



(12) 发明专利申请

(10) 申请公布号 CN 114007640 A

(43) 申请公布日 2022. 02. 01

(21) 申请号 202080045754.2

(22) 申请日 2020.04.30

(30) 优先权数据

62/841,578 2019.05.01 US

(85) PCT国际申请进入国家阶段日

2021.12.22

(86) PCT国际申请的申请数据

PCT/US2020/030875 2020.04.30

(87) PCT国际申请的公布数据

W02020/223571 EN 2020.11.05

(71) 申请人 朱诺治疗学股份有限公司

地址 美国华盛顿州

(72) 发明人 S·M·伯利 C·H·奈

(74) 专利代理机构 北京坤瑞律师事务所 11494

代理人 封新琴

(51) Int.Cl.

A61K 39/00 (2006.01)

C07K 14/725 (2006.01)

C12N 5/0783 (2006.01)

权利要求书10页 说明书210页

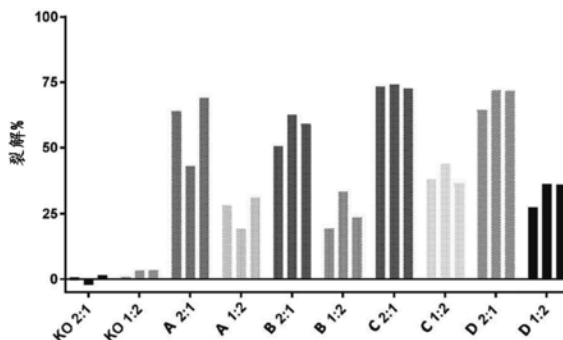
序列表91页 附图18页

(54) 发明名称

从修饰的CD247基因座表达嵌合受体的细胞、相关多核苷酸和方法

(57) 摘要

本文提供表达嵌合受体的工程化的免疫细胞,例如T细胞,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些实施方案中,所述工程化的免疫细胞含有修饰的CD247基因座,所述基因座编码所述嵌合受体或其部分。在一些实施方案中,CD3 ζ 链的至少一部分由CD247基因组基因座编码。还提供含有工程化的免疫细胞的细胞组合物、用于工程化细胞的核酸,以及用于产生所述工程化细胞的方法、试剂盒和制品,如通过靶向编码嵌合受体的一部分的转基因以供整合至CD247基因组基因座的区域中。在一些实施方案中,所述工程化细胞,例如T细胞,可以与细胞疗法结合使用,包括与癌症免疫疗法结合使用,所述癌症免疫疗法包括所述工程化细胞的过继转移。



1. 一种基因工程化的T细胞,其包含修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。

2. 根据权利要求1所述的基因工程化的T细胞,其中所述核酸序列包含编码所述嵌合受体的一部分的转基因序列,所述转基因序列已经任选地经由同源定向修复(HDR)整合于T细胞的内源CD247基因座处。

3. 根据权利要求1或权利要求2所述的基因工程化的T细胞,其中整个CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的片段是由所述内源CD247基因座的开放阅读框或其部分序列编码。

4. 根据权利要求1-3中任一项所述的基因工程化的T细胞,其中编码所述嵌合受体的所述核酸序列包含以下的框内融合物:(i) 编码所述嵌合受体的一部分的转基因序列,和(ii) 所述内源CD247基因座的开放阅读框或其部分序列。

5. 一种基因工程化的T细胞,其包含修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,其中所述核酸序列包含以下的框内融合物:(i) 编码所述嵌合受体的一部分的转基因序列,和(ii) 编码所述CD3 ζ 信号传导结构域的内源CD247基因座的开放阅读框或其部分序列。

6. 根据权利要求2-5中任一项所述的基因工程化的T细胞,其中所述转基因序列与所述内源CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框。

7. 根据权利要求2-6中任一项所述的基因工程化的T细胞,其中所述转基因序列不包含编码3'UTR的序列和/或不包含内含子。

8. 根据权利要求2-7中任一项所述的基因工程化的T细胞,其中:

所述转基因序列编码所述CD3 ζ 信号传导结构域的片段或者不编码所述CD3 ζ 信号传导结构域或其片段。

9. 根据权利要求3-8中任一项所述的基因工程化的T细胞,其中所述开放阅读框或其部分序列包含所述内源CD247基因座的至少一个内含子和至少一个外显子,和/或编码所述内源CD247基因座的3'UTR。

10. 根据权利要求2-9中任一项所述的基因工程化的T细胞,其中所述转基因序列位于所述内源CD247基因座的开放阅读框的外显子1的下游和外显子8的上游;任选地所述内源CD247基因座的开放阅读框的外显子1的下游和外显子3的上游。

11. 根据权利要求1-10中任一项所述的基因工程化的T细胞,其中所编码的嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段,任选地整个CD3 ζ 信号传导结构域,是由所述内源CD247基因座的开放阅读框或其部分序列编码,任选地其中所述CD3 ζ 信号传导结构域是由包含所述内源CD247基因座的开放阅读框的外显子2和外显子3-8的至少一部分的核苷酸序列编码;或者是由不包含所述内源CD247基因座的开放阅读框的外显子1、不包含外显子1的全长和/或不包含外显子2的全长的核苷酸序列编码。

12. 根据权利要求1-11中任一项所述的基因工程化的T细胞,其中所编码的嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

13. 根据权利要求1-12中任一项所述的基因工程化的T细胞,其中所编码的CD3 ζ 信号传

导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。

14. 根据权利要求1-13中任一项所述的基因工程化的T细胞,其中所述嵌合受体是嵌合抗原受体(CAR)。

15. 根据权利要求1-14中任一项所述的基因工程化的T细胞,其中所述嵌合受体包含含有结合结构域的细胞外区域、跨膜结构域和细胞内区域。

16. 根据权利要求15所述的基因工程化的T细胞,其中所述结合结构域是或包含抗体或其抗原结合片段。

17. 根据权利要求15或16所述的基因工程化的T细胞,其中所述结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达,任选地其中所述靶抗原是肿瘤抗原。

18. 根据权利要求17所述的基因工程化的T细胞,其中所述靶抗原选自 $\alpha v\beta 6$ 整合素(avb6整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPha2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、神经节苷脂GD3、糖蛋白100(gp100)、磷脂酰肌醇蛋白聚糖-3(GPC3)、G蛋白偶联受体C类5族成员D(GPRC5D)、Her2/neu(受体酪氨酸激酶erb-B2)、Her3(erb-B3)、Her4(erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原(HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1(HLA-A1)、人白细胞抗原A2(HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 $\alpha 2$ (IL-13R $\alpha 2$)、激酶插入结构域受体(kdr)、 κ 轻链、L1细胞粘附分子(L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A(LRRC8A)、Lewis Y、黑色素瘤相关抗原(MAGE)-A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素(MSLN)、c-Met、鼠类巨细胞病毒(CMV)、粘蛋白1(MUC1)、MUC16、自然杀伤细胞2族成员D(NKG2D)配体、黑色素A(MART-1)、神经细胞粘附分子(NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原(PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原(PSCA)、前列腺特异性膜抗原(PSMA)、受体酪氨酸激酶样孤儿受体1(ROR1)、存活蛋白、滋养层糖蛋白(TPBG,也称为5T4)、肿瘤相关糖蛋白72(TAG72)、酪氨酸酶相关蛋白1(TRP1,也称为TYRP1或gp75)、酪氨酸酶相关蛋白2(TRP2,也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体(VEGFR)、血管内皮生长因子受体2(VEGFR2)、肾母细胞瘤1(WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。

19. 根据权利要求15-18中任一项所述的基因工程化的T细胞,其中所述细胞外区域包含间隔子,任选地其中所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。

20. 根据权利要求19所述的基因工程化的T细胞,其中所述间隔子包含免疫球蛋白铰链区和/或C_H2区域和C_H3区域。

21. 根据权利要求1-20中任一项所述的基因工程化的T细胞,其中所述细胞内区域包含一个或多个共刺激信号传导结构域。

22. 根据权利要求21所述的基因工程化的T细胞,其中所述一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。

23. 根据权利要求2-22中任一项所述的基因工程化的T细胞,其中:

所述转基因序列按顺序包含:编码结合结构域、任选地单链Fv片段(scFv)的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;细胞内区域,所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域;和/或

所述修饰的CD247基因座按顺序包含:编码结合结构域、任选地scFv的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;细胞内区域,所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域;以及CD3 ζ 信号传导结构域。

24. 根据权利要求2-23中任一项所述的基因工程化的T细胞,其中所述转基因序列包含编码至少一种其他蛋白质的核苷酸序列。

25. 根据权利要求24所述的基因工程化的T细胞,其中所述至少一种其他蛋白质是替代标记,任选地其中所述替代标记是截短的受体,任选地其中所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。

26. 根据权利要求2-25中任一项所述的基因工程化的T细胞,其中所述转基因序列包含一个或多个多顺反子元件。

27. 根据权利要求26所述的基因工程化的T细胞,其中:

所述转基因序列包含编码所述嵌合受体的一部分的核苷酸序列,并且所述一个或多个多顺反子元件定位于编码所述嵌合受体的部分的核苷酸序列的上游;和/或定位于编码所述嵌合受体的部分的核苷酸序列与编码所述至少一种其他蛋白质的核苷酸序列之间;和/或

所述嵌合受体是CAR,所述CAR是多链CAR,并且所述一个或多个多顺反子元件定位于编码多链CAR的一条链的核苷酸序列与编码所述多链CAR的另一条链的核苷酸序列之间。

28. 根据权利要求26或27所述的基因工程化的T细胞,其中所述一个或多个多顺反子元件是或包含核糖体跳跃序列,任选地其中所述核糖体跳跃序列是T2A、P2A、E2A或F2A元件。

29. 根据权利要求1-28中任一项所述的基因工程化的T细胞,其中所述修饰的CD247基因座包含所述内源CD247基因座的启动子和/或调节或控制元件,所述启动子和/或调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达;或者其中所述修饰的CD247基因座包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制所述嵌合受体或其部分的表达。

30. 根据权利要求1-29中任一项所述的基因工程化的T细胞,其中所述T细胞是源自受

试者的原代T细胞,任选地其中所述受试者是人。

31. 根据权利要求1-30中任一项所述的基因工程化的T细胞,其中所述T细胞是CD8+T细胞或其亚型,或者CD4+T细胞或其亚型。

32. 一种多核苷酸,其包含:

(a) 编码嵌合受体或其部分的核酸序列;以及

(b) 与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框或其部分序列的一个或多个区域同源的序列。

33. 根据权利要求32所述的多核苷酸,其中(a)的所述核酸序列编码嵌合受体的一部分,其中由所述核酸序列编码的所述嵌合受体的部分包含含有结合结构域的细胞外区域和跨膜结构域,并且不包含细胞内区域的整个CD3zeta (CD3 ζ) 信号传导结构域。

34. 一种多核苷酸,其包含:

(a) 编码嵌合受体的部分的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,其中由所述核酸序列编码的所述嵌合受体的部分包含含有结合结构域的细胞外区域和跨膜结构域,并且不包含细胞内区域的整个CD3zeta (CD3 ζ) 信号传导结构域;以及

(b) 与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框或其部分序列的一个或多个区域同源的序列。

35. 根据权利要求32-34中任一项所述的多核苷酸,其中(a)的所述核酸序列是对于T细胞、任选地人T细胞的所述内源基因组CD247基因座的开放阅读框为外源或异源的序列。

36. 根据权利要求32-35中任一项所述的多核苷酸,其中所述开放阅读框或其部分序列包含T细胞、任选地人T细胞的所述内源CD247基因座的至少一个内含子和至少一个外显子和/或3'UTR。

37. 根据权利要求32-36中任一项所述的多核苷酸,其中在嵌合受体由引入所述多核苷酸的细胞表达时,所述CD3 ζ 信号传导结构域的至少一个片段、任选地整个CD3 ζ 信号传导结构域,是由所述内源CD247基因座的开放阅读框或其部分序列编码。

38. 根据权利要求32-37中任一项所述的多核苷酸,其中(a)的所述核酸序列编码所述CD3 ζ 信号传导结构域的片段。

39. 根据权利要求32-37中任一项所述的多核苷酸,其中(a)的所述核酸序列不编码所述CD3 ζ 信号传导结构域。

40. 根据权利要求32-39中任一项所述的多核苷酸,其中(a)的所述核酸序列不包含编码3'UTR的序列;和/或不包含内含子。

41. 根据权利要求32-40中任一项所述的多核苷酸,其中(a)的所述核酸序列包含与所述一个或多个同源臂中包含的所述CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框的核苷酸序列;任选地其中所述开放阅读框的所述一个或多个区域是或包含:位于所述CD247基因座的开放阅读框的外显子8的上游的序列;位于所述CD247基因座的开放阅读框的外显子3的上游的序列,任选地包括所述CD247基因座的开放阅读框的外显子3的序列;和/或包括所述CD247基因座的开放阅读框的外显子2的至少一部分的序列。

42. 根据权利要求32-41中任一项所述的多核苷酸,其中所述一个或多个同源臂不包含所述内源CD247基因座的开放阅读框的外显子1,不包含外显子1的全长和/或不包含外显子

2的全长。

43. 根据权利要求32-42中任一项所述的多核苷酸,其中,在由引入所述多核苷酸的细胞表达时,所述嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

44. 根据权利要求33-43中任一项所述的多核苷酸,其中所述整个CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。

45. 根据权利要求32-45中任一项所述的多核苷酸,其中所述一个或多个同源臂包含5'同源臂和3'同源臂,并且所述多核苷酸包含结构[5'同源臂]-[(a)的核酸序列]-[3'同源臂]。

46. 根据权利要求32-45中任一项所述的多核苷酸,其中所述5'同源臂和所述3'同源臂独立地具有为或约200、300、400、500、600、700或800个核苷酸的长度,或者任何前述值之间的任何值,或者具有大于或大于约300个核苷酸的长度,任选地为或约400、500或600个核苷酸的长度,或者任何前述值之间的任何值。

47. 根据权利要求32-46中任一项所述的多核苷酸,其中所述5'同源臂包含SEQ ID NO:80中所示的序列,或者展现与SEQ ID NO:80至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列;和/或所述3'同源臂包含SEQ ID NO:81中所示的序列,或者展现与SEQ ID NO:81至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。

48. 根据权利要求32-47中任一项所述的多核苷酸,其中所述嵌合受体是嵌合抗原受体(CAR)。

49. 根据权利要求33-48中任一项所述的多核苷酸,其中所述结合结构域是或包含抗体或其抗原结合片段。

50. 根据权利要求33-49中任一项所述的多核苷酸,其中所述结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达,任选地其中所述靶抗原是肿瘤抗原。

51. 根据权利要求50所述的多核苷酸,其中所述靶抗原选自 α v β 6整合素(avb6整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPha2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、神经节苷脂GD3、糖蛋白100(gp100)、磷脂酰肌醇蛋白聚糖-3(GPC3)、G蛋白偶联受体C类5族成员D(GPRC5D)、Her2/neu(受体酪氨酸激酶erb-B2)、Her3(erb-B3)、Her4(erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原(HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1(HLA-A1)、人白细胞抗原A2

(HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2 (IL-13R α 2)、激酶插入结构域受体 (kdr)、 κ 轻链、L1细胞粘附分子 (L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A (LRRC8A)、Lewis Y、黑色素瘤相关抗原 (MAGE) -A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素 (MSLN)、c-Met、鼠类巨细胞病毒 (CMV)、粘蛋白1 (MUC1)、MUC16、自然杀伤细胞2族成员D (NKG2D) 配体、黑色素A (MART-1)、神经细胞粘附分子 (NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原 (PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原 (PSCA)、前列腺特异性膜抗原 (PSMA)、受体酪氨酸激酶样孤儿受体1 (ROR1)、存活蛋白、滋养层糖蛋白 (TPBG, 也称为5T4)、肿瘤相关糖蛋白72 (TAG72)、酪氨酸酶相关蛋白1 (TRP1, 也称为TYRP1或gp75)、酪氨酸酶相关蛋白2 (TRP2, 也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体 (VEGFR)、血管内皮生长因子受体2 (VEGFR2)、肾母细胞瘤1 (WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。

52. 根据权利要求33-51中任一项所述的多核苷酸, 其中所述细胞外区域包含间隔子, 任选地其中所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。

53. 根据权利要求52所述的多核苷酸, 其中所述间隔子包含免疫球蛋白铰链区和/或C_H2区域和C_H3区域。

54. 根据权利要求33-53中任一项所述的多核苷酸, 其中所述细胞内区域包含一个或多个共刺激信号传导结构域。

55. 根据权利要求54所述的多核苷酸, 其中所述一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。

56. 根据权利要求32-55中任一项所述的多核苷酸, 其中 (a) 的所述核酸序列按顺序包含: 编码结合结构域、任选地单链Fv片段 (scFv) 的核苷酸序列; 间隔子, 所述间隔子任选地包含来自人免疫球蛋白铰链的序列, 任选地来自IgG1、IgG2或IgG4或其修饰形式的序列, 任选地还包含C_H2区域和/或C_H3区域; 以及任选地来自人CD28的跨膜结构域; 细胞内区域, 所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域。

57. 根据权利要求32-56中任一项所述的多核苷酸, 其中 (a) 的所述核酸序列包含编码至少一种其他蛋白质的核苷酸序列。

58. 根据权利要求57所述的多核苷酸, 其中所述至少一种其他蛋白质是替代标记, 任选地其中所述替代标记是截短的受体, 任选地其中所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。

59. 根据权利要求32-58中任一项所述的多核苷酸, 其中 (a) 的所述核酸序列包含一个或多个多顺反子元件。

60. 根据权利要求59所述的多核苷酸, 其中:

(a) 的所述核酸包含编码所述嵌合受体的一部分的核苷酸序列, 并且所述一个或多个多顺反子元件定位于编码所述嵌合受体的部分的核苷酸序列的上游; 和/或定位于编码所述嵌合受体的部分的核苷酸序列与编码所述至少一种其他蛋白质的核苷酸序列之间; 和/或

所述嵌合受体是CAR, 所述CAR是多链CAR, 并且所述一个或多个多顺反子元件定位于编码多链CAR的一条链的核苷酸序列与编码所述多链CAR的另一条链的核苷酸序列之间。

61. 根据权利要求59或60所述的多核苷酸,其中所述一个或多个多顺反子元件是或包含核糖体跳跃序列,任选地其中所述核糖体跳跃序列是T2A、P2A、E2A或F2A元件。

62. 根据权利要求32-61中任一项所述的多核苷酸,其中(a)的所述核酸序列包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制所述嵌合受体或其部分的表达。

63. 根据权利要求32-62中任一项所述的多核苷酸,其中所述多核苷酸包含于病毒载体中。

64. 根据权利要求63所述的多核苷酸,其中所述病毒载体是AAV载体,任选地其中所述AAV载体是AAV2或AAV6载体。

65. 根据权利要求63所述的多核苷酸,其中所述病毒载体是逆转录病毒载体,任选地慢病毒载体。

66. 根据权利要求32-63中任一项所述的多核苷酸,其为线性多核苷酸,任选地双链多核苷酸或单链多核苷酸。

67. 根据权利要求32-66中任一项所述的多核苷酸,其中所述多核苷酸具有在为或约2500与为或约5000个核苷酸之间、为或约3500与为或约4500个核苷酸之间、或者为或约3750个核苷酸与为或约4250个核苷酸之间的长度。

68. 一种产生基因工程化的T细胞的方法,所述方法包括将根据权利要求32-67中任一项所述的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中。

69. 一种产生基因工程化的T细胞的方法,所述方法包括:

(a) 将一种或多种药剂引入T细胞中,所述一种或多种药剂能够在所述T细胞的内源CD247基因座内的靶位点处诱导遗传破坏;以及

(b) 将根据权利要求32-67中任一项所述的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中。

70. 根据权利要求68或69所述的方法,其中经由同源定向修复(HDR)将编码所述嵌合受体或其部分的所述核酸序列整合于所述内源CD247基因座内。

71. 一种产生基因工程化的T细胞的方法,所述方法包括将包含编码嵌合受体或其部分的核酸序列的多核苷酸引入T细胞中,所述T细胞在所述T细胞的CD247基因座内具有遗传破坏,其中编码所述嵌合受体或其部分的所述核酸序列经由同源定向修复(HDR)整合于所述内源CD247基因座内。

72. 根据权利要求68、70和71中任一项所述的方法,其中所述遗传破坏是通过将一种或多种药剂引入T细胞中来进行,所述一种或多种药剂能够在所述T细胞的内源CD247基因座内的靶位点处诱导遗传破坏。

73. 根据权利要求68-72中任一项所述的方法,其中所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3 ζ 信号传导结构域的细胞内区域,其中所述CD3 ζ 信号传导结构域的至少一个片段是由所述内源CD247基因座的开放阅读框编码。

74. 根据权利要求71-73中任一项所述的方法,其中所述多核苷酸包含与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框的一个或多个区域同源的序列。

75. 根据权利要求71-74中任一项所述的方法,其中编码所述嵌合受体或其部分的所述核酸序列不包含编码3'UTR的序列和/或不包含内含子。

76. 根据权利要求73-75中任一项所述的方法,其中,在由引入所述多核苷酸的细胞表达时,所述嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

77. 根据权利要求73-76中任一项所述的方法,其中所编码的CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。

78. 根据权利要求74-77中任一项所述的方法,其中所述一个或多个同源臂包含5'同源臂和3'同源臂,并且所述多核苷酸包含结构[5'同源臂]-[编码嵌合受体或其部分的核酸序列]-[3'同源臂]。

79. 根据权利要求69和72-78中任一项所述的方法,其中能够诱导遗传破坏的所述一种或多种药剂包含与所述靶位点特异性结合或杂交的DNA结合蛋白或DNA结合核酸、包含DNA靶向蛋白和核酸酶的融合蛋白或者RNA指导核酸酶,任选地其中所述一种或多种药剂包含与所述靶位点特异性结合、识别所述靶位点或与所述靶位点杂交的锌指核酸酶(ZFN)、TAL效应核酸酶(TALEN)或和CRISPR-Cas9组合。

80. 根据权利要求69和72-79中任一项所述的方法,其中所述一种或多种药剂各自包含具有与至少一个靶位点互补的靶向结构域的指导RNA(gRNA)。

81. 根据权利要求80所述的方法,其中所述一种或多种药剂是作为包含gRNA和Cas9蛋白的核糖核蛋白(RNP)复合物引入,任选地其中所述RNP是经由电穿孔、粒子枪、磷酸钙转染、细胞压缩或挤压、任选地经由电穿孔引入。

82. 根据权利要求81所述的方法,其中所述RNP的浓度是为或约1、2、2.5、5、10、20、25、30、40或50 μ M,或者由任两个前述值限定的范围,任选地其中所述RNP的浓度是为或约25 μ M。

83. 根据权利要求80-82中任一项所述的方法,其中所述RNP中所述gRNA与所述Cas9分子的摩尔比是为或约为或约5:1、4:1、3:1、2:1、1:1、1:2、1:3、1:4或1:5,或者由任两个前述值限定的范围,任选地其中所述RNP中所述gRNA与所述Cas9分子的摩尔比是为或约2.6:1。

84. 根据权利要求80-83中任一项所述的方法,其中所述gRNA具有选自以下的靶向结构域序列:CACCUACACUCUCAGGAACA (SEQ ID NO:87);GAAUGACACCAUAGAUGAAG (SEQ ID NO:88);UGAAGAGGAUCCAUCAGC (SEQ ID NO:89);以及UCCAGCAGGUAGCAGAGUUU (SEQ ID NO:90)。

85. 根据权利要求80-84中任一项所述的方法,其中所述gRNA具有CACCUACACUCUCAGGAACA (SEQ ID NO:87)的靶向结构域序列。

86. 根据权利要求80-84中任一项所述的方法,其中所述gRNA具有UGAAGAGGAUCCAUCAGC (SEQ ID NO:89)的靶向结构域序列。

87. 根据权利要求68-86中任一项所述的方法,其中所述T细胞是源自受试者的原代T细胞,任选地其中所述受试者是人。

88. 根据权利要求68-87中任一项所述的方法,其中所述T细胞是CD8⁺T细胞或其亚型,或者CD4⁺T细胞或其亚型。

89. 根据权利要求68-88中任一项所述的方法,其中所述多核苷酸包含于病毒载体中。

90. 根据权利要求89所述的方法,其中所述病毒载体是AAV载体,任选地其中所述AAV载

体是AAV2或AAV6载体。

91. 根据权利要求89所述的方法,其中所述病毒载体是逆转录病毒载体,任选地慢病毒载体。

92. 根据权利要求68-88中任一项所述的方法,其中所述多核苷酸是线性多核苷酸,任选地双链多核苷酸或单链多核苷酸。

93. 根据权利要求69和72-92中任一项所述的方法,其中所述多核苷酸是在引入所述一种或多种药剂之后引入。

94. 根据权利要求93所述的方法,其中所述多核苷酸在引入所述药剂后立即引入,或者在引入所述药剂后约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、6分钟、8分钟、9分钟、10分钟、15分钟、20分钟、30分钟、40分钟、50分钟、60分钟、90分钟、2小时、3小时或4小时内引入。

95. 根据权利要求69和72-94中任一项所述的方法,其中在引入所述一种或多种药剂之前,所述方法包括在刺激或激活所述一种或多种免疫细胞的条件下将所述细胞在体外与一种或多种刺激剂一起孵育,任选地其中所述一种或多种刺激剂包含和抗CD3和/或抗CD28抗体,任选地抗CD3/抗CD28珠,任选地其中珠与细胞的比率为或为约1:1。

96. 根据权利要求69和72-95中任一项所述的方法,其中所述方法还包括在引入所述一种或多种药剂和/或引入所述多核苷酸之前、期间或之后,将所述细胞与一种或多种重组细胞因子一起孵育,任选地其中所述一种或多种重组细胞因子选自IL-2、IL-7和IL-15,任选地其中所述一种或多种重组细胞因子是以选自以下的浓度添加:为或约10U/mL至为或约200U/mL、任选地为或约50IU/mL至为或约100U/mL的浓度的IL-2;0.5ng/mL至50ng/mL、任选地为或约5ng/mL至为或约10ng/mL的浓度的IL-7;和/或0.1ng/mL至20ng/mL、任选地为或约0.5ng/mL至为或约5ng/mL的浓度的IL-15。

97. 根据权利要求95或96所述的方法,其中所述孵育是在引入所述一种或多种药剂和引入所述多核苷酸之后进行长达或大约24小时、36小时、48小时、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或21天,任选地长达或约7天。

98. 根据权利要求68-97中任一项所述的方法,其中通过所述方法生成的多个工程化细胞中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞包含在CD247基因座内至少一个靶位点的遗传破坏;和/或通过所述方法生成的多个工程化细胞中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞表达所述嵌合受体。

99. 一种基因工程化的T细胞或多个基因工程化的T细胞,其是使用根据权利要求68-98中任一项所述的方法来生成。

100. 一种组合物,其包含根据权利要求1-31和99中任一项所述的基因工程化的T细胞;或者根据权利要求1-31和99中任一项所述的多个基因工程化的T细胞。

101. 根据权利要求100所述的组合物,其中所述组合物包含CD4+T细胞和/或CD8+T细胞。

102. 根据权利要求101所述的组合物,其中所述组合物包含CD4+T细胞和CD8+T细胞,并且CD4+与CD8+T细胞的比率为从或从约1:3至3:1,任选地1:1。

103. 根据权利要求100-102中任一项所述的组合物,其中表达所述嵌合受体的细胞构

成所述组合中总细胞的或所述组合中总CD4+T细胞或CD8+T细胞的至少30%、40%、50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多。

104. 一种治疗方法,其包括将根据权利要求1-31和100-103中任一项所述的基因工程化的T细胞、多个基因工程化的T细胞或组合物施用至患有疾病或障碍的受试者。

105. 根据权利要求1-31和100-103中任一项所述的基因工程化的T细胞、多个基因工程化的T细胞或组合物用于治疗疾病或障碍的用途。

106. 根据权利要求1-31和100-103中任一项所述的基因工程化的T细胞、多个基因工程化的T细胞或组合物在制造用于治疗疾病或障碍的药物中的用途。

107. 根据权利要求1-31和100-103中任一项所述的基因工程化的T细胞、多个基因工程化的T细胞或组合物,其用于治疗疾病或障碍。

108. 根据权利要求104-107中任一项所述的方法、用途或用于所述用途的基因工程化的T细胞、多个基因工程化的T细胞或组合物,其中所述疾病或障碍是癌症或肿瘤。

109. 根据权利要求108所述的方法、用途或用于所述用途的基因工程化的T细胞、多个基因工程化的T细胞或组合物,其中所述癌症或所述肿瘤是血液恶性肿瘤,任选地淋巴瘤、白血病或浆细胞恶性肿瘤。

110. 根据权利要求108所述的方法、用途或用于所述用途的基因工程化的T细胞、多个基因工程化的T细胞或组合物,其中所述癌症或所述肿瘤是实体瘤,任选地其中所述实体瘤是非小细胞肺癌(NSCLC)或头颈鳞状细胞癌(HNSCC)。

111. 一种试剂盒,其包含:

能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂;以及
根据权利要求32-67中任一项所述的多核苷酸。

112. 一种试剂盒,其包含:

能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂;以及
包含编码嵌合受体或其部分的核酸序列的多核苷酸,其中编码所述嵌合受体或其部分的所述核酸序列被靶向以供经由同源定向修复(HDR)整合于所述靶位点处或附近;以及
用于进行根据权利要求68-98中任一项所述的方法的说明书。

从修饰的CD247基因座表达嵌合受体的细胞、相关多核苷酸和方法

相关申请的交叉引用

[0001] 本申请要求2019年5月1日提交的标题为“从修饰的CD247基因座表达嵌合受体的细胞、相关多核苷酸和方法 (CELLS EXPRESSING A CHIMERIC RECEPTOR FROM A MODIFIED CD247 LOCUS, RELATED POLYNUCLEOTIDES AND METHODS)”的美国临时申请号62/841,578的优先权,所述申请的内容通过引用以其整体并入本文。

通过引用并入序列列表

[0002] 本申请是与电子格式的序列列表一起提交。序列列表以2020年4月28日创建的标题为735042015840SeqList.txt的文件提供,其大小为172千字节。将在电子格式的序列列表中的信息通过引用以其整体并入。

技术领域

[0003] 本公开文本涉及表达嵌合受体的工程化的免疫细胞,例如T细胞,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些实施方案中,所述工程化的免疫细胞含有修饰的CD247基因座,所述基因座编码所述嵌合受体或其部分。在一些实施方案中,CD3 ζ 链的至少一部分由CD247基因组基因座编码。还提供含有工程化的免疫细胞的细胞组合物、用于工程化细胞的核酸,以及用于产生所述工程化细胞的方法、试剂盒和制品,如通过靶向编码嵌合受体的一部分的转基因以供整合至CD247基因组基因座的区域中。在一些实施方案中,所述工程化细胞,例如T细胞,可以与细胞疗法结合使用,包括与癌症免疫疗法结合使用,所述癌症免疫疗法包括所述工程化细胞的过继转移。

背景技术

[0004] 利用嵌合受体(如嵌合抗原受体(CAR))识别与疾病相关的抗原的过继细胞疗法代表用于治疗癌症和其他疾病的有吸引力的治疗方式。需要改进的策略用于工程化T细胞以表达嵌合受体,如用于过继免疫疗法,例如,用于治疗癌症、感染性疾病和自身免疫疾病。提供符合此类需要的方法以及用于所述方法中的细胞、组合物和试剂盒。

发明内容

[0005] 本文提供基因工程化的T细胞以及与基因工程化的T细胞相关的组合物、方法、用途、试剂盒和制品。在一些任何所提供的实施方案中,所述基因工程化的T细胞包含修饰的分化群247 (CD247) 基因座。在一些任何实施方案中,所述修饰的CD247基因座包含编码嵌合受体或其部分的转基因序列。在所提供的实施方案中,所述转基因序列与内源CD247基因座的开放阅读框或其部分序列同框。因此,在所提供的实施方案中,所述修饰的CD247基因座编码嵌合受体,所述嵌合受体包括从所述转基因序列编码的序列和从所述内源CD247基因座编码的序列。在特定实施方案中,所述嵌合受体含有包含CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,其中所述CD3 ζ 信号传导结构域(例如整个CD3 ζ 信号传导结构域,或所述CD3 ζ

信号传导结构域的至少一部分)是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列(例如,开放阅读框)编码。

[0006] 本文提供基因工程化的T细胞,其含有修饰的CD247基因座。在一些任何实施方案中,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些任何实施方案中,所述核酸序列包含编码所述嵌合受体的一部分的转基因序列,所述转基因序列已经整合于内源CD247基因座处。在一些任何实施方案中,所述整合通过同源定向修复(HDR)来进行。在一些任何实施方案中,所述嵌合受体中细胞内区域的所述CD3 ζ 信号传导结构域的全部或片段是由所述内源CD247基因座的开放阅读框或其部分序列编码。在一些任何实施方案中,所述核酸序列包含以下的框内融合物:(i) 编码所述嵌合受体的一部分的转基因序列,和(ii) 所述内源CD247基因座的开放阅读框或其部分序列。在特定实施方案中,所述修饰的CD247基因座编码嵌合受体,所述嵌合受体含有包含CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,其中所述CD3 ζ 信号传导结构域(例如整个CD3 ζ 信号传导结构域,或所述CD3 ζ 信号传导结构域的至少一部分)是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。

[0007] 本文提供基因工程化的T细胞,其含有修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3 ζ 信号传导结构域的细胞内区域,其中所述核酸序列包含以下的框内融合物:(i) 编码所述嵌合受体的一部分的转基因序列,和(ii) 编码所述CD3 ζ 信号传导结构域的内源CD247基因座的开放阅读框或其部分序列。在特定实施方案中,所述修饰的CD247基因座编码嵌合受体,所述嵌合受体含有包含CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。

[0008] 在一些任何实施方案中,所述转基因序列与所述内源CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框。

[0009] 在一些任何实施方案中,所述转基因序列不包含编码3'UTR的序列。在一些任何实施方案中,所述转基因序列不包含内含子。

[0010] 在一些任何实施方案中,所述转基因序列编码所述CD3 ζ 信号传导结构域的片段。例如,在特定实施方案中,所述嵌合受体的CD3 ζ 信号传导结构域或其片段由所述转基因序列的序列以及由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列(例如,开放阅读框)一起编码。

[0011] 在一些任何实施方案中,所述转基因序列不编码所述CD3 ζ 信号传导结构域或其片段。例如,在特定实施方案中,所述嵌合受体的CD3 ζ 信号传导结构域的全部或全长或其片段是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。

[0012] 在一些任何实施方案中,所述开放阅读框或其部分序列包含所述内源CD247基因座的至少一个内含子和至少一个外显子。在一些任何实施方案中,所述开放阅读框或其部分序列编码所述内源CD247基因座的3'UTR。

[0013] 在一些任何实施方案中,所述转基因序列位于所述内源CD247基因座的开放阅读

框中外显子1的下游和外显子8的上游。在一些任何实施方案中,所述转基因序列位于所述内源CD247基因座的开放阅读框中外显子1的下游和外显子3的上游。

[0014] 在一些任何实施方案中,所编码嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段(如整个CD3 ζ 信号传导结构域)是由所述内源CD247基因座的开放阅读框或其部分序列编码。在一些任何实施方案中,所述CD3 ζ 信号传导结构域是由包含所述内源CD247基因座的开放阅读框的外显子2和外显子3-8中至少一部分的核苷酸序列编码。在一些任何实施方案中,所述CD3 ζ 信号传导结构域是由不包含所述内源CD247基因座的开放阅读框的外显子1、不包含外显子1的全长和/或不包含全长外显子2的核苷酸序列编码。

[0015] 在一些任何实施方案中,所编码的嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

[0016] 在一些任何实施方案中,所编码的CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。在一些实施方案中,所编码的CD3 ζ 信号传导结构域包含SEQ ID NO:13中所示的序列。在一些实施方案中,所编码的CD3 ζ 信号传导结构域包含SEQ ID NO:14中所示的序列。在一些实施方案中,所编码的CD3 ζ 信号传导结构域包含SEQ ID NO:15中所示的序列。

[0017] 在一些任何实施方案中,所述嵌合受体是或包含功能性非T细胞受体(非TCR)抗原受体。

[0018] 在一些任何实施方案中,所述嵌合受体是嵌合抗原受体(CAR)。在一些任何实施方案中,所述嵌合受体还包含细胞外区域和/或跨膜结构域。

[0019] 在一些任何实施方案中,所述转基因序列包含编码所述嵌合受体的一个或多个区域的核苷酸序列。在一些任何实施方案中,所述转基因序列包含编码细胞外区域、跨膜结构域和/或所述细胞内区域的一部分中的一个或多个的核苷酸序列。在一些任何实施方案中,所述细胞外区域包含结合结构域。在一些任何实施方案中,所述结合结构域是抗体或其抗原结合片段。在一些任何实施方案中,所述结合结构域包含抗体或其抗原结合片段。

[0020] 在一些任何实施方案中,所述结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达。在一些任何实施方案中,所述靶抗原是肿瘤抗原。在一些任何实施方案中,所述靶抗原选自 α v β 6整合素(avb6整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPha2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、神经节苷脂GD3、糖蛋白100(gp100)、磷脂酰肌醇蛋白聚糖-3(GPC3)、G蛋白偶联受体C类5族成员D(GPRC5D)、Her2/neu(受体酪氨酸激酶erb-B2)、Her3(erb-B3)、Her4(erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原(HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1(HLA-

A1)、人白细胞抗原A2 (HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2 (IL-13R α 2)、激酶插入结构域受体 (kdr)、 κ 轻链、L1细胞粘附分子 (L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A (LRRC8A)、Lewis Y、黑色素瘤相关抗原 (MAGE) -A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素 (MSLN)、c-Met、鼠类巨细胞病毒 (CMV)、粘蛋白1 (MUC1)、MUC16、自然杀伤细胞2族成员D (NKG2D) 配体、黑色素A (MART-1)、神经细胞粘附分子 (NCAM)、癌胚抗原、黑色素瘤优先表达抗原 (PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原 (PSCA)、前列腺特异性膜抗原 (PSMA)、受体酪氨酸激酶样孤儿受体1 (ROR1)、存活蛋白、滋养层糖蛋白 (TPBG, 也称为5T4)、肿瘤相关糖蛋白72 (TAG72)、酪氨酸酶相关蛋白1 (TRP1, 也称为TYRP1或gp75)、酪氨酸酶相关蛋白2 (TRP2, 也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体 (VEGFR)、血管内皮生长因子受体2 (VEGFR2)、肾母细胞瘤1 (WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。

[0021] 在一些任何实施方案中,所述细胞外区域包含间隔子。在一些任何实施方案中,所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。在一些任何实施方案中,所述间隔子包含免疫球蛋白铰链区。在一些任何实施方案中,所述间隔子包含C_H2区域和C_H3区域。

[0022] 在一些任何实施方案中,由所述转基因序列编码的细胞内区域的部分包含一个或多个共刺激信号传导结构域。在一些任何实施方案中,所述一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。在一些实施方案中,所述共刺激信号传导结构域是人CD28的信号传导结构域。在一些实施方案中,所述共刺激信号传导结构域是人4-1BB的信号传导结构域。在一些实施方案中,所述共刺激信号传导结构域是人ICOS的信号传导结构域。在一些任何实施方案中,所述一个或多个共刺激信号传导结构域包含4-1BB (如人4-1BB) 的细胞内信号传导结构域。

[0023] 在一些任何实施方案中,所述修饰的CD247基因座编码嵌合受体,所述嵌合受体从其N末端至C末端按顺序包含:所述细胞外结合结构域、所述间隔子、所述跨膜结构域和细胞内信号传导区域。在特定实施方案中,所述细胞内区域含有CD3zeta (CD3 ζ) 信号传导结构域,其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞 (如T细胞) 的内源CD247基因座 (编码CD3 ζ 的基因组基因座) 处的基因组序列编码。

[0024] 在一些任何实施方案中,所述转基因序列按顺序包含:编码细胞外结合结构域的核苷酸序列;间隔子;和跨膜结构域;共刺激信号传导结构域。在一些任何实施方案中,所述修饰的CD247基因座按顺序包含:编码细胞外结合结构域的核苷酸序列;间隔子;和跨膜结构域;以及含有共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在特定实施方案中,所述细胞内信号传导区域含有共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域,其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞 (如T细胞) 的内源CD247基因座 (编码CD3 ζ 的基因组基因座) 处的基因组序列 (例如,开放阅读框) 编码。在一些任何实施方案中,所述转基因序列按顺序包含:编码细胞外结合结构域的核苷酸序列,所述细胞外结合结构域是scFv;间隔子,所述间隔子包括来自人免疫球蛋白铰链的序列,所述人免疫球蛋白铰链来自IgG1、IgG2或IgG4或其修饰形式,所述间隔子还包括C_H2区域和/或C_H3区域;和跨膜结构域,所述跨膜结构域来自人

CD28;共刺激信号传导结构域,所述共刺激信号传导结构域来自人4-1BB。在一些任何实施方案中,所述修饰的CD247基因座按顺序包含:编码细胞外结合结构域的核苷酸序列,所述细胞外结合结构域是scFv;间隔子,所述间隔子包括来自人免疫球蛋白铰链的序列,所述人免疫球蛋白铰链来自IgG1、IgG2或IgG4或其修饰形式,所述间隔子还包括C_H2区域和/或C_H3区域;和跨膜结构域,所述跨膜结构域来自人CD28;以及细胞内区域,所述细胞内区域含有来自人4-1BB的共刺激信号传导结构域和所述CD3 ζ 信号传导结构域。在特定实施方案中,所述细胞内区域含有共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域,其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列(例如,开放阅读框)编码。

[0025] 在一些任何实施方案中,所述嵌合受体是CAR,所述CAR是多链CAR。

[0026] 在一些任何实施方案中,所述转基因序列包含编码至少一种其他蛋白质的核苷酸序列。例如,所述至少一种其他蛋白质可以是所述CAR的另一条链。在一些例子中,所述至少一种其他蛋白质是替代标记或截短的受体,用于在细胞上与所述嵌合受体共表达。在一些任何实施方案中,所述转基因序列包含一个或多个多顺反子元件,例如它们分离所述嵌合受体与所述一种或多种其他蛋白质。在一些任何实施方案中,所述一个或多个多顺反子元件定位于编码所述嵌合受体的所述部分的核苷酸序列与编码所述至少一种其他蛋白质的核苷酸序列之间。在一些任何实施方案中,所述至少一种其他蛋白质是替代标记。在一些任何实施方案中,所述替代标记是截短的受体。在一些任何实施方案中,所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。在一些任何实施方案中,所述嵌合受体是多链CAR,并且多顺反子元件定位于编码所述多链CAR的一条链的核苷酸序列与编码所述多链CAR的另一条链的核苷酸序列之间。在一些任何实施方案中,所述一个或多个多顺反子元件位于编码所述嵌合受体的所述部分的核苷酸序列的上游。在一些任何实施方案中,所述一个或多个多顺反子元件是或包含核糖体跳跃序列。在一些任何实施方案中,所述核糖体跳跃序列是T2A、P2A、E2A或F2A元件。

[0027] 在一些任何实施方案中,所述修饰的CD247基因座包含所述内源CD247基因座的启动子和/或调节或控制元件,所述启动子和/或调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。在一些任何实施方案中,所述修饰的基因座包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。在一些任何实施方案中,所述一个或多个异源调节或控制元件包括启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、剪接接纳体序列和/或剪接供体序列。在一些任何实施方案中,所述异源启动子是或包含人延伸因子1 α (EF1 α) 启动子或MND启动子或其变体。

[0028] 在一些任何实施方案中,所述T细胞是源自受试者的原代T细胞。在一些任何实施方案中,所述受试者是人。在一些任何实施方案中,所述T细胞是CD8+T细胞或其亚型。在一些任何实施方案中,所述T细胞是CD4+T细胞或其亚型。在一些任何实施方案中,所述T细胞源自多潜能或多能细胞。在一些任何实施方案中,所述多能细胞是iPSC。在一些任何实施方案中,所述T细胞源自多潜能或多能细胞,其是iPSC。

[0029] 本文还提供多核苷酸,如可以用于将编码嵌合受体的转基因序列整合至所述CD247基因座中的多核苷酸。在一些任何实施方案中,所述多核苷酸包括(a) 编码嵌合受体

或其部分的核酸序列;以及 (b) 与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框或其部分序列的一个或多个区域同源的序列。在一些任何实施方案中,整合至所述CD247基因座中的所述多核苷酸编码包含细胞内区域(例如,包含CD3 ζ 信号传导结构域的细胞内区域)的嵌合受体,并且(a)的所述核酸序列是编码所述嵌合受体的一部分的核酸序列,其中所述部分不包括所述嵌合受体的完整细胞内区域。在一些实施方案中,所述完整细胞内区域包括CD3zeta (CD3 ζ) 信号传导结构域。在一些实施方案中,所述完整细胞内区域包括共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域。在一些实施方案中,(a)的所述核酸序列编码所述嵌合受体的一部分,所述部分不包括编码CD3zeta (CD3 ζ) 信号传导结构域的整个或全长序列。在一些实施方案中,(a)的所述核酸序列不含编码所述CD3zeta (CD3 ζ) 信号传导结构域的任何序列。在一些实施方案中,(a)的所述核酸序列编码细胞内区域,所述细胞内区域包含所述CD3zeta (CD3 ζ) 信号传导结构域的片段。在任何此类例子中,(a)的所述核酸序列可以编码所述细胞内区域的共刺激信号传导结构域。

[0030] 本文还提供多核苷酸,其含有(a)编码嵌合受体的一部分的核酸序列,所述嵌合受体包含细胞内区域(例如,包含CD3 ζ 信号传导结构域的细胞内区域),其中所述嵌合受体的所述部分包括所述嵌合受体的不完整细胞内区域;以及(b)与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框或其部分序列的一个或多个区域同源的序列。在一些实施方案中,所述多核苷酸可以用于将编码所述嵌合受体的转基因序列整合至所述CD247基因座中。在一些实施方案中,所述完整细胞内区域包括CD3zeta (CD3 ζ) 信号传导结构域。在一些实施方案中,所述完整细胞内区域包括共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域。在一些实施方案中,(a)的所述核酸序列编码所述嵌合受体的一部分,所述部分不包括编码CD3zeta (CD3 ζ) 信号传导结构域的整个或全长序列。在一些实施方案中,(a)的所述核酸序列不含编码所述CD3zeta (CD3 ζ) 信号传导结构域的任何序列。在一些实施方案中,(a)的所述核酸序列编码细胞内区域,所述细胞内区域包含所述CD3zeta (CD3 ζ) 信号传导结构域的片段。在任何此类例子中,(a)的所述核酸序列可以编码所述细胞内区域的共刺激信号传导结构域。

[0031] 在一些任何实施方案中,所述嵌合受体的完整细胞内区域包含CD3zeta (CD3 ζ) 信号传导结构域或其片段,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述细胞内区域的至少一部分由所述内源CD247基因座的开放阅读框或其部分序列编码。

[0032] 在一些任何实施方案中,编码所述嵌合受体的所述部分的所述核酸序列和所述一个或多个同源臂一起包含编码所述嵌合受体的细胞内区域的核苷酸序列的至少一个片段,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述细胞内区域的至少一部分包含由所述CD247基因座的开放阅读框或其部分序列编码的CD3 ζ 信号传导结构域或其片段。

[0033] 在一些任何实施方案中,(a)的所述核酸序列不包含编码3'UTR的序列。在一些任何实施方案中,(a)的所述核酸序列不包含内含子。

[0034] 在一些任何实施方案中,(a)的所述核酸序列编码所述CD3 ζ 信号传导结构域的片段。在此类实施方案中,在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。例如,在特定实施方案中,所述嵌合受体的CD3 ζ 信号

传导结构域或其片段是由所述转基因序列的序列以及由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列一起编码。

[0035] 在一些任何实施方案中,(a)的所述核酸序列不编码所述CD3 ζ 信号传导结构域或其片段。在此类实施方案中,在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述嵌合受体的CD3 ζ 信号传导结构域的全部或全长或其片段是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。

[0036] 在一些任何实施方案中,所述内源CD247基因座的开放阅读框或其部分序列包含所述内源CD247基因座的至少一个内含子和至少一个外显子。在一些任何实施方案中,所述开放阅读框或其部分序列编码所述内源CD247基因座的3'UTR。

[0037] 在一些任何实施方案中,在所述嵌合受体由引入所述多核苷酸的细胞表达时,所编码的嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段(如整个CD3 ζ 信号传导结构域)是由所述内源CD247基因座的开放阅读框或其部分序列编码。

[0038] 在一些任何实施方案中,(a)的所述核酸序列是如下序列,所述序列对于T细胞(如人T细胞)的所述内源基因组CD247基因座的开放阅读框是外源或异源的。

[0039] 在一些任何实施方案中,(a)的所述核酸序列包含与所述一个或多个同源臂中包含的所述CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框的核苷酸序列。

[0040] 在一些任何实施方案中,所述内源CD247基因座的开放阅读框或其部分序列的所述一个或多个区域是或包含位于所述CD247基因座的开放阅读框的外显子8上游的序列。在一些任何实施方案中,所述开放阅读框的所述一个或多个区域是或包含位于所述CD247基因座的开放阅读框的外显子3上游的序列。在一些任何实施方案中,所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列包括所述CD247基因座的开放阅读框的外显子3。在一些任何实施方案中,所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列包括所述CD247基因座的开放阅读框的外显子2的至少一部分。在一些任何实施方案中,所述一个或多个同源臂不包含所述内源CD247基因座的开放阅读框的外显子1,不包含外显子1的全长和/或不包含外显子2的全长。

[0041] 在一些任何实施方案中,在由引入所述多核苷酸的细胞表达时,所编码的嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。在一些任何实施方案中,由所述嵌合受体编码的完整细胞内区域的所述CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。在一些实施方案中,所述CD3 ζ 信号传导结构域具有SEQ ID NO:13中所示的序列。在一些实施方案中,所述CD3 ζ 信号传导结构域具有SEQ ID NO:14中所示的序列。在一些实施方案中,所述CD3 ζ 信号传导结构域具有SEQ ID NO:15中所示的序列。

[0042] 在一些任何实施方案中,所述一个或多个同源臂包含5'同源臂和3'同源臂。在一些任何实施方案中,所述多核苷酸包含结构[5'同源臂]-[(a)的核酸序列]-[3'同源臂]。

[0043] 在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有为或约50至为或约2000个核苷酸、为或约100至为或约1000个核苷酸、为或约100至为或约750个核苷酸、为或约100至为或约600个核苷酸、为或约100至为或约400个核苷酸、为或约100至为或约

300个核苷酸、为或约100至为或约200个核苷酸、为或约200至为或约1000个核苷酸、为或约200至为或约750个核苷酸、为或约200至为或约600个核苷酸、为或约200至为或约400个核苷酸、为或约200至为或约300个核苷酸、为或约300至为或约1000个核苷酸、为或约300至为或约750个核苷酸、为或约300至为或约600个核苷酸、为或约300至为或约400个核苷酸、为或约400至为或约1000个核苷酸、为或约400至为或约750个核苷酸、为或约400至为或约600个核苷酸、为或约600至为或约1000个核苷酸、为或约600至为或约750个核苷酸或者为或约750至为或约1000个核苷酸的长度。在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有为或约200、300、400、500、600、700或800个核苷酸的长度,或者任何前述值之间的任何值。在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有大于或大于约300个核苷酸的长度。在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有为或约400、500或600个核苷酸的长度,或者任何前述值之间的任何值。

[0044] 在一些任何实施方案中,所述5'同源臂包含SEQ ID NO:80中所示的序列,或者展现与SEQ ID NO:80至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。在一些实施方案中,所述5'同源臂包含SEQ ID NO:80中所示的序列。在一些实施方案中,所述5'同源臂由SEQ ID NO:80中所示的序列组成或基本上由其组成。在一些任何实施方案中,所述3'同源臂包含SEQ ID NO:81中所示的序列,或者展现与SEQ ID NO:81至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。在一些实施方案中,所述3'同源臂包含SEQ ID NO:81中所示的序列。在一些实施方案中,所述3'同源臂由SEQ ID NO:81中所示的序列组成或基本上由其组成。

[0045] 在一些任何实施方案中,所述嵌合受体是或包含功能性非T细胞受体(非TCR)抗原受体。

[0046] 在一些任何实施方案中,所述嵌合受体是嵌合抗原受体(CAR)。

[0047] 在一些任何实施方案中,(a)的所述核酸序列包含编码细胞外区域的核苷酸序列、编码跨膜结构域和/或所述细胞内区域的一部分的核苷酸序列。在一些任何实施方案中,(a)的所述核酸序列包含编码细胞外区域的核苷酸序列、编码跨膜结构域的核苷酸序列和编码所述细胞内区域的一部分的核苷酸序列。在一些任何实施方案中,所述细胞外区域包含结合结构域。在一些任何实施方案中,所述结合结构域是或包含抗体或其抗原结合片段。

[0048] 在一些任何实施方案中,所述结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达。在一些任何实施方案中,所述靶抗原是肿瘤抗原。在一些任何实施方案中,所述靶抗原选自 α v β 6整合素(α v β 6整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPHa2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、

神经节苷脂GD3、糖蛋白100 (gp100)、磷脂酰肌醇蛋白聚糖-3 (GPC3)、G蛋白偶联受体C类5族成员D (GPC5D)、Her2/neu (受体酪氨酸激酶erb-B2)、Her3 (erb-B3)、Her4 (erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原 (HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1 (HLA-A1)、人白细胞抗原A2 (HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2 (IL-13R α 2)、激酶插入结构域受体 (kdr)、 κ 轻链、L1细胞粘附分子 (L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A (LRRC8A)、Lewis Y、黑色素瘤相关抗原 (MAGE) -A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素 (MSLN)、c-Met、鼠类巨细胞病毒 (CMV)、粘蛋白1 (MUC1)、MUC16、自然杀伤细胞2族成员D (NKG2D) 配体、黑色素A (MART-1)、神经细胞粘附分子 (NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原 (PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原 (PSCA)、前列腺特异性膜抗原 (PSMA)、受体酪氨酸激酶样孤儿受体1 (ROR1)、存活蛋白、滋养层糖蛋白 (TPBG, 也称为5T4)、肿瘤相关糖蛋白72 (TAG72)、酪氨酸酶相关蛋白1 (TRP1, 也称为TYRP1或gp75)、酪氨酸酶相关蛋白2 (TRP2, 也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体 (VEGFR)、血管内皮生长因子受体2 (VEGFR2)、肾母细胞瘤1 (WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。

[0049] 在一些任何实施方案中,所述细胞外区域包含间隔子。在一些任何实施方案中,所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。在一些任何实施方案中,所述间隔子包含免疫球蛋白铰链区。在一些任何实施方案中,所述间隔子包含C_H2区域和C_H3区域。

[0050] 在一些任何实施方案中,由a)的核酸编码的所述细胞内区域的所述部分包含一个或多个共刺激信号传导结构域。在一些任何实施方案中,所述一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。在一些实施方案中,所述共刺激信号传导结构域是人CD28的信号传导结构域。在一些实施方案中,所述共刺激信号传导结构域是人4-1BB的信号传导结构域。在一些实施方案中,所述共刺激信号传导结构域是人ICOS的信号传导结构域。在一些任何实施方案中,所述一个或多个共刺激信号传导结构域包含4-1BB (如人4-1BB)的细胞内信号传导结构域。

[0051] 在一些任何实施方案中,在所述嵌合受体由引入所述多核苷酸的细胞表达时,所编码的嵌合受体从其N末端至C末端按顺序包含:所述细胞外结合结构域、所述间隔子、所述跨膜结构域和细胞内信号传导区域。在特定实施方案中,在由细胞 (如T细胞) 表达时,所编码的嵌合受体的所述细胞内区域含有CD3zeta (CD3 ζ) 信号传导结构域,其中所述整个CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述内源CD247基因座 (编码CD3 ζ 的基因组基因座) 处的基因组序列编码。

[0052] 在一些任何实施方案中,(a)的所述序列按顺序包含:编码细胞外结合结构域的核苷酸序列;间隔子;和跨膜结构域;以及共刺激信号传导结构域。在一些任何实施方案中,(a)的所述序列按顺序包含:编码细胞外结合结构域的核苷酸序列;间隔子;跨膜结构域;以及含有共刺激信号传导结构域和所述CD3 ζ 信号传导结构域的片段的细胞内信号传导区域。在特定实施方案中,在由细胞 (如T细胞) 表达时,所述多核苷酸编码具有细胞内信号传导区域的嵌合受体,所述细胞内信号传导区域含有共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域,其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是

由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。

[0053] 在一些任何实施方案中,(a)的所述核酸序列按顺序包含:编码细胞外结合结构域的核苷酸序列,所述细胞外结合结构域是scFv;间隔子,所述间隔子包括来自人免疫球蛋白铰链的序列,所述人免疫球蛋白铰链来自IgG1、IgG2或IgG4或其修饰形式,并且所述间隔子还包括C_H2区域和/或C_H3区域;和跨膜结构域,所述跨膜结构域来自人CD28;以及共刺激信号传导结构域,所述共刺激信号传导结构域来自人4-1BB。在一些任何实施方案中,(a)的所述序列按顺序包含:编码细胞外结合结构域的核苷酸序列,所述细胞外结合结构域是scFv;间隔子,所述间隔子包括来自人免疫球蛋白铰链的序列,所述人免疫球蛋白铰链来自IgG1、IgG2或IgG4或其修饰形式,并且所述间隔子还包括C_H2区域和/或C_H3区域;跨膜结构域,所述跨膜结构域来自人CD28;以及细胞内区域,所述细胞内区域含有来自人4-1BB的共刺激信号传导结构域和所述CD3 ζ 信号传导结构域的片段。在特定实施方案中,在由细胞(如T细胞)表达时,所述多核苷酸编码具有细胞内信号传导区域的嵌合受体,所述细胞内信号传导区域含有人4-1BB共刺激信号传导结构域和CD3zeta (CD3 ζ) 信号传导结构域,其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。在一些任何实施方案中,在将所述多核苷酸引入T细胞中之后,所述修饰的CD247基因座按顺序包含:编码细胞外结合结构域的核苷酸序列,所述细胞外结合结构域是scFv;间隔子,所述间隔子包括来自人免疫球蛋白铰链的序列,所述人免疫球蛋白铰链来自IgG1、IgG2或IgG4或其修饰形式,并且所述间隔子还包括C_H2区域和/或C_H3区域;和跨膜结构域,所述跨膜结构域来自人CD28;共刺激信号传导结构域,所述共刺激信号传导结构域来自人4-1BB。

[0054] 在一些任何实施方案中,所述CAR是多链CAR。在一些任何实施方案中,(a)的所述核酸序列包含编码至少一种其他蛋白质的核苷酸序列。

[0055] 在一些任何实施方案中,(a)的所述核酸序列包含一个或多个多顺反子元件。在一些任何实施方案中,所述一个或多个多顺反子元件定位于编码所述嵌合受体的所述部分的核苷酸序列与编码所述至少一种其他蛋白质的核苷酸序列之间。在一些任何实施方案中,所述至少一种其他蛋白质是替代标记。在一些任何实施方案中,所述替代标记是截短的受体。在一些任何实施方案中,所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。在一些任何实施方案中,所述嵌合受体是多链CAR,并且多顺反子元件定位于编码所述多链CAR的一条链的核苷酸序列与编码所述多链CAR的另一条链的核苷酸序列之间。在一些任何实施方案中,所述一个或多个多顺反子元件位于编码所述嵌合受体的所述部分的核苷酸序列的上游。在一些任何实施方案中,所述一个或多个多顺反子元件是或包含核糖体跳跃序列。在一些任何实施方案中,所述核糖体跳跃序列是T2A、P2A、E2A或F2A元件。

[0056] 在一些任何实施方案中,在将所述多核苷酸引入T细胞中之后,所述修饰的CD247基因座包含所述内源CD247基因座的启动子和/或调节或控制元件,所述启动子和/或调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。在一些任何实施方案中,所述修饰的基因座包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。在一些任何实施方

案中,所述一个或多个异源调节或控制元件包括启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、剪接接纳体序列和/或剪接供体序列。在一些任何实施方案中,所述异源启动子是或包含人延伸因子1 α (EF1 α) 启动子或MND启动子或其变体。

[0057] 在一些任何实施方案中,(a)的所述核酸序列包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。在一些任何实施方案中,所述一个或多个异源调节或控制元件包括启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、剪接接纳体序列和/或剪接供体序列。在一些任何实施方案中,所述异源启动子是或包含人延伸因子1 α (EF1 α) 启动子或MND启动子或其变体。

[0058] 在一些任何实施方案中,所述多核苷酸包含于病毒载体中。在一些任何实施方案中,所述病毒载体是AAV载体。在一些任何实施方案中,所述AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7或AAV8载体。在一些任何实施方案中,所述AAV载体是AAV2或AAV6载体。在一些任何实施方案中,所述病毒载体是逆转录病毒载体。在一些任何实施方案中,所述病毒载体是慢病毒载体。

[0059] 在一些任何实施方案中,所述多核苷酸是线性多核苷酸。在一些任何实施方案中,双链多核苷酸或单链多核苷酸。

[0060] 在一些任何实施方案中,所述多核苷酸具有至少或至少约2500、2750、3000、3250、3500、3750、4000、4250、4500、4760、5000、5250、5500、5750、6000、7000、7500、8000、9000或10000个核苷酸的长度,或者任何前述值之间的任何值。在一些任何实施方案中,所述多核苷酸具有在为或约2500与为或约5000个核苷酸之间、为或约3500与为或约4500个核苷酸之间、或者为或约3750个核苷酸与为或约4250个核苷酸之间的长度。

[0061] 本文还提供产生基因工程化的T细胞的方法,所述方法涉及将本文提供的任何实施方案的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中。

[0062] 本文还提供产生基因工程化的T细胞的方法,所述方法涉及:(a)将一种或多种药剂引入T细胞中,所述一种或多种药剂能够在所述T细胞的内源CD247基因座内的靶位点处诱导遗传破坏;以及(b)将本文所述的任何多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中,其中所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。

[0063] 在一些任何实施方案中,所述多核苷酸包含编码嵌合受体或其部分的核酸序列,并且编码嵌合受体或其部分的所述核酸序列经由同源定向修复(HDR)整合于所述内源CD247基因座内。

[0064] 本文还提供产生基因工程化的T细胞的方法,所述方法涉及将包含编码嵌合受体或其部分的核酸序列的多核苷酸引入T细胞中,所述T细胞在所述T细胞的CD247基因座内具有遗传破坏,其中编码所述嵌合受体或其部分的所述核酸序列经由同源定向修复(HDR)整合于所述内源CD247基因座内。

[0065] 在一些任何实施方案中,所述遗传破坏是通过将一种或多种药剂引入T细胞中来进行,所述一种或多种药剂能够在所述T细胞的内源CD247基因座内的靶位点处诱导遗传破坏。

[0066] 在一些任何实施方案中,所述方法产生修饰的CD247基因座,所述修饰的CD247基

因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。

[0067] 在一些任何实施方案中,编码嵌合受体或其部分的所述核酸序列编码所述嵌合受体的一部分。在一些任何实施方案中,所述多核苷酸还包含与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框的一个或多个区域同源的序列。

[0068] 在一些任何实施方案中,所述嵌合受体的完整细胞内区域包含CD3zeta (CD3 ζ) 信号传导结构域或其片段,其中所述细胞内区域的至少一部分是由通过所述方法生成的细胞中所述内源CD247基因座的开放阅读框或其部分序列编码。

[0069] 在一些任何实施方案中,编码所述嵌合受体的所述部分的所述核酸序列和所述一个或多个同源臂一起包含编码所述嵌合受体的细胞内区域的核苷酸序列的至少一个片段,其中所述细胞内区域的至少一部分包含由通过所述方法生成的细胞中所述CD247基因座的开放阅读框或其部分序列编码的CD3 ζ 信号传导结构域或其片段。

[0070] 在一些任何实施方案中,编码嵌合受体或其部分的所述核酸序列不包含编码3' UTR的序列。在一些任何实施方案中,在通过所述方法生成的细胞中,编码嵌合受体或其部分的所述核酸序列编码所述CD3 ζ 信号传导结构域的片段。在一些任何实施方案中,在通过所述方法生成的细胞中,编码嵌合受体或其部分的所述核酸序列不编码所述CD3 ζ 信号传导结构域或其片段。在一些任何实施方案中,所编码的嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段(如整个CD3 ζ 信号传导结构域)是由通过所述方法生成的细胞中所述内源CD247基因座的开放阅读框或其部分序列编码。

[0071] 在一些任何实施方案中,编码嵌合受体或其部分的核酸序列是如下序列,所述序列对于T细胞(如人T细胞)的所述内源基因组CD247基因座的开放阅读框是外源或异源的。

[0072] 在一些任何实施方案中,编码嵌合受体或其部分的所述核酸序列包含与所述一个或多个同源臂中包含的所述CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框的核苷酸序列。

[0073] 在一些任何实施方案中,在由引入所述多核苷酸的细胞表达时,所述嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。在一些任何实施方案中,完整细胞内区域的所述CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。在一些实施方案中,所述CD3 ζ 信号传导结构域包含SEQ ID NO:13中所示的序列。在一些实施方案中,所述CD3 ζ 信号传导结构域包含SEQ ID NO:14中所示的序列。在一些实施方案中,所述CD3 ζ 信号传导结构域包含SEQ ID NO:15中所示的序列。

[0074] 在一些任何实施方案中,所述一个或多个同源臂包含5'同源臂和3'同源臂。在一些任何实施方案中,所述多核苷酸包含结构[5'同源臂]-[编码嵌合受体或其部分的核酸序列]-[3'同源臂]。

[0075] 在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有为或约50至为或约2000个核苷酸、为或约100至为或约1000个核苷酸、为或约100至为或约750个核苷酸、为或约100至为或约600个核苷酸、为或约100至为或约400个核苷酸、为或约100至为或约

300个核苷酸、为或约100至为或约200个核苷酸、为或约200至为或约1000个核苷酸、为或约200至为或约750个核苷酸、为或约200至为或约600个核苷酸、为或约200至为或约400个核苷酸、为或约200至为或约300个核苷酸、为或约300至为或约1000个核苷酸、为或约300至为或约750个核苷酸、为或约300至为或约600个核苷酸、为或约300至为或约400个核苷酸、为或约400至为或约1000个核苷酸、为或约400至为或约750个核苷酸、为或约400至为或约600个核苷酸、为或约600至为或约1000个核苷酸、为或约600至为或约750个核苷酸或者为或约750至为或约1000个核苷酸的长度。在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有为或约200、300、400、500、600、700或800个核苷酸的长度,或者任何前述值之间的任何值。在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有大于或大于约300个核苷酸的长度。在一些任何实施方案中,所述5'同源臂和所述3'同源臂独立地具有为或约400、500或600个核苷酸的长度,或者任何前述值之间的任何值。

[0076] 在一些任何实施方案中,所述5'同源臂包含SEQ ID NO:80中所示的序列,或者展现与SEQ ID NO:80至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。在一些实施方案中,所述5'同源臂包含SEQ ID NO:80中所示的序列。在一些实施方案中,所述5'同源臂由SEQ ID NO:80中所示的序列组成或基本上由其组成。在一些任何实施方案中,所述3'同源臂包含SEQ ID NO:81中所示的序列,或者展现与SEQ ID NO:81至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。在一些实施方案中,所述3'同源臂包含SEQ ID NO:81中所示的序列。在一些实施方案中,所述3'同源臂由SEQ ID NO:81中所示的序列组成或基本上由其组成。

[0077] 在一些任何实施方案中,能够诱导遗传破坏的所述一种或多种药剂包含与所述靶位点特异性结合或杂交的DNA结合蛋白或DNA结合核酸、包含DNA靶向蛋白和核酸酶的融合蛋白或者RNA指导核酸酶。在一些任何实施方案中,所述一种或多种药剂包含与所述靶位点特异性结合、识别所述靶位点或与所述靶位点杂交的锌指核酸酶(ZFN)、TAL效应核酸酶(TALEN)或和CRISPR-Cas9组合。

[0078] 在一些任何实施方案中,所述一种或多种药剂各自包含具有与至少一个靶位点互补的靶向结构域的引导RNA(gRNA)。在一些任何实施方案中,所述一种或多种药剂是作为包含gRNA和Cas9蛋白的核糖核蛋白(RNP)复合物引入。

[0079] 在一些任何实施方案中,所述RNP是通过电穿孔、粒子枪、磷酸钙转染、细胞压缩或挤压引入。在一些任何实施方案中,所述RNP是通过电穿孔引入。

[0080] 在一些任何实施方案中,所述RNP的浓度是为或约1、2、2.5、5、10、20、25、30、40或50 μ M,或者由任两个前述值限定的范围。在一些任何实施方案中,所述RNP的浓度是为或约25 μ M。

[0081] 在一些任何实施方案中,所述RNP中gRNA与Cas9分子的摩尔比是为或约为或约5:1、4:1、3:1、2:1、1:1、1:2、1:3、1:4或1:5,或者由任两个前述值限定的范围。在一些任何实施方案中,所述RNP中gRNA与Cas9分子的摩尔比是为或约2.6:1。

[0082] 在一些任何实施方案中,所述gRNA具有选自以下的靶向结构域序列: CACCUUCACUCUCAGGAACA (SEQ ID NO:87); GAAUGACACCAUAGAUGAAG (SEQ ID NO:88); UGAAGAGGAUCCAUCAGC (SEQ ID NO:89); 以及UCCAGCAGGUAGCAGAGUUU (SEQ ID NO:90)。在

一些任何实施方案中,所述gRNA具有CACCUUCACUCUCAGGAACA (SEQ ID NO:87)的靶向结构域序列。在一些任何实施方案中,所述gRNA具有UGAAGAGGAUCCAUCAGC (SEQ ID NO:89)的靶向结构域序列。

[0083] 在一些任何实施方案中,所述T细胞是源自受试者的原代T细胞。在一些任何实施方案中,所述受试者是人。在一些任何实施方案中,所述T细胞是CD8+T细胞或其亚型。在一些任何实施方案中,所述T细胞是CD4+T细胞或其亚型。在一些任何实施方案中,所述T细胞源自多潜能或多能细胞。在一些任何实施方案中,所述多潜能或多能细胞是iPSC。在一些任何实施方案中,所述T细胞源自多潜能或多能细胞,其是iPSC。

[0084] 在一些任何实施方案中,所述多核苷酸包含于病毒载体中。在一些任何实施方案中,所述病毒载体是AAV载体。在一些任何实施方案中,所述AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7或AAV8载体。在一些任何实施方案中,所述AAV载体是AAV2或AAV6载体。在一些任何实施方案中,所述病毒载体是逆转录病毒载体。在一些任何实施方案中,慢病毒载体。

[0085] 在一些任何实施方案中,所述多核苷酸是线性多核苷酸。在一些任何实施方案中,所述线性多核苷酸是双链多核苷酸或单链多核苷酸。

[0086] 在一些任何实施方案中,所述一种或多种药剂和所述多核苷酸是同时或按任何顺序依序引入。在一些任何实施方案中,所述多核苷酸是在引入所述一种或多种药剂之后引入。

[0087] 在一些任何实施方案中,所述多核苷酸在引入所述药剂后立即引入,或者在引入所述药剂后约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、6分钟、8分钟、9分钟、10分钟、15分钟、20分钟、30分钟、40分钟、50分钟、60分钟、90分钟、2小时、3小时或4小时内引入。

[0088] 在一些任何实施方案中,在引入所述一种或多种药剂之前,所述方法包括在刺激或激活所述一种或多种免疫细胞的条件下将所述细胞在体外与一种或多种刺激剂一起孵育。在一些任何实施方案中,所述一种或多种刺激剂包含和抗CD3和/或抗CD28抗体,如抗CD3/抗CD28珠。在一些任何实施方案中,珠与细胞的比率为或为约1:1。

[0089] 在一些任何实施方案中,所述方法还包括在引入所述一种或多种药剂之前从所述一种或多种免疫细胞去除所述一种或多种刺激剂。

[0090] 在一些任何实施方案中,所述方法还包括在引入所述一种或多种药剂和/或引入所述多核苷酸之前、期间或之后,将所述细胞与一种或多种重组细胞因子一起孵育。在一些任何实施方案中,所述一种或多种重组细胞因子选自IL-2、IL-7和IL-15。在一些任何实施方案中,所述一种或多种重组细胞因子是以选自以下的浓度添加:为或约10U/mL至为或约200U/mL、如为或约50IU/mL至为或约100U/mL的浓度的IL-2;0.5ng/mL至50ng/mL、如为或约5ng/mL至为或约10ng/mL的浓度的IL-7;和/或0.1ng/mL至20ng/mL、如为或约0.5ng/mL至为或约5ng/mL的浓度的IL-15。

[0091] 在一些任何实施方案中,所述孵育是在引入所述一种或多种药剂和引入所述多核苷酸之后进行长达或大约24小时、36小时、48小时、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或21天,如长达或约7天。

[0092] 在一些任何实施方案中,通过所述方法生成的多个工程化细胞中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞(如T细胞)包含

CD247基因座内至少一个靶位点的遗传破坏。在一些任何实施方案中,通过所述方法生成的多个工程化细胞(如T细胞)中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞表达所述嵌合受体或其抗原结合片段。在一些任何实施方案中,通过所述方法生成的多个工程化细胞中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞表达所述嵌合受体或其抗原结合片段。

[0093] 在一些任何实施方案中,通过所述方法生成的多个工程化细胞(如T细胞)中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞包含CD247基因座内至少一个靶位点的遗传破坏。在一些任何实施方案中,通过所述方法生成的多个工程化细胞(如T细胞)中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞表达所述嵌合受体。在一些任何实施方案中,通过所述方法生成的多个工程化细胞(如T细胞)中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞表达所述嵌合受体,其中所述嵌合受体含有包含CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,并且其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。在一些实施方案中,所述CD3 ζ 信号传导结构域的至少一部分是由所述内源CD247基因座处的基因组序列编码。在一些实施方案中,所述嵌合受体的所述细胞内区域的整个或全部CD3 ζ 信号传导结构域是由所述内源CD247基因座处的基因组序列编码。

[0094] 还提供使用本文所述的任何方法生成的工程化T细胞或多个工程化T细胞。

[0095] 还提供包括本文所述的任何工程化T细胞的组合物。

[0096] 还提供包括多个T细胞的组合物,所述多个T细胞包括本文所述的任何工程化T细胞。在一些任何实施方案中,所述组合物中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的T细胞包含CD247基因座内至少一个靶位点的遗传破坏。在一些任何实施方案中,所述组合物中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的T细胞表达所述嵌合受体。在一些任何实施方案中,所述组合物中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的T细胞表达所述嵌合受体,其中所述嵌合受体含有包含CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,并且其中所述CD3 ζ 信号传导结构域或所述CD3 ζ 信号传导结构域的至少一部分是由所述工程化细胞(如T细胞)的内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列编码。在一些实施方案中,所述CD3 ζ 信号传导结构域的至少一部分是由所述内源CD247基因座处的基因组序列编码。在一些实施方案中,所述嵌合受体的所述细胞内区域的整个或全部CD3 ζ 信号传导结构域是由所述内源CD247基因座处的基因组序列编码。

[0097] 在一些任何实施方案中,所述组合物包含CD4⁺和/或CD8⁺T细胞。在一些任何实施方案中,所述组合物包含CD4⁺和CD8⁺T细胞,并且CD4⁺与CD8⁺T细胞的比率为从或从约1:3至3:1,如1:1。

[0098] 在一些任何实施方案中,表达所述嵌合受体的细胞构成所述组合物中总细胞的或所述组合物中总CD4⁺或CD8⁺细胞的至少30%、40%、50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多。

[0099] 本文还提供治疗方法,所述方法涉及将本文提供的任何实施方案的工程化细胞、

多个工程化细胞或组合物施用至患有疾病或障碍的受试者。

[0100] 本文还提供本文所述的任何工程化细胞、多个工程化细胞或组合物用于治疗疾病或障碍的用途。在所提供的实施方案中,由所述工程化细胞表达的所述嵌合受体指向或靶向与所述疾病或病症相关或在所述疾病或病症的细胞或组织上表达的抗原。

[0101] 本文还提供本文所述的任何工程化细胞、多个工程化细胞或组合物在制造用于治疗疾病或障碍的药物中的用途。在所提供的实施方案中,由所述工程化细胞表达的所述嵌合受体指向或靶向与所述疾病或病症相关或在所述疾病或病症的细胞或组织上表达的抗原。

[0102] 还提供来自本文提供的任何实施方案的任何工程化细胞、多个工程化细胞或组合物,其用于治疗疾病或障碍。在所提供的实施方案中,由所述工程化细胞表达的所述嵌合受体指向或靶向与所述疾病或病症相关或在所述疾病或病症的细胞或组织上表达的抗原。

[0103] 在一些任何实施方案中,所述疾病或障碍是癌症或肿瘤。在一些任何实施方案中,所述癌症或肿瘤是血液恶性肿瘤。在一些任何实施方案中,所述血液恶性肿瘤是淋巴瘤、白血病或浆细胞恶性肿瘤。在一些任何实施方案中,所述癌症是淋巴瘤,并且所述淋巴瘤是伯基特淋巴瘤、非霍奇金淋巴瘤(NHL)、霍奇金淋巴瘤、华氏巨球蛋白血症、滤泡性淋巴瘤、小无裂细胞性淋巴瘤、粘膜相关淋巴组织淋巴瘤(MALT)、边缘区淋巴瘤、脾淋巴瘤、结节单核细胞样B细胞淋巴瘤、免疫母细胞淋巴瘤、大细胞淋巴瘤、弥漫性混合细胞淋巴瘤、肺B细胞血管中心淋巴瘤、小淋巴细胞性淋巴瘤、原发性纵隔B细胞淋巴瘤、淋巴浆细胞性淋巴瘤(LPL)或套细胞淋巴瘤(MCL)。在一些任何实施方案中,所述癌症是白血病,并且所述白血病是慢性淋巴细胞性白血病(CLL)、浆细胞白血病或急性淋巴细胞性白血病(ALL)。在一些任何实施方案中,所述癌症是浆细胞恶性肿瘤,并且所述浆细胞恶性肿瘤是多发性骨髓瘤(MM)。

[0104] 在一些任何实施方案中,所述肿瘤是实体瘤。在一些任何实施方案中,所述实体瘤是非小细胞肺癌(NSCLC)或头颈鳞状细胞癌(HNSCC)。

[0105] 还提供试剂盒。在一些任何实施方案中,所述试剂盒包括能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂;以及本文提供的任何实施方案的多核苷酸。

[0106] 还提供试剂盒,所述试剂盒包括能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂;以及包含编码嵌合受体或其部分的核酸序列的多核苷酸,其中编码所述嵌合受体或其抗原结合片段或链的转基因被靶向以供经由同源定向修复(HDR)整合于所述靶位点处或附近;以及用于进行本文提供的任何实施方案的方法的说明书。

附图说明

[0107] 图1描绘T细胞中如通过流式细胞术所评估的CD3和TCR的表面表达,所述T细胞用含有四种CD247靶向gRNA(gRNA 1、2、3、4)之一的核糖核蛋白(RNP)复合物进行电穿孔,用于通过CRISPR/Cas9介导的基因编辑在内源CD247基因座处引入遗传破坏,或者所述T细胞经历不含gRNA的模拟电穿孔(模拟)作为对照。

[0108] 图2A描绘T细胞中如通过流式细胞术所评估的CD3(使用抗CD3 ϵ 抗体来检测)和抗BCMA嵌合抗原受体(CAR)(使用BCMA-Fc来检测;可溶人BCMA在其C末端与IgG的Fc区域融合)的表面表达,所述T细胞用含有CD247靶向gRNA 3的RNP复合物进行电穿孔并且与含有四种

多核苷酸(多核苷酸A、B、C、D;描述于表E1中)之一的腺相关病毒(AAV)构建体一起孵育,所述多核苷酸含有编码抗BCMA CAR或其部分的转基因序列以及调节和/或多顺反子元件;或者所述T细胞经历模拟电穿孔和转导(模拟)作为对照。图2B描绘如实施例2.B中所述工程化的示例性抗BCMA CAR的表达的变异系数(CV)(细胞群内信号的标准差除以相应群中信号的平均值)和几何平均荧光(gMFI)。

[0109] 图3A显示在使用含有四种多核苷酸(多核苷酸A、B、C、D;描述于表E1中)之一的AAV构建体工程化的CAR表达T细胞与RPMI 8226多发性骨髓瘤细胞(ATCC®CCL-155™;表达低水平的BCMA)以2:1、1:1或1:2的E:T比率共培养后来自细胞裂解活性测定的总裂解百分比,所述多核苷酸含有编码抗BCMA CAR或其部分的转基因序列和调节和/或多顺反子元件。在49小时中测量Nuclight Red (NLR) 标记的活靶细胞的损失,如通过红色荧光信号(使用IncuCyte®活细胞分析系统,Essen Bioscience)所确定。评估模拟电穿孔和转导的细胞(模拟)以及在无CAR+细胞的情况下培养的靶细胞(仅靶标)作为对照。确定裂解百分比并针对CAR+群体归一化。图3B显示在工程化的CAR表达T细胞与K562慢性髓细胞性白血病(CML)细胞(ATCC®CCL-243™;K562-BCMA,表达高水平的BCMA)以2:1、1:1或1:2的E:T比率共培养后来自细胞裂解活性测定的总裂解百分比。图3C-图3E显示在2:1(图3C)、1:1(图3D)和1:2(图3E)E:T比率下,RPMI 8226细胞随时间变化的裂解,如通过红色荧光信号所确定。图3F-图3H显示在2:1(图3F)、1:1(图3G)和1:2(图3H)E:T比率下,K562细胞随时间变化的裂解。

[0110] 图4A-图4C描绘在如实施例3中所述将使用含有四种多核苷酸(多核苷酸A、B、C、D;描述于表E1中)之一的AAV构建体工程化的CAR表达T细胞与RPMI 8226或K562靶细胞以2:1、1:1和1:2E:T的E:T比率一起孵育后,使用多重细胞因子免疫测定确定的干扰素- γ (IFN- γ ;图4A)、白介素-2 (IL-2;图4B)和肿瘤坏死因子 α (TNF- α ;图4C)的水平。评估模拟电穿孔和转导的细胞(模拟)以及在无CAR+细胞的情况下培养的靶细胞(仅靶标)作为对照。

[0111] 图5描绘T细胞中如通过流式细胞术所评估的CD3的表面表达,所述T细胞用含有CD247靶向gRNA 1或gRNA 3(各自具有Alt-R修饰, IDT Technologies;科勒尔维尔,爱荷华州)的核糖核蛋白(RNP)复合物以约2.6:1的gRNA与Cas9蛋白的比率和25 μ M的浓度进行电穿孔。

[0112] 图6A-图6B描绘来自代表性供体(供体1)的T细胞中如通过流式细胞术所评估的CD3(使用抗CD3 ϵ 抗体来检测)和抗BCMA嵌合抗原受体(CAR)(使用BCMA-Fc来检测;可溶人BCMA在其C末端与IgG的Fc区域融合)的表面表达,所述T细胞用含有CD247靶向gRNA 3的RNP复合物进行电穿孔并且与含有四种多核苷酸(多核苷酸A、B、C、D;描述于表E1中)之一的腺相关病毒(AAV)构建体一起孵育;或者所述T细胞通过慢病毒递送被工程化以表达抗BCMA CAR(慢病毒;参见图6B);所述T细胞经历模拟电穿孔和转导(模拟),或者所述T细胞仅经历模拟转导和使用CD247靶向RNP的电穿孔(仅KO)作为对照。图6C显示每组中抗BCMA CAR表达的直方图。

[0113] 图7A-图7B显示在使用含有四种多核苷酸(多核苷酸A、B、C、D;描述于表E1中,参见图7A)之一的AAV构建体工程化的CAR表达T细胞与MM.1S(ATCC®CRL-2974™)人B淋巴母细胞靶细胞以2:1或1:2的E:T比率共培养后来自细胞裂解活性测定的总裂解百分比,所述多核苷酸含有编码抗BCMA CAR的转基因序列。还评估通过慢病毒递送工程化以表达抗BCMA CAR的T细胞(慢病毒;参见图7B)和仅经历模拟转导和使用CD247靶向RNP的电穿孔的T细胞

(KO) 作为对照。将裂解%值从一式三份样品取平均值并在三个供体之间归一化。

[0114] 图8A-图8C描绘在如实施例4中所述将使用含有四种多核苷酸(多核苷酸A、B、C、D; 描述于表E1中)之一的AAV构建体工程化的CAR表达T细胞与MM.1S靶细胞以2:1和1:2的E:T比率一起孵育后,干扰素- γ (IFN- γ ; 图8A)、白介素-2 (IL-2; 图8B) 和肿瘤坏死因子 α (TNF- α ; 图8C)的水平。还评估通过慢病毒递送工程化以表达抗BCMA CAR的T细胞(LV)、仅经历模拟转导和使用CD247靶向RNP的电穿孔的T细胞(KO)以及模拟电穿孔和转导的细胞(模拟)作为对照。

具体实施方式

[0115] 本文提供基因工程化的细胞,如T细胞,其具有修饰的CD247基因座,所述基因座包括编码嵌合或重组受体(如嵌合抗原受体(CAR))或其部分的一个或多个转基因序列(下文也可互换地称为“供体”序列,例如,对于所述T细胞为外源或异源的序列)。在一些方面,所述细胞被工程化以表达嵌合受体,所述嵌合受体含有CD3zeta (CD3 ζ)链或其片段,通常存在于所述嵌合受体的C末端。在一些实施方案中,所述CD3 ζ 链或片段的至少一部分是由所述工程化细胞(如T细胞)的所述内源CD247基因座(编码CD3 ζ 的基因组基因座)处的基因组序列或其部分序列编码。在一些方面,进行例如通过同源定向修复(HDR)将转基因序列整合至内源CD247基因座中使得编码所述嵌合受体的一部分的核酸序列与内源CD247基因座的开放阅读框或其部分序列(如开放阅读框的外显子)融合(例如,框内融合)。

[0116] 还提供用于产生基因工程化的细胞的方法,所述细胞含有修饰的CD247基因座,所述基因座表达嵌合或重组受体或其部分。所提供的实施方案涉及将编码嵌合受体(例如,CAR)或其部分的转基因序列特异性靶向至内源CD247基因座。在一些情况下,所提供的实施方案涉及例如使用基因编辑方法诱导靶向遗传破坏(例如,生成DNA断裂)以及HDR,用于将编码嵌合受体的转基因序列靶向整合于内源CD247基因座处。还提供相关的细胞组合物、核酸和试剂盒,用于生成本文提供的工程化细胞和/或本文提供的方法。

[0117] 在一些实施方案中,编码所述嵌合或重组受体(例如,CAR)的一部分的转基因序列含有编码所述嵌合受体的一个或多个结构域或区域(例如,细胞外区域、跨膜结构域和细胞内区域)的核苷酸序列。在一些方面,所述细胞外区域含有提供针对期望抗原(例如,肿瘤抗原)或配体的特异性的结合结构域(例如抗原结合结构域或配体结合结构域),和/或用于将细胞外结合结构域与跨膜结构域和细胞内区域连接的间隔子。在一些方面,由转基因序列编码的细胞内区域包含一个或多个共刺激结构域和/或其他结构域。在一些实施方案中,由转基因序列(即,对于细胞为外源的引入序列)编码的细胞内区域包含小于全长的CD3 ζ 链,或者不包含编码CD3 ζ 链的序列。在将转基因序列整合至内源CD247基因座中之后,所得修饰的CD247基因座编码嵌合受体,所述嵌合受体由以下的融合物编码:通过HDR靶向的转基因序列;以及内源CD247基因座的开放阅读框或其部分序列。所编码的嵌合受体含有细胞内区域,所述细胞内区域包含CD3 ζ 链或其片段,例如,能够在T细胞中介导、激活或刺激初级胞质或细胞内信号的功能性CD3 ζ 链或其片段。所得基因工程化的细胞或细胞组合物可以用于过继细胞疗法方法中。

[0118] 基于T细胞的疗法,如过继T细胞疗法(包括涉及施用如下工程化细胞的那些疗法,所述工程化细胞表达对目的疾病或障碍具有特异性的重组、工程化或嵌合受体(如嵌合抗

原受体 (CAR) 或其他重组、工程化或嵌合受体) 可以有效治疗癌症以及其他疾病和障碍。在某些情况下, 用于生成用于过继细胞疗法的工程化细胞的其他方法可能并非总是完全令人满意的。在一些情况下, 最佳功效可以取决于所施用的细胞表达所述嵌合受体的能力, 包括具有所述受体在细胞 (如免疫细胞群和/或治疗性细胞组合物中的细胞) 之间的均匀、同质性和/或一致的表达, 以及所述嵌合受体识别并结合至受试者、肿瘤及其环境内的靶标 (例如, 靶抗原) 的能力。

[0119] 在一些情形中, 用于将嵌合受体 (如CAR) 引入细胞中的可用方法包括编码所述嵌合受体的序列的随机整合, 如通过病毒转导。在某些方面, 此类方法并非完全令人满意。在一些方面, 随机整合可能导致所述细胞中一个或多个随机遗传基因座可能的插入诱变和/或遗传破坏, 所述基因座包括对于细胞功能和活性可能重要的那些。在一些方面, 嵌合受体的表达效率在使用当前可用方法工程化的某些细胞或某些细胞群中是有限的。在一些情形中, 嵌合受体仅在细胞群中的某些细胞中表达, 并且所述嵌合受体的表达水平在所述群体中的细胞之间广泛变化。在特定方面, 嵌合受体的表达水平可能难以预测、控制和/或调节。在一些情形中, 由于将核酸序列整合至基因组中不期望的位置中, 例如, 整合至必需基因中或者在调节细胞活性方面至关重要的基因中, 将编码受体的转基因半随机或随机整合至细胞的基因组中可能在一些情形中导致不良和/或不想要的影响。

[0120] 在一些情形中, 随机整合可能导致编码重组或嵌合受体的序列的可变整合, 从而可能导致不一致的表达、所述核酸的可变拷贝数和/或受体表达在细胞组合物 (如治疗性细胞组合物) 的细胞内的变异性。在一些情形中, 编码受体的核酸序列的随机整合可能导致所述核酸序列的多变、异质性、不均匀和/或次优的表达或抗原结合、致癌的转化和转录沉默, 这取决于整合位点和/或核酸序列拷贝数。在一些方面, 在细胞群中异质性和不均匀的表达可能导致重组或嵌合受体的表达和/或抗原结合的不一致或不稳定, 工程化细胞的功能的不可预测性或功能降低, 和/或不均匀的药物产品, 从而降低所述工程化细胞的功效。在一些方面, 特定随机整合载体 (如某些慢病毒载体) 的使用需要确认, 工程化细胞不含有复制能力的病毒, 如通过有复制能力的慢病毒 (RCL) 测定的表现来确认。需要改进的策略来实现重组或嵌合受体的一致表达水平和功能, 同时使核酸的随机整合和/或在群体中的异质性表达降至最低。

[0121] 在一些方面, 用于递送编码嵌合受体的核酸序列的特定多核苷酸或载体中有效载荷 (如要插入的转基因序列或异源序列) 的大小可以具有限制性。在一些情形中, 有限的大小可能影响在细胞中引入和表达的表达和/或效率。

[0122] 所提供的实施方案涉及将细胞工程化以具有要通过同源定向修复 (HDR) 整合至细胞 (例如, T细胞) 的内源CD247基因座中的编码嵌合受体的核酸。在一些方面, HDR可以介导转基因序列 (如编码重组受体或嵌合受体或其部分、链或片段的转基因序列) 在遗传破坏靶位点 (如内源CD247基因座) 处或附近的位点特异性整合。在一些实施方案中, 遗传破坏 (例如, 在内源CD247基因座的靶位点处) 以及含有一个或多个同源臂 (例如, 含有对所述遗传破坏的周围序列为同源的核酸序列) 的多核苷酸 (例如, 模板多核苷酸) 的存在可以诱导或引导HDR, 其中同源序列用作DNA修复的模板。基于遗传破坏周围的内源基因序列与多核苷酸 (例如, 模板多核苷酸) 中包括的同源臂之间的同源性, 细胞DNA修复机构可以使用所述多核苷酸 (例如, 模板多核苷酸) 来修复遗传破坏靶位点处的DNA断裂并重新合成遗传信息, 从而

将同源臂之间的序列(如编码嵌合受体或其部分的转基因序列)有效插入或整合至遗传破坏靶位点处或附近。所提供的实施方案可以生成含有编码嵌合受体或其部分的修饰的CD247基因座的细胞,其中通过HDR将编码嵌合受体或其部分的转基因序列整合至所述内源CD247基因座中。

[0123] 在一些方面,所提供的实施方案提供产生工程化细胞方面的优点,其改进地和/或更有效地将编码嵌合或重组受体的核酸靶向至细胞中。在一些情形中,所述方法使可能的半随机或随机整合和/或异质性或多变表达和/或来自未整合核酸序列的不期望的表达降至最低,并且导致嵌合或重组受体的改进的、均匀的、同质性的、一致的、可预测的或稳定的表达,或者具有插入诱变的降低的、低的可能性或无可能性。在一些方面,与产生表达嵌合或重组受体(例如,CAR)的基因工程化的免疫细胞的其他方法相比,所提供的实施方案允许所述嵌合或重组受体更稳定的、更符合生理的、更可控的或更均匀的、一致的或同质性的表达。在一些情形中,所述方法导致生成更一致且更可预测的药物产品,例如含有工程化细胞的细胞组合物,从而可以得到对所治疗患者更安全的疗法。在一些方面,所提供的实施方案也允许在单一目的基因座或多个目的基因座处可预测且一致的整合。在一些实施方案中,所提供的实施方案还可以导致生成如下细胞群,所述细胞群具有整合于所述群体的细胞中的核酸的一致拷贝数(通常为1或2),这在一些方面提供细胞群内的嵌合或重组受体表达和内源受体基因表达的一致性。在一些情形中,所提供的实施方案不涉及使用用于整合的病毒载体,并因此可以降低对工程化细胞不含有复制能力的病毒的确认的需求,从而改进细胞组合物的安全性。

[0124] 由本文提供的工程化细胞中的修饰的CD247基因座编码的嵌合受体可以在内源或外源调节元件的控制下被编码。在一些方面,所提供的实施方案允许嵌合受体在内源CD247调节元件的控制下表达,这在一些情形中可以提供更符合生理的表达水平。在一些方面,所提供的实施方案允许编码嵌合受体的核酸在内源调节或控制元件(例如,顺式调节元件,如启动子或内源CD247基因座的5'和/或3'非翻译区(UTR))的控制下表达。因此,在一些方面,所提供的实施方案允许嵌合受体(例如,CAR)或其部分表达,和/或所述表达以与内源CD3 ζ 链类似的水平来调节。

[0125] 在一些方面,所提供的实施方案可以降低或最小化经由嵌合受体的非抗原依赖性信号传导或活性(也称为“紧张性信号传导”)。在一些情形中,非抗原依赖性信号传导可能源自所表达嵌合受体的过表达或不受控活性,并且可能导致不期望的影响,如表达所述嵌合受体的T细胞增加的分化和/或耗竭。在一些实施方案中,所提供的工程化细胞和细胞组合物可以降低非抗原依赖性信号传导的影响,所述非抗原依赖性信号传导可能源自所表达嵌合受体的过表达或不受控活性。因此,所提供的实施方案可以促进如下工程化细胞的产生,所述工程化细胞展现改进的表达、功能和表达均匀性和/或其他所需特征或特性,并最终展现更高功效。在一些实施方案中,所提供的多核苷酸、转基因和/或载体在递送至免疫细胞中时导致嵌合受体(例如,CAR)的表达,所述嵌合受体可以调节T细胞活性,并且在一些情形中,可以调节T细胞分化或稳态。

[0126] 在一些方面,所提供的实施方案允许嵌合受体在外源或异源调节或控制元件的控制下表达,这在一些方面提供更可控的表达水平。在一些方面,所提供的实施方案允许嵌合受体在不同细胞类型中的靶向且受控的表达,所述细胞类型包括在内源CD247基因座处的

内源启动子可能无活性的细胞,如通常不表达CD3 ζ 链的细胞,例如,非T细胞,如NK细胞、B细胞或某些诱导型多能干细胞(iPSC)来源的细胞。

[0127] 在一些方面,所提供的实施方案可以防止不受控表达或来自随机整合或未整合多核苷酸的表达。在一些实施方案中,所引入的多核苷酸(例如,模板多核苷酸)不含编码全长功能受体的核酸序列。在一些情形中,CD3 ζ 链的一部分并非由引入的多核苷酸编码。在一些方面,从随机整合或未整合多核苷酸转录不会产生功能受体。在一些方面,只有在整合于靶基因座(例如,内源CD247基因座)之后,才可以生成含有所有所需信号传导区域的功能受体。在一些方面,所提供的实施方案可以导致细胞组合物的改进的安全性,例如,通过防止例如随机整合或未整合多核苷酸(如未整合病毒载体序列)的不受控表达来实现。

[0128] 在一些方面,所提供的实施方案也可以导致细胞外部分CD3 ζ 的表达的减少和/或消除(例如,敲除),以降低例如应用于同种异体过继细胞疗法中的所施用细胞的免疫原性。

[0129] 所提供的实施方案还可以减小将重组CAR递送至细胞所需的转基因序列的长度,例如,以允许在同一载体(例如,病毒载体)内有足够的空间来包装另外的元件和/或转基因。在一些方面,通过利用编码CD3 ζ 链的内源基因的开放阅读框序列的一部分或全部来编码CAR的CD3 ζ 链的全部或一部分,所提供的实施方案还允许使用与现有方法相比更小的核酸序列片段进行工程化。在一些方面,与现有方法相比,所提供的实施方案为表达CAR的工程化细胞提供灵活性,因为所述方法利用编码CD3 ζ 的内源基因CD247的开放阅读框序列的一部分或全部来编码嵌合受体的CD3 ζ 或其部分。在一些情形中,这可以减小用于编码嵌合受体或其部分的序列的有效载荷空间并为编码其他组分(如其他转基因序列、同源臂、调节元件)的序列留出空间,因为对编码嵌合受体或其部分的核酸序列的长度要求被降低了。在一些方面,所提供的实施方案可以允许在所引入的多核苷酸中容纳与需要完整长度的嵌合受体(例如,CAR)的常规实施方案相比更大的同源臂,和/或允许容纳编码另外的分子的核酸序列,因为对编码所述嵌合受体(例如,CAR)的一部分的核酸序列的长度要求被降低了。在一些方面,可以使用所提供的实施方案促进或改进核酸序列(例如转基因序列)的生成、递送和/或同源定向修复(HDR)的靶向效率。在其他方面,所提供的实施方案允许容纳编码另外的分子的核酸序列用于在细胞上或细胞中表达。

[0130] 还提供用于工程化、制备和产生工程化细胞的方法,以及用于生成或产生工程化细胞的试剂盒和装置。还提供通过所述方法生成的细胞和细胞组合物。提供含有编码嵌合受体的一部分的核酸序列的多核苷酸(例如,病毒载体),以及用于将此类多核苷酸引入细胞中的方法,如通过转导或通过物理递送(如电穿孔)。还提供含有工程化细胞的组合物,以及用于将所述细胞和组合物施用至受试者(如用于过继细胞疗法)的方法、试剂盒和装置。在一些方面,所述细胞从受试者分离,工程化,并施用至同一受试者。在其他方面,所述细胞从一名受试者分离,工程化,并施用至另一受试者。所得基因工程化的细胞或细胞组合物可以用于过继细胞疗法方法中。

[0131] 本申请中提及的所有出版物(包括专利文献、科学文章和数据库)出于所有目的通过引用以其整体并入,其并入程度如同将每个单独出版物通过引用单独并入一般。如果本文所述的定义与通过引用并入本文的专利、申请、公开的申请和其他出版物中所述的定义矛盾或以其他方式不一致,则本文所述的定义优先于通过引用并入本文的定义。

[0132] 本文使用的章节标题只是出于组织的目的,而不应解释为限制所描述的主题。

I. 用于通过同源定向修复生成表达嵌合受体的细胞的方法

[0133] 本文提供生成或产生包含修饰的CD247基因座的基因工程化的细胞的方法,其中所述修饰的CD247基因座包括编码嵌合或重组受体(如嵌合抗原受体(CAR))的核酸序列。在一些方面,所述基因工程化的细胞中的所述修饰的CD247基因座包含编码嵌合受体或嵌合受体的一部分的转基因序列,所述转基因序列整合至通常编码CD3zeta(CD3 ζ)链的内源CD247基因座中。在一些实施方案中,所述方法涉及诱导靶向遗传破坏和使用含有编码嵌合或重组受体或嵌合受体的一部分的转基因的多核苷酸(例如,也称为“模板多核苷酸”)进行同源依赖修复(HDR),从而将所述转基因靶向整合于CD247基因座处。还提供通过所述方法生成的细胞和细胞组合物。在一些实施方案中,还提供含有细胞群的组合物,所述细胞群已经工程化以表达嵌合受体(例如,CAR),使得所述细胞群展现所述嵌合受体的更加改进的、均匀的、同质性的和/或稳定的表达和/或抗原结合,包括通过任何所提供方法产生的基因工程化的免疫细胞,以及用于所述方法中的多核苷酸(例如,模板多核苷酸)和试剂盒。

[0134] 在一些方面,所表达的嵌合受体包含细胞内区域,所述细胞内区域含有CD3zeta(CD3 ζ)链或其片段,如CD3 ζ 的信号传导区域或信号传导结构域。在一些实施方案中,所编码的CD3 ζ 链或其片段是功能性CD3 ζ 链或其片段,如胞质信号传导结构域或区域。在一些实施方案中,所述CD3 ζ 链或其片段位于所述受体的C末端。在一些方面,在将编码所述嵌合受体的一部分的转基因序列整合至CD247基因座中之后,CD3 ζ 链的至少一部分由基因组中CD247基因座的开放阅读框或其部分序列编码。在一些方面,所述嵌合受体是由与内源CD247基因座的开放阅读框或其部分序列融合的外源核酸序列编码。

[0135] 在一些实施方案中,所述方法采用HDR将转基因序列靶向整合至CD247基因座中。在一些情形中,所述方法涉及通过基因编辑技术将一个或多个靶向遗传破坏(例如,DNA断裂)引入内源CD247基因座,与通过HDR靶向整合编码嵌合受体或所述嵌合受体的一部分的转基因序列组合。在一些实施方案中,HDR步骤使得在靶基因组位置处的DNA中需要破坏或断裂(例如,双链断裂)。在一些实施方案中,所述DNA断裂是通过采用基因编辑方法(例如,靶向核酸酶)来诱导。

[0136] 在一些方面,所提供的方法涉及将能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂引入T细胞中;以及将包含转基因和一个或多个同源臂的多核苷酸(例如,模板多核苷酸)引入T细胞中。在一些方面,所述转基因含有编码嵌合受体或其部分的核苷酸序列。在一些实施方案中,所述核酸序列(如转基因)被靶向以供经由同源定向修复(HDR)整合于CD247基因座内。在一些方面,所提供的方法涉及将包含编码嵌合受体或其部分的转基因序列的多核苷酸引入在CD247基因座内具有遗传破坏的T细胞中,其中所述遗传破坏是通过能够诱导CD247基因座内一个或多个靶位点的遗传破坏的一种或多种药剂来诱导,并且其中所述核酸序列(如转基因)被靶向以供通过HDR整合于CD247基因座内。

[0137] 在一些方面,实施方案涉及使用基因编辑方法和/或靶向核酸酶生成靶向基因组破坏,如靶向DNA断裂,之后基于一个或多个多核苷酸(例如,一个或多个模板多核苷酸)进行HDR,所述多核苷酸含有与内源CD247基因座处的序列同源的同源性序列,所述同源性序列与编码嵌合受体的一部分的转基因序列以及(在一些实施方案中)编码其他分子的核酸序列连接,以特异性靶向所述转基因序列并将其整合于所述DNA断裂处或附近。因此,在一些方面,所述方法涉及以下步骤:诱导靶向遗传破坏(例如,通过基因编辑)并将包含转基因

序列的多核苷酸(例如,模板多核苷酸)引入细胞中(例如,通过HDR)。

[0138] 在一些实施方案中,靶向遗传破坏以及通过HDR进行的转基因序列的靶向整合发生在编码CD3zeta (CD3 ζ)链的内源CD247基因座的一个或多个靶位点处。在一些方面,所述靶向整合发生在内源CD247基因座的开放阅读框序列内。在一些方面,转基因序列的靶向整合产生转基因的编码部分与内源CD247基因座的开放阅读框的一个或多个外显子的框内融合物,例如,与整合位点处的相邻外显子框内融合物。

[0139] 在一些实施方案中,在引入能够诱导一个或多个靶向遗传破坏的一种或多种药剂之前、同时或之后,将多核苷酸(例如,模板多核苷酸)引入工程化细胞中。在一个或多个靶向遗传破坏(例如,DNA断裂)的存在下,多核苷酸可以用作DNA修复模板,以基于遗传破坏周围的内源基因序列与模板多核苷酸中包括的一个或多个同源臂(如5'和/或3'同源臂)之间的同源性,在靶向遗传破坏的位点处或附近通过HDR有效拷贝和/或整合转基因。

[0140] 在一些方面,所述两个步骤可以依序进行。在一些实施方案中,基因编辑和HDR步骤是同时和/或在一个实验反应中进行。在一些实施方案中,基因编辑和HDR步骤是在一个或多个连续实验反应中连续或依序进行。在一些实施方案中,基因编辑和HDR步骤是在单独的实验反应中同时或在不同时间进行。

[0141] 免疫细胞可以包括含有T细胞的细胞群。此类细胞可以是已经从受试者获得的细胞,如从以下获得:外周血单个核细胞(PBMC)样品、未分级的T细胞样品、淋巴细胞样品、白细胞样品、单采术产物或白细胞单采术产物。在一些实施方案中,免疫细胞(如T细胞)是原代细胞,如原代T细胞。在一些实施方案中,可以使用阳性或阴性选择和富集方法分离或选择T细胞以在群体中富集T细胞。在一些实施方案中,所述群体含有CD4⁺、CD8⁺或CD4⁺和CD8⁺T细胞。在一些实施方案中,引入多核苷酸(例如,模板多核苷酸)的步骤和引入药剂(例如Cas9/gRNA RNP)的步骤可以同时或以任何顺序依序进行。在一些实施方案中,与引入能够诱导遗传破坏的一种或多种药剂(例如Cas9/gRNA RNP)同时引入多核苷酸。在特定实施方案中,在通过引入一种或多种药剂(例如Cas9/gRNA RNP)的步骤诱导遗传破坏之后,将多核苷酸模板引入免疫细胞中。在一些实施方案中,在引入多核苷酸模板和一种或多种药剂(例如Cas9/gRNA RNP)之前、期间和/或之后,在刺激细胞扩增和/或增殖的条件下培养或孵育所述细胞。

[0142] 在所提供的方法的特定实施方案中,模板多核苷酸的引入是在引入能够诱导遗传破坏的所述一种或多种药剂之后进行。可以如所述采用用于引入一种或多种药剂的任何方法,这取决于用于诱导遗传破坏的一种或多种特定药剂。在一些方面,所述破坏是通过基因编辑来进行,如使用对所破坏的CD247基因座具有特异性的RNA指导核酸酶,如成簇的规律间隔的短回文核酸(CRISPR)-Cas系统,如CRISPR-Cas9系统。在一些方面,所述破坏是使用对CD247基因座具有特异性的CRISPR-Cas9系统来进行。在一些实施方案中,将含有Cas9和指导RNA(gRNA)的药剂引入细胞中,所述指导RNA含有靶向CD247基因座的区域的靶向结构域。在一些实施方案中,所述药剂是或包含Cas9与gRNA的核糖核蛋白(RNP)复合物,所述gRNA含有靶向CD247的靶向结构域(Cas9/gRNA RNP)。在一些实施方案中,所述引入包括使药剂或其部分与细胞在体外接触,其可以包括将所述细胞和药剂培育或孵育长达24、36或48小时或者3、4、5、6、7或8天。在一些实施方案中,所述引入还可以包括实现将药剂递送至细胞中。在各个实施方案中,根据本公开文本的方法、组合物和细胞利用例如通过电穿孔将

Cas9和gRNA的核糖核蛋白 (RNP) 复合物直接递送至细胞。在一些实施方案中,所述RNP复合物包括已经修饰以包括3'多聚A尾和5'抗反向帽类似物 (ARCA) 帽的gRNA。在一些情形中,要修饰的细胞的电穿孔包括在细胞的电穿孔之后且在铺板之前,例如在32°C下使细胞冷休克。

[0143] 在所提供的方法的此类方面,在引入例如已经通过电穿孔引入的一种或多种药剂(如Cas9/gRNA RNP)之后,将多核苷酸(例如,模板多核苷酸)引入细胞中。在一些实施方案中,在引入能够诱导遗传破坏的一种或多种药剂之后立即引入多核苷酸(例如,模板多核苷酸)。在一些实施方案中,在引入能够诱导遗传破坏的一种或多种药剂之后在为或约30秒内、在为或约1分钟内、在为或约2分钟内、在为或约3分钟内、在为或约4分钟内、在为或约5分钟内、在为或约6分钟内、在为或约6分钟内、在为或约8分钟内、在为或约9分钟内、在为或约10分钟内、在为或约15分钟内、在为或约20分钟内、在为或约30分钟内、在为或约40分钟内、在为或约50分钟内、在为或约60分钟内、在为或约90分钟内、在为或约2小时内、在为或约3小时内或者在为或约4小时内,将多核苷酸(例如,模板多核苷酸)引入细胞中。在一些实施方案中,在引入一种或多种药剂为或约15分钟与为或约4小时之间的时间,如在为或约15分钟与为或约3小时之间、在为或约15分钟与为或约2小时之间、在为或约15分钟与为或约1小时之间、在为或约15分钟与为或约30分钟之间、在为或约30分钟与为或约4小时之间、在为或约30分钟与为或约3小时之间、在为或约30分钟与为或约2小时之间、在为或约30分钟与为或约1小时之间、在为或约1小时与为或约4小时之间、在为或约1小时与为或约3小时之间、在为或约1小时与为或约2小时之间、在为或约2小时与为或约4小时之间、在为或约2小时与为或约3小时之间或者在为或约3小时与为或约4小时之间,将多核苷酸(例如,模板多核苷酸)引入细胞中。在一些实施方案中,在引入例如已经通过电穿孔引入的一种或多种药剂(如Cas9/gRNA RNP)之后为或约2小时,将多核苷酸(例如,模板多核苷酸)引入细胞中。

[0144] 可以如所述采用用于引入多核苷酸(例如,模板多核苷酸)的任何方法,这取决于用于将多核苷酸(例如,模板多核苷酸)递送至细胞的特定方法。示例性方法包括用于转移编码受体的核酸的那些,包括通过病毒(例如,逆转录病毒或慢病毒)、转导、转座子和电穿孔来进行。在特定实施方案中,采用病毒转导方法。在一些实施方案中,可以使用重组感染性病毒颗粒(如例如,源自猿猴病毒40 (SV40)、腺病毒、腺相关病毒 (AAV) 的载体)将多核苷酸转移或引入细胞中。在一些实施方案中,将重组核酸转移至T细胞中使用重组慢病毒载体或逆转录病毒载体(如 γ -逆转录病毒载体)(参见例如,Koste等人(2014) *Gene Therapy* 2014年4月3日.doi:10.1038/gt.2014.25;Carlens等人(2000) *Exp Hematol* 28(10):1137-46;Alonso-Camino等人(2013) *Mol Ther Nucl Acids* 2,e93;Park等人,Trends Biotechnol.2011年11月29日(11):550-557)。在特定实施方案中,所述病毒载体是AAV,如AAV2或AAV6。

[0145] 在一些实施方案中,在使药剂与细胞接触之前、期间或之后和/或在实现递送(例如电穿孔)之前、期间或之后,所提供的方法包括在能够诱导免疫细胞(例如T细胞)增殖、刺激或激活的细胞因子、刺激剂和/或药剂的存在下孵育所述细胞。在一些实施方案中,所述孵育的至少一部分是在刺激剂的存在下进行,所述刺激剂是或包含对CD3具有特异性的抗体、对CD28具有特异性的抗体和/或细胞因子,如抗CD3/抗CD28珠。在一些实施方案中,所述孵育的至少一部分是在细胞因子的存在下进行,所述细胞因子例如重组IL-2、重组IL-7和/

或重组IL-15中的一种或多种。在一些实施方案中,在引入一种或多种药剂(如Cas9/gRNA RNP,例如通过电穿孔)和多核苷酸(例如,模板多核苷酸)之前或之后,所述孵育持续长达8天,如长达24小时、36小时或48小时或者3、4、5、6、7或8天。

[0146] 在一些实施方案中,所述方法包括在引入所述药剂(例如Cas9/gRNA RNP)和所述多核苷酸模板之前,用刺激剂(例如抗CD3/抗CD28抗体)激活或刺激细胞。在一些实施方案中,在例如通过电穿孔引入一种或多种药剂(如Cas9/gRNA RNP)之前,在刺激剂(例如抗CD3/抗CD28)的存在下进行的孵育持续6小时至96小时,如24至48小时或24至36小时。在一些实施方案中,与刺激剂一起孵育还可以包括细胞因子的存在,所述细胞因子如重组IL-2、重组IL-7和/或重组IL-15中的一种或多种。在一些实施方案中,所述孵育是在重组细胞因子的存在下进行,所述重组细胞因子如IL-2(例如1U/mL至500U/mL,如10U/mL至200U/mL,例如至少或约50U/mL或100U/mL)、IL-7(例如0.5ng/mL至50ng/mL,如1ng/mL至20ng/mL,例如,至少或约5ng/mL或10ng/mL)或IL-15(例如0.1ng/mL至50ng/mL,如0.5ng/mL至25ng/mL,例如,至少或约1ng/mL或5ng/mL)。在一些实施方案中,在将能够诱导遗传破坏的一种或多种药剂Cas9/gRNA RNP和/或多核苷酸模板引入或递送至细胞中之前,从细胞洗涤或去除一种或多种刺激剂(例如抗CD3/抗CD28抗体)。在一些实施方案中,在引入一种或多种药剂之前,例如通过去除任何刺激或激活剂使细胞静息。在一些实施方案中,在引入一种或多种药剂之前,不去除刺激或激活剂和/或细胞因子。

[0147] 在一些实施方案中,在引入一种或多种药剂(例如Cas9/gRNA)和/或多核苷酸模板之后,在重组细胞因子(如重组IL-2、重组IL-7和/或重组IL-15中的一种或多种)的存在下孵育、培育或培养细胞。在一些实施方案中,所述孵育是在重组细胞因子的存在下进行,所述重组细胞因子如IL-2(例如1U/mL至500U/mL,如10U/mL至200U/mL,例如至少或约50U/mL或100U/mL)、IL-7(例如0.5ng/mL至50ng/mL,如1ng/mL至20ng/mL,例如,至少或约5ng/mL或10ng/mL)或IL-15(例如0.1ng/mL至50ng/mL,如0.5ng/mL至25ng/mL,例如,至少或约1ng/mL或5ng/mL)。在诱导细胞增殖或扩增的条件下孵育或培育细胞。在一些实施方案中,可以孵育或培育细胞直至实现用于收获的细胞阈值数量,例如治疗有效剂量。

[0148] 在一些实施方案中,在过程的任一部分或整个过程期间进行的孵育可以在以下温度下进行:30°C±2°C至39°C±2°C,如至少或约至少30°C±2°C、32°C±2°C、34°C±2°C或37°C±2°C。在一些实施方案中,所述孵育的至少一部分是在30°C±2°C下进行,并且所述孵育的至少一部分是在37°C±2°C下进行。

[0149] 在一些实施方案中,在靶向整合后,在修饰的CD247基因座处存在的核酸序列包含通过HDR靶向的转基因(例如,如本文所述的嵌合受体(如CAR)的一部分)与内源CD247基因座的开放阅读框或其部分序列的融合物。在一些方面,在修饰的CD247基因座处存在的核酸序列包含整合于内源CD247基因座处的转基因(例如,如本文所述的嵌合受体(如CAR)的一部分),所述基因座包含编码CD3ζ链的开放阅读框。在一些方面,在靶向整合或融合(例如,框内融合)后,转基因的外源序列的一部分和内源CD247基因座处的开放阅读框的一部分一起编码含有CD3ζ信号传导结构域或其片段的嵌合受体(例如CAR)。因此,所提供的实施方案利用内源CD247基因座的开放阅读框序列的一部分或全部来编码嵌合受体的CD3ζ信号传导结构域或其部分。在一些实施方案中,在转基因序列的靶向框内整合后,修饰的CD247基因座含有编码含有CD3ζ信号传导结构域的整体、完整或全长嵌合受体(例如CAR)的序列。

[0150] 用于在内源CD247基因座处进行遗传破坏和/或用于进行HDR以将转基因序列(如嵌合受体的一部分,例如CAR的一部分)靶向整合至CD247基因座中的示例性方法描述于以下小节中。

A. 遗传破坏

[0151] 在一些实施方案中,在内源CD247基因座处诱导一个或多个靶向遗传破坏。在一些实施方案中,在内源CD247基因座处或附近的一个或多个靶位点处诱导一个或多个靶向遗传破坏。在一些实施方案中,在内源CD247基因座的内含子中诱导靶向遗传破坏。在一些实施方案中,在内源CD247基因座的外显子中诱导靶向遗传破坏。在一些方面,一个或多个靶向遗传破坏和含有编码嵌合受体或其部分的转基因序列的多核苷酸(例如,模板多核苷酸)的存在可以导致将所述转基因序列靶向整合于内源CD247基因座的一个或多个遗传破坏(例如,靶位点)处或附近。

[0152] 在一些实施方案中,遗传破坏在基因组中的一个或多个靶位点处导致DNA断裂,如双链断裂(DSB)或切割,或者切口,如单链断裂(SSB)。在一些实施方案中,在遗传破坏(例如,DNA断裂或切口)位点,细胞DNA修复机制的作用可以导致敲除、插入、错义或移码突变(如双等位基因移码突变)、全部或部分基因的缺失;或者,在修复模板(例如,模板多核苷酸)的存在下,可以基于所述修复模板改变DNA序列,如整合或插入所述模板中所含核酸序列(如编码重组受体的全部或一部分的转基因)。在一些实施方案中,遗传破坏可以靶向至基因或其部分的一个或多个外显子。在一些实施方案中,遗传破坏可以靶向于外源序列(例如,编码嵌合受体的转基因序列)的靶向整合的期望位点附近。

