



(12) 发明专利

(10) 授权公告号 CN 113906047 B

(45) 授权公告日 2025. 02. 14

(21) 申请号 202080040738.4

(22) 申请日 2020.04.03

(65) 同一申请的已公布的文献号
申请公布号 CN 113906047 A

(43) 申请公布日 2022.01.07

(30) 优先权数据
62/829,813 2019.04.05 US
62/838,465 2019.04.25 US
62/898,053 2019.09.10 US
62/972,231 2020.02.10 US

(85) PCT国际申请进入国家阶段日
2021.12.01

(86) PCT国际申请的申请数据
PCT/US2020/026558 2020.04.03

(87) PCT国际申请的公布数据
W02020/206238 EN 2020.10.08

(73) 专利权人 根路径基因组学公司

地址 美国马萨诸塞州

(72) 发明人 陈曦 伊利·波特

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 武晶晶

(51) Int.Cl.
C07K 14/725 (2006.01)
C40B 40/06 (2006.01)
C40B 20/00 (2006.01)
C40B 30/04 (2006.01)
C12N 5/0783 (2006.01)

(56) 对比文件
W0 2018132739 A2, 2018.07.19

审查员 李梦华

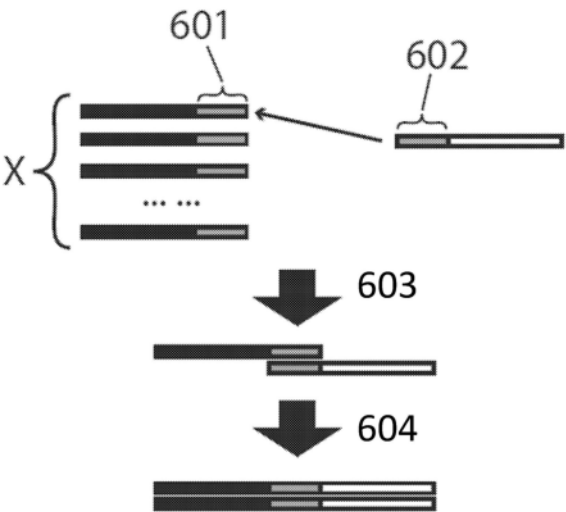
权利要求书4页 说明书62页
序列表81页 附图24页

(54) 发明名称

用于T细胞受体基因组组装的组合物和方法

(57) 摘要

本文提供了用于组装编码T细胞受体的核酸序列的组合物和方法。



1. 一种用于在同一区室中产生多种编码多种不同天然T细胞受体 (TCR) α 链或其部分的核酸分子的方法,所述方法包括:

(a) 提供第一多种核酸分子,所述第一多种核酸分子中的每种核酸分子包含编码不同的TCR α 链的CDR3 α 的序列和独特的连接序列;

(b) 提供第二多种核酸分子,所述第二多种中的每种核酸分子包含衍生自TCR α V (TRAV) 基因的序列和独特的抗连接序列,其中所述第二多种核酸分子包含衍生自至少两种不同TRAV基因的至少两种不同序列;以及

(c) 产生第三多种核酸分子,所述第三多种中的每种核酸分子包含来自来自(a)的所述第一多种核酸分子的所述编码CDR3 α 的序列中的一种的序列和来自(b)的所述第二多种核酸分子的至少两种不同TRAV基因中的一种的序列,其中产生包括将(a)的所述第一多种核酸分子与(b)的所述第二多种核酸分子在所述同一区室中接触以及将(a)的所述第一多种核酸分子中的核酸分子的独特的连接序列与包含来自所述至少两种不同TRAV基因中的单个给定TRAV基因的序列的(b)的所述第二多种核酸分子中的核酸分子的独特的抗连接序列特异性地连接,其中所述独特的抗连接序列与所述独特的连接序列互补,从而产生编码所述多种不同TCR α 链或其部分的多种所述核酸分子,并且其中所述单个给定TRAV基因与其所连接的所述CDR3 α 是同源的。

2. 如权利要求1所述的方法,其中所述来自所述单个给定TRAV基因的序列和(c)中所述与其所特异性地连接的编码所述CDR3 α 的序列来自在来自受试者的样品的T细胞中表达的同一天CR α 链。

3. 如权利要求1或2所述的方法,其中所述第一多种核酸分子中的每种核酸分子还包含TCR链的J区,并且其中所述J区和所述CDR3 α 来自同一天CR α 链。

4. 如权利要求1或2所述的方法,其中所述至少两种TRAV基因是人TRAV基因或小鼠TRAV基因。

5. 如权利要求1或2所述的方法,其中所述至少两种TRAV基因选自由人TRAV1-1、TRAV1-2、TRAV2、TRAV3、TRAV4、TRAV5、TRAV6、TRAV7、TRAV8-1、TRAV8-2、TRAV8-3、TRAV8-4、TRAV8-6、TRAV9-1、TRAV9-2、TRAV10、TRAV12-1、TRAV12-2、TRAV12-3、TRAV13-1、TRAV13-2、TRAV14、TRAV16、TRAV17、TRAV18、TRAV19、TRAV20、TRAV21、TRAV22、TRAV23、TRAV24、TRAV25、TRAV26-1、TRAV26-2、TRAV27、TRAV29、TRAV30、TRAV34、TRAV35、TRAV36、TRAV38-1、TRAV38-2、TRAV39、TRAV40和TRAV41组成的组。

6. 如权利要求1所述的方法,其中所述独特的连接序列或所述独特的抗连接序列编码与重排基因中编码CDR3 α 的序列相邻的TRAV基因的一部分。

7. 如权利要求1或2所述的方法,其中衍生自所述至少两种不同TRAV基因的所述至少两种序列中的每种序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和/或FR3的序列。

8. 如权利要求2所述的方法,其中所述受试者患有癌症。

9. 如权利要求1或2所述的方法,其中(a)的所述第一多种核酸分子中的每种核酸分子还包含编码TCR β 链的CDR3 β 的额外序列。

10. 如权利要求9所述的方法,其中(a)的所述第一多种核酸分子中的每种核酸分子包含所述TCR β 链的额外J区,并且其中所述额外J区和所述CDR3 β 来自同一天CR β 链。

11. 如权利要求9所述的方法,其中所述编码所述CDR3 α 的序列和所述编码所述CDR3 β 的

额外序列相隔至多100个核苷酸。

12. 如权利要求9所述的方法,其中所述TCR α 链和所述TCR β 链为TCR链的同源对。

13. 如权利要求1所述的方法,其中每种独特的连接序列或每种独特的抗连接序列编码与重排基因中编码所述CDR3 α 的所述序列相邻的所述TRAV基因中保守的同一氨基酸序列。

14. 如权利要求1或2所述的方法,其中(a)的所述第一多种核酸分子中的所述核酸分子和(b)的所述第二多种核酸分子中的所述核酸分子编码功能性TCR链或其部分。

15. 如权利要求1或2所述的方法,其中所述抗连接序列与(a)的所述第一多种核酸分子中的所述核酸分子的所述连接序列互补。

16. 如权利要求13所述的方法,其中每种独特的连接序列具有密码子多样化的核酸序列。

17. 如权利要求2所述的方法,其中所述来自所述单个给定TRAV基因的序列和(c)中所述与其所特异性地连接的编码所述CDR3 α 的序列来自在来自所述受试者的所述样品的肿瘤浸润性T细胞中表达的另一TCR α 链。

18. 如权利要求1或2所述的方法,其中连接包括使(a)的所述第一多种核酸分子中的所述核酸分子的所述连接序列与(b)的所述第二多种核酸分子中的所述核酸分子的所述抗连接序列杂交。

19. 如权利要求18所述的方法,其还包括(i)使用(a)的所述第一多种核酸分子中的所述核酸分子作为模板延伸所述第二多种核酸分子中的所述核酸分子的游离3'末端,和/或(ii)使用(b)的所述第二多种核酸分子中的所述核酸分子作为模板延伸(a)的所述第一多种核酸分子中的所述核酸分子的游离3'末端,以产生所述第三多种核酸分子中的核酸分子。

20. 如权利要求19所述的方法,其还包括将(a)的所述第一多种核酸分子中的所述核酸分子与(b)的所述第二多种核酸分子中的所述核酸分子连接。

21. 如权利要求20所述的方法,其还包括使所述第三多种核酸分子中的所述核酸分子与限制酶接触以产生粘性末端。

22. 如权利要求21所述的方法,其还包括使所述第三多种核酸分子中的所述核酸分子与额外的核酸分子接触。

23. 如权利要求22所述的方法,其中所述额外的核酸分子编码TCR链的恒定区或其部分。

24. 如权利要求23所述的方法,其还包括将所述第三多种核酸分子中的所述核酸分子与所述额外的核酸分子连接。

25. 如权利要求1或2所述的方法,其还包括,在(a)之前,对来自受试者的样品进行测序以获得所述受试者中表达的TCR的同源对的序列。

26. 如权利要求1或2所述的方法,其中编码至少五种不同的TCR α 链或其部分的至少五种不同的核酸分子在所述同一区室中产生。

27. 如权利要求1或2所述的方法,其中编码至少十种不同的TCR α 链或其部分的至少十种不同的核酸分子在所述同一区室中产生。

28. 如权利要求1或2所述的方法,其中所述同一区室是孔、管或液滴。

29. 如权利要求1或2所述的方法,其中所述连接序列包含独特的条形码。

30. 如权利要求29所述的方法,其中所述独特的条形码是引物结合位点。
31. 如权利要求30所述的方法,其还包括使用靶向所述独特的条形码的引物扩增所述第三多种核酸分子中的所述核酸分子。
32. 如权利要求1或2所述的方法,其还包括在细胞中表达编码(c)中产生的所述多种不同的TCR α 链或其部分的多种核酸分子中的至少一种核酸分子的序列。
33. 一种组合物,其包含
- (a) 多种核酸分子,其中所述多种核酸分子中的每种核酸分子包含(i)独特的抗连接序列、(ii)衍生自T细胞受体(TCR) V基因的序列并且(iii)不包含CDR3序列,其中所述多种中的第一核酸分子包含第一抗连接序列并且所述多种中的第二核酸分子包含第二抗连接序列,其中所述第一抗连接序列不同于所述第二抗连接序列,其中所述第一抗连接序列和所述第二抗连接序列编码同一氨基酸序列,所述氨基酸序列在与重排基因中编码所述CDR3的所述序列相邻的TRAV基因中是保守的,并且其中所述第一核酸分子和所述第二核酸分子的所述衍生自TCR V基因的序列衍生自不同TCR V基因;以及
- (b) 第一多种核酸分子,其中所述第一多种核酸分子中的每种核酸分子包含编码不同TCR链的不同CDR3的序列,其中所述第一多种核酸分子中的每种核酸分子包含独特的连接序列,所述独特的连接序列能够与(a)的所述多种核酸分子中的包含衍生自单个给定TCR V基因的序列的核酸分子的独特的抗连接序列特异性地连接,其中所述独特的抗连接序列与所述独特的连接序列互补。
34. 如权利要求33所述的组合物,其中所述组合物是液体组合物。
35. 如权利要求33或34所述的组合物,其中(a)的所述多种核酸分子和(b)的所述第一多种核酸分子在同一区室中。
36. 如权利要求33或34所述的组合物,其中所述衍生自所述TCR V基因的序列包含所述TCR V基因的至少十个核苷酸。
37. 如权利要求33或34所述的组合物,其中所述TCR V基因是TRAV基因、TRBV基因、TRGV基因或TRDV基因。
38. 如权利要求33或34所述的组合物,其中所述衍生自所述TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和/或FR3的序列。
39. 如权利要求33或34所述的组合物,其中(b)的所述第一多种核酸分子中的每种核酸分子还包含所述TCR链的J区。
40. 如权利要求33或34所述的组合物,其中(b)的所述第一多种核酸分子中的每种核酸分子还包含编码额外TCR链的额外CDR3的额外序列。
41. 如权利要求40所述的组合物,其中(b)的所述第一多种核酸分子中的每种核酸分子还包含所述额外TCR链的额外J区。
42. 如权利要求40所述的组合物,其中所述编码所述CDR3的序列和所述编码所述CDR3的额外序列相隔至多100个核苷酸。
43. 如权利要求40所述的组合物,其中所述TCR链和所述额外TCR链为TCR链的同源对。
44. 如权利要求33或34所述的组合物,其中所述第一抗连接序列或所述第二抗连接序列是密码子多样化的。
45. 如权利要求33所述的组合物,其中所述独特的连接序列或所述独特的抗连接序列

的长度大于或等于15个核苷酸并且长度小于或等于300个核苷酸。

46. 如权利要求33或34所述的组合物,其中(a)的所述多种核酸分子中的每种核酸分子不包含编码TCR链的恒定结构域的序列。

47. 如权利要求33或34所述的组合物,其中所述独特的连接序列与(a)的所述多种核酸分子中的包含衍生自单个给定TCR V基因的序列的核酸分子的所述独特的抗连接序列杂交。

48. 如权利要求33或34所述的组合物,其中(b)的所述第一多种核酸分子中的每种核酸分子包含独特的条形码。

49. 如权利要求48所述的组合物,其中所述独特的条形码是引物结合位点。

50. 如权利要求33或34所述的组合物,其中(b)的所述第一多种核酸分子中的每种核酸分子的所述独特的连接序列包含独特的条形码。

51. 如权利要求50所述的组合物,其中所述独特的条形码是引物结合位点。

52. 一种用于在同一区室中产生多种编码多种不同天然TCR链或TCR链的同源对的核酸分子的方法,其包括:

(a) 提供第一多种核酸分子,其中所述第一多种核酸分子中的每种核酸分子包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的序列和独特的连接序列,其中所述第一CDR3和所述第二CDR3来自TCR链的同源对,并且其中每种核酸分子编码来自TCR链的不同同源对的第一CDR3和第二CDR3;

(b) 提供第二多种核酸分子,其中所述第二多种核酸分子的每种核酸分子包含衍生自不同TCR V基因的序列和独特的抗连接序列;以及

(c) 使(a)的所述第一多种核酸分子与(b)的所述第二多种核酸分子接触,其中(a)的所述第一多种核酸分子中的核酸分子的独特的连接序列与(b)的所述第二多种核酸分子中的核酸分子的独特的抗连接序列特异性地连接以形成包含所述编码第一CDR3和第二CDR3的序列以及衍生自单个给定TCR V基因的序列的线性核酸分子,其中所述编码所述第一CDR3和所述第二CDR3的序列以及所述单个给定TCR V基因衍生自TCR链的同源对,并且其中所述独特的抗连接序列与所述独特的连接序列互补。

用于T细胞受体基因组装的组合物和方法

[0001] 交叉引用

[0002] 本申请要求2019年4月5日提交的美国临时专利申请号62/829,813、2019年4月25日提交的美国临时专利申请号62/838,465、2019年9月10日提交的美国临时专利申请号62/898,053和2020年2月10日提交的美国临时专利申请号62/972,231的权益,它们各自以引用的方式整体并入本文。

[0003] 序列表

[0004] 本申请含有序列表,所述序列表已以ASCII格式以电子方式提交,并以引用的方式整体并入本文。所述ASCII副本创建于2020年5月15日,名为53563-705_601_SL.txt,并且大小为69,187字节。

背景技术

[0005] T细胞受体(TCR)可负责识别抗原-主要组织相容性复合物,从而导致炎症应答的启动。存在许多T细胞子集,包括细胞毒性T细胞和辅助T细胞。细胞毒性T细胞(也称为CD8+T细胞)杀伤异常细胞,例如病毒感染的细胞或肿瘤细胞。辅助T细胞(也称为CD4+T细胞)有助于其他免疫细胞的激活和成熟。细胞毒性T细胞和辅助T细胞两者均在识别特异性靶抗原后执行它们的功能,从而触发它们各自的反应。T细胞的抗原特异性可由T细胞表面上表达的TCR来定义。T细胞受体是由两条多肽链组成的异源二聚体蛋白,最常见的是 α 链和 β 链,但是少数T细胞可以表达 γ 和 δ 链。TCR的特定氨基酸序列和所得三维结构定义了TCR抗原特异性和亲和力。任何单个T细胞的TCR链的氨基酸和编码DNA序列在生物体的整个TCR库中几乎总是独特的或丰度非常低,因为存在大量可能的TCR序列。这种巨大的序列多样性可以在T细胞发育过程中通过多种细胞机制实现,并且可能是免疫系统对多种潜在抗原作出反应的能力的关键方面。

[0006] 分析TCR库可有助于更好地了解免疫系统特征以及疾病的病因和进展,特别是那些具有未知抗原性触发物的疾病。TCR库的极度多样性和TCR的二分性质可代表主要的分析挑战。高通量测序可以允许更大的测序深度和TCR克隆型丰度的显著更精确的定量,尽管成本高于谱型分析。

发明内容

[0007] 本文提供了组装编码天然配对的T细胞受体(TCR)(或同源TCR对)的核酸序列的组合物和方法。例如,TCR可包含TCR α 链和TCR β 链,或者TCR可包含TCR γ 链和TCR δ 链。编码天然配对的TCR的序列可使用各种方法鉴定,包括但不限于使用单细胞条形码和测序技术。获得编码天然配对的TCR的序列后,本文所述的组合物和方法可用于构建或组装一种或多种核酸序列,以快速、高通量和成本有效的方式在任何给定的宿主细胞中表达天然配对的TCR。所述一种或多种核酸序列可包含大于或等于约1、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000或更多种编码不同TCR的不同序

列。

[0008] 在一个方面,本公开提供了一种用于产生编码T细胞受体 (TCR) 链或其部分的核酸分子的方法,其包括:(a) 提供至少一种核酸分子,其包含编码TCR链的CDR3的序列;(b) 提供多种核酸分子,所述多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列,其中所述多种核酸分子包含衍生自至少两种不同TCR V基因的至少两种不同序列;以及(c) 使(a) 的至少一种核酸分子与(b) 的多种核酸分子在同一区室中接触,其中(a) 的至少一种核酸分子能够与所述多种核酸分子中的核酸分子连接,以产生包含编码CDR3的序列和衍生自至少两种不同TCR V基因中的一种的序列的第三核酸分子,从而产生编码TCR链或其部分的核酸分子。在一些实施方案中,至少一种核酸分子包含至少约2、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000或更多种不同的序列。在一些实施方案中,多种核酸分子(所述多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列) 包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80或更多种衍生自不同TCR V基因的不同序列。

[0009] 在一些实施方案中,至少一种核酸分子包含第一多种核酸分子,其中所述第一多种核酸分子中的每种核酸分子包含编码TCR链的CDR3的序列。在一些实施方案中,(a) 的至少一种核酸分子能够与所述多种核酸分子中的核酸分子特异性地连接,所述多种核酸分子中的核酸分子包含衍生自至少两种不同TCR V基因中的任何单个给定TCR V基因的序列。在一些实施方案中,至少一种核酸分子还包含TCR链的J区。在一些实施方案中,第一多种核酸分子中的每种核酸分子还包含TCR链的J区。在一些实施方案中,所述至少两种TCR V基因是人TCR V基因或小鼠TCR V基因。在一些实施方案中,所述至少两种TCR V基因选自由人TRAV1-1、TRAV1-2、TRAV2、TRAV3、TRAV4、TRAV5、TRAV6、TRAV7、TRAV8-1、TRAV8-2、TRAV8-3、TRAV8-4、TRAV8-6、TRAV9-1、TRAV9-2、TRAV10、TRAV12-1、TRAV12-2、TRAV12-3、TRAV13-1、TRAV13-2、TRAV14、TRAV16、TRAV17、TRAV18、TRAV19、TRAV20、TRAV21、TRAV22、TRAV23、TRAV24、TRAV25、TRAV26-1、TRAV26-2、TRAV27、TRAV29、TRAV30、TRAV34、TRAV35、TRAV36、TRAV38-1、TRAV38-2、TRAV39、TRAV40和TRAV41组成的组。在一些实施方案中,所述至少两种TCR V基因选自由人TRBV2、TRBV3-1、TRBV4-1、TRBV4-2、TRBV4-3、TRBV5-1、TRBV5-4、TRBV5-5、TRBV5-6、TRBV5-8、TRBV6-1、TRBV6-2、TRBV6-3、TRBV6-4、TRBV6-5、TRBV6-6、TRBV6-8、TRBV6-9、TRBV7-2、TRBV7-3、TRBV7-4、TRBV7-6、TRBV7-7、TRBV7-8、TRBV7-9、TRBV9、TRBV10-1、TRBV10-2、TRBV10-3、TRBV11-1、TRBV11-2、TRBV11-3、TRBV12-3、TRBV12-4、TRBV12-5、TRBV13、TRBV14、TRBV15、TRBV16、TRBV18、TRBV19、TRBV20-1、TRBV24-1、TRBV25-1、TRBV27、TRBV28、TRBV29-1和TRBV30组成的组。在一些实施方案中,衍生自至少两种不同TCR V基因的多种序列中的每种序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和/或FR3的序列。在一些实施方案中,TCR链是TCR α 链、TCR β 链、TCR γ 链或TCR δ 链。在一些实施方案中,至少一种核酸分子还包含编码额外TCR链的额外CDR3的额外序列。在一些实施方案中,至少一种核酸分子包含额外TCR链的额外J区。在一些实施方案中,编码CDR3的序列和编码额外CDR3的额外序列相隔至多100个核苷酸。在一些实施方案中,TCR链和额外TCR链是TCR链的

同源对。在一些实施方案中,至少一种核酸分子包含连接序列,所述连接序列能够将至少一种核酸分子与多种核酸分子中的核酸分子连接以产生第三核酸分子。在一些实施方案中,至少一种核酸分子和多种核酸分子中的核酸分子编码功能性TCR链或其部分。在一些实施方案中,多种核酸分子中的核酸分子包含抗连接序列,所述抗连接序列与(a)的至少一种核酸分子的连接序列互补。在一些实施方案中,所述方法还包括将(a)的至少一种核酸分子和(b)的多种核酸分子中的核酸分子连接。在一些实施方案中,连接包括使(a)的至少一种核酸分子与(b)的多种核酸分子中的核酸分子杂交。在一些实施方案中,杂交包括使(a)的至少一种核酸分子的连接序列与(b)的多种核酸分子中的核酸分子的抗连接序列杂交。在一些实施方案中,所述方法还包括(i)使用(a)的至少一种核酸分子作为模板延伸多种核酸分子中的核酸分子的游离3'末端,和/或(ii)使用多种核酸分子中的核酸分子作为模板延伸(a)的至少一种核酸分子的游离3'末端,以产生第三核酸分子。在一些实施方案中,所述方法还包括将(a)的至少一种核酸分子与(b)的多种核酸分子中的核酸分子连接。在一些实施方案中,所述方法还包括使第三核酸分子与限制酶接触以产生粘性末端。在一些实施方案中,所述方法还包括使第三核酸分子与额外的核酸分子接触。在一些实施方案中,额外的核酸分子编码TCR链的恒定区或其部分。在一些实施方案中,所述方法还包括将第三核酸分子与额外的核酸分子连接。在一些实施方案中,多种核酸分子(每种编码不同的TCR链或其部分)在同一区室中产生。在一些实施方案中,多种核酸分子中的至少五种不同的核酸分子在同一区室中产生。在一些实施方案中,多种核酸分子中的至少十种不同的核酸分子在同一区室中产生。在一些实施方案中,多种核酸分子中的至少20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000种或更多种不同的核酸分子在同一区室中产生。在一些实施方案中,同一区室是孔、管或液滴。在一些实施方案中,至少一种核酸分子包含独特的条形码。在一些实施方案中,独特的条形码是引物结合位点。在一些实施方案中,连接序列包含独特的条形码。在一些实施方案中,独特的条形码是引物结合位点。

[0010] 在另一方面,本公开提供了一种组合物,其包含(a)多种核酸分子,其中所述多种核酸分子中的每种核酸分子包含衍生自T细胞受体(TCR)V基因的序列并且不包含CDR3序列,其中所述多种中的第一核酸分子包含第一抗连接序列并且所述多种中的第二核酸分子包含第二抗连接序列,其中所述第一抗连接序列不同于所述第二抗连接序列,并且其中所述第一核酸分子和所述第二核酸分子的衍生自TCR V基因的序列衍生自不同TCR V基因;以及(b)至少一种核酸分子,其包含编码TCR链的CDR3的序列,其中所述至少一种核酸分子还包含与所述第一抗连接序列互补的第一连接序列。

[0011] 在一些实施方案中,组合物是液体组合物。在一些实施方案中,(a)的多种核酸分子和(b)的至少一种核酸分子在同一区室中。在一些实施方案中,衍生自TCR V基因的序列包含TCR V基因的至少十个核苷酸。在一些实施方案中,TCR V基因是TRAV基因、TRBV基因、TRGV基因或TRDV基因。在一些实施方案中,衍生自TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和/或FR3的序列。在一些实施方案中,至少一种核酸分子还包含TCR链的J区。在一些实施方案中,至少一种核酸分子还包含编码额外TCR链的额外CDR3的额外序列。在一些实施方案中,至少一种核酸分子还包含额外TCR链的额外J区。在一些实施方

案中,编码CDR3的序列和编码CDR3的额外序列相隔至多100个核苷酸。在一些实施方案中,TCR链和额外TCR链是TCR链的同源对。在一些实施方案中,(b)的至少一种核酸分子包含第一多种核酸分子,并且其中所述第一多种核酸分子中的每种核酸分子包含编码TCR链的CDR3的序列。在一些实施方案中,第一多种核酸分子中的每种核酸分子编码不同TCR链的不同CDR3。在一些实施方案中,第一多种核酸分子中的每种核酸分子包含不同的连接序列,所述不同的连接序列能够与多种核酸分子中的包含衍生自任何单个给定TCR V基因的序列的核酸分子特异性地连接。在一些实施方案中,第一抗连接序列或第二抗连接序列包含TCR V基因序列。在一些实施方案中,TCR V基因序列包含与重排基因中编码CDR3的序列相邻的TCR V基因的至少三个核苷酸。在一些实施方案中,第一抗连接序列或第二抗连接序列包含预定序列。在一些实施方案中,第一连接序列与第一抗连接序列杂交。在一些实施方案中,(b)的至少一种核酸分子包含独特的条形码。在一些实施方案中,独特的条形码是引物结合位点。在一些实施方案中,至少一种核酸分子的第一连接序列包含独特的条形码。在一些实施方案中,独特的条形码是引物结合位点。

[0012] 在另一方面,本公开提供了一种用于产生多种核酸分子的方法,其包括:提供第一多种核酸分子,其中所述第一多种核酸分子中的核酸分子包含编码第一T细胞受体 (TCR) 链的第一CDR3和第二TCR链的第二CDR3的序列,其中所述第一CDR3和所述第二CDR3来自TCR链的同源对;提供第二多种核酸分子,其中所述第二多种核酸分子中的核酸分子包含衍生自TCR V基因的序列,其中所述核酸分子不包含编码恒定结构域的序列;以及使所述第一多种核酸分子与所述第二多种核酸分子接触,其中所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接以形成包含编码所述第一CDR3和所述第二CDR3的序列以及衍生自所述TCR V基因的序列的核酸分子,其中所述编码第一CDR3和第二CDR3的序列以及TCR V基因衍生自TCR链的同源对。

[0013] 在一些实施方案中,第一多种核酸分子中的每种核酸分子包含编码第一TCR链的不同第一CDR3和/或第二TCR链的不同CDR3的序列。在一些实施方案中,第一多种核酸分子包含至少约2、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000或更多种不同的序列。在一些实施方案中,第二多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。在一些实施方案中,第二多种核酸分子包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80或更多种不同的TCR V基因。在一些实施方案中,第一多种核酸分子和第二多种核酸分子在同一区室中接触。在一些实施方案中,第一多种核酸分子中的核酸分子还包含连接序列,其中所述连接序列将第一多种核酸分子中的核酸分子与第二多种核酸分子中的核酸分子连接。在一些实施方案中,第二多种核酸分子中的核酸分子还包含抗连接序列,所述抗连接序列与连接序列互补。在一些实施方案中,连接序列与抗连接序列杂交以将第一多种核酸分子中的核酸分子与第二多种核酸分子中的核酸分子连接。在一些实施方案中,连接序列是密码子多样化的,使得第一多种核酸分子中的核酸分子的连接序列不同于第一多种核酸分子中的其他核酸分子的其他连接序列。在一些实施方案中,第一多种核酸

分子中的核酸分子还包含第一TCR链的第一J区和/或第二TCR链的第二J区。在一些实施方案中, (i) 第一TCR链是TCR α 链并且第二TCR链是TCR β 链, 或(ii) 第一TCR链是TCR γ 链并且第二TCR链是TCR δ 链。在一些实施方案中, TCR V基因是TRAV基因、TRBV基因、TRGV基因或TRDV基因。在一些实施方案中, 第二多种核酸分子中的核酸分子是双链核酸分子。在一些实施方案中, 第二多种核酸分子中的核酸分子还包含编码自切割肽的一部分的序列。在一些实施方案中, 抗连接序列是第二多种核酸分子中的核酸分子的突出端。在一些实施方案中, 连接序列或抗连接序列的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100、150、200个或更多个核苷酸。在一些实施方案中, 所述方法还包括(i) 延伸与第二多种核酸分子中的核酸分子杂交的第一多种核酸分子中的核酸分子的3'末端和/或(ii) 延伸与第一多种核酸分子中的核酸分子杂交的第二多种核酸分子中的核酸分子的3'末端。在一些实施方案中, 所述方法还包括将第一多种核酸分子中的核酸分子与第二多种核酸分子中的核酸分子连接。

[0014] 在一些实施方案中, 所述方法还包括使包含编码第一CDR3和第二CDR3的序列以及衍生自TCR V基因的序列的核酸分子与限制酶接触以产生粘性末端。在一些实施方案中, 所述方法还包括使包含编码第一CDR3和第二CDR3的序列以及衍生自TCR V基因的序列的核酸分子与包含编码恒定区或其部分的序列的额外的核酸分子接触。在一些实施方案中, 所述方法还包括将包含编码第一CDR3和第二CDR3的序列以及衍生自TCR V基因的序列的核酸分子通过粘性末端与额外的核酸分子连接。在一些实施方案中, 编码第一CDR3的序列和编码第二CDR3的序列相隔至多约100、90、80、70、60、50、40、30、20、10或5个核苷酸。在一些实施方案中, 衍生自TCR V基因的序列包含编码FR1、CDR1、FR2、CDR2和FR3的序列。在一些实施方案中, 衍生自TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和FR3的序列。

[0015] 在另一方面, 本公开提供了一种组合物, 其包含: 第一多种核酸分子, 其中所述第一多种核酸分子中的每种核酸分子包含编码第一T细胞受体 (TCR) 链的第一CDR3和第二TCR链的第二CDR3的序列, 其中所述第一CDR3和第二CDR3来自TCR链的同源对; 和第二多种核酸分子, 其中所述第二多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列, 并且其中所述第二多种核酸分子中的每种核酸分子不包含编码所述第一CDR3和所述第二CDR3的序列; 其中(i) 所述第一多种核酸分子中的每种核酸分子包含编码不同的第一CDR3和/或第二CDR3的序列, 且/或(ii) 所述第二多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。在一些实施方案中, 第一多种核酸分子包含至少约2、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000或更多种不同的序列。在一些实施方案中, 第二多种核酸分子包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80或更多种不同的TCR V基因。

[0016] 在一些实施方案中, 第一多种核酸分子中的每种核酸分子还包含连接序列, 其中给定连接序列可用于将第一多种核酸分子中的给定核酸分子与第二多种核酸分子中的给

定核酸分子连接。在一些实施方案中,第二多种核酸分子中的每种核酸分子还包含抗连接序列,所述抗连接序列与连接序列互补。在一些实施方案中,连接序列是密码子多样化的,使得第一多种核酸分子中的给定核酸分子的给定连接序列不同于第一多种核酸分子中的其他核酸分子的其他连接序列。在一些实施方案中,连接序列编码氨基酸序列。在一些实施方案中,连接序列与编码第一TCR链的第一CDR3和第二TCR链的第二CDR3的序列在框内。在一些实施方案中,连接序列包含至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100、150、200个或更多个核苷酸。在一些实施方案中,连接序列包含与编码第一TCR链的第一CDR3或第二TCR链的第二CDR3的序列相邻的TCR V基因的至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100、150、200个或更多个核苷酸。在一些实施方案中,由给定连接序列编码的给定氨基酸序列与由至少一种其他连接序列编码的至少一种其他氨基酸序列相同或基本上相同。在一些实施方案中,由给定连接序列编码的给定氨基酸序列不同于由其他连接序列编码的其他氨基酸序列。在一些实施方案中,第一多种核酸分子中的每种核酸分子还包含第一TCR链的第一J区和/或第二TCR链的第二J区。在一些实施方案中,组合物是液体组合物。在一些实施方案中,第一多种核酸分子和第二多种核酸分子在同一区室内。在一些实施方案中,第一多种核酸分子中的给定核酸分子通过给定连接序列与第二多种核酸分子中的给定核酸分子连接。在一些实施方案中,第一多种核酸分子中的给定核酸分子通过与给定抗连接序列杂交的给定连接序列与第二多种核酸分子中的给定核酸分子杂交。在一些实施方案中,编码第一CDR3的序列和编码第二CDR3的序列相隔至多100个核苷酸。在一些实施方案中,衍生自TCR V基因的序列包含编码FR1、CDR1、FR2、CDR2和FR3的序列。在一些实施方案中,衍生自TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和FR3的序列。在一些实施方案中,第一多种核酸分子或第二多种分子中的每种核酸分子是化学合成的。在一些实施方案中,第一多种核酸分子中的每种核酸分子的长度为至多约250、240、230、220、210、200、190、180、170、160、150、140、130、120、110、100或50个核苷酸。

[0017] 在另一方面,本公开提供了一种包含多种核酸分子的组合物,所述多种核酸分子中的每种核酸分子包含衍生自T细胞受体(TCR) V基因序列的序列,其中所述多种核酸分子包含具有第一连接序列的第一核酸分子和具有第二连接序列的第二核酸分子,其中所述第一连接序列不同于所述第二连接序列。

[0018] 在一些实施方案中,多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。在一些实施方案中,多种核酸分子中的每种核酸分子包含不同的连接序列。在一些实施方案中,多种核酸分子中的每种核酸分子不包含编码TCR链的CDR3的序列。在一些实施方案中,多种核酸分子中的每种核酸分子不包含编码TCR链的恒定结构域的序列。在一些实施方案中,衍生自TCR V基因的序列包含TCR V基因的至少10、20、30、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000个或更多个核苷酸。在一些实施方案中,TCR V基因是TRAV基因、TRBV基因、TRGV基因或TRDV基因。

[0019] 在另一方面,本公开提供了一种包含多种核酸分子的组合物,所述多种核酸分子中的每种核酸分子编码T细胞受体(TCR)链的CDR3或其部分,其中所述多种核酸分子包含具

有第一连接序列的第一核酸分子和具有第二连接序列的第二核酸分子,其中所述第一连接序列不同于所述第二连接序列。

[0020] 在一些实施方案中,多种核酸分子中的每种核酸分子还包含TCR链的J区。在一些实施方案中,多种核酸分子中的每种核酸分子编码第一TCR链的第一CDR3或其部分和第二TCR链的第二CDR3或其部分。在一些实施方案中,多种核酸分子中的每种核酸分子还包含第一TCR链的第一J区和第二TCR链的第二J区。在一些实施方案中,多种核酸分子中的每种核酸分子编码不同TCR链的不同CDR3或其部分。在一些实施方案中,多种核酸分子中的每种核酸分子包含不同的连接序列。在一些实施方案中,多种核酸分子中的每种核酸分子不包含大于200、150、100、80、50、40、30、20或10个核苷酸的TCR V基因。在一些实施方案中,多种核酸分子中的每种核酸分子不包含编码TCR链的恒定结构域的序列。在一些实施方案中,第一连接序列或第二连接序列包含衍生自TCR V基因的序列。在一些实施方案中,衍生自TCR V基因的序列包含与重排基因中编码CDR3的序列相邻的TCR V基因的至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100、150、200个或更多个核苷酸。在一些实施方案中,第一连接序列或第二连接序列包含预定序列。在一些实施方案中,第一连接序列或第二连接序列包含与TCR V基因序列互补的序列。在一些实施方案中,组合物还包含第二多种核酸分子,所述第二多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列。在一些实施方案中,第二多种中的第一核酸分子包含第一抗连接序列,所述第一抗连接序列与第一连接序列互补。在一些实施方案中,第二多种中的第二核酸分子包含第二抗连接序列,所述第二抗连接序列与第二连接序列互补。在一些实施方案中,第二多种中的第一核酸分子的第一抗连接序列与第一多种中的第一核酸分子的第一连接序列连接。在一些实施方案中,第二多种中的第二核酸分子的第二抗连接序列与第一多种中的第二核酸分子的第二连接序列连接。

[0021] 在另一方面,本公开提供了一种包含多种核酸分子的组合物,每种包含编码T细胞受体 (TCR) 链的至少十个氨基酸 (例如,在一些情况下,编码至少约10、15、20、25、30、35、40、45、50、100、200个或更多个氨基酸) 的序列,其中所述多种核酸分子包含具有第一连接序列的第一核酸分子和具有第二连接序列的第二核酸分子,其中所述第一连接序列不同于所述第二连接序列,其中所述第一连接序列或所述第二连接序列编码TCR链的一部分,并且其中所述第一连接序列或所述第二连接序列与编码TCR链的至少十个 (例如,在一些情况下,编码至少约10、15、20、25、30、35、40、45、50、100、200个或更多个氨基酸) 氨基酸的序列在框内。

[0022] 在一些实施方案中,第一连接序列或第二连接序列包含TCR链基因的至少4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100、150、200个或更多个连续核苷酸,并且与编码TCR链的至少十个氨基酸的序列在框内。在一些实施方案中,第一连接序列和第二连接序列编码TCR链的至少两个连续氨基酸。在一些实施方案中,TCR链的一部分的TCR链与由编码至少十个氨基酸的序列编码的TCR链相同。在一些实施方案中,多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列。在一些实施方案中,多种核酸分子中的每种核酸分子编码TCR链的CDR3或其部分。在一些实施方案中,多种核酸分子中的每种核酸分子还包含

TCR链的J区。在一些实施方案中,多种核酸分子中的每种核酸分子编码第一TCR链的第一CDR3或其部分和第二TCR链的第二CDR3或其部分。在一些实施方案中,多种核酸分子中的每种核酸分子还包含第一TCR链的第一J区和第二TCR链的第二J区。在一些实施方案中,编码第一CDR3或其部分的序列和编码第二CDR3或其部分的序列相隔至多100、90、80、70、60、50、40、30、20、10或5个核苷酸。在一些实施方案中,第一连接序列或第二连接序列包含衍生自TCR V基因的序列。在一些实施方案中,第一连接序列或第二连接序列包含预定序列。在一些实施方案中,第一连接序列包含至少一个不同于第二连接序列的核苷酸的核苷酸。在一些实施方案中,第一连接序列与第二连接序列编码相同的氨基酸序列。在一些实施方案中,第一连接序列与第二连接序列编码不同的氨基酸序列。

[0023] 在另一方面,本公开提供了一种用于产生多种核酸分子的方法,所述多种中的每种核酸分子编码T细胞受体(TCR)链或其区域,所述方法包括:使第一多种核酸分子与第二多种核酸分子接触以产生第三多种核酸分子,所述第三多种核酸分子包含至少两种(例如,至少约5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000种或更多种)不同的核酸分子,其中所述至少两种不同的核酸分子中的每种具有编码不同TCR链或其区域的不同序列,并且其中所述至少两种不同的核酸分子在同一区室中产生。

[0024] 在一些实施方案中,第一多种核酸分子中的每种核酸分子包含编码TCR链的CDR3的序列。在一些实施方案中,第一多种核酸分子中的每种核酸分子包含TCR链的J区。在一些实施方案中,第二多种核酸分子中的每种核酸分子包含衍生自TCR链的TCR V基因的序列。在一些实施方案中,TCR V基因是人TCR V基因。在一些实施方案中,TCR V基因是人TRAV1-1、TRAV1-2、TRAV2、TRAV3、TRAV4、TRAV5、TRAV6、TRAV7、TRAV8-1、TRAV8-2、TRAV8-3、TRAV8-4、TRAV8-6、TRAV9-1、TRAV9-2、TRAV10、TRAV12-1、TRAV12-2、TRAV12-3、TRAV13-1、TRAV13-2、TRAV14、TRAV16、TRAV17、TRAV18、TRAV19、TRAV20、TRAV21、TRAV22、TRAV23、TRAV24、TRAV25、TRAV26-1、TRAV26-2、TRAV27、TRAV29、TRAV30、TRAV34、TRAV35、TRAV36、TRAV38-1、TRAV38-2、TRAV39、TRAV40或TRAV41。在一些实施方案中,TCR V基因是人TRBV2、TRBV3-1、TRBV4-1、TRBV4-2、TRBV4-3、TRBV5-1、TRBV5-4、TRBV5-5、TRBV5-6、TRBV5-8、TRBV6-1、TRBV6-2、TRBV6-3、TRBV6-4、TRBV6-5、TRBV6-6、TRBV6-8、TRBV6-9、TRBV7-2、TRBV7-3、TRBV7-4、TRBV7-6、TRBV7-7、TRBV7-8、TRBV7-9、TRBV9、TRBV10-1、TRBV10-2、TRBV10-3、TRBV11-1、TRBV11-2、TRBV11-3、TRBV12-3、TRBV12-4、TRBV12-5、TRBV13、TRBV14、TRBV15、TRBV16、TRBV18、TRBV19、TRBV20-1、TRBV24-1、TRBV25-1、TRBV27、TRBV28、TRBV29-1或TRBV30。在一些实施方案中,衍生自TCR V基因的序列包含编码FR1、CDR1、FR2、CDR2和FR3的序列。在一些实施方案中,衍生自TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和FR3的序列。在一些实施方案中,TCR链是TCR α 链、TCR β 链、TCR γ 链或TCR δ 链。在一些实施方案中,第一多种核酸分子中的每种核酸分子还包含编码额外TCR链的额外CDR3的额外序列。在一些实施方案中,第一多种核酸分子中的每种核酸分子包含额外TCR链的额外J区。在一些实施方案中,TCR链和额外TCR链是TCR链的同源对。在一些实施方案中,多种核酸分子中的核酸分子编码不同的TCR或其部分。在一些实施方案中,第一多种核酸分子中的给定核酸分子包含连接序列,所述连接序列可用于将第一多种核酸分子中的给定核酸分

