



# (12)发明专利申请

(10)申请公布号 CN 109996876 A

(43)申请公布日 2019.07.09

(21)申请号 201780065373.9

(22)申请日 2017.04.05

(30)优先权数据

62/378,134 2016.08.22 US

62/382,191 2016.08.31 US

62/393,948 2016.09.13 US

62/411,388 2016.10.21 US

62/419,881 2016.11.09 US

(85)PCT国际申请进入国家阶段日

2019.04.22

(86)PCT国际申请的申请数据

PCT/US2017/026232 2017.04.05

(87)PCT国际申请的公布数据

W02018/038772 EN 2018.03.01

(71)申请人 特韦斯特生物科学公司

地址 美国加利福尼亚州

(72)发明人 安东尼·考克斯 陈思远

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 贺淑东

(51)Int.Cl.

C12N 15/10(2006.01)

C12N 15/90(2006.01)

权利要求书4页 说明书57页

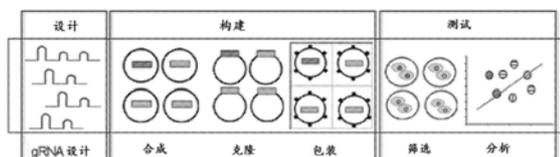
序列表11页 附图33页

(54)发明名称

从头合成的核酸文库

(57)摘要

本文公开了用于生成编码gRNA序列的寡核酸文库的方法。通过本文所述方法编码的gRNA可以是单gRNA序列或双gRNA序列。所述方法提供了以改善的准确性和均匀性生成gRNA文库作为DNA前体或作为RNA转录产物。



1. 一种核酸文库,其中所述核酸文库包含至少500个不相同的DNA分子,其中每个不相同的DNA分子编码不同的gRNA序列,并且其中所述至少500个不相同的DNA分子中的至少约80%以所述文库中每个所述不相同的DNA分子的平均频率的2倍以内的量各自存在于所述核酸文库中。

2. 根据权利要求1所述的核酸文库,其中每个不相同的DNA分子具有约20%至约85%的GC碱基含量。

3. 根据权利要求1所述的核酸文库,其中每个不相同的DNA分子具有约30%至约70%的GC碱基含量。

4. 根据权利要求1所述的核酸文库,其中所述至少500个不相同的DNA分子中的至少约90%以所述文库中每个所述不相同的DNA分子的所述平均频率的2倍以内的量各自存在于所述核酸文库中。

5. 根据权利要求1至3中任一项所述的核酸文库,其中所述至少500个不相同的DNA分子中的至少99%以所述文库中每个所述不相同的DNA分子的所述平均频率的2倍以内的量各自存在于所述核酸文库中。

6. 根据权利要求1所述的核酸文库,其中所述至少500个不相同的DNA分子包括至少2000个不相同的DNA分子。

7. 根据权利要求1所述的核酸文库,其中所述至少500个不相同的DNA分子包括至少3500个不相同的DNA分子。

8. 根据权利要求1所述的核酸文库,其中所述至少500个不相同的DNA分子包括至少100,000个不相同的DNA分子。

9. 根据权利要求1所述的核酸文库,其中每个不相同的DNA分子在长度上包含至多200个碱基。

10. 根据权利要求1至8中任一项所述的核酸文库,其中每个不相同的DNA分子在长度上包含约100至约200个碱基。

11. 根据权利要求1所述的核酸文库,其中所述至少500个不相同的DNA分子包含编码靶向生物途径中的基因的gRNA序列的不相同的DNA分子。

12. 根据权利要求1所述的核酸文库,其中所述至少500个不相同的DNA分子包含编码靶向整个基因组中的基因的gRNA序列的不相同的DNA分子。

13. 根据权利要求1所述的核酸文库,其中所述gRNA是单gRNA或双gRNA。

14. 一种核酸文库,其中所述核酸文库包含至少2000个不相同的核酸,其中每个不相同的核酸编码不同的sgRNA序列,其中每个sgRNA序列包含与真核基因互补的靶向结构域,并且其中所述至少2000个不相同的核酸中的至少约80%以所述文库中每个所述不相同的核酸的平均频率的2倍以内的量存在于所述核酸文库中。

15. 根据权利要求14所述的核酸文库,其中每个不相同的核酸具有约20%至约85%的GC碱基含量。

16. 根据权利要求14所述的核酸文库,其中每个不相同的核酸具有约30%至约70%的GC碱基含量。

17. 根据权利要求14所述的核酸文库,其中所述至少2000个不相同的核酸中的至少约90%以所述文库中每个所述不相同的核酸的所述平均频率的2倍以内的量各自存在于所述

核酸文库中。

18. 根据权利要求14至16中任一项所述的核酸文库,其中所述至少2000个不相同的核酸中的至少99%以所述文库中每个所述不相同的核酸的所述平均频率的2倍以内的量各自存在于所述核酸文库中。

19. 根据权利要求14所述的核酸文库,其中每个不相同的核酸在长度上包含至多200个碱基。

20. 根据权利要求14至18中任一项所述的核酸文库,其中每个不相同的核酸在长度上包含约100至约200个碱基。

21. 根据权利要求14所述的核酸文库,其中所述至少2000个不相同的核酸包含编码靶向生物途径中的基因的sgRNA序列的不相同的核酸。

22. 根据权利要求14所述的核酸文库,其中所述至少2000个不相同的核酸包含编码靶向整个基因组中的基因的sgRNA序列的不相同的核酸。

23. 根据权利要求14所述的核酸文库,其中每个不相同的核酸包括DNA或RNA分子。

24. 一种扩增子文库,其中所述扩增子文库包含多个不相同的DNA分子,其中每个不相同的DNA分子存在于扩增产物群体中,其中每个不相同的DNA分子编码不同的gRNA序列,并且其中所述多个不相同的DNA分子中的至少约80%以所述文库中每个所述不相同DNA分子的平均频率的2倍以内的量各自存在于所述扩增子文库中。

25. 根据权利要求24所述的扩增子文库,其中每个不相同的DNA分子具有约30%至约70%的GC碱基含量。

26. 根据权利要求24所述的扩增子文库,其中所述gRNA是单gRNA或双gRNA。

27. 一种细胞文库,其中所述细胞文库包含多个细胞群体,其中每个所述细胞群体包含编码不同gRNA序列的DNA分子,其中每个gRNA序列包含用于与基因结合的靶向区域,并且其中所述细胞群体中的至少15%在所述基因的表达方面具有至少2倍的消耗。

28. 根据权利要求27所述的细胞文库,其中所述细胞群体中的至少45%在所述基因的表达方面具有至少2倍的消耗。

29. 根据权利要求27所述的细胞文库,其中所述gRNA是单gRNA或双gRNA。

30. 根据权利要求27所述的细胞文库,其中所述多个细胞群体包含编码每个单个基因至少3个不同gRNA序列的DNA分子。

31. 根据权利要求27所述的细胞文库,其中所述多个细胞群体包含编码每个单个基因至少5个不同gRNA序列的DNA分子。

32. 根据权利要求27所述的细胞文库,其中所述多个细胞群体包含至少2000个细胞群体。

33. 根据权利要求27所述的细胞文库,其中所述多个细胞群体包含编码生物途径中的gRNA序列的DNA分子。

34. 根据权利要求27所述的细胞文库,其中所述多个细胞群体包含编码整个基因组中的gRNA序列的DNA分子。

35. 根据权利要求34所述的细胞文库,其中所述基因组是拟南芥、秀丽隐杆线虫、家犬、莱茵衣藻、斑马鱼、盘基网柄菌、黑腹果蝇、大肠杆菌、智人、猕猴、小家鼠、穴兔、褐家鼠、酿酒酵母或野猪。

36. 根据权利要求27所述的细胞文库,其中每个所述细胞群体包含原核细胞。
37. 根据权利要求27所述的细胞文库,其中每个所述细胞群体包含真核细胞。
38. 根据权利要求37所述的细胞文库,其中每个所述细胞群体包含哺乳动物细胞。
39. 根据权利要求27所述的细胞文库,其中每个所述细胞群体进一步包含外源核酸酶。
40. 根据权利要求27至39中任一项所述的细胞文库,其中所述DNA进一步包含载体序列。
41. 一种细胞文库,其中所述细胞文库包含多个细胞群体,其中每个所述细胞群体包含编码不同gRNA序列的DNA分子,其中每个gRNA序列包含用于与基因结合的靶向区域,并且其中所述细胞群体中的至多20%在所述基因的表达方面具有零或负消耗。
42. 根据权利要求41所述的细胞文库,其中所述gRNA是单gRNA或双gRNA。
43. 根据权利要求41所述的细胞文库,其中所述多个细胞群体包含编码每个单个基因至少3个不同gRNA序列的DNA分子。
44. 根据权利要求41所述的细胞文库,其中所述多个细胞群体包含编码每个单个基因至少5个不同gRNA序列的DNA分子。
45. 根据权利要求44所述的细胞文库,其中所述多个细胞群体包含至少2000个细胞群体。
46. 根据权利要求41至44中任一项所述的细胞文库,其中所述多个细胞群体包含至少10000个细胞群体。
47. 一种用于合成gRNA文库的方法,其包括:
- (a) 为至少500个不相同的DNA分子提供预定序列,其中每个不相同的DNA分子编码gRNA;
- (b) 合成所述至少500个不相同的DNA分子;以及
- (c) 转录所述至少500个不相同的DNA分子以生成gRNA文库,其中与所述至少500个不相同的DNA分子的所述预定序列相比,所述gRNA文库中至少约75%的所述gRNA没有错误。
48. 根据权利要求47所述的方法,进一步包括在所述转录步骤之前将所述至少500个不相同的DNA分子转移到细胞中。
49. 根据权利要求47所述的方法,其中由所述至少500个不相同的DNA分子编码的至少96%的所述gRNA存在于所述gRNA文库中。
50. 根据权利要求47所述的方法,其中与所述至少500个不相同的DNA分子的所述预定序列相比,所述gRNA文库中至少87%的所述gRNA没有错误。
51. 根据权利要求47所述的方法,进一步包括将所述至少500个不相同的DNA分子插入载体中。
52. 根据权利要求47所述的方法,进一步包括将所述至少500个不相同的DNA分子转移到生物体的细胞中。
53. 根据权利要求52所述的方法,其中所述生物体是拟南芥、秀丽隐杆线虫、家犬、莱茵衣藻、斑马鱼、盘基网柄菌、黑腹果蝇、大肠杆菌、智人、猕猴、小家鼠、穴兔、褐家鼠、酿酒酵母或野猪。
54. 根据权利要求47所述的方法,其中每个不相同的DNA分子编码单gRNA或双gRNA。
55. 一种用于合成gRNA文库的方法,包括:

- (a) 为多个不相同的DNA分子提供预定序列,其中每个不相同的DNA分子编码gRNA;
- (b) 提供表面,其中所述表面包含用于核酸延伸反应的座位簇;
- (c) 合成所述多个不相同的DNA分子,其中每个不相同的DNA分子从所述表面延伸;以及
- (d) 将所述多个不相同的DNA分子转移到细胞中。

56. 根据权利要求55所述的方法,其中每个簇包含约50至约500个座位。

57. 根据权利要求55所述的方法,其中每个不相同的DNA分子在长度上包含至多约200个碱基。

58. 根据权利要求55所述的方法,其中每个不相同的DNA分子编码单gRNA或双gRNA。

59. 根据权利要求55所述的方法,其中所述细胞为原核细胞。

60. 根据权利要求55所述的方法,其中所述细胞为真核细胞。

61. 根据权利要求55所述的方法,其中所述真核的为哺乳动物细胞。

62. 根据权利要求55所述的方法,其中每个所述细胞包含外源核酸酶。

## 从头合成的核酸文库

### 交叉引用

[0001] 本申请要求于2016年11月9日提交的美国临时申请号62/419,881、于2016年10月21日提交的美国临时申请号62/411,388、于2016年9月13日提交的美国临时申请号62/393,948、于2016年8月31日提交的美国临时申请号62/382,191以及于2016年8月22日提交的美国临时申请号62/378,134的权益,这些申请各自通过引用以其全文并入本文。

### 序列表

[0002] 本申请包含已经以ASCII格式电子提交的序列表,并且其通过引用以其全文并入本文。所述ASCII副本创建于2017年4月4日,命名为44854-727\_601\_SL.txt,并且大小为13,567字节。

### 发明背景

[0003] 合成生物学的基石是设计、构建和测试过程,是需要待产生的DNA能够快速且可行地生成并优化这些定制途径和生物体的迭代过程。在设计阶段,构成DNA的A、C、T和G核苷酸被规划成将包含感兴趣区域的多种序列,每种序列变体代表将进行测试的特定假设。这些变体序列代表序列空间的子集,序列空间这一概念起源于进化生物学,并且适合于构成基因、基因组、转录组和蛋白质组的全部序列。在靶向基因组编辑的背景下,需要快速生成高度准确和均匀的核酸文库,以供特异性地指导基因、基因簇、途径或整个基因组的酶促编辑。

### 援引并入

[0004] 本说明书中所提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别地且单独地指出每个单独的出版物、专利或专利申请通过引用以其全文并入。

### **发明内容**

[0005] 本文提供了用于有效从头合成和筛选高度准确的核酸文库的系统、方法和组合物。如本文所述的核酸文库包含用于特异性靶向和编辑基因、基因簇、生物途径或整个基因组的核酸。

[0006] 本文提供了核酸文库,其中所述核酸文库包含至少500个不相同的DNA分子,其中每个不相同的DNA分子编码不同的gRNA序列,并且其中所述至少500个不相同的DNA分子中的至少约80%以所述文库中每个所述不相同的DNA分子的平均频率的2倍以内的量各自存在于所述核酸文库中。本文提供了核酸文库,其中每个不相同的DNA分子具有约20%至约85%的GC碱基含量。本文提供了核酸文库,其中每个不相同的DNA分子具有约30%至约70%的GC碱基含量。本文提供了核酸文库,其中所述至少500个不相同的DNA分子中的至少约90%以所述文库中每个所述不相同的DNA分子的所述平均频率的2倍以内的量各自存在于所述核酸文库中。本文提供了核酸文库,其中所述至少500个不相同的DNA分子中的至少99%以所述文库中每个所述不相同的DNA分子的所述平均频率的2倍以内的量各自存在于所述核酸文库中。本文提供了核酸文库,其中所述至少500个不相同的DNA分子包括至少2000个不相同的DNA分子。本文提供了核酸文库,其中所述至少500个不相同的DNA分子包括

至少3500个不相同的DNA分子。本文提供了核酸文库,其中所述至少500个不相同的DNA分子包括至少100,000个不相同的DNA分子。本文提供了核酸文库,其中每个不相同的DNA分子在长度上包含至多200个碱基。本文提供了核酸文库,其中每个不相同的DNA分子在长度上包含约100至约200个碱基。本文提供了核酸文库,其中所述至少500个不相同的DNA分子包含编码靶向生物途径中的基因的gRNA序列的不相同的DNA分子。本文提供了核酸文库,其中所述至少500个不相同的DNA分子包含编码靶向整个基因组中的基因的gRNA序列的不相同的DNA分子。本文提供了核酸文库,其中所述gRNA是单gRNA或双gRNA。

[0007] 本文提供了核酸文库,其中所述核酸文库包含至少2000个不相同的核酸,其中每个不相同的核酸编码不同的sgRNA序列,其中每个sgRNA序列包含与真核基因互补的靶向结构域,并且其中所述至少2000个不相同的核酸中的至少约80%以所述文库中每个所述不相同的核酸的平均频率的2倍以内的量存在于所述核酸文库中。本文提供了核酸文库,其中每个不相同的核酸具有约20%至约85%的GC碱基含量。本文提供了核酸文库,其中每个不相同的核酸具有约30%至约70%的GC碱基含量。本文提供了核酸文库,其中所述至少2000个不相同的核酸中的至少约90%以所述文库中每个所述不相同的核酸的所述平均频率的2倍以内的量各自存在于所述核酸文库中。本文提供了核酸文库,其中所述至少2000个不相同的核酸中的至少99%以所述文库中每个所述不相同的核酸的所述平均频率的2倍以内的量各自存在于所述核酸文库中。本文提供了核酸文库,其中每个不相同的核酸在长度上包含至多200个碱基。本文提供了核酸文库,其中每个不相同的核酸在长度上包含约100至约200个碱基。本文提供了核酸文库,其中所述至少2000个不相同的核酸包含编码靶向生物途径中的基因的sgRNA序列的不相同的核酸。本文提供了核酸文库,其中所述至少2000个不相同的核酸包含编码靶向整个基因组中的基因的sgRNA序列的不相同的核酸。本文提供了核酸文库,其中每个不相同的核酸包括DNA或RNA分子。

[0008] 本文提供了扩增子文库,其中所述扩增子文库包含多个不相同的DNA分子,其中每个不相同的DNA分子存在于扩增产物群体中,其中每个不相同的DNA分子编码不同的gRNA序列,并且其中所述多个不相同的DNA分子中的至少约80%以所述文库中每个所述不相同DNA分子的平均频率的2倍以内的量各自存在于所述扩增子文库中。本文提供了扩增子文库,其中每个不相同的DNA分子具有约30%至约70%的GC碱基含量。本文提供了扩增子文库,其中所述gRNA是单gRNA或双gRNA。

[0009] 本文提供了细胞文库,其中所述细胞文库包含多个细胞群体,其中每个所述细胞群体包含编码不同gRNA序列的DNA分子,其中每个gRNA序列包含用于与基因结合的靶向区域,并且其中所述细胞群体中的至少15%在所述基因的表达方面具有至少2倍的消耗。本文提供了细胞文库,其中所述细胞群体中的至少45%在所述基因的表达方面具有至少2倍的消耗。本文提供了细胞文库,其中所述gRNA是单gRNA或双gRNA。本文提供了细胞文库,其中所述多个细胞群体包含编码每个单个基因至少3个不同gRNA序列的DNA分子。本文提供了细胞文库,其中所述多个细胞群体包含编码每个单个基因至少5个不同gRNA序列的DNA分子。本文提供了细胞文库,其中所述多个细胞群体包含至少2000个细胞群体。本文提供了细胞文库,其中所述多个细胞群体包含编码生物途径中的gRNA序列的DNA分子。本文提供了细胞文库,其中所述多个细胞群体包含编码整个基因组中的gRNA序列的DNA分子。本文提供了细胞文库,其中所述基因组是拟南芥(*Arabidopsis thaliana*)、秀丽隐杆线虫

(*Caenorhabditis elegans*)、家犬 (*Canis lupus familiaris*)、莱茵衣藻 (*Chlamydomonas reinhardtii*)、斑马鱼 (*Danio rerio*)、盘基网柄菌 (*Dictyostelium discoideum*)、黑腹果蝇 (*Drosophila melanogaster*)、大肠杆菌 (*Escherichia coli*)、智人 (*Homo sapiens*)、猕猴 (*Macaca mulatta*)、小家鼠 (*Mus musculus*)、穴兔 (*Oryctolagus cuniculus*)、褐家鼠 (*Rattus norvegicus*)、酿酒酵母 (*Saccharomyces cerevisiae*) 或野猪 (*Sus scrofa*)。本文提供了细胞文库,其中每个所述细胞群体包含原核细胞。本文提供了细胞文库,其中每个所述细胞群体包含真核细胞。本文提供了细胞文库,其中每个所述细胞群体包含哺乳动物细胞。本文提供了细胞文库,其中每个所述细胞群体进一步包含外源核酸酶。本文提供了细胞文库,其中所述DNA分子进一步包含载体序列。

[0010] 本文提供了细胞文库,其中所述细胞文库包含多个细胞群体,其中每个所述细胞群体包含编码不同gRNA序列的DNA分子,其中每个gRNA序列包含用于与基因结合的靶向区域,并且其中所述细胞群体中的至多20%在所述基因的表达方面具有零或负消耗。本文提供了细胞文库,其中所述gRNA是单gRNA或双gRNA。本文提供了细胞文库,其中所述多个细胞群体包含编码每个单个基因至少3个不同gRNA序列的DNA分子。本文提供了细胞文库,其中所述多个细胞群体包含编码每个单个基因至少5个不同gRNA序列的DNA分子。本文提供了细胞文库,其中所述多个细胞群体包含至少2000个细胞群体。本文提供了细胞文库,其中所述多个细胞群体包含至少10000个细胞群体。

[0011] 本文提供了用于合成gRNA文库的方法,包括:为至少500个不相同的DNA分子提供预定序列,其中每个不相同的DNA分子编码gRNA;合成所述至少500个不相同的DNA分子;以及转录所述至少500个不相同的DNA分子以生成gRNA文库,其中与所述至少500个不相同的DNA分子的所述预定序列相比,所述gRNA文库中至少约75%的所述gRNA没有错误。本文提供了用于合成gRNA文库的方法,进一步包括在所述转录步骤之前将所述至少500个不相同的DNA分子转移到细胞中。本文提供了用于合成gRNA文库的方法,其中由所述至少500个不相同的DNA分子编码的至少96%的所述gRNA存在于所述gRNA文库中。本文提供了用于合成gRNA文库的方法,其中与所述至少500个不相同的DNA分子的所述预定序列相比,所述gRNA文库中至少87%的所述gRNA没有错误。本文提供了用于合成gRNA文库的方法,进一步包括将所述至少500个不相同的DNA分子插入载体中。本文提供了用于合成gRNA文库的方法,进一步包括将所述至少500个不相同的DNA分子转移到生物体的细胞中。本文提供了用于合成gRNA文库的方法,其中所述生物体是拟南芥、秀丽隐杆线虫、家犬、莱茵衣藻、斑马鱼、盘基网柄菌、黑腹果蝇、大肠杆菌、智人、猕猴、小家鼠、穴兔、褐家鼠、酿酒酵母或野猪。本文提供了用于合成gRNA文库的方法,其中每个不相同的DNA分子编码单gRNA或双gRNA。

[0012] 本文提供了用于合成gRNA文库的方法,包括:为多个不相同的DNA分子提供预定序列,其中每个不相同的DNA分子编码gRNA;提供表面,其中所述表面包含用于核酸延伸反应的座位簇;合成所述多个不相同的DNA分子,其中每个不相同的DNA分子从所述表面延伸;以及将所述多个不相同的DNA分子转移到细胞中。本文提供了用于合成gRNA文库的方法,其中每个簇包含约50至约500个座位。本文提供了用于合成gRNA文库的方法,其中每个不相同的DNA分子在长度上包含至多约200个碱基。本文提供了用于合成gRNA文库的方法,其中每个不相同的DNA分子编码单gRNA或双gRNA。本文提供了用于合成gRNA文库的方法,其中所述细胞为原核细胞。本文提供了用于合成gRNA文库的方法,其中所述细胞为真核细胞。本文提供

了用于合成gRNA文库的方法,其中所述真核的为哺乳动物细胞。本文提供了用于合成gRNA文库的方法,其中每个所述细胞包含外源核酸酶。

### 附图说明

[0013] 图1A图示了成簇的规律间隔短回文重复 (CRISPR) 复合物,其包含以下组分:PAM、靶序列、CAS9酶、指导RNA (gRNA) 和供体DNA。

[0014] 图1B图示了成簇的规律间隔短回文重复 (CRISPR) 复合物,其包含以下组分:用于非同源末端连接修复 (NHEJ) 途径的PAM、靶序列、CAS9酶、指导RNA (gRNA) 和供体DNA。

[0015] 图2图示了gRNA文库筛选工作流程,包括gRNA文库的设计、合成、克隆、包装、筛选和分析。

[0016] 图3图示了用于构建文库的gRNA文库筛选工作流程,包括:在阵列上合成寡核酸文库,将寡核酸扩增并转移到载体中,以及形成用于gRNA表达的表达文库。

[0017] 图4A-图4C是各种gRNA的图。图4A是具有碱基配对区域、dCas9手柄 (handle) 和酿脓链球菌 (*S. pyogenes*) 终止子区域的sgRNA序列 (SEQ ID NO:40) 的图。图4B是单独的sgRNA的图。图4C是单独的dgRNA的图。

[0018] 图5A是模板链靶向布置中的sgRNA序列的图。

[0019] 图5B是非模板链靶向布置中的sgRNA序列的图。

[0020] 图6A是具有T7启动子的gRNA序列的图,该T7启动子在转录时产生形成发夹二级结构的gRNA序列。

[0021] 图6B是具有T7启动子的gRNA序列的图,该T7启动子在转录时产生不形成发夹二级结构的gRNA序列。

[0022] 图7描绘了用于体外Cas9介导的靶DNA切割的工作流程。

[0023] 图8图示了计算机系统的示例。

[0024] 图9是图示计算机系统的示例性架构的框图。

[0025] 图10是说明网络的示图,该网络被配置用于并入多个计算机系统、多个蜂窝电话和个人数据助理,以及网络附加存储 (NAS)。

[0026] 图11是使用共享虚拟地址存储空间的多处理器计算机系统的框图。

[0027] 图12描绘了4种sgRNA设计。图12按照出现顺序分别公开了SEQ ID NO 20、15、21、15、22、15、41、15、42和43。

[0028] 图13A-图13B是来自BioAnalyzer读数的图,其中X轴上为核苷酸碱基并且Y轴上为荧光单位。

[0029] 图14A-图14J是来自BioAnalyzer读数的图,其中X轴上为核苷酸碱基并且Y轴上为荧光单位。

[0030] 图15是256个簇的图像,每个簇具有121个座位,寡核酸从该座位延伸。

[0031] 图16A是来自从每个簇具有121个寡核酸的240个簇合成29,040个独特寡核酸的板上的寡核酸表示 (寡核酸频率与吸光度) 的图。

[0032] 图16B是每个单独簇的寡核酸频率与吸光度的测量图,其中对照簇由框标识。

[0033] 图17是四个单独簇中的寡核酸频率与吸光度的测量图。

[0034] 图18A是来自从每个簇具有121个寡核酸的240个簇合成29,040个独特寡核酸的板

上的错误率与频率的图。

[0035] 图18B是每个单独簇的寡核酸错误率与频率的测量图,其中对照簇由框标识。

[0036] 图19是四个簇中的寡核酸错误率与频率的测量图。

[0037] 图20是GC含量作为每个寡核酸的百分比与寡核酸数目的量度的图。

[0038] 图21提供了来自具有两种不同聚合酶的PCR的结果的图。每个图表描绘了“观察频率”(“0至35”,测量为每100,000个的计数)与寡核酸的数目(0至2000)。

[0039] 图22提供了记录的扩增后寡核酸群体均匀性的定量的图表。

[0040] 图23描绘了过度扩增对序列丢失的影响图。

[0041] 图24A-图24B描绘了对来自10,000个sgRNA寡核酸CRISPR文库的回收寡核酸进行测序的结果。

[0042] 图25描绘了对来自101,000个sgRNA寡核酸CRISPR文库的回收寡核酸进行测序的结果。

[0043] 图26A描绘了具有至少2倍消耗的sgRNA的百分比图。

[0044] 图26B描绘了具有零或负消耗的sgRNA的百分比图。

### 具体实施方式

[0045] 本文提供了用于有效合成和筛选高度准确的指导RNA(“gRNA”)文库的系统、方法和组合物。本文所述的从头合成方法提供了较大gRNA文库的快速且高度准确的生成,用于掺入酶系统中以供靶向基因编辑。

[0046] 定义

[0047] 在整个本公开内容中,各个实施方案以范围格式给出。应当理解,范围格式的描述只是为了方便和简明起见,而不应被解释为对任何实施方案范围的硬性限制。相应地,对范围的描述应被认为明确公开了所有可能的子范围以及该范围内精确到下限单位十分之一的单个数值,除非上下文另有明确规定。例如,诸如从1至6的范围描述应被认为已经明确公开了诸如从1至3、从1至4、从1至5、从2至4、从2至6、从3至6等子范围,以及该范围内的单个值,例如,1.1、2、2.3、5和5.9。无论范围的宽度如何,这都是适用的。这些中间范围的上限和下限可独立地包括在更小的范围内,并且也被涵盖于本发明之中,受制于所声称范围中的任何被明确排除的限值。当所声称范围包括限值中之一或全部两者时,本发明中也包括排除了这些所包括的限值中之一或全部二者的范围,除非上下文另有明确规定。

[0048] 本文使用的术语仅用于描述特定实施方案的目的,而不旨在限制任何实施方案。除非上下文另有明确规定,否则如本文所用的单数形式“一个”、“一种”和“该”也预期包括复数形式。将进一步理解的是,术语“包括”和/或“包含”在本说明书中使用指定所述特征、整体、步骤、操作、元件和/或组件的存在,但不排除存在或添加一个或多个特征、整体、步骤、操作、元件、组件和/或其群体。如本文所用,术语“和/或”包括一个或多个相关所列项目的任何和所有组合。

[0049] 除非特别说明或从上下文中明显看出,否则如本文所用,提及数字或数字范围的术语“约”理解为意指所述数字及其 $\pm 10\%$ 的数字,或者对于列出的范围值,意指低于所列下限的10%和高于所列上限的10%。

[0050] 本文提及的术语“gRNA”是指指导RNA序列并包括单指导RNA和双指导RNA序列。除

非特别说明或从上下文显而易见,否则如本文所用,本文提及的术语“dgRNA”是指双指导RNA序列:crRNA(包含与靶序列互补的种子区域的间隔序列)和单独的tracrRNA(反式激活序列),它们是部分互补的RNA。除非特别说明或从上下文显而易见,否则如本文所用,本文提及的术语“sgRNA”是指单指导RNA序列,其包含融合的crRNA和tracrRNA两者。

[0051] 除非特别说明或从上下文显而易见,否则如本文所用,术语“寡核酸”和“核酸”包括双链或三链核酸,以及单链分子。在双链或三链核酸中,核酸链无需是共延伸的(即,双链核酸无需沿两条链的整个长度是双链的)。除非另有说明,否则当提供时,核酸序列以5'至3'方向列出。本文所述的方法提供了分离的核酸的生成。本文所述的方法另外提供了分离的和纯化的核酸的生成。本文提及的“寡核酸”和“核酸”可在长度上包含至少5、10、20、30、40、50、60、70、80、90、100、125、150、175、200、225、250、275、300、325、350、375、400、425、450、475、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000个或更多个碱基。

[0052] 除非特别说明或从上下文显而易见,否则如本文所用,本文所用的术语“扩增子”是指扩增反应产物。

[0053] 除非另有说明,否则本公开内容采用本领域技术范围内的常规分子生物学技术。除非另外定义,否则本文所用的所有技术和科学术语具有与这些发明所属领域的技术人员所通常理解的含义。

[0054] gRNA文库筛选

[0055] 本文提供了用于设计、构建和筛选高度准确的gRNA文库的方法,所述gRNA用于掺入成簇的规律间隔短回文重复(CRISPR)-酶复合物中。参见例如,图1A-图1B。使用本文所述方法生成的gRNA文库包括sgRNA和dgRNA文库。本文提供了用于高度均匀合成的方法,其导致所得文库中预定gRNA的高表示。在设计阶段,设计gRNA。参见图2。设计策略包括但不限于设计跨基因的gRNA。根据所需的工作流程,从头合成的核酸是DNA或RNA碱基。

[0056] 在从头合成的DNA的情况下,合成包含核酸的文库,其中每种合成的核酸是编码gRNA(例如,sgRNA)序列作为转录产物的DNA序列。在一些情况下,随后将合成的核酸插入表达载体中。在一个示例性工作流程中,将合成的核酸插入病毒载体中,并随后包装以转导到细胞中,然后进行筛选和分析。见图2。示例性的细胞包括但不限于原核和真核细胞。示例性的真核细胞包括但不限于动物、植物和真菌细胞。示例性的动物细胞包括但不限于昆虫、鱼和哺乳动物细胞。示例性的哺乳动物细胞包括小鼠、人和灵长类动物细胞。测试的示例性细胞功能包括但不限于细胞增殖、迁移/粘附、代谢和细胞信号传导活性的变化。在从头合成的RNA的情况下,gRNA本身被合成并可用于下游应用,诸如转染到细胞中。

[0057] 寡核酸可以在位置(“座位”)的簇303内合成,用于在阵列301上延伸。见图3。与没有成簇座位布置的情况下在整个平板上合成的寡核酸的扩增产物相比,这样的布置可以提供对来自合成的寡核酸的扩增产物(称为“扩增子”)的改进的寡核酸表示。在一些情况下,由于扩增反应产物中低GC或高GC扩增子的代表性不足,合成的寡核酸在单个簇内的扩增310反映了由于重复合成具有大量GC含量的寡核酸的较大寡核酸群体而对表示的负面影响,通常称为“漂移”。在一些情况下,本文所述的单个簇包含约50-1000、75-900、100-800、125-700、150-600、200-500或300-400个离散的座位。在一些情况下,单个簇包含50-500个离散座位。在一些情况下,座位是斑点、孔、微孔、通道或柱。在一些情况下,每个簇的支持具

有相同序列的寡核酸延伸的单个特征具有至少1X、2X、3X、4X、5X、6X、7X、8X、9X、10X或更多的冗余。

[0058] 本文提供了用于插入表达载体中的gRNA文库。继续图3中的工作流程,阵列301包含用于寡核酸合成和延伸的座位的多个簇303。合成从头DNA并将其从板移除以形成寡核酸(例如,编码sgRNA的DNA)的群体305,其经历扩增310以形成扩增的寡核酸的文库320,用于插入载体330中以形成包含合成DNA的载体的文库335。一旦进入细胞,DNA就被转录成gRNA(例如,sgRNA)并可用于与基因组编辑方案(例如,基于Cas9的系统)结合。细胞可以具有编辑酶(例如,Cas9)的天然或异位表达。编辑酶(例如,Cas9)可以具有双DNA链切割活性或经修饰的活性,如切口、碱基交换或序列交换活性。用于插入载体的合成DNA可包括sgRNA、dgRNA或其片段。

[0059] 用于插入本文公开的核酸文库的表达载体包括真核或原核表达载体。示例性的表达载体包括但不限于哺乳动物表达载体:pSF-CMV-NEO-NH2-PPT-3XFLAG、pSF-CMV-NEO-COOH-3XFLAG、pSF-CMV-PURO-NH2-GST-TEV、pSF-OXB20-COOH-TEV-FLAG(R)-6His、pCEP4pDEST27、pSF-CMV-Ub-KrYFP、pSF-CMV-FMDV-daGFP、pEF1a-mCherry-N1载体、pEF1a-tdTomato载体、pSF-CMV-FMDV-Hygro、pSF-CMV-PGK-Puro、pMCP-tag(m)和pSF-CMV-PURO-NH2-CMYC;细菌表达载体:pSF-OXB20-BetaGal、pSF-OXB20-Fluc、pSF-OXB20和pSF-Tac;植物表达载体:pRI 101-AN DNA和pCambia2301;以及酵母表达载体:pTYB21和pKLAC2,以及昆虫载体:pAc5.1/V5-His A和pDEST8。

[0060] 通过本文所述方法合成的从头寡核酸文库可以在细胞中表达。在一些情况下,细胞与疾病状态相关。例如,与疾病状态相关的细胞包括但不限于细胞系、组织样品、来自受试者的原代细胞、从受试者扩充的培养细胞或模式系统中的细胞。在一些情况下,模式系统是植物或动物系统。在一些情况下,从头寡核酸文库在细胞中表达以评估细胞活动的变化。示例性的细胞活动包括但不限于增殖、周期进程、细胞死亡、粘附、迁移、繁殖、细胞信号传导、能量产生、氧利用、代谢活动、衰老、对自由基损伤的反应或其任何组合。

[0061] 本文提供了用于合成gRNA文库(或在转录时产生gRNA文库的DNA文库)的方法,其中该gRNA文库包含每个基因多个不相同的gRNA。gRNA可以编码sgRNA或dgRNA。在一些情况下,gRNA文库包含每个基因至少3、4、5、6、7、8、9、10个或多于10个不相同的gRNA。在一些情况下,gRNA文库靶向一个或多个基因。在一些情况下,gRNA文库靶向约1、5、10、20、30、40、50、60、70、80、90、100个或多于100个基因。在一些情况下,gRNA文库靶向约1-100、2-95、5-90、10-85、15-80、20-75、25-70、30-65、35-60或40-50个基因。在一些情况下,本文所述的gRNA文库靶向途径中的基因。示例性的途径包括但不限于代谢、细胞死亡、细胞周期进程、免疫细胞激活、炎症反应、血管生成、淋巴生成、缺氧和氧化应激反应、细胞粘附和细胞迁移途径。

[0062] 用于合成如本文所述的gRNA文库的方法可以提供具有与基因组的一部分互补的碱基配对区域即基因组靶区域的不相同的gRNA的合成。基因组靶区域可包含外显子、内含子、编码或非编码序列。在一些情况下,gRNA文库包含不相同的gRNA,其共同具有与整个基因组中至少或约5%的基因互补的碱基配对区域。在一些情况下,gRNA文库包含不相同的gRNA,其共同具有与整个基因组中至少或约80%的基因互补的碱基配对区域。在一些情况下,gRNA文库包含不相同的gRNA,其共同具有与整个基因组中至少或约90%的基因互补的