[0153] 在一些实施方案中,使用与至少一个靶位点之一附近的区域的序列特异性结合或杂交的DNA结合蛋白或DNA结合核酸进行靶向破坏。在一些实施方案中,可以引入模板多核苷酸(例如,包括核酸序列(如编码嵌合受体的一部分的转基因)和同源性序列的模板多核苷酸)用于通过HDR将嵌合受体编码序列靶向整合于遗传破坏位点处或附近,如本文例如在章节I.B中所述。

[0154] 在一些实施方案中,遗传破坏是通过引入能够诱导遗传破坏的一种或多种药剂来进行。在一些实施方案中,此类药剂包含与基因特异性结合或杂交的DNA结合蛋白或DNA结合核酸。在一些实施方案中,所述药剂包含多种组分,如包含DNA靶向蛋白和核酸酶的融合蛋白或RNA指导核酸酶。在一些实施方案中,所述药剂可以靶向一个或多个靶位点或靶位置。在一些方面,可以在靶位点的每一侧上生成一对单链断裂(例如,切口)。

[0155] 在所提供的实施方案中,术语“引入”涵盖在体外或在体内将核酸和/或蛋白质(如DNA)引入细胞中的多种方法,此类方法包括转化、转导、转染(例如电穿孔)和感染。载体可用于将编码分子的DNA引入细胞中。可能的载体包括质粒载体和病毒载体。病毒载体可逆转录病毒载体、慢病毒载体或其他载体,如腺病毒载体或腺相关载体。也可以使用方法(如电穿孔)将蛋白质或核糖核蛋白(RNP)(例如,含有与靶向gRNA复合的Cas9蛋白)引入或递送至目的细胞。

[0156] 在一些实施方案中,遗传破坏发生于靶位点(也称为“靶位置(target position)”、“靶DNA序列”或“靶位置(target location)”),例如,发生于内源CD247基因座处。在一些实施方案中,靶位点包括在靶DNA(例如,基因组DNA)上的位点,所述位点由能够诱导遗传破坏的所述一种或多种药剂(例如,与指定靶位点的gRNA复合的Cas9分子)修饰。

例如,靶位点可以包括在内源CD247基因座处的DNA中的位置,其中发生切割或DNA断裂。在一些方面,通过HDR整合核酸序列(如编码重组受体或其部分的转基因)可以发生在靶位点或靶序列处或附近。在一些实施方案中,靶位点可以是其中添加一个或多个核苷酸的DNA上的两个核苷酸(例如,相邻核苷酸)之间的位点。靶位点可以包含由模板多核苷酸改变的一个或多个核苷酸。在一些实施方案中,靶位点位于靶序列(例如,与gRNA结合的序列)内。在一些实施方案中,靶位点位于靶序列的上游或下游。

1. 在内源CD247基因座处的靶位点

[0157] 在一些实施方案中,遗传破坏和/或通过同源定向修复(HDR)整合编码嵌合受体的一部分的转基因靶向于内源或基因组基因座处,所述基因座编码T细胞表达糖蛋白CD3- ζ 链(也称为CD3 ζ ;CD3 ζ ;T细胞受体T3 ζ 链;CD3Z;T3Z;TCRZ;分化群247;CD247;IMD25)。在人体中,CD3 ζ 由分化群247(CD247)基因编码。在一些实施方案中,遗传破坏以及通过同源定向修复(HDR)整合编码嵌合受体的一部分的转基因靶向于人CD247基因座处。在一些方面,遗传破坏靶向于含有编码CD3 ζ 的开放阅读框的CD247基因座内的靶位点处,使得转基因序列的靶向整合、融合或插入发生于CD247基因座的遗传破坏位点处或附近。在一些方面,遗传破坏靶向于编码CD3 ζ 的开放阅读框的外显子处或附近。在一些方面,遗传破坏靶向于编码CD3 ζ 的开放阅读框的内含子处或附近。

[0158] CD3 ζ 是参与适应性免疫应答的T细胞表面上存在的TCR-CD3复合物的一部分。CD3 ζ 与T细胞受体(TCR) α/β (TCR $\alpha\beta$)或TCR γ/δ (TCR $\gamma\delta$)异二聚体、CD3- γ (CD3 γ)、CD3- δ (CD3 δ)和CD3- ϵ (CD3 ϵ)一起形成TCR-CD3复合物。CD3 ζ 在其细胞内或胞质结构域中含有免疫受体酪氨酸激活基序(ITAM)。CD3 ζ 链可以通过例如经由ITAM刺激或激活初级胞质或细胞内信号传导来将抗原识别与细胞内信号转导途径偶联。在TCR与其配体(例如,在MHC分子情况下为肽;MHC-肽复合物)接合后,ITAM基序可以被激酶(包括Src家族蛋白酪氨酸激酶LCK和FYN)磷酸化,从而导致刺激下游信号转导途径。在一些方面,CD3 ζ ITAM的磷酸化产生蛋白质激酶ZAP70的停泊位点,从而导致ZAP70的磷酸化和激活。

[0159] 示例性人CD3 ζ 前体多肽序列示于SEQ ID NO:73(亚型1;成熟多肽包括SEQ ID NO:73的残基22-164;参见Uniprot登录号P20963;NCBI参考序列:NP_932170.1;SEQ ID NO:74中所示的mRNA序列,NCBI参考序列:NM_198053.2)或SEQ ID NO:75(亚型2;成熟多肽包括SEQ ID NO:75的残基22-163;参见NCBI参考序列:NP_000725.1;SEQ ID NO:76中所示的mRNA序列,NCBI参考序列:NM_000734.3)中。示例性成熟CD3 ζ 链含有细胞外区域(包括SEQ ID NO:73或75中所示的人CD3 ζ 链前体序列的氨基酸残基22-30)、跨膜区(包括SEQ ID NO:73或75中所示的人CD3 ζ 链前体序列的氨基酸残基31-51)和细胞内区域(包括SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基52-164或者SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基52-163)。CD3 ζ 链含有三个免疫受体酪氨酸激活基序(ITAM)结构域,其位于SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-128或131-159处或者位于SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-127或130-158处。

[0160] 在人体中,CD247的示例性基因组基因座包含含有8个外显子和7个内含子的开放阅读框。根据人基因组版本GRCh38(UCSC Genome Browser on Human 2013年12月(GRCh38/hg38) Assembly),CD247的示例性mRNA转录物可以跨越对应于反向链上的染色体1:167,

430,640-167,518,610的序列。表1列出示例性人CD247基因座的转录物中开放阅读框的外显子和内含子和非翻译区的坐标。

表1. 示例性人CD247基因座的外显子和内含子的坐标 (GrCh38, 染色体1, 反向链)。

	起始 (GrCh38)	终止 (GrCh38)	长度
5'UTR和外显子1	167,518,610	167,518,408	203
内含子1-2	167,518,407	167,440,768	77,640
外显子2	167,440,767	167,440,664	104
内含子2-3	167,440,663	167,439,401	1,263
外显子3	167,439,400	167,439,344	57
内含子3-4	167,439,343	167,438,651	693
外显子4	167,438,650	167,438,570	81
内含子4-5	167,438,569	167,435,432	3,138
外显子5	167,435,431	167,435,399	33
内含子5-6	167,435,398	167,434,077	1,322
外显子6	167,434,076	167,434,020	57
内含子6-7	167,434,019	167,433,060	960
外显子7	167,433,059	167,433,024	36
内含子7-8	167,433,023	167,431,747	1,277
外显子8和3'UTR	167,431,746	167,430,640	1,107

[0161] 在一些方面,模板多核苷酸内的转基因(例如,外源核酸序列)可以用于指导靶位点和/或同源臂的定位。在一些方面,遗传破坏的靶位点可以用作设计用于HDR的模板多核苷酸和/或同源臂的指导。在一些实施方案中,遗传破坏可以靶向于转基因序列(例如,编码嵌合受体或其部分)的靶向整合的期望位点附近。在一些方面,遗传破坏是基于用于整合的转基因序列内所含编码CD3 ζ 链的序列的量来靶向。在一些方面,靶位点位于内源CD247基因座的开放阅读框的外显子内。在一些方面,靶位点位于CD247基因座的开放阅读框的内含子内。

[0162] 在一些实施方案中,选择遗传破坏的靶位点使得,在转基因序列整合后,由修饰的CD247基因座编码的嵌合受体含有功能性CD3 ζ 链或其片段,使得其能够经由所述CD3 ζ 链或其片段进行信号传导。在一些实施方案中,模板多核苷酸的一个或多个同源臂序列设计为位于遗传破坏的位点周围。在一些方面,靶位点位于内源CD247基因座的外显子内或附近,使得编码嵌合受体的一部分的转基因可以与CD247基因座的编码序列框内整合。

[0163] 在一些实施方案中,选择靶位点使得转基因的靶向整合生成所述转基因与CD247基因座的内源序列的基因融合物,其一起编码功能性CD3 ζ 链。在一些方面,内源序列可以编码功能性CD3 ζ 链,所述功能性CD3 ζ 链是CD3 ζ 链中能够介导、激活或刺激初级胞质或细胞内信号的部分,例如,CD3 ζ 链的胞质结构域,如CD3 ζ 链中包括免疫受体酪氨酸激活基序(ITAM)的部分。在一些方面,靶位点位于编码CD3 ζ 链的细胞内区域(例如,SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基52-164或SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基52-163)的内源开放阅读框序列的起点处或附近;或者位于外显子2或外显子3(例如,GrCh38中的核苷酸167,440,767-167,440,664或核苷酸167,439,400-167,439,344处或

附近的序列,如本文表1中所述)处或附近。在一些方面,靶位点位于编码CD3 ζ 链的ITAM结构域(例如,SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-128或131-159或SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-127或130-158)的内源开放阅读框序列之前或上游。

[0164] 在一些方面,靶位点位于内源CD247基因座的外显子内。在一些方面,靶位点位于内源CD247基因座的内含子内。在一些方面,靶位点位于CD247基因座的调节或控制元件(例如,启动子、5'非翻译区(UTR)或3'UTR)内。在一些实施方案中,靶位点位于本文表1中所述的CD247基因组区域序列内,或者其中所含的CD247基因组区域序列的任何外显子或内含子内。

[0165] 在一些方面,靶位点位于外显子(如对应于早期编码区的外显子)内。在一些实施方案中,靶位点位于对应于早期编码区的外显子(例如,内源CD247基因座的开放阅读框的外显子1、2或3(如本文表1中所述))内或非常靠近所述外显子,或者包括紧接在转录起始位点之后,在外显子1、2或3内,或者在外显子1、2或3的小于500、450、400、350、300、250、200、150、100或50bp内的序列。在一些方面,靶位点位于内源CD247基因座的外显子1处或附近,例如,位于外显子1的小于500、450、400、350、300、250、200、150、100或50bp内。在一些实施方案中,靶位点位于内源CD247基因座的外显子2处或附近,或者位于外显子2的小于500、450、400、350、300、250、200、150、100或50bp内。在一些方面,靶位点位于内源CD247基因座的外显子3处或附近,例如,位于外显子3的小于500、450、400、350、300、250、200、150、100或50bp内。在一些方面,靶位点位于CD247基因座的调节或控制元件(例如,启动子)内。

[0166] 在某些实施方案中,遗传破坏靶向于CD247基因座处、附近或之内。在特定实施方案中,遗传破坏靶向于CD247基因座的开放阅读框(如本文表1中所述)处、附近或之内。在某些实施方案中,遗传破坏靶向于编码CD3 ζ 链的开放阅读框处、附近或之内。在一些实施方案中,遗传破坏靶向于CD247基因座(如本文表1中所述)处、附近或之内,或者具有与CD247基因座(如本文表1中所述)的全部或部分(例如,为或至少500、1,000、1,500、2,000、2,500、3,000、3,500或4,000个连续核苷酸)为或至少70%、75%、80%、85%、90%、95%、97%、98%、99%、99.5%或99.9%序列同一性的序列处、附近或之内。

[0167] 在一些实施方案中,遗传破坏(例如,DNA断裂)靶向于CD247基因座或其开放阅读框的外显子内。在某些实施方案中,遗传破坏位于CD247基因座或其开放阅读框的第一外显子、第二外显子、第三外显子或第四外显子内。在特定实施方案中,遗传破坏位于CD247基因座或其开放阅读框的第一外显子内。在一些实施方案中,遗传破坏位于CD247基因座或其开放阅读框的第一外显子5'端下游的500个碱基对(bp)内。在特定实施方案中,遗传破坏位于外显子1的5'核苷酸与外显子1的3'核苷酸的上游之间。在某些实施方案中,遗传破坏位于CD247基因座或其开放阅读框中第一外显子5'端下游的400bp、350bp、300bp、250bp、200bp、150bp、100bp或50bp内。在特定实施方案中,遗传破坏位于CD247基因座或其开放阅读框中第一外显子5'端下游的1bp与400bp之间、50与300bp之间、100bp与200bp之间或100bp与150bp之间,每个都包含端值。在某些实施方案中,遗传破坏位于CD247基因座或其开放阅读框中第一外显子5'端下游的100bp与150bp之间,包含端值。

2. 遗传破坏的方法

[0168] 在一些方面,用于生成基因工程化的细胞的方法涉及在一个或多个靶位点(例如,

在编码CD3zeta (CD3 ζ)的CD247基因座的一个或多个靶位点)引入遗传破坏。用于生成遗传破坏的方法(包括本文所述的那些)可以涉及使用能够诱导遗传破坏的一种或多种药剂(如工程化系统)在内源或基因组DNA中的靶位点或靶位置诱导遗传破坏、切割和/或双链断裂(DSB)或切口(例如,单链断裂(SSB)),使得通过错误产生过程(error born process)(如非同源末端接合(NHEJ))修复所述断裂或使用修复模板通过HDR修复可以导致将目的序列(例如,编码嵌合受体的一部分的外源核酸序列或转基因)插入靶位点或位置处或附近。还提供能够诱导遗传破坏的一种或多种药剂用于本文提供的方法中。在一些方面,所述一种或多种药剂可以与本文提供的模板核苷酸组合用于同源定向修复(HDR)介导的转基因序列的靶向整合。

[0169] 在一些实施方案中,能够诱导遗传破坏的所述一种或多种药剂包含与基因组中的特定位点或位置(例如,靶位点或靶位置)特异性结合或杂交的DNA结合蛋白或DNA结合核酸。在一些方面,在内源CD247基因座处的靶向遗传破坏(例如,DNA断裂或切割)是使用如呈嵌合或融合蛋白的与基因编辑核酸酶偶联或复合的蛋白质或核酸来实现。在一些实施方案中,能够诱导遗传破坏的一种或多种药剂包含RNA指导核酸酶,或者包含DNA靶向蛋白和核酸酶的融合蛋白。

[0170] 在一些实施方案中,所述药剂包含多种组分,如RNA指导核酸酶或包含DNA靶向蛋白和核酸酶的融合蛋白。在一些实施方案中,靶向遗传破坏是使用与核酸酶(如内切核酸酶)融合的DNA靶向分子来进行,所述DNA靶向分子包括DNA结合蛋白,如一种或多种锌指蛋白(ZFP)或转录激活因子样效应子(TALE)。在一些实施方案中,靶向遗传破坏是使用RNA指导核酸酶来进行,所述RNA指导核酸酶如成簇的规律间隔的短回文核酸(CRISPR)相关核酸酶(Cas)系统(包括Cas和/或Cfp1)。在一些实施方案中,靶向遗传破坏是使用能够诱导遗传破坏的药剂来进行,所述药剂如序列特异性或靶向核酸酶,包括DNA结合靶向核酸酶和基因编辑核酸酶,如锌指核酸酶(ZFN)和转录激活因子样效应核酸酶(TALEN),以及RNA指导核酸酶,如CRISPR相关核酸酶(Cas)系统,所述药剂专门设计为靶向至少一个靶位点、基因序列或其部分。示例性ZFN、TALE和TALEN描述于例如Lloyd等人,Frontiers in Immunology,4(221):1-7(2013)中。

[0171] 锌指蛋白(ZFP)、转录激活因子样效应子(TALE)和CRISPR系统结合结构域可以被“工程化”以与预定核苷酸序列结合,例如通过工程化(改变一个或多个氨基酸)天然存在的ZFP或TALE蛋白的识别螺旋区。工程化的DNA结合蛋白(ZFP或TALE)是非天然存在的蛋白质。设计的合理标准包括应用替代规则和计算机化算法,以处理存储现有ZFP和/或TALE设计和结合数据的信息的数据库中的信息。参见例如,美国专利号6,140,081;6,453,242;和6,534,261;还参见WO 98/53058;WO 98/53059;WO 98/53060;WO 02/016536和WO 03/016496以及美国公开号20110301073。

[0172] 在一些实施方案中,一种或多种药剂特异性靶向CD247基因座处或附近的至少一个靶位点。在一些实施方案中,所述药剂包含与一个或多个靶位点特异性结合、识别一个或多个靶位点或与一个或多个靶位点杂交的ZFN、TALEN或CRISPR/Cas9组合。在一些实施方案中,所述CRISPR/Cas9系统包括工程化的crRNA/tracrRNA(“单一指导RNA”)以指导特异性切割。在一些实施方案中,所述药剂包含基于Argonaute系统的核酸酶(例如,来自嗜热栖热菌(*T. thermophilus*),称为“TtAgo”(Swartz等人,(2014)Nature507(7491):258-261))。使

用本文所述的任何核酸酶系统的靶向切割可以用于使用HDR或NHEJ介导的过程将核酸序列(例如,编码嵌合受体的一部分的转基因序列)插入内源CD247基因座的特定靶位置中。

[0173] 在一些实施方案中,“锌指DNA结合蛋白”(或结合结构域)是通过一个或多个锌指以序列特异性方式结合DNA的蛋白质或较大蛋白质内的结构域,所述锌指是结合结构域内的氨基酸序列区域,其结构通过锌离子的配位而稳定。术语锌指DNA结合蛋白通常缩写为锌指蛋白或ZFP。ZFP包括靶向通常长9-18个核苷酸的特定DNA序列的人工ZFP结构域,其是通过单独指状物的组装产生的。ZFP包括如下那些,其中单一指状物结构域具有大约30个氨基酸的长度,并且含有 α 螺旋,所述 α 螺旋含有经由锌与单一 β 转角两个半胱氨酸配位的两个不变组氨酸残基,且具有两个、三个、四个、五个或六个指状物。通常,ZFP的序列特异性可以通过在锌指识别螺旋上的四个螺旋位置(-1、2、3和6)进行氨基酸取代来改变。因此,例如,ZFP或含有ZFP的分子是非天然存在的,例如被工程化以结合至所选靶位点。

[0174] 在一些情况下,DNA靶向分子是或包含锌指DNA结合结构域,其与DNA切割结构域融合以形成锌指核酸酶(ZFN)。例如,融合蛋白包含来自至少一种IIS型限制性酶的切割结构域(或切割半结构域)和可以被工程化或可以不被工程化的一个或多个锌指结合结构域。在一些情况下,切割结构域来自IIS型限制性内切核酸酶FokI,其通常催化DNA的双链切割,所述切割位于一条链上距其识别位点9个核苷酸处以及另一条链上距其识别位点13个核苷酸处。参见例如,美国专利号5,356,802;5,436,150和5,487,994;Li等人(1992) Proc.Natl.Acad.Sci.USA 89:4275-4279;Li等人(1993) Proc.Natl.Acad.Sci.USA 90:2764-2768;Kim等人(1994a) Proc.Natl.Acad.Sci.USA 91:883-887;Kim等人(1994b) J.Biol.Chem.269:978-982。一些基因特异性工程化锌指可在市场上购得。例如,用于锌指构建的称为CompoZr的平台是可用的,其提供针对数千靶标的特异性靶向锌指。参见例如,Gaj等人,Trends in Biotechnology,2013,31(7),397-405。在一些情况下,使用可商购获得的锌指或者定制设计。

[0175] 在一些实施方案中,可以靶向例如在CD247基因座内的一个或多个靶位点用于通过工程化ZFN进行遗传破坏。靶向内源CD247基因座的示例性ZFN包括例如Rudemiller等人,(2014) Hypertension.63(3):559-64中所述的那些,所述文献的公开内容通过引用以其整体并入。

[0176] 转录激活因子样效应子(TALE)是来自细菌物种黄单胞菌属(Xanthomonas)的蛋白质,其包含多个重复序列,每个重复序列在位置12和13包含双残基(RVD),所述双残基对核酸靶向序列的每个核苷酸碱基具有特异性。具有类似的模块化碱基对碱基(base-per-base)核酸结合特性的结合结构域(MBBBD)也可以源自不同的细菌物种。新型模块化蛋白具有展示比TAL重复序列更高的序列变异性的优点。在一些实施方案中,与不同核苷酸的识别相关的RVD是用于识别C的HD、用于识别T的NG、用于识别A的NI、用于识别G或A的NN、用于识别A、C、G或T的NS、用于识别T的HG、用于识别T的IG、用于识别G的NK、用于识别C的HA、用于识别C的ND、用于识别C的HI、用于识别G的HN、用于识别G的NA、用于识别G或A的SN和用于识别T的YG、用于识别A的TL、用于识别A或G的VT以及用于识别A的SW。在一些实施方案中,关键氨基酸12和13可以突变为其他氨基酸残基,以调节其对核苷酸A、T、C和G的特异性,并且具体而言增强该特异性。

[0177] 在一些实施方案中,“TALE DNA结合结构域”或“TALE”是包含一个或多个TALE重复

结构域/单元的多肽。各自包含重复可变双残基 (RVD) 的重复结构域参与TALE与其同源靶DNA序列的结合。单一“重复单元” (也称为“重复序列”) 通常具有33-35个氨基酸的长度, 并且展现与天然存在的TALE蛋白内的其他TALE重复序列的至少一定序列同源性。可以使用重复单元内的典范或非典范RVD将TALE蛋白设计为与靶位点结合。参见例如, 美国专利号8, 586, 526和9, 458, 205。

[0178] 在一些实施方案中, “TALE-核酸酶” (TALEN) 是一种融合蛋白, 其包含通常源自转录激活因子样效应子 (TALE) 的核酸结合结构域和切割核酸靶序列的核酸酶催化结构域。催化结构域包含核酸酶结构域或具有内切核酸酶活性的结构域, 如例如I-TevI、ColE7、NucA和Fok-I。在特定实施方案中, TALE结构域可以与兆核酸酶 (如例如I-CreI和I-OnuI) 或其功能变体融合。在一些实施方案中, TALEN是单体TALEN。单体TALEN是特异性识别和切割无需二聚化的TALEN, 如WO 2012138927中所述的工程化TAL重复序列与I-TevI的催化结构域的融合物。已经描述TALEN且将其用于基因靶向和基因修饰 (参见例如, Boch等人 (2009) *Science* 326 (5959):1509-12; Moscou和Bogdanove (2009) *Science* 326 (5959):1501; Christian等人 (2010) *Genetics* 186 (2):757-61; Li等人 (2011) *Nucleic Acids Res* 39 (1):359-72)。在一些实施方案中, 可以靶向CD247基因座中的一个或多个位点用于通过工程化TALEN进行遗传破坏。

[0179] 在一些实施方案中, “TtAgo” 是原核Argonaute蛋白, 认为其参与基因沉默。TtAgo源自细菌嗜热栖热菌 (*Thermus thermophilus*)。参见例如, Swarts等人, (2014) *Nature* 507 (7491):258-261; G. Sheng等人, (2013) *Proc. Natl. Acad. Sci. U.S.A.* 111, 652。“TtAgo系统” 是全部所需组分, 包括例如用于通过TtAgo酶进行切割的指导DNA。

[0180] 在一些实施方案中, 在自然界中未发现工程化的锌指蛋白、TALE蛋白或CRISPR/Cas系统, 并且其产生主要源自经验过程, 如噬菌体展示、相互作用陷阱或杂交选择。参见例如, 美国专利号5, 789, 538; 美国专利号5, 925, 523; 美国专利号6, 007, 988; 美国专利号6, 013, 453; 美国专利号6, 200, 759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197和WO 02/099084。

[0181] 锌指和TALE DNA结合结构域可以被工程化以结合至预定的核苷酸序列, 例如通过工程化 (改变一个或多个氨基酸) 天然存在的锌指蛋白的识别螺旋区, 或通过工程化参与DNA结合的氨基酸 (重复可变双残基或RVD区)。因此, 工程化的锌指蛋白或TALE蛋白是非天然存在的蛋白质。用于工程化锌指蛋白和TALE的方法的非限制性例子是设计和选择。所设计的蛋白质是自然界中不存在的蛋白质, 其设计/组成主要源自合理标准。设计的合理标准包括应用替代规则和计算机化算法, 以处理存储现有ZFP或TALE设计 (典范和非典范RVD) 和结合数据的信息的数据库中的信息。参见例如, 美国专利号9, 458, 205; 8, 586, 526; 6, 140, 081; 6, 453, 242; 和6, 534, 261; 还参见WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536和WO 03/016496。

[0182] 已经描述了用于基因组DNA的靶向切割的各种方法和组合物。此类靶向切割事件可以用于例如诱导靶向诱变, 诱导细胞DNA序列的靶向缺失, 并且促进预定的染色体基因座处的靶向重组。参见例如, 美国专利号9, 255, 250; 9, 200, 266; 9, 045, 763; 9, 005, 973; 9, 150, 847; 8, 956, 828; 8, 945, 868; 8, 703, 489; 8, 586, 526; 6, 534, 261; 6, 599, 692; 6, 503, 717; 6, 689, 558; 7, 067, 317; 7, 262, 054; 7, 888, 121; 7, 972, 854; 7, 914, 796; 7, 951, 925; 8,

110,379;8,409,861;美国专利公开案20030232410;20050208489;20050026157;20050064474;20060063231;20080159996;201000218264;20120017290;20110265198;20130137104;20130122591;20130177983;20130196373;20140120622;20150056705;20150335708;20160030477和20160024474,所述文献的公开内容是通过引用以其整体并入。

a. CRISPR/Cas9

[0183] 在一些实施方案中,在编码CD3zeta (CD3 ζ)的内源基因(如人中的CD247)处的靶向遗传破坏(例如,DNA断裂)是使用成簇的规律间隔的短回文重复序列(CRISPR)和CRISPR相关(Cas)蛋白来进行。参见Sander和Joung (2014) Nature Biotechnology, 32 (4) :347-355。

[0184] 通常,“CRISPR系统”统指参与CRISPR相关(“Cas”)基因的表达或引导其活性的转录物和其他元件,包括编码Cas基因的序列、tracr (反式激活CRISPR)序列(例如tracr RNA或活性部分tracr RNA)、tracr配对序列(涵盖“同向重复序列”和在內源CRISPR系统的情况下tracr RNA处理的部分同向重复序列)、指导序列(在內源CRISPR系统的情况下,也称为“间隔子”)和/或来自CRISPR基因座的其他序列和转录物。

[0185] 在一些方面,CRISPR/Cas核酸酶或CRISPR/Cas核酸酶系统包括序列特异性结合至DNA的非编码指导RNA (gRNA)和具有核酸酶功能性的Cas蛋白(例如,Cas9)。

[0186] 还提供能够引入遗传破坏的一种或多种药剂。还提供编码能够诱导遗传破坏的一种或多种药剂的一种或多种组分的多核苷酸(例如,核酸分子)。

(i) 指导RNA (gRNA)

[0187] 在一些实施方案中,能够诱导遗传破坏的所述一种或多种药剂包含以下中的至少一种:具有与CD247基因座处的靶位点互补的靶向结构域的指导RNA (gRNA),或编码所述gRNA的至少一种核酸。

[0188] 在一些方面,“gRNA分子”是指促进gRNA分子/Cas9分子复合物特异性靶向或归巢至靶核酸(如细胞基因组DNA上的基因座)的核酸。gRNA分子可以是单分子的(具有单一RNA分子),有时在本文中称为“嵌合”gRNA;或者模块化的(包含超过一种(通常两种)单独RNA分子)。通常,指导序列(例如,指导RNA)是任何多核苷酸序列,其包含具有与靶多核苷酸序列(如在人中的CD247基因座处)的足够互补性的至少一个序列部分,以与靶位点处的靶序列杂交并引导CRISPR复合物与所述靶序列的序列特异性结合。在一些实施方案中,在形成CRISPR复合物的情况下,“靶序列”是指指导序列被设计为与其具有互补性的序列,其中靶序列与指导RNA的结构域(例如,靶向结构域)之间的杂交促进CRISPR复合物的形成。不一定需要完全互补性,条件是存在足够的互补性以引起杂交并促进CRISPR复合物的形成。通常,选择指导序列以减少指导序列内二级结构的程度。二级结构可以通过任何合适的多核苷酸折叠算法来确定。

[0189] 在一些实施方案中,将对目的靶基因座(例如在人中的CD247基因座处)具有特异性的指导RNA (gRNA)用于RNA指导核酸酶(例如,Cas),以在靶位点或靶位置处诱导DNA断裂。用于设计gRNA和示例性靶向结构域的方法可以包括例如以下文献中所述的那些:过继PCT公开号WO 2015/161276、WO 2017/193107和WO 2017/093969。

[0190] 几种示例性gRNA结构描述于WO 2015/161276中,例如描述于其中的图1A-图1G中,所述结构上指示有结构域。尽管不希望受理论束缚,关于gRNA的活性形式的三维形式或者

链内或链间相互作用,在WO 2015/161276中(例如,在其中的图1A-图1G中)和本文提供的其他描绘中,高互补性区域有时显示为双链体。

[0191] 在一些情形中,gRNA是单分子或嵌合gRNA,其从5'至3'包含:与靶核酸(如来自CD247基因的序列(SEQ ID NO:74中所示的编码序列))互补的靶向结构域;第一互补结构域;连接结构域;第二互补结构域(其与第一互补结构域互补);近端结构域;以及任选地,尾结构域。

[0192] 在其他情形中,gRNA是包含第一链和第二链的模块化gRNA。在这些情形中,第一链优选地从5'至3'包括:靶向结构域(其与靶核酸互补,所述靶核酸如来自CD247基因的序列,即SEQ ID NO:74或76中所示的编码序列)和第一互补结构域。第二链大体上从5'至3'包括:任选地,5'延伸结构域;第二互补结构域;近端结构域;以及任选地,尾结构域。

(a) 靶向结构域

[0193] 靶向结构域包含与靶核酸上的靶序列互补(例如,至少80%、85%、90%、95%、98%或99%互补,例如,完全互补)的核苷酸序列。靶核酸中包含靶序列的链在本文中称为靶核酸的“互补链”。关于选择靶向结构域的指导可以发现于例如以下文献中:Fu Y等人,Nat Biotechnol 2014(doi:10.1038/nbt.2808)和Sternberg SH等人,Nature 2014(doi:10.1038/nature13011)。靶向结构域的放置的例子包括在WO2015/161276中(例如在其中的图1A-图1G中)所述的那些。

[0194] 靶向结构域是RNA分子的一部分,因此将包含碱基尿嘧啶(U),而编码gRNA分子的任何DNA都将包含碱基胸腺嘧啶(T)。尽管不希望受理论束缚,在一些实施方案中,认为靶向结构域与靶序列的互补性有助于gRNA分子/Cas9分子复合物与靶核酸的相互作用的特异性。应理解,在靶向结构域和靶序列对中,靶向结构域中的尿嘧啶碱基将与靶序列中的腺嘌呤碱基配对。在一些实施方案中,靶结构域本身在5'至3'方向上包含任选的二级结构域和核心结构域。在一些实施方案中,核心结构域与靶序列完全互补。在一些实施方案中,靶向结构域的长度为5至50个核苷酸。靶核酸中与靶向结构域互补的链在本文中称为互补链。所述结构域的一些或所有核苷酸可以具有修饰,例如以使其更不易降解、改进生物相容性等。作为非限制性例子,靶结构域的骨架可以用硫代磷酸酯或一种或多种其他修饰来修饰。在一些情况下,靶向结构域的核苷酸可以包含2'修饰(例如2-乙酰化,例如2'甲基化)或一种或多种其他修饰。

[0195] 在各个实施方案中,靶向结构域的长度为16-26个核苷酸(即其长度为16个核苷酸,或者长度为17个核苷酸,或者长度为18、19、20、21、22、23、24、25或26个核苷酸)。

(b) 示例性靶向结构域

[0196] 在一些实施方案中,设计或鉴定是或包含靶向结构域序列的gRNA序列,所述靶向结构域序列靶向特定基因(如CD247基因座)中的靶位点。用于CRISPR基因组编辑的全基因组gRNA数据库是公众可获得的,其含有靶向人基因组或小鼠基因组中基因的组成型外显子的示例性单一指导RNA(sgRNA)序列(参见例如,genescript.com/gRNA-database.html;还参见,Sanjana等人(2014)Nat.Methods,11:783-4)。在一些方面,gRNA序列是或包含具有与非靶位点或位置的最小脱靶结合的序列。

[0197] 在一些实施方案中,靶序列(靶结构域)位于CD247基因座处或附近,如SEQ ID NO:74或76中所示CD247编码序列的任何部分。在一些实施方案中,与靶向结构域互补的靶核酸

定位于目的基因(如CD247)的早期编码区处。早期编码区的靶向可以用于目的基因的遗传破坏(即,消除其表达)。在一些实施方案中,目的基因的早期编码区包括紧接起始密码子(例如,ATG)之后或者在起始密码子的500bp内(例如,小于500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、20bp或10bp)的序列。在特定例子中,靶核酸位于起始密码子的200bp、150bp、100bp、50bp、40bp、30bp、20bp或10bp内。在一些例子中,gRNA的靶向结构域与靶核酸(如CD247基因座中的靶核酸)上的靶序列互补(例如,至少80%、85%、90%、95%、98%或99%互补,例如,完全互补)。

[0198] 在一些实施方案中,gRNA可以靶向CD247基因座处的位点,所述位点在例如编码嵌合受体的转基因序列的靶向整合的期望位点附近。在一些方面,gRNA可以基于用于整合的转基因序列内所含的编码CD3 ζ 链的序列的量来靶向位点。在一些方面,gRNA可以靶向内源CD247基因座的开放阅读框的外显子内的位点。在一些方面,gRNA可以靶向CD247基因座的开放阅读框的内含子内的位点。在一些方面,gRNA可以靶向CD247基因座的调节或控制元件(例如,启动子)内的位点。在一些方面,由gRNA靶向的在CD247基因座处的靶位点可以是本文例如在章节I.A.1中所述的任何靶位点。在一些实施方案中,gRNA可以靶向如下位点,所述位点位于对应于早期编码区的外显子(例如,内源CD247基因座的开放阅读框的外显子1、2或3)内或非常靠近所述外显子,或者包括紧接在转录起始位点之后,在外显子1、2或3内,或者在外显子1、2或3的小于500、450、400、350、300、250、200、150、100或50bp内的序列。在一些实施方案中,gRNA可以靶向如下位点,所述位点位于内源CD247基因座的外显子2处或附近,或者位于外显子2的小于500、450、400、350、300、250、200、150、100或50bp内。

[0199] 用于使用Cas9破坏人的CD247基因座的示例性靶位点序列可以包括SEQ ID NO:59-62和67-72中所示的任何序列。在一些方面,示例性靶位点序列(包括NGG PAM)包括SEQ ID NO:63-66中所示的任何序列。示例性gRNA可以包括可以结合至或靶向SEQ ID NO:59-62和67-72中任一项所示的靶位点序列的核糖核酸。示例性gRNA靶向结构域序列包括:CACCUUCACUCUCAGGAACA (SEQ ID NO:87);GAAUGACACCAUAGAUGAAG (SEQ ID NO:88);UGAAGAGGAUCCAUCAGC (SEQ ID NO:89);UCCAGCAGGUAGCAGAGUUU (SEQ ID NO:90);AGACGCCCCCGCUACCAGC (SEQ ID NO:91);GCUGACUUACGUUAUAGAGC (SEQ ID NO:92);UUUCACCGCGCCAUCUGC (SEQ ID NO:93);UAAUCGGCAACUGUGCCUGC (SEQ ID NO:94);CGGAGGCCUACAGUGAGAUU (SEQ ID NO:95);或UGGUACCCACCUUCACUCUC (SEQ ID NO:96)。用于生成内源CD247基因座(编码CD3 ζ)的遗传破坏的示例性gRNA序列描述于例如国际PCT公开号WO 2017093969中。用于内源CD247基因座(编码CD3 ζ)的基因编辑的示例性方法包括例如WO 2017093969中描述的那些。可以使用任何已知方法来靶向并生成内源CD247基因座的遗传破坏,其可以用于本文提供的实施方案中。

[0200] 在一些实施方案中,靶向结构域包括用于使用酿脓链球菌(*S. pyogenes*) Cas9或使用脑膜炎奈瑟球菌(*N. meningitidis*) Cas9在CD247基因处引入遗传破坏的那些。

[0201] 在一些实施方案中,靶向结构域包括用于使用酿脓链球菌Cas9在CD247基因处引入遗传破坏的那些。任何靶向结构域可以与生成双链断裂(Cas9核酸酶)或单链断裂(Cas9切口酶)的酿脓链球菌Cas9分子一起使用。

[0202] 在一些实施方案中,通过使用酿脓链球菌Cas9切口酶使用双重靶向在相对DNA链上产生两个切口,所述切口酶具有与相对DNA链互补的两个靶向结构域,例如,包含任何负

链靶向结构域的gRNA可以与包含正链靶向结构域的任何gRNA配对。在一些实施方案中,两种gRNA在DNA上的定向使得PAM朝外,并且所述gRNA的5'端之间的距离为0-50bp。在一些实施方案中,使用两种gRNA靶向两种Cas9核酸酶或两种Cas9切口酶,例如,使用由两种不同gRNA分子指导的一对Cas9分子/gRNA分子复合物来切割靶结构域,其中在所述靶结构域的相对链上产生两个单链断裂。在一些实施方案中,两种Cas9切口酶可以包括具有HNH活性的分子,例如,RuvC活性失活的Cas9分子,例如在D10处具有突变(例如,D10A突变)的Cas9分子;具有RuvC活性的分子,例如HNH活性失活的Cas9分子,例如在H840处具有突变(例如,H840A)的Cas9分子;或者具有RuvC活性的分子,例如HNH活性失活的Cas9分子,例如在N863处具有突变(例如,N863A)的Cas9分子。在一些实施方案中,两种gRNA中的每一种与D10A Cas9切口酶复合。

(c) 第一互补结构域

[0203] 第一互补结构域与本文所述的第二互补结构域互补,并且通常与第二互补结构域具有足够的互补性以在至少一些生理条件下形成双链体化区域。第一互补结构域通常具有5至30个核苷酸的长度,并且可以具有5至25个核苷酸的长度、7至25个核苷酸的长度、7至22个核苷酸的长度、7至18个核苷酸的长度或7至15个核苷酸的长度。在各个实施方案中,第一互补结构域的长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸。第一互补结构域的例子包括在W02015/161276中(例如,在其中的图1A-图1G中)所述的那些。

[0204] 通常,第一互补结构域与第二互补结构域靶标不具有精确的互补性。在一些实施方案中,第一互补结构域可以具有与第二互补结构域的相应核苷酸不互补的1、2、3、4或5个核苷酸。在一些实施方案中,第一互补结构域的1、2、3、4、5或6个(例如,3个)核苷酸的区段可能在双链体中不配对,并且可以形成非双链体化或环凸(looped-out)区域。在一些情况下,未配对的(或环凸)区域(例如,3个核苷酸的环凸)存在于第二互补结构域上。该未配对区域任选地开始于距第二互补结构域的5'端1、2、3、4、5或6个(例如,4个)核苷酸处。

[0205] 第一互补结构域可以包括3个子结构域,它们在5'至3'方向上是:5'子结构域、中心子结构域和3'子结构域。在一些实施方案中,5'子结构域具有4-9(例如,4、5、6、7、8或9)个核苷酸的长度。在一些实施方案中,中心子结构域具有1、2或3(例如,1)个核苷酸的长度。在一些实施方案中,3'子结构域具有3至25(例如,4-22、4-18、或4至10、或者3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25)个核苷酸的长度。

[0206] 在一些实施方案中,第一和第二互补结构域在双链体化时包含11个配对的核苷酸,例如,在gRNA序列中(一条配对的链加下划线,一条加粗):
NNNNNNNNNNNNN
NNNNNNNNGUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUA
UCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 97)。

[0207] 在一些实施方案中,第一和第二互补结构域在双链体化时包含15个配对的核苷酸,例如,在gRNA序列中(一条配对的链加下划线,一条加粗):
NNNNNNNNNNNNN
NNNNNNNNGUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUAAAAUAAGGCUA
GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 98)。

[0208] 在一些实施方案中,第一和第二互补结构域在双链体化时包含16个配对的核苷酸,例如在gRNA序列中(一条配对的链加下划线,一条加粗):
NNNNNNNNNNNNN

NNNNNNNGUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAAUAAGGCU
AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 99)。

[0209] 在一些实施方案中,第一和第二互补结构域在双链体化时包含21个配对的核苷酸,例如在gRNA序列中(一条配对的链加下划线,一条加粗):
NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUU
AAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 100)。

[0210] 在一些实施方案中,交换核苷酸以去除多聚U束,例如在gRNA序列中(交换的核苷酸加下划线):
NNNNNNNNNNNNNNNNNNNGUAUUAGAGCUAGAAAUAGCAAGUUAAUAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:101);
NNNNNNNNNNNNNNNNNNNGUUUAAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:102);
以及NNNNNNNNNNNNNNNNNNNGUAUUAGAGCUAUGCUGUAUUGAAACAAUACAGCAUAGCAAGUUAAUAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:103)。

[0211] 第一互补结构域可以与天然存在的第一互补结构域共有同源性,或者源自所述天然存在的第一互补结构域。在一些实施方案中,其具有与本文公开的第一互补结构域(例如酿脓链球菌、金黄色葡萄球菌(*S.aureus*)、脑膜炎奈瑟球菌(*N.meningitidis*)或唾液链球菌(*S.thermophilus*)第一互补结构域)的至少50%同源性。

[0212] 应注意,第一互补结构域的一个或多个或甚至所有核苷酸可以沿本文针对靶向结构域所讨论的路线具有修饰。

(d) 连接结构域

[0213] 在单分子或嵌合gRNA中,连接结构域用于将单分子gRNA的第一互补结构域与第二互补结构域连接。连接结构域可以将第一和第二互补结构域共价或非共价地连接。在一些实施方案中,连接是共价的。在一些实施方案中,连接结构域将第一和第二互补结构域共价地偶联,参见例如,W02015/161276,例如在其中的图1B-图1E中。在一些实施方案中,连接结构域是或包含插入第一互补结构域与第二互补结构域之间的共价键。通常,连接结构域包含一个或多个(例如,2、3、4、5、6、7、8、9或10个)核苷酸,但是在各个实施方案中,接头可以具有20、30、40、50或甚至100个核苷酸的长度。连接结构域的例子包括在W02015/161276中(例如,在其中的图1A-图1G中)所述的那些。

[0214] 在模块化gRNA分子中,两个分子凭借互补结构域的杂交而缔合,并且连接结构域可以不存在。参见例如,W02015/161276,例如在其中的图1A中。

[0215] 各种各样的连接结构域适用于单分子gRNA分子。连接结构域可以由共价键组成,或者短至一个或几个核苷酸,例如长度为1、2、3、4或5个核苷酸。在一些实施方案中,连接结构域具有2、3、4、5、6、7、8、9、10、15、20或25个或更多个核苷酸的长度。在一些实施方案中,连接结构域具有2至50、2至40、2至30、2至20、2至10或2至5个核苷酸的长度。在一些实施方案中,连接结构域与天然存在的序列(例如,位于第二互补结构域5'的tracrRNA的序列)共有同源性,或者源自所述天然存在的序列。在一些实施方案中,连接结构域具有与本文所公开的连接结构域的至少50%同源性。

[0216] 如本文结合第一互补结构域所讨论,连接结构域的一些或所有核苷酸可以包括修饰。

(e) 5'延伸结构域

[0217] 在一些情形中,模块化gRNA可以在第二互补结构域的5'包含另外的序列,在本文中称为5'延伸结构域。在一些实施方案中,5'延伸结构域具有2-10、2-9、2-8、2-7、2-6、2-5或2-4个核苷酸的长度。在一些实施方案中,5'延伸结构域具有2、3、4、5、6、7、8、9或10个或更多个核苷酸的长度。在一些实施方案中,5'延伸结构域的例子包括在W02015/161276中(例如在其中的图1A中)所述的那些。

(f) 第二互补结构域

[0218] 第二互补结构域与第一互补结构域互补,并且通常与第二互补结构域具有足够的互补性以在至少一些生理条件下形成双链体化区域。在一些情况下,例如如在W02015/161276中(例如,在其中的图1A-图1B中)所示,第二互补结构域可以包括与第一互补结构域缺乏互补性的序列,例如从双链体化区域环凸的序列。第二互补结构域的例子包括在W02015/161276中(例如,在其中的图1A-图1G中)所述的那些。

[0219] 第二互补结构域可以具有5至27个核苷酸的长度,并且在一些情况下可以比第一互补区长。在一些实施方案中,第二互补结构域可以具有7至27个核苷酸的长度、7至25个核苷酸的长度、7至20个核苷酸的长度或7至17个核苷酸的长度。更通常地,互补结构域可以具有5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸的长度。

[0220] 在一些实施方案中,第二互补结构域包含3个子结构域,它们在5'至3'方向上是:5'子结构域、中心子结构域和3'子结构域。在一些实施方案中,5'子结构域具有3至25(例如,4至22、4至18、或4至10、或者3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25)个核苷酸的长度。在一些实施方案中,中心子结构域具有1、2、3、4或5(例如,3)个核苷酸的长度。在一些实施方案中,3'子结构域具有4至9(例如,4、5、6、7、8或9)个核苷酸的长度。

[0221] 在一些实施方案中,第一互补结构域的5'子结构域和3'子结构域分别与第二互补结构域的3'子结构域和5'子结构域互补,例如完全互补。

[0222] 第二互补结构域可以与天然存在的第二互补结构域共有同源性,或者源自所述天然存在的第二互补结构域。在一些实施方案中,其具有与本文公开的第二互补结构域(例如,酿脓链球菌、金黄色葡萄球菌、脑膜炎奈瑟球菌或唾液链球菌第一互补结构域)至少50%同源性。

[0223] 第二互补结构域的一些或所有核苷酸可以具有修饰,例如本文所述的修饰。

(g) 近端结构域

[0224] 近端结构域的例子包括在W02015/161276中(例如,在其中的图1A-图1G中)所述的那些。在一些实施方案中,近端结构域具有5至20个核苷酸的长度。在一些实施方案中,近端结构域可以与天然存在的近端结构域共有同源性,或者源自所述天然存在的近端结构域。在一些实施方案中,其具有与本文公开的近端结构域(例如,酿脓链球菌、金黄色葡萄球菌、脑膜炎奈瑟球菌或唾液链球菌近端结构域)的至少50%同源性。

[0225] 近端结构域的一些或所有核苷酸可以沿本文所述的路线具有修饰。

(h) 尾结构域

[0226] 如通过检查W02015/161276中(例如,在其中的图1A和图1B-图1F中)的尾结构域可见,广谱的尾结构域适用于gRNA分子。在各个实施方案中,尾结构域具有0(不存在)、1、2、

3、4、5、6、7、8、9或10个核苷酸的长度。在某些实施方案中,尾结构域核苷酸来自天然存在的尾结构域的5'端的序列或与所述天然存在的尾结构域的5'端的序列共有同源性,参见例如,WO 2015/161276,例如在其中的图1D或图1E中。尾结构域还任选地包括彼此互补并且在至少一些生理条件下形成双链体化区域的序列。尾结构域的例子包括在WO 2015/161276中(例如,在其中的图1A-图1G中)所述的那些。

[0227] 尾结构域可以与天然存在的近端尾结构域共有同源性,或者源自所述天然存在的近端尾结构域。作为非限制性例子,根据本公开文本的各个实施方案的给定尾结构域可以与本文所公开的天然存在的尾结构域(例如,酿脓链球菌、金黄色葡萄球菌、脑膜炎奈瑟球菌或唾液链球菌尾结构域)共有至少50%同源性。

[0228] 在某些情形中,尾结构域在3'端包括与体外或体内转录方法有关的核苷酸。当T7启动子用于gRNA的体外转录时,这些核苷酸可以是存在于DNA模板的3'端之前的任何核苷酸。当U6启动子用于体内转录时,这些核苷酸可以是序列UUUUUU。当使用替代p_{ol}-III启动子时,这些核苷酸可以是各种数量的尿嘧啶碱基,或者可以包括替代碱基。

[0229] 作为非限制性例子,在各个实施方案中,近端结构域和尾结构域一起包含以下序列:

AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGUCU (SEQ ID NO:104)、AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGUGUC (SEQ ID NO:105)、AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCGGAUC (SEQ ID NO:106)、AAGGCUAGUCCGUUAUCAACUUGAAAAAGUG (SEQ ID NO:107)、AAGGCUAGUCCGUUAUCA (SEQ ID NO:108) 或AAGGCUAGUCCG (SEQ ID NO:109)。

[0230] 在一些实施方案中,例如,如果U6启动子用于转录,则尾结构域包含3'序列UUUUUU。在一些实施方案中,例如,如果H1启动子用于转录,则尾结构域包含3'序列UUUU。在一些实施方案中,尾结构域包含可变数量的3'U,这取决于例如所用p_{ol}-III启动子的终止信号。在一些实施方案中,如果使用T7启动子,则尾结构域包含源自DNA模板的可变3'序列。在一些实施方案中,例如,如果使用体外转录来生成RNA分子,则尾结构域包含源自DNA模板的可变3'序列。在一些实施方案中,例如,如果使用p_{ol}-II启动子来驱动转录,则尾结构域包含源自DNA模板的可变3'序列。

[0231] 在一些实施方案中,gRNA具有以下结构:5' [靶向结构域]-[第一互补结构域]-[连接结构域]-[第二互补结构域]-[近端结构域]-[尾结构域]-3',其中,靶向结构域包含核心结构域和任选地二级结构域,并且具有10至50个核苷酸的长度;第一互补结构域具有5至25个核苷酸的长度,并且在一些实施方案中具有与本文所公开的参考第一互补结构域的至少50%、60%、70%、80%、85%、90%、95%、98%或99%同源性;连接结构域具有1至5个核苷酸的长度;近端结构域具有5至20个核苷酸的长度,并且在一些实施方案中具有与本文所公开的参考近端结构域的至少50%、60%、70%、80%、85%、90%、95%、98%或99%同源性;并且尾结构域不存在或者核苷酸序列具有1至50个核苷酸的长度,并且在一些实施方案中具有与本文所公开的参考尾结构域的至少50%、60%、70%、80%、85%、90%、95%、98%或99%同源性。

(i) 示例性嵌合gRNA

[0232] 在一些实施方案中,单分子或嵌合gRNA优选地从5'至3'包含:靶向结构域,例如,

其包含15、16、17、18、19、20、21、22、23、24、25或26个核苷酸(其与靶核酸互补);第一互补结构域;连接结构域;第二互补结构域(其与第一互补结构域互补);近端结构域;以及尾结构域,其中,(a)近端结构域和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;或(c)在第二互补结构域的最后一个核苷酸(其与第一互补结构域中的其相应核苷酸互补)的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0233] 在一些实施方案中,来自(a)、(b)或(c)的序列具有与天然存在的gRNA的相应序列或与本文所述的gRNA的至少60%、75%、80%、85%、90%、95%或99%同源性。在一些实施方案中,近端结构域和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。在一些实施方案中,在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。在一些实施方案中,在第二互补结构域的最后一个核苷酸(其与第一互补结构域中的其相应核苷酸互补)的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。在一些实施方案中,靶向结构域包含,具有16、17、18、19、20、21、22、23、24、25或26个核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)或由所述核苷酸组成,所述核苷酸与靶结构域具有互补性,例如,靶向结构域具有16、17、18、19、20、21、22、23、24、25或26个核苷酸的长度。

[0234] 在一些实施方案中,单分子或嵌合gRNA分子(包含靶向结构域、第一互补结构域、连接结构域、第二互补结构域、近端结构域和任选地尾结构域)包含以下序列,其中靶向结构域被描绘为20个N,但是可以是任何序列并且长度的范围是从16至26个核苷酸,并且其中gRNA序列后接6个U,其用作U6启动子的终止信号,但是可以不存在或数量更少:NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAUAGCAAGUUAUUUUUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU(SEQ ID NO:110)。在一些实施方案中,单分子或嵌合gRNA分子是酿脓链球菌gRNA分子。

[0235] 在一些实施方案中,单分子或嵌合gRNA分子(包含靶向结构域、第一互补结构域、连接结构域、第二互补结构域、近端结构域和任选地尾结构域)包含以下序列,其中靶向结构域被描绘为20个N,但是可以是任何序列并且长度范围为16至26个核苷酸,并且其中gRNA序列之后是6个U,所述6个U用作U6启动子的终止信号,但是可以不存在或数量更少:NNNNNNNNNNNNNNNNNNNGUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUC AACUUGUUGGCGAGAUUUUUU(SEQ ID NO:111)。在一些实施方案中,单分子或嵌合gRNA分子是金黄色葡萄球菌gRNA分子。示范性嵌合gRNA的序列和结构还显示于W02015/161276中,例如在其中的图10A-图10B中。

[0236] 如本文所述的任何gRNA分子都可以与生成双链断裂或单链断裂的任何Cas9分子一起用于改变靶核酸的序列,例如,靶位置或靶遗传特征。在一些例子中,靶核酸位于CD247基因座处或附近,例如如所述的任一种。在一些实施方案中,将核糖核酸分子(如gRNA分子)和蛋白质(如Cas9蛋白或其变体)引入本文提供的任何工程化细胞中。可用于这些方法中的gRNA分子描述于下文中。

[0237] 在一些实施方案中,gRNA(例如,嵌合gRNA)配置为使得其包含以下特性中的一种或多种:

a) 例如,当靶向进行双链断裂的Cas9分子时,它可以将双链断裂定位于(i)靶位置的50、100、150、200、250、300、350、400、450或500个核苷酸内,或者(ii)足够接近以使靶位置位于末端切除区域内;

b) 它具有至少16个核苷酸的靶向结构域,例如(i) 16、(ii) 17、(iii) 18、(iv) 19、(v) 20、(vi) 21、(vii) 22、(viii) 23、(ix) 24、(x) 25或(xi) 26个核苷酸的靶向结构域;以及

c) (i) 近端和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾和近端结构域或者与其相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;

(ii) 在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌gRNA的相应序列或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;

(iii) 在第二互补结构域中与第一互补结构域的其相应核苷酸互补的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌gRNA的相应序列或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸;

(iv) 尾结构域具有至少10、15、20、25、30、35或40个核苷酸的长度,例如,其包含来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾结构域或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少10、15、20、25、30、35或40个核苷酸;或者

(v) 尾结构域包含天然存在的尾结构域(例如,天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾结构域)的15、20、25、30、35、40个核苷酸或所有相应部分。

[0238] 在一些实施方案中,gRNA配置为使得其包含特性:a和b(i)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(iii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(iv)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(v)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(vi)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(vii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(viii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(ix)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(x)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(xi)。在一些实施方案中,gRNA配置为使得其包含特性:a和c。在一些实施方案中,gRNA配置为使得其包含特性:a、b和c。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(i)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(i)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a

(i)、b(iv)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iv)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(v)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(v)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vi)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vi)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(viii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(viii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ix)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ix)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(x)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(x)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(xi)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(xi)和c(ii)。

[0239] 在一些实施方案中,gRNA(例如,嵌合gRNA)配置为使得其包含以下特性中的一种或多种:

a) 一个或两个gRNA可以例如在靶向进行单链断裂的Cas9分子时将单链断裂定位于(i)靶位置的50、100、150、200、250、300、350、400、450或500个核苷酸内,或者(ii)足够接近以使靶位置位于末端切除区域内;

b) 一个或两个具有至少16个核苷酸的靶向结构域,例如,(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25或(xi)26个核苷酸的靶向结构域;以及

c) (i) 近端和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾和近端结构域或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;

(ii) 在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌gRNA的相应序列或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;

(iii) 在第二互补结构域中与第一互补结构域的其相应核苷酸互补的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌gRNA的相应序列或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸;

(iv) 尾结构域具有至少10、15、20、25、30、35或40个核苷酸的长度,例如,其包含来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾结构域或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少10、15、20、25、30、35或40个核苷酸;或者

(v) 尾结构域包含天然存在的尾结构域(例如,天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾结构域)的15、20、25、30、35、40个核苷酸或所有相

应部分。

[0240] 在一些实施方案中,gRNA配置为使得其包含特性:a和b(i)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(iii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(iv)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(v)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(vi)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(vii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(viii)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(ix)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(x)。在一些实施方案中,gRNA配置为使得其包含特性:a和b(xi)。在一些实施方案中,gRNA配置为使得其包含特性:a和c。在一些实施方案中,gRNA配置为使得其包含特性:a、b和c。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(i)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(i)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iv)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(iv)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(v)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(v)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vi)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vi)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(vii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(viii)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(viii)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ix)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(ix)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(x)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(x)和c(ii)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(xi)和c(i)。在一些实施方案中,gRNA配置为使得其包含特性:a(i)、b(xi)和c(ii)。

[0241] 在一些实施方案中,将gRNA与具有HNH活性的Cas9切口酶分子(例如RuvC活性失活的Cas9分子,例如在D10处具有突变(例如,D10A突变)的Cas9分子)一起使用。

[0242] 在一些实施方案中,将gRNA与具有RuvC活性的Cas9切口酶分子(例如HNH活性失活的Cas9分子,例如在H840处具有突变(例如,H840A)的Cas9分子)一起使用。

[0243] 在一些实施方案中,包含第一和第二gRNA的一对gRNA(例如,一对嵌合gRNA)配置为使得它们包含以下特性中的一种或多种:

a) 一个或两个gRNA可以例如在靶向进行单链断裂的Cas9分子时将单链断裂定位于(i)靶位置的50、100、150、200、250、300、350、400、450或500个核苷酸内,或者(ii)足够接近以使靶位置位于末端切除区域内;

b) 一个或两个具有至少16个核苷酸的靶向结构域,例如,(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25或(xi)26个核苷酸的靶向结构域;

c) 对于一者或两者:

(i) 近端和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾和近端结构域或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;

(ii) 在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌gRNA的相应序列或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;

(iii) 在第二互补结构域中与第一互补结构域的其相应核苷酸互补的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,例如,来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌gRNA的相应序列或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸;

(iv) 尾结构域具有至少10、15、20、25、30、35或40个核苷酸的长度,例如,其包含来自天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾结构域或者与其相差不超过1、2、3、4、5;6、7、8、9或10个核苷酸的序列的至少10、15、20、25、30、35或40个核苷酸;或者

(v) 尾结构域包含天然存在的尾结构域(例如,天然存在的酿脓链球菌、唾液链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌尾结构域)的15、20、25、30、35、40个核苷酸或所有相应部分;

d) gRNA配置为使得在与靶核酸杂交时,它们由0-50、0-100、0-200、至少10、至少20、至少30或至少50个核苷酸隔开;

e) 由第一gRNA和第二gRNA进行的断裂位于不同链上;以及

f) PAM朝外。

[0244] 在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(i)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(ii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(iii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(iv)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(v)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(vi)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(vii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(viii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(ix)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(x)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和b(xi)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a和c。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a、b和c。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(i)和c(i)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(i)和c(ii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(i)、c和d。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(i)、c和e。

(i)、b(x)和c(ii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(x)、c和d。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(x)、c和e。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(x)、c、d和e。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(xi)和c(i)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(xi)和c(ii)。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(xi)、c和d。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(xi)、c和e。在一些实施方案中,一个或两个gRNA配置为使得其包含特性:a(i)、b(xi)、c、d和e。

[0245] 在一些实施方案中,将gRNA与具有HNH活性的Cas9切口酶分子(例如RuvC活性失活的Cas9分子,例如在D10处具有突变(例如,D10A突变)的Cas9分子)一起使用。

[0246] 在一些实施方案中,将gRNA与具有RuvC活性的Cas9切口酶分子(例如HNH活性失活的Cas9分子,例如在H840处具有突变(例如,H840A)的Cas9分子)一起使用。在一些实施方案中,将gRNA与具有RuvC活性的Cas9切口酶分子(例如HNH活性失活的Cas9分子,例如在N863处具有突变(例如,N863A)的Cas9分子)一起使用。

(j) 示例性模块化gRNA

[0247] 在一些实施方案中,模块化gRNA包含第一链和第二链。第一链优选地从5'至3'包含:靶向结构域,例如其包含15、16、17、18、19、20、21、22、23、24、25或26个核苷酸;第一互补结构域。第二链优选地从5'至3'包含:任选地5'延伸结构域;第二互补结构域;近端结构域;以及尾结构域,其中:(a)近端和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;或者(c)在第二互补结构域中与第一互补结构域的其相应核苷酸互补的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0248] 在一些实施方案中,来自(a)、(b)或(c)的序列具有与天然存在的gRNA的相应序列或与本文所述的gRNA的至少60%、75%、80%、85%、90%、95%或99%同源性。在一些实施方案中,近端结构域和尾结构域在一起时包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。在一些实施方案中,在第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0249] 在一些实施方案中,在第二互补结构域的最后一个核苷酸(其与第一互补结构域中的其相应核苷酸互补)的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0250] 在一些实施方案中,靶向结构域具有16、17、18、19、20、21、22、23、24、25或26个核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)或由所述核苷酸组成,所述核苷酸具有与靶结构域的互补性,例如,靶向结构域具有16、17、18、19、20、21、22、23、24、25或26个核苷酸的长度。

(k) 用于设计gRNA的方法

[0251] 用于设计gRNA的方法描述于本文中,包括用于选择、设计和验证靶向结构域的方法。本文还提供示例性靶向结构域。可以将本文所讨论的靶向结构域掺入本文所述的gRNA中。

[0252] 用于靶序列的选择和验证以及脱靶分析的方法描述于例如以下文献中:Mali等人,2013Science 339(6121):823-826;Hsu等人Nat Biotechnol,31(9):827-32;Fu等人,2014Nat Biotechnol,doi:10.1038/nbt.2808.PubMed PMID:24463574;Heigwer等人,2014Nat Methods 11(2):122-3.doi:10.1038/nmeth.2812.PubMed PMID:24481216;Bae等人,2014Bioinformatics PubMed PMID:24463181;Xiao A等人,2014Bioinformatics PubMed PMID:24389662。

[0253] 在一些实施方案中,可以使用软件工具优化使用者的靶序列内gRNA的选择,例如以使整个基因组中的总脱靶活性降至最低。脱靶活性可以与切割不同。例如,对于使用酿脓链球菌Cas9的每种可能的gRNA选择,软件工具可以鉴定整个基因组中所有潜在的脱靶序列(前述NAG或NGG PAM),所述潜在的脱靶序列含有最多某个数量(例如,1、2、3、4、5、6、7、8、9或10)的错配碱基对。可以例如使用源自实验的加权方案来预测每个脱靶序列处的切割效率。然后将每种可能的gRNA根据其总预测脱靶切割进行评级;最高评级的gRNA表示可能具有最高中靶和最低脱靶切割的那些。工具中还可以包括其他功能,例如用于gRNA载体构建的自动化试剂设计、用于中靶Surveyor测定的引物设计以及用于经由下一代测序进行脱靶切割的高通量检测和定量的引物设计。候选gRNA分子可以通过本领域已知的方法或如本文所述来评价。

[0254] 在一些实施方案中,使用的gRNA使用DNA序列搜索算法(例如,使用基于公共工具cas-offinder的定制gRNA设计软件)来鉴定与酿脓链球菌、金黄色葡萄球菌和脑膜炎奈瑟球菌Cas9一起(Bae等人Bioinformatics.2014;30(10):1473-1475)。定制gRNA设计软件在计算指导物的全基因组脱靶倾向后对所述指导物进行评分。通常,对于长度范围从17至24的指导物,考虑范围从完全匹配至7个错配的匹配。在一些方面,一旦通过计算确定脱靶位点,计算每种指导物的总得分,并且使用网络界面归纳于表格输出中。除了鉴定与PAM序列相邻的潜在gRNA位点,所述软件还可以鉴定与所选gRNA位点相差1、2、3个或更多个核苷酸的所有PAM相邻序列。在一些实施方案中,每个基因的基因组DNA序列是从UCSC基因组浏览器获得的,并且可以使用公众可获得的RepeatMasker程序针对重复元件筛选序列。RepeatMasker针对重复元件和低复杂度区域搜索输入DNA序列。输出是给定查询序列中存在的重复序列的详细注释。

[0255] 在鉴定后,可以基于以下中的一项或多项将gRNA评级为多层:其与靶位点的距离、其正交性和5'G的存在(基于人基因组中含有相关PAM的紧密匹配的鉴定,例如,在酿脓链球菌的情形中为NGG PAM,在金黄色葡萄球菌的情形中为NNGRR(例如,NNGRRT或NNGRRV)PAM,并且在脑膜炎奈瑟球菌的情形中为NNNGATT或NNNGCTT PAM)。正交性是指人基因组中含有与靶序列的最小错配数的序列的数量。“高正交性水平”或“良好正交性”可以例如是指20聚体靶向结构域,其在人基因组中除了预期靶标外不具有相同的序列,也不具有在靶序列中含有一个或两个错配的任何序列。选择具有良好正交性的靶向结构域以使脱靶DNA切割降至最低。应理解,这是非限制性例子,并且可以利用多种策略来鉴定与酿脓链球菌、金黄色葡萄球菌和脑膜炎奈瑟球菌或其他Cas9酶一起使用的gRNA。

[0256] 在一些实施方案中,与酿脓链球菌Cas9一起使用的gRNA可以使用公众可获得的基于网络的ZiFiT服务器来鉴定(Fu等人,Improving CRISPR-Cas nuclease specificity using truncated guide RNA.Nat Biotechnol.2014年1月26日.doi:10.1038/

nbt.2808.PubMed PMID:24463574,原始参考文献参见Sander等人,2007,NAR 35:W599-605;Sander等人,2010,NAR 38:W462-8)。除了鉴定与PAM序列相邻的潜在gRNA位点,所述软件还鉴定与所选gRNA位点相差1、2、3个或更多个核苷酸的所有PAM相邻序列。在一些方面,每种基因的基因组DNA序列可以从UCSC基因组浏览器获得,并且可以使用公众可获得的Repeat-Masker程序针对重复元件筛选序列。RepeatMasker针对重复元件和低复杂度区域搜索输入DNA序列。输出是给定查询序列中存在的重复序列的详细注释。

[0257] 在鉴定后,可以将与酿脓链球菌Cas9一起使用的gRNA评级为多层,例如5层。在一些实施方案中,第一层gRNA分子的靶向结构域是基于以下来选择:其与靶位点的距离、其正交性和5'G的存在(基于人基因组中含有NGG PAM的紧密匹配的ZiFiT鉴定)。在一些实施方案中,针对靶标设计17聚体和20聚体gRNA。在一些方面,还同时针对单gRNA核酸酶切割以及针对双重gRNA切口酶策略选择gRNA。选择gRNA以及确定哪种gRNA可以用于哪种策略的标准可以基于几种考虑因素。在一些实施方案中,鉴定同时用于单gRNA核酸酶切割和用于双重gRNA配对的“切口酶”策略的gRNA。在一些实施方案中,对于选择gRNA,包括确定哪种gRNA可以用于双重gRNA配对的“切口酶”策略,gRNA对在DNA上的定向应当使得PAM朝外并且用D10A Cas9切口酶切割将得到5'突出端。在一些方面,可以假定,用双重切口酶对切割将以合理的频率导致整个间插序列缺失。然而,用双重切口酶对切割通常还可能在仅一种gRNA的位点处导致插入缺失突变。可以针对相比于仅在一种gRNA的位点处引起插入缺失突变,它们去除整个序列的效率来测试候选对成员。

[0258] 在一些实施方案中,第一层gRNA分子的靶向结构域可以基于以下来选择:(1)与靶位置的合理距离,例如在起始密码子下游的编码序列的前500bp内,(2)高正交性水平,以及(3)5'G的存在。在一些实施方案中,对于第二层gRNA的选择,可以取消对5'G的需要,但是需要距离限制并且需要高正交性水平。在一些实施方案中,第三层选择使用相同的距离限制和对5'G的需要,但是取消了对良好正交性的需要。在一些实施方案中,第四层选择使用相同的距离限制,但是取消了对良好正交性的需要并且以5'G开始。在一些实施方案中,第五层选择取消了对良好正交性和5'G的需要,并且扫描更长的序列(例如,编码序列的其余部分,例如转录靶位点上游或下游的另外500bp)。在某些情况下,基于特定层的标准没有鉴定出gRNA。

[0259] 在一些实施方案中,针对单gRNA核酸酶切割以及针对双重gRNA配对的“切口酶”策略来鉴定gRNA。

[0260] 在一些方面,可以通过扫描基因组DNA序列中PAM序列的存在来人工鉴定与脑膜炎奈瑟球菌和金黄色葡萄球菌Cas9一起使用的gRNA。可以将这些gRNA分为两层。在一些实施方案中,对于第一层gRNA,在起始密码子下游的编码序列的前500bp内选择靶向结构域。在一些实施方案中,对于第二层gRNA,在其余编码序列(前500bp的下游)内选择靶向结构域。在某些情况下,基于特定层的标准没有鉴定出gRNA。

[0261] 在一些实施方案中,用于鉴定与酿脓链球菌、金黄色葡萄球菌和脑膜炎奈瑟球菌Cas9一起使用的指导RNA(gRNA)的另一种策略可以使用DNA序列搜索算法。在一些方面,指导RNA设计是使用基于公共工具cas-offinder的定制指导RNA设计软件来进行的(Bae等人Bioinformatics.2014;30(10):1473-1475)。所述定制指导RNA设计软件在计算指导物的全基因组脱靶倾向后对所述指导物进行评分。通常,对于长度范围从17至24的指导物,考虑范

围从完全匹配至7个错配的匹配。一旦通过计算确定脱靶位点,计算每种指导物的总得分,并且使用网络界面归纳于表格输出中。除了鉴定与PAM序列相邻的潜在gRNA位点,所述软件还鉴定与所选gRNA位点相差1、2、3个或更多个核苷酸的所有PAM相邻序列。在一些实施方案中,每种基因的基因组DNA序列是从UCSC基因组浏览器获得,并且使用公众可获得的RepeatMasker程序针对重复元件筛选序列。RepeatMasker针对重复元件和低复杂度区域搜索输入DNA序列。输出是给定查询序列中存在的重复序列的详细注释。

[0262] 在一些实施方案中,在鉴定后,基于以下将gRNA评级为多层:其与靶位点的距离或其正交性(基于人基因组中含有相关PAM的紧密匹配的鉴定,例如,在酿脓链球菌的情形中为NGG PAM,在金黄色葡萄球菌的情形中为NNGRR(例如,NNGRRT或NNGRRV) PAM,并且在脑膜炎奈瑟球菌的情形中为NNNGATT或NNNGCTT PAM)。在一些方面,选择具有良好正交性的靶向结构域以使脱靶DNA切割降至最低。

[0263] 作为例子,对于酿脓链球菌和脑膜炎奈瑟球菌靶标,可以设计17聚体或20聚体gRNA。作为另一例子,对于金黄色葡萄球菌靶标,可以设计18聚体、19聚体、20聚体、21聚体、22聚体、23聚体和24聚体gRNA。

[0264] 在一些实施方案中,鉴定同时用于单gRNA核酸酶切割和用于双重gRNA配对的“切口酶”策略的gRNA。在一些实施方案中,对于选择gRNA,包括确定哪种gRNA可以用于双重gRNA配对的“切口酶”策略,gRNA对在DNA上的定向应当使得PAM朝外并且用D10A Cas9切口酶切割将得到5'突出端。在一些方面,可以假定,用双重切口酶对切割将以合理的频率导致整个间插序列缺失。然而,用双重切口酶对切割通常还可能在仅一种gRNA的位点处导致插入缺失突变。可以针对相比于仅在一种gRNA的位点处引起插入缺失突变,它们去除整个序列的效率来测试候选对成员。

[0265] 对于设计遗传破坏策略,在一些实施方案中,用于酿脓链球菌的第1层gRNA分子的靶向结构域是基于其与靶位点的距离及其正交性(PAM是NGG)来选择。在一些情况下,第1层gRNA分子的靶向结构域是基于以下来选择:(1)与靶位置的合理距离,例如在起始密码子下游的编码序列的前500bp内;以及(2)高正交性水平。在一些方面,对于第2层gRNA的选择,不需要高正交性水平。在一些情况下,第3层gRNA取消对良好正交性的需要,并且可以扫描更长的序列(例如,编码序列的其余部分)。在某些情况下,基于特定层的标准没有鉴定出gRNA。

[0266] 对于设计遗传破坏策略,在一些实施方案中,用于脑膜炎奈瑟球菌的第1层gRNA分子的靶向结构域是在编码序列的前500bp内选择,并且具有高正交性水平。用于脑膜炎奈瑟球菌的第2层gRNA分子的靶向结构域是在编码序列的前500bp内选择,并且不需要高正交性。用于脑膜炎奈瑟球菌的第3层gRNA分子的靶向结构域是在所述500bp下游的编码序列的其余部分内选择。注意,层是非包含性的(每种gRNA仅列出一)。在某些情况下,基于特定层的标准没有鉴定出gRNA。

[0267] 对于设计遗传破坏策略,在一些实施方案中,用于金黄色葡萄球菌的第1层gRNA分子的靶向结构域是在编码序列的前500bp内选择,具有高正交性水平,并且含有NNGRRT PAM。在一些实施方案中,用于金黄色葡萄球菌的第2层gRNA分子的靶向结构域是在编码序列的前500bp内选择,不需要正交性水平,并且含有NNGRRT PAM。在一些实施方案中,用于金黄色葡萄球菌的第3层gRNA分子的靶向结构域是在下游的编码序列的其余部分内选择,并

且含有NNGRRRT PAM。在一些实施方案中,用于金黄色葡萄球菌的第4层gRNA分子的靶向结构域是在编码序列的前500bp内选择,并且含有NNGRRV PAM。在一些实施方案中,用于金黄色葡萄球菌的第5层gRNA分子的靶向结构域是在下游的编码序列的其余部分内选择,并且含有NNGRRV PAM。在某些情况下,基于特定层的标准没有鉴定出gRNA。

(ii) Cas9

[0268] 多种物种的Cas9分子可以用于本文所述的方法和组合物中。尽管酿脓链球菌、金黄色葡萄球菌、脑膜炎奈瑟球菌和唾液链球菌Cas9分子是本文中大部分公开文本的主题,也可以使用本文列出的其他物种的Cas9分子、源自所述其他物种的Cas9分子或者基于所述其他物种的Cas9蛋白的Cas9分子。换句话说,尽管本文中大部分描述使用酿脓链球菌、金黄色葡萄球菌、脑膜炎奈瑟球菌和唾液链球菌Cas9分子,但是来自其他物种的Cas9分子可以替代它们。此类物种包括:燕麦食酸菌(*Acidovorax avenae*)、胸膜肺炎放线杆菌(*Actinobacillus pleuropneumoniae*)、琥珀酸放线杆菌(*Actinobacillus succinogenes*)、猪放线杆菌(*Actinobacillus suis*)、放线菌属物种(*Actinomyces sp.*)、*Cycliphilus denitrificans*、寡食氨基酸单胞菌(*Aminomonas paucivorans*)、蜡样芽孢杆菌(*Bacillus cereus*)、史氏芽孢杆菌(*Bacillus smithii*)、苏云金芽孢杆菌(*Bacillus thuringiensis*)、拟杆菌属物种(*Bacteroides sp.*)、*Blastopirellula marina*、慢生根瘤菌属物种(*Bradyrhizobium sp.*)、侧孢短芽孢杆菌(*Brevibacillus laterosporus*)、结肠弯曲菌(*Campylobacter coli*)、空肠弯曲菌(*Campylobacter jejuni*)、红嘴鸥弯曲菌(*Campylobacter lari*)、*Candidatus puniceispirillum*、解纤维梭菌(*Clostridium cellulolyticum*)、产气荚膜梭菌(*Clostridium perfringens*)、拥挤棒杆菌(*Corynebacterium accolens*)、白喉棒杆菌(*Corynebacterium diphtheria*)、马氏棒杆菌(*Corynebacterium matruchotii*)、*Dinoroseobacter shibae*、细长真杆菌(*Eubacterium dolichum*)、 γ -变形菌(*Gammaproteobacterium*)、重氮营养葡萄糖酸醋杆菌(*Gluconacetobacter diazotrophicus*)、副流感嗜血杆菌(*Haemophilus parainfluenzae*)、*Haemophilus sputorum*、加拿大螺杆菌(*Helicobacter canadensis*)、同性恋螺杆菌(*Helicobacter cinaedi*)、雪貂螺杆菌(*Helicobacter mustelae*)、营养泥杆菌(*Ilyobacter polytropus*)、金氏金氏菌(*Kingella kingae*)、卷曲乳杆菌(*Lactobacillus crispatus*)、伊氏李斯特菌(*Listeria ivanovii*)、单核细胞增生李斯特菌(*Listeria monocytogenes*)、李斯特氏菌(*Listeriaceae*)细菌、甲基孢囊菌属物种(*Methylocystis sp.*)、甲基弯菌(*Methylosinus trichosporium*)、羞怯动弯杆菌(*Mobiluncus mulieris*)、杆菌状奈瑟球菌(*Neisseria bacilliformis*)、灰色奈瑟球菌(*Neisseria cinerea*)、金黄奈瑟球菌(*Neisseria flavescens*)、乳糖奈瑟球菌(*Neisseria lactamica*)、脑膜炎奈瑟球菌(*Neisseria meningitidis*)、奈瑟球菌属物种(*Neisseria sp.*)、瓦茨瓦尔西奈瑟球菌(*Neisseria wadsworthii*)、亚硝化单胞菌属物种(*Nitrosomonas sp.*)、*Parvibaculum lavamentivorans*、多杀巴斯德菌(*Pasteurella multocida*)、*Phascolarctobacterium succinatutens*、*Ralstonia syzygii*、沼泽红假单胞菌(*Rhodopseudomonas palustris*)、小红卵菌属物种(*Rhodovulum sp.*)、米氏西蒙斯菌(*Simonsiella muelleri*)、鞘氨醇单胞菌属物种(*Sphingomonas sp.*)、葡萄园芽孢乳杆菌(*Sporolactobacillus vineae*)、金黄色葡萄球菌、路邓葡萄球菌(*Staphylococcus lugdunensis*)、链球菌属物种(*Streptococcus*

sp.)、Subdoligranulum sp.、运动替斯崔纳菌(*Tistrella mobilis*)、密螺旋体属物种(*Treponema* sp.)或*Verminephrobacter eiseniae*。Cas9分子的例子可以包括例如以下文献中所述的那些:WO 2015/161276、WO2017/193107、WO 2017/093969、US 2016/272999和US 2015/056705。

[0269] 如该术语在本文所用,Cas9分子或Cas9多肽是指可以与gRNA分子相互作用并且与gRNA分子一齐归巢或定位至包含靶结构域和PAM序列的位点的分子或多肽。如那些术语在本文中所示,Cas9分子和Cas9多肽是指天然存在的Cas9分子,并且是指与参考序列(例如,最相似天然存在的Cas9分子)相差例如至少一个氨基酸残基的工程化的、改变的或修饰的Cas9分子或Cas9多肽。

[0270] 已经确定了两种不同的天然存在的细菌Cas9分子(Jinek等人,*Science*, 343(6176):1247997,2014)和具有指导RNA的酿脓链球菌Cas9(例如,crRNA与tracrRNA的合成融合物)(Nishimasu等人,*Cell*,156:935-949,2014;和Anders等人,*Nature*,2014,doi:10.1038/nature13579)的晶体结构。

[0271] 天然存在的Cas9分子包含两种叶:识别(REC)叶和核酸酶(NUC)叶;其各自还包含本文所述的结构域。一级结构中重要的Cas9结构域的组织示例性示意图描述于WO2015/161276中,例如在其中的图8A-图8B中。在整个本公开文本中使用的结构域命名法和每个结构域所涵盖的氨基酸残基的编号如Nishimasu等人中所述。氨基酸残基的编号参考来自酿脓链球菌的Cas9。

[0272] REC叶包含富含精氨酸的桥螺旋(BH)、REC1结构域和REC2结构域。REC叶与其他已知蛋白质没有结构相似性,表明它是Cas9特有的功能结构域。BH结构域是富含 α -螺旋和精氨酸的长区域,并且包含酿脓链球菌Cas9的序列的氨基酸60-93。REC1结构域对于识别例如gRNA或tracrRNA的重复:抗重复双链体是重要的,因此通过识别靶序列而对Cas9活性至关重要。REC1结构域在酿脓链球菌Cas9的序列的氨基酸94至179和308至717处包含两个REC1基序。这两个REC1结构域虽然在线性一级结构中由REC2结构域隔开,但是在三级结构中组装形成REC1结构域。REC2结构域或其部分也可能在重复:抗重复双链体的识别中起作用。REC2结构域包含酿脓链球菌Cas9的序列的氨基酸180-307。

[0273] NUC叶包含RuvC结构域(在本文中也称为RuvC样结构域)、HNH结构域(在本文中也称为HNH样结构域)和PAM相互作用(PI)结构域。RuvC结构域与逆转录病毒整合酶超家族成员共有结构相似性,并且切割单链,例如靶核酸分子的非互补链。RuvC结构域由分别在酿脓链球菌Cas9的序列的氨基酸1-59、718-769和909-1098处的三个分离的RuvC基序(RuvC I、RuvCII和RuvCIII,它们通常被称为RuvCI结构域或N末端RuvC结构域、RuvCII结构域和RuvCIII结构域)组装而成。与REC1结构域相似,三个RuvC基序在一级结构中由其他结构域线性隔开,然而在三级结构中,三个RuvC基序组装并形成RuvC结构域。HNH结构域与HNH内切核酸酶共有结构相似性,并且切割单链,例如靶核酸分子的互补链。HNH结构域位于RuvCII-III基序之间,并且包含酿脓链球菌Cas9的序列的氨基酸775-908。PI结构域与靶核酸分子的PAM相互作用,并且包含酿脓链球菌Cas9的序列的氨基酸1099-1368。

(a) RuvC样结构域和HNH样结构域

[0274] 在一些实施方案中,Cas9分子或Cas9多肽包含HNH样结构域和RuvC样结构域。在一些实施方案中,切割活性依赖于RuvC样结构域和HNH样结构域。Cas9分子或Cas9多肽(例如,

eaCas9分子或eaCas9多肽)可以包含以下结构域中的一种或多种:RuvC样结构域和HNH样结构域。在一些实施方案中,Cas9分子或Cas9多肽是eaCas9分子或eaCas9多肽,并且eaCas9分子或eaCas9多肽包含RuvC样结构域(例如,本文所述的RuvC样结构域)和/或HNH样结构域(例如,本文所述的HNH样结构域)。

(b) RuvC样结构域

[0275] 在一些实施方案中,RuvC样结构域切割单链,例如靶核酸分子的非互补链。Cas9分子或Cas9多肽可以包括超过一个RuvC样结构域(例如,一个、两个、三个或更多个RuvC样结构域)。在一些实施方案中,RuvC样结构域具有至少5、6、7、8个氨基酸的长度,但是长度不超过20、19、18、17、16或15个氨基酸。在一些实施方案中,Cas9分子或Cas9多肽包含约10至20个氨基酸(例如,约15个氨基酸)的长度的N末端RuvC样结构域。

(c) N末端RuvC样结构域

[0276] 一些天然存在的Cas9分子包含超过一个RuvC样结构域,且切割依赖于N末端RuvC样结构域。因此,Cas9分子或Cas9多肽可以包含N末端RuvC样结构域。

[0277] 在实施方案中,N末端RuvC样结构域具有切割能力。

[0278] 在实施方案中,N末端RuvC样结构域无切割能力。

[0279] 在一些实施方案中,N末端RuvC样结构域与本文(例如,在W02015/161276中,例如,在其中的图3A-图3B或图7A-图7B中)公开的N末端RuvC样结构域的序列相差多达1个但不超过2、3、4或5个残基。在一些实施方案中,存在W02015/161276中(例如,在其中的图3A-图3B或图7A-图7B中)鉴定的1个、2个或所有3个高度保守的残基。

[0280] 在一些实施方案中,N末端RuvC样结构域与本文(例如,在W02015/161276中,例如,在其中的图4A-图4B或图7A-图7B中)公开的N末端RuvC样结构域的序列相差多达1个但不超过2、3、4或5个残基。在一些实施方案中,存在W02015/161276中(例如,在其中的图4A-图4B或图7A-图7B中)鉴定的1个、2个、3个或所有4个高度保守的残基。

(d) 另外的RuvC样结构域

[0281] 除了N末端RuvC样结构域外,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)还可以包含一个或多个另外的RuvC样结构域。在一些实施方案中,Cas9分子或Cas9多肽可以包含两个另外的RuvC样结构域。优选地,另外的RuvC样结构域具有至少5个氨基酸的长度,且例如小于15个氨基酸的长度,例如5至10个氨基酸的长度,例如8个氨基酸的长度。

(e) HNH样结构域

[0282] 在一些实施方案中,HNH样结构域切割单链互补结构域,例如双链核酸分子的互补链。在一些实施方案中,HNH样结构域具有至少15、20、25个氨基酸的长度,但不超过40、35或30个氨基酸的长度,例如20至35个氨基酸的长度,例如25至30个氨基酸的长度。示例性HNH样结构域描述于本文中。

[0283] 在一些实施方案中,HNH样结构域具有切割能力。

[0284] 在一些实施方案中,HNH样结构域无切割能力。

[0285] 在一些实施方案中,HNH样结构域与本文(例如,在W02015/161276中,例如,在其中的图5A-图5C或图7A-图7B中)公开的HNH样结构域的序列相差多达1个但不超过2、3、4或5个残基。在一些实施方案中,存在W02015/161276中(例如,在其中的图5A-图5C或图7A-图7B中)鉴定的1个或两个高度保守的残基。

[0286] 在一些实施方案中,HNH样结构域与本文(例如,在W02015/161276中,例如,在其中的图6A-图6B或图7A-图7B中)公开的HNH样结构域的序列相差多达1个但不超过2、3、4或5个残基。在一些实施方案中,存在W02015/161276中(例如,在其中的图6A-图6B或图7A-图7B中)鉴定的1个、2个、所有3个高度保守的残基。

(f) 核酸酶和解旋酶活性

[0287] 在一些实施方案中,Cas9分子或Cas9多肽能够切割靶核酸分子。通常,野生型Cas9分子切割靶核酸分子的两条链。Cas9分子和Cas9多肽可以被工程化以改变核酸酶切割(或其他特性),例如以提供作为切口酶或缺乏切割靶核酸的能力的Cas9分子或Cas9多肽。能够切割靶核酸分子的Cas9分子或Cas9多肽在本文中称为eaCas9分子或eaCas9多肽。

[0288] 在一些实施方案中,eaCas9分子或eaCas9多肽包含以下活性中的一种或多种:切口酶活性,即切割核酸分子的单链(例如,非互补链或互补链)的能力;双链核酸酶活性,即切割双链核酸的两条链并产生双链断裂的能力,这在一些实施方案中是两种切口酶活性的存在;内切核酸酶活性;外切核酸酶活性;以及解旋酶活性,即解开双链核酸的螺旋结构的能力。

[0289] 在一些实施方案中,酶活性或eaCas9分子或eaCas9多肽切割两条链并导致双链断裂。在一些实施方案中,eaCas9分子仅切割一条链,例如与gRNA杂交的链,或者与gRNA的杂交链互补的链。在一些实施方案中,eaCas9分子或eaCas9多肽包含与HNH样结构域相关的切割活性。在一些实施方案中,eaCas9分子或eaCas9多肽包含与N末端RuvC样结构域相关的切割活性。在一些实施方案中,eaCas9分子或eaCas9多肽包含与HNH样结构域相关的切割活性和与N末端RuvC样结构域相关的切割活性。在一些实施方案中,eaCas9分子或eaCas9多肽包含有活性的或有切割能力的HNH样结构域和无活性的或无切割能力的N末端RuvC样结构域。在一些实施方案中,eaCas9分子或eaCas9多肽包含无活性的或无切割能力的HNH样结构域和有活性的或有切割能力的N末端RuvC样结构域。

[0290] 一些Cas9分子或Cas9多肽具有与gRNA分子相互作用的能力,并且与gRNA分子结合定位到核心靶结构域,但是不能切割靶核酸,或者不能以有效速率切割。没有切割活性或没有实质性切割活性的Cas9分子在本文中被称为eiCas9分子或eiCas9多肽。例如,eiCas9分子或eiCas9多肽可能缺乏切割活性,或者具有参考Cas9分子或eiCas9多肽的显著较低(例如,低于20%、10%、5%、1%或0.1%)的切割活性,如通过本文所述的测定所测量的。

(g) 靶向和PAM

[0291] Cas9分子或Cas9多肽是可以与指导RNA(gRNA)分子相互作用并且与gRNA分子一齐定位到包含靶结构域和PAM序列的位点的多肽。

[0292] 在一些实施方案中,eaCas9分子或eaCas9多肽与靶核酸相互作用并切割靶核酸的能力是PAM序列依赖性的。PAM序列是靶核酸中的序列。在一些实施方案中,靶核酸的切割发生于PAM序列的上游。来自不同细菌物种的eaCas9分子可以识别不同的序列基序(例如,PAM序列)。在一些实施方案中,酿脓链球菌的eaCas9分子识别序列基序NGG、NAG、NGA并且引导自该序列上游1至10个(例如,3至5个)碱基对的靶核酸序列的切割。参见例如,Mali等人,Science 2013;339(6121):823-826。在一些实施方案中,唾液链球菌的eaCas9分子识别序列基序NGGNG和/或NNAGAAW(W=A或T)并且引导自这些序列上游1至10个(例如,3至5个)碱基对的靶核酸序列的切割。参见例如,Horvath等人,Science 2010;327(5962):167-170;以

及Deveau等人, *J Bacteriol* 2008;190(4):1390-1400。在一些实施方案中,变形链球菌(*S.mutans*)的eaCas9分子识别序列基序NGG和/或NAAR (R=A或G)并且引导自该序列上游1至10个(例如,3至5个)碱基对的核心靶核酸序列的切割。参见例如,Deveau等人, *J Bacteriol* 2008;190(4):1390-1400。在一些实施方案中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRR (R=A或G)并且引导自该序列上游1至10个(例如,3至5个)碱基对的靶核酸序列的切割。在一些实施方案中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRRT (R=A或G)并且引导自该序列上游1至10个(例如,3至5个)碱基对的靶核酸序列的切割。在一些实施方案中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRRV (R=A或G)并且引导自该序列上游1至10个(例如,3至5个)碱基对的靶核酸序列的切割。在一些实施方案中,脑膜炎奈瑟球菌的eaCas9分子识别序列基序NNNGATT或NNNGCTT (R=A或G, V=A、G或C)并且引导自该序列上游1至10个(例如,3至5个)碱基对的靶核酸序列的切割。参见例如,Hou等人, *PNAS Early Edition* 2013,1-6。可以确定Cas9分子识别PAM序列的能力,例如使用Jinek等人, *Science*2012 337:816中所述的转化测定来确定。在前述实施方案中,N可以是任何核苷酸残基,例如A、G、C或T中的任一个。

[0293] 如本文所讨论的,Cas9分子可以被工程化以改变Cas9分子的PAM特异性。

[0294] 示例性的天然存在的Cas9分子描述于Chylinski等人, *RNA Biology* 2013 10:5, 727-737中。此类Cas9分子包括簇1-78细菌家族的Cas9分子。

[0295] 示例性的天然存在的Cas9分子包括簇1细菌家族的Cas9分子。例子包括以下的Cas9分子:酿脓链球菌(例如,菌株SF370、MGAS10270、MGAS10750、MGAS2096、MGAS315、MGAS5005、MGAS6180、MGAS9429、NZ131和SSI-1)、唾液链球菌(例如,菌株LMD-9)、伪猪链球菌(*S.pseudoporcinus*) (例如,菌株SPIN 20026)、变形链球菌(例如,菌株UA159、NN2025)、猕猴链球菌(*S.macacae*) (例如,菌株NCTC11558)、解没食子酸链球菌(*S.galloyticus*) (例如,菌株UCN34、ATCC BAA-2069)、马肠链球菌(*S.equines*) (例如,菌株ATCC 9812、MGCS 124)、停乳链球菌(*S.dysdalactiae*) (例如,菌株GGS 124)、牛链球菌(*S.bovis*) (例如,菌株ATCC 700338)、咽峡炎链球菌(*S.anginosus*) (例如,菌株F0211)、无乳链球菌(*S.agalactiae*) (例如,菌株NEM316、A909)、单核细胞增生李斯特菌(例如,菌株F6854)、无害李斯特菌(*Listeria innocua/L.innocua*,例如菌株Clip11262)、意大利肠球菌(*Enterococcus italicus*) (例如,菌株DSM 15952)或尿肠球菌(*Enterococcus faecium*) (例如,菌株1,231,408)。另一种示例性Cas9分子是脑膜炎奈瑟球菌的Cas9分子(Hou等人, *PNAS Early Edition* 2013,1-6)。

[0296] 在一些实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含如下氨基酸序列:其具有与本文所述的任何Cas9分子序列或天然存在的Cas9分子序列(例如,来自本文列出的物种的Cas9分子(例如,SEQ ID NO:112-115)或者Chylinski等人, *RNA Biology* 2013 10:5,727-737;Hou等人, *PNAS Early Edition* 2013,1-6中所述的Cas9分子)的60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%同源性;在与所述Cas9分子序列相比时相差不超过2%、5%、10%、15%、20%、30%或40%的氨基酸残基;与所述Cas9分子序列相差至少1、2、5、10或20个氨基酸,但是相差不超过100、80、70、60、50、40或30个氨基酸;或者与所述所述Cas9分子序列相同。在一些实施方案中,Cas9分子或Cas9多肽包含以下活性中的一种或多种:切口酶活性;双链切割活性(例如,内切核酸酶和/

或外切核酸酶活性);解旋酶活性;或者与gRNA分子一起归巢至靶核酸的能力。

[0297] 在一些实施方案中,Cas9分子或Cas9多肽包含W02015/161276(例如,在其中的图2A-图2G中)的共有序列的氨基酸序列,其中“*”指示在酿脓链球菌、唾液链球菌、变形链球菌和无害李斯特菌的Cas9分子的氨基酸序列的相应位置中发现的任何氨基酸,并且“-”指示任何氨基酸。在一些实施方案中,Cas9分子或Cas9多肽与SEQ ID NO:112-117的共有序列或W0 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的序列相差至少1个但不超过2、3、4、5、6、7、8、9或10个氨基酸残基。在一些实施方案中,Cas9分子或Cas9多肽包含SEQ ID NO:117的氨基酸序列或如W0 2015/161276中(例如,在其中的图7A-图7B中)所述的氨基酸序列,其中“*”指示在酿脓链球菌或脑膜炎奈瑟球菌的Cas9分子的氨基酸序列的相应位置中发现的任何氨基酸,“-”指示任何氨基酸,并且“-”指示任何氨基酸或不存在。在一些实施方案中,Cas9分子或Cas9多肽与SEQ ID NO:116或117的序列或如W0 2015/161276中(例如,在其中的图7A-图7B中)所述的序列相差至少1个但不超过2、3、4、5、6、7、8、9或10个氨基酸残基。

[0298] 对多种Cas9分子的序列的比较指示,某些区域是保守的。这些区域被标识为:区域1(残基1至180,或者在区域1'的情况下,残基120至180);区域2(残基360至480);区域3(残基660至720);区域4(残基817至900);以及区域5(残基900至960)。

[0299] 在一些实施方案中,Cas9分子或Cas9多肽包含区域1-5,其与足够的另外的Cas9分子序列一起提供生物活性分子,例如具有至少一种本文所述活性的Cas9分子。在一些实施方案中,区域1-6中的每一个独立地与本文所述(例如,SEQ ID NO:112-117所示)的Cas9分子或Cas9多肽或W02015/161276中(例如,来自其中的图2A-图2G或来自图7A-图7B)披露的序列的相应残基具有50%、60%、70%或80%同源性。

[0300] 在一个实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含称为区域1的氨基酸序列,其具有与酿脓链球菌Cas9的氨基酸序列的氨基酸1-180(编号是根据W0 2015/161276的图2A-图2G中的基序序列;W0 2015/161276的图2A-图2G中四个Cas9序列中52%的残基是保守的)的50%、60%、70%、80%、85%、90%、95%、96%、97%、98%或99%同源性;与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸1-180相差至少1、2、5、10或20个氨基酸,但是相差不超过90、80、70、60、50、40或30个氨基酸;或者与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的1-180相同。

[0301] 在一些实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含称为区域1'的氨基酸序列,其具有与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸120-180(W0 2015/161276的图2A-图2G中四个Cas9序列中55%的残基是保守的)的55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%同源性;与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸120-180相差至少1、2或5个氨基酸,但是相差不超过35、30、25、20或10个氨基酸;或者与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的120-180相同。

[0302] 在一些实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含称为区域2的氨基酸序列:其具有与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌

Cas9的氨基酸序列的氨基酸360-480(WO 2015/161276的图2A-图2G中四个Cas9序列中52%的残基是保守的)的50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%同源性;与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸360-480相差至少1、2或5个氨基酸,但是相差不超过35、30、25、20或10个氨基酸;或者与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的360-480相同。

[0303] 在一些实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含称为区域3的氨基酸序列:其具有与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸660-720(WO 2015/161276的图2A-图2G中四个Cas9序列中56%的残基是保守的)的55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%同源性;与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸660-720相差至少1、2或5个氨基酸,但是相差不超过35、30、25、20或10个氨基酸;或者与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的660-720相同。

[0304] 在一个实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含称为区域4的氨基酸序列:其具有与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸817-900(WO 2015/161276的图2A-图2G中四个Cas9序列中55%的残基是保守的)的50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%同源性;与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸817-900相差至少1、2或5个氨基酸,但是相差不超过35、30、25、20或10个氨基酸;或者与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的817-900相同。

[0305] 在一些实施方案中,Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)包含称为区域5的氨基酸序列:其具有与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸900-960(WO 2015/161276的图2A-图2G中四个Cas9序列中60%的残基是保守的)的50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%同源性;与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的氨基酸900-960相差至少1、2或5个氨基酸,但是相差不超过35、30、25、20或10个氨基酸;或者与酿脓链球菌、唾液链球菌、变形链球菌或无害李斯特菌Cas9的氨基酸序列的900-960相同。

(h) 工程化的或改变的Cas9分子和Cas9多肽

[0306] 本文所述的Cas9分子和Cas9多肽(例如,天然存在的Cas9分子)可以具有多种特性中的任一种,包括:切口酶活性;核酸酶活性(例如,内切核酸酶和/或外切核酸酶活性);解旋酶活性;在功能上与gRNA分子缔合的能力;以及靶向(或定位到)核酸上的位点的能力(例如,PAM识别和特异性)。在一些实施方案中,Cas9分子或Cas9多肽可以包括这些特性的全部或子集。在典型的实施方案中,Cas9分子或Cas9多肽具有与gRNA分子相互作用以及与gRNA分子一齐定位到核酸中的位点的能力。其他活性(例如,PAM特异性、切割活性或解旋酶活性)在Cas9分子和Cas9多肽中可能变化更广泛。

[0307] Cas9分子包括工程化的Cas9分子和工程化的Cas9多肽(如在此上下文中所用,“工

程化”仅意味着Cas9分子或Cas9多肽与参考序列不同,并且不暗示过程或来源限制)。工程化的Cas9分子或Cas9多肽可以包含改变的酶学特性,例如改变的核酸酶活性(如与天然存在的或其他参考Cas9分子相比)或改变的解旋酶活性。如本文所讨论,工程化的Cas9分子或Cas9多肽可以具有切口酶活性(如与双链核酸酶活性相反)。在一些实施方案中,工程化的Cas9分子或Cas9多肽可以具有改变其尺寸的改变,例如减小其尺寸的氨基酸序列的缺失,例如在不显著影响一种或多种或任何Cas9活性的情况下。在一些实施方案中,工程化的Cas9分子或Cas9多肽可以包含影响PAM识别的改变。例如,可以改变工程化的Cas9分子以识别除了内源野生型PI结构域所识别PAM序列以外的PAM序列。在一些实施方案中,Cas9分子或Cas9多肽的序列可以与天然存在的Cas9分子不同,但是一种或多种Cas9活性不具有显著改变。

[0308] 具有所需特性的Cas9分子或Cas9多肽可以以多种方式制备,例如通过改变亲代(例如,天然存在的)Cas9分子或Cas9多肽,以提供具有所需特性的改变的Cas9分子或Cas9多肽。例如,可以引入相对于亲代Cas9分子(例如,天然存在的或工程化的Cas9分子)的一个或多个突变或差异。此类突变和差异包含:取代(例如,保守取代或非必需氨基酸的取代);插入;或缺失。在一些实施方案中,相对于参考(例如,亲代)Cas9分子,Cas9分子或Cas9多肽可以包含一个或多个突变或差异,例如至少1、2、3、4、5、10、15、20、30、40或50个突变,但少于200、100或80个突变。

[0309] 在一些实施方案中,一个或多个突变对Cas9活性(例如,本文所述的Cas9活性)不具有实质性影响。在一些实施方案中,一个或多个突变对Cas9活性(例如,本文所述的Cas9活性)具有实质性影响。

(i) 非切割和修饰的切割Cas9分子和Cas9多肽

[0310] 在一些实施方案中,Cas9分子或Cas9多肽包含切割特性,所述切割特性与天然存在的Cas9分子不同,例如与具有最接近的同源性的天然存在的Cas9分子不同。例如,Cas9分子或Cas9多肽可以与天然存在的Cas9分子(例如,酿脓链球菌的Cas9分子)不同,如下:其调节(例如,减少或增加)双链核酸切割的能力(内切核酸酶和/或外切核酸酶活性),例如如与天然存在的Cas9分子(例如,酿脓链球菌的Cas9分子)相比;其调节(例如,减少或增加)核酸单链(例如,核酸分子的非互补链或核酸分子的互补链)的切割的能力(切口酶活性),例如如与天然存在的Cas9分子(例如,酿脓链球菌的Cas9分子)相比;或者切割核酸分子(例如,双链或单链核酸分子)的能力可能被消除。

(j) 修饰的切割eaCas9分子和eaCas9多肽

[0311] 在一些实施方案中,eaCas9分子或eaCas9多肽包含以下活性中的一种或多种:与N末端RuvC样结构域相关的切割活性;与HNH样结构域相关的切割活性;与HNH样结构域相关的切割活性和与N末端RuvC样结构域相关的切割活性。

[0312] 在一些实施方案中,eaCas9分子或eaCas9多肽包含有活性的或有切割能力的HNH样结构域和无活性的或无切割能力的N末端RuvC样结构域。示例性无活性的或无切割能力的N末端RuvC样结构域可以具有N末端RuvC样结构域中的天冬氨酸(例如,在SEQ ID NO: 112-117的共有序列或WO2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置9的天冬氨酸,或者在SEQ ID NO:117的位置10的天冬氨酸)的突变,例如可以被丙氨酸取代。在一些实施方案中,eaCas9分子或eaCas9多肽在N末端RuvC样结构域中与野生型不

同,并且不切割靶核酸或者以显著较低的效率(例如,低于参考Cas9分子的切割活性的20%、10%、5%、1%或.1%)切割,例如如通过本文所述的测定所测量的。参考Cas9分子可以是天然存在的未修饰的Cas9分子,例如天然存在的Cas9分子,如酿脓链球菌或唾液链球菌的Cas9分子。在一些实施方案中,参考Cas9分子是具有最接近的序列同一性或同源性的天然存在的Cas9分子。

[0313] 在一些实施方案中,eaCas9分子或eaCas9多肽包含无活性的或无切割能力的HNH结构域和有活性的或有切割能力的N末端RuvC样结构域。示例性无活性的或无切割能力的HNH样结构域可以在以下中的一处或多处具有突变:HNH样结构域中的组氨酸,例如显示于SEQ ID NO:112-117的共有序列或WO 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置856的组氨酸,例如可以被丙氨酸取代;以及HNH样结构域中的一个或多个天冬酰胺,例如显示于SEQ ID NO:112-117的共有序列或WO2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置870的天冬酰胺,和/或显示于SEQ ID NO:112-117的共有序列或WO 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置879的天冬酰胺,例如可以被丙氨酸取代。在一些实施方案中,eaCas9在HNH样结构域中与野生型不同,并且不切割靶核酸,或以显著较低的效率(例如,低于参考Cas9分子的切割活性的20%、10%、5%、1%或0.1%)切割,例如如通过本文所述的测定所测量的。参考Cas9分子可以是天然存在的未修饰的Cas9分子,例如天然存在的Cas9分子,如酿脓链球菌或唾液链球菌的Cas9分子。在一些实施方案中,参考Cas9分子是具有最接近的序列同一性或同源性的天然存在的Cas9分子。

[0314] 在一些实施方案中,eaCas9分子或eaCas9多肽包含无活性的或无切割能力的HNH结构域和有活性的或有切割能力的N末端RuvC样结构域。示例性无活性的或无切割能力的HNH样结构域可以在以下中的一处或多处具有突变:HNH样结构域中的组氨酸,例如显示于SEQ ID NO:112-117的共有序列或WO 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置856的组氨酸,例如可以被丙氨酸取代;以及HNH样结构域中的一个或多个天冬酰胺,例如显示于SEQ ID NO:112-117的共有序列或WO2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置870的天冬酰胺,和/或显示于SEQ ID NO:112-117的共有序列或WO 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列的位置879的天冬酰胺,例如可以被丙氨酸取代。在一些实施方案中,eaCas9在HNH样结构域中与野生型不同,并且不切割靶核酸,或以显著较低的效率(例如,低于参考Cas9分子的切割活性的20%、10%、5%、1%或0.1%)切割,例如如通过本文所述的测定所测量的。参考Cas9分子可以是天然存在的未修饰的Cas9分子,例如天然存在的Cas9分子,如酿脓链球菌或唾液链球菌的Cas9分子。在一些实施方案中,参考Cas9分子是具有最接近的序列同一性或同源性的天然存在的Cas9分子。

(k) 切割靶核酸的一条或两条链的能力的改变

[0315] 在一些实施方案中,示例性Cas9活性包括PAM特异性、切割活性和解旋酶活性中的一种或多种。一个或多个突变可以存在于例如:一个或多个RuvC样结构域,例如N末端RuvC样结构域;HNH样结构域;RuvC样结构域和HNH样结构域以外的区域。在一些实施方案中,一个或多个突变存在于RuvC样结构域(例如,N末端RuvC样结构域)中。在一些实施方案中,一个或多个突变存在于HNH样结构域中。在一些实施方案中,突变存在于RuvC样结构域(例如,

N末端RuvC样结构域)和HNH样结构域二者中。

[0316] 参考酿脓链球菌序列,可以在RuvC结构域或HNH结构域中进行的示例性突变包括:D10A、E762A、H840A、N854A、N863A和/或D986A。

[0317] 在一些实施方案中,Cas9分子或Cas9多肽是如与参考Cas9分子相比在RuvC结构域中和/或在HNH结构域中包含一个或多个差异的eiCas9分子或eiCas9多肽,并且所述eiCas9分子或eiCas9多肽不切割核酸,或以显著低于野生型的效率切割,例如,当在例如如本文所述的切割测定中与野生型相比时,以低于参考Cas9分子的50%、25%、10%或1%的效率切割,如通过本文所述的测定所测量的。

[0318] 特定序列(例如,取代)是否可以影响一种或多种活性(如靶向活性、切割活性等)可以例如通过评价突变是否是保守性的来评价或预测。在一些实施方案中,如在Cas9分子的情况下使用的“非必需”氨基酸残基是如下残基,其可以从Cas9分子(例如,天然存在的Cas9分子,例如eaCas9分子)的野生型序列改变,而不消除或更优选地不显著改变Cas9活性(例如,切割活性),而改变“必需”氨基酸残基导致活性(例如,切割活性)的实质性丧失。

[0319] 在一些实施方案中,Cas9分子或Cas9多肽包含与天然存在的Cas9分子不同,例如与具有最接近的同源性的天然存在的Cas9分子不同的切割特性。例如,Cas9分子或Cas9多肽可以与天然存在的Cas9分子(例如,金黄色葡萄球菌、酿脓链球菌或空肠弯曲菌(*C. jejuni*)的Cas9分子)不同,如下:其调节(例如,减少或增加)双链断裂切割的能力(内切核酸酶和/或外切核酸酶活性),例如如与天然存在的Cas9分子(例如,金黄色葡萄球菌、酿脓链球菌或空肠弯曲菌的Cas9分子)相比;其调节(例如,减少或增加)核酸单链(例如,核酸分子的非互补链或核酸分子的互补链)切割的能力(切口酶活性),例如如与天然存在的Cas9分子(例如,金黄色葡萄球菌、酿脓链球菌或空肠弯曲菌的Cas9分子)相比;或者切割核酸分子(例如,双链或单链核酸分子)的能力可能被消除。

[0320] 在一些实施方案中,改变的Cas9分子或Cas9多肽是包含以下活性中的一种或多种的eaCas9分子或eaCas9多肽:与RuvC结构域相关的切割活性;与HNH结构域相关的切割活性;与HNH结构域相关的切割活性和与RuvC结构域相关的切割活性。

[0321] 在一些实施方案中,改变的Cas9分子或Cas9多肽是如下eiCas9分子或eaCas9多肽,其不切割核酸分子(双链或单链核酸分子)或以显著更低的效率切割核酸分子,例如低于参考Cas9分子的切割活性的20%、10%、5%、1%或0.1%,例如如通过本文所述的测定所测量的。参考Cas9分子可以是天然存在的未修饰的Cas9分子,例如天然存在的Cas9分子,如酿脓链球菌、唾液链球菌、金黄色葡萄球菌、空肠弯曲菌或脑膜炎奈瑟球菌的Cas9分子。在一些实施方案中,参考Cas9分子是具有最接近的序列同一性或同源性的天然存在的Cas9分子。在一些实施方案中,eiCas9分子或eiCas9多肽缺乏与RuvC结构域相关的实质性切割活性和与HNH结构域相关的切割活性。

[0322] 在一些实施方案中,改变的Cas9分子或Cas9多肽是如下eaCas9分子或eaCas9多肽,其包含在WO 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列中显示的酿脓链球菌的固定的氨基酸残基,并且具有在SEQ ID NO:117中的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)或者在WO 2015/161276中(例如,在其中的图2A-图2G中)披露的共有序列中由“-”表示的残基处与酿脓链球菌的氨基酸序列不同(例如,具有取代)的一个或多个氨基酸。

[0323] 在一些实施方案中,改变的Cas9分子或Cas9多肽包含如下序列,其中:对应于W0 2015/161276的图2A-图2G中披露的共有序列的固定序列的序列与W0 2015/161276的图2A-图2G中披露的共有序列中的固定残基相差不超过1%、2%、3%、4%、5%、10%、15%或20%,对应于W0 2015/161276的图2A-图2G中披露的共有序列中由“*”标识的残基的序列与天然存在的Cas9分子(例如,酿脓链球菌Cas9分子)的相应序列中的“*”残基相差不超过1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%或40%;并且对应于W0 2015/161276的图2A-图2G中披露的共有序列中由“-”标识的残基的序列与天然存在的Cas9分子(例如,酿脓链球菌Cas9分子)的相应序列中的“-”残基相差不超过5%、10%、15%、20%、25%、30%、35%、40%、45%、55%或60%。

[0324] 在一些实施方案中,改变的Cas9分子或Cas9多肽是包含W0 2015/161276的图2A-图2G中披露的共有序列中所示唾液链球菌的固定氨基酸残基的eaCas9分子或eaCas9多肽,并且在W0 2015/161276的图2A-图2G中披露的共有序列中由“-”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有与唾液链球菌的氨基酸序列不同的一个或多个氨基酸(例如,具有取代)。

[0325] 在一些实施方案中,改变的Cas9分子或Cas9多肽包含如下序列,其中:对应于W0 2015/161276的图2A-图2G中披露的共有序列的固定序列的序列与W0 2015/161276的图2A-图2G中披露的共有序列中的固定残基相差不超过1%、2%、3%、4%、5%、10%、15%或20%,对应于W0 2015/161276的图2A-图2G中披露的共有序列中由“*”标识的残基的序列与天然存在的Cas9分子(例如,唾液链球菌Cas9分子)的相应序列中的“*”残基相差不超过1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%或40%;并且对应于W0 2015/161276的图2A-图2G中披露的共有序列中由“-”标识的残基的序列与天然存在的Cas9分子(例如,唾液链球菌Cas9分子)的相应序列中的“-”残基相差不超过5%、10%、15%、20%、25%、30%、35%、40%、45%、55%或60%。

[0326] 在一些实施方案中,改变的Cas9分子或Cas9多肽是包含W0 2015/161276的图2A-图2G中披露的共有序列中所示的变形链球菌的固定氨基酸残基的eaCas9分子或eaCas9多肽,并且在W0 2015/161276的图2A-图2G中披露的共有序列中由“-”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有与变形链球菌的氨基酸序列不同的一个或多个氨基酸(例如,具有取代)。

[0327] 在一些实施方案中,改变的Cas9分子或Cas9多肽包含如下序列,其中:对应于W0 2015/161276的图2A-图2G中披露的共有序列的固定序列的序列与W0 2015/161276的图2A-图2G中披露的共有序列中的固定残基相差不超过1%、2%、3%、4%、5%、10%、15%或20%,对应于W0 2015/161276的图2A-图2G中披露的共有序列中由“*”标识的残基的序列与天然存在的Cas9分子(例如,变形链球菌Cas9分子)的相应序列中的“*”残基相差不超过1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%或40%;并且,对应于W0 2015/161276的图2A-图2G中披露的共有序列中由“-”标识的残基的序列与天然存在的Cas9分子(例如,变形链球菌Cas9分子)的相应序列中的“-”残基相差不超过5%、10%、15%、20%、25%、30%、35%、40%、45%、55%或60%。

[0328] 在一些实施方案中,改变的Cas9分子或Cas9多肽是包含W0 2015/161276的图2A-图2G中披露的共有序列中所示的无害李斯特菌的固定氨基酸残基的eaCas9分子或eaCas9多

肽,并且在WO 2015/161276的图2A-图2G中披露的共有序列中由“-”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有与无害李斯特菌的氨基酸序列不同的一个或多个氨基酸(例如,具有取代)。在一些实施方案中,改变的Cas9分子或Cas9多肽包含如下序列,其中:对应于WO2015/161276的图2A-图2G中披露的共有序列的固定序列的序列与WO 2015/161276的图2A-图2G中披露的共有序列中的固定残基相差不超过1%、2%、3%、4%、5%、10%、15%或20%,对应于WO 2015/161276的图2A-图2G中披露的共有序列中由“*”标识的残基的序列与天然存在的Cas9分子(例如,无害李斯特菌Cas9分子)的相应序列中的“*”残基相差不超过1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%或40%;并且,对应于WO 2015/161276的图2A-图2G中披露的共有序列中由“-”标识的残基的序列与天然存在的Cas9分子(例如,无害李斯特菌Cas9分子)的相应序列中的“-”残基相差不超过5%、10%、15%、20%、25%、30%、35%、40%、45%、55%或60%。

[0329] 在一些实施方案中,改变的Cas9分子或Cas9多肽(例如,eaCas9分子)可以是例如两种或更多种不同Cas9分子或Cas9多肽(例如,不同物种的两种或更多种天然存在的Cas9分子)的融合物。例如,一种物种的天然存在的Cas9分子的片段可以与第二物种的Cas9分子的片段融合。作为例子,可以将包含N末端RuvC样结构域的酿脓链球菌的Cas9分子的片段与包含HNH样结构域的除了酿脓链球菌以外的物种(例如,唾液链球菌)的Cas9分子的片段融合。

(1) 具有改变的PAM识别或无PAM识别的Cas9分子

[0330] 天然存在的Cas9分子可以识别特定PAM序列,例如本文针对例如酿脓链球菌、唾液链球菌、变形链球菌、金黄色葡萄球菌和脑膜炎奈瑟球菌所述的PAM识别序列。

[0331] 在一些实施方案中,Cas9分子或Cas9多肽具有与天然存在的Cas9分子相同的PAM特异性。在其他实施方案中,Cas9分子或Cas9多肽具有与天然存在的Cas9分子无关的PAM特异性,或者与其具有最接近的序列同源性的天然存在的Cas9分子无关的PAM特异性。例如,可以改变天然存在的Cas9分子,例如以改变PAM识别,例如以改变Cas9分子或Cas9多肽识别的PAM序列,以减少脱靶位点和/或改进特异性;或消除PAM识别要求。在一些实施方案中,可以改变Cas9分子,例如以增加PAM识别序列的长度和/或改进Cas9特异性至高同一性水平,例如以减少脱靶位点并增加特异性。在一些实施方案中,PAM识别序列的长度为至少4、5、6、7、8、9、10或15个氨基酸的长度。

[0332] 可以使用定向进化来产生识别不同PAM序列和/或具有降低的脱靶活性的Cas9分子或Cas9多肽。可以用于Cas9分子的定向进化的示例性方法和系统描述于例如Esvelt等人Nature 2011,472(7344):499-503中。例如可以通过本文所述的方法来评价候选Cas9分子。

[0333] 在本文中讨论了介导PAM识别的PI结构域的改变。

(m) 具有改变的PI结构域的合成Cas9分子和Cas9多肽

[0334] 当前基因组编辑方法受限于可以被由所用Cas9分子识别的PAM序列靶向的靶序列的多样性。如该术语在本文中所用,合成Cas9分子(或Syn-Cas9分子)或合成Cas9多肽(或Syn-Cas9多肽)是指如下Cas9分子或Cas9多肽,其包含来自一个细菌物种的Cas9核心结构域和例如来自不同的细菌物种的功能改变的PI结构域(即,除了与Cas9核心结构域天然相关的PI结构域以外的PI结构域)。

[0335] 在一些实施方案中,改变的PI结构域识别的PAM序列与衍生出Cas9核心结构域的

天然存在的Cas9识别的PAM序列不同。在一些实施方案中,改变的PI结构域识别衍生出Cas9核心结构域的天然存在的Cas9所识别的相同,但是具有不同的亲和力或特异性的PAM序列。Syn-Cas9分子或Syn-Cas9多肽可以分别是Syn-eaCas9分子或Syn-eaCas9多肽或者Syn-eiCas9分子或Syn-eiCas9多肽。

[0336] 示例性Syn-Cas9分子或Syn-Cas9多肽包含:a) Cas9核心结构域,例如Cas9核心结构域,例如金黄色葡萄球菌、酿脓链球菌或空肠弯曲菌Cas9核心结构域;以及b) 来自物种X Cas9序列的改变的PI结构域。