子与第二多种核酸分子中的给定核酸分子连接。在一些实施方案中,第一多种核酸分子中的给定核酸分子和第二多种核酸分子中的给定核酸分子编码功能性TCR链或其部分。在一些实施方案中,第二多种核酸分子中的给定核酸分子包含抗连接序列,所述抗连接序列与第一多种核酸分子中的给定核酸分子的连接序列互补。在一些实施方案中,所述方法还包括将第一多种核酸分子中的给定核酸分子与第二多种核酸分子中的给定核酸分子连接。在一些实施方案中,连接包括使第一多种核酸分子中的给定核酸分子与第二多种核酸分子中的给定核酸分子杂交。在一些实施方案中,杂交包括使第一多种核酸分子中的给定核酸分子的连接序列与第二多种核酸分子中的给定核酸分子的抗连接序列杂交。在一些实施方案中,所述方法还包括(i)使用第一多种核酸分子中的给定核酸分子作为模板延伸第二多种核酸分子中的给定核酸分子的游离3'末端,和/或(ii)使用第二多种核酸分子中的给定核酸分子作为模板延伸第一多种核酸分子中的核酸分子的游离3'末端,以产生第三多种核酸分子中的核酸分子。在一些实施方案中,所述方法还包括将第一多种核酸分子中的给定核酸分子与第二多种核酸分子中的给定核酸分子连接。在一些实施方案中,所述方法还包括使第三多种核酸分子中的核酸分子与限制酶接触以产生粘性末端。在一些实施方案中,所述方法还包括使第三多种核酸分子中的核酸分子与额外的核酸分子接触。在一些实施方案中,额外的核酸分子编码TCR链的恒定区或其部分。在一些实施方案中,所述方法还包括将第三多种核酸分子中的核酸分子与额外的核酸分子连接。在一些实施方案中,第三多种核酸分子中的至少五种(例如,在一些情况下,至少约6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、1,000、2,000、3,000、4,000、5,000、10,000、20,000、30,000、40,000种或更多种)不同的核酸分子在同一区室中产生。在一些实施方案中,第三多种核酸分子中的至少十种不同的核酸分子在同一区室中产生。在一些实施方案中,同一区室是孔、管或液滴。

[0025] 在另一方面,本公开提供了一种用于产生多种核酸分子的方法,其包括:(a)提供第一多种核酸分子,其中所述第一多种核酸分子中的核酸分子包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的序列,其中所述第一CDR3和所述第二CDR3来自TCR链的同源对;(b)提供第二多种核酸分子,其中所述第二多种核酸分子中的核酸分子包含衍生自TCR V基因的序列;以及(c)使所述第一多种核酸分子与所述第二多种核酸分子接触,其中所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接以形成包含编码所述第一CDR3和所述第二CDR3的序列以及衍生自所述TCR V基因的序列的线性核酸分子,其中所述编码所述第一CDR3和所述第二CDR3的序列以及所述TCR V基因衍生自TCR链的同源对。在一些实施方案中,第一多种核酸分子包含至少约5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000或更多种不同的序列。在一些实施方案中,第二多种核酸分子包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80或更多种不同的TCR V基因。

[0026] 在另一方面,本公开提供了一种用于产生多种核酸分子的方法,其包括:(a)提供

第一多种核酸分子,其中所述第一多种核酸分子中的核酸分子包含(i)编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的合成序列,和(ii)编码第三T细胞受体(TCR)链的第三CDR3和第四TCR链的第四CDR3的合成序列,其中所述第一CDR3和所述第二CDR3来自TCR链的第一同源对,并且其中所述第三CDR3和所述第四CDR3来自TCR链的第二同源对;(b)提供第二多种核酸分子,其中所述第二多种核酸分子中的核酸分子包含衍生自TCR V基因的序列;以及(c)使所述第一多种核酸分子与所述第二多种核酸分子接触,其中所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接以形成包含编码所述第一CDR3和所述第二CDR3的序列以及衍生自所述TCR V基因的序列的核酸分子,其中所述编码所述第一CDR3和所述第二CDR3的序列以及所述TCR V基因衍生自TCR链的同源对。在一些实施方案中,第一多种核酸分子包含至少约2、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000或更多种不同的序列。在一些实施方案中,第二多种核酸分子包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80或更多种不同的TCR V基因。

[0027] 在另一方面,本公开提供了一种在来自受试者的组织样品中鉴定天然配对的T细胞受体(TCR)的序列的方法,其包括:(a)在从所述受试者获得的含有多个外周T细胞的样品中鉴定一种或多种天然配对的TCR的一种或多种配对序列,其中所述一种或多种配对序列中的每种包含CDR3序列;和(b)在组织样品中鉴定TCR的TCR链的组织CDR3序列,与所述TCR链天然配对的其他TCR链是未知的,其中所述组织CDR3序列与一种或多种天然配对的TCR的一种或多种配对序列中的至少一种配对序列的CDR3序列匹配,从而在组织样品中将所述至少一种配对序列鉴定为天然配对的TCR的序列。在一些实施方案中,(a)中的鉴定包括对含有多个外周T细胞的样品中的一种或多种天然配对的TCR进行测序。在一些实施方案中,测序包括单细胞测序。在一些实施方案中,单细胞测序包括将多个外周T细胞分成多个区室,每个区室包含多个外周T细胞中的单个外周T细胞。在一些实施方案中,组织样品不是体液样品。在一些实施方案中,组织样品是实体瘤样品。在一些实施方案中,组织样品是固定或冷冻样品。在一些实施方案中,含有多个外周T细胞的样品是外周血单核细胞(PBMC)样品。在一些实施方案中,所述方法还包括在(a)之前从受试者获得血液样品。在一些实施方案中,所述方法还包括在(a)之前从血液样品中分离外周血单核细胞。在一些实施方案中,组织样品包含肿瘤浸润性T细胞。

[0028] 在另一方面,本公开提供了一种鉴定靶反应性T细胞受体(TCR)的方法,其包括:(a)提供包含使用本文所述的方法鉴定的TCR的细胞;和(b)使所述细胞与抗原呈递细胞(APC)呈递的靶抗原接触,其中所述细胞通过TCR与APC呈递的靶抗原结合,从而将TCR鉴定为靶反应性TCR。在一些实施方案中,靶抗原是肿瘤抗原(例如肿瘤相关抗原或肿瘤特异性抗原)。在一些实施方案中,所述方法还包括将编码靶反应性TCR的序列递送到宿主细胞中。在一些实施方案中,所述方法还包括将宿主细胞施用于受试者。在一些实施方案中,宿主细胞是T细胞。在一些实施方案中,T细胞是自体T细胞。在一些实施方案中,T细胞是同种异体T细胞。在一些实施方案中,细胞是报告细胞系,所述报告细胞系包含在细胞与APC呈递的靶

抗原结合时表达的报告基因。

[0029] 本公开的另外的方面和优点从以下具体实施方式变得为本领域技术人员显而易见,其中仅示出并描述本公开的例示性实施方案。应当认识到的是,本公开能够具有其他以及不同的实施方案,并且其若干细节能够在各种不同方面做出修改,所有均不脱离公开内容。因此,附图和说明书应被视为在本质上是说明性的而不是限制性的。

[0030] 以引用的方式并入

[0031] 在本说明书中提及的所有公布、专利和专利申请均以引用的方式并入本文,其程度就如同每个单独的公布、专利或专利申请被具体和单独地指出以引用的方式并入一般。如果通过引用并入的出版物和专利或专利申请与说明书中包含的公开内容相矛盾,则说明书旨在取代和/或优先于任何这样的矛盾材料。

附图说明

[0032] 本发明的新颖特征在所附权利要求中具体阐述。通过参考对其中利用本发明的原理的例示性实施方案作出阐述的以下详细描述以及附图(在本文中还称为“图(Figure、Fig.和FIGURE)”)获得对本发明的特征和优点的更好理解,在附图中:

[0033] 图1A-图1C描绘了产生编码T细胞受体的核酸构建体的示例性方案。

[0034] 图2A描绘了使用本文所述的方法的示例性模拟结果。

[0035] 图2B描绘了使用本文所述的方法的示例性模拟结果。

[0036] 图3A描绘了使用本文所述的方法的示例性模拟结果。

[0037] 图3B描绘了使用本文所述的方法的示例性模拟结果。

[0038] 图4A描绘了TCR V基因的种系基因组DNA的示意图。

[0039] 图4B描绘了TCR V-J基因的重排基因组DNA的示意图。

[0040] 图4C描绘了TCR V-D-J基因的重排基因组DNA的示意图。

[0041] 图5描绘了与将CDR3-J多核苷酸与正确的V基因种系多核苷酸连接相关联的潜在挑战的方案。虚线箭头描绘了在CDR3-J多核苷酸与不正确的V基因种系多核苷酸之间可能发生连接。

[0042] 图6描绘了通过重叠引物延伸将CDR3-J多核苷酸(连接到白色框的灰色框)与指定的、预合成的V基因种系多核苷酸(连接到细箭头所指的灰色框的黑色框)连接的方案。顶部粗箭头(603)描绘了预合成的V基因种系多核苷酸上的连接序列(601)与CDR3-J多核苷酸上的连接序列(602)之间的杂交。底部粗箭头(604)描绘了引物延伸。601可以被称为连接序列,并且602可以被称为抗连接序列(或反之亦然)。

[0043] 图7描绘了使用任意连接序列(701)和抗连接序列(702)将CDR3-J多核苷酸与指定的V基因种系多核苷酸连接。

[0044] 图8描绘了TCR基因自组装的一般原理。801:预合成的V基因种系多核苷酸。802:包含CDR3-J序列的多核苷酸(例如,CDR3-J多核苷酸)。803:包含V基因种系多核苷酸序列和CDR3-J序列的核酸序列。X是各自为不同V基因种系多核苷酸的一部分的多核苷酸的数目。Y是CDR3-J多核苷酸的数目。Y可以远大于X。箭头表示批量反应,其中每个CDR3-J多核苷酸与指定的、预合成的V基因种系多核苷酸连接。

[0045] 图9A描绘了使用血液样品来鉴定肿瘤样品中的肿瘤浸润性TCR的示例工作流程。

- [0046] 图9B描绘了使用图9A所示方法鉴定的TCR的示例性应用。
- [0047] 图9C描绘了使用图9A所示方法鉴定的TCR的示例性应用。
- [0048] 图10A描绘了使用本文所述的方法的示例性模拟结果。
- [0049] 图10B描绘了使用本文所述的方法的示例性模拟结果。
- [0050] 图11A描绘了使用本文所述的方法的示例性模拟结果。
- [0051] 图11B描绘了使用本文所述的方法的示例性模拟结果。
- [0052] 图12描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。
- [0053] 图13描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。
- [0054] 图14描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。
- [0055] 图15描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。
- [0056] 图16描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。
- [0057] 图17描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。
- [0058] 图18描绘了评估本文所述的基因组组装方法的示例性下一代测序数据。

具体实施方式

[0059] 在本公开中,除非另外确切说明,否则单数的使用包括复数。同样地,除非另外说明,否则“或”的使用意指“和/或”。类似地,“包含(comprise)”、“包含(comprises)”、“包含(comprising)”、“包括(include)”、“包括(includes)”和“包括(including)”并不旨在进行限制。

[0060] 术语“约”或“近似”意指由本领域普通技术人员测定的具体值处于可接受的误差范围内,这将部分取决于所述值的测量或测定方式,即,测量系统的限制性。例如,根据本领域的实践,“约”可意指在1个或多于1个标准偏差内。或者,“约”可意指给定值的至多20%、至多10%、至多5%或至多1%的范围。或者,特别是关于生物系统或过程,所述术语可意指在值的一个数量级以内,例如在5倍以内、或在2倍以内。当特定值在申请和权利要求书中描述时,除非另外说明,否则应假定术语“约”意指在特定值的可接受误差范围内。

[0061] 当术语“至少”、“大于”或“大于或等于”在一系列两个或更多个数值中的第一个数值之前时,术语“至少”、“大于”或“大于或等于”适用于所述系列数值中的每个数值。例如,大于或等于1、2或3等同于大于或等于1、大于或等于2、或大于或等于3。

[0062] 当术语“不大于”、“小于”或“小于或等于”在一系列两个或更多个数值中的第一个数值之前时,术语“不大于”、“小于”或“小于或等于”适用于所述系列数值中的每个数值。例如,小于或等于3、2或1等同于小于或等于3、小于或等于2、或小于或等于1。

[0063] 术语“多核苷酸”、“核酸”和“寡核苷酸”在本公开中可互换使用。它们可以指各种长度的核苷酸的聚合形式。它们可以包含脱氧核糖核苷酸和/或核糖核苷酸或其类似物。多核苷酸可以包括选自腺苷(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U)或其变体的一个或多个核苷酸。核苷酸可包括核苷和至少1、2、3、4、5、6、7、8、9、10个或更多个磷酸(PO_3)基团。核苷酸可包括核碱基、五碳糖(核糖或脱氧核糖)和一个或多个磷酸基团。多核苷酸可以具有任何三维结构并且可以执行各种功能。多核苷酸可具有各种构型,诸如线性、环状、茎-环和分支。以下是多核苷酸的非限制性实例:基因或基因片段的编码或非编码区、由连锁分析确定的基因座、外显子、内含子、信使RNA(mRNA)、转运RNA(tRNA)、核糖体RNA

(rRNA)、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微小RNA(miRNA)、环状RNA、核糖酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、核酸探针以及引物。多核苷酸可以包含一个或多个修饰的核苷酸,诸如甲基化核苷酸和核苷酸类似物。如果存在,那么可在聚合物组装之前或之后赋予对核苷酸结构的修饰。核苷酸的序列可被非核苷酸组分中断。可以在聚合之后,诸如通过与标记组分缀合来进一步修饰多核苷酸。多核苷酸可以包括一个或多个核苷酸变体,包括非标准核苷酸、非天然核苷酸、核苷酸类似物和/或修饰的核苷酸。

[0064] 如本文所用,术语“序列”是指核酸分子中核苷酸的顺序,或肽的氨基酸残基的顺序。核酸序列可以是脱氧核糖核酸(DNA)序列或核糖核酸(RNA)序列;可以是线性的、环状的或分支的;并且可以是单链的或双链的。可以使序列突变,使得其不同于参考序列(例如,野生型序列)。序列可以具有任何长度,例如长度在2与1,000,000个或更多个氨基酸或核苷酸之间(或其间或其上的任何整数值),例如约100与约10,000个核苷酸之间或约200与约500个氨基酸或核苷酸之间。在一些情况下,给定核酸序列可涵盖给定核酸序列的序列信息和给定核酸序列的反向互补序列。在一些情况下,DNA序列可涵盖由DNA转录的对应RNA序列的序列信息。序列可以是多核苷酸或多肽分子的字母表示。序列可以是可由计算机处理器使用的一条信息。在一些情况下,核酸序列可用于指物理核酸分子本身。

[0065] 如本文所用,术语“平末端”是指双链核酸分子的末端,其中核酸分子的一条链末端的基本上所有的核苷酸与同一核酸分子的另一条链中的相对核苷酸碱基配对。如果核酸分子的末端包括长度为至少一个核苷酸的单链部分,则所述核酸分子不是平末端的,在本文中称为“突出端”或“粘性末端”。

[0066] 如本文所用,术语“TCR V基因”是指种系构型的T细胞受体可变(V)基因的基因组核酸序列,其包含编码前导肽的第一部分的序列(例如IMGT中定义的L-PART1)、内含子(例如IMGT中定义的V-INTRON)和外显子(例如IMGT中定义的V-EXON),具有5'UTR和3'UTR(包括重组信号序列)。重组信号序列可包含由间隔元件(例如由IMGT定义的V-SPACER)分隔的七聚体(例如IMGT中定义的V-HEPTAMER)和九聚体(例如由IMGT定义的V-NONAMER)。V-EXON涵盖编码前导肽的第二部分(L-PART2)和V-REGION的序列。TCR V基因的实例包括TCR α 可变(TRAV)基因、TCR β 可变(TRBV)基因、TCR γ 可变(TRGV)基因和TCR δ 可变(TRDV)基因。本文所述的核酸可包含衍生自TCR V基因的序列。“衍生自”意指与参考序列具有至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%或100%序列同一性的序列。衍生自TCR V基因的序列可以是如上所述的TCR V基因的基因组核酸序列的全长序列。衍生自TCR V基因的序列可以是TCR V基因的一部分,其包含TCR V基因的至少10、20、30、40、50、60、70、80、90、100、200、300、400、500个或更多个核苷酸。衍生自TCR V基因的序列可以是密码子优化的(或密码子多样化的)核酸序列。给定核酸序列的密码子优化的序列是指修饰的核酸序列,其蛋白编码区与给定核酸的蛋白编码区编码相同的氨基酸序列。修饰的核酸序列可具有不同于给定核酸序列的序列或可衍生自给定核酸。可实施密码子优化以移除限制性位点、移除多核苷酸序列中不需要的二级结构、促进CDR3-J多核苷酸与TCR V基因的指定预合成部分的正确连接,或用于其他目的。密码子优化或密码子多样化可通过改变给定核酸序列的一个或多个核苷酸来实现。例如,可以通过计算方法实现密码子优化或密码子多样化。密码子优化和密码子多样化在本公开中可互换使用。

[0067] 如本文所用,术语“V-REGION”是指种系基因组DNA或cDNA中的TCR V基因(包括V-HEPTAMER之前的1或2个核苷酸,如果存在的话)的编码区,或通常在重排的基因组DNA或cDNA中通过V-(D)-J重排在3'修剪的可变(V)区。

[0068] 如本文所用,术语“D-REGION”是指种系基因组DNA或cDNA中的TCR D基因(包括5'D-HEPTAMER之后和/或3'D-HEPTAMER之前的1或2个核苷酸,如果存在的话)的编码区,或通常在部分重排或重排的基因组DNA或cDNA中通过D-J或V-D-J重排在5'和/或3'修剪的多样性(D)区。

[0069] 如本文所用,术语“J-REGION”是指种系基因组DNA或cDNA中的TCR J基因(包括J-HEPTAMER之后的1或2个核苷酸,如果存在的话)的编码区,或通常在重排的基因组DNA或cDNA中通过V-(D)-J重排在5'修剪的连接(J)区。

[0070] 如本文所用,术语“V-J-REGION”是指在重排的基因组DNA或cDNA中包含V-REGION和J-REGION的TCR链的编码区。

[0071] 如本文所用,术语“V-D-J-REGION”是指在重排的基因组DNA或cDNA中包含V-REGION、D-REGION和J-REGION的TCR链的编码区。

[0072] 术语“连接(link)”或“连接(connect)”在本公开中可互换使用。它们是指物理连接两个或更多个核酸分子。两个或更多个核酸分子可以连接,使得所述两个或更多个核酸分子形成连续的核酸分子。两个或更多个核酸分子可以共价连接或非共价连接。连接可以通过多种方式实现,包括形成氢键、离子键和共价键或范德华力。

[0073] 相对于参考核酸序列(或肽序列)的序列同一性百分比(%)是在对齐序列并引入缺口(如果需要)以实现序列同一性最大百分比,并且不将任何保守取代视为序列同一性的一部分之后,候选序列中与参考核酸序列(或肽序列)中的核苷酸(或氨基酸残基)相同的核苷酸(或在肽序列的情况下为氨基酸残基)的百分比。用于确定序列同一性百分比的对齐可以以本领域技术范围内的各种方式实现,例如,使用可公开获得的计算机软件诸如BLAST、BLAST-2、CLUSTALW、ALIGN或Megalign(DNASTAR)软件。本领域技术人员可确定用于对齐序列的适当参数,包括在所比较的序列的全长范围内实现最大对齐所需的任何算法。

[0074] 应用于核酸或氨基酸序列的术语“基本上相同”及其语法上的等同物是指核酸或氨基酸序列包含使用上述程序(例如BLAST)使用标准参数与参考序列相比具有至少90%或更多、至少95%、至少98%或至少99%序列同一性的序列。例如,BLASTN程序(对于核苷酸序列)使用的默认值是词长(W) 11,期望值(E) 10,M=5,N=-4和两条链的比较。对于氨基酸序列,BLASTP程序使用的默认值是字长(W) 3,期望值(E) 10和BLOSUM62评分矩阵(参见Henikoff和Henikoff,Proc.Natl.Acad.Sci.USA 89:10915(1992))。

[0075] 概述

[0076] 可使用高通量、配对测序来测序T细胞受体(TCR)。例如,随着单细胞技术的发展,可将单个T细胞分配到隔离的区室中,其中来自同一T细胞的TCR α 和 β 链mRNA可附接到相同的独特条形码。这些系统中的一些是可商购获得的(例如,通过10X Genomics)。记录T细胞受体 α 可变(TRAV)基因同一性、CDR3 α 序列、T细胞受体 α 连接(TRAJ)基因同一性、T细胞受体 β 可变(TRBV)基因同一性、CDR3 β 序列和T细胞受体 β 连接(TRBJ)基因同一性的配对序列信息可以允许重建全长、可表达的TCR。然而,以可引入细胞用于功能研究或筛选的DNA或RNA形式合成此类TCR序列的技术可能是低通量的。本公开提供了多种方法和组合物,其可允许超

高通量构建编码TCR序列(例如,在一些情况下,配对的、全长的、可表达的TCR序列)的多核苷酸。

[0077] T细胞受体(TCR)

[0078] TCR可用于赋予T细胞识别与各种癌症或感染性生物体相关联的抗原的能力。TCR由两条链组成,例如 α 链和 β 链或 γ 链和 δ 链。构成这些链的蛋白质由DNA编码,所述DNA利用独特的机制产生TCR的巨大多样性。这种多亚基免疫识别受体与CD3复合物缔合并结合抗原呈递细胞(APC)表面上的MHC I类和II类蛋白呈递的肽。TCR与APC上的抗原肽的结合可以是T细胞激活的中心事件,所述结合发生在T细胞与APC接触点的免疫突触处。

[0079] 在MHC I类分子的背景下,TCR可识别T细胞表位。MHC I类蛋白可在高等脊椎动物的所有有核细胞中表达。MHC I类分子是由与12-kDa轻链 β -2微球蛋白非共价地缔合的46-kDa重链组成的异二聚体。在人类中,存在若干种MHC等位基因,例如HLA-A2、HLA-A1、HLA-A3、HLA-A24、HLA-A28、HLA-A31、HLA-A33、HLA-A34、HLA-B7、HLA-B45和HLA-Cw8。在一些实施方案中,MHC I类等位基因是HLA-A2等位基因,其在一些群体中由近似50%的群体表达。在一些实施方案中,HLA-A2等位基因可以是HLA-A*0201、*0202、*0203、*0206或*0207基因产物。在一些情况下,不同群体之间的亚型频率可能存在差异。例如,在一些实施方案中,超过95%的HLA-A2阳性白种人群体是HLA-A*0201,而在中国人群体中,据报道频率为近似23%HLA-A*0201、45%HLA-A*0207、8%HLA-A*0206和23%HLA-A*0203。

[0080] 在一些实施方案中,在MHC II类分子的背景下,TCR可识别T细胞表位。MHC II类蛋白可以在APC的子集中表达。在人类中,存在若干种MHC II类等位基因,例如像DR1、DR3、DR4、DR7、DR52、DQ1、DQ2、DQ4、DQ8和DPI。在一些实施方案中,MHC II类等位基因是HLA-DRB1*0101、HLA-DRB*0301、HLA-DRB*0701、HLA-DRB*0401或HLA-DQB1*0201基因产物。

[0081] TCR链可包含可变结构域(或可变区)和恒定结构域(或恒定区)。可变结构域可以是由IMGT独特编号系统定义的V-DOMAIN。可变结构域可对应于TCR链的V-J-REGION或V-D-J-REGION。恒定结构域可以是由IMGT独特编号系统定义的C-DOMAIN。在一些情况下,恒定结构域可以是恒定区的一部分。例如,全长恒定区可包含恒定结构域(胞外区)、连接区、跨膜区和胞质区。

[0082] TCR α 或TCR δ 链的可变结构域可以由种系中的多个可变(V)和连接(J)基因区段编码,而TCR β 或TCR γ 链的可变结构域另外由多样性(D)基因区段编码。每个基因区段的侧翼可以是重组信号序列。重组信号可包含由间隔元件分隔的七聚体和九聚体。间隔元件的长度可以是12或23bp。在V(D)J重组期间,每个基因区段的一个随机等位基因与其他等位基因重组以形成功能性可变结构域。可变结构域与恒定(C)基因区段的重组可产生功能性TCR链转录物。另外,可以在基因区段之间的连接位点处添加和/或缺失随机核苷酸。此过程可导致强组合(取决于哪些基因区域将重组)和连接多样性(取决于添加/缺失哪些和多少核苷酸),从而产生大量且高度可变的TCR库,这可确保识别出大量抗原。额外的多样性可以通过将 α 和 β 或 γ 和 δ 链配对(也称为“组装”)以形成功能性TCR来实现。通过重组、随机插入、缺失和取代,编码T细胞受体的一小组基因具有产生 10^{15} 与 10^{20} 个TCR克隆型的潜力。如本文所用,“克隆型”是指携带相同免疫受体的免疫细胞群体。例如,克隆型是指携带相同TCR的T细胞群体,或携带相同BCR(或抗体)的B细胞群体。在免疫受体多样性的背景下,“多样性”是指群体中免疫受体(例如TCR、BCR和抗体)克隆型的数目。如本文所用,“同源对组合”是指来自

T细胞的TCR的两条链(例如TCR α 和TCR β ,或TCR γ 和TCR δ)的天然组合。两条链的相同同源对组合可产生相同的TCR。例如,具有相同克隆型的T细胞具有TCR α 和TCR β 链的相同同源对组合。克隆型中的较高多样性可指示同源对组合中的较高多样性。

[0083] 每条TCR链可在其结构中含有三个高变环,称为互补决定区(CDR1-3)。CDR1和CDR2可由V基因编码,并且可能是TCR与MHC复合物相互作用所需的。然而,CDR3部分地由(1)V与J基因之间的连接区(在TCR α 或TCR γ 的情况下),或(2)V与D基因之间的连接区以及D与J基因之间的连接区(在TCR β 或TCR δ 的情况下)编码,并且因此可以是高度可变的。CDR3可以是与肽抗原直接接触的TCR的区域。CDR3可用作确定T细胞克隆型的目标区域。一个个体的T细胞的所有TCR的总和被称为TCR库或TCR谱。TCR库可以随着疾病的发作和进展而变化。因此,确定不同疾病状态(诸如癌症、自身免疫性疾病、炎性疾病和感染性疾病)下的免疫库状态,可用于疾病诊断和预后。

[0084] TCR应理解为涵盖全长TCR及其抗原结合部分或抗原结合片段(也称为MHC-肽结合片段)。在一些实施方案中,TCR是完整或全长TCR。在一些实施方案中,TCR是抗原结合部分,其小于全长TCR,但与结合MHC分子(例如MHC-肽复合物)的特异性抗原肽结合。在一些情况下,TCR的抗原结合部分或片段可仅含有全长或完整TCR的结构域的一部分,但仍能够结合完整TCR所结合的表位(例如MHC-肽复合物)。在一些情况下,TCR的抗原结合部分或片段含有TCR的可变结构域,诸如TCR的可变 α 链和可变 β 链,其足以形成用于与特定MHC-肽复合物结合的结合位点,诸如通常在每条链含有三个互补决定区的情况下。包括具有作为抗原结合结构域或与抗原结合结构域同源的结合结构域的多肽或蛋白质。

[0085] TCR分子可由 α 链(α 链或TCR α 链,由TRA基因/序列编码)和 β 链(β 链或TCR β 链,由TRB基因/序列编码)、或 γ 链(γ 链或TCR γ 链,由TRG基因/序列编码)和 δ 链(δ 链或TCR δ 链,由TRD基因/序列编码)形成。这些免疫受体链可以具有可变结构域(例如,由重排的VDJ或VJ区编码)。可变结构域的部分可以是高变的。高变区可以包括互补决定区(CDR),例如CDR1、CDR2和CDR3。在一些情况下,在一个T细胞内,可以仅表达一种功能性 α 链序列和一种功能性 β 链序列。在一些情况下,在一个T细胞内,可以仅表达一种功能性 γ 链序列和一种功能性 δ 链序列。

[0086] 基于芯片的寡核苷酸合成:机遇与挑战

[0087] 尽管基于芯片的高通量寡核苷酸合成技术可能已经发展到可以一次合成数十万或甚至数百万具有任意序列的寡核苷酸的长度,但是以这种方式合成的寡核苷酸的长度可以限制为约200至300个碱基长。相比之下,全长TCR构建体的长度可以为近两千个碱基。乍一看,基于芯片的合成似乎不足以解决TCR基因合成问题。然而,对TCR结构的检查可揭示机会。首先,TCR α 链和 β 链的恒定区(例如,TRAC和TRBC)可以是恒定的。因此,编码TCR链恒定区的多核苷酸序列可附加到TCR序列的其余部分。其次,与BCR/抗体序列不同,TCR可能不经历体细胞超突变,这意味着CDR3区外的序列可以是种系来源的。因此,可预合成各自包含衍生自TCR V基因或其部分的序列的多核苷酸。衍生自TCR V基因的序列可以是TCR V基因的一部分。衍生自TCR V基因的序列可以是密码子优化的序列或包含一个或多个修饰的核苷酸。例如,可预合成包括L-PART1(前导肽的第一部分)、L-PART2(前导肽的第二部分)、FR1、CDR1、FR2、CDR2和FR3的编码序列的衍生自TCR V基因的序列(称为L-V-REGION)。又如,可预合成包括FR1、CDR1、FR2、CDR2和FR3的编码序列的衍生自TCR V基因的序列(称为V-

REGION)。L-PART1、L-PART2、FR1、CDR1、FR2、CDR2或FR3的核酸序列区段可根据IMGT独特编号系统(<http://www.imgt.org>)来定义。在一些情况下,衍生自TCR V基因的序列可包含始于编码L-PART1的序列且终止于编码第二保守半胱氨酸的密码子(例如,由IMGT定义的第2CYS,对应于V-DOMAIN的位置104处的保守半胱氨酸的密码子)的序列。由于在人类基因组中存在约80个或更多个TCR V基因(例如TRAV和TRBV基因),因此合成这种“V基因种系多核苷酸文库”(如图8,801和括号X所示)是可行的。在一些情况下,合成物种(例如人或小鼠)的TCR V基因的子集以产生“V基因种系多核苷酸文库”。可以合成所有鉴定的TCR V基因或TCR V基因的子集。例如,可合成物种的至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80个或更多个TCR V基因以产生文库。在一些情况下,合成物种的所有鉴定的TCR V基因以产生文库。TCR V基因可以是TRAV、TRBV、TRGV或TRDV。如本文所述,在一些情况下,“V基因种系多核苷酸”是指TCR V基因的基因组或密码子优化的多核苷酸的一部分。衍生自TCR V基因的序列可以是V基因种系多核苷酸。FR3与恒定区之间的序列(例如,CDR3加上J区的剩余部分,本文称为“CDR3-J”)的长度可以为至少约10、20、30、40、50、60、70、80、90个或更多个核苷酸,或在一些情况下,长度可以为至多约90个核苷酸。TCR的 α 链和 β 链的CDR3-J序列的长度可以未至少约50、60、70、80、90、100、120、150、180个或更多个核苷酸。TCR的 α 链和 β 链的CDR3-J序列(在一些情况下,总长度为至多约180个核苷酸)可以包括在寡核苷酸(称为“配对的CDR3-J寡聚物”、“配对的CDR3-J寡核苷酸”或“配对的CDR3-J多核苷酸”,其可以互换使用)中,所述寡核苷酸可以进行基于芯片的合成(如图8,802和涵盖802的括号Y所示)。在一些情况下,配对的CDR3-J多核苷酸可包含TCR γ 链的CDR3-J序列和TCR δ 链的CDR3-J多核苷酸。如本文所用,术语“CDR3-J多核苷酸”、“CDR3-J寡核苷酸”和“CDR3-J寡聚物”(其可互换使用)是指包含一个或多个CDR3-J序列的多核苷酸序列。CDR3-J多核苷酸可以是配对的CDR3-J多核苷酸(例如,包含来自配对的TCR链的CDR3-J序列)。CDR3-J多核苷酸(例如,非配对的)可仅含有来自配对的TCR链之一的CDR3-J序列。例如,CDR3-J多核苷酸可仅含有来自TCR α 链、TCR β 链、TCR γ 链或TCR δ 链的CDR3-J序列。剩余的挑战可以是将这种配对的CDR3-J寡核苷酸以高通量(例如,在一批中构建>1,000个TCR)转化为可表达的TCR构建体。使用本文所述的方法,配对的CDR3-J寡核苷酸可以在批量反应中与其对应的V基因种系多核苷酸连接(例如,图8,803)。在一些情况下,CDR3-J多核苷酸库(例如配对或非配对)可包含至少约2、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000个或更多个不同序列。V基因种系多核苷酸文库可包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80个或更多个TCR V基因。当使用本文所述的方法时,可产生多个至少2、5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000个或更多个编码不同天然配对的TCR的不同序列。天然配

对的TCR可以在单个区室中批量产生。

[0088] 包含TCR V基因序列的核酸分子的种系或重排基因构建体的实例在图4A-图4C中示出。例如,图4A示出了TCR V基因的种系基因组DNA,其包含L-PART1、V-INTRON、V-EXON和重组信号序列(V-HEPTAMER、V-SPACER和V-NONAMER)。图4A中也示出了两个保守的半胱氨酸。在V-(D)-J重组后,重排的基因组DNA的示例性构建体在图4B或图4C中示出。CDR3可由(i)V-REGION与J-REGION之间的连接(或连接区)或(ii)V-REGION与D-REGION之间的连接以及D-REGION与J-REGION之间的连接编码。

[0089] TCR V基因可以非常多样。在人类中,已经鉴定了超过40种TRA的功能性V基因,包括例如TRAV1-1、TRAV1-2、TRAV2、TRAV3、TRAV4、TRAV5、TRAV6、TRAV7、TRAV8-1、TRAV8-2、TRAV8-3、TRAV8-4、TRAV8-6、TRAV9-1、TRAV9-2、TRAV10、TRAV12-1、TRAV12-2、TRAV12-3、TRAV13-1、TRAV13-2、TRAV14、TRAV16、TRAV17、TRAV18、TRAV19、TRAV20、TRAV21、TRAV22、TRAV23、TRAV24、TRAV25、TRAV26-1、TRAV26-2、TRAV27、TRAV29、TRAV30、TRAV34、TRAV35、TRAV36、TRAV38-1、TRAV38-2、TRAV39、TRAV40和TRAV41。在这些V基因中,它们中的一些可以被分类为相同的亚组,并且它们由紧随“TRAV”之后的相同亚组编号表示,但在“-”符号后用不同编号表示。例如,TRAV1-1和TRAV1-2来自同一亚组。如本文所用,“组”是共享相同“基因型”(例如,V、D、J或C型)并且潜在地参与相同“链型”的多肽的合成的一组基因。通过延伸,组包括相关的假基因和孤独基因(orphan)。“亚组”意指在给定物种中属于同一组,并且在核苷酸水平上(在V、D和J的种系构型中)共享至少75%同一性的一组基因。

[0090] 在人类中,已经鉴定了超过40种TRB的功能性V基因,包括,例如,TRBV2、TRBV3-1、TRBV4-1、TRBV4-2、TRBV4-3、TRBV5-1、TRBV5-4、TRBV5-5、TRBV5-6、TRBV5-8、TRBV6-1、TRBV6-2、TRBV6-3、TRBV6-4、TRBV6-5、TRBV6-6、TRBV6-8、TRBV6-9、TRBV7-2、TRBV7-3、TRBV7-4、TRBV7-6、TRBV7-7、TRBV7-8、TRBV7-9、TRBV9、TRBV10-1、TRBV10-2、TRBV10-3、TRBV11-1、TRBV11-2、TRBV11-3、TRBV12-3、TRBV12-4、TRBV12-5、TRBV13、TRBV14、TRBV15、TRBV16、TRBV18、TRBV19、TRBV20-1、TRBV24-1、TRBV25-1、TRBV27、TRBV28、TRBV29-1和TRBV30。其他物种(例如小鼠)的V基因可以在IMGT数据库中找到。

[0091] 使连接序列多样化

[0092] 将V基因种系多核苷酸与CDR3-J多核苷酸连接可以通过分子生物学技术(诸如连接和重叠引物延伸)来实现(图6)。然而,为了充分利用基于芯片的寡核苷酸合成的能力,可以在批量反应中将数千或更多个CDR3-J寡核苷酸与其对应的V基因种系多核苷酸连接(如图8的箭头,803所示)。这样做的主要挑战可以是V基因种系多核苷酸(图6,601)与CDR3-J(图6,602)之间的连接区可以是保守的FR3区。因此,在批量反应中,可能难以控制哪个V基因种系多核苷酸与哪个CDR3-J连接(如图5所示,其中实线箭头表示连接到正确的V基因种系多核苷酸,而虚线箭头表示连接到不正确的V基因种系多核苷酸)。例如,TCR序列可由与特定CDR3-J β 序列连接的TRBV4-1形成。在批量反应中,可存在TRBV4-1和TRBV4-2的V基因种系多核苷酸,并且这些TRBV基因的FR3区可高度相似。因此,此TCR的CDR3-J寡核苷酸可能与TRBV4-2种系多核苷酸不正确连接。为了缓解这个问题,可以使用密码子多样化在不同FR3序列之间产生差异。例如,连接序列可以是密码子多样化的,使得它们可以具有不同的核酸序列,即使它们可以编码相同的氨基酸序列。密码子多样化可通过计算方法(诸如实施例2中所示的方法)实现。多个核酸序列可通过随机或根据任意规则将密码子分配给氨基酸而

产生,其中每个核酸序列可编码相同的氨基酸序列。接着,可根据任意规则通过计算评价多个核酸序列以分配分数。任意规则可考虑诸如限制性位点、与不需要的序列杂交的倾向、与给定序列杂交的倾向或序列中不需要的二级结构的因素。接着,根据分数,可以从所述多个核酸序列中选择核酸序列作为密码子多样化的连接序列。密码子多样化的连接序列可用于实现CDR3-J多核苷酸与TCR V基因的指定预合成部分的正确连接。在包含一些或全部已知TCR V基因的“V基因种系多核苷酸文库”中,例如,每个不同的TCR V基因可具有不同的连接序列,其可用于正确地与对应的CDR3-J寡核苷酸连接以根据参考序列形成TCR链。参考序列可通过对TCR链的同源对进行测序来产生。然而,在一些情况下,可能不清楚连接序列可以多样化的程度,以及V基因种系多核苷酸与CDR3-J寡核苷酸之间的连接在批量反应中可以正确的程度。如实施例2所示,可以使入TCR V基因的FR3区多样化,使得实际上无法检测到任何给定CDR3-J序列的“错误连接概率”。实施例2中列出的算法可用于产生“密码子多样化的V基因种系多核苷酸”及其对应的CDR3-J序列。

[0093] 一旦发现一组多样的连接序列,使用分子生物学技术(例如连接、限制性消化、环化)的许多方法可用于将CDR3-J寡核苷酸库转化为全长、可表达的TCR库。实施例1提供了示例性工作流程。本文所提供的方法也可用于在批量反应中产生单个TCR链(例如非配对的链)的库。例如,为了产生TCR α 链的库,每个单独的CDR3-J寡核苷酸可包含来自TCR α 链的CDR3和J区,但可不包含来自TCR β 链的另一CDR3和J区,然后CDR3-J寡核苷酸可用于与对应的TRAV基因连接以形成TCR α 链。

[0094] 用于构建编码TCR的核酸分子的方法

[0095] 编码本文所述TCR的核酸分子可由两个或更多个核酸片段构建。在一些实施方案中,两个或更多个核酸片段可称为第一核酸分子、第二核酸分子、第三核酸分子、第四核酸分子等。当构建核酸分子时,可使用标准分子生物学技术,包括但不限于杂交、延伸、连接和酶消化/切割。

[0096] 本文所述的核酸片段可编码TCR链或其部分。例如,由核酸片段编码的TCR链的部分可包含大于或等于约10、15、20、25、30、35、40、45、50、60、70、80、90、100、110、120、130、140、150、200、250个或更多个氨基酸。核酸片段可包含编码功能性TCR链的序列。功能性TCR链可以是或不是全长TCR链。功能性TCR链可包含一或多个突变或修饰。在一些情况下,功能性TCR链在宿主细胞中表达时可并入TCR复合物(例如,具有TCR α 、TCR β 、CD3 γ 、CD3 δ 、CD3 ϵ 和 ζ 链的复合物)中。在一些情况下,功能性TCR可与其靶配体结合。在一些情况下,功能性TCR在宿主细胞中表达时可并入细胞膜中。在一些情况下,功能性TCR可在宿主细胞中表达。

