



(12) 发明专利申请

(10) 申请公布号 CN 114667344 A

(43) 申请公布日 2022.06.24

(21) 申请号 202080078530.1

(74) 专利代理机构 北京集佳知识产权代理有限

(22) 申请日 2020.09.11

公司 11227

(30) 优先权数据

专利代理人 张福誉 韩晓帆

62/899,625 2019.09.12 US

(51) Int.Cl.

(85) PCT国际申请进入国家阶段日

C12N 9/12 (2006.01)

2022.05.12

C12N 15/10 (2006.01)

(86) PCT国际申请的申请数据

PCT/US2020/050323 2020.09.11

(87) PCT国际申请的公布数据

W02021/050822 EN 2021.03.18

(71) 申请人 J·大卫格莱斯顿研究所-根据J·

大卫格莱斯顿遗嘱的遗嘱信托

地址 美国加利福尼亚州

申请人 哈佛大学

(72) 发明人 塞特·希普曼

权利要求书4页 说明书41页

序列表8页 附图21页

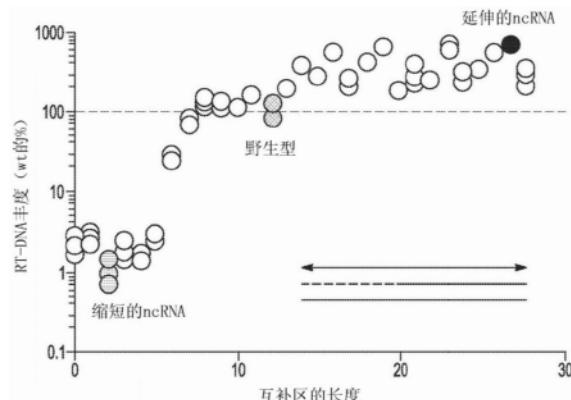
(54) 发明名称

具有增强的DNA产生的经修饰的细菌逆转录

元件

(57) 摘要

提供了经修饰以增强多拷贝单链DNA (msDNA) 的产生的工程化反转录子。此外,还公开了编码这种工程化反转录子的载体系统以及在多种应用例如CRISPR/Cas介导的基因组编辑、重组工程、细胞条码化和分子记录中使用工程化反转录子和编码其的载体系统的方法。



1. 工程化反转录子, 其包含:
 - a) msr前序列;
 - b) 编码多拷贝单链RNA (msRNA) 的msr基因;
 - c) 编码多拷贝单链DNA (msDNA) 的msd基因;
 - d) msd后序列, 其包含与所述msr前序列具有序列互补性的自身互补区, 其中所述自身互补区的长度比野生型互补区长1至50个核苷酸, 使得所述工程化反转录子能够增强所述msDNA的产生; 以及
 - e) 编码逆转录酶的ret基因。
2. 权利要求1所述的工程化反转录子, 其还包含目的异源序列。
3. 权利要求2所述的工程化反转录子, 其中所述异源序列被插入至所述msr基因或所述msd基因中。
4. 权利要求1所述的工程化反转录子, 其中由所述msd基因编码的单链DNA (msDNA) 包含msd茎环, 并且其中所述环包含目的异源序列。
5. 权利要求2所述的工程化反转录子, 其中所述异源序列编码供体多核苷酸, 所述供体多核苷酸包含与5'靶序列杂交的5'同源臂和与3'靶序列杂交的3'同源臂, 所述5'同源臂和所述3'同源臂位于包含待通过同源定向修复 (HDR) 或重组工程整合至靶基因座处的目的编辑物的核苷酸序列的侧翼。
6. 权利要求2所述的工程化反转录子, 其中所述异源序列包含CRISPR前间隔区DNA序列。
7. 权利要求6所述的工程化反转录子, 其中所述CRISPR前间隔区DNA序列包含经修饰的AAG前间区序列邻近基序 (PAM)。
8. 权利要求1所述的工程化反转录子, 其还包含条码序列。
9. 权利要求8所述的工程化反转录子, 其中所述条码序列位于所述msDNA的发夹环中。
10. 权利要求4所述的工程化反转录子, 其中所述msd茎的长度为至少14个核苷酸 (碱基对)。
11. 权利要求4所述的工程化反转录子, 其中所述msd茎可以包含一个或更多个错配的核苷酸。
12. 权利要求4所述的工程化反转录子, 其中所述msd茎具有少于五个或六个连续错配的核苷酸。
13. 权利要求4所述的工程化反转录子, 其中所述msd茎在所述msd茎基部侧翼的序列中没有任何错配或插入。
14. 权利要求4所述的工程化反转录子, 其中相对于所述野生型msd序列的任何msd茎错配或序列改变都在所述msd茎的中部。
15. 权利要求1所述的工程化反转录子, 其中所述msr基因和所述msd基因以反式排列或顺式排列提供。
16. 权利要求1所述的工程化反转录子, 其中所述ret基因相对于所述msr基因和所述msd基因以反式排列提供。
17. 权利要求1所述的工程化反转录子, 其中所述msr基因、msd基因和ret基因是经修饰的细菌反转录子msr基因、msd基因和ret基因。

18. 权利要求1所述的工程化反转录子,其中所述msr基因、msd基因和ret基因独立地为经修饰的黏细菌反转录子、经修饰的大肠杆菌(*Escherichia coli*)反转录子、经修饰的肠沙门氏菌(*Salmonella enterica*)反转录子或经修饰的霍乱弧菌(*Vibrio cholerae*)反转录子。

19. 权利要求18所述的工程化反转录子,其中所述经修饰的大肠杆菌反转录子是经修饰的EC83或经修饰的EC86。

20. 载体系统,其包含含有权利要求1所述的工程化反转录子的一种或更多种载体。

21. 权利要求20所述的载体系统,其中所述msr基因和所述msd基因通过同一载体或不同的载体提供。

22. 权利要求20所述的载体系统,其中所述msr基因、所述msd基因和所述ret基因由同一载体提供。

23. 权利要求22所述的载体系统,其中所述同一载体包含与所述msr基因和所述msd基因可操作地连接的启动子。

24. 权利要求23所述的载体系统,其中所述启动子还与所述ret基因可操作地连接。

25. 权利要求23所述的载体系统,其还包含与所述ret基因可操作地连接的第二启动子。

26. 权利要求20所述的载体系统,其中所述msr基因、所述msd基因和所述ret基因由不同的载体提供。

27. 权利要求20所述的载体系统,其中所述一种或更多种载体是病毒载体或非病毒载体。

28. 权利要求27所述的载体系统,其中所述非病毒载体是质粒。

29. 权利要求20所述的载体系统,其中所述工程化反转录子包含供体多核苷酸,所述供体多核苷酸包含与5'靶序列杂交的5'同源臂和与3'靶序列杂交的3'同源臂,所述5'同源臂和所述3'同源臂位于包含待通过同源定向修复(HDR)或重组工程整合在靶基因座处的目的编辑物的核苷酸序列的侧翼。

30. 权利要求29所述的载体系统,其还包含编码RNA指导的核酸酶的载体。

31. 权利要求30所述的载体系统,其中所述RNA指导的核酸酶是Cas核酸酶或工程化的RNA指导的FokI-核酸酶。

32. 权利要求31所述的载体系统,其中所述Cas核酸酶是Cas9或Cpf1。

33. 权利要求20所述的载体系统,其中所述工程化反转录子包含CRISPR前间隔区DNA序列。

34. 权利要求33所述的载体系统,其还包含编码Cas1和/或Cas2蛋白的载体。

35. 权利要求34所述的载体系统,其还包含含有CRISPR阵列序列的载体。

36. 权利要求20所述的载体系统,其还包含编码噬菌体同源重组蛋白的载体。

37. 权利要求36所述的载体系统,其中所述编码所述噬菌体同源重组蛋白的载体是包含exo、bet和gam基因的复制缺陷型λ原噬菌体。

38. 分离的宿主细胞,其包含权利要求1所述的工程化反转录子或权利要求20中任一项所述的载体系统。

39. 权利要求38所述的宿主细胞,其中所述宿主细胞是原核、古细菌或真核的宿主细

胞。

40. 权利要求39所述的宿主细胞,其中所述真核宿主细胞是哺乳动物宿主细胞。

41. 权利要求40所述的宿主细胞,其中所述哺乳动物宿主细胞是人宿主细胞。

42. 权利要求38所述的宿主细胞,其中所述宿主细胞是人工细胞或经遗传修饰的细胞。

43. 试剂盒,其包含权利要求1所述的工程化反转录子、权利要求20所述的载体系统或权利要求38所述的宿主细胞。

44. 权利要求43所述的试剂盒,其还包含使用该工程化反转录子对细胞进行遗传修饰的说明。

45. 对细胞进行遗传修饰的方法,其包括:

a) 用权利要求5所述的工程化反转录子转染细胞;

b) 将RNA指导的核酸酶和指导RNA引入所述细胞中,其中所述RNA指导的核酸酶与所述指导RNA形成复合体,所述指导RNA将所述复合体引导至基因组靶基因座,其中所述RNA指导的核酸酶在所述基因组靶基因座处的基因组DNA中产生双链断裂,以及由所述工程化反转录子产生的供体多核苷酸通过同源定向修复(HDR)在其5'同源臂和3'同源臂所识别的基因组靶基因座处被整合,以产生经遗传修饰的细胞。

46. 权利要求45所述的方法,其中所述RNA指导的核酸酶是Cas核酸酶或工程化的RNA指导的FokI-核酸酶。

47. 权利要求45所述的方法,其中所述Cas核酸酶是Cas9或Cpf1。

48. 权利要求45所述的方法,其中所述RNA指导的核酸酶由整合到所述细胞的基因组中的载体或重组多核苷酸提供。

49. 权利要求45所述的方法,其中所述工程化反转录子由载体提供。

50. 权利要求45所述的方法,其中所述供体多核苷酸用于产生基因置换、基因敲除、缺失、插入、倒位或点突变。

51. 通过重组工程对细胞进行遗传修饰的方法,所述方法包括:

a) 用权利要求5所述的工程化反转录子转染所述细胞;以及

b) 将噬菌体重组蛋白引入所述细胞中,其中所述噬菌体重组蛋白在靶基因座处介导同源重组,使得由所述工程化反转录子产生的供体多核苷酸在其5'同源臂和3'同源臂所识别的靶基因座处被整合,以产生经遗传修饰的细胞。

52. 权利要求51所述的方法,其中所述供体多核苷酸用于通过重组工程修饰细菌细胞中的质粒、细菌人工染色体(BAC)或细菌染色体。

53. 权利要求51所述的方法,其中所述供体多核苷酸可以产生基因置换、基因敲除、缺失、插入、倒位或点突变。

54. 权利要求51所述的方法,其中所述将噬菌体重组蛋白引入所述细胞中包括将复制缺陷型λ原噬菌体插入至细菌基因组中。

55. 权利要求54所述的方法,其中所述噬菌体包含exo、bet和gam基因。

56. 对细胞进行条码化的方法,其包括用权利要求8所述的工程化反转录子转染细胞。

57. 产生体内分子记录系统的方法,其包括:

a) 将CRISPR改编系统的Cas1蛋白或Cas2蛋白引入至宿主细胞中;

b) 将包含前导序列和至少一个重复序列的CRISPR阵列核酸序列引入至所述宿主细胞

中,其中所述CRISPR阵列核酸序列被整合到所述宿主细胞中的基因组DNA或载体中;以及

c) 将多个根据权利要求10或11所述的工程化反转录子引入至所述宿主细胞中,其中每个反转录子包含不同的前间隔区DNA序列,其可以被加工并插入至所述CRISPR阵列核酸序列中。

58. 权利要求57所述的方法,其中所述Cas1蛋白或所述Cas2蛋白由载体提供。

59. 权利要求57所述的方法,其中所述工程化反转录子由载体提供。

60. 权利要求57所述的方法,其中多个所述工程化反转录子包含至少三个不同的前间隔区DNA序列。

61. 包含体内分子记录系统的工程化细胞,其包含:

a) CRISPR改编系统的Cas1蛋白或Cas2蛋白;

b) 进入宿主细胞中的包含前导序列和至少一个重复序列的CRISPR阵列核酸序列,其中所述CRISPR阵列核酸序列被整合到所述工程化细胞中的基因组DNA或载体中;以及

c) 多个根据权利要求6所述的工程化反转录子,其中每个反转录子包含不同的前间隔区DNA序列,其可以被加工并插入至所述CRISPR阵列核酸序列中。

62. 权利要求61所述的工程化细胞,其中所述Cas1蛋白或所述Cas2蛋白由载体提供。

63. 权利要求61所述的工程化细胞,其中所述工程化反转录子由载体提供。

64. 权利要求61所述的工程化细胞,其中多个所述工程化反转录子包含至少三个不同的前间隔区DNA序列。

65. 试剂盒,其包含权利要求61所述的工程化细胞和用于体内分子记录的说明书。

66. 产生重组msDNA的方法,其包括:

a) 用权利要求1所述的工程化反转录子或权利要求20所述的载体系统转染宿主细胞;以及

b) 在合适的条件下培养所述宿主细胞,其中产生了所述msDNA。

具有增强的DNA产生的经修饰的细菌逆转录元件

[0001] 优先权申请

[0002] 本申请要求2019年9月12日提交的美国临时申请系列No.62/899,625的申请日的优先权权益,其内容通过引用其整体具体并入本文中。

[0003] 通过引用作为文本文件提供的序列表并入

[0004] 序列表作为文本文件“2072305.txt”提供,其创建于2020年9月10日,以及其大小为12,288字节。该文本文件的内容通过引用其整体并入本文中。

背景技术

[0005] 反转录子是逆转录元件,其存在于在几乎所有黏细菌中(Dhundale et al.Journal of Bacteriology 164,914-917(1985)),在大肠杆菌(E.coli)(Lampson et al.Science 243,1033-1038(1989)),V.cholerae(Inouye et al.Microbiology and Immunology 55,510-513)和其他细菌中很少见。反转录子操纵子按顺序编码RNA引物(多拷贝单链RNA,msr)、待逆转录的RNA序列(多拷贝单链DNA,msd)和逆转录酶。反转录子转录物通过自身进行折叠并部分逆转录以生成约80个碱基的单链DNA(single stranded DNA,ssDNA)。虽然反转录子来源的DNA是单链的,但其包含双链DNA的发夹。多个反转录子ssDNA也可以相互互补以形成更大的双链元件。反转录子变体具有不同的DNA长度和碱基含量,但广泛共享这种整体形式。

[0006] 由反转录子生成的ssDNA已在两种背景下用于基因组工程:细菌背景,具有用于重组工程的λRed Beta重组酶(Farzadfar et al.Science 346,1256272,(2014));和真核生物背景,作为酵母中用于Cas9编辑的同源定向修复(homology-directed repair,HDR)模板(Sharon et al.Cell 175,544-557.e516,(2018))。尽管前景广阔,但这些应用经受低于预期的效率和背景限制,其可能源于反转录子内源形式中的元件。这些包括(1)具有将ssDNA的5'端与msr RNA的2'羟基连接的磷酸二酯键的分支结构,(2)不变侧翼区,其可能是反转录子逆转录所需的,但不是修复模板的一部分,(3)有限的总长度,和(4)作为用于Pol III转录的终止子发挥作用的天然poly T延伸。

发明内容

[0007] 提供了经修饰以增强多拷贝单链DNA(multicopy single-stranded DNA,msDNA)的产生的工程化反转录子,其解决与效率和低拷贝数相关的许多现有问题。本文中还描述了编码这种工程化反转录子的载体系统以及在多种应用例如CRISPR/Cas介导的基因组编辑、重组工程、细胞的条码化和分子记录中使用工程化反转录子和载体系统的方法。

[0008] 在一个方面中,提供了工程化反转录子,该工程化反转录子包含:a)msr前序列;b)编码多拷贝单链RNA(msRNA)的msr基因;c)编码多拷贝单链DNA(msDNA)的msd基因;d)包含与msr前序列具有序列互补性的自身互补区的msd后序列,其中自身互补区的长度比野生型互补区长至少1至50个核苷酸,使得该工程化反转录子能够增强msDNA的产生;以及e)编码逆转录酶的ret基因。

[0009] 自身互补区由ncRNA的3'和5'端之间的氢键形成。在某些实施方案中，互补区的长度比野生型互补区长至少1、至少2、至少4、至少6、至少8、至少10、至少12、至少14、至少16、至少18、至少20、至少30、至少40或至少50个核苷酸。例如，自身互补区可具有比野生型互补区长1至50个核苷酸的长度，其包括该范围内的任何长度，例如长1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸。在某些实施方案中，自身互补区的长度比野生型互补区长1至16个核苷酸。

[0010] 在某些实施方案中，msr基因和msd基因以反式排列或顺式排列提供。在一些实施方案中，ret基因相对于msr基因和/或msd基因以反式排列提供。

[0011] 在某些实施方案中，msr基因、msd基因和ret基因源自细菌反转录子，其包括但不限于黏细菌反转录子（例如，Mx65、Mx162）、大肠杆菌（Escherichia coli）反转录子（例如，E67、Ec73、EC83、EC86、EC107）、肠沙门氏菌（Salmonella enterica）反转录子（例如msDNA-St85）和霍乱弧菌（Vibrio cholerae）反转录子（例如Vc81、Vc95、Vc137）。

[0012] 在某些实施方案中，工程化反转录子还包含目的异源序列。异源序列可以插入到例如msr基因或msd基因中。例如，异源序列可以插入到msd茎环的环中。在一些实施方案中，异源序列编码多肽或肽。在另一些实施方案中，异源序列编码供体多核苷酸，该供体多核苷酸包含与5'基因组靶序列杂交的5'同源臂和与3'基因组靶序列杂交的3'同源臂，所述5'同源臂和所述3'同源臂位于包含通过同源定向修复（homology directed repair, HDR）或重组工程待被整合至基因组靶基因座处的目的编辑物的核苷酸序列的侧翼。在另一些实施方案中，异源序列包含CRISPR前间隔区DNA序列。在一个实施方案中，CRISPR前间隔区DNA序列包含经修饰的“AAG”前间区序列邻近基序（protospacer adjacent motif, PAM）。

[0013] 在某些实施方案中，工程化反转录子还包含条码序列。条码序列可位于例如msDNA的发夹环中。

[0014] 在另一个方面中，提供了载体系统，该载体系统包含含有本文中所述的工程化反转录子的一个或更多个载体。在某些实施方案中，msr基因和msd基因由相同的载体或不同的载体提供。在一些实施方案中，msr基因、msd基因和ret基因由同一载体提供，其中该载体包含与msr基因和msd基因可操作地连接的启动子。在一些实施方案中，启动子还与ret基因可操作地连接。在另一些实施方案中，载体还包含与ret基因可操作地连接的第二启动子。在某些实施方案中，msr基因、msd基因和ret基因由不同的载体提供。

[0015] 在某些实施方案中，载体系统的一种或更多种载体是病毒载体或非病毒载体（例如，质粒）。

[0016] 在某些实施方案中，载体系统包含工程化反转录子，其包含编码供体多核苷酸的异源序列，所述供体多核苷酸包含在供体多核苷酸序列侧翼的与5'基因组靶序列杂交的5'同源臂和与3'基因组靶序列杂交的3'同源臂。供体多核苷酸序列例如通过同源定向修复（HDR）或重组工程可以替换或编辑基因组靶基因座。

[0017] 在某些实施方案中，载体系统还包含编码RNA指导的核酸酶的载体。示例性的RNA指导的核酸酶包括但不限于Cas核酸酶（例如Cas9、Cpf1）和工程化的RNA指导的FokI核酸酶。

[0018] 在某些实施方案中，载体系统还包含编码用于重组工程的噬菌体重组蛋白的载

体。在一些实施方案中，载体是编码噬菌体重组蛋白的复制缺陷型原噬菌体。

[0019] 在某些实施方案中，该载体系统包含工程化反转录子，其包含编码CRISPR前间隔区DNA序列的异源序列。在一些实施方案中，载体系统还包含编码Cas1或Cas2蛋白的载体。在一些实施方案中，所述载体系统还包含含有CRISPR阵列序列的载体。

[0020] 在另一个方面中，提供了分离的宿主细胞，该宿主细胞包含本文中所述的工程化反转录子或载体系统。

[0021] 在某些实施方案中，宿主细胞是原核、古细菌或真核的宿主细胞。例如，宿主细胞可以是细菌、原生生物、真菌、动物或植物的宿主细胞。在一些实施方案中，宿主细胞是哺乳动物宿主细胞。宿主细胞可以是人或非人哺乳动物宿主细胞。在另一些实施方案中，宿主细胞是人工细胞或经遗传修饰的细胞。

[0022] 在另一个方面中，提供了包含本文中所述的工程化反转录子的试剂盒，或包含这种工程化反转录子的载体系统或宿主细胞。在一些实施方案中，该试剂盒还包含关于使用工程化反转录子的方法的说明。

[0023] 在另一个方面中，提供了对细胞进行遗传修饰的方法。在一些情况下，该方法包括用工程化反转录子转染细胞。例如，该方法可以包括：a) 用包含编码供体多核苷酸的异源序列的工程化反转录子转染细胞，该供体多核苷酸包含与5'基因组靶序列杂交的5'同源臂和与3'基因组靶序列杂交的3'同源臂，所述5'同源臂和所述3'同源臂在包含通过同源定向修复(HDR)待被整合至基因组靶基因座处的目的编辑物的核苷酸序列的侧翼；以及b) 将RNA指导的核酸酶和指导RNA引入细胞中，其中RNA指导的核酸酶与指导RNA形成复合体，所述指导RNA将复合体引导至基因组靶基因座，其中RNA指导的核酸酶在基因组靶基因座处的基因组DNA中产生双链断裂，以及由工程化反转录子产生的供体多核苷酸通过同源定向修复(HDR)在其5'同源臂和3'同源臂所识别的基因组靶基因座处被整合。具有编码供体多核苷酸的工程化反转录子的HDR可用于例如产生基因替换、基因敲除、缺失、插入、倒位或点突变。在一些情况下，具有编码供体多核苷酸的工程化反转录子的HDR可用于例如修复基因、基因敲除、缺失、插入、倒位或点突变。这样的方法可以由此产生经遗传修饰的细胞。在一些实施方案中，该方法还包括对经遗传修饰的细胞进行表型分析或对经遗传修饰的细胞的基因组进行测序。

[0024] 在另一个方面中，提供了通过重组工程对细胞进行遗传修饰的方法，该方法包括：a) 用包含编码供体多核苷酸的异源序列的工程化反转录子转染细胞，该供体多核苷酸包含与5'基因组靶序列杂交的5'同源臂和与3'基因组靶序列杂交的3'同源臂，所述5'同源臂和所述3'同源臂在包含通过重组工程待被整合至基因组靶基因座处的目的编辑物的核苷酸序列的侧翼；以及b) 将噬菌体重组蛋白引入细胞中，其中噬菌体重组蛋白在靶基因座处介导同源重组，使得由工程化反转录子产生的供体多核苷酸在其5'同源臂和3'同源臂所识别的靶基因座处被整合，以产生经遗传修饰的细胞。例如，可以使用用编码供体多核苷酸的工程化反转录子的重组工程，以产生基因替换、基因敲除、缺失、插入、倒位或点突变。在某些实施方案中，供体多核苷酸可用于通过重组工程修饰质粒、细菌人工染色体(bacterial artificial chromosome, BAC)或细菌染色体(在细菌细胞中)。在一些实施方案中，该方法还包括对经遗传修饰的细胞进行表型分析或对经遗传修饰细胞的基因组进行测序。

[0025] 在某些实施方案中，通过将复制缺陷型λ原噬菌体插入细菌基因组中，将噬菌体重

组蛋白引入细菌细胞中。在一个实施方案中，噬菌体包含exo、bet和gam基因。

[0026] 在另一个方面中，提供了对细胞进行条码化的方法，该方法包括用包含条码的工程化反转录子转染细胞，如本文中所述。

[0027] 在另一个方面中，提供了产生体内分子记录系统的方法，该方法包括：a) 将CRISPR改编系统的Cas1蛋白或Cas2蛋白引入宿主细胞中；b) 将包含前导序列和至少一个重复序列的CRISPR阵列核酸序列引入宿主细胞中，其中该CRISPR阵列核酸序列被整合到宿主细胞中的基因组DNA中或整合到宿主细胞中的载体中；以及c) 将多个包含CRISPR前间隔区DNA序列的工程化反转录子引入宿主细胞中，其中每个反转录子包含不同的前间隔区DNA序列，其可以被加工并插入到CRISPR阵列核酸序列中。在某些实施方案中，Cas1蛋白或Cas2蛋白由载体提供。在某些实施方案中，工程化反转录子由载体提供。在某些实施方案中，多个工程化反转录子包含至少三个不同的前间隔区DNA序列。

[0028] 在另一个方面中，提供了包含体内分子记录系统的工程化细胞，该工程化细胞包含：a) CRISPR改编系统的Cas1蛋白或Cas2蛋白；b) 进入宿主细胞的包含前导序列和至少一个重复序列的CRISPR阵列核酸序列，其中CRISPR阵列核酸序列被整合到工程化细胞中的基因组DNA或载体中；以及c) 多个工程化反转录子，每个包含CRISPR前间隔区DNA序列，其中每个反转录子包含不同的前间隔区DNA序列，其可以被加工并插入到CRISPR阵列核酸序列中。在某些实施方案中，Cas1蛋白或Cas2蛋白由载体提供。在某些实施方案中，工程化反转录子由载体提供。在某些实施方案中，多个工程化反转录子包含至少三个不同的前间隔区DNA序列。

[0029] 在另一个方面中，提供了包含工程化细胞的试剂盒，该工程化细胞包含如本文中所述的体内分子记录系统。在一些实施方案中，该试剂盒还包含用于在体内分子记录的说明。

[0030] 在另一方面中，提供了产生重组msDNA的方法，该方法包括：a) 用本文所述的工程化反转录子或载体系统转染宿主细胞；以及b) 在合适的条件下培养宿主细胞，其中产生了msDNA。

附图说明

[0031] 图1A至1D示出了反转录子操纵子和反转录子潜在用途的示意图。图1A示出了编码msr、msd和逆转录酶的反转录子操纵子的示意图，其中逆转录酶可以合成编码多拷贝单链DNA的msd基因的一部分的DNA拷贝。图1B示出了重组工程是反转录子的潜在用途，其中Beta可以保护ssDNA并且可以促进ssDNA与互补的ssDNA靶标例如细胞中的DNA靶标的退火。图1C示出了CRISPR/Cas9基因编辑是反转录子的潜在用途，其中反转录子可以提供可以修复变体或突变靶位点的ssDNA模板。图1D示出了分子记录是反转录子的潜在用途（例如，如在W02018191525 A1中提供的，其通过引用其整体具体并入本文中）。

[0032] 图2A至2B示出了反转录子元件及其组装。图2A示出了反转录子元件：(1) msr和逆转录的msd的5'端通过2'至5'键共价键合至引发鸟苷，这种分支结构阻碍了基因组工程中的使用，(2) 不变侧翼区可以是反转录子逆转录所需的，因此其不能容易地成为修复模板的一部分，(3) 目前认为总长度有限的茎。用于基因组工程的另一个问题是作为用于Pol III转录的终止子发挥作用的天然反转录子poly T延伸。图2B示出了反转录子操纵子的非蛋白

质编码(msr-msd)部分产生具有显著二级结构的转录物，并且逆转录酶(reverse transcriptase, RT)识别该转录物中的特定起始位点，然后将该转录物部分逆转录为RT-DNA(msd)。

