



(12) 发明专利

(10) 授权公告号 CN 114555802 B

(45) 授权公告日 2025. 01. 07

(21) 申请号 202080069560.6

(22) 申请日 2020.07.30

(65) 同一申请的已公布的文献号
申请公布号 CN 114555802 A

(43) 申请公布日 2022.05.27

(30) 优先权数据
62/881,183 2019.07.31 US

(85) PCT国际申请进入国家阶段日
2022.03.31

(86) PCT国际申请的申请数据
PCT/US2020/044338 2020.07.30

(87) PCT国际申请的公布数据
W02021/022085 EN 2021.02.04

(73) 专利权人 铂赛基因组学公司
地址 美国北卡罗来纳州

(72) 发明人 查尔斯·加瓦德
杰伊·A·A·韦斯特

(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

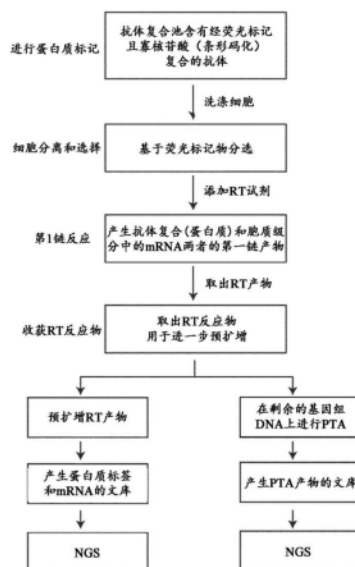
(51) Int.Cl.
C12N 15/10 (2006.01)
C12N 15/11 (2006.01)
C12Q 1/6869 (2018.01)
C12Q 1/6806 (2018.01)
C12Q 1/6844 (2018.01)
C12Q 1/686 (2018.01)

(56) 对比文件
US 2018216160 A1, 2018.08.02
WO 2018165459 A1, 2018.09.13
WO 2019028047 A1, 2019.02.07
赵利楠等. “基于单细胞靶向测序探究基因碱基突变的方法”.《遗传》.2020,第42卷(第7期),第703-712页.

审查员 樊艳爽
权利要求书3页 说明书44页
序列表9页 附图50页

(54) 发明名称
单细胞分析

(57) 摘要
本文提供了用于准确和可扩展的原代模板定向扩增(PTA)核酸扩增和测序方法的组合物和方法,以及其在研究、诊断和治疗中用于突变分析的应用。本文进一步提供了用于平行分析来自单细胞的DNA、RNA和/或蛋白质的多组学方法。本文提供了多组学单细胞分析方法,所述方法包括:(a)从细胞群体中分离单细胞;(b)对包含从来自所述单细胞的mRNA转录物扩增的多核苷酸的cDNA文库进行测序;(c)对所述单细胞的基因组进行测序。



1. 一种多组学单细胞分析的方法,所述方法包括:
 - a. 从细胞群体中分离单细胞;
 - b. 对包含从来自所述单细胞的mRNA转录物扩增的多核苷酸的cDNA文库进行测序;以及
 - c. 对所述单细胞的基因组进行测序,其中对所述基因组进行测序包括:
 - i. 使所述基因组与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制,并且其中所述至少一种终止子核苷酸为 α -硫代双脱氧核苷酸;以及
 - ii. 扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行,并且其中核酸复制的终止在链置换之前进行;
 - iii. 将在步骤(ii)中获得的分子连接至衔接子,从而产生基因组DNA文库;以及
 - iv. 对所述基因组DNA文库进行测序。
2. 如权利要求1所述的方法,其中所述mRNA转录物包括聚腺苷酸化mRNA转录物。
3. 如权利要求1所述的方法,其中所述mRNA转录物不包括聚腺苷酸化mRNA转录物。
4. 如权利要求1所述的方法,其中对cDNA文库进行测序包括用模板切换引物扩增mRNA转录物。
5. 如权利要求1所述的方法,其中所述cDNA文库中所述多核苷酸中的至少一些包含条形码。
6. 如权利要求5所述的方法,其中所述条形码包括细胞条形码或样品条形码。
7. 如权利要求1所述的方法,其中在测序之前将所述cDNA文库和所述基因组DNA文库合并。
8. 如权利要求1所述的方法,其中所述单细胞是原代细胞。
9. 如权利要求1所述的方法,其中所述单细胞源自肝脏、皮肤、肾脏、血液或肺。
10. 如权利要求1所述的方法,其中所述单细胞是癌细胞、神经元、神经胶质细胞或胎儿细胞。
11. 如权利要求1所述的方法,其中所述单细胞通过流式细胞术分离。
12. 如权利要求1所述的方法,其中所述方法还包括从所述终止的扩增产物中除去至少一种终止子核苷酸。
13. 如权利要求1所述的方法,其中所述多个终止的扩增产物的长度包含平均1000-2000个碱基。
14. 如权利要求1所述的方法,其中所述多个终止的扩增产物的长度为250-1500个碱基。
15. 如权利要求1所述的方法,其中所述多个终止的扩增产物包含所述单细胞的基因组的至少97%。
16. 如权利要求1所述的方法,其中所述扩增产物中的至少一些包含细胞条形码或样品条形码。
17. 如权利要求1所述的方法,其中对cDNA文库进行测序包括对所述单细胞的胞质裂解,以及逆转录。
18. 如权利要求1所述的方法,其中所述mRNA转录物经由模板切换逆转录进行扩增。
19. 如权利要求1所述的方法,其中所述cDNA文库包括至少10,000个基因。

20. 如权利要求1所述的方法,其中对所述单细胞的基因组进行测序还包括对所述单细胞的核裂解。

21. 如权利要求1所述的方法,其中所述方法还包括使用PCR的另外的扩增步骤。

22. 如权利要求1所述的方法,其中在所述细胞的所述基因组中鉴定出至少一个突变,其中所述突变不同于参考序列中的相应位置。

23. 如权利要求1所述的方法,其中所述至少一个突变在所述细胞群体的少于1%中发生。

24. 如权利要求1所述的方法,其中所述至少一个突变在所述细胞群体的不多于0.1%中发生。

25. 如权利要求1所述的方法,其中所述至少一个突变在所述细胞群体的不多于0.001%中发生。

26. 如权利要求1所述的方法,其中所述至少一个突变在所述扩增产物序列的不多于1%中发生。

27. 如权利要求1所述的方法,其中所述至少一个突变在所述扩增产物序列的不多于0.1%中发生。

28. 如权利要求1所述的方法,其中所述至少一个突变在所述扩增产物序列的不多于0.001%中发生。

29. 如权利要求1所述的方法,其中在最多200百万测序读段后,所涵盖的基因组的比例达到0.9。

30. 一种多组学单细胞分析的方法,所述方法包括:

a. 从细胞群体中分离单细胞;

b. 鉴定所述单细胞的表面上的至少一种蛋白质;以及

c. 对所述单细胞的基因组进行测序,其中对所述基因组进行测序包括:

i. 使所述基因组与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制,并且其中所述至少一种终止子核苷酸为 α -硫代双脱氧核苷酸;

ii. 扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行,并且其中核酸复制的终止在链置换之前进行;

iii. 将在步骤(ii)中获得的分子连接至衔接子,从而产生基因组DNA文库;以及

iv. 对所述基因组DNA文库进行测序。

31. 如权利要求30所述的方法,其中鉴定所述细胞表面上的至少一种蛋白质包括使所述细胞与结合至所述至少一种蛋白质的经标记的抗体接触。

32. 如权利要求31所述的方法,其中所述经标记的抗体包含至少一种荧光标记或质量标签。

33. 如权利要求31所述的方法,其中所述经标记的抗体包含至少一种核酸条形码。

34. 一种多组学单细胞分析的方法,所述方法包括:

a. 从细胞群体中分离单细胞;

b. 对所述单细胞的基因组进行测序,其中对所述细胞的所述基因组进行测序包括:

i. 用甲基化敏感性限制性酶消化所述基因组以产生基因组片段;

ii. 使所述基因组片段的至少一些与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制,并且其中所述至少一种终止子核苷酸为 α -硫代双脱氧核苷酸;

iii. 扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行,并且其中核酸复制的终止在链置换之前进行;

iv. 用甲基化特异性PCR扩增所述基因组片段的至少一些;

v. 将在步骤(iii和iv)中获得的分子连接至衔接子,从而产生基因组DNA文库和甲基化组DNA文库;以及

vi. 对所述基因组DNA文库和所述甲基化组文库进行测序。

单细胞分析

[0001] 交叉引用

[0002] 本申请要求于2019年7月31日提交的美国临时专利申请号62/881,183的权益,该临时申请的内容通过引用以其全文并入于此。

背景技术

[0003] 利用核酸扩增的研究方法,例如,下一代测序,提供了关于复杂样品、基因组和其他核酸来源的大量信息。在一些情况下,将这些样品以小量从单细胞获得。对于涉及小样品的研究、诊断和治疗,需要高度准确、可扩展且有效的核酸扩增和测序方法,尤其是用于同时分析RNA、DNA和蛋白质的方法。

发明内容

[0004] 本文提供了多组学单细胞分析的方法,所述方法包括:(a)从细胞群体中分离单细胞;(b)对包含从来自所述单细胞的mRNA转录物扩增的多核苷酸的cDNA文库进行测序;以及(c)对所述单细胞的基因组进行测序,其中对所述基因组进行测序包括:(i)使所述基因组与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制;和(ii)扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行;(iii)将在步骤(ii)中获得的分子连接至衔接子,从而产生基因组DNA文库;以及(iv)对所述基因组DNA文库进行测序。本文还提供了方法,其中所述mRNA转录物包括聚腺苷酸化mRNA转录物。本文还提供了方法,其中所述mRNA转录物不包括聚腺苷酸化mRNA转录物。本文还提供了方法,其中对cDNA文库进行测序包括用模板切换引物扩增mRNA转录物。本文还提供了方法,其中所述cDNA文库中所述多核苷酸中的至少一些包含条形码。本文还提供了方法,其中所述条形码包括细胞条形码或样品条形码。本文还提供了方法,其中在测序之前将所述cDNA文库和所述基因组DNA文库合并。本文还提供了方法,其中所述单细胞是原代细胞。本文还提供了方法,其中所述单细胞源自肝脏、皮肤、肾脏、血液或肺。本文还提供了方法,其中所述单细胞通过流式细胞术分离。本文还提供了方法,其中所述方法还包括从所述终止的扩增产物中除去至少一种终止子核苷酸。本文还提供了方法,其中所述多个终止的扩增产物的长度包含平均1000-2000个碱基。本文还提供了方法,其中所述多个终止的扩增产物的长度为250-1500个碱基。本文还提供了方法,其中所述多个终止的扩增产物包含所述单细胞的基因组的至少97%。本文还提供了方法,其中所述扩增产物中的至少一些包含细胞条形码或样品条形码。本文还提供了方法,其中对cDNA文库进行测序包括对所述单细胞的胞质(Cytosolic)裂解,以及逆转录。本文还提供了方法,其中所述mRNA转录物经由模板切换逆转录进行扩增。本文还提供了方法,其中所述cDNA文库包括至少10,000个基因。本文还提供了方法,其中对所述单细胞的基因组进行测序还包括对所述单细胞的核裂解。本文还提供了方法,其中所述方法还包括使用PCR的另外的扩增步骤。本文还提供了方法,其中在所述细胞的所述基因组中鉴定出至少一个突变,其中所述突变不同于参考序列中的

相应位置。本文还提供了方法,其中所述至少一个突变在所述细胞群体的少于1%中发生。本文还提供了方法,其中所述至少一个突变在所述细胞群体的不超过0.1%中发生。本文还提供了方法,其中所述至少一个突变在所述细胞群体的不超过0.001%中发生。本文还提供了方法,其中所述至少一个突变在所述扩增产物序列的不超过1%中发生。本文还提供了方法,其中所述至少一个突变在所述扩增产物序列的不超过0.1%中发生。本文还提供了方法,其中所述至少一个突变在所述扩增产物序列的不超过0.001%中发生。

[0005] 本文提供了多组学单细胞分析的方法,所述方法包括:(a)从细胞群体中分离单细胞;(b)鉴定所述单细胞的表面上的至少一种蛋白质;以及(c)对所述单细胞的基因组进行测序,其中对所述基因组进行测序包括:(i)使所述基因组与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制;(ii)扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行;(iii)将在步骤(ii)中获得的分子连接至衔接子,从而产生基因组DNA文库;以及(iv)对所述基因组DNA文库进行测序。本文还提供了方法,其中鉴定所述细胞表面上的至少一种蛋白质包括使所述细胞与结合至所述至少一种蛋白质的经标记的抗体接触。本文还提供了方法,其中所述经标记的抗体包含至少一种荧光标记或质量标签。本文还提供了方法,其中所述经标记的抗体包含至少一种核酸条形码。

[0006] 本文提供了多组学单细胞分析的方法,所述方法包括:(a)从细胞群体中分离单细胞;(b)对所述单细胞的基因组进行测序,其中对所述细胞的所述基因组进行测序包括:(i)用甲基化敏感性限制性酶消化所述基因组以产生基因组片段;(ii)使所述基因组片段的至少一些与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制;(iii)扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行;(iv)用甲基化特异性PCR扩增所述基因组片段的至少一些;(v)将在步骤(iii)和(iv)中获得的分子连接至衔接子,从而产生基因组DNA文库和甲基化组DNA文库;以及(vi)对所述基因组DNA文库和所述甲基化组文库进行测序。

[0007] 援引并入

[0008] 本说明书中提到的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别地且单独地指出每一个单独的出版物、专利或专利申请均通过引用而并入。

附图说明

[0009] 在所附权利要求书中具体阐述了本发明的新颖性特征。通过参考对在其中利用到本发明的原理的说明性实施方案加以阐述的以下详细描述和附图,将会获得对本发明的特征和优点的更好的理解,在附图中:

[0010] 图1A展示了用于分离分析来自单细胞的蛋白质、DNA和RNA的总体工作流概述。

[0011] 图1B展示了用于利用样品分拆以最小化交叉污染来分离分析来自单细胞的蛋白质、DNA和RNA的工作流。

[0012] 图1C展示了用于利用单管预扩增来分离分析来自单细胞的蛋白质、DNA和RNA的工作流。

[0013] 图1D展示了用于利用使用终止子以减小扩增子大小的单管预扩增来分离分析来自单细胞的蛋白质、DNA和RNA的工作流。

[0014] 图1E展示了用于利用共扩增来分离分析来自单细胞的蛋白质、DNA和RNA的工作流。

[0015] 图1F展示了组合来自本文所述的蛋白质/DNA/RNA单细胞实验的数据的信息学工作流。

[0016] 图1G展示了MDA与PTA-不可逆终止子方法在涉及突变传播时的比较。PTA方法导致原始DNA模板的直接拷贝数增加。

[0017] 图2A展示了扩增后进行的方法步骤,包括除去终止子、修复末端以及在衔接子连接之前进行加A尾。然后,在测序之前,可以在合并的细胞文库中对所有外显子或其他目的特定区域进行杂交介导的富集。每个读段的源细胞均通过细胞条形码(显示为绿色和蓝色序列)识别。

