 包含具有约 122 个氨基酸至 137 个氨基酸、约 9.5 个、约 10.5 个、约 11.5 个、约 12.5 个、约 13.5 个、约 14.5 个或约 15.5 个结合重复单元的 NTD、具有约 20 个氨基酸至约 85 个氨基酸的 CTD 以及 I-OnuI LHE 变体。在特定实施例中,NTD、DNA 结合结构域和 CTD 中的任一个、二个或全部可以由相同物种或不同物种以任何适合组合设计。

[0447] 在特定实施例中,本文所涵盖的 megaTAL 包含 SEQ ID NO:15-23 和 64 中任一个中所述的氨基酸序列。

[0448] 在某些实施例中,本文所涵盖的 megaTAL 是由 SEQ ID NO:40-42 和 65-68 中任一个中所述的 mRNA 序列编码。

[0449] 在特定实施例中,本文所涵盖的 megaTAL-Trex2 融合蛋白包含 SEQ ID NO:24 中所述的氨基酸序列。

[0450] 在某些实施例中, megaTAL 包含 TALE DNA 结合结构域并且 I-OnuI LHE 变体结合并裂解 SEQ ID NO:27 中所述的核苷酸序列。在特定实施例中,结合并裂解 SEQ ID NO:27 中所述的核苷酸序列的 megaTAL 包含 SEQ ID NO:15-19 中任一个中所述的氨基酸序列。

[0451] 在某些实施例中, megaTAL 包含 TALE DNA 结合结构域并且 I-OnuI LHE 变体结合并裂解 SEQ ID NO:32 中所述的核苷酸序列。在特定实施例中,结合并裂解 SEQ ID NO:32 中所述的核苷酸序列的 megaTAL 包含 SEQ ID NO:20-21 和 64 中任一个中所述的氨基酸序列。

[0452] 在某些实施例中, megaTAL包含TALEDNA结合结构域并且I-OnuI LHE变体结合并裂解SEQ ID NO:37中所述的核苷酸序列。在特定实施例中, 结合并裂解SEQ ID NO:37中所述的核苷酸序列的megaTAL包含SEQ ID NO:22-23中任一个中所述的氨基酸序列。

[0453] 3. 末端加工酶

[0454] 特定实施例中所涵盖的基因组编辑组合物和方法包含使用核酸酶变体以及末端加工酶的一个或多个拷贝编辑细胞基因组。在特定实施例中, 单一聚核苷酸编码通过连接子、自裂解肽序列如2A序列或通过IRES序列隔开的归巢核酸内切酶变体和末端加工酶。在特定实施例中, 基因组编辑组合物包含编码核酸酶变体的聚核苷酸和编码末端加工酶的独立的聚核苷酸。在特定实施例中, 基因组编辑组合物包含编码归巢核酸内切酶变体末端加工酶单一多肽融合物的聚核苷酸, 以及通过自裂解肽隔开的末端加工酶的一连串拷贝。

[0455] 术语“末端加工酶”是指修饰聚核苷酸链的暴露端的酶。所述聚核苷酸可以是双链DNA (dsDNA)、单链DNA (ssDNA)、RNA、DNA与RNA的双链杂交物以及合成DNA (例如含有除A、C、G和T外的碱基)。末端加工酶可以通过添加一个或多个核苷酸、移除一个或多个核苷酸、移除或修饰磷酸酯基和/或移除或修饰羟基来修饰暴露的聚核苷酸链两端。末端加工酶可以在核酸内切酶切割位点处或在通过其它化学或机械手段, 如剪切 (例如通过穿过精细规格针、加热、超声波处理、微型珠粒翻滚以及喷洒)、电离辐射、紫外辐射、氧自由基、化学水解以及化学治疗剂产生的末端处修饰末端。

[0456] 在特定实施例中, 特定实施例中所涵盖的基因组编辑组合物和方法包含使用归巢核酸内切酶变体或megaTAL和DNA末端加工酶编辑细胞基因组。

[0457] 术语“DNA末端加工酶”是指修饰DNA的暴露端的酶。DNA末端加工酶可以修饰钝端或交错端 (具有5'或3'突出的末端)。DNA末端加工酶可以修饰单链或双链DNA。DNA末端加工酶可以在核酸内切酶切割位点处或在通过其它化学或机械手段, 如剪切 (例如通过穿过精细规格针、加热、超声波处理、微型珠粒翻滚以及喷洒)、电离辐射、紫外辐射、氧自由基、化学水解以及化学治疗剂产生的末端处修饰末端。DNA末端加工酶可以通过添加一个或多个核苷酸、移除一个或多个核苷酸、移除或修饰磷酸酯基和/或移除或修饰羟基来修饰暴露的DNA末端。

[0458] 适用于本文所涵盖的特定实施例中的DNA末端加工酶的示例性实例包括但不限于: 5'-3'核酸外切酶、5'-3'碱性核酸外切酶、3'-5'核酸外切酶、5'瓣状核酸内切酶、解螺旋酶、磷酸酶、水解酶以及不依赖模板的DNA聚合酶。

[0459] 适用于本文所涵盖的特定实施例中的DNA末端加工酶的其他示例性实例包括但不限于Trex2、Trex1、无跨膜结构域的Trex1、Apollo、Artemis、DNA2、Exo1、ExoT、ExoIII、Fen1、Fan1、MreII、Rad2、Rad9、TdT (末端脱氧核苷酸转移酶)、PNKP、RecE、RecJ、RecQ、λ核酸外切酶、Sox、牛痘DNA聚合酶、核酸外切酶I、核酸外切酶III、核酸外切酶VII、NDK1、NDK5、NDK7、NDK8、WRN、T7-核酸外切酶基因6、禽成髓细胞瘤病毒整合蛋白 (IN)、Bloom、热敏磷酸酶、碱性磷酸酶、聚核苷酸激酶 (PNK)、ApeI、绿豆核酸酶、Hex1、TTRAP (TDP2)、Sgs1、Sae2、CUP、聚合酶μ (Pol mu)、聚合酶λ (Pol lambda)、MUS81、EME1、EME2、SLX1、SLX4以及UL-12。

[0460] 在特定实施例中, 本文所涵盖的用于编辑细胞基因组的基因组编辑组合物和方法包括含归巢核酸内切酶变体或megaTAL和核酸外切酶的多肽。术语“核酸外切酶”是指通过水解反应裂解在聚核苷酸链一端处的磷酸二酯键的酶, 所述水解反应使3'或5'端的磷酸二

酯键断裂。

[0461] 适用于本文所涵盖的特定实施例中的核酸外切酶的示例性实例包括但不限于：hExoI、酵母ExoI、大肠杆菌ExoI、hTREX2、小鼠TREX2、大鼠TREX2、hTREX1、小鼠TREX1、大鼠TREX1以及大鼠TREX1。

[0462] 在特定实施例中，DNA末端加工酶是3'至5'核酸外切酶，优选地是Trex 1或Trex2，更优选地是Trex2，并且甚至更优选地是人或小鼠Trex2。

[0463] D. 靶位点

[0464] 特定实施例中所涵盖的核酸酶变体可以被设计用于结合至任何适合的靶序列并且相较于天然存在的核酸酶，可以具有新颖的结合特异性。在特定实施例中，靶位点是基因调控区，包括但不限于启动子、增强子、阻遏子元件等。在特定实施例中，靶位点是基因编码区或剪接位点。在某些实施例中，核酸酶变体被设计用于下调或减少基因的表达。在特定实施例中，核酸酶变体和供体修复模板可以被设计用于修复或除去所希望的靶序列。

[0465] 在各种实施例中，核酸酶变体结合至并裂解程序性死亡受体1 (PD-1) 基因中的靶序列。PD-1又称为程序性细胞死亡蛋白1 (PDCD1)、全身性红斑狼疮敏感性蛋白2 (SLEB2)、CD279、HPD1、PD1、HPD-L以及HSLE1。PD-1是B7/CD28共刺激受体家族的成员。PD-1分子由细胞外配体结合IgV结构域、跨膜结构域和细胞内结构域组成，所述细胞内结构域具有位于免疫酪氨酸抑制基元 (ITIM) 和免疫受体抑制酪氨酸转换基元 (ITSM) 内的潜在磷酸化位点。PD-1是在T细胞、Treg、耗竭T细胞、B细胞、活化的单核细胞、树突状细胞 (DC)、自然杀手 (NK) 细胞以及自然杀手T (NKT) 细胞上表达的抑制性辅助受体。PD-1通过结合至其配体，即程序性死亡配体1 (PD-L1) 和程序性死亡配体2 (PD-L2) 负调控T细胞活化。PD-1结合抑制T细胞增殖，以及干扰素- γ (IFN- γ)、肿瘤坏死因子- α 和IL-2产生，并减少T细胞存活。PD-1表达是经历高水平刺激的“耗竭的”T细胞的一个标志。这一耗竭状态是在慢性感染和癌症期间出现，以T细胞功能障碍导致对感染和肿瘤的次佳控制为特征。

[0466] 在特定实施例中，归巢核酸内切酶变体或megaTAL在PD-1基因中的靶位点中引入双链断裂 (DSB)。在特定实施例中，归巢核酸内切酶变体或megaTAL在PD-1基因外显子5中，优选地在PD-1基因外显子5中的SEQ ID NO:25 (或SEQ ID NO:27) 处，并且更优选地在PD-1基因外显子5中的SEQ ID NO:25 (或SEQ ID NO:27) 中的序列“ATAC”处引入DSB。

[0467] 在一个优选实施例中，归巢核酸内切酶变体或megaTAL裂解双链DNA并将DSB引入SEQ ID NO:25或27中所述的聚核苷酸序列中。

[0468] 在特定实施例中，归巢核酸内切酶变体或megaTAL在PD-1基因外显子1中，优选地在PD-1基因外显子1中的SEQ ID NO:30 (或SEQ ID NO:32) 处，并且更优选地在PD-1基因外显子1中的SEQ ID NO:30 (或SEQ ID NO:32) 中的序列“ATCC”处引入DSB。

[0469] 在一个优选实施例中，归巢核酸内切酶变体或megaTAL裂解双链DNA并将DSB引入SEQ ID NO:30或32中所述的聚核苷酸序列中。

[0470] 在特定实施例中，归巢核酸内切酶变体或megaTAL在PD-1基因外显子2中，优选地在PD-1基因外显子2中的SEQ ID NO:35 (或SEQ ID NO:37) 处，并且更优选地在PD-1基因外显子2中的SEQ ID NO:35 (或SEQ ID NO:37) 中的序列“ACTT”处引入DSB。

[0471] 在一个优选实施例中，归巢核酸内切酶变体或megaTAL裂解双链DNA并将DSB引入SEQ ID NO:35或37中所述的聚核苷酸序列中。

[0472] 在一个优选实施例中,PD-1基因是人PD-1基因。

[0473] E. 供体修复模板

[0474] 核酸酶变体可用于在靶序列中引入DSB;所述DSB可以在一个或多个供体修复模板存在下,通过同源介导修复(HDR)机制修复。

[0475] 在各种实施例中,所述供体修复模板包含编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的一个或多个聚核苷酸。

[0476] 在各种实施例中,预期在独立地编码靶向不同免疫抑制路径的免疫效力增强剂和/或免疫抑制信号阻断剂的多个供体修复模板存在下,向细胞提供工程改造的核酸酶,产生治疗功效和持久性增加的基因组编辑过的T细胞。例如,靶向PD-1、LAG-3、CTLA-4、TIM3、IL-10R、TIGIT以及TGFBR11路径的组合的免疫效力增强剂或免疫抑制信号在特定实施例中可为优选的。

[0477] 在特定实施例中,使用了供体修复模板将序列插入基因组中。在特别优选的实施例中,使用了供体修复模板修复或修饰基因组中的序列。

[0478] 在各种实施例中,通过用包含供体修复模板的腺相关病毒(AAV)、逆转录病毒如慢病毒、IDLV等、单纯疱疹病毒、腺病毒或牛痘病毒载体转导细胞,将供体修复模板引入造血细胞,例如T细胞中。

[0479] 在特定实施例中,供体修复模板包含侧接DSB位点的一个或多个同源臂。

[0480] 如本文所使用,术语“同源臂”是指供体修复模板中与侧接靶位点处由核酸酶引入的DNA断裂的DNA序列一致或大体上一致的核酸序列。在一个实施例中,供体修复模板包含5'同源臂,其包含与在DNA断裂位点5'端的DNA序列一致或大体上一致的核酸序列。在一个实施例中,供体修复模板包含3'同源臂,其包含与在DNA断裂位点3'端的DNA序列一致或大体上一致的核酸序列。在一个优选实施例中,供体修复模板包含5'同源臂和3'同源臂。供体修复模板可与紧邻DSB位点的基因组序列同源,或与在距DSB位点多个碱基对内的基因组序列同源。在一个实施例中,供体修复模板包含与基因组序列同源的约5bp、约10bp、约25bp、约50bp、约100bp、约250bp、约500bp、约1000bp、约2500bp、约5000bp、约10000bp或更长的核酸序列,包括任何中间长度的同源序列。

[0481] 特定实施例中所涵盖的适合长度的同源臂的示例性实例可以独立地选择并且包括但不限于:约100bp、约200bp、约300bp、约400bp、约500bp、约600bp、约700bp、约800bp、约900bp、约1000bp、约1100bp、约1200bp、约1300bp、约1400bp、约1500bp、约1600bp、约1700bp、约1800bp、约1900bp、约2000bp、约2100bp、约2200bp、约2300bp、约2400bp、约2500bp、约2600bp、约2700bp、约2800bp、约2900bp或约3000bp,或更长的同源臂,包括所有中间长度的同源臂。

[0482] 适合同源臂长度的其它示例性实例包括但不限于:约100bp至约3000bp、约200bp至约3000bp、约300bp至约3000bp、约400bp至约3000bp、约500bp至约3000bp、约500bp至约2500bp、约500bp至约2000bp、约750bp至约2000bp、约750bp至约1500bp或约1000bp至约1500bp,包括所有中间长度的同源臂。

[0483] 在一个特定实施例中,5'同源臂和3'同源臂的长度独立地选自约500bp至约1500bp。在一个实施例中,5'同源臂是约1500bp并且3'同源臂是约1000bp。在一个实施例中,5'同源臂在约200bp至约600bp之间并且3'同源臂在约200bp至约600bp之间。在一个实

施例中,5'同源臂是约200bp并且3'同源臂是约200bp。在一个实施例中,5'同源臂是约300bp并且3'同源臂是约300bp。在一个实施例中,5'同源臂是约400bp并且3'同源臂是约400bp。在一个实施例中,5'同源臂是约500bp并且3'同源臂是约500bp。在一个实施例中,5'同源臂是约600bp并且3'同源臂是约600bp。

[0484] 供体修复模板还可以包含本文别处所涵盖的一个或多个聚核苷酸,如启动子和/或增强子、非翻译区(UTR)、Kozak序列、聚腺苷酸化信号、其它限制酶位点、多克隆位点、内部核糖体进入位点(IRES)、重组酶识别位点(例如LoxP、FRT和Att位点)、终止密码子、转录终止信号以及编码自裂解多肽、表位标签的聚核苷酸。

[0485] 在一个实施例中,供体修复模板包括含PD-1基因或其部分的聚核苷酸并且被设计用于在基因组PD-1序列中引入一个或多个突变以使突变型PD-1基因产物表达。在一个实施例中,突变型PD-1具有降低的配体结合和/或减弱的细胞内信号传导。

[0486] 在各种实施例中,供体修复模板包含5'同源臂;RNA聚合酶II启动子;编码免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体的一个或多个聚核苷酸;以及3'同源臂。

[0487] 在各种实施例中,用供体修复模板修饰靶位点,所述供体修复模板包含5'同源臂;编码自裂解病毒肽如T2A、免疫效力增强剂、免疫抑制信号阻断剂或工程改造的抗原受体,任选地聚腺苷酸化信号或自裂解肽的一个或多个聚核苷酸;以及3'同源臂,其中所述一个或多个聚核苷酸的表达是由内源性PD-1启动子控制。

[0488] 1. 免疫效力增强剂

[0489] 在特定实施例中,通过在编码免疫效力增强剂的供体修复模板存在下,在PD-1基因中引入DSB使本文所涵盖的基因组编辑过的免疫效应细胞更强效和/或对免疫抑制因子更具抗性。如本文所使用,术语“免疫效力增强剂”是指刺激和/或增强T细胞活化和/或功能的非天然存在的分子、免疫增强因子,以及在T细胞或其它免疫细胞中将来自肿瘤微环境的免疫抑制信号转化成免疫刺激信号的非天然存在的多肽。

[0490] 在特定实施例中,免疫效力增强剂选自由以下组成的组:双特异性T细胞接合分子(BiTE);免疫增强因子,包括但不限于细胞因子、趋化因子、细胞毒素和/或细胞因子受体;以及翻转受体。

[0491] 在一些实施例中,免疫效力增强剂、免疫增强因子或翻转受体是包含蛋白质不稳定结构域。

[0492] a. 双特异性T细胞接合分子(BiTE)

[0493] 在特定实施例中,通过在编码双特异性T细胞接合分子(BiTE)的供体修复模板存在下,在PD-1基因中引入DSB使本文所涵盖的基因组编辑过的免疫效应细胞更强效。BiTE分子是由两部分组成的分子,其包含结合靶抗原的第一结合结构域、本文别处所涵盖的连接子或间隔子以及结合免疫效应细胞上的刺激或共刺激分子的第二结合结构域。第一结合结构域和第二结合结构域可以独立地选自配体、受体、抗体或其抗原结合片段、凝集素以及碳水化合物。

[0494] 在特定实施例中,第一结合结构域和第二结合结构域是抗原结合结构域。

[0495] 在特定实施例中,第一结合结构域和第二结合结构域是抗体或其抗原结合片段。在一个实施例中,第一结合结构域和第二结合结构域是单链可变片段(ScFv)。

[0496] 在特定实施例中,可以被第一结合结构域识别和结合的靶抗原的示例性实例包括但不限于: α 叶酸受体、5T4、 α v β 6整合素、BCMA、B7-H3、B7-H6、CAIX、CD16、CD19、CD20、CD22、CD30、CD33、CD44、CD44v6、CD44v7/8、CD70、CD79a、CD79b、CD123、CD138、CD171、CEA、CSPG4、EGFR、包括ErbB2 (HER2) 的EGFR家族、EGFRvIII、EGP2、EGP40、EPCAM、EphA2、EpCAM、FAP、胎儿AchR、FR α 、GD2、GD3、磷脂酰肌醇蛋白聚糖-3 (GPC3)、HLA-A1+MAGE1、HLA-A2+MAGE1、HLA-A3+MAGE1、HLA-A1+NY-ESO-1、HLA-A2+NY-ESO-1、HLA-A3+NY-ESO-1、IL-11R α 、IL-13R α 2、Lambda、Lewis-Y、Kappa、间皮素、Muc1、Muc16、NCAM、NKG2D配体、NY-ESO-1、PRAME、PSCA、PSMA、ROR1、SSX、存活素、TAG72、TEM、VEGFR2以及WT-1。

[0497] 靶抗原的其它示例性实施例包括MHC-肽复合物,任选地其中所述肽是由以下各物加工得到: α 叶酸受体、5T4、 α v β 6整合素、BCMA、B7-H3、B7-H6、CAIX、CD16、CD19、CD20、CD22、CD30、CD33、CD44、CD44v6、CD44v7/8、CD70、CD79a、CD79b、CD123、CD138、CD171、CEA、CSPG4、EGFR、包括ErbB2 (HER2) 的EGFR家族、EGFRvIII、EGP2、EGP40、EPCAM、EphA2、EpCAM、FAP、胎儿AchR、FR α 、GD2、GD3、磷脂酰肌醇蛋白聚糖-3 (GPC3)、MAGE1、NY-ESO-1、IL-11R α 、IL-13R α 2、Lambda、Lewis-Y、Kappa、间皮素、Muc1、Muc16、NCAM、NKG2D配体、PRAME、PSCA、PSMA、ROR1、SSX、存活素、TAG72、TEM、VEGFR2以及WT-1。

[0498] 在特定实施例中,免疫效应细胞上被第二结合结构域识别并结合的刺激或共刺激分子的示例性实例包括但不限于:CD3 γ 、CD3 δ 、CD3 ϵ 、CD3 ζ 、CD28、CD134、CD137以及CD278。

[0499] 在特定实施例中,在PD-1基因中通过工程改造的核酸酶诱导DSB,并且包含BiTE的供体修复模板被引入细胞中并通过同源重组插入PD-1基因中。

[0500] b. 免疫增强因子

[0501] 在特定实施例中,通过在基因组编辑过的细胞或处于肿瘤微环境中的细胞中增加免疫增强因子,使本文所涵盖的基因组编辑过的免疫效应细胞更强效。免疫增强因子是指增强免疫效应细胞中的免疫反应的特定细胞因子、趋化因子、细胞毒素和细胞因子受体。在一个实施例中,通过在编码细胞因子、趋化因子、细胞毒素或细胞因子受体的供体修复模板存在下,在PD-1基因中引入DSB,对T细胞进行工程改造。

[0502] 在特定实施例中,供体修复模板编码选自由以下组成的组的细胞因子:IL-2、胰岛素、IFN- γ 、IL-7、IL-21、IL-10、IL-12、IL-15以及TNF- α 。

[0503] 在一个优选实施例中,供体修复模板编码选自由以下组成的组的细胞因子:IL-2、IL-7、IL-12、IL-15、IL-18以及IL-21,当供体修复模板整合于PD-1基因中的靶位点处时,将所述细胞因子可操作地连接至内源性PD-1启动子,由此使细胞因子的转录控制处于内源性PD-1启动子的控制下。

[0504] 在另一个优选实施例中,供体修复模板编码IL-12,当供体修复模板整合于PD-1基因中的靶位点处时,将细胞因子可操作地连接至内源性PD-1启动子,由此使细胞因子的转录控制处于内源性PD-1启动子的控制下。

[0505] 在特定实施例中,供体修复模板编码选自由以下组成的组的趋化因子:MIP-1 α 、MIP-1 β 、MCP-1、MCP-3和RANTES。

[0506] 在特定实施例中,供体修复模板编码选自由以下组成的组的细胞毒素:穿孔素(Perforin)、颗粒酶A和颗粒酶B。

[0507] 在特定实施例中,供体修复模板编码选自由以下组成的组的细胞因子受体:IL-2

受体、IL-7受体、IL-12受体、IL-15受体以及IL-21受体。

[0508] 在一个优选实施例中，供体修复模板编码选自自由以下组成的组的细胞因子受体：IL-2受体、IL-7受体、IL-12受体、IL-15受体、IL-18受体以及IL-21受体，当供体修复模板整合于PD-1基因中的靶位点处时，将细胞因子受体可操作地连接至内源性PD-1启动子，由此使细胞因子受体的转录控制处于内源性PD-1启动子的控制下。

[0509] 在另一个优选实施例中，供体修复模板编码IL-12受体，当供体修复模板整合于PD-1基因中的靶位点处时，将细胞因子受体可操作地连接至内源性PD-1启动子，由此使细胞因子受体的转录控制处于内源性PD-1启动子的控制下。

[0510] c. 翻转受体

[0511] 在另外的实施例中，供体修复模板编码翻转受体或其部分。如本文所使用，术语“翻转受体”是指将在T细胞中来自肿瘤微环境的免疫抑制信号转化成免疫刺激信号的非天然存在的多肽。在特定实施例中，PD-1翻转受体是指包含PD-1胞外结构域或配体结合结构域或其变体、跨膜结构域以及将免疫刺激信号转导至T细胞中的胞内结构域的多肽。在特定实施例中，PD-1翻转受体是指包含PD-1胞外结构域或配体结合结构域或其变体、PD-1跨膜结构域以及将免疫刺激信号转导至T细胞中的胞内结构域的多肽。在特定实施例中，PD-1翻转受体是指包含PD-1胞外结构域或配体结合结构域或其变体，以及来自蛋白质的将免疫刺激信号转导至T细胞中的跨膜结构域和胞内结构域的多肽。在某些实施例中，PD-1胞外结构域变体对PD-L1和/或PD-L2具有增加的结合亲和力。

[0512] 在一个实施例中，跨膜是自CD3 δ 、CD3 ϵ 、CD3 γ 、CD3 ζ 、CD4、CD5、CD8 α 、CD9、CD 16、CD22、CD27、CD28、CD33、CD37、CD45、CD64、CD80、CD86、CD 134、CD137、CD152、CD154、AMN以及PD-1分离。

[0513] 在一个实施例中，跨膜是自CD4、CD8 α 、CD8 β 、CD27、CD28、CD134、CD137、CD3多肽、IL-2受体、IL-7受体、IL-12受体、IL-15受体或IL-21受体分离。

[0514] 在一个实施例中，胞内结构域是自IL-2受体、IL-7受体、IL-12受体、IL-15受体或IL-21受体分离。

[0515] 在一个实施例中，供体修复模板包含PD-1翻转受体，所述翻转受体包含PD-1胞外结构域或配体结合结构域、跨膜结构域以及一个或多个细胞内共刺激信号传导结构域和/或初级信号传导结构域。跨膜和胞内结构域可以自相同蛋白质或不同蛋白质分离。

[0516] 在一个实施例中，供体修复模板包含PD-1翻转受体，所述翻转受体包含PD-1胞外结构域或配体结合结构域、PD-1跨膜结构域以及一个或多个细胞内共刺激信号传导结构域和/或初级信号传导结构域。

[0517] 2. 免疫抑制信号阻断剂

[0518] 困扰现有过继性细胞疗法的一个局限或问题是由肿瘤微环境介导的耗竭所引起的免疫效应细胞反应性不足。耗竭T细胞具有明显不同于原始T细胞、效应T细胞或记忆T细胞的特有分子特征。其被定义为细胞因子表达和效应功能降低的T细胞。

[0519] 在特定实施例中，通过减少或阻断免疫抑制因子的信号传导使本文所涵盖的基因组编辑过的免疫效应细胞对耗竭更具抗性。在一个实施例中，通过在编码免疫抑制信号阻断剂的供体修复模板存在下，在PD-1基因中引入DSB，对T细胞进行工程改造。

[0520] 如本文所使用，术语“免疫抑制信号阻断剂”是指减弱来自肿瘤微环境的免疫抑制

信号至T细胞的转导的非天然存在的多肽。在一个实施例中,免疫抑制信号阻断剂是结合免疫抑制因子的抗体或其抗原结合片段。在优选实施例中,免疫抑制信号阻断剂是指引起针对特定免疫抑制因子或信号传导路径的抑制、阻断或显性负性作用的多肽,因为所述阻断剂包含结合免疫抑制因子的胞外结构域,和任选地,跨膜结构域,以及任选地,经过修饰的胞内结构域(例如细胞内信号传导结构域)。

[0521] 在特定实施例中,胞外结构域是识别并结合免疫抑制因子的细胞外结合结构域。

[0522] 在特定实施例中,使经过修饰的胞内结构域突变以减弱或抑制免疫抑制信号。适合突变策略包括但不限于氨基酸取代、添加或缺失。适合突变还包括但不限于截短胞内结构域以移除信号传导结构域;使胞内结构域突变以移除对于信号传导基元活性极为重要的残基;以及使胞内结构域突变以阻断受体循环。在特定实施例中,胞内结构域当存在时不会转导免疫抑制信号,或具有基本上减弱的信号传导。

[0523] 因此,在一些实施例中,免疫抑制信号阻断剂充当来自肿瘤微环境的一个或多个免疫抑制因子的接收器(sink)并抑制T细胞中的相应免疫抑制信号传导路径。

[0524] 一种免疫抑制信号是由色氨酸分解代谢介导的。癌细胞中由吡啶胺2,3-双加氧酶(IDO)引起的色氨酸分解代谢使得产生犬尿氨酸,经显示,犬尿氨酸对肿瘤微环境中的T细胞具有免疫抑制作用。参见例如Platten等人(2012)《癌症研究(Cancer Res.)》72(21): 5435-40。

[0525] 在一个实施例中,供体修复模板包含具有犬尿氨酸酶活性的酶。

[0526] 适用于特定实施例中的具有犬尿氨酸酶活性的酶的示例性实例包括但不限于L-犬尿氨酸水解酶。

[0527] 在一个实施例中,供体修复模板包含编码免疫抑制信号阻断剂的一个或多个聚核苷酸,所述免疫抑制信号阻断剂减弱或阻断由免疫抑制因子介导的免疫抑制信号传导。

[0528] 特定实施例中所涵盖的作为免疫抑制信号阻断剂的靶的免疫抑制因子的示例性实例包括但不限于:程序性死亡配体1(PD-L1)、程序性死亡配体2(PD-L2)、转化生长因子 β (TGF β)、巨噬细胞集落刺激因子1(M-CSF1)、肿瘤坏死因子相关细胞凋亡诱导配体(TRAIL)、在SiSo细胞上表达的受体结合癌症抗原配体(RCAS1)、Fas配体(FasL)、CD47、白细胞介素-4(IL-4)、白细胞介素-6(IL-6)、白细胞介素-8(IL-8)、白细胞介素-10(IL-10)以及白细胞介素-13(IL-13)。

[0529] 在各种实施例中,免疫抑制信号阻断剂包含结合免疫抑制因子的抗体或其抗原结合片段。

[0530] 在各种实施例中,免疫抑制信号阻断剂包含结合免疫抑制因子的胞外结构域。

[0531] 在特定实施例中,免疫抑制信号阻断剂包含结合免疫抑制因子的胞外结构域和跨膜结构域。

[0532] 在另一个实施例中,免疫抑制信号阻断剂包含结合免疫抑制因子的胞外结构域、跨膜结构域以及经过修饰的胞内结构域,所述经过修饰的胞内结构域不转导免疫抑制信号或具有基本上减弱的转导免疫抑制信号的能力。

[0533] 如本文所使用,术语“胞外结构域”是指抗原结合结构域。在一个实施例中,胞外结构域是将来自肿瘤微环境的免疫抑制信号转导至T细胞中的免疫抑制受体的细胞外配体结合结构域。在特定实施例中,胞外结构域是指包含免疫受体酪氨酸抑制基元(ITIM)和/或免

疫受体酪氨酸转换基元 (ITSM) 的受体的细胞外配体结合结构域。

[0534] 适用于免疫抑制信号阻断剂的特定实施例中的胞外结构域的示例性实例包括但不限于抗体或其抗原结合片段,或从以下多肽分离的细胞外配体结合结构域:程序性细胞死亡蛋白质1 (PD-1)、淋巴细胞活化基因3蛋白质 (LAG-3)、T细胞免疫球蛋白结构域和粘蛋白结构域蛋白质3 (TIM3)、细胞毒性T淋巴细胞抗原-4 (CTLA-4)、b/T淋巴细胞衰减因子 (BTLA)、T细胞免疫球蛋白和免疫受体酪氨酸抑制基元结构域 (TIGIT)、转化生长因子 β 受体 II (TGF β R II)、巨噬细胞集落刺激因子1受体 (CSF1R)、白细胞介素4受体 (IL4R)、白细胞介素6受体 (IL6R)、趋化因子 (C-X-C基元) 受体1 (CXCR1)、趋化因子 (C-X-C基元) 受体2 (CXCR2)、白细胞介素10受体亚单元 α (IL10R)、白细胞介素13受体亚单元 α 2 (IL13R α 2)、肿瘤坏死因子相关细胞凋亡诱导配体 (TRAILR1)、在SiSo细胞上表达的受体结合癌症抗原 (RCAS1R) 以及 Fas细胞表面死亡受体 (FAS)。

[0535] 在一个实施例中,胞外结构域包含选自以下组成的组的受体的细胞外配体结合结构域:PD-1、LAG-3、TIM3、CTLA-4、IL10R、TIGIT、CSF1R以及TGF β R II。

[0536] 在特定实施例中,可以使用多个跨膜结构域。特定实施例中所涵盖的适用于免疫抑制信号阻断剂的特定实施例中的跨膜结构域的示例性实例包括但不限于以下蛋白质的跨膜结构域:T细胞受体的 α 或 β 链、CD δ 、CD3 ϵ 、CD γ 、CD3 ζ 、CD4、CD5、CD8 α 、CD9、CD 16、CD22、CD27、CD28、CD33、CD37、CD45、CD64、CD80、CD86、CD134、CD137、CD152、CD154以及PD-1。

[0537] 在特定实施例中,本文所涵盖的过继性细胞疗法包含抑制或阻断来自肿瘤微环境的免疫抑制TGF β 信号通过TGF β R II转导的免疫抑制信号阻断剂。在一个实施例中,免疫抑制信号阻断剂包括含TGF β R II细胞外配体结合的胞外结构域、TGF β R II跨膜结构域以及截短的非功能性TGF β R II胞内结构域。在另一个实施例中,免疫抑制信号阻断剂包括含TGF β R II细胞外配体结合的胞外结构域、TGF β R II跨膜结构域并且不含胞内结构域。

[0538] 3. 工程改造的抗原受体

[0539] 在特定实施例中,本文所涵盖的基因组编辑过的免疫效应细胞包含工程改造的抗原受体。在一个实施例中,通过在编码工程改造的抗原受体的供体修复模板存在下,在一个或多个PD-1基因中引入DSB,对T细胞进行工程改造。

[0540] 在特定实施例中,工程改造的抗原受体是工程改造的T细胞受体 (TCR)、嵌合抗原受体 (CAR)、Daric受体或其组分,或嵌合细胞因子受体。

[0541] a. 工程改造的TCR

[0542] 在特定实施例中,本文所涵盖的基因组编辑过的免疫效应细胞包含工程改造的TCR。在一个实施例中,通过在编码工程改造的TCR的供体修复模板存在下,在一个或多个PD-1基因中引入DSB,对T细胞进行工程改造。在一个特定实施例中,工程改造的TCR被插入单个PD-1基因中的DSB处。在另一个实施例中,工程改造的TCR的 α 链被插入一个PD-1基因中的DSB中,并且工程改造的TCR的 β 链被插入另一个PD-1基因中的DSB中。

[0543] 在一个实施例中,本文所涵盖的工程改造的T细胞包含未插入PD-1基因处的工程改造的TCR,并且以下一种或多种被插入一个或多个PD-1基因中的DSB中:免疫抑制信号阻断剂、翻转受体、工程改造的T细胞受体 (TCR) 的 α 链和/或 β 链、嵌合抗原受体 (CAR)、Daric受体或其组分、或嵌合细胞因子受体。

[0544] 天然存在的T细胞受体包含二个亚单元,即 α 链和 β 链亚单元,其各自是由每个T细

胞基因组中的重组事件所产生的特有蛋白质。可以关于其针对特定靶抗原的选择性筛选TCR文库。通过这种方式,可以选出对靶抗原具有高亲合力和反应性的天然TCR,克隆且随后引入T细胞群中用于过继性免疫疗法。

[0545] 在一个实施例中,通过在一个或多个PD-1基因中的DSB处引入供体修复模板对T细胞进行修饰,所述供体修复模板包含编码TCR亚单元的聚核苷酸,其中所述TCR亚单元能够形成TCR以使T细胞对表达靶抗原的肿瘤细胞具有特异性。在特定实施例中,相较于天然存在的亚单元,所述亚单元具有一个或多个氨基酸取代、缺失、插入或修饰,只要所述亚单元能够形成TCR并且使经过转染的T细胞能够将靶细胞归巢并参与免疫相关细胞因子信号传导。工程改造的TCR优选地还以高亲合力结合展示相关肿瘤相关肽的靶细胞,并且任选地介导对在体内呈现相关肽的靶细胞的高效杀灭作用。

[0546] 编码工程改造的TCR的核酸优选地是自其在T细胞的(天然存在的)染色体中的天然环境分离,并且可以被并入适合载体(如本文别处所描述)中。在特定实施例中,所述核酸和包含所述核酸的载体都可以被转移至细胞,优选地T细胞中。接着,经过修饰的T细胞能够表达由一个或多个转导的核酸所编码的TCR的一条或多条链。在优选实施例中,由于工程改造的TCR被引入通常不表达特定TCR的T细胞中,故所述工程改造的TCR是外源TCR。工程改造的TCR的重要方面是它对主要组织相容性复合体(MHC)或类似免疫组分所呈现的肿瘤抗原具有高亲合力。相对于工程改造的TCR,CAR被工程改造成以不依赖于MHC的方式结合靶抗原。

[0547] TCR可以被表达成具有另外的多肽附接至本发明TCR的 α 链或 β 链的氨基末端或羧基末端部分,只要所附接的另外的多肽不干扰所述 α 链或 β 链形成功能性T细胞受体的能力以及MHC依赖性抗原识别即可。

[0548] 特定实施例中所涵盖的工程改造的TCR所识别的抗原包括但不限于癌症抗原,包括在血液癌和实体肿瘤上的抗原。示例性抗原包括但不限于 α 叶酸受体、 α 叶酸受体、5T4、 α v β 6整合素、BCMA、B7-H3、B7-H6、CAIX、CD19、CD20、CD22、CD30、CD33、CD44、CD44v6、CD44v7/8、CD70、CD79a、CD79b、CD123、CD138、CD171、CEA、CSPG4、EGFR、包括ErbB2(HER2)的EGFR家族、EGFRvIII、EGP2、EGP40、EPCAM、EphA2、EpCAM、FAP、胎儿AChR、FR α 、GD2、GD3、磷脂酰肌醇蛋白聚糖-3(GPC3)、HLA-A1+MAGE1、HLA-A2+MAGE1、HLA-A3+MAGE1、HLA-A1+NY-ESO-1、HLA-A2+NY-ESO-1、HLA-A3+NY-ESO-1、IL-11R α 、IL-13R α 2、Lambda、Lewis-Y、Kappa、间皮素、Muc1、Muc16、NCAM、NKG2D配体、NY-ESO-1、PRAME、PSCA、PSMA、ROR1、SSX、存活素、TAG72、TEM、VEGFR2以及WT-1。

[0549] 在一个实施例中,供体修复模板包含编码RNA聚合酶II启动子或第一自裂解病毒肽的聚核苷酸以及编码整合至一个修饰的和/或非功能性PD-1基因中的工程改造的TCR的 α 链和/或 β 链的聚核苷酸。

[0550] 在一个实施例中,供体修复模板包含编码RNA聚合酶II启动子或第一自裂解病毒肽的聚核苷酸以及编码整合至一个修饰的和/或非功能性PD-1基因中的工程改造的TCR的 α 链和 β 链的聚核苷酸。

[0551] 在一个特定实施例中,供体修复模板自5'至3'包含编码第一自裂解病毒肽的聚核苷酸、编码工程改造的TCR的 α 链的聚核苷酸、编码第二自裂解病毒肽的聚核苷酸以及编码整合至一个修饰的和/或非功能性PD-1基因中的工程改造的TCR的 β 链的聚核苷酸。在这一

情况下,另一个PD-1基因可以具有功能,或可因DSB而具有减弱的功能或呈现非功能性并通过NHEJ修复。在一个实施例中,另一个PD-1基因已通过本文所涵盖的工程改造的核酸酶进行修饰并且可具有减弱的功能或呈现非功能性。

[0552] 在某一实施例中,两个PD-1基因都经过修饰并具有减弱的功能或为非功能的:第一个经过修饰的PD-1基因包含一种核酸,所述核酸包含编码第一自裂解病毒肽的聚核苷酸和编码工程改造的TCR的 α 链的聚核苷酸;并且第二个经过修饰的PD-1基因包含编码第二自裂解病毒肽的聚核苷酸和编码工程改造的TCR的 β 链的聚核苷酸。

[0553] b. 嵌合抗原受体 (CAR)

[0554] 在特定实施例中,本文所涵盖的工程改造的免疫效应细胞包含一个或多个嵌合抗原受体 (CAR)。在一个实施例中,通过在编码CAR的供体修复模板存在下,在一个或多个PD-1基因中引入DSB,对T细胞进行工程改造。在一个特定实施例中,CAR被插入单个PD-1基因中的DSB处。

[0555] 在一个实施例中,本文所涵盖的工程改造的T细胞包含未插入PD-1基因处的CAR,并且以下一种或多种被插入一个或多个PD-1基因中的DSB中:免疫抑制信号阻断剂、翻转受体、工程改造的T细胞受体 (TCR) 的 α 链和/或 β 链、嵌合抗原受体 (CAR)、DARIC受体或其组分、或嵌合细胞因子受体。

[0556] 在各种实施例中,基因组编辑过的T细胞表达将细胞毒性重新引导至肿瘤细胞的CAR。CAR是将抗体针对靶抗原(例如肿瘤抗原)的特异性与T细胞受体活化细胞内结构域相组合以产生展现特异性抗肿瘤细胞免疫活性的嵌合蛋白的分子。如本文所使用,术语“嵌合”描述由来自不同来源的不同蛋白质或DNA部分构成。

[0557] 在各种实施例中,CAR包含结合至特异性靶抗原(又称为结合结构域或抗原特异性结合域)的细胞外结构域、跨膜结构域和细胞内信号传导结构域。CAR的主要特征是其能够重新引导免疫效应细胞特异性,由此触发增殖、细胞因子产生、吞噬作用,或能以不依赖主要组织相容性(MHC)的方式,利用单克隆抗体、可溶性配体或细胞特异性辅助受体的细胞特异性靶向能力介导靶抗原表达细胞的细胞死亡的分子的产生。

[0558] 在特定实施例中,CAR包含特异性结合至肿瘤细胞上表达的靶多肽,例如靶抗原的细胞外结合结构域。如本文所使用,术语“结合结构域”、“细胞外结构域”、“细胞外结合结构域”、“抗原结合结构域”、“抗原特异性结合结构域”以及“细胞外抗原特异性结合结构域”可互换使用并提供能够特异性结合至目的靶抗原的嵌合受体,例如CAR或DARIC。结合结构域可以包含能够特异性识别并结合至生物分子(例如细胞表面受体或肿瘤蛋白质、脂质、多糖或其它细胞表面靶分子,或其组分)的任何蛋白质、多肽、寡肽或肽。结合结构域包括目的生物分子的任何天然存在的、合成的、半合成的或以重组方式产生的结合搭配物。

[0559] 在特定实施例中,细胞外结合结构域包含抗体或其抗原结合片段。

[0560] “抗体”是指一种结合剂,即特异性地识别并结合靶抗原如肽、脂质、多糖或含有抗原决定子(如免疫细胞所识别的那些)的核酸的表位的多肽,其包含至少一个轻链或重链免疫球蛋白可变区。抗体包括抗原结合片段,例如骆驼Ig(骆驼抗体或其VHH片段)、Ig NAR、Fab片段、Fab'片段、F(ab)'₂片段、F(ab)'₃片段、Fv、单链Fv抗体(“scFv”)、双scFv、(scFv)₂、微型抗体、双功能抗体、三功能抗体、四功能抗体、二硫键稳定的Fv蛋白质(“dsFv”)以及单结构域抗体(sdAb、纳米抗体(Nanobody))或它们的其它抗体片段。所述术语还包括基因

工程改造的形式如嵌合抗体(例如人源化鼠类抗体)、异偶联物抗体(如双特异性抗体)和其抗原结合片段。还参见《皮尔斯目录与手册(Pierce Catalog and Handbook)》,1994-1995(Pierce Chemical Co.,伊利诺伊州罗克福德(Rockford, IL));Kuby, J.,《免疫学(Immunology)》,第3版,W.H.Freeman&Co.,纽约,1997。

[0561] 在一个优选实施例中,所述结合结构域是scFv。

[0562] 在另一个优选实施例中,所述结合结构域是骆驼抗体。

[0563] 在特定实施例中,CAR包含结合选自以下组成的组的抗原的细胞外结构域: α 叶酸受体、5T4、 α v β 6整合素、BCMA、B7-H3、B7-H6、CAIX、CD16、CD19、CD20、CD22、CD30、CD33、CD44、CD44v6、CD44v7/8、CD70、CD79a、CD79b、CD123、CD138、CD171、CEA、CSPG4、EGFR、包括ErbB2(HER2)的EGFR家族、EGFRvIII、EGP2、EGP40、EPCAM、EphA2、EpCAM、FAP、胎儿AChR、FR α 、GD2、GD3、磷脂酰肌醇蛋白聚糖-3(GPC3)、HLA-A1+MAGE1、HLA-A2+MAGE1、HLA-A3+MAGE1、HLA-A1+NY-ESO-1、HLA-A2+NY-ESO-1、HLA-A3+NY-ESO-1、IL-11R α 、IL-13R α 2、Lambda、Lewis-Y、Kappa、间皮素、Muc1、Muc16、NCAM、NKG2D配体、NY-ESO-1、PRAME、PSCA、PSMA、ROR1、SSX、存活素、TAG72、TEM、VEGFR2以及WT-1。

[0564] 在特定实施例中,CAR包含细胞外结合结构域,例如结合抗原的抗体或其抗原结合片段,其中所述抗原是MHC-肽复合物,如I类MHC-肽复合物或II类MHC-肽复合物。

[0565] 在某些实施例中,CAR在各个结构域之间包含连接子残基。“可变区连接序列”是将重链可变区连接至轻链可变区的氨基酸序列,其提供与所述二个结合结构域的相互作用相容的间隔子功能,使得由此得到的多肽保持针对与包含相同轻链和重链可变区的抗体相同的靶分子的特异性结合亲和力。在特定实施例中,CAR包含一个、二个、三个、四个或五个或更多个连接子。在特定实施例中,连接子的长度是约1至约25个氨基酸、约5至约20个氨基酸,或约10至约20个氨基酸,或任何中间长度的氨基酸。在一些实施例中,连接子是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个或更多个氨基酸长。

[0566] 在特定实施例中,在CAR的结合结构域之后是一个或多个“间隔子结构域”,间隔子结构域是指移动抗原结合结构域远离效应细胞表面以便能实现适当细胞/细胞接触、抗原结合和活化的区域(Patel等人,《基因疗法(Gene Therapy)》,1999;6:412-419)。间隔子结构域可来源于天然、合成、半合成或重组来源。在某些实施例中,间隔子结构域是免疫球蛋白的一部分,包括但不限于一个或多个重链恒定区,例如CH2和CH3。间隔子结构域可以包括天然存在的免疫球蛋白铰链区或改变的免疫球蛋白铰链区的氨基酸序列。

[0567] 在一个实施例中,间隔子结构域包含IgG1、IgG4或IgD的CH2和CH3。

[0568] 在一个实施例中,CAR的结合结构域被连接至一个或多个“铰链结构域”,铰链结构域在定位抗原结合结构域远离效应细胞表面以便能实现适当细胞/细胞接触、抗原结合和活化中起到作用。CAR一般在结合结构域与跨膜结构域(TM)之间包含一个或多个铰链结构域。铰链结构域可以来源于天然、合成、半合成或重组来源。铰链结构域可以包括天然存在的免疫球蛋白铰链区或改变的免疫球蛋白铰链区的氨基酸序列。

[0569] 适用于本文所描述的CAR中的示例性铰链结构域包括来源于I型膜蛋白如CD8 α 和CD4的细胞外区域的铰链区,其可以是来自这些分子的野生型铰链区或可以被改变。在另一个实施例中,铰链结构域包含CD8 α 铰链区。

[0570] 在一个实施例中, 铰链是PD-1铰链或CD152铰链。

[0571] “跨膜结构域”是CAR中融合细胞外结合部分和细胞内信号传导结构域并将CAR锚定至免疫效应细胞的质膜的部分。TM结构域可以来源于天然、合成、半合成或重组来源。

[0572] 示例性TM结构域可以来源于(即, 至少包含) T细胞受体 α 或 β 链、CD3 δ 、CD3 ϵ 、CD3 γ 、CD3 ζ 、CD4、CD5、CD8 α 、CD9、CD 16、CD22、CD27、CD28、CD33、CD37、CD45、CD64、CD80、CD86、CD134、CD137、CD152、CD154、AMN以及PD-1的跨膜区。

[0573] 在一个实施例中, CAR包含来源于CD8 α 的TM结构域。在另一个实施例中, 本文所涵盖的CAR包含来源于CD8 α 的TM结构域以及长度优选地在1、2、3、4、5、6、7、8、9或10个氨基酸之间的短寡肽或多肽连接子, 所述连接子连接CAR的TM结构域和细胞内信号传导结构域。甘氨酸-丝氨酸连接子提供特别适合的连接子。

[0574] 在特定实施例中, CAR包含细胞内信号传导结构域。“细胞内信号传导结构域”是指CAR中参与将有效CAR结合至靶抗原的信息转导至免疫效应细胞内部以引起效应细胞功能的部分, 所述效应细胞功能例如是活化、细胞因子产生、增殖和细胞毒性活性, 包括细胞毒性因子释放至结合CAR的靶细胞中, 或由抗原结合至细胞外CAR结构域引起的其它细胞反应。

[0575] 术语“效应功能”是指细胞的特有功能。T细胞的效应功能例如可以是细胞溶解活性或帮助或活性, 包括细胞因子的分泌。因此, 术语“细胞内信号传导结构域”是指一种蛋白质中转导效应功能信号并引导细胞执行特有功能的部分。尽管通常可采用整个细胞内信号传导结构域, 但在许多情况下不必使用整个结构域。就使用细胞内信号传导结构域的截短部分来说, 可使用这类截短部分代替整个结构域, 只要其转导效应功能信号即可。术语细胞内信号传导结构域意图包括足以转导效应功能信号的细胞内信号传导结构域的任何截短部分。

[0576] 众所周知, 仅由TCR产生的信号不足以完全活化T细胞并且还需要二次或共刺激信号。因此, T细胞活化可以被认为是由二个不同类别的细胞内信号传导结构域介导: 通过TCR(例如TCR/CD3复合物) 起始抗原依赖性初级活化的初级信号传导结构域, 以及以不依赖抗原的方式起作用以提供二次或共刺激信号的共刺激信号传导结构域。在优选实施例中, CAR包含细胞内信号传导结构域, 所述细胞内信号传导结构域包含一个或多个“共刺激信号传导结构域”和一个“初级信号传导结构域”。

[0577] 初级信号传导结构域以刺激性方式或以抑制性方式调控TCR复合物的初级活化。以刺激性方式起作用的初级信号传导结构域可含有信号传导基元, 所述信号传导基元称为免疫受体酪氨酸活化基元或ITAM。

[0578] 适用于特定实施例中所涵盖的CAR中的含ITAM的初级信号传导结构域的示例性实例包括来源于FcR γ 、FcR β 、CD3 γ 、CD3 δ 、CD3 ϵ 、CD3 ζ 、CD22、CD79a、CD79b以及CD66d的那些。在特别优选的实施例中, CAR包含CD3 ζ 初级信号传导结构域和一个或多个共刺激信号传导结构域。细胞内初级信号传导结构域和共刺激信号传导结构域可以按任何次序串联连接至跨膜结构域的羧基末端。

[0579] 在特定实施例中, CAR包含一个或多个共刺激信号传导结构域, 用以增进表达CAR受体的T细胞的功效和扩增。如本文所使用, 术语“共刺激信号传导结构域”或“共刺激结构域”是指共刺激分子的细胞内信号传导结构域。

[0580] 适用于特定实施例中所涵盖的CAR中的此类共刺激分子的示例性分子包括TLR1、TLR2、TLR3、TLR4、TLR5、TLR6、TLR7、TLR8、TLR9、TLR10、CARD11、CD2、CD7、CD27、CD28、CD30、CD40、CD54 (ICAM)、CD83、CD134 (OX40)、CD137 (4-1BB)、CD278 (ICOS)、DAP10、LAT、NKD2C、SLP76、TRIM以及ZAP70。在一个实施例中，CAR包含一个或多个选自由CD28、CD137和CD134组成的组的共刺激信号传导结构域，以及一个CD3 ζ 初级信号传导结构域。

[0581] 在各种实施例中，CAR包含：结合选自由以下组成的组的抗原的细胞外结构域：BCMA、CD19、CSPG4、PSCA、ROR1和TAG72；自选自由以下组成的组的多肽分离的跨膜结构域：CD4、CD8 α 、CD154和PD-1；自选自由以下组成的组的多肽分离的一个或多个细胞内共刺激信号传导结构域：CD28、CD134和CD137；以及自选自由以下组成的组的多肽分离的信号传导结构域：FcR γ 、FcR β 、CD3 γ 、CD3 δ 、CD3 ϵ 、CD3 ζ 、CD22、CD79a、CD79b和CD66d。

[0582] c. Daric受体

[0583] 在特定实施例中，工程改造的免疫效应细胞包含一个或多个Daric受体。如本文所使用，术语“Daric受体”是指多链工程改造的抗原受体。在一个实施例中，通过在编码Daric的一种或多种组分的供体修复模板存在下，在一个或多个PD-1基因中引入DSB，对T细胞进行工程改造。在一个特定实施例中，Daric或其一种或多种组分被插入单个PD-1基因中的DSB处。

[0584] 在一个实施例中，工程改造的T细胞包含未插入PD-1基因处的Daric，并且以下一种或多种被插入一个或多个PD-1基因中的DSB中：免疫抑制信号阻断剂、翻转受体工程改造的T细胞受体 (TCR) 的 α 链和/或 β 链、嵌合抗原受体 (CAR)、或Daric受体或其组分。

[0585] Daric架构和组分的示例性实例公开于PCT公开第W02015/017214号和美国专利公开第20150266973号中，其各自以全文引用的方式并入此处。

[0586] 在一个实施例中，供体修复模板包含以下Daric组分：信号传导多肽，其包含第一多聚化结构域、第一跨膜结构域和一个或多个细胞内共刺激信号传导结构域和/或初级信号传导结构域；以及结合多肽，其包含结合结构域、第二多聚化结构域和任选地第二跨膜结构域。功能性Daric包含促进在细胞表面上形成Daric受体复合物的桥接因子，并且所述桥接因子与信号传导多肽和结合多肽的多聚化结构域缔合并安置在信号传导多肽和结合多肽的多聚化结构域之间。

[0587] 在特定实施例中，第一多聚化结构域和第二多聚化结构域与选自由以下组成的组的桥接因子缔合：雷帕霉素 (rapamycin) 或其雷帕霉素类似物 (rapalog)、库马霉素 (coumermycin) 或其衍生物、赤霉素 (gibberellin) 或其衍生物、脱落酸 (ABA) 或其衍生物、甲氨蝶呤 (methotrexate) 或其衍生物、环孢菌素A (cyclosporin A) 或其衍生物、FKCsA或其衍生物、FKBP的甲氧苄氨嘧啶 (trimethoprim) (Tnp) 合成配体 (SLF) 或其衍生物，及其任何组合。

[0588] 雷帕霉素类似物的示例性实例包括美国专利第6,649,595号中所公开的那些，所述雷帕霉素类似物的结构以全文引用的方式并入本文中。在某些实施例中，桥接因子是相较于雷帕霉素，具有明显较低的免疫抑制作用的雷帕霉素类似物。“明显较低的免疫抑制作用”是指如在临床上或在具有人免疫抑制活性的适当体外 (例如抑制T细胞增殖) 或体内替代物中所测量，雷帕霉素类似物的免疫抑制作用比关于等摩尔量雷帕霉素所观测或预期的免疫抑制作用的低至少0.1至0.005倍。在一个实施例中，“明显较低的免疫抑制作用”是指

雷帕霉素类似物在此类体外分析中的EC50值比在同一分析中关于雷帕霉素所观测到的EC50值要高至少10至250倍。

[0589] 雷帕霉素类似物的其它示例性实例包括但不限于依维莫司 (everolimus)、诺瓦莫司 (novolimus)、吡美莫司 (pimecrolimus)、地磷莫司 (ridaforolimus)、他克莫司 (tacrolimus)、坦罗莫司 (temsirolimus)、优米莫司 (umirolium) 以及佐他莫司 (zotarolimus)。

[0590] 在某些实施例中,多聚化结构域将与桥接因子缔合,所述桥接因子是雷帕霉素或其雷帕霉素类似物。例如,第一多聚化结构域和第二多聚化结构域是选自FKBP和FRB的一对。FRB结构域是能够与FKBP蛋白质和雷帕霉素或其雷帕霉素类似物形成三联复合物的多肽区域(蛋白质“结构域”)。FRB结构域存在于多种天然存在的蛋白质中,包括来自人类和其它物种的mTOR蛋白质(在文献中又称为FRAP、RAP1或RAFT);酵母蛋白质,包括Tor1和Tor2;以及念珠菌属FRAP同源物。有关这些蛋白质的核苷酸序列、克隆以及其它方面的信息是本领域中已知的。例如,人mTOR的蛋白质序列登录号是GenBank登录号L34075.1 (Brown等人,《自然 (Nature)》369:756,1994)。

[0591] 适用于本文所涵盖的特定实施例中的FRB结构域一般含有至少约85个至约100个氨基酸残基。在某些实施例中,根据GenBank登录号L34075.1的氨基酸序列,用于本公开的融合蛋白中的FRB氨基酸序列将包含93个氨基酸的序列Ile-2021至Lys-2113和T2098L突变。用于特定实施例中所涵盖的Darics中的FRB结构域将能够与FKBP蛋白质结合至雷帕霉素或其雷帕霉素类似物的复合物结合。在某些实施例中,FRB结构域的肽序列包含(a)至少跨越人mTOR中所示93个氨基酸的区域或同源蛋白质的相应区域的天然存在的肽序列;(b)天然存在的FRB的变体,其中天然存在的肽的至多约十个氨基酸、或约1至约5个氨基酸、或约1至约3个氨基酸,或在一些实施例中仅一个氨基酸已经缺失、插入或被取代;或(c)由能够与编码天然存在的FRB结构域的DNA分子选择性杂交的核酸分子或由因遗传密码的简并性而能够与编码天然存在的FRB结构域的DNA分子选择性杂交的DNA序列所编码的肽。

[0592] FKBP (FK506结合蛋白) 是针对大环内酯如FK506、FK520和雷帕霉素的胞质受体,并且在各物种界间高度保守。FKBP是能够结合至雷帕霉素或其雷帕霉素类似物并且进一步与含FRB的蛋白质或融合蛋白形成三联复合物的蛋白质或蛋白质结构域。FKBP结构域又可称为“雷帕霉素结合结构域”。有关各种FKBP物种的核苷酸序列、克隆和其它方面的信息是本领域中已知的(参见例如Staendart等人,《自然》346:671,1990 (人FKBP12);Kay,《生物化学杂志 (Biochem. J.)》314:361,1996)。在其它哺乳动物物种中、酵母中以及在其它生物体中的同源FKBP蛋白质也是本领域中已知的并且可以用于本文所公开的融合蛋白中。特定实施例中所涵盖的FKBP结构域将能够结合至雷帕霉素或其雷帕霉素类似物并且参加与含FRB蛋白质的三联复合物(可以通过用于检测此类结合的任何手段直接或间接地测定)。

[0593] 适用于特定实施例中所涵盖的Daric中的FKBP结构域的示例性实例包括但不限于:天然存在的FKBP肽序列,优选地自人FKBP12蛋白质 (GenBank登录号AAA58476.1) 分离的FKBP肽序列或由其、由另一人FKBP、由鼠类或其它哺乳动物FKBP或由某一其他动物、酵母或真菌FKBP分离的肽序列;天然存在的FKBP序列的变体,其中天然存在的肽中至多约十个氨基酸、或约1至约5个氨基酸或约1至约3个氨基酸,或在一些实施例中仅一个氨基酸已经缺失、插入或被取代;或由能够与编码天然存在的FKBP的DNA分子选择性杂交的核酸分子或由

因遗传密码的简并性而能够与编码天然存在的FKBP的DNA分子选择性杂交的DNA序列所编码的肽序列。

[0594] 适用于特定实施例中所涵盖的Daric中的多聚化结构域对的其它示例性实例包括但不限于包括来自FKBP与FRB、FKBP与钙调神经磷酸酶、FKBP与亲环蛋白、FKBP与细菌DHFR、钙调神经磷酸酶与亲环蛋白、PYL1与ABI1、或GIB1与GAI,或其变体。

[0595] 在又其它实施例中,抗桥接因子阻断信号传导多肽和结合多肽与桥接因子的缔合。例如,环孢菌素或FK506可用作抗桥接因子以滴定去除雷帕霉素并因此,由于仅结合一个多聚化结构域而停止信号传导。在某些实施例中,抗桥接因子(例如环孢菌素、FK506)是一种免疫抑制剂。例如,可以使用免疫抑制性抗桥接因子使特定实施例中所涵盖的Daric组分的功能阻断或降到最低并同时在临床环境中抑制或阻断不合需要的或病理性炎性反应。

[0596] 在一个实施例中,第一多聚化结构域包含FRB T2098L,第二多聚化结构域包含FKBP12,并且桥接因子是雷帕霉素类似物AP21967。

[0597] 在另一个实施例中,第一多聚化结构域包含FRB,第二多聚化结构域包含FKBP12,并且桥接因子是雷帕霉素、坦罗莫司或依维莫司。

[0598] 在特定实施例中,信号传导多肽第一跨膜结构域并且结合多肽包含第二跨膜结构域或GPI锚。第一跨膜结构域和第二跨膜结构域的示例性实例是从独立地选自以下组成的组的多肽分离:CD3 δ 、CD3 ϵ 、CD3 γ 、CD3 ζ 、CD4、CD5、CD8 α 、CD9、CD 16、CD22、CD27、CD28、CD33、CD37、CD45、CD64、CD80、CD86、CD 134、CD137、CD152、CD154、AMN以及PD-1。

[0599] 在一个实施例中,信号传导多肽包含一个或多个细胞内共刺激信号传导结构域和/或初级信号传导结构域。

[0600] 适用于特定实施例中所涵盖的Daric信号传导组分中的初级信号传导结构域的示例性实例包括来源于FcR γ 、FcR β 、CD3 γ 、CD3 δ 、CD3 ϵ 、CD3 ζ 、CD22、CD79a、CD79b以及CD66d的那些。在特别优选的实施例中,Daric信号传导组分包含CD3 ζ 初级信号传导结构域和一个或多个共刺激信号传导结构域。细胞内初级信号传导结构域和共刺激信号传导结构域可以按任何次序串联连接至跨膜结构域的羧基末端。

[0601] 适用于特定实施例中所涵盖的Daric信号传导组分中的这些共刺激分子的示例性实例包括TLR1、TLR2、TLR3、TLR4、TLR5、TLR6、TLR7、TLR8、TLR9、TLR10、CARD11、CD2、CD7、CD27、CD28、CD30、CD40、CD54 (ICAM)、CD83、CD134 (OX40)、CD137 (4-1BB)、CD278 (ICOS)、DAP10、LAT、NKD2C、SLP76、TRIM以及ZAP70。在一个实施例中,Daric信号传导组分包含选自CD28、CD137和CD134组成的组的一个或多个共刺激信号传导结构域,以及一个CD3 ζ 初级信号传导结构域。

[0602] 在特定实施例中,Daric结合组分包含结合结构域。在一个实施例中,所述结合结构域是抗体或其抗原结合片段。

[0603] 所述抗体或其抗原结合片段包含特异性地识别并结合靶抗原如肽、脂质、多糖或含有抗原决定子(如由免疫细胞识别的那些)的核酸的表位的至少一个轻链或重链免疫球蛋白可变区。抗体包括抗原结合片段,例如骆驼Ig(骆驼抗体或其VHH片段)、Ig NAR、Fab片段、Fab'片段、F(ab)'2片段、F(ab)'3片段、Fv、单链Fv抗体("scFv")、双scFv、(scFv)2、微型抗体、双功能抗体、三功能抗体、四功能抗体、二硫键稳定的Fv蛋白质("dsFv")以及单结构域抗体(sdAb、纳米抗体(Nanobody))或它们的其它抗体片段。所述术语还包括基因工程改

造的形式如嵌合抗体(例如人源化鼠类抗体)、异偶联物抗体(如双特异性抗体)和其抗原结合片段。还参见《皮尔斯目录与手册》,1994-1995(Pierce Chemical Co.,伊利诺伊州罗克福德);Kuby,J.,《免疫学》,第3版,W.H.Freeman&Co.,纽约,1997。

[0604] 在一个优选实施例中,所述结合结构域是scFv。

[0605] 在另一个优选实施例中,所述结合结构域是骆驼抗体。

[0606] 在特定实施例中,Daric结合组分包含结合选自以下组成的组的抗原的细胞外结构域: α 叶酸受体、5T4、 α v β 6整合素、BCMA、B7-H3、B7-H6、CAIX、CD16、CD19、CD20、CD22、CD30、CD33、CD44、CD44v6、CD44v7/8、CD70、CD79a、CD79b、CD123、CD138、CD171、CEA、CSPG4、EGFR、包括ErbB2(HER2)的EGFR家族、EGFRvIII、EGP2、EGP40、EPCAM、EphA2、EpCAM、FAP、胎儿AchR、FR α 、GD2、GD3、磷脂酰肌醇蛋白聚糖-3(GPC3)、HLA-A1+MAGE1、HLA-A2+MAGE1、HLA-A3+MAGE1、HLA-A1+NY-ESO-1、HLA-A2+NY-ESO-1、HLA-A3+NY-ESO-1、IL-11R α 、IL-13R α 2、Lambda、Lewis-Y、Kappa、间皮素、Muc1、Muc16、NCAM、NKG2D配体、NY-ESO-1、PRAME、PSCA、PSMA、ROR1、SSX、存活素、TAG72、TEM、VEGFR2以及WT-1。

[0607] 在一个实施例中,Daric结合组分包含细胞外结构域,例如结合MHC-肽复合物,如I类MHC-肽复合物或II类MHC-肽复合物的抗体或其抗原结合片段。

[0608] 在特定实施例中,本文所涵盖的Daric组分包含连接二种蛋白质、多肽、肽、结构域、区域或基元的连接子或间隔子。在某些实施例中,连接子包含约二个至约35个氨基酸、或约四个至约20个氨基酸、或约八个至约15个氨基酸、或约15个至约25个氨基酸。在其它实施例中,间隔子可以具有特定结构,如抗体CH2CH3结构域、铰链结构域等。在一个实施例中,间隔子包含IgG1、IgG4或IgD的CH2和CH3结构域。

[0609] 在特定实施例中,本文所涵盖的Daric组分包含一个或多个“铰链结构域”,所述铰链结构域在定位结构域以便能实现适当细胞/细胞接触、抗原结合和活化中起到作用。Daric可以在结合结构域与多聚化结构域和/或跨膜结构域(TM)之间或在多聚化结构域与跨膜结构域之间包含一个或多个铰链结构域。铰链结构域可以来源于天然、合成、半合成或重组来源。铰链结构域可以包括天然存在的免疫球蛋白铰链区或改变的免疫球蛋白铰链区的氨基酸序列。在特定实施例中,铰链是CD8 α 铰链或CD4铰链。

[0610] 在一个实施例中,Daric包含信号传导多肽,其包含FRB T2098L的第一多聚化结构域、CD8跨膜结构域、4-1BB共刺激结构域和CD3 ζ 初级信号传导结构域;结合多肽包含结合CD19的scFv、FKBP12的第二多聚化结构域和CD4跨膜结构域;并且桥接因子是雷帕霉素类似物AP21967。

[0611] 在一个实施例中,Daric包含信号传导多肽,其包含FRB的第一多聚化结构域、CD8跨膜结构域、4-1BB共刺激结构域和CD3 ζ 初级信号传导结构域;结合多肽包含结合CD19的scFv、FKBP12的第二多聚化结构域和CD4跨膜结构域;并且桥接因子是雷帕霉素、坦罗莫司或依维莫司。

[0612] d. ζ 因子

[0613] 在特定实施例中,本文所涵盖的工程改造的免疫效应细胞包含一个或多个嵌合细胞因子受体。在一个实施例中,通过在编码CAR的供体修复模板存在下,在一个或多个PD-1基因中引入DSB,对T细胞进行工程改造。在一个特定实施例中,嵌合细胞因子受体被插入单个PD-1基因中的DSB处。

[0614] 在一个实施例中,本文所涵盖的工程改造的T细胞包含未插入PD-1基因处的嵌合细胞因子受体,并且以下一种或多种被插入一个或多个PD-1基因中的DSB中:免疫抑制信号阻断剂、翻转受体、工程改造的T细胞受体(TCR)的 α 链和/或 β 链、嵌合抗原受体(CAR)、Daric受体或其组分、或嵌合细胞因子受体。

[0615] 在各种实施例中,基因组编辑过的T细胞表达将细胞毒性重新引导至肿瘤细胞的嵌合细胞因子受体。 ζ 因子是嵌合跨膜免疫受体,其包含细胞外结构域、跨膜区和细胞内信号传导结构域,所述细胞外结构域包含可溶性受体配体连接至能够将细胞外结构域系拴至细胞表面的载体区域。 ζ 因子当在T淋巴细胞的表面上表达时,将T细胞活性引导至表达可溶性受体配体特异性针对的受体的细胞。 ζ 因子嵌合免疫受体重新引导T细胞的抗原特异性,并且用于治疗多种癌症,特别是通过人恶性疾病利用的自分泌/旁分泌细胞因子系统进行治疗。

[0616] 在特定实施例中,所述嵌合细胞因子受体包含免疫抑制细胞因子或其细胞因子受体结合变体、连接子、跨膜结构域和细胞内信号传导结构域。

[0617] 在特定实施例中,所述细胞因子或其细胞因子受体结合变体选自由以下组成的组:白细胞介素-4(IL-4)、白细胞介素-6(IL-6)、白细胞介素-8(IL-8)、白细胞介素-10(IL-10)和白细胞介素-13(IL-13)。

[0618] 在某些实施例中,连接子包含CH2CH3结构域、铰链结构域等。在一个实施例中,连接子包含IgG1、IgG4或IgD的CH2和CH3结构域。在一个实施例中,连接子包含CD8 α 或CD4铰链结构域。

[0619] 在特定实施例中,跨膜结构域选自由以下组成的组:T细胞受体的 α 链或 β 链、CD3 δ 、CD3 ϵ 、CD3 γ 、CD3 ζ 、CD4、CD5、CD8 α 、CD9、CD 16、CD22、CD27、CD28、CD33、CD37、CD45、CD64、CD80、CD86、CD 134、CD137、CD152、CD154、AMN以及PD-1。

[0620] 在特定实施例中,细胞内信号传导结构域选自由以下组成的组:含ITAM的初级信号传导结构域和/或共刺激结构域。

[0621] 在特定实施例中,细胞内信号传导结构域选自由以下组成的组:FcR γ 、FcR β 、CD3 γ 、CD3 δ 、CD3 ϵ 、CD3 ζ 、CD22、CD79a、CD79b以及CD66d。

[0622] 在特定实施例中,细胞内信号传导结构域选自由以下组成的组:TLR1、TLR2、TLR3、TLR4、TLR5、TLR6、TLR7、TLR8、TLR9、TLR10、CARD11、CD2、CD7、CD27、CD28、CD30、CD40、CD54(ICAM)、CD83、CD134(OX40)、CD137(4-1BB)、CD278(ICOS)、DAP10、LAT、NKD2C、SLP76、TRIM以及ZAP70。

[0623] 在一个实施例中,嵌合细胞因子受体包含选自由CD28、CD137和CD134组成的组的一个或多个共刺激信号传导结构域,以及一个CD3 ζ 初级信号传导结构域。

[0624] F. 多肽

[0625] 本文涵盖各种多肽,包括但不限于归巢核酸内切酶变体、megaTAL和融合多肽。在优选实施例中,多肽包含SEQ ID NO:1-24和60-64中所述的氨基酸序列。除非作相反说明,否则“多肽”、“多肽片段”、“肽”和“蛋白质”可互换使用,且根据常规含义,即,定义为氨基酸序列。在一个实施例中,“多肽”包括融合多肽和其它变体。多肽可使用多种众所周知的重组和/或合成技术中的任一种制备。多肽不限于特定长度,例如其可以包含全长蛋白质序列、全长蛋白质的片段、或融合蛋白,并且可以包括多肽的翻译后修饰,例如糖基化、乙酰化、磷

酸化等,以及本领域中已知的天然存在的和非天然存在的其它修饰。

[0626] 如本文所使用,“分离的蛋白质”、“分离的肽”或“分离的多肽”等是指体外合成、从细胞环境以及从与细胞中其它组分的缔合分离和/或纯化的肽或多肽分子,即,它不与体内物质明显缔合。

[0627] 特定实施例中所涵盖的多肽的示例性实例包括但不限于归巢核酸内切酶变体、megaTAL、末端加工核酸酶、融合多肽以及其变体。

[0628] 多肽包括“多肽变体”。多肽变体与天然存在的多肽的不同之处可在于一个或多个氨基酸取代、缺失、添加和/或插入。这类变体可以是天然存在的或可以例如通过修饰上述多肽序列中的一个或多个氨基酸,以合成方式产生。例如,在特定实施例中,可能需要通过在归巢核酸内切酶、megaTAL等中引入一个或多个取代、缺失、添加和/或插入来改善所述多肽的生物特性,所述多肽结合并裂解人PD-1基因中的靶位点。在特定实施例中,多肽包括与本文所涵盖的参考序列中的任一个具有至少约65%、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%或99%氨基酸一致性的多肽,典型地其中变体保持参考序列的至少一种生物活性。

