



(12) 发明专利申请

(10) 申请公布号 CN 114729350 A

(43) 申请公布日 2022. 07. 08

(21) 申请号 202080077712.7

(22) 申请日 2020.11.06

(30) 优先权数据

62/933,285 2019.11.08 US

(85) PCT国际申请进入国家阶段日

2022.05.07

(86) PCT国际申请的申请数据

PCT/US2020/059419 2020.11.06

(87) PCT国际申请的公布数据

W02021/092386 EN 2021.05.14

(71) 申请人 贝克顿迪金森公司

地址 美国新泽西州

(72) 发明人 克里斯蒂娜·张

玛格丽特·纳卡莫托

(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.

C12N 15/10 (2006.01)

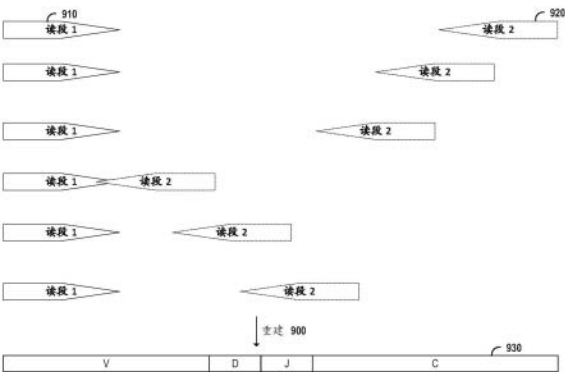
权利要求书9页 说明书61页 附图36页

(54) 发明名称

使用随机引发获得用于免疫组库测序的全长V(D)J信息

(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'末端被条形码化,并且随后在模板转换反应以及分子间和/或分子内杂交和延伸之后在5'末端被条形码化。5'条形码化的和/或3'条形码化的核酸靶可以用作扩增反应和/或随机引发和延伸反应的模板,以产生测序文库。该方法可以包括通过对齐多于一个测序读段来产生核酸靶的全长序列。在一些实施方案中还提供了免疫组库(immune repertoire)谱分析方法。



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种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与所述核酸靶杂交的靶结合区;

产生各自包含所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子;

使每种条形码化的核酸分子的所述靶结合区的互补体与以下中的一种或更多种的靶结合区杂交:

(i) 所述多于一种寡核苷酸条形码中的寡核苷酸条形码,

(ii) 所述条形码化的核酸分子自身,以及

(iii) 所述多于一种条形码化的核酸分子中的不同的条形码化的核酸分子;

延伸所述多于一种条形码化的核酸分子的3'末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子;

使用能够与所述核酸靶的序列杂交的靶特异性引物和包含所述第一通用序列的引物扩增所述多于一种延伸的条形码化的核酸分子,由此产生包含所述核酸靶的序列或其一部分的第一多于一种条形码化的扩增子;

使随机引物与所述第一多于一种条形码化的扩增子杂交并延伸所述随机引物以产生多于一种延伸产物,其中所述随机引物包含第二通用序列或其互补体;以及

使用能够与所述第一通用序列或其互补体和所述第二通用序列或其互补体杂交的引物扩增所述多于一种延伸产物,从而产生第二多于一种条形码化的扩增子。

2. 根据权利要求1所述的方法,所述方法包括:基于与所述第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来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种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与所述核酸靶杂交的靶结合区;

产生各自包含所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子;

使每种条形码化的核酸分子的所述靶结合区的互补体与以下中的一种或更多种的靶结合区杂交:

(i) 所述多于一种寡核苷酸条形码中的寡核苷酸条形码,

(ii) 所述条形码化的核酸分子自身,以及

(iii) 所述多于一种条形码化的核酸分子中的不同的条形码化的核酸分子;

延伸所述多于一种条形码化的核酸分子的3'末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子;

使用能够与所述核酸靶的序列杂交的靶特异性引物和包含所述第一通用序列的引物扩增所述多于一种延伸的条形码化的核酸分子,由此产生包含所述核酸靶的序列或其一部分的第一多于一种条形码化的扩增子;

使随机引物与所述第一多于一种条形码化的扩增子杂交并延伸所述随机引物以产生多于一种延伸产物,其中所述随机引物包含第二通用序列或其互补体;

使用能够与所述第一通用序列或其互补体和所述第二通用序列或其互补体杂交的引物扩增所述多于一种延伸产物,从而产生第二多于一种条形码化的扩增子;以及

基于与所述第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来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,其中确定所述样品中所述核酸靶的拷贝数包括:

(a) 基于与所述第二多于一种条形码化的扩增子或其产物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核酸靶的拷贝数,和/或

(b) 基于与所述第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记的数量来确定所述样品中所述核酸靶的拷贝数。

5. 根据权利要求1-4中任一项所述的方法,所述方法包括在使每种条形码化的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之前使所述多于一种条形码化的核酸分子变性:(i)所述多于一种寡核苷酸条形码中的寡核苷酸条形码,(ii)所述条形码化的核酸分子自身,和/或(iii)所述多于一种条形码化的核酸分子中的不同条形码化的核酸分子。

6. 根据权利要求1-5中任一项所述的方法,所述方法包括在扩增所述多于一种延伸的条形码化的核酸分子之前使所述多于一种延伸的条形码化的核酸分子变性。

7. 根据权利要求1-6中任一项所述的方法,其中所述多于一种核酸靶中的每一种的序列包括所述多于一种核酸靶中的每一种的子序列。

8. 根据权利要求1-7中任一项所述的方法,其中所述多于一种条形码化的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。

9. 根据权利要求1-8中任一项所述的方法,其中在延伸所述多于一种条形码化的核酸分子的3'末端之后,使所述第一分子标记与所述第二分子标记杂交。

10. 根据权利要求1-9中任一项所述的方法,其中所述延伸的条形码化的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。

11. 根据权利要求1-10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。

12. 根据权利要求1-11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、多(dT)序列或两者。

13. 根据权利要求1-12中任一项所述的方法,其中产生各自包含所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子包括:

(i) 使用所述多于一种寡核苷酸条形码对样品中的核酸靶的拷贝进行条形码化,以产生各自包含与所述核酸靶的至少一部分互补的序列、第一分子标记和所述靶结合区的多于一种条形码化的核酸分子;以及

(ii) 使包含所述靶结合区的互补体的寡核苷酸附接到所述多于一种条形码化的核酸分子以产生各自包含所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子。

14. 根据权利要求1-12中任一项所述的方法,其中产生各自包含所述靶结合区和所述

靶结合区的互补体的多于一种条形码化的核酸分子包括：

(i) 使用所述多于一种寡核苷酸条形码对样品中的核酸靶的拷贝进行条形码化以产生各自包含与所述核酸靶的至少一部分互补的序列、第一分子标记和所述靶结合区的多于一种条形码化的核酸分子；

(ii) 扩增所述多于一种条形码化的核酸分子以产生多于一种扩增的条形码化的核酸分子；以及

(iii) 使包含所述靶结合区的互补体的寡核苷酸附接到多于一种扩增的条形码化的核酸分子以产生各自包含所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子。

15. 根据权利要求13-14中任一项所述的方法，其中对样品中的核酸靶的拷贝进行条形码化包括延伸与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码以产生所述多于一种条形码化的核酸分子。

16. 根据权利要求15所述的方法，其中延伸与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码包括对与所述多于一种寡核苷酸条形码杂交的所述核酸靶的拷贝进行逆转录。

17. 根据权利要求13-16中任一项所述的方法，其中附接包含所述靶结合区的互补体的所述寡核苷酸包括将包含所述靶结合区的互补体的所述寡核苷酸连接到所述多于一种条形码化的核酸分子和/或扩增的条形码化的核酸分子。

18. 根据权利要求13-16中任一项所述的方法，其中所述靶结合区包含多(dT)序列，并且其中附接包含所述靶结合区的互补体的所述寡核苷酸包括使用末端脱氧核苷酸转移酶将多于一种腺苷单磷酸添加到所述多于一种条形码化的核酸分子和/或扩增的条形码化的核酸分子。

19. 根据权利要求1-12中任一项所述的方法，其中产生各自包含所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子包括：

在存在逆转录酶和包含所述靶结合区或其部分的模板转换寡核苷酸的情况下，延伸与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码，以产生各自包含与所述核酸靶的至少一部分互补的序列、第一分子标记、所述靶结合区和所述靶结合区的互补体的多于一种条形码化的核酸分子。

20. 根据权利要求19所述的方法，其中所述逆转录酶能够具有末端转移酶活性。

21. 根据权利要求19-20中任一项所述的方法，其中所述模板转换寡核苷酸包含一个或更多个3'核糖核苷酸，并且任选地三个3'核糖核苷酸。

22. 根据权利要求21所述的方法，其中所述3'核糖核苷酸包括鸟嘌呤。

23. 根据权利要求19-22中任一项所述的方法，其中所述逆转录酶包括病毒逆转录酶，任选地鼠白血病病毒(MLV)逆转录酶或Moloney鼠白血病病毒(MMLV)逆转录酶。

24. 根据权利要求1-23中任一项所述的方法，其中使条形码化的核酸分子的所述靶结合区的互补体与所述条形码化的核酸分子自身的所述靶结合区杂交包括使条形码化的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。

25. 根据权利要求24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。

26. 根据权利要求1-25中任一项所述的方法,其中使条形码化的核酸分子的所述靶结合区的互补体与所述多于一种寡核苷酸条形码中的寡核苷酸条形码的所述靶结合区杂交包括使条形码化的核酸分子的所述靶结合区的互补体与所述多于一种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。

27. 根据权利要求26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。

28. 根据权利要求26-27中任一项所述的方法,所述方法包括延伸与所述条形码化的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'末端以产生各自包含所述第一分子标记的互补体和第二分子标记的多于一种延伸的条形码化的核酸分子。

29. 根据权利要求28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。

30. 根据权利要求1-29中任一项所述的方法,其中使条形码化的核酸分子的所述靶结合区的互补体与所述多于一种条形码化的核酸分子中的不同条形码化的核酸分子的所述靶结合区杂交包括使条形码化的核酸分子的所述靶结合区的互补体与所述多于一种条形码化的核酸分子中的不同条形码化的核酸分子的所述靶结合区进行分子间杂交。

31. 根据权利要求30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。

32. 根据权利要求1-31中任一项所述的方法,其中所述样品包括单细胞、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。

33. 根据权利要求32所述的方法,其中单细胞包括免疫细胞或循环肿瘤细胞,任选地所述免疫细胞是B细胞或T细胞。

34. 根据权利要求1-33中任一项所述的方法,其中所述靶特异性引物与免疫受体特异性杂交,任选地所述靶特异性引物与免疫受体的恒定区、免疫受体的可变区、免疫受体的多样性区、免疫受体的可变区和多样性区的连接区或它们的组合特异性杂交。

35. 根据权利要求34所述的方法,其中所述免疫受体是T细胞受体(TCR)和/或B细胞受体(BCR);任选地所述TCR包含TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或它们的任何组合;并且还任选地所述BCR受体包含BCR重链和/或BCR轻链。

36. 根据权利要求1-35中任一项所述的方法,其中延伸所述多于一种条形码化的核酸分子的3'末端包括使用缺乏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聚合酶延伸所述多于一种条形码化的核酸分子的3'末端,并且任选地所述DNA聚合酶包括Klenow片段。

37. 根据权利要求1-36中任一项所述的方法,所述方法包括获得所述第二多于一种条形码化的扩增子或其产物的序列信息,并且任选地获得所述序列信息包括将测序衔接子附接到所述第二多于一种条形码化的扩增子或其产物。

38. 根据权利要求1-37中任一项所述的方法,其中扩增所述多于一种延伸产物包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到所述多于一种延伸产物;任选地所述测序衔接子包括P5序列、P7序列、其互补序列或其一部分;并且还任选地所述测序引物包括读段1测序引物、读段2测序引物、其互补序列或其一部分。

39. 根据权利要求37-38中任一项所述的方法,其中获得所述第二多于一种条形码化的

扩增子或其产物的序列信息包括：

获得包含所述第二多于一种条形码化的扩增子或其产物的多于一个测序读段的测序数据，

其中所述多于一个测序读段中的每一个包含 (1) 细胞标记序列，(2) 分子标记序列，和/或 (3) 所述核酸靶的子序列。

40. 根据权利要求39所述的方法，所述方法包括：

对于指示所述样品中的单细胞的每个独特细胞标记序列：

使所述核酸靶的所述多于一个测序读段中的每一个对齐以产生所述核酸靶的对齐序列。

41. 根据权利要求40所述的方法，其中所述核酸靶的对齐序列包含所述核酸靶的cDNA序列的至少50%、所述核酸靶的cDNA序列的至少70%、所述核酸靶的cDNA序列的至少90%或所述核酸靶的cDNA序列的全长。

42. 根据权利要求1-41中任一项所述的方法，其中所述核酸靶是免疫受体，任选地所述免疫受体包含BCR轻链、BCR重链、TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或它们的任何组合。

43. 根据权利要求42所述的方法，其中所述核酸靶的所述对齐序列包括互补决定区1 (CDR1)、互补决定区2 (CDR2)、互补决定区3 (CDR3)、可变区、可变区的全长或它们的组合。

44. 根据权利要求42-43中任一项所述的方法，其中所述核酸靶的所述对齐序列包含可变区、多样性区、可变区多样性区的连接区和/或恒定区或它们的任何组合。

45. 根据权利要求42-44中任一项所述的方法，其中所述方法不包括片段化、标签片段化或两者。

46. 根据权利要求37-45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，并且任选地所述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。

47. 根据权利要求46所述的方法，所述方法包括基于所获得的序列信息使所述单细胞的BCR轻链和BCR重链进行配对。

48. 根据权利要求46-47中任一项所述的方法，其中所述样品包括多于一个单细胞，所述方法包括基于所获得的序列信息使至少50%的所述单细胞的BCR轻链和BCR重链进行配对。

49. 根据权利要求37-48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TCR α 链和TCR β 链的序列信息，并且任选地所述TCR α 链和TCR β 链的序列信息包括TCR α 链和/或TCR β 链的互补决定区1 (CDR1)、CDR2、CDR3或它们的任何组合的序列。

50. 根据权利要求49所述的方法，所述方法包括基于所获得的序列信息使所述单细胞的TCR α 链和TCR β 链进行配对。

51. 根据权利要求49-50中任一项所述的方法，其中所述样品包括多于一个单细胞，所述方法包括基于所获得的序列信息使至少50%的所述单细胞的TCR α 链和TCR β 链进行配对。

52. 根据权利要求37-51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TCR γ 链和TCR δ 链的序列信息。

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

54. 根据权利要求52-53中任一项所述的方法,所述方法包括基于所获得的序列信息使所述单细胞的TCR γ 链和TCR δ 链进行配对。

55. 根据权利要求52-54中任一项所述的方法,其中所述样品包括多于一个单细胞,所述方法包括基于所获得的序列信息使至少50%的所述单细胞的TCR γ 链和TCR δ 链进行配对。

56. 根据权利要求1-55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或所述靶结合区的互补序列。

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

58. 根据权利要求1-57中任一项所述的方法,其中所述多于一种条形码化的核酸分子包括条形码化的脱氧核糖核酸(DNA)分子、条形码化的核糖核酸(RNA)分子或它们的组合。

59. 根据权利要求1-58中任一项所述的方法,其中所述核酸靶包括核酸分子,任选地所述核酸分子包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、含有多(A)尾的RNA或它们的任何组合,并且还任选地所述mRNA编码免疫受体。

60. 根据权利要求59所述的方法,其中所述核酸靶包括细胞组分结合试剂,和/或所述核酸分子与所述细胞组分结合试剂关联。

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,其中所述多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列。

63. 根据权利要求1-62中任一项所述的方法,其中所述多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。

64. 根据权利要求1-63中任一项所述的方法,其中所述多于一种寡核苷酸条形码与固体支持物关联。

65. 根据权利要求64所述的方法,其中与相同固体支持物关联的所述多于一种寡核苷酸条形码各自包含相同的样品标记。

66. 根据权利要求65所述的方法,其中所述多于一种寡核苷酸条形码的每种样品标记包含至少6个核苷酸。

67. 根据权利要求1-66中任一项所述的方法,其中所述多于一种寡核苷酸条形码各自包含细胞标记。

68. 根据权利要求67所述的方法,其中所述多于一种寡核苷酸条形码的每种细胞标记包含至少6个核苷酸。

69. 根据权利要求67-68中任一项所述的方法,其中与相同固体支持物关联的寡核苷酸条形码包含相同的细胞标记。

70. 根据权利要求67-69中任一项所述的方法,其中与不同的固体支持物关联的寡核苷酸条形码包含不同的细胞标记。

71. 根据权利要求67-70中任一项所述的方法,其中所述多于一种延伸的条形码化的核酸分子各自包含细胞标记和所述细胞标记的互补体。

72. 根据权利要求71所述的方法,其中所述细胞标记的互补体包括所述细胞标记的反

向互补序列或所述细胞标记的互补序列。

73. 根据权利要求1-72中任一项所述的方法, 所述方法包括在存在乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砷(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或它们的任何组合中的一种或更多种的情况下, 延伸与所述核酸靶的拷贝杂交的所述多于一种寡核苷酸条形码。

74. 根据权利要求64-73中任一项所述的方法, 其中所述固体支持物包括合成颗粒、平坦表面或它们的组合。

75. 根据权利要求1-74中任一项所述的方法, 其中所述样品包括单细胞, 所述方法包括使包含所述多于一种寡核苷酸条形码的合成颗粒与所述样品中的所述单细胞关联。

76. 根据权利要求75所述的方法, 所述方法包括在使所述合成颗粒与所述单细胞关联后裂解所述单细胞, 任选地裂解所述单细胞包括加热所述样品、使所述样品与去污剂接触、改变所述样品的pH或它们的任何组合。

77. 根据权利要求75-76中任一项所述的方法, 其中所述合成颗粒和所述单细胞在相同的分区中, 并且任选地所述分区是孔或液滴。

78. 根据权利要求75-77中任一项所述的方法, 其中所述多于一种寡核苷酸条形码中的至少一种被固定或部分地固定在所述合成颗粒上, 或者所述多于一种寡核苷酸条形码中的至少一种被包封或部分地包封在所述合成颗粒中。

79. 根据权利要求75-78中任一项所述的方法, 其中所述合成颗粒是可破坏的, 任选地所述合成颗粒是可破坏的水凝胶颗粒。

80. 根据权利要求75-79中任一项所述的方法, 其中所述合成颗粒包括珠, 任选地所述珠包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚(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。

81. 根据权利要求75-80中任一项所述的方法, 其中所述合成颗粒包含选自以下组成的组的材料: 聚二甲基硅氧烷(PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮及它们的任何组合。

82. 根据权利要求75-81中任一项所述的方法,

其中所述多于一种寡核苷酸条形码中的每一种包含接头官能团,

其中所述合成颗粒包含固体支持物官能团, 并且

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

83. 一种试剂盒, 所述试剂盒包含:

多于一种寡核苷酸条形码, 其中所述多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区, 并且其中所述多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;

逆转录酶;

包含所述靶结合区或其一部分的模板转换寡核苷酸; 以及

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。

84. 根据权利要求83所述的试剂盒,其中所述逆转录酶包括病毒逆转录酶,并且任选地所述病毒逆转录酶是鼠白血病病毒 (MLV) 逆转录酶或Moloney鼠白血病病毒 (MMLV) 逆转录酶。

85. 根据权利要求83-84中任一项所述的试剂盒,其中所述模板转换寡核苷酸包含一个或更多个3'核糖核苷酸,任选地所述模板转换寡核苷酸包含三个3'核糖核苷酸。

86. 根据权利要求85所述的试剂盒,其中所述3'核糖核苷酸包括鸟嘌呤。

87. 根据权利要求83-86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。

88. 一种试剂盒,所述试剂盒包含:

多于一种寡核苷酸条形码,其中所述多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中所述多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;

末端脱氧核苷酸转移酶;和

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。

89. 根据权利要求83-88中任一项所述的试剂盒,其中所述DNA聚合酶包括Klenow片段。

90. 根据权利要求83-89中任一项所述的试剂盒,所述试剂盒包含缓冲液、筒或两者。

91. 根据权利要求83-90中任一项所述的试剂盒,所述试剂盒包含一种或更多种用于逆转录反应和/或扩增反应的试剂。

92. 根据权利要求83-91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。

93. 根据权利要求83-92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。

94. 根据权利要求93所述的试剂盒,其中所述多于一种寡核苷酸条形码的每种样品标记和/或细胞标记包含至少6个核苷酸。

95. 根据权利要求83-93中任一项所述的试剂盒,其中所述多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。

96. 根据权利要求83-95中任一项所述的试剂盒,其中所述多于一种寡核苷酸条形码中的至少一种被固定或部分地固定在所述合成颗粒上,或者被包封或部分地包封在所述合成颗粒中。

97. 根据权利要求83-96中任一项所述的试剂盒,其中所述合成颗粒是可破坏的,任选地所述合成颗粒是可破坏的水凝胶颗粒。

98. 根据权利要求83-97中任一项所述的试剂盒,其中所述合成颗粒包括珠,任选地所述珠包括琼脂糖凝胶珠、链霉抗生物素蛋白珠、琼脂糖珠、磁珠、缀合珠、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、寡聚 (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。

99. 根据权利要求83-98中任一项所述的试剂盒,其中所述合成颗粒包含选自以下组成的组的材料:聚二甲基硅氧烷 (PDMS)、聚苯乙烯、玻璃、聚丙烯、琼脂糖、明胶、水凝胶、顺

磁物质、陶瓷、塑料、玻璃、甲基苯乙烯、丙烯酸聚合物、钛、乳胶、琼脂糖凝胶、纤维素、尼龙、硅酮及它们的任何组合。

100. 根据权利要求83-99中任一项所述的试剂盒，

其中所述多于一种寡核苷酸条形码中的每一种包含接头官能团，

其中所述合成颗粒包含固体支持物官能团，以及

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

使用随机引发获得用于免疫组库测序的全长V(D)J信息

[0001] 相关申请

[0002] 本申请根据35 U.S.C. §119 (e) 要求2019年11月8日提交的美国临时专利申请序列第62/933285号的权益,为了所有目的此相关申请的内容通过引用以其整体并入本文。

[0003] 背景

[0004] 领域

[0005] 本公开内容总体上涉及分子生物学领域,并且特别涉及使用分子条形码化的多组学分析。

[0006] 对相关技术的描述

[0007] 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)破译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。需要对核酸靶(例如,含V(D)J的转录物)进行全长表达谱分析的方法和技术。需要能够对核酸靶(例如,含V(D)J的转录物)进行鉴定和计数两者的组合物和方法。

[0008] 概述

[0009] 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,方法包括:使核酸靶的拷贝与多于一种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记(例如,第一分子标记)和能够与核酸靶杂交的靶结合区;产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子;使每种条形码化的核酸分子的靶结合区的互补体与以下中的一种或更多种的靶结合区杂交:(i)多于一种寡核苷酸条形码中的寡核苷酸条形码,(ii)条形码化的核酸分子自身,以及(iii)多于一种条形码化的核酸分子中的不同条形码化的核酸分子;延伸多于一种条形码化的核酸分子的3'末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子;使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子,由此产生包含核酸靶的序列或其一部分的第一多于一种条形码化的扩增子;使随机引物与第一多于一种条形码化的扩增子杂交并延伸随机引物以产生多于一种延伸产物,其中随机引物包含第二通用序列或其互补体;以及使用能够与第一通用序列或其互补体和第二通用序列或其互补体杂交的引物来扩增多于一种延伸产物,从而产生第二多于一种条形码化的扩增子。在一些实施方案中,方法包括:基于与第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来确定样品中核酸靶的拷贝数。在一些实施方案中,方法不包括片段化、标签片段化或两者。

[0010] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中,方法包括:使核酸靶的拷贝与多于一种寡核苷酸条形码接触,其中每种寡核苷酸条形码包含第一通用序列、分子标记(例如,第一分子标记)和能够与核酸靶杂交的靶结合区;产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子;使每种条形码化的核酸分子的靶结合区的互补体与以下中的一种或更多种的靶结合区杂交:(i)多于一种寡核苷酸

酸条形码中的寡核苷酸条形码, (ii) 条形码化的核酸分子自身, 以及 (iii) 多于一种条形码化的核酸分子中的不同条形码化的核酸分子; 延伸多于一种条形码化的核酸分子的3' 末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子; 使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子, 由此产生包含核酸靶的序列或其一部分的第一多于一种条形码化的扩增子; 使随机引物与第一多于一种条形码化的扩增子杂交并延伸随机引物以产生多于一种延伸产物, 其中随机引物包含第二通用序列或其互补体; 使用能够与第一通用序列和第二通用序列或其互补体杂交的引物扩增多于一种延伸产物, 从而产生第二多于一种条形码化的扩增子; 以及基于与第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来确定样品中核酸靶的拷贝数。

[0011] 在一些实施方案中, 确定样品中核酸靶的拷贝数包括: (a) 基于与第二多于一种条形码化的扩增子或其产物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核酸靶的拷贝数, 和/或 (b) 基于与第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记的数量来确定样品中核酸靶的拷贝数。在一些实施方案中, 方法包括: 在使每种条形码化的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种条形码化的核酸分子变性: (i) 多于一种寡核苷酸条形码中的寡核苷酸条形码, (ii) 条形码化的核酸分子自身, 和/或 (iii) 多于一种条形码化的核酸分子中的不同条形码化的核酸分子。

[0012] 在一些实施方案中, 方法包括: 在扩增多于一种延伸的条形码化的核酸分子之前使多于一种延伸的条形码化的核酸分子变性。在一些实施方案中, 多于一种核酸靶中的每一种的序列包括多于一种核酸靶中的每一种的子序列。在一些实施方案中, 多于一种条形码化的核酸分子中的核酸靶的序列包括核酸靶的子序列。在一些实施方案中, 在延伸多于一种条形码化的核酸分子的3' 末端之后, 使第一分子标记与第二分子标记杂交。在一些实施方案中, 延伸的条形码化的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, 靶结合区包含基因特异性序列、多 (dT) 序列或两者。

[0013] 在一些实施方案中, 产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子包括: (i) 使用多于一种寡核苷酸条形码对样品中的核酸靶的拷贝进行条形码化以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记和靶结合区的多于一种条形码化的核酸分子; 以及 (ii) 使包含靶结合区的互补体的寡核苷酸附接到多于一种条形码化的核酸分子以产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子。

[0014] 在一些实施方案中, 产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子包括: (i) 使用多于一种寡核苷酸条形码对样品中的核酸靶的拷贝进行条形码化以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记和靶结合区的多于一种条形码化的核酸分子; (ii) 扩增多于一种条形码化的核酸分子以产生多于一种扩增的条形码化的核酸分子; 以及 (iii) 使包含靶结合区的互补体的寡核苷酸附接到多于一种扩增的条形码化的核酸分子以产生各自包含靶结合区和靶结合区的互补体的多于一种条

形码化的核酸分子。

[0015] 在一些实施方案中,对样品中的核酸靶的拷贝进行条形码化包括延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码以产生多于一种条形码化的核酸分子。在一些实施方案中,延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码包括对与多于一种寡核苷酸条形码杂交的核酸靶的拷贝进行逆转录。在一些实施方案中,附接包含靶结合区的互补体的寡核苷酸包括将包含靶结合区的互补体的寡核苷酸连接到多于一种条形码化的核酸分子和/或扩增的条形码化的核酸分子。在一些实施方案中,靶结合区包含多(dT)序列,并且其中附接包含靶结合区的互补体的寡核苷酸包括使用末端脱氧核苷酸转移酶将多于一种腺苷单磷酸添加到多于一种条形码化的核酸分子和/或扩增的条形码化的核酸分子。

[0016] 在一些实施方案中,产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子包括:在存在逆转录酶和包含靶结合区或其一部分的模板转换寡核苷酸的情况下,延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码,以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记、靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子。在一些实施方案中,逆转录酶能够具有末端转移酶活性。在一些实施方案中,模板转换寡核苷酸包含一个或更多个3'核糖核苷酸,例如三个3'核糖核苷酸。在一些实施方案中,3'核糖核苷酸包括鸟嘌呤。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Moloney鼠白血病病毒(MMLV)逆转录酶。在一些实施方案中,方法包括在存在乙二醇、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情况下,延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码。

[0017] 在一些实施方案中,使条形码化的核酸分子的靶结合区的互补体与条形码化的核酸分子自身的靶结合区杂交包括使条形码化的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。

[0018] 在一些实施方案中,使条形码化的核酸分子的靶结合区的互补体与多于一种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的核酸分子的靶结合区的互补体与多于一种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,方法包括延伸与条形码化的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'末端以产生各自包含第一分子标记的互补体和第二分子标记的多于一种延伸的条形码化的核酸分子。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。

[0019] 在一些实施方案中,使条形码化的核酸分子的靶结合区的互补体与多于一种条形码化的核酸分子中的不同条形码化的核酸分子的靶结合区杂交包括使条形码化的核酸分子的靶结合区的互补体与多于一种条形码化的核酸分子中的不同条形码化的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。

[0020] 在一些实施方案中,样品包括单细胞、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或循环肿瘤细胞。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,靶特异性引物与免疫受体特异性杂交。在一些实施方案中,靶特异性引物与免疫受体的恒定区、免疫受体的可变区、免疫受体的多样性区、免疫受体的可变区和多样性区的连接区(junction)或它们的组合特异性杂交。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包含TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或它们的任何组合。在一些实施方案中,BCR受体包含BCR重链和/或BCR轻链。

[0021] 在一些实施方案中,延伸多于一种条形码化的核酸分子的3'末端包括使用缺乏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聚合酶延伸多于一种条形码化的核酸分子的3'末端。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,方法包括获得第二多于一种条形码化的扩增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第二多于一种条形码化的扩增子或其产物。在一些实施方案中,扩增多于一种延伸产物包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到多于一种延伸产物。在一些实施方案中,测序衔接子包括P5序列、P7序列、其互补序列或其一部分。在一些实施方案中,测序引物包括读段1测序引物、读段2测序引物、其互补序列或其一部分。在一些实施方案中,获得第二多于一种条形码化的扩增子或其产物的序列信息包括:获得包含第二多于一种条形码化的扩增子或其产物的多于一个测序读段的测序数据,其中所述多于一个测序读段中的每一个包含(1)细胞标记序列、(2)分子标记序列和/或(3)核酸靶的子序列。

[0022] 在一些实施方案中,方法包括:对于指示样品中的单细胞的每个独特的细胞标记序列:使核酸靶的多于一个测序读段中的每一个对齐以产生核酸靶的对齐序列。在一些实施方案中,核酸靶的对齐序列包括核酸靶的cDNA序列的至少50%、核酸靶的cDNA序列的至少70%、核酸靶的cDNA序列的至少90%或核酸靶的cDNA序列的全长。在一些实施方案中,核酸靶是免疫受体。在一些实施方案中,免疫受体包含BCR轻链、BCR重链、TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或它们的任何组合。在一些实施方案中,核酸靶的对齐序列包含互补决定区1(CDR1)、互补决定区2(CDR2)、互补决定区3(CDR3)、可变区、可变区的全长或它们的组合。在一些实施方案中,核酸靶的对齐序列包含可变区、多样性区、可变区多样性区的连接区和/或恒定区,或它们的任何组合。

[0023] 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,方法包括:基于所获得的序列信息使单细胞的BCR轻链和BCR重链进行配对。在一些实施方案中,样品包括多于一个单细胞,方法包括基于所获得的序列信息使至少50%的所述单细胞的BCR轻链和BCR重链进行配对。

[0024] 在一些实施方案中,获得序列信息包括获得单细胞的TCR α 链和TCR β 链的序列信息。在一些实施方案中,TCR α 链和TCR β 链的序列信息包括TCR α 链和/或TCR β 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,方法包括:基于所获得的序列信息使单细胞的TCR α 链和TCR β 链进行配对。在一些实施方案中,样品包括多于一个单细胞,方法包括基于所获得的序列信息使至少50%的所述单细胞的TCR α 链和TCR β 链进行配对。

[0025] 在一些实施方案中,获得序列信息包括获得单细胞的TCR γ 链和TCR δ 链的序列信息。在一些实施方案中,TCR γ 链和TCR δ 链的序列信息包括TCR γ 链和/或TCR δ 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,方法包括:基于所获得的序列信息使单细胞的TCR γ 链和TCR δ 链进行配对。在一些实施方案中,样品包括多于一个单细胞,方法包括基于所获得的序列信息使至少50%的所述单细胞的TCR γ 链和TCR δ 链进行配对。

[0026] 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或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或分子标记的互补序列。在一些实施方案中,多于一种条形码化的核酸分子包括条形码化的脱氧核糖核酸(DNA)分子、条形码化的核糖核酸(RNA)分子或它们的组合。在一些实施方案中,核酸靶包括核酸分子。在一些实施方案中,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。在一些实施方案中,mRNA编码免疫受体。在一些实施方案中,核酸靶包括细胞组分结合试剂,和/或核酸分子与细胞组分结合试剂关联。在一些实施方案中,方法包括使核酸分子和细胞组分结合试剂解离。

[0027] 在一些实施方案中,多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列。在一些实施方案中,多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。在一些实施方案中,多于一种寡核苷酸条形码与固体支持物关联。在一些实施方案中,与相同固体支持物关联的多于一种寡核苷酸条形码各自包含相同的样品标记。在一些实施方案中,多于一种寡核苷酸条形码的每种样品标记包含至少6个核苷酸。在一些实施方案中,多于一种寡核苷酸条形码中的每一种都包括细胞标记。在一些实施方案中,多于一种寡核苷酸条形码的每种细胞标记包含至少6个核苷酸。在一些实施方案中,与相同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,多于一种延伸的条形码化的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或细胞标记的互补序列。

[0028] 在一些实施方案中,固体支持物包括合成颗粒、平坦表面或它们的组合。在一些实施方案中,样品包括单细胞,方法包括将包含多于一种寡核苷酸条形码的合成颗粒与样品中的单细胞关联。方法可以包括在将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相同的分区(partition)中。在一些实施方案中,分区是孔或液滴。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被固定或部分地固定在合成颗粒上,或者多于一种寡核苷酸条形码中的至少一种被包封或部分包封在合成颗粒中。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,合成颗粒是可破坏的水凝胶颗粒。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(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、硅酮及它们的任何组合。在一些实施方案中,多于一种寡核苷酸条形码中的每一种包含接头官能团,合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮,及它们的任何组合。

[0029] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;逆转录酶;包含靶结合区或其一部分的模板转换寡核苷酸;以及缺乏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。

