



(12) 发明专利申请

(10) 申请公布号 CN 112969789 A

(43) 申请公布日 2021.06.15

(21) 申请号 201980073850.5

(22) 申请日 2019.11.07

(30) 优先权数据

62/757,757 2018.11.08 US

(85) PCT国际申请进入国家阶段日

2021.05.08

(86) PCT国际申请的申请数据

PCT/US2019/060243 2019.11.07

(87) PCT国际申请的公布数据

W02020/097315 EN 2020.05.14

(71) 申请人 贝克顿迪金森公司

地址 美国新泽西州

(72) 发明人 艾琳·夏姆

(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 刘晓杰 武晶晶

(51) Int.Cl.

C12N 15/10 (2006.01)

C12Q 1/6806 (2006.01)

C12Q 1/6804 (2006.01)

C12Q 1/6869 (2006.01)

权利要求书5页 说明书91页

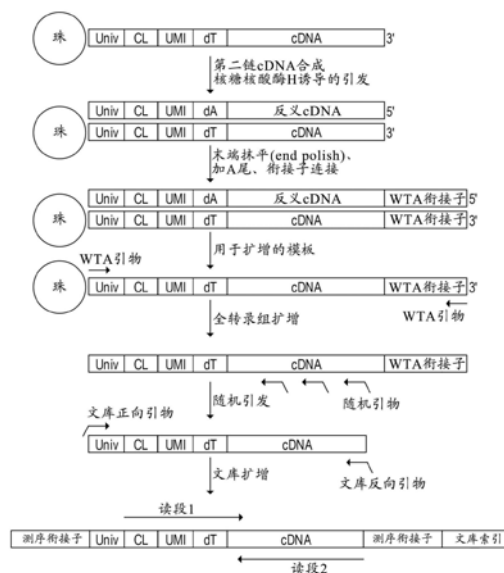
序列表3页 附图63页

(54) 发明名称

使用随机引发的单细胞全转录组分析

(57) 摘要

本文的公开内容包括用于使用随机引发和延伸(random priming and extension,RPE)进行全转录组分析(whole transcriptome analysis, WTA)的系统、方法、组合物和试剂盒。基于RPE的WTA方法可以包括使随机引物与多于一种与固体支持物关联的第一链条形码化多核苷酸杂交,并使随机引物延伸以产生多于一种延伸产物。该方法可以包括扩增多于一种延伸产物以产生测序文库。



1. 一种对样品中的核酸靶进行标记的方法,所述方法包括:

使核酸靶的拷贝与多于一种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与所述核酸靶的拷贝杂交的靶结合区;

使与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸;

使随机引物与所述多于一种第一链条形码化多核苷酸接触,其中每种所述随机引物包含第二通用序列或其互补体;以及

使与所述多于一种第一链条形码化多核苷酸杂交的所述随机引物延伸以产生多于一种延伸产物。

2. 根据权利要求1所述的方法,所述方法包括使用能够与所述第一通用序列或其互补体杂交的引物和能够与所述第二通用序列或其互补体杂交的引物扩增所述多于一种延伸产物,从而产生第一多于一种条形码化扩增子。

3. 根据权利要求2所述的方法,其中扩增所述多于一种延伸产物包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到所述多于一种延伸产物。

4. 根据权利要求2-3中任一项所述的方法,所述方法包括基于与所述第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量,确定所述样品中所述核酸靶的拷贝数。

5. 一种用于确定样品中核酸靶的数量的方法,所述方法包括:

使核酸靶的拷贝与多于一种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与所述核酸靶的拷贝杂交的靶结合区;

使与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸;

使随机引物与所述多于一种第一链条形码化多核苷酸接触,其中每种所述随机引物包含第二通用序列或其互补体;

使与所述多于一种第一链条形码化多核苷酸杂交的所述随机引物延伸以产生多于一种延伸产物;

使用能够与所述第一通用序列或其互补体杂交的引物和能够与所述第二通用序列或其互补体杂交的引物扩增所述多于一种延伸产物,从而产生第一多于一种条形码化扩增子;以及

基于与所述第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量,确定所述样品中所述核酸靶的拷贝数。

6. 根据权利要求4-5中任一项所述的方法,

其中使核酸靶的拷贝接触包括使多于一种核酸靶的拷贝与多于一种寡核苷酸条形码接触,

其中使所述多于一种寡核苷酸条形码延伸包括使与所述多于一种核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸,以及

其中确定所述样品中所述核酸靶的拷贝数包括基于与所述第一多于一种条形码化扩增子中的条形码化扩增子关联的具有不同序列的所述分子标记的数量来确定所述样品中所述多于一种核酸靶中的每一种的数量,所述第一多于一种条形码化扩增子包括所述多于

一种核酸靶中的每一种的序列。

7. 根据权利要求6所述的方法, 其中所述多于一种核酸靶中的每一种的序列包括所述多于一种核酸靶中的每一种的子序列。

8. 根据权利要求4-7中任一项所述的方法, 其中所述第一多于一种条形码化扩增子中的所述核酸靶的序列包括所述核酸靶的子序列。

9. 根据权利要求1-8中任一项所述的方法, 所述方法包括使用能够与所述第一通用序列或其互补体杂交的引物和能够与所述第二通用序列或其互补体杂交的引物扩增所述第一多于一种条形码化扩增子, 从而产生第二多于一种条形码化扩增子。

10. 根据权利要求9所述的方法, 其中扩增所述第一多于一种条形码化扩增子包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到所述第一多于一种条形码化扩增子。

11. 根据权利要求9-10中任一项所述的方法, 所述方法包括基于与所述第二多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量, 确定所述样品中所述核酸靶的拷贝数。

12. 根据权利要求9-11中任一项所述的方法,

其中使核酸靶的拷贝接触包括使多于一种核酸靶的拷贝与多于一种寡核苷酸条形码接触,

其中使所述多于一种寡核苷酸条形码延伸包括使与所述多于一种核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸, 和/或

其中确定所述样品中所述核酸靶的拷贝数包括基于与所述第二多于一种条形码化扩增子中的条形码化扩增子关联的具有不同序列的所述分子标记的数量来确定所述样品中所述多于一种核酸靶中的每一种的数量, 所述第二多于一种条形码化扩增子包括所述多于一种核酸靶中的每一种的序列。

13. 根据权利要求12所述的方法, 其中所述多于一种核酸靶中的每一种的序列包括所述多于一种核酸靶中的每一种的子序列。

14. 根据权利要求9-13中任一项所述的方法, 其中所述第二多于一种条形码化扩增子中的所述核酸靶的序列包括所述核酸靶的子序列。

15. 根据权利要求2-14中任一项的方法, 其中所述第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子中的每一种包含以下的至少一部分: 所述第一通用序列、所述第二通用序列或二者。

16. 根据权利要求1-15中任一项所述的方法, 其中所述第一通用序列和所述第二通用序列是相同的。

17. 根据权利要求1-15中任一项所述的方法, 其中所述第一通用序列和所述第二通用序列是不同的。

18. 根据权利要求1-17中任一项所述的方法, 其中所述第一通用序列和/或所述第二通用序列包含测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点。

19. 根据权利要求18所述的方法, 其中所述测序衔接子包括P5序列、P7序列、其互补序列和/或其部分。

20. 根据权利要求18所述的方法, 其中所述测序引物包括读段1测序引物、读段2测序引

物、其互补序列和/或其部分。

21. 根据权利要求1-20中任一项所述的方法,其中使与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸包括使用逆转录酶使与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸。

22. 根据权利要求21所述的方法,其中所述逆转录酶包括病毒逆转录酶。

23. 根据权利要求1-22中任一项所述的方法,其中使与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码延伸。

24. 根据权利要求1-22中任一项所述的方法,其中使与所述多于一种第一链条形码化多核苷酸杂交的所述随机引物延伸包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使与所述多于一种第一链条形码化多核苷酸杂交的所述随机引物延伸。

25. 根据权利要求22-24中任一项所述的方法,其中所述DNA聚合酶包括Klenow片段。

26. 根据权利要求2-25中任一项所述的方法,其中所述第一多于一种条形码化扩增子和/或所述第二多于一种条形码化扩增子包含全转录组扩增(WTA)产物。

27. 根据权利要求2-26中任一项所述的方法,其中所述第一多于一种条形码化扩增子和/或所述第二多于一种条形码化扩增子对应于单细胞的mRNA的至少10%。

28. 根据权利要求2-26中任一项所述的方法,其中所述第一多于一种条形码化扩增子和/或所述第二多于一种条形码化扩增子对应于单细胞的mRNA的至少50%或至少90%。

29. 根据权利要求1-28中任一项所述的方法,其中所述多于一种核酸靶中的一种或更多种包括低表达基因的mRNA。

30. 根据权利要求1-29中任一项所述的方法,其中所述随机引物包含随机核苷酸序列,任选地,所述随机核苷酸序列的长度为约4个至约30个核苷酸,并且还任选地,所述随机核苷酸序列的长度为6个或9个核苷酸。

31. 根据权利要求1-30中任一项所述的方法,其中所述靶结合区包含寡聚dT序列、随机序列、靶特异性序列,或它们的组合。

32. 根据权利要求1-31中任一项所述的方法,其中所述靶结合区包含多(dT)区,并且其中所述核酸靶包含多(dA)区。

33. 根据权利要求1-32中任一项所述的方法,所述方法包括重复以下步骤:使随机引物与所述多于一种第一链条形码化多核苷酸接触,使与所述多于一种第一链条形码化多核苷酸杂交的所述随机引物延伸,以及扩增所述多于一种延伸产物。

34. 根据权利要求1-33中任一项所述的方法,所述方法包括使用所述多于一种第一链条形码化多核苷酸作为模板合成第三多于一种条形码化扩增子以产生第三多于一种条形码化扩增子。

35. 根据权利要求34所述的方法,其中合成第三多于一种条形码化扩增子包括对所述多于一种第一链条形码化多核苷酸进行聚合酶链式反应(PCR)扩增。

36. 根据权利要求34-35中任一项所述的方法,其中合成第三多于一种条形码化扩增子包括使用能够与所述第一通用序列或其互补体杂交的引物和使用靶特异性引物进行的PCR

扩增。

37. 根据权利要求34-36中任一项所述的方法,所述方法包括获得所述第三多于一种条形码化扩增子或其产物的序列信息,并且任选地获得所述序列信息包括将测序衔接子附接到所述第三多于一种条形码化扩增子或其产物。

38. 根据权利要求36-37中任一项所述的方法,其中所述靶特异性引物与免疫受体特异性地杂交,任选地所述免疫受体是T细胞受体(TCR)和/或B细胞受体(BCR)。

39. 根据权利要求1-38中任一项所述的方法,其中所述样品包括多于一个细胞、多于一个单细胞、组织、肿瘤样品,或它们的任何组合。

40. 根据权利要求1-39中任一项所述的方法,其中所述样品包括外周血单核细胞或免疫细胞,并且任选地所述免疫细胞包括B细胞、T细胞或它们的组合。

41. 根据权利要求2-40中任一项所述的方法,其中扩增所述多于一种延伸产物不在固体支持物的存在下进行。

42. 根据权利要求1-41中任一项所述的方法,其中所述方法不包括核糖核酸酶H诱导的引发、末端修复和/或衔接子连接。

43. 根据权利要求1-42中任一项所述的方法,其中所述方法不包括片段化、标签片段化或二者。

44. 根据权利要求1-43中任一项所述的方法,其中所述第一链条形码化多核苷酸包括条形码化脱氧核糖核酸(DNA)分子、条形码化核糖核酸(RNA)分子或二者。

45. 根据权利要求1-44中任一项所述的方法,其中所述核酸靶包括核酸分子,并且任选地所述核酸分子包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、包含多(A)尾的RNA,或它们的任何组合。

46. 根据权利要求45所述的方法,其中所述mRNA编码免疫受体。

47. 根据权利要求45-46中任一项所述的方法,其中所述核酸靶包含细胞组分结合试剂。

48. 根据权利要求45所述的方法,其中所述核酸分子与所述细胞组分结合试剂关联。

49. 根据权利要求48所述的方法,所述方法包括使所述核酸分子和所述细胞组分结合试剂解离。

50. 根据权利要求1-49中任一项所述的方法,其中所述多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列。

51. 根据权利要求1-50中任一项所述的方法,其中所述多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。

52. 根据权利要求1-51中任一项所述的方法,其中所述多于一个寡核苷酸条形码与固体支持物关联。

53. 根据权利要求52所述的方法,其中与相同固体支持物关联的所述多于一种寡核苷酸条形码各自包含相同的样品标记,并且其中所述多于一种寡核苷酸条形码的每种样品标记包含至少6个核苷酸。

54. 根据权利要求1-53中任一项所述的方法,其中所述多于一种寡核苷酸条形码各自包含细胞标记,并且任选地,所述多于一种寡核苷酸条形码的每种细胞标记包含至少6个核苷酸。

55. 根据权利要求54所述的方法,其中与相同的固体支持物关联的寡核苷酸条形码包含相同的细胞标记。

56. 根据权利要求54-55中任一项所述的方法,其中与不同的固体支持物关联的寡核苷酸条形码包含不同的细胞标记。

57. 根据权利要求52-56中任一项所述的方法,其中所述固体支持物包括合成颗粒、平坦表面,或它们的组合。

58. 根据权利要求1-57中任一项所述的方法,其中所述样品包括单细胞,所述方法包括使包含所述多于一种寡核苷酸条形码的合成颗粒与所述样品中的所述单细胞关联。

59. 根据权利要求58所述的方法,所述方法包括在使所述合成颗粒与所述单细胞关联后裂解所述单细胞,并且任选地裂解所述单细胞包括加热所述样品、使所述样品与去污剂接触、改变所述样品的pH,或它们的任何组合。

60. 根据权利要求58-59中任一项所述的方法,其中所述合成颗粒和所述单细胞在相同的分区中,并且任选地所述分区是孔或液滴。

61. 根据权利要求58-60中任一项所述的方法,其中所述多于一种寡核苷酸条形码中的至少一种被固定或部分地固定在所述合成颗粒上。

62. 根据权利要求58-61中任一项所述的方法,其中所述多于一种寡核苷酸条形码中的至少一种被包封或部分地包封在所述合成颗粒中。

63. 根据权利要求58-62中任一项所述的方法,其中所述合成颗粒是可破坏的,并且任选地所述合成颗粒是可破坏的水凝胶颗粒。

64. 根据权利要求58-63中任一项所述的方法,其中所述合成颗粒包括珠,并且任选地所述珠包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、水凝胶珠、凝胶珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。

65. 根据权利要求58-64中任一项所述的方法,其中所述合成颗粒包含选自以下组成的组的材料:聚二甲基硅氧烷(PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。

66. 根据权利要求58-65中任一项所述的方法,

其中所述多于一种寡核苷酸条形码中的每一种包含接头官能团,

其中所述合成颗粒包含固体支持物官能团,和/或

其中所述支持物官能团和所述接头官能团彼此关联。

67. 根据权利要求66所述的方法,其中所述接头官能团和所述支持物官能团单独地选自以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮,以及它们的任何组合。

使用随机引发的单细胞全转录组分析

[0001] 相关申请

[0002] 本申请根据35 U.S.C. §119(e), 要求2018年11月8日提交的美国临时专利申请序列第62/757,757号的权益, 该相关申请的内容出于所有目的通过引用以其整体并入本文。

[0003] 对序列表的引用

[0004] 本申请连同电子格式的序列表一起提交。序列表被提供为题名为68EB_298705_W0的文件, 创建于2019年11月7日, 大小是4千字节。电子格式的序列表的信息通过引用以其整体并入本文。

[0005] 背景

[0006] 领域

[0007] 本公开内容总体上涉及分子生物学领域, 例如使用分子条形码化(molecular barcoding)确定细胞的基因表达谱。

[0008] 对现有技术的描述

[0009] 当前的技术允许在每个细胞与隔室(compartment)中的条形码化试剂珠共定位时通过将细胞特异性寡核苷酸条形码附接到来自个体细胞的多(A) mRNA分子而以大规模平行的方式(例如, >10000个细胞)测量单细胞的基因表达。对能够有效地定量分析细胞的基因表达的系统和方法存在需求。

[0010] 概述

[0011] 本文的公开内容包括对样品中的核酸靶进行标记的方法。在一些实施方案中, 方法包括: 使核酸靶的拷贝与多于一种寡核苷酸条形码接触, 其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区; 使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸; 使随机引物与多于一种第一链条形码化多核苷酸接触, 其中每种随机引物包含第二通用序列或其互补体; 以及使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸以产生多于一种延伸产物。

[0012] 在一些实施方案中, 方法包括: 使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增多于一种延伸产物, 从而产生第一多于一种条形码化扩增子。在一些实施方案中, 扩增多于一种延伸产物包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到多于一种延伸产物。在一些实施方案中, 方法包括: 基于与第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量, 确定样品中核酸靶的拷贝数。

[0013] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中, 方法包括: 使核酸靶的拷贝与多于一种寡核苷酸条形码接触, 其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区; 使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸; 使随机引物与多于一种第一链条形码化多核苷酸接触, 其中每种随机引物包含第二通用序列或其互补体; 使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸以产生多于一种延伸产物; 使

用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增多于一种延伸产物,从而产生第一多于一种条形码化扩增子;以及基于与第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量,确定样品中核酸靶的拷贝数。

[0014] 在一些实施方案中,使核酸靶的拷贝接触包括使多于一种核酸靶的拷贝与多于一种寡核苷酸条形码接触,使多于一种寡核苷酸条形码延伸包括使与多于一种核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸,并且确定样品中核酸靶的拷贝数包括基于与第一多于一种条形码化扩增子中的条形码化扩增子关联的具有不同序列的分子标记的数量来确定样品中多于一种核酸靶中的每一种的数量,所述第一多于一种条形码化扩增子包括多于一种核酸靶中的每一种的序列。在一些实施方案中,多于一种核酸靶中的每一种的序列包括多于一种核酸靶中的每一种的子序列。在一些实施方案中,第一多于一种条形码化扩增子中的核酸靶的序列包括核酸靶的子序列。

[0015] 在一些实施方案中,方法包括使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增第一多于一种条形码化扩增子,从而产生第二多于一种条形码化扩增子。在一些实施方案中,扩增第一多于一种条形码化扩增子包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到第一多于一种条形码化扩增子。在一些实施方案中,方法包括基于与第二多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量,确定样品中核酸靶的拷贝数。在一些实施方案中,使核酸靶的拷贝接触包括使多于一种核酸靶的拷贝与多于一种寡核苷酸条形码接触,使多于一种寡核苷酸条形码延伸包括使与多于一种核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸,和/或其中确定样品中核酸靶的拷贝数包括基于与第二多于一种条形码化扩增子中的条形码化扩增子关联的具有不同序列的分子标记的数量来确定样品中多于一种核酸靶中的每一种的数量,所述第二多于一种条形码化扩增子包括多于一种核酸靶中的每一种的序列。在一些实施方案中,多于一种核酸靶中的每一种的序列包括多于一种核酸靶中的每一种的子序列。在一些实施方案中,第二多于一种条形码化扩增子中的核酸靶的序列包括核酸靶的子序列。

[0016] 在一些实施方案中,第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子中的每一种包含以下的至少一部分:第一通用序列、第二通用序列或二者。在一些实施方案中,第一通用序列和第二通用序列是相同的。在一些实施方案中,第一通用序列和第二通用序列是不同的。在一些实施方案中,第一通用序列和/或第二通用序列包含测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点。在一些实施方案中,测序衔接子包括P5序列、P7序列、其互补序列和/或其部分。在一些实施方案中,测序引物包括读段1测序引物、读段2测序引物、其互补序列和/或其部分。

[0017] 在一些实施方案中,使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸包括使用逆转录酶使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸。在一些实施方案中,逆转录酶包括病毒逆转录酶。在一些实施方案中,使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸。在一些实施方案中,使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸包括使用缺乏5'至3'

核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸。在一些实施方案中,DNA聚合酶包含Klenow片段。

[0018] 在一些实施方案中,第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子包含全转录组扩增(WTA)产物。在一些实施方案中,第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子对应于单细胞的mRNA的至少10%。在一些实施方案中,第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子对应于单细胞的mRNA的至少50%或至少90%。在一些实施方案中,多于一种核酸靶中的一种或更多种包括低表达基因的mRNA。在一些实施方案中,所述随机引物包含随机核苷酸序列,任选地,随机核苷酸序列的长度为约4个至约30个核苷酸,并且还任选地,所述随机核苷酸序列的长度为6个或9个核苷酸。在一些实施方案中,靶结合区包含寡聚dT序列、随机序列、靶特异性序列或它们的组合。在一些实施方案中,靶结合区包含多(dT)区,并且其中核酸靶包含多(dA)区。在一些实施方案中,方法包括重复以下步骤:使随机引物与多于一种第一链条形码化多核苷酸接触,使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸,以及扩增多于一种延伸产物。

[0019] 在一些实施方案中,方法包括:使用多于一种第一链条形码化多核苷酸作为模板合成第三多于一种条形码化扩增子以产生第三多于一种条形码化扩增子。在一些实施方案中,合成第三多于一种条形码化扩增子包括对多于一种第一链条形码化多核苷酸进行聚合酶链式反应(PCR)扩增。在一些实施方案中,合成第三多于一种条形码化扩增子包括使用能够与第一通用序列或其互补体杂交的引物和使用靶特异性引物进行的PCR扩增。在一些实施方案中,方法包括:获得第三多于一种条形码化扩增子或其产物的序列信息,并且任选地,获得序列信息包括将测序衔接子附接到第三多于一种条形码化扩增子或其产物。在一些实施方案中,靶特异性引物与免疫受体特异性地杂交,任选地,免疫受体是T细胞受体(TCR)和/或B细胞受体(BCR)。

[0020] 在一些实施方案中,扩增多于一种延伸产物不在固体支持物的存在下进行。在一些实施方案中,方法不包括核糖核酸酶H诱导的引发、末端修复和/或衔接子连接。在一些实施方案中,方法不包括片段化、标签片段化(tagmentation)或二者。在一些实施方案中,第一链条形码化多核苷酸包括条形码化脱氧核糖核酸(DNA)分子、条形码化核糖核酸(RNA)分子或二者。

[0021] 在一些实施方案中,样品包括多于一个细胞、多于一个单细胞、组织、肿瘤样品或它们的任何组合。在一些实施方案中,样品包括外周血单核细胞或免疫细胞,并且任选地,免疫细胞包括B细胞、T细胞或它们的组合。在一些实施方案中,核酸靶包括核酸分子,并且任选地,核酸分子包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、包含多(A)尾的RNA,或它们的任何组合。在一些实施方案中,mRNA编码免疫受体。

[0022] 在一些实施方案中,核酸靶包含细胞组分结合试剂。在一些实施方案中,核酸分子与细胞组分结合试剂关联。在一些实施方案中,方法包括:使核酸分子和细胞组分结合试剂解离。在一些实施方案中,多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列。在一些实施方案中,多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。

[0023] 在一些实施方案中,多于一种寡核苷酸条形码与固体支持物关联。在一些实施方案中,与相同固体支持物关联的多于一种寡核苷酸条形码各自包含相同的样品标记,并且

其中多于一种寡核苷酸条形码的每种样品标记包含至少6个核苷酸。在一些实施方案中,多于一种寡核苷酸条形码各自包含细胞标记,并且任选地,所述多于一种寡核苷酸条形码的每种细胞标记包含至少6个核苷酸。在一些实施方案中,与相同的固体支持物关联的寡核苷酸条形码包含相同的细胞标记。在一些实施方案中,与不同的固体支持物关联的寡核苷酸条形码包含不同的细胞标记。在一些实施方案中,固体支持物包括合成颗粒、平坦表面,或它们的组合。

[0024] 在一些实施方案中,样品包括单细胞,方法包括将包含多于一种寡核苷酸条形码的合成颗粒与样品中的单细胞关联。在一些实施方案中,方法包括:在将合成颗粒与单细胞关联后裂解单细胞,并且任选地裂解单细胞包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。

[0025] 在一些实施方案中,合成颗粒和单细胞在相同的分区(partition)中,并且任选地,该分区是孔或液滴。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被固定或部分地固定在合成颗粒上。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被包封或部分地包封在合成颗粒中。在一些实施方案中,合成颗粒是可破坏的,并且任选地,合成颗粒是可破坏的水凝胶颗粒。

[0026] 在一些实施方案中,合成颗粒包括珠,并且任选地珠包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、水凝胶珠、凝胶珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。在一些实施方案中,合成颗粒包含选自以下组成的组的材料:聚二甲基硅氧烷(PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。在一些实施方案中,多于一种寡核苷酸条形码中的每一种包含接头官能团,合成颗粒包含固体支持物官能团,和/或支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮,以及它们的任何组合。

[0027] 附图简述

[0028] 图1图示了非限制性示例性随机条形码(stochastic barcode)。

[0029] 图2示出了随机条形码化和数字计数的非限制性示例性工作流程。

[0030] 图3是示出了用于从多于一种靶产生随机条形码化靶的索引化文库(indexed library)的非限制性示例性方法的示意性图示。

[0031] 图4示出了与寡核苷酸缀合的示例性蛋白质结合试剂(此处图示的抗体)的示意性图示,所述寡核苷酸包含用于蛋白质结合试剂的独特标识符。

[0032] 图5示出了与寡核苷酸缀合的示例性结合试剂(此处图示的抗体)的示意性图示,所述寡核苷酸包含用于样品索引化(sample indexing)以确定来自相同或不同样品的细胞的独特标识符。

[0033] 图6示出了使用寡核苷酸缀合的抗体以高通量方式同时确定蛋白质表达和基因表达的示例性工作流程的示意性图示。

[0034] 图7示出了使用寡核苷酸缀合的抗体进行样品索引化的示例性工作流程的示意性图示。

[0035] 图8是对单细胞进行全转录组分析(whole transcriptome analysis, WTA)的非限制性示例性工作流程的示意性图示。

[0036] 图9是对单细胞进行基于连接的全转录组分析的非限制性示例性工作流程的示意性图示。

[0037] 图10是使用随机引发和引物延伸(random priming and primer extension)有效地(例如,在比其他全转录组分析方法更短的时间内)对单细胞进行全转录组分析的非限制性示例性工作流程的示意性图示。

[0038] 图11A是使用随机引发和引物延伸进行全转录组分析的非限制性示例性工作流程的示意性图示。

[0039] 图11B是使用随机引发和引物延伸进行全转录组分析的非限制性示例性方法。

[0040] 图12是以一个工作流程进行靶向基因表达谱分析(profiling)、用Ab0进行样品多重化(sample multiplexing)、用Ab0进行蛋白质表达谱分析,以及全转录组分析的示意性图示。

[0041] 图13A-图13B是示出了进行基于连接的全转录组分析和用使用6mer随机多聚体(randomer)和9mer随机多聚体的随机引发和延伸进行的全转录组分析的非示例性灵敏度度量的图。

[0042] 图14是示出了与基于连接的WTA方法相比,本公开内容的基于随机引发和延伸的WTA分析的示例性性能的t-分布式随机邻域嵌入(t-Distributed Stochastic Neighbor Embedding, t-SNE)。

[0043] 图15A-图15B是示出了ID010NK6(RPE WTA九聚体)表现为比ID010UG1(基于连接的WTA方法)捕获更多的细胞表面标志物的非限制性示例性饼图。

[0044] 图16A-图16D是描绘了RPE WTA方法在Jurkat/Ramos细胞和健康供体外周血单核细胞(PBMC)中的性能的非限制性示例性生物分析仪图。

[0045] 图17A-图17D是示出了在用随机引发和延伸(采用包括一个随机引发步骤和一个PCR步骤的RPE方法)进行WTA分析的一些实施方案中具有链置换活性的Klenow Exo-优于T4 DNA聚合酶的非限制性示例性生物分析仪图(RPE PCR)。

[0046] 图18A-图18D是示出了Klenow exo-可能不具有如先前认为的那么多的链置换活性的非限制性示例性生物分析仪图(RPE PCR)。

[0047] 图19A-图19D是示出了随机引发步骤的数量和随机多聚体长度对RPE WTA方法的性能的影响的非限制性示例性生物分析仪图。

[0048] 图20A-图20F是示出了SPRI清理比率(cleanup ratio)对基于RPE的WTA方法的性能的影响的非限制性示例性数据。

[0049] 图21A-图21G是示出了不同的随机引物浓度和RPE条件对基于RPE的WTA方法的性能的影响的非限制性示例性生物分析仪图(RPE PCR, SPRI清理后)。

[0050] 图22A-图22F是示出了添加冷珠和改变珠密度对RPE WTA的性能的影响的非限制性示例性测序数据。

[0051] 图23A-图23B是示出了随机引物浓度和冷珠对RPE WTA的性能的影响的非限制性示例性测序数据。

[0052] 图24A-图24E是示出了不同的细胞数量对RPE WTA方法的影响的非限制性示例性

生物分析仪图 (RPE PCR)。

[0053] 图25A-图25E是示出了不同的细胞数量对RPE WTA方法的性能的影响的非限制性示例性生物分析仪图 (索引PCR)。

[0054] 图26A-图26F描绘了示出本公开内容的基于随机引发和延伸的WTA分析使用不同的细胞数量的示例性性能的t-分布式随机邻域嵌入 (t-SNE) 投影。

[0055] 图27A-图27D是示出了连续RPE变性 (sequential RPE denaturation) 和RPE反应体积对RPE WTA方法的性能的影响的非限制性示例性生物分析仪图。

[0056] 图28A-图28C是示出了本公开内容的基于随机引发和延伸的WTA分析使用2PCR文库产生方案 (图28A-图28B) 和1PCR文库产生方案 (图28C) 的示例性性能的示例性生物分析仪图。

[0057] 图29A-图29D示出了用于同时确定蛋白质表达和基因表达以及用于样品索引化的寡核苷酸的非限制性示例性设计。

[0058] 图30示出了用于同时确定蛋白质表达和基因表达以及用于样品索引化的非限制性示例性寡核苷酸序列的示意性图示。

[0059] 详细描述

[0060] 在以下详细描述中,参考了构成本文的一部分的附图。在附图中,除非上下文另有指示,否则相似的符号通常标识相似的部件。在详细描述、附图和权利要求书中描述的说明性实施方案并不意味着是限制性的。在不脱离本文提出的主题的精神或范围的情况下,可以利用其他实施方案,并且可以做出其他改变。将容易理解的是,如本文一般描述的以及附图中图示的本公开内容的方面能够以各种不同的配置来布置、替换、组合、分离和设计,所有这些都在本文中明确考虑并且构成本公开内容的一部分。

[0061] 本文提及的所有专利、公开的专利申请、其他出版物,和来自GenBank的序列,以及其他数据库关于相关技术通过引用以其整体并入本文。

[0062] 对少量核酸 (例如信使核糖核苷酸 (mRNA) 分子) 进行定量对于确定例如在不同发育阶段或在不同环境条件下在细胞中表达的基因是临床上重要的。然而,确定核酸分子 (例如,mRNA分子) 的绝对数量也是非常具有挑战性的,尤其是当分子数量非常小时。确定样品中分子的绝对数量的一种方法是数字聚合酶链式反应 (PCR)。理想地,PCR在每个循环产生分子的相同拷贝。然而,PCR可具有缺点使得每个分子复制具有随机概率,且此概率根据PCR循环和基因序列而变化,这导致扩增偏倚 (amplification bias) 和不准确的基因表达测量。可以将具有独特分子标记 (也称为分子索引 (MI)) 的随机条形码用于对分子的数量进行计数并且纠正扩增偏倚。诸如PreciseTM测定 (Cellular Research, Inc. (Palo Alto, CA)) 的随机条形码化可以通过使用分子标记 (ML) 在逆转录 (RT) 期间标记mRNA来纠正由PCR和文库制备步骤引起的偏倚。

[0063] PreciseTM测定可利用具有在多 (T) 寡核苷酸上的大量 (例如6561种至65536种) 独特分子标记的随机条形码的非耗尽性池 (non-depleting pool), 以在RT步骤期间与样品中的所有多 (A) -mRNA杂交。随机条形码可包括通用PCR引发位点。在RT期间,靶基因分子与随机条形码随机地反应。每种靶分子可以与随机条形码杂交,从而生成随机地条形码化的互补核糖核苷酸 (cDNA) 分子。在标记后,可将来自微孔板的微孔的随机条形码化cDNA分子汇集到单支管中用于PCR扩增和测序。可以分析原始测序数据以产生读段的数量、具有独特分

子标记的随机条形码的数量以及mRNA分子的数量。

[0064] 用于确定单细胞的mRNA表达谱的方法可以以大规模并行的方式进行。例如，PreciseTM测定可用于同时确定多于10000个细胞的mRNA表达谱。每个样品用于分析的单细胞数量（例如，100个或1000个单细胞）可以低于当前单细胞技术的容量。汇集来自不同样品的细胞能够提高对当前单细胞技术的容量的利用，从而减少试剂浪费和单细胞分析成本。本公开内容提供了用于区分用于cDNA文库制备的不同样品的细胞的样品索引化方法，所述cDNA文库制备用于细胞分析，诸如单细胞分析。汇集来自不同样品的细胞可以使不同样品的细胞的cDNA文库制备的差异最小化，从而实现更准确地比较不同样品。

[0065] 本文的公开内容包括对样品中的核酸靶进行标记的方法。在一些实施方案中，方法包括：使核酸靶的拷贝与多于一种寡核苷酸条形码接触，其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区；使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸；使随机引物与多于一种第一链条形码化多核苷酸接触，其中每种随机引物包含第二通用序列或其互补体；以及使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸以产生多于一种延伸产物。

[0066] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中，方法包括：使核酸靶的拷贝与多于一种寡核苷酸条形码接触，其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区；使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸；使随机引物与多于一种第一链条形码化多核苷酸接触，其中每种随机引物包含第二通用序列或其互补体；使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸以产生多于一种延伸产物；使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增多于一种延伸产物，从而产生第一多于一种条形码化扩增子；以及基于与第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量，确定样品中核酸靶的拷贝数。

[0067] 定义

[0068] 除非另外定义，本文使用的技术术语和科学术语具有与本公开内容所属领域的普通技术人员通常所理解的相同含义。参见，例如，Singleton等人，Dictionary of Microbiology and Molecular Biology，第2版，J.Wiley&Sons，(New York, NY 1994)；Sambrook等人，Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, NY 1989)。出于本公开内容的目的，以下术语在下文定义。

[0069] 如本文使用的，术语“衔接子”可以意指促进关联的核酸的扩增或测序的序列。关联的核酸可包括靶核酸。关联的核酸可以包含以下中的一种或更多种：空间标记、靶标记、样品标记、索引化标记(indexing label)，或条形码序列（例如，分子标记）。衔接子可以是线性的。衔接子可以是预腺苷酸化的衔接子。衔接子可以是双链或单链的。一种或更多种衔接子可以位于核酸的5' 或3' 末端。当衔接子在5' 和3' 末端包含已知序列时，已知序列可以是相同或不同的序列。位于多核苷酸的5' 和/或3' 末端的衔接子能够与固定在表面上的一种或更多种寡核苷酸杂交。在一些实施方案中，衔接子可以包含通用序列。通用序列可以是两种或更多种核酸分子共有的核苷酸序列的区域。两种或更多种核酸分子也可具有不同序

列的区域。因此,例如,5' 衔接子可以包含相同和/或通用核酸序列,并且3' 衔接子可以包含相同和/或通用序列。可以存在于多于一种核酸分子的不同成员中的通用序列可以允许使用与通用序列互补的单种通用引物复制或扩增多种不同序列。相似地,可以存在于核酸分子的集合中的不同成员中的至少一种、两种(例如,一对)或更多种通用序列可以允许使用与通用序列互补的至少一种、两种(例如,一对)或更多种单种通用引物复制或扩增多种不同序列。因此,通用引物包括可以与这样的通用序列杂交的序列。可以修饰具有靶核酸序列的分子以将通用衔接子(例如,非靶核酸序列)附接到不同靶核酸序列的一个末端或两个末端。与靶核酸附接的一种或更多种通用引物可以提供用于通用引物的杂交的位点。附接到靶核酸的一种或更多种通用引物可以彼此相同或不同。

[0070] 如本文使用的,抗体可以是全长(例如,天然存在的或通过正常免疫球蛋白基因片段重组过程形成的)免疫球蛋白分子(例如,IgG抗体)或免疫球蛋白分子的免疫活性(即,特异性结合)部分(如抗体片段)。

[0071] 在一些实施方案中,抗体是功能性抗体片段。例如,抗体片段可以是抗体的一部分,诸如F(ab')₂、Fab'、Fab、Fv、sFv等。抗体片段可以与由全长抗体识别的相同的抗原结合。抗体片段可以包括由抗体的可变区组成的分离的片段,诸如由重链和轻链的可变区组成的“Fv”片段和其中轻链和重链可变区通过肽接头连接的重组单链多肽分子(“scFv蛋白”)。示例性抗体可以包括但不限于癌细胞抗体、病毒抗体、结合至细胞表面受体(例如CD8、CD34和CD45)的抗体,和治疗性抗体。

[0072] 如本文使用的,术语“关联”或“与.....关联”可以意指两个或更多个种类(species)可以被鉴定为在某个时间点处共定位。关联可意指两个或更多个种类在或曾经在相似的容器内。关联可以是信息学关联。例如,关于两个或更多个种类的数字信息可以被存储并且可以用于确定所述种类中的一个或更多个在某个时间点处共定位。关联也可以是物理关联。在一些实施方案中,两个或更多个关联的种类彼此之间或与共同的固体或半固体表面是“连接的”、“附接的”或“固定的”。关联可以指用于将标记附接到固体或半固体支持物(诸如珠)上的共价或非共价方式。关联可以是靶与标记之间的共价键。关联可以包括两个分子(诸如靶分子和标记)之间的杂交。

[0073] 如本文使用的,术语“互补”可以指两个核苷酸之间精确配对的能力。例如,如果核酸的在给定位置的核苷酸能够与另一个核酸的核苷酸以氢键键合,则两个核酸被认为在该位置处是彼此互补的。两个单链核酸分子之间的互补性可以是“部分的”,其中所述核苷酸中仅一些核苷酸结合,或者当所述单链分子之间存在完全互补性时,这种互补性可以是完全的。如果第一核苷酸序列与第二核苷酸序列互补,则可以称第一核苷酸序列是第二序列的“互补体”。如果第一核苷酸序列与第二序列相反的序列(即,核苷酸顺序相反)互补,则可以称第一核苷酸序列是第二序列的“反向互补体”。如本文使用的,术语“互补体”、“互补”和“反向互补体”可以互换使用。从本公开内容可以理解,如果一个分子可以与另一个分子杂交,则其可以是其所杂交的分子的互补体。

[0074] 如本文使用的,术语“数字计数”可以指用于估计样品中靶分子的数量量的方法。数字计数可以包括确定已经与样品中的靶关联的独特标记的数量量的步骤。这种在性质上可以是随机的方法将对分子进行计数的问题从相同分子的定位和鉴定中的一种转化为有关检测一组预定义标记的一系列是/否数字问题。

[0075] 如本文使用的,术语(一种或更多种)“标记”可以指与样品中的靶关联的核酸编码。标记可以是例如核酸标记。标记可以是完全或部分可扩增的标记。标记可以是完全或部分可测序的标记。标记可以是可鉴定为有区别的天然核酸的一部分。标记可以是已知的序列。标记可以包括核酸序列的连接区,例如天然序列和非天然序列的连接区。如本文使用的,术语“标记”可以与术语“索引”、“标签”或“标记-标签”互换使用。标记可以传达信息。例如,在各种实施方案中,可以使用标记来确定样品的身份、样品的来源、细胞的身份,和/或靶。

[0076] 如本文使用的,术语“非耗尽性储库(non-depleting reservoir)”可以指由许多不同标记组成的条形码(例如,随机条形码)的池。非耗尽性储库可以包括大量不同的条形码,使得当非耗尽性储库与靶池关联时,每种靶可能与独特的条形码关联。每标记的靶分子的独特性可以通过随机选择的统计来确定,并且取决于与多样的标记相比在集合中相同的靶分子的拷贝数。所得的经标记的靶分子集合的大小可以通过条形码化过程的随机性质来确定,并且然后对检测到的条形码的数量的分析允许计算原始集合或样品中存在的靶分子的数量。当存在的靶分子的拷贝数量与独特的条形码的数量的比率低时,经标记的靶分子是高度独特的(即,用给定的标记来标记多于一种靶分子的概率非常低)。

[0077] 如本文使用的,术语“核酸”是指多核苷酸序列或其片段。核酸可以包括核苷酸。核酸对于细胞可以是外源的或内源的。核酸可以存在于无细胞环境中。核酸可以是基因或其片段。核酸可以是DNA。核酸可以是RNA。核酸可以包括一种或更多种类似物(例如改变的骨架、糖或核碱基)。类似物的一些非限制性实例包括:5-溴尿嘧啶、肽核酸、异源核酸(xeno nucleic acid)、吗啉代核酸(morpholinos)、锁核酸、二醇核酸、苏糖核酸、二脱氧核苷酸、虫草菌素、7-脱氮-GTP、荧光团(例如,罗丹明或与糖连接的荧光黄素)、含硫醇的核苷酸、生物素连接的核苷酸、荧光碱基类似物、CpG岛、甲基-7-鸟苷、甲基化的核苷酸、肌苷、硫代尿苷、假尿苷、二氢尿苷、癸苷(queuosine)以及怀俄苷(wyosine)。“核酸”、“多核苷酸”、“靶多核苷酸”和“靶核酸”可以互换使用。

[0078] 核酸可以包含一种或更多种修饰(例如,碱基修饰、骨架修饰),以向核酸提供新的或增强的特征(例如,改进的稳定性)。核酸可以包含核酸亲和标签。核苷可以是碱基-糖组合。核苷的碱基部分可以是杂环碱基。这样的杂环碱基的两个最常见的类别是嘌呤和嘧啶。核苷酸可以是进一步包括与核苷的糖部分共价连接的磷酸基团的核苷。对于包括呋喃戊糖的那些核苷,磷酸基团可以连接到糖的2'、3'或5'羟基部分。在形成核酸中,磷酸基团可以将相邻的核苷彼此共价连接以形成线性聚合化合物。继而此线性聚合化合物的各自末端可以进一步接合以形成环状化合物;然而,线性化合物通常是合适的。此外,线性化合物可以具有内部核苷酸碱基互补性,并且因此可以按产生完全或部分双链化合物的方式折叠。在核酸中,磷酸基团通常可以被称为形成核酸的核苷间骨架。连接或骨架可以是3'至5'磷酸二酯连接。

[0079] 核酸可以包含修饰的骨架和/或修饰的核苷间连接。修饰的骨架可以包括那些在骨架中保留磷原子的骨架和那些在骨架中不具有磷原子的骨架。其中含有磷原子的合适的修饰的核酸骨架可以包括,例如,硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其他烷基磷酸酯,诸如3'-亚烷基磷酸酯、5'-亚烷基磷酸酯、手性磷酸酯、亚磷酸酯,包括3'-氨基磷酰胺和氨基烷基磷酰胺的磷酰胺,磷酰二胺、硫代磷酰

胺、硫代烷基磷酸酯、硫代烷基磷酸三酯、硒代磷酸酯和具有正常的3'-5'连接的硼酸磷酸酯(boranophosphates)、2'-5'连接的类似物,以及具有反转极性的那些,其中一个或更多个核苷酸间连接为3'至3'、5'至5'或2'至2'连接。

[0080] 核酸可以包含由以下形成的多核苷酸骨架:短链烷基或环烷基核苷间连接、混合杂原子,和烷基或环烷基核苷间连接,或者一个或更多个短链杂原子的或杂环的核苷间连接。这些可以包括具有以下的那些:吗啉代连接(部分地从核苷的糖部分形成);硅氧烷骨架;硫化物、亚砷和砷骨架;甲乙酰基(formacetyl)和硫代甲乙酰基(thioformacetyl)骨架;亚甲基甲乙酰基和硫代甲乙酰基骨架;核糖乙酰基(riboacetyl)骨架;含有烯烃的骨架;氨基磺酸盐骨架;亚甲基亚氨基和亚甲基胍基骨架;磺酸盐和磺酰胺骨架;酰胺骨架;和具有混合的N、O、S和CH₂组成部分的其他物质。

[0081] 核酸可以包括核酸模拟物。术语“模拟物”可意在包括其中仅呋喃糖环或呋喃糖环和核苷酸间连接二者被非呋喃糖基团代替的多核苷酸,仅呋喃糖环的代替也可称为糖替代物。杂环碱基部分或修饰的杂环碱基部分可被保持以与适当的靶核酸杂交。一种这样的核酸可以是肽核酸(PNA)。在PNA中,多核苷酸的糖骨架可以被含酰胺的骨架,特别是氨基乙基甘氨酸骨架代替。核苷酸可以被保留,并直接或间接结合到骨架的酰胺部分的氮杂氮原子。PNA化合物中的骨架可以包含两个或更多个连接的氨基乙基甘氨酸单元,这为PNA提供含酰胺的骨架。杂环碱基部分可以直接或间接结合到骨架的酰胺部分的氮杂氮原子。

[0082] 核酸可以包含吗啉代骨架结构。例如,核酸可以包含代替核糖环的6元吗啉代环。在这些实施方案中的一些中,磷酰二胺或其他非磷酸二酯核苷间连接可以代替磷酸二酯连接。

[0083] 核酸可以包含具有附接到吗啉代环上的杂环碱基的连接的吗啉代单元(例如吗啉代核酸)。连接基团可以连接吗啉代核酸中的吗啉代单体单元。非离子的基于吗啉代的寡聚化合物可与细胞蛋白质具有较少的不期望的相互作用。基于吗啉代的多核苷酸可以是核酸的非离子模拟物。吗啉代类别中的各种化合物可以使用不同的连接基团连接。另一类多核苷酸模拟物可以称为环己烯基核酸(CeNA)。核酸分子中通常存在的呋喃糖环可以被环己烯基环取代。使用亚磷酰胺化学可以制备CeNA DMT保护的亚磷酰胺单体并用于寡聚化合物合成。将CeNA单体掺入核酸链中可以增加DNA/RNA杂交体的稳定性。CeNA寡核苷酸可以与核酸互补体形成稳定性与天然复合物相似的复合物。另外的修饰可以包括锁核酸(LNA),其中2'-羟基基团连接到糖环的4'碳原子,从而形成2'-C,4'-C-氧亚甲基连接,从而形成双环糖部分。连接可以是亚甲基(-CH₂),桥接2'氧原子和4'碳原子的基团,其中n是1或2。LNA和LNA类似物可以显示出与互补核酸的非常高的双链体热稳定性(T_m=+3°C至+10°C)、对3'-核酸外切酶降解的稳定性和良好的溶解度特性。

[0084] 核酸还可以包括核碱基(通常简称为“碱基”)修饰或取代。如本文使用的,“未修饰的”或“天然的”核碱基可以包括嘌呤碱基(例如腺嘌呤(A)和鸟嘌呤(G)),以及嘧啶碱基(例如胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U))。经修饰的核碱基可以包括其他合成以及天然的核碱基,诸如5-甲基胞嘧啶(5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基衍生物和其他烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基衍生物和其他烷基衍生物、2-硫代尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-卤素尿嘧啶(5-halouracil)和胞嘧啶、5-丙炔基(-C≡C-CH₃)尿嘧啶和胞嘧啶,和嘧啶碱基的其他炔基衍

生物、6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫代尿嘧啶,8-卤素、8-氨基、8-硫代、8-硫代烷基、8-羟基和其他8-取代的腺嘌呤和鸟嘌呤,5-卤素特别是5-溴、5-三氟甲基和其他5-取代的尿嘧啶和胞嘧啶,7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。修饰的核碱基可以包括三环嘧啶,诸如吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、G-夹类(G-clamps),诸如取代的吩噻嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、G-夹类,诸如取代的吩噻嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、呋唑胞苷(2H-嘧啶并(4,5-b)吡啶-2-酮)、吡啶并吡啶胞苷(H-吡啶并(3',2':4,5)吡咯并[2,3-d]嘧啶-2-酮)。

[0085] 如本文使用的,术语“样品”可以指包含靶的组合物。用于通过本文公开的方法、装置和系统进行分析的合适样品包括细胞、组织、器官或生物体。

[0086] 如本文使用的,术语“采样装置”或“装置”可以指可以取一部分样品和/或将所述部分放置在基底上的装置。采样装置可以指例如荧光激活细胞分选(FACS)机、细胞分选机、活组织检查针、活组织检查装置、组织切片装置、微流体装置、叶栅和/或超薄切片机。

[0087] 如本文使用的,术语“固体支持物”可以指可以附接多于一种条形码(例如,随机条形码)的离散固体或半固体表面。固体支持物可以包括任何类型的实心的、多孔的或空心的球体、球、承座、圆柱体或其他类似形状,由塑料、陶瓷、金属或聚合材料(例如,水凝胶)构成,其上可以固定核酸(例如,共价地或非共价地)。固体支持物可以包括可以是球形的(例如,微球)或具有非球形或不规则形状的离散颗粒,所述形状诸如立方体、长方体、锥体、圆柱体、圆锥体、椭圆形或圆盘形等。珠的形状可以是非球形的。以阵列间隔开的多于一个固体支持物可以不包括基底。固体支持物可以与术语“珠”互换使用。

[0088] 如本文使用的,术语“随机条形码”可以指包含本公开内容的标记的多核苷酸序列。随机条形码可以是可用于随机条形码化的多核苷酸序列。随机条形码可用于对样品中的靶进行定量。随机条形码可用于控制标记与靶关联后可能发生的错误。例如,随机条形码可用于评估扩增或测序错误。与靶关联的随机条形码可以称为随机条形码-靶或随机条形码-标签-靶。

[0089] 如本文使用的,术语“基因特异性随机条形码”可以指包含标记和基因特异性的靶结合区的多核苷酸序列。随机条形码可以是可用于随机条形码化的多核苷酸序列。随机条形码可用于对样品中的靶进行定量。随机条形码可用于控制标记与靶关联后可能发生的错误。例如,随机条形码可用于评估扩增或测序错误。与靶关联的随机条形码可以称为随机条形码-靶或随机条形码-标签-靶。

[0090] 如本文使用的,术语“随机条形码化”可以指核酸的随机标记(例如,条形码化)。随机条形码化可以利用递归泊松策略来关联并对与靶关联的标记进行定量。如本文使用的,术语“随机条形码化”可以与“随机进行标记”互换使用。

[0091] 如本文使用的,术语“靶”可以指可与条形码(例如,随机条形码)关联的组合物。用于通过本文公开的方法、装置和系统进行分析的示例性合适的靶包括寡核苷酸、DNA、RNA、mRNA、微RNA、tRNA等。靶可以是单链的或双链的。在一些实施方案中,靶可以是蛋白质、肽或

多肽。在一些实施方案中,靶是脂质。如本文使用的,“靶”可以与“种类”互换使用。

[0092] 如本文使用的,术语“逆转录酶”可以指具有逆转录酶活性(即,催化从RNA模板合成DNA)的一组酶。通常,这样的酶包括但不限于逆转录病毒逆转录酶、逆转录转座子逆转录酶、逆转录质粒逆转录酶、逆转录子逆转录酶(retron reverse transcriptase)、细菌逆转录酶、II型内含子衍生的逆转录酶(group II intron-derived reverse transcriptase),以及它们的突变体、变体或衍生物。非逆转录病毒逆转录酶包括非LTR逆转录转座子逆转录酶、逆转录质粒逆转录酶、逆转录子逆转录酶和II型内含子逆转录酶。II型内含子逆转录酶的实例包括乳酸乳球菌(*Lactococcus lactis*)LI.LtrB内含子逆转录酶、细长嗜热聚球藻(*Thermosynechococcus elongatus*)TeI4c内含子逆转录酶或嗜热脂肪地芽孢杆菌(*Geobacillus stearothermophilus*)GsI-IIC内含子逆转录酶。其他类别的逆转录酶可以包括许多类型的非逆转录病毒逆转录酶(即,逆转录子、II型内含子,以及多样性产生型逆转录元件,等等)。

[0093] 术语“通用衔接子引物”、“通用引物衔接子”或“通用衔接子序列”可互换地使用以指可以用于与条形码(例如,随机条形码)杂交以产生基因特异性条形码的核苷酸序列。通用衔接子序列例如可以是跨越本公开内容的方法中使用的所有条形码通用的已知序列。例如,当使用本文公开的方法标记多种靶时,每种靶特异性序列可以连接到相同的通用衔接子序列。在一些实施方案中,多于一种通用衔接子序列可以用于本文公开的方法中。例如,当使用本文公开的方法标记多个靶时,至少两种靶特异性序列连接到不同的通用衔接子序列。通用衔接子引物及其互补体可以包括在两种寡核苷酸中,其中的一种包含靶特异性序列且另一种包含条形码。例如,通用衔接子序列可以是包含靶特异性序列的寡核苷酸的一部分以产生与靶核酸互补的核苷酸序列。包含条形码以及通用衔接子序列的互补体的第二寡核苷酸可与核苷酸序列杂交并产生靶特异性条形码(例如,靶特异性随机条形码)。在一些实施方案中,通用衔接子引物具有与本公开内容的方法中使用的通用PCR引物不同的序列。

[0094] 条形码

[0095] 条形码化,诸如随机条形码化,已描述于例如US20150299784,W02015031691,以及Fu等人,Proc Natl Acad Sci U.S.A.2011May31;108(22):9026-31中,这些出版物的内容以其整体在此并入本文。在一些实施方案中,本文公开的条形码可以是随机条形码,所述随机条形码可以是可用于对靶进行随机标记(例如,条形码,标签)的多核苷酸序列。如果随机条形码中的不同条形码序列的数量与待标记的任何靶的出现次数(number of occurrence)的比率可以是以下,或是约以下,则条形码可以称为随机条形码:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1,或在这些值中的任何两个值之间的数字或范围。靶可以是包括具有相同或几乎相同序列的mRNA分子的mRNA种类。如果随机条形码中的不同条形码序列的数量与待标记的任何靶的出现次数的比率是至少以下,或至多以下,则条形码可以称为随机条形码:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1或100:1。随机条形码的条形码序列可以称为分子标记。

[0096] 条形码,例如随机条形码,可以包含一种或更多种标记。示例性标记可以包括通用

标记、细胞标记、条形码序列(例如,分子标记)、样品标记、板标记、空间标记和/或前空间标记(pre-spatial label)。图1图示了具有空间标记的示例性条形码104。条形码104可以包含可将条形码与固体支持物105连接的5' 胺。条形码可以包含通用标记、维度标记、空间标记、细胞标记和/或分子标记。条形码中不同标记(包括但不限于通用标记、维度标记、空间标记、细胞标记和分子标记)的顺序可以改变。例如,如图1中示出的,通用标记可以是最5' 侧的标记(5' -most label),且分子标记可以是最3' 侧的标记(3' -most label)。空间标记、维度标记和细胞标记可以处于任何顺序。在一些实施方案中,通用标记、空间标记、维度标记、细胞标记和分子标记是处于任何顺序的。条形码可以包含靶结合区。靶结合区可以与样品中的靶(例如,靶核酸、RNA、mRNA、DNA)相互作用。例如,靶结合区可以包括可以与mRNA的多(A)尾相互作用的寡聚(dT)序列。在一些情况下,条形码的标记(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列)可以由1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个或更多个核苷酸隔开。

[0097] 标记,例如,细胞标记,可以包括一组独特的定义长度的核酸子序列,例如各自有七个核苷酸(相当于一些汉明纠错码(Hamming error correction codes)中使用的数量),其可以设计为提供纠错能力。可以设计包含七个核苷酸序列的纠错子序列组,使得所述组中的序列的任何成对组合展现出定义的“遗传距离”(或错配碱基数),例如一组纠错子序列能被设计为展现三个核苷酸的遗传距离。在这种情况下,对于经标记的靶核酸分子的序列数据组中的纠错序列的审查(在下文更全面地描述)能允许检测或纠正扩增或测序错误。在一些实施方案中,用于产生纠错码的核酸子序列的长度可以变化,例如,它们的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、31个、40个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,其他长度的核酸子序列可以用来产生纠错码。

[0098] 条形码可以包含靶结合区。靶结合区可以与样品中的靶相互作用。靶可以是,或包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、各自含有多(A)尾的RNA,或它们的任何组合。在一些实施方案中,多于一种靶可以包括脱氧核糖核酸(DNA)。

[0099] 在一些实施方案中,靶结合区可以包含可以与mRNA的多(A)尾相互作用的寡聚(dT)序列。条形码的标记中的一种或更多种(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列(例如,分子标记))可以通过间隔序列(spacer)与条形码的剩余标记中的另一个或两个隔开。间隔序列可以是例如,1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个或更多个核苷酸。在一些实施方案中,条形码的标记中没有标记被间隔序列隔开。

[0100] 通用标记

[0101] 条形码可以包含一种或更多种通用标记。在一些实施方案中,对于条形码组中的所有条形码(附接到给定的固体支持物上的),一种或更多种通用标记可以是相同的。在一些实施方案中,对于附接到多于一个珠上的所有条形码,一种或更多种通用标记可以是相同的。在一些实施方案中,通用标记可以包括能够与测序引物杂交的核酸序列。测序引物可以用于对包含通用标记的条形码进行测序。测序引物(例如,通用测序引物)可以包括与高通量测序平台有关的测序引物。在一些实施方案中,通用标记可以包括能够与PCR引物杂交

的核酸序列。在一些实施方案中,通用标记可以包括能够与测序引物和PCR引物杂交的核酸序列。能够与测序或PCR引物杂交的通用标记的核酸序列可以被称作引物结合位点。通用标记可以包括可用于起始条形码转录的序列。通用标记可以包括可用于使条形码或条形码内的区域延伸的序列。通用标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。例如,通用标记可以包括至少约10个核苷酸。通用标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。在一些实施方案中,可裂解接头或修饰的核苷酸可以是通用标记序列的一部分,以使条形码能够从支持物上被裂解下来。

[0102] 维度标记

[0103] 条形码可以包含一种或更多种维度标记。在一些实施方案中,维度标记可以包括提供关于发生标记(例如,随机标记)的维度的信息的核酸序列。例如,维度标记可以提供关于对靶进行条形码化的时间的信息。维度标记可以与样品中条形码化(例如,随机条形码化)的时间关联。维度标记可以在标记的时间被激活。不同的维度标记可以在不同的时间被激活。维度标记提供关于靶、靶的组和/或样品被条形码化的顺序的信息。例如,在细胞周期的G0期可以对细胞群体进行条形码化。在细胞周期的G1期,可以用条形码(例如,随机条形码)再次对细胞进行脉冲处理。在细胞周期的S期,可以用条形码再次对细胞进行脉冲处理,等等。每次脉冲(例如,细胞周期的每个期)时的条形码可以包含不同的维度标记。以这种方式,维度标记提供关于哪些靶在细胞周期的哪个期被标记的信息。维度标记可以询问许多不同的生物学时间。示例性的生物学时间可以包括但不限于细胞周期、转录(例如,转录起始)和转录物降解。在另一个实例中,样品(例如,细胞、细胞群体)可以在用药物和/或疗法治疗之前和/或之后进行标记。不同靶的拷贝数的变化可以指示样品对药物和/或疗法的反应。

[0104] 维度标记可以是可激活的。可以在特定时间点激活可激活的维度标记。可激活的标记可以被例如组成性地激活(例如,不关闭)。所述可激活的维度标记可以被例如可逆地激活(例如,所述可激活的维度标记可以打开和关闭)。维度标记可以被例如可逆地激活至少1次、2次、3次、4次、5次、6次、7次、8次、9次、10次或更多次。维度标记可以被可逆地激活例如至少1次、2次、3次、4次、5次、6次、7次、8次、9次、10次或更多次。在一些实施方案中,可以用荧光、光、化学事件(例如,裂解,另一种分子的连接,修饰的添加(例如,聚乙二醇化、sumoylation)、乙酰化、甲基化、去乙酰化、去甲基化)、光化学事件(例如,光锁定(photocaging))以及引入非天然的核苷酸将所述维度标记激活。

[0105] 在一些实施方案中,维度标记对于附接到给定的固体支持物(例如,珠)上的所有条形码(例如,随机条形码)可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上的至少60%、70%、80%、85%、90%、95%、97%、99%或100%的条形码可以包含相同的维度标记。在一些实施方案中,相同固体支持物上的至少60%的条形码可以包含相同的维度标记。在一些实施方案中,相同固体支持物上的至少95%的条形码可以包含相同的维度标记。

[0106] 多于一个固体支持物(例如,珠)可以表现多达 10^6 种或更多种独特维度标记序列。维度标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30

个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。维度标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。维度标记可以包括长度在约5个至约200个之间的核苷酸、约10个至约150个之间的核苷酸或约20个至约125个之间的核苷酸。

[0107] 空间标记

[0108] 条形码可以包含一种或更多种空间标记。在一些实施方案中,空间标记可以包括提供与条形码关联的靶分子的空间定向的信息的核酸序列。空间标记可以与样品中的坐标关联。所述坐标可以是固定的坐标。例如,可以参考基底将坐标固定。空间标记可以参考二维或三维网格。可以参考界标 (landmark) 将坐标固定。在空间中界标是可被鉴定的。界标可以是可被成像的结构。界标可以是生物学结构,例如解剖学界标。界标可以是细胞界标,例如细胞器。界标可以是非天然界标,诸如具有可鉴定标识(诸如颜色编码、条形码、磁性、荧光、放射性或独特尺寸或形状)的结构。空间标记可以与物理分区(例如,孔、容器或液滴)关联。在一些实施方案中,将多于一种空间标记一起用于编码空间中的一个或更多个位置。

[0109] 空间标记对于附接到给定的固体支持物(例如,珠)上的所有条形码可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上包含相同空间标记的条形码的百分比可以是以下,或约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,相同固体支持物上包含相同空间标记的条形码的百分比可以是至少以下,或至多以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。在一些实施方案中,相同固体支持物上的至少60%的条形码可以包含相同的空间标记。在一些实施方案中,相同固体支持物上的至少95%的条形码可以包含相同的空间标记。

[0110] 多于一个固体支持物(例如,珠)可以表现多达 10^6 种或更多种独特空间标记序列。空间标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。空间标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。空间标记可以包含长度在约5个至约200个之间的核苷酸、约10个至约150个之间的核苷酸或约20个至约125个的核苷酸。

[0111] 细胞标记

[0112] 条形码(例如,随机条形码)可以包含一种或更多种细胞标记。在一些实施方案中,细胞标记可以包括提供用于确定哪种靶核酸来自哪个细胞的信息的核酸序列。在一些实施方案中,细胞标记对于附接到给定的固体支持物(例如,珠)上的所有条形码是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上包含相同细胞标记的条形码的百分比可以是以下,或约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,相同固体支持物上包含相同细胞标记的条形码的百分比可以是以下,或约以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。例如,相同固体支持物上的至少60%的条形码可以包含相同的细胞标记。作为另一个实例,相同固体支持物上的至少95%的条形码可以