[0629] 多肽变体包括生物活性“多肽片段”。生物活性多肽片段的示例性实例包括DNA结合结构域、核酸酶结构域等。如本文所使用,术语“生物活性片段”或“最小生物活性片段”是指保留至少100%、至少90%、至少80%、至少70%、至少60%、至少50%、至少40%、至少30%、至少20%、至少10%、或至少5%的天然存在的多肽活性的多肽片段。在优选实施例中,所述生物活性是针对靶序列的结合亲和力和/或裂解活性。在某些实施例中,多肽片段可包含至少5个至约1700个氨基酸长的氨基酸链。应了解,在某些实施例中,片段是至少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、37、38、39、40、41、42、43、44、45、46、47、48、49、50、55、60、65、70、75、80、85、90、95、100、110、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1100、1200、1300、1400、1500、1600、1700个或更多个氨基酸长。在特定实施例中,多肽包含归巢核酸内切酶变体的生物活性片段。在特定实施例中,本文所述的多肽可以包含一个或多个表示为“X”的氨基酸。“X”如果存在于氨基酸SEQ ID NO中,则是指任何氨基酸。一个或多个“X”残基可以存在于本文所涵盖的特定SEQ ID NO中所述氨基酸序列的N末端和C末端。如果“X”氨基酸不存在,则SEQ ID NO中所述的其余氨基酸序列可以视为生物活性片段。

[0630] 在特定实施例中,多肽包含归巢核酸内切酶变体的生物活性片段,例如SEQ ID NO:3-14和60-63,或megaTAL (SEQ ID NO:15-23和64)。所述生物活性片段可以包含N末端截短和/或C末端截短。在一个特定实施例中,生物活性片段缺乏或包含相较于相应野生型归巢核酸内切酶序列,归巢核酸内切酶变体中1、2、3、4、5、6、7或8个N末端氨基酸的缺失,更优选相较于相应野生型归巢核酸内切酶序列,归巢核酸内切酶变体中4个N末端氨基酸的缺失。在一个特定实施例中,生物活性片段缺乏或包含相较于相应野生型归巢核酸内切酶序列,归巢核酸内切酶变体中1、2、3、4或5个C末端氨基酸的缺失,更优选相较于相应野生型归巢核酸内切酶序列,归巢核酸内切酶变体中2个C末端氨基酸的缺失。在一个特别优选的实施例中,生物活性片段缺乏或包含相较于相应野生型归巢核酸内切酶序列,归巢核酸内切

酶变体中4个N末端氨基酸和2个C末端氨基酸的缺失。

[0631] 在一个特定实施例中，I-OnuI变体包含以下N末端氨基酸中1、2、3、4、5、6、7或8个的缺失：M、A、Y、M、S、R、R、E；和/或以下1、2、3、4或5个C末端氨基酸的缺失：R、G、S、F、V。

[0632] 在一个特定实施例中，I-OnuI变体包含以下N末端氨基酸中1、2、3、4、5、6、7或8个缺失或取代：M、A、Y、M、S、R、R、E；和/或以下1、2、3、4或5个C末端氨基酸的缺失或取代：R、G、S、F、V。

[0633] 在一个特定实施例中，I-OnuI变体包含以下N末端氨基酸中1、2、3、4、5、6、7或8个的缺失：M、A、Y、M、S、R、R、E；和/或以下1或2个C末端氨基酸的缺失：F、V。

[0634] 在一个特定实施例中，I-OnuI变体包含以下N末端氨基酸中1、2、3、4、5、6、7或8个缺失或取代：M、A、Y、M、S、R、R、E；和/或以下1或2个C末端氨基酸的缺失或取代：F、V。

[0635] 如上文所述，多肽可以通过不同方式改变，包括氨基酸取代、缺失、截短和插入。用于此类操作的方法是本领域中一般已知的。例如，参考多肽的氨基酸序列变体可以通过DNA中的突变制备。用于诱变和核苷酸序列改变的方法是本领域中众所周知的。参见例如Kunkel (1985,《美国国家科学院院刊》82:488-492)；Kunkel等人, (1987,《酶学方法(Methods in Enzymol)》,154:367-382)；美国专利第4,873,192号；Watson, J.D.等人(《基因分子生物学(Molecular Biology of the Gene)》,第四版, Benjamin/Cummings, 加利福尼亚州门洛帕克(Menlo Park, Calif.), 1987)以及其中引用的参考文献。关于不影响目的蛋白质的生物活性的适当氨基酸取代的指导可见于Dayhoff等人, (1978) 蛋白序列和结构图集(Atlas of Protein Sequence and Structure) (《国家生物医学研究基础(Natl. Biomed. Res. Found.)》, 华盛顿哥伦比亚特区(Washington, D.C.))的模型中。

[0636] 在某些实施例中，变体将含有一个或多个保守性取代。“保守性取代”是氨基酸被具有类似特性的另一个氨基酸取代，使得肽化学领域的技术人员能预期多肽的二级结构和亲水性质基本上不变的取代。可以对特定实施例中所涵盖的聚核苷酸和多肽的结构进行修饰，多肽包括具有至少约的多肽且仍获得编码具有所希望特征的变体或衍生物多肽的功能性分子。当希望改变多肽的氨基酸序列以产生等效或甚至改良的变体多肽时，本领域技术人员例如可以改变编码DNA序列的一个或多个密码子，例如根据表1。

[0637] 表1-氨基酸密码子

[0638]

氨基酸	单字母密码	三字母密码	密码子			
丙氨酸	A	Ala	GCA	GCC	GCG	GCU
半胱氨酸	C	Cys	UGC	UGU		
天冬氨酸	D	Asp	GAC	GAU		
谷氨酸	E	Glu	GAA	GAG		
苯丙氨酸	F	Phe	UUC	UUU		
甘氨酸	G	Gly	GGA	GGC	GGG	GGU
组氨酸	H	His	CAC	CAU		
异亮氨酸	I	Iso	AUA	AUC	AUU	
赖氨酸	K	Lys	AAA	AAG		
亮氨酸	L	Leu	UUA	UUG	CUA	CUC
甲硫氨酸	M	Met	AUG			
天冬酰胺	N	Asn	AAC	AAU		
脯氨酸	P	Pro	CCA	CCC	CCG	CCU
谷氨酰胺	Q	Gln	CAA	CAG		
精氨酸	R	Arg	AGA	AGG	CGA	CGC
丝氨酸	S	Ser	AGC	AGU	UCA	UCC
甲硫氨酸	T	Thr	ACA	ACC	ACG	ACU
缬氨酸	V	Val	GUA	GUC	GUG	GUU
色氨酸	W	Trp	UGG			

[0639]

酪氨酸	Y	Tyr	UAC	UAU
-----	---	-----	-----	-----

[0640] 可以使用本领域中众所周知的计算机程序,如DNASTAR、DNA Strider、Geneious、Mac Vector或Vector NTI软件发现有关确定可以被取代、插入或缺失而不消除生物活性的氨基酸残基的指导。优选地,本文所公开的蛋白质变体中的氨基酸变化是保守性氨基酸变化,即,带类似电荷或不带电氨基酸的取代。保守性氨基酸变化涉及侧链相关的一组氨基酸中的一个的取代。天然存在的氨基酸一般被分成四组:酸性氨基酸(天冬氨酸、谷氨酸)、碱性氨基酸(赖氨酸、精氨酸、组氨酸)、非极性氨基酸(丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、甲硫氨酸、色氨酸)和不带电极性氨基酸(甘氨酸、天冬酰胺、谷氨酰胺、半胱氨酸、丝氨酸、苏氨酸、酪氨酸)。苯丙氨酸、色氨酸和酪氨酸有时共同归类为芳香族氨基酸。在肽或蛋白质中,适合的氨基酸保守性取代是本领域技术人员已知的且一般可以在不改变所得分子的生物活性的情况下进行。本领域技术人员应认识到,一般说来,多肽非必需区中的单一氨基酸取代基本上不会改变生物活性(参见例如Watson等人,《基因分子生物学》,第4版,1987,The Benjamin/Cummings Pub.Co.,第224页)。

[0641] 在需要表达两个或更多个多肽的一个实施例中,编码所述多肽的聚核苷酸序列可通过如本文别处所公开的IRES序列隔开。

[0642] 特定实施例中涵盖的多肽包括融合多肽。在特定实施例中,提供了融合多肽和编码融合多肽的聚核苷酸。融合多肽和融合蛋白是指具有至少二个、三个、四个、五个、六个、七个、八个、九个或十个多肽区段的多肽。

[0643] 在另一个实施例中,两个或更多个多肽可表达为融合蛋白形式,所述融合蛋白包含一个或多个如本文别处所公开的自裂解多肽序列。

[0644] 在一个实施例中,本文所涵盖的融合蛋白包含一个或多个DNA结合结构域和一种或多种核酸酶,以及一个或多个连接子和/或自裂解多肽。

[0645] 在一个实施例中,本文所涵盖的融合蛋白包含核酸酶变体;连接子或自裂解肽;以及末端加工酶,包括但不限于5'-3'核酸外切酶、5'-3'碱性核酸外切酶和3'-5'核酸外切酶(例如Trex2)。

[0646] 融合多肽可包含一个或多个多肽结构域或区段,包括但不限于信号肽、细胞穿透肽结构域(CPP)、DNA结合结构域、核酸酶结构域等、表位标签(例如麦芽糖结合蛋白("MBP")、谷胱甘肽S转移酶(GST)、HIS6、MYC、FLAG、V5、VSV-G以及HA)、多肽连接子以及多肽裂解信号。融合多肽典型地是C末端连接至N末端,不过它们还可以是C末端连接至C末端、N末端连接至N末端、或N末端连接至C末端。在特定实施例中,融合蛋白的多肽可以遵循任何次序。融合多肽或融合蛋白还可以包括保守性修饰的变体、多态性变体、等位基因、突变体、子序列和种间同系物,只要保存融合多肽所希望的活性即可。融合多肽可以通过化学合成方法或通过二个部分之间的化学键联制造,或一般可以使用其它标准技术制备。构成融合多肽的连接DNA序列可操作地连接至如本文别处所公开的适合转录或翻译控制元件。

[0647] 融合多肽可任选地包含连接子,所述连接子可以用于连接一个或多个多肽或一个多肽内的一个或多个结构域。肽连接子序列可以用于将任何两种或更多种多肽组分隔开一定距离,该距离足以确保每一多肽折叠成其适当二级结构和三级结构,从而允许多肽结构域发挥其所希望的功能。此类肽连接子序列是使用本领域中的标准技术并入融合多肽中。适合肽连接子序列可以基于以下因素选择:(1)其能够呈现柔性延长构象;(2)其无法呈现能与第一多肽和第二多肽上的功能性表位相互作用的二级结构;以及(3)缺乏可能与多肽功能性表位反应的疏水性或带电残基。优选的肽连接子序列含有Gly、Asn和Ser残基。其它接近中性的氨基酸,如Thr和Ala,也可以用于连接子序列中。可以有效地用作连接子的氨基酸序列包括Maratea等人,《基因(Gene)》40:39-46,1985;Murphy等人,《美国国家科学院院刊》83:8258-8262,1986;美国专利第4,935,233号和美国专利第4,751,180号中所公开的那些。当特定融合多肽区段含有可用于隔开功能性结构域并防止空间干扰的非必需N末端氨基酸区域时,不需要连接子序列。优选的连接子典型地是作为重组融合蛋白的一部分合成的柔性氨基酸子序列。连接子多肽可以是在1个与200个氨基酸之间的长度、在1个与100个氨基酸之间的长度或在1个与50个氨基酸之间的长度,包括其间的所有整数值。

[0648] 例示性连接子包括但不限于以下氨基酸序列:甘氨酸聚合物($(G)_n$);甘氨酸-丝氨酸聚合物($(G_{1-5}S_{1-5})_n$),其中n是整数至少一、二、三、四或五;甘氨酸-丙氨酸聚合物;丙氨酸-丝氨酸聚合物;GGG(SEQ ID NO:69);DGGGS(SEQ ID NO:70);TGEKP(SEQ ID NO:71)(参见例如Liu等人,PNAS 5525-5530(1997));GRRR(SEQ ID NO:72)(Pomerantz等人,1995,前述);(GGGGS) $_n$,其中n=1、2、3、4或5(SEQ ID NO:73)(Kim等人,PNAS 93,1156-1160(1996.));EGKSSGSGSESKVD(SEQ ID NO:74)(Chaudhary等人,1990,《美国国家科学院院刊》87:1066-1070);KESGSVSSEQLAQFRSLD(SEQ ID NO:75)(Bird等人,1988,《科学》242:423-426);GGRRGGGS(SEQ ID NO:76);LRQRDGERP(SEQ ID NO:77);LRQKDGGGSERP(SEQ ID NO:78);LRQKD(GGGS) $_2$ ERP(SEQ ID NO:79)。或者,柔性连接子可使用能够建立DNA-结合位点和肽本

身的模型(Desjarlais和Berg,PNAS 90:2256-2260(1993),PNAS 91:11099-11103(1994))的计算机程序或通过噬菌体展示方法合理地设计。

[0649] 融合多肽还可以在本文所描述的每个多肽结构域之间或在内源性开放阅读框与供体修复模板所编码的多肽之间包含多肽裂解信号。此外,可将多肽裂解位点放在任何连接子肽序列中。例示性多肽裂解信号包括多肽裂解识别位点,如蛋白酶裂解位点、核酸酶裂解位点(例如罕见限制酶识别位点、自裂解核酶识别位点)和自裂解病毒寡肽(参见deFelipe和Ryan,2004.《交通(Traffic)》,5(8):616-26)。

[0650] 适合的蛋白酶裂解位点和自裂解肽是熟练技术人员已知的(参见例如Ryan等人,1997.《普通病毒学杂志(J.Gener.Virol.)》78,699-722;Scymczak等人(2004)《自然·生物技术(Nature Biotech.)》5,589-594)。例示性蛋白酶裂解位点包括但不限于马铃薯Y病毒(potyvirus)NIa蛋白酶(例如烟草蚀刻病毒蛋白酶)、马铃薯Y病毒HC蛋白酶、马铃薯Y病毒P1(P35)蛋白酶、大麦花叶病毒(byovirus)NIa蛋白酶、大麦花叶病毒RNA-2编码的蛋白酶、口疮病毒(aphthovirus)L蛋白酶、肠病毒2A蛋白酶、鼻病毒2A蛋白酶、细小核糖核酸3C蛋白酶、豇豆花叶病毒(comovirus)24K蛋白酶、线虫传多面体病毒(nepovirus)24K蛋白酶、水稻东格鲁球状病毒(rice tungro spherical virus,RTSV)3C样蛋白酶、欧防风黄点病毒(parsnip yellow fleck virus,PYVF)3C样蛋白酶、肝素、凝血酶、因子Xa以及肠激酶的裂解位点。在一个实施例中,烟草蚀刻病毒(TEV)蛋白酶裂解位点因其较高的裂解严格度而成为优选的,例如EXXYXQ(G/S)(SEQ ID NO:80),例如ENLYFQG(SEQ ID NO:81)和ENLYFQS(SEQ ID NO:82),其中X表示任何氨基酸(TEV裂解发生于Q与G或Q与S之间)。

[0651] 在某些实施例中,自裂解多肽位点包含2A或2A样位点、序列或结构域(Donnelly等人,2001.《普通病毒学杂志》82:1027-1041)。在一个特定实施例中,病毒2A肽是口疮病毒2A肽、马铃薯Y病毒2A肽或心病毒2A肽。

[0652] 在一个实施例中,病毒2A肽选自由以下组成的组:口蹄疫病毒(FMDV)2A肽、马鼻炎A病毒(ERAV)2A肽、明脉扁刺蛾B四体(Thosea asigna virus,TaV)2A肽、猪捷申病毒-1(porcine teschovirus-1,PTV-1)2A肽、泰勒病毒(Theilovirus)2A肽以及脑心肌炎病毒2A肽。

[0653] 2A位点的示例性实例提供于表2中。

[0654] 表2:例示性2A位点包括以下序列:

[0655]

SEQ ID NO: 83	GSGATNFSLLKQAGDVEENPGP
SEQ ID NO: 84	ATNFSLLKQAGDVEENPGP

[0656]

SEQ ID NO: 85	LLKQAGDVEENPGP
SEQ ID NO: 86	GSGEGRGSLTTCGDVEENPGP
SEQ ID NO: 87	EGRGSLTTCGDVEENPGP
SEQ ID NO: 88	LLTCGDVEENPGP
SEQ ID NO: 89	GSGQCTNYALLKLAGDVESNPGP
SEQ ID NO: 90	QCTNYALLKLAGDVESNPGP
SEQ ID NO: 91	LLKLAGDVESNPGP
SEQ ID NO: 92	GSGVKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 93	VKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 94	LLKLAGDVESNPGP
SEQ ID NO: 95	LLNFDLLKLAGDVESNPGP
SEQ ID NO: 96	TLNFDLLKLAGDVESNPGP
SEQ ID NO: 97	LLKLAGDVESNPGP
SEQ ID NO: 98	NFDLLKLAGDVESNPGP
SEQ ID NO: 99	QLLNFDLLKLAGDVESNPGP
SEQ ID NO: 100	APVKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 101	VTELLYRMKRAETCYCPRPLLAIHPTEARHKQKIVAPVKQT
SEQ ID NO: 102	LNFDLLKLAGDVESNPGP
SEQ ID NO: 103	LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP
SEQ ID NO: 104	EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP

[0657] G. 聚核苷酸

[0658] 在特定实施例中,提供了编码本文所涵盖的一种或多种归巢核酸内切酶变体、megaTAL、末端加工酶和融合多肽的聚核苷酸。如本文所使用,术语“聚核苷酸”或“核酸”是指脱氧核糖核酸(DNA)、核糖核酸(RNA)和DNA/RNA杂交体。聚核苷酸可以是单链或双链的并且是重组、合成或分离的。聚核苷酸包括但不限于:前信使RNA(pre-mRNA)、信使RNA(mRNA)、RNA、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微小RNA(miRNA)、核酶、基因组RNA(gRNA)、正链RNA(RNA(+))、负链RNA(RNA(-))、tracrRNA、crRNA、单向导RNA(sgRNA)、合成RNA、合成mRNA、基因组DNA(gDNA)、PCR扩增的DNA、互补DNA(cDNA)、合成DNA或重组DNA。聚核苷酸是指至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少40个、至少50个、至少100个、至少200个、至少300个、至少400个、至少500个、至少1000个、至少5000个、至少10000个或至少15000个或更多个核苷酸长度,以及所有中间长度的核苷酸,即核糖核苷酸或脱氧核糖核苷酸或任一类型核苷酸的修饰形式的聚合形式。容易理解的是,在此情形中,“中间长度”意思指在引述值之间的任何长度,如6、7、8、9等;101、102、103等;151、152、153等;201、202、203等。在特定实施例中,聚核苷酸或变体与参考序列具有至少或约50%、55%、60%、65%、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%或100%序列一致性。

[0659] 在特定实施例中,聚核苷酸可以经历密码子优化。如本文所使用,术语“密码子优化的”是指取代编码多肽的聚核苷酸中的密码子以便增加所述多肽的表达、稳定性和/或活性。影响密码子优化的因素包括但不限于以下一个或多个:(i)两种或更多种生物体或基因之间密码子偏好的变化或合成地构造的偏好表;(ii)生物体、基因或基因集内密码子偏好程度的变化;(iii)包括环境在内的密码子的系统性变化;(iv)根据解码tRNA得到的密码子

变化；(v) 根据总体或三联体一个位置中的GC%得到的密码子变化；(vi) 与参考序列，例如天然存在序列相似程度的变化；(vii) 密码子频率截止值的变化；(viii) 由DNA序列转录的mRNA的结构特性；(ix) 有关作为设计密码子取代集合的基础的DNA序列的功能的先验知识；(x) 每个氨基酸的密码子集的系统性变化；和/或(xi) 伪翻译起始位点的分离移除。

[0660] 如本文所使用，术语“核苷酸”是指与磷酸化糖的N-糖苷键联中的杂环含氮碱基。核苷酸应理解为包括天然碱基，以及多种本领域认可的经过修饰的碱基。此类碱基一般位于核苷酸糖部分的1'位处。核苷酸一般包含碱基、糖和磷酸酯基。在核糖核酸(RNA)中，糖是核糖，并且在脱氧核糖核酸(DNA)中，糖是脱氧核糖，即，不含核糖中存在的羟基的糖。例示性天然含氮碱基包括嘌呤，即腺苷(A)和胍(G)；以及嘧啶，即胞苷(C)和胸苷(T) (或在RNA情形中是尿嘧啶(U))。脱氧核糖的C-1原子键结至嘧啶的N-1或嘌呤的N-9。核苷酸通常是单磷酸酯、二磷酸酯或三磷酸酯。核苷酸可以是未修饰的或在糖、磷酸酯和/或碱基部分处经历修饰(也可互换地称为核苷酸类似物、核苷酸衍生物、经过修饰的核苷酸、非天然核苷酸和非标准核苷酸；参见例如WO 92/07065和WO 93/15187)。经过修饰的核酸碱基的实例概述于Limbach等人(1994,《核酸研究(Nucleic Acids Res.)》22,2183-2196)中。

[0661] 核苷酸也可被视为核苷的磷酸酯，并且酯化是在连接至糖的C-5的羟基上发生。如本文所使用，术语“核苷”是指与糖的N-糖苷键联中的杂环含氮碱基。核苷在本领域中公认为包括天然碱基，并且还包含众所周知的经过修饰的碱基。此类碱基一般位于核苷糖部分的1'位处。核苷一般包含碱基和糖基。核苷可以是未修饰的或在糖和/或碱基部分处经历修饰(也可互换地称为核苷类似物、核苷衍生物、经过修饰的核苷、非天然核苷或非标准核苷)。另外，如上文所述，经过修饰的核酸碱基的实例概述于Limbach等人(1994,《核酸研究》22,2183-2196)中。

[0662] 聚核苷酸的示例性实例包括但不限于编码SEQ ID NO:1-24和60-64的聚核苷酸，以及SEQ ID NO:25-59和65-68中所述的聚核苷酸序列。

[0663] 在各种示例性实施例中，本文所涵盖的聚核苷酸包括但不限于编码归巢核酸内切酶变体、megaTAL、末端加工酶、融合多肽的聚核苷酸，以及包含本文所涵盖的聚核苷酸的表达载体、病毒载体和转移质粒。

[0664] 如本文所使用，术语“聚核苷酸变体”和“变体”等是指与参考聚核苷酸序列展示相当大的序列一致性的聚核苷酸或在下文所定义的严格条件下与参考序列杂交的聚核苷酸。这些术语还涵盖因至少一个核苷酸的添加、缺失、取代或修饰而有别于参考聚核苷酸的聚核苷酸。因此，术语“聚核苷酸变体”和“变体”包括一个或多个核苷酸被添加或缺失或修饰，或被置换成不同核苷酸的聚核苷酸。就这一点而言，本领域中应充分理解，可以对参考聚核苷酸进行包括突变、添加、缺失和取代在内的某些改变，由此使改变的聚核苷酸保持参考聚核苷酸的生物功能或活性。

[0665] 在一个实施例中，聚核苷酸包含在严格条件下与靶核酸序列杂交的核苷酸序列。在“严格条件”下杂交描述使彼此至少60%一致的核苷酸序列保持杂交的杂交方案。一般来说，选择的严格条件比特定序列在既定离子强度和pH下的热熔点(T_m)低约5°C。T_m是在平衡时50%的与靶序列互补的探针与靶序列杂交的温度(在既定离子强度、pH和核酸浓度下)。由于靶序列在T_m下一般是过量存在，故在平衡时50%探针被占据。

[0666] 如本文所使用，叙述“序列一致性”或例如包含“与……50%一致的序列”是指在核

苷酸与核苷酸基础上或在氨基酸与氨基酸基础上,序列在比较窗内一致的程度。因此,“序列一致性百分比”可以通过以下方式计算:在比较窗内比较二个最佳地对准的序列,测定的一致核酸碱基(例如A、T、C、G、I)或一致氨基酸残基(例如Ala、Pro、Ser、Thr、Gly、Val、Leu、Ile、Phe、Tyr、Trp、Lys、Arg、His、Asp、Glu、Asn、Gln、Cys以及Met)在两个序列中出现的位置的数目以得到相配位置的数目,用相配位置的数目除以比较窗中的位置总数(即,窗大小)并将结果乘以100,得到序列一致性百分比。包括与本文所描述的参考序列中的任一个具有至少约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%序列一致性的核苷酸和多肽,典型地其中多肽变体保持参考多肽的至少一种生物活性。

[0667] 用于描述两个或更多个聚核苷酸或多肽之间的序列关系的术语包括“参考序列”、“比较窗”、“序列一致性”、“序列一致性百分比”和“基本上一致”。“参考序列”的长度是至少12个,但常常是15至18个且通常是至少25个单体单元,包括核苷酸和氨基酸残基在内。因为两个聚核苷酸可以各自包含(1)在两个聚核苷酸之间类似的序列(即,仅完整聚核苷酸序列的一部分);和(2)在两个聚核苷酸之间相异的序列,两个(或更多个)聚核苷酸之间的序列比较典型地通过在“比较窗”内比较该两个聚核苷酸的序列以鉴别并比较序列局部区域的相似性来进行。“比较窗”是指具有至少6个、通常约50个至约100个、更通常约100个至约150个连续位置的概念性区段,其中在最佳地对准一个序列与具有相同数目连续位置的参考序列之后,比较该两个序列。对于最佳对准的两个序列,比较窗可以包含相较于参考序列(不包含添加或缺失的序列),约20%或更低百分比的添加或缺失(即,空位)。用于比对比较窗的最佳序列对准可以通过计算机实施算法(威斯康星州遗传学软件包7.0版本(Wisconsin Genetics Software Package Release 7.0),遗传学计算机组(Genetics Computer Group),美国威斯康星州575科学驱动麦迪逊(575 Science Drive Madison,WI,USA)中的GAP、BESTFIT、FASTA和TFASTA)或通过由各种所选方法中的任一种生成的检查和最佳比对(即,在比较窗内产生最高同源性百分比)来进行。还可以参考如例如Altschul等人,1997,《核酸研究》25:3389所公开的BLAST程序家族。有关序列分析的详细论述可以见于Ausubel等人,《最新分子生物学实验方法汇编(Current Protocols in Molecular Biology)》,John Wiley&Sons Inc.,1994-1998,第15章的19.3单元。

[0668] 如本文所使用,“分离的聚核苷酸”是指已经从在天然存在的状态中侧接其的序列纯化的聚核苷酸,例如从通常邻近于DNA片段的序列移出的所述DNA片段。在特定实施例中,“分离的聚核苷酸”是指互补DNA(cDNA)、重组聚核苷酸、合成聚核苷酸或在自然界中不存在且通过人工制备的其它聚核苷酸。

[0669] 在各种实施例中,聚核苷酸包含编码本文所涵盖的多肽的mRNA,所述多肽包括但不限于归巢核酸内切酶变体、megaTAL和末端加工酶。在某些实施例中,所述mRNA包含帽、一个或多个核苷酸和聚腺苷酸尾。

[0670] 如本文所使用,术语“5'帽”或“5'帽结构”或“5'帽部分”是指并入mRNA 5'端的化学修饰。5'帽涉及到核输出、mRNA稳定性和翻译。

[0671] 在特定实施例中,本文所涵盖的mRNA包含5'帽,其包含在mRNA分子末端鸟苷帽残基与5'末端转录的有义核苷酸之间的5'-ppp-5'-三磷酸键联。接着,此5'-鸟苷酸帽可被甲基化以产生N7-甲基-鸟苷酸残基。

[0672] 适用于本文所涵盖的mRNA聚核苷酸的特定实施例中的5'帽的示例性实例包括但不限于：未甲基化的5'帽类似物，例如G(5') ppp(5') G、G(5') ppp(5') C、G(5') ppp(5') A；甲基化的5'帽类似物，例如m⁷G(5') ppp(5') G、m⁷G(5') ppp(5') C和m⁷G(5') ppp(5') A；二甲基化的5'帽类似物，例如m^{2,7}G(5') ppp(5') G、m^{2,7}G(5') ppp(5') C和m^{2,7}G(5') ppp(5') A；三甲基化的5'帽类似物，例如m^{2,2,7}G(5') ppp(5') G、m^{2,2,7}G(5') ppp(5') C和m^{2,2,7}G(5') ppp(5') A；二甲基化的对称5'帽类似物，例如m⁷G(5') ppm⁷(5') G、m⁷G(5') ppm⁷(5') C和m⁷G(5') ppm⁷(5') A；以及抗反向5'帽类似物，例如抗反向帽类似物(Anti-Reverse Cap Analog, ARCA) 帽，命名为3'-O-Me-m⁷G(5') ppp(5') G、2'-O-Me-m⁷G(5') ppp(5') G、2'-O-Me-m⁷G(5') ppp(5') C、2'-O-Me-m⁷G(5') ppp(5') A、m⁷2'd(5') ppp(5') G、m⁷2'd(5') ppp(5') C、m⁷2'd(5') ppp(5') A、3'-O-Me-m⁷G(5') ppp(5') C、3'-O-Me-m⁷G(5') ppp(5') A、m⁷3'd(5') ppp(5') G、m⁷3'd(5') ppp(5') C、m⁷3'd(5') ppp(5') A及其四磷酸酯衍生物(参见例如Jemielity等人, RNA, 9:1108-1122 (2003))。

[0673] 在特定实施例中，mRNA包含5'帽，它是7-甲基鸟苷酸("m⁷G")，通过三磷酸酯桥连接至第一个转录的核苷酸的5'端，产生m⁷G(5') ppp(5') N，其中N是任何核苷。

[0674] 在一些实施例中，mRNA包含5'帽，其中所述帽是帽0结构(帽0结构不包含附接至碱基1和2的核糖的2'-O-甲基残基)、帽1结构(帽1结构在碱基2处具有2'-O-甲基残基)或帽2结构(帽2结构具有附接至碱基2和3的2'-O-甲基残基)。

[0675] 在一个实施例中，mRNA包含m⁷G(5') ppp(5') G帽。

[0676] 在一个实施例中，mRNA包含ARCA帽。

[0677] 在特定实施例中，本文所涵盖的mRNA包含一个或多个经过修饰的核苷。

[0678] 在一个实施例中，mRNA包含一个或多个选自以下组成的组的经过修饰的核苷：假尿苷、吡啶-4-酮核苷、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-假尿苷、5-丙炔基-尿苷、1-丙炔基-假尿苷、5-牛磺酸甲基尿苷、1-牛磺酸甲基-假尿苷、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-乙酰基胞苷、5-甲酰基胞苷、N4-甲基胞苷、5-羟基甲基胞苷、1-甲基-假异胞苷、吡咯并胞苷、吡咯并假异胞苷、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-脱氮-假异胞苷、1-甲基-1-脱氮-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-脱氮-腺嘌呤、7-脱氮-8-氮杂-腺嘌呤、7-脱氮-2-氨基嘌呤、7-脱氮-8-氮杂-2-氨基嘌呤、7-脱氮-2,6-二氨基嘌呤、7-脱氮-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-羟基异戊烯基)腺苷、2-甲硫基-N6-(顺-羟基异戊烯基)腺苷、N6-甘氨酸基氨甲酰基腺苷、N6-苏氨酸基氨甲酰基腺苷、2-甲硫基-N6-苏氨酸基氨甲酰基腺苷、N6,N6-二甲基腺苷、7-甲基腺嘌呤、2-甲硫基-腺嘌呤、2-甲氧基-腺嘌呤、肌苷、1-甲基-肌苷、怀俄苷(wyosine)、怀丁苷(wybutosine)、7-脱

氮-鸟苷、7-脱氮-8-氮杂-鸟苷、6-硫代-鸟苷、6-硫代-7-脱氮-鸟苷、6-硫代-7-脱氮-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,N2-二甲基鸟苷、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-鸟苷以及N2,N2-二甲基-6-硫代-鸟苷。

[0679] 在一个实施例中,mRNA包含一个或多个选自由以下组成的组的经过修饰的核苷:假尿苷、吡啶-4-酮核苷、5-氮杂-尿苷、2-硫代-5-氮杂-尿苷、2-硫代尿核苷、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-假尿苷、5-丙炔基-尿苷、1-丙炔基-假尿苷、5-牛磺酸甲基尿苷、1-牛磺酸甲基-假尿苷、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-假尿苷以及4-甲氧基-2-硫代-假尿苷。

[0680] 在一个实施例中,mRNA包含一个或多个选自由以下组成的组的经过修饰的核苷:5-氮杂-胞苷、假异胞苷、3-甲基-胞苷、N4-乙酰基胞苷、5-甲酰基胞苷、N4-甲基胞苷、5-羟基甲基胞苷、1-甲基-假异胞苷、吡咯并胞苷、吡咯并假异胞苷、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-脱氮-假异胞苷、1-甲基-1-脱氮-假异胞苷、泽布拉林、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-假异胞苷以及4-甲氧基-1-甲基-假异胞苷。

[0681] 在一个实施例中,mRNA包含一个或多个选自由以下组成的组的经过修饰的核苷:2-氨基嘌呤、2,6-二氨基嘌呤、7-脱氮-腺嘌呤、7-脱氮-8-氮杂-腺嘌呤、7-脱氮-2-氨基嘌呤、7-脱氮-8-氮杂-2-氨基嘌呤、7-脱氮-2,6-二氨基嘌呤、7-脱氮-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-羟基异戊烯基)腺苷、2-甲硫基-N6-(顺-羟基异戊烯基)腺苷、N6-甘氨酸基氨甲酰基腺苷、N6-苏氨酸基氨甲酰基腺苷、2-甲硫基-N6-苏氨酸基氨甲酰基腺苷、N6,N6-二甲基腺苷、7-甲基腺嘌呤、2-甲硫基-腺嘌呤以及2-甲氧基-腺嘌呤。

[0682] 在一个实施例中,mRNA包含一个或多个选自由以下组成的组的经过修饰的核苷:肌苷、1-甲基-肌苷、怀俄苷、怀丁苷、7-脱氮-鸟苷、7-脱氮-8-氮杂-鸟苷、6-硫代-鸟苷、6-硫代-7-脱氮-鸟苷、6-硫代-7-脱氮-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,N2-二甲基鸟苷、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-鸟苷以及N2,N2-二甲基-6-硫代-鸟苷。

[0683] 在一个实施例中,mRNA包含一个或多个假尿苷、一个或多个5-甲基-胞嘧啶和/或一个或多个5-甲基-胞苷。

[0684] 在一个实施例中,mRNA包含一个或多个假尿苷。

[0685] 在一个实施例中,mRNA包含一个或多个5-甲基-胞苷。

[0686] 在一个实施例中,mRNA包含一个或多个5-甲基-胞嘧啶。

[0687] 在特定实施例中,本文所涵盖的mRNA包含聚腺苷酸尾以帮助防止mRNA被核酸外切酶降解,使mRNA稳定并且促进翻译。在某些实施例中,mRNA包含3'聚腺苷酸尾结构。

[0688] 在特定实施例中,聚腺苷酸尾的长度是至少约10、25、50、75、100、150、200、250、300、350、400、450个或至少约500个或更多个腺嘌呤核苷酸或任何中间数目的腺嘌呤核苷酸。在特定实施例中,聚腺苷酸尾的长度是至少约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、150、151、152、153、154、155、156、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、179、180、181、182、183、184、185、186、187、188、189、190、191、192、193、194、195、196、197、198、199、200、201、202、202、203、205、206、207、208、209、210、211、212、213、214、215、216、217、218、219、220、221、222、223、224、225、226、227、228、229、230、231、232、233、234、235、236、237、238、239、240、241、242、243、244、245、246、247、248、249、250、251、252、253、254、255、256、257、258、259、260、261、262、263、264、265、266、267、268、269、270、271、272、273、274或275个或更多个腺嘌呤核苷酸。

[0689] 在特定实施例中,聚腺苷酸尾的长度是约10至约500个腺嘌呤核苷酸、约50至约500个腺嘌呤核苷酸、约100至约500个腺嘌呤核苷酸、约150至约500个腺嘌呤核苷酸、约200至约500个腺嘌呤核苷酸、约250至约500个腺嘌呤核苷酸、约300至约500个腺嘌呤核苷酸、约50至约450个腺嘌呤核苷酸、约50至约400个腺嘌呤核苷酸、约50至约350个腺嘌呤核苷酸、约100至约500个腺嘌呤核苷酸、约100至约450个腺嘌呤核苷酸、约100至约400腺嘌呤核苷酸、约100至约350个腺嘌呤核苷酸、约100至约300个腺嘌呤核苷酸、约150至约500个腺嘌呤核苷酸、约150至约450个腺嘌呤核苷酸、约150至约400个腺嘌呤核苷酸、约150至约350个腺嘌呤核苷酸、约150至约300个腺嘌呤核苷酸、约150至约250个腺嘌呤核苷酸、约150至约200个腺嘌呤核苷酸、约200至约500个腺嘌呤核苷酸、约200至约450个腺嘌呤核苷酸、约200至约400个腺嘌呤核苷酸、约200至约350个腺嘌呤核苷酸、约200至约300个腺嘌呤核苷酸、约250至约500个腺嘌呤核苷酸、约250至约450个腺嘌呤核苷酸、约250至约400个腺嘌呤核苷酸、约250至约350个腺嘌呤核苷酸或约250至约300个腺嘌呤核苷酸或任何中间范围的腺嘌呤核苷酸。

[0690] 描述聚核苷酸取向的术语包括:5' (通常是聚核苷酸具有游离磷酸酯基的一端) 和3' (通常是聚核苷酸具有游离羟基(OH)的一端)。聚核苷酸序列可以按5'至3'取向或3'至5'取向标注。对于DNA和mRNA,5'至3'链被命名为“有义”链、“正”链或“编码”链,因为其序列与前信使(前mRNA)的序列一致[不过在RNA中是尿嘧啶(U),而在DNA中是胸腺嘧啶(T)]。对于DNA和mRNA,作为由RNA聚合酶转录的链的互补3'至5'链被命名为“模板”链、“反义”链、“负”链或“非编码”链。如本文所使用,术语“反向取向”是指以3'至5'取向书写的5'至3'序列或以5'至3'取向书写的3'至5'序列。

[0691] 术语“互补”和“互补性”是指通过碱基配对规则相关联的聚核苷酸(即,核苷酸序列)。例如,DNA序列5'A G T C A T G 3'的互补链是3'T C A G T A C 5'。后一个序列通常被写成反向互补序列,其中5'端在左侧并且3'端在右侧,即5'C A T G A C T 3'。与其反向互补序列相同的序列被称为回文序列。互补可以是“部分”的,其中仅一些核酸的碱基根据碱基配对规则匹配。或者,在核酸之间可以存在“完全”或“总体”互补性。

[0692] 如本文所使用,术语“核酸盒”或“表达盒”是指载体内可以表达RNA且随后表达多肽的基因序列。在一个实施例中,核酸盒含有目的基因,例如目的聚核苷酸。在另一个实施例中,核酸盒含有一个或多个表达控制序列,例如启动子、增强子、聚腺苷酸序列以及目的

基因,例如目的聚核苷酸。载体可以包含1、2、3、4、5、6、7、8、9或10个或更多个核酸盒。核酸盒在载体内是按位置依序地取向,使得盒中的核酸可以被转录成RNA,并且当需要时被翻译成蛋白质或多肽,经历被转化的细胞中的活性所需的适当翻译后修饰,并且通过靶向适当细胞内区室或分泌至细胞外区室中而易位至适当区室中以获得生物活性。优选地,所述盒的3'和5'端适合于迅速地插入载体中,例如其在每一端处具有限制性核酸内切酶位点。在一个优选实施例中,核酸盒含有用于治疗、预防或改善遗传病症的治疗基因序列。所述盒可以作为单一单元移出并插入质粒或病毒载体中。

[0693] 聚核苷酸包括目的聚核苷酸。如本文所使用,术语“目的聚核苷酸”是指如本文所预期,编码多肽或融合多肽的聚核苷酸或用作模板以转录抑制性聚核苷酸的聚核苷酸。

[0694] 此外,本领域的普通技术人员应了解,如本文所预期,由于遗传密码的简并性,存在许多可以编码多肽或其变体片段的核苷酸序列。这些聚核苷酸中有一些与任何天然基因的核苷酸序列具有最小同源性。尽管如此,在特定实施例中特别地涵盖由于密码子使用差异而变化的聚核苷酸,例如针对人类和/或灵长类动物密码子选择而优化的聚核苷酸。在一个实施例中,提供了包含特定等位基因序列的聚核苷酸。等位基因是因核苷酸的一个或多个突变,如缺失、添加和/或取代而改变的内源性聚核苷酸序列。

[0695] 在某一实施例中,目的聚核苷酸包含供体修复模板。

[0696] 在某一实施例中,目的聚核苷酸包含抑制性聚核苷酸,包括但不限于siRNA、miRNA、shRNA、核酶或另一抑制性RNA。

[0697] 在一个实施例中,包含抑制性RNA的供体修复模板包含一个或多个调控序列,如例如强组成型pol III,例如人或小鼠U6snRNA启动子、人和小鼠H1RNA启动子或人tRNA-val启动子,或强组成型pol II启动子,如本文别处所描述。

[0698] 特定实施例中所涵盖的聚核苷酸,不管编码序列本身的长度如何,可以与其它DNA序列组合,如本文别处所公开或如本领域中所知,所述序列是如启动子和/或增强子、非翻译区(UTR)、Kozak序列、聚腺苷酸化信号、其它限制酶位点、多个克隆位点、内部核糖体进入位点(IRES)、重组酶识别位点(例如LoxP、FRT和Att位点)、终止密码子、转录终止信号、转录后反应元件以及编码自裂解多肽、表位标签的聚核苷酸,使得其总长度可能变化极大。因此,特定实施例中预期可以采用几乎任何长度的聚核苷酸片段,并且总长度优选地受预定重组DNA方案中制备的简易性和使用限制。

[0699] 聚核苷酸可以使用本领域中已知并且可用的多种公认技术中的任一种制备、操作、表达和/或递送。为了表达所希望的多肽,可将编码所述多肽的核苷酸序列插入适当载体中。也可以通过将编码所希望多肽的mRNA递送至细胞中来表达所述多肽。

[0700] 载体的示例性实例包括但不限于质粒、自主复制序列和转座元件,例如Sleeping Beauty、PiggyBac。

[0701] 载体的其它示例性实例包括但不限于质粒;噬菌粒;黏粒;人工染色体,如酵母人工染色体(YAC)、细菌人工染色体(BAC)或P1源性人工染色体(PAC);噬菌体,如 λ 噬菌体或M13噬菌体;以及动物病毒。

[0702] 可用作载体的病毒的示例性实例包括但不限于逆转录病毒(包括慢病毒)、腺病毒、腺相关病毒、疱疹病毒(例如单纯疱疹病毒)、痘病毒、杆状病毒、乳头瘤病毒以及乳多泡病毒(例如SV40)。

[0703] 表达载体的示例性实例包括但不限于用于在哺乳动物细胞中表达的pCIneo载体(Promega);用于慢病毒介导的基因转移和在哺乳动物细胞中表达的pLenti4/V5-DEST™、pLenti6/V5-DEST™和pLenti6.2/V5-GW/lacZ (Invitrogen)。在特定实施例中,本文所公开的多肽的编码序列可连接至此类表达载体中以在哺乳动物细胞中表达所述多肽。

[0704] 在特定实施例中,载体是游离型载体或维持在染色体外的载体。如本文所使用,术语“游离型”是指载体能够复制而不必整合至宿主染色体DNA中且不会从分裂的宿主细胞逐渐损失,还意味着所述载体在染色体外或游离地复制。

[0705] 表达载体中存在的“表达控制序列”、“控制元件”或“调控序列”是载体的非翻译区域,即复制起点、选择盒、启动子、增强子、翻译起始信号(Shine Dalgarno序列或Kozak序列)内含子、转录后调控元件、聚腺苷酸化序列、5'和3'非翻译区,所述非翻译区与宿主细胞蛋白相互作用以进行转录与翻译。这些元件在其强度和特异性方面可能变化。取决于所利用的载体系统和宿主,可使用多种适合的转录与翻译元件,包括遍在型启动子和诱导型启动子。

[0706] 在特定实施例中,聚核苷酸包含载体,包括但不限于表达载体和病毒载体。载体可包含一个或多个外源性、内源性或异源控制序列,如启动子和/或增强子。“内源性控制序列”是天然地与基因组中的给定基因连接的控制序列。“外源性控制序列”是借助于基因操作(即,分子生物技术)与基因并接放置以使得所连接的增强子/启动子引导该基因的转录的控制序列。“异源控制序列”是来自与经历基因操作的细胞不同的物种的外源序列。“合成”控制序列可以包含具有一个或多个内源性和/或外源性序列和/或在体外或通过计算机模拟确定的序列的元件,其为特定疗法提供最佳的启动子和/或增强子活性。

[0707] 如本文所使用,术语“启动子”是指RNA聚合酶所结合的聚核苷酸(DNA或RNA)的识别位点。RNA聚合酶起始并且转录可操作地连接至启动子的聚核苷酸。在特定实施例中,在哺乳动物细胞中可操作的启动子包含位于转录起始位点上游约25至30个碱基处的富AT区,和/或在转录起始点上游70至80个碱基处所见的另一序列,即CNCAAT区,其中N可以是任何核苷酸。

[0708] 术语“增强子”是指含有能够提供增强的转录的序列并且在一些情况下可以独立于其相对于另一控制序列的取向起作用的一段DNA。增强子可以与启动子和/或其它增强子元件协作地或叠加地起作用。术语“启动子/增强子”是指含有能够同时提供启动子和增强子功能的序列的一段DNA。

[0709] 术语“可操作地连接”是指所描述的组分的关系允许其以其预定方式起作用的并接。在一个实施例中,所述术语是指核酸表达控制序列(如启动子和/或增强子)与第二聚核苷酸序列,例如目的聚核苷酸之间的功能性连接,其中所述表达控制序列引导对应于所述第二序列的核酸的转录。

[0710] 如本文所使用,术语“组成型表达控制序列”是指不断地或连续地使可操作地连接的序列转录的启动子、增强子或启动子/增强子。组成型表达控制序列可以是允许在多种细胞和组织类型中表达的“遍在型”启动子、增强子或启动子/增强子,或允许分别在多种限定细胞和组织类型中表达的“细胞特异性”、“细胞类型特异性”、“细胞谱系特异性”或“组织特异性”启动子、增强子或启动子/增强子。

[0711] 适用于特定实施例中的示例性遍在型表达控制序列包括但不限于巨细胞病毒

(CMV) 立即早期启动子、病毒猿猴病毒40 (SV40) (例如早期或晚期)、莫洛尼鼠白血病毒 (Moloney murine leukemia virus, MoMLV) LTR启动子、劳氏肉瘤病毒 (Rous sarcoma virus, RSV) LTR、单纯疱疹病毒 (HSV) (胸苷激酶) 启动子、来自牛痘病毒的H5、P7.5和P11启动子、短延伸因子1- α (EF1a-短) 启动子、长延伸因子1- α (EF1a-长) 启动子、早期生长反应1 (EGR1)、铁蛋白H (FerH)、铁蛋白L (FerL)、甘油醛3-磷酸脱氢酶 (GAPDH)、真核翻译起始因子4A1 (EIF4A1)、热休克70kDa蛋白质5 (HSPA5)、热休克蛋白90kDa β 成员1 (HSP90B1)、热休克蛋白70kDa (HSP70)、 β -驱动蛋白 (β -KIN)、人ROSA 26基因座 (Irions等人,《自然·生物技术 (Nature Biotechnology)》25,1477-1482 (2007))、泛素C启动子 (UBC)、磷酸甘油酸激酶-1 (PGK) 启动子、巨细胞病毒增强子/鸡 β -肌动蛋白 (CAG) 启动子、 β -肌动蛋白启动子和骨髓增生肉瘤病毒增强子、负调控区缺失、dl587rev引物结合位点取代 (MND) 的启动子 (Challita等人,《病毒学杂志 (J Virol.)》69 (2):748-55 (1995))。

[0712] 在一个特定实施例中,可能需要使用细胞、细胞类型、细胞谱系或组织特异性表达控制序列来实现所希望聚核苷酸序列的细胞类型特异性、谱系特异性或组织特异性表达 (例如仅在一小组细胞类型、细胞谱系或组织中或在特定发育阶段期间表达编码多肽的特定核酸)。

[0713] 如本文所使用,“条件性表达”可以指任何类型的条件性表达,包括但不限于诱导型表达;阻遏型表达;在具有特定生理、生物或疾病状态的细胞或组织中的表达等。这一定义不打算排除细胞类型或组织特异性表达。某些实施例提供了目的聚核苷酸的条件性表达,例如表达是通过使细胞、组织、生物体等经历使聚核苷酸表达或使目的聚核苷酸所编码的聚核苷酸的表达增加或减少的处理或条件进行控制。

[0714] 诱导型启动子/系统的示例性实例包括但不限于类固醇诱导型启动子,如编码糖皮质激素或雌激素受体的基因的启动子 (可通过用相应激素处理诱导);金属硫蛋白启动子 (可通过用各种重金属处理诱导);MX-1启动子 (可通过干扰素诱导);“GeneSwitch”米非司酮 (mifepristone) 可调控系统 (Sirin等人,2003,《基因 (Gene)》,323:67);二甲氨基二硫代甲酸铜 (cumate) 诱导型基因开关 (WO 2002/088346);四环素依赖性调控系统等。

[0715] 条件性表达还可以通过使用位点特异性DNA重组酶实现。根据某些实施例,聚核苷酸包含至少一个 (典型地两个) 用于由位点特异性重组酶介导的重组的位点。如本文所使用,术语“重组酶”或“位点特异性重组酶”包括参与涉及一个或多个 (例如二、三、四、五、六、七、八、九、十个或更多个) 重组位点的重组反应的切除型或整合型蛋白质、酶、辅因子或相关蛋白质,其可以是野生型蛋白质 (参见Landy,《生物技术新见 (Current Opinion in Biotechnology)》3:699-707 (1993)) 或其突变体、衍生物 (例如含有重组蛋白序列或其片段的融合蛋白)、片段和变体。适用于特定实施例中的重组酶的示例性实例包括但不限于: Cre、Int、IHF、Xis、Flp、Fis、Hin、Gin、 Φ C31、Cin、Tn3解离酶、TndX、XerC、XerD、TnpX、Hjc、Gin、SpCCE1以及ParA。

[0716] 聚核苷酸可以包含用于多种位点特异性重组酶中的任一种的一个或多个重组位点。应理解,除用于位点特异性重组酶的靶位点外,还需要整合载体,例如逆转录病毒载体或慢病毒载体所需的任何位点。如本文所使用,术语“重组序列”、“重组位点”或“位点特异性重组位点”是指重组酶识别并结合的特定核酸序列。

[0717] 例如,Cre重组酶的一个重组位点是loxP,它是有34个碱基对的序列,该序列包含

两个有13个碱基对的反向重复序列(充当重组酶结合位点)侧接有8个碱基对的核心序列(参见Sauer B.,《生物技术新见》5:521-527(1994)的图1)。其它例示性loxP位点包括但不限于:lox511(Hoess等人,1996;Bethke和Sauer,1997)、lox5171(Lee和Saito,1998)、lox2272(Lee和Saito,1998)、m2(Langer等人,2002)、lox71(Albert等人,1995)以及lox66(Albert等人,1995)。

[0718] 适合FLP重组酶的识别位点包括但不限于:FRT(McLeod等人,1996);F₁、F₂、F₃(Schlake和Bode,1994);F₄、F₅(Schlake和Bode,1994);FRT(LE)(Senecoff等人,1988);FRT(RE)(Senecoff等人,1988)。

[0719] 识别序列的其它实例是attB、attP、attL和attR序列,它们由重组酶 λ 整合酶,例如 ϕ -c31识别。 ϕ C31介导仅在异型位点attB(长度是34bp)与attP(长度是39bp)之间发生的重组(Groth等人,2000)。attB和attP分别是针对细菌和噬菌体基因组上的噬菌体整合酶的连接位点命名,二者都含有很可能与 ϕ C31同源二聚体结合的不完美反向重复序列(Groth等人,2000)。产物位点attL和attR对进一步 ϕ C31介导的重组实际上呈惰性(Belteki等人,2003),使得反应不可逆。对于催化插入,已经发现,与attP位点插入到基因组attB位点中相比,带有attB的DNA更容易插入到基因组attP位点中(Thyagarajan等人,2001;Belteki等人,2003)。因此,典型策略通过同源重组将带有attP的“对接位点”定位至指定基因座中,所述基因座接着与带有attB的输入序列(incoming sequence)配对以进行插入。

[0720] 在一个实施例中,本文所涵盖的聚核苷酸包含侧接一对重组酶识别位点的供体修复模板聚核苷酸。在特定实施例中,修复模板聚核苷酸侧接LoxP位点、FRT位点或att位点。

[0721] 在特定实施例中,本文所涵盖的聚核苷酸包括编码一个或多个多肽的一个或多个目的聚核苷酸。在特定实施例中,为了实现多个多肽各自的高效翻译,聚核苷酸序列可以通过一个或多个编码自裂解多肽的IRES序列或聚核苷酸序列隔开。

[0722] 如本文所使用,“内部核糖体进入位点”或“IRES”是指促进直接内部核糖体进入到顺反子(蛋白质编码区)的起始密码子如ATG中,由此引起基因的不依赖于帽的翻译的元件。参见例如Jackson等人,1990.《生化科学趋势(Trends Biochem Sci)》15(12):477-83,和Jackson和Kaminski.1995.RNA 1(10):985-1000。本领域的技术人员一般采用的IRES的实例包括美国专利第6,692,736号中所描述的那些。本领域中已知的“IRES”的其它实例包括但不限于从小核糖核酸病毒获得的IRES(Jackson等人,1990)以及从病毒或细胞mRNA源,如免疫球蛋白重链结合蛋白(BiP)、血管内皮生长因子(VEGF)(Huez等人1998.《分子与细胞生物学(Mol.Cell.Biol.)》18(11):6178-6190)、成纤维细胞生长因子2(FGF-2)和类胰岛素生长因子(IGFII)获得的IRES、翻译起始因子eIF4G及酵母转录因子TFIID和HAP4、可购自Novagen的脑心肌炎病毒(EMCV)(Duke等人,1992.《病毒学杂志》66(3):1602-9)和VEGF IRES(Huez等人,1998.《分子细胞生物学》18(11):6178-90)。IRES在小核糖核酸病毒科(Picornaviridae)、二顺反子病毒科(Dicistroviridae)和黄病毒科(Flaviviridae)物种的病毒基因组中以及在HCV、弗云德鼠白血病病毒(Friend murine leukemia virus, FrMLV)和莫洛尼鼠白血病病毒(MoMLV)中也已有报导。

[0723] 在一个实施例中,用于本文所涵盖的聚核苷酸中的IRES是EMCV IRES。

[0724] 在特定实施例中,聚核苷酸包含具有共同Kozak序列且编码所希望多肽的聚核苷

酸。如本文所使用,术语“Kozak序列”是指极大地促进mRNA与核糖体的小亚单元的初始结合并增加翻译的短核苷酸序列。共同Kozak序列是 (GCC) RCCATGG (SEQ ID NO:105),其中R是嘌呤 (A或G) (Kozak,1986.《细胞 (Cell)》.44 (2):283-92;和Kozak,1987.《核酸研究》15 (20):8125-48)。

[0725] 引导异源核酸转录物的高效终止和聚腺苷酸化的元件将增加异源基因的表达。转录终止信号一般见于聚腺苷酸化信号的下游。在特定实施例中,载体在编码待表达多肽的聚核苷酸的3'端包含聚腺苷酸化序列。如本文所使用,术语“聚腺苷酸位点 (polyA site)”、“聚腺苷酸序列 (polyA sequence)”、“聚腺苷酸位点 (poly (A) site)”或“聚腺苷酸序列 (poly (A) sequence)”表示引导由RNA聚合酶II引起的初生RNA转录物终止和聚腺苷酸化的DNA序列。聚腺苷酸化序列可以通过在编码序列的3'端添加聚腺苷酸尾来促进mRNA稳定性,并因此促使翻译效率增加。重组转录物的高效聚腺苷酸化是合乎需要的,因为不含聚腺苷酸尾的转录物不稳定并且迅速地降解。可以用于载体中的聚腺苷酸信号的示例性实例包括理想的聚腺苷酸序列 (例如AATAAA、ATTAAA、AGTAAA)、牛生长激素聚腺苷酸序列 (BGHpA)、兔 β -珠蛋白聚腺苷酸序列 (r β gpA) 或本领域中已知的另一适合的异源性或内源性聚腺苷酸序列。

[0726] 在一些实施例中,聚核苷酸或含有所述聚核苷酸的细胞利用了自杀基因,包括诱导型自杀基因,来降低直接毒性和/或不受控增殖的风险。在具体实施例中,自杀基因对含有所述聚核苷酸或细胞的宿主不具有免疫原性。可以使用的自杀基因的某一实例是半胱天冬酶-9或半胱天冬酶-8或胞嘧啶脱氨酶。半胱天冬酶-9可以使用特定的二聚化化学诱导剂 (CID) 活化。

[0727] 在某些实施例中,聚核苷酸包含使本文所涵盖的基因修饰的细胞易于在体内发生阴性选择的基因区段。“阴性选择”是指,输注的细胞可因个体体内状况的变化而消除。阴性选择表型可以通过插入赋予对所施用试剂 (例如化合物) 的敏感性的基因而产生。阴性选择基因是本领域中已知的,并且包括但不限于:赋予更昔洛韦敏感性的I型单纯疱疹病毒胸苷激酶 (HSV-I TK) 基因;细胞次黄嘌呤磷酸核糖转移酶 (HPRT) 基因、细胞腺嘌呤磷酸核糖转移酶 (APRT) 基因以及细菌胞嘧啶脱氨酶。

[0728] 在一些实施例中,基因修饰的细胞包含的聚核苷酸还包含能够在体外选择具有阴性选择表型的细胞的阳性标记物。阳性选择标记物可以是在引入宿主细胞中后表达显性表型的基因,所述显性表型允许对带有所述基因的细胞进行阳性选择。这种类型的基因在本领域中是已知的并且包括但不限于:赋予对潮霉素B (hygromycin B) 的抗性的潮霉素-B磷酸转移酶基因 (hph);编码对抗生素G418的抗性的来自Tn5的氨基糖苷磷酸转移酶基因 (neo或aph);二氢叶酸还原酶 (DHFR) 基因;腺苷脱氨酶基因 (ADA);以及多药耐药性 (MDR) 基因。

[0729] 在一个实施例中,阳性选择标记物和阴性选择元件被连接起来,使得阴性选择元件的丧失必然还伴随阳性选择标记物的丧失。在一个特定实施例中,阳性选择标记物与阴性选择标记物融合,使得其中之一的丧失强制性地导致另一个的丧失。产生赋予上述所希望阳性和阴性选择特征的多肽作为表达产物的融合聚核苷酸的一个实例是潮霉素磷酸转移酶胸苷激酶融合基因 (HyTK)。表达这一基因产生的多肽赋予潮霉素B抗性用于体阳性选择,以及赋予更昔洛韦敏感性用于体内阴性选择。还参见颁予S.D.Lupton的公开PCT US91/08442和PCT/US94/05601,描述了通过融合显性阳性选择标记物与阴性选择标记物得

到的双功能选择性融合基因的用途。

[0730] 优选的阳性选择标记物是来源于选自由以下组成的组的基因:hph、nco和gpt,并且优选的阴性选择标记物是来源于选自由以下组成的组的基因:胞嘧啶脱氨酶、HSV-I TK、VZV TK、HPRT、APRT和gpt。特定实施例中所涵盖的例示性双功能选择性融合基因包括但不限于阳性选择标记物来源于hph或neo,且阴性选择标记物来源于胞嘧啶脱氨酶或TK基因或选择性标记物的基因。

[0731] 在特定实施例中,编码一个或多个核酸酶变体、megaTAL、末端加工酶或融合多肽的聚核苷酸可以通过非病毒和病毒方法引入造血细胞,例如T细胞中。在特定实施例中,可以通过相同方法或通过不同方法,和/或通过相同载体或通过不同载体递送编码核酸酶和/或供体修复模板的一个或多个聚核苷酸。

[0732] 术语“载体”在本文中用以指能够转移或输送另一个核酸分子的核酸分子。被转移的核酸一般被连接至载体核酸分子,例如插入载体核酸分子中。载体可以包括引导在细胞中自主复制的序列,或可以包括足以允许整合至宿主细胞DNA中的序列。在特定实施例中,使用了非病毒载体将本文所涵盖的一个或多个聚核苷酸递送至T细胞中。

[0733] 非病毒载体的示例性实例包括但不限于质粒(例如DNA质粒或RNA质粒)、转座子、黏粒和细菌人工染色体。

[0734] 非病毒递送特定实施例中所涵盖的聚核苷酸的示例性方法包括但不限于:电穿孔、声致穿孔、脂质体转染、显微注射、基因枪法(biolistics)、病毒粒子、脂质体、免疫脂质体、纳米粒子、聚阳离子或脂质:核酸偶联物、裸DNA、人工病毒粒子、DEAE-葡聚糖介导的转移、基因枪(gene gun)以及热休克。

[0735] 适用于本文所涵盖的特定实施例中的聚核苷酸递送系统的示例性实例包括但不限于由Amaza Biosystems、Maxcyte, Inc.、BTX Molecular Delivery Systems、ThermoFisher Scientific和Copernicus Therapeutics Inc.所提供的那些。脂质体转染试剂是商售的(例如Transfectam™和Lipofectin™)。适于聚核苷酸的高效受体识别脂质体转染的阳离子性和中性脂质在文献中已有描述。参见例如Liu等人(2003)《基因疗法(Gene Therapy.)》10:180-187;和Balazs等人(2011)《药物递送杂志(Journal of Drug Delivery.)》2011:1-12。在特定实施例中,还涵盖抗体靶向的、细菌源性、基于纳米单元的非活体递送。

[0736] 包含特定实施例中所涵盖的聚核苷酸的病毒载体可以通过施用给各个患者,典型地如下所述,通过全身施用(例如静脉内、腹膜内、肌肉内、皮下或颅内输注)或表面施用进行体内递送。或者,可将载体递送至离体细胞,如从各个患者取出的细胞(例如流动的外周血液、淋巴细胞、骨髓穿刺液、组织活检样品等)或全适供血者造血干细胞,随后将所述细胞再植入患者体内。

[0737] 在一个实施例中,包含核酸酶变体和/或供体修复模板的病毒载体被直接施用给生物体以用于体内细胞转导。或者,可以施用裸DNA。施用是通过通常用于引入分子以与血液或组织细胞最终接触的途径中的任一种进行,包括但不限于注射、输注、表面施加和电穿孔。施用这类核酸的适合方法是本领域的技术人员可用且众所周知的,并且尽管可以使用多于一种途径施用特定组合物,但特定途径通常可以提供比另一途径更迅速且更有效的反应。

[0738] 适用于本文所涵盖的特定实施例中的病毒载体系统的示例性实例包括但不限于腺相关病毒(AAV)、逆转录病毒、单纯疱疹病毒、腺病毒和牛痘病毒载体。

[0739] 在各种实施例中,编码核酸酶变体和/或供体修复模板的一个或多个聚核苷酸是通过用包含所述一个或多个聚核苷酸的重组腺相关病毒(rAAV)转导造血细胞,例如T细胞来引入所述细胞中。

[0740] AAV是较小(约26nm)的复制缺陷型,主要是游离型的无包膜病毒。AAV可以感染分裂细胞和非分裂细胞两种并且可以将其基因组并入宿主细胞的基因组中。重组AAV(rAAV)典型地在最低限度上由转基因和其调控序列,以及5'和3' AAV反向末端重复(ITR)构成。ITR序列的长度是约145bp。包含二个ITR的rAAV载体具有约4.4kB的有效负载容量。

[0741] 自互补rAAV载体含有第三个ITR并且包装所述载体重组部分的二条链,仅留下约2.1kB用于本文所涵盖的聚核苷酸。在一个实施例中,AAV载体是一种scAAV载体。

[0742] 使用双rAAV载体策略已经实现了rAAV包装容量(约9kB)的大约两倍的扩充的包装容量。可用于制造本文所涵盖的rAAV的双载体策略包括但不限于剪接(反式剪接)、同源重组(重叠)或这两种方法的组合(混合体)。在双AAV反式剪接策略中,一个剪接供体(SD)信号被放置在5'-半载体的3'端并且一个剪接受体(SA)信号被放置在3'-半载体的5'端。在通过双AAV载体和反向末端重复(ITR)介导的两半载体头至尾连环化来共感染同一细胞时,反式剪接产生成熟mRNA和全尺寸蛋白质(Yan等人,2000)。反式剪接已被成功地用于表达肌肉和视网膜中的较大基因(Reich等人,2003;Lai等人,2005)。或者,双AAV载体中所含较大转基因表达盒的两半可以含有同源重叠序列(在5'-半载体的3'端和3'-半载体的5'端,双AAV重叠),这些序列将通过同源重组介导单一较大基因组的重构(Duan等人,2001)。这一策略取决于转基因重叠序列的致重组特性(Ghosh等人,2006)。第三种双AAV策略(混合体)是基于将来自外源基因(即,碱性磷酸酶;Ghosh等人,2008;Ghosh等人,2011)的高度致重组区域添加至反式剪接载体中。添加的区域被放置在5-半载体中SD信号的下流且在3'-半载体中S A信号的上游,以便增加双AAV之间的重组。

[0743] “混合AAV”、“混合rAAV”、“嵌合AAV”或“嵌合rAAV”是指用不同AAV血清型(并且优选地,来自一个或多个AAV ITR的不同血清型)的衣壳包装的rAAV基因组,并且可以另外地称为假型化rAAV。例如,rAAV 1型、2型、3型、4型、5型、6型、7型、8型、9型、10型、11型、12型、13型、14型、15型或16型基因组可以衣壳包装于AAV 1型、2型、3型、4型、5型、6型、7型、8型、9型、10型、11型、12型、13型、14型、15型或16型衣壳或其变体内,只要所述AAV衣壳和基因组(并且优选地,一个或多个AAV ITR)属于不同血清型即可。在某些实施例中,假型化rAAV粒子可以被称为“x/y”型,其中“x”指示ITR的来源并且“y”指示衣壳的血清型,例如2/5rAAV粒子具有来自AAV2的ITR和来自AAV6的衣壳。

[0744] 在特定实施例中,rAAV包含从AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV 12、AAV13、AAV 14、AAV15以及AAV16分离的ITR和衣壳序列。

[0745] 在一些实施例中,嵌合rAAV使用了从一种AAV血清型分离ITR序列和从不同AAV血清型分离的衣壳序列。例如,具有来源于AAV2的ITR序列和来源于AAV6的衣壳序列的rAAV称为AAV2/AAV6。在特定实施例中,rAAV载体可包含来自AAV2的ITR,以及来自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9或AAV10中任一种的衣壳蛋白。在一个优选实施例中,rAAV包含来源于AAV2的ITR序列和来源于AAV6的衣壳序列。在一个优选实施例中,rAAV

包含来源于AAV2的ITR序列和来源于AAV2的衣壳序列。

[0746] 在一些实施例中,可将工程改造和选择方法应用于AAV衣壳以使其更容易地转导目的细胞。

[0747] rAAV载体的构造、制造和其纯化已经公开于例如美国专利第9,169,494号;第9,169,492号;第9,012,224号;第8,889,641号;第8,809,058号;以及第8,784,799号中,这些专利各自以全文引用的方式并入本文中。

[0748] 在各种实施例中,编码核酸酶变体和/或供体修复模板的一个或多个聚核苷酸是通过用包含所述一个或多个聚核苷酸的逆转录病毒,例如慢病毒转导造血细胞来引入所述细胞中。

[0749] 如本文所使用,术语“逆转录病毒”是指将基因组RNA逆转录成线性双链DNA拷贝且随后将基因组DNA共价整合至宿主基因组中的RNA病毒。适用于特定实施例中的示例性逆转录病毒包括但不限于:莫洛尼鼠白血病毒(M-MuLV)、莫洛尼鼠肉瘤病毒(MoMSV)、哈维鼠(Harvey murine)肉瘤病毒(HaMuSV)、鼠乳腺肿瘤病毒(MuMTV)、长臂猿白血病毒(GaLV)、猫白血病毒(FLV)、泡沫病毒、弗兰德鼠(Friend murine)白血病毒、鼠干细胞病毒(MSCV)以及劳斯肉瘤病毒(RSV)和慢病毒。

[0750] 如本文所使用,术语“慢病毒”是指复杂逆转录病毒群(或属)。示例性慢病毒包括但不限于:人免疫缺陷病毒(HIV;包括1型HIV和2型HIV);维斯纳-梅迪病毒(visna-maedi virus,VMV)病毒;山羊关节炎-脑炎病毒(CAEV);马感染性贫血病毒(EIAV);猫免疫缺陷病毒(FIV);牛免疫缺陷病毒(BIV);以及猿猴免疫缺陷病毒(SIV)。在一个实施例中,基于HIV的载体主链(即,HIV顺式作用序列元件)是优选的。

[0751] 在各种实施例中,本文所涵盖的慢病毒载体包含一个或多个LTR,以及以下辅助元件中的一个或多个或全部:cPPT/FLAP、Psi(Ψ)包装信号、输出元件、聚腺苷酸序列,并且可以任选地包含如本文别处所论述的WPRE或HPRE、隔离子元件、选择性标记物和细胞自杀基因。

[0752] 在特定实施例中,本文所涵盖的慢病毒载体可以是整合型或非整合型或整合缺陷型慢病毒。如本文所使用,术语“整合缺陷型慢病毒”或“IDLV”是指具有不能将病毒基因组整合至宿主细胞基因组中的整合酶的慢病毒。不能整合的病毒载体在专利申请W0 2006/010834中已有描述,所述专利申请以全文引用的方式并入本文中。

[0753] HIV-1pol基因中适合于降低整合酶活性的示例性突变包括但不限于:H12N、H12C、H16C、H16V、S81R、D41A、K42A、H51A、Q53C、D55V、D64E、D64V、E69A、K71A、E85A、E87A、D116N、D116I、D116A、N120G、N120I、N120E、E152G、E152A、D35E、K156E、K156A、E157A、K159E、K159A、K160A、R166A、D167A、E170A、H171A、K173A、K186Q、K186T、K188T、E198A、R199c、R199T、R199A、D202A、K211A、Q214L、Q216L、Q221L、W235F、W235E、K236S、K236A、K246A、G247W、D253A、R262A、R263A以及K264H。

[0754] 在一个实施例中,HIV-1整合酶缺陷型pol基因包含D64V、D116I、D116A、E152G或E152A突变;D64V、D116I和E152G突变;或D64V、D116A和E152A突变。

[0755] 在一个实施例中,HIV-1整合酶缺陷型pol基因包含D64V突变。

[0756] 术语“长末端重复序列(LTR)”是指由位于逆转录病毒DNA两端处的碱基对形成的结构域,在其天然序列情形下是直接重复序列并且含有U3、R和U5区。

[0757] 如本文所使用,术语“FLAP元件”或“cPPT/FLAP”是指这样一种核酸,其序列包括逆转录病毒,例如HIV-1或HIV-2的中心多嘌呤段和中心终止序列(cPPT和CTS)。适合FLAP元件描述于美国专利第6,682,907号和Zennou等人,2000,《细胞(Cell)》,101:173中。在另一个实施例中,慢病毒载体含有在cPPT和/或CTS元件中具有一个或多个突变的FLAP元件。在又一个实施例中,慢病毒载体包含cPPT或CTS元件。在又一个实施例中,慢病毒载体不含cPPT或CTS元件。

[0758] 如本文所使用,术语“包装信号”或“包装序列”是指位于逆转录病毒基因组内的psi[Ψ]序列,这些序列是病毒RNA插入病毒衣壳或粒子中所需的,参见例如Clever等人,1995.《病毒学杂志(J.of Virology)》,第69卷,第4期;第2101-2109页。

[0759] 术语“输出元件”是指调控RNA转录物从细胞核向细胞质的输送的顺式作用转录后调控元件。RNA输出元件的实例包括但不限于人免疫缺陷病毒(HIV) rev反应元件(RRE)(参见例如Cullen等人,1991.《病毒学杂志》65:1053;和Cullen等人,1991.《细胞》58:423)和乙型肝炎病毒转录后调控元件(HPRE)。

[0760] 在特定实施例中,通过将转录后调控元件、高效的聚腺苷酸化位点和任选地转录终止信号并入病毒载体中,使这些载体中异源序列的表达增加。多种转录后调控元件可以在蛋白质水平上增加异源核酸的表达,例如土拨鼠肝炎病毒转录后调控元件(WPRE; Zufferey等人,1999,《病毒学杂志》,73:2886);存在于乙型肝炎病毒中的转录后调控元件(HPRE)(Huang等人,《分子与细胞生物学》,5:3864)等(Liu等人,1995,《基因与发育(Genes Dev.)》,9:1766)。

[0761] 慢病毒载体优选地含有因修饰LTR而引起的若干安全性增强。“自灭活”(SIN)载体是指复制缺陷型载体,例如其中称为U3区的右(3')LTR增强子-启动子区经过修饰(例如通过缺失或取代)以防止在第一轮病毒复制之外进行病毒转录。通过用异源启动子置换5'LTR的U3区以在病毒粒子产生期间驱动病毒基因组的转录,将提供额外安全性增强。可以使用的异源启动子的实例包括例如病毒猿猴病毒40(SV40)(例如早期或晚期)、巨细胞病毒(CMV)(例如立即早期)、莫洛尼鼠白血病毒(MoMLV)、劳氏肉瘤病毒(RSV)以及单纯疱疹病毒(HSV)(胸苷激酶)启动子。