[0097] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含编码CDR3的序列。

[0098] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含编码第一TCR链的第一CDR3和第二TCR链的第二CDR3的序列,其中第一CDR3和第二CDR3衍生自TCR链的同源对。在一些实施方案中,编码第一CDR3的序列和编码第二CDR3的序列相隔至多约100、90、80、70、60、50、40、30、20、10或5个核苷酸。

[0099] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含TCR V基因序列或其部分。用于构建编码TCR或其部分的核酸分子的核酸片段可包含衍生自TCR V基因序列的序列。衍生自TCR V基因的序列可包含V-REGION核酸序列。衍生自TCR V基因的序列可包含编码FR1、CDR1、FR2、CDR2和/或FR3核酸序列的序列。衍生自TCR V基因的序列可包含编码前导

肽的序列。衍生自TCR V基因的序列可包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和/或FR3核酸序列的序列。衍生自TCR V基因的序列可包含或可以是TCR V基因的一部分。TCR V基因的部分的长度可以为至少10个核苷酸。例如,TCR V基因的部分的长度可大于或等于约10、20、30、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000个或更多个核苷酸。衍生自TCR V基因的序列可包含一个或多个修饰的核苷酸。衍生自TCR V基因的序列可以是密码子优化的(或密码子多样化的),使得它具有与TCR V基因或其部分不同的序列,但它可以编码相同的氨基酸序列。衍生自TCR V基因的序列可不包含编码CDR3的一部分的序列。衍生自TCR V基因的序列可不包含重排基因的连接区的序列。

[0100] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含编码恒定结构域或其部分的序列。用于构建编码TCR或其部分的核酸分子的核酸片段可包含编码恒定区或其部分的序列。在一些情况下,恒定结构域或恒定区是TCR α 恒定结构域或恒定区、TCR β 恒定结构域或恒定区、TCR γ 恒定结构域或恒定区、或TCR δ 恒定结构域或恒定区。在一些情况下,恒定区包含恒定结构域。在一些情况下,恒定区还包含跨膜区、连接区、胞质区或其组合。

[0101] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含连接序列。连接序列可用于将一个核酸分子与另一核酸分子连接。一个核酸分子的连接序列可以与另一核酸分子的抗连接序列杂交(例如,形成一个或多个碱基对)。抗连接序列可以与连接序列互补(例如,完全或基本上互补)。抗连接序列可以在某些条件(例如,温度、缓冲条件、pH等)下与连接序列杂交。抗连接序列可以是连接序列的反向互补序列(或互补序列)。当连接序列与抗连接序列杂交时,形成的碱基对可以是至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、100个或更多个碱基对。在连接序列与抗连接序列之间形成的碱基对可以是连续的或非连续的。例如,在形成非连续的碱基对的情况下,可以存在分隔配对区的一个或多个非配对区。如果第一核酸分子包含连接序列,则第二核酸分子上的连接序列的互补序列可称为抗连接序列。连接序列(或抗连接序列)可以具有各种长度。例如,连接序列(或抗连接序列)的长度可以大于或等于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、100个或更多个核苷酸。连接序列(或抗连接序列)的长度可以小于或等于约300、250、200、150、100、90、85、80、75、70、65、60、55、50、45、40、35、30、25、20、15、14、13、12、11、10、9、8、7、6、5、4、3或2个核苷酸。连接序列(或抗连接序列)可以在核酸分子的5'末端或3'末端。连接序列(或抗连接序列)也可以是核酸分子的内部序列。例如,连接序列可以是内部连接序列,并且可以通过切割核酸分子的内部序列(例如,与内部连接序列相邻的序列)而在5'末端或3'末端暴露。实施例1中提供了内部连接序列的示例,链间连接序列(ICC)。在一些情况下,连接序列和抗连接序列用于将编码TCR链的CDR3或其部分的核酸分子与包含TCR V基因或其部分的另一核酸分子连接。在一些情况下,连接序列和抗连接序列用于将包含TCR J区的核酸分子与包含TCR V基因或其部分的另一核酸分子连接。在一些情况下,连接序列和抗连接序列用于将包含编码TCR的CDR3或其部分和J区的序列的核酸分子与包含TCR V基因或一部分的另一核酸分子连接。在一些情况下,连接序列和抗连接序列用于将包含编码CDR3或其部分、J区和TCR V基因或其部分的序列的核酸分子与编码TCR的恒定结构域或其部分的另一核酸分子连接。

[0102] 连接序列(或抗连接序列)可以是编码TCR V基因的一部分(例如,与重排基因中编码CDR3的序列相邻的TCR V基因的部分)的序列。并且在此类情况下,连接序列和连接序列

库中的一个或多个其他连接序列可以编码相同的氨基酸序列(例如,与CDR3相邻的TCR V基因的保守部分)。当连接序列编码TCR V基因的保守部分时,所述连接序列可以是密码子多样化的,使得所述连接序列可用于将核酸分子与另一核酸分子特异性连接,从而产生编码TCR的同源对的构建的核酸分子。在一些实施方案中,连接序列(或抗连接序列)包含与重排基因中编码CDR3的序列相邻的TCR V基因的至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100、150、200个或更多个核苷酸。由于连接序列和抗连接序列的特异性,可以在批量反应中(例如,在同一区室中)构建具有编码不同TCR的不同序列的核酸分子的库。连接序列可根据参考序列(例如,通过测序确定的TCR链的天然序列)指示编码CDR3的序列应与哪个TCR V基因连接。连接序列(或抗连接序列)可以是任意的(例如预定的)序列,其可以不编码TCR V基因的一部分。在此类情况下,可以在将两个核酸片段连接在一起后移除所述任意序列。

[0103] 图7描绘了使用任意连接序列(701)和抗连接序列(702)将CDR3-J多核苷酸与指定的V基因种系多核苷酸连接的实例(细箭头)。这里每个V基因种系多核苷酸具有部分双链结构。顶部链(其3'末端在此图中的右侧)在其3'末端具有单链区。连接序列和抗连接序列可以是单链的并且可以彼此杂交。连接序列和抗连接序列仅用于特异性杂交的目的,并且可能完全与TCR无关,因此是任意的。在连接序列与抗连接序列之间杂交后,V基因种系多核苷酸的顶部链的3'末端可以与CDR3-J多核苷酸杂交并且可以通过DNA聚合酶延伸。与CDR3-J多核苷酸杂交的V基因种系多核苷酸的顶部链的3'末端上核苷酸的数目可以是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15,或至多20个。

[0104] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含自切割肽。自切割肽可以是2A肽、内含肽或刺猬肽。2A肽的实例包括但不限于P2A(例如序列:ATNFSLLKQAGDVEENPGP(SEQ ID NO:284))、E2A(例如,序列QCTNYALLKLAGDVESNPGP(SEQ ID NO:285))、F2A(例如,序列VKQTLNFDLLKLAGDVESNPGP(SEQ ID NO:286)) and T2A(例如,序列EGRGSLTTCGDVEENPGP(SEQ ID NO:287))肽。

[0105] 用于构建编码TCR或其部分的核酸分子的核酸片段可包含限制酶识别位点。例如,限制酶识别位点可以是IIS型限制酶的识别位点。可用于本公开的IIS型限制酶的实例包括但不限于EarI、MnlI、PleI、AlwI、BbsI、BbvI、BcoDI、BsaI、BseRI、BsmAI、BsmBI、BspMI、Esp3I、HgaI、SapI、SfaNI、BbvI、BsmFI、BsrDI、BtsI、FokI、BseRI、HphI、MlyI和MboII。在一些情况下,在核酸构建过程中可以使用两种或更多种不同的限制酶。在一些情况下,可以使用产生4-bp 5'突出端的限制酶(例如,BbsI、BbvI、BcoDI、BsaI、BsmBI、FokI等)。在一些情况下,可以使用产生平末端或3'突出端的限制酶(例如,BseRI、BsrDI、BtsI、MlyI等)。

[0106] 用于构建编码TCR或其部分的核酸分子的核酸片段可以是环化的。例如,可以通过连接将核酸片段的两个末端连接起来而使核酸片段环化。连接可以是平末端连接。可以在使用5'至3'核酸外切酶(例如,Gibson Assembly)、3'至5'核酸外切酶(例如,序列和连接酶独立的克隆或SLIC)或USER酶混合物(例如,USER友好的DNA重组或USERec)产生粘性末端之后进行连接。环化方法的其他实例包括但不限于环状聚合酶延伸克隆(CPEC)和无缝连接克隆提取物(SLiCE)组装。或者,这两个末端可以通过重叠PCR连接。多种连接酶可用于连接,例如,包括但不限于,T4 DNA连接酶、T4 RNA连接酶、大肠杆菌DNA连接酶。

[0107] 可化学合成用于构建编码TCR链或其部分的核酸分子的核酸片段。例如,核酸片段可以通过基于芯片的合成进行预合成。在一些情况下,合成的核酸片段的长度可以等于或大于约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400个或更多个核苷酸。在一些情况下,合成的核酸片段的长度可以等于或小于约500、450、400、350、300、250、200、150、100、90、80、70、60、50、40、30、20或10个核苷酸。

[0108] 编码TCR的两条肽链的两条核酸序列可以以若干种方向构建,例如头对头、头对尾和尾对尾。如本文所述,“头”是指有义核酸链的“5'末端”并且“尾”是指有义核酸链的“3'末端”。在一些情况下,方向为头对尾,可以控制编码TCR的配对的核酸序列的顺序(例如,TRA后接TRB,或TRB后接TRA)。

[0109] 本文所述的任何核酸分子可以是双链核酸分子或单链核酸分子。在一些情况下,核酸分子可包含双链区和单链区。例如,具有连接序列或抗连接序列的核酸分子可以是具有连接序列或抗连接序列区作为单链区(例如,突出端或粘性末端)的双链核酸分子。突出端的长度可以为至少1、2、3、4、5、6、7、8、9、10个或更多个核苷酸。突出端可以位于核酸分子的5'末端或3'末端。

[0110] 本文所述的任何核酸分子可以包含一个或多个修饰的核苷酸。修饰的核苷酸的实例包括但不限于二氨基嘌呤、5-氟尿嘧啶、5-溴尿嘧啶、5-氯尿嘧啶、5-碘尿嘧啶、次黄嘌呤、黄嘌呤、4-乙酰胞嘧啶、5-(羧基羧甲基)尿嘧啶、5-羧甲基氨基甲基-2-硫尿嘧啶、5-羧甲基氨基甲基尿嘧啶、二氢尿嘧啶、 β -D-半乳糖基腺苷、肌苷、N6-异戊烯基腺嘌呤、1-甲基鸟嘌呤、1-甲基肌苷、2,2-二甲基鸟嘌呤、2-甲基腺嘌呤、2-甲基鸟嘌呤、3-甲基胞嘧啶、5-甲基胞嘧啶、N6-腺嘌呤、7-甲基鸟嘌呤、5-甲基氨基甲基尿嘧啶、5-甲氧基氨基甲基-2-硫尿嘧啶、 β -D-甘露糖基腺苷、5'-甲氧基羧甲基尿嘧啶、5-甲氧基尿嘧啶、2-甲硫基-D46-异戊烯基腺嘌呤、尿嘧啶-5-羧乙酸(v)、怀丁氧苷(wybutoxosine)、假尿嘧啶、腺苷、2-硫胞嘧啶、5-甲基-2-硫尿嘧啶、2-硫尿嘧啶、4-硫尿嘧啶、5-甲基尿嘧啶、尿嘧啶-5-羧基乙酸甲酯、尿嘧啶-5-羧基乙酸(v)、5-甲基-2-硫尿嘧啶、3-(3-氨基-3-N-2-羧丙基)尿嘧啶、(acp3)w、2,6-二氨基嘌呤等。在一些情况下,核苷酸可包含其磷酸部分的修饰,包括对三磷酸部分的修饰。此类修饰的非限制性实例包括较大长度的磷酸链(例如,具有4、5、6、7、8、9、10个或更多个磷酸部分的磷酸链)和用硫醇部分的修饰(例如, α -硫代三磷酸和 β -硫代三磷酸)。核酸分子还可以在碱基部分(例如,在通常可用于与互补核苷酸形成氢键的一个或多个原子处和/或在通常不能与互补核苷酸形成氢键的一个或多个原子处)、糖部分或磷酸骨架处进行修饰。核酸分子还可以含有胺修饰的基团,诸如氨基烯丙基1-dUTP(aa-dUTP)和氨基己基丙烯酰胺-dCTP(aha-dCTP),以允许共价附接胺反应性部分,诸如N-羧基琥珀酰亚胺酯(NHS)。本公开的寡核苷酸中的标准DNA碱基对或RNA碱基对的替代物可以提供更高的密度(单位为比特/立方毫米),更高的安全性(对天然毒素的意外或有目的的合成的抗性)、更容易区分光编程的聚合酶、或更低的二级结构。此类替代的碱基对可与用于从头合成和/或扩增合成的天然和突变聚合酶相容。

[0111] 构建编码TCR的核酸分子的示例性工作流程如图1A-图1C所示。在使用本文所述的方法构建编码TCR的核酸分子之前,可使用各种现有方法(例如,单细胞条形码化和测序)预定TCR的多个同源对。各种测序方法可用于确定配对的TCR链的序列,例如Sanger测序、高通量测序、合成测序、单分子测序、连接测序、RNA-Seq(Illumina)、下一代测序、数字基因表达

(Helicos)、克隆单微阵列(Solexa)、鸟枪法测序、Maxim-Gilbert测序或大规模平行测序。来自测序文库的配对序列可用作TCR链的同源对的参考序列,使得可通过连接序列与抗连接序列之间的特异性相互作用知道哪个CDR3与哪个V基因配对。可使用本文所述的方法批量构建编码不同TCR的多个核酸分子,但图1A-图1C中示出了一个分子的构建作为实例。包含编码第一CDR3(例如CDR3 α)和第二CDR3(例如CDR3 β)的序列的第一核酸分子可与包含衍生自第一TCR V基因(例如TRAV)的序列的第二核酸分子接触。第一核酸分子的连接序列(例如ConA#*)可与第二核酸分子的抗连接序列(ConA#)杂交以连接两个核酸分子。可进行延伸和连接以产生第三核酸分子,其包含衍生自第一TCR V基因的序列以及编码第一CDR3和第二CDR3的序列。接着,可以使用限制酶(例如,图1A的TIISRE1)产生第三核酸分子的突出端(或粘性末端)。接着,第三核酸分子可以与包含编码第一恒定区或恒定结构域(例如,TRBC)的序列的第四核酸分子接触。然后可将第三核酸分子通过突出端与第四核酸分子连接,以产生第五核酸分子,其包含衍生自第一TCR V基因的序列、编码第一CDR3和第二CDR3的序列以及编码第一恒定区的序列。可以将第五核酸分子环化并用限制酶(例如,TIISRE3)切割以暴露内部连接序列(例如,ICC)。接着,第五核酸分子可与包含衍生自第二TCR V基因(例如TRBV)的序列的第六核酸分子接触。第五核酸分子可以通过连接序列与抗连接序列之间的相互作用与第六核酸分子连接。接着,第六核酸分子可被限制酶(例如,TIISRE2)切割以产生突出端。接着,第六核酸分子可以与包含编码第二恒定区或恒定结构域(例如,TRAC)的序列的第七核酸分子接触。第六核酸分子与第七核酸分子可连接以形成包含编码配对TCR链的所有区域的第八核酸分子。可将第八核酸分子进一步构建到表达载体中,用于在宿主细胞中表达TCR链。应当理解,包含衍生自TCR V基因的序列的核酸片段可以是单链的,并且在这种情况下,编码CDR3的核酸片段的连接序列的3'末端可以在与抗连接序列杂交后延伸。

[0112] 本文所述的方法可用于产生单个TCR链的库,例如TCR α 链或TCR β 链的库。

[0113] 用于产生本文所述的多个核酸分子的方法可包括提供第一多种核酸分子(或核酸片段)。第一多种核酸分子中的核酸分子可包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的序列。第一CDR3和第二CDR3可以来自TCR链的同源对。接着,可以提供第二多种核酸分子。第二多种核酸分子中的核酸分子可包含衍生自TCR V基因的序列。核酸分子可不包含编码恒定结构域的序列。接着,可以使第一多种核酸分子与第二多种核酸分子接触。第一多种核酸分子中的核酸分子可与第二多种核酸分子中的核酸分子连接以形成包含编码第一CDR3和第二CDR3的序列以及衍生自TCR V基因的序列的核酸分子。编码第一CDR3和第二CDR3的序列以及TCR V基因可衍生自TCR链的同源对。

[0114] 用于产生多种核酸分子(所述多种中的每种核酸分子编码T细胞受体(TCR)链或其区域)的方法可包括使第一多种核酸分子与第二多种核酸分子接触,以产生包含至少两种不同核酸分子的第三多种核酸分子。至少两种不同核酸分子中的每种可具有编码不同TCR链或其区域的不同序列。至少两种不同核酸分子可以在同一区室中产生。在一些情况下,至少约5、10、20、50、100、200、300、400、500、1,000、1,500、2,000、2,500、3,000、3,500、4,000、5,000、6,000、7,000、8,000、9,000、10,000、12,000、15,000、20,000、100,000、1,000,000、10,000,000个或更多个编码不同TCR的不同序列可在同一区室中产生。

[0115] 用于产生本文所述的多种核酸分子的方法可包括提供第一多种核酸分子。第一多种核酸分子中的核酸分子可包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第

二CDR3的序列。第一CDR3和第二CDR3可以来自TCR链的同源对。接着,可以提供第二多种核酸分子。第二多种核酸分子中的核酸分子可包含衍生自TCR V基因的序列。接着,可以使第一多种核酸分子与第二多种核酸分子接触。第一多种核酸分子中的核酸分子可与第二多种核酸分子中的核酸分子连接以形成包含编码第一CDR3和第二CDR3的序列以及衍生自TCR V基因的序列的线性核酸分子。编码第一CDR3和第二CDR3的序列以及TCR V基因可衍生自TCR链的同源对。

[0116] 用于产生多种核酸分子的方法可包括提供第一多种核酸分子。第一多种核酸分子中的核酸分子可包含 (i) 编码第一T细胞受体 (TCR) 链的第一CDR3和第二TCR链的第二CDR3的合成序列,和 (ii) 编码第三T细胞受体 (TCR) 链的第三CDR3和第四TCR链的第四CDR3的合成序列。第一CDR3和第二CDR3可以来自TCR链的第一同源对,并且第三CDR3和第四CDR3可以来自TCR链的第二同源对。接着,可以提供第二多种核酸分子。第二多种核酸分子中的核酸分子可包含衍生自TCR V基因的序列。接着,可以使第一多种核酸分子与第二多种核酸分子接触。第一多种核酸分子中的核酸分子可与第二多种核酸分子中的核酸分子连接以形成包含编码第一CDR3和第二CDR3的序列以及衍生自TCR V基因的序列的核酸分子。编码第一CDR3和第二CDR3的序列以及TCR V基因可衍生自TCR链的同源对。

[0117] 用于产生编码T细胞受体 (TCR) 链或其部分的核酸分子的方法可包括提供包含编码TCR链CDR3的序列的至少一种核酸分子。接着,可以提供多种核酸分子。所述多种中的每种核酸分子可包含衍生自TCR V基因的序列。所述多种核酸分子可包含衍生自至少两种不同TCR V基因的至少两种不同序列。在一些情况下,多种核酸分子可包含衍生自至少2、5、10、15、20、25、30、35、40种或更多种不同TCR V基因的至少2、5、10、15、20、25、30、35、40种或更多种不同序列。接着,可将包含编码TCR链CDR3的序列的至少一种核酸分子与多种核酸分子(每种包含衍生自TCR V基因的序列)在同一区室中接触。包含编码TCR链的CDR3的序列的至少一种核酸分子能够与多种核酸分子中的核酸分子连接,以产生包含编码CDR3的序列和衍生自至少两种不同TCR V基因之一的序列的第三核酸分子,从而产生编码TCR链或其部分的核酸分子。

[0118] 可用于本文所述方法的本文所述组合物可包含第一多种核酸分子。第一多种核酸分子中的每种核酸分子可包含编码第一T细胞受体 (TCR) 链的第一CDR3和第二TCR链的第二CDR3的序列。第一CDR3和第二CDR3可以来自TCR链的同源对。组合物还可包含第二多种核酸分子。第二多种核酸分子中的每种核酸分子可包含衍生自TCR V基因的序列。第二多种核酸分子中的每种核酸分子可不包含编码第一CDR3和第二CDR3的序列。在此组合物中, (i) 第一多种核酸分子中的每种核酸分子可包含编码不同的第一CDR3和/或第二CDR3的序列,和/或 (ii) 第二多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。

[0119] 可用于本文所述方法的本文所述组合物可包含多种核酸分子。多种核酸分子中的每种核酸分子可包含衍生自T细胞受体 (TCR) V基因的序列。多种核酸分子可包含具有第一连接序列的第一核酸分子和具有第二连接序列的第二核酸分子。第一连接序列可以不同于第二连接序列。

[0120] 可用于本文所述方法的本文所述组合物可包含多种核酸分子。多种核酸分子中的每种核酸分子可编码T细胞受体 (TCR) 链的CDR3。所述多种中的第一核酸分子可包含第一连接序列,并且所述多种中的第二核酸分子可包含第二连接序列。第一连接序列可以不同于

第二连接序列。

[0121] 可用于本文所述方法的本文所述组合物可包含多种核酸分子。所述多种中的每种核酸分子可包含编码T细胞受体 (TCR) 链的至少十个氨基酸的序列。所述多种中的第一核酸分子可包含第一连接序列,并且所述多种中的第二核酸分子可包含第二连接序列。第一连接序列可以不同于第二连接序列。第一连接序列或第二连接序列可以编码TCR链的一部分。第一连接序列或第二连接序列可以与编码TCR链的至少十个氨基酸的序列在框内。

[0122] 可用于本文所述方法的本文所述组合物可包含多种核酸分子。所述多种核酸分子中的每种核酸分子可包含衍生自T细胞受体 (TCR) V基因的序列,并且可不包含CDR3序列。所述多种中的第一核酸分子可包含第一抗连接序列,并且所述多种中的第二核酸分子可包含第二抗连接序列。第一抗连接序列可以不同于第二抗连接序列。衍生自第一核酸分子和第二核酸分子的TCR V基因的序列可以衍生自不同的TCR V基因。所述组合物还可包含含有编码TCR链的CDR3的序列至少一种核酸分子。所述至少一种核酸分子还可包含与第一抗连接序列互补的第一连接序列。

[0123] 本公开提供了用于组装或合成包含多种TCR的TCR文库的组合物和方法。在一些情况下,从TCR文库中分离或纯化特定TCR序列(例如,目标TCR)用于进一步表征或操纵可能是有用的。为此,可在用于构建编码TCR或其部分的序列的核酸分子或片段中包含条形码。在一些情况下,包含编码CDR3的序列的核酸片段包含条形码。在一些情况下,包含编码第一TCR链的第一CDR3和第二TCR链的第二CDR3的序列的核酸片段包含条形码。例如,CDR3-J寡聚物或配对的CDR3-J寡聚物可以包含条形码。连接序列(或在一些情况下,抗连接序列)可包含条形码。CDR3-J寡聚物的链间连接序列(或ICC)可以包含条形码。条形码可以是引物结合位点,例如TCR特异性引物结合位点或DOPBS。

[0124] 例如,编码配对的CDR3-J寡聚物库(例如,图1A)中的独特配对的CDR3-J的每个序列可以包含独特的条形码(或独特的DOPBS)。可以任意设计DOPBS的序列。DOPBS的序列可被设计成避免常见的缺陷,诸如不需要的二级结构、限制性位点、与TCR基因中其他序列的相似性、或引物结合位点之间的相似性。条形码(或DOPBS)的长度可以为至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40个或更多个核苷酸。DOPBS可以是包含在配对的CDR3-J库的每个序列中的额外序列。DOPBS可以是已经包含在配对的CDR3-J库的每个序列中的序列。例如,连接序列或其部分可用作DOPBS。表3中列出的序列可以用作DOPBS。图1C的步骤(9)的产物可用作拨出PCR(dial-out PCR)中的模板,所述拨出PCR使用对应于T2A-3的正向引物 and 对应于与目标TCR相关联的DOPBS的反向引物。PCR产物可进行图1C的步骤(10)和(11)。最终产物可主要含有目标TCR。

[0125] TCR的表达

[0126] 使用本文提供的方法,可将各自编码TCR或其部分的核酸分子库进一步递送到宿主细胞中用于表达。可以将构建的核酸分子插入载体中以便在宿主细胞中表达。构建的核酸分子可作为线性或环状核酸链递送到受体细胞中。在一些情况下,构建的核酸或包含构建的核酸的载体可以通过电穿孔递送到受体细胞中。在一些情况下,构建的核酸或包含构建的核酸的载体可通过载剂(诸如阳离子聚合物)递送。

[0127] 载体可以是质粒、转座子(例如睡美人(Sleeping Beauty)、Piggy Bac)、腺病毒载

体、AAV载体、逆转录病毒载体或慢病毒载体。载体的非限制性实例包括质粒、穿梭载体、噬菌粒、粘粒、病毒粒子、逆转录病毒载体、腺病毒载体或基因疗法中常用的颗粒和/或载体。合适的质粒载体的非限制性实例包括pUC、pBR322、pET、pBluescript及其变体。此外,载体可包含额外的表达控制序列(例如,增强子序列、Kozak序列、聚腺苷酸化序列、转录终止序列等)、选择性标记序列(例如,抗生素抗性基因)、复制起点等。载体可包含允许其在宿主细胞中复制的核酸序列,诸如复制起点。载体还可以包含一种或多种选择标记基因和其他遗传元件。载体可以是表达载体,其包含与允许TCR表达的序列可操作地连接的根据本公开的编码TCR或其部分的构建的核酸序列。载体的其他实例包括但不限于病毒和非病毒载体,诸如逆转录病毒载体(包括慢病毒载体)、腺病毒载体(包括其复制型、复制缺陷型和无肠型)、腺相关病毒(AAV)载体、猿猴病毒40(SV-40)载体、牛乳头状瘤载体、爱泼斯坦-巴尔病毒载体、疱疹病毒载体、痘苗病毒载体、莫洛尼鼠白血病毒载体、Harvey鼠肉瘤病毒载体、鼠乳腺肿瘤病毒载体、劳斯肉瘤病毒载体和非病毒质粒。杆状病毒载体可适用于在昆虫细胞中表达。非病毒载体可以配制成纳米颗粒、阳离子脂质、阳离子聚合物、金属纳米聚合物、纳米棒、脂质体、胶束、微泡、细胞穿透肽或脂球。

[0128] 在一些实施方案中,载体是自扩增RNA复制子,也称为自复制(m)RNA、自复制(m)RNA、自扩增(m)RNA或RNA复制子。自扩增RNA复制子是可以自我复制的RNA。在一些实施方案中,自扩增RNA复制子可在细胞内自我复制。在一些实施方案中,自扩增RNA复制子编码RNA聚合酶和目标分子。RNA聚合酶可以是RNA依赖性RNA聚合酶(RDRP或RdRp)。自扩增RNA复制子也可编码蛋白酶或RNA加帽酶。在一些实施方案中,自扩增RNA复制子载体属于或衍生自称为甲病毒的披膜病毒科(Togaviridae)病毒,其可包括东方马脑炎病毒(EEE)、委内瑞拉马脑炎病毒(VEE)、沼泽地病毒、穆坎博病毒(Mucambo virus)、皮克孙纳病毒(Pixuna virus)、西部马脑炎病毒(WEE)、辛德毕斯病毒、南非虫媒病毒86号、塞姆利基森林病毒、米德尔堡病毒、基孔肯雅病毒、奥-奈氏病毒(Onyong-nyong virus)、罗斯河病毒、巴马森林病毒、盖他病毒、鹭山病毒、贝巴鲁病毒、马亚罗病毒、乌纳病毒、奥拉病毒、沃达罗河病毒、巴班肯病毒(Babanki virus)、克泽拉格齐病毒(Kyzylagach virus)、高地J病毒、摩根堡病毒、恩杜茂病毒、博吉河病毒(Buggy Creek virus)以及由国际病毒分类委员会(ICTV)分类为甲病毒的任何其他病毒。在一些实施方案中,自扩增RNA复制子是或含有减毒形式的甲病毒的部分,诸如VEE TC-83疫苗株。在一些实施方案中,自扩增RNA复制子载体已经在体外、体内、离体或在电脑中(in silico)进行工程化或选择,用于在宿主细胞、靶细胞或生物体中的特定功能(例如,延长的或增加的二分免疫受体表达)。例如,携带自扩增RNA复制子的不同变体的宿主细胞群体可以基于在不同时间点一种或多种目标分子(在自扩增RNA复制子或宿主基因组中编码的)的表达水平来选择。在一些实施方案中,经选择的或工程化的自扩增RNA复制子已被修饰成减少来自宿主细胞或生物体的I型干扰素反应、先天性抗病毒反应或适应性免疫反应,这导致RNA复制子的蛋白质表达在宿主细胞、靶细胞或生物体中持续更长或以更高水平表达。在一些实施方案中,这种优化的自扩增RNA复制子序列从具有所需表型性状(例如,与野生型株或疫苗株相比,目标分子的表达更高或更持久,或针对载体的先天性抗病毒免疫反应降低)的单个细胞或细胞群体获得。在一些实施方案中,携带所需或所选择的自扩增RNA复制子序列的细胞从用包含自扩增RNA复制子的治疗剂治疗后具有有益反应特征的受试者(例如人或动物)(例如完全缓解的精英反应者或受试者)获得。在一些

实施方案中,自扩增RNA复制子可含有一种或多种亚基因组序列以产生一种或多种亚基因组多核苷酸。在一些实施方案中,亚基因组多核苷酸充当功能性mRNA分子,用于通过细胞翻译机制进行翻译。亚基因组多核苷酸可以通过自扩增RNA复制子上确定序列元件(例如亚基因组启动子或SGP)的功能产生,所述确定序列元件指导聚合酶由亚基因组序列产生亚基因组多核苷酸。在一些实施方案中,SGP被RNA依赖性RNA聚合酶(RDRP或RdRp)识别。在一些实施方案中,多个SGP序列存在于单个自扩增RNA复制子上,并且可以位于编码二分免疫受体、二分免疫受体的成分或其他剂的亚基因组序列的上游。在一些实施方案中,可以修饰SGP序列的核苷酸长度或组成以改变亚基因组多核苷酸的表达特征。在一些实施方案中,不相同的SGP序列位于自扩增RNA复制子上,使得对应的亚基因组多核苷酸的比率不同于SGP序列相同的情况。在一些实施方案中,不相同的SGP序列指导TCR和其他剂(例如细胞因子)的产生,使得它们相对于彼此以一定的比率产生,从而导致TCR表达增加、靶细胞扩增增加或更快而对靶细胞或宿主无细胞毒性作用,或减弱针对RNA复制子的先天或适应性免疫反应。在一些实施方案中,亚基因组序列和SGP序列相对于彼此和基因组序列本身的位置可用于改变亚基因组多核苷酸相对于彼此的比率。在一些实施方案中,编码TCR的SGP和亚基因组序列可以位于编码其他剂的SGP和亚基因组区域的下游,使得TCR的表达相对于其他剂大幅增加。在一些实施方案中,RNA复制子或SGP已被选择或工程化成表达最佳量的细胞因子,使得细胞因子促进T细胞的扩增或增强TCR的治疗效果,但不引起严重的副作用,诸如细胞因子释放综合征、细胞因子风暴或神经毒性。

[0129] 两条链的表达可以由两个启动子或一个启动子驱动。在一些情况下,使用两个启动子。在一些情况下,两个启动子连同它们各自的两条链的蛋白编码序列可以以头对头、头对尾或尾对尾方向排列。在一些情况下,使用一个启动子。两个蛋白编码序列可以框内连接,使得一个启动子可用于表达两条链。并且在此类情况下,两个蛋白编码序列可以以头对尾方向排列并且可以与核糖体结合位点(例如内部核糖体结合位点或IRES)、蛋白酶切割位点或自加工切割位点(诸如编码2A肽的序列)连接以促进双顺反子表达。在一些情况下,两条链可以用肽接头连接,使得两条链可以表达为单链多肽。每条表达的链可以含有包含重排的V(D)J基因的完整可变结构域序列。每条表达的链可以含有完整的可变结构域序列,包括CDR1、CDR2和CDR3。每条表达的链可以含有完整的可变结构域序列,包括FR1、CDR1、FR2、CDR2、FR3和CDR3。在一些情况下,每条表达的链还可含有恒定结构域序列。

[0130] 为了产生表达载体,可以向构建的核酸分子中加入额外的序列。这些额外的序列包括载体骨架(例如,载体在靶细胞或临时宿主(诸如大肠杆菌)中复制所需的元件)、启动子、IRES、编码自切割肽的序列、终止子、辅助基因(诸如有效负载)、以及免疫受体多核苷酸的部分序列(诸如编码恒定结构域的序列的部分)。

[0131] 蛋白酶切割位点包括但不限于肠激酶切割位点:(Asp)4Lys (SEQ ID NO:288);因子Xa切割位点:Ile-Glu-Gly-Arg (SEQ ID NO:289);凝血酶切割位点,例如Leu-Val-Pro-Arg-Gly-Ser (SEQ ID NO:290);肾素切割位点,例如His-Pro-Phe-His-Leu-Val-Ile-His (SEQ ID NO:291);胶原酶切割位点,例如X-Gly-Pro(其中X是任何氨基酸);胰蛋白酶切割位点,例如Arg-Lys;病毒蛋白酶切割位点,诸如病毒2A或3C蛋白酶切割位点,包括但不限于来自小核糖核酸病毒的蛋白酶2A切割位点、甲型肝炎病毒3C切割位点、人鼻病毒2A蛋白酶切割位点、小核糖核酸病毒3蛋白酶切割位点;和半胱天冬酶蛋白酶切割位点,例如由激活

的半胱天冬酶-3识别和切割的DEVD (SEQ ID NO:292), 其中切割发生在第二个天冬氨酸残基之后。在一些实施方案中, 本公开提供了包含蛋白酶切割位点的表达载体, 其中所述蛋白酶切割位点包含细胞蛋白酶切割位点或病毒蛋白酶切割位点。在一些实施方案中, 第一蛋白酶切割位点包含由弗林蛋白酶 (furin); IPNV的VP4; 烟草蚀刻病毒 (TEV) 蛋白酶; 鼻病毒的3C蛋白酶; PC5/6蛋白酶; PACE蛋白酶、LPC/PC7蛋白酶; 肠激酶; 因子Xa蛋白酶; 凝血酶; 基因酶I; MMP蛋白酶; 芜菁花叶马铃薯Y病毒 (turnip mosaic potyvirus) 的核包含蛋白a (N1a); 登革热4型黄病毒 (Dengue type 4 flaviviruses) 的NS2B/NS3、黄热病病毒的NS3蛋白酶; 花椰菜花叶病毒的ORF V; KEX2蛋白酶; CB2; 或2A识别的位点。在一些实施方案中, 蛋白质切割位点是病毒内部可切割的信号肽切割位点。在一些实施方案中, 病毒内部可切割的信号肽切割位点包含来自丙型流感病毒、丙型肝炎病毒、汉坦病毒、黄病毒或风疹病毒的位点。

[0132] 包括在本公开的载体中的合适的IRES元件可包含能够接合真核核糖体的RNA序列。在一些实施方案中, 本公开的IRES元件为至少约250个碱基对、至少约350个碱基对或至少约500个碱基对。本公开的IRES元件可衍生自生物体的DNA, 所述生物体包括但不限于病毒、哺乳动物和果蝇。在一些情况下, 衍生IRES元件的病毒DNA包括但不限于小核糖核酸病毒互补DNA (cDNA)、脑心肌炎病毒 (EMCV) cDNA和脊髓灰质炎病毒cDNA。衍生IRES元件的哺乳动物DNA的实例包括但不限于编码免疫球蛋白重链结合蛋白 (BiP) 的DNA和编码碱性成纤维细胞生长因子 (bFGF) 的DNA。衍生IRES元件的果蝇DNA的实例包括但不限于来自黑腹果蝇的触角足 (Antennapedia) 基因。脊髓灰质炎病毒IRES元件的其他实例包括例如脊髓灰质炎病毒IRES、脑心肌炎病毒IRES或甲型肝炎病毒IRES。黄病毒IRES元件的实例包括丙型肝炎病毒IRES、GB病毒B IRES或瘟病毒IRES, 包括但不限于牛病毒性腹泻病毒IRES或经典猪瘟病毒IRES。

[0133] 自加工切割位点的实例包括但不限于内含肽序列; 修饰的内含肽; 刺猬序列; 其他猪科序列; 2A序列, 例如衍生自口蹄疫病毒 (FMDV) 的2A序列; 以及其各自的变体。

[0134] 用于重组免疫球蛋白或其他蛋白质表达的载体可以包括任何数量的启动子, 其中所述启动子是组成型的、调节型的或诱导型的、细胞类型特异性的、组织特异性的或物种特异性的。其他实例包括四环素反应性启动子。载体可以是适于表达重组构建的基因的宿主细胞的复制子, 并且它可以包括在细菌细胞 (例如大肠杆菌) 中有功能的复制子。启动子可以是组成型或诱导型的, 其中诱导与例如特定细胞类型或特定成熟水平相关联。或者, 许多病毒启动子可能是合适的。启动子的实例包括 β -肌动蛋白启动子、SV40早期和晚期启动子、免疫球蛋白启动子、人巨细胞病毒启动子、逆转录病毒启动子、延伸因子1A (EF-1A) 启动子、磷酸甘油酸激酶 (PGK) 启动子和弗里德 (Friend) 脾脏病灶形成病毒启动子。启动子可以与或不与增强子缔合, 其中增强子可以与特定启动子天然缔合或与不同启动子缔合。

[0135] 应用

[0136] 本文所述的组合物和方法可具有各种应用。示例性应用可以由测序数据 (例如单细胞测序数据) 重新构建编码天然配对的TCR的序列。在一些应用中, 人们可能希望重新构建编码从肿瘤浸润性T细胞中鉴定的天然配对的TCR的序列。在这些应用中, 来自受试者的新鲜组织样品 (例如, 新鲜实体肿瘤样品) 可用于单细胞测序, 以获得天然配对的TCR的两条TCR链的序列信息。然而, 当含有肿瘤浸润性细胞的组织样品 (例如, 不是体液样品的固体物质样品) 是冷冻样品或固定样品 (例如, FFPE样品) 时, 分离细胞以获得单细胞悬浮液可能

具有挑战性。在这些情况下,来自同一受试者的含有外周T细胞的血液样品可用于单细胞测序,以鉴定天然配对的TCR的序列。因为血液样品可能含有从组织样品释放到血流中的肿瘤浸润性T细胞,所以从血液样品获得的序列可能含有来自这些肿瘤浸润性T细胞的序列。然后,来自同一受试者的组织样品可用于批量测序。尽管组织样品的批量测序可能不提供天然配对的TCR的配对序列,但它可提供单个TCR链的CDR3序列。然后,在组织样品的批量测序中获得的CDR3序列(在本文中称为“组织CDR3序列”)可用于与在血液样品的单细胞测序中获得的配对序列进行比对。如果配对序列的CDR3序列与组织CDR3序列匹配,则可鉴定配对序列并用于任何下游应用。

[0137] 单细胞测序是指从单个细胞获得序列信息。在单细胞测序中,可以将细胞群体制成单细胞悬浮液并分隔成单独的分区。在每个分区中,从单细胞释放的序列可被条形码化并随后测序。各种单细胞测序方法可用于TCR重建(参见De Simone M,Rossetti G和Pagani M(2018)Single Cell T Cell Receptor Sequencing:Techniques and Future Challenges.Front.Immunol.9:1638)。批量测序是指从细胞群体获得序列信息。在批量测序中,可以从细胞的混合物中分离核酸分子并一起进行测序。

[0138] 图9A示出了使用血液样品来鉴定肿瘤样品中的肿瘤浸润性TCR的示例性工作流程。首先,可以从患者抽取血液样品。接着,可以从血液样品中分离含有外周血单核细胞的PBMC样品。例如,可以使用ficoll(一种分离血液各层的亲水性多糖)和梯度离心从全血中提取这些细胞。接着,可以从PBMC样品中分离T细胞。可以通过裂解红细胞并去除单核细胞,例如通过PERCOLL™梯度离心或通过逆流离心淘析,从PBMC中分离T细胞。任选地,可以通过基于标记的分选进一步富集T细胞亚群。标记可以是细胞表面标记。细胞表面标记的实例包括但不限于CD39、CD69、CD103、CD25、PD-1、TIM-3、OX-40、4-1BB、CD137、CD3、CD28、CD4、CD8、CD45RA、CD45RO、GITR和FoxP3。标记可以是细胞因子。细胞因子标记的实例包括但不限于IFN- γ 、TNF- α 、IL-17A、IL-2、IL-3、IL-4、GM-CSF、IL-10、IL-13、粒酶B和穿孔素。然后可对T细胞或T细胞亚群进行单细胞测序以获得天然配对的TCR的配对序列(例如,图9A中的信息配对的TCR序列)。肿瘤样品也可以从同一患者获得。肿瘤样品可以是固定或冷冻样品。例如,肿瘤样品可以通过固定剂(诸如甲醛)来固定。肿瘤样品可以是福尔马林固定石蜡包埋(FFPE)的组织样品。接着,可对肿瘤样品进行批量测序以获得TCR链的CDR3序列。接着,得自肿瘤样品的CDR3序列可用于与配对序列的CDR3序列比较以鉴定肿瘤浸润性TCR。肿瘤浸润性TCR可在正常T细胞或细胞系中表达,其在图9A中示出为“虚拟TIL”。

