(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112805389 B
(45)授权公告日 2024.08.06

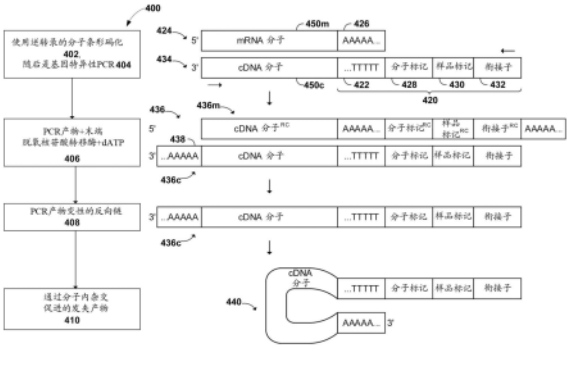
(21)申请号 201980065310.2
(22)申请日 2019.09.30
(65)同一申请的已公布的文献号
申请公布号 CN 112805389 A
(43)申请公布日 2021.05.14
(30)优先权数据
62/739,795 2018.10.01 US
(85)PCT国际申请进入国家阶段日
2021.04.01
(86)PCT国际申请的申请数据
PCT/US2019/053868 2019.09.30
(87)PCT国际申请的公布数据
WO2020/072380 EN 2020.04.09
(73)专利权人 贝克顿迪金森公司
地址 美国新泽西州
(72)发明人 克里斯蒂娜·张 艾琳·夏姆
(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 刘晓杰 武晶晶
(51)Int.Cl.
C12Q 1/6806(2006.01)
(56)对比文件
CA 2961210 A1,2016.03.24
US 2016312276 A1,2016.10.27
WO 2017079593 A1,2017.05.11
SIMONE PICELLI.Single-cell RNA-sequencing: The future of genome biology is now.《RNA BIOLOGY》.2016,第14卷(第5期),637-650.
审查员 吕健
权利要求书9页 说明书63页
序列表1页 附图35页

(54)发明名称

确定5'转录物序列

(57)摘要

本文的公开内容包括用于对样品中的核酸靶进行标记的系统、方法、组合物和试剂盒。在一些实施方案中,核酸靶(例如,mRNA)最初在3'末端进行条形码化(barcoded),并且随后在模板转换反应和分子间和/或分子内杂交和延伸之后在5'末端进行条形码化。在一些实施方案中,提供了基于5'-和基于3'-的基因表达谱分析的方法。在一些实施方案中,还提供了免疫组库谱分析(immune repertoire profiling)方法。



CN 112805389 B

1.一种用于对样品中的核酸靶进行标记的方法,所述方法包括:

使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与所述核酸靶杂交的靶结合区;

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存在下,使与所述核酸靶的拷贝杂交的所述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,每个所述条形码化核酸分子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体;

使每个条形码化核酸分子的所述靶结合区的互补体与以下的靶结合区杂交:

(i)所述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本身,和/或

(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;以及

使所述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,每个所述延伸的条形码化核酸分子包含所述第一分子标记和第二分子标记。

2.根据权利要求1所述的方法,所述方法包括:基于与所述多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定所述样品中所述核酸靶的拷贝数。

3.一种用于确定样品中核酸靶的数量的方法,所述方法包括:

使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与所述核酸靶杂交的靶结合区;

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存在下,使与所述核酸靶的拷贝杂交的所述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,每个所述条形码化核酸分子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、靶结合区和所述靶结合区的互补体;

使每个条形码化核酸分子的所述靶结合区的互补体与以下的靶结合区杂交:

(i)所述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本身,和/或

(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,以产生多于一个延伸的条形码化核酸分子,每个所述延伸的条形码化核酸分子包含所述第一分子标记和第二分子标记;以及

基于与所述多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定所述样品中所述核酸靶的拷贝数。

4.根据权利要求2-3中任一项所述的方法,

所述方法包括扩增所述多于一个延伸的条形码化核酸分子以产生多于一个单标记的核酸分子,每个所述单标记的核酸分子包含所述第一分子标记或所述第二分子标记,

其中确定所述样品中所述核酸靶的拷贝数包括:基于与所述多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定所述样品中所述核酸靶的拷贝数。

5.根据权利要求4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

所述多于一个单标记的核酸分子关联的具有不同序列的第一分子标记的数量,确定所述样品中所述核酸靶的拷贝数。

6.根据权利要求2-3中任一项所述的方法,所述方法

包括扩增所述多于一个延伸的条形码化核酸分子以产生所述多于一个延伸的条形码化核酸分子的拷贝,

其中确定所述样品中所述核酸靶的拷贝数包括:基于以下的(i)和/或(ii)确定所述样品中所述核酸靶的拷贝数:(i)与所述多于一个延伸的条形码化核酸分子或其产物的拷贝关联的具有不同序列的第一分子标记的数量,和/或(ii)与所述多于一个延伸的条形码化核酸分子或其产物的拷贝关联的具有不同序列的第二分子标记的数量。

7.一种确定样品中核酸靶的数量的方法,所述方法包括:

使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与所述核酸靶杂交的靶结合区;

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存在下,使与所述核酸靶的拷贝杂交的所述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,每个所述条形码化核酸分子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、靶结合区和所述靶结合区的互补体;

使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:

(i)所述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本身,和/或

(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,以产生多于一个延伸的条形码化核酸分子,每个所述延伸的条形码化核酸分子包含所述第一分子标记和第二分子标记;

扩增所述多于一个延伸的条形码化核酸分子以产生多于一个单标记的核酸分子,每个所述单标记的核酸分子包含所述第一分子标记或所述第二分子标记;以及

基于与所述多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定所述样品中所述核酸靶的拷贝数。

8.根据权利要求7所述的方法,所述方法包括基于与所述多于一个单标记的核酸分子关联的具有不同序列的第一分子标记的数量,确定所述样品中所述核酸靶的拷贝数。

9.根据权利要求1-8中任一项所述的方法,所述方法包括在使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交之前,使所述多于一个条形码化核酸分子变性:(i)所述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本身,和/或(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。

10.根据权利要求4-9中任一项所述的方法,所述方法包括在扩增所述多于一个延伸的条形码化核酸分子之前,使所述多于一个延伸的条形码化核酸分子变性。

11.根据权利要求7-9中任一项所述的方法,

其中确定所述核酸靶的拷贝数包括基于与所述多于一个单标记的核酸分子中的单标记的核酸分子关联的具有不同序列的第二分子标记的数量来确定所述样品中所述多于一个核酸靶中的每一个的拷贝数,所述多于一个单标记的核酸分子包含所述多于一个核酸靶中的每一个的序列。

12.根据权利要求11所述的方法,其中确定所述核酸靶的拷贝数包括基于与所述多于一个单标记的核酸分子中的单标记的核酸分子关联的具有不同序列的第一分子标记的数量来确定所述样品中所述多于一个核酸靶中的每一个的拷贝数,所述多于一个单标记的核酸分子包含所述多于一个核酸靶中的每一个的序列。

13.根据权利要求12所述的方法,其中所述多于一个核酸靶中的每一个的序列包括所述多于一个核酸靶中的每一个的子序列。

14.根据权利要求1-13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中的所述核酸靶的序列包括所述核酸靶的子序列。

15.根据权利要求1-14中任一项所述的方法,其中在使所述多于一个条形码化核酸分子的3'-末端延伸后,使所述第一分子标记与所述第二分子标记杂交。

16.根据权利要求1-15中任一项所述的方法,其中所述延伸的条形码化核酸分子各自包含所述第一分子标记、所述第二分子标记、所述靶结合区和所述靶结合区的互补体。

17.根据权利要求1-16中任一项所述的方法,其中所述靶结合区的互补体与所述靶结合区的一部分互补。

18.根据权利要求1-17中任一项所述的方法,其中所述靶结合区包含基因特异性序列和/或多(dT)序列。

19.根据权利要求1-18中任一项所述的方法,其中使条形码化核酸分子的所述靶结合区的互补体与所述条形码化核酸分子本身的靶结合区杂交包括使条形码化核酸分子内的所述靶结合区和所述靶结合区的互补体进行分子内杂交以形成茎环。

20.根据权利要求19所述的方法,其中所述第二分子标记是所述第一分子标记的互补体。

21.根据权利要求1-20中任一项所述的方法,其中使条形码化核酸分子的所述靶结合区的互补体与所述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所述靶结合区的互补体与所述多于一个寡核苷酸条形码中的寡核苷酸条形码的所述靶结合区进行分子间杂交。

22.根据权利要求21所述的方法,其中所述第二分子标记不同于所述第一分子标记,并且其中所述第二分子标记不是所述第一分子标记的互补体。

23.根据权利要求21-22中任一项所述的方法,所述方法包括使与所述条形码化核酸分子的所述靶结合区的互补体杂交的所述寡核苷酸条形码的3'末端延伸,以产生多于一个延伸的条形码化核酸分子,每个所述延伸的条形码化核酸分子包含所述第一分子标记的互补体和第二分子标记。

24.根据权利要求23所述的方法,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体。

25.根据权利要求1-24中任一项所述的方法,其中使条形码化核酸分子的所述靶结合区的互补体与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的靶结合区杂交包括使条形码化核酸分子的所述靶结合区的互补体与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的所述靶结合区进行分子间杂交。

26.根据权利要求25所述的方法,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,并且其中所述第二分子标记不是所述第一分子标记的互补体。

27.根据权利要求1-26中任一项所述的方法,其中所述逆转录酶能够发挥末端转移酶活性。

28.根据权利要求1-27中任一项所述的方法,其中所述模板转换寡核苷酸包含:

(a)一个或更多个3'核糖核苷酸,

(b)三个3'核糖核苷酸,或

(c)三个3'核糖核苷酸,其中所述3'核糖核苷酸包括鸟嘌呤。

29.根据权利要求1-28中任一项所述的方法,其中所述逆转录酶包括:

(a)病毒逆转录酶,或

(b)病毒逆转录酶,其中所述病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶或莫洛尼鼠白血病病毒(MMLV)逆转录酶。

30.根据权利要求1-29中任一项所述的方法,其中所述样品包括:

(a)单细胞,

(b)单细胞,其中所述单细胞是免疫细胞,或

(c)单细胞,其中所述单细胞是免疫细胞,并且其中所述免疫细胞是B细胞或T细胞。

31.根据权利要求1-30中任一项所述的方法,其中所述样品包括多于一个细胞、多于一个单细胞、组织、肿瘤样品,或它们的任何组合。

32.根据权利要求30-31中任一项所述的方法,其中单细胞包括循环肿瘤细胞。

33.根据权利要求1-32中任一项所述的方法,其中每个寡核苷酸条形码包含第一通用序列。

34.根据权利要求1-33中任一项所述的方法,其中所述多于一个延伸的条形码化核酸分子包含第一通用序列和所述第一通用序列的互补体。

35.根据权利要求33-34中任一项所述的方法,其中扩增所述多于一个延伸的条形码化核酸分子以产生所述多于一个延伸的条形码化核酸分子的拷贝包括使用能够与所述第一通用序列或其互补体杂交的引物。

36.根据权利要求33-34中任一项所述的方法,其中扩增所述多于一个延伸的条形码化核酸分子以产生多于一个单标记的核酸分子包括使用能够与所述第一通用序列或其互补体杂交的引物,和扩增引物。

37.根据权利要求36所述的方法,其中所述扩增引物是:

(a)靶特异性引物,或

(b)靶特异性引物,其中所述靶特异性引物与免疫受体、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和/或免疫受体的可变区和多样性区的连接区特异性杂交。

38.根据权利要求37所述的方法,其中所述免疫受体是:

(a)T细胞受体(TCR)和/或B细胞受体(BCR),或

(b)T细胞受体(TCR)和/或B细胞受体(BCR),并且其中

所述TCR包括TCRα链、TCRβ链、TCRγ链、TCRδ链,或它们的任何组合;并且

所述BCR包括BCR重链和/或BCR轻链。

39.根据权利要求1-38中任一项所述的方法,其中:

(a)使所述多于一个条形码化核酸分子的3'-末端延伸包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使所述多于一个条形码化核酸

分子的3’-末端延伸,或

(b)使所述多于一个条形码化核酸分子的3’-末端延伸包括使用缺乏5’至3’核酸外切酶活性和3’至5’核酸外切酶活性中的至少一种的DNA聚合酶使所述多于一个条形码化核酸分子的3’-末端延伸,其中所述DNA聚合酶包括Klenow片段。

40.根据权利要求1-39中任一项所述的方法,所述方法包括获得所述多于一个延伸的条形码化核酸分子或其产物的序列信息。

41.根据权利要求40所述的方法,其中获得所述序列信息包括将测序衔接子附接到所述多于一个延伸的条形码化核酸分子或其产物。

42.根据权利要求40所述的方法,其中获得所述序列信息包括将测序衔接子附接到所述多于一个单标记的核酸分子或其产物。

43.根据权利要求40-42中任一项所述的方法,其中获得所述序列信息包括获得单细胞的BCR轻链和BCR重链的序列信息。

44.根据权利要求43所述的方法,其中所述BCR轻链和所述BCR重链的序列信息包括所述BCR轻链和/或所述BCR重链的互补决定区1(CDR1)、CDR2、CDR3,或它们的任何组合的序列。

45.根据权利要求43-44中任一项所述的方法,所述方法包括基于获得的序列信息将所述单细胞的BCR轻链和BCR重链配对。

46.根据权利要求43-45中任一项所述的方法,其中所述样品包含多于一个单细胞,所述方法包括基于获得的序列信息将至少50%的所述单细胞的BCR轻链和BCR重链配对。

47.根据权利要求40-46中任一项所述的方法,其中获得所述序列信息包括获得单细胞的TCRα链和TCRβ链的序列信息。

48.根据权利要求47所述的方法,其中所述TCRα链和所述TCRβ链的序列信息包括所述TCRα链和/或所述TCRβ链的互补决定区1(CDR1)、CDR2、CDR3,或它们的任何组合的序列。

49.根据权利要求47-48中任一项所述的方法,所述方法包括基于获得的序列信息将所述单细胞的TCRα链和TCRβ链配对。

50.根据权利要求47-49中任一项所述的方法,其中所述样品包含多于一个单细胞,所述方法包括基于获得的序列信息将至少50%的所述单细胞的TCRα链和TCRβ链配对。

51.根据权利要求40-50中任一项所述的方法,其中获得所述序列信息包括获得单细胞的TCRγ链和TCRδ链的序列信息。

52.根据权利要求51所述的方法,其中所述TCRγ链和所述TCRδ链的序列信息包括所述TCRγ链和/或所述TCRδ链的互补决定区1(CDR1)、CDR2、CDR3,或它们的任何组合的序列。

53.根据权利要求51-52中任一项所述的方法,所述方法包括基于获得的序列信息将所述单细胞的TCRγ链和TCRδ链配对。

54.根据权利要求51-53中任一项所述的方法,其中所述样品包含多于一个单细胞,所述方法包括基于获得的序列信息将至少50%的所述单细胞的TCRγ链和TCRδ链配对。

55.根据权利要求1-54中任一项所述的方法,其中所述靶结合区的互补体包括所述靶结合区的反向互补序列和/或所述靶结合区的互补序列。

56.根据权利要求1-55中任一项所述的方法,其中所述分子标记的互补体包括所述分子标记的反向互补序列和/或所述分子标记的互补序列。

57.根据权利要求1-56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包括条形码化脱氧核糖核酸(DNA)分子和/或条形码化核糖核酸(RNA)分子。

58.根据权利要求1-57中任一项所述的方法,其中:

(a)所述核酸靶包括核酸分子,

(b)所述核酸靶包括核酸分子,并且其中所述核酸分子包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、包含多(A)尾的RNA,或它们的任何组合,或

(c)所述核酸靶包括核酸分子,并且其中所述核酸分子包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、包含多(A)尾的RNA,或它们的任何组合,并且其中所述mRNA编码免疫受体。

59.根据权利要求58所述的方法,其中所述核酸靶包含细胞组分结合试剂。

60.根据权利要求58所述的方法,其中所述核酸分子与细胞组分结合试剂关联。

61.根据权利要求60所述的方法,所述方法包括使所述核酸分子和所述细胞组分结合试剂解离。

62.根据权利要求1-61中任一项所述的方法,其中所述多于一个寡核苷酸条形码中的至少10个包含不同的分子标记序列。

63.根据权利要求1-62中任一项所述的方法,其中所述多于一个寡核苷酸条形码的每个分子标记包含至少6个核苷酸。

64.根据权利要求1-63中任一项所述的方法,其中:

(a)所述多于一个寡核苷酸条形码与固体支持物关联,

(b)所述多于一个寡核苷酸条形码与固体支持物关联,并且其中与相同的固体支持物关联的所述多于一个寡核苷酸条形码各自包含相同的样品标记,或

(c)所述多于一个寡核苷酸条形码与固体支持物关联,并且其中与相同的固体支持物关联的所述多于一个寡核苷酸条形码各自包含相同的样品标记,并且其中所述多于一个寡核苷酸条形码的每个样品标记包含至少6个核苷酸。

65.根据权利要求1-64中任一项所述的方法,其中:

(a)所述多于一个寡核苷酸条形码各自包含细胞标记,或

(b)所述多于一个寡核苷酸条形码各自包含细胞标记,并且其中所述多于一个寡核苷酸条形码的每个细胞标记包含至少6个核苷酸。

66.根据权利要求65所述的方法,其中与相同的固体支持物关联的寡核苷酸条形码包含相同的细胞标记。

67.根据权利要求65所述的方法,其中与不同的固体支持物关联的寡核苷酸条形码包含不同的细胞标记。

68.根据权利要求65-67中任一项所述的方法,其中

(a)所述多于一个延伸的条形码化核酸分子各自包含细胞标记和所述细胞标记的互补体,或

(b)所述多于一个延伸的条形码化核酸分子各自包含细胞标记和所述细胞标记的互补体,并且其中所述细胞标记的互补体包含所述细胞标记的反向互补序列和/或所述细胞标记的互补序列。

69.根据权利要求1-68中任一项所述的方法,所述方法包括在乙二醇、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,或它们的任何组合中的一种或更多种的存在下,使与所述核酸靶的拷贝杂交的所述多于一个寡核苷酸条形码延伸。

70.根据权利要求64或66-67中任一项所述的方法,其中所述固体支持物包括合成颗粒或平坦表面。

71.根据权利要求1-70中任一项所述的方法,其中所述样品包含单细胞,所述方法包括使包含所述多于一个寡核苷酸条形码的合成颗粒与所述样品中的所述单细胞关联。

72.根据权利要求71所述的方法,所述方法包括:

(a)在使所述合成颗粒与所述单细胞关联后裂解所述单细胞,或

(b)在使所述合成颗粒与所述单细胞关联后裂解所述单细胞,其中裂解所述单细胞包括加热所述样品、使所述样品与去污剂接触、改变所述样品的pH,或它们的任何组合。

73.根据权利要求71-72中任一项所述的方法,其中所述合成颗粒和所述单细胞在同一孔中。

74.根据权利要求71-72中任一项所述的方法,其中所述合成颗粒和所述单细胞在同一液滴中。

75.根据权利要求71-74中任一项所述的方法,其中所述多于一个寡核苷酸条形码中的至少一个被固定或部分地固定在所述合成颗粒上,或者所述多于一个寡核苷酸条形码中的至少一个被包封或部分地包封在所述合成颗粒中。

76.根据权利要求71-75中任一项所述的方法,其中所述合成颗粒是可破坏的。

77.根据权利要求71-76中任一项所述的方法,其中所述合成颗粒:

(a)包括珠,或

(b)包括珠,并且其中所述珠包括

琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合;

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、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合;或

可破坏的水凝胶颗粒。

78.根据权利要求71-77中任一项所述的方法,其中

(a)所述多于一个寡核苷酸条形码中的每一个包含接头官能基团,

所述合成颗粒包含固体支持物官能基团,并且

所述支持物官能基团和所述接头官能基团彼此关联,或

(b)所述多于一个寡核苷酸条形码中的每一个包含接头官能基团,

所述合成颗粒包含固体支持物官能基团,并且

所述支持物官能基团和所述接头官能基团彼此关联,

并且所述接头官能基团和所述支持物官能基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,以及它们的

任何组合。

79.一种试剂盒,所述试剂盒包含:

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中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个包含不同的分子标记序列;

逆转录酶;

模板转换寡核苷酸,所述模板转换寡核苷酸包含所述靶结合区或其一部分;和

DNA聚合酶,所述DNA聚合酶缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种。

80.根据权利要求79所述的试剂盒,其中所述DNA聚合酶包括Klenow片段。

81.根据权利要求79-80中任一项所述的试剂盒,其中:

(a)所述逆转录酶包括病毒逆转录酶,或

(b)所述逆转录酶包括病毒逆转录酶,并且其中所述病毒逆转录酶包括鼠白血病病毒(MLV)逆转录酶或莫洛尼鼠白血病病毒(MMLV)逆转录酶。

82.根据权利要求79-81中任一项所述的试剂盒,其中:

(a)所述模板转换寡核苷酸包含一个或更多个3'核糖核苷酸,

(b)所述模板转换寡核苷酸包含三个3'核糖核苷酸,或

(c)所述模板转换寡核苷酸包含三个3'核糖核苷酸,并且其中所述3'核糖核苷酸包括鸟嘌呤。

83.根据权利要求79-82中任一项所述的试剂盒,所述试剂盒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。

84.根据权利要求79-83中任一项所述的试剂盒,所述试剂盒包含缓冲液、药筒、一种或更多种用于逆转录反应的试剂、一种或更多种用于扩增反应的试剂,或它们的组合。

85.根据权利要求79-84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、寡聚(dT)序列、随机多聚体,或它们的任何组合。

86.根据权利要求79-85中任一项所述的试剂盒,其中所述寡核苷酸条形码包含相同的样品标记和/或相同的细胞标记。

87.根据权利要求86所述的试剂盒,其中所述多于一个寡核苷酸条形码的每个样品标记、细胞标记和/或分子标记包含至少6个核苷酸。

88.根据权利要求79-87中任一项所述的试剂盒,其中所述多于一个寡核苷酸条形码中的至少一个被固定或部分地固定在合成颗粒上,和/或所述多于一个寡核苷酸条形码中的至少一个被包封或部分地包封在合成颗粒中。

89.根据权利要求88所述的试剂盒,其中所述合成颗粒是可破坏的。

90.根据权利要求89所述的试剂盒,其中:

(a)所述合成颗粒包括珠,或

(b)所述合成颗粒包括珠,并且其中所述珠包括

琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物

素微珠、抗荧光染料微珠,或它们的任何组合;

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、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合;或

可破坏的水凝胶颗粒。

91.根据权利要求88-90中任一项所述的试剂盒,其中:

(a)所述多于一个寡核苷酸条形码中的每一个包含接头官能基团,

所述合成颗粒包含固体支持物官能基团,并且

所述支持物官能基团和所述接头官能基团彼此关联;或

(b)所述多于一个寡核苷酸条形码中的每一个包含接头官能基团,

所述合成颗粒包含固体支持物官能基团,并且

所述支持物官能基团和所述接头官能基团彼此关联,

并且所述接头官能基团和所述支持物官能基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,以及它们的任何组合。

确定5’转录物序列

[0001] 相关申请

[0002] 本申请根据35U.S.C.§119(e),要求2018年10月1日提交的美国临时专利申请序列第62/739,795号的权益,该相关申请的内容出于所有目的通过引用以其整体并入本文。

[0003] 对序列表的引用

[0004] 本申请连同电子格式的序列表一起提交。序列表被提供为题为68EB_298703_WO的文件,创建于2019年9月30日,大小是4.0千字节。电子格式的序列表的信息通过引用以其整体并入本文。

[0005] 背景

[0006] 领域

[0007] 本公开内容总体上涉及分子生物学领域,并且具体涉及使用分子条形码化(molecular barcoding)的多组学(multiomics)分析。

[0008] 相关技术描述

[0009] 分子条形码化的方法和技术可用于单细胞转录组学分析,包括使用例如逆转录、聚合酶链式反应(PCR)扩增,和下一代测序(NGS)来解密基因表达谱以确定细胞的状态。分子条形码化还可用于单细胞蛋白质组学分析。对于在5’末端和3’末端之一或二者上对核酸靶进行分子条形码化的方法和技术存在需求。

[0010] 概述

[0011] 本文的公开内容包括用于对样品中的核酸靶进行标记的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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;以及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,每个延伸的条形码化核酸分子包含第一分子标记和第二分子标记。该方法可以包括基于与多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定样品中核酸靶的拷贝数。

[0012] 本文的公开内容包括用于确定样品中核酸靶的数量的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合

区的互补体与以下的靶结合区杂交:(i)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,和/或(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,每个条形码化核酸分子包含第一分子标记和第二分子标记;以及基于与多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定样品中核酸靶的拷贝数。

[0013] 该方法可以包括扩增多于一个延伸的条形码化核酸分子,以产生多于一个单标记的核酸分子,每个单标记的核酸分子包含第一分子标记或第二分子标记,其中确定样品中核酸靶的拷贝数包括:基于与多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定样品中核酸靶的拷贝数。在一些实施方案中,确定样品中核酸靶的拷贝数包括:基于与多于一个单标记的核酸分子关联的具有不同序列的第一分子标记的数量,确定样品中核酸靶的拷贝数。该方法可以包括扩增多于一个延伸的条形码化核酸分子,以产生多于一个延伸的条形码化核酸分子的拷贝,其中确定样品中核酸靶的拷贝数包括:基于以下的(i)和/或(ii)确定样品中核酸靶的拷贝数:(i)与多于一个延伸的条形码化核酸分子或其产物的拷贝关联的具有不同序列的第一分子标记的数量,和/或(ii)与多于一个延伸的条形码化核酸分子或其产物的拷贝关联的具有不同序列的第二分子标记的数量。

[0014] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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,每个延伸的条形码化核酸分子包含第一分子标记和第二分子标记;扩增多于一个延伸的条形码化核酸分子,以产生多于一个单标记的核酸分子,每个单标记的核酸分子包含第一分子标记或第二分子标记;以及基于与多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定样品中核酸靶的拷贝数。

[0015] 在一些实施方案中,该方法包括基于与多于一个单标记的核酸分子关联的具有不同序列的第一分子标记的数量,确定样品中核酸靶的拷贝数。在一些实施方案中,该方法包括在使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交之前,使多于一个条形码化核酸分子变性:(i)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,和/或(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。在一些实施方案中,该方法包括在扩增多于一个延伸的条形码化核酸分子之前,使多于一个延伸的条形码化核酸分子变性。在一些实施方案中,确定核酸靶的拷贝数包括基于与多于一个单标记的核酸分子中的单标记的核酸分子关联的具有不同序列的第二分子标记的数量来确定样品中多于一个核酸靶中的每一个的拷贝数,所述多于一个单标记的核酸分子包括多于一个核酸靶中的每一个的序列。在一些实施方案中,确定核酸靶的拷贝数包括基于

与多于一个单标记的核酸分子中的单标记的核酸分子关联的具有不同序列的第一分子标记的数量来确定样品中多于一个核酸靶中的每一个的拷贝数,所述多于一个单标记的核酸分子包括多于一个核酸靶中的每一个的序列。在一些实施方案中,多于一个核酸靶中的每一个的序列包括多于一个核酸靶中的每一个的子序列。在一些实施方案中,多于一个条形码化核酸分子中的核酸靶的序列包括核酸靶的子序列。

[0016] 在一些实施方案中,在使多于一个条形码化核酸分子的3'-末端延伸后,使第一分子标记与第二分子标记杂交。在一些实施方案中,延伸的条形码化核酸分子各自包含第一分子标记、第二分子标记、靶结合区和靶结合区的互补体。在一些实施方案中,靶结合区的互补体与靶结合区的一部分互补。在一些实施方案中,靶结合区域包含基因特异性序列。在一些实施方案中,靶结合区域包含多(dT)序列。

[0017] 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体与条形码化核酸分子本身的靶结合区杂交包括使条形码化核酸分子内的靶结合区和靶结合区的互补体进行分子内杂交以形成茎环。在一些实施方案中,第二分子标记是第一分子标记的互补体。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体与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体与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。在一些实施方案中,第二分子标记不同于第一分子标记,并且其中第二分子标记不是第一分子标记的互补体。在一些实施方案中,该方法包括使与条形码化核酸分子的靶结合区的互补体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,每个延伸的条形码化核酸分子包含第一分子标记的互补体和第二分子标记。在一些实施方案中,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体。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体与多于一个条形码化核酸分子中的不同条形码化核酸分子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体与多于一个条形码化核酸分子中的不同条形码化核酸分子的靶结合区进行分子间杂交。在一些实施方案中,第二分子标记的序列不同于第一分子标记的序列,并且其中第二分子标记不是第一分子标记的互补体。

[0018] 在一些实施方案中,逆转录酶能够发挥末端转移酶活性。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'核糖核苷酸包括鸟嘌呤。在一些实施方案中,逆转录酶包括病毒逆转录酶,例如鼠白血病病毒(MLV)逆转录酶或莫洛尼鼠白血病病毒(MMLV)逆转录酶。

[0019] 在一些实施方案中,样品包括单细胞。在一些实施方案中,样品包括多于一个细胞、多于一个单细胞、组织、肿瘤样品,或它们的任何组合。在一些实施方案中,单细胞包括免疫细胞。在一些实施方案中,免疫细胞是B细胞或T细胞。在一些实施方案中,单细胞包括循环肿瘤细胞。在一些实施方案中,每个寡核苷酸条形码包含第一通用序列。在一些实施方案中,多于一个延伸的条形码化核酸分子包含第一通用序列和第一通用序列的互补体。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以产生多于一个延伸的条形码化核酸分子的拷贝包括使用能够与第一通用序列或其互补体杂交的引物。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以产生多于一个单标记的核酸分子包括使用能够与第一通用序列或其互补体杂交的引物,和扩增引物。在一些实施方案中,扩增引物是靶

特异性引物。在一些实施方案中,靶特异性引物与免疫受体特异性地杂交。在一些实施方案中,靶特异性引物与免疫受体的恒定区特异性地杂交。在一些实施方案中,靶特异性引物与免疫受体的可变区特异性地杂交。在一些实施方案中,靶特异性引物与免疫受体的多样性区特异性地杂交。在一些实施方案中,靶特异性引物与免疫受体的可变区和多样性区的连接区(junction)特异性地杂交。在一些实施方案中,免疫受体是T细胞受体(TCR)和/或B细胞受体(BCR)。在一些实施方案中,TCR包括TCRα链、TCRβ链、TCRγ链、TCRδ链,或它们的任何组合。在一些实施方案中,BCR包括BCR重链和/或BCR轻链。

[0020] 在一些实施方案中,使多于一个条形码化核酸分子的3'-末端延伸包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使多于一个条形码化核酸分子的3'-末端延伸。在一些实施方案中,DNA聚合酶包括Klenow片段。在一些实施方案中,该方法包括获得多于一个延伸的条形码化核酸分子或其产物的序列信息。在一些实施方案中,获得序列信息包括将测序衔接子附接到多于一个延伸的条形码化核酸分子或其产物。在一些实施方案中,获得序列信息包括将测序衔接子附接到多于一个单标记的核酸分子或其产物。在一些实施方案中,获得序列信息包括获得单细胞的BCR轻链和BCR重链的序列信息。在一些实施方案中,BCR轻链和BCR重链的序列信息包括BCR轻链和/或BCR重链的互补决定区1(CDR1)、CDR2、CDR3,或它们的任何组合的序列。在一些实施方案中,该方法包括基于获得的序列信息将单细胞的BCR轻链和BCR重链配对。在一些实施方案中,样品包含多于一个单细胞,该方法包括基于获得的序列信息将至少50%的所述单细胞的BCR轻链和BCR重链配对。在一些实施方案中,获得序列信息包括获得单细胞的TCRα链和TCRβ链的序列信息。在一些实施方案中,TCRα链和TCRβ链的序列信息包括TCRα链和/或TCRβ链的互补决定区1(CDR1)、CDR2、CDR3,或它们的任何组合的序列。在一些实施方案中,该方法包括基于获得的序列信息将单细胞的TCRα链和TCRβ链配对。在一些实施方案中,样品包含多于一个单细胞,该方法包括基于获得的序列信息将至少50%的所述单细胞的TCRα链和TCRβ链配对。在一些实施方案中,获得序列信息包括获得单细胞的TCRγ链和TCRδ链的序列信息。在一些实施方案中,TCRγ链和TCRδ链的序列信息包括TCRγ链和/或TCRδ链的互补决定区1(CDR1)、CDR2、CDR3,或它们的任何组合的序列。在一些实施方案中,该方法包括基于获得的序列信息将单细胞的TCRγ链和TCRδ链配对。在一些实施方案中,样品包含多于一个单细胞,该方法包括基于获得的序列信息将至少50%的所述单细胞的TCRγ链和TCRδ链配对。

[0021] 在一些实施方案中,靶结合区的互补体包括靶结合区的反向互补序列。在一些实施方案中,靶结合区的互补体包括靶结合区的互补序列。在一些实施方案中,分子标记的互补体包括分子标记的反向互补序列。在一些实施方案中,分子标记的互补体包括分子标记的互补序列。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。在一些实施方案中,条形码化核酸分子包括条形码化核糖核酸(RNA)分子。在一些实施方案中,核酸靶包括核酸分子。在一些实施方案中,核酸分子包括包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、含有多(A)尾的RNA,或它们的任何组合。在一些实施方案中,mRNA编码免疫受体。在一些实施方案中,核酸靶包含细胞组分结合试剂。在一些实施方案中,核酸分子与细胞组分结合试剂关联。在一些实施方案中,该方法包括使核酸分子和细胞组分结合试剂解离。在一些实施方案中,多于一个寡核苷酸条形码中的至少10个包含不同的分子标记序列。在一些实施方案中,多于一个寡核苷酸

条形码的每个分子标记包含至少6个核苷酸。

[0022] 在一些实施方案中,多于一个寡核苷酸条形码与固体支持物关联。在一些实施方案中,与相同固体支持物关联的多于一个寡核苷酸条形码各自包含相同的样品标记。在一些实施方案中,多于一个寡核苷酸条形码的每个样品标记包含至少6个核苷酸。在一些实施方案中,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,多于一个寡核苷酸条形码的每个细胞标记包含至少6个核苷酸。在一些实施方案中,与相同的固体支持物关联的寡核苷酸条形码包含相同的细胞标记。在一些实施方案中,与不同的固体支持物关联的寡核苷酸条形码包含不同的细胞标记。在一些实施方案中,多于一个延伸的条形码化核酸分子各自包含细胞标记和细胞标记的互补体。在一些实施方案中,细胞标记的互补体包括细胞标记的反向互补序列。在一些实施方案中,细胞标记的互补体包括细胞标记的互补序列。在一些实施方案中,该方法包括在乙二醇、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,或它们的任何组合中的一种或更多种的存在下,使与核酸靶的拷贝杂交的多于一个寡核苷酸条形码延伸。在一些实施方案中,固体支持物包括合成颗粒。在一些实施方案中,固体支持物包括平坦表面。

[0023] 在一些实施方案中,样品包含单细胞,该方法包括使包含多于一个寡核苷酸条形码的合成颗粒与样品中的单细胞关联。在一些实施方案中,该方法包括在使合成颗粒与单细胞关联后裂解单细胞。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,或它们的任何组合。在一些实施方案中,合成颗粒和单细胞在同一孔中。在一些实施方案中,合成颗粒和单细胞在同一液滴中。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被包封在合成颗粒中。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地包封在合成颗粒中。在一些实施方案中,合成颗粒是可破坏的。在一些实施方案中,合成颗粒包括珠。在一些实施方案中,珠包括琼脂糖凝胶珠(sepharose bead)、链霉抗生物素蛋白珠、琼脂糖珠(agarose bead)、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。在一些实施方案中,合成颗粒包含选自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、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。在一些实施方案中,合成颗粒包括可破坏的水凝胶颗粒。在一些实施方案中,多于一个寡核苷酸条形码中的每一个包含接头官能基团,合成颗粒包含固体支持物官能基团,和/或支持物官能基团和接头官能基团彼此关联。在一些实施方案中,接头官能基团和支持物官能基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,以及它们的任何组合。

[0024] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一个寡核苷酸条形码,其中多于一个寡核苷酸条形码中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个包含不同的分子标记序列;逆转录酶;模板转换寡核苷酸,所述模板转换寡核苷酸包含靶结合区或其一部分;和DNA聚合酶,所述DNA聚合酶缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种。

[0025] 在一些实施方案中,DNA聚合酶包括Klenow片段。在一些实施方案中,逆转录酶包括病毒逆转录酶。在一些实施方案中,病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶。在一些实施方案中,病毒逆转录酶是莫洛尼鼠白血病病毒(MMLV)逆转录酶。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。在一些实施方案中,模板转换寡核苷酸包含三个3'核糖核苷酸。在一些实施方案中,3'核糖核苷酸包括鸟嘌呤。在一些实施方案中,试剂盒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(7-deaza-GTP)、乙酰胺、四甲基氯化铵盐、甜菜碱,或它们的任何组合。

[0026] 在一些实施方案中,试剂盒包含缓冲液。在一些实施方案中,试剂盒包含药筒(cartridge)。在一些实施方案中,试剂盒包含用于逆转录反应的一种或更多种试剂。在一些实施方案中,试剂盒包含用于扩增反应的一种或更多种试剂。在一些实施方案中,靶结合区包含基因特异性序列、寡聚(dT)序列、随机多聚体,或它们的任何组合。在一些实施方案中,寡核苷酸条形码包含相同的样品标记和/或相同的细胞标记。在一些实施方案中,多于一个寡核苷酸条形码的每个样品标记和/或细胞标记包含至少6个核苷酸。在一些实施方案中,多于一个寡核苷酸条形码的每个分子标记包含至少6个核苷酸。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被包封在合成颗粒中。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地包封在合成颗粒中。在一些实施方案中,合成颗粒是可破坏的。在一些实施方案中,合成颗粒包括珠。在一些实施方案中,珠包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。在一些实施方案中,合成颗粒包含选自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、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。在一些实施方案中,合成颗粒包括可破坏的水凝胶颗粒。在一些实施方案中,多于一个寡核苷酸条形码中的每一个包含接头官能基团,合成颗粒包含固体支持物官能基团,和/或支持物官能基团和接头官能基团彼此关联。在一些实施方案中,接头官能基团和支持物官能基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,以及它们的任何组合。

[0027] 附图简述

[0028] 图1图示了非限制性示例性条形码。

[0029] 图2示出了条形码化和数字计数的非限制性示例性工作流程。

[0030] 图3是示出了用于从多于一个靶生成在3'末端条形码化的靶的索引文库的非限制性示例性方法的示意图。

[0031] 图4A-图4B示出了在5'-末端对核酸靶进行基因特异性标记的非限制性示例性方法的示意性图示。

[0032] 图5A-图5B示出了在5'-末端对核酸靶进行标记用于全转录组分析的非限制性示例性方法的示意性图示。

[0033] 图6A-图6K示出了使用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)的序列的非限制性示例性工作流程的示意性图示。

[0034] 图7示出了以一个工作流程进行V(D)J方案、抗体-寡核苷酸(AbO)方案和单细胞mRNA表达谱方案(例如BD Rhapsody靶向方案)的非限制性示例性的示意性图示。

[0035] 图8A-图8C示出了使用V(D)J方案对5’T细胞受体(TCR)V(D)J区进行捕获和测序的非限制性示例性实验结果。V(D)J发夹方案可以包括在相同珠上用来自捕获的单个静息外周血单核细胞(PBMC)的mRNA分子进行的3’扩增和5’扩增。TCRα链/β链的配对效率为37.9%,与其他平台诸如Clonetech的scTCR谱分析试剂盒可比较。

[0036] 图9A-图9B示出了图示使用改进的V(D)J方案提高5’V(D)J检测灵敏度的非限制性示例图。添加乙二醇以帮助减少逆转录(RT)中的二级结构。改变杂交时间、缓冲液和模板转换(TS)寡聚dT长度以提高灵敏度。生成了四个文库并一起测序:5’TCR、5’BCR、5’30重免疫小组(5’30-plex immune panel)、3’免疫反应小组(3’immune response panel)。进行了更严格的Ampure清理(0.6X)。

[0037] 图10A-图10E示出了使用参考图9描述的改进的V(D)J方案提高5’V(D)J检测灵敏度的非限制性示例性实验结果。与参考图8描述的方案相比,TCRα/β配对效率从37.9%提高到52%。成功地检测到了B细胞中的5’重链和轻链mRNA分子。

[0038] 图11A-图11F是示出了使用改进的V(D)J方案比较了使用3’扩增(图11A、图11C、图11E)和5’扩增(图11B、图11D、图11F)确定的IGHM、IGHD和IGHA的表达谱的5’B细胞重链检测的非限制性示例图。

[0039] 图12A-图12F是比较了使用3’扩增(图12A、图12C、图12E)和5’扩增(图12B、图12D、图12F)确定的CD3D、CD8A和HLA-DR的表达谱的非限制性示例图。

[0040] 详细描述

[0041] 在以下详细描述中,参考了构成本文的一部分的附图。在附图中,除非上下文另有指示,否则相似的符号通常标识相似的部件。在详细描述、附图和权利要求书中描述的说明性实施方案不意味着是限制性的。在不脱离本文提出的主题的精神或范围的情况下,可以利用其他实施方案,并且可以做出其他改变。将容易理解的是,如本文一般描述的以及附图中图示的本公开内容的方面能够以各种不同的配置来布置、替换、组合、分离和设计,所有这些都在本文中明确考虑并且构成本公开内容的一部分。

[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本文。

[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)的绝对数量也是非常具有挑战性的,尤其是当分子数量非常小时。确定样品中分子的绝对数量的一种方法是数字聚合酶链式反应(PCR)。理想地,PCR在每个循环产生分子的相同拷贝。然而,PCR可具有缺点使得每个分子复制具有随机概率,且此概率根据PCR循环和基因序列而变化,这导致扩增偏倚(amplification bias)和不准确的基因表达测量。可以将具有独特分子标记(也称为分子索引(MI))的随机条形码用于对分子的数量进行计数并且纠正扩增偏倚。诸如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™测定可利用具有在多(T)寡核苷酸上的大量(例如6561个至65536个)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。随机条形码可包括通用PCR引发位点。在RT期间,靶基因分子与随机条形码随机地反应。每个靶分子可以与随机条形码杂交,从而产生随机地条形码化的互补核糖核苷酸(cDNA)分子。在标记后,可将来自微孔板的微孔的随机条形码化cDNA分子汇集到单支管中用于PCR扩增和测序。可以分析原始测序数据以产生读段的数量、具有独特分子标记序列的随机条形码的数量以及mRNA分子的数量。

[0045] 本文的公开内容包括用于对样品中的核酸靶进行标记的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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;以及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,每个条形码化核酸分子包含第一分子标记和第二分子标记。

[0046] 本文的公开内容包括用于确定样品中核酸靶的数量的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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,每个延伸的条形码化核酸分子包含第一分子标记和第二分子标记;以及基于与多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定样品中核酸靶的拷贝数。

[0047] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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,每个延伸的条形码化核酸分子包含第一分子标记和第二分子标记;扩增多于一个延伸的条形码化核酸分子,以产生多于一个单标记的核酸分子,每个单标记的核酸分子包含第一分子标记或第二分子标记;以及基于与多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定样品中核酸靶的拷贝数。

[0048] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一个寡核苷酸条形码,其中多于一个寡核苷酸条形码中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个包含不同的分子标记序列;逆转录酶;模板转换寡核苷酸,所述模板转换寡核苷酸包含靶结合区或其一部分;和DNA聚合酶,所述DNA聚合酶缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种。