[0033] 图3A至3D。图3A示出了野生型ec86(也称为反转录子-Eco1 ncRNA)在逆转录之后的碱基结构，其中msd DNA处于顶部(SEQ ID NO:1 GTCAGAAAAACGGGTTCCCTGGTGGCTCGGA GAGCATCAGCGATGCTCTCCGTTCCAACAAGGAAAACAGACAG TAACTCAGA)以及msr RNA是较低的序列(SEQ ID NO:2 AUGCGCACCUUAGCGAGAGGUUAUCAUUAAGGUCAACCUCUGGAUGUUGUUUCGGCAUC CUGCAUUGAAUCUGAGUUACU)。图3B示出了在ec86表达之后产生的ssDNA的量，如通过qPCR分析所检测的。图3C示出了野生型和变体msd的PAGE分析。图3D示出了以下两个变体msd的碱基结构：反转录子-Eco1 v32 ncRNA，其改造自ec86野生型(GTCAGAAAAACGGGTTGTCGCCAGTCT GACTGG CGACAAACAGCTTGTAACTCAGA, SEQ ID NO:3)，以及反转录子-Eco1 v35 ncRNA，其改造自v32 ncRNA(GTCAGAAAAACGGGTGGAGAG GTTGCTGCAACCTCTCCATTCTGTAACTCAGA, SEQ ID NO:4)。

[0034] 图4A至4D示出了用于产生延伸的msd ssDNA的表达系统。图4A示出了将msr/msd与反转录子逆转录酶(RT)分离的表达构建体，这允许产生更长的(经修饰的)逆转录的msd ssDNA。图4B示出了与逆转录酶编码区分离(呈现反式)的表达盒中msr和msd的排列。图4C示出了与逆转录酶编码区分离(呈现反式)的msr/msd表达盒中msd ssDNA的数个延伸，其表明msd区可以显著扩展以包括异源序列。图4D示出了msd ssDNA的PAGE分析，其包括如图4A至4C所示产生的延伸的msd ssDNA。

[0035] 图5示出了可被修饰的反转录子特征数据(parameter)。

[0036] 图6A至6F。图6A示意性地示出了定制的测序制备路线。在存在RNA酶的情况下，用脱支RNA套索1(debranching RNA lariats 1,DBR1)处理ssDNA，然后使用不依赖模板的聚合酶(TdT)添加一系列单一类型的多核苷酸，使用含有衔接子的反向锚定引物生成互补链，连接第二个衔接子，并且然后将这种衔接子连接的双链DNA进行索引(index)并进行多重测序(SEQ ID NO:29)。图6B示出了通过TdT添加的核苷酸数量是可控的。图6C示出了有序的msd ssDNA ec86 v 32序列(GTCAGAAAAACGGGTTGTCGCCAGTCTGACTGGCGACAAACAGCTTGT AACTCAGA, SEQ ID NO:5)，其说明了通过测序的验证。图6D示出了预测的msd ssDNA ec86 v 32序列(GTCAGAAAAACGGGT GTCGCCAGTCTGACTGGCGACAAACAGCTTGTAACTCAGA, SEQ ID NO:6)，其说明了测序结果(GTCAGAAAAACGGGTGTCGCCAG TCTGACTGGCGACAAACAGCTTGAACCTCAG, SEQ ID NO:7)。图6E示出了文献野生型msd ssDNA ec86序列(GTCAGAAAAA CGGGTTCCCTGG TTGGCTCGGAGAGCATCAGCGATGCTCTCCGTTCCAACAAGGAAAACAGACAGTAACCTCAGA, SEQ ID NO:8)，其说明了测序结果(GTCAGAAAAACGGGTTCCCTGGTGGCTGGAGAGCATCAGCGATGCTCTCC GTTCCAACAAGGAAAACAGACAGTAACCTCAG, SEQ ID NO:9)。图6F示出了文献野生型msd ssDNA ec83序列(TTGAAGCCCGGAACAAACTTTGATCCGCAACCTACTGGATTGCGGCTCAAAAGTTGTCCGC A A C T G T A A A T G T A A T C , SEQ ID NO:10)，其说明了测序结果(AGCCGCGGAACAAACTTTGATCCGCAACC TACTGGATTGCGGCTCAAAAGTTGTCCGCAACTGTAAATG TAATC, SEQ ID NO:11)。

[0037] 图7A至7C示出了msd DNA的修饰。图7A示意性地示出了反转录子RNA中的变化与将在msd DNA中结束的条码的关联。图7B示出了与没有较长msd后区的野生型反转录子相比，

从具有较长msd后互补区的反转录子产生的ssDNA提高。图7C-1和7C-2示出了反转录子非编码RNA (ncRNA) 的5'和3'端区的延伸和降低。图7C-1示意性地示出了使用的基本反转录子结构, 其中ncRNA中的延伸的互补区用黑色实线标记, 而剩余的ncRNA为虚线。图7C-2图示示出了ncRNA互补区的延伸相对于野生型序列(其中野生型的丰度为100%)提高了RT-DNA的丰度, 但ncRNA互补区的缩短减少了RT-DNA的丰度。显示的数据来自每个变体的合并实验(n=3, 显示重复)。

[0038] 图8A至8B图示示出了可以通过缩短ncRNA的逆转录茎来降低ssDNA的量, 但是茎的延伸不会负面影响ssDNA的产生。图8A示意性地示出了ncRNA结构经修饰的部分, 其中茎区显示为黑色实线, 而ncRNA的其余部分显示为虚线。图8B图示示出了ncRNA区延伸约15至30个核苷酸将RT-DNA的丰度维持在与观察到的非延伸的野生型ncRNA序列大致相同的水平, 然而当ncRNA区的长度降低至小于约14个核苷酸时, 与未延伸的野生型ncRNA序列相比, 降低了通过逆转录生成的ssDNA的量。

[0039] 图9A至9B示出了断裂和固定ncRNA的逆转录茎区的影响。图9A是ncRNA的示意图, 其中ncRNA的逆转录茎区显示为黑色实线。图9B图示示出了ncRNA结构变体的逆转录DNA相对于野生型序列的丰度。数据来自每个变体的合并实验。实施例中提供了断裂茎、固定茎和可容许的断裂茎ncRNA结构变体的序列。

[0040] 图10A至10E示出了ncRNA的逆转录区中的插入和缺失对从ncRNA逆转录的DNA的丰度的影响。图10A示意性地示出了ncRNA, 其中ncRNA的逆转录区显示为黑色虚线和实线。虚线标识了msd茎侧翼的区。图10B图示示出了由一系列ncRNA变体的逆转录产生的RT-DNA丰度, 每个ncRNA变体相对于野生型序列在沿msd茎环的不同位置处具有3个碱基的缺失。缺失的位置沿x轴绘制。图10C图示示出了由一系列ncRNA变体的逆转录产生的RT-DNA丰度, 每个ncRNA变体相对于野生型序列在沿msd茎环的不同位置处具有3个碱基的插入。插入的位置沿x轴绘制。图10D图示示出了由一系列ncRNA变体的逆转录产生的RT-DNA丰度, 每个ncRNA变体相对于野生型序列在沿msd茎环的不同位置具有单个碱基变化。插入的位置沿x轴绘制。图10E图示示出了鉴于图10B至10D的结构变化和观察到的结果, msd环位置的可修饰性得分。可修饰性得分基于这些变化的平均影响, 其中数据来自每个变体的合并实验。折叠的ncRNA的茎、环和侧翼区的示意图显示在图10B至10C中。

[0041] 图11A至11B示出了使用经修饰的反转录子来改进基于CRISPR的基因组变化。图11A是说明通过CRISPR整合酶Cas1和Cas2整合反转录子RT-DNA以修饰基因组CRISPR阵列的示意图。图11B图示示出了可通过在ncRNA的5'和3'端延伸自身互补区来增强反转录子来源的间隔DNA。

[0042] 定义

[0043] 如本文中使用的术语“约”在提及可测量值例如量、长度等时, 意指涵盖给定值的±20%或±10%、更优选±5%、甚至更优选±1%、且还更优选±0.1%的变化。

[0044] 如本文中用于描述核酸分子的“重组的”意指基因组、cDNA、细菌、半合成或合成来源的多核苷酸, 由于其来源或操作, 其与自然界中与其相关的多核苷酸的全部或部分不相关。

[0045] 关于蛋白质或多肽使用的术语“重组的”意指通过重组多核苷酸的表达产生的多肽。一般而言, 如下文进一步所述的目的基因被克隆, 然后在转化的生物体中表达。宿主生

物体在表达条件下表达外源基因以产生蛋白质。

[0046] 如本文使用的“细胞”是指从原核生物、真核生物或古细菌生物体(其包括细菌、古细菌、真菌、原生生物、植物和动物)中分离的任何类型的细胞,其包括来自组织、器官和活检的细胞,以及重组细胞、来自体外培养的细胞系的细胞,以及包含核酸的细胞片段、细胞组分或细胞器。该术语还涵盖人工细胞,例如纳米颗粒、脂质体、聚合体、或微胶囊,其封装核酸。例如,本文中所述的方法可以例如在包含单一细胞或细胞群的样品上进行。该术语还包括经遗传修饰的细胞。

[0047] 术语“转化”是指将外源多核苷酸(例如,工程化反转录子)插入至宿主细胞中,与用于插入的方法无关。例如,包括直接摄取、转导或f-接合。外源多核苷酸可以保持为非整合载体,例如质粒,或者替代地可以整合到宿主基因组中。

[0048] “重组宿主细胞”、“宿主细胞”、“细胞”、“细胞系”、“细胞培养物”和其他表示微生物或高等真核细胞系(其作为单细胞实体培养)的这种术语是指可以或已经被用作重组载体或其他转移DNA的接受体,并包括已转染的原始细胞的原始后代。

[0049] “编码序列”或“编码”选定多肽的序列是核酸分子,当置于适当的调节序列(或“控制元件”)的控制下时,该核酸分子在体内被转录(在DNA的情况下)和翻译(在mRNA的情况下)成多肽。编码序列的边界可以由在5'(氨基)端的起始密码子和在3'(羧基)端的翻译终止密码子确定。编码序列可以包括但不限于来自病毒mRNA、原核mRNA或真核mRNA的cDNA、来自病毒DNA或原核DNA的基因组DNA序列,并且甚至是合成的DNA序列。转录终止序列可以位于编码序列的3'处。

[0050] 典型的“控制元件”包括但不限于转录启动子、转录增强子元件、转录终止信号、多聚腺苷酸化序列(位于翻译终止密码子的3'处)、用于优化翻译起始的序列(位于编码序列的5'处)和翻译终止序列。

[0051] “可操作地连接”指的是元件的排列,其中如此描述的组分被配置以便执行其通常功能。因此,当存在适当的酶时,与编码序列可操作地连接的给定启动子能够影响编码序列的表达。启动子不必与编码序列邻接,只要其发挥指导其表达的功能即可。因此,例如,在启动子序列和编码序列之间可以存在插入的未翻译但转录的序列,并且启动子序列仍然可以被认为与编码序列“可操作地连接”。

[0052] “由……编码”是指编码多肽或RNA序列的核酸序列。例如,多肽序列或其部分含有来自核酸序列编码的多肽的至少3至5个氨基酸、更优选至少8至10个氨基酸、甚至更优选至少15至20个氨基酸的氨基酸序列。RNA序列或其部分含有至少3至5个核苷酸,更优选至少8至10个核苷酸,甚至更优选至少15至20个核苷酸的核苷酸序列。

[0053] 术语“分离的”、“纯化的”或“生物学纯的”是指在不同程度上不含如存在于其天然状态下通常伴随其的组分的物质。“分离的”表示与原始来源或环境分离的程度。“纯度”表示高于分离的分开程度。“纯化的”或“生物学纯的”蛋白质充分不含其他物质,使得任何杂质不会实质性地影响蛋白质的生物学特性或导致其他不良后果。换言之,如果本发明的核酸或肽在通过重组DNA技术产生时基本上不含细胞物质、病毒物质或培养基,或者在化学合成时基本上不含化学前体或其他化学物质,则本发明的核酸或肽是经纯化的。纯度和均质性通常使用分析化学技术,例如聚丙烯酰胺凝胶电泳或高效液相色谱来确定。术语“纯化的”可以表示核酸或蛋白质在电泳凝胶中产生基本上一条带。对于可以进行修饰(例如磷酸

化或糖基化)的蛋白质,不同的修饰可产生不同的分离蛋白质,其可以被单独纯化。

[0054] “基本上纯化的”通常是指物质(化合物、多核苷酸、蛋白质、多肽、肽组合物)的分离,使得该物质占其所在样品的大部分百分比。通常来说,在样品中,基本上纯化的组分占样品的50%,优选80%至85%,更优选90%至95%。用于纯化目的多核苷酸和多肽的技术在本领域中是公知的,并且包括例如离子交换色谱、亲和色谱和根据密度的沉降。

[0055] “表达”是指通过细胞可检测到的基因产物的产生。基因产物可以是转录产物(即RNA),其可以被称为“基因表达”,或者基因产物可以是转录产物的翻译产物(即蛋白质),其取决于上下文。“纯化的多核苷酸”是指目的多核苷酸或其片段,其基本上不含与多核苷酸天然相关的蛋白质和/或核酸,例如,含有少于约50%、优选少于约70%、更优选少于约至少90%的与多核苷酸天然相关的蛋白质和/或核酸。用于纯化目的多核苷酸的技术在本领域中是可获得的并且包括,例如,用离液剂破坏含有多核苷酸的细胞,以及通过离子交换色谱、亲和色谱和根据密度的沉降来分离多核苷酸和蛋白质。

[0056] 术语“转染”用于指通过细胞摄取外源DNA。当外源DNA被引入细胞膜内时,细胞即被“转染”。许多转染技术在本领域中通常是熟知的。参见,例如,Graham et al.(1973)Virology,52:456,Sambrook et al.(2001)Molecular Cloning,a laboratory manual,3rd edition,Cold Spring Harbor Laboratories,New York,Davis et al.(1995)Basic Methods in Molecular Biology,2nd edition,McGraw-Hill,以及Chu et al.(1981)Gene 13:197。这种技术可用于将一种或更多种外源DNA部分引入合适的宿主细胞中。该术语指遗传物质的稳定且瞬时的摄取,以及包括肽连接或抗体连接的DNA的摄取。

[0057] “载体”能够将核酸序列转移至靶细胞(例如,病毒载体、非病毒载体、微粒载体和脂质体)。通常来说,“载体构建体”、“表达载体”和“基因转移载体”意指能够指导目的核酸表达并且可以将核酸序列转移至靶细胞的任何核酸构建体。因此,该术语包括克隆和表达载体,以及病毒载体。

[0058] 如本文中所述,“哺乳动物细胞”是指源自哺乳动物对象的任何细胞,其适合于用工程化反转录子或包含工程化反转录子的载体系统转染。细胞可以是异种的、自体的或同种异体的。该细胞可以是直接从哺乳动物对象获得的原代细胞。所述细胞也可以是源自从哺乳动物对象获得的细胞的培养和扩增中的细胞。永生化细胞也包含在该定义中。在一些实施方案中,细胞已被遗传工程化以表达重组蛋白和/或核酸。

[0059] 术语“对象”包括动物,其包括脊椎动物和无脊椎动物两者,包括但不限于无脊椎动物,例如节肢动物、软体动物、环节动物和刺胞动物;以及脊椎动物,如两栖动物,其包括青蛙、蝾螈和盲肠动物;爬行动物,其包括蜥蜴、蛇、海龟、鳄鱼和短吻鳄;鱼;哺乳动物,其包括人和非人哺乳动物,例如非人灵长类动物,其包括黑猩猩和其他猿类和猴类;实验动物,例如小鼠、大鼠、兔子、仓鼠、豚鼠和毛丝鼠;家养动物,例如狗和猫;农场动物,如绵羊、山羊、猪、马和牛;以及鸟类,例如家禽、野鸟和猎鸟,其包括鸡、火鸡和其他鹑鸡类、鸭、鹅等。在一些情况下,所公开的方法可用于实验动物、兽医应用和用于疾病动物模型的开发,其包括但不限于啮齿动物,其包括小鼠、大鼠和仓鼠;灵长类动物和转基因动物。

[0060] “基因转移”或“基因递送”是指用于将目的DNA或RNA可靠地插入宿主细胞中的方法或系统。这种方法可导致非整合的转移DNA的瞬时表达、染色体外复制和转移复制子(例如附加体)的表达,或转移的遗传物质整合到宿主细胞的基因组DNA中。基因递送表达载体

包括但不限于：病毒载体、非病毒载体、甲病毒、痘病毒和痘苗病毒、细菌质粒载体来源的载体。

[0061] 术语“源自”在本文中用于识别分子的原始来源，但并不意味着限制制备分子的方法（例如其可以通过化学合成或重组方式）。

[0062] “源自”指定序列的多核苷酸是指多核苷酸序列，其包含对应于指定核苷酸序列的区（即与之具有同一性或互补）的大约至少约6个核苷酸、优选至少约8个核苷酸、更优选至少约10至12个核苷酸、并且甚至更优选至少约15至20个核苷酸的连续序列。所得到的多核苷酸不一定物理地源自目的核苷酸序列，而是可以任何方式（其包括但不限于化学合成、复制、逆转录或转录，基于以通过多核苷酸所来源的区中碱基序列所提供的信息）生成。因此，其可代表原始多核苷酸的有义或反义取向。

[0063] “条码”是指一种或更多种核苷酸序列，其用于识别与条码相关的核酸或细胞。条码长度可以是3至1000个或更多个核苷酸，优选长度为10至250个核苷酸，并更优选长度为10至30个核苷酸，其包括这些范围内的任何长度，例如长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个核苷酸。例如，条码可用于识别核酸所来源的单个细胞、细胞亚群、集落或样品。条码也可用于识别核酸所来源的细胞、集落或样品的位置（即位置条码），例如细胞阵列中集落的位置、多孔板中孔的位置，或支架中管、烧瓶或其他容器的位置。例如，条码可用于识别核酸所来源的经遗传修饰的细胞。在一些实施方案中，条码用于识别特定类型的基因组编辑或特定类型的供体核酸。

[0064] 术语“杂交(hybridize)”和“杂交(hybridization)”是指在核苷酸序列之间形成复合体，所述核苷酸序列足够互补以通过沃森-克里克碱基配对形成复合体。

[0065] 术语“同源区”是指与另一个核酸区同源的核酸区。因此，参照相同或不同分子中的另一个核酸区确定核酸分子中是否存在“同源区”。此外，由于核酸通常是双链的，因此本文中使用的术语“同源区”是指核酸分子彼此杂交的能力。例如，单链核酸分子可以具有两个能够彼此杂交的同源区。因此，术语“同源区”包括具有互补序列的核酸区段。同源区的长度可以不同，但通常将在4至500个核苷酸之间（例如，约4至约40、约40至约80、约80至约120、约120至约160、约160至约200、约200至约240、约240至约280、约280至约320、约320至约360、约360至约400、约400至约440等）。

[0066] 如本文中使用的术语“互补”或“互补性”是指能够彼此形成碱基对的多核苷酸。碱基对通常由多核苷酸链之间反平行取向的核苷酸单元之间的氢键形成。互补的多核苷酸链可以以沃森-克里克方式（例如，A与T、A与U、C与G）或以允许形成双链体的任何其他方式进行碱基配对。如本领域技术人员所知，当使用RNA而不是DNA时，认为与腺苷互补的碱基是尿嘧啶(U)而不是胸腺嘧啶(T)。然而，除非另有说明，否则当在本发明的上下文中表示尿嘧啶时，暗示了替换胸腺嘧啶的能力。“互补性”可存在于两条RNA链、两条DNA链之间或一条RNA链和一条DNA链之间。通常理解两个或更多个多核苷酸可以是“互补的”并且能够形成双链体，尽管具有不完全或少于100%的互补性。如果每个多核苷酸序列的至少一个连续部分（包含互补区）与另外的多核苷酸进行完全碱基配对且在这种区内没有任何错配或中断，则两个序列是“完全互补的”或“100%互补的”。两个或更多个序列被认为是“完全互补的”或“100%互补的”，即使一个或两个多核苷酸包含另外的非互补序列，只要每个多核苷酸内的

连续的互补区能够与另一个完全杂交即可。“不完全”的互补性是指在这种互补性区内，少于全部的连续核苷酸能够彼此碱基配对的情况。确定两个多核苷酸序列之间的互补性百分比是本领域普通技术人员的问题。

[0067] 如本文中使用的术语“Cas9”涵盖来自任何物种的II型成簇规则间隔短回文重复 (clustered regularly interspaced short palindromic repeat, CRISPR) 系统Cas9核酸内切酶，并且还包括其保留Cas9核酸内切酶活性(即催化DNA的定点切割以生成双链断裂) 的生物活性片段、变体、类似物及衍生物。Cas9核酸内切酶在包含与其结合的指导RNA (gRNA) 互补的序列的位点处结合并切割DNA。为了靶向Cas9的目的，gRNA可包含与靶序列“互补”的序列(例如，主要或次要等位基因)，能够进行足够的碱基配对以形成双链体(即，gRNA与靶序列杂交)。此外，gRNA可包含与PAM序列互补的序列，其中gRNA还与靶DNA中的PAM序列杂交。

[0068] 术语“供体多核苷酸”是指提供通过HDR或重组工程在靶基因座处待被整合至基因组中的预期编辑序列的多核苷酸。

[0069] “靶位点”或“靶序列”是由指导RNA (gRNA) 或供体多核苷酸的同源臂识别(即，对于杂交而言充分互补)的核酸序列。靶位点可以是等位基因特异性的(例如，主要或次要等位基因)。例如，靶位点可以是旨在被修饰，例如通过插入一个或更多个核苷酸、替换一个或更多个核苷酸、删除一个或更多个核苷酸或其组合进行修饰的基因组位点。

[0070] “同源臂”意指供体多核苷酸的一部分，其负责将供体多核苷酸靶向至细胞中待编辑的基因组序列。供体多核苷酸通常包含与5'基因组靶序列杂交的5'同源臂和与3'基因组靶序列杂交的3'同源臂，所述5'同源臂和所述3'同源臂位于包含对基因组DNA的目的编辑物的核苷酸序列的侧翼。同源臂在本文中称为5'和3'(即上游和下游) 同源臂，其涉及同源臂与供体多核苷酸内包含目的编辑物的核苷酸序列的相对位置。5'和3'同源臂与待修饰的基因组DNA中靶基因座内的区杂交，在本文中分别称为“5'靶序列”和“3'靶序列”。例如，包含目的编辑物的核苷酸序列可以通过HDR或重组工程在由5'和3'同源臂识别(即，对于杂交而言充分互补)的基因组靶基因座处整合到基因组DNA中。

[0071] 一般而言，“CRISPR改编系统”是指转录物和其他元件，其参与CRISPR相关 (“Cas”) 基因的表达或指导CRISPR相关 (“Cas”) 基因的活性，包括编码Cas基因的序列和CRISPR阵列核酸序列(其包含前导序列和至少一个重复序列)。在一些实施方案中，CRISPR改编系统的一个或更多个元件源自I型、II型或III型CRISPR系统。Cas1和Cas2存在于所有三种类型的CRISPR-Cas系统中，并且其涉及间隔子获取。在大肠杆菌的I-E系统中，Cas1和Cas2形成复合体，其中Cas2二聚体桥接两个Cas1二聚体。在这个复合体中，Cas2发挥非酶支架作用，其结合侵入DNA的双链片段，而Cas1结合DNA的单链侧翼并催化其整合到CRISPR阵列中。

[0072] 在一些实施方案中，CRISPR系统的一种或更多种元件源自包含内源性CRISPR系统的特定生物体，例如酿脓链球菌 (*Streptococcus pyogenes*)。通常来说，CRISPR系统的特征在于促进在靶序列位点形成CRISPR复合体的元件(在内源性CRISPR系统的上下文中也称为前间隔区)。

[0073] 在一些实施方案中，载体包含调节元件，其可操作地连接至编码CRISPR酶的酶编码序列，例如Cas蛋白。Cas蛋白的非限制性实例包括

[0074] Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9(也被称为Csn1和

Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4,

[0075] 其同源物或其经修饰形式。

[0076] 在某些实施方案中,本公开内容提供了与称为前间区序列邻近基序 (protospacer adjacent motif,PAM) 的短(3至5bp) DNA序列相邻的前间隔区。PAM在采集期间对I型和II型系统很重要。在I型和II型系统中,前间隔区在与PAM序列相邻的位置被切除,间隔区的另一端使用标尺机制切割,从而保持CRISPR阵列中间隔区尺寸的规律性。PAM序列的保守性在CRISPR-Cas系统之间有所不同,并且可以在进化上与Cas1和前导序列相关联。

[0077] 在一些实施方案中,本公开内容提供了将在细胞内产生的所限定的合成DNA,例如通过使用细胞内的工程化反转录子系统以定向方式整合到CRISPR阵列中,该整合优先但不排除他地邻近前导序列而发生。在来自大肠杆菌的I-E型系统中,证明了与前导序列相邻的第一个同向重复序列被复制,新获得的间隔区插入在第一个和第二个同向重复序列之间。

[0078] 在一个实施方案中,前间隔区是所限定的合成DNA。在一些实施方案中,所限定的合成DNA的长度为至少3、5、10、20、30、40或50个核苷酸,或3至50个、或10至100个、或20至90个、或30至80个、或40至70个、或50至60个核苷酸。在一个实施方案中,寡核苷酸序列或所限定的合成DNA包括经修饰的“AAG”前间区序列邻近基序 (PAM)。

[0079] 在一些实施方案中,调节元件与CRISPR系统的一种或更多种元件可操作地连接以驱动CRISPR系统的一种或更多种元件的表达。通常来说,CRISPR(成簇的规则间隔短回文重复),也称为SPIDR(间隔区散在同向重复),构成通常特定于特定细菌物种的DNA基因座家族。CRISPR基因座包含不同类的散在短序列重复 (short sequence repeat, SSR),其识别于大肠杆菌(Ishino et al., J. Bacteriol., 169:5429-5433 (1987); 和 Nakata et al., J. Bacteriol., 171:3553-3556 (1989)),以及相关基因。已在地中海富盐菌(Haloferax mediterranei)、酿脓链球菌(Streptococcus pyogenes)、鱼腥藻(Anabaena)和结核分支杆菌(Mycobacterium tuberculosis)中鉴定了类似的散在SSR(参见,Groenen et al., Mol. Microbiol., 10:1057-1065 (1993); Hoe et al., Emerg. Infect. Dis., 5:254-263 (1999); Masepohl et al., Biochim. Biophys. Acta 1307:26-30 (1996); 和 Mojica et al., Mol. Microbiol., 17:85-93 (1995))。CRISPR基因座通常与其他SSR的不同之处在于重复序列的结构,其被称为短规则间隔重复 (short regularly spaced repeat, SRSR) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 (2002); 和 Mojica et al., Mol. Microbiol., 36:244-246 (2000))。通常来说,重复是出现在簇中的短元件,这些簇被具有基本恒定长度的独特插入序列规则地隔开(Mojica et al., (2000), 同上)。尽管重复序列在菌株之间是高度保守的,但散在重复和间隔区序列的数量通常因菌株而异(van Embden et al., J. Bacteriol., 182:2393-2401 (2000))。已经在超过40种原核生物中鉴定了CRISPR基因座(参见例如,Jansen et al., Mol. Microbiol., 43:1565-1575 (2002); 和 Mojica et al., (2005)),其包括但不限于气火菌属(Aeropyrum)、火棒菌属(Pyrobaculum)、硫化叶菌属(Sulfolobus)、古球状菌属(Archaeoglobus)、Halocarcula、甲烷杆菌属(Methanobacterium)、甲烷球菌属(Methanococcus)、甲烷八叠球菌属(Methanosarcina)、甲烷嗜高热菌属(Methanopyrus)、热球菌属(Pyrococcus)、嗜酸菌属(Picrophilus)、Thernioplasnia、棒状杆菌属

(Corynebacterium)、分枝杆菌属(Mycobacterium)、链霉菌属(Streptomyces)、Aquifex、Porphyrimonas、绿菌属(Chlorobium)、栖热菌属(Thermus)、芽孢杆菌属(Bacillus)、李斯特菌属(Listeria)、葡萄球菌属(Staphylococcus)、梭菌属(Clostridium)、嗜热厌氧菌属(Thermoanaerobacter)、支原体(Mycoplasma)、梭杆菌属(Fusobacterium)、Azarcus、色杆菌属(Chromobacterium)、奈瑟菌属(Neisseria)、亚硝化单胞菌属(Nitrosomonas)、脱硫弧菌属(Desulfovibrio)、地杆菌属(Geobacter)、Myrococcus、弯曲菌属(Campylobacter)、沃廉菌属(Wolinella)、不动杆菌属(Acinetobacter)、欧文氏菌属(Erwinia)、埃希氏杆菌属(Escherichia)、军团菌属(Legionella)、甲基球菌属(Methylococcus)、巴氏杆菌属(Pasteurella)、发光杆菌属(Photobacterium)、沙门菌属(Salmonella)、黄单胞菌属(Xanthomonas)、耶尔森氏菌属(Yersinia)、密螺旋体属(Treponema)和热袍菌属(Thermotoga)。