[0139] 图9B示出了虚拟TIL的示例性应用。虚拟TIL可包含报告系统,其可用于靶反应性TCR的基于报告子的T细胞选择。例如,虚拟TIL可以是包含报告基因的报告细胞,当细胞的TCR与靶抗原结合时,所述报告基因被调节以发送信号。这些虚拟TIL可以通过与负载抗原的抗原呈递细胞(APC)或人工APC接触而被激活。接着,可基于报告系统产生的信号或其他选择机制(例如,细胞表面标记或细胞因子标记),例如通过FACS,选出靶反应性T细胞。所述选择可基于细胞接触MHC结合的抗原后虚拟TIL上的细胞表面标记表达。细胞表面标记可以是CD25、CD69、CD39、CD103、CD137、以及其他T细胞激活标记、或其任何组合。所述选择可以基于钙流入。所述选择也可以基于报告基因表达。报告基因可以是荧光蛋白(诸如GFP和mCherry)。报告基因可以在受TCR信号传导调节的转录因子的控制下。这些转录因子的实例包括但不限于AP-1、NFAT、NF- κ -B、Runx1、Runx3等。所述选择可以基于使用诸如ICS和细胞

因子捕获测定的方法从激活的虚拟TIL释放的细胞因子。图9C示出了虚拟TIL的另一应用。鉴定靶反应性TCR后,靶反应性TCR可在宿主细胞诸如自体T细胞(从获得组织样品和血液样品的同一患者分离的T细胞)中递送和表达。靶反应性TCR可在同种异体T细胞中递送和表达。然后可将表达靶反应性TCR的T细胞施用于同一患者以治疗疾病,诸如癌症。

[0140] 在来自受试者的组织样品(例如,固体样品)中鉴定天然配对的T细胞受体(TCR)的序列的方法可以包括在从受试者获得的含有多个外周T细胞的样品中鉴定一种或多种天然配对的TCR的一种或多种配对序列。一种或多种配对序列中的每种可包含CDR3序列。接着,可鉴定组织样品中TCR的TCR链的组织CDR3序列,与其天然配对的其他TCR链可能是未知的。组织CDR3序列可与一种或多种天然配对的TCR的一种或多种配对序列的至少一种配对序列的CDR3序列匹配,从而将该至少一种配对序列鉴定为组织样品中天然配对的TCR的序列。本文还提供了鉴定靶反应性T细胞受体(TCR)的方法。所述方法可包括提供包含使用本文所述方法鉴定的TCR的细胞。接着,细胞可以与抗原呈递细胞(APC)呈递的靶抗原接触。细胞可通过TCR与APC呈递的靶抗原结合,从而将TCR鉴定为靶反应性TCR。

[0141] 本文所述的APC可以是专业APC,诸如树突细胞、巨噬细胞或B细胞。APC可以是单核细胞或单核细胞衍生的树突细胞。aAPC可表达T细胞受体的配体和共刺激分子,并可激活并扩增用于转移的T细胞,同时在一些情况下改善它们的效力和功能。aAPC可被工程化以表达用于T细胞激活的任何基因。aAPC可被工程化以表达用于T细胞扩增的任何基因。aAPC可以是珠、细胞、蛋白质、抗体、细胞因子或任何组合。aAPC可将信号传递给可进行基因组移植的细胞群。例如,aAPC可以传递信号1、信号2、信号3或任何组合。信号1可以是抗原识别信号。例如,信号1可以是TCR与肽-MHC复合物的连接或激动性抗体针对CD3的结合,其可导致CD3信号-转导复合物的激活。信号2可以是共刺激信号。例如,共刺激信号可以是分别与ICOS-L、CD70和4-1BBL结合的抗CD28、诱导型共刺激因子(ICOS)、CD27和4-1BB(CD137)。信号3可以是细胞因子信号。细胞因子可以是任何细胞因子。细胞因子可以是IL-2、IL-7、IL-12、IL-15、IL-21或其任何组合。

[0142] 在一些情况下,aAPC可用于激活和/或扩增细胞群体。在一些情况下,人工可能不诱导同种异型特异性。在一些情况下,aAPC可能不表达HLA。aAPC可被遗传修饰以稳定表达可用于激活和/或刺激的基因。在一些情况下,K562细胞可用于激活。K562细胞也可用于扩增。K562细胞可以是人红白血病细胞系。K562细胞可被工程化以表达目标基因。K562细胞可能不内源性表达HLA I类、II类或CD1d分子,但可表达ICAM-1(CD54)和LFA-3(CD58)。K562可以被工程化以向T细胞传递信号1。例如,K562细胞可以被工程化以表达HLA I类。在一些情况下,K562细胞可以被工程化以表达另外的分子,诸如B7、CD80、CD83、CD86、CD32、CD64、4-1BBL、抗CD3、抗CD3 mAb、抗CD28、抗CD28mAb、CD1d、抗CD2、膜结合的IL-15、膜结合的IL-17、膜结合的IL-21、膜结合的IL-2、截短的CD19或任何组合。在一些情况下,工程化的K562细胞除了表达CD80和CD83外,还可表达膜形式的抗CD3 mAb、克隆OKT3。在一些情况下,工程化的K562细胞除了表达CD80和CD83外,还可表达膜形式的抗CD3 mAb、克隆OKT3、膜形式的抗CD28 mAb。

[0143] 试剂盒

[0144] 本文所述的组合物可以提供在试剂盒中。例如,试剂盒可包含具有可用于构建多种多核苷酸分子的核酸分子库的容器,每种多核苷酸编码TCR链或TCR链的同源对或其部

分。在一些情况下,核酸分子库的每种核酸分子编码TCR链的CDR3。在一些情况下,核酸分子库的每种核酸分子编码TCR链的同源对的第一CDR3和第二CDR3。在一些情况下,核酸分子库的每种核酸分子包含衍生自TCR V基因的序列。在一些情况下,核酸分子库的每个核酸分子包含如本文所述的连接序列。连接序列可以具有与同一核酸分子库中的其他连接序列不同的序列。试剂盒可以包括一个或多个容器,每个容器含有核酸分子库。试剂盒中提供的核酸分子可以是液体形式或干燥形式(例如,冻干形式)。

[0145] 试剂盒还可包含指导使用者使用核酸分子库来构建编码TCR的多种多核苷酸分子的指导材料。

[0146] 试剂盒还可包含至少一种可用于构建核酸分子的反应的试剂(例如,缓冲液、酶、添加剂等)。

[0147] 实施例

[0148] 实施例1.将CDR3-J寡核苷酸库转化为全长、可表达的TCR库

[0149] 本实施例使用3型IIS限制酶来产生粘性末端。此类酶是可商购获得的。在本实施例中,使用产生4-bp 5'突出端的两种酶(例如,BbsI、BbvI、BcoDI、BsaI、BsmBI、FokI等)和产生平末端或3'突出端的一种限制酶(例如,BseRI、BsrDI、BtsI、MlyI等)。使用的最佳酶组可以取决于实际因素(例如,本地可用性、切割效率、星活性)并且可以容易地通过实验进行选择。这里,前两种限制酶称为TIISRE1、TIISRE2,并且最后一种限制酶称为TIISRE3。

[0150] 在本实施例中,配对的CDR3-J寡核苷酸相对于 α 和 β CDR3-J的编码序列以“头对尾”方向合成。换言之, α CDR3-J和 β CDR3-J以相同的5'至3'方向合成。所产生的全长、可表达的TCR多核苷酸也是头对尾方向。配对的CDR3-J寡核苷酸可以其他方向合成,例如头对头和尾对尾。本文所述的方法可与美国临时专利申请号62/718,227、62/725,842、62/732,898、62/818,355和62/823,831(这些申请中的每个均以引用的方式整体并入本文)中所述的方法组合,以设计配对的CDR3-J寡核苷酸并以其他方向获得全长、可表达的TCR多核苷酸。

[0151] 如图1A-图1C所示,配对的CDR3-J寡聚物按5'至3'顺序含有TRBJ、CDR3 β 、TRAJ和CDR3 α 的反向互补序列,以及下面将描述的其他插入结构域。在本文中,符号“*”表示互补性。例如,如果P是指多核苷酸序列,则P*是指P的反向互补序列。此外,在适当时,字母X用于表示A或B。例如,TRXV可用于统指TRAV和TRBV。为清楚起见,在本实施例和图1A-图1C中,TRAJ结构域和TRBJ结构域分别指编码TRAJ区和TRBJ区的部分的多核苷酸序列,它们不包括在CDR3中。

[0152] BCC代表“ β 恒定连接序列”,其功能是与TRBC序列连接。ConB#是特定TRBV序列的连接序列,其中符号#表示TRBV基因的数字ID。类似地,ConA#是特定TRAV序列的连接序列。ICC代表“链间连接序列”,其将用于将TRBV与ConB#连接,以及将TRAC与TRAJ连接。

[0153] ConB#和ConA#结构域可以被密码子多样化(参见实施例2),使得不同TRBV基因的ConX#在核苷酸水平上足够不同,使得ConX#和ConX#*仅在ConX#和ConX#*的数字ID相同时才可以高产率杂交。

[0154] 可以使用常规方法制备48个部分双链的TRAV#_{GL}多核苷酸(IMGT中每个TRAV基因有一个被注释为功能性的)的文库。可以混合所有TRAV#_{GL}多核苷酸以产生TRAV#_{GL}库。GL代表种系。每个TRAV#_{GL}多核苷酸的顶部链含有(1)P2A-3结构域,其编码自切割P2A肽的3'部分,(2)TRAV#_{GL}5结构域,其编码TRAV#的种系序列的5'部分,按此顺序包括L、FR1、CDR1、

FR2、CDR2和ConA#上游的FR3的部分,和(3)ConA#,其编码FR3的最终片段并且是密码子多样化的。每个TRAV#_GL多核苷酸的底部链含有TRAV#_GL5*和P2A-3*。因此,TRAV#_GL多核苷酸具有序列ConA#的3'突出端。可以类似地制备48个TRBV#_GL多核苷酸的文库和库。TRAV#_GL中的P2A-3结构域可以被TRBV#_GL中的T2A-3替换。T2A是另一种自切割肽。

[0155] 可以通过基于芯片的合成制备1,000至500,000个配对的CDR3-J寡核苷酸的库。

[0156] 在步骤(1)中,TRAV#_GL库可以在允许ConA#与ConA#*之间的特异性杂交的温度下与配对的CDR3-J库混合。然后,在步骤(2)中,DNA聚合酶可用于延伸TRAV#_GL的顶部链,并且连接酶可用于将配对的CDR3-J寡聚物与TRAV#_GL的底部链连接。

[0157] BCC包含TIISRE1的识别位点。在步骤(3)中,TIISRE1可用于在BCC处切割,从而在底部链留下4碱基5'突出端。在本实施例中,4个碱基是TRBC1的前4个碱基的反义。在步骤(4)中,此切割产物可以连接到预先制备的含有完整的TRBC1序列和P2A-5结构域的TRBC_P2A-5_SE,并且在TRBC1序列的开始处具有4碱基5'突出端。P2A-5结构域是P2A编码序列的5'末端部分。SE代表粘性末端。此连接产物可以在步骤(5)中进行PCR扩增。

[0158] 在步骤(6)中,可以使用美国临时专利申请号62/718,227、62/725,842、62/732,898、62/818,355和62/823,831中所述的方法,通过P2A-5与P2A-3之间的连接来环化此扩增产物。连接后,P2A-5和P2A-3形成P2A。在本实施例中,ICC含有TIISRE3的识别位点,其在步骤(7)中可用于立即切割底部链上的ConB#*的3'。顶部链上的切割位点不太重要。在步骤(8)中,可以加热此切割产物以分离顶部链和底部链。含有TRBC1的前约20个碱基的引物可用于在底部链上延伸,从而在底部链的3'末端留下单链区。在此链的3'末端的尖端是ConB#*结构域。在步骤(9)中,可以添加TRBV#_GL库,使得TRBV#_GL的顶部链上的ConB#可以与对应的ConB#*杂交。可以加入DNA聚合酶和连接酶以将杂交产物转化为完全双链的DNA。

[0159] ICC的剩余部分还含有TIISRE2的识别位点,其在步骤(10)中可用于切割ICC,留下4碱基5'突出端,其为TRAC的前4个碱基的反义序列。在步骤(11)中,可将预先制备的TRAC_SE连接到5'突出端,从而形成完整的TRAC序列,类似于上述步骤(4)。

[0160] 最终产物可以连接到用于基于CRISPR/TALEN/ZFN的敲入的慢病毒骨架或合适的“同源序列”中。

[0161] 实施例2. 使用人TRAV和TRBV序列测试密码子多样化

[0162] 在本实施例中,提供了基于热力学的算法来设计密码子多样化的ConA#和ConB#序列。所述算法采用MATLAB语言编写。在本算法中使用的一些变量和自定义函数将在下面的“注释”部分中描述,其余部分在代码的注释中描述或者对于本领域技术人员来说不言而喻。一些自定义函数依赖于使用可公开获得的热力学参数(例如,用于碱基对堆叠的 ΔH 和 ΔS)和模型(例如,作为环大小的函数的 ΔS)的DNA杂交的基于热力学的模拟。这些参数和模型已经由John SantaLucia Jr广泛地公布。技术人员可以从头开始或借助于可公开获得的软件包(诸如NUPACK)容易地编写这些函数。所述算法含有两个阶段:初始设计和密码子多样化,分别描述于脚本1和脚本2中。在初始设计中,根据原始TRAV或TRBV序列设计ConA#和ConB#序列。然后计算每个ConX#与每个ConX#*的杂交产率作为基线(图2A和图2B)。图2A示出了在没有进行密码子多样化的情况下根据原始TRAV序列设计的连接序列的杂交产率(ConA#与ConA#*)。图2B示出了在没有进行密码子多样化的情况下根据原始TRBV序列设计的连接序列的杂交产率(ConB#与ConB#*)。在密码子多样化过程中,一些TRXV#_GL的最后约

60个碱基的密码子选择是随机化的,并且选择允许特异性杂交的ConX#序列。接着,计算使用密码子多样化的序列组的每个ConX#与每个ConX#*的杂交产率,以观察密码子多样化是否成功(图3A和图3B)。图3A示出了密码子多样化的连接序列的杂交产率(ConA#与ConA#*)。图3B示出了密码子多样化的连接序列的杂交产率(ConB#与ConB#*)。

[0163] 脚本1:初始设计。

```
clear
fHybTemp = 60; %Hybridization temperature (unit: degree C)
fConcNa = 125; %Sodium ion concentration (unit: mM)
fConcMg = 3; % Magnesium ion concentration (unit: mM)
fConcQB = 5; %Concentration of ConA# or ConB#

%Store the parameters above in struPara.
struPara.fHybTemp = fHybTemp;
struPara.fConcNa = fConcNa;
struPara.fConcMg = fConcMg;
[0164] struPara.fConcQB = fConcQB;

cChain = 'B'; % The value can be A or B to design initial ConA and ConB sequences
respectively.
cFileGeneSeq = sprintf('hsTR%sV_UTR200-L-V_Sorted_FOnly.txt',cChain); % See notes

% Read V gene sequences
fidGene = fopen(cFileGeneSeq);
raGeneInfo = textscan(fidGene,'%s\t%d\t%d\t%s','Headerlines',1);
fclose(fidGene);
```

```

% Initialize the cell array (ra1on1) that stores the initially designed ConA or ConB sequences
iTotalNumOfGene = size(raGeneInfo{1},1);
ra1on1s{1} = -ones(iTotalNumOfGene,2);
ra1on1s{2} = cell(iTotalNumOfGene,4);

for iGeneNum = 1:iTotalNumOfGene
    fprintf('Designing 1on1 for gene #%u.\n',iGeneNum);
    cGeneName = raGeneInfo{1}{iGeneNum};
    cGeneSeq = raGeneInfo{4}{iGeneNum};
    iLStart = raGeneInfo{2}{iGeneNum};
    cCDS = cGeneSeq(iLStart:end);
    cAA = nt2aa(cCDS);
    iPosAAConservedC = find(cAA=='C',1,'last');
    viPosNTConservedC = iPosAAConservedC*3-2:iPosAAConservedC*3;
    vPosOfTinCys(iGeneNum) = viPosNTConservedC(1); % Position of the first nucleotide of
the codon for the conserved Cys at the N terminus of CDR3
    cSA60 = cCDS(viPosNTConservedC(1)-59:viPosNTConservedC(1));
    ra1on1ofThisGene = fun_Design1on1(cSA60,STRUPara); % See notes
    disp(ra1on1ofThisGene{1})
    ra1on1s{1}(iGeneNum,:) = ra1on1ofThisGene{1};
    ra1on1s{2}(iGeneNum,:) = ra1on1ofThisGene{2};

end

%%
cTime = datestr(now);
cTime(cTime==' ')='_';
cTime(cTime==':')='_';

save(['IniDesign1on1_',cChain,num2str(fHybTemp),'_',cTime,'.mat'],'vPosOfTinCys','cChain','
raGeneInfo','ra1on1s','STRUPara');

%% Compute cross hybridization yield
xFracBoundHyb_THyb = -ones(iTotalNumOfGene,iTotalNumOfGene);
for iSimQB = 1:iTotalNumOfGene
    for iDE = 1:iTotalNumOfGene
        if xFracBoundHyb_THyb(iSimQB,iDE) >= 0

```

```

        continue;
    end
    cSimQB = ra1on1s{2}{iSimQB,3}; % cSimQB = sequence of ConA/B
    cDE = ra1on1s{2}{iDE,4}; % cDE = sequence of ConA/B*
    fThisFracBound_THyb = NP_GetBoundFrac(cSimQB,fConcQB,cDE,fConcQB/100,...
        fHybTemp,'Na',fConcNa,'Mg',fConcMg); % See note
    xFracBoundHyb_THyb(iSimQB,iDE) = fThisFracBound_THyb;
[0166] end
end

%%
save('IniDesign.mat','xFracBoundHyb_THyb','xFracBoundHyb_THybMinus5oC');
figure;
colormap(gray);
imagesc(1-xFracBoundHyb_THyb);

```

[0167] 脚本1的注释:

[0168] 文件“hsTRAV_UTR200-L-V_Sorted_FOnly.txt”和“hsTRBV_UTR 200-L-V_Sorted_FOnly.txt”是记录在IMGT数据库中注释为“功能性”的所有TCR V基因的序列的TSV文件。每个文件有4列,第一列是V基因的名称,第4列是从起始密码子上游约200nt开始的V基因cDNA测序序列(L-PART1),第2列是起始密码子的第一个核苷酸的位置。第3列是V基因的第一个核苷酸的位置(例如,在L-PART2之后)。

[0169] 函数“fun_Designlon1”使用两个输入返回ConA或ConB序列:(1)记录TRXV#_GL的最后60个碱基的变量cSA60,以及(2)存储在变量struPara中的用于热力学建模的参数。简言之,所述函数找到满足以下陈述的在cSA60的3'末端结束的cSA60的最短连续子序列(表示为ConX):当混合5nM具有序列ConX的第一DNA寡核苷酸和0.05nM具有序列ConX*的第二DNA寡核苷酸时,预计在分别由struPara.fHybTemp、struPara.fConcNa和struPara.fConcMg定义的温度、钠离子浓度和镁离子浓度下,超过97%的第二寡核苷酸与第一寡核苷酸结合。此函数的输出(ralon1ofThisGene)是具有两个单元的单元阵列,第一单元,ralon1ofThisGene{1}是1x2向量,其中ralon1ofThisGene{1}(1)是本实施例中未使用的输出,并且ralon1ofThisGene{1}(2)是ConX的第一个碱基在cSA60上的位置。ralon1ofThisGene{2}是1x4单元阵列,其中ralon1ofThisGene{2}{1}和ralon1ofThisGene{2}{2}在本实施例中未使用,ralon1ofThisGene{2}{3}是ConX的序列,并且ralon1ofThisGene{2}{4}是ConX*的序列。本领域技术人员可以如上所述地编写此函数。

[0170] 当5nM(由fConcQB所记录的)第二核苷酸和0.05nM(由fConcQB/100所记录的)第一核苷酸在60℃(由fHybTemp所记录的)下混合并在125mM Na⁺(由fConcNa所记录的)和5mM Mg⁺⁺(由fConcMg所记录的)存在下达到平衡时,函数“NP_GetBoundFrac”返回与具有序列ConX的第二DNA寡核苷酸结合的具有序列ConX*的第一DNA寡核苷酸的分数。

[0171] 当0.05nM的ConX#*与5nM的ConX#在上述条件下混合时,由此脚本产生的图像示出

了如何预测与ConX#结合的ConX#*的分数多少的灰度热图。如图2A和图2B所示,特别是对于TRBV (图2B),存在大量交叉结合 (例如,错误连接)。

[0172] 脚本2. 密码子多样化

```
clear;
load('IniDesign.mat');
fConcQB = struPara.fConcQB;
fConcNa = struPara.fConcNa;
fConcMg = struPara.fConcMg;
fHybTemp = struPara.fHybTemp;
fCodonFreqThreshold = 0.15; % Lowest allowed codon frequency
fSSThreshold = 0.6;
fCrossAssemblyThreshold = 0.02; % Lowest allowed level of mis-connection
[0173] iMaxLengthDE = 35; % Maximum allowed length of ConX
struParaDiv.fCodonFreqThreshold = fCodonFreqThreshold;
struParaDiv.fSSThreshold = fSSThreshold;
struParaDiv.fCrossAssemblyThreshold = fCrossAssemblyThreshold;
struParaDiv.iMaxLengthDE = iMaxLengthDE;

%% Initiate
ra1on1NewDesign = cell(1,2);
iTotalNumOfGene = size(ra1on1s{1},1);
```

```

xCrossPrimeAlreadyDesigned_Hyb_THyb = -ones(iTotalNumOfGene);

%%
for iGeneToModify = 1:iTotalNumOfGene
    cSA60 = ra1on1s{2}{iGeneToModify,1};
    cAAInFrame = nt2aa(cSA60(3:59));
    fprintf('%s\n',cSA60);

    bHaveYouTriedInitialDesign = false;
    while(1)
        if ~bHaveYouTriedInitialDesign
            ra1on1ofThisGene{1} = ra1on1s{1}{iGeneToModify,:};
            ra1on1ofThisGene{2} = ra1on1s{2}{iGeneToModify,:};
            bHaveYouTriedInitialDesign = true;
        else %Randomize
            cTrialMiddle57 = fun_aa2nt(cAAInFrame,raCodonTable,fCodonFreqThreshold); %
See note
            cAAToMakeSure = nt2aa(cTrialMiddle57,'AlternativeStartCodons','false');
[0174] if ~strcmpi(cAAInFrame,cAAToMakeSure)
                error('something is wrong');
            end
            cTrialSA60 = [cSA60(1:2),cTrialMiddle57,'T'];
            fprintf('%s\n',cTrialSA60);
            ra1on1ofThisGene = fun_Design1on1(cTrialSA60,STRUPara);
        end
        iStartPosBM = ra1on1ofThisGene{1}(2);
        cTrialSimQB = ra1on1ofThisGene{2}{3};
        cTrialDE = ra1on1ofThisGene{2}{4};

        % Check DE length
        iLeftestStartPosBM = 60 - iMaxLengthDE + 1;
        if iStartPosBM < iLeftestStartPosBM
            fprintf('DE too long\n');
            continue;
        end

    %%

```

```

vRow_Thyb = -ones(1,iGeneToModify);
vColumn_Thyb = -ones(iGeneToModify,1);

for iDE = 1:iGeneToModify
    if iDE ~= iGeneToModify
        cDE = ra1on1NewDesign{2}{iDE,4};
    else
        cDE = cTrialDE;
    end
    fFracHyb_Thyb = NP_GetBoundFrac(cTrialSimQB,fConcQB,cDE,fConcQB/100,...
        fHybTemp,'Na',fConcNa,'Mg',fConcMg);
    vRow_Thyb(iDE) = fFracHyb_Thyb;
end
for iQB = 1:iGeneToModify
    if iQB ~= iGeneToModify
        cExtSimQB = ra1on1NewDesign{2}{iQB,2};
    else
[0175]     cExtSimQB = cTrialExtSimQB;
    end
    fFracHyb_Thyb =
NP_GetBoundFrac(cExtSimQB,fConcQB,cTrialDE,fConcQB/100,...
        fHybTemp,'Na',fConcNa,'Mg',fConcMg);
    vColumn_Thyb(iQB) = fFracHyb_Thyb;

end

vRowToUse = vRow_Thyb;
vColumnToUse = vColumn_Thyb;

if vRowToUse(end) >= 0.5 && sum(vRowToUse(1:end-1))<fCrossAssemblyThreshold
&& ...
    vColumnToUse(end) >= 0.5 && sum(vColumnToUse(1:end-
1))<fCrossAssemblyThreshold
    ra1on1NewDesign{1}(iGeneToModify,:) = ra1on1ofThisGene{1};
    ra1on1NewDesign{2}(iGeneToModify,:) = ra1on1ofThisGene{2};

```

```

        xCrossPrimeAlreadyDesigned_Hyb_THyb(iGeneToModify,1:iGeneToModify) =
vRow_THyb;
        xCrossPrimeAlreadyDesigned_Hyb_THyb(1:iGeneToModify,iGeneToModify) =
vColumn_THyb;

        break;
    end
[0176]
    end
end
%%
figure;
colormap(gray)
imagesc(1-xCrossPrimeAlreadyDesigned_Hyb_THyb);

```

[0177] 脚本2的注释：

[0178] 使用由输入raCodonTable提供的密码子表信息和由输入fCodonFreqThreshold提供的最低允许密码子频率,函数“fun_aa2nt”返回与输入序列cAAInFrame编码相同的多肽的多核苷酸序列。

[0179] 当0.05nM的ConX##与5nM的ConX#在密码子多样化后在上述条件下混合时,由此脚本产生的图像示出了预测与ConX#结合的ConX##的分数多少的灰度热图。如图3A和图3B所示,预测仅特异性杂交会大量发生。因此,本实施例表明密码子多样化方案是可行的,并示出了如何获得密码子多样化的ConA和ConB序列。

[0180] 实施例3. 衍生自小鼠TRAV和TRBV基因的连接序列

[0181] 本实施例提供了衍生自小鼠TRAV和TRBV基因的密码子多样化的连接序列。与上述实施例相似,ConA是特定TRAV序列的连接序列,ConB是特定TRBV序列的连接序列。使用与实施例2中所述相同的方法进行密码子多样化。表1示出了衍生自小鼠TRAV基因的密码子多样化的连接序列。表2示出了衍生自小鼠TRBV基因的密码子多样化的连接序列。在表1和表2中,在第一列中示出了每个V基因的基因名称和登录号,并且在第二列中示出了对应的连接序列。

[0182] 表1. 衍生自小鼠TRAV基因的连接序列

[0183]

SEQ ID NO :	小鼠 TRAV 基因的连接序列	连接序列
1	>Trav1 ENSMUST00000103567.5 的 ConA	TCAAAGACTCTGCCTCATACCTCT
2	>Trav2 ENSMUST00000196939.1 的 ConA	CTGAGAGACGCGAGCTGTGTATTACT
3	>Trav3-1 ENSMUST00000103569.2 的 ConA	GGGGACTCAGCCGTGTACTTCT
4	>Trav3-3 ENSMUST00000181768.2 的 ConA	GGTGACTCCGCAGCCTATTTCT
5	>Trav3-4 ENSMUST00000103670.3 的 ConA	TGGGGATAGCGCAGTCTATTTCT
6	>Trav3d-3 ENSMUST00000196023.1 的 ConA	CCGGAGACAGCGCAGTTTATTTT
7	>Trav3n-3 ENSMUST00000197557.1 的 ConA	GGTGACAGCGCCGTCTATTTTT
8	>Trav4-2 ENSMUST00000103637.5 的 ConA	TGGAGGACTCAGGCACTTACTTCT
9	>Trav4-3 ENSMUST00000103655.2 的 ConA	TGGAGGACTCTGGGACATACTTTT
10	>Trav4-4-dv10 ENSMUST00000103663.5 的 ConA	TGGAGGACTCTGGCACCTATTTTT
11	>Trav4d-3 ENSMUST00000103592.1 的 ConA	ACTCGAGGATTCCGGTACTTATTCT
12	>Trav4d-4 ENSMUST00000103600.2 的 ConA	GAAGACTCCGGGACCTACTTTT
13	>Trav4n-3 ENSMUST00000103618.1 的 ConA	GCTGGAGGATTCCGGAACCTATTCT
14	>Trav4n-4 ENSMUST00000103627.2 的 ConA	CTCGAAGATAGCGGCACATATTTT
15	>Trav5-1 ENSMUST00000103570.1 的 ConA	GCCTGGTGATAGCGCAATATACTTCT
16	>Trav5d-4 ENSMUST00000179701.1 的 ConA	TGGCGACTCTGCAATGTACTTCT
17	>Trav5n-4 ENSMUST00000179997.1 的 ConA	CCCGGAGACTCTGCTATGTATTTT
18	>Trav6-1 ENSMUST00000103571.1 的 ConA	GGAATCCGATAGCGCAGTCTATTACT
19	>Trav6-2 ENSMUST00000198058.1 的 ConA	TCCGACAGCGCTGTCTACTACT

[0184]

	1 的 ConA	
20	>Trav6-3 ENSMUST00000180549. 2 的 ConA	AAGAGATTGATAGCGCTGTTTAC TACT
21	>Trav6-4 ENSMUST00000184650. 1 的 ConA	AGGAATCTGATTCCGCAGTCTAT TTTT
22	>Trav6-5 ENSMUST00000181210. 2 的 ConA	AGAATCTGATAGCGCCGTTTATT ATT
23	>Trav6-6 ENSMUST00000103584. 3 的 ConA	GTCCGACTCCGCAGTCTACTACT
24	>Trav6-7-dv9 ENSMUST00000103 638.5 的 ConA	AGGAGTCTGATTCTGCAGTCTAC TATT
25	>Trav6d-3 ENSMUST00000181483 .2 的 ConA	CCAAGAAATAGATTCCGCAGTCT ACTACT
26	>Trav6d-4 ENSMUST00000180717 .2 的 ConA	GTCTGACAGCGCAGTCTACTTCT
27	>Trav6d-5 ENSMUST00000180687 .2 的 ConA	AGGAAAGCGATTCTGCAGTCTA TTACT
28	>Trav6d-6 ENSMUST00000197754 .1 的 ConA	AAGAGTCTGACTCCGCAGTTTAT TATT
29	>Trav6d-7 ENSMUST00000178650 .2 的 ConA	AGAATCCGACTCTGCAGTTTACT ATT
30	>Trav6n-5 ENSMUST00000103611. 1 的 ConA	GAGTCTGATAGCGCTGTGTACTA CT
31	>Trav6n-6 ENSMUST00000181793 .2 的 ConA	GAATCTGACTCTGCCGTTTACTA TT
32	>Trav6n-7 ENSMUST00000179607 .2 的 ConA	AGTCCGACTCTGCTGTGTACTAC T
33	>Trav7-1 ENSMUST00000198019. 1 的 ConA	CCATCTGATTCCGCACTGTATTT CT
34	>TRAV7-2 AC004407 的 ConA	ACTTTCTGATAGCGCTCTCTATT TTT
35	>Trav7-3 ENSMUST00000177622. 3 的 ConA	CCTTCTGATTCTGCACTGTACCT GT
36	>Trav7-4 ENSMUST00000181728. 2 的 ConA	CCAAGCGATTCTGCACTGTATTT TT
37	>Trav7-5 ENSMUST00000200609. 1 的 ConA	CCTCTGACTCTGCAGTCTACCTC T
38	>Trav7-6 ENSMUST00000103641. 5 的 ConA	CCCAGCGACTCTGCAGTTTATCT CT
39	>Trav7d-2 ENSMUST00000200127 .1 的 ConA	CCGACAGCGCACTCTACCTGT
40	>Trav7d-3 ENSMUST00000179789 .3 的 ConA	TTCCGACTCTGCACTGTATCTGT
41	>Trav7d-4 ENSMUST00000178768 .3 的 ConA	TCCGATAGCGCCCTGTATTTCT
42	>Trav7d-5 ENSMUST00000197128	CTCCGATTCCGCACTCTATCTCT

[0185]

	.1 的 ConA	
43	>Trav7D-6 ENSMUST00000196756 .1 的 ConA	CCTCCGATAGCGCTGTTTATCTC T
44	>Trav7n-4 ENSMUST00000103609 .1 的 ConA	GCGACAGCGCCCTGTACTTTT
45	>Trav7n-5 ENSMUST00000199753 .1 的 ConA	CCCTCTGATAGCGCACTGTATCT CT
46	>Trav7n-6 ENSMUST00000178100 .2 的 ConA	CTTCTGACAGCGCTGTGTATCTG T
47	>Trav8-1 ENSMUST00000103643. 3 的 ConA	GCGAGGACACAGCTGTTTACTT TT
48	>TRAV8-2 AC004096 的 ConA	AGTGCGAAGATACAGCAGTTTA CTTCT
49	>Trav8d-1 ENSMUST00000103580 .3 的 ConA	CGGTGTGAGGATACTGCTGTTTA TTTCT
50	>Trav8d-2 ENSMUST00000198439 .4 的 ConA	CGAAGATACCGCCGTCTACTTTT
51	>Trav8n-2 ENSMUST00000103632 .3 的 ConA	GGCAACTGACACAGCAGTCTAC TTTT
52	>Trav9-1 ENSMUST00000103581. 5 的 ConA	GAGCGATTCTGCCGTTTACTTCT
53	>Trav9-2 ENSMUST00000103654. 2 的 ConA	TCCGATTCCGCCGTGTATTTTT
54	>Trav9-4 ENSMUST00000103662. 5 的 ConA	TTGGTCTGATTCTGCAGTTTACT TTT
55	>Trav9d-1 ENSMUST00000178426 .3 的 ConA	GGTCTGATTCCGCTGTCTACTTT T
56	>Trav9d-2 ENSMUST00000199746 .4 的 ConA	CTGGTCTGACTCTGCTGTTTATT TTT
57	>Trav9d-3 ENSMUST00000178252 .2 的 ConA	GGTCCGACTGGGCAGTCTATTTT T
58	>Trav9d-4 ENSMUST00000200548 .1 的 ConA	TGGTCTGATTCTGCCGTCTATTT CT
59	>Trav9n-2 ENSMUST00000198913 .4 的 ConA	AGCGACTCTGCCGTGTATTTCT
60	>Trav9n-3 ENSMUST00000177705 .2 的 ConA	GAGCGATTGGGCAGTCTACTTTT
61	>Trav9n-4 ENSMUST00000103626 .2 的 ConA	TGGTCCGATTCTGCTGTCTATTT TT
62	>Trav10 ENSMUST00000103583.4 的 ConA	AGCCTGAAGATTCAGCCATCTAC TTCT
63	>Trav10d ENSMUST00000103646. 4 的 ConA	CCCGAGGACTCTGCTATTTACTT CT
64	>Trav10n ENSMUST00000103612. 1 的 ConA	ACAGCCAGAAGATTCTGCAATAT ACTTCT
65	>Trav11 ENSMUST00000103585.3	GCTCGATGACACAGCTACATACA

[0186]

	的 ConA	TCT
66	>TRAV11D AC004101 的 ConA	TCCTGGATGATACTGCAACATAC ATAT
67	>Trav12-1 ENSMUST00000200115. 1 的 ConA	CTCTCTGACTCTGCACTGTACTA CT
68	>Trav12-2 ENSMUST00000180972 .2 的 ConA	ACTGTCTGACTCTGCACTCTATT ACT
69	>Trav12-3 ENSMUST00000103657 .5 的 ConA	CTGTCCGATTCTGCACTCTACTA CT
70	>Trav12d-1 ENSMUST0000018136 0.2 的 ConA	AACTGTCTGATTCTGCTCTGTAC TATT
71	>Trav12d-2 ENSMUST0000010359 3.2 的 ConA	TCCGACTCCGCTCTGTATTTTT
72	>Trav12d-3 ENSMUST0000017770 3.2 的 ConA	AGCGACTCTGCCCTCTACTACT
73	>Trav12n-1 ENSMUST0000019868 2.1 的 ConA	TCTCTGACTCCGCTCTCTACTAC T
74	>Trav12n-2 ENSMUST0000010361 9.2 的 ConA	TCTCTGATTCTGCCCTCTACTTTT
75	>Trav12n-3 ENSMUST0000017958 3.2 的 ConA	GCTCTCCGATTCTGCTCTGTATTA TT
76	>Trav13-1 ENSMUST00000103651 .3 的 ConA	GACAACAGACTCAGGCACTTAT CTCT
77	>Trav13-2 ENSMUST00000103658 .3 的 ConA	AACAACAGACTCTGGCACATATT TTT
78	>TRAV13-3 AC003995 的 ConA	CACTGATAGCGGAACCTACCTCT
79	>Trav13-4-dv7 ENSMUST0000018 0380.2 的 ConA	TGACAGCGGCACCTACCTGT
80	>Trav13-5 ENSMUST00000103671 .3 的 ConA	ACTACAGATTCCGGCACTTACTT CT
81	>Trav13d-1 ENSMUST0000010358 8.3 的 ConA	GCCAGATAACTGATTCTGGTACT TACCTGT
82	>Trav13d-2 ENSMUST0000019795 4.1 的 ConA	ACAACAGAGCGGAACATATC TCT
83	>Trav13d-3 ENSMUST0000017951 2.2 的 ConA	AAATAACAGATAGCGGTACATAC CTGT
84	>Trav13d-4 ENSMUST0000019607 9.1 的 ConA	CCACAGATTCTGGCACCTACTTC T
85	>Trav13n-1 ENSMUST0000019835 9.1 的 ConA	ACTGACTCCGGAACCTACCTCT
86	>Trav13n-2 ENSMUST0000019694 1.1 的 ConA	ACCGACTCTGGCACTTACCTGT
87	>Trav13n-3 ENSMUST0000017958 0.2 的 ConA	AATCACAGACTCTGGAACCTATC TGT
88	>Trav13n-4 ENSMUST0000019610 5.1 的 ConA	CCAAATTACCGATTCTGGTACAT ACCTCT

[0187]

89	>Trav14-1 ENSMUST00000198297 .1 的 ConA	CGGAGATAGCGCCACATACTTTT
90	>Trav14-2 ENSMUST00000179267 .3 的 ConA	AACCTGGAGATTCTGCAACATAT TTCT
91	>Trav14-3 ENSMUST00000103589 .5 的 ConA	CTGGGGACTCTGCAACTTACTTC T
92	>Trav14d-1 ENSMUST0000018103 8.2 的 ConA	CCTGGAGACTCAGCTACCTACTT CT
93	>Trav14d-2 ENSMUST0000019680 2.1 的 ConA	CCGGGGATAGCGCTACTTATTTT T
94	>Trav14d-3-dv8 ENSMUST000001 03608.3 的 ConA	CCTGGAGATTCCGCAACTTACTT TT
95	>Trav14n-1 ENSMUST0000017757 8.1 的 ConA	CCAGGGGATTCTGCTACCTATTT TT
96	>Trav14n-2 ENSMUST0000019761 4.1 的 ConA	CCCGGAGATTCTGCCACTTATTT CT
97	>Trav14n-3 ENSMUST0000010365 2.4 的 ConA	CTGGCGACAGCGCTACTTATTTT T
98	>Trav15-1-dv6-1 ENSMUST000001 03653.2 的 ConA	AACCAGACGATTCGGGAAAGTA TTTCT
99	>Trav15-2-dv6-2 ENSMUST000001 03660.3 的 ConA	CAGAGGATTCAGGGACGTACTT CT
100	>Trav15d-1-dv6d-1 ENSMUST0000 0103616.4 的 ConA	CCAGACGACTCCGGAAAGTACT TTT
101	>Trav15d-2-dv6d-2 ENSMUST0000 0199800.1 的 ConA	CCGAGGACTCCGGTACATACTTC T
102	>Trav15n-1 ENSMUST0000010359 0.3 的 ConA	AACCCGATGACTCTGGTAAGTAT TTTT
103	>Trav15n-2 ENSMUST0000019911 2.1 的 ConA	GCCAGAAGACTCCGGTACATATT TTT
104	>Trav16 ENSMUST00000103667.5 的 ConA	TCAAATTGAAGATTCTGCAGTCT ACTTTT
105	>Trav16d-dv11 ENSMUST0000010 3606.1 的 ConA	GATTGAGGACTCGGCAGTATATT TCT
106	>Trav16n ENSMUST00000199280. 1 的 ConA	AAATCGAAGACTCTGCAGTTTA CTTTT
107	>Trav17 ENSMUST00000103672.8 的 ConA	GAGCGACTCAGCCAAGTACTTC T
108	>Trav18 ENSMUST00000103673.1 0 的 ConA	AGGGGATGCTGGGATCTACTTTT
109	>Trav19 ENSMUST00000103674.5 的 ConA	CCCGAAGATACAGCTGTCTACCT GT
110	>Trav21-dv12 ENSMUST00000180 938.2 的 ConA	AGGGACGCAGCAGTCTATCATT
111	>Trav23 ENSMUST00000199137.1 的 ConA	GCCACTCTGCCATCTACTTCTGT

