



(12) 发明专利申请

(10) 申请公布号 CN 115279919 A

(43) 申请公布日 2022. 11. 01

(21) 申请号 202180020686.9

(22) 申请日 2021.01.12

(30) 优先权数据

62/960,609 2020.01.13 US

(85) PCT国际申请进入国家阶段日

2022.09.09

(86) PCT国际申请的申请数据

PCT/US2021/013137 2021.01.12

(87) PCT国际申请的公布数据

WO2021/146219 EN 2021.07.22

(71) 申请人 贝克顿迪金森公司

地址 美国新泽西州

(72) 发明人 丹尼斯·普罗森

凯瑟琳·拉扎鲁克

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

专利代理师 刘晓杰 武晶晶

(51) Int.Cl.

C12Q 1/6806 (2006.01)

C12Q 1/6869 (2006.01)

权利要求书12页 说明书101页

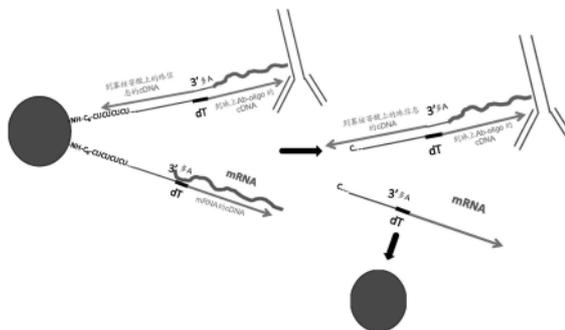
序列表3页 附图33页

(54) 发明名称

使用含有DU的寡核苷酸的细胞捕获

(57) 摘要

本文的公开内容包括用于标记核酸靶的系统、方法、组合物和试剂盒，例如用于使用细胞组分结合试剂特异性寡核苷酸和寡核苷酸条形码同时测量单细胞中的蛋白表达和基因表达。一些实施方案提供了包含裂解区的寡核苷酸条形码。裂解区可以包含一个或更多个裂解位点。在一些实施方案中，裂解剂能够实现裂解区的磷酸二酯主链的裂解。寡核苷酸条形码可以与固体支持物关联。在一些实施方案中还提供了使条形码化核酸分子与固体支持物分离的方法。



1. 一种用于同时测量细胞中蛋白表达和基因表达的方法,所述方法包括:

使多于一种细胞组分结合试剂与包含多于一种细胞组分靶和核酸靶的拷贝的多于一个细胞接触,其中所述核酸靶包括mRNA,其中所述多于一种细胞组分结合试剂中的每一种包括包含用于所述细胞组分结合试剂的独特标识符和多(A)序列的细胞组分结合试剂特异性寡核苷酸,其中所述细胞组分结合试剂特异性寡核苷酸包括DNA,并且其中所述细胞组分结合试剂能够与所述多于一种细胞组分靶中的至少一种特异性结合;

将与所述细胞组分结合试剂关联的多于一个细胞分区到多于一个分区,其中所述多于一个分区中的分区包含来自与所述细胞组分结合试剂关联的多于一个细胞的单细胞;

在包含所述单细胞的分区中,使包含多于一种寡核苷酸条形码的固体支持物与所述细胞组分结合试剂特异性寡核苷酸和所述核酸靶的拷贝接触进行杂交,其中所述寡核苷酸条形码各自包含多(T)序列、第一通用序列、第一分子标记和裂解区,其中所述裂解区位于所述第一通用序列的5'并包含一个或更多个裂解位点;

使与所述多于一种寡核苷酸条形码杂交的所述细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸,所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含所述第一分子标记的互补体、所述裂解区的互补体和所述第一通用序列的互补体;

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条形码化核酸分子,所述多于一种条形码化核酸分子各自包含与所述核酸靶的至少一部分互补的序列和所述第一分子标记;

使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸;

使所述多于一种条形码化核酸分子与所述固体支持物分离以产生多于一种分离的条形码化核酸分子;

获得所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息,以确定所述多于一个细胞中的一个或更多个中所述多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数;以及

获得多于一种分离的条形码化核酸分子或其产物的序列信息,以确定所述多于一个细胞中的一个或更多个中所述核酸靶的拷贝数。

2. 根据权利要求1所述的方法,其中所述延伸的细胞组分结合试剂特异性寡核苷酸各自包含双链裂解区,并且其中所述多于一种条形码化核酸分子各自包含单链裂解区。

3. 根据权利要求2所述的方法,其中使所述多于一种条形码化核酸分子与所述固体支持物分离以产生所述多于一种分离的条形码化核酸分子包括:

使裂解寡核苷酸与所述单链裂解区杂交以产生双链裂解区;

使所述条形码化核酸分子与能够在所述一个或更多个裂解位点处实现所述条形码化核酸分子的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子;以及

将所述多于一种裂解的条形码化核酸分子与所述固体支持物分离以产生所述多于一种分离的条形码化核酸分子。

4. 根据权利要求1-3中任一项所述的方法,其中使所述多于一种条形码化核酸分子与

所述固体支持物分离包括：

使所述条形码化核酸分子与能够在所述一个或更多个裂解位点处实现所述条形码化核酸分子的单链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子；以及

将所述多于一种裂解的条形码化核酸分子与所述固体支持物分离以产生所述多于一种分离的条形码化核酸分子。

5. 根据权利要求1-4中任一项所述的方法，其中使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离包括：

使所述延伸的细胞组分结合试剂特异性寡核苷酸与能够在所述一个或更多个裂解位点处实现所述延伸的细胞组分结合试剂特异性寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸；以及

使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离以产生所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸，

任选地使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离包括使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性。

6. 根据权利要求1-5中任一项所述的方法，其中所述一个或更多个裂解位点包含非典型核苷酸、裂解子识别序列或其任何组合。

7. 根据权利要求1-6中任一项所述的方法，其中所述一个或更多个裂解位点包括第一裂解位点和第二裂解位点，其中所述第一裂解位点包含裂解子识别序列，并且其中所述第二裂解位点包含一个或更多个非典型核苷酸。

8. 根据权利要求1-7中任一项所述的方法，其中所述一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。

9. 根据权利要求1-8中任一项所述的方法，其中所述一种或更多种裂解剂包括双链特异性裂解剂，并且任选地所述双链特异性裂解剂不能够实现对单链裂解区的磷酸二酯主链的裂解，任选地所述双链特异性裂解剂包括能够与所述裂解子识别序列特异性结合并裂解所述双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。

10. 根据权利要求9所述的方法，其中所述双链特异性裂解剂不与裂解子识别序列特异性结合，任选地所述双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。

11. 根据权利要求9-10中任一项所述的方法，其中：

(i) 所述双链特异性裂解剂能够实现所述双链裂解区的两条链的磷酸二酯主链的裂解；或

(ii) 所述双链特异性裂解剂能够实现所述双链裂解区的仅一条链的磷酸二酯主链的裂解。

12. 根据权利要求9-11中任一项所述的方法，其中所述裂解区包含裂解子识别序列和一个或更多个非典型核苷酸，所述方法包括：

使所述延伸的细胞组分结合试剂特异性寡核苷酸与双链特异性裂解剂接触以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸；

使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离；

使所述条形码化核酸分子与一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子；以及

使所述多于一种裂解的条形码化核酸分子与所述固体支持物分离。

13. 根据权利要求9-12中任一项所述的方法，所述方法包括：

使所述多于一种条形码化核酸分子和所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与一种或更多种裂解剂接触以产生：

(i) 未与所述固体支持物关联的多于一种裂解的条形码化核酸分子，和

(ii) 与所述固体支持物关联的多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸；

使所述多于一种裂解的条形码化核酸分子与所述固体支持物分离；

使与所述固体支持物关联的裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性以产生多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸；以及

使所述多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离。

14. 根据权利要求9-13中任一项所述的方法，所述方法包括：

使所述延伸的细胞组分结合试剂特异性寡核苷酸与双链特异性裂解剂接触以产生与所述固体支持物关联的多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸；

使与所述固体支持物关联的所述裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性以产生多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸；

使所述多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离；

使所述条形码化核酸分子与一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子；以及

将所述多于一种裂解的条形码化核酸分子与所述固体支持物分离以产生多于一种分离的条形码化核酸分子。

15. 根据权利要求1-14中任一项所述的方法，其中：

(i) 使所述多于一种条形码化核酸分子与所述固体支持物分离发生在使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离之前；或

(ii) 使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离发生在使所述多于一种条形码化核酸分子与所述固体支持物分离之前。

16. 根据权利要求1-15中任一项所述的方法，所述方法包括使与所述细胞组分结合试剂特异性寡核苷酸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条形码化细胞组分结合试剂特异性寡核苷酸，任选地所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸包括多于一种分离的条形码化细胞组分结合试剂特异性寡核苷酸。

17. 根据权利要求16所述的方法，其中所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸包括多于一种裂解的条形码化细胞组分结合试剂特异性寡核苷酸。

18. 根据权利要求1-17中任一项所述的方法，其中所述多于一种分离的延伸的细胞组

分结合试剂特异性寡核苷酸和所述多于一种分离的条形码化核酸分子在空间上彼此分离。

19. 根据权利要求16-18中任一项所述的方法,其中所述多于一种条形码化细胞组分结合试剂特异性寡核苷酸和所述多于一种分离的条形码化核酸分子在空间上彼此分离。

20. 根据权利要求1-19中任一项所述的方法,其中使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离和/或使所述多于一种条形码化核酸分子与所述固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

21. 根据权利要求1-20中任一项所述的方法,其中所述细胞组分结合试剂特异性寡核苷酸包含第二通用序列。

22. 根据权利要求1-21中任一项所述的方法,其中获得所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息包括:

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,扩增所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸;以及

获得所述多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的测序数据。

23. 根据权利要求1-22中任一项所述的方法,其中所述细胞组分结合试剂特异性寡核苷酸包含第二分子标记,任选地所述多于一种细胞组分结合试剂特异性寡核苷酸中的至少10种包含不同的第二分子标记序列。

24. 根据权利要求23所述的方法,其中:

(i) 至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中所述至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的;或

(ii) 至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中所述至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。

25. 根据权利要求1-24中任一项所述的方法,其中所述测序数据中与用于所述细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目指示所述多于一个细胞中的一个或更多个中所述至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与所述至少一种细胞组分靶特异性结合。

26. 根据权利要求23-25中任一项所述的方法,其中所述测序数据中与用于所述细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示所述多于一个细胞中的一个或更多个中所述至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与所述至少一种细胞组分靶特异性结合。

27. 根据权利要求1-26中任一项所述的方法,其中获得所述序列信息包括将测序衔接子附接到所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物和/或所述多于一种分离的条形码化核酸分子或其产物。

28. 根据权利要求1-27中任一项所述的方法,其中确定所述多于一个细胞中的一个或更多个中所述核酸靶的拷贝数包括基于与所述多于一种分离的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、其互补体或其组合的数目来确定所述多于一个细胞中所述核酸靶的拷贝数。

29. 一种用于样品鉴定的方法,所述方法包括:

使多于一个样品中的每一个分别与多于一种样品索引组合物中的样品索引组合物接触，

其中所述多于一个样品中的每一个包含一个或更多个细胞，每一个细胞包含一种或更多种细胞组分靶，其中所述样品索引组合物包含与样品索引寡核苷酸关联的细胞组分结合试剂，其中所述细胞组分结合试剂能够与所述一种或更多种细胞组分靶中的至少一种特异性结合，

其中所述样品索引寡核苷酸包含样品索引序列，并且其中所述多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列；

使所述多于一种样品索引组合物的所述样品索引寡核苷酸与包含多于一种寡核苷酸条形码的固体支持物接触，其中每一种寡核苷酸条形码包含第一分子标记、能够与所述样品索引寡核苷酸杂交的靶结合区和裂解区，其中所述裂解区包含一个或更多个裂解位点；

使与所述多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸以产生多于一种延伸的样品索引寡核苷酸，所述多于一种延伸的样品索引寡核苷酸各自包含所述第一分子标记的互补体和所述裂解区的互补体；

使所述多于一种延伸的样品索引寡核苷酸与所述固体支持物分离以产生多于一种分离的延伸的样品索引寡核苷酸；

获得所述多于一种分离的延伸的样品索引寡核苷酸或其产物的测序数据；以及

基于所述测序数据中所述多于一种分离的延伸的样品索引寡核苷酸或其产物中的至少一种分离的延伸的样品索引寡核苷酸或其产物的样品索引序列，鉴定所述多于一个细胞中的至少一个细胞的样品来源，

任选地使所述多于一种延伸的样品索引寡核苷酸与所述固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

30. 根据权利要求29所述的方法，其中每一种寡核苷酸条形码包含第一通用序列，任选地所述多于一种延伸的样品索引寡核苷酸包含所述第一通用序列的互补体。

31. 根据权利要求29-30中任一项所述的方法，其中所述样品索引寡核苷酸包含第二分子标记，任选地：

(i) 至少两种样品索引寡核苷酸的第二分子标记序列是不同的，并且其中所述至少两种样品索引寡核苷酸的样品索引序列是相同的；或

(ii) 至少两种样品索引寡核苷酸的第二分子标记序列是不同的，并且其中所述至少两种样品索引寡核苷酸的样品索引序列是不同的。

32. 根据权利要求29-31中任一项所述的方法，其中所述多于一种样品索引组合物中的至少10种、100种或1000种样品索引组合物中的样品索引序列包含不同的序列。

33. 根据权利要求29-32中任一项所述的方法，其中所述样品索引寡核苷酸包含第二通用序列。

34. 根据权利要求29-33中任一项所述的方法，其中鉴定所述至少一个细胞的样品来源包括鉴定所述测序数据中至少一种分离的延伸的样品索引寡核苷酸或其产物中样品索引序列的存在或不存在。

35. 根据权利要求34所述的方法，其中鉴定所述样品索引序列的存在或不存在包括：

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,扩增所述至少一种分离的延伸的样品索引寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的样品索引寡核苷酸;

获得所述多于一种扩增的分离的延伸的样品索引寡核苷酸或其产物的测序数据;以及

基于所述测序数据中所述多于一种扩增的分离的延伸的样品索引寡核苷酸或其产物中对应于所述至少一种分离的延伸的样品索引寡核苷酸或其产物的扩增的分离的延伸的样品索引寡核苷酸或其产物的样品索引序列来鉴定细胞的样品来源,

任选地扩增所述至少一种分离的延伸的样品索引寡核苷酸或其产物包括将测序衔接子附接到所述多于一种分离的延伸的样品索引寡核苷酸或其产物。

36. 根据权利要求29-35中任一项所述的方法,其中使与所述多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸包括使用逆转录酶使与所述多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸。

37. 根据权利要求29-36中任一项所述的方法,其中使与所述多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸包括使用缺乏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寡核苷酸条形码杂交的样品索引寡核苷酸延伸,任选地所述DNA聚合酶包括Klenow片段。

38. 根据权利要求29-37中任一项所述的方法,所述方法包括去除所述多于一种样品索引组合物中未结合的样品索引组合物。

39. 根据权利要求38所述的方法,其中去除所述未结合的样品索引组合物包括

(i) 用洗涤缓冲液洗涤来自所述多于一个样品中的每一个的所述一个或更多个细胞;和/或

(ii) 使用流式细胞术选择与至少一种细胞组分结合试剂结合的细胞。

40. 根据权利要求29-39中任一项所述的方法,其中:

(i) 所述样品索引寡核苷酸被配置为能够从所述细胞组分结合试剂脱离;和/或

(ii) 所述样品索引寡核苷酸通过接头与所述细胞组分结合试剂关联。

41. 根据权利要求29-40中任一项所述的方法,所述方法包括使所述样品索引寡核苷酸与所述细胞组分结合试剂解离。

42. 根据权利要求29-41中任一项所述的方法,其中所述靶结合区包含多(dT)区,任选地所述样品索引寡核苷酸包含多(dA)区,还任选地所述样品索引寡核苷酸包含与所述多(dA)区相邻的对齐序列。

43. 根据权利要求29-42中任一项所述的方法,其中所述多于一个样品中的样品包括多于一个细胞、多于一个单细胞、组织、肿瘤样品或其任何组合。

44. 根据权利要求29-43中任一项所述的方法,其中使所述多于一种延伸的样品索引寡核苷酸与所述固体支持物分离包括:使所述延伸的样品索引寡核苷酸与一种或更多种裂解剂接触。

45. 根据权利要求29-44中任一项所述的方法,其中所述延伸的细胞组分结合试剂特异性寡核苷酸各自包含双链裂解区。

46. 根据权利要求29-45中任一项所述的方法,其中使所述多于一种延伸的样品索引寡核苷酸与所述固体支持物分离包括:

使所述延伸的样品索引寡核苷酸与一种或更多种裂解剂接触以产生多于一种裂解的

延伸的样品索引寡核苷酸,所述一种或更多种裂解剂能够在所述一个或更多个裂解位点处实现所述延伸的样品索引寡核苷酸的双链裂解区的磷酸二酯主链的裂解;以及

使所述多于一种裂解的延伸的样品索引寡核苷酸与所述固体支持物分离以产生所述多于一种分离的延伸的样品索引寡核苷酸,任选地使所述多于一种裂解的延伸的样品索引寡核苷酸与所述固体支持物分离包括使所述多于一种裂解的延伸的样品索引寡核苷酸变性。

47. 一种用于测量细胞中细胞组分表达的方法,所述方法包括:

使多于一种细胞组分结合试剂与包含多于一种细胞组分靶的多于一个细胞接触,其中所述多于一种细胞组分结合试剂中的每一种包含细胞组分结合试剂特异性寡核苷酸,所述细胞组分结合试剂特异性寡核苷酸包含用于所述细胞组分结合试剂的独特标识符序列,并且其中所述细胞组分结合试剂能够与所述多于一种细胞组分靶中的至少一种特异性结合;

使所述细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触,其中每一种寡核苷酸条形码包含第一分子标记、能够与所述细胞组分结合试剂特异性寡核苷酸杂交的靶结合区和裂解区,其中所述裂解区包含一个或更多个裂解位点;

使与所述多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸,所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含所述第一分子标记的互补体和所述裂解区的互补体;

使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸;

获得所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息,以确定所述多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。

48. 根据权利要求47所述的方法,所述方法包括在延伸与所述多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸之前:

将与所述多于一种细胞组分结合试剂关联的多于一个细胞分区到多于一个分区,其中所述多于一个分区中的分区包含来自与所述细胞组分结合试剂关联的多于一个细胞的单细胞;

在包含所述单细胞的分区中,使所述细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触,任选地所述多于一个分区中的分区包含单个固体支持物。

49. 根据权利要求47-48中任一项所述的方法,其中使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

50. 根据权利要求47-49中任一项所述的方法,其中每一种寡核苷酸条形码包含第一通用序列,任选地所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸包含所述第一通用序列的互补体。

51. 根据权利要求47-50中任一项所述的方法,其中所述细胞组分结合试剂特异性寡核苷酸包含第二通用序列。

52. 根据权利要求47-51中任一项所述的方法,其中获得所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息包括:

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,扩增所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸;以及

获得所述多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的测序数据。

53. 根据权利要求47-52中任一项所述的方法,其中所述细胞组分结合试剂特异性寡核苷酸包含第二分子标记,任选地所述多于一种细胞组分结合试剂特异性寡核苷酸中的至少10种包含不同的第二分子标记序列。

54. 根据权利要求53所述的方法,其中:

(i) 至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中所述至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的;或

(ii) 至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中所述至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。

55. 根据权利要求1-54中任一项所述的方法,其中所述测序数据中与用于所述细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目指示所述多于一个细胞中的一个或多个中所述至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与所述至少一种细胞组分靶特异性结合。

56. 根据权利要求47-55中任一项所述的方法,其中所述测序数据中与用于所述细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示所述多于一个细胞中的一个或多个中所述至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与所述至少一种细胞组分靶特异性结合。

57. 根据权利要求47-56中任一项所述的方法,其中获得所述序列信息包括将测序衔接子附接到所述多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物。

58. 根据权利要求47-57中任一项所述的方法,其中使与所述多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用逆转录酶使与所述多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸。

59. 根据权利要求47-58中任一项所述的方法,其中使与所述多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中至少一种的DNA聚合酶使与所述多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸,任选地所述DNA聚合酶包括Klenow片段。

60. 根据权利要求47-59中任一项所述的方法,其中所述靶结合区包含多(dT)区,任选地所述细胞组分结合试剂特异性寡核苷酸包含多(dA)区,还任选地所述细胞组分结合试剂特异性寡核苷酸包含与所述多(dA)区相邻的对齐序列。

61. 根据权利要求47-60中任一项所述的方法,其中:

(i) 所述细胞组分结合试剂特异性寡核苷酸被配置为能够从细胞组分结合试剂脱离;和/或

(ii) 所述细胞组分结合试剂特异性寡核苷酸通过接头与所述细胞组分结合试剂关联。

62. 根据权利要求47-61中任一项所述的方法,所述方法包括使所述细胞组分结合试剂特异性寡核苷酸与所述细胞组分结合试剂解离。

63. 根据权利要求48-62中任一项所述的方法,其中将所述多于一个细胞分区包括将与所述多于一种细胞组分结合试剂关联的多于一个细胞和包含所述固体支持物的多于一个固体支持物分区到所述多于一个分区,其中所述多于一个分区中的分区包含来自与所述细胞组分结合试剂关联的多于一个细胞的单细胞和所述固体支持物,任选地所述分区是孔或液滴。

64. 根据权利要求47-63中任一项所述的方法,所述方法包括在使所述多于一种细胞组分结合试剂与所述多于一个细胞接触之后,去除所述多于一种细胞组分结合试剂中未与所述多于一个细胞接触的一种或更多种细胞组分结合试剂,任选地去除未与所述多于一个细胞接触的一种或更多种细胞组分结合试剂包括:去除未与对应的所述多于一种细胞组分靶中的至少一种靶接触的一种或更多种细胞组分结合试剂。

65. 根据权利要求47-64中任一项所述的方法,其中所述多于一个细胞包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。

66. 根据权利要求47-65中任一项所述的方法,其中使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离包括:使所述延伸的细胞组分结合试剂特异性寡核苷酸与一种或更多种裂解剂接触。

67. 根据权利要求47-66中任一项所述的方法,其中所述延伸的细胞组分结合试剂特异性寡核苷酸各自包含双链裂解区。

68. 根据权利要求47-67中任一项所述的方法,其中使所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离包括:

使所述延伸的细胞组分结合试剂特异性寡核苷酸与能够在所述一个或更多个裂解位点处实现所述延伸的细胞组分结合试剂特异性寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;以及

使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸,任选地使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与所述固体支持物分离包括使所述多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性。

69. 根据权利要求1-68中任一项所述的方法,其中所述一个或更多个裂解位点包含非典型核苷酸、裂解子识别序列或其任何组合。

70. 根据权利要求1-69中任一项所述的方法,其中所述一个或更多个裂解位点包括第一裂解位点和第二裂解位点,其中所述第一裂解位点包含裂解子识别序列,并且其中所述第二裂解位点包含一个或更多个非典型核苷酸。

71. 根据权利要求1-70中任一项所述的方法,其中所述非典型核苷酸包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1,N6-亚乙烯基腺嘌呤、8-氧-鸟嘌呤(8oxoG)或其任何组合。

72. 根据权利要求1-71中任一项所述的方法,其中所述一种或更多种裂解剂包括DNA糖苷酶、AP裂解剂或其任何组合,任选地所述DNA糖苷酶能够裂解所述非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶(AP)位点。

73. 根据权利要求72所述的方法,其中所述AP裂解剂能够裂解与所述AP位点相邻的磷酸二酯主链。

74. 根据权利要求72-73中任一项所述的方法,其中所述AP裂解剂包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合,任选地所述胺包括N,N'-二甲基乙二胺(DMED)。

75. 根据权利要求74所述的方法,其中所述AP裂解酶包括AP裂合酶、AP核酸内切酶或其任何组合。

76. 根据权利要求1-75中任一项所述的方法,其中所述一种或更多种裂解剂包括APE 1 (AP核酸内切酶1)、Endo III (核酸内切酶III)、Endo IV (核酸内切酶IV)、Endo V (核酸内切酶V)、Endo VIII (核酸内切酶VIII)、Fpg (甲酰氨基-嘧啶-DNA糖苷酶)、OGG1 (8-氧鸟嘌呤DNA糖苷酶1)、NEIL1 (核酸内切酶VIII样1)、T7 Endo I (T7核酸内切酶I)、T4 PDG (T4嘧啶二聚体DNA糖苷酶)、UDG (尿嘧啶DNA糖苷酶)、SMUG1 (单链选择性单功能尿嘧啶DNA糖苷酶)、AAG (甲基嘌呤DNA糖苷酶)或其任何组合。

77. 根据权利要求1-76中任一项所述的方法,其中所述一种或更多种裂解剂是热稳定的。

78. 根据权利要求1-77中任一项所述的方法,其中所述一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。

79. 根据权利要求78所述的方法,其中所述裂解子识别序列包含限制性位点,并且所述靶特异性裂解子包括限制性核酸内切酶。

80. 根据权利要求1-79中任一项所述的方法,其中所述靶特异性裂解子包括:

(i) TAL-效应子核酸酶(TALEN),其被配置为与所述裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链;和/或

(ii) 锌指核酸酶(ZFN),其被配置为与所述裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。

81. 根据权利要求1-80中任一项所述的方法,其中所述靶特异性裂解子包括CRISPR复合体,所述CRISPR复合体包括与互补于所述裂解子识别序列的引导RNA(gRNA)复合的CRISPR酶,任选地所述裂解子识别序列包含前间区邻近基序(PAM)位点。

82. 根据权利要求81所述的方法,其中所述CRISPR酶包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。

83. 根据权利要求1-82中任一项所述的方法,其中所述一种或更多种裂解剂包括双链特异性裂解剂,任选地所述双链特异性裂解剂不能实现单链裂解区的磷酸二酯主链的裂解。

84. 根据权利要求1-83中任一项所述的方法,其中所述双链特异性裂解剂包括能够与所述裂解子识别序列特异性结合并裂解所述双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。

85. 根据权利要求1-84中任一项所述的方法,其中所述双链特异性裂解剂不与裂解子识别序列特异性结合,任选地所述双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。

86. 根据权利要求1-85中任一项所述的方法,其中:

(i) 所述双链特异性裂解剂能够实现所述双链裂解区的两条链的磷酸二酯主链的裂解;或

(ii) 所述双链特异性裂解剂能够实现所述双链裂解区的仅一条链的磷酸二酯主链的裂解。

87. 根据权利要求1-86中任一项所述的方法,其中变性包括加热和/或碱变性。

88. 根据权利要求1-87中任一项所述的方法,其中:

(i) 所述第一通用序列、所述第二通用序列和/或所述第三通用序列是相同的;和/或

(ii) 所述第一通用序列、所述第二通用序列和/或所述第三通用序列是不同的。

89. 根据权利要求1-88中任一项所述的方法,其中所述第一通用序列、所述第二通用序列、所述第三通用序列包含测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分,任选地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,还任选地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。

90. 根据权利要求42-89中任一项所述的方法,其中

(a) 所述对齐序列包含鸟嘌呤、胞嘧啶、胸腺嘧啶、尿嘧啶或其组合;

(b) 所述对齐序列包含多(dT)序列、多(dG)序列、多(dC)序列、多(dU)序列或其组合;

(c) 所述对齐序列位于所述多(dA)区的5';和/或

(d) 所述对齐序列的长度为一个或更多个核苷酸,或长度为两个或更多个核苷酸。

91. 根据权利要求1-90中任一项所述的方法,其中所述接头包含碳链,任选地所述碳链包含2-30个碳原子,并且还任选地所述碳链包含12个碳原子,任选地所述接头包含5'氨基修饰物C12(5AmMC12)或其衍生物。

92. 根据权利要求1-91中任一项所述的方法,其中所述细胞组分靶包括碳水化合物、脂质、蛋白、细胞外蛋白、细胞表面蛋白、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、细胞内蛋白或其任何组合,任选地所述细胞组分靶在细胞表面上。

93. 根据权利要求1-92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种包含不同的第一分子标记序列。

94. 根据权利要求1-93中任一项所述的方法,其中所述多于一种寡核苷酸条形码与固体支持物关联,任选地所述固体支持物包括合成颗粒和/或平坦表面。

95. 根据权利要求1-94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,任选地所述多于一种寡核苷酸条形码的每一种细胞标记包含至少6个核苷酸,还任选地与同一固体支持物关联的寡核苷酸条形码包含相同的细胞标记,任选地与不同固体支持物关联的寡核苷酸条形码包含不同的细胞标记。

96. 根据权利要求1-95中任一项所述的方法,其中所述样品包括单细胞,所述方法包括将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,任选地所述合成颗粒和所述单细胞在同一分区中,还任选地所述分区是孔或液滴。

97. 根据权利要求1-96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

至少一种被固定在所述合成颗粒上、部分地固定在所述合成颗粒上、被包封在所述合成颗粒中或部分地包封在所述合成颗粒中,任选地所述合成颗粒是可破坏的。

98. 根据权利要求1-97中任一项所述的方法,其中所述合成颗粒包括珠,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(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠或其任何组合。

99. 根据权利要求1-98中任一项所述的方法,其中所述合成颗粒包含选自由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、胶乳、琼脂糖凝胶、纤维素、尼龙、硅酮及其任何组合,任选地所述合成颗粒包括可破坏的水凝胶颗粒。

使用含有DU的寡核苷酸的细胞捕获

[0001] 相关申请

[0002] 本申请根据35U.S.C.§119(e) 要求2020年1月13日提交的美国临时专利申请序列第62/960609号的权益,此相关申请的内容为了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。序列表被提供为题为SeqListing_68EB_298710_W0的文件,创建于2021年1月11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来鉴定不同样品的细胞并确定细胞中的蛋白表达谱。

[0008] 对现有技术的描述

[0009] 当前的技术允许在每个细胞与条形码化试剂珠在隔室(compartment)中共定位时通过将细胞特异性寡核苷酸条形码附接至来自个体细胞的多(A) mRNA分子而以大规模并行的方式(例如,>10000个细胞)测量单细胞的基因表达。为了对得到的条形码mRNA内容进行测序,在存在珠的初始PCR中对珠进行扩增。PCR反应中磁珠的存在会抑制PCR扩增。对增加初始PCR扩增效率和产量的方法和组合物存在需求。此外,基因表达可以影响蛋白表达。蛋白-蛋白相互作用可以影响基因表达和蛋白表达。对能够定量分析细胞中蛋白表达以及同时测量细胞中蛋白表达和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同时测量细胞中的蛋白表达和基因表达的方法。在一些实施方案中,该方法包括:使多于一种细胞组分结合试剂与包含多于一种细胞组分靶和核酸靶的拷贝的多于一个细胞接触,其中核酸靶包括mRNA,其中多于一种细胞组分结合试剂中的每一种包括包含用于细胞组分结合试剂的独特标识符和多(A)序列的细胞组分结合试剂特异性寡核苷酸,其中细胞组分结合试剂特异性寡核苷酸包括DNA,并且其中细胞组分结合试剂能够与多于一种细胞组分靶中的至少一种特异性结合。该方法可以包括:将与细胞组分结合试剂关联的多于一个细胞分区到多于一个分区,其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞。该方法可以包括:在包含单细胞的分区中,使包含多于一种寡核苷酸条形码的固体支持物与细胞组分结合试剂特异性寡核苷酸和核酸靶的拷贝的接触进行杂交,其中寡核苷酸条形码各自包含多(T)序列、第一通用序列、第一分子标记和裂解区,其中裂解区位于第一通用序列的5'并包含一个或更多个裂解位点。该方法可以包括:使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸,所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含第一分子标记的互补体、裂解区的互补体和第一通用序列的互补体。该方法可以包括: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条形码化核酸分子,所述多于一种条形码化核酸分子各自包含与

核酸靶的至少一部分互补的序列和第一分子标记。该方法可以包括：使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。该方法可以包括：使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。该方法可以包括：获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息，以确定多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。该方法可以包括：获得多于一种分离的条形码化核酸分子或其产物的序列信息，以确定多于一个细胞中的一个或更多个中核酸靶的拷贝数。

[0012] 在一些实施方案中，延伸的细胞组分结合试剂特异性寡核苷酸各自包含双链裂解区，并且其中多于一种条形码化核酸分子各自包含单链裂解区。在一些实施方案中，使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子包括：使裂解寡核苷酸与单链裂解区杂交以产生双链裂解区；使条形码化核酸分子与能够在在一个或更多个裂解位点处实现条形码化核酸分子的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触，以产生多于一种裂解的条形码化核酸分子；以及使多于一种裂解的条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。在一些实施方案中，使多于一种条形码化核酸分子与固体支持物分离包括：使条形码化核酸分子与能够在在一个或更多个裂解位点处实现条形码化核酸分子的单链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触，以产生多于一种裂解的条形码化核酸分子；以及使多于一种裂解的条形码化核酸分子与固体支持物分离，以产生多于一种分离的条形码化核酸分子。在一些实施方案中，使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括：使延伸的细胞组分结合试剂特异性寡核苷酸与能够在在一个或更多个裂解位点处实现延伸的细胞组分结合试剂特异性寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触，以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸；以及使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。在一些实施方案中，使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性。

[0013] 在一些实施方案中，一个或更多个裂解位点包含非典型 (non-canonical) 核苷酸、裂解子识别序列 (cleaver recognition sequence) 或其任何组合。在一些实施方案中，一个或更多个裂解位点包括第一裂解位点和第二裂解位点，其中第一裂解位点包含裂解子识别序列，并且其中第二裂解位点包含一个或更多个非典型核苷酸。在一些实施方案中，非典型核苷酸包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1,N6-亚乙烯基腺嘌呤或其任何组合。在一些实施方案中，非典型核苷酸包括8-氧代鸟嘌呤 (8oxoG)。在一些实施方案中，一种或更多种裂解剂包括DNA糖苷酶、AP裂解剂或其任何组合。在一些实施方案中，DNA糖苷酶能够裂解非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶 (apurinic/apyrimidinic, AP) 位点。在一些实施方案中，AP裂解剂能够裂解与AP位点相邻的磷酸二酯主链。在一些实施方案中，AP裂解剂包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合。在一些实施方案中，胺包括N,N'-二甲基乙二胺 (DMED)。在一些实施方

案中,AP裂解酶包括AP裂合酶、AP核酸内切酶或其任何组合。在一些实施方案中,一种或更多种裂解剂包括APE 1 (AP核酸内切酶1)、Endo III (核酸内切酶III)、Endo IV (核酸内切酶IV)、Endo V (核酸内切酶V)、Endo VIII (核酸内切酶VIII)、Fpg (甲酰氨基-嘧啶-DNA糖苷酶)、OGG1 (8-氧鸟嘌呤DNA糖苷酶1)、NEIL1 (核酸内切酶VIII样1)、T7 Endo I (T7核酸内切酶I)、T4 PDG (T4嘧啶二聚体DNA糖苷酶)、UDG (尿嘧啶DNA糖苷酶)、SMUG1 (单链选择性单功能尿嘧啶DNA糖苷酶)、AAG (甲基嘌呤DNA糖苷酶)或其任何组合。在一些实施方案中,一种或更多种裂解剂是热稳定的。在一些实施方案中,一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。在一些实施方案中,裂解子识别序列包含限制性位点,并且靶特异性裂解子包括限制性核酸内切酶。在一些实施方案中,靶特异性裂解子包括TAL-效应子核酸酶 (TALEN),所述TAL-效应子核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括锌指核酸酶 (ZFN),所述锌指核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括CRISPR复合体,所述CRISPR复合体包括与互补于裂解子识别序列的引导RNA (gRNA) 复合的CRISPR酶。在一些实施方案中,裂解子识别序列包括前间区邻近基序 (PAM) 位点。在一些实施方案中,CRISPR酶包括Cas9、Cpf1、Cas13a、其变体、其衍生物或其任何组合。在一些实施方案中,CRISPR酶包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。在一些实施方案中,一种或更多种裂解剂包括双链特异性裂解剂。在一些实施方案中,双链特异性裂解剂不能实现单链裂解区的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。在一些实施方案中,双链特异性裂解剂不特异性结合裂解子识别序列。在一些实施方案中,双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区的两条链的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区中仅一条链的磷酸二酯主链的裂解。

[0014] 在一些实施方案中,裂解区包含裂解子识别序列和一个或多个非典型核苷酸,该方法包括:使延伸的细胞组分结合试剂特异性寡核苷酸与双链特异性裂解剂接触以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离;使条形码化核酸分子与一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离。该方法可以包括:使多于一种条形码化核酸分子和多于一种延伸的细胞组分结合试剂特异性寡核苷酸与一种或更多种裂解剂接触以产生:(i) 未与固体支持物关联的多于一种裂解的条形码化核酸分子,和(ii) 与固体支持物关联的多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使多于一种裂解的条形码化核酸分子与固体支持物分离;使与固体支持物关联的裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性

以产生多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸;以及使多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离。该方法可以包括:使延伸的细胞组分结合试剂特异性寡核苷酸与双链特异性裂解剂接触以产生与固体支持物关联的多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使与固体支持物关联的裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性以产生多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离;使条形码化核酸分子与一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。

[0015] 在一些实施方案中,使多于一种条形码化核酸分子与固体支持物分离,发生在使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离之前。在一些实施方案中,使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离,发生在使多于一种条形码化核酸分子与固体支持物分离之前。该方法可以包括:使与细胞组分结合试剂特异性寡核苷酸杂交的多于一种寡核苷酸条形码延伸以产生多于一种条形码化的细胞组分结合试剂特异性寡核苷酸。在一些实施方案中,多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸包括多于一种分离的条形码化细胞组分结合试剂特异性寡核苷酸。在一些实施方案中,多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸包括多于一种裂解的条形码化细胞组分结合试剂特异性寡核苷酸。在一些实施方案中,多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸和多于一种分离的条形码化核酸分子在空间上彼此隔离。在一些实施方案中,多于一种条形码化细胞组分结合试剂特异性寡核苷酸和多于一种分离的条形码化核酸分子在空间上彼此隔离。在一些实施方案中,使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离和/或使多于一种条形码化核酸分子与固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

[0016] 在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含第二通用序列。在一些实施方案中,获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息包括: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,扩增多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸;以及获得多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的测序数据。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含第二分子标记,任选地多于一种细胞组分结合试剂特异性寡核苷酸中的至少10种包含不同的第二分子标记序列。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目指示多于一个细胞中的一个或多个中至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示

多于一个细胞中的一个或更多个中至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,获得序列信息包括将测序衔接子附接到多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物和/或多于一种分离的条形码化核酸分子或其产物。在一些实施方案中,确定多于一个细胞中的一个或更多个中核酸靶的拷贝数包括基于与多于一种分离的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、其互补体或其组合的数目来确定多于一个细胞中核酸靶的拷贝数。

[0017] 该方法可以包括:使随机引物与多于一种分离的条形码化核酸分子接触,其中随机引物中的每一种包含第三通用序列或其互补体;以及使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物。该方法可以包括: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三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的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到多于一种延伸产物。该方法可以包括: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第一分子标记的数目,确定样品中核酸靶的拷贝数。在一些实施方案中,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第一分子标记的数目来确定样品中多于一种核酸靶中的每一种的数目,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。

[0018] 该方法可以包括: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三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的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到第一多于一种条形码化扩增子。该方法可以包括: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第一分子标记的数目,确定样品中核酸靶的拷贝数。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。

[0019] 该方法可以包括:使用多于一种分离的条形码化核酸分子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分离的条形码化核酸分子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靶特异性引物进行的PCR扩增。该方法可以包括:获得第三多于一种条形码化扩增子或其产物的序列信息,并且任选地获得所述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。该方法可以包括:基于与第三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第一分子标记的数目,确定样品中核酸靶的拷贝数。

[0020] 在一些实施方案中,(i) 扩增多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物,(ii) 使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生

多于一种延伸产物和/或(iii)合成第三多于一种条形码化扩增子是分别进行的。在一些实施方案中,使与多于一种分离的条形码化核酸分子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分离的条形码化核酸分子杂交的随机引物延伸,其中使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用缺乏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寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸,和/或其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,DNA聚合酶包括Klenow片段。在一些实施方案中,使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用逆转录酶使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸,和/或其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,任选地其中病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶和/或Moloney鼠白血病病毒(MMLV)逆转录酶。

[0021] 在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含与多(dA)序列相邻的对齐序列。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸通过接头与细胞组分结合试剂关联,任选地细胞组分结合试剂特异性寡核苷酸被配置为能够从细胞组分结合试剂脱离。该方法可以包括:使细胞组分结合试剂特异性寡核苷酸与细胞组分结合试剂解离,任选地在使多于一种寡核苷酸条形码与细胞组分结合试剂特异性寡核苷酸接触之前,使细胞组分结合试剂特异性寡核苷酸与细胞组分结合试剂解离。在一些实施方案中,多于一个分区中的分区包含单个固体支持物。在一些实施方案中,将多于一个细胞分区包括将与多于一种细胞组分结合试剂关联的多于一个细胞和包含该固体支持物的多于一个固体支持物分区到多于一个分区,其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞和该固体支持物,任选地分区是孔或液滴。该方法可以包括:在使多于一种细胞组分结合试剂与多于一个细胞接触之后,去除多于一种细胞组分结合试剂中未与多于一个细胞接触的一种或更多种细胞组分结合试剂。在一些实施方案中,去除未与多于一个细胞接触的一种或更多种细胞组分结合试剂包括:去除未与多于一种细胞组分靶中对应的至少一种靶接触的一种或更多种细胞组分结合试剂。在一些实施方案中,多于一个细胞包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。

[0022] 本文的公开内容包括用于对多于一个细胞中的核酸靶进行标记的方法。在一些实施方案中,该方法包括:使多于一个细胞中的核酸靶的拷贝与包含多于一种寡核苷酸条形码的固体支持物接触,其中每一种寡核苷酸条形码包含第一分子标记、第一通用序列、能够与核酸靶杂交的靶结合区和裂解区,其中裂解区位于第一通用序列的5'并包含一个或更多个裂解位点;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条形码化核酸分子,所述多于一种条形码化核酸分子各自包含与核酸靶的至少一部分互补的序列和第一分子标记;使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子;获得多于一种分离的条形码化核酸分子或其产物的序列信息,以确

定多于一个细胞中的一个或更多个中核酸靶的拷贝数。

[0023] 在一些实施方案中,多于一种条形码化核酸分子各自包含单链裂解区。在一些实施方案中,使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子包括:使裂解寡核苷酸与单链裂解区杂交以产生双链裂解区;使条形码化核酸分子与能够在一个或更多个裂解位点处实现条形码化核酸分子的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。在一些实施方案中,使多于一种条形码化核酸分子与固体支持物分离包括:使条形码化核酸分子与能够在一个或更多个裂解位点处实现条形码化核酸分子的单链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离,以产生多于一种分离的条形码化核酸分子。