[0018] 图2B(GC)展示了MDA和PTA实验的测序碱基的GC含量比较。

[0019] 图2C展示了在单细胞经过PTA或MDA后,映射到人类基因组(p_映射的)的映射质量得分(e) (mapQ)。

[0020] 图2D在单细胞经过PTA或MDA后,映射到人类基因组(p_映射)的读段百分比。

[0021] 图2E(PCR)展示了在单细胞经过MDA和PTA后,2000万个亚取样读段中为PCR复制的读段百分比的比较。

[0022] 图2F展示了用于与PTA一起使用来RT扩增单细胞的工作流。

[0023] 图2G展示了来自通过RT获得的cDNA的文库的产生。

[0024] 图3A展示了在单细胞经过可逆或不可逆终止子PTA后,映射到人类基因组(p_映射的2)的映射质量得分(c) (mapQ2)。

[0025] 图3B展示了在单细胞经过可逆或不可逆终止子PTA后,映射到人类基因组(p_映射的2)的读段百分比。

[0026] 图3C展示了使用各种方法得到的一系列箱形图,描述了与Alu元件重叠的平均百分比读段的对齐读段。PTA具有与基因组对齐的最大读段数。

[0027] 图3D展示了使用各种方法得到的一系列箱形图,描述了与Alu元件重叠的平均百分比读段的PCR复制。

[0028] 图3E展示了使用各种方法得到的一系列箱形图,描述了与Alu元件重叠的平均百分比读段的读段GC含量。

[0029] 图3F展示了使用各种方法得到的一系列箱形图,描述了与Alu元件重叠的平均百分比读段的映射质量。PTA具有测试方法中最高的映射质量。

[0030] 图3G展示了在固定的7.5X测序深度下用不同的WGA方法时SC线粒体基因组覆盖宽度的比较。

[0031] 图4A展示了在将每个细胞下取样至4000万个配对读段后,在选择高质量MDA细胞(代表约50%细胞)后,与随机引物PTA扩增的细胞相比,对染色体1上的10千碱基窗口的平均覆盖深度。该图显示MDA的均匀性较差,更多窗口的覆盖深度比平均覆盖深度大(框A)或小(框C)两倍。由于重复区域的高GC含量和低映射质量,在着丝粒处MDA和PTA均无覆盖(框B)。

[0032] 图4B展示了MDA和PTA方法的测序覆盖与基因组位置的关系图(上图)。下部箱形图展示了与大量样品(bulk sample)相比,MDA和PTA方法的等位基因频率。

[0033] 图5A展示了基因组覆盖分数与基因组读段数目的关系图,用以评估各种方法在增加测序深度时的覆盖。PTA方法在每个深度下都接近两种大量样品,这是相对于其他测试方法的改进。

[0034] 图5B展示了基因组覆盖的变异系数与读段数目的关系图,用以评估覆盖均匀性。发现PTA方法在测试方法中具有最高的均匀性。

[0035] 图5C展示了总读段的累积分数与基因组的累积分数的洛伦兹图。发现PTA方法在测试方法中具有最高的均匀性。

[0036] 图5D展示了所测试的每种方法的计算的基尼指数的一系列箱形图,以便根据完全均匀性来估计每个扩增反应的差异。发现PTA方法比其他测试方法具有可再现的更高均匀性。

[0037] 图5E展示了判定的大量变体的分数与读段数目的关系图。在增加测序深度时,将每种方法的变体判定率与相应的大量样品进行比较。为了估计灵敏度,计算了在每个测序深度下,在每个细胞中发现的在被下取样至6.5亿个读段的相应大量样品中判定的变体百分比(图5E)。PTA的覆盖和均匀性的提高使得与灵敏度排第二的方法Q-MDA方法相比,可检测出多30%的变体。

[0038] 图5F展示了与Alu元件重叠的平均百分比读段的一系列箱形图。PTA方法显著减少了这些杂合位点的等位基因倾斜。相对于其他测试方法,PTA方法更均匀地扩增同一细胞中的两个等位基因。

[0039] 图5G展示了变体判定的特异性与读段数目的关系图,用以评估突变判定的特异性。使用各种方法发现的变体若在大量样品中未发现,则被认为是假阳性。在所测试的方法中,PTA方法的假阳性判定最低(特异性最高)。

[0040] 图5H展示了对于各种方法,每种类型碱基改变的假阳性碱基改变的比例。不受理论约束,这种模式可以是依赖于聚合酶的。

[0041] 图5I展示了对于假阳性变体判定,与Alu元件重叠的平均百分比读段的一系列箱形图。PTA方法产生了假阳性变体判定的最低等位基因频率。

[0042] 图6(A部分)展示了具有寡核苷酸的珠,该寡核苷酸附接有可切割的接头、独特细胞条形码和随机引物。B部分展示了单细胞和珠封装在同一液滴中,然后裂解细胞并切割引物。然后将液滴与包括PTA扩增混合物的另一液滴融合。C部分展示了扩增后液滴破裂,并且合并来自所有细胞的扩增子。然后将根据本公开内容的方案用于除去终止子、末端修复和加A尾,随后连接衔接子。然后,在测序之前,使合并的细胞文库经历针对目的外显子的杂交介导的富集。然后,使用细胞条形码识别每个读段的源细胞。

[0043] 图7A展示了用于使用PTA多组学(multiomic)(或多组学(polyomic))分析单细胞的工作流。步骤A:使细胞与包含荧光标记和寡核苷酸条形码标签的抗体接触。步骤B:基于荧光标记物分选细胞。步骤C:将管用结合核的抗体包被;将细胞裂解;胞质mRNA经历逆转录,而完整核结合至管壁。

[0044] 图7B展示了用于上接图7A的步骤C,使用PTA多组学分析单细胞的工作流。步骤D:在逆转录之后,取出RT级分用于测序分析。步骤E:将核裂解,并且在基因组DNA上进行PTA方

法。步骤F:PTA导致具有大约1000倍扩增的短片段cDNA池。

[0045] 图8A展示了在多组学DNA/RNA单细胞分析工作流程中用于逆转录和预扩增的引物。

[0046] 图8B展示了多组学DNA/RNA单细胞分析工作流程的逆转录和预扩增工作流程。使用了来自图8A的引物。

[0047] 图9A展示了亲本细胞系的生长速率的图,亲本细胞系用2nM奎扎替尼(2nM)处理三周的时间段以产生在FLT3抑制剂的存在下稳健生长的AML细胞系。然后将抗性单细胞和亲本单细胞(FACS富集的)通过RNA测序和低通量DNA测序分析进行分析。

[0048] 图9B展示了来自亲本培养物和抗性培养物两者的RNA表达证明了使用单锅RNAseq化学产生cDNA池(C)的能力,并且这些细胞中表达的基因产生能够通过每细胞检测的平均约10K个基因上的基因表达可视化细胞群体的不同图案。在单独的工作流程中,使用PTA方法扩增单细胞基因组。

[0049] 图9C展示了仅RNAseq的对照实验的归一化基因表达谱。

[0050] 图9D展示了通过PTA相对于不同方案产生的扩增DNA的量的图。相比于DNA,在RT步骤期间产生的转录物(R)未被PTA反应有效扩增,并且相比于来自单细胞的标准PTA扩增基因组(D、RD),使用组合方案(SC1-SC8)有效扩增单细胞中的DNA。NTC=无模板对照;R=RT步骤;D=PTA DNA步骤;RD=双重RT/PTA。

[0051] 图10A展示了针对使用低通量测序方案(约5百万个读段/细胞)的两种不同方案(双重RNAseq/PTA、标准RNAseq)的线粒体染色体量(%)。

[0052] 并且估计基因组大小大于30亿个碱基。

[0053] 图10B展示了针对使用低通量测序方案(约5百万个读段/细胞)的两种不同方案(双重RNAseq/PTA、标准RNAseq)的重复百分比。

[0054] 图10C展示了针对使用低通量测序方案(约5百万个读段/细胞)的两种不同方案(双重RNAseq/PTA、标准RNAseq)的估计基因组大小。

[0055] 图10D展示了使用双重RNAseq/PTA方案来自molm13细胞的3个scRNAseq数据集的特征分配。

[0056] 图10E展示了使用标准RNAseq方案获得的Sum159细胞系的归一化表达谱的图。P=亲本细胞。R=抗性细胞。

[0057] 图10F展示了使用双重RNAseq/PTA方案获得的Sum159细胞系的归一化表达谱的图。P=亲本细胞。R=抗性细胞。

[0058] 图11A展示了进行至25x的近似深度(K)的7个亲本和5个抗性molm13细胞的深度测序的结果。使用bwa mem将读段与Hg38比对。使用GATK4最佳操作进行质量控制和SNV判定。仅在如果SNV限于至少2个抗性细胞情况下考虑SNV,在任何亲本细胞中未判定替代性等位基因,并且对至少6个亲本细胞基因型化。所有细胞在1x覆盖率下覆盖了至少96%的基因组,并且在10x下覆盖至少76%。插图展示了在所有细胞(为清楚起见展示了4个)中检测出molm13细胞中的已知F1t3插入缺失。

[0059] 图11B展示了基因表达谱的热图,所述基因表达谱包括过表达基因GAS6,其是奎扎替尼抗性的已知机制。Gas6是AXL的配体,AXL是经历奎扎替尼治疗失败的复发患者中的临床相关抗性机制。

[0060] 图12A展示了大量样品相对于单细胞样品中覆盖的外显子的比例的图。

- [0061] 图12B展示了大量样品相对于单细胞样品中没有覆盖的外显子的比例的图。
- [0062] 图12C展示了大量样品相对于单细胞样品中选择的碱基百分比的图。
- [0063] 图12D展示了大量样品相对于单细胞样品中20X下覆盖的碱基的比例的图。
- [0064] 图13A展示了按处理分层且按样品类型加阴影的基因组中的映射的读段碱基的位置的图。
- [0065] 图13B展示了样品强度相对于捕获的插入物大小的图。
- [0066] 图14A展示了对于12-plex实验的重复百分比相对于选择的碱基百分比的图。
- [0067] 图14B展示了靶碱基数量相对于覆盖水平的图。

具体实施方式

[0068] 需要开发新的可扩展、准确且有效的核酸扩增(包括单细胞和多细胞基因组扩增)和测序方法,其将通过以可再现的方式增加序列呈现、均匀性和准确性来克服当前方法的局限性。本文提供了用于提供准确且可扩展的原代模板定向扩增(PTA)和测序的组合物和方法。本文还提供了多组学分析的方法,所述方法包括与PTA组合分析来自单细胞的蛋白质、DNA和RNA,以及相应的转录后或翻译后修饰。这些方法和组合物促进靶标(或“模板”)核酸的高精度扩增,这增加了下游应用(如下一代测序)的准确性和灵敏性。

[0069] 定义

[0070] 除非另有定义,否则本文使用的所有技术和科学术语的含义与这些发明所属领域的普通技术人员通常所理解的含义相同。

[0071] 在整个本公开内容中,数字特征以范围格式表示。应理解,范围格式的描述仅为了方便和简洁,而不应被解释为对任何实施方案的范围的严格限制。因此,除非上下文另有明确指示,否则应认为对范围的描述已具体公开了所有可能的子范围以及在该范围内直至下限单位的十分之一的单个数值。例如,对范围如从1至6的描述应视为已具体公开了子范围,如从1至3、从1至4、从1至5、从2至4、从2至6、从3至6等,以及该范围内的单个值,例如,1.1、2、2.3、5和5.9。无论范围的宽度如何,这都适用。这些中间范围的上限和下限可以独立地包括在较小的范围内,并且也包括在本发明内,受规定范围内任何明确排除的限制。当所述范围包括一个或两个限值时,除非上下文另外明确指出,否则不包含所包括的那些限值中的一个或两个的范围也包括在本发明中。

[0072] 本文所使用的术语仅出于描述特定实施方案的目的,并且不旨在限制任何实施方案。如本文所用,单数形式“一种”、“一个”和“所述”也旨在包括复数形式,除非上下文另有明确指出。还将理解,当在本说明书中使用术语“包括(comprises)”和/或“包含(comprising)”时,其指定了所述特征、整体、步骤、操作、要素和/或组件的存在,但并不排除一个或多个其他特征、整体、步骤、操作、要素、组件和/或其群组的存在或添加。如本文所用,术语“和/或”包括一个或多个相关联的所列项目的任何和所有组合。

[0073] 除非特别说明或从上下文可以明显看出,否则如本文所用,术语“约”在提及数值或数值范围时应理解是指所述数值 \pm 其10%,或对于一个范围中列出的值,则指从比列出的下限低10%到比列出的上限高10%。

[0074] 如本文所用,术语“对象”或“患者”或“个体”是指动物,包括哺乳动物,如人类、兽医动物(例如,猫、狗、牛、马、绵羊、猪等)和疾病实验动物模型(例如,小鼠、大鼠)。根据本发

明,可以在本领域技术范围内使用常规分子生物学、微生物学和重组DNA技术。这些技术在文献中有充分的解释。参见,例如,Sambrook,Fritsch和Maniatis,Molecular Cloning:A Laboratory Manual,第二版(1989)Cold Spring Harbor Laboratory Press,Cold Spring Harbor,New York(本文中为“Sambrook等人,1989”);DNA Cloning:A practical Approach,第I和II卷(D.N.Glover著,1985);Oligonucleotide Synthesis(MJ.Gait著,1984);Nucleic Acid Hybridization(B.D.Hames和S.J.Higgins著,(1985));Transcription and Translation(B.D.Hames和S.J.Higgins著,(1984));Animal Cell Culture(R.I.Freshney著,(1986));Immobilized Cells and Enzymes(1RL出版社(1986));B.Perbal,A practical Guide To Molecular Cloning(1984);F.M.Ausubel等人(著),Current Protocols in Molecular Biology,John Wiley&Sons,Inc.(1994);等等。

[0075] 术语“核酸”涵盖多链以及单链分子。在双链或三链核酸中,核酸链不需要是共延伸的(即,双链核酸不需要沿两条链的整个长度是双链的)。本文所述的核酸模板可以根据样品(从小的无细胞DNA片段至整个基因组)而具有任何大小,包括但不限于长度为50-300个碱基、100-2000个碱基、100-750个碱基、170-500个碱基、100-5000个碱基、50-10,000个碱基或50-2000个碱基。在一些情况下,模板的长度为至少50、100、200、500、1000、2000、5000、10,000、20,000、50,000、100,000、200,000、500,000、1,000,000或大于1,000,000个碱基。本文所述的方法提供了核酸如核酸模板的扩增。本文所述的方法另外提供了分离的和至少部分纯化的核酸以及核酸文库的生成。在一些情况下,本文所述的方法提供了提取的核酸(例如,从组织、细胞或培养基中提取)。核酸包括但不限于包括以下的那些:DNA、RNA、环状RNA、mtDNA(线粒体DNA)、cfDNA(无细胞DNA)、cfRNA(无细胞RNA)、siRNA(小干扰RNA)、cffDNA(无细胞胎儿DNA)、mRNA、tRNA、rRNA、miRNA(微RNA)、合成多核苷酸、多核苷酸类似物、符合本说明书的任何其他核酸,或其任何组合。当提供时,多核苷酸的长度以碱基数描述并且进行缩写,如nt(核苷酸)、bp(碱基)、kb(千碱基)或Gb(千兆碱基)。