[0188] 表2.衍生自小鼠TRBV基因的连接序列

[0189]

SEQ ID NO :	小鼠 TRBV 基因的连接序列	连接序列
112	>Trbv1 ENSMUST00000103262.2 的 ConB	GGCGCACACTGTACTGCACAT
113	>Trbv2 ENSMUST00000103263.2 的 ConB	TGATGACTCGGCCACATACTTCT
114	>Trbv3 AE000663 的 ConB	TGGAGGACTCAGCTGTGTACTTCT
115	>Trbv4 ENSMUST00000103265.4 的 ConB	ACCAGAAGATAGCGCAGTTTATCTGT
116	>Trbv5 ENSMUST00000103266.2 的 ConB	ATCCAGAAGACTCAGCTGTCTATT TTT
117	>Trbv12-1 M15614 的 ConB	TCGAAGATAGCGCCATGTACTTTT
118	>Trbv12-2 M15613 的 ConB	ACTGGAAGATAGCGCTGTGTATTT CT
119	>Trbv13-1 ENSMUST00000194399.1 的 ConB	AAGCCAGACCAGCCTCTATTTTTT
120	>Trbv13-2 ENSMUST00000103270.3 的 ConB	CCCCTCTCAGACATCAGTGTACTT CT
121	>Trbv13-3 ENSMUST00000103271.1 的 ConB	GCCAGACCGCCGTGTATTTCT
122	>Trbv14 ENSMUST00000103272.3 的 ConB	GGCGACACAGCCACCTATCTCT
123	>Trbv15 ENSMUST00000103273.2 的 ConB	GCCTAAAGACAGCGCTGTTTATCT CT
124	>Trbv16 ENSMUST00000103274.3 的 ConB	CCAGGACTCAGCGGTGTATCTTT
125	>Trbv17 ENSMUST00000103275.3 的 ConB	GCCTAGAGTATTCTGCCATGTACC TCT
126	>Trbv19 ENSMUST00000103276.2 的 ConB	AAAAATGAGATGGCAGTCTTCCTC T
127	>Trbv20 ENSMUST00000103277.1 的 ConB	CGAGGATAGGGGCCTGTATCTCT
128	>Trbv23 ENSMUST00000193997.5 的 ConB	GCAGAAGACTCAGCACTGTACTT GT
129	>Trbv24 IMGT 的 ConB	GACGACTCAGCACTGTACCTCT
130	>Trbv26 ENSMUST00000193064.1 的 ConB	GGGACTCCGCACTCTATCTCT
131	>Trbv29 ENSMUST00000103281.2 的 ConB	AAACAAACCAGACATCTGTGTAC TTCT
132	>Trbv30 ENSMUST00000103282.2 的 ConB	GGCCTGGAGACAGCAGTATCTATT TCT

[0190]

133	>Trbv31 ENSMUST00000193003.1 的 ConB	TCAGCCATAGCGGTTTTTACCTCT
-----	-------------------------------------	--------------------------

[0191] 在初始设计中,根据原始TRAV或TRBV序列设计ConA#和ConB#序列。如本文所用,符号#表示TRAV或TRBV基因的数字ID。然后计算每个ConX#与每个ConX#*的杂交产率以作为基线(图10A和图10B)。图10A示出了在没有进行密码子多样化的情况下根据原始TRAV序列设计的连接序列的杂交产率(ConA#与ConA#*)。图10B示出了在没有进行密码子多样化的情况下根据原始TRBV序列设计的连接序列的杂交产率(ConB#与ConB#*)。在密码子多样化过程中,一些TRXV#_GL的最后约60个碱基的密码子选择是随机化的,并且选择允许特异性杂交的ConX#序列。接着,计算使用密码子多样化的序列组的每个ConX#与每个ConX#*的杂交产率,以观察密码子多样化是否成功(图11A和图11B)。图11A示出了密码子多样化的连接序列的杂交产率(ConA#与ConA#*)。图11B示出了密码子多样化的连接序列的杂交产率(ConB#与ConB#*)。

[0192] 实施例4. 具有任意序列的连接序列

[0193] 表3提供了根据图7中所述的方案可用作连接CDR3-J多核苷酸和指定的V基因种系多核苷酸的连接序列的任意序列。

[0194] 表3. 具有任意序列的连接序列

[0195]	SE Q ID NO:	连接序列	SE Q ID NO:	连接序列
	134	CCGGGATTTTGTGACTCATC	209	CTTGTCCACTAAACGCAAC G
	135	GAGGATCGTATGTTTCGCAC	210	CGGGTATCACTGGGTAATGA
	136	CTTGTGTGCACTTACCGTAC	211	GGAACAGAGACCAATCCAG T
	137	TGATGCATCTCCAGTACAGG	212	GTGGGCATCCGAAATTTC G
	138	CTTCTGTGTGTACCTCGACA	213	CGCGACGACATTACCAATA G
	139	CTTGCAATCCTTTACCGTGC	214	CGTCGGAATATGCTCTCAGA

[0196]

140	GCAAGTGTGGAAAATGACC C	215	GATGCAGATCAATGAGTGG C
141	GAGTCTAGTCTCACAACCC A	216	ACTGCTTACAAGTGTCCAC G
142	GAAATGTTGAGGACTCCAC G	217	ACTGTATGCAAGCTAGTCCC
143	CCTAACAGATGCTACGTGG A	218	AGATCTCCCAAAAGTGTCC G
144	GTAGGTCCACACAGATTCC A	219	TTCCAGAACCATGTGATCCC
145	GCCAGTCACAGCAAATACA C	220	GCCTTGTCTTTCAACCTCTG
146	CCGCTACCAGTATGTACCTT	221	GATACGGATCTTCACATGCG
147	ACTGTGTTCTTGTCTTCCG	222	CGCTCATCTAGGTTGGACTA
148	TGAATGCATCTACGGTACCG	223	CGCGTTCAGATTCCAAACA G
149	GCGCTTATCAATCTTGCTCG	224	GCCTGGTTACACATGCTATC
150	CGGTCAATTCAGTAGCCACT	225	GCAAAGGTCCTACAGGTTT C
151	GGACACATGTACACTAGCC A	226	GGCTTTCCATGTCTATGCTC
152	TGGGAGCTCTACGAAAATC C	227	GCGACATTAGCAGAGTAGG T
153	GTTCTCGAGATCGTCACAC A	228	CTCGCCATACTATCTGCATG
154	CTTCTGCATTCGATCCTTGC	229	CTACTGAACACTTGGCAAG C
155	GCAGAGTTGTGTGATTGGA G	230	CTGTTCAATTCCTGTGCGAG
156	CCCATCAATTCGGAACCATC	231	CACTGAGATGGAATTTGGC G
157	ATCGTAACCCAAGTCTGTGG	232	GTCCATCACAACCTTCCACTG
158	GGCGAAATGATCCCTGAAT G	233	GGCTCAGTCTACTTTGCTTC
159	AGTGCTCAGAACTTTCAGG C	234	GTTAGCTTCCGACACAATG G
160	GATCGTTAACTCTTTGGGCG	235	CCGTGACACACTTTCATCTC
161	ACACGAGGATTGCTGTAGA G	236	CCGGGATGTCATTATGAGCA
162	TTCTACCACATTGTCTCGGG	237	GAGTGTCTTACGAGATCAG A
163	GAATGGCTAAACTGTGTCCC	238	TAACGTCTCTCTGAGTGTG G
164	CGGACTGTACGAGAAACTG A	239	ACCCTAGACAAGAGACACC T
165	CTTGCGACAACTACTCCTG	240	TGCTCAGTACTCTTCATCGC
166	CCGTTTTACTTTGTCGCCAC	241	AGCTCAATCATGGCTATCGG
167	TGGATGATATCACTTCGGCG	242	CTACACATTGCATCCAACCC
168	CCAACCTCTATATGTGCCCA	243	ACTTGTCGAATAGCTCAGG

[0197]

			C
169	TGAACAGGTATGCTCCAGA G	244	GTCTACCCTGAGAACCAGT T
170	TTGTGGATATCGTCTGGTCC	245	CAGCAACAACCTACCTTAG C
171	CTGTGGAACCTCGACTCTTGT	246	CGATTTGTTGGTACGTGTCC
172	CTCCAGGATGCACAAATTC G	247	GCCATTTCCTTTGTACCTGC
173	GGCTCATGACAAAACACAG G	248	GGCCAATAGAGAGACCACA A
174	CCGAATCCGAAAACAACAC G	249	CGGAGTCACATGGGTAGAA T
175	AGACCTAACACTGTGATGC G	250	CCCAGTACATTTGTCTGGTTG
176	CCTGGGTGAGCATAAACTTC	251	CCCACTAGCTGCTACTCAA A
177	GAGTCTTGGACGAACAAAG G	252	GGTGTGCGTCAAAGTAGA C
178	CACGTACCCATCATGTTCTG	253	TACTCCAGCTCTTACTGTGC
179	CCGTGTTAGTCAAGTGTGTG	254	GGATGAGCAGTCAACAGTT C
180	CTGGTGGCATAAATGGAAC G	255	TCAGGATCGATCAGTTGTCC
181	TGGATGTGGGTATCAATGGG	256	CCTCTCTTTTGTGCGGAAAC
182	TGTGGCTAACGTAGGACAA G	257	TGCCTAGGATTTCGAGAAC G
183	CCCTCGTTGTGAAAATGTGC	258	GGCATTGTCCTTAACTTCGC
184	TCGTCATAGGTCAGCTTACG	259	TGCATCTAACTACGATGGGC
185	CCTGATGACCTCTATGCCAA	260	CCCTAGTAGCCACACAACA T
186	CGGCAAGAATGAATAGGGT G	261	GTGCCATGAATCATCGTCTC
187	GTGCTATTGGTGGGAAATGG	262	CGCTCTGATGAAAGCTCCAT
188	GCCATGTTTGCTTACTGACG	263	CCAGCCATAGTGCATATCCT
189	CGTTGTGGCATTTCATTAGCG	264	GAGATTGTCATGTGGTCGA C
190	GCGGTAGGATTGGATCTCAT	265	CCGCAGTCTAACAGGAAAT C
191	CCTCGCAAAGCTGTTATGAC	266	CGCTTCGACTGAACCTTATG
192	GCCTTCATGTTATTGGACGC	267	CGATGCGACCAATAGAAGT G
193	AGCTGTAGTGTTCTTGAGGC	268	GCCCTTGGTACGACATATTG
194	GGTAGTGTTCGTGTGACATG	269	CAGTGATTTAGGTGACGCA G
195	CGCGGCATATGTTTCATATCC	270	GGCATGGAAGAGGTAGTTT C
196	GAGACTGGATCATGCAACA G	271	CCGATCGTATTCTGTGTCCA
197	CACAACTTCTCTGGACTCC	272	CTAAGTCAAGCACATGGGA

[0198]		A		C
	198	CGACCATGATCTGTATGCGT	273	GATCCACACTCAATCTCCTG
	199	GGTGTGACTCTTGTTCCTG	274	CCTTGTACATGCTGGTATC
	200	ACGTACATACAAGTCTGGC G	275	CGCGATTGTGGTTAATAGGC
	201	CCTCAAGGATTCACTCGCA A	276	GTAGGCAAAGTTCACCACA C
	202	CTGTATAGGATGTCCACGCA	277	GCCACGAATCGAACAAGTA C
	203	GCCTGTGATTGGTAAATGCG	278	TTGAGATCTCGATGAGCAC G
	204	CGCACTCGTAGCATCTAGAA	279	GGGCCAAGATCTATTCGTCA
	205	CGATTTGTTGTCCCTAGCTG	280	GTGGCTATAGGTATGTCCGA
	206	CCCACTTCATCTGACTCTGA	281	CCACACTTTCTGCATTCGAC
	207	CGGCATTGTACAGGTGTTAC	282	CGGCATCTCAAAGCACATA C
	208	TCTCCTATTTCCCTGAACGG	283	CGTCCACAAATTTACTGCCC

[0199] 实施例5.使用下一代测序表征组装的TCR基因

[0200] 使用本文所述的方法(例如具有一些修改的实施例1)制备编码配对的TCR的核酸序列库。编码天然配对的TCR的参考序列从可公众获得的文库获得。选择553个参考序列以在本实施例中说明。在本实施例中,分别合成了编码CDR3-J α (或CDR3-J α 片段)的核酸序列和编码CDR3-J β (或CDR3-J β 片段)的核酸序列。或者,配对的CDR3-J α 和CDR3-J β 可在一个片段上一起合成。

[0201] 合成553个CDR3-J α 片段和553个CDR3-J β 片段并将其连接(例如,通过连接、重叠PCR等)在一起以产生配对CDR3-J α -CDR3-J β 片段库。为了确保CDR3-J α 与天然配对的CDR3-J β 连接,在每个CDR3-J α 上合成任意连接序列,并设计所述任意连接序列以使其与CDR3-J α 片段库中的其他任意连接序列的交叉杂交最小化。在天然配对的CDR3-J β 上合成任意连接序列的互补序列。接着,将TRAV片段库(根据参考序列预合成)连接到配对的CDR3-J α -CDR3-J β 片段以产生TRAV-CDR3-J α -CDR3-J β 片段库,每个片段包含与其同源CDR3-J α 连接的TRAV序列。接着,将TRBC1序列附加到TRAV-CDR3-J α -CDR3-J β 片段的下游以形成TRAV-CDR3-J α -CDR3-J β -TRBC1片段。这些片段通过紧接在CDR3-J β 的上游切割而环化和再线性化,从而形成CDR3-J β -TRBC1-TRAV-CDR3-J α 片段。以框内自切割P2A序列连接TRBC1和TRAV的方式设计TRBC1和TRAV片段。接着,将TRBV片段库(根据参考序列预合成)与CDR3-J β -TRBC1-TRAV-CDR3-J α 片段连接以产生TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α ,对其进行下一代测序(NGS)以评估克隆的丰度和克隆的连接准确度。这里,NGS数据中的每个克隆都是指独特序列。由于在本实施例中使用了553个参考序列,因此在NGS数据中总共有553个克隆。对于本文所述的数据分析,CDR3-J α 序列用于代表克隆。

[0202] 图12示出了产生配对的CDR3-J α -CDR3-J β 片段后每个克隆的准确度和丰度。每个数据点对应于CDR3-J α -CDR3-J β 片段的克隆。准确度是指与同源CDR3-J β 片段连接的CDR3-J α 片段的分数。对于每个CDR3-J α ,准确度可以通过正确连接的CDR3-J β 片段的数目除以连接的CDR3-J β 片段的总数来计算。丰度是指每个克隆在总克隆库中的分数,其可以通过所述克隆的总读长数除以所有克隆的总读长数来计算。数据表明553个克隆中的497个具有高于

95%的准确度和高于0.1/553的丰度,如方框中所示。

[0203] 图13示出了产生TRAV-CDR3-J α -CDR3-J β 片段后每个克隆的准确度和丰度。每个数据点对应于TRAV-CDR3-J α -CDR3-J β 片段的克隆。准确度是指与同源TRAV片段连接的CDR3-J α -CDR3-J β 片段的分数。对于每个CDR3-J α -CDR3-J β ,准确度可以通过正确连接的TRAV片段的数目除以连接的TRAV片段的总数来计算。丰度是指每个克隆在总克隆库中的分数,其可以通过所述克隆的总读长数除以所有克隆的总读长数来计算。数据表明553个克隆中的523个具有高于95%的准确度和高于0.1/553的丰度,如方框中所示。

[0204] 图14示出了将每个TRAV映射到库中每个克隆的热图。克隆编号根据其同源TRAV基因名称进行排序。数据表明对于每个克隆,大多数读长具有正确的TRAV序列,这表明当将CDR3-J α -CDR3-J β 片段与它们的同源TRAV片段连接时具有高准确度。

[0205] 图15示出了产生TRAV-CDR3-J α -CDR3-J β 片段(例如,图15中的TRAV添加)后每个克隆的丰度相对于产生CDR3-J α -CDR3-J β 片段后的丰度。数据表明总体偏倚主要是CDR3-J α 和CDR3-J β 片段连接期间的偏倚。通过直接合成配对的CDR3-J α -CDR3-J β 片段可以减少或避免这种偏倚。

[0206] 图16示出了产生TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 片段后每个克隆的准确度和丰度。每个数据点对应于TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 片段的克隆。准确度是指与同源TRBV片段连接的CDR3-J β -TRBC1-TRAV-CDR3-J α 片段的分数。对于每个CDR3-J β -TRBC1-TRAV-CDR3-J α ,准确度可以通过正确连接的TRBV片段的数目除以连接的TRBV片段的总数来计算。丰度是指每个克隆在总克隆库中的分数,其可以通过所述克隆的总读长数除以所有克隆的总读长数来计算。数据表明553个克隆中的514个具有高于95%的准确度和高于0.1/553的丰度,如方框中所示。

[0207] 图17示出了将每个TRBV映射到库中每个克隆的热图。克隆编号根据其同源TRBV基因名称进行排序。数据表明对于每个克隆,大多数读长具有正确的TRBV序列,这表明当将CDR3-J β -TRBC1-TRAV-CDR3-J α 片段与它们的同源TRBV片段连接时具有高准确度。

[0208] 图18示出了产生TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 片段后每个克隆的总体准确度和丰度。计算每个克隆的总体准确度,乘以图12、图13和图16所示的每个步骤中的准确度。丰度通过所述克隆的总读长数除以所有克隆的总读长数来计算。

[0209] 虽然本文已经示出并描述了本发明的各种实施方案,但是本领域技术人员将显而易见此类实施方案仅以举例的方式提供。许多改变、变化和取代可由本领域技术人员想到而不脱离本发明。应理解,可使用本文所述的本发明的实施方案的各种替代方案。

[0210] 实施方案段落

[0211] [1]一种用于产生编码T细胞受体(TCR)链或其部分的核酸分子的方法,其包括:
(a)提供至少一种核酸分子,其包含编码TCR链的CDR3的序列;(b)提供多种核酸分子,所述多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列,其中所述多种核酸分子包含衍生自至少两种不同TCR V基因的至少两种不同序列;以及(c)使(a)的所述至少一种核酸分子与(b)的所述多种核酸分子在同一区室中接触,其中(a)的所述至少一种核酸分子能够与所述多种核酸分子中的核酸分子连接,以产生包含编码所述CDR3的序列和衍生自所述至少两种不同TCR V基因中的一种的序列的第三核酸分子,从而产生所述编码TCR链或其部分的核酸分子。

[0212] [2]如段落[1]所述的方法,其中所述至少一种核酸分子包括第一多种核酸分子,其中所述第一多种核酸分子中的每种核酸分子包含编码TCR链的CDR3的序列。

[0213] [3]如段落[1]或[2]所述的方法,其中(a)的所述至少一种核酸分子能够与所述多种核酸分子中的核酸分子特异性地连接,所述多种核酸分子中的核酸分子包含衍生自所述至少两种不同TCR V基因中的任何单个给定TCR V基因的序列。

[0214] [4]如段落[1]所述的方法,其中所述至少一种核酸分子还包含所述TCR链的J区。

[0215] [5]如段落[2]所述的方法,其中所述第一多种核酸分子中的每种核酸分子还包含TCR链的J区。

[0216] [6]如段落[1]-[5]中任一项所述的方法,其中所述至少两种TCR V基因是人TCR V基因或小鼠TCR V基因。

[0217] [7]如段落[1]-[6]中任一项所述的方法,其中所述至少两种TCR V基因选自由人TRAV1-1、TRAV1-2、TRAV2、TRAV3、TRAV4、TRAV5、TRAV6、TRAV7、TRAV8-1、TRAV8-2、TRAV8-3、TRAV8-4、TRAV8-6、TRAV9-1、TRAV9-2、TRAV10、TRAV12-1、TRAV12-2、TRAV12-3、TRAV13-1、TRAV13-2、TRAV14、TRAV16、TRAV17、TRAV18、TRAV19、TRAV20、TRAV21、TRAV22、TRAV23、TRAV24、TRAV25、TRAV26-1、TRAV26-2、TRAV27、TRAV29、TRAV30、TRAV34、TRAV35、TRAV36、TRAV38-1、TRAV38-2、TRAV39、TRAV40和TRAV41组成的组。

[0218] [8]如段落[1]-[6]中任一项所述的方法,其中所述至少两种TCR V基因选自由人TRBV2、TRBV3-1、TRBV4-1、TRBV4-2、TRBV4-3、TRBV5-1、TRBV5-4、TRBV5-5、TRBV5-6、TRBV5-8、TRBV6-1、TRBV6-2、TRBV6-3、TRBV6-4、TRBV6-5、TRBV6-6、TRBV6-8、TRBV6-9、TRBV7-2、TRBV7-3、TRBV7-4、TRBV7-6、TRBV7-7、TRBV7-8、TRBV7-9、TRBV9、TRBV10-1、TRBV10-2、TRBV10-3、TRBV11-1、TRBV11-2、TRBV11-3、TRBV12-3、TRBV12-4、TRBV12-5、TRBV13、TRBV14、TRBV15、TRBV16、TRBV18、TRBV19、TRBV20-1、TRBV24-1、TRBV25-1、TRBV27、TRBV28、TRBV29-1和TRBV30组成的组。

[0219] [9]如段落[1]-[8]中任一项所述的方法,其中衍生自所述至少两种不同TCR V基因的所述多种序列中的每种序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和/或FR3的序列。

[0220] [10]如段落[1]-[9]中任一项所述的方法,其中所述TCR链为TCR α 链、TCR β 链、TCR γ 链或TCR δ 链。

[0221] [11]如段落[1]-[10]中任一项所述的方法,其中所述至少一种核酸分子还包含编码额外TCR链的额外CDR3的额外序列。

[0222] [12]如段落[11]所述的方法,其中所述至少一种核酸分子包含所述额外TCR链的额外J区。

[0223] [13]如段落[11]或[12]所述的方法,其中所述编码CDR3的序列和所述编码额外CDR3的额外序列相隔至多100个核苷酸。

[0224] [14]如段落[11]-[13]中任一项所述的方法,其中所述TCR链和所述额外TCR链为TCR链的同源对。

[0225] [15]如段落[1]-[14]中任一项所述的方法,其中所述至少一种核酸分子包含连接序列,所述连接序列能够将所述至少一种核酸分子与所述多种核酸分子中的核酸分子连接以产生所述第三核酸分子。

[0226] [16]如段落[15]所述的方法,其中所述至少一种核酸分子和所述多种核酸分子中的核酸分子编码功能性TCR链或其部分。

[0227] [17]如段落[15]或[16]所述的方法,其中所述多种核酸分子中的核酸分子包含抗连接序列,所述抗连接序列与(a)的所述至少一种核酸分子的所述连接序列互补。

[0228] [18]如段落[1]-[17]中任一项所述的方法,其还包括将(a)的所述至少一种核酸分子和(b)的所述多种核酸分子中的核酸分子连接。

[0229] [19]如段落[18]所述的方法,其中连接包括使(a)的所述至少一种核酸分子与(b)的所述多种核酸分子中的核酸分子杂交。

[0230] [20]如段落[19]所述的方法,其中杂交包括使(a)的所述至少一种核酸分子的所述连接序列与(b)的所述多种核酸分子中的核酸分子的所述抗连接序列杂交。

[0231] [21]如段落[18]-[20]中任一项所述的方法,其还包括(i)使用(a)的所述至少一种核酸分子作为模板延伸所述多种核酸分子中的核酸分子的游离3'末端,和/或(ii)使用所述多种核酸分子中的核酸分子作为模板延伸(a)的所述至少一种核酸分子的游离3'末端,以产生所述第三核酸分子。

[0232] [22]如段落[1]-[21]中任一项所述的方法,其还包括将(a)的所述至少一种核酸分子与(b)的所述多种核酸分子中的核酸分子连接。

[0233] [23]如段落[1]-[22]中任一项所述的方法,其还包括使所述第三核酸分子与限制酶接触以产生粘性末端。

[0234] [24]如段落[1]-[23]中任一项所述的方法,其还包括使所述第三核酸分子与额外的核酸分子接触。

[0235] [25]如段落[24]所述的方法,其中所述额外的核酸分子编码TCR链的恒定区或其部分。

[0236] [26]如段落[24]或[25]所述的方法,其还包括将所述第三核酸分子与所述额外的核酸分子连接。

[0237] [27]如段落[1]-[26]中任一项所述的方法,其中多种核酸分子,每种编码不同的TCR链或其部分,在同一区室中产生。

[0238] [28]如段落[27]所述的方法,其中所述多种核酸分子中的至少五种不同的核酸分子在同一区室中产生。

[0239] [29]如段落[1]-[26]中任一项所述的方法,其中所述多种核酸分子中的至少十种不同的核酸分子在同一区室中产生。

[0240] [30]如段落[1]-[29]中任一项所述的方法,其中所述同一区室是孔、管或液滴。

[0241] [31]如段落[1]-[30]中任一项所述的方法,其中所述至少一种核酸分子包含独特的条形码。

[0242] [32]如段落[31]所述的方法,其中所述独特的条形码是引物结合位点。

[0243] [33]如段落[15]-[30]中任一项所述的方法,其中所述连接序列包含独特的条形码。

[0244] [34]如段落[33]所述的方法,其中所述独特的条形码是引物结合位点。

[0245] [35]一种组合物,其包含

[0246] (a)多种核酸分子,其中所述多种核酸分子中的每种核酸分子包含衍生自T细胞受

体 (TCR) V 基因的序列并且不包含 CDR3 序列, 其中所述多种中的第一核酸分子包含第一抗连接序列并且所述多种中的第二核酸分子包含第二抗连接序列, 其中所述第一抗连接序列不同于所述第二抗连接序列, 并且其中所述第一核酸分子和所述第二核酸分子的衍生自 TCR V 基因的序列衍生自不同 TCR V 基因; 以及

[0247] (b) 至少一种核酸分子, 其包含编码 TCR 链的 CDR3 的序列, 其中所述至少一种核酸分子还包含与所述第一抗连接序列互补的第一连接序列。

[0248] [36] 如段落 [35] 所述的组合物, 其中所述组合物是液体组合物。

[0249] [37] 如段落 [35] 或 [36] 所述的组合物, 其中 (a) 的所述多种核酸分子和 (b) 的所述至少一种核酸分子在同一区室中。

[0250] [38] 如段落 [35] - [37] 中任一项所述的组合物, 其中所述衍生自 TCR V 基因的序列包含所述 TCR V 基因的至少十个核苷酸。

[0251] [39] 如段落 [35] - [38] 中任一项所述的组合物, 其中所述 TCR V 基因是 TRAV 基因、TRBV 基因、TRGV 基因或 TRDV 基因。

[0252] [40] 如段落 [35] - [39] 中任一项所述的组合物, 其中所述衍生自 TCR V 基因的序列包含编码 L-PART1、L-PART2、FR1、CDR1、FR2、CDR2 和/或 FR3 的序列。

[0253] [41] 如段落 [35] - [40] 中任一项所述的组合物, 其中所述至少一种核酸分子还包含所述 TCR 链的 J 区。

[0254] [42] 如段落 [35] - [41] 中任一项所述的组合物, 其中所述至少一种核酸分子还包含编码额外 TCR 链的额外 CDR3 的额外序列。

[0255] [43] 如段落 [42] 所述的组合物, 其中所述至少一种核酸分子还包含所述额外 TCR 链的额外 J 区。

[0256] [44] 如段落 [42] 或 [43] 所述的组合物, 其中所述编码 CDR3 的序列和所述编码 CDR3 的额外序列相隔至多 100 个核苷酸。

[0257] [45] 如段落 [42] - [44] 中任一项所述的组合物, 其中所述 TCR 链和所述额外 TCR 链为 TCR 链的同源对。

[0258] [46] 如段落 [35] - [45] 中任一项所述的组合物, 其中 (b) 的所述至少一种核酸分子包含第一多种核酸分子, 并且其中所述第一多种核酸分子中的每种核酸分子包含编码 TCR 链的 CDR3 的序列。

[0259] [47] 如段落 [46] 所述的组合物, 其中所述第一多种核酸分子中的每种核酸分子编码不同 TCR 链的不同 CDR3。

[0260] [48] 如段落 [46] 或 [47] 所述的组合物, 其中所述第一多种核酸分子中的每种核酸分子包含不同的连接序列, 所述不同的连接序列能够与所述多种核酸分子中的包含衍生自任何单个给定 TCR V 基因的序列的核酸分子特异性地连接。

[0261] [49] 如段落 [35] - [48] 中任一项所述的组合物, 其中所述第一抗连接序列或所述第二抗连接序列包含 TCR V 基因序列。

[0262] [50] 如段落 [49] 所述的组合物, 其中所述 TCR V 基因序列包含与重排基因中编码 CDR3 的序列相邻的所述 TCR V 基因的至少三个核苷酸。

[0263] [51] 如段落 [35] - [50] 中任一项所述的组合物, 其中所述第一抗连接序列或所述第二抗连接序列包含预定序列。

[0264] [52]如段落[35]-[51]中任一项所述的组合物,其中所述第一连接序列与所述第一抗连接序列杂交。

[0265] [53]如段落[35]-[52]中任一项所述的组合物,其中(b)的所述至少一种核酸分子包含独特的条形码。

[0266] [54]如段落[53]所述的组合物,其中所述独特的条形码是引物结合位点。

[0267] [55]如段落[35]-[52]中任一项所述的组合物,其中所述至少一种核酸分子的所述第一连接序列包含独特的条形码。

[0268] [56]如段落[55]所述的组合物,其中所述独特的条形码是引物结合位点。

[0269] [57]一种用于产生多种核酸分子的方法,其包括:(a)提供第一多种核酸分子,其中所述第一多种核酸分子中的核酸分子包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的序列,其中所述第一CDR3和所述第二CDR3来自TCR链的同源对;(b)提供第二多种核酸分子,其中所述第二多种核酸分子中的核酸分子包含衍生自TCR V基因的序列,其中所述核酸分子不包含编码恒定结构域的序列;以及(c)使所述第一多种核酸分子与所述第二多种核酸分子接触,其中所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接以形成包含所述编码第一CDR3和第二CDR3的序列以及所述衍生自TCR V基因的序列的核酸分子,其中所述编码第一CDR3和第二CDR3的序列以及所述TCR V基因衍生自所述TCR链的同源对。

[0270] [58]如段落[57]所述的方法,其中所述第一多种核酸分子中的每种核酸分子包含编码第一TCR链的不同第一CDR3和/或第二TCR链的不同CDR3的序列。

[0271] [59]如段落[57]或[58]所述的方法,其中所述第二多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。

[0272] [60]如段落[57]-[59]中任一项所述的方法,其中所述第一多种核酸分子和所述第二多种核酸分子在同一区室中接触。

[0273] [61]如段落[57]-[60]中任一项所述的方法,其中所述第一多种核酸分子中的核酸分子还包含连接序列,其中所述连接序列将所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接。

[0274] [62]如段落[61]所述的方法,其中所述第二多种核酸分子中的核酸分子还包含抗连接序列,所述抗连接序列与所述连接序列互补。

[0275] [63]如段落[62]所述的方法,其中所述连接序列与所述抗连接序列杂交以将所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接。

[0276] [64]如段落[58]-[63]中任一项所述的方法,其中所述连接序列是密码子多样化的,使得所述第一多种核酸分子中的核酸分子的所述连接序列不同于所述第一多种核酸分子中的其他核酸分子的其他连接序列。

[0277] [65]如段落[57]-[64]中任一项所述的方法,其中所述第一多种核酸分子中的核酸分子还包含所述第一TCR链的第一J区和/或所述第二TCR链的第二J区。

[0278] [66]如段落[57]-[65]中任一项所述的方法,其中(i)所述第一TCR链是TCR α 链并且所述第二TCR链是TCR β 链,或(ii)所述第一TCR链是TCR γ 链并且所述第二TCR链是TCR δ 链。

[0279] [67]如段落[57]-[66]中任一项所述的方法,其中所述TCR V基因是TRAV基因、

TRBV基因、TRGV基因或TRDV基因。

[0280] [68]如段落[57]-[67]中任一项所述的方法,其中所述第二多种核酸分子中的核酸分子是双链核酸分子。

[0281] [69]如段落[57]-[68]中任一项所述的方法,其中所述第二多种核酸分子中的核酸分子还包含编码自切割肽的一部分的序列。

[0282] [70]如段落[62]-[69]中任一项所述的方法,其中所述抗连接序列是所述第二多种核酸分子中的核酸分子的突出端。

[0283] [71]如段落[62]-[70]中任一项所述的方法,其中所述连接序列或所述抗连接序列的长度为至少三个核苷酸。

[0284] [72]如段落[63]-[71]中任一项所述的方法,其还包括(i)延伸与所述第二多种核酸分子中的核酸分子杂交的所述第一多种核酸分子中的核酸分子的3'末端和/或(ii)延伸与所述第一多种核酸分子中的核酸分子杂交的所述第二多种核酸分子中的核酸分子的3'末端。

[0285] [73]如段落[57]-[72]中任一项所述的方法,其还包括将所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接。

[0286] [74]如段落[57]-[73]中任一项所述的方法,其还包括使包含所述编码第一CDR3和第二CDR3的序列以及所述衍生自TCR V基因的序列的所述核酸分子与限制酶接触以产生粘性末端。

[0287] [75]如段落[57]-[74]中任一项所述的方法,使包含所述编码第一CDR3和第二CDR3的序列以及所述衍生自TCR V基因的序列的所述核酸分子与包含编码恒定区或其部分的序列的额外的核酸分子接触。

[0288] [76]如段落[74]或[75]所述的方法,其还包括将包含所述编码第一CDR3和第二CDR3的序列以及所述衍生自TCR V基因的序列的所述核酸分子通过粘性末端与额外的核酸分子连接。

[0289] [77]如段落[57]-[76]中任一项所述的方法,其中所述编码第一CDR3的序列和所述编码第二CDR3的序列相隔至多100个核苷酸。

[0290] [78]如段落[57]-[77]中任一项所述的方法,其中所述衍生自TCR V基因的序列包含编码FR1、CDR1、FR2、CDR2和FR3的序列。

[0291] [79]如段落[57]-[77]中任一项所述的方法,其中所述衍生自TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和FR3的序列。

[0292] [80]一种组合物,其包含:(a)第一多种核酸分子,其中所述第一多种核酸分子中的每种核酸分子包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的序列,其中所述第一CDR3和所述第二CDR3来自TCR链的同源对;和(b)第二多种核酸分子,其中所述第二多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列,并且其中所述第二多种核酸分子中的每种核酸分子不包含编码所述第一CDR3和所述第二CDR3的序列;

[0293] 其中(i)所述第一多种核酸分子中的每种核酸分子包含编码不同的第一CDR3和/或第二CDR3的序列,且/或(ii)所述第二多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。

[0294] [81]如段落[80]所述的组合物,其中所述第一多种核酸分子中的每种核酸分子还

包含连接序列,其中给定连接序列可用于将所述第一多种核酸分子中的给定核酸分子与所述第二多种核酸分子中的给定核酸分子连接。

[0295] [82]如段落[80]或[81]所述的组合物,其中所述第二多种核酸分子中的每种核酸分子还包含抗连接序列,所述抗连接序列与所述连接序列互补。

[0296] [83]如段落[81]或[82]所述的组合物,其中所述连接序列是密码子多样化的,使得所述第一多种核酸分子中的给定核酸分子的给定连接序列不同于所述第一多种核酸分子中的其他核酸分子的其他连接序列。

[0297] [84]如段落[81]-[83]中任一项所述的组合物,其中所述连接序列编码氨基酸序列。

[0298] [85]如段落[84]所述的组合物,其中所述连接序列与所述编码第一TCR链的第一CDR3和第二TCR链的第二CDR3的序列在框内。

[0299] [86]如段落[81]-[85]中任一项所述的组合物,其中所述连接序列包含至少三个核苷酸。

[0300] [87]如段落[86]所述的组合物,其中所述连接序列包含与重排基因中编码第一TCR链的第一CDR3或第二TCR链的第二CDR3的序列相邻的所述TCR V基因的至少三个核苷酸。

[0301] [88]如段落[84]-[87]中任一项所述的组合物,其中由所述给定连接序列编码的给定氨基酸序列与由至少一种其他连接序列编码的至少一种其他氨基酸序列相同或基本上相同。

[0302] [89]如段落[84]-[87]中任一项所述的组合物,其中由所述给定连接序列编码的给定氨基酸序列不同于由其他连接序列编码的其他氨基酸序列。

[0303] [90]如段落[80]-[89]中任一项所述的组合物,其中所述第一多种核酸分子中的每种核酸分子还包含所述第一TCR链的第一J区和/或所述第二TCR链的第二J区。

[0304] [91]如段落[80]-[90]中任一项所述的组合物,其中所述组合物是液体组合物。

[0305] [92]如段落[80]-[91]中任一项所述的组合物,其中所述第一多种核酸分子和所述第二多种核酸分子在同一区室内。

[0306] [93]如段落[81]-[92]中任一项所述的组合物,其中所述第一多种核酸分子中的给定核酸分子通过所述给定连接序列与所述第二多种核酸分子中的给定核酸分子连接。

[0307] [94]如段落[93]所述的组合物,其中所述第一多种核酸分子中的给定核酸分子通过与给定抗连接序列杂交的所述给定连接序列与所述第二多种核酸分子中的给定核酸分子杂交。

[0308] [95]如段落[80]-[94]中任一项所述的组合物,其中所述编码第一CDR3的序列和所述编码第二CDR3的序列相隔至多100个核苷酸。

[0309] [96]如段落[80]-[95]中任一项所述的组合物,其中所述衍生自TCR V基因的序列包含编码FR1、CDR1、FR2、CDR2和FR3的序列。

[0310] [97]如段落[80]-[95]中任一项所述的组合物,其中所述衍生自TCR V基因的序列包含编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和FR3的序列。

[0311] [98]如段落[80]-[97]中任一项所述的组合物,其中所述第一多种核酸分子或所述第二多种分子中的每种核酸分子是化学合成的。

[0312] [99]如段落[80]-[98]中任一项所述的组合物,其中所述第一多种核酸分子中的每种核酸分子的长度为至多约250个核苷酸。

[0313] [100]一种包含多种核酸分子的组合物,所述多种核酸分子中的每种核酸分子包含衍生自T细胞受体(TCR)V基因的序列,其中所述多种核酸分子包含具有第一连接序列的第一核酸分子和具有第二连接序列的第二核酸分子,其中所述第一连接序列不同于所述第二连接序列。

[0314] [101]如段落[100]所述的组合物,所述多种核酸分子中的每种核酸分子包含衍生自不同TCR V基因的序列。

[0315] [102]如段落[100]或[101]所述的组合物,所述多种核酸分子中的每种核酸分子包含不同的连接序列。

[0316] [103]如段落[100]-[102]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子不包含编码TCR链的CDR3的序列。

[0317] [104]如段落[100]-[103]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子不包含编码TCR链的恒定结构域的序列。

[0318] [105]如段落[100]-[104]中任一项所述的组合物,其中所述衍生自TCR V基因的序列包含所述TCR V基因的至少十个核苷酸。

[0319] [106]如段落[100]-[105]中任一项所述的组合物,其中所述TCR V基因是TRAV基因、TRBV基因、TRGV基因或TRDV基因。

[0320] [107]一种包含多种核酸分子的组合物,所述多种核酸分子中的每种核酸分子编码T细胞受体(TCR)链的CDR3,其中所述多种中的第一核酸分子包含第一连接序列,并且所述多种中的第二核酸分子包含第二连接序列,其中所述第一连接序列不同于所述第二连接序列。

[0321] [108]如段落[107]所述的组合物,其中所述多种核酸分子中的每种核酸分子还包含TCR链的J区。

[0322] [109]如段落[107]所述的组合物,其中所述多种核酸分子中的每种核酸分子编码第一TCR链的第一CDR3和第二TCR链的第二CDR3。

[0323] [110]如段落[109]所述的组合物,其中所述多种核酸分子中的每种核酸分子还包含第一TCR链的第一J区和第二TCR链的第二J区。

[0324] [111]如段落[107]-[110]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子编码不同TCR链的不同CDR3。

[0325] [112]如段落[107]-[111]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子包含不同的连接序列。

[0326] [113]如段落[107]-[112]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子不包含TCR V基因的大于200个核苷酸。

[0327] [114]如段落[107]-[113]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子不包含编码TCR链的恒定结构域的序列。

[0328] [115]如段落[100]-[114]中任一项所述的组合物,其中所述第一连接序列或所述第二连接序列包含衍生自TCR V基因的序列。

[0329] [116]如段落[115]所述的组合物,其中所述衍生自TCR V基因的序列包含与重排

基因中编码CDR3的序列相邻的所述TCR V基因的至少三个核苷酸。

[0330] [117]如段落[100]-[116]中任一项所述的组合物,其中所述第一连接序列或所述第二连接序列包含预定序列。

[0331] [118]如段落[107]-[114]中任一项所述的组合物,其中所述第一连接序列或所述第二连接序列包含与TCR V基因序列互补的序列。

[0332] [119]如段落[107]-[114]和[118]中任一项所述的组合物,其中所述组合物还包含第二多种核酸分子,所述第二多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列。