[0030] 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Moloney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。试剂盒可以包含以下中的一种或更多种:乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砜(DMSO)、甘油、甲酰胺、7-脱氮-GTP、乙酰胺、四甲基氯化铵盐、甜菜碱或它们的任何组合。

[0031] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;末端脱氧核苷酸转移酶;以及缺乏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。在一些实施方案中,DNA聚合酶包括Klenow片段。试剂盒可以包含缓冲液、筒或两者。试剂盒可以包含一种或更多种用于逆转录反应和/或扩增反应的试剂。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,多于一种寡核苷酸条形码的每种样品标记和/或细胞标记包含至少6个核苷酸。在一些实施方案中,多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。

[0032] 在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被固定或部分地固定在合成颗粒上,或者被包封或部分地包封在合成颗粒中。在一些实施方案中,合成颗粒是可破坏的(例如,可破坏的水凝胶颗粒)。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(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、硅酮及它们的任何组合。在一些实施方案中,多于一种寡核苷酸条形码中的每一种包含接头官能团,合成颗粒包含固体支持物官能团,并且支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮,及它们的任何组合。

[0033] 附图简述

[0034] 图1示出了非限制性示例性条形码。

[0035] 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。

[0036] 图3是示出了用于从多于一种靶产生3'末端条形码化的靶的索引文库(indexed library)的非限制性示例性方法的示意图。

[0037] 图4A-图4B示出了在5'末端基因特异性标记核酸靶的非限制性示例性方法的示意图。

[0038] 图5A-图5B示出了在5'末端标记核酸靶用于全转录组分析的非限制性示例性方法的示意图。

[0039] 图6A-图6B示出了确定核酸靶(例如,免疫受体mRNA)的全长序列的非限制性示例性工作流程的示意图。

[0040] 图7是进行全长表达谱分析的非限制性示例性工作流程的示意图。

[0041] 图8描绘了根据本文提供的方法产生的非限制性示例性生物分析仪迹线。

[0042] 图9A-图9B示出了免疫受体mRNA的表达谱分析的非限制性示例性示意图。

[0043] 图10A-图10D描绘了与本文提供的基于随机引发的方法相关的数据。示出了与目前可用的方法(图10B和图10D)相比,用本文提供的基于随机引发的方法(图10A和图10C)获得的TCR产物(图10A-图10B)和BCR产物(图10C-图10D)的生物分析仪迹线。

[0044] 图11A-图11B描绘了与本文提供的基于随机引发的方法相关的数据。示出了与目前可用的方法相比,本文提供的基于随机引发的方法(-Pret)的全长VDJ序列(图11B)和来自细胞的VDJ序列(图11A)的百分比。

[0045] 图12A-图12D描绘了与本文提供的基于随机引发的方法相关的数据。示出了与目前可用的方法(图12B和图12D)相比,用本文提供的基于随机引发的方法(图12A和图12C)获得的BCR配对(图12A-图12B)和TCR配对(图12C-图12D)。

[0046] 详述

[0047] 在以下详细描述中,参考了构成本文的一部分的附图。在附图中,除非上下文另有指示,否则相似的符号通常标识相似的部件。在详细描述、附图和权利要求书中描述的说明性实施方案并不意味着是限制性的。在不脱离本文提出的主题的精神或范围的情况下,可以利用其他实施方案,并且可以做出其他改变。将容易理解的是,如本文一般描述的以及附图中图示的本公开内容的方面能够以各种不同的配置来布置、替换、组合、分离和设计,所有这些都在本文中明确考虑并且构成本公开内容的一部分。

[0048] 本文提及的所有专利、公布的专利申请、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。

[0049] 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)的绝对数量也可以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,这导致扩增偏差和不准确的基因表达测量。具有独特的分子标记(molecular labels,也称为分子索引(molecular indexes,MI))的随机条形码可以用

于计数分子数目和校正扩增偏倚。随机条形码化,诸如PreciseTM测定 (Cellular Research, Inc. (Palo Alto, CA)) 和RhapsodyTM测定 (Becton, Dickinson and Company (Franklin Lakes, NJ)), 可以通过在逆转录 (RT) 过程中使用分子标记 (ML) 标记mRNA来纠正由PCR和文库制备步骤引起的偏倚。

[0050] PreciseTM测定可以利用具有在多 (T) 寡核苷酸上的大量 (例如6561种至65536种) 独特分子标记序列的随机条形码的非耗尽性池 (non-depleting pool), 以在RT步骤期间与样品中的所有多 (A) -mRNA杂交。随机条形码可以包括通用PCR引发位点。在RT期间, 靶基因分子与随机条形码随机反应。每种靶分子可以与随机条形码杂交, 从而生成随机地条形码化的互补核糖核苷酸 (cDNA) 分子。在标记后, 可将来自微孔板的微孔的随机条形码化cDNA分子汇集到单支管中用于PCR扩增和测序。可以分析原始测序数据以产生读段的数量、具有独特分子标记序列的随机条形码的数量以及mRNA分子的数量。

[0051] 本文的公开内容包括对样品中的核酸靶进行标记的方法。在一些实施方案中, 方法包括: 使核酸靶的拷贝与多于一种寡核苷酸条形码接触, 其中每种寡核苷酸条形码包含第一通用序列、分子标记和能够与核酸靶杂交的靶结合区; 产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子; 使每种条形码化的核酸分子的靶结合区的互补体与以下中的一种或更多种的靶结合区杂交: (i) 多于一种寡核苷酸条形码中的寡核苷酸条形码, (ii) 条形码化的核酸分子自身, 以及 (iii) 多于一种条形码化的核酸分子中的不同条形码化的核酸分子; 延伸多于一种条形码化的核酸分子的3' 末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子; 使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子, 由此产生包含核酸靶的序列或其一部分的第一多于一种条形码化的扩增子; 使随机引物与第一多于一种条形码化的扩增子杂交并延伸随机引物以产生多于一种延伸产物, 其中随机引物包含第二通用序列或其互补体; 以及使用能够与第一通用序列或其互补体和第二通用序列或其互补体杂交的引物来扩增多于一种延伸产物, 从而产生第二多于一种条形码化的扩增子。

[0052] 本文的公开内容包括确定样品中核酸靶的数量的方法。在一些实施方案中, 方法包括: 使核酸靶的拷贝与多于一种寡核苷酸条形码接触, 其中每种寡核苷酸条形码包含第一通用序列、分子标记 (例如, 第一分子标记) 和能够与核酸靶杂交的靶结合区; 产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子; 使每种条形码化的核酸分子的靶结合区的互补体与以下中的一种或更多种的靶结合区杂交: (i) 多于一种寡核苷酸条形码中的寡核苷酸条形码, (ii) 条形码化的核酸分子自身, 以及 (iii) 多于一种条形码化的核酸分子中的不同条形码化的核酸分子; 延伸多于一种条形码化的核酸分子的3' 末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子; 使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子, 由此产生包含核酸靶的序列或其一部分的第一多于一种条形码化的扩增子; 使随机引物与第一多于一种条形码化的扩增子杂交并延伸随机引物以产生多于一种延伸产物, 其中随机引物包含第二通用序列或其互补体; 使用能够与第一通用序列或其互补体和第二通用序列或其互补体杂交的引物扩增多于一种延伸产物, 从而产生第二多于一种条形码化的扩增子; 以及基于与第二多于一种条形码化的扩增子或其产物关联的

具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来确定样品中核酸靶的拷贝数。

[0053] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;逆转录酶;包含靶结合区或其一部分的模板转换寡核苷酸;以及缺乏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。

[0054] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;末端脱氧核苷酸转移酶;以及缺乏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。

[0055] 定义

[0056] 除非另外定义,否则本文使用的技术术语和科学术语具有与本公开内容所属领域的普通技术人员通常所理解的相同含义。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)。为了本公开内容的目的,下文定义了以下术语。

[0057] 如本文使用的,术语“衔接子”可以意指促进关联的核酸的扩增或测序的序列。关联的核酸可包括靶核酸。关联的核酸可以包含以下中的一种或更多种:空间标记、靶标记、样品标记、索引化标记(indexing label)或条形码序列(例如,分子标记)。衔接子可以是线性的。衔接子可以是预腺苷酸化的衔接子。衔接子可以是双链或单链的。一种或更多种衔接子可以位于核酸的5'或3'末端。当衔接子在5'和3'末端包含已知序列时,已知序列可以是相同或不同的序列。位于多核苷酸的5'和/或3'末端的衔接子可以能够与固定在表面上的一种或更多种寡核苷酸杂交。在一些实施方案中,衔接子可以包含通用序列。通用序列可以是两个或更多个核酸分子共有的核苷酸序列的区域。两种或更多种核酸分子也可具有不同序列的区域。因此,例如,5'衔接子可以包含相同和/或通用核酸序列,并且3'衔接子可以包含相同和/或通用序列。可以存在于多于一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一通用引物复制或扩增多于一个不同序列。类似地,可以存在于核酸分子的集合中的不同成员中的至少一个、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)或更多个单一通用引物复制或扩增多于一个不同序列。因此,通用引物包括可以与这样的通用序列杂交的序列。可以修饰具有靶核酸序列的分子以将通用衔接子(例如,非靶核酸序列)附接到不同靶核酸序列的一个末端或两个末端。与靶核酸附接的一种或更多种通用引物可以提供用于通用引物的杂交的位点。附接到靶核酸的一种或更多种通用引物可以彼此相同或不同。

[0058] 如本文使用的,术语“关联(associated)”或“与.....关联(associated with)”可以意指两个或更多个种类(species)可以被鉴定为在某个时间点处共定位。关联可意指两个或更多个种类在或曾经在相似的容器内。关联可以是信息学关联。例如,关于两个或更多个种类的数字信息可以被存储并且可以用于确定所述种类中的一个或更多个在某个时间点处共定位。关联也可以是物理关联。在一些实施方案中,两个或更多个关联的种类彼此

之间或与共同的固体或半固体表面是“拴系的”、“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)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。

[0059] 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”,其中仅一些核苷酸结合,或者当所述单链分子之间存在完全互补性时,这种互补性可以是完全的。如果第一核苷酸序列与第二核苷酸序列互补,则可以称第一核苷酸序列是第二序列的“互补体”。如果第一核苷酸序列与第二序列相反的序列(即,核苷酸顺序相反)互补,则可以称第一核苷酸序列是第二序列的“反向互补体”。如本文使用的,“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,则其可以与其所杂交的分子互补或者部分互补。

[0060] 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数的方法。数字计数可以包括确定已经与样品中的靶关联的独特的标记的数量数的步骤。这种方法(其本质上可以是随机的)将计数分子的问题从相同分子的定位和鉴定之一转化为有关检测一组预定义标记的一系列是/否数字问题。

[0061] 如本文使用的,术语“一个标记(label)”或“多于一个标记(labels)”可以指与样品中的靶关联的核酸代码。标记可以是例如核酸标记。标记可以是完全或部分可扩增的标记。标记可以是完全或部分可测序的标记。标记可以是可鉴定为有区别的天然核酸的一部分。标记可以是已知的序列。标记可以包括核酸序列的连接区,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。

[0062] 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。每种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数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目与独特的条形码的数目比率低时,标记的靶分子是高度独特的(即,多于一个靶分子被给定标记标记的概率非常低)。

[0063] 如本文使用的,术语“核酸”是指多核苷酸序列或其片段。核酸可以包含核苷酸。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。核酸可以包含一种或更多种类类似物(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素)、含硫醇的核苷酸、生物素连接的核苷酸、荧光碱基类似物、CpG岛、甲基-7-鸟苷、甲基化的核苷酸、肌苷、硫代尿苷、假尿苷、二氢尿苷、癸苷(queuosine)以及怀俄苷(wyosine)。“核酸”、“多核苷酸”、“靶多核

苷酸”和“靶核酸”可以互换使用。

[0064] 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。在形成核酸时,磷酸基团可以将相邻的核苷彼此共价连接以形成线性聚合化合物。继而,此线性聚合化合物的各端可以进一步连接而形成环状化合物;然而,线性化合物通常是合适的。此外,线性化合物可以具有内部核苷酸碱基互补性,并且因此可以按产生完全或部分双链化合物的方式折叠。在核酸中,磷酸基团通常可以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。

[0065] 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)、磷酰胺(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。

[0066] 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(部分从核苷的糖部分形成);硅氧烷骨架;硫化物、亚砷和砷骨架;甲乙酰基(formacetyl)和硫代甲乙酰基(thioformacetyl)骨架;亚甲基甲乙酰基和硫代甲乙酰基骨架;核糖乙酰基(riboacetyl)骨架;含有烯烃的骨架;氨基磺酸酯骨架;亚甲基亚氨基和亚甲基胍基骨架;磺酸酯和磺酰胺骨架;酰胺骨架;和具有混合的N、O、S和CH₂组成部分的其他骨架。

[0067] 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两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。

[0068] 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。

[0069] 核酸可以包含具有附接到吗啉代环的杂环碱基的连接的吗啉代单元(例如吗啉代核酸)。连接基团可以连接吗啉代核酸中的吗啉代单体单元。非离子的基于吗啉代的寡聚化合物可与细胞蛋白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代替。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杂合体 (DNA/RNA hybrid) 的稳定性。CeNA寡核苷酸可以与核酸互补体形成稳定性与天然复合物相似的复合物。另外的修饰可以包括锁核酸 (LNA), 其中2'-羟基基团连接到糖环的4' 碳原子, 从而形成2'-C,4'-C-氧亚甲基连接, 从而形成双环糖部分。连接可以是亚甲基 (-CH₂), 桥接2' 氧原子和4' 碳原子的基团, 其中n是1或2。LNA和LNA类似物可以显示出与互补核酸的非常高的双链体热稳定性 (T_m=+3°C至+10°C)、对3'-核酸外切酶降解的稳定性和良好的溶解度特性。

[0070] 核酸还可以包括核碱基 (通常简称为“碱基”) 修饰或取代。如本文使用的, “未修饰的”或“天然的”核碱基可以包括嘌呤碱基 (例如腺嘌呤 (A) 和鸟嘌呤 (G)), 以及嘧啶碱基 (例如胸腺嘧啶 (T)、胞嘧啶 (C) 和尿嘧啶 (U))。经修饰的核碱基可以包括其他合成以及天然的核碱基, 诸如5-甲基胞嘧啶 (5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基衍生物和其他烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基衍生物和其他烷基衍生物、2-硫代尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-卤素尿嘧啶 (5-halouracil) 和胞嘧啶、5-丙炔基 (-C≡C-CH₃) 尿嘧啶和胞嘧啶以及嘧啶碱基的其他炔基衍生物、6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶 (假尿嘧啶)、4-硫代尿嘧啶、8-卤素、8-氨基、8-硫代、8-硫代烷基、8-羟基和其他8-取代的腺嘌呤和鸟嘌呤, 5-卤素特别是5-溴、5-三氟甲基和其他5-取代的尿嘧啶和胞嘧啶, 7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。修饰的核碱基可以包括三环嘧啶, 诸如吩噻嗪胞苷 (1H-嘧啶并 (5,4-b) (1,4) 苯并噻嗪-2 (3H)-酮)、吩噻嗪胞苷 (1H-嘧啶并 (5,4-b) (1,4) 苯并噻嗪-2 (3H)-酮)、G-夹类 (G-clamps), 诸如取代的吩噻嗪胞苷 (例如9-(2-氨基乙氧基)-H-嘧啶并 (5,4-b) (1,4) 苯并噻嗪-2 (3H)-酮)、吩噻嗪胞苷 (1H-嘧啶并 (5,4-b) (1,4) 苯并噻嗪-2 (3H)-酮)、G-夹类, 诸如取代的吩噻嗪胞苷 (例如9-(2-氨基乙氧基)-H-嘧啶并 (5,4-b) (1,4) 苯并噻嗪-2 (3H)-酮)、咪唑胞苷 (2H-嘧啶并 (4,5-b) 咪唑-2-酮)、吡啶并咪唑胞苷 (H-吡啶并 (3',2':4,5) 吡咯并 [2,3-d] 嘧啶-2-酮)。

[0071] 如本文使用的, 术语“样品”可以指包含靶的组合物。用于通过所公开的方法、装置和系统进行分析的合适样品包括细胞、组织、器官或生物体。

[0072] 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) 仪、细胞分选仪、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。

[0073] 如本文使用的, 术语“固体支持物”可以指可以附接多于一种条形码 (例如, 随机条形码) 的离散固体或半固体表面。固体支持物可以包括任何类型的实心的、多孔的或空心的球体、球、承座、圆柱体或其他类似形状, 由塑料、陶瓷、金属或聚合材料 (例如, 水凝胶) 构成, 其上可以固定核酸 (例如, 共价或非共价地)。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。

[0074] 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。

[0075] 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。

[0076] 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“随机标记”互换使用。

[0077] 如本文使用的,术语“靶”可以指可与条形码(例如,随机条形码)关联的组合物。用于通过所公开的方法、装置和系统进行分析的示例性合适的靶包括寡核苷酸、DNA、RNA、mRNA、微RNA、tRNA等。靶可以是单链的或双链的。在一些实施方案中,靶可以是蛋白、肽或多肽。在一些实施方案中,靶是脂质。如本文使用的,“靶”可以与“种类”互换使用。

[0078] 如本文使用的,术语“逆转录酶”可以指具有逆转录酶活性(即,催化从RNA模板合成DNA)的一组酶。通常,这样的酶包括但不限于逆转录病毒逆转录酶、逆转录转座子逆转录酶、逆转录质粒(retroplasmid)逆转录酶、逆转录子逆转录酶、细菌逆转录酶、II型内含子衍生的逆转录酶,及其突变体、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)。

[0079] 术语“通用衔接子引物”、“通用引物衔接子”或“通用衔接子序列”可互换地使用以指可以用于与条形码(例如,随机条形码)杂交以产生基因特异性条形码的核苷酸序列。通用衔接子序列例如可以是跨越本公开内容的方法中使用的所有条形码通用的已知序列。例如,当使用本文公开的方法标记多种靶时,每种靶特异性序列可以连接到相同的通用衔接子序列。在一些实施方案中,多于一种通用衔接子序列可以用于本文公开的方法中。例如,当使用本文公开的方法标记多个靶时,至少两种靶特异性序列连接到不同的通用衔接子序列。通用衔接子引物及其互补体可以包括在两种寡核苷酸中,其中的一种包含靶特异性序列且另一种包含条形码。例如,通用衔接子序列可以是包含靶特异性序列的寡核苷酸的一部分以产生与靶核酸互补的核苷酸序列。包含条形码以及通用衔接子序列的互补序列的第二寡核苷酸可与核苷酸序列杂交并产生靶特异性条形码(例如,靶特异性随机条形码)。在一些实施方案中,通用衔接子引物具有与本公开内容的方法中使用的通用PCR引物不同的序列。

[0080] 条形码

[0081] 条形码化,诸如随机条形码化,已在例如US 2015/0299784、W02015/031691和Fu等人,Proc Natl Acad Sci U.S.A.2011May31;108(22):9026-31中描述,这些公布和出版物的内容在此以其整体并入。在一些实施方案中,本文公开的条形码可以是随机条形码,所述随机条形码是可以用于对靶随机标记(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,或是至多以下,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。

[0082] 条形码,例如随机条形码,可以包含一种或更多种标记。示例性标记可以包括通用标记、细胞标记、条形码序列(例如,分子标记)、样品标记、板标记、空间标记和/或前空间标记(pre-spatial label)。图1图示了具有空间标记的示例性条形码104。条形码104可以包含可将条形码与固体支持物105连接的5' 胺。条形码可以包含通用标记、维度标记、空间标记、细胞标记和/或分子标记。条形码中不同标记(包括但不限于通用标记、维度标记、空间标记、细胞标记和分子标记)的顺序可以改变。例如,如图1中示出的,通用标记可以是最5' 侧的标记(5' -most label),且分子标记可以是最3' 侧的标记(3' -most label)。空间标记、维度标记和细胞标记可以处于任何顺序。在一些实施方案中,通用标记、空间标记、维度标记、细胞标记和分子标记是处于任何顺序的。条形码可以包含靶结合区。靶结合区可以与样品中的靶(例如,靶核酸、RNA、mRNA、DNA)相互作用。例如,靶结合区可以包括可以与mRNA的多(A)尾相互作用的寡聚(dT)序列。在一些情况下,条形码的标记(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列)可以由1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个或更多个核苷酸隔开。

[0083] 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,使得该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)能允许人们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。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。

[0084] 条形码可以包含靶结合区。靶结合区可以与样品中的靶相互作用。靶可以是,或包括核糖核酸(RNA)、信使RNA(mRNA)、微RNA、小干扰RNA(siRNA)、RNA降解产物、各自含有多

(A) 尾的RNA,或它们的任何组合。在一些实施方案中,多于一种靶可以包括脱氧核糖核酸(DNA)。

[0085] 在一些实施方案中,靶结合区可以包含可以与mRNA的多(A)尾相互作用的寡聚(dT)序列。条形码的标记中的一种或更多种(例如,通用标记、维度标记、空间标记、细胞标记和条形码序列(例如,分子标记))可以通过间隔序列(spacer)与条形码的剩余标记中的另一个或两个隔开。间隔序列可以是例如,1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个或更多个核苷酸。在一些实施方案中,条形码的标记没有被间隔序列隔开。

[0086] 通用标记

[0087] 条形码可以包含一种或更多种通用标记。在一些实施方案中,对于附接到给定的固体支持物的条形码组中的所有条形码,一种或更多种通用标记可以是相同的。在一些实施方案中,对于附接到多于一个珠的所有条形码,一种或更多种通用标记可以是相同的。在一些实施方案中,通用标记可以包括能够与测序引物杂交的核酸序列。测序引物可以用于对包含通用标记的条形码进行测序。测序引物(例如,通用测序引物)可以包括与高通量测序平台关联的测序引物。在一些实施方案中,通用标记可以包括能够与PCR引物杂交的核酸序列。在一些实施方案中,通用标记可以包括能够与测序引物和PCR引物杂交的核酸序列。能够与测序引物或PCR引物杂交的通用标记的核酸序列可以被称为引物结合位点。通用标记可以包括可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,或是约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个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,或是至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。在一些实施方案中,可裂解接头或修饰的核苷酸可以是通用标记序列的一部分,以使条形码能够从支持物上被裂解下来。

[0088] 维度标记

[0089] 条形码可以包含一种或更多种维度标记。在一些实施方案中,维度标记可以包括提供关于发生标记(例如,随机标记)的维度的信息的核酸序列。例如,维度标记可以提供关于对靶进行条形码化的时间的信息。维度标记可以与样品中条形码化(例如,随机条形码化)的时间关联。维度标记可以在标记的时间被激活。不同的维度标记可以在不同的时间被激活。维度标记提供关于靶、靶的组和/或样品被条形码化的顺序的信息。例如,在细胞周期的G0期可以将细胞的群体条形码化。在细胞周期的G1期,可以用条形码(例如,随机条形码)对细胞再次进行脉冲处理。在细胞周期的S期,可以用条形码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/或疗法的响应。

[0090] 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、类泛素化(sumoylate)、乙酰化、甲基化、去乙酰化、去甲基化)、光化学事件(例如,光锁定(photocaging))以及引入非天然的核苷酸将所述维度标记激活。

[0091] 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)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。

[0092] 多于一个固体支持物(例如,珠)可以表现多达 10^6 种或更多种独特的维度标记序列。维度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,或是至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。维度标记可以包含约5个至约200个之间的核苷酸。维度标记可以包含约10个至约150个之间的核苷酸。维度标记可以包含长度在约20个至约125个之间的核苷酸。

[0093] 空间标记

[0094] 条形码可以包含一种或更多种空间标记。在一些实施方案中,空间标记可以包括提供关于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、放射性或独特的尺寸或形状)的结构。空间标记可以与物理分区(例如,孔、容器或液滴)关联。在一些实施方案中,将多于一种空间标记一起用于编码空间中的一个或更多个位置。

[0095] 空间标记对于附接到给定的固体支持物(例如,珠)的所有条形码可以是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,包含相同空间标记的相同固体支持物上的条形码的百分比可以是以下或可以是约以下:60%、70%、80%、85%、90%、95%、97%、99%、100%或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,包含相同空间标记的相同固体支持物上的条形码的百分比可以是至少或是至多60%、70%、80%、85%、90%、95%、97%、99%或100%。在一些实施方案中,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。

[0096] 多于一个固体支持物(例如,珠)可以表现多达 10^6 种或更多种独特的空间标记序

列。空间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。空间标记可以包括长度在约5个至约200个之间的核苷酸,例如,约10个至约150个之间的核苷酸或约20个至约125个之间的核苷酸。

[0097] 细胞标记

[0098] 条形码(例如,随机条形码)可以包含一种或更多种细胞标记。在一些实施方案中,细胞标记可以包括提供用于确定哪种靶核酸来自哪个细胞的信息的核酸序列。在一些实施方案中,细胞标记对于附接到给定的固体支持物(例如,珠)上的所有条形码是相同的,但对于不同的固体支持物(例如,珠)是不同的。在一些实施方案中,相同固体支持物上包含相同细胞标记的条形码的百分比可以是以下,或是约以下: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。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,或是约以下:60%、70%、80%、85%、90%、95%、97%、99%或100%。例如,相同固体支持物上的至少60%或至少95%的条形码可以包含相同的细胞标记。

[0099] 多于一个固体支持物(例如,珠)可以表现多达 10^6 种或更多种独特的细胞标记序列。细胞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,或是至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。例如,细胞标记可以包括长度在约5个至约200个之间的核苷酸、约10个至约150个之间的核苷酸或约20个至约125个之间的核苷酸。

[0100] 条形码序列

[0101] 条形码可以包含一种或更多种条形码序列。在一些实施方案中,条形码序列可以包括提供与条形码杂交的特定类型的靶核酸种类的鉴定信息的核酸序列。条形码序列可以包括如下核酸序列,所述核酸序列提供与条形码(例如,靶结合区)杂交的靶核酸种类的特定出现的计数(例如,提供粗略近似)。

[0102] 在一些实施方案中,一组不同的条形码序列附接至给定的固体支持物(例如,珠)。在一些实施方案中,可以有以下,或有约以下的独特分子标记序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种,或这些值中的任何两个值之间的数字或范围。例如,多于一种条形码可以包括具有不同序列的约6561种条形码序列。作为另一个实例,多于一种条形码可以包括具有不同序列的约65536种条形码序列。在一些实施方案中,可以有至少以下,或至多以下的独特的条形码序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。独特分子标记序列可以附接至给定固体支持物(例如,珠)。在一些实施方案中,独特分子标记序列部分或全部被颗粒(例如,水凝胶珠)包含。

[0103] 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,或是约以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,或在这些值中的任何两个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,或是至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。

[0104] 分子标记

[0105] 条形码(例如,随机条形码)可以包含一种或更多种分子标记。分子标记可以包括条形码序列。在一些实施方案中,分子标记可以包括提供与条形码杂交的特定类型的靶核酸种类的鉴定信息的核酸序列。分子标记可以包括如下核酸序列,所述核酸序列提供与条形码(例如,靶结合区)杂交的靶核酸种类的特定出现的计数。

[0106] 在一些实施方案中,一组不同的分子标记附接到给定的固体支持物(例如,珠)。在一些实施方案中,可以有以下,或有约以下的独特分子标记序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种,或在这些值中的任何两个值之间的数字或范围。例如,多于一种条形码可以包括具有不同序列的约6561种分子标记。作为另一个实例,多于一种条形码可以包括具有不同序列的约65536种分子标记。在一些实施方案中,可以有至少以下,或有至多以下的独特分子标记序列: 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种或 10^9 种。具有独特分子标记序列的条形码可以附接到给定的固体支持物(例如,珠)。

[0107] 对于使用多于一种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,或是约以下:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1,或在这些值中的任何两个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,或是至多以下:1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1或100:1。

[0108] 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。分子标记的长度可以是至少以下,或是至多以下:1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个或300个核苷酸。

[0109] 靶结合区

[0110] 条形码可以包含一种或更多种靶结合区,诸如捕获探针。在一些实施方案中,靶结合区可以与感兴趣的靶杂交。在一些实施方案中,靶结合区可以包括与靶(例如,靶核酸、靶分子,例如待分析的细胞核酸),例如,与特定基因序列,进行特异性杂交的核酸序列。在一些实施方案中,靶结合区可以包括可附接(例如,杂交)至特定靶核酸的特定位置的核酸序列。在一些实施方案中,靶结合区可以包括能够与限制性酶位点突出端(例如,EcoRI粘性末端突出端)进行特异性杂交的核酸序列。然后条形码可以连接到包括与限制性位点突出端互补的序列的任何核酸分子。

[0111] 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可与多种靶核酸结合而不依赖靶核酸的特定序列的序列。例如,靶结合区可以包括随机多聚体序列或与mRNA分子上的多(A)尾杂交的寡聚(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,对于附接至给定珠的所有条形码,靶结合

区是相同的。在一些实施方案中,附接到给定珠的多于一种条形码的靶结合区可以包括两种或更多种不同的靶结合序列。靶结合区的长度可以是以下,或是约以下:5个、10个、15个、20个、25个、30个、35个、40个、45个、50个核苷酸,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。

[0112] 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,或是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。靶结合区的长度可以是约5-30个核苷酸。当条形码包含基因特异性靶结合区时,所述条形码在本文中可以被称为基因特异性条形码。

[0113] 定向特性(Orientation Property)

[0114] 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,样品可以经历等电聚焦,以便于将条形码定向成已知的方式。以这种方式,定向特性可以用于开发样品中条形码的已知的地图。示例性定向特性可以包括电泳迁移率(例如,基于条形码的尺寸)、等电点、自旋、电导率和/或自组装。例如,具有自组装的定向特性的条形码激活时可以自组装成特定的定向(例如,核酸纳米结构)。

[0115] 亲和特性(Affinity Property)

[0116] 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。

[0117] 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)。

[0118] 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

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。

[0119] 通用衔接子引物

[0120] 条形码可以包含一种或更多种通用衔接子引物。例如, 基因特异性条形码, 诸如基因特异性随机条形码, 可以包含通用衔接子引物。通用衔接子引物可以指跨越所有条形码通用的核苷酸序列。通用衔接子引物可以用于构建基因特异性条形码。通用衔接子引物的长度可以是以下, 或是约以下: 1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个核苷酸, 或在这些值中的任何两个值之间的数字或范围的核苷酸。通用衔接子引物的长度可以是至少以下, 或是至多以下: 1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。通用衔接子引物的长度可以是5-30个核苷酸。

[0121] 接头

[0122] 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。在某些情况下, 接头标记序列的长度是12个核苷酸。接头标记序列可以用于促进条形码的合成。接头标签可以包括纠错 (例如汉明) 码。

[0123] 固体支持物

[0124] 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, 第一条条形码序列) 的分子标记) 相差至少一个核苷酸。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相差至少一个核苷酸。例如, 第一固体支持物上的第一多于一个条形码的第一细胞标记可以具有相同的序列, 且第二固体支持物上的第二多于一个条形码的第二细胞标记可以具有相同的序列。第一固体支持物上的第一多于一个条形码的第一细胞标记和第二固体支持物上的第二多于一个条形码的第二细胞标记可以相差至少一个核苷酸。细胞标记可以是例如约5-20个核苷酸长。条形码序列可以是例如约5-20个核苷酸长。合成颗粒可以是例如珠。

[0125] 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。珠可以包含诸如以下的材料: 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。

[0126] 在一些实施方案中, 珠可以用条形码或随机条形码官能化的聚合珠, 例如可变形的珠或凝胶珠 (诸如来自10X Genomics (San Francisco, CA) 的凝胶珠)。在一些实施方式中, 凝胶珠可以包括基于聚合物的凝胶。例如, 凝胶珠可以通过将一种或更多种聚合物前体包封进液滴中来产生。在将聚合物前体暴露于促进剂 (例如, 四甲基乙二胺 (TEMED)) 后, 可以产生凝胶珠。

[0127] 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。所期望的条件可以导致聚合物珠以受控方式溶解、熔化或降解。凝胶珠可以由于化学刺激、物理刺激、生物刺激、热刺激、磁刺激、电刺激、光刺激或它们的任何组合而溶解、熔化或降解。

[0128] 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,通过寡核苷酸条形码和/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/固定可以通过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,例如通过不稳定部分(例如,通过化学交联剂,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至凝胶珠的情况下,使二硫键暴露于还原剂可以裂解二硫键并从珠释放寡核苷酸条形码。不稳定性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被包括。在一些实施方案中,多于一种条形码中的至少一种条形码可以被固定在颗粒上、部分地固定在颗粒上、被包封在颗粒中、部分地包封在颗粒中,或它们的任何组合。

[0129] 在一些实施方案中,凝胶珠可以包括宽范围的不同的聚合物,包括但不限于:聚合物、热敏聚合物、光敏聚合物、磁性聚合物、pH敏感聚合物、盐敏感聚合物、化学敏感聚合物、聚电解质、多糖、肽、蛋白和/或塑料。聚合物可以包括但不限于以下材料:诸如聚(N-异丙基丙烯酰胺)(PNIPAAm)、聚(苯乙烯磺酸酯)(PSS)、聚(烯丙基胺)(PAAm)、聚(丙烯酸)(PAA)、聚(乙烯亚胺)(PEI)、聚(双烯丙基二甲基-氯化铵)(PDADMAC)、聚(吡咯)(poly(pyrolle), PPy)、聚(乙烷基吡咯烷酮)(PVPON)、聚(乙烷基吡啶)(PVP)、聚(甲基丙烯酸)(PMAA)、聚(甲基丙烯酸甲酯)(PMMA)、聚苯乙烯(PS)、聚(四氢呋喃)(PTHF)、聚(邻苯二甲醛)(PPA)、聚(己基紫精)(PHV)、聚(L-赖氨酸)(PLL)、聚(L-精氨酸)(PARG)、聚(乳酸-共-羟基乙酸)(PLGA)。

[0130] 许多化学刺激物可以用于触发珠的破坏、溶解或降解。这些化学变化的实例可包括但不限于pH介导的珠壁变化、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)。容量变化(bulk changes)也可以用于触发珠的破坏。

[0131] 通过各种刺激物对微囊的容量或物理变化在将囊设计为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。这些过程可以包括但不限于压力引起的破裂、珠壁熔化,或珠壁的孔隙率的变化。

[0132] 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,一个实例可以包括含有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。在另一种实例中,珠包括包含纤维素的壳壁。壳聚糖水解酶的添加用作纤维素键裂解、壳壁解聚及其内部内容物释放的生物触发物。

[0133] 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,热量可以增加珠的内部组分的内部压力,使得珠破裂或爆炸。在又其他的情况下,热量可以使珠转化成收缩的脱水状态。热量还可以作用于珠壁内的热敏聚合物,从而引起珠的破坏。