包含相同的细胞标记。

[0113] 多于一个固体支持物(例如,珠)可以表现多达 10^6 种或更多种独特细胞标记序列。细胞标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。细胞标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。例如,细胞标记可以包括长度在约5个至约200个之间的核苷酸、约10个至约150个之间的核苷酸或约20个至约125个之间的核苷酸。

[0114] 条形码序列

[0115] 条形码可以包含一种或更多种条形码序列。在一些实施方案中,条形码序列可以包括提供与条形码杂交的特定类型的靶核酸种类的鉴定信息的核酸序列。条形码序列可以包括如下核酸序列,所述核酸序列提供与条形码(例如,靶结合区)杂交的靶核酸种类的特定出现的计数(例如,提供粗略近似)。

[0116] 在一些实施方案中,一组不同的条形码序列附接到给定的固体支持物(例如,珠)。在一些实施方案中,可以有以下,或有约以下的独特分子标记序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种,或在这些值中的任何两个值之间的数字或范围。例如,多于一种条形码可以包括具有不同序列的约6561种条形码序列。作为另一个实例,多于一种条形码可以包括具有不同序列的约65536种条形码序列。在一些实施方案中,可以有至少以下,或至多以下的独特条形码序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。独特分子标记序列可以附接到给定的固体支持物(例如,珠)。

[0117] 在不同实施方式中,条形码的长度可以是不同的。例如,条形码的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。作为另一个实例,条形码的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。

[0118] 分子标记

[0119] 条形码(例如,随机条形码)可以包含一种或更多种分子标记。分子标记可以包括条形码序列。在一些实施方案中,分子标记可以包括提供与条形码杂交的特定类型的靶核酸种类的鉴定信息的核酸序列。分子标记可以包括如下核酸序列,所述核酸序列提供与条形码(例如,靶结合区)杂交的靶核酸种类的特定出现的计数。

[0120] 在一些实施方案中,一组不同的分子标记附接到给定的固体支持物(例如,珠)。在一些实施方案中,可以有以下,或有约以下的独特分子标记序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种,或在这些值中的任何两个值之间的数字或范围。例如,多于一种条形码可以包括具有不同序列的约6561种分子标记。作为另一个实例,多于一种条形码可以包括具有不同序列的约65536种分子标记。在一些实施方案中,可以有至少以下,或至多以下的独特分子标记序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。具有独特分子标记序列的条形码可以附接到给定的固体支持物(例如,珠)。

[0121] 对于使用多于一种随机条形码进行的随机条形码化,不同分子标记序列的数量与任何靶的出现次数的比率可以是以下,或约以下:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:

1:10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1,或在这些值中的任何两个值之间的数字或范围。靶可以是包括具有相同或几乎相同序列的mRNA分子的mRNA种类。在一些实施方案中,不同分子标记序列的数量与任何靶的出现次数的比率是至少以下,或至多以下:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1或100:1。

[0122] 分子标记的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。分子标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。

[0123] 靶结合区

[0124] 条形码可以包含一种或更多种靶结合区,诸如捕获探针。在一些实施方案中,靶结合区可以与感兴趣的靶杂交。在一些实施方案中,靶结合区可以包括与靶(例如,靶核酸、靶分子,例如待分析的细胞核酸),例如,与特定基因序列,进行特异性杂交的核酸序列。在一些实施方案中,靶结合区可以包括可附接(例如,杂交)至特定靶核酸的特定位置的核酸序列。在一些实施方案中,靶结合区可以包括能够与限制性酶位点突出端(例如,EcoRI粘性末端突出端)进行特异性杂交的核酸序列。然后条形码可以连接到包括与限制性位点突出端互补的序列的任何核酸分子。

[0125] 在一些实施方案中,靶结合区可以包括非特异性靶核酸序列。非特异性靶核酸序列可以指不依赖靶核酸的特定序列可与多种靶核酸结合的序列。例如,靶结合区可以包括与mRNA分子上的多(A)尾杂交的随机多聚体序列或寡聚(dT)序列。随机多聚体序列可以是例如随机二聚体、三聚体、四聚体、五聚体、六聚体、七聚体、八聚体、九聚体、十聚体或任何长度的更高多聚体序列。在一些实施方案中,附接到给定珠的所有条形码的靶结合区是相同的。在一些实施方案中,附接到给定珠的多于一种条形码的靶结合区可以包括两种或更多种不同的靶结合序列。靶结合区的长度可以是以下,或是约以下:5个、10个、15个、20个、25个、30个、35个、40个、45个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。靶结合区的长度可以是至多约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更多个核苷酸。

[0126] 在一些实施方案中,靶结合区可以包括寡聚(dT),所述寡聚(dT)可以与包含多腺苷酸化末端的mRNA杂交。靶结合区可以是基因特异性的。例如,可以将靶结合区配置为与靶的特定区域杂交。靶结合区的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。靶结合区的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。靶结合区的长度可以是约5-30个核苷酸。当条形码包含基因特异性靶结合区时,所述条形码在本文中可以被称为基因特异性条形码。

[0127] 定向特性

[0128] 随机条形码(例如,随机条形码)可以包括一种或更多种可用于定向(例如,比对)条形码的定向特性。条形码可以包含用于等电聚焦的部分。不同的条形码可以包括不同的等电聚焦点。当将这些条形码引入样品中时,所述样品可以经历等电聚焦,以便于将条形码定位成已知的方式。以这种方式,定向特性可以用于开发样品中条形码的已知的地图(map)。示例性定向特性可以包括电泳迁移(例如,基于条形码的尺寸)、等电点、自旋、电导和/或自组装。例如,具有自组装的定向特性的条形码在激活后可以自组装成特定定向(例如,核酸纳米结构)。

[0129] 亲和特性

[0130] 条形码(例如,随机条形码)可以包括一种或更多种亲和特性。例如,空间标记可以包括亲和特性。亲和特性可以包括可促进条形码与另一种实体(例如,细胞受体)结合的化学和/或生物部分。例如,亲和特性可以包括抗体,例如,对样品上的特定部分(例如,受体)有特异性的抗体。在一些实施方案中,抗体可以将条形码引导至特定细胞类型或分子。在特定细胞类型或分子处的和/或在特定细胞类型或分子附近的靶可以被标记(例如,随机地标记)。在一些实施方案中,因为抗体可以将条形码引导至特定位置,亲和特性可以提供除了空间标记的核苷酸序列之外的空间信息。抗体可以是治疗性抗体,例如单克隆抗体或多克隆抗体。抗体可以是人源化的或嵌合的。抗体可以是裸抗体(naked antibody)或融合抗体。

[0131] 抗体可以是全长(即,天然存在的或通过正常免疫球蛋白基因片段重组过程形成的)免疫球蛋白分子(例如,IgG抗体)或免疫球蛋白分子的免疫活性(即,特异性结合)部分(如抗体片段)。

[0132] 抗体片段可以是例如抗体的一部分,诸如F(ab')₂、Fab'、Fab、Fv、sFv等。在一些实施方案中,抗体片段可以与由全长抗体识别的相同抗原结合。抗体片段可以包括由抗体的可变区组成的分离的片段,诸如由重链和轻链的可变区组成的“Fv”片段和其中轻链和重链的可变区通过肽接头连接的重组单链多肽分子(“scFv蛋白”)。示例性抗体可以包括但不限于癌细胞抗体、病毒抗体、结合至细胞表面受体(CD8、CD34、CD45)的抗体,和治疗性抗体。

[0133] 通用衔接子引物

[0134] 条形码可以包含一种或更多种通用衔接子引物。例如,基因特异性条形码,诸如基因特异性随机条形码,可以包含通用衔接子引物。通用衔接子引物可以指跨越所有条形码通用的核苷酸序列。通用衔接子引物可以用于构建基因特异性条形码。通用衔接子引物的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个,或在这些值中的任何两个值之间的数字或范围的核苷酸。通用衔接子引物的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。通用衔接子引物的长度可以是5-30个核苷酸。

[0135] 接头

[0136] 当条形码包含多于一种标记类型(例如,多于一种细胞标记或多于一种条形码序列,诸如一种分子标记)时,标记之间可以散布有接头标记序列。接头标记序列的长度可以是至少约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更多个核苷酸。接头标记序列的长度可以是至多约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更

多个核苷酸,例如长度为12个核苷酸。接头标记序列可用于促进条形码的合成。接头标签可以包括纠错(例如汉明)码。

[0137] 固体支持物

[0138] 在一些实施方案中,本文公开的条形码,诸如随机条形码,可以与固体支持物关联。例如,固体支持物可以是合成颗粒。在一些实施方案中,固体支持物上的多于一种条形码(例如,第一多于一种条形码)的一些或所有条形码序列(诸如,随机条形码(例如,第一条条形码序列)的分子标记)具有至少一个核苷酸的差异。相同固体支持物上的条形码的细胞标记可以是相同的。不同固体支持物上的条形码的细胞标记可以具有至少一个核苷酸的差异。例如,第一固体支持物上的第一多于一种条形码的第一细胞标记可以具有相同的序列,且第二固体支持物上的第二多于一种条形码的第二细胞标记可以具有相同的序列。第一固体支持物上的第一多于一种条形码的第一细胞标记和第二固体支持物上的第二多于一种条形码的第二细胞标记可以具有至少一个核苷酸的差异。细胞标记可以是例如约5-20个核苷酸长。条形码序列可以是例如约5-20个核苷酸长。合成颗粒可以例如是珠。

[0139] 珠可以例如是硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、交联葡聚糖/琼脂糖凝胶珠、珠状纤维素、聚苯乙烯珠,或它们的任何组合。珠可以包括诸如以下的材料:聚二甲基硅氧烷(PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,或它们的任何组合。

[0140] 在一些实施方案中,珠可以用条形码或随机条形码官能化的聚合珠,例如可变形的珠或凝胶珠(诸如来自10X Genomics(San Francisco,CA)的凝胶珠)。在一些实施方式中,凝胶珠可以包括基于聚合物的凝胶。例如,凝胶珠可以通过将一种或更多种聚合物前体包封进液滴中来产生。在将聚合物前体暴露于促进剂(例如,四甲基乙二胺(TEMED))后,可以产生凝胶珠。

[0141] 在一些实施方案中,颗粒可以是可降解的。例如,聚合珠可以例如在所期望的条件下溶解、熔化或降解。所期望的条件可以包括环境条件。所期望的条件可导致聚合珠以受控方式溶解、熔化或降解。由于化学刺激物、物理刺激物、生物刺激物、热刺激物、磁刺激物、电刺激物、光刺激物,或它们的任何组合,凝胶珠可以溶解、熔化或降解。

[0142] 分析物和/或试剂(诸如寡核苷酸条形码)例如可以偶联/固定到凝胶珠的内表面(例如,经由寡核苷酸条形码和/或用于产生寡核苷酸条形码的材料的扩散而可及的内部),和/或凝胶珠或本文描述的任何其他微囊的外表面。偶联/固定可以经由任何形式的化学键合(例如,共价键、离子键)或物理现象(例如,范德华力、偶极-偶极相互作用等)。在一些实施方案中,试剂与凝胶珠或本文描述的任何其他微囊的偶联/固定可以是可逆的,诸如,例如经由不稳定部分(例如,经由化学交联剂,包括本文描述的化学交联剂)。在施加刺激物后,不稳定部分可以被裂解并释放固定化的试剂。在一些实施方案中,不稳定部分是二硫键。例如,在经由二硫键将寡核苷酸条形码固定到凝胶珠的情况下,将二硫键暴露于还原剂可以将二硫键裂解并从珠释放寡核苷酸条形码。不稳定部分可以作为凝胶珠或微囊的一部分、作为将试剂或分析物与凝胶珠或微囊连接的化学接头的一部分,和/或作为试剂或分析物的一部分包括在内。在一些实施方案中,多于一种条形码中的至少一种条形码可固定在颗粒上、部分固定在颗粒上、包封在颗粒中、部分包封在颗粒中,或它们的任何组合。

[0143] 在一些实施方案中,凝胶珠可以包括广泛多种不同的聚合物,包括但不限于:聚合

物、热敏聚合物、光敏聚合物、磁性聚合物、pH敏感聚合物、盐敏感聚合物、化学敏感聚合物、聚电解质、多糖、肽、蛋白质和/或塑料。聚合物可包括但不限于以下材料：诸如聚(N-异丙基丙烯酰胺) (PNIPAAm)、聚(苯乙烯磺酸酯) (PSS)、聚(烯丙基胺) (PAAm)、聚(丙烯酸) (PAA)、聚(乙烯亚胺) (PEI)、聚(二烯丙基二甲基-氯化铵) (PDADMAC)、聚(吡咯) (PPy)、聚(乙烯基吡咯烷酮) (PVPON)、聚(乙烯基吡啶) (PVP)、聚(甲基丙烯酸甲酯) (PMAA)、聚(甲基丙烯酸甲酯) (PMMA)、聚苯乙烯 (PS)、聚(四氢呋喃) (PTHF)、聚(邻苯二甲醛) (PPA)、聚(己基紫精) (PHV)、聚(L-赖氨酸) (PLL)、聚(L-精氨酸) (PARG)、聚(乳酸-共-乙醇酸) (PLGA)。

[0144] 许多化学刺激物可用于触发珠的破坏、溶解或降解。这些化学改变的实例可包括但不限于pH介导的珠壁改变、经由交联键的化学裂解使珠壁崩解、珠壁的触发解聚,和珠壁转换反应(bead wall switching reactions)。容量改变(bulk changes)也可用于触发珠的破坏。

[0145] 通过各种刺激物对微囊的容量或物理变化在设计囊以释放试剂方面也提供了许多优点。在宏观尺度上发生容量或物理变化,其中珠破裂是由刺激物引起的机械-物理力的结果。这些过程可包括但不限于压力引起的破裂、珠壁熔化,或珠壁的孔隙率的改变。

[0146] 生物刺激物也可用于触发珠的破坏、溶解或降解。通常,生物触发物类似于化学触发物,但是许多实例使用生物分子或生命系统中常见的分子,诸如酶、肽、糖、脂肪酸、核酸等。例如,珠可包括具有肽交联的聚合物,所述肽交联对特定蛋白酶的裂解敏感。更具体地,一个实例可包括含有GFLGK肽交联的微囊。在添加生物触发物(诸如蛋白酶组织蛋白酶B)后,壳壁的肽交联被裂解且珠的内容物被释放。在其他情况下,蛋白酶可以是热激活的。在另一个实例中,珠包括含有纤维素的壳壁。水解酶壳聚糖的添加用作用于纤维素键的裂解、壳壁的解聚和其内部内容物的释放的生物触发物。

[0147] 还可以在施加热刺激物后诱导珠释放其内容物。温度的变化可导致珠的各种变化。热量的变化可能导致珠的熔化,使得珠壁崩解。在其他情况下,热量可能增加珠的内部组分的内部压力,使得珠破裂或爆炸。在仍然其他情况下,热量可以使珠转变成收缩的脱水状态。热量还可以作用于珠壁内的热敏聚合物,从而引起珠的破坏。

[0148] 将磁性纳米颗粒包括在微囊的珠壁中可以允许珠的触发破裂以及在阵列中引导珠。本公开内容的装置可以包括用于任一目的的磁珠。在一个实例中,将 Fe_3O_4 纳米颗粒掺入含电解质的珠中在振荡磁场刺激物的存在下触发破裂。

[0149] 珠也可能由于电刺激而破坏、溶解或降解。与先前部分中描述的磁性颗粒相似,电敏珠可以允许珠的触发破裂以及其他功能,诸如电场中的对准、电导或氧化还原反应。在一个实例中,含有电敏材料的珠在电场中对准,从而可以控制内部试剂的释放。在其他实例中,电场可以在珠壁本身内引起氧化还原反应,这可以增加孔隙率。

[0150] 也可用光刺激物使珠破坏。许多光触发物是可能的,并可以包括使用各种分子(诸如能够吸收特定波长范围的光子的纳米颗粒和发色团)的系统。例如,金属氧化物涂层可用作囊触发物。涂覆有 SiO_2 的聚电解质囊的UV照射可导致珠壁的崩解。在又一个实例中,可以将可光转换的材料,诸如偶氮苯基团,掺入珠壁中。在施加UV或可见光后,诸如这些的化学物质在吸收光子后经历可逆的顺式-至-反式异构化。在此方面,包括光转换导致珠壁在施加光触发物后可崩解或变得更为多孔。

[0151] 例如,在图2中图示的条形码化(例如,随机条形码化)的非限制性实例中,在框208

处将细胞(诸如单细胞)引入微孔阵列的多于一个微孔之后,在框212处可以将珠引入微孔阵列的多于一个微孔。每个微孔可包含一个珠。珠可以包含多于一种条形码。条形码可以包含附接到珠的5' 胺区域。条形码可以包含通用标记、条形码序列(例如,分子标记)、靶结合区,或它们的任何组合。

[0152] 本文公开的条形码可以与固体支持物(例如,珠)关联(例如,附接)。与固体支持物关联的条形码可以各自包含选自包括以下的组的条形码序列:具有独特序列的至少100种或1000种条形码序列。在一些实施方案中,与固体支持物关联的不同条形码可以包括具有不同序列的条形码。在一些实施方案中,一定百分比的与固体支持物关联的条形码包含相同的细胞标记。例如,所述百分比可以是以下,或是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或在这些值中的任何两个值之间的数字或范围。作为另一个实例,所述百分比可以是至少以下,或至多以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。在一些实施方案中,与固体支持物关联的条形码可以具有相同的细胞标记。与不同固体支持物关联的条形码可以具有选自包括以下的组的不同的细胞标记:具有独特序列的至少100种或1000种细胞标记。

[0153] 本文公开的条形码可以与固体支持物(例如,珠)关联(例如,附接)。在一些实施方案中,可以用包括与多于一种条形码关联的多于一个合成的颗粒的固体支持物对样品中的多于一种靶进行条形码化。在一些实施方案中,固体支持物可包括与多于一种条形码关联的多于一个合成的颗粒。不同固体支持物上的多于一种条形码的空间标记可以具有至少一个核苷酸的差异。固体支持物可以例如包括处于二维或三维的多于一种条形码。合成的颗粒可以是珠。珠可以是硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、交联葡聚糖/琼脂糖凝胶珠、珠状纤维素、聚苯乙烯珠,或它们的任何组合。固体支持物可包括聚合物、基质、水凝胶、针阵列装置、抗体,或它们的任何组合。在一些实施方案中,固体支持物可以自由浮动。在一些实施方案中,固体支持物可嵌入半固体或固体阵列中。条形码可以不与固体支持物关联。条形码可以是单独的核苷酸。条形码可与基底关联。

[0154] 如本文使用的,术语“拴系”、“附接”和“固定”可以互换使用,并且可以指用于将条形码附接至固体支持物的共价或非共价手段(means)。可以将各种不同的固体支持物中的任何一种用作固体支持物,以用于附接预先合成的条形码或用于条形码的原位固相合成。

[0155] 在一些实施方案中,固体支持物是珠。珠可以包括一种或更多种类型的实心的、多孔的或空心的球体、球、承座、圆柱体或其他相似形状,其上可以固定核酸(例如,共价地或非共价地)。珠可以例如由塑料、陶瓷、金属、聚合物材料,或它们的任何组合构成。珠可以是,或包括,球形的(例如,微球)或具有非球形或不规则形状的离散颗粒,所述形状是诸如立方体、长方体、锥体、圆柱体、圆锥体、椭圆形或圆盘形等。在一些实施方案中,珠的形状可以是非球形的。

[0156] 珠可以包括各种材料,包括但不限于顺磁性材料(例如,镁、钼、锂和钽)、超顺磁性材料(例如,铁氧体(Fe_3O_4 ;磁铁矿)纳米颗粒)、铁磁材料(例如,铁、镍、钴,它们的一些合金,以及一些稀土金属化合物)、陶瓷、塑料、玻璃、聚苯乙烯、二氧化硅、甲基苯乙烯、丙烯酸聚合物、钛、胶乳、交联琼脂糖、琼脂糖、水凝胶、聚合物、纤维素、尼龙,或它们的任何组合。

[0157] 在一些实施方案中,珠(例如,标记所附接的珠)是水凝胶珠。在一些实施方案中,珠包括水凝胶。

[0158] 本文公开的一些实施方案包括一个或更多个颗粒(例如,珠)。颗粒中的每一个可以包括多于一种寡核苷酸(例如,条形码)。多于一种寡核苷酸中的每一种可以包括条形码序列(例如,分子标记序列)、细胞标记和靶结合区(例如,寡聚(dT)序列、基因特异性序列、随机多聚体,或它们的组合)。多于一种寡核苷酸中的每一种的细胞标记序列可以是相同的。不同颗粒上的寡核苷酸的细胞标记序列可以是不同的,使得可以鉴定不同颗粒上的寡核苷酸。在不同实施方式中,不同细胞标记序列的数量可以是不同的。在一些实施方案中,细胞标记序列的数量可以是以下,或是约以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 、 10^9 ,在这些值中的任何两个值之间的数字或范围,或更多。在一些实施方案中,细胞标记序列的数量可以是至少以下,或至多以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 或 10^9 。在一些实施方案中,多于一个颗粒中不超过1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种或更多种包括具有相同细胞序列的寡核苷酸。在一些实施方案中,包括具有相同细胞序列的寡核苷酸的多于一个颗粒可以是至多0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%或更多。在一些实施方案中,多于一个颗粒中没有一个具有相同的细胞标记序列。

[0159] 每个颗粒上的多于一种寡核苷酸可以包含不同的条形码序列(例如,分子标记)。在一些实施方案中,条形码序列的数量可以是以下,或是约以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 、 10^9 ,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,条形码序列的数量可以是至少以下,或至多以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 或 10^9 。例如,多于一种寡核苷酸中的至少100种包含不同的条形码序列。作为另一个实例,在单个颗粒中,多于一种寡核苷酸中的至少100种、500种、1000种、5000种、10000种、15000种、20000种、50000种,这些值的任何两个之间的数字或范围,或更多种包含不同的条形码序列。一些实施方案提供了包含条形码的多于一个颗粒。在一些实施方案中,待标记的靶和不同条形码序列的出现(或拷贝或数量)的比率可以是至少1:1、1:2、1:3、1:4、1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:30、1:40、1:50、1:60、1:70、1:80、1:90或更多。在一些实施方案中,多于一种寡核苷酸中的每一种还包含样品标记、通用标记,或两者。颗粒可以是例如纳米颗粒或微颗粒。

[0160] 珠的尺寸可以变化。例如,珠的直径的范围可以是0.1微米至50微米。在一些实施方案中,珠的直径可以是以下,或是约以下:0.1微米、0.5微米、1微米、2微米、3微米、4微米、5微米、6微米、7微米、8微米、9微米、10微米、20微米、30微米、40微米、50微米,或这些值中的任何两个值之间的数字或范围。

[0161] 珠的直径可以与基底的孔的直径相关。在一些实施方案中,珠的直径可以比孔的直径长或短10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或这些值中的任何两个值之间的数字或范围,或比孔的直径长或短约10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或这些值中的任何两个值之间的数字或范围。珠的直径可以与细胞(例如,被基质的孔截留的单细胞)的直径相关。在一些实施方案中,珠的直径可以比孔的直径长或短至少或至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%。珠的直径可以与细胞(例如,被基质的孔截留的单细胞)的直径相关。在一些实施方案中,珠的直径或多于一个珠的平均直径可以比细胞的直径长或短10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或这些值中的任何两个值之间的数字或范围,或比细胞的直径长或短约10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或这些值中的任何两个值之间的数字或范围。

[0162] 珠可以附接到和/或嵌入到基底中。珠可以附接到和/或嵌入到凝胶、水凝胶、聚合物和/或基质中。珠在基底(例如,凝胶、基质、支架或聚合物)中的空间位置可以使用珠上的条形码上存在的可以用作位置地址的空间标签来鉴定。

[0163] 珠的实例包括但不限于链霉抗生物素蛋白珠、琼脂糖珠、磁珠、Dynabeads®、MACS®微珠、抗体缀合珠(例如,抗免疫球蛋白微珠)、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠和BcMag™羧基封端磁珠。

[0164] 珠可以与量子点或荧光染料关联(例如,用量子点或荧光染料浸渍),以使其在一个荧光光学通道或多个光学通道中发荧光。珠可以与氧化铁或氧化铬关联,以使其为顺磁性或铁磁性。珠是可鉴定的。例如,可以使用照相机对珠进行成像。珠可以具有与珠关联的可检测编码。例如,珠可包含条形码。例如,由于在有机或无机溶液中的膨胀,珠可以改变尺寸。珠可以是疏水的。珠可以是亲水的。珠可以是生物相容的。

[0165] 固体支持物(例如珠)可以被可视化。固体支持物可以包含可视化标签(例如,荧光染料)。固体支持物(例如珠)可以用标识符(例如数字)蚀刻。标识符可以通过对珠进行成像而可视化。

[0166] 固体支持物可以包含半溶性或不溶性物质。当固体支持物包括与其连接的接头、支架、构建单元(building block)或其他反应性部分时,固体支持物可以被称为“官能化的”,而当固体支持物缺少这样的与其连接的反应性部分时,固体支持物可以被称为“非官能化的”。固体支持物可以在溶液中自由使用,诸如在微量滴定孔形式中;在流通形式中,诸如在列中;或者在浸量尺(dipstick)中。

[0167] 固体支持物可以包括膜、纸、塑料、涂层表面、平坦表面、玻璃、载玻片、芯片,或它们的任何组合。固体支持物可以采取树脂、凝胶、微球或其他几何形状的形式。固体支持物可以包括二氧化硅芯片、微粒、纳米颗粒、板、阵列、毛细管、平坦支持物诸如玻璃纤维过滤器、玻璃表面、金属表面(钢、金、银、铝、硅和铜)、玻璃支持物、塑料支持物、硅支持物、芯片、过滤器、膜、微孔板、载玻片、塑料材料,包括多孔板或膜(例如由聚乙烯、聚丙烯、聚酰胺、聚偏二氟乙烯形成的),和/或晶片(wafers)、梳、针(pins)或针头(needles)(例如适于组合合成或分析的针阵列),或平坦表面诸如晶片(例如硅晶片)的凹点(pit)或纳升孔的阵列中的珠,具有带有或不带有过滤器底的凹点的晶片。

[0168] 固体支持物可以包含聚合物基质(例如,凝胶、水凝胶)。聚合物基质可能能够渗透细胞内空间(例如,细胞器周围)。聚合物基质可能能够被泵送到整个循环系统。

[0169] 基底和微孔阵列

[0170] 如本文使用的,基底可以指一种类型的固体支持物。基底可以指可包含本公开内容的条形码或随机条形码的固体支持物。例如,基底可以包括多于一个微孔。例如,基底可以是包括两个或更多个微孔的孔阵列。在一些实施方案中,微孔可以包括限定体积的小反应室。在一些实施方案中,微孔可以截留一个或更多个细胞。在一些实施方案中,微孔仅能截留一个细胞。在一些实施方案中,微孔可以截留一种或更多种固体支持物。在一些实施方案中,微孔仅能截留一种固体支持物。在一些实施方案中,微孔截留单细胞和单个固体支持物(例如珠)。微孔可以包含本公开内容的条形码试剂。

[0171] 条形码化的方法

[0172] 本公开内容提供了用于估计身体样品(例如,组织、器官、肿瘤、细胞)中的不同位置处的不同靶的数量的方法。所述方法可以包括将条形码(例如,随机条形码)靠近样品放置,裂解样品,将不同靶与条形码关联,对靶进行扩增和/或对靶进行数字计数。所述方法还可以包括对从条形码上的空间标记获得的信息进行分析和/或可视化。在一些实施方案中,所述方法包括使样品中的多于一种靶可视化。将多于一种靶映射到样品的地图上可以包括生成样品的二维地图或三维地图。可以在对样品中的多于一种靶进行条形码化(例如,随机条形码化)之前或之后生成二维地图和三维地图。使样品中的多于一种靶可视化可以包括将多于一种靶映射到样品的地图上。将多于一种靶映射到样品的地图上可以包括生成样品的二维地图或三维地图。可以在对样品中的多于一种靶进行条形码化之前或之后生成二维地图和三维地图。在一些实施方案中,可以在裂解样品之前或之后生成二维地图和三维地图。在生成二维地图或三维地图之前或之后裂解样品可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。

[0173] 在一些实施方案中,对多于一种靶进行条形码化包括使多于一种条形码与多于一种靶杂交以产生条形码化靶(例如,随机条形码化靶)。对多于一种靶进行条形码化可以包括生成条形码化靶的索引文库。生成条形码化靶的索引文库可以用包含多于一种条形码(例如,随机条形码)的固体支持物进行。

[0174] 使样品和条形码接触

[0175] 本公开内容提供了用于使样品(例如,细胞)与本公开内容的基底接触的方法。可以使包括例如细胞、器官或组织薄切片的样品与条形码(例如,随机条形码)接触。例如,通过重力流可以接触所述细胞,其中可以使所述细胞沉淀并且产生单层。样品可以是组织薄切片。可以将薄切片置于基底上。样品可以是一维的(例如,形成平坦表面)。样品(例如,细胞)可以散布于基底上,例如,通过在基底上生长/培养所述细胞。

[0176] 当条形码靠近靶时,靶可以与条形码杂交。条形码可以按不可耗尽的比率接触,使得每种不同的靶可以与本公开内容的不同条形码关联。为了确保靶与条形码之间的有效关联,可以将靶与条形码交联。

[0177] 细胞裂解

[0178] 在细胞和条形码的分布之后,可以裂解细胞以释放靶分子。细胞裂解可以通过各种手段中的任何一种来完成,例如通过化学或生物化学手段,通过渗透冲击,或通过热裂

解、机械裂解或光学裂解的手段。可以通过添加包含去污剂(例如,SDS、十二烷基硫酸锂、Triton X-100、吐温-20或NP-40)的细胞裂解缓冲液、有机溶剂(例如甲醇或丙酮)或消化酶(例如蛋白酶K、胃蛋白酶或胰蛋白酶),或它们的任何组合来裂解细胞。为了增加靶和条形码的关联,可通过例如降低裂解物的温度和/或增加裂解物的粘度来改变靶分子的扩散速率。

[0179] 在一些实施方案中,样品可以使用滤纸裂解。滤纸可以用滤纸之上的裂解缓冲液浸泡。滤纸可以用压力施加到样品上,该压力可以促进样品的裂解和样品的靶与基底的杂交。

[0180] 在一些实施方案中,裂解可以通过机械裂解、热裂解、光学裂解和/或化学裂解进行。化学裂解可以包括使用消化酶,诸如蛋白酶K、胃蛋白酶和胰蛋白酶。裂解可以通过将裂解缓冲液添加到基底来进行。裂解缓冲液可以包含Tris HCl。裂解缓冲液可以包含至少约0.01M、0.05M、0.1M、0.5M或1M或更多的Tris HCl。裂解缓冲液可以包含至多约0.01M、0.05M、0.1M、0.5M或1M或更多的Tris HCl。裂解缓冲液可以包含约0.1M Tris HCl。裂解缓冲液的pH可以是至少约1、2、3、4、5、6、7、8、9、10或更高。裂解缓冲液的pH可以是至多约1、2、3、4、5、6、7、8、9、10或更高。在一些实施方案中,裂解缓冲液的pH为约7.5。裂解缓冲液可以包含盐(例如LiCl)。裂解缓冲液中的盐浓度可以是至少约0.1、0.5或1M或更高。裂解缓冲液中的盐浓度可以是至多约0.1、0.5或1M或更高。在一些实施方案中,裂解缓冲液中的盐浓度为约0.5M。裂解缓冲液可以包含去污剂(例如SDS、十二烷基硫酸锂、triton X、吐温、NP-40)。裂解缓冲液中去污剂的浓度可以是至少约0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、3%、4%、5%、6%或7%或更多。裂解缓冲液中去污剂的浓度可以是至多约0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、3%、4%、5%、6%或7%或更多。在一些实施方案中,裂解缓冲液中去污剂的浓度为约1%的十二烷基硫酸锂。裂解方法中使用的时间可以取决于使用的去污剂的量。在一些实施方案中,使用的去污剂越多,裂解所需的时间越少。裂解缓冲液可以包含螯合剂(例如,EDTA、EGTA)。裂解缓冲液中螯合剂的浓度可以是至少约1mM、5mM、10mM、15mM、20mM、25mM或30mM或更高。裂解缓冲液中螯合剂的浓度可以是至多约1mM、5mM、10mM、15mM、20mM、25mM或30mM或更高。在一些实施方案中,裂解缓冲液中螯合剂的浓度是约10mM。裂解缓冲液可以包含还原剂(例如, β -巯基乙醇、DTT)。裂解缓冲液中还原剂的浓度可以是至少约1mM、5mM、10mM、15mM或20mM或更高。裂解缓冲液中还原剂的浓度可以是至多约1mM、5mM、10mM、15mM或20mM或更高。在一些实施方案中,裂解缓冲液中还原试剂的浓度是约5mM。在一些实施方案中,裂解缓冲液可以包含约0.1M TrisHCl,约pH 7.5、约0.5M LiCl、约1%十二烷基硫酸锂、约10mM EDTA和约5mM DTT。

[0181] 裂解可以在约4°C、10°C、15°C、20°C、25°C或30°C的温度进行。裂解可以进行约1分钟、5分钟、10分钟、15分钟或20分钟或更长时间。裂解的细胞可以包括至少约100000个、200000个、300000个、400000个、500000个、600000个或700000个或更多个靶核酸分子。裂解的细胞可以包括至多约100000个、200000个、300000个、400000个、500000个、600000个或700000个或更多个靶核酸分子。

[0182] 将条形码附接至靶核酸分子

[0183] 在细胞裂解和核酸分子从细胞释放之后,核酸分子可以随机地与共定位的固体支

持物的条形码关联。关联可以包括使条形码的靶识别区与靶核酸分子的互补部分杂交(例如,条形码的寡聚(dT)可以与靶的多(A)尾相互作用)。可以选择用于杂交的测定条件(例如,缓冲液pH、离子强度、温度等)以促进形成特定的稳定的杂交体。在一些实施方案中,从裂解的细胞释放的核酸分子可以与基底上的多于一个探针关联(例如,与基底上的探针杂交)。当探针包括寡聚(dT)时,mRNA分子可以与探针杂交,并且进行逆转录。寡核苷酸的寡聚(dT)部分可以充当用于cDNA分子的第一链合成的引物。例如,在图2中图示的条形码化的非限制性实例中,在框216处,mRNA分子可以与珠上的条形码杂交。例如,单链的核苷酸片段可以与条形码的靶结合区杂交。

[0184] 附接还可以包括将条形码的靶识别区与靶核酸分子的一部分连接。例如,靶结合区可以包括可能能够与限制性位点突出端(例如,EcoRI粘性末端突出端)进行特异性杂交的核酸序列。测定程序还可以包括用限制性酶(例如,EcoRI)处理靶核酸以产生限制性位点突出端。然后条形码可以连接到包括与限制性位点突出端互补的序列的任何核酸分子。连接酶(例如,T4DNA连接酶)可以用于连接两个片段。

[0185] 例如,在图2中图示的条形码化的非限制性实例中,在框220处,来自多于一个细胞(或多于一个样品)的经标记的靶(例如,靶-条形码分子)可以随后被汇集,例如汇集至管中。经标记的靶可以通过例如将条形码和/或附接靶-条形码分子的珠取回(retrieving)来汇集。

[0186] 附接的靶-条形码分子的基于固体支持物的集合的取回可以通过使用磁珠和外部施加的磁场来实施。在所述靶-条形码分子已经汇集后,所有进一步的处理可以在单个反应容器中进行。进一步的处理可以包括例如逆转录反应、扩增反应、裂解反应、解离反应和/或核酸延伸反应。进一步的处理反应可以在微孔内进行,即,不首先汇集来自多于一个细胞的经标记的靶核酸分子。

[0187] 逆转录

[0188] 本公开内容提供了使用逆转录来产生靶-条形码缀合物的方法(在图2的框224处)。靶-条形码缀合物可以包括条形码以及靶核酸(即,条形码化cDNA分子,诸如随机条形码化cDNA分子)的全部或一部分的互补序列。关联的RNA分子的逆转录可以通过添加逆转录引物连同逆转录酶一起而发生。逆转录引物可以是寡聚(dT)引物、随机六核苷酸引物或靶特异性寡核苷酸引物。寡聚(dT)引物的长度可以是12-18个核苷酸,或可以是约12-18个核苷酸,并与哺乳动物mRNA的3'末端处的内源性多(A)尾结合。随机六核苷酸引物可在多个互补位点处与mRNA结合。靶特异性寡核苷酸引物通常选择性地引发感兴趣的mRNA。

[0189] 在一些实施方案中,标记的RNA分子的逆转录可以通过添加逆转录引物而发生。在一些实施方案中,所述逆转录引物是寡聚(dT)引物、随机六核苷酸引物或靶特异性寡核苷酸引物。通常,寡聚(dT)引物的长度为12-18个核苷酸,并与哺乳动物mRNA的3'末端处的内源性多(A)尾结合。随机六核苷酸引物可以在多个互补位点处与mRNA结合。靶特异性寡核苷酸引物通常选择性地引发感兴趣的mRNA。

[0190] 逆转录可以重复地发生以产生多个经标记的cDNA分子。本文公开的方法可以包括进行至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个逆转录反应。

[0191] 扩增

[0192] 可以进行一个或更多个核酸扩增反应(例如,在图2的框228处)以产生经标记的靶核酸分子的多个拷贝。扩增可以以多重方式(in a multiplexed manner)进行,其中多种靶核酸序列同时进行扩增。扩增反应可用于将测序衔接子添加至核酸分子。扩增反应可以包括扩增样品标记(如果存在)的至少一部分。扩增反应可以包括扩增细胞标记和/或条形码序列(例如,分子标记)的至少一部分。扩增反应可以包括扩增样品标签、细胞标记、空间标记、条形码序列(例如,分子标记)、靶核酸,或它们的组合的至少一部分。扩增反应可以包括扩增多于一个核酸的0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、100%,或这些值中的任何两个值之间的数字或范围。所述方法还可以包括进行一个或更多个cDNA合成反应以产生包含样品标记、细胞标记、空间标记和/或条形码序列(例如,分子标记)的靶-条形码分子的一个或更多个cDNA拷贝。

[0193] 在一些实施方案中,可以使用聚合酶链式反应(PCR)进行扩增。如本文使用的,PCR可以指用于通过DNA的互补链的同时引物延伸使特定DNA序列体外扩增的反应。如本文使用的,PCR可以包括所述反应的派生形式,包括但不限于RT-PCR、实时PCR、巢式PCR、定量PCR、多重PCR、数字PCR和组装PCR。

[0194] 经标记的核酸的扩增可以包括非基于PCR的方法。非基于PCR的方法的实例包括但不限于多重置换扩增(MDA)、转录介导的扩增(TMA)、基于核酸序列的扩增(NASBA)、链置换扩增(SDA)、实时SDA、滚环扩增或环到环扩增(circle-to-circle amplification)。其他非基于PCR的扩增方法包括DNA依赖性RNA聚合酶驱动的RNA转录扩增或RNA指导的DNA合成和转录的多个循环以扩增DNA或RNA靶、连接酶链式反应(LCR)和QB复制酶(QB)方法、回文探针的使用、链置换扩增、使用限制性核酸内切酶的寡核苷酸驱动的扩增、使引物与核酸序列杂交并且将所得双链体在延伸反应和扩增之前裂解的扩增方法、使用缺乏5'核酸外切酶活性的核酸聚合酶的链置换扩增、滚环扩增和分支延伸扩增(RAM)。在一些实施方案中,扩增不产生环化转录物。

[0195] 在一些实施方案中,本文公开的方法还包括对经标记的核酸(例如,经标记的RNA、经标记的DNA、经标记的cDNA)进行聚合酶链式反应以产生经标记的扩增子(例如,经随机标记的扩增子)。所述经标记的扩增子可以是双链分子。双链分子可以包括双链RNA分子、双链DNA分子或者与DNA分子杂交的RNA分子。双链分子的一条或两条链可以包含样品标记、空间标记、细胞标记和/或条形码序列(例如,分子标记)。所述经标记的扩增子可以是单链分子。单链分子可以包括DNA、RNA或它们的组合。本公开内容的核酸可以包括合成的或改变的核酸。

[0196] 扩增可以包括使用一个或更多个非天然核苷酸。非天然核苷酸可以包括光不稳定或可触发的核苷酸。非天然核苷酸的实例可以包括但不限于肽核酸(PNA)、吗啉代核酸和锁核酸(LNA),以及二醇核酸(GNA)与苏糖核酸(TNA)。可以将非天然核苷酸添加至扩增反应的一个或更多个循环中。添加非天然核苷酸也可以用于鉴定扩增反应中特定循环或时间点的产物。

[0197] 进行一个或更多个扩增反应可以包括使用一种或更多种引物。一种或更多种引物可以包括例如,1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个或15

个或更多个核苷酸。一个或更多个引物可以包括至少1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个或15个或更多个核苷酸。一种或更多种引物可以包括少于12-15个核苷酸。一种或更多种引物可以退火至多于一种经标记的靶(例如,随机标记的靶)的至少一部分。一种或更多种引物可以退火至多于一种经标记的靶的3'末端和/或5'末端。一种或更多种引物可以退火至多于一种经标记的靶的内部区域。内部区域可以与多于一种经标记的靶的3'末端距离至少约50个、100个、150个、200个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸。一种或更多种引物可以包括一组固定的引物。一种或更多种引物可以包括至少一种或更多种定制引物。一种或更多种引物可以包括至少一种或更多种对照引物。一种或更多种引物可以包括至少一种或更多种基因特异性引物。

[0198] 一种或更多种引物可以包括通用引物。通用引物可以退火至通用引物结合位点。一种或更多种定制引物可以退火至第一样品标记、第二样品标记、空间标记、细胞标记、条形码序列(例如,分子标记)、靶,或它们的任何组合。一种或更多种引物可以包括通用引物和定制引物。定制引物可以被设计用于扩增一种或更多种靶。靶可以包括一个或更多个样品中总核酸的子集。靶可以包括一个或更多个样品中总的经标记的靶的子集。一种或更多种引物可以包括至少96种或更多种定制引物。一种或更多种引物可以包括至少960种或更多种定制引物。一种或更多种引物可以包括至少9600种或更多种定制引物。一种或更多种定制引物可以退火至两种或更多种不同的经标记的核酸。两种或更多种不同的经标记的核酸可以对应于一个或更多个基因。

[0199] 可以在本公开内容的方法中使用任何扩增方案。例如,在一个方案中,第一轮PCR可以使用基因特异性引物和针对通用Illumina测序引物1序列的引物来扩增附接至珠的分子。第二轮PCR可以使用侧翼为Illumina测序引物2序列的巢式基因特异性引物,和针对通用Illumina测序引物1序列的引物扩增第一PCR产物。第三轮PCR添加P5和P7以及样品索引,以使PCR产物转变为Illumina测序文库。使用150bp x 2测序进行测序可以揭示读段(read)1上的细胞标记和条形码序列(例如,分子标记)、读段2上的基因,以及索引1读段上的样品索引。

[0200] 在一些实施方案中,核酸可以使用化学裂解从基底去除。例如,存在于核酸中的化学基团或经修饰的碱基可以用于促进将其从固体支持物去除。例如,酶可以用于将核酸从基底去除。例如,核酸可以通过限制性核酸内切酶消化从基底去除。例如,使用尿嘧啶-d-糖基化酶(UDG)处理含有dUTP或ddUTP的核酸可以将核酸从基底去除。例如,核酸可以使用进行核苷酸切除的酶(诸如,碱基切除修复酶,诸如无嘌呤/无嘧啶(AP)核酸内切酶)将核酸从基底去除。在一些实施方案中,核酸可以使用可光裂解(photocleavable)基团以及光从基底去除。在一些实施方案中,可以使用可裂解接头将核酸从基底去除。例如,可裂解接头可以包括以下中的至少一种:生物素/亲和素、生物素/链霉抗生物素蛋白、生物素/中性链亲和素(neutravidin)、Ig蛋白A、光不稳定性接头、酸或碱不稳定性接头基团,或适体。

[0201] 当探针是基因特异性时,可以使分子与探针杂交,并且进行逆转录和/或扩增。在一些实施方案中,在核酸已经合成(例如,逆转录)之后,可以将其扩增。扩增可以以多重方

式进行,其中多种靶核酸序列同时进行扩增。扩增可以将测序衔接子添加至核酸。

[0202] 在一些实施方案中,可以例如用桥式扩增在基底上进行扩增。可以对cDNA加同聚物尾,以便产生相容末端,用于使用基底上的寡聚(dT)探针进行桥式扩增。在桥式扩增中,与模板核酸的3'末端互补的引物可以是共价附接到固体颗粒的每对引物的第一引物。当含有模板核酸的样品与颗粒接触并进行单个热循环时,可以使模板分子退火至第一引物,并且第一引物通过添加核苷酸而在正向方向上延伸以形成双链体分子,所述双链体分子由模板分子和与模板互补的新形成的DNA链组成。在下一循环的加热步骤中,可以使双链体分子变性,从颗粒释放模板分子,并通过第一引物将互补DNA链附接到颗粒。在随后的退火和延伸步骤的退火阶段中,互补链可以与第二引物杂交,所述第二引物在从第一引物去除的位置处与互补链的区段(segment)互补。这种杂交可以导致互补链在第一引物和第二引物之间形成桥,所述桥通过共价键固定到第一引物并通过杂交固定到第二引物。在延伸阶段,通过在相同的反应混合物中添加核苷酸,第二引物可以在反向方向上延伸,从而将桥转化为双链桥。然后开始下一个循环,并且可以使双链桥变性以产生两个单链核酸分子,每个单链核酸分子的一个末端分别经第一引物和第二引物附接到颗粒表面,其中每个单链核酸分子的另一个末端是未附接的。在该第二个循环的退火和延伸步骤中,每条链可以在相同的颗粒上与先前未使用的另外的互补引物杂交,以形成新的单链桥。使现在杂交的两个先前未使用的引物延伸,以将两个新桥转换成双链桥。

[0203] 扩增反应可包括扩增多于一种核酸的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%。

[0204] 对经标记的核酸的扩增可以包括基于PCR的方法或非基于PCR的方法。对经标记的核酸的扩增可以包括对经标记的核酸的指数式扩增。对经标记的核酸的扩增可以包括对经标记的核酸的线性扩增。扩增可以通过聚合酶链式反应(PCR)进行。PCR可以指用于通过DNA的互补链的同时引物延伸使特定DNA序列体外扩增的反应。PCR可涵盖所述反应的派生形式,包括但不限于,RT-PCR、实时PCR、巢式PCR、定量PCR、多重PCR、数字PCR、阻抑PCR、半阻抑PCR和组装PCR。

[0205] 在一些实施方案中,对经标记的核酸的扩增包括非基于PCR的方法。非基于PCR的方法的实例包括但不限于多重置换扩增(MDA)、转录介导的扩增(TMA)、基于核酸序列的扩增(NASBA)、链置换扩增(SDA)、实时SDA、滚环扩增或环到环扩增。其他非基于PCR的扩增方法包括DNA依赖性RNA聚合酶驱动的RNA转录扩增或RNA指导的DNA合成和转录的多个循环以扩增DNA或RNA靶、连接酶链式反应(LCR)、Q β 复制酶(Q β)、回文探针的使用、链置换扩增、使用限制性核酸内切酶的寡核苷酸驱动的扩增、使引物与核酸序列杂交并且将所得双链体在延伸反应和扩增之前裂解的扩增方法、使用缺乏5'核酸外切酶活性的核酸聚合酶的链置换扩增、滚环扩增和/或分支延伸扩增(RAM)。

[0206] 在一些实施方案中,本文公开的方法还包括对扩增的扩增子(例如,靶)进行巢式聚合酶链式反应。扩增子可以是双链分子。双链分子可以包括双链RNA分子、双链DNA分子或者与DNA分子杂交的RNA分子。双链分子的一条或两条链可以包含样品标签或分子标识符标记。可替代地,扩增子可以是单链分子。单链分子可以包括DNA、RNA或它们的组合。本发明的核酸可以包括合成的或改变的核酸。

[0207] 在一些实施方案中,所述方法包括反复扩增经标记的核酸以产生多种扩增子。本文公开的方法可以包括进行至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个扩增反应。可替代地,所述方法包括进行至少约25个、30个、35个、40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个扩增反应。

[0208] 扩增还可以包括将一个或多个对照核酸添加至一个或多个包括多于一个核酸的样品中。扩增还可以包括将一个或多个对照核酸添加至多于一个核酸中。对照核酸可以包括对照标记。

[0209] 扩增可以包括使用一个或多个非天然核苷酸。非天然核苷酸可以包括光不稳定和/或可触发的核苷酸。非天然核苷酸的实例包括但不限于肽核酸(PNA)、吗啉代核酸和锁核酸(LNA),以及二醇核酸(GNA)与苏糖核酸(TNA)。可以将非天然核苷酸添加至扩增反应的一个或多个循环中。添加非天然核苷酸可以用于鉴定扩增反应中特定循环或时间点的产物。

[0210] 进行一个或多个扩增反应可以包括使用一种或更多种引物。一种或更多种引物可以包括一种或更多种寡核苷酸。一种或更多种寡核苷酸可以包括至少约7-9个核苷酸。一种或更多种寡核苷酸可以包括少于12-15个核苷酸。一种或更多种引物可以退火至多于一种经标记的核酸的至少一部分。一种或更多种引物可以退火至多于一种经标记的核酸的3'末端和/或5'末端。一种或更多种引物可以退火至多于一种经标记的核酸的内部区域。内部区域可以与多于一种经标记的核酸的3'末端距离至少约50个、100个、150个、200个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸。一种或更多种引物可以包括一组固定的引物。一种或更多种引物可以包括至少一种或更多种定制引物。一种或更多种引物可以包括至少一种或更多种对照引物。一种或更多种引物可以包括至少一种或更多种管家基因引物。一种或更多种引物可以包括通用引物。通用引物可以退火至通用引物结合位点。一种或更多种定制引物可以退火至第一样品标签、第二样品标签、分子标识符标记、核酸或它们的产物。一种或更多种引物可以包括通用引物和定制引物。定制引物可以被设计成扩增一种或更多种靶核酸。靶核酸可以包括一个或多个样品中总核酸的子集。在一些实施方案中,引物是附接到本公开内容的阵列的探针。

[0211] 在一些实施方案中,使样品中的多于一种靶条形码化(例如,随机条形码化)还包括生成条形码化靶(例如,随机条形码化靶)的索引文库或所述靶的条形码化片段的索引文库。不同条形码的条形码序列(例如,不同的随机条形码的分子标记)可以彼此不同。生成条形码化靶的索引文库包括从样品中的多于一种靶生成多于一种索引多核苷酸。例如,对于包括第一索引靶和第二索引靶的条形码化靶的索引文库,第一索引多核苷酸的标记区与第二索引多核苷酸的标记区可以具有以下,具有约以下,具有至少以下,或具有至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个核苷酸的差异,或这些值中的任何两个值之间的数字或范围的核苷酸的差异。在一些实施方案中,生成条形码化靶的索引文库包括使多于一种靶(例如mRNA分子)与包括多(T)区和标记区的多于一种寡

核苷酸接触;以及使用逆转录酶进行第一链合成以产生各自包含cDNA区和标记区的单链标记的cDNA分子,其中多于一种靶包括不同序列的至少两种mRNA分子,且多于一种寡核苷酸包括不同序列的至少两种寡核苷酸。生成条形码化靶的索引文库还可以包括扩增单链标记的cDNA分子以产生双链标记的cDNA分子;以及对双链标记的cDNA分子进行巢式PCR以产生标记的扩增子。在一些实施方案中,所述方法可以包括产生衔接子标记的扩增子。

[0212] 条形码化(例如,随机条形码化)可以包括使用核酸条形码或标签以对单种核酸(例如,DNA或RNA)分子进行标记。在一些实施方案中,其涉及将DNA条形码或标签添加至cDNA分子,因为cDNA分子是从mRNA产生的。可以进行巢式PCR以使PCR扩增偏倚最小化。可以使用例如下一代测序(NGS)添加衔接子,用于测序。例如在图2的框232处,可以使用测序结果以确定靶的一个或多个拷贝的细胞标记、分子标记,和核苷酸片段的序列。

[0213] 图3是示出了生成条形码化靶(例如,随机条形码化靶)的索引文库,诸如条形码化mRNA或其片段的索引文库的非限制性示例性方法的示意图。如步骤1中示出的,逆转录过程可以编码具有独特的分子标记序列、细胞标记序列和通用PCR位点的每个mRNA分子。特别地,通过使一组条形码(例如,随机条形码)310与RNA分子302的多(A)尾区308杂交(例如,随机杂交),可以将RNA分子302逆转录以产生经标记的cDNA分子304(包括cDNA区306)。条形码310中的每一种可以包含靶结合区,例如多(dT)区312、标记区314(例如,条形码序列或分子)和通用PCR区316。

[0214] 在一些实施方案中,细胞标记序列可以包括3个至20个核苷酸。在一些实施方案中,分子标记序列可以包括3个至20个核苷酸。在一些实施方案中,多于一种随机条形码中的每种还包含通用标记和细胞标记中的一种或更多种,其中固体支持物上的多于一种随机条形码的通用标记是相同的,并且固体支持物上的多于一种随机条形码的细胞标记是相同的。在一些实施方案中,通用标记可以包括3个至20个核苷酸。在一些实施方案中,细胞标记包括3个至20个核苷酸。

[0215] 在一些实施方案中,标记区314可以包括条形码序列或分子标记318和细胞标记320。在一些实施方案中,标记区314可以包括通用标记、维度标记和细胞标记中的一种或更多种。条形码序列或分子标记318的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。细胞标记320的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。通用标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。固体支持物上的多于一种随机条形码的通用标记可以是相同的,并且固体支持物上的多于一种随机条形码的细胞标记是相同的。维度标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。

[0216] 在一些实施方案中,标记区314可以包括以下,包括约以下,包括至少以下,或包括

至多以下:1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种,或这些值中的任何两个值之间的数字或范围的不同标记,诸如条形码序列或分子标记318和细胞标记320。每种标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种,或这些值中的任何两个值之间的数字或范围的核苷酸。一组条形码或随机条形码310可以含有以下,含有约以下,含有至少以下,或可以是至多以下:10种、20种、40种、50种、70种、80种、90种、 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种、 10^{10} 种、 10^{11} 种、 10^{12} 种、 10^{13} 种、 10^{14} 种、 10^{15} 种、 10^{20} 种,或这些值中的任何两个值之间的数字或范围的条形码或随机条形码310。并且一组条形码或随机条形码310可以例如,各自含有独特标记区314。经标记的cDNA分子304可以进行纯化以去除过量条形码或随机条形码310。纯化可以包括Ampure珠纯化。

[0217] 如步骤2中示出的,来自步骤1中的逆转录过程的产物可以汇集至1支管中,并且用第1PCR引物池和第1通用PCR引物进行PCR扩增。由于独特标记区314,汇集是可能的。特别地,可以将经标记的cDNA分子304扩增以产生巢式PCR标记的扩增子322。扩增可以包括多重PCR扩增。扩增可以包括在单一反应体积中用96种多重引物进行的多重PCR扩增。在一些实施方案中,在单一反应体积中,多重PCR扩增可以利用以下,利用约以下,利用至少以下,或利用至多以下:10种、20种、40种、50种、70种、80种、90种、 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种、 10^{10} 种、 10^{11} 种、 10^{12} 种、 10^{13} 种、 10^{14} 种、 10^{15} 种、 10^{20} 种,或这些值中的任何两个值之间的数字或范围的多重引物。扩增可以包括使用包括靶向特异性基因的定制引物326A-C的第1PCR引物池324和通用引物328。定制引物326可以与经标记的cDNA分子304的cDNA部分306'内的区域杂交。通用引物328可以与经标记的cDNA分子304的通用PCR区域316杂交。

[0218] 如图3的步骤3中示出的,来自步骤2中的PCR扩增的产物可以用巢式PCR引物池和第2通用PCR引物进行扩增。巢式PCR可以使PCR扩增偏倚最小化。特别地,巢式PCR标记的扩增子322可以通过巢式PCR进行进一步扩增。巢式PCR可以包括在单个反应体积中用巢式PCR引物332a-c的巢式PCR引物池330和第2通用PCR引物328'进行的多重PCR。巢式PCR引物池330可以含有以下,含有约以下,含有至少以下,或含有至多以下:1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种,或这些值中的任何两个值之间的数字或范围的不同巢式PCR引物332。巢式PCR引物332可以含有衔接子334,并与经标记的扩增子322的cDNA部分306"内的区域杂交。通用引物328'可以含有衔接子336,并与经标记的扩增子322的通用PCR区316杂交。因此,步骤3产生衔接子标记的扩增子338。在一些实施方案中,巢式PCR引物332和第2通用PCR引物328'可以不含有衔接子334和336。相反,衔接子334和336可以连接到巢式PCR的产物以产生衔接子标记的扩增子338。

[0219] 如步骤4中示出的,来自步骤3的PCR产物可以使用文库扩增引物进行PCR扩增用于测序。特别地,衔接子334和336可以用于对衔接子标记的扩增子338进行一个或更多个另外的测定。衔接子334和336可以与引物340和342杂交。一种或更多种引物340和342可以是PCR扩增引物。一种或更多种引物340和342可以是测序引物。一种或更多种衔接子334和336可

以用于衔接子标记的扩增子338的进一步扩增。一种或更多种衔接子334和336可以用于对衔接子标记的扩增子338进行测序。引物342可以含有板索引344,使得使用同一组条形码或随机条形码310产生的扩增子可以在一轮测序反应中使用下一代测序(NGS)进行测序。

[0220] 包含与寡核苷酸关联的细胞组分结合试剂的组合物

[0221] 本文公开的一些实施方案提供了多于一种组合物,每种组合物包含与寡核苷酸缀合的细胞组分结合试剂(诸如蛋白质结合试剂),其中寡核苷酸包含用于与其缀合的细胞组分结合试剂的独特标识符。美国专利申请公布第2018/0088112号和美国专利申请第15/937,713号中描述了细胞组分结合试剂(诸如条形码化抗体)及其用途(诸如细胞的样品索引化);每一篇的内容通过引用以其整体并入本文。

[0222] 在一些实施方案中,细胞组分结合试剂能够与细胞组分靶特异性地结合。例如,细胞组分结合试剂的结合靶可以是以下或包含以下:碳水化合物、脂质、蛋白质、细胞外蛋白质、细胞表面蛋白质、细胞标志物、B细胞受体、T细胞受体、主要组织相容性复合体、肿瘤抗原、受体、整联蛋白、细胞内蛋白质,或它们的任何组合。在一些实施方案中,细胞组分结合试剂(例如,蛋白质结合试剂)能够与抗原靶或蛋白质靶特异性地结合。在一些实施方案中,每种寡核苷酸可以包含条形码,诸如随机条形码。条形码可以包括条形码序列(例如,分子标记)、细胞标记、样品标记,或它们的任何组合。在一些实施方案中,每种寡核苷酸可以包含接头。在一些实施方案中,每种寡核苷酸可以包含寡核苷酸探针的结合位点,诸如多(A)尾。例如,多(A)尾可以例如不被锚定到固体支持物或者被锚定到固体支持物。多(A)尾的长度可以是约10个至50个核苷酸,例如长度是18个核苷酸。寡核苷酸可以包含脱氧核糖核苷酸、核糖核苷酸或二者。

[0223] 独特标识符可以是例如具有任何合适长度的核苷酸序列,例如长度为约4个核苷酸至约200个核苷酸,或25个核苷酸至约45个核苷酸。在一些实施方案中,独特标识符可以具有的长度是以下,是约以下,小于以下,大于以下:4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、8个核苷酸、9个核苷酸、10个核苷酸、15个核苷酸、20个核苷酸、25个核苷酸、30个核苷酸、35个核苷酸、40个核苷酸、45个核苷酸、50个核苷酸、55个核苷酸、60个核苷酸、70个核苷酸、80个核苷酸、90个核苷酸、100个核苷酸、200个核苷酸,或在上述值中的任何两个值之间的范围。

[0224] 在一些实施方案中,独特标识符从一组不同的独特标识符选择。一组不同的独特标识符可以包括以下,或包括约以下:20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、5000种,或在这些值中的任何两个值之间的数字或范围的不同独特标识符。一组不同的独特标识符可以包括至少以下,或包括至多以下:20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种或5000种不同独特标识符。在一些实施方案中,该组独特标识符被设计成与待分析样品的DNA或RNA序列具有最小的序列同源性。在一些实施方案中,该组独特标识符的序列彼此或与其互补体相差以下或相差约以下:1个核苷酸、2个核苷酸、3个核苷酸、4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、8个核苷酸、9个核苷酸、10个核苷酸,或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,该组独特标识符的序列彼此或与其互补体相差至少以下或相差至多以下:1个核苷酸、2个核苷酸、3个核苷酸、4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、

8个核苷酸、9个核苷酸、10个核苷酸。在一些实施方案中,该组独特标识符的序列彼此或与其互补体相差至少3%、至少5%、至少8%、至少10%、至少15%、至少20%或更多。

[0225] 在一些实施方案中,独特标识符可以包括用于引物(诸如通用引物)的结合位点。在一些实施方案中,独特标识符可以包括用于引物(诸如通用引物)的至少两个结合位点。在一些实施方案中,独特标识符可以包括用于引物(诸如通用引物)的至少三个结合位点。引物可以用于扩增独特标识符,例如,通过PCR扩增。在一些实施方案中,引物可以用于巢式PCR反应。

[0226] 任何合适的细胞组分结合试剂在本公开内容中被考虑,诸如蛋白质结合试剂、抗体或其片段、适体、小分子、配体、肽、寡核苷酸等,或它们的任何组合。在一些实施方案中,细胞组分结合试剂可以是多克隆抗体、单克隆抗体、重组抗体、单链抗体(sc-Ab),或它们的片段,诸如Fab、Fv等。在一些实施方案中,多于一种细胞组分结合试剂可以包括以下,或包括约以下:20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、5000种或在这些值中的任何两个值之间的数字或范围的不同细胞组分结合试剂。在一些实施方案中,多于一种细胞组分结合试剂可以包括至少以下,或包括至多以下:20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种或5000种不同细胞组分结合试剂。

