



(12) 发明专利申请

(10) 申请公布号 CN 116157534 A

(43) 申请公布日 2023.05.23

(21) 申请号 202180060954.X

(22) 申请日 2021.07.12

(30) 优先权数据

63/051,149 2020.07.13 US

(85) PCT国际申请进入国家阶段日

2023.01.12

(86) PCT国际申请的申请数据

PCT/US2021/041327 2021.07.12

(87) PCT国际申请的公布数据

WO2022/015667 EN 2022.01.20

(71) 申请人 贝克顿迪金森公司

地址 美国新泽西州

(72) 发明人 特雷西·坎贝尔 宋慧媛

(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)

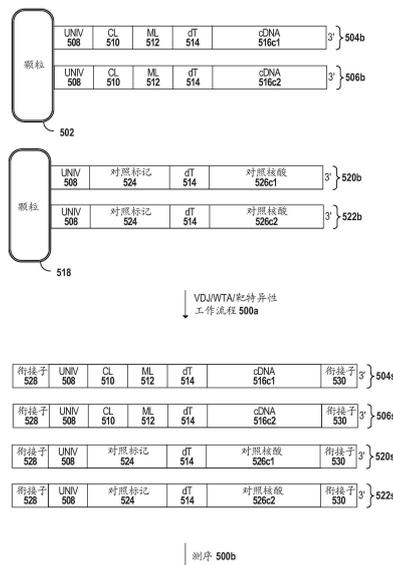
权利要求书11页 说明书49页 附图5页

(54) 发明名称

用于单细胞分析的cDNA加标对照

(57) 摘要

本文的公开内容包括用于确定单细胞mRNA测序测定工作流程失败的存在的系统、方法、组合物和试剂盒,包括确定对核酸靶的拷贝进行条形码化的失败和确定测序文库产生的失败。在一些实施方案中,还提供了用于确定测序文库成员的测序状态(例如,饱和测序或测序不足)的组合物、方法和系统。本文还提供了包含预定拷贝数的条形码化对照核酸的组合物。



1. 一种用于对样品中的核酸靶进行标记的方法,包括:

将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化以产生多于一个条形码化核酸分子,所述多于一个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;

提供多于一个一种或更多种条形码化对照核酸,其中所述一种或更多种条形码化对照核酸的每一种的拷贝数是预定的;

产生包含多于一个核酸靶文库成员和多于一个对照核酸文库成员的测序文库,其中产生测序文库包括:

将测序衔接子附接至所述多于一个条形码化核酸分子或其产物,以产生所述多于一个核酸靶文库成员;以及

将测序衔接子附接至所述多于一个一种或更多种条形码化对照核酸或其产物,以产生所述多于一个对照核酸文库成员;以及获得包含一个或多个核酸靶文库成员的多于一个测序读段和一个或多个对照核酸文库成员的多于一个测序读段的测序数据。

2. 根据权利要求1所述的方法,其中将核酸靶的拷贝用所述第一多于一个寡核苷酸条形码条形码化包括:

使所述核酸靶的拷贝与所述第一多于一个寡核苷酸条形码接触,其中所述第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、分子标记和能够与所述核酸靶杂交的靶结合区;以及

使与所述核酸靶的拷贝杂交的所述第一多于一个寡核苷酸条形码延伸以产生多于一个条形码化核酸分子,所述多于一个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。

3. 根据权利要求1-2中任一项所述的方法,其中所述多于一个条形码化核酸分子的每一个条形码化核酸分子包含第一通用序列和分子标记。

4. 根据权利要求1-3中任一项所述的方法,其中所述样品包括单细胞。

5. 根据权利要求1-4中任一项所述的方法,其中所述样品包括多于一个单细胞,所述方法包括:

在使所述核酸靶的拷贝与所述第一多于一个寡核苷酸条形码接触之前,将所述多于一个单细胞分区至多于一个分区,其中所述多于一个分区中的分区包括来自所述多于一个单细胞的单细胞;以及

在包含单细胞的分区中,使所述核酸靶的拷贝与所述第一多于一个寡核苷酸条形码接触。

6. 根据权利要求5所述的方法,其中所述分区是孔或液滴。

7. 根据权利要求1-6中任一项所述的方法,其中所述第一多于一个寡核苷酸条形码与第一固体支持物关联,所述方法包括将所述第一固体支持物与所述样品中的所述单细胞关联,并且其中所述多于一个分区中的分区包括单个第一固体支持物。

8. 根据权利要求7所述的方法,包括在所述分区步骤之后且在所述接触步骤之前裂解所述单细胞,任选地裂解所述单细胞包括加热所述样品、使所述样品与去污剂接触、改变所述样品的pH或其任何组合。

9. 根据权利要求1-8中任一项所述的方法,其中所述多于一个一种或更多种条形码化

对照核酸通过以下方式产生：

使预定拷贝数的一种或更多种对照核酸与第二多于一个寡核苷酸条形码接触，其中所述第二多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、对照标记和能够与所述一种或更多种对照核酸杂交的靶结合区；以及

使与所述一种或更多种对照核酸杂交的第二多于一个标记的寡核苷酸延伸，以产生预定拷贝数的一种或更多种条形码化对照核酸，所述一种或更多种条形码化对照核酸各自包含与所述一种或更多种条形码化对照核酸的至少一部分互补的序列。

10. 根据权利要求1-9中任一项所述的方法，其中所述第二多于一个寡核苷酸条形码与第二固体支持物关联和/或所述多于一个一种或更多种条形码化对照核酸与第二固体支持物关联。

11. 根据权利要求2-10中任一项所述的方法，其中使所述第一多于一个寡核苷酸条形码和/或所述第二多于一个寡核苷酸条形码延伸包括使用逆转录酶和/或缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶使所述第一多于一个寡核苷酸条形码和/或所述第二多于一个寡核苷酸条形码延伸，任选地所述DNA聚合酶包括Klenow片段。

12. 根据权利要求11所述的方法，其中所述逆转录酶包括病毒逆转录酶，任选地所述病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶。

13. 根据权利要求1-12中任一项所述的方法，其中所述条形码化对照核酸的每一种包含第一通用序列、对照标记和靶结合区中的一种或更多种。

14. 根据权利要求1-13中任一项所述的方法，其中所述第二多于一个寡核苷酸条形码的每一种对照标记包括至少1个、2个、3个、4个、5个、6个、7个、8个、9个或10个核苷酸。

15. 根据权利要求1-14中任一项所述的方法，其中所述一种或更多种条形码化对照核酸：

(i) 包括两种或更多种不同条形码化对照核酸；

(ii) 与至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的核酸靶至少约70%同源；和/或

(iii) 包括至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的条形码化对照核酸。

16. 根据权利要求1-15中任一项所述的方法，其中一种或更多种条形码化对照核酸：

(i) 与核酸靶至少约70%同源；

(ii) 包含持家基因的序列；

(iii) 与所述样品的基因组序列同源；

(iv) 与所述样品的基因组序列不同源；和/或

(v) 与物种的基因组序列同源，任选地所述物种是非哺乳动物物种，还任选地所述非哺乳动物物种是噬菌体物种，任选地所述噬菌体物种是T7噬菌体、PhiX噬菌体或其任何组合。

17. 根据权利要求1-16中任一项所述的方法，其中产生测序文库包括：

使随机引物与所述多于一个条形码化核酸分子和所述多于一个一种或更多种条形码化对照核酸接触，其中所述随机引物的每一个包含第二通用序列或其互补物；

使与所述多于一个条形码化核酸分子杂交的所述随机引物延伸以产生第一多于一个

延伸产物;以及

使与所述多于一个一种或更多种条形码化对照核酸分子杂交的所述随机引物延伸以产生第二多于一个延伸产物。

18. 根据权利要求17所述的方法,所述方法包括:

使用能够与所述第一通用序列或其互补物杂交的引物和能够与所述第二通用序列或其互补物杂交的引物扩增所述第一多于一个延伸产物,从而产生第一多于一个条形码化扩增子;以及

使用能够与所述第一通用序列或其互补物杂交的引物和能够与所述第二通用序列或其互补物杂交的引物扩增所述第二多于一个延伸产物,从而产生第二多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第一多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第二多于一个条形码化扩增子或其产物。

19. 根据权利要求18所述的方法,其中扩增所述第一多于一个延伸产物和所述第二多于一个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至所述第一多于一个延伸产物和所述第二多于一个延伸产物。

20. 根据权利要求18-19中任一项所述的方法,包括基于与所述第一多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目,确定所述样品中所述核酸靶的拷贝数。

21. 根据权利要求18-20中任一项所述的方法,其中产生测序文库包括:

使用能够与所述第一通用序列或其互补物杂交的引物和能够与所述第二通用序列或其互补物杂交的引物扩增所述第一多于一个条形码化扩增子,从而产生第三多于一个条形码化扩增子;以及

使用能够与所述第一通用序列或其互补物杂交的引物和能够与所述第二通用序列或其互补物杂交的引物扩增所述第二多于一个条形码化扩增子,从而产生第四多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第三多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第四多于一个条形码化扩增子或其产物。

22. 根据权利要求21所述的方法,其中扩增所述第一多于一个条形码化扩增子和所述第二多于一个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至所述第一多于一个条形码化扩增子和所述第二多于一个条形码化扩增子。

23. 根据权利要求21-22中任一项所述的方法,包括基于与所述第三多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目,确定所述样品中所述核酸靶的拷贝数。

24. 根据权利要求18-23中任一项所述的方法,其中所述第一多于一个条形码化扩增子和/或所述第三多于一个条形码化扩增子包含全转录组扩增(WTA)产物。

25. 根据权利要求1-24中任一项所述的方法,其中产生测序文库包括:

使用所述多于一个条形码化核酸分子作为模板合成第五多于一个条形码化扩增子以产生第五多于一个条形码化扩增子;以及

使用所述多于一个一种或更多种条形码化对照核酸作为模板合成第六多于一个条形码化扩增子,以产生第六多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第五多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第六多于一个条形码化扩增子或其产物。

26. 根据权利要求25所述的方法,其中合成所述第五多于一个条形码化扩增子和所述第六多于一个条形码化扩增子包括使用能够与所述第一通用序列或其互补物杂交的引物和靶特异性引物进行的PCR扩增。

27. 根据权利要求25-26中任一项所述的方法,其中合成所述第五多于一个条形码化扩增子和所述第六多于一个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至条形码化核酸分子和条形码化对照核酸。

28. 根据权利要求25-27中任一项所述的方法,包括基于与所述第五多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目,确定所述样品中所述核酸靶的拷贝数。

29. 根据权利要求1-28中任一项所述的方法,其中将核酸靶的拷贝用所述第一多于一个寡核苷酸条形码条形码化包括:

使多于一个核酸靶的拷贝与所述第一多于一个寡核苷酸条形码接触,其中所述第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、分子标记和能够与所述核酸靶的拷贝杂交的靶结合区;以及

产生各自包含所述靶结合区和所述靶结合区的互补物的多于一个条形码化核酸分子。

30. 根据权利要求29所述的方法,其中产生各自包含所述靶结合区和所述靶结合区的互补物的多于一个条形码化核酸分子包括:

在存在逆转录酶和包含所述靶结合区或其一部分的模板转换寡核苷酸的情况下,使与所述核酸靶的拷贝杂交的所述第一多于一个寡核苷酸条形码延伸,以产生多于一个条形码化核酸分子,所述多于一个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、第一分子标记、所述靶结合区和所述靶结合区的互补物。

31. 根据权利要求30所述的方法,其中所述逆转录酶具有末端转移酶活性,任选地所述逆转录酶是病毒逆转录酶,任选地鼠白血病病毒 (MLV) 逆转录酶或Moloney鼠白血病病毒 (MMLV) 逆转录酶。

32. 根据权利要求30-31中任一项所述的方法,其中所述模板转换寡核苷酸包含一个或更多个3' 核糖核苷酸,并且任选地三个3' 核糖核苷酸,还任选地所述3' 核糖核苷酸包含鸟嘌呤。

33. 根据权利要求29-32中任一项所述的方法,其中产生测序文库包括:

使每一个条形码化核酸分子的所述靶结合区的互补物与以下中的一种或更多种的靶结合区杂交:

(i) 所述第一多于一个寡核苷酸条形码中的寡核苷酸条形码,

(ii) 所述条形码化核酸分子自身,以及

(iii)所述多于一个条形码化核酸分子中的不同条形码化核酸分子;

使所述多于一个条形码化核酸分子的3'末端延伸以产生各自包含所述第一分子标记和第二分子标记的多于一个延伸的条形码化核酸分子。

34.根据权利要求33所述的方法,其中使条形码化核酸分子的靶结合区的互补物与所述第一多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交包括使条形码化核酸分子的靶结合区的互补物与所述第一多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区进行分子间杂交;并且还包含:

使与所述条形码化核酸分子的靶结合区的互补物杂交的寡核苷酸条形码的3'末端延伸,以产生多于一个延伸的条形码化核酸分子,所述多于一个延伸的条形码化核酸分子各自包含以下的互补物:第一分子标记,和第二分子标记,其中所述第二分子标记的序列不同于所述第一分子标记的序列,其中所述第二分子标记不是所述第一分子标记的互补物。

35.根据权利要求1-34中任一项所述的方法,其中所述多于一个一种或更多种条形码化对照核酸各自包含5'第一通用序列和3'所述第一通用序列的互补物。

36.根据权利要求1-35中任一项所述的方法,其中所述多于一个一种或更多种条形码化对照核酸包含5'对照标记和3'所述对照标记的互补物。

37.根据权利要求35-36中任一项所述的方法,所述方法包括:

使用能够与所述第一通用序列和/或其互补物杂交的引物扩增所述多于一个延伸的条形码化核酸分子,从而产生第七多于一个条形码化扩增子;以及

使用能够与所述第一通用序列和/或其互补物杂交的引物扩增所述多于一个一种或更多种条形码化对照核酸,从而产生第八多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第七多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第八多于一个条形码化扩增子或其产物。

38.根据权利要求37所述的方法,其中扩增所述多于一个延伸的条形码化核酸分子和扩增所述多于一个一种或更多种条形码化对照核酸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至所述多于一个延伸的条形码化核酸分子和所述多于一个一种或更多种条形码化对照核酸。

39.根据权利要求37-38中任一项所述的方法,包括基于与所述第七多于一个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、具有不同序列的第二分子标记或其组合的数目来确定所述样品中所述多于一个核酸靶的每一个的拷贝数。

40.根据权利要求35-39中任一项所述的方法,包括:

使用能够与所述核酸靶的序列杂交的靶特异性引物和能够与所述第一通用序列或其互补物杂交的引物扩增所述多于一个延伸的条形码化核酸分子,从而产生第九多于一个条形码化扩增子;以及

使用能够与所述核酸靶的序列杂交的靶特异性引物和能够与所述第一通用序列或其互补物杂交的引物扩增所述多于一个一种或更多种条形码化对照核酸,从而产生第十多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第九多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第十多于一个条形码化扩增子或其产

物。

41. 根据权利要求40所述的方法,其中扩增所述多于一个延伸的条形码化核酸分子和扩增所述多于一个一种或更多种条形码化对照核酸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至所述多于一个延伸的条形码化核酸分子和所述多于一个一种或更多种条形码化对照核酸。

42. 根据权利要求40-41中任一项所述的方法,包括基于与所述第九多于一个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、具有不同序列的第二分子标记或其组合的数目来确定所述样品中所述多于一个核酸靶的每一个的拷贝数。

43. 根据权利要求1-42中任一项所述的方法,其中所述多于一个一种或更多种条形码化对照核酸各自包含靶结合区和所述靶结合区的互补物,并且其中产生测序文库包括:

使每一个条形码化对照核酸的所述靶结合区的互补物与以下中的一种或更多种的靶结合区杂交:

(i) 所述第二多于一个寡核苷酸条形码中的寡核苷酸条形码,

(ii) 所述条形码化对照核酸本身,以及

(iii) 所述多于一个一种或更多种条形码化对照核酸的不同条形码化对照核酸;以及

使所述多于一个条形码化对照核酸的3'末端延伸以产生多于一个延伸的条形码化对照核酸,所述多于一个延伸的条形码化对照核酸各自包含第一通用序列、所述第一通用序列的互补物、对照标记和所述对照标记的互补物。

44. 根据权利要求43所述的方法,其中使条形码化对照核酸的靶结合区的互补物与所述第二多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交包括使条形码化对照核酸的靶结合区的互补物与所述第二多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区进行分子间杂交;并且还包括:

使与所述条形码化对照核酸的靶结合区的互补物杂交的寡核苷酸条形码的3'末端延伸,以产生多于一个延伸的条形码化对照核酸。

45. 根据权利要求43-44中任一项所述的方法,包括:

使用能够与所述第一通用序列和/或其互补物杂交的引物扩增所述多于一个延伸的条形码化核酸分子,从而产生第十一多于一个条形码化扩增子;以及

使用能够与所述第一通用序列和/或其互补物杂交的引物扩增所述多于一个延伸的条形码化对照核酸,从而产生第十二多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第十一多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第十二多于一个条形码化扩增子或其产物。

46. 根据权利要求45所述的方法,其中扩增所述多于一个延伸的条形码化核酸分子和扩增所述多于一个延伸的条形码化对照核酸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至所述多于一个延伸的条形码化核酸分子和所述多于一个延伸的条形码化对照核酸。

47. 根据权利要求45-46中任一项所述的方法,包括基于与所述第十一多于一个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、具有不同序列的第二分子标记或其组合的数目来确定所述样品中所述多于一个核酸靶的每一个的拷贝数。

48. 根据权利要求43-44中任一项所述的方法,所述方法包括:

使用能够与所述核酸靶的序列杂交的靶特异性引物和能够与所述第一通用序列或其互补物杂交的引物扩增所述多于一个延伸的条形码化核酸分子,从而产生第十三多于一个条形码化扩增子;以及

使用能够与所述核酸靶的序列杂交的靶特异性引物和能够与所述第一通用序列或其互补物杂交的引物扩增所述多于一个延伸的条形码化对照核酸分子,从而产生第十四多于一个条形码化扩增子,

其中所述多于一个核酸靶文库成员包括所述第十三多于一个条形码化扩增子或其产物,并且其中所述多于一个对照核酸文库成员包括所述第十四多于一个条形码化扩增子或其产物。

49. 根据权利要求48所述的方法,其中扩增所述多于一个延伸的条形码化核酸分子和扩增所述多于一个延伸的条形码化对照核酸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至所述多于一个延伸的条形码化核酸分子和所述多于一个延伸的条形码化对照核酸。

50. 根据权利要求48-49中任一项所述的方法,包括基于与所述第十三多于一个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、具有不同序列的第二分子标记或其组合的数目来确定所述样品中所述多于一个核酸靶的每一个的拷贝数。

51. 根据权利要求26-50中任一项所述的方法,其中所述靶特异性引物与免疫受体特异性杂交,任选地所述靶特异性引物与免疫受体的恒定区、免疫受体的可变区、免疫受体的多样性区、免疫受体的可变区和多样性区的连接处或其组合特异性杂交,任选地所述免疫受体是T细胞受体(TCR)和/或B细胞受体(BCR);任选地所述TCR包括TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或其任何组合;并且还任选地所述BCR受体包括BCR重链和/或BCR轻链。

52. 根据权利要求1-51中任一项所述的方法,其中所述多于一个条形码化核酸分子或其产物的多于一个测序读段的每一个包含(1)分子标记序列,和/或(2)所述核酸靶的子序列。

53. 根据权利要求1-52中任一项所述的方法,其中所述多于一个条形码化对照核酸分子或其产物的多于一个测序读段的每一个包含(1)对照标记序列,和/或(2)所述对照核酸分子子序列。

54. 根据权利要求1-53中任一项所述的方法,还包括确定所述测序数据中一个或更多个对照核酸文库成员的测序状态。

55. 根据权利要求54所述的方法,其中所述测序数据中一个或更多个对照核酸文库成员的测序状态是饱和测序或测序不足。

56. 根据权利要求55所述的方法,其中:

所述饱和测序状态由一个或更多个对照核酸文库成员具有处于或大于预定饱和阈值的测序读段数目确定;以及

所述测序不足状态由一个或更多个对照核酸文库成员具有少于预定饱和阈值的测序读段数目确定。

57. 根据权利要求56所述的方法,其中,预定饱和阈值是所述一种或更多种条形码化对照核酸的预定拷贝数的至少约1.1倍大的数字。

58. 根据权利要求56-57中任一项所述的方法,其中,预定饱和阈值是所述一种或更多种条形码化对照核酸的预定拷贝数的至少约4倍大的数字。

59. 根据权利要求56-58中任一项所述的方法,其中,如果所述测序数据中一个或更多个对照核酸文库成员的测序状态是测序不足状态,则重复获得测序数据的步骤,直到所述测序数据中一个或更多个对照核酸文库成员的测序状态是饱和测序状态。

60. 根据权利要求1-59中任一项所述的方法,还包括确定工作流程失败的存在,其中所述工作流程失败包括对所述核酸靶的拷贝进行条形码化的失败和/或测序文库产生的失败。

61. 根据权利要求60所述的方法,其中对所述核酸靶的拷贝进行条形码化的失败的存在通过所述一个或更多个对照核酸文库成员的测序读段与所述一个或更多个核酸靶文库成员的测序读段的比率超过预定条形码化阈值来确定。

62. 根据权利要求1-61中任一项所述的方法,还包括基于一个或更多个核酸靶文库成员的多于一个测序读段确定所述样品中核酸靶的拷贝数。

63. 根据权利要求62所述的方法,其中确定所述样品中所述核酸靶的拷贝数包括基于与所述一个或更多个核酸靶文库成员或其产物关联的具有不同序列的第一分子标记、其互补物或其组合的数目来确定所述样品中所述核酸靶的拷贝数。

64. 根据权利要求62-63中任一项所述的方法,其中对所述核酸靶的拷贝进行条形码化的失败的存在通过所述一个或更多个条形码化对照核酸的预定拷贝数与所述样品中所述核酸靶的拷贝数的比率超过预定条形码化阈值来确定。

65. 根据权利要求61-64中任一项所述的方法,其中所述预定条形码化阈值是至少约1、2、3、4、5、6、7、8、9或10。

66. 根据权利要求1-65中任一项所述的方法,还包括获得测序数据,所述测序数据包括预定数目的一个或更多个加标文库成员的多于一个测序读段。

67. 根据权利要求66所述的方法,其中测序文库产生的失败的存在通过预定数目的所述一个或更多个加标文库成员的测序读段与所述一个或更多个对照核酸文库成员的测序读段的比率超过预定文库产生阈值来确定,任选地,所述预定文库产生阈值是至少约1、2、3、4、5、6、7、8、9或10。

68. 根据权利要求66-67中任一项所述的方法,其中所述一个或更多个加标文库成员与所述样品的基因组序列不同源,任选地所述一个或更多个加标文库成员与物种的基因组序列同源,任选地所述物种是非哺乳动物物种,还任选地所述非哺乳动物物种是噬菌体物种,任选地所述噬菌体物种是T7噬菌体、PhiX噬菌体或其任何组合。

69. 根据权利要求2-68中任一项所述的方法,其中所述靶结合区包含多(dT)区、随机序列、靶特异性序列或其组合。

70. 根据权利要求2-69中任一项所述的方法,其中所述第一通用序列和/或所述第二通用序列包含测序引物的结合位点和/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。

71. 根据权利要求1-70中任一项所述的方法,其中所述核酸靶包括核酸分子,任选地所述核酸分子包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、

包含多(A)尾的RNA或其任何组合,并且还任选地所述mRNA编码免疫受体。

72. 根据权利要求71所述的方法,其中所述核酸靶包含细胞组分结合试剂,和/或所述核酸分子与所述细胞组分结合试剂关联,任选地所述方法包括使所述核酸分子和所述细胞组分结合试剂解离。

73. 根据权利要求1-72中任一项所述的方法,其中所述第一多于一个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。

74. 根据权利要求1-73中任一项所述的方法,其中所述第一多于一个寡核苷酸条形码各自包含细胞标记,任选地所述第一多于一个寡核苷酸条形码的每一种细胞标记包含至少6个核苷酸,还任选地与同一第一固体支持物关联的所述第一多于一个寡核苷酸条形码中的寡核苷酸条形码包含相同的细胞标记,还任选地与不同第一固体支持物关联的所述第一多于一个寡核苷酸条形码中的寡核苷酸条形码包含不同的细胞标记。

75. 根据权利要求7-74中任一项所述的方法,其中:(i)所述第一固体支持物包括第一合成颗粒、第一平坦表面或其组合;和/或(ii)所述第二固体支持物包括第二合成颗粒、第二平坦表面或其组合。

76. 根据权利要求75所述的方法,其中:

所述第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被固定或部分地固定在所述第一合成颗粒上,或者所述第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被包封或部分地包封在所述第一合成颗粒中;和/或

所述多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被固定或部分地固定在所述第二合成颗粒上,或者所述多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被包封或部分地包封在所述第二合成颗粒中。

77. 根据权利要求75-76中任一项所述的方法,其中所述第一合成颗粒和/或第二合成颗粒是可破坏的,任选地是可破坏的水凝胶颗粒。

78. 根据权利要求75-77中任一项所述的方法,其中所述第一合成颗粒和/或第二合成颗粒包括珠,任选地所述珠是琼脂糖凝胶珠、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。

79. 根据权利要求75-78中任一项所述的方法,其中所述第一合成颗粒和/或第二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。

80. 根据权利要求75-79中任一项所述的方法,

其中所述第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含接头官能团,

其中所述第一合成颗粒包含固体支持物官能团,并且

其中所述支持物官能团和所述接头官能团彼此关联,并且任选地所述接头官能团和所述支持物官能团单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

81. 根据权利要求1-80中任一项所述的方法,

其中所述多于一个一种或更多种条形码化对照核酸的每一种条形码化对照核酸包含

接头官能团，

其中所述第二合成颗粒包含固体支持物官能团，并且

其中所述支持物官能团和所述接头官能团彼此关联，并且任选地所述接头官能团和所述支持物官能团单独地选自由以下组成的组：C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

82. 一种试剂盒，包含：

多于一个一种或更多种条形码化对照核酸，其中所述一种或更多种条形码化对照核酸的每一种的拷贝数是预定的，其中所述多于一个一种或更多种条形码化对照核酸与第二固体支持物关联。

83. 根据权利要求82所述的试剂盒，包含第一多于一个寡核苷酸条形码，其中所述多于一个寡核苷酸条形码的每一个包含分子标记和靶结合区，并且其中所述多于一个寡核苷酸条形码中的至少10种包含不同的分子标记序列。

84. 根据权利要求83所述的试剂盒，其中所述第一多于一个寡核苷酸条形码与第一固体支持物关联。

85. 根据权利要求82-84中任一项所述的试剂盒，其中所述条形码化对照核酸的每一种包含第一通用序列、对照标记和靶结合区中的一种或更多种。

86. 根据权利要求82-85中任一项所述的试剂盒，其中所述一种或更多种条形码化对照核酸(i) 包含至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的条形码化对照核酸；和/或(ii) 与至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的核酸靶至少约70%同源。

87. 根据权利要求82-86中任一项所述的试剂盒，其中一种或更多种条形码化对照核酸包含持家基因的序列。

88. 根据权利要求82-87中任一项所述的试剂盒，其中一种或更多种条形码化对照核酸与物种的基因组序列同源，任选地所述物种是非哺乳动物物种，还任选地所述非哺乳动物物种是噬菌体物种，任选地所述噬菌体物种是T7噬菌体、PhiX噬菌体或其任何组合。

89. 根据权利要求83-88中任一项所述的试剂盒，其中所述靶结合区包含基因特异性序列、寡(dT)序列、随机多聚体或其任何组合。

90. 根据权利要求82-89中任一项所述的试剂盒，还包含：

(i) 逆转录酶，任选地所述逆转录酶包括病毒逆转录酶，并且还任选地所述病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶；

(ii) 包含所述靶结合区或其一部分的模板转换寡核苷酸，任选地所述模板转换寡核苷酸包含一个或更多个3'核糖核苷酸，任选地三个3'核糖核苷酸，还任选地所述3'核糖核苷酸包含鸟嘌呤；

(iii) 以下中的一种或更多种：乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砜(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或其任何组合；

(iv) 缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶，任选地其中所述DNA聚合酶包括Klenow片段；

(v) 缓冲液、筒或两者；和/或

(vi) 用于逆转录反应和/或扩增反应的一种或更多种试剂。

91. 根据权利要求83-90中任一项所述的试剂盒,其中所述第一多于一个寡核苷酸条形码各自包含细胞标记,任选地,所述第一多于一个寡核苷酸条形码的每一种细胞标记包含至少6个核苷酸。

92. 根据权利要求91所述的试剂盒,其中与同一第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码包含相同的细胞标记,并且其中与不同的第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码包含不同的细胞标记。

93. 根据权利要求84-92中任一项所述的试剂盒,其中:

所述第一固体支持物包括第一合成颗粒、第一平坦表面或其组合;和/或

所述第二固体支持物包括第二合成颗粒、第二平坦表面或其组合。

94. 根据权利要求93所述的试剂盒,其中:

所述第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被固定或部分地固定在所述第一合成颗粒上,或者所述第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被包封或部分地包封在所述第一合成颗粒中;和/或

所述多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被固定或部分地固定在所述第二合成颗粒上,或者所述多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被包封或部分地包封在所述第二合成颗粒中。

95. 根据权利要求93-94中任一项所述的试剂盒,其中所述第一合成颗粒和/或第二合成颗粒是可破坏的,任选地是可破坏的水凝胶颗粒。

96. 根据权利要求93-95中任一项所述的试剂盒,其中所述第一合成颗粒和/或第二合成颗粒包括珠,任选地所述珠是琼脂糖凝胶珠、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。

97. 根据权利要求93-96中任一项所述的试剂盒,其中所述第一合成颗粒和/或第二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。

98. 根据权利要求93-97中任一项所述的试剂盒,

其中所述第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含接头官能团,

其中所述第一合成颗粒包含固体支持物官能团,并且

其中所述支持物官能团和所述接头官能团彼此关联,并且任选地所述接头官能团和所述支持物官能团单独地选自以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及其任何组合。

99. 根据权利要求93-98中任一项所述的试剂盒,

其中所述多于一个一种或更多种条形码化对照核酸的每一种条形码化对照核酸包含接头官能团,

其中所述第二合成颗粒包含固体支持物官能团,并且

其中所述支持物官能团和所述接头官能团彼此关联,并且任选地所述接头官能团和所述支持物官能团单独地选自以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

用于单细胞分析的cDNA加标对照

[0001] 相关申请

[0002] 本申请根据35 U.S.C. §119(e) 要求2020年7月13日提交的美国临时专利申请系列第63/051,149号的权益,该相关申请的内容出于所有目的通过引用以其整体并入本文。

[0003] 背景

[0004] 领域

[0005] 本公开内容总体上涉及分子生物学领域,例如使用分子条形码化(molecular barcoding)确定基因表达。

[0006] 对相关技术的描述

[0007] 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。对于用于确定测定工作流程失败的存在的组合物、方法和系统(包括确定对核酸靶的拷贝进行条形码化的失败和确定测序文库产生的失败)存在需求。另外地,对于用于确定测序文库成员的测序状态(例如,饱和测序或测序不足(under sequencing))的组合物、方法和系统存在需求。