[0762] 如本文所使用,术语“假型”或“假型化”是指病毒包膜蛋白已经被另一具有优选特征的病毒的病毒包膜蛋白取代的病毒。例如,HIV可以用水泡性口炎病毒G蛋白(VSV-G)包膜蛋白假型化,由此允许HIV感染更多的细胞,因为HIV包膜蛋白(由env基因编码)通常使病毒靶向CD4⁺展示细胞。

[0763] 在某些实施例中,慢病毒载体是根据已知方法制造。参见例如Kutner等人,《BMC生物技术(BMC Biotechnol.)》2009;9:10.doi:10.1186/1472-6750-9-10;Kutner等人.《自然-实验手册》(Nat.Protoc.)》2009;4(4):495-505.doi:10.1038/nprot.2009.22。

[0764] 根据本文所涵盖的某些特定实施例,大部分或所有的病毒载体主链序列都来源于慢病毒,例如HIV-1。然而,应理解,许多不同来源的逆转录病毒和/或慢病毒序列也可以使用,或可以在不削弱转移载体执行本文所述功能的能力的情况下对某些慢病毒序列进行组合并作出多种取代和改变。此外,多种慢病毒载体是本领域中已知的,参见Naldini等人(1996a,1996b和1998);Zufferey等人(1997);Dull等人,1998,美国专利第6,013,516号和第5,994,136号,其中有许多都适合于制造本文所涵盖的病毒载体或转移质粒。

[0765] 在各种实施例中,编码核酸酶变体和/或供体修复模板的一个或多个聚核苷酸是通过用包含所述一个或多个聚核苷酸的腺病毒转导造血细胞来引入所述细胞中。

[0766] 基于腺病毒的载体能够在许多细胞类型中实现极高转导效率且不需要细胞分裂。利用这类载体,已获得高效价和高水平表达。此载体可以在相对简单的系统中大量制造。大部分腺病毒载体都被工程改造成使得转基因置换Ad E1a、E1b和/或E3基因;随后,将复制缺陷型载体传送于以反式提供缺失的基因功能的人293细胞中。Ad载体可以在体内转导多种类型的组织,包括非分裂、分化细胞,如在肝、肾和肌肉中发现的那些。常规的Ad载体具有较大载运容量。

[0767] 当前复制缺乏型腺病毒载体的产生和传送可利用称为293的独特辅助细胞系,这种细胞系利用Ad5DNA片段转化人胚肾细胞且组成性表达E1蛋白质(Graham等人,1977)。由于E3区在腺病毒基因组中是可有可无的(Jones和Shenk,1978),故当前的腺病毒载体借助于293细胞在E1、D3或两个区域中携带外来DNA(Graham和Prevec,1991)。腺病毒载体已被用于真核基因表达(Levrero等人,1991;Gomez-Foix等人,1992)和疫苗研究(Grunhaus和Horwitz,1992;Graham和Prevec,1992)中。有关将重组腺病毒施用至不同组织的研究包括气管灌輸(Rosenfeld等人,1991;Rosenfeld等人,1992)、肌肉注射(Ragot等人,1993)、周边静脉内注射(Herz和Gerard,1993)以及立体定位接种至脑中(Le Gal La Salle等人,1993)。在临床试验中使用Ad载体的实例包含利用肌肉内注射进行抗肿瘤免疫接种的聚核苷酸疗法(Sterman等人,《人类基因治疗(Hum.Gene Ther.)》7:1083-9(1998))。

[0768] 在各种实施例中,编码核酸酶变体和/或供体修复模板的一个或多个聚核苷酸是通过用包含所述一个或多个聚核苷酸的单纯疱疹病毒,例如HSV-1、HSV-2转导造血细胞来引入所述细胞中。

[0769] 成熟HSV病毒粒子由二十面体衣壳包覆由152kb的线性双链DNA分子组成的病毒基因组组成。在一个实施例中,基于HSV的病毒载体缺乏一个或多个必需或非必需HSV基因。在一个实施例中,基于HSV的病毒载体是复制缺陷型的。大多数复制缺陷型HSV载体都含有缺失,移除了一个或多个立即早期、早期或晚期HSV基因以防止复制。例如,HSV载体可缺乏选自以下组成的组的立即早期基因:ICP4、ICP22、ICP27、ICP47和其组合。HSV载体的优势在于,它能够进入潜伏期以引起长期DNA表达并且其具有较大的病毒DNA基因组,由此可以容纳长达25kb的外源性DNA插入序列。基于HSV的载体描述于例如美国专利第5,837,532号、第5,846,782号和第5,804,413号,以及国际专利申请W0 91/02788、W0 96/04394、W0 98/15637和W0 99/06583中,其各自以全文引用的方式并入本文中。

[0770] H. 基因组编辑过的细胞

[0771] 由特定实施例中所涵盖的方法制造的基因组编辑过的细胞在PD-1基因中包含一个或多个基因编辑并且提供用于预防、治疗或改善癌症、GVHD、感染性疾病、自身免疫性疾病、免疫缺陷或其相关病况的至少一种症状的改良的基于细胞的治疗剂。不希望受任何特定理论束缚,相信本文所涵盖的组合物和方法部分地通过使治疗性细胞对免疫抑制信号和耗竭更具抗性来增加过继性细胞疗法的功效。

[0772] 特定实施例中所涵盖的基因组编辑过的细胞可以是自体/自体的(“自身的(self)”)或非自体的(“非自身的”,例如同种异体、同基因或异种的)。如本文所使用,“自体”是指来自同一受试者的细胞。如本文所使用,“同种异体”是指来自同一物种的在基因上

与比较细胞不同的细胞。如本文所使用,“同基因”是指来自不同受试者的在基因上与比较细胞相同的细胞。如本文所使用,“异种”是指来自与比较细胞不同的物种的细胞。在优选实施例中,细胞是从哺乳动物受试者获得。在一个更优选的实施例中,细胞是从灵长类动物受试者,任选地非人类灵长类动物获得。在最优选实施例中,细胞是从人类受试者获得。

[0773] “分离的细胞”是指从体内组织或器官获得且基本上不含细胞外基质的非天然存在的细胞,例如自然界中不存在的细胞、经过修饰的细胞、工程改造的细胞、重组细胞等。

[0774] 如本文所使用,术语“细胞群”是指可以由任何数目和/或任何组合的如本文别处所描述的同质或异质细胞类型构成的多个细胞。例如,对于转导T细胞,细胞群可以自末梢血液分离或获得。细胞群可以包含约10%、约20%、约30%、约40%、约50%、约60%、约70%、约80%、约90%或约100%的待编辑的靶细胞类型。在某些实施例中,T细胞可以使用本领域中已知的方法从异质细胞群分离或纯化。

[0775] 基因组可以使用本文所涵盖的组合物和方法编辑的细胞类型的示例性实例包括但不限于细胞系、原代细胞、干细胞、祖细胞和分化细胞,及其混合物。

[0776] 在一个优选实施例中,所述基因组编辑组合物和方法被用于编辑造血细胞,更优选地编辑免疫细胞并且甚至更优选地编辑T细胞。

[0777] 术语“T细胞”或“T淋巴细胞”是本领域公认的并且意图包括胸腺细胞、免疫效应细胞、调节性T细胞、原生T淋巴细胞、不成熟T淋巴细胞、成熟T淋巴细胞、静息T淋巴细胞或活化T淋巴细胞。T细胞可以是T辅助(Th)细胞,例如T辅助1(Th1)或T辅助2(Th2)细胞。T细胞可以是辅助T细胞(HTL;CD4⁺T细胞)、CD4⁺T细胞、细胞毒性T细胞(CTL;CD8⁺T细胞)、肿瘤浸润细胞毒性T细胞(TIL;CD8⁺T细胞)、CD4⁺CD8⁺T细胞、CD4⁻CD8⁻T细胞或任何其它T细胞子集。在一个实施例中,T细胞是免疫效应细胞。在一个实施例中,T细胞是NKT细胞。适用于特定实施例中的其它示例性T细胞群包括原生T细胞和记忆T细胞。

[0778] 在各种实施例中,基因组编辑过的细胞包括含由本文所涵盖的组合物和方法编辑的PD-1基因的免疫效应细胞。“免疫效应细胞”是免疫系统具有一种或多种效应功能(例如细胞毒性细胞杀灭活性、细胞因子分泌、ADCC和/或CDC的诱导)的任何细胞。特定实施例中所涵盖的示例性免疫效应细胞是T淋巴细胞,特别是细胞毒性T细胞(CTL;CD8⁺T细胞)、TIL和辅助T细胞(HTL;CD4⁺T细胞)。在一个实施例中,免疫效应细胞包括自然杀手(NK)细胞。在一个实施例中,免疫效应细胞包括自然杀手T(NKT)细胞。

[0779] 在特定实施例中,“强效T细胞”和“幼T细胞”可互换使用并且是指T细胞能够增殖并伴随分化减少的T细胞表型。在特定实施例中,幼T细胞具有“原生T细胞”的表型。在特定实施例中,幼T细胞包含以下生物标记物中的一种或多种,或全部:CD62L、CCR7、CD28、CD27、CD122、CD127、CD197和CD38。在一个实施例中,幼T细胞包含以下生物标记物中的一种或多种,或全部:CD62L、CD127、CD197和CD38。在一个实施例中,幼T细胞不表达CD57、CD244、CD160、PD-1、CTLA4、PD-1和LAG3。

[0780] T细胞可以从多种来源获得,包括但不限于周边血液单核细胞、骨髓、淋巴结组织、脐血、胸腺组织、来自感染部位的组织、腹水、肋膜积液、脾组织以及肿瘤。

[0781] 在特定实施例中,包含免疫效应细胞或T细胞的细胞群包含编辑的PD-1基因,其中所述编辑是通过NHEJ修复的DSB。在特定实施例中,免疫效应细胞或T细胞包含编辑的PD-1基因,其中所述编辑是通过NHEJ修复的DSB。在特定实施例中,所述编辑是PD-1基因的编码

序列中、优选地在PD-1基因的外显子5、外显子1或外显子2中,更优选在PD-1基因外显子5中的SEQ ID NO:25(或SEQ ID NO:27)处、在PD-1基因外显子1中的SEQ ID NO:30(或SEQ ID NO:32)处或在PD-1基因外显子2中的SEQ ID NO:35(或SEQ ID NO:37)处约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个或更多个核苷酸的插入或缺失(INDEL)。

[0782] 在一个特定实施例中,所述编辑是PD-1基因的编码序列中、优选地在PD-1基因的外显子5、外显子1或外显子2中,更优选在PD-1基因外显子5中的SEQ ID NO:25(SEQ ID NO:27)处、在PD-1基因外显子1中的SEQ ID NO:30(或SEQ ID NO:32)处或在PD-1基因外显子2中的SEQ ID NO:35(或SEQ ID NO:37)处+1、-1、-2、-3或-4个核苷酸的缺失。

[0783] 在特定实施例中,包含免疫效应细胞或T细胞的细胞群包含在通过HDR修复的DSB处并入的供体修复模板的编辑过的PD-1基因。

[0784] 在特定实施例中,包含免疫效应细胞或T细胞的细胞群包括含供体修复模板的编辑过的PD-1基因,所述供体修复模板包含PD-1基因或其部分并且被设计成用于在基因组PD-1序列中引入一个或多个突变以改变PD-1表达或信号传导,并且优选地,减少或消除PD-1表达和/或信号传导。

[0785] 在各种实施例中,基因组编辑过的细胞包含在PD-1基因中的编辑并且还包含编码PD-1翻转受体、双特异性T细胞接合分子(BiTE);细胞因子(例如IL-2、胰岛素、IFN- γ 、IL-7、IL-21、IL-10、IL-12、IL-15以及TNF- α)、趋化因子(例如MIP-1 α 、MIP-1 β 、MCP-1、MCP-3以及RANTES)、细胞毒素(例如穿孔素、颗粒酶A和颗粒酶B)、细胞因子受体(例如IL-2受体、IL-7受体、IL-12受体、IL-15受体和IL-21受体)或工程改造的抗原受体(例如工程改造的T细胞受体(TCR)、嵌合抗原受体(CAR)、Daric受体或其组分,或嵌合细胞因子受体)的聚核苷酸。在一个实施例中,包含所述聚核苷酸和核酸酶变体的供体修复模板被引入细胞中并且所述聚核苷酸通过HDR修复来并入细胞基因组中PD-1基因中的DSB位点处。所述聚核苷酸还可以例如通过用包含所述聚核苷酸的载体转导细胞来引入细胞中除PD-1基因外的位点处。

[0786] I. 组合物和配制物

[0787] 特定实施例中所涵盖的组合物可以包含本文所涵盖的一种或多种多肽、聚核苷酸、包含其的载体以及基因组编辑组合物和基因组编辑过的细胞组合物。特定实施例中所涵盖的基因组编辑组合物和方法可用于编辑细胞或细胞群中人程序性细胞死亡蛋白1(PD-1)基因中的靶位点。在优选实施例中,基因组编辑组合物被用于编辑造血细胞,例如T细胞或免疫效应细胞中的PD-1基因。

[0788] 在各种实施例中,本文所涵盖的组合物包含核酸酶变体,并且任选地包含末端加工酶,例如3'-5'核酸外切酶(Trex2)。核酸酶变体可呈mRNA形式,以通过前述所公开的聚核苷酸递送方法,例如电穿孔、脂质纳米粒子等引入细胞中。在一个实施例中,组合物是通过前述所公开的聚核苷酸递送方法引入细胞中,所述组合物包含编码归巢核酸内切酶变体或megaTAL和任选地3'-5'核酸外切酶的mRNA。可使用所述组合物,通过易错NHEJ产生基因组编辑过的细胞或基因组编辑过的细胞群。

[0789] 在各种实施例中,本文所涵盖的组合物包含供体修复模板。所述组合物可以被递送至表达或将表达核酸酶变体和任选地末端加工酶的细胞。在一个实施例中,所述组合物可以被递送至表达或将表达归巢核酸内切酶变体或megaTAL和任选地3'-5'核酸外切酶的

细胞。可以使用在供体修复模板存在下基因编辑酶的表达,通过HDR产生基因组编辑过的细胞或基因组编辑过的细胞群。

[0790] 在特定实施例中,本文所涵盖的组合物包含细胞群、核酸酶变体和任选地,供体修复模板。在特定实施例中,本文所涵盖的组合物包含细胞群、核酸酶变体、末端加工酶和任选地,供体修复模板。核酸酶变体和/或末端加工酶可呈mRNA形式,以通过前述所公开的聚核苷酸递送方法引入细胞中。

[0791] 在特定实施例中,本文所涵盖的组合物包含细胞群、归巢核酸内切酶变体或megaTAL,和任选地,供体修复模板。在特定实施例中,本文所涵盖的组合物包含细胞群、归巢核酸内切酶变体或megaTAL、3'-5'核酸外切酶,和任选地,供体修复模板。归巢核酸内切酶变体、megaTAL和/或3'-5'核酸外切酶可呈mRNA形式,以通过前述所公开的聚核苷酸递送方法引入细胞中。

[0792] 在特定实施例中,细胞群包含经过基因修饰的免疫效应细胞。

[0793] 组合物包括但不限于药物组合物。“药物组合物”是指被配制成药学上可接受的或生理学上可接受的溶液以单独或与一种或多种其它治疗模式组合施用给细胞或动物的组合物。还应理解,必要时,所述组合物还可以与其它试剂,如细胞因子、生长因子、激素、小分子、化学治疗剂、前药、药物、抗体或其它各种药物活性剂组合施用。对于所述组合物中还可以包括的其它组分几乎无限制,只要这些另外的试剂不会不利地影响所述组合物即可。

[0794] 短语“药理学上可接受的”在本文中用以指在合理医学判断范围内,适于与人类和动物的组织接触使用而无过量毒性、刺激、过敏反应或其它问题或并发症,与合理的效益/风险比相称的那些化合物、物质、组合物和/或剂型。

[0795] 术语“药理学上可接受的载体”是指与治疗性细胞一起施用的稀释剂、佐剂、赋形剂或媒剂。药物载体的示例性实例可以是无菌液体,如细胞培养基、水和油,包括石油、动物、植物或合成来源的油,如花生油、大豆油、矿物油、芝麻油等。还可以采用生理盐水溶液和右旋糖水溶液以及甘油溶液作为液体载体,特别是用于可注射溶液。在特定实施例中,适合的药物赋形剂包括淀粉、葡萄糖、乳糖、蔗糖、明胶、麦芽、稻谷、面粉、白垩、硅胶、硬脂酸钠、单硬脂酸甘油酯、滑石、氯化钠、脱脂奶粉、甘油、丙烯、乙二醇、水、乙醇等。除非任何常规介质或试剂与活性成分不相容,否则考虑将其用于治疗组合物中。还可将补充性活性成分并入所述组合物中。

[0796] 在一个实施例中,包含药理学上可接受的载体的组合物适于施用给受试者。在一个实施例中,包含载体的组合物适合肠胃外施用,例如血管内(静脉内或动脉内)、腹膜内或肌肉内施用。在特定实施例中,包含药理学上可接受的载体的组合物适于室内、脊柱内或鞘内施用。药理学上可接受的载体包括无菌水溶液、细胞培养基或分散液。这些介质和药剂在药理学上活性物质中的使用是本领域中众所周知的。除非任何常规介质或试剂与被转导的细胞不相容,否则考虑将其用于药物组合物中。

[0797] 在特定实施例中,本文所涵盖的组合物包含经过基因修饰的T细胞和药理学上可接受的载体。包含本文所涵盖的基于细胞的组合物的组合物可以通过肠内或肠胃外施用方法单独施用或与其它适合的化合物组合施用以实现所希望的治疗目标。

[0798] 药理学上可接受的载体必须具有足够高的纯度和足够低的毒性,以使其适合施用给所治疗的人类受试者。它还应该保持或增加组合物的稳定性。药理学上可接受的载体可以是

液体或固体,并且考虑到计划的施用方式,选择的药学上可接受的载体当与组合物的其它组分组合时可提供所希望的整体(bulk)、稠度等。例如,药学上可接受的载体可以是但不限于粘合剂(例如预胶凝化玉米淀粉、聚乙烯吡咯烷酮或羟丙基甲基纤维素等)、填充剂(例如乳糖和其它糖、微晶纤维素、果胶、明胶、硫酸钙、乙基纤维素、聚丙烯酸酯、磷酸氢钙等)、润滑剂(例如硬脂酸镁、滑石、二氧化硅、胶态二氧化硅、硬脂酸、金属硬脂酸盐、氢化植物油、玉米淀粉、聚乙二醇、苯甲酸钠、乙酸钠等)、崩解剂(例如淀粉、羟基乙酸淀粉钠等)或润湿剂(例如月桂基硫酸钠等)。用于本文所涵盖的组合物的其它适合的药学上可接受的载体包括但不限于水、盐溶液、醇、聚乙二醇、明胶、直链淀粉、硬脂酸镁、滑石、硅酸、粘性石蜡、羟甲基纤维素、聚乙烯基吡咯烷酮等。

[0799] 此类载体溶液还可含有缓冲剂、稀释剂和其它适合的添加剂。如本文所使用,术语“缓冲剂”是指化学组成能中和酸或碱但不会显著改变pH的溶液或液体。本文所涵盖的缓冲剂的实例包括但不限于杜尔贝科氏磷酸盐缓冲生理盐水(Dulbecco's phosphate buffered saline,PBS)、林格氏溶液(Ringer's solution)、5%右旋糖水溶液(D5W)、生理盐水(normal/physiologic saline)(0.9%NaCl)。

[0800] 药学上可接受的载体的存在量足以将组合物的pH值维持在约7。或者,所述组合物的pH值在约6.8至约7.4范围内,例如是6.8、6.9、7.0、7.1、7.2、7.3以及7.4。在又一个实施例中,所述组合物的pH值是约7.4。

[0801] 本文所涵盖的组合物可以包含无毒的药学上可接受的介质。组合物可以是悬浮液。如本文所使用,术语“悬浮液”是指非贴壁情形,其中细胞不附着于固体载体。例如,可以搅拌或搅动保持悬浮状态的细胞并使其不粘附到载体,如培养皿。

[0802] 在特定实施例中,本文所涵盖的组合物被配制成悬浮液形式,其中基因组编辑的T细胞被分散于装在静脉内(IV)袋等中的可接受的液体介质或溶液,例如生理盐水或无血清培养基内。可接受的稀释剂包括但不限于水、PlasmaLyte、林格氏溶液、等张氯化钠(生理盐水)溶液、无血清细胞培养基以及适合低温储存的培养基,例如Cryostor®培养基。

[0803] 在某些实施例中,药学上可接受的载体基本上不含人或动物来源的天然蛋白质,并且适于储存包含基因组编辑的T细胞群的组合物。治疗性组合物意图施用给人类患者,并因此基本上不含细胞培养物组分,如牛血清白蛋白、马血清和胎牛血清。

[0804] 在一些实施例中,组合物是在药学上可接受的细胞培养基中配制。这类组合物适合施用给人类受试者。在特定实施例中,药学上可接受的细胞培养基是无血清培养基。

[0805] 与含有血清的培养基相比,无血清培养基具有几个优点,包括组成简化且更明确、污染程度降低、潜在感染物来源消除并且成本更低。在各种实施例中,无血清培养基是无动物成分的,并且可以任选地是无蛋白质的。任选地,所述培养基可以含有生物药学上可接受的重组蛋白。“无动物成分”培养基是指组分来源于非动物来源的培养基。重组蛋白置换非动物培养基中的原生动物蛋白且养分是从合成、植物或微生物来源获得。相比之下,“无蛋白质”培养基被定义为基本上不含蛋白质。

[0806] 特定组合物中所使用的无血清培养基的示例性实例包括但不限于QBSF-60(Quality Biological, Inc.)、StemPro-34(Life Technologies)和X-VIVO 10。

[0807] 在一个优选实施例中,包含基因组编辑的T细胞的组合物是在PlasmaLyte中配制。

[0808] 在各种实施例中,包含基因组编辑的T细胞的组合物是在低温保存培养基中配制。

例如,可以使用含低温保存剂的低温保存培养基在解冻后维持高细胞活力结果。特定组合物中所使用的低温保存培养基的示例性实例包括但不限于CryoStor CS10、CryoStor CS5和CryoStor CS2。

[0809] 在一个实施例中,所述组合物是在包含50:50PlasmaLyte A:CryoStor CS10的溶液中配制。

[0810] 在特定实施例中,所述组合物基本上不含支原体、内毒素和微生物污染。对于内毒素而言,“基本上不含”意思指每剂细胞中的内毒素少于FDA关于生物制剂所允许的内毒素,即每天每千克体重5EU总内毒素,对于平均70kg的人来说,就是350EU/总细胞剂量。在特定实施例中,包含用本文所涵盖的反转录病毒载体转导的造血干细胞或祖细胞的组合物含有约0.5EU/mL至约5.0EU/mL、或约0.5EU/mL、1.0EU/mL、1.5EU/mL、2.0EU/mL、2.5EU/mL、3.0EU/mL、3.5EU/mL、4.0EU/mL、4.5EU/mL或5.0EU/mL。

[0811] 在某些实施例中,所涵盖的适于递送聚核苷酸的组合物和配制物包括但不限于编码一种或多种重编程的核酸酶以及任选地末端加工酶的一个或多个mRNA。

[0812] 用于离体递送的示例性配制物还可以包括使用本领域中已知的各种转染剂,如磷酸钙、电穿孔、热休克以及各种脂质体制剂(即,脂质介导的转染)。如下文更详细地描述,脂质体是包覆一部分含水流体的脂质双层。DNA自发地与阳离子脂质体的外表面缔合(借助于其电荷),并且这些脂质体将与细胞膜相互作用。

[0813] 在特定实施例中,药学上可接受的载体溶液的配制物,以及在多种治疗方案中使用本文所描述的特定组合物的适合剂量和治疗方案的研究是本领域的技术人员众所周知的,包括例如肠内和肠胃外,例如血管内、静脉内、动脉内、骨内、室内、脑内、颅内、脊柱内、鞘内以及髓内施用和配制物。本领域技术人员应理解,本文所涵盖的特定实施例可以包含其它配制物,如药学领域中众所周知的那些,并且描述于例如《雷明顿:制药科学和实践(Remington:The Science and Practice of Pharmacy)》,第I卷和第II卷,第22版,Loyd V.Allen Jr.编辑,宾夕法尼亚州费城(Philadelphia,PA):Pharmaceutical Press;2012,其以全文引用的方式并入本文中。

[0814] J.基因组编辑过的细胞的疗法

[0815] 由本文所涵盖的组合物和方法制造的基因组编辑过的细胞提供了用于预防、治疗或改善癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病或免疫缺陷的至少一种症状的改良的药品。如本文所使用,术语“药品”是指使用本文所涵盖的组合物和方法制造的经过基因修饰的细胞。在特定实施例中,所述药品包含经过基因编辑的免疫效应细胞或T细胞。此外,特定实施例中所涵盖的基因组编辑的T细胞提供更安全且更有效的过继性细胞疗法,因为这些细胞对T细胞耗竭具有抗性并且在肿瘤微环境展示出增加的耐久性和持久性,使得可以持续进行治疗。

[0816] 在特定实施例中,将有效量的包含编辑过的PD-1基因的基因组编辑的免疫效应细胞或T细胞施用给受试者以预防、治疗或改善癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病或免疫缺陷的至少一种症状。

[0817] 在特定实施例中,PD-1编辑的细胞基本上不表达或不表达PD-1并因此缺乏或基本上缺乏功能性PD-1表达,例如缺乏增加T细胞耗竭和抑制促炎性细胞因子表达的能力。在特定实施例中,缺乏PD-1的基因组编辑过的免疫效应细胞对来自肿瘤微环境的免疫抑制信号

更具抗性并且对T细胞耗竭展示出增加的持久性和抗性。

[0818] 在特定实施例中,预防、治疗或改善癌症的至少一种症状的方法包含向受试者施用有效量的包含编辑过的PD-1基因和工程改造的TCR、CAR或Daric或其它治疗性转基因的基因组编辑的免疫效应细胞或T细胞以将细胞重新引导至肿瘤或癌症。经过基因修饰的细胞是更耐久且更持久的药品,因为这些细胞由于编辑PD-1基因以减少或消除PD-1表达而对来自肿瘤微环境的免疫抑制信号更具抗性。

[0819] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗实体肿瘤或癌症。

[0820] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗实体肿瘤或癌症,包括但不限于:肾上腺癌、肾上腺皮质癌、肛门癌、阑尾癌、星形细胞瘤、非典型畸胎样/横纹肌样肿瘤、基底细胞癌、胆管癌、膀胱癌、骨癌、脑/CNS癌症、乳癌、支气管肿瘤、心脏肿瘤、子宫颈癌、胆管癌、软骨肉瘤、脊索瘤、结肠癌、结肠直肠癌、颅咽管瘤、乳腺管原位癌(DCIS)、子宫内膜癌、室管膜瘤、食道癌、嗅神经母细胞瘤、尤文氏肉瘤(Ewing's sarcoma)、颅外生殖细胞肿瘤、性腺外生殖细胞肿瘤、眼癌、输卵管癌、纤维组织肉瘤、纤维肉瘤、胆囊癌、胃癌、胃肠类癌、胃肠道间质瘤(GIST)、生殖细胞肿瘤、神经胶质瘤、神经胶母细胞瘤、头颈癌、血管母细胞瘤、肝细胞癌、下咽癌、眼内黑素瘤、卡波西氏肉瘤(kaposi sarcoma)、肾癌、喉癌、平滑肌肉瘤、唇癌、脂肪肉瘤、肝癌、肺癌、非小细胞肺癌、肺类癌肿瘤、恶性间皮瘤、髓性癌、神经管母细胞瘤、脑膜瘤、黑素瘤、梅克尔氏细胞癌(Merkel cell carcinoma)、中线道癌、口部癌症(mouth cancer)、黏液肉瘤、骨髓发育不良综合症、骨髓增生赘瘤、鼻腔和副鼻窦癌症、鼻咽癌、神经母细胞瘤、少突神经胶质瘤、口部癌症(oral cancer)、口腔癌(oral cavity cancer)、口咽癌、骨肉瘤、卵巢癌、胰腺癌、胰岛细胞瘤、乳头状癌、副神经节瘤、副甲状腺癌、阴茎癌、咽癌、嗜铬细胞瘤、松果体瘤、垂体肿瘤、胸膜肺母细胞瘤、原发性腹膜癌、前列腺癌、直肠癌、视网膜母细胞瘤、肾细胞癌、肾盂和输尿管癌症、横纹肌肉瘤、唾液腺癌症、皮脂腺癌、皮肤癌、软组织肉瘤、鳞状细胞癌、小细胞肺癌、小肠癌、胃癌、汗腺癌、滑膜瘤、睾丸癌、喉癌、胸腺癌、甲状腺癌、尿道癌、子宫癌子宫肉瘤、阴道癌、血管癌、外阴癌以及威尔姆斯氏肿瘤(Wilms Tumor)。

[0821] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗实体肿瘤或癌症,包括但不限于肝癌、胰腺癌、肺癌、乳癌、膀胱癌、脑癌、骨癌、甲状腺癌、肾癌或皮肤癌。

[0822] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗各种癌症,包括但不限于胰腺癌、膀胱癌和肺癌。

[0823] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗液体癌或血液癌。

[0824] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗B细胞恶性病,包括但不限于:白血病、淋巴瘤和多发性骨髓瘤。

[0825] 在特定实施例中,本文所涵盖的基因组编辑过的细胞被用于治疗液体癌,包括但不限于白血病、淋巴瘤和多发性骨髓瘤:急性淋巴细胞性白血病(ALL)、急性骨髓性白血病(AML)、骨髓母细胞白血病、早幼粒细胞性白血病、骨髓单核细胞性白血病、单核细胞性白血病、红白血病、毛细胞白血病(HCL)、慢性淋巴细胞性白血病(CLL)和慢性骨髓白血病(CML)、慢性骨髓单核细胞性白血病(CMML)和真性红细胞增多症、霍奇金氏淋巴瘤(Hodgkin

lymphoma)、结节性淋巴细胞为主型霍奇金氏淋巴瘤、伯基特氏淋巴瘤 (Burkitt lymphoma)、小淋巴细胞性淋巴瘤 (SLL)、弥漫性大B细胞淋巴瘤、滤泡性淋巴瘤、成免疫细胞性大细胞淋巴瘤、前体B成淋巴细胞性淋巴瘤、套细胞淋巴瘤、边缘区淋巴瘤、蕈样真菌病、多形性大细胞淋巴瘤、塞扎莱氏综合症 (Sézary syndrome)、前体T成淋巴细胞性淋巴瘤、多发性骨髓瘤、明显多发性骨髓瘤、郁积性多发性骨髓瘤、浆细胞白血病、非分泌性骨髓瘤、IgD骨髓瘤、骨硬化性骨髓瘤、孤立性骨浆细胞瘤以及髓外浆细胞瘤。

[0826] 用于本文所涵盖的基因组编辑方法中的优选的细胞包括自体/自体的 (“自身的”) 细胞, 优选造血细胞, 更优选T细胞, 并且更优选免疫效应细胞或Treg细胞。

[0827] 在特定实施例中, 提供了向有需要患者单独或与一种或多种治疗剂组合施用治疗有效量的本文所涵盖的基因组编辑过的细胞或包含这些细胞的组合物的方法。在某些实施例中, 所述细胞被用于治疗有发展癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病或免疫缺陷的风险的患者。因此, 特定实施例包含治疗或预防或改善癌症、感染性疾病、自身免疫性疾病、炎性疾病或免疫缺陷的至少一种症状, 其包含向有需要的受试者施用治疗有效量的本文所涵盖的基因组编辑过的细胞。

[0828] 在一个实施例中, 治疗有需要的受试者的癌症、GVHD、感染性疾病、自身免疫性疾病、炎性疾病或免疫缺陷的方法包含施用有效量, 例如治疗有效量的包含本文所涵盖的基因组编辑过的细胞的组合物。施用的量和频率将由如患者的状况以及患者疾病的类型和严重程度等因素决定, 不过适当剂量可以通过临床试验确定。

[0829] 在一个示例性实施例中, 提供给受试者的基因组编辑过的细胞的有效量是至少 2×10^6 个细胞/kg、至少 3×10^6 个细胞/kg、至少 4×10^6 个细胞/kg、至少 5×10^6 个细胞/kg、至少 6×10^6 个细胞/kg、至少 7×10^6 个细胞/kg、至少 8×10^6 个细胞/kg、至少 9×10^6 个细胞/kg、或至少 10×10^6 个细胞/kg、或每公斤更多个细胞, 包括所有中间细胞剂量。

[0830] 在另一个示例性实施例中, 提供给受试者的基因组编辑过的细胞的有效量是约 2×10^6 个细胞/kg、约 3×10^6 个细胞/kg、约 4×10^6 个细胞/kg、约 5×10^6 个细胞/kg、约 6×10^6 个细胞/kg、约 7×10^6 个细胞/kg、约 8×10^6 个细胞/kg、约 9×10^6 个细胞/kg、或约 10×10^6 个细胞/kg、或每公斤更多个细胞, 包括所有中间细胞剂量。

[0831] 在另一示例性实施例中, 提供给受试者的基因组编辑过的细胞的有效量是约 2×10^6 个细胞/kg至约 10×10^6 个细胞/kg、约 3×10^6 个细胞/kg至约 10×10^6 个细胞/kg、约 4×10^6 个细胞/kg至约 10×10^6 个细胞/kg、约 5×10^6 个细胞/kg至约 10×10^6 个细胞/kg、 2×10^6 个细胞/kg至约 6×10^6 个细胞/kg、 2×10^6 个细胞/kg至约 7×10^6 个细胞/kg、 2×10^6 个细胞/kg至约 8×10^6 个细胞/kg、 3×10^6 个细胞/kg至约 6×10^6 个细胞/kg、 3×10^6 个细胞/kg至约 7×10^6 个细胞/kg、 3×10^6 个细胞/kg至约 8×10^6 个细胞/kg、 4×10^6 个细胞/kg至约 6×10^6 个细胞/kg、 4×10^6 个细胞/kg至约 7×10^6 个细胞/kg、 4×10^6 个细胞/kg至约 8×10^6 个细胞/kg、 5×10^6 个细胞/kg至约 6×10^6 个细胞/kg、 5×10^6 个细胞/kg至约 7×10^6 个细胞/kg、 5×10^6 个细胞/kg至约 8×10^6 个细胞/kg、或 6×10^6 个细胞/kg至约 8×10^6 个细胞/kg, 包括所有中间细胞剂量。

[0832] 本领域的普通技术人员应认识到, 可能需要多次施用特定实施例中所涵盖的组合物来实行所希望的疗法。例如, 组合物可以在1周、2周、3周、1个月、2个月、3个月、4个月、5个月、6个月、1年、2年、5年、10年或更长时间范围内施用1、2、3、4、5、6、7、8、9或10次或更多次。

[0833] 在某些实施例中,可能需要向受试者施用活化的T细胞且随后再抽取血液(或进行单采血液成分术),活化其中的T细胞,并将这些活化并且扩增的T细胞再输注给患者。这一过程可以每数周执行多次。在某些实施例中,可以使10cc到400cc的抽血中的T细胞活化。在某些实施例中,使20cc、30cc、40cc、50cc、60cc、70cc、80cc、90cc、100cc、150cc、200cc、250cc、300cc、350cc或400cc或更多的抽血中的T细胞活化。不受理论束缚,使用这种多次抽血/多次再输注方案可以用于选出某些T细胞群。

[0834] 特定实施例中所涵盖的组合物的可以通过任何适宜的方式施用,包括通过气雾剂吸入、注射、摄取、输注、植入或移植。在一个优选实施例中,组合物是通过肠胃外施用。如本文所使用,短语“肠胃外施用(parenteral administration)”或“通过肠胃外施用(administered parenterally)”是指除经肠和局部施用以外的施用模式,通常通过注射进行并且包括但不限于血管内、静脉内、肌肉内、动脉内、鞘内、囊内、眶内、瘤内、心内、皮内、腹膜内、经气管、皮下、表皮下、关节内、囊下、蛛网膜下、脊柱内以及胸骨内注射和输注。在一个实施例中,本文所涵盖的组合物是通过直接注射至肿瘤、淋巴结或感染部位中来施用给受试者。

[0835] 在一个实施例中,治疗被诊断患有癌症的受试者的方法包含从受试者体内取出免疫效应细胞,编辑所述免疫效应细胞的基因组并产生基因组编辑过的免疫效应细胞群,以及向同一受试者施用所述基因组编辑过的免疫效应细胞群。在一优选实施例中,免疫效应细胞包含T细胞。

[0836] 用于施用特定实施例中所涵盖的细胞组合物的方法包括可有效再引入离体基因组编辑的免疫效应细胞,或再引入基因组编辑的免疫效应细胞的祖细胞,使其在引入受试者体内后分化成成熟免疫效应细胞的任何方法。一种方法包含离体基因组编辑末梢血液T细胞并将被转到的细胞回输至受试者体内。

[0837] 本说明书中引用的所有出版物、专利申请和授权专利都以引用的方式并入本文中,其引用的程度如同特定且个别地指示每一个别出版物、专利申请或授权专利以引用的方式并入一般。

[0838] 尽管已经出于清楚理解的目的,通过说明和实例相当详细地描述了前述实施例,但根据本文所涵盖的传授内容,本领域的普通技术人员将易于了解,可以在不脱离所附权利要求书的精神或范围的情况下对其作出某些改变和修改。提供以下实例仅作为说明并且不具有限制性。本领域的技术人员将容易认识到多种可以被改变或修改以产生基本上类似的结果的非关键参数。

[0839] 实例

[0840] 实例1

[0841] 对I-ONUI重编程以破坏程序性死亡受体-1(PD-1)基因内的细胞内信号传导基元

[0842] PD-1是在抗原受体刺激和活化之后于T细胞质膜上表达的。PD-1包含信号肽、细胞外IgV样结构域、跨膜结构域以及含有免疫受体酪氨酸抑制基元(ITIM,共同序列S/I/V/LxYxxI/V/L)和免疫受体酪氨酸转换基元(ITSM,共同序列TxYxxV/I)的细胞内尾。图1A。在PD-L1/2配体结合且同时T细胞活化后,在PD-1ITSM内氨基酸位置248处的酪氨酸变得磷酸化,产生含SH2结构域的蛋白酪氨酸磷酸酶-2(SHP2)的结合底物(参见Chemnitz JM等人,《免疫学杂志(J Immunol.》)2004年7月15日;173(2):945-54。)。SHP2募集至质膜将抵消T细

胞中磷酸酪氨酸驱动的活化信号(Yokosuka T等人,《实验医学杂志(J Exp Med.)》2012年6月4日;209(6):1201-17)并抑制T细胞活化状态的持续时间。典型I-OnuI“中心-4”裂解基元ATAC涵盖磷酸化ITSM酪氨酸的密码子。图1B。已开发出一种归巢核酸内切酶变体,其靶向以PD-1基因外显子5中这一中心-4基元为中心的22bp靶序列(SEQ ID NO:25)。

[0843] 不希望受任何特定理论束缚,预期在ATAC中心-4序列中和附近的所有推定的插入/缺失事件(‘indel’)将完全消除ITSM基元的一个或多个必要特征,其中大部分看来可信的indel会破坏酪氨酸-248密码子本身。这些indel很可能产生显性负性PD-1蛋白质,这些蛋白质包含正常细胞外结构域和非功能性细胞内信号传导结构域。显性负性PD-1蛋白质可以用作PD-1配体的‘接收器’,由此减少或消除免疫抑制信号传导。此外,由于活化T细胞上调PD-L1的表达;由于在T细胞间发生顺式或反式PD-1:PD-L1相互作用;以及由于这些相互作用对于T细胞功能极为重要,故在不影响表达情况下进行基因编辑以破坏PD-1信号传导可以保持推定的PD-L1驱动的功能。

[0844] 因此,通过构建在DNA识别界面中含有可变氨基酸残基的模块文库,将I-OnuI重编程成靶向ITSM编码区。为了构建这些变体,使用寡核苷酸将简并密码子并入I-OnuI DNA结合结构域中。使用编码简并密码子的寡核苷酸作为PCR模板,在酵母菌株酿酒酵母(*S.cerevisiae*)中通过空隙重组(gap recombination)产生变体文库。每个变体文库跨N末端或C末端I-OnuI DNA识别结构域并且含有约 10^7 至 10^8 个特有转化体。如图2所示,通过流式细胞术,关于针对包含相应结构域的“半位点”的靶位点(SEQ ID NO:28-29)的裂解活性筛选由此得到的表面展示文库。

[0845] 对展示N末端和C末端结构域重编程的I-OnuI HE的酵母进行纯化并提取出质粒DNA。进行PCR反应以扩增重编程的结构域,随后将其转化至酿酒酵母中以建立重编程结构域组合的文库。从这一文库中鉴别出识别PD-1基因外显子5中的ITSM编码区中存在的完整靶位点(SEQ ID NO:25)的完全重编程的I-OnuI变体并纯化。

[0846] 使用染色体整合的荧光报告子系统测量靶向外显子5中PD-1ITSM编码区的重编程的I-OnuI HE的活性(Certo等人,2011)。将结合并裂解PD-1ITSM靶序列的完全重编程的I-OnuI HE克隆至哺乳动物表达质粒中,接着个别地转染至在编码荧光mCherry蛋白质的框外基因(out-of-frame gene)上游含有PD-1外显子5靶序列的HEK 293T纤维母细胞系中。利用所述HE裂解嵌入的靶位点并在通过非同源末端接合(NHEJ)路径修复DNA之后积累indel使得约三分之一被修复的基因座将荧光报告基因放回“框内”。因此,使用带mCherry荧光的HEK 293T细胞的百分比作为在染色体嵌入的靶序列处核酸内切酶活性的读出。靶向PD-1外显子5位点的完全重编程的I-OnuI HE(PD-1.ITSM.ex5_RD1_CV3-08,SEQ ID NO:6)在细胞染色体环境中显示出极其适中的mCherry表达效率。图3A。当通过平衡底物滴定法测量时,所述HE变体具有适中的DNA亲和力特性(图3B)。

[0847] 接下来,通过对PD-1.ITSM.ex5_RD1_CV3-08HE变体进行随机诱变以产生二级I-OnuI变体文库。在较严格的裂解条件下进行基于展示的流式分选以分离出催化效率改善的变体。这一方法鉴别出I-OnuI变体PD-1.ITSM.ex5_RD2_73(SEQ ID NO:7),相对于RD1变体,其含有四个氨基酸突变,并且具有的mCherry表达细胞的比率比RD1变体高出数倍。图4。再进行三轮活性改进筛选以增加在外显子5靶位点(PD-1.ITSM.ex5_RD3_03,SEQ ID NO:8;PD-1.ITSM.ex5_RD4_CV23,SEQ ID NO:9;和PD-1.ITSM.ex5_RD5_CV23MK,SEQ ID NO:10)处

的基因编辑效率。

[0848] 实例2

[0849] 靶向PD-1外显子5的I-ONUI变体和MEGATAL的表征

[0850] PD-1外显子5靶位点在大范围核酸酶结合位点和邻近TAL阵列结合位点中都包含CpG二核苷酸基元(图1B)。在用CD3和CD28活化并在补充有IL-2的完全培养基中培养的原代人T细胞中通过亚硫酸氢盐测序来评价这些二核苷酸的甲基化状态。3天后,分离出基因组DNA并用亚硫酸氢盐处理以将任何未甲基化的胞嘧啶碱基转化成尿嘧啶。接着,对外显子5靶位点测序以揭示各胞嘧啶的未甲基化(转化成胸腺嘧啶)与甲基化(保持胞嘧啶形式)状态。结果显示,靶位点中的两个CpG基元胞嘧啶主要呈甲基化状态(图5),符合典型‘基因体’甲基化模式。

[0851] 为了确定PD-1.ITSM.ex5_RD5_CV23MK HE变体识别并裂解甲基化外显子5靶位点,使用在靶位点位置p5处、或在p6处鸟嘌呤碱基的反向互补序列上、或在靶序列的两条链上含有5-甲基胞嘧啶的底物进行DNA结合亲和力和裂解分析,所述5-甲基胞嘧啶代表活化T细胞中靶位点的甲基化状态。图6。HE变体以类似方式结合并裂解甲基化DNA底物和未修饰的底物。

[0852] 通过将对应于11个碱基对的TAL阵列靶位点(SEQ ID NO:26)的容许5-甲基胞嘧啶的10.5个单元的TAL阵列附加至大范围核酸酶结构域的N末端,将PD-1.ITSM.ex5_RD5_CV23MK HE变体设计成megaTAL形式(SEQ ID NO:19)(如Boissel等人,2013中所描述)。megaTAL靶位点序列如SEQ ID NO:27中所述。针对TAL阵列靶位点中存在的甲基化CpG二核苷酸测试megaTAL。TAL阵列被设计成通过在相应阵列位置处并入‘N*’RVD来容许甲基化碱基。

[0853] 通过在补充有细胞因子的培养基中用抗CD3和抗CD28抗体预刺激原代人T细胞48-72小时,接着将编码PD-1.ITSM.ex5_RD5_CV23MK megaTAL的体外转录(IVT)、戴帽并聚腺苷酸化的mRNA(SEQ ID NO:40)电穿孔加入细胞中,来评估megaTAL编辑效率。此外,还添加编码3’至5’核酸外切酶Trex2的IVT-mRNA以通过非同源末端接合(NHEJ)路径增进断裂加工(参见Certo等人,2012)。电穿孔后,在补充有细胞因子的培养基中培养细胞7-10天,在此期间取出等分试样进行基因组DNA分离,随后对整个PD-1外显子5靶位点进行PCR扩增。

[0854] 使用分解跟踪Indel(Tracking of Indels by DEcomposition,TIDE;参见Brinkman等人,2014)或高通量测序测量indel的频率。在Trex2存在下PD-1.ITSM.ex5_RD5_CV23MK megaTAL的编辑效率是约60%。图7A。在靶位点处观察到的主要indel类型是+1、-1、-2、-3或-4个核苷酸。图7B。这一分析确定,PD-1.ITSM.ex5_RD5_CV23MK变体破坏相当大一部分megaTAL处理的人T细胞中的PD-1ITSM共同基元。

[0855] 实例3

[0856] 对I-ONUI大范围核酸酶结构域重编程以破坏程序性死亡受体-1(PD-1)基因内的细胞外结构域

[0857] 使用高通量酵母表面展示体外核酸内切酶分析剖析靶向PD-1外显子5靶位点(SEQ ID NO:25)的I-OnuI变体HE的中心-4特异性(Jarjour,West-Foyle等人,2009)。将编码靶向PD-1外显子5的HE变体的质粒转化至酿酒酵母中进行表面展示,接着测试针对PCR产生的包含PD-1外显子5靶位点DNA序列的双链DNA底物的裂解活性,所述PD-1外显子5靶位点DNA序

列含有256个可能的中心-4序列中的每一个。特异性剖析显示,这一重编程的I-OnuI对于中心-4靶位点具有高度选择性(图8),而且还可以裂解另外的非典型中心-4靶位点。由于在PD-1基因中别处存在极少的典型I-OnuI中心-4靶位点,故使用PD-1外显子1和外显子2中的非典型中心-4靶位点对另外的归巢核酸内切酶重编程。

[0858] 在PD-1外显子1和外显子2中的这些非典型中心-4靶位点分别是在编码信号肽和IgV结构域的区域中。不希望受任何特定理论束缚,预期通过基因编辑indel靶向这些区域可使PD-1蛋白质表达消除或不稳定,由此产生净基因敲除表型。此外,获取这些位点用于更先进的基因编辑操作,包括但不限于转基因盒的靶向插入,可以产生处于特有调控性控制下的表型。

[0859] 将I-OnuI重编程成靶向两个含非典型中心-4的靶位点(SEQ ID NO:30和35),这两个外显子/基元中各一个,并且延伸至相应TAL阵列靶位点(SEQ ID NO:26和31)和完整megaTAL靶位点(SEQ ID NO:31和37)。图9A和9B。通过构建在DNA识别界面中含有可变氨基酸残基的模块文库,将I-OnuI重编程成靶向PD-1基因的外显子1或外显子2。为了构建这些变体,使用寡核苷酸将简并密码子并入I-OnuI DNA结合结构域中。使用编码简并密码子的寡核苷酸作为PCR模板,在酵母菌株酿酒酵母中通过空隙重组产生变体文库。每个变体文库跨N末端或C末端I-OnuI DNA识别结构域并且含有约 10^7 至 10^8 个特有转化体。通过流式细胞术,关于针对包含相应结构域的“半位点”的靶位点(外显子1:SEQ ID NO:34和35;外显子2:SEQ ID NO:38和39)的裂解活性筛选所得到的表面展示文库。

[0860] 对展示N末端和C末端结构域重编程的I-OnuI HE的酵母进行纯化并提取出质粒DNA。进行PCR反应以扩增重编程的结构域,随后将其转化至酿酒酵母中以建立重编程结构域组合的文库。自这些文库中鉴别出针对在编码PD-1外显子1中的信号肽的区域中和在编码PD-1外显子2中的IgV结构域的区域中的完整靶位点的完全重编程的I-OnuI变体并纯化。

[0861] 将靶向PD-1外显子1和外显子2靶位点的重编程的I-OnuI HE克隆至哺乳动物表达质粒中,接着个别地转染至在编码荧光iRFP蛋白质的框外基因上游含有相应靶序列的HEK 293T纤维母细胞系中。

[0862] 靶向PD-1外显子1位点(PD-1.ile3.外显子1_RD1_B1,SEQ ID NO:11)的重编程的I-OnuI HE以单独HE变体形式或设计成megaTAL形式(SEQ ID NO:20)之后在细胞染色体环境中显示出适中的iRFP表达效率。通过对初始文库筛选过程中鉴别出的PD-1外显子1RD1变体进行随机诱变,产生二级I-OnuI变体文库。在较严格的裂解条件下进行基于展示的流式分选以分离出催化效率改善的变体。这一过程鉴别出两个I-OnuI变体(PD-1.ile3.外显子1_RD2_B1H8、PD-1.ile3.外显子1_RD2_B1G2,分别是SEQ ID NO:12和60)。呈单独HE变体形式以及设计成megaTAL形式(SEQ ID NO:21和SEQ ID NO:64)之后的变体相对于RD1变体含有四个(PD-1.ile3.外显子1_RD2_B1H8)或五个(PD-1.ile3.外显子1_RD2_B1G2)氨基酸突变,并且具有明显高于PD-1外显子1RD1变体的iRFP表达细胞比率。图10。

[0863] 当以单独HE变体形式或设计成megaTAL形式(SEQ ID NO:22)之后递送时,靶向PD-1外显子2位点(PD-1.IgV.外显子2_RD1_G5;SEQ ID NO:13)的重编程的I-OnuI HE显示出较高的iRFP表达效率。

[0864] 当使用相应靶序列,通过平衡底物滴定测量时,靶向外显子1和外显子2的HE变体展示出较强的DNA亲和力特性。图11A。

[0865] 再对外显子1核酸酶在接触碱基对-8、-7和-6的区域处进行改进以改善其特异性。通过使氨基酸残基68、70、78、80和82随机化来构建微型文库,并在通过流式细胞仪分选6轮之后,三个克隆(PD-1.ile3.外显子1_RD2_B1G2C4、PD-1.ile3.外显子1_RD2_B1G2C11和PD-1.ile3.外显子1_RD2_B1G2C5;分别是SEQ ID NO:61、62和63)显示出高于亲本核酸酶(PD-1.ile3.外显子1_RD2_B1G2;SEQ ID No:60)的特异性。图11B。

[0866] 使用亚硫酸氢盐测序评价PD-1外显子1和外显子2靶位点内存在的CpG基元的甲基化状态。图12。经显示,PD-1外显子1靶位点中的CpG基元在活化T细胞中是未甲基化的,而PD-1外显子2靶中的两个CpG基元都是甲基化的。进行基于展示的活性分析以确定完全CpG甲基化的PD-1外显子2靶序列被相应HE变体高效裂解。

[0867] 此外,还使用了完全CpG甲基化的外显子2靶位点鉴别针对CpG甲基化靶位点的结合和裂解活性改善的I-OnuI变体(PD-1.IgV.外显子2_RD1_PS3,SEQ ID NO:14)。

[0868] 实例4

[0869] 原代人T细胞中PD-1基因的靶向破坏将挽救PD-L1介导的T细胞功能抑制作用

[0870] 在用CD3和CD28活化并在补充有IL-2的完全培养基中培养的原代人T细胞中评价被重编程用于裂解PD-1基因中的各种靶序列的megaTAL的功能影响。用编码抗BCMA CAR的慢病毒载体转导活化的PBMC。将体外转录的编码靶向PD-1外显子5或PD-1外显子1的megaTAL的mRNA(分别是SEQ ID NO:40和41)和编码Trex2的mRNA(SEQ ID NO:43)电穿孔加入抗BCMA CAR T细胞中。对照物包括未处理的T细胞,或用编码青色荧光蛋白(CFP)或靶向CCR5的megaTAL的mRNA处理的T细胞(参见Sather等人,《科学·转化医学(Sci Transl Med.)》2015年9月30日;7(307):307ra156)。在10天扩增之后,用多克隆活化试剂佛波醇12-肉豆蔻酸酯13-乙酸酯(PMA)/离子霉素(P/I)刺激T细胞。在T细胞活化后,PD-1在细胞表面上天然地上调。在PD-1外显子1megaTAL mRNA转染后,PD-1上调受到抑制,指示这一区域中的indel将破坏PD-1蛋白质的正常制造。相比之下,用编码对照CCR5megaTAL或靶向PD-1外显子5的megaTAL的mRNA处理对PD-1表面表达无影响,不过这一megaTAL在ITSM中诱导较高比率的indel。图13A。在类似条件下用PD-1外显子1megaTAL mRNA的特异性改进形式(SEQ ID NO:65、66、67和68)再重复进行实验也在T细胞中显示出类似的PD-1表达破坏情况。图13B

[0871] 同时递送PD-1外显子1和外显子5megaTAL将明显改善PD-1细胞表面表达的破坏,不依赖于Trex2核酸外切酶的递送。图14。由此指示,同时近端DNA断裂形成是在不依赖于核酸外切酶驱动的indel增强情况下,以高效率促进大型基因缺失事件的一种机制。

[0872] 通过CAR-T细胞响应于工程改造成表达PD-1配体PD-L1的肿瘤细胞的细胞因子产生情况,分析通过靶向T细胞的表达或其信号传导功能破坏T细胞中PD-1信号传导的影响。如由上清液中测量的TNF α 和IL-17A高含量所例证,将抗BCMA CAR T细胞与BCMA表达肿瘤细胞系共培养会引起T细胞活化且随后引起炎性细胞因子分泌。BCMA表达肿瘤细胞上PD-L1的共表达抑制炎性细胞因子产生。然而,用编码PD-1外显子1或外显子5megaTAL的mRNA转染抗BCMA CAR T细胞使PD-L1介导的细胞因子抑制作用减弱,因为这些样品中炎性细胞因子产生被恢复成基线水平。图15A。

[0873] 实例5

[0874] 转基因同源重组至PD-1基因的外显子1中

[0875] 设计并构建重组腺相关病毒(rAAV)质粒,该质粒在基因靶向同源区之间含有启动子-转基因盒,其包含异源启动子、编码荧光蛋白质的转基因和聚腺苷酸化信号。用XmaI消化验证AAV ITR元件的完整性。所述转基因盒被放在两个同源区之间,长度是约600-700bp,侧接PD-1外显子1megaTAL裂解位点(SEQ ID NO:27)。5'同源臂(SEQ ID NO:54)在megaTAL裂解位点上游含有第一个PD-1外显子的一部分和其它序列。3'同源臂(SEQ ID NO:55)在megaTAL裂解位点的下游含有所述PD-1外显子的一部分和其它序列。同源区都不含完整megaTAL靶位点。这一例示性表达盒含有骨髓增生性肉瘤病毒增强子、负控制区缺失、d1587rev引物结合位点取代(MND)的启动子,所述启动子可操作地连接至编码荧光多肽,例如蓝色荧光蛋白(BFP)、红色荧光蛋白(RFP)、青色荧光蛋白(CFP)、绿色荧光蛋白(GFP)等的聚核苷酸。此外,所述表达盒含有位于转基因终止密码子下游的SV40晚期聚腺苷酸化信号。图16A。通过用质粒短暂共转染HEK 293T细胞来制备重组AAV(rAAV),所述质粒提供病毒制造所需的复制、衣壳和腺病毒辅助元件。使用超速离心法,以基于碘克沙醇(iodixanol)的梯度自共转染的HEK 293T细胞培养物中纯化出rAAV。

[0876] 在用CD3和CD28活化并在补充有IL-2的完全培养基中培养的原代人T细胞中评价MegaTAL诱导的同源重组。培养3天之后,洗涤T细胞并通过电穿孔加入体外转录的PD-1外显子1megaTAL mRNA(SEQ ID NO:41),随后用纯化的编码MND-GFP转基因盒的重组AAV(SEQ ID NO:53)转导。对照物包括用megaTAL mRNA或rAAV靶向载体处理的T细胞。在多个时间点,使用流式细胞术测量表达荧光蛋白质的T细胞的频率并将游离型rAAV靶向载体中荧光蛋白质的短暂表达与染色体整合的盒的长期表达相区别。图16B。通过测序并通过在多克隆T细胞活化后PD-1表达的损失来检测MegaTAL介导的PD-1基因破坏。

[0877] 在20-60%用megaTAL和rAAV靶向载体处理的T细胞中观察到长期转基因表达。在对照样品中,单独rAAV处理在处理过的T细胞中产生变化水平的短暂荧光蛋白质表达和极低水平(<1%)的长期荧光蛋白质表达,与没有整合至基因组中的情形相符。结果在针对从若干独立供体分离的T细胞进行的实验中得到证实。

[0878] 实例6

[0879] 编码嵌合抗原受体(CAR)的转基因同源重组至PD-1基因中

[0880] 如上文所描述,设计、构建和验证重组腺相关病毒(rAAV)质粒,不过编码抗CD19CAR(SEQ ID NO:56)或抗BCMA CAR(SEQ ID NO:57)的转基因盒被放在PD-1外显子1靶向同源区之间。CAR表达盒含有MND启动子,其可操作连接至编码CAR以及被设计成靶向PDCD1外显子1靶位点的同源臂(SEQ ID NO:54和55)的聚核苷酸,所述CAR包含CD8 α 源性信号肽、靶向CD19抗原或B细胞成熟抗原(BCMA)的单链可变片段(scFv)、CD8 α 源性铰链区和跨膜结构域、细胞内4-1BB共刺激结构域、CD3 ζ 信号传导结构域。

[0881] 如上文所描述,用CD3和CD28活化原代人T细胞并使其在补充有细胞因子的培养基中生长。使用电穿孔加入体外转录的PD-1外显子1megaTAL mRNA(SEQ ID NO:41)且随后用编码抗CD19或抗BCMA CAR的rAAV转导的活化的原代人T细胞评价CAR转基因同源重组至PD-1外显子1靶位点中的情况。在10天时间点,即电穿孔和转导之后7天时,进行流式细胞术以检测CAR表达。对照样品包括用megaTAL mRNA或rAAV靶向载体处理的T细胞。使用重组PE偶联的CD19-Fc或PE偶联的BCMA-Fc染色试剂分析CD19-CAR和BCMA-CAR表达。用megaTAL mRNA和rAAV-CAR处理的T细胞显示在30-60%的T细胞中有CD19-CAR表达并且在10-20%的T细胞

中有BCMA-CAR表达。图16B。在未处理、rAAV处理和megaTAL/rAAV CAR处理的T细胞之间观察到类似的T细胞扩增速率和无法区分的T细胞表型。

[0882] 实例7

[0883] 启动子缺少型转基因同源重组至PD-1基因中

[0884] 设计、构建并验证含有荧光报告子(mCherry)转基因和聚腺苷酸化信号但缺乏外源启动子的重组腺相关病毒(rAAV)质粒(SEQ ID NO:58)。图17。mCherry起始密码子与内源性PD-1起始密码子合并且重叠,同时PD-1外显子1的其余部分被编码mCherry蛋白质的cDNA置换。这一策略驱动由内源性PD-1启动子表达荧光蛋白质,同时还破坏PD-1蛋白质的正常表达。

[0885] 如上文所描述,活化原代人T细胞,电穿孔加入体外转录的megaTAL mRNA并用rAAV转导。对样品包括用megaTAL mRNA或rAAV靶向载体处理的T细胞。在转染后各种时间点,在存在或不存在多克隆T细胞活化情况下,使用佛波醇12-肉豆蔻酸酯13-乙酸酯(PMA)/离子霉素(P/I)通过流式细胞术分析荧光报告子表达。在用单独megaTAL mRNA或rAAV靶向载体处理的T细胞中未观察到报告子表达。在存在或不存在rAAV转导情况下观察到类似的megaTAL活性比率。在接受megaTAL和rAAV靶向载体并用P/I活化48小时的T细胞中观察到较低水平的荧光报告子表达。由内源性PD-1启动子驱动的荧光报告子表达低于异源启动子驱动的受体表达(荧光强度降低约5倍)。

[0886] 实例8

[0887] 编码荧光蛋白质的转基因同源重组至PD-1基因外显子5中之后的组成性PD-1表达

[0888] 设计、构建并验证含有异源启动子、蓝色荧光蛋白质(BFP)转基因和聚腺苷酸化信号的重组腺相关病毒(rAAV)质粒(SEQ ID NO:44)以引入PD-1外显子5中。转基因盒被放在1.3kb与1.0kb同源臂之间,侧接PD-1外显子5megaTAL裂解位点。5'同源臂(SEQ ID NO:45)包括在megaTAL裂解位点上游的PD-1外显子5部分和其它上游序列。3'同源臂(SEQ ID NO:46)含有PD-1外显子5的末端编码序列和其它下游序列。同源区都不含完整megaTAL靶位点。图18。

[0889] 用CD3和CD28活化原代人T细胞,电穿孔加入体外转录的PD-1外显子5megaTAL mRNA(SEQ ID NO:40),接着用编码荧光报告子转基因的rAAV转导。对照物包括未处理的T细胞和用单独rAAV靶向载体处理的T细胞。在转染后各种时间点,通过流式细胞术分析荧光报告子表达。用单独rAAV处理的细胞展现出短暂BFP表达,所述表达在细胞扩增期间降低至背景水平。接受PD-1外显子5megaTAL mRNA和rAAV载体两种的T细胞展现稳定的BFP表达T细胞群,该细胞群在整个离体培养过程中持续存在。

[0890] 用PD-1外显子5megaTAL mRNA和rAAV载体两种处理的T细胞具有不同的PD-1表达模式,由此在不存在活化情况下,几乎所有的BFP表达细胞还表达PD-1蛋白质。由此指示,靶向策略将PD-1基因的正常调控变为组成性表达模式,而非诱导型表达模式。此外,由于靶向PD-1外显子5,这一靶向策略中组成性表达的蛋白质是PD-1变体,该变体具有减弱的抑制性信号传导并且可以用作显性负性受体。在用单独PD-1外显子5megaTAL mRNA处理的T细胞中或在用megaTAL mRNA和包含针对外显子1的同源臂的rAAV处理的T细胞中,PD-1表达未上调。图19。

[0891] 实例9

[0892] PD-1基因中转基因的同源重组不依赖于单核苷酸多态现象 (SNP)

[0893] PD-1基因座具有遗传异质性,并在内含子和外显子区域中都存在较高的SNP发生率。megaTAL裂解位点附近发散SNP的存在可能潜在地影响PD-1基因座处同源重组的效率。设计包括异源启动子、荧光蛋白质转基因和晚期SV40聚腺苷酸化信号的重组腺相关病毒 (rAAV) 质粒以评估SNP的影响,接着如上文所描述进行构建和验证 (SEQ ID NO:47和48)。同源臂与共同PD-1序列一致或这些同源臂被设计成包括由PD-1外显子5megaTAL靶向位点附近常见的SNP引起的点突变。在这些设计中,5'同源臂 (L-SNP,SEQ ID NO:49) 含有位于megaTAL裂解位点上游220bp处的内含子G/A SNP (rs6705653),而3'同源臂 (R-SNP,SEQ ID NO:50) 含有位于PD-1megaTAL裂解位点下游68bp处的沉默C/T SNP (rs2227981)。图20。

[0894] 用CD3和CD28活化原代人T细胞,电穿孔加入体外转录的megaTAL mRNA,并用编码共同同源臂的rAAV,或用载有L-SNP或R-SNP同源臂的rAAV转导。对照物包括用单独rAAV靶向载体转导的T细胞。在转导后不同时间点,通过流式细胞术分析荧光蛋白质表达。在仅接受rAAV的样品中荧光蛋白质的表达下降至背景水平,而用megaTAL与rAAV的组合处理的T细胞展示稳定水平的荧光蛋白质表达。尽管用含非SNP的rAAV载体或含有L或R SNP变体的rAAV载体处理,但所有样品都展现类似的荧光水平,指示megaTAL靶位点附近的SNP不影响或基本上不影响靶位点处的HR。

[0895] 实例10

[0896] 编码转换受体的转基因同源重组至PDCD1基因座中

[0897] 转换受体是能够将细胞外抑制信号转化成细胞内活化信号的工程改造的嵌合分子,不过其有效性通常与其压倒天然受体的固有表达的能力相关。避开这一局限的一种方式是将转换受体嵌入原生基因座中并破坏天然受体。设计出含有异源启动子、编码PD-1细胞外配体结合结构域和细胞内CD28信号传导结构域的转基因 (PD-1转换受体,SEQ ID NO:59) 和晚期SV40聚腺苷酸化信号的重组腺相关病毒 (rAAV) 质粒以靶向PD-1外显子5,构建并验证。5'同源臂包括在含有ITIM的megaTAL裂解位点上游的PD-1外显子5部分和其它上游序列。3'同源臂含有PD-1外显子5的末端编码序列和一部分UTR区域,并缩短至约650bp以接纳转换受体cDNA。同源区都不含完整megaTAL靶位点。图21。

[0898] 用CD3和CD28活化原代人T细胞,电穿孔加入体外转录的PD-1外显子5megaTAL mRNA并用编码PD-1转换受体的rAAV靶向载体转导。对照物包括仅接受megaTAL或rAAV的样品。通过无T细胞刺激存在下,用抗PD-1抗体染色,分析PD-1转换受体的表达。在用PD-1外显子5megaTAL mRNA处理并用rAAV转导的T细胞中观察到高表达水平。

[0899] 实例11

[0900] 使用HDR长程切除PD-1细胞内信号传导区域

[0901] 已知同源重组机制涉及可横跨远端基因组序列的同源‘检索’机制。不希望受任何特定理论束缚,预期较大的靶基因或染色体区域范围可能倾向于使用megaTAL和rAAV递送的组合高效操作。设计、构建并验证侧接同源臂的重组腺相关病毒 (rAAV) 质粒以评估远端HR事件 (SEQ ID NO:52),所述质粒包括异源启动子、BFP转基因和晚期SV40聚腺苷酸化信号。所述转基因侧接1.3kb的5'同源臂,所述同源臂被设计成在PD-1外显子3的上游开始并在即将开始外显子4之前结束。3'同源臂含有PD-1外显子5的末端编码序列和一部分UTR区域。同源区都不含完整megaTAL靶位点。成功的同源重组被设计成用于除去整个PD-1细胞内

信号传导区域,仅有PD-1的细胞外部分和跨膜部分留在处理过的细胞上。图22。

[0902] 用CD3和CD28活化原代人T细胞,电穿孔加入体外转录的megaTAL mRNA并用包含远端5'同源臂的BFP编码rAAV靶向载体转导。对照物包括用megaTAL或rAAV靶向载体处理的T细胞。仅在接受rAAV和靶向PDCD1外显子5的megaTAL mRNA的样品中观察到稳定BFP表达。

[0903] 实例12

[0904] 在体外原代人T细胞中的PD-1基因破坏和抗体介导的PD1阻断

[0905] 在用CD3和CD28活化并在补充有IL-2的完全培养基中培养的原代人T细胞中评价被重编程用于裂解PD-1基因中的各种靶序列的megaTAL的功能影响。用编码抗BCMA CAR的慢病毒载体转导活化的PBMC。将体外转录的编码靶向PD-1外显子5或PD-1外显子1的megaTAL的mRNA(分别是SEQ ID NO:40和41)和编码Trex2的mRNA(SEQ ID NO:43)电穿孔加入抗BCMA CAR T细胞中。对照物包括未电穿孔加入mRNA的T细胞(模拟EP)或电穿孔加入编码megaTAL的mRNA的T细胞,所述megaTAL对无催化活性的TCR α 具有特异性。

[0906] 在扩增10天后,将T细胞与用编码BCMA-GFP的慢病毒载体转导的A549细胞共培养。将CAR T细胞与单独A549靶细胞或在20 μ g/mL抗PD-1抗体存在下共培养。将细胞共培养72小时并使用基于珠粒的分析法(Intellicyt QBeads)分析上清液中的细胞因子含量。在无抗PD-1抗体存在下,用PD1-Ex1或PD1-Ex5megaTAL处理过的细胞显示出IFN γ 和TNF α 产生相较于模拟EP和TCR α 无效对照有所升高。添加 α PD-1抗体消除此差异,使自对照细胞和PD1megaTAL处理的细胞获得的上清液中有相等水平的IFN γ 和TNF α 分泌(图15B)。

[0907] 实例13

[0908] 启动子缺少型转基因同源重组至PD-1基因中

[0909] 设计、构建并验证含有荧光报告子(mCherry)转基因和聚腺苷酸化信号但缺乏外源启动子的重组腺相关病毒(rAAV)质粒(SEQ ID NO:58)。图17.mCherry起始密码子与内源性PD-1起始密码子合并且重叠,同时PD-1外显子1的其余部分被编码mCherry蛋白质的cDNA置换。这一策略驱动由内源性PD-1启动子表达荧光蛋白质,同时还破坏PD-1蛋白质的正常表达。

[0910] 如上文所描述,活化原代人T细胞,电穿孔加入体外转录的megaTAL mRNA并用rAAV转导。对照样品包括用megaTAL mRNA或rAAV靶向载体处理的T细胞。在刺激后24小时,在存在或不存在多克隆T细胞活化情况下,使用佛波醇12-肉豆蔻酸酯13-乙酸酯(PMA)/离子霉素(P/I)通过流式细胞术分析荧光报告子表达。在未处理的T细胞中未观察到报告子表达。对于HDR,在P/I刺激之前,在接受megaTAL和rAAV靶向载体两种的T细胞中观察到较低基础水平的荧光报告子表达。二十四小时的P/I刺激使内源性PD-1驱动的荧光报告子表达从3%上调至27.3%(比较图17中的右上图与右下图)。

[0911] 实例14

[0912] 启动子缺少型细胞因子转基因同源重组至PD-1基因座中

[0913] 设计、构建并验证含有IL-12或IL-15转基因和聚腺苷酸化信号但缺乏外源启动子的重组腺相关病毒(rAAV)质粒(SEQ ID NO:58)。转基因起始密码子与内源性PD-1起始密码子合并且重叠,同时PD-1外显子1的其余部分被编码IL-12或IL-15细胞因子的cDNA置换(图23,上图)。这一策略驱动由内源性PD1启动子表达转基因,同时还破坏PD1的正常表达。

[0914] 活化原代人T细胞并在活化后24小时,用编码抗BCMA CAR的慢病毒载体转导。如上

文所描述,使细胞繁殖2天,电穿孔加入体外转录的megaTAL mRNA并用rAAV转导。对照样品包括未被转导的T细胞、仅用抗BCMA CAR处理的T细胞或用抗BCMA CAR处理并在重组IL-12存在下培养的T细胞。通过流式细胞术分析用PMA/离子霉素刺激24小时的T细胞上PD-1的表达来了解PD-1megaTAL的活性。PD-1megaTAL与IL-12AAV的组合显示PD-1表达相较于对照样品减少约40% (图23, 下图)。

[0915] 在PMA/离子霉素活化之后或在T细胞与表达抗原的K562-BCMA细胞系共培养后,通过ELISA测量IL-12表达水平。在未刺激的T细胞中IL-12产生最少并且PMA/离子霉素刺激和与抗原阳性K562细胞共培养引起IL-12分泌 (图24A)。使用重复刺激分析测量BCMA-CAR T细胞的功能耗竭。将肿瘤细胞与抗BCMA-CAR T细胞混合,培养7天并且再添加K562-BCMA靶细胞以模拟重复刺激。第1次刺激之后,所有T细胞样品中IFN γ 产生和细胞毒性水平类似。第2次刺激之后,用重组IL-12处理或用PD-1megaTAL和IL-12HDR模板处理的抗BCMA-CAR T细胞中IFN γ 产生和细胞毒性相较于对照处理的细胞有所增加 (图24B)。

[0916] 总体而言,在所附权利要求书中,所使用的术语不应解释为将权利要求项局限于本说明书和权利要求书中所公开的具体实施例,而是应解释为包括所有可能的实施例以及与这些权利要求项所授权的内容等效的全部范围。因此,权利要求书不受本公开限制。

序列表

<110> 蓝鸟生物公司(bluebird bio, Inc.)