碱基配对区域。在一些情况下，gRNA文库包含不相同的gRNA，其共同具有与整个基因组中至少或约95%的基因互补的碱基配对区域。在一些情况下，gRNA文库包含不相同的gRNA，其共同具有与整个基因组中至少或约100%的基因互补的碱基配对区域。

[0063] 本文提供了通过本文所述的方法合成的gRNA文库，其导致gRNA在不同细胞中具有至少2X的基因消耗。在一些情况下，gRNA文库包含至少或约10%、12%、15%、16%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%或更多的gRNA，所述gRNA当存在于细胞或多个细胞群体中时提供至少2X的基因消耗。在一些情况下，该基因是必需基因，即对细胞存活至关重要的基因。示例性的必需基因包括但不限于PCNA、PSMA7、RPP21和SF3B3。在一些情况下，gRNA文库包含当存在于细胞中时提供至少2X、3X、4X、5X、6X或大于6X的基因消耗的gRNA。在一些情况下，当存在于细胞或多个细胞群体中时，gRNA文库包含至多5%、10%、12%、15%或20%的具有零或负基因消耗的gRNA。在一些情况下，多个细胞群体包含至少或约2000、3000、4000、5000、6000、7000、8000、9000、10000、12000、15000、20000、25000、30000个或多于30000个细胞群体。在一些情况下，对于多个基因，gRNA文库包含具有至少2X、3X、4X、5X、6X或大于6X消耗的gRNA。在一些情况下，对于多个基因，gRNA文库包含平均至少或约40%、45%、50%、60%、65%、70%、75%、80%的或多于90%的至少2X消耗的gRNA。提供这样的基因消耗谱的gRNA可以是sgRNA或dgRNA。

[0064] 本文提供了用于合成高度均匀的寡核酸文库的方法。在一些情况下，多于90%的合成寡核酸(RNA或DNA)的表示在核酸文库的寡核酸表示的平均值的4X以内。在一些情况下，多于90%的寡核酸的表示在文库的寡核酸表示的平均值的2X内。在一些情况下，多于90%的寡核酸的表示在文库的寡核酸表示的平均值的1.5X内。在一些情况下，多于80%的寡核酸的表示在文库的寡核酸表示的平均值的1.5X内。

[0065] 与预定序列相比，通过本文所述方法从头合成的寡核酸文库包含高百分比的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于70%的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于75%的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于80%的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于85%的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于90%的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于95%的正确序列。在一些情况下，与寡核酸的预定序列相比，本文公开的从头寡核酸文库具有大于100%的正确序列。

[0066] 在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有大于70%的正确序列。在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有大于75%的正确序列。在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有大于80%的正确序列。在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有大于85%的正确序列。在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有大于90%的正确序列。在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有大于95%的正确序列。在一些情况下，与扩增反应后寡核酸的预定序列相比，本文公开的从头合成的寡核酸文库具有100%

的正确序列。

[0067] 在一些情况下,当转移到细胞中时,与寡核酸的预定序列相比,本文公开的从头合成的寡核酸文库产生大于80%的正确序列。在一些情况下,当转移到细胞中时,与寡核酸的预定序列相比,本文公开的从头合成的寡核酸文库产生大于85%的正确序列。在一些情况下,当转移到细胞中时,与寡核酸的预定序列相比,本文公开的从头合成的寡核酸文库产生大于90%的正确序列。在一些情况下,当转移到细胞中时,与寡核酸的预定序列相比,本文公开的从头合成的寡核酸文库产生大于95%的正确序列。在一些情况下,当转移到细胞中时,与寡核酸的预定序列相比,本文公开的从头合成的寡核酸文库产生100%的正确序列。

[0068] 在一些情况下,当转移到细胞中时,本文公开的从头合成的寡核酸文库产生大于80%的序列表示。在一些情况下,当转移到细胞中时,本文公开的从头合成的寡核酸文库产生大于90%的序列表示。在一些情况下,当转移到细胞中时,本文公开的从头合成的寡核酸文库产生大于95%的序列表示。在一些情况下,当转移到细胞中时,本文公开的从头合成的寡核酸文库产生100%的序列表示。

[0069] 本文所述的从头寡核酸文库可以在添加聚合酶和扩增试剂(例如,缓冲液、磷酸盐和dNTP)的情况下进行扩增反应。在一些情况下,通过PCR将从头寡核酸文库扩增至少或约6、8、10、15、20个或多于20个循环。在一些情况下,通过PCR将从头寡核酸文库扩增约6至20、7至18、8至17、9至16或10至15个循环。在一些情况下,通过PCR将从头寡核酸文库扩增约15个循环。

[0070] 在一些情况下,从头寡核酸文库的扩增提供了DNA分子的扩增子文库。在一些情况下,扩增子文库包含编码gRNA序列的不相同的核酸。在一些情况下,gRNA序列是sgRNA或dgRNA。

[0071] 在一些情况下,从头寡核酸文库包含不相同的核酸,其中每个不相同的核酸包括DNA分子。在一些情况下,DNA分子的数目为约500、2000、3500个或更多个分子。在一些情况中,DNA分子的数目为至少或约250、500、1000、1250、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000、15000、20000、50000、100000、250000、500000、750000、100万个或多于100万个分子。在一些情况中,DNA分子的数目为至多250、500、1000、1250、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000、15000、20000、50000、100000、250000、500000、750000、100万个或多于100万个分子。在一些情况下,DNA分子编码gRNA序列。在一些情况下,gRNA序列是sgRNA或dgRNA。

[0072] 在一些情况下,从头寡核酸文库包含不相同的核酸,其中每个不相同的核酸包括RNA分子。在一些情况下,RNA分子的数目为约2000个分子。在一些情况中,RNA分子的数目为至少或约250、500、1000、1250、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000、15000、20000、50000、100000、250000、500000、750000、100万个或多于100万个分子。在一些情况中,RNA分子的数目为至多250、500、1000、1250、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000、15000、20000、50000、100000、250000、500000、750000、100万个或多于100万个分子。在一些情况下,RNA分子编码gRNA序列。在一些情况下,gRNA序列是sgRNA或dgRNA。

[0073] 本文提供了扩增后具有高均匀性的从头寡核酸文库。在一些情况下,在扩增后,本文所述的从头寡核酸文库中大于80%的寡核酸的表示在整个文库的平均表示的至少约1.5X内。在一些情况下,在扩增后,本文所述的从头寡核酸文库中大于90%的寡核酸的表示在整个文库的平均表示的至少约1.5X内。在一些情况下,在扩增后,本文所述的从头寡核酸文库中大于80%的寡核酸的表示在整个文库的平均表示的至少约2X内。在一些情况下,在扩增后,本文所述的从头寡核酸文库中大于80%的寡核酸的表示在整个文库的平均表示的至少约2X内。

[0074] 使用本文所述方法的从头合成的未扩增的寡核酸群体在不相同的寡核酸序列的数目方面可有所变化。在一些情况下,不相同的寡核酸序列的数目为约2000-1000000、3000-900000、4000-800000、5000-700000、6000-600000、7000-500000、8000-400000、9000-300000、10000-200000、11000-100000、12000-75000、14000-60000和20000-50000个序列。在一些情况中,不相同的寡核酸序列的数目为约50-2000、75-1800、100-1700、150-1600、200-1500、250-1400、300-1300、400-1200、500-1100、600-1000、700-900个序列。在一些情况下,不相同的寡核酸序列的数目为2000个序列。在一些情况下,不相同的寡核酸序列的数目为大于1百万个序列。在一些情况下,不相同的寡核酸序列的数目为至少50、100、200、400、600、800、1000、1200、1400、1600、1800、2000、3000、5000、7000、10000、20000、30000、50000、100000、500000、700000、1000000、10000000、1000000000个或更多个序列。在一些情况下,不相同的寡核酸序列的数目为至多50、100、200、400、600、800、1000、1200、1400、1600、1800、2000、3000、5000、7000、10000、20000、30000、50000、100000、500000、700000、1000000个或更多个序列。在一些情况下,不相同的寡核酸序列的数目为至多50、100、200、400、600、800、1000、1200、1400、1600、1800、2000、3000、5000、7000、10000、20000、30000、50000、100000、500000、700000和1000000个序列。

[0075] 未扩增群体的寡核酸可以以不同的量存在。在一些情况下,未扩增群体的寡核酸以至少或约0.25飞摩尔的量存在。在一些情况下,未扩增群体的寡核酸以至少或约1飞摩尔的量存在。在一些情况下,未扩增群体的寡核酸以至少0.25、1、10、20、30、40、50、100、250、500、750、1000或大于1000飞摩尔的量存在。在一些情况下,未扩增群体的寡核酸以至多0.25、1、10、20、30、40、50、100、250、500、750和1000飞摩尔的量存在。

[0076] 本文提供了用于合成不相同的寡核酸的文库的方法,其中不相同的寡核酸的序列长度或平均序列长度有所变化。在一些情况下,不相同的寡核酸的序列长度或平均序列长度为至多150个碱基。在一些情况下,不相同的寡核酸的序列长度或平均序列长度在约100至约200个碱基的范围内。在一些情况下,不相同的寡核酸的序列长度或平均序列长度为至少30、50、100、125、150、200、250、300、350、400、450、500个或多于500个碱基。在一些情况下,不相同的寡核酸的序列长度或平均序列长度为至多150、200、250、300、350、400、450或500个碱基。不相同的寡核酸的示例性序列长度在约25至约150或约50至约200个碱基的范围内。在一些情况下,不相同的寡核酸的序列长度或平均序列长度在约125至约200或约150至约200个碱基的范围内。

[0077] 指导RNA序列

[0078] 本文提供了用于将基因组序列编辑酶(例如,Cas9)引导至特定靶核酸序列的单指导RNA (sgRNA) 序列。与Cas9酶复合的示例性sgRNA在图4A中示出,并且在图4B中是单独的实

例。gRNA可以是双指导RNA,如图4C所示。本文公开的指导序列包含碱基配对区域。碱基配对区域包含用于结合靶序列的种子区域以及任意的间隔区域。碱基配对区域的长度可以有所变化。例如,碱基配对区域可以在长度上包含约1至10、1至20、20至25或1至30个碱基。在一些情况下,碱基配对区域在长度上包含至少10、15、20、25、30个或更多个碱基。在一些情况下,碱基配对区域包含长度为至少10个碱基的种子区域。种子区域可以在长度上包含约8至20个碱基。在一些情况下,种子区域为约12个碱基的长度。在一些情况下,本文所述的碱基配对区域被设计用于在转录期间靶向模板链,图5A。在一些情况下,本文所述的碱基配对区域被设计用于在转录期间靶向非模板链,图5B。

[0079] 在一些情况下,sgRNA的碱基配对区域的3' 是用于与Cas9结合的Cas9手柄区域。在一些情况下,Cas9手柄区域是用于与dCas9酶结合的dCas9手柄区域。手柄区域的长度可以有所变化。例如,手柄区域可以在长度上包含约1至50、20至45或15至60个碱基。在一些情况下,手柄区域在长度上包含至少35、40、45、50个或更多个碱基。手柄区域可以在长度上包含约42个碱基。

[0080] 在一些情况下,sgRNA的手柄区域的3' 是终止子区域。在一些情况下,终止子区域是酿脓链球菌终止子区域。在一些情况下,终止子区域在长度上包含约40个碱基。在一些情况下,终止子区域在长度上包含约10至50、20至60或30至55个碱基。

[0081] 本文所述的gRNA序列的设计方案可以包括在编码gRNA序列的DNA的5' 上游处包含DNA依赖性RNA聚合酶启动子区域。示例性的DNA依赖性RNA聚合酶启动子区包括但不限于T3和T7RNA聚合酶启动子序列。例如,图6A图示了一种布置,其中T7启动子区域在gRNA的5' 上游,并且产生所得转录的gRNA,其中gRNA包含发夹。在一些布置中,gRNA被设计为缺少形成发夹二级结构的序列,图6B。发夹二级结构可能缺少Cas9手柄和/或终止子区域。

[0082] 本文提供了用于将基因组序列编辑酶(例如,Cas9)引导至特定靶核酸序列的dgRNA。在一些情况下,文库包含编码dgRNA序列的寡核酸序列。在一些情况下,文库包含核酸,其中每种合成的核酸都是编码dgRNA序列作为转录产物的DNA序列。在一些情况下,文库包含核酸,其中每种合成的核酸是RNA序列,并且dgRNA本身是合成的。在一些情况下,dgRNA文库包含crRNA和tracrRNA的寡核酸序列,其作为单独的寡核酸被合成。在一些情况下,寡核酸核酸分别编码crRNA和tracrRNA。在一些情况下,寡核酸核酸编码在转录时产生单独的crRNA序列和单独的tracrRNA序列的单个序列。crRNA和tracrRNA的示例性序列见表1。

表1.

SEQ ID NO	名称	序列
5	crRNA-sp2	5'ATAACTCAATTTGTAAAAAAGTTTTAGAGCTAT GCTGTTTTG3'
6	tracrRNA	5'GGAACCATTCAAAACAGCATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGCC ACCGAGTCGGTGCTTTTTTT3'

[0083] 本文所述的gRNA文库可用于体外筛选和分析。这样的布置的图示描绘于图7中,其中靶标双链DNA序列与gRNA序列和Cas9酶一起温育。该混合物导致双链DNA断裂。DNA断裂可

导致基因组元件的功能或表达的可测定的变化。可以经由本领域已知的各种方法将本文所述的gRNA或编码gRNA的DNA添加至细胞,所述方法包括但不限于转染、转导或电穿孔。

[0084] 在一些情况下,本文所述的gRNA文库用于体内或离体筛选和分析。用于筛选的细胞包括取自活体受试者的原代细胞或细胞系。细胞可以来自原核生物(例如,细菌和真菌)或真核生物(例如,动物和植物)。示例性的动物细胞包括但不限于来自小鼠、兔、灵长类动物和昆虫的动物细胞。在一些情况下,还可以将本文所述的gRNA文库递送至多细胞生物体。示例性的多细胞生物体包括但不限于植物、小鼠、兔、灵长类动物和昆虫。

[0085] 基因组工程化

[0086] 本文提供了包含针对于靶向特定靶核酸序列的核酸酶的核酸的文库。在一些情况下,本文所述的文库包含合成的核酸,其中该核酸是DNA、RNA、其任何类似物或衍生物。在一些情况下,靶核酸序列包括DNA、RNA、其任何类似物或衍生物。在一些情况下,核酸酶切割靶核酸序列。在一些情况下,核酸酶与靶核酸结合但不切割它。核酸酶的类型包括但不限于类转录激活因子效应物核酸酶(TALEN)、锌指核酸酶(ZFN)、大范围核酸酶、Argonaute和成簇的规律间隔短回文重复(CRISPR)-相关(Cas)蛋白。在一些情况下,核酸酶是野生型的、遗传修饰的或重组的。

[0087] 用于靶向基因编辑的模式系统包括基于Cas9的方法。当与gRNA一起表达或转移到细胞中时,Cas9允许经由具有成簇的规律间隔短回文重复(CRISPR)序列的mRNA的复合物靶向引入或删除遗传信息。如图1A-图1B所示,Cas9复合物包含与指导RNA(“gRNA”)转录物结合的Cas9蛋白。示出的gRNA含有靶序列区域、PAM区域和发夹区域。在CRISPR/Cas9过程中,gRNA将Cas9酶引导至特定的DNA片段。虽然所描绘的gRNA是sgRNA(单链指导RNA),但可以用dgRNA(双链指导RNA)形成复合物。然后Cas9切割DNA以损坏(disable)或修复基因。这里描述了对该过程的示例性修饰的非限制性列表。在CRISPR/dCas9过程中,受损的或“死亡”的Cas9(“dCas9”)不再具有剪接功能,但是通过添加另一酶活性执行不同的靶分子修饰功能。例如,将胞苷脱氨酶束缚(tether)于dCas9将C-G DNA碱基对转化为T-A碱基对。在另一dCas9过程中,束缚于dCas9的不同酶导致靶DNA中的碱基C变为T或G变为A。或者,可以通过融合转录因子来修饰dCas9过程,以阻断或激活RNA聚合酶活性,从而导致关闭(CRISPRi)或开启(CRISPRa)基因转录并由此调节基因表达。例如,通过与转录阻抑物融合来修饰dCas9过程。在一些情况下,通过与转录激活子融合来修饰dCas9过程。在一些情况下,通过与多种转录阻抑物或转录激活子融合来修饰dCas9过程。在备选的布置中,gRNA具有多个用于切割的位点,从而导致gRNA具有多个用于基因编辑的区域。在Cas9n或“切口Cas9”的情况下,RuvC或HNH切割结构域被修饰为无活性的。这种失活使得Cas9仅能够在DNA中产生仅链断裂(切口),而不是双链断裂。在一些布置中,使用两种Cas9n酶(每条链一种)来产生双链断裂。由于它们可以识别切割位点的上游和下游区域,因此消除了脱靶效应。在hfCas9的情况下,并非使用双Cas9n蛋白来生成脱靶无效Cas9切割,而是修饰的Cas9酶具有松弛的结合靶标特异性严格性,以允许在酶活性之前有不完美的匹配。在一些情况下,通过与用于检测靶核酸的标记或标签融合来修饰dCas9过程。例如,标记是用于检测靶核酸的荧光标记物(例如,GFP)。在一些情况下,dCas9与表位标签融合,并用于纯化由gRNA指定的靶核酸。

[0088] 本文提供了包含用于将核酸酶引导至特定靶核酸序列的核酸的文库。在一些情况下,靶核酸序列包括DNA。在一些情况下,靶核酸序列包括RNA。例如,生成包含用于引导C2c2

的核酸的文库,以靶向RNA序列。在一些情况下,DNA或RNA是单链或双链的。

[0089] 本文提供了包含用于靶向特定靶核酸序列的核酸酶的核酸的文库,其中该核酸酶来自但不限于链球菌属(*Streptococcus*)、弯曲杆菌属(*Campylobacter*)、*Nitratifactor*、葡萄球菌属(*Staphylococcus*)、红细菌属(*Parvibaculum*)、罗斯氏菌属(*Roseburia*)、奈瑟氏菌属(*Neisseria*)、葡糖酸醋酸杆菌属(*Gluconacetobacter*)、固氮螺菌属(*Azospirillum*)、*Sphaerochaeta*、乳杆菌属(*Lactobacillus*)、真杆菌属(*Eubacterium*)、棒杆菌属(*Corynebacter*)、肉杆菌属(*Carnobacterium*)、红细菌属(*Rhodobacter*)、利斯特氏菌属(*Listeria*)、*Paludibacter*、梭菌属(*Clostridium*)、毛螺菌科(*Lachnospiraceae*)、*Clostridiaridium*、纤毛菌属(*Leptotrichia*)、弗朗西斯氏菌属(*Francisella*)、军团菌属(*Legionella*)、脂环酸杆菌属(*Alicyclobacillus*)、甲烷嗜甲基菌(*Methanomethylophilus*)、卟啉单胞菌属(*Porphyromonas*)、普雷沃氏菌属(*Prevotella*)、拟杆菌门(*Bacteroidetes*)、创伤球菌属(*Helcococcus*)、钩端螺旋体属(*Letospira*)、脱硫弧菌属(*Desulfovibrio*)、脱硫盐碱杆菌属(*Desulfonatobium*)、硫还原球菌属(*Desulfurococcus*)、丰祐菌科(*Opitutaceae*)、肿块芽孢杆菌属(*Tuberibacillus*)、芽孢杆菌属(*Bacillus*)、短芽孢杆菌属(*Brevibacillus*)、甲基杆菌属(*Methylobacterium*)、嗜盐碱杆菌属(*Natronobacterium*)、黄杆菌属(*Flavobacterium*)、酵母属(*Saccharomyces*)、衣藻属(*Chlamydomonas*)、栖热菌属(*Thermus*)、火球菌属(*Pyrococcus*)、支原体属(*Mycoplasma*)或氨基酸球菌属(*Acidaminococcus*)等物种。示例性的核酸酶列于表2A中。本文所述的gRNA可以与来自上文所列的任何物种的核酸酶或来自其中酶允许基因组编辑功能的其他物种的核酸酶的终止子序列结合。示例性的终止子序列包括但不限于表2B中列出的终止子序列。示例性的PAM序列包括但不限于表2C中列出的PAM序列。

表2A.

名称	登录号	物种
Cas9	Q99ZW2	酿脓链球菌
Cas9	J7RUA5	金黄色葡萄球菌 ( <i>Staphylococcus aureus</i> )
Cas9	G3ECR1	嗜热链球菌 ( <i>Streptococcus thermophilus</i> )
C2c2	P0DOC6	沙氏纤毛菌 ( <i>Leptotrichia shahii</i> )
Cpf1	U2UMQ6	氨基酸球菌属的种 ( <i>Acidaminococcus sp.</i> )
C2c1	T0D7A2	酸土脂环酸杆菌 ( <i>Alicyclobacillus acidoterrestris</i> )
FokI	P14870	海床黄杆菌 ( <i>Flavobacterium okeanokoites</i> )
Acil	A0A0C5GQT3	嗜酸乳杆菌 ( <i>Lactobacillus acidophilus</i> )
I-SceI	P03882	酿酒酵母
I-CreI	P05725	莱茵衣藻
I-DmoI	1B24_A	运动硫还原球菌 ( <i>Desulfurococcus mobilis</i> )
TtAgo	Q746M7	嗜热栖热菌 ( <i>Thermus thermophilus</i> )
PfAgo	Q8U3D2	激烈火球菌 ( <i>Pyrococcus furiosus</i> )
NgAgo	A0A172MAH6	格氏嗜盐碱杆菌 ( <i>Natronobacterium gregoryi</i> )

表2B

SEQ ID NO	物种	序列
7	嗜热链球菌	5' TACTCAACTTGAAAAGGTGGCACCGATTTCGGTGTTTT 3'
8	变异链球菌 ( <i>Streptococcus mutans</i> )	5' TACACAACTTGAAAAAGTGCGCACCGATTTCGGTGCTTTT 3'
9	无害利斯特氏菌 ( <i>Listeria innocua</i> )	5' TTATCAACTTTTAATTAAGTAGCGCTGTTTCGGCGCTTTT 3'
10	运动支原体 ( <i>Mycoplasma mobile</i> )	5' TATGCCGTA ACTACTACTTATTTTCAAATAAGTAGTTTT 3'
11	空肠弯曲杆菌 ( <i>Campylobacter jejuni</i> )	5' GACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTT 3'

表2C

Cas9 的种类/变体	PAM 序列
酿脓链球菌/SpCas9	NGG
酿脓链球菌/SpCas9 D1135E 变体	NGG
酿脓链球菌/SpCas9 VRER 变体	NGCG
酿脓链球菌/SpCas9 EQR 变体	NGAG
酿脓链球菌/SpCas9 VQR 变体	NGAN 或 NGNG

Cas9 的种类/变体	PAM 序列
金黄色葡萄球菌/SaCas9	NNGRRT 或 NNGRR(N)
脑膜炎奈瑟氏球菌 ( <i>Neisseria meningitidis</i> )	NNNGATT
嗜热链球菌	NNAGAAW
齿垢密螺旋体 ( <i>Treponema denticola</i> )	NAAAAC

[0090] 本文提供了包含用于将一种或多种核酸酶靶向特定核酸序列的核酸的文库。在一些情况下,核酸酶是TALEN、ZFN、大范围核酸酶、Argonaute和Cas蛋白中的至少一种。例如,可将多于一种的核酸酶多重化以生成大的基因组缺失、一次修饰多个序列或与其他酶如切口酶一起使用。在一些情况下,对于靶核酸序列,核酸酶的数目是至少2种核酸酶。在一些情况下,对于靶核酸序列,核酸酶的数目在约2至3、2至4、2至5、2至6、2至7、2至8、2至9或2至10种核酸酶的范围。

[0091] 本文提供了包含针对于靶向特定核酸序列的切口酶的合成核酸的文库。切口酶是在核酸序列中生成单链断裂的酶。在一些情况下,合成的核酸是DNA、RNA、其任何类似物或衍生物。在一些情况下,特定核酸序列包括DNA、RNA、其任何类似物或衍生物。在一些情况下,切口酶切割特定核酸序列。在一些情况下,切口酶与特定核酸结合但不切割它。在一些情况下,切口酶是改变的核酸酶,其中该核酸酶是TALEN、ZFN、大范围核酸酶、Argonaute或Cas蛋白。在一些情况下,通过改变TALEN、ZFN、大范围核酸酶、Argonaute或Cas蛋白的核酸酶结构域生成切口酶。在一些情况下,通过改变Cas9的核酸酶结构域生成切口酶。

[0092] 在一些情况下,文库包含针对于靶向特定核酸序列的一种或多种切口酶的核酸。在一些情况下,对于特定核酸序列,切口酶的数目是至少2种切口酶。在一些情况下,对于特定核酸序列,切口酶的数目在约2至3、2至4、2至5、2至6、2至7、2至8、2至9或2至10种切口酶的范围。在一些情况下,文库包含用于将一种或多种切口酶和一种或多种核酸酶引导至特定核酸序列的核酸。

[0093] 包含用于将核酸酶靶向至本文提供的特定核酸序列的核酸的文库可导致特定核酸序列的切割。在一些情况下,核酸酶是TALEN、ZFN、大范围核酸酶、Argonaute和Cas蛋白中的至少一种。在一些情况下,核酸酶是提供除切割之外的特定核酸序列的修饰的嵌合核酸酶。例如,嵌合核酸酶导致甲基化、脱甲基化、多腺苷酸化、脱腺苷化、脱氨基或多尿苷酰化。

[0094] 类转录激活因子效应物核酸酶

[0095] 本文提供了用于合成核酸文库的方法,该核酸文库包含针对于靶向特定核酸序列的转录激活因子样效应物核酸酶 (TALEN) 的核酸。TALEN是一类可用于诱导特定靶序列处的双链断裂的工程化序列特异性核酸酶。TALEN可以通过将类转录激活因子 (TAL) 效应物DNA结合结构域或其功能部分与核酸酶的催化结构域融合而生成。TAL效应物DNA结合结构域包含一系列TAL重复,其通常是高度保守的33或34个氨基酸序列区段,每个区段包含称为重复可变双残基 (RVD) 的高度可变的第12和第13个氨基酸。每种RVD可以识别并结合特定核苷酸。因此,可以通过将包含适当RVD的TAL重复进行组合来使TAL效应物结合结构域工程化以识别特定核苷酸序列。

[0096] 本文提供了用于合成包含编码TAL效应物DNA结合结构域的不相同的核酸的核酸

文库的方法。在一些情况下,设计TAL效应物DNA结合结构域以识别特定的靶核酸序列并在特定位点诱导双链断裂。在一些情况下,TAL效应物DNA结合结构域包含许多被设计用于识别和结合特定核酸序列的TAL重复。在一些情况中,TAL重复的数目为至少或约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个或更多个TAL重复。

[0097] 在一些情况下,合成包含编码TAL效应物DNA结合结构域的不相同的核酸的核酸文库。在一些情况下,如本文所述的核酸文库在翻译时编码蛋白质文库。在一些情况下,核酸文库在细胞中表达并生成蛋白质文库。在一些情况下,合成的核酸文库被插入表达载体中。在一些情况下,合成的核酸文库被插入表达载体中并在细胞中表达。

[0098] 通过本文所述方法生成的包含编码TAL效应物DNA结合结构域的核酸的核酸文库可用于产生TALEN。在一些情况下,这通过将载体中克隆并表达的TAL效应物结合结构域文库与核酸酶混合来实现。示例性的核酸酶包括但不限于AciI、AclI、AlwI、BbvI、BccI、BceAI、BciVI、BfuAI、BmgBI、BmrI、BpmI、BpuEI、BsaI、BsmAI、BsmFI、BseRI、BspCNI、BsrI、BsgI、BsmI、BspMI、BsrBI、BsrDI、BtgZI、BtsI、BtsCI、EarI、EciI、FokI、HgaI、HphI、HpyAV、MboII、MlyI、MmeI、MnlI、NmeAIII、PleI、SfaNI、BbvCI、Bpu10I、BspQI、SapI、BaeI、BsaXI或CspCI。在一些情况下,通过连接发生混合。示例性的连接酶包括但不限于大肠杆菌连接酶、T4连接酶、哺乳动物连接酶(例如,DNA连接酶I、DNA连接酶II、DNA连接酶III、DNA连接酶IV)、热稳定连接酶和快速连接酶。可以将通过本文所述的方法生成的TALEN插入表达载体中。在一些情况下,TALEN被插入表达载体中并在细胞中表达。

[0099] 本文提供了用于合成TAL效应物DNA结合结构域文库的方法,该TAL效应物DNA结合结构域文库包含针对于原核或真核生物体的基因组中的基因的不相同的核酸序列。在一些情况下,TAL效应物DNA结合结构域文库包含针对于至少5%基因组的基因的至少3、4、5、6、7、8、9、10个或多于10个不相同的核酸序列。在一些情况下,TAL效应物DNA结合结构域文库包含针对于至少5%基因组的一个或多个基因的不相同的核酸序列。在一些情况下,TAL效应物DNA结合结构域文库包含针对于至少5%基因组的约1、5、10、20、30、40、50、60、70、80、90、100个或多于100个基因的不相同的核酸序列。在一些情况下,TAL效应物DNA结合结构域文库包含针对于至少5%基因组的约1-100、2-95、5-90、10-85、15-80、20-75、25-70、30-65、35-60或40-50个基因的不相同的核酸序列。

[0100] 锌指核酸酶

[0101] 本文提供了用于合成核酸文库的方法,该核酸文库包含针对于靶向特定核酸序列的锌指核酸酶(ZFN)的核酸。ZFN可以通过将核酸酶与DNA结合锌指结构域(ZFD)融合而生成。ZFD可以通过一个或多个锌指与靶核酸序列结合。在一些情况下,ZFD包含至少2、3、4、5、6、7、8、9、10个或更多个锌指。在一些情况下,ZFD包含至多2、3、4、5、6、7、8、9、10个或更多个锌指。在一些情况下,ZFD被设计用于识别特定的靶核酸序列并在特定位点诱导双链断裂。

[0102] 本文提供了用于合成包含在转录和翻译时编码ZFD的核酸的核酸文库的方法。在一些情况下,核酸文库在被翻译时编码蛋白质文库。在一些情况下,核酸文库在细胞中表达并生成蛋白质文库。在一些情况下,合成的核酸文库被插入表达载体中。在一些情况下,将合成的核酸文库插入表达载体中并在细胞中表达。

[0103] 包含编码通过本文所述的方法生成的ZFD的核酸的核酸文库可用于产生ZFN。在一

些情况下,这通过将克隆的并在载体中表达的ZFD与核酸酶混合来实现。示例性的核酸酶包括但不限于AciI、AclI、AlwI、BbvI、BccI、BceAI、BciVI、BfuAI、BmgBI、BmrI、BpmI、BpuEI、BsaI、BsmAI、BsmFI、BseRI、BspCNI、BsrI、BsgI、BsmI、BspMI、BsrBI、BsrDI、BtgZI、BtsI、BtsCI、EarI、EciI、FokI、HgaI、HphI、HpyAV、MboII、MlyI、MmeI、MnlI、NmeAIII、PleI、SfaNI、BbvCI、Bpu10I、BspQI、SapI、BaeI、BsaXI或CspCI。在一些情况下,通过连接发生混合。示例性的连接酶包括但不限于大肠杆菌连接酶、T4连接酶、哺乳动物连接酶(例如,DNA连接酶I、DNA连接酶II、DNA连接酶III、DNA连接酶IV)、热稳定连接酶和快速连接酶。可以将通过本文所述的方法生成的ZFN插入表达载体中。在一些情况下,ZFN被插入表达载体中并在细胞中表达。

[0104] 本文提供了用于合成ZFD文库的方法,该ZFD文库包含针对于原核或真核生物体的基因组中的基因的不相同的核酸序列。在一些情况下,ZFD文库包含针对于至少5%基因组的基因的至少3、4、5、6、7、8、9、10个或多于10个不相同的核酸序列。在一些情况下,ZFD文库包含针对于至少5%基因组的一个或多个基因的不相同的核酸序列。在一些情况下,ZFD文库包含针对于至少5%基因组的约1、5、10、20、30、40、50、60、70、80、90、100个或多于100个基因的不相同的核酸序列。在一些情况下,ZFD文库包含针对于至少5%基因组的约1-100、2-95、5-90、10-85、15-80、20-75、25-70、30-65、35-60或40-50个基因的不相同的核酸序列。

[0105] 大范围核酸酶

[0106] 本文提供了用于合成核酸文库的方法,该核酸文库包含针对于靶向特定核酸序列的大范围核酸酶的核酸。大范围核酸酶是能够识别和切割长碱基对(例如,12-40个碱基对)DNA靶标的酶。在一些情况下,大范围核酸酶被工程化以包含其他酶的结构域以赋予针对于靶核酸序列的特异性。例如,大范围核酸酶被工程化以包含TAL效应物DNA结合结构域。

[0107] 本文提供了用于合成包含核酸的核酸文库的方法,该核酸在转录和翻译时编码用于与大范围核酸酶一起使用的结合结构域。在一些情况下,核酸文库在翻译时编码蛋白质文库。在一些情况下,核酸文库在细胞中表达并生成蛋白质文库。在一些情况下,合成的核酸文库被插入表达载体中。在一些情况下,合成的核酸文库被插入表达载体中并在细胞中表达。

[0108] 包含编码通过本文所述方法生成的结构域的核酸的核酸文库可用于使大范围核酸酶工程化以靶向特定核酸序列。在一些情况下,这通过将结合结构域文库如在载体中克隆并表达的TAL效应物结合结构域文库与大范围核酸酶混合来实现。用于与本文提供的方法一起使用的示例性大范围核酸酶包括但不限于I-SceI、I-SceII、I-SceIII、I-SceIV、I-SceV、I-SceVI、I-SceVII、I-CeuI、I-CeuAIIIP、I-CreI、I-CrepsbIP、I-CrepsbIIIP、I-CrepsbIIP、I-TliI、I-PpoI、PI-PspI、F-SceI、F-SceII、F-SuvI、F-TevI、F-TevII、I-Amal、I-Anil、I-ChuI、I-Cmoel、I-CpaI、I-CpaII、I-CsmI、I-CvuI、I-CvuAIP、I-DdiI、I-DdiII、I-Dirl、I-Dmol、I-Hmul、I-HmuII、I-HsNIP、I-Llal、I-Msol、I-Naal、I-NanI、I-NcIIP、I-NgrIP、I-NitI、I-Njal、I-Nsp236IP、I-PakI、I-PboIP、I-PcuIP、I-PcuAI、I-PcuVI、I-PgrIP、I-PobIP、I-PorI、I-PorIIP、I-PbpIP、I-SpBetaIP、I-Scal、I-SexIP、I-SneIP、I-SpomI、I-SpomCP、I-SpomIP、I-SpomIIP、I-SquIP、I-Ssp6803I、I-SthPhiJP、I-SthPhiST3P、I-SthPhiSTe3bP、I-TdeIP、I-TevI、I-TevII、I-TevIII、I-UarAP、I-UarHGPAIP、I-UarHGPA13P、I-VinIP、I-ZbiIP、PI-MtuI、PI-MtuHIP、PI-MtuHIIP、PI-PfuI、

PI-PfuII、PI-PkoI、PI-PkoII、PI-Rma43812IP、PI-SpBetaIP、PI-SceI、PI-TfuI、PI-TfuII、PI-ThyI、PI-TliI、PI-TliII或其片段。在一些情况下,通过连接发生混合。示例性的连接酶包括但不限于大肠杆菌连接酶、T4连接酶、哺乳动物连接酶(例如,DNA连接酶I、DNA连接酶II、DNA连接酶III、DNA连接酶IV)、热稳定连接酶和快速连接酶。可以将通过本文所述的方法生成的工程化的大范围核酸酶插入表达载体中。在一些情况下,工程化的大范围核酸酶被插入表达载体中并在细胞中表达。

[0109] 本文提供了用于合成与大范围核酸酶一起使用的结合结构域文库的方法,该结合结构域文库包含针对于原核或真核生物体的基因组中的基因的不相同的核酸序列。在一些情况下,结构域文库包含针对于至少5%基因组的基因的至少3、4、5、6、7、8、9、10个或多于10个不相同的核酸序列。在一些情况下,结构域文库包含针对于至少5%基因组的一个或多个基因的不相同的核酸序列。在一些情况下,结构域文库包含针对于至少5%基因组的约1、5、10、20、30、40、50、60、70、80、90、100个或多于100个基因的不相同的核酸序列。在一些情况下,结构域文库包含针对于至少5%基因组的约1-100、2-95、5-90、10-85、15-80、20-75、25-70、30-65、35-60或40-50个基因的不相同的核酸序列。

[0110] Argonaute

[0111] 本文提供了用于合成核酸文库的方法,该核酸文库包含针对于靶向特定核酸序列的Argonaute的核酸。Argonaute是RNA或DNA指导的核酸酶家族。在一些情况下,Argonaute使用指导核酸鉴别靶核酸。在一些情况下,指导核酸是单指导RNA (sgRNA)。在一些情况下,指导核酸是指导DNA (gDNA)。示例性的Argonaute包括但不限于TtAgo、PfAgo和NgAgo。在一些实施方案中,Argonaute是NgAgo。