[0076] 如本文所用,术语“液滴”是指液滴致动器上的一定体积液体。在一些情况下,例如,液滴是水性或非水性的,或者可以是包括水性和非水性组分的混合物或乳液。对于可经受液滴操作的液滴流体的非限制性示例,参见,例如,国际专利申请公开号W02007/120241。在本文提出的实施方案中,可以使用任何适于形成和操纵液滴的系统。例如,在一些情况下,使用液滴致动器。对于可以使用的液滴致动器的非限制性示例,参见,例如,美国专利号6,911,132、6,977,033、6,773,566、6,565,727、7,163,612、7,052,244、7,328,979、7,547,380、7,641,779、美国专利申请公开号US20060194331、US20030205632、US20060164490、US20070023292、US20060039823、US20080124252、US20090283407、US20090192044、US20050179746、US20090321262、US20100096266、US20110048951、国际专利申请公开号W02007/120241。在一些情况下,珠在液滴中、在液滴操作间隙中或在液滴操作表面上提供。在一些情况下,珠在位于液滴操作间隙外部或与液滴操作表面分开的储器中提供,并且该储器可以与流动路径相关联,该流动路径允许包括该珠的液滴进入液滴操作间隙或与液滴操作表面接触。用于固定磁响应珠和/或非磁响应珠和/或使用珠进行液滴操作方案的液滴致动器技术的非限制性示例在美国专利申请公开号US20080053205、国际专利申请公开号W02008/098236、W02008/134153、W02008/116221、W02007/120241中描述。珠特性可以在本文所述的方法的多路复用实施方案中采用。具有适合于多路复用的特性的珠的示例,以及

检测和分析从这种珠发出的信号的方法,可以在美国专利申请公开号US20080305481、US20080151240、US20070207513、US20070064990、US20060159962、US20050277197、US20050118574中找到。

[0077] 也可以将引物和/或模板切换寡核苷酸附连至固体基底以促进mRNA多核苷酸的逆转录和模板切换。在该布置中,RT或模板切换反应的一部分发生在装置的本体溶液中,其中反应的第二步骤接近于表面进行。在其他布置中,使模板切换寡核苷酸的引物从固体基底中释放以允许整个反应在溶液中的表面上方进行。在多组学方法中,在一些情况下,将用于多级反应的引物附连至到固体基底或与珠组合以实现多级引物的组合。

[0078] 某些微流体装置也支持多组学方法。例如,以PDMS制造的装置通常具有用于每个反应步骤的连续腔室。此类多腔室化装置通常使用微型阀结构进行隔离,所述微型阀结构可以用空气或流体如水或惰性烃(即fluorinert)通过压力控制。在多组学方法中,可以将反应的每个级隔离并且允许其离散地进行。在完成特定级时,可以释放相邻腔室之间的阀,并且可以以串联方式添加用于后续反应的基底。结果是能够使用单独细胞作为输入模板材料来模拟反应的序列组,如反应的多组学(蛋白质/RNA/DNA/表观基因组学)组。各种微流体平台可用于分析单细胞。在一些情况下,通过流体动力学(液滴微流体、惯性微流体、涡旋、微型阀、微观结构(例如,微孔、微阱))、电方法(电泳(DEP)、电渗)、光学方法(光学镊子、光学诱导介电泳(ODEP)、光学热毛细)、声学方法或磁性方法操纵细胞。在一些情况下,微流体平台包括微孔。在一些情况下,微流体平台包括基于PDMS(聚二甲基硅氧烷)的装置。与本文所述的方法相容的单细胞分析平台的非限制示例是:ddSEQ单细胞分离器(Bio-Rad, Hercules, CA, USA和Illumina, San Diego, CA, USA);Chromium(10x Genomics, Pleasanton, CA, USA);Rhapsody单细胞分析系统(BD, Franklin Lakes, NJ, USA);Tapestri平台(MissionBio, San Francisco, CA, USA);Nadia Innovate(Dolomite Bio, Royston, UK);C1和Polaris(Fluidigm, South San Francisco, CA, USA);ICELL8单细胞系统(Takara);MSND(Wafergen);Puncher平台(Vycap);CellRaft AIR系统(CellMicrosystems);DEPArray NxT和DEPArray系统(Menarini Silicon Biosystems);AVISO CellCelector(ALS);以及InDrop系统(1CellBio)和TrapTx(Celldom)。

[0079] 如本文所用,术语“独特分子标识码(UMI)”是指附接于多个核酸分子中的每一个的独特核酸序列。当并入核酸分子中时,在一些情况下,UMI被用于通过直接对扩增后测序的UMI进行计数来校正后续的扩增偏倚。UMI的设计、并入和应用描述于例如以下文献中:国际专利申请公开号W0 2012/142213;Islam等人Nat.Methods(2014)11:163-166;Kivioja, T.等人Nat.Methods(2012)9:72-74;Brenner等人(2000)PNAS 97(4),1665;以及Hollas和Schuler,(2003)Conference:3rd International Workshop on Algorithms in Bioinformatics,第2812卷。

[0080] 如本文所用,术语“条形码”是指可用于鉴定核酸材料的样品或来源的核酸标签。因此,在核酸样品来自多个来源的情况下,在一些情况下,每个核酸样品中的核酸用不同的核酸标签标记,从而可以鉴定样品的来源。条形码,通常也称为索引、标签等,是本领域技术人员熟知的。任何合适的条形码或条形码组都可以使用。参见,例如,美国专利号8,053,192和国际专利申请公开号W02005/068656中提供的非限制性示例。单细胞的条形码化可以例如如美国专利申请公开号2013/0274117中描述地进行。

[0081] 本文中的术语“固体表面”、“固体支持物”和其他语法等同物是指适合于或可以被修饰以适合于本文所述的引物、条形码和序列的附接的任何材料。示例性基底包括但不限于玻璃和改性或功能化玻璃、塑料(包括丙烯酸、聚苯乙烯,和苯乙烯与其他材料的共聚物、聚丙烯、聚乙烯、聚丁烯、聚氨酯、特氟龙™等)、多糖、尼龙、硝酸纤维素、陶瓷、树脂、二氧化硅、二氧化硅基材料(例如,硅或改性硅)、碳、金属、无机玻璃、塑料、光纤束和各种其他聚合物。在一些实施方案中,固体支持物包括适于以有序模式固定引物、条形码和序列的图案化表面。

[0082] 如本文所用,术语“生物样品”包括但不限于组织、细胞、生物流体及其分离物。在一些情况下,本文所述方法中使用的细胞或其他样品是从人类患者、动物、植物、土壤或包括如细菌、真菌、原生动物等微生物的其他样品中分离的。在一些情况下,生物样品来源于人类。在一些情况下,生物样品并非来源于非人类。在一些情况下,细胞经历本文所述的PTA方法和测序。在整个基因组或特定位置检测到的变体可以与从该对象分离的所有其他细胞进行比较,以追踪细胞谱系的历史,以用于研究或诊断目的。在一些情况下,通过另外的分析方法如直接PCR测序确认变体。

[0083] 单细胞分析

[0084] 本文描述了用于分析单细胞的方法和组合物。大量细胞的分析提供有关细胞群体的一般信息,但通常无法在背景上检测低频率突变体。此类突变体可以包括重要特性,如抗药性或与癌症相关的突变。在一些情况下,平行分析来自相同单细胞的DNA、RNA和/或蛋白质。分析可以包括鉴定表观遗传的翻译后修饰(例如,糖基化、磷酸化、乙酰化、泛素化、组蛋白修饰)和/或转录后修饰(例如,甲基化、羟甲基化)。此类方法可以包括“原代模板定向扩增”(PTA),以获得用于测序的核酸文库。在一些情况下,将PTA与另外的步骤或方法组合,所述另外的步骤或方法如RT-PCR或蛋白质组/蛋白质定量技术(例如,质谱、抗体染色等)。在一些情况下,在单独分析步骤期间,细胞的各种组分在物理或空间上彼此分离。例如,在一些情况下,在工作流包括图1A中的总体步骤。首先将蛋白质用抗体标记。在一些情况下,至少一些抗体包含标签或标记物(例如,核酸/寡核苷酸标签、质量标签或荧光标签)。在一些情况下,抗体的一部分包含寡核苷酸标签。在一些情况下,抗体的一部分包含荧光标记物。在一些情况下,抗体被两个或更多个标签或标记物标记。在一些情况下,基于荧光标记物分选抗体的一部分。在RT-PCR之后,产生第一链mRNA产物,且然后取出以进行分析。然后由RT-PCR产物和存在于蛋白质特异性抗体上的条形码产生文库,随后对其进行测序。同时,对来自同一细胞的基因组DNA进行PTA,产生文库,并且进行测序。在一些情况下,将来自基因组、蛋白质组和转录物组的测序结果使用生物信息学方法合并。在一些情况下,本文所述的方法包括标记、细胞分选、亲和分离/纯化、特定细胞组分(例如,外膜、核等)的裂解、RNA扩增、DNA扩增(例如,PTA)或与蛋白质、RNA或DNA分离或分析相关的其他步骤的任何组合。在一些情况下,本文所述的方法包括一种或多种富集步骤,如外显子组富集。

[0085] 本文描述了单细胞分析的第一方法,所述第一方法包括分析来自单细胞的RNA和DNA(图1B)。在一些情况下,所述方法包括分离单细胞、裂解单细胞和逆转录(RT)。在一些情况下,用模板切换寡核苷酸(TSO)进行逆转录。在一些情况下,TSOS包含分子标签如生物素,从而允许随后下拉cDNA RT产物,并且PCR扩增RT产物以产生cDNA文库。可替代地或组合地,使用离心来将上清液中的RNA与细胞沉淀中的cDNA分离。在一些情况下,将剩余的cDNA片段

化并且用UDG(尿嘧啶DNA糖基糖酶)除去,并且使用碱性裂解来降解RNA并使基因组变性。在中和、添加引物和PTA之后,在一些情况下,将扩增产物在SPRI(固相可逆固定)珠上纯化,并且连接至衔接子以产生gDNA文库。

[0086] 本文描述了单细胞分析的第二方法,所述第二方法包括分析来自单细胞的RNA和DNA(图1C)。在一些情况下,所述方法包括分离单细胞、裂解单细胞和逆转录(RT)。在一些情况下,用模板切换寡核苷酸(TSO)进行逆转录。在一些情况下,TSOS包含分子标签如生物素,从而允许随后下拉cDNA RT产物,并且PCR扩增RT产物以产生cDNA文库。在一些情况下,然后使用碱性裂解来降解RNA并使基因组变性。在中和、添加随机引物和PTA之后,在一些情况下,将扩增产物在SPRI(固相可逆固定)珠上纯化,并且连接至衔接子以产生gDNA文库。在一些情况下,将RT产物通过下拉,如用链霉亲和素珠的下拉分离。

[0087] 本文描述了单细胞分析的第三方法,所述第三方法包括分析来自单细胞的RNA和DNA(图1D)。在一些情况下,所述方法包括分离单细胞、裂解单细胞和逆转录(RT)。在一些情况下,在终止子核苷酸的存在下用模板切换寡核苷酸(TSO)进行逆转录。在一些情况下,TSOS包含分子标签如生物素,从而允许随后下拉cDNA RT产物,并且PCR扩增RT产物以产生cDNA文库。在一些情况下,然后使用碱性裂解来降解RNA并使基因组变性。在中和、添加随机引物和PTA之后,在一些情况下,将扩增产物在SPRI(固相可逆固定)珠上纯化,并且连接至衔接子以产生DNA文库。在一些情况下,将RT产物通过下拉,如用链霉亲和素珠的下拉分离。

[0088] 本文描述了单细胞分析的第四方法,所述第四方法包括分析来自单细胞的RNA和DNA(图1E)。在一些情况下,所述方法包括分离单细胞、裂解单细胞和逆转录(RT)。在一些情况下,用模板切换寡核苷酸(TSO)进行逆转录。在一些情况下,TSOS包含分子标签如生物素,从而允许随后下拉cDNA RT产物,并且PCR扩增RT产物以产生cDNA文库。在一些情况下,然后使用碱性裂解来降解RNA并使基因组变性。在中和、添加随机引物和PTA之后,在一些情况下,使用封闭且经标记的引物使扩增产物经受RNA酶和cDNA扩增。将gDNA在SPRI(固相可逆固定)珠上纯化,并且连接至衔接子以产生gDNA文库。在一些情况下,将RT产物通过下拉,如用链霉亲和素珠的下拉分离。

[0089] 本文描述了单细胞分析的第五方法,所述第五方法包括分析来自单细胞的RNA和DNA(图7A和图7B)。将细胞群体与抗体文库接触,其中抗体被标记。在一些情况下,将抗体用荧光标记、核酸条形码或两者标记。经标记的抗体与群体中的至少一种细胞结合,并且将此类细胞分选,每个容器(例如,管、小瓶、微孔等)放置一个细胞。在一些情况下,容器包括溶剂。在一些情况下,将容器表面的区域用捕获部分包被。在一些情况下,捕获部分是能够与一个或多个细胞、细胞器或其他细胞组分结合的小分子、抗体、蛋白质或其他药剂。在一些情况下,至少一个细胞、单细胞、或其组分与容器表面的区域结合。在一些情况下,核与容器的区域结合。在一些情况下,将细胞的外膜裂解,从而将mRNA释放至容器中的溶液中。在一些情况下,含有基因组DNA的细胞的核结合至容器表面的区域。接下来,通常使用溶液中的mRNA作为模板来进行RT以产生cDNA。在一些情况下,模板切换引物从5'至3'包含TSS区域(转录起始位点)、锚定区域、RNA BC区域和聚dT尾。在一些情况下,聚dT尾与一个或多个mRNA的聚A尾结合。在一些情况下,模板切换引物从3'至5'包含TSS区域、锚定区域和聚G区域。在一些情况下,聚G区域包括核糖G。在一些情况下,聚G区域与mRNA转录物上的聚C区域结合。在一些情况下,通过末端转移酶将核糖G添加至mRNA转录物。在取出RT PCR产物以进

行后续测序之后,通过UNG除去细胞中的任何剩余RNA。然后将核裂解,并且使用随机引物和等温聚合酶对释放的基因组DNA进行PTA方法。在一些情况下,引物的长度为6-9个碱基。在一些情况下,PTA产生长度为100-5000、200-5000、500-2000、500-2500、1000-3000或300-3000个碱基的基因组扩增子。在一些情况下,PTA产生平均长度为100-5000、200-5000、500-2000、500-2500、1000-3000或300-3000个碱基的基因组扩增子。在一些情况下,PTA产生长度为250-1500个碱基的基因组扩增子。在一些情况下,本文所述的方法产生具有约500、约750、约1000、约5000或约10,000倍扩增的短片段cDNA池。在一些情况下,本文所述的方法产生具有500-5000、750-1500或250-10,000倍扩增的短片段cDNA池。任选地对PTA产物进行另外的扩增,并且进行测序。

[0090] 单细胞的样品制备和分离

[0091] 本文所述的方法可能需要分离单细胞以进行分析。任何单细胞分离方法可以与PTA一起使用,如口腔移液、微移液、流式细胞术/FACS、微流体、分选核(四倍体或其他)的方法或手动稀释。此类方法通过另外的试剂和步骤辅助,所述另外的试剂和步骤例如基于抗体的富集(例如,循环肿瘤细胞)、基于其他小分子或蛋白质的富集方法或荧光标记。在一些情况下,本文所述的多组学分析方法包括将细胞从较大组织中的机械或酶促解离。

[0092] 细胞组分的制备和分析

[0093] 包括本文所述的PTA的多组学分析方法可以包括处理细胞组分如DNA、RNA和/或蛋白质的一种或多种方法。在一些情况下,将核(包含基因组DNA)与胞质溶胶(包含mRNA)物理分离,然后是膜选择性裂解缓冲剂以溶解膜,但保持核完整。然后使用包括微移液、离心或抗体缀合的磁性微珠的方法将胞质溶胶与核分离。在另一种情况下,寡dT引物包被的磁珠结合聚腺苷酸化mRNA以便与DNA分离。在另一种情况下,同时预扩增DNA和RNA,并且然后分离以进行分析。在另一种情况下,将单细胞被拆分成两个相等部分,其中从一半处理mRNA并且从另一半处理基因组DNA。