[0049] 定义

[0050] 除非另外定义,本文使用的技术术语和科学术语具有与本公开内容所属领域的普通技术人员通常所理解的相同含义。参见,例如,Singleton等人,Dictionary of Microbiology and Molecular Biology,第2版,J.Wiley&Sons,(New York,NY 1994);Sambrook等人,Molecular Cloning,A Laboratory Manual,Cold Spring Harbor Press (Cold Spring Harbor,NY 1989)。出于本公开内容的目的,以下术语在下文定义。

[0051] 如本文使用的,术语"衔接子"可以意指促进关联的核酸的扩增或测序的序列。关联的核酸可包括靶核酸。关联的核酸可以包含以下中的一个或更多个:空间标记、靶标记、样品标记、编索引标记(indexing label),或条形码序列(例如,分子标记)。衔接子可以是线性的。衔接子可以是预腺苷酸化的衔接子。衔接子可以是双链或单链的。一个或更多个衔接子可以位于核酸的5'或3'末端。当衔接子在5'和3'末端包含已知序列时,已知序列可以是相同或不同的序列。位于多核苷酸的5'和/或3'末端的衔接子能够与固定在表面上的一个或更多个寡核苷酸杂交。在一些实施方案中,衔接子可以包含通用序列。通用序列可以是两个或更多个核酸分子共有的核苷酸序列的区域。两个或更多个核酸分子也可具有不同序列的区域。因此,例如,5'衔接子可以包含相同和/或通用核酸序列,并且3'衔接子可以包含相同和/或通用序列。可以存在于多于一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个通用引物复制或扩增多个不同序列。相似地,可以存在于核酸分子的集合中的不同成员中的至少一个、两个(例如,一对)或更多个通用序列可以允许使用与通用序列互补的至少一个、两个(例如,一对)或更多个单个通用引物复制或扩增多个不同序列。因此,通用引物包括可以与这样的通用序列杂交的序列。可以修饰具有靶核酸序列的分子以将通用衔接子(例如,非靶核酸序列)附接到不同靶核酸序列的一个末端或两个末端。与靶核酸附接的一个或更多个通用引物可以提供用于通用引物的杂交的位点。附接到靶核酸的一个或更多个通用引物可以彼此相同或不同。

[0052] 如本文使用的,术语"关联"或"与.....关联"可以意指两个或更多个种类可以被鉴定为在某个时间点处共定位。关联可意指两个或更多个种类在或曾经在相似的容器内。关联可以是信息学关联。例如,关于两个或更多个种类的数字信息可以被存储并且可以用于确定所述种类中的一个或更多个在某个时间点处共定位。关联也可以是物理关联。在一些实施方案中,两个或更多个关联的种类彼此之间或与共同的固体或半固体表面是"连接的"、"附接的"或"固定的"。关联可以指用于将标记附接到固体或半固体支持物(诸如珠)上

的共价或非共价方式。关联可以是靶与标记之间的共价键。关联可以包括两个分子(诸如靶分子和标记)之间的杂交。

[0053] 如本文使用的,术语“互补”可以指两个核苷酸之间精确配对的能力。例如,如果核酸的在给定位置的核苷酸能够与另一个核酸的核苷酸以氢键键合,则两个核酸被认为在该位置处是彼此互补的。两个单链核酸分子之间的互补性可以是“部分的”,其中所述核苷酸中仅一些核苷酸结合,或者当所述单链分子之间存在完全互补性时,这种互补性可以是完全的。如果第一核苷酸序列与第二核苷酸序列互补,则可以称第一核苷酸序列是第二序列的“互补体”。如果第一核苷酸序列与和第二序列相反的序列(即,核苷酸顺序相反)互补,则可以称第一核苷酸序列是第二序列的“反向互补体”。如本文使用的,“互补”序列可以指序列的“互补体”或“反向互补体”。从本公开内容可以理解,如果一个分子可以与另一个分子杂交,则其可以与杂交的分子互补或部分地互补。

[0054] 如本文使用的,术语“数字计数”可以指用于估计样品中靶分子的数量的方法。数字计数可以包括确定已经与样品中的靶关联的独特标记的数量的步骤。这种在性质上可以使随机的方法将对分子进行计数的问题从相同分子的定位和鉴定中的一种转化为有关检测一组预定义标记的一系列是/否数字问题。

[0055] 如本文使用的,术语(一个或更多个)“标记”可以指与样品中的靶关联的核酸编码。标记可以是例如核酸标记。标记可以是完全或部分可扩增的标记。标记可以是完全或部分可测序的标记。标记可以是可鉴定为有区别的天然核酸的一部分。标记可以是已知的序列。标记可以包括核酸序列的连接区,例如天然序列和非天然序列的连接区。如本文使用的,术语“标记”可以与术语“索引”、“标签”或“标记-标签”互换使用。标记可以传达信息。例如,在各种实施方案中,可以使用标记来确定样品的身份、样品的来源、细胞的身份,和/或靶。

[0056] 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,每个靶可能与独特的条形码关联。每个经标记的靶分子的独特性可以通过随机选择的统计来确定,并且取决于与多样的标记相比在集合中相同的靶分子的拷贝数。所得的经标记的靶分子集合的大小可以通过条形码化过程的随机性质来确定,并且然后对检测到的条形码的数量的分析允许计算原始集合或样品中存在的靶分子的数量。当存在的靶分子的拷贝数量与独特的条形码的数量的比率低时,经标记的靶分子是高度独特的(即,用给定的标记来标记多于一个靶分子的概率非常低)。

[0057] 如本文使用的,术语“核酸”是指多核苷酸序列或其片段。核酸可以包括核苷酸。核酸对于细胞可以是外源的或内源的。核酸可以存在于无细胞环境中。核酸可以是基因或其片段。核酸可以是DNA。核酸可以是RNA。核酸可以包括一种或更多种类似物(例如改变的骨架、糖或核碱基)。类似物的一些非限制性实例包括:5-溴尿嘧啶、肽核酸、异源核酸(xeno nucleic acid)、吗啉代核酸(morpholinos)、锁核酸、二醇核酸、苏糖核酸、二脱氧核苷酸、虫草菌素、7-脱氮-GTP、荧光团(例如,罗丹明或与糖连接的荧光黄素)、含硫醇的核苷酸、生物素连接的核苷酸、荧光碱基类似物、CpG岛、甲基-7-鸟苷、甲基化的核苷酸、肌苷、硫代尿苷、假尿苷、二氢尿苷、辫苷(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。

[0058] 核酸可以包含一种或更多种修饰(例如,碱基修饰、骨架修饰),以向核酸提供新的或增强的特征(例如,改进的稳定性)。核酸可以包含核酸亲和标签。核苷可以是碱基-糖组合。核苷的碱基部分可以是杂环碱基。这样的杂环碱基的两个最常见的类别是嘌呤和嘧啶。核苷酸可以是进一步包括与核苷的糖部分共价连接的磷酸基团的核苷。对于包括呋喃戊糖的那些核苷,磷酸基团可以连接到糖的2'、3'或5'羟基部分。在形成核酸中,磷酸基团可以将相邻的核苷彼此共价连接以形成线性聚合化合物。继而此线性聚合化合物的各自末端可以进一步接合以形成环状化合物;然而,线性化合物通常是合适的。此外,线性化合物可以具有内部核苷酸碱基互补性,并且因此可以按产生完全或部分双链化合物的方式折叠。在核酸中,磷酸基团通常可以被称为形成核酸的核苷间骨架。连接或骨架可以是3'至5'磷酸二酯连接。

[0059] 核酸可以包含修饰的骨架和/或修饰的核苷间连接。修饰的骨架可以包括那些在骨架中保留磷原子的骨架和那些在骨架中不具有磷原子的骨架。其中含有磷原子的合适的修饰的核酸骨架可以包括,例如,硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其他烷基磷酸酯,诸如3'-亚烷基磷酸酯、5'-亚烷基磷酸酯、手性磷酸酯、亚磷酸酯,包括3'-氨基磷酰胺和氨基烷基磷酰胺,磷酰二胺、硫代磷酰胺、硫代烷基磷酸酯、硫代烷基磷酸三酯、硒代磷酸酯和具有正常的3'-5'连接的硼酸磷酸酯(boranophosphates)、2'-5'连接的类似物,以及具有反转极性的那些,其中一个或更多个核苷酸间连接为3'至3'、5'至5'或2'至2'连接。

[0060] 核酸可以包含由以下形成的多核苷酸骨架:短链烷基或环烷基核苷间连接、混合杂原子和烷基或环烷基核苷间连接,或者一个或更多个短链杂原子的或杂环的核苷间连接。这些可以包括具有以下的那些:吗啉代连接(部分地从核苷的糖部分形成);硅氧烷骨架;硫化物、亚砜和砜骨架;甲乙酰基(formacetyl)和硫代甲乙酰基(thioformacetyl)骨架;亚甲基甲乙酰基和硫代甲乙酰基骨架;核糖乙酰基(riboacetyl)骨架;含有烯烃的骨架;氨基磺酸盐骨架;亚甲基亚氨基和亚甲基肼基骨架;磺酸盐和磺酰胺骨架;酰胺骨架;和具有混合的N、O、S和CH_2组成部分的其他物质。

[0061] 核酸可以包括核酸模拟物。术语"模拟物"可意在包括其中仅呋喃糖环或呋喃糖环和核苷酸间连接二者被非呋喃糖基团代替的多核苷酸,仅呋喃糖环的代替也可称为糖替代物。杂环碱基部分或修饰的杂环碱基部分可被保持以与适当的靶核酸杂交。一种这样的核酸可以是肽核酸(PNA)。在PNA中,多核苷酸的糖骨架可以被含酰胺的骨架,特别是氨基乙基甘氨酸骨架代替。核苷酸可以被保留,并直接或间接结合到骨架的酰胺部分的氮杂氮原子。PNA化合物中的骨架可以包含两个或更多个连接的氨基乙基甘氨酸单元,这为PNA提供含酰胺的骨架。杂环碱基部分可以直接或间接结合到骨架的酰胺部分的氮杂氮原子。

[0062] 核酸可以包含吗啉代骨架结构。例如,核酸可以包含代替核糖环的6元吗啉代环。在这些实施方案中的一些中,磷酰二胺或其他非磷酸二酯核苷间连接可以代替磷酸二酯连接。

[0063] 核酸可以包含具有附接到吗啉代环上的杂环碱基的连接的吗啉代单元(例如吗啉代核酸)。连接基团可以连接吗啉代核酸中的吗啉代单体单元。非离子的基于吗啉代的寡聚化合物可与细胞蛋白质具有较少的不期望的相互作用。基于吗啉代的多核苷酸可以是核酸的非离子模拟物。吗啉代类别中的各种化合物可以使用不同的连接基团连接。另一类多核

苷酸模拟物可以称为环己烯基核酸(CeNA)。核酸分子中通常存在的呋喃糖环可以被环己烯基环取代。使用亚磷酰胺化学可以制备CeNA DMT保护的亚磷酰胺单体并用于寡聚化合物合成。将CeNA单体掺入核酸链中可以增加DNA/RNA杂交体的稳定性。CeNA寡腺苷酸可以与核酸互补序列形成稳定性与天然复合物相似的复合物。另外的修饰可以包括锁核酸(LNA),其中2'-羟基基团连接到糖环的4'碳原子,从而形成2'-C,4'-C-氧亚甲基连接,从而形成双环糖部分。连接可以是亚甲基(-CH_2),桥接2'氧原子和4'碳原子的基团,其中n是1或2。LNA和LNA类似物可以显示出与互补核酸的非常高的双链体热稳定性(Tm=+3℃至+10℃)、对3'-核酸外切酶降解的稳定性和良好的溶解度特性。

[0064] 核酸还可以包括核碱基(通常简称为"碱基")修饰或取代。如本文使用的,"未修饰的"或"天然的"核碱基可以包括嘌呤碱基(例如腺嘌呤(A)和鸟嘌呤(G)),以及嘧啶碱基(例如胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U))。经修饰的核碱基可以包括其他合成以及天然的核碱基,诸如5-甲基胞嘧啶(5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基衍生物和其他烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基衍生物和其他烷基衍生物、2-硫代尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-卤素尿嘧啶(5-halouracil)和胞嘧啶、5-丙炔基(-C≡C-CH3)尿嘧啶和胞嘧啶,和嘧啶碱基的其他炔基衍生物、6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫代尿嘧啶,8-卤素、8-氨基、8-硫代、8-硫代烷基、8-羟基和其他8-取代的腺嘌呤和鸟嘌呤,5-卤素特别是5-溴、5-三氟甲基和其他5-取代的尿嘧啶和胞嘧啶,7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。修饰的核碱基可以包括三环嘧啶,诸如吩噁嗪胞苷(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)苯并噻嗪-2(3H)-酮)、G-夹类(G-clamps),诸如取代的吩噁嗪胞苷(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)苯并噻嗪-2(3H)-酮)、G-夹类,诸如取代的吩噁嗪胞苷(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)、咔唑胞苷(2H-嘧啶并(4,5-b)吲哚-2-酮)、吡啶并吲哚胞苷(H-吡啶并(3',2':4,5)吡咯并[2,3-d]嘧啶-2-酮)。

[0065] 如本文使用的,术语"样品"可以指包含靶的组合物。用于通过本文公开的方法、装置和系统进行分析的合适样品包括细胞、组织、器官或生物体。

[0066] 如本文使用的,术语"采样装置"或"装置"可以指可以取一部分样品和/或将所述部分放置在基底上的装置。采样装置可以指例如荧光激活细胞分选(FACS)机、细胞分选机、活组织检查针、活组织检查装置、组织切片装置、微流体装置、叶栅和/或超薄切片机。

[0067] 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、圆柱体或其他类似形状,由塑料、陶瓷、金属或聚合材料(例如,水凝胶)构成,其上可以固定核酸(例如,共价地或非共价地)。固体支持物可以包括可以是球形的(例如,微球)或具有非球形或不规则形状的离散颗粒,所述形状诸如立方体、长方体、锥体、圆柱体、圆锥体、椭圆形或圆盘形等。珠的形状可以是非球形的。以阵列间隔开的多于一个固体支持物可以不包括基底。固体支持物可以与术语"珠"互换使用。

[0068] 如本文使用的,术语"随机条形码"可以指包含本公开内容的标记的多核苷酸序

列。随机条形码可以是可用于随机条形码化的多核苷酸序列。随机条形码可用于对样品中的靶进行定量。随机条形码可用于控制标记与靶关联后可能发生的错误。例如,随机条形码可用于评估扩增或测序错误。与靶关联的随机条形码可以称为随机条形码-靶或随机条形码-标签-靶。

[0069] 如本文使用的,术语“基因特异性随机条形码”可以指包含标记和基因特异性的靶结合区的多核苷酸序列。随机条形码可以是可用于随机条形码化的多核苷酸序列。随机条形码可用于对样品中的靶进行定量。随机条形码可用于控制标记与靶关联后可能发生的错误。例如,随机条形码可用于评估扩增或测序错误。与靶关联的随机条形码可以称为随机条形码-靶或随机条形码-标签-靶。

[0070] 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。

[0071] 如本文使用的,术语“靶”可以指可与条形码(例如,随机条形码)关联的组合物。用于通过本文公开的方法、装置和系统进行分析的示例性合适的靶包括寡核苷酸、DNA、RNA、mRNA、微RNA、tRNA等。靶可以是单链的或双链的。在一些实施方案中,靶可以是蛋白质、肽或多肽。在一些实施方案中,靶是脂质。如本文使用的,“靶”可以与“种类”互换使用。

[0072] 如本文使用的,术语“逆转录酶”可以指具有逆转录酶活性(即,催化从RNA模板合成DNA)的一组酶。通常,这样的酶包括但不限于逆转录病毒逆转录酶、逆转录转座子逆转录酶、逆转录质粒逆转录酶、逆转录子逆转录酶(retron reverse transcriptase)、细菌逆转录酶、II型内含子衍生的逆转录酶(group II intron-derived reverse transcriptase),以及它们的突变体、变体或衍生物。非逆转录病毒逆转录酶包括非LTR逆转录转座子逆转录酶、逆转录质粒逆转录酶、逆转录子逆转录酶和II型内含子逆转录酶。II型内含子逆转录酶的实例包括乳酸乳球菌(Lactococcus lactis)LI.LtrB内含子逆转录酶、细长嗜热聚球藻(Thermosynechococcus elongatus)TeI4c内含子逆转录酶或嗜热脂肪地芽孢杆菌(Geobacillus stearothermophilus)GsI-IIC内含子逆转录酶。其他类别的逆转录酶可以包括许多类型的非逆转录病毒逆转录酶(即,逆转录子、II型内含子,以及多样性产生型逆转录元件,等等)。

[0073] 术语“通用衔接子引物”、“通用引物衔接子”或“通用衔接子序列”可互换地使用以指可以用于与条形码(例如,随机条形码)杂交以产生基因特异性条形码的核苷酸序列。通用衔接子序列例如可以是跨越本公开内容的方法中使用的所有条形码通用的已知序列。例如,当使用本文公开的方法标记多个靶时,每个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个靶时,至少两个靶特异性序列连接到不同的通用衔接子序列。通用衔接子引物及其互补体可以包括在两个寡核苷酸中,其中的一个包含靶特异性序列且另一个包含条形码。例如,通用衔接子序列可以是包含靶特异性序列的寡核苷酸的一部分,以产生与靶核酸互补的核苷酸序列。包含条形码以及通用衔接子序列的互补体的第二寡核苷酸可与核苷酸序列杂交并产生靶特异性条形码(例如,靶特异性随机条形码)。在一些实施方案中,通用衔接子引物具有与本公开内容的方法中使用的通用PCR引物不同的序列。

[0074] 条形码

[0075] 条形码化,诸如随机条形码化,已描述于例如US 2015/0299784,WO 2015/031691,以及Fu等人,Proc Natl Acad Sci U.S.A.2011 May 31;108(22):9026-31中,这些出版物的内容以其整体在此并入本文。在一些实施方案中,本文公开的条形码可以是随机条形码,所述随机条形码可以是可用于对靶进行随机标记(例如,条形码,标签)的多核苷酸序列。如果随机条形码中的不同条形码序列的数量与待标记的任何靶的出现次数(muber of occurrence)的比率可以是以下,或是约以下,则条形码可以称为随机条形码:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1,或在这些值中的任何两个值之间的数字或范围。靶可以是包括具有相同或几乎相同序列的mRNA分子的mRNA种类。如果随机条形码中的不同条形码序列的数量与待标记的任何靶的出现次数的比率是至少以下,或至多以下,则条形码可以称为随机条形码:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1或100:1。随机条形码的条形码序列可以称为分子标记。

[0076] 条形码,例如随机条形码,可以包含一个或更多个标记。示例性标记可以包括通用标记、细胞标记、条形码序列(例如,分子标记)、样品标记、板标记、空间标记和/或前空间标记(pre-spatial label)。图1图示了具有空间标记的示例性条形码104。条形码104可以包含可将条形码与固体支持物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)的顺序可以改变。例如,如图1中示出的,通用标记可以是最5’侧的标记(5’-most label),且分子标记可以是最3’侧的标记(3’-most label)。空间标记、维度标记和细胞标记可以处于任何顺序。在一些实施方案中,通用标记、空间标记、维度标记、细胞标记和分子标记是处于任何顺序的。条形码可以包含靶结合区。靶结合区可以与样品中的靶(例如,靶核酸、RNA、mRNA、DNA)相互作用。例如,靶结合区可以包括可以与mRNA的多(A)尾相互作用的寡聚(dT)序列。在一些情况下,条形码的标记(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列)可以由1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个或更多个核苷酸隔开。

[0077] 标记,例如,细胞标记,可以包括一组独特的定义长度的核酸子序列,例如各自有七个核苷酸(相当于一些汉明纠错码(Hamming error correction codes)中使用的数量),其可以设计为提供纠错能力。可以设计包含七个核苷酸序列的纠错子序列组,使得所述组中的序列的任何成对组合展现出定义的“遗传距离”(或错配碱基数),例如一组纠错子序列能被设计为展现三个核苷酸的遗传距离。在这种情况下,对于经标记的靶核酸分子的序列数据组中的纠错序列的审查(在下文更全面地描述)能允许检测或纠正扩增或测序错误。在一些实施方案中,用于产生纠错码的核酸子序列的长度可以变化,例如,它们的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、31个、40个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,其他长度的核酸子序列可以用来产生纠错码。

[0078] 条形码可以包含靶结合区。靶结合区可以与样品中的靶相互作用。靶可以是,或包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、各自含有多

(A)尾的RNA,或它们的任何组合。在一些实施方案中,多于一个靶可以包括脱氧核糖核酸(DNA)。

[0079] 在一些实施方案中,靶结合区可以包含可以与mRNA的多(A)尾相互作用的寡聚(dT)序列。条形码的标记中的一个或更多个(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列(例如,分子标记))可以通过间隔序列(spacer)与条形码的剩余标记中的另一个或两个隔开。间隔序列可以是例如,1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个或更多个核苷酸。在一些实施方案中,条形码的标记中没有标记被间隔序列隔开。

[0080] 通用标记

[0081] 条形码可以包含一个或更多个通用标记。在一些实施方案中,对于条形码组中的所有条形码(附接到给定的固体支持物上的),一个或更多个通用标记可以是相同的。在一些实施方案中,对于附接到多于一个珠上的所有条形码,一个或更多个通用标记可以是相同的。在一些实施方案中,通用标记可以包括能够与测序引物杂交的核酸序列。测序引物可以用于对包含通用标记的条形码进行测序。测序引物(例如,通用测序引物)可以包括与高通量测序平台有关的测序引物。在一些实施方案中,通用标记可以包括能够与PCR引物杂交的核酸序列。在一些实施方案中,通用标记可以包括能够与测序引物和PCR引物杂交的核酸序列。能够与测序或PCR引物杂交的通用标记的核酸序列可以被称为引物结合位点。通用标记可以包括可用于起始条形码转录的序列。通用标记可以包括可用于使条形码或条形码内的区域延伸的序列。通用标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。例如,通用标记可以包括至少约10个核苷酸。通用标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。在一些实施方案中,可裂解接头或修饰的核苷酸可以是通用标记序列的一部分,以使条形码能够从支持物上被裂解下来。

[0082] 维度标记

[0083] 条形码可以包含一个或更多个维度标记。在一些实施方案中,维度标记可以包括提供关于发生标记(例如,随机标记)的维度的信息的核酸序列。例如,维度标记可以提供关于对靶进行条形码化的时间的信息。维度标记可以与样品中条形码化(例如,随机条形码化)的时间关联。维度标记可以在标记的时间被激活。不同的维度标记可以在不同的时间被激活。维度标记提供关于靶、靶的组和/或样品被条形码化的顺序的信息。例如,在细胞周期的G0期可以对细胞群体进行条形码化。在细胞周期的G1期,可以用条形码(例如,随机条形码)再次对细胞进行脉冲处理。在细胞周期的S期,可以用条形码再次对细胞进行脉冲处理,等等。每次脉冲(例如,细胞周期的每个期)时的条形码可以包含不同的维度标记。以这种方式,维度标记提供关于哪些靶在细胞周期的哪个期被标记的信息。维度标记可以询问许多不同的生物学时间。示例性的生物学时间可以包括但不限于细胞周期、转录(例如,转录起始)和转录物降解。在另一个实例中,样品(例如,细胞、细胞群体)可以在用药物和/或疗法治疗之前和/或之后进行标记。不同靶的拷贝数的变化可以指示样品对药物和/或疗法的反应。

[0084] 维度标记可以是可激活的。可以在特定时间点激活可激活的维度标记。可激活的

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

[0085] 在一些实施方案中,维度标记对于附接到给定的固体支持物(例如,珠)上的所有条形码(例如,随机条形码)可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上的至少60%、70%、80%、85%、90%、95%、97%、99%或100%的条形码可以包含相同的维度标记。在一些实施方案中,相同固体支持物上的至少60%的条形码可以包含相同的维度标记。在一些实施方案中,相同固体支持物上的至少95%的条形码可以包含相同的维度标记。

[0086] 多于一个固体支持物(例如,珠)可以表现多达10^6个或更多个独特维度标记序列。维度标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。维度标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。维度标记可以包括约5个至约200个之间的核苷酸。维度标记可以包括约10个至约150个之间的核苷酸。维度标记可以包括长度在约20个至约125个之间的核苷酸。

[0087] 空间标记

[0088] 条形码可以包含一个或更多个空间标记。在一些实施方案中,空间标记可以包括提供与条形码关联的靶分子的空间定向的信息的核酸序列。空间标记可以与样品中的坐标关联。所述坐标可以是固定的坐标。例如,可以参考基底将坐标固定。空间标记可以参考二维或三维网格。可以参考界标(landmark)将坐标固定。在空间中界标是可被鉴定的。界标可以是可被成像的结构。界标可以是生物学结构,例如解剖学界标。界标可以是细胞界标,例如细胞器。界标可以是非天然界标,诸如具有可鉴定标识(诸如颜色编码、条形码、磁性、荧光、放射性或独特尺寸或形状)的结构。空间标记可以与物理分区(partition)(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。

[0089] 空间标记对于附接到给定的固体支持物(例如,珠)上的所有条形码可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上包含相同空间标记的条形码的百分比可以是以下,或约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,相同固体支持物上包含相同空间标记的条形码的百分比可以是至少以下,或至多以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。在一些实施方案中,相同固体支持物上的至少60%的条形码可以包含相同的空间标记。在一些实施方案中,相同固体支持物上的至少95%的条形码可以包含相同的空间标记。

[0090] 多于一个固体支持物(例如,珠)可以表现多达10^6个或更多个独特空间标记序列。

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

[0091] 细胞标记

[0092] 条形码(例如,随机条形码)可以包含一个或更多个细胞标记。在一些实施方案中,细胞标记可以包括提供用于确定哪个靶核酸来自哪个细胞的信息的核酸序列。在一些实施方案中,细胞标记对于附接到给定的固体支持物(例如,珠)上的所有条形码是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上包含相同细胞标记的条形码的百分比可以是以下,或约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,相同固体支持物上包含相同细胞标记的条形码的百分比可以是以下,或约以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。例如,相同固体支持物上的至少60%的条形码可以包含相同的细胞标记。作为另一个实例,相同固体支持物上的至少95%的条形码可以包含相同的细胞标记。

[0093] 多于一个固体支持物(例如,珠)可以表现多达10^6个或更多个独特细胞标记序列。细胞标记的长度可以是以下,或约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。细胞标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。例如,细胞标记可以包括约5个至约200个之间的核苷酸。作为另一个实例,细胞标记可以包括约10个至约150个之间的核苷酸。还作为另一个实例,细胞标记可以包括长度在约20个至约125个之间的核苷酸。

[0094] 条形码序列

[0095] 条形码可以包含一个或更多个条形码序列。在一些实施方案中,条形码序列可以包括提供与条形码杂交的特定类型的靶核酸种类的鉴定信息的核酸序列。条形码序列可以包括如下核酸序列,所述核酸序列提供与条形码(例如,靶结合区)杂交的靶核酸种类的特定出现的计数(例如,提供粗略近似)。

[0096] 在一些实施方案中,一组不同的条形码序列附接到给定的固体支持物(例如,珠)。在一些实施方案中,可以有以下,或有约以下的独特分子标记序列:10^2个、10^3个、10^4个、10^5个、10^6个、10^7个、10^8个、10^9个,或在这些值中的任何两个值之间的数字或范围。例如,多于一个条形码可以包括具有不同序列的约6561个条形码序列。作为另一个实例,多于一个条形码可以包括具有不同序列的约65536个条形码序列。在一些实施方案中,可以有至少以下,或至多以下的独特条形码序列:10^2个、10^3个、10^4个、10^5个、10^6个、10^7个、10^8个或10^9个。独特分子标记序列可以附接到给定的固体支持物(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)部分或全部地包围。

[0097] 在不同实施方式中,条形码的长度可以是不同的。例如,条形码的长度可以是以

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

[0098] 分子标记

[0099] 条形码(例如,随机条形码)可以包含一个或更多个分子标记。分子标记可以包括条形码序列。在一些实施方案中,分子标记可以包括提供与条形码杂交的特定类型的靶核酸种类的鉴定信息的核酸序列。分子标记可以包括如下核酸序列,所述核酸序列提供与条形码(例如,靶结合区)杂交的靶核酸种类的特定出现的计数。

[0100] 在一些实施方案中,一组不同的分子标记附接到给定的固体支持物(例如,珠)。在一些实施方案中,可以有以下,或有约以下的独特分子标记序列:10^2个、10^3个、10^4个、10^5个、10^6个、10^7个、10^8个、10^9个,或在这些值中的任何两个值之间的数字或范围。例如,多于一个条形码可以包括具有不同序列的约6561个分子标记。作为另一个实例,多于一个条形码可以包括具有不同序列的约65536个分子标记。在一些实施方案中,可以有至少以下,或至多以下的独特分子标记序列:10^2个、10^3个、10^4个、10^5个、10^6个、10^7个、10^8个或10^9个。具有独特分子标记序列的条形码可以附接到给定的固体支持物(例如,珠)。

[0101] 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率可以是以下,或约以下:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1,或在这些值中的任何两个值之间的数字或范围。靶可以是包括具有相同或几乎相同序列的mRNA分子的mRNA种类。在一些实施方案中,不同分子标记序列的数量与任何靶的出现次数的比率是至少以下,或至多以下:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1或100:1。

[0102] 分子标记的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。分子标记的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。

[0103] 靶结合区

[0104] 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,靶结合区可以包括与靶(例如,靶核酸、靶分子,例如待分析的细胞核酸),例如,与特定基因序列,进行特异性杂交的核酸序列。在一些实施方案中,靶结合区可以包括可附接(例如,杂交)至特定靶核酸的特定位置的核酸序列。在一些实施方案中,靶结合区可以包括能够与限制性酶位点突出端(例如,EcoRI粘性末端突出端)进行特异性杂交的核酸序列。然后条形码可以连接到包括与限制性位点突出端互补的序列的任何核酸分子。

[0105] 在一些实施方案中,靶结合区可以包括非特异性靶核酸序列。非特异性靶核酸序列可以指不依赖靶核酸的特定序列可与多个靶核酸结合的序列。例如,靶结合区可以包括与mRNA分子上的多(A)尾杂交的随机多聚体序列或寡聚(dT)序列。随机多聚体序列可以是

例如随机二聚体、三聚体、四聚体、五聚体、六聚体、七聚体、八聚体、九聚体、十聚体或任何长度的更高多聚体序列。在一些实施方案中,附接到给定珠的所有条形码的靶结合区是相同的。在一些实施方案中,附接到给定珠的多于一个条形码的靶结合区可以包括两个或更多个不同的靶结合序列。靶结合区的长度可以是以下,或是约以下:5个、10个、15个、20个、25个、30个、35个、40个、45个、50个,或在这些值中的任何两个值之间的数字或范围的核苷酸。靶结合区的长度可以是至多约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更多个核苷酸。

[0106] 在一些实施方案中,靶结合区可以包括寡聚(dT),所述寡聚(dT)可以与包含多腺苷酸化末端的mRNA杂交。靶结合区可以是基因特异性的。例如,可以将靶结合区配置为与靶的特定区域杂交。靶结合区的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。靶结合区的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。靶结合区的长度可以是约5-30个核苷酸。当条形码包含基因特异性靶结合区时,所述条形码在本文中可以被称为基因特异性条形码。

[0107] 定向特性

[0108] 随机条形码(例如,随机条形码)可以包括一种或更多种可用于定向(例如,比对)条形码的定向特性。条形码可以包含用于等电聚焦的部分。不同的条形码可以包括不同的等电聚焦点。当将这些条形码引入样品中时,所述样品可以经历等电聚焦,以便于将条形码定位成已知的方式。以这种方式,定向特性可以用于开发样品中条形码的已知的地图(map)。示例性定向特性可以包括电泳迁移(例如,基于条形码的尺寸)、等电点、自旋、电导和/或自组装。例如,具有自组装的定向特性的条形码在激活后可以自组装成特定定向(例如,核酸纳米结构)。

[0109] 亲和特性

[0110] 条形码(例如,随机条形码)可以包括一种或更多种亲和特性。例如,空间标记可以包括亲和特性。亲和特性可以包括可促进条形码与另一种实体(例如,细胞受体)结合的化学和/或生物部分。例如,亲和特性可以包括抗体,例如,对样品上的特定部分(例如,受体)有特异性的抗体。在一些实施方案中,抗体可以将条形码引导至特定细胞类型或分子。在特定细胞类型或分子处的和/或在特定细胞类型或分子附近的靶可以被标记(例如,随机地标记)。在一些实施方案中,因为抗体可以将条形码引导至特定位置,亲和特性可以提供除了空间标记的核苷酸序列之外的空间信息。抗体可以是治疗性抗体,例如单克隆抗体或多克隆抗体。抗体可以是人源化的或嵌合的。抗体可以是裸抗体(naked antibody)或融合抗体。

[0111] 抗体可以是全长(即,天然存在的或通过正常免疫球蛋白基因片段重组过程形成的)免疫球蛋白分子(例如,IgG抗体)或免疫球蛋白分子的免疫活性(即,特异性结合)部分(如抗体片段)。

[0112] 抗体片段可以是例如抗体的一部分,诸如F(ab')2、Fab'、Fab、Fv、sFv等。在一些实施方案中,抗体片段可以与由全长抗体识别的相同抗原结合。抗体片段可以包括由抗体的

可变区组成的分离的片段,诸如由重链和轻链的可变区组成的“Fv”片段和其中轻链和重链的可变区通过肽接头连接的重组单链多肽分子(“scFv蛋白”)。示例性抗体可以包括但不限于癌细胞抗体、病毒抗体、结合至细胞表面受体(CD8、CD34、CD45)的抗体,和治疗性抗体。

[0113] 通用衔接子引物

[0114] 条形码可以包含一个或更多个通用衔接子引物。例如,基因特异性条形码,诸如基因特异性随机条形码,可以包含通用衔接子引物。通用衔接子引物可以指跨越所有条形码通用的核苷酸序列。通用衔接子引物可以用于构建基因特异性条形码。通用衔接子引物的长度可以是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个,或在这些值中的任何两个值之间的数字或范围的核苷酸。通用衔接子引物的长度可以是至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。通用衔接子引物的长度可以是5-30个核苷酸。

[0115] 接头

[0116] 当条形码包含多于一种标记类型(例如,多于一种细胞标记或多于一种条形码序列,诸如一种分子标记)时,标记之间可以散布有接头标记序列。接头标记序列的长度可以是至少约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更多个核苷酸。接头标记序列的长度可以是至多约5个、10个、15个、20个、25个、30个、35个、40个、45个、50个或更多个核苷酸。在一些情况下,接头标记序列的长度为12个核苷酸。接头标记序列可用于促进条形码的合成。接头标签可以包括纠错(例如汉明)码。

[0117] 固体支持物

[0118] 在一些实施方案中,本文公开的条形码,诸如随机条形码,可以与固体支持物关联。例如,固体支持物可以是合成颗粒。在一些实施方案中,固体支持物上的多于一个条形码(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,第一条形码序列)的分子标记)具有至少一个核苷酸的差异。相同固体支持物上的条形码的细胞标记可以是相同的。不同固体支持物上的条形码的细胞标记可以具有至少一个核苷酸的差异。例如,第一固体支持物上的第一多于一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具有至少一个核苷酸的差异。细胞标记可以是例如约5-20个核苷酸长。条形码序列可以是例如约5-20个核苷酸长。合成颗粒可以例如是珠。

[0119] 珠可以例如是硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、交联葡聚糖/琼脂糖凝胶珠、珠状纤维素、聚苯乙烯珠,或它们的任何组合。珠可以包括诸如以下的材料:聚二甲基硅氧烷(PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,或它们的任何组合。

[0120] 在一些实施方案中,珠可以是用条形码或随机条形码官能化的聚合珠,例如可变形的珠或凝胶珠(诸如来自10X Genomics(San Francisco,CA)的凝胶珠)。在一些实施方式中,凝胶珠可以包括基于聚合物的凝胶。例如,凝胶珠可以通过将一种或更多种聚合物前体包封进液滴中来产生。在将聚合物前体暴露于促进剂(例如,四甲基乙二胺(TEMED))后,可

以产生凝胶珠。

[0121] 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,聚合珠可以例如在所期望的条件下溶解、熔化或降解。所期望的条件可以包括环境条件。所期望的条件可导致聚合珠以受控方式溶解、熔化或降解。由于化学刺激物、物理刺激物、生物刺激物、热刺激物、磁刺激物、电刺激物、光刺激物,或它们的任何组合,凝胶珠可以溶解、熔化或降解。

[0122] 分析物和/或试剂(诸如寡核苷酸条形码)例如可以偶联/固定到凝胶珠的内表面(例如,经由寡核苷酸条形码和/或用于产生寡核苷酸条形码的材料的扩散而可及的内部),和/或凝胶珠或本文描述的任何其他微囊的外表面。偶联/固定可以经由任何形式的化学键合(例如,共价键、离子键)或物理现象(例如,范德华力、偶极-偶极相互作用等)。在一些实施方案中,试剂与凝胶珠或本文描述的任何其他微囊的偶联/固定可以是可逆的,诸如,例如经由不稳定部分(例如,经由化学交联剂,包括本文描述的化学交联剂)。在施加刺激物后,不稳定部分可以被裂解并释放固定化的试剂。在一些实施方案中,不稳定部分是二硫键。例如,在经由二硫键将寡核苷酸条形码固定到凝胶珠的情况下,将二硫键暴露于还原剂可以将二硫键裂解并从珠释放寡核苷酸条形码。不稳定部分可以作为凝胶珠或微囊的一部分、作为将试剂或分析物与凝胶珠或微囊连接的化学接头的一部分,和/或作为试剂或分析物的一部分包括在内。在一些实施方案中,多于一个条形码中的至少一个条形码可固定在颗粒上、部分固定在颗粒上、包封在颗粒中、部分包封在颗粒中,或它们的任何组合。

[0123] 在一些实施方案中,凝胶珠可以包括广泛多种不同的聚合物,包括但不限于:聚合物、热敏聚合物、光敏聚合物、磁性聚合物、pH敏感聚合物、盐敏感聚合物、化学敏感聚合物、聚电解质、多糖、肽、蛋白质和/或塑料。聚合物可包括但不限于以下材料:诸如聚(N-异丙基丙烯酰胺)(PNIPAAm)、聚(苯乙烯磺酸酯)(PSS)、聚(烯丙基胺)(PAAm)、聚(丙烯酸)(PAA)、聚(乙烯亚胺)(PEI)、聚(二烯丙基二甲基-氯化铵)(PDADMAC)、聚(吡咯)(PPy)、聚(乙烯基吡咯烷酮)(PVPON)、聚(乙烯基吡啶)(PVP)、聚(甲基丙烯酸甲酯)(PMAA)、聚(甲基丙烯酸甲酯)(PMMA)、聚苯乙烯(PS)、聚(四氢呋喃)(PTHF)、聚(邻苯二甲醛)(PPA)、聚(己基紫精)(PHV)、聚(L-赖氨酸)(PLL)、聚(L-精氨酸)(PARG)、聚(乳酸-共-乙醇酸)(PLGA)。

[0124] 许多化学刺激物可用于触发珠的破坏、溶解或降解。这些化学改变的实例可包括但不限于pH介导的珠壁改变、经由交联键的化学裂解使珠壁崩解、珠壁的触发解聚,和珠壁转换反应(bead wall switching reactions)。容量改变(bulk changes)也可用于触发珠的破坏。

[0125] 通过各种刺激物对微囊的容量或物理变化在设计囊以释放试剂方面也提供了许多优点。在宏观尺度上发生容量或物理变化,其中珠破裂是由刺激物引起的机械-物理力的结果。这些过程可包括但不限于压力引起的破裂、珠壁熔化,或珠壁的孔隙率的改变。

[0126] 生物刺激物也可用于触发珠的破坏、溶解或降解。通常,生物触发物类似于化学触发物,但是许多实例使用生物分子或生命系统中常见的分子,诸如酶、肽、糖、脂肪酸、核酸等。例如,珠可包括具有肽交联的聚合物,所述肽交联对特定蛋白酶的裂解敏感。更具体地,一个实例可包括含有GFLGK肽交联的微囊。在添加生物触发物(诸如蛋白酶组织蛋白酶B)后,壳壁的肽交联被裂解且珠的内容物被释放。在其他情况下,蛋白酶可以是热激活的。在另一个实例中,珠包括含有纤维素的壳壁。水解酶壳聚糖的添加用作用于纤维素键的裂解、

壳壁的解聚和其内部内容物的释放的生物触发物。

[0127] 还可以在施加热刺激物后诱导珠释放其内容物。温度的变化可导致珠的各种变化。热量的变化可能导致珠的熔化,使得珠壁崩解。在其他情况下,热量可能增加珠的内部组分的内部压力,使得珠破裂或爆炸。在仍然其他情况下,热量可以使珠转变成收缩的脱水状态。热量还可以作用于珠壁内的热敏聚合物,从而引起珠的破坏。

[0128] 将磁性纳米颗粒包括在微囊的珠壁中可以允许珠的触发破裂以及在阵列中引导珠。本公开内容的装置可以包括用于任一目的的磁珠。在一个实例中,将Fe_3O_4纳米颗粒掺入含聚电解质的珠中在振荡磁场刺激物的存在下触发破裂。

[0129] 珠也可能由于电刺激而破坏、溶解或降解。与先前部分中描述的磁性颗粒相似,电敏珠可以允许珠的触发破裂以及其他功能,诸如电场中的对准、电导或氧化还原反应。在一个实例中,含有电敏材料的珠在电场中对准,从而可以控制内部试剂的释放。在其他实例中,电场可以在珠壁本身内引起氧化还原反应,这可以增加孔隙率。

[0130] 也可用光刺激物使珠破坏。许多光触发物是可能的,并可以包括使用各种分子(诸如能够吸收特定波长范围的光子的纳米颗粒和发色团)的系统。例如,金属氧化物涂层可用作囊触发物。涂覆有SiO_2的聚电解质囊的UV照射可导致珠壁的崩解。在又另一个实例中,可以将可光转换的材料,诸如偶氮苯基团,掺入珠壁中。在施加UV或可见光后,诸如这些的化学物质在吸收光子后经历可逆的顺式-至-反式异构化。在此方面,包括光转换导致珠壁在施加光触发物后可崩解或变得更为多孔。

[0131] 例如,在图2中图示的条形码化(例如,随机条形码化)的非限制性实例中,在框208处将细胞(诸如单细胞)引入微孔阵列的多于一个微孔之后,在框212处可以将珠引入微孔阵列的多于一个微孔。每个微孔可包含一个珠。珠可以包含多于一个条形码。条形码可以包含附接到珠的5’胺区域。条形码可以包含通用标记、条形码序列(例如,分子标记)、靶结合区,或它们的任何组合。

[0132] 本文公开的条形码可以与固体支持物(例如,珠)关联(例如,附接)。与固体支持物关联的条形码可以各自包含选自包括以下的组的条形码序列:具有独特序列的至少100个或1000个条形码序列。在一些实施方案中,与固体支持物关联的不同条形码可以包括具有不同序列的条形码。在一些实施方案中,一定百分比的与固体支持物关联的条形码包含相同的细胞标记。例如,所述百分比可以是以下,或是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%,或在这些值中的任何两个值之间的数字或范围。作为另一个实例,所述百分比可以是至少以下,或至多以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。在一些实施方案中,与固体支持物关联的条形码可以具有相同的细胞标记。与不同固体支持物关联的条形码可以具有选自包括以下的组的不同的细胞标记:具有独特序列的至少100个或1000个细胞标记。