[0134] 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,在存在振荡磁场刺激的情况下触发破裂。

[0135] 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。

[0136] 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照射可以导致珠壁的崩解。在又一个实例中,可以将可光转换的材料,诸如偶氮苯基团,掺入珠壁中。在施加UV或可见光后,诸如这些的化学物质在吸收光子后经历可逆的顺式-至-反式异构化。在此方面,掺入光子转换(photonswitches)可以导致珠壁在施加光触发物后可崩解或变得更为多孔。

[0137] 例如,在图2中示出的条形码化(例如,随机条形码化)的非限制性实例中,在框208处将细胞(诸如单细胞)引入微孔阵列的多于一个微孔上之后,在框212处可以将珠引入微孔阵列的多于一个微孔上。每个微孔可以包含一个珠。珠可以包含多于一种条形码。条形码可以包含附接到珠的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。

[0138] 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:具有独特的序列的至少100种或1000种条形码序列。在一些实施方案中,与固体支持物关联的不同条形码可以包括具有不同序列的条形码。在一些实施方案中,与固体支持物关联的条形码的一定百分比包含相同的细胞标记。例如,所述百分比可以是以下,或是约以下: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:具有独特的序列的至少100种或1000种细胞标记。

[0139] 本文公开的条形码可以与固体支持物(例如,珠)关联(例如,附接)。在一些实施方案中,可以用包括与多于一种条形码关联的多于一个合成的颗粒的固体支持物对样品中的多于一种靶进行条形码化。在一些实施方案中,固体支持物可包括与多于一种条形码关联的多于一个合成的颗粒。不同固体支持物上的多于一种条形码的空间标记可以相差至少一个核苷酸。固体支持物可以例如包括处于二维或三维的多于一种条形码。合成的颗粒可以是珠。珠可以是硅胶珠、可控孔径玻璃珠、磁珠、Dynabead、交联葡聚糖/琼脂糖凝胶珠、珠状纤维素、聚苯乙烯珠,或它们的任何组合。固体支持物可包括聚合物、基质、水凝胶、针阵列装置、抗体,或它们的任何组合。在一些实施方案中,固体支持物可以自由浮动。在一些实施

方案中,固体支持物可嵌入半固体或固体阵列中。条形码可以不与固体支持物关联。条形码可以是单独的核苷酸。条形码可以与基底关联。

[0140] 如本文使用的,术语“拴系的”、“附接的”和“固定的”可以互换使用,并且可以指用于将条形码附接至固体支持物的共价或非共价方式。可以将各种不同的固体支持物中的任何一种用作固体支持物,以用于附接预先合成的条形码或用于条形码的原位固相合成。

[0141] 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。

[0142] 珠可以包括各种材料,包括但不限于顺磁性材料(例如,镁、钼、锂和钽)、超顺磁性材料(例如,铁氧体(Fe_3O_4 ;磁铁矿)纳米颗粒)、铁磁材料(例如,铁、镍、钴,它们的一些合金,以及一些稀土金属化合物)、陶瓷、塑料、玻璃、聚苯乙烯、二氧化硅、甲基苯乙烯、丙烯酸聚合物、钛、胶乳、琼脂糖凝胶、琼脂糖、水凝胶、聚合物、纤维素、尼龙,或它们的任何组合。

[0143] 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。

[0144] 本文公开的一些实施方案包括一个或更多个颗粒(例如,珠)。颗粒中的每一个可以包括多于一种寡核苷酸(例如,条形码)。多于一种寡核苷酸中的每一种可以包括条形码序列(例如,分子标记序列)、细胞标记和靶结合区(例如,寡聚(dT)序列、基因特异性序列、随机多聚体,或它们的组合)。多于一种寡核苷酸中的每一种的细胞标记序列可以是相同的。不同颗粒上的寡核苷酸的细胞标记序列可以是不同的,使得可以鉴定不同颗粒上的寡核苷酸。在不同实施方式中,不同细胞标记序列的数量可以是不同的。在一些实施方案中,细胞标记序列的数量可以是以下,或是约以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 、 10^9 ,在这些值中的任何两个值之间的数字或范围,或更多。在一些实施方案中,细胞标记序列的数量可以是至少以下,或是至多以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 或 10^9 。在一些实施方案中,多于一个颗粒中不超过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。

[0145] 每个颗粒上的多于一种寡核苷酸可以包含不同的条形码序列(例如,分子标记)。在一些实施方案中,条形码序列的数量可以是以下,或是约以下:10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、

30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 、 10^9 ，或这些值中的任何两个值之间的数字或范围。在一些实施方案中，条形码序列的数量可以是至少以下，或是至多以下：10、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、 10^6 、 10^7 、 10^8 或 10^9 。例如，多于一种寡核苷酸中的至少100种包含不同的条形码序列。作为另一个实例，在单个颗粒中，多于一种寡核苷酸中的至少100种、500种、1000种、5000种、10000种、15000种、20000种、50000种，这些值中的任何两个之间的数字或范围，或更多种包含不同的条形码序列。一些实施方案提供了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。

[0146] 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。

[0147] 珠的直径可以与基底的孔的直径相关。在一些实施方案中，珠的直径可以比孔的直径长或短以下或约以下：10%、20%、30%、40%、50%、60%、70%、80%、90%、100%或这些值中的任何两个值之间的数字或范围。珠的直径可以与细胞（例如，被基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以下或约以下：10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%、300%或这些值中的任何两个值之间的数字或范围。在一些实施方案中，珠的直径可以比细胞的直径长或短至少或至多10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、150%、200%、250%或300%。

[0148] 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记来鉴定。

[0149] 珠的实例可以包括但不限于链霉抗生物素蛋白珠、琼脂糖珠、磁珠、**Dynabeads®**、**MACS®**微珠、抗体缀合珠（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。

[0150] 珠可以与量子点或荧光染料关联（例如，用量子点或荧光染料浸渍），以使其在一个荧光光学通道或多于一个光学通道中发荧光。珠可以与氧化铁或氧化铬关联，以使其为顺磁性或铁磁性。珠可以是可鉴定的。例如，可以使用照相机对珠进行成像。珠可以具有与珠关联的可检测编码。例如，珠可以包含条形码。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。

[0151] 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。

[0152] 固体支持物可以包含可溶性、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,诸如在柱中;或者在浸量尺(dipstick)中。

[0153] 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、针形物(pins)或针(needles)(例如适于组合合成或分析的针形物阵列),或凹点(pit)阵列中的珠或平坦表面诸如晶片(例如硅晶片)的纳升孔,具有带有或不带有过滤器底的凹点的晶片。

[0154] 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。

[0155] 基底和微孔阵列

[0156] 如本文使用的,基底可以指固体支持物类型。基底可以指可以包含本公开内容的条形码或随机条形码的固体支持物。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,微孔可以截留一个或更多个固体支持物。在一些实施方案中,微孔仅能截留一个固体支持物。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。

[0157] 条形码化的方法

[0158] 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。该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。该方法还可以包括对从条形码上的空间标记获得的信息进行分析和/或可视化。在一些实施方案中,该方法包括使样品中的多于一种靶可视化。将多于一种靶映射到样品的地图上可以包括生成样品的二维地图或三维地图。可以在对样品中的多于一种靶进行条形码化(例如,随机条形码化)之前或之后生成二维地图和三维地图。使样品中的多于一种靶可视化可以包括将多于一种靶映射到样品的地图上。将多于一种靶映射到样品的地图上可以包括生成样品的二维地图或三维地图。可以在对样品中的多于一种靶进行条形码化之前或之后生成二维地图和三维地图。在一些实施方案中,可以在裂解样品之前或之后生成二维地图和三维地图。在生成二维地图或三维地图之前或之后裂解样品可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。

[0159] 在一些实施方案中,对多于一种靶进行条形码化包括使多于一种条形码与多于一

种靶杂交以产生条形码化靶(例如,随机条形码化靶)。对多于一种靶进行条形码化可以包括生成条形码化靶的索引文库。生成条形码化靶的索引文库可以用包含多于一种条形码(例如,随机条形码)的固体支持物进行。

[0160] 使样品和条形码接触

[0161] 本公开内容提供了用于使样品(例如,细胞)与本公开内容的基底接触的方法。可以使包括例如细胞、器官或组织薄切片的样品与条形码(例如,随机条形码)接触。细胞可以例如通过重力流来接触,其中可以使细胞沉淀并且产生单层。样品可以是组织薄切片。可以将薄切片放置于基底上。样品可以是一维的(例如,形成平坦表面)。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。

[0162] 当条形码靠近靶时,靶可以与条形码杂交。条形码可以按不可耗尽的比率接触,使得每种不同的靶可以与本公开内容的不同条形码关联。为了确保靶与条形码之间的有效关联,可以将靶与条形码交联。

[0163] 细胞裂解

[0164] 在分配细胞和条形码之后,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、TritonX-100、吐温-20或NP-40)的细胞裂解缓冲液、有机溶剂(例如甲醇或丙酮)或消化酶(例如蛋白酶K、胃蛋白酶或胰蛋白酶),或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。

[0165] 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中的靶与基底的杂交。

[0166] 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0.1M、0.5M或1M或更高。裂解缓冲液中的盐浓度可以是至多约0.1M、0.5M或1M或更高。在一些实施方案中,裂解缓冲液中的盐浓度为约0.5M。裂解缓冲液可以包含去污剂(例如SDS、十二烷基硫酸锂、tritonX、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)。裂解缓冲液中的还原剂浓度可以是至少约1mM、5mM、10mM、15mM或20mM或更高。裂解缓冲液中还原剂的浓度可以是至多约1mM、5mM、10mM、15mM或20mM或更高。在一些实施方案中, 裂解缓冲液中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 Tris HCl, 约pH 7.5、约0.5M LiCl、约1%十二烷基硫酸锂、约10mM EDTA和约5mM DTT。

[0167] 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分钟或更多分钟。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。

[0168] 将条形码附接至靶核酸分子

[0169] 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。

[0170] 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。

[0171] 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。

[0172] 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, 不需要首先汇集来自多于一个细胞的经标记的靶核酸分子。

[0173] 逆转录

[0174] 本公开内容提供了使用逆转录(例如, 在图2的框224处)来产生靶-条形码缀合物

的方法。靶-条形码缀合物可以包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。

[0175] 在一些实施方案中,标记的RNA分子的逆转录可以通过添加逆转录引物而发生。在一些实施方案中,逆转录引物是寡聚(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)引物的长度是12个-18个核苷酸,并且与哺乳动物mRNA的3'末端的内源多(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。

[0176] 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次、2次、3次、4次、5次、6次、7次、8次、9次、10次、11次、12次、13次、14次、15次、16次、17次、18次、19次或20次逆转录反应。该方法可以包括进行至少约25次、30次、35次、40次、45次、50次、55次、60次、65次、70次、75次、80次、85次、90次、95次或100次逆转录反应。

[0177] 扩增

[0178] 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于一个拷贝。扩增可以以多重方式(inamultiplexed manner)进行,其中多于一种靶核酸序列同时进行扩增。扩增反应可用于将测序衔接子添加至核酸分子。扩增反应可以包括扩增样品标记(如果存在)的至少一部分。扩增反应可以包括扩增细胞标记和/或条形码序列(例如,分子标记)的至少一部分。扩增反应可以包括扩增样品标签、细胞标记、空间标记、条形码序列(例如,分子标记)、靶核酸,或它们的组合的至少一部分。扩增反应可以包括扩增多于一种核酸的0.5%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、100%,或这些值中的任何两个值之间的数字或范围。该方法还可以包括进行一个或更多个cDNA合成反应以产生包含样品标记、细胞标记、空间标记和/或条形码序列(例如,分子标记)的靶条形码分子的一个或更多个cDNA拷贝。

[0179] 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。

[0180] 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。其他非基于PCR的扩增方法包括DNA依赖性RNA聚合酶驱动的RNA转录扩增或RNA指导的DNA合成和转录的多个循环以扩增DNA或RNA靶、连接酶链式反应(LCR)和QB复制酶(QB)方法、回文探针的使用、链置换扩增、使用限制性核酸内切酶的寡核苷酸驱动的扩增、使引物与核酸序列杂交并且将所得双链体在延伸反应和扩增之前裂解的扩增方法、使用缺乏5'核酸外切酶活性的核酸聚合酶的链置换扩增、滚环扩增和分支延伸扩增(RAM)。在一些实施方案中,扩增不产生环化转录物。

[0181] 在一些实施方案中,本文公开的方法还包括对标记的核酸(例如,标记的RNA、标记的DNA、标记的cDNA)进行聚合酶链式反应以产生标记的扩增子(例如,随机标记的扩增子)。标记的扩增子可以是双链分子。双链分子可以包括双链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。

[0182] 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。添加非天然核苷酸可以用于鉴定扩增反应中特定循环或时间点的产物。

[0183] 进行一个或更多个扩增反应可以包括使用一种或更多种引物。一种或更多种引物可以包括例如,1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个或15个或更多个核苷酸。一种或更多种引物可以包括至少1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个或15个或更多个核苷酸。一种或更多种引物可以包括少于12-15个核苷酸。一种或更多种引物可以退火至多于一种经标记的靶(例如,随机标记的靶)的至少一部分。一种或更多种引物可以退火至多于一种经标记的靶的3'末端和/或5'末端。一种或更多种引物可以退火至多于一种经标记的靶的内部区域。内部区域可以与多于一种经标记的靶的3'末端距离至少约50个、100个、150个、200个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸。一种或更多种引物可以包括一组固定的引物。一种或更多种引物可以包括至少一种或更多种定制引物。一种或更多种引物可以包括至少一种或更多种对照引物。一种或更多种引物可以包括至少一种或更多种基因特异性引物。

[0184] 一种或更多种引物可以包括通用引物。通用引物可以退火至通用引物结合位点。一种或更多种定制引物可以退火至第一样品标记、第二样品标记、空间标记、细胞标记、条形码序列(例如,分子标记)、靶,或它们的任何组合。一种或更多种引物可以包括通用引物和定制引物。定制引物可以被设计用于扩增一种或更多种靶。靶可以包括一个或更多个样品中总核酸的子集。靶可以包括一个或更多个样品中总的经标记的靶的子集。一种或更多种引物可以包括至少96种或更多种定制引物。一种或更多种引物可以包括至少960种或更多种定制引物。一种或更多种引物可以包括至少9600种或更多种定制引物。一种或更多种定制引物可以退火至两种或更多种不同的经标记的核酸。两种或更多种不同的经标记的核酸可以对应于一个或更多个基因。

[0185] 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序列的引物来扩增衔接至珠的分子。第二轮PCR可以使用侧翼为Illumina测序引物2序列的巢式基因特异性引物和针对通用Illumina测序引物1序列的引物扩增第一PCR产物。第三轮PCR添加P5和P7以及样品索引,以使PCR产物变成Illumina测序文库。使用150bp×2测序的测序可以揭示读段1上的细胞标记

和条形码序列(例如,分子标记)、读段2上的基因以及索引1读段上的样品索引。

[0186] 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的核酸可以用于将核酸从基底去除。例如,核酸可以使用进行核苷酸切除的酶(诸如,碱基切除修复酶,诸如无嘌呤/无嘧啶(AP)核酸内切酶)将核酸从基底去除。在一些实施方案中,核酸可以使用可光裂解(photocleavable)基团以及光从基底去除。在一些实施方案中,可以使用可裂解接头将核酸从基底去除。例如,可裂解接头可以包括以下中的至少一种:生物素/亲和素、生物素/链霉抗生物素蛋白、生物素/中性链亲和素(neutravidin)、Ig蛋白A、光不稳定性接头、酸或碱不稳定性接头基团,或适配体。

[0187] 当探针是基因特异性探针时,可以使分子与探针杂交,并且进行逆转录和/或扩增。在一些实施方案中,在核酸已经合成(例如,逆转录)之后,可以将其扩增。扩增可以以多重方式进行,其中多种靶核酸序列同时进行扩增。扩增可以将测序衔接子添加至核酸。

[0188] 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至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链附接至颗粒。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至颗粒表面,其中每个单链核酸分子的另一个末端是未附接的。在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。

[0189] 扩增反应可以包括扩增多于一种核酸的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%。

[0190] 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。

[0191] 在一些实施方案中,对经标记的核酸的扩增包括非基于PCR的方法。非基于PCR的

方法的实例包括但不限于多重置换扩增 (MDA)、转录介导的扩增 (TMA)、基于核酸序列的扩增 (NASBA)、链置换扩增 (SDA)、实时 SDA、滚环扩增或环到环扩增。其他非基于 PCR 的扩增方法包括 DNA 依赖性 RNA 聚合酶驱动的 RNA 转录扩增或 RNA 指导的 DNA 合成和转录的多个循环以扩增 DNA 或 RNA 靶、连接酶链式反应 (LCR)、 $Q\beta$ 复制酶 ($Q\beta$) 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/或分支延伸扩增 (RAM)。

[0192] 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。

[0193] 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于一种扩增子。本文公开的方法可以包括进行至少约 1 次、2 次、3 次、4 次、5 次、6 次、7 次、8 次、9 次、10 次、11 次、12 次、13 次、14 次、15 次、16 次、17 次、18 次、19 次或 20 次扩增反应。可替代地,该方法包括进行至少约 25 次、30 次、35 次、40 次、45 次、50 次、55 次、60 次、65 次、70 次、75 次、80 次、85 次、90 次、95 次或 100 次扩增反应。

[0194] 扩增还可以包括将一个或多个对照核酸添加至一个或多个包含多于一个核酸的样品中。扩增还可以包括将一个或多个对照核酸添加至多于一个核酸中。对照核酸可以包括对照标记。

[0195] 扩增可以包括使用一个或多个非天然核苷酸。非天然核苷酸可以包括光不稳定和/或可触发的核苷酸。非天然核苷酸的实例包括但不限于肽核酸 (PNA)、吗啉代核酸和锁核酸 (LNA),以及二醇核酸 (GNA) 与苏糖核酸 (TNA)。可以将非天然核苷酸添加至扩增反应的一个或多个循环中。添加非天然核苷酸可以用于鉴定扩增反应中特定循环或时间点的产物。

[0196] 进行一个或多个扩增反应可以包括使用一种或更多种引物。一种或更多种引物可以包括一种或更多种寡核苷酸。一种或更多种寡核苷酸可以包括至少约 7-9 个核苷酸。一种或更多种寡核苷酸可以包括少于 12-15 个核苷酸。一种或更多种引物可以退火至多于一种经标记的核酸的至少一部分。一种或更多种引物可以退火至多于一种经标记的核酸的 3' 末端和/或 5' 末端。一种或更多种引物可以退火至多于一种经标记的核酸的内部区域。内部区域可以与多于一种经标记的核酸的 3' 末端距离至少约 50 个、100 个、150 个、200 个、220 个、230 个、240 个、250 个、260 个、270 个、280 个、290 个、300 个、310 个、320 个、330 个、340 个、350 个、360 个、370 个、380 个、390 个、400 个、410 个、420 个、430 个、440 个、450 个、460 个、470 个、480 个、490 个、500 个、510 个、520 个、530 个、540 个、550 个、560 个、570 个、580 个、590 个、600 个、650 个、700 个、750 个、800 个、850 个、900 个或 1000 个核苷酸。一种或更多种引物可以包括一组固定的引物、至少一种或更多种定制的引物、至少一种或更多种对照引物,或它们的组合。一个或多个引物可以包括至少一个或多个管家基因引物。一种或更多种引物可以包括通用引物。通用引物可以退火至通用引物结合位点。一种或更多种定制引物可以退火至第一样品标签、第二样品标签、分子标识符标记、核酸或它们的产物。一种或更多种引物

可以包括通用引物和定制引物。定制引物可以被设计成扩增一个或更多个靶核酸。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,引物是附接至本公开内容的阵列的探针。

[0197] 在一些实施方案中,使样品中的多于一种靶条形码化(例如,随机条形码化)还包括生成条形码化靶(例如,随机条形码化靶)的索引文库或所述靶的条形码化片段的索引文库。不同条形码的条形码序列(例如,不同的随机条形码的分子标记)可以彼此不同。生成条形码化靶的索引文库包括从样品中的多于一种靶生成多于一种索引多核苷酸。例如,对于包括第一索引靶和第二索引靶的条形码化靶的索引文库,第一索引多核苷酸的标记区与第二索引多核苷酸的标记区可以相差以下、相差约以下、相差至少以下或相差至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个核苷酸或在这些值中的任何两个之间的数字或范围的核苷酸。在一些实施方案中,产生条形码化靶的索引文库包括使多于一个靶(例如mRNA分子)与包含多(T)区和标记区的多于一个寡核苷酸接触;以及使用逆转录酶进行第一链合成以产生单链标记的cDNA分子(每个包含cDNA区和标记区),其中多于一个靶包括至少两个不同序列的mRNA分子,且多于一个寡核苷酸包括至少两个不同序列的寡核苷酸。生成条形码化靶的索引文库还可以包括扩增单链标记的cDNA分子以产生双链标记的cDNA分子;以及对双链标记的cDNA分子进行巢式PCR以产生标记的扩增子。在一些实施方案中,所述方法可以包括产生衔接子标记的扩增子。

[0198] 条形码化(例如,随机条形码化)可以包括使用核酸条形码或标签以对单种核酸(例如,DNA或RNA)分子进行标记。在一些实施方案中,其包括从mRNA产生cDNA分子时将DNA条形码或标签添加至cDNA分子。可以进行巢式PCR以使PCR扩增偏倚最小化。可以添加衔接子用于使用例如下一代测序(NGS)的测序。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。

[0199] 图3是示出了产生条形码化靶(例如,随机条形码化靶)的索引文库,诸如条形码化的mRNA或其片段的索引文库的非限制性示例性过程的示意图。如步骤1中示出的,逆转录过程可以用独特分子标记序列、细胞标记序列和通用PCR位点对每个mRNA分子进行编码。特别地,通过使一组条形码(例如,随机条形码)310与RNA分子302的多(A)尾区308杂交(例如,随机杂交),可以将RNA分子302逆转录以产生经标记的cDNA分子304(包括cDNA区306)。条形码310中的每一种可以包含靶结合区,例如多(dT)区312、标记区314(例如,条形码序列或分子)和通用PCR区316。

[0200] 在一些实施方案中,细胞标记序列可以包含3个至20个核苷酸。在一些实施方案中,分子标记序列可以包含3个至20个核苷酸。在一些实施方案中,多于一种随机条形码中的每种还包含通用标记和细胞标记中的一种或更多种,其中固体支持物上的多于一种随机条形码的通用标记是相同的,并且固体支持物上的多于一种随机条形码的细胞标记是相同的。在一些实施方案中,通用标记可以包括3个至20个核苷酸。在一些实施方案中,细胞标记包括3个至20个核苷酸。

[0201] 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个核苷酸,或这些值中的任何两个值之间的数字或范围的核苷酸。细胞标记320的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个核苷酸,或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个核苷酸,或这些值中的任何两个值之间的数字或范围的核苷酸。固体支持物上的多于一种随机条形码的通用标记可以是相同的,并且固体支持物上的多于一种随机条形码的细胞标记是相同的。维度标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个核苷酸,或这些值中的任何两个值之间的数字或范围的核苷酸。

[0202] 在一些实施方案中,标记区314可以包括以下,包括约以下,包括至少以下,或包括至多以下:1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种,或这些值中的任何两个值之间的数字或范围的不同标记,诸如条形码序列或分子标记318和细胞标记320。每种标记的长度可以是以下,可以是约以下,可以是至少以下,或可以是至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个核苷酸,或这些值中的任何两个值之间的数字或范围的核苷酸。一组条形码或随机条形码310可以含有以下,含有约以下,含有至少以下,或可以是至多以下:10种、20种、40种、50种、70种、80种、90种、 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种、 10^{10} 种、 10^{11} 种、 10^{12} 种、 10^{13} 种、 10^{14} 种、 10^{15} 种、 10^{20} 种,或这些值中的任何两个值之间的数字或范围的条形码或随机条形码310。并且一组条形码或随机条形码310可以例如,各自含有独特标记区314。经标记的cDNA分子304可以进行纯化以去除过量条形码或随机条形码310。纯化可以包括Ampure珠纯化。

[0203] 如步骤2中示出的,来自步骤1中的逆转录过程的产物可以汇集至1支管中,并且用第1PCR引物池和第1通用PCR引物进行PCR扩增。由于独特标记区314,汇集是可能的。特别地,可以将经标记的cDNA分子304扩增以产生巢式PCR标记的扩增子322。扩增可以包括多重PCR扩增。扩增可以包括在单一反应体积中用96种多重引物进行的多重PCR扩增。在一些实施方案中,在单一反应体积中,多重PCR扩增可以利用以下,利用约以下,利用至少以下,或利用至多以下:10种、20种、40种、50种、70种、80种、90种、 10^2 种、 10^3 种、 10^4 种、 10^5 种、 10^6 种、 10^7 种、 10^8 种、 10^9 种、 10^{10} 种、 10^{11} 种、 10^{12} 种、 10^{13} 种、 10^{14} 种、 10^{15} 种、 10^{20} 种,或这些值中的任何两个值之间的数字或范围的多重引物。扩增可以包括使用包括靶向特异性基因的定制引物326A-C的第1PCR引物池324和通用引物328。定制引物326可以与经标记的cDNA分子304的cDNA部分306'内的区域杂交。通用引物328可以与经标记的cDNA分子304的通用PCR区域316杂交。

[0204] 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可通过巢式PCR进行进一步扩增。巢式PCR可以包括在单个反应体积中用巢式PCR引物332a-c的巢式PCR引物池330和第2通用PCR引物328'进行的多重PCR。巢式PCR引物池330

可以包含以下,包含约以下,包含至少以下,或包含至多以下:1种、2种、3种、4种、5种、6种、7种、8种、9种、10种、20种、30种、40种、50种、60种、70种、80种、90种、100种、200种、300种、400种、500种、600种、700种、800种、900种、1000种或这些值中的任何两个值之间的数字或范围的不同的巢式PCR引物332。巢式PCR引物332可以含有衔接子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。

[0205] 如步骤4中示出的,来自步骤3的PCR产物可以使用文库扩增引物进行PCR扩增用于测序。特别地,衔接子334和336可以用于对衔接子标记的扩增子338进行一个或更多个另外的测定。衔接子334和336可以与引物340和342杂交。一种或更多种引物340和342可以是PCR扩增引物。一种或更多种引物340和342可以是测序引物。一种或更多种衔接子334和336可以用于衔接子标记的扩增子338的进一步扩增。一种或更多种衔接子334和336可以用于对衔接子标记的扩增子338进行测序。引物342可以含有板索引344,使得使用同一组条形码或随机条形码310产生的扩增子可以在一轮测序反应中使用下一代测序(NGS)进行测序。

[0206] 核酸靶5’末端上的条形码化

[0207] 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被标记的核酸靶(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互补或补充例如基于3’的转录物计数方法(例如RhapsodyTM测定(Becton,Dickinson and Company (FranklinLakes,NJ))、ChromiumTM单细胞3’解决方案(10X Genomics (SanFrancisco, CA)))。条形码化的核酸靶可用于序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或以高通量方式进行全长测序。对5’末端(相对于被标记的靶核酸靶的5’)转录物计数可以揭示在核酸分子的5’末端或较靠近核酸分子的5’末端的选择性剪接同种型和变体(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,方法可以涉及分子内杂交。

[0208] 图4A-图4B示出了在5’末端基因特异性标记核酸靶的非限制性示例性方法400的示意图。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或其他核酸靶与多腺苷酸化RNA转录物424结合,用于进行标记或条形码化(例如,进行独特标记)。条形码420可以包含分子标记(ML)428和样品标记(SL)430用于分别标记转录物424和追踪RNA转录物424的样品来源,和位于每个条形码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)用于后续反应。每个样品的条形码中分子标记的序列的组库可以足够大以对RNA转录物进行随机标记。

[0209] 在框402处的cDNA合成以产生包含RNA转录物424(或其一部分)的条形码化的cDNA分子434之后,可以使用基因特异性方法进行5’分子条形码化。在框404处的基因特异性扩增(可是是任选的)之后,在框406处可以添加末端转移酶和脱氧腺苷三磷酸(dATP)以促进3’多(dA)加尾来产生具有多(A)尾438的扩增子436。在框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)尾的条形码化的cDNA分子)的分离。在框

410处,扩增子436的反向链436c可以经由其3'末端的多(dA)尾438和该链的多(dT)区422末端进行分子内杂交以形成发夹或茎环440。然后,在框412处,聚合酶(例如,Klenow片段)可用于从多(dA)尾438延伸以复制条形码来形成延伸的条形码化的反向链442。然后可以在框414(例如,任选地)处进行基因特异性扩增以扩增感兴趣的基因来产生5'末端(相对于RNA转录物424)上具有条形码的扩增子444用于在框416处的测序。在一些实施方案中,方法400包括在框404处的条形码化的cDNA分子434的基因特异性扩增和在框414处的延伸的条形码化的反向链442的基因特异性扩增中的一种或两者。

[0210] 图5A-图5B示出了在5'末端标记核酸靶用于全转录组分析的非限制性示例性方法500的示意图。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或其他核酸靶与多腺苷酸化的RNA转录物424结合,用于进行标记或条形码化(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进行标记或条形码化。条形码420可以包含分子标记(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,无论是与抗体关联的还是已与抗体解离的)和追踪转录物424的样品来源,以及位于每个条形码420的分子标记428/样品标记430区域侧翼的一个或更多个另外的序列(例如,共有序列,诸如衔接子序列432)用于后续反应。每个样品的条形码中的分子标记428的序列的组库可以足够大以对RNA转录物424或核酸靶进行随机标记。

[0211] 在框402处的cDNA合成以产生条形码化的cDNA分子434之后,可以使用末端转移酶对条形码化的cDNA分子434的3'末端(相当于标记的RNA转录物的5'末端)进行A加尾以在框406处产生每个具有3'多(dA)尾438的cDNA分子436c。可以启动cDNA分子436c与3'多(dA)尾438的分子内杂交(例如,使用加热和冷却循环,或通过稀释具有多(dA)尾438的条形码化的cDNA分子436c),使得新的3'多(dA)尾438与相同标记的cDNA分子的多(dT)尾422退火以在框410处产生具有发夹或茎环结构440的条形码化的cDNA分子。可以添加聚合酶(例如,Klenow酶)和dNTP以促进在框412处新的3'多(dA)尾438之外的3'延伸以复制条形码(例如,在具有茎环440的标记的cDNA分子的5'末端上的分子标记428)。在框414处,可以使用镜像衔接子432、432rc或含有衔接子432、432rc的序列(或子序列)的引物来进行全转录组扩增(WTA)。可以使用诸如标签片段化或随机引发的方法来产生带有测序衔接子(例如,P5 446和P7 448序列)的扩增子444的较小片段,用于框418处的测序(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)的测序衔接子可以被直接连接以产生用于测序的扩增子。

[0212] 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,方法包括:使样品中的核酸靶424的拷贝与多于一种寡核苷酸条形码420接触,其中多于一种寡核苷酸条形码420中的每一种包含分子标记序列428和能够与核酸靶424杂交的靶结合区(例如,多(dT)序列422),并且其中多于一种寡核苷酸条形码420中的至少10种包含不同的分子标记序列428;在框402处,延伸与寡核苷酸条形码420杂交的核酸靶424的拷贝以产生各自包含与核酸靶424的至少一部分互补的序列450c的多于一种核酸分子434;在框404处,扩增多于一种条形码化的核酸分子434以产生多于一种扩增的条形码化的核酸分子

436;在框406处,使包含靶结合区422的互补体438的寡核苷酸附接到多于一种扩增的条形码化的核酸分子436以产生各自包含靶结合区422和靶结合区的互补体438的多于一种条形码化的核酸分子436c;在框410处,使多于一种条形码化的核酸分子436c中的每一种内的靶结合区422和靶结合区422的互补体438杂交以形成茎环440;在框412处,延伸各自具有茎环440的多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记的互补体428rc的多于一种延伸的条形码化的核酸分子442;在框414处,扩增多于一种延伸的条形码化的核酸分子442以产生各自包含分子标记的互补体428rc的多于一种单标记的核酸分子444c;以及基于与多于一种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的数量来确定样品中核酸靶的数量。

[0213] 在一些实施方案中,在延伸具有茎环440的多于一种条形码化的核酸分子的3'末端之后,使分子标记428与分子标记的互补体428rc杂交。方法可以包括在扩增多于一种延伸的条形码化的核酸分子442以产生多于一种单标记的核酸分子444c(其可以是扩增子444c的一部分)之前使多于一种延伸的条形码化的核酸分子442变性。使样品中的核酸靶424的拷贝接触可以包括使多于一种核酸靶424的拷贝与多于一种寡核苷酸条形码420接触。延伸核酸靶424的拷贝可以包括延伸与寡核苷酸条形码420杂交的多于一种核酸靶424的拷贝以产生各自包含与多于一种核酸靶424中的一种的至少一部分互补的序列450c的多于一种条形码化的核酸分子436c。确定核酸靶424的数量可以包括基于具有与包含多于一种核酸靶424中的每一种的序列452c的多于一种单标记的核酸分子444c中的单标记的核酸分子关联的不同序列的分子标记的互补体428rc的数量来确定样品中多于一种核酸靶424中的每一种的数量。多于一种核酸靶中的每一种的序列452c可以包括多于一种核酸靶424中的每一种的子序列(包括互补体或反向互补体)。

[0214] 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,方法包括:使用多于一种寡核苷酸条形码420对样品中的核酸靶424的拷贝进行条形码化402以产生各自包含核酸靶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)的多于一种条形码化的核酸分子434,并且其中多于一种寡核苷酸条形码420中的至少10种包含不同的分子标记序列428;使包含靶结合区422的互补体438的寡核苷酸附接406到多于一种条形码化的核酸分子434以产生各自包含靶结合区422和靶结合区422的互补体438的多于一种条形码化的核酸分子436;使多于一种条形码化的核酸分子436c中的每一种内的靶结合区422和靶结合区的互补体438进行杂交410以形成茎环440;延伸412多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记的互补体428rc的多于一种延伸的条形码化的核酸分子442;以及基于与多于一种延伸的条形码化的核酸分子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。

[0215] 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,方法包括:使用多于一种寡核苷酸条形码420对样品中的核酸靶424的拷贝进行条形码化402以产生各自包含核酸靶424的序列450c、分子标记428和靶结合区422的多于一种条形码化的核酸分子434,并且其中多于一种寡核苷酸条形码420中的至少10种包含不同的分子标记序列428;使包含靶结合区422的互补体438的寡核苷酸附接到多于一种条形码化的核酸分子434以产生各自包含靶结合区422和靶结合区422的互补体438的多于一种条