[0094] 多组学

[0095] 本文所述的方法(例如,PTA)可以用作本领域中用于单细胞测序(多组学等)的任何数量其他已知方法的替代物。PTA可以代替基因组DNA测序方法,如MDA、PicoPlex、DOP-PCR、MALBAC或靶特异性扩增。在一些情况下,PTA替代多组学方法中的标准基因组DNA测序方法,包括DR-seq(Dey等人,2015)、G&T-seq(MacAulay等人,2015)、scMT-seq(Hu等人,2016)、sc-GEM(Cheow等人,2016)、scTrio-seq(Hou等人,2016)、RNA和蛋白质同时多路复用测量(Darmanis等人,2016)、scCOOL-seq(Guo等人,2017)、CITE-seq(Stoeckius等人,2017)、REAP-seq(Peterson等人,2017)、scNMT-seq(Clark等人,2018)或SIDR-seq(Han等人,2018)。在一些情况下,本文所述的方法包括PTA和聚腺苷酸化mRNA转录物的方法。在一些情况下,本文所述的方法包括PTA和非聚腺苷酸化mRNA转录物的方法。在一些情况下,本文所述的方法包括PTA和总(聚腺苷酸化和非聚腺苷酸化)mRNA转录物的方法。

[0096] 在一些情况下,将PTA与标准RNA测序方法组合以获得基因组和转录物组数据。在一些情况下,本文所述的多组学方法包括PTA和以下之一:Drop-seq(Macosko等人2015)、mRNA-seq(Tang等人,2009)、InDrop(Klein等人,2015)、MARS-seq(Jaitin等人,2014)、Smart-seq2(Hashimshony等人,2012;Fish等人,2016)、CEL-seq(Jaitin等人,2014)、STRT-seq(Islam等人,2011)、Quartz-seq(Sasagawa等人,2013)、CEL-seq2(Hashimshony等人

2016)、cytoSeq (Fan等人, 2015)、SuPeR-seq (Fan等人, 2011)、RamDA-seq (Hayashi等人 2018)、MATQ-seq (Sheng等人, 2017) 或SMARTer (Verboom等人, 2019)。

[0097] 各种反应条件和混合物可用于产生用于转录组分析的cDNA文库。在一些情况下, 使用RT反应混合物来产生cDNA文库。在一些情况下, RT反应混合物包含拥挤试剂、至少一种引物、模板切换寡核苷酸 (TSO)、逆转录酶和dNTP混合物。在一些情况下, RT反应混合物包含RNA酶抑制剂。在一些情况下, RT反应混合物包含一种或多种表面活性剂。在一些情况下, RT反应混合物包含Tween-20和/或Triton-X。在一些情况下, RT反应混合物包含甜菜碱。在一些情况下, RT反应混合物包含一种或多种盐。在一些情况下, RT反应混合物包含镁盐 (例如, 氯化镁) 和/或四甲基氯化铵。在一些情况下, RT反应混合物包含明胶。在一些情况下, RT反应混合物包含PEG (PEG1000、PEG2000、PEG4000、PEG6000、PEG8000或其他长度的PEG)。

[0098] 本文所述的多组学方法可以提供来自单细胞的基因组信息和RNA转录物信息两者 (例如, 组合或双重方案)。在一些情况下, 来自单细胞的基因组信息从PTA方法获得, 并且RNA转录物信息从用于产生cDNA文库的逆转录获得。在一些情况下, 使用全转录物方法来获得cDNA文库。在一些情况下, 使用3' 或5' 端计数来获得cDNA文库。在一些情况下, 未使用UMI获得cDNA文库。在一些情况下, 多组学方法提供来自单细胞的关于至少500、1000、2000、5000、8000、10,000、12,000或至少15,000个基因的RNA转录物信息。在一些情况下, 多组学方法提供来自单细胞的关于约500、1000、2000、5000、8000、10,000、12,000或约15,000个基因的RNA转录物信息。在一些情况下, 多组学方法提供来自单细胞的关于100-12,000、1000-10,000、2000-15,000、5000-15,000、10,000-20,000、8000-15,000或10,000-15,000个基因的RNA转录物信息。在一些情况下, 多组学方法提供关于单细胞的基因组的至少80%、90%、92%、95%、97%、98%或至少99%的基因组序列信息。在一些情况下, 多组学方法提供关于单细胞的基因组的约80%、90%、92%、95%、97%、98%或约99%的基因组序列信息。

[0099] 多组学方法可以包括对来自细胞群体的单细胞的分析。在一些情况下, 分析了至少5、10、20、50、100、200、500、1000、2000、5000或至少8000个细胞。在一些情况下, 分析了约5、10、20、50、100、200、500、1000、2000、5000或约8000个细胞。在一些情况下, 分析了5-100、10-100、50-500、100-500、100-1000、50-5000、100-5000、500-1000、500-10000、1000-10000或5000-20,000个细胞。

[0100] 多组学方法可以基于单细胞的类型从PTA反应产生基因组DNA的产量。在一些情况下, 从单细胞产生的DNA的量是约0.1、1、1.5、2、3、5或约10微克。在一些情况下, 从单细胞产生的DNA的量是约0.1、1、1.5、2、3、5或约10飞克。在一些情况下, 从单细胞产生的DNA的量是至少0.1、1、1.5、2、3、5或至少10微克。在一些情况下, 从单细胞产生的DNA的量是至少0.1、1、1.5、2、3、5或至少10飞克。在一些情况下, 从单细胞产生的DNA的量是约0.1-10、1-10、1.5-10、2-20、2-50、1-3或0.5-3.5微克。在一些情况下, 从单细胞产生的DNA的量是约0.1-10、1-10、1.5-10、2-20、2-4、1-3或0.5-4飞克。

[0101] 甲基化组分析

[0102] 本文描述了包括PTA的方法, 其中使用PTA方法确定单细胞中甲基化DNA的位点。在一些情况下, 这些方法还包括平行分析同一细胞的转录物组和/或蛋白质组。检测甲基化基因组碱基的方法包括用甲基化敏感性内切核酸酶进行选择限制, 然后用PTA方法处理。根

据测序确定被此类酶切割的位点,并且鉴定甲基化碱基。在另一种情况下,基因组DNA文库的亚硫酸氢盐处理将未甲基化胞嘧啶转化为尿嘧啶。然后在一些情况下,用选择性地退火至甲基化序列的甲基化特异性引物扩增文库。可替代地,进行非甲基化特异性PCR,然后进行用于区分亚硫酸氢盐反应的碱基的一种或多种方法,包括直接焦磷酸测序、MS-SnuPE、HRM、COBRA、MS-SSCA或碱基特异性切割/MALDI-TOF。在一些情况下,将基因组DNA样品拆分以用于基因组(或其富集部分)和甲基化组分析的平行分析。在一些情况下,基因组和甲基化组的分析包括富集基因组片段(例如,外显子组或其他靶标)或全基因组测序。

[0103] 生物信息学

[0104] 可以将从本文所述的利用PTA的单细胞分析方法获得的数据编译成数据库。本文描述了生物信息学数据整合的方法和系统。在一些情况下将来自蛋白质组、基因组、转录物组、甲基化组或其他数据的数据组合/整合为数据库并分析。在一些情况下,生物信息学数据整合方法和系统包括蛋白质检测(FACS和/或NGS)、mRNA检测和/或基因组变异检测中的一种或多种。在一些情况下,该数据与疾病状态或病症相关。在一些情况下,将来自多个单细胞的数据编译以描述较大细胞群体(如来自特定样品、区域、生物体或组织的细胞)的特性。在一些情况下,从与细胞上的蛋白质选择性结合的经荧光标记的抗体获取蛋白质数据。在一些情况下,蛋白质检测的方法包括基于荧光标记物将细胞分组并且在分选后报告样品位置。在一些情况下,蛋白质检测的方法包括检测样品条形码、检测蛋白质条形码、与设计的序列进行比较并且基于条形码和拷贝数将细胞分组。在一些情况下,从与细胞上的蛋白质选择性结合的条形码化抗体获取蛋白质数据。在一些情况下,从样品和RNA特异性条形码获取转录物组数据。在一些情况下,mRNA检测的方法包括检测样品和RNA特异性条形码、与基因组比对、与RefSeq/Encode比对、报告外显子/内含子/基因间序列、分析外显子-外显子连接点、基于条形码和表达变异将细胞分组以及聚类分析变异和靠前可变基因。在一些情况下,从样品和DNA特异性条形码获取基因组数据。在一些情况下,基因组变异检测的方法包括检测样品和DNA特异性条形码、与基因组比对、确定基因组回收率和SNV映射率、过滤外显子-外显子连接点上的读段、产生变体判定文件(VCF)和聚类分析变异和靠前可变突变。

[0105] 突变

[0106] 在一些情况下,本文所述的方法(例如,多组学PTA)导致突变检测的较高检测灵敏度和/或较低假阳性率。在一些情况下,突变是分析序列(例如,使用本文所述的方法)与参考序列之间的差异。在一些情况下,参考序列从其他生物体、相同或相似物种的其他个体、生物体群体或同一基因组的其他区域获得。在一些情况下,突变在质粒或染色体上鉴定。在一些情况下,突变是SNV(单核苷酸变异)、SNP(单核苷酸多态性)或CNV(拷贝数变异或CNA/拷贝数异常)。在一些情况下,突变是碱基取代、插入或缺失。在一些情况下,突变是转变、颠换、无义突变、沉默突变、同义或非同义突变、非致病性突变、错义突变或移码突变(缺失或插入)。在一些情况下,当相比于诸如经由计算机模拟预测、ChIP-seq、GUIDE-seq、环状-seq、HTGTS(高通量全基因组易位测序)、IDLV(整合缺陷型慢病毒)、Digenome-seq、FISH(荧光原位杂交)或DISCOVER-seq的方法时,PTA导致突变检测的较高检测灵敏度和/或较低假阳性率。

[0107] 原代模板定向扩增

[0108] 本文描述了核酸扩增方法,如“原代模板定向扩增(PTA)”。在一些情况下,将PTA与

其他分析工作流程组合用于多组学分析。例如,本文所述的PTA方法的一个实施方案示意性地表示于图1G中。在PTA方法中,使用聚合酶(例如,链置换聚合酶)优先从原代模板(“直接拷贝”)生成扩增子。因此,与MDA相比,在随后的扩增过程中,错误以较低的速率从子扩增子传播。结果得到一种易于执行的方法,与现有的WGA方案不同,所述方法可以以准确且可再现的方式扩增低输入量的DNA(包括单细胞的基因组),且具有高覆盖范围和均匀性。此外,终止的扩增产物可以在除去终止子后进行定向连接,允许细胞条形码附接至扩增引物,从而可以在进行平行扩增反应后合并来自所有细胞的产物。在一些情况下,模板核酸未结合至固体支持物。在一些情况下,模板核酸的直接拷贝未结合至固体支持物。在一些情况下,一种或多种引物未结合至固体支持物。在一些情况下,没有引物未结合至固体支持物。在一些情况下,引物附接至第一固体支持物,并且模板核酸附接至第二固体支持物,其中第一固体支持物和第二固体支持物不是相同的。在一些情况下,使用PTA来于分析来自较大细胞群体的单细胞。在一些情况下,使用PTA来分析来自较大细胞体的多于一种细胞、或整个细胞群体。

[0109] 本文描述了使用具有链置换活性的核酸聚合酶进行扩增的方法。在一些情况下,这种聚合酶具有链置换活性和低错误率。在一些情况下,这种聚合酶具有链置换活性和校对核酸外切酶活性,如3' → 5' 校对活性。在一些情况下,核酸聚合酶与其他组分结合使用,所述其他组分如可逆或不可逆终止子,或其他链置换因子。在一些情况下,聚合酶具有链置换活性,但不具有核酸外切酶校对活性。例如,在一些情况下,这些聚合酶包括噬菌体phi29 (Φ 29) 聚合酶,其也有非常低的错误率,这是3' → 5' 校对核酸外切酶活性的结果(参见,例如,美国专利号5,198,543和5,001,050)。在一些情况下,链置换核酸聚合酶的非限制性示例包括,例如,经基因修饰的phi29 (Φ 29) DNA聚合酶、DNA聚合酶I的Klenow片段(Jacobsen等人, *Eur. J. Biochem.* 45:623-627 (1974))、噬菌体M2DNA聚合酶(Matsumoto等人, *Gene* 84:247 (1989))、噬菌体phiPRD1DNA聚合酶(Jung等人, *Proc. Natl. Acad. Sci. USA* 84:8287 (1987); Zhu和Ito, *Biochim. Biophys. Acta.* 1219:267-276 (1994))、Bst DNA聚合酶(例如, Bst大片段DNA聚合酶(Exo(-) Bst; Aliotta等人, *Genet. Anal. (Netherlands)* 12:185-195 (1996))、exo(-) Bca DNA聚合酶(Walker和Linn, *Clinical Chemistry* 42:1604-1608 (1996))、Bsu DNA聚合酶、包括Vent_R(exo-) DNA聚合酶的Vent_R DNA聚合酶(Kong等人, *J. Biol. Chem.* 268:1965-1975 (1993))、包括Deep Vent(exo-) DNA聚合酶的Deep Vent DNA聚合酶、IsoPol DNA聚合酶、DNA聚合酶I、Therminator DNA聚合酶、T5 DNA聚合酶(Chatterjee等人, *Gene* 97:13-19 (1991))、测序酶(US. Biochemicals)、T7 DNA聚合酶、T7-测序酶、T7 gp5 DNA聚合酶、PRD1DNA聚合酶、T4 DNA聚合酶(Kaboord和Benkovic, *Curr. Biol.* 5:149-157 (1995))。另外的链置换核酸聚合酶也与本文所述的方法相容。给定聚合酶进行链置换复制的能力可以被确定,例如,通过在链置换复制测定中使用该聚合酶(例如,如美国专利号6,977,148中所公开)。在一些情况下,这些测定是在适合于所用酶的最佳活性温度下进行的,例如,phi29 DNA聚合酶的该温度为32°C,exo(-) Bst DNA聚合酶的该温度为46°C至64°C,或来自超高温生物的酶的该温度为约60°C至70°C。选择聚合酶的另一有用的测定法是在Kong等人, *J. Biol. Chem.* 268:1965-1975 (1993) 中所述的引物阻断测定。该测定包括在存在或不存在寡核苷酸的情况下使用M13 ssDNA模板进行的引物延伸测定,该寡核苷酸在延伸引物的上游杂交,以阻断其进程。在该测定中,能够置换阻断引物

的其他酶在一些情况下对所公开的方法有用。在一些情况下,聚合酶以近似相等的比率并入dNTP和终止子。在一些情况下,本文所述聚合酶的dNTP和终止子的并入比率为约1:1、约1.5:1、约2:1、约3:1、约4:1、约5:1、约10:1、约20:1、约50:1、约100:1、约200:1、约500:1或约1000:1。在一些情况下,本文所述聚合酶的dNTP和终止子的并入比率为1:1至1000:1、2:1至500:1、5:1至100:1、10:1至1000:1、100:1至1000:1、500:1至2000:1、50:1至1500:1或25:1至1000:1。