贾斯迪普·曼恩(Mann, Jasdeep);

乔尔·盖伊(Gay, Joel);

乔丹·贾儒尔(Jarjour, Jordan);

张珏(Zhang, Joy)

<120> PD-1归巢核酸内切酶变体、组合物及使用方法

<130> BLBD-076/02W0

<150> US 62/414,279

<151> 2016-10-28

<150> US 62/385,079

<151> 2016-09-08

<160> 123

<170> PatentIn 3.5版

<210> 1

<211> 303

<212> PRT

<213> 新榆枯萎病菌美洲亚种(*Ophiostoma novo-ulmi* subsp. *americana*)线粒体(mitochondrion)

<400> 1

Met	Ala	Tyr	Met	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Leu	Leu	Arg	Ile	Arg	Asn	Asn
				20					25					30	
Asn	Lys	Ser	Ser	Val	Gly	Tyr	Ser	Thr	Glu	Leu	Gly	Phe	Gln	Ile	Thr
				35				40					45		
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
				50				55				60			
Lys	Val	Gly	Val	Ile	Ala	Asn	Ser	Gly	Asp	Asn	Ala	Val	Ser	Leu	Lys
65					70					75				80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
				85					90					95	
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Met	Leu	Phe	Lys	Gln
				100				105					110		
Ala	Phe	Cys	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Ile	Asn	Gly	Ile
				115				120					125		
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Leu	Asn	Trp	Gly	Leu	Thr	Asp
				130				135					140		

Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Ile	Ile	Ser	Lys	Glu	Arg	Ser	Leu
145					150					155					160
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165						170					175	
Gly	Glu	Gly	Cys	Phe	Phe	Val	Asn	Leu	Ile	Lys	Ser	Lys	Ser	Lys	Leu
			180					185					190		
Gly	Val	Gln	Val	Gln	Leu	Val	Phe	Ser	Ile	Thr	Gln	His	Ile	Lys	Asp
		195					200					205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Tyr	Ile
	210					215					220				
Lys	Glu	Lys	Asn	Lys	Ser	Glu	Phe	Ser	Trp	Leu	Asp	Phe	Val	Val	Thr
225					230					235					240
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
			245						250					255	
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
			260					265					270		
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
		275					280					285			
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Val	Phe	
	290					295					300				

<210> 2

<211> 303

<212> PRT

<213> 新榆枯萎病菌美洲亚种(*Ophiostoma novo-ulmi* subsp. *americana*) (线粒体(mitochondrion))

<400> 2

Met	Ala	Tyr	Met	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1			5						10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Leu	Leu	Arg	Ile	Arg	Asn	Asn
			20					25					30		
Asn	Lys	Ser	Ser	Val	Gly	Tyr	Ser	Thr	Glu	Leu	Gly	Phe	Gln	Ile	Thr
		35					40					45			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
	50					55					60				
Lys	Val	Gly	Val	Ile	Ala	Asn	Ser	Gly	Asp	Asn	Ala	Val	Ser	Leu	Lys
65				70					75					80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85						90				95		
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln

100	105	110
Ala Phe Ser Val Met Glu Asn Lys	Glu His Leu Lys Glu Asn Gly Ile	
115	120	125
Lys Glu Leu Val Arg Ile Lys Ala Lys	Leu Asn Trp Gly Leu Thr Asp	
130	135	140
Glu Leu Lys Lys Ala Phe Pro Glu Asn	Ile Ser Lys Glu Arg Ser Leu	
145	150	155
Ile Asn Lys Asn Ile Pro Asn Phe Lys	Trp Leu Ala Gly Phe Thr Ser	
165	170	175
Gly Glu Gly Cys Phe Phe Val Asn Leu	Ile Lys Ser Lys Ser Lys Leu	
180	185	190
Gly Val Gln Val Gln Leu Val Phe Ser	Ile Thr Gln His Ile Lys Asp	
195	200	205
Lys Asn Leu Met Asn Ser Leu Ile Thr	Tyr Leu Gly Cys Gly Tyr Ile	
210	215	220
Lys Glu Lys Asn Lys Ser Glu Phe Ser	Trp Leu Asp Phe Val Val Thr	
225	230	235
Lys Phe Ser Asp Ile Asn Asp Lys Ile	Ile Pro Val Phe Gln Glu Asn	
245	250	255
Thr Leu Ile Gly Val Lys Leu Glu Asp	Phe Glu Asp Trp Cys Lys Val	
260	265	270
Ala Lys Leu Ile Glu Glu Lys Lys His	Leu Thr Glu Ser Gly Leu Asp	
275	280	285
Glu Ile Lys Lys Ile Lys Leu Asn Met	Asn Lys Gly Arg Val Phe	
290	295	300

<210> 3

<211> 303

<212> PRT

<213> 新榆枯萎病菌美洲亚种(*Ophiostoma novo-ulmi* subsp. *americana*) (线粒体
(mitochondrion))

<220>

<221> MOD_RES

<222> (1) .. (3)

<223> 任何氨基酸或不存在

<400> 3

Xaa Xaa Xaa Met Ser Arg Arg Glu Ser	Ile Asn Pro Trp Ile Leu Thr
1	5 10 15
Gly Phe Ala Asp Ala Glu Gly Ser Phe	Leu Leu Arg Ile Arg Asn Asn
20	25 30

Asn	Lys	Ser	Ser	Val	Gly	Tyr	Ser	Thr	Glu	Leu	Gly	Phe	Gln	Ile	Thr
35				40				45							
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
50				55				60							
Lys	Val	Gly	Val	Ile	Ala	Asn	Ser	Gly	Asp	Asn	Ala	Val	Ser	Leu	Lys
65				70				75				80			
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
				85				90				95			
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
				100				105				110			
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
115				120				125							
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Leu	Asn	Trp	Gly	Leu	Thr	Asp
130				135				140							
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Ser	Leu
145				150				155				160			
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
				165				170				175			
Gly	Glu	Gly	Cys	Phe	Phe	Val	Asn	Leu	Ile	Lys	Ser	Lys	Ser	Lys	Leu
180				185				190							
Gly	Val	Gln	Val	Gln	Leu	Val	Phe	Ser	Ile	Thr	Gln	His	Ile	Lys	Asp
195				200				205							
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Tyr	Ile
210				215				220							
Lys	Glu	Lys	Asn	Lys	Ser	Glu	Phe	Ser	Trp	Leu	Asp	Phe	Val	Val	Thr
225				230				235				240			
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
				245				250				255			
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
260				265				270							
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
275				280				285							
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Val	Phe	
290				295				300							

<210> 4

<211> 303

<212> PRT

<213> 新榆枯萎病菌美洲亚种(*Ophiostoma novo-ulmi* subsp. *americana*) (线粒体(mitochondrion))

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 4

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Leu	Leu	Arg	Ile	Arg	Asn	Asn
			20					25					30		
Asn	Lys	Ser	Ser	Val	Gly	Tyr	Ser	Thr	Glu	Leu	Gly	Phe	Gln	Ile	Thr
		35					40					45			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
	50					55				60					
Lys	Val	Gly	Val	Ile	Ala	Asn	Ser	Gly	Asp	Asn	Ala	Val	Ser	Leu	Lys
65				70					75					80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85					90					95		
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
		100						105					110		
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
		115					120					125			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Leu	Asn	Trp	Gly	Leu	Thr	Asp
		130					135				140				
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Ser	Leu
145				150					155					160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165					170					175		
Gly	Glu	Gly	Cys	Phe	Phe	Val	Asn	Leu	Ile	Lys	Ser	Lys	Ser	Lys	Leu
		180						185					190		
Gly	Val	Gln	Val	Gln	Leu	Val	Phe	Ser	Ile	Thr	Gln	His	Ile	Lys	Asp
		195					200					205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Tyr	Ile
	210					215				220					
Lys	Glu	Lys	Asn	Lys	Ser	Glu	Phe	Ser	Trp	Leu	Asp	Phe	Val	Val	Thr
225				230					235					240	

Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
 245 250 255
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
 260 265 270
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
 275 280 285
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa
 290 295 300
 <210> 5
 <211> 303
 <212> PRT
 <213> 新榆枯萎病菌美洲亚种(*Ophiostoma novo-ulmi* subsp. *americana*) (线粒体
 (mitochondrion))
 <220>
 <221> MOD_RES
 <222> (1) .. (8)
 <223> 任何氨基酸或不存在
 <220>
 <221> MOD_RES
 <222> (302) .. (303)
 <223> 任何氨基酸或不存在
 <400> 5
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser Ile Asn Pro Trp Ile Leu Thr
 1 5 10 15
 Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn
 20 25 30
 Asn Lys Ser Ser Val Gly Tyr Ser Thr Glu Leu Gly Phe Gln Ile Thr
 35 40 45
 Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
 50 55 60
 Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys
 65 70 75 80
 Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
 85 90 95
 Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln
 100 105 110
 Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
 115 120 125
 Lys Glu Leu Val Arg Ile Lys Ala Lys Leu Asn Trp Gly Leu Thr Asp

130	135	140
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Ser Leu		
145	150	155
Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser		
165	170	175
Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu		
180	185	190
Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Lys Asp		
195	200	205
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile		
210	215	220
Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr		
225	230	235
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn		
245	250	255
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val		
260	265	270
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp		
275	280	285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa		
290	295	300
<210> 6		
<211> 303		
<212> PRT		
<213> 人工序列(Artificial Sequence)		
<220>		
<221>		
<223> 合成I-OnuI变体PD-1.ITSM.ex5_RD1_CV3-08		
<220>		
<221> MOD_RES		
<222> (1) .. (4)		
<223> 任何氨基酸或不存在		
<220>		
<221> MOD_RES		
<222> (302) .. (303)		
<223> 任何氨基酸或不存在		
<400> 6		
Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr		
1	5	10
		15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Gln Leu Glu Ile Arg Asn Val		
20	25	30
Asn Pro Asn Ile Pro Arg Tyr Lys Thr Arg Leu Arg Phe Glu Ile Asp		
35	40	45
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp		
50	55	60
Lys Val Gly Lys Ile Tyr Asn Gln Gly Asp Ser Tyr Val Lys Leu Arg		
65	70	75
Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys		
85	90	95
Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln		
100	105	110
Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile		
115	120	125
Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp		
130	135	140
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu		
145	150	155
Ile Asn Lys Asn Ile Pro Asn Leu Lys Trp Leu Ala Gly Phe Thr Ser		
165	170	175
Gly Glu Gly His Phe Gly Val Ile Leu Ala Lys Arg Arg Pro Ala Ser		
180	185	190
Pro Val Gln Val Arg Leu Arg Phe Ala Ile Gly Gln His Ile Arg Asp		
195	200	205
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Arg Ile		
210	215	220
Arg Glu Lys Asn Ile Ser Glu Lys Ser Trp Leu Glu Phe Glu Val Thr		
225	230	235
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn		
245	250	255
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val		
260	265	270
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp		
275	280	285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa		
290	295	300

<210> 7

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.ITSM.ex5_RD2_73

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 7

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gln	Leu	Glu	Ile	Arg	Asn	Val
			20					25					30		
Asn	Pro	Asn	Ile	Pro	Arg	Tyr	Arg	Thr	Arg	Leu	Arg	Phe	Glu	Ile	Asp
		35					40					45			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asp	Ile	Gln	Ser	Thr	Trp
	50					55					60				
Lys	Val	Gly	Lys	Ile	Tyr	Asn	Arg	Gly	Asp	Ser	Tyr	Val	Lys	Leu	Arg
65				70					75					80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85						90				95		
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
		100						105					110		
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
		115					120					125			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
	130						135				140				
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145				150					155					160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165						170				175		
Gly	Glu	Gly	His	Phe	Gly	Val	Ile	Leu	Ala	Lys	Arg	Arg	Pro	Ala	Ser
		180					185						190		
Pro	Val	Gln	Val	Arg	Leu	Arg	Phe	Ala	Ile	Gly	Gln	His	Ile	Arg	Asp
	195						200						205		

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Arg Thr
 210 215 220
 Arg Glu Lys Asn Ile Ser Glu Lys Ser Trp Leu Glu Phe Glu Val Thr
 225 230 235 240
 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
 245 250 255
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
 260 265 270
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
 275 280 285
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa
 290 295 300

<210> 8

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.ITSM.ex5_RD3_03

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 8

Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
 1 5 10 15
 Gly Phe Ala Asp Ala Glu Gly Ser Phe Gln Leu Glu Ile Arg Asn Val
 20 25 30
 Asn Pro Asn Ile Pro Arg Tyr Arg Thr Arg Leu Arg Phe Glu Ile Asp
 35 40 45
 Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asp Ile Gln Ser Thr Trp
 50 55 60
 Lys Val Gly Lys Ile Tyr Asn Gln Gly Asp Ser Tyr Val Lys Leu Arg
 65 70 75 80
 Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys

				85				90				95			
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
				100				105				110			
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
				115				120				125			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
				130				135				140			
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145				150				155				160			
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
				165				170				175			
Gly	Asp	Gly	His	Phe	Gly	Val	Ile	Leu	Ala	Lys	Arg	Arg	Pro	Ala	Ser
				180				185				190			
Pro	Val	Gln	Val	Arg	Leu	Arg	Phe	Ala	Ile	Gly	Gln	His	Ile	Arg	Asp
				195				200				205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Arg	Thr
				210				215				220			
Arg	Glu	Lys	Asn	Ile	Ser	Glu	Lys	Ser	Trp	Leu	Glu	Phe	Glu	Val	Thr
225				230				235				240			
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
				245				250				255			
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
				260				265				270			
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
				275				280				285			
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Xaa	Xaa	
290				295				300							

<210> 9

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.ITSM.ex5_RD4_CV23

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 9

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5				10					15		
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gln	Leu	Glu	Ile	Arg	Asn	Val
			20				25					30			
Asn	Pro	Asn	Ile	Pro	Arg	Tyr	Arg	Thr	Arg	Leu	Arg	Phe	Glu	Ile	Asp
		35				40					45				
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asp	Ile	Gln	Ser	Thr	Trp
	50					55				60					
Lys	Val	Gly	Lys	Ile	Tyr	Asn	Gln	Gly	Asp	Ser	Tyr	Val	Lys	Leu	Arg
65				70					75					80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85					90					95		
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
		100						105				110			
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
		115					120					125			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
	130					135				140					
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145				150					155					160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165					170					175		
Gly	Asp	Gly	His	Phe	Gly	Val	Ile	Leu	Ala	Lys	Arg	Arg	Pro	Ala	Ser
		180					185					190			
Pro	Val	Gln	Val	Arg	Leu	Arg	Phe	Ala	Ile	Gly	Gln	His	Ile	Arg	Asp
		195				200						205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Arg	Ile
	210					215				220					
Arg	Glu	Lys	Asn	Lys	Ser	Glu	Lys	Ser	Trp	Leu	Glu	Phe	Glu	Val	Thr
225				230					235					240	
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
			245					250					255		
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
		260					265					270			
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp

275	280	285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa		
290	295	300
<210> 10		
<211> 303		
<212> PRT		
<213> 人工序列(Artificial Sequence)		
<220>		
<221>		
<223> 合成I-0nuI变体PD-1.ITSM.ex5_RD5_CV23MK		
<220>		
<221> MOD_RES		
<222> (1) .. (4)		
<223> 任何氨基酸或不存在		
<220>		
<221> MOD_RES		
<222> (302) .. (303)		
<223> 任何氨基酸或不存在		
<400> 10		
Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr		
1 5 10 15		
Gly Phe Ala Asp Ala Glu Gly Ser Phe Gln Leu Glu Ile Arg Asn Val		
20 25 30		
Asn Pro Asn Ile Pro Arg Tyr Arg Thr Arg Leu Arg Phe Glu Ile Asp		
35 40 45		
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asp Ile Gln Ser Thr Trp		
50 55 60		
Lys Val Gly Lys Ile Tyr Asn Gln Gly Asp Ser Tyr Val Lys Leu Arg		
65 70 75 80		
Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys		
85 90 95		
Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln		
100 105 110		
Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile		
115 120 125		
Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp		
130 135 140		
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu		
145 150 155 160		

```

Ile Asn Lys Asn Ile Pro Asn Leu Lys Trp Leu Ala Gly Phe Thr Ser
      165                      170                      175
Gly Asp Gly His Phe Gly Val Ile Leu Ala Lys Arg Arg Pro Ala Ser
      180                      185                      190
Pro Val Gln Val Arg Leu Arg Phe Met Ile Gly Gln His Ile Lys Asp
      195                      200                      205
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Arg Ile
      210                      215                      220
Arg Glu Lys Asn Lys Ser Glu Lys Ser Trp Leu Glu Phe Glu Val Thr
      225                      230                      235                      240
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
      245                      250                      255
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
      260                      265                      270
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
      275                      280                      285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa
      290                      295                      300

```

<210> 11

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.ile3.外显子1_RD1_B1

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 11

```

Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
1                      5                      10                      15
Gly Phe Ala Asp Ala Glu Gly Ser Phe Gly Leu Ser Ile Leu Asn Arg
      20                      25                      30
Asn Arg Gly Thr Ala Arg Tyr His Thr Arg Leu Ser Phe Ala Ile Val

```

35	40	45
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp		
50	55	60
Lys Val Gly Ile Ile Thr Asn Asp Gly Asp Arg Tyr Val Arg Leu Arg		
65	70	75
Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys		
85	90	95
Tyr Pro Leu Val Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln		
100	105	110
Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile		
115	120	125
Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp		
130	135	140
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu		
145	150	155
Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser		
165	170	175
Gly Asp Gly Ser Phe Phe Val Arg Leu Arg Lys Ser Asn Val Asn Ala		
180	185	190
Arg Val Arg Val Gln Leu Val Phe Glu Ile Ser Gln His Ile Arg Asp		
195	200	205
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly His Ile		
210	215	220
Tyr Glu Gly Asn Lys Ser Glu Arg Ser Trp Leu Gln Phe Arg Val Glu		
225	230	235
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn		
245	250	255
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val		
260	265	270
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp		
275	280	285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa		
290	295	300

<210> 12

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.ile3.外显子1_RD2_B1H8

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 12

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gly	Leu	Ser	Ile	Leu	Asn	Arg
			20					25					30		
Asn	Arg	Gly	Thr	Ala	Arg	Tyr	His	Thr	Arg	Leu	Ser	Phe	Thr	Ile	Val
		35					40					45			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
	50					55				60					
Lys	Val	Gly	Ile	Ile	Thr	Asn	Asp	Gly	Asp	Arg	Tyr	Val	Arg	Leu	Cys
65					70				75					80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85					90					95		
Tyr	Pro	Leu	Val	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
		100						105					110		
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
		115					120					125			
Lys	Glu	Leu	Ala	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
	130					135				140					
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Gly	Lys	Glu	Arg	Pro	Leu
145				150					155					160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165					170					175		
Gly	Asp	Gly	Ser	Phe	Phe	Val	Arg	Leu	Arg	Lys	Ser	Asn	Val	Asn	Ala
		180					185					190			
Arg	Val	Arg	Val	Gln	Leu	Val	Phe	Glu	Ile	Ser	Gln	His	Ile	Arg	Asp
		195				200						205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	His	Ile
	210					215				220					
Tyr	Glu	Gly	Asn	Lys	Ser	Glu	Arg	Ser	Trp	Leu	Gln	Phe	Arg	Val	Glu

225	230	235	240
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn			
	245	250	255
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val			
	260	265	270
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp			
	275	280	285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa			
	290	295	300

<210> 13

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.IgV.外显子2_RD1_G5

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 13

Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr			
1	5	10	15
Gly Phe Ala Asp Ala Glu Gly Ser Phe Gln Leu Tyr Ile Ser Asn Val			
	20	25	30
Asn Asn Asn Arg Ser Arg Tyr Arg Ala Arg Leu Arg Phe Ala Ile Glu			
	35	40	45
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp			
	50	55	60
Lys Val Gly Val Ile Asn Asn Ile Gly Asp Thr Ser Val Arg Leu Ser			
65	70	75	80
Val Gly Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys			
	85	90	95
Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln			
	100	105	110

Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
115 120 125
Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp
130 135 140
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu
145 150 155 160
Ile Asn Lys Asn Ile Pro Asn Leu Lys Trp Leu Ala Gly Phe Thr Ser
165 170 175
Gly Asp Gly Asn Phe Tyr Val His Leu Lys Lys Ser Gly Arg Thr Thr
180 185 190
Arg Val Tyr Val Gln Leu Arg Phe Ser Ile Ala Gln His Ile Arg Asp
195 200 205
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile
210 215 220
Asn Glu Trp Asn Ala Ser Glu Arg Ser Ala Leu Glu Phe Arg Val Thr
225 230 235 240
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
245 250 255
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
260 265 270
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
275 280 285
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa
290 295 300

<210> 14

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-0nuI变体PD-1.IgV.外显子2_RD1_PS3

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 14

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Cys	Phe	Leu	Leu	His	Ile	Arg	Asn	Leu
			20					25					30		
Asn	Arg	Thr	Ser	Thr	Lys	Tyr	Arg	Thr	Arg	Leu	Ser	Phe	Glu	Ile	Glu
		35					40					45			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
	50					55				60					
Lys	Val	Gly	Ile	Ile	Asn	Asn	Ile	Gly	Asn	Arg	Arg	Val	Arg	Leu	Ser
65					70				75					80	
Val	Arg	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85					90				95			
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
		100						105				110			
Ala	Phe	Ser	Leu	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
	115						120					125			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
	130						135				140				
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145					150				155					160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165						170				175		
Gly	Asp	Gly	Asn	Phe	Tyr	Val	His	Leu	Lys	Lys	Ser	Gly	Arg	Thr	Thr
		180						185				190			
Arg	Val	Tyr	Val	Gln	Leu	Arg	Phe	Ser	Ile	Ser	Gln	His	Ile	Arg	Asp
	195						200					205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Tyr	Ile
	210					215					220				
Thr	Glu	Ser	Asn	Pro	Ser	Glu	Arg	Ser	Asp	Leu	Glu	Phe	Arg	Val	Thr
225					230				235					240	
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
			245					250				255			
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
		260					265					270			
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
	275						280					285			
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Xaa	Xaa	
	290					295					300				

<210> 15

<211> 870

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成megaTAL PD-1.ITSM.ex5_RD1_CV3-08

<400> 15

```

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg
1           5           10           15
Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val
20           25           30
Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe
35           40           45
Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly
50           55           60
Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala
65           70           75           80
Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
85           90           95
Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro
100          105          110
Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly
115          120          125
Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly
130          135          140
Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
145          150          155          160
Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
165          170          175
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
180          185          190
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
195          200          205
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
210          215          220
Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
225          230          235          240
Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val

```

	245		250		255
Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln					
	260		265		270
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln					
	275		280		285
Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr					
	290		295		300
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro					
305		310		315	320
Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu					
	325		330		335
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu					
	340		345		350
Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln					
	355		360		365
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His					
	370		375		380
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly					
385		390		395	400
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln					
	405		410		415
Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn					
	420		425		430
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu					
	435		440		445
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser					
	450		455		460
Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro					
465		470		475	480
Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile					
	485		490		495
Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln					
	500		505		510
Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu					
	515		520		525
Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys					
	530		535		540
Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg					
545		550		555	560

Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Ser Arg Arg		
565	570	575
Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe Ala Asp Ala Glu Gly		
580	585	590
Ser Phe Gln Leu Glu Ile Arg Asn Val Asn Pro Asn Ile Pro Arg Tyr		
595	600	605
Lys Thr Arg Leu Arg Phe Glu Ile Asp Leu His Asn Lys Asp Lys Ser		
610	615	620
Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val Gly Lys Ile Tyr Asn		
625	630	635
Gln Gly Asp Ser Tyr Val Lys Leu Arg Val Thr Arg Phe Glu Asp Leu		
645	650	655
Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Ile Thr Gln Lys		
660	665	670
Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val Met Glu Asn		
675	680	685
Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu Leu Val Arg Ile Lys		
690	695	700
Ala Lys Met Asn Trp Gly Leu Asn Asp Glu Leu Lys Lys Ala Phe Pro		
705	710	715
Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn Lys Asn Ile Pro Asn		
725	730	735
Leu Lys Trp Leu Ala Gly Phe Thr Ser Gly Glu Gly His Phe Gly Val		
740	745	750
Ile Leu Ala Lys Arg Arg Pro Ala Ser Pro Val Gln Val Arg Leu Arg		
755	760	765
Phe Ala Ile Gly Gln His Ile Arg Asp Lys Asn Leu Met Asn Ser Leu		
770	775	780
Ile Thr Tyr Leu Gly Cys Gly Arg Ile Arg Glu Lys Asn Ile Ser Glu		
785	790	795
Lys Ser Trp Leu Glu Phe Glu Val Thr Lys Phe Ser Asp Ile Asn Asp		
805	810	815
Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly Val Lys Leu		
820	825	830
Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile Glu Glu Lys		
835	840	845
Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys Ile Lys Leu		
850	855	860
Asn Met Asn Lys Gly Arg		

865 870
 <210> 16
 <211> 870
 <212> PRT
 <213> 人工序列(Artificial Sequence)
 <220>
 <221>
 <223> 合成megaTAL PD-1.ITSM.ex5_RD2_73
 <400> 16
 Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg
 1 5 10 15
 Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val
 20 25 30
 Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe
 35 40 45
 Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly
 50 55 60
 Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala
 65 70 75 80
 Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
 85 90 95
 Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro
 100 105 110
 Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly
 115 120 125
 Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly
 130 135 140
 Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 145 150 155 160
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 165 170 175
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 180 185 190
 Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 195 200 205
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 210 215 220
 Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 225 230 235 240

Leu Pro Val	Leu Cys Gln Asp His Gly	Leu Thr Pro Asp Gln Val Val
245	250	255
Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln		
260	265	270
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln		
275	280	285
Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr		
290	295	300
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro		
305	310	315
Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu		
325	330	335
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu		
340	345	350
Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln		
355	360	365
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His		
370	375	380
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly		
385	390	395
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln		
405	410	415
Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn		
420	425	430
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu		
435	440	445
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser		
450	455	460
Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro		
465	470	475
Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile		
485	490	495
Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln		
500	505	510
Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu		
515	520	525
Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys		
530	535	540
Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg		

545	550	555	560
Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Ser Arg Arg			
	565	570	575
Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe Ala Asp Ala Glu Gly			
	580	585	590
Ser Phe Gln Leu Glu Ile Arg Asn Val Asn Pro Asn Ile Pro Arg Tyr			
	595	600	605
Arg Thr Arg Leu Arg Phe Glu Ile Asp Leu His Asn Lys Asp Lys Ser			
	610	615	620
Ile Leu Glu Asp Ile Gln Ser Thr Trp Lys Val Gly Lys Ile Tyr Asn			
625	630	635	640
Arg Gly Asp Ser Tyr Val Lys Leu Arg Val Thr Arg Phe Glu Asp Leu			
	645	650	655
Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Ile Thr Gln Lys			
	660	665	670
Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val Met Glu Asn			
	675	680	685
Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu Leu Val Arg Ile Lys			
	690	695	700
Ala Lys Met Asn Trp Gly Leu Asn Asp Glu Leu Lys Lys Ala Phe Pro			
705	710	715	720
Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn Lys Asn Ile Pro Asn			
	725	730	735
Leu Lys Trp Leu Ala Gly Phe Thr Ser Gly Glu Gly His Phe Gly Val			
	740	745	750
Ile Leu Ala Lys Arg Arg Pro Ala Ser Pro Val Gln Val Arg Leu Arg			
	755	760	765
Phe Ala Ile Gly Gln His Ile Arg Asp Lys Asn Leu Met Asn Ser Leu			
	770	775	780
Ile Thr Tyr Leu Gly Cys Gly Arg Thr Arg Glu Lys Asn Ile Ser Glu			
785	790	795	800
Lys Ser Trp Leu Glu Phe Glu Val Thr Lys Phe Ser Asp Ile Asn Asp			
	805	810	815
Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly Val Lys Leu			
	820	825	830
Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile Glu Glu Lys			
	835	840	845
Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys Ile Lys Leu			
	850	855	860

Asn Met Asn Lys Gly Arg

865 870

<210> 17

<211> 870

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成megaTAL PD-1.ITSM.ex5_RD3_03

<400> 17

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg

1 5 10 15

Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val

20 25 30

Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe

35 40 45

Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly

50 55 60

Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala

65 70 75 80

Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg

85 90 95

Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro

100 105 110

Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly

115 120 125

Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly

130 135 140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

145 150 155 160

Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180 185 190

Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu

195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210 215 220

Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu

225	230	235	240
Leu Pro Val Leu Cys Gln Asp His Gly	Leu Thr Pro Asp Gln Val Val		
245	250	255	
Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln	Ala Leu Glu Thr Val Gln		
260	265	270	
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	Leu Thr Pro Asp Gln		
275	280	285	
Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln	Ala Leu Glu Thr		
290	295	300	
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	Leu Thr Pro		
305	310	315	320
Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln	Ala Leu		
325	330	335	
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	Leu		
340	345	350	
Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln			
355	360	365	
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His			
370	375	380	
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly			
385	390	395	400
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln			
405	410	415	
Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn			
420	425	430	
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu			
435	440	445	
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser			
450	455	460	
Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro			
465	470	475	480
Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile			
485	490	495	
Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln			
500	505	510	
Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu			
515	520	525	
Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys			
530	535	540	

Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg			
545	550	555	560
Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Ser Arg Arg			
	565	570	575
Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe Ala Asp Ala Glu Gly			
	580	585	590
Ser Phe Gln Leu Glu Ile Arg Asn Val Asn Pro Asn Ile Pro Arg Tyr			
	595	600	605
Arg Thr Arg Leu Arg Phe Glu Ile Asp Leu His Asn Lys Asp Lys Ser			
	610	615	620
Ile Leu Glu Asp Ile Gln Ser Thr Trp Lys Val Gly Lys Ile Tyr Asn			
625	630	635	640
Gln Gly Asp Ser Tyr Val Lys Leu Arg Val Thr Arg Phe Glu Asp Leu			
	645	650	655
Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Ile Thr Gln Lys			
	660	665	670
Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val Met Glu Asn			
	675	680	685
Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu Leu Val Arg Ile Lys			
	690	695	700
Ala Lys Met Asn Trp Gly Leu Asn Asp Glu Leu Lys Lys Ala Phe Pro			
705	710	715	720
Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn Lys Asn Ile Pro Asn			
	725	730	735
Leu Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp Gly His Phe Gly Val			
	740	745	750
Ile Leu Ala Lys Arg Arg Pro Ala Ser Pro Val Gln Val Arg Leu Arg			
	755	760	765
Phe Ala Ile Gly Gln His Ile Arg Asp Lys Asn Leu Met Asn Ser Leu			
	770	775	780
Ile Thr Tyr Leu Gly Cys Gly Arg Thr Arg Glu Lys Asn Ile Ser Glu			
785	790	795	800
Lys Ser Trp Leu Glu Phe Glu Val Thr Lys Phe Ser Asp Ile Asn Asp			
	805	810	815
Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly Val Lys Leu			
	820	825	830
Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile Glu Glu Lys			
	835	840	845
Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys Ile Lys Leu			

850	855	860
Asn Met Asn Lys Gly Arg		
865	870	
<210> 18		
<211> 870		
<212> PRT		
<213> 人工序列(Artificial Sequence)		
<220>		
<221>		
<223> 合成megaTAL PD-1.ITSM.ex5_RD4_CV23		
<400> 18		
Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg		
1 5 10 15		
Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val		
20 25 30		
Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe		
35 40 45		
Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly		
50 55 60		
Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala		
65 70 75 80		
Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg		
85 90 95		
Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro		
100 105 110		
Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly		
115 120 125		
Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly		
130 135 140		
Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn		
145 150 155 160		
Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		
165 170 175		
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
180 185 190		
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
195 200 205		
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala		
210 215 220		

Ile	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu
225						230				235					240
Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val
				245						250					255
Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln
				260						265					270
Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln
				275						280					285
Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr
				290						295					300
Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro
305						310						315			320
Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu
				325								330			335
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu
				340								345			350
Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln
				355								360			365
Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His
				370								375			380
Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly
385						390						395			400
Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln
				405								410			415
Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn
				420								425			430
Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu
				435								440			445
Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser
				450								455			460
Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro
465						470						475			480
Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile
				485								490			495
Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln
				500								505			510
Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu
				515								520			525
Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	Lys

530	535	540
Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg		
545	550	555
Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Ser Arg Arg		560
	565	570
Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe Ala Asp Ala Glu Gly		575
	580	585
Ser Phe Gln Leu Glu Ile Arg Asn Val Asn Pro Asn Ile Pro Arg Tyr		590
	595	600
Arg Thr Arg Leu Arg Phe Glu Ile Asp Leu His Asn Lys Asp Lys Ser		605
610	615	620
Ile Leu Glu Asp Ile Gln Ser Thr Trp Lys Val Gly Lys Ile Tyr Asn		
625	630	635
Gln Gly Asp Ser Tyr Val Lys Leu Arg Val Thr Arg Phe Glu Asp Leu		640
	645	650
Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Ile Thr Gln Lys		655
	660	665
Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val Met Glu Asn		670
675	680	685
Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu Leu Val Arg Ile Lys		
690	695	700
Ala Lys Met Asn Trp Gly Leu Asn Asp Glu Leu Lys Lys Ala Phe Pro		
705	710	715
Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn Lys Asn Ile Pro Asn		720
	725	730
Leu Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp Gly His Phe Gly Val		735
	740	745
Ile Leu Ala Lys Arg Arg Pro Ala Ser Pro Val Gln Val Arg Leu Arg		
	755	760
Phe Ala Ile Gly Gln His Ile Arg Asp Lys Asn Leu Met Asn Ser Leu		765
770	775	780
Ile Thr Tyr Leu Gly Cys Gly Arg Ile Arg Glu Lys Asn Lys Ser Glu		
785	790	795
Lys Ser Trp Leu Glu Phe Glu Val Thr Lys Phe Ser Asp Ile Asn Asp		800
	805	810
Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly Val Lys Leu		815
	820	825
Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile Glu Glu Lys		830
835	840	845

Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys Ile Lys Leu
 850 855 860
 Asn Met Asn Lys Gly Arg
 865 870
 <210> 19
 <211> 870
 <212> PRT
 <213> 人工序列(Artificial Sequence)
 <220>
 <221>
 <223> 合成megaTAL PD-1.ITSM.ex5_RD5_CV23MK
 <400> 19
 Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg
 1 5 10 15
 Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val
 20 25 30
 Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe
 35 40 45
 Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly
 50 55 60
 Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala
 65 70 75 80
 Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
 85 90 95
 Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro
 100 105 110
 Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly
 115 120 125
 Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly
 130 135 140
 Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 145 150 155 160
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 165 170 175
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 180 185 190
 Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 195 200 205
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210	215	220
Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu		
225	230	235
Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val		240
	245	250
Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln		255
	260	265
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln		270
	275	280
Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr		285
290	295	300
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro		
305	310	315
Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu		320
	325	330
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu		335
	340	345
Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln		350
	355	360
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His		365
370	375	380
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly		
385	390	395
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln		400
	405	410
Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn		415
	420	425
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu		430
	435	440
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser		445
450	455	460
Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro		
465	470	475
Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile		480
	485	490
Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln		495
	500	505
Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu		510
515	520	525

Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	Lys
530						535					540				
Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	Ile	Arg	Arg	Val	Asn	Arg	Arg
545					550					555				560	
Ile	Gly	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Ile	Ser	Arg	Ser	Arg	Arg
				565					570					575	
Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr	Gly	Phe	Ala	Asp	Ala	Glu	Gly
			580						585					590	
Ser	Phe	Gln	Leu	Glu	Ile	Arg	Asn	Val	Asn	Pro	Asn	Ile	Pro	Arg	Tyr
		595					600					605			
Arg	Thr	Arg	Leu	Arg	Phe	Glu	Ile	Asp	Leu	His	Asn	Lys	Asp	Lys	Ser
		610				615						620			
Ile	Leu	Glu	Asp	Ile	Gln	Ser	Thr	Trp	Lys	Val	Gly	Lys	Ile	Tyr	Asn
625					630					635					640
Gln	Gly	Asp	Ser	Tyr	Val	Lys	Leu	Arg	Val	Thr	Arg	Phe	Glu	Asp	Leu
				645					650					655	
Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys	Tyr	Pro	Leu	Ile	Thr	Gln	Lys
			660						665					670	
Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln	Ala	Phe	Ser	Val	Met	Glu	Asn
		675						680					685		
Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile	Lys	Glu	Leu	Val	Arg	Ile	Lys
		690				695						700			
Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp	Glu	Leu	Lys	Lys	Ala	Phe	Pro
705					710					715					720
Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu	Ile	Asn	Lys	Asn	Ile	Pro	Asn
				725						730				735	
Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser	Gly	Asp	Gly	His	Phe	Gly	Val
			740						745					750	
Ile	Leu	Ala	Lys	Arg	Arg	Pro	Ala	Ser	Pro	Val	Gln	Val	Arg	Leu	Arg
		755					760						765		
Phe	Met	Ile	Gly	Gln	His	Ile	Lys	Asp	Lys	Asn	Leu	Met	Asn	Ser	Leu
		770					775					780			
Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Arg	Ile	Arg	Glu	Lys	Asn	Lys	Ser	Glu
785					790					795					800
Lys	Ser	Trp	Leu	Glu	Phe	Glu	Val	Thr	Lys	Phe	Ser	Asp	Ile	Asn	Asp
				805						810				815	
Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn	Thr	Leu	Ile	Gly	Val	Lys	Leu
			820						825					830	
Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val	Ala	Lys	Leu	Ile	Glu	Glu	Lys

835	840	845
Lys His Leu Thr Glu Ser Gly	Leu Asp Glu Ile Lys	Lys Lys Ile Lys Leu
850	855	860
Asn Met Asn Lys Gly Arg		
865	870	
<210> 20		
<211> 841		
<212> PRT		
<213> 人工序列(Artificial Sequence)		
<220>		
<221>		
<223> 合成megaTAL PD-1.ile3.外显子1_RD1 B1		
<400> 20		
Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg		
1	5	10
Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val		
20	25	30
Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe		
35	40	45
Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly		
50	55	60
Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala		
65	70	75
Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg		
85	90	95
Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro		
100	105	110
Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly		
115	120	125
Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly		
130	135	140
Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn		
145	150	155
Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		
165	170	175
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
180	185	190
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
195	200	205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala			
210	215	220	
Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg			
225	230	235	240
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val			
245	250	255	
Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val			
260	265	270	
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp			
275	280	285	
Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu			
290	295	300	
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr			
305	310	315	320
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala			
325	330	335	
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly			
340	345	350	
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys			
355	360	365	
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp			
370	375	380	
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly			
385	390	395	400
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys			
405	410	415	
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn			
420	425	430	
Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val			
435	440	445	
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala			
450	455	460	
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln Leu			
465	470	475	480
Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val			
485	490	495	
Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys Lys			
500	505	510	
Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg Ile			

515	520	525
Gly Glu Arg Thr Ser His Arg	Val Ala Ile Ser Arg	Val Gly Gly Ser
530	535	540
Ser Arg Arg Glu Ser Ile Asn	Pro Trp Ile Leu Thr Gly	Phe Ala Asp
545	550	555
Ala Glu Gly Ser Phe Gly Leu	Ser Ile Leu Asn Arg	Asn Arg Gly Thr
565	570	575
Ala Arg Tyr His Thr Arg Leu	Ser Phe Ala Ile Val Leu	His Asn Lys
580	585	590
Asp Lys Ser Ile Leu Glu Asn	Ile Gln Ser Thr Trp Lys	Val Gly Ile
595	600	605
Ile Thr Asn Asp Gly Asp Arg	Tyr Val Arg Leu Arg	Val Thr Arg Phe
610	615	620
Glu Asp Leu Lys Val Ile Ile	Asp His Phe Glu Lys Tyr	Pro Leu Val
625	630	635
Thr Gln Lys Leu Gly Asp Tyr	Lys Leu Phe Lys Gln Ala	Phe Ser Val
645	650	655
Met Glu Asn Lys Glu His Leu	Lys Glu Asn Gly Ile Lys	Glu Leu Val
660	665	670
Arg Ile Lys Ala Lys Met Asn	Trp Gly Leu Asn Asp Glu	Leu Lys Lys
675	680	685
Ala Phe Pro Glu Asn Ile Ser	Lys Glu Arg Pro Leu Ile	Asn Lys Asn
690	695	700
Ile Pro Asn Phe Lys Trp Leu	Ala Gly Phe Thr Ser Gly	Asp Gly Ser
705	710	715
Phe Phe Val Arg Leu Arg Lys	Ser Asn Val Asn Ala Arg	Val Arg Val
725	730	735
Gln Leu Val Phe Glu Ile Ser	Gln His Ile Arg Asp Lys	Asn Leu Met
740	745	750
Asn Ser Leu Ile Thr Tyr Leu	Gly Cys Gly His Ile Tyr	Glu Gly Asn
755	760	765
Lys Ser Glu Arg Ser Trp Leu	Gln Phe Arg Val Glu Lys	Phe Ser Asp
770	775	780
Ile Asn Asp Lys Ile Ile Pro	Val Phe Gln Glu Asn Thr	Leu Ile Gly
785	790	795
Val Lys Leu Glu Asp Phe Glu	Asp Trp Cys Lys Val Ala	Lys Leu Ile
805	810	815
Glu Glu Lys Lys His Leu Thr	Glu Ser Gly Leu Asp Glu	Ile Lys Lys
820	825	830

Ile Lys Leu Asn Met Asn Lys Gly Arg

835

840

<210> 21

<211> 841

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成megaTAL PD-1.ile3.外显子1_RD2_B1H8

<400> 21

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg

1

5

10

15

Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val

20

25

30

Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe

35

40

45

Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly

50

55

60

Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala

65

70

75

80

Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg

85

90

95

Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro

100

105

110

Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly

115

120

125

Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly

130

135

140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

145

150

155

160

Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165

170

175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180

185

190

Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu

195

200

205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210

215

220

Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg

225	230	235	240
Leu Leu Pro Val	Leu Cys Gln Asp His Gly	Leu Thr Pro Asp Gln Val	
	245	250	255
Val Ala Ile Ala	Ser Asn Asn Gly Gly Lys Gln Ala	Leu Glu Thr Val	
	260	265	270
Gln Arg Leu Leu Pro Val	Leu Cys Gln Asp His Gly	Leu Thr Pro Asp	
	275	280	285
Gln Val Val Ala Ile Ala	Ser Asn Asn Gly Gly Lys Gln Ala	Leu Glu	
	290	295	300
Thr Val Gln Arg Leu Leu Pro Val	Leu Cys Gln Asp His Gly	Leu Thr	
305	310	315	320
Pro Asp Gln Val Val Ala Ile Ala	Ser Asn Asn Gly Gly Lys Gln Ala		
	325	330	335
Leu Glu Thr Val Gln Arg Leu Leu Pro Val	Leu Cys Gln Asp His Gly		
	340	345	350
Leu Thr Pro Asp Gln Val Val Ala Ile Ala	Ser Asn Asn Gly Gly Lys		
	355	360	365
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	Leu Cys Gln Asp		
	370	375	380
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	Ser His Asp Gly		
385	390	395	400
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	Leu Cys		
	405	410	415
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	Ser Asn		
	420	425	430
Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val			
	435	440	445
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala			
	450	455	460
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln Leu			
465	470	475	480
Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val			
	485	490	495
Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys Lys			
	500	505	510
Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg Ile			
	515	520	525
Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly Gly Ser			
	530	535	540

Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe Ala Asp			
545	550	555	560
Ala Glu Gly Ser Phe Gly Leu Ser Ile Leu Asn Arg Asn Arg Gly Thr			
	565	570	575
Ala Arg Tyr His Thr Arg Leu Ser Phe Thr Ile Val Leu His Asn Lys			
	580	585	590
Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val Gly Ile			
	595	600	605
Ile Thr Asn Asp Gly Asp Arg Tyr Val Arg Leu Cys Val Thr Arg Phe			
	610	615	620
Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Val			
	625	630	635
Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val			
	645	650	655
Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu Leu Ala			
	660	665	670
Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp Glu Leu Lys Lys			
	675	680	685
Ala Phe Pro Glu Asn Ile Gly Lys Glu Arg Pro Leu Ile Asn Lys Asn			
	690	695	700
Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp Gly Ser			
	705	710	715
Phe Phe Val Arg Leu Arg Lys Ser Asn Val Asn Ala Arg Val Arg Val			
	725	730	735
Gln Leu Val Phe Glu Ile Ser Gln His Ile Arg Asp Lys Asn Leu Met			
	740	745	750
Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly His Ile Tyr Glu Gly Asn			
	755	760	765
Lys Ser Glu Arg Ser Trp Leu Gln Phe Arg Val Glu Lys Phe Ser Asp			
	770	775	780
Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly			
	785	790	795
Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile			
	805	810	815
Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys			
	820	825	830
Ile Lys Leu Asn Met Asn Lys Gly Arg			
	835	840	

<210> 22

<211> 909

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成megaTAL PD-1.IgV.外显子2_RD1_G5

<400> 22

```

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg
1           5           10           15
Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val
20           25           30
Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe
35           40           45
Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly
50           55           60
Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala
65           70           75           80
Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
85           90           95
Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro
100          105          110
Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly
115          120          125
Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly
130          135          140
Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
145          150          155          160
Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
165          170          175
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
180          185          190
Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
195          200          205
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
210          215          220
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
225          230          235          240
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
245          250          255

```

Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val				
																260	265	270	
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp				
																275	280	285	
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu				
																290	295	300	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr				
																305	310	315	320
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala				
																325	330	335	
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly				
																340	345	350	
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys				
																355	360	365	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp				
																370	375	380	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly				
																385	390	395	400
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys				
																405	410	415	
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn				
																420	425	430	
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val				
																435	440	445	
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala				
																450	455	460	
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu				
																465	470	475	480
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala				
																485	490	495	
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg				
																500	505	510	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val				
																515	520	525	
Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile				
																530	535	540	
Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn				
																545	550	555	560
Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp				

				565					570					575			
Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	Ile	Arg	Arg	Val		
				580					585					590			
Asn	Arg	Arg	Ile	Gly	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Ile	Ser	Arg		
				595					600					605			
Val	Gly	Gly	Ser	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr		
				610					615					620			
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gln	Leu	Tyr	Ile	Ser	Asn	Val		
625					630				635					640			
Asn	Asn	Asn	Arg	Ser	Arg	Tyr	Arg	Ala	Arg	Leu	Arg	Phe	Ala	Ile	Glu		
				645					650					655			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp		
				660					665					670			
Lys	Val	Gly	Val	Ile	Asn	Asn	Ile	Gly	Asp	Thr	Ser	Val	Arg	Leu	Ser		
				675					680					685			
Val	Gly	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys		
				690					695					700			
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln		
705					710				715					720			
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile		
				725					730					735			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp		
				740					745					750			
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu		
				755					760					765			
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser		
				770					775					780			
Gly	Asp	Gly	Asn	Phe	Tyr	Val	His	Leu	Lys	Lys	Ser	Gly	Arg	Thr	Thr		
785					790				795					800			
Arg	Val	Tyr	Val	Gln	Leu	Arg	Phe	Ser	Ile	Ala	Gln	His	Ile	Arg	Asp		
				805					810					815			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Tyr	Ile		
				820					825					830			
Asn	Glu	Trp	Asn	Ala	Ser	Glu	Arg	Ser	Ala	Leu	Glu	Phe	Arg	Val	Thr		
				835					840					845			
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn		
				850					855					860			
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val		
865					870				875					880			

Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
885 890 895

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg
900 905

<210> 23
<211> 909
<212> PRT
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 合成megaTAL PD-1.IgV.外显子2_RD1_PS3
<400> 23

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg
1 5 10 15

Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val
20 25 30

Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe
35 40 45

Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly
50 55 60

Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala
65 70 75 80

Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
85 90 95

Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro
100 105 110

Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly
115 120 125

Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly
130 135 140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
145 150 155 160

Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
180 185 190

Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210	215	220
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
225	230	235
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		240
	245	250
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val		255
	260	265
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp		270
	275	280
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu		285
	290	295
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr		300
305	310	315
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala		320
	325	330
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		335
	340	345
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys		350
	355	360
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		365
	370	375
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly		380
385	390	395
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		400
	405	410
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn		415
	420	425
Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		430
	435	440
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		445
	450	455
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		460
465	470	475
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala		480
	485	490
Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		495
	500	505
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		510
	515	520
		525

Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Ser Ile		
530	535	540
Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn		
545	550	555
Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp		
	565	570
Ala Val Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val		
	580	585
Asn Arg Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg		
	595	600
Val Gly Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr		
	610	615
Gly Phe Ala Asp Ala Glu Gly Cys Phe Leu Leu His Ile Arg Asn Leu		
625	630	635
Asn Arg Thr Ser Thr Lys Tyr Arg Thr Arg Leu Ser Phe Glu Ile Glu		
	645	650
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp		
	660	665
Lys Val Gly Ile Ile Asn Asn Ile Gly Asn Arg Arg Val Arg Leu Ser		
	675	680
Val Arg Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys		
	690	695
Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln		
705	710	715
Ala Phe Ser Leu Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile		
	725	730
Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp		
	740	745
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu		
	755	760
Ile Asn Lys Asn Ile Pro Asn Leu Lys Trp Leu Ala Gly Phe Thr Ser		
	770	775
Gly Asp Gly Asn Phe Tyr Val His Leu Lys Lys Ser Gly Arg Thr Thr		
785	790	795
Arg Val Tyr Val Gln Leu Arg Phe Ser Ile Ser Gln His Ile Arg Asp		
	805	810
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile		
	820	825
Thr Glu Ser Asn Pro Ser Glu Arg Ser Asp Leu Glu Phe Arg Val Thr		

835	840	845
Lys Phe Ser Asp Ile Asn Asp	Lys Ile Ile Pro Val	Phe Gln Glu Asn
850	855	860
Thr Leu Ile Gly Val Lys Leu	Glu Asp Phe Glu Asp	Trp Cys Lys Val
865	870	875
Ala Lys Leu Ile Glu Glu Lys	Lys His Leu Thr Glu Ser	Gly Leu Asp
885	890	895
Glu Ile Lys Lys Ile Lys Leu	Asn Met Asn Lys Gly Arg	
900	905	
<210> 24		
<211> 236		
<212> PRT		
<213> 小家鼠 (Mus musculus)		
<400> 24		
Met Ser Glu Pro Pro Arg Ala	Glu Thr Phe Val Phe Leu Asp	Leu Glu
1	5	10
Ala Thr Gly Leu Pro Asn Met	Asp Pro Glu Ile Ala Glu Ile	Ser Leu
20	25	30
Phe Ala Val His Arg Ser Ser	Leu Glu Asn Pro Glu Arg Asp	Asp Ser
35	40	45
Gly Ser Leu Val Leu Pro Arg	Val Leu Asp Lys Leu Thr Leu	Cys Met
50	55	60
Cys Pro Glu Arg Pro Phe Thr	Ala Lys Ala Ser Glu Ile Thr	Gly Leu
65	70	75
Ser Ser Glu Ser Leu Met His	Cys Gly Lys Ala Gly Phe Asn	Gly Ala
85	90	95
Val Val Arg Thr Leu Gln Gly	Phe Leu Ser Arg Gln Glu Gly	Pro Ile
100	105	110
Cys Leu Val Ala His Asn Gly	Phe Asp Tyr Asp Phe Pro Leu	Leu Cys
115	120	125
Thr Gly Leu Gln Arg Leu Gly	Ala His Leu Pro Gln Asp Thr	Val Cys
130	135	140
Leu Asp Thr Leu Pro Ala Leu	Arg Gly Leu Asp Arg Ala His	Ser His
145	150	155
Gly Thr Arg Ala Gln Gly Arg	Lys Ser Tyr Ser Leu Ala Ser	Leu Phe
165	170	175
His Arg Tyr Phe Gln Ala Glu	Pro Ser Ala Ala His Ser Ala	Glu Gly
180	185	190
Asp Val His Thr Leu Leu Leu	Ile Phe Leu His Arg Ala Pro	Glu Leu

195	200	205
Leu Ala Trp Ala Asp Glu Gln	Ala Arg Ser Trp Ala His Ile Glu Pro	
210	215	220
Met Tyr Val Pro Pro Asp Gly	Pro Ser Leu Glu Ala	
225	230	235
<210> 25		
<211> 22		
<212> DNA		
<213> 智人(Homo sapiens)		
<400> 25		
aatggtggca tactccgtct gc 22		
<210> 26		
<211> 11		
<212> DNA		
<213> 智人(Homo sapiens)		
<400> 26		
tccgctagga a 11		
<210> 27		
<211> 37		
<212> DNA		
<213> 智人(Homo sapiens)		
<400> 27		
tccgctagga aagacaatgg tggcatactc cgtctgc 37		
<210> 28		
<211> 22		
<212> DNA		
<213> 智人(Homo sapiens)		
<400> 28		
aatggtggca tacaaccttt ta 22		
<210> 29		
<211> 22		
<212> DNA		
<213> 智人(Homo sapiens)		
<400> 29		
tttccactta tactccgtct gc 22		
<210> 30		
<211> 22		
<212> DNA		
<213> 智人(Homo sapiens)		

<400> 30
ggcatgcaga tcccacaggc gc 22
<210> 31
<211> 14
<212> DNA
<213> 智人(Homo sapiens)
<400> 31
ggtggggctg ctcc 14
<210> 32
<211> 37
<212> DNA
<213> 智人(Homo sapiens)
<400> 32
ggtggggctg ctccaggcat gcagatccca caggcgc 37
<210> 33
<211> 22
<212> DNA
<213> 智人(Homo sapiens)
<400> 33
ggcatgcaga tccaaccttt ta 22
<210> 34
<211> 22
<212> DNA
<213> 智人(Homo sapiens)
<400> 34
tttccactta tcccacaggc gc 22
<210> 35
<211> 22
<212> DNA
<213> 智人(Homo sapiens)
<400> 35
acgggctga cttccacatg ag 22
<210> 36
<211> 12
<212> DNA
<213> 智人(Homo sapiens)
<400> 36
gtcacacaac tg 12
<210> 37

<211> 38
 <212> DNA
 <213> 智人(Homo sapiens)
 <400> 37
 gtcacacaac tgcccaacgg gcgtgacttc cacatgag 38
 <210> 38
 <211> 22
 <212> DNA
 <213> 智人(Homo sapiens)
 <400> 38
 acgggcgtga cttaaccttt ta 22
 <210> 39
 <211> 22
 <212> DNA
 <213> 智人(Homo sapiens)
 <400> 39
 tttccactta cttccacatg ag 22
 <210> 40
 <211> 2650
 <212> RNA
 <213> 人工序列(Artificial Sequence)
 <220>
 <221>
 <223> 合成PD-1外显子5 RD5_CV23MK megaTAL mRNA
 <400> 40
 augggauccg ccccccccaa gaagaagcgc aagguggugg accugagaac ccuggggguac 60
 agccagcagc aacaggagaa gaucaaaccg aaggugcgca gcacgguggc ccagcaccau 120
 gaagcccucg ugguucaugg auucacgcac gccacaucg uggcccuguc gcagcauccg 180
 gccgcccugg gcaccguggc ggugaccuac cagcacaucg ucaccgccc ugccgaagcc 240
 acccagcagg acaucguggg agugggcaag cagugguucc gagcacgcgc ccuggaggcc 300
 cugcugaccg acgccgggga acugcgcggc ccgccucugc agcuggauac cggccaacug 360
 gucaagaucg ccaagagagg cggagugacc gcgauggagg ccguccacgc cucccggaac 420
 gcucugaccg gugccccgcu caaucugacu ccggaccagg ugguggcuau cgccagcaac 480
 ggaggaggaa aacaggcccu cgaaacagug cagaggcugc ugccuguccu uugucaagau 540
 cacgggcuga cucccgacca ggucguggcc auugccagcc acgacggcgg caagcaggcu 600
 uuggagacug ugcagcgguu ccugccagug cugugccaag aucacggucu gacccagau 660
 caggucgucg ccauugcuuc caacggaggc aaacaagcgc uggaaacggu ccaacgccug 720
 cugccggugc uuugucagga ucacggccug accccugauc aggugguggc caucgcgucc 780
 aaauacgggg ggaagcaggc acucgagacu guccagaggc uccugccugu gcucugccag 840

gaccacgggc ucacacccga ucaggucguc gcuaucgcgu cgcacgacgg uggaaagcag 900
 gcccuaggaaa ccgugcagcg ccuguugccg gugcuguguc aggaccaugg ccuacuccg 960
 gaucaggucg ucgcgaucgc aucuaauggu ggaggaaagc aggcccugga gacaguccag 1020
 cgccugcucc cgguguugug ccaagaccuau ggucuuaccc cugaccaggu ggucgcuauu 1080
 gccucgaaca ucggcggaaa gcaagcccug gaaaccgugc agcgacuucu gccgguccug 1140
 ugccaggauuc auggauugac cccagaccag gugguggcga uugccagcaa caacggcggg 1200
 aagcaagcgu uggaaaccgu ccagagacug cuuccugugc ugugccaaga ccacgguuug 1260
 accccggacc aagucgucgc caucgcuucc aacaacggag ggaagcaggc acucgaaacu 1320
 gugcaacggg ugcugcccgu gcucugucag gaucacggac ucacccuga ucagguggug 1380
 gccaucgcaa gcaacaucgg uggcaaacag gcucuggaaa cuguccaaag acugcugccc 1440
 gugcuuugcc aggaccacgg acugacuccu gaccaagugg uggcaauugc cuccaacauc 1500
 ggaggcaagc aagcgcucga auccaucgug gcgcagcuca gccggccaga ccccgcccug 1560
 gccgcccuga cuaacgauca ccugguggcc cuggcgugcc ucggcggugc ccccgcuau 1620
 gacgcgguga agaaggggcu gccccacgcc cccgagcuca uucggcgggu gaaccgccg 1680
 aucggagaaa gaaccuccca ucggguggcc aucucgagau cacggcggga auccaauaac 1740
 ccuggaucc ugacuggcuu ugccgacgcc gaggauccu uccagcucga aaucggaac 1800
 gugaacccaa acaucccccgu gnaucgcacc agacugcggu ucgagaucga ccuucacaac 1860
 aaggacaagu ccauucugga ggacauccag ucaaccugga aagugggaaa gaucuacaac 1920
 cagggggacu cauacgugaa gcugcgggug acccgcuucg aagaucuaa agugaucac 1980
 gaccauuucg agaaguaccc ccugaucacu cagaagcuug gagacuacaa acuguucaag 2040
 caggcauucu ccgugaugga gaacaaggag caccugaagg agaacgggau uaaggagcug 2100
 gugcgaauua aggcgaaaau gaacugggga uugaacgaug agcugaagaa ggcguucccg 2160
 gaaaacauuu ccaaggagcg cccgcucauc aacaagaaca ucccuaaucu gaaguggcuc 2220
 gcgggguuua ccucuggcga cggccauuuc ggagugauuc ucgcaaagcg caggcccgcc 2280
 agcccugugc aagugcggcu gcgguucaug aucggccagc acauuagga caagaaccug 2340
 augaacucgc ucaucaccua ccuuggaugc ggccggauuc gcgaaaagaa caagagcgag 2400
 aaguccuggc uggaauuuga agugaccaag uucagcgaca ucaacgacaa gaucauuccg 2460
 guguuccagg aaaacacccu gauuggcgug aagcuggagg acuucgagga uuggugcaag 2520
 guggccaagc ucaucgaaga aaagaagcac cugacugaau ccggcuugga cgagaucaag 2580
 aagauuaagc ugaauaugaa caagggaagg ugauagcgcg cuagccguac ggaaaaaaaa 2640
 aaaaaaaaaa 2650

<210> 41

<211> 2526

<212> RNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成PD-1外显子1 RD2_B1H8 megaTAL mRNA

<400> 41

augggaagcg ccccgccgaa gaagaagcgc aagguggugg aucugagaac ccugggauac 60
 agccagcagc agcaggagaa gaucaagccg aagguccggu cuaccguggc ccagcaccu 120
 gagggccuug ugggcccacgg cuucacacau gcacacaucg ucggccuguc gcagcauccc 180
 gccgcccugg ggaccguggc cgugaccuau caacacauca uuaccgccc ugcggaggcc 240
 acccacgagg acaucguggg uguggggaag caguggagcg gagccagggc acucgaagcc 300
 cuccucacug acgcuggaga acugcgcgga ccgccucucc agcuggacac cggacagcug 360
 gugaaaaucg ccaagcgggg aggagugacc gccauggaag ccgugcacgc cucgaggaac 420
 gcgcugacug gcgccccucu gaaccugacc ccugaucagg ucguggcuau cgccucaaac 480
 aacgggggua agcaggcgcu ggagacagug caacgacuuc ugccagugcu uugucaggac 540
 cauggucuga ccccgacca ggucgucgcc auugcaucca acaauggugg caagcaggca 600
 cuggagacug uccagaggcu gcucccgug cugugccagg accacgggcu caccgagc 660
 caauggucg ccaucgccuc caacggagga ggaaaacaag cucuggagac ugugcaacgc 720
 cugcugccug uguugugcca agaccacgga cugacggcg aucagguggu gggaucgca 780
 ucgaacaacg gaggaagca agcgcuggaa accgucagc gccuccugcc cguccucugc 840
 caggauacg gccugacucc ggaccaggug gucgcgaucg ccagcauaa cggggggaag 900
 caagcccucg agacugugca gcgguugcug cccgugcucu gccagaauca uggccuacc 960
 ccagaccaag ucguggccau ugcuuccaac aacgguggca aacaggcgcu cgaaaccguc 1020
 cagcggcugu ugcccugcu uuuccaggau cacggacuca cccugauca ggugguggca 1080
 auugcgucca acaacggugg aaagcaggcc cuggaaacgg ugcagcggu gcuccgguc 1140
 cugugucagg aucaugggcu gacuccgac caggucguc ccauugcauc ccacgauggg 1200
 gguaaacagg ccucgaaac agugcagaga cuccugccag uccugugcca agaccacgga 1260
 cuuaccccg aucagguggu ggccauagcc ucgaacggcg gcgggaaaca ggcucuggaa 1320
 acugugcaaa gacuccucc gguguugugu caagaccaug gacugacccc agaucaggug 1380
 guggcuauug ccucuaacaa cggcggcaag caagcacucg aaagcaucgu ggcccaguug 1440
 ucacgccccg accccgcacu ggcugcccug acgaaugacc aucugguggc gcuggccugc 1500
 cugggagggga ggccagcgau ggaugcggug aagaaggacg ugccccaugc uccggagcug 1560
 auucggagag ugaauaggcg caucggagag agaacuucac aucggguggc cauucuaaga 1620
 gugggcggca gcucccgcg cgaguccauu aaccucgga uccugaccgg cuuugccgac 1680
 gccgaagggu ccuucggccu cucgauccug aaccggaacc gggguaccgc ucgguaccac 1740
 accagacugu ccuucaccau cgugcugcac aacaaggaca agagcauccu cgaaaacauu 1800
 cagucaacgu ggaagguggg aaauauuacu aacgacggcg acagauacgu gcgccugugc 1860
 gugacccggu uugaggaccu gaaggucuu aucgaccacu ucgagaagua cccucugug 1920
 acucagaagc ugggagacua caagcuguuc aagcaggcg ucucggugau ggaaaacaag 1980
 gagcaccuga aggagaacgg caucaaggag cucgcccga ucaaggccaa gaugaacugg 2040
 ggccugaaug augaacucaa gaaggcguuc ccugaaaaca ucgguaaaga acggccccug 2100
 aucaacaaga acauccgaa cuucaagugg cuugccggau ucaccuccg cgacggaucc 2160
 uuucugucc ggcugcgcaa guccaacgug aacgcgagag ugcgggugca auuggucuuu 2220
 gaaucucac agcacaucag ggacaagaau uugaugaacu ccucaucac cuaccugggu 2280
 ugcggacaca ucuacgaagg cauaagucg gagcggagcu ggcugcaguu ucgcguggaa 2340

aaguucuccg acauuacga caagaucac ccaguguucc aggaaaacac ucugauuggc 2400
 gugaagcuug aggaauucga ggacuggugc aagguggcca agcugauuga agagaagaag 2460
 caccugaccg aguccggccu ggacgaauc aagaaaauca agcugaacau gaacaaggga 2520
 cgguga 2526

<210> 42

<211> 2730

<212> RNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成PD-1外显子2 RD1_G5 megaTAL mRNA

<400> 42

augggaagcg cccuuccgaa gaagaagcg aaggucgugg accucagaac ccuggguuac 60
 ucccagcagc agcaggagaa aaauaagccg aaggugcgcu ccaccguggc ucaacaccac 120
 gagggccucg ugggccauug auucacucac gcccauauug uggcccuguc ccagcaccgc 180
 gccgcccugg gcaccguggc ggugaccuac cagcacauca ucaccgcgcu gccugaagcc 240
 acccacgagg acaucgucgg uguggguaag cagugguccg gagccagagc ccuggaggcu 300
 cugcugaccg acgccggaga acucagaggc ccgccucugc agcuggacac cggacagcug 360
 gugaagauag ccaagagagg cggcgugacc gccauugaag ccgugcaugc gucccgaac 420
 gcacugaccg gggccccccu gaaccugacu ccagaccaag ugguggcuau cgccagcaac 480
 aauggaggaa agcaagccuu ggaaaccgug cagcggcugc ucccgguccu uugccaagac 540
 cacggccuga caccggauc ggugguggca aucgaucga auggcggcgg gaagcaggcc 600
 cucgagacug ugcagaggcu ccugccugug cugugccagg accaugggcu gacccagac 660
 cagguggucg caaucgccuc gcacgacgga ggcaagcaag ccuggaaac uguccagcgc 720
 cugcucccug uccuguguca agaucauggg cucacuccug aucaggucgu cgccaucgcc 780
 ucgaacauug guggcaagca ggcgcucgaa accguccagc gguugcugcc agugcuuugc 840
 caggaccaug gucugacccc cgaucaagug gucgcgauug ccucacacga uggcgguaag 900
 caggcuuugg aaaccgugca acggcuguug ccuguccucu gccaggacca cggcuugacu 960
 cccgaccagg ugguggccau agccucaaac aucggaggga aacaagcccu cgaaaccguc 1020
 cagaggcugc ugccgguguu gugccaggau cacggauuga cccagacca gguggucgcc 1080
 auugcuuccc acgauggggg aaagcaggcc cuggagacug ugcagcgccu ccuucccguc 1140
 cugugucagg aucauggacu uacccccgac caagucgugg cgauugcuuc caauaucgga 1200
 ggcaaacagg cccuugaaac agucgagcgc cuguugccgg ugcucugcca agaucacgga 1260
 cugaccccug accagguggu ggcgaucgag ucaaauauc ggcgcaagca ggcacuggag 1320
 acaguccaga gacuccugcc gguccucugc caggaccacg gucuuacucc ugaccaagug 1380
 guggcuauug caucccauga uggugggaaa caggcucuug aaacugugca acgccugcug 1440
 cccgugcugu gccaaagacca cggacugacu ccggaccagg ucguggccau cgcuucaaac 1500
 ggagggggaa aacaggcacu ugaacgggug cagagacugc ugccuguccu uugucaggac 1560
 cacggguuga ccccggaaca ggucguggcu auugccucga acaacggggg gaagcaagcg 1620

cucgagucca uuguggccca gcugagccgg ccugaucccg cacuggccgc gcugaccaac 1680
 gaucaccugg uggcccucgc cugucugggc ggacggccgg ccauggacgc cgugaagaag 1740
 ggacugccgc acgcgcccga gcugauccgc cgcgugaaca ggcggauugg agaacgcacc 1800
 ucccaccggg uggccaucuc gagaguggga ggauccucua gacgggaguc cauaaaccgc 1860
 uggauccuga cuggauucgc cgacgccgag ggguccuucc agcuguacau cuccaacgug 1920
 aacaacaacc gcagcagaua cagggcccgg cugcgguucg cgaucgaacu gcacaauaag 1980
 gacaagagca uccucgagaa cauucagucc acuuggaagg ugggcgugau uaacaacauc 2040
 ggcgauacca gcgugcggcu cuccgugggg cgguucgaag aucucaaggu gaucaucgac 2100
 cacuucgaga aguacccgcg gaucaccag aagcugggag acuacaagcu guucaagcaa 2160
 gcuuucagcg ugauggaaaa caaggaacau cugaaagaga acgguaucua ggaacuggug 2220
 cggaauaagg ccaagaugaa cugggggucg aacgaugaac ugaagaaggc guucccgag 2280
 aaauaucagca aggagcggcc ccugauuaac aaaaacauc caaacuca guggcucgcc 2340
 ggcuuuacuu ccggagaugg uaacuucua gugcaccuga agaaguccg aagaaccacc 2400
 cgggucuaag ugcagcugag guucuccauc gcccgacaca uccgggacaa gaacuugaug 2460
 aacucccuga ucaccuaccu cgguugcgga uacaucaacg aauggaacgc cagcgagaga 2520
 ucggcccugg aguucagggu gacuaaguuc uccgacauca acgacaagau uaucccgug 2580
 uuucaggaaa auaccucau cggcgugaag cuggaggacu uugaggacug gugcaaggug 2640
 gccaaagcuga ucgaggaaaa gaaacaucug accgagagcg gccuggacga aaucaagaag 2700
 auuaagcuga acaugaacaa gggacgauga 2730

<210> 43

<211> 711

<212> RNA

<213> 小家鼠 (*Mus musculus*)

<400> 43

augucugagc caccucgggc ugagaccuuu guauuccugg accuagaagc cacugggcuc 60
 ccaaacaugg acccugagau ugcagagaua uccuuuuug cuguucaccg cucuucccug 120
 gagaacccag aacgggauga uucugguucc uuggugcugc cccguguucu ggacaagcuc 180
 acacugugca ugugcccgga gcgccccuuu acugccaagg ccagugagau uacugguuug 240
 agcagcgaaa gccugaugca cugcgggaag gcugguuua auggcgcugu gguaaggaca 300
 cugcagggcu uccuaagccg ccaggagggc ccaucugcc uuguggccca caauggcuuc 360
 gauuaugacu ucccacugcu gugcacgggg cuacaacguc ugggugccca ucugcccca 420
 gacacugucu gccuggacac acugccugca uugcggggcc uggaccgugc ucacagccac 480
 ggcaccaggg cucaaggccg caaaagcuac agccuggcca gucucuucca ccgcuacuuc 540
 caggcugaac ccagugcugc ccuuucagca gaaggugaug ugcacaccu gcuucugauc 600
 uuccugcauc gugcuccuga gcugcucgcc ugggcagaug agcaggcccg cagcugggcu 660
 cauauugagc ccauguacgu gccaccugau gguccaagcc ucgaagccug a 711

<210> 44

<211> 7601

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子5_MND-BFP构建体

<400> 44

```

cagctgcgcg ctcgctcgct cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc 60
tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
actaggggtt ccttgtagtt aatgattaac ccgccatgct acttatctac gtagaattca 180
gtccagggct ctgtcctgca cctgggggaat ggtgaccggc atctctgtcc tctagctctg 240
gaagcacccc agcccctcta gtctgcctc acccctgacc ctgaccctcc accctgacct 300
cgtcctaacc cctgaccttt gtgcccttcc agagagaagg gcagaagtgc ccacagccca 360
ccccagcccc tcaccagggc cagccggcca gttccaaacc ctggtggttg gtgtcgtggg 420
cggcctgctg ggcagcctgg tgetgctagt ctgggtcctg gccgtcatct gctcccgggc 480
cgcacgaggt aacgtcatcc cagcccctcg gcctgcctg ccctaaccct gctggcgggc 540
ctcactcccg cctccccttc ctccaccctt ccctacccc accccacctc ccccatctc 600
cccgccaggc taagtccctg atgaaggccc ctggactaag accccccacc taggagcacg 660
gctcagggtc ggctgggtga cccaagtgt gtttctctgc agggacaata ggagccaggc 720
gcaccggcca gcccctggtg agtctcactc ttttctgca tgatccactg tgccttcctt 780
cctgggtggg cagaggtgga aggacaggct gggaccacac ggcctgcagg actcacattc 840
tattatagcc aggacccac ctccccagcc cccaggcagc aacctcaatc cctaaagcca 900
tgatctgggg ccccagccca cctgcggtct ccgggggtgc ccggcccatg tgtgtgcctg 960
cctgcggtct ccaggggtgc ctggcccacg cgtgtgcccg cctgcggtct ctgggggtgc 1020
ccggcccaca tatgtgcctg cctgcggtct ccaggtgtgc ccggcccatg cgtgtgcccc 1080
cctgcgaggg cgtgggggtg gcttgggtcat ttcttatctt acattggaga caggagagct 1140
tgaaaagtca cattttgaa tcctaaatct gcaagaatgc caggacatt tcagaggggg 1200
acattgagcc agagaggagg ggtggtgtcc ccagatcaca cagagggcag tgggtgggaca 1260
gctcagggta agcagctcat agtggggggc ccaggttcgg tgccggtact gcagccaggc 1320
tgtggagccg cgggcctcct tcctgcggtg ggccgtgggg ctgactccct ctccctttct 1380
cctcaaagaa ggaggacccc tcagccgtgc ctgtgttctc tgtggactat ggggagctgg 1440
atttccagtg gcgagagaag accccggagc ccccggtgcc ctgtgtccct gagcagacgg 1500
agtaatgcat gaacagagaa acaggagaat atgggccaaa caggatatct gtggtaaagca 1560
gttcctgccc cggctcaggg ccaagaacag ttggaacagc agaatatggg ccaaacagga 1620
tatctgtggt aagcagttcc tgccccggct cagggccaa gaacagatgg cccagatgc 1680
gggtcccgcc tcagcagttt ctagagaacc atcagatgtt tccagggtgc cccaaggacc 1740
tgaaatgacc ctgtgcctta tttgaactaa ccaatcagtt cgcttctcgc ttctgttcgc 1800
gcgcttctgc tccccgagct ctatataagc agagctcggt tagtgaaccg tcagatcgcc 1860
tgagagacgcc atccacgctg ttttgacttc catagaagga tctcgaggcc accatggcta 1920
gcgagctgat taaggagaac atgcacatga agctgtacat ggagggcacc gtggacaacc 1980
atcacttcaa gtgcacatcc gagggcgaag gcaagcccta cgagggcacc cagacatga 2040