[0080] 在一些实施方案中，编码CRISPR酶的酶编码序列被密码子优化用于在特定细胞例如真核细胞中表达。真核细胞可以是特定生物体的那些或者源自所述特定生物体，例如哺乳动物，其包括但不限于人、小鼠、大鼠、兔、狗或非人灵长类。通常来说，密码子优化是指修饰核酸序列用于在目的宿主细胞中增强表达的过程，该修饰通过用在该宿主细胞的基因中更频繁或最频繁使用的密码子来替换天然序列的至少一个密码子(例如约一个或多于约1、2、3、4、5、10、15、20、25、50或更多个密码子)，同时保持天然氨基酸序列。多种物种对特定氨基酸的某些密码子表现出特定的偏好。密码子偏倚(生物体之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关，而信使RNA(mRNA)的翻译效率又被认为取决于被翻译的密码子的性质和特定转移RNA(tRNA)分子的可用性等。细胞中所选tRNA的优势通常反映了肽合成中最常使用的密码子。因此，可以基于密码子优化来定制基因以在给定生物体中实现最佳基因表达。密码子使用表很容易获得，例如，在“密码子使用数据库”中，并且这些表可以以多种方式进行改编。参见Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases:status for the year 2000" Nucl. Acids Res. 28:292 (2000)。用于对特定序列进行密码子优化以在特定宿主细胞中表达的计算机算法也是可用的，例如Gene Forge(Aptagen; Jacobus, Pa.)也是可用的。在一些实施方案中，编码CRISPR酶的序列中的一个或更多个密码子(例如1、2、3、4、5、10、15、20、25、50或更多，或所有密码子)对应于最经常使用的用于特定氨基酸的密码子。

[0081] 向细胞“施用”核酸(例如工程化反转录子构建体或包含工程化反转录子构建体的载体)包含转导、转染、电穿孔、易位、融合、吞噬、射击(shooting)或弹道(ballistic)方法等，即，可以跨细胞膜转运核酸的任何方式。

[0082] 在进一步描述本公开内容之前，应当理解，所公开的主题不限于所描述的特定实施方案，因为其当然可变化。还应理解，本文中使用的术语仅为了描述特定实施方案的目的，而不旨在进行限制，这是由于本发明的范围将仅由所附权利要求书限制。

[0083] 当提供值的范围时，应理解，除非上下文另外明确指出，否则在该范围的上限和下限之间的每个中间值，直至下限单位的十分之一，以及在该规定范围内的任何其他规定值或中间值，涵盖在所公开的主题内。这些较小范围的上限和下限可独立地包含在较小范围内，并且也涵盖在所公开的主题内，受所规定范围中的任何明确排除的限制。当所规定范围包含一个或两个限制时，排除那些被包含的限制之一或二者的范围也包含在所公开的主题

内。

[0084] 除非另有定义,否则本文中使用的所有技术和科学术语具有与所公开主题所属领域的普通技术人员通常理解的相同含义。尽管与本文中描述的那些相似或等效的任何方法和材料也可以用于所公开主题的实践或测试中,但是现描述优选的方法和材料。本文中提及的所有出版物均通过引用并入本文中,以公开和描述与所引用出版物有关的方法和/或材料。

[0085] 必须注意,如本文中和所附权利要求书中所使用的,除非上下文另有明确指示,否则未用数量词修饰的名词包括一个/种或更多个/种。因此,例如,提及“细胞”包括多个这种细胞,提及“核酸”包括提及一种或更多种核酸及其本领域技术人员已知的其等价物,等等。还应注意,可以将权利要求书撰写为排除任何任选要素。因此,本声明旨在用作使用这种排他性术语如“唯一”、“仅”与本文中描述的任何特征或要素(其包括使用“否定”局限性)的引用相关的先行基础。

[0086] 应当理解,为清楚起见,在分开的实施方案的上下文中描述的所公开的主题的某些特征也可以在单个实施方案中组合提供。相反,为简洁起见,在单个实施方案的上下文中描述的所公开主题的多种特征也可以单独提供或以任何合适的子组合提供。与本公开内容有关的实施方案的所有组合都特别包含在所公开的主题中并在本文中公开,就好像每个和每种组合都被单独且明确地公开一样。此外,多种实施方案及其要素的所有子组合也被本公开内容特别地包括并且在本文中公开,就好像每个和每种这样的子组合在本文中单独且明确地公开一样。

[0087] 本文中所讨论的出版物仅出于其在本申请的提交日期之前的公开内容而提供。本文中的任何内容均不得理解为承认所公开主题无权先于这样的出版物。此外,提供的公布日期可能与实际公布日期有所不同,所述实际公布日期可能需要独立确认。

具体实施方式

[0088] 提供了经修饰以增强多拷贝单链DNA (multicopy single-stranded DNA, msDNA) 的产生的工程化反转录子。此外,还提供了编码这种工程化反转录子的载体系统和在多种应用例如CRISPR/Cas介导的基因组编辑、重组工程、细胞条码化和分子记录中使用工程化反转录子和编码其的载体系统的方法。

[0089] 工程化反转录子

[0090] 本公开内容提供了工程化反转录子,其被修饰以增强细胞中msDNA的产生。工程化反转录子包含msr前序列,编码多拷贝单链RNA (msRNA) 的msr基因;编码多拷贝单链DNA (msDNA) 的msd基因;msd后序列和编码逆转录酶的ret基因。通过反转录子编码的逆转录酶结果合成DNA可以提供由msd基因编码的单链DNA与由msr基因编码的单链RNA连接构成的DNA/RNA嵌合产物。反转录子msr RNA在茎环结构的末端处含有保守的鸟苷残基。msr RNA的链通过该保守的鸟苷残基2'位置处的2'-5'磷酸二酯键与msd单链DNA的5'末端连接。

[0091] 在工程化反转录子中,例如,msd后序列在其自身互补区(其与msr前序列具有序列互补性)内被修饰,其中自身互补区的长度相对于原生反转录子的相应区而加长。这种修饰导致了提供增强的msDNA产生的工程化反转录子。在某些实施方案中,互补区的长度比野生型自身互补区长至少1、至少2、至少4、至少6、至少8、至少10、至少12、至少14、至少16、至少

18、至少20、至少30、至少40或至少50个核苷酸。例如，自身互补区可具有比天然或野生型互补区长1至50个核苷酸的长度，其包括该范围内的任何长度，例如长1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸。在某些实施方案中，自身互补区的长度比野生型互补区长1至16个核苷酸。由工程化反转录子生成的单链DNA可用于多种应用。

[0092] 例如，为了创建更丰富的ssDNA，如下所示的ncRNA SEQ ID N0:12序列可以在1和169位处延伸，其中天然自身互补的3'和5'末端以粗体突出显示（在1至12和158至169位处）。

[0093]

1	TGCGCACCC	TAGCGAGAGG	TTTATCATTA	AGGTCAACCT
41	CTGGATGTTG	TTTCGGCATIC	CTGCATTGAA	TCTGAGTTAC
81	TGTCTGTTT	CCTTGTTGGA	ACGGAGAGCA	TCGCCTGATG
161	CTCTCCGAGC	CAACCAGGAA	ACCCGTTTT	TCTGACGTAA
201	GGGTGCGCA			

[0094] 例如，如下所示，对于以下工程化“延伸的ncRNA”（SEQ ID N0:13），其中延伸自身互补区的另外核苷酸以斜体并带有下划线显示。

[0095]

1	<u>TGATAAGATT</u>	<u>CCGTATGCGC</u>	ACCCTTAGCG	AGAGGTTTAT
41	CATTAAGGTC	AACCTCTGGA	TGTTGTTTCG	GCATCCTGCA
81	TTGAATCTGA	GTTACTGTCT	GTTTCCTTG	TTGGAACGGA
121	GAGCATCGCC	TGATGCTCTC	CGAGCCAACC	AGGAAACCCG
161	TTTTTCTGA	<u>CGTAAGGGTG</u>	<u>CGCA</u> <u>TACGGA</u>	<u>ATCTTATCA</u>

[0096] 在一些情况下，可以将另外的核苷酸添加至自身互补区中的任何位置，例如，在SEQ ID N0:12序列的1至12和158至169位内的任何位置。

[0097] 在某些实施方案中，工程化反转录子中使用的msr基因、msd基因和ret基因的序列可以源自细菌反转录子操纵子。代表性的反转录子是可用的，例如来自革兰氏阴性菌的反转录子，其包括但不限于黏细菌反转录子，例如黄色黏球菌(*Myxococcus xanthus*)反转录子(例如,Mx65、Mx162)和橙色标桩菌(*Stigmatella aurantiaca*)反转录子(例如,Sa163)；大肠杆菌反转录子(例如,Ec48、E67、Ec73、Ec78、EC83、EC86、EC107和Ec107)；肠沙门氏菌(*Salmonella enterica*)；霍乱弧菌(*Vibrio cholerae*)反转录子(例如,Vc81、Vc95、Vc137)；副溶血性弧菌(*Vibrio parahaemolyticus*) (例如,Vc96)；和侵蚀侏囊菌(*Nannocystis exedens*)反转录子(例如,Ne144)。反转录子msr基因、msd基因和ret基因核酸序列以及反转录子逆转录酶蛋白质序列可以源自任何来源。代表性的反转录子序列，其中包括msr基因、msd基因、ret基因核酸序列和逆转录酶蛋白序列均列于美国国家生物技术信息中心(National Center for Biotechnology Information,NCBI)数据库中。例如，参见NCBI条目：登录号EF428983、M55249、EU250030、X60206、X62583、AB299445、AB436696、AB436695、M86352、M30609、M24392、AF427793、AQ3354和AB079134；所有这些序列(如在截至本申请提交之日输入的)通过引用其整体并入本文中。任何这些反转录子序列或其包含序列的变体可包括变体核苷酸、添加的核苷酸或更少的核苷酸。例如，反转录子可以与其具有至少约80%至100%的序列同一性，其包括该范围内的任何百分比同一性，例如与本文所述的任何反转录子序列(包括由登录号限定的那些序列)具有81%、82%、83%、84%、85%、

86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的序列同一性，并且可用于构建工程化反转录子或包含工程化反转录子的载体系统，如本文所述。

[0098] 在一些实施方案中，重组反转录子构建体具有在msr基因、msd基因和ret基因之间具有非天然间隔的非天然构型。msr基因和msd基因可以以反式排列分离，而不是以内源顺式排列提供。此外，可以相对于msr基因或msd基因以反式排列提供ret基因。在一些实施方案中，以消除用于逆转录酶的隐蔽终止信号的反式排列提供ret基因，这允许从工程化反转录子构建体生成更长的单链DNA。

[0099] 在一些实施方案中，反转录子构建体相对于天然反转录子进行修饰以包括目的异源序列。在这种背景下，可以用异源序列对反转录子进行工程化以用于多种应用。例如，可以将异源序列添加至反转录子构建体以向细胞提供：编码目的蛋白质或调节性RNA的核酸，适合在基因编辑中使用（例如通过同源定向修复（homology directed repair, HDR）或重组介导的基因工程（重组工程））的供体多核苷酸，或用于分子记录的CRISPR前间隔区DNA序列，如下文进一步所讨论的。可以将这样的异源序列插入到例如msr基因或msd基因中，使得异源序列通过反转录子逆转录酶转录为msDNA产物的一部分。

[0100] 在一些情况下，可以将目的异源序列插入msd茎环的环中。

[0101] 例如，工程化反转录子可以包括独特的条码以促进多路复用（multiplexing）。条码可包含一个或更多个核苷酸序列，其可用于识别与条码相关的核酸或细胞。这样的条码可以被插入到例如msd编码的DNA的环区中。条码长度可以是3至1000个或更多个核苷酸，优选长度为10至250个核苷酸，更优选长度为10至30个核苷酸，其包括这些范围内的任何长度，例如长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个核苷酸。在一些实施方案中，条码也可用于识别反转录子起源的细胞、集落或样品的位置（即位置条码），例如细胞阵列中集落的位置、在多孔板中孔的位置、支架中试管的位置或实验室中样品的位置。特别地，条码可用于识别含有反转录子的经遗传修饰的细胞的位置。条码的使用允许将来自不同细胞的反转录子汇集在单个反应混合物中进行测序，同时仍然能够将特定的反转录子追溯到其起源的集落。

[0102] 此外，可以将衔接子序列添加至反转录子构建体以促进高通量扩增或测序。例如，可以在反转录子构建体的5'和3'端添加衔接子序列以允许通过同一组引物同时扩增或测序多个反转录子构建体。例如，可以在细胞转染或连接至载体之前进行反转录子构建体的扩增。可以使用用于扩增反转录子构建体的任何方法，其包括但不限于聚合酶链反应（polymerase chain reaction, PCR）、恒温扩增、基于核酸序列的扩增（nucleic acid sequence-based amplification, NASBA）、转录介导的扩增（transcription mediated amplification, TMA）、链置换扩增（strand displacement amplification, SDA）和连接酶链反应（ligase chain reaction, LCR）。在一个实施方案中，反转录子构建体包含共同的5'和3'引发位点，以允许与一组通用引物平行地扩增反转录子序列。在另一个实施方案中，一组选择性引物可用于从合并的混合物中选择性地扩增反转录子序列的子集。

[0103] 将工程化反转录子递送至细胞通常在使用或不使用载体的情况下完成。可以将工程化反转录子（或含有其的载体）引入任何类型的细胞中，所述细胞包括来自原核生物、真核生物或古细菌生物体（包括细菌、古细菌、真菌、原生生物、植物（例如，单子叶植物和双子

叶植物)；和动物(例如脊椎动物和无脊椎动物))的任何细胞。可以用工程化反转录子转染的动物的实例包括但不限于脊椎动物例如鱼、鸟类、哺乳动物(例如人和非人灵长类、农场动物、宠物和实验室动物)、爬行动物和两栖动物。可以用工程化反转录子转染的植物的实例包括但不限于农作物，其包括谷物例如小麦、燕麦和稻米，豆类例如大豆和豌豆，玉米，草类例如紫花苜蓿，以及棉花。工程化反转录子可以被引入至单个细胞或目的细胞群中。来自组织、器官和活检的细胞，以及重组细胞、经遗传修饰的细胞、来自体外培养的细胞系的细胞和人工细胞(例如，纳米颗粒、脂质体、聚合体或微胶囊，其包封核酸)都可以用工程化反转录子被转染。本方法也适用于细胞片段、细胞组分或细胞器(例如，动物和植物细胞中的线粒体，植物细胞和藻类中的质体(例如，叶绿体))。细胞可以在用工程化反转录子构建体转染之后进行培养或扩增。

[0104] 将核酸引入宿主细胞中的方法在本领域中是公知的。通常使用的方法包括化学诱导转化(通常使用二价阳离子(例如,CaCl₂))、右旋糖酐介导的转染、聚凝胺介导的转染、脂质体和LT-1介导的转染、电穿孔、原生质体融合、将核酸包封在脂质体中和将包含工程化反转录子的核酸直接显微注射至细胞核中。参见，例如，Sambrook et al. (2001) Molecular Cloning, a laboratory manual, 3rd edition, Cold Spring Harbor Laboratories, New York, Davis et al. (1995) Basic Methods in Molecular Biology, 2nd edition, McGraw-Hill，以及Chu et al. (1981) Gene 13:197；通过引用其整体并入本文中。

[0105] 包含工程化反转录子的载体系统

[0106] 在某些实施方案中，反转录子msr基因、msd基因和ret基因在细胞中由载体体内表达。“载体”是可用于将目的核酸递送至细胞内部的物质结构。可以将反转录子msr基因、msd基因和ret基因用单个载体或以多个单独载体引入到细胞中以在宿主对象中产生msDNA。载体通常包括与反转录子序列可操作地连接的控制元件，这允许在对象物种中体内产生msDNA。例如，反转录子msr基因、msd基因和ret基因可以与启动子可操作地连接以允许反转录子逆转录酶和msDNA产物的表达。在一些实施方案中，可将编码所期望的目的产物的异源序列(例如，编码多肽或调节性RNA的多核苷酸、用于基因编辑的供体多核苷酸或用于分子记录的前间隔区DNA)插入至msr基因或msd基因中。任何能够用包含工程化反转录子序列的载体转染的真核细胞、古细菌细胞或原核细胞都可以用于产生msDNA。可以凭经验确定构建体产生msDNA以及其他反转录子编码的产物的能力。

[0107] 在一些实施方案中，工程化反转录子由包含一种或更多种载体的载体系统产生。在载体系统中，msr基因、msd基因和ret基因可以由同一载体提供(即，所有这种反转录子元件的顺式排列)，其中载体包含与msr基因和msd基因可操作地连接的启动子。在一些实施方案中，启动子还与ret基因可操作地连接。在另一些实施方案中，载体还包含与ret基因可操作地连接的第二启动子。或者，ret基因可以由不包括msr基因和msd基因的第二载体提供(即msr-msd和ret的反式排列)。在另一些实施方案中，msr基因、msd基因和ret基因各自由不同的载体提供(即所有反转录子元件的反式排列)。许多载体是可用的，其包括但不限于线性多核苷酸、与离子或两亲化合物缔合的多核苷酸、质粒和病毒。因此，术语“载体”包括自主复制的质粒或病毒。病毒载体的实例包括但不限于腺病毒载体、腺相关病毒载体、逆转录病毒载体、慢病毒载体等。表达构建体可以在活细胞中复制，其也可以被合成地制备。为了本申请的目的，术语“表达构建体”、“表达载体”和“载体”可互换使用以在一般性、说明性

的意义上表明本发明的应用，并且不旨在限制本发明。

[0108] 在某些实施方案中，包含工程化反转录子序列的核酸在启动子的转录控制下。“启动子”是指由细胞的合成机制或引入的合成机制识别的DNA序列，是启动基因的特异性转录所需的。这里将使用术语启动子来指代聚集在RNA聚合酶I、II或III的起始位点周围的一组转录控制模块。用于哺乳动物细胞表达的典型启动子包括SV40早期启动子、CMV启动子例如CMV即早期启动子(参见美国专利号5,168,062和5,385,839,通过引用其整体并入本文中)、小鼠乳腺肿瘤病毒LTR启动子、腺病毒主要晚期启动子(*adenovirus major late promoter*, AdMLP)和单纯疱疹病毒启动子等。其他非病毒启动子，例如源自鼠金属硫蛋白基因的启动子，也可用于哺乳动物表达。这些和其他启动子可以使用本领域中公知的技术从可商购的质粒中获得。参见，例如，Sambrook et al., 同上。增强子元件可以与启动子联合使用以提高构建体的表达水平。实例包括：SV40早期基因增强子，如在Dijkema et al., EMBO J. (1985) 4: 761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., Proc. Natl. Acad. Sci. USA (1982b) 79: 6777中所述，以及来源于人巨细胞病毒的元件，如Boshart et al., Cell (1985) 41: 521所述，例如包含在CMV内含子A序列中的元件。

[0109] 在一个实施方案中，用于表达包含msr基因、msd基因和ret基因的工程化反转录子的表达载体包含与编码msr基因、msd基因和ret基因的多核苷酸“可操作地连接”的启动子。如本文所用的短语“可操作地连接”或“在转录控制下”意指启动子相对于多核苷酸处于正确的位置和取向，以控制通过RNA聚合酶的转录起始和msr基因、msd基因和ret基因的表达。

[0110] 通常来说，转录终止子/多聚腺苷酸化信号也将存在于表达构建体中。这种序列的实例包括但不限于源自SV40的序列，如Sambrook et al., 同上所述，以及牛生长激素终止子序列(参见，例如，美国专利号5,122,458)。此外，5'-UTR序列可以与编码序列相邻放置以增强其表达。这种序列可以包括包含内部核糖体进入位点(internal ribosome entry site, IRES)的UTR。

[0111] 包含IRES允许从载体翻译一个或更多个开放阅读框。IRES元件吸引真核核糖体翻译起始复合体并促进翻译起始。参见，例如，Kaufman et al., Nuc. Acids Res. (1991) 19: 4485-4490; Gurtu et al., Biochem. Biophys. Res. Comm. (1996) 229: 295-298; Rees et al., BioTechniques (1996) 20: 102-110; Kobayashi et al., BioTechniques (1996) 21: 399-402; 和Mosser et al., BioTechniques (1997) 22: 150-161。许多IRES序列是已知的并且包括这样的序列，其源自广泛多种病毒，例如来自小核糖核酸病毒例如脑心肌炎病毒(*encephalomyocarditis virus*, EMCV)UTR的前导序列(Jang et al. J. Virol. (1989) 63: 1651-1660)、脊髓灰质炎前导序列、甲型肝炎病毒前导序列、丙型肝炎病毒IRES、人鼻病毒2型IRES(Dobrikova et al., Proc. Natl. Acad. Sci. (2003) 100 (25): 15125-15130)、来自口蹄疫病毒的IRES元件(Ramesh et al., Nucl. Acid Res. (1996) 24: 2697-2700)、贾第病毒IRES(Garlapati et al., J. Biol. Chem. (2004) 279 (5): 3389-3397)等。多种非病毒IRES序列也将用于本文中，其包括但不限于来自酵母的IRES序列，以及人血管紧张素II型1受体IRES(Martin et al., Mol. Cell Endocrinol. (2003) 212: 51-61)，成纤维细胞生长因子IRES(FGF-1IRES和FGF-2IRES, Martineau et al. (2004) Mol. Cell. Biol. 24 (17): 7622-7635)，血管内皮生长因子IRES(Baranick et al. (2008) Proc. Natl. Acad. Sci. U.S.A. 105

(12) :4733-4738, Stein et al. (1998) Mol. Cell. Biol. 18 (6) :3112-3119, Bert et al. (2006) RNA 12 (6) :1074-1083), 和胰岛素样生长因子2IRES (Pedersen et al. (2002) Biochem. J. 363 (Pt 1) :37-44)。这些元件在出售的质粒中很容易商购获得,例如通过 Clontech (Mountain View, CA)、Invivogen (San Diego, CA)、Addgene (Cambridge, MA) 和 GeneCopoeia (Rockville, MD)。另见 IRESite: 经实验验证的 IRES 结构数据库 (iresite.org)。IRES 序列可以包含在载体中,例如,以表达用于重组工程的多个噬菌体重组蛋白或用于 HDR 的 RNA 指导的核酸酶(例如,Cas9)与来自表达盒的反转录子逆转录酶组合。

[0112] 或者,编码病毒T2A肽的多核苷酸可用于允许从单个载体产生多种蛋白质产物(例如,Cas9、噬菌体重组蛋白、反转录子逆转录酶)。可以在多顺反子构建体的编码序列之间插入一个或更多个2A接头肽。自切割的2A肽允许来自多顺反子构建体的共表达蛋白质以等摩尔水平产生。可以使用来自多种病毒的2A肽,其包括但不限于源自口蹄疫病毒、马鼻炎A病毒、Thesaa asigna 病毒和猪捷申病毒-1 的2A肽。参见,例如,Kim et al. (2011) PLoS One 6 (4) :e18556, Trichas et al. (2008) BMC Biol. 6:40, Provost et al. (2007) Genesis 45 (10) :625-629, Furler et al. (2001) Gene Ther. 8 (11) :864-873; 其整体通过引用并入本文中。

[0113] 在某些实施方案中,表达构建体包括适合转化细菌宿主的质粒。许多细菌表达载体是本领域技术人员已知的,并且合适载体的选择是选项问题。细菌表达载体包括但不限于 pACYC177、pASK75、pBADM、pBAT、pCal、pET、pETM、pGAT、pGEX、pHAT、pKK223、pMal、pProEx、pQE 和 pZA31。细菌质粒可包含抗生素选择标志物(例如,氨苄青霉素、卡那霉素、红霉素、羧苄青霉素、链霉素或四环素抗性)、lacZ 基因(β -半乳糖苷酶从 x-gal 底物产生蓝色色素)、荧光标志物(例如 GFP.mCherry)或其他用于选择转化的细菌的标志物。参见,例如,Sambrook et al., 同上。

[0114] 在另一些实施方案中,表达构建体包括适合转化酵母细胞的质粒。酵母表达质粒通常含有酵母特异性复制起点(origin of replication, ORI) 和营养选择标志物(例如,HIS3、URA3、LYS2、LEU2、TRP1、MET15、ura4+、leu1+、ade6+)、抗生素选择标志物(例如,卡那霉素抗性)、荧光标志物(例如,mCherry)或其他用于选择转化的酵母细胞的标志物。酵母质粒还可包含允许在细菌宿主(例如,大肠杆菌)和酵母细胞之间穿梭的组分。许多不同类型的酵母质粒是可用的,包括:酵母整合质粒(yeast integrating plasmid, YIp),其缺乏 ORI 并通过同源重组整合到宿主染色体中;酵母复制质粒(yeast replicating plasmid, YRp),其含有自主复制序列(autonomously replicating sequence, ARS) 并且可以独立复制;酵母着丝粒质粒(yeast centromere plasmid, YCp),其是含有 ARS 的部分和着丝粒序列(CEN) 的部分的低拷贝载体;以及酵母附加体质粒(yeast episomal plasmid, YEp),其是高拷贝数质粒,包含来自 2 微米圆(天然酵母质粒)的片段(允许每个细胞稳定繁殖 50 个或更多个拷贝)。

[0115] 在另一些实施方案中,表达构建体包含病毒或源自病毒基因组的工程化构建体。已经开发了许多基于病毒的系统用于将基因转移至哺乳动物细胞中。这些包括腺病毒、逆转录病毒(γ -逆转录病毒和慢病毒)、痘病毒、腺相关病毒、杆状病毒和单纯疱疹病毒(参见例如 Warnock et al. (2011) Methods Mol. Biol. 737:1-25; Walther et al. (2000) Drugs

60 (2) :249-271; 和 Lundstrom (2003) Trends Biotechnol. 21 (3) :117-122; 其通过引用整体并入本文中)。某些病毒通过受体介导的内吞作用进入细胞、整合到宿主细胞基因组中并稳定且有效地表达病毒基因的能力使其成为将外源基因转移到哺乳动物细胞中的有吸引力的候选者。

[0116] 例如,逆转录病毒为基因递送系统提供了便捷的平台。可以使用本领域已知的技术将选定的序列插入至载体中并包装到逆转录病毒颗粒中。然后可以在体内或离体分离重组病毒并将其递送至对象的细胞。已经描述了许多逆转录病毒系统(美国专利号5,219,740; Miller and Rosman (1989) BioTechniques 7:980-990; Miller, A.D. (1990) Human Gene Therapy 1:5-14; Scarpa et al. (1991) Virology 180:849-852; Burns et al. (1993) Proc. Natl. Acad. Sci. USA 90:8033-8037; Boris-Lawrie and Temin (1993) Cur. Opin. Genet. Develop. 3:102-109; 以及 Ferry et al. (2011) Curr. Pharm. Des. 17 (24) : 2516-2527)。慢病毒是一类对于将多核苷酸递送至哺乳动物细胞特别有用的逆转录病毒,因为其能够感染分裂细胞和非分裂细胞两者(参见例如Lois et al (2002) Science 295: 868-872; Durand et al. (2011) Viruses 3 (2) :132-159; 其通过引用并入本文中)。