[0227] 寡核苷酸可以通过各种机制与细胞组分结合试剂缀合。在一些实施方案中,寡核苷酸可以与细胞组分结合试剂共价地缀合。在一些实施方案中,寡核苷酸可以与细胞组分结合试剂非共价地缀合。在一些实施方案中,寡核苷酸通过接头与细胞组分结合试剂缀合。接头可以是例如可从细胞组分结合试剂和/或寡核苷酸裂解或脱离的。在一些实施方案中,接头可以包含将寡核苷酸可逆地附接到细胞组分结合试剂的化学基团。化学基团可以例如通过胺基团与接头缀合。在一些实施方案中,接头可以包含与另一个缀合到细胞组分结合试剂的化学基团形成稳定键的化学基团。例如,化学基团可以是UV光可裂解基团、二硫键、链霉亲和素、生物素、胺等。在一些实施方案中,化学基团可以通过氨基酸诸如赖氨酸上的伯胺或N-末端与细胞组分结合试剂缀合。商购可得的缀合试剂盒,诸如蛋白质-寡核苷酸缀合试剂盒(Solulink, Inc., San Diego, California)、Thunder-Link®寡核苷酸缀合系统(Innova Biosciences, Cambridge, United Kingdom)等,可以用于将寡核苷酸缀合到细胞组分结合试剂。

[0228] 寡核苷酸可以与细胞组分结合试剂(例如,蛋白质结合试剂)的任何合适的位点缀合,只要它不干扰细胞组分结合试剂与其细胞组分靶之间的特异性结合。在一些实施方案中,细胞组分结合试剂是蛋白质,诸如抗体。在一些实施方案中,细胞组分结合试剂不是抗体。在一些实施方案中,寡核苷酸可以在除了抗原结合位点以外的任何位置与抗体缀合,例如在Fc区、C_H1结构域、C_H2结构域、C_H3结构域、C_L结构域等与抗体缀合。将寡核苷酸与细胞组分结合试剂(例如抗体)缀合的方法先前已经在例如美国专利第6,531,283号中公开,该美国专利的内容通过引用以其整体在此明确并入本文。寡核苷酸与细胞组分结合试剂的化学计量可以变化。为了增加测序中检测细胞组分结合试剂特异性寡核苷酸的灵敏度,在缀合期间增加寡核苷酸与细胞组分结合试剂的比率可以是有利的。在一些实施方案中,每种细胞组分结合试剂可以与单种寡核苷酸分子缀合。在一些实施方案中,每种细胞组分结合试

剂可以与多于一种寡核苷酸分子缀合,例如,与至少以下,或至多以下的寡核苷酸分子缀合:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种或在这些值中的任何两个值之间的数字或范围,其中每种寡核苷酸分子包含相同或不同的独特标识符。在一些实施方案中,每种细胞组分结合试剂可以与多于一种寡核苷酸分子缀合,例如,与至少以下,或至多以下的寡核苷酸分子缀合:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种,其中每种寡核苷酸分子包含相同或不同的独特标识符。

[0229] 在一些实施方案中,多于一种细胞组分结合试剂能够与样品中的多于一种细胞组分靶特异性地结合,所述样品诸如单细胞、多于一个细胞、组织样品、肿瘤样品、血液样品等。在一些实施方案中,多于一种细胞组分靶包括细胞表面蛋白、细胞标志物、B细胞受体、T细胞受体、抗体、主要组织相容性复合体、肿瘤抗原、受体,或它们的任何组合。在一些实施方案中,多于一种细胞组分靶可以包括细胞内细胞组分。在一些实施方案中,多于一种细胞组分靶可以包括细胞外细胞组分。在一些实施方案中,多于一种细胞组分可以是细胞或生物体中所有细胞组分(例如蛋白质)的1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或这些值中的任何两个值之间的数字或范围,或是细胞或生物体中所有细胞组分(例如蛋白质)的约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,多于一种细胞组分可以是细胞或生物体中所有细胞组分(例如,蛋白质)的至少或至多1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%或99%。在一些实施方案中,多于一种细胞组分靶可以包括以下,或包括约以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种、10000种或这些值中的任何两个值之间的数字或范围的不同细胞组分靶。在一些实施方案中,多于一种细胞组分靶可以包括至少以下,或包括至多以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种、10000种不同的细胞组分靶。

[0230] 图4示出了与寡核苷酸缀合的示例性细胞组分结合试剂(例如抗体)的示意性图示,所述寡核苷酸包含用于抗体的独特标识符序列。与细胞组分结合试剂缀合的寡核苷酸、用于与细胞组分结合试剂缀合的寡核苷酸,或先前与细胞组分结合试剂缀合的寡核苷酸,在本文中可被称为抗体寡核苷酸(缩写为结合试剂寡核苷酸)。与抗体缀合的寡核苷酸、用于与抗体缀合的寡核苷酸,或先前与抗体缀合的寡核苷酸,在本文中可被称为抗体寡核苷酸(缩写为“AbOligo”或“AbO”)。寡核苷酸还可以包含另外的组分,包括但不限于一种或更多种接头、一种或更多种用于抗体的独特标识符、任选地一种或更多种条形码序列(例如分子标记),和多(A)尾。在一些实施方案中,寡核苷酸可以从5'至3'包含接头、独特标识符、条形码序列(例如分子标记)和多(A)尾。抗体寡核苷酸可以是mRNA模拟物。

[0231] 图5示出了与寡核苷酸缀合的示例性细胞组分结合试剂(例如抗体)的示意性图示,所述寡核苷酸包含用于抗体的独特标识符序列。细胞组分结合试剂能够与至少一种细胞组分靶,诸如抗原靶或蛋白质靶特异性地结合。结合试剂寡核苷酸(例如,样品索引化寡核苷酸或抗体寡核苷酸)可以包含用于进行本公开内容的方法的序列(例如,样品索引化序列)。例如,样品索引化寡核苷酸可以包含样品索引化序列,用于鉴定样品的一个或更多个细胞的样品来源。多于一种包含细胞组分结合试剂的组合物中的至少两种包含细胞组分结

合试剂的组合物的加索引序列(例如样品索引化序列)可以包含不同的序列。在一些实施方案中,结合试剂寡核苷酸与物种的基因组序列不同源。结合试剂寡核苷酸可以被配置为可从细胞组分结合试剂脱离或不可脱离。

[0232] 与细胞组分结合试剂缀合的寡核苷酸可以例如包含条形码序列(例如,分子标记序列)、多(A)尾,或它们的组合。与细胞组分结合试剂缀合的寡核苷酸可以是mRNA模拟物。在一些实施方案中,样品索引化寡核苷酸包含与多于一种条形码中的至少一种条形码的捕获序列互补的序列。条形码的靶结合区可以包含捕获序列。靶结合区可以例如包含多(dT)区。在一些实施方案中,与条形码的捕获序列互补的样品索引化寡核苷酸的序列可以包含多(A)尾。样品索引化寡核苷酸可以包含分子标记。

[0233] 在一些实施方案中,结合试剂寡核苷酸(例如,样品寡核苷酸)包含长度为以下的核苷酸序列或长度为约以下的核苷酸序列:6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个、110个、120个、128个、130个、140个、150个、160个、170个、180个、190个、200个、210个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、610个、620个、630个、640个、650个、660个、670个、680个、690个、700个、710个、720个、730个、740个、750个、760个、770个、780个、790个、800个、810个、820个、830个、840个、850个、860个、870个、880个、890个、900个、910个、920个、930个、940个、950个、960个、970个、980个、990个、1000个或这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,结合试剂寡核苷酸包含长度为至少以下,或至多以下的核苷酸序列:6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个、110个、120个、128个、130个、140个、150个、160个、170个、180个、190个、200个、210个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、610个、620个、630个、640个、650个、660个、670个、680个、690个、700个、710个、720个、730个、740个、750个、760个、770个、780个、790个、800个、810个、820个、830个、840个、850个、860个、870个、880个、890个、900个、910个、920个、930个、940个、950个、960个、970个、980个、990或1000个核苷酸。

[0234] 在一些实施方案中,细胞组分结合试剂包括抗体、四聚体、适体、蛋白质支架,或它们的组合。结合试剂寡核苷酸可以例如通过接头与细胞组分结合试剂缀合。结合试剂寡核苷酸可以包含接头。接头可以包含化学基团。化学基团可以可逆地或不可逆地附接到细胞组分结合试剂的分子。化学基团可以选自由以下组成的组:UV光可裂解基团、二硫键、链霉亲和素、生物素、胺,以及它们的任何组合。

[0235] 在一些实施方案中,细胞组分结合试剂可以与以下结合:ADAM10、CD156c、AN06、ATP1B2、ATP1B3、BSG、CD147、CD109、CD230、CD29、CD298、ATP1B3、CD44、CD45、CD47、CD51、CD59、CD63、CD97、CD98、SLC3A2、CLDN1、HLA-ABC、ICAM1、ITFG3、MPZL1、NA K ATP酶 α 1、

ATP1A1、NPTN、PMCA ATP酶、ATP2B1、SLC1A5、SLC29A1、SLC2A1、SLC44A2,或它们的任何组合。

[0236] 在一些实施方案中,蛋白质靶是,或包括,细胞外蛋白质、细胞内蛋白质,或它们的任何组合。在一些实施方案中,抗原或蛋白质靶是,或包括,细胞表面蛋白质、细胞标志物、B细胞受体、T细胞受体、主要组织相容性复合体、肿瘤抗原、受体、整联蛋白,或它们的任何组合。抗原或蛋白质靶可以是,或包括,脂质、碳水化合物,或它们的任何组合。蛋白质靶可以选自包含许多蛋白质靶的组。抗原靶或蛋白质靶的数量可以是以下,或是约以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000或这些值中的任何两个值之间的数字或范围。蛋白质靶的数量可以是至少以下,或至多以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000或10000。

[0237] 细胞组分结合试剂(例如,蛋白质结合试剂)可以与具有相同序列的两种或更多种结合试剂寡核苷酸(例如,样品索引化寡核苷酸)关联。细胞组分结合试剂可以与具有不同序列的两种或更多种结合试剂寡核苷酸关联。在不同的实施方式中,与细胞组分结合试剂关联的结合试剂寡核苷酸的数量可以不同。在一些实施方案中,具有相同序列或不同序列的结合试剂寡核苷酸的数量可以是以下,或是约以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000或这些值中的任何两个值之间的数字或范围。在一些实施方案中,结合试剂寡核苷酸的数量可以是至少以下,或至多以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000。

[0238] 包含细胞组分结合试剂的多于一种组合物(例如,多于一种样品索引化组合物)可以包含在本文中也称为不含结合试剂寡核苷酸的细胞组分结合试剂(诸如不含样品索引化寡核苷酸的细胞组分结合试剂)的不与结合试剂寡核苷酸缀合的一种或更多种另外的细胞组分结合试剂(诸如样品索引化寡核苷酸)。在不同的实施方式中,多于一种组合物中另外的细胞组分结合试剂的数量可以不同。在一些实施方案中,另外的细胞组分结合试剂的数量可以是以下,或是约以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100或这些值中的任何两个值之间的数字或范围。在一些实施方案中,另外的细胞组分结合试剂的数量可以是至少以下,或至多以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100。在一些实施方案中,细胞组分结合试剂和另外的细胞组分结合试剂中的任一种可以是相同的。

[0239] 在一些实施方案中,提供了包含以下的混合物:与一种或更多种结合试剂寡核苷酸(例如,样品索引化寡核苷酸)缀合的一种或更多种细胞组分结合试剂,和不与结合试剂寡核苷酸缀合的一种或更多种细胞组分结合试剂。混合物可以用于本文公开的方法的一些实施方案中,例如,以接触一种或更多种样品和/或一种或更多种细胞。在不同的实施方式中,混合物中以下(1)与(2)的比率可以不同:(1)与结合试剂寡核苷酸缀合的细胞组分结合试剂的数量,与(2)不与结合试剂寡核苷酸(例如样品索引化寡核苷酸)或一种或更多种其他结合试剂寡核苷酸缀合的另一种细胞组分结合试剂(例如相同的细胞组分结合试剂)的数量。在一些实施方案中,该比率可以是以下,或是约以下:1:1、1:1.1、1:1.2、1:1.3、1:1.4、1:1.5、1:1.6、1:1.7、1:1.8、1:1.9、1:2、1:2.5、1:3、1:4、1:5、1:6、1:7、1:8、1:9、1:10、

1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:21、1:22、1:23、1:24、1:25、1:26、1:27、1:28、1:29、1:30、1:31、1:32、1:33、1:34、1:35、1:36、1:37、1:38、1:39、1:40、1:41、1:42、1:43、1:44、1:45、1:46、1:47、1:48、1:49、1:50、1:51、1:52、1:53、1:54、1:55、1:56、1:57、1:58、1:59、1:60、1:61、1:62、1:63、1:64、1:65、1:66、1:67、1:68、1:69、1:70、1:71、1:72、1:73、1:74、1:75、1:76、1:77、1:78、1:79、1:80、1:81、1:82、1:83、1:84、1:85、1:86、1:87、1:88、1:89、1:90、1:91、1:92、1:93、1:94、1:95、1:96、1:97、1:98、1:99、1:100、1:200、1:300、1:400、1:500、1:600、1:700、1:800、1:900、1:1000、1:2000、1:3000、1:4000、1:5000、1:6000、1:7000、1:8000、1:9000、1:10000,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,该比率可以是至少以下,或至多以下:1:1、1:1.1、1:1.2、1:1.3、1:1.4、1:1.5、1:1.6、1:1.7、1:1.8、1:1.9、1:2、1:2.5、1:3、1:4、1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:21、1:22、1:23、1:24、1:25、1:26、1:27、1:28、1:29、1:30、1:31、1:32、1:33、1:34、1:35、1:36、1:37、1:38、1:39、1:40、1:41、1:42、1:43、1:44、1:45、1:46、1:47、1:48、1:49、1:50、1:51、1:52、1:53、1:54、1:55、1:56、1:57、1:58、1:59、1:60、1:61、1:62、1:63、1:64、1:65、1:66、1:67、1:68、1:69、1:70、1:71、1:72、1:73、1:74、1:75、1:76、1:77、1:78、1:79、1:80、1:81、1:82、1:83、1:84、1:85、1:86、1:87、1:88、1:89、1:90、1:91、1:92、1:93、1:94、1:95、1:96、1:97、1:98、1:99、1:100、1:200、1:300、1:400、1:500、1:600、1:700、1:800、1:900、1:1000、1:2000、1:3000、1:4000、1:5000、1:6000、1:7000、1:8000、1:9000或1:10000。

[0240] 在一些实施方案中,该比率可以是以下,或是约以下:1:1、1.1:1、1.2:1、1.3:1、1.4:1、1.5:1、1.6:1、1.7:1、1.8:1、1.9:1、2:1、2.5:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、21:1、22:1、23:1、24:1、25:1、26:1、27:1、28:1、29:1、30:1、31:1、32:1、33:1、34:1、35:1、36:1、37:1、38:1、39:1、40:1、41:1、42:1、43:1、44:1、45:1、46:1、47:1、48:1、49:1、50:1、51:1、52:1、53:1、54:1、55:1、56:1、57:1、58:1、59:1、60:1、61:1、62:1、63:1、64:1、65:1、66:1、67:1、68:1、69:1、70:1、71:1、72:1、73:1、74:1、75:1、76:1、77:1、78:1、79:1、80:1、81:1、82:1、83:1、84:1、85:1、86:1、87:1、88:1、89:1、90:1、91:1、92:1、93:1、94:1、95:1、96:1、97:1、98:1、99:1、100:1、200:1、300:1、400:1、500:1、600:1、700:1、800:1、900:1、1000:1、2000:1、3000:1、4000:1、5000:1、6000:1、7000:1、8000:1、9000:1、10000:1,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,该比率可以是至少以下,或至多以下:1:1、1.1:1、1.2:1、1.3:1、1.4:1、1.5:1、1.6:1、1.7:1、1.8:1、1.9:1、2:1、2.5:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、21:1、22:1、23:1、24:1、25:1、26:1、27:1、28:1、29:1、30:1、31:1、32:1、33:1、34:1、35:1、36:1、37:1、38:1、39:1、40:1、41:1、42:1、43:1、44:1、45:1、46:1、47:1、48:1、49:1、50:1、51:1、52:1、53:1、54:1、55:1、56:1、57:1、58:1、59:1、60:1、61:1、62:1、63:1、64:1、65:1、66:1、67:1、68:1、69:1、70:1、71:1、72:1、73:1、74:1、75:1、76:1、77:1、78:1、79:1、80:1、81:1、82:1、83:1、84:1、85:1、86:1、87:1、88:1、89:1、90:1、91:1、92:1、93:1、94:1、95:1、96:1、97:1、98:1、99:1、100:1、200:1、300:1、400:1、500:1、600:1、700:1、800:1、900:1、1000:1、2000:1、3000:1、4000:1、5000:1、6000:1、7000:1、8000:1、9000:1或10000:1。

[0241] 细胞组分结合试剂可以与结合试剂寡核苷酸(例如,样品索引化寡核苷酸)缀合,

或可以不缀合。在一些实施方案中,在包含与结合试剂寡核苷酸缀合的细胞组分结合试剂和与结合试剂寡核苷酸缀合的一种或更多种细胞组分结合试剂的混合物中,与结合试剂寡核苷酸(例如,样品索引化寡核苷酸)缀合的细胞组分结合试剂的百分比可以是以下,或是约以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,混合物中与样品索引化寡核苷酸缀合的细胞组分结合试剂的百分比可以是至少以下,或至多以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。

[0242] 在一些实施方案中,在包含与结合试剂寡核苷酸(例如,样品索引化寡核苷酸)缀合的细胞组分结合试剂和与样品索引化寡核苷酸缀合的细胞组分结合试剂的混合物中,不与结合试剂寡核苷酸(例如,样品索引化寡核苷酸)缀合的细胞组分结合试剂的百分比可以是以下,或是约以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,混合物中不与结合试剂寡核苷酸缀合的细胞组分结合试剂的百分比可以是至少以下,或至多以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、

60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。

[0243] 细胞组分混合物 (Cocktails)

[0244] 在一些实施方案中,细胞组分结合试剂的混合物(例如抗体混合物)可以用于增加本文公开的方法中的标记灵敏度。不受任何特定理论的限制,认为这可能是因为细胞组分表达或蛋白质表达可在细胞类型和细胞状态之间变化,使得寻找标记所有细胞类型的通用细胞组分结合试剂或抗体具有挑战性。例如,细胞组分结合试剂的混合物可以用于允许对更多样品类型进行更灵敏和有效的标记。细胞组分结合试剂的混合物可以包括两种或更多种不同类型的细胞组分结合试剂,例如更宽范围的细胞组分结合试剂或抗体。对不同的细胞组分靶进行标记的细胞组分结合试剂可以汇集在一起以产生足以标记所有细胞类型或一种或更多种感兴趣的细胞类型的混合物。

[0245] 在一些实施方案中,多于一种组合物(例如,样品索引化组合物)中的每一种包含细胞组分结合试剂。在一些实施方案中,多于一种组合物中的一种组合物包含两种或更多种细胞组分结合试剂,其中两种或更多种细胞组分结合试剂中的每一种与结合试剂寡核苷酸(例如,样品索引化寡核苷酸)关联,其中两种或更多种细胞组分结合试剂中的至少一种能够与一种或更多种细胞组分靶中的至少一种特异性地结合。与两种或更多种细胞组分结合试剂关联的结合试剂寡核苷酸的序列可以是相同的。与两种或更多种细胞组分结合试剂关联的结合试剂寡核苷酸的序列可以包含不同的序列。多于一种组合物中的每一种可以包含两种或更多种细胞组分结合试剂。

[0246] 在不同的实施方式中,组合物中不同类型的细胞组分结合试剂(例如,CD147抗体和CD47抗体)的数量可以不同。具有两种或更多种不同类型的细胞组分结合试剂的组合物在本文中可被称为细胞组分结合试剂混合物(例如,样品索引化组合物混合物)。混合物中不同类型的细胞组分结合试剂的数量可以不同。在一些实施方案中,混合物中不同类型的细胞组分结合试剂的数量可以是以下,或是约以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,混合物中不同类型的细胞组分结合试剂的数量可以是至少以下,或至多以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000或100000。不同类型的细胞组分结合试剂可以与具有相同或不同序列(例如,样品索引化序列)的结合试剂寡核苷酸缀合。

[0247] 细胞组分靶的定量分析方法

[0248] 在一些实施方案中,本文公开的方法还可以用于使用本文公开的组合物和寡核苷酸探针对样品中的多于一种细胞组分靶(例如,蛋白质靶)进行定量分析,所述寡核苷酸探针可以将条形码序列(例如,分子标记序列)与细胞组分结合试剂(例如,蛋白质结合试剂)的寡核苷酸关联。细胞组分结合试剂的寡核苷酸可以是以下,或包括以下:抗体寡核苷酸、样品索引化寡核苷酸、细胞鉴定寡核苷酸、对照颗粒寡核苷酸、对照寡核苷酸、相互作用确定寡核苷酸等。在一些实施方案中,样品可以是单细胞、多于一个细胞、组织样品、肿瘤样品、血液样品等。在一些实施方案中,样品可以包括细胞类型的混合物,诸如正常细胞、肿瘤细胞、血细胞、B细胞、T细胞、母体细胞、胎儿细胞等的混合物,或来自不同受试者的细胞的

混合物。

[0249] 在一些实施方案中,样品可以包括被分到个体隔室诸如微孔阵列中的微孔的多于一个单细胞。

[0250] 在一些实施方案中,多于一种细胞组分靶的结合靶(即,细胞组分靶)可以是以下,或包括以下:碳水化合物、脂质、蛋白质、细胞外蛋白质、细胞表面蛋白质、细胞标志物、B细胞受体、T细胞受体、主要组织相容性复合体、肿瘤抗原、受体、整联蛋白、细胞内蛋白质,或它们的任何组合。在一些实施方案中,细胞组分靶是蛋白质靶。在一些实施方案中,多于一种细胞组分靶包括细胞表面蛋白、细胞标志物、B细胞受体、T细胞受体、抗体、主要组织相容性复合体、肿瘤抗原、受体,或它们的任何组合。在一些实施方案中,多于一种细胞组分靶可以包括细胞内细胞组分。在一些实施方案中,多于一种细胞组分可以是生物体中所有编码的细胞组分的至少1%、至少2%、至少3%、至少4%、至少5%、至少6%、至少7%、至少8%、至少9%、至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或更多。在一些实施方案中,多于一种细胞组分靶可以包括至少2种、至少3种、至少4种、至少5种、至少10种、至少20种、至少30种、至少40种、至少50种、至少100种、至少1000种、至少10000种或更多种不同的细胞组分靶。

[0251] 在一些实施方案中,多于一种细胞组分结合试剂与样品接触,用于与多于一种细胞组分靶特异性结合。未结合的细胞组分结合试剂可以通过例如洗涤来去除。在样品包括细胞的实施方案中,可以去除不与细胞特异性结合的任何细胞组分结合试剂。

[0252] 在一些情况下,来自细胞群体的细胞可以被分离(例如,隔离)到本公开内容的基底的孔中。细胞群体可以在分离之前被稀释。可以稀释细胞群体,使得基底的孔的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%容纳单细胞。可以稀释细胞群体,使得基底的孔的至多1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%容纳单细胞。可以稀释细胞群体,使得稀释的群体中的细胞的数量是基底上的孔的数量的1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%,或是基底上的孔的数量的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。可以稀释细胞群体,使得稀释的群体中的细胞的数量是基底上的孔的数量的1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%,或是基底上的孔的数量的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。在一些情况下,稀释细胞群体,使得细胞的数量为基底中的孔的数量的约10%。

[0253] 单细胞在基底的孔中的分布可以遵循泊松分布。例如,基底的孔具有多于一个细胞的概率可以是至少0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%或更大。基底的孔具有多于一个细胞的概率可以是至少0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%或更大。单细胞在基底的孔中的分布可以是随机的。单细胞在基底的孔中的分布可以是非随机的。细胞可以被分离,使得基底的孔仅容纳一个细胞。

[0254] 在一些实施方案中,细胞组分结合试剂可以另外地与荧光分子缀合,以实现将细

胞流式分选到个体隔室中。

[0255] 在一些实施方案中,本文公开的方法提供了使多于一种组合物与样品接触,以与多于一种细胞组分靶特异性结合。应当理解,所使用的条件可以允许细胞组分结合试剂例如抗体与细胞组分靶的特异性结合。在接触步骤之后,可以去除未结合的组合物。例如,在样品包括细胞并且组合物与是细胞表面细胞组分的细胞组分靶诸如细胞表面蛋白质特异性地结合的实施方案中,未结合的组合物可以通过用缓冲液洗涤细胞来去除,使得只有特异性地结合细胞组分靶的组合物与细胞保留在一起。

[0256] 在一些实施方案中,本文公开的方法可以包括使包括条形码序列(诸如分子标记)、细胞标记、样品标记等,或它们的任何组合的寡核苷酸(例如条形码或随机条形码)与以下关联:与细胞组分结合试剂关联的多于一个寡核苷酸。例如,包含条形码的多于一种寡核苷酸探针可以用于与组合物的多于一种寡核苷酸杂交。

[0257] 在一些实施方案中,多于一种寡核苷酸探针可固定在固体支持物上。固体支持物可以是自由漂浮的,例如溶液中的珠。固体支持物可以嵌入半固体或固体阵列中。在一些实施方案中,多于一种寡核苷酸探针可以不固定在固体支持物上。当多于一种寡核苷酸探针非常接近细胞组分结合试剂的多于一种寡核苷酸时,细胞组分结合试剂的多于一种寡核苷酸可以与寡核苷酸探针杂交。寡核苷酸探针可以以不可耗尽的比率接触,使得细胞组分结合试剂的每种不同的寡核苷酸可以与具有本公开内容的不同条形码序列(例如,分子标记)的寡核苷酸探针关联。

[0258] 在一些实施方案中,本文公开的方法提供了使寡核苷酸从与细胞组分靶特异性地结合的细胞组分结合试剂脱离。脱离可以以各种方式进行以使化学基团与细胞组分结合试剂分离,诸如UV光裂解、化学处理(例如二硫苏糖醇处理)、加热、酶处理,或它们的任何组合。使寡核苷酸从细胞组分结合试剂脱离可以在使多于一种寡核苷酸探针与组合物的多于一种寡核苷酸杂交的步骤之前、之后或期间进行。

[0259] 细胞组分和核酸靶的同时定量分析方法

[0260] 在一些实施方案中,本文公开的方法也可用于使用本文公开的组合物和寡核苷酸探针对样品中的多于一种细胞组分靶(例如,蛋白质靶)和多于一种核酸靶分子进行同时定量分析,所述寡核苷酸探针可以使条形码序列(例如,分子标记序列)与细胞组分结合试剂的寡核苷酸和核酸靶分子二者关联。美国专利申请公布20180088112A1中描述了多于一种细胞组分靶和多于一种核酸靶分子的同时定量分析的其他方法;该美国专利申请公布的内容通过引用以其整体并入本文。在一些实施方案中,样品可以是单细胞、多于一个细胞、组织样品、肿瘤样品、血液样品等。在一些实施方案中,样品可以包括细胞类型的混合物,诸如正常细胞、肿瘤细胞、血细胞、B细胞、T细胞、母体细胞、胎儿细胞的混合物,或来自同一受试者或不同受试者的细胞的混合物。

[0261] 在一些实施方案中,样品可以包括被分到个体隔室诸如微孔阵列中的微孔的多于一个单细胞。

[0262] 在一些实施方案中,多于一种细胞组分靶包括细胞表面蛋白、细胞标志物、B细胞受体、T细胞受体、抗体、主要组织相容性复合体、肿瘤抗原、受体,或它们的任何组合。在一些实施方案中,多于一种细胞组分靶可以包括细胞内细胞组分。在一些实施方案中,多于一种细胞组分可以是生物体中所有细胞组分(诸如表达的蛋白质)或生物体的一个或更多个

细胞的以下,或约以下:1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,多于一种细胞组分可以是生物体中所有细胞组分(诸如可能表达的蛋白质)或生物体的一个或更多个细胞的至少以下,或至多以下:1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%或99%。在一些实施方案中,多于一种细胞组分靶可以包括以下,或包括约以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种、10000种或这些值中的任何两个值之间的数字或范围的不同的细胞组分靶。在一些实施方案中,多于一种细胞组分靶可以包括至少以下,或包括至多以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种或10000种不同的细胞组分靶。

[0263] 在一些实施方案中,多于一种细胞组分结合试剂与样品接触,用于与多于一种细胞组分靶特异性结合。未结合的细胞组分结合试剂可以通过例如洗涤来去除。在样品包括细胞的实施方案中,可以去除不与细胞特异性结合的任何细胞组分结合试剂。

[0264] 在一些情况下,来自细胞群体的细胞可以被分离(例如,隔离)到本公开内容的基底的孔中。细胞群体可以在分离之前被稀释。可以稀释细胞群体,使得基底的孔的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%容纳单细胞。可以稀释细胞群体,使得基底的孔的至多1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%容纳单细胞。可以稀释细胞群体,使得稀释的群体中的细胞的数量是基底上的孔的数量的1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%,或是基底上的孔的数量的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。可以稀释细胞群体,使得稀释的群体中的细胞的数量是基底上的孔的数量的1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%,或是基底上的孔的数量的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。在一些情况下,稀释细胞群体,使得细胞的数量为基底中的孔的数量的约10%。

[0265] 单细胞在基底的孔中的分布可以遵循泊松分布。例如,基底的孔具有多于一个细胞的概率可以是至少0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%或更大。基底的孔具有多于一个细胞的概率可以是至少0.1%、0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%或10%或更大。单细胞在基底的孔中的分布可以是随机的。单细胞在基底的孔中的分布可以是非随机的。细胞可以被分离,使得基底的孔仅容纳一个细胞。

[0266] 在一些实施方案中,细胞组分结合试剂可以另外地与荧光分子缀合,以实现将细胞流式分选到个体隔室中。

[0267] 在一些实施方案中,本文公开的方法提供了使多于一种组合物与样品接触,以与多于一种细胞组分靶特异性结合。应当理解,所使用的条件可以允许细胞组分结合试剂例如抗体与细胞组分靶的特异性结合。在接触步骤之后,可以去除未结合的组合物。例如,在样品包括细胞并且组合物与细胞表面上的细胞组分靶诸如细胞表面蛋白质特异性地结合

的实施方案中,未结合的组合物可以通过用缓冲液洗涤细胞来去除,使得只有与细胞组分靶特异性地结合的组合物与细胞保留在一起。

[0268] 在一些实施方案中,本文公开的方法可以提供使多于一种核酸靶分子从样品例如细胞释放。例如,可以裂解细胞以释放多于一种核酸靶分子。细胞裂解可以通过各种方式中的任何一种来完成,例如,通过化学处理、渗透冲击、热处理、机械处理、光学处理,或它们的任何组合。细胞可以通过添加包含去垢剂(例如,SDS、十二烷基硫酸锂、TritonX-100、吐温-20或NP-40)的细胞裂解缓冲液、有机溶剂(例如甲醇或丙酮)或消化酶(例如蛋白酶K、胃蛋白酶或胰蛋白酶)或其任何组合来裂解。

[0269] 本领域普通技术人员应理解,多于一种核酸分子可以包括各种核酸分子。在一些实施方案中,多于一种核酸分子可以包括DNA分子、RNA分子、基因组DNA分子、mRNA分子、rRNA分子、siRNA分子,或它们的组合,并且可以是双链或单链的。在一些实施方案中,多于一种核酸分子包括以下,或包括约以下:100种、1000种、10000种、20000种、30000种、40000种、50000种、100000种、1000000种或这些值中的任何两个值之间的数字或范围的种类。在一些实施方案中,多于一种核酸分子包括至少以下,或包括至多以下:100种、1000种、10000种、20000种、30000种、40000种、50000种、100000种或1000000种种类。在一些实施方案中,多于一种核酸分子可以来自样品,诸如单细胞或多于一个细胞。在一些实施方案中,多于一种核酸分子可以从多于一个样品汇集,诸如从多于一个单细胞汇集。

[0270] 在一些实施方案中,本文公开的方法可以包括使条形码(例如随机条形码)与多于一种核酸靶分子和与细胞组分结合试剂的多于一种寡核苷酸关联,所述条形码(例如随机条形码)可以包含条形码序列(诸如分子标记)、细胞标记、样品标记等,或它们的任何组合。例如,包含随机条形码的多于一种寡核苷酸探针可以用于与多于一种核酸靶分子和组合物中的多于一种寡核苷酸杂交。

[0271] 在一些实施方案中,多于一种寡核苷酸探针可固定在固体支持物上。固体支持物可以是自由漂浮的,例如溶液中的珠。固体支持物可嵌入半固体或固体阵列中。在一些实施方案中,多于一种寡核苷酸探针可以不固定在固体支持物上。当多于一种寡核苷酸探针非常接近多于一种核酸靶分子和细胞组分结合试剂的多于一种寡核苷酸时,多于一种核酸靶分子和细胞组分结合试剂的多于一种寡核苷酸可以与寡核苷酸探针杂交。寡核苷酸探针可以以不可耗尽的比率接触,使得每种不同的核酸靶分子和细胞组分结合试剂的寡核苷酸可以与具有本公开内容的不同条形码序列(例如,分子标记)的寡核苷酸探针关联。

[0272] 在一些实施方案中,本文公开的方法提供了使寡核苷酸从与细胞组分靶特异性地结合的细胞组分结合试剂脱离。脱离可以以各种方式进行以使化学基团与细胞组分结合试剂分离,诸如UV光裂解、化学处理(例如二硫苏糖醇处理)、加热、酶处理,或它们的任何组合。使寡核苷酸从细胞组分结合试剂脱离可以在使多于一种寡核苷酸探针与多于一种核酸靶和组合物中的多于一种寡核苷酸杂交的步骤之前、之后或期间进行。

[0273] 蛋白质和核酸靶的同时定量分析

[0274] 在一些实施方案中,本文公开的方法也可用于多种类型的靶分子例如蛋白质和核酸靶的同时定量分析。例如,靶分子可以是,或包括,细胞组分。图6示出了单细胞中的核酸靶和其他细胞组分靶(例如,蛋白质)二者的同时定量分析的示例性方法的示意性图示。在一些实施方案中,提供了各自包含细胞组分结合试剂诸如抗体的多于一种组合物605、

605b、605c等。与不同细胞组分靶结合的不同细胞组分结合试剂诸如抗体,与不同的独特标识符缀合。接下来,细胞组分结合试剂可以与包含多于一个细胞610的样品一起孵育。不同的细胞组分结合试剂可以特异性地结合细胞表面上的细胞组分,诸如细胞标志物、B细胞受体、T细胞受体、抗体、主要组织相容性复合体、肿瘤抗原、受体,或它们的任何组合。未结合的细胞组分结合试剂可以例如通过用缓冲液洗涤细胞来去除。然后,具有细胞组分结合试剂的细胞可以被分离到多于一个隔室诸如微孔阵列中,其中单个隔室615的尺寸适合于单细胞和单个珠620。每个珠可以包含多于一种寡核苷酸探针和条形码序列(例如分子标记序列),寡核苷酸探针可以包含珠上所有寡核苷酸探针共有的细胞标记。在一些实施方案中,每种寡核苷酸探针可以包含靶结合区,例如多(dT)序列。与细胞组分结合试剂缀合的寡核苷酸625可以使用化学、光学或其他手段从细胞组分结合试剂脱离。细胞可以被裂解635以释放细胞内的核酸,诸如基因组DNA或细胞mRNA 630。细胞mRNA 630、寡核苷酸625或二者可以被珠620上的寡核苷酸探针捕获,例如,通过与多(dT)序列杂交来捕获。逆转录酶可以用于使用细胞mRNA 630和寡核苷酸625作为模板使与细胞mRNA 630和寡核苷酸625杂交的寡核苷酸探针延伸。可以对逆转录酶产生的延伸产物进行扩增和测序。可以对测序读段进行序列的去多重化(demultiplexing)或者细胞标记、条形码(例如分子标记)、基因、细胞组分结合试剂特异性寡核苷酸(例如抗体特异性寡核苷酸)等的鉴定,这可以产生样品中每个单细胞的细胞组分和基因表达的数字表示。

[0275] 条形码的关联

[0276] 与细胞组分结合试剂(例如,抗原结合试剂或蛋白质结合试剂)和/或核酸分子关联的寡核苷酸可以与寡核苷酸探针随机地关联。与细胞组分结合试剂关联的寡核苷酸,在本文中称为结合试剂寡核苷酸,可以是或可以包括本公开内容的寡核苷酸,诸如抗体寡核苷酸、样品索引化寡核苷酸、细胞鉴定寡核苷酸、对照颗粒寡核苷酸、对照寡核苷酸、相互作用确定寡核苷酸等。关联可以例如包括寡核苷酸探针的靶结合区与靶核酸分子和/或蛋白质结合试剂的寡核苷酸的互补部分的杂交。例如,条形码(例如随机条形码)的寡聚(dT)区域可以与靶核酸分子的多(A)尾和/或蛋白质结合试剂的寡核苷酸的多(A)尾相互作用。可以选择用于杂交的测定条件(例如,缓冲液pH、离子强度、温度等)以促进形成特定的稳定杂交体。

[0277] 本公开内容提供了使用逆转录使分子标记与靶核酸和/或与细胞组分结合试剂所关联的寡核苷酸关联的方法。逆转录酶可以使用RNA和DNA二者作为模板。例如,最初缀合在细胞组分结合试剂上的寡核苷酸可以是RNA或DNA碱基,或者二者。除了结合试剂序列的序列或其一部分之外,结合试剂寡核苷酸可以被复制并连接(例如共价地连接)到细胞标记和条形码序列(例如分子标记)。作为另一个实例,除了mRNA分子的序列或其一部分之外,mRNA分子可以被复制并连接(例如共价地连接)到细胞标记和条形码序列(例如分子标记)。

[0278] 在一些实施方案中,分子标记可以通过连接寡核苷酸探针靶结合区和靶核酸分子的一部分和/或与细胞组分结合试剂关联(例如,当前或以前关联)的寡核苷酸的一部分来添加。例如,靶结合区可以包括能够与限制性位点突出端(例如,EcoRI粘性末端突出端)进行特异性杂交的核酸序列。方法还可以包括用限制性酶(例如,EcoRI)处理靶核酸和/或与细胞组分结合试剂关联的寡核苷酸以产生限制性位点突出端。连接酶(例如,T4 DNA连接酶)可以用于连接两个片段。

[0279] 确定独特分子标记序列的数量或存在

[0280] 在一些实施方案中,本文公开的方法包括确定每种独特标识符、每种核酸靶分子和/或每种结合试剂寡核苷酸(例如抗体寡核苷酸)的独特分子标记序列的数量或存在。例如,测序读段可以用于确定每种独特标识符、每种核酸靶分子和/或每种结合试剂寡核苷酸的独特分子标记序列的数量。作为另一个实例,测序读段可以用于确定分子标记序列(诸如测序读段中,与靶关联的分子标记序列、结合试剂寡核苷酸、抗体寡核苷酸、样品索引化寡核苷酸、细胞鉴定寡核苷酸、对照颗粒寡核苷酸、对照寡核苷酸、相互作用确定寡核苷酸等)的存在或不存在。

[0281] 在一些实施方案中,每种独特标识符、每种核酸靶分子和/或每种结合试剂寡核苷酸的独特分子标记序列的数量指示样品中每种细胞组分靶(例如,抗原靶或蛋白质靶)和/或每种核酸靶分子的数量。在一些实施方案中,细胞组分靶的量和其相应核酸靶分子例如mRNA分子的量可以相互比较。在一些实施方案中,可以计算细胞组分靶的量与其相应核酸靶分子例如mRNA分子的量的比率。细胞组分靶可以是,例如,细胞表面蛋白质标志物。在一些实施方案中,细胞表面蛋白质标志物的蛋白质水平和细胞表面蛋白质标志物的mRNA水平之间的比率低。

[0282] 本文公开的方法可以用于各种应用。例如,本文公开的方法可以用于样品的蛋白质组和/或转录组分析。在一些实施方案中,本文公开的方法可以用于鉴定样品中的细胞组分靶和/或核酸靶,即生物标志物。在一些实施方案中,细胞组分靶和核酸靶相互对应,即核酸靶编码细胞组分靶。在一些实施方案中,本文公开的方法可以用于鉴定具有期望的以下比率的细胞组分靶:样品中的细胞组分靶的量与其相应核酸靶分子例如mRNA分子的量之间的比率。在一些实施方案中,比率是以下,或是约以下:0.001、0.01、0.1、1、10、100、1000或这些值中的任何两个值之间的数字或范围。在一些实施方案中,比率是至少以下,或至多以下:0.001、0.01、0.1、1、10、100或1000。在一些实施方案中,本文公开的方法可以用于鉴定样品中其相应核酸靶分子在样品中的量为以下,或为约以下的细胞组分靶:1000、100、10、5、2、1、0或这些值中的任何两个值之间的数字或范围。

[0283] 在一些实施方案中,本文公开的方法可以用于鉴定样品中其相应核酸靶分子在样品中的量多于以下,或少于以下的细胞组分靶:1000、100、10、5、2、1或0。

[0284] 组合物和试剂盒

[0285] 本文公开的一些实施方案提供了用于对样品中多于一种细胞组分(例如蛋白质)和/或多于一种核酸靶分子进行同时定量分析的试剂盒和组合物。在一些实施方案中,试剂盒和组合物可以包含各自与寡核苷酸缀合的多于一种细胞组分结合试剂(例如,多于一种蛋白质结合试剂)和多于一种寡核苷酸探针,其中寡核苷酸包含用于细胞组分结合试剂的独特标识符,其中多于一种寡核苷酸探针中的每一种包含靶结合区、条形码序列(例如,分子标记序列),其中条形码序列来自一组不同的独特条形码序列。在一些实施方案中,每种寡核苷酸可以包含分子标记、细胞标记、样品标记,或它们的任何组合。在一些实施方案中,每种寡核苷酸可以包含接头。在一些实施方案中,每种寡核苷酸可以包含用于寡核苷酸探针的结合位点,诸如多(A)尾。例如,多(A)尾可以是,例如oligoA₁₈(不锚定到固体支持物)或oligoA₁₈V(锚定到固体支持物)。寡核苷酸可以包含DNA残基、RNA残基或二者。

[0286] 本文的公开内容包括多于一种样品索引化组合物。多于一种样品索引化组合物中

的每一种可以包含两种或更多种细胞组分结合试剂。两种或更多种细胞组分结合试剂中的每一种可以与样品索引化寡核苷酸关联。两种或更多种细胞组分结合试剂中的至少一种可以能够与至少一种细胞组分靶特异性地结合。样品索引化寡核苷酸可以包含样品索引化序列,用于鉴定样品的一个或更多个细胞的样品来源。多于一种样品索引化组合物中的至少两种样品索引化组合物的样品索引化序列可以包括不同的序列。

[0287] 本文的公开内容包括用于细胞鉴定的包含样品索引化组合物的试剂盒。在一些实施方案中,两种样品索引化组合物中的每一种包含与样品索引化寡核苷酸关联的细胞组分结合试剂(例如,蛋白质结合试剂),其中细胞组分结合试剂能够与一种或更多种细胞组分靶(例如,一种或更多种蛋白质靶)中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中两种样品索引化组合物的样品索引化序列包含不同的序列。在一些实施方案中,样品索引化寡核苷酸包含分子标记序列、用于通用引物的结合位点,或它们的组合。

[0288] 本文的公开内容包括用于细胞鉴定的试剂盒。在一些实施方案中,试剂盒包含:两种或更多种样品索引化组合物。两种或更多种样品索引化组合物中的每一种可以包含与样品索引化寡核苷酸关联的细胞组分结合试剂(例如,抗原结合试剂),其中细胞组分结合试剂能够与一种或更多种细胞组分靶中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中两种样品索引化组合物的样品索引化序列包含不同的序列。在一些实施方案中,样品索引化寡核苷酸包含分子标记序列、用于通用引物的结合位点,或它们的组合。本文的公开内容包括用于多重鉴定的试剂盒。在一些实施方案中,试剂盒包含两种样品索引化组合物。两种样品索引化组合物中的每一种可以包含与样品索引化寡核苷酸关联的细胞组分结合试剂(例如,抗原结合试剂),其中抗原结合试剂能够与一种或更多种细胞组分靶(例如抗原靶)中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中两种样品索引化组合物的样品索引化序列包含不同的序列。

[0289] 独特标识符(或与细胞组分结合试剂关联的寡核苷酸,诸如结合试剂寡核苷酸、抗体寡核苷酸、样品索引化寡核苷酸、细胞鉴定寡核苷酸、对照颗粒寡核苷酸、对照寡核苷酸,或相互作用确定寡核苷酸)可以具有任何合适的长度,例如,约25个核苷酸至约45个核苷酸长。在一些实施方案中,独特标识符可以具有的长度为以下,为约以下,小于以下,大于以下:4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、8个核苷酸、9个核苷酸、10个核苷酸、15个核苷酸、20个核苷酸、25个核苷酸、30个核苷酸、35个核苷酸、40个核苷酸、45个核苷酸、50个核苷酸、55个核苷酸、60个核苷酸、70个核苷酸、80个核苷酸、90个核苷酸、100个核苷酸、200个核苷酸,或上述值中的任何两个值之间的范围。

[0290] 在一些实施方案中,独特标识符选自一组不同的独特标识符。一组不同的独特标识符可以包括以下,或包括约以下:20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、5000种或这些值中的任何两个值之间的数字或范围的不同独特标识符。一组不同的独特标识符可以包括至少以下,或包括至多以下:20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种或5000种不同独特标识符。在一些实施方案中,一组独特标识符被设计成与待分析样品的DNA或RNA序列具有最小的序列

同源性。在一些实施方案中，一组独特标识符的序列彼此，或与其互补体相差以下或相差约以下：1个核苷酸、2个核苷酸、3个核苷酸、4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、8个核苷酸、9个核苷酸、10个核苷酸，或这些值中的任何两个值之间的数字或范围。在一些实施方案中，一组独特标识符的序列彼此，或与其互补体相差至少以下，或相差至多以下：1个核苷酸、2个核苷酸、3个核苷酸、4个核苷酸、5个核苷酸、6个核苷酸、7个核苷酸、8个核苷酸、9个核苷酸或10个核苷酸。

[0291] 在一些实施方案中，独特标识符可以包括用于引物（诸如通用引物）的结合位点。在一些实施方案中，独特标识符可以包括用于引物（诸如通用引物）的至少两个结合位点。在一些实施方案中，独特标识符可以包括用于引物（诸如通用引物）的至少三个结合位点。引物可以用于扩增独特标识符，例如，通过PCR扩增。在一些实施方案中，引物可以用于巢式PCR反应。

[0292] 任何合适的细胞组分结合试剂在本公开内容中被考虑，诸如任何蛋白质结合试剂（例如抗体或其片段、适体、小分子、配体、肽、寡核苷酸等，或它们的任何组合）。在一些实施方案中，细胞组分结合试剂可以是多克隆抗体、单克隆抗体、重组抗体、单链抗体（scAb），或它们的片段，诸如Fab、Fv等。在一些实施方案中，多于一种蛋白质结合试剂可以包括以下，或包括约以下：20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、5000种或这些值中的任何两个值之间的数字或范围的不同蛋白质结合试剂。在一些实施方案中，多于一种蛋白质结合试剂可以包括至少以下，或包括至多以下：20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种或5000种不同蛋白质结合试剂。

[0293] 在一些实施方案中，寡核苷酸通过接头与细胞组分结合试剂缀合。在一些实施方案中，寡核苷酸可以与蛋白质结合试剂共价地缀合。在一些实施方案中，寡核苷酸可以与蛋白质结合试剂非共价地缀合。在一些实施方案中，接头可以包含将寡核苷酸可逆或不可逆地附接到蛋白质结合试剂的化学基团。化学基团可以例如通过胺基团与接头缀合。在一些实施方案中，接头可以包含与另一个缀合到蛋白质结合试剂的化学基团形成稳定键的化学基团。例如，化学基团可以是UV光可裂解基团、二硫键、链霉亲和素、生物素、胺等。在一些实施方案中，化学基团可以通过氨基酸诸如赖氨酸上的伯胺或N-末端与蛋白质结合试剂缀合。寡核苷酸可以与蛋白质结合试剂的任何合适的位点缀合，只要它不干扰蛋白质结合试剂与其蛋白质靶之间的特异性结合。在蛋白质结合试剂是抗体的实施方案中，寡核苷酸可以在除抗原结合位点以外的任何位置与抗体缀合，例如在Fc区、C_H1结构域、C_H2结构域、C_H3结构域、C_L结构域等与抗体缀合。在一些实施方案中，每种蛋白质结合试剂可以与单种寡核苷酸分子缀合。在一些实施方案中，每种蛋白质结合试剂可以与以下，或者与约以下的寡核苷酸分子缀合：2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种或这些值中的任何两个值之间的数字或范围，其中每种寡核苷酸分子包含相同的独特标识符。在一些实施方案中，每种蛋白质结合试剂可以与多于一种寡核苷酸分子，例如，至少或至多2种、3种、4种、5种、10种、20种、30种、40种、50种、100种或1000种寡核苷酸分子缀合，其中每种寡核苷酸分子包含相同的独特标识符。

[0294] 在一些实施方案中，多于一种细胞组分结合试剂（例如，蛋白质结合试剂）能够与

样品中的多于一种细胞组分靶(例如,蛋白质靶)特异性地结合。样品可以是,或包括,单细胞、多于一个细胞、组织样品、肿瘤样品、血液样品等。在一些实施方案中,多于一种细胞组分靶包括细胞表面蛋白、细胞标志物、B细胞受体、T细胞受体、抗体、主要组织相容性复合体、肿瘤抗原、受体,或它们的任何组合。在一些实施方案中,多于一种细胞组分靶可以包括细胞内蛋白质。在一些实施方案中,多于一种细胞组分靶可以包括细胞外蛋白质。在一些实施方案中,多于一种细胞组分靶可以是生物体中所有细胞组分靶(例如,表达或可能表达的蛋白质)的1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或这些值中的任何两个值之间的数字或范围,或是生物体中所有细胞组分靶(例如,表达或可能表达的蛋白质)的约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,多于一种细胞组分靶可以是生物体中所有细胞组分靶(例如,表达或可能表达的蛋白质)的至少或至多1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%或99%。在一些实施方案中,多于一种细胞组分靶可以包括以下,或包括约以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种、10000种或这些值中的任何两个值之间的数字或范围的不同的细胞组分靶。在一些实施方案中,多于一种细胞组分靶可以包括至少以下,或包括至多以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种或10000种不同的细胞组分靶。

[0295] 使用寡核苷酸缀合的细胞组分结合试剂的样品索引化

[0296] 本文的公开内容包括用于样品鉴定的方法。在一些实施方案中,方法包括:使来自多于一个样品中的每一个的一个或更多个细胞与多于一种样品索引化组合物中的样品索引化组合物接触,其中一个或更多个细胞中的每一个包含一种或更多种细胞组分靶,其中多于一种样品索引化组合物中的每一种包含与样品索引化寡核苷酸关联的细胞组分结合试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分靶中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中多于一种样品索引化组合物中的至少两种样品索引化组合物的样品索引化序列包含不同的序列;去除多于一种样品索引化组合物中未结合的样品索引化组合物;使用多于一种条形码(例如随机条形码)对样品索引化寡核苷酸进行条形码化(例如随机条形码化)以产生多于一种条形码化样品索引化寡核苷酸;获得多于一种条形码化样品索引化寡核苷酸的测序数据;以及基于多于一种条形码化样品索引化寡核苷酸中的至少一种条形码化样品索引化寡核苷酸的样品索引化序列,鉴定一个或更多个细胞中的至少一个细胞的样品来源。

[0297] 在一些实施方案中,使用多于一种条形码对样品索引化寡核苷酸进行条形码化包括:使多于一种条形码与样品索引化寡核苷酸接触以产生与样品索引化寡核苷酸杂交的条形码;以及使与样品索引化寡核苷酸杂交的条形码延伸以产生多于一种条形码化样品索引化寡核苷酸。使条形码延伸可以包括使用DNA聚合酶使条形码延伸以产生多于一种条形码化样品索引化寡核苷酸。使条形码延伸可以包括使用逆转录酶使条形码延伸以产生多于一种条形码化样品索引化寡核苷酸。

[0298] 与抗体缀合的寡核苷酸、用于与抗体缀合的寡核苷酸,或先前与抗体缀合的寡核苷酸,在本文中被称为抗体寡核苷酸("AbOligo")。在样品索引化的上下文中,抗体寡核苷酸

酸在本文中被称为样品索引化寡核苷酸。与抗体寡核苷酸缀合的抗体在本文中被称为热抗体或寡核苷酸抗体。不与抗体寡核苷酸缀合的抗体在本文中被称为冷抗体或无寡核苷酸抗体。与结合试剂(例如,蛋白质结合试剂)缀合的寡核苷酸、用于与结合试剂缀合的寡核苷酸,或先前与结合试剂缀合的寡核苷酸,在本文中被称为试剂寡核苷酸。在样品索引化的上下文中,试剂寡核苷酸在本文中被称为样品索引化寡核苷酸。与抗体寡核苷酸缀合的结合试剂在本文中被称为热结合试剂或寡核苷酸结合试剂。不与抗体寡核苷酸缀合的结合试剂在本文中称为冷结合试剂或无寡核苷酸结合试剂。

[0299] 图7示出了使用寡核苷酸缀合的细胞组分结合试剂进行样品索引化的示例性工作流程的示意性图示。在一些实施方案中,提供了各自包含结合试剂的多于一种组合物705a、705b等。结合试剂可以是蛋白质结合试剂,诸如抗体。细胞组分结合试剂可以包括抗体、四聚体、适体、蛋白质支架,或它们的组合。多于一种组合物705a、705b的结合试剂可以与相同的细胞组分靶结合。例如,多于一种组合物705a、705b的结合试剂可以是相同的(除了与结合试剂关联的样品索引化寡核苷酸之外)。

[0300] 不同的组合物可以包括与具有不同样品索引化序列的样品索引化寡核苷酸缀合的结合试剂。在不同实施方式中,不同的组合物的数量可以是不同的。在一些实施方案中,不同的组合物的数量可以是以下,或是约以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000或这些值中的任何两个值之间的数字或范围。在一些实施方案中,不同的组合物的数量可以是至少以下,或至多以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000或10000。

[0301] 在一些实施方案中,一种组合物中结合试剂的样品索引化寡核苷酸可以包括相同的样品索引化序列。一种组合物中结合试剂的样品索引化寡核苷酸可以不相同。在一些实施方案中,一种组合物中具有相同样品索引化序列的结合试剂的样品索引化寡核苷酸的百分比可以是以下,或是约以下:50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.9%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,一种组合物中具有相同样品索引化序列的结合试剂的样品索引化寡核苷酸的百分比可以是至少以下,或至多以下:50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.9%。

[0302] 组合物705a和705b可以用于标记不同样品中的样品。例如,组合物705a中细胞组分结合试剂的样品索引化寡核苷酸可以具有一种样品索引化序列,并且可以用于对样品707a(诸如患者的样品)中的细胞710a(显示为黑色的圈)进行标记。组合物705b中细胞组分结合试剂的样品索引化寡核苷酸可以具有另一种样品索引化序列,并且可以用于对样品707b(诸如另一患者的样品或同一患者的另一样品)中的细胞710b(显示为带阴影的圈)进

行标记。细胞组分结合试剂可以特异性地结合细胞表面上的细胞组分靶或蛋白质,诸如细胞标志物、B细胞受体、T细胞受体、抗体、主要组织相容性复合体、肿瘤抗原、受体,或它们的任何组合。未结合的细胞组分结合试剂可以例如通过用缓冲液洗涤细胞来去除。

[0303] 然后,具有细胞组分结合试剂的细胞可以被分离到多于一个隔室诸如微孔阵列中,其中单个隔室715a、715b的尺寸适合于单细胞710a和单个珠720a或者单细胞710b和单个珠720b。每个珠720a、720b可以包含多于一种寡核苷酸探针和分子标记序列,所述多于一种寡核苷酸探针可以包含珠上所有寡核苷酸探针共有的细胞标记。在一些实施方案中,每种寡核苷酸探针可以包含靶结合区,例如多(dT)序列。与组合物705a的细胞组分结合试剂缀合的样品索引化寡核苷酸725a可以被配置为可从细胞组分结合试剂脱离或不可脱离。与组合物705a的细胞组分结合试剂缀合的样品索引化寡核苷酸725a可以使用化学、光学或其他手段从细胞组分结合试剂脱离。与组合物705b的细胞组分结合试剂缀合的样品索引化寡核苷酸725b可以被配置为可从细胞组分结合试剂脱离或不可脱离。与组合物705b的细胞组分结合试剂缀合的样品索引化寡核苷酸725b可以使用化学、光学或其他手段从细胞组分结合试剂脱离。

[0304] 细胞710a可以被裂解以释放细胞710a内的核酸,诸如基因组DNA或细胞mRNA 730a。裂解的细胞735a显示为加虚线的圈。细胞mRNA 730a、样品索引化寡核苷酸725a或二者可以被珠720a上的寡核苷酸探针捕获,例如,通过与多(dT)序列杂交来捕获。逆转录酶可以用于使用细胞mRNA 730a和寡核苷酸725a作为模板使与细胞mRNA 730a和寡核苷酸725a杂交的寡核苷酸探针延伸。可以对逆转录酶产生的延伸产物进行扩增和测序。

[0305] 类似地,细胞710b可以被裂解以释放细胞710b内的核酸,诸如基因组DNA或细胞mRNA 730b。裂解的细胞735b显示为加虚线的圈。细胞mRNA 730b、样品索引化寡核苷酸725b或二者可以被珠720b上的寡核苷酸探针捕获,例如,通过与多(dT)序列杂交来捕获。逆转录酶可以用于使用细胞mRNA 730b和寡核苷酸725b作为模板使与细胞mRNA 730b和寡核苷酸725b杂交的寡核苷酸探针延伸。可以对逆转录酶产生的延伸产物进行扩增和测序。

[0306] 可以对测序读段进行细胞标记、分子标记、基因身份和样品身份的去多重化(例如,根据样品索引化寡核苷酸725a和725b的样品索引化序列)。细胞标记、分子标记和基因身份的去多重化可以产生样品中每个单细胞的基因表达的数字表示。使用样品索引化寡核苷酸的样品索引化序列进行的细胞标记、分子标记和样品身份的去多重化,可以用于确定样品来源。

[0307] 在一些实施方案中,针对细胞表面上的细胞组分的细胞组分结合试剂可以与独特的样品索引化寡核苷酸的文库缀合,以允许细胞保持样品身份。例如,针对细胞表面标志物的抗体可以与独特的样品索引化寡核苷酸的文库缀合,以允许细胞保持样品身份。这将实现将多于一个样品加载到同一个Rhapsody™盒(cartridge)中,因为在整个文库制备和测序过程中,与样品来源相关的信息被保留。样品索引化可以允许在单个实验中一起运行多于一个样品,从而简化和缩短实验时间,并消除批次效应(batch effect)。

[0308] 本文的公开内容包括用于样品鉴定的方法。在一些实施方案中,方法包括:使来自多于一个样品中的每一个的一个或更多个细胞与多于一种样品索引化组合物中的样品索引化组合物接触,其中一个或更多个细胞中的每一个包含一种或更多种细胞组分靶,其中多于一种样品索引化组合物中的每一种包含与样品索引化寡核苷酸关联的细胞组分结合

试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分靶中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中多于一种样品索引化组合物中的至少两种样品索引化组合物的样品索引化序列包含不同的序列;去除多于一种样品索引化组合物中未结合的样品索引化组合物。方法可以包括使用多于一种条形码(例如,随机条形码)对样品索引化寡核苷酸进行条形码化(例如,随机条形码化)以产生多于一种条形码化样品索引化寡核苷酸;获得多于一种条形码化样品索引化寡核苷酸的测序数据;以及基于多于一种条形码化样品索引化寡核苷酸中的至少一种条形码化样品索引化寡核苷酸的样品索引化序列,鉴定一个或更多个细胞中至少一个细胞的样品来源。

[0309] 在一些实施方案中,用于样品鉴定的方法包括:使来自多于一个样品中的每一个的一个或更多个细胞与多于一种样品索引化组合物中的样品索引化组合物接触,其中一个或更多个细胞中的每一个包含一种或更多种细胞组分靶,其中多于一种样品索引化组合物中的每一种包含与样品索引化寡核苷酸关联的细胞组分结合试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分靶中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中多于一种样品索引化组合物中的至少两种样品索引化组合物的样品索引化序列包含不同的序列;去除多于一种样品索引化组合物中未结合的样品索引化组合物;以及基于多于一种样品索引化组合物中至少一种样品索引化寡核苷酸的样品索引化序列,鉴定一个或更多个细胞中至少一个细胞的样品来源。

[0310] 在一些实施方案中,鉴定至少一个细胞的样品来源包括:使用多于一种条形码(例如随机条形码)对多于一种样品索引化组合物的样品索引化寡核苷酸进行条形码化(例如随机条形码化)以产生多于一种条形码化样品索引化寡核苷酸;获得多于一种条形码化样品索引化寡核苷酸的测序数据;以及基于多于一种条形码化样品索引化寡核苷酸中的至少一种条形码化样品索引化寡核苷酸的样品索引化序列来鉴定细胞的样品来源。在一些实施方案中,使用多于一种条形码对样品索引化寡核苷酸进行条形码化以产生多于一种条形码化样品索引化寡核苷酸包括使用多于一种随机条形码对样品索引化寡核苷酸进行随机条形码化以产生多于一种随机条形码化样品索引化寡核苷酸。

[0311] 在一些实施方案中,鉴定至少一个细胞的样品来源可以包括鉴定多于一种样品索引化组合物的至少一种样品索引化寡核苷酸的样品索引化序列的存在或不存在。鉴定样品索引化序列的存在或不存在可以包括:复制至少一种样品索引化寡核苷酸以产生多于一种复制的样品索引化寡核苷酸;获得多于一种复制的样品索引化寡核苷酸的测序数据;以及基于多于一种样品索引化寡核苷酸中的复制的样品索引化寡核苷酸的样品索引化序列来鉴定细胞的样品来源,所述多于一种样品索引化寡核苷酸对应于测序数据中的至少一种条形码化样品索引化寡核苷酸。

[0312] 在一些实施方案中,复制至少一种样品索引化寡核苷酸以产生多于一种复制的样品索引化寡核苷酸包括:在复制至少一种条形码化样品索引化寡核苷酸之前,将复制衔接子连接到至少一种条形码化样品索引化寡核苷酸。复制至少一种条形码化样品索引化寡核苷酸可以包括使用连接到至少一种条形码化样品索引化寡核苷酸的复制衔接子复制至少一种条形码化样品索引化寡核苷酸以产生多于一种复制的样品索引化寡核苷酸。