[0008] 概述

[0009] 本文的公开内容包括用于标记样品中的核酸靶的方法。在一些实施方案中,该方法包括:将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化,以产生多于一个条形码化核酸分子,所述多于一个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;提供多于一个一种或更多种条形码化对照核酸,其中一种或更多种条形码化对照核酸的每一种的拷贝数是预定的;产生包含多于一个核酸靶文库成员和多于一个对照核酸文库成员的测序文库,其中产生测序文库包括:将测序衔接子附接至多于一个条形码化核酸分子或其产物,以产生多于一个核酸靶文库成员;以及将测序衔接子附接至多于一个一种或更多种条形码化对照核酸或其产物,以产生多于一个对照核酸文库成员;以及获得包含一个或更多个核酸靶文库成员的多于一个测序读段和一个或更多个对照核酸文库成员的多于一个测序读段的测序数据。

[0010] 在一些实施方案中,将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化包括:使核酸靶的拷贝与第一多于一个寡核苷酸条形码接触,其中第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶杂交的靶结合区;以及使与核酸靶的拷贝杂交的第一多于一个寡核苷酸条形码延伸,以产生多于一个条形码化核酸分子,所述多于一个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。在一些实施方案中,多于一个条形码化核酸分子的每一个条形码化核酸分子包含第一通用序列和分子标记。在一些实施方案中,样品包括单细胞。在一些实施方案中,样品包括多于一个单细胞。该方法可以包括:在使核酸靶的拷贝与第一多于一个寡核苷酸条形码接触之前,将多于一个单细胞分区至多于一个分区,其中多于一个分区中的分区包括来自多于一个单细胞的单细胞;以及在包含单细胞的分区中,使核酸靶的拷贝与第一多于一个寡核苷酸条形码接触。在一些实施方案中,分区是孔或液滴。在一些实施方案中,第一多于

一个寡核苷酸条形码与第一固体支持物关联。该方法可以包括：将第一固体支持物与样品中的单细胞关联，并且其中多于一个分区中的分区包括单个第一固体支持物。该方法可以包括：在分区步骤之后且在接触步骤之前裂解单细胞。在一些实施方案中，裂解单细胞包括加热样品、使样品与去污剂接触、改变样品的pH或其任何组合。

[0011] 在一些实施方案中，多于一个一种或更多种条形码化对照核酸通过以下方式产生：使预定拷贝数的一种或更多种对照核酸与第二多于一个寡核苷酸条形码接触，其中第二多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、对照标记和能够与一种或更多种对照核酸杂交的靶结合区；以及使与一种或更多种对照核酸杂交的第二多于一个标记的寡核苷酸延伸，以产生预定拷贝数的一种或更多种条形码化对照核酸，所述一种或更多种条形码化对照核酸各自包含与一种或更多种条形码化对照核酸的至少一部分互补的序列。在一些实施方案中，第二多于一个寡核苷酸条形码与第二固体支持物关联和/或多于一个一种或更多种条形码化对照核酸与第二固体支持物关联。在一些实施方案中，使第一多于一个寡核苷酸条形码和/或第二多于一个寡核苷酸条形码延伸包括使用逆转录酶和/或缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶（例如，Klenow片段）使第一多于一个寡核苷酸条形码和/或第二多于一个寡核苷酸条形码延伸。逆转录酶可以包括病毒逆转录酶（例如，鼠白血病病毒（MLV）逆转录酶或Moloney鼠白血病病毒（MMLV）逆转录酶）。

[0012] 在一些实施方案中，条形码化对照核酸的每一种包含第一通用序列、对照标记和靶结合区中的一种或更多种。在一些实施方案中，第二多于一个寡核苷酸条形码的每一种对照标记包括至少1个、2个、3个、4个、5个、6个、7个、8个、9个或10个核苷酸。在一些实施方案中，一种或更多种条形码化对照核酸包括两种或更多种不同的条形码化对照核酸。在一些实施方案中，一种或更多种条形码化对照核酸包括至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的条形码化对照核酸。在一些实施方案中，一种或更多种条形码化对照核酸与核酸靶至少约70%同源。在一些实施方案中，一种或更多种条形码化对照核酸与至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的核酸靶至少约70%同源。在一些实施方案中，一种或更多种条形码化对照核酸包含持家基因的序列。在一些实施方案中，一种或更多种条形码化对照核酸与样品的基因组序列同源。在一些实施方案中，一种或更多种条形码化对照核酸与样品的基因组序列不同源。在一些实施方案中，一种或更多种条形码化对照核酸与物种的基因组序列同源。物种可以是非哺乳动物物种，例如噬菌体物种（例如T7噬菌体、PhiX噬菌体或其任何组合）。

[0013] 在一些实施方案中，产生测序文库包括：使随机引物与多于一个条形码化核酸分子和多于一个一种或更多种条形码化对照核酸接触，其中随机引物的每一个包含第二通用序列或其互补物；使与多于一个条形码化核酸分子杂交的随机引物延伸，以产生第一多于一个延伸产物；以及使与多于一个一种或更多种条形码化对照核酸杂交的随机引物延伸，以产生第二多于一个延伸产物。该方法可以包括：使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第一多于一个延伸产物，从而产生第一多于一个条形码化扩增子；以及使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第二多于一个延伸产物，从而产生第二

多于一个条形码化扩增子,其中多于一个核酸靶文库成员包括第一多于一个条形码化扩增子或其产物,并且其中多于一个对照核酸文库成员包括第二多于一个条形码化扩增子或其产物。在一些实施方案中,扩增第一多于一个延伸产物和第二多于一个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至第一多于一个延伸产物和第二多于一个延伸产物。方法可以包括:基于与第一多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目来确定样品中核酸靶的拷贝数。

[0014] 在一些实施方案中,产生测序文库包括:使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第一多于一个条形码化扩增子,从而产生第三多于一个条形码化扩增子;以及使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第二多于一个条形码化扩增子,从而产生第四多于一个条形码化扩增子,其中多于一个核酸靶文库成员包括第三多于一个条形码化扩增子或其产物,并且其中多于一个对照核酸文库成员包括第四多于一个条形码化扩增子或其产物。在一些实施方案中,扩增第一多于一个条形码化扩增子和第二多于一个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至第一多于一个条形码化扩增子和第二多于一个条形码化扩增子。该方法可以包括:基于与第三多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目,确定样品中核酸靶的拷贝数。在一些实施方案中,第一多于一个条形码化扩增子和/或第三多于一个条形码化扩增子包含全转录组扩增(WTA)产物。

[0015] 在一些实施方案中,产生测序文库包括:使用多于一个条形码化核酸分子作为模板合成第五多于一个条形码化扩增子,以产生第五多于一个条形码化扩增子;以及使用多于一个一种或更多种条形码化对照核酸作为模板合成第六多于一个条形码化扩增子,以产生第六多于一个条形码化扩增子,其中多于一个核酸靶文库成员包括第五多于一个条形码化扩增子或其产物,并且其中多于一个对照核酸文库成员包括第六多于一个条形码化扩增子或其产物。在一些实施方案中,合成第五多于一个条形码化扩增子和第六多于一个条形码化扩增子包括使用能够与第一通用序列或其互补物杂交的引物和靶特异性引物进行的PCR扩增。在一些实施方案中,合成第五多于一个条形码化扩增子和第六多于一个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至条形码化核酸分子和条形码化对照核酸。该方法可以包括:基于与第五多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目,确定样品中核酸靶的拷贝数。

[0016] 在一些实施方案中,将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化包括:使多于一个核酸靶的拷贝与第一多于一个寡核苷酸条形码接触,其中第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区;以及产生多于一个条形码化核酸分子,所述多于一个条形码化核酸分子各自包含靶结合区和靶结合区的互补物。在一些实施方案中,产生各自包含靶结合区和靶结合区的互补物的多于一个条形码化的核酸分子包括:在存在逆转录酶和包含靶结合区或其一部分的模板转换寡核苷酸的情况下,使与核酸靶的拷贝杂交的第一多于一个寡核苷酸条形码延伸,以产生多于一个条形码化核酸分子,所述多于一个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、第一分子标记、靶结合区和靶结合区的互补物。在

一些实施方案中,逆转录酶具有末端转移酶活性。在一些实施方案中,模板转换寡核苷酸包含一个或更多个3'核糖核苷酸(例如三个3'核糖核苷酸)。在一些实施方案中,3'核糖核苷酸包含鸟嘌呤。在一些实施方案中,逆转录酶是病毒逆转录酶(例如鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶)。

[0017] 在一些实施方案中,产生测序文库包括:使每一个条形码化核酸分子的靶结合区的互补物与以下中的一种或更多种的靶结合区杂交:(i)第一多于一个寡核苷酸条形码中的寡核苷酸条形码,(ii)条形码化核酸分子本身,以及(iii)多于一个条形码化核酸分子中的不同条形码化核酸分子;使多于一个条形码化核酸分子的3'末端延伸以产生多于一个延伸的条形码化核酸分子,所述多于一个延伸的条形码化核酸分子各自包含第一分子标记和第二分子标记。在一些实施方案中,使条形码化核酸分子的靶结合区的互补物与第一多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交包括使条形码化核酸分子的靶结合区的互补物与第一多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区进行分子间杂交。该方法可以包括:使与条形码化核酸分子的靶结合区的互补物杂交的寡核苷酸条形码的3'末端延伸,以产生多于一个延伸的条形码化核酸分子,所述多于一个延伸的条形码化核酸分子各自包含以下的互补物:第一分子标记,和第二分子标记,其中第二分子标记的序列不同于第一分子标记的序列,其中第二分子标记不是第一分子标记的互补物。在一些实施方案中,多于一个一种或更多种条形码化对照核酸各自包含5'第一通用序列和3'第一通用序列的互补物。在一些实施方案中,多于一个一种或更多种条形码化对照核酸包含5'对照标记和3'对照标记的互补物。

[0018] 该方法可包括:使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子,从而产生第七多于一个条形码化扩增子;以及使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个一种或更多种条形码化对照核酸,从而产生第八多于一个条形码化扩增子,其中多于一个核酸靶文库成员包括第七多于一个条形码化扩增子或其产物,并且其中多于一个对照核酸文库成员包括第八多于一个条形码化扩增子或其产物。在一些实施方案中,扩增多于一个延伸的条形码化核酸分子和扩增多于一个一种或更多种条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个一种或更多种条形码化对照核酸。该方法可以包括:基于与第七多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0019] 该方法可以包括:使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子,从而产生第九多于一个条形码化扩增子;以及使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个一种或更多种条形码化对照核酸,从而产生第十多于一个条形码化扩增子,其中多于一个核酸靶文库成员包括第九多于一个条形码化扩增子或其产物,并且其中多于一个对照核酸文库成员包括第十多于一个条形码化扩增子或其产物。在一些实施方案中,扩增多于一个延伸的条形码化核酸分子和扩增多于一个一种或更多种条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个一种或更多种条形码化

对照核酸。该方法可以包括：基于与第九多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0020] 在一些实施方案中，多于一个一种或更多种条形码化对照核酸各自包含靶结合区和靶结合区的互补物，并且其中产生测序文库包括：使每一种条形码化对照核酸的靶结合区的互补物与以下一种或更多种的靶结合区杂交：(i) 第二多于一个寡核苷酸条形码中的寡核苷酸条形码，(ii) 条形码化对照核酸本身，和(iii) 多于一个一种或更多种条形码化对照核酸中的不同条形码化对照核酸；以及使多于一个条形码化对照核酸的3'末端延伸以产生多于一个延伸的条形码化对照核酸，所述多于一个延伸的条形码化对照核酸各自包含第一通用序列、第一通用序列的互补物、对照标记和对照标记的互补物。在一些实施方案中，使条形码化对照核酸的靶结合区的互补物与第二多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交包括使条形码化对照核酸的靶结合区的互补物与第二多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区进行分子间杂交。该方法可以包括：使与条形码化对照核酸的靶结合区的互补物杂交的寡核苷酸条形码的3'末端延伸，以产生多于一个延伸的条形码化对照核酸。

[0021] 该方法可以包括：使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子，从而产生第十一多于一个条形码化扩增子；以及使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个延伸的条形码化对照核酸，从而产生第十二多于一个条形码化扩增子，其中多于一个核酸靶文库成员包括第十一多于一个条形码化扩增子或其产物，并且其中多于一个对照核酸文库成员包括第十二多于一个条形码化扩增子或其产物。在一些实施方案中，扩增多于一个延伸的条形码化核酸分子和扩增多于一个延伸的条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个延伸的条形码化对照核酸。该方法可以包括：基于与第十一多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0022] 该方法可包括：使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子，从而产生第十三多于一个条形码化扩增子；以及使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个延伸的条形码化对照核酸，从而产生第十四多于一个条形码化扩增子，其中多于一个核酸靶文库成员包括第十三多于一个条形码化扩增子或其产物，并且其中多于一个对照核酸文库成员包括第十四多于一个条形码化扩增子或其产物。在一些实施方案中，扩增多于一个延伸的条形码化核酸分子和扩增多于一个延伸的条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个延伸的条形码化对照核酸。该方法可以包括：基于与第十三多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0023] 在一些实施方案中，靶特异性引物与免疫受体特异性杂交。在一些实施方案中，靶

特异性引物与免疫受体的恒定区、免疫受体的可变区、免疫受体的多样性区、免疫受体的可变区和多样性区的连接区(junction)或它们的组合特异性杂交。免疫受体可以是T细胞受体(TCR)和/或B细胞受体(BCR)。TCR可以包含TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或其任何组合。BCR受体可以包含BCR重链和/或BCR轻链。在一些实施方案中,使多于一个条形码化核酸分子的3'末端延伸和/或使多于一个条形码化对照核酸的3'末端延伸使用缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶(例如,Klenow片段)进行。该方法可以包括:在存在乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砷(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或其任何组合中的一种或更多种的情况下,使第一多于一个寡核苷酸条形码和/或第二多于一个寡核苷酸条形码延伸。

[0024] 在一些实施方案中,多于一个条形码化核酸分子或其产物的多于一个测序读段的每一个包含(1)分子标记序列,和/或(2)核酸靶的子序列。在一些实施方案中,多于一个条形码化对照核酸分子或其产物的多于一个测序读段的每一个包含(1)对照标记序列,和/或(2)对照核酸分子的子序列。该方法可以包括:确定测序数据中一个或多个对照核酸文库成员的测序状态。在一些实施方案中,测序数据中一个或多个对照核酸文库成员的测序状态是饱和测序或测序不足。在一些实施方案中,饱和测序状态由一个或多个对照核酸文库成员具有处于或大于预定饱和阈值的测序读段数目确定;并且测序不足状态由一个或多个对照核酸文库成员具有小于预定饱和阈值的测序读段数目确定。在一些实施方案中,预定饱和阈值是一种或更多种条形码化对照核酸的预定拷贝数的至少约1.1倍大的数字。在一些实施方案中,预定饱和阈值是一种或更多种条形码化对照核酸的预定拷贝数的至少约4倍大的数字。