[0024] 在一些实施方案中,一个或更多个裂解位点包含非典型核苷酸、裂解子识别序列或其任何组合。在一些实施方案中,一个或更多个裂解位点包括第一裂解位点和第二裂解位点,其中第一裂解位点包含裂解子识别序列,并且其中第二裂解位点包含一个或更多个非典型核苷酸。在一些实施方案中,非典型核苷酸包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1,N6-亚乙烯基腺嘌呤或其任何组合。

[0025] 在一些实施方案中,非典型核苷酸包括8-氧代鸟嘌呤(8oxoG)。在一些实施方案中,一种或更多种裂解剂包括DNA糖苷酶、AP裂解剂或其任何组合。

[0026] 在一些实施方案中,DNA糖苷酶能够裂解非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶(AP)位点。在一些实施方案中,AP裂解剂能够裂解与AP位点相邻的磷酸二酯主链。在一些实施方案中,AP裂解剂包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合。在一些实施方案中,胺包括N,N'-二甲基乙二胺(DMED)。在一些实施方案中,AP裂解酶包括AP裂合酶、AP核酸内切酶或其任何组合。在一些实施方案中,一种或更多种裂解剂包括APE 1 (AP核酸内切酶1)、Endo III (核酸内切酶III)、Endo IV (核酸内切酶IV)、Endo V (核酸内切酶V)、Endo VIII (核酸内切酶VIII)、Fpg (甲酰氨基-嘧啶-DNA糖苷酶)、OGG1 (8-氧鸟嘌呤DNA糖苷酶1)、NEIL1 (核酸内切酶VIII样1)、T7 Endo I (T7核酸内切酶I)、T4 PDG (T4嘧啶二聚体DNA糖苷酶)、UDG (尿嘧啶DNA糖苷酶)、SMUG1 (单链选择性单功能尿嘧啶DNA糖苷酶)、AAG (甲基嘌呤DNA糖苷酶)或其任何组合。在一些实施方案中,一种或更多种裂解剂是热稳定的。在一些实施方案中,一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。在一些实施方案中,裂解子识别序列包含限制性位点,并且靶特异性裂解子包括限制性核酸内切酶。在一些实施方案中,靶特异性裂解子包括TAL-效应子核酸酶(TALEN),所述TAL-效应子核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括锌指核酸酶(ZFN),所述锌指核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括CRISPR复合体,所述CRISPR复合体包括与互补于裂解子识别序列的引导RNA(gRNA)复合的CRISPR酶。在一些实施方案中,裂解子识别序列包括前间区邻

近基序(PAM)位点。在一些实施方案中,CRISPR酶包括Cas9、Cpf1、Cas13a、其变体、其衍生物或其任何组合。在一些实施方案中,CRISPR酶包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。在一些实施方案中,一种或更多种裂解剂包括双链特异性裂解剂。在一些实施方案中,双链特异性裂解剂不能实现单链裂解区的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。在一些实施方案中,双链特异性裂解剂不特异性结合裂解子识别序列。在一些实施方案中,双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区的两条链的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区中仅一条链的磷酸二酯主链的裂解。

[0027] 在一些实施方案中,使多于一种条形码化核酸分子与固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。在一些实施方案中,确定多于一个细胞中的一个或更多个中核酸靶的拷贝数包括基于与多于一种分离的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、其互补体或其组合的数目来确定多于一个细胞中核酸靶的拷贝数。该方法可以包括:使随机引物与多于一种分离的条形码化核酸分子接触,其中随机引物中的每一种包含第三通用序列或其互补体;以及使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物。该方法可以包括: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三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的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到多于一种延伸产物。该方法可以包括: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第一分子标记的数目,确定多于一个细胞中的一个或更多个中核酸靶的拷贝数。在一些实施方案中,确定多于一个细胞中的一个或更多个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第一分子标记的数目来确定多于一个细胞中的一个或更多个中多于一种核酸靶中的每一种的数目,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。

[0028] 该方法可以包括: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三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的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到第一多于一种条形码化扩增子。该方法可以包括: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第一分子标记的数目,确定多于一个细胞中的一个或更多个中核酸靶的拷贝

数。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。

[0029] 该方法可以包括:使用多于一种分离的条形码化核酸分子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分离的条形码化核酸分子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靶特异性引物进行的PCR扩增。该方法可以包括:获得第三多于一种条形码化扩增子或其产物的序列信息,并且任选地获得所述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。该方法可以包括:基于与第三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第一分子标记的数目,确定多于一个细胞中的一个或多个中核酸靶的拷贝数。

[0030] 在一些实施方案中,(i)使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物和/或(ii)合成第三多于一种条形码化扩增子是分别进行的。在一些实施方案中,使与多于一种分离的条形码化核酸分子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分离的条形码化核酸分子杂交的随机引物延伸,和/或其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,DNA聚合酶包括Klenow片段。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,任选地其中病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶和/或Moloney鼠白血病病毒(MMLV)逆转录酶。该方法可以包括:将多于一个细胞分区到多于一个分区,其中多于一个分区中的分区包含来自多于一个细胞的单细胞。

[0031] 在一些实施方案中,多于一个分区中的分区包含单个固体支持物。将多于一个细胞分区可以包括将多于一个细胞和包含该固体支持物的多于一个固体支持物分区到多于一个分区,其中多于一个分区中的分区包含来自多于一个细胞的单细胞和该固体支持物,任选地其中分区是孔或液滴。在一些实施方案中,多于一个细胞包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。在一些实施方案中,核酸靶包括核酸分子(例如,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、细胞组分结合试剂特异性寡核苷酸、样品索引寡核苷酸或其任何组合)。

[0032] 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,该方法包括:使多于一个样品中的每一个分别与多于一种样品索引组合物中的样品索引组合物接触,其中多于一个样品中的每一个包含一个或多个细胞,所述一个或多个细胞各自包含一种或更多种细胞组分靶,其中样品索引组合物包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列;使多于一种样品索引组合物的样品索引

寡核苷酸与包含多于一种寡核苷酸条形码的固体支持物接触,其中每一种寡核苷酸条形码包含第一分子标记、能够与样品索引寡核苷酸杂交的靶结合区和裂解区,其中裂解区包含一个或更多个裂解位点;使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸以产生多于一种延伸的样品索引寡核苷酸,所述多于一种延伸的样品索引寡核苷酸各自包含第一分子标记的互补体和裂解区的互补体;使多于一种延伸的样品索引寡核苷酸与固体支持物分离以产生多于一种分离的延伸的样品索引寡核苷酸;获得多于一种分离的延伸的样品索引寡核苷酸或其产物的测序数据;以及基于测序数据中多于一种分离的延伸的样品索引寡核苷酸或其产物中的至少一种分离的延伸的样品索引寡核苷酸或其产物的样品索引序列,鉴定多于一个细胞中的至少一个细胞的样品来源。

[0033] 在一些实施方案中,使多于一种延伸的样品索引寡核苷酸与固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。在一些实施方案中,每一种寡核苷酸条形码包含第一通用序列。在一些实施方案中,多于一种延伸的样品索引寡核苷酸包含第一通用序列的互补体。在一些实施方案中,样品索引寡核苷酸包含第二分子标记。在一些实施方案中,至少两种样品索引寡核苷酸的第二分子标记序列是不同的,并且其中至少两种样品索引寡核苷酸的样品索引序列是相同的。在一些实施方案中,至少两种样品索引寡核苷酸的第二分子标记序列是不同的,并且其中至少两种样品索引寡核苷酸的样品索引序列是不同的。在一些实施方案中,多于一种样品索引组合物中的至少10种、100种或1000种样品索引组合物的样品索引序列包含不同的序列。

[0034] 在一些实施方案中,样品索引寡核苷酸包含第二通用序列。在一些实施方案中,鉴定至少一个细胞的样品来源包括鉴定测序数据中至少一种分离的延伸的样品索引寡核苷酸或其产物中样品索引序列的存在或不存在。在一些实施方案中,鉴定样品索引序列的存在或不存在包括: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,扩增至少一种分离的延伸的样品索引寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的样品索引寡核苷酸;获得多于一种扩增的分离的延伸的样品索引寡核苷酸或其产物的测序数据;以及基于测序数据中多于一种扩增的分离的延伸的样品索引寡核苷酸或其产物中对应于至少一种分离的延伸的样品索引寡核苷酸或其产物的扩增的分离的延伸的样品索引寡核苷酸或其产物的样品索引序列来鉴定细胞的样品来源。在一些实施方案中,扩增至少一种分离的延伸的样品索引寡核苷酸或其产物包括将测序衔接子附接到多于一种分离的延伸的样品索引寡核苷酸或其产物。

[0035] 在一些实施方案中,使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸包括使用逆转录酶使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸。在一些实施方案中,使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸包括使用缺乏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寡核苷酸条形码杂交的样品索引寡核苷酸延伸。在一些实施方案中,DNA聚合酶包括Klenow片段。

[0036] 该方法可以包括:去除多于一种样品索引组合物中未结合的样品索引组合物。在一些实施方案中,去除未结合的样品索引组合物包括用洗涤缓冲液洗涤来自多于一个样品中的每一个的一个或更多个细胞。在一些实施方案中,去除未结合的样品索引组合物包括使用流式细胞术选择与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,样品

索引寡核苷酸被配置为能够从细胞组分结合试剂脱离。该方法可以包括：使样品索引寡核苷酸与细胞组分结合试剂解离。在一些实施方案中，靶结合区包含多(dT)区。在一些实施方案中，样品索引寡核苷酸包含多(dA)区。在一些实施方案中，样品索引寡核苷酸包含与多(dA)区相邻的对齐序列。在一些实施方案中，样品索引寡核苷酸通过接头与细胞组分结合试剂关联。在一些实施方案中，多于一个样品中的样品包括多于一个细胞、多于一个单细胞、组织、肿瘤样品或其任何组合。

[0037] 在一些实施方案中，使多于一种延伸的样品索引寡核苷酸与固体支持物分离包括：使延伸的样品索引寡核苷酸与一种或更多种裂解剂接触。在一些实施方案中，延伸的细胞组分结合试剂特异性寡核苷酸各自包含双链裂解区。在一些实施方案中，使多于一种延伸的样品索引寡核苷酸与固体支持物分离包括：使延伸的样品索引寡核苷酸与能够在一个或更多个裂解位点处实现延伸的样品索引寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触以产生多于一种裂解的延伸的样品索引寡核苷酸；以及使多于一种裂解的延伸的样品索引寡核苷酸与固体支持物分离以产生多于一种分离的延伸的样品索引寡核苷酸。在一些实施方案中，使多于一种裂解的延伸的样品索引寡核苷酸与固体支持物分离包括使多于一种裂解的延伸的样品索引寡核苷酸变性。

[0038] 本文的公开内容包括用于测量细胞中细胞组分表达的方法。在一些实施方案中，该方法包括：使多于一种细胞组分结合试剂与包含多于一种细胞组分靶的多于一个细胞接触，其中多于一种细胞组分结合试剂中的每一种包含细胞组分结合试剂特异性寡核苷酸，所述细胞组分结合试剂特异性寡核苷酸包含用于细胞组分结合试剂的独特标识符序列，并且其中细胞组分结合试剂能够与多于一种细胞组分靶中的至少一种特异性结合；使细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触，其中每一种寡核苷酸条形码包含第一分子标记、能够与细胞组分结合试剂特异性寡核苷酸杂交的靶结合区和裂解区，其中裂解区包含一个或更多个裂解位点；使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸，所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含第一分子标记的互补体和裂解区的互补体；使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸；获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息，以确定多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。

[0039] 该方法可以包括：在使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸之前：将与多于一种细胞组分结合试剂关联的多于一个细胞分区到多于一个分区，其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞；在包含单细胞的分区中，使细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触。在一些实施方案中，多于一个分区中的分区包含单个固体支持物。在一些实施方案中，使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

[0040] 在一些实施方案中，每一种寡核苷酸条形码包含第一通用序列。在一些实施方案中，多于一种延伸的细胞组分结合试剂特异性寡核苷酸包含第一通用序列的互补体。在一些实施方案中，细胞组分结合试剂特异性寡核苷酸包含第二通用序列。在一些实施方案中，

获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息包括：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，扩增多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物，以产生多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸；以及获得多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的测序数据。在一些实施方案中，细胞组分结合试剂特异性寡核苷酸包含第二分子标记。在一些实施方案中，多于一种细胞组分结合试剂特异性寡核苷酸中的至少10种包含不同的第二分子标记序列。在一些实施方案中，至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的，并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的。在一些实施方案中，至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的，并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。在一些实施方案中，测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目指示多于一个细胞中的一个或多个中至少一种细胞组分靶的拷贝数，所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中，测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示多于一个细胞中的一个或多个中至少一种细胞组分靶的拷贝数，所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中，获得序列信息包括将测序衔接子附接到多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物。

[0041] 在一些实施方案中，使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用逆转录酶使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸。在一些实施方案中，使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用缺乏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寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸。在一些实施方案中，DNA聚合酶包括Klenow片段。在一些实施方案中，靶结合区包含多(dT)区。

[0042] 在一些实施方案中，细胞组分结合试剂特异性寡核苷酸包含多(dA)区，任选地细胞组分结合试剂特异性寡核苷酸包含与多(dA)区相邻的对齐序列。在一些实施方案中，细胞组分结合试剂特异性寡核苷酸通过接头与细胞组分结合试剂关联。在一些实施方案中，细胞组分结合试剂特异性寡核苷酸被配置为能够从细胞组分结合试剂脱离。该方法可以包括：使细胞组分结合试剂特异性寡核苷酸与细胞组分结合试剂解离。在一些实施方案中，将多于一个细胞分区包括将与多于一种细胞组分结合试剂关联的多于一个细胞和包含该固体支持物的多于一个固体支持物分区到多于一个分区，其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞和该固体支持物。在一些实施方案中，分区是孔或液滴。该方法可以包括：在使多于一种细胞组分结合试剂与多于一个细胞接触之后，去除多于一种细胞组分结合试剂中未与多于一个细胞接触的一种或更多种细胞组分结合试剂。在一些实施方案中，去除未与多于一个细胞接触的一种或更多种细胞组分结合试剂包括：去除未与多于一种细胞组分靶中对应的至少一种靶接触的一种或更多种细胞组分结合试剂。在一些实施方案中，多于一个细胞包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。

[0043] 在一些实施方案中,使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括:使延伸的细胞组分结合试剂特异性寡核苷酸与一种或更多种裂解剂接触。在一些实施方案中,延伸的细胞组分结合试剂特异性寡核苷酸各自包含双链裂解区。在一些实施方案中,使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括:使延伸的细胞组分结合试剂特异性寡核苷酸与能够在一个或更多个裂解位点处实现延伸的细胞组分结合试剂特异性寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;以及使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。在一些实施方案中,使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性。

[0044] 在一些实施方案中,一个或更多个裂解位点包含非典型核苷酸、裂解子识别序列或其任何组合。在一些实施方案中,一个或更多个裂解位点包括第一裂解位点和第二裂解位点,其中第一裂解位点包含裂解子识别序列,并且其中第二裂解位点包含一个或更多个非典型核苷酸。在一些实施方案中,非典型核苷酸包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1,N6-亚乙烯基腺嘌呤或其任何组合。

[0045] 在一些实施方案中,非典型核苷酸包括8-氧代鸟嘌呤(8oxoG)。在一些实施方案中,一种或更多种裂解剂包括DNA糖苷酶、AP裂解剂或其任何组合。

[0046] 在一些实施方案中,DNA糖苷酶能够裂解非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶(AP)位点。在一些实施方案中,AP裂解剂能够裂解与AP位点相邻的磷酸二酯主链。在一些实施方案中,AP裂解剂包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合。在一些实施方案中,胺包括N,N'-二甲基乙二胺(DMED)。在一些实施方案中,AP裂解酶包括AP裂解酶、AP核酸内切酶或其任何组合。在一些实施方案中,一种或更多种裂解剂包括APE 1 (AP核酸内切酶1)、Endo III (核酸内切酶III)、Endo IV (核酸内切酶IV)、Endo V (核酸内切酶V)、Endo VIII (核酸内切酶VIII)、Fpg (甲酰氨基-嘧啶-DNA糖苷酶)、OGG1 (8-氧鸟嘌呤DNA糖苷酶1)、NEIL1 (核酸内切酶VIII样1)、T7 Endo I (T7核酸内切酶I)、T4 PDG (T4嘧啶二聚体DNA糖苷酶)、UDG (尿嘧啶DNA糖苷酶)、SMUG1 (单链选择性单功能尿嘧啶DNA糖苷酶)、AAG (甲基嘌呤DNA糖苷酶)或其任何组合。在一些实施方案中,一种或更多种裂解剂是热稳定的。在一些实施方案中,一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。在一些实施方案中,裂解子识别序列包含限制性位点,并且靶特异性裂解子包括限制性核酸内切酶。在一些实施方案中,靶特异性裂解子包括TAL-效应子核酸酶(TALEN),所述TAL-效应子核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括锌指核酸酶(ZFN),所述锌指核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括CRISPR复合体,所述CRISPR复合体包括与互补于裂解子识别序列的引导RNA(gRNA)复合的CRISPR酶。在一些实施方案中,裂解子识别序列包括前间区邻近基序(PAM)位点。在一些实施方案中,CRISPR酶包括Cas9、Cpf1、Cas13a、其变体、其衍生物

或其任何组合。在一些实施方案中,CRISPR酶包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。在一些实施方案中,一种或更多种裂解剂包括双链特异性裂解剂。在一些实施方案中,双链特异性裂解剂不能实现单链裂解区的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。在一些实施方案中,双链特异性裂解剂不特异性结合裂解子识别序列。在一些实施方案中,双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区的两条链的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区中仅一条链的磷酸二酯主链的裂解。

[0047] 在一些实施方案中,变性可以包括加热和/或碱变性。在一些实施方案中,第一通用序列、第二通用序列和/或第三通用序列是相同的。在一些实施方案中,第一通用序列、第二通用序列和/或第三通用序列是不同的。在一些实施方案中,第一通用序列、第二通用序列和/或第三通用序列包含测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分。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。在一些实施方案中,对齐序列的长度为一个或多个核苷酸,或长度为两个或多个核苷酸。在一些实施方案中,(a)对齐序列包含鸟嘌呤、胞嘧啶、胸腺嘧啶、尿嘧啶或其组合;(b)对齐序列包含多(dT)序列、多(dG)序列、多(dC)序列、多(dU)序列或其组合;和/或(c)对齐序列位于多(dA)区的5'。在一些实施方案中,接头包含碳链,任选地碳链包含2-30个碳原子,并且还任选地碳链包含12个碳原子。在一些实施方案中,接头包含5'氨基修饰物C12(5AmMC12)或其衍生物。在一些实施方案中,细胞组分靶包括蛋白靶。在一些实施方案中,细胞组分靶包括碳水化合物、脂质、蛋白、细胞外蛋白、细胞表面蛋白、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、细胞内蛋白或其任何组合。在一些实施方案中,细胞组分靶在细胞表面上。在一些实施方案中,多于一种寡核苷酸条形码中的至少10种包含不同的第一分子标记序列。在一些实施方案中,多于一种寡核苷酸条形码与固体支持物关联。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。在一些实施方案中,多于一种寡核苷酸条形码的每一种细胞标记包含至少6个核苷酸。在一些实施方案中,与同一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。

[0048] 在一些实施方案中,固体支持物包括合成颗粒。在一些实施方案中,固体支持物包括平坦表面。在一些实施方案中,样品包括单细胞,该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,合成颗粒和单细胞在同一分区(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在合成颗粒上、部分地固定在合成颗粒上、包封在合成颗粒中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。在一些实施方案

中,合成颗粒包括珠。在一些实施方案中,珠包括琼脂糖凝胶(Sephrose)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(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠或其任何组合。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、胶乳、琼脂糖凝胶、纤维素、尼龙、硅酮及其任何组合。在一些实施方案中,合成颗粒包括可破坏的水凝胶颗粒。

[0049] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记、靶结合区和裂解区,其中裂解区包含一个或更多个裂解位点,并且其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。

[0050] 试剂盒可以包含一种或更多种裂解剂。试剂盒可以包含DNA聚合酶或逆转录酶。试剂盒可以包含缓冲液。试剂盒可以包含筒(cartridge)。试剂盒可以包含一种或更多种用于逆转录反应的试剂。试剂盒可以包含一种或更多种用于扩增反应的试剂。

[0051] 靶结合区可以包含基因特异性序列、寡(dT)序列、随机多聚体或其任何组合。寡核苷酸条形码可以包含相同的样品标记和/或相同的细胞标记。多于一种寡核苷酸条形码的每一种样品标记和/或细胞标记可以包含至少6个核苷酸。多于一种寡核苷酸条形码的每一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中的至少10种可以包含不同的第一分子标记序列。多于一种寡核苷酸条形码可以与固体支持物关联。多于一种寡核苷酸条形码各自可以包含细胞标记。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。与同一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或不同的细胞标记。固体支持物可以包括合成颗粒或平坦表面。样品可以包括单细胞,并且该方法可以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。合成颗粒和单细胞可以在同一分区(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在合成颗粒上、部分地固定在合成颗粒上、包封在合成颗粒中或部分地包封在合成颗粒中。

[0052] 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(例如,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(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠或其任何组合)。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、胶乳、琼脂糖凝胶、纤维素、尼龙、硅酮,以及其任何组合。多于一种寡核苷酸条形码中的每一种包含接头官能团。合成颗粒可以包含固体支持物官能团。支持物官能团和接头官能团可以彼此关联。接头官能团和支持物官能团可以单独地选自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,以及其任何组合组成的组。

[0053] 在一些实施方案中,一个或更多个裂解位点包含非典型核苷酸、裂解子识别序列或其任何组合。在一些实施方案中,一个或更多个裂解位点包括第一裂解位点和第二裂解位点,其中第一裂解位点包含裂解子识别序列,并且其中第二裂解位点包含一个或更多个

非典型核苷酸。在一些实施方案中,非典型核苷酸包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1,N6-亚乙烯基腺嘌呤或其任何组合。

[0054] 在一些实施方案中,非典型核苷酸包括8-氧代鸟嘌呤(8oxoG)。在一些实施方案中,一种或更多种裂解剂包括DNA糖苷酶、AP裂解剂或其任何组合。

[0055] 在一些实施方案中,DNA糖苷酶能够裂解非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶(AP)位点。在一些实施方案中,AP裂解剂能够裂解与AP位点相邻的磷酸二酯主链。在一些实施方案中,AP裂解剂包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合。在一些实施方案中,胺包括N,N'-二甲基乙二胺(DMED)。在一些实施方案中,AP裂解酶包括AP裂解酶、AP核酸内切酶或其任何组合。在一些实施方案中,一种或更多种裂解剂包括APE 1 (AP核酸内切酶1)、Endo III (核酸内切酶III)、Endo IV (核酸内切酶IV)、Endo V (核酸内切酶V)、Endo VIII (核酸内切酶VIII)、Fpg (甲酰氨基-嘧啶-DNA糖苷酶)、OGG1 (8-氧鸟嘌呤DNA糖苷酶1)、NEIL1 (核酸内切酶VIII样1)、T7 Endo I (T7核酸内切酶I)、T4 PDG (T4嘧啶二聚体DNA糖苷酶)、UDG (尿嘧啶DNA糖苷酶)、SMUG1 (单链选择性单功能尿嘧啶DNA糖苷酶)、AAG (甲基嘌呤DNA糖苷酶)或其任何组合。在一些实施方案中,一种或更多种裂解剂是热稳定的。在一些实施方案中,一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。在一些实施方案中,裂解子识别序列包含限制性位点,并且靶特异性裂解子包括限制性核酸内切酶。在一些实施方案中,靶特异性裂解子包括TAL-效应子核酸酶(TALEN),所述TAL-效应子核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括锌指核酸酶(ZFN),所述锌指核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括CRISPR复合体,所述CRISPR复合体包括与互补于裂解子识别序列的引导RNA(gRNA)复合的CRISPR酶。在一些实施方案中,裂解子识别序列包括前间区邻近基序(PAM)位点。在一些实施方案中,CRISPR酶包括Cas9、Cpf1、Cas13a、其变体、其衍生物或其任何组合。在一些实施方案中,CRISPR酶包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。在一些实施方案中,一种或更多种裂解剂包括双链特异性裂解剂。在一些实施方案中,双链特异性裂解剂不能实现单链裂解区的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。在一些实施方案中,双链特异性裂解剂不特异性结合裂解子识别序列。在一些实施方案中,双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区的两条链的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区中仅一条链的磷酸二酯主链的裂解。

[0056] 附图简述

[0057] 图1图示了非限制性示例性随机条形码。

[0058] 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。

[0059] 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文库(indexed library)的非限制性示例性方法的示意图。

[0060] 图4示出了与包含用于蛋白结合试剂的独特标识符的寡核苷酸关联的示例性蛋白结合试剂(此处图示的抗体)的示意图。

[0061] 图5示出了与包含用于确定细胞来自同一样品或不同样品的样品索引的独特标识符的寡核苷酸关联的示例性结合试剂(此处图示的抗体)的示意图。

[0062] 图6示出了使用寡核苷酸关联的抗体以高通量方式同时确定细胞组分表达(例如,蛋白表达)和基因表达的示例性工作流程的示意图。

[0063] 图7示出了使用寡核苷酸关联的抗体进行样品索引的示例性工作流程的示意图。

[0064] 图8示出了与结合试剂(此处图示的抗体)关联的结合试剂寡核苷酸(此处图示的抗体寡核苷酸)的条形码化的非限制性示例性工作流程的示意图。

[0065] 图9A-图9D示出了用于同时确定蛋白表达和基因表达以及用于样品索引的寡核苷酸的非限制性示例性设计。

[0066] 图10示出了用于同时确定蛋白表达和基因表达以及用于样品索引的非限制性示例性寡核苷酸序列的示意图。

[0067] 图11A-图11B示出了用于同时确定蛋白表达和基因表达以及用于样品索引的寡核苷酸的非限制性示例性设计。

[0068] 图12A-图12D示出了用于标记核酸靶的非限制性示例性工作流程的示意图。

[0069] 图13A-图13L示出了用于同时测量蛋白和基因表达的非限制性示例性工作流程的示意图。

[0070] 图14A-图14B提供了用于本文提供的方法的非限制性示例性工作流程。

[0071] 详述

[0072] 以下详述中参考了形成本文的一部分的附图。在附图中,除非上下文另外指示,否则相似的符号通常标识相似的组成部分。在详述、附图和权利要求书中描述的说明性实施方案不意味着是限制性的。在不脱离本文提供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,所有这些都都在本文中明确设想并且构成本公开内容的一部分。

[0073] 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。

[0074] 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)的绝对数目也可以是非常具有挑战性的,尤其是当分子数目非常小时。确定样品中分子的绝对数目的一种方法是数字聚合酶链式反应(PCR)。理想地,PCR在每个循环中产生相同拷贝的分子。然而,PCR可具有缺点使得每个分子以随机概率复制,且此概率根据PCR循环和基因序列而变化,这导致扩增偏倚和不准确的基因表达测量。具有独特分子标记(molecular labels,也称为分子索引(molecular indexes,MI))的随机条形码可以用于

计数分子数目和校正扩增偏倚。诸如Precise™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 来校正由 PCR 和文库制备步骤引起的偏倚。

[0075] Precise™测定可以利用具有在多 (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。随机条形码可以包含通用 PCR 引发位点。在 RT 期间, 靶基因分子与随机条形码随机反应。每种靶分子可以与随机条形码杂交, 导致产生随机条形码化的互补核糖核苷酸 (cDNA) 分子。在标记后, 可以将来自微孔板微孔的随机条形码化 cDNA 分子汇集到单个管中用于 PCR 扩增和测序。可以分析原始测序数据以产生读段的数目、具有独特分子标记的随机条形码的数目以及 mRNA 分子的数目。

[0076] 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, Precise™测定可以用于同时确定多于 10000 个细胞的 mRNA 表达谱。每个样品用于分析的单细胞数目 (例如, 数百个或数千个单细胞) 可能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当前单细胞技术的容量利用, 从而降低单细胞分析的试剂浪费和成本。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。

[0077] 本文的公开内容包括同时测量细胞中的蛋白表达和基因表达的方法。在一些实施方案中, 该方法包括: 使多于一种细胞组分结合试剂与包含多于一种细胞组分靶和核酸靶的拷贝的多于一个细胞接触, 其中核酸靶包括 mRNA, 其中多于一种细胞组分结合试剂中的每一种包括包含用于细胞组分结合试剂的独特标识符和多 (A) 序列的细胞组分结合试剂特异性寡核苷酸, 其中细胞组分结合试剂特异性寡核苷酸包括 DNA, 并且其中细胞组分结合试剂能够与多于一种细胞组分靶中的至少一种特异性结合。该方法可以包括: 将与细胞组分结合试剂关联的多于一个细胞分区到多于一个分区, 其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞。该方法可以包括: 在包含单细胞的分区中, 使包含多于一种寡核苷酸条形码的固体支持物与细胞组分结合试剂特异性寡核苷酸和核酸靶的拷贝的接触进行杂交, 其中寡核苷酸条形码各自包含多 (T) 序列、第一通用序列、第一分子标记和裂解区, 其中裂解区位于第一通用序列的 5' 并包含一个或更多个裂解位点。该方法可以包括: 使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸, 所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含第一分子标记的互补体、裂解区的互补体和第一通用序列的互补体。该方法可以包括: 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条形码化核酸分子, 所述多于一种条形码化核酸分子各自包含与核酸靶的至少一部分互补的序列和第一分子标记。该方法可以包括: 使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。该方法可以包括: 使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。该方法可以包括: 获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息, 以确定多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。该方法可以包

括:获得多于一种分离的条形码化核酸分子或其产物的序列信息,以确定多于一个细胞中的一个或更多个中核酸靶的拷贝数。

[0078] 本文的公开内容包括用于对多于一个细胞中的核酸靶进行标记的方法。在一些实施方案中,该方法包括:使多于一个细胞中核酸靶的拷贝与包含多于一种寡核苷酸条形码的固体支持物接触,其中每一种寡核苷酸条形码包含第一分子标记、第一通用序列、能够与核酸靶杂交的靶结合区和裂解区,其中裂解区位于第一通用序列的5' 并包含一个或更多个裂解位点;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条形码化核酸分子,所述多于一种条形码化核酸分子各自包含与核酸靶的至少一部分互补的序列和第一分子标记;使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子;获得多于一种分离的条形码化核酸分子或其产物的序列信息,以确定多于一个细胞中的一个或更多个中核酸靶的拷贝数。

[0079] 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,该方法包括:使多于一个样品中的每一个分别与多于一种样品索引组合物中的样品索引组合物接触,其中多于一个样品中的每一个包含一个或更多个细胞,所述一个或更多个细胞各自包含一种或更多种细胞组分靶,其中样品索引组合物包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列;使多于一种样品索引组合物的样品索引寡核苷酸与包含多于一种寡核苷酸条形码的固体支持物接触,其中每一种寡核苷酸条形码包含第一分子标记、能够与样品索引寡核苷酸杂交的靶结合区和裂解区,其中裂解区包含一个或更多个裂解位点;使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸以产生多于一种延伸的样品索引寡核苷酸,所述多于一种延伸的样品索引寡核苷酸各自包含第一分子标记的互补体和裂解区的互补体;使多于一种延伸的样品索引寡核苷酸与固体支持物分离以产生多于一种分离的延伸的样品索引寡核苷酸;获得多于一种分离的延伸的样品索引寡核苷酸或其产物的测序数据;以及基于测序数据中多于一种分离的延伸的样品索引寡核苷酸或其产物中的至少一种分离的延伸的样品索引寡核苷酸或其产物的样品索引序列,鉴定多于一个细胞中的至少一个细胞的样品来源。

[0080] 本文的公开内容包括用于测量细胞中细胞组分表达的方法。在一些实施方案中,该方法包括:使多于一种细胞组分结合试剂与包含多于一种细胞组分靶的多于一个细胞接触,其中多于一种细胞组分结合试剂中的每一种包含细胞组分结合试剂特异性寡核苷酸,所述细胞组分结合试剂特异性寡核苷酸包含用于细胞组分结合试剂的独特标识符序列,并且其中细胞组分结合试剂能够与多于一种细胞组分靶中的至少一种特异性结合;使细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触,其中每一种寡核苷酸条形码包含第一分子标记、能够与细胞组分结合试剂特异性寡核苷酸杂交的靶结合区和裂解区,其中裂解区包含一个或更多个裂解位点;使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸,所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含第一分子标记的互补体和裂解区的互补体;使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸;获得多于一种分离的延

伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息,以确定多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。

[0081] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记、靶结合区和裂解区,其中裂解区包含一个或更多个裂解位点,并且其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。试剂盒可以包含一种或更多种裂解剂。试剂盒可以包含DNA聚合酶或逆转录酶。试剂盒可以包含缓冲液。试剂盒可以包含筒。试剂盒可以包含一种或更多种用于逆转录反应的试剂。试剂盒可以包含一种或更多种用于扩增反应的试剂。

[0082] 定义

[0083] 除非另外定义,否则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의相同意义。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)。为了本公开内容的目的,下文定义了以下术语。

[0084] 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。关联的核酸可以包括靶核酸。关联的核酸可以包括空间标记、靶标记、样品标记、索引标记或条形码序列(例如,分子标记)中的一种或更多种。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'端或3'端。当衔接子在5'端和3'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'端和/或3'端的衔接子可以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以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于一种不同序列。类似地,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)或更多种单一通用引物复制或扩增多于一种不同序列。因此,通用引物包含可与此类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通用引物杂交的位点。与靶核酸附接的一种或更多种通用引物可以彼此相同或不同。

[0085] 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)。

[0086] 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同一抗原结合。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。

[0087] 如本文使用的,术语“关联”或“与...关联”可以意指,两个或更多个物质可以被鉴定为在某个时间点共定位。关联可以意指,两个或更多个物质在或曾经在相似的容器内。关联可以是信息学关联。例如,关于两个或更多个物质的数字信息可以被存储并且可以用于确定一种或更多种物质在某个时间点共定位。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物质彼此“拴系”、“附接”或“固定”或与共同的固体或半固体表面“拴系”、“附接”或“固定”。关联可以指用于将标记与固体或半固体支持物(诸如珠)附接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。

[0088] 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,如果核酸在给定位置处的核苷酸能够与另一个核酸的核苷酸形成氢键,则这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”,其中只有一些核苷酸结合,或者当单链分子之间存在全部互补性时它可以是完全的。如果第一核苷酸序列与第二核苷酸序列互补,则第一核苷酸序列可以被称为第二序列的“互补体”。如果第一核苷酸序列和与第二序列相反的序列(即,核苷酸顺序相反)互补,则第一核苷酸序列可以被称为第二序列的“反向互补体”。如本文使用的,术语“互补体”、“互补的”和“反向互补体”可以互换使用。从本公开内容理解,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。

[0089] 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数目的方法。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目的步骤。这种方法(其本质上可以是随机的)将计数分子的问题从相同分子的定位和鉴定之一转化为有关检测一组预定义标记的一系列是/否数字问题。

[0090] 如本文使用的,术语“一种标记(label)”或“多于一种标记(labels)”可以指与样品中的靶关联的核酸代码。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接点,例如天然和非天然序列的接点。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,在多种实施方案中,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。

[0091] 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,每种靶可能与独特条形码关联。每种标记的靶分子的独特性可以通过随机选择的统计来确定,并且取决于与标记的多样性相比在集合中相同的靶分子的拷贝数。所得的标记的靶分子的集合的大小可以通过条形码化处理的随机性质来确定,并且然后对检测到的条形码的数目的分析允许计算原始集合或样品中存在的靶分子的数目。当存在的靶分子的拷贝数与独特条形码的数目的比率低时,标记的靶分子是高度独特的(即,多于一种靶分子被给定标记标记的概率非常低)。

[0092] 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,改变的主链、糖或核酸碱基)。类似物的一些非限制性实例包括:5-溴尿嘧啶、肽核酸、非天然核酸(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。

[0093] 核酸可以包括一种或更多种修饰(例如,碱基修饰、主链修饰),以为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。此类杂环碱基的两个最常见的类别是嘌呤和嘧啶。核苷酸可以是还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。在形成核酸时,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端可以进一步连接而形成环状化合物;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,磷酸基团通常可以称为形成核酸的核苷间主链。键(linkage)或主链(backbone)可以是3'到5'磷酸二酯键。

[0094] 核酸可以包括修饰的主链和/或修饰的核苷间键。修饰的主链可以包括那些在主链中保留磷原子的主链和那些在主链中没有磷原子的主链。合适的其中含磷原子的修饰的核酸主链可以包括,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、磷酰胺酯(phosphoramidate)(包括3'-氨基磷酰胺酯和氨基烷基磷酰胺酯、磷酸二酰胺酯(phosphorodiamidates)、硫代磷酰胺酯(thionophosphoramidates)、硫代烷基磷酸酯、硫代烷基磷酸三酯、硒代磷酸酯和硼磷酸酯,具有正常3'-5'连接、2'-5'连接的类似物以及具有反向极性的类似物(其中一个或多个核苷酸间连接是3'至3'、5'至5'或2'至2'连接)。

[0095] 核酸可以包括由短链烷基或环烷基核苷间键,混合杂原子,和烷基或环烷基核苷间键,或者一个或多个短链杂原子的或杂环的核苷间键形成的多核苷酸主链。这些可以包括具有吗啉代(morpholino)连接的那些(部分由核苷的糖部分形成);硅氧烷主链;硫化物、亚砷和砷主链;乙酰基(formacetyl)和硫代乙酰基主链;亚甲基乙酰基和硫代乙酰基主链;核糖乙酰基主链;含烯烃的主链;氨基磺酸酯主链;亚甲基亚氨基和亚甲基胍基主链;磺酸酯和磺酰胺主链;酰胺主链;和具有混合的N、O、S和CH₂组分部分的其他的那些。

[0096] 核酸可以包括核酸模拟物。术语“模拟物”可以意在包括其中仅呋喃糖环或呋喃糖环和核苷酸间键二者被非呋喃糖基团替代的多核苷酸,仅呋喃糖环的替代也可以称为糖替代物(surrogate)。可以维持杂环碱基部分或修饰的杂环碱基部分,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,多核苷酸的糖主链可以被含酰胺的主链替代,特别是被氨基乙基甘氨酸主链替代。核苷酸可以被保留,并且直接或间接与主链的酰胺部分的氮杂氮原子结合。PNA化合物中的主链可以包含两个或多个连接的氨基乙基甘

氨酸单元,这使得PNA具有含酰胺的主链。杂环碱基部分可以直接或间接与主链的酰胺部分的氮杂氮原子结合。

[0097] 核酸可以包括吗啉代主链结构。例如,核酸可以包含替代核糖环的6元吗啉代环。在这些实施方案的一些中,磷酸二酰胺酯或其他非磷酸二酯核苷间键可以替代磷酸二酯键。

[0098] 核酸可以包括具有附接至吗啉代环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。基于非离子吗啉代的寡聚化合物与细胞蛋白可以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。吗啉代类别内的各种化合物可以使用不同的连接基团来连接。另外类别的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替代。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单体掺入核酸链可以增加DNA/RNA杂合体的稳定性。CeNA寡腺苷酸酯可以与核酸互补体形成复合体,具有与天然复合体相似的稳定性。另外的修饰可以包括锁核酸(LNA),其中2'-羟基基团与糖环的4'碳原子连接,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'-外切核酸酶降解的稳定性和良好的溶解性。

[0099] 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)尿嘧啶和胞嘧啶以及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)苯并噁嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噁嗪-2(3H)-酮),G-钳(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)苯并噁嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噁嗪-2(3H)-酮),G-钳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)苯并噁嗪-2(3H)-酮)、咪唑胞苷(2H-嘧啶并(4,5-b)吡啶-2-酮)、吡啶并咪唑胞苷(H-吡啶并(3',2':4,5)吡咯并[2,3-d]嘧啶-2-酮)。

[0100] 如本文使用的,术语“样品”可以指包含靶的组合物。用于通过所公开的方法、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。

[0101] 如本文使用的,术语“采样装置”或“装置”可以指可以取样品的切片和/或将所述切片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、

活检针、活检装置、组织切片装置、微流体装置、刀片格栅和/或超薄切片机。

[0102] 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(bearing)、圆柱体或由塑料、陶瓷、金属或聚合物材料(例如,水凝胶)构成的其他类似配置,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、长方形、锥形、圆柱形、圆锥形、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。

[0103] 如本文使用的,术语“随机条形码”可以指本公开内容的包含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。随机条形码可以用于对样品中的靶定量。随机条形码可以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。

[0104] 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。随机条形码可以用于对样品中的靶定量。随机条形码可以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。

[0105] 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。

[0106] 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。用于通过所公开的方法、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、微小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、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“物质(species)”互换使用。

[0107] 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(retroplasmid)逆转录酶、逆转录子逆转录酶、细菌逆转录酶、II组内含子衍生的逆转录酶,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组内含子逆转录酶。II组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土芽孢杆菌(*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,尤其是逆转录子、II组内含子、以及多样性产生型逆转录元件)。

[0108] 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。通用衔接子序列可以例如是在本公开内容的方法中使用的遍及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于一种靶时,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于一种靶时,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被包括在两种寡核苷酸中,其中的一种寡核苷酸包含靶特异性序列且另一种寡核苷酸包含条形码。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。包含条形码和通用衔接子序列的互补序列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。

[0109] 条形码

[0110] 条形码化,诸如随机条形码化,已经在例如,Fu等人,Proc Natl Acad Sci U.S.A.,2011年5月31日,108(22):9026-31;美国专利申请公布第US2011/0160078号;Fan等人,Science,2015年2月6日,347(6222):1258367;美国专利申请公布第US2015/0299784号;和PCT申请公布第W02015/031691号中描述;这些中的每一项的内容,包括任何支持或补充信息或材料,通过引用以其整体并入本文。在一些实施方案中,本文公开的条形码可以是随机条形码,所述随机条形码是可以用于对靶进行随机标记(例如,条形码化、加标签)的多核苷酸序列。如果随机条形码的不同条形码序列的数目与待标记的任何靶的出现数目的比例可以是以下或可以是约以下: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之间的数字或范围,则条形码可以称为随机条形码。靶可以是包括具有相同或几乎相同序列的mRNA分子的mRNA物质。如果随机条形码的不同条形码序列的数目与待标记的任何靶的出现数目的比例是至少以下或是至多以下: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,则条形码可以称为随机条形码。随机条形码的条形码序列可以称为分子标记。