[0117] 还已描述了许多腺病毒载体。与整合到宿主基因组中的逆转录病毒不同,腺病毒在染色体外持续存在,因此使与插入诱变相关的风险最小化(Haj-Ahmad and Graham, J. Virol. (1986) 57:267-274; Bett et al., J. Virol. (1993) 67:5911-5921; Mittereder et al., Human Gene Therapy (1994) 5:717-729; Seth et al., J. Virol. (1994) 68:933-940; Barr et al., Gene Therapy (1994) 1:51-58; Berkner, K.L. BioTechniques (1988) 6:616-629; 以及 Rich et al., Human Gene Therapy (1993) 4:461-476)。此外,已经开发了多种腺相关病毒(adeno-associated virus, AAV)载体系统用于基因递送。可以使用本领域熟知的技术容易地构建AAV载体。参见,例如,美国专利号5,173,414和5,139,941; 国际公开号W092/01070(1992年1月23日公开)和W093/03769(1993年3月4日公开); Lebkowski et al., Molec. Cell. Biol. (1988) 8:3988-3996; Vincent et al., Vaccines 90 (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. Current Opinion in Biotechnology (1992) 3:533-539; Muzyczka, N. Current Topics in Microbiol. and Immunol. (1992) 158: 97-129; Kotin, R.M. Human Gene Therapy (1994) 5:793-801; Shelling and Smith, Gene Therapy (1994) 1:165-169; 以及 Zhou et al., J. Exp. Med. (1994) 179:1867-1875。

[0118] 可用于递送编码工程化反转录子的核酸的另一种载体系统是由Small, Jr., P.A. 等(1997年10月14日授权的美国专利号5,676,950, 其通过引用并入本文中)描述的肠内施用的重组痘病毒疫苗。

[0119] 将可用于递送目的核酸分子的另外的病毒载体包括源自痘病毒家族的病毒载体,其包括痘苗病毒和禽痘病毒。举例来说,表达目的核酸分子(例如工程化反转录子)的痘苗病毒重组体可以如下构建。首先将编码特定核酸序列的DNA插入适当的载体中,使得其与牛痘启动子相邻并为牛痘DNA序列(例如编码胸苷激酶(thymidine kinase, TK)的序列)的侧翼。然后将该载体用于转染同时感染牛痘的细胞。同源重组用于将牛痘启动子加上编码目的序列的基因插入到病毒基因组中。可以通过在存在5-溴脱氧尿苷的情况下培养细胞并挑选对其有抗性的病毒斑来选择所得的TK-重组体。

[0120] 或者,禽痘病毒,例如家禽痘病毒(fowlpox)和金丝雀痘病毒也可用于递送目的核

酸分子。禽痘载体的使用在人和其他哺乳动物物种中是特别期望的,因为禽痘属的成员只能在易感鸟类物种中进行生产性复制,因此在哺乳动物细胞中没有感染性。用于产生重组禽痘病毒的方法在本领域中是已知的,并且采用基因重组,如上文关于痘苗病毒的产生所述。参见,例如,W091/12882;W089/03429;和W092/03545。

[0121] 分子共轭载体,例如在Michael et al., J.Biol.Chem. (1993) 268:6866-6869和Wagner et al., Proc.Natl.Acad.Sci.USA (1992) 89:6099-6103中描述的腺病毒嵌合载体,也可用于基因递送。

[0122] 甲病毒属的成员,例如但不限于源自辛德比斯病毒 (Sindbis virus, SIN)、塞姆利基森林病毒 (Semliki Forest virus, SFV) 和委内瑞拉马脑炎病毒 (Venezuelan Equine Encephalitis virus, VEE) 的载体,也可用作病毒载体以递送本发明的多核苷酸。对于可用于实践本方法的源自辛德比斯病毒的载体的描述,参见Dubensky et al. (1996) J.Virol.70:508-519;和国际公开号W095/07995、W096/17072;以及1998年12月1日授权的Dubensky, Jr., T.W.等的美国专利号5,843,723和1998年8月4日授权的Dubensky, Jr., T.W.的美国专利号5,789,245,其均通过引用并入本文中。特别优选的是由源自辛德比斯病毒和委内瑞拉马脑炎病毒的序列构成的嵌合甲病毒载体。参见,例如,Perri et al. (2003) J.Virol.77:10394-10403和国际公开号W002/099035、W002/080982、W001/81609和W000/61772;其通过引用整体并入本文中。

[0123] 基于牛痘的感染/转染系统可以方便地用于在宿主细胞中提供目的核酸(例如工程化反转录子)的可诱导、瞬时表达。在该系统中,首先用编码噬菌体T7 RNA聚合酶的痘苗病毒重组体体外感染细胞。这种聚合酶显示出强烈的特异性,因为其只转录带有T7启动子的模板。在感染之后,在T7启动子的驱动下,用目的核酸转染细胞。来自痘苗病毒重组体的细胞质中表达的聚合酶将转染的DNA转录成RNA。该方法提供了大量RNA的高水平、瞬时、细胞质中的产生。参见,例如,Elroy-Stein and Moss, Proc.Natl.Acad.Sci.USA (1990) 87: 6743-6747;Fuerst et al., Proc.Natl.Acad.Sci.USA (1986) 83:8122-8126。

[0124] 作为用痘苗病毒或禽痘病毒重组体进行感染或使用其他病毒载体递送核酸的替代方法,可以使用扩增系统,其将在引入到宿主细胞中之后导致高水平表达。具体而言,可以对位于T7 RNA聚合酶编码区之前的T7RNA聚合酶启动子进行工程化。源自该模板的RNA的翻译将产生T7 RNA聚合酶,其反过来将转录更多模板。同时,会有其表达受T7启动子的控制的cDNA。因此,由扩增模板RNA的翻译生成的T7 RNA聚合酶中的一些将导致所期望基因的转录。因为需要一些T7 RNA聚合酶来启动扩增,所以可以将T7 RNA聚合酶与模板一起引入到细胞中以引发转录反应。聚合酶可以作为蛋白质或在编码RNA聚合酶的质粒上被引入。关于T7系统及其用于转化细胞的用途的进一步讨论,参见例如,国际公开号W094/26911; Studier and Moffatt, J.Mol.Biol. (1986) 189:113-130; Deng and Wolff, Gene (1994) 143:245-249; Gao et al., Biochem.Biophys.Res.Commun. (1994) 200:1201-1206; Gao and Huang, Nuc.Acids Res. (1993) 21:2867-2872; Chen et al., Nuc.Acids Res. (1994) 22: 2114-2120;以及美国专利号No.5,135,855。

[0125] 昆虫细胞表达系统,例如杆状病毒系统,也可以被使用并且是本领域技术人员已知的,并且描述于例如Baculovirus and Insect Cell Expression Protocols (Methods in Molecular Biology, D.W.Murhammer ed., Humana Press, 2nd edition, 2007) 和L.King

The Baculovirus Expression System:A laboratory guide (Springer,1992) 中。用于杆状病毒/昆虫细胞表达系统的材料和方法可以以试剂盒形式从 inter alia, Thermo Fisher Scientific (Waltham, MA) 和 Clontech (Mountain View, CA) 商购获得。

[0126] 植物表达系统也可用于转化植物细胞。通常,这样的系统使用基于病毒的载体来用异源基因转染植物细胞。对于这样的系统的描述,参见例如 Porta et al., Mol. Biotech. (1996) 5:209-221; 和 Hackland et al., Arch. Virol. (1994) 139:1-22。

[0127] 为了实现工程化反转录子构建体的表达,必须将表达构建体递送至细胞中。这种递送可以在体外完成,如在转化细胞系的实验室程序中,或在体内或离体完成,如在某些疾病状态的治疗中。一种用于递送的机制是通过病毒感染,在此情况下表达构建体被包封在感染性病毒颗粒中。

[0128] 还设想了数种将表达构建体转移至培养细胞中的非病毒方法。这些包括使用磷酸钙沉淀、DEAE-右旋糖酐、电穿孔、直接显微注射、载有DNA的脂质体、lipofectamine-DNA复合体、细胞超声处理、使用高速微弹的基因轰击和受体介导的转染(参见,例如, Graham and Van Der Eb (1973) Virology 52:456-467; Chen and Okayama (1987) Mol. Cell Biol. 7: 2745-2752; Rippe et al. (1990) Mol. Cell Biol. 10:689-695; Gopal (1985) Mol. Cell Biol. 5:1188-1190; Tur-Kaspa et al. (1986) Mol. Cell. Biol. 6:716-718; Potter et al. (1984) Proc. Natl. Acad. Sci. USA 81:7161-7165; Harland and Weintraub (1985) J. Cell Biol. 101:1094-1099; Nicolau&Sene (1982) Biochim. Biophys. Acta 721:185-190; Fraley et al. (1979) Proc. Natl. Acad. Sci. USA 76:3348-3352; Fechheimer et al. (1987) Proc. Natl. Acad. Sci. USA 84:8463-8467; Yang et al. (1990) Proc. Natl. Acad. Sci. USA 87: 9568-9572; Wu and Wu (1987) J. Biol. Chem. 262:4429-4432; Wu and Wu (1988) Biochemistry 27:887-892; 其通过引用并入本文中)。这些技术中的一些技术可以成功地适用于体内或离体使用。

[0129] 一旦表达构建体已经被递送至细胞中,包含工程化反转录子序列的核酸可以在不同位点定位和表达。在某些实施方案中,包含工程化反转录子序列的核酸可以稳定地整合到细胞的基因组中。这种整合可以通过同源重组(基因替换)在同源位置和取向上进行,或者可以在随机的非特异性位置(基因增强)中进行整合。在另一些实施方案中,核酸可以作为单独的附加体DNA片段在细胞中稳定地维持。这样的核酸片段或“附加体”编码足以允许独立于宿主细胞周期的或与宿主细胞周期同步的维持和复制的序列。表达构建体如何被递送至细胞以及在细胞中核酸保留的位置取决于所使用的表达构建体的类型。

[0130] 在又一个实施方案中,表达构建体可简单地由包含工程化反转录子的裸重组DNA或质粒组成。构建体的转移可以通过物理地或化学地使细胞膜透化的任何上述方法进行。这特别适用于体外转移,但也可应用于体内使用。Dubensky等(Proc. Natl. Acad. Sci. USA (1984) 81:7529-7533)成功地将磷酸钙沉淀形式的多瘤病毒DNA注射到成年和新生小鼠的肝脏和脾脏中,其证明了激活的病毒复制和急性感染。Benvenisty&Neshif (Proc. Natl. Acad. Sci. USA (1986) 83:9551-9555)也证明了直接腹膜内注射磷酸钙沉淀的质粒导致转染基因的表达。设想了编码目的工程化反转录子的DNA也可以以类似的方式在体内转移并表达反转录子产物。

[0131] 在另一个实施方案中,可以通过颗粒轰击将裸DNA表达构建体转移至细胞中。该方

法依赖于将包被DNA的微弹加速至高速以使其能够刺穿细胞膜并进入细胞而不杀伤细胞的能力(Klein et al. (1987) Nature 327:70-73)。已经开发了数种用于加速小颗粒的装置。一种这样的装置依靠高压放电来生成电流,而电流进而提供了动力(Yang et al. (1990) Proc.Natl.Acad.Sci.USA 87:9568-9572)。微弹可以由生物惰性物质例如钨或金珠组成。

[0132] 在另一些的实施方案中,可以使用脂质体递送表达构建体。脂质体是以磷脂双层膜和内部水性介质为特征的囊泡结构。多层脂质体具有由水性介质隔开的多个脂质层。其当磷脂悬浮在过量的水溶液中时自发形成。脂质组分在形成封闭结构之前进行自我重排,并在脂质双层之间截留水和溶解的溶质(Ghosh&Bachhawat (1991) Liver Diseases, Targeted Diagnosis and Therapy Using Specific Receptors and Ligands,Wu et al. (Eds.) ,Marcel Dekker, NY, 87-104)。还考虑使用lipofectamine-DNA复合体。

[0133] 在某些实施方案中,脂质体可以与血凝病毒(HVJ)复合。已显示这促进与细胞膜的融合并促进脂质体包封的DNA进入细胞(Kaneda et al. (1989) Science 243:375-378)。在另一些实施方案中,脂质体可以与核非组蛋白染色体蛋白(HMG-I)复合或与其结合使用(Kato et al. (1991) J.Biol.Chem.266 (6) :3361-3364)。在另一些的实施方案中,脂质体可以与HVJ和HMG-I两者复合或与其结合使用。由于这样的表达构建体已成功用于在体外和体内的核酸转移和表达,因此其适用于本发明。在DNA构建体中使用细菌启动子的情况下,还期望在脂质体中包括适当的细菌聚合酶。

[0134] 可用于将核酸递送至细胞中的其他表达构建体是受体介导的递送载剂。这些利用了几乎所有真核细胞中受体介导的内吞作用对大分子的选择性摄取。由于多种受体的细胞类型特异性分布,递送可以是高度特异性的(Wu and Wu (1993) Adv.Drug Delivery Rev.12:159-167)。

[0135] 受体介导的基因靶向载剂通常由两种组分组成:细胞受体特异性配体和DNA结合剂。数种配体已用于受体介导的基因转移。最广泛表征的配体是无唾液酸血清类黏蛋白(asialoorosomucoid, ASOR)和转铁蛋白(参见例如Wu and Wu (1987),同上;Wagner et al. (1990) Proc.Natl.Acad.Sci.USA 87 (9) :3410-3414)。一种识别与ASOR相同的受体的合成的新糖蛋白已被用作基因递送载剂(Ferkol et al. (1993) FASEB J.7:1081-1091;Perales et al. (1994) Proc.Natl.Acad.Sci.USA 91 (9) :4086-4090),并且表皮生长因子(epidermal growth factor, EGF)也已用于将基因递送至鳞状癌细胞(Myers, EP0 0273085)。

[0136] 在另一些实施方案中,递送载剂可包含配体和脂质体。例如,Nicolau等(Methods Enzymol. (1987) 149:157-176)使用乳糖基神经酰胺、半乳糖末端无唾液酸神经节苷脂,并入脂质体中并观察到肝细胞对胰岛素基因的摄取提高。因此,以下是可行的:编码特定基因的核酸也可以通过任何数量的具有或不具有脂质体的受体-配体系统而特异地递送至细胞中。此外,细胞表面抗原的抗体也可以类似地用作靶向部分。

[0137] 在一个具体的实例中,包含工程化反转录子的重组多核苷酸可以与阳离子脂质组合施用。阳离子脂质的一些实例包括但不限于lipofectin、DOTMA、DOPE和DOTAP。WO/0071096(其通过引用具体并入)的公开内容描述了不同的制剂,例如DOTAP:胆固醇或胆固醇衍生物制剂,其可以有效用于基因治疗。其他公开内容还讨论了不同的脂质或脂质体制剂(包括纳米颗粒)和施用方法;这些包括但不限于美国专利公开20030203865、

20020150626、20030032615和20040048787，其在公开核酸的制剂以及核酸的施用和递送的其他相关方面的程度上通过引用而具体并入。用于形成颗粒的方法也公开在美国专利号5,844,107、5,877,302、6,008,336、6,077,835、5,972,901、6,200,801和5,972,900中，其出于这些方面通过引用并入。

[0138] 在某些实施方案中，基因转移在离体条件下可以更容易地进行。离体基因治疗是指从对象中分离细胞，在体外将核酸递送至细胞中，并且然后将经修饰的细胞返回至对象中。这可涉及收集包含来自对象的细胞的生物样品。例如，可以通过静脉穿刺获得血液，并且可以根据本领域公知的方法通过外科技术获得实体组织样品。

[0139] 通常，但不总是，接受细胞的对象（即接受者）也是从其中收获或获得细胞的对象，这提供了捐赠的细胞是自体细胞的优势。然而，细胞可以从另一个对象（即供体）、来自供体的细胞培养物或从已建立的细胞培养系中获得。细胞可以从与待治疗的对象相同或不同的物种中获得，但优选为相同物种，并且更优选与对象具有相同的免疫学特征。例如，这样的细胞可以从包含来自近亲或匹配供体的细胞的生物样品中获得，然后用核酸（例如，包含工程化反转录子）转染，并施用于需要基因组修饰的对象，例如，用于治疗疾病或病症。

[0140] 试剂盒

[0141] 还提供了包含如本文中所述的工程化反转录子构建体的试剂盒。在一些实施方案中，试剂盒提供了工程化反转录子构建体或包含这样的反转录子构建体的载体系统。在一些实施方案中，包含在试剂盒中的工程化反转录子构建体包含异源序列，所述异源序列能够为细胞提供编码目的蛋白质或调节性RNA的核酸、细胞条码、适用于基因编辑的供体多核苷酸，例如通过同源定向修复(homology directed repair, HDR)或重组介导的基因工程(重组工程)，或用于分子记录的CRISPR前间隔区DNA序列。试剂盒中还可包括其他试剂，例如转染剂、宿主细胞、用于培养细胞的合适培养基、缓冲剂等。

[0142] 在试剂盒的背景下，试剂可以以液体形式或以任何方便包装（例如，棒状包装、剂量包装等）的出售形式提供。试剂盒的试剂可以存在于相同或分开的容器中。试剂也可以存在于同一容器中。除了上述组分之外，主题试剂盒还可以包括（在某些实施方案中）用于实施主题方法的说明。这些说明可以以多种形式存在于主题试剂盒中，其中的一种或更多种可以存在于试剂盒中。这些说明可以存在的一种形式是作为印刷信息处于合适的介质或基质（例如，在其上印刷信息的一张或更多张纸）上、在试剂盒的包装中、在包装插页中等。这些说明的又一种形式是信息已记录在其上的计算机可读介质，例如磁盘、光盘(compact disk, CD)、闪存驱动器等。这些说明可存在的另一种形式是网站地址，其可通过互联网使用以访问已删除站点上的信息。

[0143] 效用

[0144] 可以用异源序列对反转录子进行工程化以用于多种应用中。例如，可以将异源序列添加至反转录子构建体以向细胞提供编码目的蛋白质或调节性RNA的异源核酸、细胞条码、适用于基因编辑的供体多核苷酸，例如通过同源定向修复(HDR)或重组介导的基因工程(重组工程)，或用于分子记录的CRISPR前间隔区DNA序列，如下文进一步所讨论的。可以将这样的异源序列插入至例如msr基因或msd基因中，使得异源序列通过反转录子逆转录酶转录为msDNA产物的一部分。

[0145] 蛋白质或RNA的产生

[0146] 例如,通过工程化反转录子生成的单链DNA可用于在细胞中产生所期望的目的产物。在一些实施方案中,用编码目的多肽的异源序列对反转录子进行工程化,以允许从在细胞中生成的反转录子msDNA产生多肽。目的多肽可以是任何类型的蛋白质/肽,其包括但不限于酶、细胞外基质蛋白、受体、转运蛋白、离子通道或其他膜蛋白、激素、神经肽、抗体或细胞骨架蛋白质;或其片段,或者目的生物活性结构域。在一些实施方案中,蛋白质是用于治疗疾病的治疗性蛋白质或治疗性抗体。

[0147] 在另一些实施方案中,用编码目的RNA的异源序列对反转录子进行工程化以允许从细胞中的反转录子产生RNA。目的RNA可以是任何类型的RNA,其包括但不限于RNA干扰(RNA interference, RNAi)核酸或调节性RNA,例如但不限于microRNA (miRNA)、小干扰RNA (small interfering RNA, siRNA)、短发夹RNA (short hairpin RNA, shRNA)、小核RNA (small nuclear RNA, snRNA)、长非编码RNA (long non-coding RNA, lncRNA)、反义核酸等。

[0148] 基因编辑

[0149] 在一些实施方案中,用编码适合与CRISPR/Cas基因组编辑系统一起使用的供体多核苷酸的异源序列对反转录子进行工程化。供体多核苷酸包含含有预期基因组编辑的序列,该序列侧翼有一对负责将供体多核苷酸靶向至细胞中待编辑的靶基因座的同源臂。供体多核苷酸通常包含与5'基因组靶序列杂交的5'同源臂和与3'基因组靶序列杂交的3'同源臂。同源臂在本文中称为5'和3'(即上游和下游)同源臂,其与同源臂相对于供体多核苷酸内包含目的编辑物的核苷酸序列的相对位置有关。5'和3'同源臂与待修饰的基因组DNA中靶基因座内的区杂交,其在本文中分别称为“5'靶序列”和“3'靶序列”。

[0150] 同源臂必须足够互补以与靶序列杂交以介导供体多核苷酸与靶基因座处的基因组DNA之间的同源重组。例如,同源臂可包含与相应基因组靶序列具有至少约80至100%序列同一性(包括该范围内的任何百分比同一性)的核苷酸序列,例如与其具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列同一性,其中包含目的编辑物的核苷酸序列可以通过HDR在由5'和3'同源臂识别(即,具有足够互补性以进行杂交)的基因组靶基因座处被整合到基因组DNA中。

[0151] 在某些实施方案中,基因组靶序列(即“5'靶序列”和“3'靶序列”)中的相应同源核苷酸序列位于用于切割的特定位点和/或用于引入目的编辑物的特定位点的侧翼。特定切割位点与同源核苷酸序列(例如,每个同源臂)之间的距离可以是数百个核苷酸。在一些实施方案中,同源臂与切割位点之间的距离为200个核苷酸或更少(例如,0、10、20、30、50、75、100、125、150、175和200个核苷酸)。在大多数情况下,更小的距离可导致更高的基因靶向率。在一个优选的实施方案中,除了待被引入至基因组的一部分(其涵盖特定切割位点和待被改变的基因组靶序列的一部分二者)的序列改变之外,供体多核苷酸与靶基因组序列在其整个长度上基本相同。

[0152] 同源臂可以是任意长度,例如10个核苷酸或更多、15个核苷酸或更多、20个核苷酸或更多、50个核苷酸或更多、100个核苷酸或更多、250个核苷酸或更多、300个核苷酸或更多、350个核苷酸或更多、400个核苷酸或更多、450个核苷酸或更多、500个核苷酸或更多、1000个核苷酸(1kb)或更多、5000个核苷酸(5kb)或更多、10000个核苷酸(10kb)或更多等。在一些情况下,5'和3'同源臂的长度彼此基本相等。然而,在一些情况下,5'和3'同源臂的

长度彼此不一定相等。例如,一个同源臂可以比另一个同源臂短30%或更少,比另一个同源臂短20%或更少,比另一个同源臂短10%或更少,比另一个同源臂短5%或更少,比另一个同源臂短2%或更少,或仅比另一个同源臂少几个核苷酸。在另一些情况下,5'和3'同源臂在长度上彼此显著不同,例如一个可以比另一个同源臂短40%或更多、短50%或更多、有时短60%或更多、短70%或更多、短80%或更多、短90%或更多、或短95%或更多。

[0153] 供体多核苷酸与RNA指导的核酸酶组合使用,该核酸酶通过指导RNA靶向特定的基因组序列(即待修饰的基因组靶序列)。靶向特异性指导RNA包含与基因组靶序列互补的核苷酸序列,并从而通过在靶位点处的杂交来介导核酸酶-gRNA复合体的结合。例如,gRNA可以设计为与次要等位基因的序列互补的序列,以将核酸酶-gRNA复合体靶向至突变位点。突变可包含插入、缺失或替换。例如,突变可以包括单核苷酸变异、基因融合、易位、倒位、重复、移码、错义、无义或与目的表型或疾病相关的其他突变。靶向的次要等位基因可以是常见的遗传变体或罕见的遗传变体。在某些实施方案中,gRNA被设计为选择性地结合至具有单碱基对区分的次要等位基因,例如,以允许核酸酶-gRNA复合体与单核苷酸多态性(single nucleotide polymorphism, SNP)结合。特别是,gRNA可以设计为靶向与疾病相关的目的突变,以用于进行基因组编辑以从基因中去除突变的目的。或者,gRNA可以被设计为与主要或野生型等位基因的序列互补的序列,以将核酸酶-gRNA复合体靶向等位基因,以用于进行基因组编辑以将突变(例如插入、缺失或替换)引入到细胞的基因组DNA中的基因中的目的。例如,这样的经遗传修饰的细胞可用于改变表型、赋予新特性或产生用于药物筛选的疾病模型。

[0154] 在某些实施方案中,用于基因组修饰的RNA指导的核酸酶是成簇规则间隔短回文重复(CRISPR)系统Cas核酸酶。任何能够催化DNA定点切割以允许通过HDR机制整合供体多核苷酸的RNA指导Cas核酸酶都可用于基因组编辑,其包括CRISPR系统I型、II型或III型Cas核酸酶。Cas蛋白质的一些实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5e(CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9(Csn1或Csn12)、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1(CasA)、Cse2(CasB)、Cse3(CasE)、Cse4(CasC)、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4和Cu1966,及其同源物或经修饰版本。

[0155] 在某些实施方案中,使用II型CRISPR系统Cas9核酸内切酶。来自任何物种的Cas9核酸酶或保留Cas9核酸内切酶活性(即催化DNA的定点切割以生成双链断裂)的其生物活性片段、变体、类似物或衍生物可用于进行如本文所述的基因组修饰。Cas9不需要物理地源自生物体,而是可以合成地或重组地产生。来自许多细菌物种的Cas9序列在本领域中是公知的,并且列在国家生物技术信息中心(NCBI)数据库中。参见,例如,来自以下的Cas9的NCBI条目:酿脓链球菌(*Streptococcus pyogenes*) (WP_002989955、WP_038434062、WP_011528583);空肠弯曲杆菌(*Campylobacter jejuni*) (WP_022552435、YP_002344900)、大肠杆菌弯曲杆菌(*Campylobacter coli*) (WP_060786116);胎儿弯曲杆菌(*Campylobacter fetus*) (WP_059434633);溃疡棒状杆菌(*Corynebacterium ulcerans*) (NC_015683、NC_017317);白喉棒状杆菌(*Corynebacterium diphtheriae*) (NC_016782、NC_016786);粪肠球菌(*Enterococcus faecalis*) (WP_033919308);梅毒螺旋体(*Spiroplasma syrphidicola*)

(NC_021284) ; 中间普雷沃氏菌 (*Prevotella intermedia*) (NC_017861) ; 台湾螺原体 (*Spiroplasma taiwanense*) (NC_021846) ; 海豚链球菌 (*Streptococcus iniae*) (NC_021314) ; 海贝尔氏菌 (*Belliella baltica*) (NC_018010) ; 扭曲冷弯曲菌 (*Psychroflexus torquisI*) (NC_018721) ; 嗜热链球菌 (*Streptococcus thermophilus*) (YP_820832) 、变形链球菌 (*Streptococcus mutans*) (WP_061046374、WP_024786433) ; 无害李斯特菌 (*Listeria innocua*) (NP_472073) ; 单核细胞增生李斯特菌 (*Listeria monocytogenes*) (WP_061665472) ; 嗜肺军团菌 (*Legionella pneumophila*) (WP_062726656) ; 金黄色葡萄球菌 (*Staphylococcus aureus*) (WP_001573634) ; 土拉热弗朗西丝菌 (*Francisella tularensis*) (WP_032729892、WP_014548420) 、粪肠球菌 (*Enterococcus faecalis*) (WP_033919308) ; 鼠李糖乳杆菌 (*Lactobacillus rhamnosus*) (WP_048482595、WP_032965177) ; 和脑膜炎奈瑟菌 (*Neisseria meningitidis*) (WP_061704949、YP_002342100) ; 所有这些序列 (如截至本申请提交之日进入) 通过引用整体并入本文中。这些序列或其变体中的任一种, 包括与其具有至少约70%至100%序列同一性 (包括该范围内的任何百分比同一性), 例如与其具有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%的序列同一性的序列, 可用于进行基因组编辑, 如本文中所述。另参见Fonfara et al. (2014) *Nucleic Acids Res.* 42 (4) : 2577-90; Kapitonov et al. (2015) *J.Bacteriol.* 198 (5) : 797-807, Shmakov et al. (2015) *Mol.Cell.* 60 (3) : 385-397, 以及 Chylinski et al. (2014) *Nucleic Acids Res.* 42 (10) : 6091-6105) ; 用于序列比较并讨论 Cas9 的遗传多样性和系统发育分析。