[0333] [120]如段落[119]所述的组合物,其中所述第二多种中的第一核酸分子包含第一抗连接序列,所述第一抗连接序列与所述第一连接序列互补。

[0334] [121]如段落[119]或[120]所述的组合物,其中所述第二多种中的第二核酸分子包含第二抗连接序列,所述第二抗连接序列与所述第二连接序列互补。

[0335] [122]如段落[120]或[121]所述的组合物,其中所述第二多种中的第一核酸分子的第一抗连接序列与所述第一多种中的第一核酸分子的第一连接序列连接。

[0336] [123]如段落[121]或[122]所述的组合物,其中所述第二多种中的第二核酸分子的第二抗连接序列与所述第一多种中的第二核酸分子的第二连接序列连接。

[0337] [124]一种包含多种核酸分子的组合物,每种包含编码T细胞受体(TCR)链的至少十个氨基酸的序列,其中所述多种中的第一核酸分子包含第一连接序列,并且所述多种中的第二核酸分子包含第二连接序列,其中所述第一连接序列不同于所述第二连接序列,其中所述第一连接序列或所述第二连接序列编码TCR链的一部分,并且其中所述第一连接序列或所述第二连接序列与所述编码TCR链的至少十个氨基酸的序列在框内。

[0338] [125]如段落[124]所述的组合物,其中所述第一连接序列或所述第二连接序列包含TCR链基因的至少四个连续核苷酸,并且与所述编码TCR链的至少十个氨基酸的序列在框内。

[0339] [126]如段落[124]或[125]所述的组合物,其中所述第一连接序列和所述第二连接序列编码TCR链的至少两个连续氨基酸。

[0340] [127]如段落[124]-[126]中任一项所述的组合物,其中所述TCR链的一部分的TCR链与由编码至少十个氨基酸的序列编码的所述TCR链相同。

[0341] [128]如段落[124]所述的组合物,其中所述多种核酸分子中的每种核酸分子包含衍生自TCR V基因的序列。

[0342] [129]如段落[124]-[128]中任一项所述的组合物,其中所述多种核酸分子中的每种核酸分子编码所述TCR链的CDR3。

[0343] [130]如段落[129]所述的组合物,其中所述多种核酸分子中的每种核酸分子还包含所述TCR链的J区。

[0344] [131]如段落[129]所述的组合物,其中所述多种核酸分子中的每种核酸分子编码第一TCR链的第一CDR3和第二TCR链的第二CDR3。

[0345] [132]如段落[131]所述的组合物,其中所述多种核酸分子中的每种核酸分子还包含第一TCR链的第一J区和第二TCR链的第二J区。

[0346] [133]如段落[131]或[132]所述的组合物,其中编码所述第一CDR3的序列和编码

所述第二CDR3的序列相隔至多100个核苷酸。

[0347] [134]如段落[124]-[133]中任一项所述的组合物,其中所述第一连接序列或所述第二连接序列包含衍生自TCR V基因的序列。

[0348] [135]如段落[124]-[134]中任一项所述的组合物,其中所述第一连接序列或所述第二连接序列包含预定序列。

[0349] [136]如段落[100]-[135]中任一项所述的组合物,其中所述第一连接序列包含至少一个不同于所述第二连接序列的核苷酸的核苷酸。

[0350] [137]如段落[100]-[136]中任一项所述的组合物,其中所述第一连接序列与所述第二连接序列编码相同的氨基酸序列。

[0351] [138]如段落[100]-[136]中任一项所述的组合物,其中所述第一连接序列与所述第二连接序列编码不同的氨基酸序列。

[0352] [139]一种用于产生多种核酸分子的方法,所述多种中的每种核酸分子编码T细胞受体(TCR)链或其区域,所述方法包括:使第一多种核酸分子与第二多种核酸分子接触以产生第三多种核酸分子,所述第三多种核酸分子包含至少两种不同的核酸分子,其中所述至少两种不同的核酸分子中的每种具有编码不同TCR链或其区域的不同序列,并且其中所述至少两种不同的核酸分子在同一区室中产生。

[0353] [140]如段落[139]所述的方法,其中所述第一多种核酸分子中的每种核酸分子包含编码所述TCR链的CDR3的序列。

[0354] [141]如段落[140]所述的方法,其中所述第一多种核酸分子中的每种核酸分子包含所述TCR链的J区。

[0355] [142]如段落[139]-[141]中任一项所述的方法,其中所述第二多种核酸分子中的每种核酸分子包含衍生自所述TCR链的TCR V基因的序列。

[0356] [143]如段落[142]所述的方法,其中所述TCR V基因是人TCR V基因。

[0357] [144]如段落[142]或[143]所述的方法,其中所述TCR V基因是人TRAV1-1、TRAV1-2、TRAV2、TRAV3、TRAV4、TRAV5、TRAV6、TRAV7、TRAV8-1、TRAV8-2、TRAV8-3、TRAV8-4、TRAV8-6、TRAV9-1、TRAV9-2、TRAV10、TRAV12-1、TRAV12-2、TRAV12-3、TRAV13-1、TRAV13-2、TRAV14、TRAV16、TRAV17、TRAV18、TRAV19、TRAV20、TRAV21、TRAV22、TRAV23、TRAV24、TRAV25、TRAV26-1、TRAV26-2、TRAV27、TRAV29、TRAV30、TRAV34、TRAV35、TRAV36、TRAV38-1、TRAV38-2、TRAV39、TRAV40或TRAV41。

[0358] [145]如段落[142]或[143]所述的方法,其中所述TCR V基因是人TRBV2、TRBV3-1、TRBV4-1、TRBV4-2、TRBV4-3、TRBV5-1、TRBV5-4、TRBV5-5、TRBV5-6、TRBV5-8、TRBV6-1、TRBV6-2、TRBV6-3、TRBV6-4、TRBV6-5、TRBV6-6、TRBV6-8、TRBV6-9、TRBV7-2、TRBV7-3、TRBV7-4、TRBV7-6、TRBV7-7、TRBV7-8、TRBV7-9、TRBV9、TRBV10-1、TRBV10-2、TRBV10-3、TRBV11-1、TRBV11-2、TRBV11-3、TRBV12-3、TRBV12-4、TRBV12-5、TRBV13、TRBV14、TRBV15、TRBV16、TRBV18、TRBV19、TRBV20-1、TRBV24-1、TRBV25-1、TRBV27、TRBV28、TRBV29-1或TRBV30。

[0359] [146]如段落[139]-[145]中任一项所述的方法,其中所述衍生自TCR V基因的序列包括编码FR1、CDR1、FR2、CDR2和FR3的序列。

[0360] [147]如段落[139]-[145]中任一项所述的方法,其中所述衍生自TCR V基因的序

列包括编码L-PART1、L-PART2、FR1、CDR1、FR2、CDR2和FR3的序列。

[0361] [148]如段落[139]-[147]中任一项所述的方法,其中所述TCR链为TCR α 链、TCR β 链、TCR γ 链或TCR δ 链。

[0362] [149]如段落[140]-[148]中任一项所述的方法,其中所述第一多种核酸分子中的每种核酸分子还包含编码额外TCR链的额外CDR3的额外序列。

[0363] [150]如段落[149]所述的方法,其中所述第一多种核酸分子中的每种核酸分子包含所述额外TCR链的额外J区。

[0364] [151]如段落[149]或[150]所述的方法,其中所述TCR链和所述额外TCR链为TCR链的同源对。

[0365] [152]如段落[139]-[151]中任一项所述的方法,其中所述多种核酸分子中的核酸分子编码不同的TCR或其区域。

[0366] [153]如段落[139]-[152]中任一项所述的方法,其中所述第一多种核酸分子中的给定核酸分子包含连接序列,所述连接序列可用于将所述第一多种核酸分子中的给定核酸分子与所述第二多种核酸分子中的给定核酸分子连接。

[0367] [154]如段落[153]所述的方法,其中所述第一多种核酸分子中的给定核酸分子和所述第二多种核酸分子中的给定核酸分子编码功能性TCR链或其区域。

[0368] [155]如段落[153]或[154]所述的方法,其中所述第二多种核酸分子中的给定核酸分子包含抗连接序列,所述抗连接序列与所述第一多种核酸分子中的给定核酸分子的所述连接序列互补。

[0369] [156]如段落[153]-[155]中任一项所述的方法,其还包括将所述第一多种核酸分子中的给定核酸分子与所述第二多种核酸分子中的给定核酸分子连接。

[0370] [157]如段落[156]所述的方法,其中连接包括将所述第一多种核酸分子中的给定核酸分子与所述第二多种核酸分子中的给定核酸分子杂交。

[0371] [158]如段落[157]所述的方法,其中杂交包括将所述第一多种核酸分子中的给定核酸分子的所述连接序列与所述第二多种核酸分子中的给定核酸分子的所述抗连接序列杂交。

[0372] [159]如段落[156]-[158]中任一项所述的方法,其还包括(i)使用所述第一多种核酸分子中的给定核酸分子作为模板延伸所述第二多种核酸分子中的给定核酸分子的游离3'末端,和/或(ii)使用所述第二多种核酸分子中的给定核酸分子作为模板延伸所述第一多种核酸分子中的核酸分子的游离3'末端,以产生所述第三多种核酸分子中的核酸分子。

[0373] [160]如段落[139]-[159]中任一项所述的方法,其还包括将所述第一多种核酸分子中的给定核酸分子与所述第二多种核酸分子中的给定核酸分子连接。

[0374] [161]如段落[139]-[160]中任一项所述的方法,其还包括使所述第三多种核酸分子中的核酸分子与限制酶接触以产生粘性末端。

[0375] [162]如段落[139]-[161]中任一项所述的方法,其还包括使所述第三多种核酸分子中的核酸分子与额外的核酸分子接触。

[0376] [163]如段落[162]所述的方法,其中所述额外的核酸分子编码TCR链的恒定区或其部分。

[0377] [164]如段落[162]或[163]所述的方法,其还包括将所述第三多种核酸分子中的核酸分子与所述额外的核酸分子连接。

[0378] [165]如段落[139]-[164]中任一项所述的方法,其中所述第三多种核酸分子中的至少五种不同的核酸分子在同一区室中产生。

[0379] [166]如段落[139]-[165]中任一项所述的方法,其中所述第三多种核酸分子中的至少十种不同的核酸分子在同一区室中产生。

[0380] [167]如段落[139]-[166]中任一项所述的方法,其中所述同一区室是孔、管或液滴。

[0381] [168]一种用于产生多种核酸分子的方法,其包括:(a)提供第一多种核酸分子,其中所述第一多种核酸分子中的核酸分子包含编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的序列,其中所述第一CDR3和所述第二CDR3来自TCR链的同源对;(b)提供第二多种核酸分子,其中所述第二多种核酸分子中的核酸分子包含衍生自TCR V基因的序列;以及(c)使所述第一多种核酸分子与所述第二多种核酸分子接触,其中所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接以形成包含所述编码第一CDR3和第二CDR3的序列以及所述衍生自TCR V基因的序列的线性核酸分子,其中所述编码第一CDR3和第二CDR3的序列以及所述TCR V基因衍生自所述TCR链的同源对。

[0382] [169]一种用于产生多种核酸分子的方法,其包括:(a)提供第一多种核酸分子,其中所述第一多种核酸分子中的核酸分子包含(i)编码第一T细胞受体(TCR)链的第一CDR3和第二TCR链的第二CDR3的合成序列,和(ii)编码第三T细胞受体(TCR)链的第三CDR3和第四TCR链的第四CDR3的合成序列,其中所述第一CDR3和所述第二CDR3来自TCR链的第一同源对,并且其中所述第三CDR3和所述第四CDR3来自TCR链的第二同源对;(b)提供第二多种核酸分子,其中所述第二多种核酸分子中的核酸分子包含衍生自TCR V基因的序列;以及(c)使所述第一多种核酸分子与所述第二多种核酸分子接触,其中所述第一多种核酸分子中的核酸分子与所述第二多种核酸分子中的核酸分子连接以形成包含所述编码第一CDR3和第二CDR3的序列以及所述衍生自TCR V基因的序列的核酸分子,其中所述编码第一CDR3和第二CDR3的序列以及所述TCR V基因衍生自所述TCR链的同源对。

[0383] [170]一种在来自受试者的组织样品中鉴定天然配对的T细胞受体(TCR)的序列的方法,其包括:(a)在从所述受试者获得的含有多个外周T细胞的样品中鉴定一种或多种天然配对的TCR的一种或多种配对序列,其中所述一种或多种配对序列中的每种包含CDR3序列;和(b)在所述组织样品中鉴定TCR的TCR链的组织CDR3序列,与所述TCR链天然配对的其他TCR链是未知的,其中所述组织CDR3序列与所述一种或多种天然配对的TCR的所述一种或多种配对序列中的至少一种配对序列的CDR3序列匹配,从而在所述组织样品中将所述至少一种配对序列鉴定为所述天然配对的TCR的序列。

[0384] [171]如段落[170]所述的方法,其中(a)中的鉴定包括对含有所述多个外周T细胞的所述样品中的所述一种或多种天然配对的TCR进行测序。

[0385] [172]如段落[171]所述的方法,其中所述测序包括单细胞测序。

[0386] [173]如段落[172]所述的方法,其中所述单细胞测序包括将所述多个外周T细胞分成多个区室,每个区室包含所述多个外周T细胞中的单个外周T细胞。

[0387] [174]如段落[170]-[173]中任一项所述的方法,其中所述组织样品不是体液样

品。

[0388] [175]如段落[170]-[174]中任一项所述的方法,其中所述组织样品是实体瘤样品。

[0389] [176]如段落[170]-[175]中任一项所述的方法,其中所述组织样品是固定或冷冻样品。

[0390] [177]如段落[170]-[176]中任一项所述的方法,其中所述含有所述多个外周T细胞的样品是外周血单核细胞(PBMC)样品。

[0391] [178]如段落[170]-[177]中任一项所述的方法,其还包括在(a)之前从所述受试者获得血液样品。

[0392] [179]如段落[178]所述的方法,其还包括在(a)之前从所述血液样品中分离外周血单核细胞。

[0393] [180]如段落[170]-[179]中任一项所述的方法,其中所述组织样品包含肿瘤浸润性T细胞。

[0394] [181]一种鉴定靶反应性T细胞受体(TCR)的方法,其包括:(a)提供包含从段落[170]-[180]中任一项中鉴定的所述TCR的细胞;和(b)使所述细胞与抗原呈递细胞(APC)呈递的靶抗原接触,其中所述细胞通过所述TCR与所述APC呈递的所述靶抗原结合,从而将所述TCR鉴定为所述靶反应性TCR。

[0395] [182]如段落[181]所述的方法,其中所述靶抗原是肿瘤抗原。

[0396] [183]如段落[181]或[182]所述的方法,其还包括将编码所述靶反应性TCR的序列递送到宿主细胞中。

[0397] [184]如段落[183]所述的方法,其还包括将所述宿主细胞施用于所述受试者。

[0398] [185]如段落[183]或[184]所述的方法,其中所述宿主细胞是T细胞。

[0399] [186]如段落[185]所述的方法,其中所述T细胞是自体T细胞。

[0400] [187]如段落[185]所述的方法,其中所述T细胞是同种异体T细胞。

[0401] [188]如段落[181]-[187]中任一项所述的方法,其中所述细胞是报告细胞系,所述报告细胞系包含在所述细胞与所述APC呈递的所述靶抗原结合时表达的报告基因。

序列表

	<110> 根路径基因组学公司 (ROOTPATH GENOMICS, INC.)	
	<120> 用于 T 细胞受体基因组装的组合物和方法	
	<130> 53563-705.601	
	<140> PCT/US2020/026558	
	<141> 2020-04-03	
	<150> 62/972,231	
	<151> 2020-02-10	
	<150> 62/898,053	
	<151> 2019-09-10	
	<150> 62/838,465	
	<151> 2019-04-25	
[0001]	<150> 62/829,813	
	<151> 2019-04-05	
	<160> 292	
	<170> PatentIn 3.5 版	
	<210> 1	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 1	
	tcaaagactc tgcctcatac ctct	24
	<210> 2	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 2	
	ctgagagacg cagctgtgta ttact	25
	<210> 3	
	<211> 22	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 3	
	ggggactcag ccgtgtactt ct	22
	<210> 4	
	<211> 22	
	<212> DNA	
[0002]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 4	
	ggtgactccg cagcctatatt ct	22
	<210> 5	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 5	
	tggggatagc gcagtctatt tct	23
	<210> 6	

<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 6
ccggagacag cgcagtttat tttt 24

<210> 7
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 7
ggtgacagcg ccgtctat tt 22

[0003]

<210> 8
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 8
tggaggactc aggcacttac ttct 24

<210> 9
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 9

	tggaggactc tgggacatac tttt	24
	<210> 10	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 10	
	tggaggactc tggcacctat tttt	24
	<210> 11	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0004]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 11	
	actcgaggat tccgtactt atttct	26
	<210> 12	
	<211> 22	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 12	
	gaagactcgc ggacctactt tt	22
	<210> 13	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 13	
	gctggaggat tccggaacct atttct	26
	<210> 14	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 14	
	ctcgaagata gcggcacata ttttt	25
	<210> 15	
	<211> 26	
	<212> DNA	
[0005]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 15	
	gcctgtgat agcgcaatat acttct	26
	<210> 16	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 16	
	tggcgactct gcaatgtact tct	23
	<210> 17	

<211> 25	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 17	
cccgagact ctgctatgta ttttt	25
<210> 18	
<211> 26	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 18	
ggaatccgat agcgcagtct attact	26
[0006]	
<210> 19	
<211> 22	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 19	
tccgacagcg ctgtctacta ct	22
<210> 20	
<211> 27	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 20	

	aagagattga tagcgctgtt tactact	27
	<210> 21	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 21	
	aggaatctga ttccgcagtc tattttt	27
	<210> 22	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0007]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 22	
	agaatctgat agcgccgttt attatt	26
	<210> 23	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 23	
	gtccgactcc gcagtctact act	23
	<210> 24	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 24	
	aggagtctga ttctgcagtc tactatt	27
	<210> 25	
	<211> 29	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 25	
	ccaagaaata gattccgcag tetactact	29
	<210> 26	
	<211> 23	
	<212> DNA	
[0008]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 26	
	gtctgacagc gcagtctact tet	23
	<210> 27	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 27	
	aggaaagcga ttctgcagtc tattact	27
	<210> 28	

<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 28
aagagtctga ctccgcagtt tattatt 27

<210> 29
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 29
agaatccgac tctgcagttt actatt 26

[0009]

<210> 30
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 30
gagtctgata gcgctgtgta ctact 25

<210> 31
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 31

	gaatctgact ctgccgttta ctatt	25
	<210> 32	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 32	
	agtccgactc tgctgtgtac tact	24
	<210> 33	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0010]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 33	
	ccatctgatt ccgcactgta tttct	25
	<210> 34	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 34	
	accttctgat agcgtctctct attttt	26
	<210> 35	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 35	
	ccttctgatt ctgcactgta cctgt	25
	<210> 36	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 36	
	ccaagcgatt ctgcactgta ttttt	25
	<210> 37	
	<211> 24	
	<212> DNA	
[0011]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 37	
	cctctgactc tgcagtctac ctct	24
	<210> 38	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 38	
	cccagcgact ctgcagttta tctct	25
	<210> 39	

<211> 21	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 39	
ccgacagcgc actctacctg t	21
<210> 40	
<211> 23	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 40	
ttccgactct gcactgtatc tgt	23
[0012]	
<210> 41	
<211> 22	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 41	
tccgatagcg ccctgtatct ct	22
<210> 42	
<211> 23	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 42	

	ctccgattcc gcactctatc tct	23
	<210> 43	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 43	
	cctccgatag cgctgtttat ctct	24
	<210> 44	
	<211> 21	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0013]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 44	
	gcgacagcgc cctgtacttt t	21
	<210> 45	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 45	
	ccctctgata gcgcactgta tctct	25
	<210> 46	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 46	
	cttctgacag cgctgtgtat ctgt	24
	<210> 47	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 47	
	gcgaggacac agctgtttac tttt	24
	<210> 48	
	<211> 27	
	<212> DNA	
[0014]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 48	
	agtgcgaaga tacagcagtt tacttct	27
	<210> 49	
	<211> 28	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 49	
	cgggtgtgagg atactgctgt ttatttct	28
	<210> 50	

<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 50
cgaagatacc gccgtctact ttt 23

<210> 51
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 51
ggcaactgac acagcagtct actttt 26

[0015]

<210> 52
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 52
gagcgattct gccgtttact tct 23

<210> 53
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 53

	tccgattccg ccgtgtat tt	22
	<210> 54	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 54	
	ttggtctgat tctgcagttt actttt	26
	<210> 55	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0016]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 55	
	ggtctgattc cgctgtctac tttt	24
	<210> 56	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 56	
	ctggtctgac tctgtgttt attttt	26
	<210> 57	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 57	
	ggtccgactg ggcagtcctat tttt	24
	<210> 58	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 58	
	tggctctgatt ctgccgtcta tttct	25
	<210> 59	
	<211> 22	
	<212> DNA	
[0017]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 59	
	agcgactctg ccgtgtatct ct	22
	<210> 60	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 60	
	gagcgattgg gcagtctact ttt	23
	<210> 61	

<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 61
tggtcggatt ctgctgtcta ttttt 25

<210> 62
<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 62
agcctgaaga ttcagccatc tactttc 27

[0018]

<210> 63
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 63
cccaggact ctgctattta cttct 25

<210> 64
<211> 29
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 64

	acagccagaa gattctgcaa tatacttct	29
	<210> 65	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 65	
	gctcgatgac acagctacat acatct	26
	<210> 66	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0019]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 66	
	tcctggatga tactgcaaca tacatat	27
	<210> 67	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 67	
	ctctctgact ctgcactgta ctact	25
	<210> 68	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 68	
	actgtctgac tctgcactct attact	26
	<210> 69	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 69	
	ctgtccgatt ctgcactcta ctact	25
	<210> 70	
	<211> 27	
	<212> DNA	
[0020]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 70	
	aactgtctga ttctgctctg tactatt	27
	<210> 71	
	<211> 22	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 71	
	tccgactccg ctctgtatTT tt	22
	<210> 72	

<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 72
agcgactctg ccctctacta ct 22

<210> 73
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 73
tctctgactc cgctctctac tact 24

[0021]

<210> 74
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 74
tctctgattc tgcctctac tttt 24

<210> 75
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 75

	gctctccgat tctgctctgt attatt <210> 76 <211> 26 <212> DNA <213> 人工序列(Artificial Sequence) <220> <223> 人工序列的描述: 合成寡核苷酸 <400> 76 gacaacagac tcaggcactt atctct	26
	<210> 77 <211> 26 <212> DNA <213> 人工序列(Artificial Sequence) <220> <223> 人工序列的描述: 合成寡核苷酸 <400> 77 aacaactgac tctggcacat attttt	26
[0022]	<210> 78 <211> 23 <212> DNA <213> 人工序列(Artificial Sequence) <220> <223> 人工序列的描述: 合成寡核苷酸 <400> 78 cactgatagc ggaacctacc tct	23
	<210> 79 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 79	
	tgacagcggc acctacctgt	20
	<210> 80	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 80	
	actacagatt ccggcactta cttct	25
	<210> 81	
	<211> 30	
	<212> DNA	
[0023]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 81	
	gccagataac tgattctggt acttacctgt	30
	<210> 82	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 82	
	acaactgaca gcggaacata tctct	25
	<210> 83	

<211> 27	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 83	
aaataacaga tagcggtaca tacctgt	27
<210> 84	
<211> 24	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 84	
ccacagattc tggcacctac ttct	24
[0024]	
<210> 85	
<211> 22	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 85	
actgactccg gaacctacct ct	22
<210> 86	
<211> 22	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 86	

	accgactctg gcacttacct gt	22
	<210> 87	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 87	
	aatcacagac tctggaacct atctgt	26
	<210> 88	
	<211> 29	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0025]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 88	
	ccaaattacc gattctggta catacctct	29
	<210> 89	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 89	
	cggagatagc gccacatact ttt	23
	<210> 90	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 90	
	aacctggaga ttctgcaaca tatttct	27
	<210> 91	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 91	
	ctggggactc tgcaacttac ttct	24
	<210> 92	
	<211> 25	
	<212> DNA	
[0026]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 92	
	cctggagact cagctaccta cttct	25
	<210> 93	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 93	
	ccggggatag cgctacttat tttt	24
	<210> 94	

<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 94
cctggagatt ccgcaactta ctttt 25

<210> 95
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 95
ccaggggatt ctgctaccta ttttt 25

[0027]

<210> 96
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 96
cccgagatt ctgccactta tttct 25

<210> 97
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 97

	ctggcgacag cgctacttat ttct	24
	<210> 98 <211> 27 <212> DNA <213> 人工序列(Artificial Sequence)	
	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 98 aaccagacga ttcgggaaag tattttct	27
	<210> 99 <211> 24 <212> DNA <213> 人工序列(Artificial Sequence)	
[0028]	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 99 cagaggattc agggacgtac ttct	24
	<210> 100 <211> 25 <212> DNA <213> 人工序列(Artificial Sequence)	
	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 100 ccagacgact ccggaaagta ctttt	25
	<210> 101 <211> 24 <212> DNA <213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 101	
	ccgaggactc cggtagatac ttct	24
	<210> 102	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 102	
	aacccgatga ctctggaag tatTTTT	27
	<210> 103	
	<211> 26	
	<212> DNA	
[0029]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 103	
	gccagaagac tccgtacat atTTTT	26
	<210> 104	
	<211> 29	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 104	
	tcaaattgaa gattctgcag tctactTTT	29
	<210> 105	

<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 105
gattgaggac tcggcagtat atttct 26

<210> 106
<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 106
aaatcgaaga ctctgcagtt tactttt 27

[0030]

<210> 107
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 107
gagcgactca gccaaagtact tct 23

<210> 108
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 108

	aggggatgct gggatctact ttt	23
	<210> 109	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 109	
	cccgaagata cagctgtcta cctgt	25
	<210> 110	
	<211> 22	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0031]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 110	
	agggacgcag cagtctatca tt	22
	<210> 111	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 111	
	gccactctgc catctacttc tgt	23
	<210> 112	
	<211> 21	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 112	
	ggcgcacact gtactgcaca t	21
	<210> 113	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 113	
	tgatgactcg gccacatact tct	23
	<210> 114	
	<211> 24	
	<212> DNA	
[0032]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 114	
	tggaggactc agctgtgtac ttct	24
	<210> 115	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 115	
	accagaagat agcgagttt atctgt	26
	<210> 116	

<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 116
atccagaaga ctcagctgtc tatTTTT 27

<210> 117
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 117
tcgaagatag cgccatgtac tttt 24

[0033]

<210> 118
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 118
actggaagat agcgctgtgt atttct 26

<210> 119
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 119

	aagccagacc agcctctatt ttt	23
	<210> 120	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 120	
	ccccctctcag acatcagtggt acttct	26
	<210> 121	
	<211> 21	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0034]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 121	
	gccagaccgc cgtgtatttc t	21
	<210> 122	
	<211> 22	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 122	
	ggcgacacag ccacctatct ct	22
	<210> 123	
	<211> 26	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 123	
	gcctaaagac agcgctgttt atctct	26
	<210> 124	
	<211> 23	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 124	
	ccaggactca gcggtgtatc ttt	23
	<210> 125	
	<211> 27	
	<212> DNA	
[0035]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 125	
	gcctagagta ttctgccatg tacctct	27
	<210> 126	
	<211> 25	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 126	
	aaaaatgaga tggcagtctt cctct	25
	<210> 127	

<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 127
cgaggatagg ggcctgtatc tct 23

<210> 128
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 128
gcagaagact cagcactgta ctgt 25

[0036]

<210> 129
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 129
gacgactcag cactgtacct ct 22

<210> 130
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 130

	ggggactccg cactctatct ct	22
	<210> 131	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 131	
	aaacaaacca gacatctgtg tacttct	27
	<210> 132	
	<211> 27	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0037]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 132	
	ggcctggaga cagcagtatc tatttct	27
	<210> 133	
	<211> 24	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 133	
	tcagccatag cggtttttac ctct	24
	<210> 134	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 134	
	ccgggatttt gtgactcatc	20
	<210> 135	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 135	
	gaggatcgta tgtttcgcac	20
	<210> 136	
	<211> 20	
	<212> DNA	
[0038]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 136	
	cttgtgtgca cttaccgtac	20
	<210> 137	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 137	
	tgatgcatct ccagtacagg	20
	<210> 138	

<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 138	
cttctgtgtg tacctcgaca	20
<210> 139	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 139	
cttgcaatcc tttagcgtgc	20
[0039]	
<210> 140	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 140	
gcaagtgtgg aaaatgaccc	20
<210> 141	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 141	

	gagtctagtc tcacaacca	20
	<210> 142	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 142	
	gaaatgttga ggactccacg	20
	<210> 143	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0040]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 143	
	cctaacagat gctacgtgga	20
	<210> 144	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 144	
	gtaggtccac acagattcca	20
	<210> 145	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 145	
	gccagtcaca gcaaatacac	20
	<210> 146	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 146	
	ccgctaccag tatgtacctt	20
	<210> 147	
	<211> 20	
	<212> DNA	
[0041]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 147	
	actgtgttcc ttgtcttcg	20
	<210> 148	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 148	
	tgaatgcac tacggtaccg	20
	<210> 149	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 149
gcgcttatca atcttgctcg 20

<210> 150
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 150
cgggtcaattc agtagccact 20

[0042]

<210> 151
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 151
ggacacatgt acactagcca 20

<210> 152
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 152

	tgggagctct acgaaaatcc	20
	<210> 153	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 153	
	gttctcgaga tcgtcacaca	20
	<210> 154	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0043]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 154	
	cttctgcatt cgatccttgc	20
	<210> 155	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 155	
	gcagagttgt gtgattggag	20
	<210> 156	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 156	
	cccatcaatt cggaaccatc	20
	<210> 157	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 157	
	atcgtaaccc aagtctgtgg	20
	<210> 158	
	<211> 20	
	<212> DNA	
[0044]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 158	
	ggcgaaatga tccctgaatg	20
	<210> 159	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 159	
	agtgctcaga actttcaggc	20
	<210> 160	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 160
gatcgttaac tctttgggcg 20

<210> 161
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 161
acacgaggat tgctgtagag 20

[0045]

<210> 162
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 162
ttctaccaca ttgtctcggg 20

<210> 163
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 163

	gaatggctaa actgtgtccc	20
	<210> 164	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 164	
	cggactgtac gagaaactga	20
	<210> 165	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0046]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 165	
	cttgcgacaa actactcctg	20
	<210> 166	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 166	
	ccgttttact ttgtgccac	20
	<210> 167	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 167	
	tg gat gat at c act t c g g c g	20
	<210> 168	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 168	
	c c a a c c t c t a t a t g t g c c c a	20
	<210> 169	
	<211> 20	
	<212> DNA	
[0047]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 169	
	t g a a c a g g t a t g c t c c a g a g	20
	<210> 170	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 170	
	t t g t g g a t a t c g t c t g g t c c	20
	<210> 171	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 171
ctgtggaact cgactcttgt 20

<210> 172
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 172
ctccagatg cacaaattcg 20

[0048]

<210> 173
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 173
ggctcatgac aaaacacagg 20

<210> 174
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 174

	ccgaatccga aaacaacacg	20
	<210> 175 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	
	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 175 agacctaaca ctgtgatgcg	20
	<210> 176 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	
[0049]	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 176 cctgggtgag cataaacttc	20
	<210> 177 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	
	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 177 gagtcttgga cgaacaaagg	20
	<210> 178 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 178	
	cacgtaccca tcattgtctg	20
	<210> 179	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 179	
	ccgtgttagt caagtgtgtg	20
	<210> 180	
	<211> 20	
	<212> DNA	
[0050]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 180	
	ctggtggcat aaatggaacg	20
	<210> 181	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 181	
	tggatgtggg tatcaatggg	20
	<210> 182	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 182
tgtggctaac gtaggacaag 20

<210> 183
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 183
ccctcgttgt gaaaatgtgc 20

[0051]

<210> 184
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 184
tcgtcatagg tcagcttacg 20

<210> 185
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 185

	cctgatgacc tctatgccaa	20
	<210> 186 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	
	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 186 cggcaagaat gaatagggtg	20
	<210> 187 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	
[0052]	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 187 gtgctatttg tgggaaatgg	20
	<210> 188 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	
	<220> <223> 人工序列的描述: 合成寡核苷酸	
	<400> 188 gccatgtttg cttactgacg	20
	<210> 189 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 189	
	cgttgtggca ttcattagcg	20
	<210> 190	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 190	
	gcggtaggat tggatctcat	20
	<210> 191	
	<211> 20	
	<212> DNA	
[0053]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 191	
	cctcgcaaag ctgttatgac	20
	<210> 192	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 192	
	gccttcatgt tattggacgc	20
	<210> 193	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 193
agctgtagtg ttcttgaggc 20

<210> 194
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 194
ggtagtggtc gtgtgacatg 20

[0054]

<210> 195
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 195
cgcgcatat gttcatatcc 20

<210> 196
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 196

	gagactggat catgcaacag	20
	<210> 197	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 197	
	cacaacttct ctggactcca	20
	<210> 198	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0055]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 198	
	cgaccatgat ctgtatgcgt	20
	<210> 199	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 199	
	ggtgtgactc ttgtttccgt	20
	<210> 200	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 200	
	acgtacatac aagtctggcg	20
	<210> 201	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 201	
	cctcaaggat tcactcgcaa	20
	<210> 202	
	<211> 20	
	<212> DNA	
[0056]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 202	
	ctgtatagga tgtccacgca	20
	<210> 203	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 203	
	gcctgtgatt ggtaaattgcg	20
	<210> 204	

<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 204	
cgcactcgta gcatctagaa	20
<210> 205	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 205	
cgatttggtg tccctagctg	20
[0057]	
<210> 206	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 206	
cccacttcat ctgactctga	20
<210> 207	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 207	

	cggcattgta caggtgttac	20
	<210> 208	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 208	
	tctcctattt ccctgaacgg	20
	<210> 209	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0058]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 209	
	cttgccact aaacgaacg	20
	<210> 210	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 210	
	cgggtatcac tgggtaatga	20
	<210> 211	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 211	
	ggaacagaga ccaatccagt	20
	<210> 212	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 212	
	gtgggcatcc gaaatttcag	20
	<210> 213	
	<211> 20	
	<212> DNA	
[0059]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 213	
	cgcgacgaca ttaccaatag	20
	<210> 214	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 214	
	cgtcggaata tgctctcaga	20
	<210> 215	

<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 215	
gatgcagatc aatgagtggc	20
<210> 216	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 216	
actgcttaca agtgtccacg	20
[0060]	
<210> 217	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 217	
actgtatgca agctagtcce	20
<210> 218	
<211> 20	
<212> DNA	
<213> 人工序列(Artificial Sequence)	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<400> 218	

	agatctccca aaagtgtccg	20
	<210> 219	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 219	
	ttccagaacc atgtgatccc	20
	<210> 220	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0061]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 220	
	gccttgtctt tcaacctctg	20
	<210> 221	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 221	
	gatacgatc ttcacatgcg	20
	<210> 222	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 222	
	cgctcatcta ggttggacta	20
	<210> 223	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 223	
	cgcggttcaga ttccaaacag	20
	<210> 224	
	<211> 20	
	<212> DNA	
[0062]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 224	
	gcctggttac acatgetatc	20
	<210> 225	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 225	
	gcaaaggtcc tacaggtttc	20
	<210> 226	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 226
ggctttccat gtctatgctc 20

<210> 227
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 227
gcgacattag cagagtaggt 20

[0063]

<210> 228
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 228
ctcgccatac tatctgcatg 20

<210> 229
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 229

	ctactgaaca cttggcaagc	20
	<210> 230	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 230	
	ctgttcaatt cctgtgcgag	20
	<210> 231	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0064]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 231	
	cactgagatg gaatttggcg	20
	<210> 232	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 232	
	gtccatcaca acttcactg	20
	<210> 233	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 233	
	ggctcagtct actttgcttc	20
	<210> 234	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 234	
	gtagcttcc gacacaatgg	20
	<210> 235	
	<211> 20	
	<212> DNA	
[0065]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 235	
	ccgtgacaca ctttcacetc	20
	<210> 236	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 236	
	ccgggatgtc attatgagca	20
	<210> 237	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 237
gagtgtccta cgagatcaga 20

<210> 238
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 238
taacgtctct ctgagtgtgg 20

[0066]

<210> 239
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 239
accetagaca agagacacct 20

<210> 240
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 240

	tgctcagtac tcttcacgc	20
	<210> 241	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 241	
	agctcaatca tggctatcgg	20
	<210> 242	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0067]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 242	
	ctacacattg catccaaccc	20
	<210> 243	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 243	
	acttgtegaa tagctcaggc	20
	<210> 244	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 244	
	gtctaccctg agaaccagtt	20
	<210> 245	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 245	
	cagcaacaac ctaccttagc	20
	<210> 246	
	<211> 20	
	<212> DNA	
[0068]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 246	
	cgatttggtg gtacgtgtcc	20
	<210> 247	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 247	
	gccatttcct ttgtacctgc	20
	<210> 248	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 248
ggccaataga gagaccacaa 20

<210> 249
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 249
cggagtcaca tgggtagaat 20

[0069]

<210> 250
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 250
cccagtacat ttgtcggttg 20

<210> 251
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 251

	cccactagct gctactcaaa	20
	<210> 252	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 252	
	ggtgttgct caaagtagac	20
	<210> 253	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0070]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 253	
	tactccagct cttactgtgc	20
	<210> 254	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 254	
	ggatgagcag tcaacagttc	20
	<210> 255	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 255	
	tcaggatcga tcagttgtcc	20
	<210> 256	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 256	
	cctctctttt gtgcggaaac	20
	<210> 257	
	<211> 20	
	<212> DNA	
[0071]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 257	
	tgcttaggat ttcgagaacg	20
	<210> 258	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 258	
	ggcattgtcc ttaacttcgc	20
	<210> 259	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 259
tgcatctaac tacgatgggc 20

<210> 260
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 260
ccctagtagc cacacaacat 20

[0072]

<210> 261
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 261
gtgccatgaa tcatcgtctc 20

<210> 262
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 262

	cgctctgatg aaagctccat	20
	<210> 263	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 263	
	ccagccatag tgcatacct	20
	<210> 264	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0073]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 264	
	gagattgtca tgtggtcgac	20
	<210> 265	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 265	
	ccgcagtcta acaggaaatc	20
	<210> 266	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 266	
	cgcttcgact gaaccttatg	20
	<210> 267	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 267	
	cgatgcgacc aatagaagtg	20
	<210> 268	
	<211> 20	
	<212> DNA	
[0074]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 268	
	gcccttggtgta cgacatatg	20
	<210> 269	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 269	
	cagtgattta ggtgacgcag	20
	<210> 270	

<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 270
ggcatggaag aggtagtttc 20

<210> 271
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 271
ccgatcgtat tctgtgtcca 20

[0075]

<210> 272
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 272
ctaagtcaag cacatgggac 20

<210> 273
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述: 合成寡核苷酸

<400> 273

	gatccacact caatctcctg	20
	<210> 274	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 274	
	ccttgtcaca tgctgtatc	20
	<210> 275	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
[0076]	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 275	
	cgcgattgtg gttaataggc	20
	<210> 276	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 276	
	gtaggcaaag ttcaccacac	20
	<210> 277	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	

	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 277	
	gccacgaatc gaacaagtac	20
	<210> 278	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 278	
	ttgagatctc gatgagcacg	20
	<210> 279	
	<211> 20	
	<212> DNA	
[0077]	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 279	
	gggccaagat ctattcgtea	20
	<210> 280	
	<211> 20	
	<212> DNA	
	<213> 人工序列(Artificial Sequence)	
	<220>	
	<223> 人工序列的描述: 合成寡核苷酸	
	<400> 280	
	gtggctatag gtatgtccga	20
	<210> 281	

	<211> 20 <212> DNA <213> 人工序列(Artificial Sequence) <220> <223> 人工序列的描述: 合成寡核苷酸 <400> 281 ccacactttc tgcattcgac	20
	<210> 282 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence) <220> <223> 人工序列的描述: 合成寡核苷酸 <400> 282 cggcatctca aagcacatac	20
[0078]	<210> 283 <211> 20 <212> DNA <213> 人工序列(Artificial Sequence) <220> <223> 人工序列的描述: 合成寡核苷酸 <400> 283 cgtccacaaa ttactgccc	20
	<210> 284 <211> 19 <212> PRT <213> 猪捷申病毒(Porcine teschovirus) <400> 284 Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn 1 5 10 15	

Pro Gly Pro

<210> 285

<211> 20

<212> PRT

<213> 马鼻炎 A 病毒(Equine rhinitis A virus)

<400> 285

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> 286

<211> 22

<212> PRT

[0079] <213> 口蹄疫病毒(foot-and-mouth disease virus)

<400> 286

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> 287

<211> 18

<212> PRT

<213> 东亚细亚病毒(Thosea asigna virus)

<400> 287

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro
1 5 10 15

Gly Pro

<210> 288

<211> 5

<212> PRT

<213> 未知(Unknown)

<220>

<223> 未知的描述: 肠激酶切割位点

<400> 288

Asp Asp Asp Asp Lys

1 5

<210> 289

<211> 4

<212> PRT

<213> 未知(Unknown)