[0111] 条形码(例如,随机条形码)可以包括一种或更多种标记。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可以包含可以使条形码与固体支持物108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)的顺序可以变化。例如,如图1中所示,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,靶结合区可以包含可以与mRNA的多(A)尾相互作用的寡(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、20个或更多个核苷酸隔开。

[0112] 标记(例如,细胞标记)可以包含一组独特的定义长度的核酸子序列,例如每种七

个核苷酸(相当于一些汉明错误校正代码中使用的比特数目),其可被设计成提供错误校正能力。可以设计包含七个核苷酸序列的错误校正子序列组,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),例如一组错误校正子序列可被设计成展现三个核苷酸的遗传距离。在这种情况下,对于标记的靶核酸分子的序列数据组中的错误校正序列的审查(在下文更详细地描述)可允许人们检测或校正扩增错误或测序错误。在一些实施方案中,用于产生错误校正代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可以是约以下: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个核苷酸或在这些值中的任何两个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错误校正代码。

[0113] 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。靶可以是以下或包括以下:核糖核酸(RNA)、信使RNA(mRNA)、微小RNA、小干扰RNA(siRNA)、RNA降解产物、各自含有多(A)尾的RNA或其任何组合。在一些实施方案中,多于一种靶可以包括脱氧核糖核酸(DNA)。

[0114] 在一些实施方案中,靶结合区可以包括可以与mRNA的多(A)尾相互作用的寡(dT)序列。条形码的一种或更多种标记(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列(例如,分子标记))可以通过间隔区(spacer)与条形码的另一种或两种剩余标记隔开。间隔区可以是例如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区隔开。

[0115] 通用标记

[0116] 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,一种或更多种通用标记可以是对于附接至给定固体支持物的条形码组中的所有条形码相同的。在一些实施方案中,一种或更多种通用标记可以是对于附接至多于一个珠的所有条形码相同的。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括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相关的测序引物。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。能够与测序引物或PCR引物杂交的通用标记的核酸序列可以被称为引物结合位点。通用标记可以包括可以用于引发条形码转录的序列。通用标记可以包括可以用于延伸条形码或条形码内的区域的序列。通用标记的长度可以是以下或可以是约以下: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个核苷酸或在这些值中的任何两个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可以是至多以下: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。

[0117] 维度标记

[0118] 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标记(例如,随机标记)发生的维度的信息的核酸序列。例如,维度标记可以提供关于靶被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期可以将细胞的群体条形码化。在细胞周期的G1期,可以用条形码(例如,随机条形码)对细胞再次进行脉冲处理。在细胞周期的S期,可以用条形码对细胞再次进行脉冲处理,等等。每个脉冲(例如,细胞周期的每个时期)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时期被标记的信息。维度标记可以探询许多不同的生物时间。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。在另一种实例中,样品(例如,细胞、细胞的群体)可以在用药物和/或疗法治疗之前和/或之后标记。不同靶的拷贝数的变化可以指示样品对药物和/或疗法的响应。

[0119] 维度标记可以是可激活的。可激活的维度标记可以在特定时间点被激活。可激活的标记可以被例如组成性地激活(例如,不关闭)。可激活的维度标记可以被例如可逆地激活(例如,可激活的维度标记可以被打开和关闭)。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,连接另一种分子,添加修饰(例如,聚乙二醇化、类泛素化(sumoylate)、乙酰化、甲基化、去乙酰化、去甲基化)、光化学事件(例如,光囚禁(photocaging))以及引入非天然的核苷酸将维度标记激活。

[0120] 在一些实施方案中,维度标记对于附接至给定固体支持物(例如,珠)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,同一固体支持物上至少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,同一固体支持物上至少60%的条形码可以包含相同的维度标记。在一些实施方案中,同一固体支持物上至少95%的条形码可以包含相同的维度标记。

[0121] 多于一个固体支持物(例如,珠)中可以呈现多达 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可以是约以下: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个核苷酸或在这些值中的任何两个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可以是至多以下: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。维度标记可以包含约5个至约200个之间的核苷酸。维度标记可以包含约10个至约150个之间的核苷酸。维度标记可以包含长度在约20个至约125个之间的核苷酸。

[0122] 空间标记

[0123] 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,空间标记可以包含提供关于与条形码关联的靶分子的空间取向的信息的核酸序列。空间标记可以与样品中的坐标关联。坐标可以是固定的坐标。例如,坐标可以相对于基底固定。空间标记可以参考二维或三维网格。坐标可以相对于界标(landmark)固定。界标可在空间中被鉴定。界标可以是可被成像的结构。界标可以是生物结构,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(诸如色码、条形码、磁特性(magnetic property)、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。

[0124] 空间标记对于附接至给定固体支持物(例如,珠)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,同一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可以是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%或在这些值中的任何两个之间的数字或范围。在一些实施方案中,同一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或是至多60%、70%、80%、85%、90%、95%、97%、99%或100%。在一些实施方案中,同一固体支持物上至少60%的条形码可以包含相同的空间标记。在一些实施方案中,同一固体支持物上至少95%的条形码可以包含相同的空间标记。

[0125] 多于一个固体支持物(例如,珠)中可以呈现多达 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可以是约以下: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个核苷酸或在这些值中的任何两个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。空间标记可以包含约5个至约200个之间的核苷酸。空间标记可以包含约10个至约150个之间的核苷酸。空间标记可以包含长度在约20个至约125个之间的核苷酸。

[0126] 细胞标记

[0127] 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,细胞标记可以包含提供用于确定哪种靶核酸来源于哪种细胞的信息的核酸序列。在一些实施方案中,细胞标记对于附接至给定固体支持物(例如,珠)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,同一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可以是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%或在这些值中的任何两个之间的数字或范围。在一些实施方案中,同一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可以是约以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。例如,同一固体支持物上至少60%的条形码可以包含相同的细胞标记。作为另一种实例,同一固体支持物上至少95%的条形码可以包含相同的细胞标记。

[0128] 多于一个固体支持物(例如,珠)中可以呈现多达 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可以是约以下: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个核苷酸或在这些值中的任何两个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可以是至多以下: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,细胞标记可以包含约5个至约200个之间的核苷酸。作为另一种实例,细胞标记可以包含约10个至约150个之间的核苷酸。作为又另一种实例,细胞标记可以包含长度在约20个至约125个之间的核苷酸。

[0129] 条形码序列

[0130] 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,条形码序列可以包含为与条形码杂交的特定类型的靶核酸物质提供鉴定信息的核酸序列。条形码序列可以包含为与条形码(例如,靶结合区)杂交的靶核酸物质的特定出现提供计数器(例如,提供粗

略近似)的核酸序列。

[0131] 在一些实施方案中,将一组相异的(diverse)条形码序列附接至给定固体支持物(例如,珠)。在一些实施方案中,可以存在以下或可以存在约以下: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种独特分子标记序列或在这些值中的任何两个之间的数字或范围独特分子标记序列。例如,多于一种条形码可以包括约6561种具有不同序列的条形码序列。作为另一种实例,多于一种条形码可以包括约65536种具有不同序列的条形码序列。在一些实施方案中,可以存在至少以下或至多以下: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种独特条形码序列。独特分子标记序列可以附接至给定固体支持物(例如,珠)。

[0132] 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可以是约以下: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个核苷酸或在这些值中的任何两个之间的数字或范围的核苷酸。作为另一种实例,条形码的长度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、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。

[0133] 分子标记

[0134] 条形码(例如,随机条形码)可以包含一种或更多种分子标记。分子标记可以包含条形码序列。在一些实施方案中,分子标记可以包含为与条形码杂交的特定类型的靶核酸物质提供鉴定信息的核酸序列。分子标记可以包含为与条形码(例如,靶结合区)杂交的靶核酸物质的特定出现提供计数器的核酸序列。

[0135] 在一些实施方案中,将一组相异的分子标记附接至给定固体支持物(例如,珠)。在一些实施方案中,可以存在以下或可以存在约以下: 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之间的数字或范围独特分子标记序列。例如,多于一种条形码可以包括约6561种具有不同序列的分子标记。作为另一种实例,多于一种条形码可以包括约65536种具有不同序列的分子标记。在一些实施方案中,可以存在至少或至多 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。具有独特分子标记序列的条形码可以附接至给定固体支持物(例如,珠)。

[0136] 对于使用多于一种随机条形码的随机条形码化,不同分子标记序列的数目与任何靶的出现数目的比例可以是以下或可以是约以下: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物质。在一些实施方案中,不同分子标记序列的数目与任何靶的出现数目的比例是至少以下或是至多以下: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。

[0137] 分子标记的长度可以是以下或可以是约以下: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个核苷酸或在这些值中的任何两个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可以是至多以下: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。

[0138] 靶结合区

[0139] 条形码可以包含一个或更多个靶结合区,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,靶结合区可以包含与靶(例如,靶核酸、靶分子,例如待分析的细胞核酸)特异性杂交(例如,与特定基因序列特异性杂交)的核酸序列。在一些实施方案中,靶结合区可以包含可以衔接(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,靶结合区可以包含能够与限制性酶位点突出端(例如,EcoRI粘性末端突出端)特异性杂交的核酸序列。然后条形码可以连接至包含与限制性位点突出端互补的序列的任何核酸分子。

[0140] 在一些实施方案中,靶结合区可以包含非特异性靶核酸序列。非特异性靶核酸序列可以指可以独立于靶核酸的特定序列结合多于一种靶核酸的序列。例如,靶结合区可以包括随机多聚体序列或与mRNA分子上的多(A)尾杂交的寡(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,对于衔接至给定珠的所有条形码,靶结合区是相同的。在一些实施方案中,对于衔接至给定珠的多于一种条形码,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可以是约以下: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。

[0141] 在一些实施方案中,靶结合区可以包含寡(dT),所述寡(dT)可以与包含聚腺苷酸化末端的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可以是约以下: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、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,条形码在本文中 can 称为基因特异性条形码。

[0142] 定向特性(Orientation Property)

[0143] 随机条形码(例如,随机条形码)可以包含一种或更多种可以用于定向(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含不同的等电聚焦点。当将这些条形码引入样品中时,样品可以经历等电聚焦,以便于将条形码定向成已知的方式。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映射。示例性定向特性可以包括电泳迁移率(例如,基于条形码的尺寸)、等电点、自旋、电导率和/或自组装。例如,具有自组装的定向特性的条形码激活时可以自组装成特定的定向(例如,核酸纳米结构)。

[0144] 亲和特性(Affinity Property)

[0145] 条形码(例如,随机条形码)可以包含一种或更多种亲和特性。例如,空间标记可以包含亲和特性。亲和特性可以包括可以促进条形码与另一种实体(例如,细胞受体)结合的化学部分和/或生物部分。例如,亲和特性可以包括抗体,例如,对于样品上的特定部分(例

如,受体)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和/或特定细胞类型或分子附近的靶可以被标记(例如,被随机标记)。在一些实施方案中,亲和特性可以提供空间标记的核苷酸序列之外的空间信息,因为抗体可以将条形码引导至特定位置。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。抗体可以是裸抗体或融合抗体。

[0146] 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,特异性结合性)部分(如抗体片段)。

[0147] 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同一抗原结合。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、与细胞表面受体(CD8、CD34、CD45)结合的抗体和治疗性抗体。

[0148] 通用衔接子引物

[0149] 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遍及所有条形码的通用的核苷酸序列。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可以是约以下: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、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。

[0150] 接头

[0151] 当条形码包含多于一种类型的标记(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。在一些情况下,接头标记序列的长度是12个核苷酸。接头标记序列可以用于促进条形码的合成。接头标记可以包括错误校正(例如,汉明)码。

[0152] 固体支持物

[0153] 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。固体支持物可以是例如合成颗粒。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,第一条形序列)的分子标记)相差至少一个核苷酸。同一固体支持物上的条形码的细胞标记可以是相同的。不同的固体支持物上的条形码的细胞标记可以相差至少一个核苷酸。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相差至少一个核苷酸。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。合成颗粒可以是例如珠。

[0154] 珠可以是例如硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、Sephadex/琼脂糖凝胶珠、纤维素珠、聚苯乙烯珠或其任何组合。珠可以包括材料诸如聚二甲基硅氧烷 (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、胶乳、琼脂糖凝胶、纤维素、尼龙、硅酮或其任何组合。

[0155] 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物珠(例如可变形珠或凝胶珠)(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。凝胶珠可以例如通过将一种或更多种聚合物前体包封到液滴中来产生。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。

[0156] 在一些实施方案中,颗粒可以是可降解的。例如,聚合物珠可以例如在期望的条件下溶解、融化或降解。所期望的条件可以包括环境条件。所期望的条件可以导致聚合物珠以受控方式溶解、融化或降解。凝胶珠可以由于化学刺激、物理刺激、生物刺激、热刺激、磁刺激、电刺激、光刺激或其任何组合而溶解、融化或降解。

[0157] 例如,分析物和/或试剂(诸如寡核苷酸条形码)可以偶联/固定至凝胶珠的内表面(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/或凝胶珠的外表面或本文描述的任何其他微胶囊。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,本文描述的试剂与凝胶珠或任何其他微胶囊的偶联/固定可以是可逆的,诸如,例如经由不稳定型部分(例如,经由化学交联物,包括本文描述的化学交联物)。在施加刺激后,不稳定型部分可以被裂解并释放所固定的试剂。在一些实施方案中,不稳定型部分是二硫键。例如,在经由二硫键将寡核苷酸条形码固定至凝胶珠的情况下,使二硫键暴露于还原剂可以裂解二硫键并从珠释放寡核苷酸条形码。不稳定型部分可以作为凝胶珠或微胶囊的一部分、作为将试剂或分析物与凝胶珠或微胶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被包括。在一些实施方案中,多于一种条形码的至少一种条形码可以被固定在颗粒上、被部分固定在颗粒上、被包封在颗粒中、被部分包封在颗粒中或其任何组合。

[0158] 在一些实施方案中,凝胶珠可以包括宽范围的不同的聚合物,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、蛋白和/或塑料。聚合物可以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(双烯丙基二甲基-氯化铵)(PDADMAC)、聚(吡咯)(poly(pyrolle), PPy)、聚(乙烯基吡咯烷酮)(PVPON)、聚(乙烯基吡啶)(PVP)、聚(甲基丙烯酸)(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-羟基乙酸)(PLGA)。

[0159] 许多化学刺激可以用于触发珠的破坏、溶解或降解。这些化学改变的实例可以包括但不限于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。批量 (bulk) 改变也可以用于触发珠的破坏。

[0160] 通过各种刺激对微胶囊的批量或物理改变在设计胶囊以释放试剂方面也提供了许多优点。在宏观尺度上发生批量或物理改变,其中珠破裂是由刺激引起的机械-物理力的结果。这些过程可以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。

[0161] 生物刺激也可以用于触发珠的破坏、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,珠可以包含具有对特定蛋白酶的裂解敏感的肽交联的聚合物。更特别地,一种实例可以包括包含GFLGK肽交联的微胶囊。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并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。在另一种实例中,珠包括包含纤维素的壳壁。壳聚糖水解酶的添加用作纤维素键裂解、壳壁解聚及其内部内容物释放的生物触发物。

[0162] 还可以在施加热刺激后诱导珠释放其内容物。温度的改变可以引起珠的各种改变。热量的变化可以引起珠熔化,使得珠壁崩解。在其他情况下,热量可以增加珠内部组分的内部压力,使得珠破裂或爆炸。在又其他的情况下,热量可以使珠转化成收缩的脱水状态。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。

[0163] 将磁性纳米颗粒包括在微胶囊的珠壁中可以允许珠的触发破裂以及将珠引导成阵列。本公开内容的装置可以包括用于任一目的的磁珠。在一种实例中,将 Fe_3O_4 纳米颗粒掺入含聚电解质的珠中,在存在振荡磁场刺激的情况下触发破裂。

[0164] 珠也可以由于电刺激的结果被破坏、溶解或降解。与先前部分中描述的磁性颗粒类似,电敏珠可以允许珠的触发破裂以及其他功能,诸如电场中的对齐、电导率或氧化还原反应。在一种实例中,含电敏材料的珠在电场中对齐,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。

[0165] 也可以使用光刺激来破坏珠。许多光触发物是可能的,并且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,金属氧化物涂层可以用作胶囊触发物。涂覆有 SiO_2 的聚电解质胶囊的UV照射可以导致珠壁的崩解。在又另一种实例中,可以将可光切换材料(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,掺入光子开关(photon switch)产生在施加光触发物后可以崩解或变得更多孔的珠壁。

[0166] 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上之后,在框212处可以将珠引入微孔阵列的多于一个微孔上。每个微孔可以包含一个珠。珠可以包含多于一种条形码。条形码可以包含衔接至珠的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)、靶结合区或其任何组合。

[0167] 本文公开的条形码可以与固体支持物(例如,珠)关联(例如,衔接)。与固体支持物关联的条形码可各自包含选自以下组的条形码序列,该组包括至少100种或1000种具有独特序列的条形码序列。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含具有不同序列的条形码。在一些实施方案中,与固体支持物关联的条形码的一定百分比包含相同的细胞标记。例如,百分比可以是以下或可以是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%或在这些值中的任何两个之间的数字或范围。作为另一个实例,所

述百分比可以是至少以下或可以是至多以下：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下组的不同的细胞标记，该组包括至少100种或1000种具有独特序列的细胞标记。

[0168] 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种合成颗粒的固体支持物将样品中的多于一种靶条形码化。在一些实施方案中,固体支持物可以包括与多于一种条形码关联的多于一个合成颗粒。不同固体支持物上的多于一种条形码的空间标记可以相差至少一个核苷酸。固体支持物可以例如在二维或三维包括多于一种条形码。合成颗粒可以是珠。珠可以是硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、Sephadex/琼脂糖凝胶珠、纤维素珠、聚苯乙烯珠或其任何组合。固体支持物可以包括聚合物、基质、水凝胶、针阵列装置、抗体或其任何组合。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,固体支持物可以包埋到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。条形码可以与基底关联。

[0169] 如本文使用的,术语“拴系的”、“附接的”和“固定的”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方式。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原位固相合成条形码。

[0170] 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、圆柱体或可以固定核酸(例如,共价地或非共价地)的其他类似配置。珠可以由例如塑料、陶瓷、金属、聚合物材料或其任何组合构成。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,所述形状诸如立方体、长方形、锥形、圆柱形、圆锥形、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。

[0171] 珠可以包含包括但不限于以下的各种材料:顺磁材料(例如,镁、钼、锂和钽)、超顺磁材料(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、琼脂糖凝胶、琼脂糖、水凝胶、聚合物、纤维素、尼龙或其任何组合。

[0172] 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。

[0173] 本文公开的一些实施方案包括一个或更多个颗粒(例如,珠)。每个颗粒可以包含多于一种寡核苷酸(例如,条形码)。多于一种寡核苷酸中的每一种可以包含条形码序列(例如,分子标记序列)、细胞标记和靶结合区(例如,寡(dT)序列、基因特异性序列、随机多聚体或其组合)。多于一种寡核苷酸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目可以是不同的。在一些实施方案中,细胞标记序列的数目可以是以下或可以是约以下:10种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种、10000种、20000种、30000种、40000种、50000种、60000种、70000种、80000种、90000种、100000种、 10^6 种、 10^7 种、 10^8 种、 10^9 种、在这些值中的任何两个之间的数字或范围或更多。在一些实施方案中,细胞标记序列的数目可以是至少以下或可以是至多以下:10种、100种、200种、300种、

400种、500种、600种、700种、800种、900种、1000种、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种、10000种、20000种、30000种、40000种、50000种、60000种、70000种、80000种、90000种、100000种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。在一些实施方案中，多于一个颗粒中不多于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、200个、300个、400个、500个、600个、700个、800个、900个、1000个或更多个包含具有相同细胞序列的寡核苷酸。在一些实施方案中，包含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，多于一个颗粒全都不具有相同的细胞标记序列。

[0174] 在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,条形码序列的数目可以是以下或可以是约以下:10种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种、10000种、20000种、30000种、40000种、50000种、60000种、70000种、80000种、90000种、100000种、 10^6 种、 10^7 种、 10^8 种、 10^9 种或在这些值中的任何两个之间的数字或范围。在一些实施方案中,条形码序列的数目可以是至少以下或可以是至多以下:10种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种、10000种、20000种、30000种、40000种、50000种、60000种、70000种、80000种、90000种、100000种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。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。作为另一种实例,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、这些值中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多于一种包含条形码的颗粒。在一些实施方案中,待标记的靶和不同条形码序列的出现(或拷贝或数目)的比例可以是至少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或更高。在一些实施方案中,多于一种寡核苷酸的每一种还包含样品标记、通用标记或二者。颗粒可以是例如纳米颗粒或微米颗粒。

[0175] 珠的尺寸可以不同。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可以是约以下: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微米或在这些值中的任何两个之间的数字或范围。

[0176] 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以下或者约以下:10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或在这些值中的任何两个之间的数字或范围。珠的直径可以与细胞(例如,被基底的孔捕获的单细胞)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,被基底的孔捕获的单细胞)的直径相关。在一些实施方案中,珠的直径可以比细胞的直径长或短以下或者约以下:10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或在这些值中的任何两个之间的数字或范围。在一些实施方案中,珠的直径可以比细胞的直径长或短至少或至多10%、20%、30%、40%、50%、60%、70%、80%、

90%、100%、150%、200%、250%或300%。

[0177] 珠可以附接至基底和/或包埋到基底中。珠可以附接至凝胶、水凝胶、聚合物和/或基质和/或包埋到凝胶、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空间标记来鉴定,该空间标记可以用作位置地址。

[0178] 珠的实例可以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、抗体缀合的珠(例如,抗免疫球蛋白微珠)、蛋白A缀合的珠、蛋白G缀合的珠、蛋白A/G缀合的珠、蛋白L缀合的珠、寡(dT)缀合的珠、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。

[0179] 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于一个光学通道中发荧光。珠可以与氧化铁或氧化铬关联,使其成为顺磁性或铁磁性。珠可以是可鉴定的。例如,可以使用照相机对珠成像。珠可以具有与珠关联的可检测代码。例如,珠可以包含条形码。珠可以改变尺寸,例如由于在有机溶液或无机溶液中溶胀。珠可以是疏水的。珠可以是亲水的。珠可以是生物相容的。

[0180] 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)可以蚀刻有标识符(例如,数字)。标识符可以通过对珠成像来可视化。

[0181] 固体支持物可以包括可溶性、半溶性或不溶性材料。当固体支持物包括接头、支架、构建模块(building block)或其他与其附接的反应性部分时,固体支持物可以被称为“官能化的”,而当固体支持物缺少这种与其附接的反应性部分时,固体支持物可以被称为“非官能化的”。固体支持物可以以溶液中游离,诸如在微量滴定孔中的形式;以流通形式,诸如在柱中;或以浸量尺(dipstick)使用。

[0182] 固体支持物可以包括膜、纸(paper)、塑料、涂覆表面、平坦表面、玻璃、载玻片、芯片或其任何组合。固体支持物可以采取树脂、凝胶、微球或其他几何配置的形式。固体支持物可以包括二氧化硅芯片、微米颗粒、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),和/或晶片、梳、针或针头(例如,适于组合合成或分析的针阵列)或珠,平坦表面诸如晶片(例如,硅晶片)的凹陷或纳升孔阵列,具有凹陷的晶片(具有或不具有过滤器底部)。

[0183] 固体支持物可以包括聚合物基质(例如,凝胶、水凝胶)。聚合物基质可以能够渗透细胞内空间(例如,细胞器周围)。聚合物基质可以能够被泵送到整个循环系统。

[0184] 基底和微孔阵列

[0185] 如本文使用的,基底可以指固体支持物类型。基底可以指可以包含本公开内容的条形码或随机条形码的固体支持物。基底可以例如包括多于一个微孔。基底可以例如是包括两个或更多个微孔的孔阵列。在一些实施方案中,微孔可以包括定义体积的小的反应室。在一些实施方案中,微孔可以捕获一个或更多个细胞。在一些实施方案中,微孔可以仅捕获一个细胞。在一些实施方案中,微孔可以捕获一个或更多个固体支持物。在一些实施方案中,微孔可以仅捕获一个固体支持物。在一些实施方案中,微孔捕获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。

[0186] 条形码化的方法

[0187] 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)中不同位置处的不同靶的数目的方法。该方法可以包括将条形码(例如,随机条形码)紧密接近样品放置,裂解样品,将不同的靶与条形码关联,对靶进行扩增和/或对靶进行数字计数。该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,该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映射图上可以包括产生样品的二维映射图或三维映射图。可以在将样品中的多于一种靶条形码化(例如,随机条形码化)之前或之后产生二维映射图和三维映射图。将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映射图上。将多于一种靶映射到样品的映射图上可以包括产生样品的二维映射图或三维映射图。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映射图和三维映射图。在一些实施方案中,可以在裂解样品之前或之后生成二维映射图和三维映射图。在产生二维映射图或三维映射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或其任何组合。

[0188] 在一些实施方案中,将多于一种靶条形码化包括将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,随机条形码化的靶)。将多于一种靶条形码化可以包括产生条形码化靶的索引文库。产生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)的固体支持物来进行。

[0189] 使样品和条形码接触

[0190] 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。细胞可以例如通过重力流来接触,其中可以使细胞沉淀并且产生单层。样品可以是组织薄切片。可以将薄切片放置于基底上。样品可以是一维的(例如,形成平坦表面)。可以使样品(例如,细胞)分散遍及基底,例如,通过在基底上生长/培养细胞。

[0191] 当条形码紧密接近靶时,靶可以与条形码杂交。条形码可以按不可耗尽的比例接触,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。

[0192] 细胞裂解

[0193] 在细胞和条形码的分配之后,可以将细胞裂解以释放靶分子。细胞裂解可以通过各种手段中的任何一种来完成,例如通过化学或生化手段,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、Tween-20或NP-40)、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)或其任何组合的细胞裂解缓冲液来裂解细胞。为了增加靶与条形码的关联,可以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。

[0194] 在一些实施方案中,可以使用滤纸来裂解样品。可以在滤纸上部用裂解缓冲液浸泡滤纸。可以将滤纸用压力施加至样品,这可以促进样品的裂解以及样品的靶与基底的杂交。

[0195] 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/或化学裂解来

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至基底来进行。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是约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0.1M、0.5M或1M或更高。裂解缓冲液中的盐浓度可以是至多约0.1M、0.5M或1M或更高。在一些实施方案中,裂解缓冲液中的盐浓度是约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、tween、NP-40)。裂解缓冲液中的去污剂浓度可以是至少约0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、3%、4%、5%、6%或7%或更高。裂解缓冲液中的去污剂浓度可以是至多约0.0001%、0.0005%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、3%、4%、5%、6%或7%或更高。在一些实施方案中,裂解缓冲液中的去污剂浓度是约1%的十二烷基硫酸锂。裂解方法中使用的的时间可以取决于所使用的去污剂的量。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)。裂解缓冲液中的螯合剂浓度可以是至少约1mM、5mM、10mM、15mM、20mM、25mM或30mM或更高。裂解缓冲液中的螯合剂浓度可以是至多约1mM、5mM、10mM、15mM、20mM、25mM或30mM或更高。在一些实施方案中,裂解缓冲液中的螯合剂浓度是约10mM。裂解缓冲液可以包含还原剂(例如, β -巯基乙醇、DTT)。裂解缓冲液中的还原剂浓度可以是至少约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,裂解缓冲液中的还原剂浓度是约5mM。在一些实施方案中,裂解缓冲液可以包含约0.1M Tris HCl,约pH 7.5,约0.5M LiCl,约1%十二烷基硫酸锂,约10mM EDTA和约5mM DTT。

[0196] 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分钟或更多分钟。裂解的细胞可以包含至少约100000种、200000种、300000种、400000种、500000种、600000种或700000种或更多种靶核酸分子。裂解的细胞可以包含至多约100000种、200000种、300000种、400000种、500000种、600000种或700000种或更多种靶核酸分子。

[0197] 将条形码附接至靶核酸分子

[0198] 在细胞裂解和核酸分子从细胞释放之后,核酸分子可以与共定位的固体支持物的条形码随机关联。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(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,可以将裂解的细胞释放的核酸分子与基底上的多于一种探针关联(例如,与基底上的探针杂交)。当探针包含寡(dT)时,可以将mRNA分子与探针杂交并且逆转录。寡核苷酸的寡(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中框216处图示的条形码化的非限制性实例中,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。

[0199] 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

区可以包含可以能够与限制性位点突出端(例如,EcoRI粘性末端突出端)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至包含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,T4 DNA连接酶)可以用于连接两个片段。

[0200] 例如,在图2中框220处图示的条形码化的非限制性实例中,随后可以将来自多于一个细胞(或多于一个样品)的标记的靶(例如,靶-条形码分子)汇集至例如管中。标记的靶可以通过例如回收(retrieving)条形码和/或附接靶-条形码分子的珠来汇集。

[0201] 可以通过使用磁珠和外部施加的磁场来实现附接的靶-条形码分子的基于固体支持物的集合的回收。汇集靶-条形码分子后,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,不先汇集来自多于一个细胞的标记的靶核酸分子。

[0202] 逆转录

[0203] 本公开内容提供了使用逆转录(例如,在图2的框224处)来产生靶-条形码缀合物的方法。靶-条形码缀合物可以包含条形码以及靶核酸的全部或一部分的互补序列(即,条形码化的cDNA分子,诸如随机条形码化的cDNA分子)。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而发生。逆转录引物可以是寡(dT)引物、随机六核苷酸引物或靶特异性寡核苷酸引物。寡(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,并且与哺乳动物mRNA的3'端的内源多(A)尾结合。随机六核苷酸引物可以在各个互补位点处与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。

[0204] 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,逆转录引物是寡(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(dT)引物的长度是12-18个核苷酸,并且与哺乳动物mRNA的3'端的内源多(A)尾结合。随机六核苷酸引物可以在各个互补位点处与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。

[0205] 逆转录可以重复地发生以产生多于一个标记的cDNA分子。本文公开的方法可以包括进行至少约1次、2次、3次、4次、5次、6次、7次、8次、9次、10次、11次、12次、13次、14次、15次、16次、17次、18次、19次或20次逆转录反应。该方法可以包括进行至少约25次、30次、35次、40次、45次、50次、55次、60次、65次、70次、75次、80次、85次、90次、95次或100次逆转录反应。

[0206] 扩增

[0207] 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)以产生标记的靶核酸分子的多于一个拷贝。扩增可以以多重化方式进行,其中多于一种靶核酸序列同时进行扩增。扩增反应可以用于向核酸分子添加测序衔接子。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)、靶核酸或其组合的至少一部分。扩增反应可以包括扩增多于一种核酸的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%或在这些值中

的任何两个之间的范围或数字。该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。

[0208] 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的互补链的引物同时延伸使特定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衍生形式,包括但不限于RT-PCR、实时PCR、巢式PCR、定量PCR、多重化PCR、数字PCR和装配PCR。

[0209] 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。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合成和转录的多于一个循环以扩增DNA或RNA靶、连接酶链式反应(LCR)和Q β 复制酶(Q β)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。

[0210] 在一些实施方案中,本文公开的方法还包括对标记的核酸(例如,标记的RNA、标记的DNA、标记的cDNA)进行聚合酶链式反应以产生标记的扩增子(例如,随机标记的扩增子)。标记的扩增子可以是双链分子。双链分子可包括双链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)。标记的扩增子可以是单链分子。单链分子可以包括DNA、RNA或其组合。本公开内容的核酸可以包括合成的或改变的核酸。

[0211] 扩增可以包括使用一种或更多种非天然核苷酸。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)、吗啉代和锁核酸(LNA)以及乙二醇核酸(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。添加非天然核苷酸可以用于鉴定扩增反应中特定循环或时间点的产物。

[0212] 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。一种或更多种引物可以包括至少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含少于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标记的靶的3'端或5'端。一种或更多种引物可以退火至多于一种标记的靶的内部区域。内部区域可以与多于一种标记的靶的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。

[0213] 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)、靶或其任何组合。一种或更多种引物可以包括通用引物和定制引物。定制引物可以设计成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标记靶的子集。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标记的核酸。两种或更多种不同的标记的核酸可以对应于一种或更多种基因。

[0214] 可以在本公开内容的方法中使用任何扩增方案。例如,在一种方案中,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序列的引物来扩增衔接至珠的分子。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产物变成Illumina测序文库。使用150bp×2测序的测序可以揭示读段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。

[0215] 在一些实施方案中,可以使用化学裂解将核酸从基底去除。例如,存在于核酸中的化学基团或修饰的碱基可以用于促进将核酸从固体支持物去除。例如,酶可以用于将核酸从基底去除。例如,通过限制性核酸内切酶消化可以将核酸从基底去除。例如,用尿嘧啶-d-糖苷酶(UDG)处理含dUTP或ddUTP的核酸可以用于将核酸从基底去除。例如,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(apurinic/aprimidinic, AP)核酸内切酶)将核酸从基底去除。在一些实施方案中,可以使用可光裂解基团以及光将核酸从基底去除。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/抗生物素蛋白、生物素/链霉抗生物素蛋白、生物素/中性抗生物素蛋白、Ig蛋白A、光不稳定型接头、酸或碱不稳定型接头基团或适配体。

[0216] 当探针是基因特异性时,可以将分子与探针杂交,并且逆转录和/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,核酸可以被扩增。扩增可以以多重方式进行,其中多种靶核酸序列同时扩增。扩增可以将测序衔接子添加至核酸。

[0217] 在一些实施方案中,可以例如用桥接扩增在基底上进行扩增。cDNA可以加同聚物尾,以便产生相容末端,用于使用基底上的寡(dT)探针进行桥接扩增。在桥接扩增中,与模板核酸的3'末端互补的引物可以是共价地衔接至固体颗粒的每对引物中的第一引物。当包含模板核酸的样品与颗粒接触并进行单个热循环时,可以将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向前延长以形成双链体分子,所述双链体分子由模板分子和与模板互补的新形成的DNA链构成。在下一循环的加热步骤中,双链体分子可以变性,从颗粒释放模板分子并且留下通过第一引物衔接至颗粒的互补DNA链。在随后的退火和延长步骤的退火阶段中,互补链可以与第二引物杂交,第二引物在从第一引物去除的位置处与互补链的区段互补。这种杂交可导致互补链在第一引物和第二引物之间形成桥,通过共价键连接第一引物并通过杂交连接第二引物。在延长阶段,在同一反应混合物中通过添加核苷酸,第二引物可以在反向方向上延长,从而将桥转化为双链桥。然后开始下一个循环,并且

双链桥可以变性以产生两个单链核酸分子,每个单链核酸分子具有的一个末端分别经由第一引物和第二引物附接至颗粒表面,其中每个单链核酸分子的另一个末端是未附接的。在这第二个循环的退火和延长步骤中,每条链可以与同一颗粒上先前未使用的另外的互补引物杂交,以形成新的单链桥。现在杂交的两个先前未使用的引物延长从而将两个新的桥转换成双链桥。

[0218] 扩增反应可以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%。

[0219] 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。标记的核酸的扩增可以包括对标记的核酸的指数式扩增。标记的核酸的扩增可以包括对标记的核酸的线性扩增。扩增可以通过聚合酶链式反应(PCR)来进行。PCR可以指用于通过DNA的互补链的引物同时延伸使特定DNA序列体外扩增的反应。PCR可以涵盖反应的衍生形式,包括但不限于,RT-PCR、实时PCR、巢式PCR、定量PCR、多重化PCR、数字PCR、抑制PCR、半抑制PCR以及装配PCR。

[0220] 在一些实施方案中,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合成和转录的多于一个循环以扩增DNA或RNA靶、连接酶链式反应(LCR)、 $\Phi\beta$ 复制酶($\Phi\beta$)、回文探针的使用、链置换扩增、使用限制性内切核酸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'外切核酸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)。

[0221] 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。可选地,扩增子可以是单链分子。单链分子可以包括DNA、RNA或其组合。本发明的核酸可以包括合成的或改变的核酸。

[0222] 在一些实施方案中,该方法包括反复扩增标记的核酸以产生多于一个扩增子。本文公开的方法可以包括进行至少约1次、2次、3次、4次、5次、6次、7次、8次、9次、10次、11次、12次、13次、14次、15次、16次、17次、18次、19次或20次扩增反应。可选地,该方法包括进行至少约25次、30次、35次、40次、45次、50次、55次、60次、65次、70次、75次、80次、85次、90次、95次或100次扩增反应。

[0223] 扩增可以还包括将一种或更多种对照核酸添加至一个或更多个包含多于一种核酸的样品中。扩增可以还包括将一种或更多种对照核酸添加至多于一种核酸。对照核酸可以包含对照标记。

[0224] 扩增可以包括使用一种或更多种非天然核苷酸。非天然核苷酸可以包括光不稳定型和/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)、吗啉代和锁核酸(LNA)以及乙二醇核酸(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。添加非天然核苷酸可以用于鉴定扩增反应中特定循环或时间点的产

物。

[0225] 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一种或更多种寡核苷酸。一种或更多种寡核苷酸可以包含至少约7-9个核苷酸。一种或更多种寡核苷酸可以包含少于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标记的核酸的3'端和/或5'端。一种或更多种引物可以退火至多于一种标记的核酸的内部区域。内部区域可以与多于一种标记的核酸的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、核酸或其产物。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更多个样品中总核酸的子集。在一些实施方案中,引物是与本公开内容的阵列附接的探针。

[0226] 在一些实施方案中,将样品中的多于一种靶条形码化(例如,随机条形码化)还包括产生条形码化靶(例如,随机条形码化靶)或靶的条形码化片段的索引文库。不同的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。产生条形码化靶的索引文库包括从样品中的多于一种靶产生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相差以下、相差约以下、相差至少以下或相差至多以下: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个核苷酸或在这些值中的任何两个之间的数字或范围的核苷酸。在一些实施方案中,产生条形码化靶的索引文库包括使多于一种靶(例如mRNA分子)与包含多(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单链标记的cDNA分子(每种包含cDNA区和标记区),其中多于一种靶包括至少两种不同序列的mRNA分子,且多于一种寡核苷酸包括至少两种不同序列的寡核苷酸。产生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,该方法可以包括产生衔接子标记的扩增子。

[0227] 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标记单种核酸(例如,DNA或RNA)分子。在一些实施方案中,其包括在从mRNA产生cDNA分子时将DNA条形码或标签添加至cDNA分子。可以进行巢式PCR以使PCR扩增偏倚最小化。可以添加用于测序(例如下一代测序(NGS))使用的衔接子。例如在图2的框232处,可以使用测序结果来确定靶的一个或更多个拷贝的细胞标记、分子标记和核苷酸片段的序列。

[0228] 图3是示出了产生条形码化靶(例如,随机条形码化靶)的索引文库,诸如条形码化的mRNA或其片段的索引文库的非限制性示例性过程的示意图。如步骤1中示出的,逆转录过

程可以用独特分子标记、细胞标记和通用PCR位点对每个mRNA分子进行编码。具体地,通过将一组条形码(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逆转录以产生标记的cDNA分子304(包括cDNA区306)。条形码310中的每一个可以包括靶结合区,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。

[0229] 在一些实施方案中,细胞标记可以包含3个至20个核苷酸。在一些实施方案中,分子标记可以包含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一种还包括通用标记和细胞标记中的一种或更多种,其中通用标记对于固体支持物上的多于一种随机条形码是相同的,并且细胞标记对于固体支持物上的多于一种随机条形码是相同的。在一些实施方案中,通用标记可以包含3个至20个核苷酸。在一些实施方案中,细胞标记包含3个至20个核苷酸。

[0230] 在一些实施方案中,标记区314可以包含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个核苷酸,或在这些值中的任何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个核苷酸,或在这些值中的任何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个核苷酸,或在这些值中的任何之间的数字或范围的核苷酸。通用标记对于固体支持物上的多于一种随机条形码可以是相同的,并且细胞标记是对于固体支持物上的多于一种随机条形码相同的。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个核苷酸或在这些值中的任何之间的数字或范围的核苷酸。

[0231] 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、200个、300个、400个、500个、600个、700个、800个、900个、1000个不同标记或在这些值中的任何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个核苷酸或在这些值中的任何之间的数字或范围的核苷酸。一组条形码或随机条形码310可以包括以下、包括约以下、包括至少以下或可以是至多以下:10个、20个、40个、50个、70个、80个、90个、 10^2 个、 10^3 个、 10^4 个、 10^5 个、 10^6 个、 10^7 个、 10^8 个、 10^9 个、 10^{10} 个、 10^{11} 个、 10^{12} 个、 10^{13} 个、 10^{14} 个、 10^{15} 个、 10^{20} 个条形码或随机条形码310或在这些值中的任何之间的数字或范围的条形码或随机条形码310。并且条形码或随机条形码310的组可以例如,各自包含独特标记区314。标记的cDNA分子304可以进行纯化以去除过量的条形码或随机条形码310。纯化可以包括Ampure珠纯化。

[0232] 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到1个管中,且用第1PCR引物池和第1通用PCR引物进行PCR扩增。因为独特标记区314,汇集是可能的。特别地,可以将标记的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。扩增可以包括以单一反应体积用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} 种多重引物或在这些值中的任何之间的数字或范围的多重引物。扩增可以包括使用第1PCR引物池324,所述第1PCR引物池324包括靶向特定基因的定制引物326A-C和通用引物328。定制引物326可以与标记的cDNA分子304的cDNA部分306'内的区域杂交。通用引物328可以与标记的cDNA分子304的通用PCR区域316杂交。

[0233] 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引物扩增。巢式PCR可以使PCR扩增倚靠最小化。特别地,巢式PCR标记的扩增子322可通过巢式PCR进行进一步扩增。巢式PCR可以包括在单一反应体积中用巢式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可以包含以下、包含约以下、包含至少以下或包含至多以下: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种不同的巢式PCR引物332或在这些值中的任何之间的数字或范围的不同的巢式PCR引物332。巢式PCR引物332可以包含衔接子334,并与标记的扩增子322的cDNA部分306''内的区域杂交。通用引物328'可以包含衔接子336,并与标记的扩增子322的通用PCR区域316杂交。由此,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'可以不包含衔接子334和衔接子336。而是,衔接子334和衔接子336可以连接至巢式PCR的产物以产生衔接子标记的扩增子338。

[0234] 如步骤4中示出的,可以使用文库扩增引物将来自步骤3的PCR产物进行PCR扩增用于测序。特别地,可以使用衔接子334和衔接子336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和衔接子336可以与引物340和引物342杂交。一种或更多种引物340和引物342可以是PCR扩增引物。一种或更多种引物340和引物342可以是测序引物。一种或更多种衔接子334和衔接子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和衔接子336可以用于对衔接子标记的扩增子338测序。引物342可以包含板索引344,使得使用同一组条形码或随机条形码310产生的扩增子可以使用下一代测序(NGS)在一个测序反应中测序。