[0156] CRISPR-Cas 系统天然地存在于细菌和古细菌中, 在其中 CRISPR-Cas 系统在 RNA 介导的针对外源 DNA 的适应性免疫中发挥作用。细菌 II 型 CRISPR 系统使用核酸内切酶 Cas9, 其与与互补的基因组靶序列特异性杂交的指导 RNA (gRNA) 形成复合体, 其中 Cas9 核酸内切酶催化切割以产生双链断裂。Cas9 的靶向通常还依赖于 DNA 中 gRNA 结合位点处或附近的 5' 前间区序列邻近基序 (PAM) 的存在。

[0157] 基因组靶位点通常将包含与 gRNA 互补的核苷酸序列, 并且可进一步包含前间区序列邻近基序 (PAM)。在某些实施方案中, 靶位点除了 3 个碱基对的 PAM 之外还包含 20 至 30 个碱基对。通常, PAM 的第一个核苷酸可以是任何核苷酸, 而其他两个核苷酸将取决于所选的特定 Cas9 蛋白。示例性的 PAM 序列是本领域技术人员已知的并且包括但不限于 NNG、NGN、NAG 和 NGG, 其中 N 代表任何核苷酸。在某些实施方案中, 由 gRNA 靶向的等位基因包含在等位基因内产生 PAM 的突变, 其中 PAM 促进 Cas9-gRNA 复合体与等位基因的结合。

[0158] 在某些实施方案中, gRNA 的长度为 5 至 50 个核苷酸、10 至 30 个核苷酸、15 至 25 个核苷酸、18 至 22 个核苷酸或 19 至 21 个核苷酸, 或者所述范围之间的任何长度, 其包括例如长度为 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 或 35 个核苷酸。指导 RNA 可以是在单个 RNA 分子中包含 crRNA 和 tracrRNA 序列的单一指导 RNA, 或者指导 RNA 可包含两个 RNA 分子, 其中 crRNA 和 tracrRNA 序列位于分开的 RNA 分子中。

[0159] 在另一个实施方案中, 使用来自普雷沃氏菌和弗朗西丝氏菌 1 (Cpf1) 的 CRISPR 核酸酶。Cpf1 是另一 II 类 CRISPR/Cas 系统 RNA 指导的核酸酶, 其与 Cas9 相似, 并可以类似地使用。与 Cas9 不同的是, Cpf1 不需要 tracrRNA 并且仅依赖于其指导 RNA 中的 crRNA, 这提供了相

比于Cas9可以与Cpf1一起使用更短的指导RNA以进行靶向的优势。Cpf1能够切割DNA或RNA。与由Cas9识别的富含G的PAM位点形成对比,由Cpf1识别的PAM位点具有序列5'-YTN-3' (其中“Y”是嘧啶,“N”是任何核碱基)或5'-TTN-3'。DNA的Cpf1切割产生双链断裂,其具有含有4或5个核苷酸突出端的黏性末端。对于Cpf1的讨论,参见,例如,Ledford et al. (2015)

Nature.526 (7571) :17-17,Zetsche et al. (2015) Cell.163 (3) :759-771,Murovec et al. (2017) Plant Biotechnol.J.15 (8) :917-926,Zhang et al. (2017) Front.Plant Sci.8:

177,Fernandes et al. (2016) Postepy Biochem.62 (3) :315-326;其通过引用并入本文中。

[0160] C2c1是可以使用的另一II类CRISPR/Cas系统RNA指导的核酸酶。与Cas9类似,C2c1依赖于crRNA和tracrRNA二者以用于指导至靶位点。对于C2c1的描述,参见,例如Shmakov et al. (2015) Mol Cell.60 (3) :385-397,Zhang et al. (2017) Front Plant Sci.8:177;其通过引用并入本文中。

[0161] 在另一个实施方案中,可以使用工程化的RNA指导的FokI核酸酶。RNA指导的FokI核酸酶包含非活性Cas9 (dCas9) 与FokI核酸内切酶 (FokI-dCas9) 的融合体,其中dCas9部分赋予FokI指导RNA依赖性靶向。对于工程化的RNA指导的FokI核酸酶的描述,参见例如Havlicek et al. (2017) Mol.Ther.25 (2) :342-355,Pan et al. (2016) Sci Rep.6:35794,Tsai et al. (2014) Nat Biotechnol.32 (6) :569-576;其通过引用并入本文中。

[0162] RNA指导的核酸酶可以以蛋白质的形式提供,任选地,其中核酸酶与gRNA复合,或由编码RNA指导的核酸酶的核酸(例如RNA(例如,信使RNA)或DNA(表达载体))提供。在一些实施方案中, RNA指导的核酸酶和gRNA均由载体提供。两者可以由单个载体表达,或者分别在不同的载体上表达。编码RNA指导的核酸酶和gRNA的载体可以包括在包含工程化反转录子msr基因、msd基因和ret基因序列的载体系统中。

[0163] 可以优化密码子的使用以改进特定细胞或生物体中RNA指导的核酸酶和/或反转录子逆转录酶的产生。例如,编码RNA指导的核酸酶或逆转录酶的核酸可以被修饰以替代在酵母细胞、细菌细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其他目的宿主细胞中与天然存在的多核苷酸序列相比具有较高使用频率的密码子。当编码RNA指导的核酸酶或逆转录酶的核酸被引入细胞中时,蛋白质可以在细胞中瞬时、有条件地或组成性地表达。

[0164] 重组工程

[0165] 重组工程(重组介导的基因工程)可用于修饰细胞中的染色体以及附加体复制子,例如,以产生基因替换、基因敲除、缺失、插入、倒位或点突变。重组工程也可用于修饰质粒或细菌人工染色体(BAC),例如以克隆基因或插入标志物或标签。本文所述的工程化反转录子可用于重组工程应用以提供用于重组的线性单链或双链DNA。同源重组是由噬菌体蛋白质,例如来自Rac原噬菌体的RecE/RecT或来自噬菌体λ的Redαβγ介导的。线性DNA应在5'和3'端与细胞中存在的靶标DNA分子(例如,质粒、BAC或染色体)具有足够的同源性以允许重组。

[0166] 重组工程中使用的线性双链或单链DNA分子(即供体多核苷酸)包含具有待插入的目的编辑物的序列,其侧翼有两个将线性DNA分子靶向至靶位点以进行同源重组的同源臂。用于重组工程的同源臂的长度范围通常为13至300个核苷酸,或20至200个核苷酸,其包括此范围内的任何长度,例如长度为13、14、15、16、17、18、19、20、22、24、26、28、30、32、34、36、

38、40、42、44、46、48、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195或200个核苷酸。在一些实施方案中,同源臂的长度为至少15、至少20、至少30、至少40、或至少50或更多个核苷酸。长度范围为40至50个核苷酸的同源臂通常具有足够的靶向效率用于重组;然而,150至200个碱基或更多的较长同源臂可以进一步提高靶向效率。在一些实施方案中,5'同源臂和3'同源臂的长度不同。例如,线性DNA可在5'端具有约50个碱基,在3'端具有约20个碱基,其与待靶向的区具有同源性。

[0167] 噬菌体同源重组蛋白可以作为蛋白质提供给细胞或通过一种或更多种编码重组蛋白的载体提供给细胞。在一些实施方案中,一种或更多种编码噬菌体重组蛋白的载体包括在包含工程化反转录子msr基因、msd基因和ret基因序列的载体系统中。

[0168] 另外,许多含有原噬菌体重组系统的细菌菌株可用于重组工程,其包括但不限于:DY380,其含有带有重组蛋白exo、bet和gam的缺陷型λ原噬菌体;EL250,其源自DY380,其除了在DY380中发现的重组基因之外,还含有严格受控的阿拉伯糖诱导型f1pe基因(f1pe介导两个相同的frt位点之间的重组);EL350,其也源自DY380,其除了DY380中发现的重组基因之外,还含有严格受控的阿拉伯糖诱导型cre基因(cre介导两个相同的loxP位点之间的重组);SW102,其源自DY380,其使用galK阳性/阴性选择被设计用于BAC重组工程;SW105,其源自EL250,其也可用于galK阳性/阴性选择,但与EL250一样,含有ara诱导型F1pe基因;以及SW106,其源自EL350,其可用于galK阳性/阴性选择,但与EL350一样,含有ara诱导型Cre基因。重组工程可以通过用工程化反转录子转染这样的菌株的细菌细胞来进行,该工程化反转录子包含编码适合于重组工程的线性DNA的异源序列。对于重组工程系统和方案的讨论,参见例如Sharan et al. (2009) Nat Protoc. 4 (2) :206-223,Zhang et al. (1998) Nature Genetics 20:123-128,Muyrers et al. (1999) Nucleic Acids Res. 27:1555-1557,Yu et al. (2000) Proc.Natl.Acad.Sci U.S.A.97 (11) :5978-5983;其通过引用并入本文中。

[0169] 分子记录

[0170] 在一些实施方案中,工程化反转录子构建体中的异源序列包含合成的CRISPR前间隔区DNA序列以允许分子记录。内源性CRISPR Cas1-Cas2系统通常由细菌和古细菌利用以通过存储短序列(即前间隔区)来追踪源自病毒感染的外源DNA序列,所述短序列赋予了针对基于基因组的阵列内的入侵性病毒核酸的序列特异性抗性。这些阵列不仅保留了间隔子序列,还记录了获取序列的顺序,从而生成获取事件的时间记录。

[0171] 该系统可适用于将任意DNA序列以“合成前间隔区”的形式记录到基因组CRISPR阵列中,这些前间隔区使用工程化反转录子被引入细胞中。携带前间隔区序列的工程化反转录子可用于在特定基因组基因座处通过利用CRISPR系统Cas1至Cas2复合体整合合成的CRISPR前间隔区序列。分子记录可用于通过产生稳定的遗传记忆追踪代码来追踪某些生物事件。参见,例如,Shipman et al. (2016) Science 353 (6298) :aaf1175和国际专利申请公开号WO/2018/191525;其通过引用整体并入本文中。

[0172] 在一些实施方案中,利用CRISPR-Cas系统将特定和任意的DNA序列记录到细菌基因组中。DNA序列可以由细胞内的工程化反转录子产生。例如,工程化反转录子可用于在细胞内产生前间隔区,其插入细胞内的CRISPR阵列中。可以修饰细胞以包括一种或更多种工程化反转录子(或编码其的载体系统),其可以在细胞中产生一种或更多种合成的前间隔

区,其中合成的前间隔区被添加至CRISPR阵列中。可以生成限定的序列的记录,其记录了很多天,并且以多种方式记录。

[0173] 在一些实施方案中,工程化反转录子包含msd前间隔区核酸区或msr前间隔区核酸区。在msr前间隔区核酸区的情况下,前间隔区序列首先被并入msr RNA中,其被逆转录为前间隔区DNA。当具有互补序列的两个互补性前间隔区DNA序列杂交时,或者当在单链前间隔区DNA中形成双链结构(例如发夹)时(例如,单msDNA可以形成适当的发夹结构以提供双链DNA前间隔区),产生双链前间隔区DNA。

[0174] 在一些实施方案中,由第一工程化反转录子在体内产生的单链DNA可以与在体内由相同的反转录子或第二工程化反转录子产生的互补单链DNA杂交,或者可以形成发夹结构,然后用作前间隔区序列以作为间隔子序列待插入到CRISPR阵列中。工程化反转录子应该在细胞内提供足够水平的前间隔区序列以用于并入到CRISPR阵列中。使用细胞内生成的前间隔区将体内分子记录系统从仅捕获使用者已知的信息延伸到捕获使用者以前可能未知的生物或环境信息。例如,工程化反转录子构建体中的msDNA前间隔区序列可以通过位于生物现象或环境毒素的传感器通路下游的启动子驱动。CRISPR阵列中前间隔区序列的捕获和存储记录了该事件。如果多个msDNA前间隔区由不同的启动子驱动,则记录这些启动子的活性(连同可以是启动子上游的任何内容一起)以及启动子活性的相对顺序(基于CRISPR阵列中间隔子序列的相对位置)。在记录已发生之后的任何时间点,可以对CRISPR阵列进行测序以确定是否发生了给定的生物或环境事件以及多个事件的顺序,这通过CRISPR阵列中msDNA来源的间隔子的存在和相对位置给出。

[0175] 在一些实施方案中,合成的前间隔区在其5'端还包含AAG PAM序列。包括5'AAG PAM的前间隔区通过CRISPR阵列以比不包括PAM序列的前间隔区更高的效率获得。

[0176] 在一些实施方案中,Cas1和Cas2通过表达Cas1和Cas2的载体以以下水平提供:所述水平足以允许通过工程化反转录子产生的合成前间隔区序列待由细胞中的CRISPR阵列获得。这样的载体系统可用于允许在缺乏内源性Cas蛋白的细胞中进行分子记录。

[0177] 本公开内容的示例性非限制性方面

[0178] 上述本主题的一些方面(包括实施方案),单独或与一个或更多个其他方面或实施方案组合,可以是有益的。在不限制之前描述的情况下,下文提供编号为1至60的本公开内容的某些非限制性方面。如本领域技术人员在阅读本公开内容之后将是显而易见的,单独编号的方面中的每一个可以与先前或以下单独编号的方面中的任何一个一起使用或组合。这旨在为方面的所有这样的组合提供支持,并且不限于以下明确提供的方面的组合。

[0179] 1. 工程化反转录子,其包含:

[0180] a) msr前序列;

[0181] b) 编码多拷贝单链RNA(msRNA)的msr基因;

[0182] c) 编码多拷贝单链DNA(msDNA)的msd基因;

[0183] d) msd后序列,其包含与所述msr前序列具有序列互补性的自身互补区,其中所述自身互补区的长度比野生型自身互补区长至少1至50个核苷酸,使得所述工程化反转录子能够增强所述msDNA的产生;以及

[0184] e) 编码逆转录酶的ret基因。

[0185] 2. 工程化反转录子,其中自身互补区的长度比野生型自身互补区长至少5、至少

10、至少15、至少20、至少25或至少30个核苷酸。

[0186] 3. 方面1或2所述的工程化反转录子，其中所述msr基因和所述msd基因以反式排列或顺式排列提供。

[0187] 4. 方面3所述的工程化反转录子，其中所述ret基因相对于所述msr基因和所述msd基因以反式排列提供。

[0188] 5. 方面1至4中任一项所述的工程化反转录子，其中所述msr基因、msd基因和ret基因源自黏细菌反转录子、大肠杆菌反转录子或霍乱弧菌反转录子。

[0189] 6. 方面5所述的工程化反转录子，其中所述大肠杆菌反转录子是EC83或EC86。

[0190] 7. 方面1至6中任一项所述的工程化反转录子，其还包含目的异源序列。

[0191] 8. 方面7所述的工程化反转录子，其中所述异源序列被插入至所述msr基因或所述msd基因中。

[0192] 9. 方面1至7或8所述的工程化反转录子，其中异源核酸片段被插入距所述互补区的碱基9至20个碱基处。

[0193] 10. 方面7、8或9所述的工程化反转录子，其中所述异源序列在所述msd、msd后或msDNA的发夹环内。

[0194] 11. 方面1至8或9所述的工程化反转录子，其中在所述msd或所述msd后的第20至60位处插入异源核酸片段。

[0195] 12. 方面7至10或11所述的工程化反转录子，其中所述异源序列编码供体多核苷酸，所述供体多核苷酸包含与5'靶序列杂交的5'同源臂和与3'靶序列杂交的3'同源臂，所述5'同源臂和所述3'同源臂位于包含待通过同源定向修复(HDR)或重组工程整合至靶基因座处的目的编辑物的核苷酸序列的侧翼。

[0196] 13. 方面7至11或12所述的工程化反转录子，其中所述异源序列包含CRISPR前间隔区DNA序列。

[0197] 14. 方面13所述的工程化反转录子，其中所述CRISPR前间隔区DNA序列包含经修饰的AAG前间区序列邻近基序(PAM)。

[0198] 15. 方面1至14中任一项所述的工程化反转录子，其还包含条码序列。

[0199] 16. 方面15所述的工程化反转录子，其中所述条码序列位于所述msDNA的发夹环中。

[0200] 17. 载体系统，其包含含有方面1至16中任一项所述的工程化反转录子的一种或更多种载体。

[0201] 18. 方面17所述的载体系统，其中所述msr基因和所述msd基因通过同一载体或不同的载体提供。

[0202] 19. 方面17或18所述的载体系统，其中所述msr基因、所述msd基因和所述ret基因由同一载体提供。

[0203] 20. 方面17、18或19所述的载体系统，其中所述载体包含与所述msr基因和所述msd基因可操作地连接的启动子。

[0204] 21. 方面20所述的载体系统，其中所述启动子还与所述ret基因可操作地连接。

[0205] 22. 方面20或21所述的载体系统，其还包含与所述ret基因可操作地连接的第二启动子。

[0206] 23. 方面17至21或22所述的载体系统, 其中所述msr基因、所述msd基因和所述ret基因由不同的载体提供。

[0207] 24. 方面17至22或23中任一项所述的载体系统, 其中所述一种或更多种载体是病毒载体或非病毒载体。

[0208] 25. 方面24所述的载体系统, 其中所述非病毒载体是质粒。

[0209] 26. 方面17至25中任一项所述的载体系统, 其中所述工程化反转录子包含供体多核苷酸, 所述供体多核苷酸包含与5'靶序列杂交的5'同源臂和与3'靶序列杂交的3'同源臂, 所述5'同源臂和所述3'同源臂位于包含待通过同源定向修复(HDR)或重组工程整合在靶基因座处的目的编辑物的核苷酸序列的侧翼。

[0210] 27. 方面26所述的载体系统, 其还包含编码RNA指导的核酸酶的载体。

[0211] 28. 方面27所述的载体系统, 其中所述RNA指导的核酸酶是Cas核酸酶或工程化的RNA指导的FokI-核酸酶。

[0212] 29. 方面28所述的方法, 其中所述Cas核酸酶是Cas9或Cpf1。

[0213] 30. 方面17至29中任一项所述的载体系统, 其中所述工程化反转录子包含CRISPR前间隔区DNA序列。

[0214] 31. 方面30所述的载体系统, 其还包含编码Cas1或Cas2蛋白的载体。

[0215] 32. 方面30或31所述的载体系统, 其还包含含有CRISPR阵列序列的载体。

[0216] 33. 方面29至31中任一项所述的载体系统, 其还包含编码噬菌体同源重组蛋白的载体。

[0217] 34. 方面33所述的载体系统, 其中所述编码所述噬菌体同源重组蛋白的载体是包含exo、bet和gam基因的复制缺陷型λ原噬菌体。

[0218] 35. 分离的宿主细胞, 其包含方面1至16中任一项所述的工程化反转录子或方面17至34中任一项所述的载体系统。

[0219] 36. 方面35所述的宿主细胞, 其中所述宿主细胞是原核、古细菌或真核的宿主细胞。

[0220] 37. 方面36所述的宿主细胞, 其中所述真核宿主细胞是哺乳动物宿主细胞。

[0221] 38. 方面37所述的宿主细胞, 其中所述哺乳动物宿主细胞是人宿主细胞。

[0222] 39. 方面35所述的宿主细胞, 其中所述宿主细胞是人工细胞或经遗传修饰的细胞。

[0223] 40. 试剂盒, 其包含方面1至16中任一项所述的工程化反转录子、方面17至34中任一项所述的载体系统或方面35至39中任一项所述的宿主细胞。

[0224] 41. 方面40所述的试剂盒, 其还包含使用该工程化反转录子对细胞进行遗传修饰的说明。

[0225] 42. 对细胞进行遗传修饰的方法, 其包括:

[0226] a) 用方面1至15或16(例如方面12)所述的工程化反转录子转染细胞;

[0227] b) 将RNA指导的核酸酶和指导RNA引入或表达至所述细胞中, 其中所述RNA指导的核酸酶与所述指导RNA形成复合体, 所述指导RNA将所述复合体引导至基因组靶基因座, 其中所述RNA指导的核酸酶在所述基因组靶基因座处的基因组DNA中产生双链断裂, 以及由所述工程化反转录子产生的供体多核苷酸通过同源定向修复(HDR)在其5'同源臂和3'同源臂所识别的基因组靶基因座处被整合, 以产生经遗传修饰的细胞。

[0228] 43. 方面42所述的方法,其中所述RNA指导的核酸酶是Cas核酸酶或工程化的RNA指导的FokI-核酸酶。

[0229] 44. 方面43所述的方法,其中所述Cas核酸酶是Cas9或Cpf1。

[0230] 45. 方面42至44所述的方法,其中所述RNA指导的核酸酶由整合到所述细胞的基因组中的载体或重组多核苷酸提供。

[0231] 46. 方面42至45所述的方法,其中所述工程化反转录子由载体提供。

[0232] 47. 方面42至45或46所述的方法,其中所述供体多核苷酸用于产生基因置换、基因敲除、缺失、插入、倒位或点突变。

[0233] 48. 通过重组工程对细胞进行遗传修饰的方法,所述方法包括:

[0234] a) 用方面1至16(例如方面12)所述的工程化反转录子转染所述细胞;以及

[0235] b) 将噬菌体重组蛋白引入所述细胞中,其中所述噬菌体重组蛋白在靶基因座处介导同源重组,使得由所述工程化反转录子产生的供体多核苷酸在其5'同源臂和3'同源臂所识别的靶基因座处被整合,以产生经遗传修饰的细胞。

[0236] 49. 方面48所述的方法,其中所述供体多核苷酸用于通过重组工程修饰细菌细胞中的质粒、细菌人工染色体(BAC)或细菌染色体。

[0237] 50. 方面48或49所述的方法,其中使用所述供体多核苷酸以产生基因置换、基因敲除、缺失、插入、倒位或点突变。

[0238] 51. 方面48至50中任一项所述的方法,其中所述将噬菌体重组蛋白引入所述细胞中包括将复制缺陷型λ原噬菌体插入至细菌基因组中。

[0239] 52. 方面51所述的方法,其中所述噬菌体包含exo、bet和gam基因。

[0240] 53. 对细胞进行条码化的方法,其包括用方面1至15或16(例如,方面15或16)所述的工程化反转录子转染细胞。

[0241] 54. 产生体内分子记录系统的方法,其包括:

[0242] a) 将CRISPR改编系统的Cas1蛋白或Cas2蛋白引入至宿主细胞中;

[0243] b) 将包含前导序列和至少一个重复序列的CRISPR阵列核酸序列引入至所述宿主细胞中,其中所述CRISPR阵列核酸序列被整合到所述宿主细胞中的基因组DNA或载体中;以及

[0244] c) 将多个根据方面1至16(例如,方面13或14)所述的工程化反转录子引入至所述宿主细胞中,其中每个反转录子包含不同的前间隔区DNA序列,其可以被加工并插入至所述CRISPR阵列核酸序列中。

[0245] 55. 方面54所述的方法,其中所述Cas1蛋白或所述Cas2蛋白由载体提供。

[0246] 56. 方面54或55所述的方法,其中所述工程化反转录子由载体提供。

[0247] 57. 方面54至56中任一项所述的方法,其中多个所述工程化反转录子包含至少三个不同的前间隔区DNA序列。

[0248] 58. 包含体内分子记录系统的工程化细胞,其包含:

[0249] a) CRISPR改编系统的Cas1蛋白或Cas2蛋白;

[0250] b) 进入宿主细胞中的包含前导序列和至少一个重复序列的CRISPR阵列核酸序列,其中所述CRISPR阵列核酸序列被整合到所述工程化细胞中的基因组DNA或载体中;以及

[0251] c) 多个根据方面1至16(或方面13或14)所述的工程化反转录子,其中每个反转录

子包含不同的前间隔区DNA序列,其可以被加工并插入至所述CRISPR阵列核酸序列中。

[0252] 59.方面58所述的工程化细胞,其中所述Cas1蛋白或所述Cas2蛋白由载体提供。

[0253] 60.方面58或59所述的工程化细胞,其中所述工程化反转录子由载体提供。

[0254] 61.方面58至60中任一项所述的工程化细胞,其中多个所述工程化反转录子包含至少三个不同的前间隔区DNA序列。

[0255] 62.试剂盒,其包含方面58至61中任一项所述的工程化细胞和用于体内分子记录的说明书。

[0256] 60.产生重组msDNA的方法,其包括:

[0257] a)用方面1至16中任一项所述的工程化反转录子或方面17至34中任一项所述的载体系统转染宿主细胞;以及

[0258] b)在合适的条件下培养所述宿主细胞,其中产生了所述msDNA。

[0259] 实施例

[0260] 提出以下实施例以向本领域普通技术人员提供如何制造和使用所公开的主题的公开内容和描述,并且以下实施例不旨在限制本发明人认为是其发明的范围,也不旨在表示以下实验是所进行的全部或唯一实验。已经努力确保关于所使用的数字(例如量、温度等)的准确性,但应当考虑一些实验误差和偏差。除非另外指明,否则百分之几是按重量计的百分之几,分子量是重均分子量,温度是摄氏度,以及压力是大气压或接近大气压。可以使用标准缩写,例如bp,碱基对;kb,千碱基;pL,皮升;s或sec,秒;min,分钟;h或hr,小时;aa,氨基酸;kb,千碱基;bp,碱基对;nt,核苷酸;i.m.,肌内;i.p.,腹膜内;s.c.,皮下;等。

[0261] 实施例1:材料和方法

[0262] 该实施例举例说明了用于开发本发明的一些材料方法。

[0263] 细菌菌株、质粒和培养条件

[0264] 实验在BL21-AI大肠杆菌(Thermo Fisher)中进行,其含有整合的阿拉伯糖诱导型T7聚合酶、内源性CRISPR阵列、内源性反转录子,但无内源性Cas1+2或bMS.346(其是MG1655的变体,含有整合的阿拉伯糖诱导型T7聚合酶,但无内源性反转录子)。

[0265] 在图3B、3C、4D、7B、7C-2、8B、9B和10B至D中描述的基于反转录子的实验中,逆转录酶由具有红霉素诱导型启动子(mphR-ec86RT)的质粒表达(参见Rogers et al., Nucleic Acids Res. 2015 Sep 3; 43 (15): 7648-60. doi:10.1093/nar/gkv616. Epub 2015 Jul 7,其在此通过引用整体并入本文中)。msd和msr元件由诱导型T7启动子一起表达(DUET-T7-msr/msd)或者单独表达(DUET-T7-msr-T7-msd)。

[0266] 对于关于图11A至B描述的反转录子生成前间隔区实验,编码Cas1+2和经修饰的ec86msr/msd(其均由诱导型(T7/lac)启动子表达(DUET-msr/msd-Cas1+2))的质粒在每次实验之前被转化到细胞中。含有质粒的细胞在4°C下在平板上以集落形式维持长达三周。细胞在34°C下在LB培养基中生长,并使用IPTG、L-阿拉伯糖和/或红霉素进行诱导,持续指定的持续时间。

[0267] msd的电泳分析

[0268] 为了使从经修饰的反转录子产生的msd可视化,将细菌在LB中与对于表达含msr、含msd和含逆转录酶的转录物必要的所有诱导剂一起培养4至16小时。在4°C下沉淀5至25mL体积的培养物,然后使用Plasmid Plus Midi试剂盒(Qiagen)或Mini试剂盒进行制备。然后

使用RNaseA和RNaseT1的组合消化RNA，并使用ssDNA/RNA Clean&Concentrator试剂盒(Zymo Research)纯化得到的msd。通过在Novex TBE-Urea凝胶(Thermo Fisher)上运行并用SYBR Gold(Thermo Fisher)进行后染色来使msd可视化。

[0269] 变体文库构建

[0270] 反转录子ncRNA变体文库由Agilent或Twist合成为oligo库，其中每次合成运行具有多个文库。从这些oligo库中扩增单个文库，并使用NEB5a细胞作为克隆菌株使用golden gate方法将其克隆至表达载体中。这些克隆文库从克隆菌株中纯化并转移至表达菌株(BL21-AI或bMS.346)中。所有文库均通过Illumina测序进行量化。