<220>

<223> 未知的描述: 因子 Xa 切割位点

[0080]

<400> 289

Ile Glu Gly Arg

1

<210> 290

<211> 6

<212> PRT

<213> 未知(Unknown)

<220>

<223> 未知的描述: 凝血酶切割位点

<400> 290

Leu Val Pro Arg Gly Ser

1 5

<210> 291

<211> 8

<212> PRT

<213> 未知(Unknown)

<220>

<223> 未知的描述: 肾素切割位点

<400> 291

His Pro Phe His Leu Val Ile His

1

5

[0081]

<210> 292

<211> 4

<212> PRT

<213> 未知(Unknown)

<220>

<223> 未知的描述: 半胱天冬酶蛋白酶切割位点

<400> 292

Asp Glu Val Asp

1

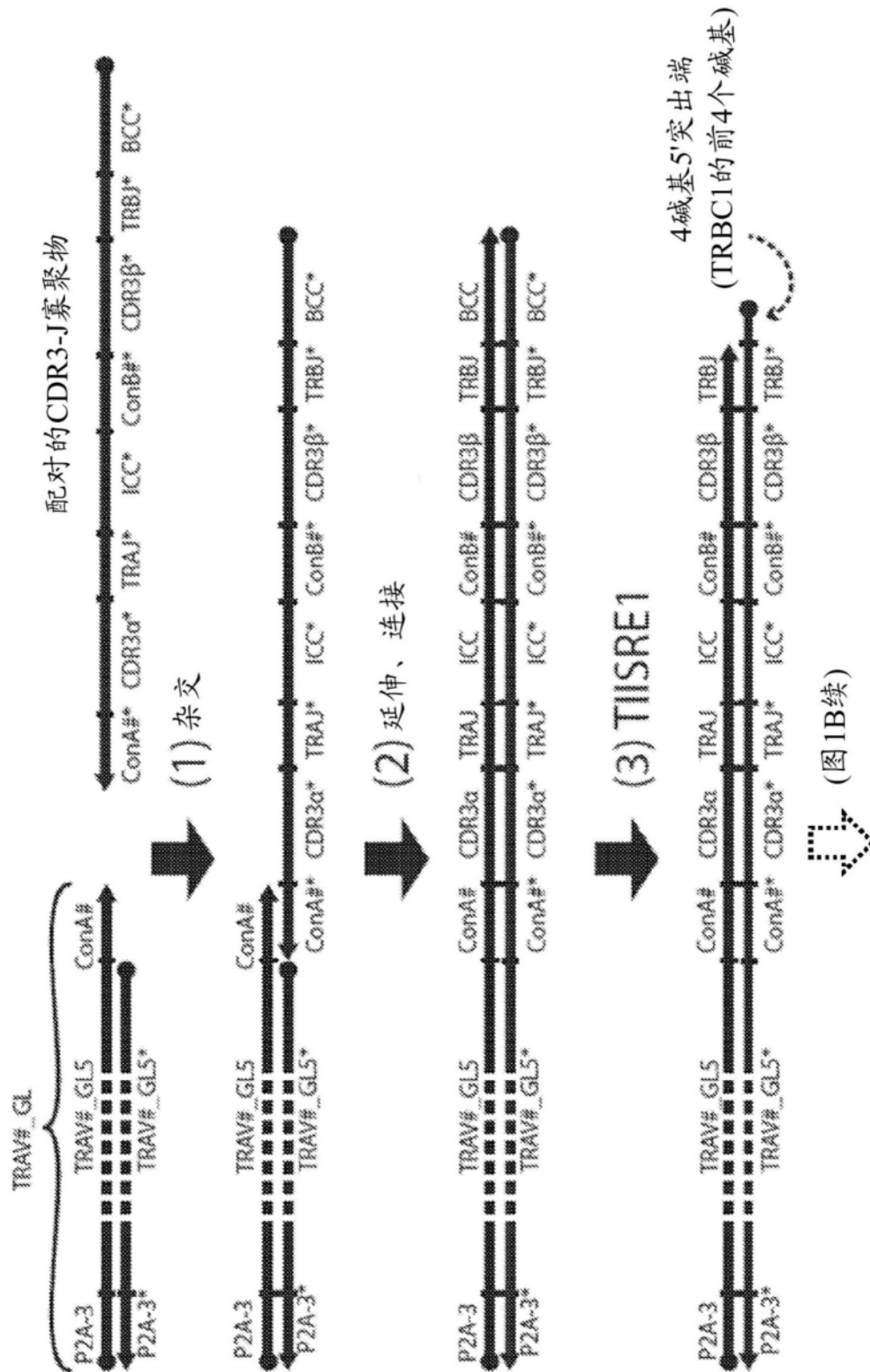


图1A

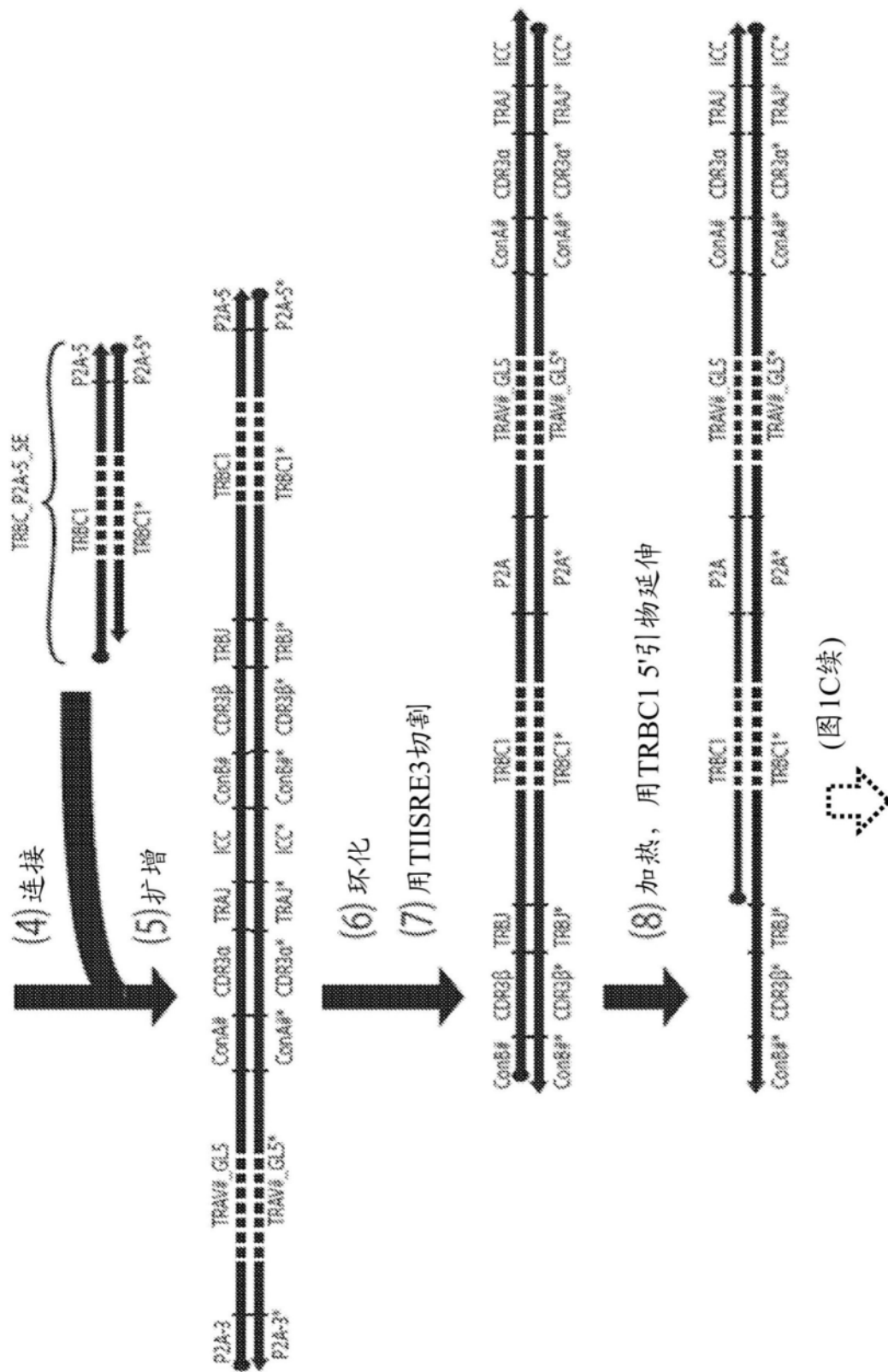


图1B

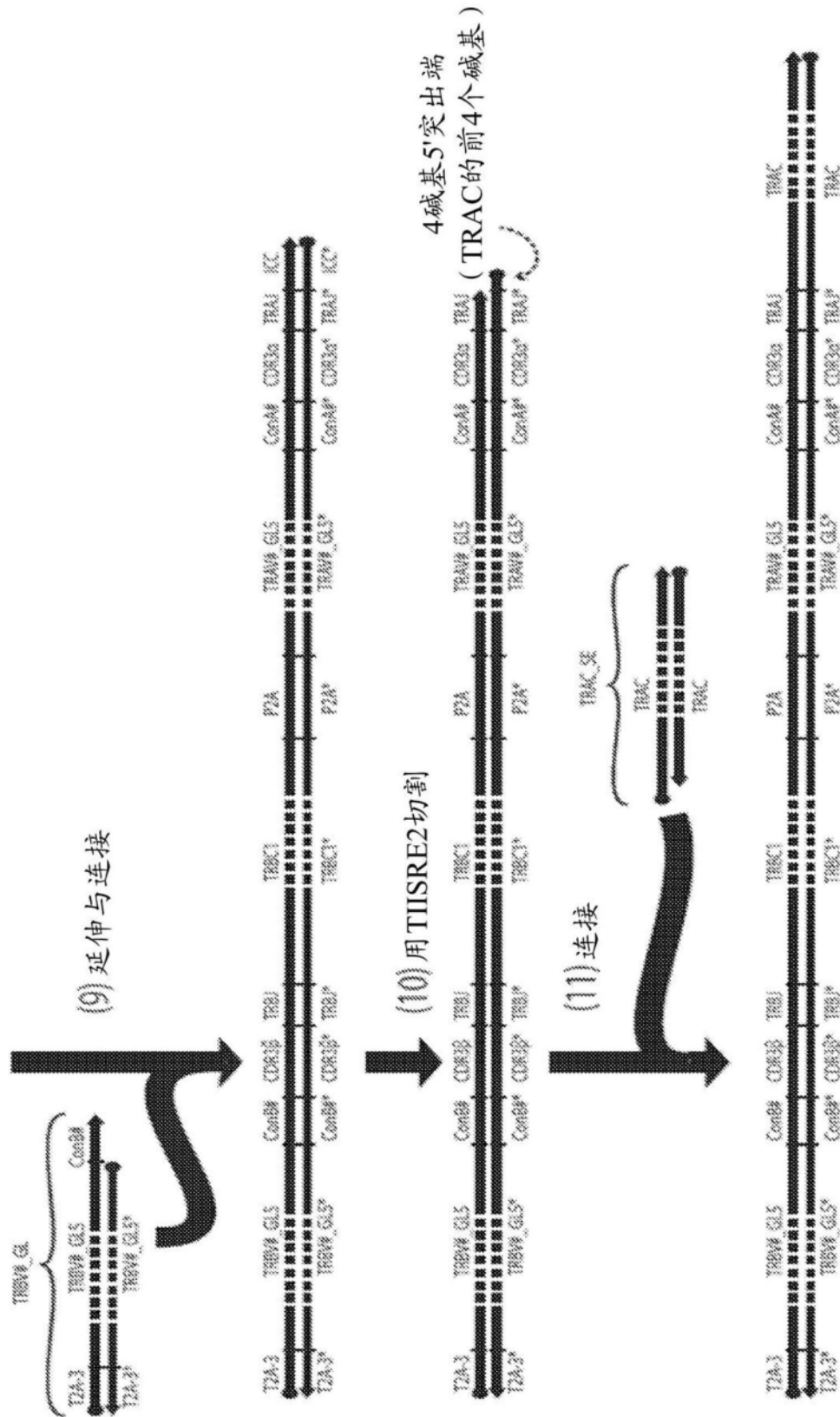


图1C

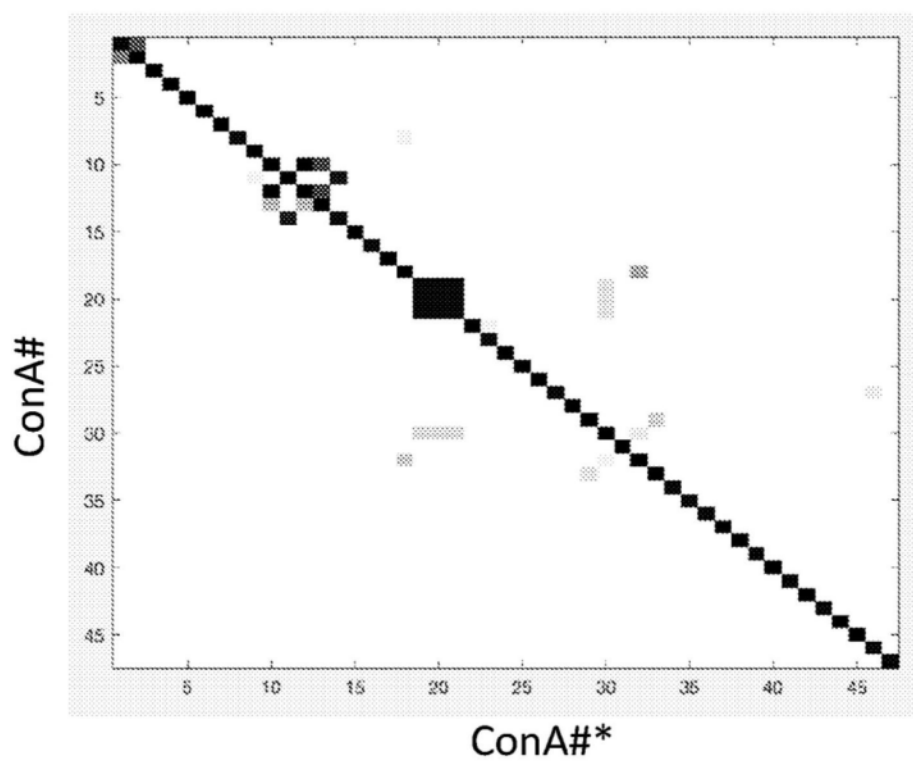


图2A

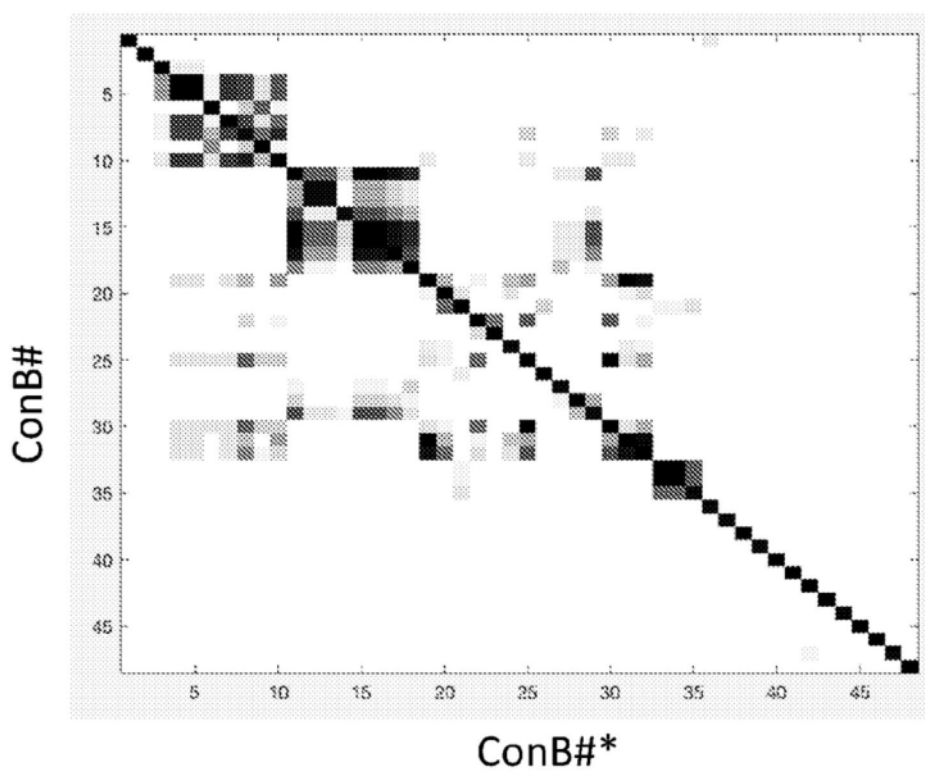


图2B

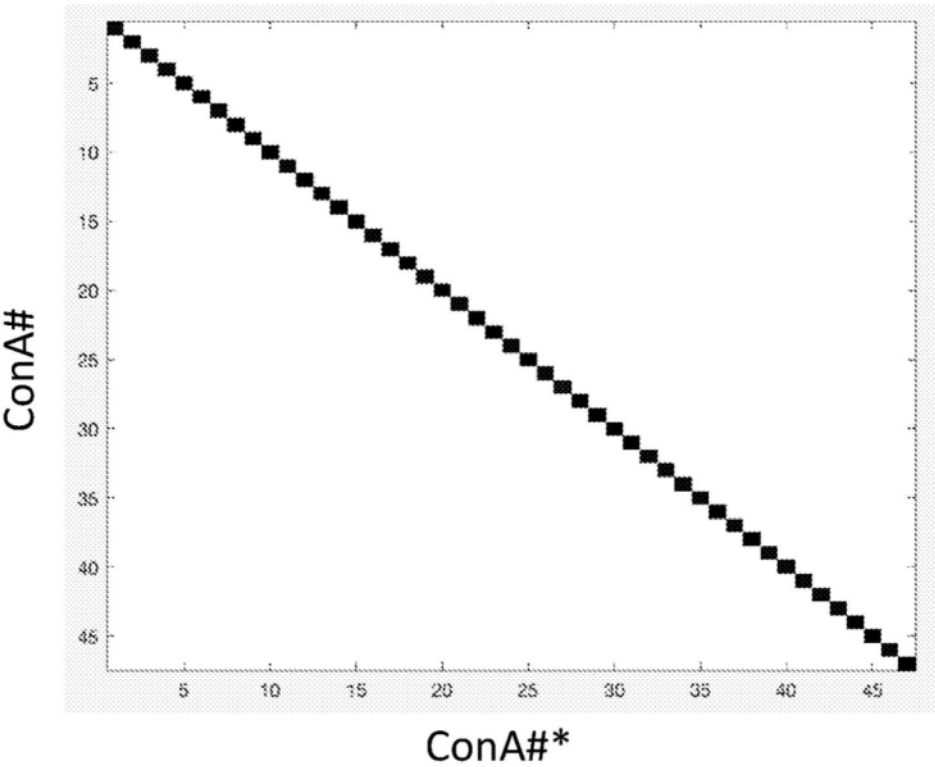


图3A

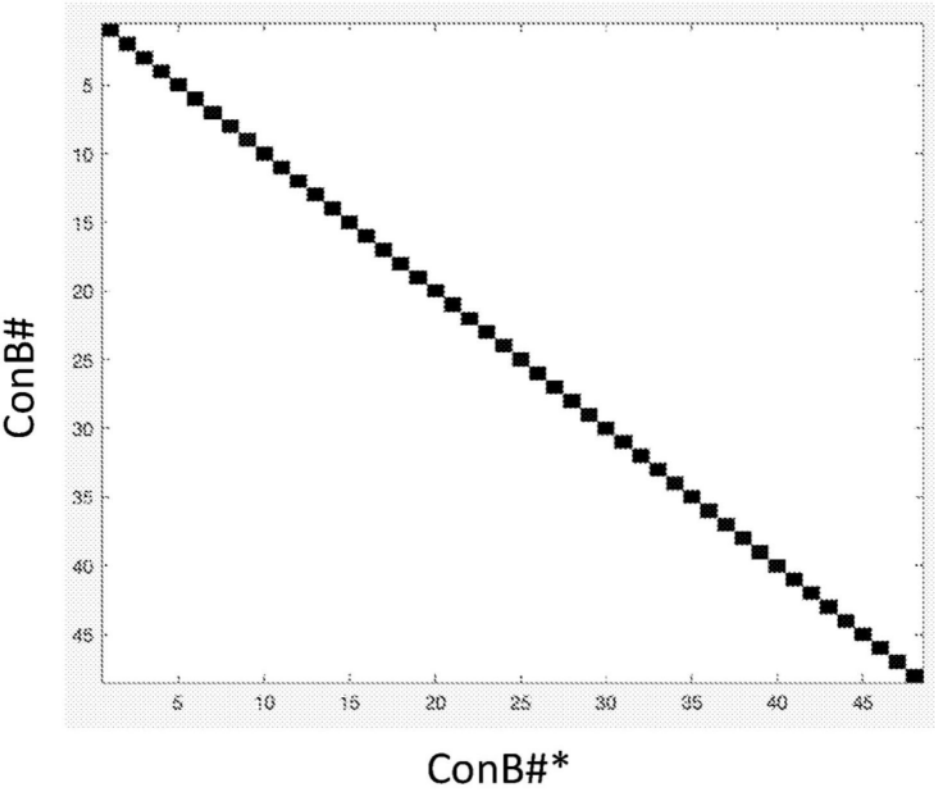


图3B

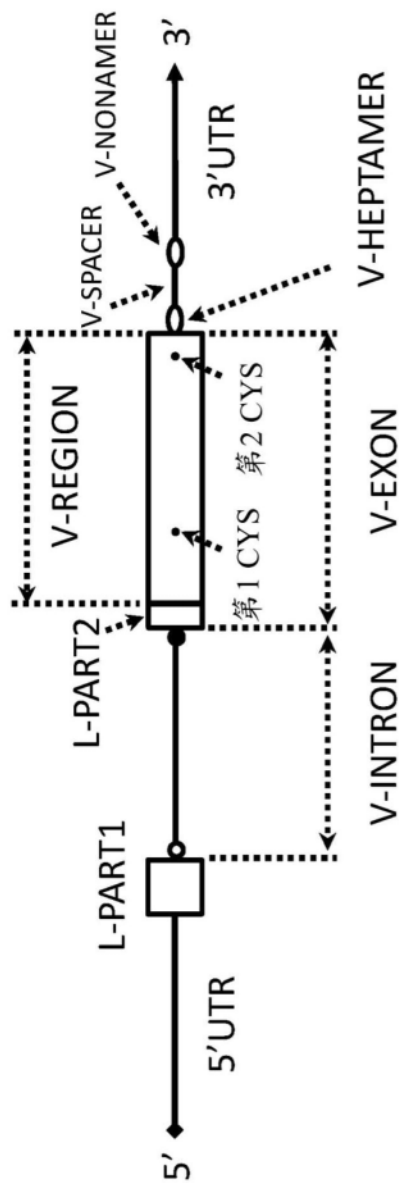


图 4A

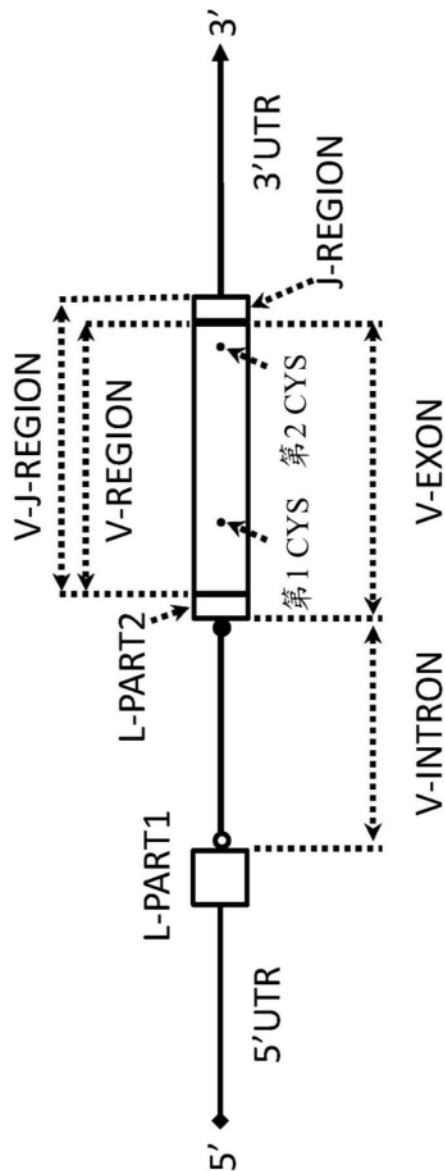