[0110] 本文描述了扩增方法,其中可以通过使用链置换因子例如解旋酶来促进链置换。在一些情况下,这些因子与另外的扩增组分结合使用,所述另外的扩增组分如聚合酶、终止子或其他组分。在一些情况下,链置换因子与不具有链置换活性的聚合酶一起使用。在一些情况下,链置换因子与具有链置换活性的聚合酶一起使用。不受理论的约束,链置换因子可以增加较小的双链扩增子被引发的速率。在一些情况下,可以在存在链置换因子的情况下进行链置换复制的任何DNA聚合酶都适用于PTA方法,即使该DNA聚合酶在不存在这种因子的情况下无法进行链置换复制。在一些情况下,可用于链置换复制的链置换因子包括(但不限于)BMRF1聚合酶辅助亚基(Tsurumi等人,J.Virology 67(12):7648-7653(1993))、腺病毒DNA结合蛋白(Zijderveld和van der Vliet,J.Virology 68(2):1158-1164(1994))、单纯疱疹病毒蛋白ICP8(Boehmer和Lehman,J.Virology 67(2):711-715(1993);Skaliter和Lehman,Proc.Natl.Acad.Sci.USA 91(22):10665-10669(1994));单链DNA结合蛋白(SSB;Rigler和Romano,J.Biol.Chem.270:8910-8919(1995));噬菌体T4基因32蛋白(Villemain和Giedroc,Biochemistry 35:14395-14404(1996));T7解旋酶-引发酶;T7gp2.5 SSB蛋白;Tte-UvrD(来自腾冲嗜热厌氧菌(*Thermoanaerobacter tengcongensis*)),小牛胸腺解旋酶(Siegel等人,J.Biol.Chem.267:13629-13635(1992));细菌SSB(例如,大肠杆菌SSB),真核生物中的复制蛋白A(RPA),人类线粒体SSB(mtSSB)和重组酶(例如,重组酶A(RecA)家族蛋白、T4 UvsX、T4UvsY、噬菌体HK620的Sak4、Rad51、Dmc1或Radb)。促进链置换和引发的因子组合也符合本文所述方法。例如,将解旋酶与聚合酶结合使用。在一些情况下,PTA方法包括使用单链DNA结合蛋白(SSB、T4gp32或其他单链DNA结合蛋白)、解旋酶和聚合酶(例如,SauDNA聚合酶、Bsu聚合酶、Bst2.0、GspM、GspM2.0、GspSSD或其他合适的聚合酶)。在一些情况下,将逆转录酶与本文所述的链置换因子结合使用。在一些情况下,将逆转录酶与本文所述的链置换因子结合使用。在一些情况下,使用聚合酶和切口酶(例如,“NEAR”(如US 9,617,586中所述的那些)进行扩增。在一些情况下,切口酶是Nt.BspQI、Nb.BbvCi、Nb.BsmI、Nb.BsrDI、Nb.BtsI、Nt.AlwI、Nt.BbvCI、Nt.BstNBI、Nt.CviPII、Nb.Bpu10I或Nt.Bpu10I。

[0111] 本文描述了扩增方法,其包括使用终止子核苷酸、聚合酶和其他因子或条件。例如,在一些情况下,这些因子在扩增过程中用于使一种或多种核酸模板或扩增子片段化。在一些情况下,这些因子包括核酸内切酶。在一些情况下,因子包括转座酶。在一些情况下,在扩增过程中使用机械剪切来使核酸片段化。在一些情况下,在扩增过程中添加核苷酸,可以通过添加其他蛋白质或条件使其片段化。例如,将尿嘧啶并入扩增子中;用尿嘧啶D-糖基化酶的处理使核酸在含尿嘧啶位置处片段化。在一些情况下,还采用了选择性核酸片段化的其他体系,例如,切割经修饰的胞嘧啶-苊碱基对的工程化DNA糖基化酶(Kwon,等人Chem Biol.2003,10(4),351)。

[0112] 本文描述了包括使用终止子核苷酸的扩增方法,该终止子核苷酸终止核酸复制,

从而减小扩增产物的大小。在一些情况下,这些终止子与本文所述的聚合酶、链置换因子或其他扩增组分结合使用。在一些情况下,终止子核苷酸减少或降低了核酸复制的效率。在一些情况下,这些终止子将延伸率减少至少99.9%、99%、98%、95%、90%、85%、80%、75%、70%或至少65%。在一些情况下,这些终止子将延伸率减少50%-90%、60%-80%、65%-90%、70%-85%、60%-90%、70%-99%、80%-99%或50%-80%。在一些情况下,终止子将平均扩增子产物长度减少至少99.9%、99%、98%、95%、90%、85%、80%、75%、70%或至少65%。在一些情况下,终止子将平均扩增子长度减少50%-90%、60%-80%、65%-90%、70%-85%、60%-90%、70%-99%、80%-99%或50%-80%。在一些情况下,包括终止子核苷酸的扩增子会形成环或发夹,从而降低聚合酶将这些扩增子用作模板的能力。在一些情况下,终止子的使用通过并入终止子核苷酸(例如,经过修饰以使其抵抗核酸外切酶从而终止DNA延伸的双脱氧核苷酸)而减慢起始扩增位点处的扩增速度,从而产生较小的扩增产物。通过比当前使用的方法产生更小的扩增产物(例如,PTA方法的平均长度为50-2000个核苷酸,而MDA方法的平均产物长度为>10,000个核苷酸),PTA扩增产物在一些情况下可直接进行连接衔接子而无需片段化,从而允许细胞条形码和独特分子标识码(UMI)的有效并入(参见图2A)。

[0113] 终止子核苷酸以各种浓度存在,取决于诸如聚合酶、模板或其他因子。例如,在一些情况下,在本文所述的方法中,终止子核苷酸的量表示为非终止子核苷酸与终止子核苷酸的比率。在一些情况下,这些浓度允许控制扩增子的长度。在一些情况下,针对存在的模板量或模板的大小改变终止子核苷酸与非终止子核苷酸的比率。在一些情况下,对于较小的样品大小(例如,飞克至皮克范围)降低终止子核苷酸与非终止子核苷酸的比率。在一些情况下,非终止子核苷酸与终止子核苷酸的比率为约2:1、5:1、7:1、10:1、20:1、50:1、100:1、200:1、500:1、1000:1、2000:1或5000:1。在一些情况下,非终止子核苷酸与终止子核苷酸的比率为2:1-10:1、5:1-20:1、10:1-100:1、20:1-200:1、50:1-1000:1、50:1-500:1、75:1-150:1或100:1-500:1。在一些情况下,在使用本文所述的方法扩增期间存在的至少一种核苷酸是终止子核苷酸。每种终止子不必以约相同的浓度存在;在一些情况下,对于特定的一组反应条件、样品类型或聚合酶,可以优化本文所述方法中存在的各种终止子的比例。不受理论的约束,每种终止子在响应与模板链上相应核苷酸的配对时,并入扩增子的生长多核苷酸链中的效率可以不同。例如,在一些情况下,与胞嘧啶配对的终止子的浓度比平均终止子浓度高约3%、5%、10%、15%、20%、25%或50%。在一些情况下,与胸腺嘧啶配对的终止子的浓度比平均终止子浓度高约3%、5%、10%、15%、20%、25%或50%。在一些情况下,与鸟嘌呤配对的终止子的浓度比平均终止子浓度高约3%、5%、10%、15%、20%、25%或50%。在一些情况下,与腺嘌呤配对的终止子的浓度比平均终止子浓度高约3%、5%、10%、15%、20%、25%或50%。在一些情况下,与尿嘧啶配对的终止子的浓度比平均终止子浓度高约3%、5%、10%、15%、20%、25%或50%。在一些情况下,能够终止通过核酸聚合酶的核酸延伸的任何核苷酸在本文所述的方法中用作终止子核苷酸。在一些情况下,可逆终止子用于终止核酸复制。在一些情况下,不可逆终止子用于终止核酸复制。在一些情况下,终止子的非限制性示例包括可逆和不可逆核酸和核酸类似物,例如,包括核苷酸的3' 封闭的可逆终止子、包括核苷酸的3' 未封闭的可逆终止子、包括脱氧核苷酸的2' 修饰的终止子、包括对脱氧核苷酸的含氮碱基的修饰的终止子或其任何组合。在一个实施方案中,终止子核苷

酸是双脱氧核苷酸。终止核酸复制并且可以适用于实施本发明的其他核苷酸修饰包括但不限于脱氧核糖的3' 碳的r基团的任何修饰,如反向双脱氧核苷酸、3' 生物素化核苷酸、3' 氨基核苷酸、3' -磷酸化核苷酸、3' -O-甲基核苷酸、包括3' C3间隔子核苷酸、3' C18核苷酸、3' 己二醇间隔子核苷酸的3' 碳间隔子核苷酸、无环核苷酸,及其组合。在一些情况下,终止子是长度为1、2、3、4或更多个碱基的多核苷酸。在一些情况下,终止子不包括可检测的部分或标签(例如,质量标签、荧光标签、染料、放射性原子或其他可检测的部分)。在一些情况下,终止子不包括允许可检测部分或标签附接的化学部分(例如,“点击”叠氮化物/炔烃、共轭加成剂或用于标签附接的其他化学处理)。在一些情况下,所有终止子核苷酸都包括相同的修饰,该修饰减少核苷酸的某区域(例如,糖部分、碱基部分或磷酸部分)处的扩增。在一些情况下,至少一种终止子具有减少扩增的不同修饰。在一些情况下,所有终止子都具有基本相似的荧光激发或发射波长。在一些情况下,未修饰磷酸基团的终止子与不具有核酸外切酶校对活性的聚合酶一起使用。终止子在与具有可以除去终止子核苷酸的3' ->5' 校对核酸外切酶活性的聚合酶(例如,phi29)一起使用时,在一些情况下,还需要进一步修饰使其抵抗核酸外切酶。例如,双脱氧核苷酸被 α -硫基团修饰,产生硫代磷酸酯键,使这些核苷酸对核酸聚合酶的3' ->5' 校对核酸外切酶活性具有抗性。在一些情况下,这种修饰使聚合酶的核酸外切酶校对活性降低至少99.5%、99%、98%、95%、90%或至少85%。在一些情况下,提供对3' ->5' 核酸外切酶活性的抗性的其他终止子核苷酸修饰的非限制性示例包括:具有对 α 基团的修饰的核苷酸,如产生硫代磷酸酯键的 α -硫代双脱氧核苷酸、C3间隔子核苷酸、锁核酸(LNA)、反向核酸、2' 氟碱基、3' 磷酸化、2' -O-甲基修饰(或其他2' -O-烷基修饰)、丙炔修饰的碱基(例如,脱氧胞嘧啶、脱氧尿苷)、L-DNA核苷酸、L-RNA核苷酸、具有反向连接的核苷酸(例如,5' -5' 或3' -3')、5' 反向碱基(例如,5' 反向的2', 3' -二脱氧dT)、甲基磷酸酯骨架和反式核酸。在一些情况下,具有修饰的核苷酸包括具有游离3' OH基团的碱基修饰的核酸(例如,2-硝基苄基烷基化的HOMedU三磷酸,包含具有大化学基团(如固体支持物或其他较大部分)的修饰的碱基)。在一些情况下,将具有链置换活性但不具有3' ->5' 核酸外切酶校对活性的聚合酶与经历或未经使其具有核酸外切酶抗性的修饰的终止子核苷酸一起使用。这些核酸聚合酶包括但不限于Bst DNA聚合酶、Bsu DNA聚合酶、Deep Vent (exo-) DNA聚合酶、Klenow片段(exo-) DNA聚合酶、Therminator DNA聚合酶和Vent_R (exo-)。

[0114] 引物和扩增子文库

[0115] 本文描述了由至少一种靶核酸分子的扩增产生的扩增子文库。在一些情况下,这些文库是使用本文所述方法生成的,如使用终止子的文库。这些方法包括使用链置换聚合酶或因子、终止子核苷酸(可逆或不可逆)或本文所述的其他特征和实施方案。在一些情况下,使用本文所述的终止子产生的扩增子文库在随后的扩增反应(例如,PCR)中进一步扩增。在一些情况下,随后的扩增反应不包括终止子。在一些情况下,扩增子文库包括多核苷酸,其中至少50%、60%、70%、80%、90%、95%或至少98%的多核苷酸包括至少一种终止子核苷酸。在一些情况下,扩增子文库包括衍生出扩增子文库的靶核酸分子。扩增子文库包括多种多核苷酸,其中至少一些多核苷酸是直接拷贝(例如,直接从靶核酸分子如基因组DNA、RNA或其他靶核酸复制)。例如,至少5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或超过95%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,至少5%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,至

少10%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,至少15%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,至少20%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,至少50%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,3%-5%、3-10%、5%-10%、10%-20%、20%-30%、30%-40%、5%-30%、10%-50%或15%-75%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝。在一些情况下,至少一些多核苷酸是靶核酸分子的直接拷贝或子代(靶核酸的第一拷贝)。例如,至少5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或超过95%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝或子代。在一些情况下,至少5%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝或子代。在一些情况下,至少10%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝或子代。在一些情况下,至少20%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝或子代。在一些情况下,至少30%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝或子代。在一些情况下,3%-5%、3%-10%、5%-10%、10%-20%、20%-30%、30%-40%、5%-30%、10%-50%或15%-75%的扩增子多核苷酸是至少一种靶核酸分子的直接拷贝或子代。在一些情况下,靶核酸的直接拷贝的长度为50-2500、75-2000、50-2000、25-1000、50-1000、500-2000或50-2000个碱基。在一些情况下,子代的长度为1000-5000、2000-5000、1000-10,000、2000-5000、1500-5000、3000-7000或2000-7000个碱基。在一些情况下,PTA扩增产物的平均长度为25-3000个核苷酸,为50-2500、75-2000、50-2000、25-1000、50-1000、500-2000或50-2000个碱基。在一些情况下,从PTA产生的扩增子的长度不超过5000、4000、3000、2000、1700、1500、1200、1000、700、500或不超过300个碱基。在一些情况下,从PTA产生的扩增子的长度为1000-5000、1000-3000、200-2000、200-4000、500-2000、750-2500或1000-2000个碱基。在一些情况下,使用本文所述方法产生的扩增子文库包括至少1000、2000、5000、10,000、100,000、200,000、500,000或超过500,000个包括独特序列的扩增子。在一些情况下,文库包括至少100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、2000、2500、3000或至少3500个扩增子。在一些情况下,长度小于1000个碱基的扩增子多核苷酸的至少5%、10%、15%、20%、25%、30%或超过30%是至少一种靶核酸分子的直接拷贝。在一些情况下,长度不超过2000个碱基的扩增子多核苷酸的至少5%、10%、15%、20%、25%、30%或超过30%是至少一种靶核酸分子的直接拷贝。在一些情况下,长度为3000-5000个碱基的扩增子多核苷酸的至少5%、10%、15%、20%、25%、30%或超过30%是至少一种靶核酸分子的直接拷贝。在一些情况下,直接拷贝扩增子与靶核酸分子的比例为至少10:1、100:1、1000:1、10,000:1、100,000:1、1,000,000:1、10,000,000:1或大于10,000,000:1。在一些情况下,直接拷贝扩增子与靶核酸分子的比例为至少10:1、100:1、1000:1、10,000:1、100,000:1、1,000,000:1、10,000,000:1或大于10,000,000:1,其中直接拷贝扩增子的长度不超过700-1200个碱基。在一些情况下,直接拷贝扩增子和子扩增子与靶核酸分子的比例为至少10:1、100:1、1000:1、10,000:1、100,000:1、1,000,000:1、10,000,000:1或大于10,000,000:1。在一些情况下,直接拷贝扩增子和子扩增子与靶核酸分子的比例为至少10:1、100:1、1000:1、10,000:1、100,000:1、1,000,000:1、10,000,000:1或大于10,000,000:1,其中直接拷贝扩增子的长度为700-1200个碱基,并且子扩增子的长度为2500-6000个碱基。在一些情况下,文库包括约50-10,000、约50-5,000、约50-2500、约50-1000、约150-2000、约250-