[0025] 该方法可以包括:如果测序数据中一个或多个对照核酸文库成员的测序状态是测序不足状态,则重复获得测序数据的步骤,直到测序数据中一个或多个对照核酸文库成员的测序状态是饱和测序状态。该方法可以包括:确定工作流程失败的存在,其中工作流程失败包括对核酸靶的拷贝进行条形码化的失败和/或测序文库产生的失败。在一些实施方案中,对核酸靶的拷贝进行条形码化的失败的存在通过一个或多个对照核酸文库成员的测序读段与一个或多个核酸靶文库成员的测序读段的比率超过预定条形码化阈值来确定。该方法可以包括:基于一个或多个核酸靶文库成员的多于一个测序读段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一个或多个核酸靶文库成员或其产物关联的具有不同序列的第一分子标记、其互补物或其组合的数目来确定样品中核酸靶的拷贝数。在一些实施方案中,对核酸靶的拷贝进行条形码化的失败的存在通过一个或多个条形码化对照核酸的预定拷贝数与样品中核酸靶的拷贝数的比率超过预定条形码化阈值来确定。在一些实施方案中,预定条形码化阈值是至少约1、2、3、4、5、6、7、8、9或10。该方法可以包括:获得包含预定数目的一个或多个加标文库成员的多于一个测序读段的测序数据。在一些实施方案中,测序文库产生的失败的存在通过一个或多个加标文库成员的预定数目的测序读段与一个或多个对照核酸文库成员的测序读段的比率超过预定文库产生阈值来确定。在一些实施方案中,预定文库产生阈值是至少约1、2、3、4、5、6、7、8、9或10。在一些实施方案中,一个或多个加标文库成员与样品的基因组序列不同源。在一些实施方案中,一个或多个加标文库成员与物种的基因组序列同源。在一些实施方案中,物种是非哺乳动物物种。在一些实施方案中,非哺乳动物物种是噬

菌体物种(例如,T7噬菌体、PhiX噬菌体或其任何组合)。

[0026] 在一些实施方案中,靶结合区包括多(dT)区、随机序列、靶特异性序列或其组合。在一些实施方案中,第一通用序列和/或第二通用序列包含测序引物的结合位点和/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,多于一个条形码化核酸分子包括条形码化脱氧核糖核酸(DNA)分子、条形码化核糖核酸(RNA)分子或它们的组合。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或其任何组合)。在一些实施方案中,mRNA编码免疫受体。在一些实施方案中,核酸靶包括细胞组分结合试剂,和/或核酸分子与细胞组分结合试剂关联。方法可以包括:使核酸分子和细胞组分结合试剂解离。

[0027] 在一些实施方案中,第一多于一个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,第一多于一个寡核苷酸条形码各自包含细胞标记。在一些实施方案中,第一多于一个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,与同一第一固体支持物关联的第一多于一个寡核苷酸条形码中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,与不同的第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码包含不同的细胞标记。在一些实施方案中,第一固体支持物包括第一合成颗粒、第一平坦表面或其组合。在一些实施方案中,第二固体支持物包括第二合成颗粒、第二平坦表面或其组合。

[0028] 在一些实施方案中,第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被固定或部分地固定在第一合成颗粒上,或者第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被包封或部分地包封在第一合成颗粒中。在一些实施方案中,多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被固定或部分地固定在第二合成颗粒上,或者多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被包封或部分地包封在第二合成颗粒中。在一些实施方案中,第一合成颗粒和/或第二合成颗粒是可破坏的(例如,可破坏的水凝胶颗粒)。在一些实施方案中,第一合成颗粒和/或第二合成颗粒包括珠(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)。在一些实施方案中,第一合成颗粒和/或第二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。在一些实施方案中,第一多于一个寡核苷酸条形码中的每一个寡核苷酸条形码包含接头官能团,第一合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由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、一种或更多种酮及其任何组合组成的组。在一些实施方案中,多于一个一种或更多种条形码化对照核酸中的每一种条形码化对照核酸包含接头官能团,第二合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由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、一种或更多种酮及其任何组合组成的组。

[0029] 本文的公开内容包括试剂盒。试剂盒可以包含：多于一个一种或更多种条形码化对照核酸，其中一种或更多种条形码化对照核酸的每一种的拷贝数是预定的。在一些实施方案中，多于一个一种或更多种条形码化对照核酸与第二固体支持物关联。试剂盒可以包含：第一多于一个寡核苷酸条形码，其中多于一个寡核苷酸条形码的每一个包含分子标记和靶结合区，并且其中多于一个寡核苷酸条形码中的至少10种包含不同的分子标记序列。

[0030] 在一些实施方案中，第一多于一个寡核苷酸条形码与第一固体支持物关联。在一些实施方案中，条形码化对照核酸的每一种包含第一通用序列、对照标记和靶结合区中的一种或更多种。在一些实施方案中，一种或更多种条形码化对照核酸包括至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的条形码化对照核酸。在一些实施方案中，一种或更多种条形码化对照核酸与至少约2种、3种、4种、6种、8种、10种、12种、15种、20种、25种、30种、35种或40种不同的核酸靶至少约70%同源。在一些实施方案中，一种或更多种条形码化对照核酸包含持家基因的序列。在一些实施方案中，一种或更多种条形码化对照核酸与物种的基因组序列同源。在一些实施方案中，物种是非哺乳动物物种。在一些实施方案中，非哺乳动物物种是噬菌体物种（例如，T7噬菌体、PhiX噬菌体或其任何组合）。在一些实施方案中，靶结合区包括基因特异性序列、寡(dT)序列、随机多聚体或其任何组合。

[0031] 试剂盒可以包含逆转录酶（例如，病毒逆转录酶，诸如例如，鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶)。该试剂盒可以包含：包含靶结合区或其一部分的模板转换寡核苷酸。在一些实施方案中，模板转换寡核苷酸包含一个或更多个3'核糖核苷酸（例如，三个3'核糖核苷酸）。在一些实施方案中，3'核糖核苷酸包含鸟嘌呤。试剂盒可以包含以下中的一种或更多种：乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砜(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或它们的任何组合。试剂盒可以包含：缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶（例如，Klenow片段）。试剂盒可以包含：缓冲液、筒(cartridge)或两者。试剂盒可以包含：一种或更多种用于逆转录反应和/或扩增反应的试剂。

[0032] 在一些实施方案中，第一多于一个寡核苷酸条形码各自包含细胞标记。在一些实施方案中，第一多于一个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，与同一第一固体支持物关联的第一多于一个寡核苷酸条形码中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，与不同的第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码包含不同的细胞标记。在一些实施方案中，第一固体支持物包括第一合成颗粒、第一平坦表面或其组合。在一些实施方案中，第二固体支持物包括第二合成颗粒、第二平坦表面或其组合。

[0033] 在一些实施方案中，第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被固定或部分地固定在第一合成颗粒上，或者第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被包封或部分地包封在第一合成颗粒中。在一些实施方案中，多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被固定或部分地固定在第二合成颗粒上，或者多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被包封或部分地包封在第二合成颗粒中。在一些实施方案中，第一合成颗粒和/或第二

合成颗粒是可破坏的(例如,可破坏的水凝胶颗粒)。在一些实施方案中,第一合成颗粒和/或第二合成颗粒包括珠(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)。在一些实施方案中,第一合成颗粒和/或第二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,第一多于一个寡核苷酸条形码中的每一个寡核苷酸条形码包含接头官能团,第一合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自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、一种或更多种酮及其任何组合组成的组。在一些实施方案中,多于一个一种或更多种条形码化对照核酸中的每一种条形码化对照核酸包含接头官能团,第二合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自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。

[0034] 附图简述

[0035] 图1图示了非限制性示例性条形码。

[0036] 图2示出条形码化和数字计数的非限制性示例性工作流程。

[0037] 图3是示出用于从多于一个靶产生在3'末端条形码化的靶的索引文库的非限制性示例性过程的示意图。

[0038] 图4是使用本文提供的条形码化对照核酸进行单细胞mRNA测序分析的非限制性示例性工作流程的示意图。

[0039] 图5是使用本文提供的条形码化对照核酸进行单细胞mRNA测序分析的非限制性示例性工作流程的示意图。

[0040] 详述

[0041] 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。在不脱离本文呈现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。

[0042] 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。

[0043] 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))和Rhapsody™测定(Becton,Dickinson and Company(Franklin Lakes,NJ))的随机条形码化可以通过使用分子标记(ML)在逆转录(RT)期间标记mRNA来纠正由PCR和文库制备步骤引起的偏倚。

[0044] 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)独特分子标记序列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。每一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、具有独特分子标记序列的随机条形码的数目以及mRNA分子的数目。

[0045] 本文的公开内容包括用于标记样品中的核酸靶的方法。在一些实施方案中,该方法包括:将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化,以产生多于一个条形码化核酸分子,所述多于一个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;提供多于一个一种或更多种条形码化对照核酸,其中一种或更多种条形码化对照核酸的每一种的拷贝数是预定的;产生包含多于一个核酸靶文库成员和多于一个对照核酸文库成员的测序文库,其中产生测序文库包括:将测序衔接子附接至多于一个条形码化核酸分子或其产物,以产生多于一个核酸靶文库成员;以及将测序衔接子附接至多于一个一种或更多种条形码化对照核酸或其产物,以产生多于一个对照核酸文库成员;以及获得包含一个或更多个核酸靶文库成员的多于一个测序读段和一个或更多个对照核酸文库成员的多于一个测序读段的测序数据。

[0046] 本文的公开内容包括试剂盒。试剂盒可以包含:多于一个一种或更多种条形码化对照核酸,其中一种或更多种条形码化对照核酸的每一种的拷贝数是预定的。在一些实施方案中,多于一个一种或更多种条形码化对照核酸与第二固体支持物关联。试剂盒可以包含:第一多于一个寡核苷酸条形码,其中多于一个寡核苷酸条形码的每一个包含分子标记和靶结合区,并且其中多于一个寡核苷酸条形码中的至少10种包含不同的分子标记序列。

[0047] 定义

[0048] 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。

[0049] 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。可以存在于多于一个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个不同序列。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个不同序列。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。与靶核酸衔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。与靶核酸衔接的一种或更多种通用引物可以彼此相同或不同。

[0050] 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”、“衔接”或“固定”或与共同的固体或半固体表面“拴系”、“衔接”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)衔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。

[0051] 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“互补物”。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“反向互补物”。如本文使用的,“互补”序列可以指序列的“互补物”或“反向互补物”。从本公开内容理解,如果一个分子可以与另一个分子杂交,则其可以与其所杂交的分子互补或部分互补。

[0052] 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。

[0053] 如本文使用的,术语“一种标记(label)”或“多于一个标记(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连接处,例如天然和非天然序列的连接处。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。

[0054] 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,每一种靶可能与独特条形码关联。每一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,多于一个靶分子被给定标记标记的概率非常低)。

[0055] 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、蔗糖核酸、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。

[0056] 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。连键(linkage)或主链可以是3'到5'磷酸二酯连键。

[0057] 核酸可以包括修饰的主链和/或修饰的核苷间连键。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)。

[0058] 核酸可以包括由短链烷基或环烷基核苷间连键,混合杂原子,和烷基或环烷基核苷间连键,或者一个或多个短链杂原子的或杂环的核苷间连键形成的多核苷酸主链。这些可以包括具有吗啉代(morpholino)连键的那些(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；甲乙酰基(formacetyl)和硫代甲乙酰基主链；亚甲基甲乙酰基和硫代甲乙酰基主链；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。

[0059] 核酸可以包括核酸模拟物。术语“模拟物”可以意图包括其中仅呋喃糖环或呋喃糖环和核苷酸间连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更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。

[0060] 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连键可以替代磷酸二酯连键。

[0061] 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复合体的稳定性。CeNA寡腺苷酸酯可以与核酸互补物形成复合体，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-氧亚甲基连键，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。

[0062] 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) 咪唑-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)。

[0063] 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。

[0064] 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。

[0065] 如本文使用的, 术语“固体支持物”可以指可以附接多于一个条形码(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。

[0066] 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。

[0067] 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。

[0068] 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。

[0069] 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。

[0070] 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地芽孢杆菌(*Geobacillus stearothermophilus*)GsI-IIC内含子逆转录酶。其他类别的逆转录酶可以包括许多类型的非逆转录病毒逆转录酶(即,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)。

[0071] 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个靶时,每一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,多于一个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个靶时,至少两种靶特异性序列连接到不同的通用衔接子序列。通用衔接子引物及其互补物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。

[0072] 条形码

[0073] 条形码化,诸如随机条形码化,已描述于以下中:例如,Fu等人,Proc Natl Acad Sci U.S.A.,2011May 31;108(22):9026-31;美国专利申请公布第US2011/0160078号;Fan等人,Science,2015February 6,347(6222):1258367;美国专利申请公布第US2015/0299784号和PCT申请公布第W02015/031691号;这些中的每一个的内容,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,所述随机条形码可以是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或这些值中的任何两个值之间的数字或范围,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。

[0074] 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可以包含可以将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处于任何顺序。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个或20个或更多个核苷酸隔开。

[0075] 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,例如每一种七个核苷酸(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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,其他长度的核酸子序列可以用来产生错误校正代码。

[0076] 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,多于一个靶可以包括脱氧核糖核酸(DNA)。

[0077] 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一个标记被间隔区隔开。

[0078] 通用标记

[0079] 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对于附接至给定固体支持物的条形码组中的所有条形码可以是相同的。在一些实施方案中,一种或更多种通用标记对于附接至多于一个珠的所有条形码可以是相同的。在一些实施方案中,通用标记可以包括能够与测序引物杂交的核酸序列。测序引物可以用于对包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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。例如,通用标记可以包括至少约10个核苷酸。通用标记的长度可以是至少以下或可以是至多以下: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。

[0080] 维度标记

[0081] 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。每次脉冲(例如,细胞周期的每个时期)时的条形码可以包含不同的维度标记。以这种方式,维度标记提供关于哪些靶在细胞周期的哪个时期被标记的信息。维度标记可以探询许多不同的生物学时间。示例性的生物学时间可以包括但不限于细胞周期、转录(例如,转录起始)和转录物降解。在另一种实例中,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。

[0082] 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,光罩(photocaging))以及引入非天然的核苷酸将维度标记激活。

[0083] 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。

[0084] 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,或这些值中的任何两个值之间的数字或范围的核苷酸。维度标记的长度可以是至少以下或可以是至多以下: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。

[0085] 空间标记

[0086] 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(identifiable identifier)(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,将多于一个空间标记一起用于编码空间中的一个或更多个位置。

[0087] 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%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,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。

[0088] 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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。空间标记的长度可以是至少以下或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。空间标记可以包含约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。

[0089] 细胞标记

[0090] 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%或这些值中的任何两个值之间的数字或范围。在一些实施方案中,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。

[0091] 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个核苷酸或这些值中的任何两个值之间的数字或范围的核

苷酸。细胞标记的长度可以是至少以下或可以是至多以下：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。

[0092] 条形码序列

[0093] 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，提供粗略估计)的核酸序列。

[0094] 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，可以有以下或可以有约以下的独特分子标记序列： 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种或这些值中的任何两个值之间的数字或范围。例如，多于一个条形码可以包括约6561种具有不同序列的条形码序列。作为另一种实例，多于一个条形码可以包括约65536种具有不同序列的条形码序列。在一些实施方案中，可以有至少以下或可以有至多以下的独特条形码序列： 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。独特分子标记序列可以附接至给定固体支持物(例如，珠)。在一些实施方案中，独特分子标记序列被颗粒(例如水凝胶珠)部分或全部包含。

[0095] 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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。作为另一种实例，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。

[0096] 分子标记

[0097] 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。

[0098] 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，可以有以下或可以有约以下的独特分子标记序列： 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种或这些值中的任何两个值之间的数字或范围。例如，多于一个条形码可以包括约6561种具有不同序列的分子标记。作为另一种实例，多于一个条形码可以包括约65536种具有不同序列的分子标记。在一些实施方案中，可以有至少以下或可以有至多以下的独特分子标记序列： 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。具有独特分子标记序列的条形码可以附接至给定固体支持物(例如，珠)。

[0099] 对于使用多于一个随机条形码进行的条形码化(例如随机条形码化)，不同分子标记序列的数量与任何靶的出现次数的比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，或这些值中的任何两个值之间的数字或范

围。靶可以是包括具有相同或几乎相同序列的mRNA分子的mRNA物质。在一些实施方案中,不同分子标记序列的数量与任何靶的出现次数的比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。

[0100] 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个或这些值中的任何两个值之间的数字或范围的核苷酸。分子标记的长度可以是至少以下或可以是至多以下: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。