[0313] 在一些实施方案中,复制至少一种样品索引化寡核苷酸以产生多于一种复制的样品索引化寡核苷酸包括:在复制至少一种条形码化样品索引化寡核苷酸之前,使捕获探针

与至少一种样品索引化寡核苷酸接触以产生与样品索引化寡核苷酸杂交的捕获探针;以及使与样品索引化寡核苷酸杂交的捕获探针延伸以产生与捕获探针关联的样品索引化寡核苷酸。复制至少一种样品索引化寡核苷酸可以包括复制与捕获探针关联的样品索引化寡核苷酸以产生多于一种复制的样品索引化寡核苷酸。

[0314] 细胞过载 (overloading) 和多重体 (multiplet) 鉴定

[0315] 本文的公开内容还包括用于鉴定细胞过载和多重体的方法、试剂盒和系统。这样的方法、试剂盒和系统可以用于本文公开的任何合适的方法、试剂盒和系统,或与本文公开的任何合适的方法、试剂盒和系统联合使用,例如用于以下或与以下联合使用:使用与寡核苷酸关联的细胞组分结合试剂来测量细胞组分表达水平(诸如蛋白质表达水平)的方法、试剂盒和系统。

[0316] 使用当前的细胞加载技术,当约20000个细胞被加载到具有~60000个微孔的微孔盒或阵列中时,具有两个或更多个细胞(称为双重体或多重体)的微孔或液滴的数量可以是最小的。然而,当加载的细胞数量增加时,具有多个细胞的微孔或液滴的数量可能显著增加。例如,当将约50000个细胞加载到微孔盒或阵列的约60000个微孔中时,具有多个细胞的微孔的百分比可以相当高,诸如11%-14%。这样将大量细胞加载到微孔中可称为细胞过载。然而,如果细胞被分为一定数量的组(例如5个组)并且每组中的细胞用具有不同样品索引化序列的样品索引化寡核苷酸进行标记,与两种或更多种样品索引化序列关联的细胞标记(例如条形码诸如随机条形码的细胞标记)可以在测序数据中鉴定并从后续处理中去除。在一些实施方案中,细胞被分为大量的组(例如10000个组)并且每组中的细胞用具有不同样品索引化序列的样品索引化寡核苷酸进行标记,与两种或更多种样品索引化序列关联的样品标记可以在测序数据中鉴定并从后续处理中去除。在一些实施方案中,不同的细胞用具有不同细胞鉴定序列的细胞鉴定寡核苷酸进行标记,与两种或更多种细胞鉴定寡核苷酸关联的细胞鉴定序列可以在测序数据中鉴定并从后续处理中去除。相对于微孔盒或阵列中的微孔数量,可以将这样更高数量的细胞加载到微孔中。

[0317] 本文的公开内容包括用于样品鉴定的方法。在一些实施方案中,方法包括:使第一多于一个细胞和第二多于一个细胞分别与两种样品索引化组合物接触,其中第一多于一个细胞中的每一个和第二多于一个细胞中的每一个包含一种或更多种细胞组分,其中两种样品索引化组合物中的每一种包含与样品索引化寡核苷酸关联的细胞组分结合试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中两种样品索引化组合物的样品索引化序列包含不同的序列;使用多于一种条形码对样品索引化寡核苷酸进行条形码化以产生多于一种条形码化样品索引化寡核苷酸,其中多于一种条形码中的每一种包含细胞标记序列、条形码序列(例如,分子标记序列)和靶结合区,其中多于一种条形码中的至少两种条形码的条形码序列包括不同的序列,并且其中多于一种条形码中的至少两种条形码包含相同的细胞标记序列;获得多于一种条形码化样品索引化寡核苷酸的测序数据;以及鉴定各自与获得的测序数据中的两种或更多种样品索引化序列关联的一种或更多种细胞标记序列;以及从获得的测序数据去除与这样的一种或更多种细胞标记序列关联的测序数据:所述一种或更多种细胞标记序列各自与两种或更多种样品索引化序列关联,和/或将与这样的一种或更多种细胞标记序列关联的测序数据从随后的分析(例如,单细胞mRNA谱分析或全转录组分析)

排除：所述一种或更多种细胞标记序列各自与两种或更多种样品索引化序列关联。在一些实施方案中，样品索引化寡核苷酸包含条形码序列（例如，分子标记序列）、用于通用引物的结合位点，或它们的组合。

[0318] 例如，方法可以用于使用样品索引化将50000个或更多个细胞（相比于10000-20000个细胞）加载。样品索引化可以使用寡核苷酸缀合的细胞组分结合试剂（例如抗体）或针对细胞组分（例如通用蛋白质标志物）的细胞组分结合试剂，以用独特样品索引对来自不同样品的细胞进行标记。当来自不同样品的两个或更多个细胞、来自样品的不同细胞群体的两个或更多个细胞，或样品的两个或更多个细胞，被捕获在同一微孔或液滴中时，组合“细胞”（或两个或更多个细胞的内容物）可以与具有不同样品索引化序列的样品索引化寡核苷酸（或具有不同细胞鉴定序列的细胞鉴定寡核苷酸）关联。在不同实施方式中，细胞的不同群体的数量可以是不同的。在一些实施方案中，不同群体的数量可以是以下，或约以下：2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100或这些值中的任何两个值之间的数字或范围。在一些实施方案中，不同群体的数量可以是至少以下，或至多以下：2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100。在不同实施方式中，每个群体中的细胞的数量或平均数量可以是不同的。在一些实施方案中，每个群体中的细胞的数量或平均数量可以是以下，或约以下：1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100或这些值中的任何两个值之间的数字或范围。在一些实施方案中，每个群体中的细胞的数量或平均数量可以是至少以下，或至多以下：1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100。当每个群体中的细胞的数量或平均数量足够小（例如，等于或少于50个、25个、10个、5个、4个、3个、2个或1个细胞/群体）时，用于细胞过载和多重体鉴定的样品索引化组合物可以称为细胞鉴定组合物。

[0319] 通过将样品的细胞等分为多个群体，可以将样品的细胞分为多个群体。基于与测序数据中的一种细胞标记序列（例如，条形码诸如随机条形码的细胞标记序列）关联的两种或更多种样品索引化序列，可以将与测序数据中的多于一种样品索引化序列关联的“细胞”鉴定为“多重体”。组合“细胞”的测序数据在本文中也称为多重体。多重体可以是双重体、三重体、四重体、五重体、六重体、七重体、八重体、九重体或它们的任何组合。多重体可以是任何n重体。在一些实施方案中，n是以下，或是约以下：2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或这些值中的任何两个值之间的范围。在一些实施方案中，n是至少以下，或至多以下：2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20。

[0320] 当确定单细胞的表达谱时，两个细胞可以被识别为一个细胞，并且两个细胞的表达谱可以被识别为一个细胞的表达谱（称为双重体表达谱）。例如，当使用条形码化（例如，随机条形码化）确定两个细胞的表达谱时，两个细胞的mRNA分子可以与具有相同细胞标记的条形码关联。作为另一个实例，两个细胞可以与一个颗粒（例如珠）关联。颗粒可以包括具有相同细胞标记的条形码。在裂解细胞后，两个细胞中的mRNA分子可以与颗粒的条形码关联，从而与相同的细胞标记关联。双重体表达谱可能扭曲对表达谱的解释。

[0321] 双重体可以指与具有不同样品索引化序列的两种样品索引化寡核苷酸关联的组合“细胞”。双重体也可以指与具有两种样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。当同一微孔或液滴中捕获到与不同序列的两种样品索引化寡核苷酸关联的两个细胞（或与具有两种不同样品索引化序列的样品索引化寡核苷酸关联的两个或更多个细胞）

时,双重体可能出现,组合“细胞”可以与具有不同样品索引化序列的两种样品索引化寡核苷酸关联。三重体可以指与全部具有不同样品索引化序列的三种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有三种不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。四重体可以指与全部具有不同样品索引化序列的四种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有四种不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。五重体可以指与全部具有不同样品索引化序列的五种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有五个不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。六重体可以指与全部具有不同样品索引化序列的六种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有六种不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。七重体可以指与全部具有不同样品索引化序列的七种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有七种不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。八重体可以指与全部具有不同样品索引化序列的八种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有八种不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。九重体可以指与全部具有不同样品索引化序列的九种样品索引化寡核苷酸关联的组合“细胞”,或者是与具有九种不同样品索引化序列的样品索引化寡核苷酸关联的组合“细胞”。当同一微孔或微滴中捕获到与不同序列的两种或更多种样品索引化寡核苷酸关联的两个或更多个细胞(或与具有两种或更多种不同样品索引化序列的样品索引化寡核苷酸关联的两个或更多个细胞)时,多重体可能出现,组合的“细胞”可以与具有两种或更多种不同样品索引化序列的样品索引化寡核苷酸关联。

[0322] 作为另一个实例,方法可以用于多重体鉴定,无论是在样品过载的情况下,还是在将细胞加载到微孔阵列的微孔上或产生含有细胞的液滴的情况下。当两个或更多个细胞被加载到一个微孔中时,从组合“细胞”(或两个或更多个细胞的内容物)得到的数据是具有异常基因表达谱的多重体。通过使用样品索引化,人们可以通过寻找细胞标记来识别这些多重体中的一些,所述细胞标记各自与以下关联或被分配给以下:具有不同样品索引化序列的两种或更多种样品索引化寡核苷酸(或具有两种或更多种样品索引化序列的样品索引化寡核苷酸)。利用样品索引化序列,本文公开的方法可以用于多重体鉴定(无论是否在样品过载的情况下,还是在将细胞加载到微孔阵列的微孔上或产生含有细胞的液滴的情况下)。在一些实施方案中,方法包括:使第一多于一个细胞和第二多于一个细胞分别与两种样品索引化组合物接触,其中第一多于一个细胞中的每一个和第二多于一个细胞中的每一个包含一种或更多种细胞组分,其中两种样品索引化组合物中的每一种包含与样品索引化寡核苷酸关联的细胞组分结合试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分中的至少一种特异性地结合,其中样品索引化寡核苷酸包含样品索引化序列,并且其中两种样品索引化组合物的样品索引化序列包含不同的序列;使用多于一种条形码对样品索引化寡核苷酸进行条形码化以产生多于一种条形码化样品索引化寡核苷酸,其中多于一种条形码中的每一种包含细胞标记序列、条形码序列(例如,分子标记序列)和靶结合区,其中多于一种条形码中的至少两种条形码的条形码序列包括不同的序列,并且其中多于一种条形码中的至少两种条形码包含相同的细胞标记序列;获得多于一种条形码化样品索引化寡核苷酸的测序数据;以及鉴定各自与获得的测序数据中的两种或更多种样品索引化序列关联的一种或更多种多重体细胞标记序列。

[0323] 可以加载到微孔盒的微孔上或加载到使用微流体装置产生的液滴中的细胞的数量可能受到多重体比率的限制。加载更多的细胞可能导致更多的多重体,这可能难以鉴定并且在单细胞数据中产生噪声。利用样品索引化,方法可以用于更准确地标记或鉴定多重体,并将多重体从测序数据或后续分析去除。能够以更高的置信度鉴定多重体,可以提高用户对多重体比率的容忍度,并将更多的细胞加载到每个微孔盒上,或者生成各自具有至少一个细胞的液滴。

[0324] 在一些实施方案中,使第一多于一个细胞和第二多于一个细胞分别与两种样品索引化组合物接触包括:使第一多于一个细胞与两种样品索引化组合物中的第一样品索引化组合物接触;以及使第一多于一个细胞与两种样品索引化组合物中的第二样品索引化组合物接触。在不同的实施方式中,多于一个细胞的数量和多于一种样品索引化组合物的数量可以不同。在一些实施方案中,多于一个细胞和/或样品索引化组合物的数量可以是以下,或是约以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000、1000000或这些值中的任何两个值之间的数字或范围。在一些实施方案中,多于一个细胞和/或样品索引化组合物的数量可以是至少以下,或至多以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000或1000000。在不同实施方式中,细胞的数量可以是不同的。在一些实施方案中,细胞的数量或平均数量可以是以下,或是约以下:1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000、1000000或这些值中的任何两个值之间的数字或范围。在一些实施方案中,细胞的数量或平均数量可以是至少以下,或至多以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000或1000000。

[0325] 在一些实施方案中,方法包括:去除两种样品索引化组合物中未结合的样品索引化组合物。去除未结合的样品索引化组合物可以包括用洗涤缓冲液洗涤第一多于一个细胞和第二多于一个细胞中的细胞。去除未结合的样品索引化组合物可以包括使用流式细胞术选择与两种样品索引化组合物的至少一种细胞组分结合试剂结合的细胞。在一些实施方案中,方法包括:将来自多于一个样品中的每一个的一个或更多个细胞裂解。

[0326] 在一些实施方案中,样品索引化寡核苷酸被配置为可从细胞组分结合试剂脱离或不可脱离。方法可以包括使样品索引化寡核苷酸从细胞组分结合试剂脱离。使样品索引化寡核苷酸脱离可以包括通过UV光裂解、化学处理(例如,使用还原剂,诸如二硫苏糖醇)、加热、酶处理或它们的任何组合使样品索引化寡核苷酸从细胞组分结合试剂脱离。

[0327] 在一些实施方案中,使用多于一种条形码对样品索引化寡核苷酸进行条形码化包括:使多于一种条形码与样品索引化寡核苷酸接触以产生与样品索引化寡核苷酸杂交的条形码;以及使与样品索引化寡核苷酸杂交的条形码延伸以产生多于一种条形码化样品索引化寡核苷酸。使条形码延伸可以包括使用DNA聚合酶使条形码延伸以产生多于一种条形码化样品索引化寡核苷酸。使条形码延伸可以包括使用逆转录酶使条形码延伸以产生多于一种条形码化样品索引化寡核苷酸。

[0328] 在一些实施方案中,方法包括:扩增多于一种条形码化样品索引化寡核苷酸以产生多于一种扩增子。扩增多于一种条形码化样品索引化寡核苷酸可以包括使用聚合酶链式反应(PCR)扩增条形码序列的至少一部分(例如,分子标记序列)和样品索引化寡核苷酸的

至少一部分。在一些实施方案中,获得多于一种条形码化样品索引化寡核苷酸的测序数据可以包括获得多于一种扩增子的测序数据。获得测序数据包括对条形码序列的至少一部分和样品索引化寡核苷酸的至少一部分进行测序。在一些实施方案中,鉴定至少一个细胞的样品来源包括基于至少一个条形码化样品索引化寡核苷酸的样品索引化序列鉴定多于一种条形码化靶的样品来源。

[0329] 在一些实施方案中,使用多于一种条形码对样品索引化寡核苷酸进行条形码化以产生多于一种条形码化样品索引化寡核苷酸包括使用多于一种随机条形码对样品索引化寡核苷酸进行随机条形码化以产生多于一种随机条形码化样品索引化寡核苷酸。

[0330] 在一些实施方案中,方法包括:使用多于一种条形码对细胞的多于一种靶进行条形码化以产生多于一种条形码化靶,其中多于一种条形码中的每一种包含细胞标记序列,并且其中多于一种条形码中的至少两种条形码包含相同的细胞标记序列;以及获取条形码化靶的测序数据。使用多于一种条形码将多于一种靶条形码化以产生多于一种条形码化靶可以包括:使靶的拷贝与条形码的靶结合区接触;以及使用多于一种条形码将多于一种靶逆转录以产生多于一种逆转录的靶。

[0331] 在一些实施方案中,方法包括:在获得多于一种条形码化靶的测序数据之前,扩增条形码化靶以产生多于一种经扩增的条形码化靶。扩增条形码化靶以产生多于一种经扩增的条形码化靶可以包括:通过聚合酶链式反应(PCR)扩增条形码化靶。使用多于一种条形码对细胞的多于一种靶进行条形码化以产生多于一种条形码化靶可以包括使用多于一种随机条形码对细胞的多于一种靶进行随机条形码化以产生多于一种随机条形码化靶。

[0332] 在一些实施方案中,用于细胞鉴定的方法包括:使第一多于一个一个或更多个细胞(a first plurality of one or more cells)和第二多于一个一个或更多个细胞(a second plurality of one or more cells)分别与两种细胞鉴定组合物接触,其中第一多于一个一个或更多个细胞中的每一个和第二多于一个一个或更多个细胞中的每一个包含一种或更多种细胞组分,其中两种细胞鉴定组合物中的每一种包含与细胞鉴定寡核苷酸关联的细胞组分结合试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分中的至少一种特异性地结合,其中细胞鉴定寡核苷酸包含细胞鉴定序列,并且其中两种细胞鉴定组合物的细胞鉴定序列包含不同的序列;使用多于一种条形码对细胞鉴定寡核苷酸进行条形码化以产生多于一种条形码化细胞鉴定寡核苷酸,其中多于一种条形码中的每一种包含细胞标记序列、条形码序列(例如,分子标记序列)和靶结合区,其中多于一种条形码中的至少两种条形码的条形码序列包括不同的序列,并且其中多于一种条形码中的至少两种条形码包含相同的细胞标记序列;获得多于一个条形码化的细胞鉴定寡核苷酸的测序数据;以及鉴定各自与获得的测序数据中的两种或更多种细胞鉴定序列关联的一种或更多种细胞标记序列;以及从获得的测序数据去除与这样的一种或更多种细胞标记序列关联的测序数据:所述一种或更多种细胞标记序列各自与两种或更多种细胞鉴定序列关联,和/或将与这样的一种或更多种细胞标记序列关联的测序数据从随后的分析(例如,单细胞mRNA谱分析或全转录组分析)排除:所述一种或更多种细胞标记序列各自与两种或更多种细胞鉴定序列关联。在一些实施方案中,细胞鉴定寡核苷酸包含条形码序列(例如,分子标记序列)、用于通用引物的结合位点,或它们的组合。

[0333] 当同一微孔或微滴中捕获到与不同序列的两种或更多种细胞鉴定寡核苷酸关联

的两个或更多个细胞(或与具有两种或更多种不同细胞鉴定序列的细胞鉴定寡核苷酸关联的两个或更多个细胞)时,多重体(例如,双重体、三重体等)可能出现,组合“细胞”可以与具有两种或更多种不同细胞鉴定序列的细胞鉴定寡核苷酸关联。

[0334] 细胞鉴定组合物可以用于多重体鉴定,无论是在细胞过载的情况下,还是在将细胞加载到微孔阵列的微孔上或产生含有细胞的液滴的情况下。当两个或更多个细胞被加载到一个微孔中时,从组合“细胞”(或两个或更多个细胞的内容物)得到的数据是具有异常基因表达谱的多重体。通过使用细胞鉴定,人们可以通过寻找细胞标记(例如,条形码诸如随机条形码的细胞标记)来识别这些多重体中的一些,所述细胞标记各自与以下关联或被分配给以下:具有不同细胞鉴定序列的两种或更多种细胞鉴定寡核苷酸(或具有两种或更多种细胞鉴定序列的细胞鉴定寡核苷酸)。利用细胞鉴定序列,本文公开的方法可以用于多重体鉴定(无论是否在样品过载的情况下,还是在将细胞加载到微孔阵列的微孔上或产生含有细胞的液滴的情况下)。在一些实施方案中,方法包括:使第一多于一个一个或更多个细胞和第二多于一个一个或更多个细胞分别与两种细胞鉴定组合物接触,其中第一多于一个一个或更多个细胞中的每一个和第二多于一个一个或更多个细胞中的每一个包含一种或更多种细胞组分,其中两种细胞鉴定组合物中的每一种包含与细胞鉴定寡核苷酸关联的细胞组分结合试剂,其中细胞组分结合试剂能够与一种或更多种细胞组分中的至少一种特异性地结合,其中细胞鉴定寡核苷酸包含细胞鉴定序列,并且其中两种细胞鉴定组合物的细胞鉴定序列包含不同的序列;使用多于一种条形码对细胞鉴定寡核苷酸进行条形码化以产生多于一种条形码化细胞鉴定寡核苷酸,其中多于一种条形码中的每一种包含细胞标记序列、条形码序列(例如,分子标记序列)和靶结合区,其中多于一种条形码中的至少两种条形码的条形码序列包括不同的序列,并且其中多于一种条形码中的至少两种条形码包含相同的细胞标记序列;获得多于一种条形码化细胞鉴定寡核苷酸的测序数据;以及鉴定各自与获得的测序数据中的两种或更多种细胞鉴定序列关联的一种或更多种多重体细胞标记序列。

[0335] 可以加载到微孔盒的微孔上或加载到使用微流体装置产生的液滴中的细胞的数量可能受到多重体比率的限制。加载更多的细胞可能导致更多的多重体,这可能难以鉴定并在单细胞数据中产生噪声。利用细胞鉴定,方法可以用于更准确地标记或鉴定多重体,并将多重体从测序数据或后续分析去除。能够以更高的置信度鉴定多重体,可以提高用户对多重体比率的容忍度,并将更多的细胞加载到每个微孔盒上,或者产生各自具有至少一个细胞的液滴。

[0336] 在一些实施方案中,使第一多于一个一个或更多个细胞和第二多于一个一个或更多个细胞分别与两种细胞鉴定组合物接触包括:使第一多于一个一个或更多个细胞与两种细胞鉴定组合物中的第一细胞鉴定组合物接触;以及使第一多于一个一个或更多个细胞与两种细胞鉴定组合物中的第二细胞鉴定组合物接触。在不同的实施方式中,多于一种细胞鉴定组合物的数量可以是不同的。在一些实施方案中,细胞鉴定组合物的数量可以是以下,或是约以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000、1000000或这些值中的任何两个值之间的数字或范围。在一些实施方案中,细胞鉴定组合物的数量可以是至少以下,或至多以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、

10000、100000或1000000。在不同的实施方式中，每一个多于一个一个或更多个细胞 (each plurality of one or more cells) 中的细胞的数量或平均数量可以不同。在一些实施方案中，每一个多于一个一个或更多个细胞中的细胞的数量或平均数量可以是以下，或是约以下：1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000、1000000或这些值中的任何两个值之间的数字或范围。在一些实施方案中，每一个多于一个一个或更多个细胞中的细胞的数量或平均数量可以是至少以下，或至多以下：2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、10000、100000或1000000。

[0337] 在一些实施方案中，方法包括：去除两种细胞鉴定组合物中未结合的细胞鉴定组合物。去除未结合的细胞鉴定组合物可以包括用洗涤缓冲液洗涤第一多于一个一个或更多个细胞和第二多于一个一个或更多个细胞中的细胞。去除未结合的细胞鉴定组合物可以包括使用流式细胞术选择与两种细胞鉴定组合物的至少一种细胞组分结合试剂结合的细胞。在一些实施方案中，方法包括：将来自多于一个样品中的每一个的一个或更多个细胞裂解。

[0338] 在一些实施方案中，细胞鉴定寡核苷酸被配置为可从细胞组分结合试剂脱离或不可脱离。方法可以包括使细胞鉴定寡核苷酸从细胞组分结合试剂脱离。使细胞鉴定寡核苷酸脱离可以包括通过UV光裂解、化学处理 (例如，使用还原剂，诸如二硫苏糖醇)、加热、酶处理或它们的任何组合使细胞鉴定寡核苷酸从细胞组分结合试剂脱离。

[0339] 在一些实施方案中，使用多于一种条形码对细胞鉴定寡核苷酸进行条形码化包括：使多于一种条形码与细胞鉴定寡核苷酸接触以产生与细胞鉴定寡核苷酸杂交的条形码；以及使与细胞鉴定寡核苷酸杂交的条形码延伸以产生多于一种条形码化细胞鉴定寡核苷酸。使条形码延伸可以包括使用DNA聚合酶使条形码延伸以产生多于一种条形码化细胞鉴定寡核苷酸。使条形码延伸可以包括使用逆转录酶使条形码延伸以产生多于一种条形码化细胞鉴定寡核苷酸。

[0340] 在一些实施方案中，方法包括：扩增多于一种条形码化细胞鉴定寡核苷酸以产生多于一种扩增子。扩增多于一种条形码化细胞鉴定寡核苷酸可以包括使用聚合酶链式反应 (PCR) 扩增条形码序列的至少一部分 (例如，分子标记序列) 和细胞鉴定寡核苷酸的至少一部分。在一些实施方案中，获得多于一种条形码化细胞鉴定寡核苷酸的测序数据可以包括获得多于一种扩增子的测序数据。获得测序数据包括对条形码序列的至少一部分和细胞鉴定寡核苷酸的至少一部分进行测序。在一些实施方案中，鉴定至少一个细胞的样品来源包括基于至少一种条形码化细胞鉴定寡核苷酸的细胞鉴定序列鉴定多于一种条形码化靶的样品来源。

[0341] 在一些实施方案中，使用多于一种条形码对细胞鉴定寡核苷酸进行条形码化以产生多于一种条形码化细胞鉴定寡核苷酸包括使用多于一种随机条形码对细胞鉴定寡核苷酸进行随机条形码化以产生多于一种随机条形码化细胞鉴定寡核苷酸。

[0342] 具有随机引发和延伸的全转录组分析 (WTA)

[0343] 本文的公开内容包括用于从cDNA产物 (例如，BD Rhapsody单细胞分析系统的cDNA产物) 产生全转录组分析 (WTA) 文库用于在相容的测序仪 (例如 Illumina®测序仪) 上测序的方法。目前，标准的BD Rhapsody RNA工作流程首先经由BD Rhapsody细胞捕获珠 (下文称为‘珠’) 上的3’捕获和cDNA合成来捕获RNA转录组，然后进行2次多重化巢式PCR扩增，以富

集感兴趣的RNA靶(靶向测定)。在方法的一些实施方案中,靶向RNA分析通常对低表达靶产生更高的灵敏度,而WTA RNA分析提供更大广度的被询问基因。在一些实施方案中,期望的靶的成功扩增可能需要彻底了解感兴趣的模型系统中的每种RNA的3'末端及其多腺苷酸化位点的使用。WTA方法,例如与BD Rhapsody一起使用,允许获得若干水平的信息,包括1)在其模型系统中识别感兴趣的RNA靶,以鉴定待设计的一组基因;2)将其模型系统中转录组的3'末端表征为用于BD Rhapsody的新一代PCR组套设计(panel design)的输入,以及3)与标准BD Rhapsody靶向方法相比,允许用户通过测定更大广度的基因来进行发现生物学。

[0344] 图8是进行单细胞的全转录组分析(WTA)的非限制性示例性工作流程800的示意性图示。对于框804处的单细胞捕获,可以使用微孔(诸如Rhapsody或ICELL8)或液滴(诸如dropseq、ddseq)。框808处的cDNA合成可以包括逆转录和第二链合成,随后在框812处通过末端修复和衔接子连接以无偏倚的方式添加5'衔接子。在一些实施方案中,框808和框812可以包括进行逆转录与使用模板转换来添加衔接子。在框814处,全转录组的扩增可以包括聚合酶链式反应(PCR)。在框816处,用于测序(例如,Illumina测序)的短插入物的产生可以包括随机引发和延伸,随后是PCR。在一些实施方案中,框816可以包括片段化、末端修复和连接,随后是PCR。

[0345] 图9是进行单细胞的基于连接的全转录组分析的非限制性示例性工作流程的示意性图示。方法可以包括使用条形码(例如,各自具有通用序列(Univ)、独特分子索引(UMI)、细胞标记(CL)和寡聚(dT)序列)的条形码)进行逆转录、第二链合成、末端制备、全长cDNA扩增、随机引发,以及随后的测序文库扩增。

[0346] 如本文描述的,过去公开的WTA方法的主要挑战是实现含有WTA PCR扩增的珠,因为该PCR产量低。例如,在一些实施方案中,本文公开的新的基于RPE的WTA方法用替换随机引发和DNA聚合酶延伸可以免去在珠上进行WTA PCR的需要。

[0347] 图10和图11A是使用随机引发和引物延伸(RPE)有效地(例如,在比其他全转录组分析方法更短的时间内)进行单细胞的全转录组分析的非限制性示例性工作流程的示意性图示。方法可以包括使用条形码(例如,各自具有通用序列("Univ",例如,读段1(R1))、独特分子索引(UMI)、细胞标记(CL)和寡聚(dT)序列的条形码)、第一链cDNA合成、随机引发(例如,使用包含Univ(例如,读段2(R2)的随机多聚体))和随机引物延伸来进行逆转录。随机引物可以与沿所有转录物的编码序列的不同位置结合,并延伸以产生延伸产物(例如线性扩增产物)。延伸产物可以包括根据随机引物的结合位点而具有不同长度的cDNA。延伸产物可以用测序文库扩增引物进行扩增。工作流程可以包括单轮PCR扩增(如图10中描绘的)或两轮或更多轮PCR扩增(诸如图11A中示出的PCR1随后为PCR2("索引PCR"))。测序文库扩增可以包括使用文库正向引物和文库反向引物,经由突出端添加测序衔接子和/或文库索引。文库扩增可以经由文库正向引物和文库反向引物中的突出端添加测序衔接子(例如P5和P7序列)和样品索引(例如i5、i7)。文库扩增子可以被测序并经受本公开内容的下游方法。用以产生150bp x 2测序读段的成对末端测序可以揭示读段1上的细胞标记、独特分子索引、多(A)尾和/或基因(或基因的部分序列),读段2上的基因(或基因的部分序列)和/或多(A)尾,以及索引1读段上的样品索引。使用随机条形码进行全转录组扩增的方法、组合物、系统、装置和试剂盒先前已经在例如美国专利公布第2016/0312276号中公开,该美国专利公布的内容通过引用以其整体在此明确并入本文。

[0348] 图11B是使用随机引发和引物延伸进行全转录组分析的非限制性示例性方法。对于框1104处的细胞捕获、逆转录和ExoI处理以去除未使用的引物,该过程的参数包括珠是否可以再使用或者珠是否与其他分析方法相容。这样的分析方法包括与寡核苷酸(“Ab0”)关联(例如,缀合)的抗体的样品索引化、追踪或多重化、与寡核苷酸关联的抗体的单细胞蛋白质表达谱分析(在本文也称为“AbSeq”)、靶向单细胞分析方法(例如BD Rhapsody™靶向测定)和用于确定5'转录物序列的方法(例如用于确定细胞的V(D)J区序列的方法)。使用Ab0来确定单细胞中的蛋白质表达谱以及样品追踪(例如,追踪样品来源)的实施方案在美国专利申请第15/715,028号(作为美国专利申请公布第2018/0088112号公布)和美国专利申请第15/937,713号中描述;每个专利申请的内容通过引用以其整体并入本文。确定5'转录物序列的实施方案已经描述于美国专利申请第62/739,795号中;该美国专利申请的内容通过引用以其整体并入本文。对于框1108处的随机引发和框1112处的延伸,该过程的参数包括珠的数量/uL、所使用的聚合酶、反应的长度、牛血清白蛋白(BSA)的浓度和Ampure清理条件。对于框1114处的PCR1和清理,该过程的参数包括所使用的聚合酶、大细胞或小细胞的循环数、Ampure清理条件和质量控制(QC)方法。对于框1116处的索引PCR和清理,该过程的参数包括所使用的聚合酶、PCR输入量、Ampure清理条件和质量控制方法。

[0349] 本文的公开内容包括对样品中的核酸靶进行标记的方法。在一些实施方案中,方法包括:使核酸靶的拷贝与多于一种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区;使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸;使随机引物与多于一种第一链条形码化多核苷酸接触,其中每种随机引物包含第二通用序列或其互补体;以及使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸以产生多于一种延伸产物。第一链条形码化多核苷酸可以包括条形码化脱氧核糖核酸(DNA)分子、条形码化核糖核酸(RNA)分子或二者。在不同实施方式中,随机引发和延伸的循环的数量可以不同。在一些实施方案中,随机引发和延伸的循环的数量可以包括以下,或包括约以下:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100个或这些值中的任何两个值之间的数字或范围的随机引发和延伸的循环。在一些实施方案中,随机引发和延伸的循环的数量可以包括至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90或100个的随机引发和延伸的循环。

[0350] 方法可以包括使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增多于一种延伸产物,从而产生第一多于一种条形码化扩增子。在不同的实施方式中,扩增(例如,线性扩增、PCR扩增)的循环的数量可以是不同的。在一些实施方案中,扩增可以包括以下,或包括约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90

个、100个或这些值中的任何两个值之间的数字或范围的扩增的循环。在一些实施方案中，线性扩增可以包括至少以下，或至多以下：1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个或100个的扩增的循环。第一多于一种条形码化扩增子可以对应于单细胞的mRNA的至少10%。第一多于一种条形码化扩增子可以对应于单细胞的mRNA的至少50%。第一多于一种条形码化扩增子可以对应于单细胞的mRNA的至少90%。在一些实施方案中，第一多于一种条形码化扩增子可以对应于单细胞的mRNA的至少5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%或这些值中的任何两个值之间的数字或范围。第一多于一种条形码化扩增子可以包含全转录组扩增(WTA)产物。扩增多于一种延伸产物可以包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到多于一种延伸产物。方法可以包括基于与第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量，确定样品中核酸靶的拷贝数。在一些实施方案中，扩增多于一种延伸产物不在固体支持物的存在下进行。在一些实施方案中，方法不包括核糖核酸酶H诱导的引发、末端修复和/或衔接子连接。在一些实施方案中，方法不包括片段化、标签片段化或二者。

[0351] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中，方法包括：使核酸靶的拷贝与多于一种寡核苷酸条形码接触，其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区；使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸；使随机引物与多于一种第一链条形码化多核苷酸接触，其中每种随机引物包含第二通用序列或其互补体；使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸以产生多于一种延伸产物；使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增多于一种延伸产物，从而产生第一多于一种条形码化扩增子；以及基于与第一多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量，确定样品中核酸靶的拷贝数。

[0352] 在一些实施方案中，使核酸靶的拷贝接触包括使多于一个核酸靶的拷贝与多于一种寡核苷酸条形码接触，使多于一种寡核苷酸条形码延伸包括使与多于一种核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸，并且确定样品中核酸靶的拷贝数包括基于与第一多于一种条形码化扩增子中的条形码化扩增子关联的具有不同序列的分子标记的数量来确定样品中多于一种核酸靶中的每一种的数量，所述第一多于一种条形码化扩增子包括多于一种核酸靶中的每一种的序列。多于一种核酸靶中的每一种的序列可以包括多于一种核酸靶中的每一种的子序列。第一多于一种条形码化

扩增子中的核酸靶的序列可以包括核酸靶的子序列。

[0353] 方法可以包括使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增第一多于一种条形码化扩增子,从而产生第二多于一种条形码化扩增子。方法可以包括使用能够与第一通用序列或其互补体杂交的引物和能够与第二通用序列或其互补体杂交的引物扩增多于一种延伸产物,从而产生第一多于一种条形码化扩增子。在不同的实施方式中,扩增(例如,线性扩增、PCR扩增)的循环的数量可以是不同的。在一些实施方案中,扩增可以包括以下,或包括约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个或这些值中的任何两个值之间的数字或范围的扩增的循环。在一些实施方案中,线性扩增可以包括至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个或100个的扩增的循环。第二多于一种条形码化扩增子可以对应于单细胞的mRNA的至少10%、单细胞的mRNA的至少50%或单细胞的mRNA的至少90%。在一些实施方案中,第二多于一种条形码化扩增子可以对应于单细胞的mRNA的至少5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%或这些值中的任何两个值之间的数字或范围。第二多于一种条形码化扩增子可以包含全转录组扩增(WTA)产物。扩增第一多于一种条形码化扩增子可以包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到第一多于一种条形码化扩增子。

[0354] 方法可以包括基于与第二多于一种条形码化扩增子或其产物关联的具有不同序列的分子标记的数量,确定样品中核酸靶的拷贝数。在一些实施方案中,使核酸靶的拷贝接触包括使多于一种核酸靶的拷贝与多于一种寡核苷酸条形码接触,使多于一种寡核苷酸条形码延伸包括使与多于一种核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸以产生多于一种第一链条形码化多核苷酸,和/或其中确定样品中核酸靶的拷贝数包括基于与第二多于一种条形码化扩增子中的条形码化扩增子关联的具有不同序列的分子标记的数量来确定样品中多于一种核酸靶中的每一种的数量,所述第二多于一种条形码化扩增子包括多于一种核酸靶中的每一种的序列。多于一种核酸靶中的每一种的序列可以包括多于一种核酸靶中的每一种的子序列。第二多于一种条形码化扩增子中的核酸靶的序列可以包括核酸靶的子序列。

[0355] 第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子中的每一种可以包含以下的至少一部分:第一通用序列、第二通用序列或二者。第一通用序列和第二通用

序列可以相同或不同。第一通用序列和/或第二通用序列可以包含测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点。测序衔接子可以包括P5序列、P7序列、其互补序列和/或其部分。测序引物可以包括读段1测序引物、读段2测序引物、其互补序列和/或其部分。

[0356] 随机引物可以包含随机核苷酸序列。随机核苷酸序列的长度可以为约4个至约30个核苷酸。在一些实施方案中,所述随机核苷酸序列的长度为6个或9个核苷酸。在不同实施方式中,随机核苷酸序列可以具有不同的长度。在一些实施方案中,随机引物内的随机核苷酸序列的长度是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个或这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,随机引物内的随机核苷酸序列的长度是至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个核苷酸。在不同的实施方式中,随机引物可以在随机引发步骤期间具有不同的浓度。在一些实施方案中,随机引物在随机引发期间的浓度是至少以下,或至多以下:1uM、2uM、3uM、4uM、5uM、6uM、7uM、8uM、9uM、10uM、11uM、12uM、13uM、14uM、15uM、16uM、17uM、18uM、19uM、20uM、25uM、30uM、35uM、40uM、45uM、50uM、60uM、70uM、80uM、90uM、100uM、110uM、120uM、128uM或这些值中的任何两个值之间的数字或范围的uM。

[0357] 使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸可以包括使用逆转录酶(例如,病毒逆转录酶)使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸。使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸可以包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使与核酸靶的拷贝杂交的多于一种寡核苷酸条形码延伸。使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸可以包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶(例如Klenow片段)使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸。在一些实施方案中,延伸酶是Klenow或Klenow exo⁻。

[0358] 在一些实施方案中,多于一种第一链条形码化多核苷酸可以被再用作另外的延伸和/或扩增反应的模板(在随机引发和延伸之后)。方法可以包括重复以下步骤:使随机引物与多于一种第一链条形码化多核苷酸接触,使与多于一种第一链条形码化多核苷酸杂交的随机引物延伸,并扩增多于一种延伸产物。方法可以包括使用多于一种第一链条形码化多核苷酸作为模板合成第三多于一种条形码化扩增子以产生第三多于一种条形码化扩增子。合成第三多于一种条形码化扩增子可以包括对多于一种第一链条形码化多核苷酸进行聚合酶链式反应(PCR)扩增。合成第三多于一种条形码化扩增子可以包括使用能够与第一通用序列或其互补体杂交的引物以及靶特异性引物进行的PCR扩增。方法可以包括获得第三多于一种条形码化扩增子或其产物的序列信息。获得序列信息可以包括将测序衔接子连接到第三多于一种条形码化扩增子或其产物。靶特异性引物可以与免疫受体(例如,T细胞受体(TCR)和/或B细胞受体(BCR))特异性地杂交。图12是以一个工作流程进行靶向基因表达谱分析、用Ab0进行样品多重化、用Ab0进行蛋白质表达谱分析,以及全转录组分析的示意性图示。在一些实施方案中,全转录组分析可以与以一个工作流程进行的靶向基因表达谱系

分析、用Ab0的样品多重化和用Ab0的蛋白质表达谱系分析中的一种或更多种相容,并且可以以一个工作流程进行。在一些实施方案中,对第一链条形码化多核苷酸的一部分(例如,在细胞捕获和条形码化之后的珠)进行二次取样以用于RPE WTA,并且将剩余的第一链条形码化多核苷酸用于靶向(例如,基因特异性)扩增和分析(例如,靶向PCR2和靶向索引PCR)。在一些实施方案中,在RPE WTA之后,第一链条形码化多核苷酸可以使用本文公开的多重化引物组再用于靶向(例如,基因特异性)扩增和分析。在一些实施方案中,从本文提供的基于RPE的WTA方法获得的测序信息用于产生用于本文公开的靶向(例如,基因特异性)扩增和分析的一组引物。

[0359] 样品可以包括多于一个细胞、多于一个单细胞、组织、肿瘤样品,或它们的任何组合。样品可以包括外周血单核细胞或免疫细胞。免疫细胞可以包括B细胞、T细胞,或它们的组合。核酸靶可以包括核酸分子。核酸分子可以包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、含有多(A)尾的RNA,或它们的任何组合。mRNA编码免疫受体。多于一种核酸靶中的一种或更多种可以包括低表达基因的mRNA。核酸靶可以包含细胞组分结合试剂。核酸分子可以与细胞组分结合试剂关联。方法可以包括使核酸分子和细胞组分结合试剂解离。靶结合区可以包含寡聚dT序列、随机序列、靶特异性序列或它们的组合。靶结合区可以包含多(dT)区,并且核酸靶可以包含多(dA)区。

[0360] 方法可以包括多于一种延伸产物、第一多于一种条形码化扩增子和/或第二多于一种条形码化扩增子的清理(例如,ampure清理)。在不同的实施方式中,SPRI清理比率可以是不同的。在一些实施方案中,SPRI清理比率为约0.1x至约3.0x(例如,0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2.0、2.1、2.2、2.5、2.7、3.0和其间的范围)。清理可以是单侧或双侧的。

[0361] 多于一种寡核苷酸条形码中的至少10种可以包含不同的分子标记序列。多于一种寡核苷酸条形码的每种分子标记可以包含至少6个核苷酸。多于一种寡核苷酸条形码可以与固体支持物关联。与相同固体支持物关联的多于一种寡核苷酸条形码可以各自包含相同的样品标记。多于一种寡核苷酸条形码的每种样品标记可以包含至少6个核苷酸。多于一种寡核苷酸条形码可以各自包含细胞标记。多于一种寡核苷酸条形码的每种细胞标记可以包含至少6个核苷酸。与相同固体支持物关联的寡核苷酸条形码可以包含相同的细胞标记。与不同固体支持物关联的寡核苷酸条形码可以包含不同的细胞标记。固体支持物可以包括合成颗粒、平坦表面,或它们的组合。在本文提供的反应的不同实施方式中,固体支持物(例如珠)可以具有不同的密度。在一些实施方案中,在随机引发和延伸反应期间,固体支持物的密度是以下,或是约以下:1个固体支持物/u1、2个固体支持物/u1、3个固体支持物/u1、4个固体支持物/u1、5个固体支持物/u1、6个固体支持物/u1、7个固体支持物/u1、8个固体支持物/u1、9个固体支持物/u1、10个固体支持物/u1、11个固体支持物/u1、12个固体支持物/u1、13个固体支持物/u1、14个固体支持物/u1、15个固体支持物/u1、16个固体支持物/u1、17个固体支持物/u1、18个固体支持物/u1、19个固体支持物/u1、20个固体支持物/u1、25个固体支持物/u1、30个固体支持物/u1、35个固体支持物/u1、40个固体支持物/u1、45个固体支持物/u1、50个固体支持物/u1、60个固体支持物/u1、70个固体支持物/u1、80个固体支持物/u1、90个固体支持物/u1、100个固体支持物/u1、110个固体支持物/u1、120个固体支持物/u1、128个固体支持物/u1、130个固体支持物/u1、140个固体支持物/u1、150个固体支持物/

u1、160个固体支持物/u1、170个固体支持物/u1、180个固体支持物/u1、190个固体支持物/u1、200个固体支持物/u1、210个固体支持物/u1、220个固体支持物/u1、230个固体支持物/u1、240个固体支持物/u1、250个固体支持物/u1、260个固体支持物/u1、270个固体支持物/u1、280个固体支持物/u1、290个固体支持物/u1、300个固体支持物/u1、310个固体支持物/u1、320个固体支持物/u1、330个固体支持物/u1、340个固体支持物/u1、350个固体支持物/u1、360个固体支持物/u1、370个固体支持物/u1、380个固体支持物/u1、390个固体支持物/u1、400个固体支持物/u1、410个固体支持物/u1、420个固体支持物/u1、430个固体支持物/u1、440个固体支持物/u1、450个固体支持物/u1、460个固体支持物/u1、470个固体支持物/u1、480个固体支持物/u1、490个固体支持物/u1、500个固体支持物/u1、510个固体支持物/u1、520个固体支持物/u1、530个固体支持物/u1、540个固体支持物/u1、550个固体支持物/u1、560个固体支持物/u1、570个固体支持物/u1、580个固体支持物/u1、590个固体支持物/u1、600个固体支持物/u1、610个固体支持物/u1、620个固体支持物/u1、630个固体支持物/u1、640个固体支持物/u1、650个固体支持物/u1、660个固体支持物/u1、670个固体支持物/u1、680个固体支持物/u1、690个固体支持物/u1、700个固体支持物/u1、710个固体支持物/u1、720个固体支持物/u1、730个固体支持物/u1、740个固体支持物/u1、750个固体支持物/u1、760个固体支持物/u1、770个固体支持物/u1、780个固体支持物/u1、790个固体支持物/u1、800个固体支持物/u1、810个固体支持物/u1、820个固体支持物/u1、830个固体支持物/u1、840个固体支持物/u1、850个固体支持物/u1、860个固体支持物/u1、870个固体支持物/u1、880个固体支持物/u1、890个固体支持物/u1、900个固体支持物/u1、910个固体支持物/u1、920个固体支持物/u1、930个固体支持物/u1、940个固体支持物/u1、950个固体支持物/u1、960个固体支持物/u1、970个固体支持物/u1、980个固体支持物/u1、990个固体支持物/u1、1000个固体支持物/u1或这些值中的任何两个值之间的数字或范围的固体支持物/u1。在一些实施方案中,所有固体支持物用于随机引发和延伸反应。在一些实施方案中,仅一部分固体支持物用于随机引发和延伸反应(例如,二次取样)中。在一些这样的实施方案中,二次取样包括将不包含第一链条形码化多核苷酸的固体支持物(例如,“冷”珠)添加到二次取样的固体支持物。

[0362] 样品可以包含单细胞,方法包括将包含多于一种寡核苷酸条形码的合成颗粒与样品中的单细胞关联。方法可以包括在将合成颗粒与单细胞关联后裂解单细胞。裂解单细胞可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。合成颗粒和单细胞可以在相同的分区(例如,孔或液滴)中。多于一种寡核苷酸条形码中的至少一种可以被固定或部分地固定在合成颗粒上。多于一种寡核苷酸条形码中的至少一种可以被包封或部分地包封在合成颗粒中。合成颗粒可以是可破坏的(例如,可破坏的水凝胶颗粒)。合成颗粒可以包括珠。珠可以包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、水凝胶珠、凝胶珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。合成颗粒可以包含选自以下组成的组的材料:聚二甲基硅氧烷(PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。多于一种寡核苷酸条形码中的每一种可以包含接头官能团,合成颗粒可以包含固体支持物官能团,和/或支持物官能团

和接头官能团彼此关联。接头官能团和支持物官能团可以单独地选自由以下组成的组：C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮，以及它们的任何组合。

[0363] 基于RPE的WTA性能

[0364] 本文公开的基于RPE的WTA方法可以导致低丰度转录物的灵敏度提高、检测新的转录物、检测新的细胞类型和/或降低测序成本。与可选的非基于RPE的WTA方法（例如，基于连接或基于片段化的WTA方法）相比，本文提供的基于RPE的WTA方法的一些实施方案导致低丰度转录物的灵敏度提高、检测新的转录物、检测新的细胞类型和/或降低了测序成本。如本文使用的，“靶”可以与“种类”互换使用。

[0365] 如本文使用的，“种类 (species)”是指彼此相同、彼此为互补体或反向互补体，或者能够彼此杂交，或者是来自相同遗传基因座的转录物，或者编码相同的蛋白质或其片段，等等的多核苷酸（例如，单链多核苷酸，诸如cDNA分子、用于样品追踪的样品索引化寡核苷酸和用于确定蛋白质表达谱的蛋白质特异性寡核苷酸）。如本文使用的，“靶”可以与“种类”互换使用。在一些实施方案中，种类的成员（例如，拷贝或出现物 (occurrences)）彼此或与其互补体至少80%、至少90%、至少95%、至少98%、至少99%或100%同源。在一些实施方案中，种类的成员是来自相同遗传基因座的转录物，并且转录物可以具有相同或不同的长度。在一些实施方案中，种类是cDNA或mRNA。

[0366] 如本文使用的，“高丰度种类”是指大量存在的种类，例如该种类可以是以下，是约以下，是至少以下，或是至少约以下：1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%或更多。在一些实施方案中，样品可以包含至少1种、至少2种、至少3种、至少4种、至少5种、至少10种、至少20种、至少50种、至少100种、至少200种、至少500种、至少1000种或更多种高丰度种类。在一些实施方案中，所有高丰度种类的总和代表样品中核酸分子或种类的至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%或更多。在一些实施方案中，高丰度种类可以包括编码一种或更多种核糖体蛋白的多核苷酸。在一些实施方案中，高丰度种类可以包括编码一种或更多种线粒体蛋白质的多核苷酸。在一些实施方案中，高丰度种类可以包括编码一种或更多种管家蛋白的多核苷酸。在一些实施方案中，高丰度种类可以包括样品索引化寡核苷酸的序列（在本文也称为样品标签序列或样品索引化索引 (sample indexing indices)）。在一些实施方案中，高丰度种类可以包括用于样品追踪的寡核苷酸序列。在一些实施方案中，高丰度种类可以包括样品索引化序列，用于鉴定样品的一个或多个细胞的样品来源。在一些实施方案中，高丰度种类可以包括用于细胞组分结合试剂（例如，抗体）的独特标识符序列。在一些实施方案中，高丰度种类可以包括与细胞组分结合试剂缀合（或预先缀合）的寡核苷酸序列。在一些实施方案中，高丰度种类可以包括在本文中可以被称为抗体寡核苷酸（缩写为“AbOligo”或“AbO”）的与抗体缀合或预先缀合的寡核苷酸序列。

[0367] 如本文使用的，“中丰度种类”是指以低于至少一种种类且高于至少一种其他种类的量存在的种类。在一些实施方案中，中丰度种类可以是以下，是至少以下，是约以下，或是至少约以下：样品中核酸分子或种类的10%、5%、4%、3%、2%、1%、0.1%、0.01%，或这些值中的任何两个值之间的范围。在一些实施方案中，样品可以包括至少1种、至少2种、至少3种、至少4种、至少5种、至少10种、至少20种、至少50种、至少100种、至少200种、至少500种、

至少1000种或更多种中丰度种类。在一些实施方案中,所有中丰度种类的总量代表样品中核酸分子的约1%、约2%、约3%、约4%、约5%、约10%、约20%、约30%或上述值中的任何两个值之间的范围。在一些实施方案中,中丰度种类可以包括编码一种或更多种管家蛋白的多核苷酸。在一些实施方案中,中丰度种类可以包括用于样品追踪的寡核苷酸序列。在一些实施方案中,中丰度种类可以包括样品索引化序列,用于鉴定样品的一个或更多细胞的样品来源。在一些实施方案中,中丰度种类可以包括用于细胞组分结合试剂(例如,抗体)的独特标识符序列。在一些实施方案中,中丰度种类可以包括与细胞组分结合试剂缀合(或预先缀合)的寡核苷酸序列。在一些实施方案中,中丰度种类可以包括在本文中可以被称为抗体寡核苷酸(缩写为“AbOligo”或“AbO”)的与抗体缀合或预先缀合的寡核苷酸序列。

[0368] 如本文使用的,“低丰度种类”是指以低量存在的种类,例如该种类可以是以下,是约以下,少于以下,或少于约以下:样品中核酸分子或种类的1%、0.1%、0.01%、0.001%、0.0001%或更少。在一些实施方案中,样品可以包括至少1种、至少2种、至少3种、至少4种、至少5种、至少10种、至少20种、至少50种、至少100种、至少200种、至少500种、至少1000种或更多种低丰度种类。在一些实施方案中,所有低丰度种类的总量代表样品中核酸分子的少于20%、少于10%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.1%或更少。在一些实施方案中,低丰度种类可以包括编码一种或更多种转录因子的多核苷酸。在一些实施方案中,低丰度种类可以包括编码一种或更多种T细胞受体的多核苷酸。在一些实施方案中,低丰度种类可以包括编码一种或更多种抗体的多核苷酸。

[0369] 在一些实施方案中,多于一种核酸靶中的两种或更多种包括低丰度种类。在一些实施方案中,多于一种核酸靶中的两种或更多种包括低丰度种类和/或中丰度种类。在一些实施方案中,多于一种核酸靶中的两种或更多种包括低表达基因的mRNA。在一些实施方案中,代表低丰度种类的核酸靶的百分比可以是以下,或是约以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,代表低丰度种类的核酸靶的百分比可以是至少以下,或至多以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。

[0370] 在一些实施方案中,RPE WTA文库包括本文公开的RPE WTA文库制备的产物。在一些实施方案中,RPE WTA文库包括本公开内容的一种或更多种下游方法的产物,所述方法采用RPE WTA文库制备的产物作为模板。在一些实施方案中,与通过可选的非基于RPE的WTA方法(例如,基于连接或基于片段化的WTA方法)产生的文库的测序相比,通过本公开内容的RPE WTA方法产生的RPE WTA文库的测序产生了一种或更多种低丰度种类(例如,核酸靶、低表达的mRNA)的读段数量的增加。在一些实施方案中,与非基于RPE的WTA文库相比,RPE WTA文库中一种或更多种低丰度种类的读段数量的增加为至少2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%或更高,以及其中的重叠范围)。在一些实施方案中,与非RPE WTA文库的测序相比,通过本公开内容的方法产生的RPE WTA文库的测序产生一种或更多种低丰度种类(例如,核酸靶、低表达的mRNA)的读段数量相对于总读段数量的增加。在一些实施方案中,与非RPE WTA文库相比,一种或更多种低丰度种类的读段数量相对于RPE WTA文库中总读段数量的增加为至少2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%或更高,以及其中的重叠范围)。在一些实施方案中,与非RPE WTA文库的测序相比,通过本公开内容的方法产生的RPE WTA文库的测序产生了一种或更多种低丰度种类(例如,核酸靶、低表达的mRNA)的分子索引(MI)的数量的增加。在一些实施方案中,与非RPE WTA文库相比,一种或更多种低丰度种类的MI数量相对于RPE WTA文库中总读段数量的增加为至少2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%或更高,以及其中的重叠范围)。

[0371] 在一些实施方案中,本文公开的方法可以在RPE WTA文库制备后(与非RPE WTA文库制备,例如基于连接或基于片段化的WTA方法相比)增加一种或更多种低丰度种类的相对丰度。例如,本文公开的方法可以在RPE WTA文库制备后(与非RPE WTA文库制备相比)增加至少1种、至少2种、至少3种、至少4种、至少5种、至少10种、至少20种、至少50种、至少100种、至少200种、至少500种、至少1000种或更多种低丰度种类的相对丰度。在一些实施方案中,本文公开的方法可以在RPE WTA文库制备后(与非RPE WTA文库制备相比)使一种或更多种低丰度种类中的每一种的相对丰度增加至少约2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%、150%、200%、250%、500%、1000%或更高,以及其中的重叠范围)。在一些实施方案中,本文公开的方法可以在RPE WTA文库制备后(与非RPE WTA文库制备相比)使一种或更多种低丰度种类中的至少一种的相对丰度增加至少约2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%、150%、200%、250%、500%、1000%或更高,以及其中的重叠范围)。在一些实施方案中,本文公开的方法可以在RPE WTA文库制备后(与非RPE WTA文库制备相比)使总低丰度种类的相对丰度增加至少约2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%、150%、200%、250%、500%、1000%或更高,以及其中的重叠范围)。

[0372] 在一些实施方案中,本文公开的方法可以在RPE WTA文库制备后(与制备非RPE

WTA文库相比)选择性地增加一种或更多种低丰度种类的相对丰度,使得在RPE WTA文库制备后,低丰度种类的丰度相对于一种或更多种高丰度种类增加。在一些实施方案中,本文公开的方法和组合物可以在RPE WTA文库制备后(与非RPE WTA文库相比)使一种或更多种低丰度种类中的每一种的相对丰度增加、增加约、增加至少或增加至少约2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%、150%、200%、250%、500%、1000%或更高,以及其中的重叠范围),使得在RPE WTA文库制备后低丰度种类的丰度相对于一种或更多种高丰度种类增加。

[0373] 在一些实施方案中,与非RPE WTA文库的测序相比,通过本公开内容的方法产生的RPE WTA文库的测序产生了一种或更多种靶的相似基因表达谱和/或蛋白质表达谱(例如,基因表达组结果)。在一些实施方案中,RPE WTA文库和非RPE WTA文库之间的一种或更多种靶的基因表达谱和/或蛋白质表达谱的差异可以是至多0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%,或者这些值中的任何两个值之间的数字或范围。在一些实施方案中,RPE WTA文库和非RPE WTA文库的表达谱之间的 R^2 相关性可以是0.6、0.7、0.8、0.9、0.990、0.999、1.0以及其中的重叠范围。对文库结果的表达谱之间的相关性进行测试的方法是本领域中良好理解的(例如,tSNE图的重叠)。

[0374] 将理解的是,通过增加RPE WTA文库中的中丰度种类或低丰度种类的测序读段,和/或减少非RPE WTA文库中的高丰度种类的测序读段,RPE WTA文库可以提高测序效率。在通过本公开内容的方法产生的RPE WTA文库中,与非RPE WTA文库相比,可以更容易地鉴定不太丰富(例如,更罕见)的核酸靶。与非RPE WTA文库相比,RPE WTA文库中不太丰富的靶的测序读段可以构成总读段的更大部分。与非RPE WTA文库中相同靶的读段相比,RPE WTA文库中一种或更多种不太丰富的靶的测序读段可以包括至少50%、100%、150%、200%、250%、300%、350%、400%、450%或500%或更多的读段。归一化文库中一种或更多种不太丰富的靶的测序读段可以是非RPE WTA文库中相同靶的测序读段的至少1倍、2倍、3倍、4倍、5倍或6倍或更多倍。

[0375] 本文描述的用于产生RPE WTA文库的方法可以提高多于一种核酸(例如核酸文库)中的中丰度核酸靶和/或低丰度核酸靶的测序读段。例如,多于一种中丰度核酸靶的测序读段可以是RPE WTA文库的总测序读段的至少30%、至少20%、至少10%、至少5%。在一些实施方案中,多于一种低丰度核酸靶的测序读段可以是RPE WTA文库的总测序读段的至少50%、至少40%、至少30%、至少20%、至少10%、至少5%、至少4%、至少3%、至少2%、至少1%。在一些实施方案中,多于一种低丰度核酸靶的测序读段是RPE WTA文库的总测序读段的至少5%。在一些实施方案中,多于一种低丰度核酸靶的测序读段是RPE WTA文库的总测序读段的至少10%。在一些实施方案中,多于一种低丰度核酸靶的测序读段是RPE WTA文库的总测序读段的至少20%。在一些实施方案中,多于一种低丰度核酸靶的测序读段是RPE WTA文库的总测序读段的至少30%。在一些实施方案中,多于一种低丰度核酸靶的测序读段是RPE WTA文库的总测序读段的至少40%。在一些实施方案中,多于一种低丰度核酸靶的测序读段是RPE WTA文库的总测序读段的至少50%。

[0376] 与可选的非基于RPE的WTA方法(例如,基于连接或基于片段化的WTA方法)相比,本

文公开的基于RPE的WTA方法的一些实施方案有利地具有更短的方案时间。在一些实施方案中,与基于非RPE的WTA方法(例如,基于连接或基于片段化的WTA方法)相比,从单细胞捕获到cDNA合成的时间和/或从文库制备到QC的时间减少了至少约2%(例如,2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、75%、100%或更高,以及其中的重叠范围)。

实施例

[0377] 上文讨论的实施方案的一些方面在以下实施例中进一步详细公开,其并不是以任何方式意在限制本公开内容的范围。

[0378] 用于全转录组分析(WTA)的随机引发和延伸(RPE)方案

[0379] 下文提供的该方案概述了使用随机引发方法用于从Rhapsody珠产生全转录组文库的说明性实施方案。在实施例1中采用了与下文提供的方案类似的程序。

[0380] 试剂

[0381] BD Rhapsody靶向mRNA和AbSeq扩增试剂盒(BD Biosciences,633774);SPRIselect试剂(Beckman Coulter Life Sciences,B23318);100%乙醇;无核酸酶水;Klenow片段(3'→5' exo-) (New England Biolabs,M0212L);10mM dNTP(New England Biolabs);和Nonamer 2(5' TCA GAC GTG TGC TCT TCC GAT CTNNNNNNNNN 3';SEQ ID NO: 1)预稀释至100uM的浓度(通过手动混合达到混合碱基为25%/核苷酸含量)。

[0382] 耗材

[0383] 移液器(P10,P20,P200,P1000);低保留过滤移液器吸头;1.5mL微量离心管(Eppendorf,022431021);DNA LoBind管,1.5mL(Eppendorf,0030108051);和0.2mL PCR联排管(strip tubes)。

[0384] 设备

[0385] 用于1.5-2.0mL管的微量离心机;用于0.2mL管的微量离心机;涡旋器;Pipet-Aid;用于1.5mL管的6-管磁分离架(New England Biolabs,S1506S);用于0.2mL,8-联排管的低剖面磁分离支架(V&P Scientific,Clontech;VP772F4-1,635011);数字定时器;多通道移液器(P20,P200);加热块;和热混合仪。

[0386] 程序

[0387] 产生不具有mRNA的珠(“冷”珠)

[0388] 1.将一管Rhapsody珠(包含~600,000个珠—足够添加到3套盒珠的冷珠)与磁体孵育。去除所有液体,并且用750μl样品缓冲液重悬沉淀。

[0389] 2.将管与磁体孵育,并弃去所有液体。

[0390] 3.将珠与42μl 1x裂解缓冲液(具有5mM DTT)孵育2分钟。具有5mM DTT的裂解缓冲液通过将10uL的1M DTT和2mL裂解缓冲液混合而制成。

[0391] 4.然后各自用1mL 1x裂解缓冲液(具有5mM DTT)稀释,与磁体孵育,并去除所有液体。

[0392] 5.添加1.0mL珠洗涤缓冲液。在与磁体混合和孵育后,去除所有液体。

[0393] 6.添加1.0mL珠洗涤缓冲液。

[0394] 7.如表1中描绘地制备RT反应缓冲液混合物。

[0395] 表1.RT缓冲液混合物

组分	终浓度	1x	1.05x	2.1x
5x RT缓冲液 (PN 650000067)	1x	120.0	126	252
100mM DTT (PN 650000068)	5mM	30.0	31.5	63
20mg/mL BSA (NEB B9000S)	1.2mg/mL	36.0	37.8	75.6
无核酸酶水		414.0	434.7	869.4
总计		600.0	630	1260