形码化的核酸分子436c;使多于一种条形码化的核酸分子436c中的每一种内的靶结合区422和靶结合区422的互补体438杂交410以形成茎环440;以及延伸412多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记428的互补体428rc的多于一种延伸的条形码化的核酸分子442。在一些实施方案中,方法包括:基于与多于一种延伸的条形码化的核酸分子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来确定样品中核酸靶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或其互补体428rc中的一种或两者来确定。

[0216] 在一些实施方案中,方法包括:对多于一种靶424的拷贝进行条形码化402包括:使核酸靶424的拷贝与多于一种寡核苷酸条形码420接触,其中多于一种寡核苷酸条形码420中的每一种包括能够与核酸靶424杂交的靶结合区422;以及延伸402与寡核苷酸条形码420杂交的核酸靶424的拷贝以产生多于一种条形码化的核酸分子434。

[0217] 在一些实施方案中,方法包括:扩增404多于一种条形码化的核酸分子434以产生多于一种扩增的条形码化的核酸分子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种扩增的条形码化的核酸分子以产生各自包含靶结合区422和靶结合区的互补体438的多于一种条形码化的核酸分子436r。

[0218] 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多于一种延伸的条形码化的核酸分子442以产生各自包含分子标记428的互补体428rc的多于一种单标记的核酸分子444c。当含有单标记的核酸分子的扩增子444变性时,可以产生单标记的核酸分子444c。确定样品中核酸靶424的数量可以包括:基于与多于一种单标记的核酸分子444c关联的具有不同序列的分子标记428的互补体428rc的数量来确定样品中核酸靶424的数量。

[0219] 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多于一种延伸的条形码化的核酸分子442以产生多于一种延伸的条形码化的核酸分子的拷贝444c。确定样品中核酸靶424的数量包括:基于与多于一种延伸的条形码化的核酸分子的拷贝444c关联的具有不同序列的分子标记428的互补体428rc的数量来确定样品中核酸靶424的数量。当含有多于一种延伸的条形码化的核酸分子的扩增子444变性时,可以形成多于一种延伸的条形码化的核酸分子的拷贝444c。

[0220] 在一些实施方案中,多于一种条形码化的核酸分子中核酸靶的序列包括核酸靶的子序列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种条形码化的核酸分子434。

[0221] 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种腺苷单磷酸添加到多于一种条形码化的核酸分子434。

[0222] 在一些实施方案中,延伸与寡核苷酸条形码420杂交的核酸靶424的拷贝可以包括逆转录与寡核苷酸条形码420杂交的核酸靶424的拷贝以产生多于一种条形码化的互补脱氧核糖核酸(cDNA)分子434。延伸与寡核苷酸条形码420杂交的核酸靶424的拷贝可以包括使用缺乏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聚合酶延伸

402与寡核苷酸条形码420杂交的核酸靶424的拷贝。DNA聚合酶可以包括Klenow片段。

[0223] 在一些实施方案中,方法包括:获得多于一种延伸的条形码化的核酸分子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)附接到多于一种延伸的条形码化的核酸分子442。

[0224] 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。

[0225] 在一些实施方案中,多于一种条形码化的核酸分子434可以包括条形码化的脱氧核糖核酸(DNA)分子。条形码化的核酸分子434可以包括条形码化的核糖核酸(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。

[0226] 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。与核酸靶关联的细胞结合试剂(例如,抗体寡核苷酸,诸如样品索引寡核苷酸)已在US2018/0088112和2018年3月27日提交的美国申请第15/937,713号中描述;这些申请中的每一项的内容通过引用以其整体并入本文。在一些实施方案中,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、甲基组学、转录组学和蛋白组学。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。

[0227] 在一些实施方案中,多于一种寡核苷酸条形码420的每种分子标记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。多于一种寡核苷酸条形码420的每种样品标记可以包含至少6个核苷酸。寡核苷酸条形码420可以包含相同的细胞标记。多于一种寡核苷酸条形码420的每种细胞标记可以包含至少6个核苷酸。

[0228] 在一些实施方案中,当使多于一种条形码化的核酸分子中的每一种内的靶结合区和靶结合区的互补体杂交410以形成茎环时,多于一种条形码化的核酸分子436c中的至少一种与固体支持物关联。当使多于一种条形码化的核酸分子436c中的每一种内的靶结合区422和靶结合区422的互补体438杂交410以形成茎环440时,多于一种条形码化的核酸分子436c中的至少一种可以从固体支持物解离。当使多于一种条形码化的核酸分子436c中的每一种内的靶结合区422和靶结合区的互补体438杂交410以形成茎环440时,多于一种条形码化的核酸分子436c中的至少一种可以与固体支持物关联。

[0229] 在一些实施方案中,当延伸412多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记的互补体428rc的多于一种延伸的条形码化的核酸分子442时,多于一种条形码化的核酸分子中的至少一种与固体支持物关联。当延伸412多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记的互补体428rc的多于一种延伸的条形码化的核酸分子442时,多于一种条形码化的核酸分子中的至少一种可以从固体支持物解离。当延伸412多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记的互补体428rc的多于一种延伸的条形码化的核酸分子442时,多于一种条形码化的核酸分子436c中的至少一种可以与固体支持物关联。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上平坦的表面(例如,载玻片,诸如显微镜载玻片或盖玻片)。

[0230] 在一些实施方案中,当使多于一种条形码化的核酸分子436c中的每一种内的靶结合区422和靶结合区422的互补体438杂交410以形成茎环440时,多于一种条形码化的核酸分子436c中的至少一种处于溶液中。例如,当溶液中多于一种条形码化的核酸分子436c的浓度足够低时,可以发生这样的分子内杂交。当延伸412多于一种条形码化的核酸分子的3'末端以延伸茎环440来产生各自包含分子标记428和分子标记的互补体428rc的多于一种延伸的条形码化的核酸分子442时,多于一种条形码化的核酸分子中的至少一种可以处于溶液中。

[0231] 在一些实施方案中,样品包括单细胞,方法包括将包含多于一种寡核苷酸条形码420的合成颗粒454与样品中的单细胞关联。方法可以包括:在将合成颗粒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相同的孔中。合成颗粒和单细胞可以在相同的液滴中。

[0232] 在一些实施方案中,多于一种寡核苷酸条形码420中的至少一种可以被固定在合成颗粒454上。多于一种寡核苷酸条形码420中的至少一种可以被部分地固定在合成颗粒454上。多于一种寡核苷酸条形码420中的至少一种可以被包封在合成颗粒454中。多于一种寡核苷酸条形码420中的至少一种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。多于一种寡核苷酸条形码420中的每一种可以包含接头官能团。合成颗粒454可以包括固体支持物官能团。支持物官能团和接头官能团可以彼此关联。接头官能团和支持物官能团可以单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

[0233] 用于对核酸靶5'末端进行条形码化的试剂盒

[0234] 本文的公开内容包括用于将寡核苷酸条形码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的数量的试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码420,其中多于一种寡核苷酸条形码420中的每一种包含分子标记428和靶结合区(例如,多(dT)序列422),并且其中多于一种寡核苷酸条形码420中的至少10种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;以及缺乏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。DNA聚合酶可以包括Klenow片段。试剂盒可以包含缓冲液和/或筒。试剂盒可以包含一种或更多种用于逆转录反应的试剂。试剂盒可以包含一种或更多种用于扩增反应的试剂。

[0235] 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。多于一种寡核苷酸条形码的每种样品标记和/或细胞标记可以包含至少6个核苷酸。多于一种寡核苷酸条形码的每种分子标记可以包含至少6个核苷酸。

[0236] 在一些实施方案中,多于一种寡核苷酸条形码420中的至少一种被固定在合成颗粒454上。多于一种寡核苷酸条形码420中的至少一种可以被部分地固定在合成颗粒454上。多于一种寡核苷酸条形码420中的至少一种可以被包封在合成颗粒454中。多于一种寡核苷酸条形码420中的至少一种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。多于一种寡核苷酸条形码中的每一种可以包含接头官能团。合成颗粒454可以包括固体支持物官能团。支持物官能团和接头官能团可以彼此关联。接头官能团和支持物官能团可以单独地选自以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

[0237] 全长表达谱分析

[0238] 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,这可能限制了剪接变体、选择性转录起始位点和由于重排造成的高度可变基因座诸如T细胞和B细胞受体和抗体的VDJ连接区的分析。使用5'条形码化和/或3'条形码化确定核酸靶(例如,免疫受体的V(D)J区)序列的方法在2019年9月30日提交的美国专利申请第16/588,405号中描述;其内容通过引用以其整体并入本文。图9A是免疫受体mRNA的5'表达谱分析的非限制性示例性示意图。虽然对免疫受体V(D)J区的5'表达谱分析可以改进TCR/BCR的覆盖率,并且可以获得CDR3序列信息,但这种方法可能会遗漏V区(在图9A中用虚线框指示)。对于T细胞和B细胞两者,目前可用的基于C引发的方法可以读取V(D)J但遗漏上游V区。因此,目前可用的方法可能限制获得全长核酸靶(例如,含V(D)J的转录物)信息的能力。本领域中的特定问题是,我们需要知道作为更长的读段的VDJ序列,因为存在由于可能的大量重组事件造成的许多VDJ。需要对序列(例如,含V(D)J的转录物)计数和鉴定所述序列(特别是全长序列鉴定)的方法。

[0239] 在一些实施方案中,提供了使用随机引发方法获得全长V(D)J信息(例如,通过Rhapsody系统上的Illumina测序)的方法。T细胞受体和B细胞受体包含V区段、D区段(仅对于TCR β 链和BCR重链)、J区段以及mRNA3'引发末端的恒定区。CDR3由V(D)J连接区构成,含有大部分组库多样性并且足够短以在Illumina短读段平台上进行测序。然而,全长V区段信息也是有用的,并且没有长读段测序技术而不能容易地获得,因为Illumina的短读段能力限制了获得全长V(D)J信息的能力。本文提供的方法可以采用随机引发和延伸来产生包含全长V(D)J扩增子以及仅含有部分V区段序列的较短扩增子两者的文库,这将使用户能够从单个文库获得CDR3信息以及全长V区段序列两者,并且使得能够进行与Illumina测序仪兼容的测序运行。因此,本文提供的方法的一些实施方案产生全长免疫受体mRNA序列。

[0240] 目前可用的5'VDJ测定可以使用片段化方法获得全长信息,这在Rhapsody上也是可能的,但需要更高成本的酶和试剂。此外,与本文提供的随机引发方法(仅涉及随机引发和延伸)相比,这种基于片段化的方法涉及更多的酶促步骤(例如,片段化、末端修复、加A

尾、连接)。片段化还可使得使用酶来添加引物位点成为必要,而本文提供的基于随机引发的方法的一些实施方案不需要连接已知的一个或更多个引物位点。

[0241] 图7是进行全长表达谱分析的非限制性示例性工作流程的示意图。方法可以包括对富集TCR/BCR的PCR产物进行随机引发。富集TCR/BCR的PCR产物可以是本文提供的延伸的条形码化的核酸分子的扩增产物(例如,第一多于一种条形码化的扩增子)。富集TCR/BCR的PCR产物可以是使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子的结果。随机引发的进行可以产生较短的产物。有义链引发和反义链引发均可以产生随机引发延伸产物。方法可以包括进行PCR3。对含有细胞条形码和UMI的有义链产物的PCR3扩增可以产生测序文库。但是,没有对不含细胞条形码和UMI的反义链产物的扩增。取决于随机引物的结合位点,可以产生一系列的延伸产物。图8描绘了根据本文提供的方法产生的非限制性示例性生物分析仪迹线。因此,本文提供了获得V区较长读段的手段(以便鉴定特定样品的VDJ)。Illumina从C区读取,并且正常读段长度为约150bp,其产生的测序数据一直到V的末端(如果从本文提供的延伸的条形码化的核酸分子起始)。PCR3引物可以经由突出端添加测序衔接子(例如,P5和P7)和样品索引(例如,i5、i7)。方法可以包括测序和生物信息学重建全长V(D)J(例如,如图9B中描绘的对齐多于一个测序读段)。在一些实施方案中,本文提供的方法既鉴定核酸靶(例如,含V(D)J的转录物),又对所述核酸靶的拷贝数计数。

[0242] 在一些实施方案中,提供了用于获得全长核酸靶(例如,转录物)信息的组合物、方法、系统和试剂盒。在一些实施方案中,根据本文提供的方法产生的延伸的条形码化的核酸分子被用作进行随机引发和延伸的模板。在一些实施方案中,扩增延伸的条形码化的核酸分子(例如,用靶特异性引物和包含第一通用序列的引物)以产生第一多于一种条形码化的扩增子。在一些实施方案中,第一多于一种条形码化的扩增子被用于进行随机引发和延伸(例如,使用包含第二通用序列或其互补体的随机物)以产生多于一种延伸产物的模板。随机引物可以与沿所有转录物的编码序列的不同位置结合,并延伸以产生多于一种延伸产物(例如,线性扩增产物)。延伸产物可以包括根据随机引物的结合位点而具有不同长度的cDNA。延伸产物可以用测序文库扩增引物扩增以产生第二多于一种条形码化的扩增子。测序文库扩增可以包括使用文库正向引物和文库反向引物,经由突出端添加测序衔接子和/或文库索引。文库扩增可以经由文库正向引物和文库反向引物中的突出端添加测序衔接子(例如P5和P7序列)和样品索引(例如i5、i7)。方法可以包括获得第二多于一种条形码化的扩增子(或其产物)的序列信息。可以对第二多于一种条形码化的扩增子进行测序并对其进行本公开内容的下游方法。用以产生150bp x 2测序读段的成对末端测序可以揭示读段1上的细胞标记、独特分子索引、多(A)尾和/或基因(或基因的部分序列),读段2上的基因(或基因的部分序列)和/或多(A)尾,以及索引1读段上的样品索引。获得第二多于一种条形码化的扩增子(或其产物)的序列信息可以包括获得包括第二多于一种条形码化的扩增子(或其产物)的多于一个测序读段的测序数据。方法可以包括通过对齐多于一个测序读段中的每一个以产生核酸靶的全长序列。

[0243] 本公开内容的方法可用于鉴定B细胞受体(BCR)、T细胞受体(TCR)和抗体的VDJ区。VDJ重组又称为体细胞重组,是免疫系统产生免疫球蛋白(Ig)(例如,BCR)和T细胞受体(TCR)的早期阶段中的遗传重组机制。VDJ重组可以几乎随机地组合可变(V)基因区段、多样

(D) 基因区段和连接(J) 基因区段。由于其在选择不同基因时的随机性,它能够多样地编码蛋白以匹配来自细菌、病毒、寄生虫、功能失调的细胞(诸如肿瘤细胞)和花粉的抗原。

[0244] VDJ区可以包含3Mb的大基因座,该3Mb的大基因座包含可变(V) 基因、多样性(D) 基因和连接(J) 基因。这些都是可以参与VDJ重组的区段。可以存在不进行VDJ重组的恒定基因。该基因座的VDJ重组中的第一个事件可以是,D基因中的一个重排到J基因中的一个。其后,V基因中的一个可以附加到该DJ重排以形成功能性的VDJ重排的基因,该功能性的VDJ重排的基因然后编码重链蛋白的可变区段。这两个步骤都可以由可以使间插DNA缺失的重组酶类催化。

[0245] 该重组过程在祖B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细胞可能只产生一种抗体(例如,BCR)。这种特异性可以通过等位基因排斥来实现,使得一个等位基因信号进行功能性重排以阻止第二个等位基因的进一步重组。

[0246] 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、髓样祖细胞、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、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、树突状细胞和/或巨噬细胞,或它们的任何组合。

[0247] T细胞可以是T细胞克隆,T细胞克隆可以指源自单个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,T细胞系可以包括T细胞克隆和具有不同TCR的T细胞的混合群体,所有这些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、脾组织和肿瘤。T细胞可以诸如使用Ficoll分离从采集自受试者的单位血液获得。个体的循环血液中的细胞可以通过单采血液成分术(apheresis)或白细胞单采术(leukapheresis)获得。单采血液成分术产品可以包括淋巴细胞(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。细胞可以被洗涤并在培养基中重悬以分离感兴趣的细胞。

[0248] T细胞可以通过裂解红细胞和耗尽单核细胞(例如,通过PERCOLL™梯度离心)从外周血淋巴细胞中分离。可以通过正选择或负选择技术进一步分离特定的T细胞亚群,诸如CD28+T细胞、CD4+T细胞、CD45RA+T细胞、和CD45RO+T细胞。例如,T细胞可以通过与抗CD3/抗CD28(即,3×28)-缀合珠诸如DYNABEADS®M-450 CD3/CD28 T或XCYTE DYNABEADS™孵育足以对于合意的T细胞进行正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,肿瘤特异性的)。

[0249] 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、来自淋巴结的活化的B细胞、类淋巴母细胞、静息B细胞或赘生性B细胞,例如来自淋巴瘤。APC可指在其表面表达至少一种BCRC蛋白的B细胞或滤泡树突状细胞。

[0250] 本公开内容的方法可以用于追踪单个T细胞的分子表型。不同亚型的T细胞可以通过不同分子标志物的表达来区分。T细胞表达来自不同的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链的异源二聚体组成;每一个功能链都可能是T细胞发育过程中体细胞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可以用来定义个体T细胞的身份,允许在免疫应答过程中对T细胞克隆扩增的谱系进行追踪。本公开内容的免疫学方法可以以多种方式使用,包括但不限于,鉴定单个T细胞中独特的TCRα链和TCRβ链配对,在单细胞水平上定量TCR和标志物表达,鉴定个体中TCR多

样性,表征不同T细胞群体中表达的TCR组库,确定TCR的 α 链等位基因和 β 链等位基因的功能性,以及鉴定免疫应答过程中T细胞的克隆扩增。

[0251] T细胞受体链配对

[0252] T细胞受体 (TCR) 是存在于T淋巴细胞表面的识别分子。在T细胞表面发现的T细胞受体可以由两个糖蛋白亚基组成,这两个亚基被称为 α 链和 β 链。两条链可以包括约40kDa的分子量并具有可变结构域和恒定结构域。编码 α 链和 β 链的基因可以在V、D和J区的文库中被组织,基因从V、D和J区通过遗传重排形成。TCR可以识别由抗原呈递细胞呈递的抗原,该抗原是与由组织相容性基因编码的特定自身分子组成的复合体的一部分。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体由MHC/肽配体组成。

[0253] 在一些实施方案中,本公开内容的方法、设备和系统可用于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 链和 β 链进行测序,使 α 链和 β 链进行配对,和/或确定T细胞受体 α 链的功能性拷贝。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。珠可以包含随机标记,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TCR α 分子和TCR β 分子可以经历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TCR α 链和 β 链可以被认为来自相同的单细胞,从而使TCR的 α 链和 β 链进行配对。

[0254] 抗体组库中的重链和轻链配对

[0255] 本公开内容的方法、装置和系统可以用于使BCR受体和抗体的重链和轻链进行配对。本公开内容的方法允许确定单个有机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)。这两种免疫受体类型都可以包含两条多肽链。免疫球蛋白可以包含可变重 (VH) 链和可变轻 (VL) 链。可以有两种TCR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含恒定区和可变区。可变区可以由B细胞或T细胞染色体上基因片段的重组和末端连接重排导致。在B细胞中,可变区的另外的多样化可以通过体细胞高频突变发生。

[0256] 免疫系统有大量的受体,并且由淋巴细胞表达的任何给定受体对都可以由一对分别的、独特的转录物编码。知晓单细胞中表达的免疫受体链对的序列可以用来确定给定个体或细胞群体的免疫组库 (immune repertoire)。

[0257] 在一些实施方案中,本公开内容的方法、装置和系统可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(例如,在B细胞中的)抗体重链和轻链进行测序,和/或对重链和轻链进行配对。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。

[0258] 珠可以包含随机标记,随机标记可以结合到抗体(例如,B细胞中的)的重链和/或轻链内的特定位置。与固体支持物关联的重链分子和轻链分子可以经历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相同的单细胞,从而使抗体的重链和轻链进行配对。

[0259] 全长表达谱分析的方法

[0260] 在一些实施方案中,提供了标记样品中的核酸靶的方法。在一些实施方案中,方法

包括：使核酸靶的拷贝与多于一种寡核苷酸条形码接触，其中每种寡核苷酸条形码包含第一通用序列、分子标记（例如，第一分子标记）和能够与核酸靶杂交的靶结合区；产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子；使每种条形码化的核酸分子的靶结合区的互补体与以下中的一种或更多种的靶结合区杂交：(i) 多于一种寡核苷酸条形码中的寡核苷酸条形码，(ii) 条形码化的核酸分子自身，以及(iii) 多于一种条形码化的核酸分子中的不同条形码化的核酸分子；延伸多于一种条形码化的核酸分子的3' 末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子；使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子，由此产生包含核酸靶的序列或其一部分的第一多于一种条形码化的扩增子；使随机引物与第一多于一种条形码化的扩增子杂交并延伸随机引物以产生多于一种延伸产物，其中随机引物包含第二通用序列或其互补体；以及使用能够与第一通用序列或其互补体和第二通用序列或其互补体杂交的引物来扩增多于一种延伸产物，从而产生第二多于一种条形码化的扩增子。在一些实施方案中，该方法包括：基于与第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来确定样品中的核酸靶的拷贝数。在一些实施方案中，第一通用序列和第二通用序列是相同的。在一些实施方案中，第一通用序列和第二通用序列是不同的。

[0261] 在一些实施方案中，提供了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种寡核苷酸条形码接触，其中每种寡核苷酸条形码包含第一通用序列、分子标记（例如，第一分子标记）和能够与核酸靶杂交的靶结合区；产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子；使每种条形码化的核酸分子的靶结合区的互补体与以下中的一种或更多种的靶结合区杂交：(i) 多于一种寡核苷酸条形码中的寡核苷酸条形码，(ii) 条形码化的核酸分子自身，以及(iii) 多于一种条形码化的核酸分子中的不同条形码化的核酸分子；延伸多于一种条形码化的核酸分子的3' 末端以产生各自包含第一分子标记和第二分子标记的多于一种延伸的条形码化的核酸分子；使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物扩增多于一种延伸的条形码化的核酸分子，由此产生包含核酸靶的序列或其一部分的第一多于一种条形码化的扩增子；使随机引物与第一多于一种条形码化的扩增子杂交并延伸随机引物以产生多于一种延伸产物，其中随机引物包含第二通用序列或其互补体；使用能够与第一通用序列和第二通用序列或其互补体杂交的引物扩增多于一种延伸产物，从而产生第二多于一种条形码化的扩增子；以及基于与第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记、具有不同序列的第二分子标记或它们的组合的数量来确定样品中核酸靶的拷贝数。

[0262] 确定样品中核酸靶的拷贝数可以包括：(a) 基于与第二多于一种条形码化的扩增子或其产物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核酸靶的拷贝数，和/或(b) 基于与第二多于一种条形码化的扩增子或其产物关联的具有不同序列的第一分子标记的数量来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：在使每种条形码化的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种条形码化的核酸分子变性：(i) 多于一种寡核苷酸条形码中的寡核苷酸条形码，(ii) 条形码化的核酸

分子自身,和/或(iii)多于一种条形码化的核酸分子中的不同条形码化的核酸分子。

[0263] 方法可以包括在扩增多于一种延伸的条形码化的核酸分子之前使多于一种延伸的条形码化的核酸分子变性。多于一种核酸靶中的每一种的序列可以包括多于一种核酸靶中的每一种的子序列。多于一种条形码化的核酸分子中的核酸靶的序列可以包括核酸靶的子序列。在一些实施方案中,在延伸多于一种条形码化的核酸分子的3'末端之后,使第一分子标记与第二分子标记杂交。延伸的条形码化的核酸分子可以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。靶结合区可以包含基因特异性序列、多(dT)序列或两者。

[0264] 产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子可以包括:(i)使用多于一种寡核苷酸条形码对样品中的核酸靶的拷贝进行条形码化以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记和靶结合区的多于一种条形码化的核酸分子;以及(ii)使包含靶结合区的互补体的寡核苷酸附接到多于一种条形码化的核酸分子以产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子。

[0265] 产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子可以包括:(i)使用多于一种寡核苷酸条形码对样品中的核酸靶的拷贝进行条形码化以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记和靶结合区的多于一种条形码化的核酸分子;(ii)扩增多于一种条形码化的核酸分子以产生多于一种扩增的条形码化的核酸分子;以及(iii)使包含靶结合区的互补体的寡核苷酸附接到多于一种扩增的条形码化的核酸分子以产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子。

[0266] 对样品中核酸靶的拷贝进行条形码化可以包括延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码以产生多于一种条形码化的核酸分子。延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码可以包括逆转录与多于一种寡核苷酸条形码杂交的核酸靶的拷贝。附接包含靶结合区的互补体的寡核苷酸可以包括将包含靶结合区的互补体的寡核苷酸连接到多于一种条形码化的核酸分子和/或扩增的条形码化的核酸分子。在一些实施方案中,靶结合区包含多(dT)序列,并且其中附接包含靶结合区的互补体的寡核苷酸包括使用末端脱氧核苷酸转移酶将多于一种腺苷单磷酸添加到多于一种条形码化的核酸分子和/或扩增的条形码化的核酸分子。产生各自包含靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子可以包括:在存在逆转录酶和包含靶结合区或其一部分的模板转换寡核苷酸的情况下,延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码,以产生各自包含与核酸靶的至少一部分互补的序列、第一分子标记、靶结合区和靶结合区的互补体的多于一种条形码化的核酸分子。在一些实施方案中,延伸多于一种条形码化的核酸分子的3'末端包括使用缺乏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聚合酶延伸多于一种条形码化的核酸分子的3'末端。在一些实施方案中,DNA聚合酶包括Klenow片段。

[0267] 在不同实施方式中,随机引发和延伸的循环的数量可以不同。在一些实施方案中,随机引发和延伸的循环数可以包括以下,或包括约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90

个、100个随机引发和延伸的循环,或在这些值中的任何两个值之间的数字或范围的随机引发和延伸的循环。在一些实施方案中,随机引发和延伸的循环数可以包括至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个或100个随机引发和延伸的循环。延伸与第一多于一种条形码化的扩增子杂交的随机引物可以包括使用缺乏5'至3'核酸外切酶活性和3'至5'核酸外切酶活性中至少一个的DNA聚合酶(例如,Klenow片段)。在一些实施方案中,延伸酶是Klenow或Klenow exo⁻。

[0268] 随机引物可以包含随机核苷酸序列。随机核苷酸序列的长度可以为约4个至约30个核苷酸。在一些实施方案中,所述随机核苷酸序列的长度为6个或9个核苷酸。在不同实施方式中,随机核苷酸序列可以具有不同的长度。在一些实施方案中,随机引物内的核苷酸的随机序列的长度是以下,或是约以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个核苷酸,或在这些值中的任何两个值之间的数字或范围的核苷酸。在一些实施方案中,随机引物内的核苷酸的随机序列的长度是至少以下,或至多以下:1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个核苷酸。在不同的实施方式中,随机引物可以在随机引发步骤期间具有不同的浓度。在一些实施方案中,随机引物在随机引发期间的浓度是至少以下,或至多以下:1uM、2uM、3uM、4uM、5uM、6uM、7uM、8uM、9uM、10uM、11uM、12uM、13uM、14uM、15uM、16uM、17uM、18uM、19uM、20uM、25uM、30uM、35uM、40uM、45uM、50uM、60uM、70uM、80uM、90uM、100uM、110uM、120uM、128uM,或在这些值中的任何两个值之间的数字或范围的uM。

[0269] 使条形码化的核酸分子的靶结合区的互补体与条形码化的核酸分子自身的靶结合区杂交可以包括使条形码化的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。

[0270] 使条形码化的核酸分子的靶结合区的互补体与多于一种寡核苷酸条形码中的寡核苷酸条形码的靶结合区杂交可以包括使条形码化的核酸分子的靶结合区的互补体与多于一种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,方法包括延伸与条形码化的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'末端以产生各自包含第一分子标记的互补体和第二分子标记的多于一种延伸的条形码化的核酸分子。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。

[0271] 使条形码化的核酸分子的靶结合区的互补体与多于一种条形码化的核酸分子中的不同条形码化的核酸分子的靶结合区杂交可以包括使条形码化的核酸分子的靶结合区的互补体与多于一种条形码化的核酸分子中的不同条形码化的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。

[0272] 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)到条形码化的核酸分子的末端(例如,3'末端)。在一些实施方案中,方法包括寡核苷酸条形码(或其产物,诸如例如另一种条形码化的核酸分子,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,随后进行延伸以产生延伸的条形码化的核酸分子。延伸的条形码化的核酸分子可以在3'末端和5'末端两者上被条形码化。在一些实施方案中,条形码化的分子的分子内杂交在3'多(dT)捕获珠上与捕获mRNA转录物形成发夹环。mRNA分子可以经由多(A)尾与寡核苷酸条形码的靶结合区结合而被捕获到珠上。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)尾可以与在相同的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。延伸之后,mRNA分子可以在3'末端和5'末端两者上被条形码化。这允许产生可以在例如Illumina测序平台上进行测序的3'条形码化的和5'条形码化的转录物两者。获取条形码化的5'序列可以允许检测T细胞受体(TCR)和B细胞受体(BCR)的可变区,以及发生在转录物的5'末端的剪接变体和序列变异。

[0273] 图6A-图60示出使用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)的全长序列的非限制性示例性工作流程的示意图。BD[®]Rhapsody[™]珠是固体条形码化的珠,其通过广泛的物理和化学操作保持完整性。在珠上进行mRNA多(A)捕获后,可以进行逆转录和模板转换以在条形码化的cDNA的3'末端添加多(dA)尾。添加的多(dA)尾允许结合珠的cDNA与相同的珠上的条形码(例如,随机条形码)的寡聚(dT)区域自杂交,形成桥环结构。桥环的Klenow延伸可以产生新的条形码化的cDNA分子,该cDNA分子来自相同的mRNA转录物,具有与第一条条形码化的cDNA相反的方向,允许连接分子条形码的3'末端和5'末端。

[0274] 本文公开的方法可以允许基于3'和/或基于5'的全长序列确定。该方法可以实现为序列测定提供灵活性。在一些实施方案中,方法能够实现在Rhapsody[™]系统上对诸如小鼠样品和人类样品的样品的T细胞和B细胞两者进行免疫组库谱分析,除了所使用的引物外不改变方案或产品配置。在一些实施方案中,可以进行基于3'和/或5'的V(D)J全长基因表达谱分析。在一些实施方案中,可以研究单细胞平台中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,高频突变)。

[0275] 本文描述的方法和系统可与使用与寡核苷酸(本文也称为Ab0或Ab0ligo)关联(例如,与寡核苷酸附接或缀合)的抗体的方法和系统一起使用。使用Ab0来确定单细胞中的蛋白表达谱和追踪样品来源的实施方案在美国专利申请第15/715,028号(作为美国专利申请公布第2018/0088112号公布)和美国专利申请第15/937,713号中描述;每项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对T细胞和B细胞的全长V(D)J谱分析、3'靶向扩增、5'靶向扩增、3'全转录组扩增(WTA)、5' WTA、用Ab0进行蛋白表达谱分析和/或单个实验中的样品多重化分析。

[0276] 模板转换反应

[0277] 图6A-图60示出使用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)的全长序列的非限制性示例性工作流程的示意图。条形码(例如,随机条形码、寡核苷酸条形码602)可以包含靶结合区(例如,多(dT) 604),该靶结合区可以经由多(dA)尾608与核酸靶(例如,多腺苷酸化的RNA转录物606)结合,或与其他核酸靶结合,用于标记或条形

码化(例如,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。多于一种条形码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。

[0278] 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过程中,当到达寡核苷酸条形码602的末端后,酶(例如,逆转录酶,诸如Moloney鼠白血病病毒(MMLV))的末端转移酶活性使一些另外的核苷酸(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可以作为模板转换寡核苷酸(例如,模板转换寡核苷酸)616的锚定位点起作用,它包括与加尾序列(例如,rGrGrG 618)互补的序列。模板转换寡核苷酸616可以包含至少一部分靶结合区604。在rGrGrG 618和附加的脱氧胞苷链段(stretch)612之间碱基配对后,酶“转换”模板链,从寡核苷酸条形码602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,条形码化的核酸分子620)含有模板转换寡核苷酸616的反向互补体序列并且因此可以包含靶结合区(例如,多(dA) 608)的互补体(例如,反向互补体)。条形码化的核酸分子620可以包括cDNA614c (RNA序列614r的反向互补序列)。该反应可以在存在一种或更多种配置为减少二级结构的添加剂(例如,乙二醇)的情况下进行。条形码化的核酸分子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,无论是与抗体关联的或还是已与抗体解离的)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),以及用于后续反应的位于每个条形码602的第一分子标记622/细胞标记624区域侧翼的一个或更多个另外的序列,诸如,例如,第一通用序列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足够大以对RNA转录物进行随机标记。在一些实施方案中,样品标记是分区标记。在一些实施方案中,样品标记是细胞标记。条形码化的核酸分子620可以经历变性步骤600b(例如,变性),从而产生单链条形码化的核酸分子621。

[0279] 在一些实施方案中,在延伸多于一种条形码化的核酸分子的3'末端之后,使第一分子标记与第二分子标记杂交。在一些实施方案中,延伸的条形码化的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,靶结合区包含基因特异性序列。在一些实施方案中,靶结合区包含多(dT)序列。

[0280] 术语“模板转换”可以指逆转录酶将初始核酸序列模板转换为新的核酸序列模板的3'末端的能力,该新的核酸序列模板的3'末端与由初始模板合成的核酸的3'末端具有很小或没有互补性。模板转换的实例是逆转录酶将初始核酸序列模板/引物底物转换为新的核酸序列模板的3'末端的能力,该新的核酸序列模板的3'末端与核酸引物链的3'末端几乎不互补或不互补。模板转换允许例如使用逆转录酶制备DNA拷贝,所述逆转录酶将初始核酸序列模板转换为新的核酸序列模板的3'末端,该新的核酸序列模板的3'末端与由初始模板

合成的DNA的3'末端具有很小或没有互补性,由此允许将衔接子序列与靶寡核苷酸序列直接相连(无连接)的连续产物DNA的合成。模板转换可以包括连接衔接子、均聚物加尾(例如,多腺苷酸化)、随机引物或聚合酶可与之关联的寡核苷酸。在上述任何实施方案中,模板转换可用于引入靶结合区或其互补体。

[0281] 在一些实施方案中,逆转录酶能够具有末端转移酶活性。在一些实施方案中,模板转换寡核苷酸包含一个或更多个3'核糖核苷酸。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Moloney鼠白血病病毒(MMLV)逆转录酶。