[0337] 在一些实施方案中,所述改变的PI结构域的RKR基序(PAM结合基序)包含:在1、2或3个氨基酸残基处的差异;在氨基酸序列中在第一、第二或第三位置处的差异;在氨基酸序列中在第一和第二位置、第一和第三位置或第二和第三位置处的差异;如相比于与Cas9核心结构域相关的天然或内源PI结构域的RKR基序的序列。

[0338] 在一些实施方案中,Syn-Cas9分子或Syn-Cas9多肽还可以进行尺寸优化,例如,Syn-Cas9分子或Syn-Cas9多肽包含一个或多个缺失以及任选地布置在侧接所述缺失的氨基酸残基之间的一个或多个接头。在一些实施方案中,Syn-Cas9分子或Syn-Cas9多肽包含REC缺失。

(n) 尺寸优化的Cas9分子和Cas9多肽

[0339] 本文所述的工程化的Cas9分子和工程化的Cas9多肽包括如下Cas9分子或Cas9多肽,其包含减小分子尺寸的缺失,同时仍保留所需的Cas9特性,例如基本上天然的构象、Cas9核酸酶活性和/或靶核酸分子识别。在所提供的实施方案的情况下使用的Cas9分子或Cas9多肽可以包含一个或多个缺失以及任选地一个或多个接头,其中接头布置在侧接所述缺失的氨基酸残基之间。

[0340] 具有缺失的Cas9分子(例如,金黄色葡萄球菌、酿脓链球菌或空肠弯曲菌Cas9分子)小于(例如,具有减少的氨基酸数量)相应的天然存在的Cas9分子。Cas9分子的较小尺寸允许增加递送方法的灵活性,从而增加用于基因组编辑的实用性。Cas9分子或Cas9多肽可以包含不显著影响或降低本文所述的所得Cas9分子或Cas9多肽的活性的一个或多个缺失。在如本文所述包含缺失的Cas9分子或Cas9多肽中保留的活性包括以下中的一种或多种:切口酶活性,即切割核酸分子的单链(例如,非互补链或互补链)的能力;双链核酸酶活性,即切割双链核酸的两条链并产生双链断裂的能力,这在一些实施方案中是两种切口酶活性的存在;内切核酸酶活性;外切核酸酶活性;解旋酶活性,即解开双链核酸的螺旋结构的能力;以及核酸分子(例如,靶核酸或gRNA)的识别活性。

[0341] 本文所述的Cas9分子或Cas9多肽的活性可以使用本文所述或者已知的活性测定来评估。

(o) 鉴定适合于缺失的区域

[0342] Cas9分子中适合于缺失的区域可以通过多种方法来鉴定。可以将来自各个细菌物种的天然存在的直系同源Cas9分子建模至酿脓链球菌Cas9的晶体结构上(Nishimasu等人,Cell,156:935-949,2014),以检查整个所选Cas9直系同源物中关于蛋白质的三维构象的保守水平。空间定位远离参与Cas9活性的区域(例如,与靶核酸分子和/或gRNA相互作用)的较不保守或不保守的区域表示在不显著影响或降低Cas9活性的情况下的缺失的候选区域或结构域。

(p) REC优化的Cas9分子和Cas9多肽

[0343] 如该术语在本文中所示,REC优化的Cas9分子或REC优化的Cas9多肽是指如下Cas9分子或Cas9多肽,其在REC2结构域和REC1_{CT}结构域中的一个或两个中包含缺失(统称为REC缺失),其中所述缺失包含同源结构域中至少10%的氨基酸残基。REC优化的Cas9分子或Cas9多肽可以是eaCas9分子或eaCas9多肽或者eiCas9分子或eiCas9多肽。示例性REC优化的Cas9分子或REC优化的Cas9多肽包含:a)选自以下的缺失:i) REC2缺失;ii) REC1_{CT}缺失;或iii) REC1_{SUB}缺失。

[0344] 任选地,接头布置在侧接缺失的氨基酸残基之间。在一些实施方案中,Cas9分子或Cas9多肽包括仅一个缺失,或者仅两个缺失。Cas9分子或Cas9多肽可以包含REC2缺失和REC1_{CT}缺失。Cas9分子或Cas9多肽可以包含REC2缺失和REC1_{SUB}缺失。

[0345] 通常,缺失将含有同源结构域中至少10%的氨基酸,例如,REC2缺失将包括REC2结构域中至少10%的氨基酸。缺失可以包含:其同源结构域的至少10%、20%、30%、40%、50%、60%、70%、80%或90%的氨基酸残基;其同源结构域的所有氨基酸残基;其同源结构域以外的氨基酸残基;其同源结构域以外的多个氨基酸残基;紧接其同源结构域N末端的氨基酸残基;紧接其同源结构域C末端的氨基酸残基;紧接其同源结构域N末端的氨基酸残基和紧接其同源结构域C末端的氨基酸残基;其同源结构域N末端的多个(例如,多达5、10、15或20个)氨基酸残基;其同源结构域C末端的多个(例如,多达5、10、15或20个)氨基酸残基;其同源结构域N末端的多个(例如,多达5、10、15或20个)氨基酸残基和其同源结构域C末端的多个(例如,多达5、10、15或20个)氨基酸残基。

[0346] 在一些实施方案中,缺失不会延伸超过:其同源结构域;其同源结构域的N末端氨基酸残基;其同源结构域的C末端氨基酸残基。

[0347] REC优化的Cas9分子或REC优化的Cas9多肽可以包括布置在侧接缺失的氨基酸残基之间的接头。适用于侧接REC优化的Cas9分子中的REC缺失的氨基酸残基之间的接头描述于本文中。

[0348] 在一些实施方案中,REC优化的Cas9分子或REC优化的Cas9多肽包含如下氨基酸序列,除了任何REC缺失和相关接头以外,其与天然存在的Cas9(例如,金黄色葡萄球菌Cas9分子、酿脓链球菌Cas9分子或空肠弯曲菌Cas9分子)的氨基酸序列具有至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、99%或100%同源性。

[0349] 在一些实施方案中,REC优化的Cas9分子或REC优化的Cas9多肽包含如下氨基酸序列,除了任何REC缺失和相关接头以外,其与天然存在的Cas9(例如,金黄色葡萄球菌Cas9分子、酿脓链球菌Cas9分子或空肠弯曲菌Cas9分子)的氨基酸序列相差不超过1、2、3、4、5、6、7、8、9、10、15、20或25个氨基酸残基。

[0350] 在一些实施方案中,REC优化的Cas9分子或REC优化的Cas9多肽包含如下氨基酸序列,除了任何REC缺失和相关接头以外,其与天然存在的Cas9(例如,金黄色葡萄球菌Cas9分子、酿脓链球菌Cas9分子或空肠弯曲菌Cas9分子)的氨基酸序列相差不超过1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%或25%的氨基酸残基。

[0351] 对于序列比较,通常一个序列用作参考序列,将测试序列与其相比。在使用序列比较算法时,将测试和参考序列输入电脑,如果需要则指定子序列坐标,并指定序列算法程序参数。可以使用默认的程序参数,或者可以指定替代的参数。然后序列比较算法基于所述程

序参数来计算测试序列相对于参考序列的序列同一性百分比。用于比较的序列比对方法是熟知的。用于比较的序列的最佳比对可以例如通过以下方法来进行:Smith和Waterman, (1970) *Adv. Appl. Math.* 2:482c的局部同源性算法;Needleman和Wunsch, (1970) *J. Mol. Biol.* 48:443的同源性比对算法;Pearson和Lipman, (1988) *Proc. Nat'l Acad. Sci. USA* 85:2444的相似度搜索方法;这些算法的计算机化实现(Wisconsin Genetics软件包中的GAP、BESTFIT、FASTA和TFASTA, Genetics Computer Group, 575 Science Dr., Madison, 威斯康辛州);或者人工比对和目测检查(参见例如, Brent等人, (2003) *Current Protocols in Molecular Biology*)。

[0352] 适合于确定序列同一性和序列相似度百分比的算法的两个例子是BLAST和BLAST 2.0算法,它们分别描述于Altschul等人, (1977) *Nuc. Acids Res.* 25:3389-3402;和Altschul等人, (1990) *J. Mol. Biol.* 215:403-410。用于执行BLAST分析的软件可通过国家生物技术信息中心(National Center for Biotechnology Information)公开获得。

[0353] 两个氨基酸序列之间的同一性百分比也可以使用E. Meyers和W. Miller, (1988) *Comput. Appl. Biosci.* 4:11-17的算法来确定,所述算法已经并入ALIGN程序(2.0版)中,使用PAM120加权残基表,空位长度罚分为12且空位罚分为4。另外,两个氨基酸序列之间的同一性百分比可以使用Needleman和Wunsch (1970) *J. Mol. Biol.* 48:444-453的算法来确定,所述算法已经并入GCG软件包(可在www.gcg.com获得)的GAP程序中,使用Blossom 62矩阵或PAM250矩阵,并且空位权重为16、14、12、10、8、6或4,并且长度权重为1、2、3、4、5或6。

[0354] 提供了例如国际PCT公开号WO 2015/161276、WO 2017/193107和WO 2017/093969中所述的83种天然存在的Cas9直系同源物的示例性REC缺失的序列信息。

(q) 编码Cas9分子的核酸

[0355] 编码Cas9分子或Cas9多肽(例如, eaCas9分子或eaCas9多肽)的核酸可以与本文提供的任何实施方案结合使用。

[0356] 编码Cas9分子或Cas9多肽的示例性核酸描述于以下文献中: Cong等人, *Science* 2013, 399 (6121): 819-823; Wang等人, *Cell* 2013, 153 (4): 910-918; Mali等人, *Science* 2013, 399 (6121): 823-826; Jinek等人, *Science* 2012, 337 (6096): 816-821; 以及WO2015/161276, 例如在其中的图8中。

[0357] 在一些实施方案中, 编码Cas9分子或Cas9多肽的核酸可以是合成核酸序列。例如, 可以对合成核酸分子进行化学修饰。在一些实施方案中, Cas9 mRNA具有以下特性中的一种或多种(例如, 所有): 其被加帽, 多腺苷酸化, 被5-甲基胞苷和/或假尿苷取代。

[0358] 另外或可替代地, 可以对合成核酸序列进行密码子优化, 例如, 至少一个非常见密码子或较不常见密码子已经被常见密码子替代。例如, 合成核酸可以引导优化的信使mRNA的合成, 例如针对例如本文所述的哺乳动物表达系统中的表达进行优化。

[0359] 另外或可替代地, 编码Cas9分子或Cas9多肽的核酸可以包含核定位序列(NLS)。核定位序列是已知的。

[0360] 在一些实施方案中, Cas9分子由是或包含SEQ ID NO: 121、123或125中任一项的序列或者展现与SEQ ID NO: 121、123或125中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、92%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列编码。在一些实施方案中, Cas9分子是或包含SEQ ID NO: 122、124或125中的任一项或者展现与SEQ ID

N0:122、123或125中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、92%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列。SEQ ID NO:121是编码酿脓链球菌的Cas9分子的示例性的经密码子优化的核酸序列。SEQ ID NO:122是酿脓链球菌Cas9分子的相应氨基酸序列。SEQ ID NO:123是编码脑膜炎奈瑟球菌的Cas9分子的示例性的经密码子优化的核酸序列。SEQ ID NO:124是脑膜炎奈瑟球菌Cas9分子的相应氨基酸序列。SEQ ID NO:125是编码金黄色葡萄球菌Cas9的Cas9分子的示例性的经密码子优化的核酸序列。SEQ ID NO:126是金黄色葡萄球菌Cas9分子的氨基酸序列。

[0361] 如果任何前述Cas9序列在C末端与肽或多肽融合,应理解,将去除终止密码子。

(r) 其他Cas分子和Cas多肽

[0362] 可以使用各种类型的Cas分子或Cas多肽来实践本文公开的发明。在一些实施方案中,使用II型Cas系统的Cas分子。在其他实施方案中,使用其他Cas系统的Cas分子。例如,可以使用I型或III型Cas分子。示例性Cas分子(和Cas系统)描述于例如Haft等人, PLoS Computational Biology 2005,1(6):e60和Makarova等人, Nature Review Microbiology 2011,9:467-477中,将两个参考文献的内容通过引用以其整体并入本文。示例性Cas分子(和Cas系统)也示于表2中。

表2.Cas系统

基因名称 [‡]	系统类型或亚型	来自Haft等人的名称 [§]	所编码蛋白质的结构(PDB登录) [¶]	所编码蛋白质的家族(和超家族) ^{***}	代表
<i>cas1</i>	• I型 • II型 • III型	<i>cas1</i>	3GOD、 3LFX 和 2YZS	COG1518	SERP2463、 SPy1047 和 <i>ygbT</i>
<i>cas2</i>	• I型 • II型 • III型	<i>cas2</i>	2IVY、 2I8E 和 3EXC	COG1343 和 COG3512	SERP2462、 SPy1048、 SPy1723(N 末端结构域)和 <i>ygbF</i>
<i>cas3'</i>	• I型 ^{**}	<i>cas3</i>	NA	COG1203	APE1232 和 <i>ygcB</i>
<i>cas3''</i>	• I-A亚型 • I-B亚型	NA	NA	COG2254	APE1231 和 BH0336
<i>cas4</i>	• I-A亚型 • I-B亚型 • I-C亚型 • I-D亚型 • II-B亚型	<i>cas4</i> 和 <i>csa1</i>	NA	COG1468	APE1239 和 BH0340

基因名称 [‡]	系统类型或亚型	来自Haft等人的名称 [§]	所编码蛋白质的结构(PDB登录) [¶]	所编码蛋白质的家族(和超家族) ^{***}	代表
<i>cas5</i>	• I-A亚型 • I-B亚型 • I-C亚型 • I-E亚型	<i>cas5a</i> 、 <i>cas5d</i> 、 <i>cas5e</i> 、 <i>cas5h</i> 、 <i>cas5p</i> 、 <i>cas5t</i> 和 <i>cmx5</i>	3KG4	COG1688(RAMP)	APE1234、BH0337、 <i>devS</i> 和 <i>ygcI</i>
<i>cas6</i>	• I-A亚型 • I-B亚型 • I-D亚型 • III-A亚型 • III-B亚型	<i>cas6</i> 和 <i>cmx6</i>	3I4H	COG1583和COG5551(RAMP)	PF1131和slr7014
<i>cas6e</i>	• I-E亚型	<i>cse3</i>	1WJ9	(RAMP)	<i>ygcH</i>
<i>cas6f</i>	• I-F亚型	<i>csy4</i>	2XLJ	(RAMP)	y1727
<i>cas7</i>	• I-A亚型 • I-B亚型 • I-C亚型 • I-E亚型	<i>csa2</i> 、 <i>csd2</i> 、 <i>cse4</i> 、 <i>csh2</i> 、 <i>csp1</i> 和 <i>cst2</i>	NA	COG1857和COG3649(RAMP)	<i>devR</i> 和 <i>ygcJ</i>
<i>cas8a1</i>	• I-A亚型 ^{**}	<i>cmx1</i> 、 <i>cst1</i> 、 <i>csx8</i> 、 <i>csx13</i> 和CXXC-CXXC	NA	BH0338样	LA3191 ^{§§} 和PG2018 ^{§§}
<i>cas8a2</i>	• I-A亚型 ^{**}	<i>csa4</i> 和 <i>csx9</i>	NA	PH0918	AF0070、AF1873、MJ0385、PF0637、PH0918和SSO1401
<i>cas8b</i>	• I-B亚型 ^{**}	<i>csh1</i> 和TM1802	NA	BH0338样	MTH1090和TM1802
<i>cas8c</i>	• I-C亚型 ^{**}	<i>csd1</i> 和 <i>csp2</i>	NA	BH0338样	BH0338
<i>cas9</i>	• II型 ^{**}	<i>csn1</i> 和 <i>csx12</i>	NA	COG3513	FTN_0757和SPy1046
<i>cas10</i>	• III型 ^{**}	<i>cmr2</i> 、 <i>csm1</i> 和 <i>csx11</i>	NA	COG1353	MTH326、Rv2823c ^{§§} 和TM1794 ^{§§}
<i>cas10d</i>	• I-D亚型 ^{**}	<i>csc3</i>	NA	COG1353	slr7011
<i>csy1</i>	• I-F亚型 ^{**}	<i>csy1</i>	NA	y1724样	y1724
<i>csy2</i>	• I-F亚型	<i>csy2</i>	NA	(RAMP)	y1725
<i>csy3</i>	• I-F亚型	<i>csy3</i>	NA	(RAMP)	y1726
<i>cse1</i>	• I-E亚型 ^{**}	<i>cse1</i>	NA	YgcL样	<i>ygcL</i>
<i>cse2</i>	• I-E亚型	<i>cse2</i>	2ZCA	YgcK样	<i>ygcK</i>
<i>csc1</i>	• I-D亚型	<i>csc1</i>	NA	alr1563样(RAMP)	alr1563

基因名称 [‡]	系统类型或亚型	来自Haft等人的名称 [§]	所编码蛋白质的结构(PDB登录) [¶]	所编码蛋白质的家族(和超家族) ^{***}	代表
<i>csc2</i>	• I-D亚型	<i>csc1</i> 和 <i>csc2</i>	NA	COG1337(RAMP)	slr7012
<i>csa5</i>	• I-A亚型	<i>csa5</i>	NA	AF1870	AF1870、 MJ0380、 PF0643 和 SSO1398
<i>csn2</i>	• II-A亚型	<i>csn2</i>	NA	SPy1049样	SPy1049
<i>csm2</i>	• III-A亚型 ^{‡‡}	<i>csm2</i>	NA	COG1421	MTH1081 和 SERP2460
<i>csm3</i>	• III-A亚型	<i>csc2</i> 和 <i>csm3</i>	NA	COG1337(RAMP)	MTH1080 和 SERP2459
<i>csm4</i>	• III-A亚型	<i>csm4</i>	NA	COG1567(RAMP)	MTH1079 和 SERP2458
<i>csm5</i>	• III-A亚型	<i>csm5</i>	NA	COG1332(RAMP)	MTH1078 和 SERP2457
<i>csm6</i>	• III-A亚型	APE2256 和 <i>csm6</i>	2WTE	COG1517	APE2256 和 SSO1445
<i>cmr1</i>	• III-B亚型	<i>cmr1</i>	NA	COG1367(RAMP)	PF1130
<i>cmr3</i>	• III-B亚型	<i>cmr3</i>	NA	COG1769(RAMP)	PF1128
<i>cmr4</i>	• III-B亚型	<i>cmr4</i>	NA	COG1336(RAMP)	PF1126
<i>cmr5</i>	• III-B亚型 ^{‡‡}	<i>cmr5</i>	2ZOP 和 2OEB	COG3337	MTH324 和 PF1125
<i>cmr6</i>	• III-B亚型	<i>cmr6</i>	NA	COG1604 (RAMP)	PF1124
<i>csb1</i>	• I-U亚型	GSU0053	NA	(RAMP)	Balac_1306 和 GSU0053
<i>csb2</i>	• I-U亚型 ^{§§}	NA	NA	(RAMP)	Balac_1305 和 GSU0054
<i>csb3</i>	• I-U亚型	NA	NA	(RAMP)	Balac_1303 ^{§§}
<i>csx17</i>	• I-U亚型	NA	NA	NA	Btus_2683
<i>csx14</i>	• I-U亚型	NA	NA	NA	GSU0052
<i>csx10</i>	• I-U亚型	<i>csx10</i>	NA	(RAMP)	Caur_2274
<i>csx16</i>	• III-U亚型	VVA1548	NA	NA	VVA1548
<i>csaX</i>	• III-U亚型	<i>csaX</i>	NA	NA	SSO1438
<i>csx3</i>	• III-U亚型	<i>csx3</i>	NA	NA	AF1864

基因名称 [‡]	系统类型或亚型	来自Haft等人的名称 [§]	所编码蛋白质的结构(PDB登录) [¶]	所编码蛋白质的家族(和超家族) ^{***}	代表
<i>csx1</i>	• III-U亚型	<i>csa3</i> 、 <i>csx1</i> 、 <i>csx2</i> 、DXTHG、NE0113 和 TIGR02710	1XMX 和 2I71	COG1517 和 COG4006	MJ1666、NE0113、PF1127 和 TM1812
<i>csx15</i>	• 未知	NA	NA	TTE2665	TTE2665
<i>csf1</i>	• U型	<i>csf1</i>	NA	NA	AFE_1038
<i>csf2</i>	• U型	<i>csf2</i>	NA	(RAMP)	AFE_1039
<i>csf3</i>	• U型	<i>csf3</i>	NA	(RAMP)	AFE_1040
<i>csf4</i>	• U型	<i>csf4</i>	NA	NA	AFE_1037

(iii) Cpf1

[0363] 在一些实施方案中,指导RNA或gRNA促进RNA指导的核酸酶(如Cas9或Cpf1)与靶序列(如细胞中的基因组或附加体序列)的特异性关联靶向。通常,gRNA可以是单分子的(包含单一RNA分子,并且可替代地称为嵌合的),或者模块化的(包含超过一种、并且通常两种单独RNA分子(如crRNA和tracrRNA),其通常(在一些实施方案中)通过双链体化彼此缔合)。gRNA和其组成部分描述于多个文献中,在一些实施方案中描述于以下文献中:Briner等人(Molecular Cell 56(2),333-339,2014年10月23日(Briner),其是通过引用并入),以及Cotta-Ramusino。

[0364] 无论是单分子的还是模块化的,指导RNA通常都包括与靶标完全或部分互补的靶向结构域,并且通常具有10-30个核苷酸的长度,并且在某些实施方案中具有16-24个核苷酸的长度(在一些实施方案中,16、17、18、19、20、21、22、23或24个核苷酸的长度)。在一些方面,在Cas9 gRNA的情况下,靶向结构域在gRNA的5'末端处或附近,并且在Cpf1 gRNA的情况下,靶向结构域在gRNA的3'末端处或附近。尽管前述描述集中于用于与Cas9一起使用的gRNA,应当了解,已经(或在将来可能)发现或发明其他RNA指导的核酸酶,所述核酸酶利用以某些方式与针对这一点所述的那些不同的gRNA。在一些实施方案中,Cpf1(“来自普雷沃菌属和弗朗西斯菌属的CRISPR 1(CRISPR from *Prevotella* and *Franciscella* 1)”)是最近发现的RNA指导的核酸酶,其不需要tracrRNA起作用。(Zetsche等人,2015,Cell 163,759-771 2015年10月22日(Zetsche I),通过引用并入本文)。用于Cpf1基因组编辑系统的gRNA通常包括靶向结构域和互补结构域(可替代地称为“手柄(handle)”)。还应当注意,在用于与Cpf1一起使用的gRNA中,靶向结构域通常存在于3'端处或附近,而不是如上文结合Cas9 gRNA所述在5'端处或附近(手柄在Cpf1 gRNA的5'端处或附近)。

[0365] 尽管在来自不同原核物种的gRNA之间或在Cpf1与Cas9 gRNA之间可能存在结构差异,但是gRNA的作用原理大体上一致。由于这种作用一致性,在广义上,gRNA可以依据其靶向结构域序列来定义,并且熟练技术人员将了解,可以将给定的靶向结构域序列掺入任何合适的gRNA中,包括单分子或嵌合gRNA,或者包括一种或多种化学修饰和/或序列修饰(取代、另外的核苷酸、截短等)的gRNA。因此,在本公开文本中的一些方面,gRNA可以仅根据其靶向结构域序列来描述。

[0366] 更通常地,本公开文本的一些方面涉及可以使用多种RNA指导的核酸酶实施的系统、方法和组合物。除非另有说明,否则应当将术语gRNA理解为涵盖可以与任何RNA指导的核酸酶一起使用的任何合适的gRNA,而不仅仅是与特定种类的Cas9或Cpf1相容的那些gRNA。通过说明的方式,在某些实施方案中,术语gRNA可以包括用于与存在于2类CRISPR系统(如II型或V型或CRISPR系统)中的任何RNA指导的核酸酶或者从所述核酸酶衍生或修改的RNA指导的核酸酶一起使用的gRNA。

[0367] 此部分中讨论的某些示例性修饰可以被包括在gRNA序列内的任何位置处,包括但不限于在5'端处或附近(例如,在5'端的1-10、1-5或1-2个核苷酸内)和/或在3'端处或附近(例如,在3'端的1-10、1-5或1-2个核苷酸内)。在一些情况下,修饰位于功能性基序内,所述功能性基序如Cas9 gRNA的重复-抗-重复双链体、Cas9或Cpf1 gRNA的茎环结构和/或gRNA的靶向结构域。

[0368] RNA指导的核酸酶包括但不限于天然存在的2类CRISPR核酸酶(如Cas9和Cpf1)以及从所述核酸酶衍生或获得的其他核酸酶。从功能上讲,RNA指导的核酸酶被定义为那些核酸酶,它们:(a)与gRNA相互作用(例如与其形成复合物);并且(b)gRNA一起与DNA的靶区域缔合,并且任选地对其进行切割或修饰,所述靶区域包括(i)与gRNA的靶向结构域互补的序列和任选地(ii)将在下文更详细地描述的称为“原型间隔子邻近基序”或“PAM”的另外的序列。如以下例子将说明,在广义上,RNA指导的核酸酶可以依据其PAM特异性和切割活性来定义,即使共有相同PAM特异性或切割活性的单独的RNA指导的核酸酶之间可能存在差异。熟练技术人员将理解,本公开文本的一些方面涉及可以使用具有某种PAM特异性和/或切割活性的任何合适的RNA指导的核酸酶实施的系统、方法和组合物。因此,除非另有说明,否则术语RNA指导的核酸酶应当理解为通用术语,并且不限于RNA指导的核酸酶的任何特定类型(例如Cas9与Cpf1)、物种(例如酿脓链球菌与金黄色葡萄球菌)或变异(例如全长的与截短的或分离的;天然存在的PAM特异性与工程化的PAM特异性等)。

[0369] 除了识别PAM和原间隔子的特定顺序取向,在一些实施方案中,RNA指导的核酸酶还可以识别特定PAM序列。在一些实施方案中,金黄色葡萄球菌Cas9通常识别NNGRRT或NNGRRV的PAM序列,其中N残基紧接gRNA靶向结构域识别的区域的3'。酿脓链球菌Cas9通常识别NGG PAM序列。并且新凶手弗朗西斯菌(*F. novicida*) Cpf1通常识别TTN PAM序列。

[0370] Yamano等人(*Cell*.2016年5月5日;165(4):949-962(Yamano),通过引用并入本文)已经解析了与crRNA和包括TTN PAM序列的双链(ds)DNA靶标复合的氨基酸球菌属物种(*Acidaminococcus* sp.) Cpf1的晶体结构。Cpf1像Cas9一样具有两种叶:REC(识别)叶和NUC(核酸酶)叶。REC叶包括REC1和REC2结构域,它们与任何已知的蛋白质结构都缺乏相似性。同时,NUC叶包括三个RuvC结构域(RuvC-I、RuvC-II和RuvC-III)和BH结构域。然而,与Cas9相反,Cpf1 REC叶缺乏HNH结构域,并且包括也与已知蛋白质结构缺乏相似性的其他结构域:结构独特的PI结构域、三个Wedge(WED)结构域(WED-I、WED-II和WED-III)和核酸酶(Nuc)结构域。

[0371] 尽管Cas9和Cpf1在结构和功能上共有相似性,但应当了解,某些Cpf1活性是由与任何Cas9结构域都不类似的结构域介导的。在一些实施方案中,靶DNA的互补链的切割似乎是由Nuc结构域介导的,所述Nuc结构域与Cas9的HNH结构域在顺序和空间上有所不同。另外,Cpf1 gRNA的非靶向部分(手柄)采用假结结构,而不是Cas9 gRNA中由重复:抗重

复双链体形成的茎环结构。

[0372] 本文提供了编码RNA指导的核酸酶(例如,Cas9、Cpf1或其功能片段)的核酸。先前已经描述了编码RNA指导的核酸酶的示例性核酸(参见例如,Cong 2013;Wang 2013;Mali 2013;Jinek 2012)。

b. 基因组编辑方法

[0373] 通常,应理解,根据本文所述方法的任何基因的改变都可以通过任何机制介导,并且任何方法都不限于特定机制。可以与基因改变相关的示例性机制包括但不限于非同源末端连接(例如,经典的或替代的)、微同源介导的末端连接(MMEJ)、同源定向修复(例如,内源供体模板介导的)、合成依赖性链退火(SDSA)、单链退火、单链侵入、单链断裂修复(SSBR)、错配修复(MMR)、碱基切除修复(BER)、链间交联(ICL)跨损伤合成(TLS)或无错复制后修复(PRR)。本文描述用于靶向敲除CD247基因座中的一个或所有的一个或两个等位基因的示例性方法。

1) 用于基因靶向的NHEJ方法

[0374] 如本文所述,核酸酶诱导的非同源末端连接(NHEJ)可以用于靶向基因特异性敲除。核酸酶诱导的NHEJ也可以用于去除(例如,缺失)目的基因中的序列插入。

[0375] 尽管不希望受到理论的束缚,认为在一些实施方案中,与本文所述方法相关的基因组改变依赖于核酸酶诱导的NHEJ和NHEJ修复途径的易错性质。NHEJ通过将两个末端连接在一起来修复DNA中的双链断裂;然而,通常,只有完美连接两个相容末端(在它们恰好由双链断裂形成时)才能恢复原始序列。双链断裂的DNA末端经常是酶加工的对象,从而导致在重新连接末端之前在一条或两条链处添加或去除核苷酸。这导致在DNA序列中在NHEJ修复位点处存在插入和/或缺失(indel)突变。这些突变的三分之二通常改变阅读框,因此产生非功能性蛋白质。此外,维持阅读框但插入或缺失大量序列的突变可能破坏蛋白质的功能性。这是基因座依赖性的,因为关键功能结构域中的突变可能比蛋白质非关键区域中的突变耐受性低。NHEJ产生的插入缺失突变本质上是不可预测的;然而,在给定的断裂位点处,某些插入缺失序列受到青睐,并且在群体中有过高的代表,这可能是由于微同源性的小区域所致。缺失的长度可以广泛地变化;最常见地在1-50bp的范围内,但是它们可以容易地达到大于100-200bp。插入往往较短,并且通常包括紧邻断裂位点周围的序列的短重复。然而,有可能获得大的插入,并且在这些情况下,所插入的序列通常被追踪到基因组的其他区域或细胞中存在的质粒DNA。

[0376] 因为NHEJ是诱变过程,它也可以用于缺失小序列基序,只要不需要产生特定的最终序列。如果双链断裂靶向在短靶序列附近,则由NHEJ修复引起的缺失突变通常跨越,因此去除不需要的核苷酸。对于较大的DNA区段的缺失,在序列的每一侧上引入两个双链断裂可以在末端之间导致NHEJ,并且去除整个间插序列。在一些实施方案中,一对gRNA可以用于引入两个双链断裂,从而导致缺失两个断裂之间的间插序列。

[0377] 这两种方法均可以用于缺失特定的DNA序列;然而,NHEJ的易错性质仍可能在修复位点处产生插入缺失突变。

[0378] 双链切割eaCas9分子和单链或切口酶eaCas9分子均可以用于本文所述的方法和组合中,以产生NHEJ介导的插入缺失。靶向目的基因(例如,基因的编码区,例如早期编码区)的NHEJ介导的插入缺失可以用于敲除目的基因(即,消除其表达)。例如,目的基因的早

期编码区包括紧接转录起始位点之后、在编码序列的第一外显子内或在转录起始位点的500bp内(例如,小于500、450、400、350、300、250、200、150、100或50bp)的序列。

[0379] 在一些实施方案中,将NHEJ介导的插入缺失引入一个或多个T细胞表达的基因(如CD247基因座)中。提供了靶向所述基因的单独的gRNA或gRNA对以及Cas9双链核酸酶或单链切口酶。

(1) 双链或单链断裂相对于靶位置的放置

[0380] 在gRNA和Cas9核酸酶生成双链断裂以用于诱导NHEJ介导的插入缺失的目的的一些实施方案中,gRNA(例如,单分子(或嵌合)或模块化gRNA分子)被配置将双链断裂定位地非常靠近靶位置的核苷酸。在一些实施方案中,切割位点在距离靶位置0-30bp之间(例如,距离靶位置小于30、25、20、15、10、9、8、7、6、5、4、3、2或1bp)。

[0381] 在与Cas9切口酶复合的两种gRNA诱导两个单链断裂以用于诱导NHEJ介导的插入缺失的目的的一些实施方案中,两种gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置为定位两个单链链断裂以提供NHEJ修复靶位置的核苷酸。在一些实施方案中,gRNA被配置为将切割定位在不同链上的相同位置处或彼此的几个核苷酸内,从而本质上模拟双链断裂。在一些实施方案中,较近的切口在距离靶位置0-30bp之间(例如,距离靶位置小于30、25、20、15、10、9、8、7、6、5、4、3、2或1bp),并且两个切口在彼此25-55bp内(例如,在25至50、25至45、25至40、25至35、25至30、50至55、45至55、40至55、35至55、30至55、30至50、35至50、40至50、45至50、35至45或40至45bp之间)且彼此相距不超过100bp(例如,不超过90、80、70、60、50、40、30、20或10bp)。在一些实施方案中,gRNA被配置为将单链断裂放置在靶位置的核苷酸的任一侧上。

[0382] 双链切割Cas9分子和单链或切口酶Cas9分子均可以用于本文所述的方法和组合物中,以在靶位置的两侧产生断裂。可以在靶位置的两侧上产生双链或成对的单链断裂,以去除两个切割之间的核酸序列(例如,缺失两个断裂之间的区域)。在一些实施方案中,两种gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置为将双链断裂定位在靶位置的两侧上。在替代实施方案中,三种gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置为将双链断裂(即,一种gRNA与Cas9核酸酶复合)和两个单链断裂或成对的单链断裂(即,两种gRNA与Cas9切口酶复合)定位在靶位置的任一侧上。在另一个实施方案中,四种gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置为在靶位置的任一侧上产生两对单链断裂(即,两对两种gRNA与Cas9切口酶复合)。所述一个或多个双链断裂或一对中的两个单链切口中的较近者将理想地在靶位置的0-500bp内(例如,距离靶位置不超过450、400、350、300、250、200、150、100、50或25bp)。当使用切口酶时,一对中的两个切口在彼此的25-55bp内(例如,在25至50、25至45、25至40、25至35、25至30、50至55、45至55、40至55、35至55、30至55、30至50、35至50、40至50、45至50、35至45或40至45bp之间)且彼此相距不超过100bp(例如,不超过90、80、70、60、50、40、30、20或10bp)。

2) 靶向敲低

[0383] 与通过在DNA水平上突变基因永久地消除或减少表达的CRISPR/Cas介导的基因敲除不同,CRISPR/Cas敲低允许通过使用人工转录因子暂时减少基因表达。突变Cas9蛋白的两个DNA切割域中的关键残基(例如,D10A和H840A突变)导致产生催化失活的Cas9(eiCas9,其也称为死Cas9或dCas9)。催化失活的Cas9与gRNA复合并且定位到由所述gRNA的靶向结构

域指定的DNA序列,然而,它不切割靶DNA。dCas9与效应子结构域(例如,转录阻遏结构域)的融合使得能够将效应子募集到由gRNA指定的任何DNA位点。尽管已经显示eiCas9自身在募集到编码序列中的早期区域时可以阻断转录,可以通过将转录阻遏结构域(例如KRAB、SID或ERD)与Cas9融合并将其募集到基因的启动子区来实现更稳健的阻遏。可能的是,靶向启动子的DNA酶I超敏区可能产生更有效的基因阻遏或激活,因为这些区域更可能是Cas9蛋白可及的,并且也更可能包含内源转录因子的位点。尤其是对于基因阻遏,本文考虑了阻断内源转录因子的结合位点将有助于下调基因表达。在另一个实施方案中,可以将eiCas9与染色质修饰蛋白融合。改变染色质状态可以导致靶基因表达降低。

[0384] 在一些实施方案中,可以将gRNA分子靶向已知的转录应答元件(例如,启动子、增强子等)、已知的上游激活序列(UAS)和/或怀疑能够控制靶DNA的表达的具有未知或已知功能的序列。

[0385] 在一些实施方案中,CRISPR/Cas介导的基因敲低可以用于减少一个或多个T细胞表达的基因的表达。在本文所述的eiCas9或eiCas9融合蛋白用于敲低CD247基因座的一些实施方案中,与所述eiCas9或eiCas9融合蛋白一起提供靶向两个或所有基因的单独的gRNA或gRNA对。

3) 单链退火

[0386] 单链退火(SSA)是另一种DNA修复过程,其修复靶核酸中存在的两个重复序列之间的双链断裂。SSA途径利用的重复序列的长度通常大于30个核苷酸。在断裂末端进行切除以揭示靶核酸的两条链上的重复序列。在切除之后,用RPA蛋白包被含有重复序列的单链突出端,以防止重复序列不适当地退火,例如自身退火。RAD52与突出端上的每个重复序列结合并对齐序列,以使得互补重复序列能够退火。在退火之后,切割突出端的单链瓣(flap)。新DNA合成填充任何缺口,并且连接恢复DNA双链体。作为加工的结果,缺失两个重复之间的DNA序列。缺失的长度可以取决于许多因素,包括所利用的两个重复的位置以及切除的途径或进行性(processivity)。

[0387] 与HDR途径相反,SSA不需要模板核酸来改变或校正靶核酸序列。相反,利用互补重复序列。

4) 其他DNA修复途径

A) SSB(单链断裂修复)

[0388] 基因组中的单链断裂(SSB)通过SSBR途径修复,这是与上文讨论的DSB修复机制不同的机制。SSBR途径具有四个主要阶段:SSB检测、DNA末端加工、DNA缺口填充和DNA连接。更详细的解释给出于Caldecott,Nature Reviews Genetics 9,619-631(2008年8月),并且这里给出了概述。

[0389] 在第一阶段,当形成SSB时,PARP1和/或PARP2识别断裂并募集修复机构。PARP1在DNA断裂处的结合和活性是短暂的,并且似乎通过促进损伤处SSBr蛋白复合物的局灶积累或稳定性来加速SSBr。可以说,这些SSBr蛋白中最重要的是XRCC1,它起分子支架的作用,所述分子支架与SSBr过程的多种酶促组分(包括负责清洁DNA 3'和5'末端的蛋白质)相互作用,对其进行稳定并对其进行刺激。在一些实施方案中,XRCC1与促进末端加工的几种蛋白质(DNA聚合酶 β 、PNK和三种核酸酶APE1、APTX和APLF)相互作用。APE1具有内切核酸酶活性。APLF展现内切核酸酶和3'至5'外切核酸酶活性。APTX具有内切核酸酶和3'至5'外切核酸酶

活性。

[0390] 此末端加工是SSBR的重要阶段,因为大多数(如果不是全部)SSB的3'末端和/或5'末端“受损”。末端加工通常涉及将受损的3'末端恢复为羟基化状态和/或将受损的5'末端恢复为磷酸部分,使得末端变得有连接能力。可以加工受损的3'末端的酶包括PNKP、APE1和TDP1。可以加工受损的5'末端的酶包括PNKP、DNA聚合酶 β 和APTX。LIG3(DNA连接酶III)也可以参与末端加工。一旦末端得到了清洁,便会发生缺口填充。

[0391] 在DNA缺口填充阶段,通常存在的蛋白质是PARP1、DNA聚合酶 β 、XRCC1、FEN1(瓣状内切核酸酶1)、DNA聚合酶 δ/ϵ 、PCNA和LIG1。缺口填充有两种方式,即短补丁修复和长补丁修复。短补丁修复涉及缺失的单个核苷酸的插入。在一些SSB处,“缺口填充”可能继续置换两个或更多个核苷酸(已经报道了置换多达12个碱基)。FEN1是去除置换的5'残基的内切核酸酶。多种DNA聚合酶(包括Po1 β)参与SSB的修复,其中DNA聚合酶的选择受SSB的来源和类型的影响。

[0392] 在第四阶段,DNA连接酶(如LIG1(连接酶I)或LIG3(连接酶III))催化末端的连接。短补丁修复使用连接酶III,并且长补丁修复使用连接酶I。

[0393] 有时,SSBR是复制偶联的。此途径可以涉及CtIP、MRN、ERCC1和FEN1中的一种或多种。可以促进SSBR的另外的因子包括:aPARP、PARP1、PARP2、PARG、XRCC1、DNA聚合酶b、DNA聚合酶d、DNA聚合酶e、PCNA、LIG1、PNK、PNKP、APE1、APTX、APLF、TDP1、LIG3、FEN1、CtIP、MRN和ERCC1。

B) MMR(错配修复)

[0394] 细胞含有三种切除修复途径:MMR、BER和NER。切除修复途径具有的共同特征在于它们通常识别DNA一条链上的损伤,然后外切核酸酶/内切核酸酶去除损伤并留下1-30个核苷酸的缺口,其随后被DNA聚合酶填充并最终被连接酶密封起来。更完整的画面给出于Li, Cell Research(2008)18:85-98,并且这里提供了概述。错配修复(MMR)对错配的DNA碱基进行操作。

[0395] MSH2/6或MSH2/3复合物均具有ATP酶活性,这在错配识别和修复启动中起重要作用。MSH2/6优先识别碱基-碱基错配并鉴定1或2个核苷酸的错配,而MSH2/3优先识别较大的ID错配。

[0396] hMLH1与hPMS2异二聚体形成hMutL α ,其具有ATP酶活性并且对于MMR的多个步骤是重要的。它具有PCNA/复制因子C(RFC)依赖性内切核酸酶活性,这在涉及EXO1的3'切口指导的MMR中起重要作用。(EXO1是HR和MMR两者的参与者。)它调节错配引起的切除的终止。连接酶I是此途径的相关连接酶。可以促进MMR的另外的因子包括:EXO1、MSH2、MSH3、MSH6、MLH1、PMS2、MLH3、DNA Po1 d、RPA、HMGB1、RFC和DNA连接酶I。

C) 碱基切除修复(BER)

[0397] 碱基切除修复(BER)途径在整个细胞周期中都是活跃的;它主要负责从基因组中去除小的非螺旋扭曲的碱基损伤。相反,相关的核苷酸切除修复途径(在下一章中讨论)修复庞大的螺旋扭曲损伤。更详细的解释给出于Caldecott, Nature Reviews Genetics 9, 619-631(2008年8月),并且这里给出了概述。

[0398] 一旦DNA碱基受损,便启动碱基切除修复(BER),并且所述过程可以简化为五个主要步骤:(a)去除受损的DNA碱基;(b)切下随后的碱基位点;(c)清理DNA末端;(d)将正确的

核苷酸插入修复缺口中;以及(e)连接DNA骨架中的其余切口。这些最后步骤类似于SSBR。

[0399] 在第一步,受损特异性DNA糖基化酶通过切割将碱基与糖磷酸骨架连接的N-糖苷键来切除受损的碱基。然后,AP内切核酸酶-1(APE1)或具有相关裂解酶活性的双功能DNA糖基化酶切开磷酸二酯骨架,以产生DNA单链断裂(SSB)。BER的第三步涉及清理DNA末端。BER中的第四步由Pol β 进行,它将新的互补核苷酸添加到修复缺口,并且在最终步骤,XRCC1/连接酶III将DNA骨架中的其余切口密封起来。这样就完成了短补丁BER途径,其中大部分(约80%)受损的DNA碱基得到了修复。然而,如果步骤3中的5'末端对末端加工活性具有抗性,则在通过Pol β 插入一个核苷酸后,然后将聚合酶转换为复制性DNA聚合酶Pol δ/ϵ ,其然后再将约2-8核苷酸添加到DNA修复缺口中。这产生了5'瓣结构,其被瓣状内切核酸酶-1(FEN-1)联合进行性因子增殖细胞核抗原(PCNA)识别并切除。然后,DNA连接酶I将DNA骨架中的其余切口密封起来,并且完成了长补丁BER。可以促进BER途径的另外的因子包括:DNA糖基化酶、APE1、Pol β 、Pol δ 、Pol ϵ 、XRCC1、连接酶III、FEN-1、PCNA、RECQL4、WRN、MYH、PNKP和APTX。

D) 核苷酸切除修复(NER)

[0400] 核苷酸切除修复(NER)是重要的切除机制,其从DNA中去除庞大的螺旋扭曲的损伤。有关NER的另外的细节给出于Marteijn等人,Nature Reviews Molecular Cell Biology 15,465-481(2014),并且这里给出了概述。NER的广泛途径涵盖两个较小的途径:全基因组NER(GG-NER)和转录偶联修复NER(TC-NER)。GG-NER和TC-NER使用不同的因子来识别DNA受损。然而,它们使用相同的机构进行损伤切除、修复和连接。

[0401] 一旦识别出受损,细胞便去除含有损伤的短单链DNA区段。内切核酸酶XPF/ERCC1和XPG(由ERCC5编码)通过切割损伤任一侧上的受损链来去除损伤,从而产生22-30个核苷酸的单链缺口。接下来,细胞进行DNA缺口填充合成和连接。参与此过程的是:PCNA、RFC、DNA Pol δ 、DNA Pol ϵ 或DNA Pol κ 和DNA连接酶I或XRCC1/连接酶III。复制细胞往往使用DNA pol ϵ 和DNA连接酶I,而非复制细胞往往使用DNA Pol δ 、DNA Pol κ 和XRCC1/连接酶III复合物进行连接步骤。

[0402] NER可能涉及以下因子:XPA-G、POLH、XPF、ERCC1、XPA-G和LIG1。转录偶联NER(TC-NER)可能涉及以下因子:CSA、CSB、XPB、XPD、XPG、ERCC1和TTDA。可以促进NER修复途径的另外的因子包括XPA-G、POLH、XPF、ERCC1、XPA-G、LIG1、CSA、CSB、XPA、XPB、XPC、XPD、XPF、XPG、TTDA、UVSSA、USP7、CETN2、RAD23B、UV-DDB、CAK亚复合物、RPA和PCNA。

E) 链间交联(ICL)

[0403] 称为ICL修复途径的专用途径修复链间交联。在复制或转录期间可能发生链间交联或不同DNA链中碱基之间的共价交联。ICL修复涉及多个修复过程(特别是溶核活性、跨损伤合成(TLS)和HDR)的协调。募集核酸酶以切除交联碱基任一侧上的ICL,同时协调TLS和HDR以修复切割链。ICL修复可能涉及以下因子:内切核酸酶(例如,XPF和RAD51C)、内切核酸酶(如RAD51)、跨损伤聚合酶(例如,DNA聚合酶 ζ 和Rev1)以及范科尼贫血(FA)蛋白(例如,FancJ)。

F) 其他途径

[0404] 在哺乳动物中存在几种其他DNA修复途径。跨损伤合成(TLS)是用于修复有缺陷的复制事件之后留下的单链断裂的途径,并且涉及跨损伤聚合酶,例如DNA pol ζ 和Rev1。无错

复制后修复 (PRR) 是用于修复有缺陷的复制事件之后留下的单链断裂的另一种途径。

(2) 用于基因编辑的药剂的功能分析

[0405] 可以通过本领域已知的方法或如本文所述的评价任何Cas9分子、gRNA分子、Cas9分子/gRNA分子复合物。例如,用于评价Cas9分子的内切核酸酶活性的示例性方法描述于例如Jinek等人,Science 2012,337 (6096):816-821。

(a) 结合和切割测定:测试Cas9分子的内切核酸酶活性

[0406] 可以在质粒切割测定中评价Cas9分子/gRNA分子复合物结合至并切割靶核酸的能力。在此测定中,通过加热至95℃并缓慢冷却至室温,将合成的或体外转录的gRNA分子在反应前预退火。将天然的或限制性消化线性化的质粒DNA (300ng (约8nM)) 与经纯化的Cas9蛋白分子 (50-500nM) 和gRNA (50-500nM, 1:1) 在含或不含10mM MgCl₂的Cas9质粒切割缓冲液 (20mM HEPES pH 7.5, 150mM KCl, 0.5mM DTT, 0.1mM EDTA) 中在37℃下一起孵育60min。将反应用5X DNA上样缓冲液 (30%甘油, 1.2% SDS, 250mM EDTA) 终止,通过0.8%或1%琼脂糖凝胶电泳解析,并且通过溴化乙锭可视化。所得切割产物指示了Cas9分子是切割两条DNA链还是仅切割两条链之一。例如,线性DNA产物指示了两条DNA链的切割。带切口的开放圆形产物表明仅切割了两条链之一。

[0407] 可替代地,可以在寡核苷酸DNA切割测定中评价Cas9分子/gRNA分子复合物结合至并切割靶核酸的能力。在此测定中,将DNA寡核苷酸 (10pmol) 通过在50μL反应中与5个单位的T4多核苷酸激酶和1X T4多核苷酸激酶反应缓冲液中的约3-6pmol (约20-40mCi) [γ-³²P]-ATP在37℃下一起孵育30min来进行放射性标记。在热灭活 (65℃持续20min) 之后,将反应物通过柱纯化,以去除未掺入的标记。通过如下方式产生双链体底物 (100nM): 将经标记的寡核苷酸与等摩尔量的未经标记的互补寡核苷酸在95℃下退火3min,然后缓慢冷却至室温。对于切割测定,通过如下方式使gRNA分子退火:加热至95℃持续30s,然后缓慢冷却至室温。将Cas9 (终浓度为500nM) 与退火的gRNA分子 (500nM) 在切割测定缓冲液 (20mM HEPES pH 7.5, 100mM KCl, 5mM MgCl₂, 1mM DTT, 5%甘油) 中预孵育,总体积为9μL。通过添加1μL靶DNA (10nM) 来起始反应,并在37℃下孵育1h。通过添加20μL上样染料 (甲酰胺中的5mM EDTA、0.025% SDS、5%甘油) 终止反应,并加热至95℃持续5min。在含有7M尿素的12%变性聚丙烯酰胺凝胶上解析切割产物,并通过磷成像进行可视化。所得切割产物表明切割了互补链、非互补链还是两者。

[0408] 这些测定之一或两者可以用于评价所提供的任何gRNA分子或Cas9分子的适合性。

G) 结合测定:测试Cas9分子与靶DNA的结合

[0409] 用于评价Cas9分子与靶DNA的结合的示例性方法描述于例如Jinek等人,Science 2012;337 (6096):816-821。

[0410] 例如,在电泳迁移率变动测定中,通过如下方式形成靶DNA双链体:将每条链 (10nmol) 在去离子水中混合,加热至95℃持续3min并且缓慢冷却至室温。将全部DNA在含有1X TBE的8%天然凝胶上纯化。将DNA条带通过UV遮蔽可视化,切下并通过将凝胶碎片浸泡在DEPC处理的H₂O中进行洗脱。将经洗脱的DNA进行乙醇沉淀,并且溶解在DEPC处理的H₂O中。使用T4多核苷酸激酶将DNA样品5'端用[γ-³²P]-ATP在37℃下标记30min。使多核苷酸激酶在65℃下热变性20min,并使用柱去除未掺入的放射性标记。在总体积为10μL的含有20mM HEPES pH 7.5、100mM KCl、5mM MgCl₂、1mM DTT和10%甘油的缓冲液中进行结合测定。将

Cas9蛋白分子用等摩尔量的预退火gRNA分子编程,并且从100pM滴定至1 μ M。添加放射性标记的DNA至终浓度为20pM。将样品在37 $^{\circ}$ C下孵育1h,并且在4 $^{\circ}$ C下在含有1X TBE和5mM MgCl₂的8%天然聚丙烯酰胺凝胶上解析。将凝胶干燥,并且将DNA通过磷成像可视化。

H) 用于测量Cas9/gRNA复合物的热稳定性的技术

[0411] 可以通过差示扫描荧光测定法(DSF)和其他技术来检测Cas9-gRNA核糖核蛋白(RNP)复合物的热稳定性。蛋白质的热稳定性可以在有利的条件(如添加结合性RNA分子,例如gRNA)下增加。因此,关于Cas9/gRNA复合物的热稳定性的信息对于确定复合物是否稳定是有用的。

I) 差示扫描荧光测定法(DSF)

[0412] 可以经由DSF测量Cas9-gRNA核糖核蛋白(RNP)复合物的热稳定性。如下所述,RNP复合物包括核糖核苷酸序列(如RNA或gRNA)和蛋白质(如Cas9蛋白或其变体)。此技术测量蛋白质的热稳定性,其可以在有利的条件(如添加结合性RNA分子,例如gRNA)下增加。

[0413] 所述测定能以许多方式应用。示例性方案包括但不限于确定用于RNP形成的所需溶液条件的方案(测定1,参见下文)、测试gRNA:Cas9蛋白的所需化学计量比的方案(测定2,参见下文)、筛选Cas9分子(例如,野生型或突变型Cas9分子)的有效gRNA分子的方案(测定3,参见下文)以及在靶DNA的存在下检查RNP形成的方案(测定4)。在一些实施方案中,使用两种不同的方案进行测定,一种用于测试gRNA:Cas9蛋白的最佳化学计量比,并且另一种用于确定RNP形成的最佳溶液条件。

[0414] 为了确定形成RNP复合物的最佳溶液,将Cas9在水+10x SYPRO Orange[®] (Life Technologies目录号S-6650)中的2 μ M溶液分配到384孔板中。然后添加等摩尔量的在具有不同pH和盐的溶液中稀释的gRNA。在室温下孵育10'并短暂离心以去除任何气泡之后,使用具有Bio-Rad CFX Manager软件的Bio-Rad CFX384[™]实时系统C1000 Touch[™]热循环仪运行从20 $^{\circ}$ C至90 $^{\circ}$ C的梯度,其中温度每10秒升高1 $^{\circ}$ 。

[0415] 第二测定由将各种浓度的gRNA与2 μ M Cas9在来自上述方法1的最佳缓冲液中混合并且在384孔板中于室温下孵育10'组成。添加等体积的最佳缓冲液+10x SYPRO Orange[®] (Life Technologies目录号S-6650),并且将板用Microseal[®]B粘合剂(MSB-1001)密封起来。短暂离心以去除任何气泡后,使用具有Bio-Rad CFX Manager软件的Bio-Rad CFX384[™]实时系统C1000 Touch[™]热循环仪运行从20 $^{\circ}$ C至90 $^{\circ}$ C的梯度,其中温度每10秒升高1 $^{\circ}$ 。

[0416] 在第三测定中,纯化目的Cas9分子(例如Cas9蛋白,例如Cas9变体蛋白)。合成变体gRNA分子文库,并且将其重悬至20 μ M的浓度。在5x SYPRO Orange[®] (Life Technologies目录号S-6650)的存在下,将Cas9分子与gRNA分子(各自的终浓度为1 μ M)在预定的缓冲液中一起孵育。在室温下孵育10分钟并在2000rpm下离心2分钟以去除任何气泡之后,使用具有Bio-Rad CFX Manager软件的Bio-Rad CFX384[™]实时系统C1000Touch[™]热循环仪运行从20 $^{\circ}$ C至90 $^{\circ}$ C的梯度,其中温度每10秒升高1 $^{\circ}$ 。

[0417] 在第四测定中,对以下样品进行DSF实验:仅Cas9蛋白、Cas9蛋白与gRNA、Cas9蛋白与gRNA和靶DNA以及Cas9蛋白与靶DNA。混合组分的顺序为:反应溶液、Cas9蛋白、gRNA、DNA和SYPRO Orange。反应溶液含有10mM HEPES pH 7.5、100mM NaCl,不存在或存在MgCl₂。在2000rpm下离心2分钟以去除任何气泡后,使用具有Bio-Rad CFX Manager软件的Bio-Rad CFX384[™]实时系统C1000 Touch[™]热循环仪运行从20 $^{\circ}$ C至90 $^{\circ}$ C的梯度,其中温度每10秒升高

1°。

3. 用于遗传破坏的药剂的递送

[0418] 在一些实施方案中,人体中内源CD247基因座(编码CD3ζ)的靶向遗传破坏(例如,DNA断裂)是通过将能够诱导遗传破坏的一种或多种药剂(例如,Cas9和/或gRNA组分)递送或引入至细胞中来进行,其使用用于引入或转移至细胞的多种已知递送方法或媒介物中的任一种(例如,使用病毒(例如,慢病毒)递送载体),或者用于递送Cas9分子和gRNA的任何已知方法或媒介物。示例性方法描述于例如以下文献中:Wang等人(2012) *J. Immunother.* 35 (9):689-701;Cooper等人(2003) *Blood.* 101:1637-1644;Verhoeyen等人(2009) *Methods Mol Biol.* 506:97-114;以及Cavalieri等人(2003) *Blood.* 102 (2):497-505。在一些实施方案中,例如通过本文所述或已知的将核酸引入细胞中的任何方法将编码能够诱导遗传破坏(例如,DNA断裂)的一种或多种药剂的一种或多种组分的核酸序列引入细胞中。在一些实施方案中,可以将编码能够诱导遗传破坏的一种或多种药剂(如CRISPR指导RNA和/或Cas9酶)的组分的载体递送至细胞中。

[0419] 在一些实施方案中,将能够诱导遗传破坏的所述一种或多种药剂(例如,作为Cas9/gRNA的一种或多种药剂)作为核糖核蛋白(RNP)复合物引入细胞中。RNP复合物包括核糖核苷酸序列(如RNA或gRNA分子)和蛋白质(如Cas9蛋白或其变体)。例如,例如使用电穿孔或其他物理递送方法,将Cas9蛋白作为RNP复合物递送,所述复合物包含Cas9蛋白和靶向靶序列的gRNA分子。在一些实施方案中,经由电穿孔或其他物理方式(例如,粒子枪、磷酸钙转染、细胞压缩或挤压)将RNP递送至细胞中。在一些实施方案中,RNP可以穿过细胞的质膜而无需另外的递送剂(例如,小分子药剂、脂质等)。在一些实施方案中,将能够诱导遗传破坏的所述一种或多种药剂(例如,CRISPR/Cas9)作为RNP递送提供如下优点:例如在引入RNP的细胞中短暂发生靶向破坏,而不将所述药剂传播到细胞后代。例如,通过RNP进行的递送使被遗传到其后代的药剂最小化,从而减小后代中脱靶遗传破坏的可能性。在此类情况下,遗传破坏和转基因的整合可以被后代细胞遗传,但是可能会进一步引入脱靶遗传破坏的药剂本身不被传递给后代细胞。

[0420] 使用多种递送方法和配制品(如表3和表4中所示)或者例如WO 2015/161276;US 2015/0056705;US 2016/0272999;US 2017/0211075;或US 2017/0016027中描述的方法可以将能够诱导遗传破坏的一种或多种药剂和组分(例如,Cas9分子和gRNA分子)以多种形式引入靶细胞中。如本文进一步描述,可以在本文所述方法的先前或随后步骤中使用所述递送方法和配制品将模板多核苷酸和/或其他药剂(如工程化细胞所需的那些)递送至细胞。当Cas9或gRNA组分被编码为DNA用于递送时,DNA通常可以但不一定包括控制区,例如包含启动子,以实现表达。Cas9分子序列的有用启动子包括例如CMV、EF-1α、EFS、MSCV、PGK或CAG启动子。gRNA的有用启动子包括例如H1、EF-1α、tRNA或U6启动子。可以选择具有相似或不相似强度的启动子来调整组分的表达。编码Cas9分子的序列可以包含核定位信号(NLS),例如SV40 NLS。在一些实施方案中,Cas9分子或gRNA分子的启动子可以独立地是可诱导的、组织特异性的或细胞特异性的。在一些实施方案中,能够诱导遗传破坏的药剂是引入的RNP复合物。在一些实施方案中,gRNA含有修饰,如A1t-R修饰(IDT Technologies;科勒尔维尔,爱荷华州)。

表3. 示例性递送方法

元件		注释
一种或多种 Cas9分子	一种或多种 gRNA分子	
DNA	DNA	在此实施方案中，Cas9分子和gRNA是从DNA转录。在此实施方案中，它们是在单独分子上编码。
DNA		在此实施方案中，Cas9分子和gRNA是从DNA转录，这里是从单一分子转录。
DNA	RNA	在此实施方案中，Cas9分子是从DNA转录，并且gRNA是作为体外转录的或合成的RNA来提供。
mRNA	RNA	在此实施方案中，Cas9分子是从体外转录的mRNA翻译，并且gRNA是作为体外转录的或合成的RNA来提供。
mRNA	DNA	在此实施方案中，Cas9分子是从体外转录的mRNA翻译，并且gRNA是从DNA转录。
蛋白质	DNA	在此实施方案中，Cas9分子是作为蛋白质来提供，并且gRNA是从DNA转录。
蛋白质	RNA	在此实施方案中，Cas9分子是作为蛋白质来提供，并且gRNA是作为转录的或合成的RNA来提供。

表4. 示例性递送方法的比较

递送载体/方式		递送至非分裂细胞中	表达的持续时间	基因组整合	所递送分子的类型
物理的(例如，电穿孔、粒子枪、磷酸钙转染、细胞压缩或挤压)		是	短暂的	否	核酸和蛋白质
病毒的	逆转录病毒	否	稳定的	是	RNA
	慢病毒	是	稳定的	是/在有修饰的情况下，否	RNA
	腺病毒	是	短暂的	否	DNA
	腺相关病毒(AAV)	是	稳定的	否	DNA
	牛痘病毒	是	非常短暂	否	DNA
	单纯疱疹病毒	是	稳定的	否	DNA

递送载体/方式		递送至非分裂细胞中	表达的持续时间	基因组整合	所递送分子的类型
非病毒的	阳离子脂质体	是	短暂的	取决于递送什么	核酸和蛋白质
	聚合纳米颗粒	是	短暂的	取决于递送什么	核酸和蛋白质
生物非病毒递送媒介物	减毒细菌	是	短暂的	否	核酸
	工程化的噬菌体	是	短暂的	否	核酸
	哺乳动物病毒样颗粒	是	短暂的	否	核酸
	生物脂质体：红细胞血影和外来体	是	短暂的	否	核酸

[0421] 在一些实施方案中，可以通过已知或如本文所述的方法将编码Cas9分子和/或gRNA分子的DNA或包含Cas9分子和/或gRNA分子的RNP复合物递送至细胞中。例如，可以例如通过载体(例如，病毒或非病毒载体)、基于非载体的方法(例如，使用裸DNA或DNA复合物)或其组合来递送Cas9编码DNA和/或gRNA编码DNA。在一些实施方案中，通过载体(例如，病毒载

体/病毒或质粒)递送含有一种或多种药剂和/或其组分的多核苷酸。载体可以是本文所述的任何载体。

[0422] 在一些方面,将与指导序列组合(并且任选地与其复合)的CRISPR酶(例如Cas9核酸酶)递送至细胞中。例如,CRISPR系统的一个或多个元件源自I型、II型或III型CRISPR系统。例如,CRISPR系统的一个或多个元件源自包含内源CRISPR系统的特定生物体,如酿脓链球菌、金黄色葡萄球菌或脑膜炎奈瑟球菌。

[0423] 在一些实施方案中,将Cas9核酸酶(例如,由来自金黄色葡萄球菌或来自酿脓链球菌的mRNA编码,例如pCW-Cas9, Addgene#50661, Wang等人(2014) Science, 3:343-80-4;或者可从Applied Biological Materials (ABM;加拿大)以目录号K002、K003、K005或K006获得的核酸酶或切口酶慢病毒载体)和对靶基因(例如人体中的CD247基因座)具有特异性的指导RNA引入细胞中。

[0424] 在一些实施方案中,通过基于非载体的方法(例如,使用裸DNA或DNA复合物)递送含有一种或多种药剂和/或其组分的多核苷酸或RNP复合物。例如,DNA或RNA或蛋白质或其组合(例如,核糖核蛋白(RNP)复合物)可以例如通过以下方式递送:有机改性的二氧化硅或硅酸盐(Ormosil)、电穿孔、瞬时细胞压缩或挤压(如以下文献中所述:Lee,等人(2012) Nano Lett 12:6322-27;Kollmannsperger等人(2016) Nat Comm 7,10372)、基因枪、声孔效应、磁转染、脂质介导的转染、树枝状聚合物、无机纳米颗粒、磷酸钙或其组合。

[0425] 在一些实施方案中,经由电穿孔递送包括将细胞与Cas9编码DNA和/或gRNA编码DNA或RNP复合物在筒、室或比色皿中混合,并且施加具有限定的持续时间和幅度的一次或多次电脉冲。在一些实施方案中,经由电穿孔递送是使用如下系统来进行,其中将细胞与Cas9编码DNA和/或gRNA编码DNA在与装置(例如,泵)连接的容器中混合,所述装置将混合物进料至筒、室或比色皿中,其中施加限定持续时间和幅度的一次或多次电脉冲,之后将细胞递送至第二容器。

[0426] 在一些实施方案中,递送媒介物是非病毒载体。在一些实施方案中,非病毒载体是无机纳米颗粒。示例性无机纳米颗粒包括例如磁性纳米颗粒(例如,Fe₃MnO₂)和二氧化硅。纳米颗粒的外表面可以与带正电荷的聚合物(例如,聚乙烯亚胺、聚赖氨酸、聚丝氨酸)缀合,其允许有效载荷的附着(例如,缀合或截留)。在一些实施方案中,非病毒载体是有机纳米颗粒。示例性有机纳米颗粒包括例如SNALP脂质体,其含有阳离子脂质和被聚乙二醇(PEG)包被的中性辅助脂质;以及被脂质包被的鱼精蛋白-核酸复合物。用于基因转移的示例性脂质示于下表5中。

表5. 用于基因转移的脂质

脂质	缩写	特征
1,2-二油酰基-sn-甘油-3-磷脂酰胆碱	DOPC	辅助者
1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助者
胆固醇		辅助者
<i>N</i> -[1-(2,3-二油烯基氧基)丙基] <i>N,N,N</i> -三甲基氯化铵	DOTMA	阳离子的
1,2-二油酰基氧基-3-三甲基铵-丙烷	DOTAP	阳离子的
双十八烷基酰胺基甘氨酸精胺	DOGS	阳离子的
<i>N</i> -(3-氨基丙基)- <i>N,N</i> -二甲基-2,3-双(十二烷基氧基)-1-溴化丙铵	GAP-DLRIE	阳离子的
十六烷基三甲基溴化铵	CTAB	阳离子的
6-月桂氧基己基鸟氨酸酯	LHON	阳离子的
1-(2,3-二油酰基氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子的
2,3-二油酰基氧基- <i>N</i> -[2(精胺甲酰胺基-乙基)- <i>N,N</i> -二甲基-1-丙铵三氟乙酸盐]	DOSPA	阳离子的
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子的
<i>N</i> -(2-羟基乙基)- <i>N,N</i> -二甲基-2,3-双(十四烷基氧基)-1-溴化丙铵	MDRIE	阳离子的
二肉豆蔻氧基丙基二甲基羟乙基溴化铵	DMRI	阳离子的
3β-[<i>N</i> -(<i>N,N'</i> -二甲基氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子的
双胍鎓-三氮乙基胺 (tren)-胆固醇	BGTC	阳离子的
1,3-二脱氧-2-(6-羧基-精胺基 (spermyl))-丙基酰胺	DOSPER	阳离子的
二甲基十八烷基溴化铵	DDAB	阳离子的
双十八烷基酰氨基甘氨酸亚精胺	DSL	阳离子的
外消旋-[2(3-二十八烷基氧基丙基)(2-羟基乙基)]-二甲基氯化铵	CLIP-1	阳离子的
外消旋-[2(2,3-二十六烷基氧基丙基-氧基甲基氧基)乙基]三甲基溴化铵	CLIP-6	阳离子的
乙基二肉豆蔻酰基磷脂酰胆碱	EDMPC	阳离子的
1,2-二硬脂酰基氧基- <i>N,N</i> -二甲基-3-氨基丙烷	DSDMA	阳离子的
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子的
<i>O,O'</i> -二肉豆蔻基- <i>N</i> -赖氨酸天冬氨酸酯	DMKE	阳离子的
1,2-二硬脂酰基-sn-甘油-3-乙基磷酸胆碱	DSEPC	阳离子的
<i>N</i> -棕榈酰基 D-赤型-鞘氨醇 (sphingosyl) 氨基甲酰基-精胺	CCS	阳离子的
<i>N</i> -叔丁基- <i>N</i> -(十四烷基-3-十四烷基氨基丙基)	二C14-咪	阳离子的
十八碳烯醇基氧基[乙基-2-十七碳烯基-3-羟基乙基]氯化咪唑鎓	DOTIM	阳离子的
<i>NI</i> -胆固醇基氧基羧基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子的
2-(3-[双(3-氨基-丙基)-氨基]丙基氨基)- <i>N</i> -二十四烷基氨基甲酰基甲-乙基-乙酰胺	RPR209120	阳离子的
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子的
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子的
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子的

[0427] 用于基因转移的示例性聚合物示于下表6中。

表6. 用于基因转移的聚合物

聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP

聚合物	缩写
二甲基-3,3'-二硫代双丙亚氨酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的PLL	
聚(N-乙基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰胺基胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚(β -氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙胺)	
聚(α -[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(N-乙基-4-乙基溴化吡啶鎓)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(氨基磷酸酯)	PPA
聚(N-2-羟基丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲基氨基)乙基甲基丙烯酸酯)	pDMAEMA
聚(2-氨基乙基丙烯磷酸酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-十二烷基化壳聚糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[0428] 在一些实施方案中,媒介物具有靶向修饰以增加纳米颗粒和脂质体(例如,细胞特异性抗原、单克隆抗体、单链抗体、适体、聚合物、糖和细胞穿透肽)的靶细胞更新。在一些实施方案中,媒介物使用促融合且内体去稳定的肽/聚合物。在一些实施方案中,媒介物经历酸触发的构象变化(例如,加快负荷的内体逃逸)。在一些实施方案中,使用刺激物可切割的聚合物,例如,用于细胞区室中释放。例如,可以使用在还原性细胞环境中切割的基于二硫化物的阳离子聚合物。

[0429] 在一些实施方案中,递送媒介物是生物非病毒递送媒介物。在一些实施方案中,媒介物是减毒细菌(例如,被天然或人工工程化已具有侵入性,但进行减毒以防止发病,并且表达转基因(例如,单核细胞增生李斯特菌、某些沙门菌属(*Salmonella*)菌株、长双歧杆菌(*Bifidobacterium longum*)、和经修饰的大肠杆菌(*Escherichia coli*))、具有营养和组织特异性向性以靶向特定细胞的细菌、具有经修饰的表面蛋白以改变靶细胞特异性的细菌)。在一些实施方案中,媒介物是基因修饰的噬菌体(例如,具有大包装容量、较低免疫原性、含有哺乳动物质粒维持序列且具有掺入的靶向配体的工程化噬菌体)。在一些实施方案中,媒介物是哺乳动物病毒样颗粒。例如,可以产生经修饰的病毒颗粒(例如,通过纯化“空”颗粒,然后将病毒与所需负荷离体组装)。媒介物还可以被工程化以掺入靶向配体,以改变靶组织特异性。在一些实施方案中,媒介物是生物脂质体。例如,生物脂质体是源自人细胞的基于磷脂的颗粒(例如,红细胞血影,其是源自受试者的分解成球形结构的红细胞(例如,组织靶向可以通过附着各种组织或细胞特异性配体实现))或分泌性外来体—内吞起源的受试者

衍生的膜结合纳米囊泡 (30-100nm) (例如,可以从各种细胞类型产生,因此可以被细胞摄取而不需要靶向配体)。

[0430] 在一些实施方案中,可以通过已知方法或如本文所述将编码Cas9分子和/或gRNA分子的RNA递送至细胞(例如,本文所述的靶细胞)中。例如,Cas9编码RNA和/或gRNA编码RNA可以例如通过以下来实现:显微注射、电穿孔、瞬时细胞压缩或挤压(如Lee,等人(2012) Nano Lett 12:6322-27中所述)、脂质介导的转染、肽介导的递送(例如,细胞穿膜肽)或其组合。

[0431] 在一些实施方案中,经由电穿孔递送包括将细胞与编码Cas9分子和/或gRNA分子的RNA在筒、室或比色皿中混合,并且施加具有限定的持续时间和幅度的一次或多次电脉冲。在一些实施方案中,经由电穿孔递送是使用如下系统来进行,其中将细胞与编码Cas9分子和/或gRNA分子的RNA在与装置(例如,泵)连接的容器中混合,所述装置将混合物进料至筒、室或比色皿中,其中施加限定持续时间和幅度的一次或多次电脉冲,之后将细胞递送至第二容器。

[0432] 在一些实施方案中,可以通过已知方法或如本文所述将Cas9分子递送至细胞中。例如,Cas9蛋白分子可以例如通过以下方式递送:显微注射、电穿孔、瞬时细胞压缩或挤压(如Lee,等人(2012) Nano Lett 12:6322-27中所述)、脂质介导的转染、肽介导的递送或其组合。递送可以伴随编码gRNA的DNA或伴随gRNA。

[0433] 在一些实施方案中,将能够引入切割的所述一种或多种药剂(例如,Cas9/gRNA)作为核糖核蛋白(RNP)复合物引入细胞中。RNP复合物包括核糖核苷酸序列(如RNA或gRNA分子)和蛋白质(如Cas9蛋白或其变体)。例如,例如使用电穿孔或其他物理递送方法,将Cas9蛋白作为RNP复合物递送,所述复合物包含Cas9蛋白和靶向靶序列的gRNA分子。在一些实施方案中,经由电穿孔或其他物理方式(例如,粒子枪、磷酸钙转染、细胞压缩或挤压)将RNP递送至细胞中。

[0434] 在一些实施方案中,所述一种或多种药剂是或包含核糖核蛋白(RNP)复合物。在一些实施方案中,与用于工程化的细胞一起孵育、添加至所述细胞或与所述细胞接触的RNP的浓度是以下浓度:为或约0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、2.2、2.5、3、4、5、6、7、7.5、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40或50 μ M,或者由任两个前述值限定的范围。在一些实施方案中,与用于工程化的细胞一起孵育、添加至所述细胞或与所述细胞接触的RNP的浓度是以下浓度:为或约1、2、2.5、5、10、20、25、30、40或50 μ M,或者由任两个前述值限定的范围。在一些实施方案中,RNP的浓度为2 μ M。在一些实施方案中,RNP的浓度为25 μ M。在一些实施方案中,在RNP复合物中,gRNA与Cas9分子或其他核酸酶的比率(例如,摩尔比)是为或约5:1、4:1、3:1、2:1、1:1、1:2、1:3、1:4或1:5,或者由任两个前述值限定的范围。在一些实施方案中,在RNP复合物中,gRNA与Cas9分子或其他核酸酶的比率(例如,摩尔比)是为或约3:1、2.9:1、2.8:1、2.7:1、2.6:1、2.5:1、2.4:1、2.3:1、2.2:1、2.1:1、2:1或1:1,或者由任两个前述值限定的范围。在一些实施方案中,在RNP复合物中,gRNA与Cas9分子或其他核酸酶的摩尔比是为或约2.6:1。

[0435] 在一些实施方案中,经由电穿孔递送包括将细胞与Cas9分子在具有或不具有gRNA分子的情况下在筒、室或比色皿中混合,并且施加具有限定的持续时间和幅度的一次或多次电脉冲。在一些实施方案中,经由电穿孔递送是使用如下系统来进行,其中将细胞与Cas9

分子在具有或不具有gRNA分子的情况下在与装置(例如,泵)连接的容器混合,所述装置将混合物进料至筒、室或比色皿中,其中施加限定持续时间和幅度的一次或多次电脉冲,之后将细胞递送至第二容器。

[0436] 在一些实施方案中,经由电穿孔进行的递送包括将细胞与Cas9分子(例如,eaCas9分子、eiCas9分子或eiCas9融合蛋白)、与或不与gRNA分子在盒、室或比色皿中混合,并且施加具有限定的持续时间和振幅的一个或多个电脉冲。在一些实施方案中,经由电穿孔递送是使用如下系统来进行:其中将细胞与Cas9分子(例如,eaCas9分子、eiCas9分子或eiCas9融合蛋白)混合。

[0437] 在一些实施方案中,通过基于载体的方法与基于非载体的方法的组合递送含有一种或多种药剂和/或其组分的多核苷酸。例如,病毒体包含与失活的病毒(例如,HIV或流感病毒)组合的脂质体,其可以导致比单独的病毒或脂质体方法更有效的基因转移。

[0438] 在一些实施方案中,将超过一种药剂或其组分递送至细胞中。例如,在一些实施方案中,将能够诱导基因组中的两个或更多个位置(如在CD247基因座(编码CD3 ζ)内的两个或更多个位点处)的遗传破坏的一种或多种药剂递送至细胞。在一些实施方案中,使用一种方法递送一种或多种药剂及其组分。例如,在一些实施方案中,用于诱导CD247基因座的遗传破坏的一种或多种药剂是作为编码用于遗传破坏的组分的多核苷酸来递送。在一些实施方案中,一种多核苷酸可以编码靶向CD247基因座的药剂。在一些实施方案中,两种或更多种不同多核苷酸可以编码靶向CD247基因座的药剂。在一些实施方案中,能够诱导遗传破坏的药剂可以作为核糖核蛋白(RNP)复合物来递送,并且两种或更多种不同的RNP复合物可以作为混合物一起递送或分开递送。

[0439] 在一些实施方案中,递送除了能够诱导遗传破坏的所述一种或多种药剂和/或其组分(例如,Cas9分子组分和/或gRNA分子组分)以外的一种或多种核酸分子,如用于HDR引导的整合的模板多核苷酸(如本文例如在章节I.B.2中所述的任何模板多核苷酸)。在一些实施方案中,在与Cas系统的一种或多种组分相同的时间递送核酸分子(例如,模板多核苷酸)。在一些实施方案中,在递送Cas系统的一种或多种组分之前或之后(例如,小于约1分钟、5分钟、10分钟、15分钟、30分钟、1小时、2小时、3小时、6小时、9小时、12小时、1天、2天、3天、1周、2周或4周)递送核酸分子。在一些实施方案中,通过与Cas系统的一种或多种组分(例如,Cas9分子组分和/或gRNA分子组分)不同的方式来递送核酸分子(例如,模板多核苷酸)。核酸分子(例如,模板多核苷酸)可以通过本文所述的任何递送方法来递送。例如,核酸分子(例如,模板多核苷酸)可以通过病毒载体(例如,逆转录病毒或慢病毒)来递送,并且Cas9分子组分和/或gRNA分子组分可以通过电穿孔来递送。在一些实施方案中,核酸分子(例如,模板多核苷酸)包括一个或多个外源序列,例如,编码嵌合受体或其部分的序列和/或其他外源基因核酸序列。

B. 经由同源定向修复(HDR)进行靶向整合

[0440] 在一些方面,所提供的实施方案涉及将多核苷酸的特定部分(如模板多核苷酸中含有编码嵌合受体或其部分的转基因序列的部分)靶向整合于基因组中在编码CD3 ζ (CD3 ζ)的内源CD247基因座处的特定位置(如靶位点或靶位置)。在一些方面,同源定向修复(HDR)可以介导转基因序列在靶位点处的位点特异性整合。在一些实施方案中,遗传破坏(例如,DNA断裂,如章节I.A中所述)和含有一个或多个同源臂(例如,含有与遗传破坏周围

的序列同源的核酸序列)的模板多核苷酸的存在可以诱导或引导HDR,其中同源序列用作DNA修复的模板。基于遗传破坏周围的内源基因序列与模板多核苷酸中包括的5'和/或3'同源臂之间的同源性,细胞DNA修复机构可以使用模板多核苷酸来修复DNA断裂并重合成(例如,拷贝)遗传破坏的位点处的遗传信息,从而将转基因序列有效地插入或整合于模板多核苷酸中的遗传破坏位点处或附近。在一些实施方案中,在编码CD3 ζ 的内源CD247基因座处的遗传破坏可以通过本文例如在章节I.A中描述的用于生成靶向遗传破坏的任何方法来生成。

[0441] 还提供多核苷酸,例如,本文所述的模板多核苷酸,以及包括此类多核苷酸的试剂盒。在一些实施方案中,所提供的多核苷酸和/或试剂盒可以用于本文所述的方法(例如,涉及HDR)中,以将编码嵌合受体的一部分的转基因序列靶向于编码CD3 ζ 的内源CD247基因座处。

[0442] 在一些实施方案中,模板多核苷酸是或包含如下多核苷酸,所述多核苷酸含有编码嵌合受体或其部分(例如,嵌合受体的一个或多个区域或结构域)的转基因(如外源或异源核酸序列),以及与编码CD3 ζ 的内源CD247基因座的内源基因组位点处或附近的序列同源的同源性序列(例如,同源臂)。在一些方面,所述模板多核苷酸中的转基因序列包含编码嵌合受体的一部分的核苷酸序列。在一些方面,在编码嵌合受体的一部分的转基因序列的靶向整合后,工程化细胞中的CD247基因座被修饰,使得修饰的CD247基因座含有所述转基因序列与所述内源CD247基因座的序列的融合物,所述融合物编码嵌合受体,例如,嵌合抗原受体(CAR)。

[0443] 在一些方面,模板多核苷酸是作为线性DNA片段引入或包含于载体中。在一些方面,诱导遗传破坏的步骤和用于靶向整合的步骤(例如,通过引入模板多核苷酸)是同时或依序进行的。

1. 同源定向修复 (HDR)

[0444] 在一些实施方案中,同源定向修复(HDR)可以用于将一个或多个核酸序列(例如,编码嵌合受体或其部分的转基因序列)靶向整合或插入于基因组中CD247基因座的一个或多个靶位点处。在一些实施方案中,核酸酶诱导的HDR可以用于改变靶序列,将转基因序列整合于特定靶位置处,和/或编辑或修复特定靶基因中的突变。

[0445] 靶位点处的核酸序列的改变可以通过HDR用外源提供的多核苷酸(例如,模板多核苷酸(也称为“供体多核苷酸”或“模板序列”))来进行。例如,模板多核苷酸提供靶序列的改变,如含于所述模板多核苷酸内的转基因序列的插入。在一些实施方案中,可以使用质粒或载体作为同源重组的模板。在一些实施方案中,可以使用线性DNA片段作为同源重组的模板。在一些实施方案中,可以使用单链模板多核苷酸作为通过靶序列与模板多核苷酸之间的同源定向修复的替代方法(例如,单链退火)来改变靶序列的模板。模板多核苷酸实现的靶序列的改变依赖于通过核酸酶(例如,靶向核酸酶,如CRISPR/Cas9)进行的切割。通过核酸酶进行的切割可以包括双链断裂或两个单链断裂。

[0446] 在一些实施方案中,“重组”包括两个多核苷酸之间交换遗传信息的过程。在一些实施方案中,“同源重组(HR)”包括这种交换的特化形式,其在例如经由同源定向修复机制修复细胞中的双链断裂期间进行。这个过程需要核苷酸序列同源性,使用模板多核苷酸对靶DNA(即,经历双链断裂的靶DNA,如内源基因中的靶位点)进行模板修复,并且因为其导致

遗传信息从模板多核苷酸转移至靶标而被不同地称为“非交叉型基因转换”或“短束基因转换”。在一些实施方案中,该转移可以涉及在断裂的靶标与模板多核苷酸之间形成的异源双链DNA的错配校正,和/或“合成依赖性链退火”(其中使用模板多核苷酸重合成将成为靶标的一部分的遗传信息),和/或相关过程。这种特化的HR通常导致靶分子序列的改变,使得模板多核苷酸的部分或全部序列掺入靶多核苷酸中。

[0447] 在一些实施方案中,经由非同源性依赖性机制将多核苷酸(如模板多核苷酸,例如,含有转基因的多核苷酸)的一部分整合至细胞基因组中。所述方法包括在细胞基因组中产生双链断裂(DSB),并使用核酸酶切割模板多核苷酸分子,使得模板多核苷酸整合于DSB的位点处。在一些实施方案中,模板多核苷酸是经由非同源性依赖性方法(例如,NHEJ)来整合。在体内切割后,模板多核苷酸可以以靶向方式整合至细胞基因组中的DSB位置处。模板多核苷酸可以包括用于产生DSB的一种或多种核酸酶的一个或多个相同靶位点。因此,模板多核苷酸可以通过用于切割期望整合至其中的内源基因的一种或多种相同核酸酶来切割。在一些实施方案中,模板多核苷酸包括与用于诱导DSB的核酸酶不同的核酸酶靶位点。如本文所述,可以通过任何已知方法或本文所述的任何方法(如ZFN、TALEN、CRISPR/Cas9系统或TtAgo核酸酶)来产生靶位点或靶位置的遗传破坏。

[0448] 在一些实施方案中,DNA修复机制可以通过核酸酶在以下之后诱导:(1)单一双链断裂;(2)两个单链断裂;(3)两个双链断裂,在靶位点的每一侧上发生断裂;(4)一个双链断裂和两个单链断裂,在靶位点的每一侧上发生双链断裂和两个单链断裂;(5)四个单链断裂,在靶位点的每一侧上发生一对单链断裂;或者(6)一个单链断裂。在一些实施方案中,使用单链模板多核苷酸,并且可以通过替代性HDR改变靶位点。

[0449] 模板多核苷酸实现的靶位点的改变依赖于通过核酸酶分子进行的切割。通过核酸酶进行的切割可以包括切口、双链断裂、或两个单链断裂,例如在靶位点处DNA的每条链上一个断裂。在靶位点上引入断裂之后,在断裂末端处进行切除,得到单链突出的DNA区域。

[0450] 在典范HDR中,引入双链模板多核苷酸,其包含靶位点的同源序列,所述同源序列将被直接掺入所述靶位点中,或者用作模板以插入转基因或校正所述靶位点的序列。在断裂处切除后,修复可以通过不同途径进行,例如通过双霍利迪连接体模型(double Holliday junction model)(或双链断裂修复(DSBR)途径)或合成依赖性链退火(SDSA)途径进行。

[0451] 在双霍利迪连接体模型中,发生靶位点的两个单链突出端链侵入至模板多核苷酸的同源序列中,导致形成具有两个霍利迪结点的中间体。随着从侵入链的末端合成新DNA以填充切除产生的缺口,结点移行。新合成的DNA的末端连接至切除的末端,并且结点被分解,导致在靶位点处的插入,例如在模板多核苷酸中插入转基因。与模板多核苷酸的交叉可以在结点分解后进行。

[0452] 在SDSA途径中,仅一个单链突出端侵入模板多核苷酸,并且从侵入链的末端合成新DNA,以填充切除产生的缺口。然后新合成的DNA与剩余的单链突出端退火,合成新DNA以填充于缺口中,并且连接所述链以产生修饰的DNA双链体。

[0453] 在替代性HDR中,引入单链模板多核苷酸,例如模板多核苷酸。在靶位点处用于改变所需靶位点的切口、单链断裂或双链断裂是由核酸酶分子介导的,并且进行在断裂处切除以露出单链突出端。掺入模板多核苷酸的序列以校正或改变DNA的靶位点通常是通过

SDSA途径进行的,如本文所述。

[0454] 在一些实施方案中,“替代性HDR”或替代性同源定向修复是指使用同源核酸(例如,内源同源序列,例如姐妹染色单体;或外源核酸,例如模板多核苷酸)修复DNA损伤的过程。替代性HDR与典范HDR的不同之处在于,所述过程利用与典范HDR不同的途径,并且可能被典型HDR介体RAD51和BRCA2抑制。替代性HDR也使用单链或带切口的同源核酸进行断裂的修复。在一些实施方案中,“典范HDR”或典范同源定向修复是指使用同源核酸(例如,内源同源序列,例如姐妹染色单体;或外源核酸,例如模板核酸)修复DNA损伤的过程。典范HDR通常在双链断裂处已经存在显著切除时发挥作用,形成DNA的至少一个单链部分。在正常细胞中,HDR通常涉及一系列步骤,如识别断裂、稳定断裂、切除、稳定单链DNA、形成DNA交叉中间体、分解交叉中间体以及连接。所述过程需要RAD51和BRCA2,并且同源核酸通常是双链的。除非另有指示,否则术语“HDR”在一些实施方案中涵盖典范HDR和替代性HDR。

[0455] 在一些实施方案中,双链切割是通过核酸酶来实现的,所述核酸酶是例如具有与HNH样结构域相关的切割活性和与RuvC样结构域(例如,N末端RuvC样结构域)相关的切割活性的Cas9分子,例如野生型Cas9。此类实施方案仅需要单一gRNA。

[0456] 在一些实施方案中,一个单链断裂或切口是通过具有切口酶活性的核酸酶分子(例如,Cas9切口酶)实现的。在靶位点处带切口的DNA可以是替代性HDR的底物。

[0457] 在一些实施方案中,两个单链断裂或切口是通过具有切口酶活性(例如,与HNH样结构域相关的切割活性或与N末端RuvC样结构域相关的切割活性)的核酸酶(例如,Cas9分子)实现的。此类实施方案通常需要两种gRNA,每个单链断裂的放置需要一种。在一些实施方案中,具有切口酶活性的Cas9分子切割gRNA所杂交的链,但不切割与gRNA所杂交的链互补的链。在一些实施方案中,具有切口酶活性的Cas9分子不切割gRNA所杂交的链,而是切割与gRNA所杂交的链互补的链。在一些实施方案中,切口酶具有HNH活性,例如RuvC活性失活的Cas9分子,例如具有D10处的突变(例如,D10A突变)的Cas9分子。D10A使RuvC失活;因此,Cas9切口酶(仅)具有HNH活性,并且将在gRNA所杂交的链(例如,互补链,其上不具有NGG PAM)上进行切割。在一些实施方案中,可以使用具有H840(例如,H840A)突变的Cas9分子作为切口酶。H840A使HNH失活;因此,Cas9切口酶(仅)具有RuvC活性,并且在非互补链(例如,具有NGG PAM并且其序列与gRNA相同的链)上进行切割。在一些实施方案中,Cas9分子是N末端RuvC样结构域切口酶,例如,Cas9分子包含N863处的突变,例如N863A。

[0458] 在使用切口酶和两种gRNA定位两个单链切口的一些实施方案中,一个切口在靶DNA的+链上,并且一个切口在-链上。PAM朝外。可以选择gRNA使得所述gRNA相隔约0-50、0-100或0-200个核苷酸。在一些实施方案中,在与两种gRNA的靶向结构域互补的靶序列之间没有重叠。在一些实施方案中,gRNA不重叠,并且相隔多达50、100或200个核苷酸。在一些实施方案中,使用两种gRNA可以例如通过减少脱靶结合来增加特异性(Ran等人,Cell 2013)。

[0459] 在一些实施方案中,可以使用单一切口来诱导HDR,例如替代性HDR。在本文中考虑,可以使用单一切口增加给定切割位点(例如,靶位点)处的HR与NHEJ的比率。在一些实施方案中,在靶位点处在与所述gRNA的靶向结构域互补的DNA的链中形成单链断裂。在一些实施方案中,在靶位点处在除了与所述gRNA的靶向结构域互补的链以外的DNA的链中形成单链断裂。

[0460] 在一些实施方案中,细胞可以采用其他DNA修复途径(如单链退火(SSA)、单链断裂

修复 (SSBR)、错配修复 (MMR)、碱基切除修复 (BER)、核苷酸切除修复 (NER)、链间交联 (ICL)、跨损伤合成 (TLS)、无误性复制后修复 (PRR) 来修复核酸酶产生的双链或单链断裂。

[0461] 靶向整合导致转基因 (例如,同源臂之间的序列) 整合至基因组中的CD247基因座中。转基因可以被整合于基因组中至少一个靶位点或位点中的一个位点处或附近的任何位置。在一些实施方案中,转基因被整合于至少一个靶位点中的一个位点处或附近,例如在切割位点上游或下游300、250、200、150、100、50、10、5、4、3、2、1个或更少碱基对内,如在靶位点任一侧的100、50、10、5、4、3、2、1个碱基对内,如在靶位点任一侧的50、10、5、4、3、2、1个碱基对内。在一些实施方案中,包含转基因的整合的序列不包括任何载体序列 (例如,病毒载体序列)。在一些实施方案中,整合的序列包括载体序列 (例如,病毒载体序列) 的一部分。

[0462] 双链断裂或一条链中的单链断裂 (如靶位点) 应与靶整合位点 (例如,用于靶向整合的位点) 足够接近,使得在所需区域中产生改变,如发生转基因的插入或突变的校正。在一些实施方案中,距离不超过10、25、50、100、200、300、350、400或500个核苷酸。在一些实施方案中,认为断裂应当与靶整合位点足够接近,使得断裂位于在末端切除期间经历外切核酸酶介导的去除的区域内。在一些实施方案中,靶向结构域被配置使得切割事件 (例如,双链或单链断裂) 定位于期望改变的区域 (例如,靶向插入位点) 的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、300、350、400或500个核苷酸内。断裂 (例如,双链或单链断裂) 可以定位于期望改变的区域 (例如,靶向插入位点) 的上游或下游。在一些实施方案中,断裂定位于期望改变的区域,例如由至少两个突变体核苷酸限定的区域内。在一些实施方案中,断裂的定位紧邻期望改变的区域,例如紧接靶整合位点的上游或下游。

[0463] 在一些实施方案中,单链断裂伴随通过第二gRNA分子定位的另外的单链断裂。例如,靶向结构域被配置使得切割事件 (例如,两个单链断裂) 定位于靶整合位点的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、300、350、400或500个核苷酸内。在一些实施方案中,第一和第二gRNA分子被配置使得,在指导Cas9切口酶时,单链断裂将伴随通过第二gRNA定位的另外的单链断裂,它们彼此足够接近以导致所需区域的改变。在一些实施方案中,第一和第二gRNA分子被配置使得,例如在Cas9是切口酶时,通过所述第二gRNA定位的单链断裂位于通过所述第一gRNA分子定位的断裂的10、20、30、40或50个核苷酸内。在一些实施方案中,两个gRNA分子被配置为将切割定位于不同链上的相同位置,或者彼此的几个核苷酸内,例如,从而基本上模拟双链断裂。

[0464] 在出于诱导HDR介导转基因的插入或校正的目的,gRNA (单分子 (或嵌合) 或模块化gRNA) 和Cas9核酸酶诱导双链断裂的一些实施方案中,切割位点 (如靶位点) 位于远离靶整合位点0至200bp之间 (例如,0至175、0至150、0至125、0至100、0至75、0至50、0至25、25至200、25至175、25至150、25至125、25至100、25至75、25至50、50至200、50至175、50至150、50至125、50至100、50至75、75至200、75至175、75至150、75至125、75至100bp)。在一些实施方案中,切割位点 (如靶位点) 位于远离靶向整合位点的0至100bp之间 (例如,0至75、0至50、0至25、25至100、25至75、25至50、50至100、50至75或75至100bp)。

[0465] 在一些实施方案中,可以通过使用切口酶产生具有突出端的断裂来促进HDR。在一些实施方案中,与例如NHEJ相反,突出端的单链性质可以增强细胞通过HDR修复断裂的可能性。

[0466] 具体而言,在一些实施方案中,通过选择第一gRNA和第二gRNA来促进HDR,所述第

一gRNA将第一切口酶靶向第一靶位点,并且所述第二gRNA将第二切口酶靶向第二靶位点,所述第二靶位点在与第一靶位点相对的DNA链上并且偏离第一切口。在一些实施方案中,gRNA分子的靶向结构域被配置为将切割事件定位于足够远离预选的核苷酸(例如,编码区的核苷酸),使得所述核苷酸不发生改变。在一些实施方案中,gRNA分子的靶向结构域被配置为将内含子切割事件定位于足够远离内含子/外显子边界或天然存在的剪接信号,以避免外显子序列的改变或不期望的剪接事件。在一些实施方案中,gRNA分子的靶向结构域被配置为定位于早期外显子中,以允许转基因序列框内整合于至少一个靶位点之一处或附近。

[0467] 在一些实施方案中,双链断裂可以伴随通过第二gRNA分子定位的另外的双链断裂。在一些实施方案中,双链断裂可以伴随通过第二gRNA分子和第三gRNA分子定位的两个另外的单链断裂。

[0468] 在一些实施方案中,两种gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置为将双链断裂定位于靶整合位点(例如,靶向整合的位点)的两侧上。

2. 模板多核苷酸

[0469] 在一些实施方案中,可以采用模板多核苷酸作为修复模板,例如,含有以下的多核苷酸:转基因,如外源或异源核酸序列,所述转基因包括编码嵌合受体、重组受体或其部分的一条或多条链的核苷酸序列,以及同源性序列(例如,同源臂),所述同源性序列与用于靶向整合的内源基因组位点处或附近的序列同源,细胞DNA修复过程(如同源重组)中涉及的分子和机构。在一些方面,与内源DNA中一个或多个靶位点处或附近的序列具有同源性的模板多核苷酸可以用于改变靶DNA(如内源CD247基因座处的靶位点)的结构,用于转基因或外源序列(例如,编码嵌合受体或其部分的外源核酸序列)的靶向插入。还提供用于本文提供的方法中的多核苷酸(例如,模板多核苷酸),例如,作为模板用于同源定向修复(HDR)介导的转基因序列的靶向整合。在一些实施方案中,所述多核苷酸包括编码嵌合受体或其部分的一条或多条链的核酸序列(如转基因);以及与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框的一个或多个区域同源的序列。在一些实施方案中,所述多核苷酸包括编码嵌合受体的一部分的核酸序列,所述嵌合受体包含细胞内区域,其中所述嵌合受体的所述部分包括所述嵌合受体的不完整细胞内区域(例如,小于整个的CD3 ζ 信号传导结构域);以及与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247的基因座的开放阅读框的一个或多个区域同源的序列。

[0470] 在一些实施方案中,模板多核苷酸含有一个或多个同源性序列(例如,同源臂)连接至和/或侧接包括编码嵌合受体或其部分的一条或多条链的核苷酸序列的转基因(外源或异源核酸序列)。在一些实施方案中,所述同源性序列用于将外源序列靶向于内源CD247基因座处。在一些实施方案中,模板多核苷酸包括在同源臂之间的核酸序列(转基因序列),用于插入或整合至细胞基因组中。模板多核苷酸中的转基因可以包含编码功能多肽(例如,cDNA)的一个或多个序列,其具有或不具有启动子或其他调节元件。

[0471] 在一些实施方案中,模板多核苷酸是指可以与能够引入遗传破坏的一种或多种药剂结合用于改变靶位点的结构的核酸序列。在一些实施方案中,模板多核苷酸通过同源定向修复事件改变靶位点的结构,例如插入转基因。

[0472] 在一些实施方案中,模板多核苷酸改变靶位点的序列,例如,导致将同源臂之间的转基因序列插入或整合至细胞基因组中。在一些方面,靶向整合导致转基因序列的编码部分与内源CD247基因座的开放阅读框的一个或多个外显子的框内整合,例如,与整合位点处的相邻外显子同框。例如,在一些情形中,框内整合导致表达内源开放阅读框的一部分和所述转基因编码的嵌合受体的所述部分。

[0473] 在一些实施方案中,模板多核苷酸包括与例如通过能够引入遗传破坏的一种或多种药剂切割的靶序列上的位点对应或同源的序列。在一些实施方案中,模板多核苷酸包括与在能够引入遗传破坏的第一药剂中切割的靶序列上的第一位点和在能够引入遗传破坏的第二药剂中切割的靶序列的第二位点二者对应或同源的序列。

[0474] 在一些实施方案中,模板多核苷酸包含以下组分:[5'同源臂]-[转基因序列(外源或异源核酸序列,例如,编码嵌合受体或其部分的一条或多条链)]-[3'同源臂]。在一些实施方案中,编码嵌合受体的核酸序列包含编码嵌合受体的一部分的转基因序列。同源臂提供重组至染色体中,由此将例如编码嵌合受体或其部分的转基因有效插入或整合至切割位点(如一个或多个靶位点)处或附近的基因组DNA中。在一些实施方案中,同源臂侧接遗传破坏的靶位点处的序列。

[0475] 在一些实施方案中,模板多核苷酸是双链的。在一些实施方案中,模板多核苷酸是单链的。在一些实施方案中,模板多核苷酸包含单链部分和双链部分。在一些实施方案中,模板多核苷酸包含于载体中。在一些实施方案中,模板多核苷酸是DNA。在一些实施方案中,模板多核苷酸是RNA。在一些实施方案中,模板多核苷酸是双链DNA。在一些实施方案中,模板多核苷酸是单链DNA。在一些实施方案中,模板多核苷酸是双链RNA。在一些实施方案中,模板多核苷酸是单链RNA。在一些实施方案中,模板多核苷酸包含单链部分和双链部分。在一些实施方案中,模板多核苷酸包含于载体中。

[0476] 在某些实施方案中,多核苷酸(例如,模板多核苷酸)含有和/或包括编码嵌合受体的一条或多条链的部分和/或片段(例如,CAR或其部分)的转基因。在特定实施方案中,转基因被靶向于位于编码CD3zeta(CD3 ζ)链或其片段的内源基因、基因座或开放阅读框内的一个或多个靶位点处。在一些实施方案中,转基因被靶向以供框内整合于内源CD247开放阅读框内,如以得到编码含有CD3zeta(CD3 ζ)链的完整、全部和/或全长CAR的编码序列。

[0477] 用于插入的多核苷酸也可以被称为“转基因”或“外源序列”或“供体”多核苷酸或分子。模板多核苷酸可以是单链和/或双链的DNA,并且可以以线性或环状形式引入细胞中。

[0478] 模板多核苷酸可以是单链和/或双链的DNA,并且可以以线性或环状形式引入细胞中。模板多核苷酸可以是单链和/或双链RNA,并且可以作为RNA分子(例如,RNA病毒的一部分)引入。还参见美国专利公开号20100047805和20110207221。模板多核苷酸也可以以DNA形式引入,可以将其以环状或线性形式引入细胞中。如果以线性形式引入,则可以通过已知方法保护模板多核苷酸的末端(例如,防止核酸外切降解)。例如,将一个或多个双脱氧核苷酸残基添加至线性分子的3'末端,和/或将自身互补的寡核苷酸连接至一个或两个末端。参见例如,Chang等人(1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls等人(1996) Science 272:886-889。用于保护外源多核苷酸免于降解的另外的方法包括但不限于添加一个或多个末端氨基以及使用修饰的核苷酸间连接(例如硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基)。如果以双链形式引入,则模板多核苷酸可以包括一个或多个核酸

酶靶位点,例如侧接要整合至细胞基因组中的转基因的核酸酶靶位点。参见例如,美国专利公开号20130326645。

[0479] 在一些实施方案中,双链模板多核苷酸包括长度大于1kb(例如在2与200kb之间、在2与10kb之间(或其间的任何值))的序列(也称为转基因)。例如,双链模板多核苷酸还包括至少一个核酸酶靶位点。在一些实施方案中,例如对于一对ZFN或TALEN,模板多核苷酸包括至少2个靶位点。通常,核酸酶靶位点在转基因序列的外部,例如在转基因序列的5'和/或3',用于切割转基因。一个或多个核酸酶切割位点(如一个或多个靶位点)可以针对任何一种或多种核酸酶。在一些实施方案中,含于双链模板多核苷酸中的一个或多个核酸酶靶位点是针对用于切割内源靶标的相同的一种或多种核酸酶,将切割的模板多核苷酸经由非同源性依赖性方法整合至所述内源靶标中。

[0480] 在一些实施方案中,模板多核苷酸是单链核酸。在一些实施方案中,模板多核苷酸是双链核酸。在一些实施方案中,模板多核苷酸包含例如一个或多个核苷酸的核苷酸序列,其将被添加至靶DNA或将作为靶DNA中变化的模板。在一些实施方案中,模板多核苷酸包含可以用于修饰靶位点的核苷酸序列。在一些实施方案中,模板多核苷酸包含例如一个或多个核苷酸的核苷酸序列,其对应于靶DNA(例如,靶位点)的野生型序列。

[0481] 在一些实施方案中,模板多核苷酸是线性双链DNA。长度可以为例如约200至约5000个碱基对,例如,约200、300、400、500、600、700、800、900、1000、1200、1400、1600、1800、2000、2500、3000、4000或5000个碱基对。长度可以为例如至少200、300、400、500、600、700、800、900、1000、1200、1400、1600、1800、2000、2500、3000、4000或5000个碱基对。在一些实施方案中,长度不大于200、300、400、500、600、700、800、900、1000、1200、1400、1600、1800、2000、2500、3000、4000或5000个碱基对。在一些实施方案中,双链模板多核苷酸的长度为约160个碱基对,例如,约200至4000、300至3500、400至3000、500至2500、600至2000、700至1900、800至1800、900至1700、1000至1600、1100至1500或1200至1400个碱基对。

[0482] 含于本文所述的模板多核苷酸上的转基因可以使用已知的标准技术如PCR从质粒、细胞或其他来源分离。用于使用的模板多核苷酸可以包括各种类型的拓扑学,包括环状超螺旋的、环状松弛的、线性的等。可替代地,它们可以使用标准寡核苷酸合成技术化学合成。另外,模板多核苷酸可以被甲基化或缺乏甲基化。模板多核苷酸可以呈细菌或酵母人工染色体(BAC或YAC)的形式。

[0483] 模板多核苷酸可以是线性单链DNA。在一些实施方案中,模板多核苷酸是(i)可以与靶DNA的带切口的链退火的线性单链DNA,(ii)可以与靶DNA的完整链退火的线性单链DNA,(iii)可以与靶DNA的转录链退火的线性单链DNA,(iv)可以与靶DNA的非转录链退火的线性单链DNA,或多于一种的前述项。

[0484] 长度可以为例如约200至5000个核苷酸,例如,约200、300、400、500、600、700、800、900、1000、1200、1400、1600、1800、2000、2500、3000、4000或5000个核苷酸。长度可以为例如至少200、300、400、500、600、700、800、900、1000、1200、1400、1600、1800、2000、2500、3000、4000或5000个核苷酸。在一些实施方案中,长度不大于200、300、400、500、600、700、800、900、1000、1200、1400、1600、1800、2000、2500、3000、4000或5000个核苷酸。在一些实施方案中,单链模板多核苷酸的长度为约160个核苷酸,例如,约200至4000、300至3500、400至3000、500至2500、600至2000、700至1900、800至1800、900至1700、1000至1600、1100至1500

或1200至1400个核苷酸。

[0485] 在一些实施方案中,模板多核苷酸是环状双链DNA,例如,质粒。在一些实施方案中,模板多核苷酸在转基因和/或靶位点的任一侧上包含约500至1000个同源性碱基对。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000个同源性碱基对。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000个同源性碱基对。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含不超过10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000个同源性碱基对。

[0486] 在一些实施方案中,模板多核苷酸中的转基因序列包含与一个或多个同源臂中所含的CD247基因座的开放阅读框的一个或多个外显子同框的核苷酸序列。在一些实施方案中,所述开放阅读框的一个或多个区域是或包含如下序列,所述序列位于CD247基因座的开放阅读框的外显子8的上游。在一些实施方案中,所述开放阅读框的一个或多个区域是或包含如下序列,所述序列位于CD247基因座的开放阅读框的外显子3的上游。在一些实施方案中,所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列包括CD247基因座的开放阅读框的外显子3。在一些实施方案中,所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列包括CD247基因座的开放阅读框的外显子2的至少一部分。

[0487] 在一些实施方案中,模板多核苷酸中的一个或多个同源臂不包含CD247基因座的开放阅读框的外显子1的全长。在一些实施方案中,一个或多个同源臂不包含CD247基因座的开放阅读框的外显子1和/或不包含外显子2的全长。

a. 转基因序列

[0488] 在一些实施方案中,模板多核苷酸含有编码嵌合受体或其部分的一条或多条链(如嵌合受体(如本文例如在章节III.B中所述的任何嵌合受体)的一个或多个区域或结构域,或这种嵌合受体一个或多个区域或结构域或链)的转基因序列或外源序列。在一些方面,在含有与一个或多个同源臂连接的转基因序列的模板多核苷酸存在下进行的HDR得到编码嵌合受体的修饰的CD247基因座,所述一个或多个同源臂与内源CD247基因座处的靶位点附近的序列同源。在一些实施方案中,转基因序列编码嵌合受体或其部分,如嵌合受体的一个或多个结构域、区域或链,包括细胞外结合区、跨膜结构域和/或细胞内区域的一部分。在一些实施方案中,转基因序列不包含内含子。在一些方面,转基因序列是对于T细胞(任选地人T细胞)的内源基因组CD247基因座的开放阅读框为外源或异源的序列。

[0489] 在一些方面,由转基因序列编码的嵌合受体是或包含功能性非T细胞受体(非TCR)抗原受体。在一些实施方案中,嵌合受体是嵌合抗原受体(CAR)。在一些实施方案中,转基因序列编码本文例如在章节III.B中所述的任何嵌合受体或其部分。在一些方面,在将转基因序列整合至内源CD247基因座中之后,所得修饰的CD247基因座编码嵌合受体,如本文例如在章节III.B中所述的任何嵌合受体。在一些实施方案中,转基因序列编码本文例如在章节III.B中所述的嵌合受体的一部分,如含有包含CD3 ζ 链或其片段(例如,CD3 ζ 链的细胞内区域)的细胞内区域的嵌合受体的一部分。在一些实施方案中,转基因序列编码作为多链CAR

(如本文在章节III.B.2中所述的多链CAR)的嵌合受体的链(如含有CD3 ζ 链或其片段的多链CAR的链)的一部分。

[0490] 在一些实施方案中,由修饰的CD247基因座编码的嵌合受体包含例如含有CD3 ζ 信号传导结构域的细胞内区域,并且转基因序列编码嵌合受体的一部分,所述部分不包括所述嵌合受体的完整细胞内区域。在一些实施方案中,由修饰的CD247基因座编码的嵌合受体包含CD3 ζ 信号传导结构域,并且转基因序列不编码整个CD3 ζ 信号传导结构域。在一些实施方案中,在将转基因序列整合至内源CD247基因座中之后,CD3 ζ 链的至少一部分(如CD3 ζ 信号传导结构域的片段或整个CD3 ζ 信号传导结构域)是由内源CD247基因座的开放阅读框序列或其部分序列编码。在一些方面,含有编码嵌合受体的一部分的核酸序列和一个或多个同源臂的模板多核苷酸一起包含编码嵌合受体的细胞内区域(例如,包含CD3 ζ 信号传导结构域)的核苷酸序列的至少一个片段,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述细胞内区域的至少一部分包含由CD247基因座的开放阅读框或其部分序列编码的CD3 ζ 信号传导结构域或其片段。

[0491] 在一些方面,被插入或整合于基因组中的靶位置处的转基因序列(其为编码嵌合受体或其部分的一条或多条链的目的核酸序列)(包括编码和/或非编码序列和/或其部分编码序列)也可以称为“转基因”、“转基因序列”、“外源核酸序列”、“异源序列”或“供体序列”。在一些方面,转基因是对于T细胞(例如,人T细胞)的内源基因组序列(如基因组中特定靶基因座或靶定位处的内源基因组序列)为外源或异源的核酸序列。在一些方面,转基因是与T细胞(例如,人T细胞)的靶基因座或靶定位处的内源基因组序列相比被修饰或不同的序列。在一些方面,转基因是源自不同的基因、物种和/或来源的核酸序列,或者与来自不同基因、物种和/或来源的核酸序列相比被修饰的核酸序列。在一些方面,转基因是源自相同物种的不同基因座(例如,不同基因组区域或不同基因)的序列的序列。在一些方面,示范性嵌合受体包括本文例如在章节III.B中所述的任一种。

[0492] 在一些实施方案中,核酸酶诱导的HDR导致插入转基因(也称为“外源序列”或“转基因序列”),用于表达用于靶向插入的转基因。模板多核苷酸序列通常与其所在的基因组序列不同。模板多核苷酸序列可以含有侧翼为两个同源性区域的非同源序列,以允许在目的位置进行有效HDR。另外,模板多核苷酸序列可以包含载体分子,所述载体分子含有与细胞染色质中的目的区域不同源的序列。模板多核苷酸序列可以含有与细胞染色质具有同源性的几个不连续区域。例如,对于通常在目的区域中不存在的序列的靶向插入,所述序列可以存在于转基因中并且侧翼为与目的区域中的序列具有同源性的区域。

[0493] 在一些方面,转基因是嵌合序列,其包含通过接合来自不同基因、物种和/或来源的不同核酸序列生成的序列。在一些方面,转基因含有接合或连接的编码来自不同基因、编码序列或外显子或其部分的不同区域或结构域或其部分的核苷酸序列。在一些方面,用于靶向整合的转基因序列编码多肽(例如,融合多肽)或其片段。

[0494] 在一些方面,由转基因编码的多肽是嵌合多肽。在一些方面,转基因还含有非编码调节或控制序列,例如,允许、调节和/或调整所编码多肽或其片段的表达所需的序列,或者修饰多肽所需的序列。在一些实施方案中,如果转基因源自基因组序列,则如与基因组中的相应核酸相比,转基因不包含内含子或者缺少一个或多个内含子。在一些实施方案中,转基因序列不包含内含子。在一些实施方案中,转基因含有编码嵌合受体或其部分的序列,其中

转基因序列的全部或部分例如针对在人细胞中表达而进行密码子优化。

[0495] 在一些实施方案中,转基因序列(包括编码区和非编码区)的长度为在或在约100至约10,000个碱基对之间,如约100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、3500、4000、4500、5000、6000、7000、8000、9000或10000个碱基对。在一些实施方案中,转基因序列的长度受到可以制备、合成或组装和/或引入细胞中的多核苷酸的最大长度或病毒载体的容量的限制。在一些方面,转基因序列的长度可以根据模板多核苷酸的最大长度和/或所需的一个或多个同源臂的长度而变化。

[0496] 在一些实施方案中,遗传破坏诱导的HDR导致转基因序列插入或整合于基因组中的靶位置处。模板多核苷酸序列通常与其被靶向的基因组序列不同。模板多核苷酸序列可以含有侧翼为两个同源性区域的转基因序列,以允许在目的位置进行有效HDR。模板多核苷酸序列可以含有与基因组DNA具有同源性的几个不连续区域。例如,对于通常在目的区域中不存在的序列的靶向插入,所述序列可以存在于转基因中并且侧翼为与目的区域中的序列具有同源性的区域。在一些实施方案中,转基因序列编码嵌合受体或其部分,例如,细胞外结合区、跨膜结构域和/或细胞内区域的一部分中的一个或多个。

[0497] 在一些方面,在通过HDR靶向整合转基因后,细胞基因组含有包含编码嵌合受体或其部分的核酸序列的修饰的CD247基因座。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有转基因与内源CD247基因座的开放阅读框或其部分序列的融合物,例如,基因融合物。在一些方面,所述融合物是关于来自不同来源的两种或更多种核酸分子的融合物:例如,由于HDR介导的靶向整合而发生的转基因序列与基因组DNA的融合。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有整合至内源CD247基因座的开放阅读框内的位点中的转基因。在一些实施方案中,在靶向整合后,含有转基因序列与内源CD247基因座的序列的融合物(例如,基因融合物)的修饰的CD247基因座编码嵌合受体,例如,嵌合抗原受体(CAR)。在一些方面,嵌合受体的某些部分是由转基因编码,并且所述嵌合受体的其他部分是由内源CD247基因座的开放阅读框或其部分序列编码。在一些实施方案中,转基因序列包含与一个或多个同源臂中所含的CD247基因座的开放阅读框的一个或多个外显子同框的核苷酸序列。在一些方面,整个嵌合受体是由转基因序列编码。在一些方面,转基因序列还含有编码其他分子或多链嵌合受体的其他链的核苷酸序列,和/或调节或控制元件(例如,外源启动子)和/或多顺反子元件。在一些方面,示例性嵌合受体包括本文例如在章节III.B中所述的任一种。

[0498] 在一些方面,由于靶向整合的转基因序列包括编码嵌合受体的序列,所述嵌合受体是嵌合受体,如嵌合抗原受体(CAR)或嵌合自身抗体受体(CAAR)。在一些方面,转基因含有接合或连接的编码嵌合受体的不同区域或结构域或部分的核苷酸序列,所述区域或结构域或部分可以来自不同基因、编码序列或外显子或其部分。

[0499] 在一些实施方案中,转基因序列编码嵌合受体的各个区域、结构域或链的全部或一些或一部分,如章节III.B中所述的嵌合受体或各个区域、结构域或链。在一些实施方案中,转基因序列编码嵌合受体的各个区域、结构域或链的一部分。在一些实施方案中,转基因序列编码多链嵌合受体的多肽链或其部分。在一些实施方案中,所编码的嵌合受体含有CAR的各个区域或结构域。在一些实施方案中,所编码的嵌合受体含有一个或多个区域或结构域,如细胞外区域(例如,含有一个或多个细胞外结合结构域和/或间隔子)、跨膜结构域

和/或细胞内区域(例如,含有初级信号传导区域或结构域和/或一个或多个共刺激信号传导结构域)中的一个或多个。在一些方面,所编码的CAR还含有其他结构域,如多聚化结构域或接头。

[0500] 在一些实施方案中,转基因包括编码细胞内区域的核苷酸序列。在一些实施方案中,转基因还包括编码跨膜区或膜结合区的核苷酸序列。在一些实施方案中,转基因还包括编码细胞外区域的核苷酸序列。在一些实施方案中,嵌合受体包含细胞外区域和/或跨膜结构域。在一些实施方案中,转基因序列包含编码嵌合受体的一个或多个区域的核苷酸序列,任选地其中所述转基因序列包含编码细胞外区域、跨膜结构域和/或细胞内区域的一部分中的一个或多个的核苷酸序列。

[0501] 在一些方面,在转基因中,编码细胞外区域的核苷酸序列位于信号序列与编码间隔子的核苷酸之间。在一些方面,在转基因中,编码细胞外多聚化结构域的核苷酸序列位于编码结合结构域的核苷酸序列与编码间隔子的核苷酸序列之间。在一些方面,编码间隔子的核苷酸序列位于编码结合结构域的核苷酸序列与编码跨膜结构域的核苷酸序列之间。

[0502] 在一些实施方案中,转基因以5'至3'顺序包括编码细胞外区域的核苷酸序列、编码跨膜结构域(或膜结合结构域)的核苷酸序列和编码细胞内区域的核苷酸序列。在一些实施方案中,转基因以5'至3'顺序包括编码跨膜结构域(或膜结合结构域)的核苷酸序列和编码细胞内区域的核苷酸序列。在一些实施方案中,转基因以5'至3'顺序包括编码细胞外区域的核苷酸序列、编码跨膜结构域的核苷酸序列和编码细胞内区域的核苷酸序列。

[0503] 在一些方面,嵌合受体的一些区域或结构域是由转基因的序列(即,异源或外源序列)编码。例如,转基因序列可以包括编码细胞外区域、跨膜结构域和可以包含共刺激信号传导结构域的细胞内区域以及其他结构域或其部分中的一个或多个的核苷酸序列。在一些方面,嵌合受体的一些区域或结构域是由CD247基因座的内源序列的序列编码。例如,CD3 ζ 链或其片段的全部或一部分可以由内源CD247基因座的开放阅读框序列或其部分序列编码,和/或CD3 ζ 链的一部分可以由转基因编码。因此,在转基因的靶向整合后,所编码的嵌合受体是由包含整合的转基因和CD247基因座处的内源序列的基因融合物编码。