[0235] 包含与寡核苷酸关联的细胞组分结合试剂的组合物

[0236] 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。已在美国专利申请公布第US2018/0088112号和美国专利申请公布第US2018/0346970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);这些中的每一项的内容通过引用以其整体并入本文。

[0237] 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结合。例如,细胞组分结合试剂的结合靶可以是以下或包括以下:碳水化合物、脂质、蛋白、细胞外蛋白、细胞

表面蛋白、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、细胞内蛋白或其任何组合。在一些实施方案中，细胞组分结合试剂(例如，蛋白结合试剂)能够与抗原靶或蛋白靶特异性结合。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、样品标记或其任何组合。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。在一些实施方案中，多(A)尾的长度可以是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。

[0238] 独特标识符可以是，例如，具有任何合适长度例如约4个核苷酸至约200个核苷酸的核苷酸序列。在一些实施方案中，独特标识符是长度为25个核苷酸至约45个核苷酸的核苷酸序列。在一些实施方案中，独特标识符可以具有是以下、是约以下、小于以下、大于以下的长度：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个核苷酸或以上值中的任何两个之间的范围。

[0239] 在一些实施方案中，独特标识符选自一组相异的独特标识符。一组相异的独特标识符可以包括以下或可以包括约以下：20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、2000个、5000个或在这些值中的任何两个之间的数字或范围的不同的独特标识符。一组相异的独特标识符可以包括至少以下或包括至多以下：20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、2000个或5000个不同的独特标识符。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个核苷酸或在这些值中的任何两个之间的数字或范围。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、10个核苷酸。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3%、至少5%、至少8%、至少10%、至少15%、至少20%或更多。

[0240] 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。

[0241] 本公开内容中设想了任何合适的细胞组分结合试剂，诸如蛋白结合试剂、抗体或其片段、适配体、小分子、配体、肽、寡核苷酸等，或其任何组合。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)或其片段，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可以包括约以

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之间的数字或范围的不同细胞组分试剂。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可以包括至多以下: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、5000种不同的细胞组分试剂。

[0242] 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。接头可以是例如能够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至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、链霉抗生物素蛋白、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。可以使用商购可得的缀合试剂盒,诸如Protein-Oligo缀合试剂盒(Solulink, Inc., San Diego, California)、Thunder-Link® oligo缀合系统(Innova Biosciences, Cambridge, United Kingdom)等将寡核苷酸与细胞组分结合试剂缀合。

[0243] 寡核苷酸可以与细胞组分结合试剂(例如,蛋白结合试剂)的任何合适的位点缀合,条件为寡核苷酸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,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,寡核苷酸可以与抗体在除了抗原结合位点之外的任何地方(例如Fc区、C_H1结构域、C_H2结构域、C_H3结构域、C_L结构域等)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,其内容特此通过引用以其整体明确并入。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,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、1000种寡核苷酸分子,其中每种寡核苷酸分子包含相同或不同的独特标识符。

[0244] 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、受体或其任何组合。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

可以包括细胞内细胞组分。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)的以下或是约以下:1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或在这些值中的任何两个之间的数字或范围。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)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可以包括约以下:2、3、4、5、10、20、30、40、50、100、1000、10000种或在这些值中的任何两个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可以包括至多以下:2、3、4、5、10、20、30、40、50、100、1000、10000种不同的细胞组分靶。

[0245] 图4示出了与包含用于抗体的独特标识符序列的寡核苷酸关联(例如,缀合)的示例性细胞组分结合试剂(例如,抗体)的示意图。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可以被称为抗体寡核苷酸(缩写为“AbOligo”或“Ab0”)。寡核苷酸还可以包含另外的组分,所述另外的组分包括但不限于一种或更多种接头、用于抗体的一种或更多种独特标识符、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。

[0246] 图5示出了与包含用于抗体的独特标识符序列的寡核苷酸关联(例如,缀合)的示例性细胞组分结合试剂(例如,抗体)的示意图。细胞组分结合试剂可以能够与至少一种细胞组分靶(诸如抗原靶或蛋白靶)特异性结合。结合试剂寡核苷酸(例如,样品索引寡核苷酸或抗体寡核苷酸)可以包含用于进行本公开内容方法的序列(例如,样品索引序列)。例如,样品索引寡核苷酸可以包含样品索引序列,用于鉴定样品的一个或更多细胞的样品来源。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两种细胞组分结合试剂的组合物(例如样品索引组合物)的索引序列(例如样品索引序列)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(或可以为)能够从细胞组分结合试剂脱离或不能够从细胞组分结合试剂脱离。

[0247] 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)尾或其组合。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寡核苷酸的序列可以包含多(A)尾。样品索引寡核苷酸可以包含分子标记。

[0248] 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个核苷酸或在这些值中的任何两个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长度为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个或1000个核苷酸。

[0249] 在一些实施方案中,细胞组分结合试剂包括抗体、四聚体、适配体、蛋白支架或其组合。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至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、链霉抗生物素蛋白、生物素、胺及其任何组合。

[0250] 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、CLDND1、HLA-ABC、ICAM1、ITFG3、MPZL1、NA K ATP酶 α 1、ATP1A1、NPTN、PMCA ATP酶、ATP2B1、SLC1A5、SLC29A1、SLC2A1、SLC44A2或其任何组合。

[0251] 在一些实施方案中,蛋白靶是或包括细胞外蛋白、细胞内蛋白或其任何组合。在一些实施方案中,抗原或蛋白靶是或包括细胞表面蛋白、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,或其任何组合。抗原或蛋白靶可以是或包括脂质、碳水化合物或其任何组合。蛋白靶可以选自包含许多蛋白靶的组。抗原靶或蛋白靶的数目可以是以下或可以是约以下: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、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种、10000种或在这些值中的任何两个之间的数字或范围。蛋白靶的数目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、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种或10000种。

[0252] 细胞组分结合试剂(例如,蛋白结合试剂)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目可以是不同的。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目可以是以下或可以是约以下: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种或在这些值中的任何两个之间的数字或范围。在一些实施方案中,结合试剂寡核苷酸的数目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、200种、300种、400种、500种、600种、700种、800种、900种或1000种。

[0253] 多于一种包含细胞组分结合试剂的组合物(例如,多于一种样品索引组合物)可以包含一种或更多种未与结合试剂寡核苷酸(诸如样品索引寡核苷酸)缀合的另外的细胞组分结合试剂,其在本文也被称为不含结合试剂寡核苷酸的细胞组分结合试剂(诸如不含样品索引寡核苷酸的细胞组分结合试剂)。在不同实施方式中,多于一种组合物中的另外的细胞组分结合试剂的数目可以是不同的。在一些实施方案中,另外的细胞组分结合试剂的数目可以是以下或可以是约以下: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种或在这些值中的任何两个之间的数字或范围。在一些实施方案中,另外的细胞组分结合试剂的数目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种或100种。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。

[0254] 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,样品索引寡核苷酸)缀合的一种或更多种细胞组分结合试剂,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。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目,(2)不与结合试剂寡核苷酸(例如样品索引寡核苷酸)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)的数目。在一些实施方案中,比率可以是以下或可以是约以下: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或在所述值中的任何两个之间的数字或范围。在一些实施方案中,比率可以是至少以下或可以是至多以下: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。

[0255] 在一些实施方案中,比率可以是以下或可以是约以下: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或在所述值中的任何两个之间的数字或范围。在一些实施方案中,比率可以是至少以下或可以是至多以下: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。

[0256] 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可以是约以下: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%或在这

些值中的任何两个之间的数字或范围。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可以是至多以下: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%。

[0257] 在一些实施方案中,在包含与结合试剂寡核苷酸(例如,样品索引寡核苷酸)缀合的细胞组分结合试剂和与样品索引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可以是约以下: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%或在这些值中的任何两个之间的数字或范围。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可以是至多以下: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%。

[0258] 细胞组分混合物(Cocktails)

[0259] 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。

[0260] 在一些实施方案中,多于一种组合物(例如,样品索引组合物)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都与结合试剂寡核苷酸(例如,样品索引寡核苷酸)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。

[0261] 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)的数目可以不同。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目可以不同。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目可以是以下或可以是约以下: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、10000种、100000种或在这些值中的任何两个之间的数字或范围。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目可以是至少以下或可以是至多以下: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、10000种或100000种。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,样品索引序列)的结合试剂寡核苷酸缀合。

[0262] 细胞组分靶的定量分析方法

[0263] 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或包括抗体寡核苷酸、样品索引寡核苷酸、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。

[0264] 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。

[0265] 在一些实施方案中,多于一种细胞组分结合试剂的结合靶(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、细胞外蛋白、细胞表面蛋白、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、细胞内蛋白或其任何组合。在一些实施方案中,细胞组分靶是蛋白靶。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、受体或其任何组合。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。

[0266] 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。在样品包含细胞的实施方案中,可以去除不与细胞特异性结合的任何细胞组分结合试剂。

[0267] 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,使得稀释的群体的细胞数目是基底上的孔的数目的以下或至少以下: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,使得稀释的群体的细胞数目是基底上的孔的数目的以下或至多以下: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,使得细胞的数目是基底中的孔的数目的约10%。

[0268] 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多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,使得基底的一个孔仅容纳一个细胞。

[0269] 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。

[0270] 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,在样品包含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特异性结合的实施方案中,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结合的组合物与细胞一起保留。

[0271] 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、样品标记等或其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种寡核苷酸。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。

[0272] 在一些实施方案中,多于一种寡核苷酸探针可以被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被固定在固体支持物上。当多于一种寡核苷酸探针紧密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。

[0273] 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、酶处理或其任何组合。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。

[0274] 细胞组分和核酸靶的同时定量分析方法

[0275] 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。美国专利申请公布第US2018/0088112号和美国专利申请公布第US2018/0346970号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;这些美国专利申请的每一项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(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)的混合物或来自不同受试者的细胞的混合物。

[0276] 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。

[0277] 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、受体或其任何组合。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是生物体或生物体的一个或多个细胞中的所有细胞组分靶(诸如,表达的蛋白)的以下或可以是约以下:1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%、99%或在这些值中的任何两个之间的数字或范围。在一些实施方案中,多于一种细胞组分靶可以是可以在生物体或生物体的一个或多个细胞中表达的所有细胞组分(诸如,蛋白)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可以包括约以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种、10000种,或在这些值中的任何两个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可以包括至多以下: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。

[0278] 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。在样品包含细胞的实施方案中,可以去除不与细胞特异性结合的任何细胞组分结合试剂。

[0279] 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目是以下或是至少以下:基底上的孔的数目的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目是以下或是至多以下:基底上的孔的数目的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,使得细胞的数目为基底中的孔的数目的约10%。

[0280] 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多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,使得基底的一个孔仅容纳一个细胞。

[0281] 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。

[0282] 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,在样品包含细胞并且组合物特异性结合的细胞组分靶在细胞表面上(诸如细胞表面蛋白)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,使得仅与细胞组分靶特异性结合的组合物与细胞一起保留。

[0283] 在一些实施方案中,本文公开的方法可以提供从样品(例如,细胞)中释放多于一种核酸靶分子。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、光学处理或其任何组合。细胞可以通过添加包含去垢剂(例如,SDS、十二烷基硫酸锂、Triton X-100、吐温-20或NP-40)、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)或其任何组合的细胞裂解缓冲液来裂解。

[0284] 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分子或其组合,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种或在这些值中的任何两个之间的数字或范围的物质。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种物质。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。

[0285] 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、样品标记等或其任何组合。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。

[0286] 在一些实施方案中,多于一种寡核苷酸探针可以被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被固定在固体支持物上。当多于一种寡核苷酸探针与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紧密接近时,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。

[0287] 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、酶处理或其任何组合。将寡核苷酸从细胞组分结合试剂脱离可以在将多于一个寡核苷酸探针与多于一个核酸靶分子和组合物中的多于一个寡核苷酸杂交的步骤之前、之后或期间进行。

[0288] 蛋白和核酸靶的同时定量分析

[0289] 在一些实施方案中,本文公开的方法也可以用于多于一种类型靶分子例如蛋白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示出了在单细胞中同时定量分析核酸靶和其他细胞组分靶(例如,蛋白)二者的示例性方法的示意图。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605a、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与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、受体或其任何组合特异性结合。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。

[0290] 条形码的关联

[0291] 与细胞组分结合试剂(例如,抗原结合试剂或蛋白结合试剂)和/或核酸分子关联的寡核苷酸可以与寡核苷酸探针(例如,条形码,诸如随机条形码)随机关联。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,可以是或包括本公开内容的寡核苷酸,诸如抗体寡核苷酸、样品索引寡核苷酸、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。关联可以例如包括使寡核苷酸探针的靶结合区与靶核酸分子和/或蛋白结合试剂的寡核苷酸的互补部分杂交。例如,条形码(例如随机条形码)的寡(dT)区域可以与靶核酸分子的多(A)尾和/或蛋白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的杂交体。

[0292] 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,细胞组分结合试剂上最初缀合的寡核苷酸可以是RNA碱基或DNA碱基或二者。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拷贝并连接(例如,共价地连接)至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分子或其一部分的序列之外,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)。

[0293] 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。该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。

[0294] 确定独特分子标记序列的数目或存在

[0295] 在一些实施方案中,本文公开的方法包括确定用于每种独特标识符、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目或存在。例如,测序读段可以用于确定用于每种独特标识符、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目。作为另一种实例,测序读段可以用于确定分子标记序列的存在或不存在(诸如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测序读段中的存在或不存在)。

[0296] 在一些实施方案中,用于每种独特标识符、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目指示样品中每种细胞组分靶(例如,抗原靶或蛋白靶)和/或每种核酸靶分子的数量。在一些实施方案中,细胞组分靶的数量和其相应核酸靶分子例如mRNA分子的数量可以相互比较。在一些实施方案中,可以计算细胞组分靶的数量与其相应核酸靶分子例如mRNA分子的数量之比。细胞组分靶可以是,例如,细胞表面蛋白标志物。

[0297] 在一些实施方案中,细胞表面蛋白标志物的蛋白水平和细胞表面蛋白标志物的mRNA水平之间的比率低。

[0298] 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分子的量之间的期望比率的细胞组分靶。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或在这些值中的任何两个之间的数字或范围。在一些实施方案中,比率是至少以下或是至多以下: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的细胞组分靶,样品中所述细胞组分靶对应的核酸靶分子的量是以下或是约以下:1000种、100种、10种、5种、2种、1种、0种或在这些值中的任何两个之间的数字或范围。在一些实施方案中,本文公开的方法可以用于鉴定样品中这样的细胞组分靶,样品中该细胞组分靶对应的核酸靶分子的量多于以下或小于以下:1000种、100种、10种、5种、2种、1种或0种。

[0299] 组合物和试剂盒

[0300] 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)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,多于一种蛋白结合试剂),其中寡核苷酸包含用于细胞组分结合试剂的独特标识符;和多于一种寡核苷酸探针,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相异的独特条形码序列。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、样品标记或其任何组合。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(未锚定至固体支持物)或oligoA₁₈V(锚定至固体支持物)。寡核苷酸可以包含DNA残基、RNA残基或二者。

[0301] 本文的公开内容包括多于一种样品索引组合物。多于一种样品索引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结合。样品索引寡核苷酸可以包含用于鉴定样品中的一个或更多细胞的样品来源的样品索引序列。多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列可以包括不同的序列。

[0302] 本文的公开内容包括用于细胞鉴定的包含样品索引组合物的试剂盒。在一些实施方案中,两种样品索引组合物中的每一种包含与样品索引寡核苷酸关联的细胞组分结合试剂(例如,蛋白结合试剂),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,一种或更多种蛋白靶)中的至少一种特异性结合,其中样品索引寡核苷酸包含样品索引序列,并且其中两种样品索引组合物的样品索引序列包含不同的序列。在一些实施方案中,样品索引寡核苷酸包含分子标记序列、用于通用引物的结合位点或其组合。

[0303] 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组合物。两种或更多种样品索引组合物中的每一种可以包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中两种样品索引组合物的样品索引序列包含不同的序列。在一些实施方案

中,样品索引寡核苷酸包含分子标记序列、用于通用引物的结合位点或其组合。本文的公开内容包括用于多重体(multiplet)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组合物。两种样品索引组合物中的每一种可以包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中两种样品索引组合物的样品索引序列包含不同的序列。

[0304] 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是以下、是约以下、小于以下、大于以下的长度:4个、5个、6个、7个、8个、9个、10个、15个、20个、25个、30个、35个、40个、45个、50个、55个、60个、70个、80个、90个、100个、200个核苷酸或以上值中的任何两个之间的范围。

[0305] 在一些实施方案中,独特标识符选自一组相异的独特标识符。一组相异的独特标识符可以包括以下或可以包括约以下:20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、2000个、5000个或在这些值中的任何两个之间的数字或范围的不同的独特标识符。一组相异的独特标识符可以包括至少以下或包括至多以下:20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、2000个或5000个不同的独特标识符。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、2个、3个、4个、5个、6个、7个、8个、9个、10个,或在这些值中的任何两个之间的数字或范围。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、2个、3个、4个、5个、6个、7个、8个、9个或10个核苷酸。

[0306] 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。

[0307] 本公开内容中设想了任何合适的细胞组分结合试剂,诸如任何蛋白结合试剂(例如抗体或其片段、适配体、小分子、配体、肽、寡核苷酸等或其任何组合)。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)或其片段,诸如Fab和Fv。在一些实施方案中,多于一种蛋白结合试剂可以包括以下或可以包括约以下: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之间的数字或范围的不同的蛋白结合试剂。在一些实施方案中,多于一种蛋白结合试剂可以包括至少以下或可以包括至多以下: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种不同的蛋白结合试剂。

[0308] 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结合试剂共价缀合。在一些实施方案中,寡核苷酸可以与蛋白结合试剂非共价缀合。在一些实施方案中,接头可以包含将寡核苷酸与蛋白结合试剂可逆地或不可逆地附接的化学基团。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至蛋白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、链霉抗生物素蛋白、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-末端与蛋白结合试剂缀合。寡核苷酸可以与蛋白结合试剂的任何合适的位点缀合,只要它不干扰蛋白结合试剂与其蛋白靶之间的特异性结合。在蛋白结合试剂是抗体的实施方案中,寡核苷酸可以与抗体在除了抗原结合位点之外的任何地方(例如Fc区、C_H1结合域、C_H2结合域、C_H3结合域、C_L结合域等)缀合。在一些实施方案中,每种蛋白结合试剂可以与单个寡核苷酸分子缀合。在一些实施方案中,每种蛋白结合试剂可以与以下或与约以下的寡核苷酸分子缀合:2个、3个、4个、5个、10个、20个、30个、40个、50个、100个、1000个或在这些值中的任何两个之间的数字或范围,其中每个寡核苷酸分子包含相同的独特标识符。在一些实施方案中,每种蛋白结合试剂可以与多于一个寡核苷酸分子缀合,例如,至少或至多2个、3个、4个、5个、10个、20个、30个、40个、50个、100个或1000个寡核苷酸分子,其中每个寡核苷酸分子包含相同的独特标识符。

[0309] 在一些实施方案中,多于一种细胞组分结合试剂(例如,蛋白结合试剂)能够与样品中的多于一种细胞组分靶(例如,蛋白靶)特异性结合。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、受体或其任何组合。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。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。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)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%或在这些值中的任何两个之间的数字或范围。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,表达或可以表达的蛋白)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可以包括约以下:2种、3种、4种、5种、10种、20种、30种、40种、50种、100种、1000种、10000种或在这些值中的任何两个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可以包括至多以下: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。

[0310] 使用寡核苷酸缀合的细胞组分结合试剂的样品索引

[0311] 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,该方法包括: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组合物中的样品索引组合物接触,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组合物中的每一种包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列;去除多于一种样品索引组合物中未结合的样品

索引组合物;使用多于一种条形码(例如,随机条形码)使样品索引寡核苷酸条形码化(例如,随机条形码化)以产生多于一种条形码化样品索引寡核苷酸;获得多于一种条形码化样品索引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寡核苷酸中的至少一种条形码化样品索引寡核苷酸的样品索引序列,鉴定一个或更多细胞中的至少一个细胞的样品来源。

[0312] 在一些实施方案中,使用多于一种条形码对样品索引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寡核苷酸接触以产生与样品索引寡核苷酸杂交的条形码;以及使与样品索引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寡核苷酸。

[0313] 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”)。在样品索引的上下文中,抗体寡核苷酸在本文中被称为样品索引寡核苷酸。与抗体寡核苷酸缀合的抗体在本文中被称为热抗体(hot antibody)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(cold antibody)或无寡核苷酸抗体。与结合试剂(例如,蛋白结合试剂)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被称为试剂寡核苷酸。在样品索引的上下文中,试剂寡核苷酸在本文中被称为样品索引寡核苷酸。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。

[0314] 图7示出了使用寡核苷酸关联的细胞组分结合试剂进行样品索引的示例性工作流程的示意图。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、适配体、蛋白支架或其组合。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寡核苷酸之外)。

[0315] 不同的组合物可以包括与具有不同样品索引序列的样品索引寡核苷酸缀合的结合试剂。在不同实施方式中,不同的组合物的数目可以是不同的。在一些实施方案中,不同组合物的数目可以是以下或可以是约以下: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、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种、10000种或在这些值中的任何两个之间的数字或范围。在一些实施方案中,不同组合物的数目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、200种、300种、400种、500种、600种、700种、800种、900种、1000种、2000种、3000种、4000种、5000种、6000种、7000种、8000种、9000种或10000种。

[0316] 在一些实施方案中,一种组合物中结合试剂的样品索引寡核苷酸可以包含相同的样品索引序列。一种组合物中结合试剂的样品索引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序列的结合试剂的样品索引寡核苷酸的百分比可以是

以下或可以是约以下: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%或在这些值中的任何两个之间的数字或范围。在一些实施方案中,一种组合物中具有相同样品索引序列的结合试剂的样品索引寡核苷酸的百分比可以是至少以下或可以是至多以下: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%。

[0317] 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寡核苷酸可以具有一种样品索引序列,并且可以用于对样品707a(诸如患者的样品)中的细胞710a(以黑色圆圈示出)进行标记。组合物705b中细胞组分结合试剂的样品索引寡核苷酸可以具有另一种样品索引序列,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(以带阴影的圆圈示出)进行标记。细胞组分结合试剂可以与细胞表面上的细胞组分靶或蛋白诸如细胞标志物、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、受体或其任何组合特异性结合。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。

[0318] 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寡核苷酸725a可以被配置为(或可以)能够从细胞组分结合试剂脱离或不能够从细胞组分结合试剂脱离。与组合物705a的细胞组分结合试剂缀合的样品索引寡核苷酸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寡核苷酸725b可以被配置为(或可以为)能够从细胞组分结合试剂脱离或不能够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。

[0319] 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以虚线的圆圈示出。细胞mRNA 730a、样品索引寡核苷酸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。

[0320] 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以虚线的圆圈示出。细胞mRNA 730b、样品索引寡核苷酸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。

[0321] 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,根据样品索引寡核苷酸725a和725b的样品索引序列)。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。使用样品索引寡核苷酸的样品索引序列对细胞标记、分子标记和样品身份进行去多重化可以用于确定样品来源。

[0322] 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样品索引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样品索引寡核苷酸的文库缀合,以允许细胞保持样品身份。这将使多个样品能够加载到同一个Rhapsody™筒(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。样品索引可以允许在单个实验中一起运行多个样品,从而简化和缩短实验时间,并消除批次效应(batch effect)。

[0323] 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,该方法包括: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组合物中的样品索引组合物接触,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组合物中的每一种包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列;去除多于一种样品索引组合物中未结合的样品索引组合物。该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寡核苷酸;获得多于一种条形码化样品索引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寡核苷酸中的至少一种条形码化样品索引寡核苷酸的样品索引序列,鉴定一个或更多个细胞中至少一个细胞的样品来源。

[0324] 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组合物中的样品索引组合物接触,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组合物中的每一种包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列;去除多于一种样品索引组合物中未结合的样品索引组合物;以及基于多于一种样品索引组合物的至少一种样品索引寡核苷酸的样品索引序列,鉴定一个或更多个细胞中至少一个细胞的样品来源。

[0325] 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组合物的样品索引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寡核苷酸;获得多于一种条形码化样品索引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寡核苷酸中的至少一种条形码化样品索引寡核苷酸的样品索引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寡核苷酸包括使用多于一种随机条形码对样品索引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寡核苷酸。

[0326] 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组合物的至少一种样品索引寡核苷酸的样品索引序列的存在或不存在。鉴定样品索引序列的存在或不存在可以包括:复制至少一种样品索引寡核苷酸以产生多于一种复制的样品索引寡核苷酸;获得多于一种复制的样品索引寡核苷酸的测序数据;以及基于多于一种样品索引寡核苷酸中的复制的样品索引寡核苷酸的样品索引序列来鉴定细胞的样品来源,所述多于一种样品索引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寡核苷酸。

[0327] 在一些实施方案中,复制至少一种样品索引寡核苷酸以产生多于一种复制的样品索引寡核苷酸包括:在复制至少一种条形码化样品索引寡核苷酸之前,将复制衔接子连接到至少一种条形码化样品索引寡核苷酸。复制至少一种条形码化样品索引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寡核苷酸的复制衔接子复制至少一种条形码化样品索引寡核苷酸以产生多于一种复制的样品索引寡核苷酸。

[0328] 在一些实施方案中,复制至少一种样品索引寡核苷酸以产生多于一种复制的样品索引寡核苷酸包括:在复制至少一种条形码化样品索引寡核苷酸之前,使捕获探针与至少一种样品索引寡核苷酸接触以产生与样品索引寡核苷酸杂交的捕获探针;以及使与样品索引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寡核苷酸。复制至少一种样品索引寡核苷酸可以包括复制与捕获探针关联的样品索引寡核苷酸以产生多于一种复制的样品索引寡核苷酸。

[0329] 细胞过载和多重体鉴定

[0330] 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(诸如蛋白表达水平)的方法、试剂盒和系统。

[0331] 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)的微孔或液滴的数目可以是极小的。然而,当加载的细胞数目增加时,具有多于一个细胞的微孔或液滴的数目可能显著增加。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,具有多于一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目的组(例如5个组)并且每组中的细胞用具有不同样品索引序列的样品索引寡核苷酸进行标记,与两种或更多种样品索引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序列的样品索引寡核苷酸进行标记,与两种或更多种样品索引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目,可以将这样更高数目的细胞加载到微孔中。

[0332] 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,该方法包括: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组合物接触,其中第一多于一个

细胞中的每一个和第二个多于一个细胞中的每一个包含一种或更多种细胞组分,其中两种样品索引组合物中的每一种包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中两种样品索引组合物的样品索引序列包含不同的序列;使用多于一种条形码对样品索引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序列关联。在一些实施方案中,样品索引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其组合。

[0333] 例如,该方法可以用于使用样品索引加载50000个或更多个细胞(相比于10000-20000个细胞)。样品索引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,组合的“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序列的样品索引寡核苷酸(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目可以是不同的。在一些实施方案中,不同群体的数目可以是以下或可以是约以下:2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100或在这些值中的任何两个之间的数字或范围。在一些实施方案中,不同群体的数目可以是至少以下或可以是至多以下: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,每个群体中细胞的数目或平均数目可以是不同的。在一些实施方案中,每个群体中细胞的数目或平均数目可以是以下或可以是约以下: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个或在这些值中的任何两个之间的数字或范围。在一些实施方案中,每个群体中细胞的数目或平均数目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个或100个。当每个群体中细胞的数目或平均数目足够小(例如,等于或小于每个群体50个、25个、10个、5个、4个、3个、2个或1个细胞)时,用于细胞过载和多重体鉴定的样品索引组合物可以被称为细胞鉴定组合物。

[0334] 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序列,可以将与测序数据中的多于一种样品索引序列关联的“细胞”鉴定为“多重体”。组合的“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、九重体或其任何组合。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或在这些值中的任何两个之间的范围。在一些实施方案中，n是至少以下或是至多以下：2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20。

[0335] 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。

[0336] 双重体可以指与具有不同样品索引序列的两种样品索引寡核苷酸关联的组的“细胞”。双重体也可以指与具有两种样品索引序列的样品索引寡核苷酸关联的组的“细胞”。当同一微孔或微滴中捕获到与不同序列的两种样品索引寡核苷酸关联的两个细胞(或与具有两种不同样品索引序列的样品索引寡核苷酸关联的两个或更多个细胞)时，双重体可能出现，组的“细胞”可以与具有不同样品索引序列的两种样品索引寡核苷酸关联。三重体可以指与全部具有不同样品索引序列的三种样品索引寡核苷酸关联的组的“细胞”，或者是与具有三种不同样品索引序列的样品索引寡核苷酸关联的组的“细胞”。四重体可以指与全部具有不同样品索引序列的四种样品索引寡核苷酸关联的组的“细胞”，或者是与具有四种不同样品索引序列的样品索引寡核苷酸关联的组的“细胞”。五重体可以指与全部具有不同样品索引序列的五种样品索引寡核苷酸关联的组的“细胞”，或者是与具有五种不同样品索引序列的样品索引寡核苷酸关联的组的“细胞”。六重体可以指与全部具有不同样品索引序列的六种样品索引寡核苷酸关联的组的“细胞”，或者是与具有六种不同样品索引序列的样品索引寡核苷酸关联的组的“细胞”。七重体可以指与全部具有不同样品索引序列的七种样品索引寡核苷酸关联的组的“细胞”，或者是与具有七种不同样品索引序列的样品索引寡核苷酸关联的组的“细胞”。八重体可以指与全部具有不同样品索引序列的八种样品索引寡核苷酸关联的组的“细胞”，或者是与具有八种不同样品索引序列的样品索引寡核苷酸关联的组的“细胞”。九重体可以指与全部具有不同样品索引序列的九种样品索引寡核苷酸关联的组的“细胞”，或者是与具有九种不同样品索引序列的样品索引寡核苷酸关联的组的“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序列的样品索引寡核苷酸关联的两个或更多个细胞)时，多重体可能出现，组的“细胞”可以与具有两种或更多种不同样品索引序列的样品索引寡核苷酸关联。

[0337] 作为另一个实例，该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，从组的“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，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：具有不同样品索引序列的两种或更多种样品索引寡核苷酸(或具有两种或更多种样品索引序列的样品索引寡核苷酸)。利用样品索引序列，本文公开的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)。在一些实施

方案中,该方法包括: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,其中两种样品索引组合物中的每一种包含与样品索引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结合,其中样品索引寡核苷酸包含样品索引序列,并且其中两种样品索引组合物的样品索引序列包含不同的序列;使用多于一种条形码对样品索引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序列关联的一种或更多种多重体细胞标记序列。

[0338] 可以加载到微孔筒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目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。利用样品索引,该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,可以提高使用者对多重体比率的容忍度,并将更多的细胞加载到每个微孔筒上,或者产生各自具有至少一个细胞的液滴。

[0339] 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组合物接触包括:使第一多于一个细胞与两种样品索引组合物中的第一样品索引组合物接触;以及使第一多于一个细胞与两种样品索引组合物中的第二样品索引组合物接触。在不同的实施方式中,多于一个细胞的数目和多于一种样品索引组合物的数目可以不同。在一些实施方案中,多于一个细胞和/或样品索引组合物的数目可以是以下或可以是约以下: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或在这些值中的任何两个之间的数字或范围。在一些实施方案中,多于一个细胞和/或样品索引组合物的数目可以是至少以下或可以是至多以下: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,细胞的数目可以是不同的。在一些实施方案中,细胞的数目或平均数目可以是以下或可以是约以下: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、200个、300个、400个、500个、600个、700个、800个、900个、1000个、10000个、100000个、1000000个或在这些值中的任何两个之间的数字或范围。在一些实施方案中,细胞的数目或平均数目可以是至少以下或可以是至多以下: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、200个、300个、400个、500个、600个、700个、800个、900个、1000个、10000个、100000个或1000000个。

[0340] 在一些实施方案中,该方法包括:去除两种样品索引组合物中未结合的样品索引组合物。去除未结合的样品索引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组合物可以包括使用流式细胞术选择与两种样品索引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,该方法包括: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。

[0341] 在一些实施方案中,样品索引寡核苷酸被配置为(或可以为)能够从细胞组分结合试剂脱离或不能够从细胞组分结合试剂脱离。该方法可以包括使样品索引寡核苷酸从细胞组分结合试剂脱离。使样品索引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、酶处理或其任何组合使样品索引寡核苷酸从细胞组分结合试剂脱离。

[0342] 在一些实施方案中,使用多于一种条形码对样品索引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寡核苷酸接触以产生与样品索引寡核苷酸杂交的条形码;以及使与样品索引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寡核苷酸。

[0343] 在一些实施方案中,该方法包括:扩增多于一种条形码化样品索引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样品索引寡核苷酸的样品索引序列鉴定多于一种条形码化靶的样品来源。

[0344] 在一些实施方案中,使用多于一种条形码对样品索引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寡核苷酸包括使用多于一种随机条形码对样品索引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寡核苷酸。

[0345] 在一些实施方案中,该方法包括: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。

[0346] 在一些实施方案中,该方法包括: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。

[0347] 在一些实施方案中,用于细胞鉴定的方法包括:使第一多于一个或更多细胞(a first plurality of one or more cells)和第二多于一个或更多细胞(a second plurality of one or more cells)分别与两种细胞鉴定组合物接触,其中第一多于一个或更多细胞中的每一个和第二多于一个或更多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其组合。

[0348] 当同一微孔或微滴中捕获到与具有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)可能出现,组合的“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。

[0349] 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,从组合的“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,该方法包括:使第一多于一个或更多个细胞和第二多于一个或更多个细胞分别与两种细胞鉴定组合物接触,其中第一多于一个或更多个细胞中的每一个和第二多于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。

[0350] 可以加载到微孔筒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

目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,该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,可以提高使用者对多重体比率的容忍度,并将更多的细胞加载到每个微孔筒上,或者产生各自具有至少一个细胞的液滴。

[0351] 在一些实施方案中,使第一多于一个或更多细胞和第二多于一个或更多细胞分别与两种细胞鉴定组合物接触包括:使第一多于一个或更多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;以及使第二多于一个或更多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目可以是不同的。在一些实施方案中,细胞鉴定组合物的数目可以是以下或可以是约以下: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、10000种、100000种、1000000种或在这些值中的任何两个之间的数字或范围。在一些实施方案中,细胞鉴定组合物的数目可以是至少以下或可以是至多以下: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、10000种、100000种或1000000种。在不同的实施方式中,每一个多于一个或更多细胞(each plurality of one or more cells)中的细胞的数目或平均数目可以不同。在一些实施方案中,每一个多于一个或更多细胞中的细胞的数目或平均数目可以是以下或可以是约以下: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、200个、300个、400个、500个、600个、700个、800个、900个、1000个、10000个、100000个、1000000个或在这些值中的任何两个之间的数字或范围。在一些实施方案中,每一个多于一个或更多细胞中的细胞的数目可以是至少以下或可以是至多以下: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、200个、300个、400个、500个、600个、700个、800个、900个、1000个、10000个、100000个或1000000个。

[0352] 在一些实施方案中,该方法包括: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或更多细胞和第二多于一个或更多细胞中的细胞。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,该方法包括裂解来自多于一个样品的每一个中的一个或更多细胞。

[0353] 在一些实施方案中,细胞鉴定寡核苷酸被配置为(或可以为)能够从细胞组分结合试剂脱离或不能够从细胞组分结合试剂脱离。该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、酶处理或其任何组合使细胞鉴定寡核苷酸从细胞组分结合试剂脱离。

[0354] 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。

[0355] 在一些实施方案中,该方法包括: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。

[0356] 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。

[0357] 寡核苷酸缀合的抗体

[0358] 独特分子标记序列

[0359] 在一些实施方案中,与细胞组分结合试剂关联的寡核苷酸(例如,抗体寡核苷酸(“AbOligo”或“Ab0”)、结合试剂寡核苷酸、细胞组分结合试剂特异性寡核苷酸、样品索引寡核苷酸)包含独特分子标记序列(也称为分子索引(MI)、“分子条形码”或独特分子标识符(UMI))。在一些实施方案中,包含如本文描述的分子条形码的结合试剂寡核苷酸物质通过增加灵敏度、降低相对标准误差或增加灵敏度和/或减少标准误差来减少偏倚。分子条形码可以包含独特序列,使得当多于一个样品核酸(可以彼此相同和/或不同)与分子条形码一一关联时,不同的样品核酸可以通过分子条形码彼此区分。这样,即使样品包含具有相同序列的两种核酸,这两种核酸中的每一种可以用不同的分子条形码标记,使得群体中的核酸可以被定量,甚至在扩增之后被定量。分子条形码可以包含以下的核酸序列:至少5个核苷酸,例如至少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个或50个核苷酸,包括在所列值中的任何两个之间的范围,例如5-50个、5-45个、5-40个、5-35个、5-30个、5-25个、5-20个、5-15个、5-14个、5-13个、5-12个、5-11个、5-10个、5-9个、5-8个、5-7个、5-6个、6-50个、6-45个、6-40个、6-35个、6-30个、6-25个、6-20个、6-15个、6-14个、6-13个、6-12个、6-11个、6-10个、6-9个、6-8个、6-7个、7-50个、7-45个、7-40个、7-35个、7-30个、7-25个、7-20个、7-15个、7-14个、7-13个、7-12个、7-11个、7-10个、7-9个、7-8个、8-50个、8-45个、8-40个、8-35个、8-30个、8-25个、8-20个、8-15个、8-14个、8-13个、8-12个、8-11个、8-10个、8-9个、9-50个、9-45个、9-40个、9-35个、9-30个、9-25个、9-20个、9-15个、9-14个、9-13个、9-12个、9-11个、9-10个、10-50个、10-45个、10-40个、10-35个、10-30个、10-25个、10-20个、10-15个、10-14个、10-13个、10-12个或10-11个核苷酸。在一些实施方案中,分子条形码的核酸序列包含例如独特序列,使得组合物中的每种独特寡核苷酸物质包含不同的分子条形码。在一些实施方案中,两种或更多种独特寡核苷酸物质可以包含相同的分

子条形码,但是仍然彼此不同。例如,如果独特寡核苷酸物质包含样品条形码,则具有特定样品条形码的每种独特寡核苷酸物质可以包含不同的分子条形码。在一些实施方案中,包含独特寡核苷酸物质的组合物包含至少1000种不同分子条形码的分子条形码多样性,并且因此包含至少1000种独特寡核苷酸物质。在一些实施方案中,包含独特寡核苷酸物质的组合物包含至少6,500种不同分子条形码的分子条形码多样性,并且因此包含至少6,500种独特寡核苷酸物质。在一些实施方案中,包含独特寡核苷酸物质的组合物包含至少65,000种不同分子条形码的分子条形码多样性,并且因此包含至少65,000种独特寡核苷酸物质。

[0360] 在一些实施方案中,独特分子标记序列位于独特标识符序列的5',在独特分子标记序列与独特标识符序列之间没有任何间插序列。在一些实施方案中,独特分子标记序列位于间隔区的5',所述间隔区位于独特标识符序列的5',使得间隔区位于独特分子标记序列与独特标识符序列之间。在一些实施方案中,独特标识符序列位于独特分子标记序列的5',在独特标识符序列与独特分子标记序列之间没有任何间插序列。在一些实施方案中,独特标识符序列位于间隔区的5',所述间隔区位于独特分子标记序列的5',使得间隔区位于独特标识符序列与独特分子标记序列之间。

[0361] 独特分子标记序列可以包含以下的核酸序列:至少3个核苷酸,例如至少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个核苷酸,包括在所列值中的任何两个之间的范围,例如3-50个、3-45个、3-40个、3-35个、3-30个、3-25个、3-20个、3-15个、3-14个、3-13个、3-12个、3-11个、3-10个、3-9个、3-8个、3-7个、3-6个、3-5个、3-4个、4-50个、4-45个、4-40个、4-35个、4-30个、4-25个、4-20个、4-15个、4-14个、4-13个、4-12个、4-11个、4-10个、4-9个、4-8个、4-7个、4-6个、4-5个、5-50个、5-45个、5-40个、5-35个、5-30个、5-25个、5-20个、5-15个、5-14个、5-13个、5-12个、5-11个、5-10个、5-9个、5-8个、5-7个、5-6个、6-50个、6-45个、6-40个、6-35个、6-30个、6-25个、6-20个、6-15个、6-14个、6-13个、6-12个、6-11个、6-10个、6-9个、6-8个、6-7个、7-50个、7-45个、7-40个、7-35个、7-30个、7-25个、7-20个、7-15个、7-14个、7-13个、7-12个、7-11个、7-10个、7-9个、7-8个、8-50个、8-45个、8-40个、8-35个、8-30个、8-25个、8-20个、8-15个、8-14个、8-13个、8-12个、8-11个、8-10个、8-9个、9-50个、9-45个、9-40个、9-35个、9-30个、9-25个、9-20个、9-15个、9-14个、9-13个、9-12个、9-11个、9-10个、10-50个、10-45个、10-40个、10-35个、10-30个、10-25个、10-20个、10-15个、10-14个、10-13个、10-12个或10-11个核苷酸。在一些实施方案中,独特分子标记序列的长度是2-20个核苷酸。

[0362] 在一些实施方案中,结合试剂寡核苷酸的独特分子标记序列包含为双重体“VN”和/或“NV”的至少三个重复的序列(其中每个“V”是A、C或G中的任何一个,并且其中“N”是A、G、C或T中的任何一个),至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。双重体“VN”的多个重复的实例包括VN、VNVN、VNVNVN和VNVNVNVN。要注意的是,虽然式“VN”和“NV”描述了对基本内容的约束,但不是每个V或每个N都必须相同或不同。例如,如果组合物中独特寡核苷酸物质的分子条形码包含VNVNVN,一种分子条形码可以包含序列ACGGCA,而另一种分子条形码可以包含序列ATACAT,而另一种分子条形码可以包含序

列ATACAC。要注意的是,任何数目的重复的双重体“VN”将具有不多于50%的T含量。在一些实施方案中,包含至少1000种独特寡核苷酸物质的组合物的至少95%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少1000种独特寡核苷酸物质的组合物的至少99%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少1000种独特寡核苷酸物质的组合物的至少99.9%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少6500种独特寡核苷酸物质的组合物的至少95%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少6500种独特寡核苷酸物质的组合物的至少99%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少65,000种独特寡核苷酸物质的组合物的至少95%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少65,000种独特寡核苷酸物质的组合物的至少99%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,包含至少65,000种独特寡核苷酸物质的组合物的至少99.9%的独特寡核苷酸物质包含含有双重体“VN”和/或“NV”的至少三个重复的分子条形码,至少三个重复例如至少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个重复,包括在所列值中的任何两个之间的范围。在一些实施方案中,组合物由或基本上由至少1000、6500或65,000种独特的寡核苷酸物质组成,每种寡核苷酸物质具有包含序列VNVNVN的分子条形码。在一些实施方案中,组合物由或基本上由至少1000、6500或65,000种独特的寡核苷酸物质组成,每种寡核苷酸物质具有包含序列VNVNVNVN的分子条形码。在一些实施方案中,如本文描述的组合物至少95%、99%或99.9%的条形码区域包含如本文描述的双重体“VN”和/或“NV”的至少三个重复。在