3000、约50-2000、约500-2000或约500-1500个扩增子,这些是靶核酸分子的直接拷贝。在一些情况下,文库包括约50-10,000、约50-5,000、约50-2500、约50-1000、约150-2000、约250-3000、约50-2000、约500-2000或约500-1500个扩增子,这些是靶核酸分子的直接拷贝或子扩增子。在一些情况下,直接拷贝数可以通过PCR扩增循环数控制。在一些情况下,使用不超过30、25、20、15、13、11、10、9、8、7、6、5、4或3个PCR循环来产生靶核酸分子的拷贝。在一些情况下,使用约30、25、20、15、13、11、10、9、8、7、6、5、4或约3个PCR循环来产生靶核酸分子的拷贝。在一些情况下,使用3、4、5、6、7或8个PCR循环来产生靶核酸分子的拷贝。在一些情况下,使用2-4、2-5、2-7、2-8、2-10、2-15、3-5、3-10、3-15、4-10、4-15、5-10或5-15个PCR循环来产生靶核酸分子的拷贝。在一些情况下,使用本文所述方法生成的扩增子文库经受另外步骤,如衔接子连接和进一步的PCR扩增。在一些情况下,这些另外步骤在测序步骤之前。

[0116] 本文所述的方法可以另外包括一个或多个富集或纯化步骤。在一些情况下,在本文所述的方法期间富集一种或多种多核苷酸(如cDNA、PTA扩增子或其他多核苷酸)。在一些情况下,使用多核苷酸探针捕获一种或多种多核苷酸。在一些情况下,探针被配置为捕获一个或多个基因组外显子。在一些情况下,探针文库包括至少1000、2000、5000、10,000、50,000、100,000、200,000、500,000或超过1百万个不同的序列。在一些情况下,探针文库包括能够与至少10、20、50、100、200、500、1000、2000、5000、10,000或超过10,000个基因结合的序列。在一些情况下,探针包含用于固体支持物的捕获的部分,如生物素。在一些情况下,富集步骤在PTA步骤之后进行。在一些情况下,富集步骤在PTA步骤之前进行。在一些情况下,探针被配置为结合基因组DNA文库。在一些情况下,探针被配置为结合cDNA文库。

[0117] 在一些情况下,由本文所述的PTA方法和组合物(终止子、聚合酶等)产生的多核苷酸的扩增子文库具有增加的均匀性。在一些情况下,均匀性使用洛伦兹曲线(例如,图5C)或其他类似方法来描述。在一些情况下,这种增加使得覆盖所期望的靶核酸分子(例如,基因组DNA、RNA或其他靶核酸分子)所需的测序读段更少。例如,多核苷酸的累积分数的不超过50%包括靶核酸分子的序列的累积分数的至少80%的序列。在一些情况下,多核苷酸的累积分数的不超过50%包括靶核酸分子的序列的累积分数的至少60%的序列。在一些情况下,多核苷酸的累积分数的不超过50%包括靶核酸分子的序列的累积分数的至少70%的序列。在一些情况下,多核苷酸的累积分数的不超过50%包括靶核酸分子的序列的累积分数的至少90%的序列。在一些情况下,均匀性使用基尼指数描述(其中指数0表示文库的完全相等,指数1表示完全不等)。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.55、0.50、0.45、0.40或0.30。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.50。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.40。在一些情况下,这种均匀性度量取决于所获得的读段数。例如,获得的读段不超过1亿、2亿、3亿、4亿或不超过5亿。在一些情况下,读取的长度为约50、75、100、125、150、175、200、225或约250个碱基。在一些情况下,均匀性度量取决于靶核酸的覆盖深度。例如,平均覆盖深度为约10X、15X、20X、25X或约30X。在一些情况下,平均覆盖深度为10-30X、20-50X、5-40X、20-60X、5-20X或10-20X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.55,其中获得了约3亿个读段。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.50,其中获得了约3亿个读段。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.45,其中获得了约3亿个读段。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.55,其中获得了不超过3亿

个读段。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.50,其中获得了不超过3亿个读段。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.45,其中获得了不超过3亿个读段。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.55,其中测序覆盖的平均深度为约15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.50,其中测序覆盖的平均深度为约15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.45,其中测序覆盖的平均深度为约15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.55,其中测序覆盖的平均深度为至少15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.50,其中测序覆盖的平均深度为至少15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.45,其中测序覆盖的平均深度为至少15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.55,其中测序覆盖的平均深度不超过15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.50,其中测序覆盖的平均深度不超过15X。在一些情况下,本文所述的扩增子文库的基尼指数不超过0.45,其中测序覆盖的平均深度不超过15X。在一些情况下,使用本文所述方法生成的均匀扩增子文库需要经受另外步骤,如衔接子连接和进一步的PCR扩增。在一些情况下,这些另外步骤在测序步骤之前。

[0118] 引物包括用于引发本文所述的扩增反应的核酸。在一些情况下,这些引物包括但不限于具有或不具有使其抵抗核酸外切酶的修饰的任何长度的随机脱氧核苷酸,具有或不具有使其抵抗核酸外切酶的修饰的任意长度的随机核糖核苷酸,经修饰的核酸,如锁核酸、靶向特定基因组区域的DNA或RNA引物,以及由诸如引发酶的酶引发的反应。在全基因组PTA的情况下,优选使用具有随机或部分随机核苷酸序列的一组引物。在非常复杂的核酸样品中,不需要知道样品中存在的具体核酸序列,并且不需要将引物设计为与任何特定序列互补。相反,核酸样品的复杂性导致样品中大量不同的杂交靶序列,它们将与随机或部分随机序列的各种引物互补。在一些情况下,用于PTA的引物的互补部分是完全随机的,仅包括随机的部分,或是选择性地随机的。在一些情况下,例如,引物的互补部分中的随机碱基位置数为引物的互补部分中的核苷酸总数的20%至100%。在一些情况下,引物的互补部分中的随机碱基位置数为引物的互补部分中的核苷酸总数的10%至90%、15-95%、20%-100%、30%-100%、50%-100%、75-100%或90-95%。在一些情况下,引物的互补部分中的随机碱基位置数为引物的互补部分中的核苷酸总数的至少10%、20%、30%、40%、50%、60%、70%、80%或至少90%。在一些情况下,使用标准技术合成具有随机或部分随机序列的引物组,这是通过允许在每个位置随机添加任何核苷酸。在一些情况下,引物组由具有相似的长度和/或杂交特性的引物组成。在一些情况下,术语“随机引物”是指在每个位置均可表现出四倍简并性的引物。在一些情况下,术语“随机引物”是指在每个位置均可表现出三倍简并性的引物。在一些情况下,本文所述方法中使用的随机引物包括长度为3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20或更多个碱基的随机序列。在一些情况下,引物包括长度为3-20、5-15、5-20、6-12或4-10个碱基的随机序列。引物也可以包括不可延伸的元件,其限制产生的扩增子的后续扩增。例如,在一些情况下,具有不可延伸元件的引物包括终止子。在一些情况下,引物包括终止子核苷酸,如1、2、3、4、5、10或超过10种终止子核苷酸。引物不需要局限于从外部添加至扩增反应的组分。在一些情况下,通过添加促进引发的核苷酸和蛋白质来原位产生引物。例如,在一些情况下,将与核苷酸组合的引发酶样酶用于产生本文

所述方法的随机引物。在一些情况下,引发酶样酶是DnaG或AEP酶超家族的成员。在一些情况下,引发酶样酶是TthPrimPol。在一些情况下,引发酶样酶是T7 gp4解旋酶-引发酶。在一些情况下,这些引发酶与本文所述的聚合酶或链置换因子一起使用。在一些情况下,引发酶用脱氧核糖核苷酸启动引发。在一些情况下,引发酶用核糖核苷酸启动引发。

[0119] 在PTA扩增后可以选择扩增子的特定子集。在一些情况下,这种选择取决于大小、亲和力、活性、与探针的杂交或本领域中其他已知的选择因子。在一些情况下,选择在本文所述的另外步骤如衔接子连接和/或文库扩增之前或之后进行。在一些情况下,选择基于扩增子的大小(长度)。在一些情况下,选择较小的扩增子,其不太可能经历指数扩增,从而富集了从原代模板衍生的产物,同时进一步将扩增从指数形式转化为拟线性扩增过程(图1A)。在一些情况下,选择长度为50-2000、25-5000、40-3000、50-1000、200-1000、300-1000、400-1000、400-600、600-2000或800-1000个碱基的扩增子。在一些情况下,通过使用方案来进行大小选择,例如,在羧化顺磁珠上使用固相可逆固定化(SPRI)以富集特定大小的核酸片段,或本领域技术人员已知的其他方案。任选地或组合地,选择通过在制备测序文库的同时在PCR期间优先连接和扩增较小片段来进行,也作为在测序(例如,合成测序、纳米孔测序或其他测序方法)期间优先由较小测序文库片段形成簇的结果来进行。选择较小片段的其他策略也符合本文所述的方法并且包括但不限于在凝胶电泳后分离特定大小的核酸片段,使用结合特定大小的核酸片段的硅胶柱,以及使用可以更强地富集较小片段的其他PCR策略。可以将任何数量的文库制备方案与本文所述的PTA方法一起使用。在一些情况下,将通过PTA产生的扩增子连接至衔接子(任选地在除去终止子核苷酸的情况下)。在一些情况下,通过PTA产生的扩增子包含由基于转座酶的片段化产生的同源区域,其用作引发位点。在一些情况下,通过以机械或酶促方式片段化核酸来制备文库。在一些情况下,经由转座体使用标签化来制备文库。在一些情况下,经由衔接子,如Y-衔接子、通用衔接子或环状衔接子的连接来制备文库。

[0120] PTA中使用的引物的非互补部分可以包括可用于进一步操纵和/或分析扩增序列的序列。这种序列的一个示例是“检测标签”。检测标签具有与检测探针互补的序列,并使用其同源检测探针进行检测。引物上可以有一个、两个、三个、四个或四个以上的检测标签。除引物的大小外,对引物上可能存在的检测标签的数目没有基本限制。在一些情况下,引物上只有一个检测标签。在一些情况下,引物上有两个检测标签。当有多个检测标签时,它们可以具有相同的序列,也可以具有不同的序列,每个不同的序列与不同的检测探针互补。在一些情况下,多个检测标签具有相同的序列。在一些情况下,多个检测标签具有不同的序列。

[0121] 可以包括在引物的非互补部分中的序列的另一个示例是“地址标签”,所述地址标签可以编码扩增子的其他细节,如组织切片中的位置。在一些情况下,细胞条形码包括地址标签。地址标签具有与地址探针互补的序列。地址标签被并入扩增链的末端。如果存在,引物上可以有一个或多个地址标签。除引物的大小外,引物上可能存在的地址标签的数目没有基本限制。当有多个地址标签时,它们可以具有相同的序列,也可以具有不同的序列,每个不同的序列都与不同的地址探针互补。地址标签部分可以是支持地址标签与地址探针之间特异性且稳定的杂交的任何长度。在一些情况下,来自一个以上来源的核酸可以并入可变标签序列。该标签序列的长度可以高达100个核苷酸,优选地长度为1至10个核苷酸,最优选地为4、5或6个核苷酸,并且包括核苷酸的组合。在一些情况下,标签序列的长度为1-20、

2-15、3-13、4-12、5-12或1-10个核苷酸。例如,如果选择六个碱基对形成标签并且使用四个不同核苷酸的排列,则可以制成总共4096个核酸锚(例如,发夹),每个锚具有独特的6碱基标签。

[0122] 本文所述的引物可以存在于溶液中或固定在固体支持物上。在一些情况下,带有样品条形码和/或UMI序列的引物可以固定在固体支持物上。例如,固体支持物可以是一个或多个珠。在一些情况下,使个体细胞与一个或多个具有独特的一组样品条形码和/或UMI序列的珠接触,以鉴定个体细胞。在一些情况下,将来自个体细胞的裂解物与一个或多个具有独特的一组样品条形码和/或UMI序列的珠接触,以鉴定个体细胞裂解物。在一些情况下,将来自个体细胞的提取的核酸与一个或多个具有独特的一组样品条形码和/或UMI序列的珠接触,以鉴定来自个体细胞的提取的核酸。珠可以以本领域已知的任何合适方式来操纵,例如,使用本文所述的液滴致动器。珠可以是任何合适的大小,包括例如,微珠、微粒、纳米珠和纳米颗粒。在一些实施方案中,珠是磁响应的;在其他实施方案中,珠没有明显的磁响应。合适的珠的非限制性示例包括流式细胞术微珠、聚苯乙烯微粒和纳米颗粒、官能化的聚苯乙烯微粒和纳米颗粒、包覆的聚苯乙烯微粒和纳米颗粒、二氧化硅微珠、荧光微球和纳米球、官能化的荧光微球和纳米球、包覆的荧光微球和纳米球、颜色染色微粒和纳米颗粒、磁性微粒和纳米颗粒、超顺磁性微粒和纳米颗粒(例如,可从Invitrogen Group, Carlsbad, CA获得的DYNABEADS®)、荧光微粒和纳米颗粒、包覆的磁性微粒和纳米颗粒、铁磁性微粒和纳米颗粒、包覆的铁磁性微粒和纳米颗粒,以及在美国专利申请公开号US20050260686、US20030132538、US20050118574、20050277197、20060159962中所描述的。珠可以与抗体、蛋白质或抗原、DNA/RNA探针或任何其他对所需靶标具有亲和力的分子预偶联。在一些实施方案中,带有样品条形码和/或UMI序列的引物可以在溶液中。在某些实施方案中,可以提供多个液滴,其中多个液滴中的每个液滴都具有对于液滴来说独特的样品条形码和对于分子来说独特的UMI,从而使得UMI在液滴集合内重复多次。在一些实施方案中,使个体细胞与具有独特的一组样品条形码和/或UMI序列的液滴接触,以鉴定个体细胞。在一些实施方案中,使来自个体细胞的裂解物与具有独特的一组样品条形码和/或UMI序列的液滴接触,以鉴定个体细胞裂解物。在一些实施方案中,将来自个体细胞的提取的核酸与具有独特的一组样品条形码和/或UMI序列的液滴接触,以鉴定来自个体细胞的提取的核酸。

[0123] PTA引物可以包括序列特异性或随机的引物、细胞条形码和/或独特分子标识码(UMI)(参见,例如,图10A(线性引物)和图10B(发夹引物))。在一些情况下,引物包括序列特异性引物。在一些情况下,引物包括随机引物。在一些情况下,引物包括细胞条形码。在一些情况下,引物包括样品条形码。在一些情况下,引物包括独特分子标识码。在一些情况下,引物包括两个或更多个细胞条形码。在一些情况下,这些条形码鉴定独特的样品来源或独特的工作流程。在一些情况下,这些条形码或UMI的长度为5、6、7、8、9、10、11、12、15、20、25、30或超过30个碱基。在一些情况下,引物包括至少1000、10,000、50,000、100,000、250,000、500,000、 10^6 、 10^7 、 10^8 、 10^9 或至少 10^{10} 个独特条形码或UMI。在一些情况下,引物包括至少8、16、96或384个独特条形码或UMI。在一些情况下,然后在测序前将标准衔接子连接至扩增产物上;测序后,首先根据细胞条形码将读段分配给特定细胞。可以与PTA方法一起使用的合适衔接子包括,例如,可从Integrated DNA Technologies (IDT)获得的xGen® Dual Index UMI衔接子。然后,使用UMI将来自每个细胞的读段分组,并将具有相同UMI的读段合