[0101] 靶结合区

[0102] 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附接(例如,杂交)至特定靶核酸的特定位置的核酸序列。在一些实施方案中,靶结合区可以包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。

[0103] 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个靶核酸的序列。例如,靶结合区可以包含随机多聚体序列、多(dA)序列、多(dT)序列、多(dG)序列、多(dC)序列或其组合。例如,靶结合区可以是与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,对于附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,对于附接至给定珠的多于一个条形码,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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。靶结合区的长度可以是至多约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更多个核苷酸。例如,可以使用逆转录酶诸如Moloney鼠白血病病毒(MMLV)逆转录酶逆转录mRNA分子,以产生具有多(dC)尾的cDNA分子。条形码可以包括具有多(dG)尾的靶结合区。在条形码的多(dG)尾和cDNA分子的多(dC)尾之间碱基配对后,逆转录酶将模板链从细胞RNA分子转换到条形码,并继续向条形码的5'末端复制。通过这样做,得到的cDNA分子在该cDNA分子3'末端上含有条形码序列(诸如分子标记)。

[0104] 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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。靶结合区的长度可以是至少以下或可以是至多以下: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,条形码在本文中可以称为基因特异性条形码。

[0105] 定向特性(Orientation Property)

[0106] 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)。

[0107] 亲和特性(Affinity Property)

[0108] 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。

[0109] 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)。

[0110] 抗体片段可以是例如抗体的一部分,诸如F(ab')₂、Fab'、Fab、Fv、sFv等。在一些实施方案中,抗体片段可以与由全长抗体识别的相同抗原结合。抗体片段可以包括由抗体的可变区组成的分离的片段,诸如由重链和轻链的可变区组成的“Fv”片段和其中轻链和重链可变区通过肽接头连接的重链多肽分子(“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。

[0111] 通用衔接子引物

[0112] 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个或这些值中的任何两个值之间的数字或范围的核苷酸。通用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。

[0113] 接头

[0114] 当条形码包含多于一个类型的标记(例如,多于一个细胞标记或多于一个条形码

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。

[0115] 固体支持物

[0116] 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个条形码(例如,第一多于一个条形码)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个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个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个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个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。

[0117] 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。

[0118] 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。

[0119] 在一些实施方案中,颗粒可以是可破坏的(例如,可溶解的、可降解的)。例如,聚合物珠可以例如在期望的条件下溶解、熔化或降解。所期望的条件可以包括环境条件。所期望的条件可以导致聚合物珠以受控方式溶解、熔化或降解。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、熔化或降解。

[0120] 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,多于一个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。

[0121] 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)。

[0122] 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。

[0123] 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。

[0124] 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。

[0125] 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,热量可以增加珠的内部组的内部压力,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。

[0126] 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。在一个实例中,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。

[0127] 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。

[0128] 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。

[0129] 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。珠可以包含多于一个条形码。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。

[0130] 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,所述百分比可以是以下或可以是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或这些值中的任何两个值之间的数字或范围。作为另一个实例,所述百分比可以是至少以下或可以是至多以下: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以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。

[0131] 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)。在一些实施方案中,可以用包括与多于一个条形码关联的多于一个合成颗粒的固体支持物将样品中的多于一个靶条形码化。在一些实施方案中,固体支持物可以包括与多于一个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个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个条形码。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。

[0132] 如本文使用的,术语“拴系的”、“衔接的”和“固定的”可以互换使用,并且可以指用于将条形码衔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,以用于衔接预先合成的条形码或用于条形码的原位固相合成。

[0133] 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,所述形状诸如立方形、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。

[0134] 珠可以包括各种材料,包括但不限于顺磁性材料(例如,镁、钼、锂和钽)、超顺磁性材料(例如,铁氧体(Fe_3O_4 ;磁铁矿)纳米颗粒)、铁磁材料(例如,铁、镍、钴,它们的一些合金,以及一些稀土金属化合物)、陶瓷、塑料、玻璃、聚苯乙烯、二氧化硅、甲基苯乙烯、丙烯酸聚合物、钛、胶乳、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。

[0135] 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。

[0136] 本文公开的一些实施方案包括一个或多个颗粒(例如,珠)。每个颗粒可以包含多于一个寡核苷酸(例如,条形码)。多于一个寡核苷酸中的每一个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)。多于一个寡核苷酸的每一个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、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 、 10^9 、这些值中的任何两个值之间的数字或范围或更多。在一些实施方案中,细胞标记序列的数目可以是至少以下或可以是至多以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 或 10^9 。在一些实施方案中,多于一个颗粒中的不超过1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000或更多颗粒包括具有相同细胞序列的寡核苷酸。在一些实施方案中,包括具有相同细胞序列的寡核苷酸的多于一个颗粒可以是至多0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%或更多。在一些实施方案中,多于一个颗粒全都不具有相同的细胞标记序列。

[0137] 在每个颗粒上的多于一个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、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 、 10^9 ,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,条形码序列的数目可以是至少以下或可以是至多以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 或 10^9 。例如,多于一个寡核苷酸中的至少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,多于一个寡核苷酸中的至少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值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个包含条形码的颗粒。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,多于一个寡核苷酸的每一个还包含样品标记、通用标记或二者。颗粒可以是例如纳米颗粒或微米颗粒。

[0138] 珠的尺寸可以不同。例如,珠的直径范围可以为从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微米或这些值中

的任何两个值之间的数字或范围。

[0139] 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%或这些值中的任何两个值之间的数字或范围。珠的直径可以与细胞(例如,被基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约以下:10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或这些值中的任何两个值之间的数字或范围。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%。

[0140] 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。

[0141] 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。

[0142] 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。

[0143] 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。

[0144] 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。

[0145] 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)。

[0146] 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。

[0147] 基底和微孔阵列

[0148] 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。

[0149] 条形码化的方法

[0150] 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/或将所述信息可视化。在一些实施方案中,该方法包括使样品中的多于一个靶可视化。将多于一个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个靶条形码化(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。将样品中的多于一个靶可视化可以包括将多于一个靶映射到样品的映射图上。将多于一个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个靶进行条形码化之前或之后产生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产生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。

[0151] 在一些实施方案中,将多于一个靶条形码化包括将多于一个条形码与多于一个靶杂交以产生条形码化靶(例如,随机条形码化靶)。将多于一个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个条形码(例如,随机条形码)的固体支持物来进行。

[0152] 使样品和条形码接触

[0153] 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。

[0154] 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,使得每一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。

[0155] 细胞裂解

[0156] 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。

[0157] 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。

[0158] 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。裂解缓冲液中的还原剂浓度可以是至多约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。

[0159] 裂解可以在约4°C、10°C、15°C、20°C、25°C或30°C的温度进行。裂解可以进行约1分钟、5分钟、10分钟、15分钟或20分钟或更长时间。裂解的细胞可以包括至少约100000个、200000个、300000个、400000个、500000个、600000个或700000个或更多个靶核酸分子。裂解的细胞可以包括至多约100000个、200000个、300000个、400000个、500000个、600000个或700000个或更多个靶核酸分子。

[0160] 将条形码附接至靶核酸分子

[0161] 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个探针关联(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中图示的条形码化的非限制性实例中,在框216处,mRNA分子可以与珠上的条形码杂交。例如,单链的核苷酸片段可以与条形码的靶结合区杂交。

[0162] 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。

[0163] 例如,在图2中图示的条形码化的非限制性实例中,在框220处,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。

[0164] 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。

[0165] 逆转录或核酸延伸

[0166] 本公开内容提供了使用逆转录(例如,在图2的框224处)或核酸延伸来产生靶-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。

[0167] 在一些实施方案中,mRNA分子向标记的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。

[0168] 在一些实施方案中,靶是cDNA分子。例如,可以使用逆转录酶诸如Moloney鼠白血病病毒(MMLV)逆转录酶逆转录mRNA分子,以产生具有多(dC)尾的cDNA分子。条形码可以包括具有多(dG)尾的靶结合区。在条形码的多(dG)尾和cDNA分子的多(dC)尾之间碱基配对后,逆转录酶将模板链从细胞RNA分子转换到条形码,并继续向条形码的5'末端复制。通过

这样做,得到的cDNA分子在该cDNA分子3'末端上含有条形码序列(诸如分子标记)。

[0169] 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。

[0170] 扩增

[0171] 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,其中多于一个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个核酸的0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、100%或这些值中的任何两个值之间的范围或数字。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。

[0172] 在一些实施方案中,扩增可以使用聚合酶链式反应(PCR)来进行。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以及组装PCR。

[0173] 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QB复制酶(QB)方法、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。

[0174] 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。

[0175] 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。

[0176] 进行一个或多个扩增反应可以包括使用一种或更多种引物。一种或更多种引物可以包括例如,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个标记的靶(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个标记的靶的3'端或5'端。一种或更多种引物可以退火至多于一个标记的靶的内部区域。内部区域可以与多于一个标记的靶的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。

[0177] 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。定制引物可以被设计成扩增一种或更多种靶。靶可以包括一个或多个样品中总核酸的子集。靶可以包括一个或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。

[0178] 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。

[0179] 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经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,通过限制性内切核酸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)内切核酸酶)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。

[0180] 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。

[0181] 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附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附接至颗粒的互补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,通过在同一反应混合物中添加核苷酸,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。

[0182] 扩增反应可以包括扩增多于一个核酸的至少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%。

[0183] 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以及组装PCR。

[0184] 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)、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)。

[0185] 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。

[0186] 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,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。

[0187] 扩增还可以包括将一种或更多种对照核酸添加至一个或多个包含多于一个核酸的样品中。扩增还可以包括将一种或更多种对照核酸添加至多于一个核酸。对照核酸可以包含对照标记。

[0188] 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。

[0189] 进行一个或多个扩增反应可以包括使用一种或更多种引物。一种或更多种引物可以包括一种或更多种寡核苷酸。一种或更多种寡核苷酸可以包含至少约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个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个标记的核酸的3'端和/或5'端。一种或更多种引物可以退火至多于一个标记的核酸的内部区域。内部区域可以与多于一个标记的核酸的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。

[0190] 在一些实施方案中,将样品中的多于一个靶条形码化(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个靶产生多于一个索引多核苷酸。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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,产生条形码化靶的索引文库包括使多于一个靶(例如mRNA分子)与包含多(T)区和标记区的多于一个寡核苷酸接触;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(每一种包含cDNA区和标记区),其中多于一个靶包括至少两种不同序列的mRNA分子,且多于一个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。

[0191] 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个体核酸(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。

[0192] 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序列、细胞标记序列和通用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。

[0193] 在一些实施方案中,细胞标记序列可以包含3个至20个核苷酸。在一些实施方案中,分子标记序列可以包含3个至20个核苷酸。在一些实施方案中,多于一个随机条形码中的每一个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个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个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。

[0194] 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个或这些值中的任何两个值之间的数字或范围的核苷酸。细胞标记320的长度可以是以下、可以是约以下、可以是至少以下或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个或这些值中的任何两个值之间的数字或范围的核苷酸。通用标记的长度可以是以下、可以是约以下、可以是至少以下或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个或这些值中的任何两个值之间的数字或范围的核苷酸。通用标记对于固体支持物上的多于一个随机条形码可以是相同的,并且细胞标记对于固体支持物上的多于一个随机条形码是相同的。维度标记的长度可以是以下、可以是约以下、可以是至少以下或可以是至多以下:1个、2个、3个、4个、5个、6个、

7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个或这些值中的任何两个值之间的数字或范围的核苷酸。

[0195] 在一些实施方案中,标记区314可以包括以下、可以包括约以下、可以包括至少以下或可以包括至多以下:1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种或这些值中的任何两个值之间的数字或范围的不同标记,诸如条形码序列或分子标记318和细胞标记320。每一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个核苷酸或这些值中的任何两个值之间的数字或范围的核苷酸。一组条形码或随机条形码310可以含有以下、可以含有约以下、可以含有至少以下或可以含有至多以下:10种、20种、40种、50种、70种、80种、90种、 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种、 10^{10} 种、 10^{11} 种、 10^{12} 种、 10^{13} 种、 10^{14} 种、 10^{15} 种、 10^{20} 种或这些值中的任何两个值之间的数字或范围的条形码或随机条形码310。并且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。

[0196] 如步骤2中示出的,来自步骤1中的逆转录过程的产物可以汇集至1支管中,并且用第1PCR引物池和第1通用PCR引物进行PCR扩增。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、利用约、利用至少或利用至多10、20、40、50、70、80、90、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、 10^9 、 10^{10} 、 10^{11} 、 10^{12} 、 10^{13} 、 10^{14} 、 10^{15} 、 10^{20} 个或这些值中的任何两个值之间的数字或范围的多重引物。扩增可以包括使用包括靶向特定基因的定制引物326A-C的第1PCR引物池324和通用引物328。定制引物326可以与标记的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。

[0197] 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扩增倚倚最小化。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可以包括在单个反应体积中用巢式PCR引物332a-c的巢式PCR引物池330和第2通用PCR引物328'进行的多重PCR。巢式PCR引物池330可以包含以下、可以包含约以下、可以包含至少以下或可以包含至多以下: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种或这些值中的任何两个值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。

[0198] 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。

[0199] 条形码化对照核酸

[0200] 在一些实施方案中,提供了加标标准品(例如,Rhapsody珠加标标准品)。在一些实施方案中,加标标准品是用于逆转录下游测定步骤的标准化cDNA加标对照。在一些实施方案中,提供了固体支持物(例如,对照珠),所述固体支持物(例如,对照珠)与转录到珠上的标准受控量的DNA或RNA关联。珠关联的寡核苷酸可以包含细胞标记(或一组细胞标记),诸如对对照部分号(control part number)特异性的对照标记。细胞标记的存在可以使用户能够分离来源于所述对照珠的测序读段(例如,在多重测序反应中与样品来源的测序读段区分开)。控制珠上的cDNA量可以允许用户针对对照结果来衡量他们的测定是否成功,并且可以帮助确定样品是否已经足够深入地测序。本文提供的方法和组合物使用目前不可用的组合物(例如ERCC加标对照)解决长期以来觉察到的需求。本文公开的方法和组合物(例如,固体支持物相关的条形码化对照核酸)包含已知的条形码组(例如,对照标记)以在测序分析期间区分对照序列与真实样品序列,允许更复杂的核酸和更好的细胞近似。在一些实施方案中,提供了对照固体支持物(与包含对照标记和能够与一种或更多种对照核酸杂交的靶结合区的寡核苷酸条形码关联)。在所述对照固体支持物(例如,对照珠)上,标准(例如,预定)量的标准品(例如,对照)核酸(例如,DNA或RNA)可以杂交和逆转录。这些标准品可以在用户使用之前进行验证。在一些实施方案中,提供了自动分离和定量这些对照的分析管线。

[0201] 对于用户来说,试图使目前可用的标准品核酸杂交和条形码化并试图标准化珠上的质量(mass)将是昂贵和不切实际的。除了其他问题以外,这将要求用户开发方法来验证与他们的珠杂交的量。此外,用户将必须以某种方式将另一个转录组添加至他们的分析管线中。如本文描述的,使用对我们的对照珠特异性的条形码,使得有机会超越目前可用的标准品(例如,ERCC)分支至具有更高多样性和更能代表实际细胞的核酸。此外,在一些实施方案中,对照核酸的增加的复杂性是优选的,或者甚至是必要的,诸如,例如,如果对照被用作如本文公开的VDJ分析的对照。

[0202] 在一些实施方案中,条形码化颗粒(例如,Rhapsody珠)包含细胞标记和能够与mRNA(随后被逆转录)杂交的靶结合区。所述对照珠上产生的cDNA的量可以基于输入量、珠批次捕获效率和逆转录反应的效率而变化。在一些实施方案中,提供了包含指示对照序列的对照标记的条形码化颗粒。在一些实施方案中,严格控制量的DNA或RNA附接(例如,逆转录)到所述珠上。在一些实施方案中,珠寡聚体(oligo)密度和/或核酸输入是标准化的,并且可以用荧光探针来验证。对于本文描述的对照珠具有单独条形码可以允许使用复杂的核酸环境。