[0133] 本文公开的条形码可以与固体支持物(例如,珠)关联(例如,附接)。在一些实施方案中,可以用包括与多于一个条形码关联的多于一个合成的颗粒的固体支持物对样品中的多于一个靶进行条形码化。在一些实施方案中,固体支持物可包括与多于一个条形码关联的多于一个合成的颗粒。不同固体支持物上的多于一个条形码的空间标记可以具有至少一个核苷酸的差异。固体支持物可以例如包括处于二维或三维的多于一个条形码。合成的颗粒可以是珠。珠可以是硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、交联葡聚糖/琼脂糖凝胶

珠、珠状纤维素、聚苯乙烯珠，或它们的任何组合。固体支持物可包括聚合物、基质、水凝胶、针阵列装置、抗体，或它们的任何组合。在一些实施方案中，固体支持物可以自由浮动。在一些实施方案中，固体支持物可嵌入半固体或固体阵列中。条形码可以不与固体支持物关联。条形码可以是单独的核苷酸。条形码可与基底关联。

[0134] 如本文使用的，术语“拴系”、“附接”和“固定”可以互换使用，并且可以指用于将条形码附接至固体支持物的共价或非共价手段(means)。可以将各种不同的固体支持物中的任何一种用作固体支持物，以用于附接预先合成的条形码或用于条形码的原位固相合成。

[0135] 在一些实施方案中，固体支持物是珠。珠可以包括一种或更多种类型的实心的、多孔的或空心的球体、球、承座、圆柱体或其他相似形状，其上可以固定核酸(例如，共价地或非共价地)。珠可以例如由塑料、陶瓷、金属、聚合物材料，或它们的任何组合构成。珠可以是，或包括，球形的(例如，微球)或具有非球形或不规则形状的离散颗粒，所述形状是诸如立方体、长方体、锥体、圆柱体、圆锥体、椭圆形或圆盘形等。在一些实施方案中，珠的形状可以是非球形的。

[0136] 珠可以包括各种材料，包括但不限于顺磁性材料(例如，镁、钼、锂和钽)、超顺磁性材料(例如，铁氧体(Fe_3O_4;磁铁矿)纳米颗粒)、铁磁材料(例如，铁、镍、钴，它们的一些合金，以及一些稀土金属化合物)、陶瓷、塑料、玻璃、聚苯乙烯、二氧化硅、甲基苯乙烯、丙烯酸聚合物、钛、胶乳、交联琼脂糖、琼脂糖、水凝胶、聚合物、纤维素、尼龙，或它们的任何组合。

[0137] 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。

[0138] 本文公开的一些实施方案包括一个或更多个颗粒(例如，珠)。颗粒中的每一个可以包括多于一个寡核苷酸(例如，条形码)。多于一个寡核苷酸中的每一个可以包括条形码序列(例如，分子标记序列)、细胞标记和靶结合区(例如，寡聚(dT)序列、基因特异性序列、随机多聚体，或它们的组合)。多于一个寡核苷酸中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量可以是不同的。在一些实施方案中，细胞标记序列的数量可以是以下，或是约以下：10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10^6、10^7、10^8、10^9，在这些值中的任何两个值之间的数字或范围，或更多。在一些实施方案中，细胞标记序列的数量可以是至少以下，或至多以下：10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10^6、10^7、10^8或10^9。在一些实施方案中，多于一个颗粒中不超过1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个或更多个包括具有相同细胞序列的寡核苷酸。在一些实施方案中，包括具有相同细胞序列的寡核苷酸的多于一个颗粒可以是至多0.1％、0.2％、0.3％、0.4％、0.5％、0.6％、0.7％、0.8％、0.9％、1％、2％、3％、4％、5％、6％、7％、8％、9％、10％或更多。在一些实施方案中，多于一个颗粒中没有一个具有相同的细胞标记序列。

[0139] 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,条形码序列的数量可以是以下,或是约以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10^6、10^7、10^8、10^9,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,条形码序列的数量可以是至少以下,或至多以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、10^6、10^7、10^8或10^9。例如,多于一个寡核苷酸中的至少100个包含不同的条形码序列。作为另一个实例,在单个颗粒中,多于一个寡核苷酸中的至少100个、500个、1000个、5000个、10000个、15000个、20000个、50000个,这些值的任何两个之间的数字或范围,或更多个包含不同的条形码序列。一些实施方案提供了包含条形码的多于一个颗粒。在一些实施方案中,待标记的靶和不同条形码序列的出现(或拷贝或数量)的比率可以是至少1:1、1:2、1:3、1:4、1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:30、1:40、1:50、1:60、1:70、1:80、1:90或更多。在一些实施方案中,多于一个寡核苷酸中的每一个还包含样品标记、通用标记,或两者。颗粒可以是例如纳米颗粒或微颗粒。

[0140] 珠的尺寸可以变化。例如,珠的直径的范围可以是0.1微米至50微米。在一些实施方案中,珠的直径可以是以下,或是约以下:0.1微米、0.5微米、1微米、2微米、3微米、4微米、5微米、6微米、7微米、8微米、9微米、10微米、20微米、30微米、40微米或50微米,或这些值中的任何两个值之间的数字或范围。

[0141] 珠的直径可以与基底的孔的直径相关。在一些实施方案中,珠的直径可以比孔的直径长或短10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或这些值中的任何两个值之间的数字或范围,或比孔的直径长或短约10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或这些值中的任何两个值之间的数字或范围。珠的直径可以与细胞(例如,被基质的孔截留的单细胞)的直径相关。在一些实施方案中,珠的直径可以比孔的直径长或短至少或至多10%、20%、30%、40%、50%、60%、70%、80%、90%或100%。珠的直径可以与细胞(例如,被基质的孔截留的单细胞)的直径相关。在一些实施方案中,珠的直径可以比细胞的直径长或短10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或这些值中的任何两个值之间的数字或范围,或比细胞的直径长或短约10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或这些值中的任何两个值之间的数字或范围。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%。

[0142] 珠可以附接到和/或嵌入到基底中。珠可以附接到和/或嵌入到凝胶、水凝胶、聚合物和/或基质中。珠在基底(例如,凝胶、基质、支架或聚合物)中的空间位置可以使用珠上的条形码上存在的可以用作位置地址的空间标签来鉴定。

[0143] 珠的实例包括但不限于链霉抗生物素蛋白珠、琼脂糖珠、磁珠、Dynabeads®、MACS®微珠、抗体缀合珠(例如,抗免疫球蛋白微珠)、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠和BcMag™羧基封端磁珠。

[0144] 珠可以与量子点或荧光染料关联(例如,用量子点或荧光染料浸渍),以使其在一个荧光光学通道或多个光学通道中发荧光。珠可以与氧化铁或氧化铬关联,以使其为顺磁性或铁磁性。珠是可鉴定的。例如,可以使用照相机对珠进行成像。珠可以具有与珠关联的可检测编码。例如,珠可包含条形码。例如,由于在有机或无机溶液中的膨胀,珠可以改变尺寸。珠可以是疏水的。珠可以是亲水的。珠可以是生物相容的。

[0145] 固体支持物(例如珠)可以被可视化。固体支持物可以包含可视化标签(例如,荧光染料)。固体支持物(例如珠)可以用标识符(例如数字)蚀刻。标识符可以通过对珠进行成像而可视化。

[0146] 固体支持物可以包含半溶性或不溶性物质。当固体支持物包括与其连接的接头、支架、构建单元(building block)或其他反应性部分时,固体支持物可以被称为“官能化的”,而当固体支持物缺少这样的与其连接的反应性部分时,固体支持物可以被称为“非官能化的”。固体支持物可以在溶液中自由使用,诸如在微量滴定孔形式中;在流通形式中,诸如在列中;或者在浸量尺(dipstick)中。

[0147] 固体支持物可以包括膜、纸、塑料、涂层表面、平坦表面、玻璃、载玻片、芯片,或它们的任何组合。固体支持物可以采取树脂、凝胶、微球或其他几何形状的形式。固体支持物可以包括二氧化硅芯片、微粒、纳米颗粒、板、阵列、毛细管、平坦支持物诸如玻璃纤维过滤器、玻璃表面、金属表面(钢、金、银、铝、硅和铜)、玻璃支持物、塑料支持物、硅支持物、芯片、过滤器、膜、微孔板、载玻片、塑料材料,包括多孔板或膜(例如由聚乙烯、聚丙烯、聚酰胺、聚偏二氟乙烯形成的),和/或晶片(wafers)、梳、针(pins)或针头(needles)(例如适于组合合成或分析的针阵列),或平坦表面诸如晶片(例如硅晶片)的凹点(pit)或纳升孔的阵列中的珠,具有带有或不带有过滤器底的凹点的晶片。

[0148] 固体支持物可以包含聚合物基质(例如,凝胶、水凝胶)。聚合物基质可能能够渗透细胞内空间(例如,细胞器周围)。聚合物基质可能能够被泵送到整个循环系统。

[0149] 基底和微孔阵列

[0150] 如本文使用的,基底可以指一种类型的固体支持物。基底可以指可包含本公开内容的条形码或随机条形码的固体支持物。例如,基底可以包括多于一个微孔。例如,基底可以是包括两个或更多个微孔的孔阵列。在一些实施方案中,微孔可以包括限定体积的小反应室。在一些实施方案中,微孔可以截留一个或更多个细胞。在一些实施方案中,微孔仅能截留一个细胞。在一些实施方案中,微孔可以截留一种或更多种固体支持物。在一些实施方案中,微孔仅能截留一种固体支持物。在一些实施方案中,微孔截留单细胞和单个固体支持物(例如珠)。微孔可以包含本公开内容的条形码试剂。

[0151] 条形码化的方法

[0152] 本公开内容提供了用于估计身体样品(例如,组织、器官、肿瘤、细胞)中的不同位置处的不同靶的数量的方法。所述方法可以包括将条形码(例如,随机条形码)靠近样品放置,裂解样品,将不同靶与条形码关联,对靶进行扩增和/或对靶进行数字计数。所述方法还可以包括对从条形码上的空间标记获得的信息进行分析和/或可视化。在一些实施方案中,所述方法包括使样品中的多于一个靶可视化。将多于一个靶映射到样品的地图上可以包括生成样品的二维地图或三维地图。可以在对样品中的多于一个靶进行条形码化(例如,随机条形码化)之前或之后生成二维地图和三维地图。使样品中的多于一个靶可视化可以包括

将多于一个靶映射到样品的地图上。将多于一个靶映射到样品的地图上可以包括生成样品的二维地图或三维地图。可以在对样品中的多于一个靶进行条形码化之前或之后生成二维地图和三维地图。在一些实施方案中,可以在裂解样品之前或之后生成二维地图和三维地图。在生成二维地图或三维地图之前或之后裂解样品可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。

[0153] 在一些实施方案中,对多于一个靶进行条形码化包括使多于一个条形码与多于一个靶杂交以产生条形码化的靶(例如,随机条形码化的靶)。对多于一个靶进行条形码化可以包括生成经条形码化的靶的索引文库。生成经条形码化的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)的固体支持物进行。

[0154] 使样品和条形码接触

[0155] 本公开内容提供了用于使样品(例如,细胞)与本公开内容的基底接触的方法。可以使包括例如细胞、器官或组织薄切片的样品与条形码(例如,随机条形码)接触。例如,通过重力流可以接触所述细胞,其中可以使所述细胞沉淀并且产生单层。样品可以是组织薄切片。可以将薄切片置于基底上。样品可以是一维的(例如,形成平坦表面)。样品(例如,细胞)可以散布于基底上,例如,通过在基底上生长/培养所述细胞。

[0156] 当条形码靠近靶时,靶可以与条形码杂交。条形码可以按不可耗尽的比率接触,使得每个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。

[0157] 细胞裂解

[0158] 在细胞和条形码的分布之后,可以裂解细胞以释放靶分子。细胞裂解可以通过各种手段中的任何一种来完成,例如通过化学或生物化学手段,通过渗透冲击,或通过热裂解、机械裂解或光学裂解的手段。可以通过添加包含去污剂(例如,SDS、十二烷基硫酸锂、Triton X-100、吐温-20或NP-40)的细胞裂解缓冲液、有机溶剂(例如甲醇或丙酮)或消化酶(例如蛋白酶K、胃蛋白酶或胰蛋白酶),或它们的任何组合来裂解细胞。为了增加靶和条形码的关联,可通过例如降低裂解物的温度和/或增加裂解物的粘度来改变靶分子的扩散速率。

[0159] 在一些实施方案中,样品可以使用滤纸裂解。滤纸可以用滤纸之上的裂解缓冲液浸泡。滤纸可以用压力施加到样品上,该压力可以促进样品的裂解和样品的靶与基底的杂交。

[0160] 在一些实施方案中,裂解可以通过机械裂解、热裂解、光学裂解和/或化学裂解进行。化学裂解可以包括使用消化酶,诸如蛋白酶K、胃蛋白酶和胰蛋白酶。裂解可以通过将裂解缓冲液添加到基底来进行。裂解缓冲液可以包含Tris HCl。裂解缓冲液可以包含至少约0.01M、0.05M、0.1M、0.5M或1M或更多的Tris HCl。裂解缓冲液可以包含至多约0.01M、0.05M、0.1M、0.5M或1M或更多的Tris HCL。裂解缓冲液可以包含约0.1M Tris HCl。裂解缓冲液的pH可以是至少约1、2、3、4、5、6、7、8、9、10或更高。裂解缓冲液的pH可以是至多约1、2、3、4、5、6、7、8、9、10或更高。在一些实施方案中,裂解缓冲液的pH为约7.5。裂解缓冲液可以包含盐(例如LiCl)。裂解缓冲液中的盐浓度可以是至少约0.1、0.5或1M或更高。裂解缓冲液中的盐浓度可以是至多约0.1、0.5或1M或更高。在一些实施方案中,裂解缓冲液中的盐浓度为约0.5M。裂解缓冲液可以包含去污剂(例如SDS、十二烷基硫酸锂、triton X、吐温、NP-

40)。裂解缓冲液中去污剂的浓度可以是至少约0.0001％、0.0005％、0.001％、0.005％、0.01％、0.05％、0.1％、0.5％、1％、2％、3％、4％、5％、6％或7％或更多。裂解缓冲液中去污剂的浓度可以是至多约0.0001％、0.0005％、0.001％、0.005％、0.01％、0.05％、0.1％、0.5％、1％、2％、3％、4％、5％、6％或7％或更多。在一些实施方案中,裂解缓冲液中去污剂的浓度为约1％的十二烷基硫酸锂。裂解方法中使用的时间可以取决于使用的去污剂的量。在一些实施方案中,使用的去污剂越多,裂解所需的时间越少。裂解缓冲液可以包含螯合剂(例如,EDTA、EGTA)。裂解缓冲液中螯合剂的浓度可以是至少约1mM、5mM、10mM、15mM、20mM、25mM或30mM或更高。裂解缓冲液中螯合剂的浓度可以是至多约1mM、5mM、10mM、15mM、20mM、25mM或30mM或更高。在一些实施方案中,裂解缓冲液中螯合剂的浓度是约10mM。裂解缓冲液可以包含还原剂(例如,β-巯基乙醇、DTT)。裂解缓冲液中还原剂的浓度可以是至少约1mM、5mM、10mM、15mM或20mM或更高。裂解缓冲液中还原剂的浓度可以是至多约1mM、5mM、10mM、15mM或20mM或更高。在一些实施方案中,裂解缓冲液中还原试剂的浓度是约5mM。在一些实施方案中,裂解缓冲液可以包含约0.1M TrisHCl,约pH 7.5、约0.5M LiCl、约1％十二烷基硫酸锂、约10mM EDTA和约5mM DTT。

[0161] 裂解可以在约4℃、10℃、15℃、20℃、25℃或30℃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。

[0162] 将条形码附接至靶核酸分子

[0163] 在细胞裂解和核酸分子从细胞释放之后,核酸分子可以随机地与共定位的固体支持物的条形码关联。关联可以包括使条形码的靶识别区与靶核酸分子的互补部分杂交(例如,条形码的寡聚(dT)可以与靶的多(A)尾相互作用)。可以选择用于杂交的测定条件(例如,缓冲液pH、离子强度、温度等)以促进形成特定的稳定的杂交体。在一些实施方案中,从裂解的细胞释放的核酸分子可以与基底上的多于一个探针关联(例如,与基底上的探针杂交)。当探针包括寡聚(dT)时,mRNA分子可以与探针杂交,并且进行逆转录。寡核苷酸的寡聚(dT)部分可以充当用于cDNA分子的第一链合成的引物。例如,在图2中图示的条形码化的非限制性实例中,在框216处,mRNA分子可以与珠上的条形码杂交。例如,单链的核苷酸片段可以与条形码的靶结合区杂交。

[0164] 附接还可以包括将条形码的靶识别区与靶核酸分子的一部分连接。例如,靶结合区可以包括可能能够与限制性位点突出端(例如,EcoRI粘性末端突出端)进行特异性杂交的核酸序列。测定程序还可以包括用限制性酶(例如,EcoRI)处理靶核酸以产生限制性位点突出端。然后条形码可以连接到包括与限制性位点突出端互补的序列的任何核酸分子。连接酶(例如,T4DNA连接酶)可以用于连接两个片段。

[0165] 例如,在图2中图示的条形码化的非限制性实例中,在框220处,来自多于一个细胞(或多于一个样品)的经标记的靶(例如,靶-条形码分子)可以随后被汇集,例如汇集至管中。经标记的靶可以通过例如将条形码和/或附接靶-条形码分子的珠取回(retrieving)来汇集。

[0166] 附接的靶-条形码分子的基于固体支持物的集合的取回可以通过使用磁珠和外部

施加的磁场来实施。在所述靶-条形码分子已经汇集后,所有进一步的处理可以在单个反应容器中进行。进一步的处理可以包括例如逆转录反应、扩增反应、裂解反应、解离反应和/或核酸延伸反应。进一步的处理反应可以在微孔内进行,即,不首先汇集来自多于一个细胞的经标记的靶核酸分子。

[0167] 逆转录

[0168] 本公开内容提供了使用逆转录来产生靶-条形码缀合物的方法(在图2的框224处)。靶-条形码缀合物可以包括条形码以及靶核酸(即,条形码化cDNA分子,诸如随机条形码化cDNA分子)的全部或一部分的互补序列。关联的RNA分子的逆转录可以通过添加逆转录引物连同逆转录酶一起而发生。逆转录引物可以是寡聚(dT)引物、随机六核苷酸引物或靶特异性寡核苷酸引物。寡聚(dT)引物的长度可以是12-18个核苷酸,或可以是约12-18个核苷酸,并与哺乳动物mRNA的3’末端处的内源性多(A)尾结合。随机六核苷酸引物可在多个互补位点处与mRNA结合。靶特异性寡核苷酸引物通常选择性地引发感兴趣的mRNA。

[0169] 在一些实施方案中,标记的RNA分子的逆转录可以通过添加逆转录引物而发生。在一些实施方案中,所述逆转录引物是寡聚(dT)引物、随机六核苷酸引物或靶特异性寡核苷酸引物。通常,寡聚(dT)引物的长度为12-18个核苷酸,并与哺乳动物mRNA的3’末端处的内源性多(A)尾结合。随机六核苷酸引物可以在多个互补位点处与mRNA结合。靶特异性寡核苷酸引物通常选择性地引发感兴趣的mRNA。

[0170] 逆转录可以重复地发生以产生多个经标记的cDNA分子。本文公开的方法可以包括进行至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个逆转录反应。所述方法可以包括进行至少约25个、30个、35个、40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个逆转录反应。

[0171] 扩增

[0172] 可以进行一个或更多个核酸扩增反应(例如,在图2的框228处)以产生经标记的靶核酸分子的多个拷贝。扩增可以以多重方式(in a multiplexed manner)进行,其中多个靶核酸序列同时进行扩增。扩增反应可用于将测序衔接子添加至核酸分子。扩增反应可以包括扩增样品标记(如果存在)的至少一部分。扩增反应可以包括扩增细胞标记和/或条形码序列(例如,分子标记)的至少一部分。扩增反应可以包括扩增样品标签、细胞标记、空间标记、条形码序列(例如,分子标记)、靶核酸,或它们的组合的至少一部分。扩增反应可以包括扩增多于一个核酸的0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、100%,或这些值中的任何两个值之间的数字或范围。所述方法还可以包括进行一个或更多个cDNA合成反应以产生包含样品标记、细胞标记、空间标记和/或条形码序列(例如,分子标记)的靶-条形码分子的一个或更多个cDNA拷贝。

[0173] 在一些实施方案中,可以使用聚合酶链式反应(PCR)进行扩增。如本文使用的,PCR可以指用于通过DNA的互补链的同时引物延伸使特定DNA序列体外扩增的反应。如本文使用的,PCR可以包括所述反应的派生形式,包括但不限于RT-PCR、实时PCR、巢式PCR、定量PCR、多重PCR、数字PCR和组装PCR。

[0174] 经标记的核酸的扩增可以包括非基于PCR的方法。非基于PCR的方法的实例包括但不限于多重置换扩增(MDA)、转录介导的扩增(TMA)、基于核酸序列的扩增(NASBA)、链置换

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

[0175] 在一些实施方案中,本文公开的方法还包括对经标记的核酸(例如,经标记的RNA、经标记的DNA、经标记的cDNA)进行聚合酶链式反应,以产生经标记的扩增子(例如,经随机标记的扩增子)。所述经标记的扩增子可以是双链分子。双链分子可以包括双链RNA分子、双链DNA分子或者与DNA分子杂交的RNA分子。双链分子的一条或两条链可以包含样品标记、空间标记、细胞标记和/或条形码序列(例如,分子标记)。所述经标记的扩增子可以是单链分子。单链分子可以包括DNA、RNA或它们的组合。本公开内容的核酸可以包括合成的或改变的核酸。

[0176] 扩增可以包括使用一个或更多个非天然核苷酸。非天然核苷酸可以包括光不稳定或可触发的核苷酸。非天然核苷酸的实例可以包括但不限于肽核酸(PNA)、吗啉代核酸和锁核酸(LNA),以及二醇核酸(GNA)与苏糖核酸(TNA)。可以将非天然核苷酸添加至扩增反应的一个或更多个循环中。添加非天然核苷酸也可以用于鉴定扩增反应中特定循环或时间点的产物。

[0177] 进行一个或更多个扩增反应可以包括使用一个或更多个引物。一个或更多个引物可以包括例如,1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个或15个或更多个核苷酸。一个或更多个引物可以包括至少1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个或15个或更多个核苷酸。一个或更多个引物可以包括少于12-15个核苷酸。一个或更多个引物可以退火至多于一个经标记的靶(例如,随机标记的靶)的至少一部分。一个或更多个引物可以退火至多于一个经标记的靶的3'末端和/或5'末端。一个或更多个引物可以退火至多于一个经标记的靶的内部区域。内部区域可以与多于一个经标记的靶的3'末端距离至少约50个、100个、150个、200个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸。一个或更多个引物可以包括一组固定的引物。一个或更多个引物可以包括至少一个或更多个定制引物。一个或更多个引物可以包括至少一个或更多个对照引物。一个或更多个引物可以包括至少一个或更多个基因特异性引物。

[0178] 一个或更多个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。一个或更多个定制引物可以退火至第一样品标记、第二样品标记、空间标记、细胞标记、条形码序列(例如,分子标记)、靶,或它们的任何组合。一个或更多个引物可以包括通用引物和定制引物。定制引物可以被设计用于扩增一个或更多个靶。靶可以包括一个或更多个样品中总核酸的子集。靶可以包括一个或更多个样品中总的经标记的靶的子集。一个或更多个引物可以包括至少96个或更多个定制引物。一个或更多个引物可以包括至少960个或更

多个定制引物。一个或更多个引物可以包括至少9600个或更多个定制引物。一个或更多个定制引物可以退火至两个或更多个不同的经标记的核酸。两个或更多个不同的经标记的核酸可以对应于一个或更多个基因。

[0179] 可以在本公开内容的方法中使用任何扩增方案。例如,在一个方案中,第一轮PCR可以使用基因特异性引物和针对通用Illumina测序引物1序列的引物来扩增附接至珠的分子。第二轮PCR可以使用侧翼为Illumina测序引物2序列的巢式基因特异性引物,和针对通用Illumina测序引物1序列的引物扩增第一PCR产物。第三轮PCR添加P5和P7以及样品索引,以使PCR产物转变为Illumina测序文库。使用150bp x 2测序进行测序可以揭示读段(read)1上的细胞标记和条形码序列(例如,分子标记)、读段2上的基因,以及索引1读段上的样品索引。

[0180] 在一些实施方案中,核酸可以使用化学裂解从基底去除。例如,存在于核酸中的化学基团或经修饰的碱基可以用于促进将其从固体支持物去除。例如,酶可以用于将核酸从基底去除。例如,核酸可以通过限制性核酸内切酶消化从基底去除。例如,使用尿嘧啶-d-糖基化酶(UDG)处理含有dUTP或ddUTP的核酸可以将核酸从基底去除。例如,核酸可以使用进行核苷酸切除的酶(诸如,碱基切除修复酶,诸如无嘌呤/无嘧啶(AP)核酸内切酶)将核酸从基底去除。在一些实施方案中,核酸可以使用可光裂解(photocleavable)基团以及光从基底去除。在一些实施方案中,可以使用可裂解接头将核酸从基底去除。例如,可裂解接头可以包括以下中的至少一种:生物素/亲和素、生物素/链霉抗生物素蛋白、生物素/中性链亲和素(neutravidin)、Ig蛋白A、光不稳定性接头、酸或碱不稳定性接头基团,或衔接子。

[0181] 当探针是基因特异性时,可以使分子与探针杂交,并且进行逆转录和/或扩增。在一些实施方案中,在核酸已经合成(例如,逆转录)之后,可以将其扩增。扩增可以以多重方式进行,其中多个靶核酸序列同时进行扩增。扩增可以将测序衔接子添加至核酸。

[0182] 在一些实施方案中,可以例如用桥式扩增在基底上进行扩增。可以对cDNA加同聚物尾,以便产生相容末端,用于使用基底上的寡聚(dT)探针进行桥式扩增。在桥式扩增中,与模板核酸的3’末端互补的引物可以是共价附接到固体颗粒的每对引物的第一引物。当含有模板核酸的样品与颗粒接触并进行单个热循环时,可以使模板分子退火至第一引物,并且第一引物通过添加核苷酸而在正向方向上延伸以形成双链体分子,所述双链体分子由模板分子和与模板互补的新形成的DNA链组成。在下一循环的加热步骤中,可以使双链体分子变性,从颗粒释放模板分子,并通过第一引物将互补DNA链附接到颗粒。在随后的退火和延伸步骤的退火阶段中,互补链可以与第二引物杂交,所述第二引物在从第一引物去除的位置处与互补链的区段(segment)互补。这种杂交可以导致互补链在第一引物和第二引物之间形成桥,所述桥通过共价键固定到第一引物并通过杂交固定到第二引物。在延伸阶段,通过在相同的反应混合物中添加核苷酸,第二引物可以在反向方向上延伸,从而将桥转化为双链桥。然后开始下一个循环,并且可以使双链桥变性以产生两个单链核酸分子,每个单链核酸分子的一个末端分别经第一引物和第二引物附接到颗粒表面,其中每个单链核酸分子的另一个末端是未附接的。在该第二个循环的退火和延伸步骤中,每条链可以在相同的颗粒上与先前未使用的另外的互补引物杂交,以形成新的单链桥。使现在杂交的两个先前未使用的引物延伸,以将两个新桥转换成双链桥。

[0183] 扩增反应可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％。

[0184] 对经标记的核酸的扩增可以包括基于PCR的方法或非基于PCR的方法。对经标记的核酸的扩增可以包括对经标记的核酸的指数式扩增。对经标记的核酸的扩增可以包括对经标记的核酸的线性扩增。扩增可以通过聚合酶链式反应(PCR)进行。PCR可以指用于通过DNA的互补链的同时引物延伸使特定DNA序列体外扩增的反应。PCR可涵盖所述反应的派生形式,包括但不限于,RT-PCR、实时PCR、巢式PCR、定量PCR、多重PCR、数字PCR、阻抑PCR、半阻抑PCR和组装PCR。

[0185] 在一些实施方案中,对经标记的核酸的扩增包括非基于PCR的方法。非基于PCR的方法的实例包括但不限于多重置换扩增(MDA)、转录介导的扩增(TMA)、基于核酸序列的扩增(NASBA)、链置换扩增(SDA)、实时SDA、滚环扩增或环到环扩增。其他非基于PCR的扩增方法包括DNA依赖性RNA聚合酶驱动的RNA转录扩增或RNA指导的DNA合成和转录的多个循环以扩增DNA或RNA靶、连接酶链式反应(LCR)、Qβ复制酶(Qβ)、回文探针的使用、链置换扩增、使用限制性核酸内切酶的寡核苷酸驱动的扩增、使引物与核酸序列杂交并且将所得双链体在延伸反应和扩增之前裂解的扩增方法、使用缺乏5’核酸外切酶活性的核酸聚合酶的链置换扩增、滚环扩增和/或分支延伸扩增(RAM)。

[0186] 在一些实施方案中,本文公开的方法还包括对扩增的扩增子(例如,靶)进行巢式聚合酶链式反应。扩增子可以是双链分子。双链分子可以包括双链RNA分子、双链DNA分子或者与DNA分子杂交的RNA分子。双链分子的一条或两条链可以包含样品标签或分子标识符标记。可替代地,扩增子可以是单链分子。单链分子可以包括DNA、RNA或它们的组合。本发明的核酸可以包括合成的或改变的核酸。

[0187] 在一些实施方案中,所述方法包括反复扩增经标记的核酸以产生多个扩增子。本文公开的方法可以包括进行至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个扩增反应。可替代地,所述方法包括进行至少约25个、30个、35个、40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个扩增反应。

[0188] 扩增还可以包括将一个或更多个对照核酸添加至一个或更多个包括多于一个核酸的样品中。扩增还可以包括将一个或更多个对照核酸添加至多于一个核酸中。对照核酸可以包括对照标记。

[0189] 扩增可以包括使用一个或更多个非天然核苷酸。非天然核苷酸可以包括光不稳定和/或可触发的核苷酸。非天然核苷酸的实例包括但不限于肽核酸(PNA)、吗啉代核酸和锁核酸(LNA),以及二醇核酸(GNA)与苏糖核酸(TNA)。可以将非天然核苷酸添加至扩增反应的一个或更多个循环中。添加非天然核苷酸可以用于鉴定扩增反应中特定循环或时间点的产物。

[0190] 进行一个或更多个扩增反应可以包括使用一个或更多个引物。一个或更多个引物可以包括一个或更多个寡核苷酸。一个或更多个寡核苷酸可以包括至少约7-9个核苷酸。一个或更多个寡核苷酸可以包括少于12-15个核苷酸。一个或更多个引物可以退火至多于一个经标记的核酸的至少一部分。一个或更多个引物可以退火至多于一个经标记的核酸的3’末端和/或5’末端。一个或更多个引物可以退火至多于一个经标记的核酸的内部区域。内部

区域可以与多于一个经标记的核酸的3’末端距离至少约50个、100个、150个、200个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸。一个或更多个引物可以包括一组固定的引物。一个或更多个引物可以包括至少一个或更多个定制引物。一个或更多个引物可以包括至少一个或更多个对照引物。一个或更多个引物可以包括至少一个或更多个管家基因引物。一个或更多个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。一个或更多个定制引物可以退火至第一样品标签、第二样品标签、分子标识符标记、核酸或它们的产物。一个或更多个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个或更多个靶核酸。靶核酸可以包括一个或更多个样品中总核酸的子集。在一些实施方案中,引物是附接到本公开内容的阵列的探针。

[0191] 在一些实施方案中,使样品中的多于一个靶条形码化(例如,随机条形码化)还包括生成条形码化的靶(例如,随机条形码化的靶)的索引文库或所述靶的条形码化的片段的索引文库。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。生成条形码化的靶的索引文库包括从样品中的多于一个靶生成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的靶的索引文库,第一索引多核苷酸的标记区与第二索引多核苷酸的标记区可以具有以下,具有约以下,具有至少以下,或具有至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个核苷酸的差异,或这些值中的任何两个值之间的数字或范围的核苷酸的差异。在一些实施方案中,生成条形码化的靶的索引文库包括使多于一个靶(例如mRNA分子)与包括多(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各自包含cDNA区和标记区的单链标记的cDNA分子,其中多于一个靶包括不同序列的至少两个mRNA分子,且多于一个寡核苷酸包括不同序列的至少两个寡核苷酸。生成条形码化的靶的索引文库还可以包括扩增单链标记的cDNA分子以产生双链标记的cDNA分子;以及对双链标记的cDNA分子进行巢式PCR以产生标记的扩增子。在一些实施方案中,所述方法可以包括产生衔接子标记的扩增子。

[0192] 条形码化(例如,随机条形码化)可以包括使用核酸条形码或标签以对单个核酸(例如,DNA或RNA)分子进行标记。在一些实施方案中,其涉及将DNA条形码或标签添加至cDNA分子,因为cDNA分子是从mRNA产生的。可以进行巢式PCR以使PCR扩增偏倚最小化。可以使用例如下一代测序(NGS)添加衔接子,用于测序。例如在图2的框232处,可以使用测序结果以确定靶的一个或更多个拷贝的细胞标记、分子标记,和核苷酸片段的序列。

[0193] 图3是示出了生成条形码化的靶(例如,随机条形码化的靶)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方法的示意图。如步骤1中示出的,逆转录过程可以编码具有独特的分子标记序列、细胞标记序列和通用PCR位点的每个mRNA分子。特别地,通过使一组条形码(例如,随机条形码)310与RNA分子302的多(A)尾区308杂交(例如,随机杂交),可以将RNA分子302逆转录以产生经标记的cDNA分子304(包括cDNA区306)。条形码310中的每一个可以包含靶结合区,例如多(dT)区312、标记区314(例如,条形码序列或分子)和通用PCR区316。

[0194] 在一些实施方案中,细胞标记序列可以包括3个至20个核苷酸。在一些实施方案中,分子标记序列可以包括3个至20个核苷酸。在一些实施方案中,多于一个随机条形码中的每个还包含通用标记和细胞标记中的一个或更多个,其中固体支持物上的多于一个随机条形码的通用标记是相同的,并且固体支持物上的多于一个随机条形码的细胞标记是相同的。在一些实施方案中,通用标记可以包括3个至20个核苷酸。在一些实施方案中,细胞标记包括3个至20个核苷酸。

[0195] 在一些实施方案中,标记区314可以包括条形码序列或分子标记318和细胞标记320。在一些实施方案中,标记区314可以包括通用标记、维度标记和细胞标记中的一个或更多个。条形码序列或分子标记318的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。细胞标记320的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。通用标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。固体支持物上的多于一个随机条形码的通用标记可以是相同的,并且固体支持物上的多于一个随机条形码的细胞标记是相同的。维度标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个,或这些值中的任何两个值之间的数字或范围的核苷酸。

[0196] 在一些实施方案中,标记区314可以包括以下,包括约以下,包括至少以下,或包括至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个,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。每个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。一组条形码或随机条形码310可以含有以下,含有约以下,含有至少以下,或可以是至多以下:10个、20个、40个、50个、70个、80个、90个、10^2个、10^3个、10^4个、10^5个、10^6个、10^7个、10^8个、10^9个、10^{10}个、10^{11}个、10^{12}个、10^{13}个、10^{14}个、10^{15}个、10^{20}个,或这些值中的任何两个值之间的数字或范围的条形码或随机条形码310。并且一组条形码或随机条形码310可以例如,各自含有独特标记区314。经标记的cDNA分子304可以进行纯化以去除过量条形码或随机条形码310。纯化可以包括Ampure珠纯化。

[0197] 如步骤2中示出的,来自步骤1中的逆转录过程的产物可以汇集至1支管中,并且用第1PCR引物池和第1通用PCR引物进行PCR扩增。由于独特标记区314,汇集是可能的。特别地,可以将经标记的cDNA分子304扩增以产生巢式PCR标记的扩增子322。扩增可以包括多重PCR扩增。扩增可以包括在单一反应体积中用96种多重引物进行的多重PCR扩增。在一些实施方案中,在单一反应体积中,多重PCR扩增可以利用以下,利用约以下,利用至少以下,或

利用至多以下:10个、20个、40个、50个、70个、80个、90个、10^2个、10^3个、10^4个、10^5个、10^6个、10^7个、10^8个、10^9个、10^{10}个、10^{11}个、10^{12}个、10^{13}个、10^{14}个、10^{15}个、10^{20}个,或这些值中的任何两个值之间的数字或范围的多重引物。扩增可以包括使用包括靶向特异性基因的定制引物326A-C的第1PCR引物池324和通用引物328。定制引物326可以与经标记的cDNA分子304的cDNA部分306'内的区域杂交。通用引物328可以与经标记的cDNA分子304的通用PCR区域316杂交。

[0198] 如图3的步骤3中示出的,来自步骤2中的PCR扩增的产物可以用巢式PCR引物池和第2通用PCR引物进行扩增。巢式PCR可以使PCR扩增偏倚最小化。特别地,巢式PCR标记的扩增子322可以通过巢式PCR进行进一步扩增。巢式PCR可以包括在单个反应体积中用巢式PCR引物332a-c的巢式PCR引物池330和第2通用PCR引物328'进行的多重PCR。巢式PCR引物池330可以含有以下,含有约以下,含有至少以下,或含有至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个,或这些值中的任何两个值之间的数字或范围的不同巢式PCR引物332。巢式PCR引物332可以含有衔接子334,并与经标记的扩增子322的cDNA部分306"内的区域杂交。通用引物328'可以含有衔接子336,并与经标记的扩增子322的通用PCR区316杂交。因此,步骤3产生衔接子标记的扩增子338。在一些实施方案中,巢式PCR引物332和第2通用PCR引物328'可以不含有衔接子334和336。相反,衔接子334和336可以连接到巢式PCR的产物以产生衔接子标记的扩增子338。

[0199] 如步骤4中示出的,来自步骤3的PCR产物可以使用文库扩增引物进行PCR扩增用于测序。特别地,衔接子334和336可以用于对衔接子标记的扩增子338进行一个或更多个另外的测定。衔接子334和336可以与引物340和342杂交。一个或更多个引物340和342可以是PCR扩增引物。一个或更多个引物340和342可以是测序引物。一个或更多个衔接子334和336可以用于衔接子标记的扩增子338的进一步扩增。一个或更多个衔接子334和336可以用于对衔接子标记的扩增子338进行测序。引物342可以含有板索引344,使得使用同一组条形码或随机条形码310产生的扩增子可以在一轮测序反应中使用下一代测序(NGS)进行测序。

[0200] 在核酸靶的5'末端的条形码化

[0201] 本文的公开内容包括用于将具有分子标记(或分子索引)的条形码(例如随机条形码)附接到被条形码化或标记的核酸靶(例如脱氧核糖核酸分子和核糖核酸分子)的5'-末端的系统、方法、组合物和试剂盒。本文公开的基于5'的转录物计数方法可以补足或补充例如基于3'的转录物计数方法(例如Rhapsody™测定(Becton,Dickinson and Company(Franklin Lakes,NJ))、Chromium™单细胞3'方案(10X Genomics(San Francisco,CA)))。条形码化的核酸靶可以以高通量方式用于序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序。对5'-末端(相对于被标记的核酸靶的5')的转录物计数可以揭示在核酸分子的5'末端或更靠近核酸分子的5'末端的选择性剪接异构体和变体(包括但不限于剪接变体、单核苷酸多态性(SNP)、插入、缺失、取代)。在一些实施方案中,所述方法可以包括分子内杂交。

[0202] 图4A-图4B示出了在5'-末端对核酸靶进行基因特异性标记的非限制性示例性方法400的示意性图示。具有靶结合区(例如,多(dT)尾422)的条形码420(例如,随机条形码)可以经由多(dA)尾426与多腺苷酸化RNA转录物424或其他核酸靶结合,用于标记或条形码

化(例如,进行独特标记)。条形码420可以包括分别用于标记转录物424和追踪RNA转录物424的样品来源的分子标记(ML)428和样品标记(SL)430,以及位于每个条形码420的分子标记428/样品标记430区域侧翼的一起用于随后的反应的一个或更多个另外的序列(例如,共有序列,诸如衔接子序列432)。每个样品的条形码中分子标记的序列的组库可以是对于对RNA转录物进行随机标记而言足够大的。

[0203] 在框402处的cDNA合成以产生包含RNA转录物424(或其一部分)的条形码化cDNA分子434之后,基因特异性方法可以用于5'分子条形码化。在框404处的基因特异性扩增(可以是任选的)之后,可以在框406处添加末端转移酶和脱氧腺苷三磷酸(dATP),以促进加3'多(dA)尾,以产生具有多(A)尾438的扩增子436。框408处的短变性步骤允许扩增子436的正向链436m和反向链436c(例如,具有多(dA)尾的条形码化cDNA分子)的分离。扩增子436的反向链436c可以经由其3'末端的多(dA)尾438和链的5'末端的多(dT)区422进行分子内杂交,以在框410处形成发夹或茎环440。然后,在框412处,聚合酶(例如Klenow片段)可以用于从多(dA)尾438延伸,以复制条形码,从而形成延伸的条形码化反向链442。然后可以在框414处(例如,任选地)进行基因特异性扩增,以扩增感兴趣的基因,从而产生在5'末端(相对于RNA转录物424)具有条形码的扩增子444,用于在框416处进行测序。在一些实施方案中,方法400包括以下中的一种或两种:框404处对条形码化cDNA分子434的基因特异性扩增,和框414处对延伸的条形码化反向链442的基因特异性扩增。

[0204] 图5A-图5B示出了在5'-末端上对核酸靶进行标记用于全转录组分析的非限制性示例性方法500的示意性图示。具有靶结合区(例如,多(dT)尾422)的条形码420(例如,随机条形码)可以经由多(dA)尾426与多腺苷酸RNA转录物424或其他核酸靶结合,用于标记或条形码化(例如,进行独特标记)。例如,具有靶结合区的条形码420可以与核酸靶结合,用于标记或条形码化。条形码420可以包括分子标记(ML)428和样品标记(SL)430。分子标记428和样品标记430可以分别用于标记转录物424或核酸靶(例如,抗体寡核苷酸,无论其与抗体缔和还是已经与抗体解离)和追踪转录物424的样品来源,与位于每个条形码420的分子标记428/样品标记430区域侧翼的一个或更多个另外的序列(例如,共有序列,诸如衔接子序列432)一起用于随后的反应。每个样品的条形码中分子标记428的序列的组库可以是对于对RNA转录物424或核酸靶进行随机标记而言足够大的。

[0205] 在框402处的cDNA合成以产生条形码化cDNA分子434后,末端转移酶可以用于对条形码化cDNA分子434的3'末端(相当于标记的RNA转录物的5'末端)进行加A尾,以在框406处产生各自具有3'多(dA)尾438的cDNA分子436c。在框410处,可以启动cDNA分子436c与3'多(dA)尾438的分子内杂交(例如,通过热和冷却循环,或者通过用多(dA)尾438稀释条形码化cDNA分子436c),使得新的3'多(dA)尾438与相同的标记的cDNA分子的多(dT)尾422退火,以产生具有发夹或茎环结构440的条形码化cDNA分子。在框412处,可以添加聚合酶(例如Klenow酶)和dNTP,以促进新的3'多(dA)尾438之外的3'延伸,从而复制具有茎环440的标记的cDNA分子的5'末端上的条形码(例如,分子标记428)。在框414处,可以使用成镜像的衔接子432、432rc或包含衔接子432、432rc的序列(或子序列)的引物进行全转录组扩增(WTA)。方法,诸如标记片段化(tagmentation)或随机引发,可以用于产生具有测序衔接子(例如P5 446和P7 448序列)的扩增子444的较小片段,用于在框418处测序(例如,使用Illumina(San Diego,CA,U.S.)测序仪)。在一些实施方案中,用于其他测序方法或测序仪(例如,来自