[0271] 测序和分析

[0272] 为了量化文库实验中的RT-DNA丰度，在含有反转录子变体文库的细胞中表达之后，如上所述纯化逆转录的DNA。在4°C下沉淀5至25ml体积的培养物，然后使用Plasmid Plus Midi试剂盒(Qiagen)或Mini试剂盒进行制备。然后使用RNaseA和RNaseT1的组合消化RNA，并使用ssDNA/RNA Clean&Concentrator试剂盒(Zymo Research)纯化得到的msd。在修饰在逆转录元件之外的变体文库中(例如图7C-2)，逆转录DNA环中的条码化区在表达之后被扩增并制备以用于Illumina测序。在修饰位于逆转录元件内的变体文库中(例如，图8B)，纯化的逆转录DNA的库使用末端脱氧核苷酸转移酶(Terminal deoxynucleotidyl transferase, TdT)在3'末端用单核苷酸进行延伸，其中另外的核苷酸的长度受TdT孵育时间的控制(图6B)。接下来，使用由衔接子序列、与用于延伸的核苷酸互补的核苷酸区段和锚定核苷酸(每个碱基与用于延伸的核苷酸不互补)构成的反向引物以使用Klenow片段(3' → 5' exo-) (其在5'末端留下A突出端)创建第二链。此突出的A用于TA连接，以连接在相对的5'末端具有氨基修饰的双链衔接子(图6A)。然后对两端添加有衔接子的该核苷酸库进行索引并准备以用于Illumina测序。在所有变体文库中，通过扩增可变区并对该区进行Illumina测序来对细胞中存在的质粒库进行量化。然后通过比较反转录DNA中变体的比率与变体质粒的比率(相对于共表达的野生型反转录子归一化)来计算不同反转录DNA的相对丰度。

[0273] 为了分析间隔子的获取，通过加热至95°C持续5分钟来裂解细菌，然后使用位于前导重复序列连接点侧翼并另外包含与Illumina兼容的衔接子的引物对其基因组阵列进行PCR。基于侧翼重复序列的存在在生物信息学上提取间隔子序列，并将其与预先存在的间隔子序列进行比较以确定延伸的阵列的百分比以及新获得的间隔子的位置和序列。针对基因组和质粒序列对新的间隔子进行blast(NCBI)，并针对预期的前间隔区序列进行另外的比较以确定前间隔区的来源。此分析是使用Python中的自定义编写脚本进行的。

[0274] 实施例2：用于增强细胞中DNA产生的细菌逆转录元件(反转录子)

[0275] 重写活细胞的基因组需要新的DNA。目前，工作人员外源地合成了这种DNA并将其递送至细胞，在那里其作为用于基因组工程的模板或作为标志细胞或细胞事件的条码。然而，递送足够丰富的外源DNA以克服同源定向修复(HDR)和整合过程中的低效率(尤其是在复杂组织中)，仍然具有极大的挑战性。此外，没有办法通过细胞类型或细胞状态来控制(gate)递送以使得能够靶向具有特定DNA序列的细胞亚群。

[0276] 如果我们能够在我们选择的细胞内丰富地且按需生产设计的DNA序列-就像我们生产RNA和蛋白质一样-我们就可以克服HDR的低效率，并解锁将不同模板递送至不同细胞的能力。我们可以使用这些局部产生的DNA模板以从编辑基因组转变为写入基因组。DNA按

需将为一系列DNA修饰蛋白提供供给,其包括为λRedβ提供供给以重新编码细菌基因组,为Cas9提供供给以将治疗性修饰写入人基因组中,以及为CRISPR整合酶Cas1+2提供供给以创建在活细胞内记录分子事件时间的分子装置。

[0277] 如本文中所述,逆转录酶是产生丰富的DNA的解决方案,其包括具有不同序列的不同DNA。其不仅可以生成丰富的DNA,而且其活性可以随时间和空间而受到控制,以与我们目前控制RNA和蛋白质表达的方式一样。因此,广泛递送的逆转录酶可以在目标细胞亚群中生成丰富的模板DNA。

[0278] 一类特别有吸引力的逆转录酶来自细菌并被称为反转录子 (Inouye&Inouye, Annual review of microbiology 45, 163-186 (1991)) (参见,例如图1A)。其是紧凑的、模块化的、与真核细胞正交的,已被证明产生其他蛋白质可及的DNA,并在原核细胞 (Farzadfard&Lu Science 346, 1256272 (2014)) 和真核细胞 (Sharon et al. Cell 175, 544-557 (2018)) 二者中用作基因组编辑的模板DNA (图1B,1C)。然而,关于反转录子的生物学我们仍有很多不了解,并且这种知识空白使我们无法使用反转录子在细胞内生成完全设计的DNA序列。

[0279] 在此,我们通过对反转录子进行进一步表征和工程化来直接解决反转录子的局限性,以在细胞内产生高丰度的CRISPR兼容的任意DNA序列。大部分工程化是在大肠杆菌中进行的,以实现高通量,但本文描述的经修饰的反转录子和系统可用于真核细胞(包括人)中,以在基因组写入的背景下提供改进。

[0280] 例如,在一些情况下,反转录子-Eco1被用作示例性的反转录子。该转录物被逆转录酶识别并被部分逆转录成RT-DNA,如图2A所示。逆转录的反转录子-Eco1 DNA (RT-DNA) 的序列在下文中显示为SEQIDNO:14。

1 GTCAGAAAAAA ACGGGTTTCC TGGTTGGCTC GGAGAGCATC
[0281] 41 AGGCGATGCT CTCCGTTCCA ACAAGGAAAA CAGACAGTAA
81 CTCAGA

[0282] 实施例3:筛选工程化反转录子变体

[0283] 本发明人在大肠杆菌中表达反转录子-Eco1(也称为ec86或反转录子-Eco1 ncRNA)。这种野生型反转录子-Eco1 ncRNA的序列在下文中显示为SEQIDNO:15。

1 TGCACCCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
[0284] 41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CCTTGTTGGA ACGGAGAGCA TCGCCTGATG
121 CTCTCCGAGC CAACCAGGAA ACCCGTTTT TCTGACGTAA
161 GGGTGCGCA

[0285] 定量PCR (qPCR) 显示所表达的反转录子-Eco1在大肠杆菌中的表达产生约800至1,000个ssDNA拷贝/细胞(图3B)。如图3C所示,如此产生的ssDNA可以在变性凝胶上进行可视化、量化和纯化。

[0286] 发明人还制备了编码多种反转录子元件的构建体。例如,逆转录酶与msr/msd(引物-模板)分离,从而允许msr和msd以反式提供给逆转录酶(而不是典型的顺式排列)(图4A, 4B)。这种反式排列消除了逆转录酶的隐蔽终止信号。反转录子-Eco1仅msr区的序列在下文中显示为SEQIDNO:16。

[0287] 1 TGCGCACCCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
 41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
 81 AAGCTGTTG TCGCCAG

[0288] 反转录子-Eco1仅msd区的序列在下文中显示为SEQIDNO:17。

[0289] 1 AATCTGAGTT ACAAGCTGTT TGTCGCCAGT CAGACTGGCG
 41 ACAACCCGTT TTTTCTGACG TAAGGGTGCG CA

[0290] 使用低通量实验,逆转录酶容许的msd元件进行了改变。例如,产生了两种变体,反转录子-Eco1 v32 ncRNA和反转录子-Eco1 v35 ncRNA。反转录子-Eco1 v32 ncRNA的序列在下文中显示为SEQIDNO:18。

[0291] 1 TGCGCACCCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
 41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
 81 AAGCTGTTG TCGCCAGTCA GACTGGCGAC AACCCGTTTT
 121 TTCTGACGTA AGGGTGCGCA

[0292] 反转录子-Eco1 v35 ncRNA的序列在下文中显示为SEQIDNO:19。

[0293] 1 TGCGCACCCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
 41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
 81 AAGCTGTTG TCGCCAGTCA GACTGGCGAC AACCCGTTTT
 121 TTCTGACGTA AGGGTGCGCA

[0294] 反转录子-Eco1 v32 ncRNA和反转录子-Eco1 v35 ncRNA的关键部分显示在图3D中。

[0295] 对反转录子-Eco1 ncRNA进行了大量修饰。但是,并非每次尝试的修饰都已成功。在一些情况下,大多数经修饰的版本不会在细胞中产生ssDNA。

[0296] 为了更好地理解从反转录子产生msd的决定因素,本发明人已采用了基于文库的方法。合成了数以万计的反转录子变体以系统地测试反转录子的每个结构参数(图5)。基于golden-gate的克隆策略(Engler et al., PLOS One (Nov. 5, 2008))用于克隆这些变体,并然后在多重实验中与逆转录酶一起表达大的经修饰反转录子库。通过纯化由这些细胞产生的所有msd、对其进行测序并将其丰度与表达菌株中其反转录子/质粒的原始丰度进行比较,量化了反转录子特定参数的影响,因为其与ssDNA的产生有关。使用了ec86反转录子以及其他反转录子,包括ec83反转录子,其具有内部分支结构。

[0297] 反转录子逆转录酶通常以非标准方式引发以创建分支的RNA-DNA杂合体,其通过磷酸二酯键将2'位置处的msr RNA与msd ssDNA的5'端连接起来(Inouye&Inouye, Annual review of microbiology 45, 163-186 (1991))。大肠杆菌不具有切割这种键的酶。因此,当在大肠杆菌中时,ec86反转录子保持分支。然而,还已报道了ec83通过未知的机制进行加工,该机制对反转录子是固有的,其消除了2'-5'键并释放了ssDNA(Lim, Molecular microbiology 6, 3531-3542 (1992))。这样的分离可有利于基因组工程中的多种应用。

[0298] 实施例4:对工程化反转录子变体进行测序

[0299] 将反转录子来源的ssDNA作为实验的读出进行测序引入了显著的复杂性,因为纯化的ssDNA的库通过设计而包含未知部分(例如,不同的末端)。这些ssDNA无法被准备用于使用传统流程进行的多重测序。为了解决此挑战,本发明人已经开发了定制的测序流程,其涉及纯化ssDNA、用RNAase处理ssDNA以及使反转录子来源的ssDNA脱支。然后使用模板非依

赖性聚合酶(TdT)将纯化和脱支的反转录子ssDNA用一串单一类型的多核苷酸进行加尾。然后使用含有衔接子的反向锚定引物生成ssDNA的互补链(图6A,6B)。将第二衔接子连接至该双链DNA,然后对该衔接子连接的双链DNA进行索引并进行多重测序。

[0300] 该流程已使用合成寡核苷酸、野生型和经修饰的ec86以及野生型ec83进行了验证。该方法可靠地确定了合成寡核苷酸的正确序列。

[0301] 有趣的是,使用这种多路的单分子方法,ec86反转录子来源的ssDNA通常比文献中报到的使用较旧的批量方法(例如Maxam-Gilbert测序)早一个碱基终止。还证实了ec83的切割和预测的内源性外切核酸酶加工(图6C至6F)。由于对反转录子来源的ssDNA进行了测序,因此可以直接读取反转录子的msd(模板)部分的修饰。

[0302] 本发明人还旨在理解非逆转录的ncRNA的参数。为了读出这些参数,将非逆转录区的变体与插入到msd的环区中的条码相关联(图7A)。即使没有对所述变体进行测序,但这种方法阐明了序列变异的影响,例如,对ssDNA的产生的影响。

[0303] 实施例5:提高细胞中DNA产生的修饰

[0304] 该实施例说明,分离msr和msd转录物可以允许产生更长的RT-DNA,并且改变反转录子自身互补的非编码RNA区的长度可以提高通过反转录子生成的逆转录DNA的丰度。

[0305] 如实施例3中所述,编码反转录子-Eco1 msr和msd元件的反式构建体的表达消除了逆转录酶的隐蔽终止信号,并允许生成待生成的更长的ssDNA(图4C,4D)。

[0306] 延伸的反式反转录子-Eco1 msd序列的一个实例被称为反转录子-Eco1 msd+50,并在下文中显示为SEQIDNO:20。

[0307]

```

1 TGGACAATAT TGAATGGAGT CTGATCAACC TTCACACCGA
41 TCTAGAACATCG GAATCTGAGT TACAAAGCTGT TTGTCGCCAG
81 TCAGACTGGC GACAACCCGT TTTTCTGAC GTAAGGGTGC
121 GCA

```

[0308] 如图7所示,反转录子-Eco1 ncRNA在5'和3'端的自身互补区的延伸导致细胞中由反转录子产生的RT-DNA的丰度大幅提高。延伸的反转录子-Eco1 ncRNA的一个实例在下文中显示为SEQIDNO:21。

[0309]

```

1 TGATAAGATT CCGTATGCGC ACCCTTAGCG AGAGGTTTAT
41 CATTAAGGTC AACCTCTGGA TGTTGTTTCG GCATCCTGCA
81 TTGAATCTGA GTTACTGTCT GTTTCCCTTG TTGGAACGGA
121 GAGCATCGCC TGATGCTCTC CGAGCCAACC AGGAAACCCG
161 TTTTTCTGA CGTAAGGGTG CGCATAACGGA ATCTTATCA

```

[0310] 具有稍微不同序列的延伸的反转录子-Eco1 v35 ncRNA的一个实例如下所示(SEQIDNO:22)。

[0311]

```

1 TGATAAGATT CCGTATGCGC ACCCTTAGCG AGAGGTTTAT
41 CATTAAGGTC AACCTCTGGA TGTTGTTTCG GCATCCTGCA
81 TTGAATCTGA GTTACAAGCT GTTTGTCGCC AGTCAGACTG
121 GCGACAACCC GTTTTTCTG ACGTAAGGGT GCGCATACGG
161 AATCTTATCA

```

[0312] 如图7B至7C所示,反转录子-Eco1 ncRNA在5'和3'端的自身互补区的延伸导致细胞中由反转录子产生的RT-DNA的丰度大幅提高。例如,通过提高msd序列自身互补区的长度,可以使ssDNA的相对量提高10倍。

[0313] 相反,ncRNA自身互补碱基的减少大大减弱了RT-DNA的产生。例如,具有较短自身互补序列的反转录子-Eco1 ncRNA的一个序列在下文中显示为SEQIDNO:23。

[0314]

```

1 TAGCGAGAGG TTTATCATT AAGGTCAACCT CTGGATGTTG
41 TTTCGGCATC CTGCATTGAA TCTGAGTTAC TGTCTGTTT
81 CCTTGTGGA ACGGAGAGCA TCGCCTGATG CTCTCCGAGC
121 CAACCAGGAA ACCCGTTTT TCTGACGTA

```

[0315] 如图7C所示,这些互补碱基的数量的减少大大减弱了RT-DNA的产生。

[0316] 因此,msr前/msd后自身互补区的延伸可以提高ssDNA的库。更大的逆转录ssDNA库可用于遗传修饰,并可以提高细菌(重组工程)、酵母(CRISPEY)和哺乳动物细胞中基因组编辑的效率。为了在活细胞中产生丰富的DNA,这些具有延伸的自身互补区的变体是优选的。

[0317] 实施例6:Msd茎区容许部分而非全部的修饰

[0318] 对反转录子-Eco1 ncRNA区的msd茎区进行了修饰,以破坏茎二级结构(双链键合)。该实施例说明了可以在哪些地方对msd茎进行修饰而不会对由反转录子产生的逆转录ssDNA的丰度产生不利影响。

[0319] 沿msd茎经修饰的位置在图8A至8B和9A至9B中举例说明。

[0320] 对msd茎结构长度的改变可以创建更短的RT-DNA序列。例如,具有短茎短、仍然提供野生型水平的ssDNA的反转录子-Eco1 ncDNA的一个序列(反转录子-Eco1茎短OK,参见图8B)在下文显示为SEQIDNO:24。

[0321]

```

1 TGCGCACCCCT TAGCGAGAGG TTTATCATT AAGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTT CCTTGTGGA AGCCTAGCCA ACCAGGAAAC
121 CCGTTTTTC TGACGTAAGG GTGCGCA

```

[0322] 然而,随着msd茎长度减少至少于14个碱基,产生的RT-DNA的丰度受到负面影响(图8B)。具有太短的茎的序列反转录子-Eco1 ncDNA的一个实例在下文显示为SEQIDNO:25(图8B,茎太短)。

[0323]

```

1 TGCGCACCCCT TAGCGAGAGG TTTATCATT AAGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTG CCTAACCGT TTTTCTGAC GTAAGGGTGC
121 GCA

```

[0324] 相比之下,当反转录子的茎区被破坏并随后修复时,生成的ssDNA的量与野生型相同或略高于其。“破坏”意指茎的碱基配对被破坏,例如,通过引入非互补的核苷酸。如图9B所示,当通过连续改变五个碱基来破坏ncRNA茎的碱基时,RT-DNA的丰度显著降低。这样的具有五个错配碱基的反转录子-Eco1 ncRNA的序列(断裂茎,图9B)在下文中显示为SEQIDNO:26。

[0325]

```

1 TGCGCACCCCT TAGCGAGAGG TTTATCATT AAGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTT CGAACATGGA ACGGAGAGCA TCGCCTGATG
121 CTCTCCGAGC CAACCAGGAA ACCCGTTTT TCTGACGTAA
161 GGGTGCGCA

```

[0326] 然而,如果这些碱基通过茎另一侧的互补变化来补偿以保留茎二级结构,则保留RT-DNA丰度(参见例如,固定茎,图9B)。如图9B所示的“固定茎”反转录子-Eco1 ncRNA的序

列在下文显示为SEQIDNO:27。

```

1 TGCGCACCC TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
[0327] 81 TGTCTGTTT CGAACATGGA ACGGAGAGCA TCGCCTGATG
121 CTCTCCGAGC CATGGTCGAA ACCCGTTTT TCTGACGTAA
141 GGGTGCGCA

```

[0328] 在距茎基部(base)9至20个碱基的区中对茎进行修饰是容许的,即使其破坏了茎结构(图9B,参见可容许的断裂茎)。这样的可容许的断裂茎对茎的中部进行了修饰(错配)。具有图9B的可容许断裂茎的反转录子-Eco1 ncRNA的序列的一个实例在下文显示为SEQIDNO:28。

```

1 TGCGCACCC TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
[0329] 81 TGTCTGTTT CCTTGGTGGG ACGCTCTCCA TCGCCTGATG
121 CTCTCCGAGC CAACCAGGAA ACCCGTTTT TCTGACGTAA
161 GGGTGCGCA

```

[0330] 因此,为了在活细胞中产生DNA的目的,可以修饰ncRNA的序列,只要保留茎结构的基部即可。对msd茎中部(距茎的基部约9至20个碱基)的修饰是容许的,并且不会对ssDNA的逆转录产生不利影响。

[0331] 实施例7:ncRNA msd茎区中心更容许修饰

[0332] 对反转录子-Eco1 ncRNA区的逆转录(msd)区内的不同位置进行了小的修饰,并测量了这些修饰对从不同ncRNA变体逆转录的ssDNA的量的影响。

[0333] 如图10所示,对整个ncRNA逆转录区的小修饰的容限是可变的。图10B示出了从沿msd互补(茎)区的不同位置缺失三个碱基的影响。如所示出的,从msd茎的中部缺失三个碱基没有不利影响,并且仍然导致高水平的ssDNA产生(图10B)。然而,当靠近互补(茎)区的基部或在茎的侧翼区中产生三个碱基缺失时,观察到较低水平的ssDNA产生(图10B)。对于在msd区的中部和侧翼部分中的三个碱基的插入(图10C)和单个碱基变化(图10D),观察到类似的影响。虽然msd茎区的中部对插入和/或缺失几个核苷酸(例如少于5个核苷酸)是容许的,以使得没有观察到ssDNA产生的显著性降低,但对msd茎基部处的侧翼序列进行这样的修饰是不容许的。msd茎的基部和侧翼区的修饰导致ssDNA的逆转录降低。

[0334] 图10E以图表形式示出了基于图10B至10D中的数据计算的msd茎区内的位置的可修饰性得分。为了在活细胞中产生更高水平的DNA的目的,ncRNA的序列应在可修饰性得分高的区内进行修饰,而应避免对可修饰性得分低的区进行修饰。

[0335] 实施例8:工程化反转录子的应用

[0336] 在具有工程化反转录子的细胞中按需创建DNA能够从在活细胞中编辑基因组转变为写入基因组。这种转变将使我们能够在治疗性地修饰细胞,而不受先前存在的序列的限制。目前,新的DNA HDR模板必须作为随时间递减的推注递送。这种推注递送的低效率意味着其不能原位写入,而是必须在体外写入,随后进行选择和延伸。并非所有实验都与此策略兼容,并且少数治疗剂与此策略兼容。

[0337] 为了充分发挥基于CRISPR的治疗剂的潜力,发明人提供了设计的DNA序列,以便在需要的时间和地点准确地重写基因组。可以提供如图11A所示的设计的序列。修饰的效率通过在ncRNA的5'和3'末端延伸自身互补区来提高,并且反转录子逆转录酶可被调用以从中

产生大量所期望的ssDNA。

[0338] 按需产生DNA还实现了基因组工程的新应用,其目的不是直接的治疗剂,而是了解疾病的生物学。这是分子记录领域,其中对细胞基因组的修饰用于在活系统中写入数据。对于使用CRISPR整合酶Cas1+2来捕获CRISPR阵列中的DNA短序列(按时间顺序记录以存储事件记录)的该方法的描述,参见,例如,Shipman et al. (2017) Nature 547 (7663) :345-349,和Shipman et al. (2016) Science 353 (6298) :aaf1175;其通过引用整体并入本文中。这种方法可以用作细胞中数据采集装置的后端。然而,其需要在细胞中生成DNA条码(通过生物事件(如转录)驱动)的前端。反转录子来源的DNA可以实现这些类型的技术,并使我们能够以前所未有的详细程度了解复杂的生物过程。

[0339] 最终,驯化反转录子扩展了我们的针对基因组工程的全部分子生物技术。反转录子可以用模块化组分设计,以按需在活细胞中制造任意DNA序列,其影响广泛延伸到对基因治疗、细胞控制和细胞工程感兴趣的科学家。

[0340] 参考文献

[0341] 1 Inouye,M.&Inouye,S.msDNA and bacterial reverse transcriptase.Annual review of microbiology45,163-186,doi:10.1146/annurev.mi.45.100191.001115 (1991) .

[0342] 2 Farzadfar,F.&Lu,T.K.Synthetic biology.Genomically encoded analog memory with precise in vivo DNA writing in living cell populations.Science346,1256272,doi:10.1126/science.1256272 (2014) .

[0343] 3 Sharon,E.et al.Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing.Cell 175,544-557.e516,doi:10.1016/j.cell.2018.08.057 (2018) .

[0344] 4 Dhundale,A.R.,Furuichi,T.,Inouye,S.&Inouye,M.Distribution of multicopy single-stranded DNA among myxobacteria and related species.Journal of bacteriology164,914-917 (1985) .

[0345] 5 Lampson,B.C.et al.Reverse transcriptase in a clinical strain of Escherichia coli:production of branched RNA-linked msDNA.Science 243,1033-1038 (1989) .

[0346] 6 Inouye,K.,Tanimoto,S.,Kamimoto,M.,Shimamoto,T.&Shimamoto,T.Two novel retror elements are replaced with retror-Vc95 in Vibrio cholerae.Microbiology and immunology55,510-513,doi:10.1111/j.1348-0421.2011.00342.x (2011) .

[0347] 7 Lim,D.Structure and biosynthesis of unbranched multicopy single-stranded DNA by reverse transcriptase in a clinical Escherichia coli isolate.Molecular microbiology 6,3531-3542 (1992) .

[0348] 8 Shipman,S.L.,Nivala,J.,Macklis,J.D.&Church,G.CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria.Nature, doi:10.1038/nature23017 (2017) .

[0349] 9 Shipman,S.L.,Nivala,J.,Macklis,J.D.&Church,G.M.Molecular recordings

by directed CRISPR spacer acquisition. Science, doi:10.1126/science.aaf1175 (2016).

[0350] 虽然已经参考本公开内容的一些具体实施方案描述了本公开内容,但是本领域技术人员应理解,在不脱离本公开内容的真实精神和范围的情况下,可以进行多种改变并且可以替换成等同方案。另外,可以进行许多修饰以使特定情况、材料、物质组成、工艺、一个或更多个工艺步骤适应本公开内容的目的、精神和范围。所有这样的修饰都旨在在所附权利要求书的范围内。

序列表

<110>	Shipman, Seth	
	The J. David Gladstone Institutes, a testamentary trust	
	Established under the Will of J. David Gladstone	
	President and Fellows of Harvard College	
<120>	具有增强的DNA产生的经修饰的细菌逆转录元件	
<130>	3730.112W01	
<150>	US 62/899,625	
<151>	2019-09-12	
<160>	29	
<170>	FastSEQ for Windows Version 4.0	
<210>	1	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	1	
	gtcagaaaaaa acgggttcc tggttgctc ggagagcatc aggcgatgct ctccgttcca	60
	acaaggaaaa cagacagttaa ctcaga	86
<210>	2	
<211>	69	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	2	
	augcgcaccc uuagcgagag guuuaucauu aagguaaacc ucuggauguu guuucggcau	60
	ccugcauug	69
<210>	3	
<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	3	
	gtcagaaaaaa acgggttgc gccagtctga ctggcgacaa acagcttgta actcaga	57
<210>	4	

<211>	55	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	4	
gtcagaaaaaa acgggtggag gttgctgcaa cctctccatt ttcttgtaac tcaga		55
<210>	5	
<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	5	
gtcagaaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcaga		57
<210>	6	
<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	6	
gtcagaaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcaga		57
<210>	7	
<211>	56	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	7	
gtcagaaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcag		56
<210>	8	
<211>	88	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	8	
gtcagaaaaaa acgggtttcc tgggtggctc ggagagcatc aggcgatgct ctctccgttc		60

caacaaggaa aacagacagt aactcaga	88
<210> 9	
<211> 87	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 9	
gtcagaaaaa acggtttcc tggggctc ggagagcatc aggcgatgct ctctccgttc	60
caacaaggaa aacagacagt aactcag	87
<210> 10	
<211> 83	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 10	
ttgaagccgc ggaacaaact tttgatccg caacctactg gattgcggct caaaaagttt	60
gttccgcaac tgtaatgta atc	83
<210> 11	
<211> 79	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 11	
agcccgccgaa caaactttt gatccgcaac ctactggatt gcggctaaa aagtttgttc	60
cgcaactgta aatgtaatc	79
<210> 12	
<211> 169	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 12	
tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcgccatc	60
ctgcattgaa tctgagttac tgtctgttt ccttggatggaa acggagagca tcgcctgatg	120
ctctccgagc caaccaggaa acccgaaaa tctgacgtaa gggtgcgca	169
<210> 13	

<211>	199	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	13	
tgataagatt ccgtatgcgc acccttagcg agaggttat cattaaggtc aacctctgga		60
tgttgttgc gcattcgtca ttgaatctga gttactgtct gtttccttg ttggAACGGA		120
gagcatcgcc tcatgtctc cgagccaacc aggaaACCCG tttttctga cgtaagggtg		180
cgcatacgga atcttatca		199
<210>	14	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	14	
gtcagaaaaa acgggttcc tggtggctc ggagagcatc aggcgtatgt ctccgttcca		60
acaaggaaaa cagacagtaa ctcaga		86
<210>	15	
<211>	169	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	15	
tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttgc ttccggcatc		60
ctgcattgaa tctgagttac tgtctgttt ccttggatggaa acggagagca tcgcctgtatg		120
ctctccgagc caaccaggaa acccggtttt tctgacgtaa gggtgcgca		169
<210>	16	
<211>	97	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	16	
tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttgc ttccggcatc		60
ctgcattgaa tctgagttac aagctgtttt tcggccag		97
<210>	17	

<211>	72	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	17	
aatctgagtt acaagctgtt tgcgccagt cagactggcg acaaccgtt tttctgacg		60
taagggtgcg ca		72
<210>	18	
<211>	140	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	18	
tgcgcacccct tagcgagagg ttatcatta aggtcaacct ctggatgtt ttcggcatc		60
ctgcattgaa tctgagttac aagctgttg tcgcaggta gactggcgac aaccgtttt		120
ttctgacgta agggtgcgca		140
<210>	19	
<211>	140	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	19	
tgcgcacccct tagcgagagg ttatcatta aggtcaacct ctggatgtt ttcggcatc		60
ctgcattgaa tctgagttac aagctgttg tcgcaggta gactggcgac aaccgtttt		120
ttctgacgta agggtgcgca		140
<210>	20	
<211>	123	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	20	
tggacaatat tgaatggagt ctgatcaacc ttcacaccga tctagaatcg gaatctgagt		60
tacaagctgt ttgtcgccag tcagactggc gacaaccgt ttttctgac gtaagggtgc		120
gca		123
<210>	21	