[0397] 8. 将珠与磁体孵育, 并去除所有液体。

[0398] 9. 将600μL RT缓冲液混合物添加到珠, 并通过移液进行混合。

[0399] 10. 将珠在热混合仪中以1200RPM摇动在37℃孵育20分钟。

[0400] 11. 将管短暂置于冰上后, 将它们与磁体孵育, 并去除所有液体。

[0401] 12. 如表2中描绘地制备核酸外切酶I混合物。

[0402] 表2. 核酸外切酶I反应混合物

组分	终浓度	1x	1.05x	2.05x
10x ExoI缓冲液 (NEB)	1x	60.0	63	123
20U/μL核酸外切酶I (NEB)	1U/μL	30.0	31.5	61.5
无核酸酶水		510.0	535.5	1045.5
总计		600.0	630	1230

[0404] 13. 将600μL Exo I混合物添加到珠, 并通过移液进行混合。

[0405] 14. 将珠在热混合仪中以1200RPM摇动在37℃孵育30分钟。

[0406] 15. 将珠移至80℃持续20分钟不摇动, 并且然后置于冰上1分钟。

[0407] 16. 将珠与磁体孵育, 并去除所有液体。

[0408] 17. 每个珠沉淀用600μL珠重悬缓冲液重悬。所得浓度为大致1000个珠/μL。在一些情况下, 使用扫描仪上的珠计数程序确定最终的珠浓度。

[0409] 在具有cDNA的BD Rhapsody珠上的随机引发和延伸 (RPE)

[0410] 1. 在预扩增工作区, 在一个新的1.5mL LoBind管中, 按所示顺序移取表3中列出的组分。

[0411] 表3. 随机引物混合物

组分	终浓度	1x (200uL)	1.2x	4.4x
水		135	162	594
10x NEBuffer 2	1x	20	24	88
100 μM Randomer 2 (Nonamer)	10 μM	20	24	88
总计		175	211.2	774.4

[0413] 2. 通过温和移液将随机引物混合物混合。

[0414] 3. 在一些情况下, 对经核酸外切酶I处理的BD Rhapsody细胞捕获珠的整个样品进行随机引发和延伸, 并且程序跳至步骤7。在一些情况下, 对经核酸外切酶I处理的BD Rhapsody细胞捕获珠的子样品进行随机引发和延伸, 并且用户继续进行到步骤4。

[0415] 4. 对于经核酸外切酶I处理的BD Rhapsody细胞捕获珠的二次取样, 基于于珠上捕

获的有活力的细胞在最终珠重悬体积中的预期数量,确定待二次取样用于靶向测序的珠的体积。在一些情况下,将剩余的珠储存在4℃长达3个月。通过移液管混合将珠完全重悬,并且然后将计算的珠悬浮液体积移液至新的1.5mL LoBind管中。当使用二次取样的珠时,用珠重悬缓冲液将总体积升至100μL。

[0416] 5.将来自盒实验的珠与“冷”珠合并,使得最终珠体积为200uL(珠总数为约200,000个)。例如,在40uL的珠具有2000个细胞的情况下,添加160uL冷珠。可选地,在一些情况下,计算如下进行:待添加的冷珠数量=200000*(1-(待二次取样的细胞/珠洗涤后的总细胞))。

[0417] 6.将具有合并的珠的管置于95℃的加热块中2分钟。然后立即将具有BD Rhapsody细胞捕获珠的管置于1.5mL磁体上,并且立即取出上清液并弃去。避免了BD Rhapsody细胞捕获珠变干。

[0418] 7.将管从磁体上取下,并且然后用低保留尖端将175μL随机引物混合物移液至管中。进行移液管混合10次以使珠重悬。

[0419] 8.将管置于95℃的加热块中2分钟,在37℃于1,200rpm持续5分钟,在25℃于1,200rpm持续5分钟,并且然后将样品旋转沉降(spindown)。

[0420] 9.将25μL引物延伸反应混合物(如表4中示出的)移液至含有珠的样品管中(总体积为200μL)。进行移液管混合10次。

[0421] 表4. 引物延伸反应混合物

组分	终浓度	1x (200uL)	1.2x	4.4x
10 mM dNTP	0.45 mM	9	10.8	39.6
20 mg/mL BSA (NEB B9000S)	1.0 mg/mL	10	12	44
5000 U/mL Klenow Exo- (NEB M0212S)	150 U/mL	6	7.2	26.4
总计		25	30	110

[0423] 10.将管置于热混合仪中在37℃于1,200rpm持续30分钟。

[0424] 11.将管置于95℃的加热块中2分钟,在热混合仪上摇动~20秒,置于1.5mL磁体中,并将上清液(随机引物延伸产物,即RPE产物)立即转移到新的1.5mL LoBind管中,并将该管保存在冰上。

[0425] 12.在一些情况下,将珠重悬于200μL珠重悬缓冲液(PN 650000066)中。将管置于95℃的加热块中2分钟,置于1.5mL磁体中,并将上清液立即转移到新管中。

[0426] 13.在一些情况下,将200μL冷珠重悬缓冲液(PN 650000066)与剩余的珠一起移液至管中。珠仅通过移液管混合温和重悬(并且不涡旋)。将珠储存在4℃在预扩增工作区中。剩余的珠可以用于BD Rhapsody靶向的和BD AbSeq文库生成。

[0427] 纯化RPE产物

[0428] 1.在新的5.0mL LoBind管中,通过将4.0mL无水乙醇移液至1.0mL无核酸酶水中,制备5mL新鲜的80%(v/v)乙醇。将管涡旋10秒。制备新鲜的80%乙醇并在不到24小时内使用。

[0429] 2.将SPRI Select珠在使用前以高速涡旋1分钟。珠看起来颜色均匀且一致。

[0430] 3.将320μL(1.6x) SPRI Select珠移液至含有200μL RPE产物上清液的管中。

[0431] 4.进行移液管混合至少10次,并且然后短暂离心。

- [0432] 5.将悬浮液在室温孵育10分钟,并且然后将其置于1.5mL管磁体上5分钟。
- [0433] 6.去除上清液。
- [0434] 7.将原始管与磁体上的珠保持在一起,将600μL新鲜的80%乙醇温和移液至管中。
- [0435] 8.将样品在磁体上孵育30秒。
- [0436] 9.在磁体上时,小心地去除上清液,并适当地弃去,而不干扰SPRI Select珠。
- [0437] 10.重复乙醇洗涤步骤,总共进行两次80%乙醇洗涤。
- [0438] 11.在磁体上时,使用小体积移液管将任何残留的上清液从管去除。
- [0439] 12.将管在磁体上打开,以在室温干燥SPRI Select珠10-15分钟,或者直到珠看起来不再有光泽。
- [0440] 13.将管从磁体取下,并将31μL洗脱缓冲液(PN 650000075)移液到管中。对悬浮液进行移液管混合至少10次,直到溶液看起来呈悬浮状且颜色均匀。
- [0441] 14.将样品在室温孵育2分钟。
- [0442] 15.将管短暂离心,以收集处于底部的内容物。
- [0443] 16.将管置于磁体上,直到溶液澄清(通常≤30秒)。
- [0444] 17.将全部洗出物(~30μL)移液至新的PCR管中。这是纯化的BD Rhapsody RPE产物。将其保持在冰上或4℃,直到PCR。

[0445] RPE PCR

- [0446] 1.在预扩增工作区,在新的1.5mL LoBind管中,将表5中描绘的组分按所示顺序进行移液。

[0447] 表5.RPE PCR混合物

组分	终浓度	1x	1.2x	4.4x
随机引发1产物		30.0		
2x PCR MasterMix		50.0	60	220
通用寡聚物(10μM)	1μM	10.0	12	44
AbSeq PCR1引物(10μM)	1μM	10.0	12	44
总计		100.0	84	308

- [0449] 2.将70μL RPE PCR混合物与30μL纯化的RPE产物合并。
- [0450] 3.将RPE PCR反应分到2个PCR管中,每管50μL反应混合物。
- [0451] 4.运行表6中描绘的PCR程序,其中根据表7确定PCR循环数。

[0452] 表6.PCR程序

步骤	循环	温度	时间
热启动	1	95℃	3 min
变性	参见表 7	95℃	30 s
退火		60℃	1 min
延伸		72℃	1 min
最终延伸	1	72℃	2 min
保持	1	4℃	∞

[0455] 表7.PCR循环数

	扩增中的细胞数量	对于静息 PBMC, 建议的启动 PCR 循环
	500 个细胞	17
[0456]	1,000 个细胞	16
	2,500 个细胞	15
	5,000 - 10,000 个细胞	14
	10,000 个细胞	13

[0457] 5. 当RPE PCR反应完成时,将管短暂离心以收集处于管底的内容物。

[0458] 6. 将这两个反应物合并到新的0.2mL PCR管中。

[0459] RPE PCR扩增产物的纯化(单侧清理)

[0460] 1. 在一些情况下,取2μL的RPE PCR反应产物进行生物分析仪分析(稍后与纯化的产物进行比较)。

[0461] 2. 将SPRI Select珠置于室温,并将SPRI Select珠以高速涡旋1分钟。珠看起来颜色均匀且一致。

[0462] 3. 将具有最终扩增产物的管短暂离心。

[0463] 4. 将100.0μL (1.0x) SPRI Select珠移液至最终扩增产物中。

[0464] 5. 进行移液管混合至少10次,并且然后将样品短暂离心。

[0465] 6. 将悬浮液在室温孵育5分钟,并且然后置于联排管磁体上3-5分钟。

[0466] 7. 在磁体上时,用户小心地取出并适当地仅弃去上清液,而不干扰SPRI Select珠。

[0467] 8. 将管保持在磁体上,将200μL新鲜的80%乙醇温和移液。

[0468] 9. 将样品在磁体上孵育30秒。

[0469] 10. 在磁体上时,用户小心地取出并适当地仅弃去上清液,而不干扰SPRI Select珠。

[0470] 11. 重复80%乙醇洗涤,总共进行两次洗涤。

[0471] 12. 将管保持在磁体上,使用小体积移液管任何残留的上清液从管去除。

[0472] 13. 将管在磁体上打开,以在室温干燥SPRI Select珠至少5分钟,或者直到珠看起来不再有光泽。

[0473] 14. 将30μL洗脱缓冲液(PN 650000075)移液至管中,并进行脉冲涡流,以完全重悬SPRI Select珠。

[0474] 15. 将样品在室温孵育2分钟。

[0475] 16. 将管短暂离心,以收集处于底部的内容物。

[0476] 17. 将管置于磁体上,直到溶液澄清(通常≤30秒)。

[0477] 18. 将全部洗出物(~30μL)移液至新的1.5mL LoBind管中。RPE PCR洗出物现已准备好进行索引PCR。在一些情况下,这一步骤是停止点,并且RPE PCR文库可以在-20℃储存≤6个月。

[0478] 19. 用户使用Qubit dsDNA HS测定用Qubit™荧光计和使用Agilent高灵敏度DNA试剂盒用Agilent 2100生物分析仪对RPE PCR产物进行定量和进行质量控制。生物分析仪迹

线适当地显示了200bp到1000bp的宽峰。200bp到1000bp的浓度用于计算多少模板添加在索引PCR中。在一些情况下,可以观察到约70bp的小峰。在一些情况下,如果70bp峰是显著的,索引PCR后的产量受到影响。

[0479] WTA索引PCR

[0480] 1.将RPE PCR洗出物稀释,使得200-1000bp峰的浓度为~1-2nM。例如,如果200-1000bp峰的生物分析仪测量值为~6nM,则将样品以1:3稀释至2nM。

[0481] 2.将表8中描绘的组分按所示顺序移液至新的1.5mL管中。

[0482] 表8.WTA索引PCR混合物

	组分	终浓度	1X	1.2X	4.4X
[0483]	RPE PCR 洗出物(模板)	~1-2 nM 的 200-1000 bp 峰 (如由生物分析仪测量的)	3		
	2 x PCR MasterMix	1 x	25	30	110
[0484]	文库正向引物(10 μ M)	1 μ M	5	6	22
	文库反向引物(10 μ M)	1 μ M	5	6	22
	水		12	14.4	52.8
	总计		50	56.4	184.8

[0485] 3.将WTA索引PCR混合物与PCR1模板如下进行组合:(a)对于1个样品,将47 μ L的WTA索引PCR混合物与3 μ L的1-2nM的RPE PCR洗出物(模板)组合;或者(b)对于具有不同反向文库索引的多个样品,将WTA索引PCR混合物与5 μ L反向索引和3 μ L RPE PCR洗出物(模板)组合。

[0486] 4.运行表9中描绘的PCR程序。

[0487] 表9.索引PCR程序

	步骤	循环	温度	时间
[0488]	热启动	1	95 $^{\circ}$ C	5 min
	变性	6-8	95 $^{\circ}$ C	30 s
	退火		60 $^{\circ}$ C	30 s
	延伸		72 $^{\circ}$ C	30 s
	最终延伸	1	72 $^{\circ}$ C	1 min
	保持	1	4 $^{\circ}$ C	∞

[0489] 5.当WTA索引PCR完成时,将管短暂离心以收集处于管底的内容物。

[0490] WTA索引PCR产物的纯化(双侧清理)

[0491] 1.在一些情况下,取2 μ L的WTA索引PCR反应产物进行生物分析仪分析(稍后与纯化的产物进行比较)。

[0492] 2.将SPRI Select珠置于室温,并将SPRI Select珠以高速涡旋1分钟。珠看起来颜色均匀且一致。

[0493] 3.将具有WTA索引PCR产物的管短暂离心。

[0494] 4.将50 μ L水添加到WTA索引PCR产物,最终体积为100 μ L。

[0495] 5.将55 μ L (0.55x) SPRI Select珠移液至WTA索引PCR产物中。

[0496] 6.进行移液管混合至少10次,并且然后将样品短暂离心。

- [0497] 7. 将悬浮液在室温孵育5分钟,并且然后置于0.2mL联排管磁体上3分钟。
- [0498] 8. 在步骤5中的联排管仍在磁体上时,用户小心地将155 μ L上清液取出并转移到新的0.2mL联排管中并进行移液管混合10次,而不干扰磁珠。
- [0499] 9. 将15 μ L (0.70x) SPRI Select珠移液至含有上清液的新联排管中。
- [0500] 10. 将悬浮液在室温孵育5分钟,并且然后将新管置于0.2mL管磁体上5分钟。
- [0501] 11. 在磁体上时,用户小心地取出并适当地仅弃去上清液,而不干扰SPRI Select珠。
- [0502] 12. 将管保持在磁体上,将200 μ L新鲜的80%乙醇温和移液。
- [0503] 13. 将样品在磁体上孵育30秒。
- [0504] 14. 在磁体上时,用户小心地取出并适当地仅弃去上清液,而不干扰SPRI Select珠。
- [0505] 15. 重复80%乙醇洗涤,总共进行两次洗涤。
- [0506] 16. 将管保持在磁体上,使用小体积移液管将任何残留的上清液从管去除。
- [0507] 17. 将管在磁体上打开,以在室温干燥SPRI Select珠5分钟。
- [0508] 18. 将30 μ L洗脱缓冲液 (PN 650000075) 移液至管中,并进行脉冲涡流,以完全重悬SPRI Select珠。
- [0509] 19. 将样品在室温孵育2分钟。
- [0510] 20. 将管短暂离心,以收集处于底部的内容物。
- [0511] 21. 将管置于磁体上,直到溶液澄清 (通常 \leq 30秒)。
- [0512] 22. 将全部洗出物 (\sim 30 μ L) 移液至新的1.5mL LoBind管中。WTA索引PCR洗出液是最终的测序文库。在一些情况下,这一步是停止点。索引PCR文库可以在-20 $^{\circ}$ C储存 \leq 6个月,直到测序。
- [0513] 23. 用户使用Qubit dsDNA HS测定用QubitTM荧光计和使用Agilent高灵敏度DNA试剂盒用Agilent 2100生物分析仪对索引PCR文库进行定量和进行质量控制。
- [0514] 测序
- [0515] 在一些情况下,测序以50,000个读段/细胞和加载到测序运行中的20%PhiX载量来进行。
- [0516] 实施例1
- [0517] 用于进行全转录组分析 (WTA) 的基于随机引发和延伸 (RPE) 的方法的开发和改进
- [0518] 部分地基于表10中提供的非限制性示例性WTA性能度量以及当前可得的WTA方法的比较,开发并优化了本文提供的用于进行全转录组分析 (WTA) 的基于随机引发和延伸 (RPE) 的方法。例如,考虑了以下 (1) 和 (2) 来开发基于RPE的WTA方法: (1) 常见的WTA度量 (例如每个细胞的基因/UMI), 尽管由于不同的WTA方法具有来自不同基因类型的固有偏倚,这些度量可能有偏倚;和 (2) 另外的度量,诸如例如,跨越不同表达水平基因 (核糖体基因、线粒体基因、管家基因等) 和偏倚的WTA丢失率 (dropout rates)。在一些情况下,主要免疫细胞类型 (B细胞、NK、T细胞和髓细胞) 的鉴定使用BD数据视图上的tSNE和特征选择默认设置 (Feature Selection default settings) 完成。
- [0519] 表10. WTA性能度量

[0520]	性能度量	标准
	偏倚	核糖体%读段/样品
		线粒体%读段/样品
	灵敏度	流水线完成后, 至少 8000 个读段/细胞的中值基因/细胞
		非核糖体基因、非线粒体基因、非假基因的中值基因/细胞
	WTA 丢失率	管家基因阳性的%总细胞
		CD4 阳性的%T 细胞
		CD8A 阳性的%T 细胞
		TRBC2 阳性的%T 细胞
		TRAC 阳性的%T 细胞
		CD3D 阳性的%T 细胞
		TBX21 阳性的%T 细胞
		IGLC3 阳性的%B 细胞
[0521]	文库产量和与 Illumina 测序仪的相容性	最终文库 nM
		最终文库大小(bp)

[0522] 基于RPE的WTA方法与基于连接的WTA方法的比较

[0523] 本文提供的使用六聚体 (ID010HK6) 或九聚体 (ID010NK6) 用于进行WTA的基于RPE的方法, 与目前可得的基于连接的WTA方法 (ID010UG1) 在Jurkat/Ramos细胞 (表11) 和PMBC (表12) 中进行比较。图13A-图13B是示出了进行基于连接的全转录组分析和用使用6mer随机多聚体和9mer随机多聚体的随机引发和延伸进行的全转录组分析的非示例性灵敏度度量的图。与基于连接的WTA方法相比, 使用基于RPE的WTA方法在Jurkat/Ramos细胞中检测到更多的基因和分子。此外, 与基于六聚体的方法相比, 使用基于九聚体的RPE方法检测到更多的基因和分子。一些标志物在Jurkat细胞中没有表达。注意到, 基于连接的方法具有更好的IGLC3检测。

[0524] 表11. 比较Jurkat/Ramos细胞中WTA性能

[0525]	标准	RPE 六聚体	RPE 九聚体	基于连接的 WTA 方案
	检测到的总基因	21576	21986	21149
	核糖体%读段/样品	11.02%	11.13%	13.09%
	线粒体%读段/样品	15.14%	15.61%	11.60%
	流水线完成后, 至少 8000 个读段/细胞的中值基因/细胞	2495	2698	2154
	非核糖体基因、非线粒体基因、非假基因的中值基因/细胞	2362	2556	2003
	管家基因阳性的%总细胞	-	-	-
	CD4 阳性的%T 细胞	10.1%	10.1%	7.2%
	CD8A 阳性的%T 细胞	-	-	-
	TRBC2 阳性的%T 细胞	76.6%	81.0%	74.1%
	TRAC 阳性的%T 细胞	43.0%	45.8%	46.3%
	CD3D 阳性的%T 细胞	91.3%	94.3%	93.5%
	TBX21 阳性的%T 细胞	0.5%	0.3%	0.1%
	IGLC3 阳性的%B 细胞	22.1%	65.8%	72.4%

[0526] 表12. 比较PBMC中WTA性能

[0527]	标准	RPE 六聚体	RPE 九聚体	基于连接的 WTA 方案
	检测到的总基因	18419	18999	18958
	核糖体%读段/样品	11.91%	12.05%	27.01%
	线粒体%读段/样品	19.53%	20.27%	5.27%
	流水线完成后, 至少 8000 个读段/细胞的中值基因/细胞	569	651	391
	非核糖体基因、非线粒体基因、非假基因的中值基因/细胞	564	489	276
	管家基因阳性的%总细胞	-	-	-
	CD4 阳性的%T 细胞	12.1%	17.0%	8.3%
	CD8A 阳性的%T 细胞	18.7%	21.1%	9.7%
	TRBC2 阳性的%T 细胞	65.6%	67.4%	36.5%
	TRAC 阳性的%T 细胞	44.8%	51.6%	28.1%
	CD3D 阳性的%T 细胞	16.1%	16.2%	22.1%
	TBX21 阳性的%T 细胞	7.0%	11.2%	5.9%
	IGLC3 阳性的%B 细胞	0.0%	0.7%	8.5%

[0528] 图14是示出了与基于连接的WTA方法相比, 本公开内容的基于随机引发和延伸的WTA分析的示例性性能的t-分布式随机邻域嵌入(t-SNE)。tSNE投影集中于预先定义的基因集中的664个免疫相关基因。ID010HK6 (RPE六聚体) 和ID010NK6 (RPE九聚体) 样品具有相似的基因表达模式, 并且表现为比ID010UG1 (基于连接的WTA方法) 在细胞簇之间具有更高的可变性, 尽管所有样品在更近的2D空间中表现为扁平的。

[0529] 图15A-图15B是示出了ID010NK6 (RPE WTA九聚体) 表现为比ID010UG1 (基于连接的WTA方法) 捕获更多的细胞表面标志物的非限制性示例性饼图。图15A是示出了在RPE WTA九聚体方法或基于连接的WTA方法中表达特定基因的细胞数量较高的饼图。图15B是示出了在

RPE九聚体方法或基于连接的WTA方法中每个表达特定基因的细胞的UMI较高的饼图。此外，如表13和表14中描绘的，基于九聚体RPE的WTA方法比基于连接的细胞WTA方法捕获更多的细胞表面标志物。

[0530] 因此，这些数据证明，与目前可得的基于连接的WTA方法相比，本文提供的用于进行WTA的基于RPE的方法(使用六聚体或非六聚体)跨越多种细胞类型和多种性能度量的性能增强。

[0531] 此外，本文公开的基于RPE的WTA方法可以提供时间更短的WTA方案。对于现有的基于液滴的WTA方法、现有的基于连接的WTA方法，以及基于RPE的WTA方法，从单细胞捕获到cDNA合成完成分别需要85分钟、130分钟和130分钟。对于现有的基于液滴的WTA方法、现有的基于连接的WTA方法，以及基于RPE的WTA方法，从文库制备到QC分别需要375分钟、520分钟和330分钟。因此，本文提供的进行WTA的新型方法有利地缩短了方案时间。

[0532] 表13. 基因本体术语

GO 术语	基因
GO:0009897-质膜外侧	ENOX2, CD8A, ATP6AP2, CD88, CXCR3, TRDC1, IL7R, ADA, ALCAM, IFNG, CD4, IL2RG, CD24...
GO:0098552-膜侧	ENOX2, CD8A, ATP6AP2, CD88, CXCR3, TRDC, IL7R, ADA, BTK, ALCAM, IFNG, IL2RG, CD4...
GO:0009986-细胞表面	ENOX2, CD8A, ATP6AP2, CDBB, CXCR3, TRDC, IL7R, TGFB1, ADA, HMMR, ALCAM, CD44...
GO:0043231-细胞内膜结合细胞器	HCCS, PNMA3, SLC9A7, SLC9A6, RP2, TRMT28, AURKA, HMGN5, USP51, SLC35A2, PRKX...
GO:0005622-细胞内	HCCS, PNMA3, SLC9A7, SLC9A6, RP2, TRMT28, AURKA, HMGN5, SLC35A2, USP51, PRKX...
GO:0005768-内体	PRF1, SLC9A7, SLC9A6, RAB9A, MMGT1, CD88, ATP6AP1, DIAPH2, TLR7, TLR8, RRAGB...
GO:0044459-质膜部分	PLXNA3, ENOX2, CD8A, CD88, ATP6AP2, RP2, TSPAN6, CASK, TSPAN7, CXCR3, TLR7, BTK...
GO:0044424-细胞内部分	HCCS, PNMA3, SLC9A7, SLC9A6, RP2, TRMT28, AURKA, HMGN5, SLC35A2, USP51, PRKX...
GO:0005634-核	PNMA3, AURKA, HMGN5, USP51, SLC35A2, PRKX, BTK, DNASE1L1, PRIM1, MAGED1, ZFP9...
GO:0043227-膜结合细胞器	HCCS, PNMA3, SLC9A7, SLC9A6, RP2, TRMT28, AURKA, HMGN5, SLC35A2, USP51, PRKX...
GO:0005773-液泡	PRF1, SLC9A7, ARSD, SLC9A6, MMGT1, RAB9A, CD88, ATP6AP1, DIAPH2, UBQLN2, TLR7...
GO:0044464-细胞部分	HCCS, SLC9A7, SLC9A6, RP2, USP51, BTK, MAGED1, SYP, CDCA7, CCR10, VMA21, PDHA1...

[0534] 表14. 基因本体术语

GO 术语	计数	%	P 值	基因
GO:0042571-免疫球蛋白复合物, 循环	7	7.142857143	8.75644E-10	IGHG1, IGHG3, IGHD, IGHA1, IGHA2, IGHM, IGLC3

[0536]

GO:0019814-免疫球蛋白复合物	7	7.142857143	1.97463E-09	IGHG1, IGHG3, IGHD, IGHA1, IGHA2, IGHM, IGLC3
GO:0009897-质膜外侧	11	11.2244898	2.41164E-07	IGHG1, SELP, IGHG3, IL2RA, IGHD, IGHA1, IGHA2, IGHM, ENG, LAG3, IGLC3
GO:0098552-膜侧	12	12.24489796	7.5028E-06	IGHG1, SELP, IGHG3, IL2RA, IGHD, IGHA1, IGHA2, CDH1, IGHM, ENG, LAG3, IGLC3
GO:0072562-血液微粒	8	8.163265306	8.2479E-06	IGHG 1, IGHG3, IGHD, IGHA 1, IGHA2, IGHM, ENG, IGLC3
GO:0005615-细胞外间隙	21	21.42857143	9.56996E-06	MUC1, IGHG1, SELP, IGHG3, VPRE83, LGALS1, PRDX4, IGHM, RPL39, TIMP1, AZU1, CFP...
GO:0032991-大分子复合物	41	41.83673469	6.22817E-05	IGHG1, IGHG3, RPS26P11, ERBB2, TIMM178, COX78, BEX1, CDH1, IGHM, RPL39, RPA4...
GO:0043234-蛋白复合物	36	36.73469388	0.000127154	IGHG1, IGHG3, ER882, TIMM17B, COX78, BEX1, CDH1, IGHM, RPA4I, TIMP1, GATA1, NAA10...
GO:0009986-细胞表面	12	12.24489796	0.000920806	IGHG1, SELP, IGHG3, IL2RA, LGALS1, IGHD, IGHA1, IGHA2, IGHM, ENG, LAG3, IGLC3
GO:0044421-细胞外区部分	31	31.63265306	0.002477963	IGHG1, IGHG3, VPRE83, PRDX4, CDH1, IGHM, OCRL, RPL39, TIMP1, AZU1, CFP, SERPINE1...
GO:0005576-细胞外区	34	34.69387755	0.005252573	IGHG1, IGHG3, VPRE83, FAM3A, PRDX4, CDH1, IGHM, RPL39, OCRL1, TIMP1, AZU1, CFP...

[0537] 模型系统

[0538] 本文提供的使用六聚体或九聚体用于进行WTA的基于RPE的方法在PBMC和Jurkat/Ramos细胞中进行了测试。

[0539] 图16A-图16D是示出了以下的非限制性示例性生物分析仪图：当采用包括一个随机引发步骤和一个PCR步骤的RPE方法时，Jurkat/Ramos比健康供体外周血单核细胞(PBMC)具有更高的RNA含量(当比较UMI计数/细胞时为~5x)，并且PBMC文库产生165bp的污染物，由于可能对这些峰进行测序，这些污染物可能是随机多聚体中的二聚体。图16A-图16B分别描绘了Jurkat/Ramos和PBMC中的六聚体RPE生物分析仪图，而图16C-图16D分别描绘了Jurkat/Ramos和PBMC中的九聚体RPE生物分析仪图。

[0540] 表15描述了基于六聚体和基于九聚体的RPE WTA方法在Jurkat和PBMC中的WTA性能。一些标志物在Jurkat细胞中没有表达。因为RNA含量低，PBMC是更具挑战性的模型系统。在PBMC中开发方法更注重灵敏度分析。

[0541] 表15. Jurkat细胞和PBMC中基于RPE的WTA性能

[0542]

标准	Jurkat 上的 RPE 六聚体	Jurkat 上的 RPE 九聚体	PBMC 上的 RPE 六聚体	PBMC 上的 RPE 九聚体
检测到的总基因	21576	21986	18419	18999
核糖体%读段/样品	11.02%	11.13%	11.91%	12.05%
线粒体%读段/样品	15.14%	15.61%	19.53%	20.27%
流水线完成后, 至少 8000 个读段/细胞 的中值基因/细胞	2495	2698	569	651
非核糖体基因、非线 粒体基因、非假基因 的中值基因/细胞	2362	2556	564	489
管家基因阳性的%总 细胞	-	-	-	-
CD4 阳性的%T 细胞	10.1%	10.1%	12.1%	17.0%
CD8A 阳性的%T 细胞	-	-	18.7%	21.1%
TRBC2 阳性的%T 细 胞	76.6%	81.0%	65.6%	67.4%
TRAC 阳性的%T 细 胞	43.0%	45.8%	44.8%	51.6%
CD3D 阳性的%T 细胞	91.3%	94.3%	16.1%	16.2%
TBX21 阳性的%T 细 胞	0.5%	0.3%	7.0%	11.2%
IGLC3 阳性的%B 细 胞	22.1%	65.8%	0.0%	0.7%

[0543] 延伸酶

[0544] 研究了延伸酶在本文提供的使用六聚体或九聚体进行WTA的基于RPE的方法中的使用。

[0545] 图17A-图17D是示出了在用随机引发和延伸(采用包括一个随机引发步骤和一个PCR步骤的RPE方法)进行WTA分析的一些实施方案中具有链置换活性的Klenow Exo- 优于T4 DNA聚合酶的非限制性示例性生物分析仪图(RPE PCR)。测试了不同的聚合酶。二次取样的Jurkat/Ramos细胞数为1000。图17A-图17B分别描绘了使用六聚体和九聚体的Klenow Exo-的生物分析仪图,而图17C-图17D分别描绘了使用六聚体和九聚体的T4DNA聚合酶的生物分析仪图。这些图显示出Klenow Exo-具有中等的链置换活性,并且T4 DNA聚合酶没有链置换活性。两种酶之间的核酸外切酶活性也有差异。Klenow Exo-由于能够产生多种随机引发延伸产物,因此六聚体和九聚体两者的产量都更高。在一些实施方案中,Klenow用于延伸。在一些实施方案中,Klenow exo-用于延伸。

[0546] 图18A-图18D是示出了Klenow exo-可能不具有如先前认为的那么多的链置换活性的非限制性示例性生物分析仪图(RPE PCR)。二次取样的PBMC数为1000。在PCR1显示产量改变之前,测试了变性方法的RPE。测试了以下变量:(i)在95℃变性2min,然后立即取出上清液(图18A);(ii)在95℃变性2min,然后等待1min以取出上清液(图18B);(iii)用0.1N NaOH处理5min,然后用Tris Cl pH7中和(图18C);和(iv)不变性(图18D)。图18D显示出低的RPE PCR产量表明大部分产物保持与cDNA珠结合,表明该酶可能具有很少的链置换活性。在一些实施方案中,可以使用其他酶,诸如phi29酶。

[0547] RPE引物长度和随机引发步骤

[0548] 研究了本文提供的使用六聚体或九聚体进行WTA的基于RPE的方法的随机引发步骤的数量和随机多聚体长度。

[0549] 图19A-图19D是示出了随机引发步骤的数量和随机多聚体长度对RPE WTA方法的影响的非限制性示例性生物分析仪图。二次取样的Jurkat/Ramos细胞数为1000。使用六聚体和九聚体(分别为图19A和图19B)测试了包括双重随机引发(进行两次随机引发和延伸以及两个PCR循环)的基于RPE的WTA方案。此外,使用六聚体和九聚体(分别为图19C和图19D)测试了包括单个随机引发(进行一个随机引发和延伸以及一个PCR循环)的基于RPE的WTA方案。使用单次随机引发,可以在索引PCR后生成如图所示的可测序文库。观察到165bp的污染物。表16描绘了使用六聚体和九聚体的上文的单次随机引发和双重随机引发RPE变化形式的WTA性能。虽然单个随机引发步骤要快得多,但2个随机引发步骤表现为具有更高的灵敏度。九聚体表现为对更多的基因有更好的灵敏度。本文提供的其余实验采用了1个随机引发步骤。

[0550] 表16. 使用1个或2个随机引发步骤以及六聚体或非聚体的RPE WTA性能

[0551]

标准	六聚体, 2 个随机引发 步骤	六聚体, 1 个随机引发 步骤	九聚体, 2 个随机引发 步骤	九聚体, 1 个随机引发 步骤
检测到的总基因	17327	17380	17984	17474
核糖体%读段/样品	10.26%	11.48%	10.16%	11.30%
线粒体%读段/样品	14.15%	15.12	14.45%	16.53%
流水线完成后, 至少 8000 个读段/细胞 的中值基因/细胞	2579	2069	2637	1897
非核糖体基因、非线 粒体基因、非假基因 的中值基因/细胞	2458	1951	2503	1779
管家基因阳性的%总 细胞	-	-	-	
CD4 阳性的%T 细胞	5.4%	6.3%	3.0%	3.9%
CD8A 阳性的%T 细胞	-	-	-	-
TRBC2 阳性的%T 细 胞	81.3%	75.3%	86.7%	78.3%
TRAC 阳性的%T 细 胞	40.1%	28.9%	42.1%	28.1%
CD3D 阳性的%T 细胞	97.6%	83.7%	97.5%	78.3%
TBX21 阳性的%T 细 胞	0.6%	0.0%	0.6%	0.0%
IGLC3 阳性的%B 细 胞	18.6%	16.3%	55.9%	41.9%

[0552] RPE清理

[0553] 研究了本文提供的用于进行WTA的基于RPE的方法的最佳SPRI清理比率。

[0554] 图20A-图20F是示出了SPRI清理比率对基于RPE的WTA方法的性能的影响的非限制性示例性数据。图20A-图20B是比较1.6x(图20A)和1.0x(图20B)SPRI清理比率的非限制性

示例性生物分析仪图。此外,针对1.6x和1.0x的SPRI清理比率,比较了读段/细胞(图20C)、中值基因/细胞(图20D)、中值分子/细胞(图20E)和%核糖体和线粒体分子(图20F)。这些数据证明,1.6x RPE清理比率比1.0x清理比率产生更高的产量,并且1.6x清理比率尽管具有更少的读段/细胞,也比1.0x清理比率产生更多的检测到的分子和基因/细胞。

[0555] 随机引物浓度和RPE条件

[0556] 研究了用于本文提供的用于进行WTA的基于RPE的方法的不同随机引物浓度和RPE条件。

[0557] 图21A-图21G是示出了不同的随机引物浓度和RPE条件对基于RPE的WTA方法的影响的非限制性示例性生物分析仪图(RPE PCR,SPRI清理后)。具体而言,实验比较了以下条件:ID019-1(图21A,10uM随机引物;未添加冷珠,6.186nM)、ID019-2(图21B,10uM随机引物,添加冷珠,2.138nM)、ID019-3(图21C,5uM随机引物,添加冷珠,0.538nM)、ID019-4(图21D,1uM随机引物,添加冷珠,0.314nM)、ID019-5(图21E,0.5uM随机引物,添加冷珠,0.438nM)、ID019-6(图21F,1uM随机引物,添加冷珠,0.652nM)和ID019-7(图21G,0.5uM随机引物,添加冷珠,0.242nM)。比较图21A和图21B,观察到添加冷珠降低了RPE PCR期望产物的产量。虽然ID017(如下文描述的)具有相反的PCR产量结果,但这可能是由于PCR条件不同,并且在ID019中RPE清理条件被修改为1侧清理。

[0558] 图22A-图22F是示出了添加冷珠和改变珠密度对RPE WTA的影响的非限制性示例性测序数据。如图22A-图22C(ID019)所示,添加冷珠降低了RPE WTA测定的灵敏度。测定了基因/细胞的中值数(图22A,非核糖体基因、非线粒体基因、非假基因)、平均读段/细胞(图22B)和平均分子/细胞(图22C)。此外,如图22D-图22F(ID017)所示,RPE中较高的珠密度降低了RPE WTA测定的灵敏度。测定了中值分子/细胞(图22D)、读段/细胞(图22E)和中值基因/细胞(图22F)。

[0559] 图23A-图23B是示出了随机引物浓度和冷珠对RPE WTA的影响的非限制性示例性测序数据。图23A示出中值基因/细胞(ID017,不具有冷珠),并且图23B示出具有冷珠时的基因/细胞的中值数(非核糖体基因、非线粒体基因、非假基因)(ID019)。在ID017和ID019单细胞实验两者中,10uM随机引物浓度始终产生最高的灵敏度。表17提供了现有的WTA方法与采用不同随机引物浓度和珠密度的RPE WTA方法的WTA性能度量的比较。

[0560] 表17. 使用不同随机引物浓度和珠密度的RPE WTA性能

标准	现有 WTA 方案平均度量	RPE WTA 度量(随机引物和珠浓度)				
		10 uM 引物 (600 个珠/ ul)	10 uM 引物 (150 个珠/ ul)	5 uM 引物 (150 个珠/ ul)	1 uM 引物 (150 个珠/ ul)	0.5 uM 引物 (150 个珠/ ul)
检测到的总基因	14,757-19,919 (平均值 16,485)	16,766	19,109	18,584	16,521	15,454
核糖体 % 读段/样品	27.38%-44.79% (平均值 37%)	17.95	18.76	17.89	15.3	13.71
线粒体 % 读段/样品	2.24%-4.15% (平均值 3%)	22.26	19.62	20.93	23.49	24.76
管家基因阳性的 % 总细胞	65.06%-81.76% (平均值 75%)	57.84	66.88	59.76	36.83	30.61
CD4 阳性的 % T 细胞	1.9%-10% (平均值 6%)	23.1	27	20.5	12.6	7.7
CD8A 阳性的 % T 细胞	7.5%-19.7% (平均值 14%)	15.1	18.5	14.6	8.7	4.4
TRBC2 阳性的 % T 细胞	44.5%-75.4% (平均值 59%)	67.6	72.8	67.7	46.8	43.4
TRAC 阳性的 % T 细胞	51.1%-85.9% (平均值 71%)	63.5	67.9	65.9	51.9	36
CD3D 阳性的 % T 细胞	65.7%-84.5% (平均值 75%)	27.8	38.7	31.1	15.8	8.2
TBX21 阳性的 % T 细胞	1.5%-5.9% (平均值 4%)	3.8	5.8	3	2.9	2.5
IGLC3 阳性的 % B 细胞	26.1%-43.9% (平均值 35%)	5.5	11.9	8.8	3.9	5.3

[0562] 图24A-图24E是示出了不同的细胞数对RPE WTA方法的影响的非限制性示例性生物分析仪图(RPE PCR)。具体而言,实验比较了以下条件:ID018-1(图24A,7970个细胞,14个循环,2.62ng/uL)、ID018-2(图24B,872个细胞,18个循环,1.14ng/uL)、ID018-3(图24C,75个细胞,18个循环,0.774ng/uL)、ID018-4(图24D,872个二次取样的细胞并添加冷珠,18个循环,0.892ng/uL)和ID018-5(图24E,75个二次取样的细胞并添加冷珠,18个循环,0.572ng/uL)。箭头表示宽峰,可能是过度扩增。图25A-图25E是示出了不同的细胞数对RPE WTA方法的影响的非限制性示例性生物分析仪图(索引PCR)。具体而言,实验比较了以下条件:ID018-1(图25A,7970个细胞,6.36ng/uL)、ID018-2(图25B,872个细胞,1.14ng/uL)、ID018-3(图25C,75个细胞,0.774ng/uL)、ID018-4(图25D,872个二次取样的细胞并添加冷珠,0.892ng/uL)和ID018-5(图25E,75个二次取样的细胞并添加冷珠,0.572ng/uL)。发现约100个细胞的细胞文库在一些情况下不具有良好的质量。表18示出了上文提及的条件的WTA

性能度量。发现当细胞与珠的比率低时,文库质量有降低的趋势。较高的珠浓度可能降低测定的灵敏度,并且用户可能希望增加RPE体积。当细胞:珠的比率低时,文库质量倾向于更差。发现10uM的随机引物浓度具有最高的灵敏度,并且在一些情况下可能期望甚至更高的浓度。

[0563] 表18. 使用不同的细胞数的RPE WTA性能

[0564]

	#盒中捕 获的细 胞	推定的 #细胞	#读段/细 胞	#分配给 细胞标记 的细胞读 段	%独特地 分配给转 录组的细 胞读段	中值分 子 / 细 胞	中值基 因 / 细 胞
ID018-1	7,970	6597	10657	74.22	39.74	965	531
ID018-2	872	789	5917	28.2	9.26	1114	572
ID018-3	75	72	2947.71	16.75	2	896	441.5
ID018-4	872	640	11820	42.55	18.74	819.5	450.5
ID018-5	75	57	4622.32	18.14	2.5	487	297

[0565] 图26A-图26F描绘了示出本公开内容的基于随机引发和延伸的WTA分析使用不同的细胞数的示例性性能的t-分布式随机邻域嵌入(t-SNE)投影。图26B描绘了ID018-1(10,000个细胞),图26C描绘了ID018-2(1,000个细胞),图26D描绘了ID018-3(100个细胞),图26E描绘了ID018-4(具有二次采样和添加了冷珠的1,000个细胞),图26F描绘了ID018-5(具有二次采样和添加了冷珠的100个细胞),以及图26A描绘了所有以上提及的条件叠加。观察到很少的批次效应。

[0566] 图27A-图27D是示出连续RPE变性和RPE反应体积对RPE WTA方法的影响的非限制性示例性生物分析仪图。将图27A和图27C中示出的单次变性方案(包括在RPE反应混合物中以95℃进行变性)与图27B和图27D中示出的连续变性方案(还包括在珠重悬缓冲液中以95℃进行第二次变性)进行比较。此外,比较了使用200uL RPE体积(图27A-图27B)和使用1mL RPE体积(图27C-图27D)。实验在以下条件下进行:图27A(12971个细胞;13个循环;4.64ng/uL);图27B(12971个细胞;17个循环;68.07ng/uL),图27C(12522个细胞;13个循环;5.44ng/uL);和图27D(12522个细胞;17个循环;49.30ng/uL)。在1mL和200uL的RPE体积两者,发现第二次变性产生大致相同量的PCR材料。

[0567] PCR扩增的数量

[0568] 图28A-图28C是示出了本公开内容的基于随机引发和延伸的WTA分析使用2PCR文库生成方案(图28A-图28B)和1PCR文库生成方案(图28C)的示例性性能的示例性生物分析仪图。图28A描绘了2PCR文库生成方案的PCR1(例如RPE PCR),并且图28B描绘了2PCR文库生成方案的PCR2(例如索引PCR),而图28C描绘了单个PCR步骤的结果。在2PCR文库生成方案的第一个PCR(RPE PCR)中,使用了读段1&2引物。在1PCR文库生成方案的第一个PCR中,使用了索引P5/P7引物。使用索引引物产生了高的165bp峰,16个循环后<1ng/uL(在30uL中洗脱)。在PCR1中使用索引引物(1PCR文库生成方案)导致大量的165bp峰(由图28C中的箭头指示)。该峰在读段1/读段2样品中不太突出(图28A)(显示为71bp)。因此,在PCR1中使用R1/R2引物产生较少的165bp峰,并且在PCR1中产生略少的产物,但是用户可以在第二PCR中产生更多的产量。因此,2PCR文库生成方案具有较少的165bp污染物峰和更多的总文库产量。因此,基于2个PCR的RPE WTA方法可以允许产生更洁净的文库。

[0569] 实施例2

[0570] 用于与蛋白质结合试剂关联的寡核苷酸

[0571] 本实施例展示了可以与蛋白质结合试剂缀合的寡核苷酸的设计。寡核苷酸可用于同时确定蛋白质表达和基因表达。寡核苷酸也可用于样品索引化,以确定相同或不同样品的细胞。

[0572] 95mer寡核苷酸设计

[0573] 以下方法用于产生候选寡核苷酸序列和相应的引物序列,所述候选寡核苷酸序列和相应的引物序列用于蛋白质表达和基因表达的同时确定或样品索引化。

[0574] 1.序列生成和消除

[0575] 以下方法用于产生候选寡核苷酸序列,所述候选寡核苷酸序列用于蛋白质表达和基因表达的同时确定或样品索引化。

[0576] 步骤1a.随机产生许多具有期望长度(45bp)的候选序列(50000种序列)。

[0577] 步骤1b.将转录调节物LSRR序列附加到所产生的序列的5'末端,并将多(A)序列(25bp)附加到所产生的序列的3'末端。

[0578] 步骤1c.将不具有40%至50%范围中的GC含量的所产生和附加的序列去除。

[0579] 步骤1d.将各自具有一个或多个发夹结构的剩余序列去除。

[0580] 剩余的候选寡核苷酸序列的数量是423。

[0581] 2.引物设计

[0582] 以下方法用于设计用于剩余的423种候选寡核苷酸序列的引物。

[0583] 2.1 N1引物:使用通用N1序列:5'-GTTGTCAAGATGCTACCGTTCAGAG-3'(LSRR序列;SEQ ID NO.4)作为N1引物。

[0584] 2.2 N2引物(用于扩增特异性样品索引寡核苷酸;例如,图29B-图29D中的N2引物):

[0585] 2.2a.将不从N1序列下游开始的候选N2引物去除。

[0586] 2.2b.将候选寡核苷酸序列最后35bp中重叠的候选N2引物去除。

[0587] 2.2c.将与使用寡核苷酸进行研究的细胞的物种的转录组(例如,人类转录组或小鼠转录组)对齐的候选引物去除。

[0588] 2.2d.使用ILR2序列作为默认对照(ACACGACGCTCTCCGATCT;SEQ ID NO.5)以最小化或避免引物-引物相互作用。

[0589] 在423种候选寡核苷酸序列中,设计了用于390种候选物的N2引物。

[0590] 3.过滤

[0591] 以下方法用于过滤剩余的390种候选引物序列。

[0592] 3a.消除具有以A结尾的随机序列(即多(A)序列的有效长度大于25bp)的任何候选寡核苷酸序列,以保持多(A)尾对于所有条形码而言长度相同。

[0593] 3b.消除具有4个或更多个连续G(>3G)的任何候选寡核苷酸序列,这是由于G的运行的寡核苷酸合成中的额外成本和潜在的较低产量。

[0594] 图29A示出了使用上文的方法产生的非限制性示例性候选寡核苷酸序列。

[0595] 200mer寡核苷酸设计

[0596] 以下方法用于产生候选寡核苷酸序列和相应的引物序列,所述候选寡核苷酸序列和相应的引物序列用于蛋白质表达和基因表达的同时确定以及样品索引化。

[0597] 1. 序列生成和消除

[0598] 以下用于产生候选寡核苷酸序列,所述候选寡核苷酸序列用于蛋白质表达和基因表达的同时确定以及样品索引化。

[0599] 1a. 随机产生许多具有期望长度(128bp)的候选序列(100000种序列)。

[0600] 1b. 将转录调节物LSRR序列和一个另外的非人类、非小鼠的锚序列附加到所产生的序列的5'末端,并将多(A)序列(25bp)附加到所产生的序列的3'末端。

[0601] 1c. 将不具有40%至50%范围中的GC含量的所产生和附加的序列去除。

[0602] 1d. 基于发夹结构评分分选剩余的候选寡核苷酸序列。

[0603] 1e. 选择具有最低发夹评分的1000种剩余的候选寡核苷酸序列。

[0604] 2. 引物设计

[0605] 以下方法用于设计用于具有最低发夹评分的400种候选寡核苷酸序列的引物。

[0606] 2.1 N1引物: 使用通用N1序列: 5'-GTTGTCAAGATGCTACCGTTCAGAG-3' (LSRR序列; SEQ ID NO.4) 作为N1引物。

[0607] 2.2 N2引物 (用于扩增特异性样品索引寡核苷酸; 例如, 图29B和图29C中的N2引物):

[0608] 2.2a. 将不从N1序列下游23nt开始的候选N2引物 (锚序列是跨越所有候选寡核苷酸序列通用的) 去除。

[0609] 2.2b. 将靶序列最后100bp中重叠的候选N2引物去除。所得的候选引物可以在靶序列的第48个核苷酸和第100个核苷酸之间。

[0610] 2.2c. 将与使用寡核苷酸进行研究的细胞的物种的转录组 (例如, 人类转录组或小鼠转录组) 对齐的候选引物去除。

[0611] 2.2d. 使用ILR2序列5'-ACACGACGCTCTCCGATCT-3' (SEQ ID NO.5) 作为默认对照, 以最小化或避免引物-引物相互作用。

[0612] 2.2e. 将靶序列最后100bp中重叠的候选N2引物去除。

[0613] 在400种候选寡核苷酸序列中, 设计了用于392种候选寡核苷酸的N2引物。

[0614] 3. 过滤

[0615] 以下用于过滤剩余的392种候选引物序列。

[0616] 3a. 消除具有以A结尾的随机序列 (即多(A)序列的有效长度大于25bp) 的任何候选寡核苷酸序列, 以保持多(A)尾对于所有条形码而言长度相同。

[0617] 3b. 消除具有4个或更多个连续G(>3G) 的任何候选寡核苷酸序列, 这是由于G运行的寡核苷酸合成中的额外成本和潜在的较低产量。

[0618] 图29B示出了使用上文的方法产生的非限制性示例性候选寡核苷酸序列。图29B中示出的巢式N2引物可以结合抗体或样品特异性序列用于靶向扩增。图29C示出了具有对应于用于靶向扩增的锚序列的巢式通用N2引物的相同非限制性示例性候选寡核苷酸序列。图29D示出了用于一步靶向扩增的具有N2引物的相同非限制性示例性候选寡核苷酸序列。

[0619] 总之, 这些数据表明, 不同长度的寡核苷酸序列可以被设计用于蛋白质表达和基因表达的同时确定或样品索引化。寡核苷酸序列可以包括通用引物序列、抗体特异性寡核苷酸序列或样品索引化序列以及多(A)序列。

[0620] 实施例3

[0621] 寡核苷酸关联的抗体的工作流程

[0622] 本实施例展示了使用寡核苷酸缀合的抗体来确定蛋白质靶的表达谱的工作流程。

[0623] 将受试者的冷冻细胞(例如,冷冻外周血单核细胞(PBMC))解冻。将解冻的细胞用寡核苷酸缀合的抗体(例如0.06 μ g/100 μ l的抗CD4抗体(寡核苷酸缀合的抗体储备物的1:333稀释))在一定温度染色持续一段时间(例如在室温持续20分钟)。将寡核苷酸缀合的抗体与1种、2种或3种寡核苷酸(“抗体寡核苷酸”)缀合。抗体寡核苷酸的序列在图30中示出。洗涤细胞以去除未结合的寡核苷酸缀合的抗体。细胞任选地用钙黄绿素AM(BD(Franklin Lake, New Jersey))和Draq7TM(Abcam(Cambridge, United Kingdom))染色,用于使用流式细胞术分选以获得感兴趣的细胞(例如活细胞)。任选地洗涤细胞以去除过量的钙黄绿素AM和Draq7TM。被钙黄绿素AM染色的单细胞(活细胞)和不被Draq7TM染色的单细胞(未死亡或透化的细胞)使用流式细胞术分选到BD RhapsodyTM盒中。

[0624] 在含有单细胞和珠的孔中,孔中的单细胞(例如3500个活细胞)在裂解缓冲液(例如具有5mM DTT的裂解缓冲液)中裂解。靶(例如CD4)的mRNA表达谱使用BD RhapsodyTM珠确定。使用BD RhapsodyTM珠和抗体寡核苷酸确定靶(例如CD4)的蛋白质表达谱。简而言之,mRNA分子在细胞裂解后被释放。RhapsodyTM珠与条形码(例如随机条形码)关联,每种条形码包含分子标记、细胞标记和寡聚(dT)区域。从裂解的细胞释放的mRNA分子的多(A)区与随机条形码的多(T)区杂交。抗体寡核苷酸的多(dA)区与条形码的寡聚(dT)区杂交。使用条形码对mRNA分子进行逆转录。使用条形码复制抗体寡核苷酸。逆转录和复制任选地同时发生在一个样品等分试样中。

[0625] 逆转录的产物和复制的产物使用引物来进行PCR扩增,用于使用N1引物确定感兴趣的基因的mRNA表达谱,以及使用抗体寡核苷酸N1引物确定靶的蛋白质表达谱。例如,逆转录的产物和复制的产物可以在60度退火温度使用引物进行15个循环的PCR扩增,用于使用血液小组N1引物确定488个血液小组基因的mRNA表达谱,以及使用抗体寡核苷酸N1引物(“PCR 1”)确定CD4蛋白的表达谱。多余的条形码任选地通过Ampure清理来去除。来自PCR 1的产物任选地分成两个等份试样,一个等份试样用于使用用于感兴趣的基因的N2引物确定感兴趣的基因的mRNA表达谱,并且一个等份试样用于使用抗体寡核苷酸N2引物(“PCR 2”)确定感兴趣的靶的蛋白质表达谱。两个等分试样进行PCR扩增(例如,在60度退火温度进行15个循环)。基于如图30中图示的抗体寡核苷酸(“PCR 2”),确定细胞中靶的蛋白质表达。在测序衔接子添加(“PCR 3”),诸如测序衔接子连接后获得和分析测序数据。基于感兴趣的基因的mRNA表达谱确定细胞类型。

[0626] 总之,本实施例描述了使用寡核苷酸缀合的抗体用于确定感兴趣的靶的蛋白质表达谱。本实施例还描述了可以同时确定感兴趣的靶的蛋白质表达谱和感兴趣的基因的mRNA表达谱。

[0627] 在至少一些先前描述的实施方案中,在一种实施方案中使用的一个或更多个要素可以互换地用于另一种实施方案中,除非这样的替换在技术上不可行。本领域技术人员将理解,在不脱离所要求保护的的主题的范围的情况下,可以对上文描述的方法和结构进行各种其他的省略、添加和修改。所有这样的修改和改变都意在落入由所附权利要求书限定的主题的范围内。

[0628] 关于本文中使用的基本上任何复数和/或单数术语,在对于背景和/或应用适当的情

况下,本领域技术人员可以从复数转换为单数和/或从单数转换为复数。为了清楚起见,可以在本文明确阐述各种单数/复数排列。如本说明书和所附权利要求书中使用的,除非上下文另有清楚指示,否则单数形式“a(一)”、“an(一)”和“该(the)”包括复数的提示物。除非另外说明,否则在本文中对“或”的任何提及意在涵盖“和/或”。

[0629] 本领域技术人员将理解,一般来说,本文使用的术语,并且尤其是所附权利要求(例如,所附权利要求的主体)中的术语,通常意在作为“开放式”术语(例如,术语“包括(including)”应解释为“包括但不限于(including but not limited to)”,术语“具有(having)”应解释为“具有至少(having at least)”,术语“包括(includes)”应解释为“包括但不限于(includes but is not limited to)”,等等)。本领域技术人员将进一步理解,如果所引入的权利要求陈述的特定数目是所预期,这样的预期将明确地陈述于权利要求中,并且在不存在这样的陈述的情况下,不存在这样的预期。例如,作为对理解的帮助,以下所附权利要求可以包含前置词“至少一个/至少一种(at least one)”和“一个或更多个/一种或更多种(one or more)”的使用,以引入权利要求陈述。然而,这样的短语的使用不应理解为暗含通过不定冠词“一个”或“一种”(“a” or “an”)引入权利要求陈述会将包含这样的引入的权利要求陈述的任何具体权利要求限制到包含仅一个这样的陈述的实施方案中,甚至在相同的权利要求包括前置词“一个或更多个/一种或更多种”或“至少一个/至少一种”以及不定冠词诸如“一个”或“一种”时也是如此(例如,“一个”和/或“一种”应解释为意指“至少一个/至少一种”或“一个或更多个/一种或更多种”);这对于使用定冠词来引入权利要求陈述同样适用。此外,即使明确地陈述了所引入的权利要求陈述的特定数目,本领域技术人员将认识到,这样的陈述应解释为意指至少所陈述的数目(例如,仅陈述“两个陈述”而没有其他修饰词意指至少两个陈述,或两个或更多个陈述)。此外,在使用类似于“A、B和C等中的至少一个”的惯例的那些情况下,通常这样的句法结构以本领域技术人员将理解该惯例的意义被预期(例如,“具有A、B和C中的至少一个的系统”将包括但不限于仅具有A,仅具有B,仅具有C,A和B一起,A和C一起,B和C一起,和/或A、B和C一起等的系统)。在使用类似于“A、B或C等中的至少一个”的惯例的那些情况下,通常这样的句法结构以本领域技术人员将理解该惯例的意义被预期(例如,“具有A、B或C中的至少一个的系统”将包括但不限于仅具有A,仅具有B,仅具有C,A和B一起,A和C一起,B和C一起,和/或A、B和C一起等的系统)。本领域技术人员将进一步理解,实际上,无论在说明书、权利要求书还是在附图中,呈现两个或更多个替代术语的任何转折性(disjunctive)词语和/或短语应被理解为考虑到包括术语之一、任一术语或两个术语的可能性。例如,短语“A或B”将被理解为包括“A”或“B”或者“A和B”的可能性。

[0630] 此外,当本公开内容的特征或方面以马库什群组(Markush group)描述时,本领域技术人员将意识到本公开内容还由此以马库什群组的任何单独的成员或成员子组描述。

[0631] 如本领域技术人员将理解的,出于任何和所有目的,诸如在提供书面描述方面,本文公开的所有范围还包括任何和所有可能的它的子范围和子范围组合。任何列出的范围都可以很容易地被识别为充分描述并使相同的范围能被分解为至少相等的一半,三分之一,四分之一,五分之一,十分之一等。作为非限制性实例,本文讨论的每个范围可以容易地分解为下三分之一,中三分之一和上三分之一等。如本领域技术人员还将理解的,所有语言,诸如“多达”、“至少”、“大于”、“小于”等包括所陈述的数字,并且指可以随后分解为如上文

讨论的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个单独的成员。因此,例如,具有1-3个物品的组是指具有1个、2个或3个物品的组。类似地,具有1-5个物品的组是指具有1个、2个、3个、4个或5个物品的组,等等。

[0632] 尽管本文已经公开了各种方面和实施方案,但其他方面和实施方案对本领域技术人员将是明显的。本文公开的各种方面和实施方案用于说明的目的而并不意在限制由以下权利要求所指出的实际范围和精神。

序列表

<110> 细胞研究有限责任公司

艾琳·夏姆

<120> 使用随机引发的单细胞全转录组分析

<130> 68EB-298705-W0

<150> 62/757,757

<151> 2018-11-08

<160> 9

<170> PatentIn version 3.5

<210> 1

<211> 32

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<220>

<221> misc_feature

<222> (24) .. (32)

<223> n是a、c、g或t

<400> 1

tcagacgtgt gctcttccga tctnnnnnnn nn 32

<210> 2

<211> 95

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<220>

<221> 5'AmMC6

<222> (1) .. (1)

<223> 5'氨基修饰物C6

<400> 2

gttgtcaaga tgctaccgtt cagagtacgt ggagttggtg gcccgacccc gagcgctacg 60

agccccccgg aaaaaaaaaa aaaaaaaaaa aaaaa 95

<210> 3

<211> 200

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<220>

<221> 5AmMC6

<222> (1) .. (1)

<223> 5'氨基修饰物C6

<400> 3

gttgtcaaga tgctaccgtt cagagctact gtccgaagtt accgtgtatc taccacgggt 60
ggtttttcga atccggaaaa gatagtaata agtgtttttag ttggaataag tcgcaacttt 120
tgtagacggt tacctctcaa tttttctgat ccgtaggccc cccgatctcg gcctcaaaaa 180
aaaaaaaaaa aaaaaaaaaa 200

<210> 4

<211> 25

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<400> 4

gttgtcaaga tgctaccgtt cagag 25

<210> 5

<211> 20

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<400> 5

acacgacgct cttccgatct 20

<210> 6

<211> 95

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<400> 6

gttgtcaaga tgctaccgtt cagagcccca tgtctagtag ctattgggtcc cctatcctca 60
gattcgttta aaaaaaaaaa aaaaaaaaaa aaaaa 95