Pacific Biosciences of California,Inc.(Menlo Park,CA,US)或Oxford Nanopore Technologies Limited(Oxford,UK)的测序仪)的测序衔接子可以直接连接以产生用于测序的扩增子。

[0206] 本文的公开内容包括用于确定样品中核酸靶的数量的方法。在一些实施方案中,该方法包括:使样品中核酸靶424的拷贝与多于一个寡核苷酸条形码420接触,其中多于一个寡核苷酸条形码420中的每一个包含分子标记序列428和能够与核酸靶424杂交的靶结合区(例如,多(dT)序列422),并且其中多于一个寡核苷酸条形码420中的至少10个包含不同的分子标记序列428;在框402处,使与寡核苷酸条形码420杂交的核酸靶424的拷贝延伸,以产生多于一个核酸分子434,每个核酸分子434包含与核酸靶424的至少一部分互补的序列450c;在框404处,扩增多于一个条形码化核酸分子434,以产生多于一个扩增的条形码化核酸分子436;在框406处,将包含靶结合区422的互补体438的寡核苷酸附接到多于一个扩增的条形码化核酸分子436,以产生多于一个条形码化核酸分子436c,每个条形码化核酸分子436c包含靶结合区422和靶结合区的互补体438;在框410处,使多于一个条形码化核酸分子436c中的每一个内的靶结合区422和靶结合区422的互补体438杂交,以形成茎环440;在框412处,使各自具有茎环440的多于一个条形码化核酸分子的3'-末端延伸,以使茎环440延伸,从而产生多于一个延伸的条形码化核酸分子442,每个延伸的条形码化核酸分子442包含分子标记428和分子标记的互补体428rc;在框414处,扩增多于一个延伸的条形码化核酸分子442,以产生多于一个单标记的核酸分子444c,每个单标记的核酸分子444c包含分子标记的互补体428rc;以及基于与多于一个单标记的核酸分子关联的具有不同序列的分子标记的互补体428rc的数量,确定样品中核酸靶的数量。

[0207] 在一些实施方案中,在使具有茎环440的多于一个条形码化核酸分子的3'末端延伸后,分子标记428与分子标记的互补体428rc杂交。该方法可以包括使多于一个延伸的条形码化核酸分子442变性,然后扩增多于一个延伸的条形码化核酸分子442,以产生多于一个单标记的核酸分子444c(其可以是扩增子444c的一部分)。使样品中核酸靶424的拷贝接触可以包括使多于一个核酸靶424的拷贝与多于一个寡核苷酸条形码420接触。使核酸靶424的拷贝延伸可以包括使与寡核苷酸条形码420杂交的多于一个核酸靶424的拷贝延伸,以产生多于一个条形码化核酸分子436c,每个条形码化核酸分子436c包括与多于一个核酸靶424中的一个的至少一部分互补的序列450c。确定核酸靶424的数量可以包括基于分子标记的互补体428rc的数量来确定样品中多于一个核酸靶424中的每一个的数量,所述分子标记与多于一个单标记的核酸分子444c中的单标记的核酸分子关联、具有不同序列,所述单标记的核酸分子444c包括多于一个核酸靶424中的每一个的序列452c。多于一个核酸靶中的每一个的序列452c可以包括多于一个核酸靶424中的每一个的子序列(包括互补体或反向互补体)。

[0208] 本文的公开内容包括用于确定样品中靶的数量的方法。在一些实施方案中,该方法包括:使用多于一个寡核苷酸条形码420对样品中的核酸靶424的拷贝进行条形码化402,以产生多于一个条形码化核酸分子434,每个条形码化核酸分子434包含核酸靶424的序列450c(例如互补序列、反向互补序列,或它们的组合)、分子标记428和靶结合区(例如多(dT)区422),并且其中多于一个寡核苷酸条形码420中的至少10个包含不同的分子标记序列428;将包含靶结合区422的互补体438的寡核苷酸附接406到多于一个条形码化核酸分子

434,以产生多于一个条形码化核酸分子436,每个条形码化核酸分子436包含靶结合区422和靶结合区422的互补体438;使多于一个条形码化核酸分子436c中的每一个内的靶结合区422和靶结合区的互补体438杂交410以形成茎环440;使多于一个条形码化核酸分子的3'末端延伸412,以使茎环440延伸,从而产生多于一个延伸的条形码化核酸分子442,每个延伸的条形码化核酸分子442包含分子标记428和分子标记的互补体428rc;以及基于与多于一个延伸的条形码化核酸分子442关联的具有不同序列的分子标记的互补体428rc的数量来确定样品中核酸靶424的数量。

[0209] 本文的公开内容包括用于将寡核苷酸条形码附接到样品中的靶的方法。在一些实施方案中,该方法包括:使用多于一个寡核苷酸条形码420对样品中的核酸靶424的拷贝进行条形码化402,以产生多于一个条形码化核酸分子434,每个条形码化核酸分子434包含核酸靶424的序列450c、分子标记428和靶结合区422,并且其中多于一个寡核苷酸条形码420中的至少10个包含不同的分子标记序列428;将包含靶结合区422的互补体438的寡核苷酸附接到多于一个条形码化核酸分子434,以产生多于一个条形码化核酸分子436c,每个条形码化核酸分子436c包含靶结合区422和靶结合区422的互补体438;使多于一个条形码化核酸分子436c中的每一个内的靶结合区422和靶结合区422的互补体438杂交410以形成茎环440;以及使多于一个条形码化核酸分子的3'末端延伸412,以使茎环440延伸,从而产生多于一个延伸的条形码化核酸分子442,每个延伸的条形码化核酸分子442包含分子标记428和分子标记428的互补体428rc。在一些实施方案中,该方法包括:基于与多于一个延伸的条形码化核酸分子442关联的具有不同序列的分子标记428、其互补体428rc,或它们的组合的数量,确定样品中核酸靶424的数量。例如,核酸靶424的数量可以基于具有不同序列的分子标记428、其互补体428rc中的一个或二者来确定。

[0210] 在一些实施方案中,该方法包括:对多于一个靶424的拷贝进行条形码化402,包括:使核酸靶424的拷贝与多于一个寡核苷酸条形码420接触,其中多于一个寡核苷酸条形码420中的每一个包含能够与核酸靶424杂交的靶结合区422;以及使与寡核苷酸条形码420杂交的核酸靶424的拷贝延伸402,以产生多于一个条形码化的核酸分子434。

[0211] 在一些实施方案中,该方法包括:扩增404多于一个条形码化核酸分子434以产生多于一个扩增的条形码化核酸分子436c,其中附接包含靶结合区422的互补体438的寡核苷酸包括:将包含靶结合区的互补体438的寡核苷酸附接到多于一个扩增的条形码化核酸分子以产生多于一个条形码化核酸分子436c,每个条形码化核酸分子436c包含靶结合区422和靶结合区的互补体438。

[0212] 基因特异性分析。在一些实施方案中,方法(例如,方法400)包括:扩增414多于一个延伸的条形码化核酸分子442,以产生多于一个单标记的核酸分子444c,每个单标记的核酸分子444c包含分子标记428的互补体428rc。单标记的核酸分子444c可以在包含它们的扩增子444变性时产生。确定样品中核酸靶424的数量可以包括:基于分子标记428的互补体428rc的数量确定样品中核酸靶424的数量,分子标记428与多于一个单标记的核酸分子444c关联、具有不同序列。

[0213] 全转录组分析。在一些实施方案中,方法(例如,方法500)包括:扩增414多于一个延伸的条形码化核酸分子442,以产生多于一个延伸的条形码化核酸分子的拷贝444c。确定样品中核酸靶424的数量包括:基于分子标记428的互补体428rc的数量确定样品中核酸靶

424的数量,分子标记428与多于一个延伸的条形码化核酸分子的拷贝444c关联、具有不同序列。多于一个延伸的条形码化核酸分子的拷贝444c可以在包含它们的扩增子444变性时形成。

[0214] 在一些实施方案中,多于一个条形码化核酸分子中的核酸靶的序列包含核酸靶的子序列452c。靶结合区可以包含基因特异性序列。附接406包含靶结合区422的互补体438的寡核苷酸可以包括将包含靶结合区422的互补体438的寡核苷酸连接到多于一个条形码化核酸分子434。

[0215] 在一些实施方案中,靶结合区可以包含多(dT)序列422。附接包含靶结合区422的互补体438的寡核苷酸包括:使用末端脱氧核苷酸转移酶将多于一个腺苷单磷酸添加到多于一个条形码化核酸分子434。

[0216] 在一些实施方案中,使与寡核苷酸条形码420杂交的核酸靶424的拷贝延伸可以包括将与寡核苷酸条形码420杂交的核酸靶424的拷贝逆转录,以产生多于一个条形码化互补脱氧核糖核酸(cDNA)分子434。使与寡核苷酸条形码420杂交的核酸靶424的拷贝延伸可以包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶使与寡核苷酸条形码420杂交的核酸靶424的拷贝延伸402。DNA聚合酶可以包括Klenow片段。

[0217] 在一些实施方案中,该方法包括:获得多于一个延伸的条形码化核酸分子442的序列信息。获得序列信息可以包括将测序衔接子(例如P5 446和P7 448衔接子)连接到多于一个延伸的条形码化核酸分子442。

[0218] 在一些实施方案中,靶结合区的互补体438可以包括靶结合区的反向互补序列。靶结合区的互补体438可以包括靶结合区的互补序列。分子标记的互补体428rc可以包括分子标记的反向互补序列。分子标记的互补体可以包括分子标记的互补序列。

[0219] 在一些实施方案中,多于一个条形码化核酸分子434可以包括条形码化脱氧核糖核酸(DNA)分子。条形码化核酸分子434可以包括条形码化核糖核酸(RNA)分子。核酸靶424可以包括核酸分子。核酸分子可以包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、含有多(A)尾的RNA,或它们的任何组合。

[0220] 抗体寡核苷酸。在一些实施方案中,核酸靶可以包含细胞组分结合试剂。在US2018/0088112和2018年3月27日提交的美国申请第15/937,713号中已经描述了与核酸靶相关的细胞结合试剂(例如,抗体寡核苷酸,诸如样品编索引寡核苷酸);这些申请中的每一个的内容通过引用以其整体并入本文。在一些实施方案中,可以使用本公开内容的5'条形码化方法获得单细胞的多组学信息,诸如基因组学、染色质可及性(chromatin accessibility)、甲基化组学(methylomics)、转录组学和蛋白质组学。核酸分子可以与细胞组分结合试剂关联。方法可以包括:使核酸分子和细胞组分结合试剂解离。

[0221] 在一些实施方案中,多于一个寡核苷酸条形码420的每个分子标记428包含至少6个核苷酸。寡核苷酸条形码420可以包含相同的样品标记430。多于一个寡核苷酸条形码420的每个样品标记430可以包含至少6个核苷酸。寡核苷酸条形码420可以包含相同的细胞标记。多于一个寡核苷酸条形码420的每个细胞标记可以包含至少6个核苷酸。

[0222] 在一些实施方案中,当使多于一个条形码化核酸分子中的每一个内的靶结合区和靶结合区的互补体杂交410以形成茎环时,多于一个条形码化核酸分子436c中的至少一个

与固体支持物关联。当使多于一个条形码化核酸分子436c中的每一个内的靶结合区422和靶结合区的互补体438杂交410以形成茎环440时,多于一个条形码化核酸分子436c中的至少一个可以与固体支持物解离。当使多于一个条形码化核酸分子436c中的每一个内的靶结合区422和靶结合区的互补体438杂交410以形成茎环440时,多于一个条形码化核酸分子436c中的至少一个可以与固体支持物关联。

[0223] 在一些实施方案中,当使多于一个条形码化核酸分子的3'-末端延伸412以使茎环440延伸从而产生多于一个延伸的条形码化核酸分子442时,多于一个条形码化核酸分子中的至少一个与固体支持物关联,每个延伸的条形码化核酸分子442包含分子标记428和分子标记的互补体428rc。当使多于一个条形码化核酸分子的3'-末端延伸412以使茎环440延伸从而产生多于一个延伸的条形码化核酸分子442时,多于一个条形码化核酸分子中的至少一个可以与固体支持物解离,每个延伸的条形码化核酸分子442包含分子标记428和分子标记的互补体428rc。当使多于一个条形码化核酸分子的3'-末端延伸412以使茎环440延伸从而产生多于一个延伸的条形码化核酸分子442时,多于一个条形码化核酸分子436c中的至少一个可以与固体支持物关联,每个延伸的条形码化核酸分子442包含分子标记428和分子标记的互补体428rc。固体支持物可以包括合成颗粒454。固体支持物可以包括平坦表面或基本平坦的表面(例如载玻片,诸如显微镜载玻片或盖玻片)。

[0224] 在一些实施方案中,当使多于一个条形码化核酸分子436c中的每一个内的靶结合区422和靶结合区422的互补体438杂交410以形成茎环440时,多于一个条形码化核酸分子436c中的至少一个在溶液中。例如,当溶液中多于一个条形码化核酸分子436c的浓度足够低时,可以发生这样的分子内杂交。当使多于一个条形码化核酸分子的3'-末端延伸412以使茎环440延伸从而产生多于一个延伸的条形码化核酸分子442时,多于一个条形码化核酸分子中的至少一个可以在溶液中,每个延伸的条形码化核酸分子442包含分子标记428和分子标记的互补体428rc。

[0225] 在一些实施方案中,样品包含单细胞,该方法包括使包含多于一个寡核苷酸条形码420的合成颗粒454与样品中的单细胞关联。该方法可以包括:在使合成颗粒454与单细胞关联后裂解单细胞。裂解单细胞可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。合成颗粒和单细胞可以在同一孔中。合成颗粒和单细胞可以在同一液滴中。

[0226] 在一些实施方案中,多于一个寡核苷酸条形码420中的至少一个可以被固定在合成颗粒454上。多于一个寡核苷酸条形码420中的至少一个可以被部分地固定在合成颗粒454上。多于一个寡核苷酸条形码420中的至少一个可以被包封在合成颗粒454中。多于一个寡核苷酸条形码420中的至少一个可以被部分地包封在合成颗粒454中。合成颗粒454可以是可破坏的。合成颗粒454可以包括珠。珠可以包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。合成颗粒454可以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、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。合成颗粒454可以包括可破坏的水凝胶颗粒。多于一个寡核苷酸条形码420中的每一个可以包含接头官能基团。合成

颗粒454可以包含固体支持物官能基团。支持物官能基团和接头官能基团可以彼此关联。接头官能基团和支持物官能基团可以单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮,以及它们的任何组合。

[0227] 用于核酸靶的5'末端上的条形码化的试剂盒

[0228] 本文的公开内容包括用于以下的试剂盒:将寡核苷酸条形码420附接到样品中的靶424、确定样品中靶424的数量和/或确定样品中核酸靶424的数量。在一些实施方案中,试剂盒包含:多于一个寡核苷酸条形码420,其中多于一个寡核苷酸条形码420中的每一个包含分子标记428和靶结合区(例如,多(dT)序列422),并且其中多于一个寡核苷酸条形码420中的至少10个包含不同的分子标记序列428;末端脱氧核苷酸转移酶或连接酶;和DNA聚合酶,所述DNA聚合酶缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种。DNA聚合酶可以包括Klenow片段。试剂盒可以包含缓冲液。试剂盒可以包含药筒。试剂盒可以包含用于逆转录反应的一种或更多种试剂。试剂盒可以包含用于扩增反应的一种或更多种试剂。

[0229] 在一些实施方案中,靶结合区包含基因特异性序列、寡聚(dT)序列、随机多聚体,或它们的任何组合。寡核苷酸条形码可以包含相同的样品标记和/或相同的细胞标记。多于一个寡核苷酸条形码的每个样品标记和/或细胞标记可以包含至少6个核苷酸。多于一个寡核苷酸条形码的每个分子标记可以包含至少6个核苷酸。

[0230] 在一些实施方案中,多于一个寡核苷酸条形码420中的至少一个被固定在合成颗粒454上。多于一个寡核苷酸条形码420中的至少一个可以被部分地固定在合成颗粒454上。多于一个寡核苷酸条形码420中的至少一个可以被包封在合成颗粒454中。多于一个寡核苷酸条形码420中的至少一个可以被部分地包封在合成颗粒454中。合成颗粒454可以是可破坏的。合成颗粒454可以包括珠。珠可以包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。合成颗粒可以包含选自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、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。合成颗粒454可以包括可破坏的水凝胶颗粒。多于一个寡核苷酸条形码中的每一个可以包含接头官能基团。合成颗粒454可以包含固体支持物官能基团。支持物官能基团和接头官能基团可以彼此关联。接头官能基团和支持物官能基团可以单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮,以及它们的任何组合。

[0231] 确定5'转录物序列

[0232] 高通量单细胞RNA测序改变了对复杂和异质生物样品的理解。然而,大多数方法只能对mRNA转录物信息进行3'分析,这可能限制对剪接变体、选择性转录起始位点和由于重排(诸如T细胞和B细胞受体以及抗体的VDJ连接)产生的高度可变基因座的分析。如本文公开的,使用BD Rhapsody平台以高通量方式捕获mRNA分子并生成转录物3'和5'末端两者的测序文库。

[0233] 本公开内容的方法可以用于鉴定B细胞受体(BCR)、T细胞受体(TCR)和抗体的VDJ区域。VDJ重组,也称为体细胞重组,是在免疫系统的免疫球蛋白(Ig)(例如BCR)和T细胞受

体(TCR)产生的早期中的遗传重组机制。VDJ重组几乎可以随机地将可变(V)基因区段、多样性(D)基因区段和连接(J)基因区段组合。由于它在选择不同基因时的随机性,它能够不同地编码蛋白质,以匹配来自细菌、病毒、寄生虫、功能障碍细胞诸如肿瘤细胞以及花粉的抗原。

[0234] VDJ区域可以包含大的3Mb基因座,该基因座包含可变(V)基因、多样性(D)基因和连接(J)基因。这些是可以参与VDJ重组的区段。可能有不经历VDJ重组的恒定基因。这个基因座的VDJ重组的第一个事件可能是一个D基因重排至一个J基因。在此之后,一个V基因可以附加到该DJ重排,以形成功能性VDJ重排基因,然后编码重链蛋白的可变区段。这两个步骤都可以由重组酶催化,重组酶可以使间插DNA缺失。

[0235] 这种重组过程在B祖细胞中以逐步的方式发生,以产生抗体组库所需的多样性。每个B细胞只能产生一种抗体(例如BCR)。这种特异性可以通过等位基因排斥实现,使得一个等位基因的功能性重排发出阻止第二个等位基因进一步重组的信号。

[0236] 在一些实施方案中,样品包括免疫细胞。免疫细胞可以包括,例如,T细胞、B细胞、淋巴样干细胞、骨髓祖细胞、淋巴细胞、粒细胞、B细胞祖细胞、T细胞祖细胞、自然杀伤细胞、Tc细胞、Th细胞、浆细胞、记忆细胞、嗜中性粒细胞、嗜酸性粒细胞、嗜碱性粒细胞、肥大细胞、单核细胞、树突细胞和/或巨噬细胞,或它们的任何组合。

[0237] T细胞可以是T细胞克隆,它可以指源自于单个T细胞的T细胞或具有相同TCR的T细胞。T细胞可以是T细胞系的一部分,该细胞系可以包括T细胞克隆和具有不同TCR的T细胞混合群体,所有所述不同TCR都可以识别相同的靶(例如,抗原、肿瘤、病毒)。T细胞可以从许多来源获得,包括外周血单核细胞、骨髓、淋巴结组织、脾脏组织和肿瘤。T细胞可以从收集自受试者的血液单位获得,诸如使用Ficoll分离获得。来自个体的循环血液的细胞可以通过单采血液成分术(apheresis)或单采白细胞术(leukapheresis)获得。单采血液成分术产物可以包括淋巴细胞,包括T细胞、单核细胞、粒细胞、B细胞、其他有核白细胞、红细胞和血小板。细胞可以被洗涤并重悬在培养基中,以分离感兴趣的细胞。

[0238] T细胞可以通过裂解红细胞和耗尽单核细胞从外周血淋巴细胞分离,例如,通过PERCOLL™梯度离心进行分离。特定的T细胞亚群,诸如CD28+、CD4+、CDC、CD45RA+和CD45RO+T细胞还可以通过阳性或阴性选择技术分离。例如,T细胞可以通过与抗CD3/抗CD28(即3×28)缀合珠诸如DYNABEADS®M-450CD3/CD28 T或XCYTE DYNABEADS™一起孵育足以对期望的T细胞进行阳性选择的时间段来分离。免疫细胞(例如T细胞和B细胞)可以是抗原特异性的(例如对肿瘤有特异性)。

[0239] 在一些实施方案中,细胞可以是抗原呈递细胞(APC),诸如B细胞、来自淋巴结的活化B细胞、淋巴母细胞样细胞(lymphoblastoid cell)、静息B细胞或赘生性B细胞(例如来自淋巴瘤)。APC可以指在其表面上表达至少一种BCRC蛋白的B细胞或滤泡树突细胞。

[0240] 本公开内容的方法可以用于追踪单个T细胞的分子表型。不同的T细胞亚型可以通过表达不同的分子标志物进行区分。T细胞表达来自多种多样的TCR组库的独特的T细胞受体(TCR)。在大多数T细胞中,TCR可以由α链和β链的异二聚体组成;每条功能链可以是T细胞发育期间体细胞DNA重组事件的产物,允许在单个个体中表达超过一百万种不同的TCR。TCR可以用于定义单个T细胞的身份,允许对免疫反应期间的T细胞克隆扩增进行谱系追踪。本公开内容的免疫学方法可以以各种方式使用,包括但不限于,在单个T细胞中鉴定独特的

TCRα和TCRβ链配对,在单个细胞水平定量TCR和标志物表达,鉴定个体中的TCR多样性,表征在不同T细胞群体中表达的TCR组库,确定TCR的α和β链等位基因的功能,以及鉴定免疫反应期间T细胞的克隆扩增。

[0241] T细胞受体链配对

[0242] T细胞受体(TCR)是存在于T淋巴细胞表面上的识别分子。在T细胞表面上发现的T细胞受体可以由称为α链和β链的两个糖蛋白亚基组成。两条链均可以包括约40kDa的分子量,并具有可变结构域和恒定结构域。编码α和β链的基因可以被组织在由基因重排形成的基因的V、D和J区的文库中。TCR可以将由抗原呈递细胞呈递的抗原识别为与组织相容性基因编码的特定自身分子形成的复合物的一部分。最有效的组织相容性基因被称为主要组织相容性复合体(MHC)。因此,被T细胞受体识别的复合物由MHC/肽配体组成。

[0243] 在一些实施方案中,本公开内容的方法、装置和系统可以用于T细胞受体测序和配对。本公开内容的方法、装置和系统可以用于对T细胞受体α和β链进行测序,将α和β链配对,和/或确定T细胞受体α链的功能拷贝。单个细胞可以被容纳在具有单个固体支持物(例如珠)的单个分区(例如孔)中。细胞可以被裂解。珠可以包含随机标记,该随机标记可以结合到TCR的α和/或β链内的特定位置。与固体支持物关联的TCRα和β分子可以经受本公开内容的分子生物学方法,包括逆转录、扩增和测序。包含相同细胞标记的TCRα和β链可以被认为来自同一单细胞,从而使TCR的α和β链配对。

[0244] 抗体组库中的重链和轻链配对

[0245] 本公开内容的方法、装置和系统可以用于BCR受体和抗体的重链和轻链配对。本公开内容的方法允许确定单个生物体或细胞群体中的免疫受体和抗体的组库。本公开内容的方法可以帮助确定组成免疫受体的多肽链对。B细胞和T细胞各自表达免疫受体;B细胞表达免疫球蛋白和BCR,并且T细胞表达T细胞受体(TCR)。两种类型的免疫受体都可以包含两条多肽链。免疫球蛋白可以包含可变重链(VH)和可变轻链(VL)。TCR可以有两种类型:一种由α链和β链组成,并且另一种由δ链和γ链组成。免疫受体中的多肽可以包括恒定区和可变区。可变区可以由B或T细胞的染色体上基因片段的重组和末端连接重排产生。在B细胞中,可变区的另外的多样化可以通过体细胞超突变发生。

[0246] 免疫系统具有大的受体组库,并且淋巴细胞表达的任何给定受体对可以由一对单独的、独特转录物编码。知晓在单个细胞中表达的免疫受体链对的序列可以用于确定给定个体或细胞群体的免疫组库。

[0247] 在一些实施方案中,本公开内容的方法、装置和系统可以用于抗体测序和配对。本公开内容的方法、装置和系统可以用于对抗体重链和轻链(例如,在B细胞中)进行测序,和/或将重链和轻链配对。单细胞可以被容纳在具有单个固体支持物(例如珠)的单个分区(例如孔)中。细胞可以被裂解。

[0248] 珠可以包含随机标记,该随机标记可以结合到抗体(例如,在B细胞中)重链和/或轻链内的特定位置。与固体支持物关联的重链和轻链分子可以经受本公开内容的分子生物学方法,包括逆转录、扩增和测序。包含相同细胞标记的抗体重链和轻链可以被认为来自同一单细胞,从而将抗体的重链和轻链配对。

[0249] 在一些实施方案中,提供了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:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码

包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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;以及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,每个延伸的条形码化核酸分子包含第一分子标记和第二分子标记。该方法可以包括基于与多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定样品中核酸靶的拷贝数。

[0250] 在一些实施方案中,提供了用于确定样品中核酸靶的数量的方法。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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,每个延伸的条形码化核酸分子包含第一分子标记和第二分子标记;以及基于与多于一个延伸的条形码化核酸分子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记,或它们的组合的数量,确定样品中核酸靶的拷贝数。

[0251] 在一些实施方案中,提供了产生和分析单标记的核酸分子的方法。该方法可以包括扩增多于一个延伸的条形码化核酸分子,以产生多于一个单标记的核酸分子,每个单标记的核酸分子包含第一分子标记或第二分子标记,其中确定样品中核酸靶的拷贝数包括:基于与多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定样品中核酸靶的拷贝数。在一些实施方案中,确定样品中核酸靶的拷贝数包括:基于与多于一个单标记的核酸分子关联的具有不同序列的第一分子标记的数量,确定样品中核酸靶的拷贝数。该方法可以包括扩增多于一个延伸的条形码化核酸分子,以产生多于一个延伸的条形码化核酸分子的拷贝,其中确定样品中核酸靶的拷贝数包括:基于以下的(i)和/或(ii)确定样品中核酸靶的拷贝数:(i)与多于一个延伸的条形码化核酸分子或其产物的拷贝关联的具有不同序列的第一分子标记的数量,和/或(ii)与多于一个延伸的条形码化核酸分子或其产物的拷贝关联的具有不同序列的第二分子标记的数量。

[0252] 本文还提供了用于确定样品中核酸靶的数量的方法、系统、组合物和试剂盒。在一些实施方案中,该方法包括:使核酸靶的拷贝与多于一个寡核苷酸条形码接触,其中每个寡核苷酸条形码包含分子标记和能够与核酸靶杂交的靶结合区;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存在下,使与核酸靶的拷贝杂交的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,每个条形码化核酸分子包含与核酸靶的至

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体;使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交:(i)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,和/或(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,每个条形码化核酸分子包含第一分子标记和第二分子标记;扩增多于一个延伸的条形码化核酸分子,以产生多于一个单标记的核酸分子,每个单标记的核酸分子包含第一分子标记或第二分子标记;以及基于与多于一个单标记的核酸分子关联的具有不同序列的第二分子标记的数量,确定样品中核酸靶的拷贝数。

[0253] 本文提供的方法的一些实施方案包括基于与多于一个单标记的核酸分子关联的具有不同序列的第一分子标记的数量,确定样品中核酸靶的拷贝数。在一些实施方案中,该方法包括在使每个条形码化核酸分子的靶结合区的互补体与以下的靶结合区杂交之前,使多于一个条形码化核酸分子变性:(i)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,和/或(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。该方法可以包括在扩增多于一个延伸的条形码化核酸分子之前,使多于一个延伸的条形码化核酸分子变性。确定核酸靶的拷贝数可以包括基于与多于一个单标记的核酸分子中的单标记的核酸分子关联的具有不同序列的第二分子标记的数量来确定样品中多于一个核酸靶中的每一个的拷贝数,所述多于一个单标记的核酸分子包括多于一个核酸靶中的每一个的序列。确定核酸靶的拷贝数可以包括基于与多于一个单标记的核酸分子中的单标记的核酸分子关联的具有不同序列的第一分子标记的数量来确定样品中多于一个核酸靶中的每一个的拷贝数,所述多于一个单标记的核酸分子包括多于一个核酸靶中的每一个的序列。多于一个核酸靶中的每一个的序列可以包括多于一个核酸靶中的每一个的子序列。多于一个条形码化核酸分子中的核酸靶的序列可以包括核酸靶的子序列。

[0254] 在一些实施方案中,方法包括将靶结合区的互补体添加(例如,通过模板转换反应)到条形码化核酸分子的末端(例如,3'末端)。在一些实施方案中,该方法包括寡核苷酸条形码(或其产物,诸如例如另一个条形码化核酸分子或其扩增子)的靶结合区的i)分子内杂交和/或ii)分子间杂交,随后延伸以产生延伸的条形码化核酸分子。延伸的条形码化核酸分子可以在3'末端和5'末端两者都进行条形码化。在一些实施方案中,条形码化分子的分子内杂交与3'多(dT)捕获珠上的捕获mRNA转录物形成发夹环。mRNA分子可以通过多(A)尾结合到寡核苷酸条形码的靶结合区而被捕获到珠上。杂交后,模板转换可以用于在捕获的转录物的5'末端处附接多(dA)尾。然后,新的多(dA)尾可以与同一珠上的游离捕获寡核苷酸(例如条形码,诸如随机条形码)杂交。延伸后,mRNA分子可以在3'末端和5'末端两者都进行条形码化。这允许产生3'和5'条形码化转录物,其可以在例如Illumina测序平台上测序。对条形码化5'序列的访问可以允许检测T细胞受体(TCR)和B细胞受体(BCR)的可变区,以及转录物的5'末端中出现的剪接变体和序列变异。

[0255] 图6A-图6K示出了使用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)的序列的非限制性示例性工作流程的示意性图示。BD®Rhapsody™珠是经过广泛的物理和化学操纵保持完整性的固体条形码化珠。在珠上对mRNA进行多(A)捕获后,可以进行逆转录和模板转换,以将多(dA)尾添加到条形码化cDNA的3'末端。添加的多(dA)尾允许珠结合的cDNA与同一珠上的条形码(例如随机条形码)的寡聚(dT)区自杂交,形成桥-环结

构。桥-环的Klenow延伸可以产生新的条形码化cDNA分子,所述新的条形码化cDNA分子分子来自同一mRNA转录物,其方向与第一个条形码化cDNA相反,允许连接到分子条形码的3’末端和5’末端。

[0256] 本文公开的方法可以允许基于3’和/或基于5’的序列确定。该方法能够实现为序列确定提供灵活性。在一些实施方案中,该方法能够实现在Rhapsody™系统上对样品(诸如小鼠和人类样品)的T细胞和B细胞两者进行免疫组库谱分析,除了使用的引物之外不改变方案或产物配置。在一些实施方案中,可以进行V(D)J的3’和/或5’基因表达谱分析。在一些实施方案中,可以在单细胞平台中研究T细胞和B细胞的表型标志物和V(D)J序列两者。在一些实施方案中,它们的转录物的3’和5’信息两者都可以在单个实验中捕获。本文公开的方法可以允许T细胞和B细胞两者的V(D)J检测(例如,超突变)。

[0257] 本文描述的方法和系统可以与使用与寡核苷酸关联(例如,附接或缀合)的抗体(本文也称为AbO或AbOligo)的方法和系统一起使用。使用AbO来确定单细胞中的蛋白质表达谱和追踪样品来源的实施方案已经在美国专利申请第15/715,028号(作为美国专利申请公布第2018/0088112号公布)和美国专利申请第15/937,713号中描述;每个专利申请的内容通过引用以其整体并入本文。在一些实施方案中,本文公开的方法允许T细胞和B细胞的V(D)J谱分析、3’靶向、5’靶向、3’全转录组扩增(WTA)、5’WTA、用AbO进行蛋白质表达谱分析,和/或在单个实验中进行样品多重化。图7示出了进行V(D)J工作流程、抗体-寡核苷酸(AbO)工作流程和单细胞mRNA表达谱工作流程(例如BD Rhapsody靶向工作流程)的非限制性示例性的示意性图示。

[0258] 模板转换反应

[0259] 图6A-图6K示出了使用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)的序列的非限制性示例性工作流程的示意性图示。条形码(例如随机条形码、寡核苷酸条形码602)可以包含靶结合区(例如多(dT)604),所述靶结合区可以经由多(dA)尾608或其他核酸靶与核酸靶(例如多腺苷酸化的RNA转录物606)结合,用于标记或条形码化(例如,进行独特标记)。靶结合区可以包含基因特异性序列、寡聚(dT)序列、随机多聚体,或它们的任何组合。在一些实施方案中,条形码与固体支持物(例如,颗粒610)关联。多于一个条形码602可以与颗粒610关联。在一些实施方案中,颗粒是珠。珠可以是聚合物珠,例如可变形的珠或凝胶珠,用条形码或随机条形码官能化(诸如来自10X Genomics(San Francisco,CA)的凝胶珠)。在一些实施方式中,凝胶珠可以包括基于聚合物的凝胶。例如,可以通过将一种或更多种聚合物前体包封到液滴中来产生凝胶珠。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。

[0260] 图6A描绘了逆转录反应600a的非限制性示例性实施方案。在逆转录600a期间,当到达寡核苷酸条形码602的末端时,酶(例如逆转录酶,诸如莫洛尼鼠白血病病毒(MMLV))的末端转移酶活性将一些另外的核苷酸(例如脱氧胞苷,CCC 612)添加到新合成的cDNA序列链614c(RNA序列614r的反义序列)的3’末端。这些CCC碱基612可以作为包含与加尾的序列互补的序列(例如,rGrGrG 618)的模板转换寡核苷酸(例如,模板转换寡核苷酸)616的锚定位点。模板转换寡核苷酸616可以包含靶结合区604的至少一部分。在rGrGrG 618和附加的脱氧胞苷链段(appended deoxycytidine stretch)612之间进行碱基配对时,酶将模板链从寡核苷酸条形码602“转换”到模板转换寡核苷酸616,并继续复制到模板转换寡核苷酸

616的5’末端。因此,得到的第一链标记的cDNA(例如,条形码化核酸分子620)包含模板转换寡核苷酸616的反向互补体序列(reverse complement sequence),并且因此可以包含靶结合区的互补体(例如,反向互补体)(例如,多(dA)608)。条形码化核酸分子620可以包含cDNA 614c(RNA序列614r的反向互补序列)。该反应可以在一种或更多种配置用于还原二级结构的添加剂(例如乙二醇)存在下进行。条形码化核酸分子620也可以包含许多标记。寡核苷酸条形码602可以包含分别用于标记转录物606和追踪RNA转录物606(或核酸靶,诸如例如抗体寡核苷酸,无论其与抗体关联还是已经与抗体解离)的样品来源的第一分子标记(ML1)622和样品标记(例如,分区标记、细胞标记(CL)624),以及在每个条形码602的第一分子标记622/细胞标记624区域侧翼的一起用于随后的反应的一个或更多个另外的序列,诸如例如第一通用序列反应626(例如,读段1序列)。每个样品的寡核苷酸条形码中分子标记的序列的组库可以是对于对RNA转录物进行随机标记而言足够大的。在一些实施方案中,样品标记是分区标记。在一些实施方案中,样品标记是细胞标记。条形码化核酸分子620可以经历变性步骤600b(例如,变性),从而产生单链条形码化核酸分子621。

[0261] 在一些实施方案中,在使多于一个条形码化核酸分子的3’-末端延伸后,使第一分子标记与第二分子标记杂交。在一些实施方案中,延伸的条形码化核酸分子各自包含第一分子标记、第二分子标记、靶结合区和靶结合区的互补体。在一些实施方案中,靶结合区的互补体与靶结合区的一部分互补。在一些实施方案中,靶结合区域包含基因特异性序列。在一些实施方案中,靶结合区域包含多(dT)序列。

[0262] 术语“模板转换”可以指逆转录酶将初始核酸序列模板转换到与从初始模板合成的核酸的3’末端具有很少或没有互补性的新核酸序列模板的3’末端的能力。模板转换的一个实例是逆转录酶将初始核酸序列模板/引物基底转换到与核酸引物链的3’末端很少或没有互补性的新核酸序列模板的3’末端的能力。模板转换允许例如使用逆转录酶制备DNA拷贝,该逆转录酶将初始核酸序列模板转换到与从初始模板合成的DNA的3’末端具有很少或没有互补性的新核酸序列模板的3’末端,从而允许合成连续产物DNA,该连续产物DNA在不进行连接的情况下直接将衔接子序列与靶寡核苷酸序列相连。模板转换可以包括衔接子、加同聚物尾(例如,多腺苷酸化)、随机引物或聚合酶可以与之缔和的寡核苷酸的连接。在上文描述的实施方案的任一种中,模板转换可以用于引入靶结合区或其互补体。

[0263] 在一些实施方案中,逆转录酶能够发挥末端转移酶活性。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。在一些实施方案中,模板转换寡核苷酸包含三个3’核糖核苷酸。在一些实施方案中,3’核糖核苷酸包括鸟嘌呤。在一些实施方案中,逆转录酶包括病毒逆转录酶。在一些实施方案中,病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶。在一些实施方案中,病毒逆转录酶是莫洛尼鼠白血病病毒(MMLV)逆转录酶。在一些实施方案中,模板转换寡核苷酸包括SEQ ID NO:1。

[0264] 靶结合区的互补体可以包括靶结合区的反向互补序列,或者可以包括靶结合区的互补序列。分子标记的互补体可以包括分子标记的反向互补序列,或者可以包括分子标记的互补序列。在一些实施方案中,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)分子和/或条形码化核糖核酸(RNA)分子。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、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。核酸靶可以包含

细胞组分结合试剂。在一些实施方案中,核酸分子与细胞组分结合试剂关联。方法可以包括使核酸分子和细胞组分结合试剂解离。在一些实施方案中,多于一个寡核苷酸条形码中的至少10个包含不同的分子标记序列。多于一个寡核苷酸条形码的每个分子标记可以包含至少6个核苷酸。

[0265] 在一些实施方案中,多于一个寡核苷酸条形码与固体支持物关联。与相同固体支持物关联的多于一个寡核苷酸条形码可以各自包含相同的样品标记。多于一个寡核苷酸条形码的每个样品标记可以包含至少6个核苷酸。多于一个寡核苷酸条形码可以各自包含细胞标记。多于一个寡核苷酸条形码的每个细胞标记可以包含至少6个核苷酸。与相同固体支持物关联的寡核苷酸条形码可以包含相同的细胞标记。与不同固体支持物关联的寡核苷酸条形码可以包含不同的细胞标记。多于一个延伸的条形码化核酸分子可以各自包含细胞标记和细胞标记的互补体。细胞标记的互补体可以包括细胞标记的反向互补序列或细胞标记的互补序列。该方法可以包括在以下中的一种或更多种的存在下,使与核酸靶的拷贝杂交的多于一个寡核苷酸条形码延伸: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。在一些实施方案中,固体支持物可以包括合成颗粒。在一些实施方案中,固体支持物可以包括平坦表面。

[0266] 样品可以包含单细胞,并且该方法可以包括使包含多于一个寡核苷酸条形码的合成颗粒与样品中的单细胞关联。该方法可以包括在使合成颗粒与单细胞关联后裂解单细胞。裂解单细胞可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。在一些实施方案中,合成颗粒和单细胞在同一孔中。在一些实施方案中,合成颗粒和单细胞在同一液滴中。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地固定在合成颗粒上。多于一个寡核苷酸条形码中的至少一个可以被包封在合成颗粒中。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地包封在合成颗粒中。在一些实施方案中,合成颗粒是可破坏的。合成颗粒可以包括珠。珠可以包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。合成颗粒可以包含选自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、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。在一些实施方案中,合成颗粒可以包括可破坏的水凝胶颗粒。多于一个寡核苷酸条形码中的每一个可以包含接头官能基团,合成颗粒可以包含固体支持物官能基团,和/或支持物官能基团和接头官能基团可以彼此关联。在一些实施方案中,接头官能基团和支持物官能基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,以及它们的任何组合。条形码化核酸分子的分子内杂交

[0267] 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体与条形码化核酸分子本身的靶结合区杂交包括使条形码化核酸分子内的靶结合区和靶结合区的互补体进行分子内杂交以形成茎环。在一些实施方案中,第二分子标记是第一分子标记的互补体。

[0268] 工作流程可以包括单链条形码化核酸分子621的分子内杂交,如非限制性示例性

图6B的示意性图示中描绘的。该工作流程可以包括单链条形码化核酸分子621内的靶结合区604和靶结合区的互补体608的分子内杂交600c1,以形成茎环。该工作流程可以包括使单链条形码化核酸分子621的茎环的3’末端延伸600c2,以产生延伸的条形码化核酸分子620c。延伸的条形码化核酸分子620c可以包含第一分子标记的互补体(例如,反向互补体)622rc、细胞标记的互补体(例如,反向互补体)624rc和/或第一通用序列的互补体(例如,反向互补体)626rc。该工作流程可以包括使延伸的条形码化核酸分子620c变性600c3,以产生单链延伸的条形码化核酸分子620cd。在一些实施方案中,分子间杂交600c1和/或延伸600c2在高盐缓冲液和/或PEG的存在下进行。在一些实施方案中,延伸使用缺乏5’至3’核酸外切酶活性和3’至5’核酸外切酶活性中的至少一种的DNA聚合酶(例如Klenow片段)进行。

[0269] 单链延伸的条形码化核酸分子620cd可以在靶核酸分子(例如转录物)的5’末端和3’末端两者都包含条形码(例如细胞标记和分子标记),从而与在序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序方面对在一个末端仅具有一个条形码的靶核酸分子的分析相比,实现对靶核酸分子的更广泛分析。单链延伸的条形码化核酸分子620cd可以用作一个或更多个扩增反应(例如,PCR)的模板,诸如例如,图6C-图6D中描绘的非限制性示例性扩增方案的模板。扩增可以包括靶特异性(例如,基因特异性)的cDNA扩增。例如,单链延伸的条形码化核酸分子620cd可以经历第一轮扩增(“PCR1”)600c4,所述第一轮扩增(“PCR1”)600c4使用包含第一通用序列的序列(或其互补体)的通用寡核苷酸引物646和靶特异性引物(例如,靶特异性引物648和/或靶特异性引物650)。PCR1 600c4可以包括用通用寡核苷酸引物646和靶特异性引物648扩增单链延伸的条形码化核酸分子620cd的5’区域,从而产生包含第一分子标记622、细胞标记624、第一通用序列626和部分cDNA 614c1(其长度取决于靶特异性引物648在cDNA 614c内的结合位点)的单标记的核酸分子620c1。PCR1 600c4可以包括用通用寡核苷酸引物646和靶特异性引物650扩增单链延伸的条形码化核酸分子620cd的3’区域,从而产生包含第一分子标记622、细胞标记624、第一通用序列626和部分反义cDNA 614cas1(其长度取决于靶特异性引物650在cDNA 614c内的结合位点)的单标记的核酸分子620cas1。PCR1 600c4可以包括1-30个循环(例如,15个循环)。