<211>	199	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	21	
tgataagatt ccgtatgcgc acccttagcg agaggtttat cattaaggtc aacctctgga	60	
tgttgttgc gcacccctgca ttgaatctga gttactgtct gtttcccttg ttggAACGGA	120	
gagcatcgcc tggatgtctc cgagccaacc aggaaacccg tttttctga cgttaagggtg	180	
cgcatacgga atcttatca	199	
<210>	22	
<211>	170	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	22	
tgataagatt ccgtatgcgc acccttagcg agaggtttat cattaaggtc aacctctgga	60	
tgttgttgc gcacccctgca ttgaatctga gttacaagct gtttgcgcc agtcagactg	120	
gcaacaaccc gtttttctg acgttaagggt ggcatacgga aatcttatca	170	
<210>	23	
<211>	149	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	23	
tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc ctgcattgaa	60	
tctgagttac tgtctgtttt ccttggatggaa acggagagca tcgcctgatg ctctccgagc	120	
caaccaggaa acccgaaaa tctgacgta	149	
<210>	24	
<211>	147	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的寡核苷酸	
<400>	24	
tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc	60	
ctgcattgaa tctgagttac tgtctgtttt ccttggatggaa agcctagcca accagggaaac	120	

ccgtttttc tgacgtaagg gtgcgca	147
<210> 25	
<211> 123	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 25	
tgcgcacccct tagcgagagg tttatcatta aggtcaacct ctggatgttgc 60 ctgcattgaa tctgagttac tgtctgttt cctaaccgt ttttctgac gtaagggtgc 120 gca 123	
<210> 26	
<211> 169	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 26	
tgcgcacccct tagcgagagg tttatcatta aggtcaacct ctggatgttgc 60 ctgcattgaa tctgagttac tgtctgttt cgaacatgga acggagagca tcgcctgatg 120 ctctccgagc caaccaggaa acccgaaaa tctgacgtaa gggtgcgca 169	
<210> 27	
<211> 169	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 27	
tgcgcacccct tagcgagagg tttatcatta aggtcaacct ctggatgttgc 60 ctgcattgaa tctgagttac tgtctgttt cgaacatgga acggagagca tcgcctgatg 120 ctctccgagc catggcgaa acccgaaaa tctgacgtaa gggtgcgca 169	
<210> 28	
<211> 169	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 28	
tgcgcacccct tagcgagagg tttatcatta aggtcaacct ctggatgttgc 60	

ctgcattgaa tctgagttac tgtctgttt ccttggatc acgctctcca tcgcctgatg	120
ctctccgagc caaccaggaa acccgaaaa tctgacgtaa gggtgcgca	169
<210> 29	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的寡核苷酸	
<400> 29	
cccccccccc cc	12