[0504] 在一些方面,细胞外区域可以包括结合结构域和/或间隔子。在一些实施方案中,细胞外区域可以包括细胞外多聚化结构域。在一些方面,由转基因编码的细胞内区域包含一个或多个共刺激结构域和/或多聚化结构域和其他结构域。在一些实施方案中,由转基因序列编码的细胞内区域包含小于全长的CD3 ζ 链或CD3 ζ 链的一部分。在一些方面,转基因不含编码CD3 ζ 链或其片段的核苷酸序列。在一些实施方案中,转基因序列还包括编码信号肽的信号序列、调节或控制元件(如启动子)和/或一个或多个多顺反子元件(例如,核糖体跳跃元件或内部核糖体进入位点(IRES))。在一些实施方案中,信号序列可以位于编码细胞外区域的核苷酸序列的5'。在一些实施方案中,转基因还包含一个或多个多顺反子元件,例如,核糖体跳跃序列和/或内部核糖体进入位点(IRES)。在一些方面,转基因还包括通常位于转基因序列的最5'部分(例如,信号序列的5')的调节或控制元件(如启动子)。在一些方面,在多核苷酸的转基因部分中可以包括编码一种或多种另外的分子的核苷酸序列。在一些方面,编码一种或多种另外的分子的核苷酸序列位于编码嵌合受体的一个或多个区域或结构域或链的核苷酸序列的5'。在一些方面,编码一种或多种另外的分子或另外地结构域、区域或链的核苷酸序列位于编码嵌合受体的一个或多个区域的核苷酸序列的上游。

[0505] 由转基因序列编码的示例性区域或结构域示于下文中,并且还包括本文的章节III.B中所述的任何区域或结构域。在特定实施方案中,转基因序列包括编码以下的核苷酸序列:信号肽、结合结构域(例如抗原结合结构域,如scFv)、间隔子、跨膜结构域和含有共刺激信号传导结构域的细胞内信号传导区域以及CD3 ζ 链或CD3 ζ 链的一部分。

(i) 信号序列

[0506] 在一些实施方案中,转基因包括编码信号肽的信号序列。在一些方面,信号序列可以编码异源或非天然信号肽,例如,来自不同基因或物种的信号肽或者与内源CD247基因座的信号肽不同的信号肽。在一些方面,示例性信号序列包括SEQ ID NO:24中所示的GMCSFR α 链的信号序列以及编码SEQ ID NO:25中所示的信号肽或SEQ ID NO:26中所示的CD8 α 信号肽的信号序列。在所表达嵌合受体的成熟形式中,从所述多肽的其余部分切割信号序列。在一些方面,信号序列位于调节或控制元件(例如,启动子,如异源启动子,例如,并非源自CD247基因座的启动子)的3'。在一些方面,信号序列位于一个或多个多顺反子元件(例如,编码核糖体跳跃序列和/或内部核糖体进入位点(IRES)的核苷酸序列)的3'。在一些方面,信号序列可以位于转基因中编码细胞外区域的一个或多个组分的核苷酸序列的5'。在一些实施方案中,信号序列位于转基因中存在的最5'区域,并且与同源臂之一连接。在一些方面,由转基因序列编码的信号序列包括本文例如在章节III.B中所述的任何信号序列。

(ii) 结合结构域

[0507] 在一些实施方案中,转基因编码对特定抗原(或配体)具有特异性的嵌合受体(如CAR)的一部分,所述抗原如在特定细胞类型的表面上表达的抗原。在一些实施方案中,如与例如健康细胞或组织中的正常或非靶向细胞或组织相比,所述抗原在疾病或病症的细胞(例如,肿瘤或病原细胞)上选择性表达或过表达。

[0508] 在一些方面,转基因编码嵌合受体的细胞外区域。在一些实施方案中,转基因序列编码细胞外结合结构域,如特异性结合抗原或配体的结合结构域。

[0509] 在一些实施方案中,所述结合结构域是或包含多肽、配体、受体、配体结合结构域、受体结合结构域、抗原、表位、抗体、抗原结合结构域、表位结合结构域、抗体结合结构域、标签结合结构域或前述任一种的片段。在其他实施方案中,所述抗原在正常细胞上表达和/或在工程化细胞上表达。在一些方面,所述抗原由结合结构域(如配体结合结构域或抗原结合结构域)识别。在一些方面,转基因编码含有一个或多个结合结构域的细胞外区域。在一些实施方案中,由转基因编码的示例性结合结构域包括抗体及其抗原结合片段,包括scFv或sdAb。在一些实施方案中,抗原结合片段包含通过柔性接头接合的抗体可变区。

[0510] 在一些实施方案中,结合结构域是或包含单链可变片段(scFv)。在一些实施方案中,结合结构域是或包含单一结构域抗体(sdAb)。在一些实施方案中,结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达。在一些实施方案中,所述疾病、障碍或病症是感染性疾病或障碍、自身免疫疾病、炎性疾病或者肿瘤或癌症。在一些实施方案中,靶抗原是肿瘤抗原。

[0511] 由转基因序列编码的示例性抗原和抗原结合结构域或配体结合结构域包括本文的章节III.B.1中所述的那些。在一些方面,所编码的嵌合受体含有结合结构域,所述结合结构域是或包含特异性识别作为主要组织相容性复合物(MHC)-肽复合物存在于细胞表面上的细胞内抗原(如肿瘤相关抗原)的TCR样抗体或其片段(如scFv)。在一些方面,转基因序

列可以编码作为TCR样抗体或其片段的结合结构域。因此,所编码的嵌合受体是TCR样CAR,如本文在章节III.B.1中所述的任一种。在一些实施方案中,结合结构域是多特异性(如双特异性)结合结构域。在一些实施方案中,所编码的嵌合受体含有作为与自身抗体结合的抗原的结合结构域。在一些实施方案中,嵌合受体是嵌合自身抗体受体(CAAR),如本文在章节III.B.3中所述的任一种。

[0512] 在一些方面,编码一个或多个结合结构域的核苷酸序列可以位于转基因中的信号序列(如果存在)的3'。在一些方面,编码一个或多个结合结构域的核苷酸序列可以位于转基因中编码一种或多种调节或控制元件的核苷酸序列的3'。在一些方面,编码一个或多个结合结构域的核苷酸序列可以位于转基因中编码间隔子(如果存在)的核苷酸序列的5'。在一些方面,编码一个或多个结合结构域的核苷酸序列可以位于转基因中编码跨膜结构域的核苷酸序列的5'。

(iii) 间隔子和跨膜结构域

[0513] 在一些实施方案中,转基因包括编码间隔子的序列和/或编码跨膜结构域或其部分的序列。在一些实施方案中,所编码的嵌合受体的细胞外区域包含间隔子,任选地其中所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。在一些方面,间隔子和/或跨膜结构域可以连接含有配体(例如,抗原)结合结构域的细胞外部分与嵌合受体的其他区域或结构域,如细胞内区域(例如,含有一个或多个共刺激信号传导结构域、细胞内多聚化结构域和/或CD3 ζ 链或其片段)。

[0514] 在一些实施方案中,转基因还包括编码将抗原结合结构域与跨膜结构域隔开的间隔子和/或铰链区的核苷酸序列。在一些方面,间隔子可以是或包括免疫球蛋白恒定区的至少一部分或其变体或修饰形式,如铰链区(例如,IgG4铰链区)和/或C_H1/C_L和/或Fc区。在一些实施方案中,恒定区或部分是人IgG(如IgG4或IgG1)的。在一些方面,恒定区的所述部分用作结合结构域(例如,scFv)与跨膜结构域之间的间隔子区域。可以由转基因编码的示例性间隔子包括单独的IgG4铰链、与C_H2和C_H3结构域连接的IgG4铰链或者与C_H3结构域连接的IgG4铰链,以及以下文献中所述的那些:Hudecek等人(2013) Clin. Cancer Res., 19:3153; Hudecek等人(2015) Cancer Immunol Res. 3(2):125-135或国际专利申请公开号W0 2014031687,或者在本文的章节III.B.1中所述的任一种。

[0515] 在一些方面,编码间隔子的核苷酸序列可以位于转基因中编码一个或多个结合结构域的核苷酸序列的3'。在一些方面,编码间隔子的核苷酸序列可以位于转基因中编码跨膜结构域的核苷酸序列的5'。在一些实施方案中,编码间隔子的核苷酸序列位于编码一个或多个结合结构域的核苷酸序列与编码跨膜结构域的核苷酸序列之间。

[0516] 在一些实施方案中,转基因编码跨膜结构域,所述跨膜结构域可以连接例如含有一个或多个结合结构域和/或间隔子的细胞外区域与例如含有一个或多个共刺激信号传导结构域、细胞内多聚化结构域和/或CD3 ζ 链或其片段的细胞内区域。在一些实施方案中,转基因包含编码跨膜结构域的核苷酸序列,任选地其中所述跨膜结构域是人的或者包含来自人蛋白的序列。在一些实施方案中,跨膜结构域是或包含源自CD4、CD28或CD8、任选地源自人CD4、人CD28或人CD8的跨膜结构域。在一些实施方案中,跨膜结构域是或包含源自CD28、任选地源自人CD28的跨膜结构域。

[0517] 在一些实施方案中,编码跨膜结构域的核苷酸序列与编码细胞外区域的核苷酸序

列融合。在一些实施方案中,编码跨膜结构域的核苷酸序列与编码细胞内区域的核苷酸序列融合。在一些方面,编码跨膜结构域的核苷酸序列可以位于转基因中编码一个或多个结合结构域和/或间隔子的核苷酸序列的3'。在一些方面,编码跨膜结构域的核苷酸可以位于转基因中编码细胞内区域(例如,含有一个或多个共刺激信号传导结构域、细胞内多聚化结构域和/或CD3 ζ 链或其片段)的核苷酸序列的5'。在一些方面,由转基因序列编码的跨膜结构域包括本文例如在章节III.B.1中所述的任何跨膜结构域。

[0518] 在一些实施方案中,在所编码的嵌合受体包含含有CD3 ζ 链的细胞内区域但不包含跨膜结构域和/或细胞外区域的情形中,转基因可以包括编码膜结合结构域的核苷酸序列,如本文例如在章节III.B中所述的任一种。

(iv) 细胞内区域

[0519] 在一些实施方案中,转基因包括编码细胞内区域的核苷酸序列。在一些方面,所述细胞内区域包含一个或多个次级或共刺激信号传导区域。在一些方面,编码跨膜结构域的核苷酸序列可以位于转基因中编码一个或多个结合结构域和/或间隔子的核苷酸序列的3'。在一些方面,编码一个或多个共刺激信号传导结构域的核苷酸序列可以位于编码CD3 ζ 链的一部分的核苷酸序列的5'。在一些方面,编码一个或多个共刺激信号传导结构域的核苷酸序列是转基因中的最3'区域,其随后与同源臂序列之一(例如,3'同源臂序列)连接。例如,在一些情形中,转基因不包括编码CD3 ζ 链或其片段的核苷酸序列,并且因此转基因中与同源臂连接的最3'区域是编码一个或多个共刺激信号传导结构域的核苷酸序列。在一些方面,编码一个或多个共刺激信号传导结构域的核苷酸序列可以位于转基因中编码跨膜结构域的核苷酸序列的3'。在一些方面,由转基因序列编码的共刺激信号传导区域或者CD3 ζ 或其部分包括本文例如在章节III.B.1中所述的任何共刺激信号传导区域或者CD3 ζ 或其部分。

(a) 共刺激信号传导结构域

[0520] 在一些实施方案中,转基因包含编码细胞内区域的一部分的核苷酸序列,所述细胞内区域可以包括一个或多个共刺激信号传导结构域。在一些实施方案中,一个或多个共刺激信号传导结构域包含T细胞共刺激分子的细胞内信号传导结构域或其信号传导部分,任选地其中所述T细胞共刺激分子或其信号传导部分是人的。

[0521] 在一些实施方案中,一个或多个共刺激信号传导结构域包含T细胞共刺激分子的细胞内信号传导结构域或其信号传导部分。在一些实施方案中,T细胞共刺激分子或其信号传导部分是人的。在一些实施方案中,由转基因编码的示例性共刺激信号传导结构域包括来自一种或多种共刺激受体(如CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D、ICOS和/或其他共刺激受体)的信号传导区域或结构域,如本文在本文的章节III.B中所述的任一种。在一些实施方案中,一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。在一些实施方案中,一个或多个共刺激信号传导结构域包含人CD28、人4-1BB、人ICOS的信号传导结构域或其信号传导部分。在一些实施方案中,一个或多个共刺激信号传导结构域包含人4-1BB的细胞内信号传导结构域。

(b) CD3 ζ 链或CD3 ζ 链的一部分

[0522] 在一些实施方案中,转基因包括编码CD3 ζ 链或其片段(如CD3 ζ 的胞质结构域或其

部分)的核苷酸序列。在一些实施方案中,转基因仅编码CD3 ζ 链的一部分。在一些方面,在将转基因整合至内源CD247基因座中之后,所得修饰的CD247基因座编码含有CD3 ζ 链或其片段(如CD3 ζ 的细胞内区域)的嵌合受体(例如,CAR)。在一些实施方案中,在由引入所述多核苷酸的细胞表达时,嵌合受体能够经由CD3 ζ 信号传导结构域进行信号传导。在一些实施方案中,所编码的嵌合受体是如本文例如在章节III.B中所述的任一种。

[0523] 在一些方面,多核苷酸的转基因序列部分不含编码全长CD3 ζ 链的核苷酸序列。因此,在一些方面,所编码的嵌合受体中CD3 ζ 链的至少一部分是由存在于内源CD247基因座中的序列编码。在一些实施方案中,转基因序列不包括编码CD3 ζ 链的任何部分的核酸序列。在一些实施方案中,转基因仅编码CD3 ζ 链的不超过1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸的一部分。在一些方面,在转基因序列的整合后,嵌合受体的编码CD3 ζ 链或其片段的一些或所有核酸序列源自或来源于内源CD247基因座的开放阅读框序列或其部分序列。在一些实施方案中,转基因不包括编码CD3 ζ 链或其片段的核苷酸序列,或者仅包括编码细胞内区域中不包括CD3 ζ 链或其片段的一部分(part)或一部分(portion)的核苷酸序列。

[0524] 在一些实施方案中,转基因包括编码小于全长的CD3 ζ 链或CD3 ζ 链的一部分的核苷酸序列。在一些方面,转基因包括编码CD3 ζ 链的细胞内区域或其部分序列的核苷酸序列。在一些实施方案中,转基因在编码CD3 ζ 链的部分(例如,CD3 ζ 链的细胞内区域)的序列中不包含内含子。

[0525] 在一些实施方案中,转基因的靶向整合生成转基因与CD247基因座的内源序列的基因融合物,它们一起编码功能性CD3 ζ 链,例如,CD3 ζ 链中能够介导、激活或刺激初级胞质或细胞内信号的部分,例如,CD3 ζ 链的胞质结构域,或者CD3 ζ 链中包括免疫受体酪氨酸激活基序(ITAM)的部分。

[0526] 在一些方面,由转基因与CD247基因座的内源序列的基因融合物编码的示例性CD3 ζ 链或其片段包括CD3 ζ 链的细胞内区域的全部或一部分,例如,SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基52-164或SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基52-163,或者展现与SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基52-164或SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基52-163或其部分序列的至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。在一些方面,由转基因与CD247基因座的内源序列的基因融合物编码的示例性CD3 ζ 链或其片段包括SEQ ID NO:13、14或15中所示的氨基酸序列,或者展现与SEQ ID NO:13、14或15或其部分序列的至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。由转基因与CD247基因座的内源序列的基因融合物编码的示例性CD3 ζ 链或其片段包括CD3 ζ 链的一个或多个ITAM结构域,例如,SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-128或131-159或SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-127或130-158,或者含有来自CD3 ζ 链的一个或多个ITAM结构域并且展现与SEQ ID NO:73或SEQ ID NO:75中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-127或130-158的至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。

[0527] 在一些实施方案中,在由引入所述多核苷酸的细胞表达时,嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。在一些任何实施方案中,所编码的嵌合受体(例如,由修饰的CD247基因座编码的嵌合受体)包含能够进行信号传导或信号转导的CD3 ζ 信号传导结构域,如整个CD3 ζ 信号传导结构域。在一些任何实施方案中,整个CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列。在一些方面,全部或完整CD3 ζ 信号传导结构域(例如,整个CD3 ζ 信号传导结构域)或CD3 ζ 信号传导结构域的一部分(例如,所述CD3 ζ 信号传导结构域的片段)是由所提供的工程化细胞中细胞(例如,T细胞)的内源CD247基因座的开放阅读框编码。在一些实施方案中,由编码嵌合受体的修饰的CD247基因座编码的完整细胞内信号传导结构域的CD3 ζ 信号传导结构域(例如,整个CD3 ζ 信号传导结构域)包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。在任何此类例子中,CD3 ζ 信号传导结构域包含SEQ ID NO:13中所示的序列。在任何此类例子中,CD3 ζ 信号传导结构域由SEQ ID NO:13中所示的序列组成或基本上由其组成。在任何此类例子中,CD3 ζ 信号传导结构域包含SEQ ID NO:14中所示的序列。在任何此类例子中,CD3 ζ 信号传导结构域由SEQ ID NO:14中所示的序列组成或基本上由其组成。在任何此类例子中,CD3 ζ 信号传导结构域包含SEQ ID NO:15中所示的序列。在任何此类例子中,CD3 ζ 信号传导结构域由SEQ ID NO:15中所示的序列组成或基本上由其组成。

[0528] 在特定实施方案中,转基因是或含有编码小于全长的CD3 ζ 链(例如,小于整个的CD3 ζ 信号传导结构域)的核苷酸序列。在某些实施方案中,转基因含有编码CD3 ζ 链的一部分的核苷酸序列,所述部分是或包括CD247开放阅读框的小于4个外显子、3个完整外显子、小于3个外显子、2个完整外显子、小于2个外显子、1个外显子或小于一个外显子。在特定实施方案中,转基因含有编码CD3 ζ 链的一部分的核苷酸序列,所述部分的长度为或小于100、90、80、70、60、50、40、30、20、10、9、8、7、6、5、4、3或2个核苷酸。在一些实施方案中,转基因含有编码CD3 ζ 链的一部分的核苷酸序列,所述部分具有与SEQ ID NO:74或76中所示核酸序列的全部或部分具有为或至少70%、75%、80%、85%、90%、95%、97%、98%、99%、99.5%或99.9%序列同一性的序列的为、约或小于100、90、80、70、60、50、40、30、20、10、9、8、7、6、5、4、3或2个连续核苷酸。

[0529] 在一些方面,由转基因编码的CD3 ζ 链的部分包括由选自外显子1、2、3、4或5的一个或多个外显子编码的部分,或者由连续序列中包含外显子1-5或2-5的部分序列(例如,没有内含子序列)编码的部分。在一些方面,由转基因编码的CD3 ζ 链的部分包括由选自外显子1、2或3的一个或多个外显子编码的部分,或者由连续序列中包含外显子1-3或2-3的部分序列(例如,没有内含子序列)编码的部分。在一些方面,由转基因编码的CD3 ζ 链的部分包括由外显子2或其部分序列编码的部分。在一些方面,转基因包括内源CD247基因座的开放阅读框序列的外显子2的最后3、6、9、12、15或18个核苷酸。在一些实施方案中,转基因包括编码由内源CD247基因座的开放阅读框的外显子2编码的最后1、2、3、4、5或6个氨基酸残基的序列。在一些方面,转基因包括外显子2的最后9个核苷酸和/或编码由内源CD247基因座的开放阅读框序列的外显子2编码的最后3个氨基酸残基的序列。在一些方面,转基因包括外显子3的

前3、6、9、12、15或18个核苷酸和/或编码由内源CD247基因座的开放阅读框的外显子3编码的前1、2、3、4、5或6个氨基酸残基的序列。

[0530] 在特定实施方案中,转基因不含任何内含子。在特定实施方案中,转基因中编码CD3 ζ 链的部分的核苷酸序列不含任何内含子或其部分。

[0531] 在一些方面,转基因不包含编码CD3 ζ 链的任何部分的核酸序列。在一些方面,编码CD3 ζ 链的部分的核苷酸序列(如果存在于转基因内)通常与同源臂序列之一(例如,3'同源臂序列)连接。在一些方面,如果存在于转基因中,是转基因中的最3'区域,其随后与同源臂序列之一(例如,3'同源臂序列)连接。

(v) 另外的结构域,例如,多聚化结构域

[0532] 在一些实施方案中,转基因还包括编码一个或多个多聚化结构域(例如,二聚化结构域)的核苷酸序列。在一些方面,所编码的多聚化结构域可以是细胞外的或细胞内的。在一些实施方案中,所编码的多聚化结构域是细胞外的。在一些实施方案中,所编码的多聚化结构域是细胞内的。在一些实施方案中,由转基因序列编码的细胞内区域的部分包含多聚化结构域,任选地二聚化结构域。在一些实施方案中,转基因包含编码细胞外区域的核苷酸序列。在一些实施方案中,细胞外区域包含多聚化结构域,任选地二聚化结构域。在一些实施方案中,多聚化结构域能够在与诱导物结合后发生二聚化。

[0533] 在一些方面,嵌合受体是多链嵌合受体,如多链CAR。在一些实施方案中,多链嵌合受体或其部分的一条或多条链是由转基因序列编码。在一些实施方案中,借助嵌合受体的每条链中包括的多聚化结构域的多聚化,多链嵌合受体的一条或多条链可以一起形成功能性或活性嵌合受体。

[0534] 在一些方面,编码多聚化结构域的核苷酸序列位于其他结构域的5'或3'。例如,在一些实施方案中,所编码的多聚化结构域是细胞外的,并且编码多聚化结构域的序列位于编码间隔子的序列的5'。在一些实施方案中,所编码的多聚化结构域是细胞内的,并且编码多聚化结构域的序列位于编码CD3 ζ 链或其片段的序列的5'。在一些实施方案中,多聚化结构域是细胞内的,并且编码多聚化结构域的序列位于编码一个或多个共刺激信号传导结构域的序列的5'或3'。在一些实施方案中,所编码的多聚化结构域可以在结合诱导物后发生多聚化(例如,二聚化)。示例性编码的多聚化结构域包括本文例如在本文的章节III.B中所述的任何多聚化结构域。

(vi) 另外的分子,例如,标记

[0535] 在一些实施方案中,转基因还包括核苷酸序列编码一种或多种另外的分子,如抗体、抗原、另外的嵌合或多链嵌合受体(例如,多链CAR、嵌合共刺激受体、抑制性受体、可调节嵌合抗原受体或者本文例如在章节III.B.2中所述的多链嵌合受体系统的其他组分;或者重组T细胞受体(TCR))的另外的多肽链、转导标记或替代标记(例如,截短的细胞表面标记)、酶、因子、转录因子、抑制性肽、生长因子、核受体、激素、淋巴因子、细胞因子、趋化因子、可溶受体、可溶细胞因子受体、可溶趋化因子受体、受体、前述任一种的功能片段或功能变体以及前述的组合。在一些方面,编码一种或多种另外的分子的这种核苷酸序列可以位于编码嵌合受体的区域或结构域的核苷酸序列的5'。在一些方面,编码一种或多种其他分子的序列和编码嵌合受体的区域或结构域的核苷酸序列是由调节序列(如2A核糖体跳跃元件和/或启动子序列)隔开。

[0536] 在一些实施方案中,转基因还包括编码一种或多种另外的分子的核苷酸序列。在一些方面,一种或多种另外的分子包括一种或多种标记。在一些实施方案中,所述一种或多种标记包括转导标记、替代标记和/或选择标记。在一些实施方案中,转基因还包括可以例如通过促进转移细胞的活力和/或功能来改进疗法功效的核酸序列;提供用于选择和/或评价细胞(如用于在体内评估存活或定位)的遗传标记的核酸序列;例如通过使细胞对体内阴性选择敏感来提高安全性的核酸序列,如以下文献中所述:Lupton S.D.等人,Mol. and Cell Biol.,11:6(1991);和Riddell等人,Human Gene Therapy 3:319-338(1992);还参见WO 1992008796和WO 1994028143(其描述使用源自融合显性阳性选择标记与阴性选择标记的双功能可选融合基因)以及美国专利号6,040,177。在一些方面,标记包括本文例如在此章节或章节II或III.B中所述的任何标记,或者本文例如在章节III.B.2中所述的任何另外的分子和/或受体多肽。在一些实施方案中,另外的分子是替代标记,任选地截短的受体,任选地其中所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。

[0537] 在一些实施方案中,标记是转导标记或替代标记。转导标记或替代标记可以用于检测已经引入多核苷酸(例如,编码嵌合受体或其部分的多核苷酸)的细胞。在一些实施方案中,转导标记可以指示或确认对细胞的修饰。在一些实施方案中,替代标记是制备为在细胞表面上与嵌合受体或其部分(例如CAR)共表达的蛋白质。在特定实施方案中,这种替代标记是已经被修饰的以具有极小或无活性的表面蛋白。在某些实施方案中,替代标记是在编码嵌合受体或其部分的相同多核苷酸上编码。在一些实施方案中,编码嵌合受体或其部分的核酸序列与编码标记的核酸序列可操作地连接,任选地由内部核糖体进入位点(IRES)或编码自切割肽或引起核糖体跳跃的肽(如2A序列,如T2A、P2A、E2A或F2A)的核酸隔开。在一些情形中,外在标记基因可以与工程化细胞结合用于允许检测或选择细胞,并且在一些情形中,还可以用于促进细胞消除和/或细胞自杀。

[0538] 示例性替代标记可以包括细胞表面多肽的截短的形式,如以下截短的形式,其是非功能性的,并且不转导或不能转导信号或通常由细胞表面多肽的全长形式转导的信号,和/或不内化或不能内化。示例性截短的细胞表面多肽包括截短形式的生长因子或其他受体,如截短的人表皮生长因子受体2(tHER2)、截短的表皮生长因子受体(tEGFR,SEQ ID NO:7或16所示的示例性tEGFR序列)或前列腺特异性膜抗原(PSMA)或其修饰形式。tEGFR可以含有抗体西妥昔单抗(Erbitux®)或其他治疗性抗EGFR抗体或结合分子识别的表位,其可以用于鉴定或选择已用tEGFR构建体和编码的外源蛋白质工程化的细胞,和/或用于消除或分离表达所编码的外源蛋白质的细胞。参见美国专利号8,802,374和Liu等人,Nature Biotech.2016年4月;34(4):430-434。在一些方面,标记(例如替代标记)包括全部或部分(例如截短形式的)CD34、NGFR、CD19或截短的CD19(例如截短的非人CD19)或表皮生长因子受体(例如tEGFR)。

[0539] 在一些实施方案中,标记是或包含可检测蛋白,如荧光蛋白,如绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(EGFP)(如超折叠GFP(sfGFP))、红色荧光蛋白(RFP)(如tdTomato、mCherry、mStrawberry、AsRed2、DsRed或DsRed2)、青色荧光蛋白(CFP)、蓝绿色荧光蛋白(BFP)、增强型蓝色荧光蛋白(EBFP)和黄色荧光蛋白(YFP)及其变体,包括荧光蛋白的物种变体、单体变体、密码子优化的、稳定的和/或增强的变体。在一些实施方案中,标记

是或包含酶(如萤光素酶)、来自大肠杆菌(E.coli)的lacZ基因、碱性磷酸酶、分泌的胚胎碱性磷酸酶(SEAP)、氯霉素乙酰转移酶(CAT)。示例性发光报告基因包括萤光素酶(luc)、 β -半乳糖苷酶、氯霉素乙酰转移酶(CAT)、 β -葡糖醛酸糖苷酶(GUS)或其变体。在一些方面,所述酶的表达可以通过添加底物来检测,所述底物可以根据所述酶的表达和功能活性来检测。

[0540] 在一些实施方案中,标记是选择标记。在一些实施方案中,选择标记是或包含赋予针对外源药剂或药物的抗性的多肽。在一些实施方案中,选择标记是抗生素抗性基因。在一些实施方案中,选择标记是向哺乳动物细胞赋予抗生素抗性的抗生素抗性基因。在一些实施方案中,选择标记是或包含嘌呤霉素抗性基因、潮霉素抗性基因、杀稻瘟菌素抗性基因、新霉素抗性基因、遗传霉素抗性基因或博莱霉素抗性基因或其变体。

[0541] 在一些实施方案中,所述分子是非自身分子,例如非自身蛋白,即不被将被过继转移细胞的宿主免疫系统识别为“自身”的分子。

[0542] 在一些实施方案中,所述标记不提供任何治疗功能和/或不产生除了用作基因工程化标记(例如,用于选择成功工程化的细胞)以外的作用。在其他实施方案中,所述标记可以是治疗性分子或以其他方式发挥一些所需作用的分子,如在体内会遇到的细胞的配体,如用于在过继转移和遇到配体时增强和/或减弱细胞应答的共刺激或免疫检查点分子。

[0543] 在一些实施方案中,转基因包括编码作为免疫调节剂的一种或多种另外的分子的序列。在一些实施方案中,免疫调节分子选自免疫检查点调节剂、免疫检查点抑制剂、细胞因子或趋化因子。在一些实施方案中,免疫调节剂是能够抑制或阻断免疫检查点分子的功能或涉及免疫检查点分子的信号传导途径的免疫检查点抑制剂。在一些实施方案中,免疫检查点分子选自PD-1、PD-L1、PD-L2、CTLA-4、LAG-3、TIM3、VISTA、腺苷受体或细胞外腺苷,任选地腺苷2A受体(A2AR)或腺苷2B受体(A2BR)、或涉及前述任一种的腺苷或途径。其他示例性的另外的分子包括表位标签、可检测分子(如荧光或发光蛋白)或者介导增强的细胞生长和/或基因扩增的分子(例如,二氢叶酸还原酶)。表位标签包括例如一个或多个拷贝的FLAG、His、myc、Tap、HA或任何可检测氨基酸序列。在一些实施方案中,另外的分子可以包括非编码序列、抑制性核酸序列(如反义RNA、RNAi、shRNA和微小RNA(miRNA))或核酸酶识别序列。

[0544] 在一些方面,另外的分子可以包括本文所述的任何另外的受体多肽,如多链嵌合受体的任何另外的受体多肽链,如在章节III.B.2中所述。

(vii) 多顺反子元件和调节或控制元件

[0545] 在一些实施方案中,转基因(例如,外源核酸序列)还含有并非或不同于内源CD247基因座的调节或控制元件的一个或多个异源或外源调节或控制元件,例如,顺式调节元件。在一些方面,异源调节或控制元件包括例如启动子、增强子、内含子、隔离子、多腺苷酸化信号、转录终止序列、Kozak共有序列、多顺反子元件(例如,内部核糖体进入位点(IRES)、2A序列)、对应于信使RNA(mRNA)的非翻译区(UTR)的序列,以及剪接接纳体或供体序列,如并非或不同于CD247基因座处的调节或控制元件的那些。在一些实施方案中,异源调节或控制元件包括启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、剪接接纳体序列和/或剪接供体序列。在一些实施方案中,转基因包含异源启动子和/或通常在靶位点处或附近不存在的启动子。在一些方面,调节或控制元件包括在整合于CD247基因座处时,调节或控制嵌合受体的表达所需的元件。在一些实施方案中,转基因序列包括对应于异源基因或基因座

的5'和/或3'非翻译区(UTR)的序列。在一些方面,转基因序列可以包括本文所述的任何调节或控制元件,包括在此章节和章节II中所述的那些。

[0546] 可以插入转基因(包括编码嵌合受体或其部分的转基因),使得其表达由整合位点处的内源启动子(即驱动内源CD247基因表达的启动子)驱动。在多肽编码序列无启动子的一些实施方案中,则通过内源启动子或目的区域中的其他控制元件驱动的转录来确保整合的转基因的表达。例如,编码嵌合受体的一部分的转基因可以在没有启动子的情况下插入,但是与内源CD247基因座的编码序列同框,使得整合的转基因的表达受整合位点处的内源启动子和/或其他调节元件的转录的控制。在一些实施方案中,多顺反子元件,如核糖体跳跃元件/自切割元件(例如,2A元件或内部核糖体进入位点(IRES)),位于编码嵌合受体的一部分的转基因的上游,使得所述多顺反子元件与CD247基因座的内源开放阅读框的一个或多个外显子同框定位,使得编码嵌合受体的转基因的表达与内源CD247启动子可操作地连接。在一些实施方案中,转基因序列不包含编码3'UTR的序列。在一些实施方案中,在将转基因整合至内源CD247基因座中之后,所述转基因整合于内源CD247基因座的3'UTR的上游,使得编码嵌合受体的信息含有内源CD247基因座的3'UTR,例如,来自内源CD247基因座的开放阅读框或其部分序列的3'UTR。在一些实施方案中,编码嵌合受体的其余部分的开放阅读框或其部分序列包含内源CD247基因座的3'UTR。

[0547] 在一些实施方案中,将“串联”盒整合至所选位点中。在一些实施方案中,一个或多个“串联”盒编码一种或多种多肽或因子,它们各自独立地受调节元件控制或全部作为多顺反子表达系统而受控。在一些实施方案,如多核苷酸含有第一和第二核酸序列的那些实施方案中,编码不同多肽链中的每一种的编码序列可以与相同或不同的启动子操作性地连接。在一些实施方案中,核酸分子可以含有驱动两条或更多条不同多肽链表达的启动子。在一些实施方案中,此类核酸分子可以是多顺反子的(双顺反子的或三顺反子的,参见例如,美国专利号6,060,273)。在一些实施方案中,转录单元可以被工程化为含有IRES(内部核糖体进入位点)的双顺反子单元,其允许通过来自单一启动子的信息共表达基因产物。可替代地,在一些情况下,单一启动子可以引导在单一开放阅读框(ORF)中含有两个或三个多肽的RNA的表达,所述多肽通过编码自我切割肽(例如,2A序列)或蛋白酶识别位点(例如,弗林蛋白酶)的序列彼此隔开,如本文所述。ORF由此编码单一多肽,所述多肽在翻译期间(在2A的情况下)或在翻译后被加工为单独蛋白质。在一些实施方案中,“串联盒”包括所述盒的包含无启动子序列的第一组分、之后的转录终止序列、以及编码自主表达盒或多顺反子表达序列的第二序列。在一些实施方案中,串联盒编码两种或更多种不同的多肽或因子,例如嵌合受体的两个或更多个链或结构域。在一些实施方案中,将编码嵌合受体的两个或更多个链或结构域的核酸序列作为串联表达盒或者双顺反子或多顺反子盒引入一个靶DNA整合位点中。

[0548] 在一些情形中,多顺反子元件(如T2A)可以引起核糖体跳过(核糖体跳跃)2A元件C末端肽键的合成,从而导致2A序列的末端与下游的下一个肽之间的分离(参见例如,de Felipe, *Genetic Vaccines and Ther.* 2:13 (2004) 以及 de Felipe等人 *Traffic* 5:616-626 (2004);也称为自切割元件)。这允许插入的转基因受到整合位点处的内源启动子(如CD247启动子)的转录的控制。示例性多顺反子元件包括来自以下的2A序列:口蹄疫病毒(F2A,例如,SEQ ID NO:21)、马鼻炎A病毒(E2A,例如,SEQ ID NO:20)、明脉扁刺蛾β四体病毒(T2A,

例如,SEQ ID NO:6或17)和猪捷申病毒-1(P2A,例如,SEQ ID NO:18或19),如美国专利公开号20070116690中所述。在一些实施方案中,模板多核苷酸包括转基因(例如,编码嵌合受体或其部分的核酸)上游的P2A核糖体跳跃元件(SEQ ID NO:18或19中所示的序列)。

[0549] 在一些实施方案中,编码嵌合受体或其部分的一条或多条链的转基因和/或编码另外的分子的序列独立地包含一个或多个多顺反子元件。在一些实施方案中,所述一个或多个多顺反子元件位于编码其嵌合受体部分的核酸序列和/或编码另外的分子的序列的上游。在一些实施方案中,所述一个或多个多顺反子元件位于编码其嵌合受体部分的核酸序列和/或编码另外的分子的序列之间。在一些实施方案中,所述一个或多个多顺反子元件位于编码嵌合受体的部分或链的核酸序列之间。

[0550] 在一些实施方案中,异源调节或控制元件包括异源启动子。在一些实施方案中,异源启动子选自组成型启动子、诱导型启动子、阻抑型启动子和/或组织特异性启动子。在一些实施方案中,调节或控制元件是启动子和/或增强子,例如组成型启动子或诱导型或组织特异性启动子。在一些实施方案中,启动子选自RNA pol I、pol II或pol III启动子。

在一些实施方案中,启动子由RNA聚合酶II识别(例如,CMV、SV40早期区域或腺病毒主要晚期启动子)。在一些实施方案中,启动子由RNA聚合酶III识别(例如,U6或H1启动子)。在一些实施方案中,启动子是或包含组成型启动子。示例性组成型启动子包括例如猿病毒40早期启动子(SV40)、巨细胞病毒立即早期启动子(CMV)、人泛素C启动子(UBC)、人延伸因子1 α 启动子(EF1 α)、小鼠磷酸甘油酸激酶1启动子(PGK)、和与CMV早期增强子偶联的鸡 β -肌动蛋白启动子(CAGG)。在一些实施方案中,异源启动子是或包含人延伸因子1 α (EF1 α)启动子或MND启动子或其变体。

[0551] 在一些实施方案中,启动子是受调节的启动子(例如,诱导型启动子)。在一些实施方案中,启动子是诱导型启动子或阻抑型启动子。在一些实施方案中,启动子包含Lac操纵基因序列、四环素操纵基因序列、半乳糖操纵基因序列或多西环素操纵基因序列,或者是其类似物,或者能够被Lac阻遏物或四环素阻遏物或其类似物结合或识别。在一些实施方案中,启动子是组织特异性启动子。在一些情况下,启动子仅在特定细胞类型中表达(例如,T细胞或B细胞或NK细胞特异性启动子)。

[0552] 在一些实施方案中,启动子是或包含组成型启动子。示例性组成型启动子包括例如猿病毒40早期启动子(SV40)、巨细胞病毒立即早期启动子(CMV)、人泛素C启动子(UBC)、人延伸因子1 α 启动子(EF1 α)、小鼠磷酸甘油酸激酶1启动子(PGK)、和与CMV早期增强子偶联的鸡 β -肌动蛋白启动子(CAGG)。在一些实施方案中,组成型启动子是合成或修饰的启动子。在一些实施方案中,启动子是或包含MND启动子,它是一种合成启动子,含有修饰的MoMuLV LTR的U3区和骨髓增生性肉瘤病毒增强子(参见Challita等人(1995)J.Virol.69(2):748-755)。在一些实施方案中,启动子是组织特异性启动子。在一些情况下,启动子仅在特定细胞类型中驱动表达(例如,T细胞或B细胞或NK细胞特异性启动子)。

[0553] 在一些实施方案中,启动子是病毒启动子。在一些实施方案中,启动子是非病毒启动子。在一些情形中,启动子选自人延伸因子1 α (EF1 α)启动子(如SEQ ID NO:77或118中所示)或其修饰的形式(具有HTLV1增强子的EF1 α 启动子;如SEQ ID NO:119中所示)或MND启动子(如SEQ ID NO:131中所示)。在一些实施方案中,多核苷酸不包括异源或外源调节元件,例如,启动子。在一些实施方案中,启动子是双向启动子(参见例如,W02016/022994)。

[0554] 在一些实施方案中,转基因序列还可以包括剪接接纳体序列。示例性已知剪接接纳体位点序列包括例如CTGACCTCTTCTCTTCCTCCCACAG (SEQ ID NO:78) (来自人HBB基因) 和TTTCTCTCCACAG (SEQ ID NO:79) (来自人IgG基因)。

[0555] 在一些实施方案中,转基因序列还可以包括转录终止和/或多腺苷酸化信号所需的序列。在一些方面,示例性多腺苷酸化信号选自SV40、hGH、BGH和rbGlob转录终止序列和/或多腺苷酸化信号。在一些实施方案中,转基因包括SV40多腺苷酸化信号。在一些实施方案中,如果在转基因内存在,转录终止序列和/或多腺苷酸化信号通常是所述转基因内的最3'序列,并且与同源臂之一连接。在一些方面,转基因序列不包含编码3'UTR或转录终止子的序列。在一些实施方案中,在将转基因整合至内源CD247基因座中之后,所述转基因整合于内源CD247基因座的3'UTR和/或转录终止子的上游,使得编码嵌合受体的信息含有内源CD247基因座的3'UTR,例如,来自内源CD247基因座的开放阅读框或其部分序列的3'UTR。因此,在一些实施方案中,在整合编码嵌合受体的一部分的转基因序列之后,编码嵌合受体的核酸序列被可操作地连接以处于内源CD247基因座的3'UTR、转录终止子和/或其他调节元件的控制下。

(viii) 示例性转基因序列

[0556] 在一些实施方案中,示例性转基因以5'至3'顺序包括各自编码以下的核苷酸序列:跨膜结构域(或膜结合结构域)和细胞内区域。在一些实施方案中,示例性转基因以5'至3'顺序包括各自编码以下的核苷酸序列:细胞外区域、跨膜结构域和细胞内区域。

[0557] 在一些实施方案中,编码细胞外区域的示例性转基因以5'至3'顺序包括编码细胞外结合结构域的核苷酸序列和编码间隔子的核苷酸序列。在一些实施方案中,示例性转基因还包括编码一个或多个细胞外多聚化结构域的核苷酸序列,所述核苷酸序列可以位于编码结合结构域和/或间隔子的任何核苷酸序列的5'或3'和/或编码跨膜结构域的核苷酸序列的5'。在一些方面,示例性转基因序列还包括通常位于编码细胞外区域的核苷酸序列的5'的信号序列。

[0558] 在一些方面,在示例性转基因中,编码结合结构域的核苷酸序列位于信号序列与编码间隔子的核苷酸之间。在一些方面,在示例性转基因中,编码细胞外多聚化结构域的核苷酸序列位于编码结合结构域的核苷酸序列与编码间隔子的核苷酸序列之间。在一些方面,编码间隔子的核苷酸序列位于编码结合结构域的核苷酸序列与编码跨膜结构域的核苷酸序列之间。

[0559] 在一些实施方案中,示例性转基因含有编码细胞内区域的核苷酸序列,所述核苷酸序列可以按5'至3'顺序包括编码一个或多个共刺激信号传导结构域的核苷酸序列和任选地编码CD3 ζ 链的一部分的核苷酸序列。在一些实施方案中,示例性转基因还包括编码一个或多个细胞内多聚化结构域的核苷酸序列,所述核苷酸序列可以位于一个或多个共刺激结构域中任一个的5'或3';和/或编码CD3 ζ 链的一部分的核苷酸序列(如果存在)或与多核苷酸中的示例性转基因相邻的3'同源臂序列的5'。在一些实施方案中,示例性转基因含有编码细胞内区域的核苷酸序列,所述核苷酸序列可以按5'至3'顺序包括编码细胞内多聚化结构域的核苷酸序列、编码一个或多个共刺激信号传导结构域的核苷酸序列以及任选地编码CD3 ζ 链的一部分的核苷酸序列。在一些方面,在示例性转基因中,编码一个或多个共刺激信号传导结构域的核苷酸序列位于编码跨膜结构域的核苷酸序列与编码CD3 ζ 链的一部分

的核苷酸序列(如果存在)或与多核苷酸中的示例性转基因相邻的3'同源臂序列之间。在一些方面,在示例性转基因中,编码细胞内多聚化结构域的核苷酸序列位于编码跨膜结构域的核苷酸序列与编码一个或多个共刺激信号传导结构域的核苷酸序列之间;或者位于编码一个或多个共刺激信号传导结构域的核苷酸序列与编码CD3 ζ 链的一部分的核苷酸序列(如果存在)或与多核苷酸中的示例性转基因相邻的3'同源臂序列之间。

[0560] 在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和包含CD3 ζ 链的一部分的细胞内区域。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和共刺激信号传导结构域。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和两个共刺激信号传导结构域。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和三个共刺激信号传导结构域。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和共刺激信号传导结构域和CD3 ζ 链的一部分。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和两个共刺激信号传导结构域和CD3 ζ 链的一部分。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:信号肽、细胞外结合结构域、间隔子、跨膜结构域和三个共刺激信号传导结构域和CD3 ζ 链的一部分。

[0561] 在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:跨膜结构域(或膜结合结构域)、细胞内多聚化结构域、任选地一个或多个共刺激信号传导结构域以及任选地CD3 ζ 链的一部分。在一些实施方案中,示例性转基因序列在5'至3'方向上包含各自编码以下的核苷酸序列:细胞外多聚化结构域、跨膜结构域、任选地一个或多个共刺激信号传导结构域以及任选地CD3 ζ 链的一部分。

[0562] 在一些实施方案中,示例性转基因序列还可以包含位于编码信号肽和/或细胞外区域的序列的5'的多顺反子元件(例如,2A元件或内部核糖体进入位点(IRES))和/或调节或控制元件(例如,启动子)。在一些实施方案中,示例性转基因序列还可以包含另外的序列,例如,编码一种或多种另外的分子的核苷酸序列,所述另外的分子如标记、另外的嵌合受体、抗体或其抗原结合片段、免疫调节分子、配体、细胞因子或趋化因子。在一些方面,编码一种或多种其他分子的序列和编码嵌合受体的区域或结构域的核苷酸序列是由调节序列(如2A核糖体跳跃元件和/或启动子序列)隔开。在一些方面,在示例性转基因中,编码一种或多种另外的分子的核苷酸序列位于编码信号肽和/或细胞外区域的序列的5'。在一些实施方案中,编码一种或多种另外的分子的核苷酸序列位于多顺反子元件和/或调节或控制元件与编码嵌合受体的区域或结构域的核苷酸序列之间。在一些实施方案中,编码一种或多种另外的分子的核苷酸序列位于两个援建和/或调节或控制元件之间。在一些实施方案中,示例性转基因序列按5'至3'方向上包含:多顺反子元件和/或调节元件、编码另外的分子的核苷酸序列、多顺反子元件和/或调节元件、信号肽、编码嵌合受体的区域或结构域(例如,细胞外区域、跨膜结构域、细胞内区域)的核酸序列。

[0563] 在一些实施方案中,转基因序列按顺序包含:编码细胞外结合结构域、任选地scFv

的核苷酸序列；间隔子，所述间隔子任选地包含来自人免疫球蛋白铰链的序列，任选地来自IgG1、IgG2或IgG4或其修饰形式的序列，任选地还包含C_H2区域和/或C_H3区域；以及任选地来自人CD28的跨膜结构域；细胞内区域，所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域；以及任选地CD3 ζ 信号传导结构域的一部分。在一些实施方案中，嵌合受体的所编码的细胞内区域从其N末端至C末端按顺序包含：一个或多个共刺激信号传导结构域和CD3 ζ 链或其片段。

b. 同源臂

[0564] 在一些实施方案中，模板多核苷酸在5'和/或3'端上含有一个或多个同源性序列（也称为“同源臂”），所述一个或多个同源性序列与编码嵌合受体或其部分的一条或多条链的转基因序列连接，侧接所述转基因序列或在所述转基因序列周围。在一些实施方案中，一个或多个同源臂包括5'和/或3'同源臂。同源臂允许DNA修复机制（例如，同源重组机构）识别同源性以及使用模板多核苷酸作为修复模板，并且将同源臂之间的核酸序列拷贝至所修复的DNA中，从而将转基因序列有效插入或整合至基因组中同源性位置之间的靶整合位点中。

[0565] 在一些方面，在整合转基因序列后，转基因序列包含与一个或多个同源臂中包含的CD247基因座的开放阅读框的一个或多个外显子同框的核苷酸序列。在一些方面，嵌合受体的一部分是由转基因序列编码，并且所述嵌合受体的其余部分（例如，CD3 ζ 信号传导结构域的一部分或整个CD3 ζ 信号传导结构域）是由内源CD247基因座的一个或多个外显子编码。

[0566] 在一些实施方案中，同源臂序列包括与遗传破坏（例如，CD247基因座内的靶位点）周围的基因组序列同源的序列。在一些实施方案中，模板多核苷酸包含以下组分：[5'同源臂]-[转基因序列（例如，编码嵌合受体的一条或多条链的外源或异源核酸序列）]-[3'同源臂]。在一些实施方案中，5'同源臂序列包括与位于5'侧上的遗传破坏附近的序列同源的连续序列。在一些实施方案中，3'同源臂序列包括与位于3'侧上的遗传破坏附近的序列同源的连续序列。在一些方面，靶位点是通过能够引入遗传破坏的一种或多种药剂（例如，靶向CD247基因座内的特定定位点的Cas9和gRNA）的靶向来确定。

[0567] 在一些方面，模板多核苷酸内的转基因序列可以用于指导靶位点和/或同源臂的定位。在一些方面，遗传破坏的靶位点可以用作设计用于HDR的模板多核苷酸和/或同源臂的指导。在一些实施方案中，遗传破坏可以靶向于转基因序列的所需靶向整合位点附近。在一些方面，同源臂设计为靶向整合于内源CD247基因座的开放阅读框的外显子内，并且同源臂序列是基于遗传破坏周围的所需整合位置（包括遗传破坏周围的外显子和内含子序列）来确定。在一些实施方案中，靶位点的位置、一个或多个同源臂的相对位置和用于插入的转基因（外源核酸序列）可以根据有效靶向的需要以及可以使用的模板多核苷酸或载体的长度来设计。在一些方面，同源臂设计为靶向整合于CD247基因座的开放阅读框的内含子内。在一些方面，同源臂设计为靶向整合于CD247基因座的开放阅读框的外显子内。

[0568] 在一些方面，CD247基因座内的靶整合位点（靶向整合的位点）位于编码CD3 ζ 链的内源CD247基因座的开放阅读框内。在一些实施方案中，靶整合位点位于本文例如在章节I.A中所述的任何靶位点处或附近。在一些方面，靶整合位置位于遗传破坏的靶位点处或附近，例如，位于遗传破坏的靶位点的小于500、450、400、350、300、250、200、150、100或50bp内。

[0569] 在一些方面,靶整合位点位于内源CD247基因座的开放阅读框的外显子内。在一些方面,靶整合位点位于CD247基因座的开放阅读框的内含子内。在一些方面,靶整合位点位于CD247基因座的调节或控制元件(例如,启动子)内。在一些实施方案中,靶整合位点位于对应于早期编码区的外显子(例如,内源CD247基因座的开放阅读框的外显子1、2或3)内或非常靠近所述外显子,或者包括紧接在转录起始位点之后,在外显子1、2或3(如本文表1中所述)内,或者在外显子1、2或3的小于500、450、400、350、300、250、200、150、100或50bp内的序列。在一些实施方案中,所述整合靶向于内源CD247基因座的外显子2处或附近,或者位于外显子2的小于500、450、400、350、300、250、200、150、100或50bp内。在一些方面,靶整合位点位于内源CD247基因座的外显子1处或附近,例如,位于外显子1的小于500、450、400、350、300、250、200、150、100或50bp内。在一些实施方案中,靶整合位点位于内源CD247基因座的外显子2处或附近,或者位于外显子2的小于500、450、400、350、300、250、200、150、100或50bp内。在一些方面,靶整合位点位于内源CD247基因座的外显子3处或附近,例如,位于外显子3的小于500、450、400、350、300、250、200、150、100或50bp内。

[0570] 在一些实施方案中,5'同源臂序列包括位于遗传破坏的靶位点的5'的大约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个碱基对的连续序列,其始于内源CD247基因座的靶位点附近。在一些实施方案中,3'同源臂序列包括位于遗传破坏的靶位点的3'的大约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个碱基对的连续序列,其始于内源CD247基因座的靶位点附近。因此,在经由HDR整合后,转基因序列被靶向以供整合于遗传破坏的靶位点(例如,内源CD247基因座的外显子或内含子内的靶位点)处或附近。

[0571] 在一些方面,同源臂含有与内源CD247基因座的开放阅读框序列的一部分同源的序列。在一些方面,同源臂序列含有与内源CD247基因座的开放阅读框序列的连续部分(包括外显子和内含子)同源的序列。在一些方面,同源臂含有与内源CD247基因座的开放阅读框序列的连续部分(包括外显子和内含子)相同的序列。

[0572] 在一些实施方案中,模板多核苷酸含有用于将转基因序列靶向整合于内源CD247基因座处的同源臂(描述于本文表1中的示例性基因组基因座序列;示于SEQ ID NO:74 (NCBI参考序列:NM_198053.2)和SEQ ID NO:76 (NCBI参考序列:NM_000734.3)中的示例性mRNA序列)。在一些实施方案中,使用用于遗传破坏的任何药剂(例如,本文所述的靶向核酸酶和/或gRNA)引入遗传破坏。在一些实施方案中,模板多核苷酸在通过靶向核酸酶和/或gRNA引入的遗传破坏的任一侧上包含约500至1000(例如,500至900或600至700)个同源性碱基对。在一些实施方案中,模板多核苷酸包含5'同源臂序列的约500、600、700、800、900或1000个碱基对,其与CD247基因座处的遗传破坏的5'序列的500、600、700、800、900或1000个碱基对同源;转基因;以及3'同源臂序列的约500、600、700、800、900或1000个碱基对,其与CD247基因座处的遗传破坏的3'序列的500、600、700、800、900或1000个碱基对同源。

[0573] 在一些方面,转基因与一个或多个同源臂序列之间的边界设计为使得在HDR和转基因序列的靶向整合之后,转基因内编码一个或多个多肽(例如,嵌合受体的一条或多条链、一个或多个结构域或者一个或多个区域)的序列与内源CD247基因座处的开放阅读框序列的一个或多个外显子同框整合,和/或生成编码多肽的转基因与内源CD247基因座处的开放阅读框序列的一个或多个外显子的框内融合物。

[0574] 在一些实施方案中,一个或多个同源臂序列包括如下序列,所述序列与在位于内源CD247基因座处的开放阅读框序列内的靶位点周围或侧接所述靶位点的序列同源、基本上相同或相同。在一些实施方案中,一个或多个同源臂包含CD247基因座的开放阅读框的至少一个内含子和至少一个外显子。在一些方面,一个或多个同源臂序列含有内源CD247基因座处的开放阅读框的部分序列的内含子和外显子。在一些方面,5'同源臂序列与转基因的边界使得在转基因不含异源启动子的情形中,转基因序列的编码部分与内源CD247基因座的开放阅读框的上游外显子或其部分(例如,外显子1、2或3,根据靶向整合的位置而定)框内融合。在一些方面,3'同源臂序列与转基因的边界使得,内源CD247基因座的开放阅读框的下游外显子或其部分(例如,外显子2、3、4、5、6、7或8)与转基因序列的编码部分框内融合。因此,在靶向整合、转录和翻译后,从转基因与内源CD247基因座的开放阅读框序列的融合DNA序列产生作为连续多肽的所编码的嵌合受体。在一些方面,所编码的嵌合受体中由融合DNA序列产生的部分是CD3 ζ 链或其片段。在一些方面,所编码的嵌合受体能够经由CD3 ζ 链或其部分进行信号传导。在一些实施方案中,一个或多个同源臂不包含CD247基因座的开放阅读框的外显子1的全长。在一些实施方案中,一个或多个同源臂不包含CD247基因座的开放阅读框的外显子1和/或不包含外显子2的全长。

[0575] 在一些实施方案中,用于靶向整合于内源CD247基因座处的示例性5'同源臂包含SEQ ID NO:80中所示的序列,或者展现与SEQ ID NO:80至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。在一些实施方案中,所述5'同源臂包含SEQ ID NO:80中所示的序列。在一些实施方案中,5'同源臂由SEQ ID NO:80中所示的序列组成或基本上由其组成。

[0576] 在一些实施方案中,用于靶向整合于内源CD247基因座处的示例性3'同源臂包含SEQ ID NO:81中所示的序列,或者展现与SEQ ID NO:81至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。在一些实施方案中,所述3'同源臂包含SEQ ID NO:81中所示的序列。在一些实施方案中,3'同源臂由SEQ ID NO:81中所示的序列组成或基本上由其组成。

[0577] 在一些方面,靶位点可以决定同源臂的相对位置和序列。同源臂通常可以延伸至至少远至其中可以在引入遗传破坏(例如,DSB)后通过DNA修复机制进行末端切除的区域,例如,以允许切除的单链突出端找到模板多核苷酸内的互补区。总体长度可能受限于诸如以下的参数:质粒大小、病毒包装极限或构建体大小极限。

[0578] 在一些实施方案中,同源臂在内源基因的靶位点的任一侧上包含为或约500至1000(例如,600至900或700至800)个同源性碱基对。在一些实施方案中,同源臂在CD247基因座的靶位点的5'、靶位点的3'或靶位点的5'和3'两侧包含为或约至少或至少约或者少于或约200、300、400、500、600、700、800、900或1000个同源性碱基对。

[0579] 在一些实施方案中,同源臂在CD247基因座的靶位点的3'包含为或约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个同源性碱基对。在一些实施方案中,同源臂在转基因和/或CD247基因座的靶位点的3'包含为或约100至500、200至400或250至350个同源性碱基对。在一些实施方案中,同源臂在CD247基因座的靶位点的5'包含少于约100、90、80、70、60、50、40、30、20、15或10个同源性碱基对。

[0580] 在一些实施方案中,同源臂在CD247基因座的靶位点的5'包含为或约10、20、30、

40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个同源性碱基对。在一些实施方案中，同源臂在转基因和/或CD247基因座的靶位点的5'包含为或约100至500、200至400或250至350个同源性碱基对。在一些实施方案中，同源臂在CD247基因座的靶位点的3'包含少于约100、90、80、70、60、50、40、30、20、15或10个同源性碱基对。

[0581] 在一些实施方案中，5'同源臂的3'端是与转基因的5'端相邻的位置。在一些实施方案中，5'同源臂可以从转基因的5'端向5'延伸至少或至少约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸。

[0582] 在一些实施方案中，3'同源臂的5'端是与转基因的3'端相邻的位置。在一些实施方案中，3'同源臂可以从转基因的3'端向3'延伸至少或至少约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸。

[0583] 在一些实施方案中，对于靶向插入，同源臂（例如，5'和3'同源臂）可以各自包含侧接最远端靶位点的序列的约1000个碱基对（bp）（例如，在突变任一侧上的序列的1000bp）。

[0584] 示例性同源臂长度包括至少或至少约50、100、200、250、300、400、500、600、700、750、800、900、1000、2000、3000、4000或5000个核苷酸。在一些实施方案中，同源臂长度是为或约50-100、100-250、250-500、500-750、750-1000、1000-2000、2000-3000、3000-4000或4000-5000个核苷酸。示例性同源臂长度包括少于或少于约或者为或为约50、100、200、250、300、400、500、600、700、750、800、900、1000、2000、3000、4000或5000个核苷酸。在一些实施方案中，同源臂长度是为或约50-100、100-250、250-500、500-750、750-1000、1000-2000、2000-3000、3000-4000或4000-5000个核苷酸。示例性同源臂长度包括从为或约100至为或约1000个核苷酸、从为或约100至为或约750个核苷酸、从为或约100至为或约600个核苷酸、从为或约100至为或约400个核苷酸、从为或约100至为或约300个核苷酸、从为或约100至为或约200个核苷酸、从为或约200至为或约1000个核苷酸、从为或约200至为或约750个核苷酸、从为或约200至为或约600个核苷酸、从为或约200至为或约400个核苷酸、从为或约200至为或约300个核苷酸、从为或约300至为或约1000个核苷酸、从为或约300至为或约750个核苷酸、从为或约300至为或约600个核苷酸、从为或约300至为或约400个核苷酸、从为或约400至为或约1000个核苷酸、从为或约400至为或约750个核苷酸、从为或约400至为或约600个核苷酸、从为或约600至为或约1000个核苷酸、从为或约600至为或约750个核苷酸或者750至为或约1000个核苷酸。

[0585] 在一些任何此类实施方案中，转基因是通过引入多个T细胞的每一个中的模板多核苷酸来整合。在特定实施方案中，模板多核苷酸包含结构[5'同源臂]-[转基因]-[3'同源臂]。在某些实施方案中，5'同源臂和3'同源臂包含与至少或至少约一个靶位点周围的核酸序列同源的核酸序列。在一些实施方案中，5'同源臂包含与靶位点5'的核酸序列同源的核酸序列。在特定实施方案中，3'同源臂包含与靶位点3'的核酸序列同源的核酸序列。在某些实施方案中，5'同源臂和3'同源臂独立地具有至少或至少约或者至少或至少约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸，或者少于或少于约10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸。在一些实施方案中，5'同源臂和3'同源臂独立地具有在为或约50与为或约100之间、100与为或约250之间、250与为或约500之间、500与为或约750之间、750与为或约1000之间、1000与为或约2000之间的核苷酸。在一些任何此类实施方案中，5'同源臂和3'同源臂独立

地具有在为或约50与为或约100个核苷酸之间的长度、为或约100与为或约250个核苷酸之间的长度、为或约250与为或约500个核苷酸之间的长度、为或约500与为或约750个核苷酸之间的长度、为或约750与为或约1000个核苷酸之间的长度或者为或约1000与为或约2000个核苷酸之间的长度。

[0586] 在特定实施方案中,5'同源臂和3'同源臂独立地具有从为或约100至为或约1000个核苷酸、从为或约100至为或约750个核苷酸、从为或约100至为或约600个核苷酸、从为或约100至为或约400个核苷酸、从为或约100至为或约300个核苷酸、从为或约100至为或约200个核苷酸、从为或约200至为或约1000个核苷酸、从为或约200至为或约750个核苷酸、从为或约200至为或约600个核苷酸、从为或约200至为或约400个核苷酸、从为或约200至为或约300个核苷酸、从为或约300至为或约1000个核苷酸、从为或约300至为或约750个核苷酸、从为或约300至为或约600个核苷酸、从为或约300至为或约400个核苷酸、从为或约400至为或约1000个核苷酸、从为或约400至为或约750个核苷酸、从为或约400至为或约600个核苷酸、从为或约600至为或约1000个核苷酸、从为或约600至为或约750个核苷酸或者从为或约750至为或约1000个核苷酸。在特定实施方案中,5'同源臂和3'同源臂独立地具有从为或约100至为或约1000个核苷酸、从为或约100至为或约750个核苷酸、从为或约100至为或约600个核苷酸、从为或约100至为或约400个核苷酸、从为或约100至为或约300个核苷酸、从为或约100至为或约200个核苷酸、从为或约200至为或约1000个核苷酸、从为或约200至为或约750个核苷酸、从为或约200至为或约600个核苷酸、从为或约200至为或约400个核苷酸、从为或约200至为或约300个核苷酸、从为或约300至为或约1000个核苷酸、从为或约300至为或约750个核苷酸、从为或约300至为或约600个核苷酸、从为或约300至为或约400个核苷酸、从为或约400至为或约1000个核苷酸、从为或约400至为或约750个核苷酸、从为或约400至为或约600个核苷酸、从为或约600至为或约1000个核苷酸、从为或约600至为或约750个核苷酸或者从为或约750至为或约1000个核苷酸的长度。在一些实施方案中,5'同源臂和3'同源臂独立地具有为或约200、300、400、500、600、700或800个核苷酸的长度,或者任何前述值之间的任何值。在一些实施方案中,5'同源臂和3'同源臂独立地具有大于或大于约300个核苷酸的长度,任选地其中5'同源臂和3'同源臂独立地具有为或约400、500或600个核苷酸的长度或者任何前述值之间的任何值。在一些实施方案中,5'同源臂和3'同源臂独立地具有大于或大于约300个核苷酸的长度。

[0587] 在一些实施方案中,一个或多个同源臂含有与编码CD3 ζ 链或其片段的序列同源的核苷酸序列。在一些实施方案中,一个或多个同源臂与编码嵌合受体或其部分的转基因序列框内相连或连接。因此,在一些实施方案中,一个或多个同源臂和转基因一起编码大于仅由所述转基因编码的CD3 ζ 链的部分的CD3 ζ 链或其片段。在一些实施方案中,一个或多个同源臂与转基因的组合一起含有与编码CD3 ζ 链的内源基因、基因座或开放阅读框(CD247)的完整外显子同源的序列。在一些实施方案中,一个或多个同源臂含有与编码CD3 ζ 链的内源基因、基因座或开放阅读框(CD247)的内含子的全部或一部分同源的核苷酸序列。

[0588] 在一些实施方案中,采用替代性HDR。在一些实施方案中,当模板多核苷酸具有与靶位点的5'延伸同源性(即,在靶位点链的5'方向上)时,替代性HDR更有效地继续进行。因此,在一些实施方案中,模板多核苷酸具有较长的同源臂和较短的同源臂,其中较长的同源臂可以与靶位点的5'退火。在一些实施方案中,可以与靶位点5'退火的臂距所述靶位点或

者转基因的5'或3'端至少25、50、75、100、125、150、175、或200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸。在一些实施方案中,可以与靶位点5'退火的臂比可以与靶位点3'退火的臂长至少10%、20%、30%、40%或50%。在一些实施方案中,可以与靶位点5'退火的臂比可以与靶位点3'退火的臂长至少2x、3x、4x或5x。根据ssDNA模板是可以与完整链退火还是可以与靶向链退火,与靶位点5'退火的同源臂分别可以位于ssDNA模板的5'端或ssDNA模板的3'端处。

[0589] 类似地,在一些实施方案中,模板多核苷酸具有5'同源臂、转基因和3'同源臂,使得模板多核苷酸含有与靶位点的5'的延伸同源性。例如,5'同源臂和3'同源臂可以具有基本上相同的长度,但是转基因向靶位点的5'的延伸可以比向靶位点的3'的延伸更远。在一些实施方案中,同源臂向靶位点5'端的延伸比向靶位点3'端的延伸远至少10%、20%、30%、40%、50%、2x、3x、4x或5x。

[0590] 在一些实施方案中,在模板多核苷酸以靶位点为中心时,替代性HDR更有效地继续进行。因此,在一些实施方案中,模板多核苷酸具有两个尺寸基本上相同的同源臂。在一些实施方案中,模板多核苷酸的第一同源臂(例如,5'同源臂)的长度在模板多核苷酸的第二同源臂(例如,3'同源臂)的10%、9%、8%、7%、6%、5%、4%、3%、2%或1%内。

[0591] 类似地,在一些实施方案中,模板多核苷酸具有5'同源臂、转基因和3'同源臂,使得模板多核苷酸在靶位点的任一侧上延伸基本上相同的距离。例如,同源臂可以具有不同的长度,但是可以选择转基因以对此进行补偿。例如,转基因向靶位点5'的延伸可以远于其向靶位点3'的延伸,但是靶位点5'的同源臂短于靶位点3'的同源臂以进行补偿。反过来也是可能的,例如,转基因向靶位点3'的延伸可以远于其向靶位点5'的延伸,但是靶位点3'的同源臂短于靶位点5'的同源臂以进行补偿。

[0592] 在一些实施方案中,模板多核苷酸(包括转基因序列和一个或多个同源臂)的长度为在或在约1000至约20,000个碱基对之间,如约1000、1500、2000、2500、3000、3500、4000、4500、5000、6000、7000、8000、9000、10000、11000、12000、13000、14000、15000、16000、17000、18000、19000或20000个碱基对。在一些实施方案中,模板多核苷酸的长度受到以下的限制:可以制备、合成或组装和/或引入细胞中的多核苷酸的最大长度或病毒载体的容量,以及多核苷酸或载体的类型。在一些方面,模板多核苷酸的有限容量可以决定转基因序列和/或一个或多个同源臂的长度。在一些方面,转基因序列与一个或多个同源臂的组合总长度必须在多核苷酸或载体的最大长度或容量内。例如,在一些方面,模板多核苷酸的转基因部分具有约1000、1500、2000、2500、3000、3500或4000个碱基对,并且如果模板多核苷酸的最大长度为约5000个碱基对,则所述序列的其余部分可以划分在一个或多个同源臂之间,例如,使得3'或5'同源臂可以具有大约500、750、1000、1250、1500、1750或2000个碱基对。

[0593] 在一些方面,所提供的实施方案允许使用与现有方法相比更小或更短的核酸序列片段进行工程化。例如,在一些实施方案中,转基因序列编码嵌合受体的CD3 ζ 或其部分的一部分或不编码所述CD3 ζ 或其部分。通过利用内源CD247基因的开放阅读框序列的一部分或全部来编码嵌合受体的CD3 ζ 或其部分,所提供的实施方案允许使用更小的转基因序列。在一些实施方案中,编码嵌合受体的一部分的转基因序列与内源CD247开放阅读框的一个或多个外显子框内融合,并且因此可以处于内源CD247调节元件的控制下。因此,在一些情形

中,不需要异源调节或控制元件,并且与需要表达调节或控制元件的现有方法相比,转基因序列可以更小。在一些方面,实施方案允许使用比现有方法中使用的转基因序列(例如,其包括初级信号传导区域(例如,CD3 ζ 链或其片段)的整个长度和/或含有异源调节元件)更小或更短的转基因序列,例如,小大约100至1000个碱基对的转基因序列,例如,小约100、200、300、400、500、600、700、800、900或1000个碱基对的转基因序列。因此,在多核苷酸的最大长度受限的情形中,所提供的实施方案允许容纳更大的同源臂和/或允许容纳编码另外的分子的核酸序列,因为对编码嵌合受体的一部分的转基因序列的长度需要减小了。在一些方面,与其他方法相比,可以促进或改进核酸序列(例如,转基因序列)的生成、递送和/或通过同源定向修复(HDR)的靶向效率。

3. 模板多核苷酸的递送

[0594] 在一些实施方案中,将多核苷酸(如含有编码嵌合受体或其部分的一条或多条链的转基因序列的模板多核苷酸(例如,如本文章节I.B.2中所述))以核苷酸形式(例如,作为多核苷酸或载体)引入细胞中。在特定实施方案中,多核苷酸含有编码嵌合受体的一部分的转基因序列和一个或多个同源臂,并且可以被引入细胞中用于同源定向修复(HDR)介导的转基因序列的整合。

[0595] 在一些方面,所提供的实施方案细胞的基因工程化,通过引入能够诱导遗传破坏的一种或多种药剂或其组分和模板多核苷酸,以诱导HDR和转基因序列的靶向整合。在一些方面,一种或多种药剂和模板多核苷酸是同时递送。在一些方面,一种或多种药剂和模板多核苷酸是依序递送。在一些实施方案中,一种或多种药剂是在递送多核苷酸之前递送。

[0596] 在一些实施方案中,除了能够诱导靶向遗传破坏的一种或多种药剂(例如核酸酶和/或gRNA)以外,还将模板多核苷酸引入细胞中用于工程化。在一些实施方案中,可以在将能够诱导靶向遗传破坏的一种或多种药剂的一种或多种组分引入细胞中之前、同时或之后递送一种或多种模板多核苷酸。在一些实施方案中,一种或多种模板多核苷酸是与所述药剂同时递送。在一些实施方案中,模板多核苷酸是在所述药剂之前递送,例如,在模板多核苷酸之前数秒至数小时至数天,包括但不限于在所述药剂之前1至60分钟(或其间的任何时间)、在所述药剂之前1至24小时(或其间的任何时间)或者在所述药剂之前超过24小时。在一些实施方案中,模板多核苷酸是在所述药剂之后递送,在模板多核苷酸之后数秒至数小时至数天,包括紧接在递送所述药剂之后,例如,在递送所述药剂之后30秒至4小时之间,如约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、6分钟、8分钟、9分钟、10分钟、15分钟、20分钟、30分钟、40分钟、50分钟、60分钟、90分钟、2小时、3小时或4小时,和/或优选地在递送所述药剂的4小时内。在一些实施方案中,模板多核苷酸是在递送所述药剂之后超过4小时递送。在一些实施方案中,模板多核苷酸是在引入一种或多种药剂之后为或约2小时引入。

[0597] 在一些实施方案中,模板多核苷酸可以使用与能够诱导靶向遗传破坏的一种或多种药剂(例如,核酸酶和/或gRNA)相同的递送系统来递送。在一些实施方案中,模板多核苷酸可以使用与能够诱导靶向遗传破坏的一种或多种药剂(例如,核酸酶和/或gRNA)不同的递送系统来递送。在一些实施方案中,模板多核苷酸是与一种或多种药剂同时递送。在其他实施方案中,模板多核苷酸是在递送一种或多种药剂之前或之后的不同时间递送。本文在章节I.A.3中(例如,在表3和表4中)所述的用于递送能够诱导靶向遗传破坏的一种或多种药剂(例如,核酸酶和/或gRNA)中的核酸的任何递送方法都可以用于递送模板多核苷酸。

[0598] 在一些实施方案中,一种或多种药剂和模板多核苷酸是以相同的形式或方法来递送。例如,在一些实施方案中,一种或多种药剂和模板多核苷酸都包含于载体(例如,病毒载体)中。在一些实施方案中,模板多核苷酸在与Cas9和gRNA相同的载体骨架(例如AAV基因组、质粒DNA)上编码。在一些方面,一种或多种药剂和模板多核苷酸呈不同的形式,例如用于Cas9-gRNA药剂的核糖核酸-蛋白质复合物(RNP)以及用于模板多核苷酸的线性DNA,但它们是使用相同的方法递送。

[0599] 在一些实施方案中,模板多核苷酸是线性或环状核酸分子,如线性或环状DNA或线性RNA,并且可以使用本文在章节I.A.3(例如,本文的表3和表4)中描述的用于将核酸分子递送至细胞中的任何方法来递送。

[0600] 在特定实施方案中,将多核苷酸(例如,模板多核苷酸)以核苷酸形式(例如,作为非病毒载体或在非病毒载体内)引入细胞中。在一些实施方案中,所述非病毒载体是或包括多核苷酸,例如DNA或RNA多核苷酸,其适合于通过用于基因递送的任何合适的和/或已知的非病毒方法进行转导和/或转染,所述非病毒方法例如但不限于显微注射、电穿孔、瞬时细胞压缩或挤压(如Lee等人(2012) Nano Lett 12:6322-27所述)、脂质介导的转染、肽介导的递送(例如细胞穿透肽)、或其组合。在一些实施方案中,通过本文所述的非病毒方法,如本文在表4中列出的非病毒方法,将非病毒多核苷酸递送至细胞中。

[0601] 在一些实施方案中,模板多核苷酸序列可以包含于载体分子中,所述载体分子包含与基因组DNA中的目的区域不同源的序列。在一些实施方案中,病毒是DNA病毒(例如,dsDNA或ssDNA病毒)。在一些实施方案中,病毒是RNA病毒(例如,ssRNA病毒)。示例性病毒载体/病毒包括例如逆转录病毒、慢病毒、腺病毒、腺相关病毒(AAV)、牛痘病毒、痘病毒和单纯疱疹病毒、或本文其他地方所述的任何病毒。可以将多核苷酸作为载体分子的部分引入细胞中,所述载体分子具有另外的序列,如例如复制起点、启动子和编码抗生素抗性的基因。此外,模板多核苷酸可以作为裸核酸来引入,作为与诸如脂质体、纳米颗粒或泊洛沙姆的材料复合的核酸来引入,或者可以通过病毒(例如,腺病毒、AAV、疱疹病毒、逆转录病毒、慢病毒和整合酶缺陷慢病毒(IDLV))来递送。

[0602] 在一些实施方案中,可以使用重组感染性病毒颗粒(如例如源自猿猴病毒40(SV40)、腺病毒、腺相关病毒(AAV)的载体)将模板多核苷酸转移到细胞中。在一些实施方案中,使用重组慢病毒载体或逆转录病毒载体将模板多核苷酸转移至T细胞中,所述病毒载体如 γ -逆转录病毒载体(参见例如,Koste等人(2014) Gene Therapy 2014年4月3日.doi:10.1038/gt.2014.25;Carlens等人(2000) Exp Hematol 28(10):1137-46;Alonso-Camino等人(2013) Mol Ther Nucl Acids 2,e93;Park等人,Trends Biotechnol.2011年11月29日(11):550-557)或源自HIV-1的慢病毒载体。

[0603] 在其他方面,通过病毒和/或非病毒基因转移方法递送模板多核苷酸。在一些实施方案中,将模板多核苷酸经由腺相关病毒(AAV)递送至细胞。可以使用任何AAV载体,包括但不限于AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8及其组合。在一些情况下,与衣壳血清型(例如,具有AAV5、AAV6或AAV8衣壳的AAV2 ITR)相比,AAV包括异源血清型的LTR。模板多核苷酸可以使用与用于递送核酸酶的相同基因转移系统(包括在相同载体上)来递送,或者可以使用与用于核酸酶不同的递送系统来递送。在一些实施方案中,使用病毒载体(例如,AAV)来递送模板多核苷酸,并且以mRNA形式来递送一种或多种核酸酶。还可以在递送病毒

载体(例如,携带一种或多种核酸酶和/或模板多核苷酸)之前、同时和/或之后,用抑制病毒载体与如本文所述的细胞表面受体的结合的一种或多种分子来处理细胞。

[0604] 在一些实施方案中,逆转录病毒载体具有长末端重复序列(LTR),例如源自莫洛尼鼠白血病病毒(MoMLV)、骨髓增生性肉瘤病毒(MPSV)、鼠胚胎干细胞病毒(MESV)、鼠干细胞病毒(MSCV)或脾病灶形成病毒(SFFV)的重组逆转录病毒载体。大多数逆转录病毒载体源自鼠逆转录病毒。在一些实施方案中,逆转录病毒包括源自任何禽类或哺乳动物细胞来源的那些。逆转录病毒通常是双嗜性的,这意味着它们能够感染包括人在内的若干个物种的宿主细胞。在一个实施方案中,要表达的基因替代逆转录病毒gag、pol和/或env序列。已经描述了多种例示性逆转录病毒系统(例如,美国专利号5,219,740、6,207,453、5,219,740; Miller和Rosman(1989) *BioTechniques* 7:980-990; Miller, A.D. (1990) *Human Gene Therapy* 1:5-14; Scarpa等人(1991) *Virology* 180:849-852; Burns等人(1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; 以及 Boris-Lawrie和Temin(1993) *Cur. Opin. Genet. Develop.* 3:102-109)。

[0605] 在一些实施方案中,使用AAV载体递送模板多核苷酸,并且以不同的形式(如以编码核酸酶和/或gRNA的mRNA)递送能够诱导靶向遗传破坏的一种或多种药剂(如核酸酶和/或gRNA)。在一些实施方案中,使用相同类型的方法(如病毒载体)但是在分开的载体上递送模板多核苷酸和核酸酶。在一些实施方案中,在与能够诱导遗传破坏的药剂(如核酸酶和/或gRNA)不同的递送系统中递送模板多核苷酸。用于递送的核酸和载体的类型包括本文在章节III中描述的那些类型中的任一种。

[0606] 在一些实施方案中,模板多核苷酸和核酸酶可以位于相同的载体(例如AAV载体(如AAV6))上。在一些实施方案中,使用AAV载体递送模板多核苷酸,并且以不同的形式(如以编码核酸酶和/或gRNA的mRNA)递送能够诱导靶向遗传破坏的一种或多种药剂(如核酸酶和/或gRNA)。在一些实施方案中,使用相同类型的方法(如病毒载体)但是在分开的载体上递送模板多核苷酸和核酸酶。在一些实施方案中,在与能够诱导遗传破坏的药剂(如核酸酶和/或gRNA)不同的递送系统中递送模板多核苷酸。在一些实施方案中,在体内从载体骨架切除模板多核苷酸,如所述模板多核苷酸的侧翼为gRNA识别序列。在一些实施方案中,模板多核苷酸位于作为Cas9和gRNA的单独多核苷酸分子上。在一些实施方案中,将Cas9和gRNA以核糖核蛋白(RNP)复合物的形式引入,并且将模板多核苷酸作为例如在载体或线性核酸分子(如线性DNA)中的多核苷酸分子引入。用于递送的核酸和载体的类型包括本文在章节II中描述的那些类型中的任一种。

[0607] 在一些实施方案中,模板多核苷酸是腺病毒载体,例如AAV载体,例如ssDNA分子,其长度和序列允许将其包装于AAV衣壳中。载体可以为例如小于5kb,并且可以含有促进包装至衣壳中的ITR序列。载体可以是整合缺陷的。在一些实施方案中,模板多核苷酸在转基因和/或靶位点的任一侧上包含约150至1000个同源性核苷酸。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含约100、150、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含至少100、150、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转

基因的3'、或同时在靶位点或转基因的5'和3'包含至多100、150、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸。

[0608] 在一些实施方案中,模板多核苷酸是慢病毒载体,例如,IDLV(整合缺陷性慢病毒)。在一些实施方案中,模板多核苷酸在转基因和/或靶位点的任一侧上包含约500至1000个同源性碱基对。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含约300、400、500、600、700、800、900、1000、1500或2000个同源性碱基对。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含至少300、400、500、600、700、800、900、1000、1500或2000个同源性碱基对。在一些实施方案中,模板多核苷酸在靶位点或转基因的5'、在靶位点或转基因的3'、或同时在靶位点或转基因的5'和3'包含不超过300、400、500、600、700、800、900、1000、1500或2000个同源性碱基对。在一些实施方案中,模板多核苷酸包含防止Cas9识别并切割模板多核苷酸的一个或多个突变(例如,沉默突变)。相对于要改变的细胞的基因组中的相应序列,模板多核苷酸可以包含例如至少1、2、3、4、5、10、20或30个沉默突变。在一些实施方案中,相对于要改变的细胞的基因组中的相应序列,模板多核苷酸包含至多2、3、4、5、10、20、30或50个沉默突变。在一些实施方案中,cDNA包含防止Cas9识别并切割模板多核苷酸的一个或多个突变(例如,沉默突变)。相对于要改变的细胞的基因组中的相应序列,模板多核苷酸可以包含例如至少1、2、3、4、5、10、20或30个沉默突变。在一些实施方案中,相对于要改变的细胞的基因组中的相应序列,模板多核苷酸包含至多2、3、4、5、10、20、30或50个沉默突变。

[0609] 本文所述的双链模板多核苷酸可以包括一个或多个非天然碱基和/或骨架。特定地,具有甲基化胞嘧啶的模板多核苷酸的插入可以使用本文所述的方法来进行,以在目的区域中实现转录静止的状态。

II. 核酸、载体和递送

[0610] 在一些实施方案中,将多核苷酸(如多核苷酸,如编码嵌合受体或其部分的一条或多条链的模板多核苷酸)以核苷酸形式(如多核苷酸或载体)引入细胞中。在特定实施方案中,多核苷酸含有编码嵌合受体或其部分的转基因。在某些实施方案中,将用于遗传破坏的一种或多种药剂或其组分以核酸形式(如多核苷酸和/或载体)引入细胞中。在一些实施方案中,用于工程化的组分可以使用多种递送方法以多种形式来递送,所述递送方法包括如本文的章节I.A.3和表3和表4中所述的用于递送一种或多种药剂的任何合适的方法。还提供编码能够诱导遗传破坏的所述一种或多种药剂的一种或多种组分的一种或多种多核苷酸(如核酸分子)(例如,本文在章节I.A中所述的任一种)。还提供含有转基因序列的一种或多种模板多核苷酸(例如,本文在章节I.B.2中所述的任一种)。还提供包括一种或多种此类多核苷酸(如模板多核苷酸或编码能够诱导遗传破坏的一种或多种药剂的一种或多种组分的多核苷酸)的载体(如用于基因工程化细胞以供靶向整合转基因的载体)。

[0611] 在一些实施方案中,提供多核苷酸,如用于将转基因靶向于特定基因组靶位置处(如在CD247基因座处)的模板多核苷酸。在一些实施方案中,提供本文在章节I.B中所述的任何模板多核苷酸。在一些实施方案中,模板多核苷酸含有包括编码嵌合受体或其部分或其他多肽和/或因子的核酸序列的转基因,以及用于靶向整合的同源臂。在一些实施方案中,模板多核苷酸可以含于载体中。

[0612] 在一些实施方案中,能够诱导遗传破坏的药剂可以编码于一种或多种多核苷酸中。在一些实施方案中,所述药剂的组分(如Cas9分子和/或gRNA分子)可以编码于一种或多种多核苷酸中,并被引入细胞中。在一些实施方案中,编码药剂的一种或多种组分的多核苷酸可以包括于载体中。

[0613] 在一些实施方案中,载体可以包含编码Cas9分子和/或gRNA分子的序列和/或模板多核苷酸。在一些方面,载体还可以包含如与Cas9分子序列融合的编码信号肽(如用于核定位、核仁定位、线粒体定位)的序列。例如,载体可以包含与编码Cas9分子的序列融合的核定位序列(如来自SV40)。在一些实施方案中,提供用于基因工程化细胞以供靶向整合含于多核苷酸(如章节I.B.2中所述的模板多核苷酸)中的转基因序列的载体。

[0614] 在特定实施方案中,一种或多种调节/控制元件(如启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、内部核糖体进入位点(IRES)、2A序列和剪接受体或供体)可以包括于载体中。在一些实施方案中,启动子选自RNA pol I、pol II或pol III启动子。在一些实施方案中,启动子由RNA聚合酶II识别(如CMV、SV40早期区域或腺病毒主要晚期启动子)。在另一个实施方案中,启动子由RNA聚合酶III识别(如U6或H1启动子)。

[0615] 在某些实施方案中,启动子是受调节的启动子(如诱导型启动子)。在一些实施方案中,启动子是诱导型启动子或阻抑型启动子。在一些实施方案中,启动子包含Lac操纵基因序列、四环素操纵基因序列、半乳糖操纵基因序列或多西环素操纵基因序列,或者是其类似物,或者能够被Lac阻遏物或四环素阻遏物或其类似物结合或识别。

[0616] 在一些实施方案中,启动子是或包含组成型启动子。示例性组成型启动子包括例如猿病毒40早期启动子(SV40)、巨细胞病毒立即早期启动子(CMV)、人泛素C启动子(UBC)、人延伸因子1 α 启动子(EF1 α)、小鼠磷酸甘油酸激酶1启动子(PGK)、和与CMV早期增强子偶联的鸡 β -肌动蛋白启动子(CAGG)。在一些实施方案中,组成型启动子是合成或修饰的启动子。在一些实施方案中,启动子是或包含MND启动子,它是合成启动子,含有修饰的MoMuLV LTR的U3区和骨髓增生性肉瘤病毒增强子(SEQ ID NO:18或126所示的序列;参见Challita等人(1995) J.Virol.69(2):748-755)。在一些实施方案中,启动子是组织特异性启动子。在另一个实施方案中,启动子是病毒启动子。在另一个实施方案中,启动子是非病毒启动子。在一些实施方案中,示例性启动子可以包括但不限于人延伸因子1 α (EF1 α)启动子(如SEQ ID NO:77或118中所示)或其修饰的形式(具有HTLV1增强子的EF1 α 启动子;如SEQ ID NO:119中所示)或MND启动子(如SEQ ID NO:131中所示)。在一些实施方案中,多核苷酸和/或载体不包括调节元件(例如启动子)。

[0617] 在特定实施方案中,将多核苷酸(例如,编码嵌合受体或其部分的多核苷酸)以核苷酸形式(例如,作为非病毒载体或在非病毒载体内)引入细胞中。在一些实施方案中,多核苷酸是DNA或RNA多核苷酸。在一些实施方案中,多核苷酸是双链或单链多核苷酸。在一些实施方案中,非病毒载体是或包括多核苷酸(例如DNA或RNA多核苷酸),其适合于通过用于基因递送的任何合适的和/或已知的非病毒方法进行转导和/或转染,所述非病毒方法例如但不限于显微注射、电穿孔、瞬时细胞压缩或挤压(如Lee等人(2012) Nano Lett 12:6322-27所述)、脂质介导的转染、肽介导的递送或其组合。在一些实施方案中,通过本文所述的非病毒方法(如在表4中列出的非病毒方法)将非病毒多核苷酸递送至细胞中。

[0618] 在一些实施方案中,载体或递送媒介物是病毒载体(例如,用于产生重组病毒)。在

一些实施方案中,病毒是DNA病毒(例如,dsDNA或ssDNA病毒)。在一些实施方案中,病毒是RNA病毒(例如,ssRNA病毒)。示例性病毒载体/病毒包括例如逆转录病毒、慢病毒、腺病毒、腺相关病毒(AAV)、牛痘病毒、痘病毒和单纯疱疹病毒、或本文其他地方所述的任何病毒。

[0619] 在一些实施方案中,病毒感染分裂细胞。在另一个实施方案中,病毒感染非分裂细胞。在另一个实施方案中,病毒感染分裂和非分裂两种细胞。在另一个实施方案中,病毒可以整合至宿主基因组中。在另一个实施方案中,病毒被工程化以具有例如在人体内降低的免疫性。在另一个实施方案中,病毒有复制能力。在另一个实施方案中,病毒是复制缺陷型的,例如另外几轮病毒体复制和/或包装所需的基因的一个或多个编码区被其他基因替代或缺失。在另一个实施方案中,出于短暂诱导遗传破坏的目的,病毒引起Cas9分子和/或gRNA分子的短暂表达。在另一个实施方案中,病毒引起Cas9分子和/或gRNA分子的长期(例如,至少1周、2周、1个月、2个月、3个月、6个月、9个月、1年、2年)或永久表达。病毒的包装容量可以例如从至少约4kb到至少约30kb变化,例如至少约5kb、10kb、15kb、20kb、25kb、30kb、35kb、40kb、45kb或50kb。

[0620] 在一些实施方案中,含有一种或多种药剂的多核苷酸和/或模板多核苷酸是通过重组逆转录病毒来递送。在另一个实施方案中,逆转录病毒(例如,莫洛尼鼠白血病毒)包含例如允许整合至宿主基因组中的逆转录酶。在一些实施方案中,逆转录病毒有复制能力。在另一个实施方案中,逆转录病毒是复制缺陷型的,例如另外几轮病毒体复制和包装所必需的基因的一个或多个编码区被其他基因替代或缺失。

[0621] 在一些实施方案中,含有一种或多种药剂的多核苷酸和/或模板多核苷酸是通过重组慢病毒来递送。例如,慢病毒是复制缺陷型的,例如不包含病毒复制所需的一个或多个基因。

[0622] 在一些实施方案中,含有一种或多种药剂的多核苷酸和/或模板多核苷酸是通过重组腺病毒来递送。在另一个实施方案中,腺病毒被工程化以在人体内具有降低的免疫性。

[0623] 在一些实施方案中,含有一种或多种药剂的多核苷酸和/或模板多核苷酸是通过重组AAV来递送。在一些实施方案中,AAV可以将其基因组掺入宿主细胞(例如,如本文所述的靶细胞)的基因组中。在另一个实施方案中,AAV是自身互补的腺相关病毒(scAAV),例如包装一起退火以形成双链DNA的两条链的scAAV。可以用于所公开的方法中的AAV血清型包括AAV1、AAV2、修饰的AAV2(例如,在Y444F、Y500F、Y730F和/或S662V处的修饰)、AAV3、修饰的AAV3(例如,在Y705F、Y731F和/或T492V处的修饰)、AAV4、AAV5、AAV6、修饰的AAV6(例如,在S663V和/或T492V处的修饰)、AAV7、AAV8、AAV 8.2、AAV9、AAV.rh10、修饰的AAV.rh10、AAV.rh32/33、修饰的AAV.rh32/33、AAV.rh43、修饰的AAV.rh43、AAV.rh64R1、修饰的AAV.rh64R1,并且假型化AAV(如AAV2/8、AAV2/5和AAV2/6)也可以用于所公开的方法中。

[0624] 在一些实施方案中,含有一种或多种药剂的多核苷酸和/或模板多核苷酸是通过杂合病毒(例如,本文所述一种或多种病毒的杂合体)来递送。

[0625] 包装细胞用于形成能够感染靶细胞的病毒颗粒。这种细胞包括可以包装腺病毒的293细胞和可以包装逆转录病毒的 ψ 2细胞或PA317细胞。用于基因疗法中的病毒载体通常由将核酸载体包装到病毒颗粒中的生产细胞系产生。载体通常含有包装和随后整合至宿主或靶细胞中(如果适用)所需的最小病毒序列,并且其他病毒序列被编码要表达的蛋白质(例如Cas9)的表达盒替代。例如,用于基因疗法中的AAV载体通常仅具有来自AAV基因组的反向

末端重复 (ITR) 序列,所述序列为在宿主或靶细胞中进行包装和基因表达所需。失去的病毒功能由包装细胞系反式提供。此后,将病毒DNA包装在细胞系中,所述细胞系含有编码其他 AAV基因 (即,rep和cap) 但缺乏 ITR序列的辅助质粒。细胞系还被作为辅助者的腺病毒感染。辅助病毒促进 AAV载体的复制和从辅助质粒表达 AAV基因。由于缺少 ITR序列,辅助质粒并未被大量包装。腺病毒的污染可以通过例如热处理来减少,腺病毒对热处理比 AAV更敏感。

[0626] 在一些实施方案中,病毒载体具有细胞类型识别的能力。例如,病毒载体可以用不同的/替代的病毒包膜糖蛋白假型化;用细胞类型特异性受体工程化 (例如,病毒包膜糖蛋白的遗传修饰以掺入靶向配体,如肽配体、单链抗体、生长因子);和/或工程化以具有分子桥,所述分子桥具有双重特异性,其一端识别病毒糖蛋白,并且另一端识别靶细胞表面的部分 (例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合)。

[0627] 在一些实施方案中,病毒载体实现细胞类型特异性表达。例如,可以构建组织特异性启动子以将能够引入遗传破坏的药剂 (例如,Cas9和gRNA) 的表达限制在仅特定靶细胞中。载体的特异性也可以通过微小RNA依赖性表达控制来介导。在一些实施方案中,病毒载体具有增加的融合病毒载体与靶细胞膜的效率。例如,可以掺入融合蛋白 (如有融合能力的血凝素 (HA)),以增加病毒摄取到细胞中。在一些实施方案中,病毒载体具有核定位能力。例如,可以改变需要核膜分解 (在细胞分裂期间) 并因此不感染非分裂细胞的病毒,以在病毒的基质蛋白中掺入核定位肽,从而使得能够转导非增殖细胞。III. 表达嵌合受体的工程化细胞和细胞组合物

[0628] 本文提供包含修饰的CD247基因座的基因工程化的细胞,所述基因座包含编码嵌合受体 (如嵌合抗原受体 (CAR)) 或其部分的一条或多条链的核酸序列 (如转基因)。在一些方面,所述基因工程化的细胞中修饰的CD247基因座包含整合至内源CD247基因座中的编码嵌合受体或其部分的一条或多条链的外源核酸序列 (例如,转基因序列)。在一些方面,所提供的工程化细胞是使用本文所述的方法来产生,例如,所述方法涉及通过采用用于诱导遗传破坏 (例如,章节I.A中所述) 的一种或多种药剂和含有用于修复的转基因序列的模板多核苷酸 (例如,章节I.B中所述) 进行同源依赖修复 (HDR)。在一些方面,可以靶向所提供的多核苷酸 (如章节I.B中所述的任何模板多核苷酸) 的部分 (例如,连续区段) 用于整合于内源CD247基因座处,以生成含有包含编码嵌合受体的核酸序列的修饰的CD247基因座的细胞。在一些方面,所编码的嵌合受体包括包含CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些实施方案中,通过HDR整合至内源CD247基因座中的模板多核苷酸的部分包括转基因序列部分,如本文例如在章节I.B中所述的模板多核苷酸的任何转基因序列部分。

[0629] 在一些方面,将细胞工程化以表达嵌合受体,如嵌合抗原受体 (CAR)。在一些方面,嵌合受体是由存在于工程化细胞中修饰的CD247基因座处的核酸序列编码。在一些方面,所述细胞是通过经由HDR整合编码嵌合受体的全部或一部分的转基因序列来生成。在一些实施方案中,嵌合受体含有结合至或识别配体或抗原 (例如,与疾病或障碍相关的抗原) 的结合结构域。在一些方面,嵌合受体含有细胞内区域,所述细胞内区域含有信号传导区域,例如,包含CD3zeta (CD3 ζ) 链或其片段的信号传导区域,如CD3 ζ 的信号传导区域或信号传导结构域。在一些实施方案中,由所述细胞表达的嵌合受体通常在所述受体的C末端含有CD3 ζ 或其部分。在一些方面,存在于修饰的CD247基因座处的编码嵌合受体或其部分的核酸序列包含与内源CD247基因座的开放阅读框或其部分序列融合的外源核酸序列。在一些实施方案

中,嵌合受体中的CD3 ζ 链的至少一部分是由基因组中的内源CD247基因座的序列编码。在一些方面,CD3 ζ 链或其部分是功能性CD3 ζ 链或部分。在一些方面,嵌合受体能够经由CD3 ζ 链或其片段进行信号传导。

[0630] 在一些方面,工程化细胞是免疫细胞,如T细胞。在一些方面,将免疫细胞工程化以表达嵌合受体,例如,嵌合抗原受体或修饰的嵌合受体,如本文所述的任一种。

[0631] 在一些实施方案中,本文提供的方法、组合物、制品和/或试剂盒可用于生成、制造或产生基因工程化的细胞,例如,基因工程化的免疫细胞和/或T细胞,所述细胞具有或含有修饰的CD247基因座用于表达含有CD3zeta (CD3 ζ) 信号传导结构域的嵌合受体。在特定实施方案中,本文提供的方法得到基因工程化的细胞,所述细胞具有或含有修饰的CD247基因座。在特定实施方案中,修饰的基因座是或含有转基因(例如,章节I.B中所述的转基因)与内源CD247基因的开放阅读框的融合物。在某些实施方案中,转基因编码嵌合受体的包括CD3zeta (CD3 ζ) 链或其片段的部分,并且被框内插入内源CD247基因的开放阅读框中,从而得到编码完整嵌合受体的修饰的基因座。在一些实施方案中,嵌合受体是嵌合抗原受体(CAR)。

[0632] 在一些情形中,将细胞工程化以表达一种或多种另外的分子,例如,另外的因子和/或辅助分子,如本文所述的任何另外的分子,包括治疗性分子。在一些实施方案中,另外的分子可以包括标记、另外的嵌合受体多肽链、抗体或其抗原结合片段、免疫调节分子、配体、细胞因子或趋化因子。在一些实施方案中,另外的因子是可溶分子。在一些实施方案中,另外的因子是膜结合分子。在一些方面,另外的因子可以用于克服或抵消免疫抑制性环境(如肿瘤微环境(TME))的影响。在一些方面,示例性另外的分子包括细胞因子、细胞因子受体、嵌合共刺激受体、共刺激配体和T细胞功能或活性的其他调节剂。在一些实施方案中,由工程化细胞表达的另外的分子包括IL-7、IL-12、IL-15、CD40配体(CD40L)和4-1BB配体(4-1BBL)。在一些方面,另外的分子是结合不同分子的另外的受体,例如,膜结合受体。例如,在一些实施方案中,另外的分子是细胞因子受体或趋化因子受体,例如,IL-4受体或CCL2受体。在一些情况下,工程化细胞被称为“装甲CAR”或针对通用细胞因子杀伤重定向的T细胞(TRUCK)。

[0633] 还提供含有多个工程化细胞的组合物。在一些方面,与使用其他工程化方法(如其中将嵌合受体随机引入细胞基因组中的方法)生成的细胞或细胞组合物相比,含有工程化细胞的组合物展现改进的、均匀的、同质的和/或稳定的嵌合受体的表达和/或抗原结合。在一些实施方案中,工程化细胞或包含所述工程化细胞的组合物可以用于疗法(例如,过继细胞疗法)中。在一些实施方案中,所提供的细胞或细胞组合物可以用于本文所述的任何治疗方法中或者用于本文所述的治疗性用途。

A. 修饰的CD247基因座

[0634] 在一些方面,提供基因工程化的细胞,其包含修饰的CD247基因座。在一些实施方案中,修饰的CD247基因座包含编码嵌合受体的一条或多条链的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些实施方案中,核酸序列包含编码嵌合受体的一条或多条链的一部分的转基因序列,所述转基因序列已经任选地经由同源定向修复(HDR)整合于内源CD247基因座处。在一些实施方案中,CD3 ζ 信号传导结构域的全部(例如,整个或全部CD3 ζ 信号传导结构域)或片段是由内源CD247基因座的开放阅读框或

其部分序列编码。在一些实施方案中,核酸序列包含编码嵌合受体的一部分的转基因序列与内源CD247基因座的开放阅读框或其部分序列的融合物。

[0635] 在一些方面,修饰的CD247基因座是如经由HDR方法通过遗传破坏和转基因序列(例如外源或异源核酸序列)的整合来生成,所述转基因序列包括编码嵌合受体或其部分的核苷酸序列。在一些方面,存在于修饰的CD247基因座处的核酸序列包括整合于内源CD247基因座中的区域处的一个或多个转基因序列(如外源序列),所述区域通常会包括编码全长CD3 ζ 的开放阅读框。在一些方面,在通过HDR靶向整合转基因后,细胞的基因组含有修饰的CD247基因座,所述基因座包含编码嵌合受体或其部分的核酸序列。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有整合至内源CD247基因座的开放阅读框内的位点中的,使得从所述工程化细胞表达嵌合受体的部分,并且也从包含编码CD3 ζ 链的开放阅读框的内源CD247基因座表达CD3 ζ 的一部分。

[0636] 在一些方面,在通过HDR靶向整合转基因后,细胞基因组含有包含编码嵌合受体或其部分的核酸序列的修饰的CD247基因座。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有转基因与内源CD247基因座的开放阅读框或其部分序列的融合物。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有整合至内源CD247基因座的开放阅读框内的位点中的转基因。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有编码完整、全部和/或全长嵌合受体的核酸序列(例如,DNA序列),所述嵌合受体的一部分是由转基因(例如,由整合的转基因或异源序列)编码,并且所述嵌合受体的其余部分是由内源CD247基因座的开放阅读框的一部分(例如,内源CD247基因座的开放阅读框的一个或多个外显子)编码。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有编码嵌合受体的核酸序列(例如,DNA序列),所述嵌合受体包含完整、全部、全长和/或整个CD3 ζ 信号传导结构域。例如,CD3 ζ 信号传导结构域是功能性的并且能够进行信号转导。在此类例子中,所述CD3 ζ 信号传导结构域的一部分是由转基因(例如,由整合的转基因或异源序列)编码,并且所述CD3 ζ 信号传导结构域的其余部分是由内源CD247基因座的开放阅读框的一部分(例如,内源CD247基因座的开放阅读框的一个或多个外显子)编码。因此,转基因序列和CD247基因座一起编码嵌合受体的CD3 ζ 信号传导结构域。在其他例子中,所编码的嵌合受体的整个CD3 ζ 信号传导结构域是存在于内源CD247基因座的开放阅读框中的所编码序列,如由内源CD247基因座的开放阅读框的一个或多个外显子编码。

[0637] 在一些实施方案中,所编码的嵌合受体是含有细胞内区域的受体,所述细胞内区域包含CD3 ζ 链的全部或一部分(例如,CD3 ζ 链的细胞内区域,例如CD3 ζ 信号传导结构域)。在一些任何实施方案中,所编码的嵌合受体(例如,由修饰的CD247基因座编码)的细胞内区域包含CD3zeta (CD3 ζ) 信号传导结构域,如整个或全部CD3 ζ 信号传导结构域(例如,全长CD3 ζ 信号传导结构域,在一些例子中,包含选自SEQ ID NO:13-15中任一项的序列)。在一些任何实施方案中,整个CD3 ζ 信号传导结构域能够进行信号传导或信号转导。在一些方面,所编码的嵌合受体包含整个CD3 ζ 信号传导结构域(例如,包含选自SEQ ID NO:13-15中任一项的序列),其部分或完全由内源CD247基因座的开放阅读框或其部分序列编码。在一些实施方案中,内源CD247基因座的开放阅读框或其部分序列编码嵌合受体的CD3 ζ 信号传导结构域。在一些实施方案中,核酸序列包含编码嵌合受体一部分的转基因序列,所述部分任选地编码CD3 ζ 信号传导结构域的片段,并且其中所述开放阅读框或其部分序列编码CD3 ζ 信号传导

结构域,或者任选地CD3 ζ 信号传导结构域的另一片段。在一些实施方案中,转基因含有仅编码CD3 ζ 链的一部分或CD3 ζ 链的细胞内区域的一部分的核苷酸序列。在一些实施方案中,转基因不包含全长CD3 ζ 链。在一些实施方案中,转基因不包括编码CD3 ζ 链的核酸序列。在一些方面,在整合转基因后,编码嵌合受体的CD3 ζ 链或其片段的核酸序列的一些或全部源自或来源于内源CD247基因座的开放阅读框序列或其部分序列。因此,在一些实施方案中,转基因的整合生成转基因与CD247基因座的内源序列的基因融合物,它们可以一起编码含有CD3 ζ 链或其片段的嵌合受体。

[0638] 在一些实施方案中,转基因包含编码嵌合受体(例如,CAR)的细胞外区域、跨膜结构域和细胞内区域中的一个或多个的核苷酸序列。在一些方面,细胞内区域的一部分包含CD3 ζ 链或其片段,并且CD3 ζ 链的仅一部分是由转基因编码。在一些实施方案中,由转基因编码的细胞内区域不包含编码CD3 ζ 链的核苷酸序列。在一些实施方案中,内源CD247基因座的开放阅读框序列或其部分序列编码存在于嵌合受体中的CD3 ζ 链的全部或一部分。例如,在一些实施方案中,在存在于修饰的CD247基因座中的编码嵌合受体的核酸中,编码CD3 ζ 链或其片段的序列的一些或全部源自或来源于内源CD247基因座和/或同源臂序列,所述同源臂序列含有与内源CD247基因座同源的序列。

[0639] 在一些实施方案中,转基因的整合生成转基因与CD247基因座的内源序列的基因融合物,它们一起编码含有CD3 ζ 链或其片段的嵌合受体。

[0640] 在一些实施方案中,在靶向整合后,修饰的CD247基因座含有编码嵌合受体或其部分的核酸序列,其包含内源CD247基因座的至少一个内含子和至少一个外显子。在一些实施方案中,在整合后,嵌合受体的部分是由内源CD247基因座的开放阅读框的部分序列编码,所述部分序列包括内源CD247基因座的至少一个内含子和至少一个外显子。在一些实施方案中,在整合后,嵌合受体的部分是由内源CD247基因座的开放阅读框的部分序列编码,所述部分序列包括内源CD247基因座的至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个内含子。在一些实施方案中,整合的转基因序列不包含编码3'UTR的序列。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有编码嵌合受体的核酸序列,并且所述核酸序列编码内源CD247基因座的3'UTR。

在一些实施方案中,开放阅读框或其部分序列包含内源CD247基因座的3'UTR。在一些实施方案中,在靶向整合后,修饰的CD247基因座含有不包含内含子的转基因。

[0641] 在某些实施方案中,转基因编码嵌合受体(如CAR)的一部分,并且被框内插入编码CD3 ζ 链的CD247基因座的内源开放阅读框内。在一些实施方案中,修饰的基因座编码全长嵌合受体。在特定实施方案中,所编码的嵌合受体的一部分是由存在于转基因中的核酸序列编码,并且所述嵌合受体的其余部分是由存在于内源CD247基因座的开放阅读框中的核酸序列编码。在特定实施方案中,修饰的基因座的转录得到编码嵌合受体的mRNA。在特定实施方案中,所述mRNA的一部分是从存在于转基因中的核酸序列转录,并且所述mRNA的其余部分是从存在于内源CD247基因座的开放阅读框中的核酸序列转录。在一些实施方案中,转基因被整合于紧邻编码CD3 ζ 链的内源CD247基因座的开放阅读框的一个或多个外显子上游且与所述外显子同框的靶位点处。

[0642] 在一些实施方案中,从修饰的基因座转录的mRNA含有3'UTR,所述3'UTR是由内源CD247基因座编码和/或与从内源CD247基因座转录的mRNA的3'UTR相同。在一些实施方案

中,转基因含有编码CAR的部分的核酸序列上游(例如,紧邻上游)的核糖体跳跃元件。在某些实施方案中,修饰的基因座的转录得到编码全长CAR的mRNA。在一些实施方案中,编码CAR的mRNA含有5'UTR,所述5'UTR是由内源基因编码和/或与从内源CD247基因座转录的mRNA的5'UTR相同。

[0643] 在某些实施方案中,修饰的基因座含有编码嵌合受体的核酸序列,并且所述核酸序列含有至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个内含子。在一些实施方案中,开放阅读框或其部分序列包含内源CD247基因座的至少一个内含子和至少一个外显子。在一些实施方案中,所述内含子并不位于整合的转基因序列内。在一些实施方案中,开放阅读框或其部分序列编码内源CD247基因座的3'UTR。在一些实施方案中,转基因序列不包含内含子。在一些实施方案中,转基因序列与内源CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框。

[0644] 在一些实施方案中,由修饰的CD247基因座编码的嵌合受体(例如,CAR)是功能性CAR。在一些实施方案中,由修饰的基因座编码的CAR结合至和/或能够结合至靶抗原。在一些实施方案中,靶抗原与和疾病、障碍或病症相关的细胞或组织相关、为所述细胞或组织所特有和/或在所述细胞或组织上表达。在一些实施方案中,由修饰的CD247基因座编码的嵌合受体(例如,CAR)是能够在T细胞中刺激和/或诱导初级激活信号的功能性CAR、T细胞受体(TCR)组分的信号传导结构域和/或包含免疫受体酪氨酸激活基序(ITAM)的信号传导结构域,如经由CD3-zeta(CD3 ζ)链的细胞内信号传导结构域或区域或其功能变体或信号传导部分。

[0645] 在一些实施方案中,在靶向整合后,修饰的CD247基因座含有转基因序列(例如,编码多肽,例如,嵌合受体的一个或多个结构域或一个或多个区域),并且转基因(外源序列)的编码部分是内源CD247基因座的开放阅读框序列的一个或多个外显子同框整合。在一些实施方案中,转基因序列被整合或插入于内源CD247基因座的开放阅读框的外显子1的下游和外显子8的上游。在一些实施方案中,转基因序列被整合或插入于内源CD247基因座的开放阅读框的外显子1的下游和外显子3的上游。

[0646] 在一些实施方案中,在靶向整合后,修饰的CD247基因座含有编码包含细胞内区域的嵌合受体的核酸序列,其中所述细胞内区域的至少一部分包含由内源CD247基因座的开放阅读框或其部分序列编码的CD3zeta(CD3 ζ)链或其片段。在一些实施方案中,编码细胞内区域的核苷酸序列至少包含内源CD247基因座的开放阅读框的外显子3-8。在一些实施方案中,编码细胞内区域的核苷酸序列包含内源CD247基因座的开放阅读框的外显子2和外显子3-8的至少一部分。在一些实施方案中,编码细胞内信号传导区域的核苷酸序列包含内源CD247基因座的开放阅读框的小于全长的外显子1。在一些实施方案中,编码细胞内区域的核苷酸序列不包含内源CD247基因座的开放阅读框的外显子1和/或包含小于全长的外显子2。

[0647] 在一些方面,所编码的嵌合受体能够经由由转基因序列与CD247基因座的内源序列的融合物编码的CD3 ζ 链或其片段进行信号传导。在一些实施方案中,所编码的CD3 ζ 链或其片段包含人CD3 ζ (登录号:P20963.2)的亚型3的112个AA的细胞内或胞质结构域或如美国专利号7,446,190或美国专利号8,911,993中所述的CD3 ζ 信号传导结构域。在一些实施方案中,所编码的CD3 ζ 链或其片段包含SEQ ID NO:13、14或15中所示的氨基酸序列或者展现与

SEQ ID NO:13、14或15至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列,或其部分序列。

[0648] 在一些实施方案中,转基因序列位于内源CD247基因座的开放阅读框的外显子1的下游和外显子8的上游。在一些实施方案中,转基因序列位于内源CD247基因座的开放阅读框的外显子1的下游和外显子3的上游。在一些实施方案中,所编码的嵌合受体的CD3 ζ 信号传导结构域的至少一个片段(任选地整个CD3 ζ 信号传导结构域)是由内源CD247基因座的开放阅读框或其部分序列编码。在一些实施方案中,CD3 ζ 信号传导结构域是由至少包含内源CD247基因座的开放阅读框的外显子3-8的核苷酸序列编码。在一些实施方案中,CD3 ζ 信号传导结构域是由包含内源CD247基因座的开放阅读框的外显子2和外显子3-8的至少一部分的核苷酸序列编码。在一些实施方案中,CD3 ζ 信号传导结构域是由不包含内源CD247基因座的开放阅读框的全长外显子1的核苷酸序列编码。在一些实施方案中,CD3 ζ 信号传导结构域是由不包含内源CD247基因座的开放阅读框的外显子1和/或不包含全长外显子2的核苷酸序列编码。

[0649] 在一些实施方案中,编码嵌合受体的一部分的转基因序列按顺序包含:编码细胞外结合结构域、任选地scFv的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;细胞内区域,所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域;和/或在转基因序列的整合后,修饰的CD247基因座按顺序包含:编码细胞外结合结构域、任选地scFv的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;细胞内区域,所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域;以及CD3 ζ 信号传导结构域。在一些实施方案中,嵌合受体的所编码的细胞内区域从其N末端至C末端按顺序包含:一个或多个共刺激信号传导结构域和CD3 ζ 链或其片段。

B. 所编码的嵌合受体

[0650] 在一些实施方案中,由本文提供的工程化细胞或根据本文提供的方法生成的工程化细胞编码的嵌合受体包括含有细胞内区域的嵌合受体,所述细胞内区域包含CD3 ζ 链或其片段或部分(例如,CD3 ζ 链的细胞内区域,如CD3 ζ 信号传导结构域)。在一些方面,嵌合受体的至少一部分是由存在于本文提供的多核苷酸(如上文章节I.B.2中所述的任何模板多核苷酸)中的转基因序列编码。在一些方面,含于多核苷酸中的编码嵌合受体的一部分的转基因序列被整合于工程化细胞的内源CD247基因座处,以得到编码嵌合受体(如本文所述的任何嵌合受体)的修饰的CD247基因座。在一些方面,所编码的嵌合受体含有CD3 ζ 链或其片段或部分,如CD3 ζ 信号传导结构域,例如整个或全部CD3 ζ 信号传导结构域。在一些实施方案中,CD3 ζ 链的至少一部分(如CD3 ζ 信号传导结构域的一部分或整个CD3 ζ 信号传导结构域)是由内源CD247基因座的开放阅读框序列或其部分序列编码。在一些实施方案中,工程化细胞还可以表达一种或多种另外的分子,例如,标记、另外的嵌合受体多肽、抗体或其抗原结合片段、免疫调节分子、配体、细胞因子或趋化因子。在一些方面,含于多核苷酸中的编码嵌合受体或其部分的转基因序列被整合于工程化细胞的内源CD247基因座处,以得到编码嵌合受体或其部分(如本文所述的任何嵌合受体,包括多链嵌合受体的一条或多条多肽链)的修

饰的CD247基因座。

[0651] 在一些实施方案中,提供工程化细胞,如免疫细胞,如T细胞,所述细胞表达一种或多种嵌合受体。嵌合受体包括嵌合受体或抗原受体以及含有其一种或多种组分的受体。嵌合受体可以包括嵌合受体(如含有配体结合结构域或其结合片段和细胞内信号传导结构域或区域的那些)、功能性非TCR抗原受体、嵌合抗原受体(CAR)、嵌合自身抗体受体(CAAR)以及前述任一种的一个或多个区域、一条或多条链、一个或多个结构域或者一种或多种组分。

[0652] 本文提供的基因工程化的细胞重编码的嵌合受体(如CAR)通常含有多个区域或结构域,如细胞外区域(例如,含有一个或多个细胞外结合结构域和/或间隔子)、跨膜结构域和/或细胞内区域(例如,含有CD3zeta(CD3 ζ)链或其片段和/或一个或多个共刺激信号传导结构域)中的一个或多个。在一些方面,所编码的嵌合受体还含有其他结构域(如多聚化结构域)、接头和/或调节元件。在一些实施方案中,所编码的嵌合受体从其N末端至C末端按顺序包含:细胞外结合结构域、间隔子、跨膜结构域和细胞内区域。

[0653] 在一些实施方案中,从工程化细胞表达的示例性嵌合受体包含两个或更多个受体多肽,在一些情形中,所述受体多肽含有不同的组分、结构域或区域。在一些方面,嵌合受体是多链受体或多链嵌合受体系统,其包含两个或更多个多肽,所述多肽一起构成功能性嵌合受体,或者一个或多个所述多肽可以调节、修饰或控制另一个受体多肽的表达、活性或功能。在一些方面,嵌合受体是双链受体,其包含两个多肽,所述多肽一起构成功能性嵌合受体。在一些方面,多链受体可以允许在空间或时间上调节或控制所述受体的特异性、活性、抗原(或配体)结合、功能和/或表达。在一些此类实施方案中,由修饰的CD247基因座处的核酸序列编码的嵌合受体多肽链是嵌合受体多肽链,其包括包含CD3zeta(CD3 ζ)链或其片段的细胞内区域。

[0654] 在一些实施方案中,在本文提供的基因工程化的细胞中编码的嵌合受体含有跨膜结构域或膜结合结构域。在一些方面,嵌合受体还含有细胞外区域。在一些方面,嵌合受体还含有细胞内区域。在一些实施方案中,本文提供的基因工程化的细胞重编码的嵌合受体通常含有多个区域或结构域,如细胞外区域(例如,含有一个或多个细胞外结合结构域和/或间隔子)、跨膜结构域和细胞内区域(例如,含有细胞内信号传导区域和/或一个或多个共刺激信号传导结构域)中的一个或多个。在一些情形中,间隔子隔开细胞外区域(例如细胞外结合结构域)与跨膜结构域或定位于所述二者之间。在一些方面,所编码的嵌合受体还含有其他结构域(如多聚化结构域)、接头和/或调节元件。

[0655] 在一些实施方案中,从工程化细胞表达的示例性嵌合受体包括含有两个或更多个受体多肽的多链受体,在一些情形中,所述受体多肽含有不同的组分、结构域或区域。在一些方面,嵌合受体含有两个或更多个多肽,所述多肽一起构成功能性嵌合受体。在一些方面,多链受体是双链受体,其包含两个多肽,所述多肽一起构成功能性嵌合受体。在一些实施方案中,嵌合受体是多链受体,其中一个或多个多肽调节、修饰或控制另一个受体多肽的表达、活性或功能。在一些方面,多链受体允许在空间或时间上调节或控制所述受体的特异性、活性、抗原(或配体)结合、功能和/或表达。

[0656] 在一些实施方案中,所编码的嵌合受体是嵌合受体,如CAR。示例性编码的CAR序列包含:细胞外结合结构域、间隔子、跨膜结构域和细胞内区域,所述细胞内区域包含初级信号传导结构域或区域和一个或多个共刺激信号传导结构域。在一些实施方案中,示例性编