<210> 7

<211> 26

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<400> 7

ttttttttttt ttttttttttt tttttt 26

<210> 8

<211> 20

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<400> 8

aaaaaaaaaa aaaaaaaaaa 20

<210> 9

<211> 20

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成的寡核苷酸

<400> 9

ttttttttttt ttttttttttt 20

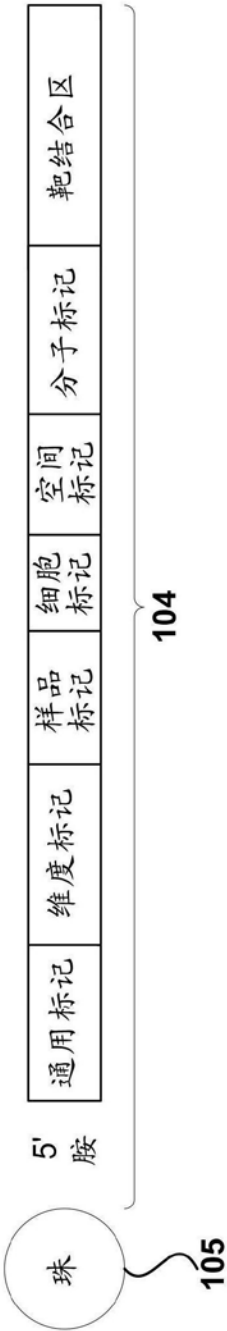


图1

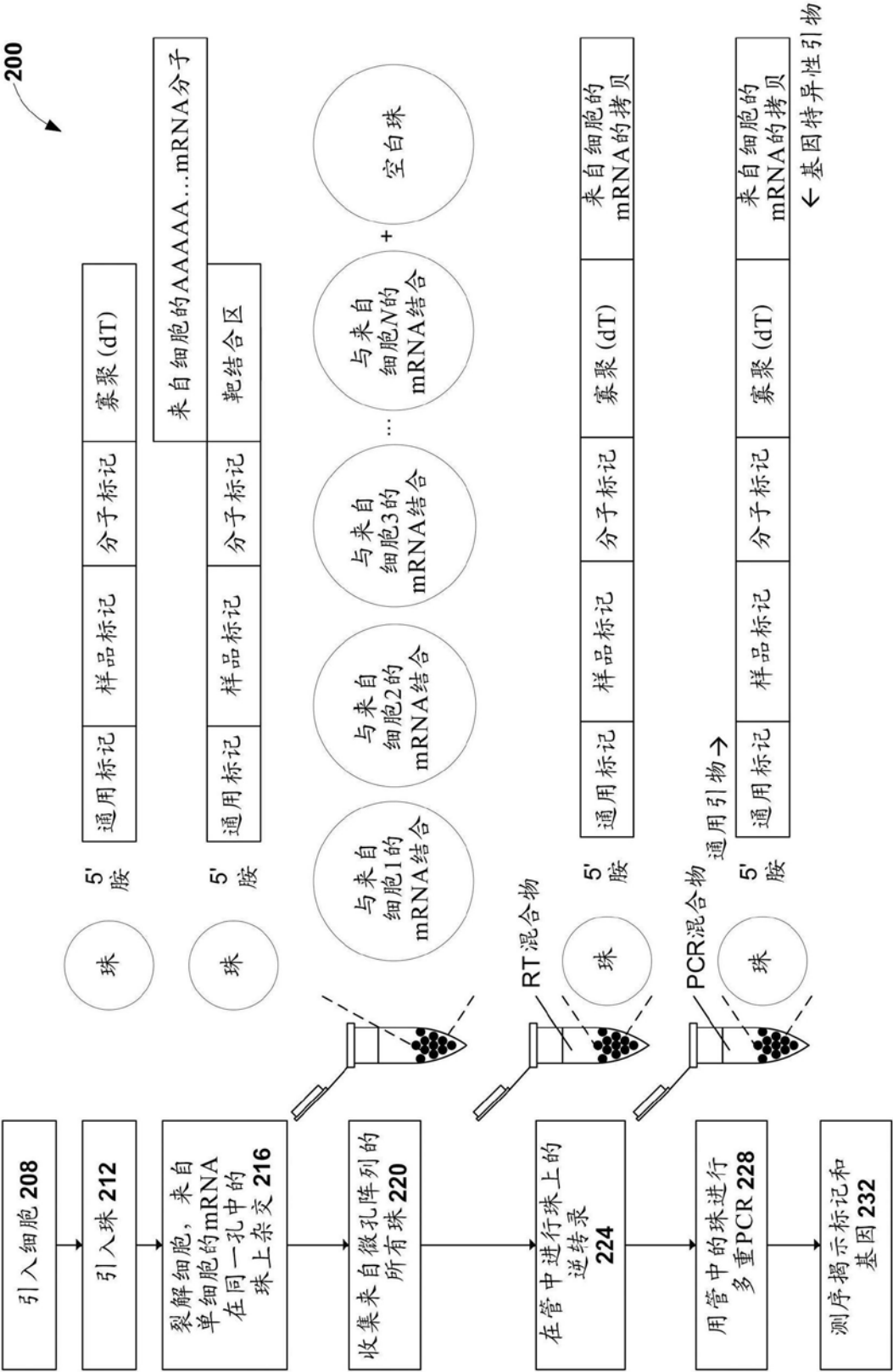


图2

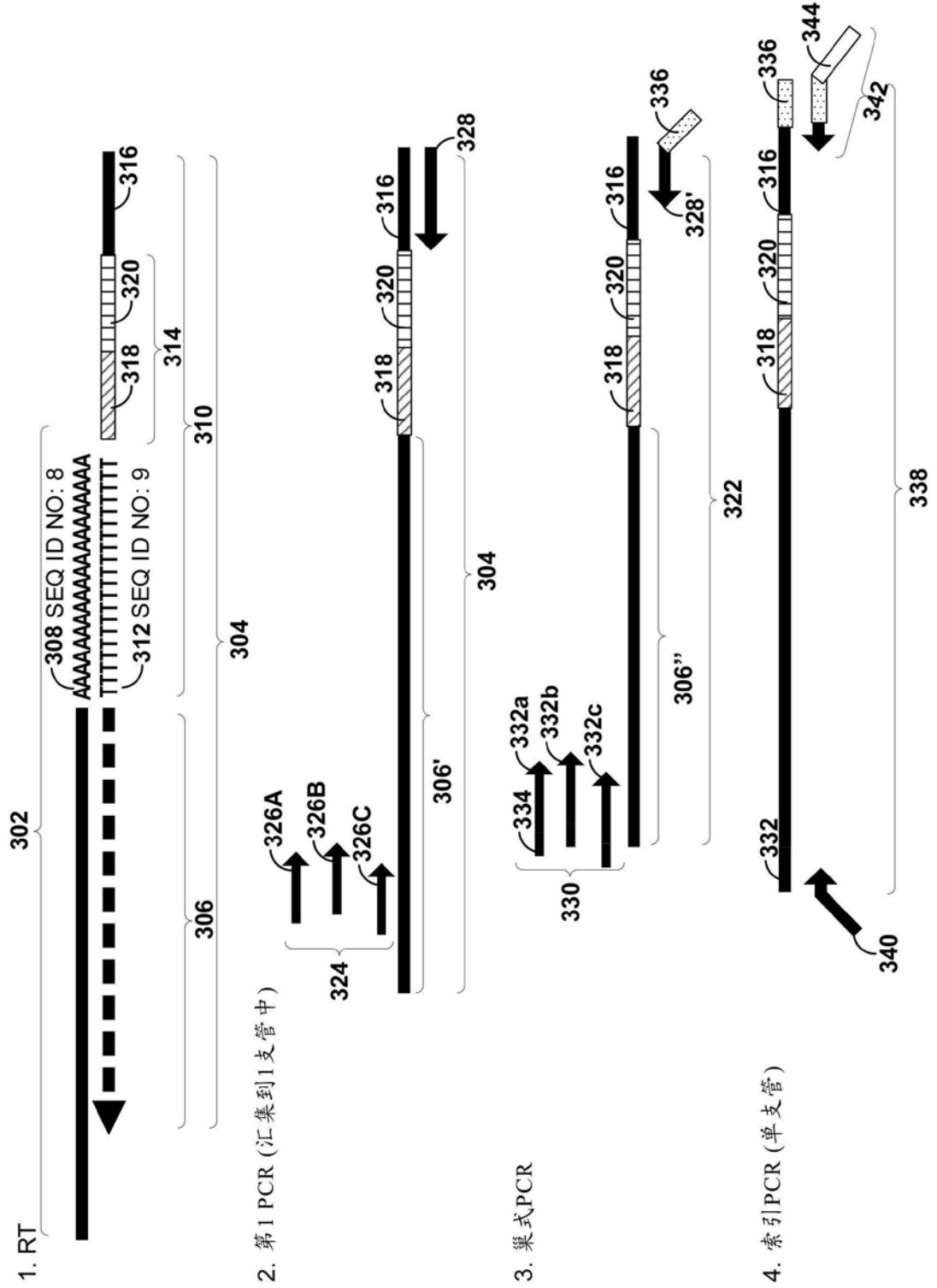


图3

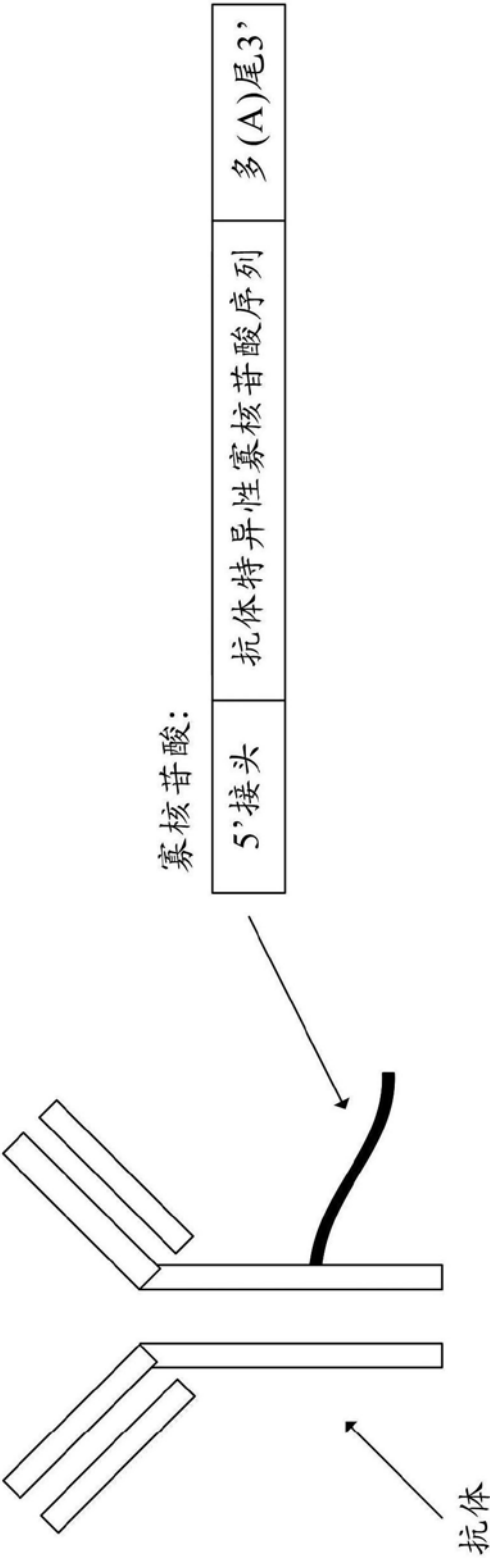


图4

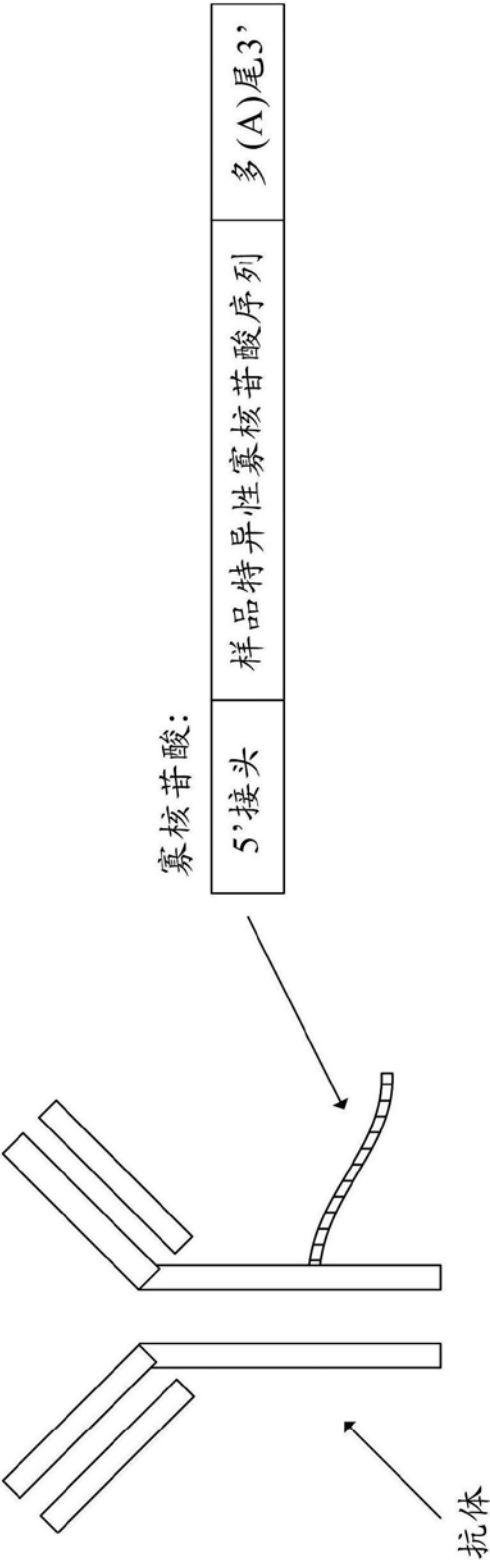


图5

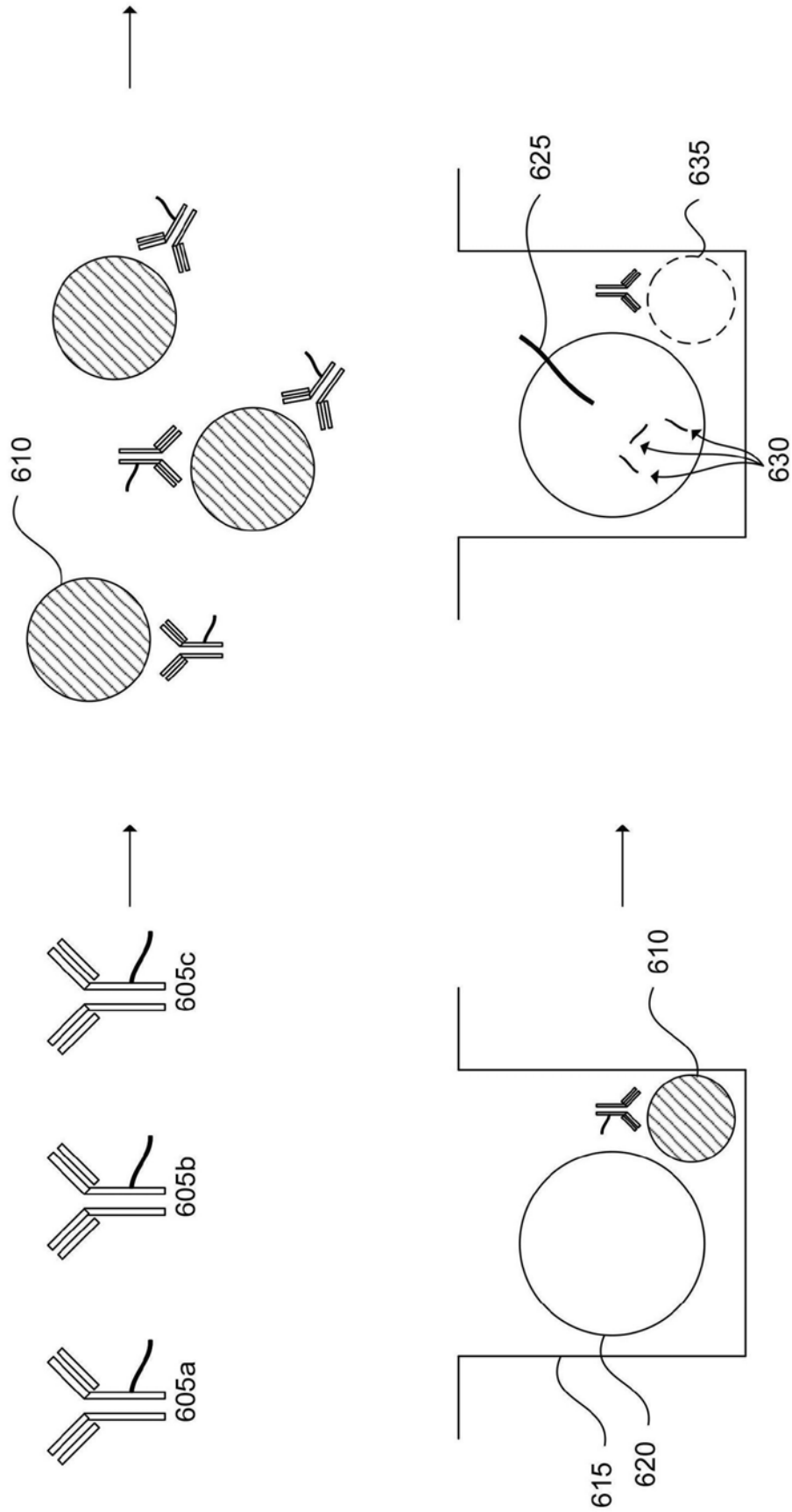


图6

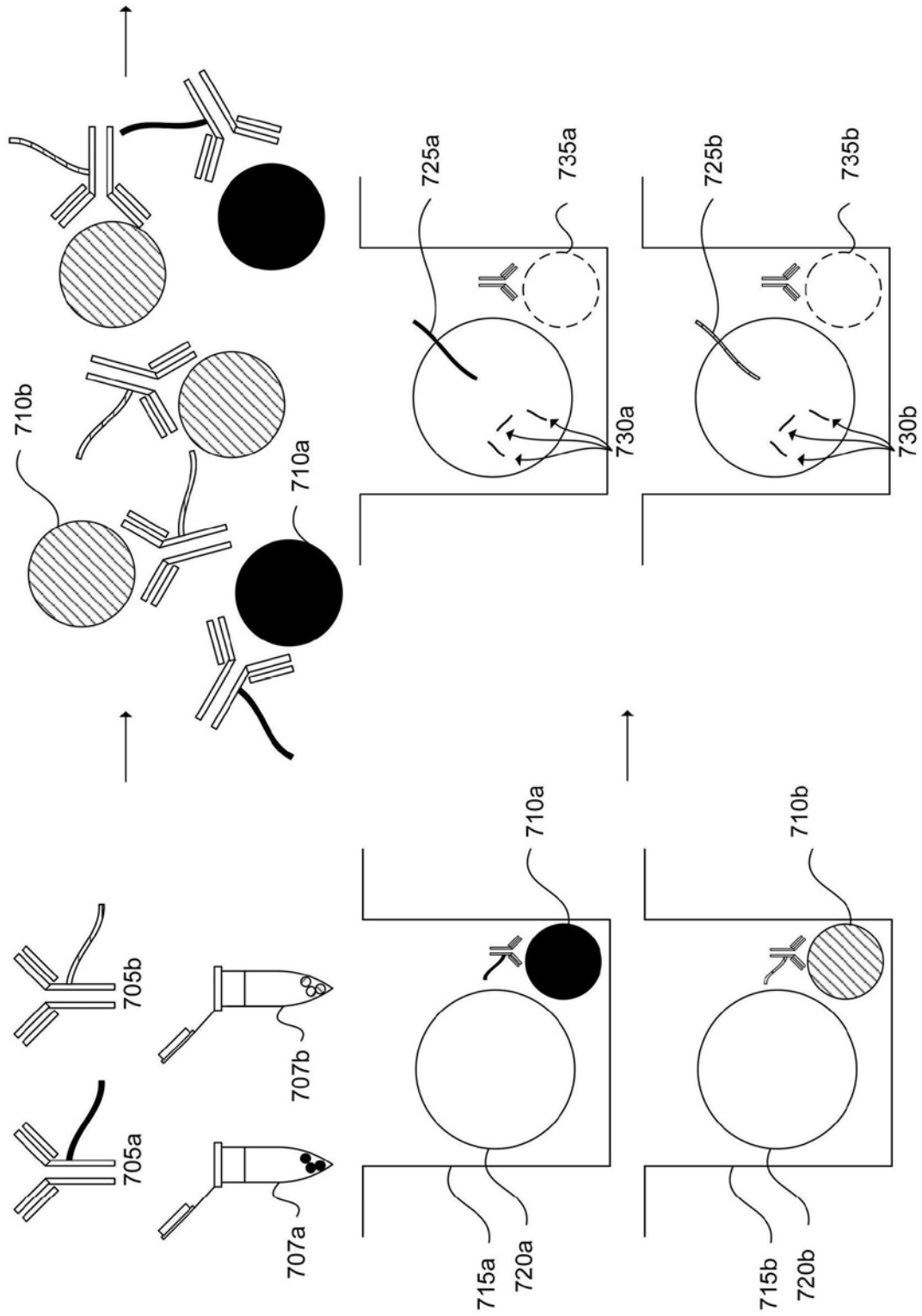


图7

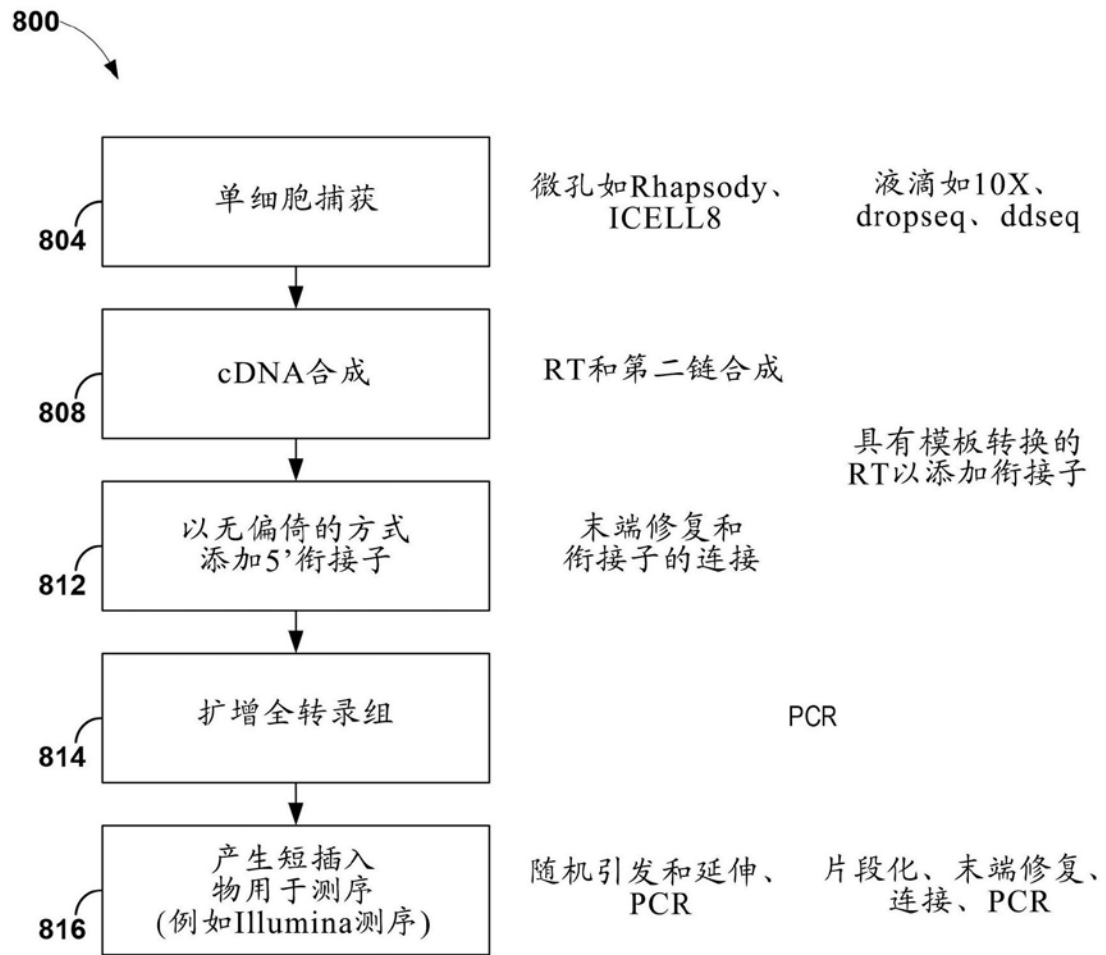


图8

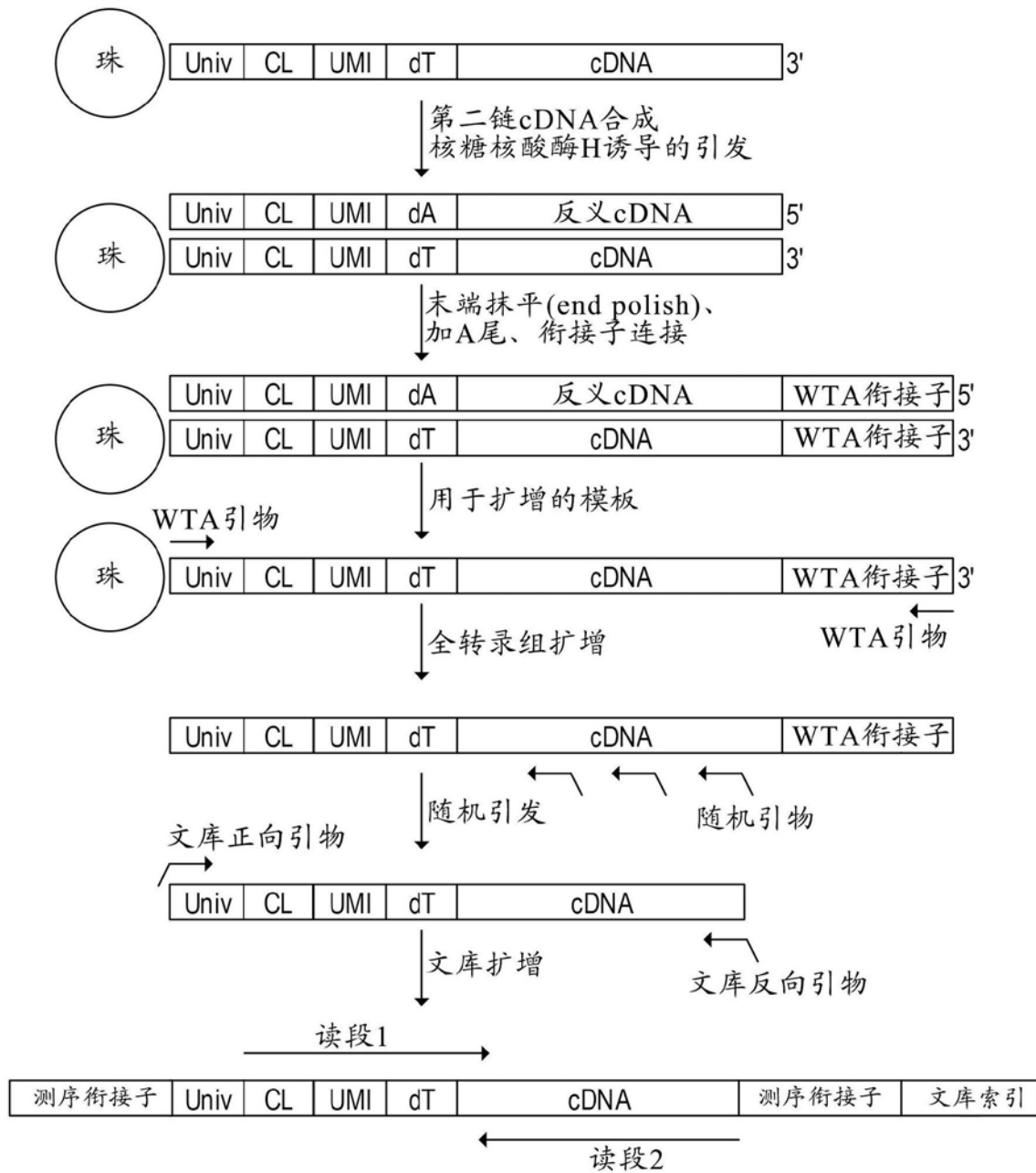


图9

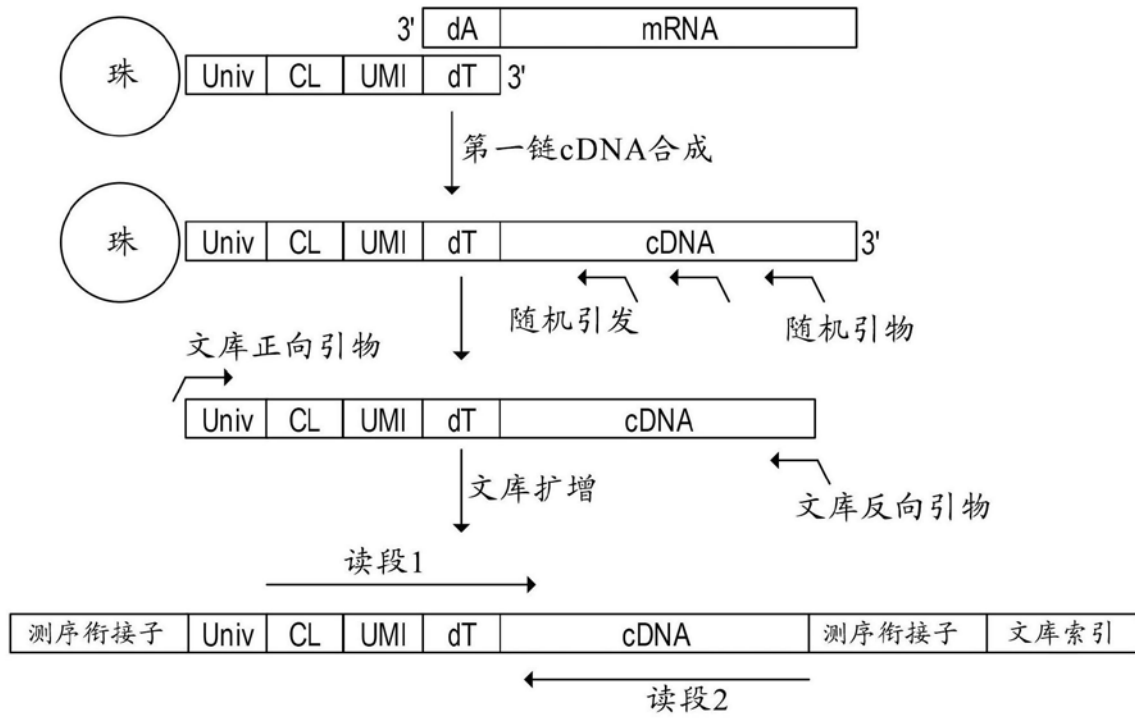


图10

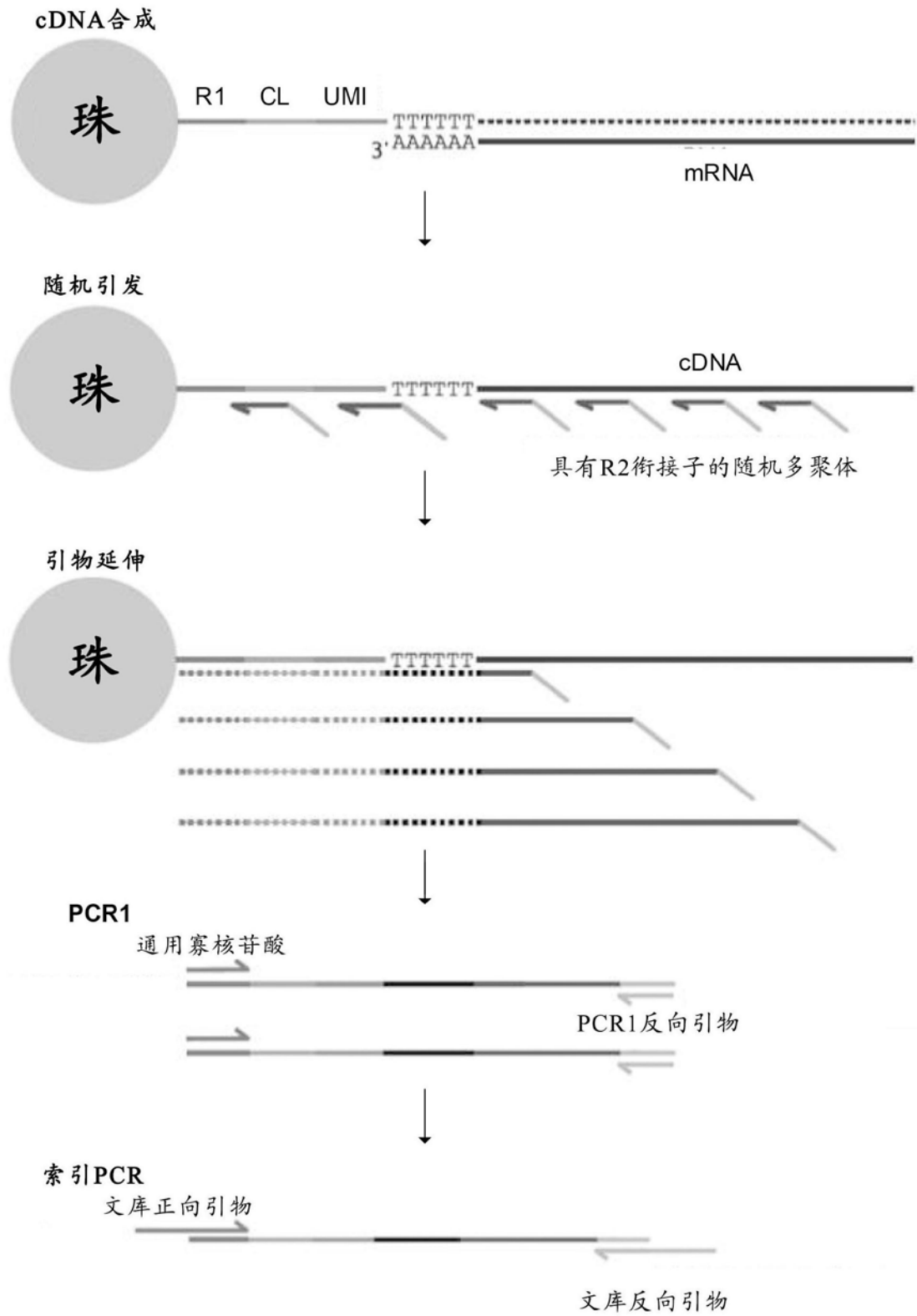


图11A

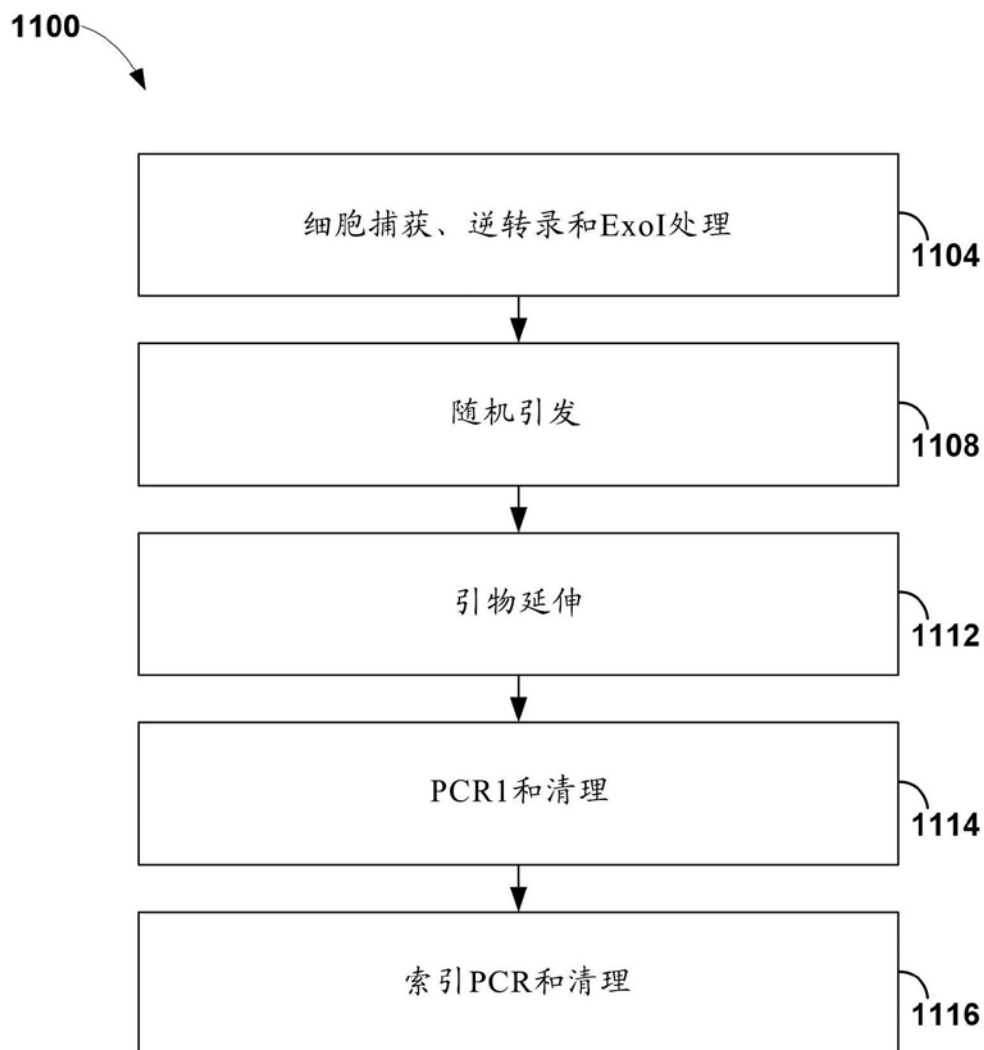


图11B

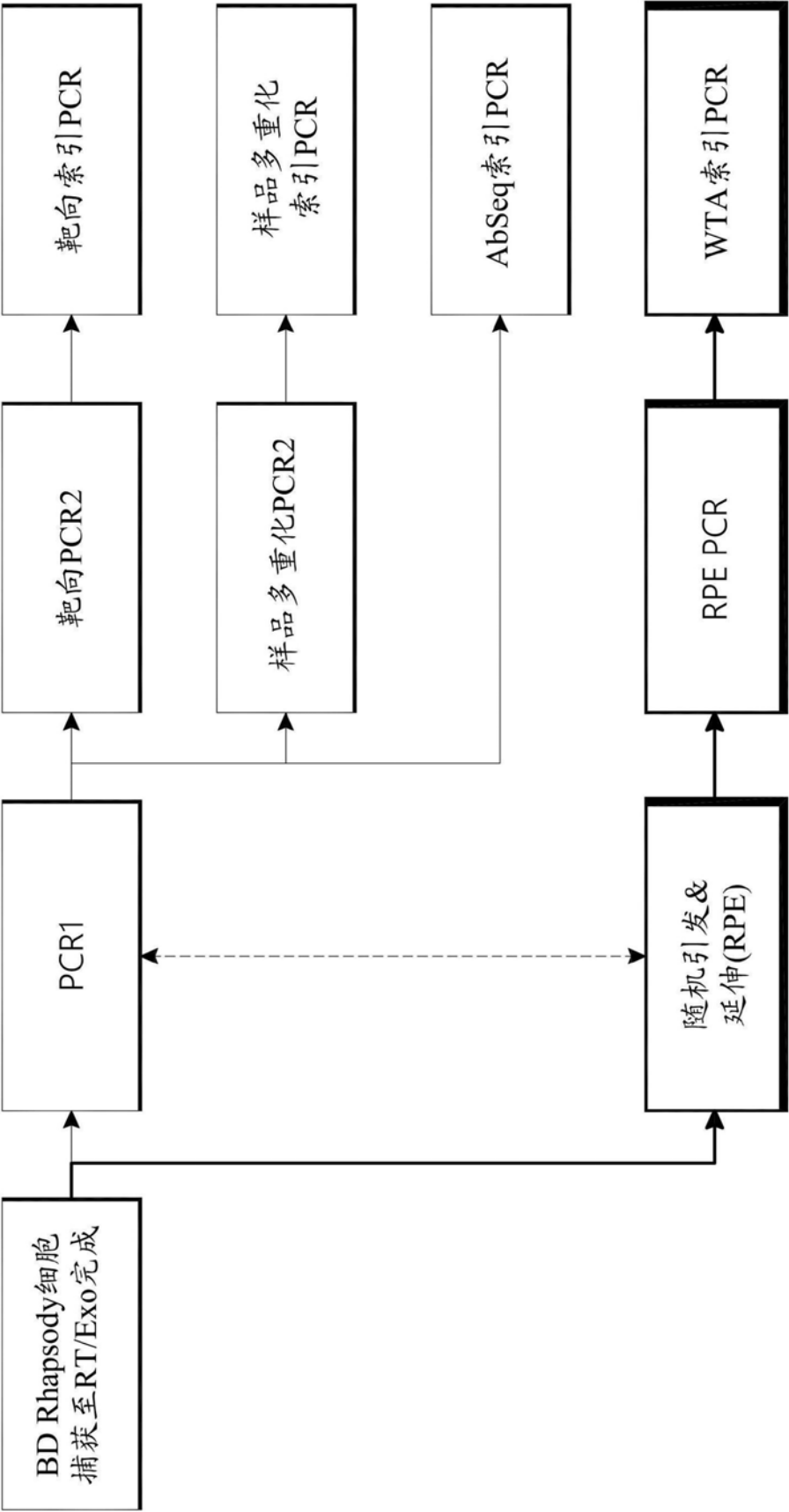


图12

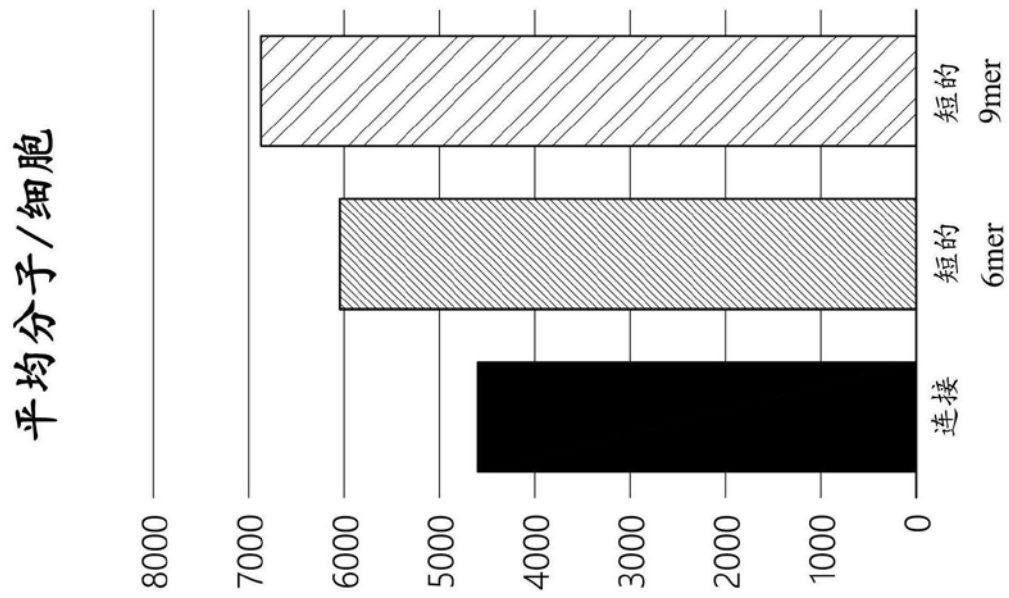


图13A

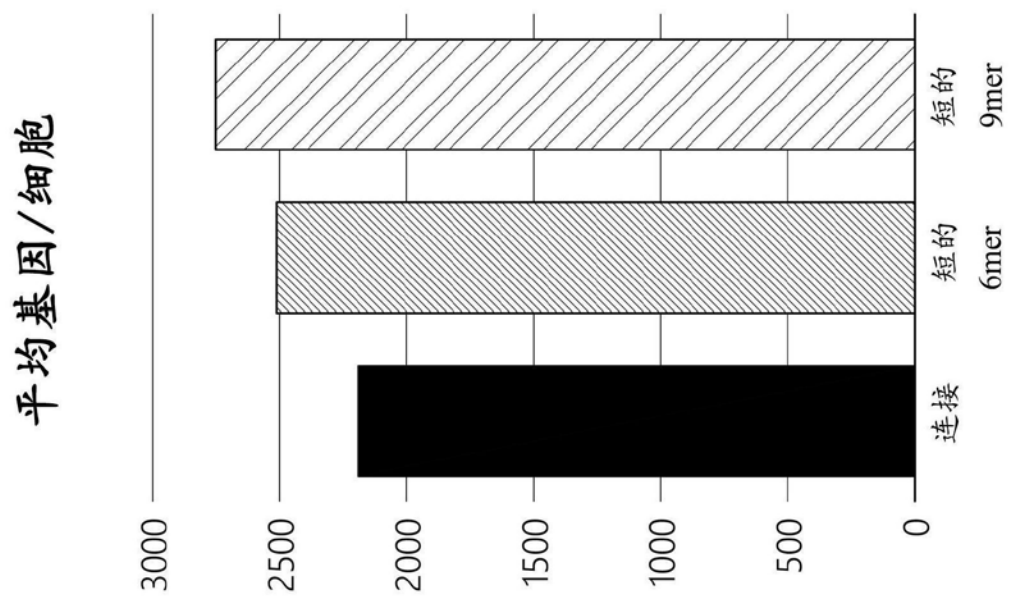


图13B

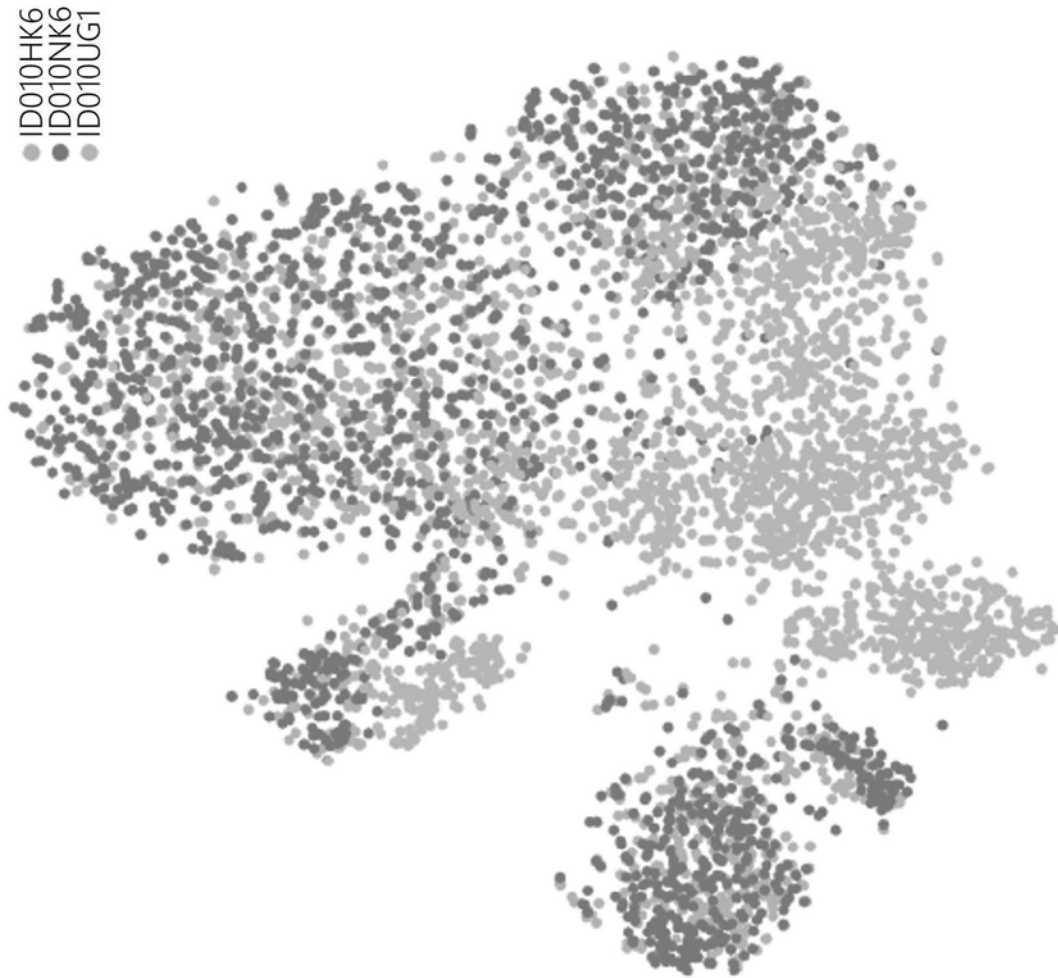


图14

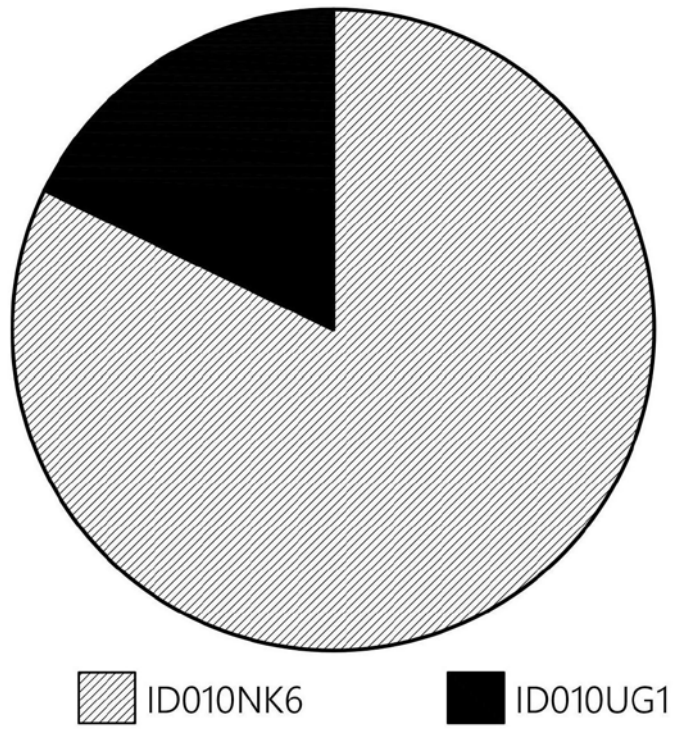


图15A

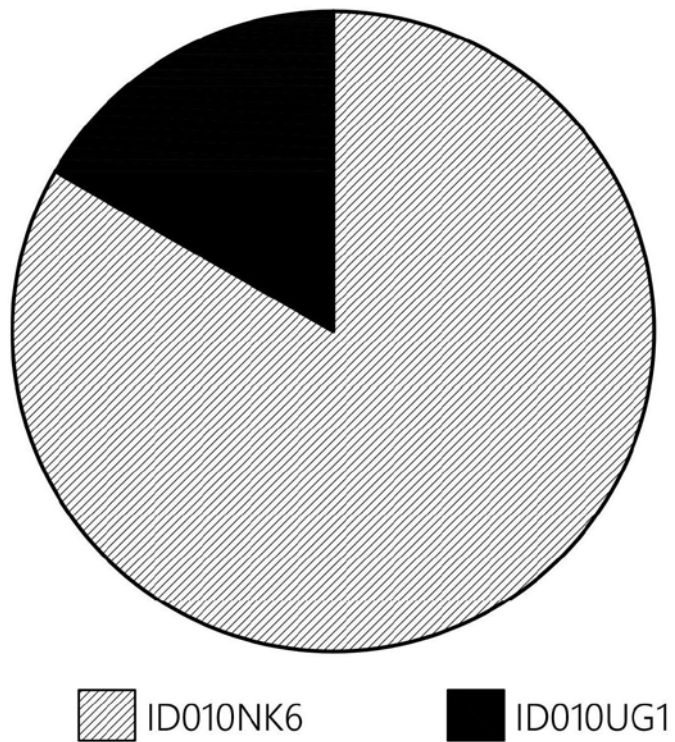


图15B

HK7 最终

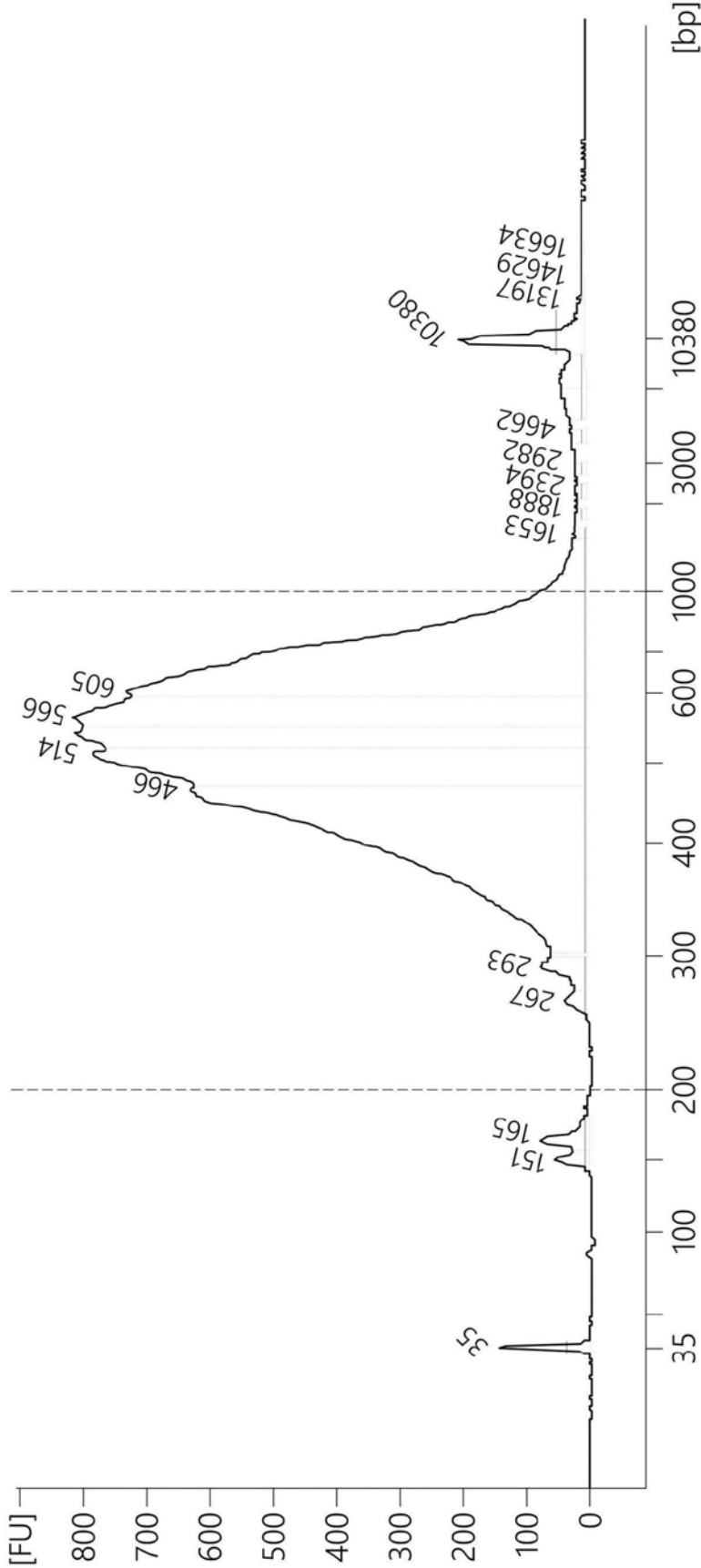


图16A

HK6最终

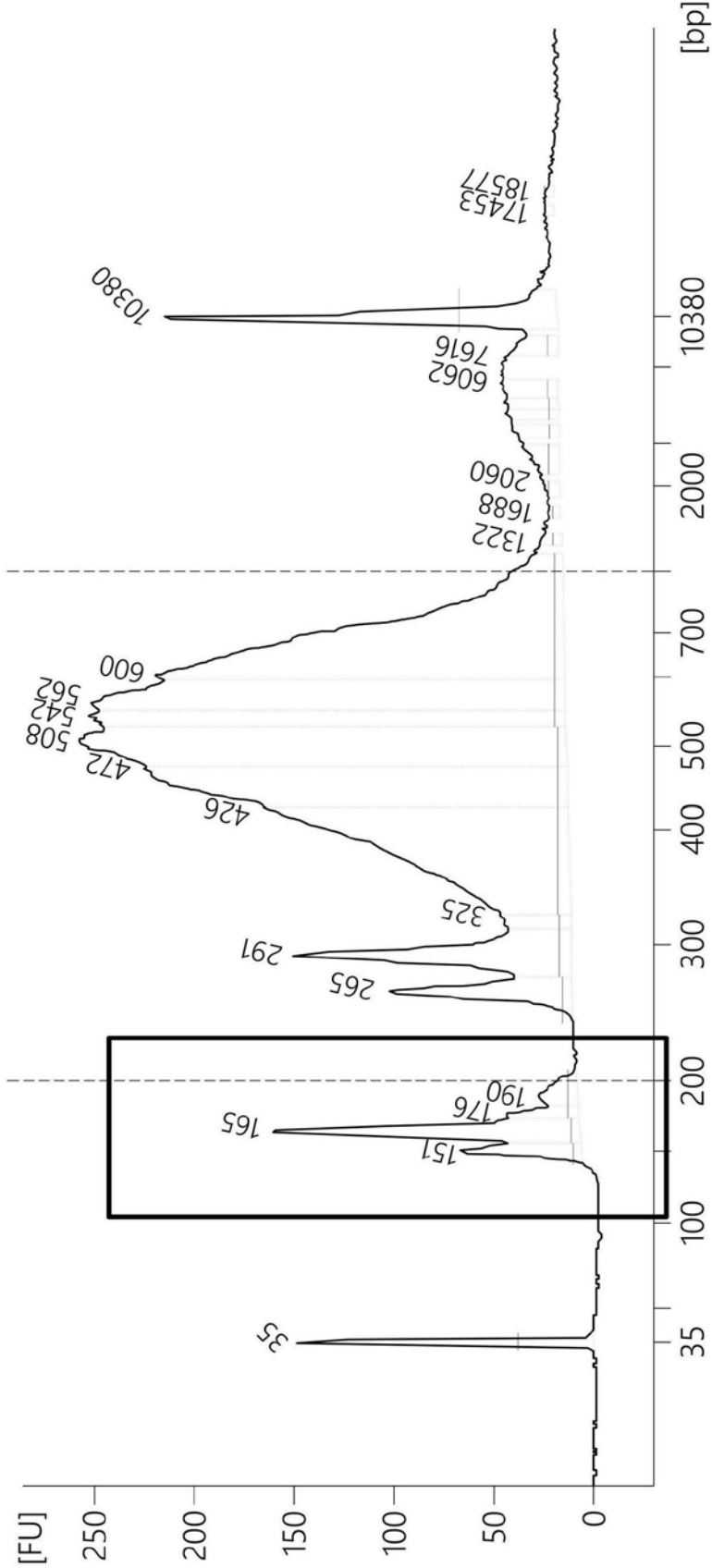


图16B

NK7最终

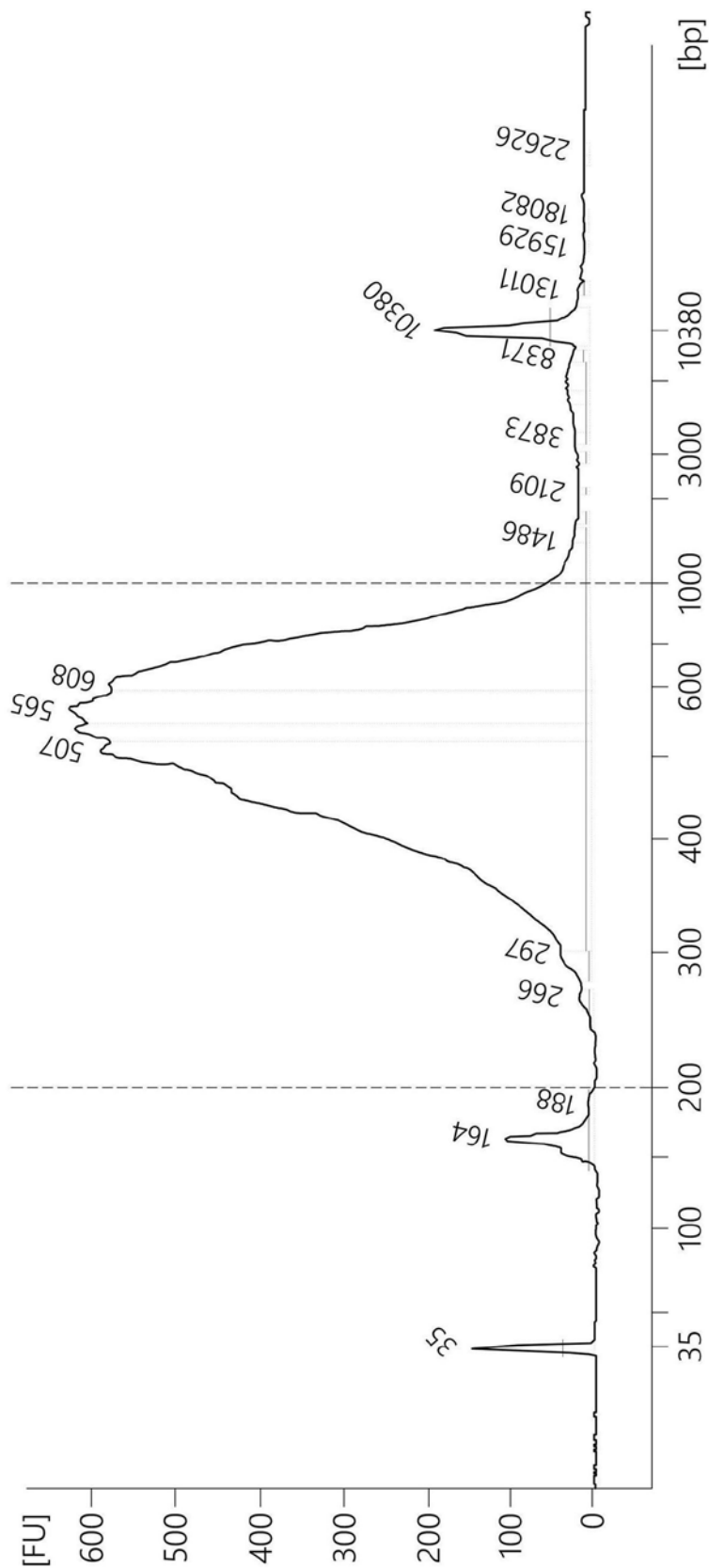


图16C

NK6 最终

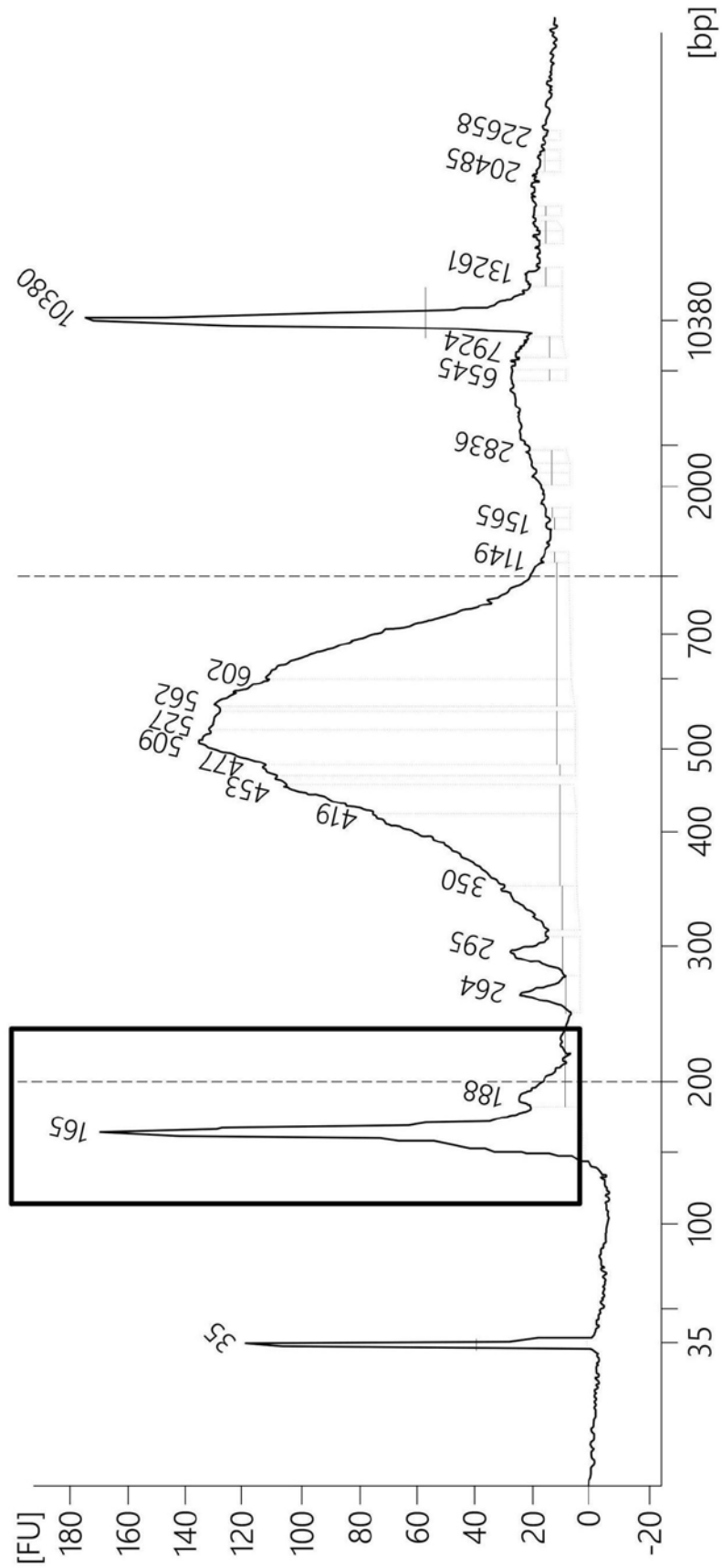


图16D

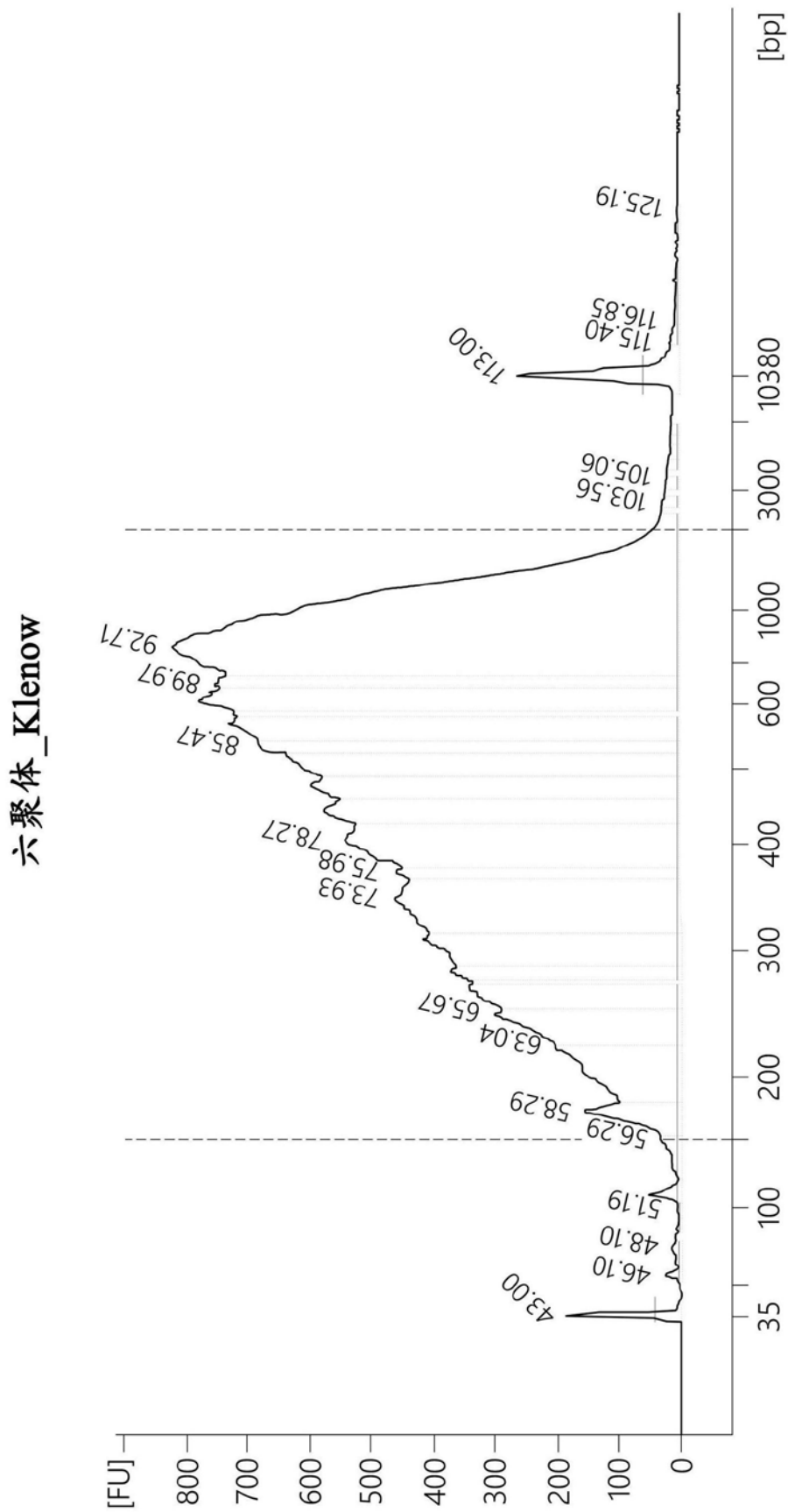


图17A

九聚体_Klenow