[0112] 本文提供了用于合成指导核酸文库的方法,该指导核酸文库包含针对于原核或真核生物体的基因组中的基因的不相同的核酸序列。在一些情况下,指导核酸文库是sgRNA文库。在一些情况下,指导核酸文库是dgRNA文库。在一些情况下,指导核酸文库包含针对于至少5%基因组的基因的至少3、4、5、6、7、8、9、10个或多于10个不相同的核酸序列。在一些情况下,指导核酸文库包含针对于至少5%基因组的一个或多个基因的不相同的核酸序列。在一些情况下,指导核酸文库包含针对于至少5%基因组的约1、5、10、20、30、40、50、60、70、80、90、100个或多于100个基因的不相同的核酸序列。在一些情况下,指导核酸文库包含针对于至少5%基因组的约1-100、2-95、5-90、10-85、15-80、20-75、25-70、30-65、35-60或40-50个基因的不相同的核酸序列。

[0113] CRISPR相关蛋白

[0114] 本文提供了用于合成核酸文库的方法,该核酸文库包含编码针对于靶向特定核酸序列的CRISPR相关(Cas)蛋白的gRNA的核酸。在一些情况下,Cas蛋白是Cpf1、C2c1、C2c2、Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (Csn1或Csx12)、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、其同源物及其修饰形式中的至少一种。在一些情况下,Cas蛋白是Cas9。

[0115] 本文提供了用于合成gRNA文库的方法,该gRNA文库包含针对于原核或真核生物体的基因组中的基因的不相同的核酸序列。在一些情况下,gRNA文库包含针对于至少5%基因

组的基因的至少3、4、5、6、7、8、9、10个或多于10个不相同的核酸序列。在一些情况下，gRNA文库包含针对于至少5%基因组的一个或多个基因的不相同的核酸序列。在一些情况下，gRNA文库包含针对于至少5%基因组的约1、5、10、20、30、40、50、60、70、80、90、100个或多于100个基因的不相同的核酸序列。在一些情况下，gRNA文库包含针对于至少5%基因组的约1-100、2-95、5-90、10-85、15-80、20-75、25-70、30-65、35-60或40-50个基因的不相同的核酸序列。gRNA文库可以编码sgRNA或dgRNA。

#### [0116] 变体文库合成

[0117] 本文提供了用于合成变体核酸文库的方法，该变体核酸文库通过将编码完整或部分基因序列的核酸与gRNA和核酸酶（例如，Cas9酶或Cas9变体酶）进行组合而生成。片段可以共同地间隔开基因的整个区域。在一些情况下，文库编码DNA或RNA。在一些情况下，文库编码单个基因或至多整个基因组。例如，编码包含约20,000个基因的基因组的每个基因5个gRNA的gRNA文库将产生约100,000个gRNA。这样的文库可用于选择性沉默或修饰单个基因、基因途径或单个基因组中的所有基因。在一些布置中，gRNA缺乏同源序列并且发生随机末端连接。这样的过程导致非同源末端连接（“NHEJ”）。在一些情况下，在NHEJ之后，发生插入、缺失、移码或单碱基交换。参见图1B。

[0118] 本文所述的合成文库可用于应用于CRISPR-Cas9功能，其中使用生成的gRNA序列来破坏细胞中或包含靶DNA和Cas9酶的混合物中的靶DNA序列的表达，或改变靶DNA序列的表达产物序列。在一些实施方案中，每种变体编码密码子，从而在翻译期间产生不同的氨基酸。表3提供了变体位点每种可能的密码子（和代表性氨基酸）的列表。

表3. 密码子和氨基酸列表

氨基酸	单字母代码	三字字母代码	密码子			
丙氨酸	A	Ala	GCA	GCC	GCG	GCT
半胱氨酸	C	Cys	TGC	TGT		
天冬氨酸	D	Asp	GAC	GAT		
谷氨酸	E	Glu	GAA	GAG		
苯丙氨酸	F	Phe	TTC	TTT		
甘氨酸	G	Gly	GGA	GGC	GGG	GGT
组氨酸	H	His	CAC	CAT		
异亮氨酸	I	Iso	ATA	ATC	ATT	
赖氨酸	K	Lys	AAA	AAG		

亮氨酸	L	Leu	TTA	TTG	CTA	CTC	CTG	CTT
甲硫氨酸	M	Met	ATG					
天冬酰胺	N	Asn	AAC	AAT				
脯氨酸	P	Pro	CCA	CCC	CCG	CCT		
谷氨酰胺	Q	Gln	CAA	CAG				
精氨酸	R	Arg	AGA	AGG	CGA	CGC	CGG	CGT
丝氨酸	S	Ser	AGC	AGT	TCA	TCC	TCG	TCT
苏氨酸	T	Thr	ACA	ACC	ACG	ACT		
缬氨酸	V	Val	GTA	GTC	GTG	GTT		
色氨酸	W	Trp	TGG					
酪氨酸	Y	Tyr	TAC	TAT				

[0119] 本文提供了用于合成变体核酸文库的方法,该变体核酸文库通过将编码完整或部分基因序列的核酸与核酸酶进行组合而生成,其中该核酸酶是TALEN、ZFN或工程化的大范围核酸酶。在一些情况下,用于合成变体核酸文库的方法,该变体核酸文库通过将编码完整或部分基因序列的核酸与引导核酸如sgRNA与核酸酶减小组合而产生,其中该核酸酶是Argonate或Cas蛋白。本文所述的合成文库可用于应用于核酸酶功能,其中使用生成的核酸序列来破坏细胞中或包含靶DNA和核酸酶的混合物中靶DNA序列的表达,或改变靶DNA序列的表达产物序列。在一些实施方案中,每种变体编码密码子,从而在翻译期间产生不同的氨基酸。

[0120] 如本文所述的变体核酸文库包含sgRNA或dgRNA,用于改变编码至少单个密码子的靶核酸序列,使得通过标准翻译过程生成后续由合成核酸编码的蛋白质中的单个残基的多种不同变体。在一些情况下,每种变体编码密码子,从而产生蛋白质结构域的不同氨基酸。例如,蛋白质结构域是保守结构域或催化结构域。在一些实施方案中,蛋白质结构域是但不限于激酶结构域、ATP结合结构域、GTP结合结构域、鸟嘌呤核苷酸交换因子(GEF)结构域、GTP酶激活蛋白(GAP)结构域、水解酶结构域、内切核酸酶结构域、外切核酸酶结构域、蛋白酶结构域、磷酸酶结构域、磷脂酶结构域、pleckstrin同源结构域、Src同源结构域和泛素结合结构域。在一些情况下,变体核酸文库包含sgRNA或dgRNA,以便于靶向编码至少1、2、3、4、5、6、7、8、9、10个或多于10个蛋白质结构域中的变异的核酸序列。

[0121] 在一些实施方案中,变体编码具有特定活性的蛋白质的氨基酸。例如,变体编码包含甲基转移酶活性、脱甲基酶活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光解酶活性、糖基化酶活性、乙酰基转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、脱遍在蛋白活性、腺苷酸化活性、脱腺苷化活性、类泛素化活性、脱类泛素化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性、重塑活性、蛋白酶活性、氧化

还原酶活性、转移酶活性、水解酶活性、裂合酶活性、异构酶活性、合酶活性、合成酶活性或脱豆蔻酰化活性的蛋白质。

[0122] 通过同源定向修复 (HDR) 生成的变异

[0123] 在用于变体核酸文库生成的示例性过程中, 包含Cas9切割和同源重组以生成靶DNA文库的多样性。首先, 合成gRNA文库 (通过从头合成RNA, 或者通过从头合成DNA然后进行转录 (体内或体外) 以生成gRNA), 其中该文库包含每个基因多个gRNA分子。例如, gRNA文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个gRNA。将gRNA文库与Cas9酶和靶DNA文库混合, 其中靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如, 靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下, 靶DNA文库包含来自途径中的多个基因或来自生物体中所有基因的序列。还向混合物添加替换序列, 该替换序列包括同源序列和变体核酸序列, 使得变异被引入靶DNA链。所得靶DNA文库将会包含多个变体DNA序列。在一些情况下, 变异在靶DNA序列中引入缺失、移码或插入。在一些情况下, 变体DNA序列导致每个基因或基因片段至少一个密码子的变异。在一些情况下, 将一部分基因插入靶DNA中, 或者备选地从靶DNA去除一部分靶DNA序列 (即基因的片段或整个基因)。在一些情况下, 变体DNA序列导致与基因或基因片段相关的至少一种转录调控序列例如启动子、UTR或终止子序列的变异。

[0124] 在一些情况下, 对于变体核酸文库生成, 包含核酸酶切割和同源重组以生成靶DNA文库的多样性, 其中该核酸酶是TALEN、ZFN、大范围核酸酶、Cas或Argonaute。在一些情况下, 当核酸酶是TALEN时, 合成TAL效应物DNA结合结构域文库 (通过从头合成RNA, 或者通过从头合成DNA然后进行转录和翻译 (体内或体外)), 其中该文库包含每个基因多个TAL效应物DNA结合结构域分子。例如, TAL效应物DNA结合结构域文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个TAL效应物DNA结合结构域分子。然后可以将TAL效应物DNA结合结构域文库与核酸酶混合以生成TALEN。在一些情况下, 将TALEN与靶DNA文库组合, 其中靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如, 靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下, 靶DNA文库包含来自途径中的多个基因或来自生物体中的所有基因的序列。在一些情况下, 还向混合物中添加替换序列, 该替换序列包括同源序列和变体核酸序列, 使得变异被引入靶DNA链。所得靶DNA文库将会包含多个变体DNA序列。在一些情况下, 变异在靶DNA序列中引入缺失、移码或插入。在一些情况下, 变体DNA序列导致每个基因或基因片段至少一个密码子的变异。在一些情况下, 将一部分基因插入靶DNA中, 或者备选地从靶DNA去除一部分靶DNA序列 (即基因的片段或整个基因)。在一些情况下, 变体DNA序列导致与基因或基因片段相关的至少一种转录调控序列例如启动子、UTR或终止子序列的变异。

[0125] 通过修饰的Cas9酶生成的变异

[0126] 在用于变体核酸文库生成的第二示例性过程中, 掺入修饰的Cas9酶以生成变体靶DNA文库。首先, 合成gRNA文库 (通过从头合成RNA, 或者通过从头合成DNA然后进行转录以生成gRNA), 其中文库包含每个基因多个gRNA分子。例如, gRNA文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个gRNA。将gRNA文库与修饰的Cas9酶和靶DNA文库混合, 其中靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如, 靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下, 靶DNA文库包含来自

途径中的多个基因或来自生物体中所有基因的序列。修饰的Cas9酶具有与其束缚的另一具有核酸序列修饰能力的酶。示例性的修饰的Cas9酶包括dCas9过程,其中损坏的或“死亡”的Cas9(“dCas9”)不再具有剪接功能,但是通过添加另一酶活性执行不同的靶分子修饰功能。例如,将胞苷脱氨酶束缚于dCas9将C-G DNA碱基对转化为T-A碱基对。在另一dCas9过程中,束缚于dCas9的不同酶导致在靶DNA中将碱基C变为T或将G变为A。所得靶DNA文库包含多种变体靶DNA序列。在一些情况下,变异在靶DNA序列中引入缺失、移码或插入。在一些情况下,变体DNA序列导致每个基因或基因片段至少一个密码子的变异。在一些情况下,变体DNA序列导致与基因或基因片段相关的至少一种转录调控序列例如启动子、UTR或终止子序列的变异。

[0127] 通过修饰的核酸酶生成的变异

[0128] 本文提供了包含修饰的核酸酶的用于生成变体核酸文库的方法,掺入修饰的核酸酶以生成变体靶DNA文库。在一些情况下,该核酸酶是BAL-31。在一些情况下,合成TAL效应物DNA结合结构域文库(通过从头合成RNA,或者通过从头合成DNA然后进行转录和翻译以生成TAL效应物DNA结合结构域文库),其中该文库包含每个基因多个不相同的核酸序列。例如,TAL效应物DNA结合结构域文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个不相同的核酸序列。然后将TAL效应物DNA结合结构域文库与核酸酶混合以生成TALEN。在一些情况下,随后将TALEN与靶DNA文库混合,其中该靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如,靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下,靶DNA文库包含来自途径中的多个基因或来自生物体中所有基因的序列。

[0129] 在一些情况下,该核酸酶是ZFN。在一些情况下,合成ZFD文库(通过从头合成RNA,或者通过从头合成DNA然后进行转录和翻译以生成ZFD文库),其中该文库包含每个基因多个不相同的核酸序列。例如,ZFD文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个不相同的核酸序列。然后将ZFD文库与核酸酶混合以生成ZFN。在一些情况下,然后将ZFN与靶DNA文库混合,其中该靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如,靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下,靶DNA文库包含来自途径中的多个基因或来自生物体中所有基因的序列。

[0130] 在一些情况下,该核酸酶是大范围核酸酶。在一些情况下,合成结合结构域文库如TAL效应物DNA结合结构域文库,以便将大范围核酸酶靶向特定核酸序列(通过从头合成RNA,或者通过从头合成DNA然后进行转录以生成结合结构域文库),其中结合结构域文库包含每个基因多个不相同的核酸序列。例如,结合结构域文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个不相同的核酸序列。然后将结合结构域文库与大范围核酸酶混合以生成工程化的大范围核酸酶。在一些情况下,然后将工程化的大范围核酸酶与靶DNA文库混合,其中该靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如,靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下,靶DNA文库包含来自途径中的多个基因或来自生物体中所有基因的序列。

[0131] 在一些情况下,该核酸酶是Argonaute。在一些情况下,合成指导核酸文库(gRNA或gDNA)(通过从头合成RNA,或者通过从头合成DNA然后进行转录以生成指导核酸文库),其中该指导核酸文库包含每个基因多个不相同的核酸序列。例如,指导核酸文库可以包含每个

基因1、2、3、4、5、6、7、8、9、10个或更多个不相同的核酸序列。将指导核酸文库与修饰的Argonaute酶和靶DNA文库混合,其中该靶DNA文库包含编码至少一个基因片段或至少一个基因的核酸序列。例如,靶DNA文库可包含1、2、3、4、5、6、7、8、9、10个或更多个基因或者基因片段。在一些情况下,靶DNA文库包含来自途径中的多个基因或来自生物体中所有基因的序列。

[0132] 在一些情况下,修饰的核酸酶具有与其束缚的另一具有核酸序列修饰能力的酶。示例性的修饰能力包括但不限于甲基化、脱甲基化、多腺苷酸化、脱腺苷化、脱氨基和多尿苷酰化。在一些情况下,包含多个变体靶DNA序列的靶DNA文库导致变异。在一些情况下,变异在靶DNA序列中引入缺失、移码或插入。在一些情况下,变体DNA序列导致每个基因或基因片段的至少一个密码子的变异。在一些情况下,变体DNA序列导致与基因或基因片段相关的至少一种转录调控序列例如启动子、UTR或终止子序列的变异。

[0133] 用于靶向模式系统的基因的gRNA文库合成

[0134] 本文提供了用本文所述的核酸文库筛选模式系统的方法。在一些情况下,核酸文库是本文所述的gRNA文库。在一些情况下,核酸文库是本文所述的DNA文库,其在转录时产生gRNA序列的转录。表4中提供了模式生物的非限制性示例性列表。

表4.生物和基因数

模式系统	编码蛋白质的基因*
拟南芥	27000
秀丽隐杆线虫	20000
家犬	19000
莱茵衣藻	14000
斑马鱼	26000
盘基网柄菌	13000
黑腹果蝇	14000
大肠杆菌	4300
猕猴	22000
小家鼠	20000
穴兔	27000
褐家鼠	22000
酿酒酵母	6600
野猪	21000
智人	21000

\*此处的数字反映了编码蛋白质的基因数,且排除tRNA和非编码RNA。Ron Milo&Rob Phillips,Cell Biology by the Numbers 286(2015)。

[0135] 合成gRNA的文库(通过从头合成RNA,或者通过从头合成DNA然后进行转录以生成gRNA),其中该文库包含每个基因多个gRNA分子。例如,本文所述的文库可以包含每个基因1、2、3、4、5、6、7、8、9、10个或更多个gRNA。在一些情况下,从头合成的文库内的核酸编码每个单个基因至少或约3个不相同的gRNA的序列。在一些情况下,核酸编码每个单个基因约1至约10个不相同的gRNA的序列。在一些情况下,核酸编码每个单个基因至少或约1个不相同的gRNA的序列。在一些情况下,核酸编码每个单个基因至多10个不相同的gRNA的序列。在一些情况下,核酸编码每个单个基因1至2、1至3、1至4、1至5、1至6、1至7、1至8、1至9或1至10个不相同的gRNA的序列。在一些情况下,gRNA是sgRNA。在一些情况下,gRNA是dgRNA。

[0136] 在一些情况下,本文所述的gRNA文库包含每个生物体基因一个或多个不相同的gRNA。在一些情况下,gRNA文库包含每个生物体基因1、2、3、4、5、6、7、8、9、10个或更多个不相同的gRNA。示例性的生物体包括但不限于拟南芥、秀丽隐杆线虫、家犬、莱茵衣藻、斑马鱼、盘基网柄菌、黑腹果蝇、大肠杆菌、猕猴、小家鼠、穴兔、褐家鼠、酿酒酵母、野猪和智人。在一些情况下,gRNA是sgRNA。在一些情况下,gRNA是dgRNA。在一些情况下,gRNA文库包含针对于生物体的整个基因组的至少或约5%的不相同的gRNA。在一些情况下,gRNA文库包含针对于生物体的整个基因组的约5%至约100%的不相同的gRNA。在一些情况下,gRNA文库包含针对于生物体的整个基因组的至少或约80%的不相同的gRNA。在一些情况下,sgRNA文库包含针对于生物体的整个基因组的至少或约90%的不相同的gRNA。在一些情况下,gRNA文库包含针对于生物体的整个基因组的至少或约95%的不相同的gRNA。在一些情况下,gRNA文库包含针对于生物体的整个基因组的至少或约100%的不相同的gRNA。在一些情况下,gRNA文库包含针对于生物体的整个基因组的约5%至10%、5%至20%、5%至30%、5%至40%、5%至50%、5%至60%、5%至70%、5%至80%、5%至90%、5%至95%、5%至100%、10%至20%、10%至30%、10%至40%、10%至50%、10%至60%、10%至70%、10%至80%、10%至90%、10%至95%、10%至100%、20%至30%、20%至40%、20%至50%、20%至60%、20%至70%、20%至80%、20%至90%、20%至95%、20%至100%、30%至40%、30%至50%、30%至60%、30%至70%、30%至80%、30%至90%、30%至95%、30%至100%、40%至50%、40%至60%、40%至70%、40%至80%、40%至90%、40%至95%、40%至100%、50%至60%、50%至70%、50%至80%、50%至90%、50%至95%、50%至100%、60%至70%、60%至80%、60%至90%、60%至95%、60%至100%、70%至80%、70%至90%、70%至95%、70%至100%、80%至90%、80%至95%、80%至100%、90%至95%、90%至100%或95%至100%的不相同的gRNA。在一些情况下,gRNA文库包含来自途径中的多个基因或来自生物体中所有基因的序列。在表4中列出的生物体中,gRNA的数目可以包括每个基因至少2X、3X、5X或10X。在一些情况下,gRNA文库靶向基因、一组基因(例如,3-10个基因)、途径(例如,10-100个基因)或底盘(chassis)(例如,100-1000个基因)中的至少一种。

[0137] 高度平行的从头核酸合成

[0138] 本文描述了利用端对端过程的小型化、平行化和垂直整合从寡核酸合成到硅上纳米孔内基因组组装的平台方法,以创建突破性的合成平台。本文所述的与96孔板具有相同占地面积的装置提供的硅合成平台与常规合成方法相比能够将通量提高100至1,000倍,在单

个高度平行化的运行中产生至多约1,000,000个寡核酸。在一些情况下,本文所述的单个硅板提供约6100个不相同的寡核酸的合成。在一些情况下,每个不相同的寡核酸位于簇内。簇可包含50至500个不相同的寡核酸。

[0139] 在一些情况下,与DNA的预定序列相比,编码本文所述gRNA文库的DNA文库具有小于1:500的错误率。在一些情况下,与DNA的预定序列相比,本文公开的从头寡核酸文库具有小于1:500、1:1000、1:1500、1:2000、1:3000、1:5000或更小的总错误率。在一些情况下,与DNA的预定序列相比,总错误率小于1:1000。错误率可以是总错误率或平均错误率。

[0140] 在一些情况下,与RNA的预定序列相比,编码本文所述gRNA文库的RNA文库具有小于1:500的错误率。在一些情况下,与RNA的预定序列相比,本文公开的从头寡核酸文库具有小于1:500、1:1000、1:1500、1:2000、1:3000、1:5000、1:10,000或更小的总错误率。在一些情况下,与RNA的预定序列相比,总错误率小于1:1000。

[0141] 基底

[0142] 在一些情况下,本文描述了包含多个簇的基底,其中每个簇包含支持寡核酸的附接和合成的多个座位。如本文所用的术语“座位”是指结构上的离散区域,其向编码单个预定序列的寡核酸提供支持以从表面延伸。在一些实施方案中,座位在二维表面上,例如,基本上平面的表面上。在一些实施方案中,座位在三维表面上,例如,孔、微孔、通道或柱上。在一些实施方案中,座位的表面包含积极官能化以与至少一个核苷酸附接以供寡核酸合成,或优选地与相同核苷酸群体附接以供合成寡核酸群体的材料。在一些实施方案中,寡核酸是指编码相同核酸序列的寡核酸群体。在一些情况下,基底的表面包括基底的一个或多个表面。使用所提供的系统和方法合成的文库内的寡核酸的平均错误率通常小于1/1000、小于约1/2000、小于约1/3000或更小。

[0143] 在一些实施方案中,基底包含支持在共同支持物上的可寻址位置处合成具有不同预定序列的多个寡核酸的表面。在一些实施方案中,基底向多于50、100、200、400、600、800、1000、1200、1400、1600、1800、2,000;5,000;10,000;20,000;50,000;100,000;200,000;300,000;400,000;500,000;600,000;700,000;800,000;900,000;1,000,000;1,200,000;1,400,000;1,600,000;1,800,000;2,000,000;2,500,000;3,000,000;3,500,000;4,000,000;4,500,000;5,000,000;10,000,000个或更多个不相同的寡核酸的合成提供支持。在一些情况下,基底向编码不同序列的多于50、100、200、400、600、800、1000、1200、1400、1600、1800、2,000;5,000;10,000;20,000;50,000;100,000;200,000;300,000;400,000;500,000;600,000;700,000;800,000;900,000;1,000,000;1,200,000;1,400,000;1,600,000;1,800,000;2,000,000;2,500,000;3,000,000;3,500,000;4,000,000;4,500,000;5,000,000;10,000,000个或更多个寡核酸的合成提供支持。在一些实施方案中,至少一部分寡核酸具有相同的序列或被配置成用相同的序列合成。在一些实施方案中,基底像具有至少80、90、100、120、150、175、200、225、250、275、300、325、350、375、400、425、450、475、500个碱基或更多个碱基的寡核酸的生长提供表面环境。

[0144] 在一些实施方案中,在基底的不同座位上合成寡核酸,其中每个座位支持寡核酸群体的合成。在一些情况下,每个座位支持寡核酸群体的合成,该寡核酸群体与在另一座位上生长的寡核酸群体具有不同的序列。在一些实施方案中,基底的座位位于多个簇内。在一些情况下,基底包含至少10、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、

10000、11000、12000、13000、14000、15000、20000、30000、40000、50000个或更多个簇。在一些实施方案中,基底包含多于2,000;5,000;10,000;100,000;200,000;300,000;400,000;500,000;600,000;700,000;800,000;900,000;1,000,000;1,100,000;1,200,000;1,300,000;1,400,000;1,500,000;1,600,000;1,700,000;1,800,000;1,900,000;2,000,000;300,000;400,000;500,000;600,000;700,000;800,000;900,000;1,000,000;1,200,000;1,400,000;1,600,000;1,800,000;2,000,000;2,500,000;3,000,000;3,500,000;4,000,000;4,500,000;5,000,000;或10,000,000个或更多个不同的座位。在一些实施方案中,基底包含约10,000个不同的座位。在不同的实施方案中,单个簇内的座位的量有所变化。在一些情况下,每个簇包含1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、120、130、150、200、300、400、500个或更多个座位。在一些实施方案中,每个簇包含约50-500个座位。在一些实施方案中,每个簇包含约100-200个座位。在一些实施方案中,每个簇包含约100-150个座位。在一些实施方案中,每个簇包含约109、121、130或137个座位。在一些实施方案中,每个簇包含约19、20、61、64个或更多个座位。

[0145] 本文提供了在硅板上合成不相同的寡核酸的方法。在一些情况下,硅板包含约1-10、1-50或50-500个簇。在一些情况下,硅板包含多于约50、100、250、500、2500、5000、6000、6150、10000个或更多个簇。在一些情况下,每个簇包含121个座位。在一些情况下,每个簇包含约50-500、50-200、100-150个座位。在一些情况下,每个簇包含至少约50、100、150、200、500、1000个或更多个座位。在一些情况下,单个板包含约100、500、10000、20000、30000、50000、100000、500000、700000、1000000个或更多个座位。

[0146] 在一些实施方案中,在基底上合成的不同寡核酸的数目取决于基底中可用的不同座位的数目。在一些实施方案中,基底的簇内的座位密度为至少或约1个座位/ $\text{mm}^2$ 、10个座位/ $\text{mm}^2$ 、25个座位/ $\text{mm}^2$ 、50个座位/ $\text{mm}^2$ 、65个座位/ $\text{mm}^2$ 、75个座位/ $\text{mm}^2$ 、100个座位/ $\text{mm}^2$ 、130个座位/ $\text{mm}^2$ 、150个座位/ $\text{mm}^2$ 、175个座位/ $\text{mm}^2$ 、200个座位/ $\text{mm}^2$ 、300个座位/ $\text{mm}^2$ 、400个座位/ $\text{mm}^2$ 、500个座位/ $\text{mm}^2$ 、1,000个座位/ $\text{mm}^2$ 或更大。在一些情况下,基底包含每 $\text{mm}^2$ 至约500 $\text{mm}^2$ 约10个座位、每 $\text{mm}^2$ 至约400 $\text{mm}^2$ 约25个座位、每 $\text{mm}^2$ 至约500 $\text{mm}^2$ 约50个座位、每 $\text{mm}^2$ 至约500 $\text{mm}^2$ 约100个座位、每 $\text{mm}^2$ 至约500 $\text{mm}^2$ 约150个座位、每 $\text{mm}^2$ 至约250 $\text{mm}^2$ 约10个座位、每 $\text{mm}^2$ 至约250 $\text{mm}^2$ 约50个座位、每 $\text{mm}^2$ 至约200 $\text{mm}^2$ 约10个座位或每 $\text{mm}^2$ 至约200 $\text{mm}^2$ 约50个座位。在一些实施方案中,簇内两个相邻座位的中心之间的距离为约10 $\mu\text{m}$ 至约500 $\mu\text{m}$ 、约10 $\mu\text{m}$ 至约200 $\mu\text{m}$ 或约10 $\mu\text{m}$ 至约100 $\mu\text{m}$ 。在一些情况下,相邻座位的两个中心之间的距离大于约10 $\mu\text{m}$ 、20 $\mu\text{m}$ 、30 $\mu\text{m}$ 、40 $\mu\text{m}$ 、50 $\mu\text{m}$ 、60 $\mu\text{m}$ 、70 $\mu\text{m}$ 、80 $\mu\text{m}$ 、90 $\mu\text{m}$ 或100 $\mu\text{m}$ 。在一些情况下,两个相邻座位的中心之间的距离小于约200 $\mu\text{m}$ 、150 $\mu\text{m}$ 、100 $\mu\text{m}$ 、80 $\mu\text{m}$ 、70 $\mu\text{m}$ 、60 $\mu\text{m}$ 、50 $\mu\text{m}$ 、40 $\mu\text{m}$ 、30 $\mu\text{m}$ 、20 $\mu\text{m}$ 或10 $\mu\text{m}$ 。在一些情况下,每个座位具有约0.5 $\mu\text{m}$ 、1 $\mu\text{m}$ 、2 $\mu\text{m}$ 、3 $\mu\text{m}$ 、4 $\mu\text{m}$ 、5 $\mu\text{m}$ 、6 $\mu\text{m}$ 、7 $\mu\text{m}$ 、8 $\mu\text{m}$ 、9 $\mu\text{m}$ 、10 $\mu\text{m}$ 、20 $\mu\text{m}$ 、30 $\mu\text{m}$ 、40 $\mu\text{m}$ 、50 $\mu\text{m}$ 、60 $\mu\text{m}$ 、70 $\mu\text{m}$ 、80 $\mu\text{m}$ 、90 $\mu\text{m}$ 或100 $\mu\text{m}$ 的宽度。在一些情况下,每个座位具有约0.5 $\mu\text{m}$ 至100 $\mu\text{m}$ 、约0.5 $\mu\text{m}$ 至50 $\mu\text{m}$ 、约10 $\mu\text{m}$ 至75 $\mu\text{m}$ 或约0.5 $\mu\text{m}$ 至50 $\mu\text{m}$ 的宽度。

[0147] 在一些实施方案中,基底内簇的密度为至少或约每100 $\text{mm}^2$  1个簇、每10 $\text{mm}^2$  1个簇、每5 $\text{mm}^2$  1个簇、每4 $\text{mm}^2$  1个簇、每3 $\text{mm}^2$  1个簇、每2 $\text{mm}^2$  1个簇、每1 $\text{mm}^2$  1个簇、每1 $\text{mm}^2$  2个簇、每1 $\text{mm}^2$  3个簇、每1 $\text{mm}^2$  4个簇、每1 $\text{mm}^2$  5个簇、每1 $\text{mm}^2$  10个簇、每1 $\text{mm}^2$  50个簇或更达。在一些实施方案中,基底包含约每10 $\text{mm}^2$  1个簇至约每1 $\text{mm}^2$  10个簇。在一些实施方案中,两个相邻簇的中心之间的距离小于约50 $\mu\text{m}$ 、100 $\mu\text{m}$ 、200 $\mu\text{m}$ 、500 $\mu\text{m}$ 、1000 $\mu\text{m}$ 或2000 $\mu\text{m}$ 或5000 $\mu\text{m}$ 。在一些

情况下,两个相邻簇的中心之间的距离为约50 $\mu\text{m}$ 至约100 $\mu\text{m}$ 、约50 $\mu\text{m}$ 至约200 $\mu\text{m}$ 、约50 $\mu\text{m}$ 至约300 $\mu\text{m}$ 、约50 $\mu\text{m}$ 至约500 $\mu\text{m}$ 和约100 $\mu\text{m}$ 至约2000 $\mu\text{m}$ 。在一些情况下,两个相邻簇的中心之间的距离为约0.05mm至约50mm、约0.05mm至约10mm、约0.05mm至约5mm、约0.05mm至约4mm、约0.05mm至约3mm、约0.05mm至约2mm、约0.1mm至10mm、约0.2mm至10mm、约0.3mm至约10mm、约0.4mm至约10mm、约0.5mm至10mm、约0.5mm至约5mm或约0.5mm至约2mm。在一些情况下,每个簇具有约0.5至2mm、约0.5至1mm或约1至2mm的横截面。在一些情况下,每个簇具有约0.5、0.6、0.7、0.8、0.9、1、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9或2mm的横截面。在一些情况下,每个簇具有约0.5、0.6、0.7、0.8、0.9、1、1.1、1.15、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9或2mm的内部横截面。

[0148] 在一些实施方案中,基底为约标准96孔板的大小,例如约100至200mm乘以约50至150mm。在一些实施方案中,基底具有小于或等于约1000mm、500mm、450mm、400mm、300mm、250mm、200mm、150mm、100mm或50mm的直径。在一些实施方案中,基底的直径为约25mm至1000mm、约25mm至约800mm、约25mm至约600mm、约25mm至约500mm、约25mm至约400mm、约25mm至约300mm或约25mm至约200mm。基底大小的非限制性实例包括约300mm、200mm、150mm、130mm、100mm、76mm、51mm和25mm。在一些实施方案中,基底具有至少约100mm<sup>2</sup>; 200mm<sup>2</sup>; 500mm<sup>2</sup>; 1,000mm<sup>2</sup>; 2,000mm<sup>2</sup>; 5,000mm<sup>2</sup>; 10,000mm<sup>2</sup>; 12,000mm<sup>2</sup>; 15,000mm<sup>2</sup>; 20,000mm<sup>2</sup>; 30,000mm<sup>2</sup>; 40,000mm<sup>2</sup>; 50,000mm<sup>2</sup>或更大的平面表面积。在一些实施方案中,基底的厚度为约50mm至约2000mm、约50mm至约1000mm、约100mm至约1000mm、约200mm至约1000mm或约250mm至约1000mm。基底厚度的非限制性实例包括275mm、375mm、525mm、625mm、675mm、725mm、775mm和925mm。在一些情况下,基底的厚度随直径变化并且取决于基底的组成。例如,包含除硅之外的材料的基底具有与相同直径的硅基底不同的厚度。基底厚度可以通过所用材料的机械强度来决定,并且基底必须足够厚以支持其自身重量而不会在处理期间破裂。

#### [0149] 表面材料

[0150] 本文提供的基底、装置和反应器由适用于本文所述方法和组合物的任何种类的材料制造。在某些实施方案中,基底材料被制造以展现出低水平的核苷酸结合。在一些情况下,基底材料被修饰以生成展现出高水平核苷酸结合的不同表面。在一些实施方案中,基底材料对可见光和/或UV光是透明的。在一些实施方案中,基底材料是足够导电的,例如,能够在整个基底或一部分基底上形成均匀的电场。在一些实施方案中,导电材料与电接地连接。在一些情况下,基底是导热的或绝缘的。在一些情况下,材料具有耐化学性和耐热性以支持化学或生物化学反应,例如寡核酸合成反应过程。在一些实施方案中,基底包含柔性材料。柔性材料包括但不限于改性尼龙、未改性尼龙、硝化纤维素、聚丙烯等。在一些实施方案中,基底包含刚性材料。刚性材料包括但不限于玻璃、熔融石英、硅、二氧化硅、氮化硅、塑料(例如,聚四氟乙烯、聚丙烯、聚苯乙烯、聚碳酸酯及其共混物等)和金属(例如,金、铂等)。在一些实施方案中,基底由包括硅、聚苯乙烯、琼脂糖、葡聚糖、纤维素聚合物、聚丙烯酰胺、聚二甲基硅氧烷(PDMS)、玻璃或其任何组合的材料制造。在一些情况下,用本文列出的材料的组合或本领域已知的任何其他合适的材料制造基底。

#### [0151] 表面架构

[0152] 在各个实施方案中,基底包括凸起和/或凹陷的特征。具有这样的特征的一个好处是表面积增加以支持寡核酸合成。在一些实施方案中,具有凸起和/或凹陷特征的基底被称

为三维基底。在一些情况下,三维基底包含一个或多个通道。在一些情况下,一个或多个座位包含通道。在一些情况下,通道可经由诸如寡核酸合成仪等沉积装置进行试剂沉积。在一些情况下,试剂和/或流体收集在与一个或多个通道流体连通的较大的孔中。例如,基底包含与伴有簇的多个座位对应的多个通道,并且该多个通道与该簇的一个孔流体连通。在一些方法中,寡核酸的文库在簇的多个座位中合成。

[0153] 在一些实施方案中,该结构被配置成允许用于表面上的寡核酸合成的受控流动和质量传递路径。在一些实施方案中,基底的配置允许在寡核酸合成期间的质量传递路径、化学暴露时间和/或洗涤功效的受控和均匀的分布。在一些实施方案中,基底的配置允许提高的扫描效率,例如通过向生长寡核酸提供足够的体积,使得通过生长寡核酸排除的体积占据可用于或适用于生长寡核酸的最初可用体积的不超过50%、45%、40%、35%、30%、25%、20%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或更少。在一些实施方案中,三维结构允许被管理的流体的流动,以允许化学暴露的快速交换。

[0154] 在一些实施方案中,通过物理结构实现隔离。在一些实施方案中,通过将表面进行差异官能化产生对于寡核酸合成活性和惰性的区域来实现隔离。还可以通过在基底表面上交替疏水性来实现差异官能化,从而产生引起沉积的试剂的成珠或润湿的水接触角效应。采用较大的结构可以减少邻近斑点的试剂对不同的寡核酸合成位置的飞溅和交叉污染。在一些情况下,使用诸如寡核酸合成仪等装置将试剂沉积到不同的寡核酸合成位置。以允许以低错误率(例如,小于约1:500、1:1000、1:1500、1:2,000;1:3,000;1:5,000;或1:10,000)合成较大数目的寡核酸(例如,多于约10,000个)的方式配置具有三维特征的基底。在一些情况下,基底包含密度为约或大于约1、5、10、20、30、40、50、60、70、80、100、110、120、130、140、150、160、170、180、190、200、300、400或500个特征/ $\text{mm}^2$ 的特征。