[0203] 图4是使用本文提供的条形码化对照核酸进行单细胞mRNA测序分析的非限制性示例性工作流程的示意图。工作流程可以包括分区(例如,孔、液滴)中的单细胞捕获402。工作

流程可以包括将来自所述单细胞中的每一个的转录物逆转录404到与珠关联的寡核苷酸条形码上,以产生多于一个条形码化核酸分子。工作流程可以包括条形码化对照核酸的加标406。条形码化对照核酸中的每一种的拷贝数可以是预定的。条形码化对照核酸可以与固体支持物(例如珠)关联。工作流程可以包括使条形码化核酸分子和条形码化对照核酸经历一个或更多个延伸和/或扩增反应408,诸如靶特异性扩增、全转录组扩增或VDJ分析,从而产生包含多于一个核酸靶文库成员和多于一个对照核酸文库成员的测序文库。工作流程可以包括获得包含一个或更多个核酸靶文库成员的多于一个测序读段和一个或更多个对照核酸文库成员的多于一个测序读段的测序数据410。在一些实施方案中,如果对照(例如,条形码化对照核酸)起作用而用户样品失败,则失败点在条形码化对照核酸的加标之前(例如,逆转录或更早)。在一些实施方案中,如果对照失败,则失败点在测定内。用户可以将对照的已知灵敏度与他们的检测进行比较,以确定他们是否从其样品中获得了最大收益。在一些实施方案中,用户可以保存先前成功的对照珠PCR模板来控制下游测序步骤。

[0204] 图5是使用本文提供的条形码化对照核酸进行单细胞mRNA测序分析的非限制性示例性工作流程的示意图。逆转录反应可以产生衔接(例如,缀合、共价衔接、非共价衔接)至固体支持物502(例如,珠)的条形码化核酸分子504b和506b。条形码化核酸分子504b和506b可以包含条形码(例如,随机条形码)。条形码可以包含靶结合区(例如,多(dT)尾514),其可以与RNA分子(例如,经由多(dA)尾与多腺苷酸化mRNA转录物)或其他核酸靶结合,用于标记或条形码化(例如,独特标记)。条形码可以包含许多标记,诸如独特分子索引(UMI) 512、细胞标记(CL) 510和通用PCR柄(universal PCR handle, Univ) 508(其可以包括或可以是例如测序文库扩增引物,诸如读段1测序引物的结合位点)。通用PCR柄可以包括第一通用引物、其互补序列、其部分序列或其组合。条形码化核酸分子504b和506b可以包含分别来源于核酸靶的逆转录的cDNA 516c1和cDNA 516c2。工作流程可以包括衔接(例如,缀合、共价衔接、非共价衔接)至固体支持物518(例如,珠)的条形码化对照核酸520b和522b的加标。条形码化对照核酸520b和522b的每一种可以包含通用PCR柄(Univ) 508(其可以包括或可以是例如测序文库扩增引物,诸如读段1测序引物的结合位点)。条形码化对照核酸520b和522b的每一种可以包含对照标记524。条形码化对照核酸520b和522b可以包括分别来源于(例如,逆转录自)对照核酸的对照核酸526c1(例如cDNA)和对照核酸526c2(例如,cDNA)。条形码化对照核酸520b和522b可以包含靶结合区(例如,多(dT)尾514),其可以与用于标记或条形码化(例如,独特标记)的对照核酸结合。条形码化对照核酸520b和522b的拷贝数可以是预定的。工作流程可以包括使条形码化核酸分子和条形码化对照核酸经历一个或更多个延伸和/或扩增反应500a,诸如靶特异性扩增、全转录组扩增或VDJ分析,从而产生包含核酸靶文库成员504s和506s以及多于一个对照核酸文库成员520s和522s的测序文库。产生测序文库500a可以添加测序衔接子528和530(例如,P5和P7序列),并且在一些实施方案中,可以添加样品索引(例如,i5、i7)。工作流程可以包括获得包含一个或更多个核酸靶文库成员的多于一个测序读段和一个或更多个对照核酸文库成员的多于一个测序读段的测序数据500b。在一些实施方案中,转录的对照核酸的数目落在可接受值的范围内。在一些实施方案中,将珠用外切核酸酶处理以去除过量的珠寡核苷酸(例如,未与对照核酸结合并被逆转录的珠关联的寡核苷酸)。测序数据中太少的标准品对照分子(例如,对照核酸)可以向用户指示他们的测定中出了问题或/或他们应该更深入地测序。

[0205] 在一些实施方案中,提供了用于标记样品中核酸靶的方法。在一些实施方案中,该方法包括:将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化,以产生多于一个条形码化核酸分子,所述多于一个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;提供多于一个一种或更多种条形码化对照核酸,其中一种或更多种条形码化对照核酸的每一种的拷贝数是预定的;产生包含多于一个核酸靶文库成员和多于一个对照核酸文库成员的测序文库,其中产生测序文库包括:将测序衔接子附接至多于一个条形码化核酸分子或其产物,以产生多于一个核酸靶文库成员;以及将测序衔接子附接至多于一个一种或更多种条形码化对照核酸或其产物,以产生多于一个对照核酸文库成员;以及获得包含一个或更多个核酸靶文库成员的多于一个测序读段和一个或更多个对照核酸文库成员的多于一个测序读段的测序数据。

[0206] 在一些实施方案中,将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化包括:使核酸靶的拷贝与第一多于一个寡核苷酸条形码接触,其中第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶杂交的靶结合区;以及使与核酸靶的拷贝杂交的第一多于一个寡核苷酸条形码延伸,以产生多于一个条形码化核酸分子,所述多于一个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。多于一个条形码化核酸分子的每一个条形码化核酸分子可以包含第一通用序列和分子标记。样品可以包括单细胞。样品可以包括多于一个单细胞。该方法可以包括:在使核酸靶的拷贝与第一多于一个寡核苷酸条形码接触之前,将多于一个单细胞分区至多于一个分区,其中多于一个分区中的分区包括来自多于一个单细胞的单细胞;以及在包含单细胞的分区中,使核酸靶的拷贝与第一多于一个寡核苷酸条形码接触。分区可以是孔或液滴。第一多于一个寡核苷酸条形码可以与第一固体支持物关联。该方法可以包括:将第一固体支持物与样品中的单细胞关联,并且其中多于一个分区中的分区包括单个第一固体支持物。该方法可以包括:在分区步骤之后且在接触步骤之前裂解单细胞。裂解单细胞可以包括加热样品、使样品与去污剂接触、改变样品的pH或其任何组合。

[0207] 在一些实施方案中,多于一个一种或更多种条形码化对照核酸通过以下方式产生:使预定拷贝数的一种或更多种对照核酸与第二多于一个寡核苷酸条形码接触,其中第二多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、对照标记和能够与一种或更多种对照核酸杂交的靶结合区;以及使与一种或更多种对照核酸杂交的第二多于一个标记的寡核苷酸延伸,以产生预定拷贝数的一种或更多种条形码化对照核酸,所述一种或更多种条形码化对照核酸各自包含与一种或更多种条形码化对照核酸的至少一部分互补的序列。在一些实施方案中,第二多于一个寡核苷酸条形码与第二固体支持物关联和/或多于一个一种或更多种条形码化对照核酸与第二固体支持物关联。使第一多于一个寡核苷酸条形码和/或第二多于一个寡核苷酸条形码延伸可以包括使用逆转录酶和/或缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶(例如,Klenow片段)使第一多于一个寡核苷酸条形码和/或第二多于一个寡核苷酸条形码延伸。逆转录酶可以包括病毒逆转录酶(例如,鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶)。

[0208] 条形码化对照核酸的每一种可以包含第一通用序列、对照标记和靶结合区中的一种或更多种。第二多于一个寡核苷酸条形码的每一个对照标记的长度可以是以下、可以是

约以下、可以是至少以下或可以是至多以下：1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个或这些值中的任何两个值之间的数字或范围的核苷酸。对于与一个固体支持物关联的所有寡核苷酸条形码，对照标记可以是相同的。与不同固体支持物关联的寡核苷酸条形码之间的对照标记可以是相同的或不同的。一种或更多种条形码化对照核酸可以包括两种或更多种不同的条形码化对照核酸。多于一个一种或更多种条形码化对照核酸可以包括以下的不同的条形码化对照核酸：约1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、11种、12种、13种、14种、15种、16种、17种、18种、19种、20种、25种、30种、35种、40种、45种、50种、60种、70种、80种、90种、100种、110种、120种、128种、130种、140种、150种、160种、170种、180种、190种、200种、210种、220种、230种、240种、250种、260种、270种、280种、290种、300种、310种、320种、330种、340种、350种、360种、370种、380种、390种、400种、410种、420种、430种、440种、450种、460种、470种、480种、490种、500种、510种、520种、530种、540种、550种、560种、570种、580种、590种、600种、610种、620种、630种、640种、650种、660种、670种、680种、690种、700种、710种、720种、730种、740种、750种、760种、770种、780种、790种、800种、810种、820种、830种、840种、850种、860种、870种、880种、890种、900种、910种、920种、930种、940种、950种、960种、970种、980种、990种、1000种或这些值中的任何两个值之间的数字或范围。在一些实施方案中，多于一个一种或更多种条形码化对照核酸中不同条形码化对照核酸的数目是至少以下或者是至多以下：1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、11种、12种、13种、14种、15种、16种、17种、18种、19种、20种、25种、30种、35种、40种、45种、50种、60种、70种、80种、90种、100种、110种、120种、128种、130种、140种、150种、160种、170种、180种、190种、200种、210种、220种、230种、240种、250种、260种、270种、280种、290种、300种、310种、320种、330种、340种、350种、360种、370种、380种、390种、400种、410种、420种、430种、440种、450种、460种、470种、480种、490种、500种、510种、520种、530种、540种、550种、560种、570种、580种、590种、600种、610种、620种、630种、640种、650种、660种、670种、680种、690种、700种、710种、720种、730种、740种、750种、760种、770种、780种、790种、800种、810种、820种、830种、840种、850种、860种、870种、880种、890种、900种、910种、920种、930种、940种、950种、960种、970种、980种、990种或1000种不同的条形码化对照核酸。

[0209] 在一些实施方案中，多于一个一种或更多种条形码化对照核酸可以包括至少20种、至少30种、至少40种、至少50种、至少60种、至少70种、至少80种、至少90种、至少100种、至少200种、至少300种、至少400种、至少500种、至少600种、至少700种、至少800种、至少900种、至少1,000种、至少2,000种、至少5,000种或更多种不同的条形码化对照核酸。一种或更多种条形码化对照核酸可以与核酸靶至少约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%同源，或这些值中的任何两个值之间的数字或范围。

[0210] 一种或更多种条形码化对照核酸可以与至少约1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种或这些值中的任何两个值之间的数字或范围的不同核酸靶至少约70%同源。一种或更多种条形码化对照核酸可以包含持家基因的序

列。一种或更多种条形码化对照核酸可以与样品的基因组序列同源。一种或更多种条形码化对照核酸可以与样品的基因组序列不同源。一种或更多种条形码化对照核酸可以与物种的基因组序列同源。该物种可以是非哺乳动物物种。非哺乳动物物种可以是噬菌体物种(例如, T7噬菌体、PhiX噬菌体或其任何组合)。

[0211] 在一些实施方案中, 产生测序文库包括: 使随机引物与多于一个条形码化核酸分子和多于一个一种或更多种条形码化对照核酸接触, 其中随机引物的每一个包含第二通用序列或其互补物; 使与多于一个条形码化核酸分子杂交的随机引物延伸, 以产生第一多于一个延伸产物; 以及使与多于一个一种或更多种条形码化对照核酸杂交的随机引物延伸, 以产生第二多于一个延伸产物。该方法可以包括: 使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第一多于一个延伸产物, 从而产生第一多于一个条形码化扩增子; 以及使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第二多于一个延伸产物, 从而产生第二多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第一多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第二多于一个条形码化扩增子或其产物。扩增第一多于一个延伸产物和第二多于一个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至所第一多于一个延伸产物和第二多于一个延伸产物。方法可以包括: 基于与第一多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目来确定样品中核酸靶的拷贝数。

[0212] 在一些实施方案中, 产生测序文库包括: 使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第一多于一个条形码化扩增子, 从而产生第三多于一个条形码化扩增子; 以及使用能够与第一通用序列或其互补物杂交的引物和能够与第二通用序列或其互补物杂交的引物扩增第二多于一个条形码化扩增子, 从而产生第四多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第三多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第四多于一个条形码化扩增子或其产物。扩增第三多于一个条形码化扩增子和第四多于一个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至第三多于一个条形码化扩增子和第四多于一个条形码化扩增子。该方法可以包括: 基于与第三多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目, 确定样品中核酸靶的拷贝数。第一多于一个条形码化扩增子和/或第三多于一个条形码化扩增子可以包含全转录组扩增(WTA)产物。

[0213] 在一些实施方案中, 产生测序文库包括: 使用多于一个条形码化核酸分子作为模板合成第五多于一个条形码化扩增子, 以产生第五多于一个条形码化扩增子; 以及使用多于一个一种或更多种条形码化对照核酸作为模板合成第六多于一个条形码化扩增子, 以产生第六多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第五多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第六多于一个条形码化扩增子或其产物。合成第五多于一个条形码化扩增子和第六多于一个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物杂交的引物和靶特异性引物进行的PCR扩增。合成第五多于一个条形码化扩增子和第六多于一个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至条形码化核酸分子和条形码化对

照核酸。该方法可以包括：基于与第五多于一个条形码化扩增子或其产物关联的具有不同序列的分子标记的数目，确定样品中核酸靶的拷贝数。

[0214] 在一些实施方案中，多于一个靶核酸包括免疫受体。本公开内容的方法可以与鉴定B细胞受体(BCR)、T细胞受体(TCR)和抗体的VDJ区的测定联合使用。VDJ重组又称为体细胞重组，是免疫系统产生免疫球蛋白(Ig)(例如，BCR)和T细胞受体(TCR)的早期阶段中的遗传重组机制。VDJ重组可以几乎随机地组合可变(V)基因区段、多样(D)基因区段和连接(J)基因区段。在一些实施方案中，本文公开的方法允许对T细胞和B细胞的V(D)J谱分析、3' 靶向扩增、5' 靶向扩增、3' 全转录组扩增(WTA)、5' WTA、用AbO进行的蛋白表达谱分析和/或单个实验中的样品多重化分析。使用5' 条形码化和/或3' 条形码化确定核酸靶(例如，免疫受体的V(D)J区)的序列的方法在US2020/0109437中描述；该文献的内容通过引用以其整体并入本文。用于在核酸靶的5' 末端上进行分子条形码化的系统、方法、组合物和试剂盒已经在例如US2019/0338278中描述，该文献的内容通过引用以其整体并入本文。在一些实施方案中，本文提供的系统、方法、组合物和试剂盒可以与使用组合的5' 条形码化和随机引发方法获得全长V(D)J信息(例如，在Rhapsody系统上通过Illumina测序)的方法协同采用，所述方法已在2020年11月6日提交的题为“USING RANDOM PRIMING TO OBTAIN FULL-LENGTH V(D)J INFORMATION FOR IMMUNE REPERTOIRE SEQUENCING”的美国专利申请第17/091,639号中描述，该专利申请的内容通过引用以其整体并入本文。在一些实施方案中，本文提供的系统、方法、组合物和试剂盒可以与基于随机引发和延伸(random priming and extension, RPE)的全转录组分析方法和组合物(其已经在美国专利申请第16/677,012号中描述；该专利申请的内容通过引用以其整体并入本文)协同采用。在一些实施方案中，本文提供的用于基于5' 的基因表达谱分析的系统、方法、组合物和试剂盒可以与阻断剂寡核苷酸协同采用，所述阻断剂寡核苷酸在于2021年1月29日提交的题为“MESOPHILIC DNA POLYMERASE EXTENSION BLOCKERS”的美国专利申请第17/163,177号中描述，该专利申请的内容通过引用以其整体并入本文。在一些实施方案中，本文提供的系统、方法、组合物和试剂盒可以与合成颗粒(例如，条形码化珠)协同采用，所述合成颗粒(例如，条形码化珠)在于2021年6月1日提交的题为“OLIGONUCLEOTIDES AND BEADS FOR 5PRIME GENE EXPRESSION ASSAY”的美国专利申请第17/336,055号中描述，该专利申请的内容通过引用以其整体并入本文。