[0282] 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,多于一种条形码化的核酸分子可以包括条形码化的脱氧核糖核酸(DNA)分子和/或条形码化的核糖核酸(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,多于一种寡核苷酸条形码中的至少10种可以包含不同的分子标记序列。多于一种寡核苷酸条形码的每种分子标记可以包含至少6个核苷酸。

[0283] 在一些实施方案中,多于一种寡核苷酸条形码与固体支持物关联。与相同固体支持物关联的多于一种寡核苷酸条形码可以各自包含相同的样品标记。多于一种寡核苷酸条形码的每种样品标记可以包含至少6个核苷酸。多于一种寡核苷酸条形码可以各自包含细胞标记。多于一种寡核苷酸条形码的每种细胞标记可以包含至少6个核苷酸。与相同固体支持物关联的寡核苷酸条形码可以包含相同的细胞标记。与不同固体支持物关联的寡核苷酸条形码可以包含不同的细胞标记。多于一种延伸的条形码化的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。方法可以包括在存在乙二醇、聚乙二醇、1,2-丙二醇、二甲基亚砜(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情况下,延伸与核酸靶的拷贝杂交的多于一种寡核苷酸条形码。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。

[0284] 样品可以包括单细胞,并且方法可以包括将包含多于一种寡核苷酸条形码的合成颗粒与样品中的单细胞关联。方法可以包括在将合成颗粒与单细胞关联后裂解单细胞。裂解单细胞可以包括加热样品、使样品与去污剂接触、改变样品的pH,或它们的任何组合。在一些实施方案中,合成颗粒和单细胞在相同的孔中。在一些实施方案中,合成颗粒和单细胞在相同的液滴中。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被固定在合成颗粒上。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被部分地固定在合成颗粒上。多于一种寡核苷酸条形码中的至少一种可以被包封在合成颗粒中。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被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。多于一种寡核苷酸条形码中的每一种可以包含接头官能团,合成颗粒可以包含固体支持物官能团,和/或支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自由以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

[0285] 条形码化的核酸分子的分子内杂交

[0286] 在一些实施方案中,使条形码化的核酸分子的靶结合区的互补体与条形码化的核酸分子自身的靶结合区杂交包括使条形码化的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。

[0287] 工作流程可以包括使单链条形码化的核酸分子621进行分子内杂交,如在非限制性示例性图6B示意图中描绘的。工作流程可以包括使单链条形码化的核酸分子621内的靶结合区604和靶结合区的互补体608进行分子内杂交600c1形成茎环。工作流程可以包括延伸600c2单链条形码化的核酸分子621的茎环的3'末端以产生延伸的条形码化的核酸分子620c。延伸的条形码化的核酸分子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。工作流程可以包括使延伸的条形码化的核酸分子620c变性600c3以产生单链延伸的条形码化的核酸分子620cd。在一些实施方案中,分子间杂交600c1和/或延伸600c2在存在高盐缓冲液和/或PEG的情况下进行。在一些实施方案中,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)进行延伸。

[0288] 单链延伸的条形码化的核酸分子620cd可以包含靶核酸分子(例如,转录物)的5'末端和3'末端两者上的条形码(例如,细胞标记和分子标记),从而与在一端仅有一种条形码的靶核酸分子的涉及序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序的分析相比,实现靶核酸分子更广泛的分析。单链延伸的条形码化的核酸分子620cd可以用作一个或更多个延伸反应(例如,随机引发和延伸)和/或扩增反应(例如,PCR)的模板,诸如例如,图6C-图6E中描绘的非限制性示例性方案。一次或更多次扩增可以包括靶特异性(例如,基因特异性)cDNA扩增。例如,单链延伸的条形码化的核酸分子620cd可以经历采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650)的第一轮扩增("PCR1")600c4,从而产生各自包含分子标记(例如,第一分子标记)、细胞标记、第一通用序列和部分反义cDNA 614cas1的第一多于一种扩增子620cas1(其长度取决于cDNA 614c内的靶特异性引物650的结合位点)。PCR1 600c4可以包括1-30个循环(例如,15个循环)。

[0289] 工作流程可以包括随机引发和延伸600c5。随机引物670可以与第一多于一种条形

码化的扩增子620cas1杂交,并且随机引物670可以被延伸以产生多于一种延伸产物620e2c1和620e2c2。随机引物670可以包括突出端,突出端可以包含或者是例如第二通用序列638(或其互补体,例如反向互补体638rc)(例如,读段2序列、通用PCR手柄)。延伸产物620e2c1可以包含第一通用序列、第二通用序列、细胞标记和分子标记(例如,第一分子标记),或其互补体。延伸产物620e2c1可以包含,例如,部分cDNA614c2a、614c2b、614c2c和614c2d(每种延伸产物的长度取决于部分cDNA内的随机引物670的结合位点)。延伸产物620e2c2可以包含第二通用序列(或其互补体)。延伸产物620e2c2可以包含,例如,部分反义cDNA 614cas2e、614cas2f、614cas2g和614cas2h(每种延伸产物的长度取决于部分cDNA内的随机引物670的结合位点)。

[0290] 工作流程可以包括文库扩增(“文库PCR”)600c6。文库PCR 600c6可以包括用测序文库扩增引物656和658对延伸产物620e2c1进行文库扩增。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)。文库PCR 600c6可以经由测序文库扩增引物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PCR扩增子620cL(例如,第二多于一种条形码化的扩增子)可以包含一定范围的cDNA长度(例如,部分反义cDNA 614cas2a、614cas2b、614cas2c、614cas2d),该长度可以集体地包括核酸靶的mRNA序列的一些或全部。文库扩增子620cL可以被测序并经历本公开内容的下游方法。使用150bp x 2测序的测序600c7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。文库PCR 600c6可以包括1-30个循环(例如,15个循环)。方法可以包括通过如图9B中描绘的对齐多于一个读段1 910和读段2 920读段对核酸靶(例如,免疫受体mRNA 930)的全长序列进行生物信息重建900。

[0291] 在一些实施方案中,可以对免疫受体的V(D)J区进行基于3'和/或5'的全长表达谱分析。在一些实施方案中,可以研究单细胞平台中T细胞和/或B细胞的表型标志物和一种或更多种免疫受体V(D)J序列两者。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,高频突变)。在一些实施方案中,扩增延伸的条形码化的核酸分子620cd的3'区域和5'区域两者。在一些实施方案中,仅扩增延伸的条形码化的核酸分子620cd的5'区域。在一些实施方案中,仅扩增延伸的条形码化的核酸分子620cd的3'区域。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,延伸的条形码化的核酸分子620cd的3'和5'区域两者可以同时扩增(例如,多重PCR)。在一些实施方案中,工作流程包括采用一组靶特异性PCR1引物的多重PCR。在一些实施方案中,靶包括BCR、TCR和/或免疫相关转录物。

[0292] 条形码化的核酸分子与条形码化的核酸分子的分子间杂交

[0293] 在一些实施方案中,使条形码化的核酸分子的靶结合区的互补体与多于一种条形码化的核酸分子中的不同条形码化的核酸分子的靶结合区杂交包括使条形码化的核酸分子的靶结合区的互补体与多于一种条形码化的核酸分子中的不同条形码化的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。

[0294] 工作流程可以包括使单链条形码化的核酸分子621与独特条形码化的核酸分子628进行分子间杂交,如在非限制性示例性图6F-图6G示意图中描绘的。不同条形码化的核酸分子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

码化的核酸分子628的第二分子标记632的序列可以不同于单链条形码化的核酸分子621的第一分子标记622的序列(例如,不是互补体)。条形码化的核酸分子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的核酸分子621的靶结合区604、细胞标记624和/或第一通用序列626相同(或是其互补体)。在一些实施方案中,工作流程可以包括使单链条形码化的核酸分子621的靶结合区的互补体608与条形码化的核酸分子628的靶结合区604进行分子间杂交600d1。工作流程可以包括延伸600d2单链条形码化的核酸分子621的3'末端以产生延伸的条形码化的核酸分子620d。延伸的条形码化的核酸分子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。工作流程可以包括使延伸的条形码化的核酸分子620d变性600d3以产生单链延伸的条形码化的核酸分子620dd。在一些实施方案中,分子间杂交600d1和/或延伸600d2在存在高盐缓冲液和/或PEG的情况下进行。在一些实施方案中,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)进行延伸。

[0295] 单链延伸的条形码化的核酸分子620dd可以包含靶核酸分子(例如,转录物)的5'末端和3'末端两者上的条形码(例如,细胞标记和分子标记),从而与在一端仅有一种条形码的靶核酸分子的涉及序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序的分析相比,实现靶核酸分子更广泛的分析。单链延伸的条形码化的核酸分子620dd可以用作一个或更多个延伸反应(例如,随机引发和延伸)和/或扩增反应(例如,PCR)的模板,诸如例如,图6H-图6J中描绘的非限制性示例性方案。一次或更多次扩增可以包括靶特异性(例如,基因特异性)cDNA扩增。例如,单链延伸的条形码化的核酸分子620dd可以经历采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650)的第一轮扩增("PCR1")600d4,从而产生各自包含分子标记(例如,第二分子标记)、细胞标记、第一通用序列和部分反义cDNA614cas1的第一多于一种扩增子620das1(其长度取决于cDNA 614c内的靶特异性引物650的结合位点)。PCR1 600d4可以包括1-30个循环(例如,15个循环)。

[0296] 工作流程可以包括随机引发和延伸600d5。随机引物670可以与第一多于一种条形码化的扩增子620das1杂交,并且随机引物670可以被延伸以产生多于一种延伸产物620e2d1和620e2d2。随机引物670可以包括突出端,突出端可以包含或者是例如第二通用序列638(或其互补体,例如反向互补体638rc)(例如,读段2序列、通用PCR手柄)。延伸产物620e2d1可以包含第一通用序列、第二通用序列、细胞标记和分子标记(例如,第一分子标记),或其互补体。延伸产物620e2d1可以包括,例如,部分cDNA614c2a、614c2b、614c2c和614c2d(每种延伸产物的长度取决于部分cDNA内随机引物670的结合位点)。延伸产物620e2d2可以包含第二通用序列(或其互补体)。延伸产物620e2d2可以包含,例如,部分反义cDNA 614cas2e、614cas2f、614cas2g和614cas2h(每种延伸产物的长度取决于部分cDNA内的随机引物670的结合位点)。

[0297] 工作流程可以包括文库扩增("文库PCR")600d6。文库PCR 600d6可以包括用测序文库扩增引物656和658对延伸产物620e2d1进行文库扩增。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)。文库PCR 600d6可以由测序文库扩增引物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PCR扩增子620dL(例如,第二多于一种条形码化的扩增子)可以包含一定范围的cDNA长度(例如,部分反义cDNA 614cas2a、614cas2b、614cas2c、614cas2d),该长度可以集体地包括核酸靶的mRNA序列的一些或全部。文库扩增子620dL可以被测序并经历本公开内容的下游方法。使用150bp x 2测序的测序600d7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。文库PCR 600d6可以包括1-30个循环(例如,15个循环)。方法可以包括通过如图9B中描绘的对齐多于一个读段1 910和读段2 920读段对核酸靶(例如,免疫受体mRNA 930)的全长序列进行生物信息重建900。

[0298] 在一些实施方案中,可以对免疫受体的V(D)J区进行基于3'和/或5'的全长表达谱分析。在一些实施方案中,可以研究单细胞平台中T细胞和/或B细胞的表型标志物和一种或更多种免疫受体V(D)J序列两者。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,高频突变)。在一些实施方案中,扩增延伸的条形码化的核酸分子620dd的3'区域和5'区域两者。在一些实施方案中,仅扩增延伸的条形码化的核酸分子620dd的5'区域。在一些实施方案中,仅扩增延伸的条形码化的核酸分子620dd的3'区域。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,延伸的条形码化的核酸分子620dd的3'区域和5'区域两者可以同时扩增(例如,多重PCR)。在一些实施方案中,工作流程包括采用一组靶特异性PCR1引物的多重PCR。在一些实施方案中,靶包括BCR、TCR和/或免疫相关转录物。

[0299] 条形码化的核酸分子与寡核苷酸条形码的分子间杂交

[0300] 在一些实施方案中,使条形码化的核酸分子的靶结合区的互补体与多于一种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的核酸分子的靶结合区的互补体与多于一种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,方法包括延伸与条形码化的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'末端以产生各自包含第一分子标记的互补体和第二分子标记的多于一种延伸的条形码化的核酸分子。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。

[0301] 工作流程可以包括使单链条形码化的核酸分子621与独特的寡核苷酸条形码634进行分子间杂交,如在非限制性示例性图6K-图6L示意图中描绘的。独特的寡核苷酸条形码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的核酸分子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的核酸分子621的靶结合区604、细胞标记624和/或第一通用序列626相同(或是其互补体)。在一些实施方案中,工作流程可以包括使单链条形码化的核酸分子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进行分子间杂交600e1。工作流程可以包括延伸600e2单链条形码化的核酸分子621的3'末端以产生延伸的条形码化的核酸分子620e1。延伸的条形码化的核酸分子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。工作流程可以包括使延伸的条形码化的核酸分子620e1变性600e3以产生单链延伸的条形码化的核酸分子620e1d。工作流程可

以包括延伸600e2寡核苷酸条形码634的3'末端以产生延伸的条形码化的核酸分子620e2。延伸的条形码化的核酸分子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。工作流程可以包括使延伸的条形码化的核酸分子620e2变性600e3以产生单链延伸的条形码化的核酸分子620e2d。在一些实施方案中,分子间杂交600e1和/或延伸600e2在存在高盐缓冲液和/或PEG的情况下进行。在一些实施方案中,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)进行延伸。

[0302] 单链延伸的条形码化的核酸分子620e1d和单链延伸的条形码化的核酸分子620e2d可以包含靶核酸分子(例如,转录物)的5'末端和3'末端两者上的条形码(例如,细胞标记和分子标记),从而与在一端仅有一个条形码的靶核酸分子的涉及序列鉴定、转录物计数、选择性剪接分析、突变筛选和/或全长测序的分析相比,实现靶核酸分子更广泛的分析。单链延伸的条形码化的核酸分子620e1d和单链延伸的条形码化的核酸分子620e2d可以用作一个或多个延伸反应(例如,随机引发和延伸)和/或扩增反应(例如,PCR)的模板,诸如例如,图6M-图6O中描绘的非限制性示例性方案。一次或多次扩增可以包括靶特异性(例如,基因特异性)cDNA扩增。例如,单链延伸的条形码化的核酸分子620e2d可以经历采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650)的第一轮扩增("PCR1")600e4,从而产生各自包含分子标记(例如,第二分子标记)、细胞标记、第一通用序列和部分反义cDNA 614cas1的第一多于一种扩增子620eas1(其长度取决于cDNA 614cas内的靶特异性引物650的结合位点)。PCR1 600e4可以包括1-30个循环(例如,15个循环)。

[0303] 工作流程可以包括随机引发和延伸600e5。随机引物670可以与第一多于一种条形码化的扩增子杂交,并且随机引物670可以被延伸以产生多于一种延伸产物620e2e1和620e2e2。随机引物670可以包括突出端,突出端可以包含或者是例如第二通用序列638(或其互补体,例如反向互补体638rc)(例如,读段2序列、通用PCR手柄)。延伸产物620e2e1可以包含第一通用序列、第二通用序列、细胞标记和分子标记(例如,第一分子标记),或其互补体。延伸产物620e2e1可以包含,例如,部分cDNA 614c2a、614c2b、614c2c和614c2d(每种延伸产物的长度取决于部分cDNA内的随机引物670的结合位点)。延伸产物620e2e2可以包含第二通用序列(或其互补体)。延伸产物620e2e2可以包含,例如,部分反义cDNA 614cas2e、614cas2f、614cas2g和614cas2h(每种延伸产物的长度取决于部分cDNA内的随机引物670的结合位点)。

[0304] 工作流程可以包括文库扩增("文库PCR")600e6。文库PCR 600e6可以包括用测序文库扩增引物656和658对延伸产物620e2e1进行文库扩增。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)。文库PCR 600e6可以由测序文库扩增引物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PCR扩增子620e2L(例如,第二多于一种条形码化的扩增子)可以包含一定范围的cDNA长度(例如,部分反义cDNA 614cas2a、614cas2b、614cas2c、614cas2d),该长度可以集体地包括核酸靶的mRNA序列的一些或全部。文库扩增子620e2L可以被测序并经历本公开内容的下游方法。使用150bp x 2测序的测序600e7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。文库PCR600e6可以包括1-30个循环(例如,15个循环)。方法可以包括通过如图9B中描绘的对齐多于一个读段1 910和读段2 920读段对核酸靶(例如,免疫受体mRNA 930)的全长序列进行生物信息重建900。

[0305] 在一些实施方案中,可以对免疫受体的V(D)J区进行基于3'和/或5'的全长表达谱分析。在一些实施方案中,可以研究单细胞平台中T细胞和/或B细胞的表型标志物和一种或更多种免疫受体V(D)J序列两者。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,高频突变)。在一些实施方案中,扩增一种或更多种延伸的条形码化的核酸分子620e1d和/或620e2d的3'和5'区域两者。在一些实施方案中,仅扩增一种或更多种延伸的条形码化的核酸分子620e1d和/或620e2d的5'区域。在一些实施方案中,仅扩增一种或更多种延伸的条形码化的核酸分子620e1d和/或620e2d的3'区域。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,一种或更多种延伸的条形码化的核酸分子620e1d和/或620e2d的3'区域和5'区域两者可以同时扩增(例如,多重PCR)。在一些实施方案中,工作流程包括采用一组靶特异性PCR1引物的多重PCR。在一些实施方案中,靶包括BCR、TCR和/或免疫相关转录物。

[0306] 免疫组库谱分析

[0307] 在一些实施方案中,提供了用于对免疫受体的V(D)J区域进行全长表达谱分析的方法。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。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,每种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,多于一种延伸的条形码化的核酸分子包含第一通用序列和第一通用序列的互补体。扩增多于一种延伸的条形码化的核酸分子可以包括使用能够与核酸靶的序列杂交的靶特异性引物和包含第一通用序列的引物。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TCR α 链、TCR β 链、TCR γ 链、TCR δ 链或它们的任何组合。BCR可以包含BCR重链和/或BCR轻链。

[0308] 方法可以包括获得第二多于一种条形码化的扩增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第二多于一种条形码化的扩增子或其产物。扩增多于一种延伸产物可以包括将测序引物和/或测序衔接子、其互补序列和/或其部分的结合位点的序列添加到多于一种延伸产物。测序衔接子可以包括P5序列、P7序列、其互补序列或其一部分。测序引物可以包括读段1测序引物、读段2测序引物、其互补序列或其一部分。

[0309] 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。方法可以包括基于所获得的序列信息使单细胞的BCR轻链和BCR重链进行配对。样品可以包括多于一个单细胞,并且方法可以包括基于所获得的序列信息使至少50%的单细胞的BCR轻链和BCR重链进行配对。在一些实施方案中,其中BCR轻链和BCR重

链按照本文提供的方法配对的样品的单细胞百分比可以是以下,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,其中BCR轻链和BCR重链按照本文提供的方法配对的样品的单细胞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%。

[0310] 获得序列信息可以包括获得单细胞的TCR α 链和TCR β 链的序列信息。在一些实施方案中,TCR α 链和TCR β 链的序列信息可以包括TCR α 链和/或TCR β 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,方法可以包括基于所获得的序列信息使单细胞的TCR α 链和TCR β 链进行配对。在一些实施方案中,样品可以包括多于一个单细胞,并且方法可以包括基于所获得的序列信息使至少50%的单细胞的TCR α 链和TCR β 链进行配对。在一些实施方案中,其中TCR α 链和TCR β 链根据本文提供的方法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,其中TCR α 链和TCR β 链根据本文提供的方法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%。

[0311] 获得序列信息可以包括获得单细胞的TCR γ 链和TCR δ 链的序列信息。TCR γ 链和TCR δ 链的序列信息可以包括TCR γ 链和/或TCR δ 链的互补决定区1 (CDR1)、CDR2、CDR3或它们的任何组合的序列。方法可以包括基于所获得的序列信息使单细胞的TCR γ 链和TCR δ 链进行配对。样品可以包括多于一个单细胞,并且方法可以包括基于所获得的序列信息使至少50%的单细胞的TCR γ 链和TCR δ 链进行配对。在一些实施方案中,其中TCR δ 链和TCR γ 链根据本文提供的方法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,其中TCR δ 链和TCR γ 链根据本文提供的方法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%。

[0312] 获得第二多于一种条形码化的扩增子或其产物的序列信息可以包括:获得包含第二多于一种条形码化的扩增子或其产物的多于一个测序读段的测序数据,其中多于一个测序读段中的每一个包含(1)细胞标记序列、(2)分子标记序列和/或(3)核酸靶的子序列。在一些实施方案中,方法包括:对于指示样品中的单细胞的每个独特的细胞标记序列:使核酸靶的多于一个测序读段中的每一个对齐以产生核酸靶的对齐序列。如图9B中所描绘的,测序数据可以包括多于一个测序读段,诸如例如读段1 910和读段2 920读段。在一些实施方案中,作为随机引发步骤的结果,读段1 910和/或读段2 920读段可以集体地包含整个核酸靶(例如,免疫受体转录物)。方法可以包括通过对齐多于一个读段1 910和读段2 920对核酸靶(例如,免疫受体mRNA 930)的全长序列进行生物信息重建900。有利的是,本文提供的组合物和方法可以提供核酸靶(例如,含V(D)J的转录物)的鉴定和计数两者。

[0313] 在一些实施方案中,核酸靶的对齐序列包含核酸靶的cDNA序列的至少50%、核酸

靶的cDNA序列的至少70%、核酸靶的cDNA序列的至少90%或核酸靶的cDNA序列的全长。在一些实施方案中,核酸靶的对齐序列可以是核酸靶的cDNA序列的以下,或是约以下:1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、100%,或在这些值中的任何两个值之间的数字或范围。在一些实施方案中,核酸靶的对齐序列可以是核酸靶的cDNA序列的至少以下,或至多以下:1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。核酸靶可以是免疫受体。

[0314] 在一些实施方案中,核酸靶的对齐序列包含互补决定区1(CDR1)、互补决定区2(CDR2)、互补决定区3(CDR3)、可变区、可变区的全长或它们的组合。核酸靶的对齐序列可以包含可变区、多样性区、可变区多样性区的连接区和/或恒定区或它们的任何组合。在一些实施方案中,方法不包括片段化、标签片段化或两者。

[0315] 用于获得核酸靶全长信息的试剂盒

[0316] 本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;逆转录酶;包含靶结合区或其一部分的模板转换寡核苷酸;以及缺乏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。本文的公开内容包括试剂盒。在一些实施方案中,试剂盒包含:多于一种寡核苷酸条形码,其中多于一种寡核苷酸条形码中的每一种包含分子标记和靶结合区,并且其中多于一种寡核苷酸条形码中的至少10种包含不同的分子标记序列;末端脱氧核苷酸转移酶;以及缺乏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。

[0317] 试剂盒可以包含随机引物。随机引物可以包含随机核苷酸序列。随机核苷酸序列的长度可以为约4个至约30个核苷酸。在一些实施方案中,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Moloney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。

[0318] 在一些实施方案中,试剂盒包含缓冲液。在一些实施方案中,试剂盒包含筒。在一些实施方案中,试剂盒包含一种或更多种用于逆转录反应的试剂。在一些实施方案中,试剂盒包含一种或更多种用于扩增反应的试剂。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,多于一种寡核苷酸条形码的每种样品标记和/或细胞标记包含至少6个核苷酸。在一些实施方案中,多于一种寡核苷酸条形码的每种分子标记包含至少6个核苷酸。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被固定在合成颗粒上。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被部分地固定在合成颗粒上。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被包封在合成颗粒中。在一些实施方案中,多于一种寡核苷酸条形码中的至少一种被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,多于一种寡核苷酸条形码中的每一种包含接头官能团,合成颗粒包含固体支持物官能团,和/或支持物官能团和接头官能团彼此关联。在一些实施方案中,接头官能团和支持物官能团单独地选自以下组成的组:C6、生物素、链霉抗生物素蛋白、一种或更多种伯胺、一种或更多种醛、一种或更多种酮及它们的任何组合。

实施例

[0319] 本文描述的实施方案的一些方面在以下实施例中进一步详细公开,其并不是以任何方式意在限制本公开内容的范围。

[0320] 实施例1

[0321] 全长V(D)J序列的基于随机引发的确定

[0322] 本实施例展示了本文提供的核酸靶(例如,含V(D)J的转录物)的全长表达谱分析的基于随机引发的方法。如在TCR产物(图10A-图10B)和BCR产物(图10C-图10D)的生物分析仪迹线中所观察到的,与目前可用的方法(图10B和图10D)相比,用本文提供的基于随机引发的方法(图10A和图10C)获得更小的片段。接下来,进行测序以确定全长(图11B)和来自细胞(图11A)的VDJ序列的百分比。在与VDJ对齐的读段中,成功地组装成重叠群,并且该重叠群是有效的VDJ序列(具有CDR3),在图11A中示出的百分比与来自推定细胞的那些读段的数量相关。在推定的细胞中,对于被鉴定为优势的重叠群,它们之中全长的百分比在图11B中示出。全长通过该重叠群是否具有针对以下链的所有主要部分所鉴定的序列来确定:FR1、CDR1、FR2、CDR2、FR3、CDR3、FR4。此外,如图12A-图12D中示出的,在正确的细胞类型(B细胞中的BCR、T细胞中的TCR)中发现全长配对链,其中BCR配对(图12A-图12B)和TCR配对(图12C-图12D)通过优于目前可用的方法(图12B和图12D)的本文提供的基于随机引发的方法(图12A和图12B)获得。这些结果为本文提供的基于随机引发的方法和组合物提供了原理论

证。

[0323] 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真正范围和精神。

[0324] 本领域的技术人员将理解,对于本文公开的这个和其他过程和方法,在该过程和方法中执行的功能可以以不同的顺序实现。此外,所概述的步骤和操作仅作为示例提供,并且该步骤和操作中的一些可以是任选的,组合成较少的步骤和操作,或者扩展成另外的步骤和操作,而不偏离所公开的实施方案的本质。

[0325] 关于本文中基本上任何复数和/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。

[0326] 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解释为暗含通过不定冠词“一个(a)”或“一种(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为包含仅一个这样的陈述的实施方案,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。

[0327] 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,本领域技术人员将认识到,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。

[0328] 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涵盖任何和所有可能的它的子范围和子范围组合。任何列举的范围可以被容易地认为充分地描述了并且使得同一范围能够被分成至少相等的二分之一、三分之一、四分之一、五分之一、十分之一等。作为非限制性实例,本文讨论的每个范围可以被容易地分成下三分之一、中三分之一和上三分之一等。如本领域技术人员还将理解的,所有语言诸如“多达 (up to)”、“至少”等包括所述及的数字并且指随后可以被分成如以上讨论的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个单独的成员。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。

[0329] 从前述内容,应当理解,本文出于说明的目的已经描述了本公开内容的各种实施方案,并且可以在不脱离本公开内容的范围和精神的情况下进行各种修改。因此,本文公开的各种实施方案并不旨在进行限制,真正的范围和精神由以下权利要求来指示。

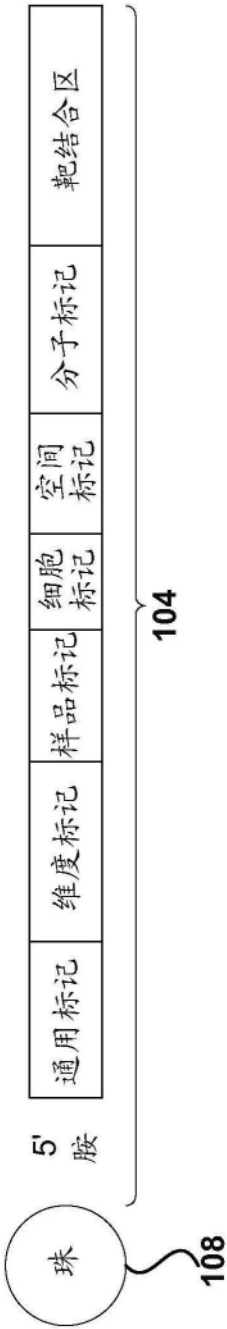


图1

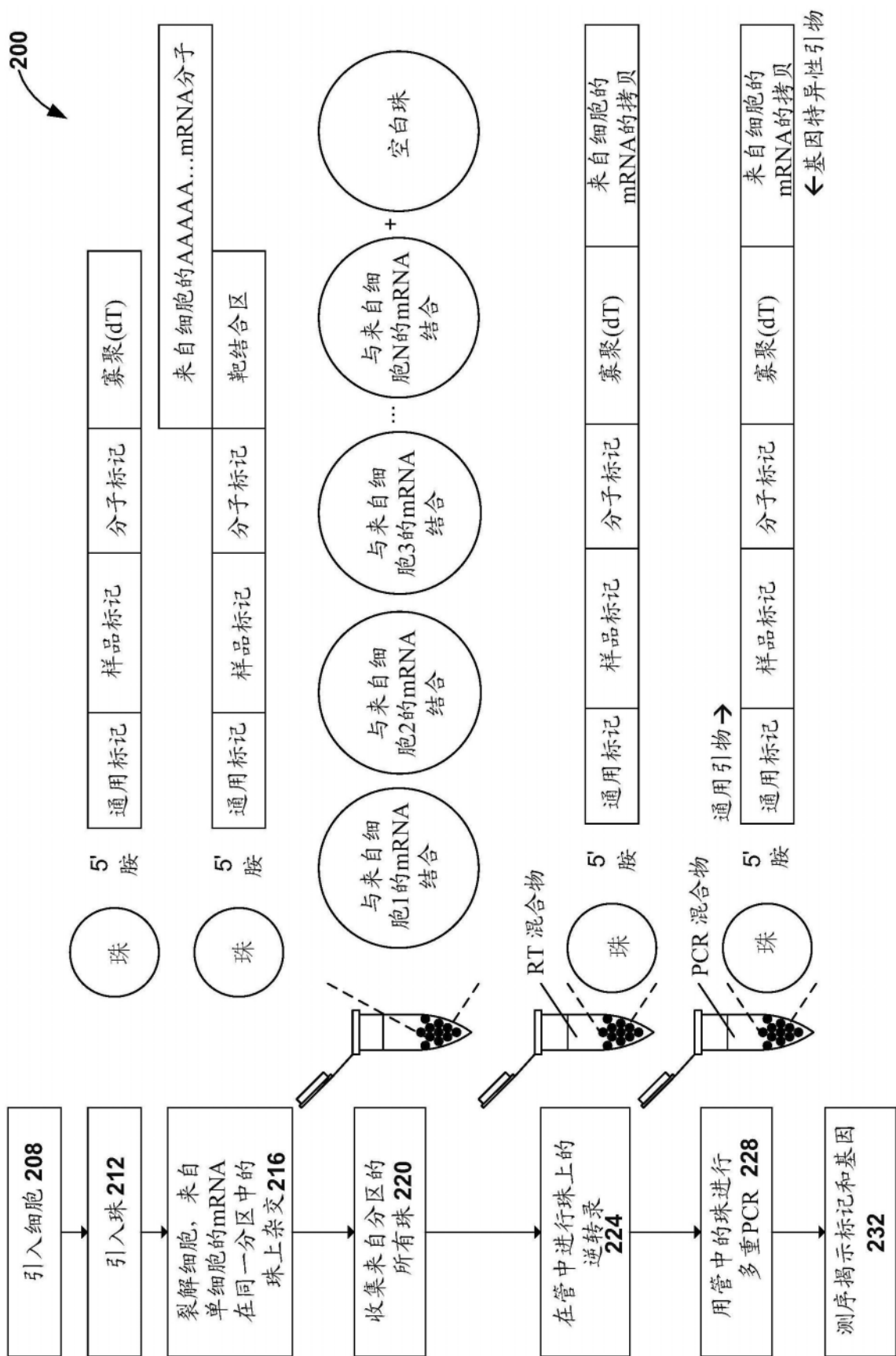
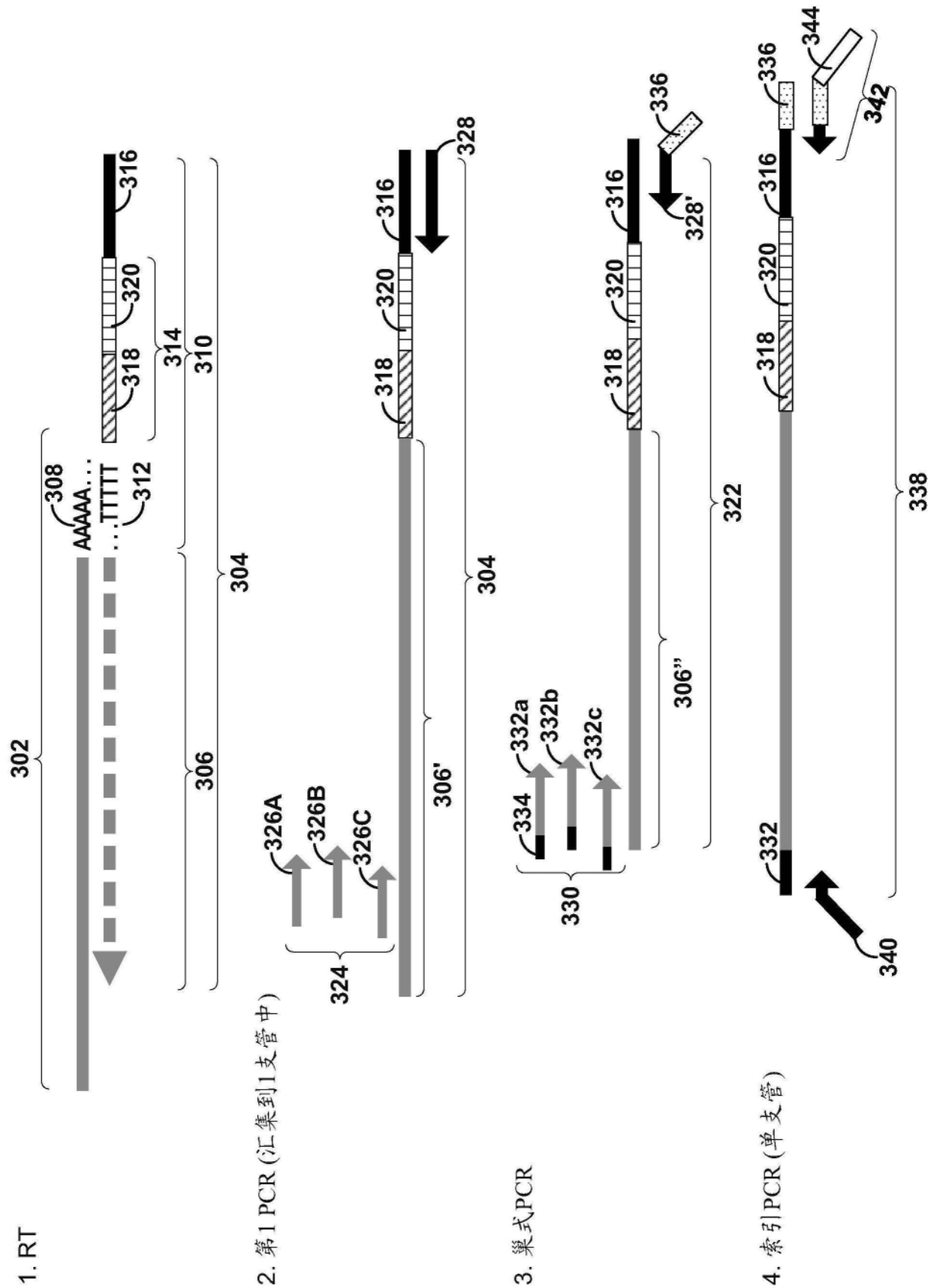