码的CAR序列包含:细胞外结合结构域、间隔子、跨膜结构域和一个或多个共刺激信号传导结构域和初级信号传导结构域或区域。

[0657] 在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:跨膜结构域(或膜结合结构域)和包含CD3 ζ 链或其片段的细胞内区域,其中编码所述嵌合受体的核酸序列存在于修饰的CD247基因座中。在一些实施方案中,细胞内区域的至少一部分包含由内源CD247基因座的开放阅读框或其部分序列编码的CD3zeta (CD3 ζ) 链或其片段。在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:细胞外区域、跨膜结构域和包含CD3 ζ 链或其片段的细胞内区域。

[0658] 在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:信号肽、细胞外结合结构域、间隔子、跨膜结构域和包含CD3 ζ 链或其片段的细胞内区域,其中编码所述嵌合受体的核酸序列存在于修饰的CD247基因座中。在一些实施方案中,细胞内区域的至少一部分包含由内源CD247基因座的开放阅读框或其部分序列编码的CD3zeta (CD3 ζ) 链或其片段。在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:信号肽、细胞外结合结构域、间隔子、跨膜结构域和共刺激信号传导结构域和CD3 ζ 链或其片段。在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:信号肽、细胞外结合结构域、间隔子、跨膜结构域和两个共刺激信号传导结构域和CD3 ζ 链或其片段。在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:信号肽、细胞外结合结构域、间隔子、跨膜结构域和三个共刺激信号传导结构域和CD3 ζ 链或其片段。

[0659] 在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:跨膜结构域(或膜结合结构域)、细胞内多聚化结构域、任选地一个或多个共刺激信号传导结构域和CD3 ζ 链或其片段,其中编码所述嵌合受体的核酸序列存在于修饰的CD247基因座中。在一些实施方案中,细胞内区域的至少一部分包含由内源CD247基因座的开放阅读框或其部分序列编码的CD3zeta (CD3 ζ) 链或其片段。在一些实施方案中,示例性编码的嵌合受体按其N末端至C末端的顺序包含:细胞外多聚化结构域、跨膜结构域、任选地一个或多个共刺激信号传导结构域和CD3 ζ 链或其片段。

[0660] 在一些实施方案中,示例性编码的CAR序列按顺序包含:编码细胞外结合结构域、任选地scFv的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;细胞内区域,所述细胞内区域包含任选地来自人4-1BB的共刺激信号传导结构域;以及初级信号传导结构域或区域,如CD3 ζ 链的细胞内信号传导区域。在一些实施方案中,嵌合受体的所编码的细胞内区域从其N末端至C末端按顺序包含:一个或多个共刺激信号传导结构域和初级信号传导结构域或区域(如含有CD3 ζ 链或其片段)。

1. 嵌合抗原受体 (CAR)

[0661] 在一些实施方案中,由修饰的CD247基因座编码的嵌合受体是嵌合抗原受体(CAR)。在一些实施方案中,工程化细胞(如T细胞)表达对特定抗原(或标记或配体)(如在特定细胞类型的表面上表达的抗原)具有特异性的嵌合受体(如CAR)。在一些方面,本文所述的任何CAR(包括多链或可调节CAR)的至少一部分编码于转基因序列中。在一些方面,编码本文所述的CAR或其部分的转基因序列可以是章节I.B.2中所述的任一种。在一些方面,在

经由HDR整合转基因序列后,所得修饰的CD247基因座含有编码CAR(如本文所述的任何CAR,包括多链或可调节CAR)的核酸序列。

[0662] 在一些实施方案中,由修饰的CD247基因座编码的嵌合受体(例如,CAR)含有细胞外区域(例如,含有一个或多个细胞外结合结构域和/或间隔子)、跨膜结构域和/或细胞内区域(例如,含有初级信号传导区域或结构域和/或一个或多个共刺激信号传导结构域)中的一个或多个。在一些方面,所编码的嵌合受体还含有其他结构域,如多聚化结构域。在一些方面,修饰的CD247基因座含有编码接头和/或调节元件的序列。在一些实施方案中,所编码的嵌合受体从其N末端至C末端按顺序包含:细胞外结合结构域、跨膜结构域和细胞内区域(例如,包含初级信号传导区域或结构域或其部分和/或共刺激信号传导结构域)。在一些实施方案中,所编码的嵌合受体从其N末端至C末端按顺序包含:细胞外结合结构域、间隔子、跨膜结构域和细胞内区域(例如,包含初级信号传导区域或结构域或其部分和/或共刺激信号传导结构域)。

[0663] 在一些实施方案中,由修饰的CD247基因座编码的嵌合受体(例如,CAR)含有细胞外区域(例如,含有一个或多个细胞外结合结构域和/或间隔子)、跨膜结构域和/或细胞内区域(例如,含有CD3zeta (CD3 ζ)链或其片段(如CD3 ζ 链的细胞内区域)和/或一个或多个共刺激信号传导结构域)中的一个或多个。在一些实施方案中,嵌合受体在嵌合受体的C末端包括包含CD3zeta (CD3 ζ)链或其片段的细胞内区域。在一些方面,所编码的嵌合受体还含有其他结构域,如多聚化结构域。在一些方面,修饰的CD247基因座含有编码接头和/或调节元件的序列。在一些实施方案中,所编码的嵌合受体从其N末端至C末端按顺序包含:细胞外结合结构域、间隔子、跨膜结构域和细胞内区域(例如,包含CD3zeta (CD3 ζ)链或其片段)。

a. 结合结构域

[0664] 在一些实施方案中,所编码的嵌合受体的细胞外区域包含结合结构域。在一些实施方案中,结合结构域是细胞外结合结构域。在一些实施方案中,结合结构域是或包含多肽、配体、受体、配体结合结构域、受体结合结构域、抗原、表位、抗体、抗原结合结构域、表位结合结构域、抗体结合结构域、标签结合结构域或前述任一种的片段。

在一些实施方案中,结合结构域是配体结合结构域或抗原结合结构域。

[0665] 在一些方面,细胞外结合结构域(如一个或多个配体(例如,抗原)结合区或结构域)和一个或多个细胞内区域或结构域是经由一个或多个接头和/或跨膜结构域来连接或相连。在一些实施方案中,嵌合抗原受体包括布置在细胞外区域与细胞内区域之间的跨膜结构域。

[0666] 在一些实施方案中,抗原(例如,结合嵌合受体的结合结构域的抗原)是多肽。在一些实施方案中,抗原是碳水化合物或其他分子。在一些实施方案中,与正常或非靶向细胞或组织(例如,在健康细胞或组织中)相比,抗原在疾病、障碍或病症的细胞(例如肿瘤细胞或致病细胞)上选择性表达或过表达。在一些实施方案中,所述疾病、障碍或病症是感染性疾病或障碍、自身免疫疾病、炎性疾病或者肿瘤或癌症。在一些实施方案中,抗原在正常细胞上表达和/或在工程化细胞上表达。在一些方面,嵌合受体(例如,CAR)包括选自细胞外配体(例如,抗原)结合区域或结构域(例如,本文所述的任何抗体或片段)和细胞内区域的一个或多个区域或结构域。在一些实施方案中,配体(例如,抗原)结合区或结构域是或包括scFv或单一结构域V_H抗体,并且细胞内区域包含含有CD3-zeta (CD3 ζ)链或其片段的细胞内信号

传导区域或结构域。

[0667] 示例性编码的嵌合受体(包括CAR)包括例如以下文献中所述的那些:国际专利申请公开号WO 2000/14257、WO 2013/126726、WO 2012/129514、WO 2014/031687、WO 2013/166321、WO 2013/071154、WO 2013/123061;美国专利申请公开号US 2002131960、US 2013287748、US 20130149337;美国专利号6,451,995、7,446,190、8,252,592、8,339,645、8,398,282、7,446,179、6,410,319、7,070,995、7,265,209、7,354,762、7,446,191、8,324,353和8,479,118;以及欧洲专利申请号EP 2537416,和/或由以下文献所述的那些:Sadelain等人,Cancer Discov.2013年4月;3(4):388-398;Davila等人(2013)PLoS ONE 8(4):e61338;Turtle等人,Curr.Opin.Immunol.,2012年10月;24(5):633-39;以及Wu等人,Cancer,2012年3月18(2):160-75。在一些方面,抗原受体包括如美国专利号7,446,190中所述的CAR,以及国际专利申请公开号WO 2014/055668中所述的那些。CAR的例子包括如任何前述参考文献中披露的CAR,所述参考文献如WO 2014/031687、US 8,339,645、US 7,446,179、US 2013/0149337、US 7,446,190、US 8,389,282;Kochenderfer等人,2013,Nature Reviews Clinical Oncology,10,267-276(2013);Wang等人(2012)J.Immunother.35(9):689-701;和Brentjens等人,Sci Transl Med.2013 5(177)。

[0668] 在一些实施方案中,所编码的嵌合受体(例如,抗原受体)含有与抗原、配体和/或标记结合(例如,特异性结合)的细胞外结合结构域,如抗原结合结构域或配体结合结构域。抗原受体包括功能性非TCR抗原受体,如嵌合抗原受体(CAR)。在一些实施方案中,抗原受体是含有与抗原特异性结合的细胞外抗原识别结构域的CAR。在一些实施方案中,CAR被构建为具有对特定抗原、标记或配体的特异性,所述特定抗原、标记或配体是例如在由过继疗法靶向的特定细胞类型中表达的抗原(例如癌症标记)和/或旨在诱导衰减反应的抗原(例如在正常或未患病细胞类型上表达的抗原)。因此,CAR通常在其细胞外部分中包括一种或多种配体(例如,抗原)结合分子,例如一种或多种抗原结合片段、结构域或部分,或一种或多种抗体可变结构域,和/或抗体分子。在一些实施方案中,CAR包括抗体分子的一个或多个抗原结合部分,如源自单克隆抗体(mAb)的可变重链(V_H)和可变轻链(V_L)的单链抗体片段(scFv)、或单结构域抗体(sdAb)(如sdFv、纳米抗体、 V_H 和 V_{NAR})。在一些实施方案中,抗原结合片段包含通过柔性接头接合的抗体可变区。

[0669] 在一些实施方案中,所编码的CAR含有抗体或抗原结合片段(例如scFv),其特异性识别在细胞表面上表达的抗原或配体(如完整抗原)。在一些实施方案中,抗原或配体是在细胞表面上表达的蛋白质。在一些实施方案中,抗原或配体是多肽。在一些实施方案中,其为碳水化合物或其他分子。在一些实施方案中,如与正常或非靶向细胞或组织相比,抗原或配体在疾病或病症的细胞(例如,肿瘤或致病细胞)上选择性表达或过表达。在其他实施方案中,抗原在正常细胞上表达和/或在工程化细胞上表达。

[0670] 在一些实施方案中,嵌合受体靶向的抗原包括在通过过继细胞疗法靶向的疾病、病症或细胞类型的情况下表达的抗原。疾病和病症包括增殖性、肿瘤性和恶性疾病和障碍,包括癌症和肿瘤,包括血液恶性肿瘤、免疫系统癌症,如淋巴瘤、白血病和/或骨髓瘤,如B型白血病、T型白血病和髓系白血病、淋巴瘤和多发性骨髓瘤。

[0671] 在一些实施方案中,抗原或配体是肿瘤抗原或癌症标记。在一些实施方案中,与疾病或障碍相关的抗原是或包括 $\alpha v\beta 6$ 整合素($\alpha v\beta 6$ 整合素)、B细胞成熟抗原(BCMA)、B7-H3、

B7-H6、碳酸酐酶9 (CA9, 也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B (CTAG, 也称为NY-ESO-1和LAGE-2)、癌胚抗原 (CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1 (CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4 (CSPG4)、表皮生长因子蛋白 (EGFR)、表皮生长因子受体III型突变体 (EGFR vIII)、上皮糖蛋白2 (EPG-2)、上皮糖蛋白40 (EPG-40)、肝配蛋白B2、肝配蛋白受体A2 (EPHa2)、雌激素受体、Fc受体样蛋白5 (FCRL5; 也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体 (胎儿AChR)、叶酸结合蛋白 (FBP)、叶酸受体 α 、神经节苷脂GD2、0-乙酰化GD2 (OGD2)、神经节苷脂GD3、糖蛋白100 (gp100)、磷脂酰肌醇蛋白聚糖-3 (GPC3)、G蛋白偶联受体C类5族成员D (GPRC5D)、Her2/neu (受体酪氨酸激酶erb-B2)、Her3 (erb-B3)、Her4 (erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原 (HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1 (HLA-A1)、人白细胞抗原A2 (HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2 (IL-13R α 2)、激酶插入结构域受体 (kdr)、 κ 轻链、L1细胞粘附分子 (L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A (LRRC8A)、Lewis Y、黑色素瘤相关抗原 (MAGE)-A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素 (MSLN)、c-Met、鼠类巨细胞病毒 (CMV)、粘蛋白1 (MUC1)、MUC16、自然杀伤细胞2族成员D (NKG2D) 配体、黑色素A (MART-1)、神经细胞粘附分子 (NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原 (PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原 (PSCA)、前列腺特异性膜抗原 (PSMA)、受体酪氨酸激酶样孤儿受体1 (ROR1)、存活蛋白、滋养层糖蛋白 (TPBG, 也称为5T4)、肿瘤相关糖蛋白72 (TAG72)、酪氨酸酶相关蛋白1 (TRP1, 也称为TYRP1或gp75)、酪氨酸酶相关蛋白2 (TRP2, 也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体 (VEGFR)、血管内皮生长因子受体2 (VEGFR2)、肾母细胞瘤1 (WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。在一些实施方案中, 受体靶向的抗原包括与B细胞恶性肿瘤相关的抗原, 如许多已知B细胞标记中的任何一种。在一些实施方案中, 抗原是或包括CD20、CD19、CD22、ROR1、CD45、CD21、CD5、CD33、Ig κ 、Ig λ 、CD79a、CD79b或CD30。

[0672] 在一些实施方案中, 抗原是或包括病原体特有的或病原体表达的抗原。在一些实施方案中, 抗原是病毒抗原 (如来自HIV、HCV、HBV等的病毒抗原)、细菌抗原和/或寄生虫抗原。

[0673] 在一些实施方案中, 抗体或抗原结合片段 (例如scFv或V_H结构域) 特异性识别抗原 (如CD19)。在一些实施方案中, 抗体或抗原结合片段源自与CD19特异性结合的抗体或抗原结合片段或者是其变体。

[0674] 在一些实施方案中, 抗原是CD19。在一些实施方案中, scFv含有源自对CD19具有特异性的抗体或抗体片段的V_H和V_L。在一些实施方案中, 结合CD19的抗体或抗体片段是小鼠来源的抗体, 如FMC63和SJ25C1。在一些实施方案中, 抗体或抗体片段是人抗体, 例如如美国专利公开号US 2016/0152723中所述。

[0675] 在一些实施方案中, scFv源自FMC63。FMC63通常是指针对表达人源CD19的Nalm-1和-16细胞产生的小鼠单克隆IgG1抗体 (Ling, N.R., 等人 (1987). Leucocyte typing III. 302)。在一些实施方案中, FMC63抗体包含分别在SEQ ID NO: 38和39中所示的CDR-H1和CDR-H2, 和SEQ ID NO: 40或54中所示的CDR-H3; 以及SEQ ID NO: 35中所示的CDR-L1和SEQ

ID NO:36或55中所示的CDR-L2和SEQ ID NO:37或56中所示的CDR-L3。在一些实施方案中，FMC63抗体包含含有SEQ ID NO:41的氨基酸序列的重链可变区(V_H)和含有SEQ ID NO:42的氨基酸序列的轻链可变区(V_L)。

[0676] 在一些实施方案中，scFv包含含有SEQ ID NO:35的CDR-L1序列、SEQ ID NO:36的CDR-L2序列和SEQ ID NO:37的CDR-L3序列的可变轻链和/或含有SEQ ID NO:38的CDR-H1序列、SEQ ID NO:39的CDR-H2序列和SEQ ID NO:40的CDR-H3序列的可变重链。在一些实施方案中，scFv包含如SEQ ID NO:41所示的可变重链区和如SEQ ID NO:42所示的可变轻链区。在一些实施方案中，可变重链和可变轻链通过接头连接。在一些实施方案中，接头示于SEQ ID NO:58中。在一些实施方案中，scFv按顺序包含 V_H 、接头和 V_L 。在一些实施方案中，scFv按顺序包含 V_L 、接头和 V_H 。在一些实施方案中，scFv由SEQ ID NO:57所示的核苷酸序列或展现与SEQ ID NO:57至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的序列编码。在一些实施方案中，scFv包含SEQ ID NO:43所示的氨基酸序列或展现与SEQ ID NO:43至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的序列。

[0677] 在一些实施方案中，scFv源自SJ25C1。SJ25C1是针对人来源的表达CD19的NaIm-1和NaIm-16细胞产生的小鼠单克隆IgG1抗体(Ling, N.R.等人(1987). Leucocyte typing III. 302)。在一些实施方案中，SJ25C1抗体包含分别如SEQ ID NO:47-49中所示的CDR-H1、CDR-H2和CDR-H3序列，以及分别如SEQ ID NO:44-46中所示的CDR-L1、CDR-L2和CDR-L3序列。在一些实施方案中，SJ25C1抗体包含含有SEQ ID NO:50的氨基酸序列的重链可变区(V_H)和含有SEQ ID NO:51的氨基酸序列的轻链可变区(V_L)。

[0678] 在一些实施方案中，scFv包含含有SEQ ID NO:44的CDR-L1序列、SEQ ID NO:45的CDR-L2序列和SEQ ID NO:46的CDR-L3序列的可变轻链和/或含有SEQ ID NO:47的CDR-H1序列、SEQ ID NO:48的CDR-H2序列和SEQ ID NO:49的CDR-H3序列的可变重链。在一些实施方案中，scFv包含SEQ ID NO:50中所示的可变重链区和SEQ ID NO:51中所示的可变轻链区。在一些实施方案中，可变重链和可变轻链通过接头连接。在一些实施方案中，接头示于SEQ ID NO:52中。在一些实施方案中，scFv按顺序包含 V_H 、接头和 V_L 。在一些实施方案中，scFv按顺序包含 V_L 、接头和 V_H 。在一些实施方案中，scFv包含SEQ ID NO:53所示的氨基酸序列或展现与SEQ ID NO:53至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的序列。

[0679] 在一些实施方案中，抗原是CD20。在一些实施方案中，scFv含有源自对CD20具有特异性的抗体或抗体片段的 V_H 和 V_L 。在一些实施方案中，结合CD20的抗体或抗体片段是或源自利妥昔单抗的抗体，例如是利妥昔单抗scFv。

[0680] 在一些实施方案中，抗原是CD22。在一些实施方案中，scFv含有源自对CD22具有特异性的抗体或抗体片段的 V_H 和 V_L 。在一些实施方案中，结合CD22的抗体或抗体片段是或源自m971的抗体，例如是m971 scFv。

[0681] 在一些实施方案中，抗原是BCMA。在一些实施方案中，scFv含有源自对BCMA具特异性的抗体或抗体片段的 V_H 和 V_L 。在一些实施方案中，结合BCMA的抗体或抗体片段是或含有来自国际专利申请公开号WO 2016/090327和WO 2016/090320中阐述的抗体或抗体片段的 V_H 和 V_L 。

[0682] 在一些实施方案中,抗原是GPCR5D。在一些实施方案中,scFv含有源自对GPCR5D具特异性的抗体或抗体片段的 V_H 和 V_L 。在一些实施方案中,结合GPCR5D的抗体或抗体片段是或含有来自国际专利申请公开号WO 2016/090329和WO 2016/090312中阐述的抗体或抗体片段的 V_H 和 V_L 。

[0683] 在一些方面,所编码的CAR含有结合或识别(例如特异性结合)通用标签或通用表位的配体(例如,抗原)结合结构域。在一些方面,结合结构域可以结合分子、标签、多肽和/或表位,所述分子、标签、多肽和/或表位可以连接至识别与疾病或障碍相关的抗原的不同结合分子(例如,抗体或抗原结合片段)。示例性标签或表位包括染料(例如,异硫氰酸荧光素)或生物素。在一些方面,结合分子(例如,抗体或抗原结合片段)与标签连接,所述标签识别与具有表达对所述标签具有特异性的CAR的工程化细胞的疾病或障碍相关的抗原(例如,肿瘤抗原),以实现所述工程化细胞的细胞毒性或其他效应子功能。在一些方面,CAR对与疾病或障碍相关的抗原的特异性由带标签的结合分子(例如,抗体)提供,并且不同的带标签的结合分子可以用于靶向不同抗原。对通用标签或通用表位具有特异性示例性CAR包括例如在U.S.9,233,125,WO 2016/030414,Urbanska等人,(2012)Cancer Res 72:1844-1852,以及Tamada等人,(2012).Clin Cancer Res 18:6436-6445中所描述的那些。

[0684] 在一些实施方案中,所编码的CAR含有TCR样抗体,例如抗体或抗原结合片段(例如scFv),其特异性识别作为主要组织相容性复合物(MHC)-肽复合物存在于细胞表面上的细胞内抗原(例如肿瘤相关抗原)。在一些实施方案中,识别MHC-肽复合物的抗体或其抗原结合部分可以作为嵌合受体(如抗原受体)的部分在细胞上表达。抗原受体包括功能性非T细胞受体(TCR)抗原受体,如嵌合抗原受体(CAR)。在一些实施方案中,含有展现针对肽-MHC复合物的TCR样特异性的抗体或抗原结合片段的CAR也可以称为TCR样CAR。在一些实施方案中,CAR是TCR样CAR,并且抗原是被处理的肽抗原,如细胞内蛋白的肽抗原,所述抗原像TCR一样,在MHC分子的背景下在细胞表面上被识别。在一些实施方案中,在一些方面,对TCR样CAR的MHC-肽复合物具有特异性的细胞外抗原结合结构域经由接头和/或一个或多个跨膜结构域与一种或多种细胞内信号传导组分连接。在一些实施方案中,此类分子通常可以模拟或接近通过天然抗原受体(如TCR)的信号,并且任选地模拟或接近通过这种受体与共刺激受体的组合的信号。

[0685] 在一些实施方案中,主要组织相容性复合物(MHC)包括含有多态性肽结合位点或结合沟的蛋白质,通常是糖蛋白,在一些情形中,所述蛋白质可以与多肽的肽抗原(包括由细胞机构加工的肽抗原)复合。在一些情况下,MHC分子可以在细胞表面上展示或表达,包括作为与肽的复合物,即MHC-肽复合物,用于呈递具有T细胞上的抗原受体(如TCR或TCR样抗体)可识别的构象的抗原。通常,MHC I类分子是具有跨膜 α 链(在一些情况下具有三个 α 结构域)和非共价缔合的 β_2 微球蛋白的异二聚体。通常,MHC II类分子由两种跨膜糖蛋白 α 和 β 构成,两者通常都跨越膜。MHC分子可以包括MHC的有效部分,其含有用于结合肽的一个或多个抗原结合位点以及由适当抗原受体识别所需的序列。在一些实施方案中,MHC I类分子将源自胞质溶胶的肽递送至细胞表面,在此处MHC-肽复合物被T细胞(如通常是 $CD8^+$ T细胞,但是在一些情况下是 $CD4^+$ T细胞)识别。在一些实施方案中,MHC II类分子将源自囊泡系统的肽递送至细胞表面,其中所述肽通常被 $CD4^+$ T细胞识别。通常,MHC分子由一组连锁基因座编码,它们在小鼠中统称为H-2,并且在人中统称为人白细胞抗原(HLA)。因此,通常人MHC也可

以称为人白细胞抗原 (HLA)。

[0686] 术语“MHC-肽复合物”或“肽-MHC复合物”或其变体是指肽抗原与MHC分子的复合物或缔合物,例如通常通过所述肽在MHC分子的结合沟或裂缝中的非共价相互作用来形成。在一些实施方案中,MHC-肽复合物存在或展示于细胞表面上。在一些实施方案中,MHC-肽复合物可以由抗原受体(如TCR、TCR样CAR或其抗原结合部分)特异性识别。

[0687] 在一些实施方案中,多肽的肽(如肽抗原或表位)可以与MHC分子缔合,例如用于由抗原受体识别。通常,肽源自或基于更长生物分子(如多肽或蛋白质)的片段。在一些实施方案中,肽通常具有约8至约24个氨基酸的长度。在一些实施方案中,对于MHC II类复合物中的识别,肽的长度为从或从约9至22个氨基酸。在一些实施方案中,对于MHC I类复合物中的识别,肽的长度为从或从约8至13个氨基酸。在一些实施方案中,在识别MHC分子(如MHC-肽复合物)的背景中的肽后,抗原受体(如TCR或TCR样CAR)产生或触发激活信号至T细胞,从而诱导T细胞应答,如T细胞增殖、细胞因子产生、细胞毒性T细胞应答或其他应答。

[0688] 在一些实施方案中,TCR样抗体或抗原结合部分是已知的或可通过已知方法产生(参见例如,美国专利申请公开号US 2002/0150914;US 2003/0223994;US 2004/0191260;US 2006/0034850;US 2007/00992530;US 20090226474;US 20090304679;和国际申请公开号WO 03/068201)。

[0689] 在一些实施方案中,与MHC-肽复合物特异性结合的抗体或其抗原结合部分可以通过用有效量的含有特定MHC-肽复合物的免疫原对宿主进行免疫来产生。在一些情况下,MHC-肽复合物的肽是能够与MHC结合的抗原的表位,所述抗原如肿瘤抗原,例如通用肿瘤抗原、骨髓瘤抗原或如本文所述的其他抗原。在一些实施方案中,然后向宿主施用有效量的免疫原以用于引发免疫应答,其中免疫原保持其三维形式持续足以引发针对所述肽在MHC分子的结合沟中的三维呈递的免疫应答的时间段。然后测定从宿主中收集的血清以确定是否产生了识别MHC分子结合沟中的肽的三维呈递的所需抗体。在一些实施方案中,可以评估所产生的抗体以确认所述抗体可以区分MHC-肽复合物与单独的MHC分子、单独的目的肽以及MHC与无关肽的复合物。然后可以分离所需抗体。

[0690] 在一些实施方案中,与MHC-肽复合物特异性结合的抗体或其抗原结合部分可以通过采用抗体文库展示方法(如噬菌体抗体文库)来产生。在一些实施方案中,可以产生突变体Fab、scFv或其他抗体形式的噬菌体展示文库,例如,其中所述文库的成员在一个或多个CDR的一个或多个残基处发生突变。参见例如,美国专利申请公开号US 20020150914、US 20140294841;以及Cohen CJ.等人(2003) *J Mol. Recogn.* 16:324-332。

[0691] 本文中的术语“抗体”在最广泛的意义上使用,并且包括多克隆和单克隆抗体,包括完整抗体和功能性(抗原结合)抗体片段,包括片段抗原结合(Fab)片段、F(ab')₂片段、Fab'片段、Fv片段、重组IgG(rIgG)片段、能够特异性结合抗原的可变重链(V_H)区、单链抗体片段(包括单链可变片段(scFv))以及单结构域抗体(例如,sdAb、sdFv、纳米抗体、V_HH或V_{NAR})或片段。所述术语涵盖免疫球蛋白的基因工程化的和/或以其他方式修饰的形式,如胞内抗体、肽体、嵌合抗体、全人抗体、人源化抗体和异缀合抗体、多特异性(例如,双特异性)抗体、双抗体、三抗体和四抗体、串联二-scFv、串联三-scFv。除非另有说明,否则术语“抗体”应当理解为涵盖其功能性抗体片段。所述术语还涵盖完整或全长抗体,包括任何类别或亚类(包括IgG及其亚类、IgM、IgE、IgA和IgD)的抗体。在一些方面,CAR是双特异性CAR,例如

含有两个具有不同特异性的抗原结合结构域。

[0692] 在一些实施方案中,抗原结合蛋白、抗体及其抗原结合片段特异性识别全长抗体的抗原。在一些实施方案中,抗体的重链和轻链可以是全长的或者可以是抗原结合部分(Fab、F(ab')₂、Fv或单链Fv片段(scFv))。在其他实施方案中,抗体重链恒定区选自例如IgG1、IgG2、IgG3、IgG4、IgM、IgA1、IgA2、IgD和IgE,特别是选自例如IgG1、IgG2、IgG3和IgG4,更特别是IgG1(例如,人IgG1)。在一些实施方案中,抗体轻链恒定区选自例如κ或λ,特别是κ。

[0693] 所编码的嵌合受体的结合结构域是抗体片段。“抗体片段”是指不同于完整抗体的分子,其包含完整抗体的结合完整抗体所结合的抗原的一部分。抗体片段的例子包括但不限于Fv、Fab、Fab'、Fab'-SH、F(ab')₂;双抗体;线性抗体;可变重链(V_H)区、单链抗体分子(如scFv)和单结构域V_H单个抗体;和由抗体片段形成的多特异性抗体。在特定实施方案中,抗体是包含可变重链区和/或可变轻链区的单链抗体片段,如scFv。

[0694] 术语“可变区”或“可变结构域”是指抗体重链或轻链中参与抗体与抗原的结合的结构域。天然抗体的重链和轻链的可变结构域(分别为V_H和V_L)通常具有相似的结构,每个结构域包含四个保守的框架区(FR)和三个CDR。(参见例如,Kindt等人Kuby Immunology,第6版,W.H.Freeman and Co.,第91页(2007)。)单个V_H或V_L结构域可能足以赋予抗原结合特异性。此外,可以使用来自结合抗原的抗体的V_H或V_L结构域分离结合所述特定抗原的抗体,以分别筛选互补的V_L或V_H结构域的文库。参见例如,Portolano等人,J.Immunol.150:880-887(1993);Clarkson等人,Nature 352:624-628(1991)。

[0695] 单结构域抗体(sdAb)是包含抗体的重链可变结构域的全部或一部分或轻链可变结构域的全部或一部分的抗体片段。在某些实施方案中,单结构域抗体是人单结构域抗体。在一些实施方案中,CAR包含特异性地结合抗原的抗体重链结构域,所述抗原例如癌症标记或要靶向的细胞或疾病(例如肿瘤细胞或癌细胞)的细胞表面抗原,例如本文所述或已知的任何靶抗原。示例性的单结构域抗体包括sdFv、纳米抗体、V_H或V_{NAR}。

[0696] 抗体片段可以通过各种技术制备,包括但不限于完整抗体的蛋白水解消化以及通过重组宿主细胞产生。在一些实施方案中,抗体是重组产生的片段,如包含天然不存在的排列的片段(如具有通过合成接头(例如,肽接头)连接的两个或更多个抗体区域或链的那些),和/或可能并非通过天然存在的完整抗体的酶消化产生的片段。在一些实施方案中,抗体片段是scFv。

[0697] “人源化”抗体是如下抗体,其中所有或基本上所有CDR氨基酸残基都衍生自非人CDR并且所有或基本上所有FR氨基酸残基都衍生自人FR。人源化抗体任选地可以包括源自人抗体的抗体恒定区的至少一部分。非人抗体的“人源化形式”是指非人抗体的变体,其已经经历人源化以通常降低对人的免疫原性,同时保留亲代非人抗体的特异性和亲和力。在一些实施方案中,人源化抗体中的一些FR残基被来自非人抗体(例如,CDR残基所来源的抗体)的相应残基取代,例如以恢复或改善抗体特异性或亲和力。

[0698] 因此,在一些实施方案中,所编码的嵌合抗原受体(包括TCR样CAR)包括含有抗体或抗体片段的细胞外部分。在一些实施方案中,抗体或片段包括scFv。在一些方面,抗体或抗原结合片段可以通过筛选多个抗原结合片段或分子(例如其文库)来获得,例如通过针对与特定抗原或配体的结合筛选scFv文库来获得。

[0699] 在一些实施方案中,所编码的CAR是多特异性CAR,例如,含有可以结合和/或识别(例如,特异性结合)多种不同抗原的多个配体(例如,抗原)结合结构域。在一些方面,所编码的CAR是例如如通过含有具有不同特异性的两个抗原结合结构域而靶向两种抗原的双特异性CAR。在一些实施方案中,CAR含有双特异性结合结构域,例如,双特异性抗体或其片段,其含有结合至靶细胞上的不同表面抗原(例如,选自如本文所述任何列示抗原,例如,CD19和CD22或CD19和CD20)的至少一个抗原结合结构域。在一些实施方案中,双特异性结合结构域与其每一表位或抗原的结合可以导致刺激T细胞的功能、活性和/或应答,例如,细胞毒性活性和随后对靶细胞的裂解。这种示例性双特异性结合结构域可以包括:在一些情况下经由例如柔性接头彼此融合的串联scFv分子;双抗体及其衍生物,包括串联双抗体(Holliger等人,Prot Eng 9,299-305(1996);Kipriyanov等人,J Mol Biol 293,41-66(1999));双重亲和力重靶向(DART)分子,其可以包括具有C末端二硫桥的双抗体形式;双特异性T细胞接合器(BiTE)分子,其含有通过柔性接头融合的串联scFv分子(参见例如,Nagorsen和Bauerle,Exp Cell Res 317,1255-1260(2011));或者三功能抗体(triomab),其包括完整杂合小鼠/大鼠IgG分子(Seimetz等人,Cancer Treat Rev 36,458-467(2010))。在本文所述的任何CAR中可以含有任何此类结合结构域。

b. 间隔子和跨膜结构域

[0700] 在一些方面,所编码的嵌合受体(例如,嵌合抗原受体)包括细胞外部分,其含有一个或多个配体(例如,抗原)结合结构域(如抗体或其片段);和一个或多个细胞内信号传导区或结构域(也可互换地称为胞质信号传导结构域或区域)。在一些方面,嵌合受体(例如CAR)还包括间隔子和/或跨膜结构域或部分。在一些方面,间隔子和/或跨膜结构域可以连接含有配体(例如,抗原)结合结构域的细胞外部分和一个或多个细胞内信号传导区域或结构域。

[0701] 在一些实施方案中,所编码的嵌合受体(例如CAR)还包括间隔子,所述间隔子可以是或包括免疫球蛋白恒定区的至少一部分或其变体或经修饰形式,例如铰链区(例如,IgG4铰链区)和/或C_H1/C_L:和/或Fc区。在一些实施方案中,嵌合受体还包含间隔子和/或铰链区。在一些实施方案中,恒定区或部分是人IgG(如IgG4、IgG2或IgG1)的。在一些方面,恒定区的所述部分用作抗原识别组分(例如,scFv)与跨膜结构域之间的间隔子区。如与不存在间隔子的情况下相比,间隔子的长度可以提供抗原结合后增强的细胞反应性。在一些例子中,间隔子具有为或约12个氨基酸的长度或者具有不超过12个氨基酸的长度。示例性间隔子包括具有以下的那些:至少约10至229个氨基酸、约10至200个氨基酸、约10至175个氨基酸、约10至150个氨基酸、约10至125个氨基酸、约10至100个氨基酸、约10至75个氨基酸、约10至50个氨基酸、约10至40个氨基酸、约10至30个氨基酸、约10至20个氨基酸或约10至15个氨基酸,并且包括任何列出的范围的端点之间的任何整数。在一些实施方案中,间隔子区具有约12个或更少的氨基酸、约119个或更少的氨基酸或约229个或更少的氨基酸。在一些实施方案中,间隔子具有小于250个氨基酸的长度、小于200个氨基酸的长度、小于150个氨基酸的长度、小于100个氨基酸的长度、小于75个氨基酸的长度、小于50个氨基酸的长度、小于25个氨基酸的长度、小于20个氨基酸的长度、小于15个氨基酸的长度、小于12个氨基酸的长度或小于10个氨基酸的长度。在一些实施方案中,间隔子具有从或从约10至250个氨基酸的长度、10至150个氨基酸的长度、10至100个氨基酸的长度、10至50个氨基酸的长度、10至25个氨基

酸的长度、10至15个氨基酸的长度、15至250个氨基酸的长度、15至150个氨基酸的长度、15至100个氨基酸的长度、15至50个氨基酸的长度、15至25个氨基酸的长度、25至250个氨基酸的长度、25至100个氨基酸的长度、25至50个氨基酸的长度、50至250个氨基酸的长度、50至150个氨基酸的长度、50至100个氨基酸的长度、100至250个氨基酸的长度、100至150个氨基酸的长度或150至250个氨基酸的长度。示例性间隔子包括单独的IgG4铰链、与C_H2和C_H3结构域连接的IgG4铰链或与C_H3结构域连接的IgG4铰链。示例性间隔子包括但不限于以下文献中所述的那些:Hudecek等人.(2013) Clin.Cancer Res.,19:3153;Hudecek等人(2015) Cancer Immunol Res.3(2):125-135;或国际专利申请公开号W0 2014031687。

[0702] 在一些实施方案中,间隔子可以全部或部分地源自IgG4和/或IgG2。在一些实施方案中,间隔子可以是含有源自IgG4、IgG2和/或IgG2和IgG4的铰链、C_H2和/或C_H3序列中的一种或多种的嵌合多肽。在一些实施方案中,间隔子可以含有突变,如在一个或多个结构域中的一个或多个单氨基酸突变。在一些例子中,氨基酸修饰是在IgG4的铰链区中脯氨酸(P)对丝氨酸(S)的取代。在一些实施方案中,氨基酸修饰是用谷氨酰胺(Q)取代天冬酰胺(N)以降低糖基化异质性,如对应于SEQ ID NO:128所示的IgG4重链恒定区序列的C_H2区中的位置177的位置(Uniprot登录号P01861;对应于依据EU编号的位置297和SEQ ID NO:4所示的铰链-C_H2-C_H3间隔子序列的位置79的位置)处的N到Q取代,或者对应于SEQ ID NO:127所示的IgG2重链恒定区序列的C_H2区中的位置176的位置(Uniprot登录号P01859;对应于依据EU编号的位置297的位置)处的N到Q取代。

[0703] 在一些方面,间隔子仅含有IgG的铰链区,如IgG4、IgG2或IgG1的仅铰链,如SEQ ID NO:1中所示的仅铰链间隔子,并且是由SEQ ID NO:2中所示的序列编码。在其他实施方案中,间隔子是与C_H2和/或C_H3结构域连接的Ig铰链,例如IgG4铰链。在一些实施方案中,间隔子是与C_H2和C_H3结构域连接的Ig铰链,例如IgG4铰链,如SEQ ID NO:3所示。在一些实施方案中,间隔子是仅与C_H3结构域连接的Ig铰链,例如IgG4铰链,如SEQ ID NO:4所示。在一些实施方案中,间隔子是或包含富含甘氨酸-丝氨酸的序列或其他柔性接头,如已知的柔性接头。在一些实施方案中,恒定区或部分为IgD的。在一些实施方案中,间隔子具有SEQ ID NO:5所示的序列。在一些实施方案中,间隔子具有展现与SEQ ID NO:1、3、4和5中的任何一个至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。

[0704] 在一些方面,间隔子是多肽间隔子,如选自以下的一种或多种:(a)包含全部或部分免疫球蛋白铰链或其修饰形式或由其组成或包含约15个氨基酸或更少,并且不包含CD28细胞外区域或CD8细胞外区域,(b)包含全部或部分免疫球蛋白铰链(任选IgG4铰链)或其修饰形式或由其组成和/或包含约15个氨基酸或更少,并且不包含CD28细胞外区域或CD8细胞外区域,或(c)是为或约12个氨基酸的长度和/或包含全部或部分免疫球蛋白铰链(任选IgG4铰链)或其修饰形式或由其组成;或(d)由以下项组成或包含以下项:SEQ ID NO:1、3-5或27-34所示的氨基酸序列或与其具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的前述任一项的变体;或(e)包含式X₁PPX₂P(其中X₁是甘氨酸、半胱氨酸或精氨酸并且X₂是半胱氨酸或苏氨酸)或由其组成。

[0705] 示例性间隔子包括含有免疫球蛋白恒定区的一个或多个部分的那些,如含有Ig铰

链(如IgG铰链结构域)的那些。在一些方面,间隔子包括单独的IgG铰链、与C_H2和C_H3结构域中的一个或多个连接的IgG铰链或者与C_H3结构域连接的IgG铰链。在一些实施方案中,IgG铰链、C_H2和/或C_H3可以全部或部分源自IgG4或IgG2。在一些实施方案中,间隔子可以是含有源自IgG4、IgG2和/或IgG2和IgG4的铰链、C_H2和/或C_H3序列中的一种或多种的嵌合多肽。在一些实施方案中,铰链区包含IgG4铰链区和/或IgG2铰链区的全部或一部分,其中所述IgG4铰链区任选地是人IgG4铰链区,并且所述IgG2铰链区任选地是人IgG2铰链区;所述C_H2区包含IgG4 C_H2区和/或IgG2 C_H2区的全部或一部分,其中所述IgG4 C_H2区任选地是人IgG4 C_H2区,并且所述IgG2 C_H2区任选地是人IgG2 C_H2区;和/或所述C_H3区包含IgG4 C_H3区和/或IgG2 C_H3区的全部或一部分,其中所述IgG4 C_H3区任选地是人IgG4 C_H3区,并且所述IgG2 C_H3区任选地是人IgG2 C_H3区。在一些实施方案中,铰链C_H2和C_H3包含来自IgG4的铰链区C_H2和C_H3中的每一个的全部或一部分。在一些实施方案中,铰链区是嵌合的并且包含来自人IgG4和人IgG2的铰链区;所述C_H2区是嵌合的并且包含来自人IgG4和人IgG2的C_H2区;和/或所述C_H3区是嵌合的并且包含来自人IgG4和人IgG2的C_H3区。在一些实施方案中,间隔子包含IgG4/2嵌合铰链或含有与人IgG4铰链区相比的至少一个氨基酸替代的经修饰的IgG4铰链;人IgG2/4嵌合C_H2区;和人IgG4 C_H3区。

[0706] 在一些实施方案中,间隔子可以全部或部分源自IgG4和/或IgG2,并且可以含有突变,例如在一个或多个结构域中的一个或多个单氨基酸突变。在一些例子中,氨基酸修饰是在IgG4的铰链区中脯氨酸(P)对丝氨酸(S)的取代。在一些实施方案中,氨基酸修饰是用谷氨酰胺(Q)取代天冬酰胺(N)以降低糖基化异质性,如在SEQ ID NO:128中所示全长IgG4 Fc序列的C_H2区中的位置177处的N177Q突变,或者在SEQ ID NO:127中所示全长IgG2 Fc序列的C_H2区中的位置176处的N176Q。在一些实施方案中,间隔子是或包含IgG4/2嵌合铰链或修饰的IgG4铰链;IgG2/4嵌合C_H2区;以及IgG4 C_H3区,并且任选地具有约228个氨基酸的长度;或者SEQ ID NO:129中所示的间隔子。在一些实施方案中,CAR的配体(例如,抗原)结合或识别结构域与细胞内区域连接,例如,所述细胞内区域含有一种或多种细胞内信号传导组分,如细胞内信号传导区域或结构域和/或模拟经由抗原受体复合物(如TCR复合物)激活和/或经由另一细胞表面受体进行信号传导的信号传导组分。因此,在一些实施方案中,细胞外区域(例如,其含有结合结构域,如抗原结合组分(例如,抗体))与一个或多个跨膜和细胞内区域或结构域连接。在一些实施方案中,跨膜结构域与细胞外区域融合。在一些实施方案中,使用与受体(例如,CAR)中的结构域之一缔合的跨膜结构域。在一些情形中,通过氨基酸取代选择或修饰跨膜结构域以避免此类结构域与相同或不同表面膜蛋白的跨膜结构域结合,以最小化与受体复合物的其他成员的相互作用。

[0707] 在一些实施方案中,跨膜结构域源自天然或合成来源。在来源是天然的情况下,所述结构域在一些方面源自任何膜结合蛋白或跨膜蛋白。跨膜区域包括源自T细胞受体的 α 、 β 或 ζ 链、CD28、CD3 ϵ 、CD45、CD4、CD5、CD8、CD9、CD16、CD22、CD33、CD37、CD64、CD80、CD86、CD134、CD137(4-1BB)或CD154(即,至少包含它们的一个或多个跨膜区域)的跨膜区域。可替代地,在一些实施方案中,跨膜结构域是合成的。在一些方面,合成跨膜结构域主要包含疏水性残基,例如亮氨酸和缬氨酸。在一些方面,将在合成跨膜结构域的每个末端发现苯丙氨酸、色氨酸和缬氨酸的三联体。在一些实施方案中,所述连接是通过接头、间隔子和/或一个或多个跨膜结构域来实现。在一些方面,跨膜结构域含有CD28的跨膜部分或其变体。细胞外

区域和跨膜可以直接或间接连接。在一些实施方案中,细胞外区域和跨膜通过间隔子(如本文描述的任何间隔子)连接。

[0708] 在一些实施方案中,受体(例如CAR)的跨膜结构域是人CD28的跨膜结构域或其变体,例如人CD28(登录号:P10747.1)的27个氨基酸的跨膜结构域,或者是包含SEQ ID NO:8中所示氨基酸序列或展现与SEQ ID NO:8至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列的跨膜结构域;在一些实施方案中,含有嵌合受体的部分的跨膜结构域包含SEQ ID NO:9中所示的氨基酸序列或具有与SEQ ID NO:9的至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。

c. 细胞内区域

[0709] 在一些方面,在修饰的CD247基因座中编码的嵌合受体(例如,CAR)包括包含信号传导区域或结构域的细胞内区域(也称为胞质区域)。在一些实施方案中,细胞内区域包含细胞内信号传导区域或结构域。在一些实施方案中,细胞内信号传导区域或结构域是或包含初级信号传导区域、能够在T细胞中刺激和/或诱导初级激活信号的信号传导结构域、T细胞受体(TCR)组分的信号传导结构域(例如CD3-Zeta(CD3 ζ)链的细胞内信号传导结构域或区域或其功能变体或信号传导部分)和/或包含基于免疫受体酪氨酸的激活基序(ITAM)的信号传导结构域。

[0710] 在一些实施方案中,嵌合受体(例如,CAR)包括至少一种或多种细胞内信号传导组分,如细胞内信号传导区或结构域。细胞内信号传导区包括模拟或接近以下的那些:经由天然抗原受体的信号、经由这种受体与共刺激受体的组合的信号和/或仅经由共刺激受体的信号。在一些实施方案中,存在短的寡肽或多肽接头,例如长度在2与10个氨基酸之间的接头(如含有甘氨酸和丝氨酸的接头,例如甘氨酸-丝氨酸双联体),并且在CAR的跨膜结构域与胞质信号传导结构域之间形成连接。

[0711] 在一些实施方案中,在连接CAR后,所述CAR的胞质(或细胞内)结构域或区域(例如,细胞内信号传导区域)刺激和/或激活免疫细胞(例如,工程化以表达CAR的T细胞)的正常效应子功能或应答中的至少一种。例如,在一些情境下,CAR诱导T细胞的功能,如细胞溶解活性或T辅助活性,如细胞因子或其他因子的分泌。在一些实施方案中,使用抗原受体组分或共刺激分子的细胞内信号传导区域或结构域的截短部分代替完整的免疫刺激链,例如如果其转导效应子功能信号的话。在一些实施方案中,细胞内信号传导区域(例如,包含一个或多个细胞内结构域)包括T细胞受体(TCR)的胞质序列,并且在一些方面还包括共受体(其在天然背景中与这种受体协同起作用以在抗原受体接合后启动信号转导)和/或此类分子的任何衍生物或变体的那些,和/或具有相同功能能力的任何合成序列。在一些实施方案中,例如包含一个或多个细胞内结构域的细胞内信号传导区域包括参与提供共刺激信号的区域或结构域的胞质序列。

(i) 共刺激信号传导结构域

[0712] 在一些实施方案中,为了促进完全刺激和/或激活,在所编码的CAR中包括用于生成次级或共刺激信号的一种或多种组分。在其他实施方案中,所编码的CAR不包括用于生成共刺激信号的组分。在一些方面,另外的受体多肽或其部分在相同细胞中表达并且提供用于生成次级或共刺激信号的组分。

[0713] 在一些实施方案中,所编码的CAR包括共刺激受体(例如CD28、4-1BB、OX40(CD134)、CD27、DAP10、DAP12、ICOS和/或其他共刺激受体)的信号传导区域和/或跨膜部分。在一些方面,相同的CAR包括初级胞质信号传导区域和共刺激信号传导组分。

[0714] 在一些实施方案中,一种或多种不同的嵌合受体可以含有一个或多个不同的细胞内信号传导区域或结构域。在一些实施方案中,初级胞质信号传导区域被包括于一种所编码的CAR内,而共刺激组分由另一种受体(例如,识别另一种抗原的另一种CAR)提供。在一些实施方案中,所编码的CAR包括在同一细胞上表达的激活或刺激CAR和共刺激CAR(参见WO 2014/055668)。

[0715] 在某些实施方案中,细胞内信号传导区域包含与CD3(例如,CD3 ζ)细胞内区域或结构域连接的CD28跨膜和信号传导结构域。在一些实施方案中,细胞内区域包含与CD3 ζ 细胞内区域或结构域连接的嵌合CD28和CD137(4-1BB, TNFRSF9)共刺激结构域。

[0716] 在一些实施方案中,所编码的CAR在胞质部分中包含一个或多个(例如两个或更多个)共刺激结构域和初级胞质信号传导区域。示例性CAR包括CD3 ζ 、CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D和/或ICOS的细胞内组分,如一个或多个细胞内信号传导区或结构域。在一些实施方案中,嵌合抗原受体含有T细胞共刺激分子的细胞内信号传导区域或结构域,例如来自CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D和/或ICOS,在一些情况下,在跨膜结构域与细胞内信号传导区域或结构域之间。在一些方面,T细胞共刺激分子是CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D和/或ICOS中的一种或多种。在一些实施方案中,共刺激分子是人共刺激分子。

[0717] 在一些实施方案中,细胞内信号传导区域或结构域包含人CD28的细胞内共刺激信号传导结构域或其功能变体或部分,如其41个氨基酸的结构域,和/或在天然CD28蛋白的位置186-187处具有LL至GG取代的这种结构域。在一些实施方案中,细胞内信号传导区域和/或结构域可以包含SEQ ID NO:10或11所示的氨基酸序列,或展现与SEQ ID NO:10或11至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。在一些实施方案中,细胞内区域包含CD137(4-1BB)的细胞内共刺激信号传导结构域或区域或其功能变体或部分,如人4-1BB(登录号Q07011.1)的42个氨基酸的胞质结构域或其功能变体或部分,如SEQ ID NO:12所示的氨基酸序列或展现与SEQ ID NO:12至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。

[0718] 在一些情况下,所编码的CAR被称为第一代、第二代、第三代或第四代CAR。在一些方面,第一代CAR是在抗原结合后例如经由CD3链诱导的信号仅提供初级刺激或激活信号的CAR;在一些方面,第二代CAR是提供这种信号和共刺激信号的CAR,如包括来自一种或多种共刺激受体如CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D、ICOS和/或其他共刺激受体的一个或多个细胞内信号传导区或结构域的CAR;在一些方面,第三代CAR是包括不同共刺激受体(例如选自CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D、ICOS和/或其他共刺激受体)的多个共刺激结构域的CAR;在一些方面,第四代CAR是包括不同共刺激受体(例如选自CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D、ICOS和/或其他共刺激受体)的三个或更多个共刺激结构域的CAR。

(ii) CD3 ζ 链

[0719] 在一些实施方案中,所编码的嵌合受体包括TCR复合物的细胞内组分,如介导T细胞激活和细胞毒性的TCR CD3链,例如CD3 ζ 链。因此,在一些方面,抗原结合或抗原识别结构域连接至一个或多个细胞信号传导模块。在一些实施方案中,细胞信号传导模块包括CD3跨膜结构域、CD3细胞内信号传导结构域和/或其他CD跨膜结构域。在一些实施方案中,所编码的嵌合受体(例如,CAR)还包括一种或多种另外的分子(如Fc受体 γ (FcR γ)、CD8 α 、CD8 β 、CD4、CD25或CD16)的一部分。例如,在一些方面,CAR包括CD3 zeta (CD3 ζ)与CD8 α 、CD8 β 、CD4、CD25或CD16中的一种或多种之间的嵌合分子。

[0720] 在天然TCR的情况下,完全刺激通常不仅需要经由TCR进行信号传导,还需要共刺激信号。在一些方面,T细胞刺激可以由两个类别的胞质信号传导序列来介导:经由TCR起始抗原依赖性初级激活的那些(一个或多个初级胞质信号传导区域或结构域),以及以非抗原依赖性方式作用以提供次级或共刺激信号的那些(一个或多个次级胞质信号传导区域或结构域)。在一些方面,CAR包括此类信号传导组分中的一种或两种。

[0721] 在一些方面,所编码的CAR包括包含初级胞质信号传导区域的细胞内区域,所述初级胞质信号传导区域调节TCR复合物的初级刺激和/或激活。在一些实施方案中,所编码的嵌合受体包括包含CD3 ζ 信号传导结构域(如整个或全部CD3 ζ 信号传导结构域)的细胞内区域,例如,所述CD3 ζ 信号传导结构域能够在嵌合受体接合后进行信号传导或信号转导来介导T细胞激活和/或细胞毒性。在一些方面,所编码的嵌合受体的初级胞质信号传导区域的至少一个片段(例如,含有CD3 ζ 链或其片段,如CD3 ζ 的细胞内信号传导结构域,或任选地整个CD3 ζ 信号传导结构域)是由内源CD247基因座的开放阅读框或其部分序列编码。以刺激性方式作用的一个或多个初级胞质信号传导区域可以含有例如源自CD3 zeta (CD3 ζ)的信号传导基序(其被称为免疫受体酪氨酸激活基序或ITAM)。在一些实施方案中,CAR含有源自CD3 ζ 的胞质信号传导结构域、其片段或部分或序列。在一些实施方案中,细胞内(或胞质)信号传导区域包含人CD3 ζ 链或其片段或部分,包括CD3 ζ 的细胞内或胞质刺激性信号传导结构域或其功能变体,如人CD3 ζ (登录号:P20963.2)的亚型3的112个AA的胞质结构域或如美国专利号7,446,190或美国专利号8,911,993中所述的CD3 ζ 信号传导结构域。在一些实施方案中,所编码的嵌合受体的细胞内区域包含SEQ ID NO:13、14或15中所示的氨基酸序列或者展现与SEQ ID NO:13、14或15至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列,或其部分序列。在一些实施方案中,由转基因与CD247基因座的内源序列的基因融合物编码的示例性CD3 ζ 链或其片段包括CD3 ζ 链的ITAM结构域,例如,SEQ ID NO:73中所示的人CD3 ζ 链前体序列的氨基酸残基61-89、100-128或131-159,或者含有来自CD3 ζ 链的一个或多个ITAM结构域并且展现与SEQ ID NO:73至少或至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的氨基酸序列。

[0722] 在一些实施方案中,将细胞工程化以表达一种或多种另外的分子(例如,多肽,如另外的嵌合受体多肽或其部分),用于调整、控制或调节所编码的CAR的功能和/或活性。示例性多链嵌合受体(如多链CAR)描述于本文中,例如章节III.B.2中。

[0723] 在一些实施方案中,所编码的CAR含有抗体(例如,抗体片段)、跨膜结构域(其是或含有CD28的跨膜部分或其功能变体)以及含有CD28的信号传导部分或其功能变体和CD3 ζ 的信号传导区域或其功能变体的细胞内信号传导结构域。在一些实施方案中,CAR含有抗体

(例如,抗体片段)、跨膜结构域(其是或含有CD28的跨膜部分或其功能变体)以及细胞内信号传导结构域(含有4-1BB的信号传导部分或其功能变体和CD3 ζ 的信号传导部分或其功能变体)。在一些此类实施方案中,受体还包括含有Ig分子(如人Ig分子)的一部分(如Ig铰链,例如IgG4铰链)的间隔子,如仅铰链间隔子。在一些实施方案中,嵌合受体在所述受体的C末端包含CD3 zeta (CD3 ζ)。

2. 多链CAR

[0724] 在一些实施方案中,由修饰的CD247基因座的核酸序列编码的嵌合受体可以是多链CAR。在一些实施方案中,如果包含两条或更多条多肽链的多链CAR在细胞中表达,则至少一条所述多肽链包括CD3zeta (CD3 ζ)链或其片段或部分并且由修饰的CD247基因座编码。在一些方面,用于引入编码多链CAR的一条或多条链的核酸序列的多核苷酸可以包括本文在章节I.B中所述的任一种。在一些方面,多核苷酸(例如,模板多核苷酸)含有编码多链CAR或其部分的至少一条链(如含有CD3zeta (CD3 ζ)链或其片段或部分的多链CAR的至少一个多肽的至少一部分)的转基因序列。在一些方面,转基因序列还包括编码不同的或另外的多肽(例如,多链CAR的另一条链或另外的链)或另外的分子(如本文在章节I.B.2.(vi)中所述的那些)的序列。在一些方面,可以引入另外的多核苷酸(例如,另外的模板多核苷酸),其编码多链CAR的另外的组分。在一些方面,另外的多核苷酸可以是本文例如在章节I.B.2中所述的任何多核苷酸或其修饰的形式,如包含不同同源臂以供将用于整合的核酸靶向于不同基因组基因座的的多核苷酸。

[0725] 在一些实施方案中,所提供的工程化细胞包括表达多链受体(如多链CAR)的细胞。在一些实施方案中,在所述细胞上示例性多链CAR可以含有两种或更多种基因工程化的受体,它们可以一起构成功能性嵌合受体。在一些方面,组合中的各个多肽链可以执行CAR的功能或活性,和/或调整、控制或调节CAR的功能和/或活性。在一些方面,多链CAR可以含有两条或更多条多肽链,其各自识别相同或不同的抗原,并且通常各自包括不同的区域或结构域,如不同的细胞内信号传导组分。在一些方面,至少一种基因工程化的受体的细胞内信号传导组分包括CD3zeta (CD3 ζ)链或其片段或部分。在一些方面,修饰的CD247基因座可以包括编码多链受体的至少一条链(如含有CD3zeta (CD3 ζ)链或其片段或部分的受体多肽)的核酸序列。

[0726] 在一些实施方案中,嵌合受体是包含两条或更多条多肽链的多链CAR或双链CAR。在一些实施方案中,多链受体是可调节CAR、条件活性CAR或诱导型CAR。在一些方面,嵌合受体(如双链CAR)的两个或更多个多肽允许在空间或时间上调节或控制所述嵌合受体的特异性、活性、抗原(或配体)结合、功能和/或表达。在一些此类实施方案中,由修饰的CD247基因座处的核酸序列编码的嵌合受体可以包括双链或多链受体的一条或多条链。在一些方面,在双链CAR的仅一条由修饰的CD247基因座编码的情形中,另一条链可以由整合于不同基因组位置或作为附加体的单独核酸分子编码。

[0727] 在一些实施方案中,多链CAR可以包括激活CAR与共刺激CAR的组合。例如,在一些实施方案中,多链CAR可以包括编码靶向两种不同抗原的CAR的两种多肽,所述抗原在非靶细胞(例如,正常细胞)上单独存在,但是仅在要治疗的疾病或病症的细胞上一起存在。在一些实施方案中,多链CAR可以包括激活和抑制性CAR,例如如下的那些:其中激活CAR结合至在正常或未患病细胞和要治疗的疾病或病症的细胞两者上表达的一种抗原,并且抑制性

CAR结合至仅在正常细胞或不期望治疗的细胞上表达的另一种抗原。在一些方面,多链CAR可以包括编码CAR的一种或多种多肽,所述CAR能够被调整、调节或控制。

[0728] 在一些实施方案中,多链CAR包括一条或多条多肽链,其编码CAR的一个或多个结构域或区域。在一些方面,组合中的各个多肽链可以包含CAR。在一些实施方案中,在CAR中存在一个或多个另外的结构域或区域。在一些实施方案中,使用存在于多链CAR的一条或多条多肽链中的各个结构域或区域来调整、控制或调节所述CAR的功能和/或活性。在一些实施方案中,工程化细胞表达含有不同的组分、结构域或区域的两条或更多条多肽链。在一些方面,两条或更多条多肽链允许在空间或时间上调节或控制嵌合受体的特异性、活性、抗原(或配体)结合、功能和/或表达。在包括多于一个多肽(例如,2个或更多个多肽)的多链CAR的一些实施方案中,靶向编码至少一个多肽的核酸序列(如编码在细胞内区域中(例如,在受体的C末端中)含有CD3 zeta (CD3 ζ)的多肽链的核酸序列)以供整合于内源CD247基因座处。在一些实施方案中,可以将编码另外的分子或多肽(例如,多链CAR的另外的多肽链或另外的分子)的核酸序列靶向于相同基因座处,例如借助放置在用于靶向的相同多核苷酸上。在一些中,将编码另外的分子或多肽的核酸序列靶向于不同基因座处或者通过不同方法来递送。

[0729] 在一些方面,编码CAR的结构域或区域的一条或多条多肽链可以靶向一种或多种抗原示例性多链CAR或其他多重靶向策略包括例如以下文献中所述的那些:国际专利申请公开号W0 2014055668;或Fedorov等人,Sci.Transl.Medicine,Sci Transl Med. (2013) 5 (215):215ra172;Sadelain,Curr Opin Immunol. (2016) 41:68-76;Wang等人(2017) Front.Immunol.8:1934;Mirzaei等人(2017) Front.Immunol.8:1850;Marin-Acevedo等人(2018) Journal of Hematology&Oncology 11:8;Fesnak等人(2016) Nat Rev Cancer.16 (9):566-581;以及Abate-Daga和Davila, (2016) Molecular Therapy-Oncolytics 3, 16014。

[0730] 在一些实施方案中,工程化细胞可以表达嵌合受体(例如,CAR)的第一多肽链,其通常在与所述第一多肽链识别的抗原(例如,第一抗原)特异性结合后,能够诱导激活或刺激信号至细胞。在一些实施方案中,所述细胞还可以表达嵌合受体(例如,CAR,在一些情形中称为嵌合共刺激受体)的第二多肽链,其通常在与所述第二多肽链识别的第二抗原特异性结合后,能够诱导共刺激信号至免疫细胞。在一些实施方案中,第一抗原和第二抗原是相同的。在一些实施方案中,第一抗原和第二抗原是不同的。

[0731] 在一些实施方案中,第一和/或第二多肽链能够诱导激活或刺激信号至细胞。在一些实施方案中,受体包括含有ITAM或ITAM样基序的细胞内信号传导组分。在一些实施方案中,由第一多肽链诱导的激活涉及细胞中的信号转导或蛋白质表达的变化,导致启动免疫应答(如ITAM磷酸化)和/或启动ITAM介导的信号转导级联、形成免疫突触和/或所结合受体(例如,CD4或CD8等)附近的分子的聚簇、激活一种或多种转录因子(如NF- κ B和/或AP-1)、和/或诱导因子(如细胞因子)的基因表达、增殖和/或存活。在一些实施方案中,激活结构域包括于至少一种多链CAR(如由修饰的CD247基因座编码的多肽链)内,而共刺激组分是由识别另一种抗原的另一种多肽提供。在一些实施方案中,工程化细胞可以包括多链CAR,包括激活或刺激性CAR、共刺激CAR,二者在相同细胞上表达(参见W0 2014/055668)。在一些方面,细胞表达一种或多种刺激性或激活CAR(如由修饰的CD247基因座编码的那些,如本文例

如在章节III.A中所述)和/或共刺激CAR。

[0732] 在一些实施方案中,第一和/或第二多肽链包括共刺激受体(如CD28、CD137(4-1BB)、OX40(CD134)、CD27、DAP10、DAP12、NKG2D、ICOS和/或其他共刺激受体)的细胞内信号传导区域或结构域。在一些实施方案中,第一和第二多肽链可以含有不同共刺激受体的一个或多个细胞内信号传导结构域。在一个实施方案中,第一多肽链含有CD28共刺激信号传导结构域并且第二多肽链含有4-1BB共刺激信号传导区域,或反之亦然。

[0733] 在一些实施方案中,第一和/或第二多肽链包括含有ITAM或ITAM样基序的细胞内信号传导结构域(如来自CD3zeta(CD3 ζ)链或其片段或部分的那些,如CD3 ζ 细胞内信号传导结构域)和共刺激受体的细胞内信号传导结构域二者。在一些实施方案中,第一多肽链含有包含ITAM或ITAM样基序的细胞内信号传导结构域,并且第二多肽链含有共刺激受体的细胞内信号传导结构域。与同一细胞中诱导的激活或刺激信号组合的共刺激信号是导致免疫应答的信号,所述免疫应答如稳健且持续的免疫应答,如增加的基因表达、细胞因子和其他因子的分泌以及T细胞介导的效应子功能(如细胞杀伤)。

[0734] 在一些实施方案中,仅连接第一多肽链和仅连接第二多肽链都没有诱导稳健的免疫应答。在一些方面,如果仅连接一种受体,则细胞变得耐受抗原或对抗原无应答,或被抑制,和/或不被诱导增殖或分泌因子或执行效应子功能。然而,在一些此类实施方案中,在连接多条多肽链时,如在遇到表达第一和第二抗原的细胞时,实现所需应答,如完全免疫激活或刺激,例如通过一种或多种细胞因子的分泌、增殖、持久性和/或执行免疫效应子功能(如靶细胞的细胞毒性杀伤)所指示。

[0735] 在一些实施方案中,多链CAR的一条或多条链可以包括抑制性CAR(iCAR,参见Fedorov等人,Sci.Transl.Medicine,5(215)(2013)),如识别除了与疾病或病症相关和/或为疾病或病症所特有的抗原以外的抗原的CAR,借此经由疾病靶向CAR递送的激活信号由于所述抑制性CAR与其配体的结合而被减小或被抑制,例如,以减小脱靶效应。在一些实施方案中,抑制性CAR可以由与刺激或激活CAR相同的多核苷酸(例如,含有CD3zeta(CD3 ζ)链或其片段或部分,如整合于CD247基因座处的多核苷酸)或由不同的多核苷酸编码。

[0736] 在一些实施方案中,多链CAR的两条多肽链分别诱导激活和抑制信号至细胞,使得一条多肽链与其抗原的连接激活细胞或诱导应答,但是第二多肽链(例如,抑制性受体)与其抗原的连接诱导抑制或减弱该应答的信号。例子是激活CAR与抑制性CAR(iCAR)的组合。例如,这种策略可以用于例如降低在如下背景下的脱靶效应的可能性,在所述背景中激活CAR结合在疾病或病症中表达但也在正常细胞上表达的抗原,并且抑制性受体结合在正常细胞上表达但不在疾病或病症的细胞上表达的单抗原。

[0737] 在一些方面,在细胞中表达的另外的受体多肽还包括抑制性CAR(例如iCAR),并且包括减弱或抑制细胞中的免疫应答(如ITAM和/或共刺激促进的应答)的细胞内组分。示例性的此类细胞内信号传导组分是在免疫检查点分子上发现的那些,包括PD-1、CTLA4、LAG3、BTLA、OX2R、TIM-3、TIGIT、LAIR-1、PGE2受体、EP2/4腺苷受体(包括A2AR)。在一些方面,工程化细胞包括抑制性CAR,所述抑制性CAR包括这种抑制性分子的信号传导结构域或源自这种抑制性分子的信号传导结构域,使得其起到减弱细胞的应答(例如,由激活和/或共刺激性CAR诱导)的作用。

[0738] 在一些实施方案中,在与特定疾病或病症相关的抗原在未患病细胞上表达和/或

在工程化细胞本身上表达的情况下,可以采用多链CAR,所述表达是瞬时的(例如,在与基因工程化相关的刺激后)或永久的。在此类情形中,由于需要连接两个分开的且单独特异性的多肽,可以改进特异性、选择性和/或功效。

[0739] 在一些实施方案中,多种抗原(例如,第一和第二抗原)在所靶向的细胞、组织或者疾病或病症上(如在癌细胞上)表达。在一些方面,所述细胞、组织、疾病或病症是多发性骨髓瘤或多发性骨髓瘤细胞。在一些实施方案中,多种抗原中的一种或多种通常也在不需要用细胞疗法靶向的细胞(例如正常或未患病细胞或组织)和/或工程化细胞本身上表达。在此类实施方案中,因为需要连接多种受体以实现细胞的反应,实现特异性和/或功效。

[0740] 在一些实施方案中,第一和/或第二多肽链中的一条可以调节另一条多肽链的表达、抗原结合和/或活性。

[0741] 在一些方面,双多肽链系统可以用于调节至少一条多肽链的表达。在一些实施方案中,第一多肽链含有连接至经由可调节切割元件连接的调节分子(如转录因子)的第一配体(例如,抗原)结合结构域。在一些方面,可调节切割元件源自修饰的Notch受体(例如,synNotch),它能够在接合第一配体(例如,抗原)结合结构域后切割并释放细胞内结构域。在一些方面,第二多肽链含有连接至能够诱导激活或刺激信号至细胞的细胞内信号传导组分(如含有ITAM的细胞内信号传导结构域)的第二配体(例如,抗原)结合结构域。在一些方面,编码第二多肽链的核酸序列与转录调节元件(例如,启动子)可操作地连接,所述转录调节元件能够由特定转录因子(例如,由第一多肽链编码的转录因子)调节。在一些方面,配体或抗原与第一配体(例如,抗原)结合结构域的接合导致转录因子的蛋白水解释放,这又可以诱导第二多肽链的表达(参见Roybal等人(2016)Cell164:770-779;Morsut等人(2016)Cell 164:780-791)。在一些实施方案中,第一抗原和第二抗原是不同的。

[0742] 在一些情况下,嵌合受体(例如,CAR)能够被调节、控制、诱导或抑制,可以期望优化使用所述嵌合受体的疗法的安全性和功效。在一些实施方案中,多链CAR是可调节CAR。在一些方面,本文提供工程化细胞,其包含能够被调节的CAR。能够被调节的嵌合受体(本文中也称为“可调节嵌合受体”或“可调节CAR”)是指多个多肽,如一组至少两条多肽链,其在工程化细胞(例如,工程化T细胞)中表达时为所述工程化细胞提供在诱导物的控制下生成细胞内信号的能力。

[0743] 在一些实施方案中,可调节CAR的多肽含有能够与另一多聚化结构域发生多聚化的多聚化结构域。在一些实施方案中,多聚化结构域能够在与诱导物结合后发生多聚化。例如,多聚化结构域可以结合诱导物(如化学诱导物),其借助所述多聚化结构域的多聚化导致可调节CAR的多肽的多聚化,从而产生可调节CAR。

[0744] 在一些实施方案中,可调节CAR的一个多肽包含配体(例如,抗原)结合结构域,并且可调节CAR的不同多肽包含细胞内信号传导区域,其中两个多肽借助多聚化结构域的多聚化而进行的多聚化产生包含配体结合结构域和细胞内信号传导区域的可调节CAR。在一些实施方案中,多聚化可以诱导、调节、激活、介导和/或促进含有可调节CAR的工程化细胞中的信号。在一些实施方案中,诱导物与可调节CAR的至少一个多肽的多聚化结构域结合并诱导可调节CAR的构象变化,其中所述构象变化激活信号传导。在一些实施方案中,配体与此类嵌合受体的结合诱导多肽链中的构象变化,在一些情形中,所述构象变化包括多肽链寡聚化,其可以使受体能够进行细胞内信号传导。

[0745] 在一些实施方案中,诱导物发挥功能以偶联或多聚化(例如,二聚化)在工程化细胞中表达的可调节CAR的一组至少两条多肽链,以使所述可调节CAR产生所需细胞内信号,如在可调节CAR与靶抗原的相互作用期间。通过诱导物使可调节CAR的至少两个多肽发生偶联或多聚化是在诱导物与多聚化结构域结合后实现的。例如,在一些实施方案中,工程化细胞中的第一多肽和第二多肽可以各自包含能够结合诱导物的多聚化结构域。在多聚化结构域与诱导物结合后,第一多肽和第二多肽偶联在一起以产生所需细胞内信号。在一些实施方案中,多聚化结构域位于多肽的细胞内部分上。在一些实施方案中,多聚化结构域位于多肽的细胞外部分上。

[0746] 在一些实施方案中,可调节CAR的一组至少两个多肽包含两个、三个、四个或五个或更多个多肽。在一些实施方案中,所述组的至少两个多肽是相同多肽,例如,包含细胞内信号传导区域和多聚化结构域的两个、三个或更多个相同多肽。在一些实施方案中,所述组的至少两个多肽是不同多肽,例如,第一多肽包含配体(例如,抗原)结合结构域和多聚化结构域,并且第二多肽包含细胞内信号传导区域和多聚化结构域。在一些实施方案中,在诱导物的存在下生成细胞内信号。在一些实施方案中,细胞内信号是在不存在诱导物的情况下生成的,例如,诱导物干扰可调节CAR的至少两个多肽的多聚化,从而阻止通过所述可调节CAR进行细胞内信号传导。

[0747] 在一些实施方案中,多链CAR,例如通过HDR将编码至少一条多肽链的核酸序列整合至内源CD247基因座中。在嵌合受体(例如,多链CAR)的一些实施方案中,例如通过HDR将编码嵌合受体多肽链的核酸序列整合至内源CD247基因座中,所述嵌合受体多肽链(在一些情形中)在C末端含有CD3zeta (CD3 ζ) 链或其片段。在一些实施方案中,可以将编码两条或更多条嵌合受体多肽链中的其他多肽链(例如,不含有CD3zeta (CD3 ζ) 链或其片段的多肽链)的核酸序列靶向于相同基因座内(例如,相同转基因序列内,通常位于编码含有CD3zeta (CD3 ζ) 链或其片段的嵌合受体多肽链的核酸序列的5')或不同基因座处。在一些方面,编码两种或更多种单独的嵌合受体中的其他嵌合受体(例如,不含有CD3zeta (CD3 ζ) 链或其片段的嵌合受体)的核酸序列的引入可以通过不同的递送方法来进行,例如,通过瞬时递送方法或作为附加体核酸分子来进行。

[0748] 在一些实施方案中,多链CAR中的一条或多条多肽链可以包括多聚化结构域。在一些实施方案中,一种或多种受体多肽(如所提供的嵌合受体多肽的一部分)可以包括多聚化结构域。在一些实施方案中,多聚化结构域可以在结合诱导物后发生多聚化(例如,二聚化)。本文考虑的诱导物包括但不限于化学诱导物或蛋白质(例如,半胱天冬酶)。在一些实施方案中,诱导物选自雌激素、糖皮质激素、维生素D、类固醇、四环素、环孢菌素、雷帕霉素、香豆霉素、赤霉素、FK1012、FK506、FKCsA、rimiducid或HaXS或其类似物或衍生物。在一些实施方案中,诱导物是AP20187或AP20187类似物,如AP1510。

[0749] 在一些实施方案中,多聚化结构域可以在结合诱导物(如本文提供的诱导物)后发生多聚化(例如,二聚化)。在一些实施方案中,多聚化结构域可以来自FKBP、亲环蛋白受体、类固醇受体、四环素受体、雌激素受体、糖皮质激素受体、维生素D受体、钙神经素A、CyP-Fas、mTOR的FRB结构域、GyrB、GAI、GID1、Snap-tag和/或HaloTag或其部分或衍生物。在一些实施方案中,多聚化结构域是FK506结合蛋白(FKBP)或其衍生物、或其片段和/或多聚体,如FKBP12v36。在一些实施方案中,FKBP包含氨基酸序列GVQVETISPGDGRTPKRGQTCVVHYTGMLE

DGKKMDSSRDRNKPFKMLGKQEVIRGWEEGVAQMSVGRRAKLTISPDIYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO:82)。在一些实施方案中,FKBP12v36包含氨基酸序列GVQVETISPGDGRTPKRGQT CVVHYTGMLLEDGKKVDSSRDRNKPFKMLGKQEVIRGWEEGVAQMSVGRRAKLTISPDIYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO:83)。

[0750] 示例性诱导物和相应的多聚化结构域是已知的,例如,如以下文献中所述:美国专利申请公开号2016/0046700;Clackson等人(1998)Proc Natl Acad Sci U S A.95(18):10437-42;Spencer等人(1993)Science 262(5136):1019-24;Farrar等人(1996)Nature383(6596):178-81;Miyamoto等人(2012)Nature Chemical Biology 8(5):465-70;Erhart等人(2013)Chemistry and Biology 20(4):549-57。在一些实施方案中,诱导物是rimiducid(也称为AP1903;CAS索引名:2-哌啶甲酸1-[(2S)-1-氧代-2-(3,4,5-三甲氧基苯基)丁基]-,1,2-乙二基双[亚氨基(2-氧代-2,1-乙二基)氧基-3,1-亚苯基[(1R)-3-(3,4-二甲氧基苯基)亚丙基]]酯,[2S-[1(R*),2R*[S*[S*[1(R*),2R]]]]-(9C1);CAS登记号:195514-63-7;分子式: $C_{78}H_{98}N_4O_{20}$;分子量:1411.65),并且多聚化结构域是FK506结合蛋白(FKBP)。

[0751] 在一些实施方案中,工程化细胞的细胞膜对于诱导物是不可渗透的。在一些实施方案中,工程化细胞的细胞膜对于诱导物是可渗透的。

[0752] 在一些实施方案中,在不存在诱导物的情况下,可调节CAR不是多聚物或二聚物的部分。在结合诱导物后,多聚化结构域可以发生多聚化,例如,二聚化。在一些方面,多聚化结构域的多聚化导致可调节CAR的多肽与可调节CAR的另一多肽的多聚化,例如可调节CAR的至少两个多肽的多聚复合物。在一些实施方案中,多聚化结构域的多聚化可以借助诱导信号传导组分的物理靠近或者多聚体或二聚体的形成来诱导、调节、激活、介导和/或促进信号传导。在一些实施方案中,在结合诱导物后,多聚化结构域的多聚化还诱导与多聚化结构域直接或间接连接的信号传导结构域的多聚化。在一些实施方案中,多聚化诱导、调节、激活、介导和/或促进经由信号传导结构域或区域的信号传导。在一些实施方案中,与多聚化结构域连接的信号传导结构域或区域是细胞内信号传导区。

[0753] 在一些实施方案中,多聚化结构域在细胞内或者与工程化细胞(例如,工程化T细胞)的细胞内或胞质侧上的细胞膜缔合。在一些方面,细胞内多聚化结构域与膜缔合结构域(例如,脂质连接结构域,如豆蔻酰化结构域、棕榈酰化结构域、异戊烯化结构域或跨膜结构域)直接或间接连接。在一些实施方案中,多聚化结构域在细胞内,并且经由跨膜结构域与细胞外配体(例如,抗原)结合结构域连接。在一些实施方案中,细胞内多聚化结构域与细胞内信号传导区直接或间接连接。在一些方面,诱导的多聚化结构域的多聚化也使细胞内信号传导区彼此靠近,从而允许多聚化(例如,二聚化),并刺激细胞内信号传导。在一些实施方案中,可调节CAR的多肽包含跨膜结构域、一个或多个细胞内信号传导区以及一个或多个多聚化结构域,它们各自直接或间接连接。

[0754] 在一些实施方案中,多聚化结构域在细胞外或者与工程化细胞(例如,工程化T细胞)的细胞外侧上的细胞膜缔合。在一些方面,细胞外多聚化结构域与膜缔合结构域(例如,脂质连接结构域,如豆蔻酰化结构域、棕榈酰化结构域、异戊烯化结构域或跨膜结构域)直接或间接连接。在一些实施方案中,细胞外多聚化结构域与配体结合结构域(例如,抗原结合结构域)直接或间接连接,如用于结合至与疾病相关的抗原。在一些实施方案中,多聚化结构域在细胞外,并且经由跨膜结构域与细胞内信号传导区连接。

[0755] 在一些方面,膜缔合结构域是现有跨膜蛋白的跨膜结构域。在一些例子中,膜缔合结构域是本文所述的任何跨膜结构域。在一些方面,膜缔合结构域含有蛋白质-蛋白质相互作用基序或跨膜序列。

[0756] 在一些方面,膜缔合结构域是酰化结构域,如豆蔻酰化结构域、棕榈酰化结构域、异戊烯化结构域(即,法尼基化、牻牛儿基-牻牛儿基化、CAAX盒)。例如,膜缔合结构域可以是存在于蛋白质的N末端或C末端的酰化序列基序。此类结构域含有可以由酰基转移酶识别的特定序列基序,所述酰基转移酶将酰基部分转移至含有所述结构域的多肽。例如,酰化基序可以被单一酰基部分修饰(在一些情况下,所述酰基部分之后有几个带正电的残基(例如人c-*Src*:MGSNKSQPKDASQRRR (SEQ ID NO:84),以改进与阴离子脂质头基的缔合)。在其他方面,乙酰化基序能够被多个酰基部分修饰。例如,双酰化区位于某些蛋白激酶的N末端区域内,所述蛋白激酶如*Src*家族成员的子集(例如,*Yes*、*Fyn*、*Lck*)和G蛋白 α 亚基。示例性双酰化区含有序列基序Met-Gly-Cys-Xaa-Cys (SEQ ID NO:85),其中Met被切割,Gly发生N-酰化,并且一个Cys残基发生S-酰化。Gly通常发生豆蔻酰化,并且Cys可以发生棕榈酰化。

[0757] 其他示例性酰化区包括序列基序Cys-Ala-Ala-Xaa(所谓的“CAAX盒”;SEQ ID NO:86),其可以被C15或O10异戊烯基部分修饰,并且是已知的(参见例如,Gauthier-Campbell等人(2004) *Molecular Biology of the Cell* 15:2205-2217;Glabati等人(1994) *Biochem. J.* 303:697-700;以及Zlakine等人(1997) *J. Cell Science* 110:673-679;ten Klooster等人(2007) *Biology of the Cell* 99:1-12;Vincent等人(2003) *Nature Biotechnology* 21:936-40)。在一些实施方案中,酰基部分是C1-C20烷基、C2-C20烯基、C2-C20炔基、C3-C6环烷基、C1-C4卤代烷基、C4-C12环烷基烷基、芳基、取代的芳基或芳基(C1-C4)烷基。在一些实施方案中,含有酰基的部分是脂肪酸,并且脂肪酸部分的例子是丙基(C3)、丁基(C4)、戊基(C5)、己基(C6)、庚基(C7)、辛基(C8)、壬基(C9)、癸基(C10)、十一烷基(C11)、月桂基(C12)、肉豆蔻基(C14)、棕榈基(C16)、硬脂酰基(C18)、二十烷基(C20)、山萘基(C22)和木蜡基部分(C24),并且每部分可以含有0、1、2、3、4、5、6、7或8个不饱和键(即,双键)。在一些例子中,酰基部分是脂质分子,如磷脂酰基脂质(例如,磷脂酰丝氨酸、磷脂酰肌醇、磷脂酰乙醇胺、磷脂酰胆碱)、鞘脂(例如,鞘磷脂、鞘氨醇、神经酰胺、神经节苷脂、脑苷脂)或其修饰的形式。在某些实施方案中,一个、两个、三个、四个或五个或更多个酰基部分与膜缔合结构域连接。

[0758] 在一些方面,膜缔合结构域是促进糖脂(也称为糖基磷脂酰肌醇或GPI)的添加的结构域。在一些方面,GPI分子通过转酰胺反应翻译后附着至蛋白质靶标,从而导致羧基末端GPI信号序列的切割(参见例如,White等人(2000) *J. Cell Sci.* 113:721)且同时将已合成的GPI锚分子转移至新形成的羧基末端氨基酸(参见例如,Varki A等人,编辑. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1999. 第10章, *Glycophospholipid Anchors*. 可从以下网址获得:<https://www.ncbi.nlm.nih.gov/books/NBK20711/>)。在某些实施方案中,膜结合结构域是GPI信号序列。

[0759] 在一些实施方案中,如本文提供的多聚化结构域连接至细胞内信号传导区,例如,初级信号传导区域和/或共刺激信号传导结构域。在一些实施方案中,多聚化结构域在细胞外,并且经由跨膜结构域与细胞内信号传导区连接。在一些实施方案中,多聚化结构域在细

胞内,并且经由跨膜结构域与配体(例如,抗原)结合结构域连接。配体结合结构域和跨膜结构域可以直接或间接连接。在一些实施方案中,配体结合结构域与跨膜通过间隔子(如本文所述的任一种)连接。在一些实施方案中,多聚化结构域是FK506结合蛋白(FKBP)或其衍生物或片段,如FKBP12v36。在一些例子中,在引入诱导物如rimiducid之后,可调节CAR的多肽发生多聚化,例如,二聚化,从而刺激与多聚化结构域缔合的信号传导结构域并形成多聚复合物。多聚复合物的形成导致诱导、调节、刺激、激活、介导和/或促进经由细胞内信号传导区的信号。

[0760] 在一些实施方案中,经由可调节CAR的信号传导可以以条件方式经由条件多聚化来调节。例如,可调节CAR的多肽的多聚化结构域可以结合诱导物以进行多聚化,并且诱导物可以从外源提供。在一些方面,在结合诱导物后,多聚化结构域发生多聚化并诱导、调节、激活、介导和/或促进经由信号传导结构域的信号传导。例如,可以从外源施用诱导物,从而控制提供至含有所述可调节CAR的工程化细胞的信号的位置和持续时间。在一些实施方案中,可调节CAR的多肽的多聚化结构域可以结合诱导物以进行多聚化,并且诱导物可以内源提供。例如,诱导物可以由工程化细胞(例如,工程化T细胞)在可诱导或条件启动子的控制下从重组表达载体或从所述工程化细胞的基因组内源产生,从而控制提供至含有可调节CAR的工程化细胞的信号的位置和持续时间。

[0761] 在一些实施方案中,使用自杀开关控制可调节CAR。示例性嵌合受体利用可诱导半胱天冬酶-9(iCasp9)系统,包含人半胱天冬酶-9与修饰的FKBP二聚化结构域的融合物,从而允许与诱导物(例如,AP1903)结合后的条件二聚化。在通过结合诱导物进行二聚化之后,半胱天冬酶-9被激活并导致表达嵌合受体的细胞的细胞凋亡和细胞死亡(参见例如,Di Stasi等人(2011)N.Engl.J.Med.365:1673-1683)。

[0762] 在一些实施方案中,示例性可调节CAR包括:(1)可调节CAR的第一多肽,其包含:(i)细胞内信号传导区;以及(ii)能够结合诱导物的至少一个多聚化结构域;以及(2)可调节CAR的第二多肽,其包含:(i)配体(例如,抗原)结合结构域;(ii)跨膜结构域;以及(iii)能结合诱导物的至少一个多聚化结构域。在一些实施方案中,示例性可调节CAR包括:(1)可调节CAR的第一多肽,其包含:(i)跨膜结构域或酰化结构域;(ii)细胞内信号传导区;以及(iii)能结合诱导物的至少一个多聚化结构域;以及(2)可调节CAR的第二多肽,其包含:(i)配体(例如,抗原)结合结构域;(ii)跨膜结构域;以及(iii)能结合诱导物的至少一个多聚化结构域。在一些实施方案中,细胞内信号传导区还包含共刺激信号传导结构域。在一些实施方案中,第二多肽还包含共刺激信号传导结构域。在一些实施方案中,两个多肽上的至少一个多聚化结构域在细胞内。在一些实施方案中,两个多肽上的至少一个多聚化结构域在细胞外。

[0763] 在一些实施方案中,示例性可调节CAR包括:(1)可调节CAR的第一多肽,其包含:(i)能结合诱导物的至少一个细胞外多聚化结构域;(ii)跨膜结构域;以及(iii)细胞内信号传导区;以及(2)可调节CAR的第二多肽,其包含:(i)配体(例如,抗原)结合结构域;(ii)能结合诱导物的至少一个细胞外多聚化结构域,以及(iii)跨膜结构域、酰化结构域或GPI信号序列。在一些实施方案中,细胞内信号传导区还包含共刺激信号传导结构域。在一些实施方案中,第二多肽还包含共刺激信号传导结构域。

[0764] 在一些实施方案中,示例性可调节CAR包括:(1)可调节CAR的第一多肽,其包含:

(i) 跨膜结构域或酰化结构域; (ii) 至少一个共刺激结构域; (iii) 能够结合诱导物的多聚化结构域, 以及 (iv) 细胞内信号传导区; 以及 (iii) 至少一个共刺激结构域; 以及 (2) 可调节 CAR 的第二多肽, 其包含: (i) 配体 (例如, 抗原) 结合结构域; (ii) 跨膜结构域; (iii) 至少一个共刺激结构域; 以及 (iv) 能结合诱导物的至少一个细胞外多聚化结构域。

[0765] 在一些方面, 示例性可调节 CAR 中所述的任何区域和/或结构域可以按多种不同顺序排序。在一些方面, 一种或多种可调节 CAR 的各种多肽含有在细胞膜同一侧上的多聚化结构域, 例如, 两种或更多种多肽中的多聚化结构域都在细胞内或都在细胞外。

[0766] 可调节 CAR 的变异是已知的, 例如描述于以下文献中: 美国专利申请公开号 2014/0286987、美国专利申请公开号 2015/0266973、国际专利申请公开号 WO 2014/127261 和国际专利申请公开号 WO 2015/142675。

3. 嵌合自身抗体受体 (CAAR)

[0767] 在一些实施方案中, 由修饰的 CD247 基因座编码的嵌合受体是嵌合自身抗体受体 (CAAR)。在一些方面, CAAR 包括 CD3zeta (CD3 ζ) 链或其片段或部分, 如 CD3 ζ 信号传导结构域。在一些方面, CD3 ζ 信号传导结构域的全部 (例如, 整个 CD3 ζ 信号传导结构域) 或片段是由内源 CD247 基因座的开放阅读框或其部分序列编码。在一些实施方案中, CAAR 结合 (例如特异性结合) 或识别自身抗体。在一些实施方案中, CAAR 表达细胞 (如工程化以表达 CAAR 的 T 细胞) 可以用于结合至并杀伤表达自身抗体的细胞, 而不是表达正常抗体的细胞。在一些实施方案中, CAAR 表达细胞可以用于治疗与自身抗原的表达相关的自身免疫性疾病, 如自身免疫性疾病。在一些实施方案中, 表达 CAAR 的细胞可以靶向最终产生自身抗体并在其细胞表面上展示自身抗体的 B 细胞, 将这些 B 细胞标记为用于治疗性干预的疾病特异性靶标。在一些实施方案中, 表达 CAAR 的细胞可以用于通过使用抗原特异性嵌合自身抗体受体靶向引起疾病的 B 细胞, 有效靶向和杀死自身免疫疾病中的致病性 B 细胞。在一些实施方案中, 嵌合受体是 CAAR, 如美国专利申请公开号 US 2017/0051035 中所述的任一种。

[0768] 在一些实施方案中, CAAR 包含自身抗体结合结构域、跨膜结构域和一个或多个细胞内信号传导区域或结构域 (也可互换地称为胞质信号传导结构域或区域)。在一些实施方案中, 细胞内信号传导区域包含细胞内信号传导结构域。在一些实施方案中, 细胞内信号传导结构域是或包含初级信号传导区、能够在 T 细胞中刺激和/或诱导初级激活信号的信号传导结构域、T 细胞受体 (TCR) 组分的信号传导结构域 (例如 CD3-Zeta (CD3 ζ) 链的细胞内信号传导结构域或区域或其功能变体或信号传导部分) 和/或包含基于免疫受体酪氨酸的激活基序 (ITAM) 的信号传导结构域。

[0769] 在一些实施方案中, 自身抗体结合结构域包含自身抗原或其片段。自身抗原的选择可以取决于所靶向的自身抗体的类型。例如, 自身抗原的选择可能是由于其识别与特定疾病状态 (例如自身免疫性疾病, 例如自身抗体介导的自身免疫性疾病) 相关的靶细胞 (例如 B 细胞) 上的自身抗体。在一些实施方案中, 自身免疫性疾病包括寻常型天疱疮 (PV)。示例性自身抗原包括桥粒芯糖蛋白 1 (Dsg1) 和 Dsg3。

C. 用于基因工程化的细胞和细胞的制备

[0770] 在一些实施方案中, 提供工程化细胞 (例如, 基因工程化的或修饰的细胞) 和工程化细胞的方法, 包括包含修饰的 CD247 基因座的基因工程化的细胞, 所述基因座包含编码重组受体或其部分的转基因序列。在一些实施方案中, 将含有包含编码嵌合受体的一部分的

转基因序列的核酸序列和/或一种或多种另外的分子的多核苷酸(例如,模板多核苷酸,如本文例如在章节I.B.2中所述的任何模板多核苷酸)引入一种细胞中用于工程化(例如根据本文所述的工程化方法)。在一些方面,使用本文提供的任何方法将细胞工程化。在一些实施方案中,工程化细胞含有修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些方面,工程化细胞的修饰的CD247基因座包括本文在章节III.A中所述的那些。

[0771] 在一些方面,多核苷酸(如模板多核苷酸,例如,本文在章节I.B.2中所述)和/或其部分中的转基因序列(外源或异源核酸序列,如本文在章节I.B.2中所述的任一种)是异源的,即,通常不存在于细胞或从细胞获得的样品中,如从另一生物体或细胞获得的转基因序列,所述转基因序列例如通常不在被工程化的细胞和/或衍生出这种细胞的生物体中被发现。在一些实施方案中,所述核酸序列不是天然存在的,如在自然界中没有发现的核酸序列,或者从在自然界中发现的核酸序列进行修饰,包括包含编码来自多种不同细胞类型的各种结构域的核酸的嵌合组合的核酸序列。

[0772] 在一些方面,提供产生基因工程化的T细胞的方法,所述方法涉及将例如本文在章节I.B.2中所述的任何所提供的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中。在一些方面,所述遗传破坏是通过用于引入靶向遗传破坏的任何药剂或方法(包括本文例如在章节I.A中所述的任一种)来引入。在一些方面,所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些方面,提供产生基因工程化的T细胞的方法,其涉及将一种或多种药剂引入T细胞中,所述一种或多种药剂能够在T细胞的内源CD247基因座内的靶位点诱导遗传破坏;以及将例如本文在章节I.B.2中所述的任何所提供的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中,其中所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些实施方案中,核酸序列包含编码嵌合受体的一部分的转基因序列,并且所述转基因序列经由同源定向修复(HDR)被靶向以供整合于内源CD247基因座内。

[0773] 在一些实施方案中,提供产生基因工程化的T细胞的方法,其涉及将包含编码嵌合受体或其部分的核酸序列的多核苷酸引入T细胞中,所述T细胞在T细胞的CD247基因座内具有遗传破坏,其中编码嵌合受体或其部分的核酸序列经由同源定向修复(HDR)被靶向以供整合于内源CD247基因座内。在一些实施方案中,所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。在一些实施方案中,所述核酸序列包含编码嵌合受体的一部分的转基因序列,如本文例如在章节I.B.2中所述的任一种。在一些实施方案中,在进行所述方法后,基因工程化的T细胞中的CD3 ζ 信号传导结构域的全部(例如,整个或全部CD3 ζ 信号传导结构域)或片段是由内源CD247基因座的开放阅读框或其部分序列编码。在一些实施方案中,核酸序列包含编码嵌合受体的一部分的转基因序列,所述部分任选地编码CD3 ζ 信号传导结构域的片段(如整个CD3 ζ 的一部分;如小于整个的CD3 ζ 信号传导结构域或全长CD3 ζ 信号传导结构域),并且其中开放阅读框或其部分序列编码CD3 ζ 信号传导结构域(如整个或全部CD3 ζ 信号传导结构域)或任选地CD3 ζ 信号传导结构域的另一片段。在一些实施方案

中,所编码的嵌合受体的CD3 ζ 信号传导结构域的至少一个片段(任选地整个CD3 ζ 信号传导结构域)是由内源CD247基因座的开放阅读框或其部分序列编码。

[0774] 细胞通常是真核细胞,例如哺乳动物细胞,并且通常是人细胞。在一些实施方案中,细胞源自血液、骨髓、淋巴或淋巴器官,是免疫系统的细胞,如先天免疫或适应性免疫的细胞,例如骨髓或淋巴细胞,包括淋巴细胞,通常是T细胞和/或NK细胞。其他示例性细胞包括干细胞,如多潜能干细胞和多能干细胞,包括诱导的多能干细胞(iPSC)。细胞通常是原代细胞,如直接从受试者分离和/或从受试者分离并冷冻的那些原代细胞。在一些实施方案中,细胞包括T细胞或其他细胞类型的一个或多个子集,如整个T细胞群、CD4+T细胞、CD8+T细胞及其亚群,如由以下各项所定义的那些亚群:功能、激活状态、成熟度、分化的可能性、扩增、再循环、定位和/或持久能力、抗原特异性、抗原受体类型、在特定器官或区室中的存在、标记或细胞因子分泌特征和/或分化程度。关于要治疗的受试者,细胞可以是同种异体的和/或自体的。所述方法包括现成的方法。在一些方面,如对于现成的技术,细胞是多能的和/或多潜能的,如干细胞,如iPSC。在一些实施方案中,所述方法包括从受试者分离细胞、将其制备、处理、培养和/或工程化,并在低温保存之前或之后将它们重新引入同一受试者体内。

[0775] T细胞和/或CD4+和/或CD8+T细胞的亚型和亚群包括幼稚T(T_N)细胞、效应T细胞(T_{EFF})、记忆T细胞及其亚型(如干细胞记忆T(T_{SCM})、中央记忆T(T_{CM})、效应记忆T(T_{EM})或终末分化的效应记忆T细胞)、肿瘤浸润淋巴细胞(TIL)、不成熟T细胞、成熟T细胞、辅助T细胞、细胞毒性T细胞、粘膜相关不变T(MAIT)细胞、天然存在和适应性调节性T(Treg)细胞、辅助T细胞(如TH1细胞、TH2细胞、TH3细胞、TH17细胞、TH9细胞、TH22细胞、滤泡性辅助T细胞)、 α/β T细胞和 δ/γ T细胞。

[0776] 在一些实施方案中,细胞是自然杀伤(NK)细胞。在一些实施方案中,细胞是单核细胞或粒细胞,例如,骨髓细胞、巨噬细胞、嗜中性粒细胞、树突细胞、肥大细胞、嗜酸性粒细胞和/或嗜碱性粒细胞。在一些实施方案中,细胞包括经由基因工程引入的一种或多种核酸,从而表达此类核酸的重组或基因工程化产物。在一些实施方案中,核酸是异源的,即通常不存在于细胞或从细胞获得的样品中,如从另一种生物体或细胞获得的核酸,例如,所述核酸通常不在被工程化的细胞和/或衍生出这种细胞的生物体中发现。在一些实施方案中,核酸不是天然存在的,如在自然界中未发现的核酸,包括包含编码来自多种不同细胞类型的多种结构域的核酸的嵌合组合的核酸。

[0777] 在一些实施方案中,工程化细胞的制备包括一个或多个培养和/或制备步骤。用于引入编码转基因受体(如CAR)的核酸的细胞可以从样品分离,所述样品如生物样品,例如,从受试者获得或源自受试者的样品。在一些实施方案中,从中分离细胞的受试者是患有疾病或病症或需要细胞疗法或将被施用细胞疗法的受试者。在一些实施方案中,受试者是需要特定治疗性干预(如过继细胞疗法,其中细胞被分离、处理和/或工程化)的人。

[0778] 因此,在一些实施方案中,细胞是原代细胞,例如,原代人细胞。样品包括直接取自受试者的组织、流体和其他样品,以及从一个或多个处理步骤得到的样品,所述处理步骤如分离、离心、基因工程化(例如,用病毒载体转导)、洗涤和/或孵育。生物样品可以是直接从生物来源获得的样品或经过加工的样品。生物样品包括但不限于体液(如血液、血浆、血清、脑脊液、滑液、尿液和汗液)、组织和器官样品,包括源自它们的处理样品。

[0779] 在一些方面,衍生出或分离细胞的样品是血液或源自血液的样品,或者是或源自单采术或白细胞单采术产物。示例性样品包括全血、外周血单个核细胞(PBMC)、白细胞、骨髓、胸腺、组织活检、肿瘤、白血病、淋巴瘤、淋巴结、肠相关淋巴组织、粘膜相关淋巴组织、脾、其他淋巴组织、肝、肺、胃、肠、结肠、肾、胰腺、乳房、骨、前列腺、子宫颈、睾丸、卵巢、扁桃体或其他器官和/或源自它们的细胞。在细胞疗法(例如,过继细胞疗法)的情况下,样品包括来自自体 and 同种异体来源的样品。

[0780] 在一些实施方案中,细胞源自细胞系,例如T细胞系。在一些实施方案中,细胞从异种来源(例如,从小鼠、大鼠、非人灵长类动物和猪)获得。

[0781] 在一些实施方案中,细胞的分离包括一个或多个制备步骤和/或非基于亲和力的细胞分离步骤。在一些例子中,将细胞在一种或多种试剂的存在下洗涤、离心和/或孵育,例如以去除不需要的组分、富集所需组分、裂解或去除对特定试剂敏感的细胞。在一些例子中,细胞基于一种或多种特性来分离,所述特性如密度、粘附特性、大小、对特定组分的敏感性和/或抗性。

[0782] 在一些例子中,来自受试者的循环血液的细胞是例如通过单采术或白细胞单采术获得。在一些方面,样品含有淋巴细胞,包括T细胞、单核细胞、粒细胞、B细胞、其他有核白细胞、红细胞和/或血小板,并且在一些方面含有除了红细胞和血小板以外的细胞。

[0783] 在一些实施方案中,洗涤从受试者收集的血细胞,例如以去除血浆级分并将细胞置于适当的缓冲液或介质中以用于后续处理步骤。在一些实施方案中,用磷酸盐缓冲盐水(PBS)洗涤细胞。在一些实施方案中,洗涤溶液缺少钙和/或镁和/或多种或全部二价阳离子。在一些方面,根据制造商的说明书,通过半自动“流通式”离心机(例如,Cobe 2991细胞处理器,Baxter)完成洗涤步骤。在一些方面,洗涤步骤是根据制造商的说明书通过切向流过滤(TFF)完成的。在一些实施方案中,在洗涤后将细胞重悬于多种生物相容的缓冲液中,所述缓冲液,如例如不含Ca⁺⁺/Mg⁺⁺的PBS。在某些实施方案中,去除血细胞样品的组分并且将细胞直接重悬于培养基中。

[0784] 在一些实施方案中,所述方法包括基于密度的细胞分离方法,如通过裂解红细胞并通过Percoll或Ficoll梯度离心而从外周血制备白细胞。

[0785] 在一些实施方案中,分离方法包括基于一种或多种特定分子(如表面标记(例如,表面蛋白)、细胞内标记或核酸)在细胞中的表达或存在分离不同细胞类型。在一些实施方案中,可以使用任何已知的基于此类标记的分离方法。在一些实施方案中,分离是基于亲和力或基于免疫亲和力的分离。例如,在一些方面,所述分离包括基于一种或多种标记(通常是细胞表面标记)在细胞中的表达或表达水平来分离细胞和细胞群体,例如通过与特异性结合至此类标记的抗体或结合配偶体一起孵育,之后通常进行洗涤步骤并将已结合所述抗体或结合配偶体的细胞与尚未结合至所述抗体或结合配偶体的那些细胞分离。

[0786] 此类分离步骤可以基于阳性选择(其中保留已结合试剂的细胞以供进一步使用)和/或阴性选择(其中保留尚未结合至所述抗体或结合配偶体的细胞)。在一些例子中,保留两种级分以供进一步使用。在一些方面,在无法获得特异性鉴定异质性群体中的细胞类型的抗体,使得最好基于除了所需群体以外的细胞表达的标记来进行分离的情况下,阴性选择可以特别有用。

[0787] 分离无需导致特定细胞群或表达特定标记的细胞的100%富集或去除。例如,对特

定类型的细胞(如表达标记的那些)的阳性选择或富集是指增加此类细胞的数量或百分比,但不需要使不表达所述标记的细胞完全不存在。同样,对特定类型的细胞(如表达标记的那些)的阴性选择、去除或耗尽是指减少此类细胞的数量或百分比,但不需要使所有此类细胞完全去除。

[0788] 在一些例子中,进行多轮分离步骤,其中使来自一个步骤的阳性或阴性选择的级分经受另一个分离步骤,如随后的阳性或阴性选择。在一些例子中,单个分离步骤可以同时耗尽表达多种标记的细胞,如通过将细胞与多种抗体或结合配偶体(每种抗体或结合配偶体对被靶向以供阴性选择的标记具有特异性)一起孵育。同样,通过将细胞与在各种细胞类型上表达的多种抗体或结合配偶体一起孵育,可以同时多个细胞类型进行阳性选择。

[0789] 例如,在一些方面,T细胞的特定亚群(如对一种或多种表面标记呈阳性或高水平表达的细胞(例如,CD28⁺、CD62L⁺、CCR7⁺、CD27⁺、CD127⁺、CD4⁺、CD8⁺、CD45RA⁺和/或CD45RO⁺T细胞))通过阳性或阴性选择技术来分离。

[0790] 例如,可以使用抗CD3/抗CD28缀合的磁珠(例如,DYNABEADS®M-450CD3/CD28 T细胞扩增器)阳性选择CD3⁺、CD28⁺T细胞。

[0791] 在一些实施方案中,通过以下方式进行分离:通过阳性选择富集特定细胞群体,或者通过阴性选择耗尽特定细胞群体。在一些实施方案中,阳性或阴性选择通过将细胞与一种或多种抗体或其他结合剂一起孵育来完成,所述一种或多种抗体或其他结合剂与分别在阳性或阴性选择的细胞上表达(标记⁺)或以相对较高水平表达(标记^高)的一种或多种表面标记特异性结合。

[0792] 在一些实施方案中,通过阴性选择非T细胞(如B细胞、单核细胞或其他白细胞,如CD14)上表达的标记,将T细胞与PBMC样品分离。在一些方面,CD4⁺或CD8⁺选择步骤用于分离CD4⁺辅助T细胞和CD8⁺细胞毒性T细胞。通过对在一种或多种幼稚、记忆和/或效应T细胞亚群上表达或以相对较高程度表达的标记的阳性或阴性选择,可以将此类CD4⁺和CD8⁺群体还分类成亚群。

[0793] 在一些实施方案中,如通过基于与相应亚群相关的表面抗原进行阳性或阴性选择,将CD8⁺细胞针对幼稚、中央记忆、效应记忆和/或中央记忆干细胞进一步富集或耗尽。在一些实施方案中,进行中枢记忆T(T_{CM})细胞的富集以增加功效,如以改进施用后的长期存活、扩增和/或移植物植入,这在一些方面在此类亚群中特别稳健。参见Terakura等人(2012)Blood.1:72-82;Wang等人(2012)J Immunother.35(9):689-701。在一些实施方案中,组合富含T_{CM}的CD8⁺T细胞与CD4⁺T细胞进一步增强功效。

[0794] 在实施方案中,记忆T细胞存在于CD8⁺外周血淋巴细胞的CD62L⁺和CD62L⁻两个子集中。可以例如使用抗CD8和抗CD62L抗体将PBMC针对CD62L⁻CD8⁺和/或CD62L⁺CD8⁺级分进行富集或耗尽。

[0795] 在一些实施方案中,对中枢记忆T(T_{CM})细胞的富集是基于CD45RO、CD62L、CCR7、CD28、CD3和/或CD127的阳性或高表面表达;在一些方面,它是基于表达或高度表达CD45RA和/或颗粒酶B的细胞的阴性选择。在一些方面,富含T_{CM}细胞的CD8⁺群体的分离是通过耗尽表达CD4、CD14、CD45RA的细胞以及对表达CD62L的细胞的阳性选择或富集来进行的。在一个方面,中枢记忆T(T_{CM})细胞的富集从基于CD4表达所选择的阴性细胞级分开始进行,所述阴性细胞级分基于CD14和CD45RA的表达进行阴性选择且基于CD62L进行阳性选择。在一些方

面,此类选择是同时进行的,并且在其他方面,是以任何顺序依序进行的。在一些方面,用于制备CD8⁺细胞群或亚群的相同的基于CD4表达的选择步骤也用于生成CD4⁺细胞群或亚群,使得来自基于CD4的分离的阳性和阴性级分被保留并用于所述方法的后续步骤中,任选地在 一个或多个其他阳性或阴性选择步骤之后。

[0796] 在特定例子中,PBMC样品或其他白细胞样品进行CD4⁺细胞的选择,其中保留了阴性和阳性级分。然后所述阴性级分基于CD14和CD45RA或CD19的表达进行阴性选择,并基于中枢记忆T细胞(如CD62L或CCR7)的标记特征进行阳性选择,其中以任何顺序进行所述阳性和阴性选择。

[0797] 通过鉴定具有细胞表面抗原的细胞群,将CD4⁺T辅助细胞分类为幼稚、中枢记忆和效应细胞。CD4⁺淋巴细胞可以通过标准方法获得。在一些实施方案中,幼稚CD4⁺T淋巴细胞是CD45RO⁻、CD45RA⁺、CD62L⁺、CD4⁺T细胞。在一些实施方案中,中枢记忆CD4⁺细胞是CD62L⁺且CD45RO⁺。在一些实施方案中,效应CD4⁺细胞是CD62L⁻且CD45RO⁻。

[0798] 在一个例子中,为了通过阴性选择富集CD4⁺细胞,单克隆抗体混合剂通常包括针对CD14、CD20、CD11b、CD16、HLA-DR和CD8的抗体。在一些实施方案中,抗体或结合配偶体结合至固体支持物或基质(如磁珠或顺磁珠),以允许分离细胞用于阳性和/或阴性选择。例如,在一些实施方案中,使用免疫磁性(或亲和磁性)分离技术来分开或分离细胞和细胞群(综述于以下文献中:Methods in Molecular Medicine,第58卷:Metastasis Research Protocols,第2卷:Cell Behavior In Vitro and In Vivo,第17-25页S.A.Brooks和U.Schumacher编辑© Humana Press Inc.,新泽西州托托瓦)。

[0799] 在一些方面,将要分离的细胞的样品或组合物与小的可磁化或磁响应材料(如磁响应颗粒或微粒,如顺磁珠(例如像DynaBeads或MACS珠))一起孵育。磁响应材料(例如,颗粒)通常直接或间接地附着至结合配偶体(例如,抗体),所述结合配偶体与期望分离(例如,期望阴性或阳性选择)的一个细胞、多个细胞或细胞群上存在的分子(例如,表面标记)特异性结合。

[0800] 在一些实施方案中,磁粒或磁珠包含与特异性结合成员(如抗体或其他结合配偶体)结合的磁响应材料。有多种熟知的磁响应材料用于磁分离方法中。合适的磁性颗粒包括在Molday的美国专利号4,452,773和欧洲专利说明书EP 452342B中所述的那些,所述专利通过引用特此并入。胶体大小的颗粒(如在Owen美国专利号4,795,698;和Liberti等人的美国专利号5,200,084中所述的那些)是其他的例子。

[0801] 孵育通常在如下条件下进行:由此抗体或结合配偶体或者与附着于磁性颗粒或珠的此类抗体或结合配偶体特异性结合的分子(如二抗或其他试剂)与细胞表面分子(如果存在于所述样品内的细胞上的话)特异性结合。

[0802] 在一些方面,将样品置于磁场中,并且附着有磁响应或可磁化颗粒的那些细胞将被吸引到磁体并与未标记的细胞分离。对于阳性选择,保留被吸引到磁体的细胞;对于阴性选择,保留未被吸引的细胞(未标记的细胞)。在一些方面,在同一选择步骤期间进行阳性和阴性选择的组合,其中保留阳性和阴性级分并进一步处理或使其经历进一步分离步骤。

[0803] 在某些实施方案中,磁响应颗粒被包被在一抗或其他结合配偶体、二抗、凝集素、酶或链霉抗生物素蛋白中。在某些实施方案中,磁性颗粒通过对一种或多种标记具有特异性的一抗的包被而附着于细胞。在某些实施方案中,将细胞而不是珠用一抗或结合配偶体

标记,之后添加细胞类型特异性二抗或其他结合配偶体(例如,链霉亲和素)包被的磁粒。在某些实施方案中,链霉亲和素包被的磁粒是与生物素化的一抗或二抗结合使用。

[0804] 在一些实施方案中,磁响应颗粒保持附着于细胞,所述细胞随后被孵育,培养和/或工程化;在一些方面,颗粒保持附着于细胞以用于施用至患者。在一些实施方案中,从细胞去除可磁化或磁响应颗粒。从细胞去除可磁化颗粒的方法是已知的,并且包括例如使用竞争性非标记抗体和与可切割接头缀合的可磁化颗粒或抗体。在一些实施方案中,可磁化颗粒是可生物降解的。

[0805] 在一些实施方案中,基于亲和力的选择是通过磁激活细胞分选(MACS)(Miltenyi Biotec,加利福尼亚州奥本)来进行。磁激活细胞分选(MACS)系统能够高纯度选择具有附着至其上的磁化颗粒的细胞。在某些实施方案中,MACS以如下模式操作,其中在施加外部磁场之后依次洗脱非靶标和靶标种类。也就是说,附着至磁化颗粒的细胞被保持在适当的位置,而未附着的种类被洗脱。然后,在完成这个第一洗脱步骤后,以某种方式释放被捕获在磁场中并被阻止洗脱的种类,使得它们可以被洗脱和回收。在某些实施方案中,非靶细胞被标记并从异质细胞群耗尽。

[0806] 在某些实施方案中,使用如下系统、装置或设备进行分离或分开,所述系统、装置或设备进行所述方法的分离、细胞制备、分开、处理、孵育、培养和/或制备步骤中的一个或多个。在一些方面,所述系统用于在封闭或无菌环境中进行这些步骤中的每一个,例如以使错误、用户操作和/或污染降至最低。在一个例子中,所述系统是如国际专利申请公开号WO 2009/072003或US 20110003380中所述的系统。

[0807] 在一些实施方案中,系统或设备在集成或独立系统中和/或以自动化或可编程方式进行分离、处理、工程化和配制步骤中的一个或多个(例如,全部)。在一些方面,所述系统或设备包括与所述系统或设备通信的计算机和/或计算机程序,其允许用户对处理、分离、工程化和配制步骤的各个方面进行编程、控制、结果评估和/或调整。

[0808] 在一些方面,使用CliniMACS系统(Miltenyi Biotec)进行分离和/或其他步骤,例如以用于在封闭和无菌系统中在临床规模水平上自动分离细胞。部件可以包括集成微计算机、磁分离单元、蠕动泵和多种夹管阀。在一些方面,所述集成计算机控制所述仪器的所有部件并引导所述系统以标准化顺序执行重复程序。在一些方面,磁分离单元包括可移动的永磁体和用于选择柱的支架。蠕动泵控制整个管组的流速,并与夹管阀一起确保缓冲液通过所述系统的受控流动和细胞的连续悬浮。

[0809] 在一些方面,CliniMACS系统使用在无菌、无热原溶液中供应的抗体偶联的可磁化颗粒。在一些实施方案中,在用磁性颗粒标记细胞后,洗涤细胞以去除过量的颗粒。然后将细胞制备袋连接到管组,再将所述管组连接到含有缓冲液的袋和细胞收集袋。管组由预装配的无菌管路(包括预柱和分离柱)组成,并且仅供一次性使用。在启动分离程序后,系统将细胞样品自动上样到分离柱上。标记的细胞保留在柱内,而未标记的细胞通过一系列洗涤步骤去除。在一些实施方案中,用于与本文所述的方法一起使用的细胞群体未被标记并且不保留在柱中。在一些实施方案中,用于与本文所述的方法一起使用的细胞群体被标记并保留在柱中。在一些实施方案中,用于与本文所述方法一起使用的细胞群在去除磁场后从柱中洗脱,并被收集在细胞收集袋内。

[0810] 在某些实施方案中,使用CliniMACS Prodigy系统(Miltenyi Biotec)进行分离

和/或其他步骤。在一些方面,CliniMACS Prodigy系统配备有细胞处理联合体,其允许自动化洗涤和通过离心分级分离细胞。CliniMACS Prodigy系统还可以包括机载相机和图像识别软件,从而通过辨别源细胞产物的宏观层来确定最佳的细胞分级分离终点。例如,将外周血自动分离成红细胞、白细胞和血浆层。CliniMACS Prodigy系统还可以包括集成细胞培育室,所述集成细胞培育室完成细胞培养方案,如例如细胞分化和扩增、抗原加载和长期细胞培养。输入端口可以允许培养基的无菌去除和补充,并且可以使用集成显微镜监测细胞。参见例如,Klebanoff等人(2012) *J Immunother.* 35 (9) :651-660;Terakura等人(2012) *Blood.* 1:72-82;以及Wang等人(2012) *J Immunother.* 35 (9) :689-701。

[0811] 在一些实施方案中,经由流式细胞术收集并富集(或耗尽)本文所述的细胞群体,其中流体流中携带针对多种细胞表面标记染色的细胞。在一些实施方案中,通过制备规模(FACS)分选来收集和富集(或耗尽)本文所述的细胞群。在某些实施方案中,通过使用微机电系统(MEMS)芯片结合基于FACS的检测系统来收集和富集(或耗尽)本文描述的细胞群(参见例如,WO 2010/033140,Cho等人(2010) *Lab Chip* 10,1567-1573;和Godin等人(2008) *J Biophoton.* 1 (5) :355-376。在两种情况下,细胞可以用多种标记来标记,从而允许以高纯度分离明确定义的T细胞子集。

[0812] 在一些实施方案中,用一种或多种可检测的标记来标记抗体或结合配偶体,以促进用于阳性和/或阴性选择的分离。例如,分离可以基于与荧光标记的抗体的结合。在一些例子中,基于对一种或多种细胞表面标记具有特异性的抗体或其他结合配偶体的结合来分离细胞是在流体流中进行,如通过荧光激活细胞分选(FACS)(包括制备规模(FACS))和/或微机电系统(MEMS)芯片,例如与流式细胞检测系统组合。此类方法允许同时基于多种标记进行阳性和阴性选择。

[0813] 在一些实施方案中,制备方法包括在分离、孵育和/或工程化之前或之后冷冻(例如,冷冻保存)细胞的步骤。在一些实施方案中,冷冻和后续解冻步骤去除细胞群中的粒细胞,并且在一定程度上去除单核细胞。在一些实施方案中,例如,在洗涤步骤去除血浆和血小板之后,使细胞悬浮于冷冻溶液中。在一些方面,可以使用多种已知的冷冻溶液和参数中的任一种。一个例子包括使用含有20%DMSO和8%人血清白蛋白(HSA)的PBS,或其他合适的细胞冷冻培养基。然后用培养基将其1:1稀释,使得DMSO和HSA的最终浓度分别为10%和4%。然后通常将所述细胞以1°/分钟的速率冷冻至-80℃并储存在液氮储罐的气相中。

[0814] 在一些实施方案中,在基因工程化之前或与其结合地孵育和/或培养细胞。孵育步骤可以包括培养、培育、刺激、激活和/或繁殖。孵育和/或工程化可以在培养容器中进行,所述培养容器是如单元、室、孔、柱、管、管组、阀、小瓶、培养皿、袋子或者用于培养或培育细胞的其他容器。在一些实施方案中,在刺激条件或刺激剂的存在下孵育组合物或细胞。此类条件包括针对以下而设计的那些条件:诱导群体中细胞的增殖、扩增、激活和/或存活,模拟抗原暴露,和/或引发细胞用于基因工程化(如用于引入重组抗原受体)。