[0270] 工作流程可以包括第二轮扩增(“PCR2”)600c5,所述第二轮扩增(“PCR2”)600c5使用通用寡核苷酸引物646和巢式靶特异性引物(例如,靶特异性引物652和/或靶特异性引物654)。靶特异性引物652和/或靶特异性引物654可以包括突出端,所述突出端可以包括或是例如第二通用序列638(例如,读段2序列,所述读段2序列为通用PCR柄(universal PCR handle))。PCR2 600c5可以包括用通用寡核苷酸引物646和巢式靶特异性引物654扩增单标记的核酸分子620c1,从而产生包含第一分子标记622、细胞标记624、第一通用序列626、第二通用序列638和部分cDNA 614c2(其长度取决于巢式靶特异性引物654在部分cDNA 614c1内的结合位点)的单标记的核酸分子620c2。PCR2 600c5可以包括用通用寡核苷酸引物646和巢式靶特异性引物652扩增单标记的核酸分子620cas1,从而产生包含第一分子标记622、细胞标记624、第一通用序列626、第二通用序列638和部分反义cDNA 614cas2(其长度取决于巢式靶特异性引物652在部分反义cDNA 614cas1内的结合位点)的单标记的核酸分子620cas2。PCR2 600c5可以包括1-30个循环(例如,15个循环)。在一些实施方案中,靶特异性引物648、650、652和/或654结合免疫受体的恒定区、可变区、多样性区和/或连接区。

[0271] 工作流程可以包括第三轮扩增(“PCR3”)600c6。PCR3 600c6可以包括用测序文库

扩增引物656和658对单标记的核酸分子620cas2和/或单标记的核酸分子620c2进行的文库扩增。测序文库扩增引物656和658可以分别退火至第一通用序列626和第二通用序列638(或其互补体)。PCR3 600c6可以通过测序文库扩增引物656和658中的突出端添加测序衔接子(例如P5 640和P7 642)和样品索引644(例如i5、i7)。文库扩增子620cas3和/或620c3可以被测序并经受本公开内容的下游方法。使用150bp x 2测序进行的测序可以揭示读段1上的细胞标记、独特分子标记和/或基因(或基因的部分序列)、读段2上的基因(或基因的部分序列),以及索引1读段和/或索引2读段上的样品索引。PCR3 600c6可以包括1-30个循环(例如,15个循环)。

[0272] 在一些实施方案中,可以对免疫受体的V(D)J区进行3'和/或5'表达谱分析。在一些实施方案中,可以在单细胞平台中研究T细胞和/或B细胞的表型标志物和一个或更多个免疫受体V(D)J序列两者。在一些实施方案中,它们的转录物的3'信息和5'信息两者都可以在单个实验中捕获。本文公开的方法可以允许T细胞和B细胞两者的V(D)J检测(例如,超突变)。在一些实施方案中,延伸的条形码化核酸分子620cd的3'区域和5'区域两者都被扩增。在一些实施方案中,仅延伸的条形码化核酸分子620cd的5'区域被扩增。在一些实施方案中,仅延伸的条形码化核酸分子620cd的3'区域被扩增。在一些实施方案中,一个或更多个扩增反应包括多重PCR。例如,延伸的条形码化核酸分子620cd的3'区域和5'区域两者可以同时被扩增(例如多重PCR)。在一些实施方案中,该工作流程包括多重PCR,该多重PCR采用一组靶特异性PCR1引物和/或一组靶特异性PCR2引物。在一些实施方案中,靶包括BCR、TCR和/或免疫相关的转录物。

[0273] 条形码化核酸分子与条形码化核酸分子的分子间杂交

[0274] 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体与多于一个条形码化核酸分子中的不同条形码化核酸分子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体与多于一个条形码化核酸分子中的不同条形码化核酸分子的靶结合区进行分子间杂交。在一些实施方案中,第二分子标记的序列不同于第一分子标记的序列,并且其中第二分子标记不是第一分子标记的互补体。

[0275] 工作流程可以包括单链条形码化核酸分子621与不同的条形码化核酸分子628的分子间杂交,如非限制性示例性图6E-图6F的示意性图示中描绘的。不同的条形码化核酸分子628可以包含cDNA 630c、第二分子标记632、细胞标记624和第一通用序列626。条形码化核酸分子628的第二分子标记632的序列可以不同于单链条形码化核酸分子621的第一分子标记622的序列(例如,不是互补体)。条形码化核酸分子628的靶结合区604、细胞标记624和/或第一通用序列626可以与单链条形码化核酸分子621的靶结合区604、细胞标记624和/或第一通用序列626相同(或为其互补体)。在一些实施方案中,该工作流程可以包括单链条形码化核酸分子621的靶结合区的互补体608与条形码化核酸分子628的靶结合区604的分子间杂交600d1。该工作流程可以包括使单链条形码化核酸分子621的3'末端延伸600d2,以产生延伸的条形码化核酸分子620d。延伸的条形码化核酸分子620d可以包含第二分子标记的互补体(例如,反向互补体)632rc、细胞标记的互补体(例如,反向互补体)624rc和/或第一通用序列的互补体(例如,反向互补体)626rc。该工作流程可以包括使延伸的条形码化核酸分子620d变性600d3,以产生单链延伸的条形码化核酸分子620dd。在一些实施方案中,分子间杂交600d1和/或延伸600d2在高盐缓冲液和/或PEG的存在下进行。在一些实施方案中,

延伸使用缺乏5’至3’核酸外切酶活性和3’至5’核酸外切酶活性中至少一种的DNA聚合酶(例如Klenow片段)进行。

[0276] 单链延伸的条形码化核酸分子620dd可以在靶核酸分子(例如转录物)的5’末端和3’末端两者都包含条形码(例如细胞标记和分子标记),从而与在序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序方面对在一个末端仅具有一个条形码的靶核酸分子的分析相比,实现对靶核酸分子的更广泛分析。单链延伸的条形码化核酸分子620dd可以用作一个或更多个扩增反应(例如,PCR)的模板,诸如例如,图6G-图6H中描绘的非限制性示例性扩增方案的模板。扩增可以包括靶特异性(例如,基因特异性)的cDNA扩增。例如,单链延伸的条形码化核酸分子620dd可以经历第一轮扩增(“PCR1”)600d4,所述第一轮扩增(“PCR1”)600d4使用包含第一通用序列(或其互补体)的序列的通用寡核苷酸引物646和靶特异性引物(例如,靶特异性引物648和/或靶特异性引物650)。PCR1 600d4可以包括用通用寡核苷酸引物646和靶特异性引物648扩增单链延伸的条形码化核酸分子620dd的5’区域,从而产生包含第一分子标记622、细胞标记624、第一通用序列626和部分cDNA 614c1(其长度取决于靶特异性引物648在cDNA 614c内的结合位点)的单标记的核酸分子620d1。PCR1 600d4可以包括用通用寡核苷酸引物646和靶特异性引物650扩增单链延伸的条形码化核酸分子620dd的3’区域,从而产生包含第二分子标记632、细胞标记624、第一通用序列626和部分反义cDNA 614cas1(其长度取决于靶特异性引物650在cDNA 614c内的结合位点)的单标记的核酸分子620das1。PCR1 600d4可以包括1-30个循环(例如,15个循环)。

[0277] 工作流程可以包括第二轮扩增(“PCR2”)600d5,所述第二轮扩增(“PCR2”)600d5使用通用寡核苷酸引物646和巢式靶特异性引物(例如,靶特异性引物652和/或靶特异性引物654)。靶特异性引物652和/或靶特异性引物654可以包括突出端,所述突出端可以包括或是例如第二通用序列638(例如,读段2序列,所述读段2序列为通用PCR柄)。PCR2 600d5可以包括用通用寡核苷酸引物646和巢式靶特异性引物654扩增单标记的核酸分子620d1,从而产生包含第一分子标记622、细胞标记624、第一通用序列626、第二通用序列638和部分cDNA 614c2(其长度取决于巢式靶特异性引物654在cDNA 614c1内的结合位点)的单标记的核酸分子620d2。PCR2 600d5可以包括用通用寡核苷酸引物646和巢式靶特异性引物652扩增单标记的核酸分子620das1,从而产生包含第二分子标记632、细胞标记624、第一通用序列626、第二通用序列638和部分反义cDNA 614cas2(其长度取决于巢式靶特异性引物652在部分反义cDNA 614cas1内的结合位点)的单标记的核酸分子620das2。PCR2 600d5可以包括1-30个循环(例如,15个循环)。在一些实施方案中,靶特异性引物648、650、652和/或654结合免疫受体的恒定区、可变区、多样性区和/或连接区。

[0278] 工作流程可以包括第三轮扩增(“PCR3”)600d6。PCR3 600d6可以包括用测序文库扩增引物656和658对单标记的核酸分子620das2和/或单标记的核酸分子620d2进行的文库扩增。测序文库扩增引物656和658可以分别退火至第一通用序列626和第二通用序列638(或其互补体)。PCR3 600d6可以通过测序文库扩增引物656和658中的突出端添加测序衔接子(例如P5 640和P7 642)和样品索引644(例如i5、i7)。文库扩增子620das3和/或620d3可以被测序并经受本公开内容的下游方法。使用150bp x 2测序的测序可以揭示读段1上的细胞标记、独特分子标记和/或基因(或基因的部分序列)、读段2上的基因(或基因的部分序列),以及索引1读段和/或索引2读段上的样品索引。PCR3 600d6可以包括1-30个循环(例

如,15个循环)。

[0279] 在一些实施方案中,可以对免疫受体的V(D)J区进行3'和/或5'表达谱分析。在一些实施方案中,可以在单细胞平台中研究T细胞和/或B细胞的表型标志物和免疫受体V(D)J序列两者。在一些实施方案中,它们的转录物的3'信息和5'信息两者都可以在单个实验中捕获。本文公开的方法可以允许T细胞和B细胞两者的V(D)J检测(例如,超突变)。在一些实施方案中,延伸的条形码化核酸分子620dd的3'区域和5'区域两者都被扩增。在一些实施方案中,仅延伸的条形码化核酸分子620dd的5'区域被扩增。在一些实施方案中,仅延伸的条形码化核酸分子620dd的3'区域被扩增。在一些实施方案中,一个或更多个扩增反应包括多重PCR。例如,延伸的条形码化核酸分子620dd的3'区域和5'区域两者可以同时被扩增(例如多重PCR)。在一些实施方案中,该工作流程包括多重PCR,该多重PCR采用一组靶特异性PCR1引物和/或一组靶特异性PCR2引物。在一些实施方案中,靶包括BCR、TCR和/或免疫相关的转录物。

[0280] 条形码化核酸分子与寡核苷酸条形码的分子间杂交

[0281] 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体与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体与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。在一些实施方案中,第二分子标记不同于第一分子标记,并且其中第二分子标记不是第一分子标记的互补体。在一些实施方案中,该方法包括使与条形码化核酸分子的靶结合区的互补体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,每个延伸的条形码化核酸分子包含第一分子标记的互补体和第二分子标记。在一些实施方案中,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体。

[0282] 工作流程可以包括单链条形码化核酸分子621与不同的寡核苷酸条形码634的分子间杂交,如非限制性示例性图6I-图6J的示意性图示中描绘的。不同的寡核苷酸条形码634可以包含第二分子标记636、细胞标记624和第一通用序列626。寡核苷酸条形码634的第二分子标记636的序列可以不同于单链条形码化核酸分子621的第一分子标记622的序列(例如,不是互补体)。寡核苷酸条形码634的靶结合区604、细胞标记624和/或第一通用序列626可以与单链条形码化核酸分子621的靶结合区604、细胞标记624和/或第一通用序列626相同(或为其互补体)。在一些实施方案中,该工作流程可以包括单链条形码化核酸分子621的靶结合区的互补体608与寡核苷酸条形码634的靶结合区604的分子间杂交600e1。该工作流程可以包括使单链条形码化核酸分子621的3'末端延伸600e2,以产生延伸的条形码化核酸分子620e1。延伸的条形码化核酸分子620e1可以包含第二分子标记的互补体(例如,反向互补体)636rc、细胞标记的互补体(例如,反向互补体)624rc,第一通用序列的互补体(例如,反向互补体)626rc,和/或cDNA 614c。该工作流程可以包括使延伸的条形码化核酸分子620e1变性600e3,以产生单链延伸的条形码化核酸分子620e1d。该工作流程可以包括使寡核苷酸条形码634的3'末端延伸600e2,以产生延伸的条形码化核酸分子620e2。延伸的条形码化核酸分子620e2可以包含第一分子标记的互补体(例如,反向互补体)622rc、细胞标记的互补体(例如,反向互补体)624rc,第一通用序列的互补体(例如,反向互补体)626rc,和/或反义cDNA 614cas。该工作流程可以包括使延伸的条形码化核酸分子620e2变性600e3,以

产生单链延伸的条形码化核酸分子620e2d。在一些实施方案中,分子间杂交600e1和/或延伸600e2在高盐缓冲液和/或PEG的存在下进行。在一些实施方案中,延伸使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种的DNA聚合酶(例如Klenow片段)进行。

[0283] 单链延伸的条形码化核酸分子620e1d和单链延伸的条形码化核酸分子620e2d可以在靶核酸分子(例如转录物)的5'末端和3'末端两者都包含条形码(例如细胞标记和分子标记),从而与在序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序方面对在一个末端仅具有一个条形码的靶核酸分子的分析相比,实现对靶核酸分子的更广泛分析。单链延伸的条形码化核酸分子620e1d和单链延伸的条形码化核酸分子620e2d可以用作一个或更多个扩增反应(例如,PCR)的模板。扩增可以包括靶特异性(例如,基因特异性)的cDNA扩增。在一些实施方案中,单链延伸的条形码化核酸分子620e1d和/或单链延伸的条形码化核酸分子620e2d可以经历两轮或更多轮的PCR扩增(例如,如图6G-图6H中描绘的的PCR1 600d4、PCR2 600d5和/或PCR3 600d6)。在一些实施方案中,单链延伸的条形码化核酸分子620e1d和/或单链延伸的条形码化核酸分子620e2d可以用作单次扩增的模板,诸如例如,图6K中描绘的非限制性示例性扩增方案(PCR 600e4)的模板。PCR 600e4可以通过引物660、662和664中的突出端添加测序衔接子(例如P5 640和P7 642)和样品索引644(例如i5、i7)。PCR 600e4可以包括用引物664(退火至第一通用序列或其互补体)和靶特异性引物660扩增单链延伸的条形码化核酸分子620e1d,从而产生包含第一分子标记622、细胞标记624、第一通用序列626和部分cDNA 614c1e(其长度取决于靶特异性引物660在cDNA 614c内的结合位点)的单标记的核酸分子620e1c。PCR 600e4可以包括用引物664(退火至第一通用序列或其互补体)和靶特异性引物662扩增单链延伸的条形码化核酸分子620e2d,从而产生包含第二分子标记636、细胞标记624、第一通用序列626和部分反义cDNA 614cas1e(其长度取决于靶特异性引物662在反义cDNA 614cas内的结合位点)的单标记的核酸分子620e2c。文库扩增子620e1c和/或620e2c可以被测序并经受本公开内容的下游方法。使用150bp x 2测序进行的测序可以揭示读段1上的细胞标记、独特分子标记和/或基因(或基因的部分序列)、读段2上的基因(或基因的部分序列),以及索引1读段和/或索引2读段上的样品索引。PCR 600e4可以包括1-30个循环(例如,15个循环)。在一些实施方案中,靶特异性引物660和/或662结合免疫受体的恒定区、可变区、多样性区和/或连接区。

[0284] 在一些实施方案中,可以对免疫受体的V(D)J区进行3'和/或5'表达谱分析。在一些实施方案中,可以在单细胞平台中研究T细胞和/或B细胞的表型标志物和免疫受体V(D)J序列两者。在一些实施方案中,转录物的3'信息和5'信息两者都可以在单个实验中捕获。本文公开的方法可以允许T细胞和B细胞两者的V(D)J检测(例如,超突变)。在一些实施方案中,延伸的条形码化核酸分子620e1d和/或620e2d的3'区域和5'区域两者都被扩增。在一些实施方案中,仅延伸的条形码化核酸分子620e1d和/或620e2d的5'区域被扩增。在一些实施方案中,仅延伸的条形码化核酸分子620e1d和/或620e2d的3'区域被扩增。在一些实施方案中,一个或更多个扩增反应包括多重PCR。例如,延伸的条形码化核酸分子620e1d和/或620e2d的3'区域和5'区域两者可以同时被扩增(例如多重PCR)。在一些实施方案中,该工作流程包括多重PCR,该多重PCR采用一组靶特异性PCR1引物和/或一组靶特异性PCR2引物。在一些实施方案中,靶包括BCR、TCR和/或免疫相关的转录物。

[0285] 免疫组库谱分析

[0286] 在一些实施方案中,提供了对免疫受体的V(D)J区进行3’和/或5’表达谱分析的方法。在一些实施方案中,样品包含单细胞。在一些实施方案中,样品包含多于一个细胞、多于一个单细胞、组织、肿瘤样品,或它们的任何组合。单细胞可以包括免疫细胞。在一些实施方案中,免疫细胞是B细胞或T细胞。在一些实施方案中,单细胞可以包括循环肿瘤细胞。在一些实施方案中,每个寡核苷酸条形码可以包含第一通用序列。在一些实施方案中,多于一个延伸的条形码化核酸分子包含第一通用序列和第一通用序列的互补体。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以产生多于一个延伸的条形码化核酸分子的拷贝包括使用能够与第一通用序列或其互补体杂交的引物。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以产生多于一个单标记的核酸分子包括使用能够与第一通用序列或其互补体杂交的引物,和扩增引物。在一些实施方案中,扩增引物是靶特异性引物。在一些这样的实施方案中,靶特异性引物与免疫受体特异性地杂交。例如,靶特异性引物可以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,或它们的任何组合特异性地杂交。免疫受体可以是T细胞受体(TCR)和/或B细胞受体(BCR)。TCR可以包括TCRα链、TCRβ链、TCRγ链、TCRδ链,或它们的任何组合。BCR可以包括BCR重链和/或BCR轻链。

[0287] 该方法可以包括获得多于一个延伸的条形码化核酸分子或其产物的序列信息。获得序列信息可以包括将测序衔接子附接到多于一个延伸的条形码化核酸分子或其产物。获得序列信息可以包括将测序衔接子附接到多于一个单标记的核酸分子或其产物。

[0288] 获得序列信息可以包括获得单细胞的BCR轻链和BCR重链的序列信息。BCR轻链和BCR重链的序列信息可以包括BCR轻链和/或BCR重链的互补决定区1(CDR1)、CDR2、CDR3或它们的任何组合的序列。该方法可以包括基于获得的序列信息将单细胞的BCR轻链和BCR重链配对。样品可以包含多于一个单细胞,并且该方法可以包括基于获得的序列信息将至少50%的单细胞的BCR轻链和BCR重链配对。在一些实施方案中,其中BCR轻链和BCR重链根据本文提供的方法进行配对的样品的单细胞的百分比可以是以下,或是约以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,其中BCR轻链和BCR重链根据本文提供的方法进行配对的样品的单细胞的百分比可以是至少以下,或至多以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、

48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。

[0289] 获得序列信息可以包括获得单细胞的TCRα链和TCRβ链的序列信息。在一些实施方案中,TCRα链和TCRβ链的序列信息可以包括TCRα链和/或TCRβ链的互补决定区1(CDR1)、CDR2、CDR3或它们的任何组合的序列。在一些实施方案中,该方法可以包括基于获得的序列信息将单细胞的TCRα链和TCRβ链配对。在一些实施方案中,样品可以包含多于一个单细胞,并且该方法可以包括基于获得的序列信息将至少50%的单细胞的TCRα链和TCRβ链配对。在一些实施方案中,其中TCRα链和TCRβ链根据本文提供的方法进行配对的样品的单细胞的百分比可以是以下,或是约以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,其中TCRα链和TCRβ链根据本文提供的方法进行配对的样品的单细胞的百分比可以是至少以下,或至多以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。

[0290] 获得序列信息可以包括获得单细胞的TCRγ链和TCRδ链的序列信息。TCRγ链和TCRδ链的序列信息可以包括TCRγ链和/或TCRδ链的互补决定区1(CDR1)、CDR2、CDR3或它们的任何组合的序列。该方法可以包括基于获得的序列信息将单细胞的TCRγ链和TCRδ链配对。样品可以包含多于一个单细胞,并且该方法可以包括基于获得的序列信息将至少50%的单细胞的TCRγ链和TCRδ链配对。在一些实施方案中,其中TCRδ链和TCRγ链根据本文提供的方法进行配对的样品的单细胞的百分比可以是以下,或是约以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、

63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%,或这些值中的任何两个值之间的数字或范围。在一些实施方案中,其中TCRδ链和TCRγ链根据本文提供的方法进行配对的样品的单细胞的百分比可以是至少以下,或至多以下:0.000000001%、0.00000001%、0.0000001%、0.000001%、0.00001%、0.0001%、0.001%、0.01%、0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。用于在核酸靶的5'和3'末端的条形码化的试剂盒

[0291] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一个寡核苷酸条形码,其中多于一个寡核苷酸条形码中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个包含不同的分子标记序列;逆转录酶;模板转换寡核苷酸,所述模板转换寡核苷酸包含靶结合区或其一部分;和DNA聚合酶,所述DNA聚合酶缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中的至少一种。在一些实施方案中,DNA聚合酶包括Klenow片段。在一些实施方案中,逆转录酶包括病毒逆转录酶。在一些实施方案中,病毒逆转录酶是鼠白血病病毒(MLV)逆转录酶。在一些实施方案中,病毒逆转录酶是莫洛尼鼠白血病病毒(MMLV)逆转录酶。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'核糖核苷酸包括鸟嘌呤。在一些实施方案中,试剂盒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。

[0292] 在一些实施方案中,试剂盒包含缓冲液。在一些实施方案中,试剂盒包含药筒。在一些实施方案中,试剂盒包含用于逆转录反应的一种或更多种试剂。在一些实施方案中,试剂盒包含用于扩增反应的一种或更多种试剂。在一些实施方案中,靶结合区包括基因特异性序列、寡聚(dT)序列、随机多聚体,或它们的任何组合。在一些实施方案中,寡核苷酸条形码包含相同的样品标记和/或相同的细胞标记。在一些实施方案中,多于一个寡核苷酸条形码的每个样品标记和/或细胞标记包含至少6个核苷酸。在一些实施方案中,多于一个寡核苷酸条形码的每个分子标记包含至少6个核苷酸。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地固定在合成颗粒上。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被包封在合成颗粒中。在一些实施方案中,多于一个寡核苷酸条形码中的至少一个被部分地包封在合成颗粒中。在一些实施方案中,合成颗粒是可破坏的。在一些实施方案中,合成颗粒包括珠,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、寡聚(dT)缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光染料微珠,或它们的任何组合。在一些实施方案中,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、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮,以及它们的任何组合。在一些实施方案中,合成颗粒包括可破坏的水凝胶颗粒。在一些实施方案中,多于一个寡核苷酸条形码中的每一个包含接头官能基团,合成颗粒包含固体支持物官能基团,和/或支持物官能基团和接头官能基团彼此关联。在一些实施方案中,接头官能基团和支持物官能基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,以及它们的任何组合。

实施例

[0293] 上文讨论的实施方案的一些方面在以下实施例中进一步详细公开,其不是意在以任何方式限制本公开内容的范围。

[0294] 下文描述的非限制性示例性V(D)J方案用于证明靶向小组(targeted panel)的mRNA靶的3'和5'末端二者的测序文库的产生。

[0295] BD*Rhapsody*®细胞捕获和逆转录

[0296] 1.制备样品细胞的单细胞悬液。

[0297] 2.遵循标准BD Rhapsody®方案以通过取回和珠洗涤进行单细胞捕获,并将珠置于冰上。

[0298] 3.根据下表1制作模板转换反应混合物。

[0299] 表1.模板转换反应混合物

[0300]

试剂	200 uL 反应	终浓度
水	68	
5X SSIV 缓冲液(ThermoFisher)	40	1X
dNTP (10mM, NEB N0447L)	20	1 mM
0.1M DTT	10	5 mM
100 uM 模板转换寡核苷酸(25T)	5	2.5 uM 5' TTT TTT TTT TTT TTT TTT rG rG rG 3' (SEQ ID NO: 1)
25 mM $MgCl_2$	24	3 mM
20 mg/ml BSA	1	100 ng/ul
RNA 酶抑制剂(40U/ul)	10	2 U/ul
乙二醇(1113.3 mg/ml)	12	66.8 ug/ul
SSIV (200 U/ul, ThermoFisher)	10	10 U/ul
总计	200	

[0301]

[0302] 4.将珠置于磁体上,去除上清液并将珠重悬浮于200uL的反应混合物中。

[0303] 5.将管置于25℃的热混合仪上30min,然后在42℃、1200rpm放置1.5h。反应结束后置于冰上。

[0304] 6.将珠置于磁体上,并去除上清液。

[0305] 7.将珠重悬浮于1mL TE缓冲液中。

[0306] 8.将珠加热到95℃持续2分钟,以使mRNA变性。

[0307] 9.将珠置于磁力架上,并去除上清液。

[0308] 10.将珠重悬浮于1mL TE缓冲液中。

[0309] 11.将珠加热到95℃持续2分钟,以使mRNA变性。

[0310] 12.将珠置于磁力架上,并去除上清液。

[0311] 13.将珠重悬浮于2mL预加温(37℃)的HT1缓冲液(Illumina,San Diego,CA)中。

[0312] 自杂交

[0313] 1.将管在37℃以1200rpm摇动5min,然后在25℃摇动25min。之后置于冰上。

[0314] 2.用1mL HT1缓冲液洗涤珠一次。

[0315] Klenow延伸

[0316] 1.制备下表2中示出的Klenow延伸反应混合物。

[0317] 表2.Klenow延伸反应混合物

[0318]

试剂	200uL rxn
水	150
10X Klenow缓冲液	20
dNTP(10mM)	20
Klenow片段外切-(NEB M0212S,5U/ul)	10

[0319] 2.将珠置于磁力架上,并去除上清液。

[0320] 3.将珠重悬浮于200uL Klenow延伸反应混合物中。

[0321] 4.置于37℃的热混合仪中30分钟,1200rpm。

[0322] 5.用1mL TE洗涤一次。

[0323] ExoI处理

[0324] 1.根据下表3制备ExoI反应混合物。

[0325] 表3.ExoI反应混合物

[0326]

组分	1个文库(uL)	1.2X
水	170.0	204.0
10X核酸外切酶I缓冲液	20.0	24.0
核酸外切酶I	10.0	12.0
总计	200.0	240.0

[0327] 2.将珠置于磁体上,并去除上清液。

[0328] 3.将珠重悬浮于200uL ExoI反应混合物中。

[0329] 4.将管置于37℃的热混合仪中30分钟,1200rpm。

[0330] 5.将管转移到80℃热混合仪中20分钟,不摇动。

[0331] 6.将管置于冰上～1分钟。

[0332] 7.将珠置于磁体上。

[0333] 8.去除上清液,并将珠重悬浮于200uL珠重悬浮缓冲液中。

[0334] PCR1扩增

[0335] 1.根据下表4制备PCR1主混合物:

[0336] 表4.PCR1主混合物(TCR+IR+5’IR+BCR)

[0337]

PCR1	1x	终浓度
PCR 级水	3.4	
解析 PCR 主混合物(Resolve PCR Mastermix) (2x KAPA2G)	100	1x
免疫反应-Hs	40	
TCRa N1 引物 – 10uM	1.2	60nM
TCRb N1 引物 – 10uM	1.2	60nM
BCR 池 N1 - 20uM	4.2	60nM
5' IR 30 重 - 20uM	18	60nM
通用寡核苷酸(ILR2, 10uM)	20	1uM
20mg/ml BSA	12	
总计	200	

[0338] 2.(任选的)对珠进行二次取样。

[0339] 3.将具有珠的管置于磁体上,并去除上清液。

[0340] 4.将珠重悬浮于200uL PCR1反应混合物中。轻轻移液以彻底混合。

[0341] 5.在(4)个0.2ml PCR管中均匀分配(即~50ul±5ul/管)。

[0342] 6.在后PCR室(Post-PCR room)中,运行以下PCR方案:95℃3min,15个循环(95℃30s,60℃3min,72℃1min),72℃5min。4℃保持。

[0343] 7.PCR后,将PCR1产物和珠合并到LoBind 1.5ml微量离心管中。

[0344] 8.将试管置于1.5ml磁体上,并将PCR1产物移液到新管中。

[0345] PCR1清理

[0346] 1.向PCR1产物添加200ul Ampure XP珠(PCR产物体积的1x)。充分混合。

[0347] 2.在室温孵育5min。

[0348] 3.制备80%的新鲜乙醇(例如,800ul乙醇和200ul无DNA酶/RNA酶的水)。

[0349] 4.将具有Ampure珠的管置于1.5ml管状磁体上持续大约1-2分钟。

[0350] 5.在管的侧面收集所有珠后,去除上清液。

[0351] 6.当管在磁体上时,添加500ul 80%的乙醇以洗涤珠沉淀。

[0352] 7.尽可能多地去除乙醇。

[0353] 8.重复80%乙醇洗涤一次,总计两次洗涤。

[0354] 9.打开盖子,使Ampure珠在磁体上风干,直到没有明显的液滴存在(约3-5分钟)。

[0355] 10.当管在磁体上时,添加500ul 80%的乙醇以洗涤珠沉淀。

[0356] 11.尽可能多地去除乙醇。

[0357] 12.重复80%乙醇洗涤一次,总计两次洗涤。

[0358] 13.打开盖子,使Ampure珠在磁体上风干,直到没有明显的液滴存在(约3-5分钟)。

[0359] 14.将Ampure珠重悬浮于30ul洗脱缓冲液中。

[0360] 15.置于1.5ml管状磁体上。

[0361] 16.将上清液转移到新的1.5ml管中。这是纯化的PCR1产物。如果在同一天进行下一步,则储存在4℃或冰上,或储存在-20℃直到使用。

[0362] PCR2扩增

[0363] 1.在前-PCR区域(pre-PCR area),制备以下反应混合物:

[0364] 表5.TCR反应混合物

[0365]

	1x	终浓度
解析PCR主混合物(2x KAPA2G)	25	1x
TCRa N2引物-1uM	3	60nM每种引物
TCRb N2引物-1uM	3	
通用寡核苷酸(ILR2,10uM)	2	400nM
PCR级水	12	
总计	45	

[0366] 表6.BCR反应混合物

[0367]

	1x	终浓度
解析PCR主混合物(2x KAPA2G)	25	1x
BCR N2引物20uM	1	60nM每种引物
通用寡核苷酸(ILR2,10uM)	2	400nM
PCR级水	22	
总计	45	

[0368] 表7.免疫反应5'小组

[0369]

	1x	终浓度
解析PCR主混合物(2x KAPA2G)	25	1x
5'IR 30重-20uM	4.5	60nM每种引物
通用寡核苷酸(ILR2,10uM)	2	400nM
PCR级水	13.5	
总计	45	

[0370] 表8.免疫反应3'小组

[0371]

	1x	终浓度
免疫反应-Hs	10	60nM每种引物
通用寡核苷酸(ILR2,10uM)	2	400nM
PCR级水	8	
总计	45	

[0372] 2.将反应混合物放至后PCR区域。

[0373] 3.向45ul反应混合物添加5ul经清理的PCR1产物。

[0374] 4.在处于后PCR区域的热循环仪中运行以下PCR方案:95℃3min,15个循环(95℃30s,60℃3min,72℃1min),72℃5min。

[0375] PCR2清理

[0376] 1.对于TCR产物和BCR产物,向PCR1产物添加30ul Ampure XP珠(PCR产物体积的0.6x)。对于IR 3'和5',添加50ul Ampure XP珠(体积的1X)。充分混合。

[0377] 2.在室温孵育5min。

[0378] 3.制备80%的新鲜乙醇(例如,800ul乙醇和200ul无DNA酶/RNA酶的水)。

[0379] 4.将具有Ampure珠的管置于1.5ml管状磁体上持续大约1-2分钟。在管的侧面收集

所有珠后,去除上清液。

[0380] 5.当管在磁体上时,添加200ul 80%的乙醇以洗涤珠沉淀。

[0381] 6.尽可能多地去除乙醇。

[0382] 7.重复80%乙醇洗涤一次,总计两次洗涤。

[0383] 8.打开盖子,使Ampure珠在磁体上风干,直到没有明显的液滴存在。

[0384] 9.将珠重悬浮于30ul洗脱缓冲液中。

[0385] 10.置于1.5ml管状磁体上。

[0386] 11.将上清液转移到新的1.5ml管中。这是纯化的PCR2产物。如果在同一天进行下一步,则储存在4℃或冰上,或储存在-20℃直到使用。

[0387] 12.使用Qubit DNA HS测定测量洗脱的DNA量,以评价下一次PCR是否需要稀释产物。在进行最终PCR之前,必须使用洗脱缓冲液将PCR2产物稀释至≤10ng/ul,以避免过度扩增。

[0388] 编索引PCR

[0389] 1.在前-PCR区域,制备表9中示出的以下反应混合物。

[0390] 表9.反应混合物

[0391]

	1x	4.4X	终浓度
解析PCR主混合物(2x KAPA2G)	25	110	1x
解析文库正向引物(P5,10uM)	2	8.8	400nM
解析文库反向引物**(P7,10uM)	2	8.8	400nM
PCR级水	18	79.2	
总计	47	-	

[0392] 2.将反应混合物放至后PCR区域。

[0393] 3.向47ul反应混合物添加3ul经清理的PCR2产物。

[0394] 4.在后PCR区域运行以下PCR方案:95℃持续5min,8个循环(98℃持续15s,60℃持续30s,72℃持续30s),72℃持续1min。

[0395] 最终PCR清理

[0396] 1.向PCR产物添加30ul Ampure XP珠(PCR产物体积的0.6x)。充分混合。

[0397] 2.在室温孵育5min。

[0398] 3.制备80%的新鲜乙醇(例如,800ul乙醇和200ul无DNA酶/RNA酶的水)。

[0399] 4.将具有Ampure珠的管置于1.5ml管状磁体上持续大约1-2分钟。在管的侧面收集所有珠后,去除上清液。

[0400] 5.当管在磁体上时,添加200ul 80%的乙醇以洗涤珠沉淀。

[0401] 6.尽可能多地去除乙醇。

[0402] 7.重复80%乙醇洗涤一次,总计两次洗涤。

[0403] 8.打开盖子,使Ampure珠在磁体上风干,直到没有明显的液滴存在。

[0404] 9.将珠重悬浮于30ul洗脱缓冲液中。

[0405] 10.置于1.5ml管状磁体上。

[0406] 11.将上清液转移到新的1.5ml管中。这是纯化的PCR2产物。如果在同一天进行下一步,则储存在4℃或冰上,或储存在-20℃直到使用。

[0407] 12.使用Qubit DNA HS测定测量洗脱的DNA量,以评价下一次PCR是否需要稀释产物。在进行最终PCR之前,必须使用洗脱缓冲液将PCR2产物稀释至≤10ng/ul,以避免过度扩增。

[0408] 实施例1

[0409] V(D)J方案

[0410] 本实施例展示了生成靶向小组的mRNA靶的3’末端和5’末端二者的测序文库。

[0411] 在本实施例中,生成了靶向小组的mRNA靶的3’和5’末端二者的测序文库,所述测序文库也包括T细胞受体和免疫球蛋白基因的V(D)J区。除了同时分析5’和3’通用分子索引(UMI)计数外,还鉴定了健康供体外周血单核细胞的淋巴细胞中的CDR3重排模式。

[0412] 图8A-图8C示出了使用V(D)J方案对5’T细胞受体(TCR)V(D)J区进行捕获和测序的非限制性示例性实验结果。V(D)J发夹方案可以包括在相同珠上用来自捕获的单个静息外周血单核细胞(PBMC)的mRNA分子进行的3’扩增和5’扩增。图8A描绘了细胞类型注释。图8B和图8C分别描绘了使用以这种V(D)J方案进行的5’扩增得到的TCRα和TCRβ的表达谱。TCRα链/β链的配对效率为37.9%,与其他平台诸如Clonetech的scTCR谱分析试剂盒可比较。表10描绘了使用这种V(D)J方案实现的TCRα/β配对。

[0413] 表10.TCRα/β配对(PBMC)

[0414]

	TCRα	TCRβ	二者
3’	82.6%	83.7%	71.2%
5’	56.7%	69%	37.9%

[0415] 图9A-图9B示出了图示使用改进的V(D)J方案来提高5’V(D)J检测灵敏度的非限制性示例图。添加乙二醇以帮助减少逆转录(RT)中的二级结构。改变杂交时间、缓冲液和模板转换(TS)寡聚dT长度以提高灵敏度。生成了四个文库并一起测序:5’TCR、5’BCR、5’30重免疫小组、3’免疫反应小组。进行了更严格的Ampure清理(0.6X)。图9B描绘了使用改进的V(D)J方案获得的生物分析仪图,而图9A描绘了使用参考图8描述的V(D)J方案获得的生物分析仪图。

[0416] 图10A-图10E示出了使用参考图9描述的改进的V(D)J方案来提高5’V(D)J检测灵敏度的非限制性示例性实验结果。图10A描绘了细胞类型注释。图10B-图10E描绘了使用改进的V(D)J方案获得的TCRα、TCRβ、IGKC和IGLC的表达谱。与参考图8描述的方案相比,TCRα/β配对效率从37.9%提高到52%。成功地检测到了B细胞中的5’重链和轻链mRNA分子。表11和表12分别描绘了使用改进的V(D)J方案的TCRα/β配对和BCR重/轻配对。

[0417] 表11.TCRα/β配对

[0418]

	TCRα	TCRβ	二者	未检测到
3’	82.8%	88.6%	73.4%	2%
5’	65.8%	80.5%	52.8%	6.5%

[0419] 表12.BCR重链/轻链配对

[0420]

	IGKC	KGLC	未检测到
3’	47.7%	51.9%	0.4%
5’	52.6%	40.6%	6.8%

[0421] 图11A-图11F是示出了使用具有提高的灵敏度的改进的V(D)J方案比较了使用3'扩增(图11A、图11C、图11E)和5'扩增(图11B、图11D、图11F)确定的IGHM、IGHD和IGHA的表达谱的5'B细胞重链检测的非限制性示例图。

[0422] 图12A-图12F是比较了使用3'扩增(图12A、图12C、图12E)和5'扩增(图12B、图12D、图12F)确定的CD3D、CD8A和HLA-DR的表达谱的非限制性示例图。

[0423] 总之,数据表明,对mRNA转录物信息的3'和5'两者进行谱分析的能力可以扩展3'单细胞RNA测序平台的灵活性和能力。

[0424] 尽管本文已经公开了各种方面和实施方案,但其他方面和实施方案对本领域技术人员将是明显的。本文公开的各种方面和实施方案用于说明的目的而并不意在限制由以下权利要求所指出的实际范围和精神。

[0425] 本领域技术人员将理解,对于本文公开的这个和其他的过程和方法,在过程和方法中执行的功能可以以不同的顺序实施。此外,所概述的步骤和操作仅作为示例提供,并且一些步骤和操作可以是任选的,可以组合成更少的步骤和操作,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。

[0426] 关于本文中使用基本上任何复数和/或单数术语,在对于背景和/或应用适当的情况下,本领域技术人员可以从复数转换为单数和/或从单数转换为复数。为了清楚起见,可以在本文明确阐述各种单数/复数排列。

[0427] 本领域技术人员将理解,一般来说,本文使用的术语,并且尤其是所附权利要求(例如,所附权利要求的主体)中的术语,通常意在作为"开放式"术语(例如,术语"包括(including)"应解释为"包括但不限于(including but not limited to)",术语"具有(having)"应解释为"具有至少(having at least)",术语"包括(includes)"应解释为"包括但不限于(includes but is not limited to)",等等)。本领域技术人员将进一步理解,如果所引入的权利要求陈述的特定数目是所预期,这样的预期将明确地陈述于权利要求中,并且在不存在这样的陈述的情况下,不存在这样的预期。例如,作为对理解的帮助,以下所附权利要求可以包含前置词"至少一个/至少一种(at least one)"和"一个或更多个/一种或更多种(one or more)"的使用,以引入权利要求陈述。然而,这样的短语的使用不应理解为暗含通过不定冠词"一个"或"一种"("a"or"an")引入权利要求陈述会将包含这样的引入的权利要求陈述的任何具体权利要求限制到包含仅一个这样的陈述的实施方案中,甚至在相同的权利要求包括前置词"一个或更多个/一种或更多种"或"至少一个/至少一种"以及不定冠词诸如"一个"或"一种"时也是如此(例如,"一个"和/或"一种"应解释为意指"至少一个/至少一种"或"一个或更多个/一种或更多种");这对于使用定冠词来引入权利要求陈述同样适用。此外,即使明确地陈述了所引入的权利要求陈述的特定数目,本领域技术人员将认识到,这样的陈述应解释为意指至少所陈述的数目(例如,仅陈述"两个陈述"而没有其他修饰词意指至少两个陈述,或两个或更多个陈述)。此外,在使用类似于"A、B和C等中的至少一个"的惯例的那些情况下,通常这样的句法结构以本领域技术人员将理解该惯例的意义被预期(例如,"具有A、B和C中的至少一个的系统"将包括但不限于仅具有A,仅具有B,仅具有C,A和B一起,A和C一起,B和C一起,和/或A、B和C一起等的系统)。在使用类似于"A、B或C等中的至少一个"的惯例的那些情况下,通常这样的句法结构以本领域技术人员将理解该惯例的意义被预期(例如,"具有A、B或C中的至少一个的系统"将包括但不限于仅具有A,

仅具有B,仅具有C,A和B一起,A和C一起,B和C一起,和/或A、B和C一起等的系统)。本领域技术人员将进一步理解,实际上,无论在说明书、权利要求书还是在附图中,呈现两个或更多个替代术语的任何转折性(disjunctive)词语和/或短语应被理解为考虑到包括术语之一、任一术语或两个术语的可能性。例如,短语“A或B”将被理解为包括“A”或“B”或者“A和B”的可能性。

[0428] 此外,当本公开内容的特征或方面以马库什群组(Markush group)描述时,本领域技术人员将意识到本公开内容还由此以马库什群组的任何单独的成员或成员子组描述。

[0429] 如本领域技术人员将理解的,出于任何和所有目的,诸如在提供书面描述方面,本文公开的所有范围还包括任何和所有可能的它的子范围和子范围组合。任何列出的范围都可以很容易地被识别为充分描述并使相同的范围能被分解为至少相等的一半,三分之一,四分之一,五分之一,十分之一等。作为非限制性实例,本文讨论的每个范围可以容易地分解为下三分之一,中三分之一和上三分之一等。如本领域技术人员还将理解的,所有语言,诸如“多达”、“至少”等包括所陈述的数字,并且指可以随后分解为如上文讨论的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个单独的成员。因此,例如,具有1-3个细胞的组是指具有1个、2个或3个细胞的组。类似地,具有1-5个细胞的组是指具有1个、2个、3个、4个或5个细胞的组,等等。

[0430] 从前述将理解,出于说明的目的,本文已经描述了本公开内容的各种实施方案,并且可以进行各种修改而不偏离本公开内容的精神和范围。因此,本文公开的各种实施方案不意在限制由以下权利要求所指出的实际范围和精神。