[0155] 基底的孔可以具有与基底的另一孔相同或不同的宽度、高度和/或体积。基底的通道可以具有与基底的另一通道相同或不同的宽度、高度和/或体积。在一些实施方案中,簇的直径或包含簇的孔的直径或两者为约0.05mm至约50mm、约0.05mm至约10mm、约0.05mm至约5mm、约0.05mm至约4mm、约0.05mm至约3mm、约0.05mm至约2mm、约0.05mm至约1mm、约0.05mm至约0.5mm、约0.05mm至约0.1mm、约0.1mm至10mm、约0.2mm至10mm、约0.3mm至约10mm、约0.4mm至约10mm、约0.5mm至10mm、约0.5mm至约5mm或约0.5mm至约2mm。在一些实施方案中,簇或孔或两者的直径为小于或约为5mm、4mm、3mm、2mm、1mm、0.5mm、0.1mm、0.09mm、0.08mm、0.07mm、0.06mm或0.05mm。在一些实施方案中,簇或孔或两者的直径为约1.0至1.3mm。在一些实施方案中,簇或孔的或两者直径为约1.150mm。在一些实施方案中,簇或孔或两者的直径为约0.08mm。簇的直径是指二维或三维基底内的簇。

[0156] 在一些实施方案中,孔的高度为约20 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 、约50 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 、约100 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 、约200 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 、约300 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 、约400 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 或约500 $\mu\text{m}$ 至约1000 $\mu\text{m}$ 。在一些情况下,孔的高度为小于约1000 $\mu\text{m}$ 、小于约900 $\mu\text{m}$ 、小于约800 $\mu\text{m}$ 、小于约700 $\mu\text{m}$ 或小于约600 $\mu\text{m}$ 。

[0157] 在一些实施方案中,基底包含与簇内多个座位相对应的多个通道,其中通道的高度或深度为约5 $\mu\text{m}$ 至约500 $\mu\text{m}$ 、约5 $\mu\text{m}$ 至约400 $\mu\text{m}$ 、约5 $\mu\text{m}$ 至约300 $\mu\text{m}$ 、约5 $\mu\text{m}$ 至约200 $\mu\text{m}$ 、约5 $\mu\text{m}$ 至约100 $\mu\text{m}$ 、约5 $\mu\text{m}$ 至约50 $\mu\text{m}$ 或约10 $\mu\text{m}$ 至约50 $\mu\text{m}$ 。在一些情况下,通道的高度小于100 $\mu\text{m}$ 、小于

80um、小于60um、小于40um或小于20um。

[0158] 在一些实施方案中,通道、座位(例如,在基本上平面的基底中)或通道和座位两者(例如,在其中座位与通道相对应的三维基底中)的直径为约1um至约1000um、约1um至约500um、约1um至约200um、约1um至约100um、约5um至约100um或约10um至约100um,例如,约90um、80um、70um、60um、50um、40um、30um、20um或10um。在一些实施方案中,通道、座位或通道和座位两者的直径小于约100um、90um、80um、70um、60um、50um、40um、30um、20um或10um。在一些实施方案中,两个相邻通道的中心之间、座位之间或通道与座位之间的距离为约1um至约500um、约1um至约200um、约1um至约100um、约5um至约200um、约5um至约100um、约5um至约50um或约5um至约30um,例如,约20um。

[0159] 表面修饰

[0160] 在各个实施方案中,采用表面修饰,以通过添加或减去过程对表面进行化学和/或物理改变,从而改变基底表面或者基底表面的选定位点或区域的一种或多种化学和/或物理性质。例如,表面修饰包括但不限于(1)改变表面的润湿性质,(2)对表面进行官能化,即提供、修饰或取代表面官能团,(3)对表面进行去官能化,即去除表面官能团,(4)以其他方式改变表面的化学组成,例如通过蚀刻,(5)增大或减小表面粗糙度,(6)在表面上提供涂层,例如展现出与表面的润湿性质不同的润湿性质的涂层,和/或(7)在表面上沉积颗粒。

[0161] 在一些情况下,在表面上方添加化学层(称为粘合促进剂)有助于基底表面上座位的结构化图案化。用于应用粘合促进剂的示例性表面包括但不限于玻璃、硅、二氧化硅和氮化硅。在一些情况下,粘合促进剂是具有高表面能的化学品。在一些实施方案中,第二化学层沉积在基底的表面上。在一些情况下,第二化学层具有低表面能。在一些情况下,涂覆在表面上的化学层的表面能支持液滴在表面上的定位。取决于所选择的图案化布置,座位的接近度和/或座位处的流体接触面积是可变的。

[0162] 在一些实施方案中,其上沉积有核酸或其他部分(例如用于寡核酸合成)的基底表面或分离的(resolved)座位是光滑的或基本上平面的(例如,二维的),或者具有不规则性如凸起或凹陷的特征(例如,三维特征)。在一些实施方案中,用一种或多种不同的化合物层修饰基底表面。这样的感兴趣的修饰层包括但不限于无机层和有机层,诸如金属、金属氧化物、聚合物、小有机分子等。非限制性聚合物层包括肽、蛋白质、核酸或其模拟物(例如,肽核酸等)、多糖、磷脂、聚氨酯、聚酯、聚碳酸酯、聚脲、聚酰胺、聚乙烯胺、聚芳硫醚、聚硅氧烷、聚酰亚胺、聚乙酸酯以及本文所述或本领域另外已知的任何其他合适的化合物。在一些情况下,聚合物是杂聚的。在一些情况下,聚合物是均聚的。在一些情况下,聚合物包含官能部分或是缀合的。

[0163] 在一些实施方案中,用一个或多个增加和/或降低表面能的部分使基底的分离座位官能化。在一些情况下,部分是化学惰性的。在一些情况下,部分被配置用于支持期望的化学反应,例如,寡核酸合成反应中的一个或多个过程。表面的表面能或疏水性是用于确定核苷酸附接在该表面上的亲和力的因素。在一些实施方案中,用于基底官能化的方法包括:(a)提供具有包含二氧化硅的表面的基底;以及(b)使用本文所述或本领域另外已知的合适的硅烷化剂,例如有机官能烷氧基硅烷分子使表面硅烷化。在一些情况下,有机官能烷氧基硅烷分子包括二甲基氯-十八烷基-硅烷、甲基二氯-十八烷基-硅烷、三氯-十八烷基-硅烷、三甲基-十八烷基-硅烷、三乙基-十八烷基-硅烷或其任何组合。在一些实施方案中,基底表

面包含用聚乙烯/聚丙烯(通过 $\gamma$ 辐射或铬酸氧化官能化,并还原成羟烷基表面)、高度交联的聚苯乙烯-二乙烯基苯(通过氯甲基化衍生化,并胺化成苄胺官能表面)、尼龙(末端氨基己基基团是直接反应性的)官能化,或用还原的聚四氟乙烯蚀刻。其他方法和官能化剂描述于美国专利号5474796中,其通过引用以其全文并入本文。

[0164] 在一些实施方案中,通常经由存在于基底表面上的反应性亲水部分,在将硅烷有效偶联至基底表面的反应条件下,通过与含有硅烷混合物的衍生化组合物接触来使基底表面官能化。硅烷化通常通过用有机官能烷氧基硅烷分子的自组装来覆盖表面。如本领域目前已知的,可以进一步使用各种硅氧烷官能化试剂,例如用于降低或增加表面能。有机官能烷氧基硅烷根据其有机官能而分类。硅氧烷官能化试剂的非限制性实例包括羟烷基硅氧烷(甲硅烷基化表面,用乙硼烷使其官能化并用过氧化氢将醇氧化)、二醇(二羟基烷基)硅氧烷(甲硅烷基化表面,并水解成二醇)、氨基烷基硅氧烷(胺不需要中间官能化步骤)、环氧丙氧基硅烷(3-环氧丙氧基丙基-二甲基-乙氧基硅烷、环氧丙氧基-三甲氧基硅烷)、巯基硅烷(3-巯基丙基-三甲氧基硅烷、3-4-环氧环己基-乙基三甲氧基硅烷或3-巯基丙基-甲基-二甲氧基硅烷)、双环庚基-三氯硅烷、丁基-醛-三甲氧基硅烷或二聚仲氨基烷基硅氧烷。示例性的羟烷基硅氧烷包括变为3-羟基丙基的烯丙基三氯硅烷,或变为8-羟基辛基的7-辛-1-烯基三氯硅烷。二醇(二羟基烷基)硅氧烷包括环氧丙基三甲氧基硅烷衍生的(2,3-二羟基丙基氧基)丙基(GOPS)。氨基烷基硅氧烷包括变为3-氨基丙基的3-氨基丙基三甲氧基硅烷(3-氨基丙基-三乙氧基硅烷、3-氨基丙基-二乙氧基-甲基硅烷、3-氨基丙基-二甲基-乙氧基硅烷或3-氨基丙基-三甲氧基硅烷)。示例性的二聚仲氨基烷基硅氧烷包括变为二(甲硅氧基丙基)胺的二(3-三甲氧基甲硅烷基丙基)胺。在一些实施方案中,官能化剂包括11-乙酰氧基十一烷基三乙氧基硅烷、正癸基三乙氧基硅烷、(3-氨基丙基)三甲氧基硅烷、(3-氨基丙基)三乙氧基硅烷、环氧丙氧基丙基/三甲氧基硅烷和N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺。

#### [0165] 寡核酸合成

[0166] 在各个实施方案中,用于寡核酸合成的方法包括涉及亚磷酰胺化学的方法。在一些实施方案中,寡核酸合成包括使碱基与亚磷酰胺偶联。在一些实施方案中,寡核酸合成包括在偶联条件下通过亚磷酰胺沉积来偶联碱基,其中相同的碱基任选地与亚磷酰胺沉积多于一次,即双偶联。在一些实施方案中,寡核酸合成包括对未反应的位点进行加帽。在一些情况下,加帽是任选的。在一些实施方案中,寡核酸合成包括氧化。在一些实施方案中,寡核酸合成包括去封闭或脱三苯甲基。在一些实施方案中,寡核酸合成包括硫化。在一些情况下,寡核酸合成包括氧化或硫化。在一些实施方案中,在寡核酸合成反应期间的一个或每个步骤之间,使用例如四唑或乙腈来洗涤基底。亚磷酰胺合成方法中任何一步的时间范围包括短于约2min、1min、50sec、40sec、30sec、20sec和10sec。

[0167] 使用亚磷酰胺方法合成寡核酸包括随后将亚磷酰胺构成单元(例如,核苷亚磷酰胺)添加至生长的寡核酸链,以形成亚磷酸三酯键。亚磷酰胺寡核酸合成以3'至5'方向进行。亚磷酰胺寡核酸合成允许每个合成循环将一个核苷酸受控地添加至生长的核酸链。在一些实施方案中,每个合成循环包括偶联步骤。亚磷酰胺偶联涉及在活化的核苷亚磷酰胺和与基底结合的核苷之间形成亚磷酸三酯键,例如,经由连接体。在一些实施方案中,将核苷亚磷酰胺提供给活化的基底。在一些实施方案中,将核苷亚磷酰胺与激活剂一起提供给

基底。在一些实施方案中,将核苷亚磷酰胺以相对于与基底结合的核苷1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、50、60、70、80、90、100倍过量或更多提供给基底。在一些实施方案中,核苷亚磷酰胺的添加在无水环境中例如在无水乙醇中进行。在添加核苷亚磷酰胺后,任选地洗涤基底。在一些实施方案中,将偶联步骤再重复一次或多次,任选地在将核苷亚磷酰胺添加到基底之间进行洗涤步骤。在一些实施方案中,本文使用的寡核酸合成方法包括1、2、3个或更多个连续的偶联步骤。在许多情况下,在偶联之前通过去除保护基团使与基底结合的核苷脱保护,其中该保护基团起到防止聚合的作用。常见的保护基团是4,4'-二甲氧基三苯甲基(DMT)。

[0168] 偶联后,亚磷酰胺寡核酸合成方法任选地包括加帽步骤。在加帽步骤中,用加帽剂处理生长的寡核酸。加帽步骤可用于在偶联后封闭未反应的基底结合的5'-OH基团不进行进一步链延伸,从而防止形成具有内部碱基缺失的寡核酸。此外,用1H-四唑激活的亚磷酰胺可以在很小程度上与鸟苷的O6位置反应。不受理论束缚,在用I<sub>2</sub>/水氧化后,该副产物可经历脱嘌呤,这可能经由O6-N7迁移。在寡核酸的最终脱保护过程中,脱嘌呤位点可能最终被切割,从而降低了全长产物的产率。在用I<sub>2</sub>/水氧化之前,可以通过用加帽剂处理来去除O6修饰。在一些实施方案中,与没有加帽的合成相比,在寡核酸合成期间包括加帽步骤降低了错误率。作为实例,加帽步骤包括用乙酸酐和1-甲基咪唑的混合物处理基底结合的寡核酸。在加帽步骤后,任选地洗涤基底。

[0169] 在一些实施方案中,在添加核苷亚磷酰胺后,且任选地在加帽和一个或多个洗涤步骤之后,将基底结合的生长核酸进行氧化。氧化步骤包括将亚磷酸三酯氧化成四配位的磷酸三酯,其是天然存在的磷酸二酯核苷间键的受保护的前体。在一些情况下,通过用碘和水处理,任选地在弱碱(例如,吡啶、二甲基吡啶、三甲基吡啶)存在下,实现生长的寡核酸的氧化。氧化可以在无水条件下使用例如叔丁基氢过氧化物或(1S)-(+)-(10-樟脑磺酰基)-氧杂吡丙啶(CSO)进行。在一些方法中,在氧化后进行加帽步骤。第二加帽步骤允许基底干燥,因为可能继续存在的来自氧化的残余水可抑制随后的偶联。在氧化后,任选地洗涤基底和生长的寡核酸。在一些实施方案中,氧化步骤被硫化步骤代替以获得寡核苷酸硫代磷酸酯,其中可在硫化后进行任何加帽步骤。许多试剂能够进行有效的硫转移,包括但不限于3-(二甲基氨基亚甲基)氨基)-3H-1,2,4-二噻唑-3-硫酮、DDTT、3H-1,2-苯并二硫-3-酮1,1-二氧化物(也称为Beaucage试剂)和二硫化N,N,N',N'-四乙基秋兰姆(TETD)。

[0170] 为了使核苷掺入的后续循环通过偶联而发生,移除基底结合的生长寡核酸的受保护的5'末端,使得伯羟基基团与下一个核苷亚磷酰胺反应。在一些实施方案中,保护基团是DMT,并且在二氯甲烷中用三氯乙酸进行去封闭。进行延长时间的脱三苯甲基化或用强于推荐的酸溶液进行脱三苯甲基化可导致固体支持物结合的寡核苷酸的脱嘌呤增加,并因此降低期望的全长产物的产率。本文所述的本发明的方法和组合物提供了受控去封闭条件,从而限制不需要的脱嘌呤反应。在一些情况下,在去封闭后洗涤基底结合的寡核苷酸。在一些情况下,去封闭后的有效洗涤有助于合成的寡核酸具有低错误率。

[0171] 用于合成寡核酸的方法通常包括以下步骤的迭代顺序:将受保护的单体施加于活性官能化的表面(例如,座位)以与活化的表面、连接体或与先前脱保护的单体连接;使所施加的单体脱保护,使其与随后施加的保护单体反应;以及施加另一受保护的单体进行连接。一个或多个中间步骤包括氧化或硫化。在一些情况下,一个或多个洗涤步骤在一个或所有

步骤之前或之后。

[0172] 用于基于亚磷酰胺的寡核酸合成的方法包括一系列化学步骤。在一些实施方案中,合成方法的一个或多个步骤涉及试剂循环,其中该方法的一个或多个步骤包括向基底施加可用于该步骤的试剂。例如,通过一系列液相沉积和真空干燥步骤使试剂循环。对于包含三维特征如孔、微孔、通道等的基底,试剂任选地经由孔和/或通道穿过基底的一个或多个区域。

[0173] 在各个实施方案中,使用本文所述的方法和/或基底合成的寡核酸包含至少约20、30、40、50、60、70、75、80、90、100、120、150个或更多个碱基。在一些实施方案中,在座位内合成至少约1pmol、10pmol、20pmol、30pmol、40pmol、50pmol、60pmol、70pmol、80pmol、90pmol、100pmol、150pmol、200pmol、300pmol、400pmol、500pmol、600pmol、700pmol、800pmol、900pmol、1nmol、5nmol、10nmol、100nmol或更多的寡核酸。本文提供的用于在表面上合成寡核酸的方法允许快速合成。例如,每小时合成至少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、35、40、45、50、55、60、70、80、90、100、125、150、175、200个核苷酸或更多。核苷酸包括腺嘌呤、鸟嘌呤、胸腺嘧啶、胞嘧啶、尿苷构成单元或其类似物/修饰形式。在一些实施方案中,在基底上平行合成寡核酸文库。例如,包含约或至少约100;1,000;10,000;100,000;1,000,000;2,000,000;3,000,000;4,000,000;或5,000,000个分离座位的基底能够支持至少相同数目的不同寡核酸的合成,其中编码不同序列的寡核酸在分离的座位上合成。在一些实施方案中,寡核酸文库在少于约三个月、两个月、一个月、三周、15、14、13、12、11、10、9、8、7、6、5、4、3、2天、24小时或更短的时间内以本文所述的低错误率在基底上合成。在一些实施方案中,由使用本文所述的基底和方法以低错误率合成的寡核酸文库组装的较大核酸在少于约三个月、两个月、一个月、三周、15、14、13、12、11、10、9、8、7、6、5、4、3、2天、24小时或更短的时间内制备。

[0174] 一旦用于生成的大寡核酸被选择,设计用于从头合成的预定寡核酸文库。已知各种合适的方法用于生成高密度寡核酸阵列。在工作流程实例中,提供了基底表面层。在该实例中,改变表面的化学性质以改善寡核酸合成过程。生成低表面能的区域以排斥液体,同时生成高表面能的区域以吸引液体。表面本身可以是平面表面的形式或包含形状的变化,如增加表面积的突出物或微孔。在工作流程实例中,所选择的高表面能分子可用于支持DNA化学的双重功能,如国际专利申请公开W0/2015/021080中所公开的,其通过引用以其全文并入本文。

[0175] 寡核酸阵列的原位制备在固体支持物上生成,并利用单核苷酸延伸过程平行延伸多个寡聚体。沉积装置如寡核酸合成仪被设计成以逐步方式释放试剂,使得多个寡核酸平行地一次延伸一个残基,以生成具有预定核酸序列的寡聚体。在一些情况下,在该阶段从表面切割寡核酸。切割包括气体切割,例如利用氨或甲胺。

[0176] 计算机系统

[0177] 本文所述的任何系统可以可操作地连接到计算机,并且可以本地或远程地通过计算机自动化。在各个实施方案中,本发明的方法和系统可以进一步包括计算机系统上的软件程序及其用途。因此,针对于分配/真空/再填充功能的同步化如编排并同步化材料沉积装置移动、分配行动和真空致动的计算机控制都在本发明的范围内。计算机系统可以被编程用于在用户指定的碱基序列与材料沉积装置的位置之间进行接口,以将正确的试剂递送

到基底的指定区域。

[0178] 图8中图示的计算机系统800可以被理解为可以从媒体811和/或网络端口805读取指令的逻辑设备,其可以任选地连接到具有固定媒体812的服务器809。如图8所示的系统可包括CPU 801、磁盘驱动器803、任选的输入装置如键盘815和/或鼠标816、以及任选的监视器807。可以通过与指示的通信媒介实现与本地或远程位置处的服务器的数据通信。通信媒介可以包括发送和/或接收数据的任何手段。例如,通信媒介可以是网络连接、无线连接或互联网连接。这样的连接可以通过万维网提供通信。预期与本公开内容相关的数据可以通过这样的网络或连接传输,以供由如图8所示的一方822接收和/或审阅。

[0179] 图9是图示可以与本发明的示例性实施方案结合使用的计算机系统900的第一示例性架构的框图。如图9所示,示例性计算机系统可包括用于处理指令的处理器902。处理器的非限制性示例包括: Intel Xeon™处理器、AMD Opteron™处理器、Samsung 32位RISC ARM 1176JZ (F)-S v1.0™处理器、ARM Cortex-A8 Samsung S5PC100™处理器、ARM Cortex-A8 Apple A4™处理器、Marvell PXA 930™处理器或功能上等效的处理器。多个执行线程可用于平行处理。在一些实施方案中,还可以使用多个处理器或具有多个核的处理器,无论在单个计算机系统中、在群集中,还是通过包含多个计算机、蜂窝电话和/或个人数据助理装置的网络跨系统分布。

[0180] 如图9所示,高速缓冲存储器904可以连接到处理器902或者包含在处理器902中,以向处理器902最近或经常使用的指令或数据提供高速存储器。处理器902通过处理器总线908连接到北桥906。北桥906通过存储器总线912连接到随机存取存储器 (RAM) 910,并管理处理器902对RAM 910的访问。北桥906还通过芯片组总线916连接到南桥914。南桥914转而连接到外围总线918。外围总线可以是例如PCI、PCI-X、PCI Express或其他外围总线。北桥和南桥通常被称为处理器芯片组,并管理处理器、RAM与外围总线918上的外围组件之间的数据传输。在一些替代架构中,北桥的功能可以合并到处理器中,而不是使用单独的北桥芯片。在一些实施方案中,系统900可包括附接到外围总线918的加速器卡922。加速器可以包括现场可编程门阵列 (FPGA) 或用于加速某些处理的其他硬件。例如,加速器可用于自适应数据重构或评价扩展集处理中使用的代数表达式。

[0181] 软件和数据存储在外部存储924中,并且可以加载到RAM 910和/或缓冲存储器904中以供处理器使用。系统900包括用于管理系统资源的操作系统;操作系统的非限制性示例包括:Linux、Windows™、MACOS™、BlackBerry OS™、iOSTM和其他功能上等效的操作系统,以及在该操作系统之上运行以管理根据本发明示例性实施方案的数据存储和优化的应用软件。在该实例中,系统900还包括连接到外围总线用于向外部存储提供网络接口的网络接口卡 (NIC) 920和921,该外部存储如网络附加存储 (NAS) 和可用于分布式平行处理的其他计算机系统。

[0182] 图10是示出具有多个计算机系统1002a和1002b、多个蜂窝电话和个人数据助理1002c、以及网络附加存储 (NAS) 1004a和1004b的网络1000的图像。在示例性实施方案中,系统1002a、1002b和1002c可以管理数据存储并优化对网络附加存储 (NAS) 1004a和1004b中存储的数据的数据访问。可以对数据使用数学模型,并且可以使用跨计算机系统1002a和1002b以及蜂窝电话和个人数据助理系统1002c的分布式平行处理来评价该数学模型。计算机系统1002a和1002b以及蜂窝电话和个人数据助理系统1002c还可以提供用于网络附加存

储 (NAS) 1004a和1004b中存储的数据的自适应数据重构的平行处理。图10仅图示了实例,并且可以结合本发明的各个实施方案使用各种其他计算机架构和系统。例如,可以使用刀片服务器提供平行处理。处理器刀片可以通过背板连接以提供平行处理。存储还可以连接到背板或通过单独的网络接口作为网络附加存储 (NAS)。在一些示例性实施方案中,处理器可以维持单独的存储空间并通过网络接口、背板或其他连接器传输数据,以供其他处理器平行处理。在其他实施方案中,一些或所有处理器可以使用共享的虚拟地址存储空间。

[0183] 图11是根据示例性实施方案的使用共享虚拟地址存储空间的多处理器计算机系统的框图。该系统包括可以访问共享存储器子系统1104的多个处理器1102a-1102f。该系统在存储器子系统1104中包含多个可编程硬件存储器算法处理器 (MAP) 1106a-1106f。每个MAP 1106a-1106f可以包含存储器1108a-1108f和一个或多个现场可编程门阵列 (FPGA) 1110a-1110f。MAP提供可配置的功能单元,并且可以向FPGA 1110a-1110f提供特定的算法或算法部分,以便与相应处理器紧密协调地进行处理。例如,在示例性实施方案中,MAP可用于评价关于数据模型的代数表达式并且执行自适应数据重构。在该实例中,出于这些目的,所有处理器可以全局访问每个MAP。在一种配置中,每个MAP可以使用直接存储器访问 (DMA) 来访问相关联的存储器1108a-1108f,从而允许其独立于相应的微处理器1102a-1102f并且与相应的微处理器1102a-1102f异步地执行任务。在该配置中,MAP可以将结果直接提供给另一个MAP,以进行流水线操作和算法的平行执行。

[0184] 上述计算机架构和系统仅是实例,并且可以结合示例性实施方案使用各种其他计算机、蜂窝电话和个人数据助理架构以及系统,包括使用通用处理器、协处理器、FPGA和其他可编程逻辑器件的任意组合的系统、片上系统 (SOC)、专用集成电路 (ASIC) 以及其他处理和逻辑元件。在一些实施方案中,计算机系统的全部或部分可以在软件或硬件中实现。可以结合示例性实施方案使用任何种类的数据存储介质,包括随机存取存储器、硬盘驱动器、闪存存储器、磁带驱动器、磁盘阵列、网络附加存储 (NAS) 以及其他本地或分布式数据存储装置和系统。

[0185] 在示例性实施方案中,可以使用在任何上述或其他计算机架构和系统上执行的软件模块来实现计算机系统。在其他实施方案中,系统的功能可以部分或完全地在固件可编程逻辑装置中实现,该固件可编程逻辑装置如图11中提到的现场可编程门阵列 (FPGA)、片上系统 (SOC)、专用集成电路 (ASIC) 或其他处理和逻辑元件。例如,可以通过使用硬件加速器卡如图9中所示的加速器卡922利用硬件加速来实现设置处理器和优化器。

[0186] 提出以下实施例是为了更清楚地说明本文公开的实施方案的原理和实践,但不应解释为限制任何要求保护的实施方案的范围。除非另有说明,否则所有份数和百分比均以重量计。

#### 实施例

[0187] 给出以下实施例是出于说明本发明的各种实施方案的目的,而并不意味着以任何方式限制本发明。本实施例以及目前代表优选实施方案的本文所述的方法是示例性的,并不旨在作为对本发明范围的限制。本领域技术人员将想到其中的变化和包含在由权利要求范围限定的本发明精神内的其他用途。

[0188] 实施例1:基底表面的官能化

[0189] 使基底官能化以支持寡核酸文库的附接和合成。首先使用包含90% $H_2SO_4$ 和10%

H<sub>2</sub>O<sub>2</sub>的食人鱼溶液(piranha solution)将基底表面湿法清洁20分钟。将基底在几个烧杯中用DI水漂洗,在DI水鹅颈式水龙头下保持5min,并用N<sub>2</sub>干燥。随后将基底在NH<sub>4</sub>OH(1:100; 3mL:300mL)中浸泡5min,使用手动枪用DI水漂洗,在连续三个烧杯中用DI水浸泡各自1min,然后再次使用手动枪用DI水漂洗。随后通过将基底表面暴露于O<sub>2</sub>来对基底进行等离子体清洁。使用SAMCO PC-300仪器在下游模式下以250瓦进行O<sub>2</sub>等离子体蚀刻1min。

[0190] 用包含N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺的溶液使用具有以下参数的YES-1224P气相沉积烘箱系统使清洁的基底表面主动官能化:0.5至1torr,60min,70°C,135°C的蒸发器。使用Brewer Science 200X旋涂机对基底表面进行抗蚀剂涂覆。将SPR™ 3612光刻胶以2500rpm在基底上旋涂40sec。将基底在Brewer热板上在90°C下预烘烤30min。使用Karl Suss MA6掩模对准器仪器使基底经历光刻。将基底暴露2.2sec并在MSF 26A中显影1min。用手动枪漂洗冲洗剩余的显影剂,并将基底在水中浸泡5min。将基底在烘箱中在100°C下烘烤30min,然后使用Nikon L200目视检查光刻缺陷。使用SAMCO PC-300仪器以250瓦进行O<sub>2</sub>等离子体蚀刻1min,从而使用清洁过程去除残留的抗蚀剂。

[0191] 用全氟辛基三氯硅烷与10μL轻质矿物油混合的100μL溶液使基底表面被动官能化。将基底置于室中,泵送10min,然后将阀关闭至泵并静置10min。将室通风排气。通过在70°C下在500mL NMP中进行两次5min的浸泡,并以最大功率(在Crest系统上为9)进行超声处理,来剥离基底的抗蚀剂。然后在室温下将基底在500mL异丙醇中浸泡5min,并以最大功率进行超声波处理。将基底浸入300mL无水乙醇中并用N<sub>2</sub>吹干。激活功能化表面以用作寡核酸合成的支持物。

[0192] 实施例2:在寡核酸合成装置上合成50-聚体序列

[0193] 将二维寡核酸合成装置组装到流动池中,其与流动池(Applied Biosystems“ABI394DNA合成仪”)连接。使用通过N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺(Gelest)均匀官能化的二维寡核酸合成装置,使用本文所述的寡核酸合成方法合成50bp的示例性寡核酸(“50-聚体寡核酸”)。

[0194] 所述50-聚体的序列如SEQ ID NO.:1所述。

5' AGACAATCAACCATTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTTTTTT3' (SEQ ID NO.:1),其中#表示胸苷-琥珀酰己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244),它是能够在脱保护过程中使寡核酸从表面上释放的可切割的连接体。

[0195] 根据表5中的方案,使用标准DNA合成化学法(偶联、加帽、氧化和去封闭)和ABI合成仪完成合成。

表5:

表 5		
一般 DNA 合成过程名称	过程步骤	时间 (sec)
洗涤 (乙腈洗涤流)	乙腈系统冲洗	4
	乙腈至流动池	23
	N2 系统冲洗	4
	乙腈系统冲洗	4
DNA 碱基添加 (亚磷酸胺+激活剂流)	激活剂歧管冲洗	2
	激活剂至流动池	6
	激活剂+亚磷酸胺至流动池	6
	激活剂至流动池	0.5
	激活剂+亚磷酸胺至流动池	5
	激活剂至流动池	0.5
	激活剂+亚磷酸胺至流动池	5

表 5		
一般 DNA 合成过程名称	过程步骤	时间 (sec)
	池	
	激活剂至流动池	0.5
	激活剂+亚磷酰胺至流动池	5
	温育 25 sec	25
洗涤 (乙腈洗涤流)	乙腈系统冲洗	4
	乙腈至流动池	15
	N2 系统冲洗	4
	乙腈系统冲洗	4
DNA 碱基添加 (亚磷酰胺+激活剂流)	激活剂歧管冲洗	2
	激活剂至流动池	5
	激活剂+亚磷酰胺至流动池	18
	温育 25 sec	25
洗涤 (乙腈洗涤流)	乙腈系统冲洗	4
	乙腈至流动池	15
	N2 系统冲洗	4
	乙腈系统冲洗	4
加帽 (帽 A+B, 1:1, 流)	帽 A+B 至流动池	15
洗涤 (乙腈洗涤流)	乙腈系统冲洗	4
	乙腈至流动池	15
	乙腈系统冲洗	4
氧化 (氧化剂流)	氧化剂至流动池	18
洗涤 (乙腈洗涤流)	乙腈系统冲洗	4
	N2 系统冲洗	4
	乙腈系统冲洗	4
	乙腈至流动池	15

表 5		
一般 DNA 合成过程名称	过程步骤	时间 (sec)
	乙腈系统冲洗	4
	乙腈至流动池	15
	N2 系统冲洗	4
	乙腈系统冲洗	4
	乙腈至流动池	23
	N2 系统冲洗	4
	乙腈系统冲洗	4
去封闭 (去封闭流)	去封闭至流动池	36
洗涤 (乙腈洗涤流)	乙腈系统冲洗	4
	N2 系统冲洗	4
	乙腈系统冲洗	4
	乙腈至流动池	18
	N2 系统冲洗	4.13
	乙腈系统冲洗	4.13
	乙腈至流动池	15

[0196] 亚磷酰胺/活化剂组合以类似于本体试剂通过流动池的递送的方式进行递送。当在整个时间中环境保持被试剂“润湿”时,不进行干燥步骤。

[0197] 从ABI 394合成仪中移除限流器以实现更快的流动。在没有限流器的情况下,amidite (0.1M的ACN溶液)、激活剂 (0.25M苯甲酰基硫代四唑 (“BTT”;来自GlenResearch的30-3070-xx)的ACN溶液)和0x (0.02M I<sub>2</sub>的20%吡啶、10%水和70%THF溶液)的流速约为~100uL/sec,乙腈 (“ACN”)和加帽试剂 (帽A和帽B的1:1混合物,其中帽A是乙酸酐的THF/吡啶溶液,并且帽B为16%1-甲基咪唑的THF溶液)的流速约为~200uL/sec,Deblock (3%二氯乙酸的甲苯溶液)的流速约为~300uL/sec (相比之下,在具有限流器的情况下,所有试剂的流速为~50uL/sec)。观察完全推出氧化剂的时间,相应地调整化学流动时间的时间安排并在不同化学品之间引入额外的ACN洗涤。在寡核酸合成后,将芯片在75psi下在气态氨中脱保护过夜。将五滴水施加到表面以回收寡核酸。然后在BioAnalyzer小RNA芯片上分析回收的寡核酸 (数据未示出)。

[0198] 实施例3:在寡核酸合成装置上合成100-聚体序列

[0199] 使用与实施例2中描述的用于合成50-聚体序列相同的方法在两个不同的硅芯片上合成100-聚体寡核酸 (“100-聚体寡核酸”;5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCATTGCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTT3',其中#表示胸苷-琥珀酰己酰胺CED亚磷酰胺 (来自ChemGenes的CLP-2244);

SEQ ID NO.:2),用N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺将第一个芯片均匀官能化,并且用11-乙酰氧基十一烷基三乙氧基硅烷和正癸基三乙氧基硅烷的5/95混合物将第二个芯片官能化,并在BioAnalyzer仪器上分析从表面提取的寡核酸(数据未示出)。

[0200] 使用以下热循环程序,使用50uL PCR混合物(25uL NEB Q5 mastermix,2.5uL 10uM正向引物,2.5uL 10uM反向引物,1uL从表面提取的寡核酸,以及至多50uL的水)中的正向(5' ATGCGGGTTCTCATCATC3';SEQ ID NO.:3)和反向(5' CGGGATCCTTATCGTCATCG3';SEQ ID NO.:4)引物进一步对来自两个芯片的所有10个样品进行PCR扩增:

98°C,30sec

98°C,10sec;63°C,10sec;72°C,10sec;重复12个循环72°C,2min

[0201] PCR产物还在BioAnalyzer上运行(数据未示出),在100-聚体位置处显示尖锐峰。然后,对PCR扩增的样品进行克隆,并进行Sanger测序。表6总结了从来自芯片1的斑点1-5采集的样品和从来自芯片2的斑点6-10采集的样品的Sanger测序结果。

表6:

斑点	错误率	循环效率
1	1/763bp	99.87%
2	1/824bp	99.88%
3	1/780bp	99.87%
4	1/429bp	99.77%
5	1/1525bp	99.93%
6	1/1615bp	99.94%
7	1/531bp	99.81%
8	1/1769bp	99.94%
9	1/854bp	99.88%
10	1/1451bp	99.93%

[0202] 因此,合成的寡核酸的高质量和均匀性在具有不同表面化学性质的两个芯片上得以重复。总体而言,被测序的89%是没有错误的正确序列,对应于262个100-聚体中的233个。

[0203] 最后,表7总结了从来自斑点1-10的寡核酸样品获得的序列的关键错误特征。

表7:

样品 ID/斑点号	OSA_0046/	OSA_0047/	OSA_0048/	OSA_0049/	OSA_0050/	OSA_0051/	OSA_0052/	OSA_0053/	OSA_0054/	OSA_0055/1
	1	2	3	4	5	6	7	8	9	0
全序列	32	32	32	32	32	32	32	32	32	32
测序质量	25/28	27/27	26/30	21/23	25/26	29/30	27/31	29/31	28/29	25/28
寡核酸质量	23/25	25/27	22/26	18/21	24/25	25/29	22/27	28/29	26/28	20/25
ROI 匹配计数	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI 突变	2	2	1	3	1	0	2	1	2	1
ROI 多碱基缺失	0	0	0	0	0	0	0	0	0	0
ROI 小插入	1	0	0	0	0	0	0	0	0	0
ROI 单碱基缺失	0	0	0	0	0	0	0	0	0	0

大缺失计数	0	0	1	0	0	1	1	0	0	0
突变: G>A	2	2	1	2	1	0	2	1	2	1
突变: T>C	0	0	0	1	0	0	0	0	0	0
ROI 错误计数	3	2	2	3	1	1	3	1	2	1
ROI 错误率	错误率: ~1/834	错误率: ~1/1350	错误率: ~1/1282	错误率: ~1/708	错误率: ~1/2500	错误率: ~1/2667	错误率: ~1/876	错误率: ~1/2900	错误率: ~1/1400	错误率: ~1/2349
POI-引物错误率	MP 错误率: 率: ~1/763	MP 错误率: ~1/824	MP 错误率: ~1/780	MP 错误率: ~1/429	MP 错误率: ~1/1525	MP 错误率: ~1/1615	MP 错误率: ~1/531	MP 错误率: ~1/1769	MP 错误率: ~1/854	MP 错误率: ~1/1451

#### [0204] 实施例4:sgRNA设计

[0205] 在5' 端具有可变区的嵌合sgRNA序列被设计用于由Cas9蛋白进行直接序列特异性切割。参见图4A。sgRNA序列具有用于特异性DNA结合的20个碱基的碱基配对区域,其包含12个碱基的种子区域。碱基配对区域的5' 端被设计为转录起始位点。碱基配对区域的3' 近端是用于Cas9结合的dCas9手柄区域,其长度为42个碱基。dCas9手柄区域的3' 近端是酿脓链球菌终止子区域,其长度为40个碱基。dCas9手柄区域和终止子区域各自被设计为包含将产生发夹结构的序列。

[0206] 还设计了sgRNA以靶向模板(T) 或非模板(NT) DNA链,图5A-图5B。被设计用于靶向模板DNA链的sgRNA包含与转录序列具有相同序列同一性的sgRNA的碱基配对区域。被设计用于靶向非模板DNA链的sgRNA包含是转录序列的反向互补的sgRNA的碱基配对区域。

[0207] 在额外的布置中,紧邻可变碱基配对区域的上游设计有T7启动子。参见图6A-图6B。添加T7启动子区域以使用T7聚合酶实现sgRNA的体外产生。

#### [0208] 实施例5:编码sgRNA的DNA的合成-设计和聚合酶分析

[0209] 将DNA寡核酸设计为在连接时编码sgRNA序列的片段。图12。设计sgRNA以在紧邻可

变序列区域1233的上游包含T7启动子。在从头合成DNA寡核酸后,进行扩增反应以连接并延伸重叠片段。

[0210] 在1201处转录DNA寡核酸导致用T7聚合酶从DNA模板体外产生sgRNA。

[0211] 设计如表8所示的设计1 1220、设计2 1222、设计3 1224和设计4 1226的序列。设计1、设计2、设计3和设计4中每个的序列包含T7启动子、可变序列部分和恒定序列区(手柄和终止子)(表8)。具体地,如图12所示的恒定序列区包含含有碱基配对区1211、1213、1215、1217、1223和1225的Cas9手柄发夹,以及含有碱基配对区1219和1221的终止子发夹。

表8.