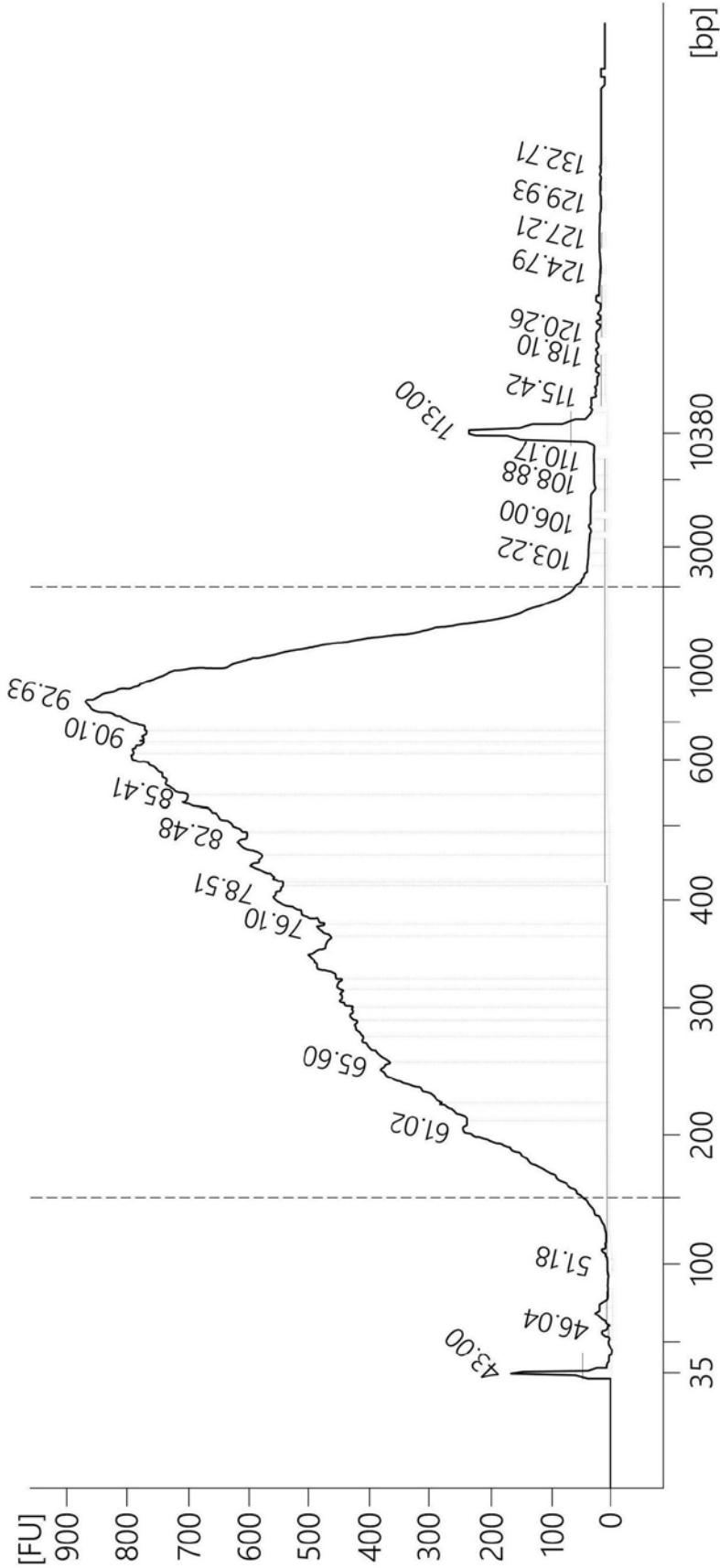


图17B

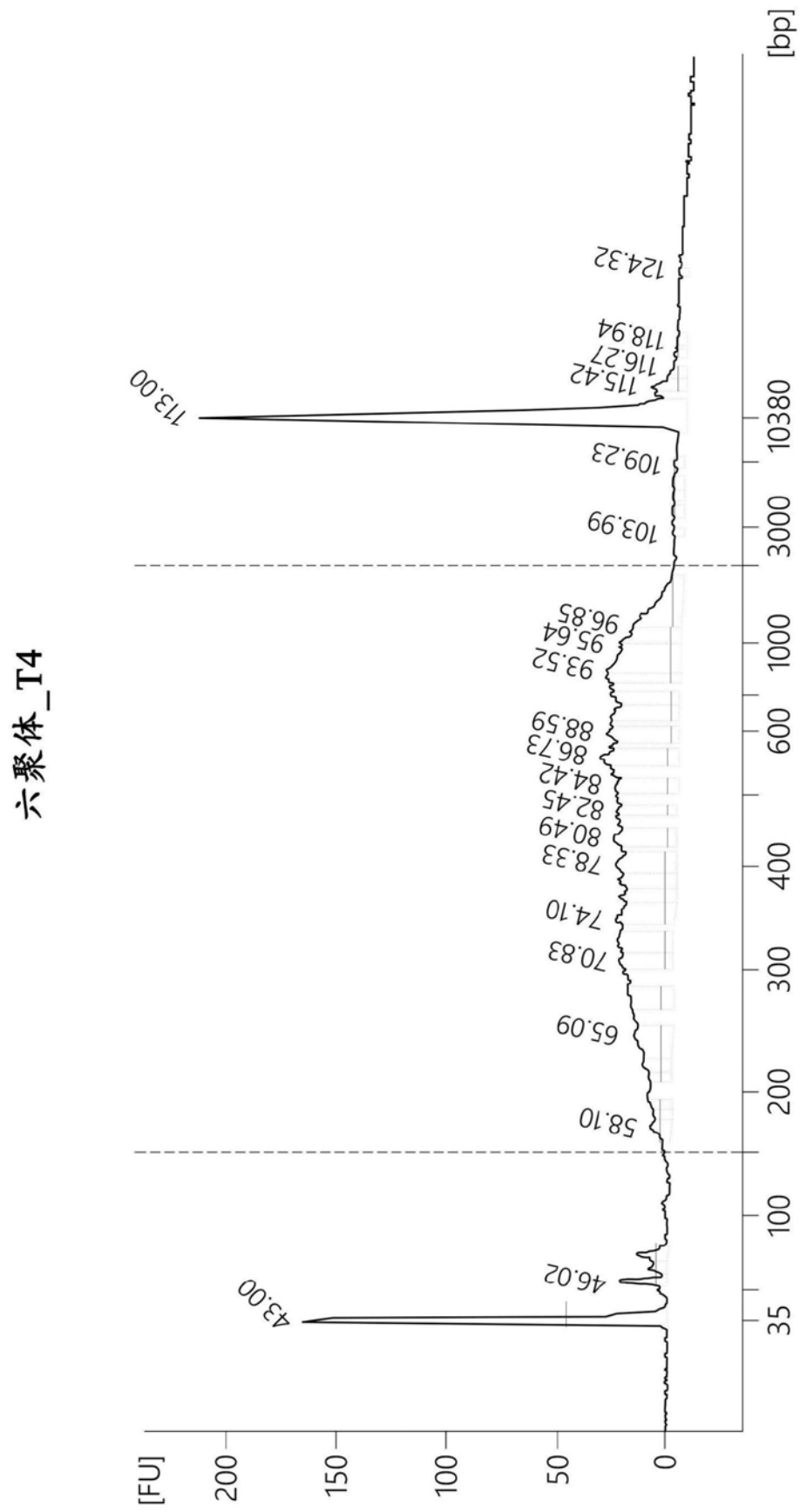


图17C

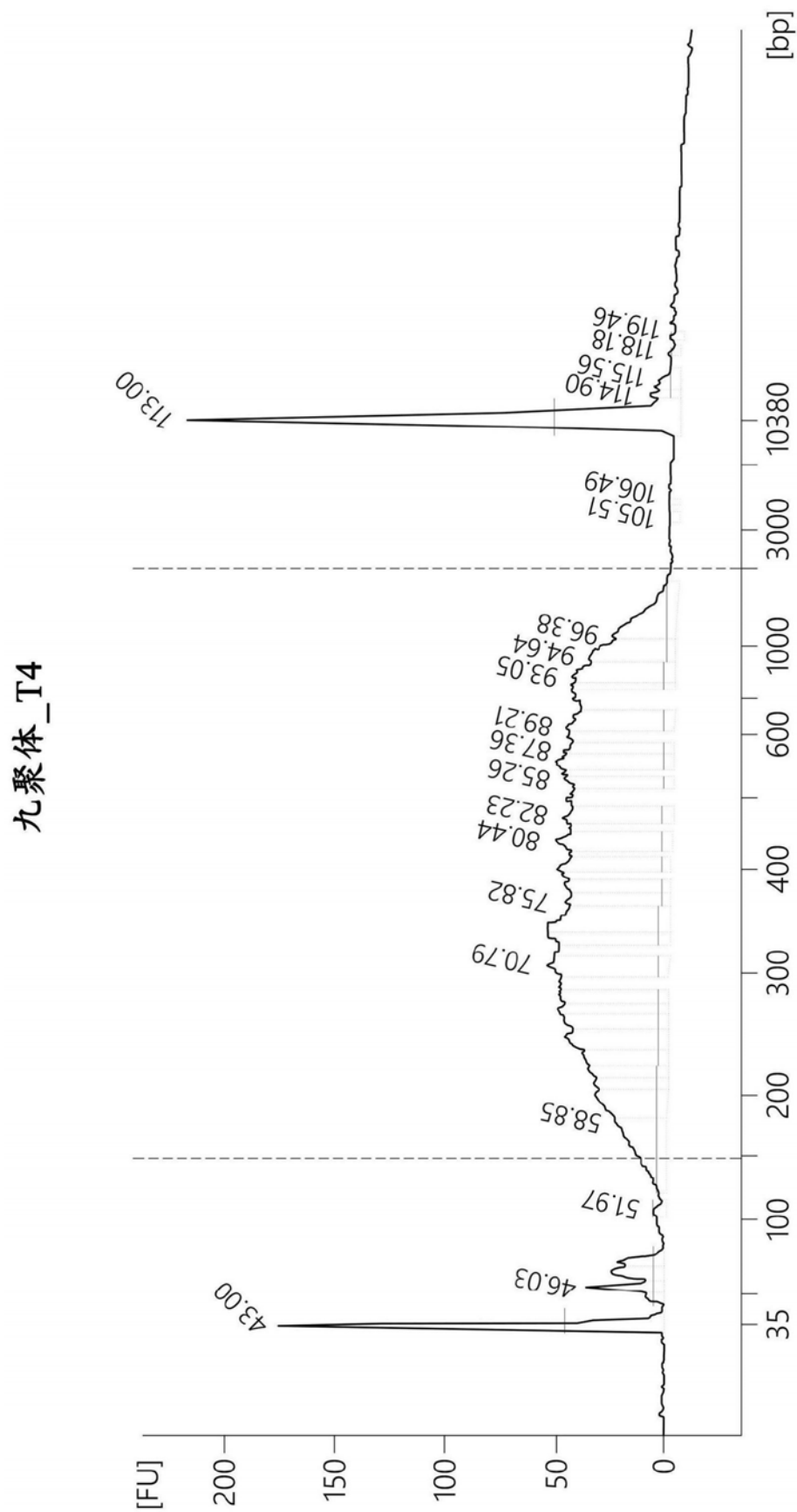


图17D

ID017c-1

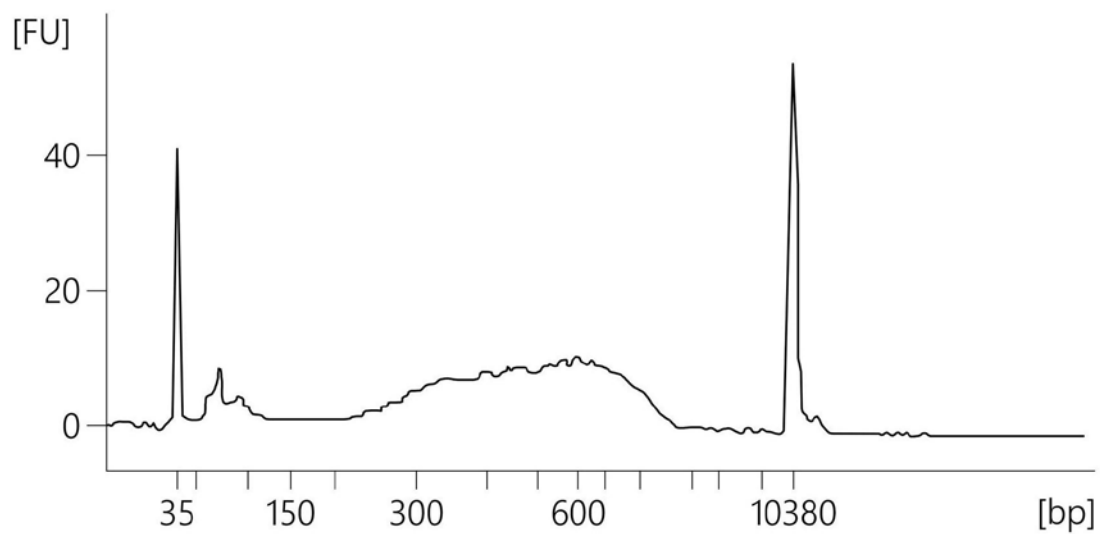


图18A

ID017c-2

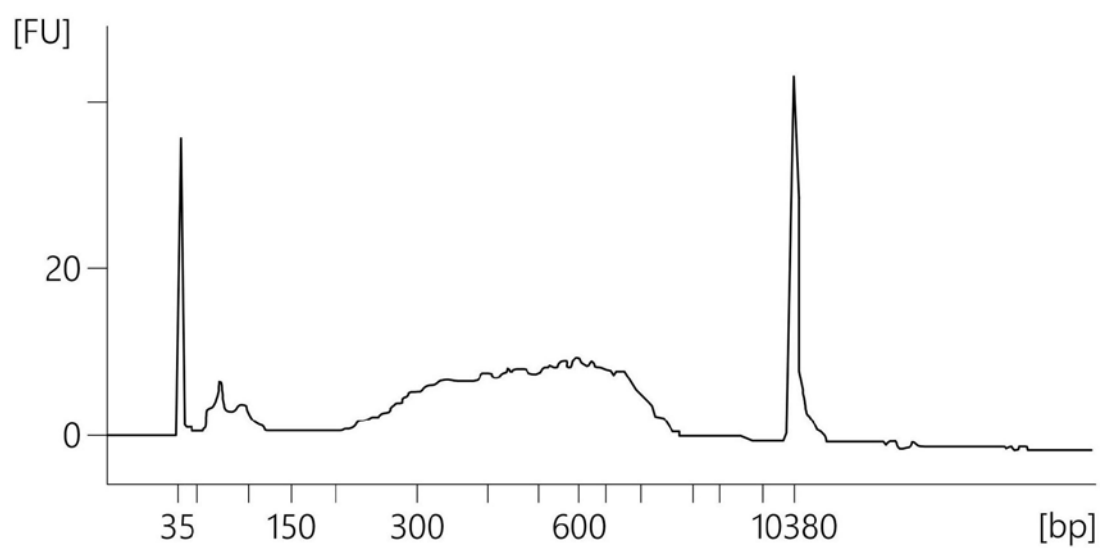


图18B

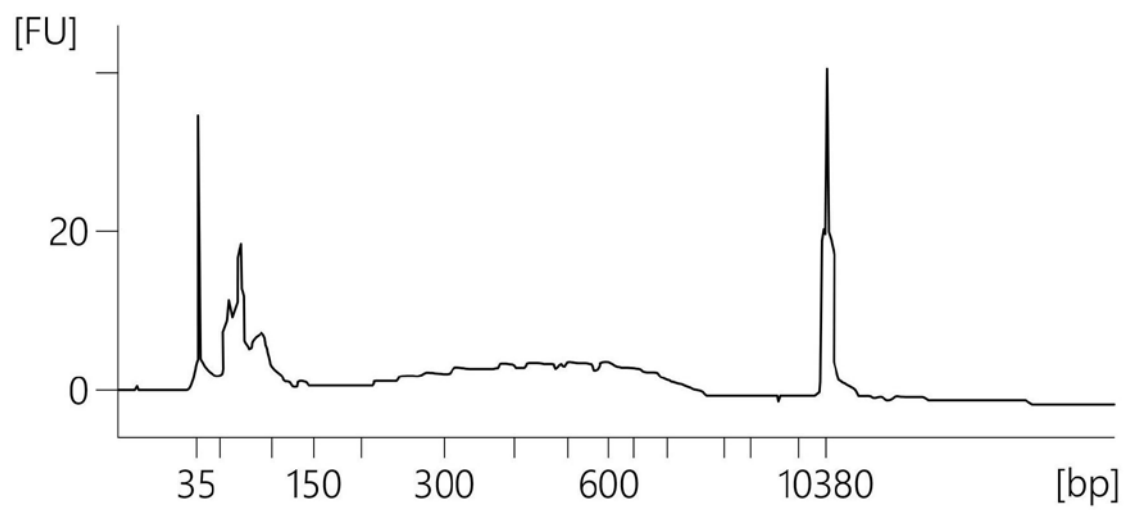
ID017c-3

图18C

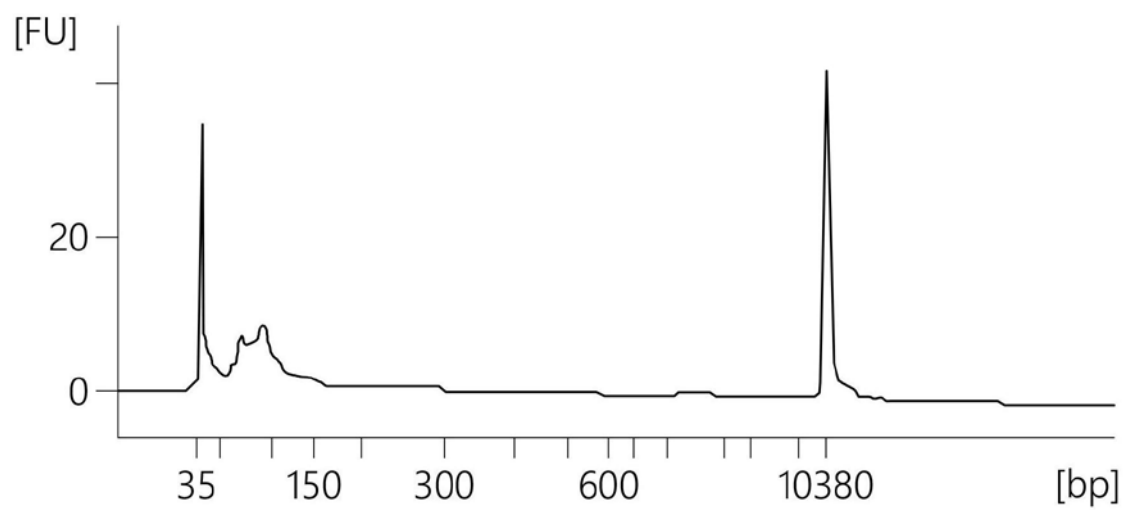
ID017c-4

图18D

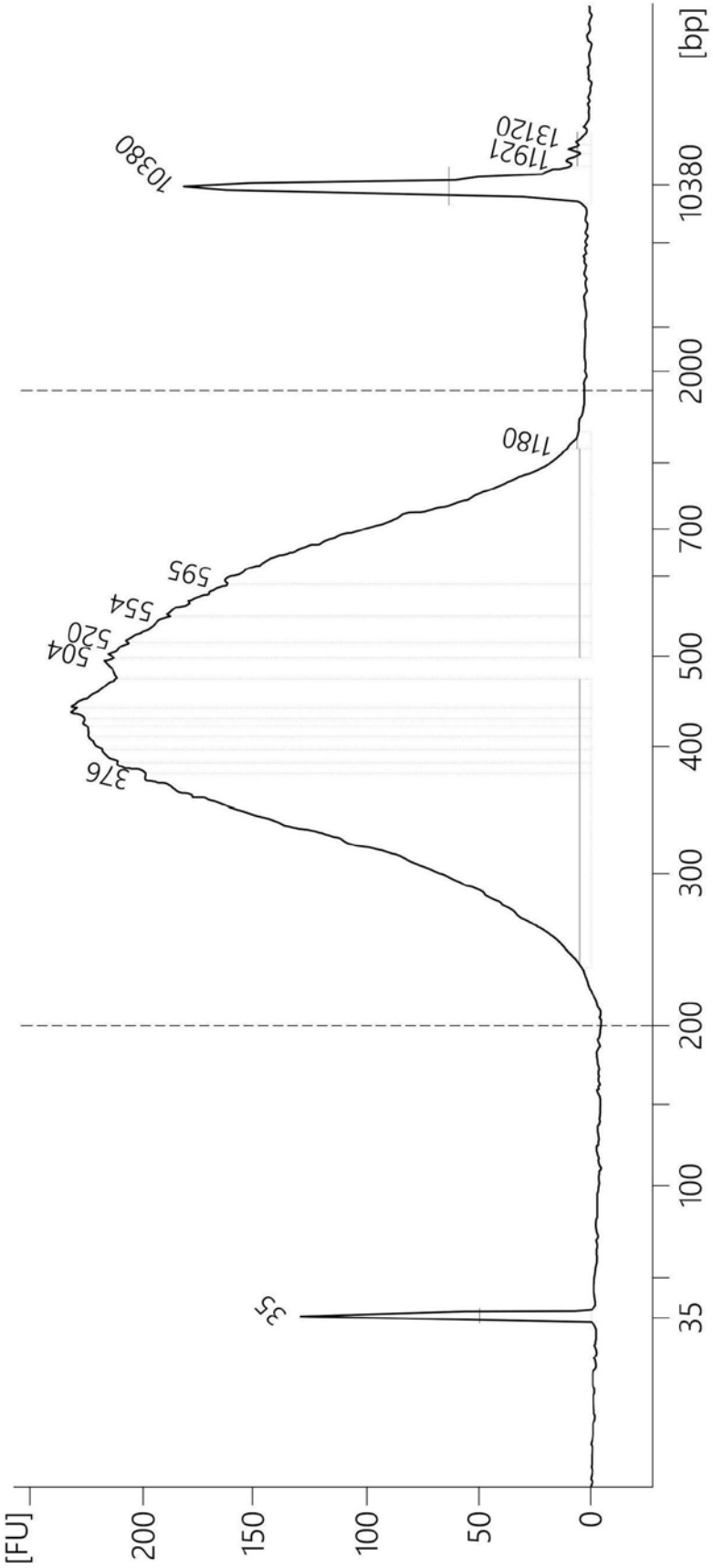


图19A

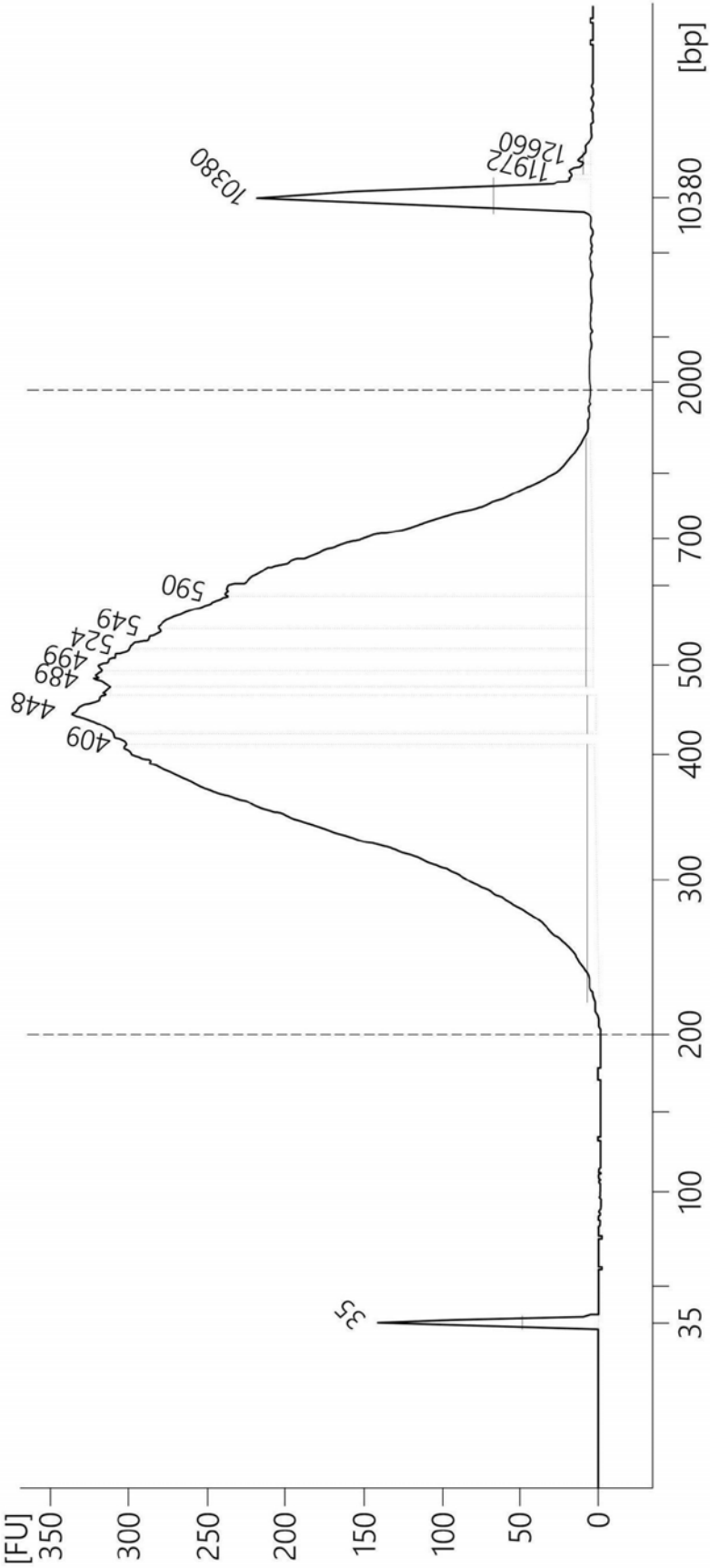


图19B

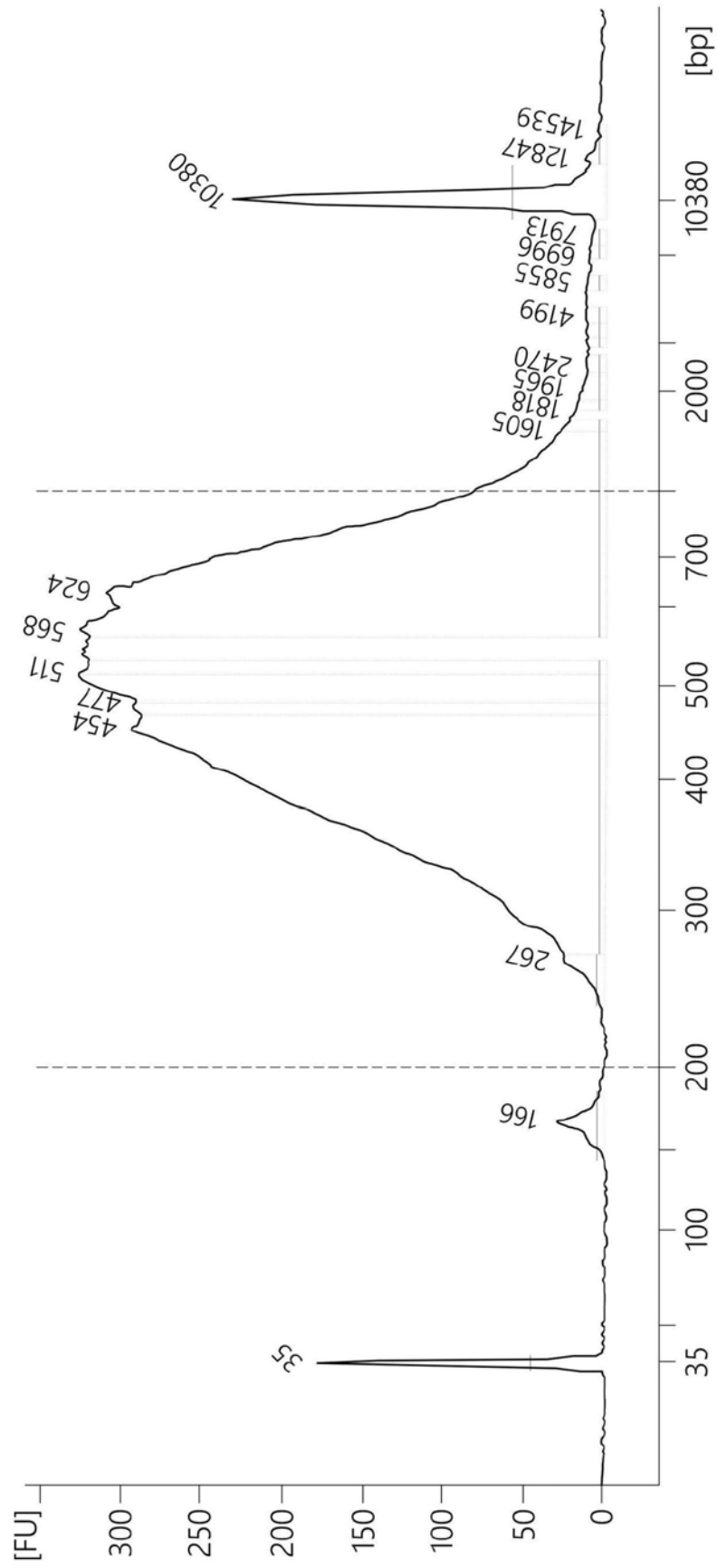


图19C

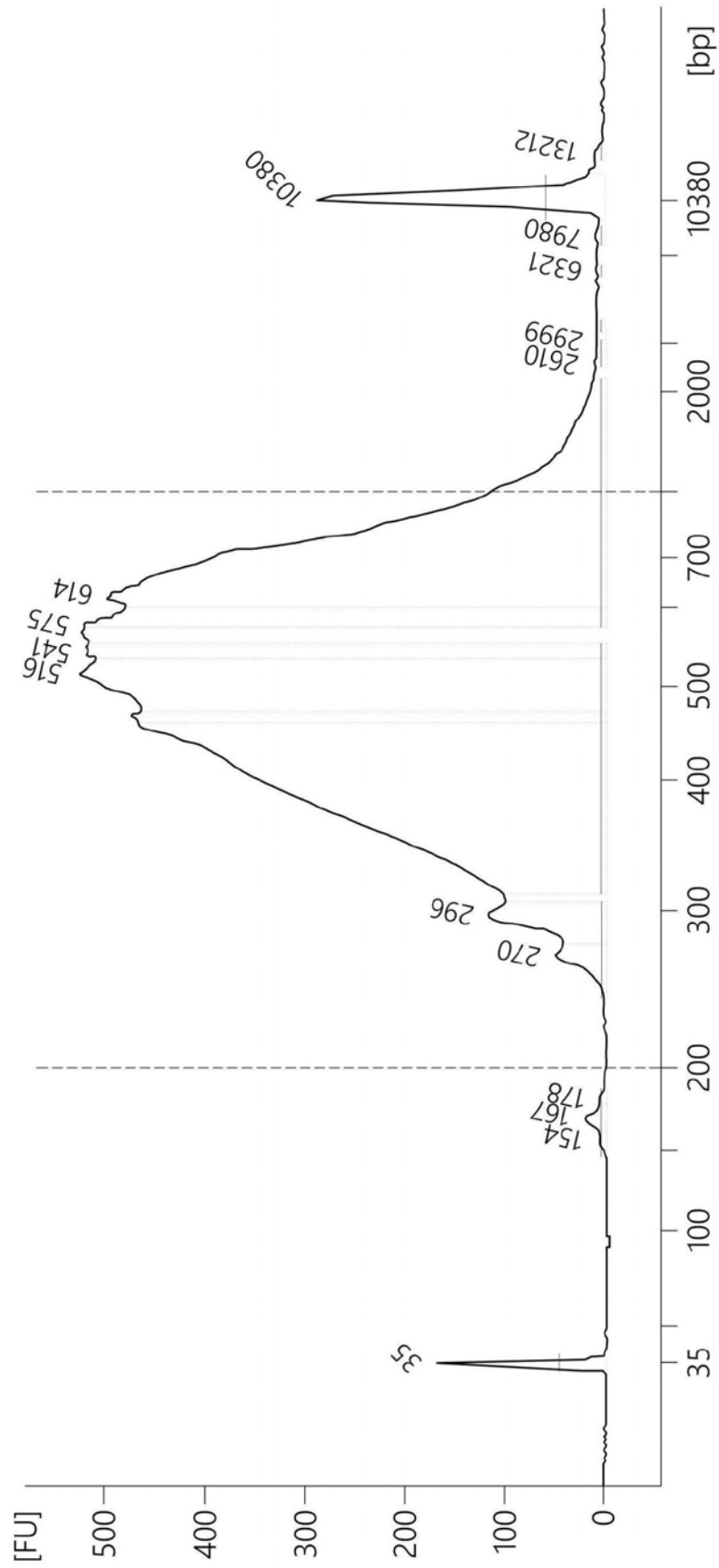


图19D

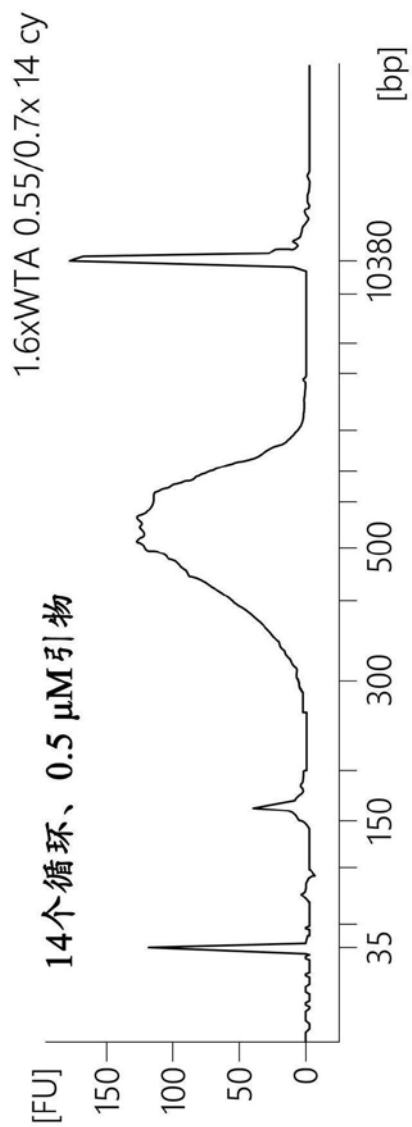


图20A

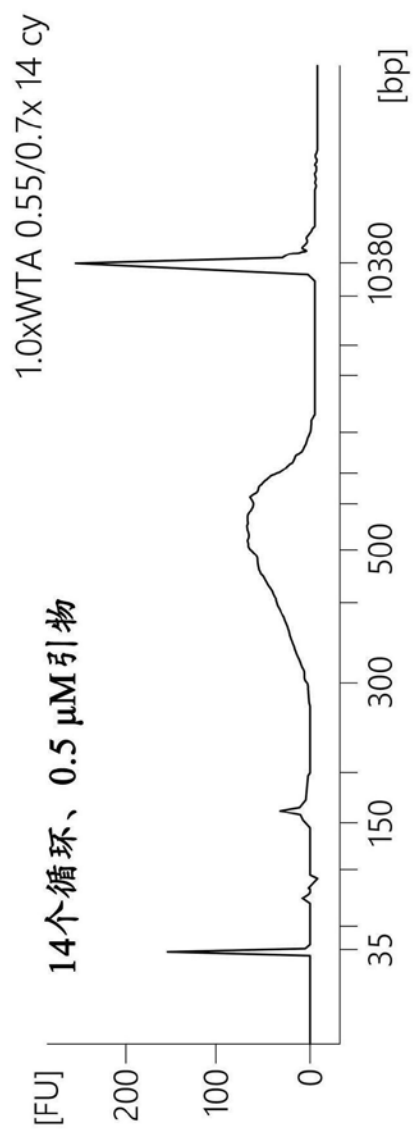


图20B

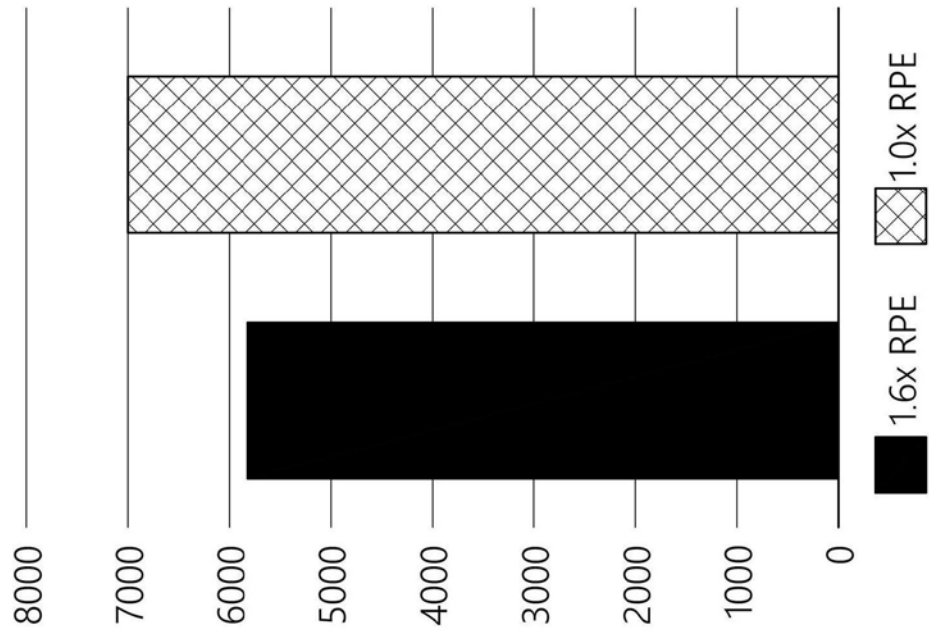


图20C

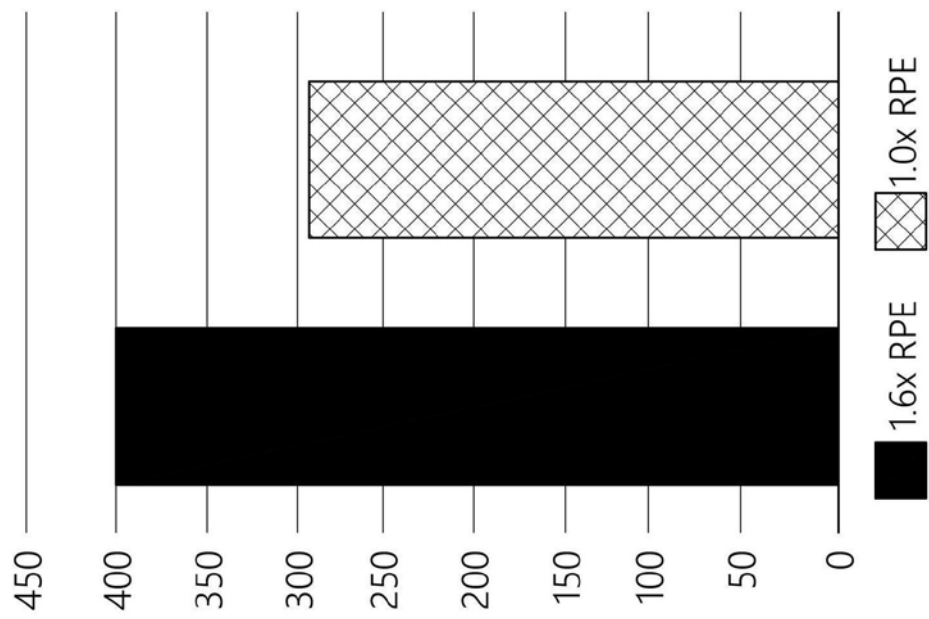


图20D

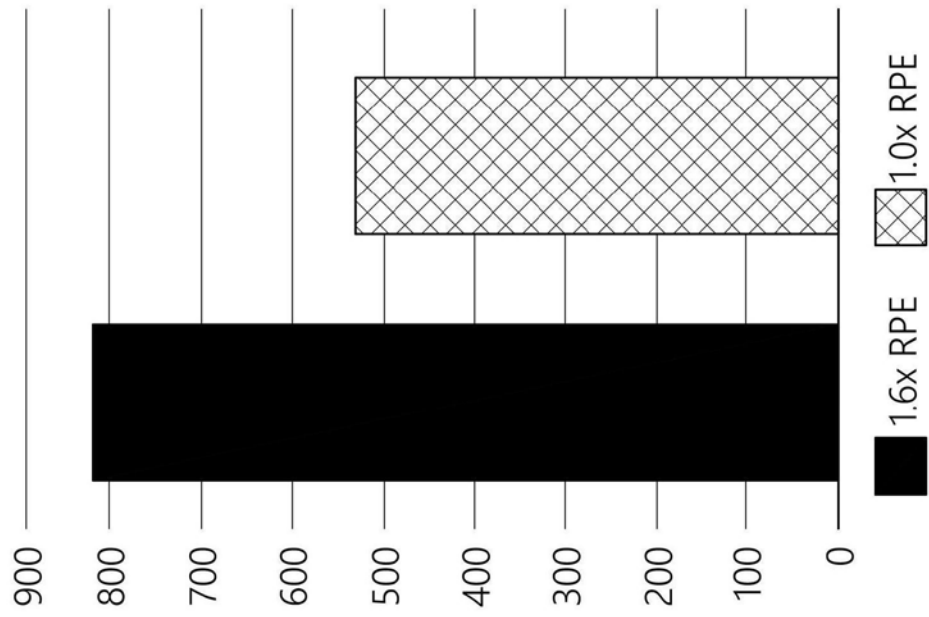


图20E

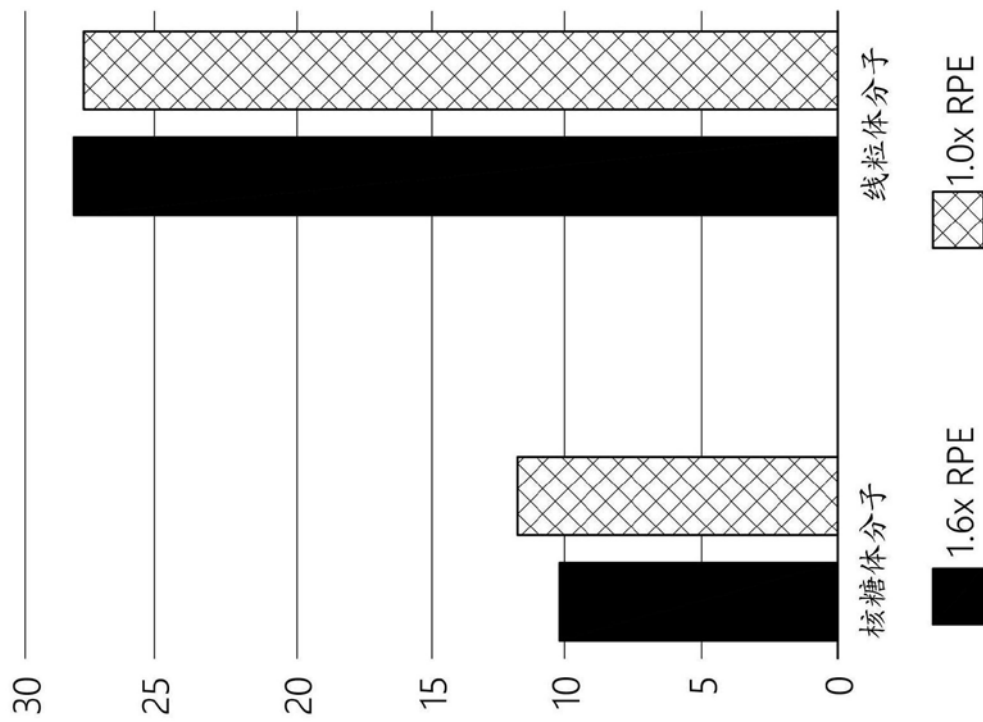


图20F

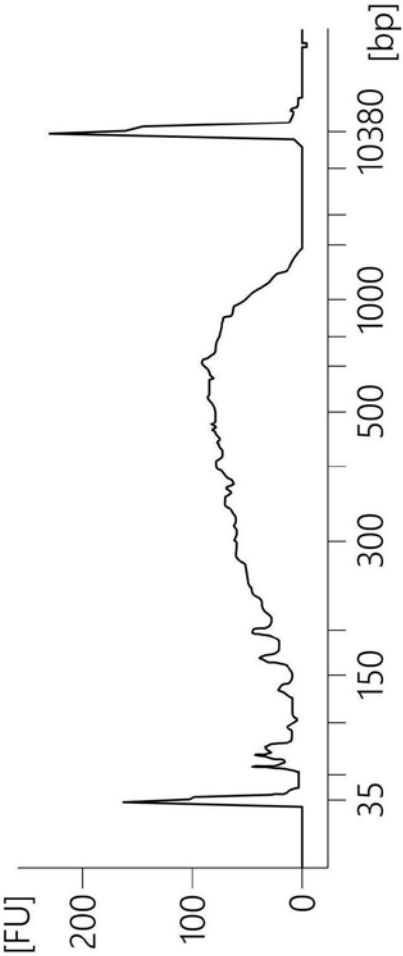


图21A

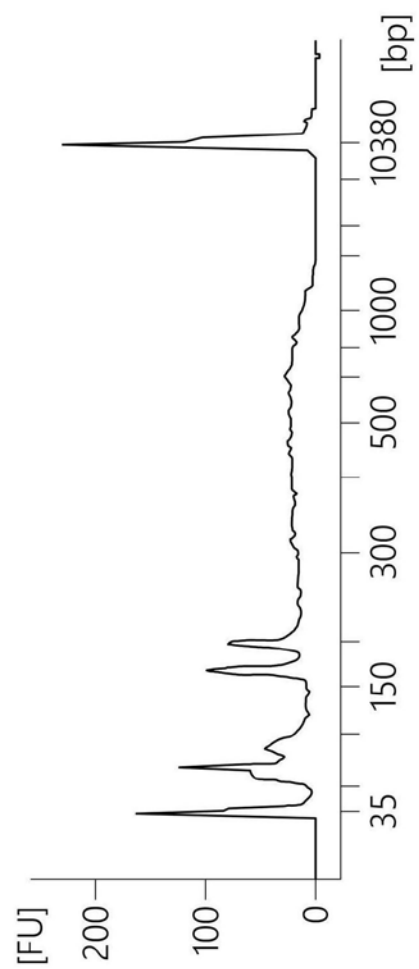


图21B

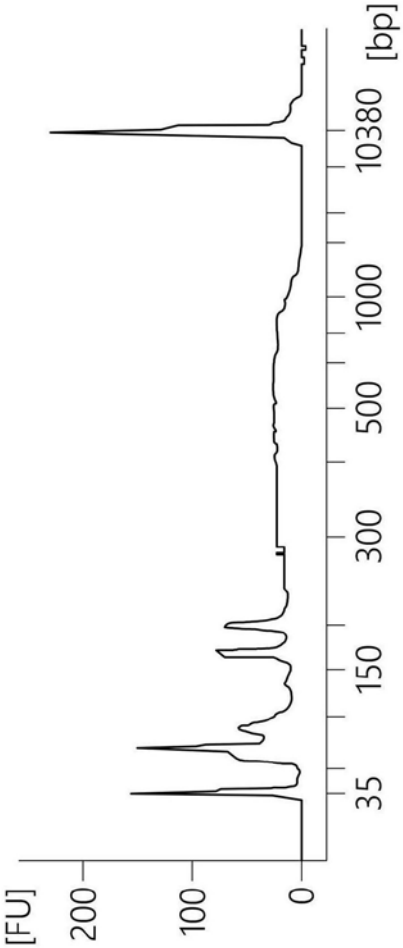


图21C

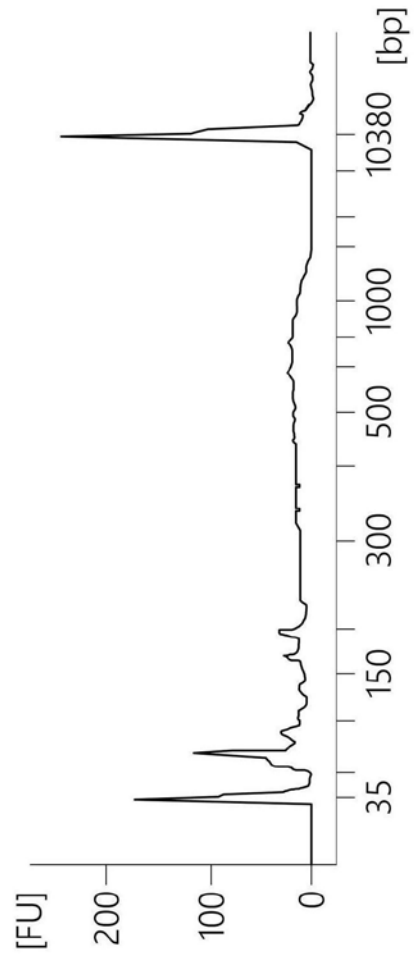


图21D

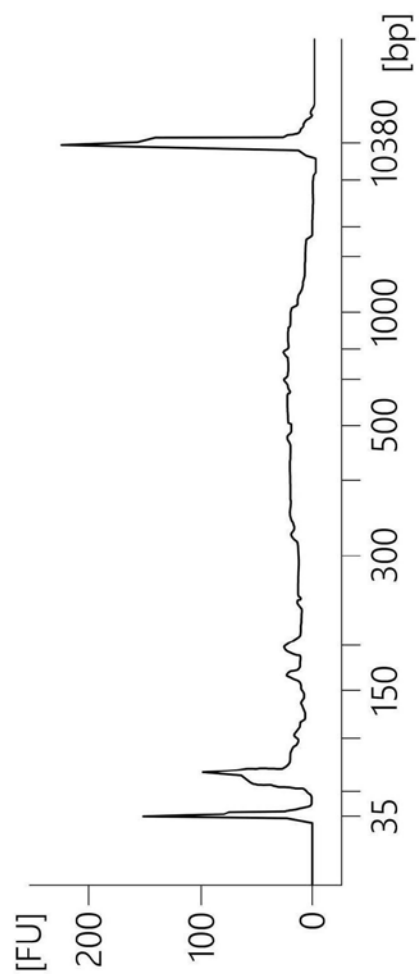


图21E

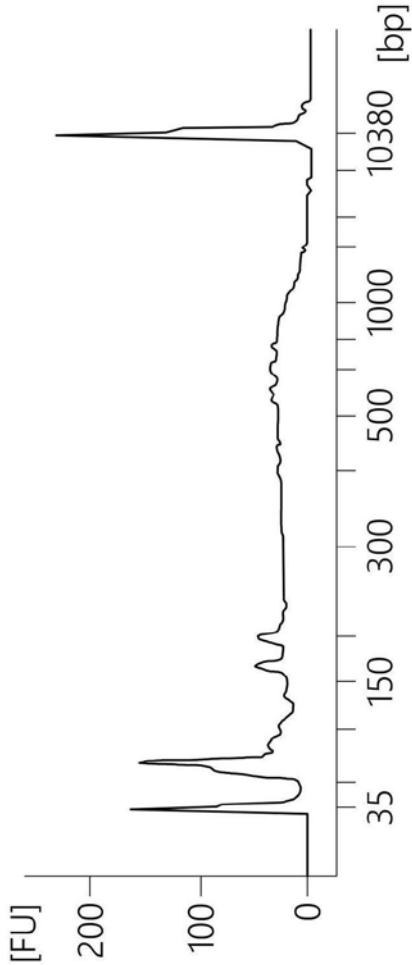


图21F

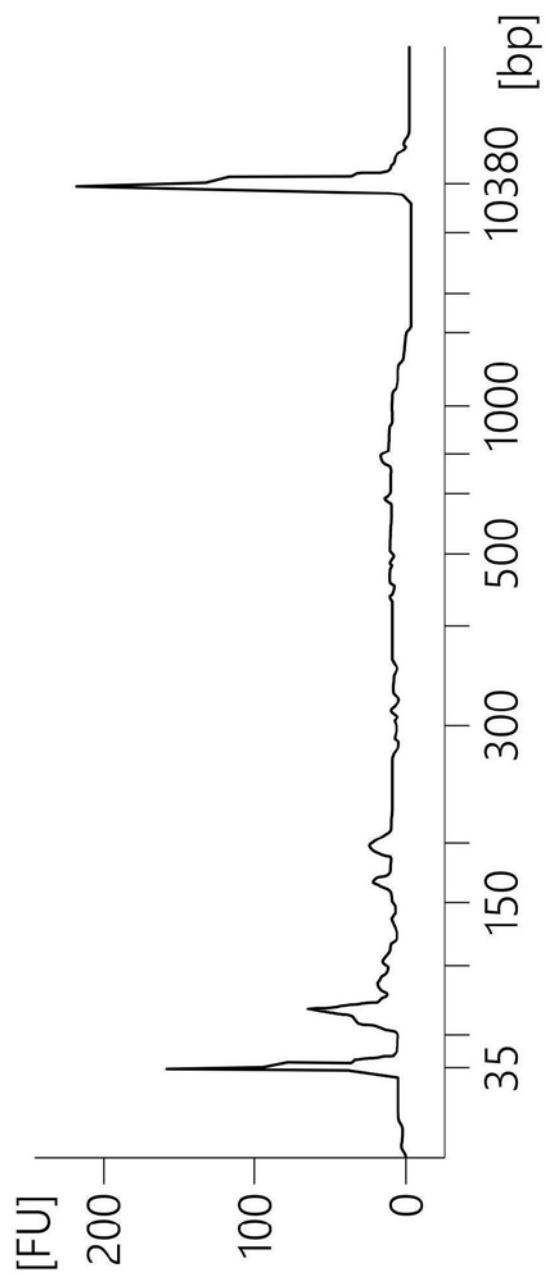


图21G

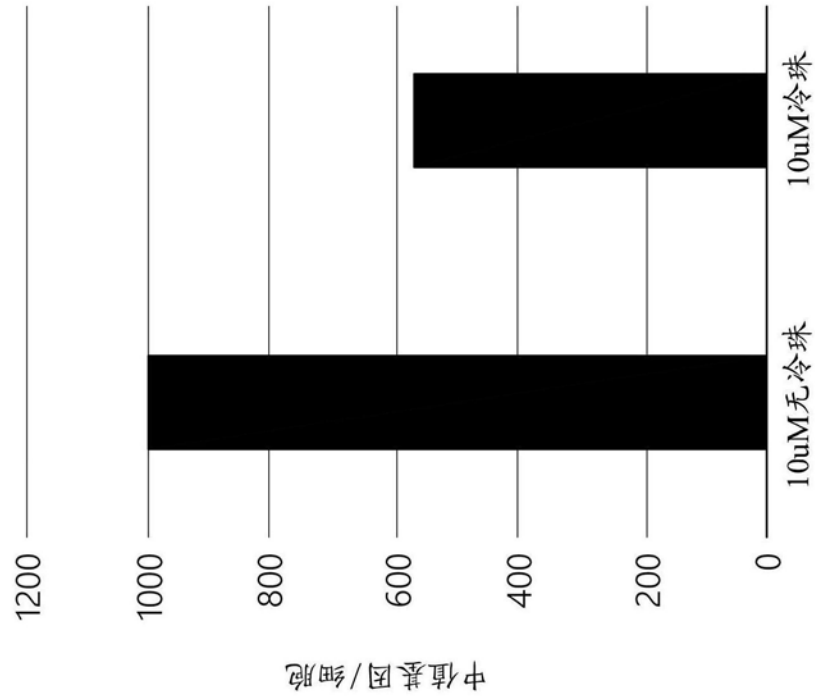


图22A

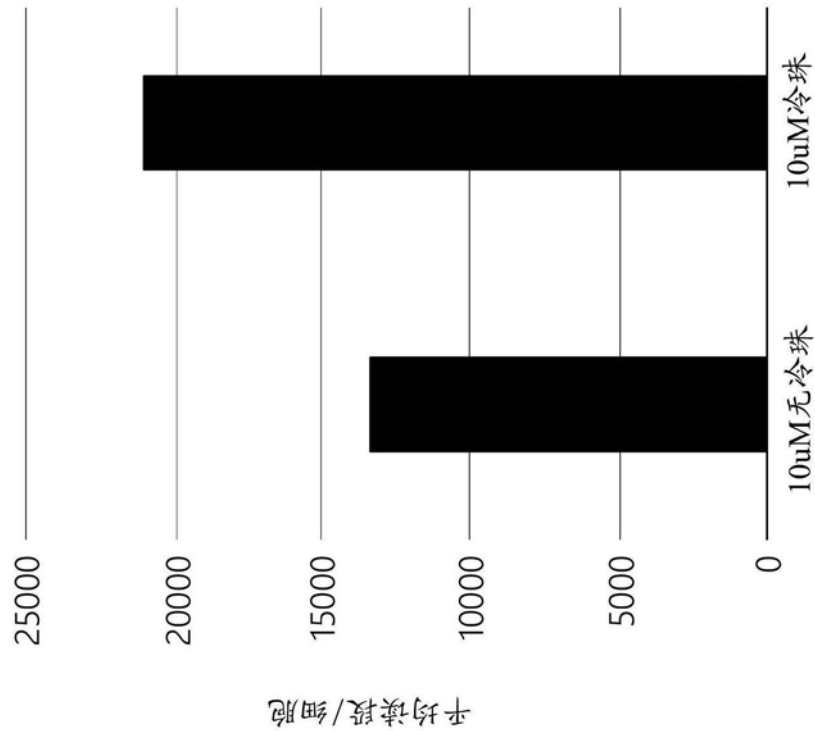


图22B

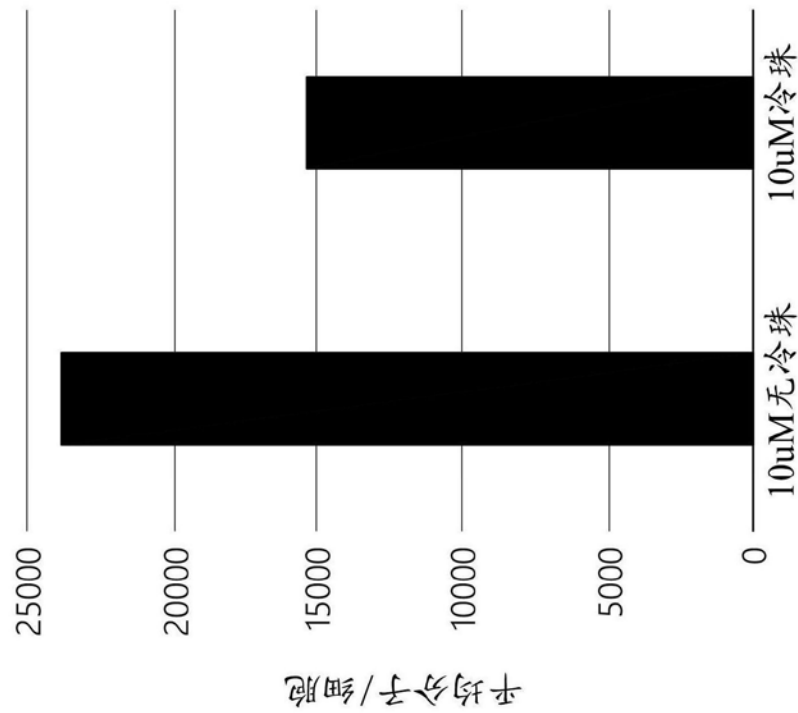


图22C

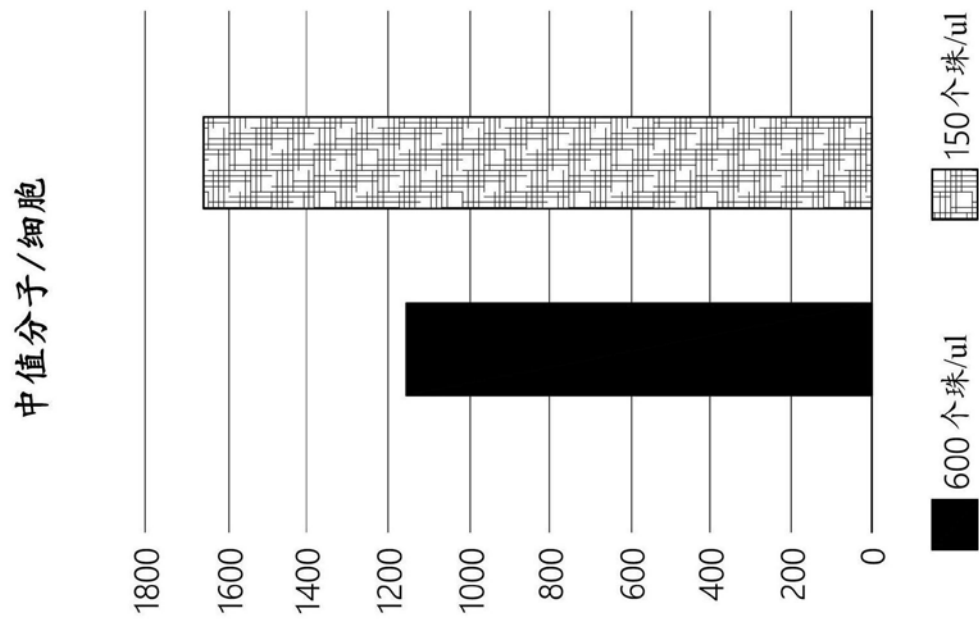


图22D

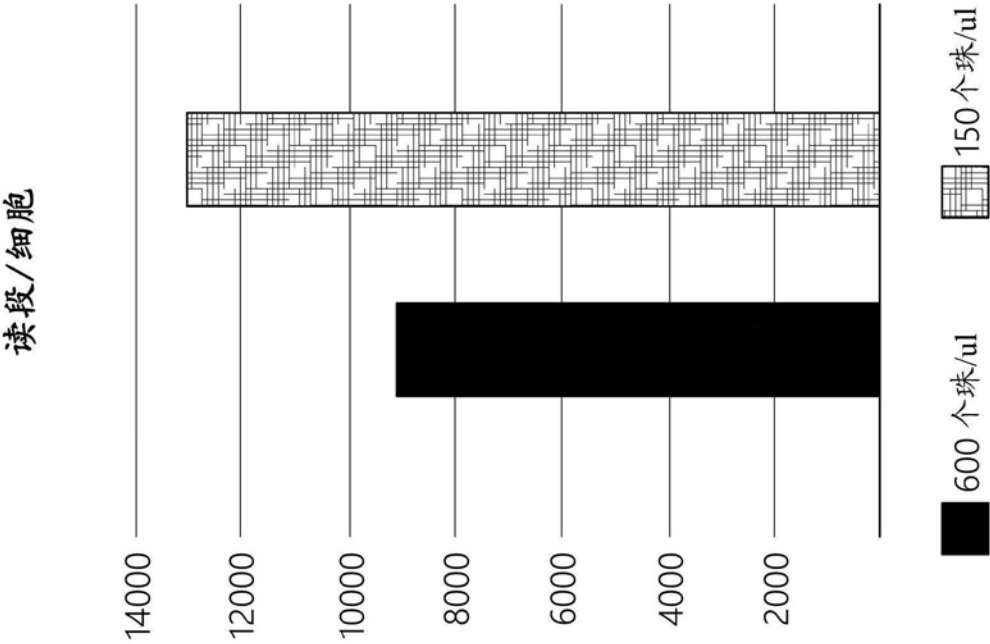


图22E

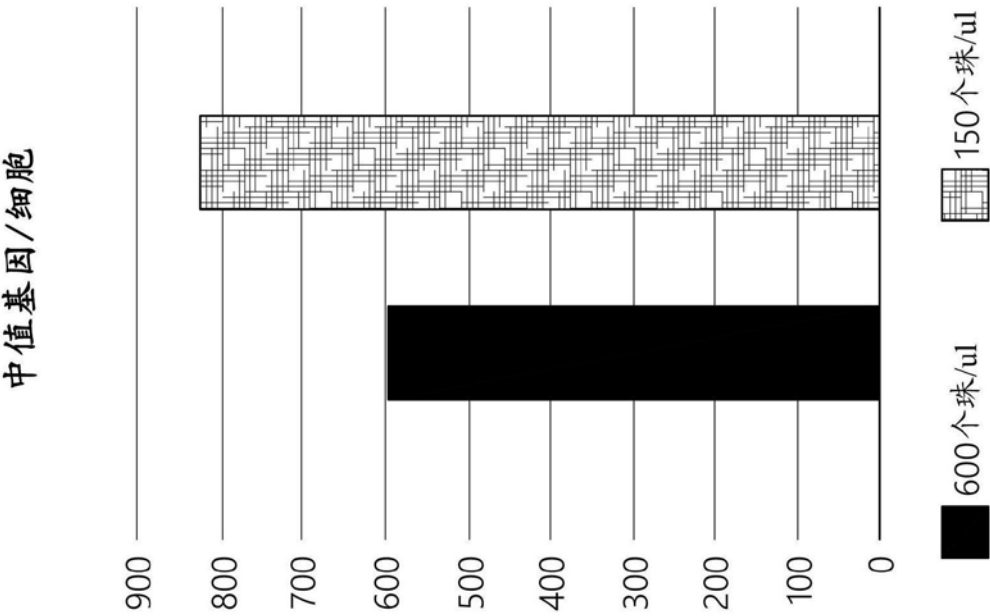