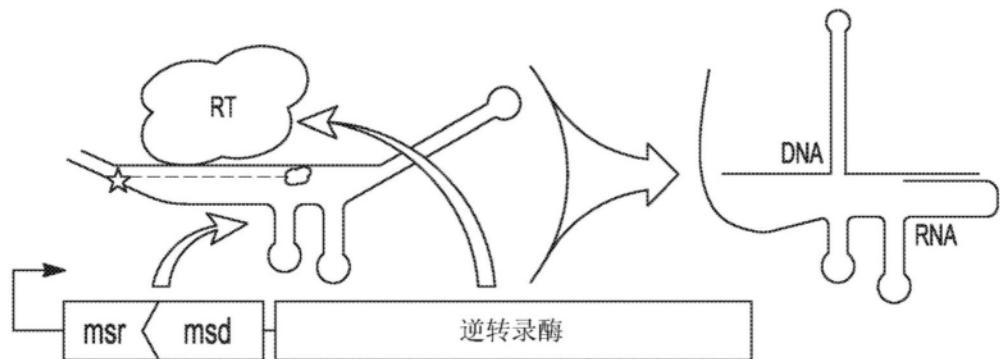


图1A

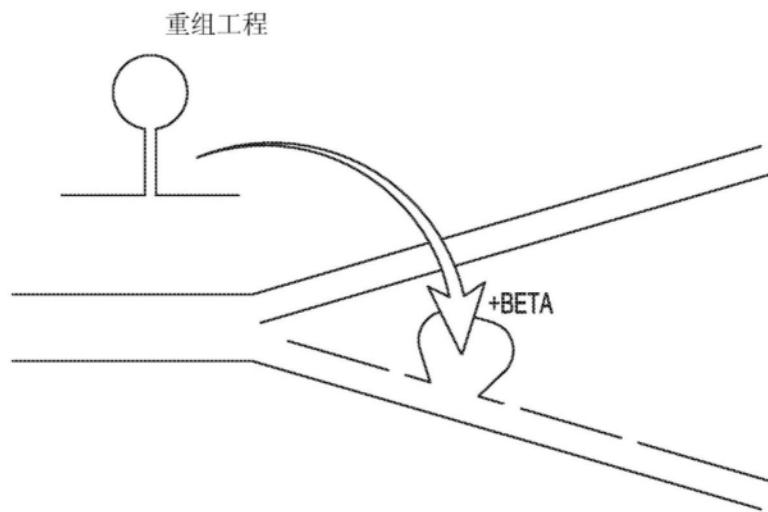


图1B

CRISPR 写入

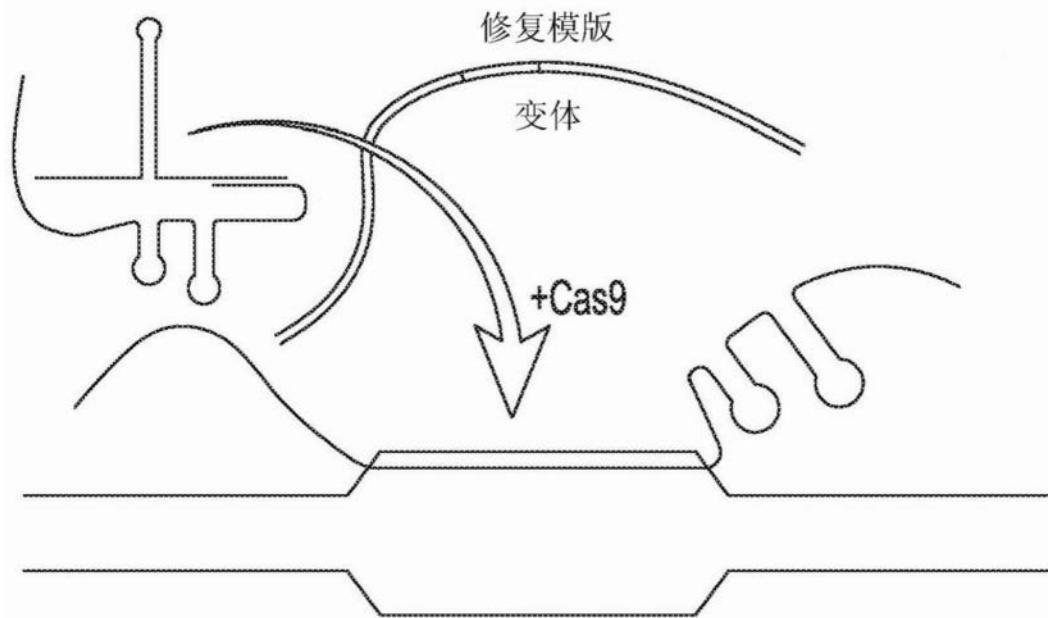


图1C

分子记录

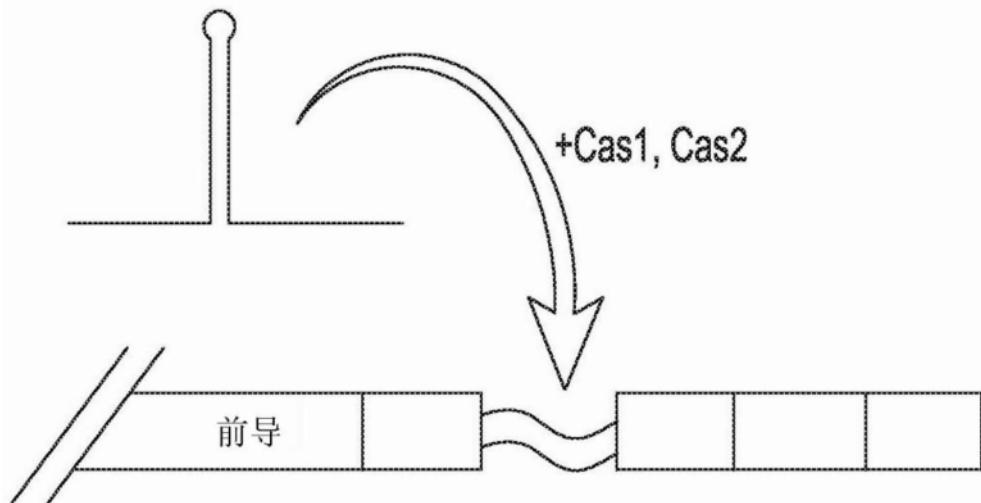


图1D

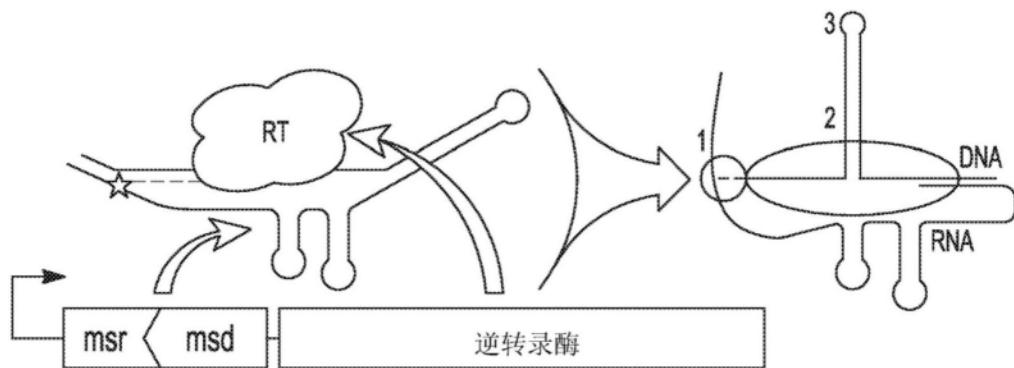


图2A

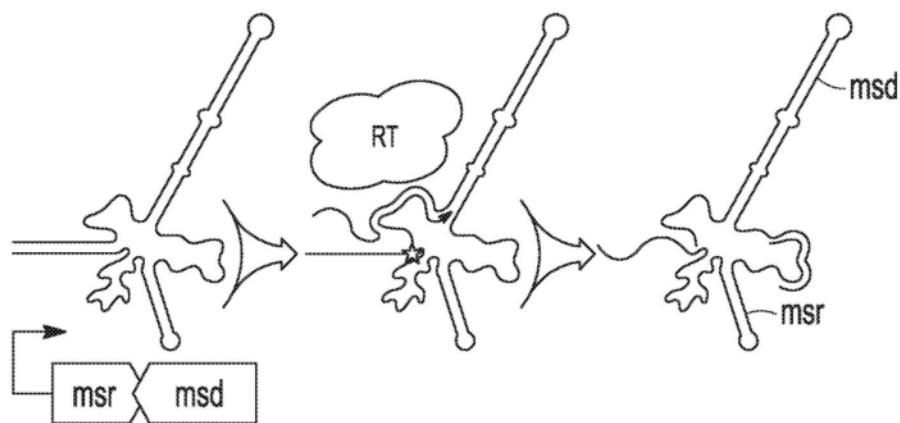


图2B

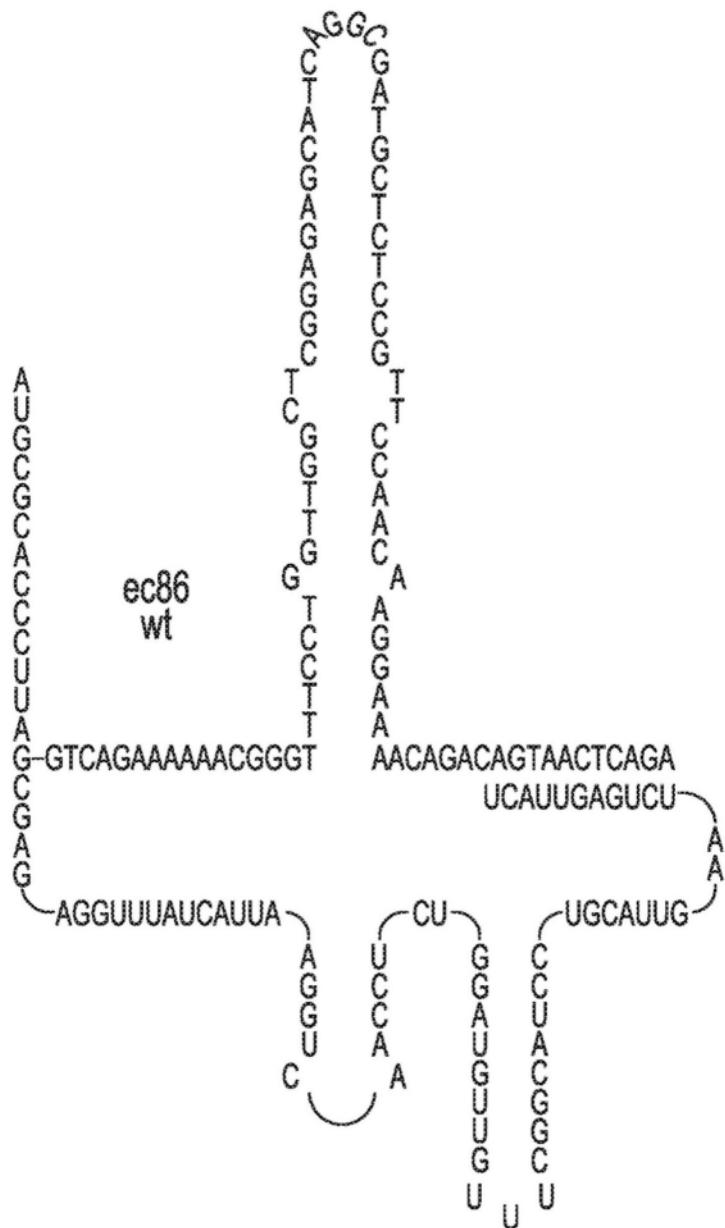


图3A

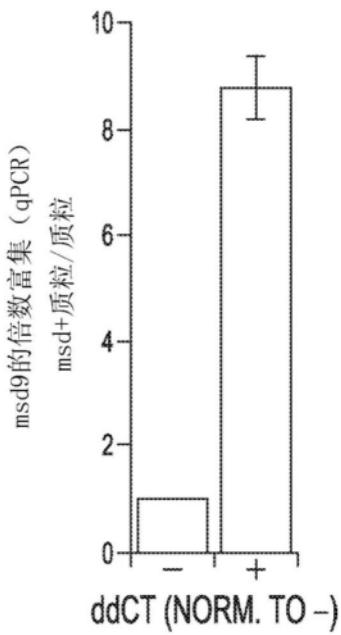


图3B

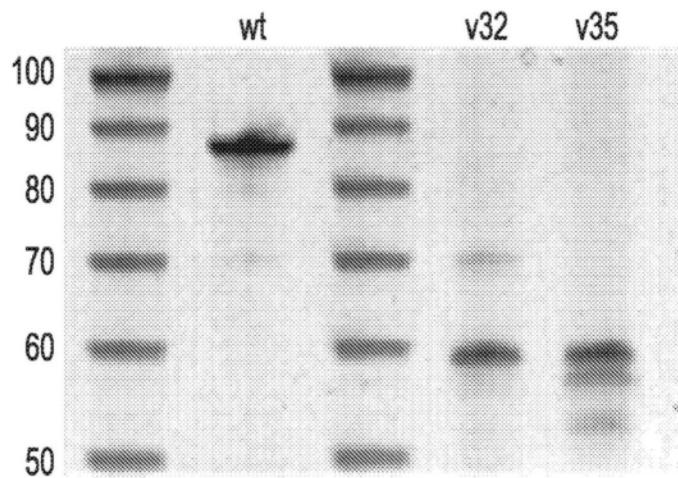


图3C



图3D

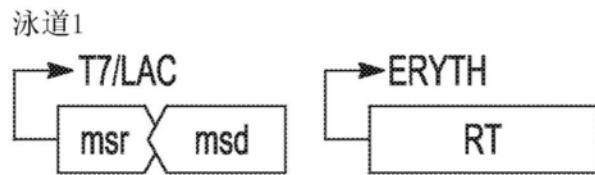


图4A

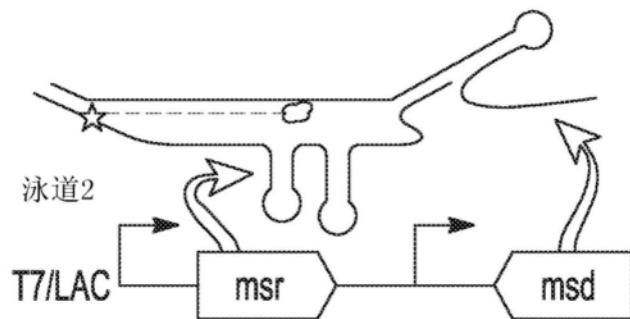


图4B

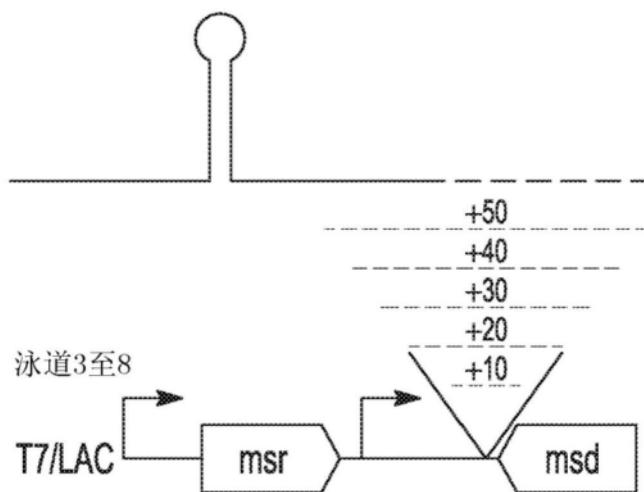


图4C

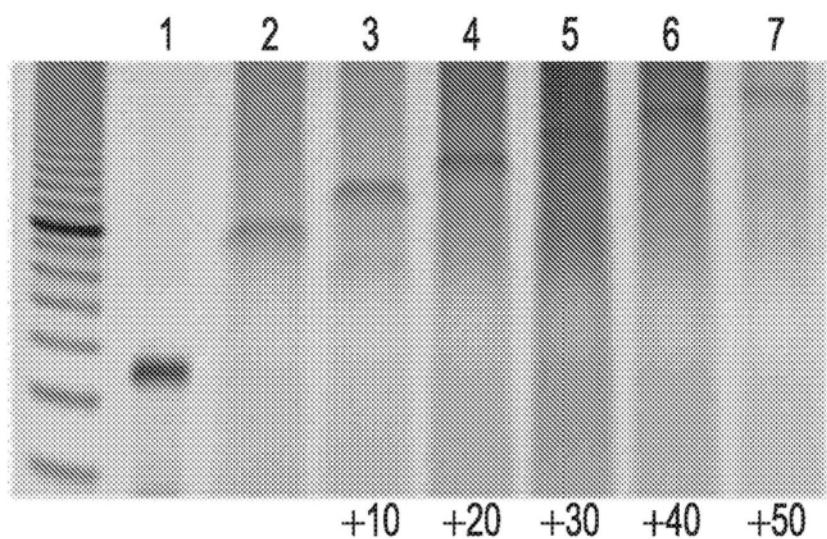


图4D

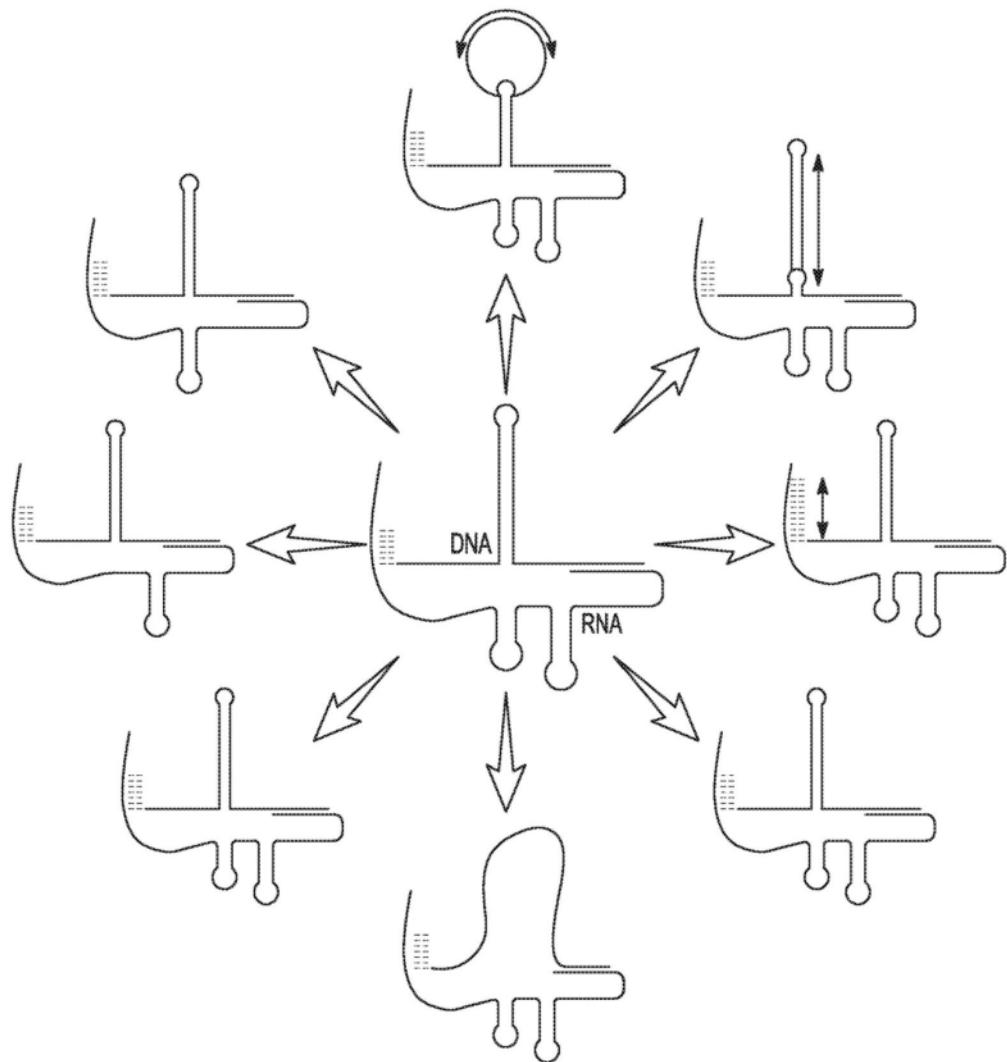


图5

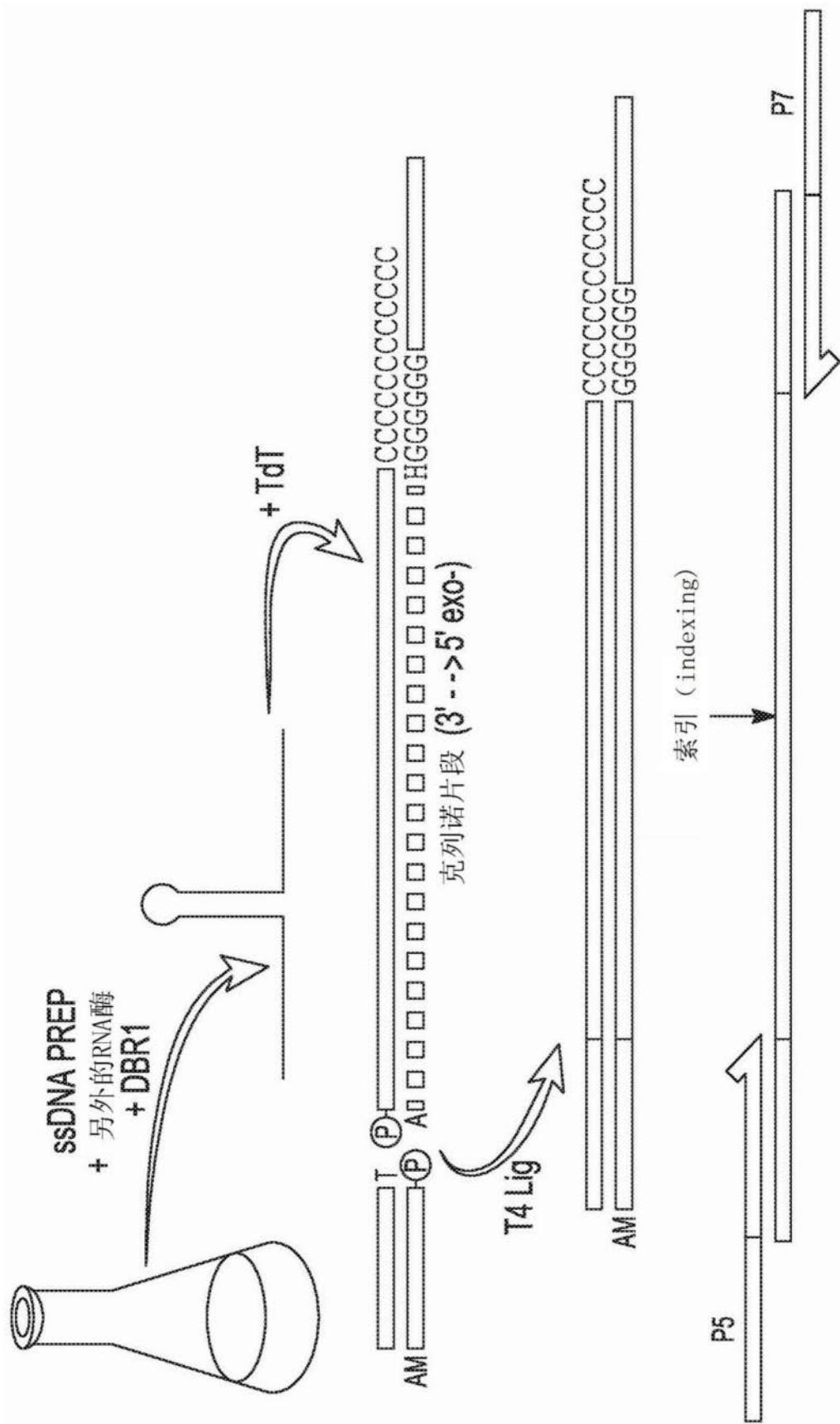


图6A

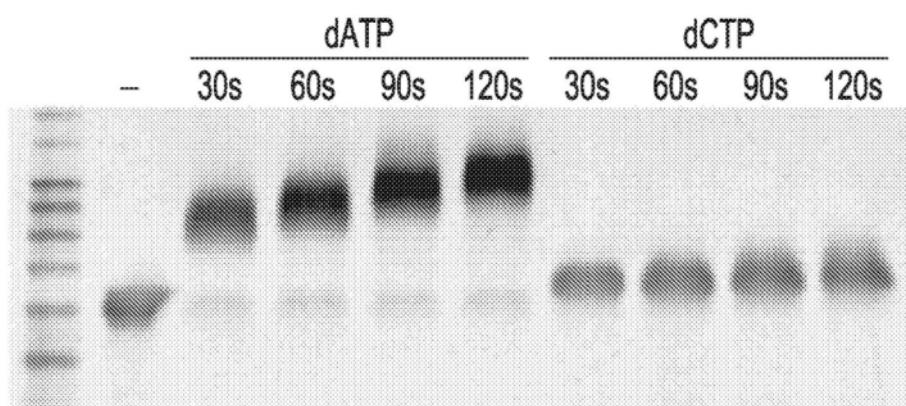


图6B



图6C

图6D展示了两个序列比对结果，上方是经测序的序列，下方是预测的序列。两个序列都包含以下部分：

↑ GTCAGAAAAACGGGT
↓ AACAGCTTGTAACTCAGA
↑ GACCCGCGTCT
↓ CTTGCT-GGCGACCA
↑ C T G A
↓ G A C C C G C T C T

右侧标注了“经测序的”和“预测的”，上方还有一个指向测序序列的箭头。

图6D

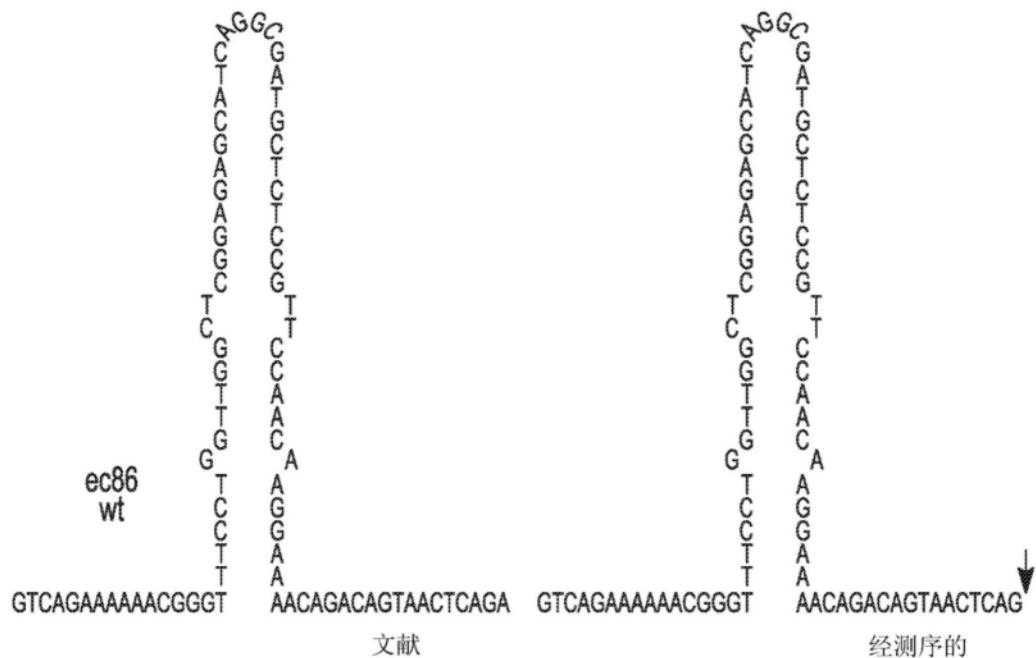


图 6E

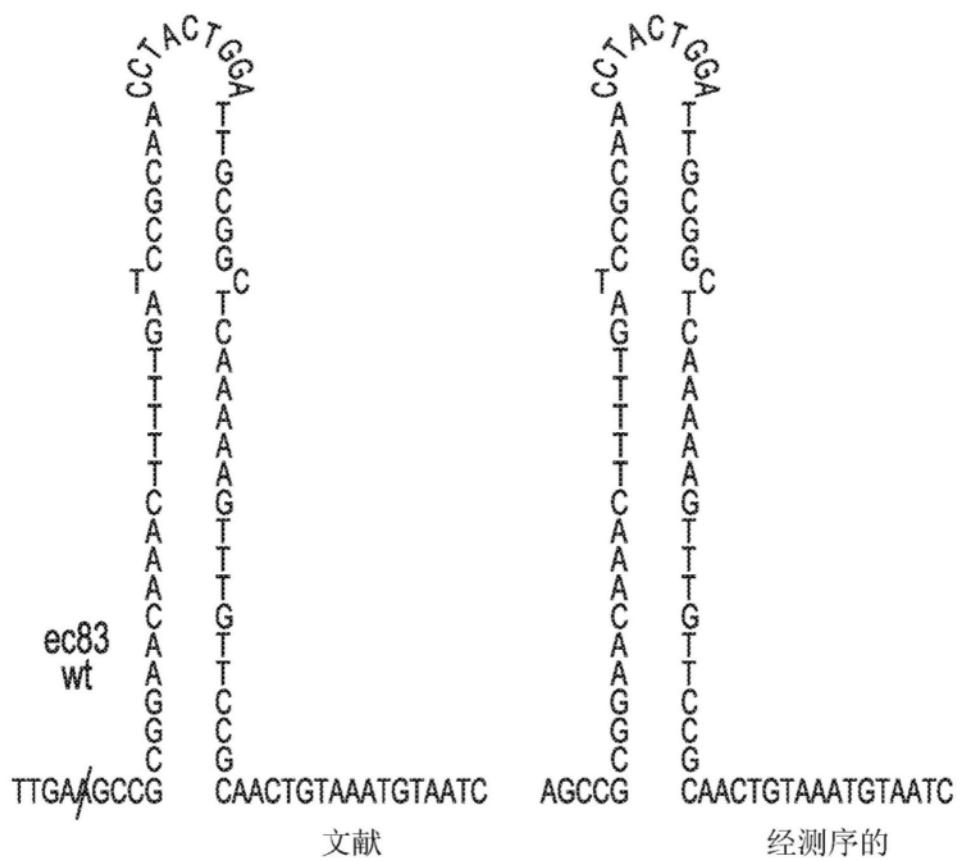


图6F

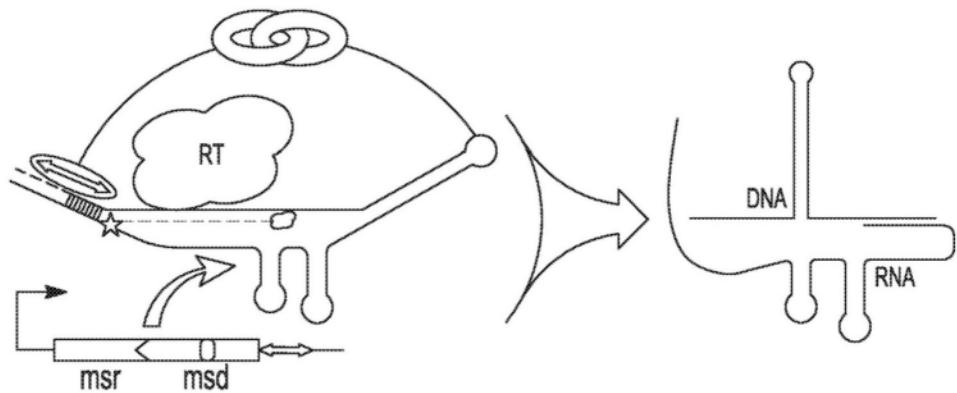


图7A

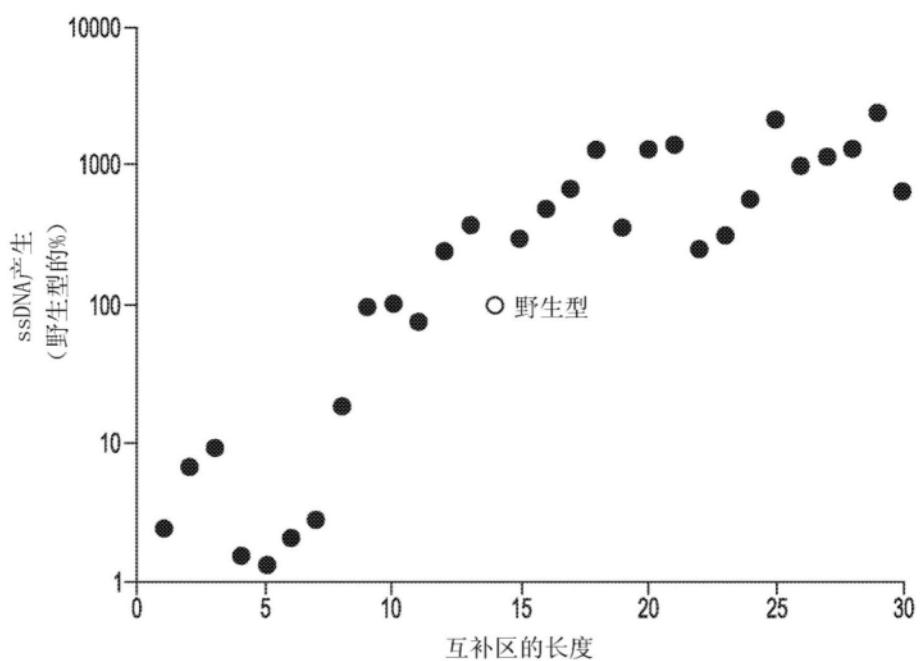


图7B

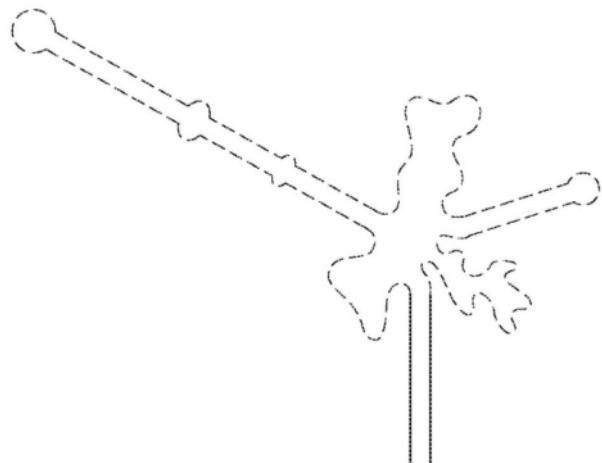


图7C-1

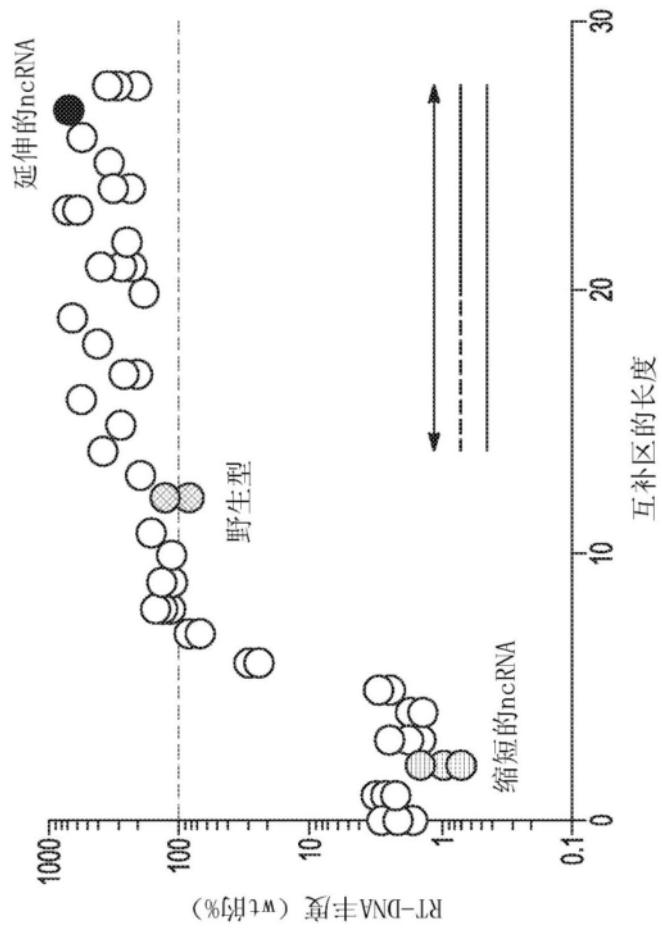


图7C-2

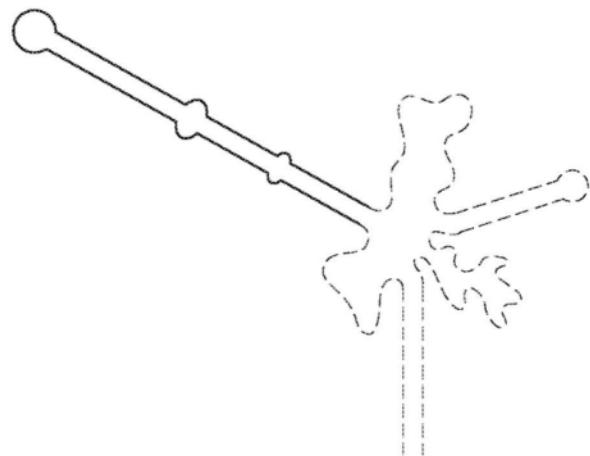


图8A

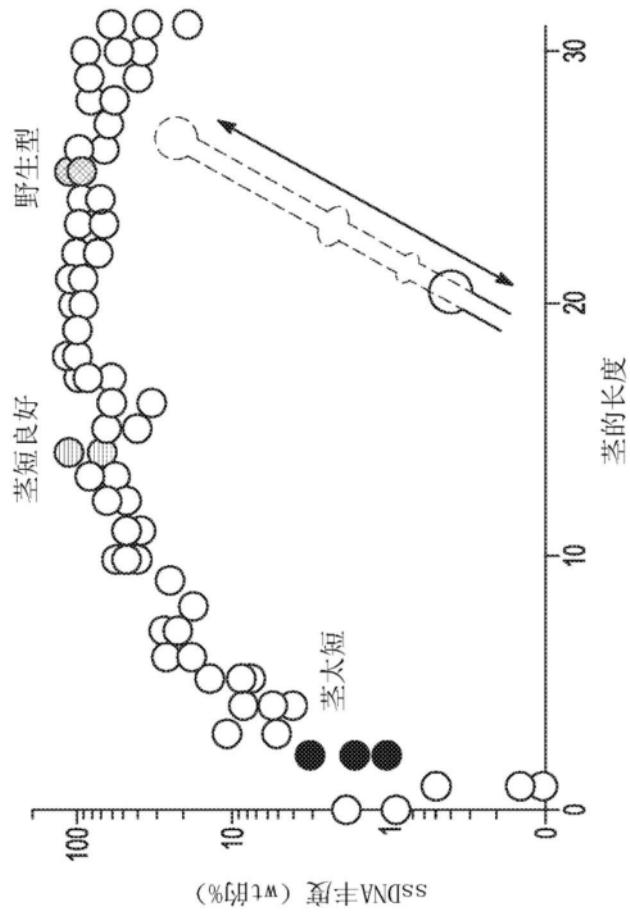


图8B

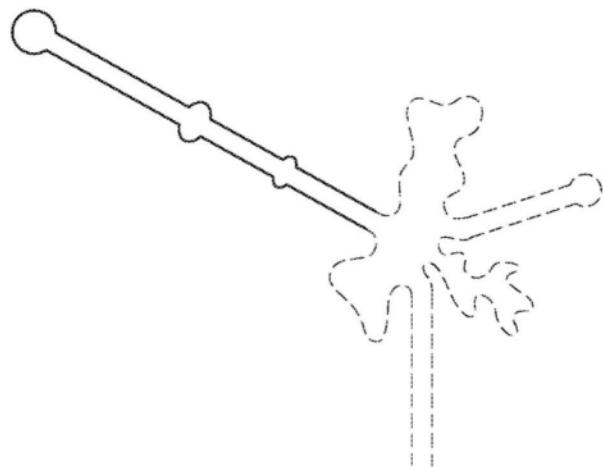


图9A

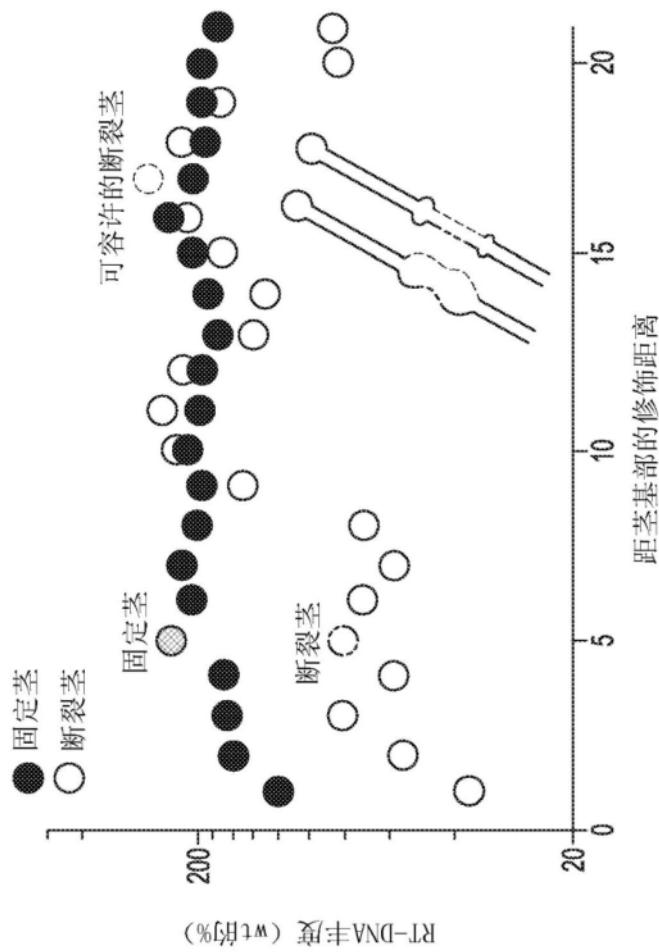


图9B

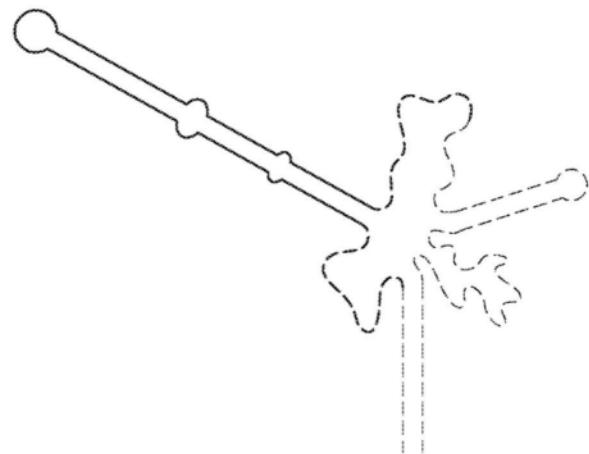


图10A

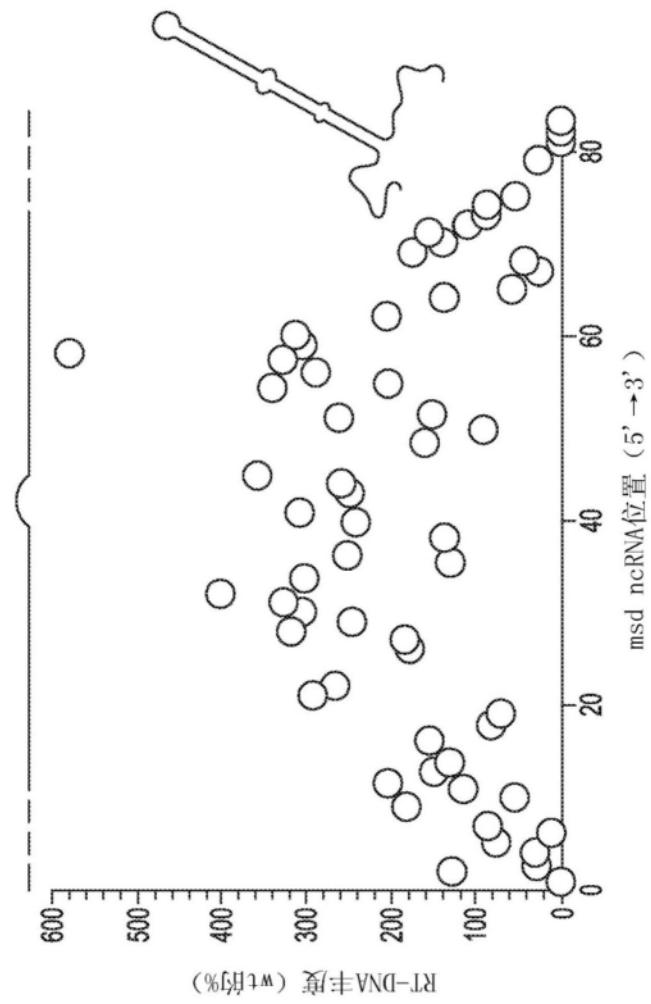


图10B

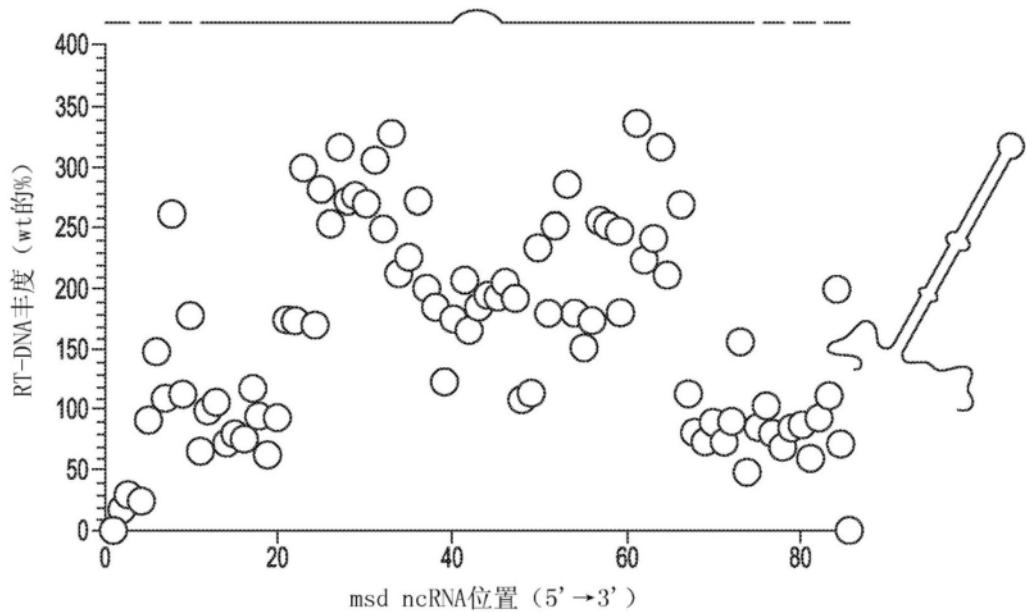


图10C

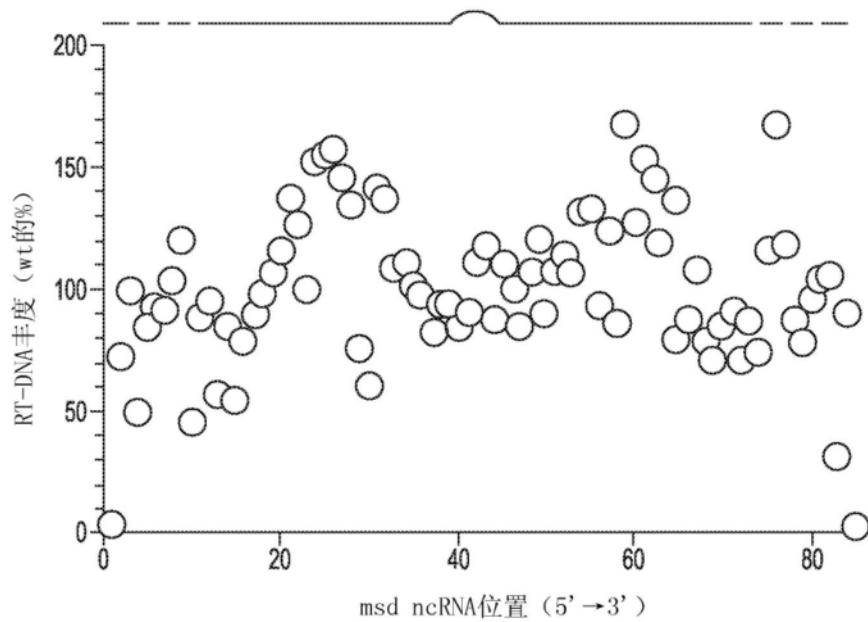


图10D

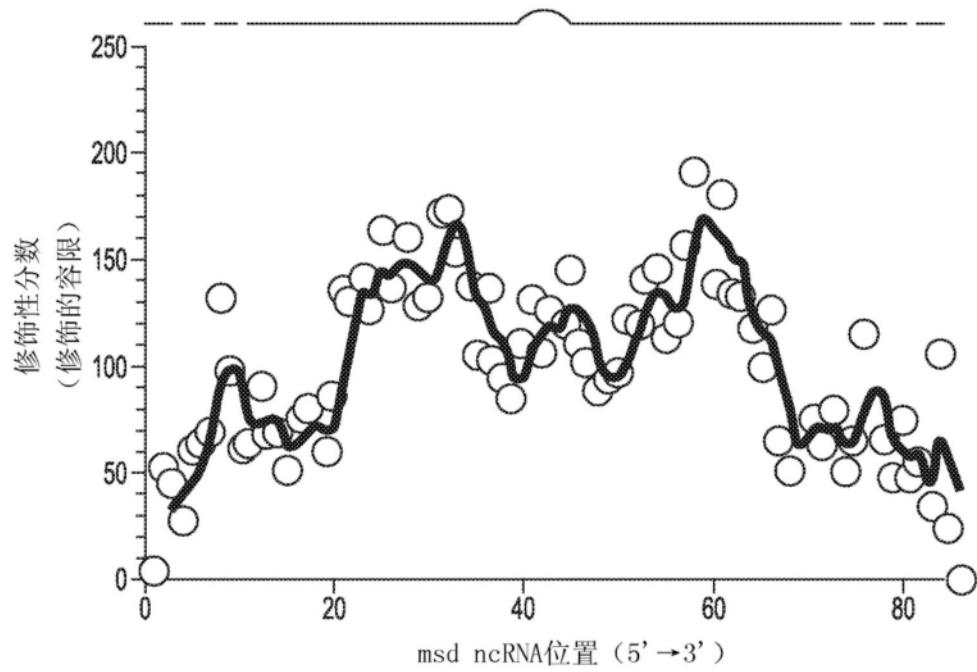


图10E

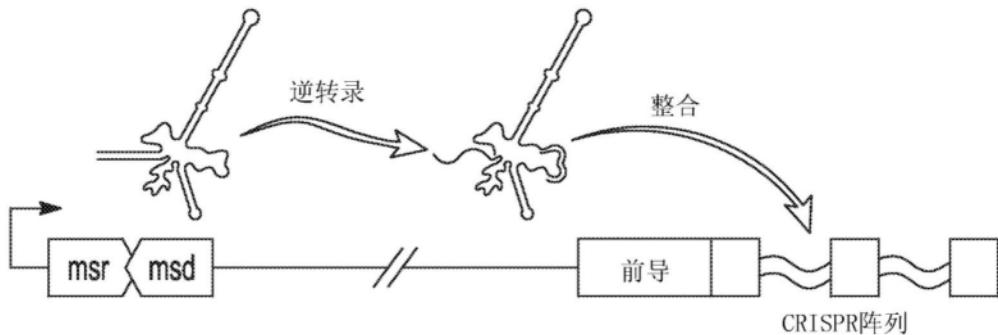


图11A

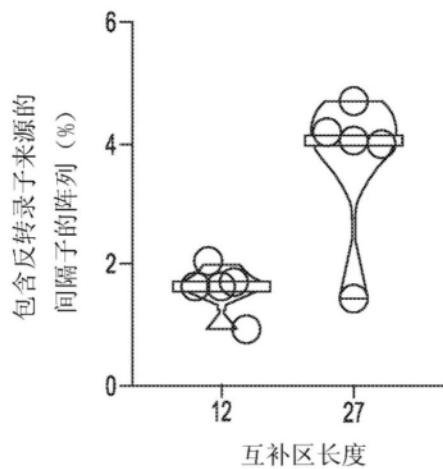


图11B