SEQ ID NO	名称	序列
12	设计 1	TAATACGACTCACTATAGGGGATGCGCGCAGTT GTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTT
13	设计 1 的 T7 启动子	TAATACGACTCACTATA
14	设计 1 的可变序列	GGATGCGCGCAGTTGTCC
15	设计 1 的手柄和终止子	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCTTTT
44	设计 2	GAAATTAATACGACTCACTATAGGGGATGCGCG CAGTTGTCCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAA GTGGCACCGAGTCGGTGCTTTT
13	设计 2 的 T7 启动子	TAATACGACTCACTATA
14	设计 2 的可变序列	GGATGCGCGCAGTTGTCC
15	设计 2 的手柄和终止子	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCTTTT

16	设计 3	GAGCTAATACGACTCACTATAGGGGATGCGCGC AGTTGTCCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAA GTGGCACCGAGTCGGTGCTTTT
13	设计 3 的 T7 启动子	TAATACGACTCACTATA
14	设计 3 的可变序列	GGATGCGCGCAGTTGTCC
15	设计 3 的手柄和终止子	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCTTTT
17	设计 4	CGAGCTAATACGACTCACTATAGGGGATGCGCG CAGTTGTCCGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAA GTGGCACCGAGTCGGTGCTTTT
13	设计 4 的 T7 启动子	TAATACGACTCACTATA
14	设计 4 的可变序列	GGATGCGCGCAGTTGTCC
15	设计 4 的手柄和终止子	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCTTTT

[0212] 对于体外分析,注意到T7RNA聚合酶启动子区域应当是双链的,以供通过T7RNA聚合酶识别。使用反义寡核酸进行杂交:5'-TAATACGACTCACTATAGG-3' (SEQ ID NO:18)。此外,表9提供了用于分析4组不同的模板和扩增寡核酸的引物的列表。参见图12。

表9.

SEQ ID NO	名称 (#碱基-解链温度)	序列
19	SgR1-R1 (80bp)	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTG ATAACGGACTAGCCTTATTTAACTTGCTATTCT

		AGCTCTAAAAC
20	SgR1-F1 (58bp - 51°C/54°C)	TAATACGACTCACTATAGGGGATGCGCGCAGTTG TCCGTTTTAGAGCTAGAAATAGCA
21	SgR1-F2 (65bp - 56°C/56°C)	GAAATTAATACGACTCACTATAGGGGATGCGCGC AGTTGTCCGTTTTAGAGCTAGAAATAGCAAG
22	SgR1-F3 (66bp - 59°C/58°C)	GAGCTAATACGACTCACTATAGGGGATGCGCGCA GTTGTCCGTTTTAGAGCTAGAAATAGCAAGTT
23	SgR1-F4 (78bp - 62°C/62°C)	GAGCTAATACGACTCACTATAGGGGATGCGCGCA GTTGTCCGTTTTAGAGCTAGAAATAGCAAGTTAA ATAAAGG
24	SgR1-AR1 (14bp - 52°C)	AAAAGCACCGACTC
25	SgR1-AR2 (15bp - 56°C)	AAAAGCACCGACTCG
26	SgR1-AR3 (16bp - 60°C)	AAAAGCACCGACTCGG
27	SgR1-AR4 (17bp - 62°C)	AAAAGCACCGACTCGGT
18	sgR1-AF1 (19bp - 51°C)	TAATACGACTCACTATAGG
28	SgR1-AF2 (24bp - 56°C)	GAAATTAATACGACTCACTATAGG
29	SgR1-AF3 (25bp - 59°C)	GAGCTAATACGACTCACTATAGG
30	SgR1-AF4 (25bp - 62°C)	GCGAGCTAATACGACTCACTATAGG

[0213] 在多种条件下分析4组不同的模板和扩增寡核酸以优化全长模板的纯度和产率。用100fMol的模板寡核酸 (SgR1-R1与SgR1-F1、SgR1-F2、SgR1-F3、SgR1-F4) 和聚合酶-1PCR浓度为600nMol的各组扩增引物进行10ul PCR反应。使用Eppendorf Mastercycler上的梯度,使用两种高保真DNA聚合酶 (聚合酶1和3) 和标准DNA聚合酶 (聚合酶2) 在25个循环的PCR中评价3种退火温度 (50°C、55°C、60°C)。表10提供了反应条件的总结,且表11提供了扩增方案。

表10.

试剂 (10 ul 反应)		聚合酶 1		聚合酶 2		聚合酶 3
		1x	15x		1x	15x

聚合酶		0.1	1.5		0.1	1.5
缓冲液		2	30		1	15
dNTP		0.2	3		0.2	3
扩增引物 (10 uM)		0.6			0.6	
模板寡核酸 (100 nM)		1			1	
H <sub>2</sub> O		6.1	91.5		7.1	106.5
总计		10	150		10	150

表11.

聚合酶 1、2			聚合酶 3		
98	30sec		95	3min	
98	10sec	25x	98	10sec	25x
50 / 55 / 60	15sec		50 / 55 / 60	15sec	
72	10sec		72	10sec	
72	30sec		72	30sec	
4	保持		4	保持	

[0214] 在BioAnalyzer上运行聚合酶-1PCR反应的结果(数据未示出)以估计产率,并总结于表12中。DNA产率以ng/ul表示(表12)。寡核酸设计3和4各自导致比寡核酸设计1和2更高的DNA产率。较高的退火温度也导致产率提高,60℃时具有较高产率。

表12.

聚合酶1	50℃	55℃	60℃
寡核酸设计1	0	0	0
寡核酸设计2	2.9	4.5	3.7
寡核酸设计3	6.8	9.2	10
寡核酸设计4	9.9	13	15.3

产率以ng/ul列出。

[0215] 在BioAnalyzer上运行聚合酶2PCR反应的结果(数据未示出)以估计产率,并总结于表13中。DNA产率以ng/ul表示(表13)。同样,寡核酸设计3和4各自导致比寡核酸设计1和2更高的DNA产率。较高的退火温度也导致产率提高,60℃时具有较高产率。

表13.

聚合酶2	50℃	55℃	60℃
寡核酸设计1	0	0	0
寡核酸设计2	7.6	5.9	6.9

寡核酸设计3	6.1	8.5	10.5
寡核酸设计4	7.4	11.1	19.4

产率以ng/u1列出。

[0216] 在BioAnalyzer上运行聚合酶3PCR反应的结果(数据未示出)以估计产率,并总结于表14中。DNA产率以ng/u1表示(表14)。寡核酸设计3和4各自导致比寡核酸设计1和2更高的DNA产率。较高的退火温度也导致产率提高,60℃时具有较高产率。

表14.

聚合酶3	50℃	55℃	60℃
寡核酸设计1	10	13	12.1
寡核酸设计2	12.4	14.3	15.9
寡核酸设计3	13.2	26.1	28.8
寡核酸设计4	16.1	13.2	18.5

产率以ng/u1列出。

[0217] 总之,寡核酸设计3和4导致所有三种聚合酶的DNA产率提高。此外,60℃的较高退火温度导致DNA产率提高。

[0218] 实施例6:CRISPR sgRNA合成-温度分析

[0219] 使用实施例5的寡核酸引物,在运行如实施例5所述的PCR反应后分析提高退火温度条件的影响。在BioAnalyzer上运行扩增产物(数据未显示)以估计产率,并总结于表15中。DNA产率以ng/u1表示(表15)。总之,聚合酶3提供了提高的DNA产率,并且60℃的退火温度导致DNA产率提高。

表15.

	寡核酸设计 3				寡核酸设计 4			
	聚合酶 3		聚合酶 1		聚合酶 3		聚合酶 1	
	60℃	65℃	60℃	65℃	60℃	65℃	60℃	65℃
25 个循环	27.3	15.2	11.6	3.9	28.4	29.5	13.8	10.2

产率以ng/u1列出。

[0220] 实施例7:sgRNA生成-无结构RNA

[0221] 设计两种装配寡核酸用于生成具有T7启动子序列和终止子但没有含有tracrRNA发夹的序列的修饰的sgRNA模板(120bp)。参见表16。

表16.

SEQ ID NO	名称	序列
31	sgR2-模板	CGAGCTAATACGACTCACTATAGGGGCACAACGTGG AGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGC AAAACACTCCAATCGGTGATGGTCCTGTTGCACCGAG TCGGTGCTTTT
32	sgR2-F	CGAGCTAATACGACTCACTATAGGGGCACAACGTGG AGGATGGCAGCGTGCAGCTGGCTGATCACTACCAG
33	sgR2-R	AAAGCACCGACTCGGTGCAACAGGACCATCACCGAT TGGAGTGTTTTGCTGGTAGTGATCAGCCAGCTG

[0222] 用与扩增实施例5中的sgRNA所使用的相同的引物寡核酸扩增组装寡核酸。表17中总结了使用的反应条件。

表17.

成分	1x		PCR 条件	时间	循环
聚合酶 1	0.1		95°C	3 min	
缓冲液	2		98°C	10 sec	25 x
dNTP	0.2		65°C	15 sec	
扩增引物 (10 uM)	0.6		72°C	10 sec	
模板寡核酸 (100 nm)	1		72°C	30 sec	
H2O	6.1		4°C	保持	
总计	10				

[0223] 使用T7RNA聚合酶的转录预期产生80bp的没有二级结构的RNA产物。用体外转录试剂盒 (NEB HiScribe) 进行扩增产物的转录。在BioAnalyzer上分析反应混合物。参见图13A-图13B。与具有含tracrRNA发夹的序列的sgRNA (图13A) 相比, 经修饰的sgRNA产物更简洁 (clean), 具有无结构设计 (图13B)。

[0224] 实施例8: sgRNA引导的Cas9切割

[0225] 设计具有T7启动子区三种sgRNA序列, 并且每种序列具有针对720bp GFP编码序列的区域的不同的识别序列。每种sgRNA序列由两种寡核酸的PCR组装而成。表18中提供了sgRNA骨架和引物。

表18.

SEQ ID NO	名称	序列
34	sgRNA骨架	CGAGCTAATACGACTCACTATAggNNNNNNNNNNNN NNNNNGTTTTAGAGCTATGCTGAAAAGCATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTGAAAAAG TGGCACCGAGTCGGTGCTTTT
35	GFP	ATGcgtAAAggcGAAgagCTGttcACTggtGTCgtcCCTattCTG gtgGAActgGATggtGATgtcAACggtCATaagTTTtccGTGcgtG  GCgagGGTgaaGGTgacGCAactAATggtAAAActgACGctgAAG ttcATCtgtACTactGGTaaaCTGccgGTAcctTGGccgACTctgGT AacgACGctgACTtatGGTggtCAGtgcTTTgctCGTtatCCGgacC ATatgAAGcagCATgacTTCttcAAGtccGCCatgCCGgaaGGCta tGTGcagGAacgACGattTCCtttAAGgatGACggcACGtacAA AacgCGTgcgGAAgtgAAAtttGAAggcGATaccCTGgtaAACcg cATTgagCTGaaaGGCattGACtttAAAgaagACggcAATatcCT GggcCATaagCTGgaaTACaatTTTaaacAGCcacAATggtTACatc ACCgccGATaaaCAAaaaAATggcATTaaaGCGaatTTTaaaAT TcgcCACaacGTGgagGATggcAGCgtgCAGctgGCTgatCACta cCAGcaaAACactCCAateGGTgatGGTectGTTctgCTGccaGA CaatCACtatCTGagcACGcaaAGCgttCTGtctAAAgatCCGaac GAGaaaCGCgatCATatgTTctgCTGgagTTCgtaACCgcaGCG ggcATCacgCATggtATGgatGAActgTACaaaTGAtaa
36	sgR35-F	CGAGCTAATACGACTCACTATAGGAAcgcACGattTCCttt AGTTTTAGAGCTATGCTGAAAAGCATAGC
37	sgR36-F	CGAGCTAATACGACTCACTATAGGCattGACtttAAAgaagG AGTTTTAGAGCTATGCTGAAAAGCATAGC
38	sgR37-F	CGAGCTAATACGACTCACTATAGGagGATggcAGCgtgC AGcGTTTTAGAGCTATGCTGAAAAGCATAGC
39	sgR3-R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATA ACGGACTAGCCTTATTTAACTTGCTATGCTTTTCAGC ATAGCTCTAAAAC
30	SgR1-AF4	GCGAGCTAATACGACTCACTATAGG
27	SgR1-AR4	AAAAGCACCGACTCGGT

[0226] 在表19中总结的反应条件下扩增组装寡核酸。

表19.

成分	1x		PCR 条件	时间	循环
聚合酶 3	0.1		95°C	3 min	
缓冲	2		98°C	10 sec	25 x
dNTP	0.2		95°C	15 sec	
扩增引物 (10 um)	0.6		60°C	15 sec	
模板寡核酸 (100 nm)	1		72°C	10 sec	
H2O	6.1		72°C	30 sec	
总计	10		4°C	保持	

[0227] 在BioAnalyzer上分析来自每个sgRNA组装反应的样品(图14A-图14C)。使用T7RNA聚合酶PCR扩增产物进行转录反应。在BioAnalyzer上分析来自每个反应的样品(图14D-图14F)。

[0228] 使用GFP扩增产物、Cas9和转录的sgRNA制备Cas9消化物。与对照的单峰相比,所有三种消化物观察到2个峰(图14G-图14J)。表20中列出了使用3种合成的sgRNA进行Cas9切割的预期和所得片段。

表20.

sgRNA	预测片段1	所得片段1	预测片段2	所得片段2
sgR35	321	324	439	451
sgR36	342	350	418	430
sgR37	208	137	552	560

[0229] 纯化Cas9消化样品并再次在BioAnalyzer上分析(数据未示出)。表21中总结了纯化样品的结果。

表21.

sgRNA	预测片段 1	所得片段 1	预测片段 2	所得片段 2
sgR35	321	323	439	451
sgR36	342	353	418	427
sgR37	208	220	552	560

[0230] 实施例9:29,040种独特寡核酸的平行组装

[0231] 如图15所示,制造在平面硅板上包含各自含有121个座位的256个簇1505的结构。簇的扩展视图在1510中示出,具有121个座位。256个簇中的240个的座位为具有不同序列的

寡核酸的合成提供了附接和支持。使用实施例3的一般方法,通过亚磷酰胺化学法进行寡核酸合成。256个簇中的16个的座位是对照簇。图16A示出了合成的29,040种独特寡核酸(240个非对照簇×每簇121个寡核酸群体)的整体分布。NGS测序证实了被选择用于合成的设计的寡核酸的100%表示。测定每个簇的分布,如图16B所示。图17示出了在4个代表性簇中合成的独特寡核酸的分布。在整体水平上,被设计用于合成的所有寡核酸均存在,并且99%的寡核酸具有平均值的2倍以内的丰度,指示了较高的合成均匀性。同样的观察结果在每个簇的水平上是一致的。

[0232] 使用Illumina MiSeq基因测序仪测定每种寡核酸的错误率。29,040种独特寡核酸的错误率分布如图18A所示,平均值为约1/500个碱基,一些错误率低至1/800个碱基。测定每个簇的分布,如图18B所示。图19示出了四个代表性簇中独特寡核酸的错误率分布。29,040种独特寡核酸的文库在不到20小时内合成。GC百分比与所有29,040种独特寡核酸的寡核酸表示的分析显示,无论GC含量(每种寡核酸为约20%至85%的GC)如何,合成是均匀的,图20。

[0233] 实施例10:编码sgRNA的从头合成DNA文库的PCR扩增分析

[0234] 设计并在实施例9中所述的具有相似布置的结构上合成具有20-80%GC的不同GC含量的长度为100个碱基的随机序列的9,996个寡核酸。为了确定PCR扩增对GC表示的影响,用高保真DNA聚合酶(DNA聚合酶1)将寡核酸群体扩增6个或20个循环。或者,使用两种其他高保真PCR酶将寡核酸群体扩增6、8、10或15个循环,以确定聚合酶选择是否对扩增后的总体序列表示有影响。在PCR扩增后,准备样品用于下一代测序并在Illumina MiSeq平台上进行测序。生成150个碱基对SE读取,读取覆盖率约为100X。分析原始FASTQ文件。图21中描绘了用任一种聚合酶进行6、10或15个循环的寡核酸表示。针对各种条件评估通过测序读取中的表示频率所测定的寡核酸均匀性,并总结于表22中。

表22.

	循环	1.5x 内%	2x 内%
聚合酶 1	6	72.1%	92.6%
	8	76.1%	90.3%
	10	70.9%	86.6%
	15	64.1%	82.7%
聚合酶 2	6	91.9%	98.9%
	8	89.9%	98.1%
	10	90.1%	98.4%
	15	89.2%	97.9%

[0235] 如图22所示量化每个扩增的寡核酸群体的丢失(dropout)数,即扩增循环相对于低于平均值的10%的阈值的群体分数。聚合酶1丢失迅速增长,而聚合酶2丢失保持相对稳定。

[0236] 评估过度扩增对GC分布的影响,图23。通常,GC含量为30%至70%的寡核酸遵循趋势线, $Y=X$ ,并且频率随着循环增多而增加。GC含量大于70%的寡核酸通常在20个循环后略微更频繁,而GC含量低于30%的寡核酸通常在6个循环后略微更频繁。

[0237] 实施例11:人表观遗传的CRISPR筛选

[0238] 使用CRISPR-Cas9进行sgRNA筛选,以将突变引入编码功能结构域的外显子中。在实施例9中所述的具有相似布置的结构上在实施例1中所述的硅芯片上使用与实施例3中所述类似的方法从头合成约10,000个DNA寡核酸。总体而言,寡核酸的总错误率为约1:500或更低。每个寡核酸为至多200个碱基的长度,并且每个寡核酸种类生成至少1fmole。将寡核酸进行PCR扩增、克隆到载体中,并电穿孔到培养的细胞中以进行sgRNA转录。从细胞中分离核酸并使用下一代测序进行测序。

[0239] 测序结果显示高度准确和均匀的文库合成,其具有最小偏差和高保真度的sgRNA产生。与竞争者集合相比,利用最小测序的每个指导序列的读数更多,且用于下游筛选的正确序列的sgRNA回收率提高30%。参见表23。汇总测序结果显示,与基于阵列的竞争者集合的6个log值相比,每个指导序列的读取更多且读取分布更紧密(4个log值)。参见图24A-图24B。与市售的基于阵列的集合(图24B)相比,克隆的测序验证显示100%sgRNA回收率和更高的序列精确性(图24A)。在测序的克隆中,回收的显著更多的克隆具有正确的sgRNA序列。参见表23。寡核酸群体中100%的预定序列有表示。sgRNA克隆的基于NGS验证示出,与竞争者的群体相比,每个克隆的sgRNA回收率为100%,合成精确性高13%(数据未示出)。

表23.

	合成的寡核酸群体	市售的寡核酸群体
sgRNA 寡核酸回收	100%	95.5%
正确序列率 (MiSeq)	约 87%	约 74%
正确序列 (Sanger 10 克隆)	约 100%	约 70%
克隆寡核酸群体中每个 sgRNA 的平均读取 (100x 标准化)	约 256	约 1024

[0240] 实施例12:全基因组sgRNA文库

[0241] DNA文库被设计为包含编码sgRNA的DNA,用于生成101,000种不同的寡核酸(20200个基因靶标,每个靶标5个sgRNA)的克隆。使用与实施例3中所述类似的方法在实施例9中所述的具有相似布置的结构上在实施例1中所述的硅芯片上从头合成101,000个寡核酸。将合成的寡核酸进行PCR扩增、消化并克隆到慢病毒载体中,并转化到细胞中。从细胞中分离核酸并使用下一代测序进行测序。或者,对合成的寡核酸进行PCR扩增以形成基于扩增子的文库并进行测序。

[0242] 下一代测序读取相对于回收的sgRNA数的绘图示出,随着寡核酸集合大小的增加,寡核酸群体在整个文库上保持更均匀更紧密的读取分布,与市售的基于阵列的参考寡核酸群体相比具有最小的尾部。图25。

[0243] 实施例13:具有改善的靶向和活性的sgRNA文库的设计

[0244] 设计sgRNA文库并使用与实施例3中所述类似的方法在实施例9中所述的具有相似布置的结构上在实施例1中所述的硅芯片上从头合成sgRNA文库。将合成的寡核酸进行PCR扩增、消化并克隆到载体中,并转移到细胞中以进行包括筛选和分析在内的下游应用。

[0245] 比较不同的sgRNA设计参数包括以NAG PAM、NGG PAM、高活性、低脱靶和过滤为特征的文库。与其他市售的gRNA系统相比,通过本文所述方法设计的sgRNA文库提供了导致基因表达的至少2倍消耗的较高百分比的sgRNA,即约16%的sgRNA。图26A。与其他市售的gRNA系统相比,sgRNA文库还提供了导致在基因表达方面零或负消耗的较低百分比的sgRNA,即约17%。图26B。

[0246] 还评估了sgRNA介导的消耗所必需的基因表达水平,其中以下基因被sgRNA靶向:PCNA、PSMA7、RPP21和SF3B3。分析展现出至少2倍消耗的sgRNA的数目,与比较物1、比较物2和比较物3相比,sgRNA文库具有更高百分比的消耗必需基因的sgRNA。参见表24。

表24.

基因名称	比较物1	比较物2	比较物3	CRISPR文库
PCNA	1/5	2/6	5/9	5/5
PSMA7	1/5	0/6	2/9	3/5
RPP21	3/5	1/6	4/9	2/5
SF3B3	0/5	0/6	4/9	3/5
平均(%)	25%	12.5%	42%	65%

[0247] 实施例14:针对于MS2的CRISPR文库

[0248] 包含编码sgRNA的不相同DNA序列的DNA文库被设计用于由C2c2蛋白进行序列特异性切割。该文库包含用于靶向噬菌体MS2基因组的C2c2的所有可能的间隔区序列。因为来自沙氏纤毛菌的C2c2的成熟crRNA包含28个核苷酸的最大间隔长度,所以在噬菌体基因组中平铺所有可能的28个核苷酸靶位点产生约3500个间隔区序列的文库。

[0249] 使用与实施例3中所述类似的方法在实施例9中所述的具有相似布置的结构上在实施例1中所述的硅芯片上从头合成约3500个不相同的寡核酸。将约3500个序列的文库插入载体中并转化到大肠杆菌中。使用MS2的三种稀释液,用MS2噬菌体感染大肠杆菌细胞。然后筛选文库的赋予大肠杆菌对MS2感染的抗性的序列。

[0250] 发现许多间隔区序列赋予抗性。比较间隔区表示(crRNA频率),许多间隔区序列在MS2感染的三种稀释液中表现出超过 $1.25 \log_2$ 倍的富集,而没有发现非靶向间隔区序列被富集。

[0251] 实施例15:针对于斑马鱼的sgRNA文库

[0252] DNA文库被设计具有编码约130,000个sgRNA的序列。平均而言,针对于每个斑马鱼基因设计了约5个sgRNA模板。使用与实施例3中所述类似的方法在如实施例9所述的具有相似布置的结构上在实施例1中所述的硅芯片上从头合成寡核酸。从头合成产生130,000个寡核酸,每个寡核酸从硅板表面上的不同位点延伸。从板中移除寡核酸,通过PCR进行扩增,并将其克隆到表达载体中。使每个模板进行测序。将sgRNA文库注射到斑马鱼胚胎中。将斑马鱼养至成年。然后将精子冷冻保存并通过测序筛选以鉴别种系传递的插入和缺失的序列。在种系筛选后,通过竞争性等位基因特异性PCR对精子进行基因型分型。

[0253] 实施例16:针对于小鼠的gRNA文库

[0254] DNA文库被设计具有编码约100,000个sgRNA的序列。平均而言,针对于每个小鼠基因设计了约5个sgRNA模板。使用与实施例3中所述类似的方法在如实施例9所述的具有相似布置的结构上在实施例1中所述的硅芯片上从头合成寡核酸。从头合成编码sgRNA序列的sgRNA文库以生成100,000个寡核酸。从头合成产生100,000个寡核酸,每个寡核酸从硅板表面上的不同位点延伸。从板中移除寡核酸,通过PCR进行扩增,并将其克隆到载体中。使每个模板进行测序。通过surveyor核酸酶测定或测序验证sgRNA中靶效率。然后将sgRNA显微注射到具有所需遗传背景的小鼠受精卵中。或者,在验证sgRNA效率后,将sgRNA包装到病毒载体如腺伴随病毒(AAV)中。然后将sgRNA立体定向地递送到小鼠中的所需位置。在从小鼠收集的组织中观察预先选择的靶基因的表达水平。

[0255] 实施例17:针对于受体酪氨酸激酶的gRNA文库

[0256] DNA寡核酸文库被设计具有编码靶向表25中列出的58种人受体酪氨酸激酶的基因每种基因5个sgRNA的序列,共290种不同的DNA寡核酸。使用与实施例3中所述类似的方法在如实施例9所述的具有相似布置的结构上在实施例1中所述的硅芯片上从头合成寡核酸。从板中移除寡核酸,通过PCR进行扩增,将其克隆到载体中并转移到预先选择的细胞群体中。在每个预先选择的细胞群体中将表25中列出的预先选择的基因的表达水平与暴露于没有激酶特异性sgRNA的对照载体的对照细胞群体进行比较。

表25.

基因	Hs NT ACC #	Hs PROT ACC#
ALK	NM_004304	NP_004295
LTK	NM_002344	NP_002335
AXL	NM_001699	NP_001690
MER	NM_006343	NP_006334
TYRO3	NM_006293	NP_006284
DDR1	NM_013993	NP_001945
DDR2	NM_006182	NP_006173
EGFR	NM_005228	NP_005219

ERBB2	NM_004448	NP_004439
ERBB3	NM_001982	NP_001973
ERBB4	NM_005235	NP_005226
EPHA1	NM_005232	NP_005223
EPHA2	NM_004431	NP_004422
EPHA3	NM_005233	NP_005224
EPHA4	NM_004438	NP_004429
EPHA5	L36644	P54756
EPHA6	AL133666	
EPHA7	NM_004440	NP_004431
EPHA8	AB040892	CAB81612
EPHB1	NM_004441	NP_004432
EPHB2	AF025304	AAB94602
EPHB3	NM_004443	NP_004434
EPHB4	NM_004444	NP_004435
EPHB6	NM_004445	NP_004436
EPHX		
FGFR1	M34641	AAA35835
FGFR2	NM_000141	NP_000132
FGFR3	NM_000142	NP_000133
FGFR4	NM_002011	NP_002002
IGF1R	NM_000875	NP_000866
INSR	NM_000208	NP_000199
INSRR	J05046	AAC31759
MET	NM_000245	NP_000236
RON	NM_002447	NP_002438
MUSK	NM_005592	NP_005583
CSF1R	NM_005211	NP_005202
FLT3	NM_004119	NP_0041110

KIT	NM_000222	NP_000213
PDGFRA	NM_006206	NP_006197
PDGFRB	NM_002609	NP_002600
PTK7	NM_002821	NP_002812
RET	X12949	P07949
ROR1	NM_005012	NP_005003
ROR2	NM_004560	NP_004551
ROS1	NM_002944	NP_002935
RYK	S59184	AAB26341
TEK	NM_000459	NP_000450
TIE	NM_005424	NP_005415
NTRK1	NM_002529	NP_002520
NTRK2	NM_006180	NP_006171
NTRK3	NM_002530	NP_002521
VEGFR1	NM_002019	NP_002010
VEGFR2		AAB88005
VEGFR3	NM_002020	NP_002011
AATYK	NM_004920	NP_004911
AATYK2	NM_014916	NP_055731
AATYK3		
DKFZp761P1010	NM_018423	NP_060893

[0257] 实施例18:针对于人激酶组的gRNA文库

[0258] DNA寡核酸文库被设计具有编码靶向518种人激酶的基因每种基因5个sgRNA的序列,共2,590种不同的DNA寡核酸。从板中移除寡核酸,通过PCR进行扩增,将其克隆到载体中并转移到预先选择的细胞群体中。在每个预先选择的细胞群体中将预先选择的518个基因的表达水平与暴露于没有激酶特异性sgRNA的对照载体的对照细胞群体进行比较。

[0259] 实施例19:针对于人磷酸酶组(Phosphatome)的gRNA文库

[0260] DNA寡核酸文库被设计具有编码靶向200种人磷酸酶组的基因每种基因5个sgRNA的序列,共1000种不同的DNA寡核酸。从板中移除寡核酸,通过PCR进行扩增,将其克隆到载体中并转移到预先选择的细胞群体中。在每个预先选择的细胞群体中将200个预先选择的基因的表达水平与暴露于没有激酶特异性sgRNA的对照载体的对照细胞群体进行比较。

[0261] 虽然本文中已经示出并描述了本发明的优选实施方案,但是对于本领域技术人员显而易见的是,这些实施方案仅以示例的方式提供。本领域技术人员在不脱离本发明的情

况下现将想到多种变化、改变和替换。应当理解,本文所述的本发明实施方案的各种替代方案可用于实施本发明。以下述权利要求旨在限定本发明的范围,并由此涵盖这些权利要求范围内的方法和结构及其等同物。

## 序列表

<110> 特韦斯特生物科学公司

<120> 从头合成的核酸文库

<130> 44854-727.601

<140>

<141>

<150> 62/419,881

<151> 2016-11-09

<150> 62/411,338

<151> 2016-10-21

<150> 62/393,948

<151> 2016-09-13

<150> 62/382,191

<151> 2016-08-31

<150> 62/378,134

<151> 2016-08-22

<160> 44

<170> PatentIn version 3.5

<210> 1

<211> 62

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:合成寡核苷酸

<220>

<221> 修饰的碱基

<222> (51) .. (52)

<223> 胸苷-琥珀酰己酰胺CED亚磷酰胺

<400> 1

agacaatcaa ccatttgggg tggacagcct tgacctctag acttcggcat tttttttttt 60

tt 62

<210> 2

<211> 112

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:合成多核苷酸

<220>

- <221> 修饰的碱基  
<222> (101) .. (102)  
<223> 胸苷-琥珀酰己酰胺CED亚磷酰胺  
<400> 2  
cgggatcctt atcgtcatcg tcgtacagat cccgacccat ttgctgtcca ccagtcacgc 60  
tagccataacc atgatgatga tgatgatgag aaccccgcat tttttttttt tt 112
- <210> 3  
<211> 19  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成引物  
<400> 3  
atgcgggggtt ctcatcatc 19
- <210> 4  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成引物  
<400> 4  
cgggatcctt atcgtcatcg 20
- <210> 5  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 5  
ataactcaat ttgtaaaaaa gttttagagc tatgctgttt tg 42
- <210> 6  
<211> 86  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 6  
ggaaccattc aaaacagcat agcaagttaa aataaggcta gtccgttatc aacttgaaaa 60  
agtggcaccg agtcggtgct tttttt 86

- <210> 7  
<211> 38  
<212> DNA  
<213> 嗜热链球菌  
<400> 7  
tactcaactt gaaaaggtgg caccgattcg gtgttttt 38
- <210> 8  
<211> 39  
<212> DNA  
<213> 变异链球菌  
<400> 8  
tacacaactt gaaaaagtgc gcaccgattc ggtgctttt 39
- <210> 9  
<211> 40  
<212> DNA  
<213> 无害利斯特氏菌  
<400> 9  
ttatcaactt ttaattaagt agcgctgttt cggcgctttt 40
- <210> 10  
<211> 40  
<212> DNA  
<213> 运动支原体  
<400> 10  
tatgccgtaa ctactactta ttttcaaat aagtagtttt 40
- <210> 11  
<211> 36  
<212> DNA  
<213> 空肠弯曲杆菌  
<400> 11  
gactctgcgg gttacaate ccctaaaacc gctttt 36
- <210> 12  
<211> 117  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成多核苷酸  
<400> 12  
taatacgact cactataggg gatgcgcgca gttgtccggt ttagagctag aaatagcaag 60  
ttaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttt 117

- <210> 13  
<211> 17  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 13  
taatacgact cactata 17
- <210> 14  
<211> 18  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 14  
ggatgcgcgc agttgtcc 18
- <210> 15  
<211> 80  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 15  
gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60  
ggcaccgagt cgggtgctttt 80
- <210> 16  
<211> 121  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成多核苷酸  
<400> 16  
gagctaatac gactcactat aggggatgcg cgcagttgtc cgtttagag ctagaaatag 60  
caagttaaaa taaggctagt ccgttatcaa cttgaaaaag tggcaccgag tcgggtgcttt 120  
t 121
- <210> 17  
<211> 122  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的描述:合成多核苷酸  
<400> 17  
cgagctaata cgactcacta taggggatgc gcgcagttgt ccgtttttaga gctagaaata 60  
gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 120  
tt 122  
<210> 18  
<211> 19  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 18  
taatacgact cactatagg 19  
<210> 19  
<211> 80  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 19  
aaaagcaccg actcgggtgcc actttttcaa gttgataacg gactagcctt attttaactt 60  
gctattttcta gctctaaaac 80  
<210> 20  
<211> 58  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 20  
taatacgact cactataggg gatgcgcgca gttgtccggt ttagagctag aaatagca 58  
<210> 21  
<211> 65  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 21  
gaaattaata cgactcacta taggggatgc gcgcagttgt ccgtttttaga gctagaaata 60

gcaag 65  
<210> 22  
<211> 66  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 22  
gagctaatac gactcactat aggggatgcg cgcagttgtc cgtttttagag ctagaatag 60  
caagtt 66  
<210> 23  
<211> 75  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 23  
gagctaatac gactcactat aggggatgcg cgcagttgtc cgtttttagag ctagaatag 60  
caagttaaaa taagg 75  
<210> 24  
<211> 14  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 24  
aaaagcaccg actc 14  
<210> 25  
<211> 15  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 25  
aaaagcaccg actcg 15  
<210> 26  
<211> 16  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 26  
aaaagcaccg actcgg 16  
<210> 27  
<211> 17  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 27  
aaaagcaccg actcgg 17  
<210> 28  
<211> 24  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 28  
gaaattaata cgactcacta tagg 24  
<210> 29  
<211> 23  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 29  
gagctaatac gactcactat agg 23  
<210> 30  
<211> 25  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 30  
gcgagctaatac gactcact atagg 25  
<210> 31  
<211> 120  
<212> DNA

- <213> 人工序列  
<220>  
<223> 人工序列的描述:合成多核苷酸  
<400> 31  
cgagctaata cgactcacta taggggcaca acgtggagga tggcagcgtg cagctggctg 60  
atcactacca gcaaaacact ccaatcggtg atggtcctgt tgcaccgagt cggtgctttt 120  
<210> 32  
<211> 71  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 32  
cgagctaata cgactcacta taggggcaca acgtggagga tggcagcgtg cagctggctg 60  
atcactacca g 71  
<210> 33  
<211> 69  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 33  
aaagcaccga ctcggtgcaa caggaccatc accgattgga gtgttttgct ggtatgatc 60  
agccagctg 69  
<210> 34  
<211> 130  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成多核苷酸  
<220>  
<221> 修饰的碱基  
<222> (25) .. (42)  
<223> a、c、t、g、未知或其他  
<400> 34  
cgagctaata cgactcacta taggnnnnnn nnnnnnnnnn nngttttaga gctatgctga 60  
aaagcatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt 120  
cggtgctttt 130  
<210> 35