```

gaatcaaggt ggctcagggc ggccctctcc ccttcgcctt cgacatcctg gctactagct 2100
 tcctctacgg cagcaagacc ttcatcaacc acaccaggg catccccgac ttcttcaagc 2160
 agtccttccc tgagggcttc acatgggaga gagtaccac atacaggac gggggcgtgc 2220
 tgaccgctac ccaggacacc agcctccagg acggctgcct catctacaac gtcaagatca 2280
 gaggggtgaa cttcacatcc aacggccctg tgatgcagaa gaaaacactc ggctgggagg 2340
 ccttcaccga gacgctgtac cccgctgacg gcggcctgga aggcagaaac gacatggccc 2400
 tgaagctcgt gggcgggagc catctgatcg caaacatcaa gaccacatat agatccaaga 2460
 aacccgctaa gaacctcaag atgcctggcg tctactatgt ggactacaga ctggaaagaa 2520
 tcaaggaggc caacaacgag acctacgtcg agcagcacga ggtggcagtg gccagatact 2580
 gcgacctccc tagcaaaactg gggcacaagc taaattgaaa gctttgcttt atttgtgaaa 2640
 tttgtgatgc tattgcttta tttgtaacca ttataagctg caataaacia gtttaacaac 2700
 aacaattgca ttcatTTTTat gtttcaggtt cagggggagg tgtgggaggT tttttaagT 2760
 cgaccacca ttgtctttcc tagcggaatg ggcacctcat ccccgcccg caggggctca 2820
 gctgacggcc ctcgagtgC ccagccactg aggcctgagg atggacactg ctcttgcccc 2880
 ctctgaccgg cttccttggc caccagtgtt ctgcagacc tccaccatga gcccggtca 2940
 gcgcatttcc tcaggagaag caggcagggt gcaggccatt gcaggccgtc caggggctga 3000
 gctgcctggg ggcgaccggg gctccagcct gcacctgcac caggcacagc cccaccacag 3060
 gactcatgtc tcaatgcccA cagtgagccc aggcagcagg tgtcaccgtc ccctacaggg 3120
 agggccagat gcagtcactg cttcaggtcc tgccagcaca gagctgcctg cgtccagctc 3180
 cctgaatctc tgctgtctgt gctgtctgtg ctgtgtctgc ctgcggcccc gggctgaagg 3240
 cgccgtggcc ctgcctgacg ccccgagcc tctgcctga acttgggggc tggttggaga 3300
 tggccttgga gcagccaagg tgcccctggc agtggcatcc cgaaacgccc tggacgcagg 3360
 gccaagact gggcacagga gtgggaggta catggggctg gggactcccc aggagttatc 3420
 tgctccctgc aggcctagag aagtttcagg gaaggtcaga agagctcctg gctgtggtgg 3480
 gcagggcagg aaacccctcc acctttacac atgcccaggc agcacctcag gccctttgtg 3540
 gggcagggaa gctgaggcag taagcgggca ggcagagctg gaggcctttc aggccagcc 3600
 agcactctgg cctcctgccg ccgcattcca cccagcccc tcacaccact cgggagaggg 3660
 acatcctacg gtcccaaggt caggagggca gggctggggT tgactcaggc ccctcccagc 3720
 tgtggccacc tgggtgttgg gagggcagaa gtgcaggcac ctagggcccc ccatgtgccc 3780
 accctgggag ctctccttgg aaccattcc tgaattatt taaaggggtt ggccggacta 3840
 gttacgtaga taagtagcat ggcgggttaa tcattaacta caaggaaccc ctagtgatgg 3900
 agttggccac tccctctctg cgctctctg cgctcactga ggccgggcga ccaaaggctc 3960
 cccgacgccc gggctttgcc cgggcggcct cagtgagcga gcgagcgcg cagctggcgt 4020
 aatagcgaag agggccgcac cgatcgccct tccaacagt tgcgcagcct gaatggcgaa 4080
 tggcgattcc gttgcaatgg ctggcggtaa tattgttctg gatattacca gcaaggccga 4140
 tagtttgagt tcttctactc aggcaagtga tgttattact aatcaaagaa gtattgcgac 4200
 aacggttaat ttgcgtgatg gacagactct ttactcgtt ggcctcactg attataaaaa 4260
 cacttctcag gattctggcg taccgttcct gtctaaaatc cttttaatcg gcctcctgtt 4320
 tagctcccgC tctgattcta acgaggaaaG cacgttatac gtgctcgtca aagcaacat 4380

```

agtacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 4440
ccgctacact tgccagcgcc ctacgccccg ctcttttcgc tttcttcctt tcctttctcg 4500
ccacgttcgc cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat 4560
ttagtgcttt acggcacctc gacccccaaa aacttgatta gggatgatgt tcacgtagt 4620
ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata 4680
gtggactctt gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt 4740
tataagggat tttgccgatt tcggcctatt ggtaaaaaa tgagctgatt taacaaaaat 4800
ttaacgcgaa ttttaacaaa atattaacgt ttacaattta aatatttgct tatacaatct 4860
tcctgttttt ggggcttttc tgattatcaa ccgggggtaca tatgattgac atgctagt 4920
tacgattacc gttcatcgat tctcttggtt gctccagact ctcaggcaat gacctgatag 4980
cctttgtaga gacctctcaa aaatagctac cctctccggc atgaatttat cagctagaac 5040
ggttgaatat catattgatg gtgatttgac tgtctccggc ctttctcacc cgtttgaatc 5100
tttacctaca cattactcag gcattgcatt taaaatata gagggttcta aaaattttta 5160
tccttgcggt gaaataaagg cttctccgcg aaaagtatta cagggtcata atgtttttgg 5220
tacaaccgat ttagctttat gctctgagcg tttattgctt aattttgcta attctttgcc 5280
ttgcctgtat gatttattgg atgttggaat cgcctgatgc ggtattttct ccttacgcat 5340
ctgtgcggta tttcacaccg catatgggtg actctcagta caatctgctc tgatgccgca 5400
tagttaagcc agccccgaca cccgccaaca cccgctgacg cgcctgacg ggcttgtctg 5460
ctcccgcat ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg 5520
ttttcacctg catcacgaa acgcgcgaga cgaaaggcc tcgtgatacg cctattttta 5580
taggttaatg tcatgataat aatggtttct tagacgtcag gtggcacttt tcggggaaat 5640
gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg 5700
agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa 5760
catttccgtg tcgcccttat tccctttttt gcggcatttt gccttcctgt ttttgctcac 5820
ccagaaacgc tgggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 5880
atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt 5940
ccaatgatga gcacttttaa agttctgcta tgtggcgcggt tattatcccg tattgacgcc 6000
gggcaagagc aactcggctc cgcatacac tattctcaga atgacttggt tgagtactca 6060
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtctgc 6120
ataaccatga gtgataaac tgccggccaac ttacttctga caacgatcgg aggaccgaag 6180
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctcgccttga tcgttgggaa 6240
ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg 6300
gcaacaacgt tgcgcaaact attaaactggc gaactactta ctctagcttc ccggcaacaa 6360
ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttcgc 6420
gctggctggg ttattgctga taaatctgga gccggtgagc gtgggtctcg cggtatcatt 6480
gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 6540
caggcaacta tggatgaacg aaatagacag atcgtgaga taggtgcctc actgattaag 6600
cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat 6660
ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct 6720

```

taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 6780
 tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca 6840
 gcggtgggtt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt aactggcttc 6900
 agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 6960
 aagaactctg tagcaccgcc tacatactc gctctgctaa tcctgttacc agtggctgct 7020
 gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag 7080
 gcgcagcggc cgggctgaac ggggggttcg tgcacacagc ccagcttgga gcgaacgacc 7140
 tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct tcccgaaggg 7200
 agaaaaggcgg acaggtatcc ggtaagcggc agggtcggaa caggagagcg cagcaggagg 7260
 cttccagggg gaaacgcctg gtatctttat agtctgtcg ggtttcgcca cctctgactt 7320
 gagcgtcgat ttttgtgatg ctctgcaggg gggcgaggcc tatggaaaaa cgccagcaac 7380
 gcggcctttt tacggttcct ggcttttgc tggccttttg ctcacatgtt ctttctgctg 7440
 ttatccccctg attctgtgga taaccgtatt accgccttg agtgagctga taccgctcgc 7500
 cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga gcgccaata 7560
 cgcaaaccgc ctctccccgc gcgttgccg attcattaat g 7601

<210> 45

<211> 1325

<212> DNA

<213> 智人(Homo sapiens)

<400> 45

agtccagggc tctgtcctgc acctggggaa tggtagccgg catctctgtc ctctagctct 60
 ggaagcacc cagccccctt agtctgccct caccctgac cctgaccctc caccctgacc 120
 ccgtcctaac cctgacatt tgtgcccttc cagagagaag ggcagaagtg cccacagccc 180
 accccagccc ctcaccagg ccagccggcc agttccaaac cctgggtggtt ggtgtcgtgg 240
 gcggcctgct gggcagcctg gtgctgctag tctgggtcct ggccgtcatc tgctcccggg 300
 ccgcacgagg taacgtcatc ccagccctc ggctgccct gccctaacc tgctggcggc 360
 cctcaactcc gcctccccct cctccaccct tccctaccc caccacact ccccccatt 420
 ccccgccagg ctaagtccct gatgaaggcc cctggactaa gacccccac ctaggagcac 480
 ggctcagggt cggcctggtg accccaagtg tgtttctctg cagggacaat aggagccagg 540
 cgcaccggcc agccccgtgt gactctcact ctttctctgc atgatccact gtgccttctt 600
 tcctgggtgg gcagaggtgg aaggacaggc tgggaccaca cggcctgcag gactcacatt 660
 ctattatagc caggacccca cctccccagc cccaggcag caacctcaat ccctaaagcc 720
 atgatctggg gccccagccc acctgcggtc tccgggggtg cccggcccat gtgtgtgcct 780
 gcctgcggtc tccaggggtg cctggccac gcgtgtgccc gcctgcggtc tctgggggtg 840
 cccggccac atatgtgcct gcctgcggtc tccaggtgtg cccggcccat gcgtgtgccc 900
 acctgcgagg gcgtgggggt ggcttggtca tttcttatct tacattggag acaggagagc 960
 ttgaaaagtc acattttgga atcctaaatc tgcaagaatg ccaggacat ttcagagggg 1020
 gacattgagc cagagaggag ggggtggtgc cccagatcac acagagggca gtggtgggac 1080
 agctcagggt aagcagctca tagtgggggg cccaggttcg gtgccgttac tgcagccagg 1140

ctgtggagcc gcgggcctcc ttcctgcggt gggccgtggg gctgactccc tctccctttc 1200
 tcctcaaaga aggaggaccc ctcagccgtg cctgtgttct ctgtggacta tggggagctg 1260
 gatttccagt ggcgagagaa gaccccgag ccccccgtgc cctgtgtccc tgagcagacg 1320
 gagta 1325

<210> 46

<211> 1072

<212> DNA

<213> 智人(Homo sapiens)

<400> 46

ccaccattgt ctttcctagc ggaatgggca cctcatcccc cgcccgcagg ggctcagctg 60
 acggccctcg gagtgcacag ccaactgaggc ctgaggatgg aactgtctct tggccctctt 120
 gaccggcttc cttggccacc agtggttctgc agacctcca ccatgagccc gggtcagcgc 180
 atttcctcag gagaagcagg cagggtgcag gccattgcag gccgtccagg ggctgagctg 240
 cctggggggcg accgggggtc cagcctgcac ctgcaccagg cacagcccca ccacaggact 300
 catgtctcaa tgcccacagt gagcccaggc agcaggtgtc accgtcccct acagggaggg 360
 ccagatgcag tcaactgttc aggtcctgcc agcacagagc tgccctgcgtc cagctccctg 420
 aatctctgct gctgctgctg ctgctgctgc tgctgcctgc ggcccggggc tgaaggcgcc 480
 gtggccctgc ctgacgcccc ggagcctcct gcctgaactt gggggctggt tggagatggc 540
 cttggagcag ccaaggtgcc cctggcagtg gcatcccga acgccctgga cgcaggggcc 600
 aagactgggc acaggagtgg gaggtacatg gggctgggga ctccccagga gttatctgct 660
 ccctgcaggc ctagagaagt ttcagggaag gtcagaagag ctccctggctg tgggtgggcag 720
 ggcaggaaac ccctccacct ttacacatgc ccaggcagca cctcaggccc tttgtggggc 780
 aggggaagctg aggcagtaag cgggcaggca gagctggagg cctttcaggc ccagccagca 840
 ctctggcctc ctgccgccgc attccacccc agcccctcac accactcggg agagggacat 900
 cctacggtcc caaggtcagg agggcagggc tggggttgac tcaggcccct ccagctgtg 960
 gccacctggg tgttgggagg gcagaagtgc aggcacctag ggcccccat gtgccaccc 1020
 tgggagctct ccttgaacc cattcctgaa attatttaa ggggttggcc gg 1072

<210> 47

<211> 7601

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子5_MND-BFP 3'-同源臂SNP构建体

<400> 47

cagctgcgcg ctcgctcgt cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc 60
 tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
 actaggggtt cctttagatt aatgattaac ccgccatgct acttatctac gtagaattca 180
 gtccagggct ctgtcctgca cctggggaat ggtgaccggc atctctgtcc tctagctctg 240

```

gaagcacccc agccccctcta gtctgccctc acccctgacc ctgaccctcc accctgaccc 300
cgctcctaacc cctgaccttt gtgcccttcc agagagaagg gcagaagtgc ccacagccca 360
ccccagcccc tcacccaggc cagccggcca gttccaaacc ctggtggttg gtgtcgtggg 420
cggcctgctg ggcagcctgg tgctgctagt ctgggtcctg gccgtcatct gctcccgggc 480
cgcacgaggt aacgtcatcc cagccccctg gcctgccctg ccctaaccct gctggcggcc 540
ctcactcccc cctccccctt ctcacacctt ccctacccc accccacctc ccccatctc 600
cccgccaggc taagtccctg atgaaggccc ctggactaag accccccacc taggagcacg 660
gctcagggtc ggcttggtga cccaagtgt gtttctctgc agggacaata ggagccaggc 720
gcaccggcca gcccctggtg agtctcactc ttttctgca tgatccactg tgccttcctt 780
cctgggtggg cagaggtgga aggacaggct gggaccacac ggctgcagg actcacattc 840
tattatagcc aggacccac ctcaccagcc ccaggcagc aacctcaatc cctaaagcca 900
tgatctgggg cccagccca cctgcggtct ccgggggtgc ccggcccatg tgtgtgcctg 960
cctgcggtct ccagggtgc ctggcccacg cgtgtgccc cctgcggtct ctgggggtgc 1020
ccggcccaca tatgtgcctg cctgcggtct ccagggtgc ccggcccatg cgtgtgccc 1080
cctgcgaggg cgtgggtgg gcttggtcat ttcttatctt acattggaga caggagagct 1140
tgaaaagtca cattttgaa tcctaaatct gcaagaatgc caggacatt tcagaggggg 1200
acattgagcc agagaggagg ggtggtgtcc ccagatcaca cagagggcag tggtgggaca 1260
gctcagggtg agcagctcat agtggggggc ccaggttcgg tgccgtact gcagccaggc 1320
tgtggagccg cgggcctcct tcctgcggtg ggccgtgggg ctgactccct ctccctttct 1380
cctcaaagaa ggaggacccc tcagccgtgc ctgtgttctc tgtggactat ggggagctgg 1440
atttcagtg gcgagagaag accccggagc ccccggtgc ctgtgtccct gagcagacgg 1500
agtaatgcat gaacagagaa acaggagaat atgggccaaa caggatatct gtggtaaagca 1560
gttcctgccc cggctcaggg ccaagaacag ttggaacagc agaatatggg ccaaacagga 1620
tatctgtggt aagcagttcc tgccccggct cagggccaaag aacagatggt cccagatgc 1680
ggctccgccc tcagcagttt ctagagaacc atcagatgtt tccagggtgc cccaaggacc 1740
tgaaatgacc ctgtgcctta tttgaactaa ccaatcagtt cgcttctcgc ttctgttcgc 1800
gcgcttctgc tccccgagct ctatataagc agagctcggt tagtgaaccg tcagatcgcc 1860
tgagacgcc atccacgctg ttttgacttc catagaagga tctcaggcc accatggcta 1920
gcgagctgat taaggagaac atgcacatga agctgtacat ggagggcacc gtggacaacc 1980
atcacttcaa gtgcacatcc gagggcgaag gcaagcccta cgagggcacc cagaccatga 2040
gaatcaaggt ggtcgagggc ggccctctcc cttcgcctt cgacatcctg gctactagct 2100
tcctctacgg cagcaagacc ttcatcaacc acaccaggg catccccgac ttcttcaagc 2160
agtccttccc tgagggttc acatgggaga gagtcaccac atacaggac gggggcgtgc 2220
tgaccgtac ccaggacacc agcctccagg acggtgcct catctacaac gtcaagatca 2280
gaggggtgaa cttcacatcc aacggccctg tgatgcagaa gaaaacactc ggctgggagg 2340
ccttaccga gacgtgtac cccgtgacg gcggcctgga aggcagaaac gacatggccc 2400
tgaagctcgt gggcgggagc catctgatcg caaacatcaa gaccacatat agatccaaga 2460
aaccgctaa gaacctcaag atgcctggcg tctactatgt ggactacaga ctggaaagaa 2520
tcaaggaggc caacaacgag acctacgtcg agcagcacga ggtggcagtg gccagatact 2580

```

gcgacctccc tagcaaactg gggcacaagc taaattgaaa gcttttgcttt atttgtgaaa 2640
tttgtgatgc tattgcttta tttgtaacca ttataagctg caataaaciaa gtttaacaac 2700
aacaattgca ttcatTTTTat gtttcagggt cagggggagg tgtgggaggt tttttaaagt 2760
cgacccacca ttgtctttcc tagcggaatg ggcacctcat ccccgccccg caggggctca 2820
gccgacggcc ctccgagtg cccagccactg aggcctgagg atggacactg ctcttgcccc 2880
ctctgaccgg ctcccttggc caccagtgtt ctgcagacct tccacctga gcccggtca 2940
gcgcattttcc tcaggagaag caggcagggt gcaggccatt gcaggccgtc caggggctga 3000
gctgcctggg ggcgaccggg gctccagcct gcacctgcac caggcacagc cccaccacag 3060
gactcatgtc tcaatgcccc cagtgcagcc aggcagcagg tgtcacctgc ccctacaggg 3120
agggccagat gcagtcactg ctccaggtcc tgccagcaca gagctgcctg cgtccagctc 3180
cctgaatctc tgctgtctgt gctgtctgtg ctgtctgtgc ctgcggcccc gggctgaagg 3240
cgccgtggcc ctgcctgacg ccccgaggcc tcttgctga acttgggggc tgggttgaga 3300
tgcccttgga gcagccaagg tgccctggc agtggcatcc cgaaacgccc tggacgcagg 3360
gccaagact gggcacagga gtgggaggta catggggctg gggactcccc aggagttatc 3420
tgctccctgc aggcctagag aagtttcagg gaaggtcaga agagctctg gctgtggtgg 3480
gcagggcagg aaacccctcc acctttacac atgcccaggc agcacctcag gccctttgtg 3540
gggcagggaa gctgaggcag taagcgggca ggcagagctg gaggcctttc aggccagcc 3600
agcactctgg cctcctgccg ccgcattcca cccagcccc tcacaccact cgggagaggg 3660
acatcctacg gtcccaaggc caggagggca gggctggggt tgactcaggc ccctcccagc 3720
tgtggccacc tgggtgttgg gagggcagaa gtgcaggcac ctagggcccc ccatgtgccc 3780
accctgggag ctctccttgg aaccattcc tgaaattatt taaagggtt ggccggacta 3840
gttacgtaga taagtagcat ggcgggttaa tcattaacta caaggaacct ctagtgatgg 3900
agttggccac tccctctctg cgcgtctgct cgctcactga ggccgggcca ccaaaggctc 3960
cccagcggcc gggctttgcc cgggcggcct cagtgcagca gcgagcgcg cagctggcgt 4020
aatagcgaag agggccgcac cgatcgccct tccaacagt tgccgagcct gaatggcgaa 4080
tgccgattcc gttgcaatgg ctggcggtaa tattgttctg gatattacca gcaaggccga 4140
tagtttgagt tcttctactc aggcaagtga tgttattact aatcaaagaa gtattgcgac 4200
aacggttaat ttgcgtgatg gacagactct ttactcggg gccctcactg attataaaaa 4260
cacttctcag gattctggcg taccgttccg gtctaaaatc cctttaatcg gcctcctgtt 4320
tagctcccg cctgattcta acgaggaaag caggttatac gtgctcgtca aagcaacct 4380
agtacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 4440
ccgctacact tgccagcgcc ctacgccccg ctcttttcgc tttcttccct tcttttctcg 4500
ccacgttcgc cggctttccc cgtcaagctc taaatcgagg gctcccttta gggttccgat 4560
ttagtgttt acggcacctc gaccccaaaa aacttgatta ggggtgatgg tccagtagtg 4620
ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttttttaata 4680
gtggactctt gttccaaaact ggaacaacac tcaacctat ctcggtctat tcttttgatt 4740
tataagggat tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat 4800
ttaacgcgaa ttttaacaaa atattaacgt ttacaattta aatatttgct tatacaatct 4860
tctgtttttt ggggcttttc tgattatcaa ccggggtaca tatgattgac atgctagttt 4920

```

tacgattacc gttcatcgat tctcttgttt gctccagact ctcaggcaat gacctgatag 4980
cctttgtaga gacctctcaa aaatagctac cctctccggc atgaatttat cagctagaac 5040
ggttgaatat catattgatg gtgatttgac tgtctccggc ctttctcacc cgtttgaatc 5100
tttacctaca cattactcag gcattgcatt taaaatatat gagggttcta aaaattttta 5160
tccttgcgtt gaaataaagg cttctcccg cccaggtatta cagggtcata atgttttttg 5220
tacaaccgat ttagctttat gctctgaggc tttattgctt aattttgcta attccttgcc 5280
ttgcctgtat gatttattgg atgttggaat cgcctgatgc ggtattttct ccttacgcat 5340
ctgtgcggta tttcacaccg catatggtgc actctcagta caatctgctc tgatgccgca 5400
tagttaagcc agccccgaca cccgccaca cccgtgacg cgcctgacg ggcttgtctg 5460
ctccccgcat ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg 5520
ttttcaccgt catcaccgaa acgcgcgaga cgaaagggcc tcgtgatacg cctattttta 5580
taggttaatg tcatgataat aatggtttct tagacgtcag gtggcacttt tcggggaaat 5640
gtgcgcggaa cccctatttg tttattttcc taaatacatt caaatatgta tccgctcatg 5700
agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa 5760
catttccgtg tcgcccttat tccctttttt gcggcatttt gccttctctg ttttgctcac 5820
ccagaaacgc tggtgaaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 5880
atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt 5940
ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc 6000
gggcaagagc aactcggctg ccgcatacac tattctcaga atgacttggt tgagtactca 6060
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtctgcc 6120
ataacatga gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaag 6180
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga tcgttgggaa 6240
ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg 6300
gcaacaacgt tgcgcaaact attaaactggc gaactactta ctctagcttc ccggaacaa 6360
ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg 6420
gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg cggtatcatt 6480
gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 6540
caggcaacta tggatgaacg aaatagacag atcgtgaga taggtgcctc actgattaag 6600
cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat 6660
ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct 6720
taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 6780
tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca 6840
gcggtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaagg t aactggcttc 6900
agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 6960
aagaactctg tagcaccgcc tacatactc gctctgctaa tcctgttacc agtggctgct 7020
gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag 7080
gcgcagcggc cgggctgaac ggggggttcg tgcacacagc ccagcttga gcgaacgacc 7140
tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct tcccgaagg 7200
agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa caggagagcg cacgaggag 7260

```

cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt 7320
 gagcgtcgat ttttgtgatg ctctgcaggg gggcggagcc tatggaaaaa cgccagcaac 7380
 gcggcctttt tacggttcct ggccttttgc tggccttttg ctacatgtt ctttcctgcg 7440
 ttatccccctg attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc 7500
 cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga gcgccaata 7560
 cgcaaaccgc ctctccccgc gcgttggccg attcattaat g 7601

<210> 48

<211> 7601

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子5_MND-BFP 5'-同源臂SNP构建体

<400> 48

cagctgcgcg ctgctcgt cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc 60
 tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
 actaggggtt ccttgtagt aatgattaac ccgccatgct acttatctac gtagaattca 180
 gtccagggct ctgtcctgca cctggggaat ggtgaccggc atctctgtcc tctagctctg 240
 gaagcacccc agccccctta gtctgccctc acccctgacc ctgaccctcc accctgacct 300
 cgtcctaacc cctgaccttt gtgcccttcc agagagaagg gcagaagtgc ccacagccca 360
 cccagacccc tcaccagggc cagccggcca gttccaaacc ctggtggttg gtgtcgtggg 420
 cggcctgctg ggcagcctgg tgctgctagt ctgggtcctg gccgtcatct gctcccgggc 480
 cgcacgaggt aacgtcatcc cagccccctg gcctgccctg ccctaaccct gctggcggcc 540
 ctactccccg cctccccctc ctccaccctt cctcacccc accccacctc ccccatctc 600
 cccgccaggc taagtccctg atgaaggccc ctggactaag accccccacc taggagcacg 660
 gctcagggtc ggccctggtg cccaagtgt gtttctctgc agggacaata ggagccaggc 720
 gcaccggcca gccctggtg agtctcactc ttttctgca tgatccactg tgccttcctt 780
 cctgggtggg cagaggtgga aggacaggct gggaccacac ggccctgcagg actcacattc 840
 tattatagcc aggacccccc ctccccagcc cccaggcagc aacctcaatc cctaaagcca 900
 tgatctgggg cccagcccca cctgcggtct ccgggggtgc ccggcccatg tgtgtgcctg 960
 cctgcggtct ccaggggtgc ctggcccacg cgtgtgcccg cctgcggtct ctgggggtgc 1020
 ccggcccaca tatgtgcctg cctgcggtct ccagggtgtgc ccggcccatg cgtgtgcccc 1080
 cctgcgaggg cgtgggtgg gcttggtcat ttcttatctt acattggaga caggagagct 1140
 tgaagagtca cattttgga tcctaaatct gcaagaatgc caggacatt tcagaggggg 1200
 acattgagcc agagaggagg ggtggtgtcc ccagatcaca cagagggcag tgggtgggaca 1260
 gctcagggtg agcagctcgt agtggggggc ccaggttcgg tgccggtact gcagccaggc 1320
 tgtggagccg cgggcctcct tcctgcggtg ggccgtgggg ctgactccct ctccccttct 1380
 cctcaaagaa ggaggacccc tcagccgtgc ctgtgttctc tgtggactat ggggagctgg 1440
 atttccagtg gcgagagaag accccggagc ccccggtgcc ctgtgtccct gagcagacgg 1500

```

agtaatgcat gaacagagaa acaggagaat atgggccaaa caggatatct gtggttaagca 1560
gttcctgccc cggctcaggg ccaagaacag ttggaacagc agaatatggg ccaaacagga 1620
tatctgtggt aagcagttcc tgccccggct cagggccaag aacagatggt cccagatgc 1680
gggtcccgccc tcagcagttt ctagagaacc atcagatggt tccagggtgc cccaaggacc 1740
tgaaatgacc ctgtgcctta tttgaactaa ccaatcagtt cgcttctcgc ttctgttcgc 1800
gcgcttctgc tccccgagct ctatataagc agagctcggt tagtgaaccg tcagatcgcc 1860
tgagagacgcc atccacgctg ttttgacttc catagaagga tctcgaggcc accatggcta 1920
gcgagctgat taaggagaac atgcacatga agctgtacat ggagggcacc gtggacaacc 1980
atcacttcaa gtgcacatcc gagggcgaag gcaagcccta cgagggcacc cagaccatga 2040
gaatcaaggt ggtcgagggc ggccctctcc ctttcgctt cgacatcctg gctactagct 2100
tcctctacgg cagcaagacc ttcatcaacc acaccaggg catccccgac ttcttcaagc 2160
agtccttccc tgagggttc acatgggaga gagtcaccac atacaggac gggggcgctgc 2220
tgaccgtac ccaggacacc agcctccagg acggtgcct catctacaac gtcaagatca 2280
gaggggtgaa cttcacatcc aacggccctg tgatgcagaa gaaaacactc ggctgggagg 2340
ccttcaccga gacgtgtac ccgctgacg gcggcctgga aggcagaaac gacatggccc 2400
tgaagctcgt gggcgggagc catctgatcg caaacatcaa gaccacatat agatccaaga 2460
aaccgcgtaa gaacctcaag atgcctggcg tctactatgt ggactacaga ctggaaagaa 2520
tcaaggaggc caacaacgag acctacgtcg agcagcacga ggtggcagtg gccagatact 2580
gcgacctccc tagcaaactg gggcacaagc taaattgaaa gctttgcttt atttgtgaaa 2640
tttgtgatgc tattgcttta tttgtaacca ttataagctg caataaacia gttaacaac 2700
aacaattgca ttcattttat gtttcaggtt cagggggagg tgtgggaggt tttttaaggt 2760
cgaccaccca ttgtctttcc tagcggaatg ggcacctcat ccccgcccg caggggctca 2820
gctgacggcc ctcgagtgcc ccagccactg aggcctgagg atggacactg ctcttgcccc 2880
ctctgaccgg cttccttgcc caccagtgtt ctgcagacct tccaccatga gcccggtca 2940
gcgcatttcc tcaggagaag caggcagggt gcaggccatt gcaggccgtc caggggctga 3000
gctgcctggg ggcgaccggg gctccagcct gcacctgcac caggcacagc cccaccacag 3060
gactcatgtc tcaatgcccc cagttagccc aggcagcagg tgtcacctgc ccctacaggg 3120
agggccagat gcagtcactg cttcaggtcc tgccagcaca gagctgcctg cgtccagctc 3180
cctgaatctc tgctgtgtgt gctgtgtgtc ctgtgtgtgc ctgcggcccc gggctgaagg 3240
cgccgtggcc ctgcctgacg ccccgagacc tctgcctga acttgggggc tggttggaga 3300
tggccttgga gcagccaagg tgccctggc agtggcatcc cgaaacgccc tggacgcagg 3360
gccaagact gggcacagga gtgggaggta catggggctg gggactcccc aggagttatc 3420
tgctccctgc aggcctagag aagtttcagg gaaggtcaga agagctcctg gctgtggtgg 3480
gcagggcagg aaaccctcc acctttacac atgccaggc agcacctcag gccctttgtg 3540
gggcagggaa gctgaggcag taagcgggca ggcagagctg gaggcctttc agggccagcc 3600
agcactctgg cctcctgccg ccgcattcca cccagcccc tcacaccact cgggagaggg 3660
acatcctacg gtccaaggt caggagggca gggctgggtg tgactcaggc ccctcccagc 3720
tgtggccacc tgggtgttgg gagggcagaa gtgcaggcac ctagggcccc ccatgtgccc 3780
accctgggag ctctccttgg aaccattcc tgaaattatt taaagggtt ggccggacta 3840

```

gttacgtaga taagtagcat ggcggttaa tcattaacta caaggaaccc ctagtgatgg 3900
agttggccac tccctctctg cgcgctcgct cgctcactga ggccgggcca ccaaggtcg 3960
cccgcgccc gggctttgcc cggcgggcct cagtgcgca gcgagcgcgc cagctggcgt 4020
aatagcgaag aggcccgac cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa 4080
tggcgattcc gttgcaatgg ctggcggtta tattgttctg gatattacca gcaaggccga 4140
tagtttgagt tcttctactc aggcaagtga tgttattact aatcaaagaa gtattgcgac 4200
aacggttaat ttgcgtgatg gacagactct tttactcggt ggcctcactg attataaaaa 4260
cacttctcag gattctggcg taccgttcct gtctaaaatc cctttaatcg gcctcctgtt 4320
tagctcccg cctgattcta acgaggaaag caggttatac gtgctcgtca aagcaaccat 4380
agtacgcgcc ctgtagcggc gcattaagcg cggcggtgt ggtggttacg cgcagcgtga 4440
ccgctacact tgccagcgcc ctagcgcccg ctcttttcgc tttcttcctc tcttttctcg 4500
ccacgttcgc cggtttccc cgtcaagctc taaatcgagg gctcccttta gggttccgat 4560
ttagtgcttt acggcacctc gacccccaaa aacttgatta gggatgatgt tcacgtagt 4620
ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttttttaata 4680
gtggactctt gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt 4740
tataagggat tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat 4800
ttaacgcgaa ttttaacaaa atattaacgt ttacaattta aatatttgct tatacaatct 4860
tctgtttttt ggggcttttc tgattatcaa ccgggttaca tatgattgac atgctagttt 4920
tacgattacc gttcatcgat tctcttgttt gctccagact ctcaggcaat gacctgatag 4980
cctttgtaga gacctctcaa aaatagctac cctctccggc atgaatttat cagctagaac 5040
ggttgaatat catattgatg gtgatttgac tgtctccggc ctttctcacc cgtttgaatc 5100
tttacctaca cattactcag gcattgcatt taaaatatat gagggttcta aaaattttta 5160
tccttgcggt gaaataaagg cttctccgcg aaaagtatta cagggtcata atgttttttg 5220
tacaaccgat ttagctttat gctctgaggc tttattgctt aattttgcta attctttgcc 5280
ttgcctgtat gatttattgg atgttggaat cgctgatgc ggtattttct cttacgcatt 5340
ctgtgcggta tttcacaccg catatggtgc actctcagta caatctgctc tgatgccgca 5400
tagttaagcc agccccgaca cccgccaaca cccgtgacg cgccctgacg ggcttgtctg 5460
ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg 5520
ttttcaccgt catcaccgaa acgcgcgaga cgaaagggcc tcgtgatacg cctattttta 5580
taggttaatg tcatgataat aatggtttct tagacgtcag gtggcacttt tcggggaaat 5640
gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg 5700
agacaataac cctgataaat gttcaataa tattgaaaaa ggaagagtat gagtattcaa 5760
catttccgtg tcgcccttat tccctttttt gcggcatttt gccttctgt ttttgctcac 5820
ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 5880
atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaagttttt 5940
ccaatgatga gcacttttaa agttctgcta tgtggcgcggt tattatcccg tattgacgcc 6000
gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt tgagtactca 6060
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtgcc 6120
ataacatga gtgataacac tgcgccaac ttacttctga caacgatcgg aggaccgaag 6180

gagctaaccg cttttttgca caacatgggg gatcatgtaa ctcgccttga tcgttgggaa 6240
 ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg 6300
 gcaacaacgt tgcgcaaact attaaactggc gaactactta ctctagcttc ccggcaacaa 6360
 ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg 6420
 gctggctggg ttattgctga taaatctgga gccggtgagc gtgggtctcg cggtatcatt 6480
 gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 6540
 caggcaacta tggatgaacg aaatagacag atcgtgaga taggtgcctc actgattaag 6600
 cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat 6660
 ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct 6720
 taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 6780
 tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca 6840
 gcggtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaagg t aactggcttc 6900
 agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 6960
 aagaactctg tagcaccgcc tacatactc gctctgctaa tcctgttacc agtggctgct 7020
 gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag 7080
 gcgcagcggc cgggctgaac ggggggttcg tgcacacagc ccagcttga gcgaacgacc 7140
 tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct tcccgaagg 7200
 agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa caggagagcg cacgaggag 7260
 cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt 7320
 gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa cgccagcaac 7380
 gcggcctttt tacggttcct ggccttttgc tggccttttg ctcacatgtt ctttctgctg 7440
 ttatccccctg attctgtgga taaccgtatt accgccttg agtgagctga taccgctcgc 7500
 cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga gcgccaata 7560
 cgcaaaccgc ctctccccgc gcgttggccg attcattaat g 7601

<210> 49

<211> 1325

<212> DNA

<213> 智人(Homo sapiens)

<400> 49

agtccagggc tctgtcctgc acctggggaa tggtagaccg catctctgtc ctctagctct 60
 ggaagcacc cagccccctct agtctgcct caccctgac cctgaccctc caccctgacc 120
 ccgtcctaac ccctgacctt tgtgcccttc cagagagaag ggcagaagtg cccacagccc 180
 accccagccc ctaccccagg ccagccggcc agttccaaac cctgggtggtt ggtgtcgtgg 240
 gcggcctgct gggcagcctg gtgtgctag tctgggtcct ggccgtcatc tgctcccggg 300
 ccgcacgagg taacgtcatc ccagccctc ggctgacct gccctaacc tgctggcggc 360
 cctcactccc gcctccccct cctccacct tccctaccc caccctacct ccccatct 420
 ccccgccagg ctaagtccct gatgaaggcc cctggactaa gacccccac ctaggagcac 480
 ggctcagggt cggcctgggt accccaagt gtgttctctg cagggacaat aggagccagg 540
 cgcaccggcc agccctgggt gagtctcact ctttctctgc atgatccact gtgccttct 600

tcctgggtgg gcagaggtgg aaggacaggc tgggaccaca cggcctgcag gactcacatt 660
 ctattatagc caggacccca cctccccagc cccaggcag caacctcaat ccctaaagcc 720
 atgatctggg gccccagccc acctgcggtc tccgggggtg cccggcccat gtgtgtgcct 780
 gcctgcggtc tccaggggtg cctggcccac gcgtgtgccc gcctgcggtc tctgggggtg 840
 cccggcccac atatgtgcct gcctgcggtc tccaggtgtg cccggcccat gcgtgtgccc 900
 acctgcgagg gcgtgggggtg ggcttggtca tttcttatct tacattggag acaggagagc 960
 ttgaaaagtc acatttttga atcctaaatc tgcaagaatg ccagggacat ttcagagggg 1020
 gacattgagc cagagaggag gggtggtgtc cccagatcac acagagggca gtggtgggac 1080
 agctcagggt aagcagctcg tagtgggggg cccaggttcg gtgccgttac tgcagccagg 1140
 ctgtggagcc gcgggcctcc ttctgcggt gggccgtggg gctgactccc tctccctttc 1200
 tcctcaaaga aggaggaccc ctccagccgtg cctgtgttct ctgtggacta tggggagctg 1260
 gatttccagt ggcgagagaa gaccccgagg ccccccgtgc cctgtgtccc tgagcagacg 1320
 gagta 1325

<210> 50

<211> 1072

<212> DNA

<213> 智人(Homo sapiens)

<400> 50

ccaccattgt ctttcctagc ggaatgggca cctcatcccc cgcccgcagg ggctcagccg 60
 acggccctcg gagtggccag ccactgaggc ctgaggatgg aactgtctct tggccctctt 120
 gaccggcttc ctggccacc agtgttctgc agacctcca ccatgagccc gggtcagcgc 180
 atttcctcag gagaagcagg cagggtgcag gccattgcag gccgtccagg ggctgagctg 240
 cctgggggcg accgggggctc cagcctgcac ctgcaccagg cacagcccca ccacaggact 300
 catgtctcaa tgcccacagt gagcccaggc agcaggtgtc accgtcccct acagggaggg 360
 ccagatgcag tcaactgctc aggtcctgcc agcacagagc tgccctgcgtc cagctccctg 420
 aatctctgct gctgctgctg ctgctgctgc tgctgcctgc ggcccggggc tgaaggcgcc 480
 gtggccctgc ctgacgcccc ggagcctcct gcctgaactt gggggctggt tggagatggc 540
 cttggagcag ccaaggtgcc cctggcagtg gcatcccga acgccctgga cgcaggggcc 600
 aagactgggc acaggagtgg gaggtacatg gggctgggga ctccccagga gttatctgct 660
 ccctgcaggc ctagagaagt ttcagggaag gtcagaagag ctctggctg tgggtggcag 720
 ggcaggaaac ccctccacct ttacacatgc ccaggcagca cctcaggccc tttgtggggc 780
 agggaagctg aggcagtaag cgggcaggca gagctggagg cctttcaggc ccagccagca 840
 ctctggcctc ctgccgccgc attccacccc agcccctcac accactcggg agagggacat 900
 cctacgttcc caaggtcagg agggcagggc tggggttgac tcaggccctt cccagctgtg 960
 gccacctggg tgttgggagg gcagaagtgc aggcacctag ggccccccat gtgcccaccc 1020
 tgggagctct ccttgaacc cattcctgaa attattttaa ggggttggcc gg 1072

<210> 51

<211> 7880

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子5_MND-PD-1-CD28转换受体构建体

<400> 51

```

cagctgcgcg ctcgctcgct cactgaggcc gcccgggcaa agccccgggcg tcgggcgacc 60
tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
actaggggtt ccttgtagtt aatgattaac ccgccatgct acttatctac gtagaattca 180
gtccagggct ctgtcctgca cctgggggaat ggtgaccggc atctctgtcc tctagctctg 240
gaagcacccc agccccctta gtctgcctc acccctgacc ctgaccctcc accctgacct 300
cgtcctaacc cctgaccttt gtgcccttcc agagagaagg gcagaagtgc ccacagccca 360
ccccagcccc tcacccaggc cagccggcca gttccaaacc ctggtggttg gtgtcgtggg 420
cggcctgctg ggcagcctgg tgetgctagt ctgggtcctg gccgtcatct gctcccgggc 480
cgcacgaggt aacgtcatcc cagccccctg gcctgcctg ccctaaccct gctggcgggc 540
ctcactcccc cctccccctt ctccacctt ccctacccc accccacctc cccccatctc 600
cccgccaggc taagtccctg atgaaggccc ctggactaag accccccacc taggagcacg 660
gctcagggtc ggctgggtga cccaagtgt gtttctctgc agggacaata ggagccaggc 720
gcaccggcca gcccctggtg agtctcactc ttttctgca tgatccactg tgccttcctt 780
cctgggtggg cagaggtgga aggacaggct gggaccacac ggcctgcagg actcacattc 840
tattatagcc aggacccac ctccccagcc cccaggcagc aacctcaatc cctaaagcca 900
tgatctgggg ccccagccca cctgcggtct ccgggggtgc ccggcccatg tgtgtgcctg 960
cctgcggtct ccaggggtgc ctggcccacg cgtgtgcccg cctgcggtct ctgggggtgc 1020
ccggcccaca tatgtgcctg cctgcggtct ccaggtgtgc ccggcccatg cgtgtgcccc 1080
cctgcgaggg cgtgggggtg gcttgggtcat ttcttatctt acattggaga caggagagct 1140
tgaaaagtca cattttgtaa tcctaaatct gcaagaatgc caggacatt tcagaggggg 1200
acattgagcc agagaggagg ggtggtgtcc ccagatcaca cagagggcag tgggtgggaca 1260
gctcagggta agcagctcat agtggggggc ccaggttcgg tgccggtact gcagccaggc 1320
tgtggagccg cgggcctcct tcctgcggtg ggccgtgggg ctgactccct ctccctttct 1380
cctcaaagaa ggaggacccc tcagccgtgc ctgtgttctc tgtggactat ggggagctgg 1440
atttccagtg gcgagagaag accccggagc ccccggtgcc ctgtgtccct gagcagacgg 1500
agtaatgcat gaacagagaa acaggagaat atgggccaaa caggatatct gtggtaaagca 1560
gttcctgccc cggctcaggg ccaagaacag ttggaacagc agaatatggg ccaaacagga 1620
tatctgtggt aagcagttcc tgccccggct cagggccaaag aacagatggt cccagatgc 1680
gggtcccgcc tcagcagttt ctagagaacc atcagatgtt tccagggtgc cccaaggacc 1740
tgaaatgacc ctgtgcctta tttgaactaa ccaatcagtt cgcttctcgc ttctgttcgc 1800
gcgcttctgc tccccgagct ctatataagc agagctcggt tagtgaaccg tcagatcgcc 1860
tgagagacgcc atccacgctg ttttgacttc catagaagga tctcagggcc accatgcaga 1920
tcccgcaagc gccctggcca gtcgtctggg cggtgctaca actgggctgg cggccaggat 1980
ggttcttaga ctccccagac aggccctgga accccccac cttctcccca gccctgctcg 2040

```

```

tggtgaccga aggggacaac gccaccttca cctgcagctt ctccaacaca tcggagagct 2100
tcgtgctaaa ctggtaccgc atgagcccca gcaaccagac ggacaagctg gccgccttcc 2160
ccgaggaccg cagccagccc ggccaggact gccgcttccg tgtcacacaa ctgcccacacg 2220
ggcgtgactt ccacatgagc gtggtcaggg cccggcgcaa tgacagcggc acctacctct 2280
gtggggccat ctccctggcc cccaaggcgc agatcaaaga gagcctgcgg gcagagctca 2340
gggtgacaga gagaaggga gaagtgccca cagcccaccc cagcccctca cccaggccag 2400
ccggccagtt ccaaaccctg gtggttggtg tcgtgggcgg cctgctgggc agcctggtgc 2460
tgctagtctg ggtcctggcc gtcacagga gtaagaggag caggctcctg cacagtgact 2520
acatgaacat gactccccgc cgccccgggc ccaccgcaa gcattaccag ccctatgccc 2580
caccacgcga cttcgcagcc tatcgtccg gtgaggcgag aggaagtctt ctaacatgcg 2640
gtgacgtgga ggagaatccg ggccctgtga gcaaggcgga ggaggataac tccgccatca 2700
tcaaggagtt cctgcgttc aaggtgcaca tggagggtc cgtgaacggc cagagttcg 2760
agatcgaggg cgaggcgag ggccgccct acgagggcac ccagaccgcc aagctgaagg 2820
tgaccaaggg tggccccctg cccttcgcct gggacatcct gtccccctag ttcatgtacg 2880
gctccaaggc ctacgtgaag caccgcccg acatccccga ctacttgaag ctgtccttcc 2940
ccgagggtt caagtgggag cgctgatga acttcagga cggcggcgtg gtgaccgtga 3000
cccaggactc ctctctgcag gacggcgagt tcatctacaa ggtgaagctg cgcgccacca 3060
acttccccctc cgacggcccc gtaatgcaga agaagaccat gggctgggag gcctcctccg 3120
agcggatgta ccccgaggac ggcgccctga agggcgagat caagcagagg ctgaagctga 3180
aggacggcgg ccactacgac gctgaggtca agaccaccta caaggccaag aagcccgtgc 3240
agctgcccgg cgctacaaac gtcaacatca agttggacat cacctccac aacaggact 3300
acaccatcgt ggaacagtac gaacgcgcc agggccgcca ctccaccggc ggcattggacg 3360
agctgtacaa gtgatgaaag ctttgcttta tttgtgaaat ttgtgatgct attgctttat 3420
ttgtaacat tataagctgc aataaacaag tttacaaca acaattgcat tcattttatg 3480
tttcaggttc agggggaggt gtgggaggtt tttaaagtc gacgtggccc tgcctgacgc 3540
cccggagcct cctgcctgaa cttgggggct ggttgagat ggccttgag cagccaaggt 3600
ggccctggca gtggcatccc gaaacgccct ggacgcaggg ccaagactg ggcacaggag 3660
tgggaggtac atggggctgg ggactcccc ggagttatct gctccctgca ggcctagaga 3720
agtttcaggg aaggtcagaa gagctcctgg ctgtgggtgg cagggcagga aaccctcca 3780
cctttacaca tgcccaggca gcacctcagg ccctttgtgg ggcagggaag ctgaggcagt 3840
aagcgggcag gcagagctgg aggcctttca ggcccagcca gcactctggc ctctgcccgc 3900
cgcatccac cccagcccct cacaccactc gggagaggga catcctacgg tccaaggctc 3960
aggagggcag ggctggggtt gactcaggcc cctcccagct gtggccacct ggggtgttggg 4020
agggcagaag tgcaggcacc tagggcccc catgtgccc ccctgggagc tctccttgga 4080
accattcct gaaaattat taaaggggtt gccggactag ttacgtagat aagtagcatg 4140
gcgggttaat cattaactac aaggaacccc tagtgatgga gttggccact ccctctctgc 4200
gcgctcgtc gctcactgag gccgggcgac caaagtcgc ccgacgccc ggctttgccc 4260
gggcggcctc agtgagcgag cgagcgcgcc agctggcgta atagcgaaga ggcccgcacc 4320
gatcgcctt cccaacagtt gcgcagcctg aatggcgaat ggcgattccg ttgcaatggc 4380

```

tggcggtaat attgttcttg atattaccag caaggccgat agtttgagtt cttctactca 4440
ggcaagtgat gttattacta atcaaagaag tattgcgaca acggttaatt tgcgtgatgg 4500
acagactcctt ttactcgggtg gcctcactga ttataaaaac acttctcagg attctggcgt 4560
accgttcctg tctaaaatcc ctttaatcgg cctcctgttt agctcccgt ctgattctaa 4620
cgaggaaagc acgttatagc tgctcgtcaa agcaaccata gtacgcgccc tgtagcggcg 4680
cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcggcc 4740
tagcggccgc tcctttcgct ttcttccctt cttttctcgc cacgttcgcc ggctttcccc 4800
gtcaagctct aaatcggggg ctcccttttag ggttccgatt tagtgcttta cggcacctcg 4860
accccaaaaa acttgattag ggtgatggtt cacgtagtgg gccatcgccc tgatagacgg 4920
tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg ttccaaactg 4980
gaacaacact caaccctatc tcggtctatt cttttgattt ataagggtt ttgccgattt 5040
cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat ttttaacaaa 5100
tattaacgtt tacaatttaa atatttgctt atacaatctt cctgtttttg gggtttttct 5160
gattatcaac cggggtacat atgattgaca tgctagtttt acgattaccg ttcacgatt 5220
ctcttgtttg ctccagactc tcaggcaatg acctgatagc ctttgtagag acctctcaa 5280
aatagctacc ctctccggca tgaatttatc agctagaacg gttgaatata atattgatgg 5340
tgatttgact gtctccggcc tttctcaccg gtttgaatct ttacctacac attactcagg 5400
cattgcattt aaaatatatg agggttctaa aaatttttat ctttgcgttg aaataaaggc 5460
ttctcccga aaagtattac agggtcataa tgtttttggt acaaccgatt tagctttatg 5520
ctctgaggct ttattgctta attttgctaa ttctttgcct tgcctgtatg atttattgga 5580
tgttggaatc gcctgatgcg gtattttctc cttacgcate tgtgcggtat ttcacaccgc 5640
atatggtgca ctctcagtac aatctgctct gatgccgat agttaagcca gccccgacac 5700
ccgccaacac ccgtgacgc gccctgacgg gcttgctcgc tcccggcatc cgcttacaga 5760
caagctgtga ccgtctccgg gagctgcatg tgcagaggt tttaccgctc atcaccgaaa 5820
cgcgcgagac gaaagggcct cgtgatacgc ctatttttat aggttaatgt catgataata 5880
atggtttctt agacgtcagg tggcactttt cggggaaatg tgcgcggaac ccctatttgt 5940
ttatttttct aaatacatc aaatatgtat ccgtcatga gacaataacc ctgataaatg 6000
cttcaataat attgaaaaag gaagagtatg agtattcaac atttccgtgt cgcccttatt 6060
cccttttttg cggcattttg cttcctgtt tttgctcacc cagaaacgct ggtgaaagta 6120
aaagatgctg aagatcagtt ggggtgcacga gtgggttaca tcgaactgga tctcaacagc 6180
ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc caatgatgag cactttttaa 6240
gttctgctat gtggcgcggt attatcccgt attgacgccg ggcaagagca actcggtcgc 6300
cgcatacact attctcagaa tgacttggtt gactactcac cagtcacaga aaagcatctt 6360
acggatggca tgacagtaag agaattatgc agtctgccaa taaccatgag tgataacact 6420
gcggccaact tacttctgac aacgatcgga ggaccgaagg agctaaccgc ttttttgac 6480
aacatggggg atcatgtaac tcgccttgat cgttggaac cggagctgaa tgaagccata 6540
ccaaacgacg agcgtgacac cacgatgcct gtagcaatgg caacaacgtt gcgcaacta 6600
ttaactggcg aactacttac tctagcttcc cggcaacaat taatagactg gatggaggcg 6660
gataaagttg caggaccact tctgcgctcg gcccttcgg ctggctggtt tattgctgat 6720

aaatctggag cccgtgagcg tgggtctcgc ggtatcattg cagcactggg gccagatggt 6780
 aagccctccc gtatcgtagt tatctacacg acggggagtc aggcaactat ggatgaacga 6840
 aatagacaga tcgctgagat aggtgcctca ctgattaagc attggttaact gtcagaccaa 6900
 gtttactcat atatacttta gattgattta aaacttcatt ttttaatttaa aaggatctag 6960
 gtgaagatcc tttttgataa tctcatgacc aaaatccctt aacgtgagtt ttcgttccac 7020
 tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccttt ttttctgcgc 7080
 gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag cgggtggtttg tttgccggat 7140
 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca gataccaaat 7200
 actgtccttc tagtgtagcc gtagttaggc caccacttca agaactctgt agcaccgcct 7260
 acatacctcg ctctgctaata cctgttacca gtggctgctg ccagtggcga taagtcgtgt 7320
 cttaccgggt tggactcaag acgatatgta ccggataagg cgcagcggtc gggctgaacg 7380
 gggggttcgt gcacacagcc cagcttggag cgaacgacct acaccgaact gagataccta 7440
 cagcgtgagc tatgagaaaag cgccacgctt cccgaaggga gaaaggcgga caggtatccg 7500
 gtaagcggca gggtcggaac aggagagcgc acgaggagc ttccaggggg aaacgcctgg 7560
 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc 7620
 tcgtcagggg ggccggagcct atggaaaaac gccagcaacg cggccttttt acggttcctg 7680
 gcctttttgct ggctttttgc tcacatgttc tttcctgcgt tatcccctga ttctgtggat 7740
 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac gaccgagcgc 7800
 agcgagtcag tgagcgagga agcggaagag cgcccaatac gcaaaccgcc tctccccgcg 7860
 cgttggccga ttcattaatg 7880

<210> 52

<211> 7593

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子5_MND-BFP_远端5'-同源臂构建体

<400> 52

cagctgcgcg ctgcctcgt cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc 60
 tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
 actaggggtt ccttgtagtt aatgattaac ccgccatgct acttatctac gtagaattcg 180
 gccccactgc ccaactgcca gggcagcaat gcccatacca cgttggtcca gctccgagct 240
 tgtcctgaaa agggggcaaa gactggaccc tgagcctgcc aagggggccac actcctccca 300
 gggctggggc ctccatgggc agccccccac ccaccagac cagttacact cccctgtgcc 360
 agagcagtcg agacaggacc aggccaggat gcccagggt caggggctgg ggatgggtag 420
 cccccaaaca gccctttctg ggggactggc ctcaacgggg aaggggggtga aggctcttag 480
 taggaaatca gggagaccca agtcagagcc aggcgctgtg cagaagctgc agcctcacgt 540
 agaaggaaga gcctctgcag tggaggccag tgcccatccc cgggtggcag aggccccagc 600
 agagacttct caatgacatt ccagctgggg tggcccttcc agagcccttg ctgcccagag 660

```

gatgtgagca ggtggccggg gaggctttgt ggggccaccc agcccccttc tcacctctct 720
ccatctctca gactccccag acaggccctg gaaccccccc accttctccc cagccctgct 780
cgtggtgacc gaaggggaca acgccacctt cacctgcagc ttctccaaca catcgagag 840
cttcgtgcta aactggtacc gcatgagccc cagcaaccag acggacaagc tggccgcctt 900
ccccgaggac cgcagccagc ccggccagga ctgccgcttc cgtgtcacac aactgcccac 960
cgggcgtgac ttccacatga gcgtggtcag ggccccggcg aatgacagcg gcacctacct 1020
ctgtggggcc atctccctgg cccccaaggc gcagatcaaa gagagcctgc gggcagagct 1080
cagggtgaca ggtgcggcct cggaggcccc ggggcagggg tgagctgagc cggtcctggg 1140
gtgggtgtcc cctcctgcac aggatcagga gctccagggt cgtagggcag ggacccccca 1200
gctccagtc agggctctgt cctgcacctg gggaatgggt accggcatct ctgtcctcta 1260
gctctggaag caccacagc cctctagtct gccctaccc ctgacctga cctccaccc 1320
tgaccccgct ctaaccctg acctttgtgc cttccagag agaagggcag aagtgccac 1380
agccaccccc agccctcac ccaggccagc cggccagttc caaacctgg tggttggtgt 1440
cgtgggcggc ctgctgggca gcctggtgct gctagtctgg gtcttgccg tcctctatgc 1500
atgaacagag aaacaggaga atatgggcca aacaggatat ctgtggtgta cagttcctgc 1560
cccggctcag ggccaagaac agttggaaca gcagaatat ggccaaacag gatctctgtg 1620
gtaagcagtt cctgccccgg ctgaggcca agaacagatg gtccccagat gcggtccgc 1680
cctcagcagt ttctagagaa ccatcagatg tttccagggt gcccgaagga cctgaaatga 1740
ccctgtgcct tatttgaact aaccaatcag ttcgcttctc gttctgttc gcgcgcttct 1800
gctccccgag ctctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg 1860
ccatccacgc tgttttgact tccatagaag gatctcgagg ccaccatggc tagcgagctg 1920
attaaggaga acatgcacat gaagctgtac atggagggca ccgtggacaa ccatcacttc 1980
aagtgcacat ccgagggcga aggcaagccc tacgagggca ccagaccat gagaatcaag 2040
gtggctgagg gcggccctct ccccttcgcc ttcgacatcc tggctactag cttcctctac 2100
ggcagcaaga ccttcatcaa ccacaccag ggcacccccg acttcttcaa gcagtccttc 2160
cctgagggct tcacatggga gagagtcacc acatacgagg acgggggcgt gctgaccgct 2220
accagagaca ccagcctcca ggacggctgc ctcatctaca acgtcaagat cagaggggtg 2280
aacttcacat ccaacggccc tgtgatgcag aagaaaacac tcggctggga ggccttcacc 2340
gagacgctgt accccgctga cggcggcctg gaaggcagaa acgacatggc cctgaagctc 2400
gtgggcggga gccatctgat cgaaacatc aagaccacat atagatccaa gaaacccgct 2460
aagaacctca agatgcctgg cgtctactat gtggactaca gactggaaag aatcaaggag 2520
gccaacaacg agacctacgt cgagcagcac gaggtggcag tggccagata ctgcgacctc 2580
cctagcaaac tggggcacao gctaaattga aagctttgct ttatttgtga aatttgtgat 2640
gctattgctt tatttgaac cattataagc tgcaataaac aagttaaca acaacaattg 2700
cattcatttt atgtttcagg ttcaggggga ggtgtgggag gttttttaa gtcgaccac 2760
cattgtcttt ctagcgga tgggcacctc atccccgcc cgcaggggct cagctgacgg 2820
ccctcggagt gccagccac tgaggcctga ggatggacac tgctcttggc ccctctgacc 2880
ggcttccttg gccaccagt ttctgcagac cctccacat gagcccggt cagcgcatct 2940
cctcaggaga agcaggcagg gtgcaggcca ttgcaggccg tccaggggct gagctgcctg 3000

```

```

ggggcgaccg gggctccagc ctgcacctgc accaggcaca gccccaccac aggactcatg 3060
tctcaatgcc cacagtgagc ccaggcagca ggtgtcaccg tcccctacag ggaggggccag 3120
atgcagtcac tgcttcaggt cctgccagca cagagctgcc tgcgtccagc tccctgaatc 3180
tctgctgctg ctgctgctgc tgctgctgct gcctgcggcc cggggctgaa ggcgccgtgg 3240
ccctgcctga cgccccggag cctcctgcct gaacttgggg gctggttgga gatggccttg 3300
gagcagccaa ggtgccccctg gcagtggcat cccgaaacgc cctggacgca gggcccaaga 3360
ctgggcacag gagtgggagg tacatggggc tggggactcc ccaggagtta tctgctccct 3420
gcaggcctag agaagtttca gggaaggtea gaagagctcc tggctgtggt gggcagggca 3480
ggaaacccct ccacctttac acatgccag gcagcacctc aggccctttg tggggcaggg 3540
aagctgaggc agtaagcggg caggcagagc tggaggcctt tcaggcccag ccagcactct 3600
ggcctcctgc cgccgcattc caccacagcc ctcacacca ctcgaggag ggacatccta 3660
cgggtcccaag gtcaggaggg cagggtggg gttgactcag gcccctccca gctgtggcca 3720
cctgggtgtt gggagggcag aagtgcaggc acctagggcc ccccatgtgc ccacctggg 3780
agctctcctt ggaaccatt cctgaaatta tttaaagggg ttggccggac tagttacgta 3840
gataagtagc atggcgggtt aatcattaac tacaaggaac ccctagtgat ggagttagcc 3900
actccctctc tgcgcgctcg ctgcctcact gaggccgggc gaccaaaggc cgcccgacgc 3960
ccgggctttg cccgggcggc ctcagtgagc gagcgagcgc gccagctggc gtaatagcga 4020
agaggcccgcc accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatggcgatt 4080
ccgttgcaat ggctggcggc aatattgttc tggatattac cagcaaggcc gatagtttga 4140
gttcttctac tcaggcaagt gatgttatta ctaatcaaag aagtattgcg acaacggtta 4200
atttgctga tggacagact cttttactcg gtggcctcac tgattataaa aacacttctc 4260
aggattctgg cgtaccgttc ctgtctaaaa tccctttaat cggcctcctg tttagctccc 4320
gctctgattc taacaggaa agcacgttat acgtgctcgt caaagcaacc atagtacgcg 4380
ccctgtagcg gcgcattaag cgcggcgggt gtggtgggta cgcgcacgct gaccgctaca 4440
cttgccagcg ccctagcgcc cgctcctttc gtttcttcc cttcctttct cgccacgttc 4500
gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct 4560
ttacggcacc tcgaccccaa aaaacttgat tagggtgatg gttcacgtag tgggccatcg 4620
ccctgataga cggtttttcg ccctttgacg ttggagtcca cgttctttta tagtggactc 4680
ttgttccaaa ctggaacaac actcaaccct atctcggtct attcttttga tttataaggg 4740
attttgccga tttcgcccta ttggttaaaa aatgagctga tttacaaaaa atttaacgcg 4800
aattttaaca aaatattaac gtttacaatt taaatatttg cttatacaat cttcctgttt 4860
ttggggcttt tctgattatc aaccggggta catatgattg acatgctagt tttacgatta 4920
ccgttcatcg attctcttgt ttgctccaga ctctcaggca atgacctgat agcctttgta 4980
gagacctctc aaaaatagct accctctccg gcatgaattt atcagctaga acggttgaat 5040
atcatattga tggtgatttg actgtctccg gcctttctca cccgtttgaa tctttacctc 5100
cacattactc aggcatcgca tttaaaatat atgagggttc taaaaatctt tctcttgccg 5160
ttgaaataaa ggcttctccc gcaaaagtat tacagggtca taatgttttt ggtacaaccg 5220
attagcttt atgctctgag gctttattgc ttaattttgc taattctttg cttgcctgt 5280
atgatttatt ggatgttgga atcgctgat gcggtatttt ctccttacgc atctgtgcgg 5340

```

tattttcacac cgcatatggt gcactctcag tacaatctgc tctgatgccg catagttaag 5400
 ccagccccga caccgcgcaa caccgctga cgcgcctga cgggcttgct tgctcccggc 5460
 atccgcttac agacaagctg tgaccgtctc cgggagctgc atgtgtcaga ggttttcacc 5520
 gtcattcaccg aaacgcgcga gacgaaaggg cctcgtgata cgcctatttt tataggttaa 5580
 tgtcatgata ataattggtt cttagacgtc aggtggcact tttcggggaa atgtgcgcgg 5640
 aacccctatt tgtttatttt tctaaataca ttcaaataat tatccgctca tgagacaata 5700
 accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg 5760
 tgtcgcctt attccctttt ttgcggcatt ttgccttctt gtttttgctc acccagaaac 5820
 gctggtgaaa gtaaaaagatg ctgaagatca gttgggtgca cgagtgggtt acatcgaact 5880
 ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt ttccaatgat 5940
 gagcactttt aaagtctctg tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga 6000
 gcaactcggc cgccgcatac actattctca gaatgacttg gttgagtact caccagtcac 6060
 agaaaagcat ctacggatg gcatgacagt aagagaatta tgcagtgtcg ccataaccat 6120
 gagtgataac actgcggcca acttacttct gacaacgac ggaggaccga aggagctaac 6180
 cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct 6240
 gaatgaagcc ataccaaagc acgagcgtga caccacgatg cctgtagcaa tggcaacaac 6300
 gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac aattaataga 6360
 ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc cggctggctg 6420
 gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca ttgcagcact 6480
 ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga gtcaggcaac 6540
 tatggatgaa cgaaatagac agatcgctga gataggtgcc tcaactgatta agcatttgta 6600
 actgtcagac caagtttact catatatact ttagattgat ttaaaacttc atttttaatt 6660
 taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc cttaacgtga 6720
 gttttcgttc cactgagcgt cagaccccggt agaaaagatc aaaggatctt cttgagatcc 6780
 tttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac cagcgggtgt 6840
 ttgtttgccg gatcaagagc taccaactct ttttcgaag gtaactggct tcagcagagc 6900
 gcagatacca aatactgtcc ttctagtgtg gccgtagtta ggccaccact tcaagaactc 6960
 tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg ctgccagtgg 7020
 cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg 7080
 gtcgggctga acgggggggtt cgtgcacaca gccagcttg gagcgaacga cctacaccga 7140
 actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag ggagaaaggc 7200
 ggacaggtat ccggtaaagc gcagggtcgg aacaggagag cgcacgaggg agcttccagg 7260
 gggaaacgcc tggatctttt atagtctgtt cgggtttcgc cacctctgac ttgagcgtcg 7320
 atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcggcctt 7380
 tttacggttc ctggcctttt gctggccttt tgctcacatg ttctttcttg cgttatcccc 7440
 tgattctgtg gataaccgta ttaccgcctt tgagttagct gataccgctc gccgcagccg 7500
 aacgaccgag cgcagcgagt cagtgagcga ggaagcgga gagcgcccaa tacgcaaacc 7560
 gcctctcccc gcgcgttggc cgattcatta atg 7593

<210> 53

<211> 6718

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子1_MND-GFP构建体

<400> 53

```

cagctgcgcg ctcgctcgcg cactgaggcc gcccgggcaa agccccgggcg tcgggcgacc 60
tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
actaggggtt ccttgtagtt aatgattaac ccgccatgct acttatctac gtaaggctgt 180
tgcaggcatc acacggtgga aagatctgga actgtggcca tgggtgtgagg ccatccacaa 240
ggtggaagct ttgaggggga gccgattagc catggacagt tgtcattcag tagggtcacc 300
tgtgccccag cgaaggggga tgggccggga aggagaggc caggcacctg cccccagcag 360
gggcagaggc tgtgggcagc cgggaggctc ccagaggctc cgacagaatg ggagtggggg 420
tgagcccacc cctcactgca gccccaggaa ctgagcccag agggggccac ccaccttccc 480
caggcaggga ggccccggcc ccaggagat gggggggatg ggggaggaga agggcctgcc 540
cccacccggc agcctcagga ggggcagctc gggcgggata tggaaagagg ccacagcagt 600
gagcagagac acagaggagg aaggggccct gagctgggga gacccccacg gggtagggcg 660
tgggggccac gggcccacct cctccccatc tcctctgtct ccctgtctct gtctctctct 720
ccctccccca ccctctcccc agtcctaccc cctcctcacc cctcctcccc cagcactgcc 780
tctgtcactc tcgcccacgt ggatgtggag gaagaggggg cgggagcaag gggcgggcac 840
cctcccttca acctgacctg ggacagtttc cttccgctc acctccgcct gagcagtgga 900
gaaggcggca ctctggtggg gctgctccaa cgcgtgaaca gagaaacagg agaatatggg 960
ccaaacagga tatctgtggt aagcagttcc tgccccggct cagggccaaag aacagttgga 1020
acagcagaat atgggccaaa caggatatct gtggtaaaga gttcctgccc cggctcaggg 1080
ccaagaacag atggtcccca gatgcggtcc cgccctcagc agtttctaga gaaccatcag 1140
atgtttccag ggtgccccaa ggacctgaaa tgacctgtg cttattttga actaaccaat 1200
cagttcgctt ctcgcttctg ttcgcgcgct tctgctcccc gagctctata taagcagagc 1260
tcgttttagtg aaccgtcaga tcgcctggag acgccatcca cgctgttttg acttccatag 1320
aaggatctcg aggccaccat ggtgagcaag ggcgaggagc tgttcaccgg ggtggtgccc 1380
atcctggtcg agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cggcgagggc 1440
gaggcgatg ccacctacgg caagctgacc ctgaagtcca tctgcaccac cggcaagctg 1500
cccgtgccct ggcccaccct cgtgaccacc ctgacctacg gcgtgcagtg cttcagccgc 1560
taccgccgacc acatgaagca gcacgaattc ttcaagtccg ccatgcccga aggctacgtc 1620
caggagcgca ccatcttctt caaggacgac ggcaactaca agaccgcgc cgaggtgaag 1680
ttcgagggcg acacctggt gaaccgcac gagctgaagg gcatcgactt caaggaggac 1740
ggcaacatcc tggggcacaa gctggagtac aactacaaca gccacaacgt ctatatcatg 1800
gccgacaagc agaagaacgg catcaaggtg aacttcaaga tccgccacaa catcgaggac 1860
ggcagcgtgc agctcgccga ccactaccag cagaacaccc ccatcggcga cggccccgtg 1920