一些实施方案中,包含重复的双重体“VN”和/或“NV”的独特分子标记序列可以产生低偏倚,同时在减少偏倚与维持相对大量的可用核苷酸序列之间提供折中,使得可以在相对短的序列中获得相对高的多样性,同时仍然使偏倚最小化。在一些实施方案中,包含重复的双重体“VN”和/或“NV”的独特分子标记序列可以通过增加灵敏度、降低相对标准误差或增加灵敏度和减少标准误差来减少偏倚。在一些实施方案中,包含重复的双重体“VN”和/或“NV”的独特分子标记序列通过充当地理标志物(geomarker)来改进信息学分析。在一些实施方案中,本文描述的重复的双重体“VN”和/或“NV”减少了独特分子标记序列内均聚物的发生率。在一些实施方案中,本文描述的重复的双重体“VN”和/或“NV”中断均聚物。

[0363] 在一些实施方案中,样品索引寡核苷酸包含第一分子标记序列。在一些实施方案中,至少两种样品索引寡核苷酸的第一分子标记序列是不同的,并且至少两种样品索引寡核苷酸的样品索引序列是相同的。在一些实施方案中,至少两种样品索引寡核苷酸的第一分子标记序列是不同的,并且至少两种样品索引寡核苷酸的样品索引序列是不同的。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含第二分子标记序列。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示多于一个细胞中的一个或更多个细胞中至少一种细胞组分靶的拷贝数,细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,以下(1)和(2)的组合(例如,最小值、平均值和最大值)指示多于一个细胞中的一个或更多个细胞中至少一种细胞组分靶的拷贝数:(1)测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目,细胞组分结合试剂能够与至少一种细胞组分靶特异性结合;和(2)测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目,细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。

[0364] 对齐序列

[0365] 在一些实施方案中,结合试剂寡核苷酸包含与多(dA)区相邻的对齐序列(例如,参考图8描述的对齐序列825bb)。对齐序列的长度可以是1个或更多个核苷酸。对齐序列的长度可以是2个核苷酸。对齐序列可以包含鸟嘌呤、胞嘧啶、胸腺嘧啶、尿嘧啶或其组合。对齐序列可以包含多(dT)区、多(dG)区、多(dC)区、多(dU)区或其组合。在一些实施方案中,对齐序列位于多(dA)区的5'。有利地,在一些实施方案中,对齐序列的存在使得结合试剂寡核苷酸中的每一种的多(A)尾具有相同的长度,导致更大的性能均匀性。在一些实施方案中,在多于一种结合试剂寡核苷酸(其中的每一种结合试剂寡核苷酸包含对齐序列)中具有相同多(dA)区长度的结合试剂寡核苷酸的百分比可以是以下或可以是约以下:80%、90%、91%、93%、95%、97%、99.9%、99.9%、99.99%或100%或在这些值中的任何两个之间的数字或范围。在一些实施方案中,在多于一种结合试剂寡核苷酸(其中的每一种结合试剂寡核苷酸包含对齐序列)中具有相同多(dA)区长度的结合试剂寡核苷酸的百分比可以是至少以下或可以是至多以下:80%、90%、91%、93%、95%、97%、99.9%、99.9%、99.99%或100%。

[0366] 在不同实施方式中,对齐序列的长度可以是不同的。在一些实施方案中,对齐序列的长度可以是以下或可以是约以下: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、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或在这些值中的任何两个之间的数字或范围。在一些实施方案中,对齐序列的长度可以是至少以下或是至多以下: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、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。在不同实施方式中,对齐序列中鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶的数目可以是不同的。鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶的数目可以是以下或可以是约以下: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、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、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个、65个、66个、67个、68个、69个、70个、71个、72个、73个、74个、75个、76个、77个、78个、79个、80个、81个、82个、83个、84个、85个、86个、87个、88个、89个、90个、91个、92个、93个、94个、95个、96个、97个、98个、99个、100个或在这些值中的任何两个之间的数字或范围。鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶的数目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、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、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个、65个、66个、67个、68个、69个、70个、71个、72个、73个、74个、75个、76个、77个、78个、79个、80个、81个、82个、83个、84个、85个、86个、87个、88个、89个、90个、91个、92个、93个、94个、95个、96个、97个、98个、99个或100个。在一些实施方案中,样品索引寡核苷酸包含对齐序列。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含对齐序列。

[0367] 接头

[0368] 结合试剂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,结合试剂寡核苷酸可以与细胞组分结合试剂共价缀合。在一些实施方案中,结合试剂寡核苷酸可以与细胞组分结合试剂非共价缀合。在一些实施方案中,结合试剂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,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至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、链霉抗生物素蛋白、生物素、胺及其任何组合。接头可以包含碳链。碳链可以包含例如5-50个碳原子。在不同实施方案中,碳链可以具有不同数目的碳原子。在一些实施方案中,碳链中碳原子的数目可以是以下或可以是约以下: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个或在这些值中的任何两个之间的数字或范围。在一些实施方案中,碳链中碳原子的数目可以是至少以下或可以是至多以下: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个或50个。在一些实施方案中,碳链包含2-30个碳原子。在一些实施方案中,碳链包含12个碳原子。在一些实施方案中,用于结合试剂寡核苷酸的氨基修饰物可以与细胞组分结合试剂缀合。在一些实施方案中,接头包含5'氨基修饰物C6(5AmMC6)。在一些实施方案中,接头包含5'氨基修饰物C12(5AmMC12)。在一些实施方案中,接头包含5AmMC12的衍生物。在一些实施方案中,较长的接头实现了更高的缀合效率。在一些实施方案中,较长的接头在缀合之前实现了更高的修饰效率。在一些实施方案中,增加功能性胺与DNA序列之间的距离产生更高的缀合效率。在一些实施方案中,增加功能性胺与DNA序列之间的距离在缀合之前产生更高的修饰效率。在一些实施方案中,使用5AmMC12作为接头比使用5AmMC6作为接头产生更高的修饰效率(在缀合之前)。在一些实施方案中,使用5AmMC12作为接头比使用5AmMC6作为接头产生更高的缀合效率。在一些实施方案中,样品索引寡核苷酸通过接头与细胞组分结合试剂关联。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸通过接头与细胞组分结合试剂关联。

[0369] 抗体特异性条形码序列

[0370] 在数种实施方案中,本文公开了对结合试剂寡核苷酸的独特标识符序列(例如,抗体特异性条形码序列)的设计的改进。在一些实施方案中,独特标识符序列(例如,样品索引序列、细胞组分结合试剂特异性寡核苷酸)被设计成具有大于3的汉明距离。在一些实施方案中,独特标识符序列的汉明距离可以是以下,或是约以下:1、2、3、4、5、6、7、8、9、10,或在这些值中的任何两个之间的数字或范围。在一些实施方案中,独特标识符序列具有在40%至60%范围内的GC含量,并且不具有预测的二级结构(例如,发夹)。在一些实施方案中,独特标识符序列不包含计算机模拟预测与小鼠和/或人类转录物结合的任何序列。在一些实施方案中,独特标识符序列不包含计算机模拟预测与Rhapsody和/或SCMK系统引物结合的任何序列。在一些实施方案中,独特标识符序列不包含均聚物。

[0371] 引物衔接子

[0372] 在一些实施方案中,结合试剂寡核苷酸包含引物衔接子。在一些实施方案中,引物衔接子包含以下的序列:第一通用引物、其互补序列、其部分序列或其组合。在一些实施方案中,第一通用引物包括扩增引物、其互补序列、其部分序列或其组合。在一些实施方案中,第一通用引物包括测序引物、其互补序列、其部分序列或其组合。在一些实施方案中,测序引物包括Illumina测序引物。在一些实施方案中,测序引物包括Illumina测序引物的一部分。在一些实施方案中,测序引物包括P7测序引物。在一些实施方案中,测序引物包括P7测序引物的一部分。在一些实施方案中,引物衔接子包括用于Illumina P7的衔接子。在一些实施方案中,引物衔接子包括用于Illumina P7的部分衔接子。在一些实施方案中,扩增引物是Illumina P7序列或其子序列。在一些实施方案中,测序引物是Illumina R2序列或其子序列。在一些实施方案中,第一通用引物的长度是5-50个核苷酸。在一些实施方案中,引物衔接子可以包含以下的核酸序列:至少5个核苷酸,例如至少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个或50个核苷酸,包括在所列值中的任何两个之间的范围,例如5-50个、5-45个、5-40个、5-35个、5-30个、5-25个、5-20个、5-15个、5-14个、5-13个、5-12个、5-11个、5-10个、5-9个、5-8个、5-7个、5-6个、6-50个、6-45个、6-40个、6-35个、6-30个、6-25个、6-20个、6-15个、6-14个、6-13个、6-12个、6-11个、6-10个、6-9个、6-8个、6-7个、7-50个、7-45个、7-40个、7-35个、7-30个、7-25个、7-20个、7-15个、7-14个、7-13个、7-12个、7-11个、7-10个、7-9个、7-8个、8-50个、8-45个、8-40个、8-35个、8-30个、8-25个、8-20个、8-15个、8-14个、8-13个、8-12个、8-11个、8-10个、8-9个、9-50个、9-45个、9-40个、9-35个、9-30个、9-25个、9-20个、9-15个、9-14个、9-13个、9-12个、9-11个、9-10个、10-50个、10-45个、10-40个、10-35个、10-30个、10-25个、10-20个、10-15个、10-14个、10-13个、10-12个或10-11个核苷酸。引物衔接子可以包含以下的核酸序列:第一通用引物、扩增引物、测序引物,其互补序列、其部分序列,或其组合的序列的至少5个核苷酸,例如至少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个或50个核苷酸,包括在所列值中的任何两个之间的范围,例如第一通用引物、扩增引物、测序引物,其互补序列、其部分序列,或其组合的序列的5-50个、5-45个、5-40个、5-35个、5-30个、5-25个、5-20个、5-15个、5-14个、5-13个、5-12个、5-11个、5-10个、5-9个、5-8个、5-7个、5-6个、6-50个、6-45个、6-40个、6-35个、6-30个、6-25个、6-20个、6-15个、6-14个、6-13个、6-12个、6-11个、6-10个、6-9个、6-8个、6-7个、7-50个、7-45个、7-40个、7-35个、7-30个、7-25个、7-20个、7-15个、7-14个、7-13个、7-12个、7-11个、7-10个、7-9个、7-8个、8-50个、8-45个、8-40个、8-35个、8-30个、8-25个、8-20个、8-15个、8-14个、8-13个、8-12个、8-11个、8-10个、8-9个、9-50个、9-45个、9-40个、9-35个、9-30个、9-25个、9-20个、9-15个、9-14个、9-13个、9-12个、9-11个、9-10个、10-50个、10-45个、10-40个、10-35个、10-30个、10-25个、10-20个、10-15个、10-14个、10-13个、10-12个或10-11个核苷酸。

[0373] 用于测序文库制备的常规扩增工作流程可以采用三轮PCR,诸如,例如:第一轮(“PCR 1”)采用靶特异性引物和针对通用Illumina测序引物1序列的引物;第二轮(“PCR 2”)使用侧翼为Illumina测序引物2序列的巢式靶特异性引物和针对通用Illumina测序引物1序列的引物;和第三轮(“PCR 3”)添加Illumina P5和P7和样品索引。有利地,在一些实施方案中,与如果起始模板(例如,附接至珠的样品索引寡核苷酸)不具有引物衔接子相比,本文公开的引物衔接子能够在文库制备中实现更短且更简单的工作流程。在一些实施方案中,引物衔接子将模板的测序前PCR扩增减少一轮(与如果不包含引物衔接子的模板相比)。在一些实施方案中,引物衔接子将模板的测序前PCR扩增减少至一轮(与如果不包含引物衔接子的模板相比)。在一些实施方案中,包含引物衔接子的模板不需要用于附接Illumina测序衔接子的PCR扩增步骤,如果模板不包含引物衔接子,则需要预测序。在一些实施方案中,引物衔接子序列(或其子序列)不是包含引物衔接子序列的测序模板的测序读出的一部分,并且因此不影响包含引物衔接子的模板的读段质量。在一些实施方案中,与如果不包含引物衔接子的模板相比,包含引物衔接子的模板具有减少的测序多样性。

[0374] 在一些实施方案中,样品索引寡核苷酸包含引物衔接子。在一些实施方案中,复制样品索引寡核苷酸、条形码化样品索引寡核苷酸或其产物包括使用第一通用引物、包含第一通用引物序列的第一引物或其组合,以产生多于一种复制的样品索引寡核苷酸。在一些实施方案中,复制一种样品索引寡核苷酸、条形码化样品索引寡核苷酸或其产物包括使用第一通用引物、包含第一通用引物序列的第一引物、第二通用引物、包含第二通用引物序列的第二引物或其组合,以产生多于一种复制的样品索引寡核苷酸。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含引物衔接子。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸包含以下序列:第一通用引物、其互补序列、其部分序列或其组合。

[0375] 结合试剂寡核苷酸条形码化

[0376] 图8示出了与结合试剂805(此处图示的抗体)关联的结合试剂寡核苷酸825(此处图示的抗体寡核苷酸)的条形码化的非限制性示例性工作流程的示意图。结合试剂寡核苷酸825可以通过接头8251与结合试剂805关联。结合试剂寡核苷酸825可以使用化学、光学或其他手段从结合试剂脱离。结合试剂寡核苷酸825可以是mRNA模拟物。结合试剂寡核苷酸825可以包含引物衔接子825pa、抗体分子标记825am(例如,独特分子标记序列)、抗体条形码825ab(例如,独特标识符序列)、对齐序列825bb和多(A)尾825a。在一些实施方案中,引物衔接子825pa包含以下序列:第一通用引物、其互补序列、其部分序列或其组合。在一些实施方案中,引物衔接子825pa对于所有或一些结合试剂寡核苷酸825可以是相同的。在一些实施方案中,抗体条形码825ab对于所有或一些结合试剂寡核苷酸825可以是相同的。在一些实施方案中,不同结合试剂寡核苷酸825的抗体条形码825ab是不同的。在一些实施方案中,不同结合试剂寡核苷酸825的抗体分子标记825am是不同的。

[0377] 结合试剂寡核苷酸825可以使用多于一种条形码815(例如,与颗粒诸如珠810关联的条形码815)进行条形码化,以产生多于一种条形码化结合试剂寡核苷酸840。在一些实施方案中,条形码815可以包含用于与结合试剂寡核苷酸825,任选地分子标记815m(例如,用于确定结合试剂寡核苷酸的出现数目)、细胞标记815c和通用标记815u结合的多(dT)区815t。在一些实施方案中,条形码815与结合试剂寡核苷酸825的多(dT)区815t杂交。在一些实施方案中,条形码化结合试剂寡核苷酸840通过使与结合试剂寡核苷酸825杂交的条形码815延伸(例如,通过逆转录)而产生。在一些实施方案中,条形码化结合试剂寡核苷酸840包含引物衔接子825pa、抗体分子标记825am(例如,独特分子标记序列)、抗体条形码825ab(例如,独特标识符序列)、对齐序列825bb、多(dT)区815t、分子标记815m、细胞标记815c和通用标记815u。

[0378] 在一些实施方案中,本文公开的条形码化结合试剂寡核苷酸包含两种独特分子标记序列:来源于条形码的分子标记序列(例如,分子标记815m)和来源于结合试剂寡核苷酸的分子标记序列(例如,抗体分子标记825am、样品索引寡核苷酸的第一分子标记序列、细胞组分结合试剂特异性寡核苷酸的第二分子标记序列)。如本文使用的,“双重分子索引”是指采用包含第一独特分子标记序列和第二独特分子标记序列(或其互补序列)的条形码化结合试剂寡核苷酸(或其产物)的本文公开的方法和组合物。在一些实施方案中,本文公开的样品鉴定和细胞组分靶的定量分析的方法可以包括获得条形码分子标记序列和/或结合试剂寡核苷酸分子标记序列的信息序列。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的条形码分子标记序列的数目指示多于一个细胞中的一个

或更多个细胞中至少一种细胞组分靶的拷贝数,细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的结合试剂寡核苷酸分子标记序列的数目指示多于一个细胞中的一个或更多个细胞中至少一种细胞组分靶的拷贝数,细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的结合试剂寡核苷酸分子标记序列和条形码分子标记序列二者的数目指示多于一个细胞中的一个或更多个细胞中至少一种细胞组分靶的拷贝数,细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。

[0379] 在开始测序方案之前使用PCR来扩增物质的量增加了伪影(artifacts)的可能性,诸如当过早终止产物引发随后轮合成时,在扩增期间发生人为重组。在一些实施方案中,本文提供的双重分子索引的方法允许在给定足够测序深度的情况下鉴定PCR嵌合体。另外地,在一些实施方案中,向结合试剂寡核苷酸添加独特分子标记序列增加了随机标记的复杂性。因此,在一些实施方案中,结合试剂寡核苷酸中独特分子标记序列的存在可以克服UMI多样性的限制。在一些实施方案中,与如果不使用该方法 and 组合物相比,本文提供的双重分子索引方法将测序后分子覆盖计算期间标记为“饱和”的细胞组分靶的数目降低了至少约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%、150%、200%、250%、500%、1000%或更高,以及其中的重叠范围)。

[0380] 用于标记核酸靶的方法

[0381] 本文的公开内容包括用于标记核酸靶的系统、方法、组合物和试剂盒。在一些实施方案中,提供了包含裂解区的寡核苷酸条形码。裂解区可以包含一个或更多个裂解位点。在一些实施方案中,提供了能够实现裂解区(或其互补体)的磷酸二酯主链的裂解的裂解剂。寡核苷酸条形码可以与固体支持物关联。在一些实施方案中提供了使条形码化核酸分子与固体支持物分离的方法。

[0382] 在一些实施方案中,提供了使用含有dU的寡核苷酸作为磁珠细胞捕获寡核苷酸序列的一部分以便在PCR1之前用UNG酶从磁珠裂解第一链cDNA的方法。在含有细胞标记和UMI的磁珠捕获mRNA后,可以进行第一链cDNA合成以将每个细胞的mRNA内容编码到代表单细胞的单个珠上。为了对得到的捕获的mRNA内容进行测序,可以在存在珠的初始PCR中对含有mRNA的珠进行扩增。然而,本领域的一个突出问题是,在PCR反应中磁珠的存在可能对完全有效的PCR有抑制性。例如,在一些实施方案中,由于珠中存在能够抑制PCR的铁氧体(Fe_3O_4),当前PCR1不允许高效PCR。为了解决这个问题,在一些实施方案中,提供了包含含有一个或更多个裂解位点(例如,5'端附近的dU)的裂解区的寡核苷酸条形码,所述裂解区使得能够在初始PCR之前使条形码化核酸分子(例如,cDNA)从珠裂解。在一些实施方案中,使用在磁珠细胞捕获寡核苷酸中的包含一个或更多个裂解位点(例如,dU)的裂解区能够实现更高的PCR效率和产量。如本文公开的,在PCR1之前使细胞标记-cDNA从珠裂解的优点是,不存在珠的溶液相PCR可以更有效,并因此允许对mRNA分子/细胞更高的检测灵敏度。

[0383] 与先前使用UNG作为用于在后续PCR反应之前对PCR扩增产物进行去污染的方法不同,本文提供了在PCR之前使用UNG和含有dU的寡核苷酸的方法,具体地该方法在PCR之前使细胞标记从珠脱落以便允许不存在潜在抑制珠的溶液相PCR反应。在一些实施方案中,提供

了在寡核苷酸1的5'端处含有dUTP的Rhapsody珠,并裂解含有U的DNA以释放细胞标记-UMI-cDNA构建体(逆转录后),以允许不存在磁珠的溶液相PCR1反应。

[0384] 在一些实施方案中,在对条形码化反应产物的PCR扩增之前进行DNA的尿嘧啶N-糖苷酶裂解。在本文提供的方法的一些实施方案中,限制性酶和限制性酶被用作可选的裂解位点以实现与dU-UNG组合相同的效果。在采用限制性酶的一些实施方案中,与限制性位点互补的引物与裂解区退火以产生双链限制性结合位点。在一些实施方案中,限制性酶不使单链DNA(例如,cDNA或UMI)裂解。在本文提供的一些实施方案中,使用II型限制性酶进行裂解。

[0385] 在本文提供的方法和组合物的一些实施方案中,使DNA细胞组分结合试剂特异性寡核苷酸(例如,抗体寡核苷酸)与寡核苷酸条形码杂交并延伸以实现用于对从同一珠进行蛋白定量和mRNA定量的分开但并行的工作流程,如2020年1月13日提交的,题为“METHODS AND COMPOSITIONS FOR QUANTITATION OF PROTEINS AND RNA”的美国临时专利申请序号62/960603中描述的,该美国临时专利申请的内容通过引用以其整体并入本文。

[0386] 图14A描绘了采用包含不含有裂解区的寡核苷酸条形码的固体支持物(例如,Rhapsody™珠)的条形码化方案。固体支持物存在于PCR1反应中,如本文公开的,其抑制PCR反应效率。图14B描绘了与本文提供的寡核苷酸条形码关联的固体支持物(例如,具有在C6(或C12)后的CUCUCUCU接头的Rhapsody™珠)。在一些实施方案中,在进行PCR1之前,用UNG(并且在一些实施方案中,另外的裂解剂,诸如例如,AP核酸内切酶)处理珠,并使具有编码的mRNA/AbSeq的珠寡核苷酸从珠裂解。磁珠可以在PCR1之前被去除,并因此通过本文提供的组合物和方法克服了上述珠对PCR的抑制。在一些实施方案中,用UNG(尿嘧啶DNA糖苷酶)处理珠以在dUTP位点裂解。在一些实施方案中,通过用磁体去除(或本文提供的其他去除方法)来防止珠干扰后续反应。

[0387] 在一些实施方案中,提供了用于标记多于一个细胞中的核酸靶的方法。在一些实施方案中,该方法包括:使多于一个细胞中核酸靶的拷贝与包含多于一种寡核苷酸条形码的固体支持物接触,其中每一种寡核苷酸条形码包含第一分子标记、第一通用序列、能够与核酸靶杂交的靶结合区和裂解区,其中裂解区位于第一通用序列的5'并包含一个或更多个裂解位点;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条形码化核酸分子,所述多于一种条形码化核酸分子各自包含与核酸靶的至少一部分互补的序列和第一分子标记;使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子;获得多于一种分离的条形码化核酸分子或其产物的序列信息,以确定多于一个细胞中的一个或更多个中核酸靶的拷贝数。

[0388] 图12A-图12D示出了用于标记核酸靶的非限制性示例性工作流程的示意图。条形码(例如,随机条形码、寡核苷酸条形码1210)可以包含靶结合区(例如,多(dT)1218),所述靶结合区可以经由多(dA)尾1206结合到核酸靶(例如,多腺苷酸化RNA转录物1202或其他核酸靶,诸如例如,抗体寡核苷酸,无论是与抗体关联还是已与抗体解离的),或其他核酸靶,用于加标记或条形码化(例如,加独特标记)。靶结合区可以包含基因特异性序列、寡(dT)序列、随机多聚体或其任何组合。寡核苷酸条形码1210可以包含裂解区1220。裂解区1220可以包含一个或更多个裂解位点,诸如例如,一个或更多个非典型核苷酸1224(例如,dU)和/或裂解识别位点1222(例如,限制性位点)。寡核苷酸条形码1210也可以包含多个标记。寡核苷

酸条形码1210可以包含第一分子标记(ML) 1216和样品标记(例如,分区标记、细胞标记(CL) 1214)用于分别标记转录物和/或追踪RNA转录物(或核酸靶,诸如例如,抗体寡核苷酸,无论与抗体关联或已与抗体解离)的样品来源,以及每种条形码1210的第一分子标记1216/细胞标记1214区域侧翼的一个或更多个另外的序列(诸如例如,第一通用序列1212(例如,读段1序列))用于后续反应。每个样品的寡核苷酸条形码中分子标记的序列的库可以足够大以对RNA转录物进行随机标记。在一些实施方案中,样品标记是分区标记。在一些实施方案中,样品标记是细胞标记。在一些实施方案中,条形码与固体支持物(例如,颗粒1208)关联。多于一种条形码1210可以与颗粒1208关联。在一些实施方案中,颗粒是珠。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。凝胶珠可以例如通过将一种或更多种聚合物前体包封到液滴中来产生。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。多腺苷酸化RNA转录物1202可以包含RNA序列1204r和多(dA)尾1206。

[0389] 该工作流程可以包括使多腺苷酸化RNA转录物1202和寡核苷酸条形码1210杂交1200a。该工作流程可以包括使与多腺苷酸化RNA转录物1202杂交的寡核苷酸条形码1202延伸1200b以产生条形码化核酸分子1226,所述条形码化核酸分子1226包含cDNA 1204c (RNA序列1204r的反向互补序列)。该工作流程可以包括1200c RNA酶H处理和/或使多腺苷酸化转录物变性(例如,加热和/或碱变性)和与条形码化核酸分子1226分离。

[0390] 该工作流程可以包括如图12C中描绘的使条形码化核酸分子1226与裂解剂(例如, DNA糖苷酶,诸如UNG)接触1200d,产生包含AP位点1230的条形码化核酸分子1228。该工作流程可以包括使条形码化核酸分子1228与AP裂解剂接触1200e,从而产生裂解的条形码化核酸分子1232。该工作流程可以包括使裂解的条形码化核酸分子1232与固体支持物分离1200f。该工作流程还可以包括下游1200g引物延伸和/或扩增反应和1200h测序,如本文公开的。

[0391] 该工作流程可以包括使裂解寡核苷酸1234(例如,裂解寡核苷酸)与条形码化核酸分子1226杂交1200i。裂解寡核苷酸1234可以包含裂解识别位点的互补体1236。使裂解寡核苷酸与单链裂解区杂交可以产生双链裂解区。该工作流程可以包括使条形码化核酸分子与一种或更多种裂解剂(例如,能够实现条形码化核酸分子的双链裂解区的磷酸二酯主链的裂解)诸如限制性酶接触1200j,从而产生裂解的条形码化核酸分子1226。该工作流程可以包括1200f使裂解的条形码化核酸分子1226与固体支持物分离。该工作流程还可以包括下游1200g引物延伸和/或扩增反应和1200h测序,如本文公开的。

[0392] 多于一种条形码化核酸分子各自可以包含单链裂解区。使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子可以包括:使裂解寡核苷酸与单链裂解区杂交以产生双链裂解区;使条形码化核酸分子与能够在一个或更多个裂解位点处实现条形码化核酸分子的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。使多于一种条形码化核酸分子与固体支持物分离可以包括:使条形码化核酸分子与能够在一个或更多个裂解位点处实现条形码化核酸分子的单链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接

触,以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离,以产生多于一种分离的条形码化核酸分子。使多于一种条形码化核酸分子与固体支持物分离可以包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

[0393] 确定多于一个细胞中的一个或多个中核酸靶的拷贝数可以包括基于与多于一种分离的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、其互补体或其组合的数目来确定多于一个细胞中核酸靶的拷贝数。该方法可以包括:使随机引物与多于一种分离的条形码化核酸分子接触,其中随机引物中的每一种包含第三通用序列或其互补体;以及使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物。该方法可以包括: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三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。扩增多于一种延伸产物可以包括将测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到多于一种延伸产物。该方法可以包括: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第一分子标记的数目,确定多于一个细胞中的一个或多个中核酸靶的拷贝数。确定多于一个细胞中的一个或多个中核酸靶的拷贝数可以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第一分子标记的数目来确定多于一个细胞中的一个或多个中多于一种核酸靶中的每一种的数目,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。

[0394] 该方法可以包括: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三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。扩增第一多于一种条形码化扩增子可以包括将测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到第一多于一种条形码化扩增子。该方法可以包括: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第一分子标记的数目,确定多于一个细胞中的一个或多个中核酸靶的拷贝数。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可以包含全转录组扩增(WTA)产物。

[0395] 该方法可以包括:使用多于一种分离的条形码化核酸分子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分离的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。该方法可以包括:获得第三多于一种条形码化扩增子或其产物的序列信息,并且任选地获得所述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。该方法可以包括:基于与第三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第一分子标记的数目,确定多于一个细胞中的一个或多个中核酸靶的拷贝数。

[0396] 在一些实施方案中,(i)使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物和/或(ii)合成第三多于一种条形码化扩增子是分别进行的。在一些实施方案中,使与多于一种分离的条形码化核酸分子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

多于一种分离的条形码化核酸分子杂交的随机引物延伸,和/或其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。DNA聚合酶可以包括Klenow片段。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,诸如鼠白血病病毒(MLV)逆转录酶和/或Moloney鼠白血病病毒(MMLV)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。

[0397] 该方法可以包括:将多于一个细胞分区到多于一个分区。多于一个分区中的分区可以包含来自多于一个细胞的单细胞。多于一个分区中的分区可以包含单个固体支持物。将多于一个细胞分区可以包括将多于一个细胞和包含该固体支持物的多于一个固体支持物分区到多于一个分区,其中多于一个分区中的分区包含来自多于一个细胞的单细胞和该固体支持物。分区可以是孔或液滴。多于一个细胞可以包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。核酸靶可以包括核酸分子(例如,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、细胞组分结合试剂特异性寡核苷酸、样品索引寡核苷酸或其任何组合)。

[0398] 裂解位点和裂解剂

[0399] 本文公开的一个或更多个裂解位点可以包含非典型核苷酸、裂解子识别序列或其任何组合。一个或更多个裂解位点可以包括第一裂解位点和第二裂解位点。第一裂解位点可以包含裂解子识别序列,并且第二裂解位点可以包含一个或更多个非典型核苷酸。非典型核苷酸可以包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1, N6-亚乙烯基腺嘌呤或其任何组合。非典型核苷酸可以包括8-氧代鸟嘌呤(8oxoG)。

[0400] 一种或更多种裂解剂可以包括DNA糖苷酶、AP裂解剂或其任何组合。DNA糖苷酶可以能够裂解非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶(AP)位点。AP裂解剂可以能够裂解与AP位点相邻的磷酸二酯主链。AP裂解剂可以包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合。胺可以包括N,N'-二甲基乙二胺(DMED)。AP裂解酶可以包括AP裂解酶、AP核酸内切酶或其任何组合。一种或更多种裂解剂可以包括APE 1(AP核酸内切酶1)、Endo III(核酸内切酶III)、Endo IV(核酸内切酶IV)、Endo V(核酸内切酶V)、Endo VIII(核酸内切酶VIII)、Fpg(甲酰氨基-嘧啶-DNA糖苷酶)、OGG1(8-氧鸟嘌呤DNA糖苷酶1)、NEIL1(核酸内切酶VIII样1)、T7 Endo I(T7核酸内切酶I)、T4 PDG(T4嘧啶二聚体DNA糖苷酶)、UDG(尿嘧啶DNA糖苷酶)、SMUG1(单链选择性单功能尿嘧啶DNA糖苷酶)、AAG(甲基嘌呤DNA糖苷酶)或其任何组合。一种或更多种裂解剂可以是热稳定的。一种或更多种裂解剂可以包括USER酶(例如,尿嘧啶DNA糖苷酶(UDG)和DNA糖苷酶-裂解酶核酸内切酶VIII的混合物)。

[0401] 一种或更多种裂解剂可以包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。裂解子识别序列可以包含限制性位点,并且靶特异性裂解子可以包含限制性核酸内切酶。靶特异性裂解子可以包括TAL-效应子核酸酶(TALEN),所述TAL-效应子核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。靶特异性裂解子可以包括锌指核酸酶(ZFN),所述锌指核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单

链裂解区的磷酸二酯主链。靶特异性裂解子可以包括CRISPR复合体,所述CRISPR复合体包括与互补于裂解子识别序列的引导RNA (gRNA) 复合的CRISPR酶。裂解子识别序列可以包括前间区邻近基序 (PAM) 位点。CRISPR酶可以包括Cas9、Cpf1、Cas13a、其变体、其衍生物或其任何组合。CRISPR酶可以包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。

[0402] 一种或更多种裂解剂可以包括双链特异性裂解剂。双链特异性裂解剂可以不能实现单链裂解区的磷酸二酯主链的裂解。双链特异性裂解剂可以包括能够与裂解子识别序列特异性结合并裂解双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。在一些实施方案中,双链特异性裂解剂不特异性结合裂解子识别序列。双链特异性裂解剂可以包括DNA糖苷酶和双链特异性AP核酸内切酶。双链特异性裂解剂可以能够实现双链裂解区的两条链的磷酸二酯主链的裂解。双链特异性裂解剂可以能够实现双链裂解区的仅一条链的磷酸二酯主链的裂解。

[0403] 同时测量细胞中的蛋白表达和基因表达的方法

[0404] 本文的公开内容包括同时测量细胞中的蛋白表达和基因表达的方法。在一些实施方案中,该方法包括:使多于一种细胞组分结合试剂与包含多于一种细胞组分靶和核酸靶的拷贝的多于一个细胞接触,其中核酸靶包括mRNA,其中多于一种细胞组分结合试剂中的每一种包括包含用于细胞组分结合试剂的独特标识符和多(A)序列的细胞组分结合试剂特异性寡核苷酸,其中细胞组分结合试剂特异性寡核苷酸包括DNA,并且其中细胞组分结合试剂能够与多于一种细胞组分靶中的至少一种特异性结合。该方法可以包括:将与细胞组分结合试剂关联的多于一个细胞分区到多于一个分区,其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞。该方法可以包括:在包含单细胞的分区中,使包含多于一种寡核苷酸条形码的固体支持物与细胞组分结合试剂特异性寡核苷酸和核酸靶的拷贝的接触进行杂交,其中寡核苷酸条形码各自包含多(T)序列、第一通用序列、第一分子标记和裂解区,其中裂解区位于第一通用序列的5'并包含一个或更多个裂解位点。该方法可以包括:使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸,所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含第一分子标记的互补体、裂解区的互补体和第一通用序列的互补体。该方法可以包括: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条形码化核酸分子,所述多于一种条形码化核酸分子各自包含与核酸靶的至少一部分互补的序列和第一分子标记。该方法可以包括:使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。该方法可以包括:使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。该方法可以包括:获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息,以确定多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。该方法可以包括:获得多于一种分离的条形码化核酸分子或其产物的序列信息,以确定多于一个细胞中

的一个或多个中核酸靶的拷贝数。

[0405] 延伸的细胞组分结合试剂特异性寡核苷酸可以各自包含双链裂解区,并且其中多于一种条形码化核酸分子各自包含单链裂解区。使多于一种条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子可以包括:使裂解寡核苷酸与单链裂解区杂交以产生双链裂解区;使条形码化核酸分子与能够在一个或多个裂解位点处实现条形码化核酸分子的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子。使多于一种条形码化核酸分子与固体支持物分离可以包括:使条形码化核酸分子与能够在一个或多个裂解位点处实现条形码化核酸分子的单链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离,以产生多于一种分离的条形码化核酸分子。使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括:使延伸的细胞组分结合试剂特异性寡核苷酸与能够在一个或多个裂解位点处实现延伸的细胞组分结合试剂特异性寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;以及使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。

[0406] 图13A-图13L示出了用于同时测量蛋白和基因表达的非限制性示例性工作流程的示意图。条形码(例如,随机条形码、寡核苷酸条形码1320)可以包含靶结合区(例如,多(dT)1328),所述靶结合区可以经由多(dA)尾1306结合到核酸靶(例如,多腺苷酸化RNA转录物1302或其他核酸靶,诸如例如,抗体寡核苷酸1308,无论是与抗体关联还是已与抗体解离的),或其他核酸靶,用于加标记或条形码化(例如,加独特标记)。靶结合区可以包含基因特异性序列、寡(dT)序列、随机多聚体或其任何组合。寡核苷酸条形码1320可以包含裂解区1330。裂解区1330可以包含一个或多个裂解位点,诸如例如,一个或多个非典型核苷酸1334(例如,dU)和/或裂解识别位点1332(例如,限制性位点)。寡核苷酸条形码1320也可以包含多个标记。寡核苷酸条形码1320可以包含第一分子标记(ML)1326和样品标记(例如,分区标记、细胞标记(CL)1324)用于分别标记转录物和/或追踪RNA转录物(或核酸靶,诸如例如,抗体寡核苷酸,无论与抗体关联或已与抗体解离)的样品来源,以及每种条形码1320的第一分子标记1326/细胞标记1324区域侧翼的一个或多个另外的序列(诸如例如,第一通用序列1322(例如,读段1序列))用于后续反应。每个样品的寡核苷酸条形码中分子标记的序列的库可以足够大以对RNA转录物进行随机标记。在一些实施方案中,样品标记是分区标记。在一些实施方案中,样品标记是细胞标记。在一些实施方案中,条形码与固体支持物(例如,颗粒1318)关联。多于一种条形码1320可以与颗粒1318关联。在一些实施方案中,颗粒是珠。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。凝胶珠可以例如通过将一种或更多种聚合物前体包封到液滴中来产生。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。多腺苷酸化RNA转录物1302可以包含RNA序列1304r和多(dA)尾1306。细胞组分结合试剂特异性寡核苷酸1308可以包含第二通用序列1312、分子标记(例如,第二分子标记1314)、独特标识符序列

1316、与靶结合区互补的序列(例如,多(A)尾1306),或其互补体。在一些实施方案中,细胞组分结合试剂特异性寡核苷酸1308与细胞组分结合试剂(例如,抗体1310)关联。

[0407] 该工作流程可以包括使细胞组分结合试剂特异性寡核苷酸1308和寡核苷酸条形码1320杂交1300a。该工作流程可以包括使多腺苷酸化RNA转录物1302和寡核苷酸条形码1320杂交1300a。该工作流程可以包括使与寡核苷酸条形码1320杂交的细胞组分结合试剂特异性寡核苷酸1308延伸1300b以产生延伸的细胞组分结合试剂特异性寡核苷酸1336,所述延伸的细胞组分结合试剂特异性寡核苷酸1336包含第一分子标记的互补体1326rc、细胞标记的互补体1324rc和第一通用序列的互补体1322rc、一个或更多个非典型核苷酸的互补体1334rc和裂解识别位点的互补体1332rc。在一些实施方案中,延伸反应1300b可以包括使与细胞组分结合试剂特异性寡核苷酸1336杂交的寡核苷酸条形码1320延伸以产生第一条条形码化核酸分子1338,所述第一条条形码化核酸分子1338包含独特标识符序列的互补体1316rc、第二分子标记的互补体1314rc和第二通用序列的互补体1312rc。在一些实施方案中,延伸反应1300b可以包括使与多腺苷酸化RNA转录物1302杂交的寡核苷酸条形码1320延伸以产生第二条条形码化核酸分子1340,所述第二条条形码化核酸分子1340包含cDNA 1304c(RNA序列1304r的反向互补序列)。

[0408] 该方法可以包括:使延伸的细胞组分结合试剂特异性寡核苷酸与双链特异性裂解剂接触以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离;使条形码化核酸分子与一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离(例如,图13C-图13D中描绘的工作流程)。

[0409] 该方法可以包括:使多于一种条形码化核酸分子和多于一种延伸的细胞组分结合试剂特异性寡核苷酸与一种或更多种裂解剂接触以产生:(i)未与固体支持物关联的多于一种裂解的条形码化核酸分子,和(ii)与固体支持物关联的多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使多于一种裂解的条形码化核酸分子与固体支持物分离;使与固体支持物关联的裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性以产生多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸;以及使多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离(例如,图13F-图13I中描绘的工作流程)。

[0410] 该方法可以包括:使延伸的细胞组分结合试剂特异性寡核苷酸与双链特异性裂解剂接触以产生与固体支持物关联的多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使与固体支持物关联的裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性以产生多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸;使多于一种变性的裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离;使条形码化核酸分子与一种或更多种裂解剂接触以产生多于一种裂解的条形码化核酸分子;以及使多于一种裂解的条形码化核酸分子与固体支持物分离以产生多于一种分离的条形码化核酸分子(例如,图13F和图13J-图13L中描绘的工作流程)。

[0411] 在一些实施方案中,使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性。

[0412] 使多于一种条形码化核酸分子与固体支持物分离,可以发生在使多于一种延伸的

细胞组分结合试剂特异性寡核苷酸与固体支持物分离之前。使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离,可以发生在使多于一种条形码化核酸分子与固体支持物分离之前。该方法可以包括:使与细胞组分结合试剂特异性寡核苷酸杂交的多于一种寡核苷酸条形码延伸以产生多于一种条形码化细胞组分结合试剂特异性寡核苷酸。多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸可以包括多于一种分离的条形码化细胞组分结合试剂特异性寡核苷酸。多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸可以包括多于一种裂解的条形码化细胞组分结合试剂特异性寡核苷酸。多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸和多于一种分离的条形码化核酸分子可以在空间上彼此分离。多于一种条形码化细胞组分结合试剂特异性寡核苷酸和多于一种分离的条形码化核酸分子可以在空间上彼此分离。使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离和/或使多于一种条形码化核酸分子与固体支持物分离可以包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

[0413] 细胞组分结合试剂特异性寡核苷酸可以包含第二通用序列。获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息可以包括: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,扩增多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸;以及获得多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的测序数据。细胞组分结合试剂特异性寡核苷酸可以包含第二分子标记。多于一种细胞组分结合试剂特异性寡核苷酸中的至少10种可以包含不同的第二分子标记序列。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目指示多于一个细胞中的一个或多个中至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示多于一个细胞中的一个或多个中至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。获得序列信息可以包括将测序衔接子附接到多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物和/或多于一种分离的条形码化核酸分子或其产物。

[0414] 确定多于一个细胞中的一个或多个中核酸靶的拷贝数可以包括基于与多于一种分离的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、其互补体或其组合的数目来确定多于一个细胞中核酸靶的拷贝数。

[0415] 该方法可以包括:使随机引物与多于一种分离的条形码化核酸分子接触,其中随机引物中的每一种包含第三通用序列或其互补体;以及使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物。该方法可以包括: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三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。扩增多于一种延伸产物可以包括将