图22F

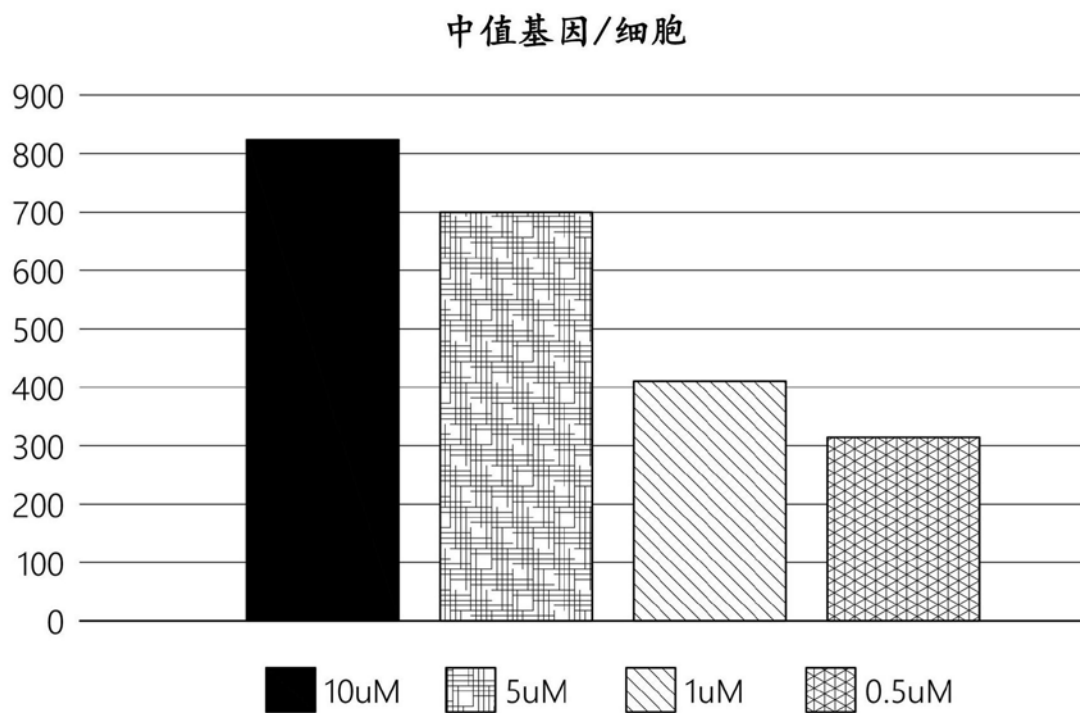


图23A

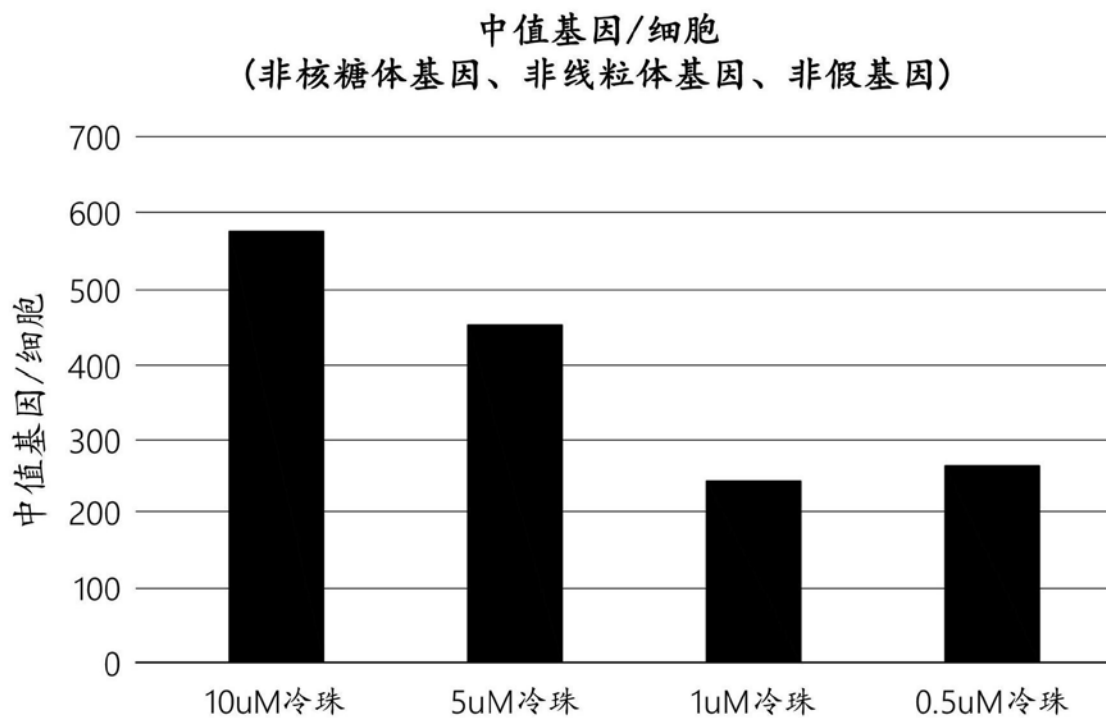


图23B

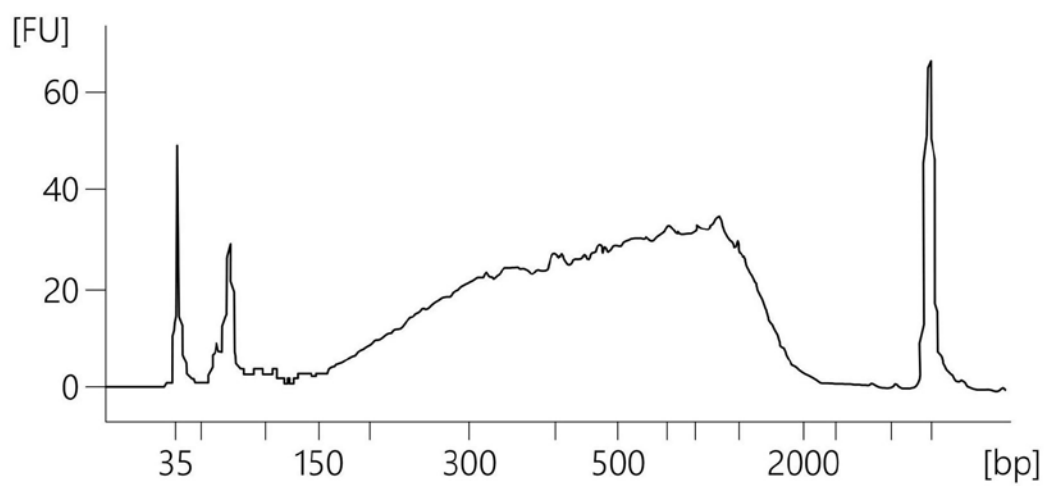


图24A

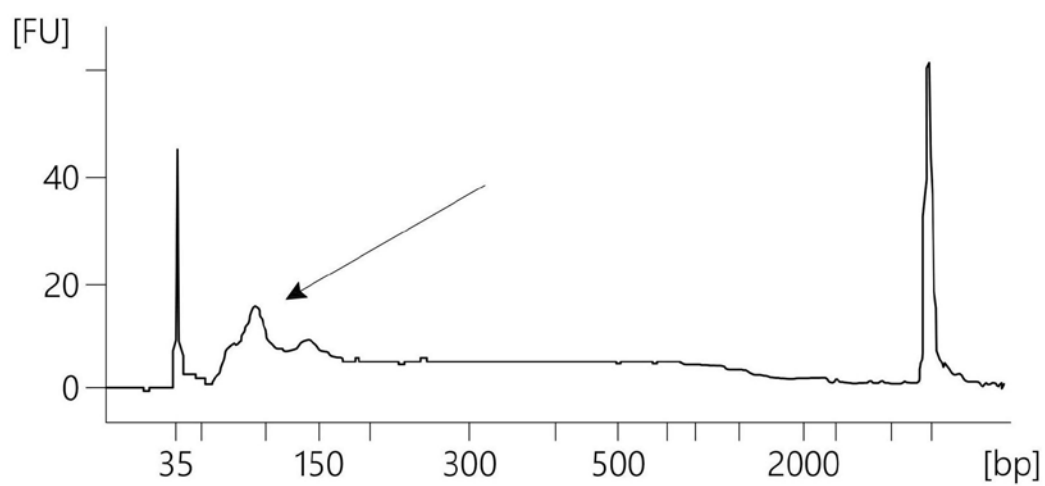


图24B

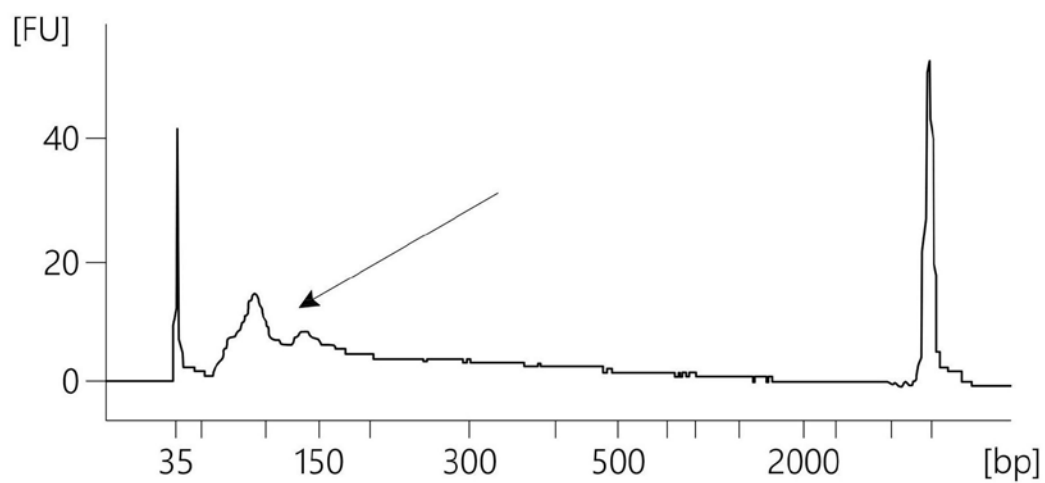


图24C

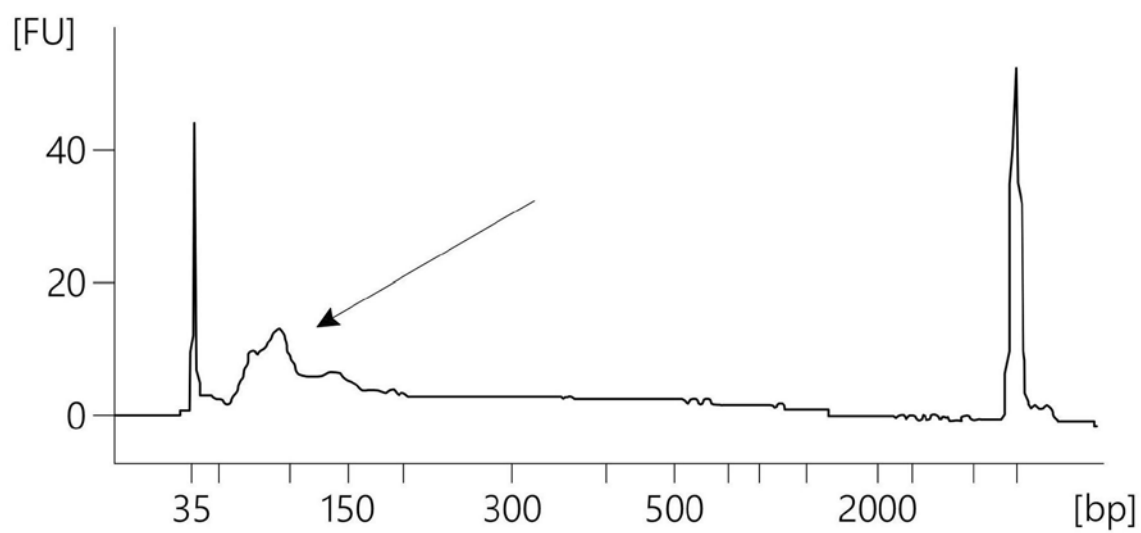


图24D

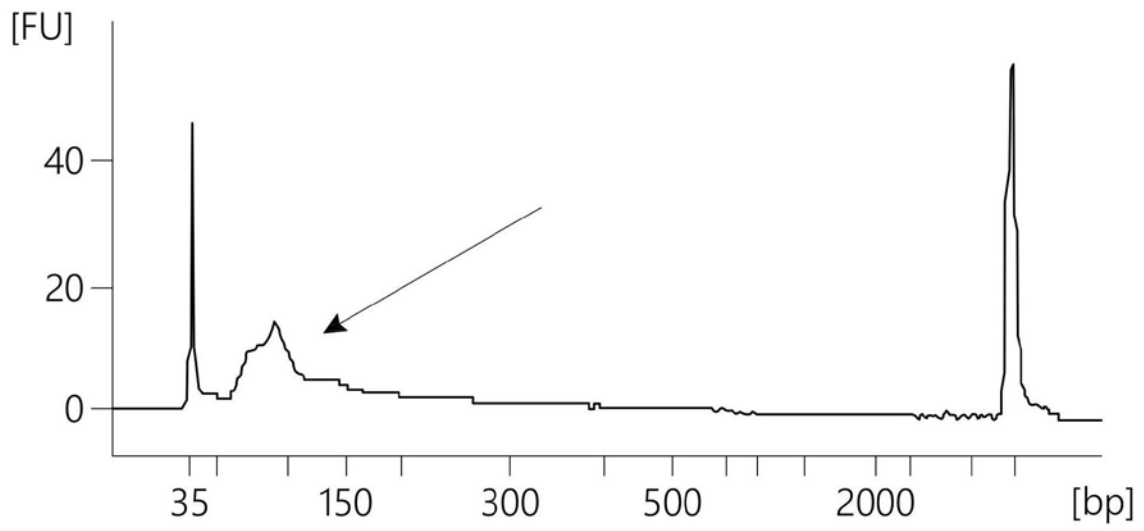


图24E

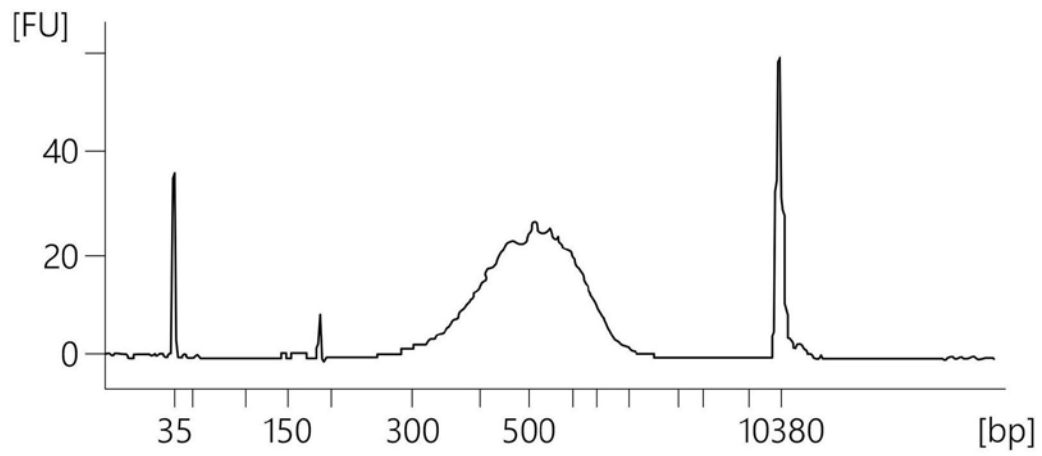


图25A

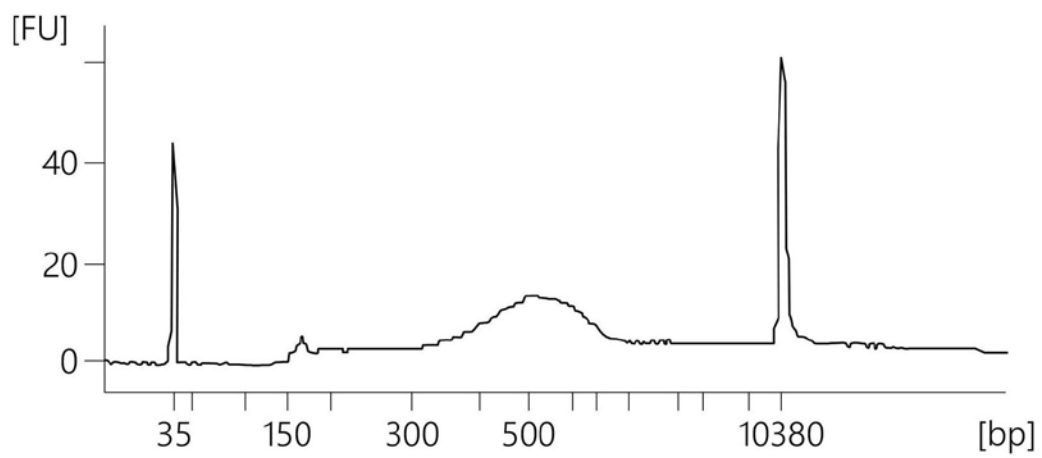


图25B

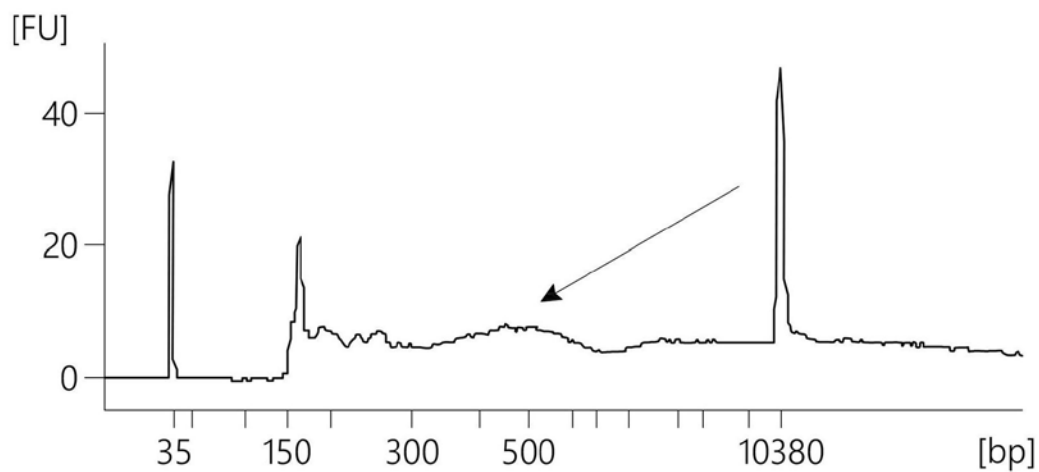


图25C

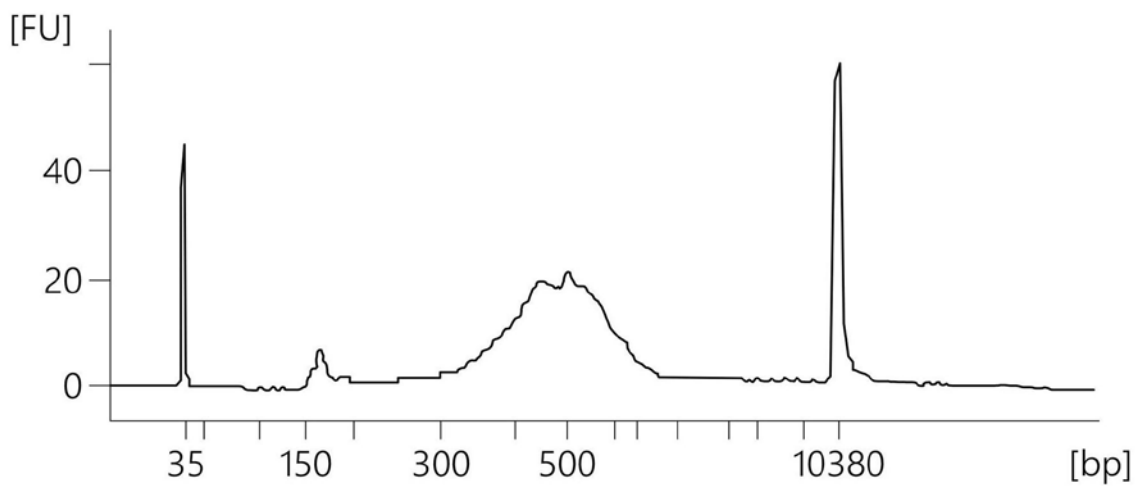


图25D

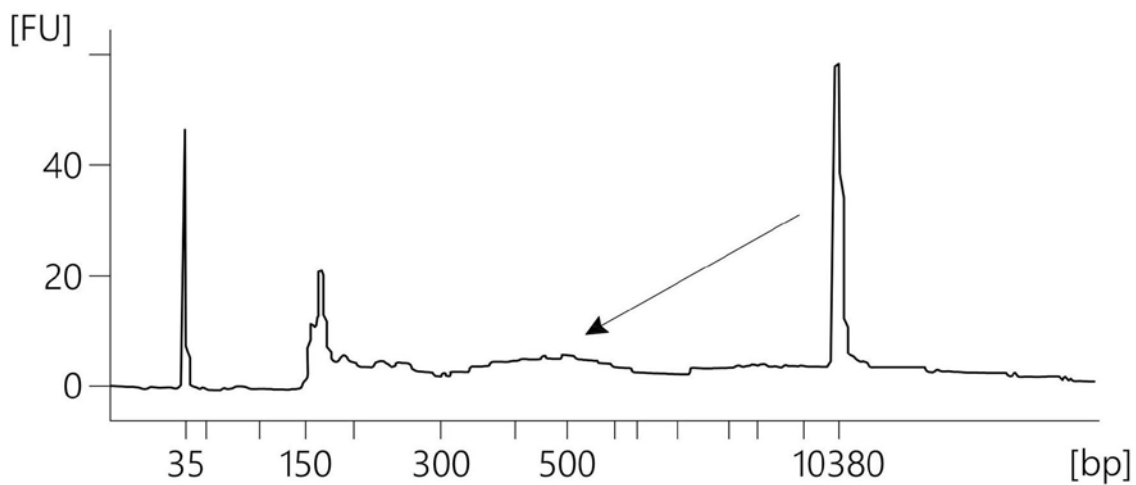


图25E

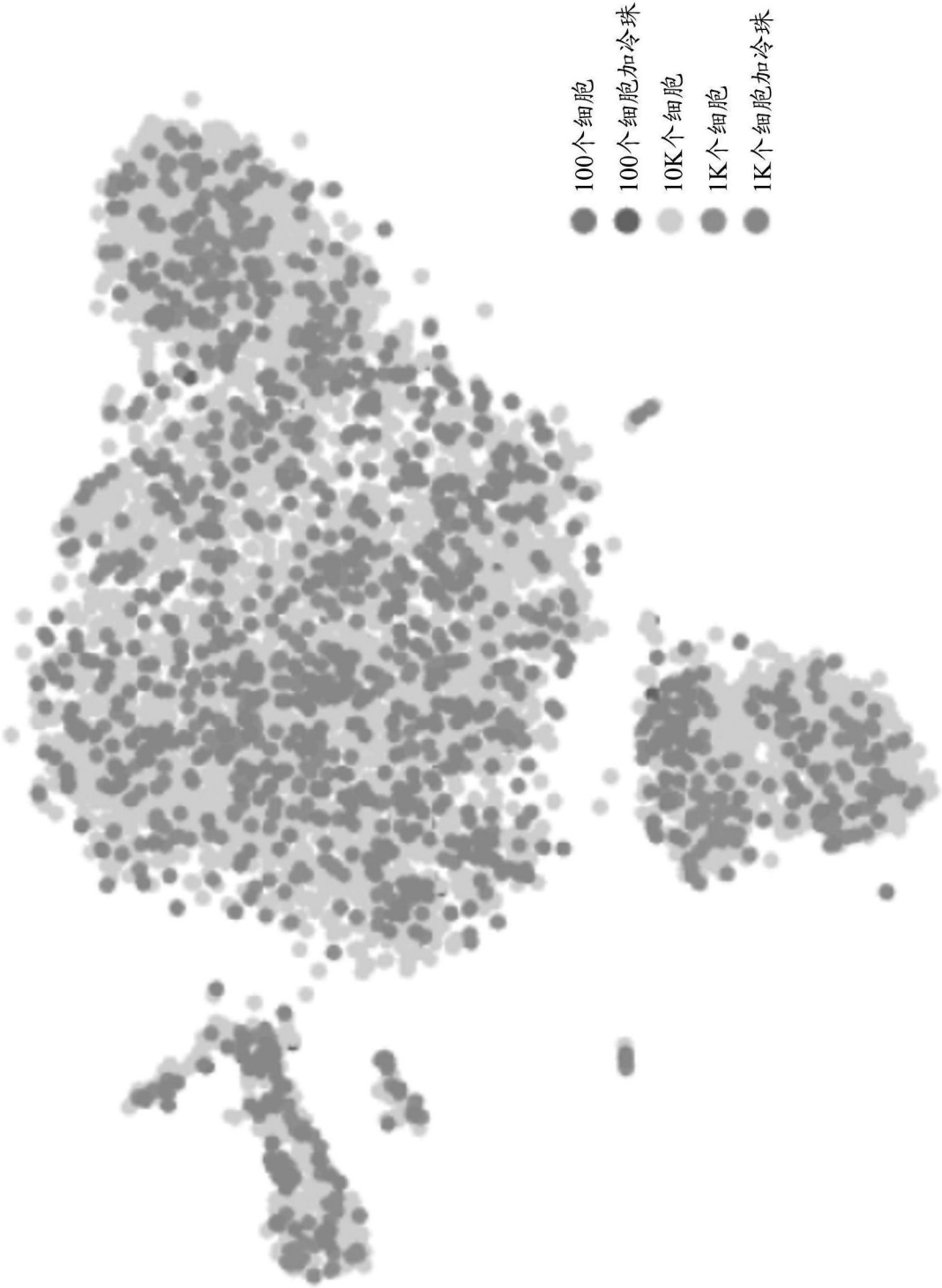


图26A



图26B



图26C



图26D



图26E



图26F

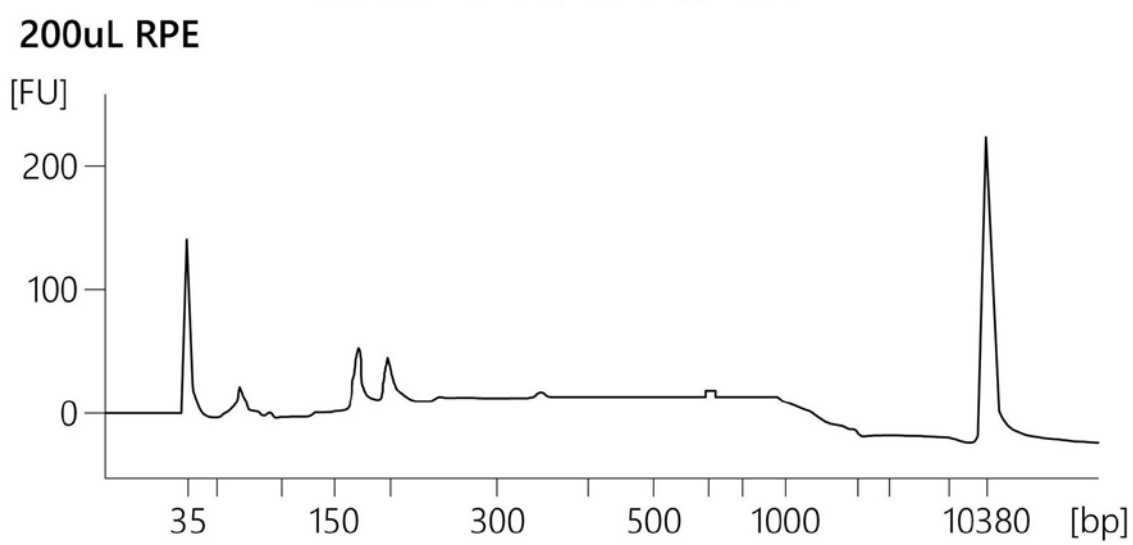
GL ID021-1 第1次变性 1:2

图27A

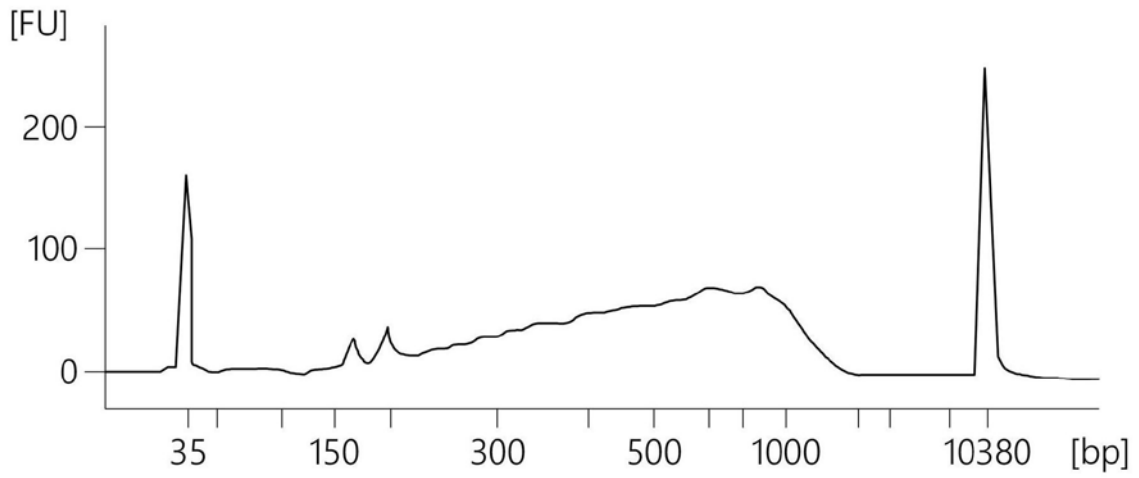
GL ID021-1 第2次变性 1:20**200uL RPE**

图27B

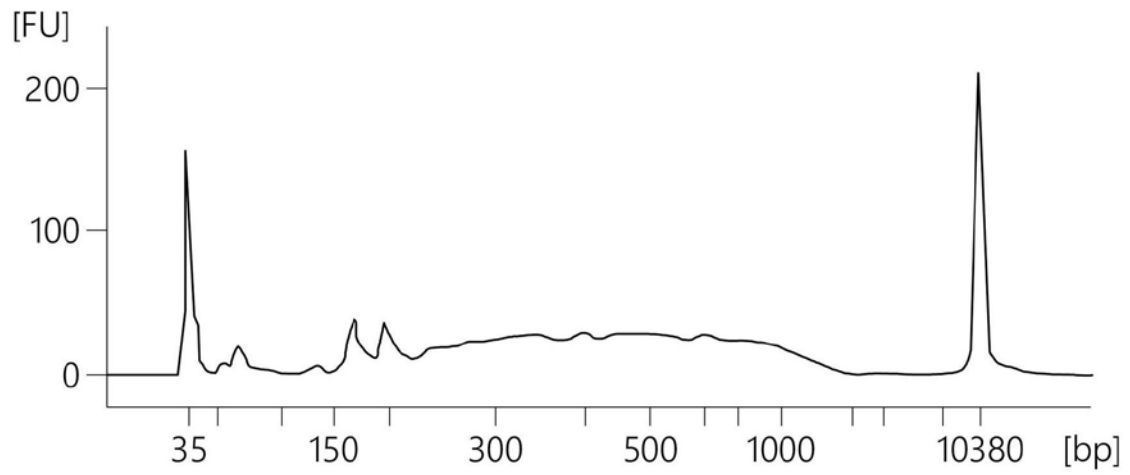
GL ID021-2 第1次变性 1:2**1mL RPE**

图27C

GL ID021-2 第2次变性1:20

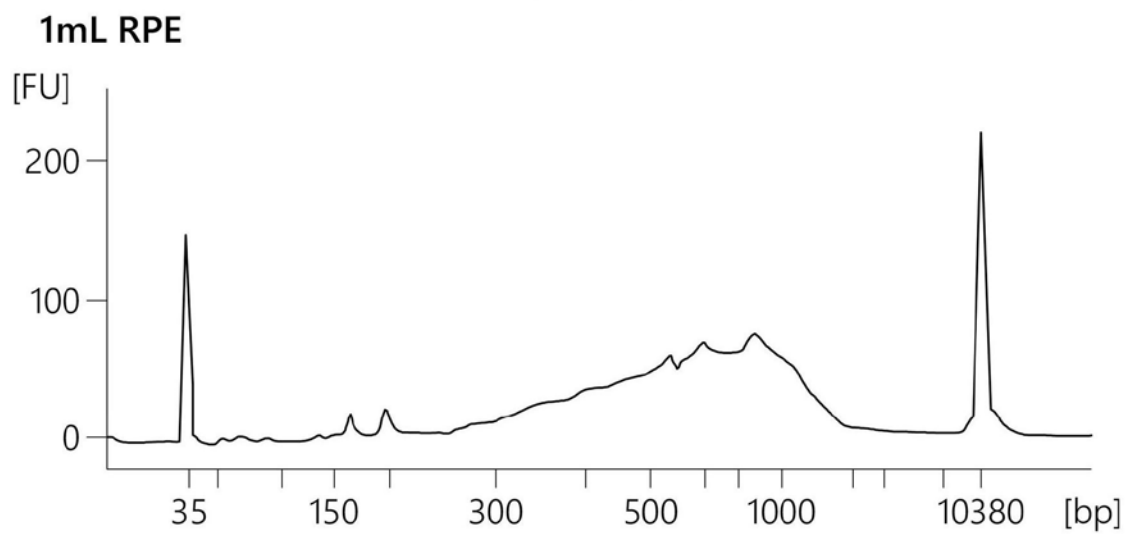


图27D

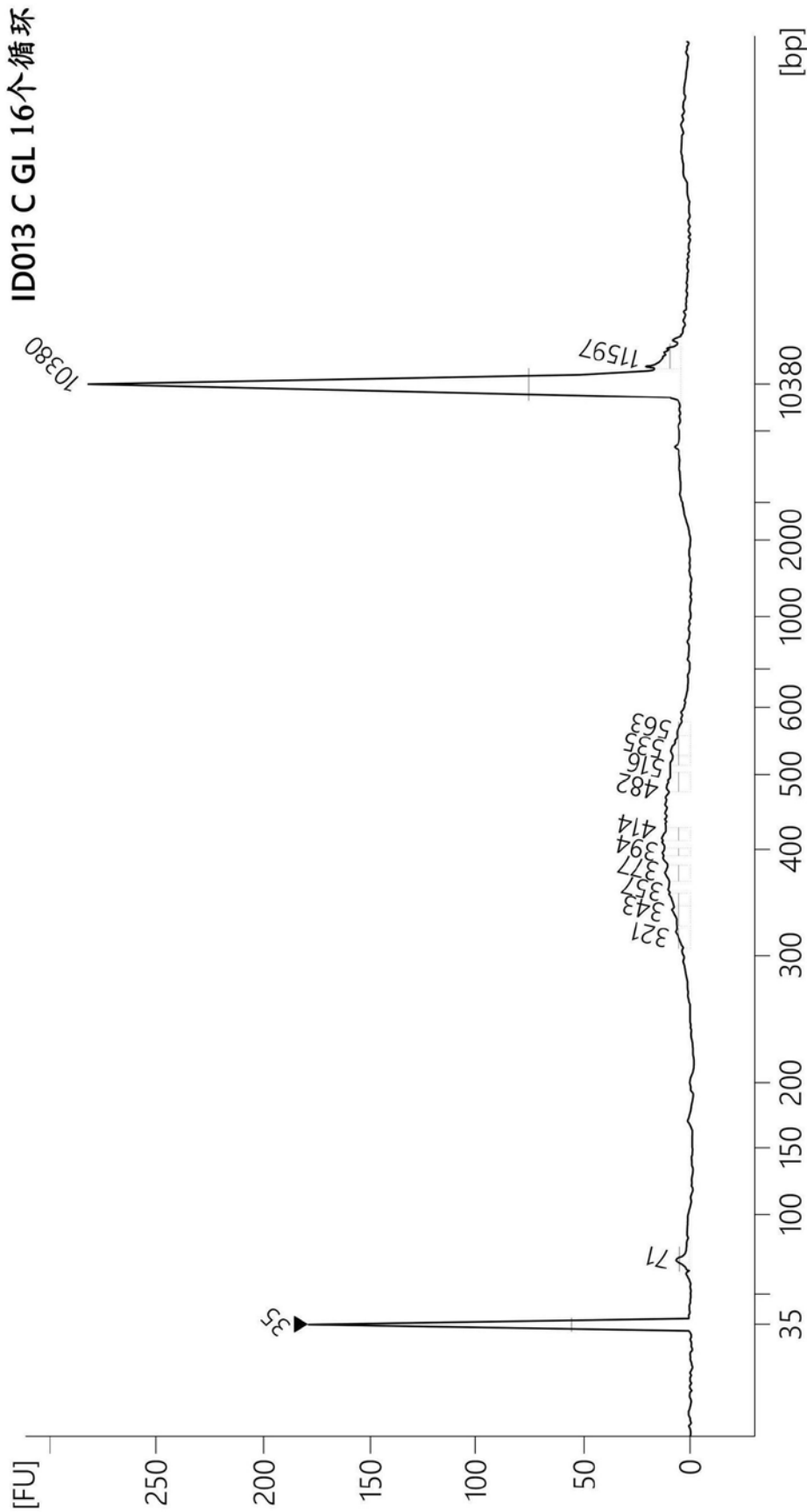


图28A

C PCR2 1:5 [GL R1/R2]

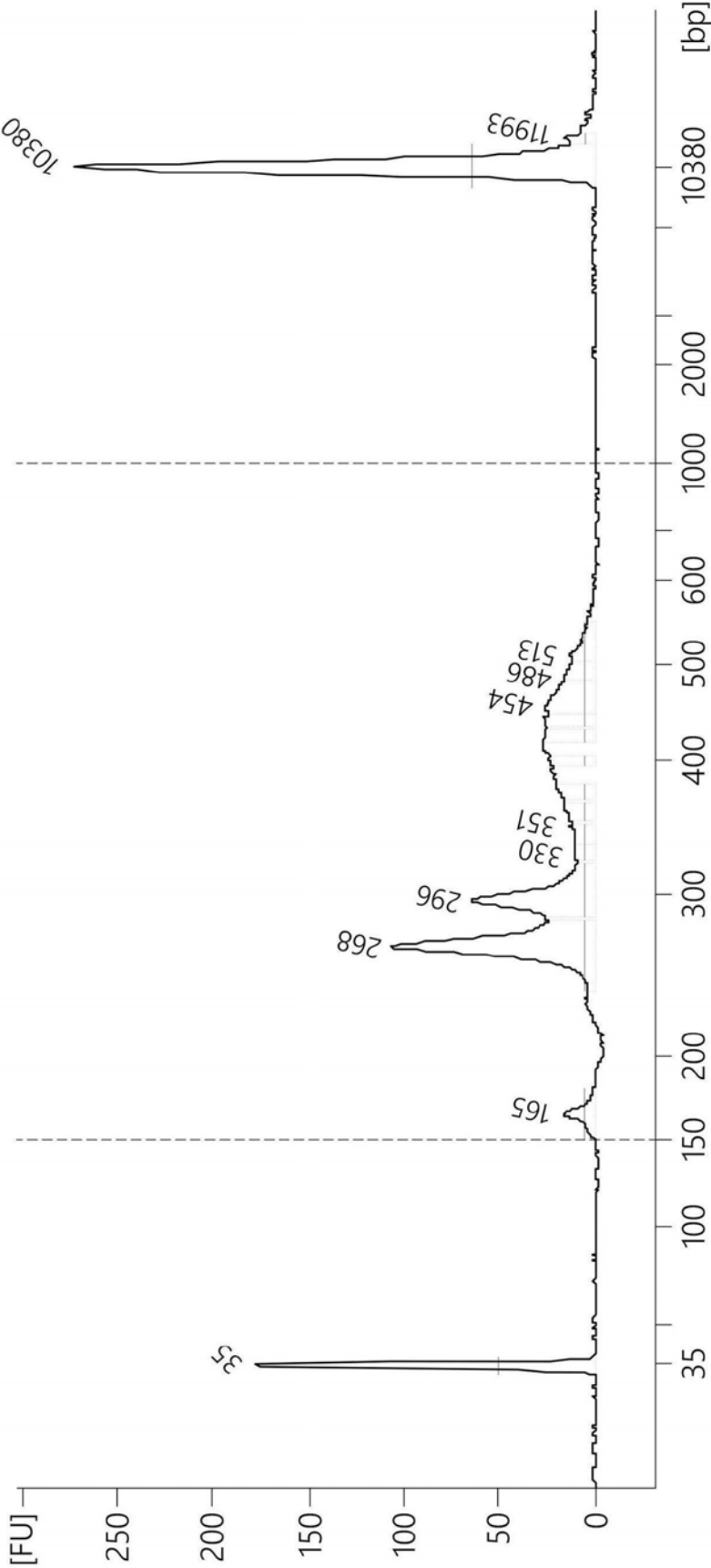


图28B

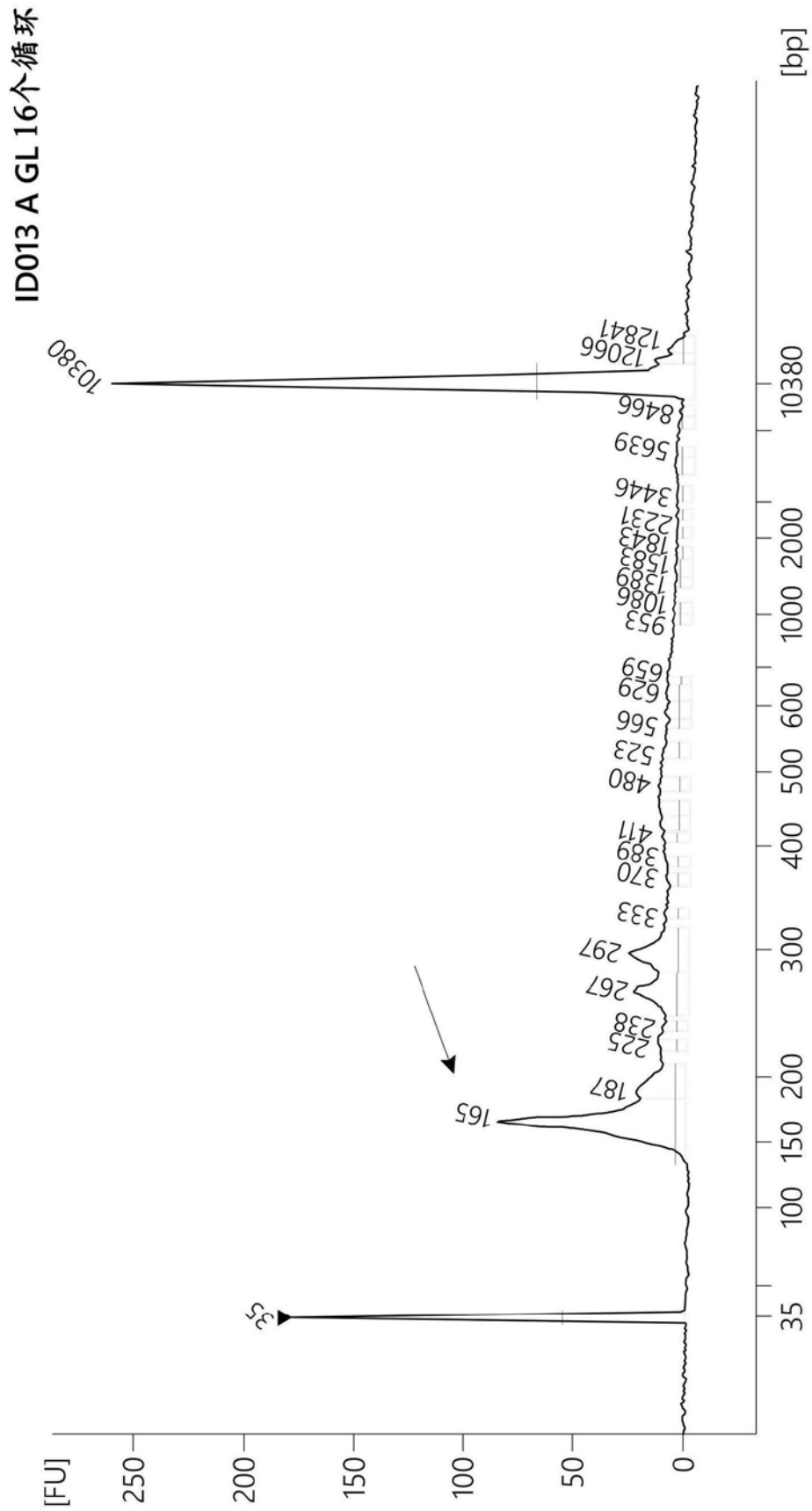


图28C

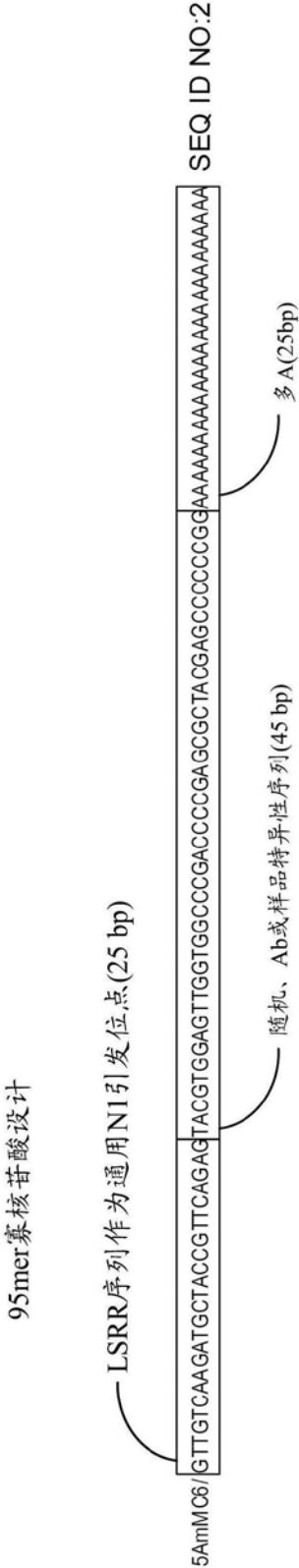


图29A

200mer寡核苷酸设计。靶向扩增方案(巢式特异性)

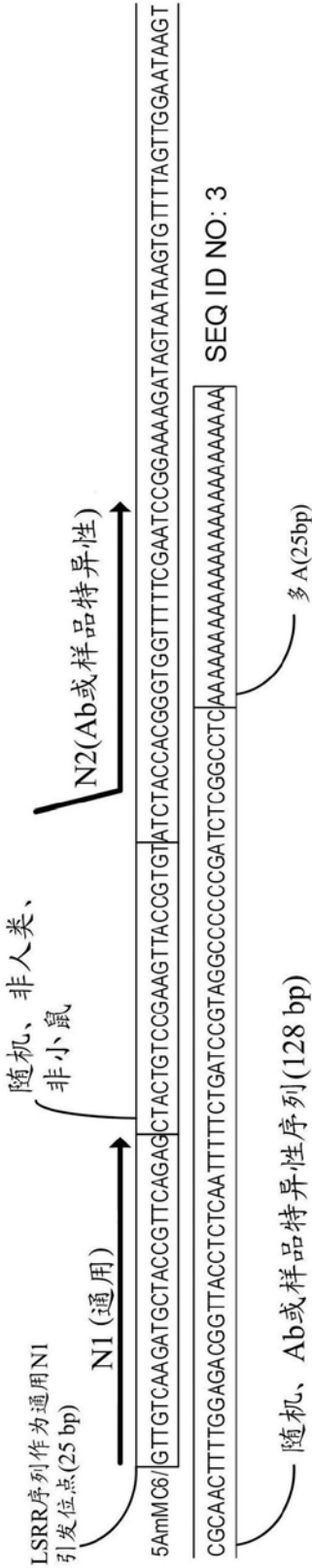


图29B

200mer寡核苷酸设计。靶向扩增方案(巢式通用)

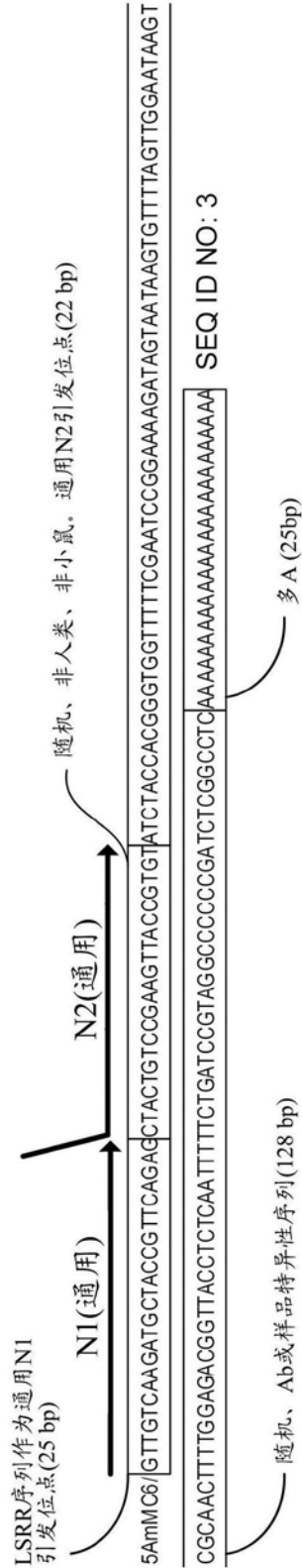


图29C

200mer寡核苷酸设计。靶向扩增方案(一步)

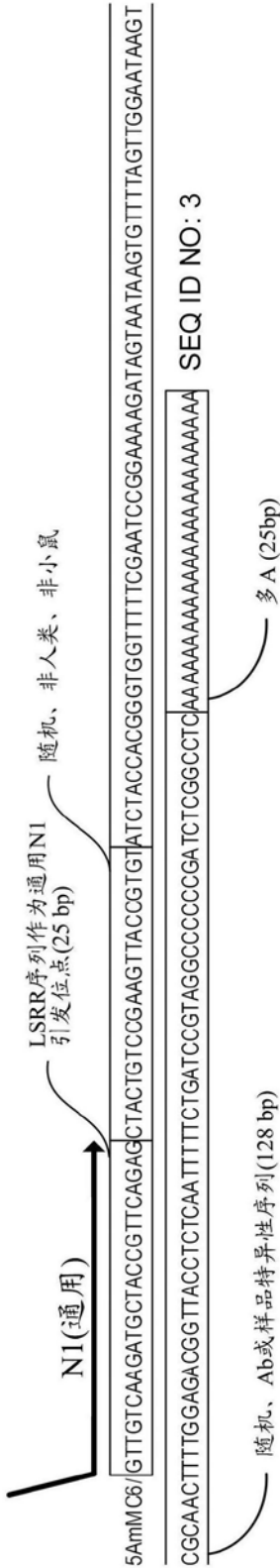


图29D

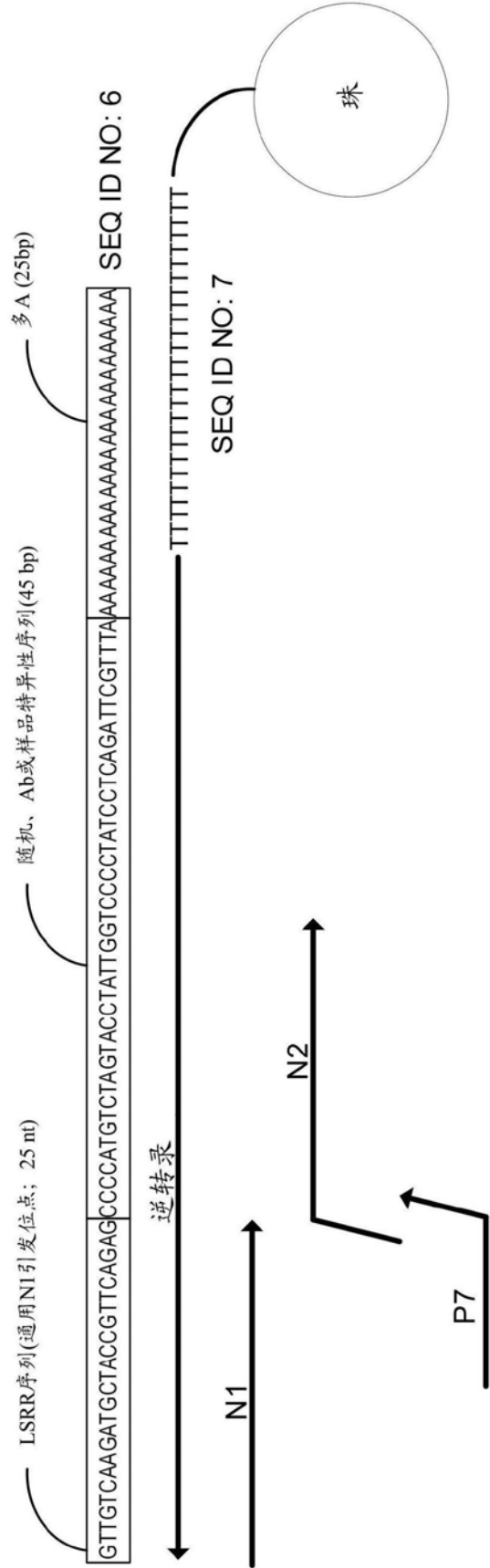


图30