```

ctgctgccccg acaaccacta cctgagcacc cagtccgccc tgagcaaaga ccccaacgag 1980
 aagcgcgatac acatggtcct gctggagttc gtgaccgccc ccgggatcac tctcgcatg 2040
 gacgagctgt acaagtaagc ggccgcgctt tatttgtgaa atttgtgatg ctattgcttt 2100
 atttgttaacc attataagct gcaataaaca agttaacaac aacaattgca ttcattttat 2160
 gtttcagggtt cagggggaga tgtgggaggt tttttaaagc ctcaccggtt ctgggcggtg 2220
 ctacaactgg gctggcggcc aggatggttc ttaggtaggt ggggtcggcg gtcagggtgc 2280
 ccagagccag gggctctggag ggaccttcca ccctcagtcc ctggcaggtc ggggggtgct 2340
 gaggcgggcc tggccctggc agcccagggg tcccggagcg aggggtctgg agggacctt 2400
 cactctcagt ccctggcagg tcgggggggtg ctgtggcagg ccagccttg gccccagct 2460
 ctgcccccta ccctgagctg tgtggctttg ggcagctcga actcctgggt tcctctctgg 2520
 gccccaaactc ctccccctggc ccaagtcctt tctttgtctc tgggcaggca ggacctctgt 2580
 cccctctcag ccggtccttg gggtgcgtg tttctgtaga atgacgggtc aggctggcca 2640
 gaaccccaaa ccttggccgt ggggagctct cgtggcggt ctgccttgcc caggcatact 2700
 tggctctcac tcgagttttc ctaaggatgg gatgagcccc atgtgggact aaccttggt 2760
 ttacgacgtc aaagttaga tgagctggtg atatttttct cattatatcc aaagtgtacc 2820
 tgttcgagtg aggacagttc ttctgtctcc aggatccctc ctgggtgggg attgtgcccc 2880
 cctgggtctc tgcccagatt ccagggtctt ccccgagccc tgttcagacc atccgtgggg 2940
 gaggccttgg cctcactctt acgtagataa gtagcatggc gggttaatca ttaactacaa 3000
 ggaaccctta gtgatggagt tggccactcc ctctctgcgc gctcgtcgc tcaactaggc 3060
 cgggcgacca aaggtcgccc gacggcggg ctttgcccg gcggcctcag tgagcgagcg 3120
 agcgcgccag ctggcgtaat agcgaagagg cccgcaccga tcgcccttcc caacagttgc 3180
 gcagcctgaa tggcgaatgg cgattccgtt gcaatggctg gcggtaatat tgttctggat 3240
 attaccagca aggccgatag tttgagttct tctactcagg caagtgatgt tattactaat 3300
 caaagaagta ttgcgacaac ggttaatttg cgtgatggac agactctttt actcgggtggc 3360
 ctcactgatt ataaaaacac ttctcaggat tctggcgtag cgttcctgtc taaaatccct 3420
 ttaatcggcc tcctgttttag ctcccgctct gattctaacg aggaaagcac gttatacgtg 3480
 ctcgtcaaag caaccatagt acgcgcctg tagcggcgca ttaagcgcg cggtgtggt 3540
 ggttacgcgc agcgtgaccg ctacacttgc cagcgcccta gcgcccgtc ctttcgcttt 3600
 cttcccttcc tttctcgcca cgttcgcccgt ctttccccgt caagctctaa atcgggggct 3660
 ccttttaggg ttccgattta gtgctttacg gcacctcgac cccaaaaaac ttgattaggg 3720
 tgatggttca cgtagtgggc catcgccctg atagacggtt tttcgccctt tgacgttgga 3780
 gtccacgttc tttaatagtg gactcttggt ccaactgga acaacactca accctatctc 3840
 ggtctattct tttgatttat aagggtttt gccgatttcg gcctattggt taaaaaatga 3900
 gctgatttaa caaaaattta acgcgaattt taacaaaata ttaacgttta caatttaaat 3960
 atttgcttat acaatcttcc tgtttttggg gcttttctga ttatcaaccg gggtacatat 4020
 gattgacatg ctagttttac gattaccgtt catcgattct cttgtttgct ccagactctc 4080
 aggcaatgac ctgatagcct ttgtagagac ctctcaaaaa tagctaccct ctccggcatg 4140
 aatttatcag ctagaacggt tgaatatcat attgatggtg atttgactgt ctccggcctt 4200
 tctcaccctg ttgaatcttt acctacacat tactcaggca ttgcatttaa aatatatgag 4260

```

ggttctaaaa atttttatcc ttgcgttgaa ataaaggctt ctcccgcaaa agtattacag 4320
ggtcataatg ttttttgtac aaccgattta gctttatgct ctgaggcttt attgcttaat 4380
tttgctaatt ctttgccttg cctgtatgat ttattggatg ttggaatcgc ctgatgcggt 4440
attttctcct tacgcatctg tgcggtatth cacaccgcat atggtgcact ctcagtacaa 4500
tctgctctga tgccgcatag ttaagccagc cccgacaccc gccaacaccc gctgacgcgc 4560
cctgacgggc ttgtctgctc ccggcatccg cttacagaca agctgtgacc gtctccggga 4620
gctgcatgtg tcagaggtht tcaccgtcat caccgaaacg cgcgagacga aagggcctcg 4680
tgatacgctt atttttatag gttaatgtca tgataataat ggtttcttag acgtcagggt 4740
gcacttttct gggaaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa 4800
atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat tgaaaaagga 4860
agagtatgag tattcaacat ttccgtgtcg cccttattec cttttttgcg gcattttgcc 4920
ttcctgtttt tgctcaccca gaaacgctgg tgaaagtaa agatgctgaa gatcagttgg 4980
gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt gagagtttct 5040
gccccgaaga acgttttcca atgatgagca cttttaaaagt tctgctatgt ggcgcggtat 5100
tatcccgat tgacgccggg caagagcaac tcggtcgccg catacactat tctcagaatg 5160
acttggttga gtactacca gtacagaaa agcatcttac ggatggcatg acagtaagag 5220
aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta cttctgacaa 5280
cgatcggagg accgaaggag ctaaccgctt ttttgacaa catgggggat catgtaactc 5340
gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca 5400
cgatgcctgt agcaatggca acaacgttgc gcaaaactatt aactggcgaa ctacttactc 5460
tagcttcccg gcaacaatta atagactgga tggaggcgga taaagttgca ggaccacttc 5520
tgcgctcggc ccttccggct ggctggttta ttgctgataa atctggagcc ggtgagcgtg 5580
ggctctcgcg tatcattgca gactggggc cagatggtaa gccctcccgt atcgtagtta 5640
tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag 5700
gtgcctcact gattaagcat tggtaactgt cagaccaagt ttactcatat atactttaga 5760
ttgattttaa acttcattht taatttaaaa ggatctaggt gaagatcctt tttgataatc 5820
tcatgaccaa aatcccctta cgtgagtht cgttccactg agcgtcagac cccgtagaaa 5880
agatcaaagg atcttcttga gatccttht ttctgcgcgt aatctgctgc ttgcaaacaa 5940
aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca actctthtct 6000
cgaaggtaac tggcttcagc agagcgcaga taccaatac tgtccttcta gtgtagccgt 6060
agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgt ctgctaattc 6120
tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg gactcaagac 6180
gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca 6240
gcttggagcg aacgacctac accgaactga gatactaca gcgtgagcta tgagaaagcg 6300
ccacgcttcc cgaagggaga aaggcggaca ggtatccgtt aagcggcagg gtcggaacag 6360
gagagcgcac gagggagctt ccagggggaa acgcctggtt tctttatagt cctgtcgggt 6420
ttcgccacct ctgacttgag cgtcgattht tgtgatgctc gtcagggggg cggagcctat 6480
ggaaaaacgc cagcaacgcg gcctthttac ggttccctgg cttttgctgg cttttgctc 6540
acatgttctt tcctgcgtta tcccctgatt ctgtggataa ccgtattacc gcctttgagt 6600

```

gagctgatac cgctcgccgc agccgaacga ccgagcgag cgagtcagtg agcgaggaag 6660
cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt cattaatg 6718

<210> 54

<211> 756

<212> DNA

<213> 智人(Homo sapiens)

<400> 54

aggctgttgc aggcatcaca cggtggaaag atctggaact gtggccatgg tgtgaggcca 60
tccacaaggt ggaagctttg agggggagcc gattagccat ggacagttgt cattcagtag 120
ggtcacctgt gccccagcga agggggatgg gccgggaagg cagaggccag gcacctgccc 180
ccagcagggg cagaggtgtg gggcagccgg gaggtccca gaggtccga cagaatggga 240
gtggggttga gcccaccct cactgcagcc caggaacctg agcccagagg gggccacca 300
ccttccccag gcagggaggc ccggcccca gggagatggg ggggatgggg gaggagaagg 360
gcctgcccc acccggcagc ctgaggagg gcagctcggg cgggatatgg aaagaggcca 420
cagcagtgag cagagacaca gaggaggaag gggccctgag ctggggagac cccacgggg 480
tagggcgtgg gggccacggg cccacctcct ccccatctcc tctgtctccc tgtctctgtc 540
tctctctccc tccccaccc tctccccagt cctacccct cctcaccct cctccccag 600
cactgcctct gtcactctcg cccacgtgga tgtggaggaa gagggggcgg gagcaagggg 660
cgggcaccct cccttcaacc tgacctgga cagtttcct tccgctcacc tccgcctgag 720
cagtggagaa ggcggcactc tggtagggct gctcca 756

<210> 55

<211> 750

<212> DNA

<213> 智人(Homo sapiens)

<400> 55

tctgggcggt gctacaactg ggctggcggc caggatggtt cttaggtagg tggggtcggc 60
ggtcaggtgt cccagagcca ggggtctgga gggaccttc accctcagtc cctggcaggt 120
cggggggtgc tgaggcgggc ctggccctgg cagcccagg gtcccggagc gaggggtctg 180
gaggacctt tcaactctag tccctggcag gtcgggggt gctgtggcag gccagcctt 240
ggcccccagc tctgccccct accctgagct gtgtggcttt gggcagctcg aactcctggg 300
ttcctctctg ggccccaaact cctccccctg cccaagtccc ctctttgctc ctgggcaggc 360
aggacctctg tcccctctca gccggtcctt ggggtgcgt gtttctgtag aatgacgggt 420
caggctggcc agaaccctaa accttgccg tggggagtct gcgtggcggc tctgccttgc 480
ccaggcatcc ttgtcctca ctgagtttt cctaaggat ggatgagccc catgtgggac 540
taaccttggc ttacgacgt caaagtttag atgagctggt gatatttttc tcattatata 600
caaagtgtac ctgttcgagt gaggacagtt cttctgtctc caggatccct cctgggtggg 660
gattgtgccc gcctgggtct ctgccagat tccagggtc tccccagcc ctgttcagac 720
catcgtggg ggaggcctt gcctcactct 750

<210> 56

<211> 7609

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子1_MND-CD19CAR构建体

<400> 56

```

cagctgcgcg ctcgctcgct cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc 60
tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
actaggggtt ccttgtagtt aatgattaac ccgccatgct acttatctac gtaaggctgt 180
tgcaggcatc acacggtgga aagatctgga actgtggcca tgggtgtgagg ccatccacaa 240
ggtggaagct ttgaggggga gccgattagc catggacagt tgtcattcag tagggtcacc 300
tgtgccccag cgaaggggga tgggccggga aggagaggc caggcacctg cccccagcag 360
gggcagaggc tgtgggcagc cgggaggctc ccagaggctc cgacagaatg ggagtggggt 420
tgagcccacc cctcactgca gcccaggaac ctgagcccag agggggccac ccaccttccc 480
caggcaggga ggcccggccc ccaggagat gggggggatg ggggaggaga agggcctgcc 540
cccacccggc agcctcagga ggggcagctc gggcgggata tggaaagagg ccacagcagt 600
gagcagagac acagaggagg aaggggccct gagctgggga gacccccacg gggtagggcg 660
tgggggccac gggcccacct cctccccatc tcctctgtct ccctgtctct gtctctctct 720
ccctccccc cctctctccc agtcctaccc cctcctcacc cctcctcccc cagcactgcc 780
tctgtcactc tcgcccacgt ggatgtggag gaagaggggg cgggagcaag gggcgggcac 840
cctcccttca acctgacctg ggacagtttc cttccgctc acctccgcct gagcagtgga 900
gaaggcggca ctctggtggg gctgtctcaa cgcgtgatcc atcgattagt ccaatttgtt 960
aaagacagga tatcagtggg ccaggctcta gttttgactc aacaatatca ccagctgaag 1020
cctatagagt acgagccata gatagaataa aagattttat ttagtctcca gaaaaagggg 1080
ggaatgaaag accccacctg taggtttggc aagctaggat caaggtagg aacagagaga 1140
cagcagaata tgggccaaac aggatatctg tggtaagcag ttccctgcccc ggctcagggc 1200
caagaacagt tggaacagca gaatatgggc caaacaggat atctgtggta agcagttcct 1260
gccccggctc agggccaaga acagatggtc ccagatgct gtcccgccct cagcagtttc 1320
tagagaacca tcagatgttt ccagggtgcc ccaaggacct gaaatgacct tgtgccttat 1380
ttgaactaac caatcagttc gtttctcgct tctgttcgcg cgcttctgct ccccgagctc 1440
aataaaaagag cccacaaccc ctcaactcggc gcgacgcgtc atagccacca tggccttacc 1500
agtgaccgcc ttgctcctgc cgtggcctt gctgctccac gccgccaggc cggacatcca 1560
gatgacacag actacatcct ccctgtctgc ctctctggga gacagagtca ccatcagttg 1620
cagggcaagt caggacatta gtaaatatct aaattggtat cagcagaaac cagatggaac 1680
tggttaaactc ctgatctacc atacatcaag attacactca ggagtcccat caaggttcag 1740
tggcagtggg tctggaacag attattctct caccattagc aacctggagc aagaagatat 1800
tgccacttac ttttgccaac agggtaatac gcttccgtac acgttcggag gggggaccaa 1860
gctggagatc acaggtggcg gtggctccgg cggtggtggg tctggtggcg gcggaagcga 1920

```

```

ggtgaaactg caggagtcag gacctggcct ggtggcgccc tcacagagcc tgtccgtcac 1980
atgcaactgtc tcaggggtct cattacccga ctatgggtga agctggattc gccagcctcc 2040
acgaaagggg ctggagtggtc tgggagtaat atggggtagt gaaaccacat actataattc 2100
agctctcaaaa tccagactga ccatcatcaa ggacaactcc aagagccaag ttttcttaaa 2160
aatgaacagt ctgcaaaactg atgacacagc catttactac tgtgccaaac attattacta 2220
cggtggtagc tatgctatgg actactgggg tcaaggaacc tcggtcaccg tctcctcaac 2280
cacgacgcca gcgccgcgac caccaacacc ggcgcccacc atcgcgtcgc agcccctgtc 2340
cctgcgcccc gaggcgtgcc ggccagcggc gggggcgcca gtgcacacga gggggctgga 2400
cttcgcctgt gatattctaca tctgggcgcc cttggccggg acttgtgggg tcttctcct 2460
gtcaactgtg atcacccttt actgcaaacg gggcagaaag aaactcctgt atatattcaa 2520
acaaccattt atgagaccag tacaactac tcaagaggaa gatggctgta gctgccgatt 2580
tccagaagaa gaagaaggag gatgtgaact gagagtgaag ttcagcagga gcgcagacgc 2640
ccccgcgtac cagcagggcc agaaccagct ctataacgag ctcaatctag gacgaagaga 2700
ggagtacgat gttttggaca agagacgtgg ccgggacctt gagatggggg gaaagccgag 2760
aaggaagaac cctcaggaag gcctgtacaa tgaactgcag aaagataaga tggcggaggc 2820
ctacagttag attgggatga aaggcgagcg ccggaggggc aaggggcacg atggccttta 2880
ccagggtctc agtacagcca ccaaggacac ctacgacgcc cttcacatgc aggcctgtcc 2940
ccctcgctaa gcggccgcgc tttattttgtg aaattttgtga tgctattgct ttattttgtaa 3000
ccattataag ctgcaataaa caagttaaca acaacaattg cattcatttt atgtttcagg 3060
ttcaggggga gatgtgggag gttttttaa gctcaccggt tctgggcggt gctacaactg 3120
ggctggcggc caggatgggt cttaggtagg tggggtcggc ggtcaggtgt ccagagcca 3180
gggggtctgga gggaccttc accctcagtc cctggcaggt cggggggtgc tgaggcgggc 3240
ctggccctgg cagcccagg gtcccggagc gaggggtctg gagggacctt tcaactctcag 3300
tccctggcag gtcggggggt gctgtggcag gccagacctt ggccccagc tctgcccctt 3360
accctgagct gtgtggcttt gggcagctcg aactcctggg ttcctctctg ggccccaaact 3420
cctccccctg cccaagtccc ctctttgctc ctgggcaggc aggacctctg tcccctctca 3480
gccggtcctt ggggctgcgt gtttctgtag aatgacgggt caggctggcc agaaccccaa 3540
accttggccg tggggagtct gcgtggcggc tctgccttgc ccaggcatcc ttggtcctca 3600
ctcgagtttt cctaaggatg ggatgagccc catgtgggac taaccttggc ttacgacgt 3660
caaagttag atgagctggt gatatttttc tcattatata caaagtgtac ctgttcgagt 3720
gaggacagtt cttctgtctc caggatccct cctgggtggg gattgtgccc gcctgggtct 3780
ctgcccagat tccagggtc tccccagacc ctgttcagac catccgtggg ggaggccttg 3840
gcctcactct tacgtagata agtagcatgg cgggttaatc attactaca aggaaccct 3900
agtgatggag ttggccactc cctctctgcg cgctcgtcgc ctcaactagg ccgggcgacc 3960
aaaggctgcc cgacgcccgg gctttgcccg ggcgccctca gtgagcgagc gagcgcgcca 4020
gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga 4080
atggcgaaat gcgattccgt tgcaatggct ggcggttaata ttgttctgga tattaccagc 4140
aaggccgata gtttgagttc ttctactcag gcaagtgatg ttattactaa tcaaagaagt 4200
attgcgacaa cggttaattt gcgtgatgga cagactcttt tactcggtgg cctcactgat 4260

```

tataaaaaca cttctcagga ttctggcgta ccgttcctgt ctaaaatccc tttaatcggc 4320
ctcctgttta gctcccgctc tgattctaac gaggaagca cgttatacgt gctcgtcaaa 4380
gcaaccatag tacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg 4440
cagcgtgacc gctacacttg ccagcgccct agcgcccgt cctttcgctt tcttccttc 4500
ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc tccctttagg 4560
gttccgattt agtgctttac ggcacctga ccccaaaaaa cttgattagg gtgatggttc 4620
acgtagtggg ccatcgccct gatagacggg ttttcgccct ttgacgttgg agtccacgtt 4680
ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct cggctctattc 4740
ttttgattta taagggattt tgccgatttc ggctattgg ttaaaaaatg agctgattta 4800
acaaaaattt aacgcgaatt ttaacaaaat attaacgttt acaatttaaa tatttgctta 4860
tacaatcttc ctgtttttgg ggcttttctg attatcaacc ggggtacata tgattgacat 4920
gctagtttta cgattaccgt tcatcgattc tcttgtttgc tccagactct caggcaatga 4980
cctgatagcc tttgtagaga cctctcaaaa atagctacce tctccggcat gaatttatca 5040
gctagaacgg ttgaatatca tattgatggg gatttgactg tctccggcct ttctcaccgc 5100
tttgaatctt tacctacaca ttactcaggc attgcattta aatatatga gggttctaaa 5160
aatttttatc cttgcgttga aataaaggct tctcccgcaa aagtattaca gggtcataat 5220
gtttttggta caaccgattt agctttatgc tctgaggctt tattgcttaa ttttgctaat 5280
tctttgcctt gcctgtatga tttattggat gttggaatcg cctgatgcgg tattttctcc 5340
ttacgcatct gtgcggtatt tcacaccgca tatggtgcac tctcagtaca atctgctctg 5400
atgccgcata gttaagccag ccccgacacc cgccaacacc cgctgacgcg ccctgacggg 5460
cttgtctgct cccggcatcc gcttacagac aagctgtgac cgtctccggg agctgcatgt 5520
gtcagagggt ttcaccgtca tcaccgaaac gcgcgagacg aaagggcctc gtgatacgcc 5580
tatttttata ggftaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 5640
ggggaaatgt gcgcggaacc cctatttggt tatttttcta aatacattca aatatgtatc 5700
cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 5760
gtattcaaca tttccgtgtc gcccttattc ctttttttgc ggcattttgc cttcctgttt 5820
ttgctcacc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 5880
tggtttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 5940
aacgttttcc aatgatgagc acttttaaaag ttctgctatg tggcgcggta ttatcccgt 6000
ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 6060
agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 6120
gtgctgccat aaccatgagt gataaactg cggccaactt acttctgaca acgatcgag 6180
gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 6240
gttggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 6300
tagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 6360
ggcaacaatt aatagactgg atggaggcgg ataaagtgc aggaccactt ctgcgctcgg 6420
cccttccggc tggctggttt attgctgata aatctggagc cggtagcgt gggctctcgcg 6480
gtatcattgc agcactgggg ccagatggta agccctccc tatcgtagtt atctacacga 6540
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 6600

tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 6660
 aacttcatatt ttaattttaa aggatctagg tgaagatcct ttttgataat ctcatgacca 6720
 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 6780
 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 6840
 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa 6900
 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 6960
 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaata ctgttaccag 7020
 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 7080
 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttgagac 7140
 gaacgacctt caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 7200
 ccgaaggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 7260
 cgaggagct tccaggggga aacgcctggt atctttatag tctgtcggg tttcgccacc 7320
 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 7380
 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 7440
 ttctgcgtt atccccctgat tctgtggata accgtattac cgcctttgag tgagctgata 7500
 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 7560
 gcccaatacg caaaccgcct ctccccgcgc gttggccgat tcattaatg 7609

<210> 57

<211> 7492

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子1_MND-BCMACAR构建体

<400> 57

cagctgcgcg ctgctcgct cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc 60
 tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
 actaggggtt cctttagatt aatgattaac ccgccatgct acttatctac gtaaggctgt 180
 tgcaggcatc acacggtgga aagatctgga actgtggcca tgggtgtgagg ccatccacaa 240
 ggtggaagct ttgaggggga gccgattagc catggacagt tgtcattcag tagggtcacc 300
 tgtgccccag cgaaggggga tgggccggga aggcagaggc caggcacctg cccccagcag 360
 gggcagaggc tgtgggcagc cgggaggtc ccagaggctc cgacagaatg ggagtggggt 420
 tgagcccacc cctcactgca gcccaggaac ctgagcccag agggggccac ccaccttccc 480
 caggcaggga ggccccggcc ccaggagat ggggggggatg ggggaggaga agggcctgcc 540
 cccacccggc agcctcagga ggggcagctc gggcgggata tggaaagagg ccacagcagt 600
 gagcagagac acagaggagg aaggggccct gagctgggga gacccccacg gggtagggcg 660
 tggggggccac gggcccacct cctccccatc tcctctgtct ccctgtctct gtctctctct 720
 cctccccca cctctcccc agtcctacc cctcctcacc cctcctcccc cagcactgcc 780
 tctgtcactc tcgccacgt ggatgtggag gaagaggggg cgggagcaag gggcgggcac 840

```

cctcccttca acctgacctg ggacagtttc ctttccgctc acctccgcct gagcagtgga 900
gaaggcggca ctctgggtgg gctgctccaa cgcgtaatga aagaccccac ctgtaggttt 960
ggcaagctag gatcaagggt aggaacagag agacagcaga atatgggcca aacaggatat 1020
ctgtggtaag cagttcctgc cccggctcag ggccaagaac agttggaaca gcagaatatg 1080
ggccaaacag gatatctgtg gtaagcagtt cctgccccgg ctcagggccca agaacagatg 1140
gtccccagat gcggtcccgc cctcagcagt ttctagagaa ccatcagatg tttccagggt 1200
gccccaaagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag ttcgcttctc 1260
gcttctgttc gcgcgcttct gctccccgag ctcaataaaa gagcccacaa cccctcactc 1320
ggcgcgattc acctgacgcg tctacgccac catggcactc cccgtcacccg cccttctctt 1380
gccccctgcc ctgctgctgc atgctgccag gcccgacatt gtgctcactc agtcacctcc 1440
cagcctggcc atgagcctgg gaaaaagggc caccatctcc thtagagcca gtgagtccgt 1500
cacaatcttg gggagccatc ttattcactg gtatcagcag aagcccgggc agcctccaac 1560
ccttcttatt cagctcgcgt caaacgtcca gacgggtgta cctgccagat tttctggtag 1620
cgggtcccgc actgatttta cactgaccat agatccagtg gaagaagacg atgtggccgt 1680
gtattattgt ctgcagagca gaacgattcc tcgcacattt ggtgggggta ctaagctgga 1740
gattaaggga agcacgtccg gctcaggga gcccggctcc ggcgaggga gcacgaagg 1800
gcaaatcag ctggtccaga gcggaacctga gctgaaaaaa cccggcgaga ctgttaagat 1860
cagttgtaaa gcatctggct ataccttcac cgactacagc ataaattggg tgaaacgggc 1920
ccctggaaag ggctcaaat ggatgggttg gatcaatacc gaaactaggg agcctgctta 1980
tgcatatgac ttccgcggga gattcgcctt ttactcagag acatctgcct ctactgctta 2040
cctccaaata aacaacctca agtatgaaga tacagccact tacttttgcg ccctcgacta 2100
tagttacgcc atggactact ggggacaggg aacctccgtt accgtcagtt ccgcggccgc 2160
aaccacaaca cctgctccaa ggccccccac acccgctcca actatagcca gccaaccatt 2220
gagcctcaga cctgaagctt gcaggccgc agcaggaggc gccgtccata cgcgaggcct 2280
ggacttcgcg tgtgatattt atatttgggc ccctttggcc ggaacatgtg ggtgttgct 2340
tctctccctt gtgatcactc tgtattgtaa gcgcgggaga aagaagctcc tgtacatctt 2400
caagcagcct tttatgcgac ctgtgcaaac cactcaggaa gaagatgggt gttcatgccg 2460
cttccccgag gaggaagaag gaggggtgta actgagggtg aaattttcta gaagcgccga 2520
tgctcccgca tatcagcagg gtcagaatca gctctacaat gaattgaatc tcggcaggcg 2580
agaagagtac gatgttctgg acaagagacg gggcagggat cccgagatgg ggggaaagcc 2640
ccggagaaaa aatcctcagg aggggttgta caatgagctg cagaaggaca agatggctga 2700
agcctatagc gagatcgga tgaaaggcga aagacgcaga ggcaaggggc atgacggtct 2760
gtaccagggt ctctctacag ccaccaagga cacttatgat gcgttgcata tgcaagcctt 2820
gccacccgc taagcggccg cgctttattt gtgaaatttg tgatgctatt gctttatttg 2880
taaccattat aagctgcaat aaacaagtta acaacaaca ttgcattcat tttatgtttc 2940
aggttcaggg ggagatgtgg gaggtttttt aaagctcacc ggttctgggc ggtgctacaa 3000
ctgggctggc ggccaggatg gttcttaggt aggtggggtc ggcggtcagg tgtcccagag 3060
ccagggtct ggagggacct tccacctca gtccctggca ggtcggggg tgctgaggcg 3120
ggcctggccc tggcagccca ggggtcccgc agcgagggt ctggaggac ctttactct 3180

```

cagtcacctgg caggtcgggg ggtgctgtgg caggcccagc cttggccccc agctctgccc 3240
 cttacctga gctgtgtggc tttgggcagc tcgaactcct gggttcctct ctgggccccca 3300
 actcctcccc tggcccaagt cccctctttg ctctctgggca ggcaggacct ctgtcccctc 3360
 tcagccggtc cttggggctg cgtgtttctg tagaatgacg ggtcaggctg gccagaaccc 3420
 caaaccttgg ccgtggggag tctgcgtggc ggctctgcct tgcccaggca tccttggtcc 3480
 tcactcgagt tttcctaagg atgggatgag ccccatgtgg gactaacctt ggctttacga 3540
 cgtcaaagtt tagatgagct ggtgatattt ttctcattat atccaaagtg tacctgttcg 3600
 agtgaggaca gttcttctgt ctccaggatc cctcctgggt ggggattgtg cccgcctggg 3660
 tctctgcccc gattccaggg ctctccccga gccctgttca gaccatccgt gggggaggcc 3720
 ttggcctcac tcttacgtag ataagtagca tggcgggtta atcattaact acaaggaacc 3780
 cctagtgatg gaggttggcca ctccctctct gcgcgtctgc tcgctcactg aggccgggcg 3840
 accaaaggtc gcccgcagcc cgggctttgc ccgggcggcc tcagtgcgcg agcgagcgcg 3900
 ccagctggcg taatagcgaa gagggccgca ccgatcgcc ttcccaacag ttgcgcagcc 3960
 tgaatggcga atggcgattc cgttgcaatg gctggcggtta atattgttct ggatattacc 4020
 agcaaggccg atagtgttag ttcttctact caggcaagtg atgttattac taatcaaaga 4080
 agtattgcga caacggttaa tttgcgtgat ggacagactc ttttactcgg tggcctcact 4140
 gattataaaa acacttctca ggattctggc gtaccgttcc tgtctaaaat ccctttaatc 4200
 ggctcctgt ttagctcccg ctctgattct aacgaggaaa gcacgttata cgtgctcgtc 4260
 aaagcaacca tagtacgcgc cctgtagcgg cgcattaagc gcggcgggtg tgggtggttac 4320
 gcgcagcgtg accgctacac ttgccagcgc cctagcgccc gctcctttcg ctttcttccc 4380
 ttcttttctc gccacgttcg ccggctttcc ccgtcaagct ctaaateggg ggctcccttt 4440
 agggttccga tttagtgtt tacggcacct cgaccccaaa aaacttgatt aggggtgatgg 4500
 ttcacgtagt gggccatcgc cctgatagac ggtttttcgc cctttgacgt tggagtccac 4560
 gttctttaat agtggactct tgttccaaac tggaacaaca ctcaacccta tctcgggtcta 4620
 ttcttttgat ttataaggga ttttgccgat ttcggcctat tggttaaaaa atgagctgat 4680
 ttaacaaaaa tttaacgcga attttaacaa aatattaacg ttacaattt aaatatattgc 4740
 ttatacaatc ttcctgtttt tggggctttt ctgattatca accgggggtac atatgattga 4800
 catgctagtt ttacgattac cgttcacga ttctcttgtt tgctccagac tctcaggcaa 4860
 tgacctgata gcctttgtag agacctctca aaaatagcta ccctctccgg catgaattta 4920
 tcagctagaa cggttgaata tcatattgat ggtgatttga ctgtctccgg cttttctcac 4980
 ccgtttgaat ctttacctac acattactca ggcattgcat ttaaaatata tgagggttct 5040
 aaaaattttt atccttgctg tgaataaaag gcttctcccg caaaagtatt acagggtcat 5100
 aatgtttttg gtacaaccga tttagcttta tgctctgagg ctttattgct taattttgct 5160
 aattctttgc cttgcctgta tgatttattg gatgttgga tcgcctgatg cggatatttc 5220
 tccttacgca tctgtgcggt atttcacacc gcatatggtg cactctcagt acaatctgct 5280
 ctgatgccgc atagttaagc cagccccgac acccgccaac acccgctgac gcgcctgac 5340
 gggcttgtct gctcccggca tccgcttaca gacaagctgt gaccgtctcc gggagctgca 5400
 tgtgtcagag gttttcaccg tcatcaccga aacgcgcgag acgaaagggc ctcgtgatac 5460
 gcctattttt atagggttaat gtcattgataa taatggtttc ttagacgtca ggtggcactt 5520

```

ttcggggaaa tgtgcgcgga acccctatTT gtttattttt ctaaatacat tcaaatatgt 5580
atccgctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta 5640
tgagtattca acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttcctg 5700
tttttgcTca ccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac 5760
gagtgggtta catcgaactg gatctcaaca gcggtaaagt ccttgagagt tttcgccccg 5820
aagaacgttt tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc 5880
gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg 5940
ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat 6000
gcagtgtctg cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg 6060
gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccctg 6120
atcgTtggga accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc 6180
ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt 6240
cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct 6300
cggcccttcc ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc 6360
gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca 6420
cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgtgag ataggtgcct 6480
cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt 6540
taaaacttca tttttaattt aaaaggatct aggtgaagat cttttttgat aatctcatga 6600
ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca 6660
aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac 6720
caccgctacc agcggTggtt tgtttgccgg atcaagagct accaactctt tttccgaagg 6780
taactggctt cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag 6840
gccaccactt caagaactct gtagcacgcg ctacatacct cgctctgcta atcctgttac 6900
cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 6960
taccggataa ggcgCagcgg tcgggctgaa cgggggggttc gtgcacacag ccagcttgg 7020
agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc 7080
ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc 7140
gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc 7200
acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa 7260
acgccagcaa cgcgcccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt 7320
tctttcctgc gttatcccct gattctgttg ataaccgtat taccgccttt gagtgagctg 7380
ataccgctcg ccgCagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag 7440
agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tg 7492

```

<210> 58

<211> 6147

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成rAAV PD-1外显子1_ATG-mCherry构建体

<400> 58

```

cagctgcgcg ctcgctcgct cactgaggcc gcccgggcaa agccccggcg tcgggcgacc 60
tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag agggagtggc caactccatc 120
actaggggtt ccttgtagtt aatgattaac ccgccatgct acttatctac gtaaggctgt 180
tgcaggcatc acacggtgga aagatctgga actgtggcca tgggtgtgagg ccatccacaa 240
ggtggaagct ttgaggggga gccgattagc catggacagt tgtcattcag tagggtcacc 300
tgtgccccag cgaaggggga tgggccggga aggcagaggc caggcacctg cccccagcag 360
gggcagaggc tgtgggcagc cgggaggctc ccagaggctc cgacagaatg ggagtggggt 420
tgagcccacc cctcactgca gcccaggaac ctgagcccag agggggccac ccaccttccc 480
caggcaggga ggccccggcc ccaggagat gggggggatg ggggaggaga agggcctgcc 540
cccacccggc agcctcagga ggggcagctc gggcgggata tggaaagagg ccacagcagt 600
gagcagagac acagaggagg aaggggccct gagctgggga gacccccacg gggtagggcg 660
tggggggcac ggccccacct cctccccatc tctctgtctt cctgtctctt gtctctctct 720
ccctccccca ccctctcccc agtcctaccc cctcctcacc cctcctcccc cagcactgcc 780
tctgtcactc tcgccacagt ggatgtggag gaagaggggg cgggagcaag gggcgggcac 840
cctcccttca acctgacctg ggacagtttc ctttcgctc acctccgctt gagcagtgga 900
gaaggcggca ctctggtggg gctgtccag gcatgcaggt gagcaagggc gaggaggata 960
actccgccat catcaaggag ttcttgcgct tcaagtgca catggagggc tccgtgaacg 1020
gccacgagtt cgagatcgag ggcgaggcg agggccgccc ctacgagggc acccagaccg 1080
ccaagctgaa ggtgaccaag ggtggcccc tgcccttcgc ctgggacatc ctgtcccctc 1140
agttcatgta cggctccaag gcctacgtga agcaccgcc cgacatcccc gactacttga 1200
agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag gacggcggcg 1260
tggtgaccgt gaccagagc tcctctctgc aggacggcga gttcatctac aaggtgaagc 1320
tgcgcggcac caacttcccc tccgacggcc ccgtaatgca gaagaagacc atgggctggg 1380
aggcctcctc cgagcggatg taccgagagg acggcgccct gaaggcgag atcaagcaga 1440
ggctgaagct gaaggacggc ggccactacg acgtgaggt caagaccacc tacaaggcca 1500
agaagcccgt gcagctgccc ggcgctaca acgtcaacat caagttggac atcacctccc 1560
acaacgagga ctacaccatc gtggaacagt acgaacgcgc cgagggccgc cactccaccg 1620
gcggcatgga cgagctgtac aagtgtgaa agctttgctt tttttgtgaa atttgtgatg 1680
ctatgcttta tttgtaacca ttataagctg caataaaca gtttaacaac aacaattgca 1740
ttcattttat gtttcaggtt cagggggagg tgtgggaggt tttttaagt aggtggggtc 1800
ggcggtcagg tgtcccagag ccagggtctt ggaggacct tccacctca gtccctggca 1860
ggtcgggggg tgctgaggcg ggctggccc tggcagccca ggggtcccgg agcgagggtt 1920
ctggagggac ctttactctt cagtccttgg caggtcgggg ggtgctgtgg caggcccagc 1980
cttggcccc agctctgccc cttacctga gctgtgtggc tttgggcagc tcgaactcct 2040
gggttcctct ctgggcccc actcctcccc tggcccaagt cccctctttg ctctgggca 2100
ggcaggacct ctgtcccctc tcagccggtc cttggggctg cgtgtttctg tagaatgacg 2160
ggtcaggctg gccagaacct caaaccttgg ccgtggggag tctgcgtggc ggctctgcct 2220

```

tgcccaggca tccttgggtcc tcaactcgagt tttcctaagg atgggatgag ccccatgtgg 2280
gactaacctt ggctttacga cgtcaaagtt tagatgagct ggtgatattt ttctcattat 2340
atccaaagtg tacctgttcg agtgaggaca gttcttctgt ctccaggata cgtagataag 2400
tagcatggcg ggtaaatcat taactacaag gaacccttag tgatggagtt ggccactccc 2460
tctctgcgcg ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccc acgcccgggc 2520
tttgcccggg cggcctcagt gagcgagcga gcgcgccagc tggcgtaata gcgaagaggc 2580
ccgcaccgat cgcccttccc aacagttgcg cagcctgaat ggcgaaatggc gattccgttg 2640
caatggctgg cggtaaatatt gttctggata ttaccagcaa ggccgatagt ttgagttctt 2700
ctactcaggc aagtgatgtt attactaate aaagaagtat tgcgacaacg gttaattttgc 2760
gtgatggaca gactctttta ctcggtggcc tcaactgatta taaaaacact tctcaggatt 2820
ctggcgtagc gttcctgtct aaaatccctt taatcggcct cctgttttagc tcccgtctctg 2880
attctaacga ggaaagcacg ttatacgtgc tcgtcaaagc aaccatagta cgcgcctgt 2940
agcggcgcat taagcgcggc ggggtgtgtg gttacgcga gcgtgaccgc tacacttgcc 3000
agcgccttag cgcccgtcc tttcgcttcc ttcccttctt ttctcgccac gttegcgggc 3060
tttccccgtc aagctctaaa tcgggggctc ctttaggtt tccgatttag tgctttacgg 3120
cacctcgacc ccaaaaaact tgattagggt gatggttcac gtagtgggcc atcgccctga 3180
tagacggttt ttgcctctt gacgttgag tccacgttct ttaatagtgg actcttgttc 3240
caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata agggattttg 3300
ccgatttcgg cctattgggt aaaaaatgag ctgatttaac aaaaatttaa cgcgaatttt 3360
aacaaaatat taacgtttac aatttaaata tttgcttata caatcttctt gtttttgggg 3420
cttttctgat tatcaaccgg ggtacatatg attgacatgc tagttttacg attaccgttc 3480
atcgattctc ttgtttgctc cagactctca ggcaatgacc tgatagcctt tgtagagacc 3540
tctcaaaaat agctaccctc tccggcatga atttatcagc tagaacggtt gaatatcata 3600
ttgatgggtga tttgactgtc tccggccttt ctacccgtt tgaatcttta cctacacatt 3660
actcaggcat tgcatttaaa atatatgagg gttctaaaaa tttttatcct tgcgttgaaa 3720
taaaggcttc tcccgaaaa gtattacagg gtcataatgt ttttggtaacc accgatttag 3780
ctttatgctc tgaggcttta ttgcttaatt ttgctaattc tttgccttgc ctgtatgatt 3840
tattggatgt tggaatcgcc tgatgcggta tttctcctt acgcatctgt gcggtatttc 3900
acaccgcata tgggtgcactc tcagtacaat ctgctctgat gccgcatagt taagccagcc 3960
ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct tgtctgctcc cggcatccgc 4020
ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt caccgtcatc 4080
accgaaacgc gcgagacgaa agggcctcgt gatagccta tttttatagg ttaatgtcat 4140
gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc 4200
tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg 4260
ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc 4320
ccttattccc ttttttgagg cattttgcct tctgttttt gctcaccag aaacgctggt 4380
gaaagtaaaa gatgctgaag atcagttggg tgcacgagt gggtacatcg aactggatct 4440
caacagcggg aagatccttg agagttttcg ccccgaagaa cgtttttcaa tgatgagcac 4500
ttttaaagtt ctgctatgtg gcgcggtatt atcccgtatt gacgccgggc aagagcaact 4560

cggtcgccgc atacactatt ctcagaatga cttggttgag tactcaccag tcacagaaaa 4620
 gcatcttacg gatggcatga cagtaagaga attatgcagt gctgccataa ccatgagtga 4680
 taacactgcg gccaaacttac ttctgacaac gatcggagga ccgaaggagc taaccgcttt 4740
 tttgcacaac atgggggagc atgtaactcg cttgatcgt tgggaaccgg agctgaatga 4800
 agccatacca aacgacgagc gtgacaccac gatgcctgta gcaatggcaa caacgttgcg 4860
 caaactatta actggcgaac tacttactct agcttcccgg caacaattaa tagactggat 4920
 ggaggcggat aaagttgcag gaccacttct gcgctcggcc cttccggctg gctggtttat 4980
 tgctgataaa tctggagccg gtgagcgtgg gtctcgcggt atcattgcag cactggggcc 5040
 agatggtaag ccctcccgtg tcgtagttat ctacacgacg gggagtcagg caactatgga 5100
 tgaacgaaat agacagatcg ctgagatagg tgccctactg attaagcatt ggtaactgtc 5160
 agaccaagtt tactcatata tacttttagat tgatttaaaa cttcattttt aatttaaaag 5220
 gatctaggtg aagatccttt ttgataatct catgaccaa atcccttaac gtgagttttc 5280
 gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag atcctttttt 5340
 tctgcgcgta atctgctgct tgcaaaaaaaa aaaaccaccg ctaccagcgg tggtttgttt 5400
 gccggatcaa gagctaccaa ctctttttcc gaaggttaact ggcttcagca gacgcagat 5460
 accaaatact gtccttctag ttagaccgta gttaggccac cacttcaaga actctgtagc 5520
 accgcctaca tacctcgtc tgctaatacct gttaccagt gctgctgcca gtggcgataa 5580
 gtcgtgtctt accgggttg actcaagacg atagttaccg gataaggcgc agcggtcggg 5640
 ctgaacgggg gggtcgtgca cacagcccag cttggagcga acgacctaca ccgaactgag 5700
 atacctacag cgtgagctat gagaaagcgc cacgcttccc gaaggagaa aggcgacag 5760
 gtatccggta agcggcagg tcggaacagg agagcgcac agggagcttc cagggggaaa 5820
 cgcttggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt 5880
 gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg cttttttacg 5940
 gttcctggcc ttttgctggc cttttgctca catgttcttt cctgcgttat cccctgattc 6000
 tgtggataac cgtattaccg ctttgagt agctgatacc gctcgccgca gccgaacgac 6060
 cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc ccaatacgca aaccgcctct 6120
 ccccgcgctg tggccgattc attaatg 6147

<210> 59

<211> 232

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 实验室制备

<400> 59

Met Gln Ile Pro Gln Ala Pro Trp Pro Val Val Trp Ala Val Leu Gln

1 5 10 15

Leu Gly Trp Arg Pro Gly Trp Phe Leu Asp Ser Pro Asp Arg Pro Trp

20 25 30

```

Asn Pro Pro Thr Phe Ser Pro Ala Leu Leu Val Val Thr Glu Gly Asp
    35                      40                      45
Asn Ala Thr Phe Thr Cys Ser Phe Ser Asn Thr Ser Glu Ser Phe Val
    50                      55                      60
Leu Asn Trp Tyr Arg Met Ser Pro Ser Asn Gln Thr Asp Lys Leu Ala
    65                      70                      75                      80
Ala Phe Pro Glu Asp Arg Ser Gln Pro Gly Gln Asp Cys Arg Phe Arg
    85                      90                      95
Val Thr Gln Leu Pro Asn Gly Arg Asp Phe His Met Ser Val Val Arg
    100                     105                     110
Ala Arg Arg Asn Asp Ser Gly Thr Tyr Leu Cys Gly Ala Ile Ser Leu
    115                     120                     125
Ala Pro Lys Ala Gln Ile Lys Glu Ser Leu Arg Ala Glu Leu Arg Val
    130                     135                     140
Thr Glu Arg Arg Ala Glu Val Pro Thr Ala His Pro Ser Pro Ser Pro
    145                     150                     155                     160
Arg Pro Ala Gly Gln Phe Gln Thr Leu Val Val Gly Val Val Gly Gly
    165                     170                     175
Leu Leu Gly Ser Leu Val Leu Leu Val Trp Val Leu Ala Val Ile Arg
    180                     185                     190
Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn Met Thr Pro
    195                     200                     205
Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr Ala Pro Pro
    210                     215                     220
Arg Asp Phe Ala Ala Tyr Arg Ser
    225                     230
<210> 60
<211> 303
<212> PRT
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 合成I-0nuI变体PD-1.ile3.外显子1_RD2_B1G2
<220>
<221> MOD_RES
<222> (1) .. (4)
<223> 任何氨基酸或不存在
<220>
<221> MOD_RES

```

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 60

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gln	Leu	Glu	Ile	Arg	Asn	Val
			20					25					30		
Asn	Pro	Asn	Ile	Pro	Arg	Tyr	Lys	Thr	Arg	Leu	Arg	Phe	Glu	Ile	Asp
			35					40					45		
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
			50					55				60			
Lys	Val	Gly	Lys	Ile	Tyr	Asn	Gln	Gly	Asp	Ser	Tyr	Val	Lys	Leu	Arg
65					70				75					80	
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
				85					90					95	
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
				100					105					110	
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
			115						120					125	
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
			130						135					140	
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145					150					155				160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Leu	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
				165					170					175	
Gly	Glu	Gly	His	Phe	Gly	Val	Ile	Leu	Ala	Lys	Arg	Arg	Pro	Ala	Ser
			180						185					190	
Pro	Val	Gln	Val	Arg	Leu	Arg	Phe	Ala	Ile	Gly	Gln	His	Ile	Arg	Asp
			195						200					205	
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Arg	Ile
			210						215					220	
Arg	Glu	Lys	Asn	Ile	Ser	Glu	Lys	Ser	Trp	Leu	Glu	Phe	Glu	Val	Thr
225					230					235				240	
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
				245					250					255	
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
			260						265					270	
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
			275						280					285	

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa
 290 295 300
 <210> 61
 <211> 303
 <212> PRT
 <213> 人工序列(Artificial Sequence)
 <220>
 <221>
 <223> 合成I-OnuI变体PD-1.ile3.外显子1_RD3_B1G2C4
 <220>
 <221> MOD_RES
 <222> (1) .. (4)
 <223> 任何氨基酸或不存在
 <220>
 <221> MOD_RES
 <222> (302) .. (303)
 <223> 任何氨基酸或不存在
 <400> 61
 Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
 1 5 10 15
 Gly Phe Ala Asp Ala Glu Gly Ser Phe Gly Leu Ser Ile Leu Asn Arg
 20 25 30
 Asn Arg Gly Thr Ala Arg Tyr His Thr Arg Leu Ser Phe Thr Ile Met
 35 40 45
 Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
 50 55 60
 Lys Val Gly Ser Ile Leu Asn Asn Gly Asp His Tyr Val Ser Leu Val
 65 70 75 80
 Val Tyr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
 85 90 95
 Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln
 100 105 110
 Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
 115 120 125
 Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp
 130 135 140
 Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu
 145 150 155 160
 Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser

				165				170				175			
Gly	Asp	Gly	Ser	Phe	Phe	Val	Arg	Leu	Arg	Lys	Ser	Asn	Val	Asn	Ala
				180				185				190			
Arg	Val	Arg	Val	Gln	Leu	Val	Phe	Glu	Ile	Ser	Gln	His	Ile	Arg	Asp
				195				200				205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	His	Ile
				210				215				220			
Tyr	Glu	Gly	Asn	Lys	Ser	Glu	Arg	Ser	Trp	Leu	Gln	Phe	Arg	Val	Glu
225				230				235				240			
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
				245				250				255			
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
				260				265				270			
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
				275				280				285			
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Xaa	Xaa	
290				295				300							

62

303

<212> PRT

〈213〉 人工序列(Artificial Sequence)

 $\langle 220 \rangle$

〈221〉

〈223〉 合成I-OnuI变体PD-1.ile3.外显子1 RD3 B1 G2C11

 $\langle 220 \rangle$ $\langle 221 \rangle$ MOD RES $\langle 222 \rangle \quad (1) \dots (4)$

〈223〉 任何氨基酸或不存在

 $\langle 220 \rangle$ $\langle 221 \rangle \text{ MOD RES}$ $\langle 222 \rangle \quad (302) \dots (303)$

〈223〉 任何氨基酸或不存在

<400> 62

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gly	Leu	Ser	Ile	Leu	Asn	Arg
			20					25					30		
Asn	Arg	Gly	Thr	Gly	Arg	Tyr	His	Thr	Arg	Leu	Ser	Phe	Thr	Ile	Met
		35					40					45			

Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
50						55					60				
Lys	Val	Gly	Ser	Ile	Thr	Asn	Asn	Gly	Asp	His	Tyr	Val	Ser	Leu	Val
65					70				75					80	
Val	Tyr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85					90					95		
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
		100						105					110		
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
		115					120					125			
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
		130				135				140					
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145				150					155					160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165					170					175		
Gly	Asp	Gly	Ser	Phe	Phe	Val	Arg	Leu	Arg	Lys	Ser	Asn	Val	Asn	Ala
		180					185						190		
Arg	Val	Arg	Val	Gln	Leu	Val	Phe	Glu	Ile	Ser	Gln	His	Ile	Arg	Asp
		195					200					205			
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	His	Ile
		210				215					220				
Tyr	Glu	Gly	Asn	Lys	Ser	Glu	Arg	Ser	Trp	Leu	Gln	Phe	Arg	Val	Glu
225				230					235					240	
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn
			245					250					255		
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val
		260					265						270		
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp
		275					280					285			
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Xaa	Xaa	
		290				295					300				

<210> 63

<211> 303

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成I-OnuI变体PD-1.ile3.外显子1_RD3_B1 G2C5

<220>

<221> MOD_RES

<222> (1) .. (4)

<223> 任何氨基酸或不存在

<220>

<221> MOD_RES

<222> (302) .. (303)

<223> 任何氨基酸或不存在

<400> 63

Xaa	Xaa	Xaa	Xaa	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	Leu	Thr
1				5					10					15	
Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gly	Leu	Ser	Ile	Leu	Asn	Arg
			20					25					30		
Asn	Arg	Gly	Thr	Ala	Arg	Tyr	His	Thr	Arg	Leu	Ser	Phe	Thr	Ile	Met
			35				40					45			
Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	Thr	Trp
			50				55					60			
Lys	Val	Gly	Ser	Ile	Tyr	Asn	Asn	Gly	Asp	His	Tyr	Val	Ser	Leu	Glu
65				70					75					80	
Val	Phe	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys
			85						90					95	
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln
			100						105					110	
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile
			115						120					125	
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp
			130						135					140	
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu
145				150						155				160	
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ser
			165						170					175	
Gly	Asp	Gly	Ser	Phe	Phe	Val	Arg	Leu	Arg	Lys	Ser	Asn	Val	Asn	Ala
			180						185					190	
Arg	Val	Arg	Val	Gln	Leu	Val	Phe	Glu	Ile	Ser	Gln	His	Ile	Arg	Asp
			195						200					205	
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	His	Ile
			210						215					220	
Tyr	Glu	Gly	Asn	Lys	Ser	Glu	Arg	Ser	Trp	Leu	Gln	Phe	Arg	Val	Glu
225				230						235				240	

Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
 245 250 255
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
 260 265 270
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
 275 280 285
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa
 290 295 300

<210> 64

<211> 943

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成megaTAL PD-1.ile3.外显子1_RD1_B1G2构建体

<400> 64

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg
 1 5 10 15
 Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val
 20 25 30
 Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe
 35 40 45
 Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly
 50 55 60
 Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala
 65 70 75 80
 Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
 85 90 95
 Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro
 100 105 110
 Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly
 115 120 125
 Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly
 130 135 140
 Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 145 150 155 160
 Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 165 170 175
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

	180		185		190
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu					
195		200		205	
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala					
210		215		220	
Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg					
225		230		235	240
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val					
245		250		255	
Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val					
260		265		270	
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp					
275		280		285	
Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu					
290		295		300	
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr					
305		310		315	320
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala					
325		330		335	
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly					
340		345		350	
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys					
355		360		365	
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp					
370		375		380	
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly					
385		390		395	400
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys					
405		410		415	
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn					
420		425		430	
Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val					
435		440		445	
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala					
450		455		460	
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu					
465		470		475	480
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala					
485		490		495	

Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	500	505	510
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	515	520	525
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	530	535	540
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	545	550	555
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	565	570	575
Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	580	585	590
Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	595	600	605
Met	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	Ile	Arg	610	615	620
Arg	Val	Asn	Arg	Arg	Ile	Gly	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Ile	625	630	635
Ser	Arg	Val	Gly	Gly	Ser	Ser	Arg	Arg	Glu	Ser	Ile	Asn	Pro	Trp	Ile	645	650	655
Leu	Thr	Gly	Phe	Ala	Asp	Ala	Glu	Gly	Ser	Phe	Gly	Leu	Ser	Ile	Leu	660	665	670
Asn	Arg	Asn	Arg	Gly	Thr	Ala	Arg	Tyr	His	Thr	Arg	Leu	Ser	Phe	Thr	675	680	685
Ile	Met	Leu	His	Asn	Lys	Asp	Lys	Ser	Ile	Leu	Glu	Asn	Ile	Gln	Ser	690	695	700
Thr	Trp	Lys	Val	Gly	Ile	Ile	Thr	Asn	Asn	Gly	Asp	His	Tyr	Val	Thr	705	710	715
Leu	Arg	Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	725	730	735
Glu	Lys	Tyr	Pro	Leu	Val	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	740	745	750
Lys	Gln	Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	755	760	765
Gly	Ile	Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	770	775	780
Asn	Asp	Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	785	790	795
Pro	Leu	Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe			

	805		810		815
Thr Ser Gly Asp Gly Ser Phe Phe Val Arg Leu Arg Lys Ser Asn Val					
	820		825		830
Asn Ala Arg Val Arg Val Gln Leu Val Phe Glu Ile Ser Gln His Ile					
	835		840		845
Arg Asp Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly					
	850		855		860
His Ile Tyr Glu Gly Asn Lys Ser Glu Arg Ser Trp Leu Gln Phe Arg					
865		870		875	880
Val Glu Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln					
	885		890		895
Glu Asn Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys					
	900		905		910
Lys Val Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly					
	915		920		925
Leu Asp Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg					
	930		935		940

<210> 65

<211> 2832

<212> RNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成PD-1外显子1 RD2_B1G2 megaTAL mRNA构建体

<400> 65

```

augggaagcg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac 60
agucagcagc agcaagagaa gaucaaaccg aaggugcguu cgacaguggc gcagcaccac 120
gaggcacugg ugggccauug guuuacacac gcgcacauug uugcgcucag ccaacacccg 180
gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg 240
acacacgaag acaucguugg cgucggcaaa cagugguccg gcgcacgcgc ccuggaggcc 300
uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaacuu 360
gugaagauug caaaacgugg cggcgugacc gcaauaggagg cagugcaugc aucgcgcaau 420
gcacugacgg gugccccccu gaaccuaacc ccugaucagg uagucgcuau agcuucaaac 480
aacgggggca agcaagcacu ggagaccguu caacgacucc ugccagugcu cugccaagac 540
cacggacuua cgccagauca ggugguugcu auugccucca acaauggcgg gaaacaagcg 600
uuggaaacug ugcagagacu guuaccuguc uugugucaag accacggccu cagccagau 660
caggugguag ccuagcgcg gaauggaggu gguaagcaag ccuugaaac gguccagcgu 720
cuucugccgg uguugugcca ggaccacgga cuaacgccgg aucaggucgu agccauugcu 780
ucaaauaacg gcggcaaaca ggcgcuagag acaguccagc gccucuugcc uguguuauugc 840

```

caggauacacg gcuaaaccac agaccaaguu guggcuauug caucuaacaa ugguggcaaa 900
 caagccuugg agacagugca acgauuacug ccugucuuau gucaggauca uggccugacg 960
 cccgaucagg uaguggcaau cgcaucuaau aauggaggua agcaagcacu ggagacuguc 1020
 cagagauugu uacccguacu augucaagau caugguuuga cgccugauca gguuguugcg 1080
 auagccagca acaacggagg gaaacaggcu cuugaaaccg uacagcgacu ucuccaguc 1140
 uuugccaaag aucacgggcu uacuccugau caagucguag cuaucgccag ccacgacggu 1200
 gggaaacagg ccugggaaac cguacaacgu cuccuccag uacuuguca agaccacggg 1260
 uuaguccgg aucaagucgu cgcgaucgag agcaauagg gggggaagca ggcgcuggaa 1320
 acuguucaga gacugcugcc uguacuugu caggaccaug gucugacacc ugaccaaguu 1380
 guggcgauag ccaguaacaa ugggggaaaa caggcacuag agacgguuca aagguuguug 1440
 cccguucugu gccaggacca cggcuugaca ccggaucagg ugguauguau cguucacac 1500
 gauggcgga aacaggcuuu agaaacaguc caaagacuuc ucccaguccu uuugcaggac 1560
 cacggauga cuccagauca agucguugcu auugcaagua augguggugg uaagcaagcu 1620
 uuagaaaccg uacagaggcu uuugccagug cugugccagg accauggacu gacccugau 1680
 caagugguag caauugcauc ucaugaugga ggaaaacaag cucuggaaag cauugggcc 1740
 cagcugagcc ggccugaucc ggcuuggcc gcguugacca acgaccacu cgucgccuug 1800
 gccugccucg gcggacgucc ugccauggau gcagugaaaa agggauugcc gcacgcgccg 1860
 gaauugauca gaagagucaa ucgccguuu gggaacgca cguccaucg cguugcgaua 1920
 ucuagagugg gaggaagcuc ucgcagagag uccaauaacc cauggauucu gacugguuuc 1980
 gcugaugccg aaggaucauu cgggcuaagc auccucaaca gaaacagagg uacugcuaga 2040
 uaccacacuc gacugucauu cacaaucaug cugcacaaca aggacaauc gauucggag 2100
 aaauuccagu cgacuuggaa ggucggcaua aucaccaaca acggcgacca cuacgucacc 2160
 cugcgcuca cccguuucga agauuugaaa gugauuauug accacuucga gaaauuccg 2220
 cugguaacac agaaauggg cgauuacaag uuuuuaaac aggcauucag cgucauggag 2280
 aacaaagaac aucuaagga gaaugggauu aaggagcucg uacgaaucaa agcuagaug 2340
 aauggggguc ucaaugacga auugaaaaa gcauuuccag agaacaauag caaagagcg 2400
 ccccuuauc auaagaacau uccgaauuuc aaauggcugg cuggauucac aucuggugau 2460
 ggcuccuucu ucgugcgccu aagaaagucu aauguuaug cuagaguacg ugugcaacug 2520
 guauucgaga ucucacagca caucagagac aagaaccuga ugaauucauu gaaacauac 2580
 cuaggcugug gucacaucua cgagggaac aaucugagc gcaguuggcu ccaauucaga 2640
 guagaaaaau ucagcgauau caacgacaag aucauuccg uauuccagga aaauacucug 2700
 auuggcguca aacucgagga cuuugaagau uggugcaagg uugccaaau gaucgaagag 2760
 aagaaacacc ugaccgaauc cgguuuggau gagauuaaga aaaucaagcu gaacaugaac 2820
 aaaggucguu ga 2832
 <210> 66
 <211> 2832
 <212> RNA
 <213> 人工序列(Artificial Sequence)
 <220>

<221>

<223> 合成PD-1外显子1 RD2_B1G2C4 megaTAL mRNA构建体

<400> 66

```

augggaagcg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac 60
agucagcagc agcaagagaa gaucaaaccg aaggugcguu cgacaguggc gcagcaccac 120
gaggcacugg ugggccauug guuuacacac gcgcacaucg uugcgcucag ccaacacccg 180
gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg 240
acacacgaag acaucguugg cgucggcaaa cagugguccg gcgcacgcgc ccuggaggcc 300
uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaacuu 360
gugaagauug caaaacgugg cggcgugacc gcaauggagg cagugcaugc aucgcgcaau 420
gcacugacgg gugccccccu gaaccuaacc ccugaucagg uagucgcuau agcuucaaac 480
aacgggggca agcaagcacu ggagaccguu caacgacucc ugccagugcu cugccaagac 540
cacggacuua cgccagauga ggugguugcu auugccucca acaauggcgg gaaacaagcg 600
uuggaaacug ugcagagacu guuaccuguc uugugucaag accacggccu cacgccagau 660
caggugguag ccuagcguu gaauggaggu gguaagcaag cccuugaaac gguccagcgu 720
cuucugccgg uguugugcca ggaccacgga cuaacgccgg aucaggucgu agccauugcu 780
ucaaauaacg gcggcaaaaca ggcgcuagag acaguccagc gccucuugcc uguguuaugc 840
caggauacag gcuaaaccac agaccaaguu guggcuauug caucuaacaa ugguggcaaa 900
caagccuugg agacagugca acgauuacug ccugucuuau gucaggauca uggccugacg 960
cccgauacagg uaguggcaau cgcaucuaau aauggaggua agcaagcacu ggagacuguc 1020
cagagauugu uaccguacu augucaagau caugguuuga cgccugauga gguuguugcg 1080
auagccagca acaacggagg gaaacaggcu cuugaaaccg uacagcgacu ucuccaguc 1140
uugugccaag aucacgggcu uacuccugau caagucguag cuaucgccag ccacgacgg 1200
gggaaacagg ccugggaaac cguacaacgu cuccucccag uacuuguca agaccacggg 1260
uugacuccgg aucaagucgu cgcgaucgag agcaauagg gggggaagca ggcgcuggaa 1320
acuguucaga gacugcugcc uguacuugu caggaccaug gucugacacc ugaccaaguu 1380
guggcgauag ccaguaacaa ugggggaaaa caggcacuag agacggauca aagguuguug 1440
cccgauucugu gccaggacca cggcuugaca ccggaucagg ugguaagcuau cguucacac 1500
gauggcggaa aacaggcuuu agaaacaguc caaagacuuc ucccaguccu uugucaggac 1560
cacggauga cuccagauga agucguugcu auugcaagua augguggugg uaagcaagcu 1620
uuagaaaccg uacagaggcu uuugccagug cugugccagg accauggacu gacccugau 1680
caagugguag caauugcauc ucaugaugga ggaaaacaag cucuggaaag cauugggcc 1740
cagcugagcc ggccugauc ggcgauuggc gcguugacca acgaccaccu cgucgccuug 1800
gccugccucg gcggacgucc ugccauggau gcagugaaaa agggauugcc gcacgcgccg 1860
gaauugauga gaagagucua ucgccguuu ggcgaaacga cguccaucg cguugcgaua 1920
ucuagagugg gaggaagcuc ucgcagagag uccaucacac cauggauucu gacugguuuc 1980
gcugaugccg aaggaucauu cgggcuaagc auccucaaca gaaacagagg uacugcuaga 2040
uaccacacuc gacugucauu cacaaucaug cugcacaaca aggacaauc gauucggag 2100
aauauccagu cgacuuggaa ggucggcagc auccucaaca auggcgacca cuacgucucg 2160

```

cugguggucu accguuucga agauuugaaa gugauuauug accacuucga gaaauauccg 2220
 cugauaacac agaaaauuggg cgauuacaag uuguuuuaac aggcauucag cgucauggag 2280
 aacaaagaac aucuuaagga gaaugggauu aaggagcucg uacgaaucaa agcuaagaug 2340
 aauggggguc ucaaugacga auugaaaaaa gcauuuccag agaacauuag caaagagcgc 2400
 ccccuuauca auagaacau uccgaauuuc aaugggcugg cuggauucac aucuggugau 2460
 ggcuccuucu ucgugcgccu aagaaagucu aauguuaaug cuagaguacg ugugcaacug 2520
 guauucgaga ucucacagca caucagagac aagaaccuga ugaauucauu gauaacauac 2580
 cuaggcugug gucacauca cgagggaac aaucugagc gcaguuggcu ccaauucaga 2640
 guagaaaaau ucagcgauau caacgacaag aucauuccgg uauuccagga aaauacucug 2700
 auuggcgua aacucgagga cuuugaagau uggugcaagg uugccaaau gaucgaagag 2760
 aagaaacacc ugaccgaauc cgguuuggau gagauuaaga aaaucaagcu gaacaugaac 2820
 aaaggucguu ga 2832

<210> 67

<211> 2832

<212> RNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成PD-1外显子1 RD2_B1G2C11 megaTAL mRNA构建体

<400> 67

augggaagcg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac 60
 agucagcagc agcaagagaa gaucaaaccg aaggugcguu cgacaguggc gcagcaccac 120
 gaggcacugg ugggccauug guuuacacac gcgcacauug uugcgcucag ccaacacccg 180
 gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg 240
 acacacgaag acaucguugg cgucggcaaa cagugguccg gcgcacgcgc ccuggaggcc 300
 uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaacuu 360
 gugaagauug caaaacgugg cggcgugacc gcauaggagg cagugcaugc aucgcgcaau 420
 gcacugacgg gugcccccu gaaccuaacc ccugaucagg uagucgcuau agcuucaaac 480
 aacgggggca agcaagcacu ggagaccguu caacgacucc ugccagugcu cugccaagac 540
 cacggacuua cgccagauca ggugguugcu auugccucca acauaggcgg gaaacaagcg 600
 uuggaaacug ugcagagacu guuaccuguc uuuguucaag accacggccu cacgccagau 660
 caggugguag ccuagcguu gaauggaggu gguaagcaag cccuugaaac gguccagcgu 720
 cuucugccgg uguugugcca ggaccacgga cuaacgccgg aucaggucgu agccauugcu 780
 ucaauaacg gcggcaaaac ggcgcuagag acaguccagc gccucuugcc uguguuauugc 840
 caggauacg gcuaaaccac agaccaaguu guggcuauug caucuaacaa ugguggcaaa 900
 caagccuugg agacagugca acgauuacug ccugucuuau gucaggauca uggccugacg 960
 ccgcaucagg uaguggcaau cgcaucuaau aauggaggua agcaagcacu ggagacuguc 1020
 cagagauugu uaccguacu augucaagau caugguuuga cgccugauca gguuguugcg 1080
 auagccagca acaacggagg gaaacaggcu cuugaaaccg uacagcgacu ucuccaguc 1140

uugugccaag aucacgggcu uacuccugau caagucguag cuaucgccag ccacgacggu 1200
 gggaaacagg cccuggaaac cguacaacgu cuccucccag uacuuuguca agaccacggg 1260
 uugacuccgg aucaagucgu cgcgaucgcg agcaauaggag gggggaagca ggcgcuggaa 1320
 acuguucaga gacugcugcc uguacuugu caggaccaug gucugacacc ugaccaaguu 1380
 guggcgauag ccaguaacaa ugggggaaaa caggcacuag agacgguuca aagguuguug 1440
 cccguucugu gccaggacca cggcuugaca ccggaucagg ugguagcuau cgcuucacac 1500
 gauggcggaa aacaggcuuu agaaacaguc caaagacuuc ucccaguccu uugucaggac 1560
 cacggauuga cuccagauga agucguugcu auugcaagua augguggugg uaagcaagcu 1620
 uuagaaaccg uacagaggcu uuugccagug cugugccagg accauggacu gaccccugau 1680
 caaugguag caauugcauc ucaugaugga ggaaaacaag cucuggaaag cauuguggcc 1740
 cagcugagcc ggccugauc ggcuuggcc gcguugacca acgaccaccu cgucgccuug 1800
 gccugccucg gcggacgucc ugccauggau gcagugaaaa agggauugcc gcacgcgccg 1860
 gaauugauga gaagagucua ucgccguuu gggaacgca cguccaucg cguugcgaua 1920
 ucuagagugg gaggaagcuc ucgcagagag uccaucaacc cauggauucu gacugguuuc 1980
 gcugaugccg aaggaucauu cgggcuaagc auccucaaca gaaacagagg uacugguaga 2040
 uaccacacuc gacugucauu cacaaucaug cugcacaaca aggacaauc gauucuggag 2100
 aaauuccagu cgacuuggaa ggucggcucg aucacgaaca acggcgacca cuacgucagc 2160
 cuggucgucu accguuucga agauuugaaa gugauuauug accacuucga gaaauauccg 2220
 cugauaacac agaaauuggg cgauuacaag uuguuuuaac aggcauucag cguauggag 2280
 aacaaagaac aucuuaagga gaaugggauu aaggagcucg uacgaaucaa agcuagaug 2340
 aauggggguc ucaaugacga auugaaaaaa gcauuuccag agaacaauag caaagagcgc 2400
 ccccuuauca auagaacau uccgaauuuc aaauggcugg cuggauucac aucuggugau 2460
 ggcuccuucu ucgugcgccu aagaaagucu aauguuaaug cuagaguacg ugugcaacug 2520
 guauucgaga ucucacagca caucagagac aagaaccuga ugaauucauu gauaacauac 2580
 cuaggcugug gucacauca cgagggaac aaauucgagc gcaguuggcu ccaauucaga 2640
 guagaaaaau ucagcgauau caacgacaag aucauuccgg uauuccagga aaauacucug 2700
 auuggcguca aacucgagga cuuugaagau uggugcaagg uugccaaau gaucgaagag 2760
 aagaaacacc ugaccgaauc cgguuuggau gagauuaaga aaaucaagcu gaacaugaac 2820
 aaaggucguu ga 2832

<210> 68

<211> 2832

<212> RNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 合成PD-1外显子1 RD2_B1G2C5 megaTAL mRNA构建体

<400> 68

augggaagcg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac 60
 agucagcagc agcaagagaa gaucaaaccg aaggugcguu cgacaguggc gcagcaccac 120

gaggcacugg ugggccauagg guuuacacac gcgcacaucg uugcgcucag ccaacacccg 180
 gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg 240
 acacacgaag acaucguugg cgucggcaaa cagugguccg gcgcacgcgc ccuggaggcc 300
 uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaacuu 360
 gugaagauug caaaacgugg cggcgugacc gcaauaggagg cagugcaugc aucgcgcaau 420
 gcacugacgg gugccccccu gaaccuaacc ccugaucagg uagucgcuau agcuucaaac 480
 aacgggggca agcaagcacu ggagaccguu caacgacucc ugccagugcu cugccaagac 540
 cacggacuua cgccagauca ggugguugcu auugccucca acauaggcgg gaaacaagcg 600
 uuggaaacug ugcagagacu guuaccuguc uugugucaag accacggccu cagccagau 660
 caggugguag ccuagcguu gaauggaggu gguaagcaag cccuugaaac gguccagcgu 720
 cuucugccgg uguugugcca ggaccacgga cuaacgccgg aucaggucgu agccauugcu 780
 ucaaaauaac gcggcaaaaca ggcgcuagag acaguccagc gccucuugcc uguguuauugc 840
 caggaucacg gcuaaaccac agaccaaguu guggcuauug caucuaacaa ugguggcaaa 900
 caagccuugg agacagugca acgauuacug ccugucuauu gucaggauca uggccugacg 960
 cccgaucagg uaguggcaau cgcaucuaau aauggaggua agcaagcacu ggagacuguc 1020
 cagagauugu uacccguacu augucaagau caugguuuga cgccugauca gguuguugcg 1080
 auagccagca acaacggagg gaaacaggcu cuugaaaccg uacagcgacu ucuccaguc 1140
 uuuguccaag aucacgggcu uacuccugau caagucguag cuaucgccag ccacgacggu 1200
 gggaaacagg cccuggaaac cguacaacgu cuccucccag uacuuguca agaccacggg 1260
 uugacuccgg aucaagucgu cgcgauccgc agcaauaggag gggggaagca ggcgcuggaa 1320
 acuguucaga gacugcugcc uguacuuguu caggaccaug gucugacacc ugaccaaguu 1380
 guggcgauag ccaguaacaa ugggggaaaa caggcacuag agacgguuca aagguuguug 1440
 cccguucugu gccaggacca cggcuugaca ccggaucagg ugguaagcuau cgcuucacac 1500
 gauggcggaa aacaggcuuu agaaacaguc caaagacuuc ucccaguccu uugucaggac 1560
 cacggauuga cuccagauca agucguugcu auugcaagua augguggugg uaagcaagcu 1620
 uuagaaaccg uacagaggcu uuugccagug cugugccagg accauggacu gacccugau 1680
 caagugguag caauugcauc ucaugaugga ggaaaacaag cucuggaaag cauugggcc 1740
 cagcugagcc ggccugaucc ggcuuuggcc gcguugacca acgaccaccu cgucgccuug 1800
 gccugccucg gcggacgucc ugccauggau gcagugaaaa agggauugcc gcacgcgccg 1860
 gaauugauca gaagagucaa ucgccguuuu ggcgaacgca cgucccaucg cguugcgaua 1920
 ucuagagugg gaggaagcuc ucgcagagag uccaucacac cauggauucu gacugguuuc 1980
 gcugaugccg aaggaucauu cgggcuaagc auccucaaca gaaacagagg uacugcuaga 2040
 uaccacacuc gacugucauu cacaaucaug cugcacaaca aggacaauc gauucuggag 2100
 aaauuccagu cgacuuggaa ggucggcucg aucuacaaca acggcgacca cuacgucucg 2160
 cuggaggucu uccguuucga agauuugaaa gugauuauug accacuucga gaaauuccg 2220
 cugauaacac agaaaauuggg cgaauacaag uuguuuuuaac aggcauucag cguauggag 2280
 aacaaagaac aucuuuagga gaaugggauu aaggagcucg uacgaaucaa agcuaagaug 2340
 aauggggguc ucaaugacga auugaaaaaa gcauuuccag agaacaauag caaagagcgc 2400
 ccccuuauca auuagaacau uccgaauuuc aaauggcugg cuggauucac aucuggugau 2460

ggcuccuucu ucgugcgccu aagaaagucu aauguuaaug cuagaguacg ugugcaacug 2520
guauucgaga ucucacagca caucagagac aagaaccuga ugaauucauu gauaacauac 2580
cuaggcugug gucacauca cagaggaaac aaaucugagc gcaguuggcu ccaauucaga 2640
guagaaaaau ucagcgauau caacgacaag aucauuccgg uauuccagga aaauacucug 2700
auuggcguca aacucgagga cuuugaagau uggugcaagg uugccaaaau gaucgaagag 2760
aagaaacacc ugaccgaauc cgguuuggau gagauaaga aaaucaagcu gaacaugaac 2820
aaaggucguu ga 2832

<210> 69

<211> 3

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 69

Gly Gly Gly

1

<210> 70

<211> 5

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 70

Asp Gly Gly Gly Ser

1

5

<210> 71

<211> 5

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 71

Thr Gly Glu Lys Pro

1

5

<210> 72

<211> 4

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 72

Gly Gly Arg Arg

1

<210> 73

<211> 5

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 73

Gly Gly Gly Gly Ser

1

5

<210> 74

<211> 14

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 74

Glu Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Val Asp

1

5

10

<210> 75

<211> 18

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 75

Lys Glu Ser Gly Ser Val Ser Ser Glu Gln Leu Ala Gln Phe Arg Ser

1

5

10

15

Leu Asp

<210> 76

<211> 8

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 76

Gly Gly Arg Arg Gly Gly Gly Ser

1 5

<210> 77

<211> 9

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 77

Leu Arg Gln Arg Asp Gly Glu Arg Pro

1 5

<210> 78

<211> 12

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 78

Leu Arg Gln Lys Asp Gly Gly Gly Ser Glu Arg Pro

1 5 10

<210> 79

<211> 16

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 例示性连接子序列

<400> 79

Leu Arg Gln Lys Asp Gly Gly Gly Ser Gly Gly Gly Ser Glu Arg Pro

1	5	10	15
<210> 80			
<211> 7			
<212> PRT			
<213> 人工序列(Artificial Sequence)			
<220>			
<221>			
<223> TEV蛋白酶裂解序列			
<220>			
<221> SITE			
<222> (2) .. (3)			
<223> Xaa是任何氨基酸			
<220>			
<221> SITE			
<222> (5) .. (5)			
<223> Xaa是任何氨基酸			
<220>			
<221> SITE			
<222> (7) .. (7)			
<223> Xaa = Gly或Ser			
<400> 80			
Glu Xaa Xaa Tyr Xaa Gln Xaa			
1	5		
<210> 81			
<211> 7			
<212> PRT			
<213> 人工序列(Artificial Sequence)			
<220>			
<221>			
<223> TEV蛋白酶裂解序列			
<400> 81			
Glu Asn Leu Tyr Phe Gln Gly			
1	5		
<210> 82			
<211> 7			
<212> PRT			
<213> 人工序列(Artificial Sequence)			
<220>			
<221>			