测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到多于一种延伸产物。该方法可以包括：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第一分子标记的数目，确定样品中核酸靶的拷贝数。确定样品中核酸靶的拷贝数可以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第一分子标记的数目来确定样品中多于一种核酸靶中的每一种的数目，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。

[0416] 该方法可以包括：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三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。扩增第一多于一种条形码化扩增子可以包括将测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分的序列添加到第一多于一种条形码化扩增子。该方法可以包括：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第一分子标记的数目，确定样品中核酸靶的拷贝数。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可以包含全转录组扩增 (WTA) 产物。

[0417] 该方法可以包括：使用多于一种分离的条形码化核酸分子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分离的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。该方法可以包括：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附接到第三多于一种条形码化扩增子或其产物。该方法可以包括：基于与第三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第一分子标记的数目，确定样品中核酸靶的拷贝数。

[0418] 在一些实施方案中，(i) 扩增多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物，(ii) 使与多于一种分离的条形码化核酸分子杂交的随机引物延伸以产生多于一种延伸产物和/或 (iii) 合成第三多于一种条形码化扩增子是分别进行的。在一些实施方案中，使与多于一种分离的条形码化核酸分子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分离的条形码化核酸分子杂交的随机引物延伸，其中使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用缺乏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寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸，和/或其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。DNA聚合酶可以包括Klenow片段。在一些实施方案中，使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸包括使用逆转录酶使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸，和/或其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（例如，病毒逆转录酶，诸如鼠白血病病毒 (MLV) 逆转录酶和/或Moloney鼠白血病病毒 (MMLV) 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。

[0419] 细胞组分结合试剂特异性寡核苷酸可以包含与多 (dA) 序列相邻的对齐序列。细胞组分结合试剂特异性寡核苷酸可以通过接头与细胞组分结合试剂关联。细胞组分结合试剂特异性寡核苷酸可以被配置为可从细胞组分结合试剂脱离。该方法可以包括：使细胞组分结合试剂特异性寡核苷酸与细胞组分结合试剂解离，任选地在使多于一种寡核苷酸条形码与细胞组分结合试剂特异性寡核苷酸接触之前，使细胞组分结合试剂特异性寡核苷酸与细胞组分结合试剂解离。

[0420] 多于一个分区中的分区可以包含单个固体支持物。将多于一个细胞分区可以包括将与多于一种细胞组分结合试剂关联的多于一个细胞和包含该固体支持物的多于一个固体支持物分区到多于一个分区，其中多于一个分区中的分区可以包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞和该固体支持物。分区可以是孔或液滴。该方法可以包括：在使多于一种细胞组分结合试剂与多于一个细胞接触之后，去除多于一种细胞组分结合试剂中未与多于一个细胞接触的一种或更多种细胞组分结合试剂。去除未与多于一个细胞接触的一种或更多种细胞组分结合试剂包括：去除未与多于一种细胞组分靶中对应的至少一种接触的一种或更多种细胞组分结合试剂。多于一个细胞可以包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。

[0421] 用于样品鉴定和用于测量细胞中细胞组分表达的方法

[0422] 在一些实施方案中，提供了用于样品鉴定的方法。在一些实施方案中，该方法包括：使多于一个样品中的每一个分别与多于一种样品索引组合物中的样品索引组合物接触，其中多于一个样品中的每一个包含一个或更多个细胞，所述一个或更多个细胞各自包含一种或更多种细胞组分靶，其中样品索引组合物包含与样品索引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结合，其中样品索引寡核苷酸包含样品索引序列，并且其中多于一种样品索引组合物中的至少两种样品索引组合物的样品索引序列包含不同的序列；使多于一种样品索引组合物的样品索引寡核苷酸与包含多于一种寡核苷酸条形码的固体支持物接触，其中每一种寡核苷酸条形码包含第一分子标记、能够与样品索引寡核苷酸杂交的靶结合区和裂解区，其中裂解区包含一个或更多个裂解位点；使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸以产生多于一种延伸的样品索引寡核苷酸，所述多于一种延伸的样品索引寡核苷酸各自包含第一分子标记的互补体和裂解区的互补体；使多于一种延伸的样品索引寡核苷酸与固体支持物分离以产生多于一种分离的延伸的样品索引寡核苷酸；获得多于一种分离的延伸的样品索引寡核苷酸或其产物的测序数据；以及基于测序数据中多于一种分离的延伸的样品索引寡核苷酸或其产物中的至少一种分离的延伸的样品索引寡核苷酸或其产物的样品索引序列，鉴定多于一个细胞中的至少一个细胞的样品来源。

[0423] 使多于一种延伸的样品索引寡核苷酸与固体支持物分离包括：使延伸的样品索引寡核苷酸与一种或更多种裂解剂接触。延伸的细胞组分结合试剂特异性寡核苷酸各自可以包含双链裂解区。使多于一种延伸的样品索引寡核苷酸与固体支持物分离包括：使延伸的样品索引寡核苷酸与能够在一个或更多个裂解位点处实现延伸的样品索引寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触以产生多于一种裂解的延伸的样品索引寡核苷酸；以及使多于一种裂解的延伸的样品索引寡核苷酸与固体支持物分离以产生多于一种分离的延伸的样品索引寡核苷酸。使多于一种裂解的延伸的样品索引寡核

昔酸与固体支持物分离可以包括使多于一种裂解的延伸的样品索引寡核苷酸变性。

[0424] 使多于一种延伸的样品索引寡核苷酸与固体支持物分离可以包括磁去除、离心、过滤、色谱、沉淀或其任何组合。每一种寡核苷酸条形码可以包含第一通用序列。多于一种延伸的样品索引寡核苷酸可以包含第一通用序列的互补体。样品索引寡核苷酸可以包含第二分子标记。在一些实施方案中,至少两种样品索引寡核苷酸的第二分子标记序列是不同的,并且其中至少两种样品索引寡核苷酸的样品索引序列是相同的。在一些实施方案中,至少两种样品索引寡核苷酸的第二分子标记序列是不同的,并且其中至少两种样品索引寡核苷酸的样品索引序列是不同的。多于一种样品索引组合物中的至少10种、100种或1000种样品索引组合物的样品索引序列可以包含不同的序列。样品索引寡核苷酸可以包含第二通用序列。

[0425] 鉴定至少一个细胞的样品来源可以包括鉴定测序数据中至少一种分离的延伸的样品索引寡核苷酸或其产物中样品索引序列的存在或不存在。鉴定样品索引序列的存在或不存在可以包括: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,扩增至少一种分离的延伸的样品索引寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的样品索引寡核苷酸;获得多于一种扩增的分离的延伸的样品索引寡核苷酸或其产物的测序数据;以及基于测序数据中多于一种扩增的分离的延伸的样品索引寡核苷酸或其产物中对应于至少一种分离的延伸的样品索引寡核苷酸或其产物的扩增的分离的延伸的样品索引寡核苷酸或其产物的样品索引序列来鉴定细胞的样品来源。扩增至少一种分离的延伸的样品索引寡核苷酸或其产物可以包括将测序衔接子附接到多于一种分离的延伸的样品索引寡核苷酸或其产物。

[0426] 使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸可以包括使用逆转录酶使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸。使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸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聚合酶使与多于一种寡核苷酸条形码杂交的样品索引寡核苷酸延伸。DNA聚合酶可以包括Klenow片段。该方法可以包括:去除多于一种样品索引组合物中未结合的样品索引组合物。去除未结合的样品索引组合物可以包括用洗涤缓冲液洗涤来自多于一个样品的每一个中的一个或更多个细胞。去除未结合的样品索引组合物可以包括使用流式细胞术选择与至少一种细胞组分结合试剂结合的细胞。样品索引寡核苷酸可以被配置为可从细胞组分结合试剂脱离。该方法可以包括:使样品索引寡核苷酸与细胞组分结合试剂解离。

[0427] 靶结合区可以包含多(dT)区。样品索引寡核苷酸可以包含多(dA)区。样品索引寡核苷酸可以包含与多(dA)区相邻的对齐序列。样品索引寡核苷酸可以通过接头与细胞组分结合试剂关联。多于一个样品中的样品可以包括多于一个细胞、多于一个单细胞、组织、肿瘤样品或其任何组合。

[0428] 在一些实施方案中,提供了用于测量细胞中细胞组分表达的方法。在一些实施方案中,该方法包括:使多于一种细胞组分结合试剂与包含多于一种细胞组分靶的多于一个细胞接触,其中多于一种细胞组分结合试剂中的每一种包含细胞组分结合试剂特异性寡核苷酸,所述细胞组分结合试剂特异性寡核苷酸包含用于细胞组分结合试剂的独特标识符序列,并且其中细胞组分结合试剂能够与多于一种细胞组分靶中的至少一种特异性结合;使

细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触,其中每一种寡核苷酸条形码包含第一分子标记、能够与细胞组分结合试剂特异性寡核苷酸杂交的靶结合区和裂解区,其中裂解区包含一个或更多个裂解位点;使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸以产生多于一种延伸的细胞组分结合试剂特异性寡核苷酸,所述多于一种延伸的细胞组分结合试剂特异性寡核苷酸各自包含第一分子标记的互补体和裂解区的互补体;使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸;获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息,以确定多于一个细胞中的一个或更多个中多于一种细胞组分靶中的至少一种细胞组分靶的拷贝数。

[0429] 使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离包括:使延伸的细胞组分结合试剂特异性寡核苷酸与一种或更多种裂解剂接触。延伸的细胞组分结合试剂特异性寡核苷酸各自可以包含双链裂解区。使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离可以包括:使延伸的细胞组分结合试剂特异性寡核苷酸与能够在一个或更多个裂解位点处实现延伸的细胞组分结合试剂特异性寡核苷酸的双链裂解区的磷酸二酯主链的裂解的一种或更多种裂解剂接触,以产生多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸;以及使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离以产生多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸。使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离可以包括使多于一种裂解的延伸的细胞组分结合试剂特异性寡核苷酸变性。

[0430] 该方法可以包括:在使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸之前:将与多于一种细胞组分结合试剂关联的多于一个细胞分区到多于一个分区,其中多于一个分区中的分区包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞;在包含单细胞的分区中,使细胞组分结合试剂特异性寡核苷酸与多于一种寡核苷酸条形码接触。多于一个分区中的分区可以包含单个固体支持物。使多于一种延伸的细胞组分结合试剂特异性寡核苷酸与固体支持物分离可以包括磁去除、离心、过滤、色谱、沉淀或其任何组合。

[0431] 每一种寡核苷酸条形码可以包含第一通用序列。多于一种延伸的细胞组分结合试剂特异性寡核苷酸可以包含第一通用序列的互补体。细胞组分结合试剂特异性寡核苷酸可以包含第二通用序列。获得多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的序列信息可以包括: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,扩增多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物,以产生多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸;以及获得多于一种扩增的分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物的测序数据。

[0432] 细胞组分结合试剂特异性寡核苷酸可以包含第二分子标记。多于一种细胞组分结合试剂特异性寡核苷酸中的至少10种可以包含不同的第二分子标记序列。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是相同的。在一些实施方案中,至少两种细胞组分结合试剂特异性寡核苷酸的第二分子标记序列是不同的,并且其中

至少两种细胞组分结合试剂特异性寡核苷酸的独特标识符序列是不同的。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第一分子标记序列的数目指示多于一个细胞中的一个或更多个中至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。在一些实施方案中,测序数据中与用于细胞组分结合试剂的独特标识符序列关联的独特第二分子标记序列的数目指示多于一个细胞中的一个或更多个中至少一种细胞组分靶的拷贝数,所述细胞组分结合试剂能够与至少一种细胞组分靶特异性结合。获得序列信息可以包括将测序衔接子附接到多于一种分离的延伸的细胞组分结合试剂特异性寡核苷酸或其产物。

[0433] 使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸可以包括使用逆转录酶使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸。使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸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聚合酶使与多于一种寡核苷酸条形码杂交的细胞组分结合试剂特异性寡核苷酸延伸。DNA聚合酶可以包括Klenow片段。靶结合区可以包含多(dT)区。细胞组分结合试剂特异性寡核苷酸可以包含多(dA)区,任选地细胞组分结合试剂特异性寡核苷酸可以包含与多(dA)区相邻的对齐序列。细胞组分结合试剂特异性寡核苷酸可以通过接头与细胞组分结合试剂关联。细胞组分结合试剂特异性寡核苷酸可以被配置为可从细胞组分结合试剂脱离。该方法可以包括:使细胞组分结合试剂特异性寡核苷酸与细胞组分结合试剂解离。

[0434] 将多于一个细胞分区可以包括将与多于一种细胞组分结合试剂关联的多于一个细胞和包含该固体支持物的多于一个固体支持物分区到多于一个分区,其中多于一个分区中的分区可以包含来自与细胞组分结合试剂关联的多于一个细胞的单细胞和该固体支持物。分区可以是孔或液滴。该方法可以包括:在使多于一种细胞组分结合试剂与多于一个细胞接触之后,去除多于一种细胞组分结合试剂中未与多于一个细胞接触的一种或更多种细胞组分结合试剂。去除未与多于一个细胞接触的一种或更多种细胞组分结合试剂包括去除未与多于一种细胞组分靶中对应的至少一种接触的一种或更多种细胞组分结合试剂。多于一个细胞可以包括T细胞、B细胞、肿瘤细胞、髓样细胞、血细胞、正常细胞、胎儿细胞、母体细胞或其混合物。

[0435] 变性可以包括加热和/或碱变性。在一些实施方案中,第一通用序列、第二通用序列和/或第三通用序列可以是相同的或不同的。在一些实施方案中,第一通用序列、第二通用序列和/或第三通用序列可以包含测序引物的结合位点和/或测序衔接子、其互补序列和/或其部分。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。

[0436] 对齐序列的长度可以为一个或更多个核苷酸,或长度可以为两个或更多个核苷酸。对齐序列可以包含鸟嘌呤、胞嘧啶、胸腺嘧啶、尿嘧啶或其组合。对齐序列可以包含多(dT)序列、多(dG)序列、多(dC)序列、多(dU)序列或其组合。对齐序列可以在多(dA)区的5'。接头可以包含碳链,任选地碳链可以包含2-30个碳原子,并且还任选地碳链可以包含12个碳原子。接头可以包括5'氨基修饰物C12(5AmMC12)或其衍生物。细胞组分靶可以包括蛋白靶。细胞组分靶可以包括碳水化合物、脂质、蛋白、细胞外蛋白、细胞表面蛋白、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、胞内蛋白或其任何组合。

细胞组分靶可以在细胞表面上。

[0437] 多于一种寡核苷酸条形码中的至少10种可以包含不同的第一分子标记序列。多于一种寡核苷酸条形码可以与固体支持物关联。多于一种寡核苷酸条形码各自可以包含细胞标记。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。与同一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或不同的细胞标记。固体支持物可以包括合成颗粒或平坦表面。样品可以包括单细胞,并且该方法可以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。合成颗粒和单细胞可以在同一分区(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在合成颗粒上、部分地固定在合成颗粒上、包封在合成颗粒中或部分地包封在合成颗粒中。

[0438] 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(例如,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(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,或其任何组合)。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、胶乳、琼脂糖凝胶、纤维素、尼龙、硅酮,以及其任何组合。

[0439] 试剂盒

[0440] 在一些实施方案中,提供了试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记、靶结合区和裂解区,其中裂解区包含一个或多个裂解位点,并且其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。试剂盒可以包含一种或更多种裂解剂。试剂盒可以包含DNA聚合酶或逆转录酶。试剂盒可以包含缓冲液。试剂盒可以包含筒。试剂盒可以包含一种或更多种用于逆转录反应的试剂。试剂盒可以包含一种或更多种用于扩增反应的试剂。试剂盒可以包含USER酶(例如,尿嘧啶DNA糖苷酶(UDG)和DNA糖苷酶-裂合酶核酸内切酶VIII的混合物)。试剂盒可以包含一种或更多种如本文描述的裂解寡核苷酸。

[0441] 在一些实施方案中,一个或多个裂解位点包含非典型核苷酸、裂解子识别序列或其任何组合。在一些实施方案中,一个或多个裂解位点包括第一裂解位点和第二裂解位点,其中第一裂解位点包含裂解子识别序列,并且其中第二裂解位点包含一个或多个非典型核苷酸。在一些实施方案中,非典型核苷酸包括脱氧尿苷、脱氧肌苷、脱氧-7-甲基鸟苷、脱氧-5,6-二羟基胸苷、脱氧-3-甲基腺苷、5-甲基-脱氧胞苷、0-6-甲基-脱氧鸟苷、5-碘-脱氧尿苷、8-氧基-脱氧鸟嘌呤、1,N6-亚乙烯基腺嘌呤或其任何组合。在一些实施方案中,非典型核苷酸包括8-氧代鸟嘌呤(8oxoG)。在一些实施方案中,一种或更多种裂解剂包括DNA糖苷酶、AP裂解剂或其任何组合。在一些实施方案中,DNA糖苷酶能够裂解非典型核苷酸的碱基部分以产生无嘌呤/无嘧啶(AP)位点。在一些实施方案中,AP裂解剂能够裂解与AP位点相邻的磷酸二酯主链。在一些实施方案中,AP裂解剂包括AP裂解酶、胺、加热、碱性条件、酸性条件或其任何组合。在一些实施方案中,胺包括N,N'-二甲基乙二胺(DMED)。在一些实施方案中,AP裂解酶包括AP裂合酶、AP核酸内切酶或其任何组合。在一些实施方案中,一种或更多种裂解剂包括APE 1(AP核酸内切酶1)、Endo III(核酸内切酶III)、Endo IV(核酸内切酶IV)、Endo V(核酸内切酶V)、Endo VIII(核酸内切酶VIII)、Fpg(甲酰氨基-嘧啶-DNA

糖苷酶)、OGG1(8-氧鸟嘌呤DNA糖苷酶1)、NEIL1(核酸内切酶VIII样1)、T7 Endo I(T7核酸内切酶I)、T4 PDG(T4嘧啶二聚体DNA糖苷酶)、UDG(尿嘧啶DNA糖苷酶)、SMUG1(单链选择性单功能尿嘧啶DNA糖苷酶)、AAG(甲基嘌呤DNA糖苷酶)或其任何组合。在一些实施方案中,一种或更多种裂解剂是热稳定的。在一些实施方案中,一种或更多种裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链的靶特异性裂解子。在一些实施方案中,裂解子识别序列包含限制性位点,并且靶特异性裂解子包括限制性核酸内切酶。在一些实施方案中,靶特异性裂解子包括TAL-效应子核酸酶(TALEN),所述TAL-效应子核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括锌指核酸酶(ZFN),所述锌指核酸酶被配置为与裂解子识别序列特异性结合并裂解双链裂解区和/或单链裂解区的磷酸二酯主链。在一些实施方案中,靶特异性裂解子包括CRISPR复合体,所述CRISPR复合体包括与互补于裂解子识别序列的引导RNA(gRNA)复合的CRISPR酶。在一些实施方案中,裂解子识别序列包括前间区邻近基序(PAM)位点。在一些实施方案中,CRISPR酶包括Cas9、Cpf1、Cas13a、其变体、其衍生物或其任何组合。在一些实施方案中,CRISPR酶包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(也被称为Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、Cpf1、C2c1、C2c3、Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas13a、Cas13b、Cas13c、其变体、其衍生物或其任何组合。在一些实施方案中,一种或更多种裂解剂包括双链特异性裂解剂。在一些实施方案中,双链特异性裂解剂不能实现单链裂解区的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂包括能够与裂解子识别序列特异性结合并裂解双链裂解区的磷酸二酯主链的双链特异性靶特异性裂解子。在一些实施方案中,双链特异性裂解剂不特异性结合裂解子识别序列。在一些实施方案中,双链特异性裂解剂包括DNA糖苷酶和双链特异性AP核酸内切酶。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区的两条链的磷酸二酯主链的裂解。在一些实施方案中,双链特异性裂解剂能够实现双链裂解区中仅一条链的磷酸二酯主链的裂解。在一些实施方案中,裂解区包含裂解子识别序列和一个或多个非典型核苷酸。靶结合区可以包含基因特异性序列、寡(dT)序列、随机多聚体或其任何组合。寡核苷酸条形码可以包含相同的样品标记和/或相同的细胞标记。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。多于一种寡核苷酸条形码的每一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中的至少10种可以包含不同的第一分子标记序列。多于一种寡核苷酸条形码可以与固体支持物关联。多于一种寡核苷酸条形码各自可以包含细胞标记。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。与同一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或不同的细胞标记。固体支持物可以包括合成颗粒或平坦表面。样品可以包括单细胞,并且该方法可以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。合成颗粒和单细胞可以在同一分区(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在合成颗粒上、部分地固定在合成颗粒上、包封在合成颗粒中或部分地包封在合成颗粒中。

[0442] 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(例

如,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(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠或其任何组合)。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、胶乳、琼脂糖凝胶、纤维素、尼龙、硅酮,以及其任何组合。多于一种寡核苷酸条形码中的每一种包含接头官能团。合成颗粒可以包含固体支持物官能团。支持物官能团和接头官能团可以彼此关联。接头官能团和支持物官能团可以单独地选自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,以及其任何组合组成的组。

实施例

[0443] 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图限制本公开内容的范围。

[0444] 实施例1

[0445] 用于与蛋白结合试剂关联的寡核苷酸

[0446] 本实施例展示了可以与蛋白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表达和基因表达。寡核苷酸也可用于样品索引,以确定相同或不同样品的细胞。

[0447] 95mer寡核苷酸设计

[0448] 以下方法用于产生用于蛋白表达和基因表达的同时确定或样品索引的候选寡核苷酸序列和相应的引物序列。

[0449] 1. 序列产生和消除

[0450] 以下方法用于产生用于蛋白表达和基因表达的同时确定或样品索引的候选寡核苷酸序列。

[0451] 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)。

[0452] 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。

[0453] 步骤1c. 去除产生和附接的不具有40%至50%范围内的GC含量的序列。

[0454] 步骤1d. 去除各自具有一个或更多个发夹结构的剩余序列。

[0455] 剩余的候选寡核苷酸序列的数目是423。

[0456] 2. 引物设计

[0457] 使用以下方法设计用于剩余的423种候选寡核苷酸序列的引物。

[0458] 2.1N1引物:使用通用N1序列:5'-GTTGTCAAGATGCTACCGTTCAGAG-3'(LSRR序列;SEQ ID NO.3)作为N1引物。

[0459] 2.2N2引物(用于扩增特定样品索引寡核苷酸;例如,图9B-图9D中的N2引物):

[0460] 2.2a. 去除不从N1序列下游开始的候选N2引物。

[0461] 2.2b. 去除在候选寡核苷酸序列最后35bp中重叠的候选N2引物。

[0462] 2.2c. 去除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。

[0463] 2.2d. 使用ILR2序列作为缺省对照(ACACGACGCTCTCCGATCT;SEQ ID NO.4)以最小

化或避免引物-引物相互作用。

[0464] 在423种候选寡核苷酸序列中,设计了用于390种候选寡核苷酸的N2引物。

[0465] 3.过滤

[0466] 以下方法用于过滤剩余的390种候选引物序列。

[0467] 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。

[0468] 3b. 消除具有4个或更多个连续G(>3个G)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。

[0469] 图9A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。

[0470] 200mer寡核苷酸设计

[0471] 使用以下方法产生用于蛋白表达和基因表达的同时确定以及样品索引的候选寡核苷酸序列和相应的引物序列。

[0472] 1.序列产生和消除

[0473] 使用以下产生用于蛋白表达和基因表达的同时确定以及样品索引的候选寡核苷酸序列。

[0474] 1a. 随机产生许多具有期望长度(128bp)的候选序列(100000种序列)。

[0475] 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。

[0476] 1c. 去除产生和附接的不具有40%至50%范围内的GC含量的序列。

[0477] 1d. 基于发夹结构评分分选剩余的候选寡核苷酸序列。

[0478] 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。

[0479] 2.引物设计

[0480] 使用以下方法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。

[0481] 2.1N1引物:使用通用N1序列:5'-GTTGTCAAGATGCTACCGTTCAGAG-3'(LSRR序列;SEQ ID NO.3)作为N1引物。

[0482] 2.2 N2引物(用于扩增特异性样品索引寡核苷酸;例如,图9B和图9C中的N2引物):

[0483] 2.2a. 去除不从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)。

[0484] 2.2b. 去除在靶序列最后100bp中重叠的候选N2引物。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。

[0485] 2.2c. 去除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。

[0486] 2.2d. 使用ILR2序列5'-ACACGACGCTCTCCGATCT-3'(SEQ ID NO.4)作为缺省对照,以最小化或避免引物-引物相互作用。

[0487] 2.2e. 去除在靶序列最后100bp中重叠的候选N2引物。

[0488] 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。

[0489] 3.过滤

[0490] 以下用于过滤剩余的392种候选引物序列。

[0491] 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。

[0492] 3b. 消除具有4个或更多个连续G(>3个G)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。

[0493] 图9B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9B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9C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9D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。

[0494] 总之,这些数据表明,具有不同长度的寡核苷酸序列可以被设计用于蛋白表达和基因表达的同时确定或样品索引。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序列以及多(A)序列。

[0495] 实施例2

[0496] 寡核苷酸关联的抗体的工作流程

[0497] 本实施例展示了使用寡核苷酸缀合的抗体来确定蛋白靶的表达谱的工作流程。

[0498] 将受试者的冷冻细胞(例如,冷冻外周血单个核细胞(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)。寡核苷酸缀合的抗体缀合有1种、2种或3种寡核苷酸(“抗体寡核苷酸”)。抗体寡核苷酸的序列在图10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。使用流式细胞术将被钙黄绿素AM(活细胞)而非Draq7TM(未死亡或透化的细胞)染色的单细胞分选到BD RhapsodyTM筒中。

[0499] 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。靶(例如,CD4)的蛋白表达谱使用BD RhapsodyTM珠和抗体寡核苷酸确定。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,每个条形码包含分子标记、细胞标记和寡(dT)区。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(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。

[0500] 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分试样,一个等分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,并且一个等分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10中图示的抗体寡核苷酸(“PCR 2”),确定细胞中靶的蛋白表达。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。

[0501] 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表达谱和感兴趣的基因的mRNA表达谱。

[0502] 实施例3

[0503] 细胞组分结合试剂寡核苷酸

[0504] 图11A-图11B示出了用于同时确定蛋白表达和基因表达以及用于样品索引的寡核苷酸的非限制性示例性设计。图11A示出了非限制性示例性细胞组分结合试剂寡核苷酸(SEQ ID NO:7),其包含用于抗体缀合的5'氨基修饰物C6(5AmMC6)接头(例如,可以在抗体缀合之前修饰)、通用PCR手柄、抗体特异性条形码序列和多(A)尾。虽然该实施方案描述了长度为25个核苷酸的多(A)尾,但是多(A)尾的长度可以变化。在一些实施方案中,抗体特异性条形码序列是用于蛋白表达谱分析的方法的抗体克隆特异性条形码。在一些实施方案中,抗体特异性条形码序列是用于样品索引的方法的样品标签序列。在一些实施方案中,抗体特异性条形码序列的示例性设计特征是汉明距离大于3,GC含量在40%至60%的范围内,并且不存在预测的二级结构(例如,发夹)。在一些实施方案中,通用PCR手柄用于在文库制备期间将Illumina测序衔接子附接至扩增子的靶向的PCR扩增。在一些实施方案中,高质量的测序读段可以通过减少测序多样性来实现。

[0505] 图11B示出了非限制性示例性细胞组分结合试剂寡核苷酸(SEQ ID NO:8),其包含用于抗体缀合的5'氨基修饰物C12(5AmMC12)接头、引物衔接子(例如,Illumina P7的部分衔接子)、抗体独特分子标识符(UMI)、抗体特异性条形码序列、对齐序列和多(A)尾。虽然该实施方案描述了长度为25个核苷酸的多(A)尾,但是在一些实施方案中,多(A)尾的长度可以在18-30个核苷酸的范围内。除了图11A中描述的那些之外,在一些实施方案中,抗体特异性条形码序列的示例性设计特征(其中“X”表示任何核苷酸)包括,不存在均聚物和不存在计算机模拟预测结合人类转录物、小鼠转录物、Rhapsody系统引物和/或SCMK系统引物的序列。在一些实施方案中,对齐序列包含序列BB(其中B是C、G或T)。在一些实施方案中,提供长度为1个核苷酸和长度多于2个核苷酸的对齐序列。在一些实施方案中,与较短的接头(例如,5AmMC6)相比,5AmMC12接头可以实现更高的效率(例如,对于抗体缀合或抗体缀合之前的修饰)。抗体UMI序列可以包含“VN”和/或“NV”双重体(其中每个“V”是A、C或G中的任何一个,并且其中“N”是A、G、C或T中的任何一个),在一些实施方案中,这通过充当地理标志物和/或减少均聚物的发生率来改进信息学分析。在一些实施方案中,结合试剂寡核苷酸上独特分子标记序列的存在增加了随机标记的复杂性。在一些实施方案中,引物衔接子包含以下序列:第一通用引物、其互补序列、其部分序列或其组合。在一些实施方案中,引物衔接子消除了对通常在测序之前需要的用于附接Illumina测序衔接子的PCR扩增步骤的需要。在一些实施方案中,引物衔接子序列(或其子序列)不是包含引物衔接子序列的测序模板的测序读出的一部分,并且因此不影响包含引物衔接子的模板的读段质量。

[0506] 术语

[0507] 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,除非这种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,可以对上述方法和结构进行各种其他的省略、添加和修改。所有此类修改和改变都旨在落在由所附权利要求书限定的主题的范围

[0508] 关于本文中基本上任何复数和/或单数术语的使用,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,单数形式“一(a)”、“一(an)”和“所述/该(the)”包括复数指代物,除非上下文另外明确指示。除非另外说明,否则在本文中对“或”的任何提及旨在涵盖“和/或”。

[0509] 本领域技术人员将理解,通常,本文使用的术语,并且特别是所附权利要求书(例如,所附权利要求书的主体)中的术语,通常旨在作为“开放性的”术语(例如,术语“包括/包含(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/包含(includes)”应解释为“包括但不限于(includes but is not limited to)”等)。本领域技术人员将进一步理解,如果意图所介绍的权利要求陈述的特定数字,则这样的意图将明确地陈述于权利要求中,并且在这种陈述不存在的情况下,不存在这种意图。例如,作为对理解的帮助,以下所附权利要求书可以包含介绍性措辞“至少一种”和“一种或更多种”的使用,以介绍权利要求陈述。然而,此类措辞的使用不应解读为意味着由不定冠词“一(a)”或“一(an)”介绍权利要求陈述会将任何包含这种介绍的权利要求陈述的具体权利要求限制到包含仅一种这种陈述的实施方案中,甚至当相同的权利要求包括介绍性措辞“一种或更多种”或“至少一种”以及不定冠词诸如“一(a)”或“一(an)”时也是如此(例如,“一(a)”和/或“一(an)”应解释为意指“至少一种”或“一种或更多种”);这对于使用定冠词来介绍权利要求陈述同样适用。此外,即使明确地陈述了介绍的权利要求陈述的特定数字,本领域技术人员将认识到,这种陈述应解释为意指至少所陈述的数字(例如,仅陈述“两种陈述”而没有其他修饰词意指至少两种陈述或两种或更多种陈述)。此外,在使用类似于“A、B和C等中的至少一种”的惯例的那些情况下,通常这种句法结构意在为本领域技术人员将理解该惯例的意义(例如,“具有A、B和C中的至少一种的系统”将包括但不限于具有单独的A,具有单独的B,具有单独的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等中的至少一种”的惯例的那些情况下,通常这种句法结构意在为本领域技术人员将理解该惯例的意义(例如,“具有A、B或C中的至少一种的系统”将包括但不限于具有单独的A,具有单独的B,具有单独的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分离性词语和/或措辞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”应被理解为包括“A”或“B”或“A和B”的可能性。

[0510] 此外,当本公开内容的特征或方面以马库什组(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组的任何个体成员或成员子组描述。

[0511] 如本领域技术人员将理解的,为了任何和所有目的,诸如在提供书面描述方面,本文公开的所有范围还包括该范围的任何和所有可能的子范围和子范围组合。任何列举的范围可以被容易地认为充分地描述了并且使得同一范围能够被分成至少相等的二分之一、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被

容易地分成下三分之一、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 (up to)”、“至少”“大于”“少于”等包括所述及的数字并且指随后可以被分成如以上讨论的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个个体的成员。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。

[0512] 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旨在限制,实际范围和精神由以下权利要求指出。

序列表

<110> 贝克顿迪金森公司

丹尼斯·普罗森

凯瑟琳·拉扎鲁克

<120> 使用含有DU的寡核苷酸的细胞捕获

<130> 68EB-298710-W0

<150> 62/960609

<151> 2020-01-13

<160> 8

<170> PatentIn version 3.5

<210> 1

<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> 1

gttgtcaaga tgctaccggt cagagtacgt ggagttggtg gcccgacccc gagcgctacg 60

agccccccgg aaaaaaaaaa aaaaaaaaaa aaaaa 95

<210> 2

<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> 2

gttgtcaaga tgctaccggt cagagctact gtccgaagtt accgtgtatc taccacgggt 60

ggtttttcga atccggaaaa gatagtaata agtgtttttag ttggaataag tcgcaacttt 120

tggagacggt tacctctcaa tttttctgat ccgtaggccc cccgatctcg gcctcaaaaa 180

aaaaaaaaaa aaaaaaaaaa 200

- <210> 3
<211> 25
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成的寡核苷酸
<400> 3
gttgtcaaga tgctaccggt cagag 25
- <210> 4
<211> 20
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成的寡核苷酸
<400> 4
acacgacgct cttccgatct 20
- <210> 5
<211> 95
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成的寡核苷酸
<400> 5
gttgtcaaga tgctaccggt cagagcccca tgtctagtac ctattggtcc cctatcctca 60
gattcgttta aaaaaaaaaa aaaaaaaaaa aaaaa 95
- <210> 6
<211> 26
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成的寡核苷酸
<400> 6
tttttttttt tttttttttt tttttt 26
- <210> 7
<211> 95
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成的寡核苷酸

<220>
 <221> 5AmMC12
 <222> (1) .. (1)
 <223> 5'氨基修饰物C12
 <400> 7
 gttgtcaaga tgctaccggt cagagattca agggcagccg cgtcacgatt ggatacgact 60
 gttggaccgg aaaaaaaaaa aaaaaaaaaa aaaaa 95
 <210> 8
 <211> 102
 <212> DNA
 <213> 人工序列 (Artificial Sequence)
 <220>
 <223> 合成的寡核苷酸
 <220>
 <221> 5'AmMC12
 <222> (1) .. (1)
 <223> 5'氨基修饰物C12
 <220>
 <221> misc_feature
 <222> (24) .. (25)
 <223> n是a、c、g或t
 <220>
 <221> misc_feature
 <222> (27) .. (30)
 <223> n是a、c、g或t
 <220>
 <221> misc_feature
 <222> (32) .. (33)
 <223> n是a、c、g或t
 <220>
 <221> misc_feature
 <222> (35) .. (75)
 <223> n是a、c、g或t
 <400> 8
 cagacgtgtg ctcttccgat ctvnnvnnnn vnnvnnnnnn nnnnnnnnnn nnnnnnnnnn 60
 nnnnnnnnnn nnnnnbbaaa aaaaaaaaaa aaaaaaaaaa aa 102