[0001] 序列表

[0002] <110> 细胞研究有限责任公司

[0003] 克里斯蒂娜·张

[0004] 艾琳·夏姆

[0005] <120> 确定5'转录物序列

[0006] <130> 68EB-298703-WO / P-16246.WO01

[0007] <150> 62/739,795

[0008] <151> 2018-10-01

[0009] <160> 1

[0010] <170> PatentIn version 3.5

[0011] <210> 1

[0012] <211> 21

[0013] <212> DNA

[0014] <213> 人工序列(Artificial Sequence)

[0015] <220>

[0016] <223> 模板转换寡核苷酸

[0017] <220>

[0018] <221> misc_feature

[0019] <222> (19)..(21)

[0020] <223> n是核糖鸟苷

[0021] <400> 1

[0022] tttttttttt tttttttttnn n 21

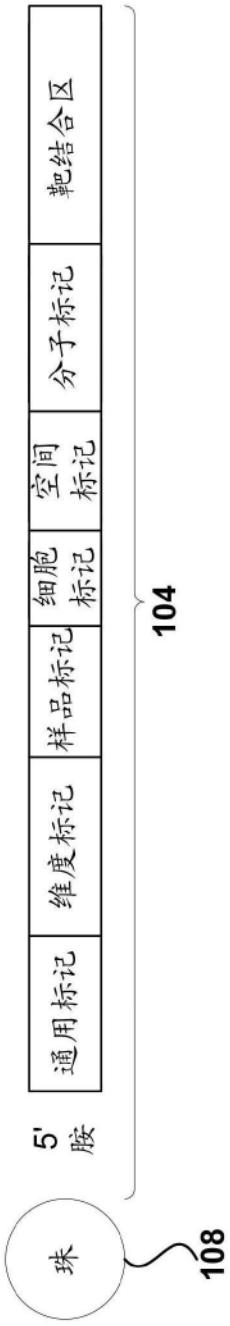


图1

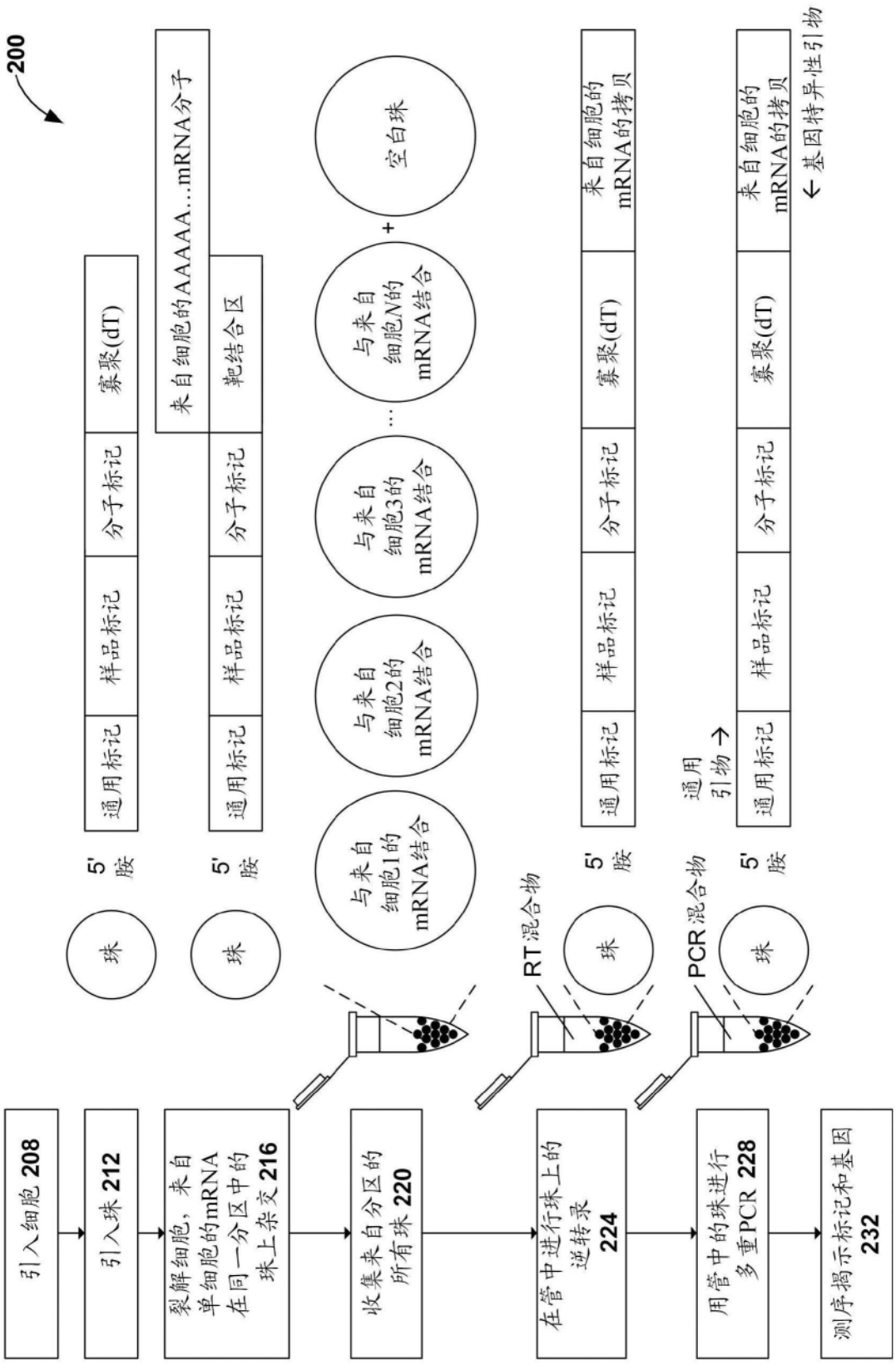

200
引入细胞 208
引入珠 212
裂解细胞，来自单细胞的mRNA在同一分区中的珠上杂交 216
收集来自分区的所有珠 220
在管中进行珠上的逆转录 224
用管中的珠进行多重PCR 228
测序揭示标记和基因 232
珠
5' 胺
通用标记
样品标记
分子标记
寡聚(dT)
来自细胞的AAAAA...mRNA分子
靶结合区
与来自细胞1的mRNA结合
与来自细胞2的mRNA结合
与来自细胞3的mRNA结合
与来自细胞N的mRNA结合
空白珠
RT混合物
PCR混合物
通用引物 →
来自细胞的mRNA的拷贝
← 基因特异性引物

图2

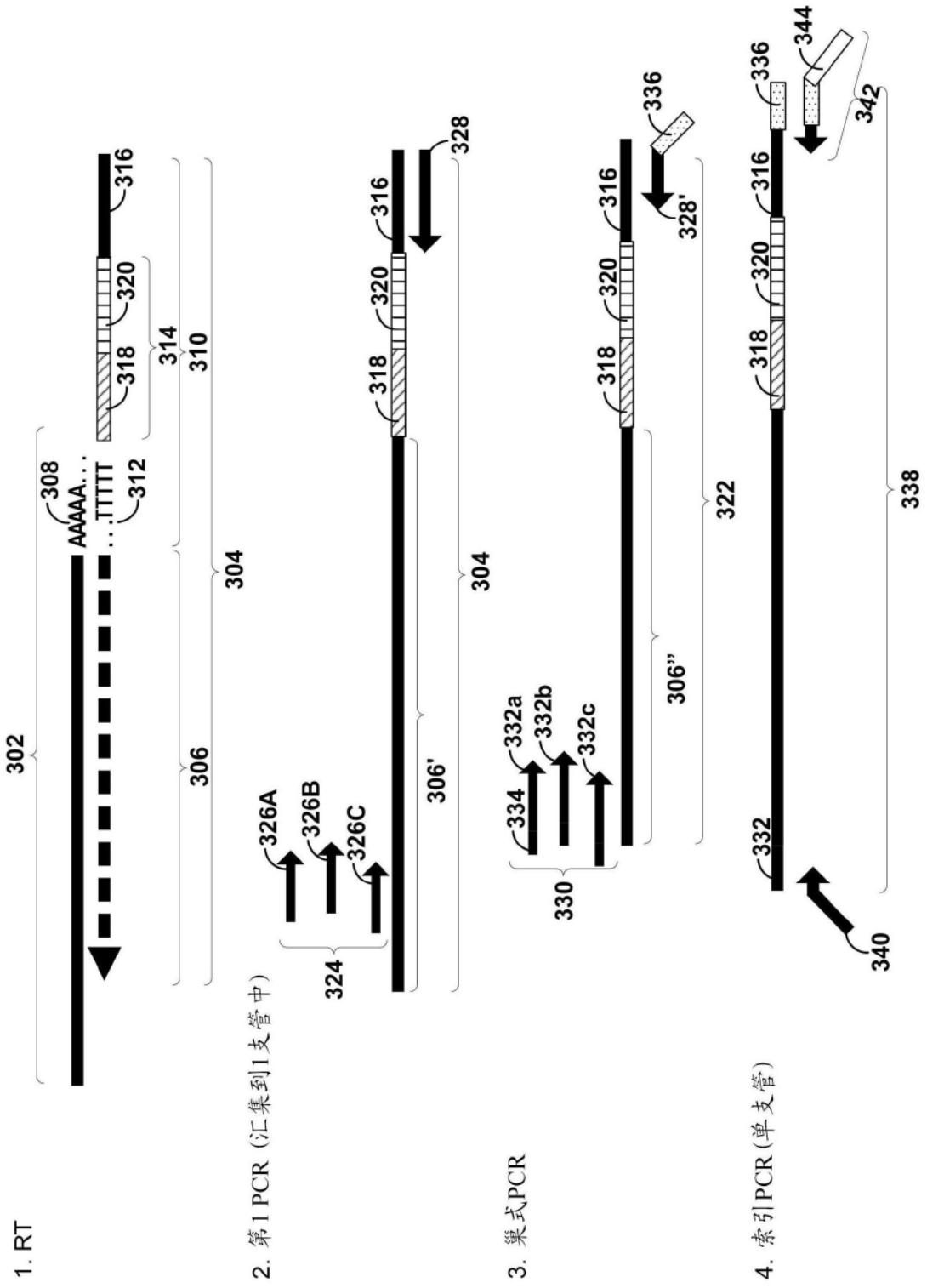

1. RT
2. 第1 PCR (汇集到1支管中)
3. 巢式PCR
4. 索引PCR (单支管)
302
308
AAAAA...
...TTTTT
312
316
320
318
314
310
304
306
306'
326A
326B
326C
324
328
328'
306''
322
330
332a
332b
332c
334
332
336
338
340
342
344

图3

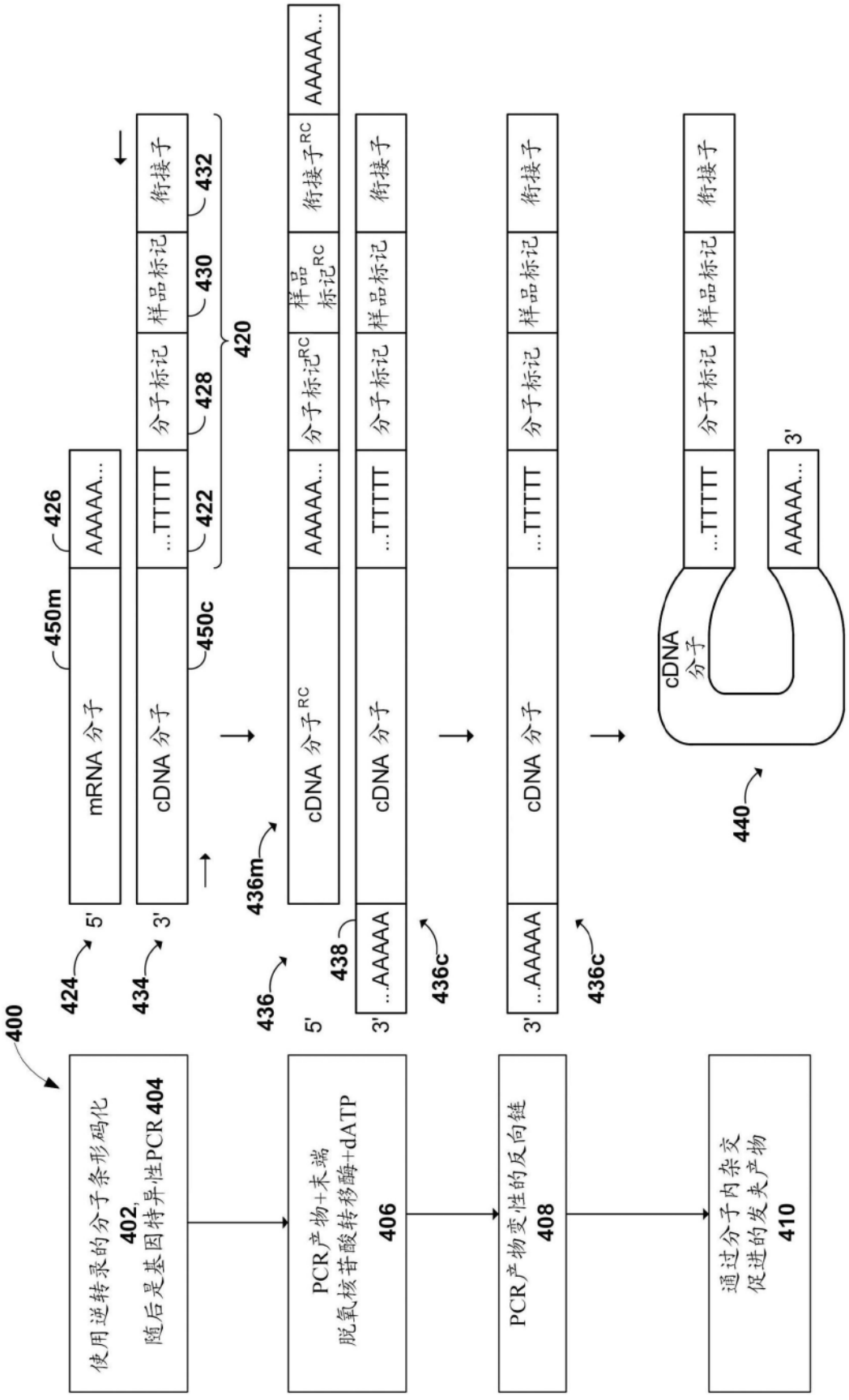

400
使用逆转录的分子条形码化
402,
随后是基因特异性PCR 404
PCR产物+末端
脱氧核苷酸转移酶+dATP
406
PCR产物变性的反向链
408
通过分子内杂交
促进的发夹产物
410
mRNA 分子
cDNA 分子
cDNA 分子RC
cDNA
分子
AAAAA...
...TTTTT
...AAAAA
分子标记
样品标记
衔接子
分子标记RC
样品标记RC
衔接子RC
AAAAA...
5'
3'
424
434
436
436m
436c
438
440
426
450m
450c
420
422
428
430
432

图4A

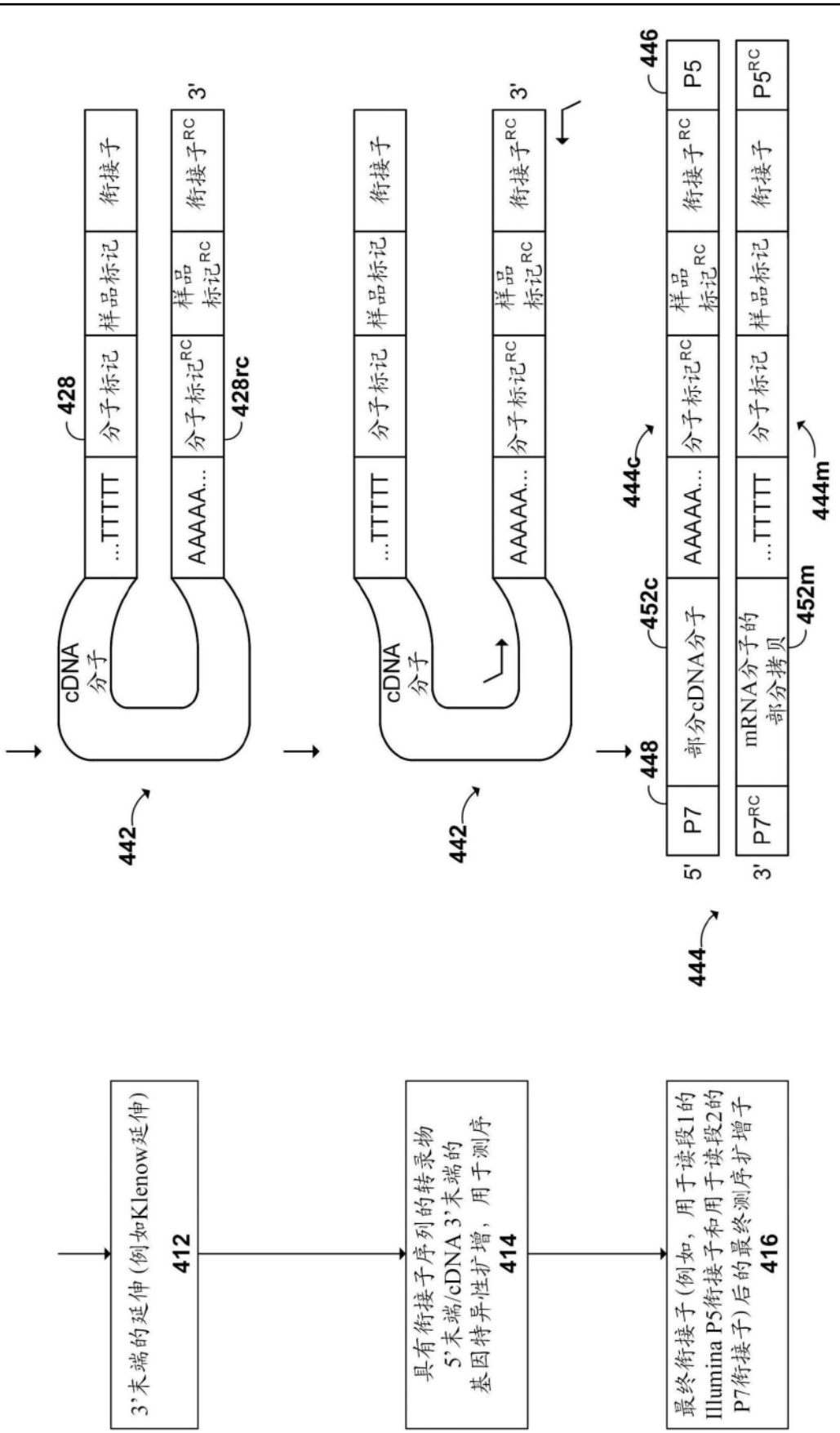

3'末端的延伸(例如Klenow延伸)
412
具有衔接子序列的转录物
5'末端/cDNA 3'末端的
基因特异性扩增，用于测序
414
最终衔接子(例如，用于读段1的
Illumina P5衔接子和用于读段2的
P7衔接子)后的最终测序扩增子
416
cDNA
分子
...TTTTT
AAAAA...
分子标记
样品标记
衔接子
分子标记RC
样品标记RC
衔接子RC
3'
428
428rc
442
444
444c
444m
446
448
452c
452m
5'
P7
P7RC
部分cDNA分子
mRNA分子的
部分拷贝
P5
P5RC

图4B

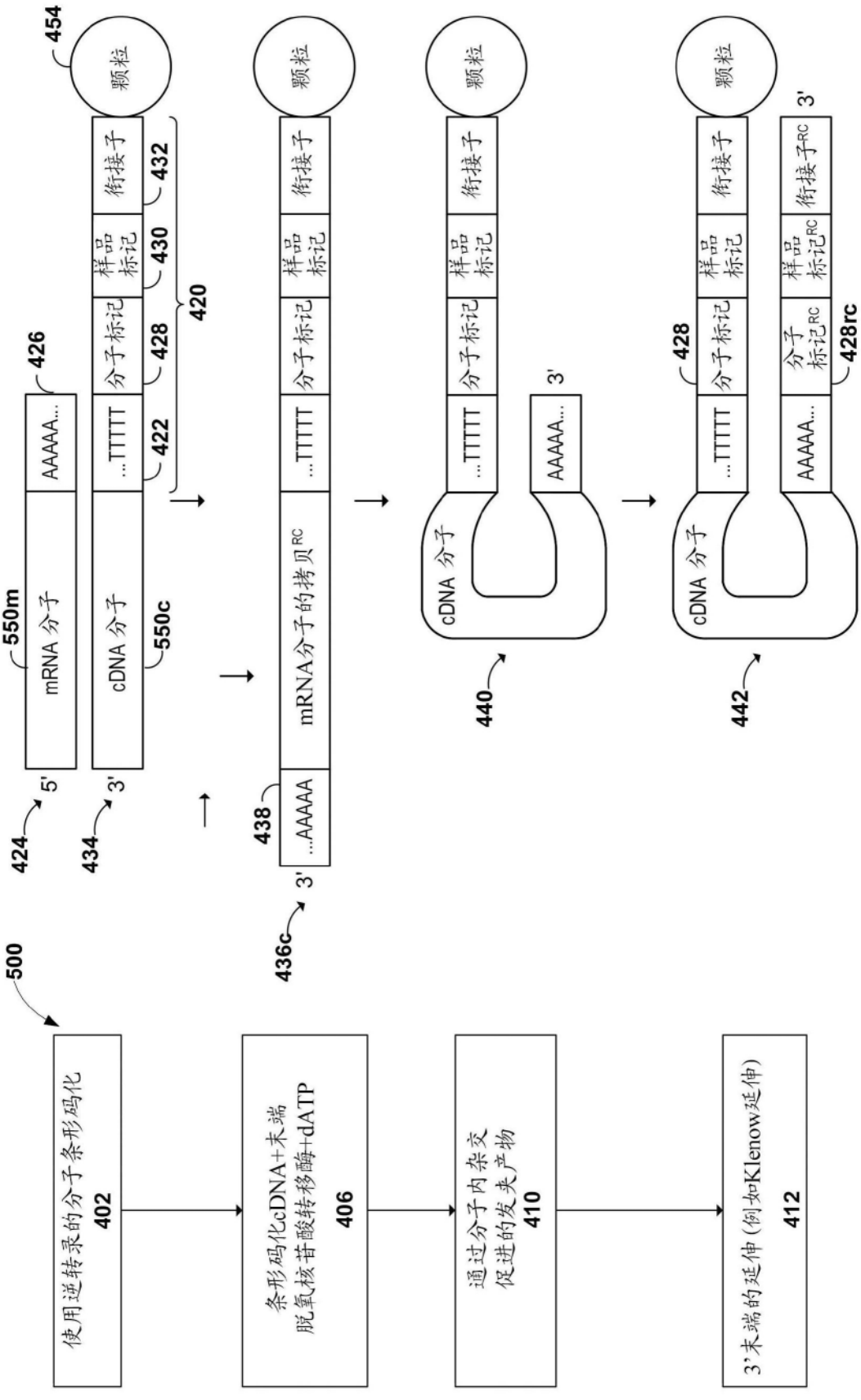

500
使用逆转录的分子条形码化
402
条形码化cDNA+末端
脱氧核苷酸转移酶+dATP
406
通过分子内杂交
促进的发夹产物
410
3'末端的延伸(例如Klenow延伸)
412
436c
438
...AAAAA
3'
mRNA分子的拷贝RC
...TTTTT
分子标记
样品标记
衔接子
颗粒
454
424
5'
mRNA 分子
AAAAA...
426
434
3'
cDNA 分子
550m
550c
420
422
428
430
432
440
442
428rc
分子标记RC
样品标记RC
衔接子RC

图5A

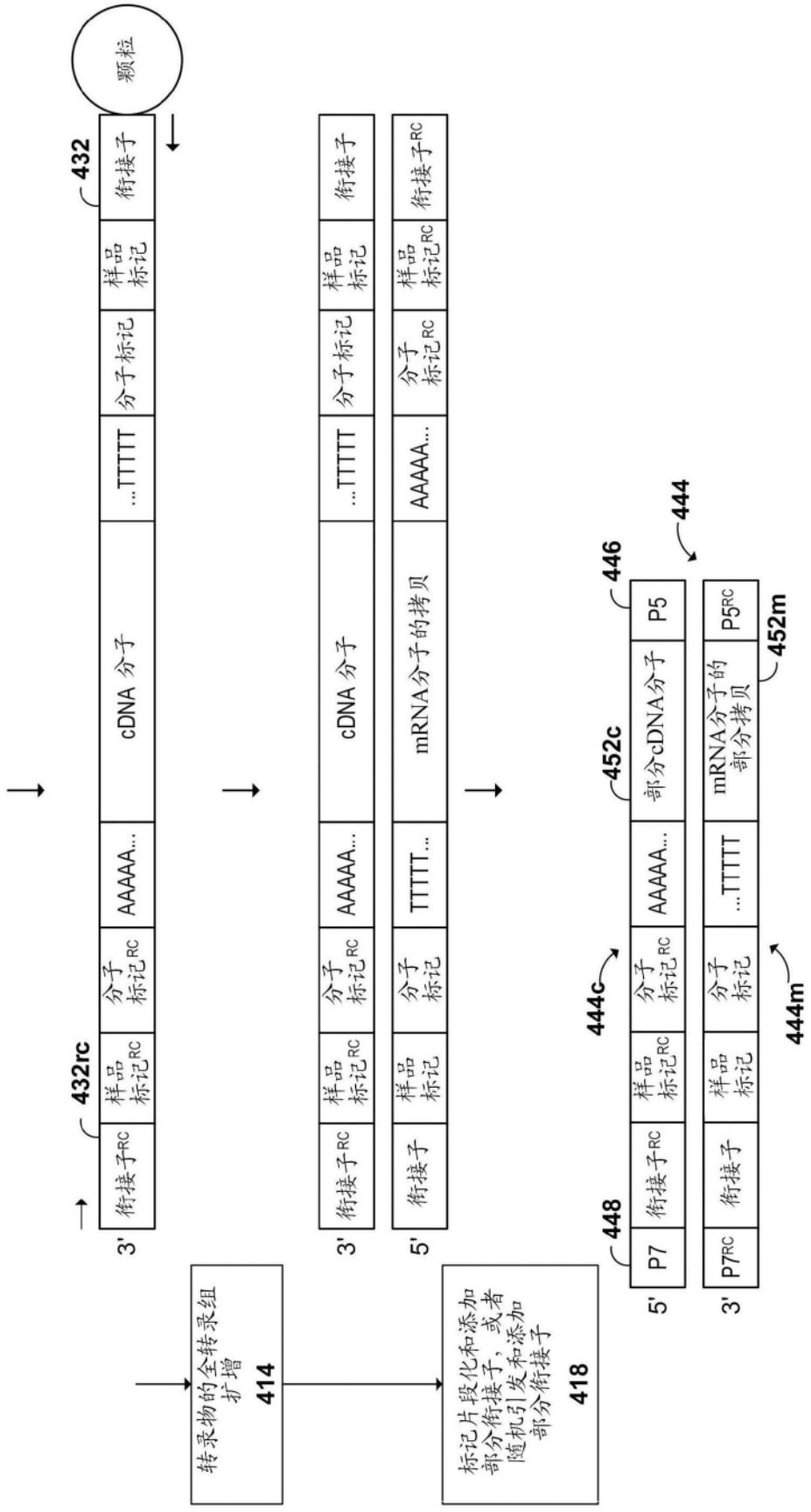

颗粒
432
3'
衔接子
样品标记
分子标记
AAAAA...
cDNA 分子
...TTTTT
分子标记
样品标记
衔接子
432rc
3'
衔接子RC
样品标记RC
分子标记RC
AAAAA...
cDNA 分子
...TTTTT
分子标记
样品标记
衔接子
转录物的全转录组扩增
414
5'
衔接子
样品标记
分子标记
TTTTT...
mRNA分子的拷贝
AAAAA...
分子标记RC
样品标记RC
衔接子RC
标记片段化和添加部分衔接子，或者随机引发和添加部分衔接子
418
444
446
452c
444c
448
452m
444m
5'
P7
衔接子RC
样品标记RC
分子标记RC
AAAAA...
部分cDNA分子
P5
3'
P7RC
衔接子
样品标记
分子标记
...TTTTT
mRNA分子的部分拷贝
P5RC

图5B

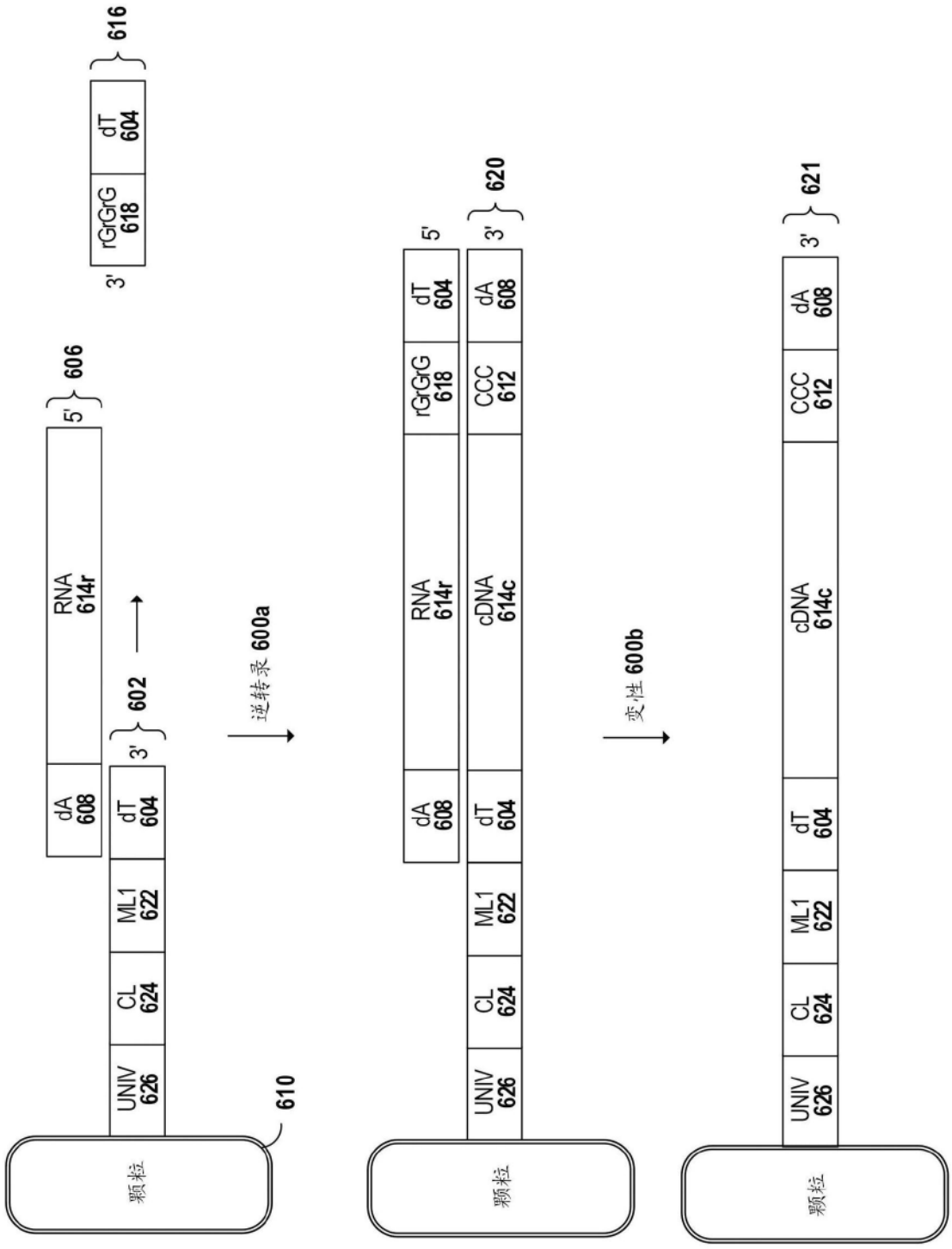

颗粒
UNIV 626
CL 624
ML1 622
dT 604
dA 608
3'
602
RNA 614r
5'
606
逆转录 600a
610
3'
rGrGrG 618
dT 604
616
变性 600b
cDNA 614c
CCC 612
620
621

图6A

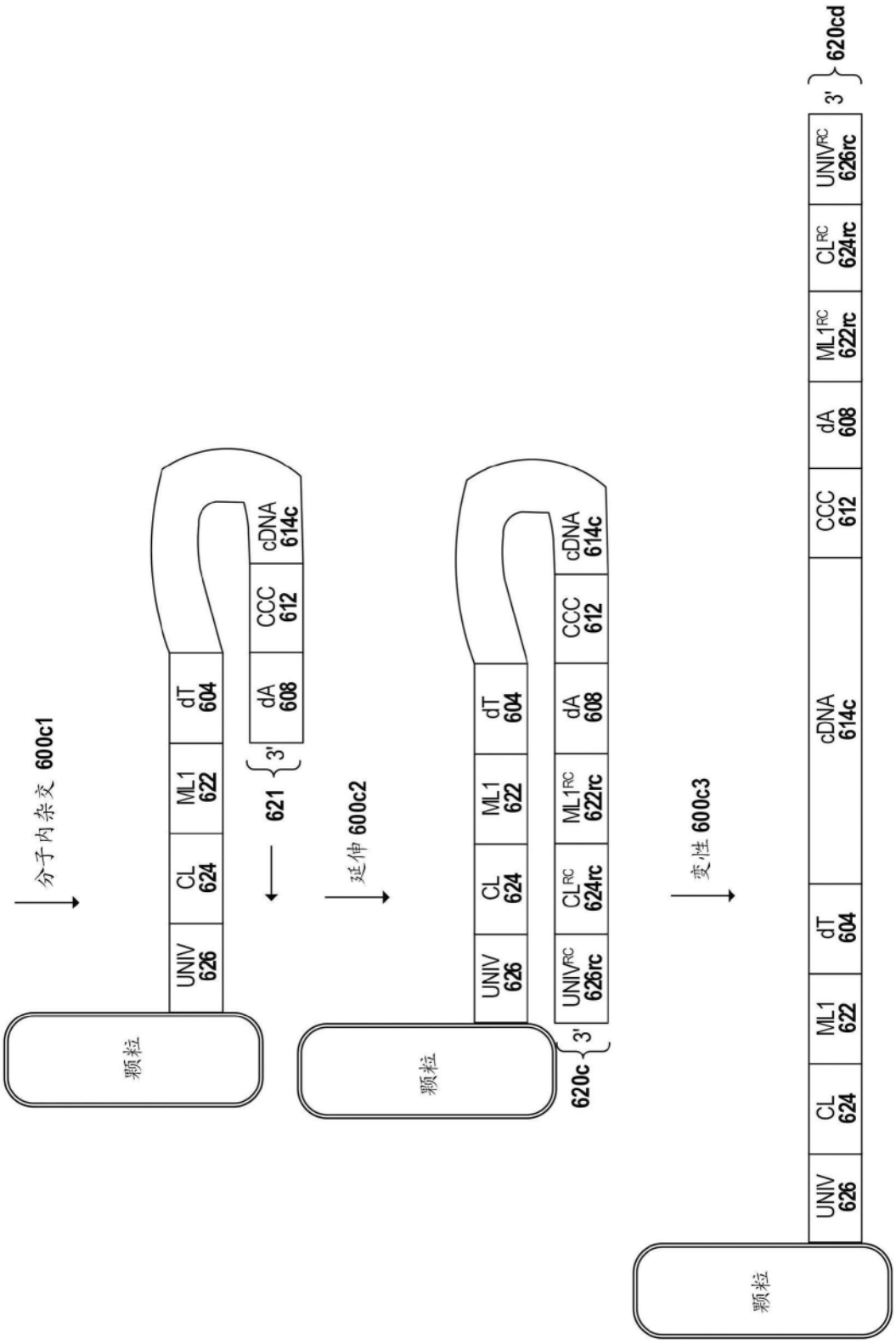

分子内杂交 600c1
延伸 600c2
变性 600c3
颗粒
UNIV 626
CL 624
ML1 622
dT 604
dA 608
CCC 612
cDNA 614c
3'
621
620c
620cd
UNIV^RC 626rc
CL^RC 624rc
ML1^RC 622rc

图6B

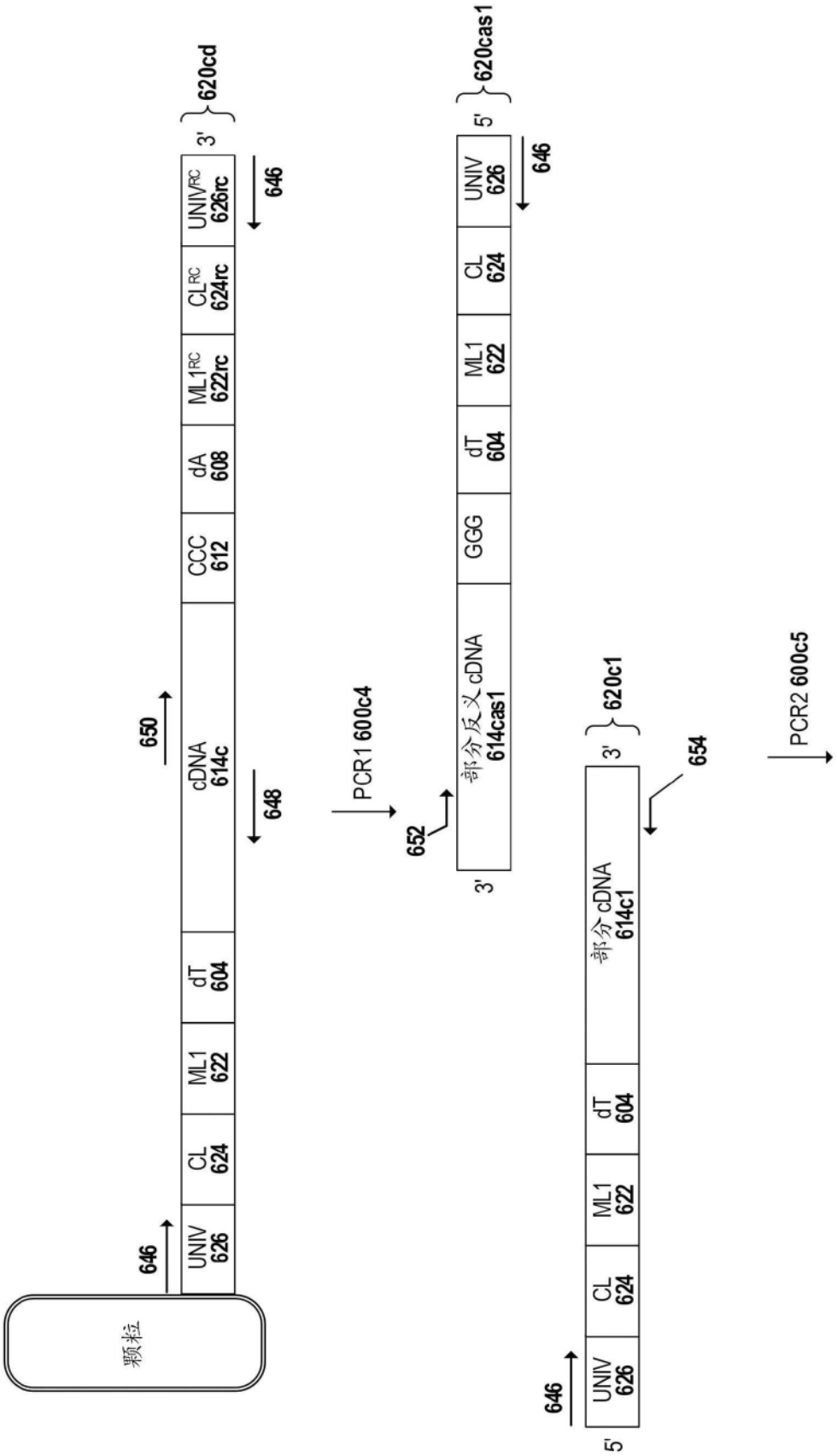

颗粒
646
UNIV 626
CL 624
ML1 622
dT 604
cDNA 614c
650
648
CCC 612
dA 608
ML1RC 622rc
CLRC 624rc
UNIVRC 626rc
3'
620cd
646
PCR1 600c4
652
3'
部分反义cDNA 614cas1
GGG
dT 604
ML1 622
CL 624
UNIV 626
5'
620cas1
646
PCR2 600c5
654
3'
部分cDNA 614c1
620c1
646
5'
UNIV 626
CL 624
ML1 622
dT 604

图6C

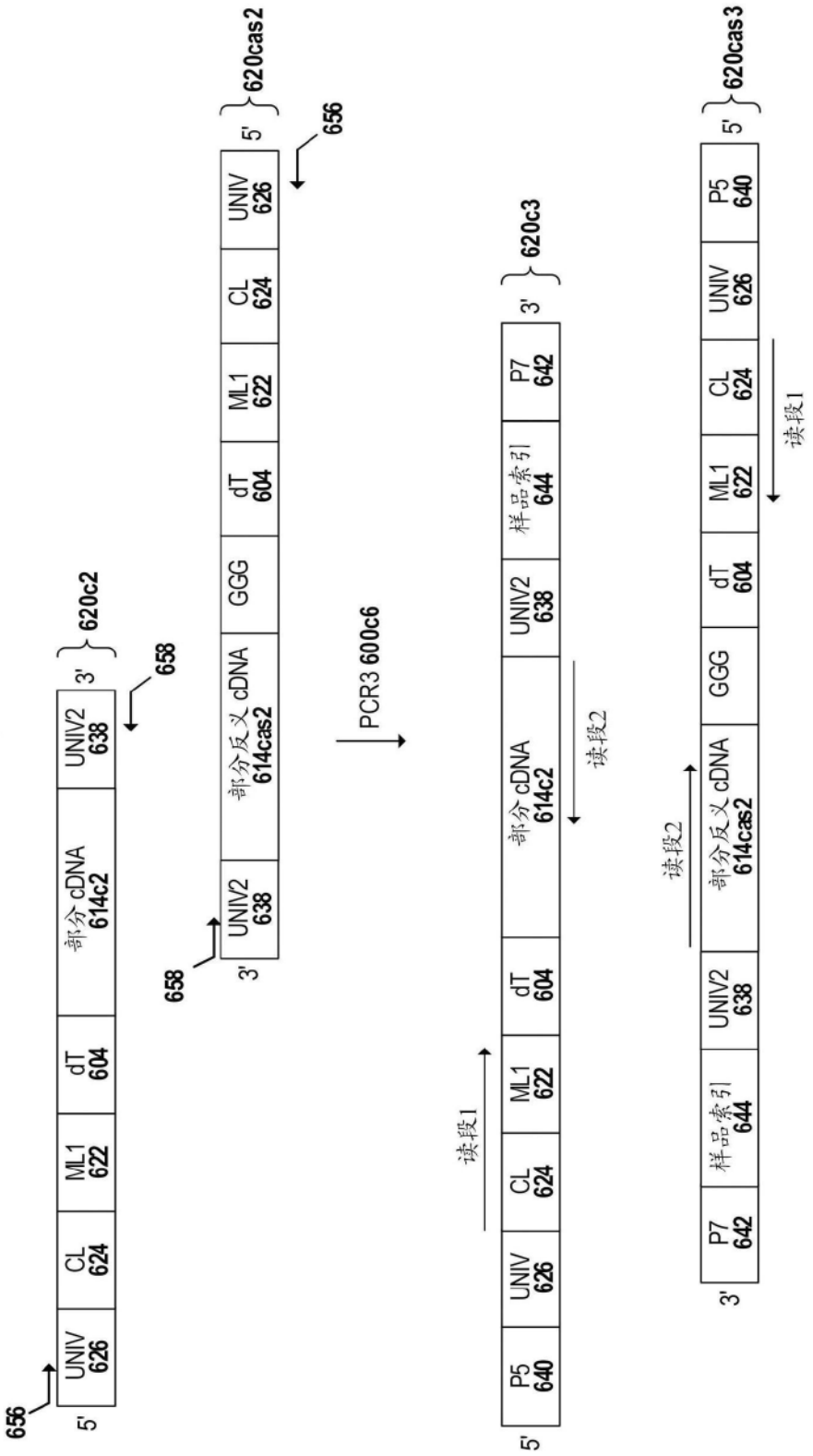

656
5'
UNIV 626
CL 624
ML1 622
dT 604
部分cDNA 614c2
UNIV2 638
3'
620c2
658
658
3'
UNIV2 638
部分反义cDNA 614cas2
GGG
dT 604
ML1 622
CL 624
UNIV 626
5'
620cas2
656
PCR3 600c6
5'
P5 640
UNIV 626
CL 624
ML1 622
dT 604
部分cDNA 614c2
UNIV2 638
样品索引 644
P7 642
3'
620c3
读段1
读段2
读段2
3'
P7 642
样品索引 644
UNIV2 638
部分反义cDNA 614cas2
GGG
dT 604
ML1 622
CL 624
UNIV 626
P5 640
5'
620cas3
读段1