[0815] 所述条件可以包括以下中的一种或多种:特定培养基、温度、氧含量、二氧化碳含量、时间、药剂(例如,营养素、氨基酸、抗生素、离子和/或刺激因子(如细胞因子、趋化因子、抗原、结合配偶体、融合蛋白、重组可溶受体和设计为激活细胞的任何其他药剂))。

[0816] 在一些实施方案中,刺激条件或刺激剂包括能够刺激或激活TCR复合物的细胞内信号传导结构域的一种或多种药剂(例如,配体)。在一些方面,所述药剂在T细胞中开启或

启动TCR/CD3细胞内信号传导级联。此类药剂可包括抗体如对于TCR具有特异性的抗体,例如抗CD3。在一些实施方案中,刺激条件包括一种或多种药剂例如配体,其能够刺激共刺激受体,例如抗CD28。在一些实施方案中,这种药剂和/或配体可结合至固体支撑物如珠和/或一种或多种细胞因子。任选地,扩增方法可以还包括向培养基中添加抗CD3和/或抗CD28抗体(例如,以至少约0.5ng/mL的浓度)的步骤。在一些实施方案中,刺激剂包括IL-2、IL-15和/或IL-7。在一些方面,IL-2浓度为至少约10单位/mL。

[0817] 在一些方面,根据多种技术进行孵育,如以下文献中所述的那些:美国专利号6,040,177;Klebanoff等人(2012) *J Immunother.* 35 (9) :651-660;Terakura等人(2012) *Blood.* 1:72-82;和/或Wang等人(2012) *J Immunother.* 35 (9) :689-701。

[0818] 在一些实施方案中,通过以下方法扩增T细胞:向培养起始组合物中添加饲养细胞(如非分裂外周血单个核细胞(PBMC)) (例如,使得对于要扩增的初始群体中的每个T淋巴细胞,所得细胞群含有至少约5、10、20或40个或更多个PBMC饲养细胞);以及孵育培养物(例如,持续足以扩增所述T细胞数量的时间)。在一些方面,非分裂饲养细胞可以包含 γ 辐照的PBMC饲养细胞。在一些实施方案中,将所述PBMC用范围为约3000至3600拉德的 γ 射线辐照,以防止细胞分裂。在一些方面,在添加所述T细胞群体之前将所述饲养细胞添加至培养基。

[0819] 在一些实施方案中,刺激条件包括适合于人T淋巴细胞生长的温度,例如至少约25摄氏度,通常至少约30摄氏度,并且通常在或在约37摄氏度。任选地,孵育还可以包括添加非分裂的EBV转化的类淋巴母细胞(LCL)作为饲养细胞。可以用约6000至10,000拉德范围内的 γ 射线辐照LCL。在一些方面,LCL饲养细胞是以任何合适的量来提供,如LCL饲养细胞与初始T淋巴细胞的比率为至少约10:1。

[0820] 在实施方案中,通过用抗原刺激幼稚或抗原特异性T淋巴细胞获得抗原特异性T细胞,如抗原特异性CD4+和/或CD8+T细胞。例如,可以通过从被感染的受试者分离T细胞并在体外用相同抗原刺激所述细胞,针对巨细胞病毒抗原来生成抗原特异性T细胞系或克隆。

[0821] 用于引入基因工程化的组分(例如,用于诱导遗传破坏的药剂和/或编码嵌合受体(例如,CAR)的核酸)的各种方法是已知的,并且可以与所提供的方法和组合物一起使用。示例性方法包括用于转移编码多肽或受体的核酸的那些方法,包括经由病毒载体,例如逆转录病毒或慢病毒、非病毒载体或转座子(例如睡美人(Sleeping Beauty)转座子系统)。基因转移方法可以包括转导、电穿孔或导致将基因转移至细胞中的其他方法,或者本文在第I.A节中描述的任何递送方法。用于转移编码重组产物的核酸的其他方法和载体是描述于例如WO 2014055668和美国专利号7,446,190中的那些。

[0822] 在一些实施方案中,经由电穿孔将重组核酸转移到T细胞中(参见例如,Chicaybam等人,(2013) *PLoS ONE* 8 (3) :e60298和Van Tedeloo等人(2000) *Gene Therapy* 7 (16) :1431-1437)。在一些实施方案中,通过转座将重组核酸转移至T细胞中(参见例如,Manuri等人(2010) *Hum Gene Ther* 21 (4) :427-437;Sharma等人(2013) *Molec Ther Nucl Acids* 2, e74;和Huang等人(2009) *Methods Mol Biol* 506:115-126)。在免疫细胞中引入并表达遗传物质的其他方法包括磷酸钙转染(如以下文献中所述:Current Protocols in Molecular Biology, John Wiley&Sons, 纽约州纽约市)、原生质体融合、阳离子脂质体介导的转染;钨粒子促进的微粒轰击(Johnston, *Nature*, 346:776-777 (1990));和磷酸锶DNA共沉淀(Brash等人, *Mol. Cell Biol.*, 7:2031-2034 (1987))。

[0823] 在一些实施方案中,通过以下方式完成基因转移:首先刺激细胞,如通过将其与诱导反应(如增殖、存活和/或激活)的刺激物进行组合,例如如通过细胞因子或激活标记的表达测量的,然后转导激活的细胞,并且在培养中扩增至足以用于临床应用的数目。

[0824] 在一些情境下,可能需要防止刺激因子(例如,淋巴因子或细胞因子)的过表达可能潜在地导致受试者中不希望的结果或较低的功效(如受试者中与毒性相关的因素)的可能性。因此,在一些情境下,所述工程化细胞包括导致细胞在体内(如在过继免疫疗法中施用)对阴性选择易感的基因区段。例如,在一些方面,工程化细胞,使得它们可以由于施用它们的患者的体内状况的改变而被消除。阴性选择性表型可以由赋予对所施用的药剂(例如,化合物)的敏感性的基因的插入而产生。阴性选择性基因包括单纯疱疹病毒I型胸苷激酶(HSV-I TK)基因(Wigler等人,Cell 11:223,1977),其赋予更昔洛韦敏感性;细胞次黄嘌呤磷酸核糖基转移酶(HPRT)基因;细胞腺嘌呤磷酸核糖基转移酶(APRT)基因;细菌胞嘧啶脱氨酶(Mullen等人,Proc.Natl.Acad.Sci.USA.89:33(1992))。

[0825] 在一些实施方案中,细胞(例如,T细胞)可以在扩增期间或之后进行工程化。例如,用于引入所需多肽或受体的基因的这种工程化可以用任何合适的逆转录病毒载体来进行。然后可以使遗传修饰的细胞群摆脱初始刺激物(例如,CD3/CD28刺激物),并随后用第二类型的刺激物(例如,经由从头引入的受体)进行刺激。该第二类型的刺激物可以包括呈肽/MHC分子形式的抗原刺激物、遗传引入的受体的同源(交联)配体(例如CAR的天然配体)或在新受体的框架内直接结合(例如通过识别受体内的恒定区)的任何配体(如抗体)。参见例如,Cheadle等人,“Chimeric antigen receptors for T-cell based therapy”Methods Mol Biol.2012;907:645-66;或Barrett等人,Chimeric Antigen Receptor Therapy for Cancer Annual Review of Medicine第65卷:333-347(2014)。

[0826] 另外的核酸(例如,用于引入的基因)包括用于改善治疗功效的那些,如通过促进转移细胞的活力和/或功能;用于提供选择和/或评估细胞的遗传标记的基因,如以评估体内存活或定位;改善安全性的基因,例如通过使细胞在体内对阴性选择易感,如Lupton S.D.等人,Mol.and Cell Biol.,11:6(1991);和Riddell等人,Human Gene Therapy3:319-338(1992)所述;还参见Lupton等人的PCT/US91/08442和PCT/US94/05601的出版物,其描述了通过将显性阳性选择性标记与阴性选择性标记融合而得到的双功能选择性融合基因的使用。参见例如,Riddell等人,美国专利号6,040,177,第14-17栏。

[0827] 如本文所述,在一些实施方案中,在基因工程化之前或结合基因工程化来孵育和/或培养细胞。孵育步骤可以包括培养、培育、刺激、激活、繁殖和/或冷冻以保存(例如冷冻保存)。

D. 表达嵌合受体的细胞的组合物

[0828] 还提供多个工程化细胞或工程化细胞的群体、含有此类细胞和/或富含此类细胞的组合物。在一些方面,所提供的工程化细胞和/或工程化细胞的组合物包括本文所述的任一种,例如,其包含含有编码重组受体或其部分的转基因序列的修饰的CD247基因座,和/或是通过本文所述的方法产生的。在一些方面,所述多个工程化细胞或工程化细胞的群体含有本文例如在本文的章节III.C中所述的任何工程化细胞。在一些方面,所提供的细胞和细胞组合物可以使用本文所述的任何方法来工程化,例如,使用用于引入遗传破坏的一种或多种药剂或方法(例如,如本文在章节I.A中所述)和/或使用多核苷酸(如本文例如在章节

I.B.2中所述的模板多核苷酸)通过同源定向修复(HDR)来进行。在一些方面,本文提供的这种细胞群和/或组合物包含于药物组合物或者用于治疗用途或方法的组合物中,例如如本文在章节V中所述。

[0829] 在一些实施方案中,与使用其他方法生成的细胞群和/或组合物的表达和/或抗原结合相比,所提供的细胞群和/或含有工程化细胞的组合物包括展现更加改进的、均匀的、同质的和/或稳定的嵌合受体的表达和/或抗原结合(例如,展现减小的变异系数)的细胞群。在一些实施方案中,与使用其他方法(例如,编码嵌合受体的序列的随机整合)生成的相应群体相比,细胞群和/或组合物展现降低至少100%、95%、90%、80%、70%、60%、50%、40%、30%、20%或10%的嵌合受体的表达和/或嵌合受体的抗原结合的变异系数。所述变异系数定义为目的核酸(例如,编码嵌合受体或其部分的转基因序列)在细胞(例如,CD4+和/或CD8+T细胞)的群体中的表达的标准差除以相应目的核酸在相应细胞群中的表达平均值。在一些实施方案中,当在已经使用本文所提供的方法工程化的CD4+和/或CD8+T细胞中测量时,所述细胞群和/或组合物展现低于0.70、0.65、0.60、0.55、0.50、0.45、0.40、0.35或0.30或更低的变异系数。

[0830] 在一些实施方案中,所提供的细胞群和/或含有工程化细胞的组合物包括展现最小或减少的编码嵌合受体或其部分的转基因的随机整合的细胞群。在一些方面,由于将转基因整合至基因组中的不期望位置(例如整合至必需基因或调节细胞活性的关键基因中)和/或受体的失调或不受控表达,将转基因随机整合至细胞基因组中可能导致不良影响或细胞死亡。在一些方面,与使用其他方法生成的细胞群相比,转基因的随机整合减少至少或大于50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多。

[0831] 在一些实施方案中,提供包括多个表达嵌合受体的工程化免疫细胞的细胞群和/或组合物,其中编码嵌合受体的核酸序列存在于CD247基因座处,例如通过经由同源定向修复(HDR)将编码嵌合受体的一部分的转基因整合于CD247基因座处来实现。在一些实施方案中,组合物中至少或大于30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞和/或组合物中在CD247基因座处含有遗传破坏的细胞包含整合于CD247基因座处的编码嵌合受体的一部分的转基因。

[0832] 在一些实施方案中,所提供的组合物含有细胞,如其中表达嵌合受体的细胞构成组合物中的总细胞或某一类型的细胞(如T细胞或CD8+或CD4+细胞)的至少30%、40%、50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多。在一些实施方案中,所提供的组合物含有细胞,如其中表达嵌合受体的细胞构成组合物中在CD247基因座处含有遗传破坏的总细胞的至少30%、40%、50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多。

IV. 治疗方法

[0833] 本文提供治疗方法,例如,其包括施用任何本文所述的工程化细胞或含有工程化细胞的组合物,例如,包含含有编码重组受体或其部分的转基因的修饰的CD247基因座的工程化细胞。在一些方面,还提供将任何本文所述的工程化细胞或含有工程化细胞的组合物施用至受试者(如患有疾病或障碍的受试者)的方法。本文所述的表达嵌合受体(如嵌合抗原受体(CAR))的工程化细胞或包含所述工程化细胞的组合物可用于多种治疗性、诊断性和

预防性情形。例如,工程化细胞或包含工程化细胞的组合物可用于治疗受试者的多种疾病和障碍。此类方法和用途包括治疗性方法和用途,例如涉及将工程化细胞或含有工程化细胞的组合物施用至患有疾病、病症或障碍(如肿瘤或癌症)的受试者。在一些实施方案中,以有效量施用工程化细胞或包含工程化细胞的组合物以实现疾病或障碍的治疗。用途包括工程化细胞或组合物在此类方法和治疗中以及在制备药物以实施此类治疗方法中的用途。在一些实施方案中,所述方法通过向患有或怀疑患有疾病或病症的受试者施用工程化细胞或包含所述工程化细胞的组合物来进行。在一些实施方案中,所述方法从而治疗受试者的疾病或病症或障碍。还提供了用于将所述细胞和组合物施用至受试者(例如,患者)的治疗方法。

[0834] 用于过继细胞疗法的细胞的施用方法是已知的,并且可以与所提供的方法和组合物结合使用。例如,过继T细胞治疗方法描述于例如以下文献中:Gruenberg等人的美国专利申请公开号2003/0170238;Rosenberg的美国专利号4,690,915;Rosenberg(2011)Nat Rev Clin Oncol.8(10):577-85。参见例如,Themeli等人(2013)Nat Biotechnol.31(10):928-933;Tsukahara等人(2013)Biochem Biophys Res Commun 438(1):84-9;Davila等人(2013)PLoS ONE 8(4):e61338。

[0835] 所治疗的疾病或病症可以是任何疾病或病症,其中抗原的表达与疾病、病症或障碍的病因相关和/或参与所述病因,例如引起、加重或以其他方式参与这种疾病、病症或障碍。示例性疾病和病症可以包括与恶性肿瘤或细胞转化(例如癌症)、自身免疫性疾病或炎性疾病或者例如由细菌、病毒或其他病原体引起的感染性疾病相关的疾病或病症。本文中描述示例性抗原,其包括与可以治疗的各种疾病和病症相关的抗原。在特定实施方案中,所述嵌合抗原受体特异性结合至与所述疾病或病症相关的抗原。

[0836] 疾病、病症和障碍包括肿瘤,包括实体瘤、血液恶性肿瘤和黑色素瘤,并且包括局部和转移性肿瘤;感染性疾病,如病毒或其他病原体的感染,例如HIV、HCV、HBV、CMV、HPV和寄生虫病;以及自身免疫性和炎性疾病。在一些实施方案中,疾病、障碍或病症是肿瘤、癌症、恶性肿瘤、肿瘤或其他增殖性疾病或障碍。此类疾病包括但不限于白血病、淋巴瘤,例如,急性髓样(或髓性)白血病(AML)、慢性髓样(或髓性)白血病(CML)、急性淋巴细胞(或成淋巴细胞)白血病(ALL)、慢性淋巴细胞白血病(CLL)、毛细胞白血病(HCL)、小淋巴细胞淋巴瘤(SLL)、套细胞淋巴瘤(MCL)、边缘区淋巴瘤、伯基特淋巴瘤、霍奇金淋巴瘤(HL)、非霍奇金淋巴瘤(NHL)、间变性大细胞淋巴瘤(ALCL)、滤泡性淋巴瘤、难治性滤泡性淋巴瘤、弥漫性大B细胞淋巴瘤(DLBCL)和多发性骨髓瘤(MM)。在一些实施方案中,疾病或病症是选自以下的B细胞恶性肿瘤:急性成淋巴细胞性白血病(ALL)、成人ALL、慢性成淋巴细胞性白血病(CLL)、非霍奇金淋巴瘤(NHL)和弥漫性大B细胞淋巴瘤(DLBCL)。在一些实施方案中,疾病或病症是NHL,并且NHL选自侵袭性NHL、弥漫性大B细胞淋巴瘤(DLBCL)NOS型(从头的和从惰性转化的)、原发性纵隔大B细胞淋巴瘤(PMBCL)、富含T细胞/组织细胞的大B细胞淋巴瘤(TCHRBCL)、伯基特淋巴瘤、套细胞淋巴瘤(MCL)和/或滤泡性淋巴瘤(FL)(任选地3B级滤泡性淋巴瘤(FL3B))。

[0837] 在一些实施方案中,疾病或障碍是多发性骨髓瘤(MM)。在一些实施方案中,施用所提供的细胞(例如,含有编码嵌合受体(如CAR)的修饰的CD247基因座的工程化细胞)可以导致受试者的疾病或病症(如MM)的治疗和/或改善。在一些实施方案中,受试者患有或怀疑患

有与肿瘤相关抗原(如B细胞成熟抗原(BCMA))的表达相关的MM。

[0838] 在一些实施方案中,疾病或障碍是慢性淋巴细胞性白血病(CLL)。在一些实施方案中,施用所提供的细胞(例如,含有编码嵌合受体(如CAR)的修饰的CD247基因座的工程化细胞)可以导致受试者的疾病或病症(如CLL)的治疗和/或改善。在一些实施方案中,受试者患有或怀疑患有与肿瘤相关抗原(如受体酪氨酸激酶样孤儿受体1(ROR1))的表达相关的CLL。

[0839] 在一些实施方案中,疾病或障碍是实体瘤,或者与非血液肿瘤相关的癌症。在一些实施方案中,疾病或障碍是实体瘤,或者与实体瘤相关的癌症。在一些实施方案中,疾病或障碍是胰腺癌、膀胱癌、结直肠癌、乳腺癌、前列腺癌、肾癌、肝细胞癌、肺癌、卵巢癌、宫颈癌、胰腺癌、直肠癌、甲状腺癌、子宫癌、胃癌、食道癌、头颈癌、黑色素瘤、神经内分泌癌、CNS癌症、脑肿瘤、骨癌或软组织肉瘤。在一些实施方案中,疾病或障碍是膀胱癌、肺癌、脑癌、黑色素瘤(例如小细胞肺癌、黑色素瘤)、乳腺癌、宫颈癌、卵巢癌、结直肠癌、胰腺癌、子宫内膜癌、直肠癌、食道癌、肾癌、肝癌、前列腺癌、皮肤癌、甲状腺癌或子宫癌。在一些实施方案中,疾病或障碍是胰腺癌、膀胱癌、结直肠癌、乳腺癌、前列腺癌、肾癌、肝细胞癌、肺癌、卵巢癌、宫颈癌、胰腺癌、直肠癌、甲状腺癌、子宫癌、胃癌、食道癌、头颈癌、黑色素瘤、神经内分泌癌、CNS癌症、脑肿瘤、骨癌或软组织肉瘤。

[0840] 在一些实施方案中,疾病或障碍是非小细胞肺癌(NSCLC)。在一些实施方案中,施用所提供的细胞(例如,含有编码嵌合受体(如CAR)的修饰的CD247基因座的工程化细胞)可以导致受试者的疾病或病症(如NSCLC)的治疗和/或改善。在一些实施方案中,受试者患有或怀疑患有与肿瘤相关抗原(如受体酪氨酸激酶样孤儿受体1(ROR1))的表达相关的NSCLC。

[0841] 在一些实施方案中,疾病或病症是感染性疾病或病症,例如但不限于病毒、逆转录病毒、细菌和原生动物感染、免疫缺陷、巨细胞病毒(CMV)、EB病毒(EBV)、腺病毒、BK多瘤病毒。在一些实施方案中,疾病或病症是自身免疫性或炎性疾病或病症,如关节炎(例如类风湿性关节炎(RA))、I型糖尿病、系统性红斑狼疮(SLE)、炎性肠病、银屑病、硬皮病、自身免疫性甲状腺疾病、格雷夫斯病、克罗恩病、多发性硬化症、哮喘和/或与移植相关的疾病或病症。

[0842] 在一些实施方案中,与疾病或障碍相关的抗原是或包括 α v β 6整合素(α v β 6整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPHa2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、神经节苷脂GD3、糖蛋白100(gp100)、磷脂酰肌醇蛋白聚糖-3(GPC3)、G蛋白偶联受体C类5族成员D(GPRC5D)、Her2/neu(受体酪氨酸激酶erb-B2)、Her3(erb-B3)、Her4(erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原(HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1(HLA-A1)、人白细胞抗原A2(HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2(IL-13R α 2)、激酶插入结构域受体(kdr)、 κ 轻链、L1细胞粘附分子(L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列

的蛋白8家族成员A (LRRC8A)、Lewis Y、黑色素瘤相关抗原 (MAGE) -A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素 (MSLN)、c-Met、鼠类巨细胞病毒 (CMV)、粘蛋白1 (MUC1)、MUC16、自然杀伤细胞2族成员D (NKG2D) 配体、黑色素A (MART-1)、神经细胞粘附分子 (NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原 (PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原 (PSCA)、前列腺特异性膜抗原 (PSMA)、受体酪氨酸激酶样孤儿受体1 (ROR1)、存活蛋白、滋养层糖蛋白 (TPBG, 也称为5T4)、肿瘤相关糖蛋白72 (TAG72)、酪氨酸酶相关蛋白1 (TRP1, 也称为TYRP1或gp75)、酪氨酸酶相关蛋白2 (TRP2, 也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体 (VEGFR)、血管内皮生长因子受体2 (VEGFR2)、肾母细胞瘤1 (WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。在一些实施方案中,受体靶向的抗原包括与B细胞恶性肿瘤相关的抗原,如许多已知B细胞标记中的任何一种。在一些实施方案中,抗原是或包括CD20、CD19、CD22、ROR1、CD45、CD21、CD5、CD33、Ig κ 、Ig λ 、CD79a、CD79b或CD30。

[0843] 在一些实施方案中,抗原是或包括病原体特有的或病原体表达的抗原。在一些实施方案中,抗原是病毒抗原(如来自HIV、HCV、HBV等的病毒抗原)、细菌抗原和/或寄生虫抗原。

[0844] 在一些方面,嵌合受体(如CAR)特异性地结合至与疾病或病症相关的抗原或在与B细胞恶性肿瘤相关的病灶环境的细胞中表达的抗原。在一些实施方案中,受体靶向的抗原包括与B细胞恶性肿瘤相关的抗原,如许多已知B细胞标记中的任何一种。在一些实施方案中,受体靶向的抗原是CD20、CD19、CD22、ROR1、CD45、CD21、CD5、CD33、Ig κ 、Ig λ 、CD79a、CD79b或CD30、或其组合。

[0845] 在一些实施方案中,疾病或病症是骨髓瘤,如多发性骨髓瘤。在一些方面,嵌合受体(如CAR)特异性地结合至与疾病或病症相关的抗原或在与多发性骨髓瘤相关的病灶环境的细胞中表达的抗原。在一些实施方案中,受体靶向的抗原包括与多发性骨髓瘤相关的抗原。在一些方面,抗原(例如第二或另外的抗原,如疾病特异性抗原和/或相关抗原,例如B细胞成熟抗原 (BCMA)、G蛋白偶联受体C型家族5D (GPCR5D)、CD38 (环状ADP核糖水解酶)、CD138 (多配体聚糖-1、多配体聚糖、SYN-1)、CS-1 (CS1、CD2子集1、CRACC、SLAMF7、CD319和19A24)、BAFF-R、TACI和/或FcRH5)在多发性骨髓瘤上表达。其他示例性多发性骨髓瘤抗原包括CD56、TIM-3、CD33、CD123、CD44、CD20、CD40、CD74、CD200、EGFR、 β 2-微球蛋白、HM1.24、IGF-1R、IL-6R、TRAIL-R1和IIA型激活素受体 (ActRIIA)。参见Benson和Byrd, *J. Clin. Oncol.* (2012) 30 (16):2013-15; Tao和Anderson, *Bone Marrow Research* (2011):924058; Chu等人, *Leukemia* (2013) 28 (4):917-27; Garfall等人, *Discov Med.* (2014) 17 (91):37-46。在一些实施方案中,抗原包括存在于淋巴瘤、骨髓瘤、AIDS相关的淋巴瘤和/或移植后淋巴组织增生上的那些抗原,如CD38。针对此类抗原的抗体或抗原结合片段是已知的并且包括例如以下文献中所述的那些:美国专利号8,153,765;8,603,477、8,008,450;美国公开号US 20120189622或US20100260748;和/或国际PCT公开号WO 2006099875、WO 2009080829或WO 2012092612或WO 2014210064。在一些实施方案中,此类抗体或其抗原结合片段(例如scFv)含于多特异性抗体、多特异性嵌合受体(如多特异性CAR)和/或多特异性细胞中。

[0846] 在一些实施方案中,疾病或障碍与G蛋白偶联受体C类5族成员D (GPCR5D)的表达和/或B细胞成熟抗原 (BCMA)的表达相关。

[0847] 在一些实施方案中,疾病或障碍是B细胞相关障碍。在所提供方法的一些任何所提供的实施方案中,与BCMA相关的疾病或障碍是自身免疫性疾病或障碍。在所提供方法的一些任何所提供的实施方案中,自身免疫性疾病或障碍是系统性红斑狼疮(SLE)、狼疮性肾炎、炎性肠病、类风湿性关节炎、ANCA相关性血管炎、特发性血小板减少性紫癜(ITP)、血栓性血小板减少性紫癜(TTP)、自身免疫性血小板减少症、查加斯病、格雷夫斯病、韦格纳氏肉芽肿病、结节性多动脉炎、舍格伦综合征、寻常型天疱疮、硬皮病、多发性硬化症、牛皮癣、IgA肾病、IgM多发性神经病变、血管炎、糖尿病、雷诺综合征、抗磷脂综合征、古德帕斯丘病、川崎病、自身免疫性溶血性贫血、重症肌无力或进行性肾小球肾炎。

[0848] 在一些实施方案中,疾病或障碍是癌症。在一些实施方案中,癌症是表达GPCR5D的癌症。在一些实施方案中,癌症是浆细胞恶性肿瘤,并且所述浆细胞恶性肿瘤是多发性骨髓瘤(MM)或浆细胞瘤。在一些实施方案中,癌症是多发性骨髓瘤(MM)。在一些实施方案中,癌症是复发性/难治性多发性骨髓瘤。

[0849] 在一些实施方案中,抗原是ROR1,并且疾病或障碍是CLL。在一些实施方案中,抗原ROR1,并且疾病或障碍是NSCLC。

[0850] 在一些实施方案中,抗体或抗原结合片段(例如scFv或V_H结构域)特异性识别抗原(如CD19、BCMA、GPCR5D或ROR1)。在一些实施方案中,抗体或抗原结合片段源自与CD19、BCMA、GPCR5D或ROR1特异性结合的抗体或抗原结合片段或者是其变体。

[0851] 在一些实施方案中,细胞疗法(例如,过继T细胞疗法)通过自体转移进行,其中从接受细胞疗法的受试者或从源自这种受试者的样品中分离和/或以其他方式制备细胞。因此,在一些方面,细胞源自需要治疗的受试者(例如,患者),并且在分离和处理后将细胞施用于同一受试者。

[0852] 在一些实施方案中,细胞疗法(例如,过继T细胞疗法)通过同种异体转移进行,其中从将要接受或最终接受细胞疗法的受试者以外的受试者(例如,第一受试者)分离和/或以其他方式制备细胞。在此类实施方案中,然后将细胞施用于相同物种的不同受试者,例如第二受试者。在一些实施方案中,第一和第二受试者在遗传上是相同的。在一些实施方案中,第一和第二受试者在遗传上是相似的。在一些实施方案中,第二受试者表达与第一受试者相同的HLA类别或超类型。

[0853] 可以将细胞通过任何合适的方式施用,例如通过推注输注,通过注射例如静脉内或皮下注射、眼内注射、眼周注射、视网膜下注射、玻璃体内注射、经中隔注射、巩膜下注射、脉络膜内注射、前房注射、结膜下(subconjunctival)注射、结膜下(subconjunctival)注射、眼球筋膜囊下(sub-Tenon)注射、眼球后注射、眼球周注射或后近巩膜(posterior juxtасleral)递送。在一些实施方案中,将它们通过肠胃外、肺内和鼻内以及(如果需要用于局部治疗的话)病灶内施用来施用。肠胃外输注包括肌内、静脉内、动脉内、腹膜内或皮下施用。在一些实施方案中,给定剂量是通过细胞的单次推注施用来施用。在一些实施方案中,给定剂量通过例如在不超过3天的时间段内细胞的多次推注施用,或通过细胞的连续输注施用来施用。在一些实施方案中,细胞剂量或任何其他疗法(例如,淋巴细胞清除疗法、干预疗法和/或组合疗法)的施用是经由门诊递送进行的。

[0854] 对于疾病的预防或治疗,适当的剂量可取决于要治疗的疾病类型、细胞或嵌合受体的类型、疾病的严重程度和病程、是针对预防目的还是针对治疗目的而施用细胞、先前治

疗、受试者的临床病史和对细胞的反应以及主治医师的决断。在一些实施方案中,适合将组合物和细胞一次或在一系列治疗中施用至受试者。

[0855] 在一些实施方案中,将细胞作为组合治疗的一部分施用,如与另一种治疗性干预(如抗体或工程化细胞或受体或药剂(如细胞毒性剂或治疗剂))同时施用或以任何顺序依序施用。在一些实施方案中,将细胞与一种或多种另外的治疗剂共同施用或与另一种治疗性干预联合施用(同时或以任何顺序依次施用)。在一些情境下,将细胞与另一种疗法在时间上足够接近地共同施用,使得细胞群增强一种或多种另外的治疗剂的效果,或反之亦然。在一些实施方案中,在一种或多种另外的治疗剂之前施用细胞。在一些实施方案中,在一种或多种另外的治疗剂之后施用细胞。在一些实施方案中,一种或多种另外的药剂包括细胞因子(例如IL-2),例如以增强持久性。在一些实施方案中,所述方法包括施用化学治疗剂。

[0856] 在一些实施方案中,所述方法包括在所述施用之前施用化学治疗剂(例如,调理性化学治疗剂),例如以减小肿瘤负荷。

[0857] 在一些方面,用免疫清除(例如,淋巴细胞清除)疗法预调理受试者可以改善过继细胞疗法(ACT)的效果。

[0858] 因此,在一些实施方案中,所述方法包括在开始细胞疗法之前将预调理剂施用至受试者,所述预调理剂如淋巴细胞清除剂或化学治疗剂,如环磷酰胺、氟达拉滨或其组合。例如,可以在开始细胞疗法之前至少2天(如之前至少3、4、5、6或7天),向受试者施用预调理剂。在一些实施方案中,在开始细胞疗法之前不超过7天(如之前不超过6、5、4、3或2天),向受试者施用预调理剂。

[0859] 在一些实施方案中,用环磷酰胺以在或在约20mg/kg与100mg/kg之间,如在或在约40mg/kg与80mg/kg之间的剂量对受试者进行预调理。在一些方面,用或用约60mg/kg的环磷酰胺对受试者进行预调理。在一些实施方案中,可以将环磷酰胺按单剂量施用或者可以按多个剂量施用,如每日给予、每隔一天给予或每三天给予。在一些实施方案中,将环磷酰胺每日施用一次,持续一天或两天。在一些实施方案中,如果淋巴细胞清除剂包含环磷酰胺,那么按以下剂量向受试者施用环磷酰胺:在或在约100mg/m²与500mg/m²之间,例如在或在约200mg/m²与400mg/m²之间,或者250mg/m²与350mg/m²之间,包含端值。在一些情形中,向受试者施用约300mg/m²的环磷酰胺。在一些实施方案中,可以将环磷酰胺按单剂量施用或者可以按多个剂量施用,如每日给予、每隔一天给予或每三天给予。在一些实施方案中,每天施用环磷酰胺,如持续1-5天,例如持续3至5天。在一些情况下,在开始细胞疗法之前每天向受试者施用约300mg/m²的环磷酰胺,持续3天。

[0860] 在一些实施方案中,当淋巴细胞清除剂包含氟达拉滨时,向受试者施用剂量在或在约1mg/m²与100mg/m²之间、如在或在约10mg/m²与75mg/m²之间、15mg/m²与50mg/m²之间、20mg/m²与40mg/m²之间或24mg/m²与35mg/m²之间的氟达拉滨(包含端值)。在一些情形中,向受试者施用约30mg/m²的氟达拉滨。在一些实施方案中,可以将氟达拉滨按单一剂量施用或者可以按多个剂量施用,如每日给予、每隔一天给予或每三天给予。在一些实施方案中,每日施用氟达拉滨,如持续1-5天,例如持续3至5天。在一些情况下,在开始细胞疗法之前每天向受试者施用约30mg/m²的氟达拉滨,持续3天。

[0861] 在一些实施方案中,淋巴细胞清除剂包含药剂的组合,如环磷酰胺与氟达拉滨的组合。因此,药剂的组合可以包括按任何剂量或施用时间表(如本文所述的那些)的环磷酰

胺以及按任何剂量或施用时间表(如本文所述的那些)的氟达拉滨。例如,在一些方面,在第一剂量或后续剂量之前,向受试者施用60mg/kg(约2g/m²)的环磷酰胺和3至5个剂量的25mg/m²氟达拉滨。

[0862] 在一些实施方案中,在施用细胞后,例如通过许多已知方法中的任一种测量工程化细胞群的生物学活性。要评估的参数包括工程化或天然T细胞或其他免疫细胞与抗原的特异性结合,其在体内例如通过成像进行评估,或离体例如通过ELISA或流式细胞术进行评估。在某些实施方案中,工程化细胞破坏靶细胞的能力可以使用任何合适的已知方法来测量,所述方法例如描述于例如以下文献中的细胞毒性测定:Kochenderfer等人,J.Immunotherapy,32(7):689-702(2009),以及Herman等人J.Immunological Methods,285(1):25-40(2004)。在一些实施方案中,通过测定一种或多种细胞因子(如CD107a、IFN γ 、IL-2和TNF)的表达和/或分泌来测量细胞的生物学活性。在一些方面,通过评估临床结局(如肿瘤负荷或负担的减少)来测量生物活性。

[0863] 在某些实施方案中,将工程化细胞以任何数量的方式进一步修饰,使得其治疗或预防功效增加。例如,可以将群体表达的工程化CAR或TCR直接或通过接头间接缀合至靶向部分。将化合物(例如,CAR)与靶向部分缀合的实践在本领域是已知的。参见例如,Wadwa等人,J.Drug Targeting 3:1 1 1(1995),和美国专利5,087,616。

[0864] 在一些实施方案中,将细胞作为组合治疗的一部分施用,如与另一种治疗性干预(如抗体或工程化细胞或受体或药剂(如细胞毒性剂或治疗剂))同时施用或以任何顺序依序施用。在一些实施方案中,将细胞与一种或多种另外的治疗剂共同施用或与另一种治疗性干预联合施用(同时或以任何顺序依次施用)。在一些情境下,将细胞与另一种疗法在时间上足够接近地共同施用,使得细胞群增强一种或多种另外的治疗剂的效果,或反之亦然。在一些实施方案中,在一种或多种另外的治疗剂之前施用细胞。在一些实施方案中,在一种或多种另外的治疗剂之后施用细胞。在一些实施方案中,一种或多种另外的药剂包括细胞因子(如IL-2),例如以增强持久性。

[0865] 在一些实施方案中,根据所提供方法和/或用所提供制品或组合物将一定剂量的细胞施用至受试者。在一些实施方案中,剂量的大小或时间安排根据受试者的特定疾病或病症确定。在一些情况下,可以根据所提供的描述凭经验确定用于特定疾病的剂量的大小或时间安排。

[0866] 在一些实施方案中,细胞剂量包含在为或约 2×10^5 个细胞/kg与为或约 2×10^6 个细胞/kg之间,如在为或约 4×10^5 个细胞/kg与为或约 1×10^6 个细胞/kg之间或在为或约 6×10^5 个细胞/kg与为或约 8×10^5 个细胞/kg之间。在一些实施方案中,细胞的剂量包含不超过 2×10^5 个细胞(例如抗原表达细胞,如CAR表达细胞)/公斤受试者体重(细胞/kg),如不超过或不超过约 3×10^5 个细胞/kg、不超过或不超过约 4×10^5 个细胞/kg、不超过或不超过约 5×10^5 个细胞/kg、不超过或不超过约 6×10^5 个细胞/kg、不超过或不超过约 7×10^5 个细胞/kg、不超过或不超过约 8×10^5 个细胞/kg、不超过或不超过约 9×10^5 个细胞/kg、不超过或不超过约 1×10^6 个细胞/kg、或者不超过或不超过约 2×10^6 个细胞/kg。在一些实施方案中,细胞的剂量包含至少或至少约或为或约 2×10^5 个细胞(例如抗原表达细胞,如CAR表达细胞)/公斤受试者体重(细胞/kg),如至少或至少约或为或约 3×10^5 个细胞/kg、至少或至少约或为或约 4×10^5 个细胞/kg、至少或至少约或为或约 5×10^5 个细胞/kg、至少或至少约或为或约 6×10^5 个细胞/kg、至少或至少

约或为或约 7×10^5 个细胞/kg、至少或至少约或为或约 8×10^5 个细胞/kg、至少或至少约或为或约 9×10^5 个细胞/kg、至少或至少约或为或约 1×10^6 个细胞/kg、或至少或至少约或为或约 2×10^6 个细胞/kg。

[0867] 在某些实施方案中,将细胞或细胞亚型的单独群体按以下施用至受试者:在为或约10万至为或约1000亿个细胞的范围和/或每公斤受试者体重该量的细胞,如例如为或约10万至为或约500亿个细胞(例如,为或约500万个细胞、为或约2500万个细胞、为或约5亿个细胞、为或约10亿个细胞、为或约50亿个细胞、为或约200亿个细胞、为或约300亿个细胞、为或约400亿个细胞或由任两个前述值限定的范围)、为或约100万至为或约500亿个细胞(例如,为或约500万个细胞、为或约2500万个细胞、为或约5亿个细胞、为或约10亿个细胞、为或约50亿个细胞、为或约200亿个细胞、为或约300亿个细胞、为或约400亿个细胞或由任两个前述值限定的范围),如为或约1000万至为或约1000亿个细胞(例如,为或约2000万个细胞、为或约3000万个细胞、为或约4000万个细胞、为或约6000万个细胞、为或约7000万个细胞、为或约8000万个细胞、为或约9000万个细胞、为或约100亿个细胞、为或约250亿个细胞、为或约500亿个细胞、为或约750亿个细胞、为或约900亿个细胞或由任两个前述值限定的范围),并且在一些情况下,为或约1亿个细胞至为或约500亿个细胞(例如,为或约1.2亿个细胞、为或约2.5亿个细胞、为或约3.5亿个细胞、为或约6.5亿个细胞、为或约8亿个细胞、为或约9亿个细胞、为或约30亿个细胞、为或约300亿个细胞、为或约450亿个细胞)或在这些范围和/或每公斤受试者体重的这些范围之间的任何值。剂量可以根据疾病或障碍和/或患者和/或其他治疗特有的属性而变化。在一些实施方案中,这些值是指表达嵌合受体的细胞的数量;在其他实施方案中,它们是指施用的T细胞或PBMC或总细胞的数量。

[0868] 在一些实施方案中,例如,在受试者是人的情况下,剂量包括少于约 5×10^8 个总嵌合受体(例如,CAR)表达细胞、T细胞或外周血单个核细胞(PBMC),例如,在为或约 1×10^6 至为或约 5×10^8 个此类细胞的范围内,如为或约 2×10^6 、 5×10^6 、 1×10^7 、 5×10^7 、 1×10^8 、 1.5×10^8 或 5×10^8 个总此类细胞,或者在任两个前述值之间的范围。例如,在一些实施方案中,在受试者是人的情况下,剂量包括多于为或约 1×10^6 个总嵌合受体(例如CAR)表达细胞、T细胞或外周血单个核细胞(PBMC)且少于为或约 2×10^9 个总嵌合受体(例如CAR)表达细胞、T细胞或外周血单个核细胞(PBMC),例如,在为或约 2.5×10^7 至为或约 1.2×10^9 个此类细胞的范围内,如为或约 2.5×10^7 、 5×10^7 、 1×10^8 、 1.5×10^8 、 8×10^8 、或 1.2×10^9 个总此类细胞、或任两个前述值之间的范围。

[0869] 在一些实施方案中,基因工程化的细胞的剂量包含从为或约 1×10^5 至为或约 5×10^8 个总CAR表达(CAR⁺)T细胞、从为或约 1×10^5 至为或约 2.5×10^8 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 1×10^8 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 5×10^7 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 2.5×10^7 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 1×10^7 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 5×10^6 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 2.5×10^6 个总CAR⁺T细胞、从为或约 1×10^5 至为或约 1×10^6 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 5×10^8 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 2.5×10^8 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 1×10^8 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 5×10^7 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 2.5×10^7 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 1×10^7 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 5×10^6 个总CAR⁺T细胞、从为或约 1×10^6 至为或约 2.5×10^6 个总CAR⁺T细胞、从为或约 2.5×10^6 至为或约

5x10⁸个总CAR⁺T细胞、从为或约2.5x10⁶至为或约2.5x10⁸个总CAR⁺T细胞、从为或约2.5x10⁶至为或约1x10⁸个总CAR⁺T细胞、从为或约2.5x10⁶至为或约5x10⁷个总CAR⁺T细胞、从为或约2.5x10⁶至为或约2.5x10⁷个总CAR⁺T细胞、从为或约2.5x10⁶至为或约1x10⁷个总CAR⁺T细胞、从为或约2.5x10⁶至为或约5x10⁶个总CAR⁺T细胞、从为或约5x10⁶至为或约5x10⁸个总CAR⁺T细胞、从为或约5x10⁶至为或约2.5x10⁸个总CAR⁺T细胞、从为或约5x10⁶至为或约1x10⁸个总CAR⁺T细胞、从为或约5x10⁶至为或约5x10⁷个总CAR⁺T细胞、从为或约5x10⁶至为或约2.5x10⁷个总CAR⁺T细胞、从为或约5x10⁶至为或约1x10⁷个总CAR⁺T细胞、从为或约1x10⁷至为或约5x10⁸个总CAR⁺T细胞、从为或约1x10⁷至为或约2.5x10⁸个总CAR⁺T细胞、从为或约1x10⁷至为或约1x10⁸个总CAR⁺T细胞、从为或约1x10⁷至为或约5x10⁷个总CAR⁺T细胞、从为或约1x10⁷至为或约2.5x10⁷个总CAR⁺T细胞、从为或约2.5x10⁷至为或约5x10⁸个总CAR⁺T细胞、从为或约2.5x10⁷至为或约1x10⁸个总CAR⁺T细胞、从为或约2.5x10⁷至为或约5x10⁷个总CAR⁺T细胞、从为或约5x10⁷至为或约5x10⁸个总CAR⁺T细胞、从为或约5x10⁷至为或约2.5x10⁸个总CAR⁺T细胞、从为或约5x10⁷至为或约1x10⁸个总CAR⁺T细胞、从为或约1x10⁸至为或约5x10⁸个总CAR⁺T细胞、从为或约1x10⁸至为或约2.5x10⁸个总CAR⁺T细胞、从为或约2.5x10⁸至为或约5x10⁸个总CAR⁺T细胞。在一些实施方案中,基因工程化的细胞的剂量包含从或从约2.5x10⁷至为或约1.5x10⁸个总CAR⁺T细胞,如从或从约5x10⁷至或至约1x10⁸个总CAR⁺T细胞。

[0870] 在一些实施方案中,基因工程化的细胞的剂量包含至少或至少约1x10⁵个CAR⁺细胞、至少或至少约2.5x10⁵个CAR⁺细胞、至少或至少约5x10⁵个CAR⁺细胞、至少或至少约1x10⁶个CAR⁺细胞、至少或至少约2.5x10⁶个CAR⁺细胞、至少或至少约5x10⁶个CAR⁺细胞、至少或至少约1x10⁷个CAR⁺细胞、至少或至少约2.5x10⁷个CAR⁺细胞、至少或至少约5x10⁷个CAR⁺细胞、至少或至少约1x10⁸个CAR⁺细胞、至少或至少约1.5x10⁸个CAR⁺细胞、至少或至少约2.5x10⁸个CAR⁺细胞或者至少或至少约5x10⁸个CAR⁺细胞。

[0871] 在一些实施方案中,细胞疗法包括施用一定剂量,所述剂量包含如下数量的细胞:从或从约1x10⁵至或至约5x10⁸个总嵌合受体表达细胞、总T细胞或总外周血单个核细胞(PBMC),从或从约5x10⁵至或至约1x10⁷个总嵌合受体表达细胞、总T细胞或总外周血单个核细胞(PBMC)或者从或从约1x10⁶至或至约1x10⁷个总嵌合受体表达细胞、总T细胞或总外周血单个核细胞(PBMC),每个都包含端值。在一些实施方案中,细胞疗法包含施用一定剂量的细胞,所述剂量包含如下数量的细胞:至少或至少约1x10⁵个总嵌合受体表达细胞、总T细胞或总外周血单个核细胞(PBMC),如至少或至少1x10⁶、至少或至少约1x10⁷、至少或至少约1x10⁸个此类细胞。在一些实施方案中,所述数量是关于CD3⁺或CD8⁺的总数量,在一些情况下也是关于嵌合受体表达(例如CAR⁺)细胞。在一些实施方案中,细胞疗法包含施用一定剂量,所述剂量包含以下数量的细胞:从或从约1x10⁵至或至约5x10⁸个CD3⁺或CD8⁺总T细胞或CD3⁺或CD8⁺嵌合受体表达细胞、从或从约5x10⁵至或至约1x10⁷个CD3⁺或CD8⁺总T细胞或CD3⁺或CD8⁺嵌合受体表达细胞、或从或从约1x10⁶至或至约1x10⁷个CD3⁺或CD8⁺总T细胞或CD3⁺或CD8⁺嵌合受体表达细胞,每个都包含端值。在一些实施方案中,细胞疗法包括施用包含如下数量的细胞的剂量:从或从约1x10⁵至或至约5x10⁸个总CD3⁺/CAR⁺或CD8⁺/CAR⁺细胞、从或从约5x10⁵至或至约1x10⁷个总CD3⁺/CAR⁺或CD8⁺/CAR⁺细胞、或从或从约1x10⁶至或至约1x10⁷个总CD3⁺/CAR⁺或CD8⁺/CAR⁺细胞,每个都包含端值。

[0872] 在一些实施方案中,剂量的T细胞包括CD4⁺T细胞、CD8⁺T细胞或CD4⁺和CD8⁺T细胞。

[0873] 例如,在一些实施方案中,如果受试者是人,则剂量的CD8⁺T细胞(包括在包括CD4⁺和CD8⁺T细胞的剂量中)包括在为或约 1×10^6 与为或约 5×10^8 个之间的总嵌合受体(例如CAR)表达CD8⁺细胞,例如在以下范围内:从为或约 5×10^6 至为或约 1×10^8 个此类细胞,如 1×10^7 、 2.5×10^7 、 5×10^7 、 7.5×10^7 、 1×10^8 、 1.5×10^8 或 5×10^8 个总此类细胞,或任两个前述值之间的范围。在一些实施方案中,向患者施用多个剂量,并且每个剂量或总剂量可以在任何前述值内。在一些实施方案中,细胞的剂量包括施用从或从约 1×10^7 至或至约 0.75×10^8 个总嵌合受体表达CD8⁺T细胞、从或从约 1×10^7 至或至约 5×10^7 个总嵌合受体表达CD8⁺T细胞、从或从约 1×10^7 至或至约 0.25×10^8 个总嵌合受体表达CD8⁺T细胞(每个都包含端值)。在一些实施方案中,细胞的剂量包括施用为或约 1×10^7 、 2.5×10^7 、 5×10^7 、 7.5×10^7 、 1×10^8 、 1.5×10^8 、 2.5×10^8 或 5×10^8 个总嵌合受体表达CD8⁺T细胞。

[0874] 在一些实施方案中,细胞(例如,嵌合受体表达T细胞)的剂量是作为单一剂量施用至受试者,或者在两周、一个月、三个月、六个月、1年或更长的时间段内仅施用一次。在过继细胞疗法的情况下,施用给定“剂量”涵盖施用作为单一组合物和/或单次不间断施用(例如,作为单次注射或连续输注)的给定量或数量的细胞,并且还涵盖在指定时间段中(如在不超过3天中)施用在多种单独组合物或输注中提供的作为分割剂量或作为多种组合物的给定量或数量的细胞。因此,在一些情况下,剂量是指定数量的细胞的单次或连续施用,在单个时间点给予或开始。然而,在一些情况下,剂量在不超过三天的时间段内以多次注射或输注的方式施用,例如每天一次持续三天或两天或者通过在一天的时间内多次输注。

[0875] 因此,在一些方面,剂量的细胞以单一药物组合物施用。在一些实施方案中,剂量的细胞以共同含有所述剂量的细胞的多种组合物施用。

[0876] 在一些实施方案中,术语“分割剂量”是指被分割从而可在超过一天中施用的剂量。这种类型的给药包括在本发明方法中并且被认为是单一剂量。

[0877] 因此,细胞的剂量可以作为分割剂量施用,例如随时间施用的分割剂量。例如,在一些实施方案中,可以在2天中或在3天中将剂量施用至受试者。用于分割给药的示例性方法包括在第一天施用25%的剂量并在第二天施用剩余的75%的剂量。在其他实施方案中,可以在第一天施用33%的剂量,并且在第二天施用剩余的67%。在一些方面,在第一天施用10%的剂量,在第二天施用30%的剂量,并且在第三天施用60%的剂量。在一些实施方案中,分割剂量持续不超过3天。

[0878] 在一些实施方案中,剂量的细胞可以通过施用多种组合物或溶液(如第一和第二,任选地更多)来施用,每种组合物或溶液含有所述剂量的一些细胞。在一些方面,任选地在某一时间段内,分开地或独立地施用各自含有不同细胞群和/或细胞亚型的多个组合物。例如,细胞群体或细胞亚型可以分别包括CD8⁺和CD4⁺T细胞,和/或分别包含富集CD8⁺和CD4⁺的群体,例如CD4⁺和/或CD8⁺T细胞,其各自单独地包括基因工程化以表达嵌合受体的细胞。在一些实施方案中,剂量的施用包括施用第一组合物,其包含CD8⁺T细胞的剂量或CD4⁺T细胞的剂量;以及施用第二组合物,其包含CD4⁺T细胞和CD8⁺T细胞的剂量中的另一者。

[0879] 在一些实施方案中,组合物或剂量的施用(例如,多种细胞组合物的施用)涉及分开施用细胞组合物。在一些方面,分开施用同时或按任何顺序依序进行。在一些实施方案中,剂量包含第一组合物和第二组合物,并且将所述第一组合物和第二组合物相隔从为或

约0至为或约12小时、相隔从为或约0至为或约6小时或相隔从为或约0至为或约2小时施用。在一些实施方案中,开始施用第一组合物和开始施用第二组合物是相隔不超过或不超过约2小时、不超过或不超过约1小时或不超过或不超过约30分钟,相隔不超过或不超过约15分钟、不超过或不超过约10分钟或不超过或不超过约5分钟进行的。在一些实施方案中,开始和/或完成施用第一组合物和完成和/或开始施用第二组合物是相隔不超过或不超过约2小时、不超过或不超过约1小时或不超过或不超过约30分钟,相隔不超过或不超过约15分钟、不超过或不超过约10分钟或不超过或不超过约5分钟进行的。

[0880] 在一些组合物中,第一组合物(例如,剂量的第一组合物)包含CD4+T细胞。在一些组合物中,第一组合物(例如,剂量的第一组合物)包含CD8+T细胞。在一些实施方案中,第一组合物是在第二组合物之前施用。

[0881] 在一些实施方案中,细胞的剂量或组合物包括限定或目标比率的表达嵌合受体的CD4+细胞与表达嵌合受体的CD8+细胞和/或CD4+细胞与CD8+细胞,所述比率任选地为大约1:1,或者在大约1:3与大约3:1之间,如大约1:1。在一些方面,具有目标或所需比率的不同细胞群(例如CD4+:CD8+比率或CAR+CD4+:CAR+CD8+比率,例如1:1)的组合物或剂量的施用涉及施用含有一种群体的细胞组合物,和随后施用包含另一种群体的单独细胞组合物,其中所述施用为或大约为目标或所需比率。在一些方面,定义比率的细胞的剂量或组合物的施用导致改善T细胞疗法的扩增、持久性和/或抗肿瘤活性。

[0882] 在一些实施方案中,受试者接受细胞的多个剂量,例如,两个或更多个剂量或多个连续剂量。在一些实施方案中,将两个剂量施用至受试者。在一些实施方案中,受试者接受连续剂量,例如,在第一剂量后大约4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或21天施用第二剂量。在一些实施方案中,在第一剂量后施用多个连续剂量,使得在施用所述连续剂量后施用另外的一个或多个剂量。在一些方面,在另外的剂量中施用至受试者的细胞数量与第一剂量和/或连续剂量相同或相似。在一些实施方案中,另外的一个或多个剂量大于先前剂量。

[0883] 在一些方面,第一和/或连续剂量的大小基于一个或多个标准来确定,如受试者对先前治疗(例如化学疗法)的反应、受试者的疾病负荷(如肿瘤负担、体积、大小或程度)、转移的范围或类型、分期和/或受试者发生毒性结果(例如CRS、巨噬细胞激活综合征、肿瘤溶解综合征、神经毒性和/或针对所施用的细胞和/或嵌合受体的宿主免疫应答)的可能性或发生率。

[0884] 在一些方面,第一剂量的施用与连续剂量的施用之间的时间为约9至约35天、约14至约28天或15至27天。在一些实施方案中,连续剂量的施用是在施用第一剂量后大于约14天且小于约28天的时间点。在一些方面,第一剂量与连续剂量之间的时间为约21天。在一些实施方案中,在施用连续剂量后施用另外的一个或多个剂量(例如,连续剂量)。在一些方面,在施用先前剂量后至少约14天且小于约28天施用另外的一个或多个连续剂量。在一些实施方案中,在先前剂量后小于约14天(例如,在先前剂量后4、5、6、7、8、9、10、11、12或13天)施用另外的剂量。在一些实施方案中,在先前剂量后小于约14天不施用剂量,和/或在先前剂量后超过约28天不施用剂量。

[0885] 在一些实施方案中,细胞(例如,嵌合受体表达细胞)的剂量包含两个剂量(例如,双剂量),包含T细胞的第一剂量和T细胞的连续剂量,其中第一剂量和第二剂量中的一个或

两个包括施用T细胞的分割剂量。

[0886] 在一些实施方案中,细胞的剂量通常足够大以有效减轻疾病负荷。

[0887] 在一些实施方案中,细胞以所需剂量施用,所述所需剂量在一些方面包括所需剂量或数量的细胞或一种或多种细胞类型和/或所需比率的细胞类型。因此,细胞的剂量在一些实施方案中是基于总细胞数(或每kg体重的数量)以及单独群体或亚型的所需比率,如CD4⁺与CD8⁺的比率。在一些实施方案中,细胞的剂量是基于单独群体中或单独细胞类型的的细胞的所需总数(或每kg体重的细胞数量)。在一些实施方案中,剂量是基于此类特征的组合,如总细胞的所需数量、所需比率和单独群体中细胞的所需总数。

[0888] 在一些实施方案中,以总细胞的所需剂量(如T细胞的所需剂量)的耐受差异或在所述耐受差异内施用细胞的群体或亚型(如CD8⁺和CD4⁺T细胞)。在一些方面,所需剂量是细胞的所需数量或被施用细胞的受试者的每单位体重细胞的所需数量,例如,细胞/kg。在一些方面,所需剂量等于或高于细胞的最小数量或每单位体重细胞的最小数量。在一些方面,在以所需剂量施用的总细胞中,单独群体或亚型以等于或接近所需输出比率(如CD4⁺与CD8⁺比率)存在,例如在这种比率的一定耐受差异或误差内。

[0889] 在一些实施方案中,细胞以所需剂量的一种或多种单独细胞群或亚型的耐受差异或在所述耐受差异之内施用,如所需剂量的CD4⁺细胞和/或所需剂量的CD8⁺细胞。在一些方面,所需剂量是亚型或群体的细胞的所需数量或被施用所述细胞的受试者的每单位体重此类细胞的所需数量,例如细胞/kg。在一些方面,所需剂量等于或高于群体或亚型的细胞的最小数量或每单位体重群体或亚型的细胞的最小数量。

[0890] 因此,在一些实施方案中,剂量是基于总细胞的所需固定剂量和所需比率,和/或基于一种或多种(例如,每一种)单独亚型或亚群的所需固定剂量。因此,在一些实施方案中,剂量是基于T细胞的所需固定或最小剂量和CD4⁺与CD8⁺细胞的所需比率,和/或是基于CD4⁺和/或CD8⁺细胞的所需固定或最小剂量。

[0891] 在一些实施方案中,细胞在多种细胞群或亚型(如CD4⁺和CD8⁺细胞或亚型)的所需输出比率的耐受范围下或耐受范围内施用。在一些方面,所需比率可以是特定比率,或者可以是一系列比率。例如,在一些实施方案中,所需比率(例如,CD4⁺与CD8⁺细胞的比率)是在为或约1:5与为或约5:1之间(或大于约1:5且小于约5:1)、或在为或约1:3与为或约3:1之间(或大于约1:3且小于约3:1),如在为或约2:1与为或约1:5之间(或大于约1:5且小于约2:1,如为或约5:1、4.5:1、4:1、3.5:1、3:1、2.5:1、2:1、1.9:1、1.8:1、1.7:1、1.6:1、1.5:1、1.4:1、1.3:1、1.2:1、1.1:1、1:1、1:1.1、1:1.2、1:1.3、1:1.4、1:1.5、1:1.6、1:1.7、1:1.8、1:1.9、1:2、1:2.5、1:3、1:3.5、1:4、1:4.5或1:5。在一些方面,耐受差异在所需比率的约1%、约2%、约3%、约4%、约5%、约10%、约15%、约20%、约25%、约30%、约35%、约40%、约45%、约50%内,包括这些范围之间的任何值。

[0892] 在特定实施方案中,细胞的数量和/或浓度是指嵌合受体(例如,CAR)表达细胞的数量。在其他实施方案中,细胞的数量和/或浓度是指施用的所有细胞、T细胞或外周血单个核细胞(PBMC)的数量或浓度。

[0893] 在一些方面,剂量的大小是基于一个或多个标准来确定,如受试者对先前治疗(例如,化学疗法)的反应、受试者的疾病负荷(如肿瘤负担、体积、大小或程度)、转移的程度或类型、分期和/或受试者发生毒性结果(例如,CRS、巨噬细胞激活综合征、肿瘤溶解综合征、

神经毒性和/或针对所施用的细胞和/或嵌合受体的宿主免疫应答)的可能性或发生率。

[0894] 在一些实施方案中,所述方法还包括施用一个或多个另外的剂量的表达嵌合抗原受体(CAR)的细胞和/或淋巴细胞清除疗法,和/或重复所述方法的一个或多个步骤。在一些实施方案中,一个或多个另外的剂量与初始剂量相同。在一些实施方案中,一个或多个另外的剂量不同于初始剂量,例如更高,如是初始剂量的2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍或10倍高或更高;或者更低,如是初始剂量的2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍或10倍低或更低。在一些实施方案中,一个或多个另外的剂量的施用是基于以下来确定:受试者对初始治疗或任何先前治疗的反应、受试者的疾病负荷(如肿瘤负担、体积、大小或程度)、转移的程度或类型、分期和/或受试者发生毒性结果(例如,CRS、巨噬细胞激活综合征、肿瘤溶解综合征、神经毒性和/或针对所施用的细胞的宿主免疫应答)的可能性或发生率。

V. 药物组合物和配制品

[0895] 还提供用于施用(如用于过继细胞疗法)的组合物(如药物组合物)和配制品。在一些方面,药物组合物含有任何本文所述的工程化细胞或含有工程化细胞的组合物,例如,其包含含有编码嵌合受体的一部分的转基因序列的修饰的CD247基因座。在一些实施方案中,包含用嵌合受体(例如CAR)工程化的细胞的细胞剂量是作为组合物或配制品(如药物组合物或配制品)来提供。此类组合物可以根据所提供的方法来使用和/或与所提供的制品或组合物一起使用,如用于预防或治疗疾病、病症和障碍,或者用于检测、诊断和预后方法中。

[0896] 术语“药物配制品”是指如下制剂,其处于使得其中所含活性成分的生物活性有效的形式,并且不含对施用配制品的受试者具有不可接受的毒性的另外的组分。

[0897] “药学上可接受的载体”是指药物配制品中除了活性成分之外对受试者无毒的成分。药学上可接受的载体包括但不限于缓冲液、赋形剂、稳定剂或防腐剂。

[0898] 在一些方面,载体的选择部分地由特定细胞或药剂和/或通过施用方法确定。因此,存在多种合适的配制品。例如,药物组合物可以含有防腐剂。合适的防腐剂可以包括例如对羟基苯甲酸甲酯、对羟基苯甲酸丙酯、苯甲酸钠和苯扎氯铵。在一些方面,使用两种或更多种防腐剂的混合物。防腐剂或其混合物通常以总组合物重量的约0.0001%至约2%的量存在。载体描述于例如Remington's Pharmaceutical Sciences第16版,Osol, A. 编辑(1980)中。药学上可接受的载体在所用剂量和浓度下通常对接受者无毒,并且包括但不限于:缓冲液,如磷酸盐、柠檬酸盐和其他有机酸;抗氧化剂,包括抗坏血酸和甲硫氨酸;防腐剂(如十八烷基二甲基苄基氯化铵;六甲氯铵;苯扎氯铵;苄索氯铵;苯酚、丁醇或苄醇;烷基对羟基苯甲酸酯,如对羟基苯甲酸甲酯或对羟基苯甲酸丙酯;儿茶酚;间苯二酚;环己醇;3-戊醇;和间甲酚);低分子量(少于约10个残基)多肽;蛋白质,如血清白蛋白、明胶或免疫球蛋白;亲水性聚合物,如聚乙烯吡咯烷酮;氨基酸,如甘氨酸、谷氨酰胺、天冬酰胺、组氨酸、精氨酸或赖氨酸;单糖、二糖和其他碳水化合物,包括葡萄糖、甘露糖或糊精;螯合剂,如EDTA;糖类,如蔗糖、甘露醇、海藻糖或山梨醇;成盐抗衡离子,如钠;金属络合物(例如锌-蛋白质络合物);和/或非离子表面活性剂,如聚乙二醇(PEG)。

[0899] 在一些方面,在组合物中包括缓冲剂。合适的缓冲剂包括例如柠檬酸、柠檬酸钠、磷酸、磷酸钾以及各种其他酸和盐。在一些方面,使用两种或更多种缓冲剂的混合物。缓冲剂或其混合物通常以按总组合物的重量计约0.001%至约4%的量存在。用于制备可施用的药物组合物的方法是已知的。示例性方法更详细地描述于例如Remington: The Science

and Practice of Pharmacy, Lippincott Williams&Wilkins; 第21版(2005年5月1日)中。

[0900] 配制品或组合物还可含有多于一种活性成分,其可用于用细胞或药剂预防或治疗的特定适应症、疾病或病症,其中各自的活性不会相互产生不利影响。此类活性成分以对既定目的有效的量以合适的方式组合存在。因此,在一些实施方案中,药物组合物进一步包含其他药物活性剂或药物如化学治疗剂,例如天冬酰胺酶、白消安、卡铂、顺铂、柔红霉素、多柔比星、氟尿嘧啶、吉西他滨、羟基脲、甲氨蝶呤、紫杉醇、利妥昔单抗、长春碱、长春新碱等。在一些实施方案中,所述药剂或细胞以盐(例如药学上可接受的盐)的形式施用。合适的药学上可接受的酸加成盐包括源自无机酸(如盐酸、氢溴酸、磷酸、偏磷酸,硝酸和硫酸)和有机酸(如酒石酸、乙酸、柠檬酸、苹果酸、乳酸、富马酸、苯甲酸、乙醇酸、葡萄糖酸、琥珀酸和芳基磺酸,例如对甲苯磺酸)的那些盐。

[0901] 在一些实施方案中,药物组合物含有有效治疗或预防疾病或病症的量(如治疗有效量或预防有效量)的药剂或细胞。在一些实施方案中,通过定期评估所治疗的受试者来监测治疗或预防功效。对于数天或更长时间的重复施用,取决于病症,重复所述治疗直至出现所需疾病症状的抑制。然而,其他剂量方案可能有用并且可以被确定。所需剂量可以通过单次推注施用所述组合物、通过多次推注施用所述组合物或通过连续输注施用所述组合物来递送。

[0902] 药剂或细胞可以通过任何合适的方式施用,例如通过推注输注,通过注射例如静脉内或皮下注射、眼内注射、眼周注射、视网膜下注射、玻璃体内注射、经中隔注射、巩膜下注射、脉络膜内注射、前房注射、结膜下(subconjunctival)注射、结膜下(subconjunctival)注射、眼球筋膜囊下(sub-Tenon)注射、眼球后注射、眼球周注射或后近巩膜(posterior juxtасcleral)递送。在一些实施方案中,将它们通过肠胃外、肺内和鼻内以及(如果需要用于局部治疗的话)病灶内施用来施用。肠胃外输注包括肌内、静脉内、动脉内、腹膜内或皮下施用。在一些实施方案中,给定剂量通过细胞或药剂的单次推注给药来施用。在一些实施方案中,其是通过例如在不超过3天的时间段内对细胞或药剂的多次推注施用或通过细胞或药剂的连续输注施用来施用。

[0903] 对于疾病的预防或治疗,适当的剂量可以取决于待治疗的疾病类型、一种或多种药剂的类型、细胞或嵌合受体的类型、疾病的严重程度和病程、施用药剂或细胞是用于预防目的还是治疗目的、先前疗法、受试者的临床病史和对药剂或细胞的反应以及主治医师的决断。在一些实施方案中,将组合物一次性或在一系列治疗中合适地施用至受试者。

[0904] 可以使用标准施用技术、配制品和/或设备施用细胞或药剂。提供了用于储存和施用组合物的配制品和装置(如注射器和小瓶)。关于细胞,施用可以是自体的或异源的。在一些方面,所述细胞从受试者分离,工程化,并施用至同一受试者。在其他方面,所述细胞从一名受试者分离,工程化,并施用至另一受试者。例如,免疫应答细胞或祖细胞可以获得自一名受试者,并且施用至同一受试者或不同的相容受试者。源自外周血的免疫应答细胞或其后代(例如,体内、离体或体外衍生的)可以经由局部注射施用,包括导管施用、全身注射、局部注射、静脉内注射或肠胃外施用。当施用治疗性组合物(例如,含有基因修饰的免疫反应性细胞或治疗或改善神经毒性症状的药剂的药物组合物)时,通常将其配制成单位剂量可注射形式(溶液、悬浮液、乳液)。

[0905] 配制品包括用于口服、静脉内、腹膜内、皮下、经肺、透皮、肌内、鼻内、经颊、舌下或

栓剂施用的那些。在一些实施方案中,肠胃外施用药剂或细胞群。如本文所用术语“肠胃外”包括静脉内、肌内、皮下、直肠、阴道和腹膜内施用。在一些实施方案中,使用通过静脉内、腹膜内或皮下注射的外周全身递送向受试者施用药剂或细胞群。

[0906] 在一些实施方案中,组合物是作为无菌液体制剂(例如,等渗水性溶液、悬浮液、乳液、分散液或粘性组合物)来提供,在一些方面可以将其缓冲至所选pH。液体制剂一般比凝胶、其他粘性组合物和固体组合物更易于制备。另外地,液体组合物稍微更方便施用,特别是通过注射。在另一方面,粘性组合物可以配制在适当的粘度范围内,以提供与特定组织的更长的接触时间。液体或粘性组合物可以包含载体,所述载体可以是溶剂或分散介质,所述溶剂或分散介质含有例如水、盐水、磷酸盐缓冲盐水、多元醇(例如,甘油、丙二醇、液体聚乙二醇)及其合适的混合物。

[0907] 无菌可注射溶液可以通过以下方式来制备:将所述药剂或细胞掺入溶剂中,如掺入与合适的载体、稀释剂或赋形剂(如无菌水、生理盐水、葡萄糖、右旋糖等)的混合物中。

[0908] 用于体内施用的配制品通常是无菌的。可以例如通过经无菌滤膜过滤容易地实现无菌性。

VI. 试剂盒和制品

[0909] 还提供可用于进行所提供的实施方案的制品、系统、装置和试剂盒。在一些实施方案中,所提供的制品或试剂盒含有能够诱导遗传破坏的一种或多种药剂的一种或多种组分和/或一种或多种模板多核苷酸(例如,含有编码嵌合受体或其部分的转基因序列的模板多核苷酸)。在一些实施方案中,制品或试剂盒可以用于工程化T细胞以表达嵌合受体和/或其他分子的方法中,如通过经由同源依赖修复(HDR)整合编码嵌合受体或其部分的转基因序列,例如,以生成包含含有编码嵌合受体的核酸序列的修饰的CD247基因座的工程化细胞,所述嵌合受体包含含有CD3zeta(CD3 ζ)信号传导结构域的细胞内区域。

[0910] 在一些实施方案中,制品或试剂盒包括可用于进行所提供的方法的多肽、核酸、载体和/或多核苷酸。在一些实施方案中,制品或试剂盒包括能够例如在CD247基因座处诱导遗传破坏的一种或多种药剂(如本文在章节I.A中所述的那些)。在一些实施方案中,制品或试剂盒包括一种或多种核酸分子(例如,质粒或DNA片段),所述核酸分子编码能够诱导遗传破坏的一种或多种药剂的一种或多种组分和/或包含一种或多种模板多核苷酸(例如,用于经由HDR将转基因序列靶向至细胞中,如本文在章节I.B.2中所述的那些)。在一些实施方案中,本文提供的制品或试剂盒含有对照载体。

[0911] 在一些实施方案中,本文提供的制品或试剂盒含有一种或多种药剂,其中所述一种或多种药剂中的每一种能够独立地诱导CD247基因座内的靶位点的遗传破坏;以及模板多核苷酸,所述模板多核苷酸包含编码嵌合受体的一部分的转基因,其中所述转基因被靶向以供经由同源定向修复(HDR)整合于所述靶位点处或附近。在一些方面,能够诱导遗传破坏的一种或多种药剂是本文所述的任一种。在一些方面,一种或多种药剂是包含Cas9/gRNA复合物的核糖核蛋白(RNP)复合物。在一些方面,RNP中包括的gRNA靶向CD247基因座中的靶位点,如本文所述的任何靶位点。在一些方面,模板多核苷酸是本文所述的任何模板多核苷酸。

[0912] 在一些实施方案中,制品或试剂盒包括一个或多个容器(通常是多个容器)、包装材料和位于所述一个或多个容器和/或包装上或与所述容器和/或包装相连的标签或包装

说明书,所述标签或包装说明书通常包括使用说明书,例如,用于将组分引入细胞中以供工程化的说明书。

[0913] 本文提供的制品含有包装材料。用于包装所提供材料的包装材料是熟知的。参见例如,美国专利号5,323,907、5,052,558和5,033,252,其各自以其整体并入本文。包装材料的例子包括但不限于泡罩包装、瓶子、管、吸入器、泵、袋子、小瓶、容器、注射器、一次性实验室用品(例如,移液管尖端和/或塑料板)或瓶子。制品或试剂盒可以包括装置,以便有助于分配材料或有助于以高通量或大规模方式使用,例如以有助于在机器人设备中使用。通常,包装对其中所含的组合物无反应性。

[0914] 在一些实施方案中,能够诱导遗传破坏的一种或多种药剂和/或一种或多种模板多核苷酸是分开包装的。在一些实施方案中,每个容器可以具有单一区室。在一些实施方案中,制品或试剂盒的其他组分是分开包装的,或者一起包装于单一区室中。

[0915] 还提供用于施用例如用于疗法或治疗的所提供细胞和/或细胞组合物的制品、系统、设备和试剂盒。在一些实施方案中,本文提供的制品或试剂盒含有T细胞和/或T细胞组合物,如本文所述的任何T细胞和/或T细胞组合物。在一些方面,本文提供的制品或试剂盒可以用于施用T细胞或T细胞组合物,并且可以包括使用说明书。

[0916] 在一些实施方案中,本文提供的制品或试剂盒含有T细胞和/或T细胞组合物,如本文所述的任何T细胞和/或T细胞组合物。在一些实施方案中,T细胞和/或T细胞组合物任何修饰的T细胞使用本文所述的筛选方法。在一些实施方案中,本文提供的制品或试剂盒含有对照或未修饰的T细胞和/或T细胞组合物。在一些实施方案中,制品或试剂盒包括用于施用用于疗法的工程化细胞和/或细胞组合物的一个或多个说明书。

[0917] 含有用于疗法的细胞或细胞组合物的制品和/或试剂盒可以包括容器以及所述容器上或与所述容器相连的标签或包装说明书。合适的容器包括例如瓶子、小瓶、注射器、IV溶液袋等。容器可以由各种材料(例如玻璃或塑料)形成。在一些实施方案中,容器容纳组合物,所述组合物自身或与另一种组合物组合地有效治疗、预防和/或诊断病症。在一些实施方案中,容器具有无菌接入端口。示例性容器包括静脉内溶液袋、小瓶(包括具有可被注射针刺穿的塞子的那些溶液袋、小瓶)或用于口服施用剂的瓶子或小瓶。标签或包装说明书可以指示将组合物用于治疗疾病或病症。制品可以包括(a)其中含有组合物的第一容器,其中所述组合物包括表达嵌合受体的工程化细胞;和(b)其中含有组合物的第二容器,其中所述组合物包括第二药剂。在一些实施方案中,制品可以包括(a)其中含有第一组合物的第一容器,其中所述组合物包括表达嵌合受体的工程化细胞的亚型;和(b)其中含有组合物的第二容器,其中所述组合物包括表达嵌合受体的工程化细胞的不同亚型。制品还可以包括包装说明书,所述包装说明书指示组合物可以用于治疗特定病症。可替代地或另外地,制品还可以包括另一种或相同的容器,所述另一种或相同的容器包含药学上可接受的缓冲液。所述容器还可以包括其他材料,如其他缓冲剂、稀释剂、过滤器、针和/或注射器。

VII. 定义

[0918] 除非另有定义,否则本文使用的所有技术术语、符号以及其他技术和科学术语或用辞意图具有与所要求保护的主体所属领域的普通技术人员通常理解的含义相同的含义。在一些情形中,为了清楚和/或为了便于引用而在本文中定义具有通常理解的含义的术语,并且本文中包含此类定义不一定应当被解释为代表与本领域通常所理解的含义有实质性

差异。

[0919] 除非上下文另有明确规定,否则如本文所用,单数形式“一个/一种(a)”、“一个/一种(an)”和“所述(the)”包括复数指代物。例如,“一个/一种(a)”或“一个/一种(an)”意指“至少一个/至少一种”或“一个或多个/一种或多种”。应理解,本文描述的方面和变化包括“由方面和变化组成”和/或“基本上由方面和变化组成”。

[0920] 在整个本公开文本中,要求保护的主题的各个方面以范围形式呈现。应当理解,范围形式的描述仅仅是为了方便和简洁,并且不应被解释为对所要求保护的主题的范围的僵硬限制。因此,应当将范围的描述视为已明确公开所有可能的子范围以及该范围内的单独数值。例如,在提供值的范围的情况下,应理解,在该范围的上限和下限之间的每个中间值以及在所陈述范围内的任何其他所述值或中间值都涵盖在所要求保护的主题内。这些较小范围的上限和下限可以独立地被包括在所述较小范围内,并且也涵盖在所要求保护的主题内,受制于在所陈述范围内任何确切排除的限制。在所陈述范围包括所述限值中的之一或两者的情况下,排除那些所包括限值中的任何一个或两个的范围也被包括在所要求保护的主题内。无论范围的宽度如何,这都适用。

[0921] 如本文所用术语“约”是指容易知道的相应值的通常误差范围。本文提及“约”一值或参数时包括(并描述)涉及该值或参数本身的实施方案。例如,提及“约X”的描述包括“X”的描述。在一些实施方案中,“约”可以指代 $\pm 25\%$ 、 $\pm 20\%$ 、 $\pm 15\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 或 $\pm 1\%$ 。

[0922] 如本文所用,叙述核苷酸或氨基酸位置“对应于”所公开序列中的核苷酸或氨基酸位置(例如序列中所述)是指,在使用标准比对算法(例如GAP算法)与所公开序列比对以使同一性最大化之后,所鉴定的核苷酸或氨基酸位置。通过比对序列,可以例如使用保守且相同的氨基酸残基作为指导来鉴定相应残基。通常,为了鉴定对应位置,比对氨基酸序列使得获得最高阶匹配(参见例如:Computational Molecular Biology,Lesk,A.M.编辑,Oxford University Press,纽约,1988;Biocomputing:Informatics and Genome Projects,Smith,D.W.编辑,Academic Press,纽约,1993;Computer Analysis of Sequence Data,Part I,Griffin,A.M.和Griffin,H.G.编辑,Humana Press,新泽西州,1994;Sequence Analysis in Molecular Biology,von Heinje,G.,Academic Press,1987;和Sequence Analysis Primer,Gribskov,M.和Devereux,J.编辑,M Stockton Press,纽约,1991;Carrillo等人(1988)SIAM J Applied Math 48:1073)。

[0923] 如本文所用,术语“载体”是指能够传播与其连接的另一核酸的核酸分子。所述术语包括作为自我复制核酸结构的载体以及掺入已将其引入的宿主细胞基因组中的载体。某些载体能够引导与它们操作性地连接的核酸的表达。此类载体在本文中称为“表达载体”。载体包括病毒载体,例如逆转录病毒(例如 γ 逆转录病毒和慢病毒)载体。

[0924] 术语“宿主细胞”、“宿主细胞系”和“宿主细胞培养物”可互换使用,并且是指已引入外源核酸的细胞,包括此类细胞的后代。宿主细胞包括“转化体”和“转化细胞”,其包括原代转化细胞和由其衍生的后代,不考虑传代次数。后代在核酸含量上可能与亲代细胞不完全相同,但可能含有突变。本文包括如在初始转化细胞中所筛选或选择,具有相同功能或生物活性的突变体后代。

[0925] 如本文所用,细胞或细胞群体针对特定标记物呈“阳性”的陈述是指,特定标记物(通常是表面标记物)在细胞上或细胞中的可检测存在。当提及表面标记物时,所述术语是

指如通过流式细胞术检测到的,表面表达的存在,例如通过用与所述标记物特异性地结合的抗体进行染色并检测所述抗体,其中所述染色通过流式细胞术以如下水平是可检测的,所述水平基本上高于在其他方面相同的条件下用同种型匹配对照进行相同程序检测到的染色,和/或所述水平与已知对所述标记物呈阳性的细胞的水平基本上相似,和/或所述水平基本上高于已知对所述标记物呈阴性的细胞的水平。

[0926] 如本文所用,细胞或细胞群体对特定标记物呈“阴性”的陈述是指,特定标记物(通常是表面标记物)在细胞上或细胞中不存在基本上可检测的存在。当提及表面标记物时,所述术语是指如通过流式细胞术检测到的,表面表达的不存在,例如通过用与所述标记物特异性地结合的抗体进行染色并检测所述抗体,其中所述染色通过流式细胞术以如下水平没有检测到,所述水平基本上高于在其他方面相同的条件下用同种型匹配对照进行相同程序检测到的染色,和/或所述水平基本上低于已知对所述标记物呈阳性的细胞的水平,和/或所述水平与已知对所述标记物呈阴性的细胞的水平相比是基本上相似的。

[0927] 如本文所用,在关于氨基酸序列(参考多肽序列)使用时,“氨基酸序列同一性百分比(%)”和“同一性百分比”定义为,在比对序列并在必要时引入空位以实现最大序列同一性百分比并且不将任何保守取代视作序列同一性的一部分后,候选序列(例如,主题抗体或片段)中与参考多肽序列中的氨基酸残基相同的氨基酸残基的百分比。用于确定氨基酸序列同一性百分比的目的的比对可以以各种已知方式来实现,在一些实施方案中,使用公众可获得的计算机软件,如BLAST、BLAST-2、ALIGN或Megalign (DNASTAR) 软件。可以确定用于比对序列的适当参数,包括在所比较的序列的全长上实现最大比对所需的任何算法。

[0928] 在一些实施方案中,“可操作地连接”可以包括组分(如DNA序列,例如异源核酸)与一个或多个调节序列的缔合,所述缔合的方式允许在适当分子(例如转录激活蛋白)与所述调节序列结合时进行基因表达。因此,它意味着,所述组分所处的关系允许它们以其预期方式起作用。

[0929] 氨基酸取代可以包括用一个氨基酸替代多肽中的另一个氨基酸。取代可以是保守氨基酸取代或非保守氨基酸取代。可以将氨基酸取代引入感兴趣的结合分子(例如抗体)、和针对所希望的活性(例如保留/改进的抗原结合、降低的免疫原性或改进的ADCC或CDC)筛选的产物中。

[0930] 通常可以根据以下常见的侧链特性将氨基酸进行分组:

- (1) 疏水性: 正亮氨酸、Met、Ala、Val、Leu、Ile;
- (2) 中性亲水性: Cys、Ser、Thr、Asn、Gln;
- (3) 酸性: Asp、Glu;
- (4) 碱性: His、Lys、Arg;
- (5) 影响链取向的残基: Gly、Pro;
- (6) 芳香族: Trp、Tyr、Phe。

[0931] 在一些实施方案中,保守取代可能包括将这些类别之一的成员交换为同一类别的另一个成员。在一些实施方案中,非保守氨基酸取代可能涉及将这些类别之一的成员交换为另一个类别。

[0932] 如本文所用,组合物是指两种或更多种产物、物质或化合物(包括细胞)的任何混合物。它可以是溶液、悬浮液、液体、粉末、糊剂、水性的、非水性的或其任何组合。

[0933] 如本文所用,“受试者”是哺乳动物,如人或其他动物,并且通常是人。

VIII. 示例性实施方案

[0934] 所提供的实施方案包括:

1. 一种基因工程化的T细胞,其包含修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。

2. 根据实施方案1所述的基因工程化的T细胞,其中所述核酸序列包含编码所述嵌合受体的一部分的转基因序列,所述转基因序列已经任选地经由同源定向修复 (HDR) 整合于所述内源CD247基因座处。

3. 根据实施方案1或实施方案2所述的基因工程化的T细胞,其中所述CD3 ζ 信号传导结构域的全部或片段是由所述内源CD247基因座的开放阅读框或其部分序列编码。

4. 根据实施方案1-3中任一项所述的基因工程化的T细胞,其中所述核酸序列包含以下的框内融合物:(i) 编码所述嵌合受体的一部分的转基因序列,和(ii) 所述内源CD247基因座的开放阅读框或其部分序列。

5. 一种基因工程化的T细胞,其包含修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域,其中所述核酸序列包含以下的框内融合物:(i) 编码所述嵌合受体的一部分的转基因序列,和(ii) 编码所述CD3 ζ 信号传导结构域的内源CD247基因座的开放阅读框或其部分序列。

6. 根据实施方案2、3和5中任一项所述的基因工程化的T细胞,其中所述转基因序列与所述内源CD247基因座的所述开放阅读框或其部分序列的一个或多个外显子同框。

7. 根据实施方案2-6中任一项所述的基因工程化的T细胞,其中所述转基因序列不包含编码3'UTR的序列。

8. 根据实施方案2-7中任一项所述的基因工程化的T细胞,其中所述转基因序列不包含内含子。

9. 根据实施方案2-8中任一项所述的基因工程化的T细胞,其中所述转基因序列编码所述CD3 ζ 信号传导结构域的片段。

10. 根据实施方案2-8中任一项所述的基因工程化的T细胞,其中所述转基因序列不编码所述CD3 ζ 信号传导结构域或其片段。

11. 根据实施方案3-10中任一项所述的基因工程化的T细胞,其中所述开放阅读框或其部分序列包含所述内源CD247基因座的至少一个内含子和至少一个外显子。

12. 根据实施方案3-11中任一项所述的基因工程化的T细胞,其中所述开放阅读框或其部分序列编码所述内源CD247基因座的3'UTR。

13. 根据实施方案2-12中任一项所述的基因工程化的T细胞,其中所述转基因序列位于所述内源CD247基因座的所述开放阅读框的外显子1的下游和外显子8的上游。

14. 根据实施方案2-13中任一项所述的基因工程化的T细胞,其中所述转基因序列位于所述内源CD247基因座的所述开放阅读框的外显子1的下游和外显子3的上游。

15. 根据实施方案3-14中任一项所述的基因工程化的T细胞,其中所编码的嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段,任选地整个CD3 ζ 信号传导结构域,是由所

述内源CD247基因座的所述开放阅读框或其部分序列编码。

16. 根据实施方案1-15中任一项所述的基因工程化的T细胞,其中所述CD3 ζ 信号传导结构域是由包含所述内源CD247基因座的所述开放阅读框的外显子2和外显子3-8的至少一部分的核苷酸序列编码。

17. 根据实施方案1-16中任一项所述的基因工程化的T细胞,其中所述CD3 ζ 信号传导结构域是由不包含所述内源CD247基因座的所述开放阅读框的外显子1、不包含外显子1的全长和/或不包含外显子2的全长的核苷酸序列编码。

18. 根据实施方案1-17中任一项所述的基因工程化的T细胞,其中所编码的嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

19. 根据实施方案1-18中任一项所述的基因工程化的T细胞,其中所编码的CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。

20. 根据实施方案1-19中任一项所述的基因工程化的T细胞,其中所述嵌合受体是或包含功能性非T细胞受体(非TCR)抗原受体。

21. 根据实施方案1-20中任一项所述的基因工程化的T细胞,其中所述嵌合受体是嵌合抗原受体(CAR)。

22. 根据实施方案1-21中任一项所述的基因工程化的T细胞,其中所述嵌合受体还包含细胞外区域和/或跨膜结构域。

23. 根据实施方案2-22中任一项所述的基因工程化的T细胞,其中所述转基因序列包含编码所述嵌合受体的一个或多个区域的核苷酸序列,任选地其中所述转基因序列包含编码细胞外区域、跨膜结构域和/或所述细胞内区域的一部分中的一个或多个的核苷酸序列。

24. 根据实施方案23所述的基因工程化的T细胞,其中所述细胞外区域包含结合结构域。

25. 根据实施方案24所述的基因工程化的T细胞,其中所述结合结构域是或包含抗体或其抗原结合片段。

26. 根据实施方案24或实施方案25所述的基因工程化的T细胞,其中所述结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达。

27. 根据实施方案26所述的基因工程化的T细胞,其中所述靶抗原是肿瘤抗原。

28. 根据实施方案26或实施方案27所述的基因工程化的T细胞,其中所述靶抗原选自 α v β 6整合素(α v β 6整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPha2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆

碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、神经节苷脂GD3、糖蛋白100(gp100)、磷脂酰肌醇蛋白聚糖-3(GPC3)、G蛋白偶联受体C类5族成员D(GPRC5D)、Her2/neu(受体酪氨酸激酶erb-B2)、Her3(erb-B3)、Her4(erb-B4)、erbB二聚体、人高分子量黑色素瘤相关抗原(HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1(HLA-A1)、人白细胞抗原A2(HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2(IL-13R α 2)、激酶插入结构域受体(kdr)、 κ 轻链、L1细胞粘附分子(L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A(LRRC8A)、Lewis Y、黑色素瘤相关抗原(MAGE)-A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素(MSLN)、c-Met、鼠类巨细胞病毒(CMV)、粘蛋白1(MUC1)、MUC16、自然杀伤细胞2族成员D(NKG2D)配体、黑色素A(MART-1)、神经细胞粘附分子(NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原(PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原(PSCA)、前列腺特异性膜抗原(PSMA)、受体酪氨酸激酶样孤儿受体1(ROR1)、存活蛋白、滋养层糖蛋白(TPBG,也称为5T4)、肿瘤相关糖蛋白72(TAG72)、酪氨酸酶相关蛋白1(TRP1,也称为TYRP1或gp75)、酪氨酸酶相关蛋白2(TRP2,也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体(VEGFR)、血管内皮生长因子受体2(VEGFR2)、肾母细胞瘤1(WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。

29. 根据实施方案22-27中任一项所述的基因工程化的T细胞,其中所述细胞外区域包含间隔子,任选地其中所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。

30. 根据实施方案29所述的基因工程化的T细胞,其中所述间隔子包含免疫球蛋白铰链区。

31. 根据实施方案29或实施方案30所述的基因工程化的T细胞,其中所述间隔子包含C_H2区域和C_H3区域。

32. 根据实施方案23-31中任一项所述的基因工程化的T细胞,其中所述细胞内区域的所述部分包含一个或多个共刺激信号传导结构域。

33. 根据实施方案32所述的基因工程化的T细胞,其中所述一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。

34. 根据实施方案32或实施方案33所述的基因工程化的T细胞,其中所述一个或多个共刺激信号传导结构域包含4-1BB的细胞内信号传导结构域。

35. 根据实施方案24-34中任一项所述的基因工程化的T细胞,其中所述修饰的CD247基因座编码嵌合受体,所述嵌合受体从其N末端至C末端按顺序包含:所述细胞外结合结构域、所述间隔子、所述跨膜结构域和细胞内信号传导区域。

36. 根据实施方案1-35中任一项所述的基因工程化的T细胞,其中:

所述转基因序列按顺序包含:编码细胞外结合结构域、任选地scFv的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;任选地来自人4-1BB的共刺激信号传导结构域;和/或

所述修饰的CD247基因座按顺序包含:编码细胞外结合结构域、任选地scFv的核苷酸序列;间隔子,所述间隔子任选地包含来自人免疫球蛋白铰链的序列,任选地来自IgG1、

IgG2或IgG4或其修饰形式的序列,任选地还包含C_H2区域和/或C_H3区域;以及任选地来自人CD28的跨膜结构域;任选地来自人4-1BB的共刺激信号传导结构域;以及CD3ζ信号传导结构域。

37. 根据实施方案21-36中任一项所述的基因工程化的T细胞,其中所述CAR是多链CAR。

38. 根据实施方案2-37中任一项所述的基因工程化的T细胞,其中所述转基因序列包含编码至少一种其他蛋白质的核苷酸序列。

39. 根据实施方案2-38中任一项所述的基因工程化的T细胞,其中所述转基因序列包含一个或多个多顺反子元件。

40. 根据实施方案39所述的基因工程化的T细胞,其中所述一个或多个多顺反子元件定位于编码所述嵌合受体的一部分的核苷酸序列与编码所述至少一种其他蛋白质的核苷酸序列之间。

41. 根据实施方案38-40中任一项所述的基因工程化的T细胞,其中所述至少一种其他蛋白质是替代标记,任选地其中所述替代标记是截短的受体,任选地其中所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。

42. 根据实施方案39所述的基因工程化的T细胞,其中所述嵌合受体是多链CAR,并且多顺反子元件定位于编码所述多链CAR的一条链的核苷酸序列与编码所述多链CAR的另一条链的核苷酸序列之间。

43. 根据实施方案39-42中任一项所述的基因工程化的T细胞,其中所述一个或多个多顺反子元件位于编码所述嵌合受体的一部分的核苷酸序列的上游。

44. 根据实施方案39-43中任一项所述的基因工程化的T细胞,其中所述一个或多个多顺反子元件是或包含核糖体跳跃序列,任选地其中所述核糖体跳跃序列是T2A、P2A、E2A或F2A元件。

45. 根据实施方案1-44中任一项所述的基因工程化的T细胞,其中所述修饰的CD247基因座包含所述内源CD247基因座的启动子和/或调节或控制元件,所述启动子和/或调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。

46. 根据实施方案1-45中任一项所述的基因工程化的T细胞,其中所述修饰的基因座包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。

47. 根据实施方案46所述的基因工程化的T细胞,其中所述一个或多个异源调节或控制元件包括启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、剪接接纳体序列和/或剪接供体序列。

48. 根据实施方案47所述的基因工程化的T细胞,其中所述异源启动子是或包含人延伸因子1α(EF1α)启动子或MND启动子或其变体。

49. 根据实施方案1-48中任一项所述的基因工程化的T细胞,其中所述T细胞是源自受试者的原代T细胞,任选地其中所述受试者是人。

50. 根据实施方案1-49中任一项所述的基因工程化的T细胞,其中所述T细胞是CD8⁺T细胞或其亚型。

51. 根据实施方案1-49中任一项所述的基因工程化的T细胞,其中所述T细胞是CD4

+T细胞或其亚型。

52. 根据实施方案1-51中任一项所述的基因工程化的T细胞,其中所述T细胞源自多潜能或多能细胞,其任选地是iPSC。

53. 一种多核苷酸,其包含:

(a) 编码嵌合受体或其部分的核酸序列;以及

(b) 与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框或其部分序列的一个或多个区域同源的序列。

54. 根据实施方案53所述的多核苷酸,其中所述嵌合受体包含细胞内区域,并且(a)的所述核酸序列是编码所述嵌合受体的一部分的核酸序列,所述部分不包括所述嵌合受体的完整细胞内区域。

55. 一种多核苷酸,其包含:

(a) 编码嵌合受体的部分的核酸序列,所述嵌合受体包含细胞内区域,其中所述嵌合受体的所述部分包括所述嵌合受体的不完整细胞内区域;以及

(b) 与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框或其部分序列的一个或多个区域同源的序列。

56. 根据实施方案54或实施方案55所述的多核苷酸,其中所述嵌合受体的完整细胞内区域包含CD3zeta (CD3 ζ) 信号传导结构域或其片段,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述细胞内区域的至少一部分是由所述内源CD247基因座的开放阅读框或其部分序列编码。

57. 根据实施方案53-56中任一项所述的多核苷酸,其中编码所述嵌合受体的所述部分的核酸序列和所述一个或多个同源臂一起包含编码所述嵌合受体的细胞内区域的核苷酸序列的至少一个片段,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,所述细胞内区域的至少一部分包含由所述CD247基因座的所述开放阅读框或其部分序列编码的CD3 ζ 信号传导结构域或其片段。

58. 根据实施方案53-57中任一项所述的多核苷酸,其中(a)的所述核酸序列不包含编码3'UTR的序列。

59. 根据实施方案53-58中任一项所述的多核苷酸,其中(a)的所述核酸序列不包含内含子。

60. 根据实施方案53-59中任一项所述的多核苷酸,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,(a)的所述核酸序列编码所述CD3 ζ 信号传导结构域的片段。

61. 根据实施方案53-59中任一项所述的多核苷酸,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,(a)的所述核酸序列不编码所述CD3 ζ 信号传导结构域或其片段。

62. 根据实施方案53-61中任一项所述的多核苷酸,其中所述开放阅读框或其部分序列包含所述内源CD247基因座的至少一个内含子和至少一个外显子。

63. 根据实施方案53-62中任一项所述的多核苷酸,其中所述开放阅读框或其部分序列编码所述内源CD247基因座的3'UTR。

64. 根据实施方案53-63中任一项所述的多核苷酸,其中在所述嵌合受体由引入所述多核苷酸的细胞表达时,所编码的嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段、任选地整个CD3 ζ 信号传导结构域,是由所述内源CD247基因座的所述开放阅读框或其部

分序列编码。

65. 根据实施方案53-64中任一项所述的多核苷酸,其中(a)的所述核酸序列是对于T细胞、任选地人T细胞的所述内源基因组CD247基因座的开放阅读框为外源或异源的序列。

66. 根据实施方案53-65中任一项所述的多核苷酸,其中(a)的所述核酸序列包含与所述一个或多个同源臂中包含的所述CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框的核苷酸序列。

67. 根据实施方案53-66中任一项所述的多核苷酸,其中所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列位于所述CD247基因座的开放阅读框的外显子8的上游。

68. 根据实施方案53-67中任一项所述的多核苷酸,其中所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列位于所述CD247基因座的开放阅读框的外显子3的上游,任选地包括所述CD247基因座的开放阅读框的外显子3的序列。

69. 根据实施方案53-68中任一项所述的多核苷酸,其中所述开放阅读框的所述一个或多个区域是或包含如下序列,所述序列包括所述CD247基因座的开放阅读框的外显子2的至少一部分。

70. 根据实施方案53-69中任一项所述的多核苷酸,其中所述一个或多个同源臂不包含所述内源CD247基因座的开放阅读框的外显子1,不包含外显子1的全长和/或不包含外显子2的全长。

71. 根据实施方案53-70中任一项所述的多核苷酸,其中,在由引入所述多核苷酸的细胞表达时,所述嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

72. 根据实施方案53-71中任一项所述的多核苷酸,其中所述CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。

73. 根据实施方案53-72中任一项所述的多核苷酸,其中所述一个或多个同源臂包含5'同源臂和3'同源臂。

74. 根据实施方案53-73中任一项所述的多核苷酸,其中所述多核苷酸包含结构[5'同源臂]-[(a)的核酸序列]-[3'同源臂]。

75. 根据实施方案73或实施方案74所述的多核苷酸,其中所述5'同源臂和所述3'同源臂独立地具有为或约50至为或约2000个核苷酸、为或约100至为或约1000个核苷酸、为或约100至为或约750个核苷酸、为或约100至为或约600个核苷酸、为或约100至为或约400个核苷酸、为或约100至为或约300个核苷酸、为或约100至为或约200个核苷酸、为或约200至为或约1000个核苷酸、为或约200至为或约750个核苷酸、为或约200至为或约600个核苷酸、为或约200至为或约400个核苷酸、为或约200至为或约300个核苷酸、为或约300至为或约1000个核苷酸、为或约300至为或约750个核苷酸、为或约300至为或约600个核苷酸、为或约300至为或约400个核苷酸、为或约400至为或约1000个核苷酸、为或约400至为或约750个核苷酸、为或约400至为或约600个核苷酸、为或约600至为或约1000个核苷酸、为或约600至为或约750个核苷酸或者为或约750至为或约1000个核苷酸的长度。

76. 根据实施方案73-75中任一项所述的多核苷酸,其中所述5'同源臂和所述3'同源臂独立地具有为或约200、300、400、500、600、700或800个核苷酸的长度,或者任何前述值之间的任何值。

77. 根据实施方案73-76中任一项所述的多核苷酸,其中所述5'同源臂和所述3'同源臂独立地具有大于或大于约300个核苷酸的长度,任选地其中所述5'同源臂和所述3'同源臂独立地具有为或约400、500或600个核苷酸的长度,或者任何前述值之间的任何值。

78. 根据实施方案53-77中任一项所述的多核苷酸,其中所述5'同源臂包含SEQ ID NO:80中所示的序列,或者展现与SEQ ID NO:80至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。

79. 根据实施方案53-78中任一项所述的多核苷酸,其中所述3'同源臂包含SEQ ID NO:81中所示的序列,或者展现与SEQ ID NO:81至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。

80. 根据实施方案53-79中任一项所述的多核苷酸,其中所述嵌合受体是或包含功能性非T细胞受体(非TCR)抗原受体。

81. 根据实施方案53-80中任一项所述的多核苷酸,其中所述嵌合受体是嵌合抗原受体(CAR)。

82. 根据实施方案53-81中任一项所述的多核苷酸,其中(a)的所述核酸序列包含编码细胞外区域的核苷酸序列、编码跨膜结构域和/或所述细胞内区域的一部分的核苷酸序列。

83. 根据实施方案82所述的多核苷酸,其中所述细胞外区域包含结合结构域。

84. 根据实施方案83所述的多核苷酸,其中所述结合结构域是或包含抗体或其抗原结合片段。

85. 根据实施方案83或实施方案84所述的多核苷酸,其中所述结合结构域能够与靶抗原结合,所述靶抗原与疾病、障碍或病症相关,为疾病、障碍或病症所特有,和/或在疾病、障碍或病症的细胞或组织上表达。

86. 根据实施方案85所述的多核苷酸,其中所述靶抗原是肿瘤抗原。

87. 根据实施方案85或实施方案86所述的多核苷酸,其中所述靶抗原选自 $\alpha v\beta 6$ 整合素($\alpha v\beta 6$ 整合素)、B细胞成熟抗原(BCMA)、B7-H3、B7-H6、碳酸酐酶9(CA9,也称为CAIX或G250)、癌症-睾丸抗原、癌症/睾丸抗原1B(CTAG,也称为NY-ESO-1和LAGE-2)、癌胚抗原(CEA)、细胞周期蛋白、细胞周期蛋白A2、C-C基序趋化因子配体1(CCL-1)、CD19、CD20、CD22、CD23、CD24、CD30、CD33、CD38、CD44、CD44v6、CD44v7/8、CD123、CD133、CD138、CD171、硫酸软骨素蛋白聚糖4(CSPG4)、表皮生长因子蛋白(EGFR)、表皮生长因子受体III型突变体(EGFR vIII)、上皮糖蛋白2(EPG-2)、上皮糖蛋白40(EPG-40)、肝配蛋白B2、肝配蛋白受体A2(EPhA2)、雌激素受体、Fc受体样蛋白5(FCRL5;也称为Fc受体同源物5或FCRH5)、胎儿乙酰胆碱受体(胎儿AChR)、叶酸结合蛋白(FBP)、叶酸受体 α 、神经节苷脂GD2、O-乙酰化GD2(OGD2)、神经节苷脂GD3、糖蛋白100(gp100)、磷脂酰肌醇蛋白聚糖-3(GPC3)、G蛋白偶联受体C类5族成员D(GPRC5D)、Her2/neu(受体酪氨酸激酶erb-B2)、Her3(erb-B3)、Her4(erb-B4)、erbB二

聚体、人高分子量黑色素瘤相关抗原 (HMW-MAA)、乙型肝炎表面抗原、人白细胞抗原A1 (HLA-A1)、人白细胞抗原A2 (HLA-A2)、IL-22受体 α (IL-22R α)、IL-13受体 α 2 (IL-13R α 2)、激酶插入结构域受体 (kdr)、 κ 轻链、L1细胞粘附分子 (L1-CAM)、L1-CAM的CE7表位、含有富亮氨酸重复序列的蛋白8家族成员A (LRRC8A)、Lewis Y、黑色素瘤相关抗原 (MAGE) -A1、MAGE-A3、MAGE-A6、MAGE-A10、间皮素 (MSLN)、c-Met、鼠类巨细胞病毒 (CMV)、粘蛋白1 (MUC1)、MUC16、自然杀伤细胞2族成员D (NKG2D) 配体、黑色素A (MART-1)、神经细胞粘附分子 (NCAM)、癌胚胎抗原、黑色素瘤优先表达抗原 (PRAME)、孕酮受体、前列腺特异性抗原、前列腺干细胞抗原 (PSCA)、前列腺特异性膜抗原 (PSMA)、受体酪氨酸激酶样孤儿受体1 (ROR1)、存活蛋白、滋养层糖蛋白 (TPBG, 也称为5T4)、肿瘤相关糖蛋白72 (TAG72)、酪氨酸酶相关蛋白1 (TRP1, 也称为TYRP1或gp75)、酪氨酸酶相关蛋白2 (TRP2, 也称为多巴色素互变异构酶、多巴色素 δ 异构酶或DCT)、血管内皮生长因子受体 (VEGFR)、血管内皮生长因子受体2 (VEGFR2)、肾母细胞瘤1 (WT-1)、病原体特有的或病原体表达的抗原、或与通用标签相关的抗原、和/或生物素化分子、和/或由HIV、HCV、HBV或其他病原体表达的分子。

88. 根据实施方案82-87中任一项所述的多核苷酸, 其中所述细胞外区域包含间隔子, 任选地其中所述间隔子可操作地连接于所述结合结构域与所述跨膜结构域之间。

89. 根据实施方案88所述的多核苷酸, 其中所述间隔子包含免疫球蛋白铰链区。

90. 根据实施方案88或实施方案89所述的多核苷酸, 其中所述间隔子包含C_H2区域和C_H3区域。

91. 根据实施方案82-90中任一项所述的多核苷酸, 其中所述细胞内区域的所述部分包含一个或多个共刺激信号传导结构域。

92. 根据实施方案91所述的多核苷酸, 其中所述一个或多个共刺激信号传导结构域包含CD28、4-1BB或ICOS的细胞内信号传导结构域或其信号传导部分。

93. 根据实施方案91或实施方案92所述的多核苷酸, 其中所述一个或多个共刺激信号传导结构域包含4-1BB的细胞内信号传导结构域。

94. 根据实施方案82-93中任一项所述的多核苷酸, 其中在所述嵌合受体由引入所述多核苷酸的细胞表达时, 所编码的嵌合受体从其N末端至C末端按顺序包含: 所述细胞外结合结构域、所述间隔子、所述跨膜结构域和细胞内信号传导区域。

95. 根据实施方案53-94中任一项所述的多核苷酸, 其中:

(a) 的所述核酸序列按顺序包含: 编码细胞外结合结构域、任选地scFv的核苷酸序列; 间隔子, 所述间隔子任选地包含来自人免疫球蛋白铰链的序列, 任选地来自IgG1、IgG2或IgG4或其修饰形式的序列, 任选地还包含C_H2区域和/或C_H3区域; 以及任选地来自人CD28的跨膜结构域; 任选地来自人4-1BB的共刺激信号传导结构域; 和/或

所述修饰的CD247基因座按顺序包含: 编码细胞外结合结构域、任选地scFv的核苷酸序列; 间隔子, 所述间隔子任选地包含来自人免疫球蛋白铰链的序列, 任选地来自IgG1、IgG2或IgG4或其修饰形式的序列, 任选地还包含CH2区域和/或CH3区域; 以及任选地来自人CD28的跨膜结构域; 任选地来自人4-1BB的共刺激信号传导结构域; 以及CD3 ζ 信号传导结构域。

96. 根据实施方案81-95中任一项所述的多核苷酸, 其中所述CAR是多链CAR。

97. 根据实施方案53-96中任一项所述的多核苷酸, 其中 (a) 的所述核酸序列包含

编码至少一种其他蛋白质的核苷酸序列。

98. 根据实施方案53-97中任一项所述的多核苷酸,其中(a)的所述核酸序列包含一个或多个多顺反子元件。

99. 根据实施方案98所述的多核苷酸,其中所述一个或多个多顺反子元件定位于编码所述嵌合受体的一部分的核苷酸序列与编码所述至少一种其他蛋白质的核苷酸序列之间。

100. 根据实施方案97-99中任一项所述的多核苷酸,其中所述至少一种其他蛋白质是替代标记,任选地其中所述替代标记是截短的受体,任选地其中所述截短的受体缺少细胞内信号传导结构域和/或在与其配体结合时不能介导细胞内信号传导。

101. 根据实施方案100所述的多核苷酸,其中所述嵌合受体是多链CAR,并且多顺反子元件定位于编码所述多链CAR的一条链的核苷酸序列与编码所述多链CAR的另一条链的核苷酸序列之间。

102. 根据实施方案98-101中任一项所述的多核苷酸,其中所述一个或多个多顺反子元件位于编码所述嵌合受体的一部分的核苷酸序列的上游。

103. 根据实施方案98-102中任一项所述的多核苷酸,其中所述一个或多个多顺反子元件是或包含核糖体跳跃序列,任选地其中所述核糖体跳跃序列是T2A、P2A、E2A或F2A元件。

104. 根据实施方案53-103中任一项所述的多核苷酸,其中所述修饰的CD247基因座包含所述内源CD247基因座的启动子和/或调节或控制元件,所述启动子和/或调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。

105. 根据实施方案53-104中任一项所述的多核苷酸,其中所述修饰的基因座包含一个或多个异源调节或控制元件,所述一个或多个异源调节或控制元件可操作地连接以控制编码所述嵌合受体的核酸序列的表达。

106. 根据实施方案105所述的多核苷酸,其中所述一个或多个异源调节或控制元件包括启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、剪接接纳体序列和/或剪接供体序列。

107. 根据实施方案106所述的多核苷酸,其中所述异源启动子是或包含人延伸因子1 α (EF1 α) 启动子或MND启动子或其变体。

108. 根据实施方案53-107中任一项所述的多核苷酸,其中所述多核苷酸包含于病毒载体中。

109. 根据实施方案108所述的多核苷酸,其中所述病毒载体是AAV载体。

110. 根据实施方案109所述的多核苷酸,其中所述AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7或AAV8载体。

111. 根据实施方案109或实施方案110所述的多核苷酸,其中所述AAV载体是AAV2或AAV6载体。

112. 根据实施方案108所述的多核苷酸,其中所述病毒载体是逆转录病毒载体,任选地慢病毒载体。

113. 根据实施方案53-112中任一项所述的多核苷酸,其为线性多核苷酸,任选地双链多核苷酸或单链多核苷酸。

114. 根据实施方案52-113中任一项所述的多核苷酸,其中所述多核苷酸具有至少或至少约2500、2750、3000、3250、3500、3750、4000、4250、4500、4760、5000、5250、5500、5750、6000、7000、7500、8000、9000或10000个核苷酸的长度,或者任何前述值之间的任何值。

115. 根据实施方案52-114中任一项所述的多核苷酸,其中所述多核苷酸具有在为或约2500与为或约5000个核苷酸之间、为或约3500与为或约4500个核苷酸之间、或者为或约3750个核苷酸与为或约4250个核苷酸之间的长度。

116. 一种产生基因工程化的T细胞的方法,所述方法包括将根据实施方案53-115中任一项所述的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中。

117. 一种产生基因工程化的T细胞的方法,所述方法包括:

(a) 将一种或多种药剂引入T细胞中,所述一种或多种药剂能够在所述T细胞的内源CD247基因座内的靶位点处诱导遗传破坏;以及

(b) 将根据实施方案53-115中任一项所述的多核苷酸引入在CD247基因座处包含遗传破坏的T细胞中,其中所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。

118. 根据实施方案117所述的方法,其中所述多核苷酸包含编码嵌合受体或其部分的核酸序列,并且编码嵌合受体或其部分的所述核酸序列经由同源定向修复 (HDR) 整合于所述内源CD247基因座内。

119. 一种产生基因工程化的T细胞的方法,所述方法包括将包含编码嵌合受体或其部分的核酸序列的多核苷酸引入T细胞中,所述T细胞在所述T细胞的CD247基因座内具有遗传破坏,其中编码所述嵌合受体或其部分的所述核酸序列经由同源定向修复 (HDR) 整合于所述内源CD247基因座内。

120. 根据实施方案116或实施方案119所述的方法,其中所述遗传破坏是通过将一种或多种药剂引入T细胞中来进行,所述一种或多种药剂能够在所述T细胞的内源CD247基因座内的靶位点处诱导遗传破坏。

121. 根据实施方案116-120中任一项所述的方法,其中所述方法产生修饰的CD247基因座,所述修饰的CD247基因座包含编码嵌合受体的核酸序列,所述嵌合受体包含含有CD3zeta (CD3 ζ) 信号传导结构域的细胞内区域。

122. 根据实施方案119-121中任一项所述的方法,其中所述核酸序列包含编码嵌合受体或其部分编码所述嵌合受体的一部分的核酸序列。

123. 根据实施方案119-122中任一项所述的方法,其中所述多核苷酸还包含与所述核酸序列连接的一个或多个同源臂,其中所述一个或多个同源臂包含与CD247基因座的开放阅读框的一个或多个区域同源的序列。

124. 根据实施方案116-123中任一项所述的方法,其中所述嵌合受体的完整细胞内区域包含CD3zeta (CD3 ζ) 信号传导结构域或其片段,其中所述细胞内区域的至少一部分是由通过所述方法生成的细胞中所述内源CD247基因座的开放阅读框或其部分序列编码。

125. 根据实施方案116-124中任一项所述的方法,其中编码所述嵌合受体的所述部分的所述核酸序列和所述一个或多个同源臂一起包含编码所述嵌合受体的细胞内区域

的核苷酸序列的至少一个片段,其中所述细胞内区域的至少一部分包含由通过所述方法生成的细胞中所述CD247基因座的开放阅读框或其部分序列编码的所述CD3 ζ 信号传导结构域或其片段。

126. 根据实施方案119-125中任一项所述的方法,其中编码嵌合受体或其部分的所述核酸序列不包含编码3'UTR的序列。

127. 根据实施方案119-126中任一项所述的方法,其中在通过所述方法生成的细胞中,编码嵌合受体或其部分的所述核酸序列编码所述CD3 ζ 信号传导结构域的片段。

128. 根据实施方案119-126中任一项所述的方法,其中在通过所述方法生成的细胞中,编码嵌合受体或其部分的所述核酸序列不编码所述CD3 ζ 信号传导结构域或其片段。

129. 根据实施方案119-128中任一项所述的方法,其中所编码的嵌合受体的所述CD3 ζ 信号传导结构域的至少一个片段、任选地整个CD3 ζ 信号传导结构域,是由通过所述方法生成的细胞中所述内源CD247基因座的开放阅读框或其部分序列编码。

130. 根据实施方案119-129中任一项所述的方法,其中编码嵌合受体或其部分的所述核酸序列是对于T细胞、任选地人T细胞的所述内源基因组CD247基因座的开放阅读框为外源或异源的序列。

131. 根据实施方案119-130中任一项所述的方法,其中编码嵌合受体或其部分的所述核酸序列包含与所述一个或多个同源臂中包含的所述CD247基因座的开放阅读框或其部分序列的一个或多个外显子同框的核苷酸序列。

132. 根据实施方案119-131中任一项所述的方法,其中,在由引入所述多核苷酸的细胞表达时,所述嵌合受体能够经由所述CD3 ζ 信号传导结构域进行信号传导。

133. 根据实施方案119-132中任一项所述的方法,其中所述CD3 ζ 信号传导结构域包含选自SEQ ID NO:13-15中任一项的序列,或者展现与SEQ ID NO:13-15中任一项至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列,或者其片段。

134. 根据实施方案119-133中任一项所述的方法,其中所述一个或多个同源臂包含5'同源臂和3'同源臂。

135. 根据实施方案119-134中任一项所述的方法,其中所述多核苷酸包含结构[5'同源臂]-[编码嵌合受体或其部分的核酸序列]-[3'同源臂]。

136. 根据实施方案134或实施方案135所述的方法,其中所述5'同源臂和所述3'同源臂独立地具有为或约50至为或约2000个核苷酸、为或约100至为或约1000个核苷酸、为或约100至为或约750个核苷酸、为或约100至为或约600个核苷酸、为或约100至为或约400个核苷酸、为或约100至为或约300个核苷酸、为或约100至为或约200个核苷酸、为或约200至为或约1000个核苷酸、为或约200至为或约750个核苷酸、为或约200至为或约600个核苷酸、为或约200至为或约400个核苷酸、为或约200至为或约300个核苷酸、为或约300至为或约1000个核苷酸、为或约300至为或约750个核苷酸、为或约300至为或约600个核苷酸、为或约300至为或约400个核苷酸、为或约400至为或约1000个核苷酸、为或约400至为或约750个核苷酸、为或约400至为或约600个核苷酸、为或约600至为或约1000个核苷酸、为或约600至为或约750个核苷酸或者为或约750至为或约1000个核苷酸的长度。

137. 根据实施方案134-136中任一项所述的方法,其中所述5'同源臂和所述3'同

源臂独立地具有为或约200、300、400、500、600、700或800个核苷酸的长度,或者任何前述值之间的任何值。

138. 根据实施方案134-137中任一项所述的方法,其中所述5'同源臂和所述3'同源臂独立地具有大于或大于约300个核苷酸的长度,任选地其中所述5'同源臂和所述3'同源臂独立地具有为或约400、500或600个核苷酸的长度,或者任何前述值之间的任何值。

139. 根据实施方案134-138中任一项所述的方法,其中所述5'同源臂包含SEQ ID NO:80中所示的序列,或者展现与SEQ ID NO:80至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。

140. 根据实施方案134-139中任一项所述的方法,其中所述3'同源臂包含SEQ ID NO:81中所示的序列,或者展现与SEQ ID NO:81至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高序列同一性的序列或其部分序列。

141. 根据实施方案117和120-140中任一项所述的方法,其中能够诱导遗传破坏的所述一种或多种药剂包含与所述靶位点特异性结合或杂交的DNA结合蛋白或DNA结合核酸、包含DNA靶向蛋白和核酸酶的融合蛋白或者RNA指导核酸酶,任选地其中所述一种或多种药剂包含与所述靶位点特异性结合、识别所述靶位点或与所述靶位点杂交的锌指核酸酶(ZFN)、TAL效应核酸酶(TALEN)或和CRISPR-Cas9组合。

142. 根据实施方案117和120-141中任一项所述的方法,其中所述一种或多种药剂各自包含具有与至少一个靶位点互补的靶向结构域的指导RNA(gRNA)。

143. 根据实施方案117和120-142中任一项所述的方法,其中所述一种或多种药剂是作为包含gRNA和Cas9蛋白的核糖核蛋白(RNP)复合物引入。

144. 根据实施方案143所述的方法,其中所述RNP是通过电穿孔、粒子枪、磷酸钙转染、细胞压缩或挤压、任选地通过电穿孔来引入。

145. 根据实施方案143或实施方案144所述的方法,其中所述RNP的浓度是为或约1、2、2.5、5、10、20、25、30、40或50 μ M,或者由任两个前述值限定的范围,任选地其中所述RNP的浓度是为或约25 μ M。

146. 根据实施方案143-145中任一项所述的方法,其中所述RNP中所述gRNA与所述Cas9分子的摩尔比是为或约为或约5:1、4:1、3:1、2:1、1:1、1:2、1:3、1:4或1:5,或者由任两个前述值限定的范围,任选地其中所述RNP中所述gRNA与所述Cas9分子的摩尔比是为或约2.6:1。

147. 根据实施方案142-146中任一项所述的方法,其中所述gRNA具有选自以下的靶向结构域序列:CACCUUCACUCUCAGGAACA (SEQ ID NO:87);GAAUGACACCAUAGAUGAAG (SEQ ID NO:88);UGAAGAGGAUCCAUCAGC (SEQ ID NO:89);以及UCCAGCAGGUAGCAGAGUUU (SEQ ID NO:90)。

148. 根据实施方案142-147中任一项所述的方法,其中所述gRNA具有CACCUUCACUCUCAGGAACA (SEQ ID NO:87)的靶向结构域序列。

149. 根据实施方案142-147中任一项所述的方法,其中所述gRNA具有UGAAGAGGAUCCAUCAGC (SEQ ID NO:89)的靶向结构域序列。

150. 根据实施方案116-149中任一项所述的方法,其中所述T细胞是源自受试者的原代T细胞,任选地其中所述受试者是人。

151. 根据实施方案116-150中任一项所述的方法,其中所述T细胞是CD8⁺T细胞或其亚型。

152. 根据实施方案116-150中任一项所述的方法,其中所述T细胞是CD4⁺T细胞或其亚型。

153. 根据实施方案116-152中任一项所述的方法,其中所述T细胞源自多潜能或多能细胞,其任选地是iPSC。

154. 根据实施方案119-153中任一项所述的方法,其中所述多核苷酸包含于病毒载体中。

155. 根据实施方案154所述的方法,其中所述病毒载体是AAV载体。

156. 根据实施方案155所述的方法,其中所述AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7或AAV8载体。