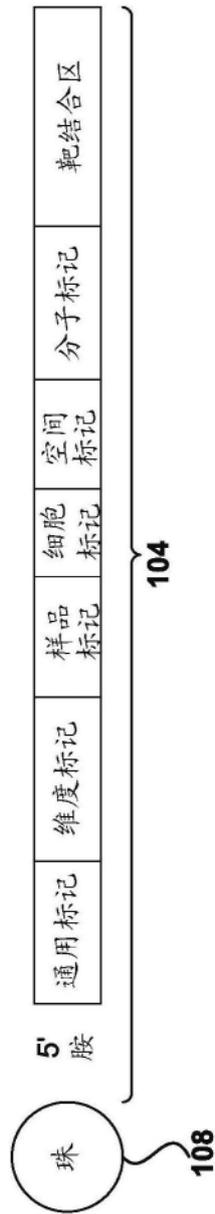


图1

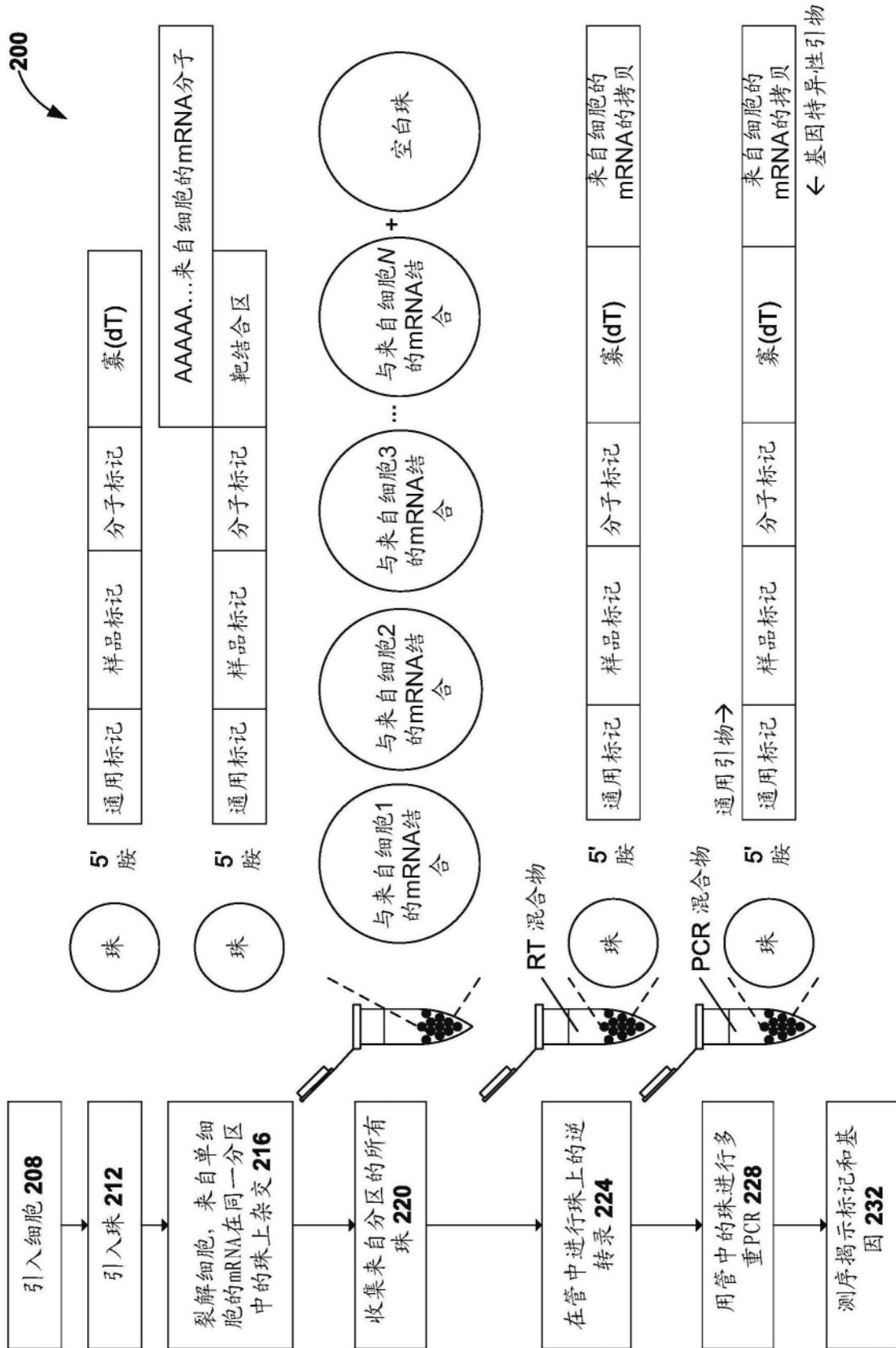


图2

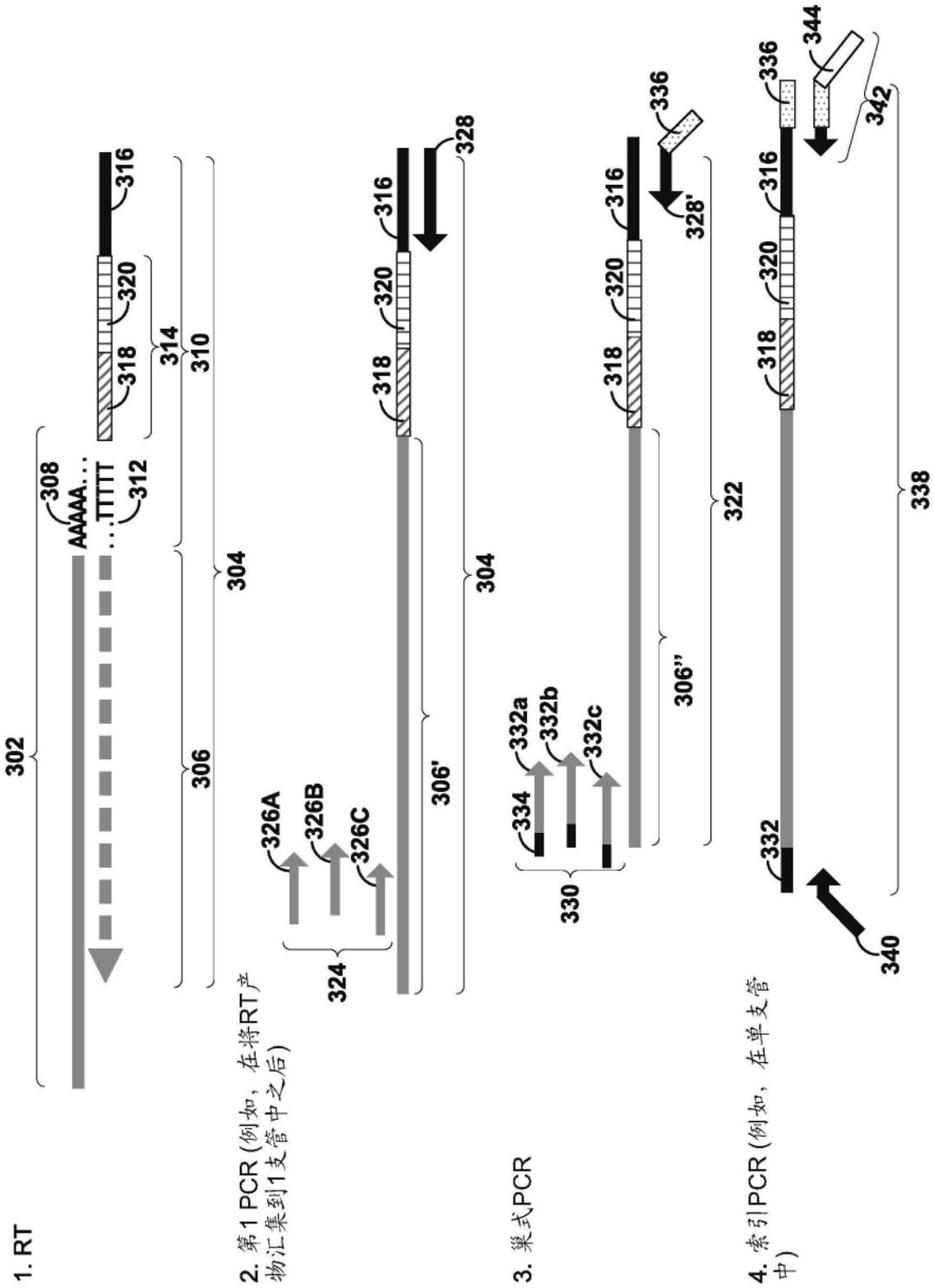


图3

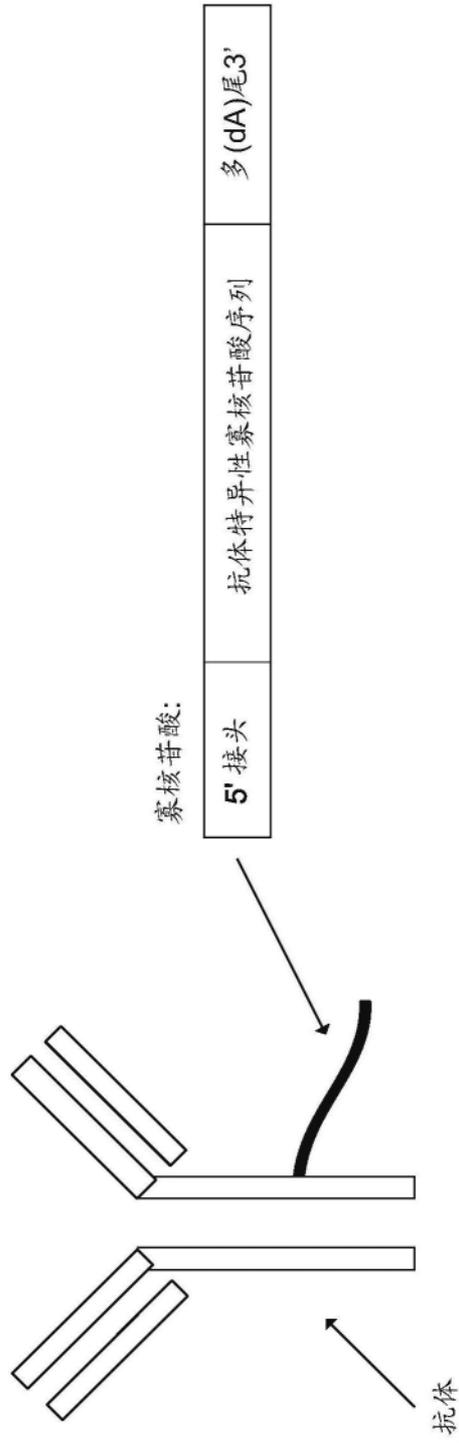


图4

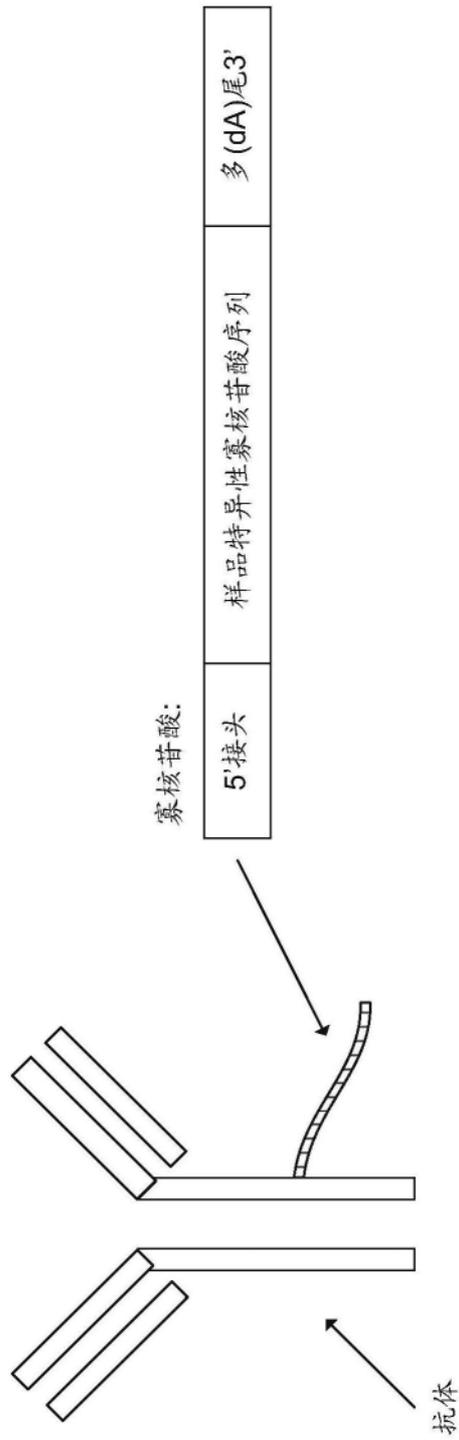


图5

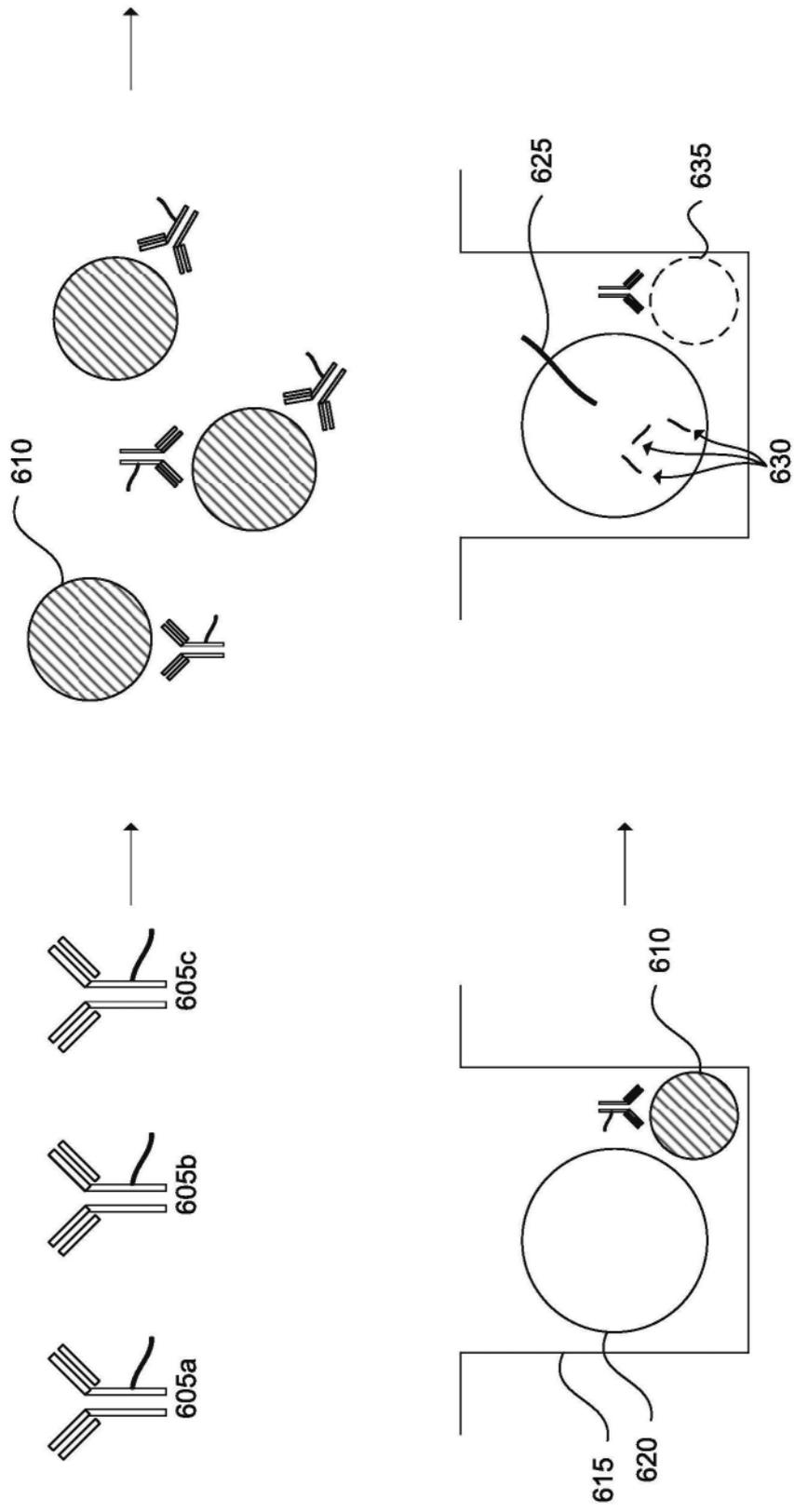


图6

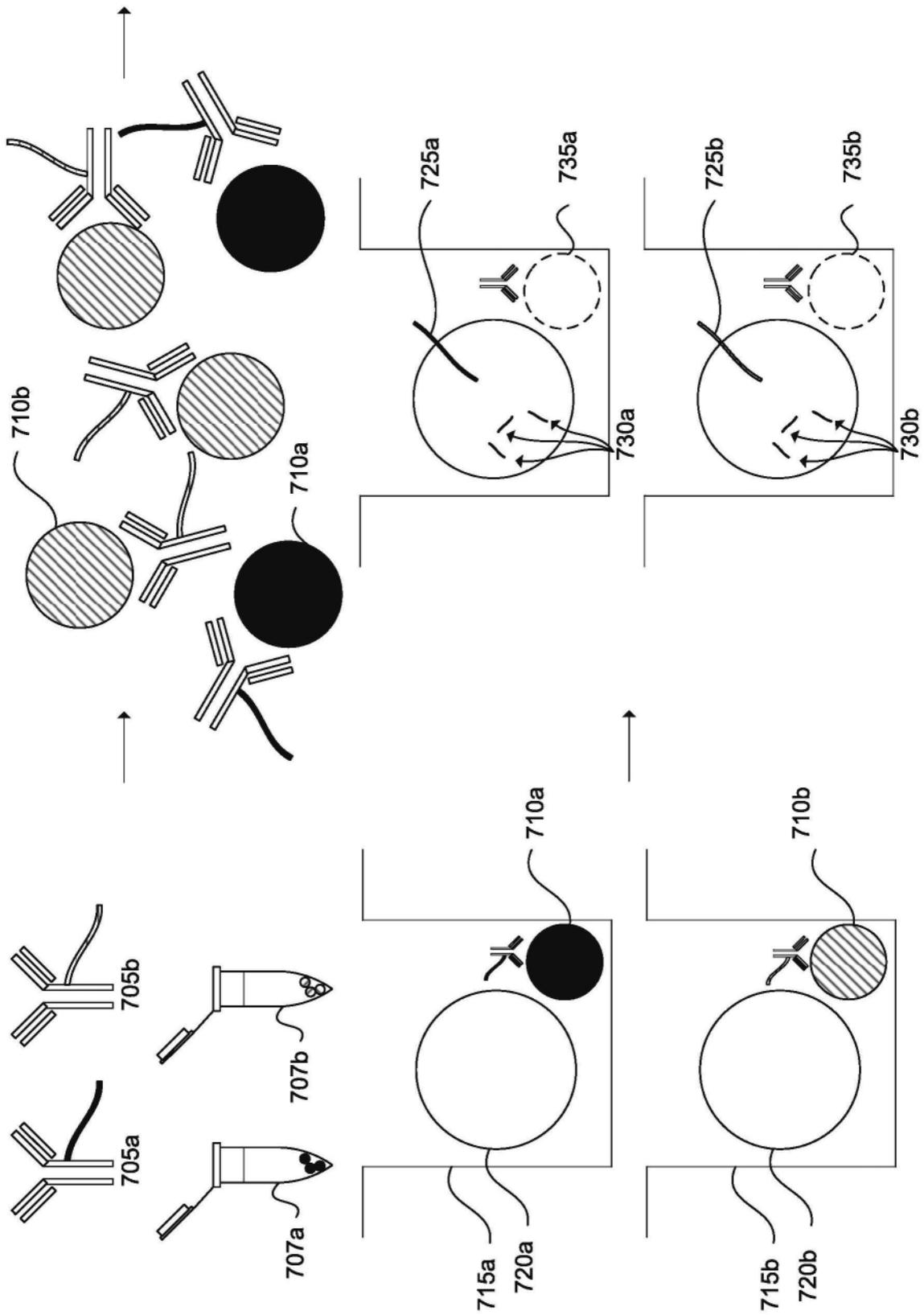


图7

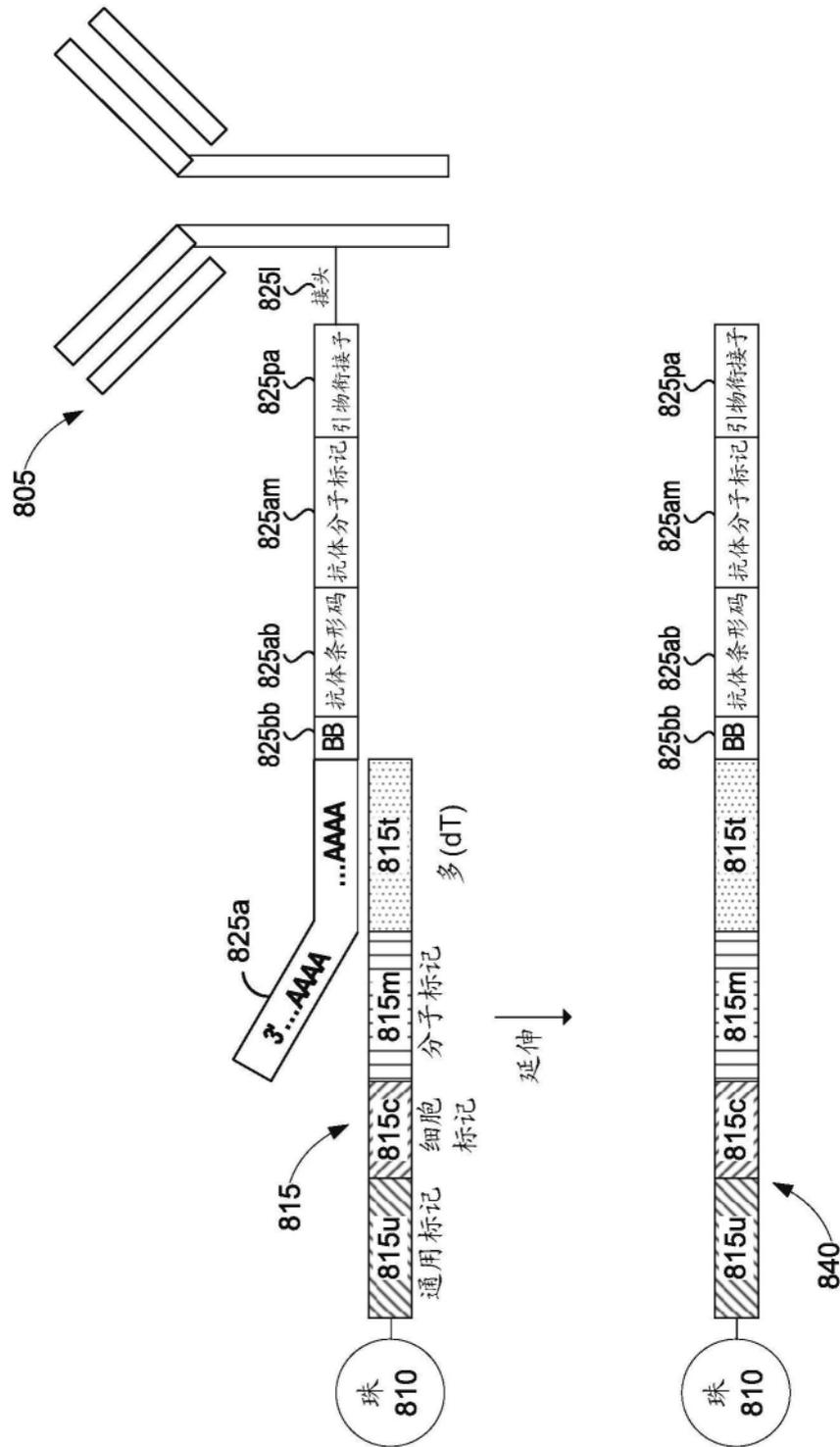


图8

95mer寡核苷酸设计

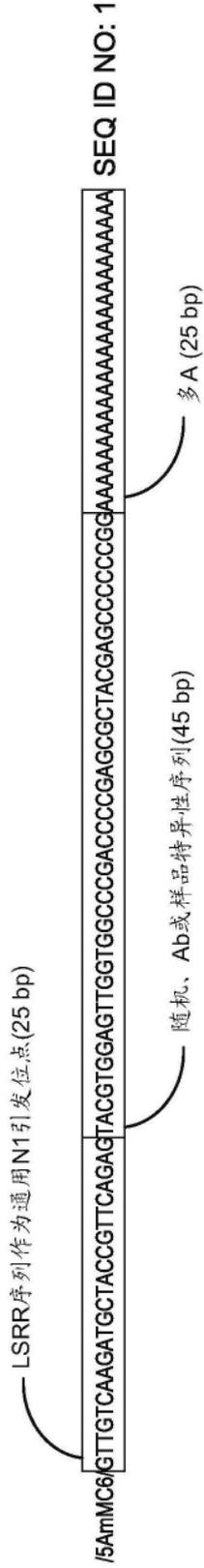


图9A

200mer寡核苷酸设计。靶向扩增方案(巢式特异性)

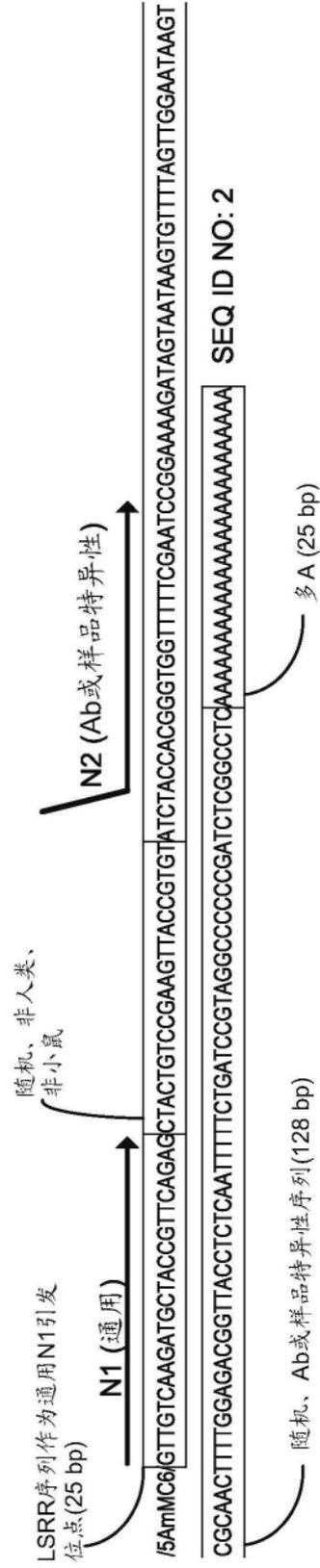


图9B

95mer寡核苷酸设计

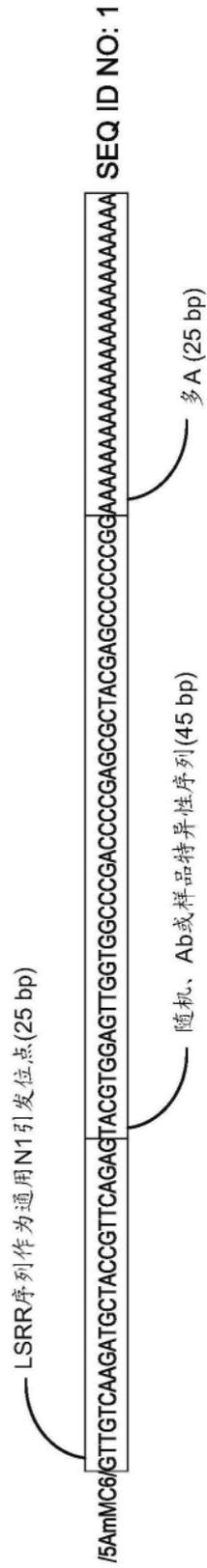


图9C

200mer寡核苷酸设计。靶向扩增方案(巢式特异性)