<211> 720

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:合成多核苷酸

<400> 35

```
atgcgtaaag gcgaagagct gttcactggg gtcgtcccta ttctgggtgga actggatggt 60
gatgtcaacg gtcataagtt ttccgtgctg ggcgagggtg aaggtgacgc aactaatggt 120
aaactgacgc tgaagttcat ctgtactact ggtaaactgc cggtagcttg gccgactctg 180
gtaacgacgc tgacttatgg tgttcagtgc tttgctcgtt atccggacca tatgaagcag 240
catgacttct tcaagtccgc catgccggaa ggctatgtgc aggaacgcac gatttccttt 300
aaggatgacg gcacgtacaa aacgcgtgcg gaagtgaat ttgaaggcga taccctggta 360
aaccgcattg agctgaaagg cattgacttt aaagaagacg gcaatatect gggccataag 420
ctggaataca attttaacag ccacaatggt tacatcaccg ccgataaaca aaaaaatggc 480
attaaagcga attttaaaat tcgccacaac gtggaggatg gcagcgtgca gctggctgat 540
cactaccagc aaaacactcc aatcgggtgat ggtcctgttc tgctgccaga caatcactat 600
ctgagcacgc aaagcgttct gtctaaagat ccgaacgaga aacgcgatca tatggttctg 660
ctggagttcg taaccgcagc gggcatcacg catggtatgg atgaactgta caaatgataa 720
```

<210> 36

<211> 70

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:合成寡核苷酸

<400> 36

```
cgagctaata cgactcacta taggaacgca cgatttcctt tagttttaga gctatgctga 60
aaagcatagc 70
```

<210> 37

<211> 70

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:合成寡核苷酸

<400> 37

```
cgagctaata cgactcacta taggcattga ctttaaagaa gagttttaga gctatgctga 60
aaagcatagc 70
```

<210> 38

<211> 70

<212> DNA

- <213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 38  
cgagctaata cgactcacta taggaggatg gcagcgtgca gcgttttaga gctatgctga 60  
aaagcatagc 70  
<210> 39  
<211> 88  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 39  
aaaagcaccg actcggtgcc actttttcaa gttgataacg gactagcctt attttaactt 60  
gctatgcttt tcagcatagc tctaaaac 88  
<210> 40  
<211> 103  
<212> RNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成多核苷酸  
<220>  
<221> 修饰的碱基  
<222> (1) .. (20)  
<223> a、c、u、g、未知或其他  
<400> 40  
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60  
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103  
<210> 41  
<211> 76  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 41  
cgagctaata cgactcacta taggggatgc gcgcagttgt ccgttttaga gctagaaata 60  
gcaagttaaa ataagg 76  
<210> 42  
<211> 30

<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 42  
taaggctagt ccgttatcaa cttgaaaaag 30  
<210> 43  
<211> 27  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成寡核苷酸  
<400> 43  
ctagtccggt atcaacttga aaaagtg 27  
<210> 44  
<211> 122  
<212> DNA  
<213> 人工序列  
<220>  
<223> 人工序列的描述:合成多核苷酸  
<400> 44  
gaaattaata cgactcacta taggggatgc gcgcagttgt ccgttttaga gctagaaata 60  
gcaagttaaa ataaggctag tccgttatca acttgaaaa gtggcaccga gtcggtgctt 120  
tt 122