[0215] 在一些实施方案中，将核酸靶的拷贝用第一多于一个寡核苷酸条形码条形码化包括：使多于一个核酸靶的拷贝与第一多于一个寡核苷酸条形码接触，其中第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶的拷贝杂交的靶结合区；以及产生各自包含靶结合区和靶结合区的互补物的多于一个条形码化核酸分子。在一些实施方案中，产生各自包含靶结合区和靶结合区的互补物的多于一个条形码化核酸分子包括：在存在逆转录酶和包含靶结合区或其一部分的模板转换寡核苷酸的情况下，使与核酸靶的拷贝杂交的第一多于一个寡核苷酸条形码延伸，以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记、靶结合区和靶结合区的互补物的多于一个条形码化核酸分子。在一些实施方案中，逆转录酶具有末端转移酶活性。模板转换寡核苷酸可以包含一个或更多个3' 核糖核苷酸(例如三个3' 核糖核苷酸)。3' 核糖核苷酸可以包含鸟嘌呤。逆转录酶可以是病毒逆转录酶(例如，鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶)。

[0216] 在一些实施方案中,产生测序文库包括:使每一个条形码化核酸分子的靶结合区的互补物与以下中的一种或更多种的靶结合区杂交:(i)第一多于一个寡核苷酸条形码中的寡核苷酸条形码,(ii)条形码化核酸分子本身,以及(iii)多于一个条形码化核酸分子中的不同条形码化核酸分子;使多于一个条形码化核酸分子的3'末端延伸以产生多于一个延伸的条形码化核酸分子,所述多于一个延伸的条形码化核酸分子各自包含第一分子标记和第二分子标记。在一些实施方案中,使条形码化核酸分子的靶结合区的互补物与第一多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交包括使条形码化核酸分子的靶结合区的互补物与第一多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区进行分子间杂交。该方法可以包括:使与条形码化核酸分子的靶结合区的互补物杂交的寡核苷酸条形码的3'末端延伸,以产生多于一个延伸的条形码化核酸分子,所述多于一个延伸的条形码化核酸分子各自包含以下的互补物:第一分子标记,和第二分子标记,其中第二分子标记的序列不同于第一分子标记的序列,其中第二分子标记不是第一分子标记的互补物。多于一个一种或更多种延伸的条形码化核酸各自可以包含5'第一通用序列和3'第一通用序列的互补物。多于一个一种或更多种条形码化对照核酸可以包含5'对照标记和3'对照标记的互补物。

[0217] 本文提供的方法可以包括使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子,从而产生第七多于一个条形码化扩增子;以及使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个一种或更多种条形码化对照核酸,从而产生第八多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第七多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第八多于一个条形码化扩增子或其产物。扩增多于一个延伸的条形码化核酸分子和扩增多于一个一种或更多种条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个一种或更多种条形码化对照核酸。该方法可以包括:基于与第七多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。该方法可以包括:使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子,从而产生第九多于一个条形码化扩增子;以及使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个一种或更多种条形码化对照核酸,从而产生第十多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第九多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第十多于一个条形码化扩增子或其产物。扩增多于一个延伸的条形码化核酸分子和扩增多于一个一种或更多种条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个一种或更多种条形码化对照核酸。该方法可以包括:基于与第九多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0218] 多于一个一种或更多种条形码化对照核酸各自可以包含靶结合区和靶结合区的互补物。产生测序文库可以包括:使每一种条形码化对照核酸的靶结合区的互补物与以下

中的一种或更多种的靶结合区杂交：(i) 第二多于一个寡核苷酸条形码中的寡核苷酸条形码，(ii) 条形码化对照核酸本身，和(iii) 多于一个一种或更多种条形码化对照核酸中的不同条形码化对照核酸；以及使多于一个条形码化对照核酸的3' 末端延伸以产生多于一个延伸的条形码化对照核酸，所述多于一个延伸的条形码化对照核酸各自包含第一通用序列、第一通用序列的互补物、对照标记和对照标记的互补物。在一些实施方案中，使条形码化对照核酸的靶结合区的互补物与第二多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交包括使条形码化对照核酸的靶结合区的互补物与第二多于一个寡核苷酸条形码中的寡核苷酸条形码的靶结合区进行分子间杂交。该方法可以包括：使与条形码化对照核酸的靶结合区的互补物杂交的寡核苷酸条形码的3' 末端延伸，以产生多于一个延伸的条形码化对照核酸。

[0219] 本文提供的方法可以包括使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子，从而产生第十一多于一个条形码化扩增子；以及使用能够与第一通用序列和/或其互补物杂交的引物扩增多于一个延伸的条形码化对照核酸，从而产生第十二多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第十一多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第十二多于一个条形码化扩增子或其产物。扩增多于一个延伸的条形码化核酸分子和扩增多于一个延伸的条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个延伸的条形码化对照核酸。该方法可以包括：基于与第十一多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0220] 本文公开的方法可以包括：使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个延伸的条形码化核酸分子，从而产生第十三多于一个条形码化扩增子；以及使用能够与核酸靶序列杂交的靶特异性引物和能够与第一通用序列或其互补物杂交的引物扩增多于一个延伸的条形码化对照核酸，从而产生第十四多于一个条形码化扩增子。多于一个核酸靶文库成员可以包含第十三多于一个条形码化扩增子或其产物。多于一个对照核酸文库成员可以包含第十四多于一个条形码化扩增子或其产物。扩增多于一个延伸的条形码化核酸分子和扩增多于一个延伸的条形码化对照核酸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至多于一个延伸的条形码化核酸分子和多于一个延伸的条形码化对照核酸。该方法可以包括：基于与第十三多于一个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、具有不同序列的第二分子标记或其组合的数目来确定样品中多于一个核酸靶的每一个的拷贝数。

[0221] 在一些实施方案中，靶特异性引物与免疫受体特异性杂交。在一些实施方案中，靶特异性引物与免疫受体的恒定区、免疫受体的可变区、免疫受体的多样性区、免疫受体的可变区和多样性区的连接区(junction)或它们的组合特异性杂交。免疫受体可以是T细胞受体(TCR)和/或B细胞受体(BCR)。TCR可以包含TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或其任何组合。BCR受体可以包含BCR重链和/或BCR轻链。使多于一个条形码化核酸分子的3' 末端延伸和/或使多于一个条形码化对照核酸的3' 末端延伸可以使用缺乏5' 至3' 外切核酸酶活性和

3'至5'外切核酸酶活性中的至少一种的DNA聚合酶(例如,Klenow片段)进行。该方法可以包括:在存在乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砷(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或其任何组合中的一种或更多种的情况下,使第一多于一个寡核苷酸条形码和/或第二多于一个寡核苷酸条形码延伸。

[0222] 靶结合区可以包含多(dT)区、随机序列、靶特异性序列或其组合。第一通用序列和/或第二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。多于一个条形码化核酸分子可以包括条形码化脱氧核糖核酸(DNA)分子、条形码化核糖核酸(RNA)分子或其组合。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、含有多(A)尾的RNA,或它们的任何组合)。在一些实施方案中,mRNA编码免疫受体。在一些实施方案中,核酸靶包括细胞组分结合试剂,和/或核酸分子与细胞组分结合试剂关联。方法可以包括:使核酸分子和细胞组分结合试剂解离。

[0223] 第一多于一个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。第一多于一个寡核苷酸条形码各自可以包含细胞标记。第一多于一个寡核苷酸条形码的每一种细胞标记可以包含至少6个核苷酸。与同一第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码可以包含相同的细胞标记。与不同的第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码可以包含不同的细胞标记。第一固体支持物可以包括第一合成颗粒、第一平坦表面或其组合。第二固体支持物可以包括第二合成颗粒、第二平坦表面或其组合。在一些实施方案中,第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被固定或部分地固定在第一合成颗粒上,或者第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被包封或部分地包封在第一合成颗粒中。在一些实施方案中,多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被固定或部分地固定在第二合成颗粒上,或者多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被包封或部分地包封在第二合成颗粒中。第一合成颗粒和/或第二合成颗粒可以是可破坏的(例如,可破坏的水凝胶颗粒)。第一合成颗粒和/或第二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)。第一合成颗粒和/或第二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、硅酮及其任何组合。在一些实施方案中,第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含接头官能团,第一合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由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。在一些实施方案中,多于一个一种或更多种条形码化对照核酸的每一种条形码化对照核酸包含接头官能团,第二合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由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。

[0224] 测序状态确定

[0225] 多于一个条形码化核酸分子或其产物的多于一个测序读段的每一个可以包含(1)分子标记序列,和/或(2)核酸靶的子序列。多于一个条形码化对照核酸分子或其产物的多于一个测序读段的每一个可以包含(1)对照标记序列,和/或(2)对照核酸分子的子序列。该方法可以包括:确定测序数据中一个或多个对照核酸文库成员的测序状态。测序数据中一个或多个对照核酸文库成员的测序状态可以是饱和测序或测序不足。在一些实施方案中,饱和测序状态由一个或多个对照核酸文库成员具有处于或大于预定饱和阈值的测序读段数目确定;并且测序不足状态由一个或多个对照核酸文库成员具有小于预定饱和阈值的测序读段数目确定。预定饱和阈值可以是一种或更多种条形码化对照核酸的预定拷贝数的至少约1.05倍大的数字(例如,至少1.05倍、至少1.1倍、至少1.2倍、至少1.3倍、至少1.5倍、至少1.7倍、至少2倍、至少2.5倍、至少3倍、至少3.5倍、至少4倍、至少5倍、至少7倍、至少10倍、至少20倍或至少50倍大,或这些值中的任何两个值之间的数字或范围)。本文公开的方法可以包括:如果测序数据中一个或多个对照核酸文库成员的测序状态是测序不足状态,则重复获得测序数据的步骤,直到测序数据中一个或多个对照核酸文库成员的测序状态是饱和测序状态。

[0226] 工作流程失败确定

[0227] 本文提供了确定工作流程失败的存在的方法,其中工作流程失败包括对核酸靶的拷贝进行条形码化的失败和/或测序文库产生的失败。对核酸靶的拷贝进行条形码化的失败的存在可以通过一个或多个对照核酸文库成员的测序读段与一个或多个核酸靶文库成员的测序读段的比率超过预定条形码化阈值来确定。该方法可以包括:基于一个或多个核酸靶文库成员的多于一个测序读段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一个或多个核酸靶文库成员或其产物关联的具有不同序列的第一分子标记、其互补物或其组合的数目来确定样品中核酸靶的拷贝数。对核酸靶的拷贝进行条形码化的失败的存在可以通过一个或多个条形码化对照核酸的预定拷贝数与样品中核酸靶的拷贝数的比率超过预定条形码化阈值来确定。预定条形码化阈值可以是以下、可以是约以下、可以是至少以下或者可以是至多以下:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、60、70、80、90、100、110、120、128、130、140、150、160、170、180、190、200、210、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、610、620、630、640、650、660、670、680、690、700、710、720、730、740、750、760、770、780、790、800、810、820、830、840、850、860、870、880、890、900、910、920、930、940、950、960、970、980、990、1000、1050、1100、1150、1200、1250、1300、1350、1400、1450、1500、1550、1600、1650、1700、1750、1800、1850、1900、1950、2000、2050、2100、2200、2250、2300、2350、2400、2450、2500、2550、2600、2650、2700、2750、2800、2850、2900、2950、3000、3050、4000、4500、5000,或这些值中的任何两个值之间的数字或范围。

[0228] 本文提供的方法可以包括:获得包含预定数目的一个或多个加标文库成员的多于一个测序读段的测序数据。测序文库产生的失败的存在可以通过一个或多个加标文库成员的预定数目的测序读段与一个或多个对照核酸文库成员的测序读段的比率超过预

定文库产生阈值来确定。预定文库产生阈值可以是以下、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、25、30、35、40、45、50、60、70、80、90、100、110、120、128、130、140、150、160、170、180、190、200、210、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、610、620、630、640、650、660、670、680、690、700、710、720、730、740、750、760、770、780、790、800、810、820、830、840、850、860、870、880、890、900、910、920、930、940、950、960、970、980、990、1000、1050、1100、1150、1200、1250、1300、1350、1400、1450、1500、1550、1600、1650、1700、1750、1800、1850、1900、1950、2000、2050、2100、2200、2250、2300、2350、2400、2450、2500、2550、2600、2650、2700、2750、2800、2850、2900、2950、3000、3050、4000、4500、5000，或这些值中的任何两个值之间的数字或范围。一个或多个加标文库成员可以与样品的基因组序列不同源。一个或多个加标文库成员可以与物种的基因组序列同源。该物种可以是非哺乳动物物种。非哺乳动物物种可以是噬菌体物种（例如，T7噬菌体、PhiX噬菌体或其任何组合）。

[0229] 试剂盒

[0230] 在一些实施方案中提供了试剂盒。试剂盒可以包含：多于一个一种或更多种条形码化对照核酸，其中一种或更多种条形码化对照核酸的每一种的拷贝数是预定的。多于一个一种或更多种条形码化对照核酸可以与第二固体支持物关联。试剂盒可以包含：第一多于一个寡核苷酸条形码，其中多于一个寡核苷酸条形码的每一个包含分子标记和靶结合区。多于一个寡核苷酸条形码中的至少10种可以包含不同的分子标记序列。第一多于一个寡核苷酸条形码可以与第一固体支持物关联。条形码化对照核酸的每一种可以包含第一通用序列、对照标记和靶结合区中的一种或更多种。在一些实施方案中，多于一个一种或更多种条形码化对照核酸可以包含至少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、1,000种、2,000种、5,000种或更多种不同的条形码化对照核酸。一种或更多种条形码化对照核酸可以与至少约2种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种或这些值中的任何两个值之间的数字或范围的不同的核酸靶至少约70%同源。一种或更多种条形码化对照核酸可以包含持家基因的序列。一种或更多种条形码化对照核酸可以与物种的基因组序列同源。该物种可以是非哺乳动物物种。非哺乳动物物种可以是噬菌体物种（例如，T7噬菌体、PhiX噬菌体或其任何组合）。

[0231] 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。试剂盒可以包含逆转录酶（例如，病毒逆转录酶，诸如例如，鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶)。该试剂盒可以包含：包含靶结合区或其一部分的模板转换寡核苷酸。模板转换寡核苷酸可以包含一个或多个3'核糖核苷酸（例如，三个3'核糖核苷酸）。3'核糖核苷酸可以包含鸟嘌呤。试剂盒可以包含以下中的一种或更多种：乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砜(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或它们的任何组合。试剂盒可以包含：缺乏5'至3'外切核酸酶活性和3'至5'外切核酸酶活性中的至少一种的DNA聚合酶（例如，Klenow片段）。试剂盒可以包含：缓冲液、筒或两者。试剂盒可以包含：一种或更多种用于逆转录反应和/或扩增反应的试剂。