图2



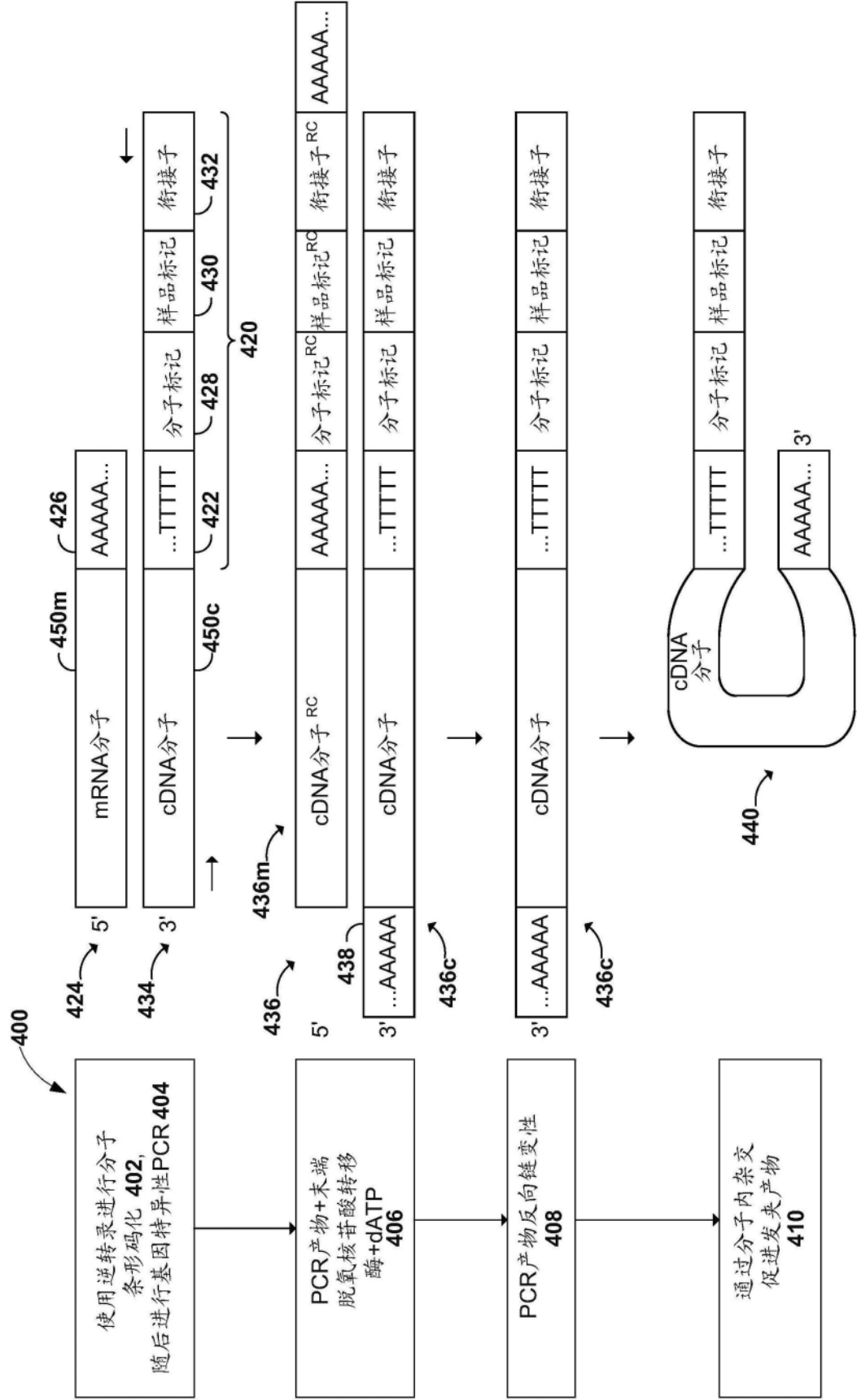


图4A

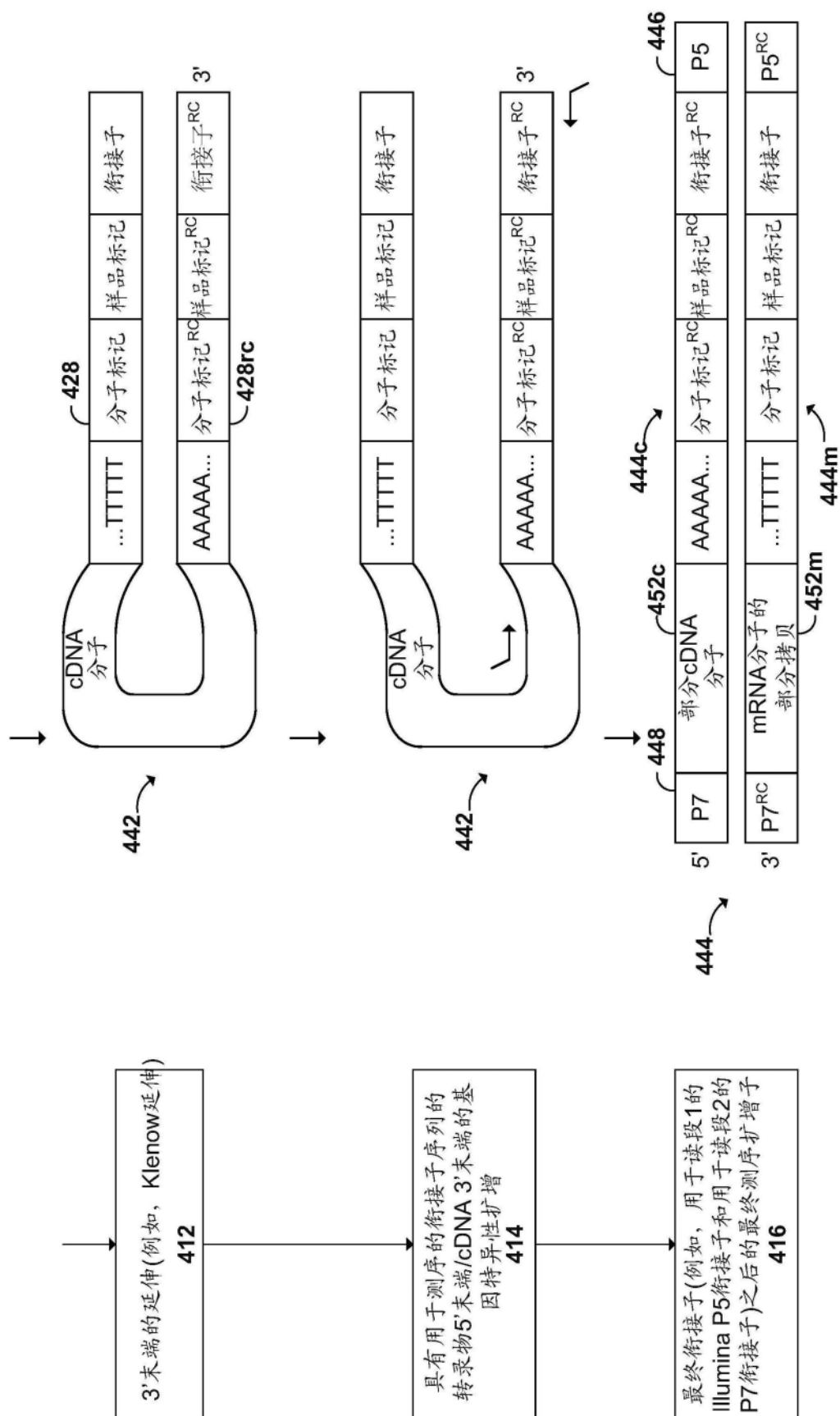


图4B

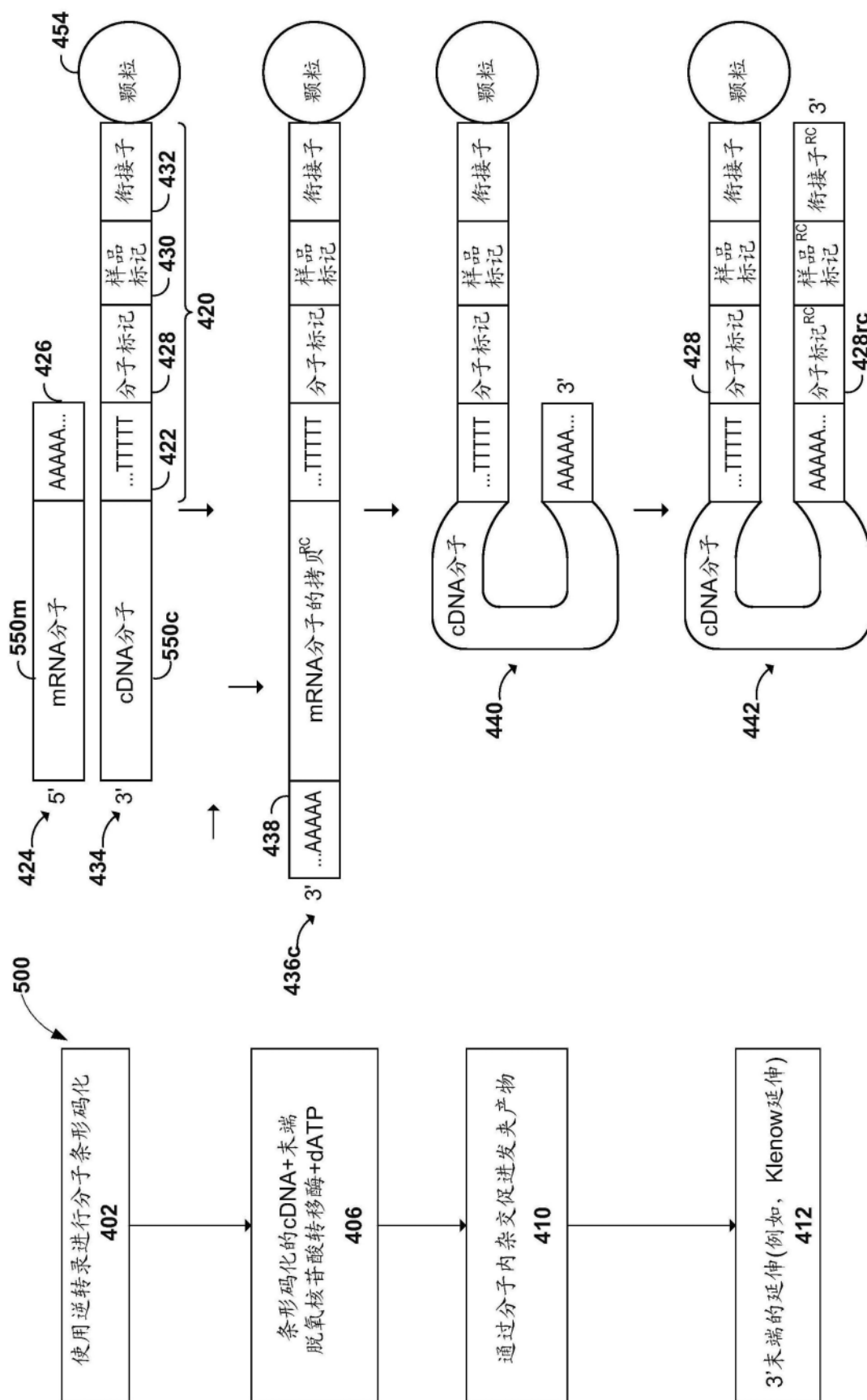


图5A

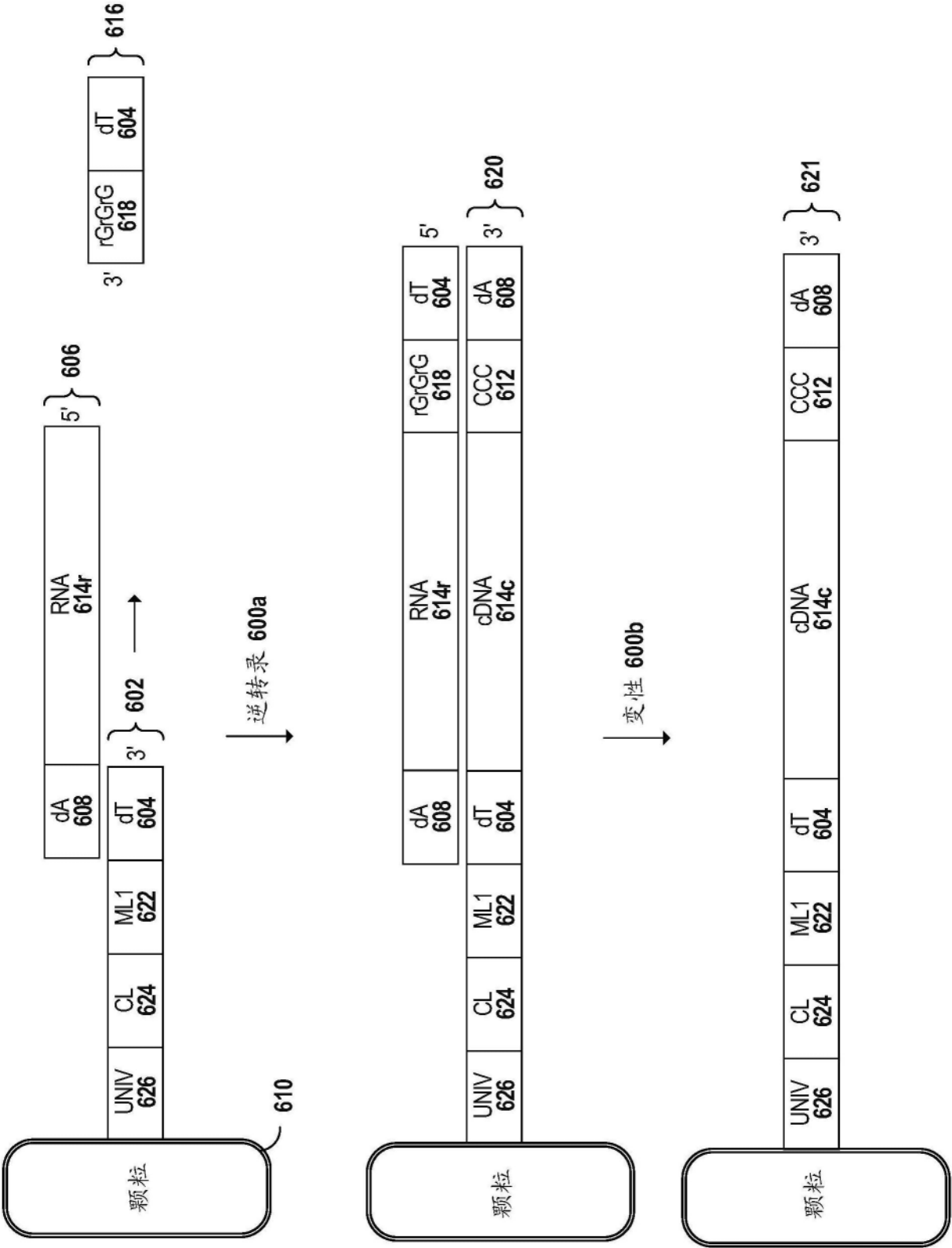


图6A

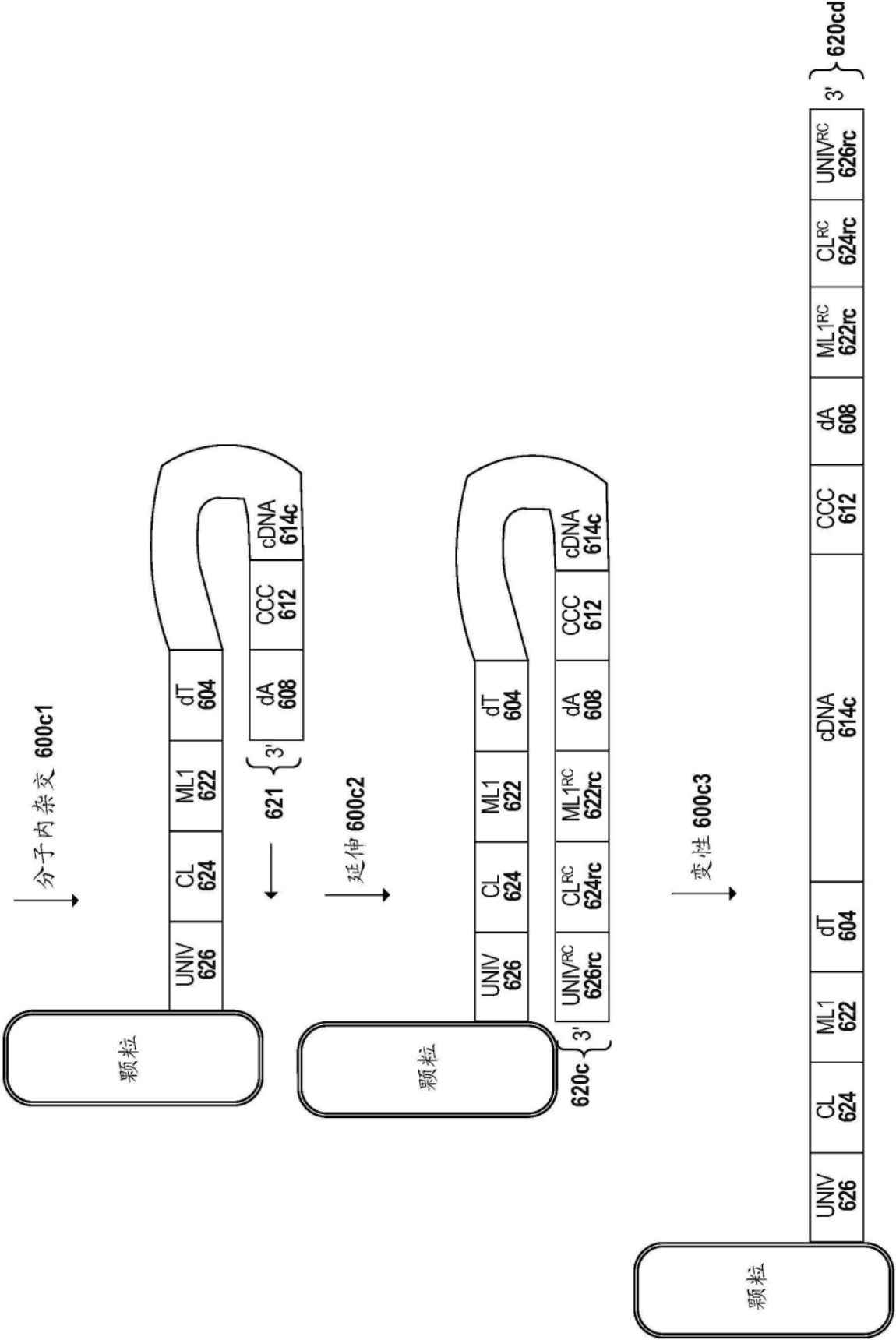


图6B

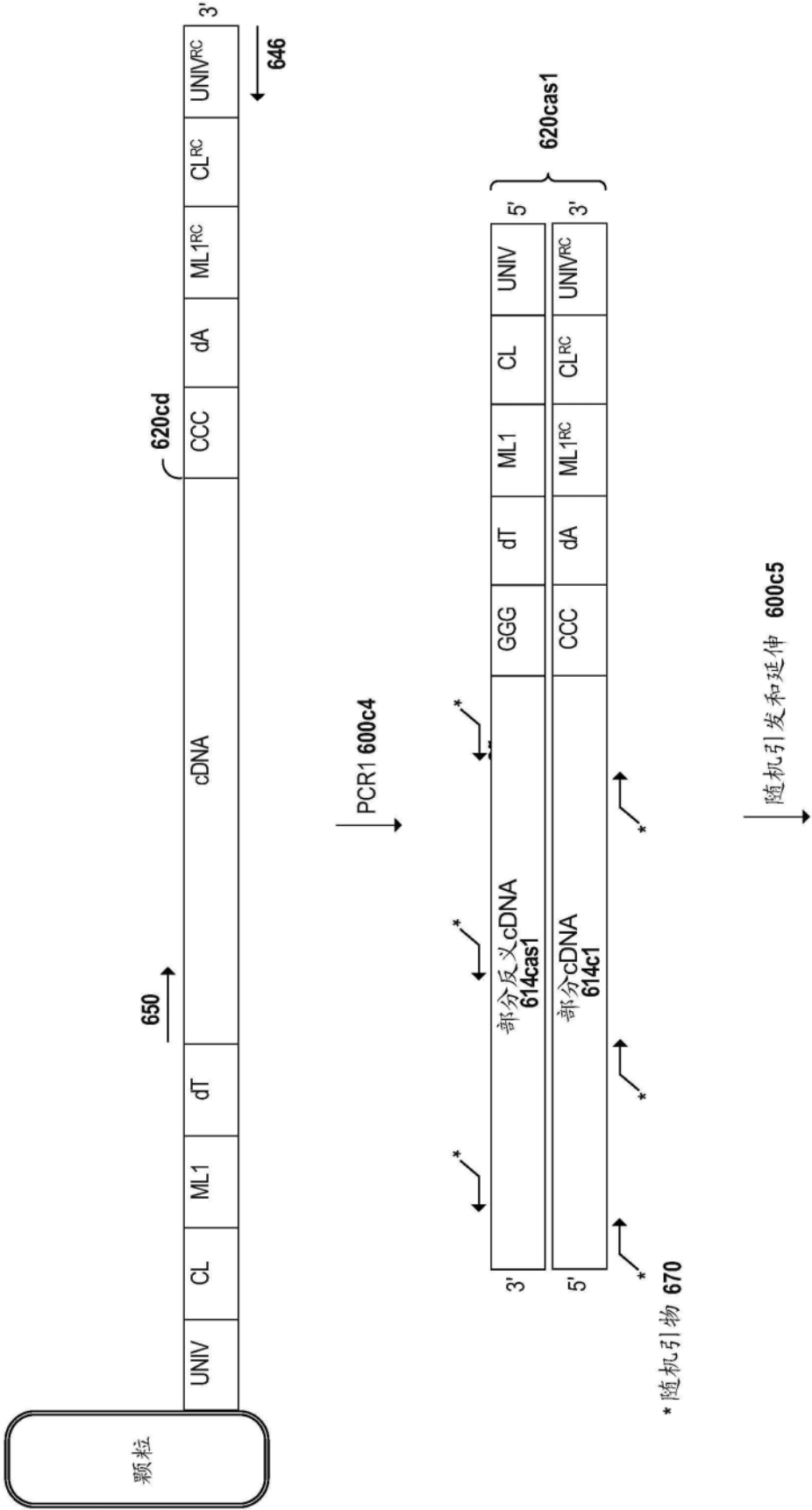


图6C

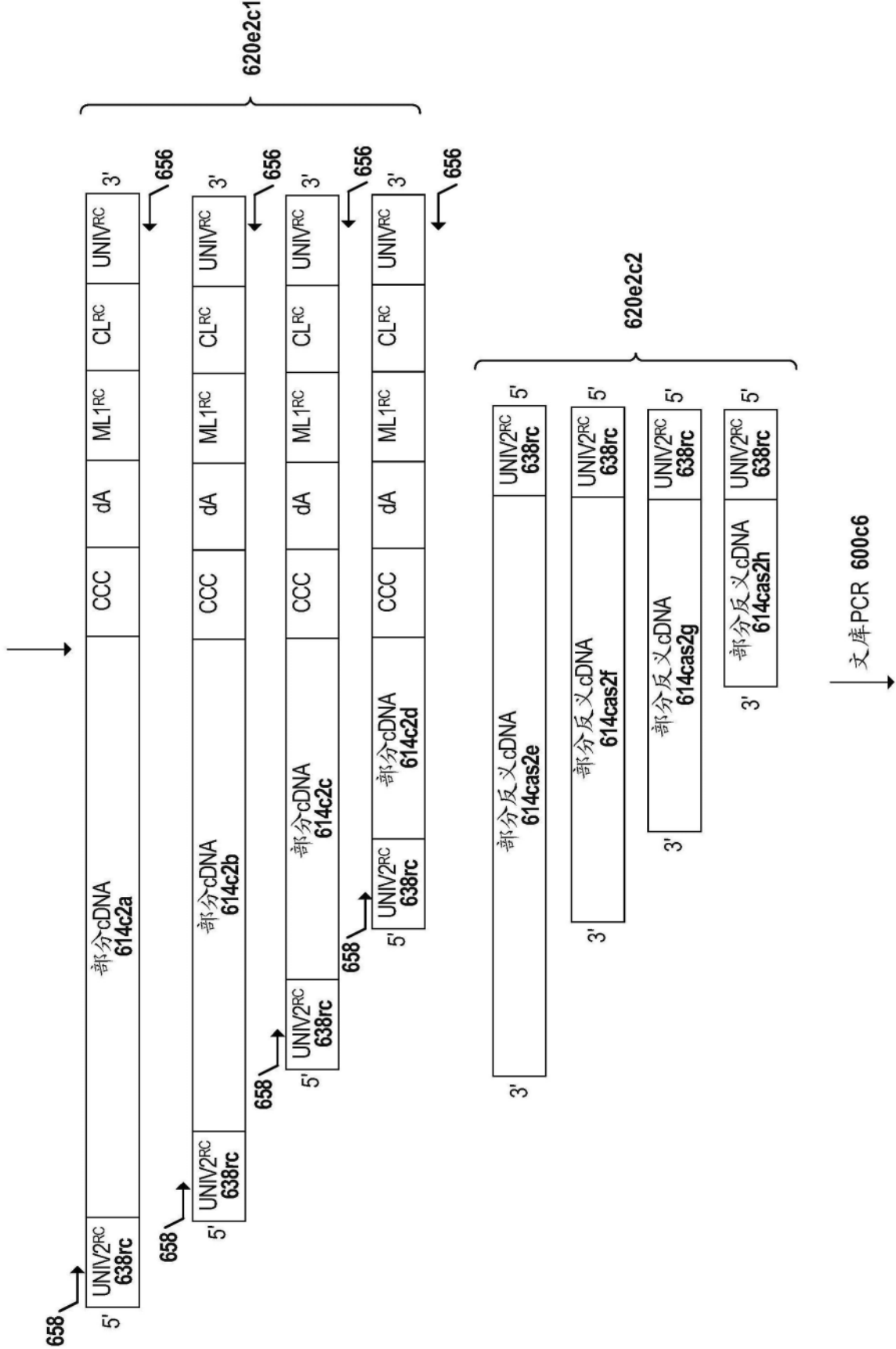


图6D

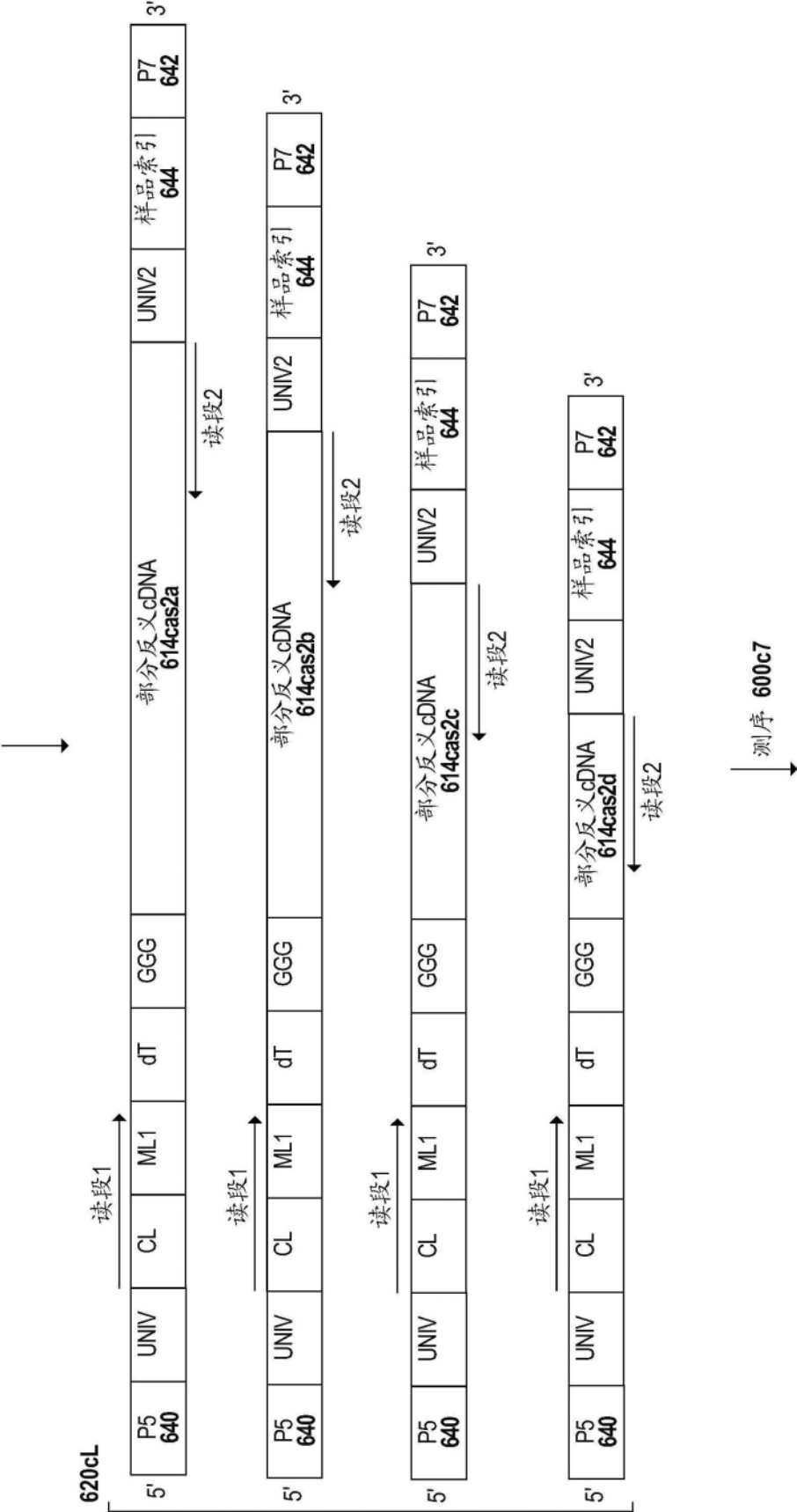


图6E

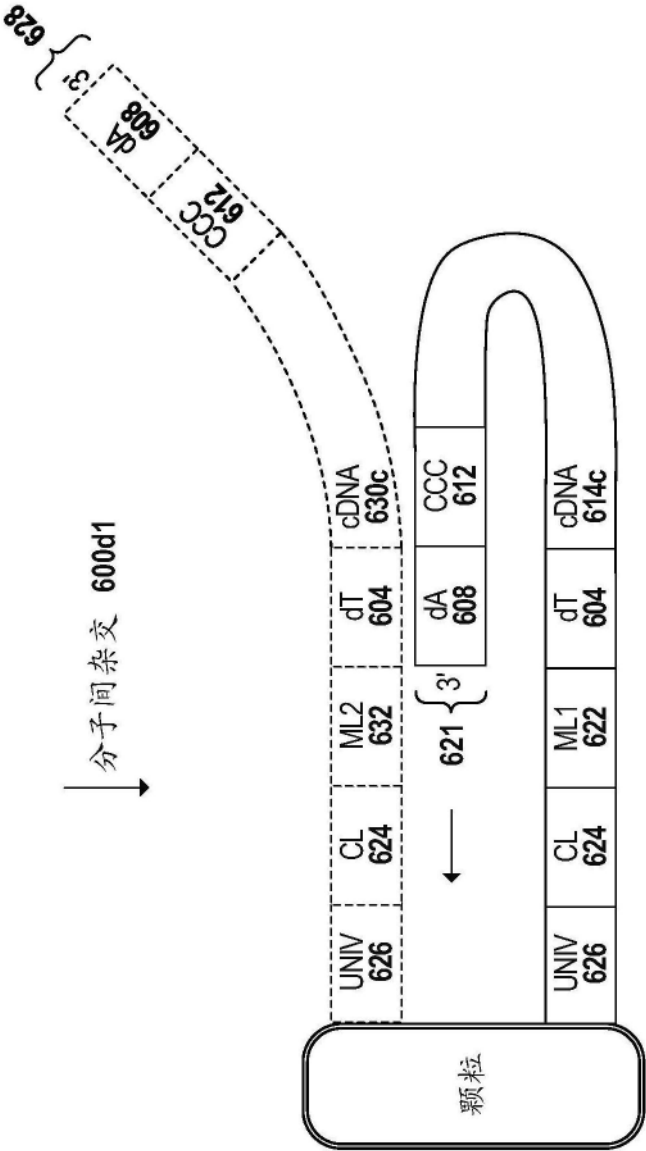


图6F

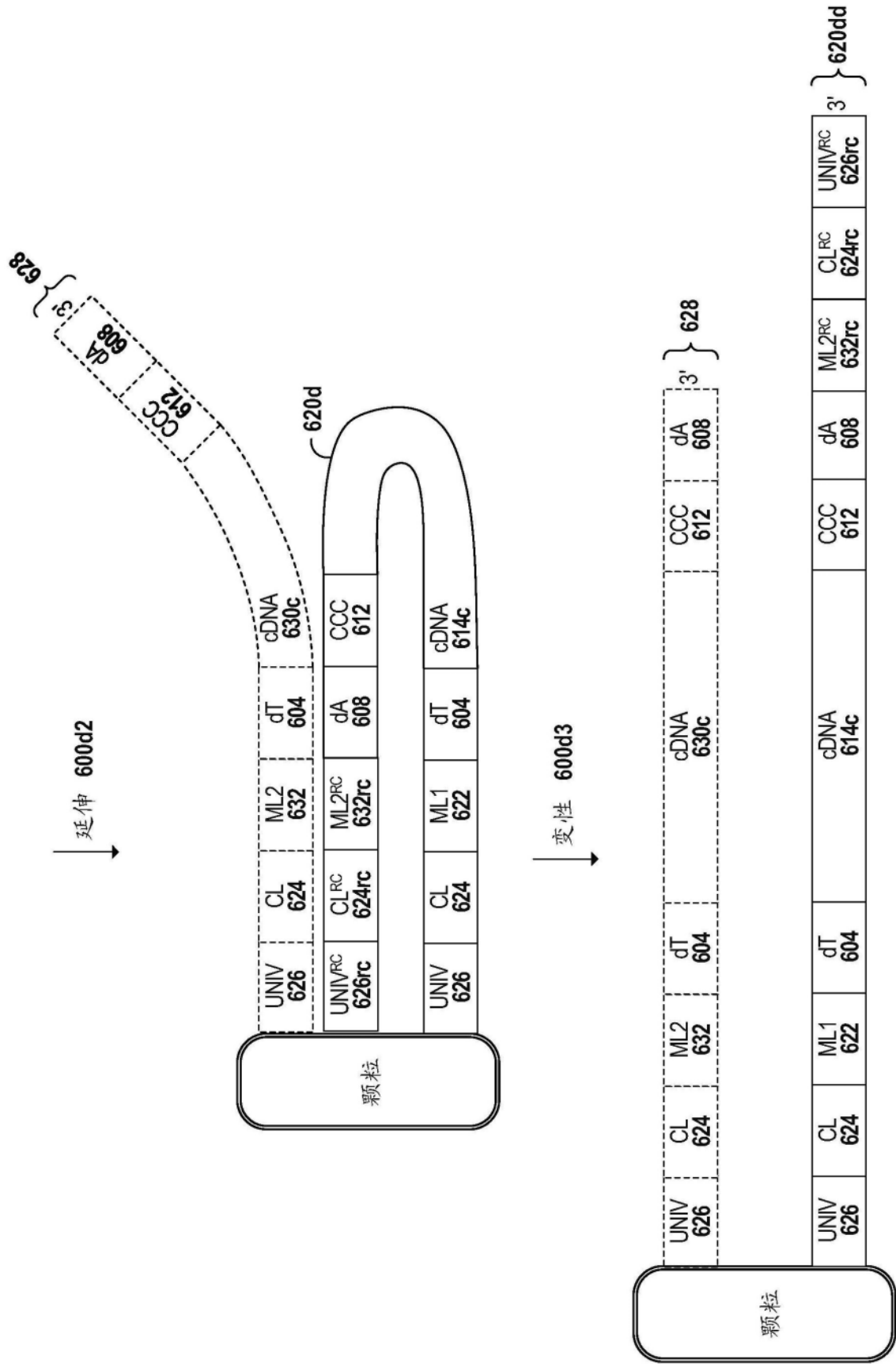


图6G

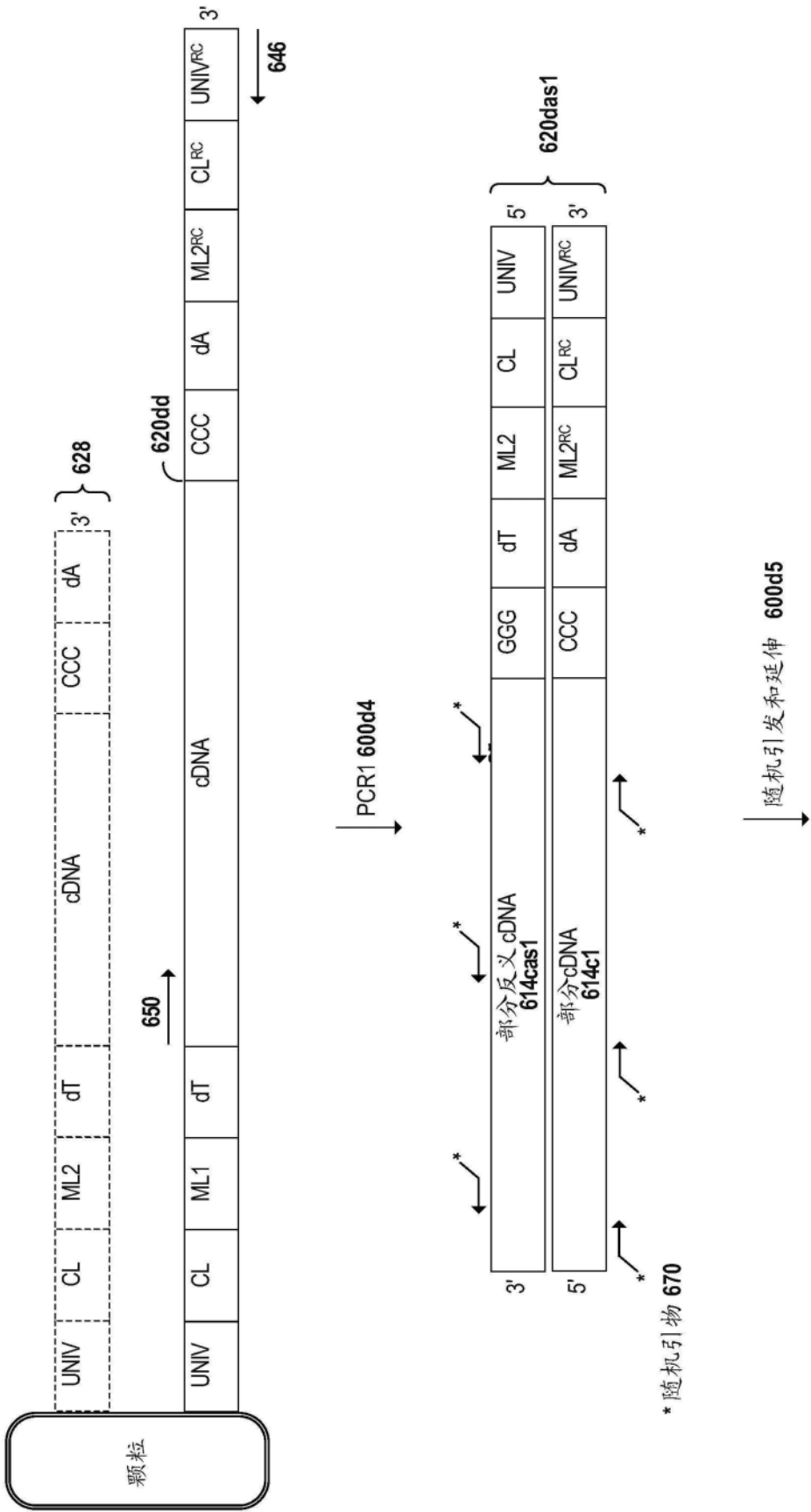