图6D

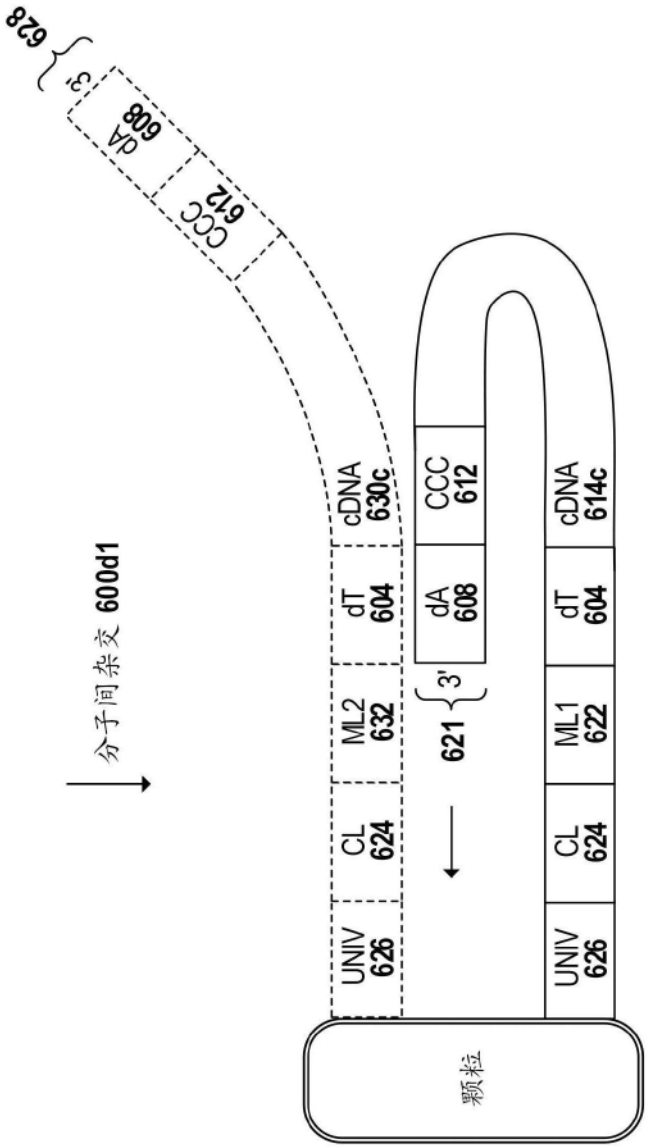

628
3'
dA
608
CCC
612
cDNA
630c
dT
604
ML2
632
CL
624
UNIV
626
分子间杂交 600d1
3'
621
CCC
612
dA
608
dT
604
ML1
622
CL
624
UNIV
626
cDNA
614c
颗粒

图6E

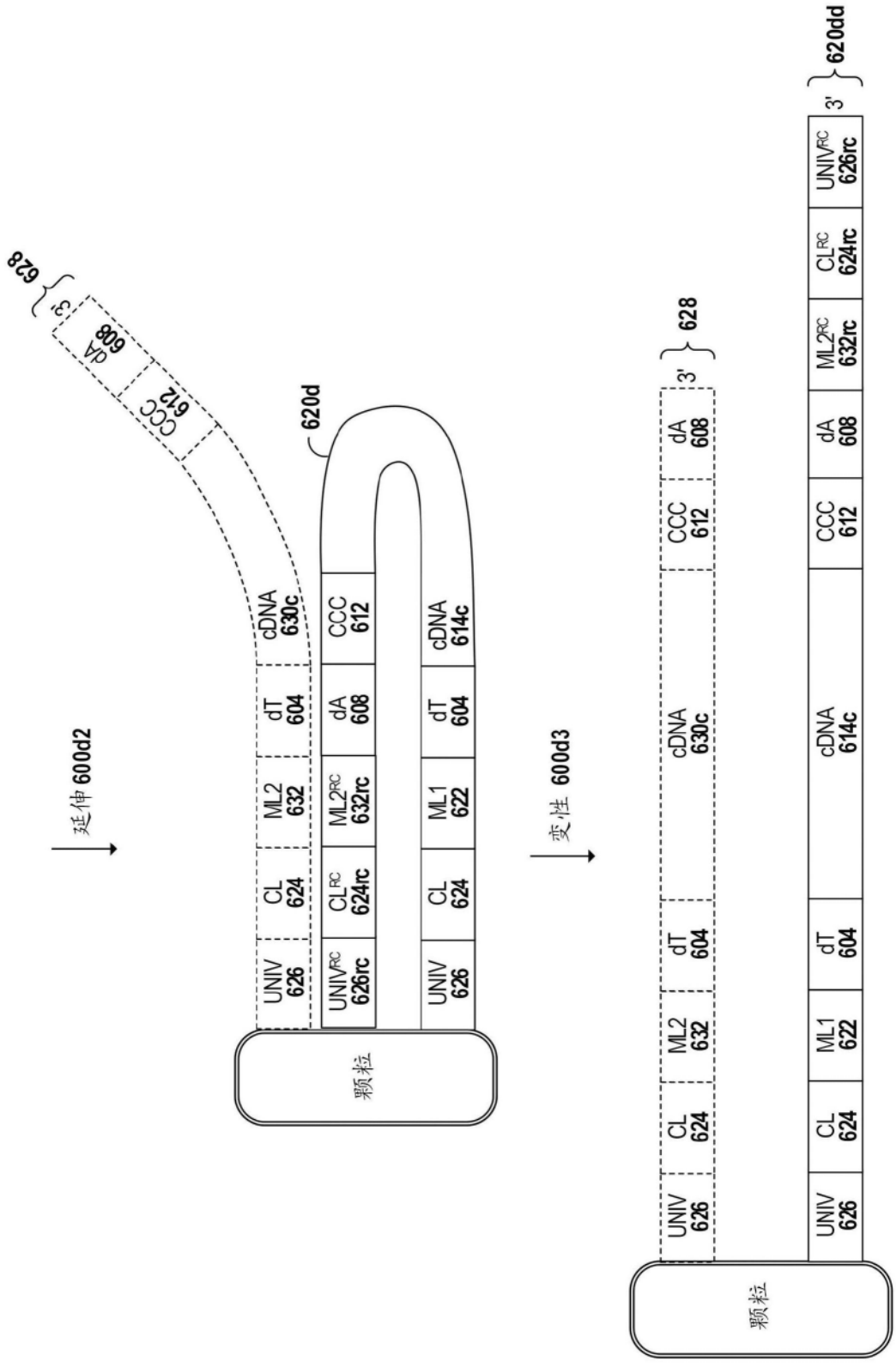

延伸 600d2
变性 600d3
628
3'
dA 608
CCC 612
620d
cDNA 630c
dT 604
ML2 632
CL 624
UNIV 626
ML2RC 632rc
CLRC 624rc
UNIVRC 626rc
cDNA 614c
ML1 622
620dd
UNIVRC 626rc
CLRC 624rc
ML2RC 632rc
颗粒

图6F

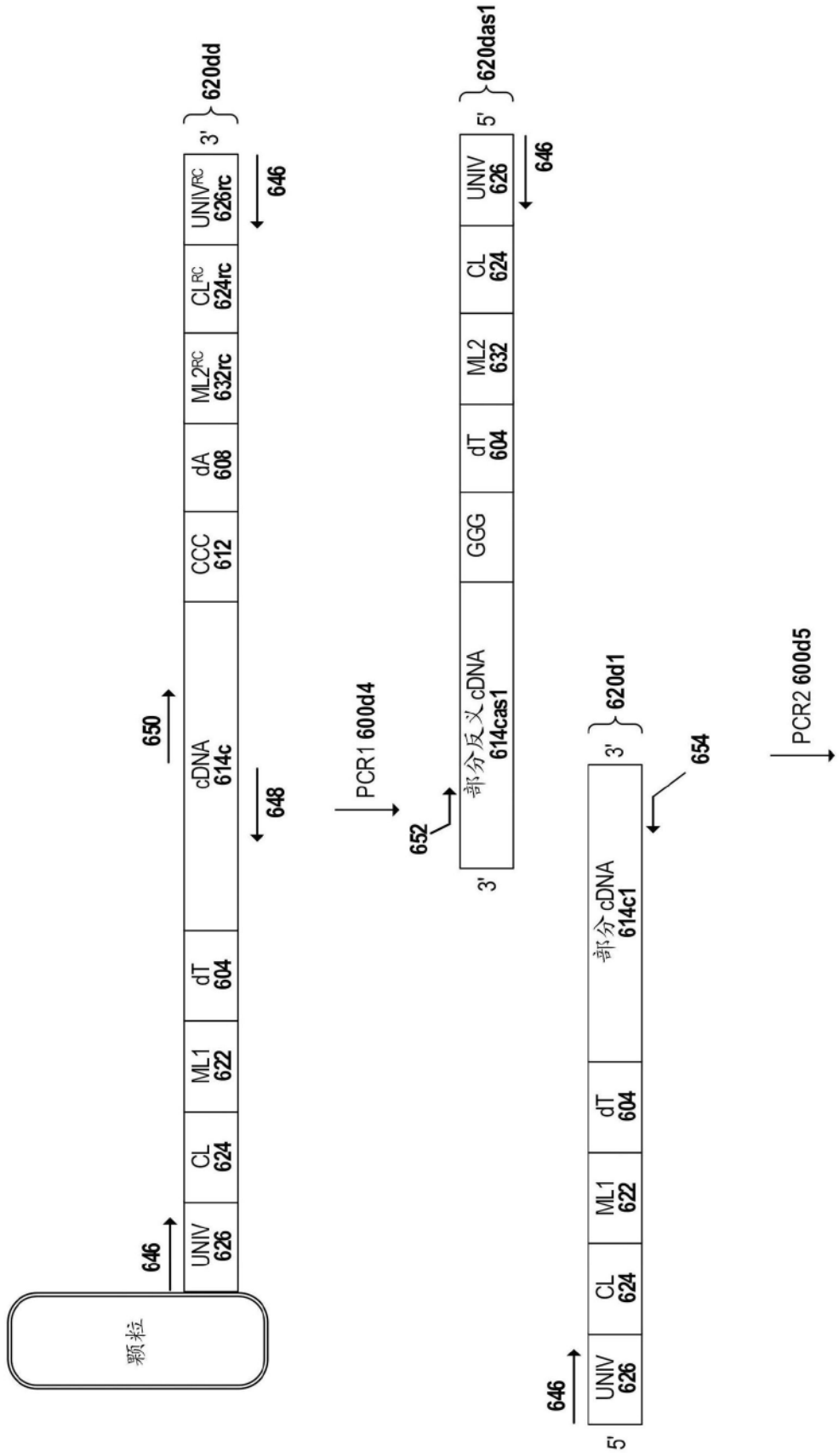

颗粒
646
UNIV 626
CL 624
ML1 622
dT 604
cDNA 614c
650
CCC 612
dA 608
ML2RC 632rc
CLRC 624rc
UNIVRC 626rc
3'
620dd
646
648
PCR1 600d4
652
3'
部分反义cDNA 614cas1
GGG
dT 604
ML2 632
CL 624
UNIV 626
5'
620das1
646
PCR2 600d5
654
3'
620d1
部分cDNA 614c1
646
5'
UNIV 626
CL 624
ML1 622
dT 604

图6G

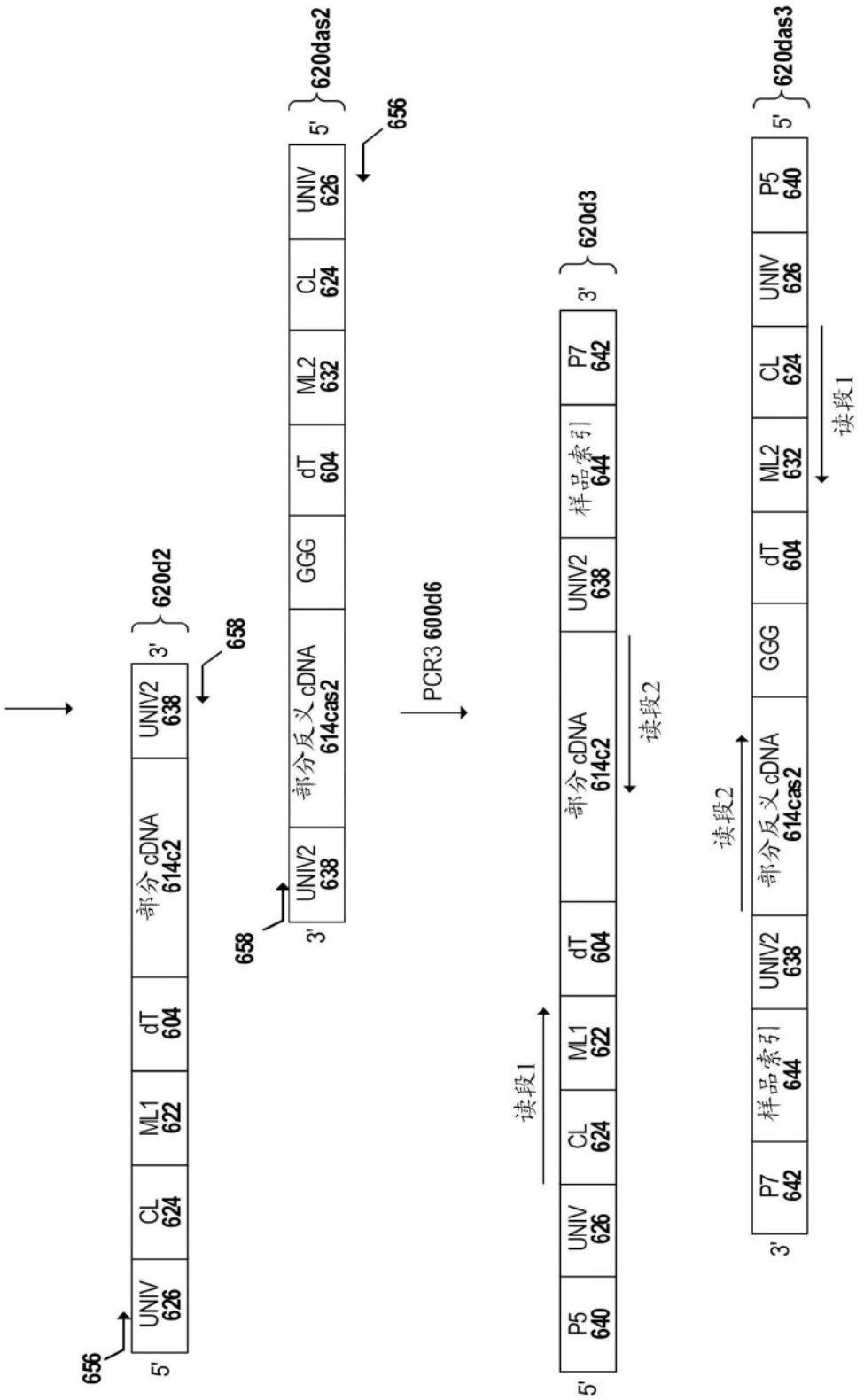

656
5'
UNIV 626
CL 624
ML1 622
dT 604
部分cDNA 614c2
UNIV2 638
3'
620d2
658
658
3'
UNIV2 638
部分反义cDNA 614cas2
GGG
dT 604
ML2 632
CL 624
UNIV 626
5'
620das2
656
PCR3 600d6
5'
P5 640
UNIV 626
CL 624
ML1 622
dT 604
部分cDNA 614c2
UNIV2 638
样品索引 644
P7 642
3'
620d3
读段1
读段2
读段2
3'
P7 642
样品索引 644
UNIV2 638
部分反义cDNA 614cas2
GGG
dT 604
ML2 632
CL 624
UNIV 626
P5 640
5'
620das3
读段1

图6H

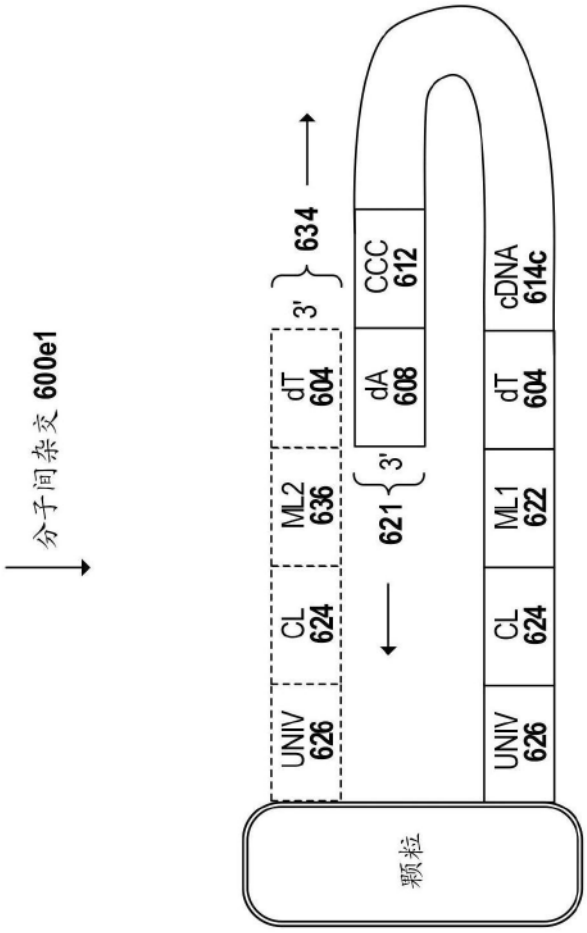

分子间杂交 600e1
UNIV 626
CL 624
ML2 636
dT 604
3'
621
dA 608
CCC 612
634
颗粒
UNIV 626
CL 624
ML1 622
dT 604
cDNA 614c

图6I

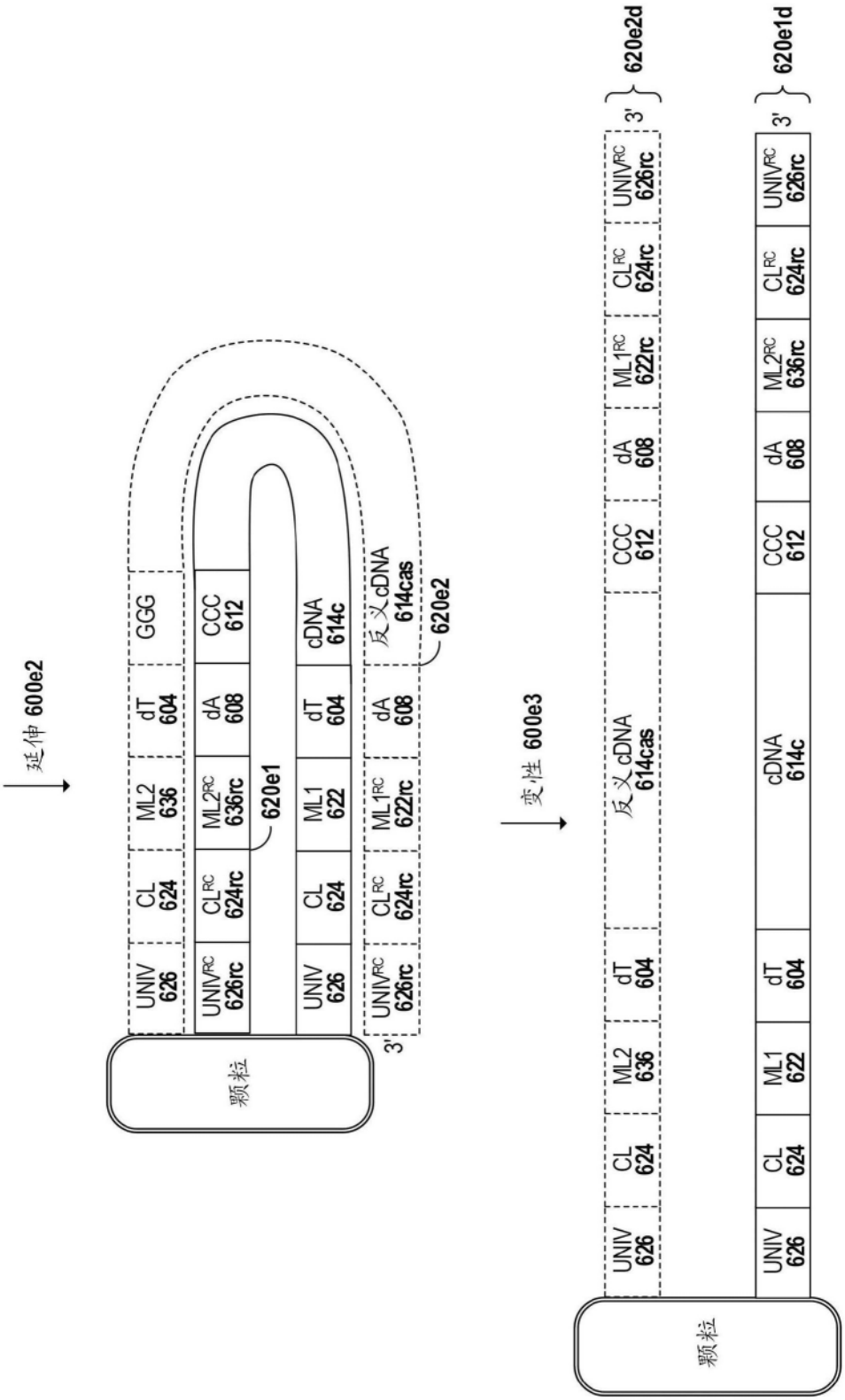

延伸 600e2
UNIV 626
CL 624
ML2 636
dT 604
GGG
UNIV^RC 626rc
CL^RC 624rc
ML2^RC 636rc
dA 608
CCC 612
620e1
ML1 622
dT 604
cDNA 614c
ML1^RC 622rc
dA 608
反义cDNA 614cas
620e2
3'
颗粒
变性 600e3
反义cDNA 614cas
cDNA 614c
ML1^RC 622rc
ML2^RC 636rc
CL^RC 624rc
620e2d
620e1d

图6J

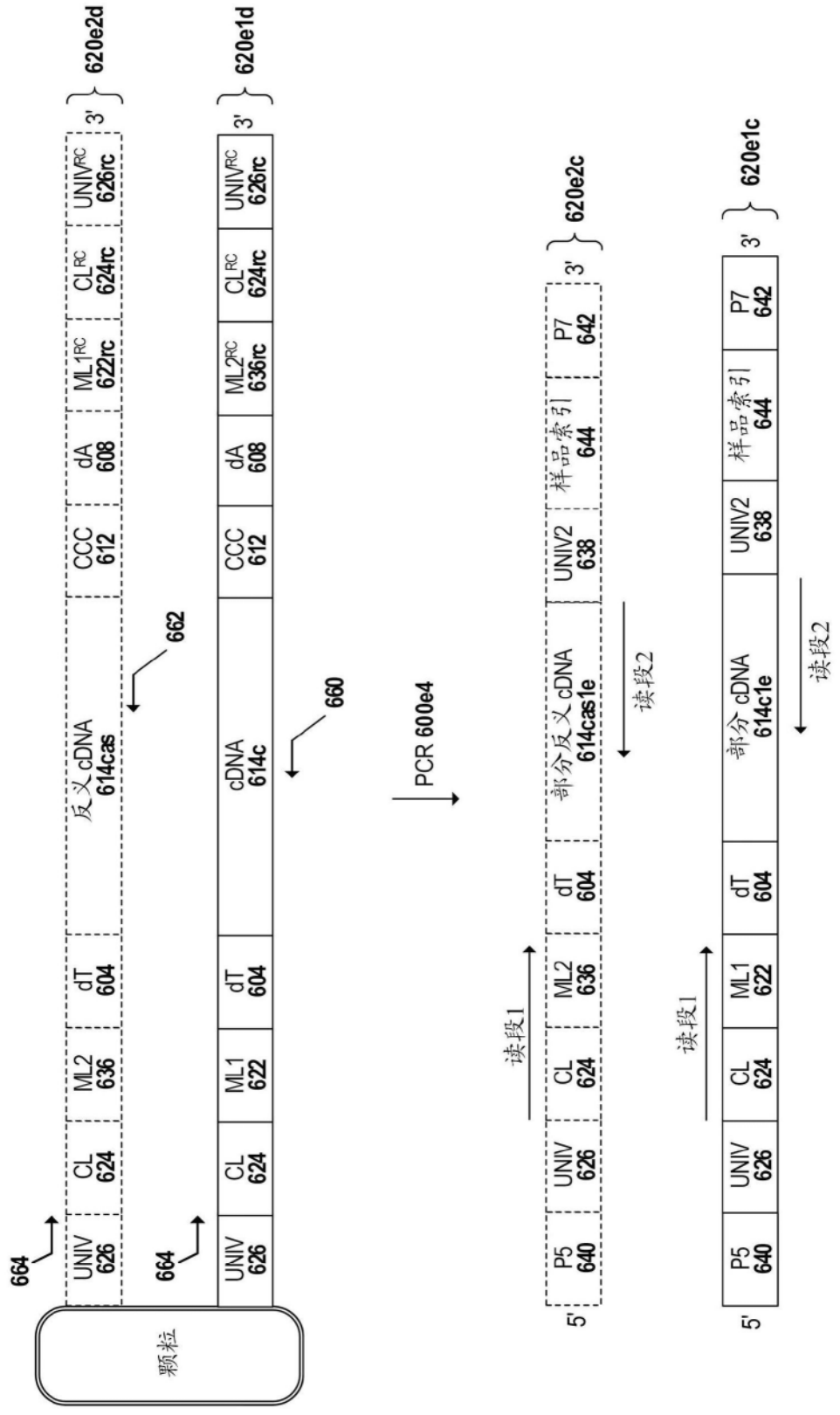

颗粒
664
UNIV 626
CL 624
ML2 636
dT 604
反义cDNA 614cas
662
CCC 612
dA 608
ML1RC 622rc
CLRC 624rc
UNIVRC 626rc
3'
620e2d
664
UNIV 626
CL 624
ML1 622
dT 604
cDNA 614c
660
CCC 612
dA 608
ML2RC 636rc
CLRC 624rc
UNIVRC 626rc
3'
620e1d
PCR 600e4
5'
P5 640
UNIV 626
CL 624
ML2 636
dT 604
部分反义cDNA 614cas1e
UNIV2 638
样品索引 644
P7 642
3'
620e2c
读段1
读段2
5'
P5 640
UNIV 626
CL 624
ML1 622
dT 604
部分cDNA 614c1e
UNIV2 638
样品索引 644
P7 642
3'
620e1c
读段1
读段2

图6K

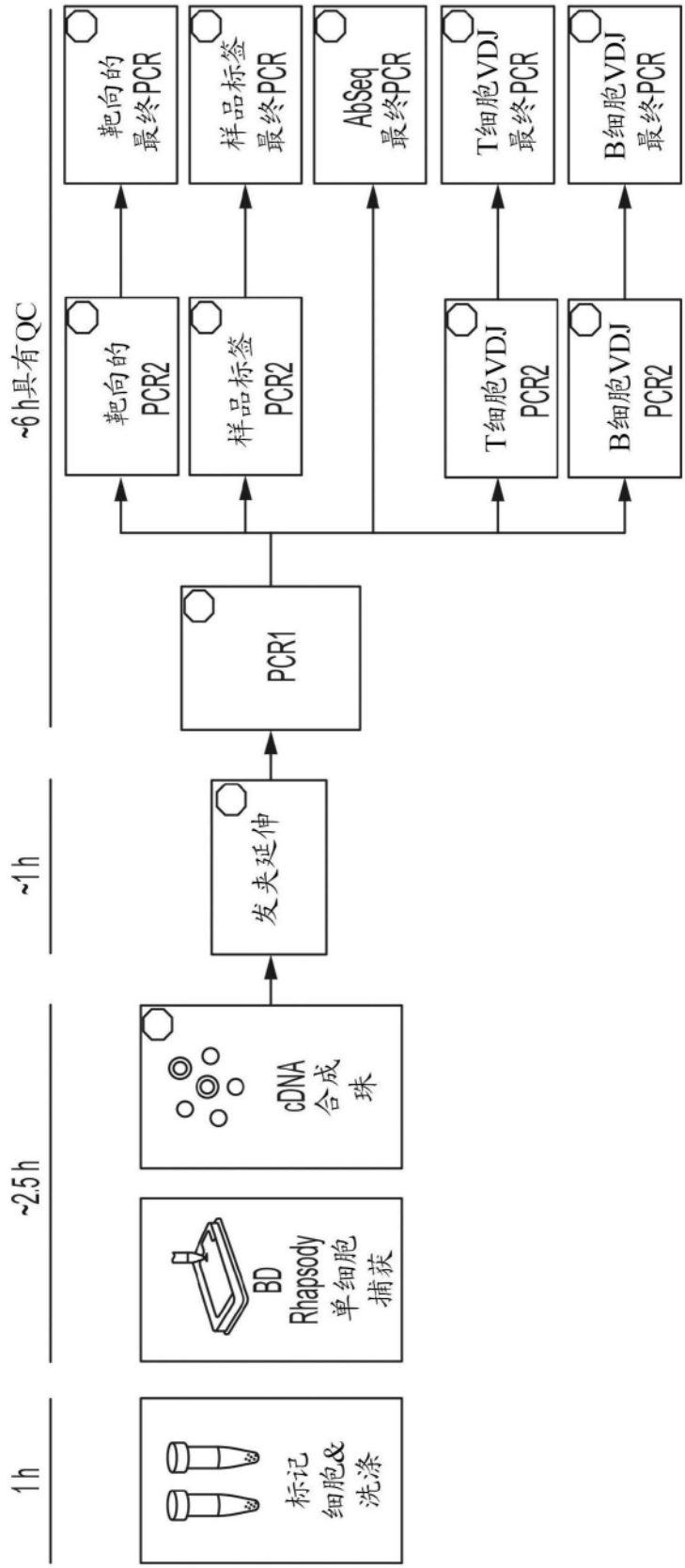

1h
~2.5h
~1h
~6h具有QC
标记细胞&洗涤
BD Rhapsody单细胞捕获
cDNA合成珠
发夹延伸
PCR1
靶向的PCR2
样品标签PCR2
T细胞VDJ PCR2
B细胞VDJ PCR2
靶向的最终PCR
样品标签最终PCR
AbSeq最终PCR
T细胞VDJ最终PCR
B细胞VDJ最终PCR

图7

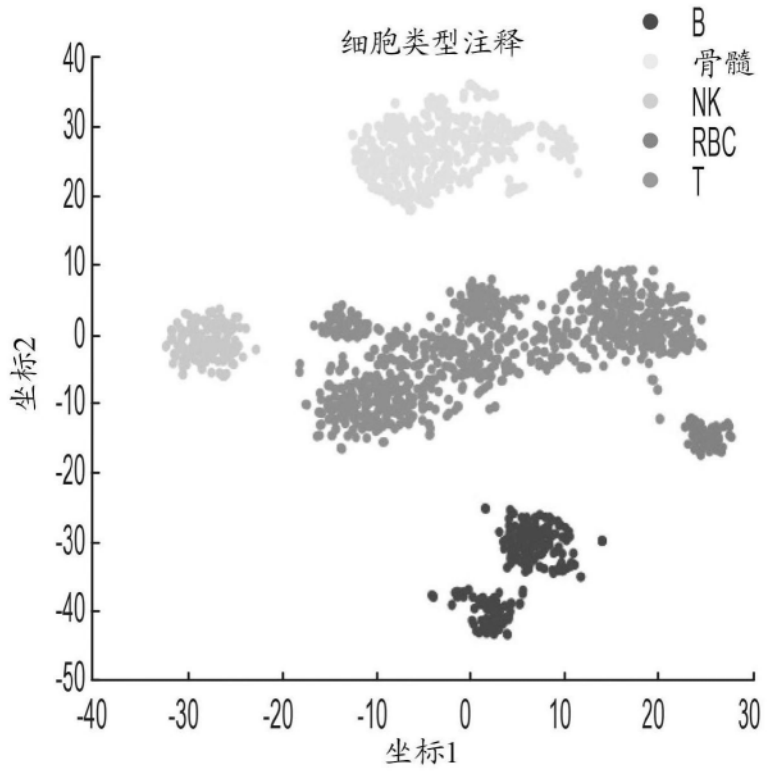

细胞类型注释
B
骨髓
NK
RBC
T
坐标2
坐标1
40
30
20
10
0
-10
-20
-30
-40
-50
-40
-30
-20
-10
0
10
20
30

图8A

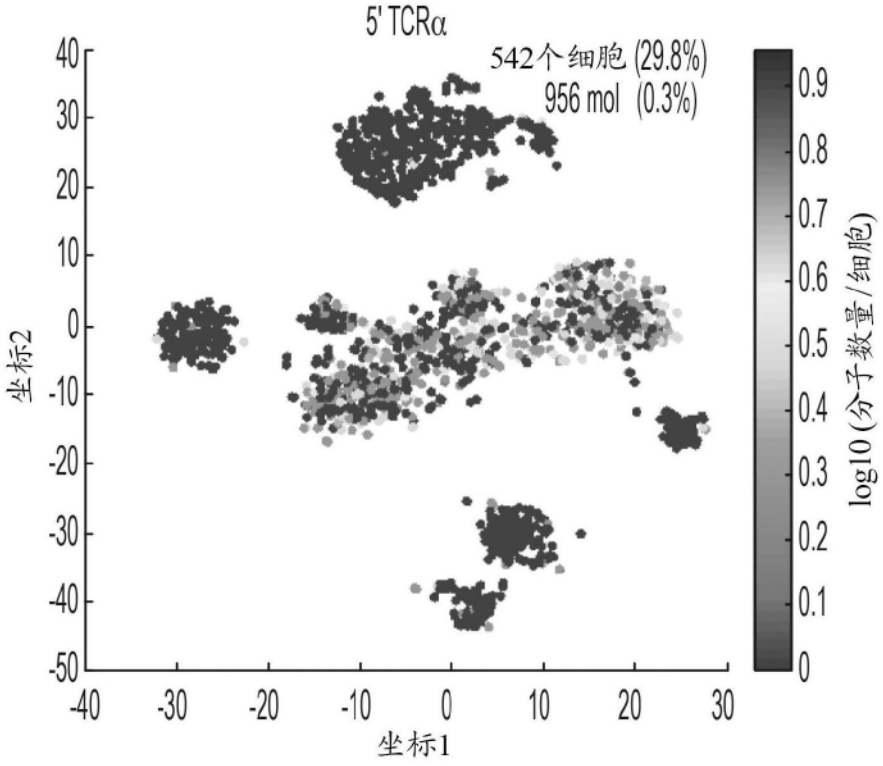

5'TCRα
542个细胞 (29.8%)
956 mol (0.3%)
log10 (分子数量/细胞)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
坐标2
坐标1
40
30
20
10
0
-10
-20
-30
-40
-50
-40
-30
-20
-10
0
10
20
30

图8B

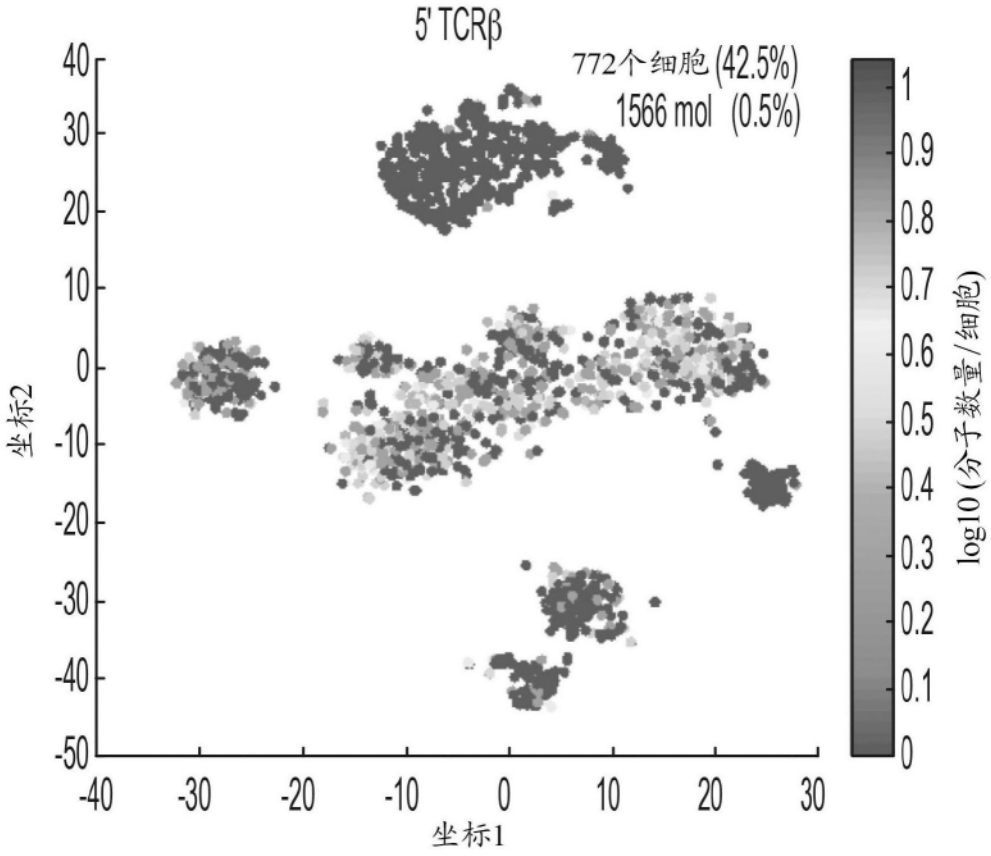

5' TCRβ
772个细胞 (42.5%)
1566 mol (0.5%)
坐标2
坐标1
log10 (分子数量/细胞)
40
30
20
10
0
-10
-20
-30
-40
-50
-40
-30
-20
-10
0
10
20
30
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

图8C

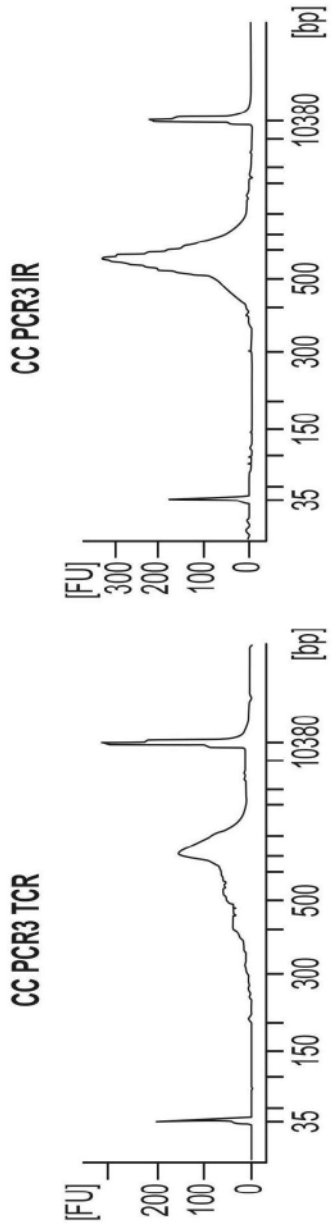

CC PCR3 IR
[FU]
300
200
100
0
35
150
300
500
10380
[bp]
CC PCR3 TCR
[FU]
200
100
0
35
150
300
500
10380
[bp]