<223> TEV蛋白酶裂解序列

<400> 82

Glu Asn Leu Tyr Phe Gln Ser

1 5

<210> 83

<211> 22

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 83

Gly Ser Gly Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val

1 5 10 15

Glu Glu Asn Pro Gly Pro

20

<210> 84

<211> 19

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 84

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn

1 5 10 15

Pro Gly Pro

<210> 85

<211> 14

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 85

Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn Pro Gly Pro

1 5 10

<210> 86

<211> 21

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 86

Gly Ser Gly Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu

1 5 10 15

Glu Asn Pro Gly Pro

20

<210> 87

<211> 18

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 87

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro

1 5 10 15

Gly Pro

<210> 88

<211> 13

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 88

Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro Gly Pro

1 5 10

<210> 89

<211> 23

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 89

Gly Ser Gly Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp
1 5 10 15

Val Glu Ser Asn Pro Gly Pro
20

<210> 90

<211> 20

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 90

Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp Val Glu Ser
1 5 10 15

Asn Pro Gly Pro
20

<210> 91

<211> 14

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 91

Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly Pro
1 5 10

<210> 92

<211> 25

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 92

Gly Ser Gly Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala
1 5 10 15

Gly Asp Val Glu Ser Asn Pro Gly Pro
20 25

<210> 93

<211> 22

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 93

Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val

1 5 10 15

Glu Ser Asn Pro Gly Pro

20

<210> 94

<211> 14

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 94

Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly Pro

1 5 10

<210> 95

<211> 19

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 95

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn

1 5 10 15

Pro Gly Pro

<210> 96

<211> 19

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 96

Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn
1 5 10 15

Pro Gly Pro

<210> 97

<211> 14

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 97

Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly Pro
1 5 10

<210> 98

<211> 17

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 98

Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly
1 5 10 15

Pro

<210> 99

<211> 20

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 99

Gln Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser
1 5 10 15

Asn Pro Gly Pro
20

<210> 100

<211> 24

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 100

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly

1 5 10 15

Asp Val Glu Ser Asn Pro Gly Pro

20

<210> 101

<211> 40

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 101

Val Thr Glu Leu Leu Tyr Arg Met Lys Arg Ala Glu Thr Tyr Cys Pro

1 5 10 15

Arg Pro Leu Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys

20 25 30

Ile Val Ala Pro Val Lys Gln Thr

35 40

<210> 102

<211> 18

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 102

Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro

1 5 10 15

Gly Pro

<210> 103

<211> 40

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 103

```

Leu Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys Ile Val
1           5           10           15
Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly
           20           25           30
Asp Val Glu Ser Asn Pro Gly Pro
           35           40

```

<210> 104

<211> 33

<212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 包含2A位点的自裂解多肽

<400> 104

```

Glu Ala Arg His Lys Gln Lys Ile Val Ala Pro Val Lys Gln Thr Leu
1           5           10           15
Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly
           20           25           30
Pro

```

<210> 105

<211> 10

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 共同Kozak序列

<400> 105

gccrccatgg 10

<210> 106

<211> 8

<212> PRT

<213> 智人(Homo sapiens)

<400> 106

```

Glu Gln Thr Glu Tyr Ala Thr Ile
1           5

```

<210> 107
<211> 43
<212> DNA
<213> 智人(Homo sapiens)
<400> 107
tgagcagacg gagtatgccca ccattgtctt tcctagcgga atg 43
<210> 108
<211> 43
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<400> 108
cattccgcta ggaaagacaa tgggtggcata ctccgtctgc tca 43
<210> 109
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<220>
<221> misc_feature
<222> (14) .. (22)
<223> n是A、C、G或T
<400> 109
aatggtggca tacnnnnnnnn nn 22
<210> 110
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<220>
<221> misc_feature
<222> (1) .. (9)
<223> n是a、c、g或t

<400> 110
nnnnnnnnna tactccgtct gc 22
<210> 111
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<400> 111
aatggtggca tactccgtct gc 22
<210> 112
<211> 43
<212> DNA
<213> 智人(Homo sapiens)
<400> 112
attccgctag gaaagacaat ggtggcatac tccgtctgct cag 43
<210> 113
<211> 43
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<400> 113
atttcgttag gaaagataat ggtggtatat ttcgtttggt tag 43
<210> 114
<211> 7
<212> PRT
<213> 智人(Homo sapiens)
<400> 114
Met Gln Ile Pro Gln Ala Pro
1 5
<210> 115
<211> 44
<212> DNA
<213> 智人(Homo sapiens)
<400> 115
ctctggtggg gctgctccag gcatgcagat cccacaggcg ccct 44

<210> 116
<211> 44
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<400> 116
agggcgcctg tgggatctgc atccctggag cagccccacc agag 44
<210> 117
<211> 8
<212> PRT
<213> 智人(Homo sapiens)
<400> 117
Asn Gly Arg Asp Phe His Met Ser
1 5
<210> 118
<211> 43
<212> DNA
<213> 智人(Homo sapiens)
<400> 118
gtgtcacaca actgccaac gggcgtgact tccacatgag cgt 43
<210> 119
<211> 43
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<221>
<223> 实验室制备
<400> 119
acgctcatgt ggaagtcacg cccgttgggc agttgtgtga cac 43
<210> 120
<211> 32
<212> DNA
<213> 智人(Homo sapiens)
<400> 120
gctccaggca tgcagatccc acaggcgccc tg 32
<210> 121
<211> 32

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 实验室制备

<400> 121

actccaaaca tacaaatccc acaaacaccc ta 32

<210> 122

<211> 31

<212> DNA

<213> 智人(Homo sapiens)

<400> 122

gcccacggg cgtgacttcc acatgagcgt g 31

<210> 123

<211> 31

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<221>

<223> 实验室制备

<400> 123

acccaacgaa cgtaacttcc acataaacgt a 31

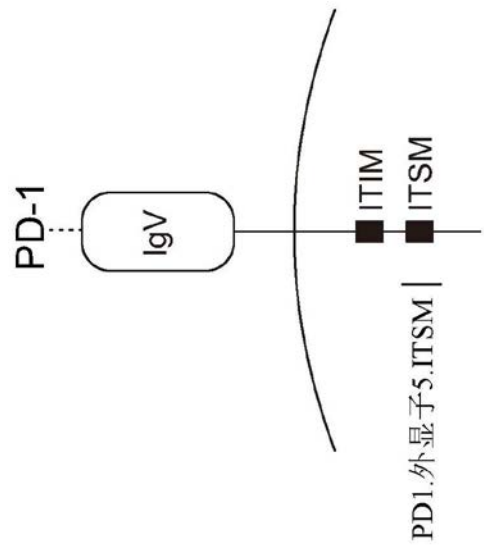


图1A

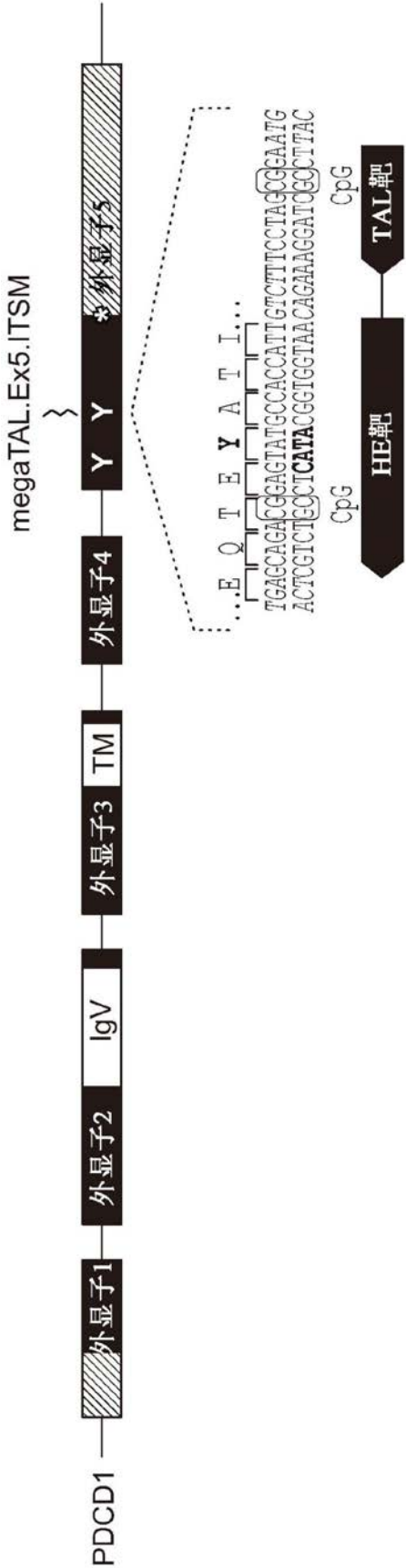


图1B

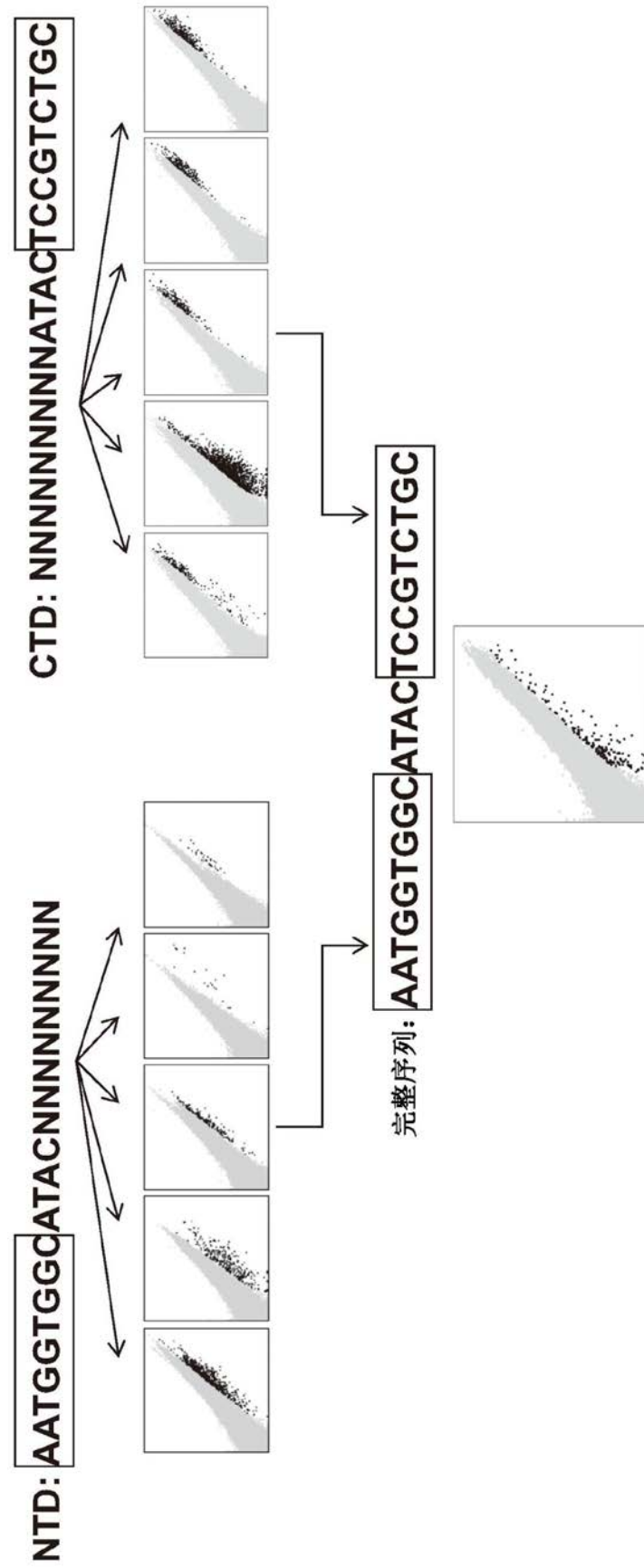


图2

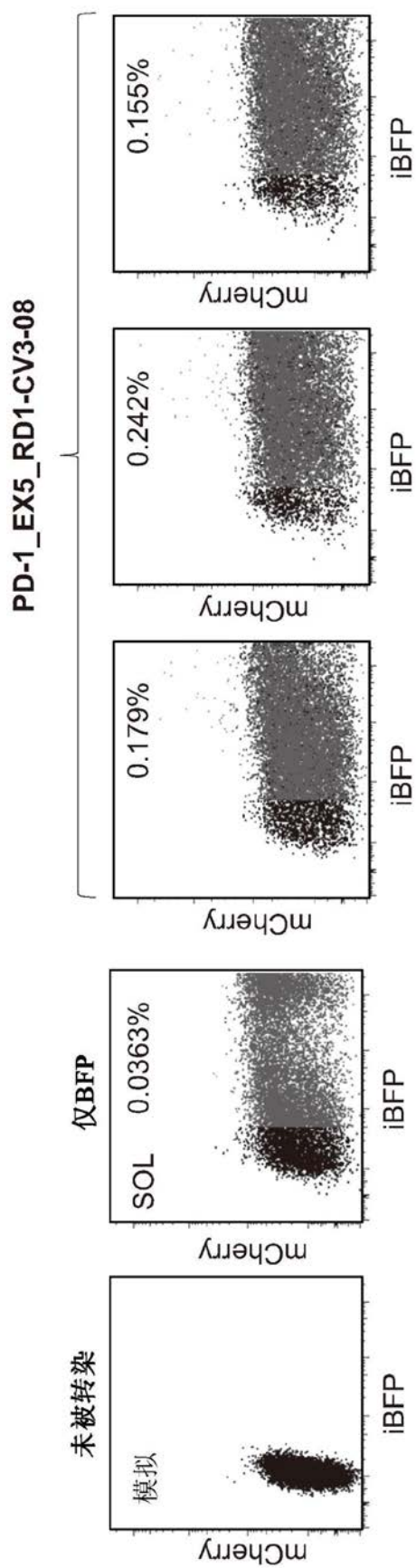


图3A

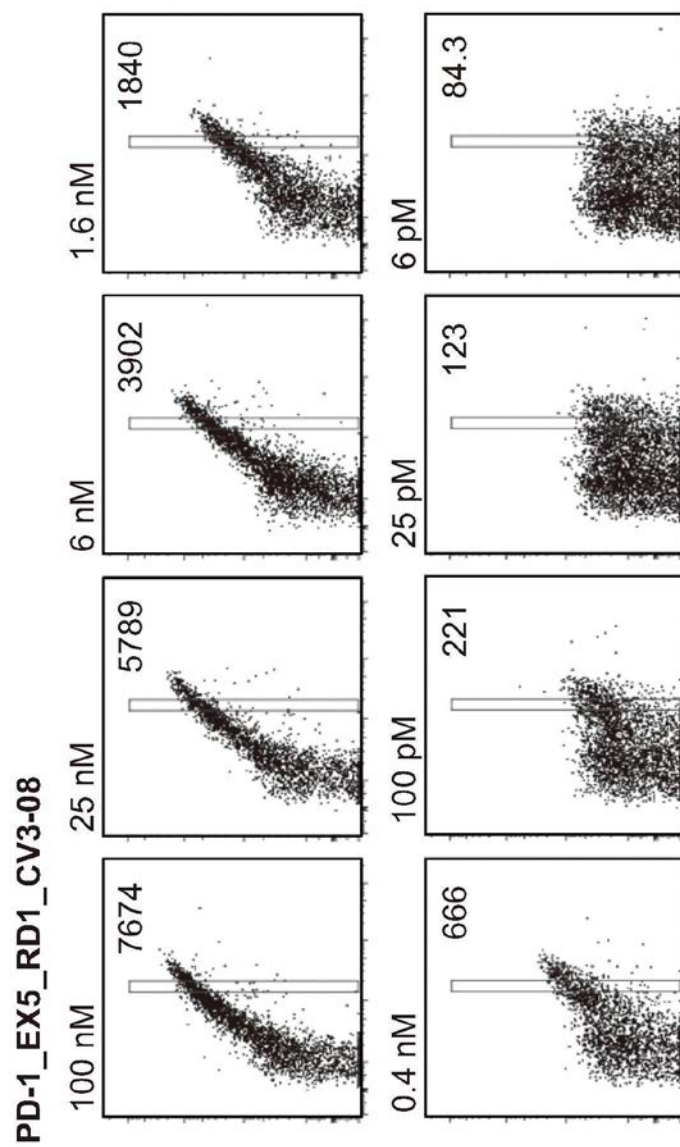


图3B

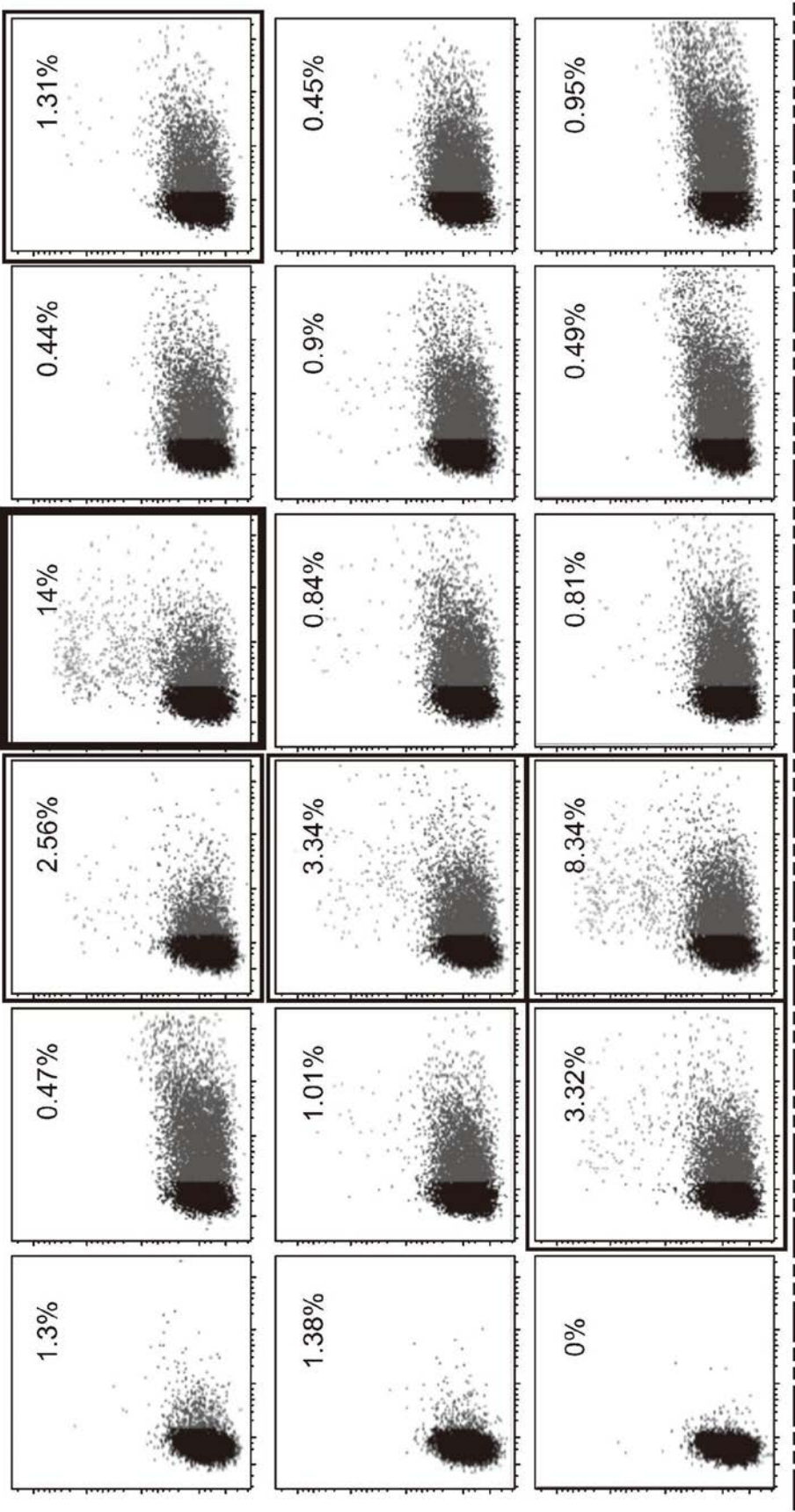


图4

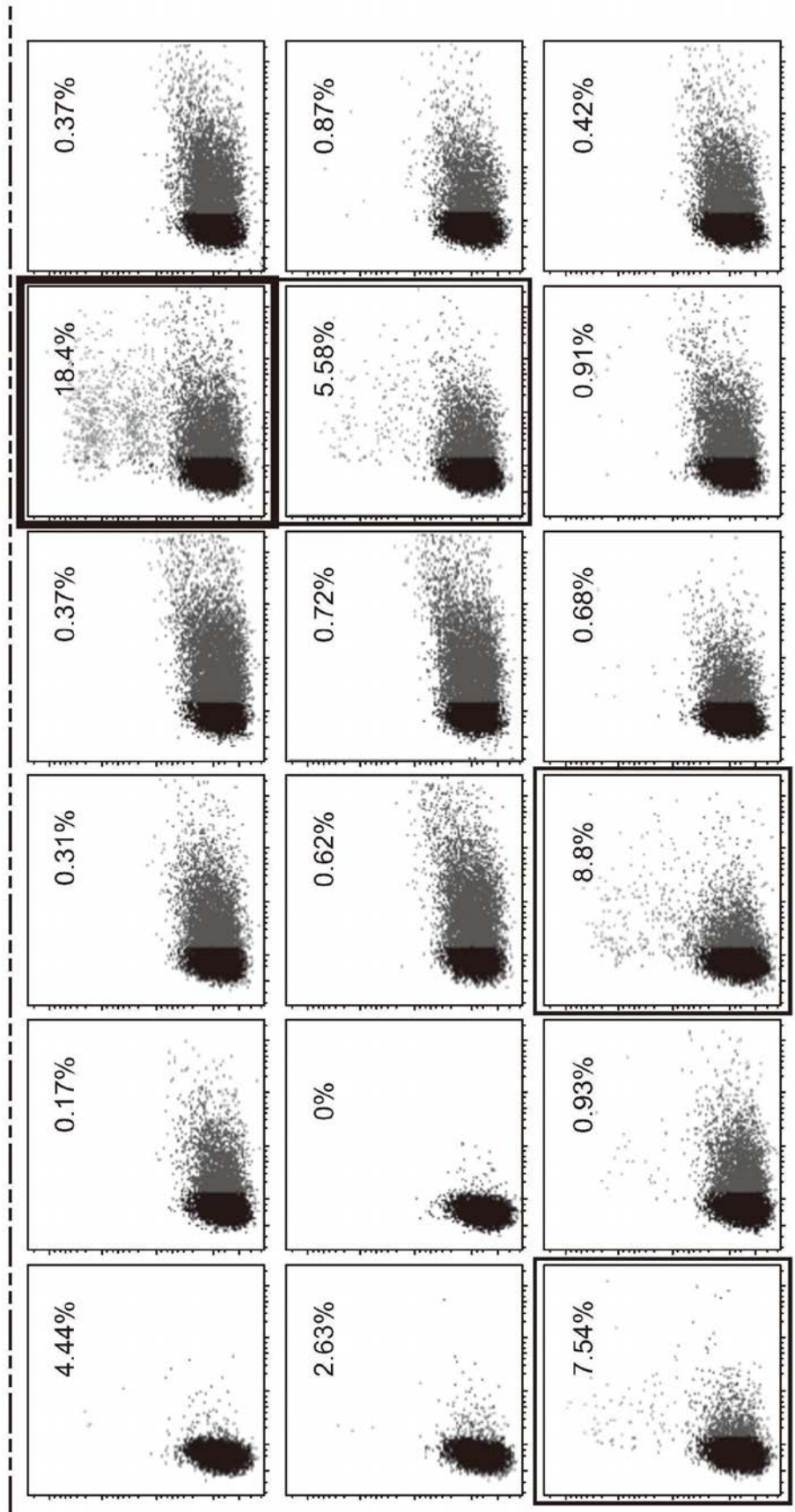


图4(续)

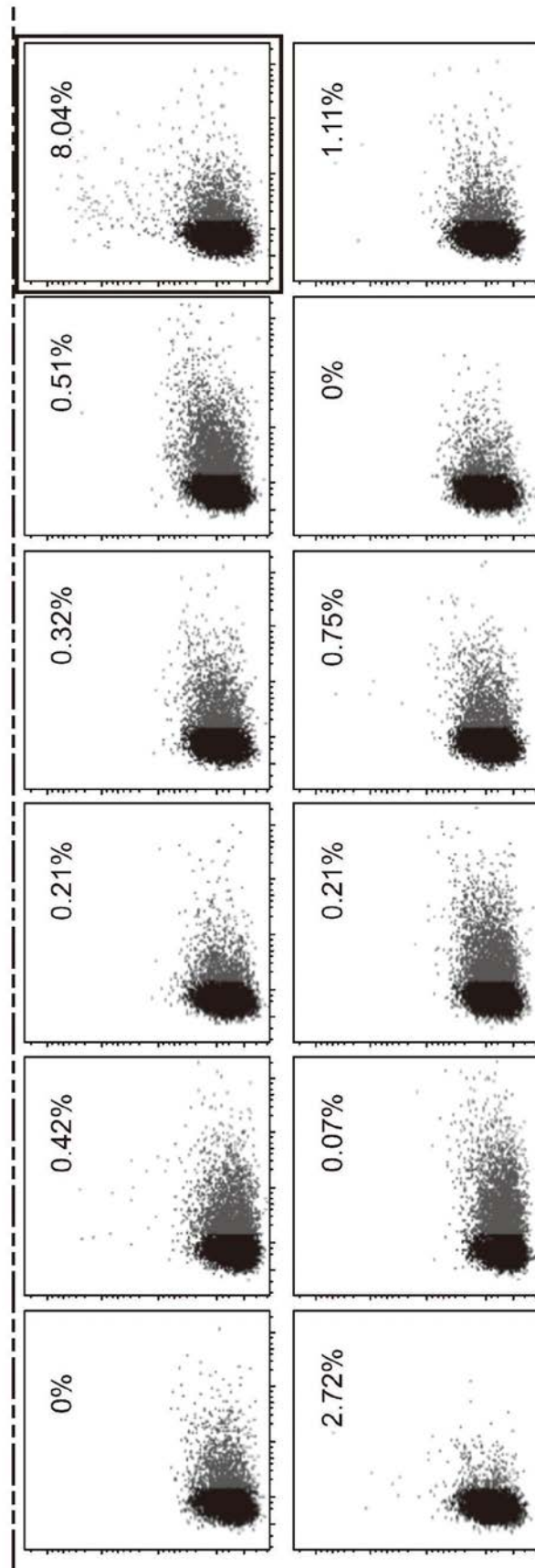


图4(续)

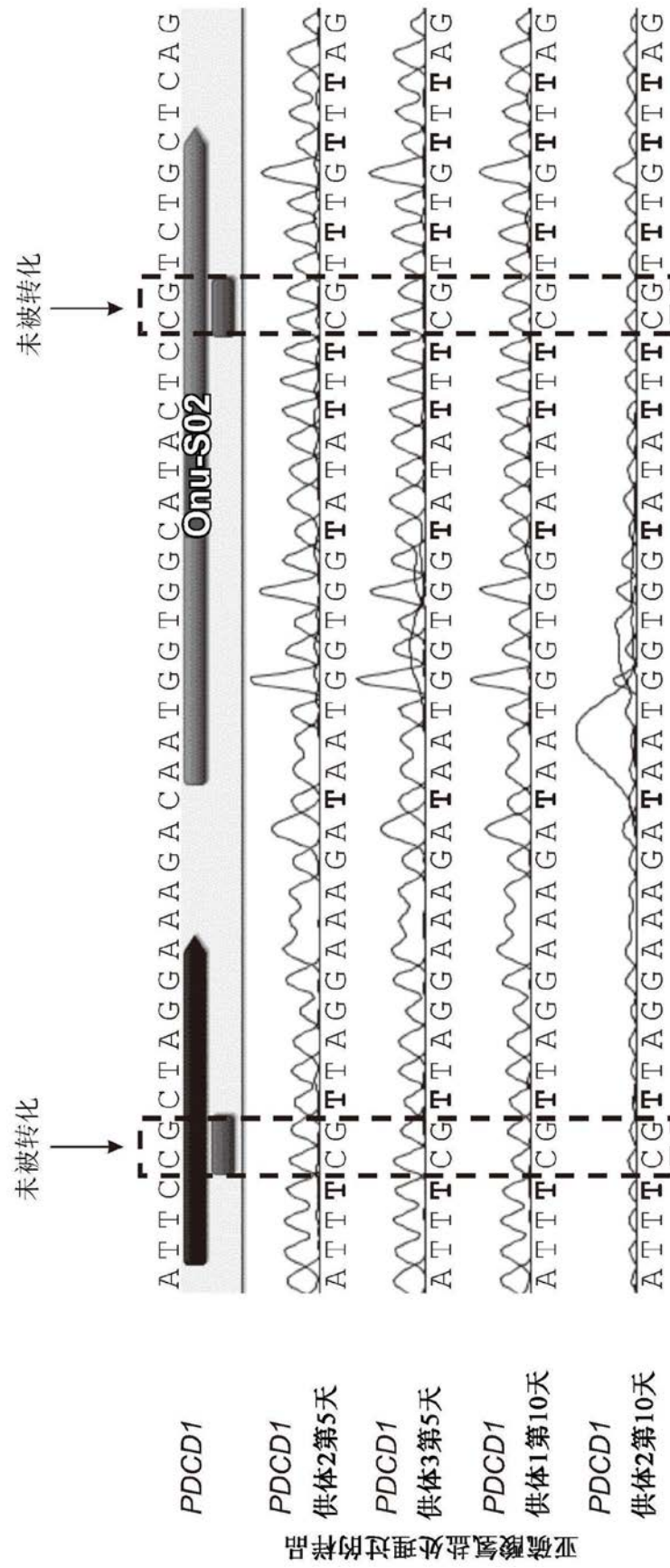


图5

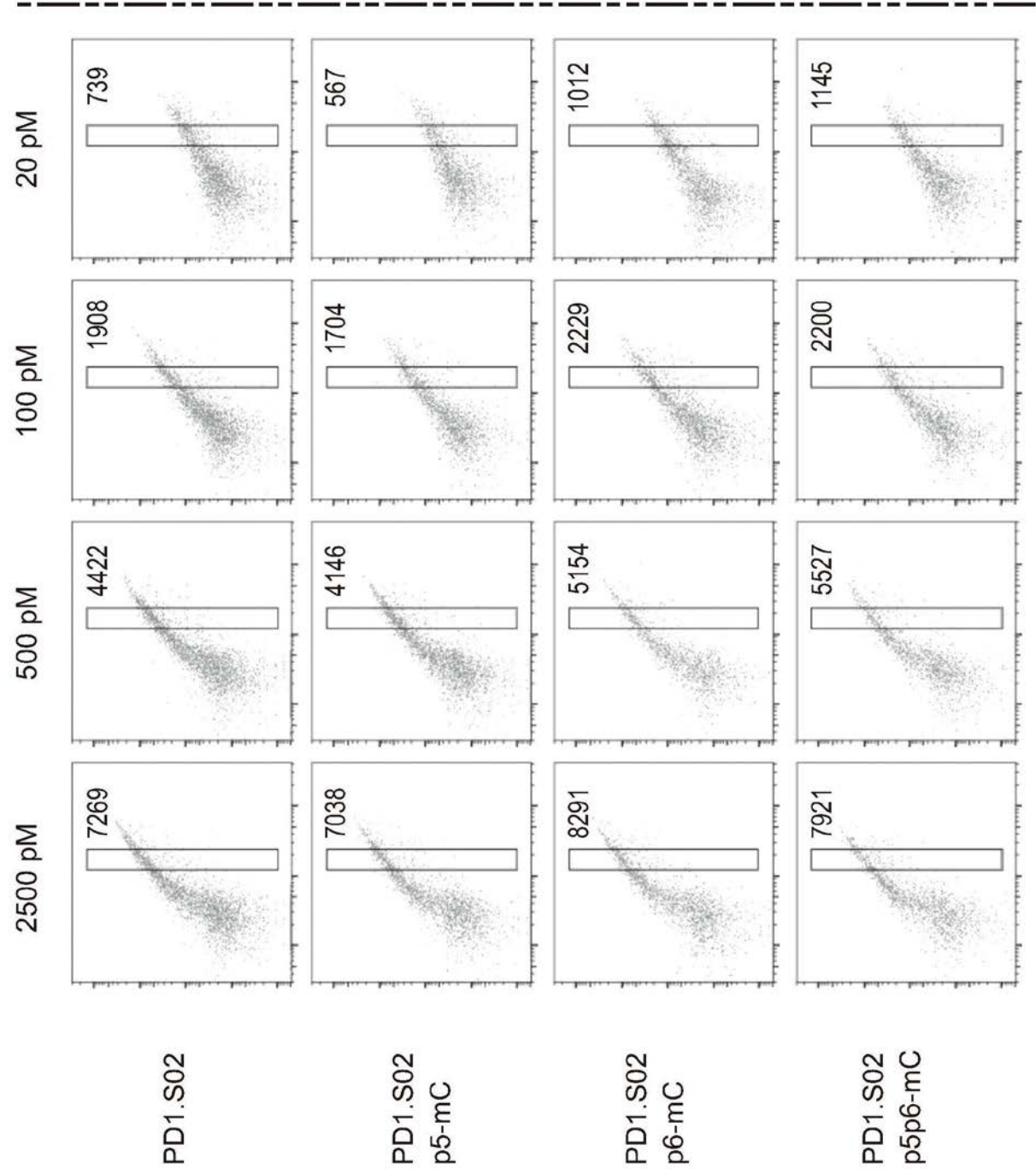


图6

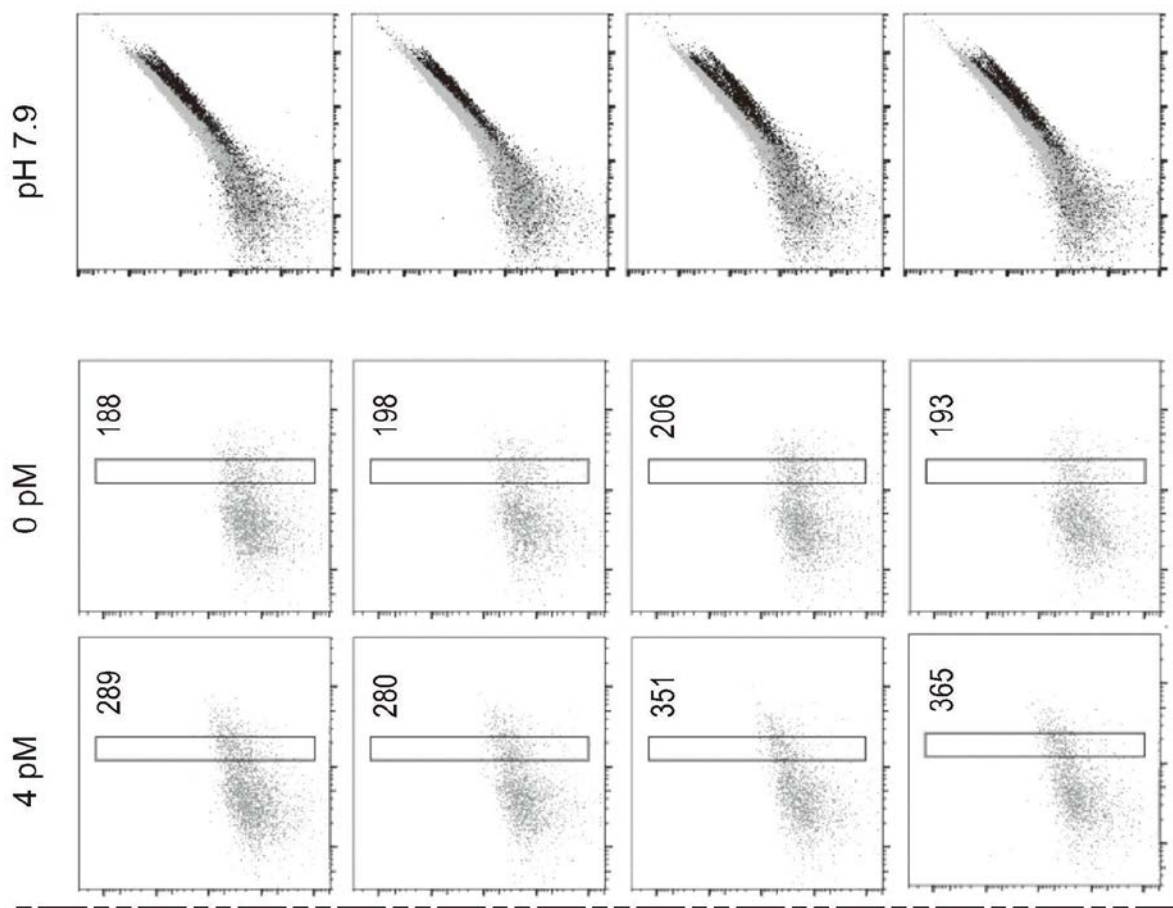


图6(续)

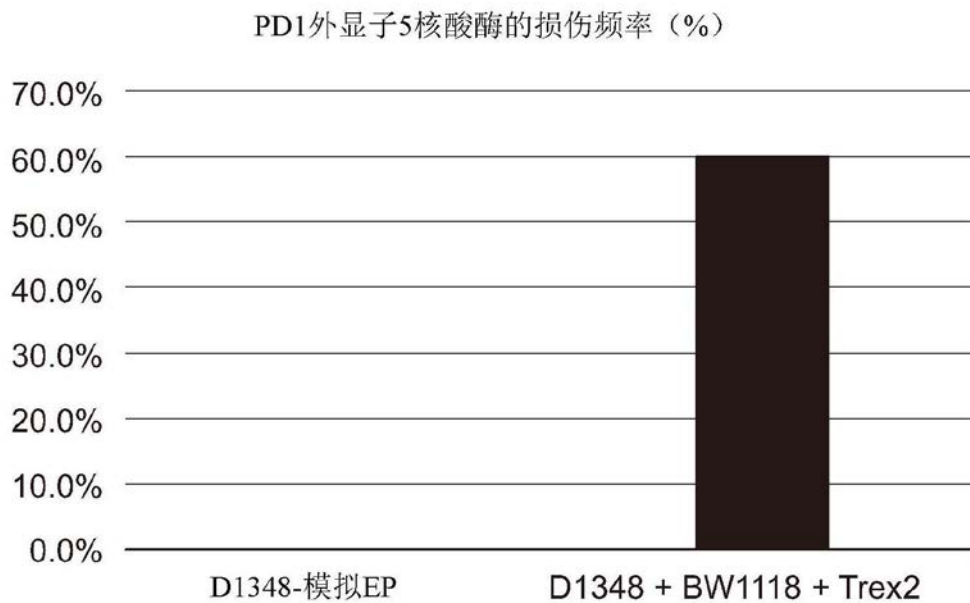


图7A

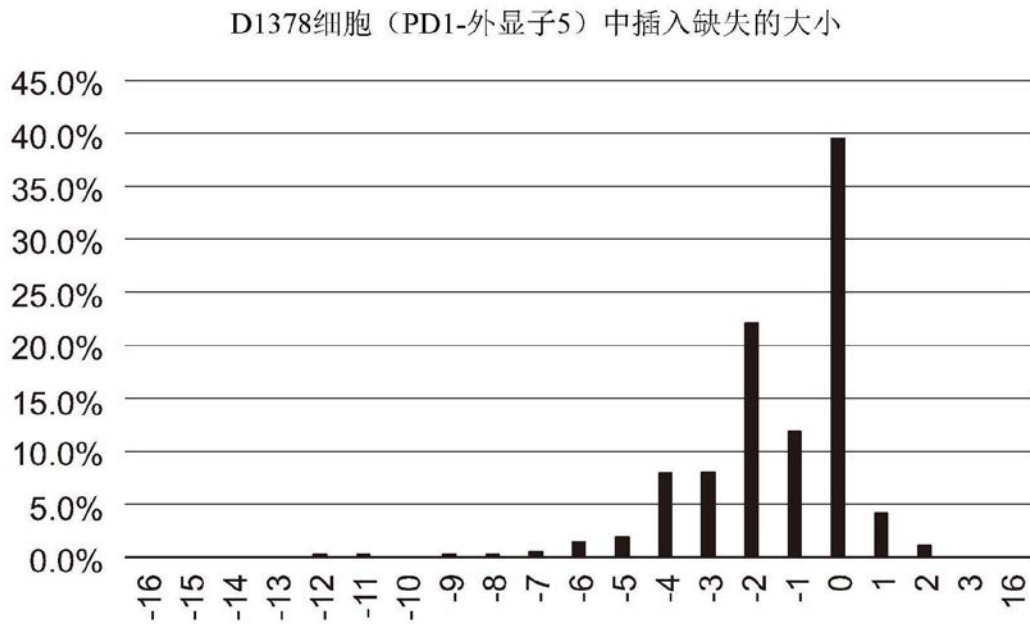


图7B

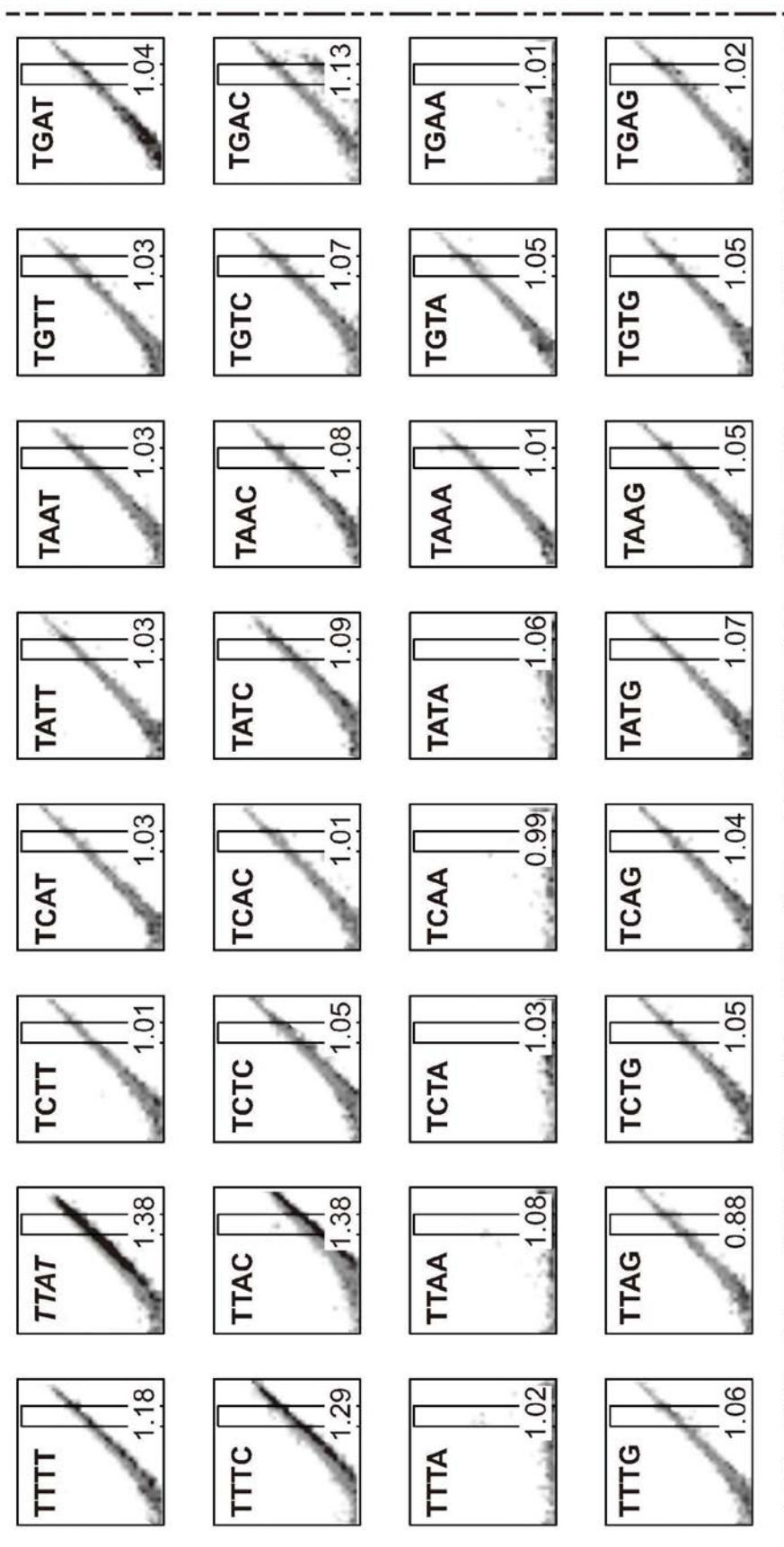


图8

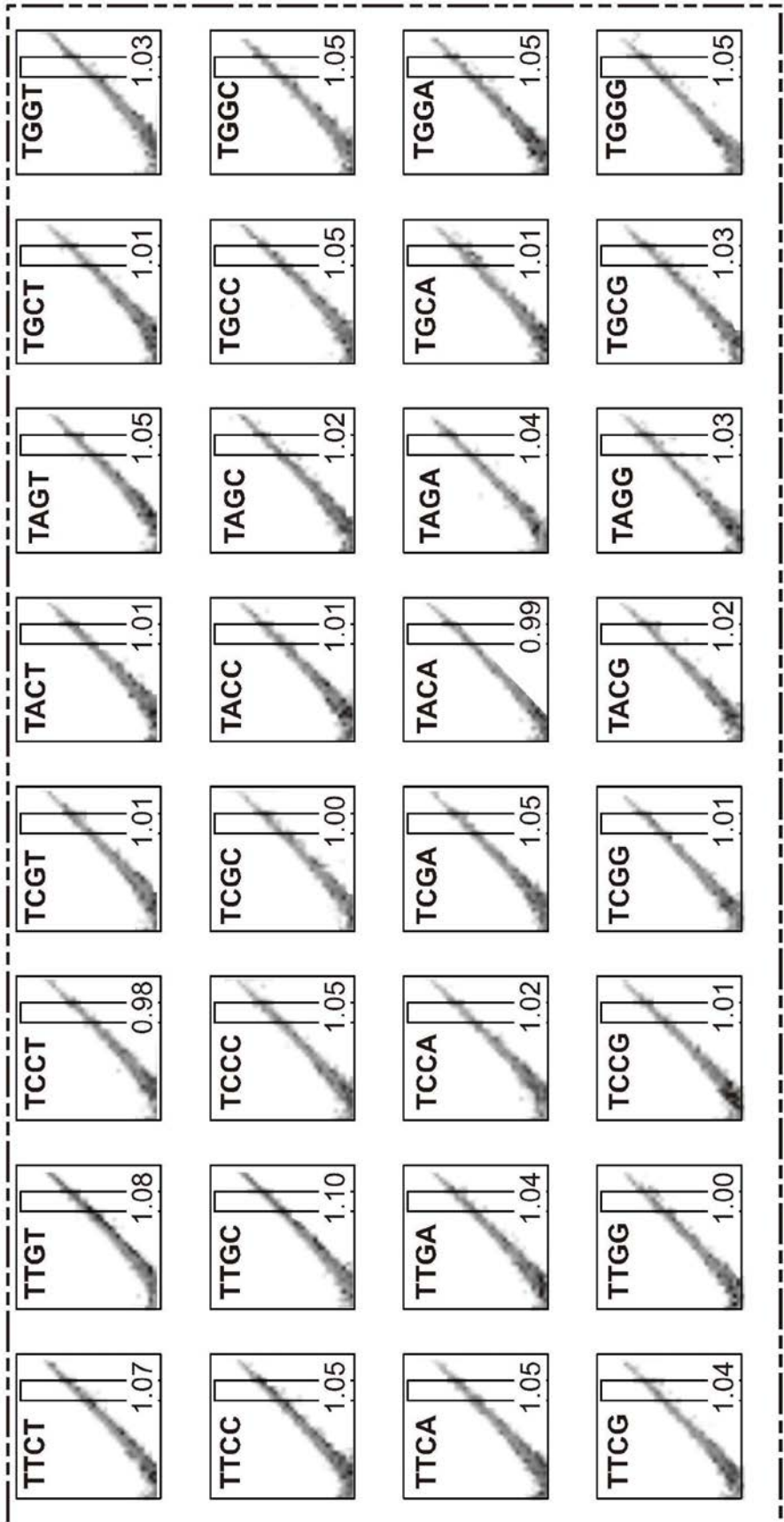


图8(续)

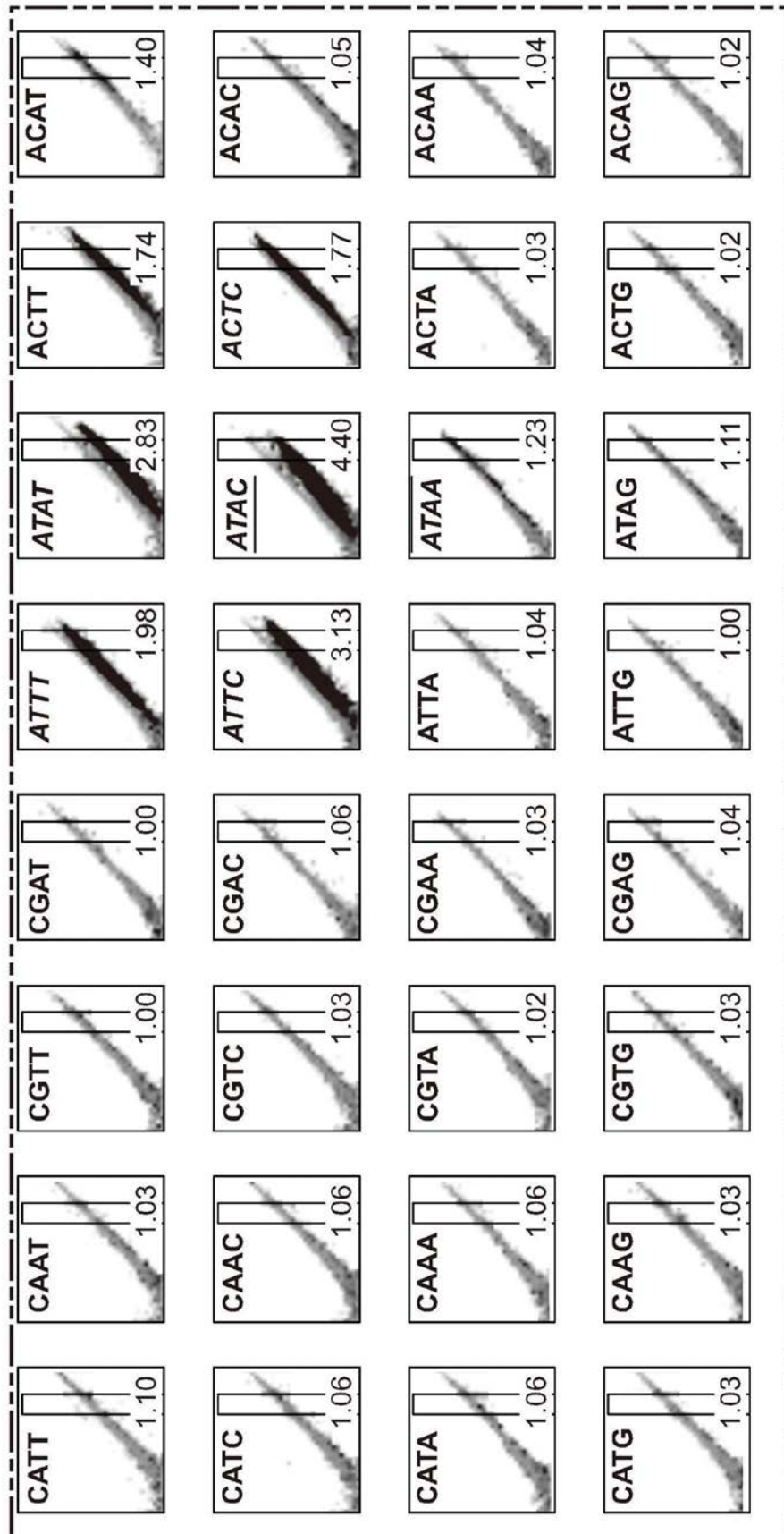


图8(续)

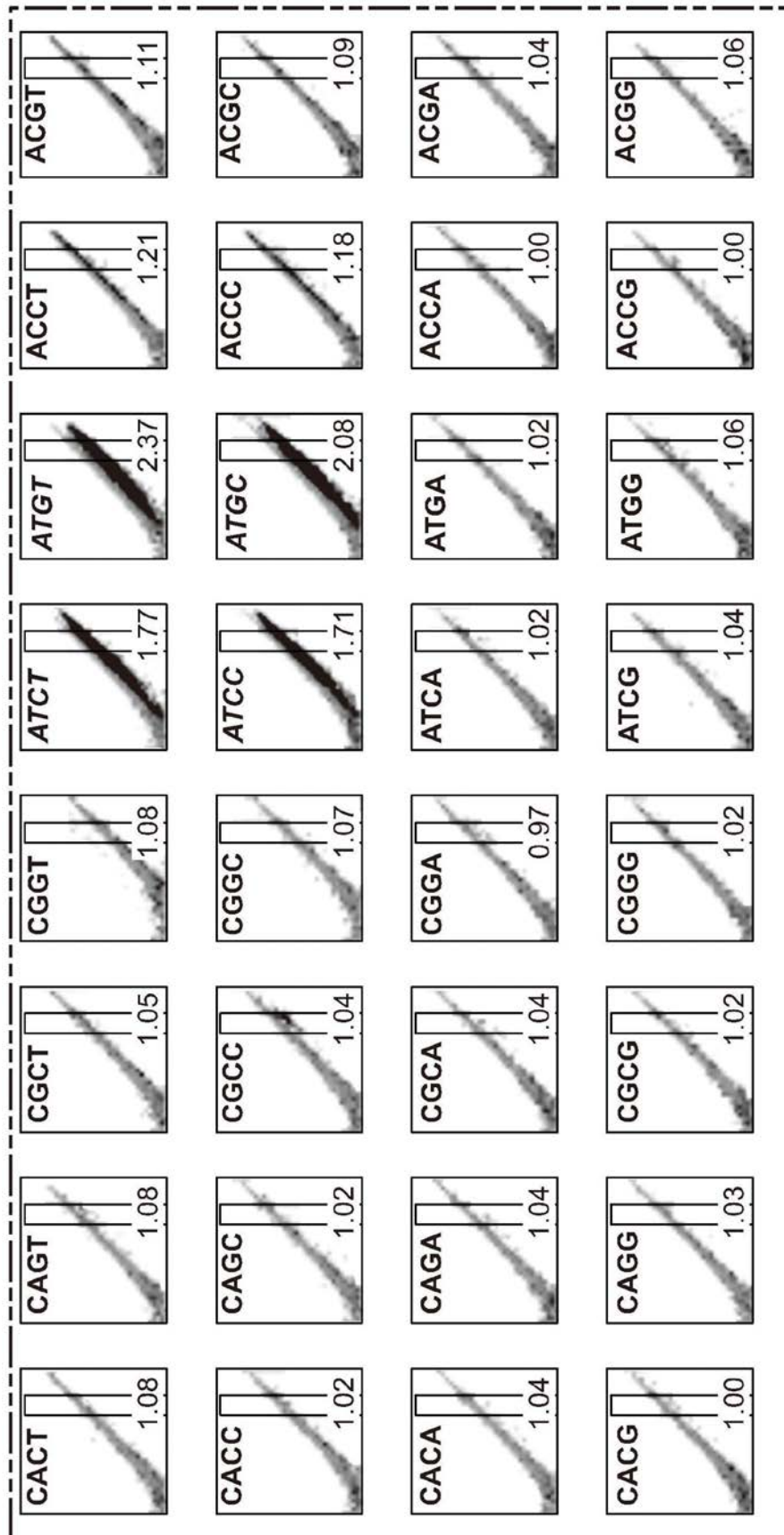


图8(续)

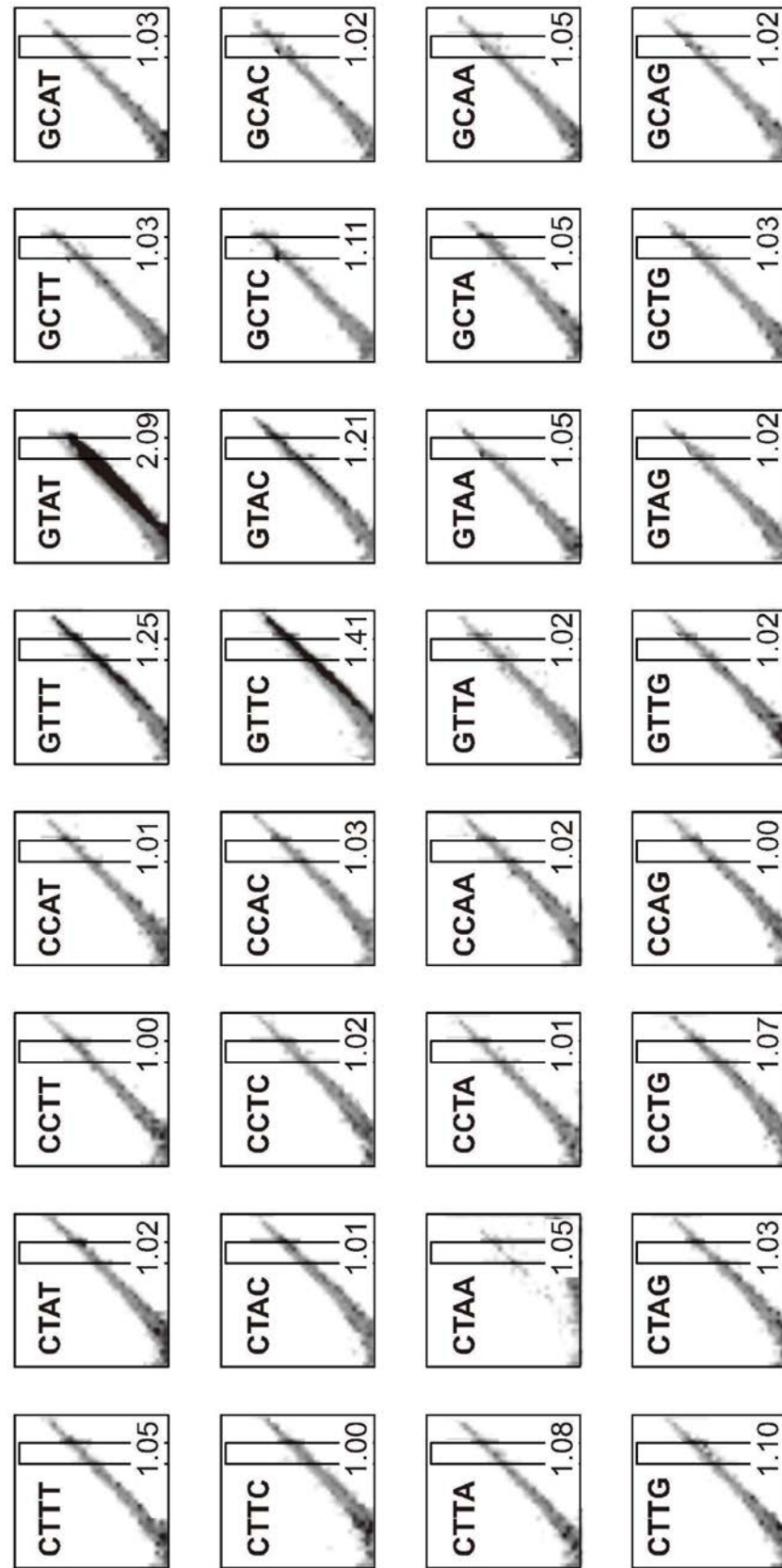


图8(续)

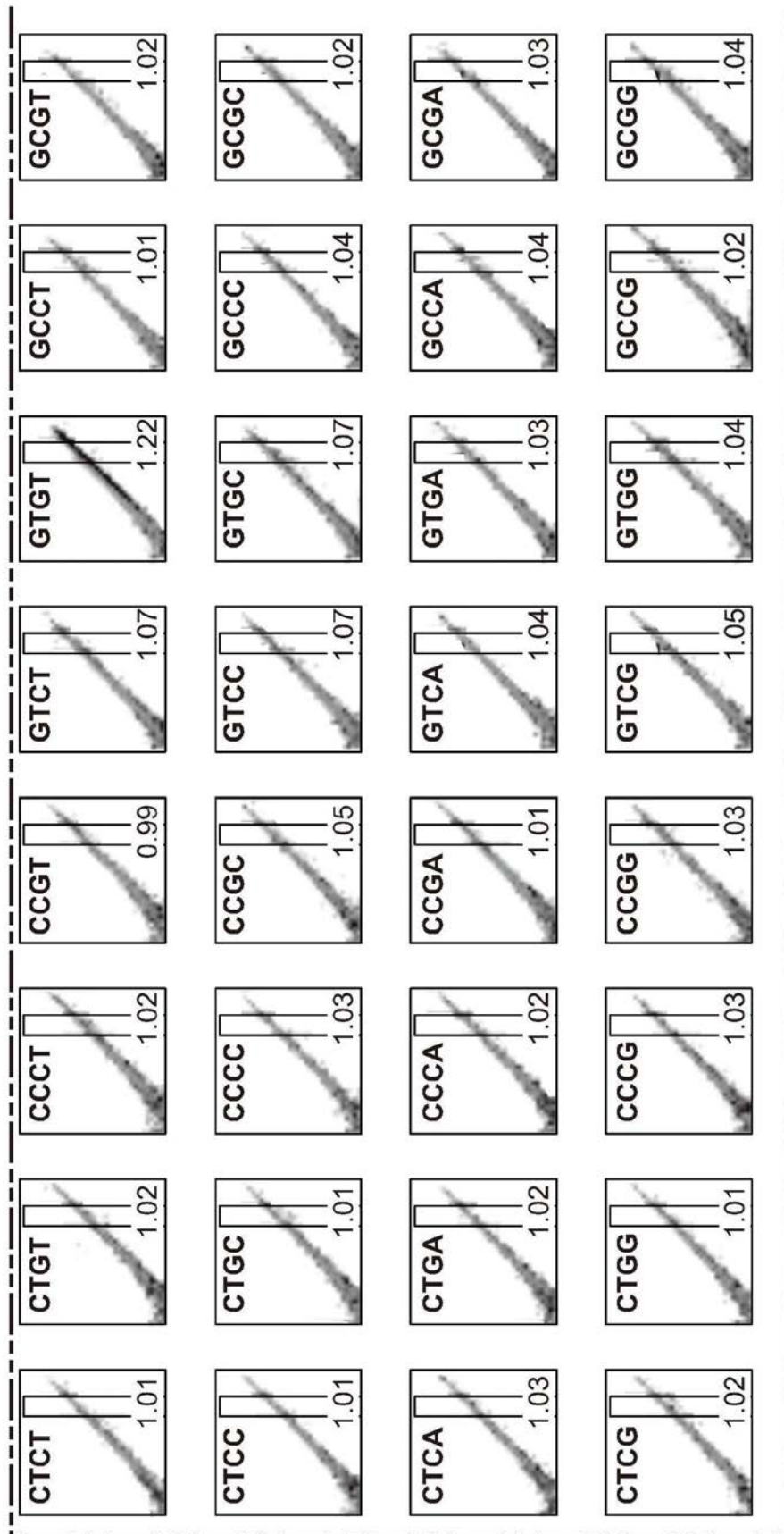


图8(续)

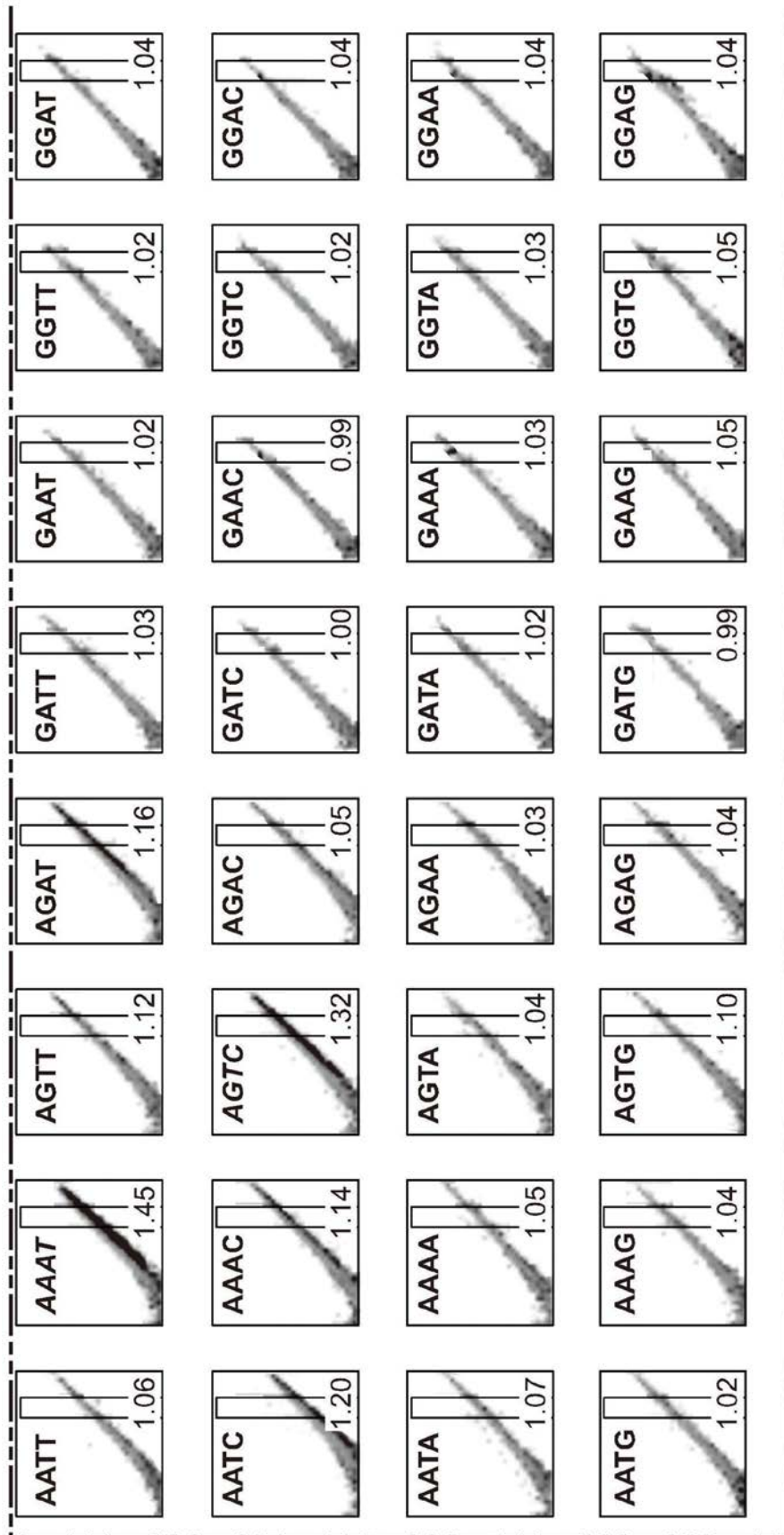


图8(续)

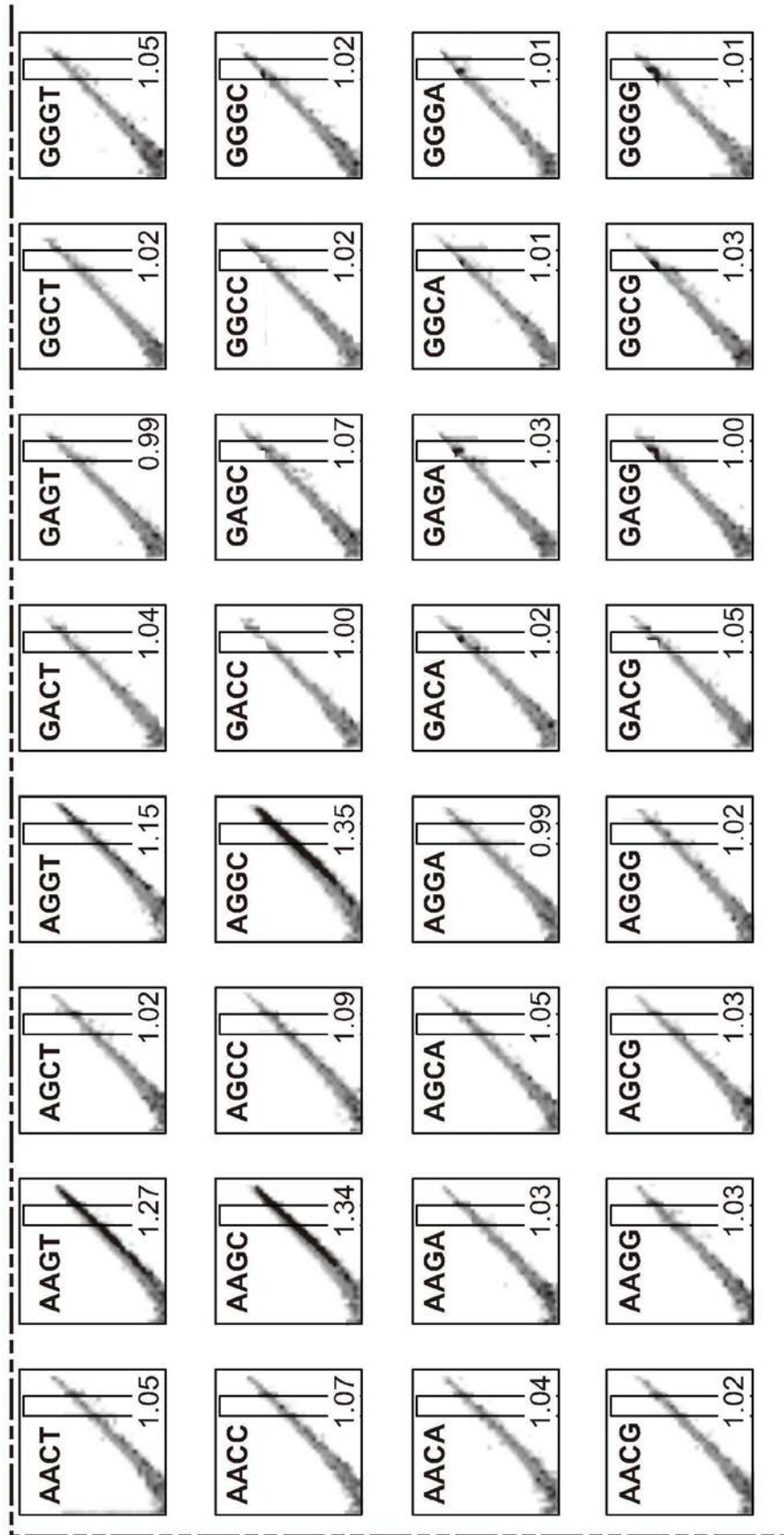


图8(续)