图 4B

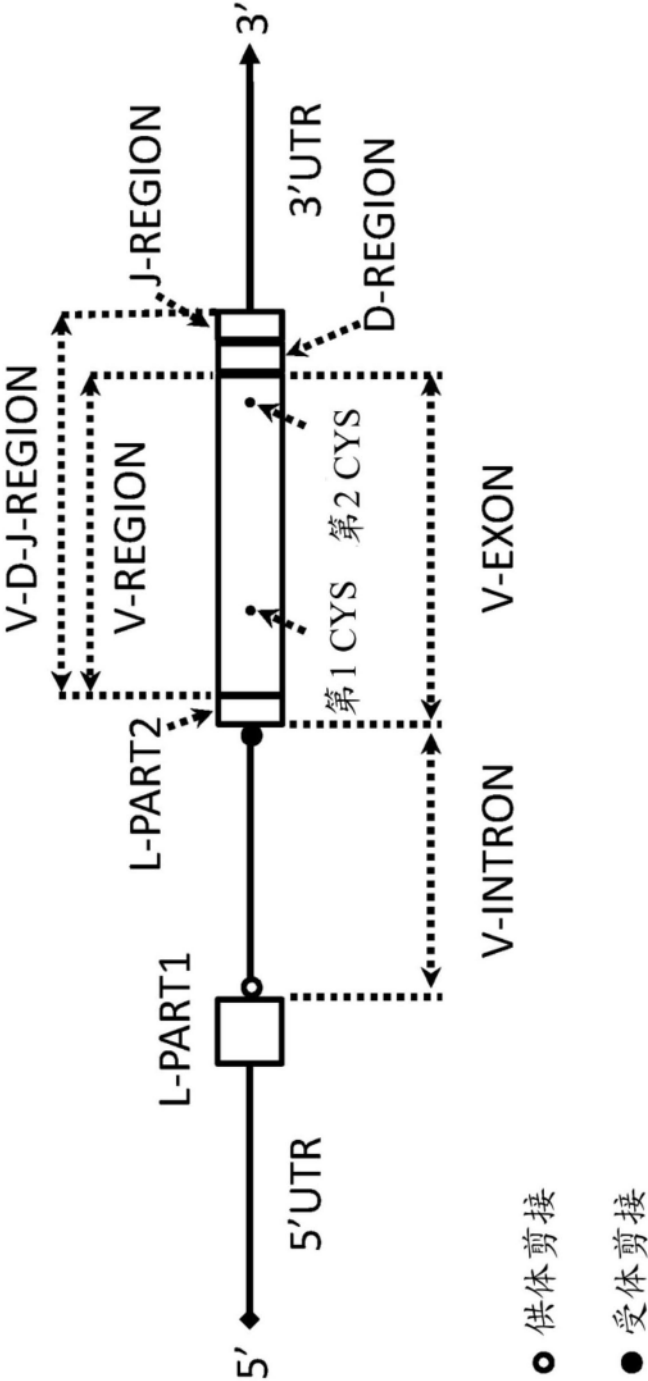


图4C

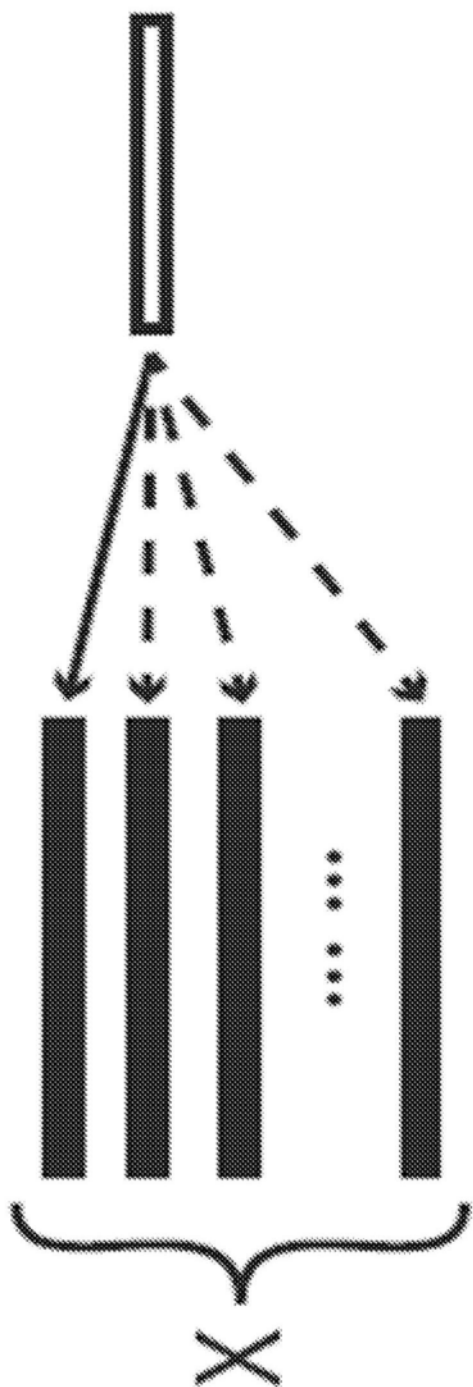


图5

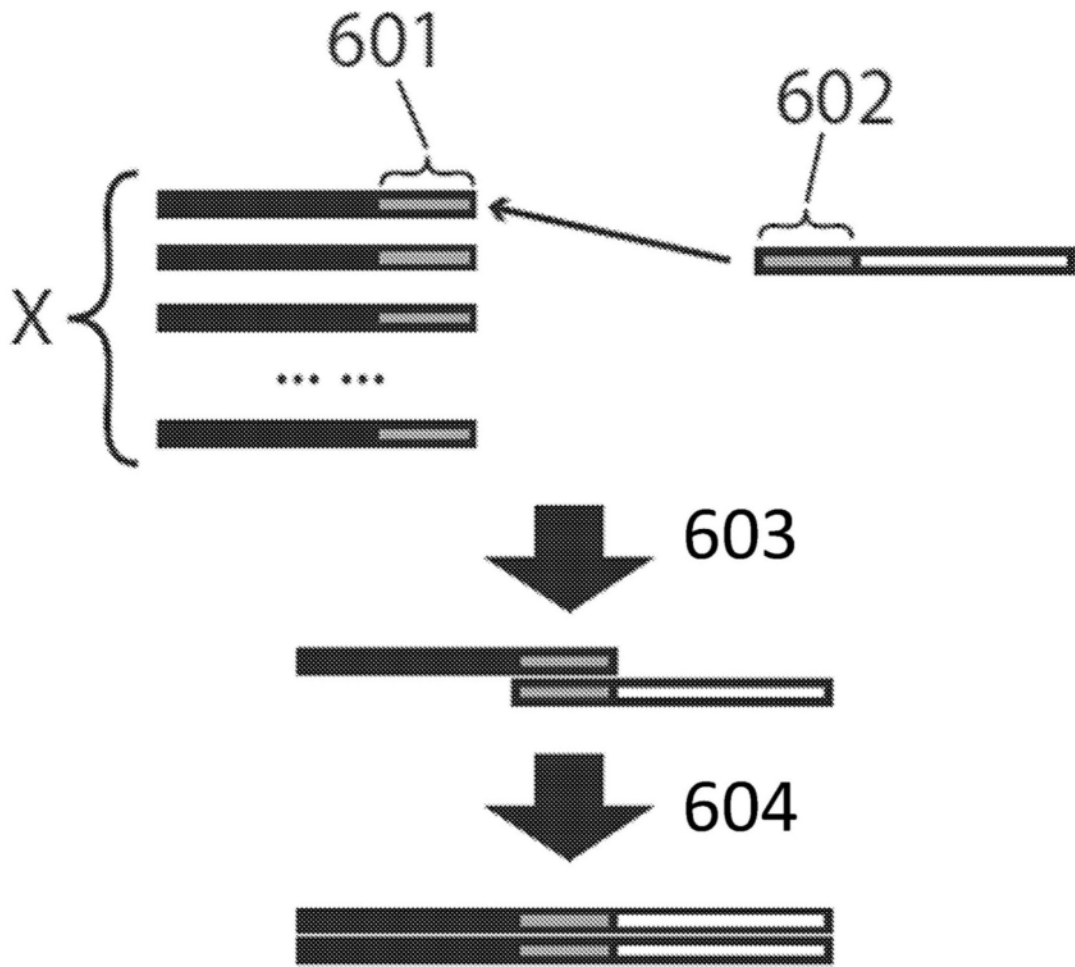


图6

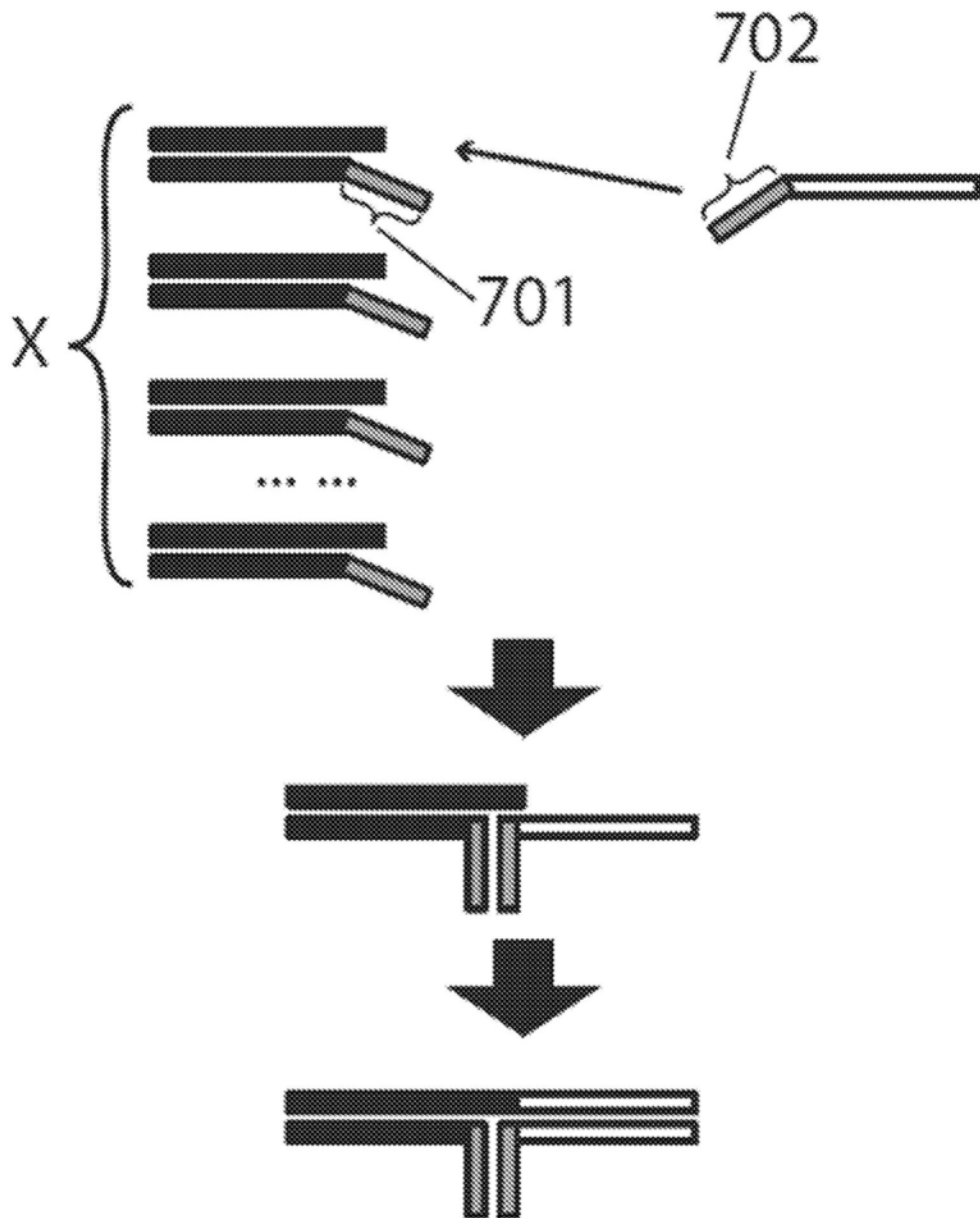


图7

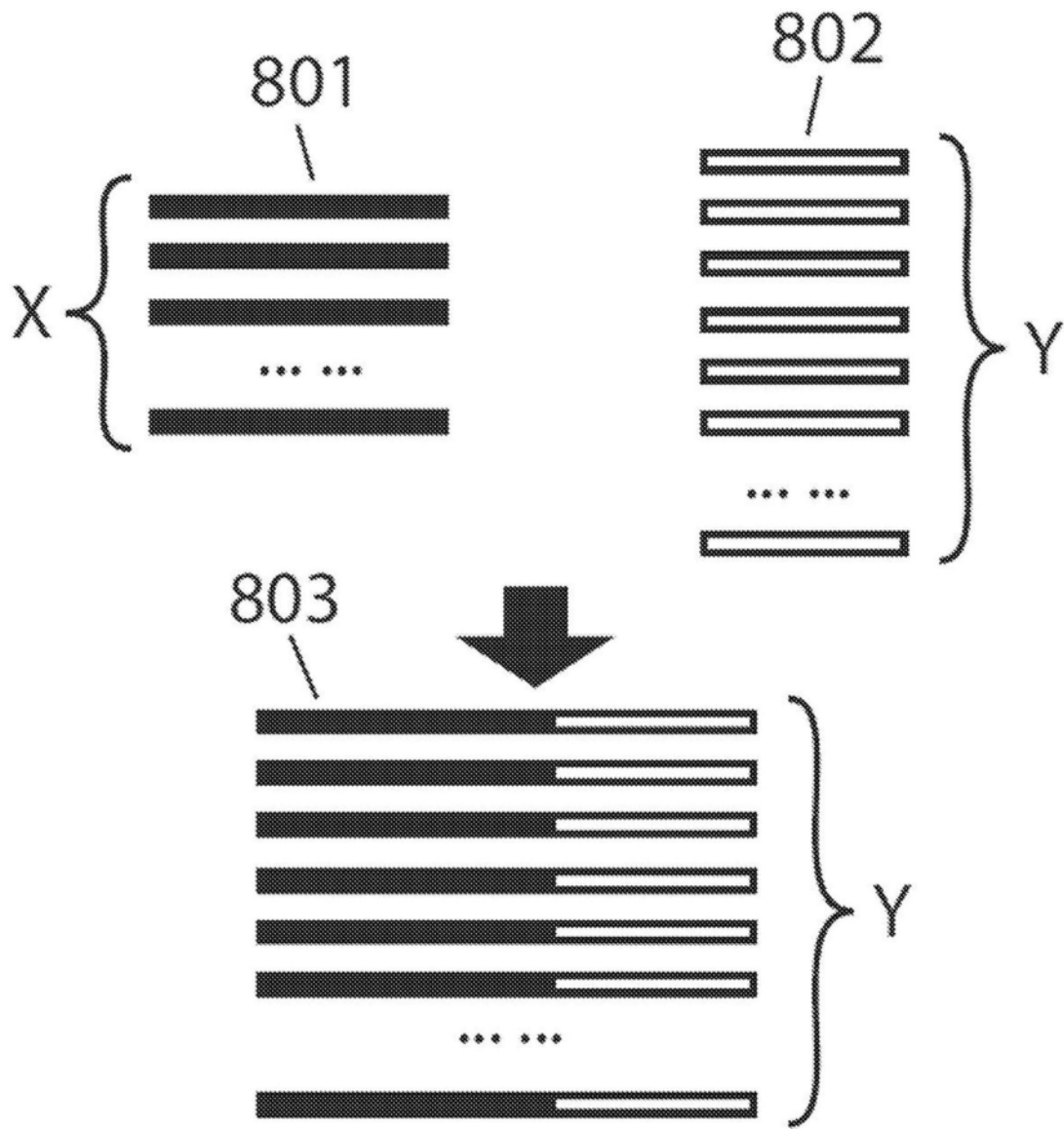


图8

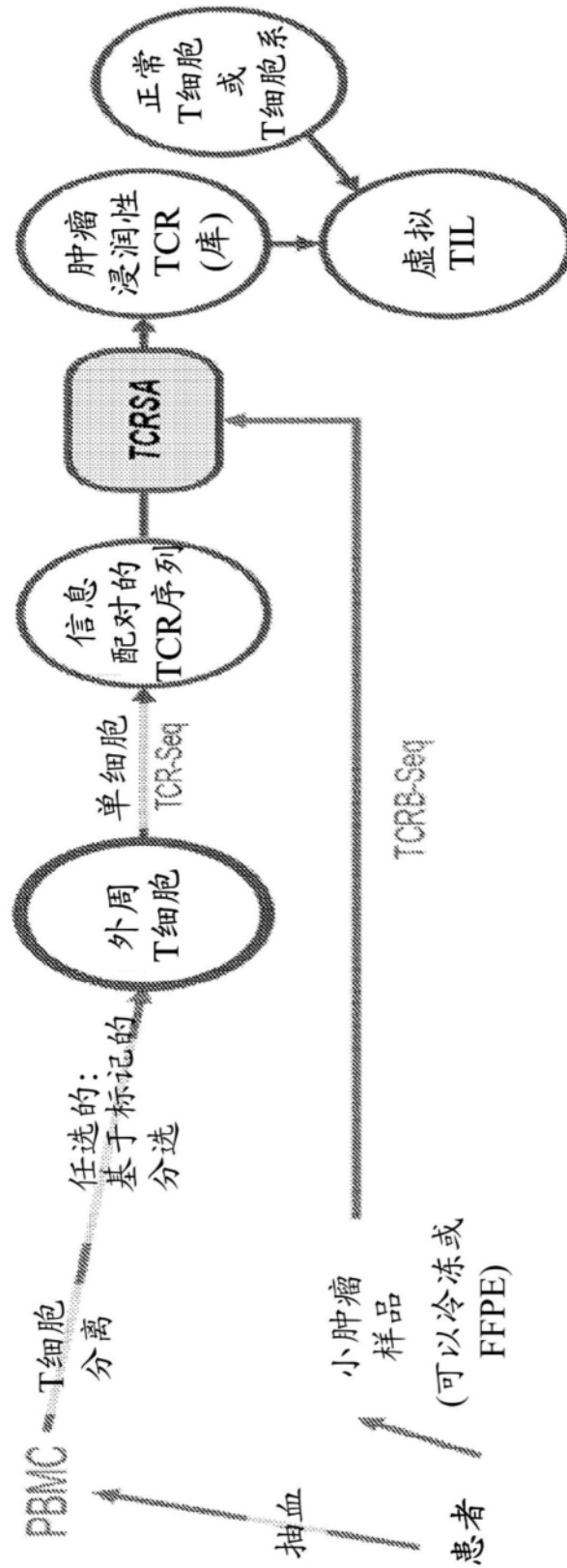


图9A

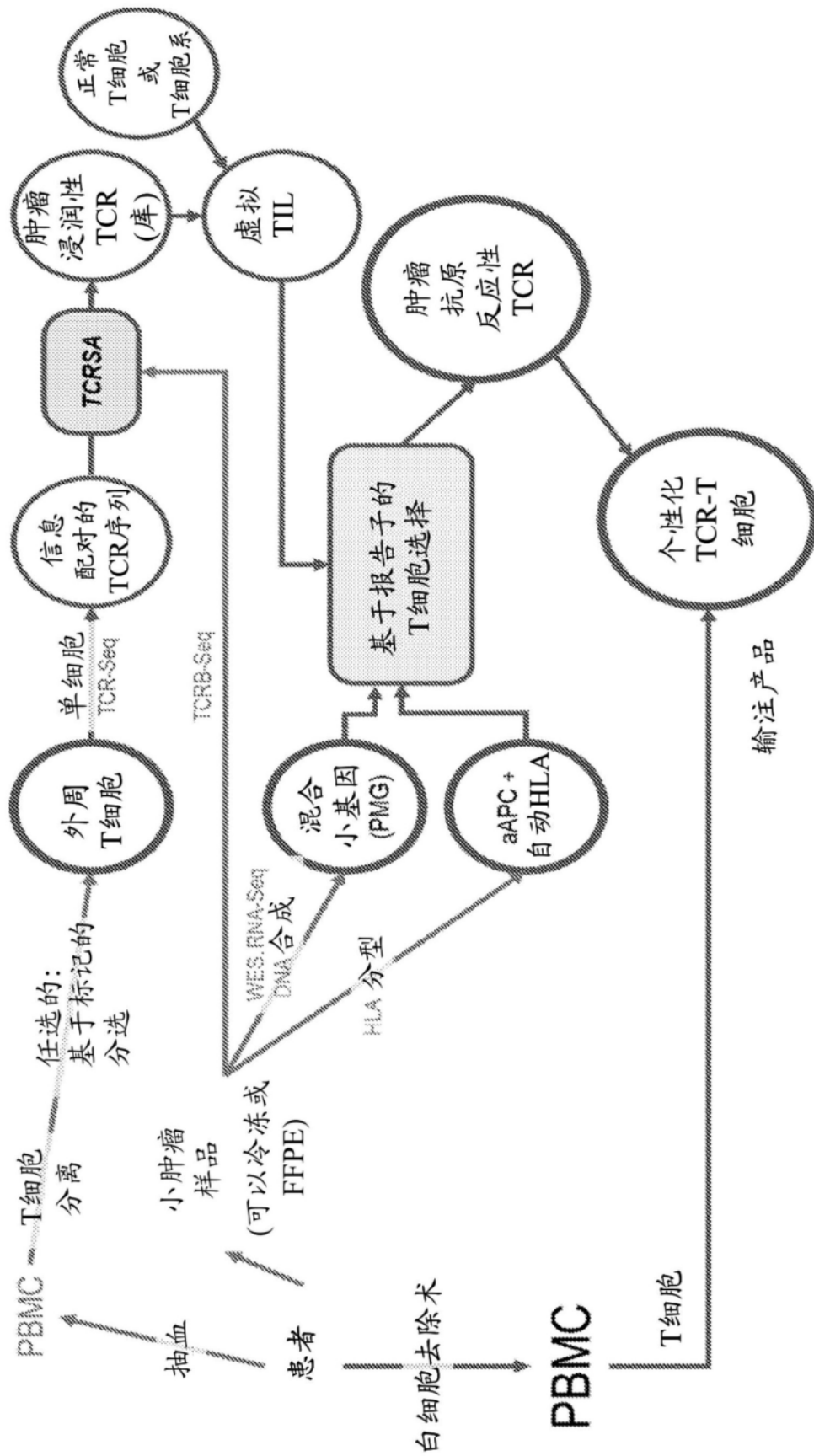


图9C

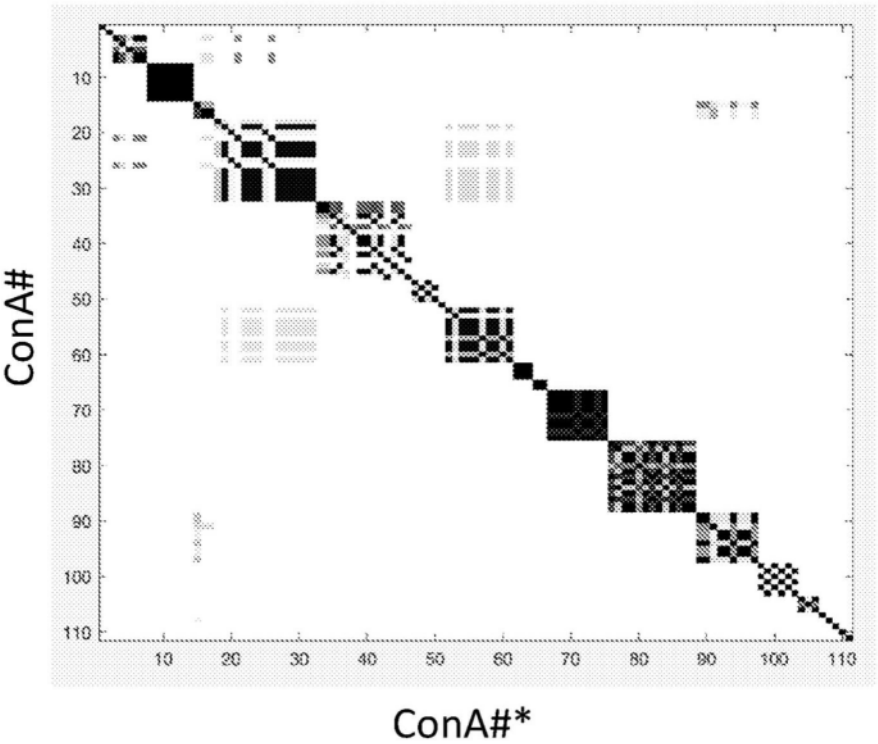


图10A

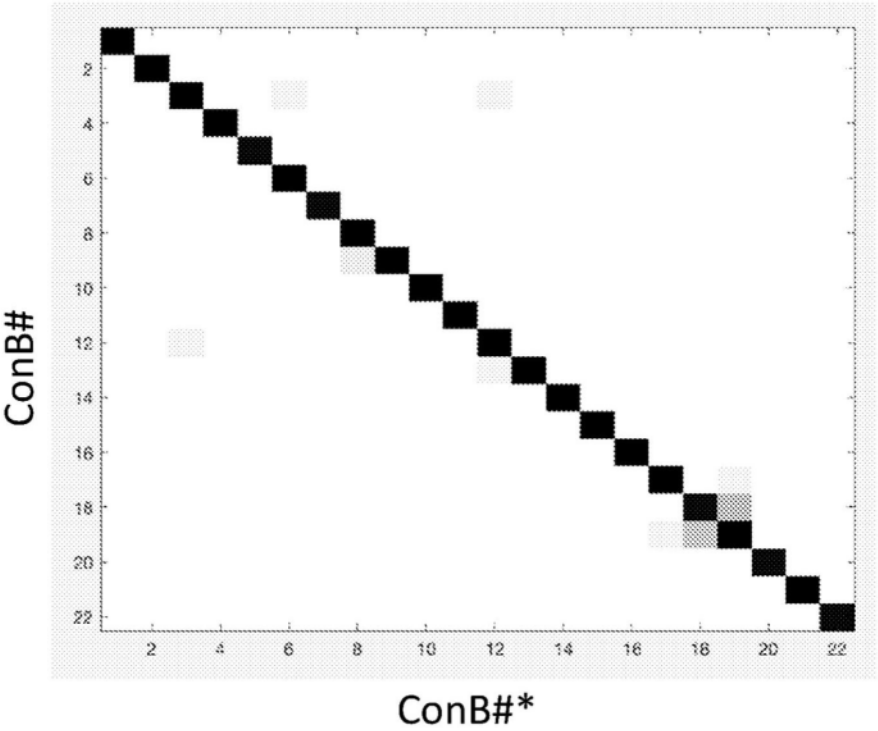


图10B

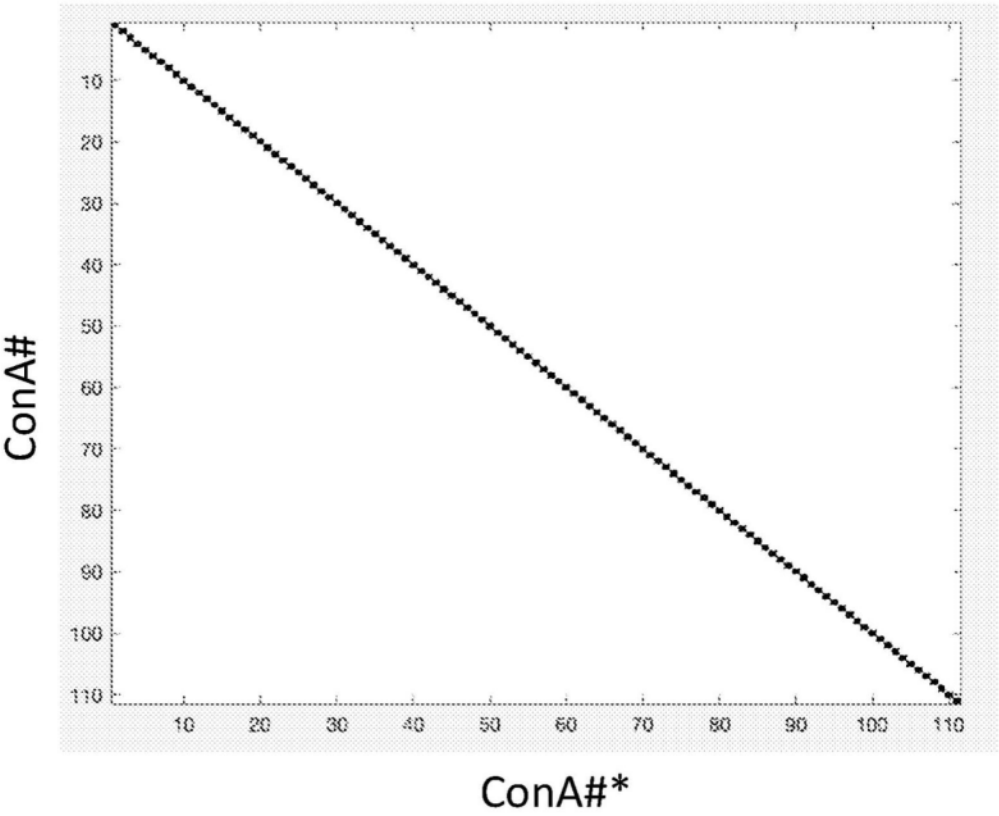


图11A

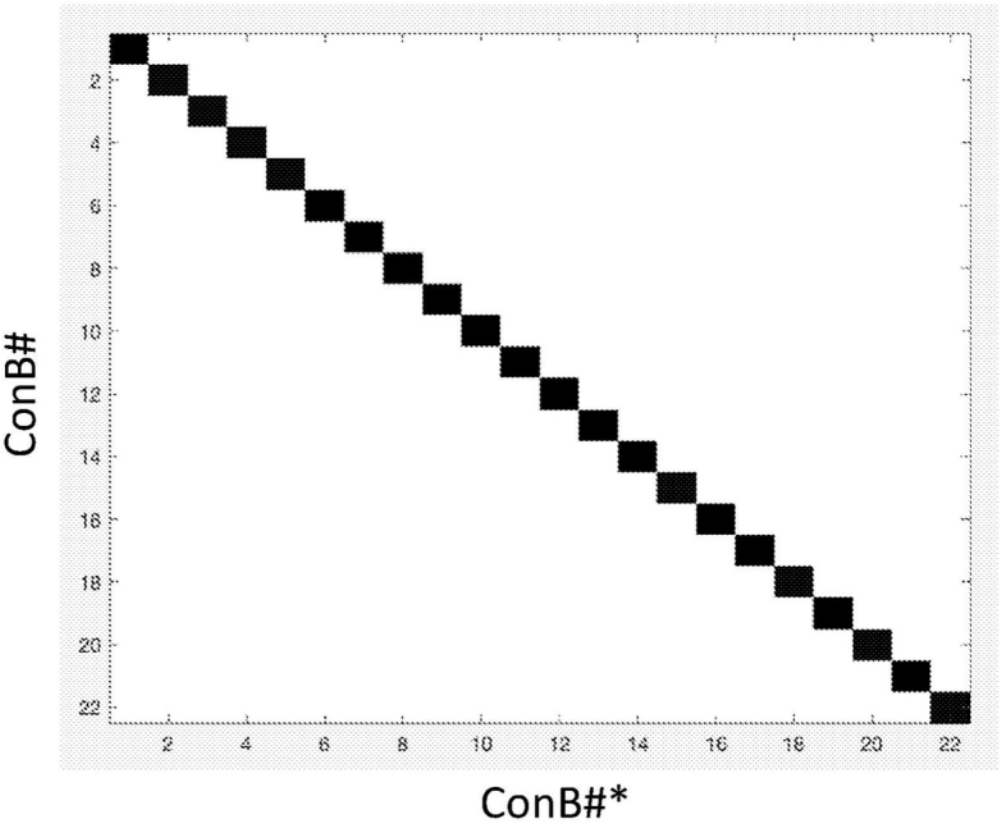


图11B

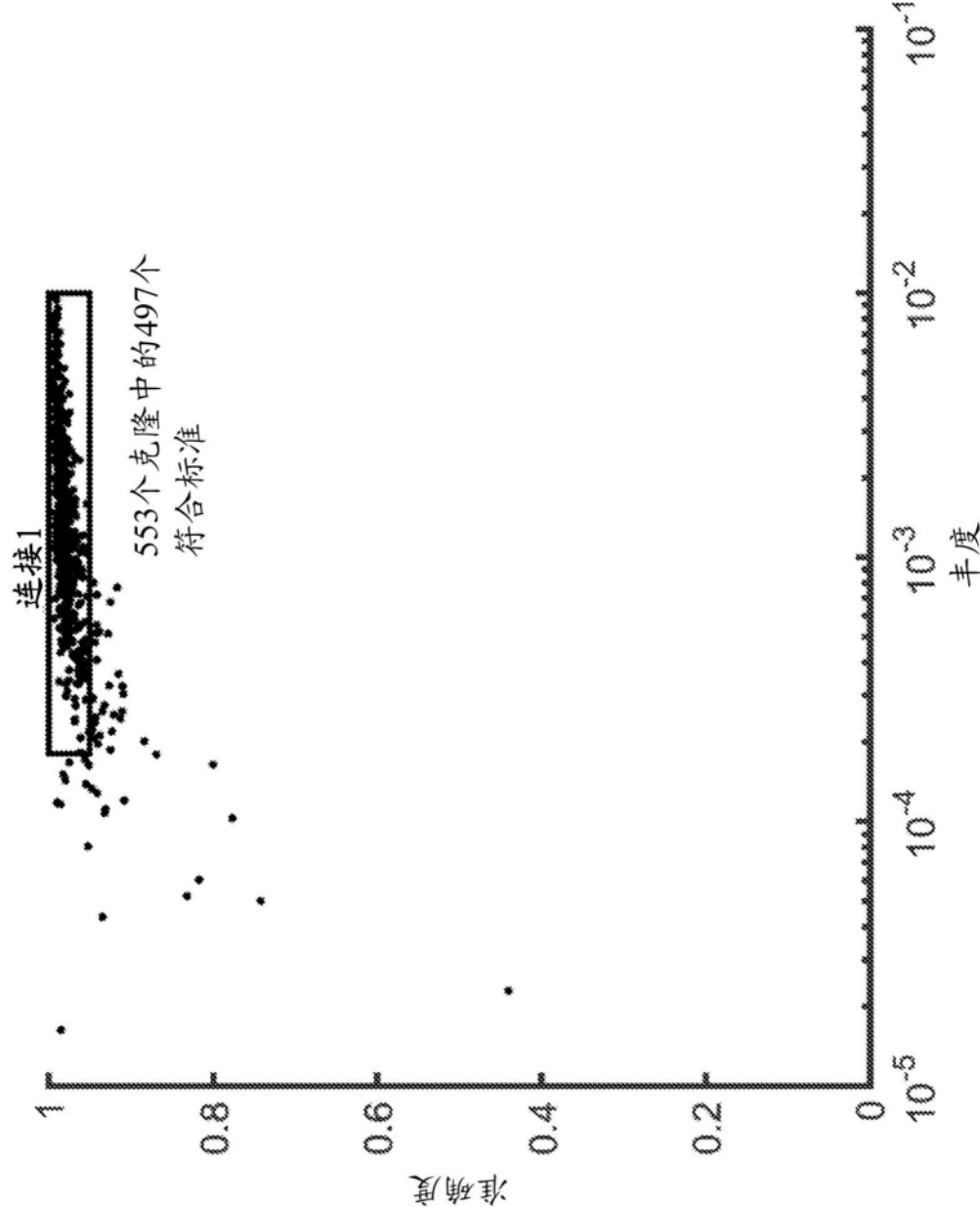


图12

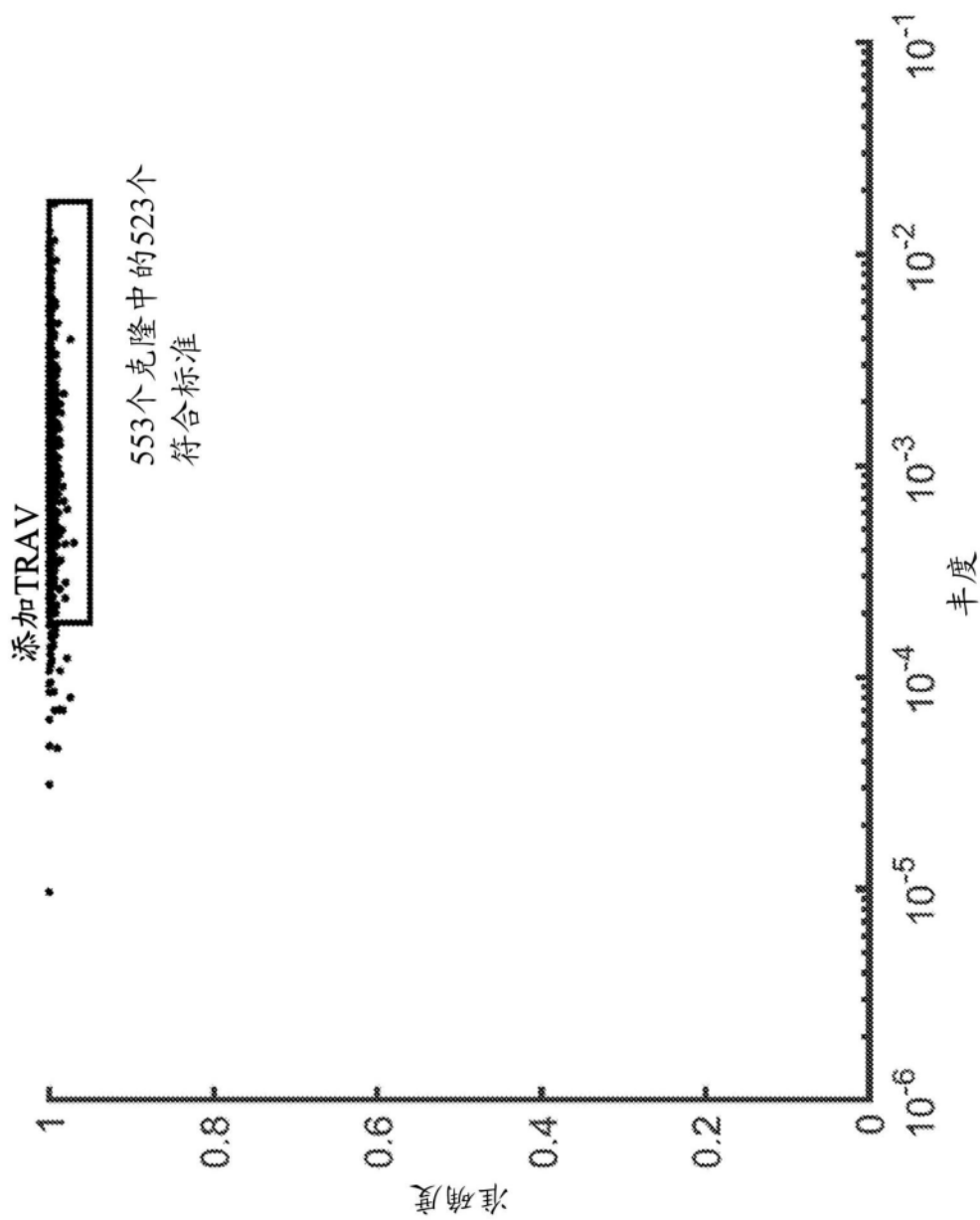


图13

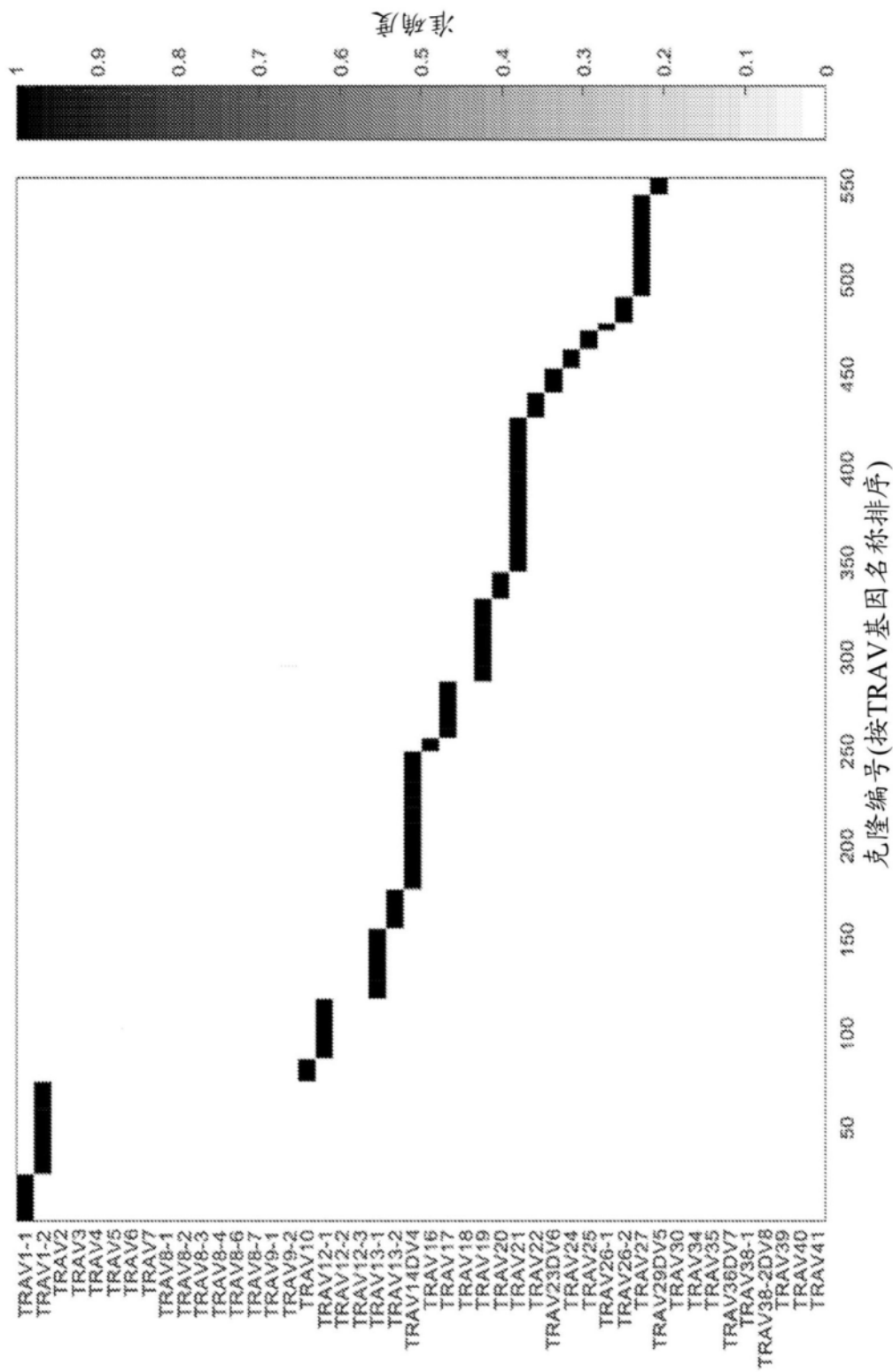


图14

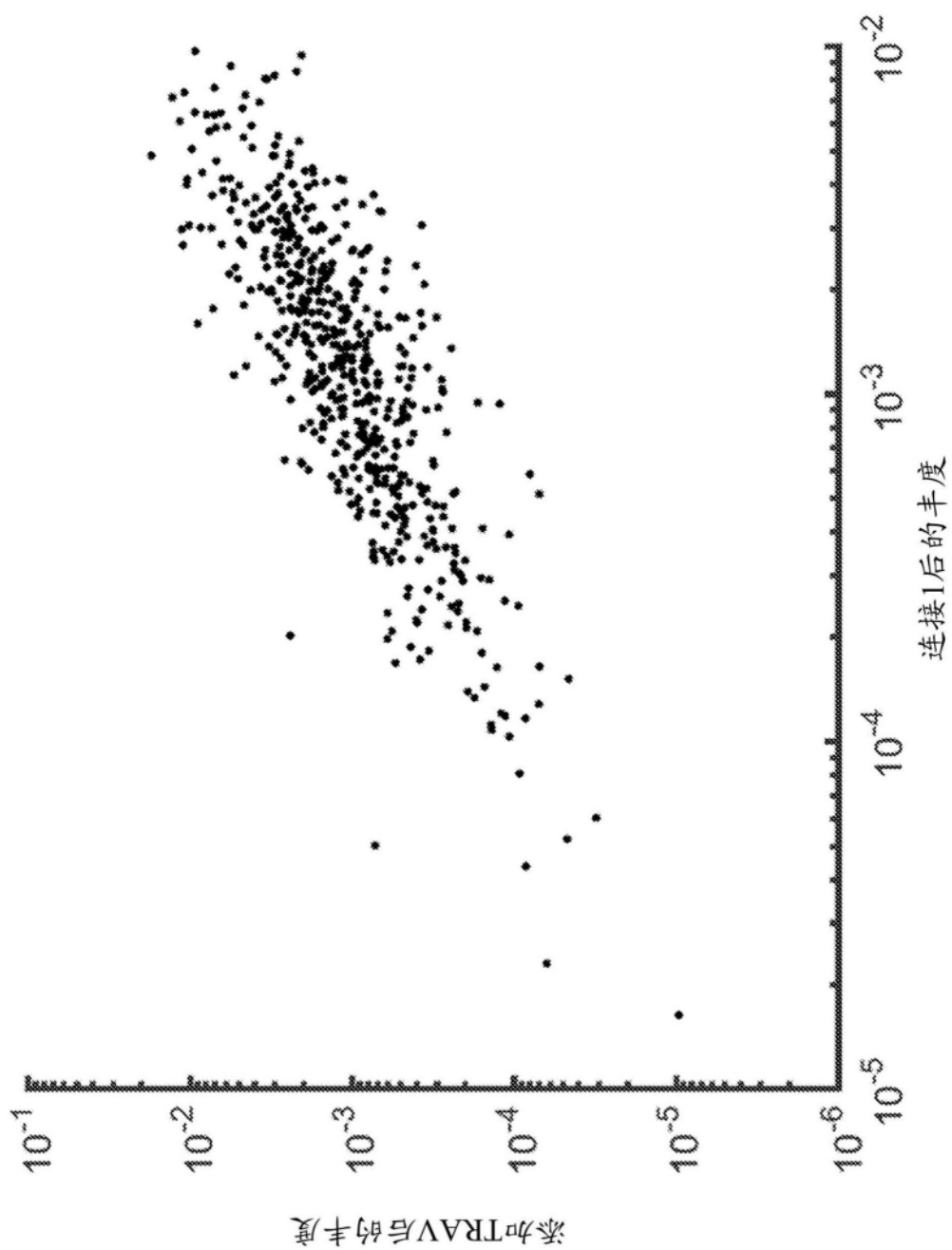


图15

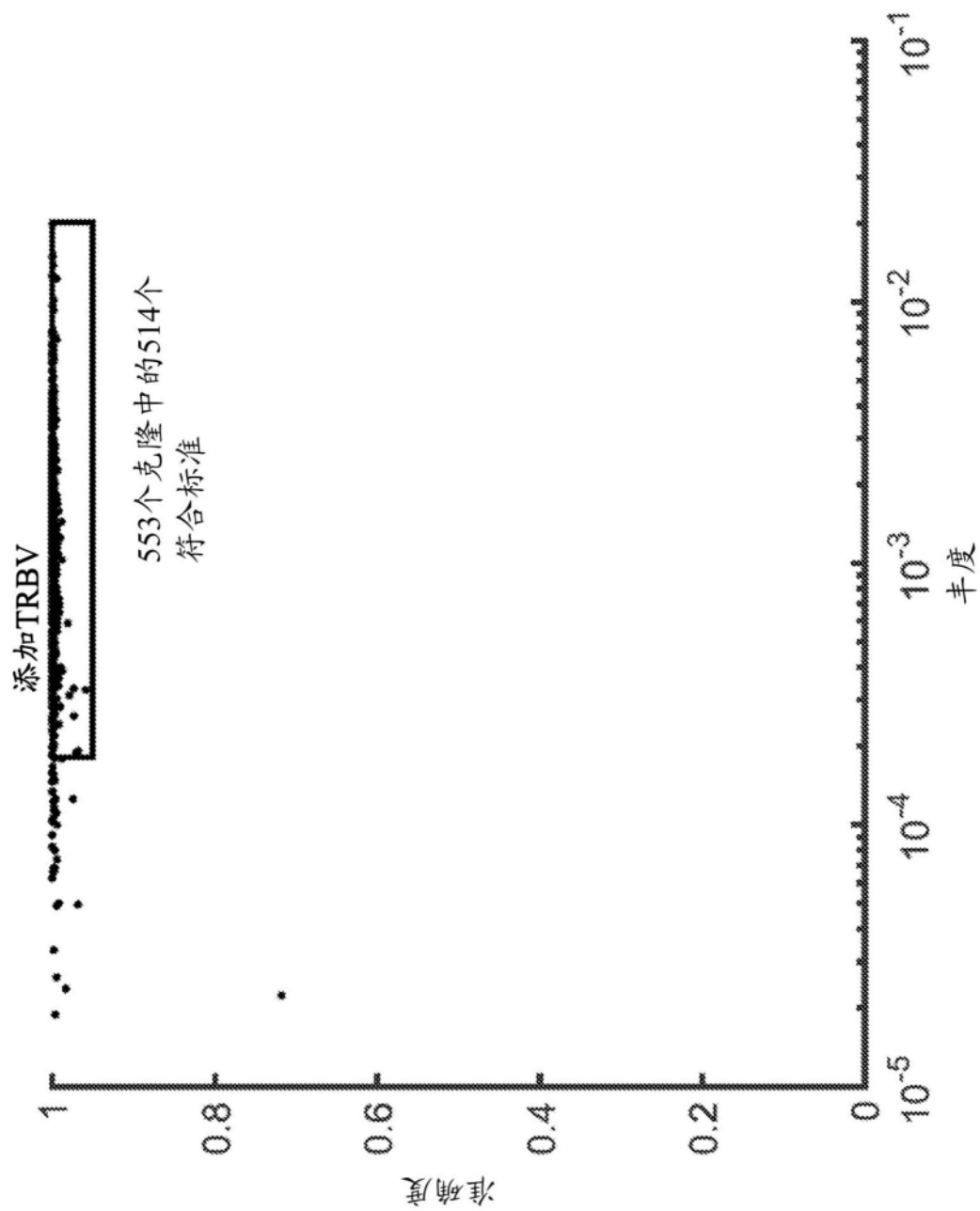


图16

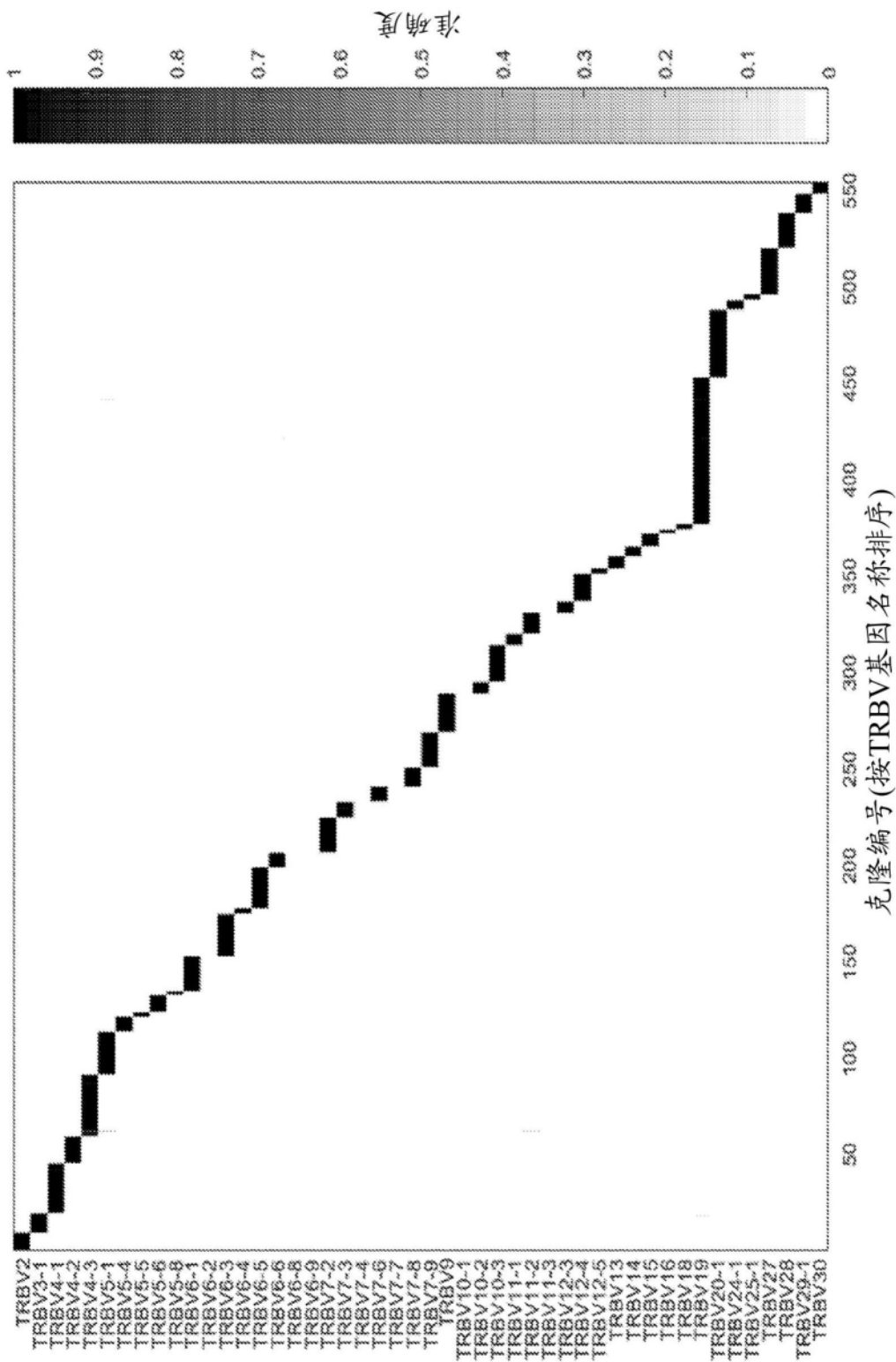


图17

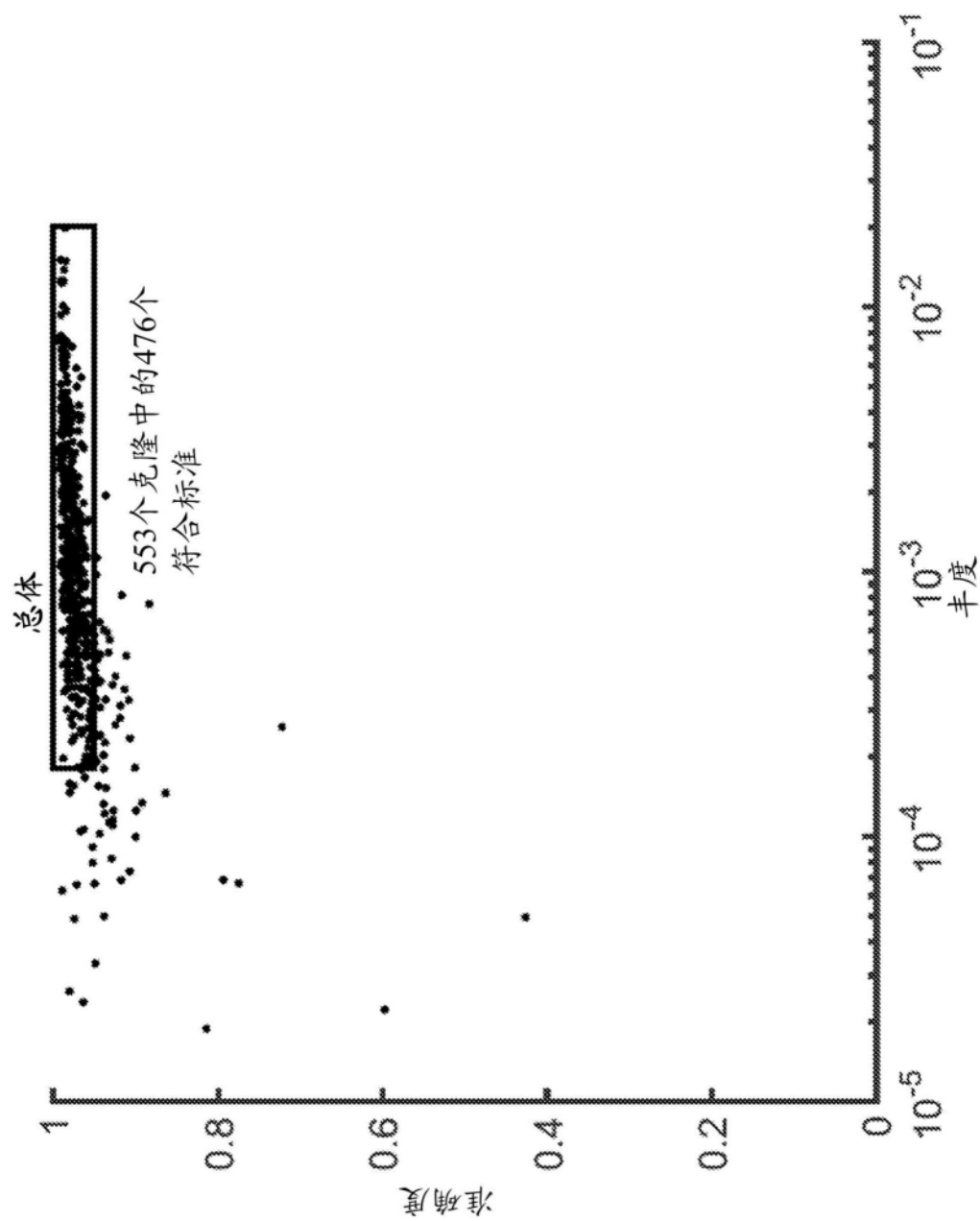


图18