图6H

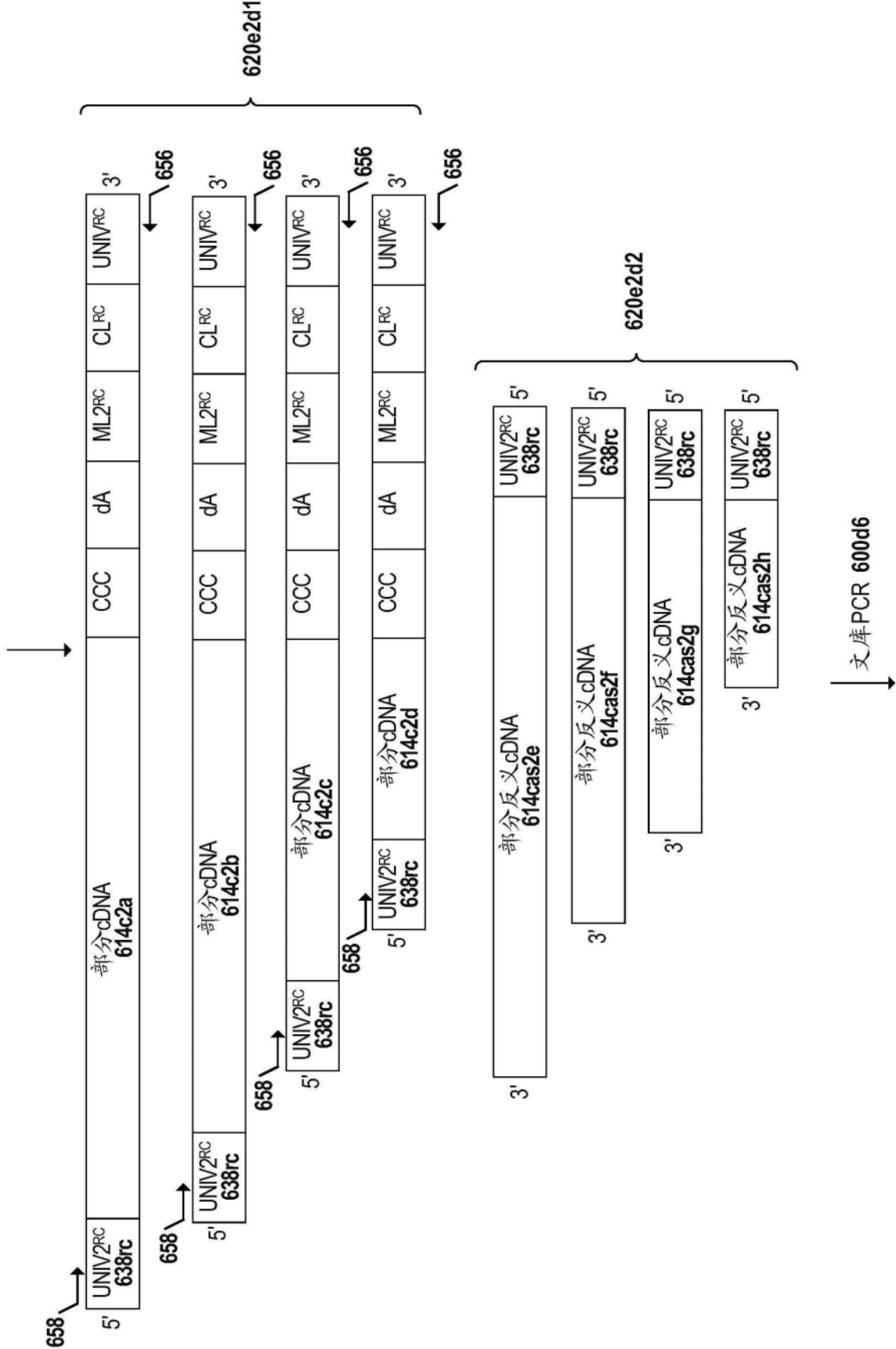


图6I

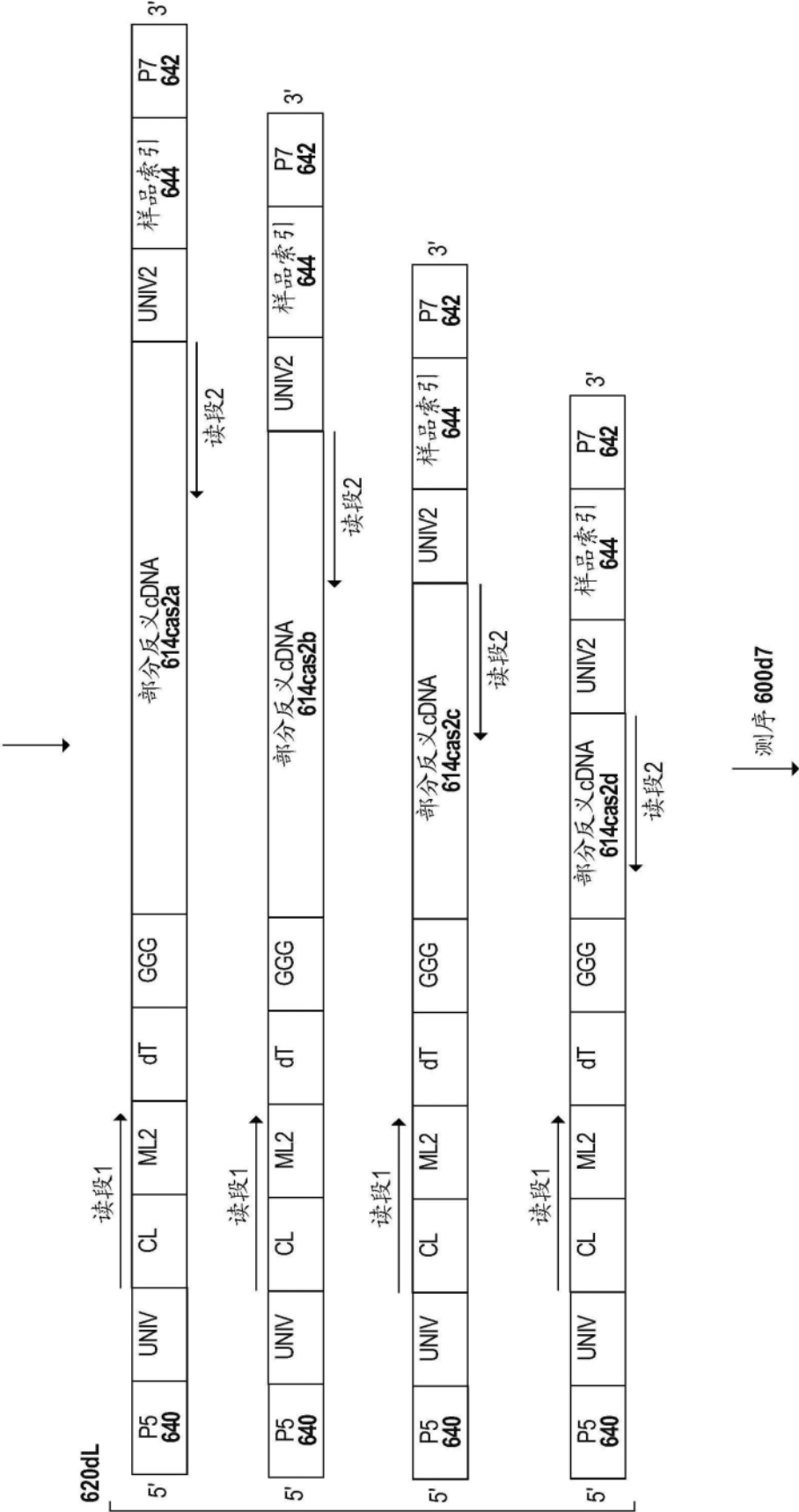


图6J

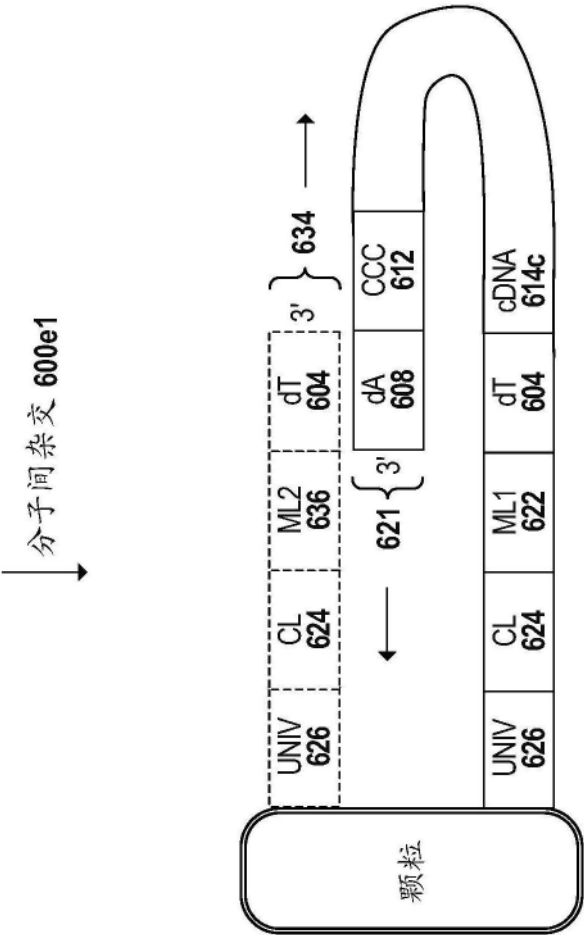


图6K

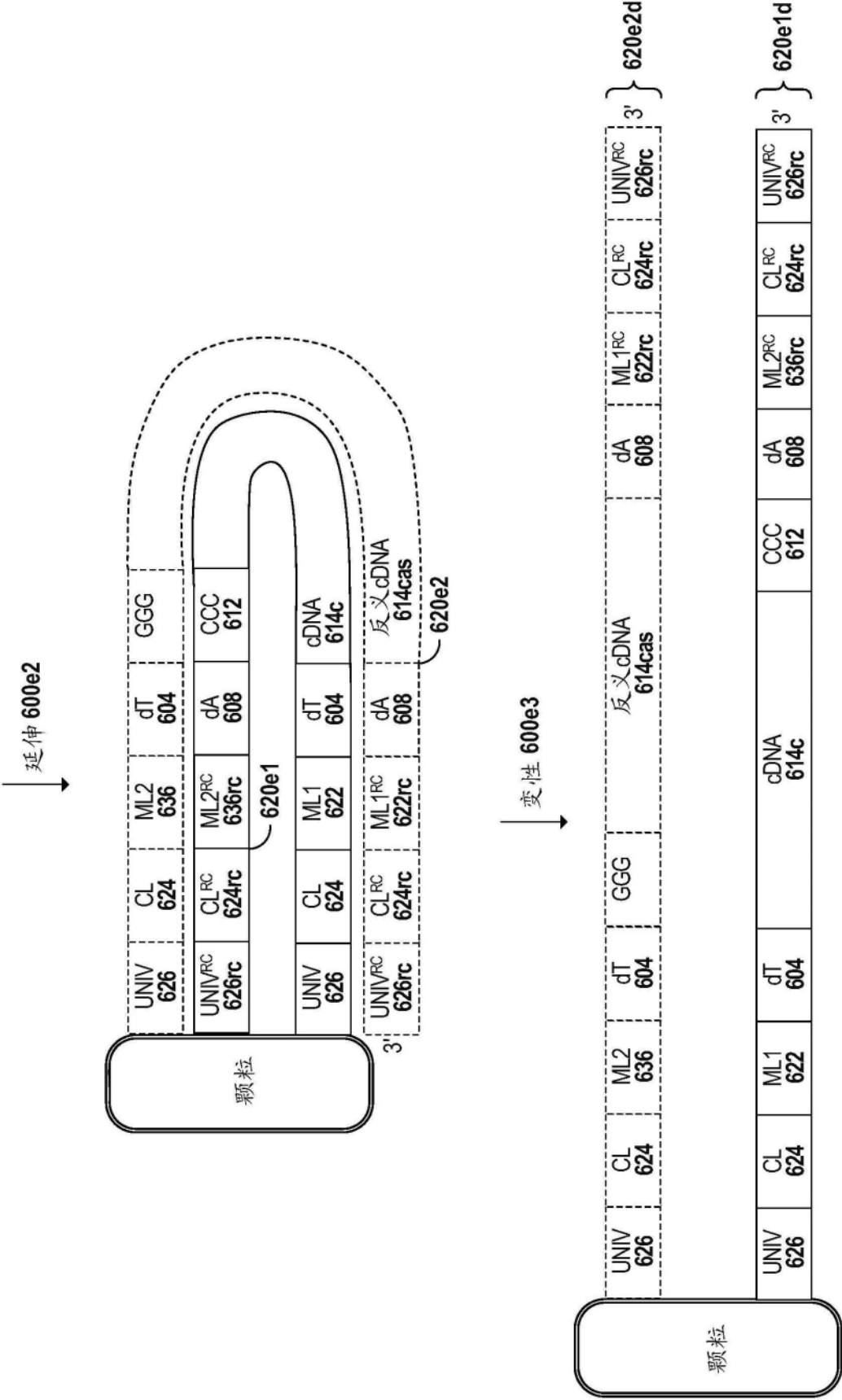


图6L

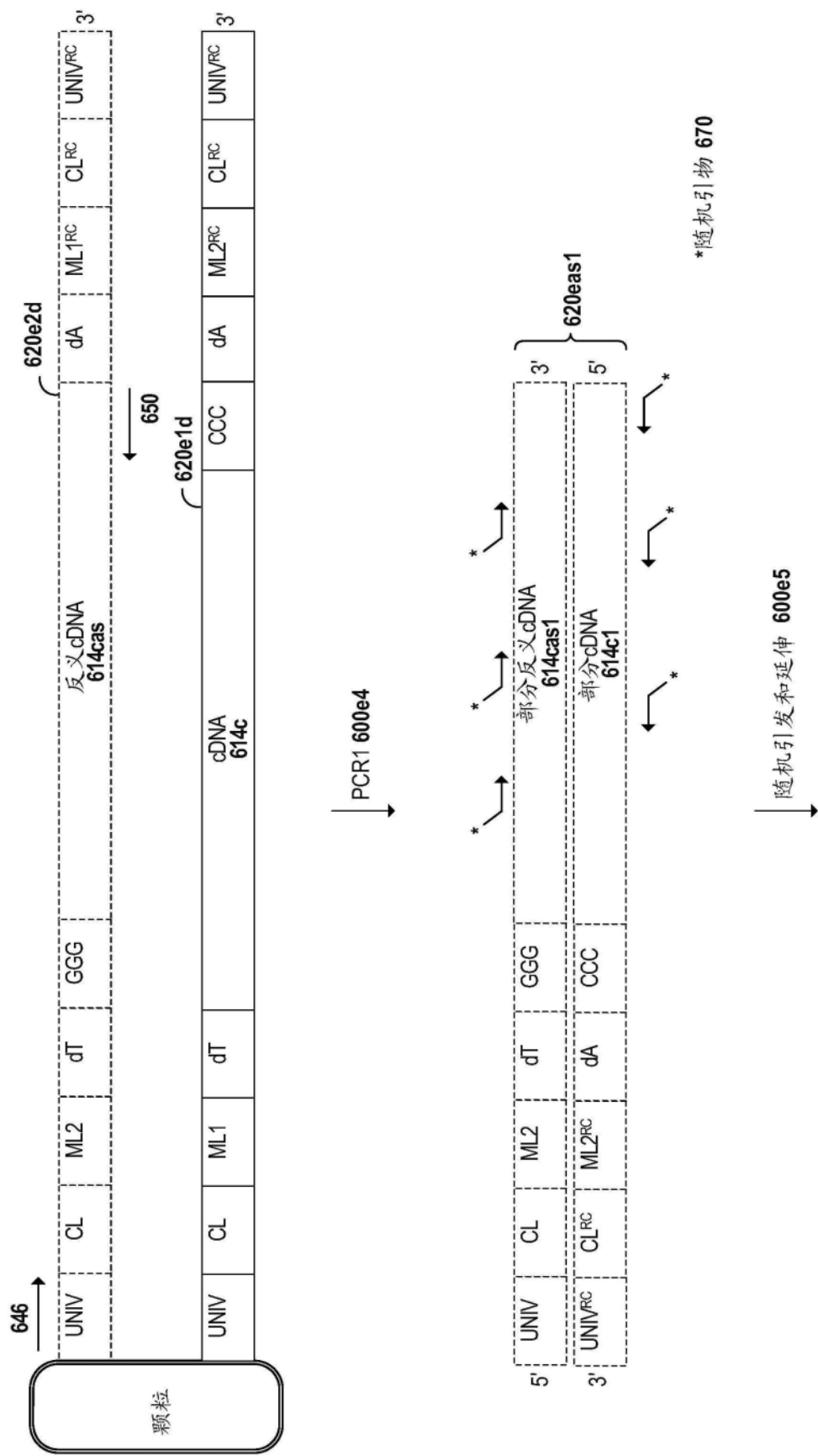


图6M

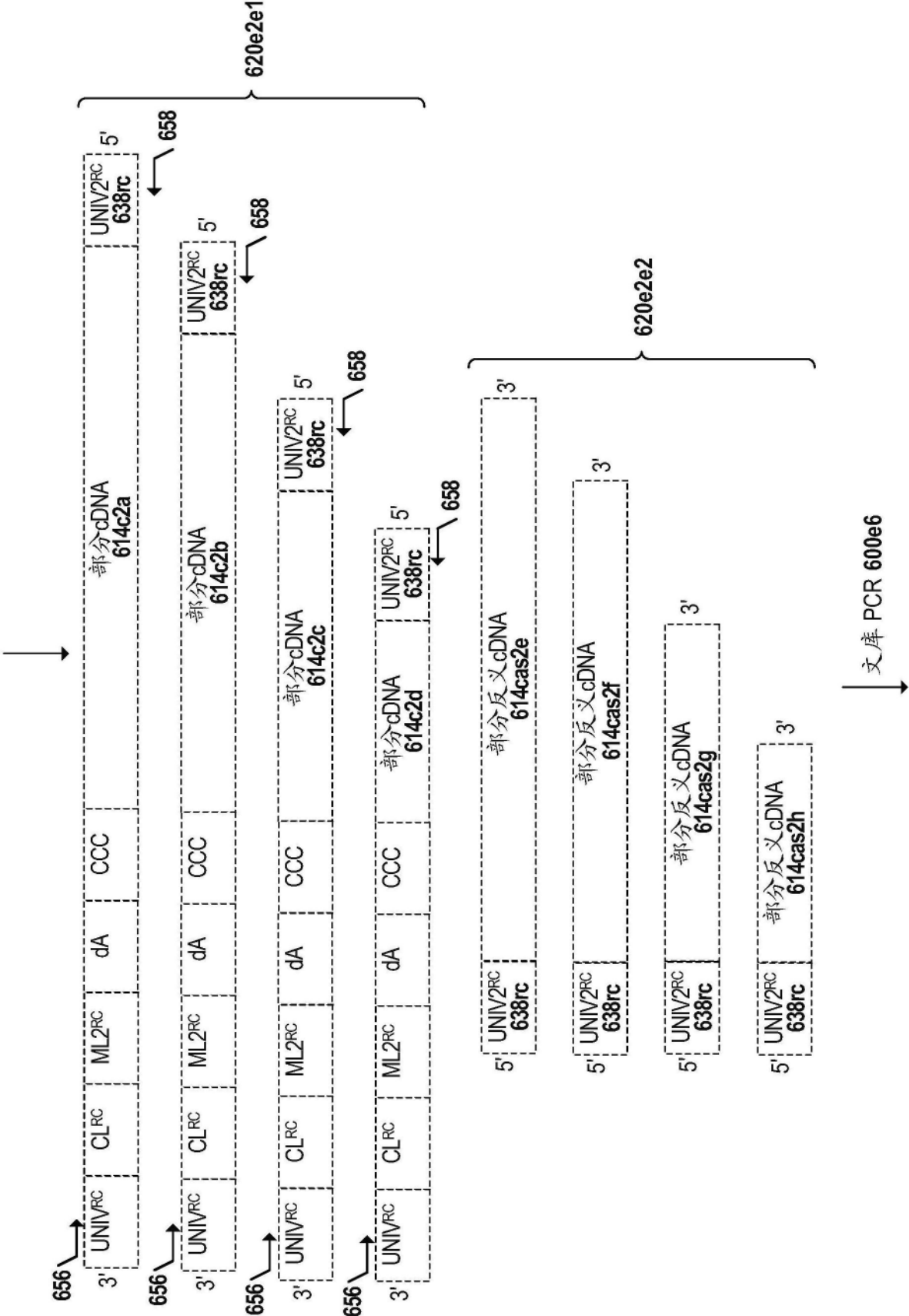


图6N

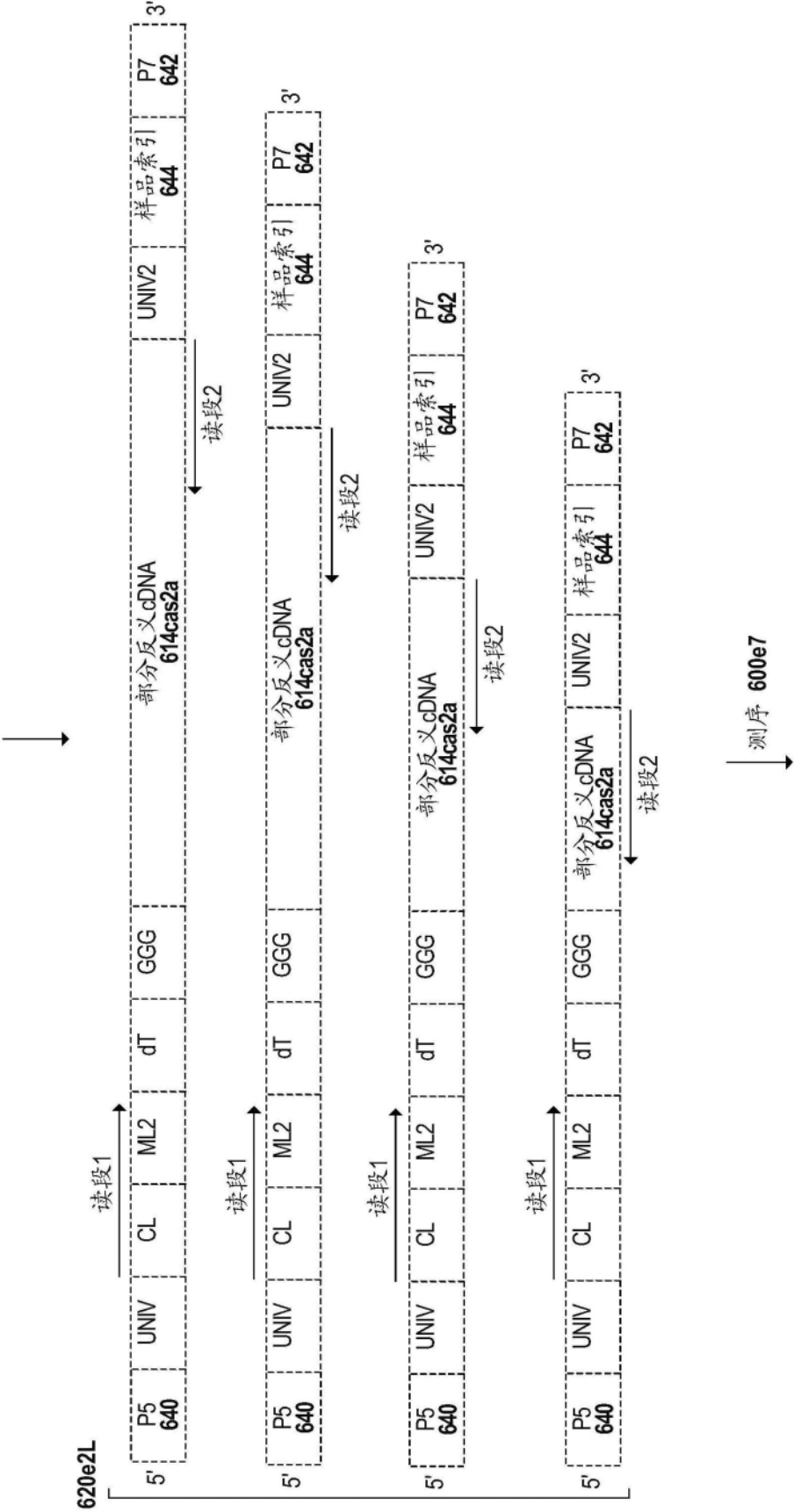


图60

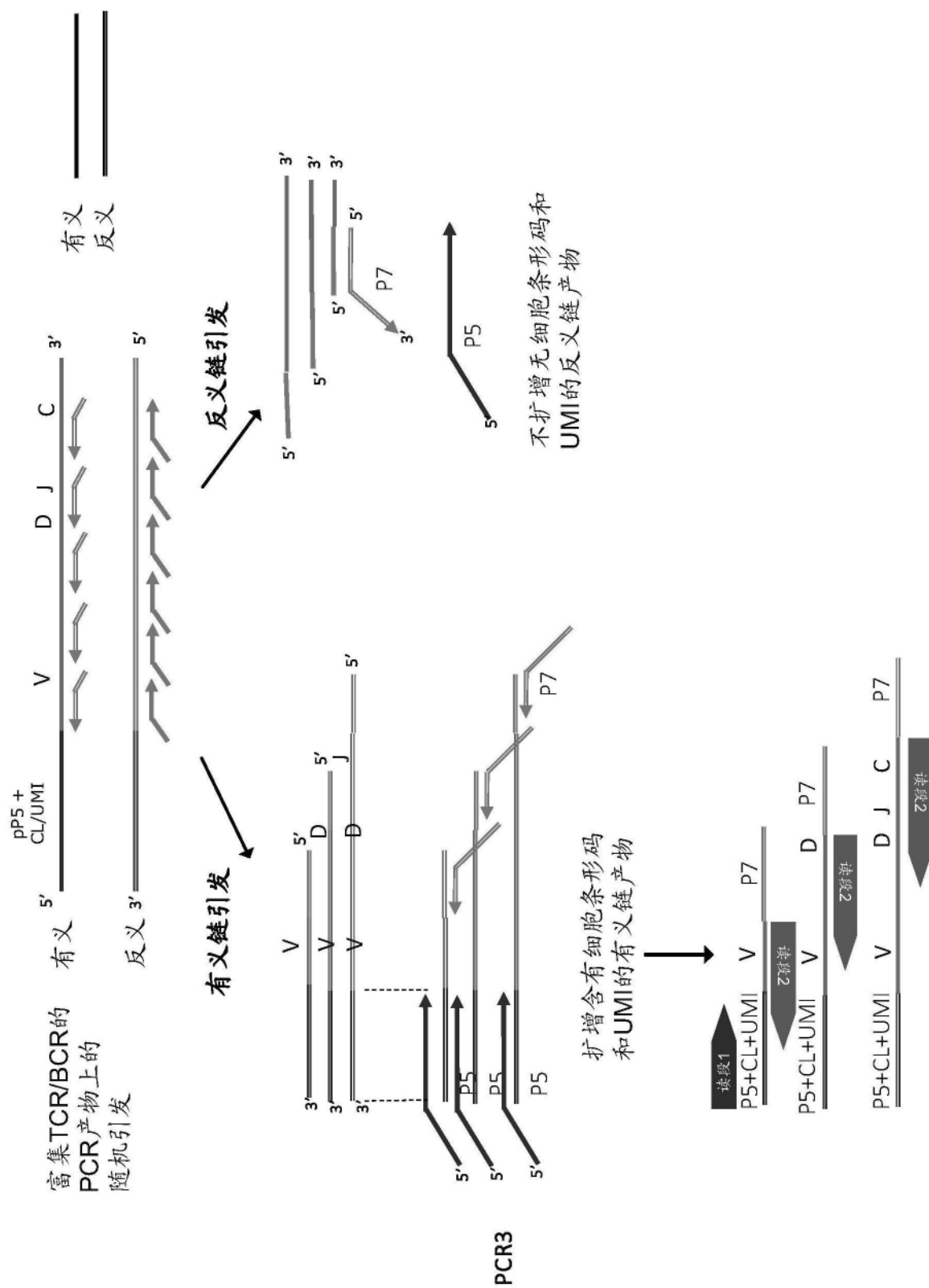


图7

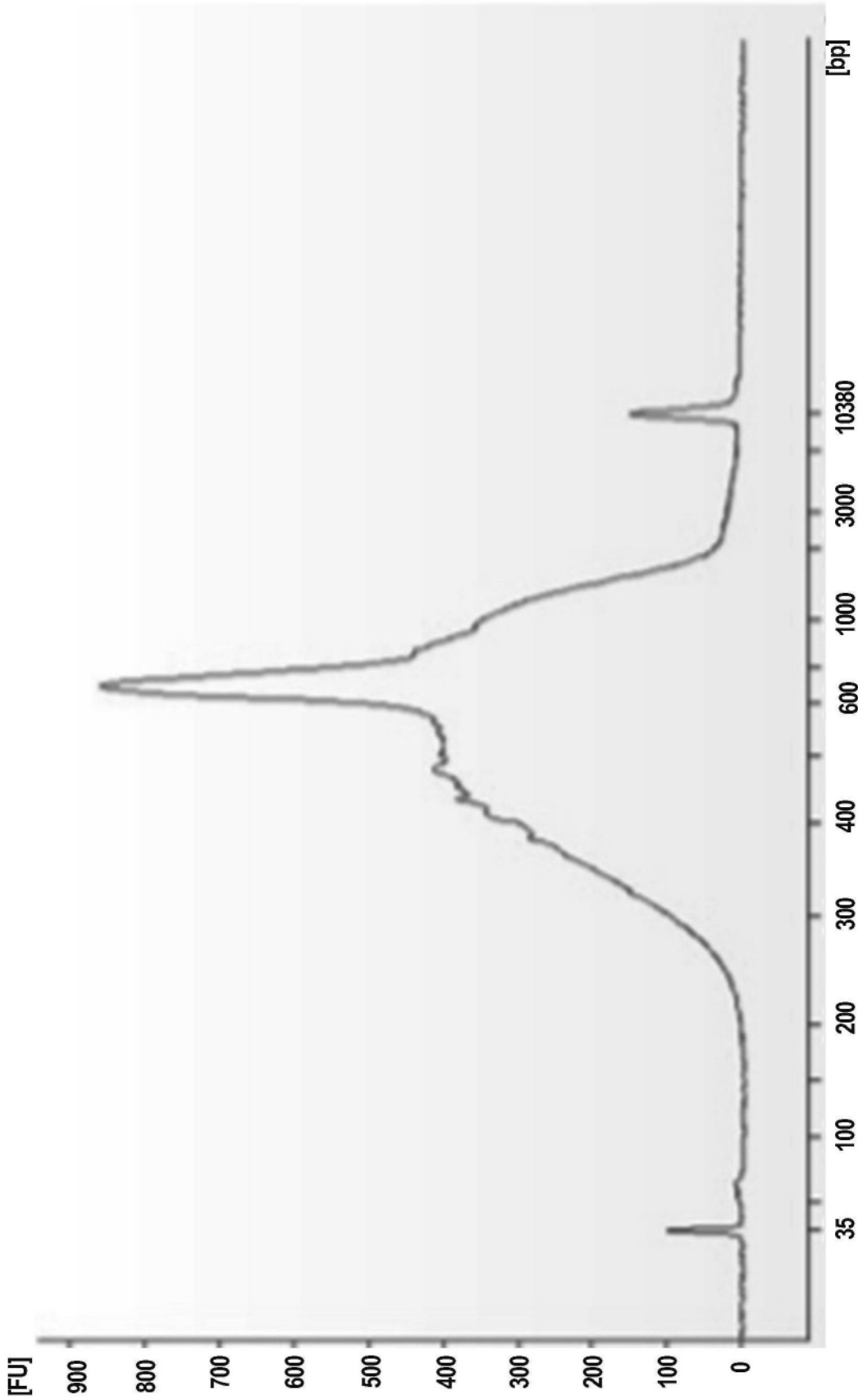


图8

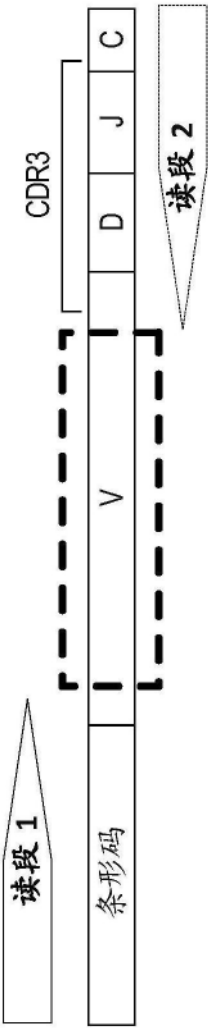


图9A