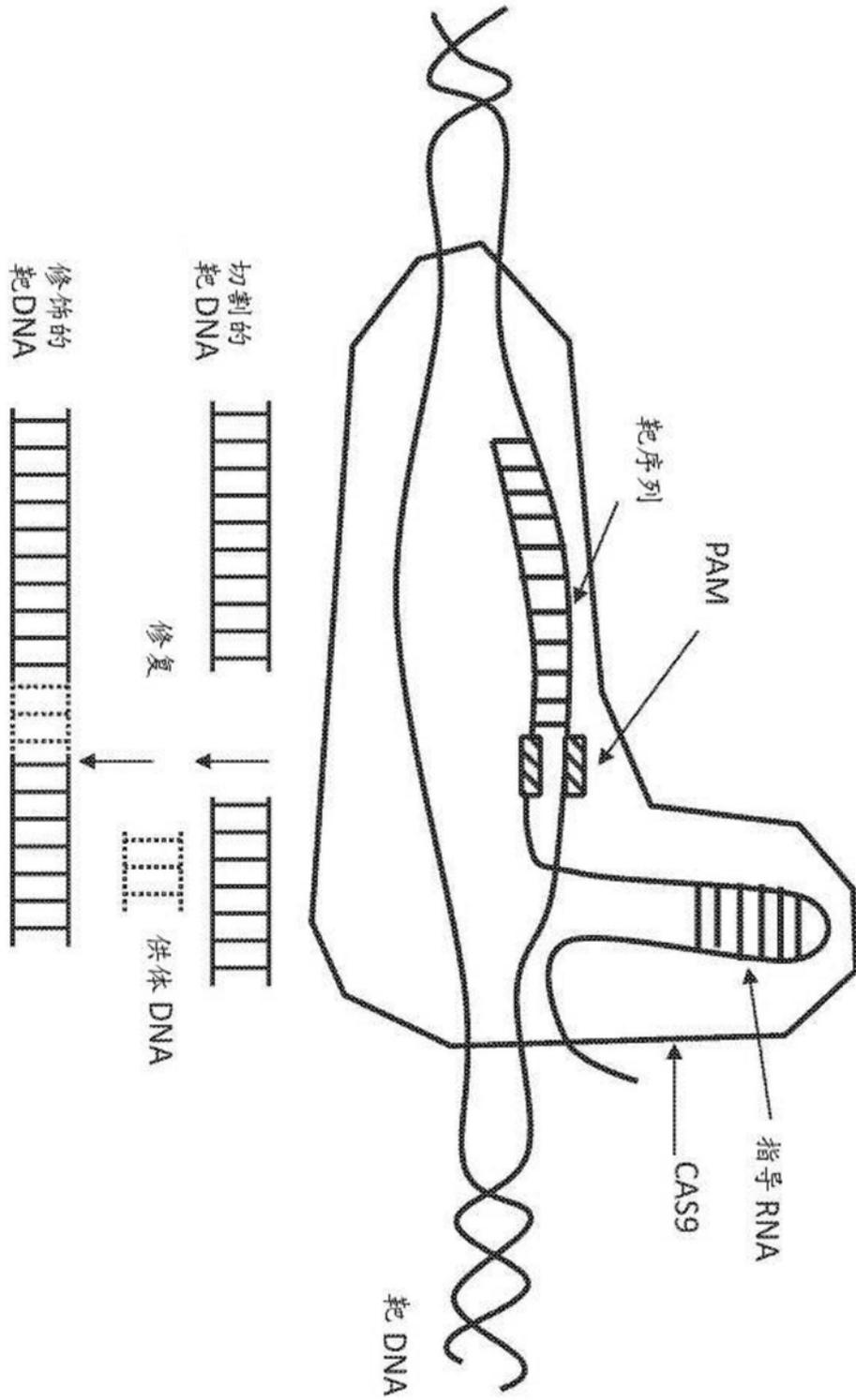


图1A

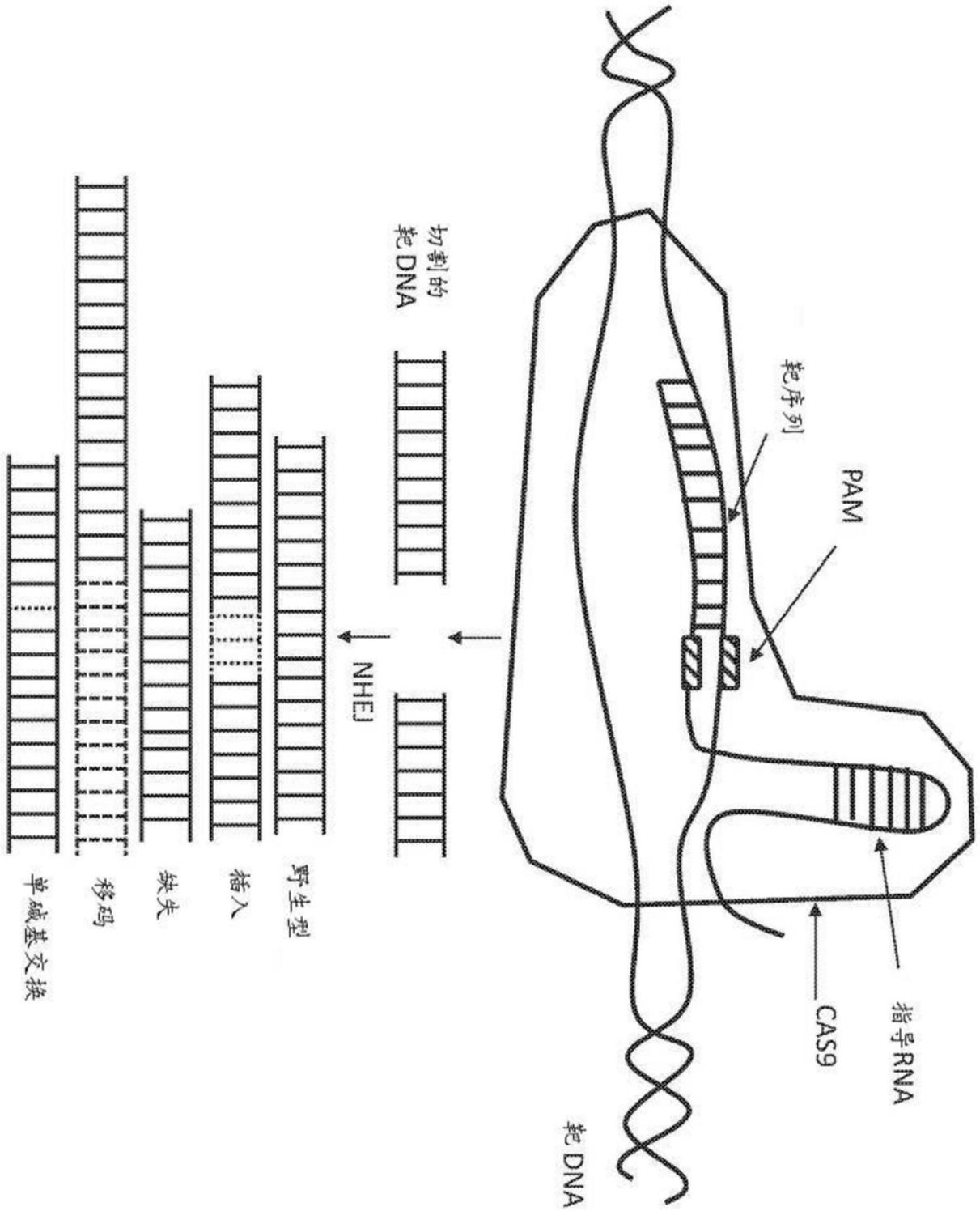


图1B

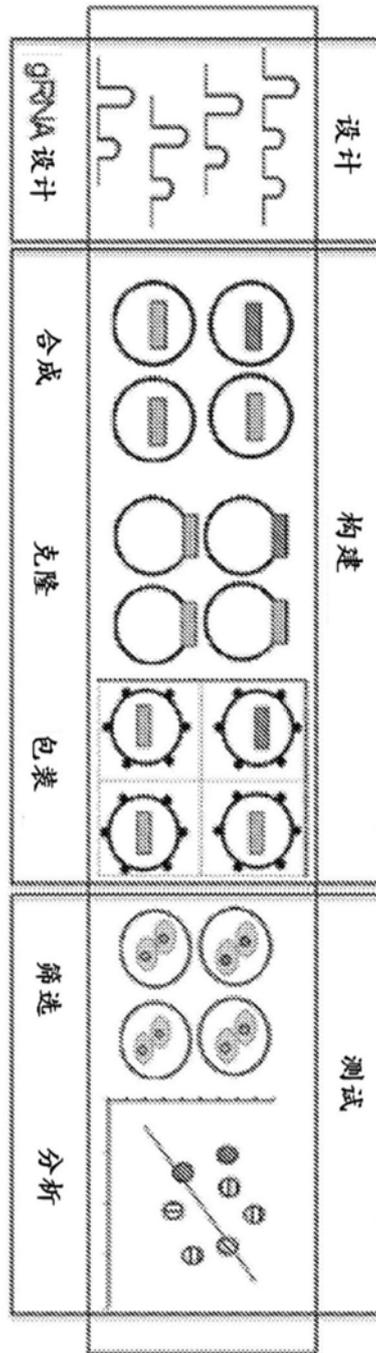


图2

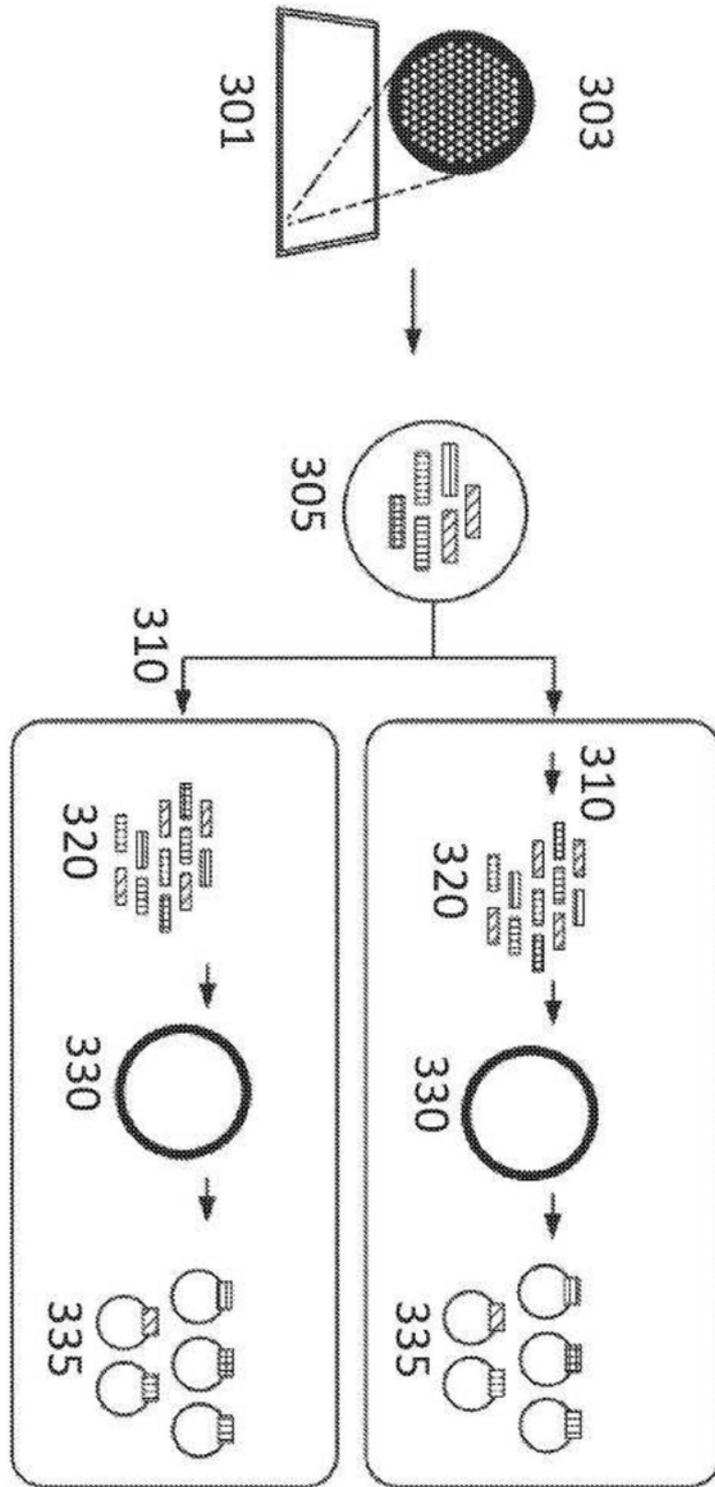


图3



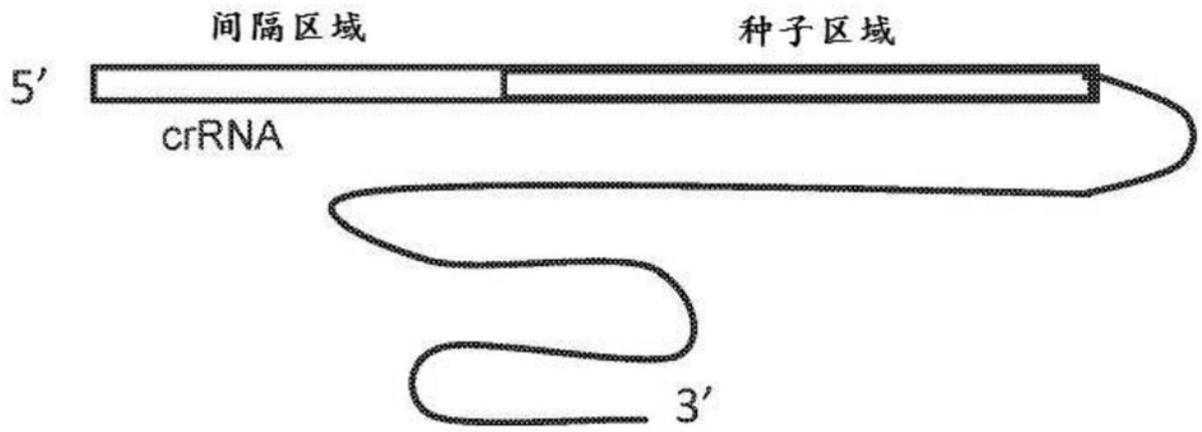


图4B

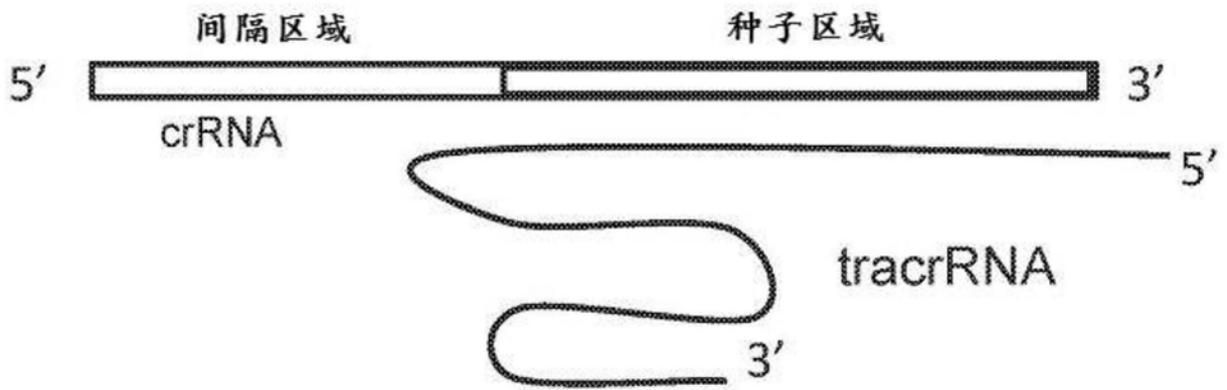


图4C

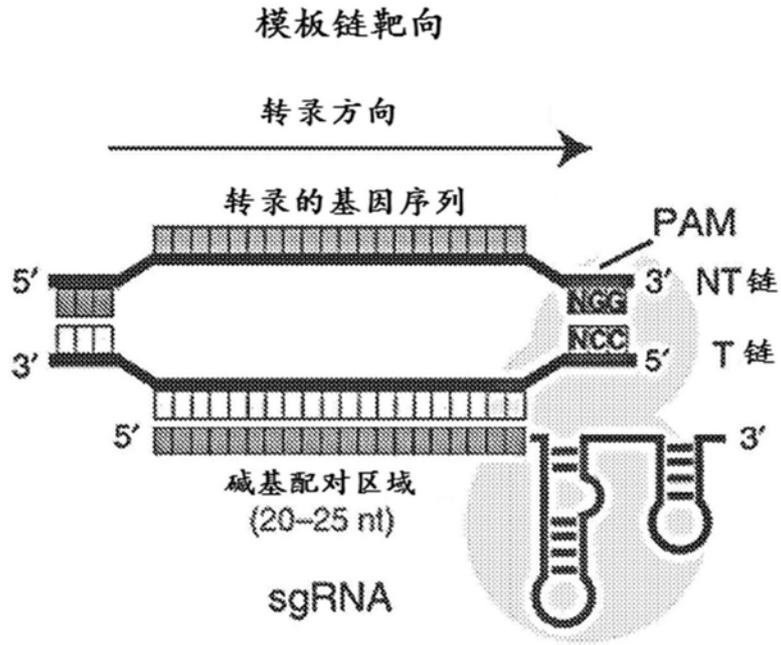


图5A

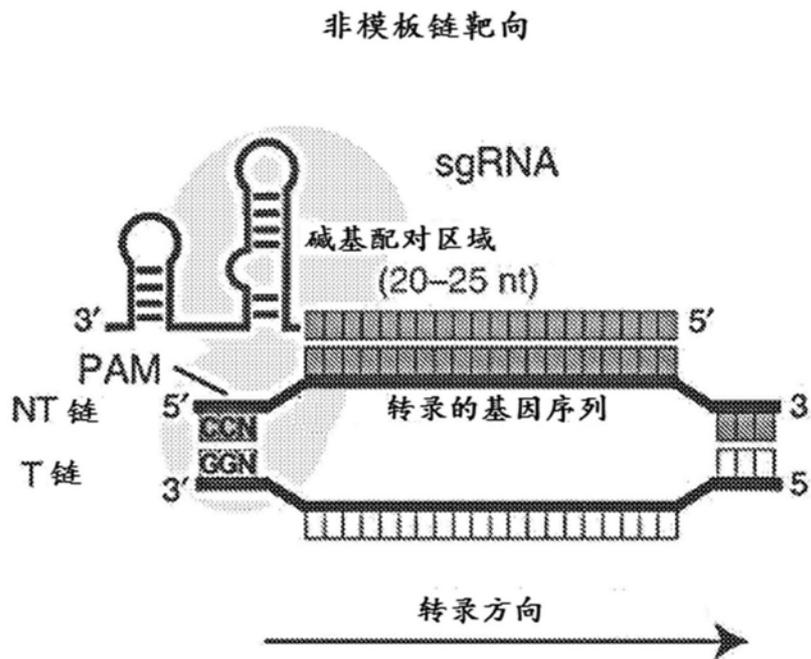


图5B

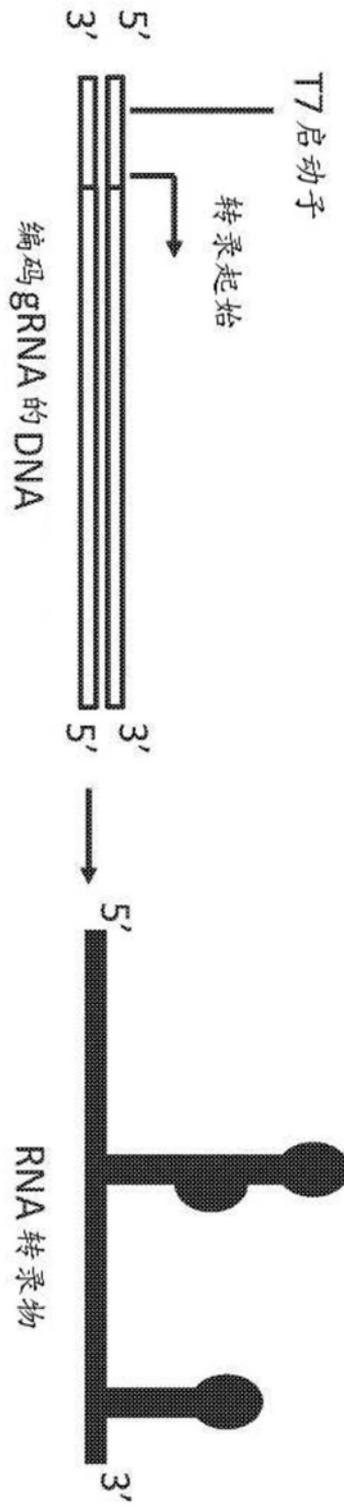


图6A

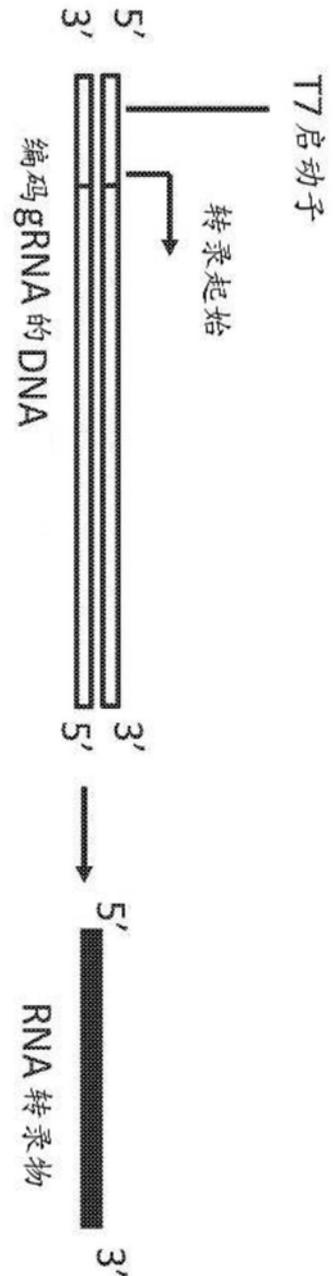


图6B

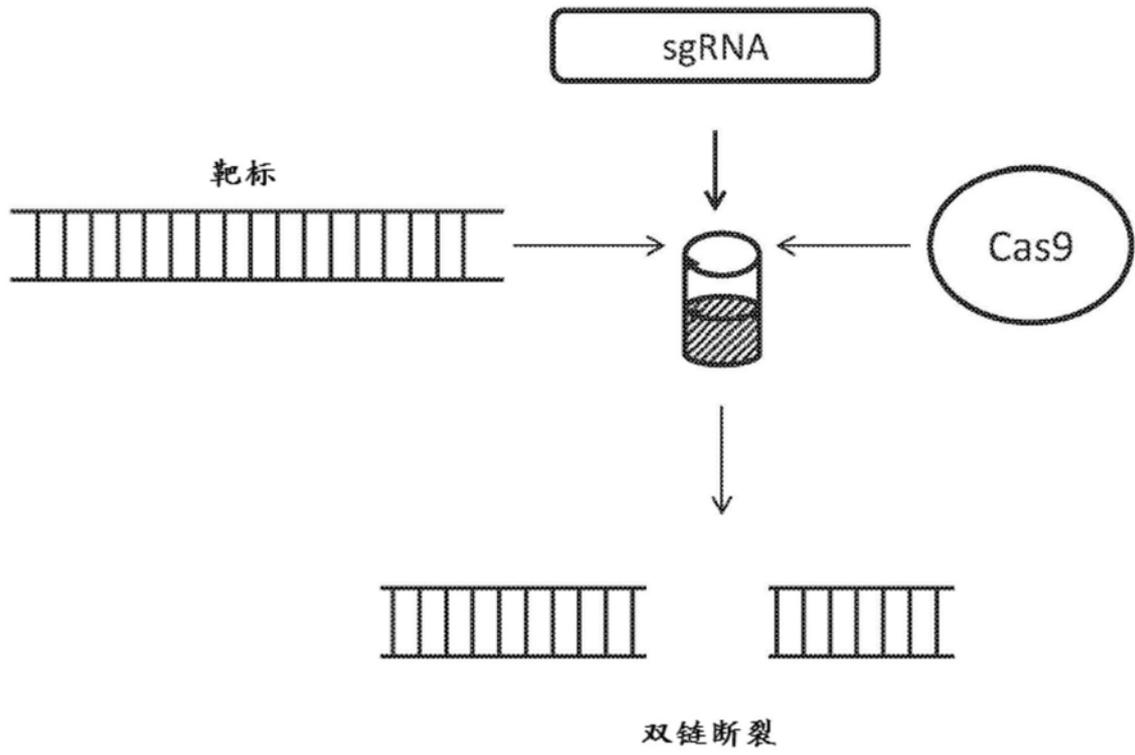


图7

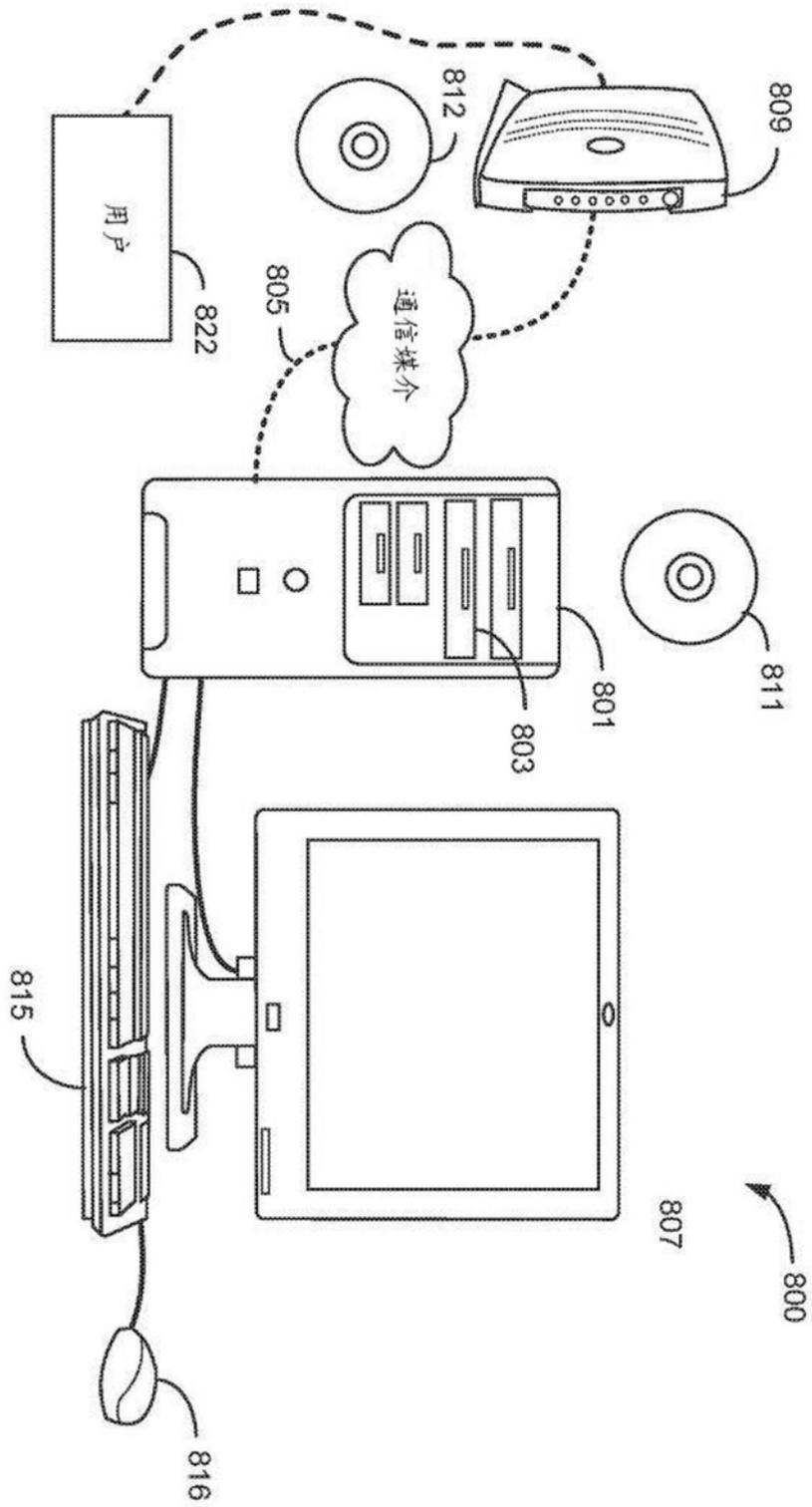


图8

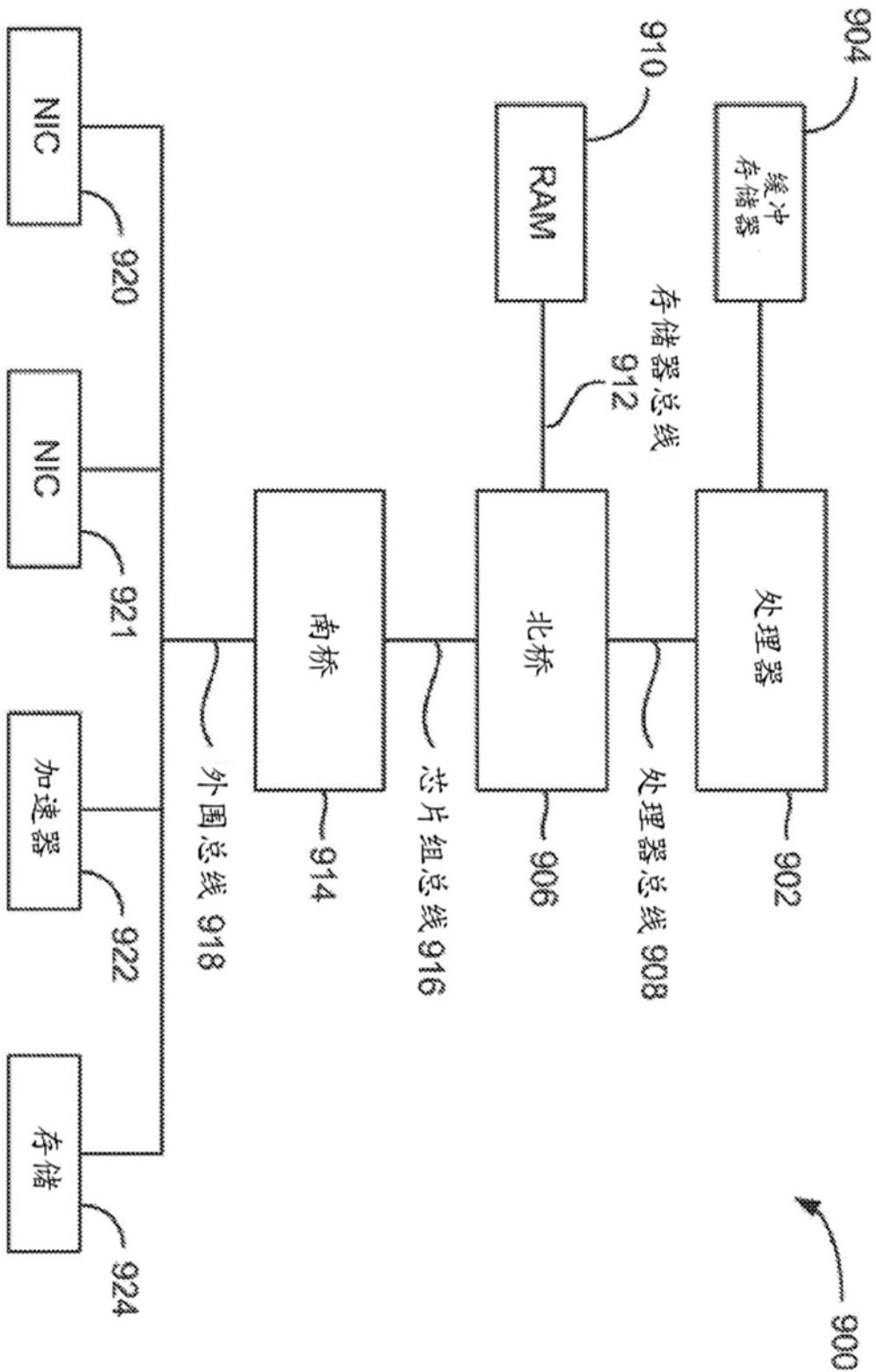


图9

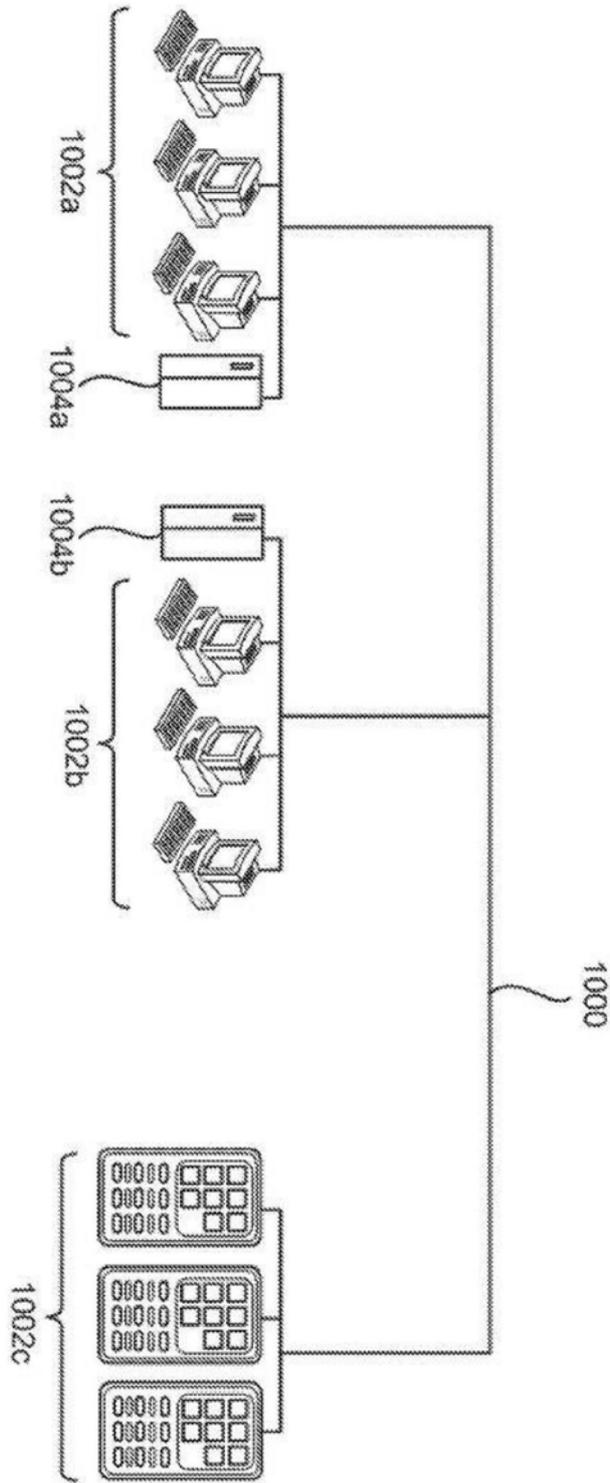


图10

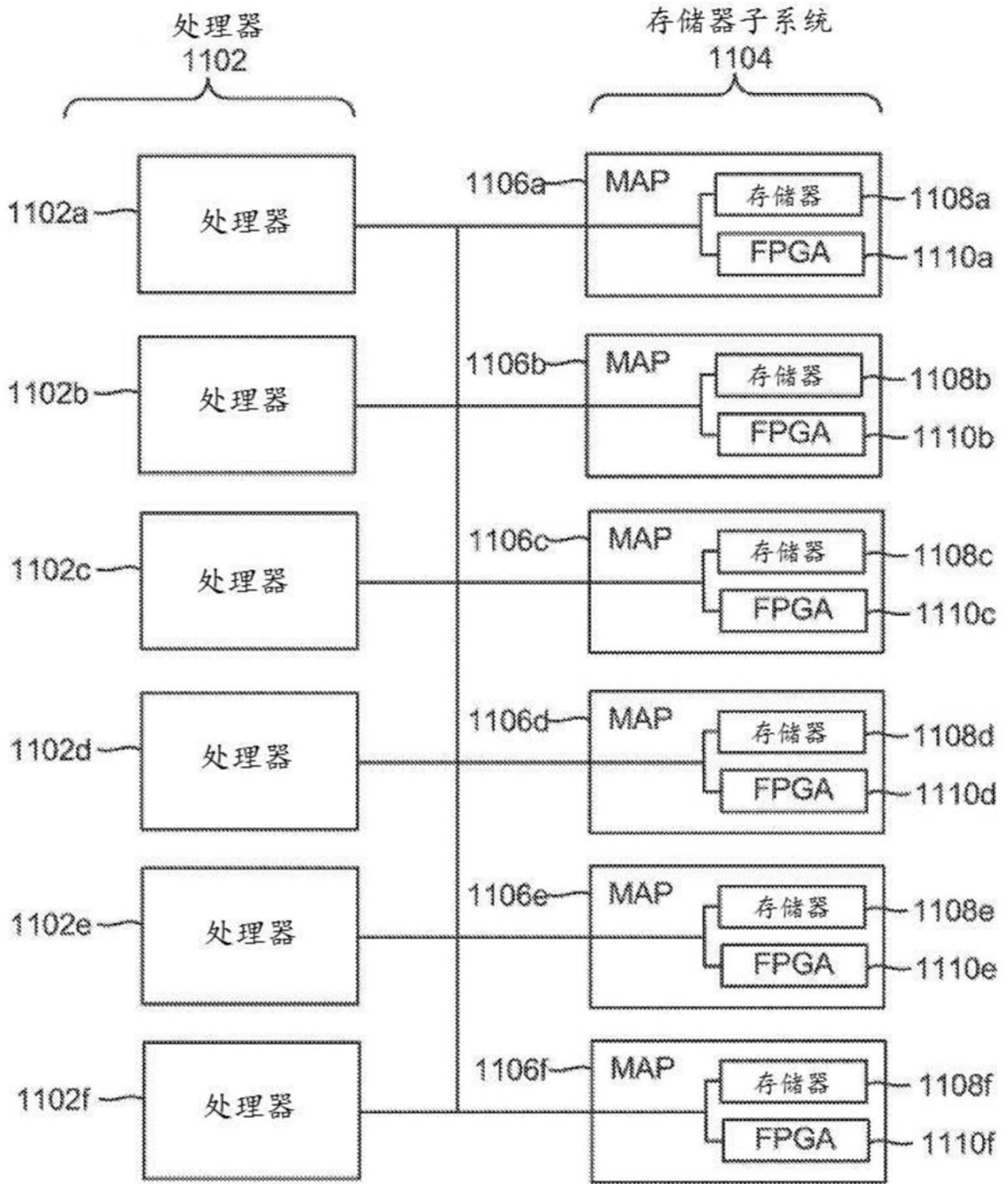


图11

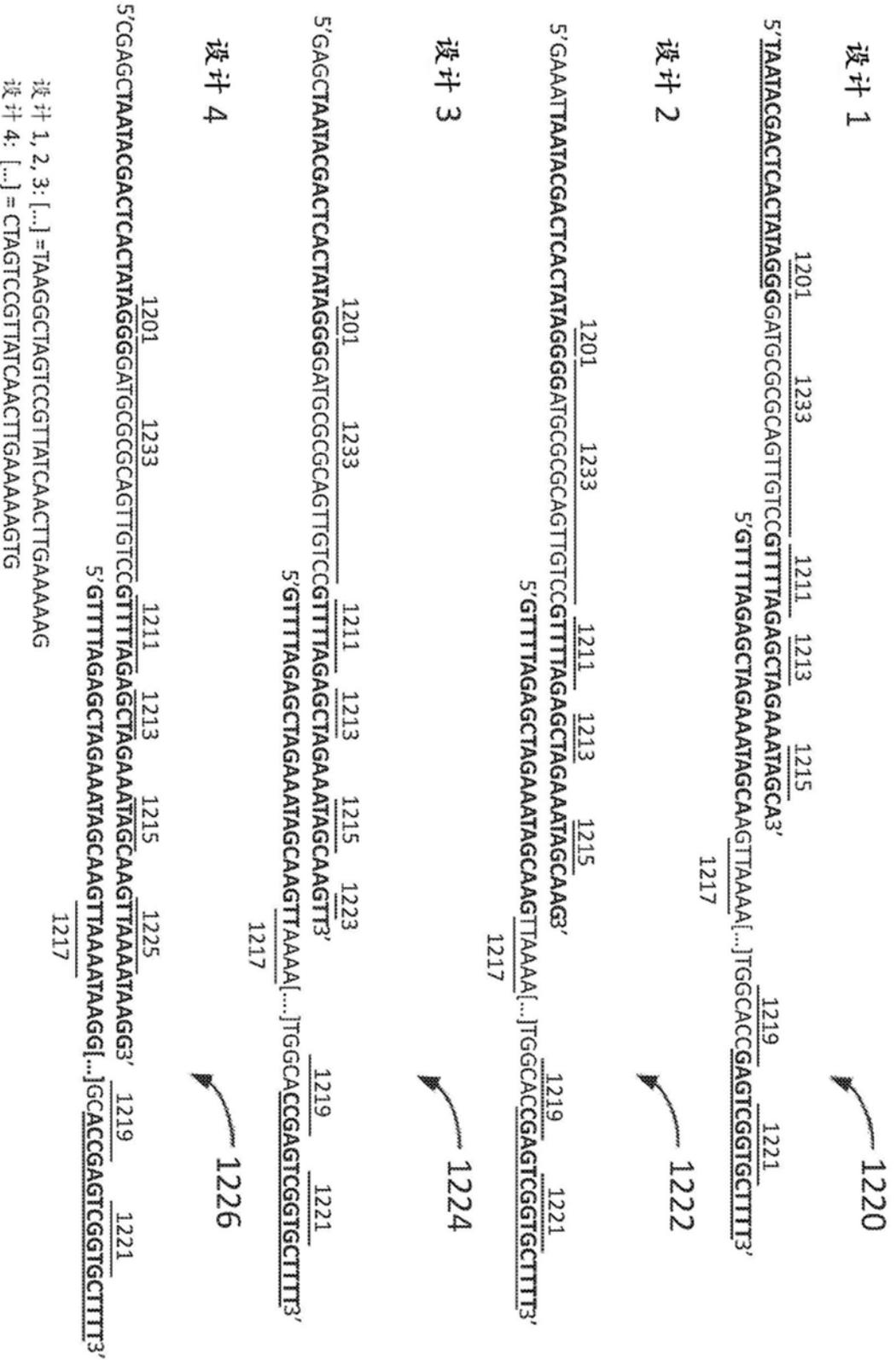


图12

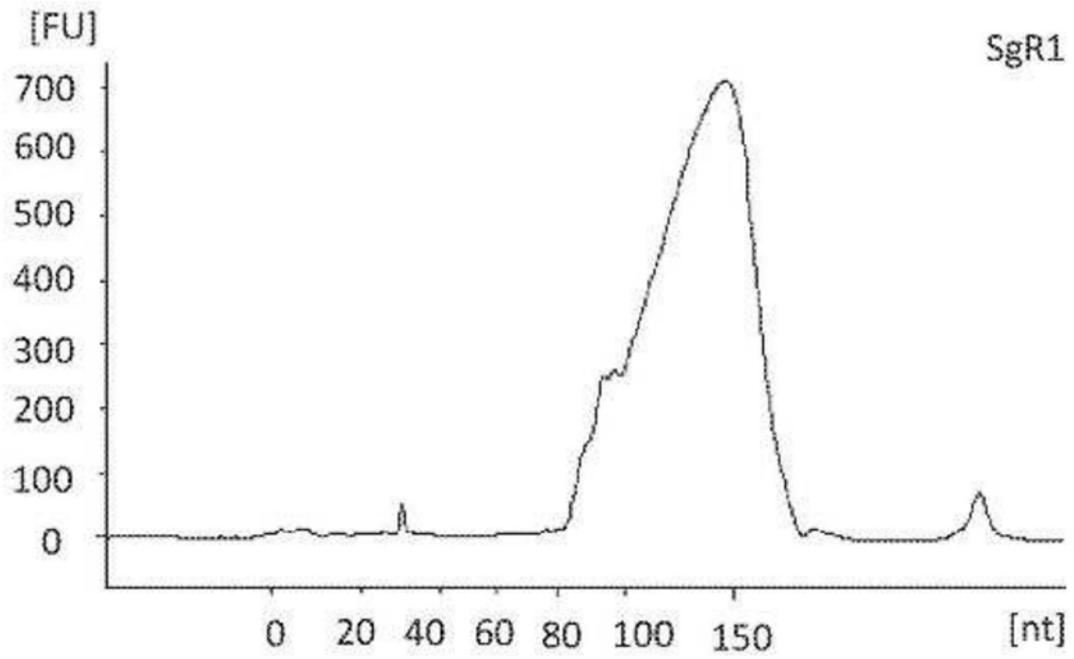


图13A

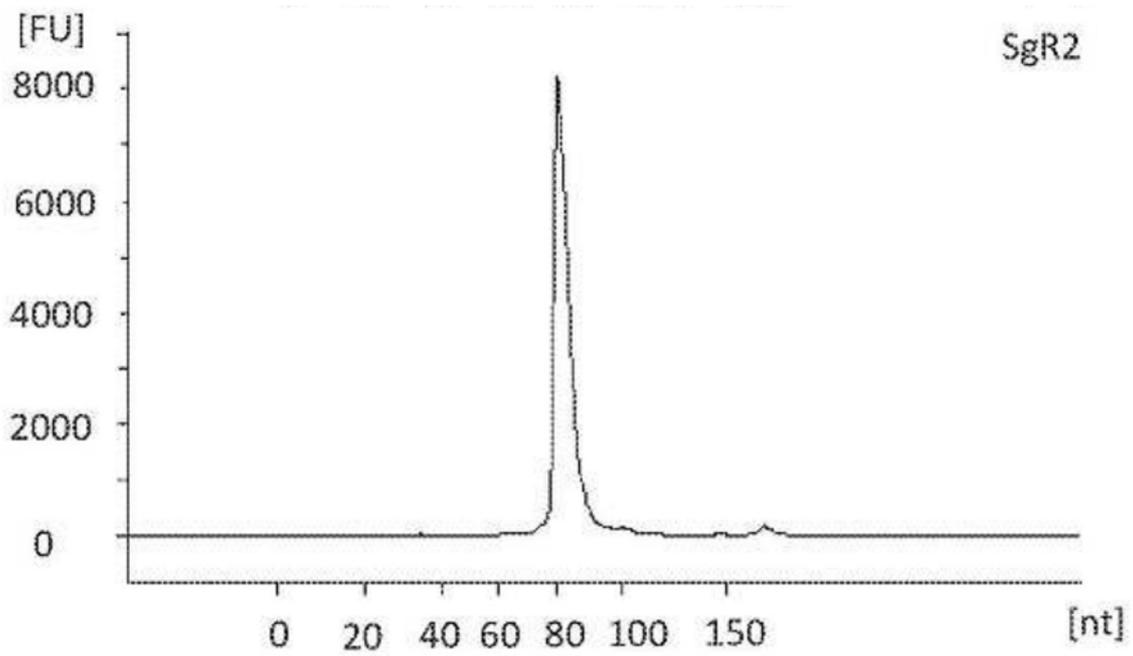
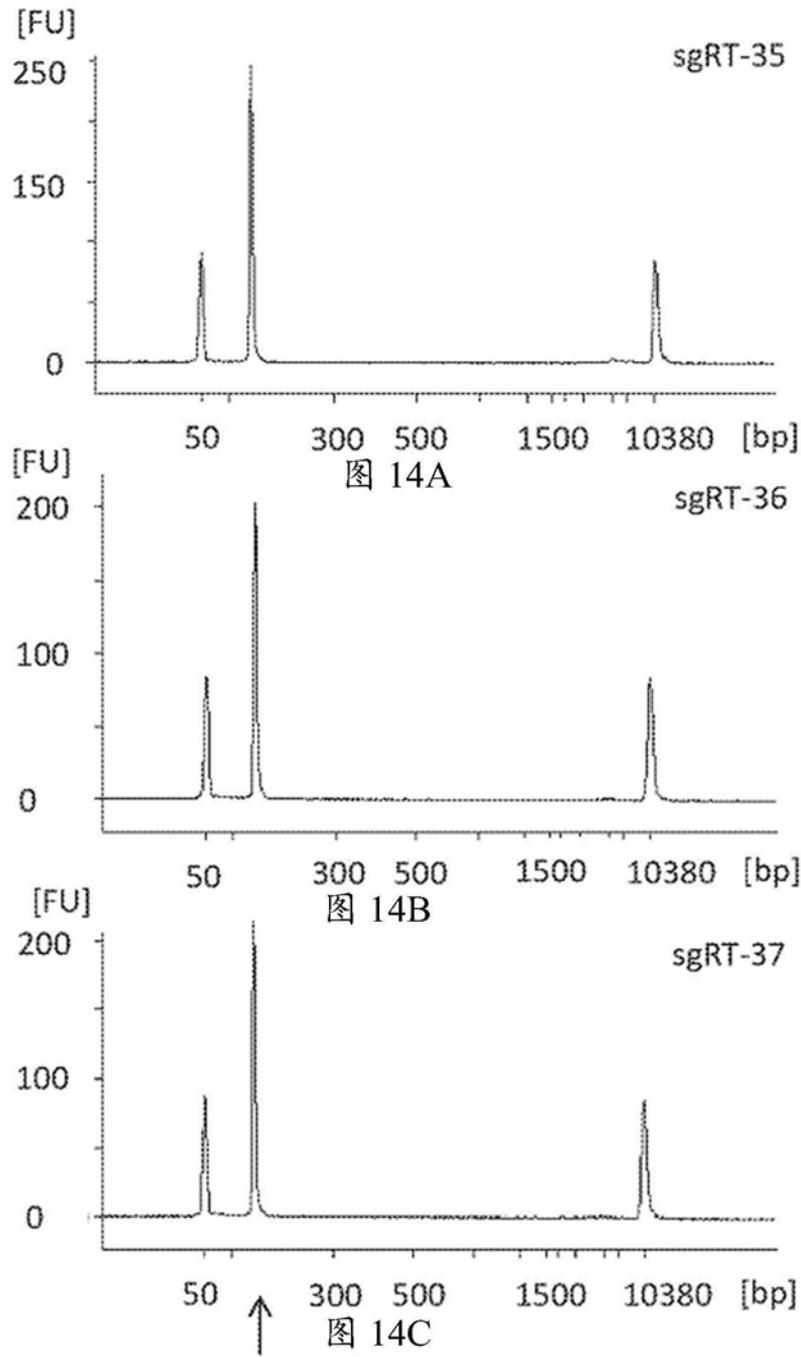


图13B

sgRNA 模板组装



sgRNA 转录后

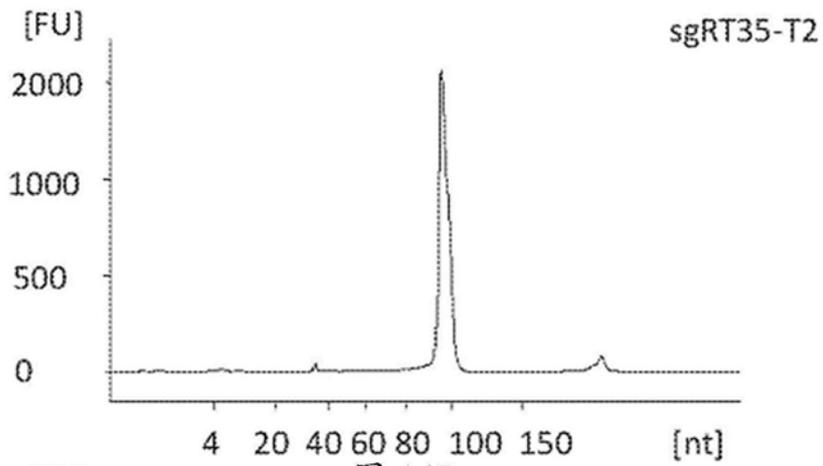


图 14D

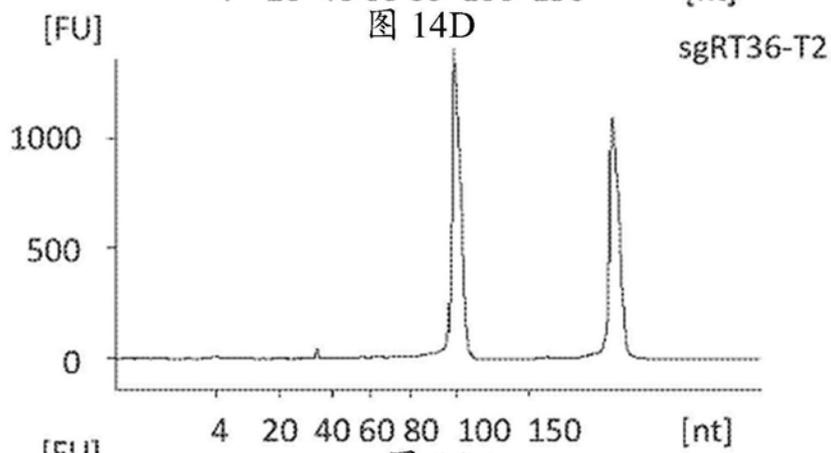


图 14E

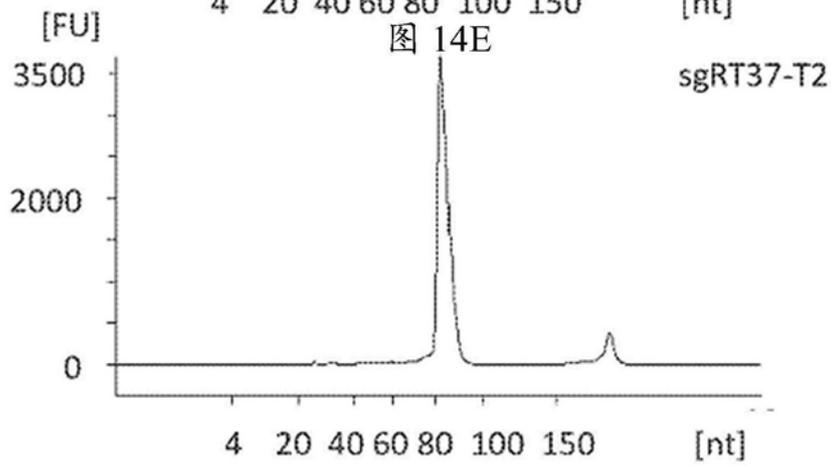


图 14F

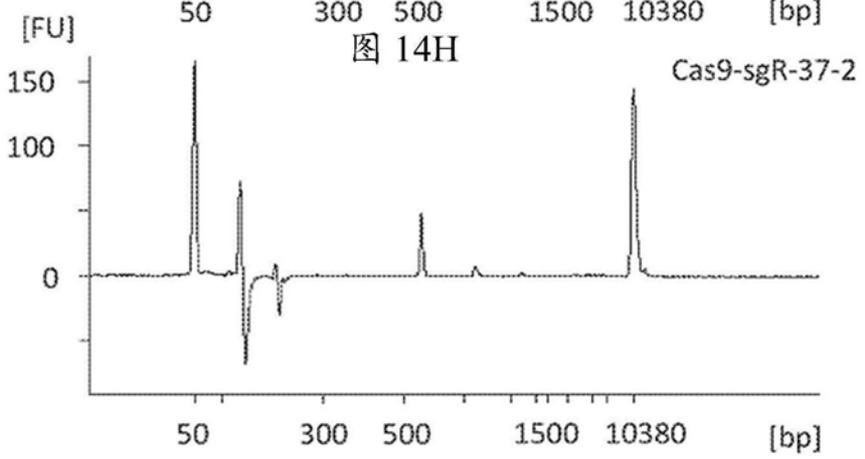
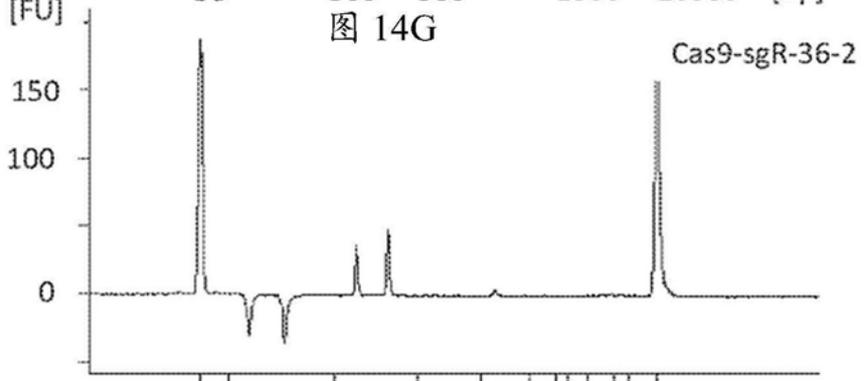
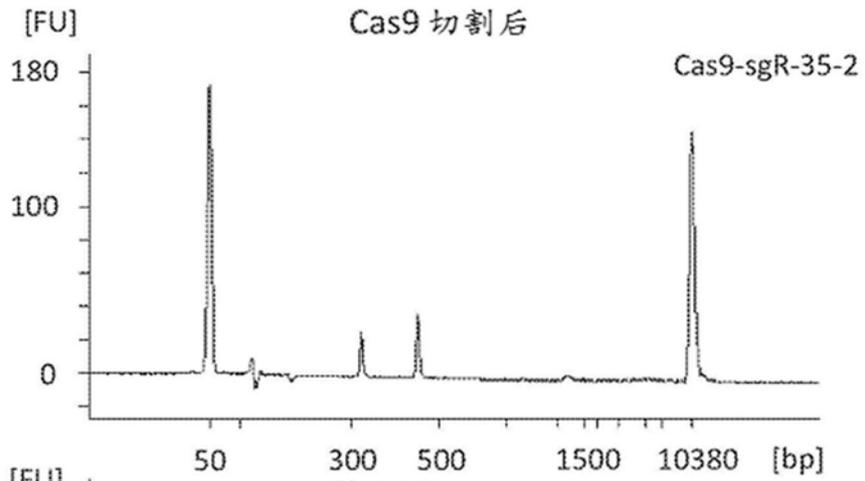