[0232] 第一多于一个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。第一多于一个寡核苷酸条形码各自可以包含细胞标记。第一多于一个寡核苷酸条形码的每一种细胞标记可以包含至少6个核苷酸。与同一第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码可以包含相同的细胞标记。与不同的第一固体支持物关联的第一多于一个寡核苷酸条形码中的寡核苷酸条形码可以包含不同的细胞标记。第一固体支持物可以包括第一合成颗粒、第一平坦表面或其组合。第二固体支持物可以包括第二合成颗粒、第二平坦表面或其组合。在一些实施方案中,第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被固定或部分地固定在第一合成颗粒上,或者第一多于一个寡核苷酸条形码中的至少一种寡核苷酸条形码被包封或部分地包封在第一合成颗粒中。在一些实施方案中,多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被固定或部分地固定在第二合成颗粒上,或者多于一个一种或更多种条形码化对照核酸中的至少一种条形码化对照核酸被包封或部分地包封在第二合成颗粒中。第一合成颗粒和/或第二合成颗粒可以是可破坏的(例如,可破坏的水凝胶颗粒)。第一合成颗粒和/或第二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)。第一合成颗粒和/或第二合成颗粒可以包含选自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。在一些实施方案中,第一多于一个寡核苷酸条形码的每一个寡核苷酸条形码包含接头官能团,第一合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自自由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。在一些实施方案中,多于一个一种或更多种条形码化对照核酸的每一种条形码化对照核酸包含接头官能团,第二合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自自由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。

[0233] 术语

[0234] 在至少一些先前描述的实施方案中,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之内。

[0235] 本领域的技术人员将理解,对于本文公开的这个和其他过程和方法,在该过程和方法中执行的功能可以以不同的顺序实现。此外,所概述的步骤和操作仅作为示例提供,并且该步骤和操作中的一些可以是任选的,组合成较少的步骤和操作,或者扩展成另外的步骤和操作,而不偏离所公开的实施方案的本质。

[0236] 关于本文中基本上任何复数和/或单数术语的使用,在对于上下文和/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”。

[0237] 本领域技术人员将理解,一般来说,本文使用的术语,并且尤其是所附权利要求(例如,所附权利要求的主体)中的术语,通常意在作为“开放式”术语(例如,术语“包括(including)”应解释为“包括但不限于(including but not limited to)”,术语“具有(having)”应解释为“具有至少(having at least)”,术语“包括(includes)”应解释为“包括但不限于(includes but is not limited to)”,等等)。本领域技术人员将进一步理解,如果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此类介绍的权利要求陈述的具体权利要求限制为包含仅一种此类陈述的实施方案,甚至当同一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,此类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。

[0238] 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。

[0239] 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的组合。任何列出的范围都可以很容易地被识别为充分描述并使相同的范围能被分解为至少相等的一半、三分之一、四分之一、五分之一、十分之一等。作为非限制性实例,本文讨论的每个范围可以容易地分解为下三分之一、中三分之一和上三分之一等。如本领域技术人员还将理解的,所有语言,诸如“多达”、“至少”、“大于”、“小于”等包括所陈述的数字,并且指可以随后分解为如上文讨论的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个个体的成员。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。

[0240] 从前述内容,应当理解,本文出于说明的目的已经描述了本公开内容的各种实施

方案,并且可以在不脱离本公开内容的范围和精神的情况下进行各种修改。因此,本文公开的各种实施方案并不旨在进行限制,真正的范围和精神由以下权利要求来指示。

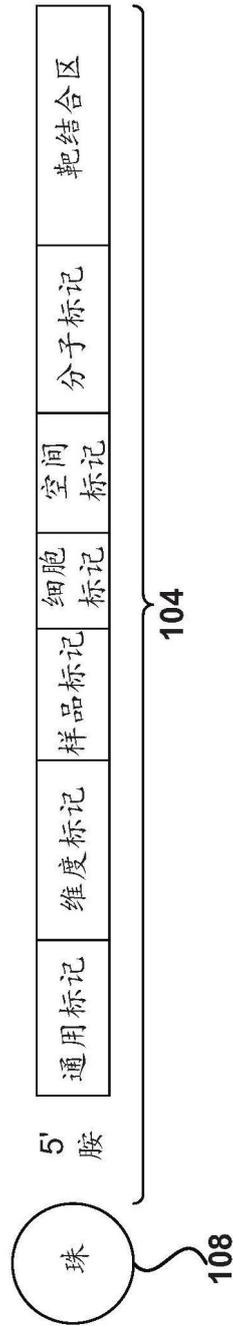


图1

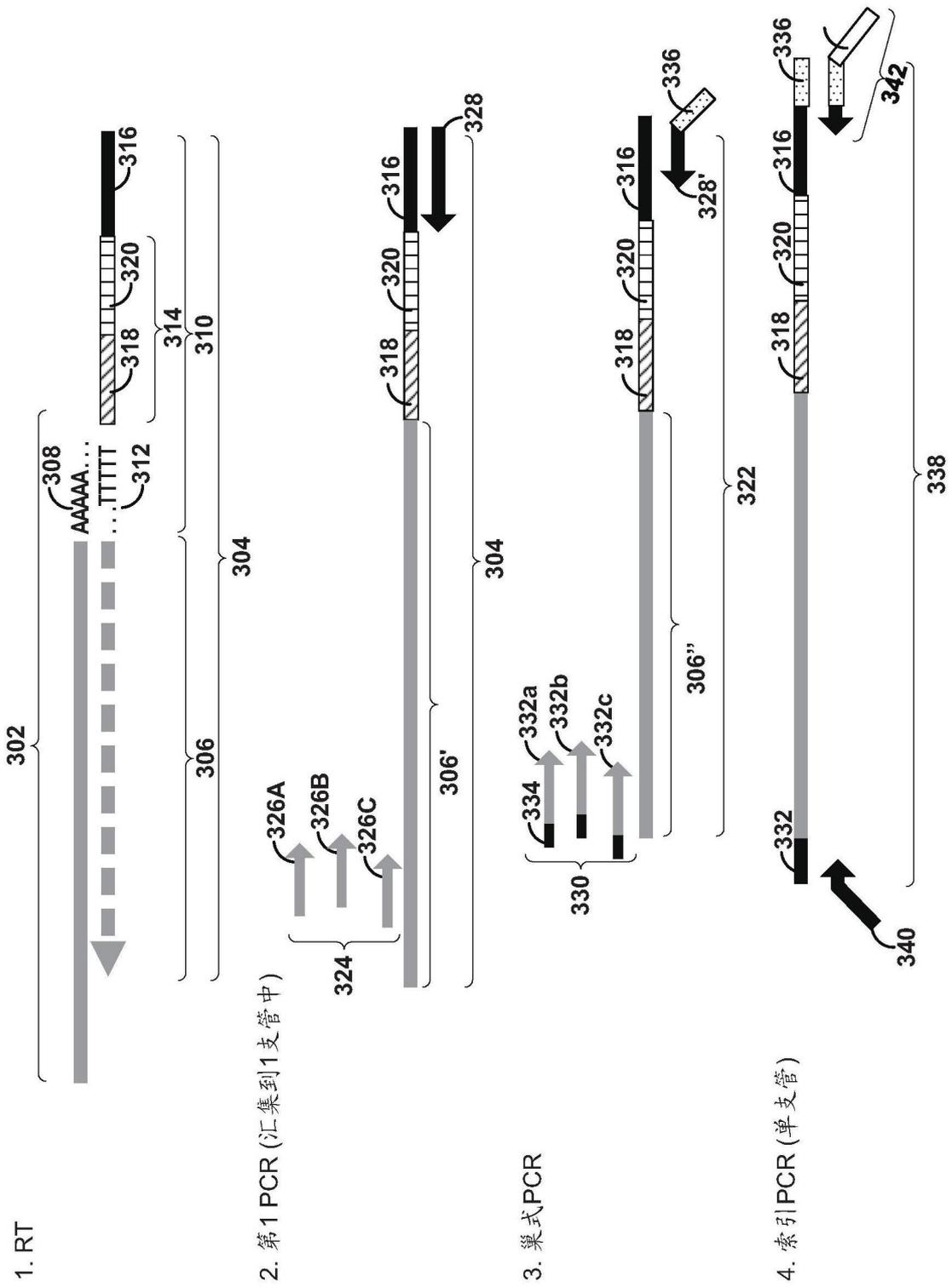


图3

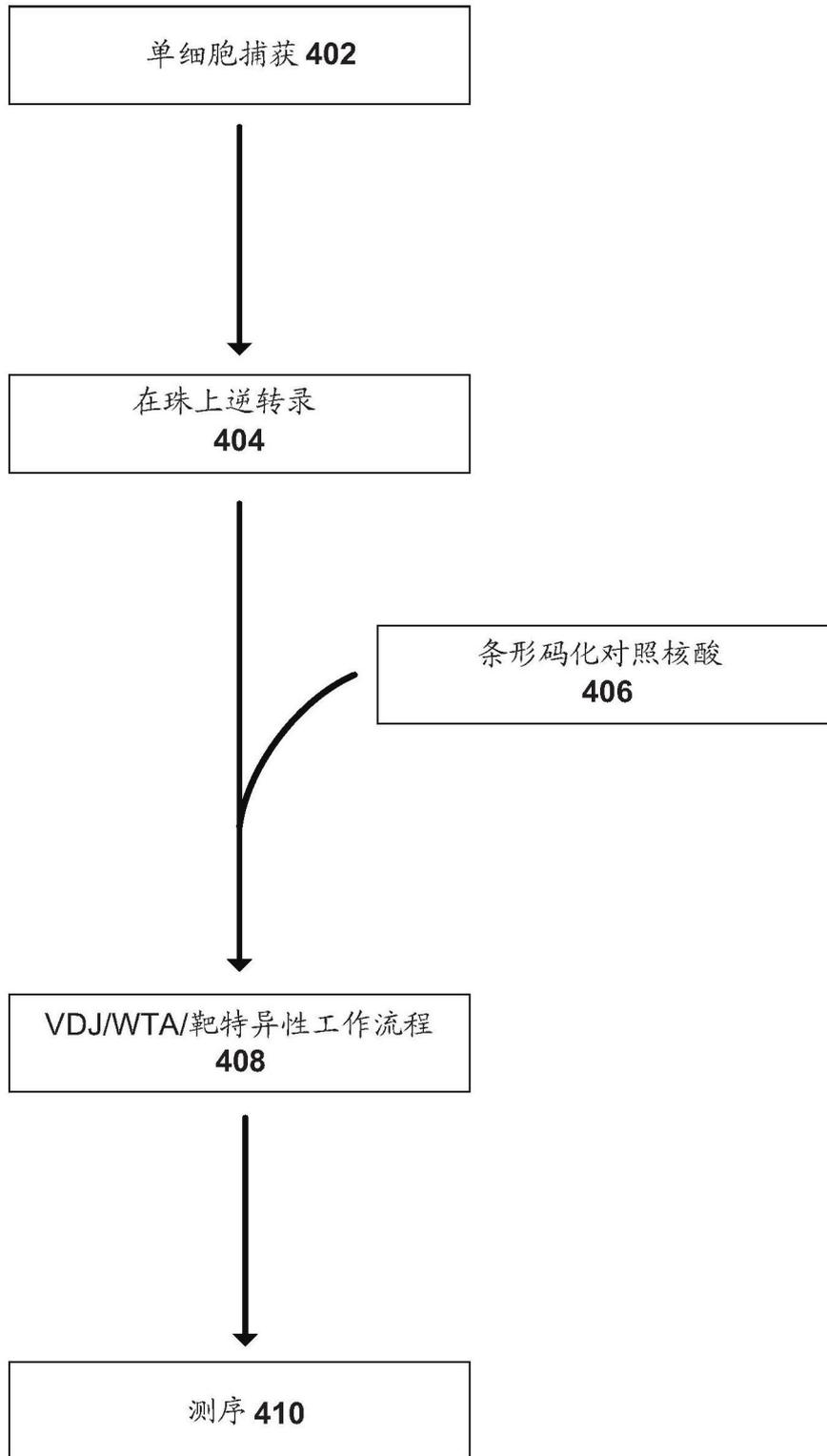


图4

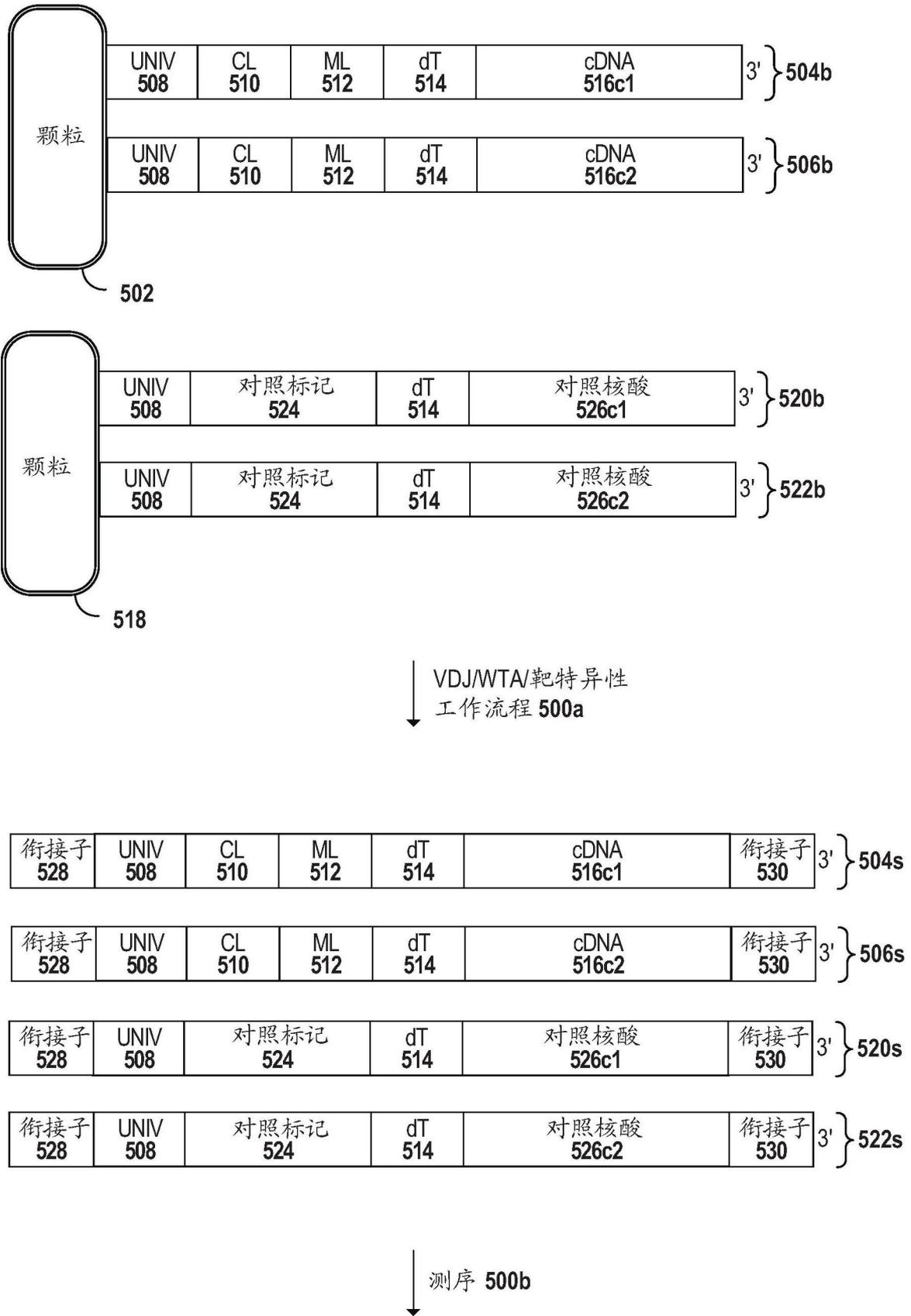


图5