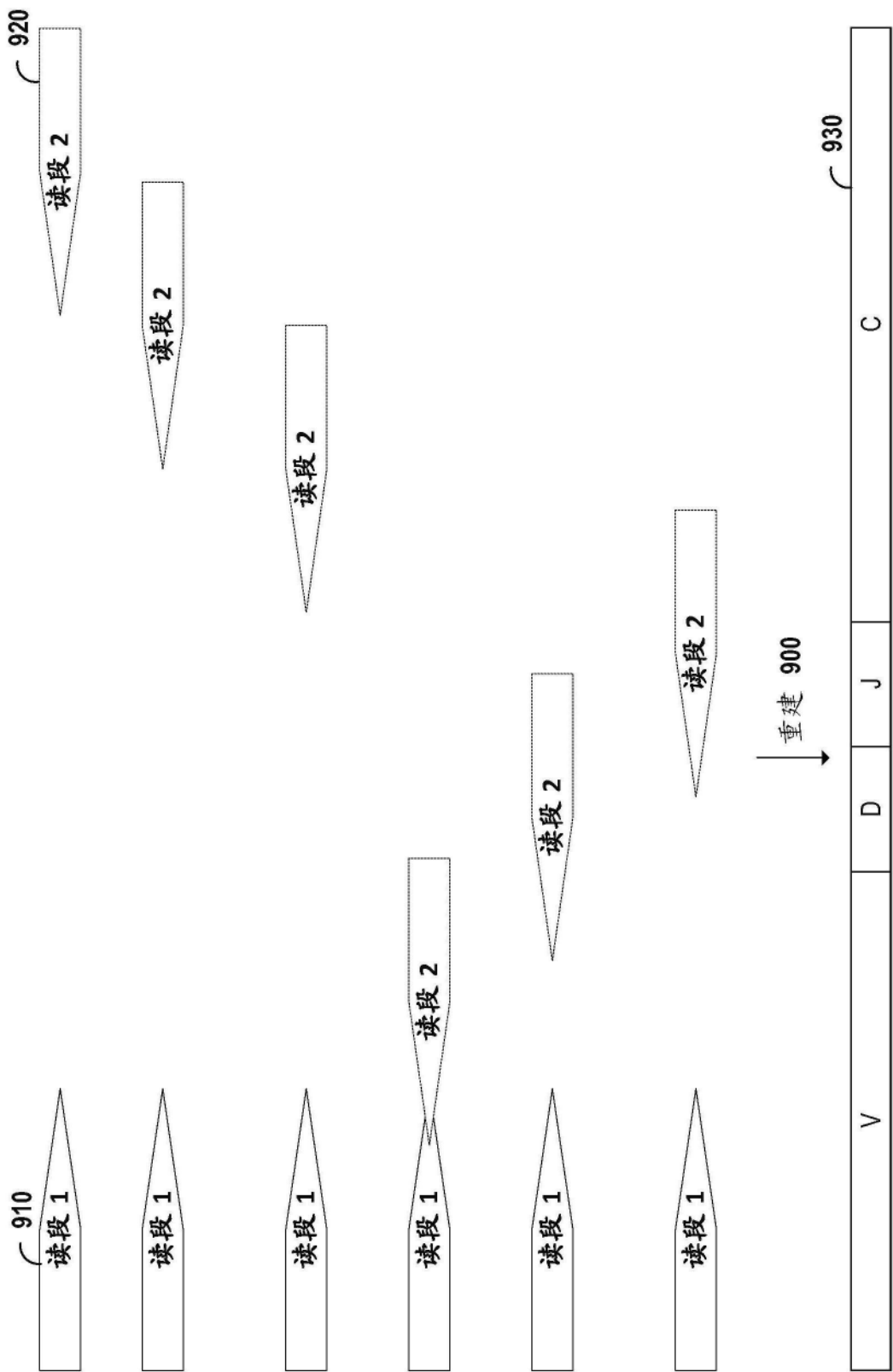


图9B

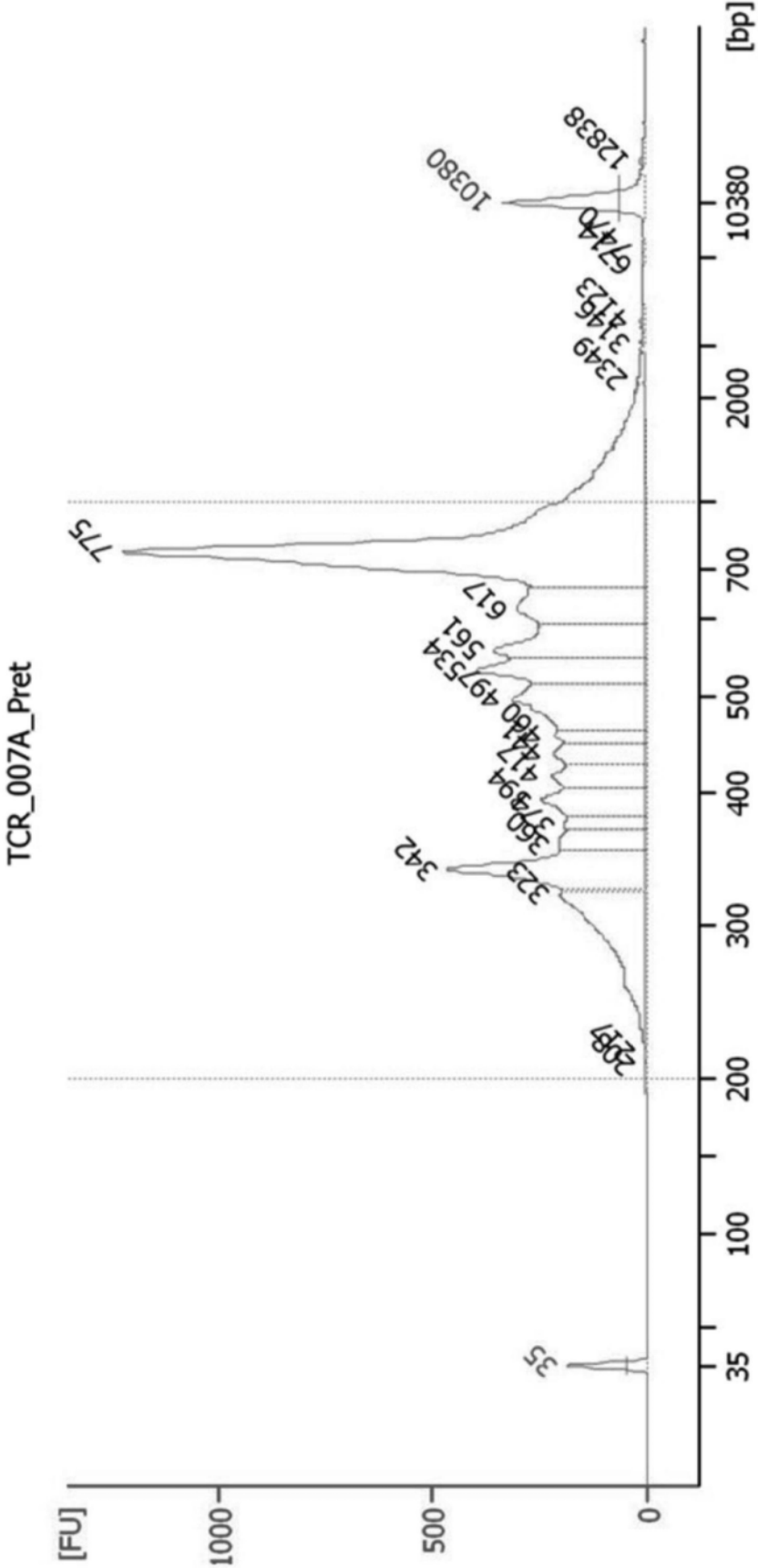


图10A

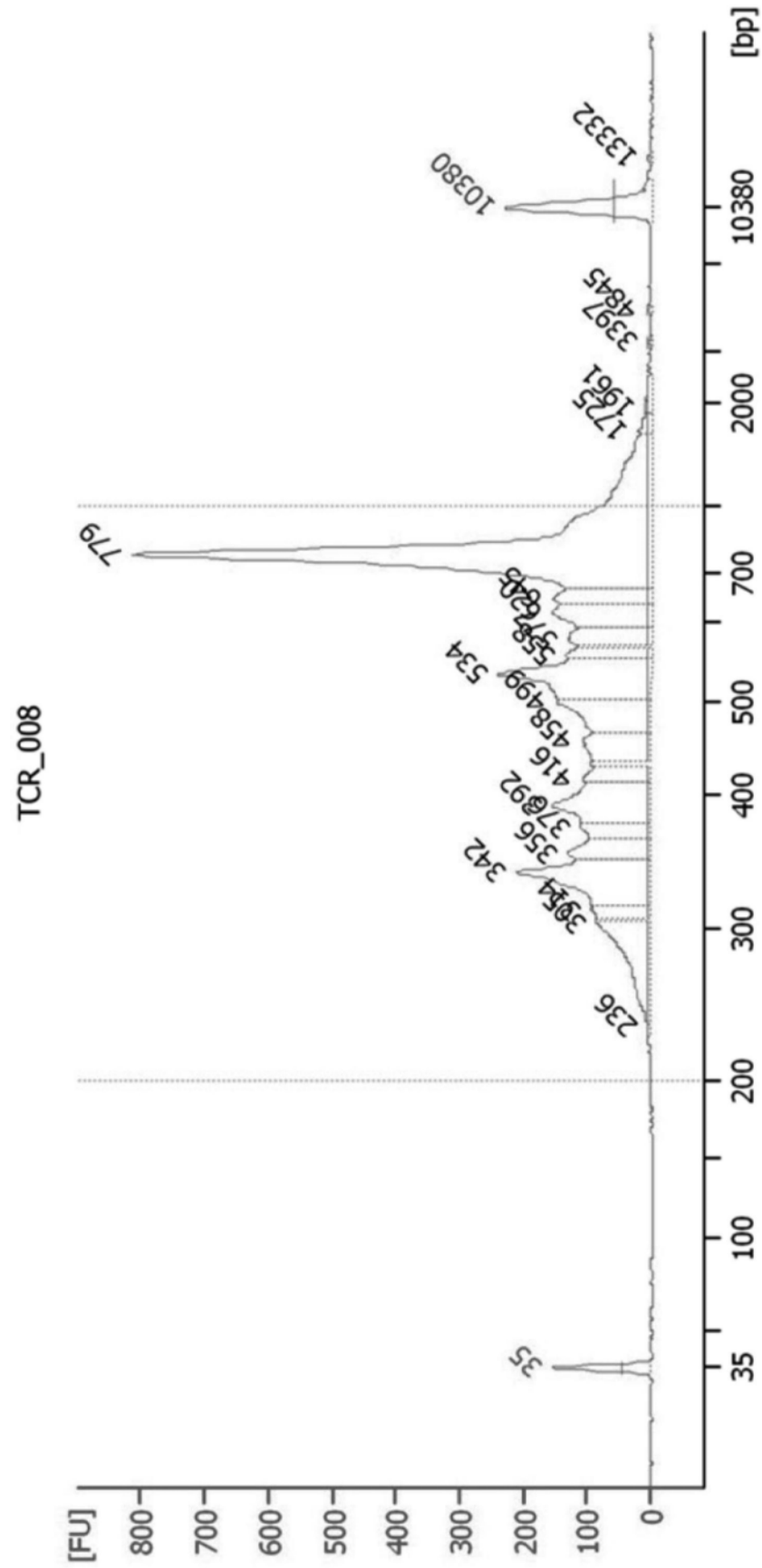


图10B

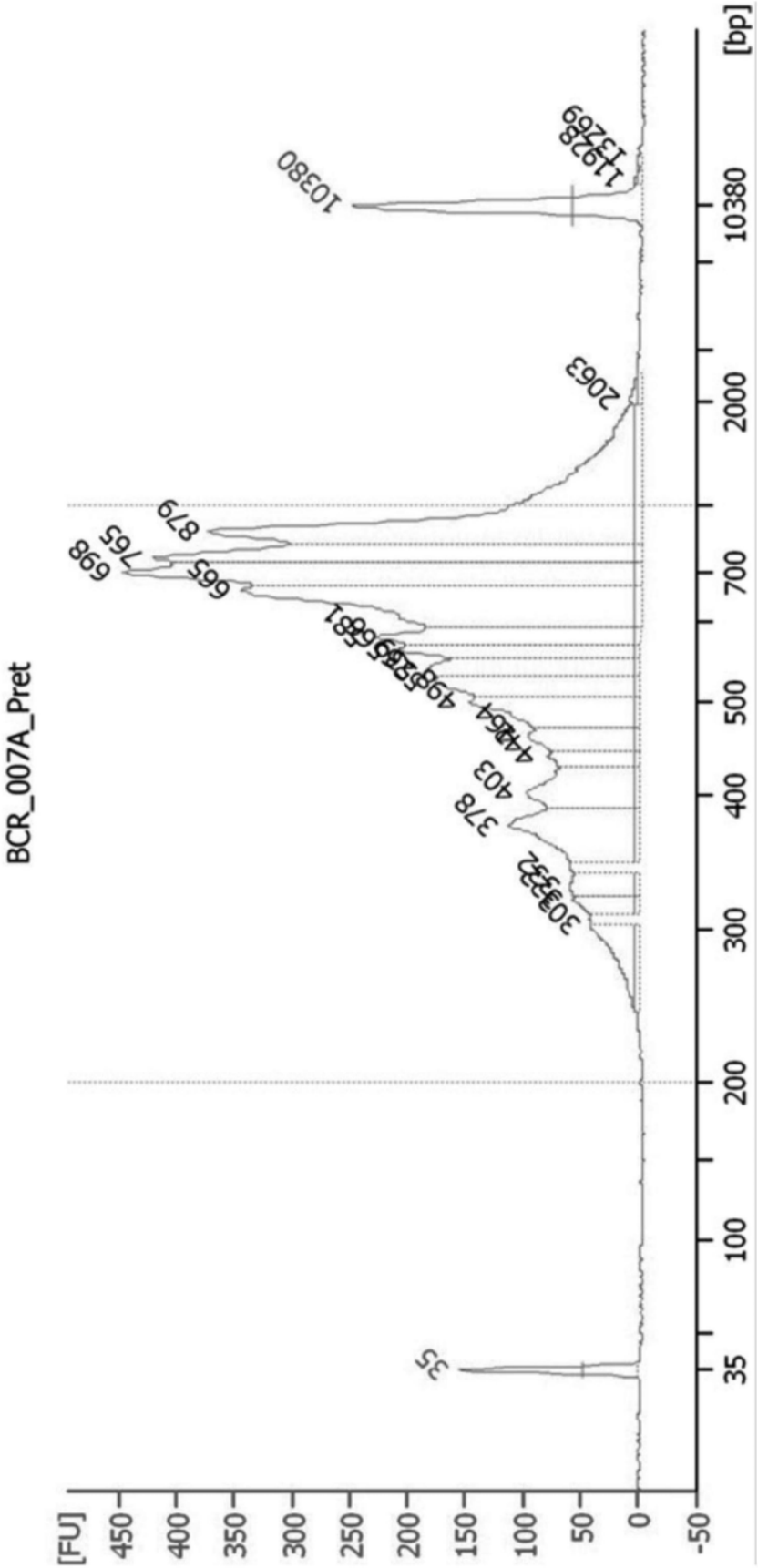


图10C

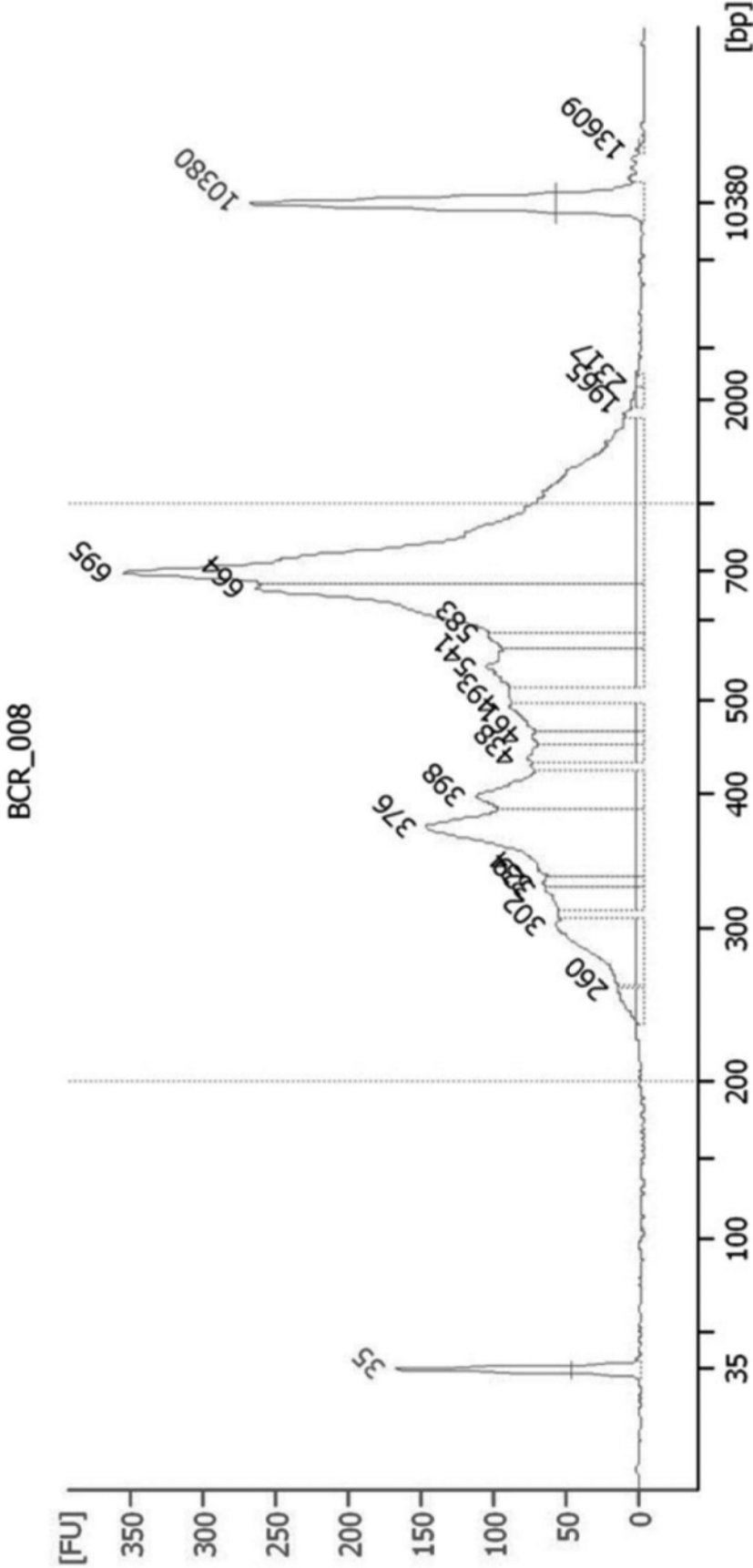


图10D

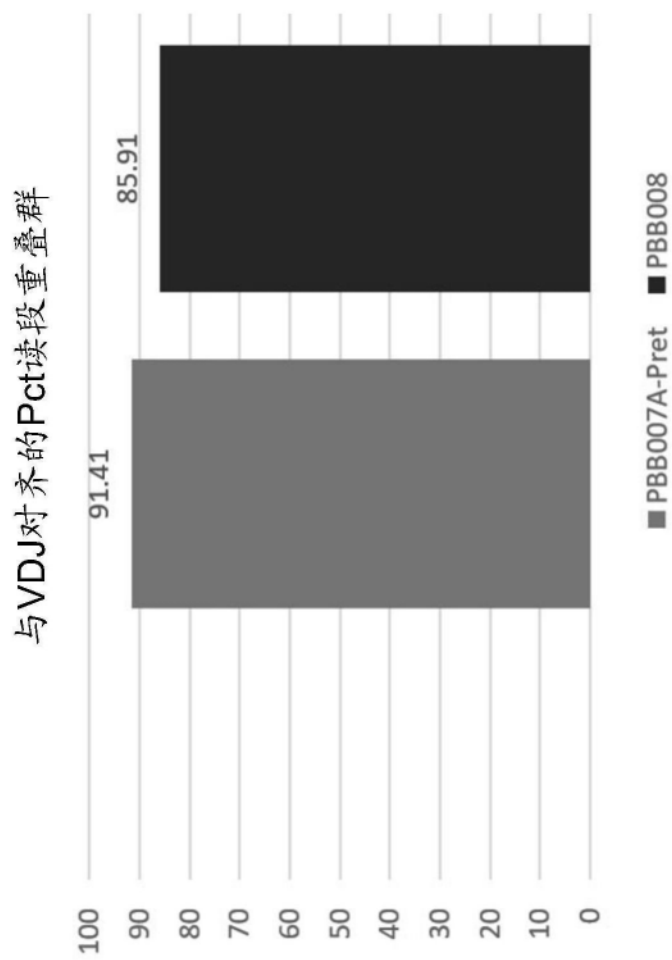


图11A

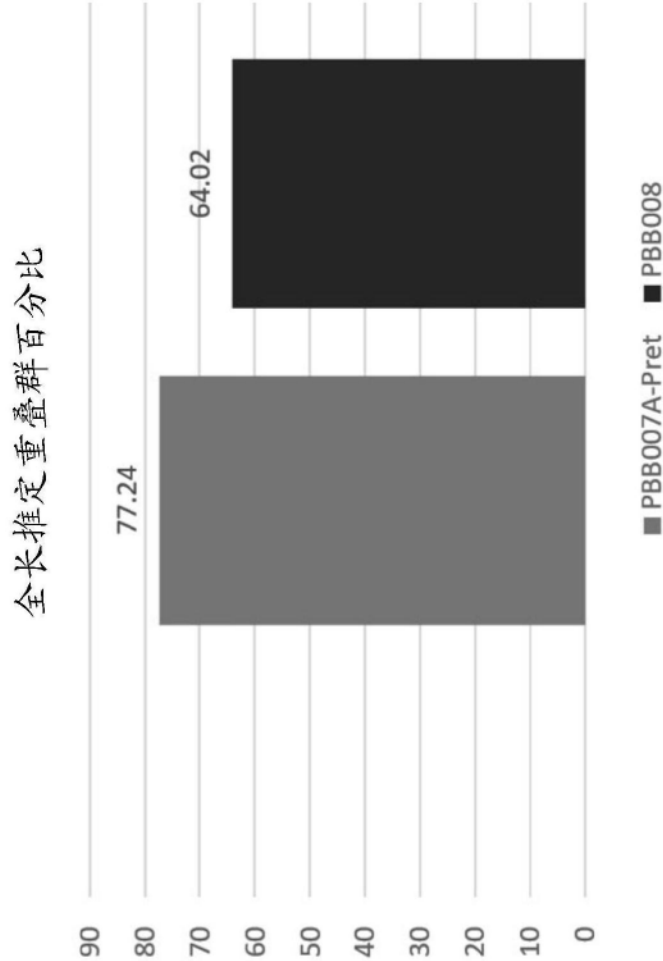


图11B

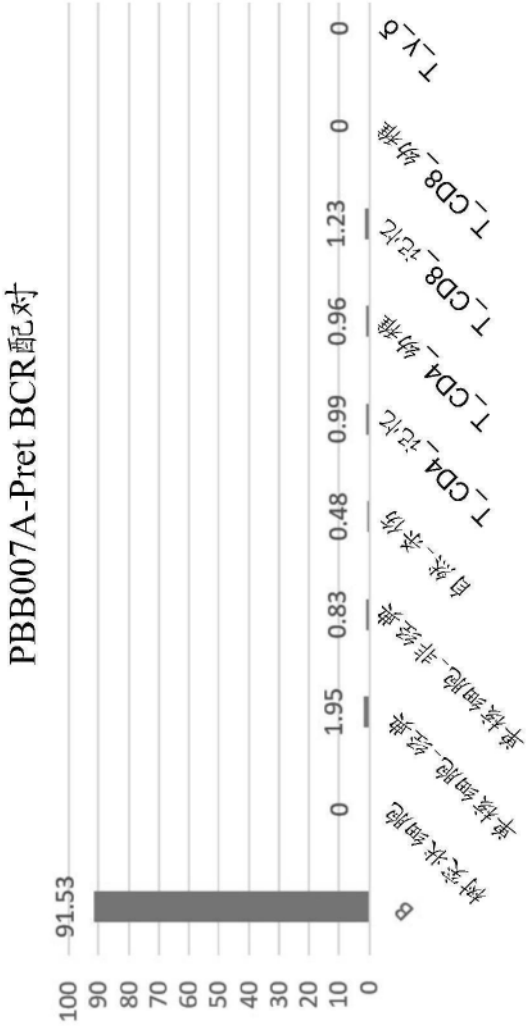


图12A

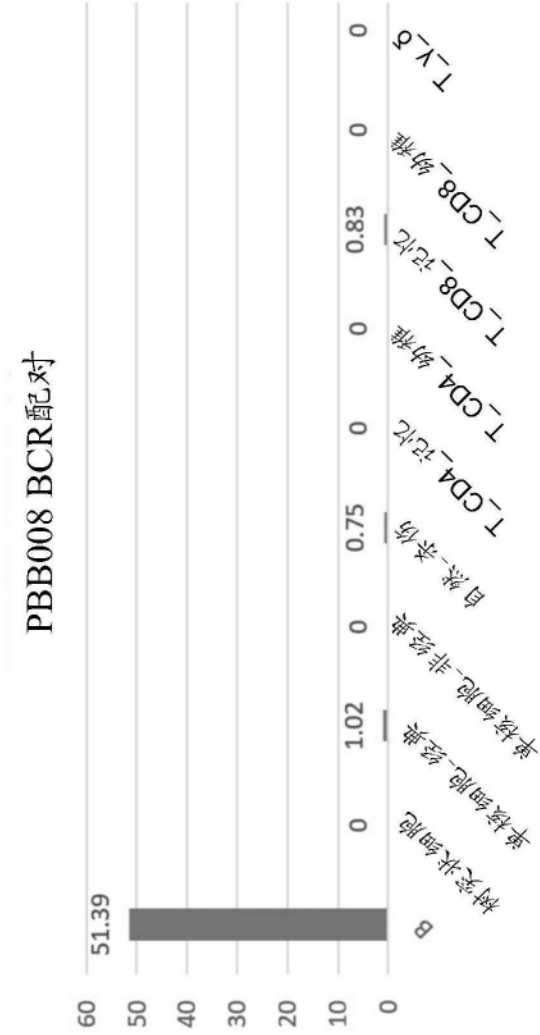


图12B

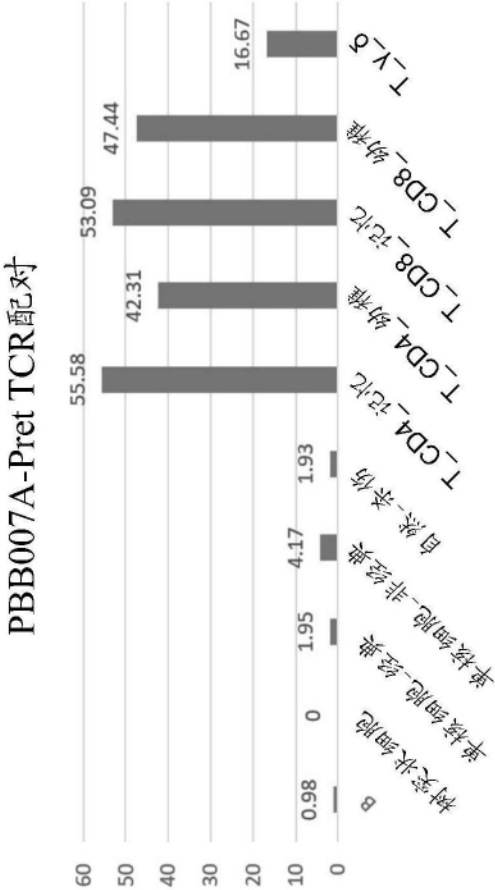


图12C

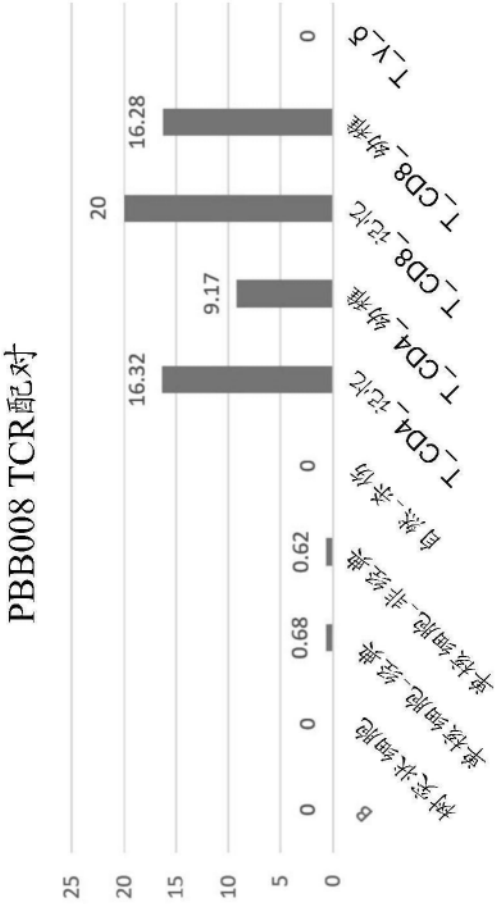


图12D