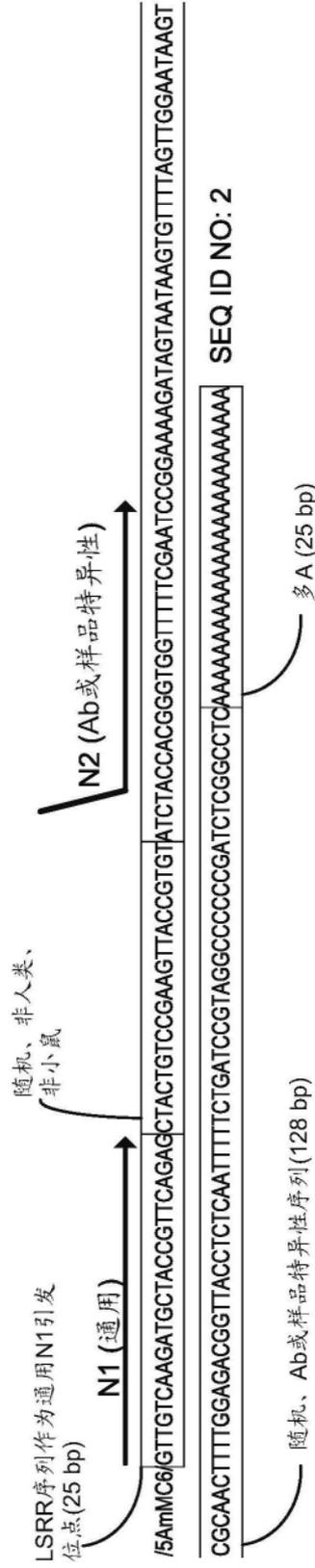


图9D

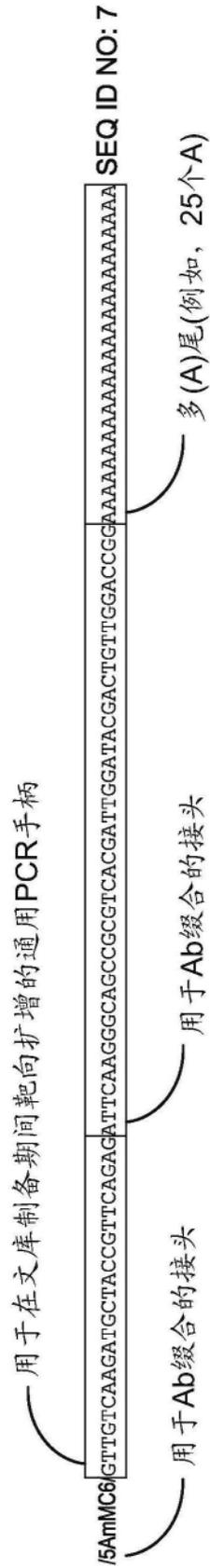


图11A

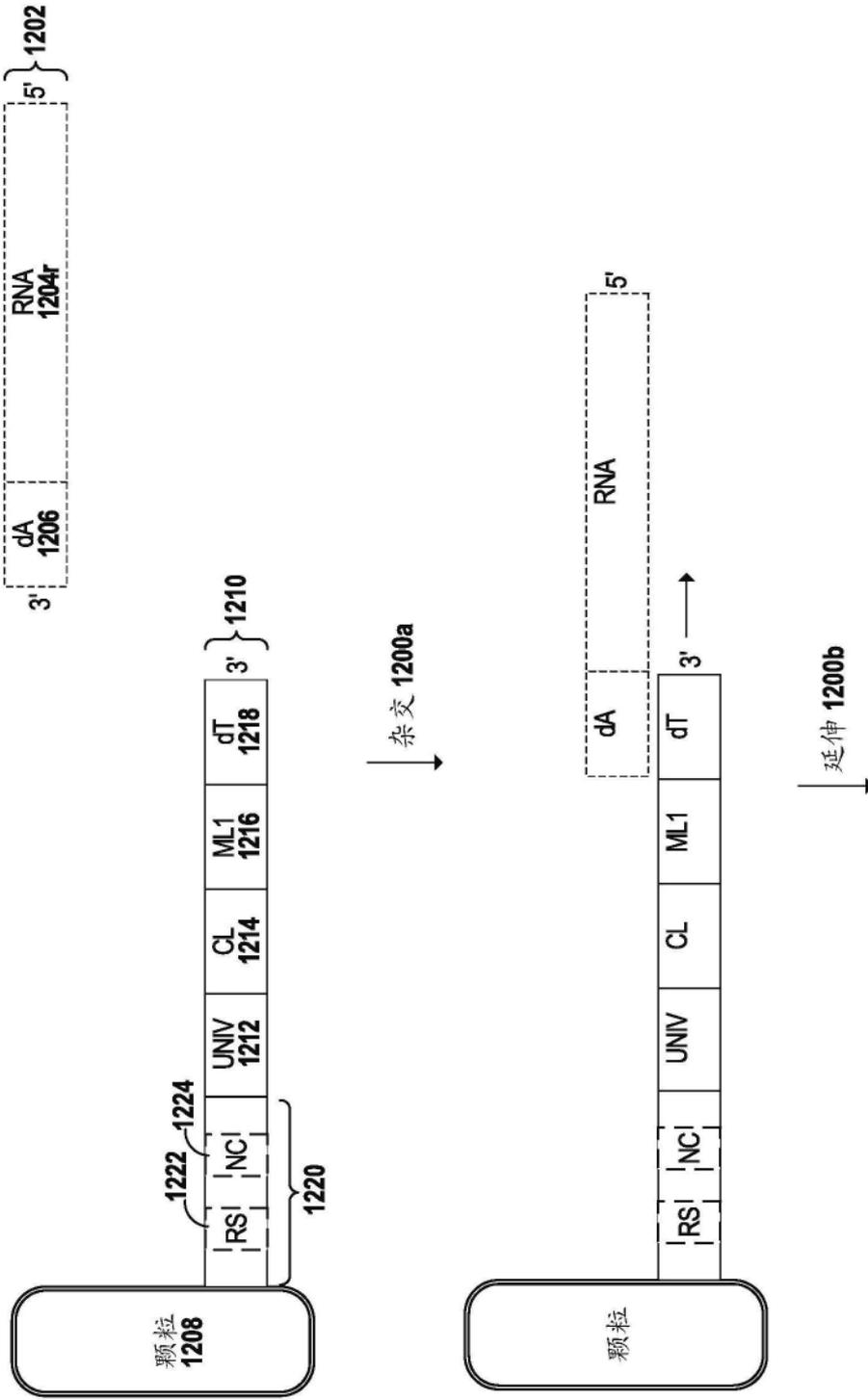


图12A

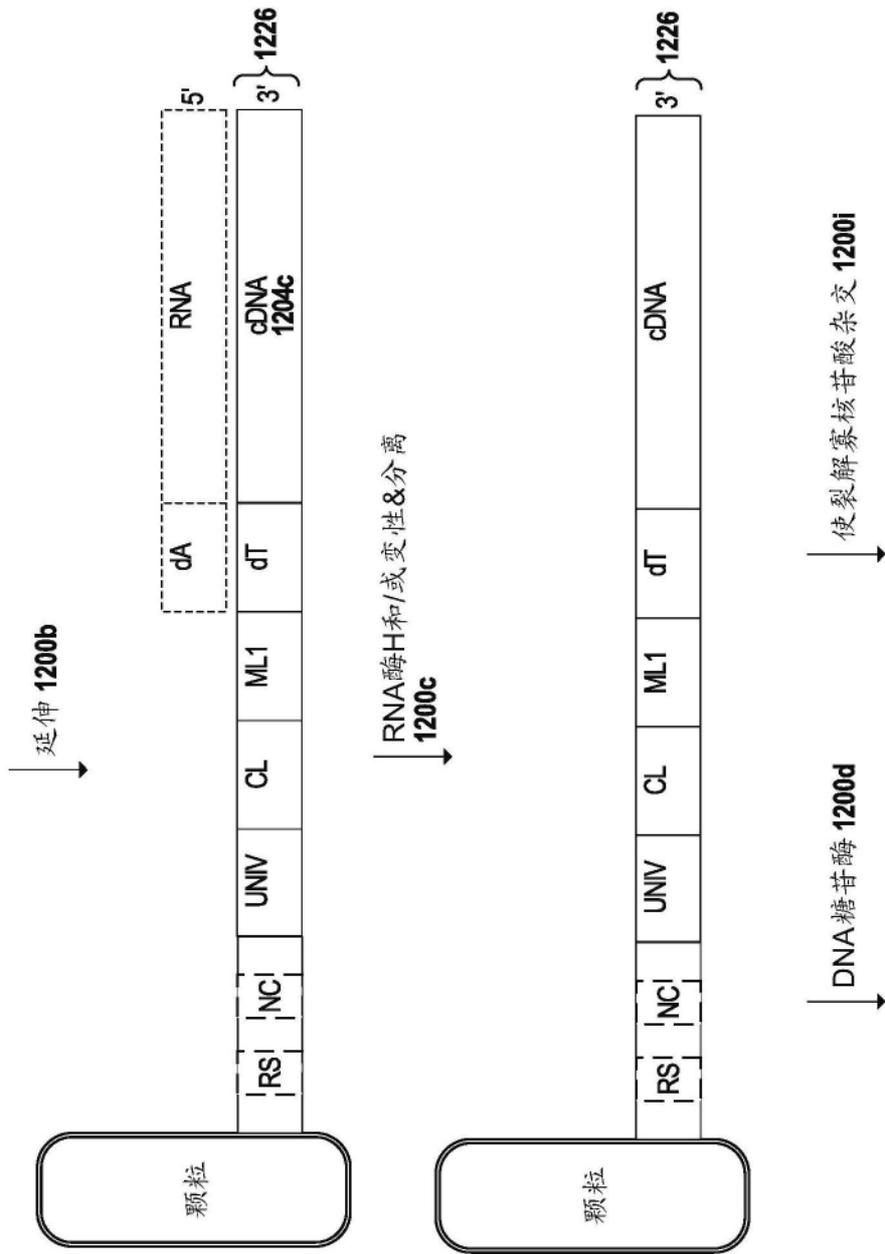


图12B

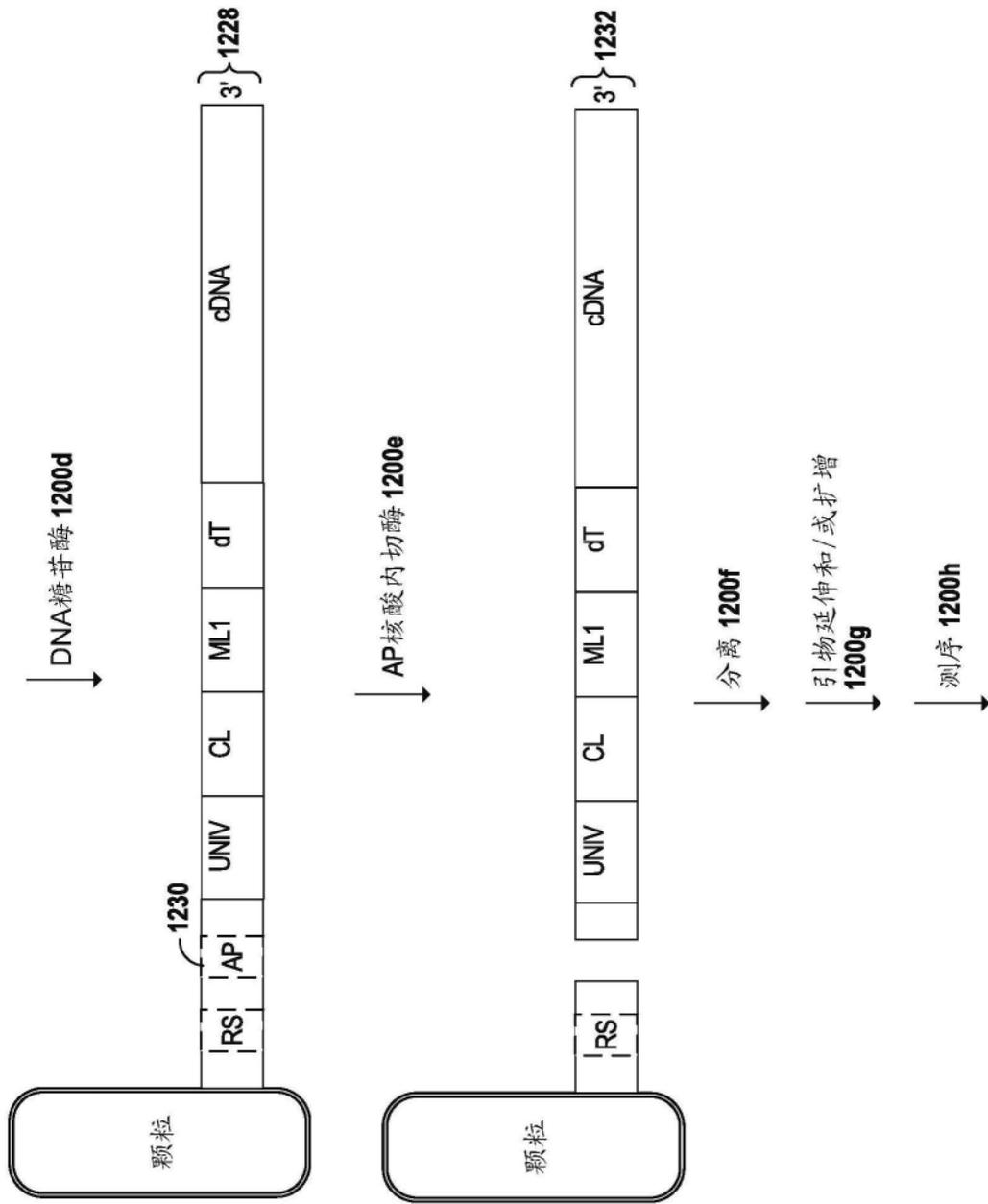


图12C

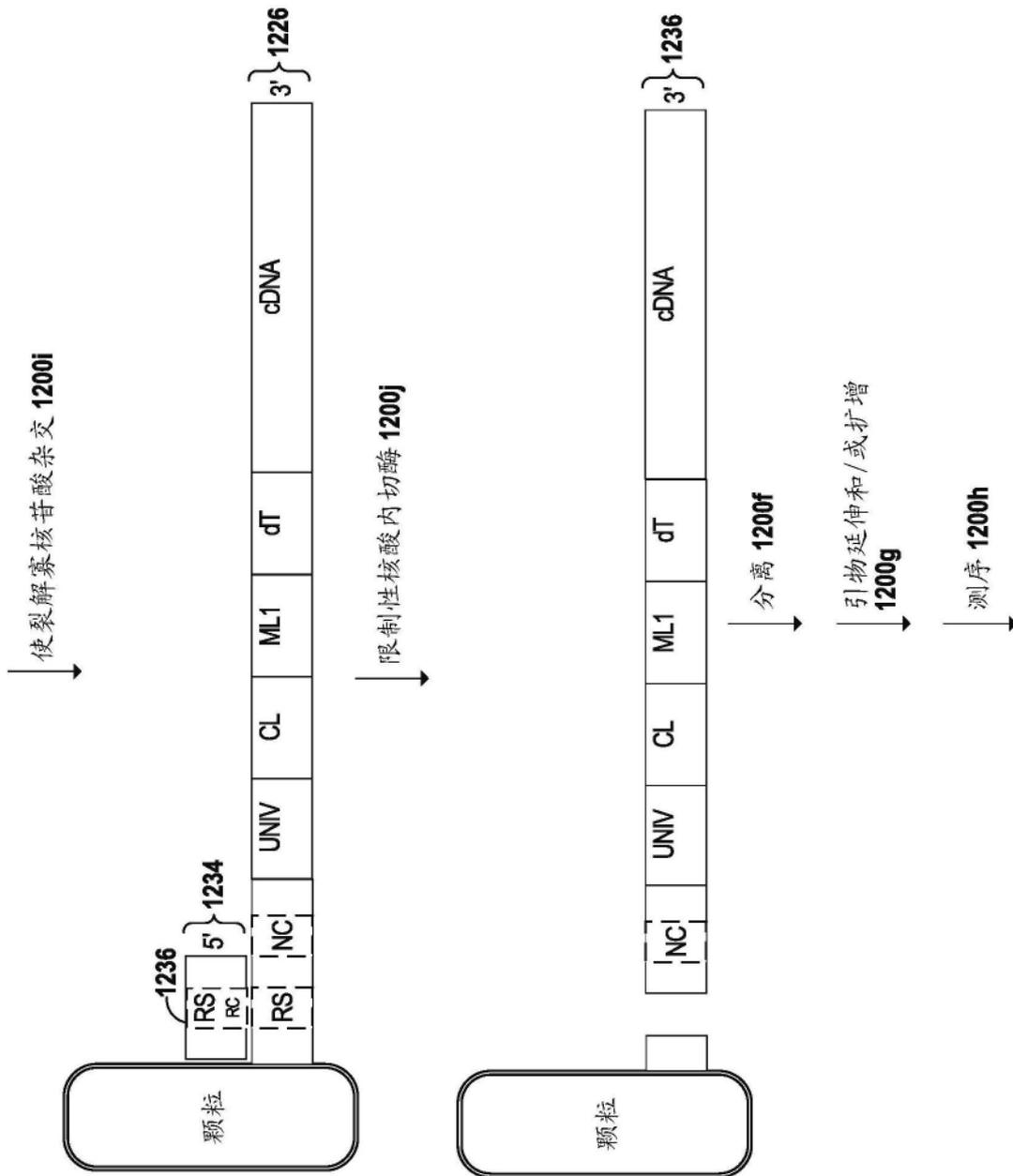


图12D

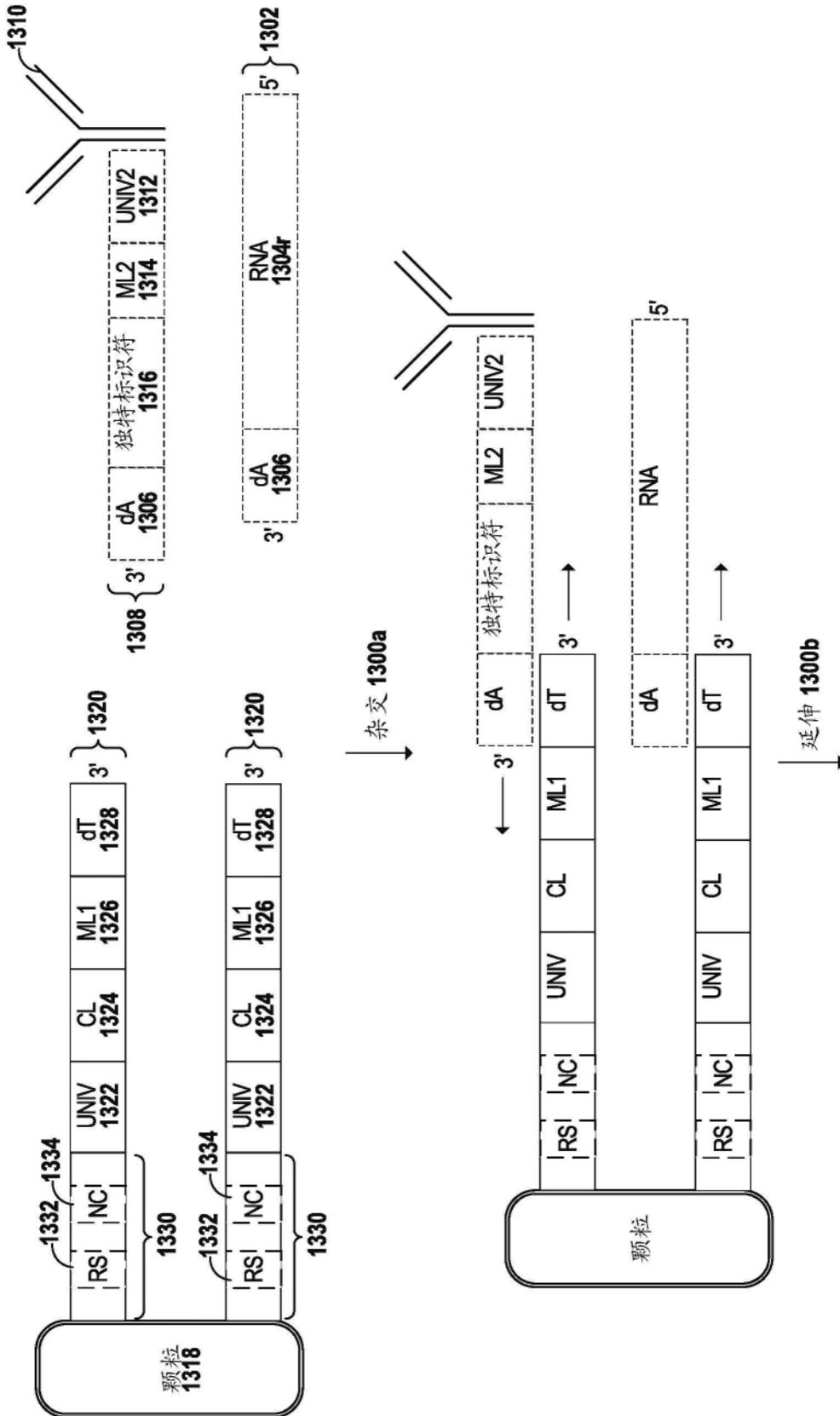


图13A

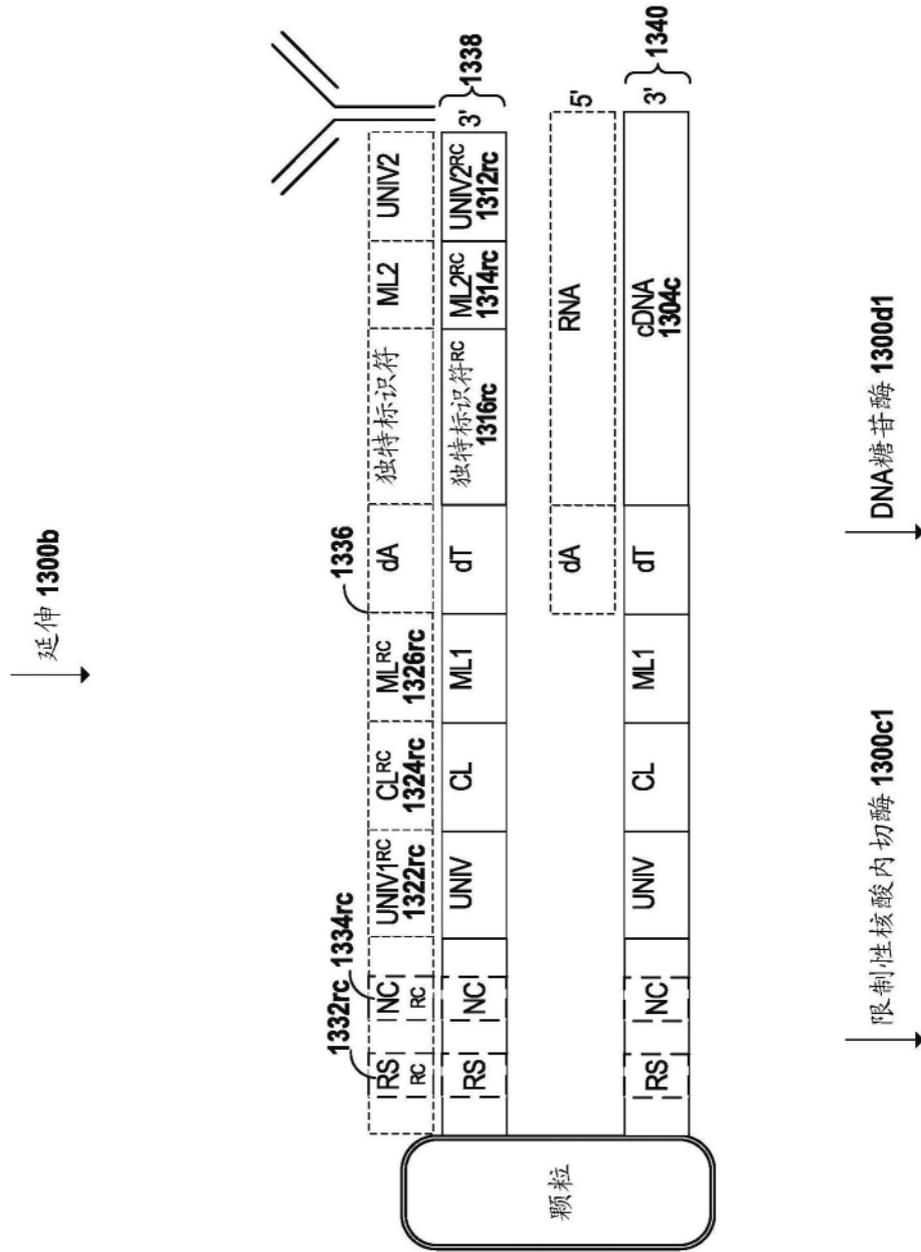


图13B

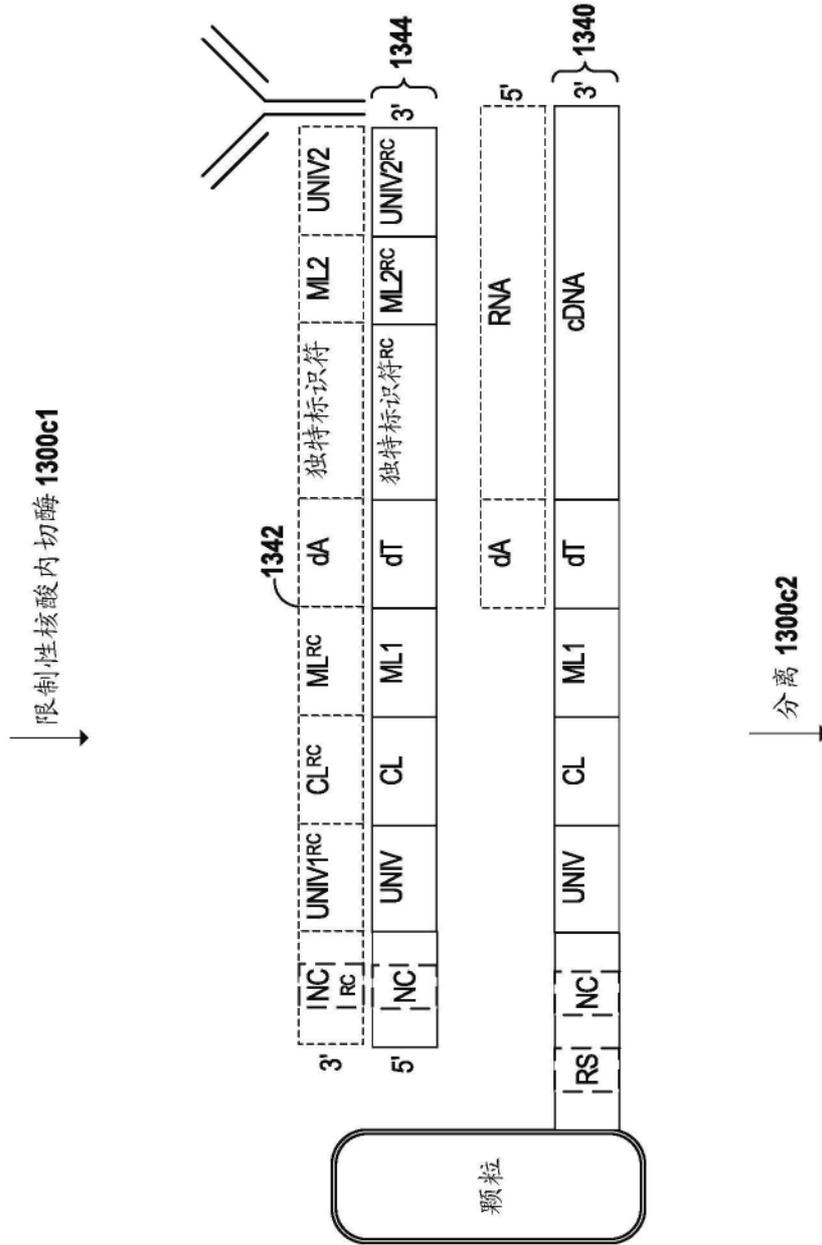


图13C

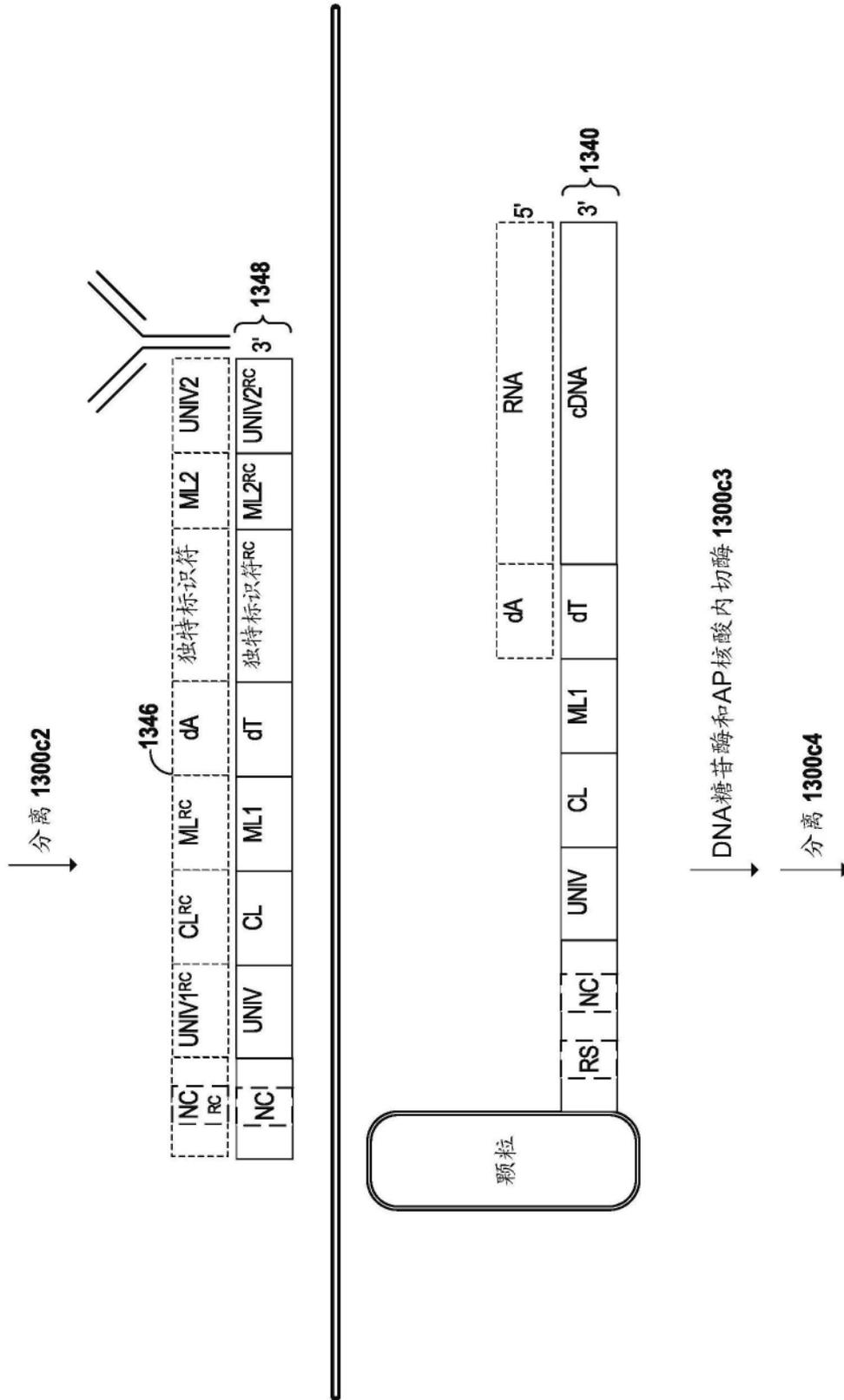


图13D

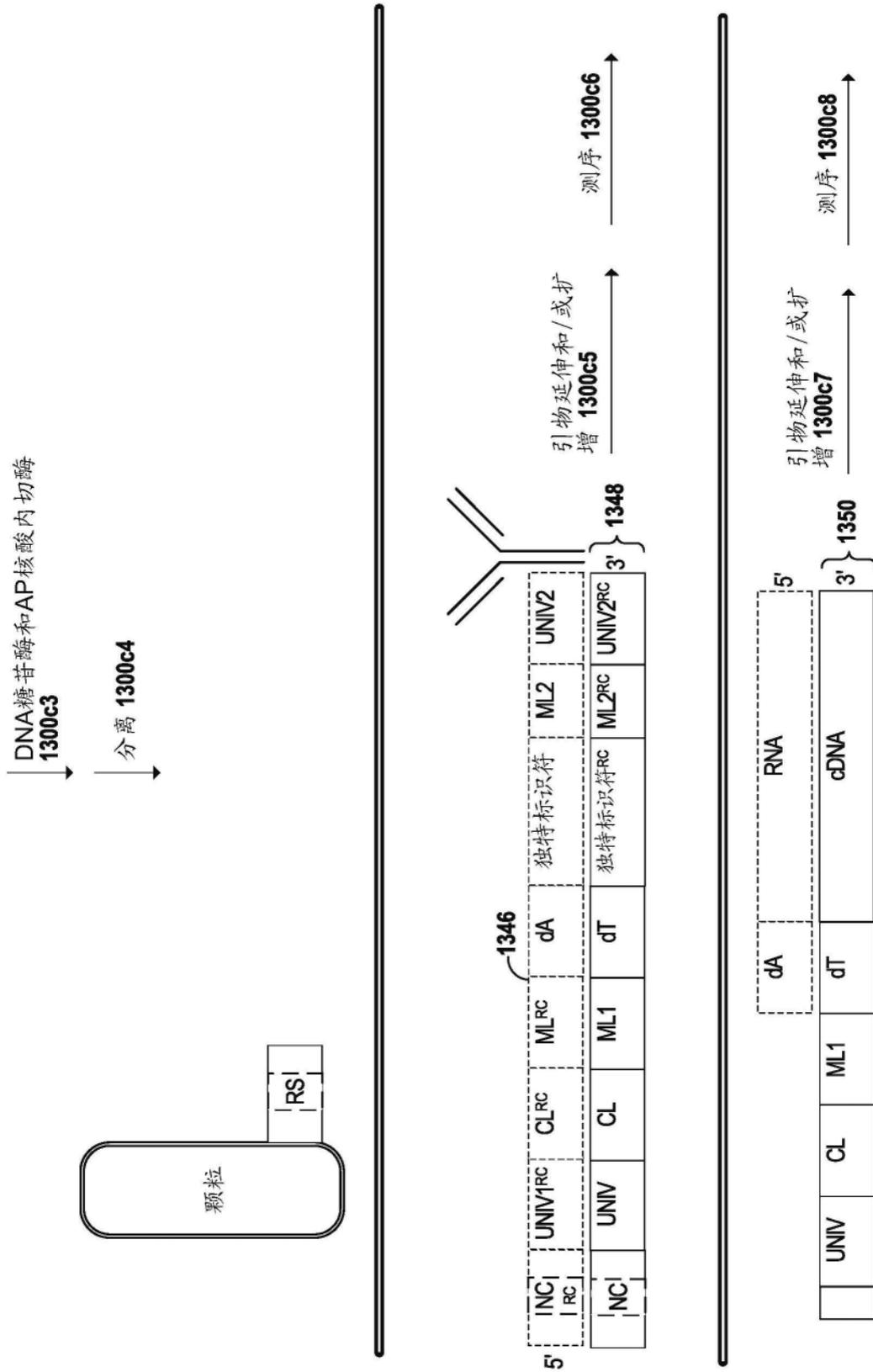


图13E

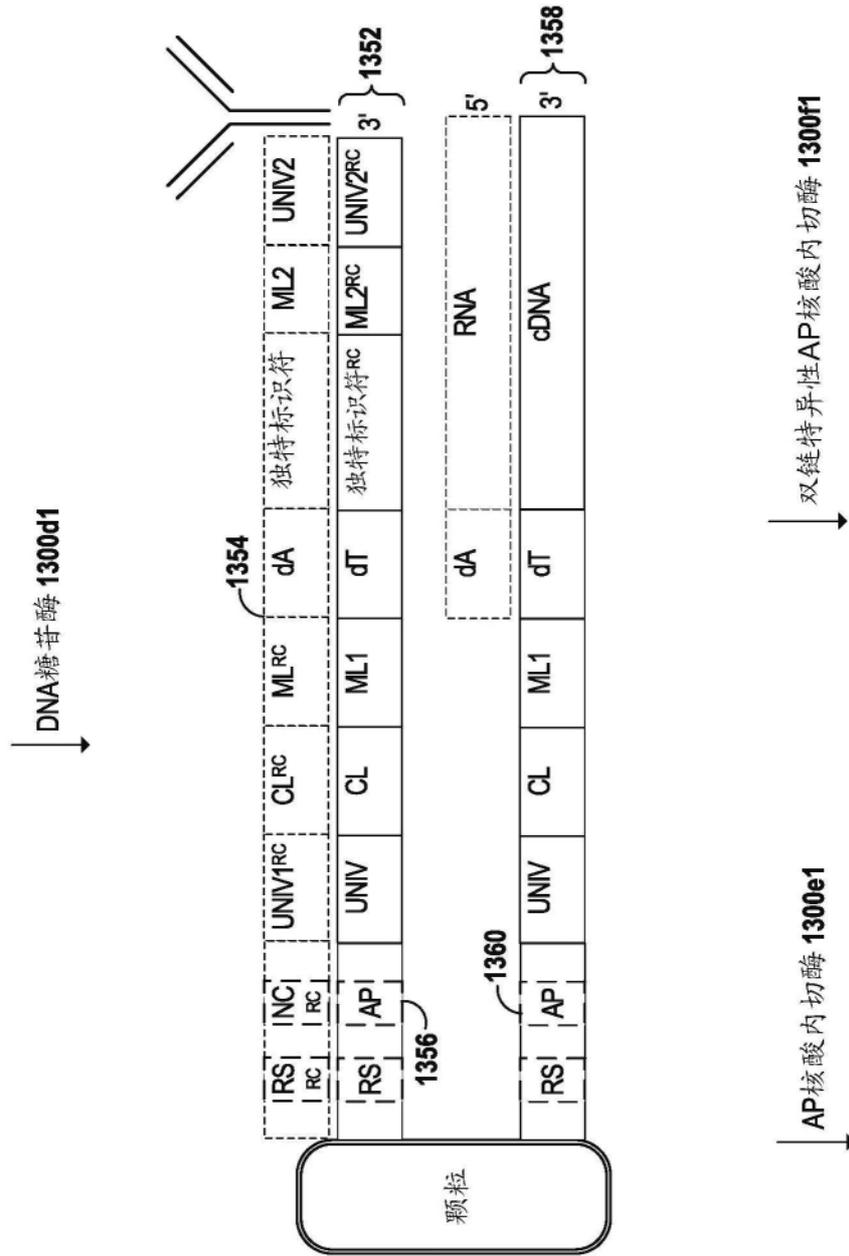


图13F

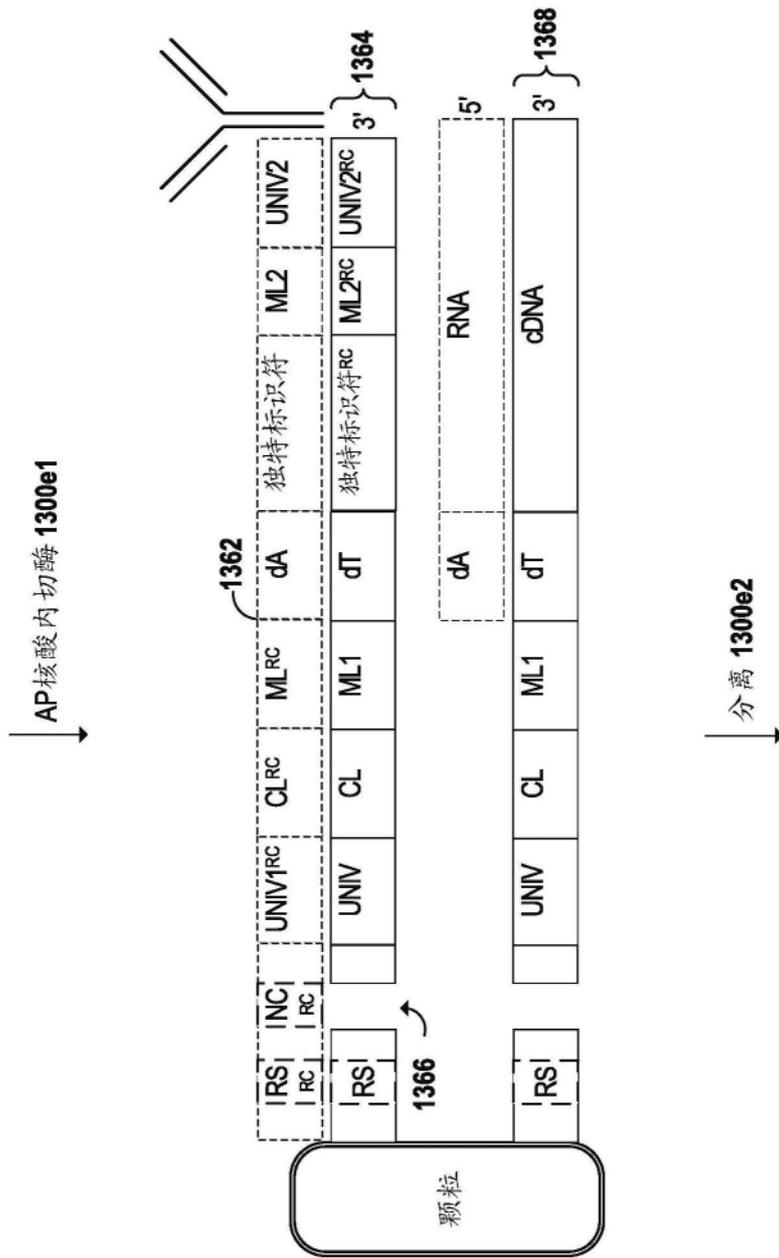


图13G

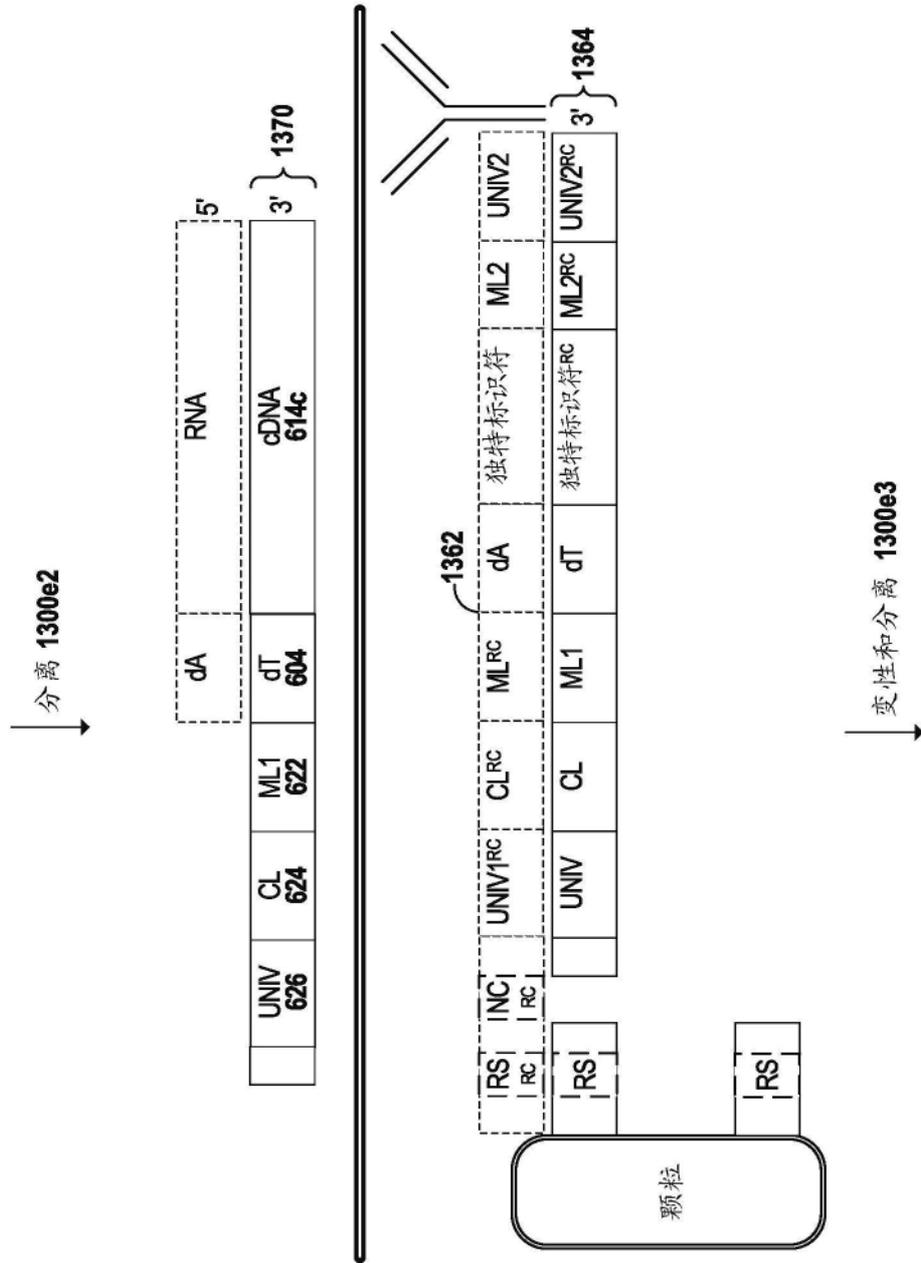


图13H

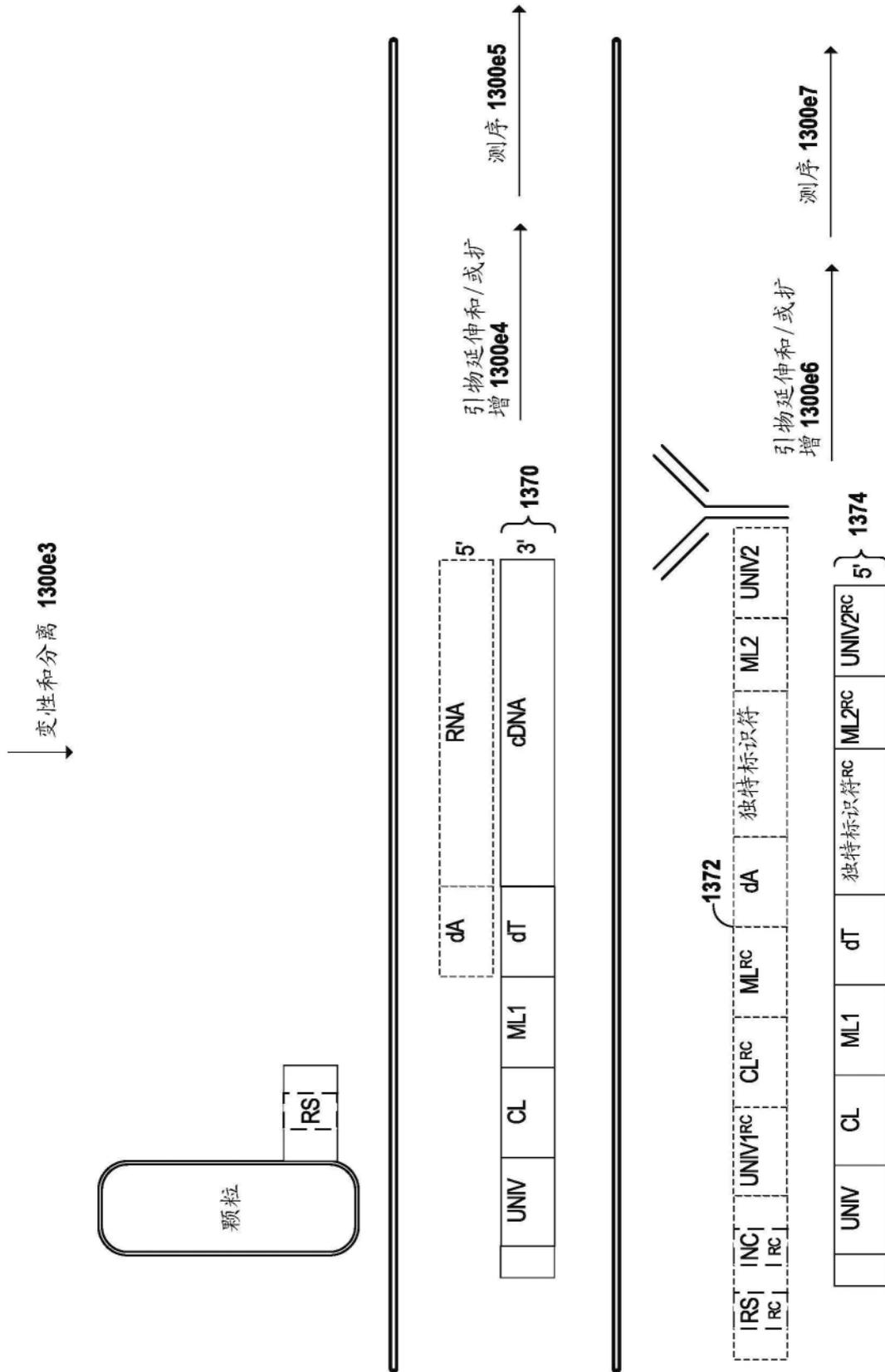


图131

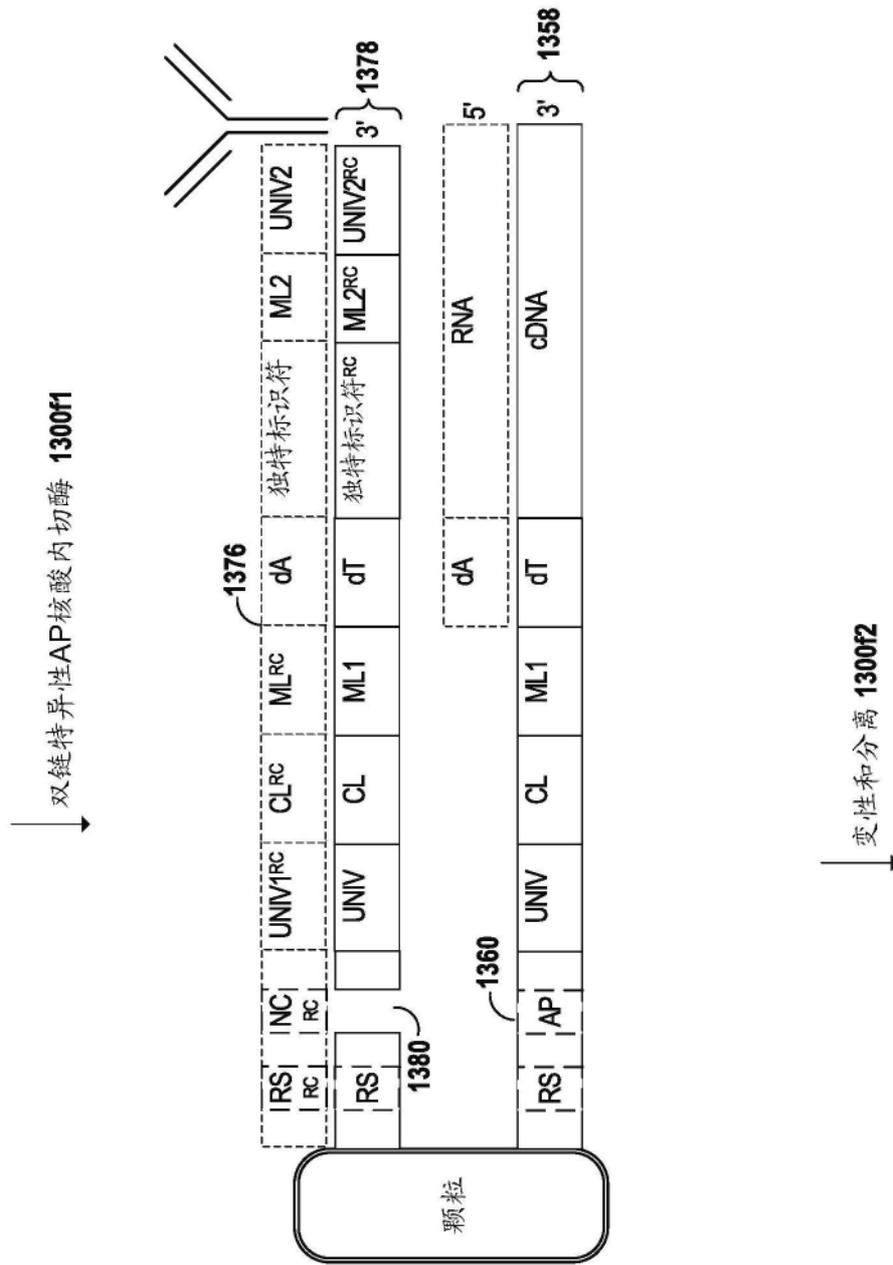


图13J

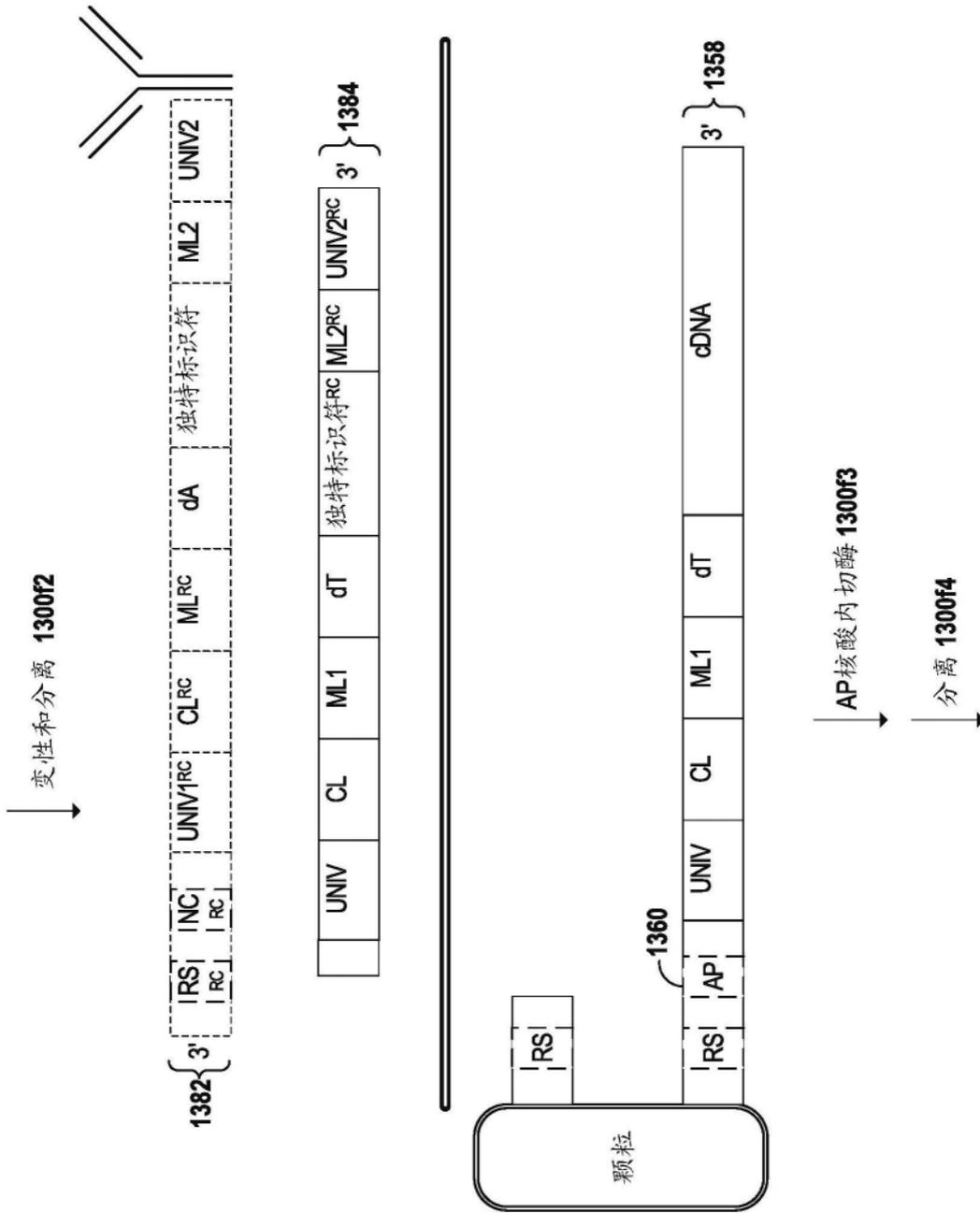


图13K

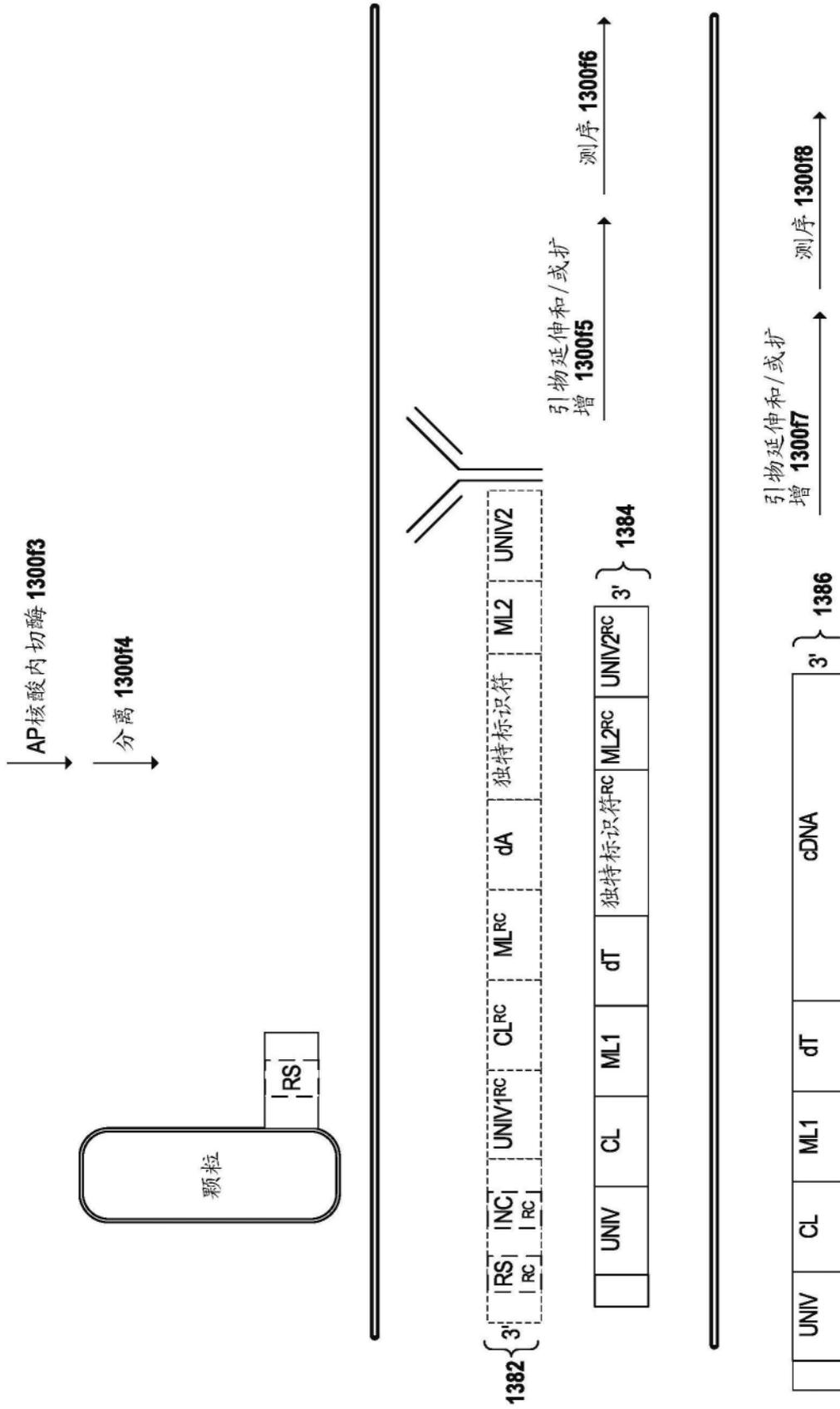


图13L

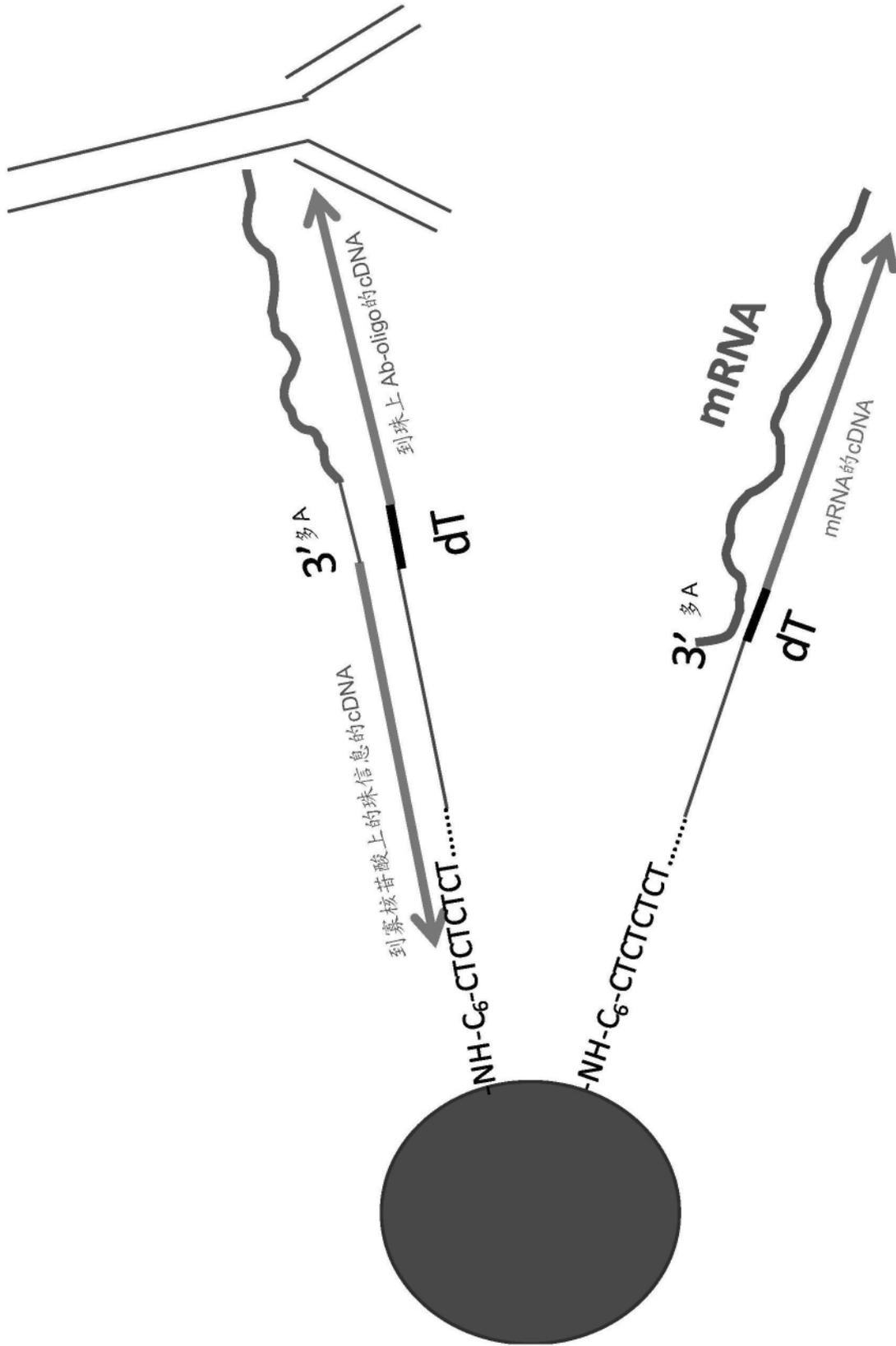


图14A