并为共有读段。使用细胞条形码允许在制备文库之前合并所有细胞,因为它们之后可以通过细胞条形码鉴定。在一些情况下,使用UMI形成共有读段校正PCR偏倚,从而改善拷贝数变异(CNV)检测(图11A和图11B)。此外,可以通过要求来自同一分子的固定百分比的读段在每个位置具有相同的检测到的碱基变化来校正测序错误。这种方法已被用于改善CNV检测并校正大量样品中的测序错误。在一些情况下,UMI与本文所述的方法一起使用,例如,美国专利号8,835,358公开了在附接随机可扩增条形码后的数字计数原理。Schmitt.等人和Fan等人公开了校正测序错误的类似方法。在一些情况下,产生文库以使用引物进行测序。在一些情况下,文库包括长度为200-700个碱基、100-1000、300-800、300-550、300-700或200-800个碱基的片段。在一些情况下,文库包括长度为至少50、100、150、200、300、500、600、700、800或至少1000个碱基的片段。在一些情况下,文库包括长度为约50、100、150、200、300、500、600、700、800或约1000个碱基的片段。

[0124] 本文所述的方法还可以包括另外的步骤,包括对样品或模板进行的步骤。在一些情况下,这些样品或模板在PTA之前要经过一个或多个步骤。在一些情况下,对包括细胞的样品进行预处理步骤。例如,使用冻融、Triton X-100、Tween 20和蛋白酶K的组合对细胞进行裂解和蛋白水解,以增加染色质的可及性。其他裂解策略也适用于实施本文所述的方法。这些策略包括但不限于使用洗涤剂 and/或溶菌酶和/或蛋白酶处理,以及/或细胞物理破坏如超声和/或碱裂解和/或低渗裂解的其他组合进行裂解。在一些情况下,对原代模板或一种或多种靶分子进行预处理步骤。在一些情况下,使用氢氧化钠使原代模板(或靶标)变性,然后中和溶液。其他变性策略也可适用于实施本文所述的方法。这些策略可以包括但不限于将碱裂解与其他碱性溶液组合,提高样品温度和/或改变样品中的盐浓度,添加添加剂如溶剂或油,其他修饰或其任何组合。在一些情况下,另外的步骤包括按大小对样品、模板或扩增子进行分类、过滤或分离。在一些情况下,将细胞用机械(例如,高压均化器、珠研磨)或非机械(物理、化学或生物)裂解。在一些情况下,物理裂解方法包括加热、渗透压休克和/或空化。在一些情况下,化学裂解包括碱和/或洗涤剂。在一些情况下,生物裂解包括使用酶。裂解方法的组合也与本文所述的方法相容。裂解酶的非限制示例包括重组溶菌酶、丝氨酸蛋白酶和细菌溶素。在一些情况下,用酶的裂解包括使用溶菌酶、溶葡萄球菌酶、消解酶、纤维素、蛋白酶或聚糖酶。例如,在用本文所述的方法扩增后,扩增子文库富集具有期望长度的扩增子。在一些情况下,扩增子文库富含长度为50-2000、25-1000、50-1000、75-2000、100-3000、150-500、75-250、170-500、100-500或75-2000个碱基的扩增子。在一些情况下,扩增子文库富含长度不超过75、100、150、200、500、750、1000、2000、5000或不超过10,000个碱基的扩增子。在一些情况下,扩增子文库富含长度为至少25、50、75、100、150、200、500、750、1000或至少2000个碱基的扩增子。

[0125] 本文所述的方法和组合物可包括缓冲剂或其他制剂。在一些情况下,此类缓冲剂用于PTA、RT或本文所述的其他方法。在一些情况下,这些缓冲剂包括表面活性剂/洗涤剂或变性剂(Tween-20、DMSO、DMF,包括疏水基团的聚乙二醇化聚合物或其他表面活性剂)、盐(磷酸钾或磷酸钠(一元或二元)、氯化钠、氯化钾)、TrisHCl、氯化镁或硫酸镁、铵盐,如磷酸盐、硝酸盐或硫酸盐、EDTA)、还原剂(DTT、THP、DTE、 β -巯基乙醇、TCEP或其他还原剂)或其他组分(甘油、亲水性聚合物,如PEG)。在一些情况下,将缓冲剂与诸如聚合酶、链置换因子、终止子或本文所述的其他反应组分等组分结合使用。在一些情况下,将缓冲剂与诸如聚合酶、

链置换因子、终止子或本文所述的其他反应组分等组分结合使用。缓冲剂可以包含一种或多种拥挤剂。在一些情况下,拥挤试剂包括聚合物。在一些情况下,拥挤试剂包括聚合物如多元醇。在一些情况下,拥挤试剂包括聚乙二醇聚合物(PEG)。在一些情况下,拥挤试剂包括多糖。非限制地,拥挤试剂的示例包括ficoll(例如,ficoll PM 400、ficoll PM 70或其他分子量ficoll)、PEG(例如,PEG1000、PEG 2000、PEG4000、PEG6000、PEG8000或其他分子量PEG)、右旋糖酐(右旋糖酐6、右旋糖酐10、右旋糖酐40、右旋糖酐70、右旋糖酐6000、右旋糖酐138k或其他分子量右旋糖酐)。

[0126] 根据本文所述方法扩增的核酸分子可以使用本领域技术人员已知的方法进行测序和分析。在一些情况下,使用的测序方法的非限制性示例包括,例如,杂交测序(SBH)、连接测序(SBL)(Shendure等人(2005) *Science* 309:1728)、定量增量荧光核苷酸添加测序(QIFNAS)、逐步连接和切割、荧光共振能量转移(FRET)、分子信标、TaqMan报告基因探针消化、焦磷酸测序、荧光原位测序(FISSEQ)、FISSEQ珠(美国专利号7,425,431)、摆动测序(国际专利申请公开号W02006/073504)、多重测序(美国专利申请公开号US2008/0269068; Porreca等人,2007, *Nat. Methods* 4:931)、聚合酶克隆(POLONY)测序(美国专利号6,432,360、6,485,944和6,511,803,以及国际专利申请公开号W02005/082098)、纳米网格滚环测序(ROLONY)(美国专利号9,624,538)、等位基因特异性寡核苷酸连接测定(例如,寡核苷酸连接测定(OLA),使用连接的线性探针和滚环扩增(RCA)读出的单模板分子OLA,连接的挂锁探针和/或使用连接的圆形挂锁探针和滚环扩增(RCA)读出的单模板分子OLA)、高通量测序方法,例如,使用Roche 454、Illumina Solexa、AB-SOLiD、Helicos、Polonator平台等的方法,以及基于光的测序技术(Landegren等人(1998) *Genome Res.* 8:769-76; Kwok(2000) *Pharmacogenomics* 1:95-100;以及Shi(2001) *Clin. Chem.* 47:164-172)。在一些情况下,将扩增的核酸分子进行鸟枪法测序。在一些情况下,测序文库的测序用任何适当的测序技术进行,所述测序技术包括但不限于单分子实时(SMRT)测序、Polony测序、连接测序、可逆终止子测序、质子检测测序、离子半导体测序、纳米孔测序、电子测序、焦磷酸测序、Maxam-Gilbert测序、链终止(例如,Sanger)测序、+S序列或合成测序(基于阵列/集落基或基于纳米球)。

[0127] 可以对使用本文所述的方法(例如,PTA或RNAseq)产生的测序文库进行测序以获得所需数量的测序读段。在一些情况下,从单细胞或包含单细胞的样品产生文库(单独或多组学工作流的一部分)。在一些情况下,对文库进行测序以获得至少0.1、0.2、0.4、0.5、0.7、0.8、0.9、1、1.1、1.2、1.5、2.5或至少10百万个读段。在一些情况下,对文库进行测序以获得不超过0.1、0.2、0.4、0.5、0.7、0.8、0.9、1、1.1、1.2、1.5、2.5或不超过10百万个读段。在一些情况下,对文库进行测序以获得约0.1、0.2、0.4、0.5、0.7、0.8、0.9、1、1.1、1.2、1.5、2.5或约10百万个读段。在一些情况下,对文库进行测序以每样品获得0.1-10、0.1-5、0.1-1、0.2-1、0.3-1.5、0.5-1、1-5或0.5-5百万个读段。在一些情况下,读段数目取决于基因组的大小。在一些情况下,对包含细菌基因组的样品进行测序以获得0.5-1百万个读段。在一些情况下,对文库进行测序以获得至少2、4、10、20、50、100、200、300、500、700或至少900百万个读段。在一些情况下,对文库进行测序以获得不超过2、4、10、20、50、100、200、300、500、700或不超过900百万个读段。在一些情况下,对文库进行测序以获得约2、4、10、20、50、100、200、300、500、700或约900百万个读段。在一些情况下,对包含哺乳动物基因组的样品进行

测序以获得500-600百万个读段。在一些情况下,在测序期间鉴定测序文库的类型(cDNA文库或基因组文库)。在一些情况下,在测序期间用独特的条形码鉴定cDNA文库和基因组文库。

[0128] 当关于聚合酶介导的扩增反应使用时,术语“循环”在本文中用于描述以下步骤:双链核酸(例如,来自扩增子的模板、或双链模板)的至少一部分的解离(变性)、引物的至少一部分与模板的杂交(退火)和引物的延伸以产生扩增子。在一些情况下,在扩增循环期间温度保持恒定(例如,等温反应)。在一些情况下,循环数与产生的扩增子数直接相关。在一些情况下,等温反应的循环数通过使反应进行的时间量控制。

[0129] 方法和应用

[0130] 本文描述了用多组学分析PTA方法如单细胞鉴定细胞中的突变的方法。在一些情况下使用PTA方法会导致对已知方法(例如,MDA)的改进。在一些情况下,与MDA方法相比,PTA的假阳性和假阴性变体判定率较低。在一些情况下,将基因组,如NA12878铂基因组,用于确定PTA的更大的基因组覆盖和均匀性是否会导致较低的假阴性变体判定率。不受理论的约束,可以确定PTA中错误传播的缺失降低了假阳性变体判定率。在一些情况下,通过比较已知阳性基因座处杂合突变判定的等位基因频率来估算用这两种方法在等位基因之间的扩增平衡。在一些情况下,通过PCR进一步扩增使用PTA产生的扩增子文库。在一些情况下,将PTA与另外的分析方法,如RNAseq、甲基化组分析或本文所述的其他方法一起用于工作流程中。

[0131] 在一些情况下,使用本文所述的方法分析的细胞包括肿瘤细胞。例如,循环肿瘤细胞可以从取自患者的体液中分离,如但不限于血液、骨髓、尿液、唾液、脑髓液、胸膜液、心包液、腹水或房水。然后,使细胞经历本文所述的方法(例如,PTA)并测序,以确定每个细胞的突变负荷和突变组合。在一些情况下,这些数据用于诊断特定疾病或用作预测治疗应答的工具。类似地,在一些情况下,恶性潜能未知的细胞是从取自患者的体液中分离的,所述体液如但不限于血液、骨髓、尿液、唾液、脑髓液、胸膜液、心包液、腹水、房水、囊胚腔液或培养中细胞周围的收集培养基。在一些情况下,样品从胚细胞周围的收集培养基获得。在利用本文所述的方法和测序后,这些方法还用于确定每个细胞的突变负荷和突变组合。在一些情况下,这些数据用于诊断特定疾病或用作预测恶性前状态发展为显性恶性肿瘤的工具。在一些情况下,可以从原发性肿瘤样品中分离细胞。然后,细胞可以进行PTA和测序,以确定每个细胞的突变负荷和突变组合。这些数据可以用于诊断特定疾病或用作预测患者恶性肿瘤对可用抗癌药物的抗药性的工具。通过将样品暴露于不同的化疗药物,已发现主要和次要克隆对特定药物具有不同的敏感性,这些敏感性不一定与已知的“驱动突变”的存在相关,这表明克隆种群中的突变组合决定了它对特定化疗药物的敏感性。不受理论的约束,这些发现表明,如果检测到尚未扩展并且可演化为基因组修饰数目增加的克隆而使其更有可能对治疗产生抗性的癌前病变,则可能更容易根除该恶性肿瘤。参见, Ma等人, 2018, “Pan-cancer genome and transcriptome analyses of 1,699 pediatric leukemias and solid tumors”。在一些情况下,单细胞基因组学方案用于检测从患者样品中分离的正常和恶性细胞混合物内的单个癌细胞或克隆型中的体细胞遗传变体组合。在一些情况下,该技术还用于鉴定在体外和/或患者体内暴露于药物后经历阳性选择的克隆型。如图6A所示,通过比较暴露于化疗的存活克隆与诊断时鉴定的克隆,可以创建癌症克隆型目录,该目录记

录它们对特定药物的抗性。在一些情况下,PTA方法检测由多种克隆型组成的样品中的特定克隆对现有药物或新药及其组合的敏感性,其中该方法可以检测特定克隆对药物的敏感性。在一些情况下,这种方法显示了药物对特定克隆的功效,而当前的药物敏感性测量在一次测量中考虑了所有癌症克隆的敏感性,因此可能无法检测到这种功效。当将本文所述的PTA应用于诊断时收集的患者样品以检测给定患者癌症中的癌症克隆型时,可以随后使用药物敏感性目录来查找这些克隆,从而告知肿瘤学家哪种药物或药物组合无效,哪种药物或药物组合最有可能对患者的癌症有效。PTA可用于分析包含成群细胞的样品。在一些情况下,样品包含神经元或神经胶质细胞。在一些情况下,样品包含核。

[0132] 本文描述了测量与环境因子致突变性组合时的基因表达改变的方法。例如,细胞(单细胞或细胞群体)暴露于潜在的环境条件。例如,在一些情况下,该方法使用诸如源自器官(肝脏、胰腺、肺、结肠、甲状腺或其他器官)、组织(皮肤或其他组织)、血液或其他生物来源的细胞。在一些情况下,环境条件包括热、光(例如,紫外线)、辐射、化学物质或其任何组合。在一定量的环境条件暴露后,在一些情况下,这是几分钟、几小时、几天或更长时间,分离出单细胞并进行PTA方法。在一些情况下,分子条形码和独特分子标识码用于标记样品。对样品进行测序,然后进行分析,以鉴定基因表达改变和或暴露于环境条件所导致的突变。在一些情况下,将这种突变与对照环境条件进行比较,如已知的非诱变物质、媒介物/溶剂或缺乏环境条件。在一些情况下,这种分析不仅提供了由环境条件引起的突变总数,而且还提供了这种突变的位置和性质。在一些情况下,模式从数据中鉴定,并且可以用于诊断疾病或病症。在一些情况下,模式用于预测未来的疾病状态或病症。在一些情况下,本文所述的方法测量在暴露于环境药剂,例如,潜在的诱变剂或致畸剂后细胞的突变负荷、位置和模式。在一些情况下,该方法用于评估给定药剂的安全性,包括其诱发可能导致疾病发展的突变的可能性。例如,该方法可用于预测暴露于特定浓度的特定药剂后该药剂对特定细胞类型的致癌性或致畸性。

[0133] 本文描述了鉴定与已经历基因组编辑(例如,使用CRISPR技术)的动物、植物或微生物细胞中的突变组合时的基因表达改变的方法。在一些情况下,可以将这些细胞分离并进行PTA和测序,以确定每个细胞的突变负荷和突变组合。在一些情况下,将由基因组编辑方案产生的每细胞突变率和突变位置用于评估给定基因组编辑方法的安全性。