图9A

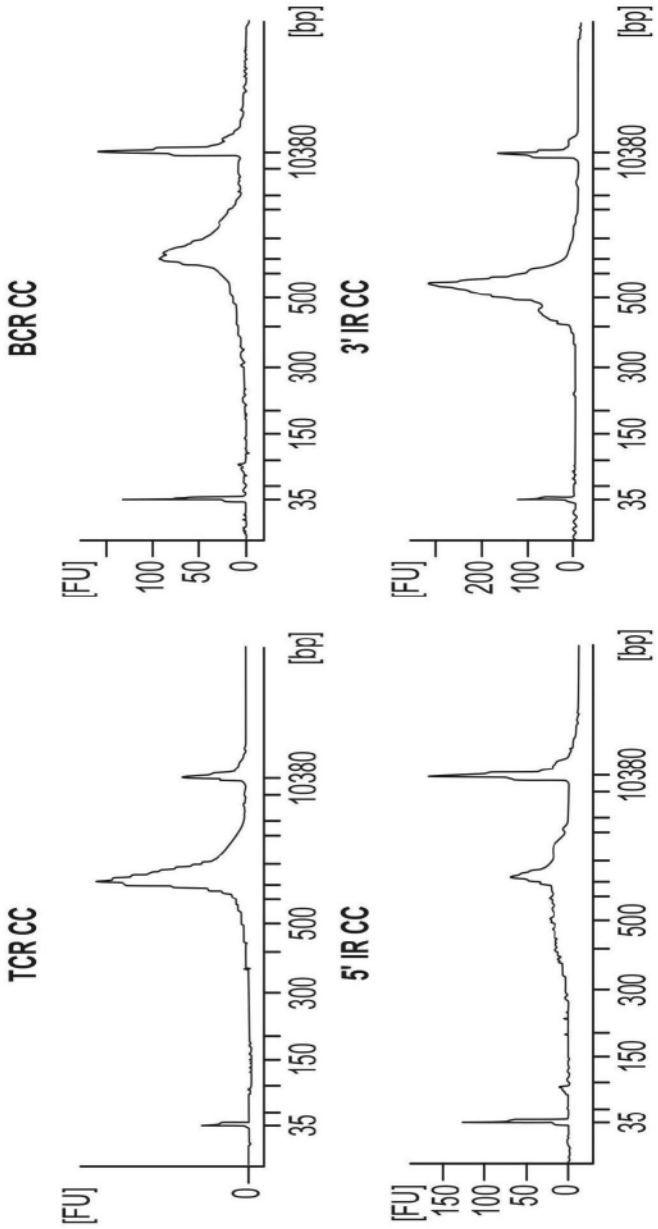

BCR CC
3' IR CC
TCR CC
5' IR CC
[FU]
[bp]
0
35
150
300
500
10380
50
100
200
150

图9B

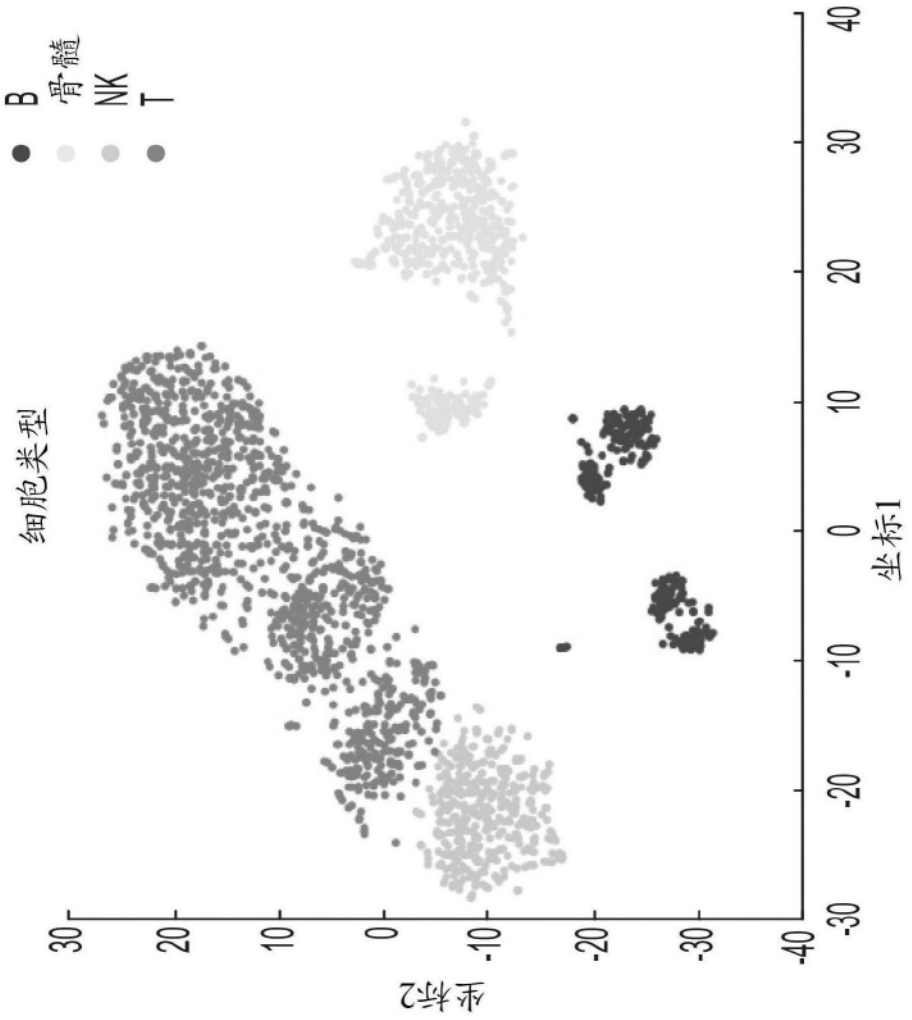


图10A

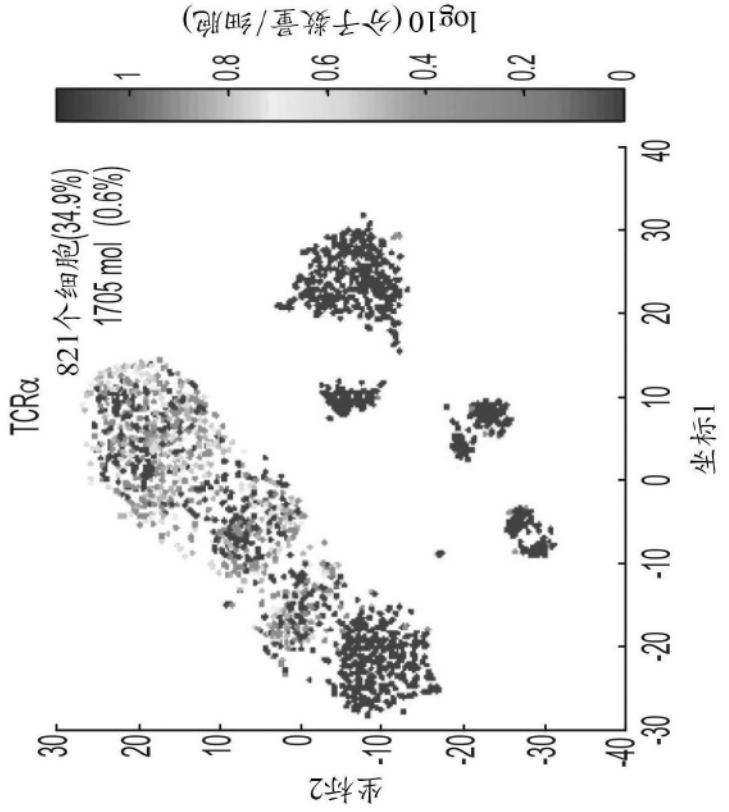

log10 (分子数量/细胞)
1
0.8
0.6
0.4
0.2
0
TCRα
821个细胞(34.9%)
1705 mol (0.6%)
坐标2
30
20
10
0
-10
-20
-30
-40
坐标1
40
30
20
10
0
-10
-20
-30

图10B

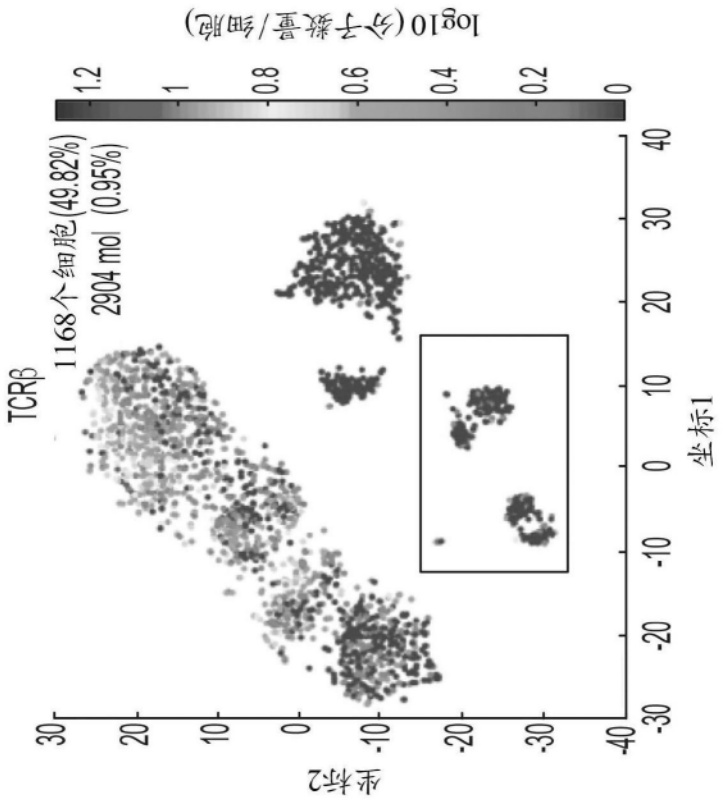

log10(分子数量/细胞)
1.2
1
0.8
0.6
0.4
0.2
0
TCRβ
1168个细胞(49.82%)
2904 mol (0.95%)
30
20
10
0
-10
-20
-30
-40
坐标2
40
30
20
10
0
-10
-20
-30
坐标1

图10C

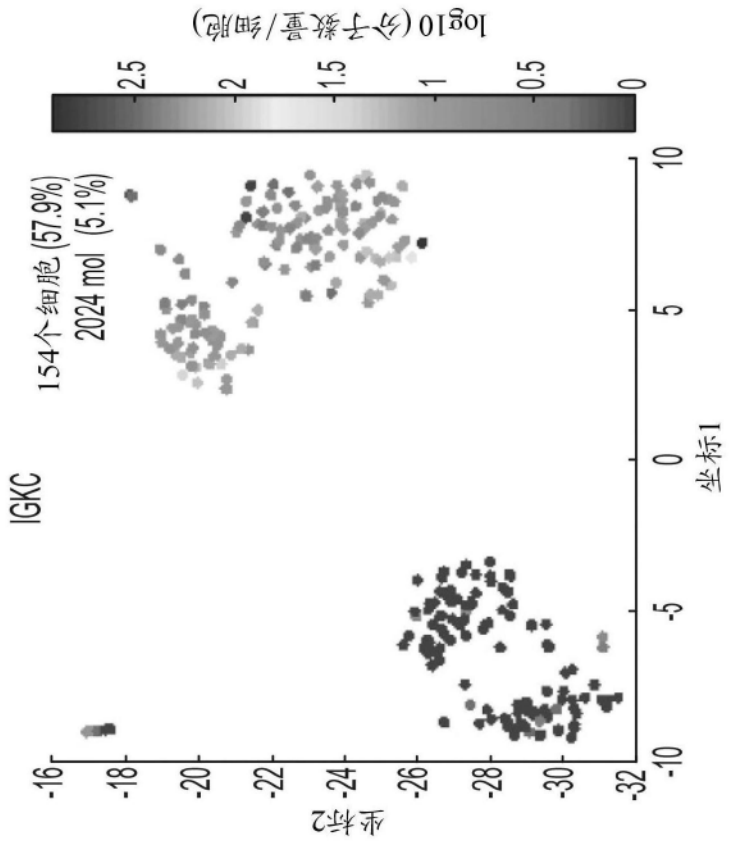

log10 (分子数量/细胞)
2.5
2
1.5
1
0.5
0
154个细胞 (57.9%)
2024 mol (5.1%)
IGKC
10
5
0
-5
-10
坐标1
-16
-18
-20
-22
-24
-26
-28
-30
-32
坐标2

图10D

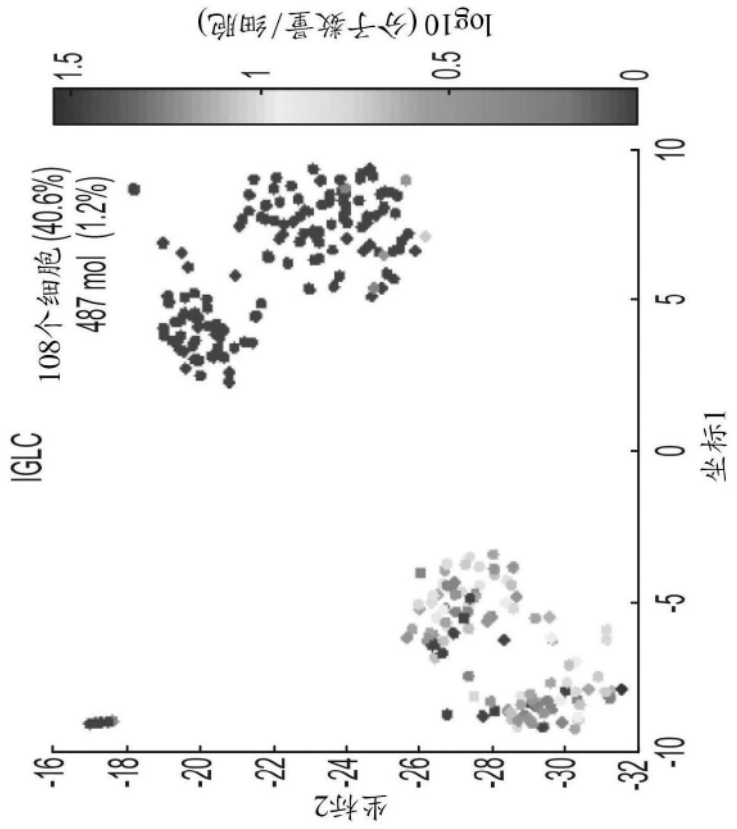

log10 (分子数量/细胞)
1.5
1
0.5
0
IGLC
108个细胞 (40.6%)
487 mol (1.2%)
坐标1
坐标2
10
5
0
-5
-10
-16
-18
-20
-22
-24
-26
-28
-30
-32

图10E

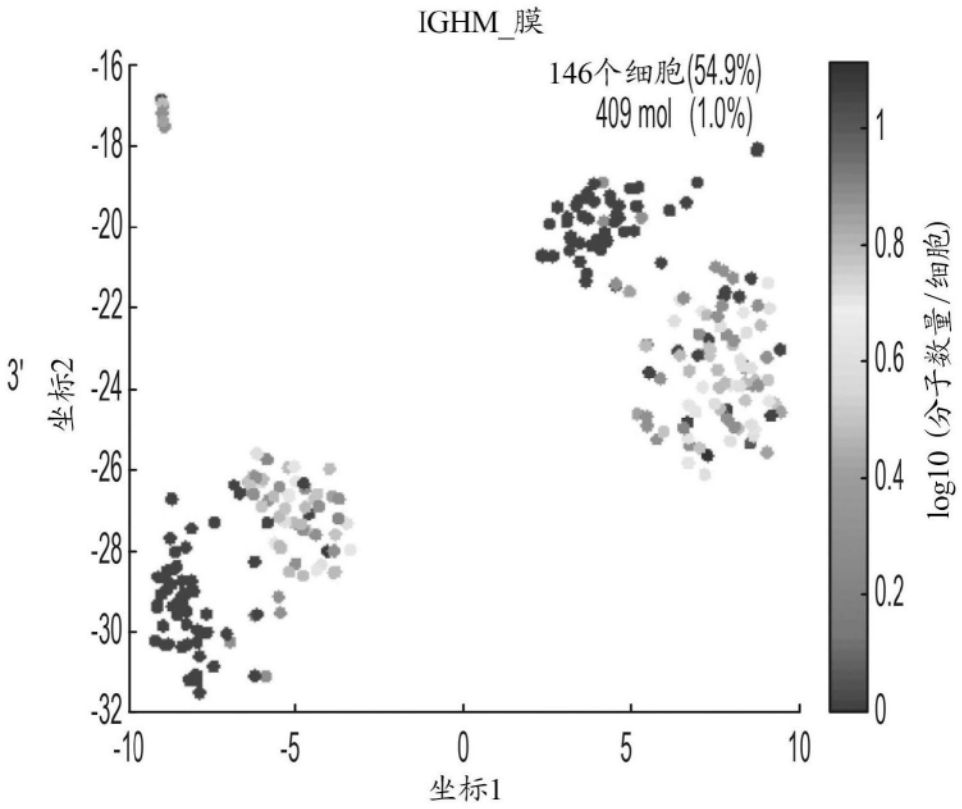

IGHM_膜
146个细胞(54.9%)
409 mol (1.0%)
3'
坐标2
坐标1
log10 (分子数量/细胞)

图11A

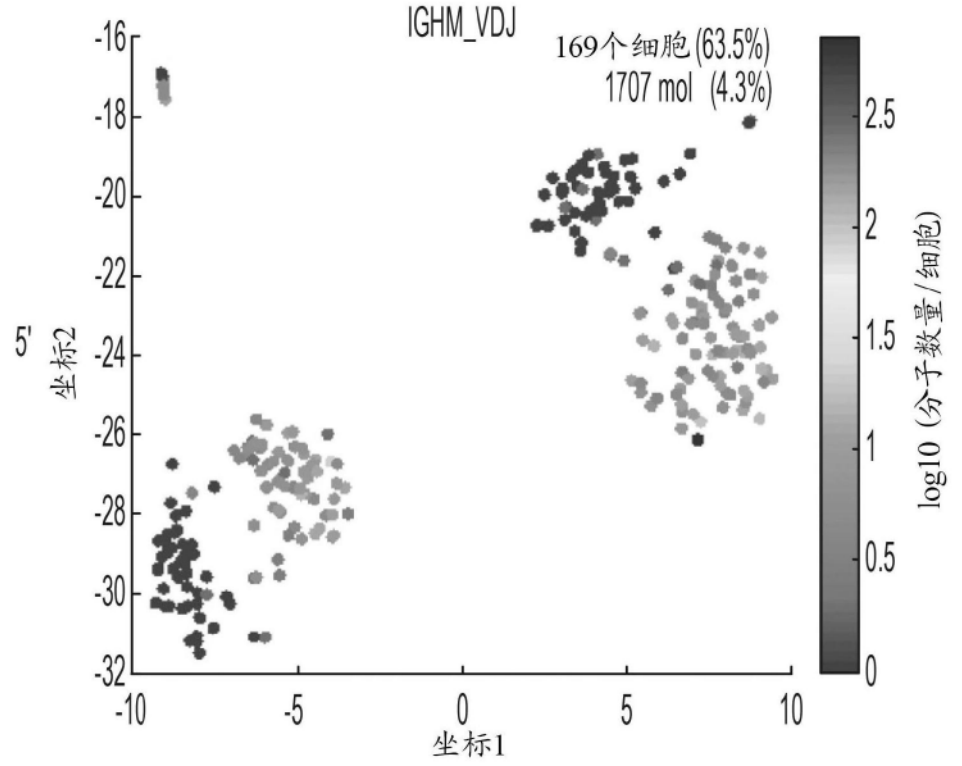

IGHM_VDJ
169个细胞(63.5%)
1707 mol (4.3%)
5'
坐标2
坐标1
log10 (分子数量/细胞)

图11B

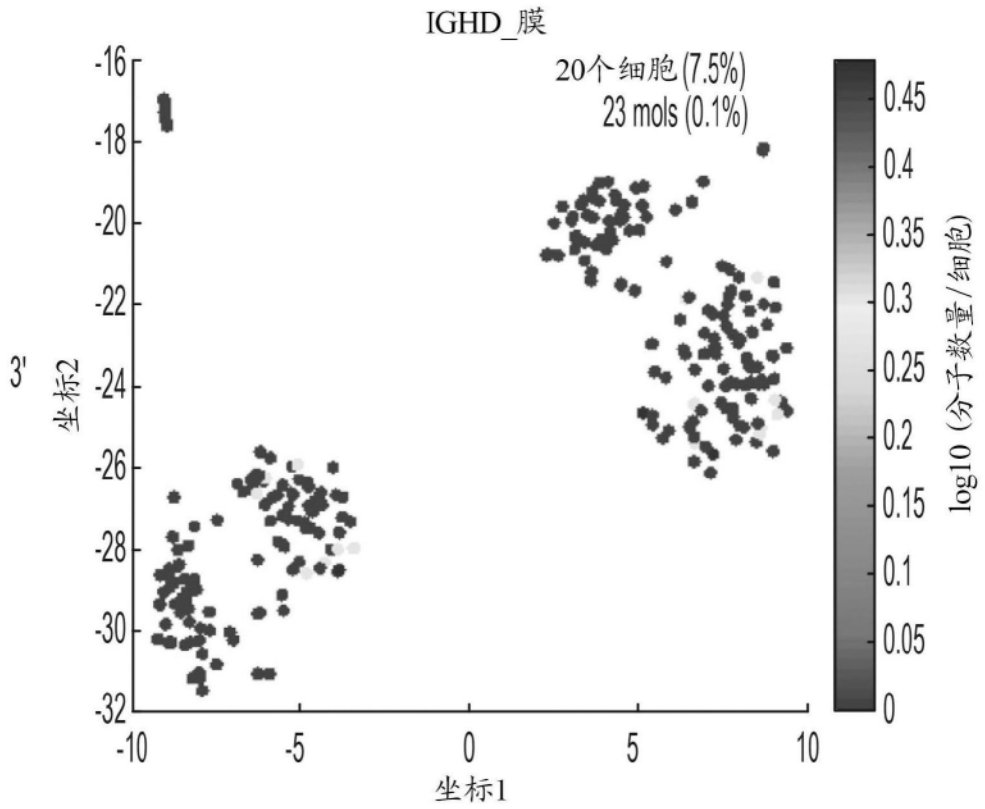

IGHD_膜
20个细胞(7.5%)
23 mols (0.1%)
3'
坐标2
坐标1
log10 (分子数量/细胞)
-16
-18
-20
-22
-24
-26
-28
-30
-32
-10
-5
0
5
10
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0

图11C

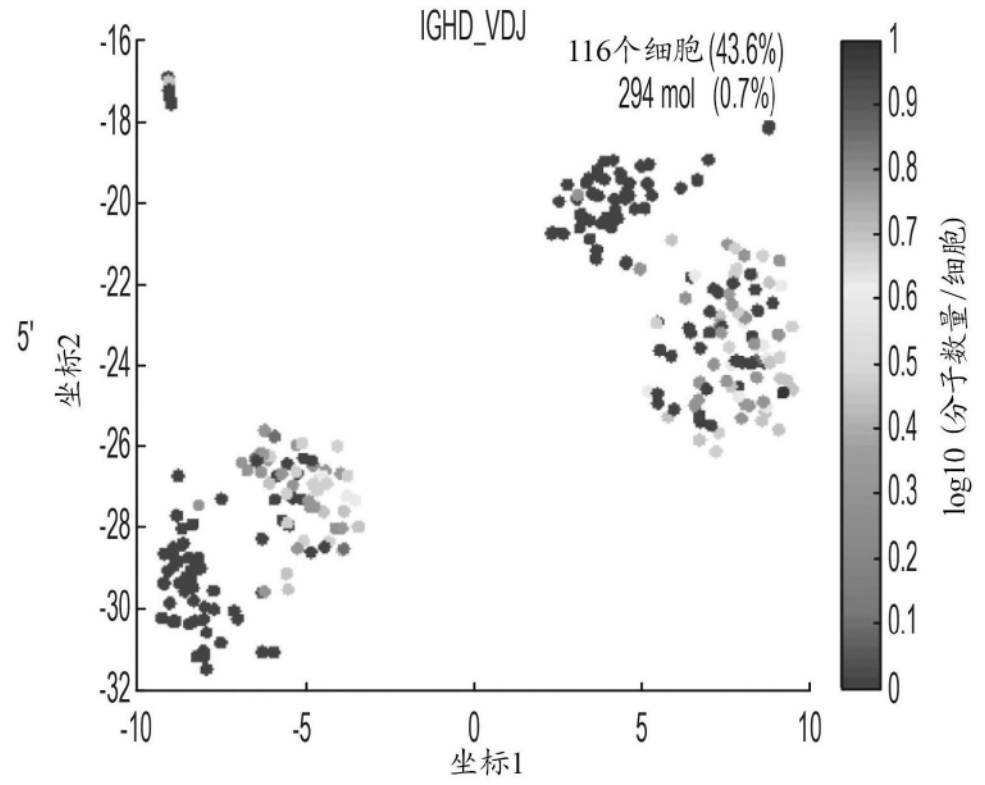

IGHD_VDJ
116个细胞(43.6%)
294 mol (0.7%)
5'
坐标2
坐标1
log10 (分子数量/细胞)
-16
-18
-20
-22
-24
-26
-28
-30
-32
-10
-5
0
5
10
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

图11D

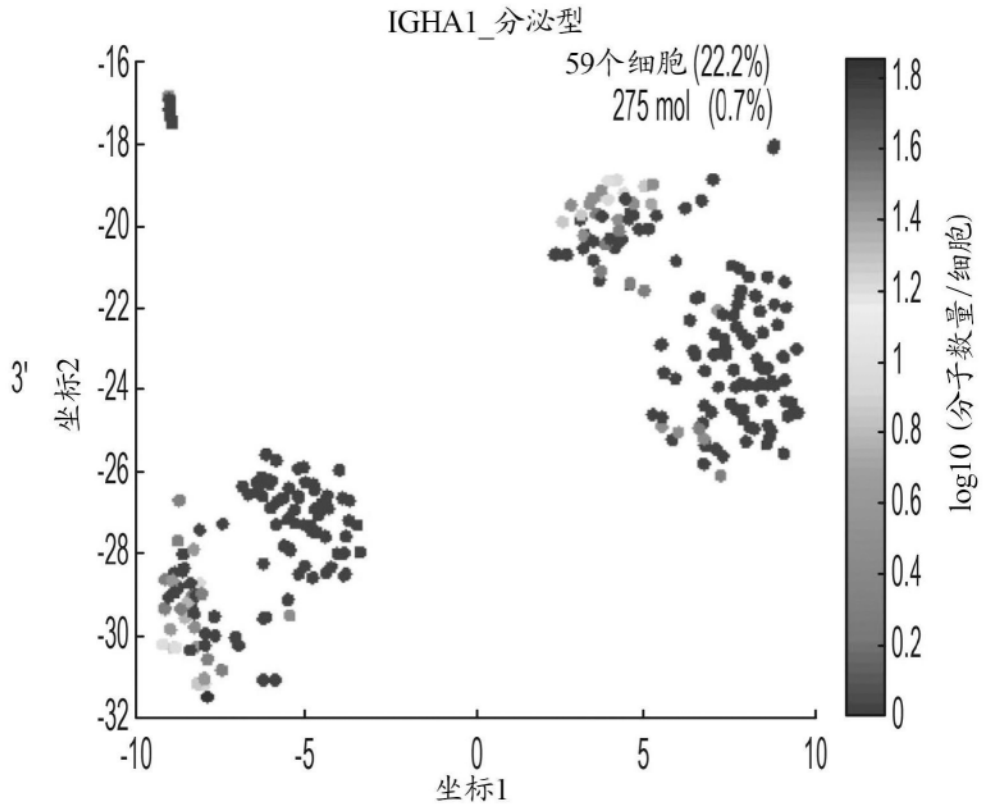

IGHA1_分泌型
59个细胞(22.2%)
275 mol (0.7%)
3'
坐标2
坐标1
log10 (分子数量/细胞)
-16
-18
-20
-22
-24
-26
-28
-30
-32
-10
-5
0
5
10
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0

图11E

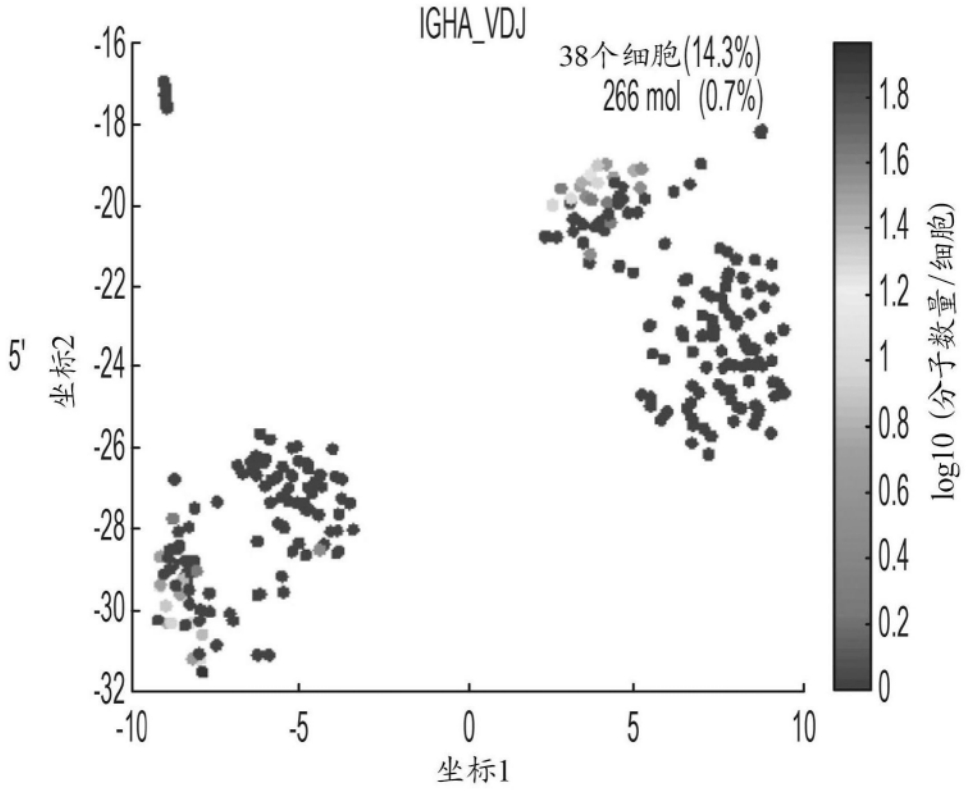

IGHA_VDJ
38个细胞(14.3%)
266 mol (0.7%)
5'
坐标2
坐标1
log10 (分子数量/细胞)

图11F

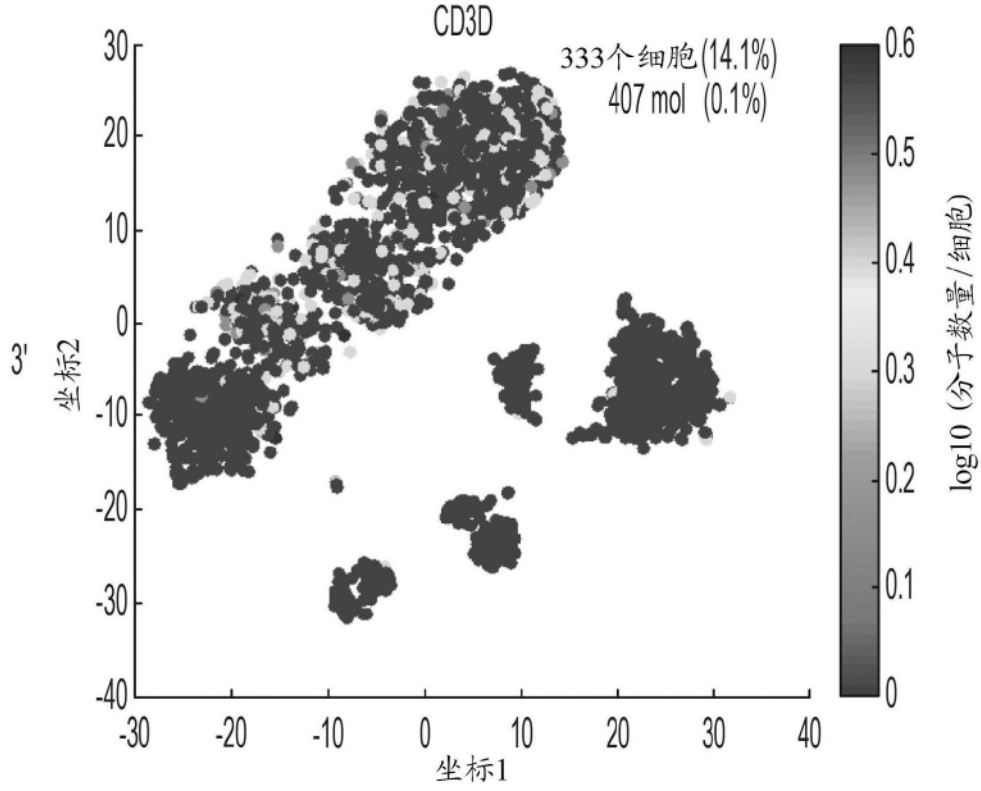

CD3D
333个细胞(14.1%)
407 mol (0.1%)
3'
坐标2
坐标1
log10 (分子数量/细胞)

图12A

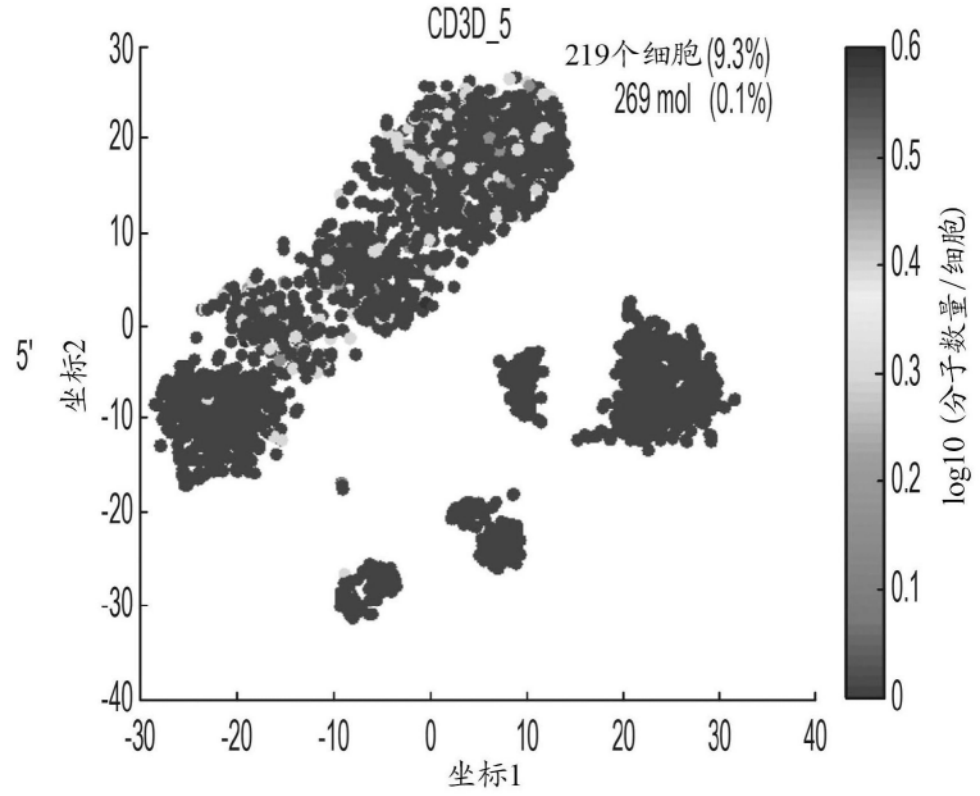

CD3D_5
219个细胞(9.3%)
269 mol (0.1%)
5'
坐标2
坐标1
log10 (分子数量/细胞)

图12B

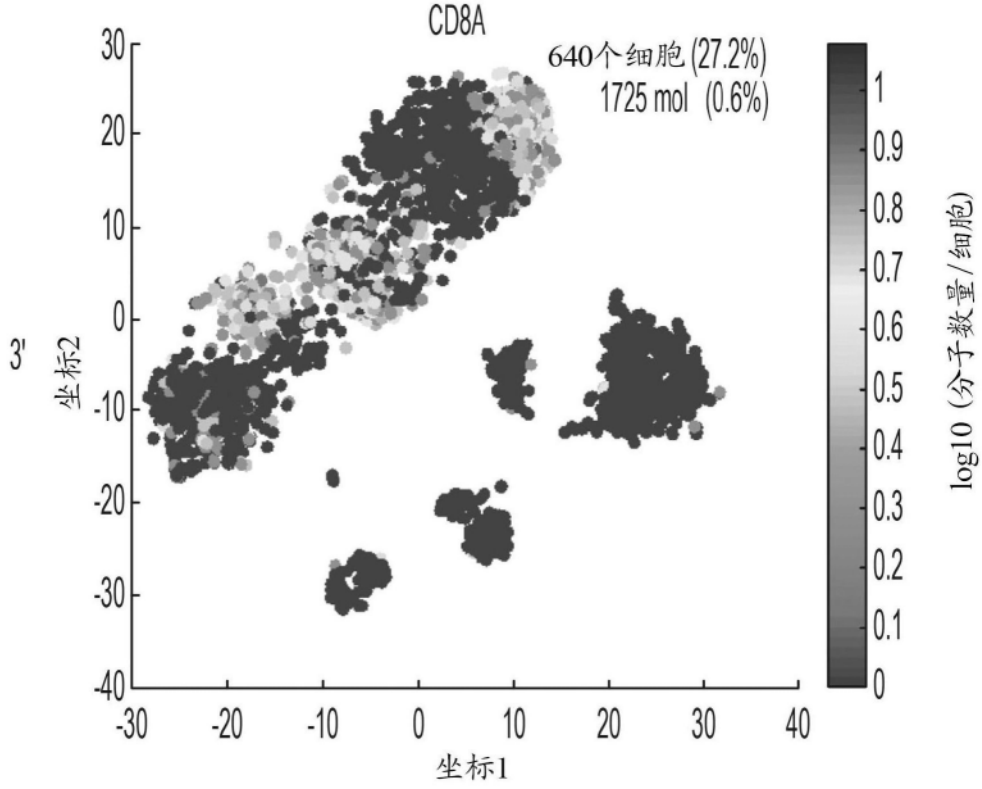

CD8A
640个细胞(27.2%)
1725 mol (0.6%)
3'
坐标2
坐标1
log10 (分子数量/细胞)

图12C

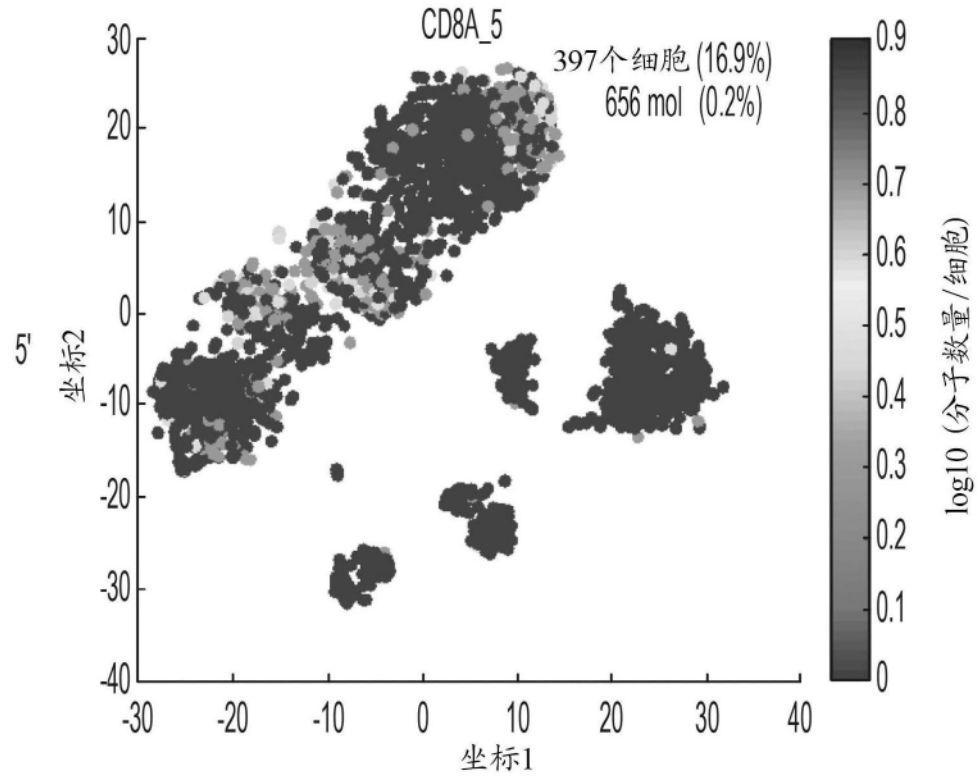

CD8A_5
397个细胞(16.9%)
656 mol (0.2%)
5'
坐标2
坐标1
log10 (分子数量/细胞)

图12D

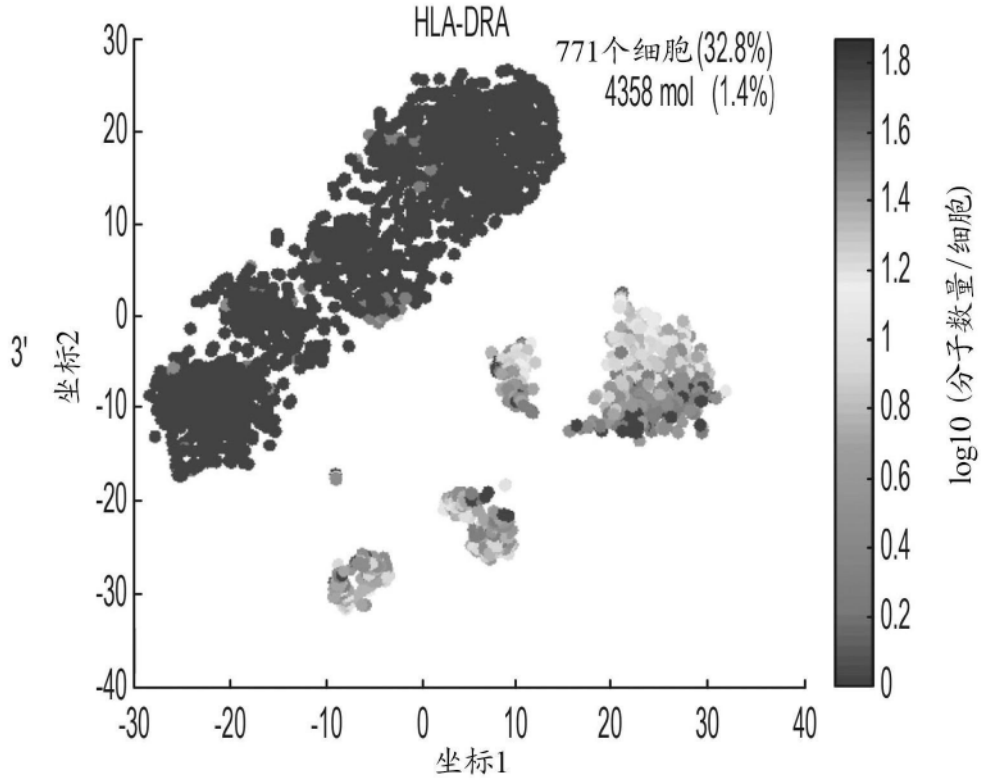

HLA-DRA
771个细胞(32.8%)
4358 mol (1.4%)
3'
坐标2
坐标1
log10 (分子数量/细胞)

图12E

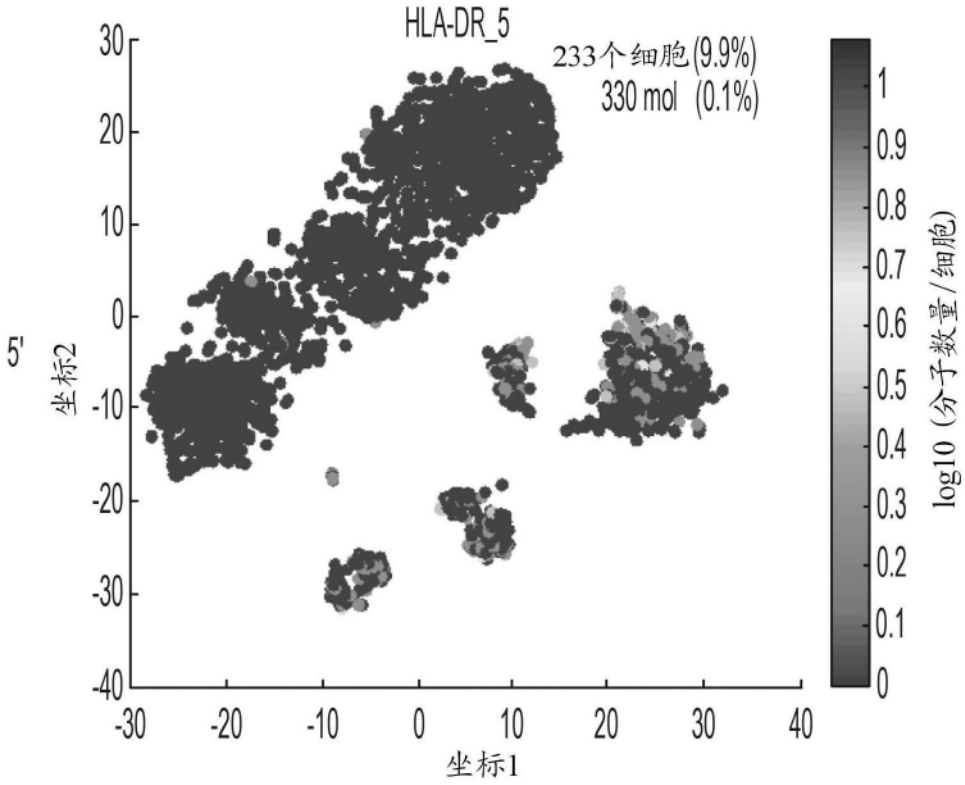

HLA-DR_5
233个细胞(9.9%)
330 mol (0.1%)
5'
坐标2
坐标1
log10 (分子数量/细胞)
30
20
10
0
-10
-20
-30
-40
-30
-20
-10
0
10
20
30
40
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

图12F