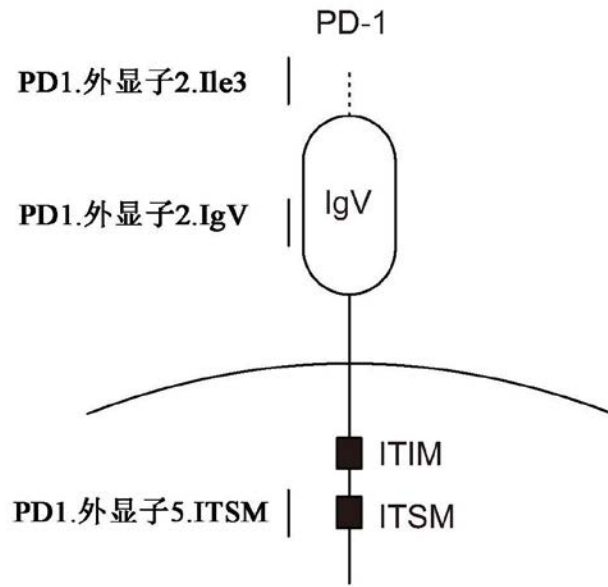


图9A

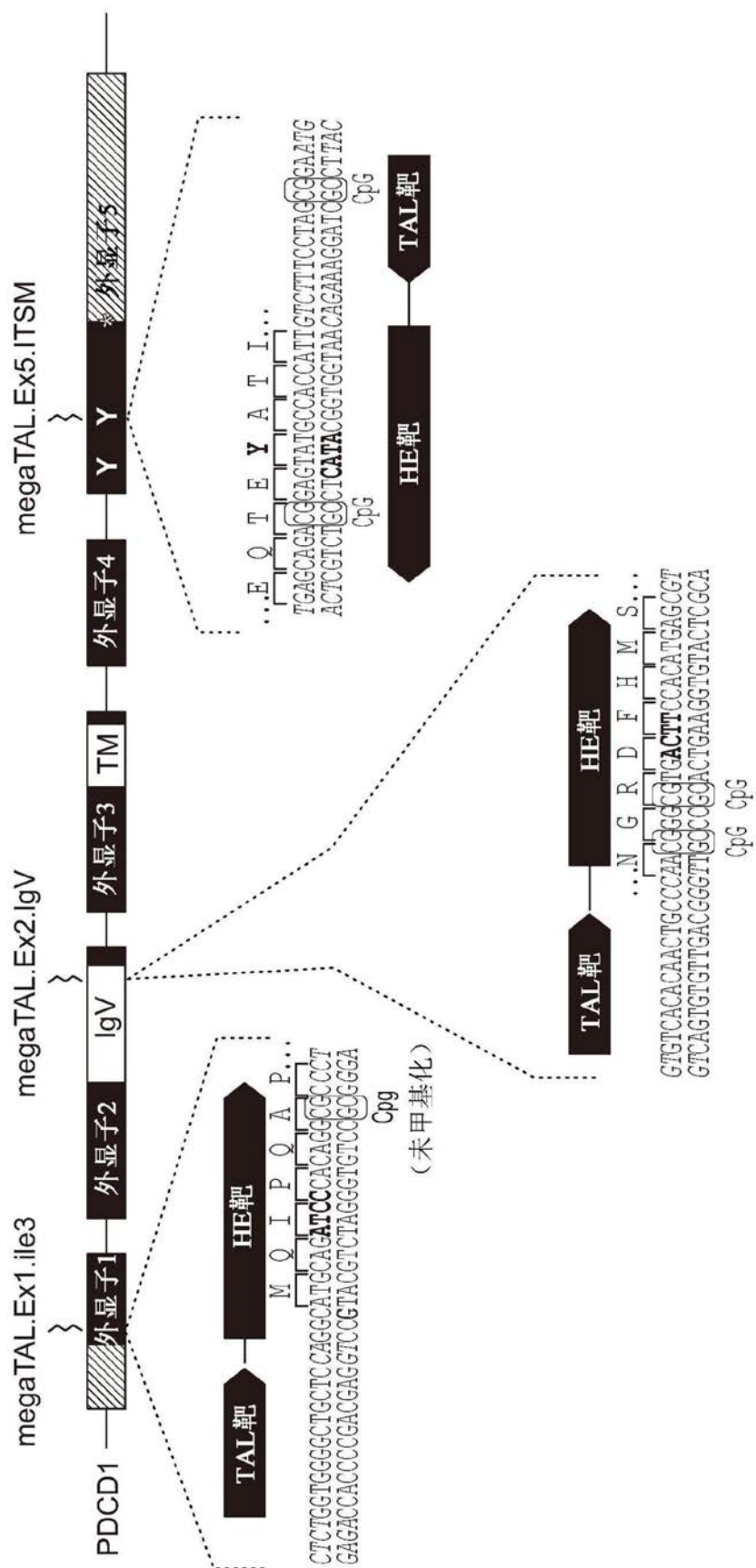


图9B

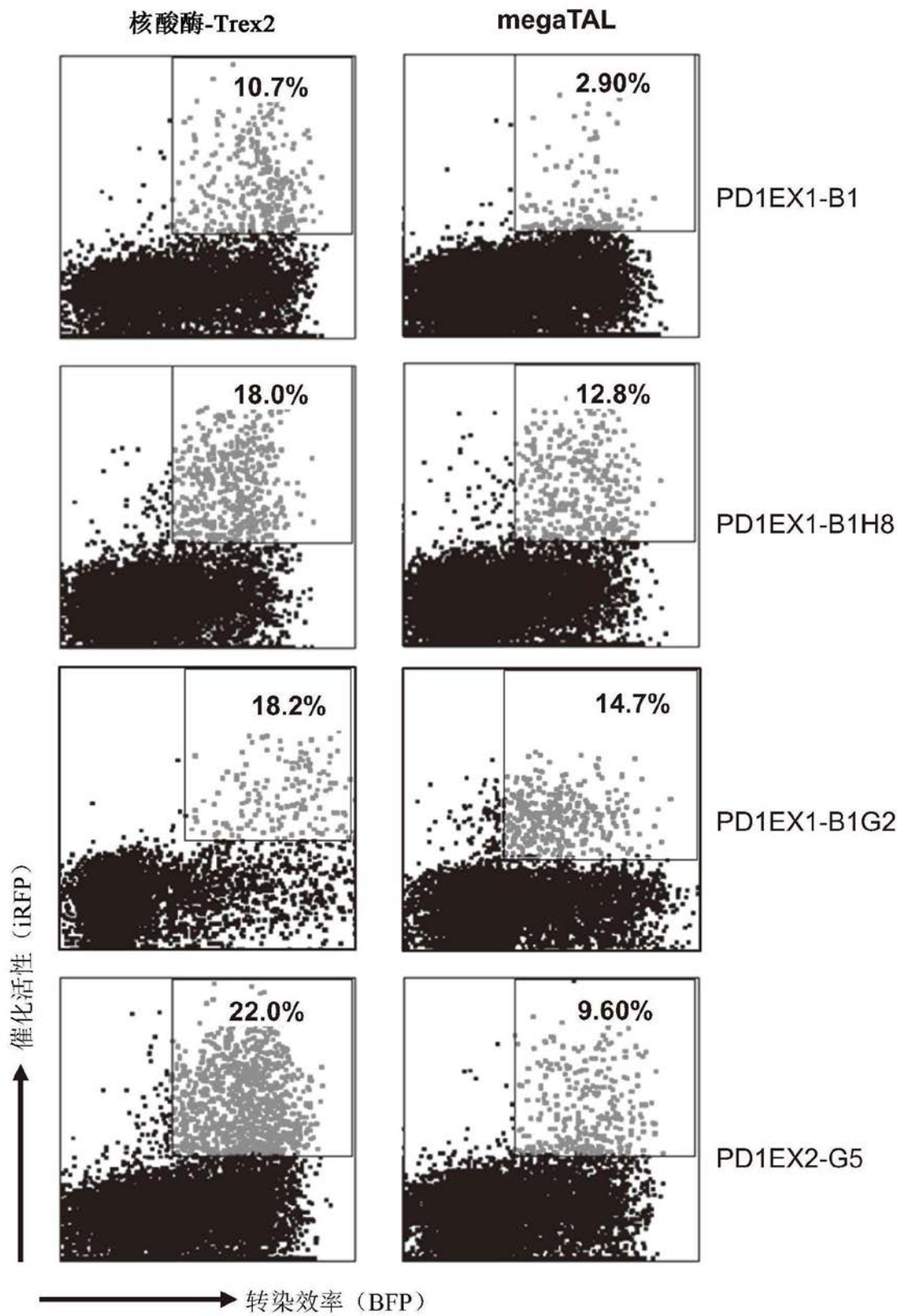


图10

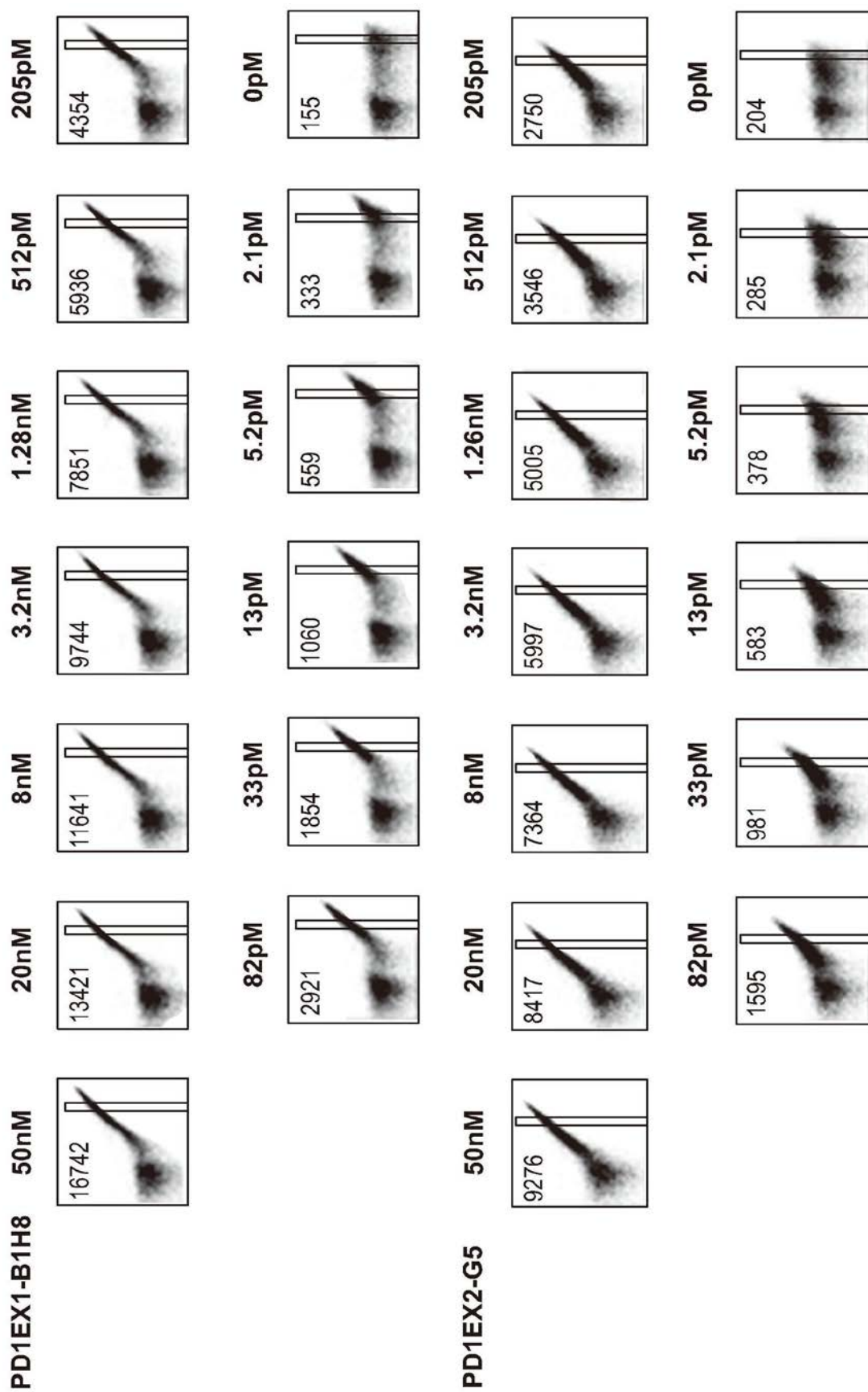


图11A

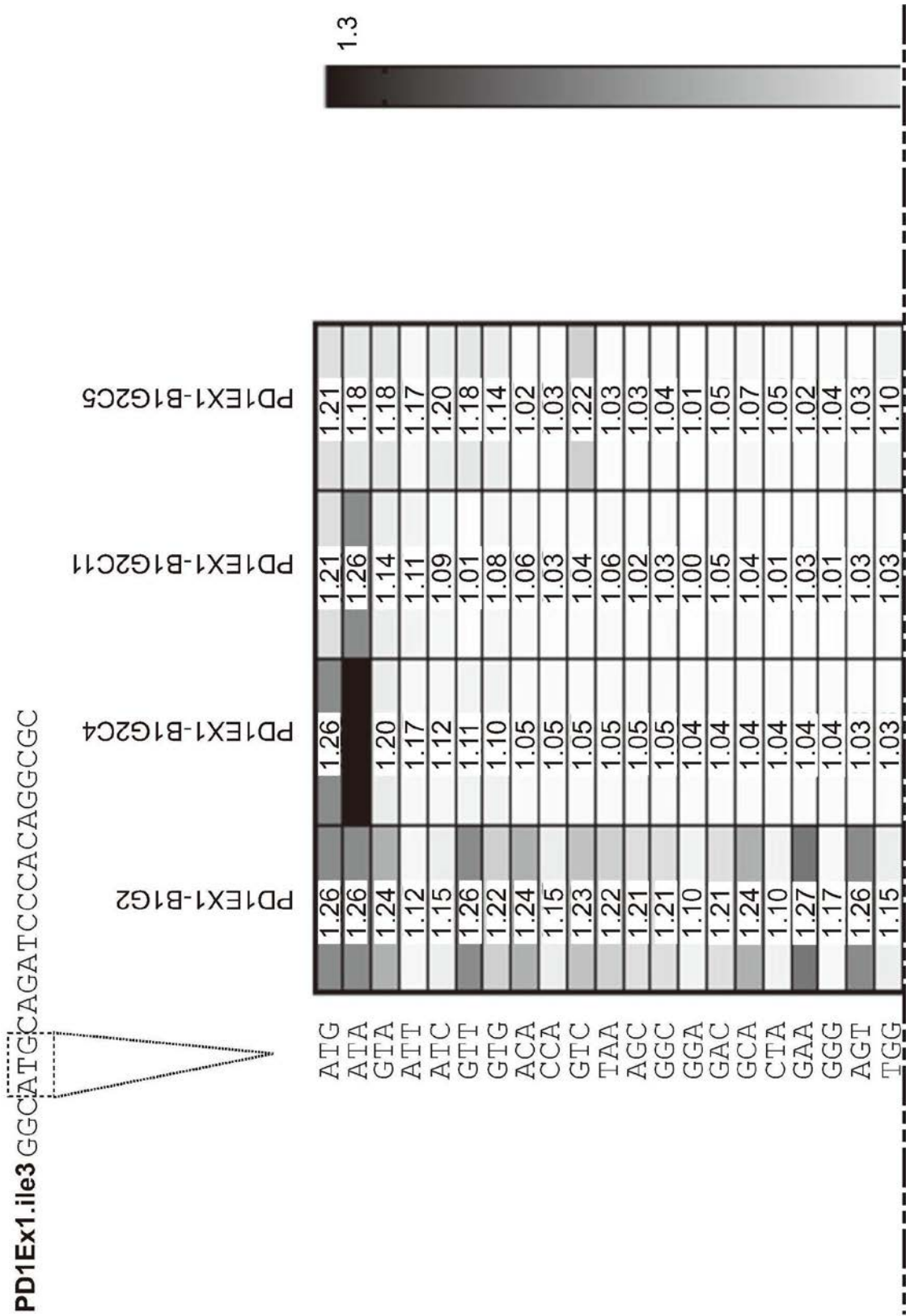


图11B

TTG	1.07	1.03	1.03	1.03	1.11	1.2
ACG	1.25	1.03	1.03	1.01	1.01	
TCA	1.24	1.03	1.03	1.07	1.06	1.1
TTC	1.06	1.03	1.03	1.00	1.05	
CGC	1.13	1.03	1.03	1.03	1.01	
GAT	1.25	1.03	1.03	1.03	1.05	
TCG	1.22	1.03	1.03	0.99	1.07	
CGA	1.19	1.03	1.03	1.00	1.00	
CTG	1.05	1.03	1.03	1.02	1.04	
CGG	1.15	1.02	1.02	1.04	1.03	
GGT	1.21	1.02	1.02	1.03	1.04	
AGA	1.21	1.02	1.02	1.04	1.02	
GAG	1.25	1.02	1.02	1.00	1.00	
CGT	1.12	1.02	1.02	1.01	1.04	
AAG	1.25	1.02	1.02	1.04	1.04	
TAC	1.23	1.02	1.02	1.03	1.06	
CCG	1.07	1.02	1.02	1.03	1.01	
TTT	1.02	1.02	1.02	1.01	1.03	
AGG	1.28	1.02	1.02	1.02	1.02	
TAT	1.21	1.02	1.02	1.02	1.01	
GCC	1.05	1.02	1.02	1.03	1.03	
TTA	1.10	1.02	1.02	1.03	1.05	
CAC	1.10	1.02	1.02	1.01	1.04	
CTC	1.04	1.02	1.02	1.04	1.05	
TCT	1.19	1.02	1.02	1.05	1.06	
TCC	1.19	1.01	1.01	1.04	1.14	
GCG	1.23	1.01	1.01	1.00	1.04	
CAG	1.13	1.01	1.01	1.02	1.03	
CCT	1.03	1.01	1.01	1.02	1.01	
CTT	1.02	1.01	1.01	1.01	1.03	

图11B(续)

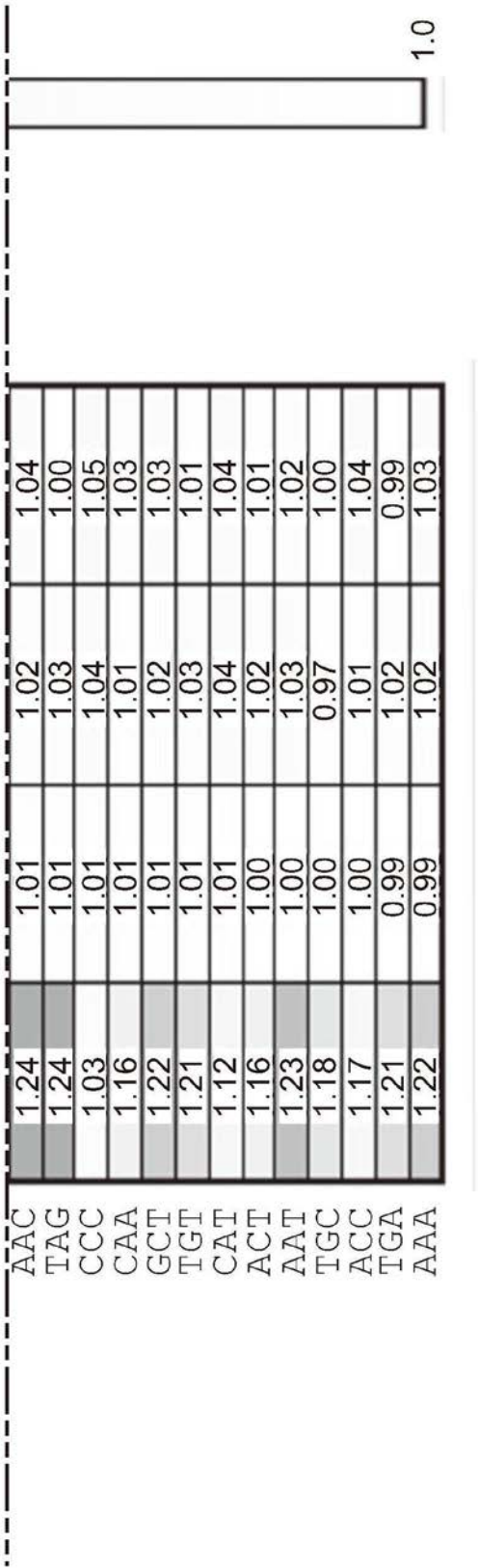


图11B (续)

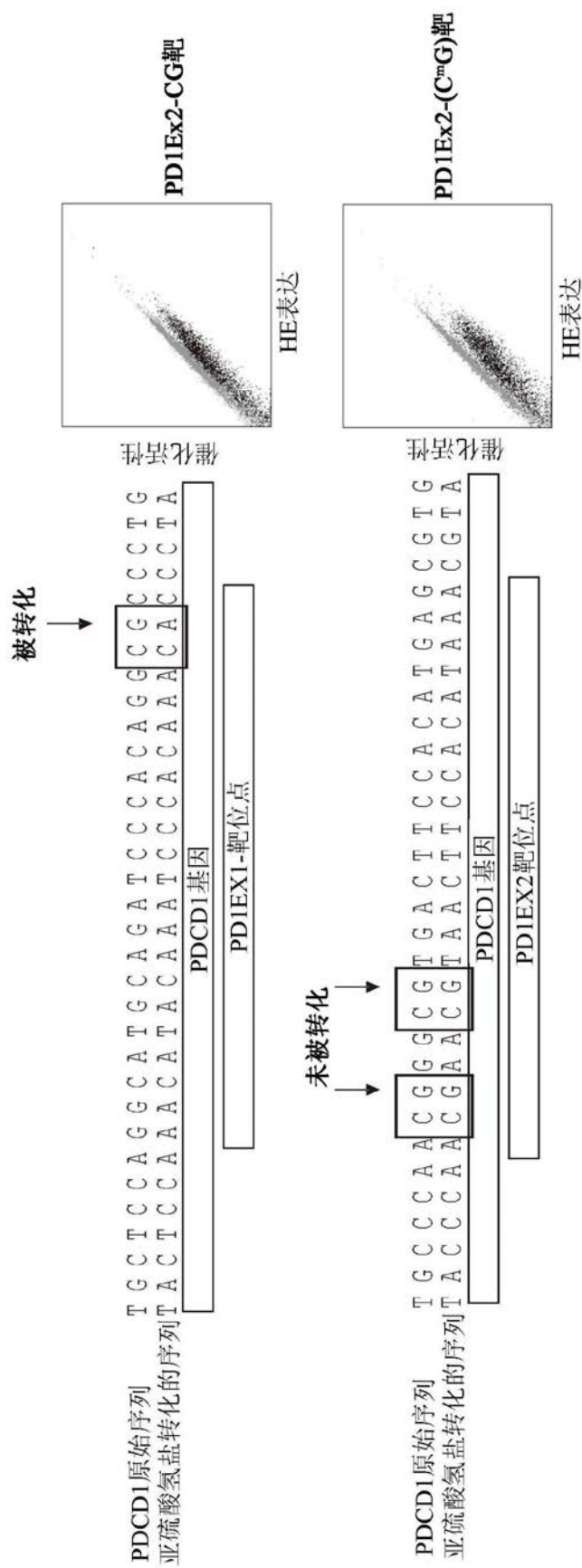


图12

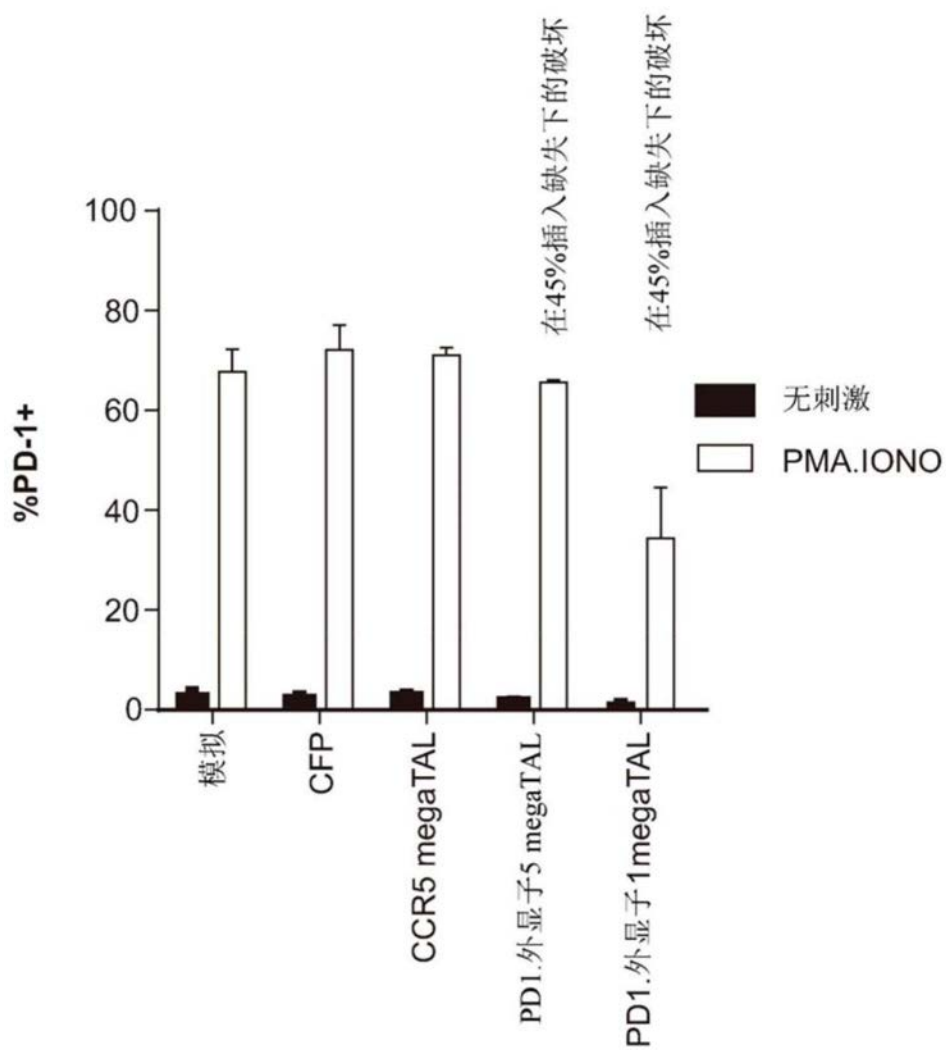


图13A

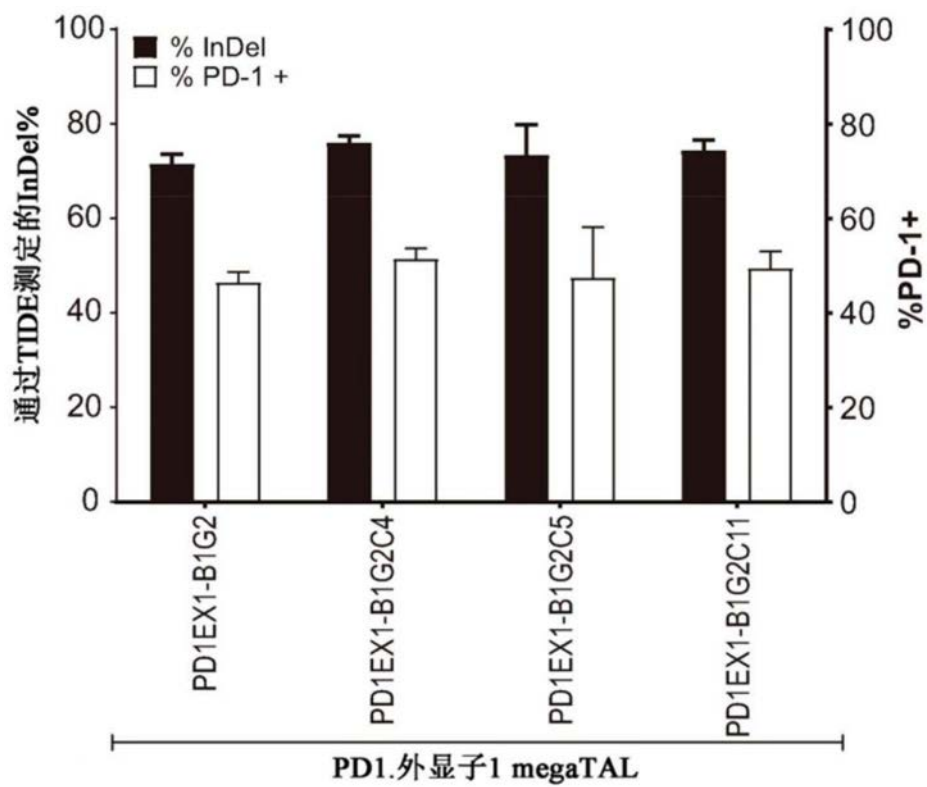


图13B

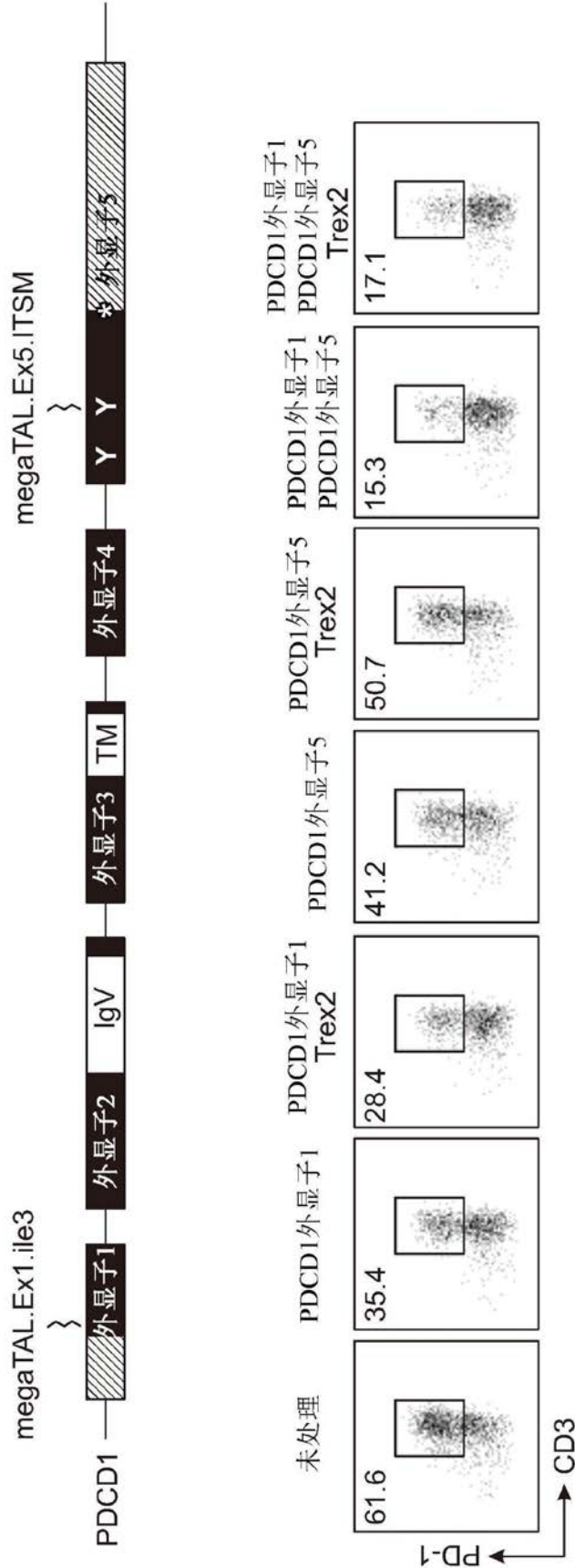


图14

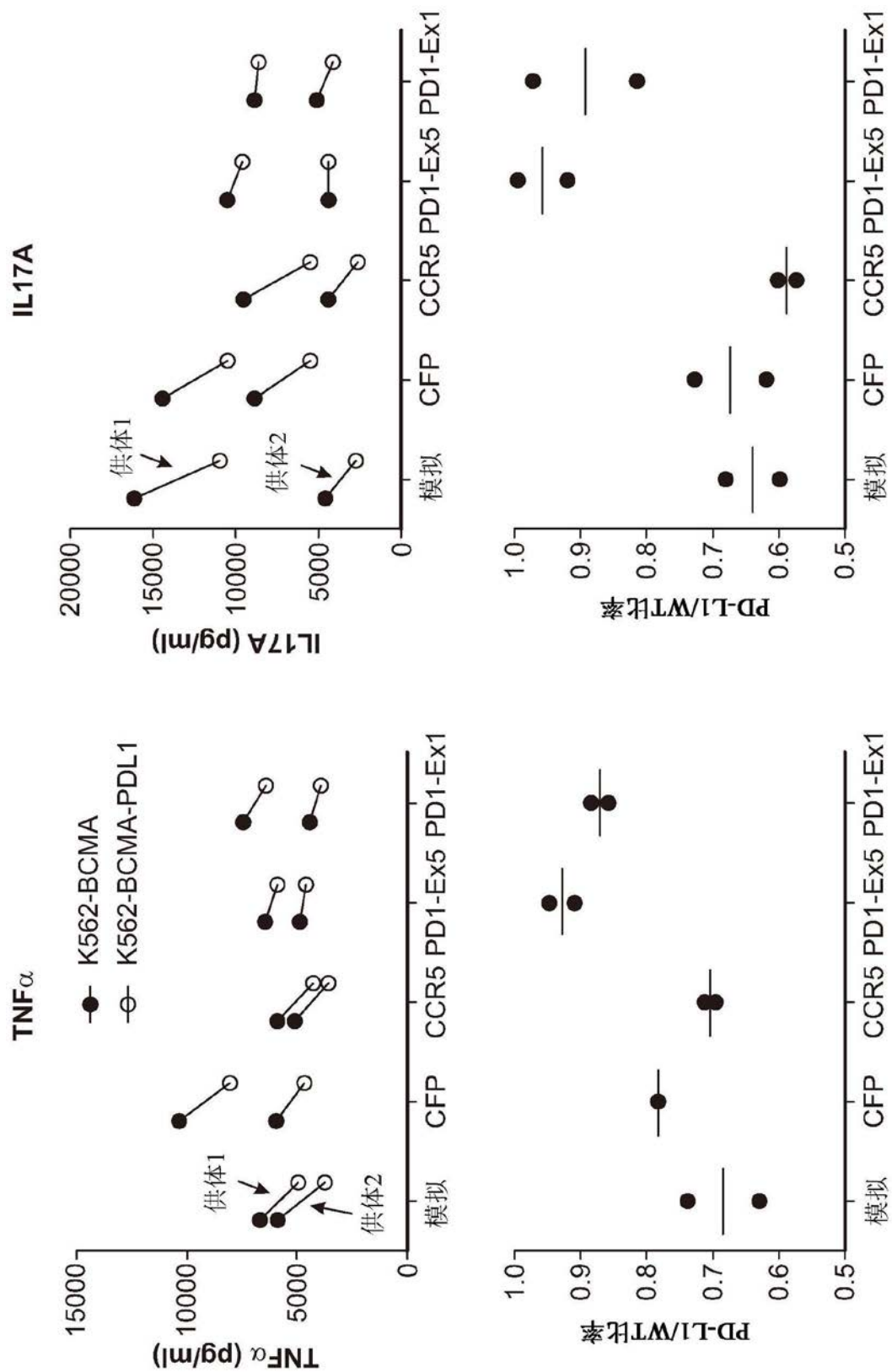


图15A

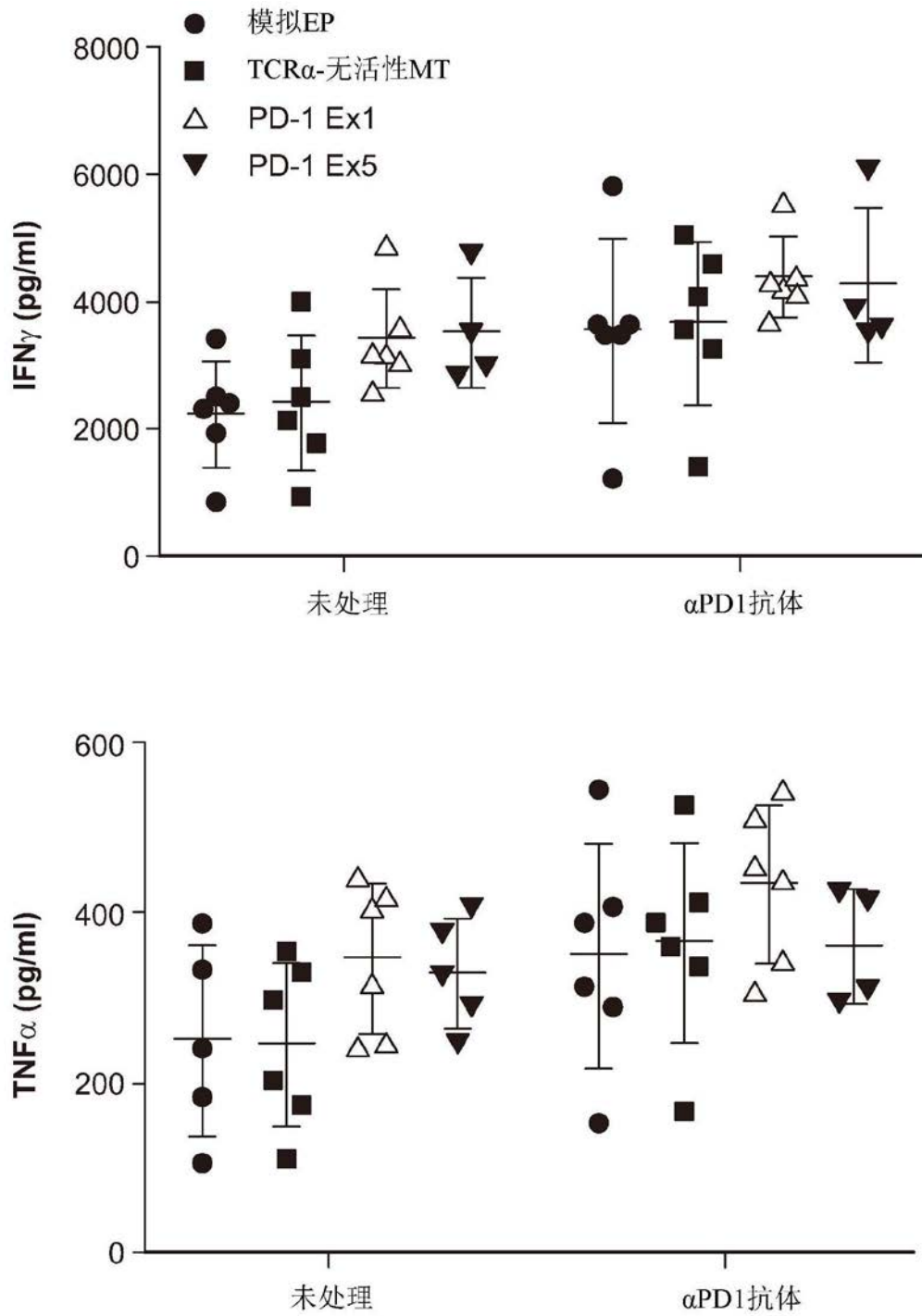


图15B

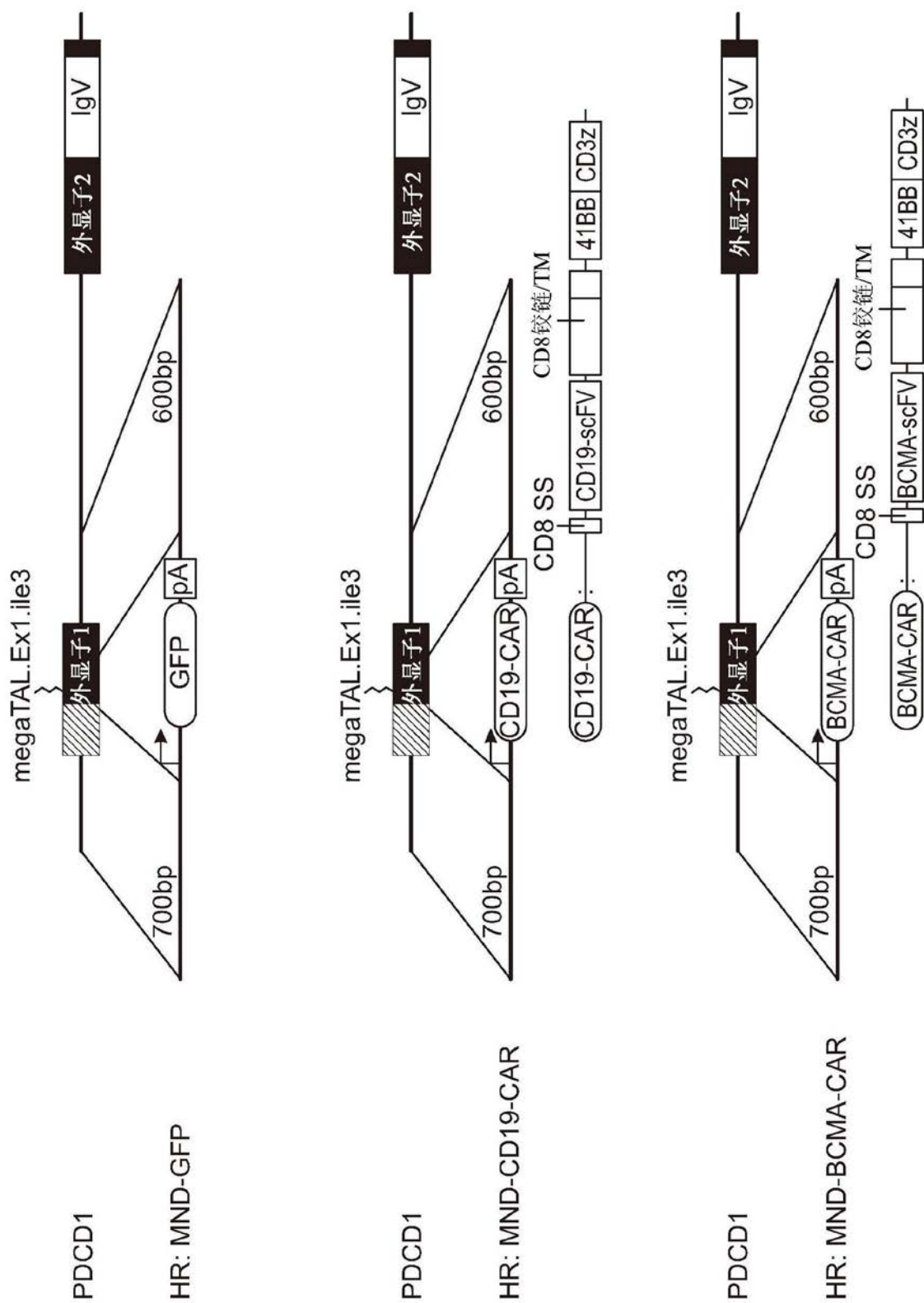


图16A

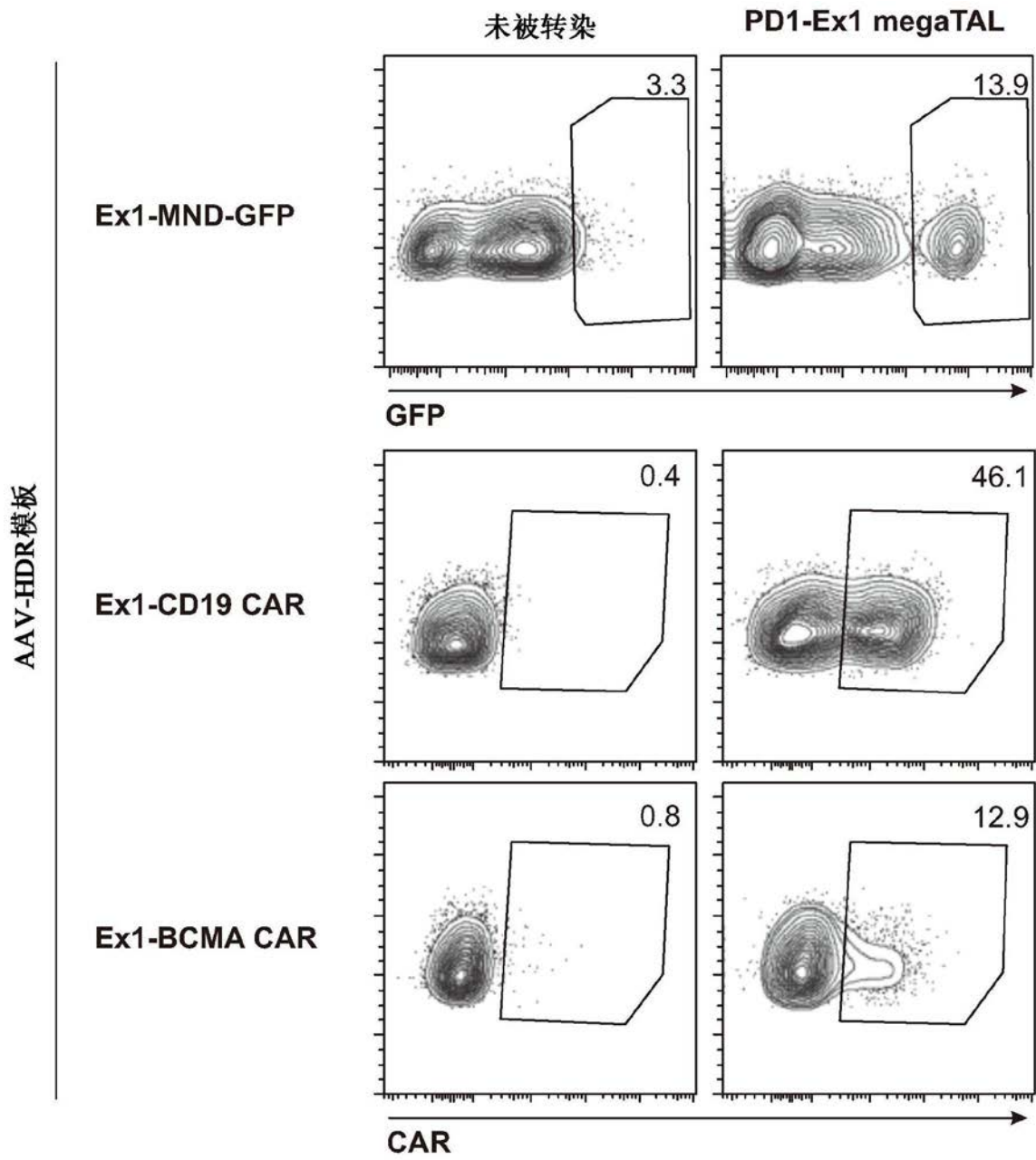


图16B

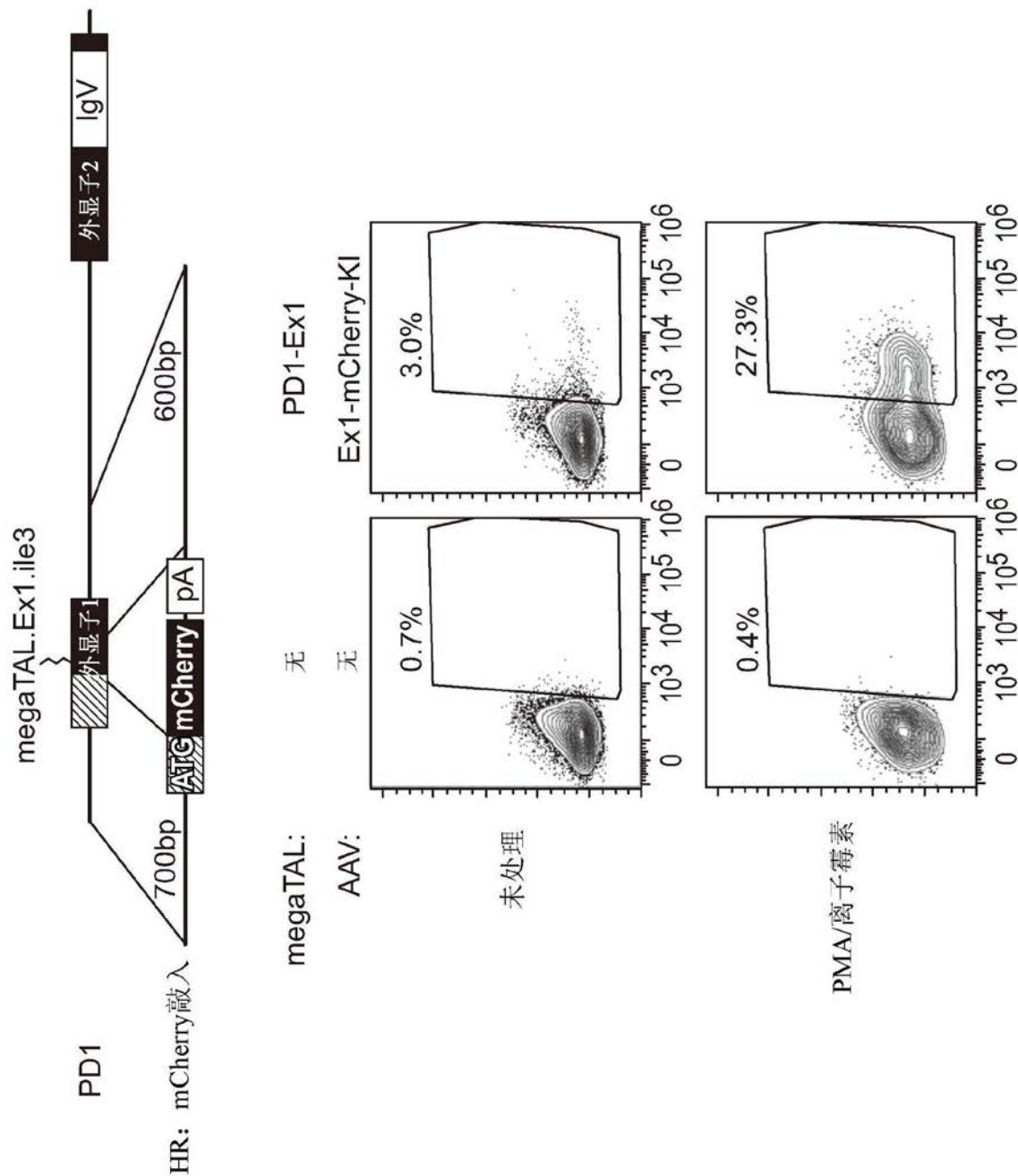


图17

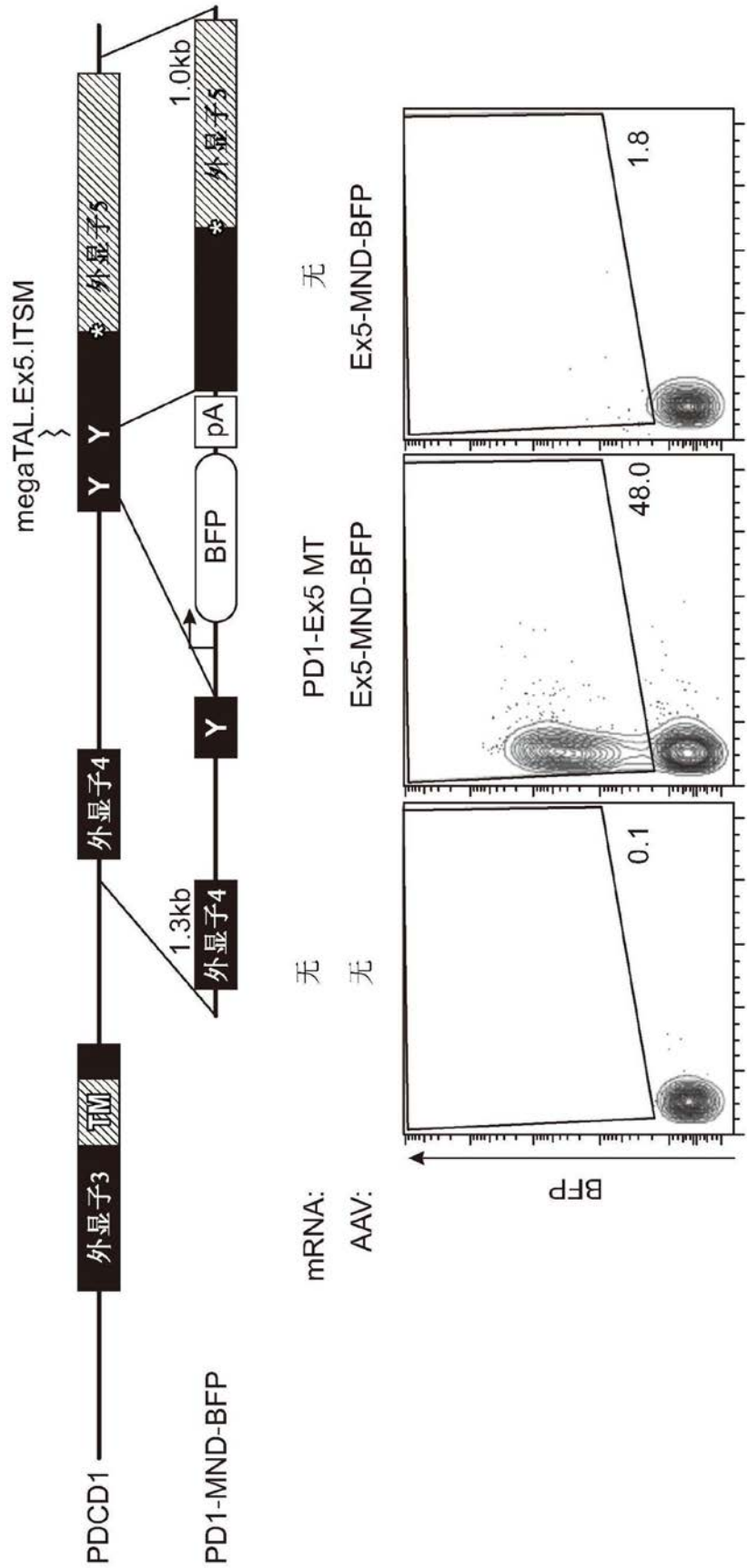


图18

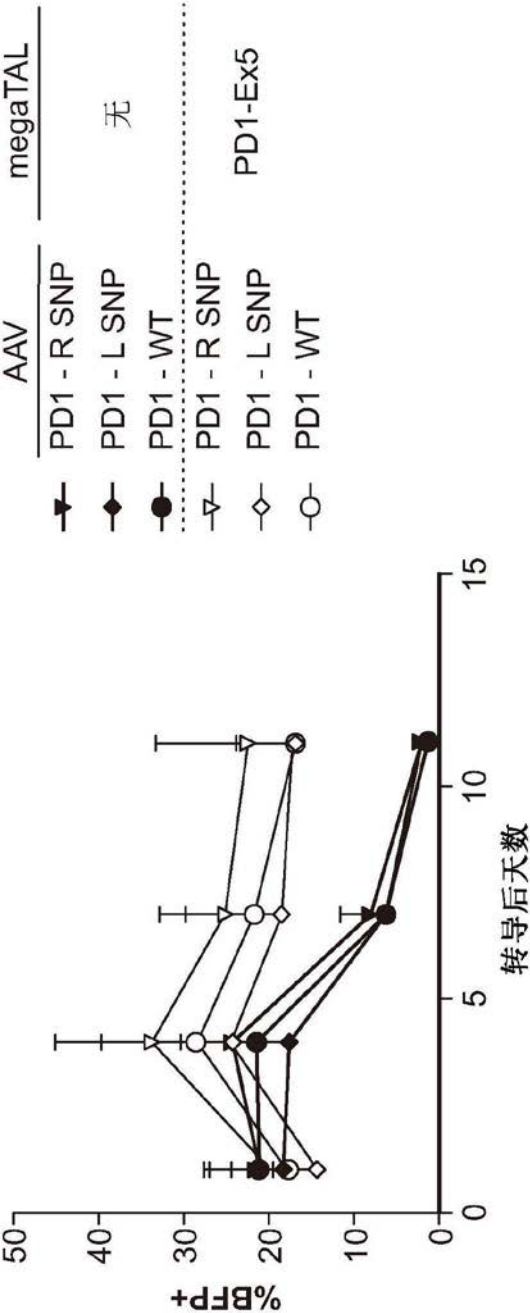
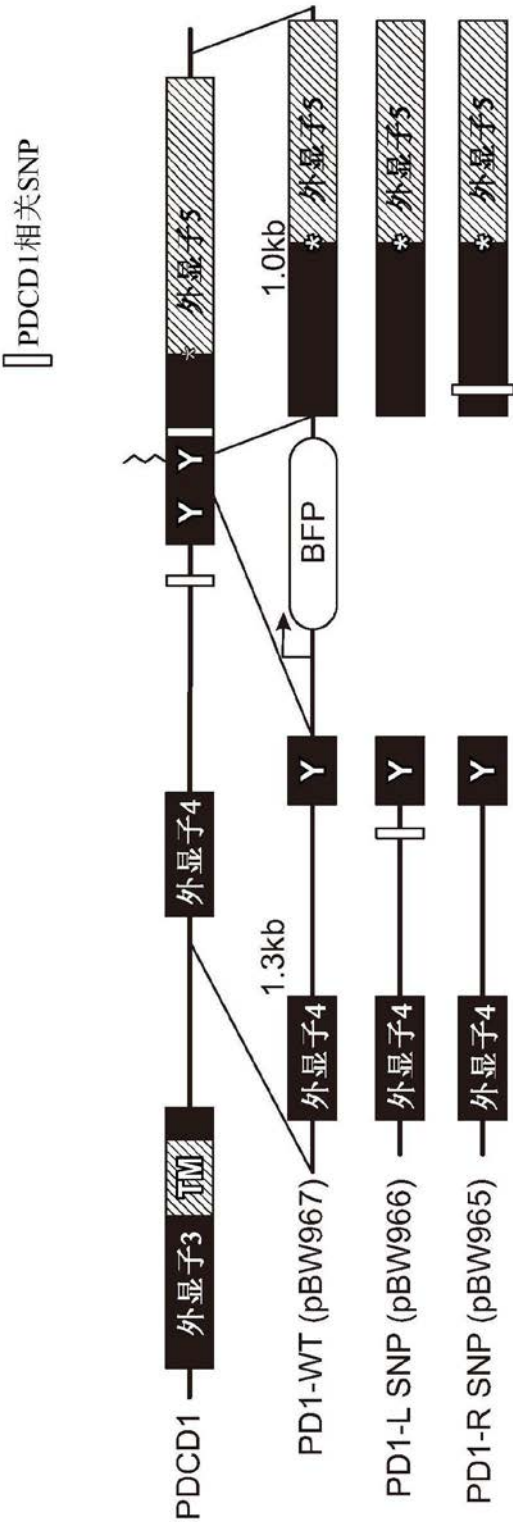


图20

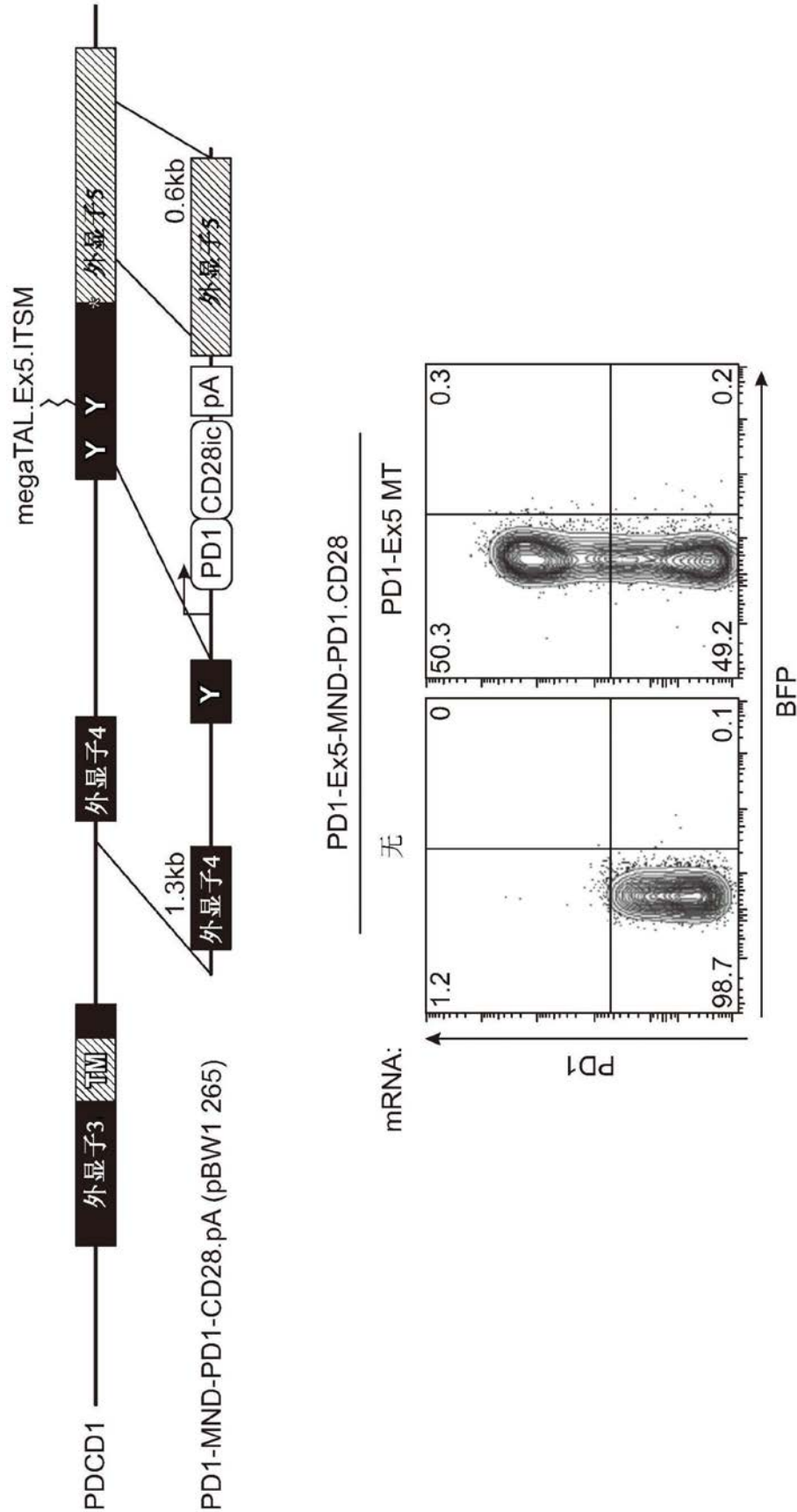


图21

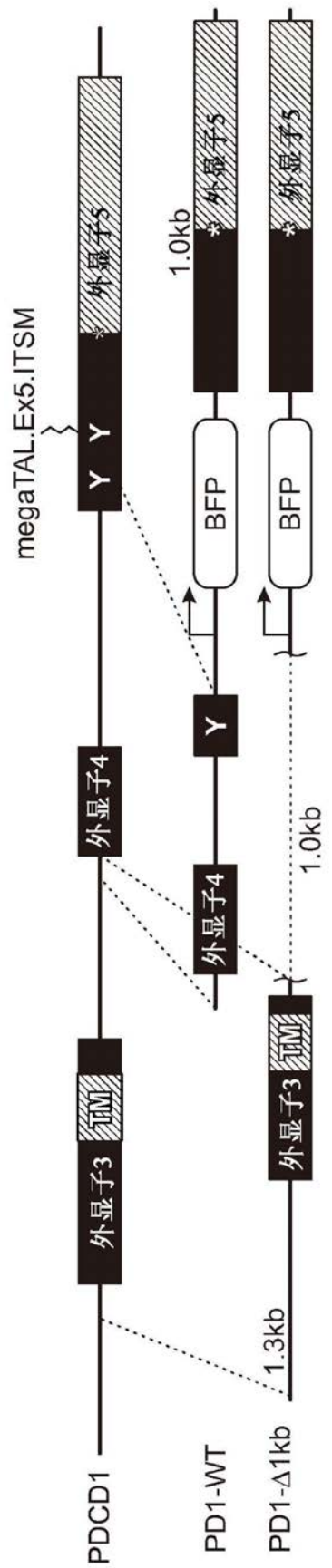


图22

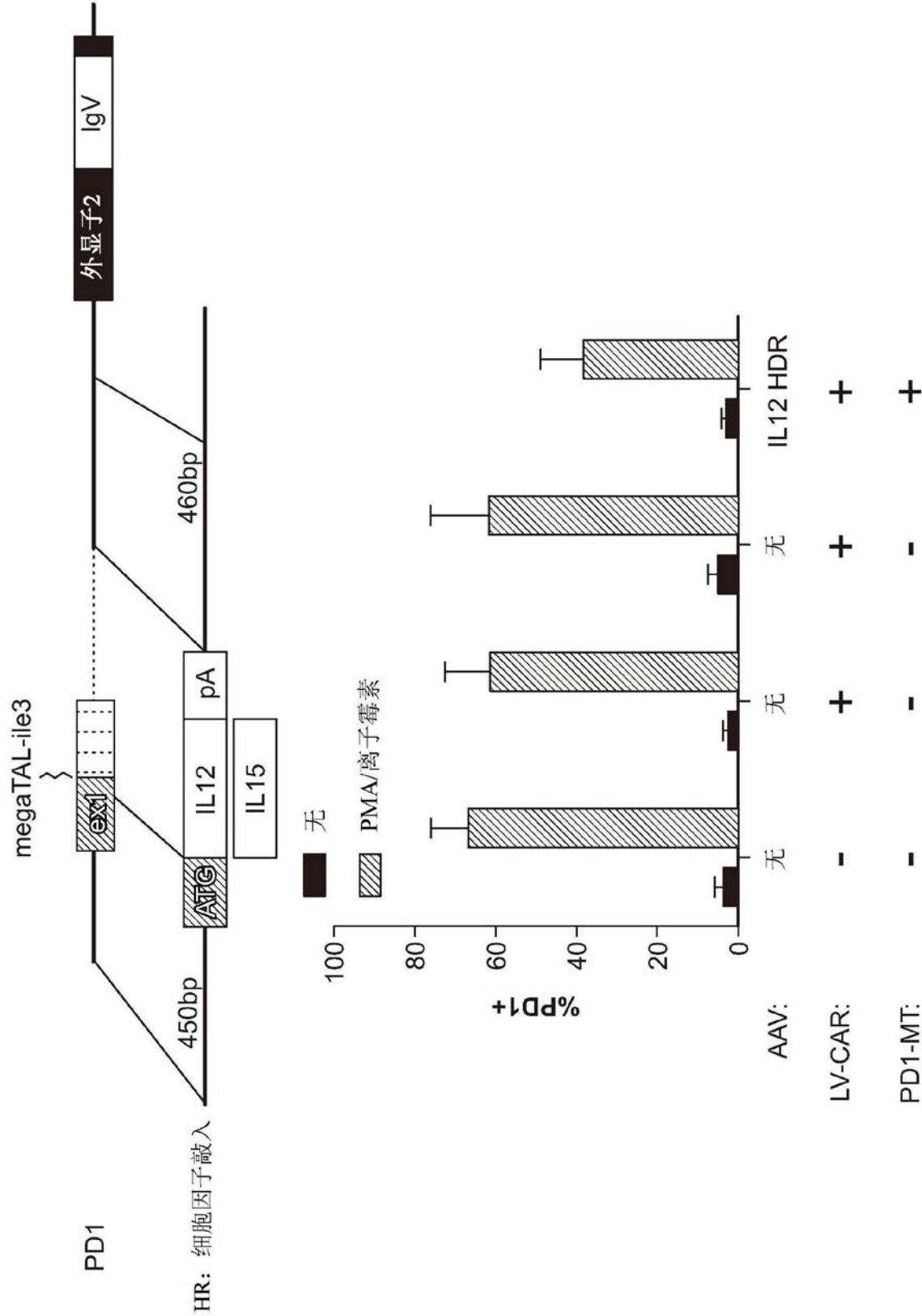


图23

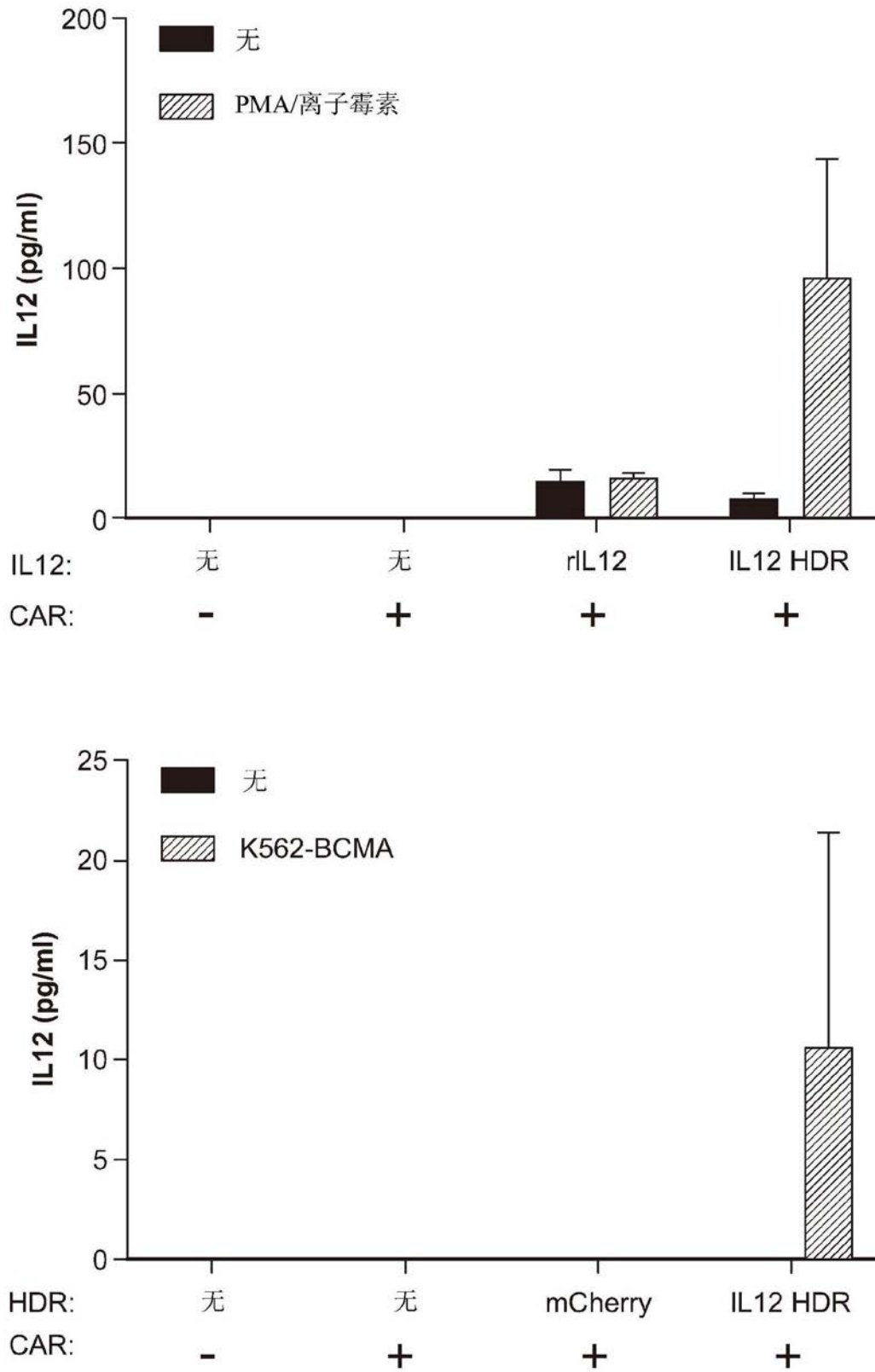


图24A

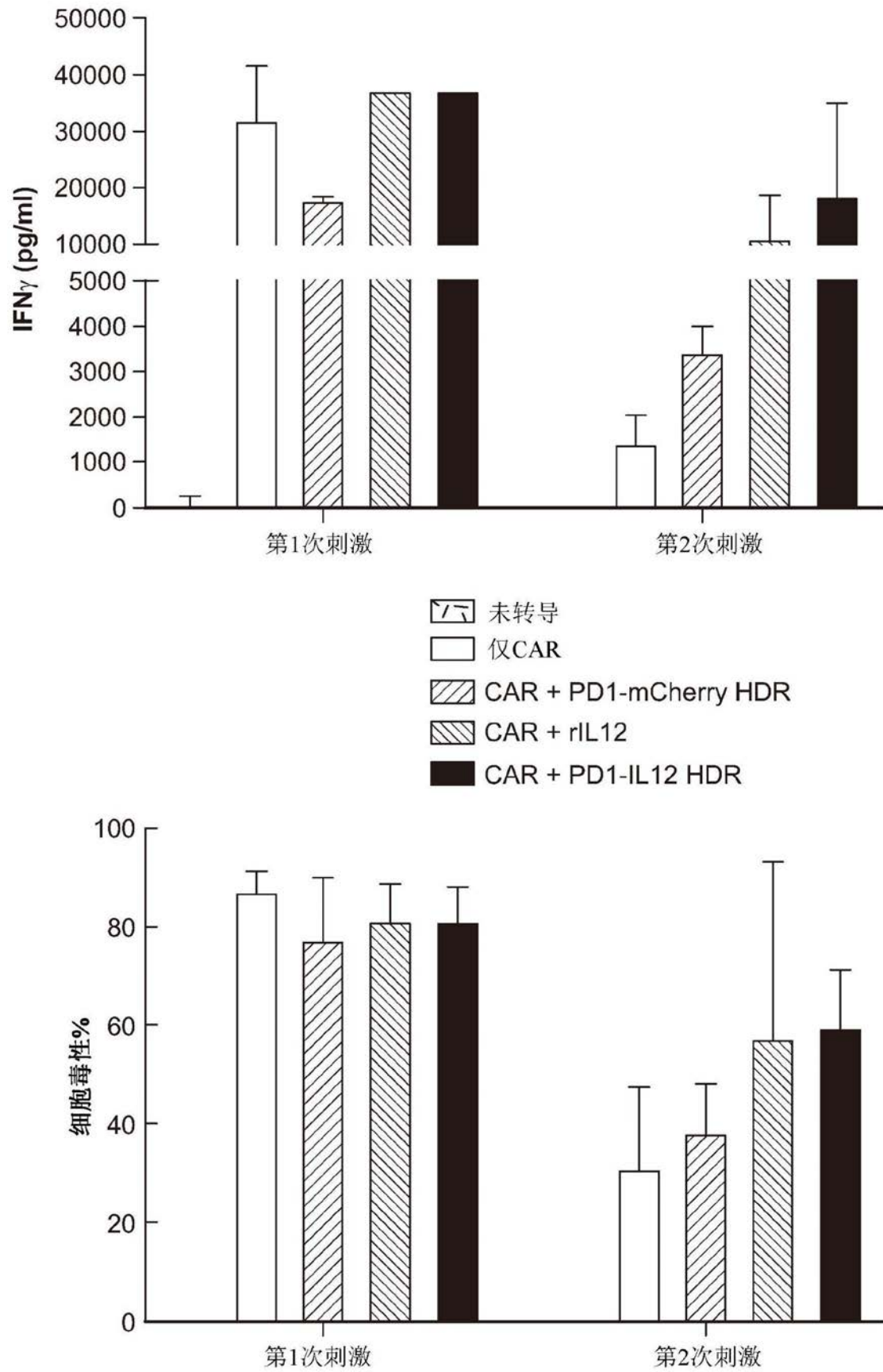


图24B