157. 根据实施方案155或实施方案156所述的方法,其中所述AAV载体是AAV2或AAV6载体。

158. 根据实施方案154所述的方法,其中所述病毒载体是逆转录病毒载体,任选地慢病毒载体。

159. 根据实施方案119-153中任一项所述的方法,其中所述多核苷酸是线性多核苷酸,任选地双链多核苷酸或单链多核苷酸。

160. 根据实施方案117和120-159中任一项所述的方法,其中所述一种或多种药剂和所述多核苷酸是同时或按任何顺序依序引入。

161. 根据实施方案117和120-160中任一项所述的方法,其中所述多核苷酸是在引入所述一种或多种药剂之后引入。

162. 根据实施方案161所述的方法,其中所述多核苷酸在引入所述药剂后立即引入,或者在引入所述药剂后约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、6分钟、8分钟、9分钟、10分钟、15分钟、20分钟、30分钟、40分钟、50分钟、60分钟、90分钟、2小时、3小时或4小时内引入。

163. 根据实施方案117和120-162中任一项所述的方法,其中在引入所述一种或多种药剂之前,所述方法包括在刺激或激活所述一种或多种免疫细胞的条件下将所述细胞在体外与一种或多种刺激剂一起孵育。

164. 根据实施方案163所述的方法,其中所述一种或多种刺激剂包含和抗CD3和/或抗CD28抗体,任选地抗CD3/抗CD28珠,任选地其中珠与细胞的比率为或为约1:1。

165. 根据实施方案163或实施方案164所述的方法,其包括在引入所述一种或多种药剂之前,从所述一种或多种免疫细胞去除所述一种或多种刺激剂。

166. 根据实施方案117和120-165中任一项所述的方法,其中所述方法还包括在引入所述一种或多种药剂和/或引入所述多核苷酸之前、期间或之后,将所述细胞与一种或多种重组细胞因子一起孵育,任选地其中所述一种或多种重组细胞因子选自IL-2、IL-7和IL-15。

167. 根据实施方案166所述的方法,其中所述一种或多种重组细胞因子是以选自

以下的浓度添加:为或约10U/mL至为或约200U/mL、任选地为或约50IU/mL至为或约100U/mL的浓度的IL-2;0.5ng/mL至50ng/mL、任选地为或约5ng/mL至为或约10ng/mL的浓度的IL-7;和/或0.1ng/mL至20ng/mL、任选地为或约0.5ng/mL至为或约5ng/mL的浓度的IL-15。

168. 根据实施方案166或实施方案167所述的方法,其中所述孵育是在引入所述一种或多种药剂和引入所述多核苷酸之后进行长达或大约24小时、36小时、48小时、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或21天,任选地长达或约7天。

169. 根据实施方案116-168中任一项所述的方法,其中通过所述方法生成的多个工程化细胞中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞包含CD247基因座内至少一个靶位点的遗传破坏。

170. 根据实施方案116-169中任一项所述的方法,其中通过所述方法生成的多个工程化细胞中至少或大于35%、40%、45%、50%、55%、60%、65%、70%、75%、80%或90%的细胞表达所述嵌合受体或其抗原结合片段。

171. 一种工程化T细胞或多个工程化T细胞,其是使用根据实施方案116-170中任一项所述的方法来生成。

172. 一种组合物,其包含根据实施方案1-52和171中任一项所述的工程化T细胞。

173. 一种组合物,其包含多个根据实施方案1-52和171中任一项所述的工程化T细胞。

174. 根据实施方案172或实施方案173所述的组合物,其中所述组合物包含CD4+和/或CD8+T细胞。

175. 根据实施方案172-174中任一项所述的组合物,其中所述组合物包含CD4+和CD8+T细胞,并且CD4+与CD8+T细胞的比率为从或从约1:3至3:1,任选地1:1。

176. 根据实施方案172-175中任一项所述的组合物,其中表达所述嵌合受体的细胞构成所述组合物中总细胞的或所述组合物中总CD4+或CD8+细胞的至少30%、40%、50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多。

177. 一种治疗方法,其包括将根据实施方案1-52和171-176中任一项所述的工程化细胞、多个工程化细胞或组合物施用至患有疾病或障碍的受试者。

178. 根据实施方案1-52和171-176中任一项所述的工程化细胞、多个工程化细胞或组合物用于治疗疾病或障碍的用途。

179. 根据实施方案1-52和171-176中任一项所述的工程化细胞、多个工程化细胞或组合物在制造用于治疗疾病或障碍的药物的用途。

180. 根据实施方案1-52和171-176中任一项所述的工程化细胞、多个工程化细胞或组合物,其用于治疗疾病或障碍。

181. 根据实施方案177-180中任一项所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述疾病或障碍是癌症或肿瘤。

182. 根据实施方案181所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述癌症或所述肿瘤是血液恶性肿瘤,任选地淋巴瘤、白血病或浆细胞恶性肿瘤。

183. 根据实施方案181或实施方案182所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述癌症是淋巴瘤,并且所述淋巴瘤是伯基特淋巴

瘤、非霍奇金淋巴瘤 (NHL)、霍奇金淋巴瘤、华氏巨球蛋白血症、滤泡性淋巴瘤、小无裂细胞性淋巴瘤、粘膜相关淋巴组织淋巴瘤 (MALT)、边缘区淋巴瘤、脾淋巴瘤、结节单核细胞样B细胞淋巴瘤、免疫母细胞淋巴瘤、大细胞淋巴瘤、弥漫性混合细胞淋巴瘤、肺B细胞血管中心淋巴瘤、小淋巴细胞性淋巴瘤、原发性纵隔B细胞淋巴瘤、淋巴浆细胞性淋巴瘤 (LPL) 或套细胞淋巴瘤 (MCL)。

184. 根据实施方案181或实施方案182所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述癌症是白血病,并且所述白血病是慢性淋巴细胞性白血病 (CLL)、浆细胞白血病或急性淋巴细胞性白血病 (ALL)。

185. 根据实施方案181或实施方案182所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述癌症是浆细胞恶性肿瘤,并且所述浆细胞恶性肿瘤是多发性骨髓瘤 (MM)。

186. 根据实施方案181所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述肿瘤是实体瘤。

187. 根据实施方案186所述的方法、用途或者用于所述用途的工程化细胞、多个工程化细胞或组合物,其中所述实体瘤是非小细胞肺癌 (NSCLC) 或头颈鳞状细胞癌 (HNSCC)。

188. 一种试剂盒,其包含:

能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂;以及根据实施方案53-115中任一项所述的多核苷酸。

189. 一种试剂盒,其包含:

能够在CD247基因座内的靶位点处诱导遗传破坏的一种或多种药剂;以及

包含编码嵌合受体或其部分的核酸序列的多核苷酸,其中编码所述嵌合受体或其抗原结合片段或链的转基因被靶向以供经由同源定向修复 (HDR) 整合于所述靶位点处或附近;以及

用于进行根据实施方案116-170中任一项所述的方法的说明书。

IX. 实施例

[0935] 以下实施例仅出于说明性目的而被包括,并且不意图限制本发明的范围。

实施例1在内源CD247基因座处的遗传破坏

[0936] 针对通过CRISPR/Cas9介导的基因编辑在内源CD247基因座 (编码CD3 ζ) 内的遗传破坏评估指导RNA。

[0937] 通过基于免疫亲和力的选择从获自健康供体的人外周血单个核细胞 (PBMC) 分离原代人CD4⁺和CD8⁺T细胞。通过与抗CD3/抗CD28试剂一起培养,将所得CD4⁺和CD8⁺细胞 (1:1比率) 刺激72小时。去除抗CD3/抗CD28试剂,并且用2 μ M核糖核蛋白 (RNP) 复合物对刺激的细胞进行电穿孔,以供通过CRISPR/Cas9介导的基因编辑在内源CD247基因座处引入遗传破坏。所述RNP复合物含有酿脓链球菌Cas9以及四 (4) 种示例性gRNA中的一种,所述示例性gRNA具有如下靶向结构域序列: CACCUUCACUCUCAGGAACA (CD247 gRNA 1; SEQ ID NO: 87); GAAUGACACCAUAGAUGAAG (CD247 gRNA 2; SEQ ID NO: 88); UGAAGAGGAUCCAUCAGC (CD247 gRNA 3; SEQ ID NO: 89); 或者UCCAGCAGGUAGCAGAGUUU (CD247 gRNA 4; SEQ ID NO: 90), 或者进行模拟处理 (模拟)。gRNA与Cas9蛋白的比率为约2.0:1。将所述细胞培养3天并通过流式细胞术进行评估,用抗CD3抗体和抗T细胞受体 (TCR) 抗体染色。

[0938] 如图1中所示,在用含有靶向CD247的gRNA的RNP复合物电穿孔的细胞中敲除CD3表达,范围为大约12.6%至76.1%的所述细胞被敲除了CD3表达。CD247 gRNA 3(具有SEQ ID NO:89中所示的靶向结构域序列)显示最高敲除效率。

实施例2编码嵌合抗原受体(CAR)的一部分的转基因序列在T细胞中的内源CD247基因座处的靶向整合

[0939] 将人T细胞工程化以表达嵌合抗原受体(CAR),其中编码CAR或其部分的核酸序列被靶向以供经由同源依赖修复(HDR)整合于内源CD247基因座(编码CD3 ζ)处。

A. 用于HDR介导的靶向的模板多核苷酸

[0940] 生成示例性模板多核苷酸用于转基因序列的靶向整合,所述转基因序列包括编码CAR的一部分的核酸序列,所述CAR包括对示例性抗原BCMA具有特异性的scFv、免疫球蛋白来源的间隔子、源自CD28的跨膜结构域和源自4-1BB的共刺激区域。

[0941] 所述转基因序列还包括a)人延伸因子1 α (EF1 α)启动子,以驱动CAR编码序列在异源启动子控制下的表达;或者b)位于编码抗BCMA CAR的部分的核酸序列上游的P2A核糖体跳跃元件,以在HDR介导的框内靶向整合至CD247开放阅读框中之后,驱动从内源CD247基因座表达CAR。转基因序列还包括i)编码CD3 ζ 链的细胞内部分的异源核酸序列和异源多腺苷酸化信号(SV40多腺苷酸化序列;SV40pA),或者ii)无编码CD3 ζ 链的一部分的异源序列,使得在靶向整合后,编码CAR的转录物含有CD3 ζ 链的内源3'UTR。

[0942] 模板多核苷酸还含有同源臂以引导将转基因序列靶向整合至T细胞的内源人CD247基因座中。示例性模板多核苷酸的大体结构如下:[5'同源臂]-[转基因序列]-[3'同源臂]。5'同源臂含有大约600bp的序列,所述序列与内源人CD247基因座的第一内含子和第二外显子的一部分同源(SEQ ID NO:80中所示的5'同源臂序列)。3'同源臂含有大约600bp序列,所述序列与第二外显子的一部分(包括编码由第二外显子编码的最后3个氨基酸的序列)和第二内含子的一部分同源(SEQ ID NO:81中所示的3'同源臂序列)。

[0943] 表E1显示生成的模板多核苷酸的结构。

表E1. 示例性模板多核苷酸。

多核苷酸	示例性模板多核苷酸的组分
A	[5'同源臂]-[P2A - 抗BCMA CAR(具有异源CD3 ζ)- SV40pA]-[3'同源臂] (内源CD3 ζ 启动子; 异源3' UTR)
B	[5'同源臂]-[P2A - 抗BCMA CAR(没有异源CD3 ζ)-[3'同源臂] (内源CD3 ζ 启动子; 内源3' UTR)
C	[5'同源臂]-[EF1 α 启动子 - 抗BCMA CAR(具有异源CD3 ζ)- SV40pA]-[3'同源臂] (异源EF1 α 启动子; 异源3' UTR)
D	[5'同源臂]-[EF1 α 启动子 - 抗BCMA CAR(没有异源CD3 ζ)-[3'同源臂] (异源EF1 α 启动子; 内源3' UTR)

B. 通过HDR进行的工程化T细胞的生成以及CAR的表达

[0944] 对于通过HDR靶向整合,生成含有表E1中所示四种多核苷酸(多核苷酸A、B、C、D)之一的腺相关病毒(AAV)载体构建体。通过AAV载体的三重转染以供将模板多核苷酸、血清型辅助质粒和腺病毒辅助质粒引入293T细胞系中来产生AAV原液。收集转染的细胞,将其裂解,并收集AAV原液用于细胞转导。

[0945] 从人供体受试者分离原代人CD4⁺和CD8⁺T细胞,进行刺激,并用2 μ M的RNP复合物对刺激的细胞进行电穿孔,所述RNP复合物含有CD247靶向gRNA 3和酿脓链球菌Cas9,gRNA与

Cas9蛋白为约2.0:1,如实施例1中所述。在电穿孔后,立即用以5%v/v含有多核苷酸A、B、C、D之一的AAV制剂转导细胞。作为对照,通过慢病毒转导引入编码抗BCMA CAR的核酸序列或进行模拟处理。

[0946] 将细胞培养3天并通过流式细胞术进行评估,用抗CD3 epsilon (抗CD3 ϵ)抗体和BCMA-Fc (在其C末端与IgG的Fc区融合的可溶人BCMA)融合多肽染色,以检测抗BCMA CAR的表达。

[0947] 如图2A中所示,4种模板多核苷酸(多核苷酸A、B、C、D)中每一种的引入都导致将抗BCMA CAR靶向整合至CD247基因座中,如通过抗BCMA CAR在细胞表面上的表达所证实,并且少于20%的细胞表达CD3。图2B描绘在每种条件下示例性抗BCMA CAR的表达的变异系数(CV)(细胞群内信号的标准差除以相应群中信号的平均值)和几何平均荧光(gMFI)。如图所示,对于某些所测试的多核苷酸,经由HDR将所述核酸靶向整合于内源CD247基因座处导致较低的变异系数。

实施例3通过将编码嵌合抗原受体(CAR)的一部分的转基因序列靶向整合于T细胞中的内源CD247基因座处来工程化的CAR表达细胞的功能活性的评估

[0948] 针对抗原依赖性刺激后的细胞裂解活性和细胞因子产生来评估如实施例2中所述生成的工程化细胞。

[0949] 为了评估细胞裂解活性,在具有RPMI 8226多发性骨髓瘤细胞(ATCC®CCL-155™;表达低水平的BCMA)或BCMA转导的K562慢性髓细胞性白血病(CML)细胞(ATCC®CCL-243™;K562-BCMA,表达高水平的BCMA)的96孔板中培养工程化的CAR表达T细胞,并且以2:1、1:1或1:2的E:T比率进行孵育。评估模拟电穿孔和转导的细胞(模拟)以及在没有任何CAR+细胞的情况下培养的靶细胞(仅靶标)作为对照。在49小时中测量NucLight Red (NLR)标记的活靶细胞的损失,如通过红色荧光信号(使用IncuCyte®活细胞分析系统,Essen Bioscience)所确定。确定裂解百分比并针对CAR+群体归一化。

[0950] 还在将工程化的抗BCMA CAR表达细胞与靶细胞共培养后评估细胞因子产生。收集来自共培养的细胞的上清液,用于使用多重细胞因子免疫测定测量干扰素- γ (IFN- γ)、肿瘤坏死因子 α (TNF- α)和白介素-2 (IL-2)。

[0951] 针对RPMI 8226细胞,在每个评估的E:T比率下,每组工程化细胞在实验期间的总裂解百分比显示于图3A中。在2:1、1:1和1:2E:T比率下随时间观察到的RPMI 8226细胞的裂解分别显示于图3C、图3D和图3E中。如图所示,针对示例性RPMI 8226细胞,通过经由HDR靶向整合于CD247基因座处来工程化以表达示例性抗BCMA CAR的细胞展现与通过慢病毒递送用CAR工程化的细胞可比的活性。

[0952] 针对表达相对更高BCMA的替代性K562-BCMA靶细胞的结果显示于图3B和图3F-图3H中。在每个评估的E:T比率下,每组工程化细胞在实验期间的总裂解百分比显示于图3B中,并且在2:1、1:1和1:2E:T比率下随时间观察的K562靶细胞的裂解分别显示于图3F、图3G和图3H中。在此实验中,与使用慢病毒转导工程化的细胞相比,在各个E:T比率下,通过经由HDR靶向整合于CD247基因座处工程化以表达示例性抗BCMA CAR的细胞展现可比的或更大的针对K562-BCMA靶细胞的细胞裂解活性。特定地,特别是在较低E:T比率下,在通过使用多核苷酸D(使用EF1 α 启动子表达并且含有内源CD3 ζ 3'UTR)靶向整合CAR来工程化的细胞中观察到最大细胞裂解活性,从而表明所述多核苷酸D在用作用于HDR介导的靶向整合的构建体

时的优良活性。

[0953] 如图4A中所示,针对两种BCMA表达细胞系,与使用慢病毒转导工程化的细胞相比,使用HDR工程化的细胞产生可比的或更高的IFN- γ 。在此研究中,通过使用多核苷酸D靶向整合CAR工程化的细胞展现最高的IFN- γ 的一般产生。如图4B-图4C中所示,在使用HDR将CAR靶向整合至CD247基因座中来工程化的细胞中,IL-2和TNF- α 产生较低。

[0954] 结果与以下一致:与使用慢病毒转导工程化的细胞相比,通过HDR介导的将编码CAR或其部分的核酸序列整合于内源CD247基因座处来工程化以表达示例性CAR的原代T细胞的活性是可比的或改进的。

实施例4通过使用各种模板多核苷酸构建体的HDR介导的整合或通过慢病毒递送在T细胞中表达CAR的比较

[0955] 通过经由HDR使用如上所述的各种构建体靶向编码CAR的全部或一部分的核酸序列以供整合于内源CD247基因座(编码CD3 ζ)处或通过慢病毒递送来将T细胞工程化以表达示例性CAR,并且评估CAR的表达和活性。

[0956] 如上文实施例1中所述刺激原代人T细胞,并用含有合成的gRNA 1或gRNA 3(各自具有Alt-R修饰, IDT Technologies;科勒尔维尔,爱荷华州)的RNP复合物对刺激的细胞进行电穿孔。RNP复合物以约2.6:1的比率含有gRNA与Cas9蛋白,并且用RNP以25 μ M的浓度对细胞进行电穿孔。如图5中所示,使用gRNA 1和gRNA 3分别敲除大约87.9%和95.5%的细胞的CD3表达。这些结果支持,gRNA的修饰、gRNA与Cas9蛋白的比率和/或更高浓度的RNP导致更高的CD247基因座的敲除效率。

[0957] 对于通过HDR生成CAR表达细胞,从三个人供体受试者(供体1-3)分离原代人CD4⁺和CD8⁺T细胞,如上所述对所述细胞进行刺激并用25 μ M的含有CD247靶向修饰的gRNA 1(供体1)或gRNA 3(供体2和3)的RNP复合物进行电穿孔。在电穿孔后,立即将细胞与含有多核苷酸A、B、C、D(如上文表E1中所示)的AAV制剂一起孵育。作为对照,通过经由慢病毒转导引入编码抗BCMA CAR的核酸序列来工程化细胞,或者用CD247靶向RNP进行电穿孔但不与AAV制剂接触,或者进行模拟处理(模拟)。在电穿孔和/或转导后四天,将细胞用CD3 ϵ (抗CD3 ϵ)抗体和BCMA-Fc(在其C末端与IgG的Fc区融合的可溶人BCMA)融合多肽染色,以检测抗BCMA CAR的表达。

[0958] 将细胞与标记的MM.1S(ATCC®CRL-2974TM)人B淋巴母细胞靶细胞以2:1或1:2的效应子:靶标(E:T)比率一起孵育,并且在三天中测量NucLight Red(NLR)标记的活靶细胞的损失,如通过红色荧光信号(使用IncuCyte®活细胞分析系统,Essen Bioscience)所确定。确定裂解的细胞的百分比(裂解%),在三个供体之间进行归一化,并从一式三份样品取平均值。通过以下方式获得细胞因子产生:在24小时从共培养的细胞收集上清液,并使用多重细胞因子免疫测定测量IFN- γ 、TNF- α 和IL-2。

[0959] 来自示例性供体的代表性结果显示于图6A-图6C中。在此实验中如通过gRNA3中的CD3 ϵ ⁻群体显示的CD3敲除效率为大约90%。引入4种模板多核苷酸(多核苷酸A、B、C、D)中的每一种用于通过HDR靶向整合于CD247基因座处导致大约49%至65%的细胞在细胞表面上表达抗BCMA CAR。在被靶向以供通过HDR整合至CD247基因座中的细胞中(参见图6A的多核苷酸A、B、C和D图),表达抗BCMA CAR的细胞中的CD3表面表达的损失与一些细胞中CD247的敲除是一致的。如图6B中所示,慢病毒递送导致大约58%的细胞在细胞表面上表达抗BCMA

CAR。在来自另外两个供体(供体2和3,使用gRNA 3进行CD3 ζ 敲除)的结果中,在流式细胞术图中的象限中的细胞百分比(在图6A-图6B中类似地门控)显示于下文表E2和表E3中。

表E2.对于供体2工程化细胞,在流式细胞术图中的每个所指示象限中呈现的细胞百分比

细胞	CD3+/BCMA- 象限	CD3+/BCMA + 象限	CD3-/BCMA+ 象限	CD3-/BCMA- 象限
模拟	98.6	0.45	8.32E-3	0.95
仅KO	16.3	0.045	0.096	83.6
多核苷酸A	7.76	1.89	42.5	48.1
多核苷酸B	8.75	1.46	31.0	58.8
多核苷酸C	8.55	1.44	29.1	60.9
多核苷酸D	8.46	1.66	36.1	53.7
慢病毒	38.2	60.6	0.65	0.52

表E3.对于供体3工程化细胞,在流式细胞术图中的每个所指示象限中呈现的细胞百分比

细胞	CD3+/BCMA- 象限	CD3+/BCMA + 象限	CD3-/BCMA+ 象限	CD3-/BCMA- 象限
模拟	95.8	0.72	0.021	3.44
仅KO	7.57	0.043	0.19	92.2
多核苷酸A	1.41	1.33	66.7	30.6
多核苷酸B	11.0	2.13	47.0	39.9
多核苷酸C	2.46	1.59	54.7	41.2
多核苷酸D	3.35	1.84	55.6	39.3
慢病毒	51.4	44.7	2.14	1.79

[0960] 如图6C中所示,与其中HDR模板将异源EF1 α 启动子和完整外源CAR整合至CD247基因座中的策略相比,在使用其中将CAR工程化以在异源EF1 α 启动子的控制下表达且含有内源CD3 ζ 3'UTR的策略(即,比较多核苷酸D(内源CD3 ζ 3'UTR)与多核苷酸C(异源3'UTR))通过HDR介导的靶向整合来工程化的细胞中,CAR表达的水平分布更狭窄且均匀。与通过慢病毒递送工程化以表达CAR的细胞中相比,CAR表达水平也更紧缩。此结果与通过HDR工程化的细胞中CAR表达水平的更低的变异系数和更紧缩的范围一致。

[0961] 如图7A中所示,在通过HDR使用不同多核苷酸工程化的细胞中,与通过HDR工程化而具有在内源CD247启动子控制下表达的CAR的细胞(多核苷酸A和B)相比,在通过HDR工程化而具有在异源EF1 α 启动子控制下表达的CAR的细胞中(多核苷酸C和D),如通过裂解%表示的细胞裂解活性通常更高。如图7B中所示,如通过裂解%表示的细胞裂解活性与在通过慢病毒递送工程化的细胞中观察到的细胞裂解活性是类似的。

[0962] 如图8A-图8C中所示,与使用慢病毒递送工程化的细胞相比,通过HDR工程化的细胞的IFN- γ (图8A)、TNF- α (图8B)和IL-2(图8C)产生水平通常是类似的或更低的。在一个供体中,在通过HDR工程化的细胞中,具有使用EF1 α 启动子表达且含有内源CD3 ζ 3'UTR的CAR的细胞(多核苷酸D)产生最高IL-2,并且具有使用EF1 α 启动子和SV40pA表达的CAR的细胞(多核苷酸C)展现最高IFN- γ 产生。

[0963] 结果与如下原代T细胞一致,所述原代T细胞通过HDR介导的将编码CAR或其部分的核酸序列(例如,使用含有异源EF1 α 启动子且含有内源CD3 ζ 3'UTR的构建体)整合于内源CD247基因座处来工程化以表达示例性CAR,从而导致与通过慢病毒递送工程化的细胞相比,表达功能性CAR的细胞具有可比的或改进的细胞裂解活性和细胞因子产生。结果支持使

用HDR介导的将编码CAR或其部分的核酸序列整合于CD247基因座(编码CD3ζ)处以供表达具有内源CD247基因座的3'UTR的CAR,用于工程化T细胞,所述T细胞展现高且均匀的CAR表达以及一致的功能活性。

[0964] 本发明的范围不意图受限于具体公开的实施方案,所述实施方案的提供例如是为了说明本发明的各个方面。根据本文的描述和传授,对所述组合物和方法的各种修改将变得清楚。可以在不脱离本公开文本的真实范围和精神的情况下实践这些变化,并且这些变化旨在落入本公开文本的范围内。

序列

#	序列	注释
1	ESKYGPPCPPCP	间隔子 (IgG4 铰链)(aa)
2	GAATCTAAGTACGGACCGCCCTGCCCCCTTGCCCT	间隔子 (IgG4 铰链)(nt)
3	ESKYGPPCPPCPGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMEALHNHYTQKSLSLGK	铰链-C _H 3 间隔子
4	ESKYGPPCPPCPAPEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSDQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVSVLTVHLQDQLWLNQKEYCKVSNKGLPSSIEKTIISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMEALHNHYTQKSLSLGK	铰链 -CH ₂ -C _H 3间隔子
5	RWPESPKAQASSVPTAQQAEGSLAKATTAPATTRNTGRGGEKKKEKEKEEQEERETKTPECPSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGLLERHSNGSQHSRLTLPRSLWNAGTSVCTLNHPSLPPQRLMALREPAQAQPVKLSLNLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPGSTTFWAWSVLRVPAPPSQPATYTCVSHEDSRTLLNASRSLEVSVDH	IgD-铰链 -Fc
6	LEGGEGRGSLLTCGDVEENPGPR	T2A
7	MLLLVTSLLLCELPHPAFLLIPKVCNGIGIGEFKDSLSINATNIKHFNCTSIGDLHLIPVAFRGDSFTHTPPLDPQELDILKTVEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNLLEGEPREFVENSECICQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMGENNTLVWKYADAGHVCHLCHPNCTYGCCTGPGLEGCPNGPKIPSIATGMVGAALLLVVALGIGLFM	tEGFR
8	FWVLVVVGGVLACYSLLVTVAFIIFWV	CD28(Un iprot P10747的 aa 153-179)
9	IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPFPGPSKP FWVLVVVGGVLACYSLLVTVAFIIFWV	CD28(Un iprot P10747的 aa 114-179)
10	RSKRSRLLHSDYMNMTPRRPGPTRKHYPYAPPRDFAAYRS	CD28(Un iprot P10747的 aa 180-220)
11	RSKRSRGGHSDYMNMTPRRPGPTRKHYPYAPPRDFAAYRS	CD28(LL 至GG)
12	KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEGGCEL	4-1BB(Q 07011.1 的 aa 214-255)
13	RVKFSRSADAPAYQQGQNQLYNELNLRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGDGLYQGLSTATKDTYDALHMQUALPPR	CD3ζ
14	RVKFSRSAEPPAYQQGQNQLYNELNLRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGDGLYQGLSTATKDTYDALHMQUALPPR	CD3ζ
15	RVKFSRSADAPAYKQGQNQLYNELNLRREEYDVLDKRRGRDPEMGGKPRRKNPQE	CD3ζ

	GLYNELQKDKMAEAYSEIGMKGERRRGKGHGDLGYQGLSTATKDTYDALHMQALPPR	
16	RKVCNGIGIGEFKDSLSINATNIKHFKNCTSIGDLHILPVAFRGDSFTHTPPLDP QELDILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGGQFSLAVVSLNI TSLGLRSLKEISGDVVISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENSCKA TGQVCHALCSPEGCWGPEDRCVSCRNVSRGRCVCKCNLLEGEPEFVENSECICQ CHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMGENNTLVWKYADA GHVCHLCHPNCTYGCTGPGLEGCPNGPKIPSIATGMVGALLLLLVLVALGIGLFM	tEGFR
17	EGRGSLLTCDGVEENPGP	T2A
18	GSGATNFSLLKQAGDVEENPGP	P2A
19	ATNFSLLKQAGDVEENPGP	P2A
20	QCTNYALLKLAGDVESNPGP	E2A
21	VKQTLNFDLLKLAGDVESNPGP	F2A
22	-PGGG-(SGGGG)5-P-, 其中P是脯氨酸, G是甘氨酸并且S是丝氨酸	接头
23	GSADDAKKDAAKKDGKS	接头
24	atgcttctcctggtgacaagccttctgctctgtgagttaccacaccagcattcct cctgatccca	GMCSFR α链信号 序列
25	MLLLVTSLLLCELPHPAFLLIP	GMCSFR α链信号 序列
26	MALPVTALLLPLALLLHA	CD8 α信 号肽
27	EPKSCDKTHTCPPCP	铰链
28	ERKCCVECP	铰链
29	ELKTPVLDGTHTCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPEPKSCDTPPP CPRCP	铰链
30	ESKYGPPCPSCP	铰链
31	X ₁ PPX ₂ P X ₁ 是甘氨酸、半胱氨酸或精氨酸 X ₂ 是半胱氨酸或苏氨酸	铰链
32	Tyr Gly Pro Pro Cys Pro Pro Cys Pro	铰链
33	Lys Tyr Gly Pro Pro Cys Pro Pro Cys Pro	铰链
34	Glu Val Val Val Lys Tyr Gly Pro Pro Cys Pro Pro Cys Pro	铰链
35	RASQDISKYLN	CDR L1
36	SRLHSGV	CDR L2
37	GNTLPYTFG	CDR L3
38	DYGVS	CDR H1
39	VIWGSETTYNSALKS	CDR H2
40	YAMDYWG	CDR H3
41	EVKLVQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVWIGSET TYNSALKSRLTI IKDNSKSQVFLKMNSLQTDDETAIYYCAKHYYYGGSYAMDYWGQ GTSVTVSS	VH
42	DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHS GVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEIT	VL
43	DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHS GVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITGSTSG SGKPGSGEGSTKGEVQLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPR KLEWLGVWIGSETTYNSALKSRLTI IKDNSKSQVFLKMNSLQTDDETAIYYCAKH YYYGGSYAMDYWGQGTSVTVSS	scFv
44	KASQNVGTNVA	CDR L1
45	SATYRNS	CDR L2
46	QQYNRYPYT	CDR L3
47	SYWMN	CDR H1
48	QIYPGDGDTNYNGKFKG	CDR H2
49	KTISSVDFYFDY	CDR H3
50	EVKLQQSGAELVRPGSSVKISCKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDG	VH

	DTNYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFCARKTISSVVDIFYFDYW GQGTTVTVSS	
51	DIELTQSPKFMSTSVGDRVSVTCKASQNVGTNVAWYQQKPGQSPKPLIYSATYRNS GVPDRFTGSGSGTDFTLTITNVQSKDLADYFCQQYNRYPYTSGGGTKLEIKR	VL
52	GGGGSGGGSGGGGS	接头
53	EVKLQQSGAELVRPGSSVKISCKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPGDG DTNYNGKFKGQATLTADKSSSTAYMQLSGLTSEDSAVYFCARKTISSVVDIFYFDYW GQGTTVTVSSGGGGSGGGSGGGSDIELTQSPKFMSTSVGDRVSVTCKASQNVGT NVAWYQQKPGQSPKPLIYSATYRNSGVPDRFTGSGSGTDFTLTITNVQSKDLADYF CQQYNRYPYTSGGGTKLEIKR	scFv
54	HYYYGGSYAMDY	HC-CDR 3
55	HTSRLHS	LC-CDR 2
56	QQGNTLPYT	LC-CDR 3
57	gacatccagatgaccagaccacctccagcctgagcgccagcctgggcgaccgggt gaccatcagctgccgggcccagccaggacatcagcaagtacctgaaactggtatcagc agaagcccagcggcaccgtcaagctgctgatctaccacaccagccgctgcacagc ggcgtgccagccgggttagcggcagcggctccggcaccgactacagcctgacat ctccaacctggaacaggaagatatgccacctacttttgccagcagggcaacacac tgccctacacctttggcggcgaacaaagctggaatcaccggcagcacctccggc agcggcaagcctggcagcggcgagggcagcaccaggggcgaggtgaagctgcagga aagcggccctggcctggcggccccagccagagcctgagcgtgacctgcaccgtga gcggcgtgagcctgccgactacggcgtgagctggatccggcagccccccaggaag ggcctggaatggcctggcgtgatctggggcagcagaccacctactacaacagcgc cctgaagagccggctgaccatcatcaaggacaacagcaagagccaggtgttcctga agatgaacagcctgcagaccgacgacaccgccatctactactgcgccaaagcactac tactacggcggcagctacgccatggactactggggccagggcaccagcgtgaccgt gagcagc	编码scFv 的序列
58	GSTSGSGKPGSGEGSTKG	接头
59	CACCTTCACTCTCAGGAACA	CD247(C D3z)靶序 列1
60	GAATGACACCATAGATGAAG	CD247(C D3z)靶序 列2
61	TGAAGAGGATTCCATCCAGC	CD247(C D3z)靶序 列3
62	TCCAGCAGGTAGCAGAGTTT	CD247(C D3z)靶序 列4
63	CACCTTCACTCTCAGGAACAGG	CD247(C D3z)靶序 列 + PAM 1
64	GAATGACACCATAGATGAAGAGG	CD247(C D3z)靶序 列 + PAM 2
65	TGAAGAGGATTCCATCCAGCAGG	CD247(C D3z)靶序 列 + PAM 3
66	TCCAGCAGGTAGCAGAGTTTGGG	CD247(C D3z)靶序

		列 + PAM 4
67	AGACGCCCCCGGTACCAGC	CD247(C D3z)靶序 列5
68	GCTGACTTACGTTATAGAGC	CD247(C D3z)靶序 列6
69	TTTCACCGCGGCCATCCTGC	CD247(C D3z)靶序 列7
70	TAATCGGCAACTGTGCCTGC	CD247(C D3z)靶序 列8
71	CGGAGGCCTACAGTGAGATT	CD247(C D3z)靶序 列9
72	TGGTACCCACCTTCACTCTC	CD247(C D3z)靶序 列10
73	MKWKALFTAAILQAQLPITEAQSFGLLDPKLCYLLDGILFIYGVILTALFLRVKFS RSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKQRRKNPQEGLYN ELQKDKMAEAYSEIGMKGERRRGKGHDLGYQLSTATKDTYDALHMQUALPPR	CD247 亚 型 1 前体 蛋白序列 (NCBI 参 考序列: NP_9321 70.1)
74	tgcttttctcaaaggccccacagtcctccacttccctggggaggtagctgcagaataa aaccagcagagactccttttctcctaaccgtcccggccaccgctgcctcagcctct gcctcccagcctctttctgaggaaaggacaagatgaagtggaaggcgcttttccac cgggccatcctgcaggcacagttgccgattacagaggcacagagctttggcctgc tggatcccaaactctgctacctgctggatggaatcctcttcatctatggtgtcatt ctcactgccttgttccctgagagtgaagttcagcaggagcgcagacgccccgcgta ccagcaggggccagaaccagctctataacgagctcaatctaggacgaagagaggagt acgatgttttgacaagagacgtggccgggaccctgagatggggggaaagccgcag agaaggaagaaccctcaggaaggcctgtacaatgaactgcagaaagataagatggc atggcctttaccagggtctcagtacagccaccaaggacacctacgacgcccctcac atgcaggccctgccccctcgctaacagccaggggatttcaccactcaaaggccaga cctgcagacgcccagattatgagacacaggatgaagcatttacaaccgggttact cttctcagccactgaagtattcccctttatgtacaggatgctttggttatatttag ctccaaaccttcacacacagactgttgtccctgcactctttaaggagtgactcc cagggcttacggccctggccttgggcccctctggtttgcccgtggtgcaggtagacc tgtctcctggcggttcctcgttctccctgggaggcgggcgactgcctctcacagc tgagttggtgagctctgttttgtaaagtccccagagaaagcgcagatgctagacat gcctaatgtctgtatcactctgtgtctgagtggttcaactcctgctgtaaatttg gcttctgttgtcaccttcacctcctttcaaggtaactgtactgggccatgttgtgc ctccctggtgagaggccgggcagaggggcagatggaaaggagcctaggccaggtg caaccaggagctgcagggcatgggaaggtgggcccaggggaggggtcagccag ggcctgaggggcagcgggagcctccctgcctcaggcctctgtgccgaccattga actgtaccatgtgctacagggccagaagatgaacagactgacctgatgagctgt gcacaaagtggcataaaaaacatgtgggttacacagtgatgaataaagtgctgaggag caagaggaggccgttgatcacttcacgctttcagcgaatgacaaaatcatcttg tgaaggcctgcaggaagacccaacacatgggacctataactgccagcggacagt ggcaggacaggaaaaaccgtcaatgtactaggatactgctgctcattacagggc acaggccatggatggaaaacgctctctgctctgctttttttctactgttttaattt atactggcatgctaaggccttccatattttgcataataaatgcttcagtgaaaatgc	CD247 亚 型 1 前体 mRNA 序 列 (NCBI 参 考 序 列 : NM_198 053.2)

	aaaaaaaaa	
75	MKWKALFTAAILQAQLPITEAQSFGLLDPKLCYLLDGILFIYGVILTALFLRVKFS RSADAPAYQQGQNQLYNELNLGRREYDVLDKRRGRDPEMGGKPRRKNPQEGLYNE LQKDKMAEAYSEIGMKGERRRGKGDGLYQGLSTATKDTYDALHMQUALPPR	CD247 亚 型 2 前体 蛋白序列 (NCBI 参 考序列: NP_0007 25.1)
76	tgctttctcaaaggccccacagtcctccacttccctggggaggtagctgcagaataa aaccagcagagactccttttctcctaaccgtcccggccaccgctgcctcagcctct gcctcccagcctctttctgaggaaaggacaagatgaagtggaaggcgtttttcac cgcgcccatcctgcaggcacagttgccgattacagaggcacagagctttggcctgc tggatcccaaactctgctacctgctggatggaatcctcttcatctatggtgtcatt ctcactgccttgttccctgagagtgaagttcagcaggagcgcagacgccccgcgta ccagcagggccagaaccagctctataacgagctcaatctaggacgaagagaggagt acgatgttttgacaagagacgtggccgggaccctgagatggggggaaagccgaga aggaagaaccctcaggaaggcctgtacaatgaactgcagaaagataagatggcgga ggcctacagtgagattgggatgaaaggcgagcgcggaggggcaaggggcacgatg gcctttaccagggtctcagtacagccaccaaggacacctacgacgacctcaccatg caggccctgccccctcgttaacagccaggggatttcaccactcaaaggccagacct gcagacgcccagattatgagacacaggatgaagcatttacaaccgggttccactctt ctcagccactgaagtattcccctttatgtacaggatgctttggttatatttagctc caaaccttcacacacagactgttgtccctgcactctttaaggagtgactcccag ggcttacggccctggccttgggcccctctggtttgcccgtggtgcaggtagacctgt ctcctggcgggttccctcgtttctccctgggaggcggggcgcactgcctctcacagctga gttgttgagtctgtttttaaagtcccagagaaagcgcagatgcttagcacatgccc ctaattgtctgtatcactctgtgtctgagtggttccactcctgctgtaaattttggct ctgtgttcaccttcacctccttcaaggtaactgtactgggccaatgttgtgctc cctggtgagaggccggcagaggggagatggaaggagcctaggccaggtgcaa ccagggagctgcaggggcatgggaagggtgggcccggcaggggaggtcagccagggc ctgagggggcagcgggagcctccctgcctcaggcctctgtgccgcaccattgaact gtaccatgtgctacaggggccaagaagatgaacagactgaccttgatgagctgtgca caaagtggcataaaaaacatgtggttacacagtgtaataaagtgctgcccagcagaa gaggaggccgttgattcacttcacgctttcagcgaatgacaaaatcatctttgtga aggcctcgcaggaagacccaacacatgggacctataactgccagcggacagtgcc aggacaggaaaaaccgctcaatgtactaggatactgctgcgtcattacagggcaca ggccatggatggaaaaagcctctctgctctgcttttttttactgttttaatttata ctggcatgctaaggccttccatatttgcataataaatgcttcagtgaaaaatgcaaa aaaaaaa	CD247 亚 型 2 前体 mRNA 序 列 (NCBI 参 考 序 列 : NM_000 734.3)
77	cgtgaggctccggtgccctcagtgggcagagcgcacatcgcccacagtcccagag aagtgggggggaggggtcggcaattgaaccgggtgcctagagaagggtggcgggggt aaactgggaaagtgatgtcgtgactggctccgctttttcccaggggtgggggag aacgctatataagtgcagtagtcgcgctgaacgttctttttcgcaacgggtttgcc gccagaacacaggttaagtgccgtgtgtggttcccgcgggctggcctctttacggg ttatggcccttgctgcttgaattacttccacgcccctggctgcagtaagtgatt cttgatcccagacttcgggttggaaagtgggtgggagagttcgaggccttgccgtta aggagccccttcgectcgtgcttgagttgaggcctggcctgggcgctggggccgccc gcgtgcgaatctgggtggcacccttcgcccctgtctcgtgcttccgataaagtctcta gccatttaaaatttttgatgacctgctgcgacgctttttttctggcaagatagctt tgtaaatgcccggccaagatctgcacactggatatttcgggtttttggggccgcccgg gagacggggcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcg cggccaccgagaatcggacgggggtagctctcaagctggccggcctgctctgggtgcc tggcctcgcgcccgcctgtatcgcgcccccctgggcccgaaggctggcccggctcgg caccagttgctgagcggaaagatggccgcttcccggccctgctgcagggagctca aaatggaggacgcccgcctcgggagagcgggggggtgagtcacccacacaaaggaa aaggccctttccgtcctcagccctcgttcatgtgactccaaggagtaccggggcgc cgtccaggcacctcgattagttctcagagcttttggagtagctcgtctcttaggttg gggggggggttttatgcatggagtttccccacactgagtggtggagactgaagt taggccagcttggcacttgatgtaattctccttggaaatttgcctttttgagtttg	EF1a 启 动 子 (GenBan k : J04617.1)

	gatcttgggttcattctcaagcctcagacagtggttcaaagttttttcttccatttcaggtgtcgtgaa	
78	CTGACCTCTTCTCTCCCTCCCACAG	人HBB剪接纳体位点
79	TTTCTCTCCACAG	人IgG剪接纳体位点
80	AGATCCCCTGTCTTAGCGGGAGAGTGCTTGGCACTGAGGAGGCAGGGAGTTGGGGGAGAGTTAACCAGATTCTCCCTGTCTAGTTAACTGTCAGATATTGAAATGATCTCATTGACCATCATTGACCTATTGTCTCCCTGTGGGTAGGCCTCAGAGCCACACACCTCAGGCCAGGAGTACCATTTCATCCAGACGTGAACATCTCCCAGGCTTCCAGAGTTCTTGGTTCACACCGGGGCTAACATGGCTGGGCTTCTGCTGCAGTGGCAGGAGCTCTGTGCACAGAGAACAGCCTCATCTGCTCGCCTTGTTCACCTCCCCTCCCATTGCCCCAGGTTCTTTGGCCCCACAGCGGCCACATCTGCCGTGGTGCCAATAGGTTTCCAGGAGCTGGTTGAGGTGGGAGGGAGGGAGGGGTTGTGATCAGGCTGAGGCA TGGGGATTGGATATAGTCTCCGTGTCATGATTTATTTGGTCAGTCAGTCCTAGTGC CACCCTGGGGTAATGGGGATGTGTTCTCGTCACCTGGGCCCTGGCTGACCAGCTTT ATCTCTTGGCACAGAGGCACAGAGCTTTGGCCTGCTGGAT	5'同源臂
81	AGAGTGAAGGTGGGTACCACTGGGCTTTGGGAGGAGGGCACGGGGTCCCCACTTG ATGGATGTTTCAAGGGGCTTGGTCTTGAAGGTCTCAAGCTCGGGTGGTGCCTGG GGCTTGGTATCCAGGAGCAAAGCAAGGACCAGCCAAGTGTGTGCCTTGAGTGGGCT GAGGAGGAGGTGGCAGTGTCTGGCTGAGATGGACAGGGTAGGAGGGAGAGCCTGGT GCTAGGCACCTCCATGACAAGCCGTACAAATGTGTGCACATCAGAGTGTCCCAGGG AAGCGATGCTACTGGTGACAAAGGGGCTTACACTCAGGCAGAGGTCCCTTCTTTCC AAGTGTGAATGAAGGCCATGTTAGCCTTTCTCTTGAAAAGGCCCTTTCTCATCTG TAACTGGGGAGCTGACCAGAGCGTGGGTTTTTCACTTGGTGCCTTGACAGCCCTGG ATTTCTTCGTGGGGCTGGTCATGGTGTGGGCAGAGATTGGAAGAATTTAGGAGTAA AGGGGAGAATCCAGTCCCAGTCATCACTTCACTGCTTCTCAACCCATTTCTTGTTT GAATTTGGACTTTGGCATAATATTTATTTGAAAACCA	3'同源臂
82	GVQVETISPGDGRTPFKRGQTCVVHYTGMLDGGKMDSSRDNRNPKFKFMLGKQEVI RGWEEGVAQMSVGRQAKLTISPDYAYGATGHPGIIIPPHATLVFDVELLKLE	FKBP
83	GVQVETISPGDGRTPFKRGQTCVVHYTGMLDGGKVDSSRDNRNPKFKFMLGKQEVI RGWEEGVAQMSVGRQAKLTISPDYAYGATGHPGIIIPPHATLVFDVELLKLE	FKBP12v36
84	MGSNKSXPKDASQRRR	人 C-Src 酰化基序
85	MGCXC	双重酰化基序
86	CAAX	CAAX 基序
87	CACCUUCACUCUCAGGAACA	CD247(CD3z)靶向结构域1
88	GAAUGACACCAUAGAUGAAG	CD247(CD3z)靶向结构域2
89	UGAAGAGGAUUCCAUCCAGC	CD247(CD3z)靶向结构域3
90	UCCAGCAGGUAGCAGAGUUU	CD247(CD3z)靶向结构域4
91	AGACGCCCCCGGUACCAGC	CD247(CD3z)靶向结构域5
92	GCUGACUUACGUUAUAGAGC	CD247(C

		D3z)靶向结构域6
93	UUUCACCGCGGCCAUCCUGC	CD247(C D3z)靶向结构域7
94	UAAUCGGCAACUGUGCCUGC	CD247(C D3z)靶向结构域8
95	CGGAGGCCUACAGUGAGAUU	CD247(C D3z)靶向结构域9
96	UGGUACCCACCUUCACUCUC	CD247(C D3z)靶向结构域10
97	NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA互补结构域
98	NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCUGAAA AGCAUAGCAAGUAAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA互补结构域
99	NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCUGGAAA CAGCAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA互补结构域
100	NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCUGUUUUGGAAA CAAAACAGCAUAGCAAGUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA互补结构域
101	NNNNNNNNNNNNNNNNNNNGU <u>U</u> UAGAGCUAGAAAUAGCAAGUUAA <u>U</u> AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA
102	NNNNNNNNNNNNNNNNNNNGUU <u>U</u> AAGAGCUAGAAAUAGCAAGUU <u>U</u> AAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA
103	NNNNNNNNNNNNNNNNNNNGU <u>U</u> UAGAGCUAUGCUGU <u>U</u> UGGAAACAA <u>U</u> ACAGCAUAGCAAGUUAA <u>U</u> AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC	示例性gRNA
104	AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU	示例性近端和尾结构域
105	AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC	示例性近端和尾结构域
106	AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCGGAUC	示例性近端和尾结构域
107	AAGGCUAGUCCGUUAUCAACUUGAAAAAGUG	示例性近端和尾结构域
108	AAGGCUAGUCCGUUAUCA	示例性近端和尾结构域
109	AAGGCUAGUCCG	示例性近端和尾结构域
110	NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU	示例性嵌合gRNA

111	NNNNNNNNNNNNNNNNNNNGUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAA GGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGCGAGAUUUUUU	示范性嵌 合gRNA
112	KKPYSIGLDIGTNSVGWAVVTDDYKVPKAKMKVLGNTDKSHIEKNLLGALLFDSGN TAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFHRLSDSFLVTEDEKRG ERHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIKFRGHFL IEGKFDTRNNDVQRLFQEF LAVYDNTFENSSLQEONVQVEEILTDKISKSAKKDRV LKLFPNEKSNRFAEFLKLVGNQADFKKHFELEEKAPLQFSKDTYEEELVLLAQ IGDNYAELFLSAKKLYDSILLSGILTVTDVGTAKPLSASMIQRYNEHQMDLAQLKQ FIRQKLSDKYNEVFSVSKDGYAGYIDGKTNQEAIFYKYLKGLLNKIEGSGYFLDKI EREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFY PFLADNQDRIEKLITFRIP YYVGPLARGKSDFAWLSRKSADKITPWNFDEIVDKESSAEAFINRMTNYDLYLPNQ KVLPKHSLLYEKFVTVNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKD KLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKIL EDIVLTLTLFEDREMIRKRELENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRN KESRKTILDYLI DDGNSNRNFMQLINDDALSFKEEI AKAQVIGETDNLNQVVS DIA GSPAIAKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRNSQQRLKGLT DSIKEFGSQILKEHPVENSQ LQNDRLFLLYLLQNGRDMYTGEELDI DYLSDYDIDHI IPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKF DNLTKAERGGLTDDD KAGFIKRQLVETRQITKHVARILDERFNTETDENNKIRQV KIVTLKSNLVS NFRKEFELYK VREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFV YGDYPHFHGHKENKATAKFFYSNIMNFFKDDVVRTDKNGEIIWKKDEHISNIKKV LSYPQVNI VKKVEEQ TGGFSKESILPKGNSDKLIPRKT KKFYWDTKKYGGFDSPIV AYSLVIADIEKGSKK LKTVKALVGVTIMEKMTFERDPVAFLEKRGYRNVQEENI IKLPKYSLFKLENGRKRLLASARELQKGN EIVLPNHLGTLTYHAKNIHKVDEPKHL DYVDKHKDEFKELLDVVS NFKKYTLAEGNLEKIKELYAQNGEDLKE LASSFINL LTFTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQ SITGLYETRIDLNLKGGD	变形链球 菌Cas9
113	DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRNRI CYLQEIFSNEMAKVDDSFHRLSDSFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFL IEGDLNPDNSDVKLFIQLVQTYNQ LFEENPINASGVDAKAIL SARLSKSRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDYYDDDLNLLAQ IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIP YYVGPLARGNSRFAMTRKSEETITPWNFEVVDK GASAQSFIERMTNFDKNLPNE KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTV KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILE DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTI LDFLKS DGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIAKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKQKNSRERMKRIE EGIKELG SQILKEHPVENTQLQNEKLYLYLLQNGRDMYVDQELDINRLSDYDVDHI VPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREV KVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALI KKYPKLESEFV YGDYKVDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEK GKSKKLSVKEL LGITIMERS SFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFEL ENGRKRLASAGELQKGNELALPSKYV NFLYLASHYEKLGKSPEDNEQKQLFVEQH KHYLDEIEQISEFSKRVI LADANLKD VLSAYNKHRDKPIREQAENIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	酿脓链球 菌Cas9
114	TKPYSIGLDIGTNSVGWAVTTDNYKVP SKKMKVLGNTSKKYIKKNLLGVLLFDSGI TAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQLDSDSFLVPDDKRD SKYPIFGNLVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFL IEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRI LKLFPGEKNSGIFSEFLKLVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGY IGDDYSDVFLKAKKLYDAILLSGFLT VTDNETEAPLSSAMI KRYNEHKEDLALLKE YIRNISLKT YNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKLLAEFEGADYFLEKI DREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFY PFLAKNKERIEKILTFRIP YYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEE	唾液链球 菌Cas9

	<p>KVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTD KDIIEYLHAIYGYDGIELKGIKQFNSSLSTYHDLNLIINDKEFLDSSNEAIEE IIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRHYTGWGLSAKLINGIRDEK SGNTILDYLDIDGINSRNFQLIHDDALSFKKKIQKAQIIGDEDKNIKEVVKSLP GSPAIIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQORLKR EKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLDIDRLS NYDIDHIIPQAFKDNSIDNKVLVSSASNRGKSDVPSLEVVKRRTFWYQLLKSK LISQRKFDNLTKAERGGSPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDEN NRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFFHHAHDAYLNAVVASALLKKYP KLEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEV NEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSK PKPNSNENLVGAKEYLDPKYGAGISNSFTVLVKGTIEKGAKKKITNVLEFQGI SILDRINRDKLNFLEKGYKDIELIIELPKYSLFELSDGSRRLMASILSTNNKR GEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKFEFEELFYIILEFNE NYVGAKKNGKLLNSAFQSWQNSIDELCSSFIGPTGSEKGLFELTSRGSAADFEF LGVKIPRYRDTTPSSLLKDATLIHQSVTGLYETRIDLAKLGE</p>	
115	<p>KKPYTIGLDIGTNSVGWAVLTDQYDLVLRKMKIAGDSEKKQIKKNFWGVRLFDEGQ TAADRRMARTARRRIERRRNRISYLGQIFAEEMSKTDANFFCRLSDSFYVDNEKRN SRHPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFL IEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRK EKLERILKLYPGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYLEDL ESSLALIGDEYAEFVAAKNAYSAVVLSIIITVAETETNAKLSASMIERFDTHEED LGELKAFIKLHLPKHYYEIFSNTTEKHGYAGYIDGKTKQADFYKYMKMTLENIEGAD YFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSL VTFRIPIYFVGPLANGQSEFAWLTRKADGEIRPWNIEEKVDFGKSAVDFIEKMTNKD TYLPKENVLPKHSLCYQKYLIVYNELTKVRYINDQGKTSYFSGQEKEQIFNDLQKQ RKVKKKDLLEFLRNMESHVESPTIEGLEDSFNSSYSTYHDLKLVGIKQEIILDNPVNT EMLNIVKILTVFEDKRMIEKQLQFSDVLDGVVLLKLERRHYTGWGRLSAKLLMG IRDKQSHLTIIDYLMNDDGLNRNLMQLINDSNLSFKSIEKEQVTADKDIQSIVA DLAGSPAIIKKGILQSLKIVDELVSVMGYPPQTIIVEMARENQTTGKGNNSRPRYK SLEKAIKEFGSQILKEHPTDNQELRNNRLYLYYLQNGKDMYTGQDLDIHNLNSYDI DHIVPQSFITDNSIDNLVLTSSAGNREKGDVPPLEIVRKRKVFWEKLYQGNLMSK RKFDYLTKAERGGLEADKARFIHRQLVETRQITKNVANIHLQRFNYEKDDHGNTM KQVRIIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEP EFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKKYLDTVK KVMSYRQMNIVKKTIEQKGEFSKATIKPKGNSSKLI PRKTNWDPMKYGGLDSPNMA YAVVIEYAKGKNKLVFEKKIIRVTIMERKAFEFKDEKAFLEEQGYRQPKVLAKLPKY TLYECEEGRRMLASANEAQKGNQVLPNHLVTLHHAANCEVSDGKSLDYIESNR EMPAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIIKAIQSFVDLMAFNAMG APASLFFETTIERKRYNNLKELLNSTIIYQSITGLYESRRLDD</p>	<p>无害李斯特菌 Cas9</p>
116	<p>MAAFKPN SINYLGLDIGIASVGWAMVEIDEEENPIRLIDLGVRFERAEPKTDG SLAMARRLARSVRRLTRRAHRLRLTRRLKREGVLQAANFDENGLIKSLPNTPWQ LRAAALDRKLTPLEWSAVLLHLIKHRGYSQRKNEGETADKELGALLKGVAGNAHA LQTGDFRTPAELALNKFEKESGHIRNQSDYSHTFSRKDLQAEIILLFEKQKEFGN PHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAPKAAKNTYTAERFIWL TKLNNLRILEQGSRPLTDTERATLMDEPYRKSCLTYAQARKLLGLEDTAFFKGLR YGKDNAEASTLMEMKAYHAISRALKEGLKDKKSPNLSPQLQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKVFQISLALRRIVPLMEQGKRYDEACAEI YGDHYGKKNTEEKIYLPPIPADEIRNPVVLRLALSQARKVINGVRRYGSPIRIHIE TAREVGSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYE QQHGKCLYSGKEINLGRLEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGN QTPYEFYNGKDNSREWQEFKARVETSRFPRSKQRILLQKFDGDFKERNLNDTRY VNRFLCQFVADRMRLTGKGGKRVFASNGQITNLLRGFWGLRQVRAENDRHHDALDAV VVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVM IRVFGKPDGKPEFEEADTLEKLRLLAEKLSRPEAVHEVYVPLFVSRAPNRKMSG QGHMETVKSARLDEGVSVLRVPLTQLKLDLEKMNREPERPKLYEALKARLEAHK DDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTVWVRNHNGIADNATMVRVDV FEKGDKYLLVPIYSWQVAKGILPDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLVE VITKKARMFYGFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFOQYQID ELGKEIRPCRLKRRPPVR</p>	<p>脑膜炎奈瑟球菌 Cas9</p>

<p>117</p>	<p>MDKKYSIGLDIGTNSVGVAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLEN LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI PYYVGPLARGNSRFAMTRKSEETITPWNFEVVDK GASAQSFIERMTNFDKNLNP EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDIL EDIVLTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRI EEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDH IVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRK FDNLTKAERGGSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLTKSLVSDFRKDFQFYKVINNYHHAHDAYLNAVVG TALIKKYPKLESEF VYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVL SMPQVNIKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEK GKSKKLSVKELLGITIMERSSEK NPIDFLEAKGYKEVKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLGSPEDNEQQLFVEQHKHYLDEIEQISEFSKRVI LADANLD KVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSI TGLYETRIDLSQLGGD</p>	<p>酿脓链球 菌Cas9</p>
<p>118</p>	<p>CGTGAGGCTCCGGTGCCGTCAGTGGGCAGAGCGCACATCGCCACAGTCCCGGAG AAGTTGGGGGGAGGGGTCCGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGT AAAC TGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAG ACCCGATATAAGTGCCTAGTCCCGTGAACGTTCTTTTTCGCAACGGGTATTGCC GACAGAACACAGGTAAGTCCCGTGTGTGGTTCGCGGGCCCTGGCCTCTTTACGGG TTATGGCCCTTGCGTGCCTGAATTACTTCCACCTGGCTGCAGTACGTGATTCTTG ATCCCGAGCTTCGGGTGGAAAGTGGGTGGGAGAGTTCGTGGCCTTGCGCTTAAGGA GCCCTTCGCTCGTGTGAGTTGTGGCCTGGCCTGGGCGCTGGGGCCGCCGCT GCGAATCTGGTGGCACCTTCGCGCTGTCTCGCTGCTTCGATAAGTCTCTAGCCA TTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTA AATGCGGGCCAAGATCAGCACACTGGTATTTCCGTTTTTTGGGGCCGCGGGCGGCGA CGGGGCCCGTGCCTCCAGCGCACATGTTCCGGCAGGGCGGGGCTGCGAGCGCGGC CACCGAGAATCGGACGGGGGTAGTCTCAAGCTGCCCGCCCTGCTCTGGTGCCTGGC CTCGCGCCCGCTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGTCCGACC AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGCCCTGCTGCAGGGAGCACAAAAT GGAGGACGCGCGCTCGGGAGAGCGGGCGGGT GAGTACCCACACAAAAGAAAAGG GCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTC CAGGCACCTCGATTAGTTCTCCAGCTTTTGGAGTACGTCGCTTTAGGTTGGGGGG AGGGGTTTTATGCGATGGAGTTCCCCACACTGAGTGGGTGGAGACTGAAGTTAGG CCAGCTTGGCACTTGATGTAATTCTCCTTGAATTTGCCCTTTTGGAGTTTGGATC TTGGTTCAATCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTT CAGG TGTCGTGAAAAC TACCCCTAAAAGCCAAA</p>	<p>Ef1α 启 动子</p>
<p>119</p>	<p>GGATCTGCGATCGCTCCGGTGCCGTCAGTGGGCAGAGCGCACATCGCCACAGTC CCCGAGAAGTTGGGGGAGGGGTCCGCAATTGAACCGGTGCCTAGAGAAGGTGGCG CGGGGTAAACTGGGAAAGTGATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTG GGGGAGAACCGTATATAAGTGCAGTAGTCCCGTGAACGTTCTTTTTCGCAACGGG TTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGGCTCGCATCTCTCTTACGCGC CCGCCGCCCTACCTGAGGCCCATCCACGCCGTTGAGTCGCGTCTTGCCGCTC CCGCCTGTGGTGCCTCTGAAGTCCGCTCCGCCGCTTAGGTAAGTTAAAGCTCAGG TCGAGACCGGGCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCT CTCCACGCTTTGCCTGACCTGCTTGCTCAACTCTACGCTTTGTTTTCGTTTTCTG TTCTGCGCCGTTACAGATCCAAGCTGTGACCGGCGCCTAC</p>	<p>具 有 HTLV1 增强子的 Ef1α启动 子</p>
<p>120</p>	<p>GGATCTGGAGCGACGAATTTTAGTCTACTGAAACAAGCGGGAGACGTGGAGGAAAA CCCTGGACCT</p>	<p>P2A 核苜 酸序列</p>

121	<p>atggataaaaagtacagcatcgggctggacatcgggtacaaaactcagtgggggtgggc cgtgattacggacgagtacaaggtaccctccaaaaatttaaagtgctgggtaaca cggacagacactctataaagaaaaatcttattggagccttgcgttcgactcaggc gagacagccgaagccacaaggttgaagcggaccgccaggaggggtataaccaggag aaagaaccgcataatgctacctgcaagaaatcttcagtaacgagatggcaaaggttg acgatagctttttccatcgctggaagaatcctttcttggttgaggaagacaagaag cacgaacggcaccccatctttggcaatattgtcgacgaagtggcatatcacgaaaa gtacccgactatctaccacctcaggaagaagctgggtggactctaccgataaggcgg acctcagacttatttatttggcactcgcccacatgattaaatttagaggacatttc ttgatcggggcgacctgaaccggacaacagtgacgtcgataagctggtcatcca acttgtgcagacctacaatcaactggttcgaagaaaacctataaatgcttcaggag tcgacgctaaagcaatcctgtccgcgcctctcaaaatctagaagacttgagaat ctgattgctcagttgccggggaaaagaaaaatggattggttggcaacctgatcgc cctcagctcctggactgaccccaaatttcaaaagtaacttcgacctggccgaagacg ctaagctccagctgtccaaggacacatacagatgacgacctcgacaatctgctggcc cagattggggatcagtacccgatctctttttggcagcaaagaacctgtccgacgc catcctggttgagcgatatcttgagagtgaacaccgaaattactaaagcaccctta gcgcatctatgatcaagcggtagcagcgatcatcaggatctgacctgctgaag gctcttgtgaggcaacagctccccgaaaaatacaaggaaatcttctttgaccagag caaaaacggctacgctggctatatagatgggtggggccagtcaggaggaattctata aattcatcaagcccattctcgagaaaatggacggcacagaggagtgtgctggtcaaa cttaacagggaggacctgctgcggaagcagcggacctttgacaacgggtctatccc ccaccagattcatctggcgcaactgcacgcaatcctgaggaggcaggaggattttt atccttttcttaaagataaccgcgagaaaatagaaaagattcttacattcaggatc ccgtactacgtgggacctctcgccggggcaattcacggtttgctggatgacaag gaagtacagaggagactattacaccttggaaacttcgaagaagtgggtggacaaggggtg catctgcccagcttttcatcgagcggatgacaaaattttgacaagaacctcccta gagaaggtgctgcccacaactctctgctctacgagtactttaccgtctacaatga actgactaaagtcaagtacgtcaccgaggggaatgaggaagccggcattccttagtg gagaacagaagaaggcagattgtagacctgttgttcaagaccaacaggaaggtgact gtgaagcaacttaagaagactacttaagaagatcgaatgttttgacagtggtga aatttcaggggttgaagaccgcttcaatgcgtcattggggacttaccatgatcttc tcaagatcataaaggacaagacttctggacaacgaagaaaatgaggatattctc gaagacatcgtcctcaccctgacctgttcgaagacagggaaatgatagaagagcg cttgaaaacctatgccacctcttcgacgataaagttatgaagcagctgaagcgca ggagatacacaggatggggaagattgtcaaggaagctgatcaatggaattagggat aaacagagtggaagaccatactggatttctcctcaaatctgatggcttcgccaatag gaacttcatgcaactgattcacgatgactctcttacctcaaggaggacattcaaa aggctcaggtgagcggcaggagactcccttcatgaacacatcgcaatgtggca ggttccccgctatttaaaagggcactccttcaaaactgtcaaggtgggtggatgaatt ggtcaaggtaatgggcagacataagccagaaaatattgtgacgagatggcccg aaaaccagaccacacagaagggccagaaaaatagtagagagcggatgaagaggatc gaggagggcatcaaagagctgggatctcagattctcaaagaacaccccgtagaaaa cacacagctgcagaacgaaaaattgtacttgtactatctgcagaacggcagagaca tgtacgtcgaccaagaacttgatattaatagactgtccgactatgacgtagaccat atcgtgccccagctcctcctgaaggacgactccattgatacaaaagtcttgacaag aagcgacaagaacaggggtaaaagtgataatgtgccttagcaggagggtgggtgaaaa aaatgaagaactactggcgacagctgcttaatgcaaagctcattacacaacggaag ttcgataatctgacgaaagcagagagaggtggcttgcctgagttggacaagggcagg gtttattaagcggcagctgggtggaactaggcagatcacaaagcagctggcgcaga ttttggacagccggatgaacacaaaatacgcagcaaaaatgataaactgatacagag gtcaaagttatcacgctgaaaagcaagctgggtgtccgattttcgaaagacttcca gttctacaaagttcgcgagattaataactaccatcatgctcacgatgcgtacctga acgctgttgcggaccgcttgataaagaagtacccaaagctggaatccgagttc gtatacggggattacaaagtgtacgatgtgaggaaaaatgatagccaagctccgagca ggagattggaagggccacagctaagtacttcttttatttcaacatcatgaattttt ttaagacggaaattaccctggccaacggagagatcagaaagcggccccttatagag acaaatgggtgaaacaggtgaaatcgtctgggataagggcagggatttcgctactgt gaggaaggtgctgagtatgccacaggtaaatatcgtgaaaaaacggaagtacaga ccggaggattttccaaggaaagcattttgcctaaaagaaactcagacaagctcatc</p>	<p>酿酒链球 菌Cas9密 码子优化 的核酸序 列</p>
-----	--	--

	<p>gaggacatcaccggccgctgaaggaccgcatccagcccgagatcctggaggccct gctgaagcacatcagcttcgacaagttcgtgcagatcagcctgaaggccctgccc gcatcgtgcccctgatggagcagggcaagcgctacgacgaggcctgcccagatc tacggcgaccactacggcaagaagaacaccgaggagaagatctacctcctctat ccccgccgacgagatccgcaaccccggtggtgctgcccgcctgagccaggcccga aggtgatcaacggcgtggtgcccgcctacggcagcccccgcctccacatcgag accgcccgcgaggtgggcaagagcttcaaggaccgcaaggagatcgagaagcgcca ggaggagaaccgcaaggaccgagagaaggccgcccgaagtcccgcgagtacttcc ccaacttcgtgggagagcccaagagcaaggacatcctgaagctgcccgtgacgag cagcagcacggcaagtgcctgtacagcggcaaggagatcaacctgggcccctgaa cgagaagggctacgtggagatcgaccagccctgcccctcagccgacctgggacg acagcttcaacaacaaggtgctggtgctgggagcgcgagaaccagaacaagggcaac cagacccccctacgagtacttcaacggcaaggacaacagccgagtggtgaggagtt caaggcccgcgtggagaccagccgcttccccgcagcaagaagcagcgcacccctgc tgcagaagttcgacgaggacggcttcaaggagcgaacctgaacgacaccgctac gtgaaccgcttccctgtgcccagttcgtggccgaccgcatgcccctgaccggcaaggg caagaagcgcgtgttcgcccagcaacggccagatcaccaacctgctgcccggcttct ggggcctgcccgaaggtgcccgcgagagaacgaccgccaccacgcccctggacgcccgtg gtggtggcctgacgaccgctggccatgacgacagaagatcacccgcttcgtgcccga caagagatgaacgccttcgacggtaaaaccatcgacaaggagaccggcaggggtgc tgcaccagaagaccacttccccagccctgggagttcttcgcccagggaggtgatg atccgctgttcggcaagcccgcagcgaagcccagttcgaggaggccgacacccc cgagaagctgcccaccctgctggccgagaagctgagcagcccctgaggccgtgc acgagtacgtgactcctctgttcgtgagccgcccaccccaaccgcaagatgagcgggt caggggtcacatggagaccgtgaagagcgcgaagcgcctggacgagggcgtgagcgt gctgcccgtgcccctgaccagctgaagctgaaggaccctggagaagatggtgaacc gagcagcgcgagcccaagctgtacgaggccctgaaggcccgcctggaggcccacaag gacgacccccgcaagccttcgcccagcccttctacaagtaacgacaaggccggca ccgcaaccagcaggtgaaggccgtgcccgtggagcaggtgcagaagaccggcgtgt gggtgcccgaaccacaacggcatcgccgacaacgccaccatggtgcccgtggagctg ttcgagaagggcgacaagtaactacctgggtgcccatctacagctggcaggtggccaa gggcatcctgcccagccgcccgtggtgagggcaaggacgaggaggactggcagc tgatcgacgacagcttcaacttcaagttcagcctgacccccaacgacctggtggag gtgatcaccaagaaggcccgcctgttcggctacttccgagctgcccaccgcccac cggcaacatcaacatccgcacccagcctggaccacaagatcggaagaacggca tccctggagggcatcggcgtgaagaccgcccctgagcttccagaagtaccagatcgac gagctgggcaaggagatccgcccctgcccctgaagaagcgcctcctgtgcccga a</p>	
<p>124</p>	<p>MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRFERAEVPKTGD SLAMARRLARSVRRLTRRAHRLLRARRLLKREGVLQAADFENGLIKSLPNTPWQ LRAAALDRKLTPLEWSAVLLHLIKHRGYSQRKNEGETADKELGALLKGVADNAHA LQTGDFRTPAELALNKFEKESGHIRNQGDYSHTFSRKDLQAEILLFEKQKEFGN PHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAPKAAKNTYTAERFIWL TKLNNLRILEQGSERPLTDERATLMDEPYRKSCLTYAQARKLLGLEDTAFFKGLR YGKDNAEASTLMEMKAYHAI SRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKQVQISLKLRRIVPLMEQGKRYDEACAEI YGDHYGKKNTEEKIYLPPIPADEIRNPVVLRLSQAARKVINGVVRVYGSPIRIHIE TAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKRLYE QQHGKCLYSGKEINLGRLEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGN QTPYEYFNGKDNSREWEFKARVETSRFPRSKKQRI LLQKFDEDGFKERNLNDTRY VNRFLCQFVADRMRLTGKGGKRVFASNGQITNLLRGFWGLRKRVAENDRHHALDAV VVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVM IRVFGKPDGKPEFEEADTPEKLRLLAELKSSRPEAVHEVYVTPFVSRAPNRKMSG QGHMETVKSARLDEGVSVLRVPLTQLKLDLEKMNREPKLYEALKARLEAHK DDPAKAFAPFYKYDKAGNRTQQVKAVRVEQVQKTGVVVRNHNHGIADNATMVRVDV FEKGDKYLVPIYSWQVAKGILPDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLVE VITTKARMFYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALS FQKYQID ELGKEIRPCRLKRRPPVR</p>	<p>脑膜炎奈 瑟球菌 Cas9</p>
<p>125</p>	<p>atgaaaaggaactacattctggggctggacatcgggattacaagcgtgggggtatgg gattattgactatgaaacaaggagcgtgatcgacgcaggcgtcagactgttcaagg</p>	<p>金黄色葡 萄球菌</p>

	<p>aggccaacgtggaaaacaatgaggagcggagaagcaagaggggagccaggcgcctg aaacgacgggagaaggcacagaatccagaggggtgaagaaactgctgttcgattacaa cctgctgaccgaccattctgagctgagtggaattaatccttatgaagccagggtga aaggcctgagtcagaagctgtcagaggaagagttttccgcagctctgctgcacctg gctaagcgccgaggagtgataacgtcaatgaggtggaagaggacaccggcaacga gctgtctacaaaggaacagatctcacgcaatagcaaagctctggaagagaagtatg tcgagagctgcagctggaacggctgaagaaagatggcgaggtgagaggggtcaatt aataggttcaagacaagcgactacgtcaaagaagccaagcagctgctgaaagtga gaaggcttaccaccagctggatcagagcttcatcgatacttataatcgacctgctgg agactcgggagaacctactatgagggaccaggagaagggagccccttcggatggaaa gacatcaaggaatggtacgagatgctgatgggacattgcacctattttccagaaga gctgagaagcgtcaagtacgcttataacgcagatctgtacaacgcctgaatgacc tgaacaacctggctcatcaccagggatgaaaacgagaaactggaatactatgagaag ttccagatcatcgaaaacgtgtttaagcagaagaaaaagcctacactgaaacagat tgctaaggagatcctggcaacgaagaggacatcaagggtaccgggtgacaagca ctggaaaaccagagttaccaatctgaaagtgtatcacgatattaaggacatcaca gcacggaaagaaatcatgagaacgcgcaactgctggatcagattgctaagatcct gactatctaccagagctccgaggacatccaggaagagctgactaacctgaacagcg agctgaccagggaagagatcgaaacagattagtaatctgaaggggtacaccggaa cacaacctgtccctgaaagctatcaatctgattctggatgagctgtggcatacaaa cgacaatcagattgcaatctttaaccggctgaagctgggtcccaaaaaggtggacc tgagtcagcagaaagagatcccaaccacactgggtggacgatttcatctgtcacc gtgggtcaagcggagcttcatccagagcatcaaagtgtatcaacgccatcatcaagaa gtacggcctgcccattgatatactatcgagctggctagggagaagaacagcaagg acgcacagaagatgatcaatgagatgcagaaacgaaaccggcagaccaatgaacgc attgaagagattatccgaactaccgggaaagagaacgcaaagtacctgattgaaaa aatcaagctgcacgatatgcaggagggaaagtgtctgtattctctggaggccatcc ccctggaggacctgctgaacaatccattcaactacgaggtcgatcatattatcccc agaagcgtgtccttcgacaattcctttaacaacaaggtgctggtaacgaggaaga gaactctaaaaagggcaataggactcctttccagtagctgtctagttcagattcca agatctcttacgaaaacctttaaaaagcacattctgaaatctggccaaaaggaagggc cgcatcagcaagacaaaaaggagtacctgctggaagagcgggacatcaacagatt ctccgtccagaaggattttatataaccggaatctgggtggacacaagatacgtactc ggggcctgatgaaatctgctgcgactcctatcttccgggtgaacaatctggatgtgaa gtcaagtccatcaacggcgggttcacatctttctgaggcgcaaatggaagttaa aaaggagcgaacaaaggttacaagcaccatgccgaagatgctctgattatcgcaa atgccgacttcatcttaaggagtggaaaaagctggacaaagccaagaaagtgatg gagaaccagatgttcgaagagaagcaggccgaatctatgcccgaatcgagacaga acaggagtacaaggagattttcatcactcctcaccagatcaagcatatcaaggatt tcaaggactacaagtaactctcaccgggtggataaaaagcccaacagagagctgatc aatgacaccctgtatagtacaagaaaagacgataaggggaaataccctgattgtgaa caatctgaacggactgtacgacaaagataatgacaagctgaaaaagctgatcaaca aaagtcccagagaagctgctgatgtaccaccatgatcctcagacatatcagaaactg aagctgattatggagcagtagcggcagcagagaagaaccactgtataagtactatga agagactgggaactacctgaccaagtatagcaaaaaggataatggccccgtgatca agaagatcaagtaactatgggaacaagctgaatgcccatctggacatcacagacgat taccctaacagctcgcaacaaggtgggtcaagctgtcactgaagccatacagattcga tgtctatctggacaacggcgtgtataaaatctgtgactgtcaagaatctggatgtca tcaaaaaggagaactactatgaagtgaatagcaagtgctacgaagaggctaaaaag ctgaaaaagatttagcaaccaggcagagttcatcgccctccttttacaacaacgacct gattaagatcaatggcgaactgtatagggtcatcgggggtgaacaatgatctgctga accgcattgaagtgaatatgattgacatcacttaccgagagtatctggaaaaacatg aatgataagcgcctccctcgattatcaaaaacaattgcctctaagactcagagtat caaaaagtaactcaaccgacattctgggaaacctgtatgaggtgaagagcaaaaagc acctcagattatcaaaaagggc</p>	<p>Cas9 密码子优化的核酸序列</p>
<p>126</p>	<p>MKRNYYILGLDIGITSVGYGII DYETRDVIDAGVRLFKEANVENNEGRRSKRGARRL KRRRRHRIQRVKLLFDYNLLTDHSELGINPYEARVKGLS QKLSEEEFSAALLHL AKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSI NRFKTS DYVKEAKQLLVQKAYHQLDQSFIDTYIDLLETRRTRTYEGPGE GSPFGWK DIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNLNVI TRDENEKLEYEYK</p>	<p>金黄色葡萄球菌 Cas9</p>

	FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKIDIT ARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLNSSELTQEIEIQISNLKGYTGT HNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP VVKRSFIQSIVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNER IEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAI PLEDLLNPFNYEVDHIIP RSVSFDNSFNKVLVKQEENSCKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKG RISKTKKEYLLEERDINRFVQKDF INRNLDVTRYATRGLMNLRSYFRVNNLDVK VKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI IANADFI FKWKKLDKAKKVM ENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELI NDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKL KLIMEQYGDENPLYKYEBETGNLYTKYSKKNNGPVIKKIKYGNKLNALHLDITDD YPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKK LKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDDLNRIEVNMIDITYREYLENM NDRKPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
127	ASTKGPSVFLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPVAV LQSSGLYSLSSVTVPSNFGTQTYTCNVDHKPSNTKVDKTVKCCVECPPCPAP PVAGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHNAKT KPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPRE PQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPMLDSD GSFFLYSKLTVDKSRWQGNVFSCSVMHEALHNHYTQKSLSLSPGK	人 IgG2 Fc(Unipr ot P01859)
128	ASTKGPSVFLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPVAV LQSSGLYSLSSVTVPSSSLGKTYTCNVDHKPSNTKVDKRVESKYGPPCPCPAP EFLGGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSDQEDPEVQFNWYVDGVEVHNAK TKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPR EPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDS DGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK	人 IgG4 Fc(Unipr ot P01861)
129	ESKYGPPCPPCPAPPVAGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSDQEDPEVQF NWYVDGVEVHNAKTKPREEQFQSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSS IEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQP ENNYKTTTPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSL SLGK	间隔子
130	tgcctttatattgtgaaattgtgatgctattgctttatattgttaaccattataagctg caataaacaagttaacaacaacaattgcattcattttatgtttcagggttcagggggg agggtgtgggagggttttttaaa	SV40 多 聚A信号
131	gaacagagaaacaggagaatatgggccaacaggataatctgtggtaagcagttcct gccccggctcagggccaagaacagttggaacagcagaatatgggccaacaggata tctgtggtaagcagttcctgccccggctcagggccaagaacagatgggtccccagat gcggtccccgccctcagcagtttctagagaacctcagatggttccaggggtgcccc aggacctgaaatgacctgtgccttatttgaactaaccaatcagttcgccttctcgc ttctgttcgcgcgcttctgctccccgagctctatataagcagagctcgttttagtga accgtcagatc	MND 启 动子

Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp
 35 40 45
 Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys
 50 55 60
 Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser
 65 70 75 80
 Arg Leu Thr Val Asp Lys Ser Arg Trp Gln Glu Gly Asn Val Phe Ser
 85 90 95
 Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser
 100 105 110
 Leu Ser Leu Ser Leu Gly Lys
 115

<210> 4

<211> 229

<212> PRT

<213> 智人(Homo sapiens)

<220>

<223> 铰链-CH2-CH3间隔子

<400> 4

Glu Ser Lys Tyr Gly Pro Pro Cys Pro Pro Cys Pro Ala Pro Glu Phe
 1 5 10 15
 Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr
 20 25 30
 Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val
 35 40 45
 Ser Gln Glu Asp Pro Glu Val Gln Phe Asn Trp Tyr Val Asp Gly Val
 50 55 60
 Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Phe Asn Ser
 65 70 75 80
 Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu
 85 90 95
 Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Gly Leu Pro Ser
 100 105 110
 Ser Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro
 115 120 125
 Gln Val Tyr Thr Leu Pro Pro Ser Gln Glu Glu Met Thr Lys Asn Gln
 130 135 140
 Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala
 145 150 155 160

Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr
 165 170 175
 Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Arg Leu
 180 185 190
 Thr Val Asp Lys Ser Arg Trp Gln Glu Gly Asn Val Phe Ser Cys Ser
 195 200 205
 Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser
 210 215 220
 Leu Ser Leu Gly Lys
 225
 <210> 5
 <211> 282
 <212> PRT
 <213> 人工序列
 <220>
 <223> IgD-铰链-Fc
 <400> 5
 Arg Trp Pro Glu Ser Pro Lys Ala Gln Ala Ser Ser Val Pro Thr Ala
 1 5 10 15
 Gln Pro Gln Ala Glu Gly Ser Leu Ala Lys Ala Thr Thr Ala Pro Ala
 20 25 30
 Thr Thr Arg Asn Thr Gly Arg Gly Gly Glu Glu Lys Lys Lys Glu Lys
 35 40 45
 Glu Lys Glu Glu Gln Glu Glu Arg Glu Thr Lys Thr Pro Glu Cys Pro
 50 55 60
 Ser His Thr Gln Pro Leu Gly Val Tyr Leu Leu Thr Pro Ala Val Gln
 65 70 75 80
 Asp Leu Trp Leu Arg Asp Lys Ala Thr Phe Thr Cys Phe Val Val Gly
 85 90 95
 Ser Asp Leu Lys Asp Ala His Leu Thr Trp Glu Val Ala Gly Lys Val
 100 105 110
 Pro Thr Gly Gly Val Glu Glu Gly Leu Leu Glu Arg His Ser Asn Gly
 115 120 125
 Ser Gln Ser Gln His Ser Arg Leu Thr Leu Pro Arg Ser Leu Trp Asn
 130 135 140
 Ala Gly Thr Ser Val Thr Cys Thr Leu Asn His Pro Ser Leu Pro Pro
 145 150 155 160
 Gln Arg Leu Met Ala Leu Arg Glu Pro Ala Ala Gln Ala Pro Val Lys
 165 170 175

Leu Ser Leu Asn Leu Leu Ala Ser Ser Asp Pro Pro Glu Ala Ala Ser
 180 185 190
 Trp Leu Leu Cys Glu Val Ser Gly Phe Ser Pro Pro Asn Ile Leu Leu
 195 200 205
 Met Trp Leu Glu Asp Gln Arg Glu Val Asn Thr Ser Gly Phe Ala Pro
 210 215 220
 Ala Arg Pro Pro Pro Gln Pro Gly Ser Thr Thr Phe Trp Ala Trp Ser
 225 230 235 240
 Val Leu Arg Val Pro Ala Pro Pro Ser Pro Gln Pro Ala Thr Tyr Thr
 245 250 255
 Cys Val Val Ser His Glu Asp Ser Arg Thr Leu Leu Asn Ala Ser Arg
 260 265 270
 Ser Leu Glu Val Ser Tyr Val Thr Asp His
 275 280

<210> 6

<211> 24

<212> PRT

<213> 人工序列

<220>

<223> T2A

<400> 6

Leu Glu Gly Gly Gly Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp
 1 5 10 15
 Val Glu Glu Asn Pro Gly Pro Arg
 20

<210> 7

<211> 357

<212> PRT

<213> 人工序列

<220>

<223> tEGFR

<400> 7

Met Leu Leu Leu Val Thr Ser Leu Leu Leu Cys Glu Leu Pro His Pro
 1 5 10 15
 Ala Phe Leu Leu Ile Pro Arg Lys Val Cys Asn Gly Ile Gly Ile Gly
 20 25 30
 Glu Phe Lys Asp Ser Leu Ser Ile Asn Ala Thr Asn Ile Lys His Phe
 35 40 45
 Lys Asn Cys Thr Ser Ile Ser Gly Asp Leu His Ile Leu Pro Val Ala

<210> 8
 <211> 27
 <212> PRT
 <213> 智人 (Homo sapiens)
 <220>
 <223> CD28
 <300>
 <308> Uniprot P10747
 <309> 1989-07-01
 <400> 8
 Phe Trp Val Leu Val Val Val Gly Gly Val Leu Ala Cys Tyr Ser Leu
 1 5 10 15
 Leu Val Thr Val Ala Phe Ile Ile Phe Trp Val
 20 25

<210> 9
 <211> 66
 <212> PRT
 <213> 智人 (Homo sapiens)
 <220>
 <223> CD28
 <300>
 <308> Uniprot P10747
 <309> 1989-07-01
 <400> 9
 Ile Glu Val Met Tyr Pro Pro Pro Tyr Leu Asp Asn Glu Lys Ser Asn
 1 5 10 15
 Gly Thr Ile Ile His Val Lys Gly Lys His Leu Cys Pro Ser Pro Leu
 20 25 30
 Phe Pro Gly Pro Ser Lys Pro Phe Trp Val Leu Val Val Val Gly Gly
 35 40 45
 Val Leu Ala Cys Tyr Ser Leu Leu Val Thr Val Ala Phe Ile Ile Phe
 50 55 60
 Trp Val
 65

<210> 10
 <211> 41
 <212> PRT
 <213> 智人 (Homo sapiens)
 <220>

<223> CD28

<300>

<308> Uniprot P10747

<309> 1989-07-01

<400> 10

Arg Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn Met Thr
 1 5 10 15
 Pro Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr Ala Pro
 20 25 30
 Pro Arg Asp Phe Ala Ala Tyr Arg Ser
 35 40

<210> 11

<211> 41

<212> PRT

<213> 智人 (Homo sapiens)

<220>

<223> CD28 (LL至GG)

<400> 11

Arg Ser Lys Arg Ser Arg Gly Gly His Ser Asp Tyr Met Asn Met Thr
 1 5 10 15
 Pro Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr Ala Pro
 20 25 30
 Pro Arg Asp Phe Ala Ala Tyr Arg Ser
 35 40

<210> 12

<211> 42

<212> PRT

<213> 智人 (Homo sapiens)

<220>

<223> 4-1BB

<300>

<308> Uniprot Q07011.1

<309> 1995-02-01

<400> 12

Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe Lys Gln Pro Phe Met
 1 5 10 15
 Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly Cys Ser Cys Arg Phe
 20 25 30
 Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu

	35	40
<p><210> 13 <211> 112 <212> PRT <213> 智人 (Homo sapiens) <220> <223> CD3ζ <400> 13</p>		
Arg Val Lys Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly		
1	5	10
Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr		15
	20	25
Asp Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys		30
	35	40
Pro Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys		45
	50	55
Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg		60
65	70	75
Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala		80
	85	90
Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg		95
	100	105
		110
<p><210> 14 <211> 112 <212> PRT <213> 智人 (Homo sapiens) <220> <223> CD3ζ <400> 14</p>		
Arg Val Lys Phe Ser Arg Ser Ala Glu Pro Pro Ala Tyr Gln Gln Gly		
1	5	10
Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr		15
	20	25
Asp Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys		30
	35	40
Pro Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys		45
	50	55
Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg		60
65	70	75
		80

Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala
 85 90 95
 Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg
 100 105 110

<210> 15

<211> 112

<212> PRT

<213> 智人 (Homo sapiens)

<220>

<223> CD3ζ

<400> 15

Arg Val Lys Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Lys Gln Gly
 1 5 10 15
 Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr
 20 25 30
 Asp Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys
 35 40 45
 Pro Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys
 50 55 60
 Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg
 65 70 75 80
 Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala
 85 90 95
 Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg
 100 105 110

<210> 16

<211> 335

<212> PRT

<213> 人工序列

<220>

<223> tEGFR

<400> 16

Arg Lys Val Cys Asn Gly Ile Gly Ile Gly Glu Phe Lys Asp Ser Leu
 1 5 10 15
 Ser Ile Asn Ala Thr Asn Ile Lys His Phe Lys Asn Cys Thr Ser Ile
 20 25 30
 Ser Gly Asp Leu His Ile Leu Pro Val Ala Phe Arg Gly Asp Ser Phe
 35 40 45
 Thr His Thr Pro Pro Leu Asp Pro Gln Glu Leu Asp Ile Leu Lys Thr

50	55	60
Val Lys Glu Ile Thr Gly Phe Leu Leu Ile Gln Ala Trp Pro Glu Asn		
65	70	75
Arg Thr Asp Leu His Ala Phe Glu Asn Leu Glu Ile Ile Arg Gly Arg		80
	85	90
Thr Lys Gln His Gly Gln Phe Ser Leu Ala Val Val Ser Leu Asn Ile		95
	100	105
Thr Ser Leu Gly Leu Arg Ser Leu Lys Glu Ile Ser Asp Gly Asp Val		110
	115	120
Ile Ile Ser Gly Asn Lys Asn Leu Cys Tyr Ala Asn Thr Ile Asn Trp		125
	130	140
Lys Lys Leu Phe Gly Thr Ser Gly Gln Lys Thr Lys Ile Ile Ser Asn		
145	150	155
Arg Gly Glu Asn Ser Cys Lys Ala Thr Gly Gln Val Cys His Ala Leu		
	165	170
Cys Ser Pro Glu Gly Cys Trp Gly Pro Glu Pro Arg Asp Cys Val Ser		
	180	185
Cys Arg Asn Val Ser Arg Gly Arg Glu Cys Val Asp Lys Cys Asn Leu		190
	195	200
Leu Glu Gly Glu Pro Arg Glu Phe Val Glu Asn Ser Glu Cys Ile Gln		205
	210	215
Cys His Pro Glu Cys Leu Pro Gln Ala Met Asn Ile Thr Cys Thr Gly		
225	230	235
Arg Gly Pro Asp Asn Cys Ile Gln Cys Ala His Tyr Ile Asp Gly Pro		
	245	250
His Cys Val Lys Thr Cys Pro Ala Gly Val Met Gly Glu Asn Asn Thr		255
	260	265
Leu Val Trp Lys Tyr Ala Asp Ala Gly His Val Cys His Leu Cys His		270
	275	280
Pro Asn Cys Thr Tyr Gly Cys Thr Gly Pro Gly Leu Glu Gly Cys Pro		285
	290	295
Thr Asn Gly Pro Lys Ile Pro Ser Ile Ala Thr Gly Met Val Gly Ala		
305	310	315
Leu Leu Leu Leu Leu Val Val Ala Leu Gly Ile Gly Leu Phe Met		320
	325	330
		335

<210> 17

<211> 18

<212> PRT

<213> 人工序列

<220>

<223> T2A

<400> 17

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro
1 5 10 15

Gly Pro

<210> 18

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> P2A

<400> 18

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> 19

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> P2A

<400> 19

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn
1 5 10 15

Pro Gly Pro

<210> 20

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> E2A

<400> 20

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> 21

<211> 22
 <212> PRT
 <213> 人工序列
 <220>
 <223> F2A
 <400> 21
 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> 22
 <211> 10
 <212> PRT
 <213> 人工序列
 <220>
 <223> 接头
 <220>
 <221> 重复
 <222> (5) ... (9)
 <223> SGGGG重复5次
 <400> 22
 Pro Gly Gly Gly Ser Gly Gly Gly Gly Pro
 1 5 10
 <210> 23
 <211> 17
 <212> PRT
 <213> 人工序列
 <220>
 <223> 接头
 <400> 23
 Gly Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Gly Lys
 1 5 10 15
 Ser
 <210> 24
 <211> 66
 <212> DNA
 <213> 人工序列
 <220>
 <223> GMCSFR α 链信号序列

<400> 24

atgcttctcc tggtgacaag ccttctgctc tgtgagttac cacaccaccagc attcctcctg 60
atccca 66

<210> 25

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> GMCSFR α 链信号序列

<400> 25

Met Leu Leu Leu Val Thr Ser Leu Leu Leu Cys Glu Leu Pro His Pro

1 5 10 15

Ala Phe Leu Leu Ile Pro

20

<210> 26

<211> 18

<212> PRT

<213> 人工序列

<220>

<223> CD8 α 信号肽

<400> 26

Met Ala Leu Pro Val Thr Ala Leu Leu Leu Pro Leu Ala Leu Leu Leu

1 5 10 15

His Ala

<210> 27

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 27

Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro

1 5 10 15

<210> 28

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 28

Glu Arg Lys Cys Cys Val Glu Cys Pro Pro Cys Pro

1 5 10

<210> 29

<211> 61

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 29

Glu Leu Lys Thr Pro Leu Gly Asp Thr His Thr Cys Pro Arg Cys Pro

1 5 10 15

Glu Pro Lys Ser Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Glu

20 25 30

Pro Lys Ser Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Glu Pro

35 40 45

Lys Ser Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro

50 55 60

<210> 30

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 30

Glu Ser Lys Tyr Gly Pro Pro Cys Pro Ser Cys Pro

1 5 10

<210> 31

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<220>

<221> 变体

<222> (1) ... (1)

<223> Xaa是甘氨酸、半胱氨酸或精氨酸

<220>

<221> 变体

<222> (4) ... (4)

<223> Xaa是半胱氨酸或苏氨酸

<400> 31

Xaa Pro Pro Xaa Pro

1 5

<210> 32

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 32

Tyr Gly Pro Pro Cys Pro Pro Cys Pro

1 5

<210> 33

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 33

Lys Tyr Gly Pro Pro Cys Pro Pro Cys Pro

1 5 10

<210> 34

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 铰链

<400> 34

Glu Val Val Val Lys Tyr Gly Pro Pro Cys Pro Pro Cys Pro

1 5 10

<210> 35

<211> 11

<212> PRT

<213> 人工序列

<220>

<223> CDR L1

<400> 35

Arg Ala Ser Gln Asp Ile Ser Lys Tyr Leu Asn

1 5 10

<210> 36

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> CDR L2

<400> 36

Ser Arg Leu His Ser Gly Val

1 5

<210> 37

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> CDR L3

<400> 37

Gly Asn Thr Leu Pro Tyr Thr Phe Gly

1 5

<210> 38

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> CDR H1

<400> 38

Asp Tyr Gly Val Ser

1 5

<210> 39

<211> 16

<212> PRT

<213> 人工序列

<220>

<223> CDR H2

<400> 39

Val Ile Trp Gly Ser Glu Thr Thr Tyr Tyr Asn Ser Ala Leu Lys Ser

1 5 10 15

<210> 40

1	5	10	15
Asp Arg Val Thr Ile Ser Cys Arg Ala Ser Gln Asp Ile Ser Lys Tyr			
	20	25	30
Leu Asn Trp Tyr Gln Gln Lys Pro Asp Gly Thr Val Lys Leu Leu Ile			
	35	40	45
Tyr His Thr Ser Arg Leu His Ser Gly Val Pro Ser Arg Phe Ser Gly			
	50	55	60
Ser Gly Ser Gly Thr Asp Tyr Ser Leu Thr Ile Ser Asn Leu Glu Gln			
65	70	75	80
Glu Asp Ile Ala Thr Tyr Phe Cys Gln Gln Gly Asn Thr Leu Pro Tyr			
	85	90	95
Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Thr			
	100	105	
<210> 43			
<211> 245			
<212> PRT			
<213> 人工序列			
<220>			
<223> scFv			
<400> 43			
Asp Ile Gln Met Thr Gln Thr Thr Ser Ser Leu Ser Ala Ser Leu Gly			
1	5	10	15
Asp Arg Val Thr Ile Ser Cys Arg Ala Ser Gln Asp Ile Ser Lys Tyr			
	20	25	30
Leu Asn Trp Tyr Gln Gln Lys Pro Asp Gly Thr Val Lys Leu Leu Ile			
	35	40	45
Tyr His Thr Ser Arg Leu His Ser Gly Val Pro Ser Arg Phe Ser Gly			
	50	55	60
Ser Gly Ser Gly Thr Asp Tyr Ser Leu Thr Ile Ser Asn Leu Glu Gln			
65	70	75	80
Glu Asp Ile Ala Thr Tyr Phe Cys Gln Gln Gly Asn Thr Leu Pro Tyr			
	85	90	95
Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Thr Gly Ser Thr Ser Gly			
	100	105	110
Ser Gly Lys Pro Gly Ser Gly Glu Gly Ser Thr Lys Gly Glu Val Lys			
	115	120	125
Leu Gln Glu Ser Gly Pro Gly Leu Val Ala Pro Ser Gln Ser Leu Ser			
	130	135	140
Val Thr Cys Thr Val Ser Gly Val Ser Leu Pro Asp Tyr Gly Val Ser			

1 5
 <210> 47
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> CDR H1
 <400> 47
 Ser Tyr Trp Met Asn
 1 5
 <210> 48
 <211> 17
 <212> PRT
 <213> 人工序列
 <220>
 <223> CDR H2
 <400> 48
 Gln Ile Tyr Pro Gly Asp Gly Asp Thr Asn Tyr Asn Gly Lys Phe Lys
 1 5 10 15
 Gly
 <210> 49
 <211> 13
 <212> PRT
 <213> 人工序列
 <220>
 <223> CDR H3
 <400> 49
 Lys Thr Ile Ser Ser Val Val Asp Phe Tyr Phe Asp Tyr
 1 5 10
 <210> 50
 <211> 122
 <212> PRT
 <213> 人工序列
 <220>
 <223> VH
 <400> 50
 Glu Val Lys Leu Gln Gln Ser Gly Ala Glu Leu Val Arg Pro Gly Ser
 1 5 10 15
 Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Ala Phe Ser Ser Tyr

	20		25		30												
Trp	Met	Asn	Trp	Val	Lys	Gln	Arg	Pro	Gly	Gln	Gly	Leu	Glu	Trp	Ile		
	35						40					45					
Gly	Gln	Ile	Tyr	Pro	Gly	Asp	Gly	Asp	Thr	Asn	Tyr	Asn	Gly	Lys	Phe		
	50					55						60					
Lys	Gly	Gln	Ala	Thr	Leu	Thr	Ala	Asp	Lys	Ser	Ser	Ser	Thr	Ala	Tyr		
65					70					75					80		
Met	Gln	Leu	Ser	Gly	Leu	Thr	Ser	Glu	Asp	Ser	Ala	Val	Tyr	Phe	Cys		
				85					90					95			
Ala	Arg	Lys	Thr	Ile	Ser	Ser	Val	Val	Asp	Phe	Tyr	Phe	Asp	Tyr	Trp		
			100					105					110				
Gly	Gln	Gly	Thr	Thr	Val	Thr	Val	Ser	Ser								
	115						120										

<210> 51

<211> 108

<212> PRT

<213> 人工序列

<220>

<223> VL

<400> 51

Asp	Ile	Glu	Leu	Thr	Gln	Ser	Pro	Lys	Phe	Met	Ser	Thr	Ser	Val	Gly		
1				5					10					15			
Asp	Arg	Val	Ser	Val	Thr	Cys	Lys	Ala	Ser	Gln	Asn	Val	Gly	Thr	Asn		
			20					25					30				
Val	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Gln	Ser	Pro	Lys	Pro	Leu	Ile		
			35					40					45				
Tyr	Ser	Ala	Thr	Tyr	Arg	Asn	Ser	Gly	Val	Pro	Asp	Arg	Phe	Thr	Gly		
			50			55						60					
Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	Ile	Thr	Asn	Val	Gln	Ser		
65					70					75				80			
Lys	Asp	Leu	Ala	Asp	Tyr	Phe	Cys	Gln	Gln	Tyr	Asn	Arg	Tyr	Pro	Tyr		
				85					90					95			
Thr	Ser	Gly	Gly	Gly	Thr	Lys	Leu	Glu	Ile	Lys	Arg						
			100					105									

<210> 52

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 接头

<400> 52

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> 53

<211> 245

<212> PRT

<213> 人工序列

<220>

<223> scFv

<400> 53

Glu Val Lys Leu Gln Gln Ser Gly Ala Glu Leu Val Arg Pro Gly Ser
1 5 10 15

Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Ala Phe Ser Ser Tyr
 20 25 30

Trp Met Asn Trp Val Lys Gln Arg Pro Gly Gln Gly Leu Glu Trp Ile
 35 40 45

Gly Gln Ile Tyr Pro Gly Asp Gly Asp Thr Asn Tyr Asn Gly Lys Phe
 50 55 60

Lys Gly Gln Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr
65 70 75 80

Met Gln Leu Ser Gly Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
 85 90 95

Ala Arg Lys Thr Ile Ser Ser Val Val Asp Phe Tyr Phe Asp Tyr Trp
 100 105 110

Gly Gln Gly Thr Thr Val Thr Val Ser Ser Gly Gly Gly Gly Ser Gly
 115 120 125

Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Glu Leu Thr Gln Ser
 130 135 140

Pro Lys Phe Met Ser Thr Ser Val Gly Asp Arg Val Ser Val Thr Cys
145 150 155 160

Lys Ala Ser Gln Asn Val Gly Thr Asn Val Ala Trp Tyr Gln Gln Lys
 165 170 175

Pro Gly Gln Ser Pro Lys Pro Leu Ile Tyr Ser Ala Thr Tyr Arg Asn
 180 185 190

Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe
 195 200 205

Thr Leu Thr Ile Thr Asn Val Gln Ser Lys Asp Leu Ala Asp Tyr Phe
 210 215 220

atcagctgcc gggccagcca ggacatcagc aagtacctga actggtatca gcagaagccc 120
gacggcaccg tcaagctgct gatctaccac accagccggc tgcacagcgg cgtgcccagc 180
cggtttagcg gcagcggctc cggcaccgac tacagcctga ccatctcaa cctggaacag 240
gaagatatcg ccacctactt ttgccagcag ggcaacacac tgcctacac ctttggcggc 300
ggaacaaagc tggaaatcac cggcagcacc tccggcagcg gcaagcctgg cagcggcgag 360
ggcagcacca agggcgaggt gaagctgcag gaaagcggcc ctggcctggt ggccccagc 420
cagagcctga gcgtgacctg caccgtgagc ggcgtgagcc tgcccgacta cggcgtgagc 480
tggatccggc agccccccag gaagggcctg gaatggctgg gcgtgatctg gggcagcgag 540
accacctact acaacagcgc cctgaagagc cggtgacca tcatcaagga caacagcaag 600
agccaggtgt tcctgaagat gaacagcctg cagaccgagc acaccgcat ctactactgc 660
gccaagcact actactacgg cggcagctac gccatggact actggggcca gggcaccagc 720
gtgaccgtga gcagc 735

<210> 58

<211> 18

<212> PRT

<213> 人工序列

<220>

<223> 接头

<400> 58

Gly Ser Thr Ser Gly Ser Gly Lys Pro Gly Ser Gly Glu Gly Ser Thr

1

5

10

15

Lys Gly

<210> 59

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> CD247 (CD3z) 靶序列1

<400> 59

caccttcaact ctcaggaaca 20

<210> 60

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> CD247 (CD3z) 靶序列2

<400> 60

gaatgacacc atagatgaag 20

<210> 61

<211> 20
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列3
<400> 61
tgaagaggat tccatccagc 20
<210> 62
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列4
<400> 62
tccagcaggt agcagagttt 20
<210> 63
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列 + PAM 1
<400> 63
caccttcaact ctcaggaaca gg 22
<210> 64
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列 + PAM 2
<400> 64
gaatgacacc atagatgaag agg 23
<210> 65
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列 + PAM 3
<400> 65
tgaagaggat tccatccagc agg 23

<210> 66
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列 + PAM 4
<400> 66
tccagcaggt agcagagttt ggg 23
<210> 67
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列5
<400> 67
agacgcccc gcgtaccagc 20
<210> 68
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列6
<400> 68
gctgacttac gttatagagc 20
<210> 69
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列7
<400> 69
tttcaccgcg gccatcctgc 20
<210> 70
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶序列8
<400> 70

Gly Gly Lys Pro Gln Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn
 100 105 110
 Glu Leu Gln Lys Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met
 115 120 125
 Lys Gly Glu Arg Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly
 130 135 140
 Leu Ser Thr Ala Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala
 145 150 155 160

Leu Pro Pro Arg

<210> 74

<211> 1690

<212> DNA

<213> 智人 (Homo sapiens)

<220>

<223> CD247亚型1前体mRNA序列

<300>

<308> NM_198053.2

<309> 2020-01-13

<400> 74

tgcttttctca aaggccccac agtcctccac ttcttgggga ggtagctgca gaataaaacc 60
 agcagagact ccttttctcc taaccgtccc ggccaccgct gcctcagcct ctgcctccca 120
 gcctcttttct gagggaaagg acaagatgaa gtggaaggcg cttttcaccg cggccatcct 180
 gcaggcacag ttgccgatta cagaggcaca gagctttggc ctgctggatc ccaaactctg 240
 ctacctgctg gatggaatcc tcttcatcta tgggtgcatt ctactgcct tgttctgag 300
 agtgaagttc agcaggagcg cagacgcccc cgcgtaccag cagggccaga accagctcta 360
 taacgagctc aatctaggac gaagagagga gtacgatgtt ttggacaaga gacgtggccg 420
 ggaccctgag atggggggaa agccgcagag aaggaagaac cctcaggaag gcctgtacaa 480
 tgaactgcag aaagataaga tggcggaggc ctacagttag attgggatga aaggcgagcg 540
 ccggagggggc aaggggcacg atggccttta ccagggtctc agtacagcca ccaaggacac 600
 ctacgacgcc cttcacatgc aggcctgcc cctcgtctaa cagccagggg atttcaccac 660
 tcaaaggcca gacctgcaga cgcccagatt atgagacaca ggatgaagca tttacaacc 720
 ggttcaactct tctcagccac tgaagtattc ccttttatgt acaggatgct ttggttatat 780
 ttagctccaa accttcacac acagactgtt gtcctgcac tctttaaggg agtgtactcc 840
 cagggttac ggccctggcc ttgggcectc tggtttgccg gtggtgcagg tagacctgtc 900
 tcttggcggt tctctgttct ccctgggagg cgggcgcaact gcctctcaca gctgagttgt 960
 tgagtctggt ttgtaaagtc cccagagaaa gcgcagatgc tagcacatgc cctaatgtct 1020
 gtatcaactct gtgtctgagt ggettcactc ctgctgtaaa tttggcttct gttgtcacct 1080
 tcacctcctt tcaaggtaac tgtactgggc catgtttgtc ctccctggtg agagggccgg 1140
 gcagagggggc agatggaaaag gagcctaggc caggtgcaac caggagctg caggggcatg 1200

ggaaggtggg cgggcagggg agggtcagcc agggcctgcg agggcagcgg gaggcctcct 1260
gcctcaggcc tctgtgccgc accattgaac tgtacatgt gctacagggg ccagaagatg 1320
aacagactga ccttgatgag ctgtgcacaa agtggcataa aaaacatgtg gttacacagt 1380
gtgaataaag tgctgcggag caagaggagg ccgttgattc acttcacgct ttcagcgaat 1440
gacaaaatca tctttgtgaa ggctcgcag gaagaccaa cacatgggac ctataactgc 1500
ccagcggaca gtggcaggac aggaaaaacc cgtcaatgta ctaggatact gctgcgtcat 1560
tacagggcac agccatgga tggaaaacgc tctctgctct gctttttttc tactgtttta 1620
atttatactg gcatgctaaa gccttctat tttgcataat aaatgcttca gtgaaaatgc 1680
aaaaaaaaaa 1690

<210> 75

<211> 163

<212> PRT

<213> 智人 (Homo sapiens)

<220>

<223> CD247亚型2前体蛋白序列

<300>

<308> NO_000725.1

<309> 2020-02-20

<400> 75

Met	Lys	Trp	Lys	Ala	Leu	Phe	Thr	Ala	Ala	Ile	Leu	Gln	Ala	Gln	Leu	1	5	10	15
Pro	Ile	Thr	Glu	Ala	Gln	Ser	Phe	Gly	Leu	Leu	Asp	Pro	Lys	Leu	Cys	20	25	30	
Tyr	Leu	Leu	Asp	Gly	Ile	Leu	Phe	Ile	Tyr	Gly	Val	Ile	Leu	Thr	Ala	35	40	45	
Leu	Phe	Leu	Arg	Val	Lys	Phe	Ser	Arg	Ser	Ala	Asp	Ala	Pro	Ala	Tyr	50	55	60	
Gln	Gln	Gly	Gln	Asn	Gln	Leu	Tyr	Asn	Glu	Leu	Asn	Leu	Gly	Arg	Arg	65	70	75	80
Glu	Glu	Tyr	Asp	Val	Leu	Asp	Lys	Arg	Arg	Gly	Arg	Asp	Pro	Glu	Met	85	90	95	
Gly	Gly	Lys	Pro	Arg	Arg	Lys	Asn	Pro	Gln	Glu	Gly	Leu	Tyr	Asn	Glu	100	105	110	
Leu	Gln	Lys	Asp	Lys	Met	Ala	Glu	Ala	Tyr	Ser	Glu	Ile	Gly	Met	Lys	115	120	125	
Gly	Glu	Arg	Arg	Arg	Gly	Lys	Gly	His	Asp	Gly	Leu	Tyr	Gln	Gly	Leu	130	135	140	
Ser	Thr	Ala	Thr	Lys	Asp	Thr	Tyr	Asp	Ala	Leu	His	Met	Gln	Ala	Leu	145	150	155	160

Pro Pro Arg

<210> 76

<211> 1687

<212> DNA

<213> 智人 (Homo sapiens)

<220>

<223> CD247亚型2前体mRNA序列

<400> 76

```

tgcttttctca aaggccccac agtcctccac ttcttgggga ggtagctgca gaataaaacc 60
agcagagact ccttttctcc taaccgtccc ggccaccgct gcctcagcct ctgcctccca 120
gcctcttttct gagggaaaagg acaagatgaa gtggaaggcg cttttcaccg cggccatcct 180
gcaggcacag ttgccgatta cagaggcaca gagctttggc ctgctggatc ccaaactctg 240
ctacctgctg gatggaatcc tcttcateta tgggtgcatt ctcactgcct tgttctctgag 300
agtgaagttc agcaggagcg cagacgcccc cgcgtaccag cagggccaga accagctcta 360
taacgagctc aatctaggac gaagagagga gtacgatgtt ttggacaaga gacgtggccg 420
ggaccctgag atggggggaa agccgagaag gaagaaccct caggaaggcc tgtacaatga 480
actgcagaaa gataagatgg cggaggccta cagtgagatt gggatgaaag gcgagcgccg 540
gaggggcaag gggcacgatg gcctttacca gggctctcagt acagccacca aggacaccta 600
cgacgccctt cacatgcagg ccctgcccc tcgctaacag ccaggggatt tcaccactca 660
aaggccagac ctgcagacgc ccagattatg agacacagga tgaagcattt acaaccgggt 720
tcaactcttct cagccactga agtattcccc tttatgtaca ggatgctttg gttatattta 780
gctccaaacc ttcacacaca gactgttgtc cctgcactct ttaagggagt gtactcccag 840
ggcttacggc cctggccttg ggccctctgg tttgccggtg gtgcaggtag acctgtctcc 900
tggcggttcc tcgttctccc tgggaggcgg gcgcaactgc tctcacagct gagttgttga 960
gtctgttttg taaagtcccc agagaaagcg cagatgctag cacatgccct aatgtctgta 1020
tcaactctgtg tctgagtggc ttcactcctg ctgtaaattt ggcttctggt gtcaccttca 1080
cctcctttca aggtaactgt actgggccat gttgtgcctc cctgggtgaga gggccgggca 1140
gaggggcaga tggaaaggag cctaggccag gtgcaaccag ggagctgcag gggcatggga 1200
aggtgggagg gcaggggagg gtcagccagg gcctgcgagg gcagcgggag cctccctgcc 1260
tcaggcctct gtgccgcacc attgaaactgt accatgtgct acaggggcca gaagatgaac 1320
agactgacct tgatgagctg tgcacaaagt ggcataaaaa acatgtggtt acacagtgtg 1380
aataaagtgc tgcggagcaa gaggaggccg ttgattcact tcacgctttc agcgaatgac 1440
aaaatcatct ttgtgaaggc ctgcagggaa gacccaacac atgggaccta taactgcccc 1500
gcggacagtg gcaggacagg aaaaaccgct caatgtacta ggatactgct gcgtcattac 1560
agggcacagg ccatggatgg aaaacgctct ctgctctgct ttttttctac tgttttaatt 1620
tatactggca tgctaaagcc ttctatttt gcataataaa tgcttcagtg aaaatgcaaa 1680
aaaaaaa 1687

```

<210> 77

<211> 1189

<212> DNA

<213> 智人 (Homo sapiens)

<220>

<223> EF1 α 启动子

<300>

<308> J04617.1

<309> 1994-11-07

<400> 77

```

cgtgaggctc cggtgcccgt cagtgggcag agcgcacatc gccacagtc cccgagaagt 60
tggggggagg ggtcggcaat tgaaccggtg cctagagaag gtggcgcggg gtaaactggg 120
aaagtgatgt cgtgtaactg ctccgccttt ttcccgaggg tgggggagaa ccgtatataa 180
gtgcagtagt cgccgtgaac gttctttttc gcaacgggtt tgccgccaga acacaggtaa 240
gtgccgtgtg tggttcccgc gggcctggcc tctttacggg ttatggcctt tgcgtgcctt 300
gaattacttc cacgcccctg gctgcagtac gtgattcttg atcccgagct tcgggttggg 360
agtgggtggg agagttcgag gccttgcgct taaggagccc ctctgcctcg tgcttgagtt 420
gaggcctggc ctgggcgctg gggccgccgc gtgcgaatct ggtggcacct tcgcgcctgt 480
ctcgtgctt tcgataagtc tctagccatt taaaattttt gatgacctgc tgcgacgctt 540
tttttctggc aagatagtct tgtaaatgcg ggccaagatc tgcacactgg tatttcggtt 600
tttggggccg cgggcggcga cggggcccgt gcgtcccagc gcacatgttc ggcgaggcgg 660
ggcctgagag cgcgccacc gagaatcgga cgggggtagt ctcaagctgg ccggcctgct 720
ctgggtgcctg gcctcgcgcc gccgtgtatc gccccgcctt gggcggcaag gctggcccgg 780
tcggcaccag ttgcgtgagc ggaaagatgg ccgcttcccg gccctgctgc aggagactca 840
aaatggagga cgcggcgctc gggagagcgg gcgggtgagt cacccacaca aaggaaaagg 900
gcctttccgt cctcagccgt cgcttcatgt gactccacgg agtaccgggc gccgtccagg 960
cacctcgatt agttctcgag cttttggagt acgtcgtctt taggttgggg ggaggggttt 1020
tatgcatgag agtttcccca cactgagtgg gtggagactg aagttaggcc agcttggcac 1080
ttgatgtaat tctccttggg atttgccctt ttgagtttg gatcttggtt cattctcaag 1140
cctcagacag tggttcaaag ttttttctt ccatttcagg tgtcgtgaa 1189

```

<210> 78

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> 人HBB剪接接纳体位点

<400> 78

```
ctgacctctt ctcttcctcc cacag 25
```

<210> 79

<211> 13

<212> DNA

<213> 人工序列

<220>

<223> 人IgG剪接接纳体位点

<400> 79

tttctctcca cag 13

<210> 80

<211> 600

<212> DNA

<213> 人工序列

<220>

<223> 5'同源臂

<400> 80

agatcccact gtccctaggcg ggagagtgct tggcactgag gaggcaggga gttgggggag 60
 agttaacca gattctccct gtccctagtta actgtcagat attgaaatga tctcatttga 120
 ccatcatttg acctattgtc tcctgtggg taggcctcag agccacacac ctcaggccag 180
 gagtaccatt catccagacg tgaacatctt cccgaggctt ccagagttct tggttcacac 240
 cggggctaac atggctgggc ttctgtgca gtggcaggag ctctgtgcac agagaacagc 300
 ctcatctgct cgccttgttt ccacctccc tccattgcc ccaggttctt tggccccaca 360
 gcggccacat ctgccgttgg tgccaatagg ttttcagga gctggttgag gtgggaggga 420
 gggagagggt tgtgatcagg ctgaggcatg gggattggat atagtctccg tgtcatgatt 480
 tatttggta gtcagtccta gtgccaccct ggggtaatgg ggatgtgttc tcgtcacctt 540
 gggcctggct gaccagcttt atctcttggc acagaggcac agagctttgg cctgctggat 600

<210> 81

<211> 600

<212> DNA

<213> 人工序列

<220>

<223> 3'同源臂

<400> 81

agagtgaagg tgggtaccac tgggctttgg gaggaggga cggggtcccc cacttgatgg 60
 atgttcagag gggccttggc cttggaaggt ctcaagctcg ggtggtgcct ggggcttggc 120
 atccaggagc aaagcaagga ccagccaagt gtgtgccttg agtgggctga ggaggaggtg 180
 gcagtgtctg gctgagatgg acaggtagg agggagagcc tgggtctagg cacctccatg 240
 acaagccgta caaatgtgtg cacatcagag tgtcccaggg aaggcgatgc tactggtgac 300
 aaaggggctt aactcaggc agaggtcctt ctttccaagt gtgaatgaag gccatgttag 360
 cctttctctt gaaaaggccc tttctctcgc tgtaactggg gagctgacca gacgtgggt 420
 ttttcaactg gtgccttgcg gaccctggat ttcttcgtgg ggctggtcat ggtgtgggca 480
 gagattggaa gaatttagga gtaaagggga gaatccagtc ccagtcatca cttcagtgtc 540
 tctcaacca tttcttggtt gaattttgga ctttggcata atatatttatt tgaaaaacca 600