图 14I

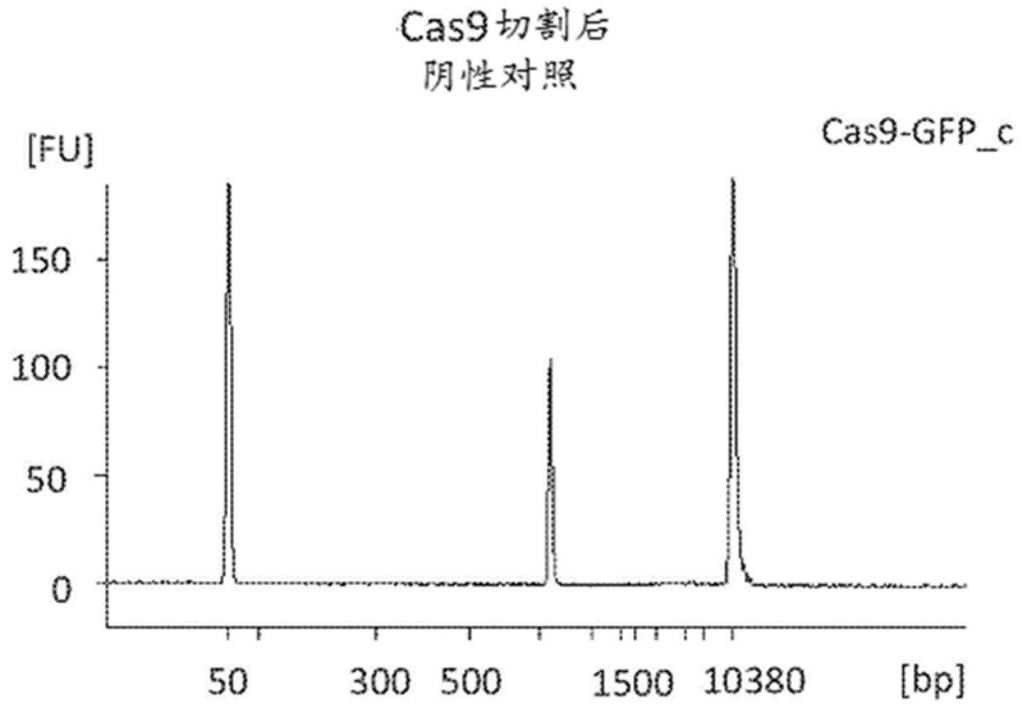


图14J

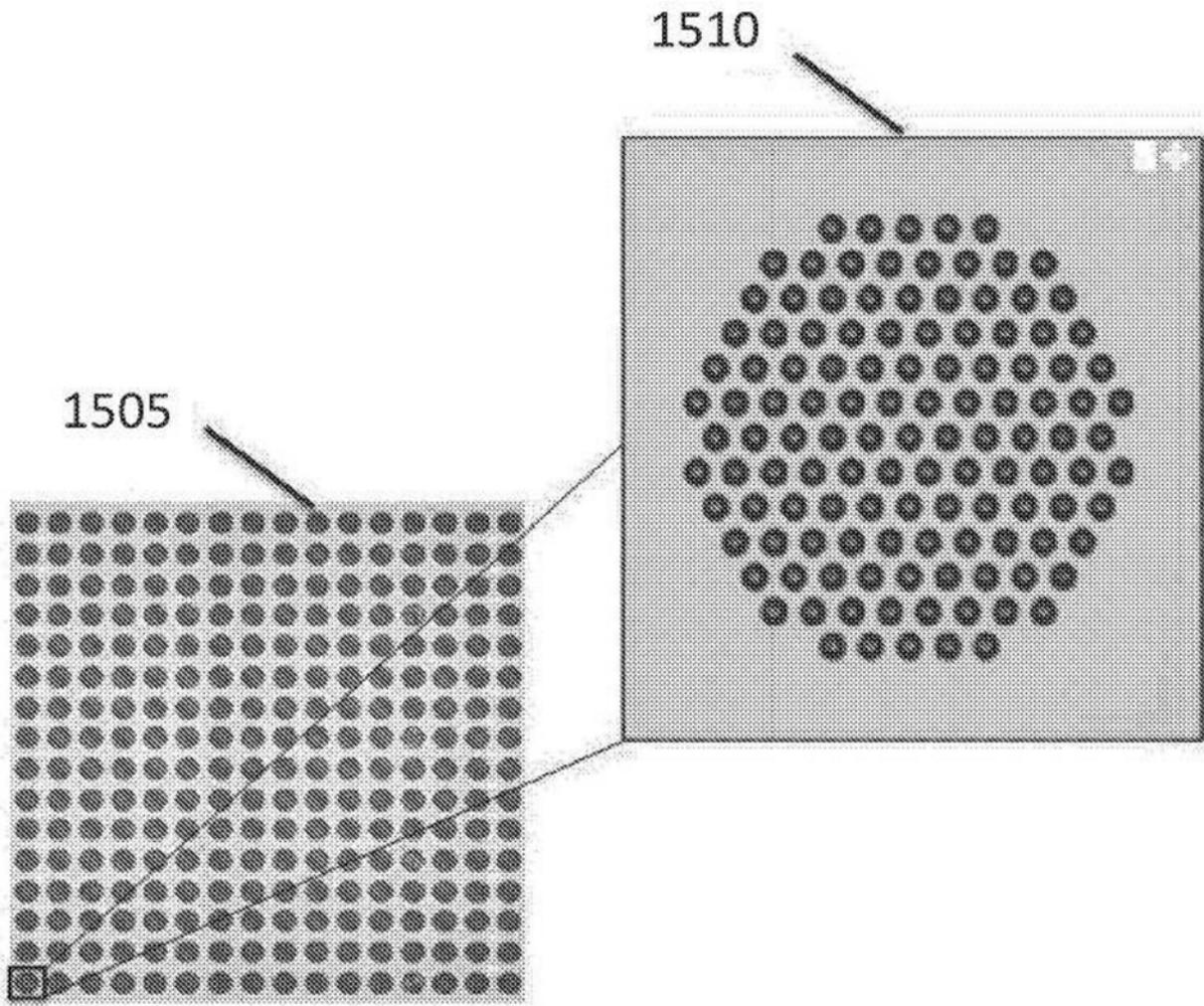


图15

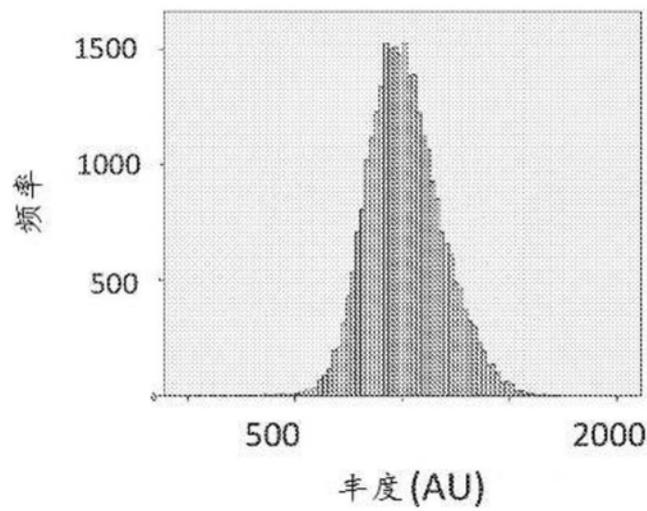


图16A

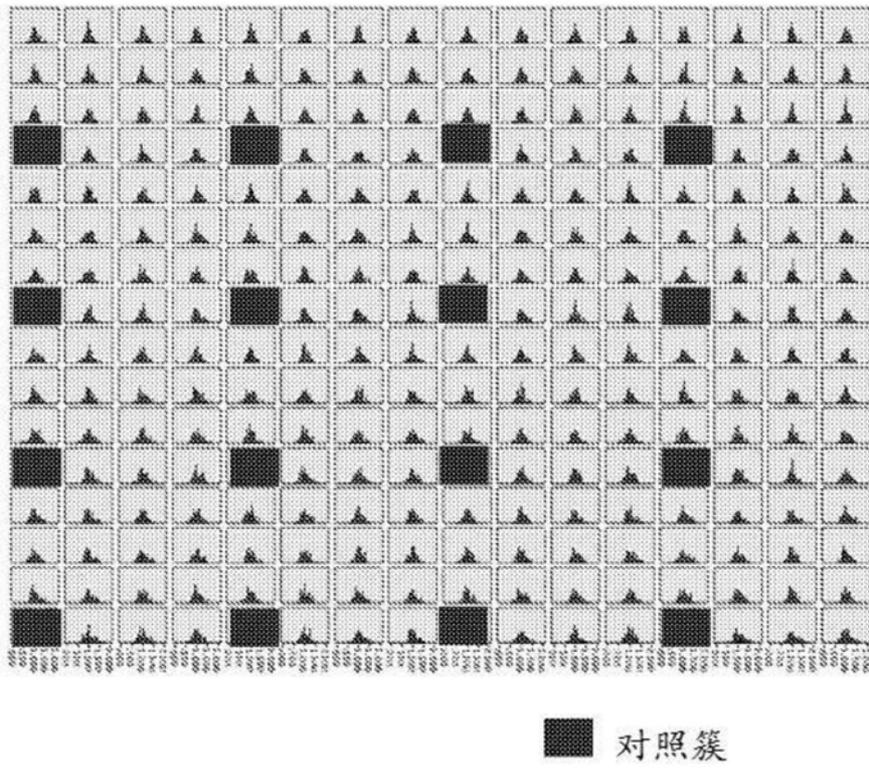


图16B

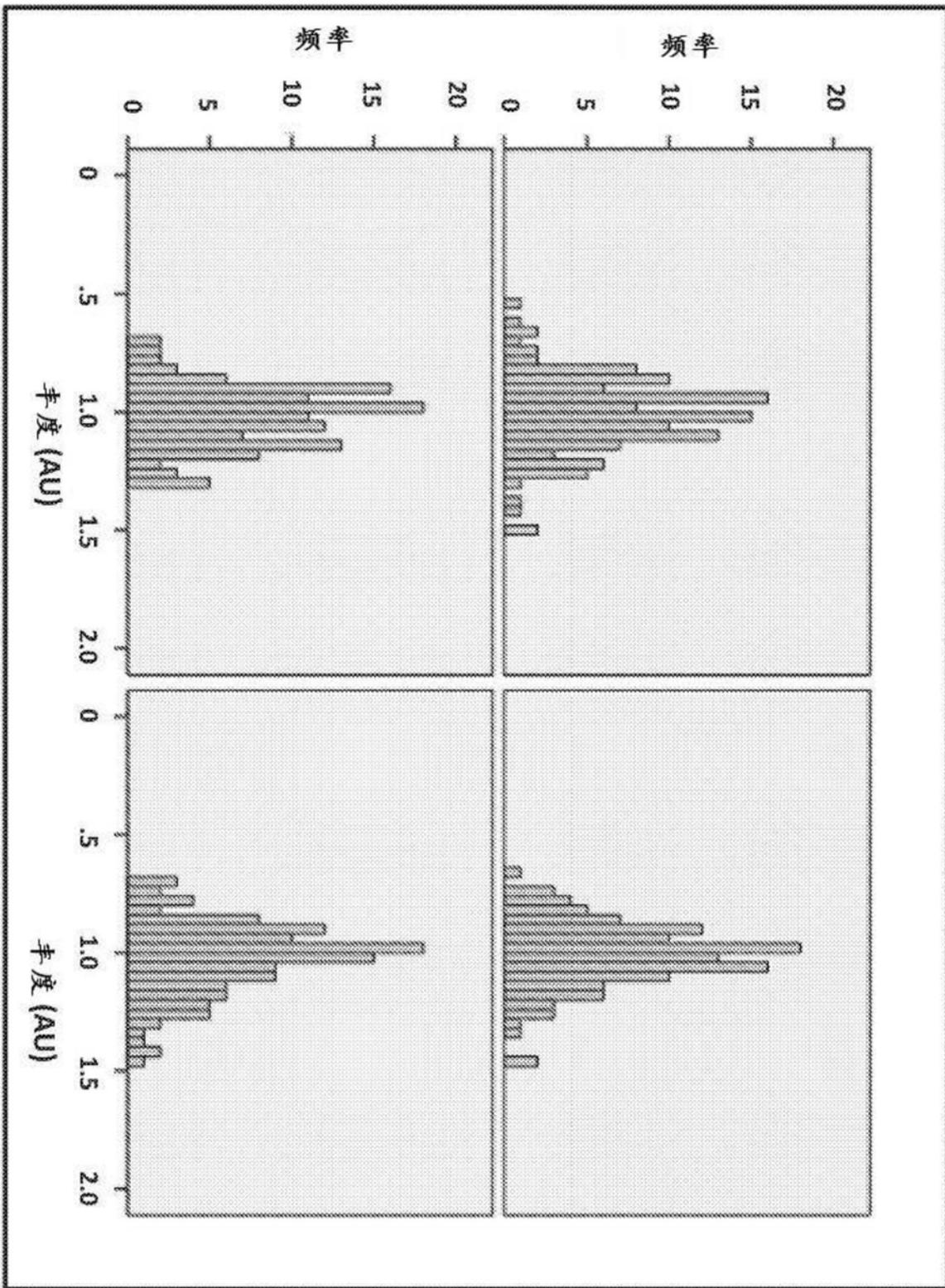


图17

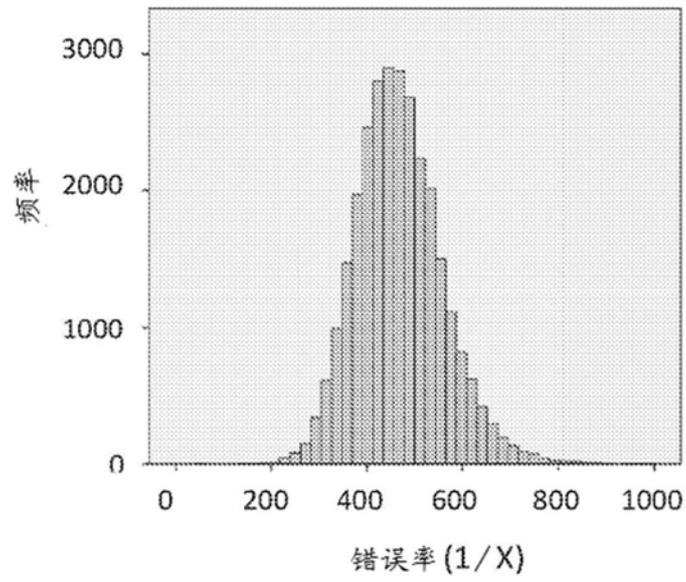


图18A

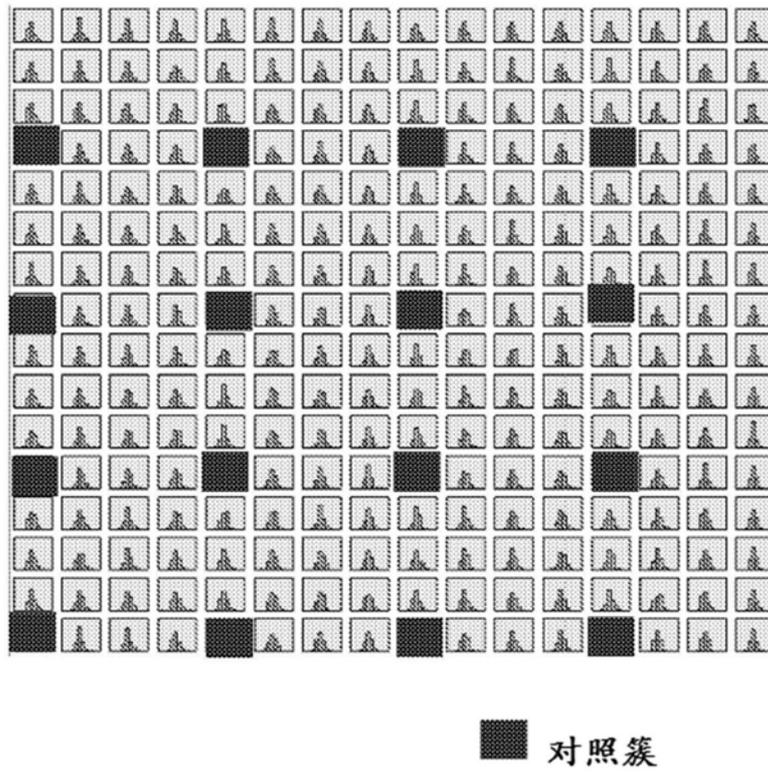


图18B

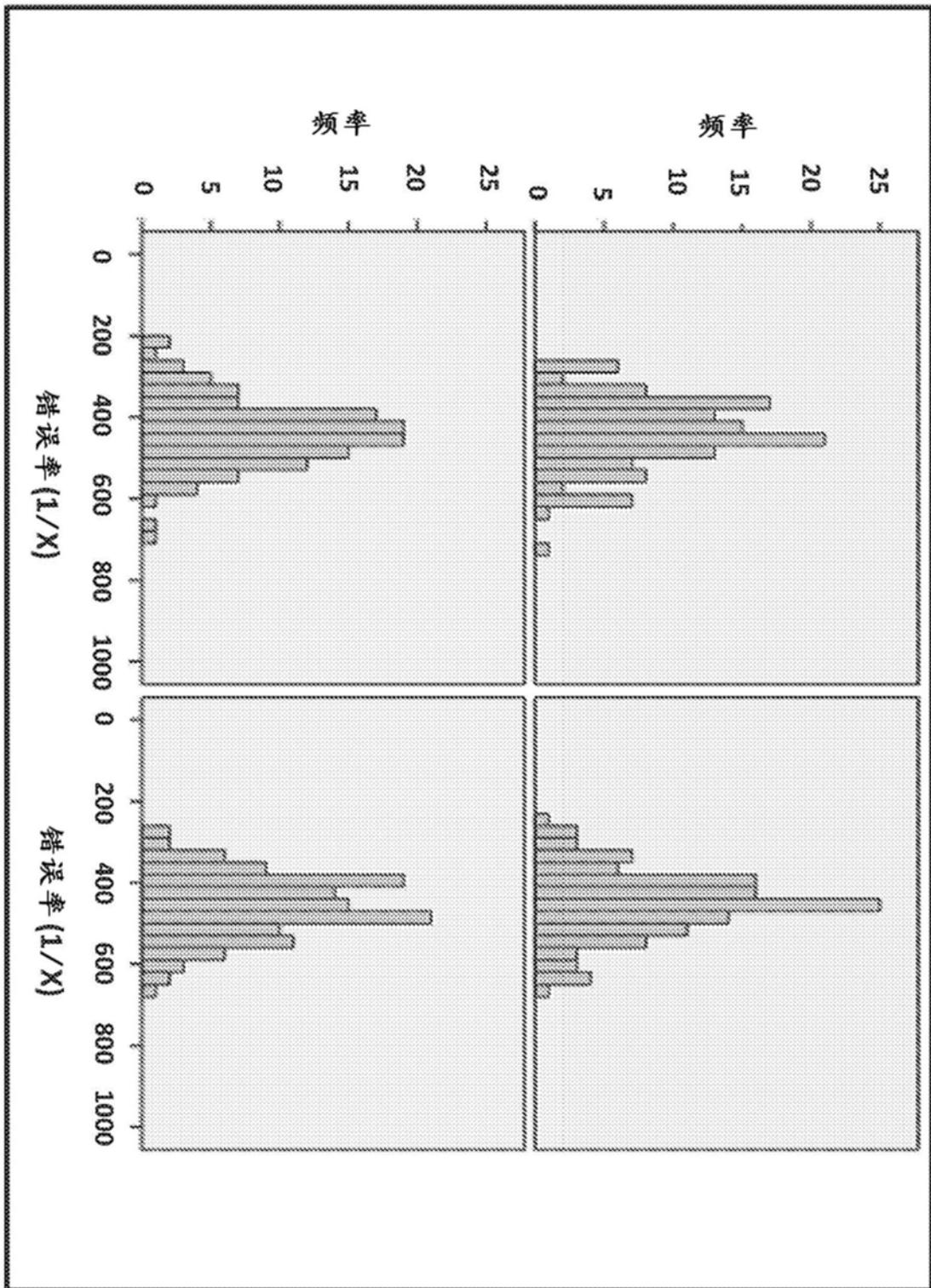


图19

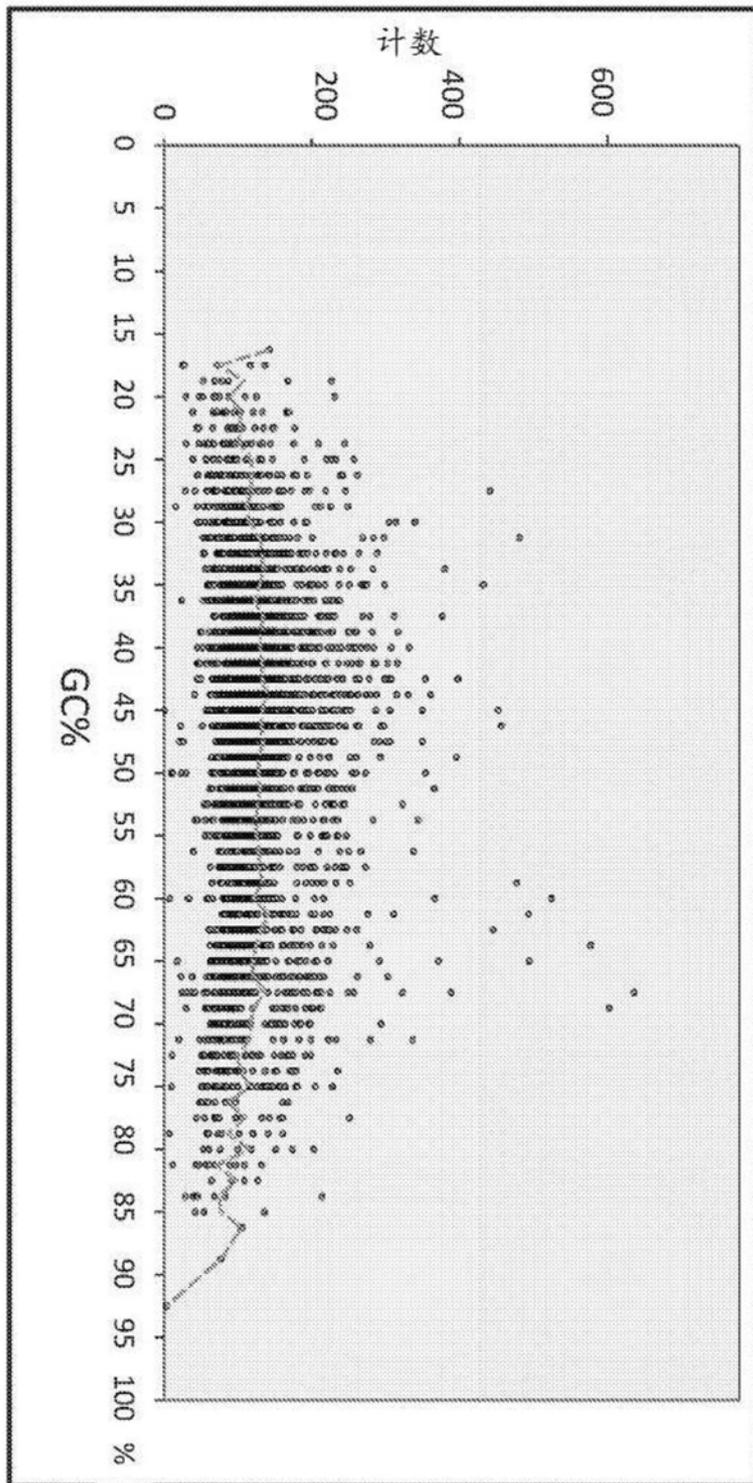


图20

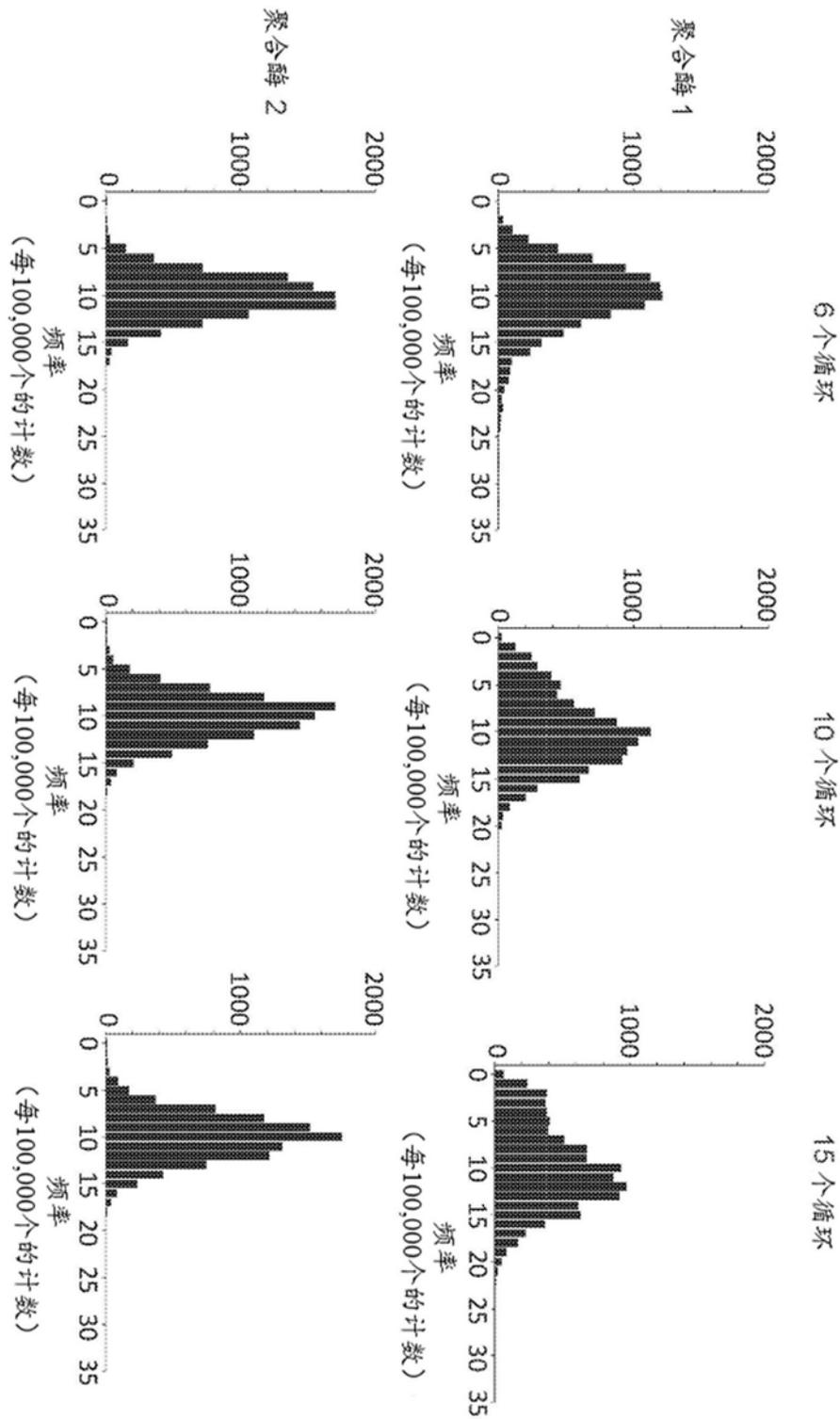


图21

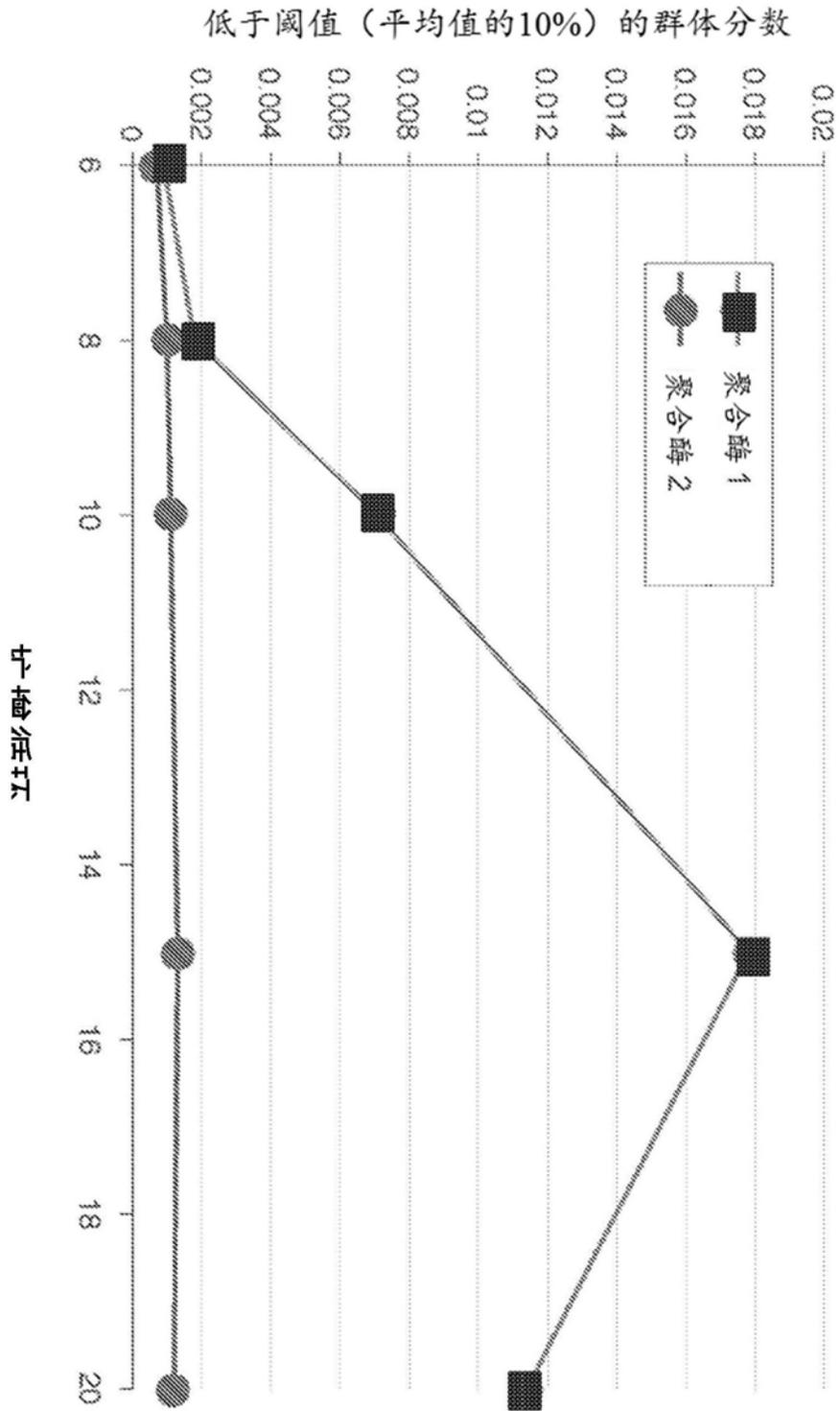


图22

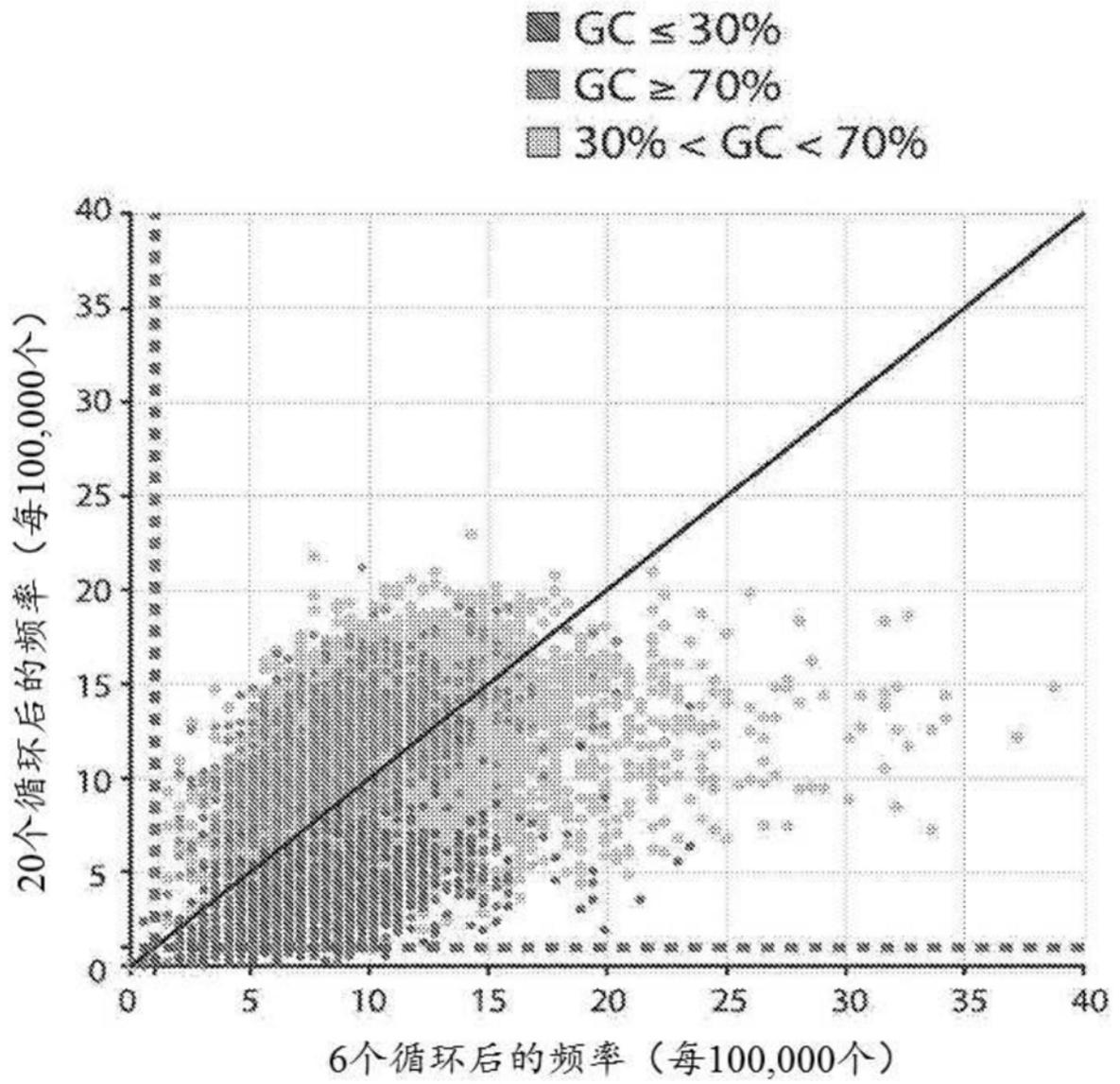


图23

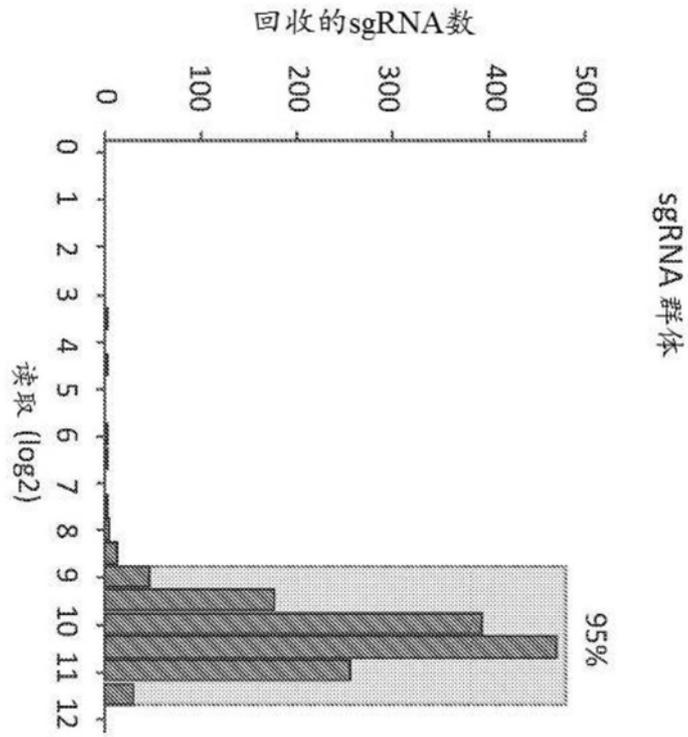


图24A

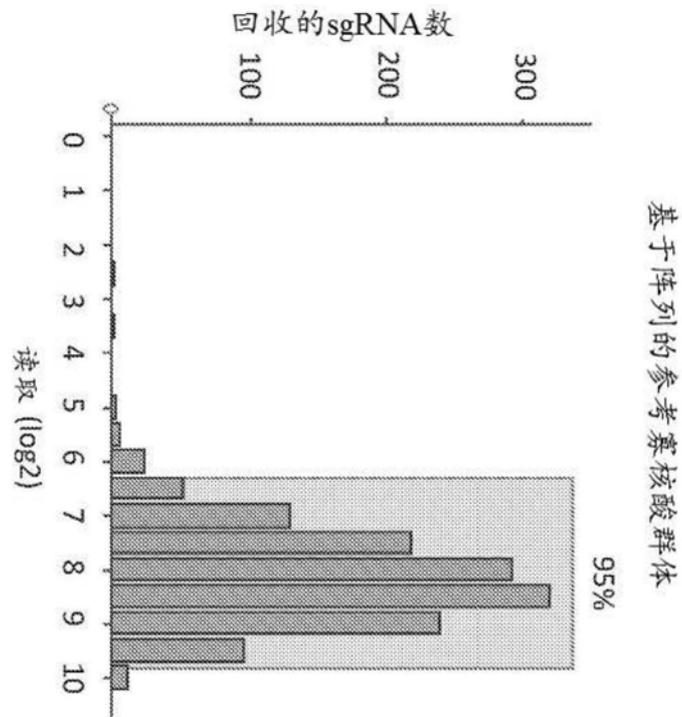


图24B

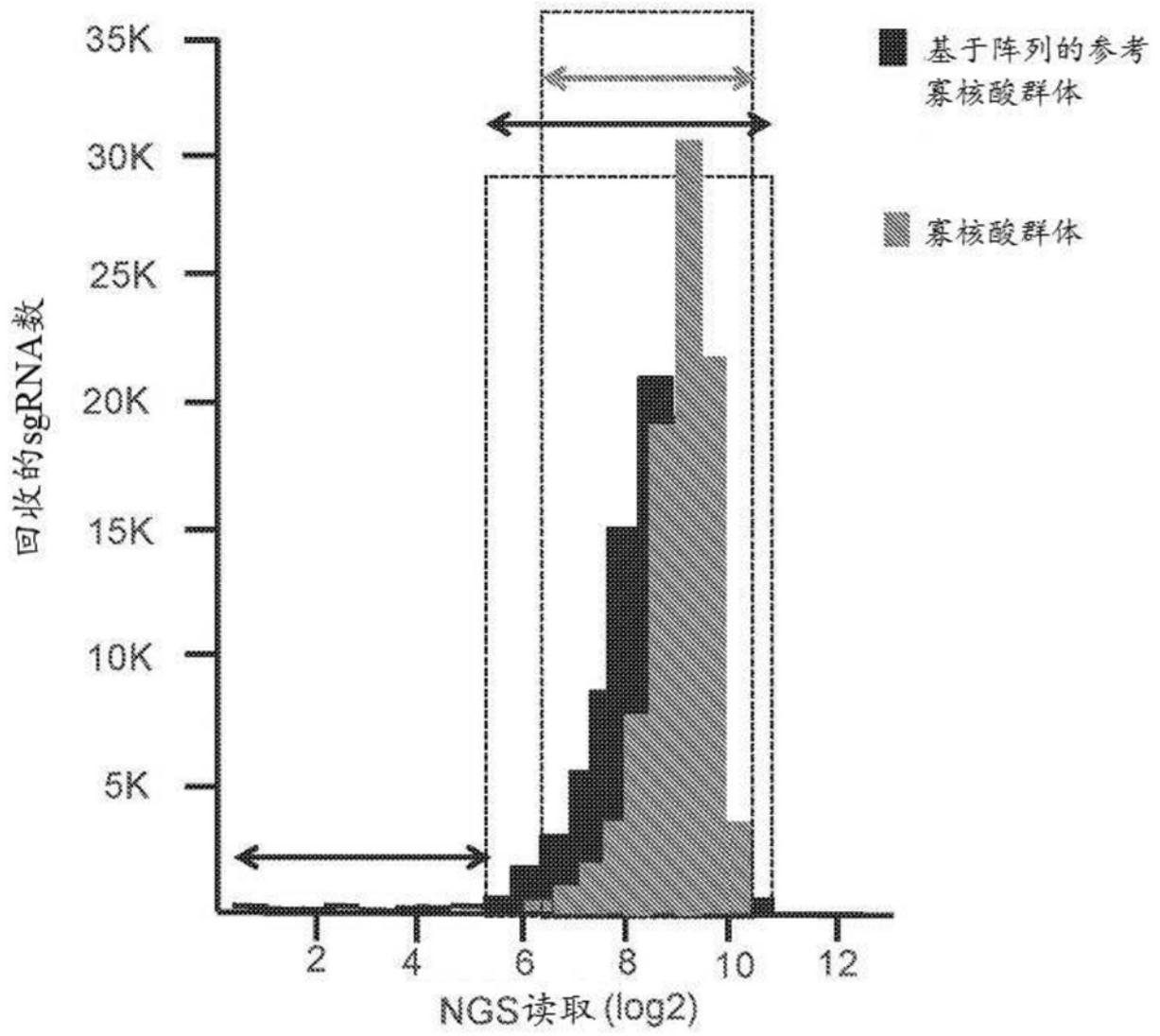


图25

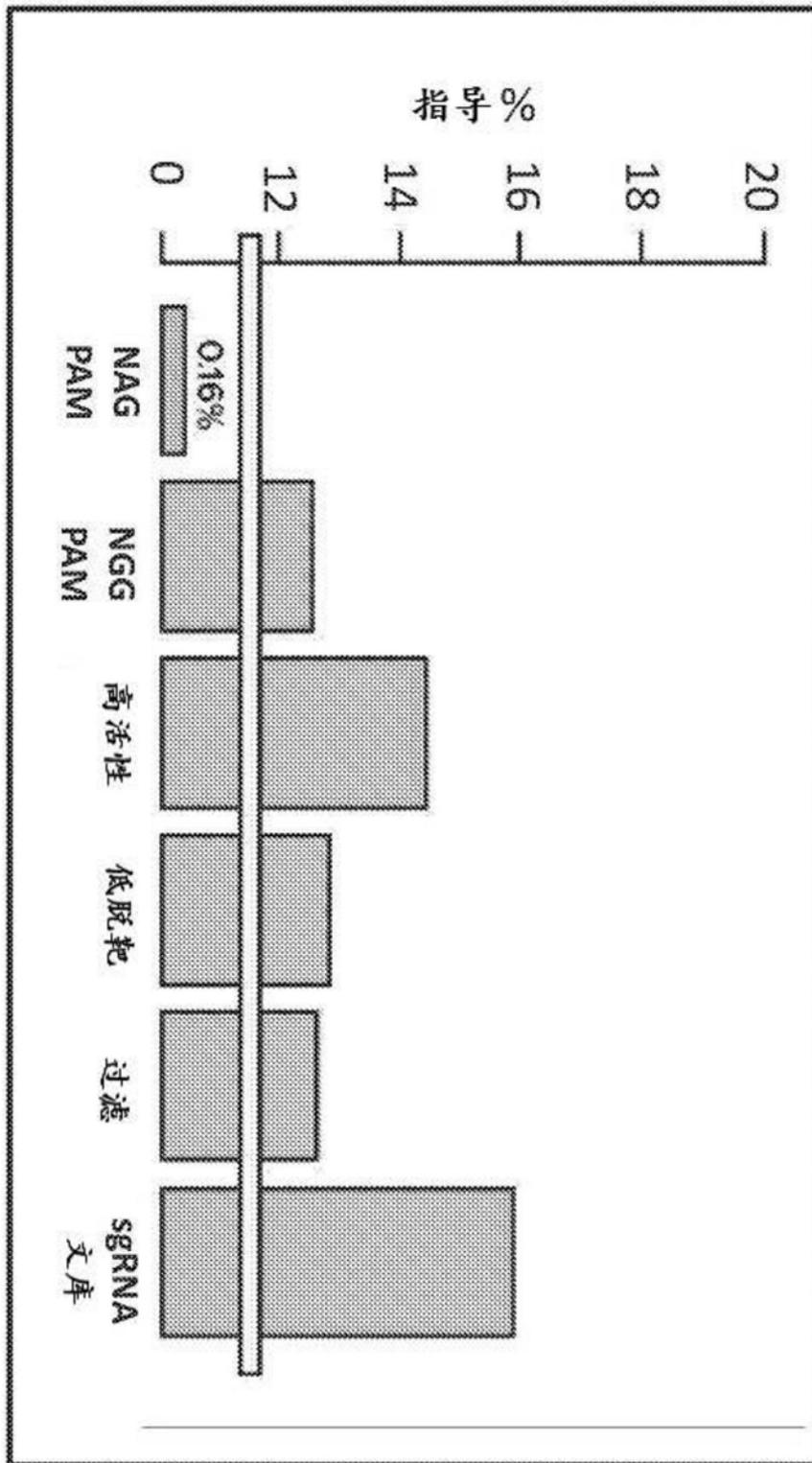


图26A

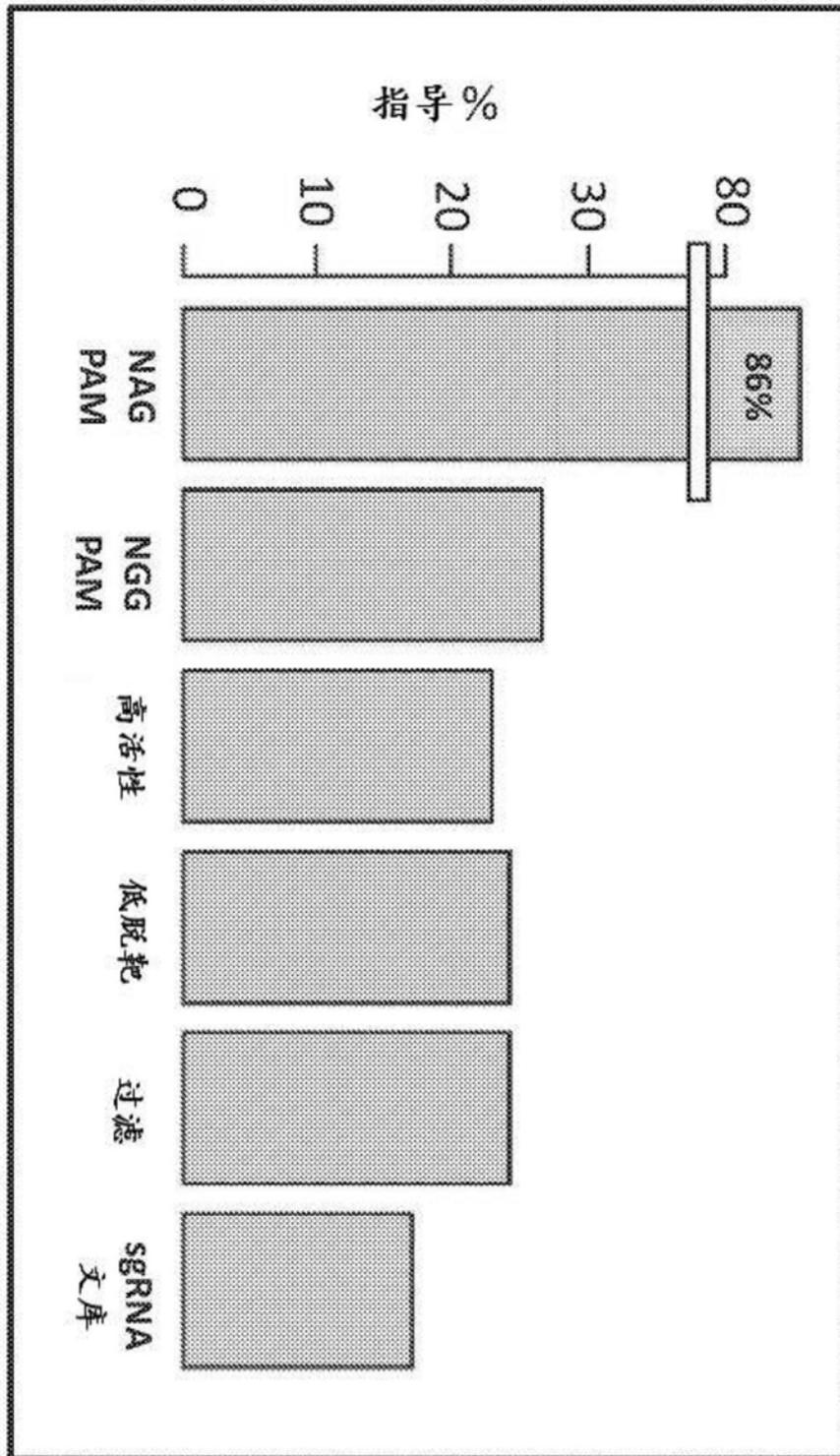


图26B