<210> 82

<211> 107

<212> PRT

<213> 人工序列

<220>

<223> FKBP

<400> 82

Gly	Val	Gln	Val	Glu	Thr	Ile	Ser	Pro	Gly	Asp	Gly	Arg	Thr	Phe	Pro
1				5					10					15	
Lys	Arg	Gly	Gln	Thr	Cys	Val	Val	His	Tyr	Thr	Gly	Met	Leu	Glu	Asp
			20					25					30		
Gly	Lys	Lys	Met	Asp	Ser	Ser	Arg	Asp	Arg	Asn	Lys	Pro	Phe	Lys	Phe
			35				40					45			
Met	Leu	Gly	Lys	Gln	Glu	Val	Ile	Arg	Gly	Trp	Glu	Glu	Gly	Val	Ala
			50				55				60				
Gln	Met	Ser	Val	Gly	Gln	Arg	Ala	Lys	Leu	Thr	Ile	Ser	Pro	Asp	Tyr
65					70					75					80
Ala	Tyr	Gly	Ala	Thr	Gly	His	Pro	Gly	Ile	Ile	Pro	Pro	His	Ala	Thr
					85					90					95
Leu	Val	Phe	Asp	Val	Glu	Leu	Leu	Lys	Leu	Glu					
					100					105					

<210> 83

<211> 107

<212> PRT

<213> 人工序列

<220>

<223> FKBP12v36

<400> 83

Gly	Val	Gln	Val	Glu	Thr	Ile	Ser	Pro	Gly	Asp	Gly	Arg	Thr	Phe	Pro
1				5					10					15	
Lys	Arg	Gly	Gln	Thr	Cys	Val	Val	His	Tyr	Thr	Gly	Met	Leu	Glu	Asp
			20					25					30		
Gly	Lys	Lys	Val	Asp	Ser	Ser	Arg	Asp	Arg	Asn	Lys	Pro	Phe	Lys	Phe
			35				40					45			
Met	Leu	Gly	Lys	Gln	Glu	Val	Ile	Arg	Gly	Trp	Glu	Glu	Gly	Val	Ala
			50				55				60				
Gln	Met	Ser	Val	Gly	Gln	Arg	Ala	Lys	Leu	Thr	Ile	Ser	Pro	Asp	Tyr
65					70					75					80
Ala	Tyr	Gly	Ala	Thr	Gly	His	Pro	Gly	Ile	Ile	Pro	Pro	His	Ala	Thr

	85	90	95
Leu Val Phe Asp Val Glu Leu Leu Lys Leu Glu			
	100	105	
<210> 84			
<211> 16			
<212> PRT			
<213> 人工序列			
<220>			
<223> 人C-Src酰化基序			
<400> 84			
Met Gly Ser Asn Lys Ser Lys Pro Lys Asp Ala Ser Gln Arg Arg Arg			
1	5	10	15
<210> 85			
<211> 5			
<212> PRT			
<213> 人工序列			
<220>			
<223> 双重酰化基序			
<220>			
<221> 变体			
<222> (4) ... (4)			
<223> Xaa是任何氨基酸			
<400> 85			
Met Gly Cys Xaa Cys			
1	5		
<210> 86			
<211> 4			
<212> PRT			
<213> 人工序列			
<220>			
<223> CAAX基序			
<220>			
<221> 变体			
<222> (4) ... (4)			
<223> Xaa是任何氨基酸			
<400> 86			
Cys Ala Ala Xaa			
1			
<210> 87			

<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域1
<400> 87
caccuucacu cucaggaaca 20
<210> 88
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域2
<400> 88
gaaugacacc auagaugaag 20
<210> 89
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域3
<400> 89
ugaagaggau uccauccagc 20
<210> 90
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域4
<400> 90
uccagcaggu agcagaguuu 20
<210> 91
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域5
<400> 91
agacgcccc gcguaccagc 20

<210> 92
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域6
<400> 92
gcugacuuac guuauagagc 20
<210> 93
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域7
<400> 93
uuucaccgcg gccauccugc 20
<210> 94
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域8
<400> 94
uaaucggcaa cugugccugc 20
<210> 95
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域9
<400> 95
cggaggccua cagugagauu 20
<210> 96
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> CD247 (CD3z) 靶向结构域10
<400> 96

ugguacccac cuucacucuc 20
 <210> 97
 <211> 96
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性gRNA互补结构域
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> a、c、u、g、未知或其他
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> n是a、c、g或u
 <400> 97
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugc 96
 <210> 98
 <211> 104
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性gRNA互补结构域
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> a、c、u、g、未知或其他
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> n是a、c、g或u
 <400> 98
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcugaaa agcauagcaa guuaaaauaa 60
 ggcuaguccg uuaucaacuu gaaaaagugg caccgagucg gugc 104
 <210> 99
 <211> 106
 <212> RNA
 <213> 人工序列

<220>
 <223> 示例性gRNA互补结构域
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> a、c、u、g、未知或其他
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> n是a、c、g或u
 <400> 99
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuggaa acagcauagc aaguuaaaau 60
 aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 106
 <210> 100
 <211> 116
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性gRNA互补结构域
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> a、c、u、g、未知或其他
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> n是a、c、g或u
 <400> 100
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu uggaacaaa acagcauagc 60
 aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 116
 <210> 101
 <211> 96
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性gRNA
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)

- <223> a、c、u、g、未知或其他
<220>
<221> 修饰的碱基
<222> (1) ... (20)
<223> n是a、c、g或u
<400> 101
nnnnnnnnnn nnnnnnnnnn guauuagagc uagaaauagc aaguuaauau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugc 96
<210> 102
<211> 96
<212> RNA
<213> 人工序列
<220>
<223> 示例性gRNA
<220>
<221> 修饰的碱基
<222> (1) ... (20)
<223> a、c、u、g、未知或其他
<220>
<221> 修饰的碱基
<222> (1) ... (20)
<223> n是a、c、g或u
<400> 102
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuuuuuu aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugc 96
<210> 103
<211> 116
<212> RNA
<213> 人工序列
<220>
<223> 示例性gRNA
<220>
<221> 修饰的碱基
<222> (1) ... (20)
<223> a、c、u、g、未知或其他
<220>
<221> 修饰的碱基
<222> (1) ... (20)
<223> n是a、c、g或u

- <400> 103
nnnnnnnnnn nnnnnnnnnn guauuagagc uaugcuguau uggaacaau acagcauagc 60
aaguuauau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 116
- <210> 104
<211> 47
<212> RNA
<213> 人工序列
<220>
<223> 示例性近端和尾结构域
- <400> 104
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcu 47
- <210> 105
<211> 49
<212> RNA
<213> 人工序列
<220>
<223> 示例性近端和尾结构域
- <400> 105
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cgguggugc 49
- <210> 106
<211> 51
<212> RNA
<213> 人工序列
<220>
<223> 示例性近端和尾结构域
- <400> 106
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcggau c 51
- <210> 107
<211> 31
<212> RNA
<213> 人工序列
<220>
<223> 示例性近端和尾结构域
- <400> 107
aaggcuaguc cguuaucaac uugaaaaagu g 31
- <210> 108
<211> 18
<212> RNA
<213> 人工序列

<220>
 <223> 示例性近端和尾结构域
 <400> 108
 aaggcuaguc cguuauca 18
 <210> 109
 <211> 12
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性近端和尾结构域
 <400> 109
 aaggcuaguc cg 12
 <210> 110
 <211> 102
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性嵌合gRNA
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> a、c、u、g、未知或其他
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> n是a、c、g或u
 <400> 110
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uu 102
 <210> 111
 <211> 102
 <212> RNA
 <213> 人工序列
 <220>
 <223> 示例性嵌合gRNA
 <220>
 <221> 修饰的碱基
 <222> (1) ... (20)
 <223> a、c、u、g、未知或其他

195	200	205
Glu Ile Leu Thr Asp Lys Ile Ser Lys Ser Ala Lys Lys Asp Arg Val		
210	215	220
Leu Lys Leu Phe Pro Asn Glu Lys Ser Asn Gly Arg Phe Ala Glu Phe		
225	230	235
240		
Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe Glu		
245	250	255
Leu Glu Glu Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Glu Glu		
260	265	270
Glu Leu Glu Val Leu Leu Ala Gln Ile Gly Asp Asn Tyr Ala Glu Leu		
275	280	285
Phe Leu Ser Ala Lys Lys Leu Tyr Asp Ser Ile Leu Leu Ser Gly Ile		
290	295	300
Leu Thr Val Thr Asp Val Gly Thr Lys Ala Pro Leu Ser Ala Ser Met		
305	310	315
320		
Ile Gln Arg Tyr Asn Glu His Gln Met Asp Leu Ala Gln Leu Lys Gln		
325	330	335
Phe Ile Arg Gln Lys Leu Ser Asp Lys Tyr Asn Glu Val Phe Ser Asp		
340	345	350
Val Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn Gln		
355	360	365
Glu Ala Phe Tyr Lys Tyr Leu Lys Gly Leu Leu Asn Lys Ile Glu Gly		
370	375	380
Ser Gly Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg Lys		
385	390	395
400		
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gln		
405	410	415
Glu Met Arg Ala Ile Ile Arg Arg Gln Ala Glu Phe Tyr Pro Phe Leu		
420	425	430
Ala Asp Asn Gln Asp Arg Ile Glu Lys Leu Leu Thr Phe Arg Ile Pro		
435	440	445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Lys Ser Asp Phe Ala Trp Leu		
450	455	460
Ser Arg Lys Ser Ala Asp Lys Ile Thr Pro Trp Asn Phe Asp Glu Ile		
465	470	475
480		
Val Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr Asn		
485	490	495
Tyr Asp Leu Tyr Leu Pro Asn Gln Lys Val Leu Pro Lys His Ser Leu		
500	505	510

Leu Tyr Glu Lys Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
 515 520 525
 Lys Thr Glu Gln Gly Lys Thr Ala Phe Phe Asp Ala Asn Met Lys Gln
 530 535 540
 Glu Ile Phe Asp Gly Val Phe Lys Val Tyr Arg Lys Val Thr Lys Asp
 545 550 555 560
 Lys Leu Met Asp Phe Leu Glu Lys Glu Phe Asp Glu Phe Arg Ile Val
 565 570 575
 Asp Leu Thr Gly Leu Asp Lys Glu Asn Lys Val Phe Asn Ala Ser Tyr
 580 585 590
 Gly Thr Tyr His Asp Leu Cys Lys Ile Leu Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Ser Lys Asn Glu Lys Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Arg Lys Arg Leu Glu Asn Tyr Ser
 625 630 635 640
 Asp Leu Leu Thr Lys Glu Gln Val Lys Lys Leu Glu Arg Arg His Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Ala Glu Leu Ile His Gly Ile Arg Asn
 660 665 670
 Lys Glu Ser Arg Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Asn
 675 680 685
 Ser Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Ala Leu Ser Phe
 690 695 700
 Lys Glu Glu Ile Ala Lys Ala Gln Val Ile Gly Glu Thr Asp Asn Leu
 705 710 715 720
 Asn Gln Val Val Ser Asp Ile Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735
 Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val Lys Ile Met Gly
 740 745 750
 His Gln Pro Glu Asn Ile Val Val Glu Met Ala Arg Glu Asn Gln Phe
 755 760 765
 Thr Asn Gln Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys Gly Leu Thr
 770 775 780
 Asp Ser Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys Glu His Pro Val
 785 790 795 800
 Glu Asn Ser Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Leu Gln
 805 810 815
 Asn Gly Arg Asp Met Tyr Thr Gly Glu Glu Leu Asp Ile Asp Tyr Leu

820	825	830
Ser Gln Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp		
835	840	845
Asn Ser Ile Asp Asn Arg Val Leu Thr Ser Ser Lys Glu Asn Arg Gly		
850	855	860
Lys Ser Asp Asp Val Pro Ser Lys Asp Val Val Arg Lys Met Lys Ser		
865	870	875
Tyr Trp Ser Lys Leu Leu Ser Ala Lys Leu Ile Thr Gln Arg Lys Phe		
885	890	895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr Asp Asp Asp Lys		
900	905	910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys		
915	920	925
His Val Ala Arg Ile Leu Asp Glu Arg Phe Asn Thr Glu Thr Asp Glu		
930	935	940
Asn Asn Lys Lys Ile Arg Gln Val Lys Ile Val Thr Leu Lys Ser Asn		
945	950	955
Leu Val Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg Glu		
965	970	975
Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile		
980	985	990
Gly Lys Ala Leu Leu Gly Val Tyr Pro Gln Leu Glu Pro Glu Phe Val		
995	1000	1005
Tyr Gly Asp Tyr Pro His Phe His Gly His Lys Glu Asn Lys Ala Thr		
1010	1015	1020
Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Lys Asp		
1025	1030	1035
Asp Val Arg Thr Asp Lys Asn Gly Glu Ile Ile Trp Lys Lys Asp Glu		
1045	1050	1055
His Ile Ser Asn Ile Lys Lys Val Leu Ser Tyr Pro Gln Val Asn Ile		
1060	1065	1070
Val Lys Lys Val Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile		
1075	1080	1085
Leu Pro Lys Gly Asn Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys Lys		
1090	1095	1100
Phe Tyr Trp Asp Thr Lys Lys Tyr Gly Gly Phe Asp Ser Pro Ile Val		
1105	1110	1115
Ala Tyr Ser Ile Leu Val Ile Ala Asp Ile Glu Lys Gly Lys Ser Lys		
1125	1130	1135

Lys Leu Lys Thr Val Lys Ala Leu Val Gly Val Thr Ile Met Glu Lys
 1140 1145 1150
 Met Thr Phe Glu Arg Asp Pro Val Ala Phe Leu Glu Arg Lys Gly Tyr
 1155 1160 1165
 Arg Asn Val Gln Glu Glu Asn Ile Ile Lys Leu Pro Lys Tyr Ser Leu
 1170 1175 1180
 Phe Lys Leu Glu Asn Gly Arg Lys Arg Leu Leu Ala Ser Ala Arg Glu
 1185 1190 1195 1200
 Leu Gln Lys Gly Asn Glu Ile Val Leu Pro Asn His Leu Gly Thr Leu
 1205 1210 1215
 Leu Tyr His Ala Lys Asn Ile His Lys Val Asp Glu Pro Lys His Leu
 1220 1225 1230
 Asp Tyr Val Asp Lys His Lys Asp Glu Phe Lys Glu Leu Leu Asp Val
 1235 1240 1245
 Val Ser Asn Phe Ser Lys Lys Tyr Thr Leu Ala Glu Gly Asn Leu Glu
 1250 1255 1260
 Lys Ile Lys Glu Leu Tyr Ala Gln Asn Asn Gly Glu Asp Leu Lys Glu
 1265 1270 1275 1280
 Leu Ala Ser Ser Phe Ile Asn Leu Leu Thr Phe Thr Ala Ile Gly Ala
 1285 1290 1295
 Pro Ala Thr Phe Lys Phe Phe Asp Lys Asn Ile Asp Arg Lys Arg Tyr
 1300 1305 1310
 Thr Ser Thr Thr Glu Ile Leu Asn Ala Thr Leu Ile His Gln Ser Ile
 1315 1320 1325
 Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Asn Lys Leu Gly Gly Asp
 1330 1335 1340
 <210> 113
 <211> 1367
 <212> PRT
 <213> 酿脓链球菌
 <220>
 <223> Cas9
 <400> 113
 Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
 1 5 10 15
 Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
 20 25 30
 Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
 35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
 50 55 60
 Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
 65 70 75 80
 Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
 85 90 95
 Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
 100 105 110
 Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
 115 120 125
 Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
 130 135 140
 Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
 145 150 155 160
 Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
 165 170 175
 Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
 180 185 190
 Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
 195 200 205
 Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
 210 215 220
 Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
 225 230 235 240
 Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
 245 250 255
 Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
 260 265 270
 Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
 275 280 285
 Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
 290 295 300
 Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
 305 310 315 320
 Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
 325 330 335
 Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
 340 345 350
 Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355	360	365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly		
370	375	380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys		
385	390	395
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly		
405	410	415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu		
420	425	430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro		
435	440	445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met		
450	455	460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val		
465	470	475
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn		
485	490	495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu		
500	505	510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr		
515	520	525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys		
530	535	540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val		
545	550	555
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser		
565	570	575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr		
580	585	590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn		
595	600	605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu		
610	615	620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His		
625	630	635
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr		
645	650	655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys		
660	665	670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
 690 695 700
 Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
 705 710 715 720
 Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
 725 730 735
 Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
 740 745 750
 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
 755 760 765
 Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
 770 775 780
 Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
 785 790 795 800
 Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
 805 810 815
 Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
 820 825 830
 Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
 835 840 845
 Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
 850 855 860
 Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
 865 870 875 880
 Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
 885 890 895
 Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
 900 905 910
 Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
 915 920 925
 His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
 930 935 940
 Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
 945 950 955 960
 Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
 965 970 975
 Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

	980	985	990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val			
	995	1000	1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser			
	1010	1015	1020
Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn			
1025	1030	1035	1040
Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile			
	1045	1050	1055
Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val			
	1060	1065	1070
Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met			
	1075	1080	1085
Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe			
1090	1095	1100	
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala			
1105	1110	1115	1120
Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro			
	1125	1130	1135
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys			
	1140	1145	1150
Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met			
	1155	1160	1165
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys			
1170	1175	1180	
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr			
1185	1190	1195	1200
Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala			
	1205	1210	1215
Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val			
	1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro			
	1235	1240	1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr			
1250	1255	1260	
Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile			
1265	1270	1275	1280
Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His			
	1285	1290	1295

Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe
 1300 1305 1310
 Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
 1315 1320 1325
 Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
 1330 1335 1340
 Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
 1345 1350 1355 1360
 Leu Ser Gln Leu Gly Gly Asp
 1365
 <210> 114
 <211> 1387
 <212> PRT
 <213> 唾液链球菌
 <220>
 <223> Cas9
 <400> 114
 Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
 1 5 10 15
 Trp Ala Val Thr Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met Lys
 20 25 30
 Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile Lys Lys Asn Leu Leu Gly
 35 40 45
 Val Leu Leu Phe Asp Ser Gly Ile Thr Ala Glu Gly Arg Arg Leu Lys
 50 55 60
 Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu Tyr
 65 70 75 80
 Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala Phe
 85 90 95
 Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg Asp
 100 105 110
 Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Ala Tyr His
 115 120 125
 Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp Ser
 130 135 140
 Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His Met
 145 150 155 160
 Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser Lys
 165 170 175

Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr Asn
 180 185 190
 Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu Glu
 195 200 205
 Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg Ile
 210 215 220
 Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu Phe
 225 230 235 240
 Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe Asn
 245 250 255
 Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp Glu
 260 265 270
 Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp Val
 275 280 285
 Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly Phe
 290 295 300
 Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala Met
 305 310 315 320
 Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys Glu
 325 330 335
 Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys Asp
 340 345 350
 Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn Gln
 355 360 365
 Glu Asp Phe Tyr Val Tyr Leu Lys Lys Leu Leu Ala Glu Phe Glu Gly
 370 375 380
 Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg Lys
 385 390 395 400
 Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu Gln
 405 410 415
 Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe Leu
 420 425 430
 Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
 435 440 445
 Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp Ser
 450 455 460
 Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp Val
 465 470 475 480
 Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr Ser

	485	490	495
Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser Leu			
	500	505	510
Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu Leu Thr Lys Val Arg Phe			
	515	520	525
Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe Leu Asp Ser Lys Gln Lys			
	530	535	540
Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp Lys Arg Lys Val Thr Asp			
545	550	555	560
Lys Asp Ile Ile Glu Tyr Leu His Ala Ile Tyr Gly Tyr Asp Gly Ile			
	565	570	575
Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ser Ser Leu Ser Thr Tyr			
	580	585	590
His Asp Leu Leu Asn Ile Ile Asn Asp Lys Glu Phe Leu Asp Asp Ser			
	595	600	605
Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile His Thr Leu Thr Ile Phe			
	610	615	620
Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ser Lys Phe Glu Asn Ile			
625	630	635	640
Phe Asp Lys Ser Val Leu Lys Lys Leu Ser Arg Arg His Tyr Thr Gly			
	645	650	655
Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Glu Lys			
	660	665	670
Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Ile Ser Asn			
	675	680	685
Arg Asn Phe Met Gln Leu Ile His Asp Asp Ala Leu Ser Phe Lys Lys			
	690	695	700
Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp Glu Asp Lys Gly Asn Ile			
705	710	715	720
Lys Glu Val Val Lys Ser Leu Pro Gly Ser Pro Ala Ile Lys Lys Gly			
	725	730	735
Ile Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met Gly			
	740	745	750
Gly Arg Lys Pro Glu Ser Ile Val Val Glu Met Ala Arg Glu Asn Gln			
	755	760	765
Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg Leu			
	770	775	780
Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn Ile			
785	790	795	800

Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp Arg
 805 810 815
 Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Asp
 820 825 830
 Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile Ile
 835 840 845
 Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu Val
 850 855 860
 Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Leu Glu
 865 870 875 880
 Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser Lys
 885 890 895
 Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
 900 905 910
 Gly Leu Ser Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu Val
 915 920 925
 Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu Lys
 930 935 940
 Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val Lys
 945 950 955 960
 Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp Phe
 965 970 975
 Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His Asp
 980 985 990
 Ala Tyr Leu Asn Ala Val Val Ala Ser Ala Leu Leu Lys Lys Tyr Pro
 995 1000 1005
 Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr Asn Ser
 1010 1015 1020
 Phe Arg Glu Arg Lys Ser Ala Thr Glu Lys Val Tyr Phe Tyr Ser Asn
 1025 1030 1035 1040
 Ile Met Asn Ile Phe Lys Lys Ser Ile Ser Leu Ala Asp Gly Arg Val
 1045 1050 1055
 Ile Glu Arg Pro Leu Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val
 1060 1065 1070
 Trp Asn Lys Glu Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser Tyr
 1075 1080 1085
 Pro Gln Val Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly Leu
 1090 1095 1100
 Asp Arg Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser Lys

1105	1110	1115	1120
Pro Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu			
	1125	1130	1135
Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe Thr			
	1140	1145	1150
Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys Ile Thr			
	1155	1160	1165
Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg Ile Asn Tyr			
	1170	1175	1180
Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly Tyr Lys Asp Ile			
1185	1190	1195	1200
Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu Phe Glu Leu Ser Asp			
	1205	1210	1215
Gly Ser Arg Arg Met Leu Ala Ser Ile Leu Ser Thr Asn Asn Lys Arg			
	1220	1225	1230
Gly Glu Ile His Lys Gly Asn Gln Ile Phe Leu Ser Gln Lys Phe Val			
	1235	1240	1245
Lys Leu Leu Tyr His Ala Lys Arg Ile Ser Asn Thr Ile Asn Glu Asn			
	1250	1255	1260
His Arg Lys Tyr Val Glu Asn His Lys Lys Glu Phe Glu Glu Leu Phe			
1265	1270	1275	1280
Tyr Tyr Ile Leu Glu Phe Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn			
	1285	1290	1295
Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile			
	1300	1305	1310
Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys			
	1315	1320	1325
Gly Leu Phe Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe			
	1330	1335	1340
Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu			
1345	1350	1355	1360
Leu Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu			
	1365	1370	1375
Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly			
	1380	1385	

<210> 115

<211> 1333

<212> PRT

<213> 无害李斯特菌

Asp Glu Tyr Ala Glu Leu Phe Val Ala Ala Lys Asn Ala Tyr Ser Ala
 290 295 300
 Val Val Leu Ser Ser Ile Ile Thr Val Ala Glu Thr Glu Thr Asn Ala
 305 310 315 320
 Lys Leu Ser Ala Ser Met Ile Glu Arg Phe Asp Thr His Glu Glu Asp
 325 330 335
 Leu Gly Glu Leu Lys Ala Phe Ile Lys Leu His Leu Pro Lys His Tyr
 340 345 350
 Glu Glu Ile Phe Ser Asn Thr Glu Lys His Gly Tyr Ala Gly Tyr Ile
 355 360 365
 Asp Gly Lys Thr Lys Gln Ala Asp Phe Tyr Lys Tyr Met Lys Met Thr
 370 375 380
 Leu Glu Asn Ile Glu Gly Ala Asp Tyr Phe Ile Ala Lys Ile Glu Lys
 385 390 395 400
 Glu Asn Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ala Ile Pro
 405 410 415
 His Gln Leu His Leu Glu Glu Leu Glu Ala Ile Leu His Gln Gln Ala
 420 425 430
 Lys Tyr Tyr Pro Phe Leu Lys Glu Asn Tyr Asp Lys Ile Lys Ser Leu
 435 440 445
 Val Thr Phe Arg Ile Pro Tyr Phe Val Gly Pro Leu Ala Asn Gly Gln
 450 455 460
 Ser Glu Phe Ala Trp Leu Thr Arg Lys Ala Asp Gly Glu Ile Arg Pro
 465 470 475 480
 Trp Asn Ile Glu Glu Lys Val Asp Phe Gly Lys Ser Ala Val Asp Phe
 485 490 495
 Ile Glu Lys Met Thr Asn Lys Asp Thr Tyr Leu Pro Lys Glu Asn Val
 500 505 510
 Leu Pro Lys His Ser Leu Cys Tyr Gln Lys Tyr Leu Val Tyr Asn Glu
 515 520 525
 Leu Thr Lys Val Arg Tyr Ile Asn Asp Gln Gly Lys Thr Ser Tyr Phe
 530 535 540
 Ser Gly Gln Glu Lys Glu Gln Ile Phe Asn Asp Leu Phe Lys Gln Lys
 545 550 555 560
 Arg Lys Val Lys Lys Lys Asp Leu Glu Leu Phe Leu Arg Asn Met Ser
 565 570 575
 His Val Glu Ser Pro Thr Ile Glu Gly Leu Glu Asp Ser Phe Asn Ser
 580 585 590
 Ser Tyr Ser Thr Tyr His Asp Leu Leu Lys Val Gly Ile Lys Gln Glu

595	600	605
Ile Leu Asp Asn Pro Val Asn Thr Glu Met Leu Glu Asn Ile Val Lys		
610	615	620
Ile Leu Thr Val Phe Glu Asp Lys Arg Met Ile Lys Glu Gln Leu Gln		
625	630	635
Gln Phe Ser Asp Val Leu Asp Gly Val Val Leu Lys Lys Leu Glu Arg		
	645	650
Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Ala Lys Leu Leu Met Gly		
	660	670
Ile Arg Asp Lys Gln Ser His Leu Thr Ile Leu Asp Tyr Leu Met Asn		
675	680	685
Asp Asp Gly Leu Asn Arg Asn Leu Met Gln Leu Ile Asn Asp Ser Asn		
690	695	700
Leu Ser Phe Lys Ser Ile Ile Glu Lys Glu Gln Val Thr Thr Ala Asp		
705	710	715
Lys Asp Ile Gln Ser Ile Val Ala Asp Leu Ala Gly Ser Pro Ala Ile		
	725	730
Lys Lys Gly Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val Ser		
	740	745
Val Met Gly Tyr Pro Pro Gln Thr Ile Val Val Glu Met Ala Arg Glu		
755	760	765
Asn Gln Thr Thr Gly Lys Gly Lys Asn Asn Ser Arg Pro Arg Tyr Lys		
770	775	780
Ser Leu Glu Lys Ala Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys Glu		
785	790	795
His Pro Thr Asp Asn Gln Glu Leu Arg Asn Asn Arg Leu Tyr Leu Tyr		
	805	810
Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Gln Asp Leu Asp Ile		
	820	825
His Asn Leu Ser Asn Tyr Asp Ile Asp His Ile Val Pro Gln Ser Phe		
	835	840
Ile Thr Asp Asn Ser Ile Asp Asn Leu Val Leu Thr Ser Ser Ala Gly		
850	855	860
Asn Arg Glu Lys Gly Asp Asp Val Pro Pro Leu Glu Ile Val Arg Lys		
865	870	875
Arg Lys Val Phe Trp Glu Lys Leu Tyr Gln Gly Asn Leu Met Ser Lys		
	885	890
Arg Lys Phe Asp Tyr Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr Glu		
	900	905
		910

Ala Asp Lys Ala Arg Phe Ile His Arg Gln Leu Val Glu Thr Arg Gln
 915 920 925
 Ile Thr Lys Asn Val Ala Asn Ile Leu His Gln Arg Phe Asn Tyr Glu
 930 935 940
 Lys Asp Asp His Gly Asn Thr Met Lys Gln Val Arg Ile Val Thr Leu
 945 950 955 960
 Lys Ser Ala Leu Val Ser Gln Phe Arg Lys Gln Phe Gln Leu Tyr Lys
 965 970 975
 Val Arg Asp Val Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn
 980 985 990
 Gly Val Val Ala Asn Thr Leu Leu Lys Val Tyr Pro Gln Leu Glu Pro
 995 1000 1005
 Glu Phe Val Tyr Gly Asp Tyr His Gln Phe Asp Trp Phe Lys Ala Asn
 1010 1015 1020
 Lys Ala Thr Ala Lys Lys Gln Phe Tyr Thr Asn Ile Met Leu Phe Phe
 1025 1030 1035 1040
 Ala Gln Lys Asp Arg Ile Ile Asp Glu Asn Gly Glu Ile Leu Trp Asp
 1045 1050 1055
 Lys Lys Tyr Leu Asp Thr Val Lys Lys Val Met Ser Tyr Arg Gln Met
 1060 1065 1070
 Asn Ile Val Lys Lys Thr Glu Ile Gln Lys Gly Glu Phe Ser Lys Ala
 1075 1080 1085
 Thr Ile Lys Pro Lys Gly Asn Ser Ser Lys Leu Ile Pro Arg Lys Thr
 1090 1095 1100
 Asn Trp Asp Pro Met Lys Tyr Gly Gly Leu Asp Ser Pro Asn Met Ala
 1105 1110 1115 1120
 Tyr Ala Val Val Ile Glu Tyr Ala Lys Gly Lys Asn Lys Leu Val Phe
 1125 1130 1135
 Glu Lys Lys Ile Ile Arg Val Thr Ile Met Glu Arg Lys Ala Phe Glu
 1140 1145 1150
 Lys Asp Glu Lys Ala Phe Leu Glu Glu Gln Gly Tyr Arg Gln Pro Lys
 1155 1160 1165
 Val Leu Ala Lys Leu Pro Lys Tyr Thr Leu Tyr Glu Cys Glu Glu Gly
 1170 1175 1180
 Arg Arg Arg Met Leu Ala Ser Ala Asn Glu Ala Gln Lys Gly Asn Gln
 1185 1190 1195 1200
 Gln Val Leu Pro Asn His Leu Val Thr Leu Leu His His Ala Ala Asn
 1205 1210 1215
 Cys Glu Val Ser Asp Gly Lys Ser Leu Asp Tyr Ile Glu Ser Asn Arg

	1220	1225	1230
Glu Met Phe	Ala Glu Leu Leu	Ala His Val Ser	Glu Phe Ala Lys Arg
	1235	1240	1245
Tyr Thr Leu	Ala Glu Ala Asn Leu	Asn Lys Ile Asn	Gln Leu Phe Glu
	1250	1255	1260
Gln Asn Lys	Glu Gly Asp Ile Lys	Ala Ile Ala Gln	Ser Phe Val Asp
1265	1270	1275	1280
Leu Met Ala	Phe Asn Ala Met Gly	Ala Pro Ala Ser	Phe Lys Phe Phe
	1285	1290	1295
Glu Thr Thr	Ile Glu Arg Lys Arg	Tyr Asn Asn Leu	Lys Glu Leu Leu
	1300	1305	1310
Asn Ser Thr	Ile Ile Tyr Gln Ser	Ile Thr Gly Leu	Tyr Glu Ser Arg
	1315	1320	1325
Lys Arg Leu	Asp Asp		
	1330		
<210>	116		
<211>	1082		
<212>	PRT		
<213>	脑膜炎奈瑟球菌		
<220>			
<223>	Cas9		
<400>	116		
Met Ala Ala	Phe Lys Pro Asn Ser	Ile Asn Tyr Ile	Leu Gly Leu Asp
1	5	10	15
Ile Gly Ile	Ala Ser Val Gly Trp	Ala Met Val Glu	Ile Asp Glu Glu
	20	25	30
Glu Asn Pro	Ile Arg Leu Ile Asp	Leu Gly Val Arg	Val Phe Glu Arg
	35	40	45
Ala Glu Val	Pro Lys Thr Gly Asp	Ser Leu Ala Met	Ala Arg Arg Leu
	50	55	60
Ala Arg Ser	Val Arg Arg Leu Thr	Arg Arg Arg Ala	His Arg Leu Leu
65	70	75	80
Arg Thr Arg	Arg Leu Leu Lys Arg	Glu Gly Val Leu	Gln Ala Ala Asn
	85	90	95
Phe Asp Glu	Asn Gly Leu Ile Lys	Ser Leu Pro Asn	Thr Pro Trp Gln
	100	105	110
Leu Arg Ala	Ala Ala Leu Asp Arg	Lys Leu Thr Pro	Leu Glu Trp Ser
	115	120	125
Ala Val Leu	Leu His Leu Ile Lys	His Arg Gly Tyr	Leu Ser Gln Arg

130	135	140
Lys Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys		
145	150	155
Gly Val Ala Gly Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr		
	165	170
Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile		
	180	185
Arg Asn Gln Arg Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu		
	195	200
Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn		
	210	220
Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met		
225	230	235
Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly		
	245	250
His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr		
	260	265
Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile		
	275	280
Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr		
	290	295
Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala		
305	310	315
Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg		
	325	330
Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala		
	340	345
Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys		
	355	360
Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr		
	370	375
Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys		
385	390	395
Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser		
	405	410
Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val		
	420	425
Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile		
	435	440
		445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
 450 455 460
 Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
 465 470 475 480
 Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
 485 490 495
 Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
 500 505 510
 Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
 515 520 525
 Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
 530 535 540
 Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
 545 550 555 560
 Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
 565 570 575
 Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
 580 585 590
 Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
 595 600 605
 Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn
 610 615 620
 Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
 625 630 635 640
 Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
 645 650 655
 Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
 660 665 670
 Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
 675 680 685
 Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
 690 695 700
 Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
 705 710 715 720
 Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
 725 730 735
 Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
 740 745 750
 Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln

755	760	765
Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met		
770	775	780
Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala		
785	790	795
Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser		
805	810	815
Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg		
820	825	830
Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys		
835	840	845
Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu		
850	855	860
Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg		
865	870	875
Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys		
885	890	895
Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys		
900	905	910
Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val		
915	920	925
Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn		
930	935	940
Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr		
945	950	955
Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp		
965	970	975
Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp		
980	985	990
Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu		
995	1000	1005
Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys His		
1010	1015	1020
Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp His Lys		
1025	1030	1035
Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys Thr Ala Leu		
1045	1050	1055
Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys Glu Ile Arg Pro		
1060	1065	1070

Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
 1075 1080
 <210> 117
 <211> 1368
 <212> PRT
 <213> 酿脓链球菌
 <220>
 <223> Cas9
 <400> 117
 Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15
 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30
 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45
 Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60
 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95
 Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110
 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125
 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220
 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545	550	555	560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp			
	565	570	575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly			
	580	585	590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp			
	595	600	605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr			
	610	615	620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala			
625	630	635	640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr			
	645	650	655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp			
	660	665	670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe			
	675	680	685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe			
	690	695	700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu			
705	710	715	720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly			
	725	730	735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly			
	740	745	750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln			
	755	760	765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile			
	770	775	780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro			
785	790	795	800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu			
	805	810	815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg			
	820	825	830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys			
	835	840	845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg			
	850	855	860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
 1010 1015 1020
 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
 1025 1030 1035 1040
 Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
 1045 1050 1055
 Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
 1060 1065 1070
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
 1075 1080 1085
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
 1090 1095 1100
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
 1105 1110 1115 1120
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
 1125 1130 1135
 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
 1140 1145 1150
 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
 1155 1160 1165
 Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1170	1175	1180	
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys			
1185	1190	1195	1200
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser			
	1205	1210	1215
Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr			
	1220	1225	1230
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser			
	1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His			
	1250	1255	1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val			
1265	1270	1275	1280
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys			
	1285	1290	1295
His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu			
	1300	1305	1310
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp			
	1315	1320	1325
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp			
	1330	1335	1340
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile			
1345	1350	1355	1360
Asp Leu Ser Gln Leu Gly Gly Asp			
	1365		

<210> 118

<211> 1205

<212> DNA

<213> 人工序列

<220>

<223> EF1 α 启动子

<400> 118

```

cgtgaggctc cgggtccccgt cagtgggcag agcgcacatc gcccacagtc cccgagaagt 60
tggggggagg ggtcggcaat tgaaccggtg cctagagaag gtggcgcggg gtaaactggg 120
aaagtgatgt cgtgtactgg ctccgccttt ttcccagggg tgggggagaa ccgtatataa 180
gtgcactagt cgccgtgaac gttctttttc gcaacggggtt tgccgccaga acacaggtaa 240
gtgccgtgtg tggttcccgc gggcctggcc tctttacggg ttatggcctt tgcgtgcctt 300
gaattacttc cacctggctg cagtacgtga ttcttgatcc cgagcttcgg gttggaagtg 360
gggtgggagag ttcgtggcct tgcgcttaag gagccccttc gcctcgtgct tgagttgtgg 420

```

cctggcctgg gcgctggggc cgccgcgtgc gaatctggtg gcaccttcgc gcctgtctcg 480
 ctgctttcga taagtctcta gccatttaa atttttgatg acctgctgcg acgctttttt 540
 tctggcaaga tagtcttgta aatgcgggcc aagatcagca cactggtatt tcggtttttg 600
 gggccgcggg cggcgacggg gcccggtgct cccagcgcac atgttcggcg aggcggggcc 660
 tgcgagcgcg gccaccgaga atcggacggg ggtagtctca agctgcccgg cctgctctgg 720
 tgcctggcct cgcgccgccc tgtatcgccc cgccctgggc ggcaaggctg gcccggtcgg 780
 caccagtgc gtgagcggaa agatggccgc ttcccggccc tgctgcaggg agcacaaaat 840
 ggaggacgcg gcgctcggga gagcgggagg gtgagtcacc cacacaaagg aaaagggcct 900
 ttccgtcctc agccgtcgtc tcatgtgact ccacggagta ccgggcgccc tccaggcacc 960
 tcgattagtt ctccagcttt tggagtacgt cgtcttttagg ttggggggag gggttttatg 1020
 cgatggagtt tccccacact gagtgggtgg agactgaagt taggccagct tggcacttga 1080
 tgtaattctc cttggaattt gccctttttg agtttggatc ttggttcatt ctcaagcctc 1140
 agacagtgg tcaaagtttt tttcttccat ttcaggtgct gtgaaaacta ccctaaaag 1200
 ccaaa 1205

<210> 119

<211> 544

<212> DNA

<213> 人工序列

<220>

<223> 具有HTLV1增强子的Ef1 α 启动子

<400> 119

ggatctgcga tcgctccggt gcccgtcagt gggcagagcg cacatcgccc acagtccccg 60
 agaagttggg gggagggggtc ggcaattgaa ccggtgccta gagaagggtg cgcggggtaa 120
 actgggaaaag tgatgtcgtg tactggctcc gcctttttcc cgagggtggg ggagaaccgt 180
 atataagtgc agtagtcgcc gtgaacgtt ttttcgcaa cgggtttgcc gccagaacac 240
 agctgaagct tcgaggggct cgcattcttc cttcacgcbc ccgccgccct acctgaggcc 300
 gccatccacg ccggttgagt cgcgttctgc cgctcccgc ctgtggtgcc tcctgaactg 360
 cgtccgccgt ctaggtaagt ttaaagctca ggtcgagacc gggcctttgt ccggcgctcc 420
 cttggagcct acctagactc agccggtctt ccacgctttg cctgaccctg cttgctcaac 480
 tctacgtctt tgtttcgttt tctgttctgc gccgttacag atccaagctg tgaccggcgc 540
 ctac 544

<210> 120

<211> 66

<212> DNA

<213> 人工序列

<220>

<223> P2A核苷酸序列

<400> 120

ggatctggag cgacgaattt tagtctactg aaacaagcgg gagacgtgga ggaaaaccct 60

ggacct 66
<210> 121
<211> 4107
<212> DNA
<213> 酿脓链球菌
<220>
<223> Cas9密码子优化的核酸序列
<400> 121
atggataaaa agtacagcat cgggctggac atcggtaaca actcagtggg gtgggccgtg 60
attacggacg agtacaaggt accctccaaa aaattttaaag tgctgggtaa cacggacaga 120
cactctataa agaaaaatct tattggagcc ttgctgttcg actcaggcga gacagccgaa 180
gccacaaggt tgaagcggac cgccaggagg cggatataca ggagaaagaa ccgcatatgc 240
tacctgcaag aaatcttcag taacgagatg gcaaaggttg acgatagctt tttccatcgc 300
ctggaagaat cctttcttgt tgaggaagac aagaagcacg aacggcacc cctctttggc 360
aatattgtcg acgaagtggc atatcacgaa aagtaccga ctatctacca ctcaggaag 420
aagctgggtg actctaccga taaggcggac ctgagactta tttatttggc actcgccac 480
atgattaaat ttagaggaca tttcttgatc gagggcgacc tgaaccgga caacagtgc 540
gtcgataagc tgttcatcca acttgtgcag acctacaatc aactgttcga agaaaacct 600
ataaatgctt caggagtcga cgctaaagca atcctgtccg cgcgcctctc aaaatctaga 660
agacttgaga atctgattgc tcagttgcc ggggaaaaga aaaatggatt gtttggcaac 720
ctgatcgccc tcagtctcgg actgaccca aatttcaaaa gtaacttga cctggccgaa 780
gacgctaagc tccagctgtc caaggacaca tacgatgacg acctcgaaa tctgctggcc 840
cagattgggg atcagtacgc cgatctctt ttggcagcaa agaacctgtc cgacgccatc 900
ctgttgagcg atatcttgag agtgaacacc gaaattacta aagcaccct tagcgcattct 960
atgatcaagc ggtacgacga gcatcatcag gatctgacce tgctgaaggc tcttgtgagg 1020
caacagctcc ccgaaaaata caaggaaatc ttctttgacc agagcaaaaa cggctacgct 1080
ggctatatag atggtggggc cagtcaggag gaattctata aattcatcaa gccattctc 1140
gagaaaatgg acggcacaga ggagtgtctg gtcaaactta acagggagga cctgctgcgg 1200
aagcagcggc cctttgacaa cgggtctatc cccaccaga ttcattctggg cgaactgcac 1260
gcaatcctga ggaggcagga ggatttttat cttttctta aagataaccg cgagaaaata 1320
gaaaagattc ttacattcag gatcccgtac tacgtgggac ctctcgccc gggcaattca 1380
cggtttgcct ggatgacaag gaagtcagag gagactatta caccttgga cttcgaagaa 1440
gtggtggaca aggtgcatc tgcccagtct ttcacgagc ggatgacaaa ttttgacaag 1500
aacctcccta atgagaaggt gctgccc aaa cattctctgc tctacgagta ctttaccgtc 1560
tacaatgaac tgactaaagt caagtacgtc accgaggaa tgaggaagcc ggcattcctt 1620
agtggagaac agaagaagc gattgtagac ctgttgttca agaccaacag gaaggtgact 1680
gtgaagcaac ttaaagaaga ctactttaag aagatcgaat gttttgacag tgtggaatt 1740
tcaggggttg aagaccgctt caatgcgtca ttggggactt accatgatct tctcaagatc 1800
ataaaggaca aagacttctt ggacaacgaa gaaaatgagg atattctcga agacatcgtc 1860

ctcaccctga ccctgttcga agacagggaa atgatagaag agcgcttgaa aacctatgcc 1920
 cacctcttcg acgataaagt tatgaagcag ctgaagcgca ggagatacac aggatgggga 1980
 agattgtcaa ggaagctgat caatggaatt agggataaac agagtggcaa gaccatactg 2040
 gatttcctca aatctgatgg cttcgccaat aggaacttca tgcaactgat tcacgatgac 2100
 tctcttacct tcaaggagga cattcaaaag gctcaggtga gcgggcagg agactccctt 2160
 catgaacaca tcgcgaatth ggcaggttcc cccgctatta aaaaggcat cttcaaact 2220
 gtcaaggtgg tggatgaatt ggtcaaggta atgggcagac ataagccaga aaatattgtg 2280
 atcgagatgg cccgcgaaaa ccagaccaca cagaagggcc agaaaaatag tagagagcgg 2340
 atgaagagga tcgaggagg catcaaagag ctgggatctc agatttctca agaacacccc 2400
 gtagaaaaca cacagctgca gaacgaaaaa ttgtacttgt actatctgca gaacggcaga 2460
 gacatgtacg tcgaccaaga acttgatatt aatagactgt ccgactatga cgtagacat 2520
 atcgtgcccc agtccttctt gaaggacgac tccattgata acaaagtctt gacaagaagc 2580
 gacaagaaca ggggtaaaag tgataatgtg cctagcgagg aggtggtgaa aaaaatgaag 2640
 aactactggc gacagctgct taatgcaaag ctcatcac aacggaagtt cgataatctg 2700
 acgaaagcag agagaggtgg cttgtctgag ttggacaagg cagggtttat taagcggcag 2760
 ctggtggaaa ctaggcagat cacaaagcac gtggcgcaga ttttgacag ccgatgaac 2820
 acaaaatagc acgaaaatga taaactgata cgagaggtca aagttatcac gctgaaaagc 2880
 aagctggtgt ccgattttcg gaaagacttc cagttctaca aagttcgcga gattaataac 2940
 taccatcatg ctcacgatgc gtacctgaac gctgttctc ggaccgcctt gataaagaag 3000
 tacccaaagc tggaatccga gttcgtatac ggggattaca aagtgtacga tgtgaggaaa 3060
 atgatagcca agtccgagca ggagattgga aaggccacag ctaagtactt cttttattct 3120
 aacatcatga attttttaa gacggaaatt accctggcca acggagagat cagaaagcgg 3180
 ccccttatag agacaaatgg tgaacaggt gaaatcgtct gggataaggg cagggatttc 3240
 gctactgtga ggaaggtgct gagtatgcca caggtaaata tcgtgaaaa aaccgaagta 3300
 cagaccggag gattttccaa ggaaagcatt ttgcctaaaa gaaactcaga caagctcatc 3360
 gcccgaaga aagattggga ccctaagaaa tacgggggat ttgactcacc caccgtagcc 3420
 tattctgtgc tgggtgtagc taagtgga aaaggaaagt ctaagaagct gaagtcctg 3480
 aaggaactct tgggaatcac tatcatgga agatcatcct ttgaaaagaa ccctatcgat 3540
 ttcttgagg ctaagggtta caaggagtc aagaaagacc tcatcattaa actgccaaaa 3600
 tactctctct tcgagctgga aatggcagg aagagaatgt tggccagcgc cggagagctg 3660
 caaaagggaa acgagcttgc tctgcctcc aatatgtta attttctcta tctcgcttcc 3720
 cactatgaaa agctgaaaag gtctcccgaa gataacgagc agaagcagct gttcgtcgaa 3780
 cagcacaagc actatctgga tgaaataatc gaacaataa gcgagttcag caaaagggtt 3840
 atcctggcgg atgctaattt ggacaaagta ctgtctgctt ataacaagca ccgggataag 3900
 cctattaggg aacaagccga gaatataatt cacctcttta cactcacgaa tctcggagcc 3960
 cccgccct tcaaataact tgatacgaact atcgaccgga aacggtatac cagtaccaa 4020
 gaggtcctcg atgccacct catccaccag tcaattactg gcctgtacga aacacgatc 4080
 gacctctctc aactgggcgg cgactag 4107

<210> 122

260	265	270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp		
275	280	285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp		
290	295	300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser		
305	310	315
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys		
325	330	335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe		
340	345	350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser		
355	360	365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp		
370	375	380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg		
385	390	395
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu		
405	410	415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe		
420	425	430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile		
435	440	445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp		
450	455	460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu		
465	470	475
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr		
485	490	495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser		
500	505	510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys		
515	520	525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln		
530	535	540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr		
545	550	555
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp		
565	570	575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685
 Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700
 Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 705 710 715 720
 His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735
 Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750
 Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 755 760 765
 Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 770 775 780
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800
 Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
 805 810 815
 Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
 820 825 830
 Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
 835 840 845
 Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 850 855 860
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

	885	890	895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp			
	900	905	910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr			
	915	920	925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp			
	930	935	940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser			
945	950	955	960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg			
	965	970	975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val			
	980	985	990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe			
	995	1000	1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys			
1010	1015	1020	
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser			
1025	1030	1035	1040
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu			
	1045	1050	1055
Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile			
	1060	1065	1070
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser			
	1075	1080	1085
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly			
1090	1095	1100	
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile			
1105	1110	1115	1120
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser			
	1125	1130	1135
Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly			
	1140	1145	1150
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile			
	1155	1160	1165
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala			
1170	1175	1180	
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys			
1185	1190	1195	1200

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
 1205 1210 1215
 Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
 1220 1225 1230
 Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
 1235 1240 1245
 Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
 1250 1255 1260
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
 1265 1270 1275 1280
 Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
 1285 1290 1295
 His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
 1300 1305 1310
 Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
 1315 1320 1325
 Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
 1330 1335 1340
 Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
 1345 1350 1355 1360
 Asp Leu Ser Gln Leu Gly Gly Asp
 1365

<210> 123

<211> 3249

<212> DNA

<213> 脑膜炎奈瑟球菌

<220>

<223> Cas9密码子优化的核酸序列

<400> 123

atggccgcct tcaagcccaa cccatcaac tacatcctgg gcttgacat cggcatcgcc 60
 agcgtgggct gggccatggt ggagatcgac gaggacgaga acccatctg cctgatcgac 120
 ctgggtgtgc gcgtgttcga gcgcgctgag gtgccaaga ctggtgacag tctggctatg 180
 gctcgccggc ttgctcgtc tgttcggcgc ctactcgcc ggcgcgctca ccgccttctg 240
 cgcgctcgcc gcctgctgaa gcgcgagggt gtgctgcagg ctgccgactt cgacgagaac 300
 ggctgatca agagcctgcc caaactcct tggcagctgc gcgctgccgc tctggaccgc 360
 aagctgactc ctctggagtg gagcgccgtg ctgctgcacc tgatcaagca ccgcggctac 420
 ctgagccagc gcaagaacga gggcgagacc gccgacaagg agctgggtgc tctgctgaag 480
 ggcgtggccg acaacgcca cgccctgcag actggtgact tccgactcc tgctgagctg 540
 gccctgaaca agttcgagaa ggagagcggc cacatccga accagcgcgg cgactacagc 600

cacaccttca gccgcaagga cctgcaggcc gagctgatcc tgctgttcga gaagcagaag 660
gagttcggca acccccacgt gagcggcggc ctgaaggagg gcatcgagac cctgctgatg 720
acccagcgcc ccgccctgag cggcgacgcc gtgcagaaga tgctgggcca ctgcaccttc 780
gagccagccg agcccaaggc cgccaagaac acctacaccg ccgagcgcctt catctggctg 840
accaagctga acaacctgcg catcctggag cagggcagcg agcgccccct gaccgacacc 900
gagcgcgcca ccctgatgga cgagccctac cgcaagagca agctgacctt cgcccaggcc 960
cgcaagctgc tgggtctgga ggacaccgcc ttcttcaagg gcctgcgcta cggcaaggac 1020
aacgccgagg ccagcacctt gatggagatg aaggcctacc acgcatcag ccgcgccctg 1080
gagaaggagg gcctgaagga caagaagagt cctctgaacc tgagccccga gctgcaggac 1140
gagatcggca ccgccttcag cctgttcaag accgacgagg acatcaccgg ccgcctgaag 1200
gaccgcatcc agcccagat cctggaggcc ctgctgaagc acatcagctt cgacaagttc 1260
gtgcagatca gcctgaaggc cctgcgccgc atcgtgcccc tgatggagca gggcaagcgc 1320
tacgacgagg cctgcgccga gatctacggc gaccactacg gcaagaagaa caccgaggag 1380
aagatctacc tgccctctat ccccgccgac gagatccgca acccctggtt gctgcgccgc 1440
ctgagccagg cccgcaaggt gatcaacggc gtggtgcgcc gctacggcag ccccgccccg 1500
atccacatcg agaccgcccg cgaggtgggc aagagcttca aggaccgcaa ggagatcgag 1560
aagcgcagg aggagaaccg caaggaccgc gagaaggccg ccgccaagtt ccgaggtac 1620
ttcccaact tcgtgggcca gcccaagagc aaggacatcc tgaagctgcg cctgtacgag 1680
cagcagcacg gcaagtgcct gtacagcggc aaggagatca acctgggccc cctgaacgag 1740
aagggctacg tggagatcga ccacgccctg cccttcagcc gcacctggga cgacagcttc 1800
aacaacaagg tgctggtgct gggcagcgag aaccagaaca agggcaacca gaccccctac 1860
gagtacttca acggcaagga caacagccgc gagtggcagg agttcaaggc ccgctggag 1920
accagccgct tccccgcag caagaagcag cgcacctctg tgacagaagt cgacgaggac 1980
ggcttcaagg agcgcacact gaacgacacc cgctacgtga accgcttctt gtgccagttc 2040
gtggccgacc gcatgcgcct gaccggcaag ggcaagaagc gcgtgttcgc cagcaacggc 2100
cagatcacca acctgctgcg cggcttctgg ggctgcgca aggtgcgcgc cgagaacgac 2160
cgccaccacg ccctggacgc cgtggtggtg gcctgcagca ccgtggccat gcagcagaag 2220
atcacccgct tcgtgcgcta caaggagatg aacgccttcg acggtaaaac catcgacaag 2280
gagaccggcg aggtgctgca ccagaagacc cacttcccc agccctggga gttcttcgcc 2340
caggaggtga tgatccgcgt gttcggcaag cccgacggca agcccagatt cgaggaggcc 2400
gacacccccg agaagctgcg cacctgctg gccgagaagc tgagcagccg ccctgaggcc 2460
gtgcacgagt acgtgactcc tctgttcgtg agccgcgccc ccaaccgcaa gatgagcggc 2520
cagggtcaca tggagaccgt gaagagcgc aagcgcctgg acgagggcgt gagcgtgctg 2580
cgcgtgcccc tgaccagct gaagctgaag gacctggaga agatggtgaa ccgagcgcgc 2640
gagcccaagc tgtacgagc cctgaaggcc cgctggagg cccacaagga cgacccccgc 2700
aaggccttcg ccgagccctt ctacaagtac gacaaggccg gcaaccgca ccagcaggtg 2760
aaggccgtgc gcgtggagca ggtgcagaag accggcgtgt ggggtgcgca ccacaacggc 2820
atcgccgaca acgccaccat ggtgcgcgtg gacgtgttcg agaagggcga caagtactac 2880
ctggtgcccc tctacagctg gcaggtggcc aaggcctcc tgcccagccg cgccgtggtg 2940

cagggcaagg acgaggagga ctggcagctg atcgacgaca gcttcaactt caagttcagc 3000
 ctgcacccca acgacctggt ggaggtgatc accaagaagg cccgcatggt cggctacttc 3060
 gccagctgcc accgcggcac cggcaacatc aacatccgca tccacgacct ggaccacaag 3120
 atcggcaaga acggcatcct ggagggcatc ggcgtaaga ccgccctgag cttccagaag 3180
 taccagatcg acgagctggg caaggagatc cgcccctgcc gcctgaagaa gcgccctcct 3240
 gtgcgctaa 3249

<210> 124

<211> 1082

<212> PRT

<213> 脑膜炎奈瑟球菌

<220>

<223> Cas9

<400> 124

Met	Ala	Ala	Phe	Lys	Pro	Asn	Pro	Ile	Asn	Tyr	Ile	Leu	Gly	Leu	Asp	1	5	10	15
Ile	Gly	Ile	Ala	Ser	Val	Gly	Trp	Ala	Met	Val	Glu	Ile	Asp	Glu	Asp	20	25	30	
Glu	Asn	Pro	Ile	Cys	Leu	Ile	Asp	Leu	Gly	Val	Arg	Val	Phe	Glu	Arg	35	40	45	
Ala	Glu	Val	Pro	Lys	Thr	Gly	Asp	Ser	Leu	Ala	Met	Ala	Arg	Arg	Leu	50	55	60	
Ala	Arg	Ser	Val	Arg	Arg	Leu	Thr	Arg	Arg	Arg	Ala	His	Arg	Leu	Leu	65	70	75	80
Arg	Ala	Arg	Arg	Leu	Leu	Lys	Arg	Glu	Gly	Val	Leu	Gln	Ala	Ala	Asp	85	90	95	
Phe	Asp	Glu	Asn	Gly	Leu	Ile	Lys	Ser	Leu	Pro	Asn	Thr	Pro	Trp	Gln	100	105	110	
Leu	Arg	Ala	Ala	Ala	Leu	Asp	Arg	Lys	Leu	Thr	Pro	Leu	Glu	Trp	Ser	115	120	125	
Ala	Val	Leu	Leu	His	Leu	Ile	Lys	His	Arg	Gly	Tyr	Leu	Ser	Gln	Arg	130	135	140	
Lys	Asn	Glu	Gly	Glu	Thr	Ala	Asp	Lys	Glu	Leu	Gly	Ala	Leu	Leu	Lys	145	150	155	160
Gly	Val	Ala	Asp	Asn	Ala	His	Ala	Leu	Gln	Thr	Gly	Asp	Phe	Arg	Thr	165	170	175	
Pro	Ala	Glu	Leu	Ala	Leu	Asn	Lys	Phe	Glu	Lys	Glu	Ser	Gly	His	Ile	180	185	190	
Arg	Asn	Gln	Arg	Gly	Asp	Tyr	Ser	His	Thr	Phe	Ser	Arg	Lys	Asp	Leu	195	200	205	

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
 210 215 220
 Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
 225 230 235 240
 Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
 245 250 255
 His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
 260 265 270
 Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
 275 280 285
 Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
 290 295 300
 Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
 305 310 315 320
 Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
 325 330 335
 Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
 340 345 350
 Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
 355 360 365
 Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
 370 375 380
 Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
 385 390 395 400
 Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
 405 410 415
 Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
 420 425 430
 Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
 435 440 445
 Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
 450 455 460
 Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
 465 470 475 480
 Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
 485 490 495
 Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
 500 505 510
 Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys

515	520	525
Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe		
530	535	540
Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu		
545	550	555
Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly		
565	570	575
Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe		
580	585	590
Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly		
595	600	605
Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn		
610	615	620
Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu		
625	630	635
Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys		
645	650	655
Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr		
660	665	670
Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr		
675	680	685
Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn		
690	695	700
Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp		
705	710	715
Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala		
725	730	735
Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala		
740	745	750
Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln		
755	760	765
Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met		
770	775	780
Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala		
785	790	795
Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser		
805	810	815
Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg		
820	825	830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
835 840 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
850 855 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865 870 875 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
885 890 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
900 905 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
915 920 925

Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
930 935 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
945 950 955 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
965 970 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
980 985 990

Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
995 1000 1005

Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys His
1010 1015 1020

Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp His Lys
1025 1030 1035 1040

Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys Thr Ala Leu
1045 1050 1055

Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys Glu Ile Arg Pro
1060 1065 1070

Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1075 1080

<210> 125
<211> 3159
<212> DNA
<213> 金黄色葡萄球菌
<220>
<223> Cas9密码子优化的核酸序列
<400> 125

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt 60
attgactatg aaacaagga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac 120
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga 180
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgacat 240
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg 300
tcagaggaag agttttccgc agctctgctg cacctggcta agcgcggagg agtgcataac 360
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc 420
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa 480
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc 540
aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact 600
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc 660
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatfff 720
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat 780
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag 840
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct 900
aaggagatcc tggatcaacga agaggacatc aagggtacc gggtgacaag cactggaaaa 960
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acgaaagaa 1020
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc 1080
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgacca ggaagagatc 1140
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc 1200
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg 1260
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg 1320
gtggacgatt tcattctgtc acccgtggtc aagcggagct tcattccagag catcaaagt 1380
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg 1440
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag 1500
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg 1560
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc 1620
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc 1680
agaagcgtgt ccttcgacaa ttctttaac aacaaggtgc tggatcaagca ggaagagaac 1740
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct 1800
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag 1860
accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat 1920
tttattaacc ggaatctggt ggacacaaga tacgtactc gcggcctgat gaatctgctg 1980
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc 2040
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac 2100
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag 2160
ctggacaaaag ccaagaaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct 2220
atgcccgaaa tcgagacaga acaggagtac aaggagatft tcactactcc tcaccagatc 2280
aagcatatca aggatttcaa ggactacaag tactctcacc ggggtgataa aaagcccaac 2340

agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg 2400
 attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac 2460
 aacaaaagtc ccgagaagct gctgatgtac caccatgac ctcagacata tcagaaactg 2520
 aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag 2580
 actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagac 2640
 aagtactatg ggaacaagct gaatgcccc ctggacatca cagacgatta ccctaacagt 2700
 cgcaacaagg tggcaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac 2760
 ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat 2820
 gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca 2880
 gagttcatcg cctccttita caacaacgac ctgattaaga tcaatggcga actgtatagg 2940
 gtcacgggg tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact 3000
 taccgagagt atctggaaaa catgaatgat aagcgecccc ctgaattat caaacaatt 3060
 gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag 3120
 gtgaagagca aaaagcaccc tcagattatc aaaaagggc 3159

<210> 126

<211> 1053

<212> PRT

<213> 金黄色葡萄球菌

<220>

<223> Cas9

<400> 126

Met	Lys	Arg	Asn	Tyr	Ile	Leu	Gly	Leu	Asp	Ile	Gly	Ile	Thr	Ser	Val
1			5					10						15	
Gly	Tyr	Gly	Ile	Ile	Asp	Tyr	Glu	Thr	Arg	Asp	Val	Ile	Asp	Ala	Gly
			20					25						30	
Val	Arg	Leu	Phe	Lys	Glu	Ala	Asn	Val	Glu	Asn	Asn	Glu	Gly	Arg	Arg
			35				40						45		
Ser	Lys	Arg	Gly	Ala	Arg	Arg	Leu	Lys	Arg	Arg	Arg	Arg	His	Arg	Ile
			50				55					60			
Gln	Arg	Val	Lys	Lys	Leu	Leu	Phe	Asp	Tyr	Asn	Leu	Leu	Thr	Asp	His
65					70					75					80
Ser	Glu	Leu	Ser	Gly	Ile	Asn	Pro	Tyr	Glu	Ala	Arg	Val	Lys	Gly	Leu
					85					90					95
Ser	Gln	Lys	Leu	Ser	Glu	Glu	Glu	Phe	Ser	Ala	Ala	Leu	Leu	His	Leu
			100							105				110	
Ala	Lys	Arg	Arg	Gly	Val	His	Asn	Val	Asn	Glu	Val	Glu	Glu	Asp	Thr
			115							120				125	
Gly	Asn	Glu	Leu	Ser	Thr	Lys	Glu	Gln	Ile	Ser	Arg	Asn	Ser	Lys	Ala
			130											140	

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
 145 150 155 160
 Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
 165 170 175
 Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
 180 185 190
 Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
 195 200 205
 Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
 210 215 220
 Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
 225 230 235 240
 Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
 245 250 255
 Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
 260 265 270
 Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
 275 280 285
 Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
 290 295 300
 Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
 305 310 315 320
 Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
 325 330 335
 Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
 340 345 350
 Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
 355 360 365
 Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
 370 375 380
 Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
 385 390 395 400
 Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
 405 410 415
 Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
 420 425 430
 Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
 435 440 445
 Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile

450	455	460
Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg		
465	470	475
Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys		480
	485	490
Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr		495
	500	505
Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp		510
	515	520
Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu		525
	530	535
Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro		540
545	550	555
Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys		560
	565	570
Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu		575
	580	585
Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile		590
	595	600
Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu		605
	610	615
Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp		620
625	630	635
Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu		640
	645	650
Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys		655
	660	665
Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp		670
	675	680
Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp		685
	690	695
Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys		700
705	710	715
Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys		720
	725	730
Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu		735
	740	745
Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp		750
	755	760
		765

Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
 770 775 780
 Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
 785 790 795 800
 Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
 805 810 815
 Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
 820 825 830
 Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
 835 840 845
 Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
 850 855 860
 Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile
 865 870 875 880
 Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
 885 890 895
 Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
 900 905 910
 Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
 915 920 925
 Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
 930 935 940
 Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
 945 950 955 960
 Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
 965 970 975
 Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
 980 985 990
 Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
 995 1000 1005
 Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr
 1010 1015 1020
 Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu
 1025 1030 1035 1040
 Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
 1045 1050

<210> 127

<211> 326

<212> PRT

<213> 智人 (Homo sapiens)

<220>

<223> 人IgG2 Fc

<300>

<308> Uniprot P01859

<309> 2008-12-16

<400> 127

Ala	Ser	Thr	Lys	Gly	Pro	Ser	Val	Phe	Pro	Leu	Ala	Pro	Cys	Ser	Arg
1				5					10					15	
Ser	Thr	Ser	Glu	Ser	Thr	Ala	Ala	Leu	Gly	Cys	Leu	Val	Lys	Asp	Tyr
			20					25					30		
Phe	Pro	Glu	Pro	Val	Thr	Val	Ser	Trp	Asn	Ser	Gly	Ala	Leu	Thr	Ser
			35				40					45			
Gly	Val	His	Thr	Phe	Pro	Ala	Val	Leu	Gln	Ser	Ser	Gly	Leu	Tyr	Ser
			50				55					60			
Leu	Ser	Ser	Val	Val	Thr	Val	Pro	Ser	Ser	Asn	Phe	Gly	Thr	Gln	Thr
65					70					75					80
Tyr	Thr	Cys	Asn	Val	Asp	His	Lys	Pro	Ser	Asn	Thr	Lys	Val	Asp	Lys
				85						90					95
Thr	Val	Glu	Arg	Lys	Cys	Cys	Val	Glu	Cys	Pro	Pro	Cys	Pro	Ala	Pro
				100						105				110	
Pro	Val	Ala	Gly	Pro	Ser	Val	Phe	Leu	Phe	Pro	Pro	Lys	Pro	Lys	Asp
				115						120				125	
Thr	Leu	Met	Ile	Ser	Arg	Thr	Pro	Glu	Val	Thr	Cys	Val	Val	Val	Asp
130															
135															
140															
Val	Ser	His	Glu	Asp	Pro	Glu	Val	Gln	Phe	Asn	Trp	Tyr	Val	Asp	Gly
145															
150															
155															
160															
Val	Glu	Val	His	Asn	Ala	Lys	Thr	Lys	Pro	Arg	Glu	Glu	Gln	Phe	Asn
165															
170															
175															
Ser	Thr	Phe	Arg	Val	Val	Ser	Val	Leu	Thr	Val	Val	His	Gln	Asp	Trp
180															
185															
190															
Leu	Asn	Gly	Lys	Glu	Tyr	Lys	Cys	Lys	Val	Ser	Asn	Lys	Gly	Leu	Pro
195															
200															
205															
Ala	Pro	Ile	Glu	Lys	Thr	Ile	Ser	Lys	Thr	Lys	Gly	Gln	Pro	Arg	Glu
210															
215															
220															
Pro	Gln	Val	Tyr	Thr	Leu	Pro	Pro	Ser	Arg	Glu	Glu	Met	Thr	Lys	Asn
225															
230															
235															
240															
Gln	Val	Ser	Leu	Thr	Cys	Leu	Val	Lys	Gly	Phe	Tyr	Pro	Ser	Asp	Ile
245															
250															
255															

Ser Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr
 260 265 270
 Thr Pro Pro Met Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys
 275 280 285
 Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
 290 295 300
 Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu
 305 310 315 320
 Ser Leu Ser Pro Gly Lys
 325

<210> 128

<211> 327

<212> PRT

<213> 智人 (Homo sapiens)

<220>

<223> 人IgG4 Fc

<300>

<308> Uniprot P01861

<309> 1986-07-21

<400> 128

Ala Ser Thr Lys Gly Pro Ser Val Phe Pro Leu Ala Pro Cys Ser Arg
 1 5 10 15
 Ser Thr Ser Glu Ser Thr Ala Ala Leu Gly Cys Leu Val Lys Asp Tyr
 20 25 30
 Phe Pro Glu Pro Val Thr Val Ser Trp Asn Ser Gly Ala Leu Thr Ser
 35 40 45
 Gly Val His Thr Phe Pro Ala Val Leu Gln Ser Ser Gly Leu Tyr Ser
 50 55 60
 Leu Ser Ser Val Val Thr Val Pro Ser Ser Ser Leu Gly Thr Lys Thr
 65 70 75 80
 Tyr Thr Cys Asn Val Asp His Lys Pro Ser Asn Thr Lys Val Asp Lys
 85 90 95
 Arg Val Glu Ser Lys Tyr Gly Pro Pro Cys Pro Ser Cys Pro Ala Pro
 100 105 110
 Glu Phe Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys
 115 120 125
 Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val
 130 135 140
 Asp Val Ser Gln Glu Asp Pro Glu Val Gln Phe Asn Trp Tyr Val Asp

65	70	75	80
Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn			
	85	90	95
Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Gly Leu Pro Ser Ser			
	100	105	110
Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln			
	115	120	125
Val Tyr Thr Leu Pro Pro Ser Gln Glu Glu Met Thr Lys Asn Gln Val			
	130	135	140
Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val			
145	150	155	160
Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro			
	165	170	175
Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Arg Leu Thr			
	180	185	190
Val Asp Lys Ser Arg Trp Gln Glu Gly Asn Val Phe Ser Cys Ser Val			
	195	200	205
Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu			
	210	215	220

Ser Leu Gly Lys

225

<210> 130

<211> 133

<212> DNA

<213> 人工序列

<220>

<223> SV40多聚A信号

<400> 130

tgctttatatt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat 60
 aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg 120
 gaggtttttt aaa 133

<210> 131

<211> 347

<212> DNA

<213> 人工序列

<220>

<223> MND启动子

<400> 131

gaacagagaa acaggagaat atgggcaaaa caggatatct gtgtaagca gttcctgccc 60

cggctcaggg ccaagaacag ttggaacagc agaatatggg ccaaacagga tatctgtggt 120
aagcagttcc tgccccggct caggccaag aacagatggt cccagatgc ggtcccgccc 180
tcagcagttt ctagagaacc atcagatggt tccagggtgc ccaaggacc tgaatgacc 240
ctgtgcctta ttgaaactaa ccaatcagtt cgcttctgc ttctgttcgc gcgcttctgc 300
tccccgagct ctatataagc agagctcggt tagtgaaccg tcagatc 347

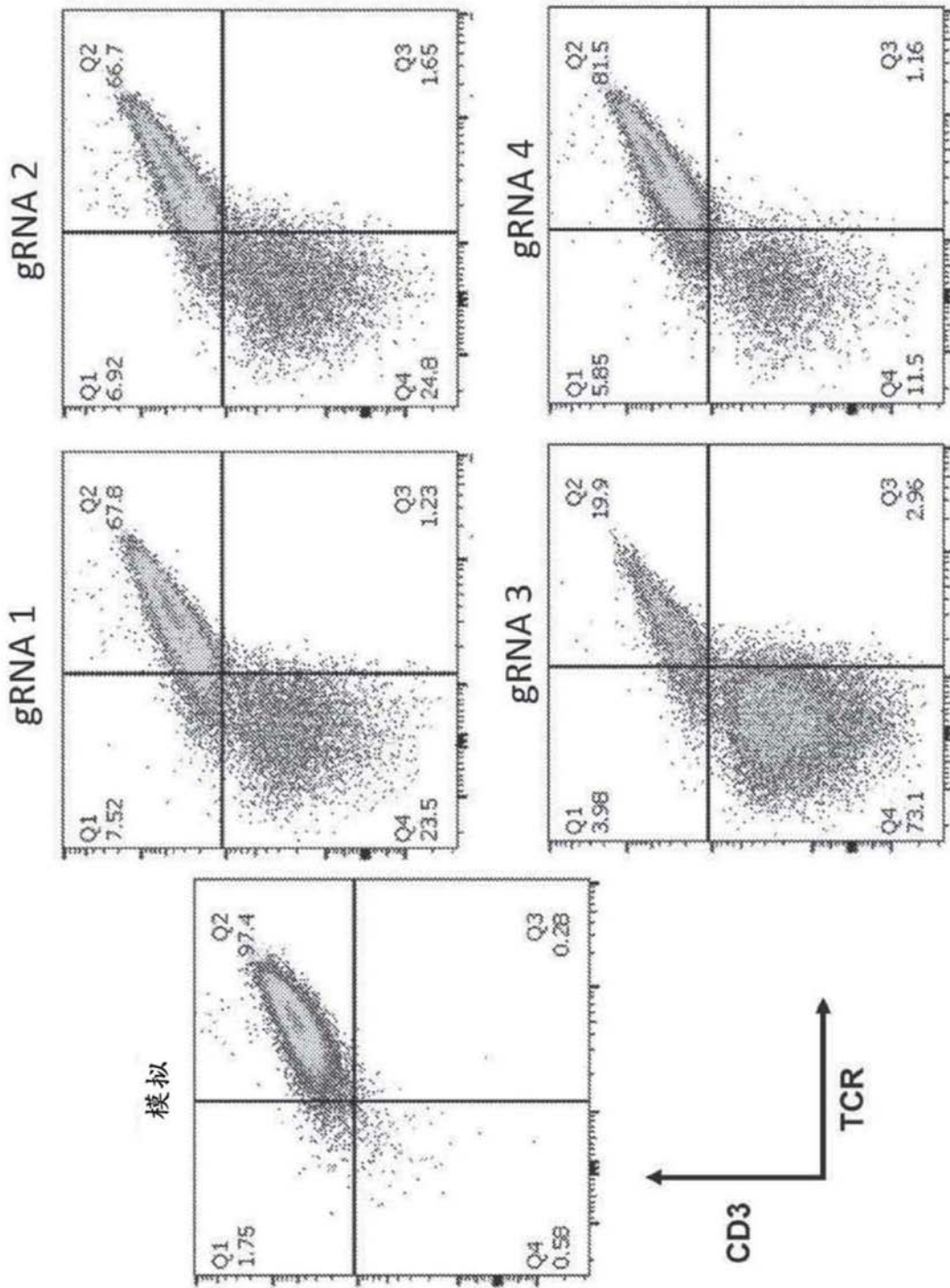


FIG.1

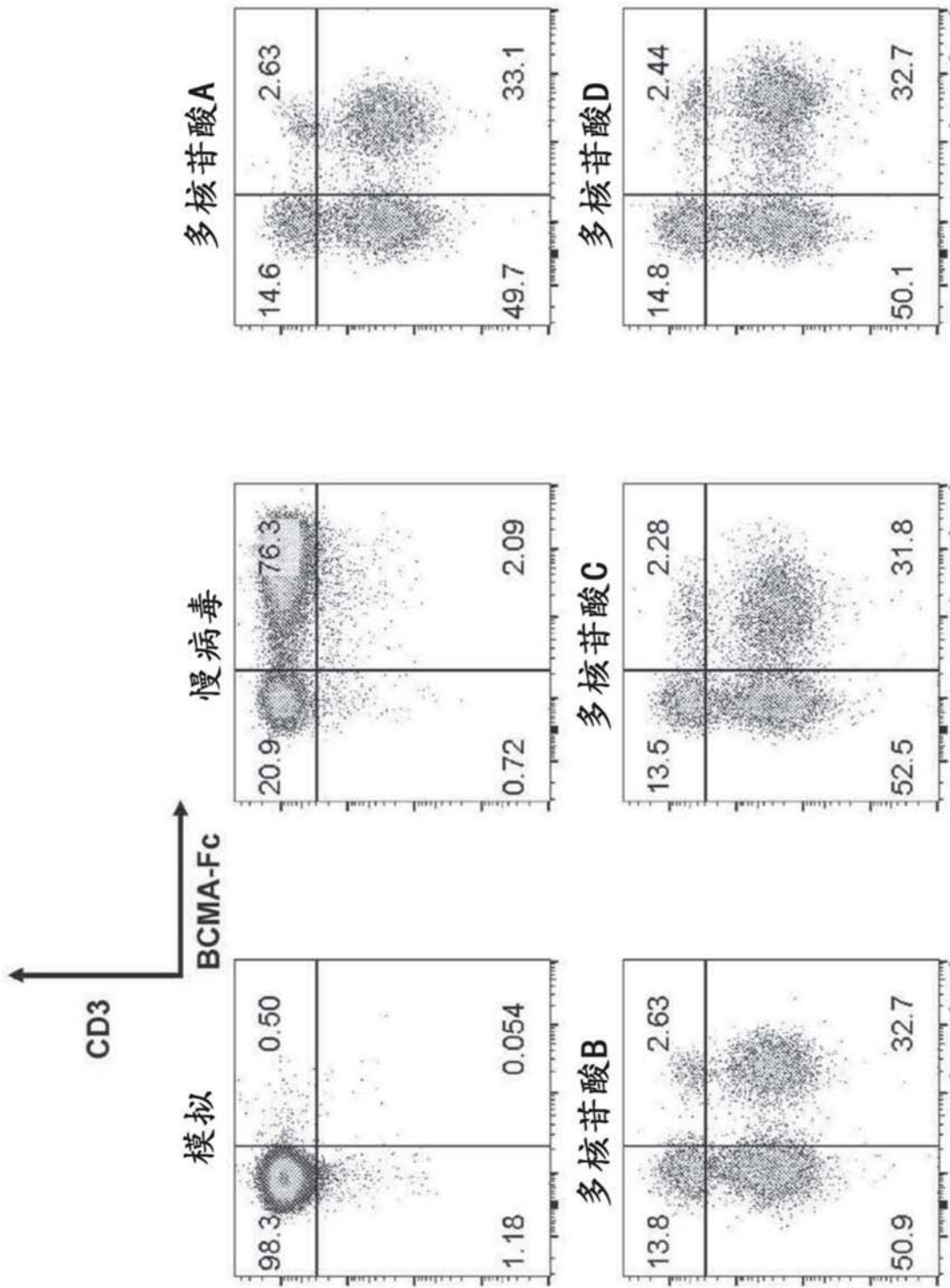


FIG. 2A

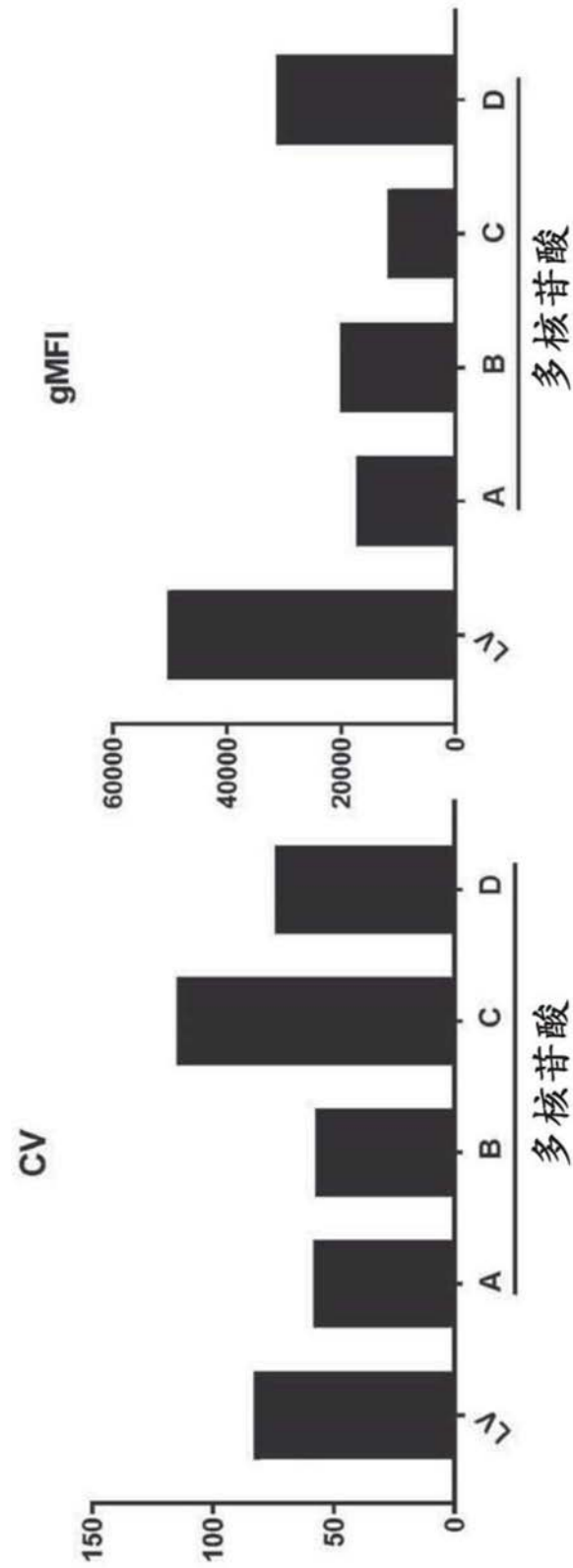


FIG.2B

FIG. 3B

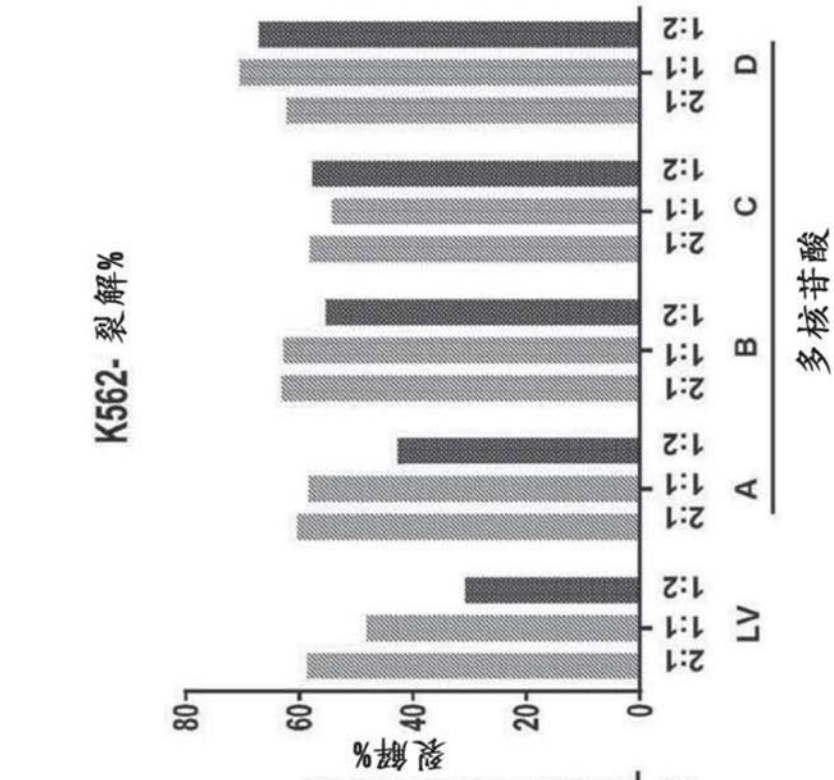
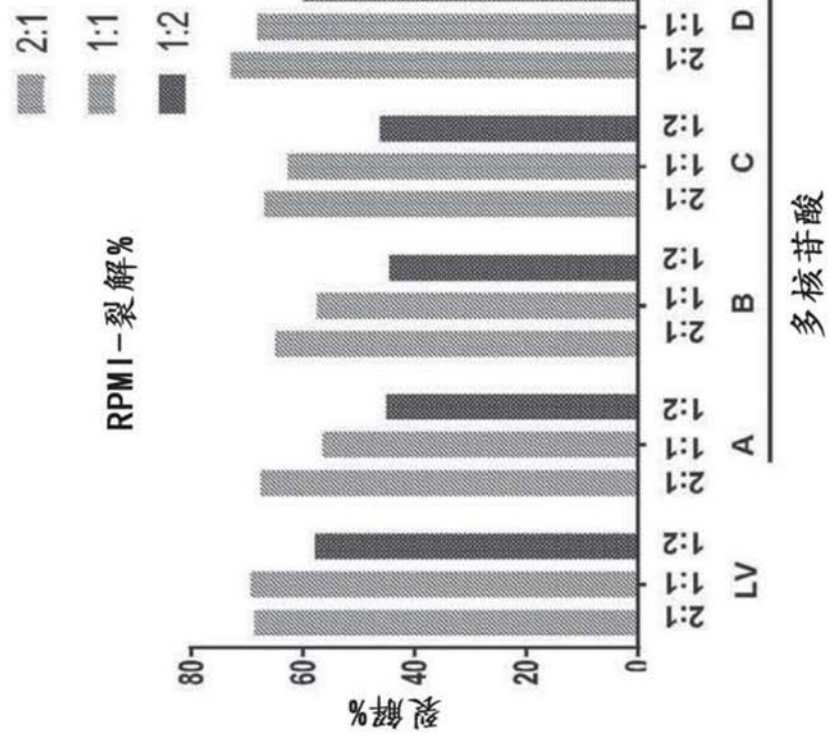


FIG. 3A



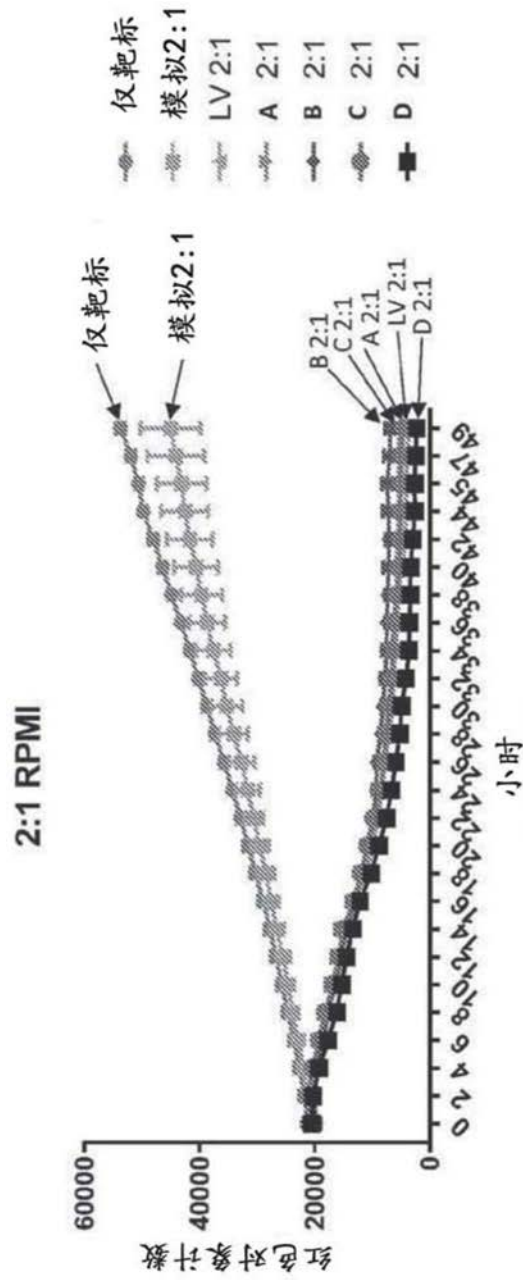


FIG.3C

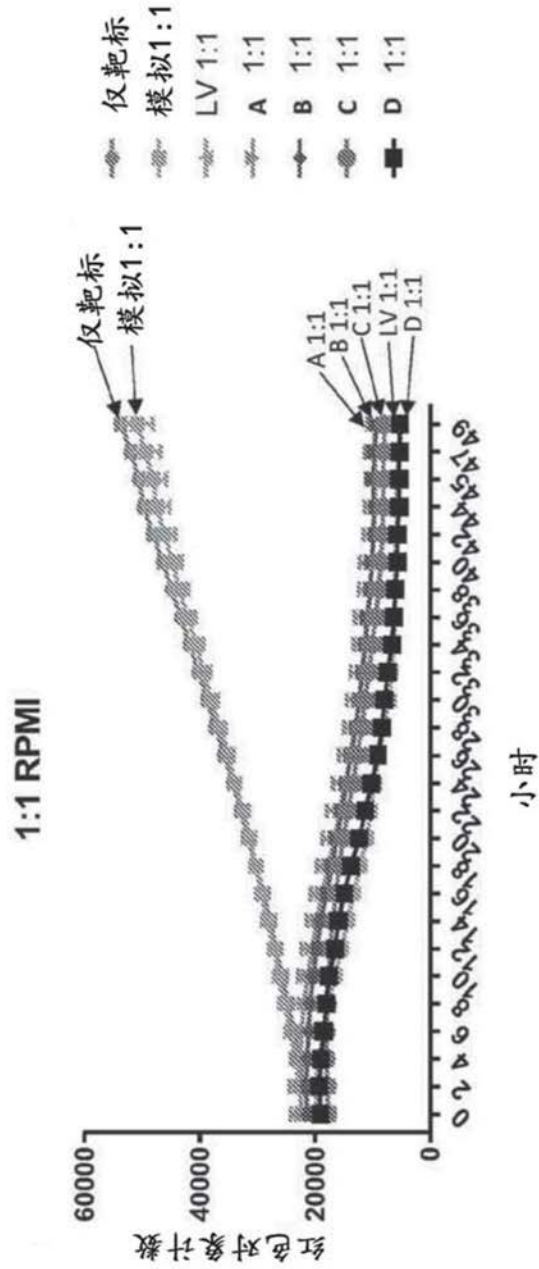


FIG. 3D

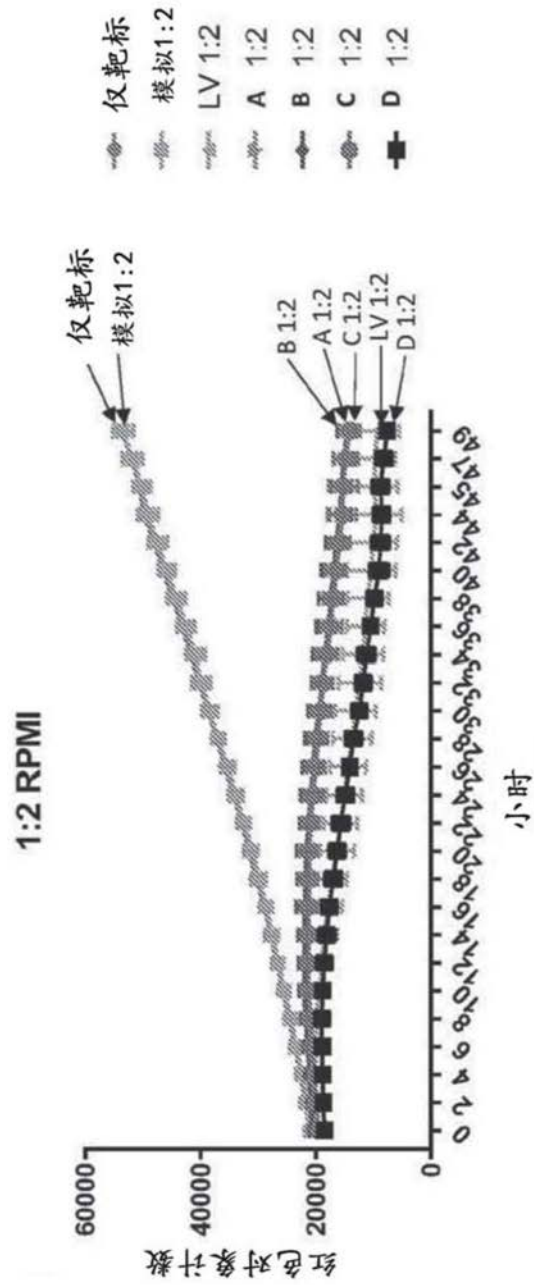


FIG. 3E

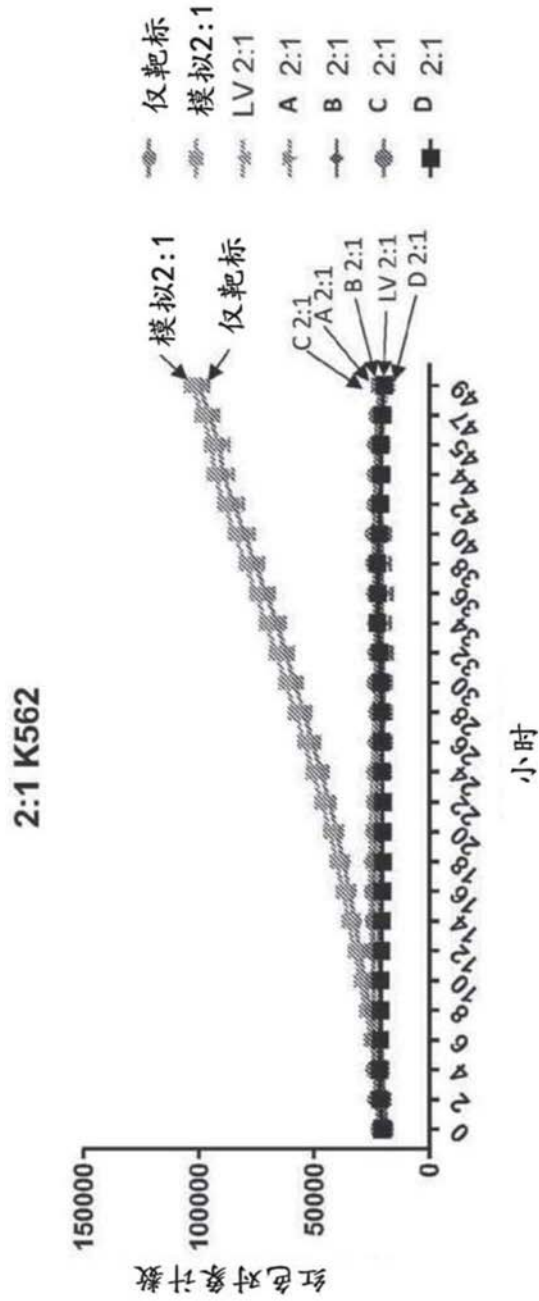


FIG. 3F

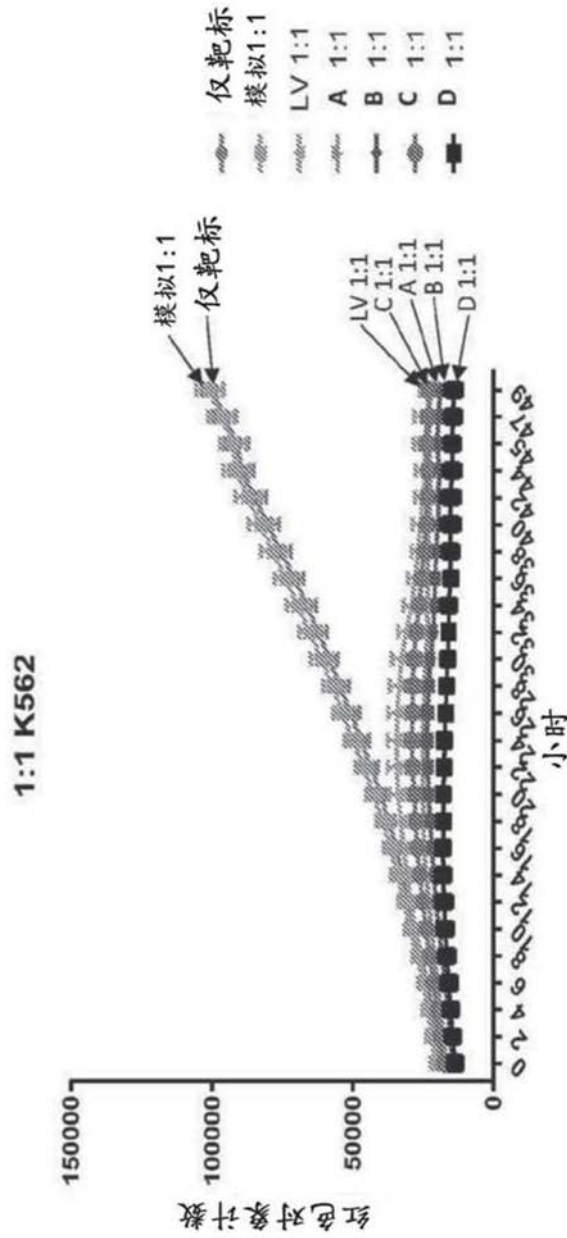


FIG. 3G

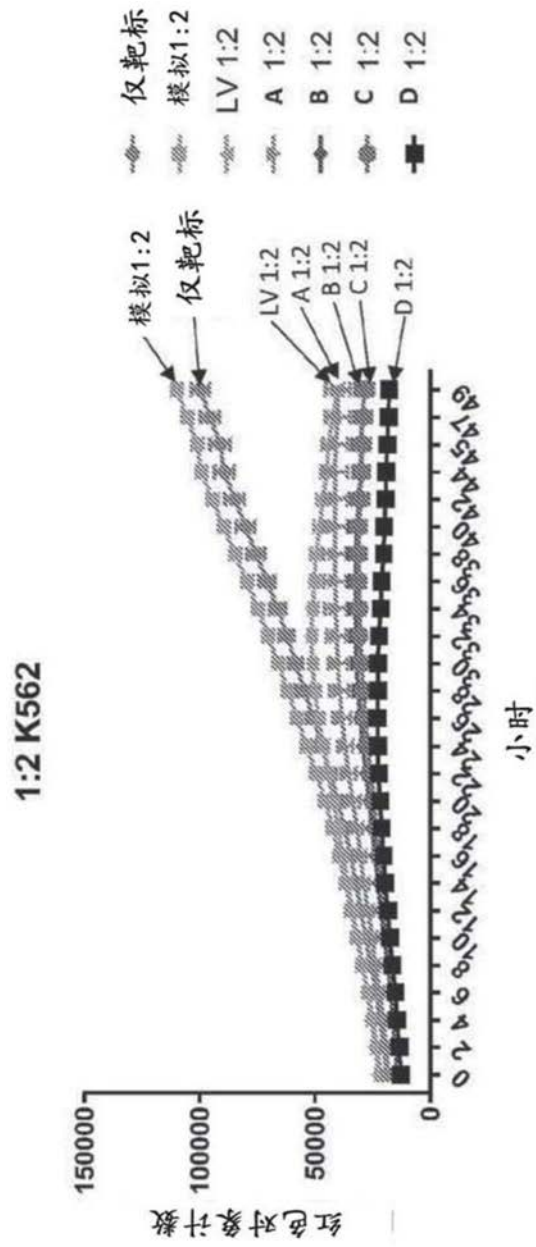


FIG. 3H

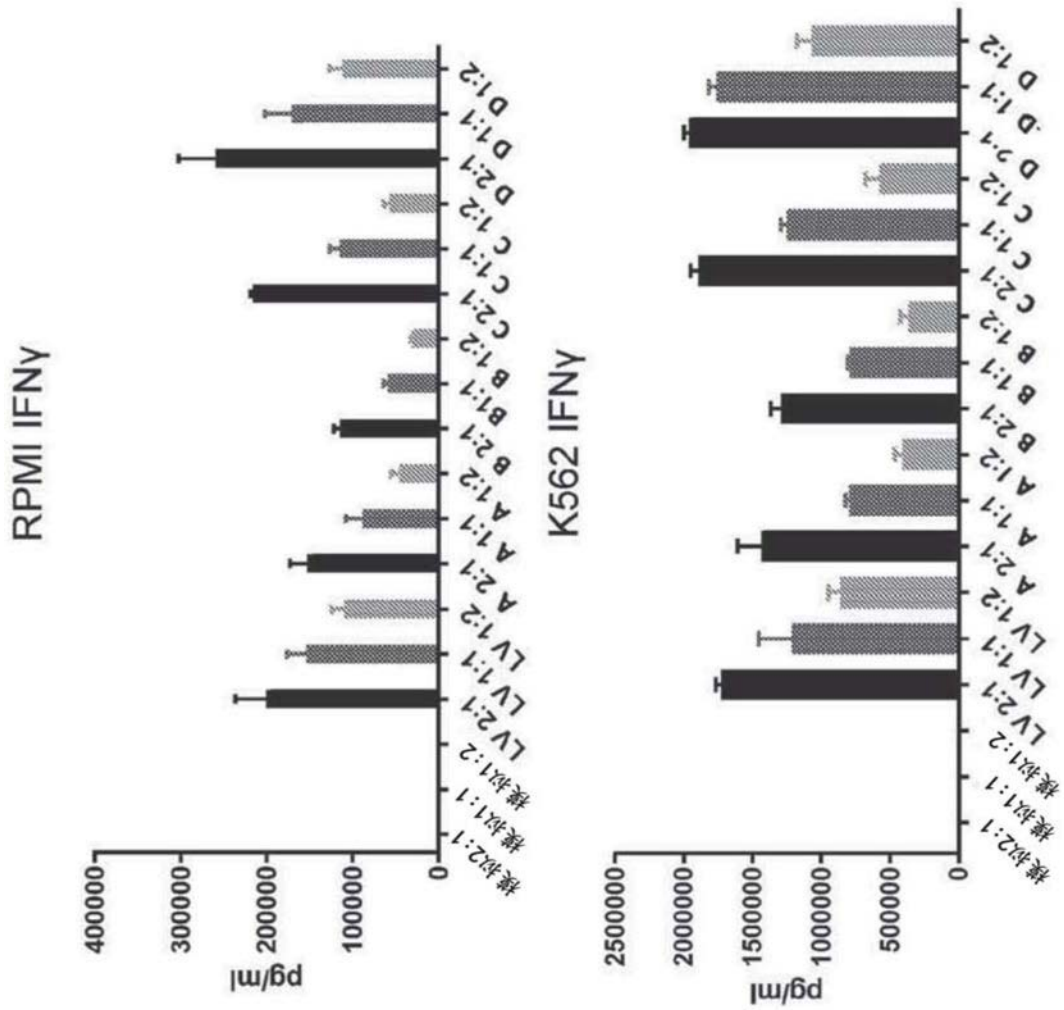


FIG.4A

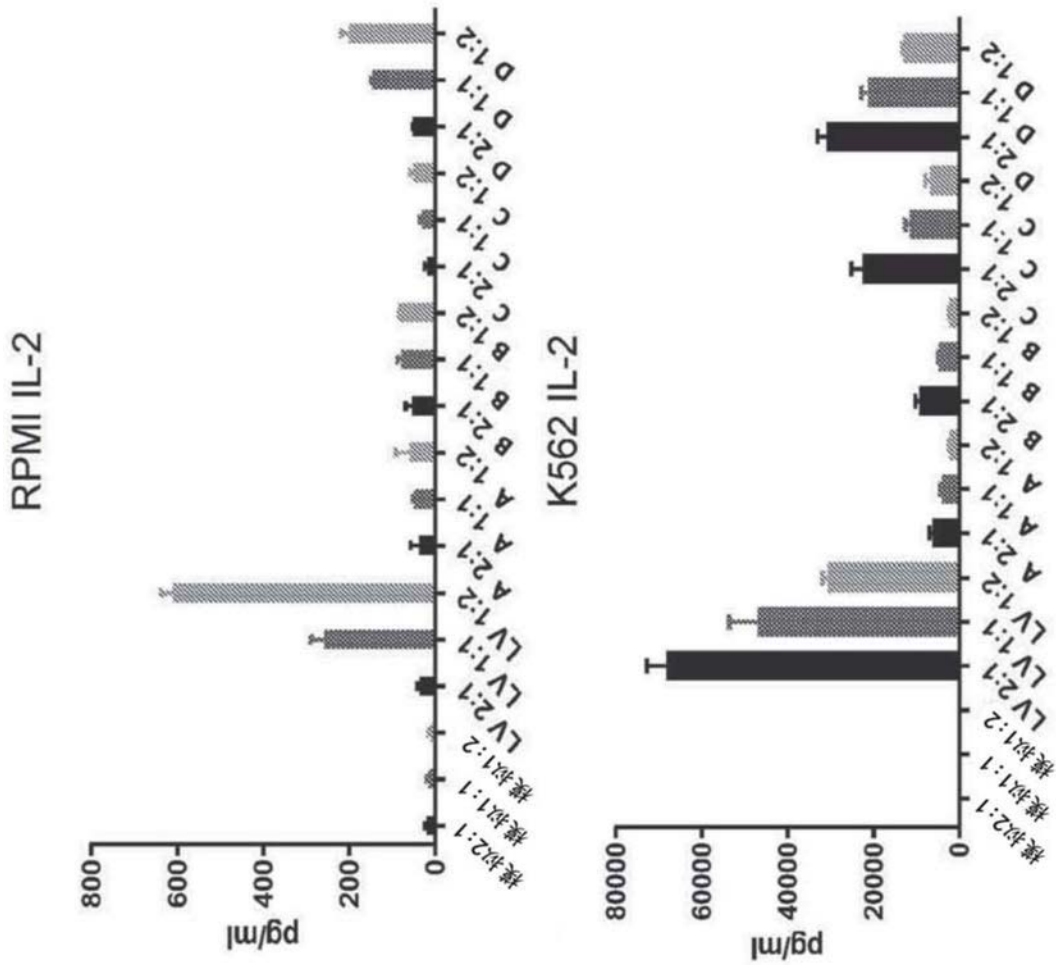


FIG. 4B

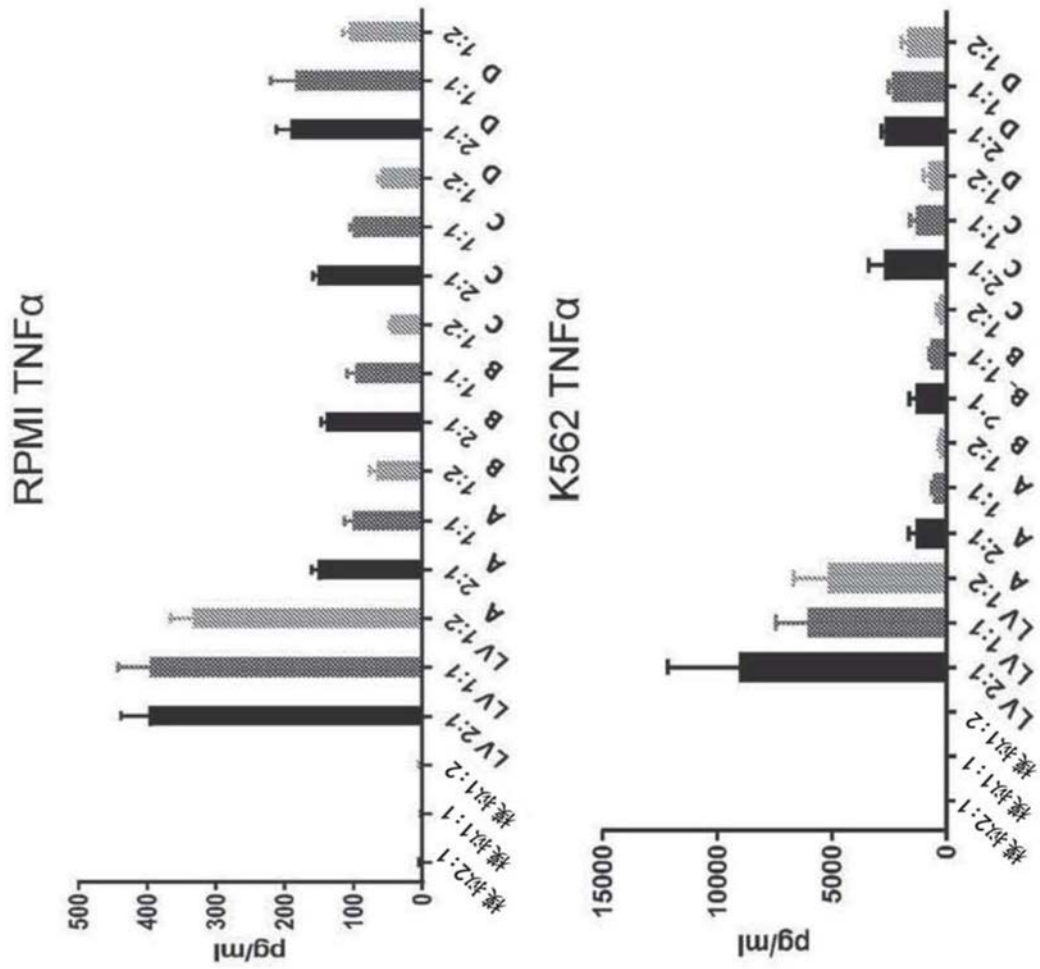


FIG.4C

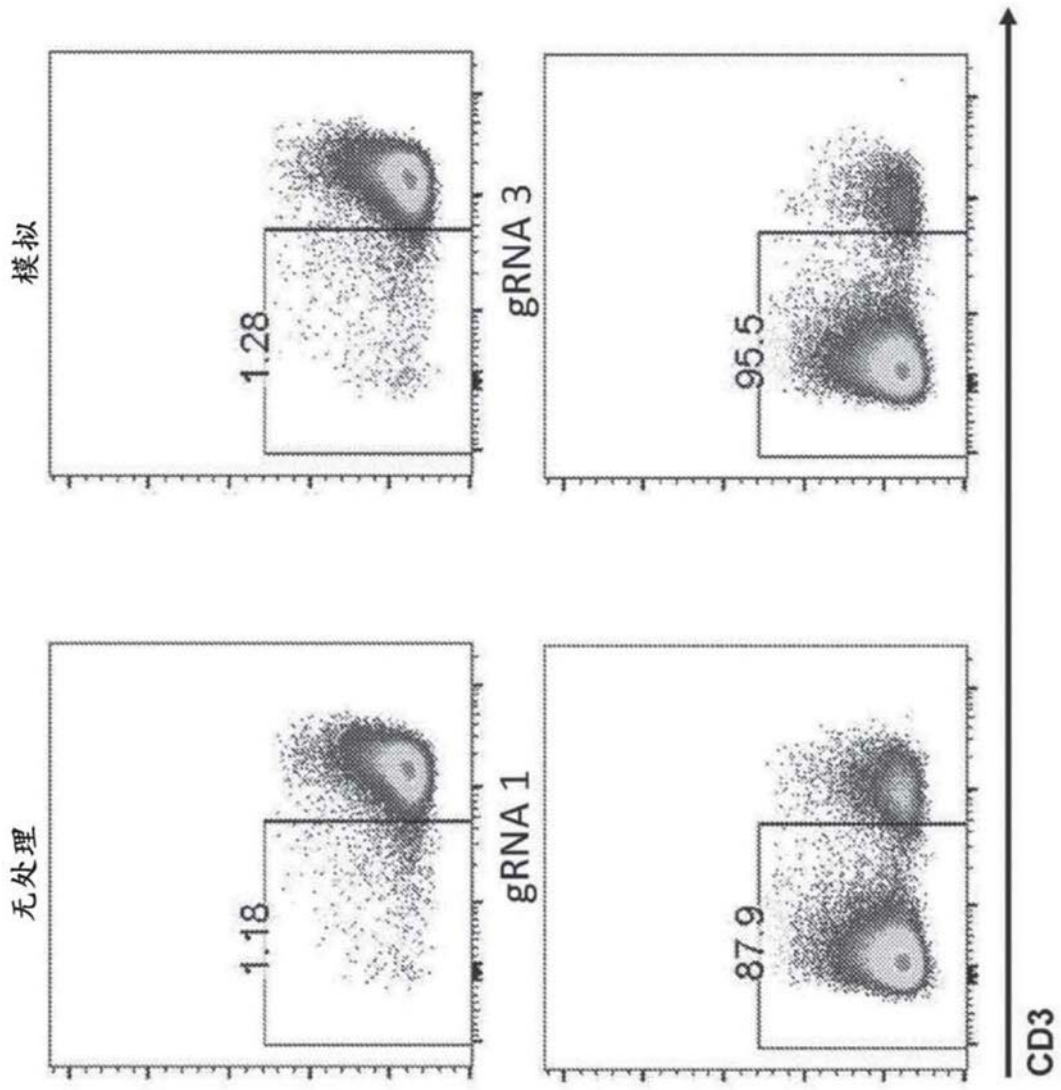


FIG. 5

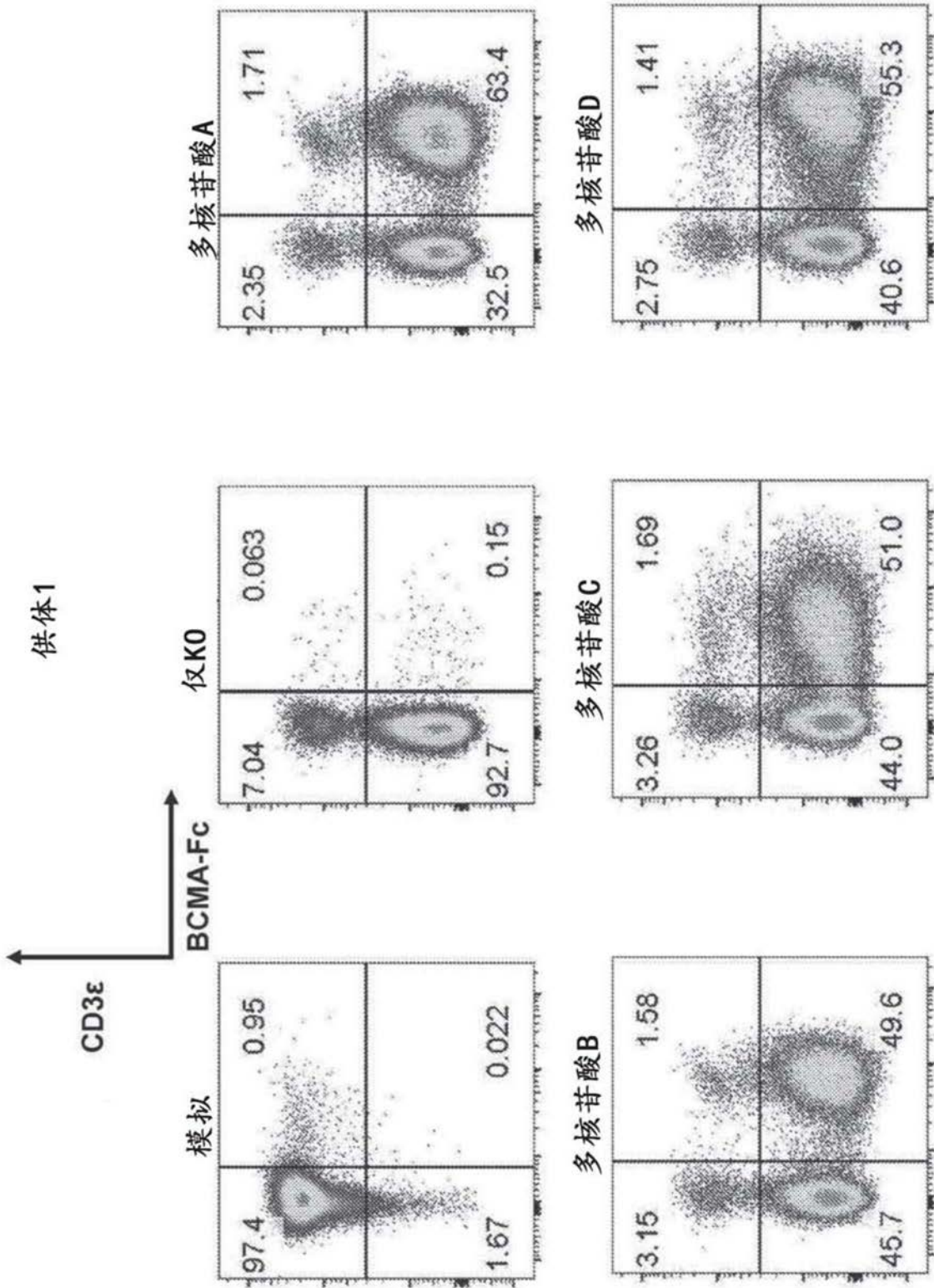


FIG.6A

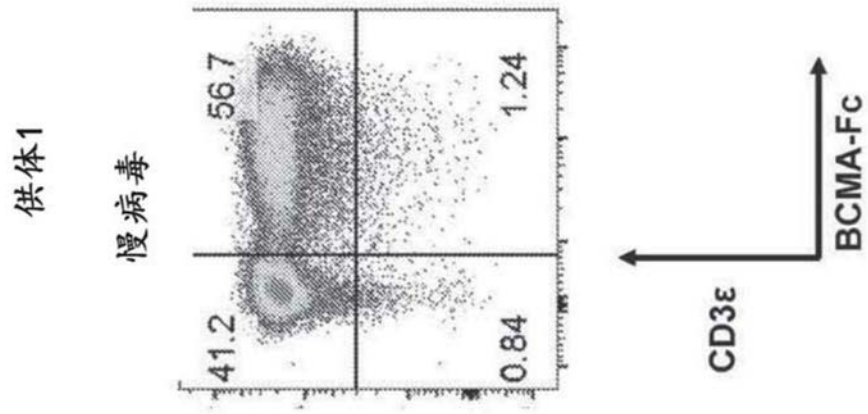


FIG.6B

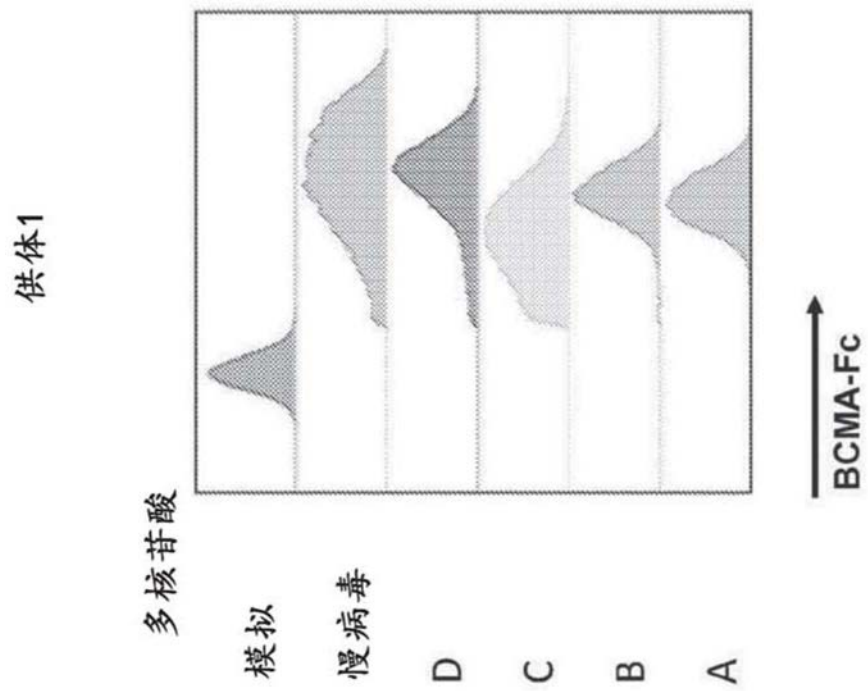


FIG.6C

FIG. 7B

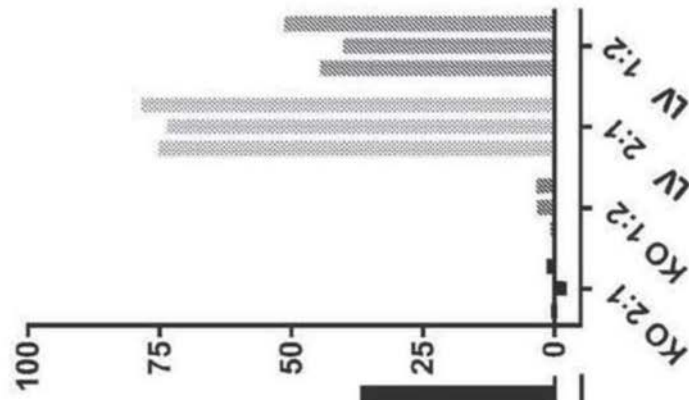


FIG. 7A