[0134] 本文描述了确定与用于细胞疗法的细胞中的突变组合时的基因表达改变的方法,该细胞疗法诸如但不限于诱导多能干细胞的移植,尚未被操纵的造血细胞或其他细胞的移植,或经过基因组编辑的造血细胞或其他细胞的移植。然后,细胞可以进行PTA和测序,以确定每个细胞中的突变负荷和突变组合。细胞疗法产品中的每细胞突变率和突变位置可用于评估产品的安全性和潜在功效。

[0135] 用于与PTA方法一起使用的细胞可以是胎儿细胞,如胚细胞。在一些实施方案中,将PTA与非侵入性植入前选遗传学测试(NIPGT)结合使用。在另一个实施方案中,细胞可以从体外受精产生的卵裂球分离。然后,细胞可以经历PTA和测序,以确定每个细胞中潜在疾病易感遗传变体的负荷和组合。然后,可以将细胞的基因表达改变与突变谱的组合用于在植入前推断卵裂球对特定疾病的遗传易感性。在一些情况下,培养中的胚胎脱落核酸,将所述核酸用于使用低通量基因组测序评估胚胎的健康状态。在一些情况下,将胚胎冷冻解冻。在一些情况下,核酸从胚细胞培养条件培养基(BCCM)、囊胚腔液(BF)或其组合获得。在一些

情况下,使用胎儿细胞的PTA分析来检测染色体畸变,如胎儿非整倍体。在一些情况下,使用PTA来检测诸如唐氏(Down's)或帕陶(Patau)综合症的疾病。在一些情况下,在获得用于分析的核酸(例如,培养基、BF或细胞活检)之前,将冷冻胚细胞解冻并培养一段时间。在一些情况下,在获得用于分析的核酸之前,将胚细胞培养不超过4、6、8、12、16、24、36、48小时或不超过64小时。

[0136] 在另一个实施方案中,微生物细胞(例如,细菌、真菌、原生动物)可以从植物或动物(例如,微生物群样品[例如,GI微生物群、皮肤微生物群等]或体液,例如,血液、骨髓、尿液、唾液、脑脊液、胸膜液、心包液、腹水或房水)分离。此外,微生物细胞可以从留置的医疗装置分离,如但不限于,静脉导管、导尿管、脑脊髓分流器、假体瓣膜、人工关节或气管导管。然后,细胞可以经历PTA和测序,以确定特定微生物的身份,并检测预测对特定抗菌剂的应答(或抗性)的微生物遗传变体的存在。这些数据可用于诊断特定的传染病和/或用作预测治疗应答的工具。

[0137] 本文描述了使用本文所述的PTA方法从包括短核酸的样品中产生扩增子文库的方法。在一些情况下,PTA可以提高短核酸扩增的保真度和均匀性。在一些情况下,核酸的长度不超过2000个碱基。在一些情况下,核酸的长度不超过1000个碱基。在一些情况下,核酸的长度不超过500个碱基。在一些情况下,核酸的长度不超过200、400、750、1000、2000或5000个碱基。在一些情况下,包括短核酸片段的样品包括但不限于古DNA(年龄为数百年、数千年、数百万甚至数十亿年)、FFPE(福尔马林固定的石蜡包埋的)样品、无细胞DNA或其他包括短核酸的样品。

[0138] 实施方案

[0139] 本文描述了扩增靶核酸分子的方法,该方法包括:a)使包括靶核酸分子的样品,一种或多种扩增引物,核酸聚合酶以及包含一种或多种终止子核苷酸的核苷酸混合物接触,所述终止子核苷酸终止通过聚合酶的核酸复制,以及b)在促进靶核酸分子复制的条件下温育样品,以获得多个终止的扩增产物,其中复制通过链置换复制进行。在上述任何方法的一个实施方案中,该方法还包括从多个终止的扩增产物中分离长度为约50至约2000个核苷酸的产物。在上述任何方法的一个实施方案中,该方法还包括从多个终止的扩增产物中分离长度为约400至约600个核苷酸的产物。在上述任何方法的一个实施方案中,该方法还包括:c)修复末端和加A-尾,以及d)将步骤(c)中获得的分子连接至衔接子,从而产生扩增产物文库。在一些实施方案中,所述方法还包括从终止的扩增产物中除去终止子核苷酸。在上述任何方法的一个实施方案中,该方法还包括对扩增产物进行测序。在上述任何方法的一个实施方案中,扩增在基本等温的条件下进行。在上述任何方法的一个实施方案中,核酸聚合酶是DNA聚合酶。

[0140] 在上述任何方法的一个实施方案中,DNA聚合酶是链置换DNA聚合酶。在上述任何方法的一个实施方案中,核酸聚合酶选自噬菌体phi29(Φ 29)聚合酶、基因修饰的phi29(Φ 29)DNA聚合酶、DNA聚合酶I的Klenow片段、噬菌体M2 DNA聚合酶、噬菌体phiPRD1 DNA聚合酶、Bst DNA聚合酶、Bst大片段DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent DNA聚合酶、Deep Vent(exo-)DNA聚合酶、IsoPol DNA聚合酶、DNA聚合酶I、Therminator DNA聚合酶、T5 DNA聚合酶、测序酶、T7 DNA聚合酶、T7-测序酶和T4 DNA聚合酶。在上述任何方法的一个实施方案

中,核酸聚合酶具有3' ->5' 核酸外切酶活性,并且终止子核苷酸抑制这种3' ->5' 核酸外切酶活性。在一个特定的实施方案中,终止子核苷酸选自带有 α 基团修饰的核苷酸(例如,产生硫代磷酸酯键的 α -硫代双脱氧核苷酸)、C3间隔子核苷酸、锁核酸(LNA)、反向核酸、2' 氟核苷酸、3' 磷酸化核苷酸、2' -O-甲基修饰核苷酸和反式核酸。在上述任何方法的一个实施方案中,核酸聚合酶不具有3' ->5' 核酸外切酶活性。在一个特定的实施方案中,聚合酶选自Bst DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent(exo-)DNA聚合酶、Klenow片段(exo-)DNA聚合酶和Therminator DNA聚合酶。在一个特定的实施方案中,终止子核苷酸包括脱氧核糖的3' 碳的r基团的修饰。在一个特定的实施方案中,终止子核苷酸选自包括核苷酸的3' 封闭的可逆终止子、包括核苷酸的3' 未封闭的可逆终止子、包括脱氧核苷酸的2' 修饰的终止子、包括对脱氧核苷酸的含氮碱基的修饰的终止子及其组合。在一个特定的实施方案中,终止子核苷酸选自双脱氧核苷酸、反向双脱氧核苷酸、3' 生物素化核苷酸、3' 氨基核苷酸、3' -磷酸化核苷酸、3' -O-甲基核苷酸、包括3' C3间隔子核苷酸、3' C18核苷酸、3' 己二醇间隔子核苷酸的3' 碳间隔子核苷酸、无环核苷酸,及其组合。在上述任何方法的一个实施方案中,扩增引物的长度为4至70个核苷酸。在上述任何方法的一个实施方案中,扩增产物的长度为约50至约2000个核苷酸。在上述任何方法的一个实施方案中,靶核酸是DNA(例如,cDNA或基因组DNA)。在上述任何方法的一个实施方案中,扩增引物是随机引物。在上述任何方法的一个实施方案中,扩增引物包括条形码。在一个特定实施方案中,条形码包括细胞条形码。在一个特定实施方案中,条形码包括样品条形码。在上述任何方法的一个实施方案中,扩增引物包括独特分子标识码(UMI)。在上述任何方法的一个实施方案中,该方法包括在初始引物退火之前使靶核酸或基因组DNA变性。在一个特定的实施方案中,变性在碱性条件下进行,然后中和。在上述任何方法的一个实施方案中,样品、扩增引物、核酸聚合酶和核苷酸混合物包含在微流体装置中。在上述任何方法的一个实施方案中,样品、扩增引物、核酸聚合酶和核苷酸混合物包含在液滴中。在上述任何方法的一个实施方案中,样品选自一种或多种组织的样品、细胞、生物流体样品(例如,血液、尿液、唾液、淋巴液、脑脊液(CSF)、羊水、胸膜液、心包液、腹水、房水)、骨髓样品、精液样品、活检样品、癌症样品、肿瘤样品、细胞裂解物样品、法医样品、考古样品、古生物样品、感染样品、生产样品、整株植物、植物部分、微生物群样品、病毒制剂、土壤样品、海洋样品、淡水样品、家庭或工业样品,及其组合和分离物。在上述任何方法的一个实施方案中,样品是细胞(例如,动物细胞[例如,人类细胞]、植物细胞、真菌细胞、细菌细胞和原生动细胞)。在一个特定的实施方案中,细胞在复制之前裂解。在一个特定的实施方案中,细胞裂解伴随蛋白水解。在一个特定的实施方案中,该细胞选自来自植入前胚胎的细胞、干细胞、胎儿细胞、肿瘤细胞、疑似癌细胞、癌细胞、经过基因编辑程序的细胞、来自致病生物的细胞、从法医样品获得的细胞、从考古样品获得的细胞和从古生物样品获得的细胞。在上述任何方法的一个实施方案中,样品是来自来自植入前胚胎(例如卵裂球[例如,通过体外受精产生的八细胞期胚胎获得的卵裂球])的细胞。在一个特定的实施方案中,该方法还包括确定在胚胎细胞中疾病易感种系或体细胞变体的存在。在上述任何方法的一个实施方案中,样品是来自致病生物(例如,细菌、真菌、原生动物)的细胞。在一个特定的实施方案中,致病生物细胞从取自患者的体液、微生物群样品(例如GI微生物群样品、阴道微生物群样品、皮肤微生物群样品等)或留置医疗装置(例如,静脉导管、导尿管、脑脊髓分流器、假体

瓣膜、人工关节、气管导管等)获得的。在一个特定的实施方案中,该方法还包括确定致病生物的身份的步骤。在一个特定的实施方案中,该方法还包括确定导致致病生物对治疗有抗性的遗传变体的存在。在上述任何方法的一个实施方案中,样品是肿瘤细胞、疑似癌细胞或癌细胞。在一个特定的实施方案中,该方法还包括确定一种或多种诊断或预后突变的存在。在一个特定的实施方案中,该方法还包括确定导致对治疗有抗性的种系或体细胞变体的存在。在上述任何方法的一个实施方案中,样品是经过基因编辑程序的细胞。在一个特定的实施方案中,该方法还包括确定由基因编辑过程引起的计划外突变的存在。在上述任何方法的一个实施方案中,该方法还包括确定细胞谱系历史。在相关方面,本发明提供了上述任何方法在鉴定低频序列变体(例如,构成总序列的 $\geq 0.01\%$ 的变体)中的用途。

[0141] 在相关方面,本发明提供了包含核酸聚合酶、一种或多种扩增引物、包括一种或多种终止子核苷酸的核苷酸混合物以及任选的使用说明的试剂盒。在本发明试剂盒的一个实施方案中,核酸聚合酶是链置换DNA聚合酶。在本发明试剂盒的一个实施方案中,核酸聚合酶选自噬菌体phi29 ($\Phi 29$) 聚合酶、基因修饰的phi29 ($\Phi 29$) DNA聚合酶、DNA聚合酶I的Klenow片段、噬菌体M2 DNA聚合酶、噬菌体phiPRD1DNA聚合酶、Bst DNA聚合酶、Bst大片段DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent DNA聚合酶、Deep Vent(exo-)DNA聚合酶、IsoPol DNA聚合酶、DNA聚合酶I、Therminator DNA聚合酶、T5 DNA聚合酶、测序酶、T7 DNA聚合酶、T7-测序酶和T4 DNA聚合酶。在本发明试剂盒的一个实施方案中,核酸聚合酶具有3' →5' 核酸外切酶活性,并且终止子核苷酸抑制这种3' →5' 核酸外切酶活性(例如,带有 α 基因修饰的核苷酸[例如, α -硫代双脱氧核苷酸]、C3间隔子核苷酸、锁核酸(LNA)、反向核酸、2' 氟核苷酸、3' 磷酸化核苷酸、2' -O-甲基修饰的核苷酸、反式核酸)。在本发明试剂盒的一个实施方案中,核酸聚合酶不具有3' →5' 核酸外切酶活性(例如,Bst DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent(exo-)DNA聚合酶、Klenow片段(exo-)DNA聚合酶、Therminator DNA聚合酶)。在一个特定的实施方案中,终止子核苷酸包括脱氧核糖的3' 碳的r基团的修饰。在一个特定的实施方案中,终止子核苷酸选自包括核苷酸的3' 封闭的可逆终止子、包括核苷酸的3' 未封闭的可逆终止子、包括脱氧核苷酸的2' 修饰的终止子、包括对脱氧核苷酸的含氮碱基的修饰的终止子及其组合。在一个特定的实施方案中,终止子核苷酸选自双脱氧核苷酸、反向双脱氧核苷酸、3' 生物素化核苷酸、3' 氨基核苷酸、3' -磷酸化核苷酸、3' -O-甲基核苷酸、包括3' C3间隔子核苷酸、3' C18核苷酸、3' 己二醇间隔子核苷酸的3' 碳间隔子核苷酸、无环核苷酸,及其组合。

[0142] 本文描述了扩增基因组的方法,所述方法包括:a)使包含基因组的样品、多种扩增引物(例如,两种或更多种引物)、核酸聚合酶以及包含一种或多种终止子核苷酸的核苷酸混合物接触,所述终止子核苷酸终止聚合酶的核酸复制,和b)在促进基因组复制的条件下温育样品,以获得多个终止的扩增产物,其中复制通过链置换复制进行。在上述任何方法的一个实施方案中,该方法还包括从多个终止的扩增产物中分离长度为约50至约2000个核苷酸的产物。在上述任何方法的一个实施方案中,该方法还包括从多个终止的扩增产物中分离长度为约400至约600个核苷酸的产物。在上述任何方法的一个实施方案中,该方法还包括:c)修复末端和加A-尾,以及d)将步骤(c)中获得的分子连接至衔接子,从而产生扩增产物文库。在上述任何方法的一个实施方案中,该方法还包括对扩增产物进行测序。在上述任

何方法的一个实施方案中,扩增在基本等温的条件下进行。在上述任何方法的一个实施方案中,核酸聚合酶是DNA聚合酶。

[0143] 在上述任何方法的一个实施方案中,DNA聚合酶是链置换DNA聚合酶。在上述任何方法的一个实施方案中,核酸聚合酶选自噬菌体phi29(Φ 29)聚合酶、基因修饰的phi29(Φ 29)DNA聚合酶、DNA聚合酶I的Klenow片段、噬菌体M2 DNA聚合酶、噬菌体phiPRD1 DNA聚合酶、Bst DNA聚合酶、Bst大片段DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent DNA聚合酶、Deep Vent(exo-)DNA聚合酶、IsoPol DNA聚合酶、DNA聚合酶I、Therminator DNA聚合酶、T5 DNA聚合酶、测序酶、T7 DNA聚合酶、T7-测序酶和T4 DNA聚合酶。在上述任何方法的一个实施方案中,核酸聚合酶具有3'→5'核酸外切酶活性,并且终止子核苷酸抑制这种3'→5'核酸外切酶活性。在一个特定的实施方案中,终止子核苷酸选自带有 α 基团修饰的核苷酸(例如,产生硫代磷酸酯键的 α -硫代双脱氧核苷酸)、C3间隔子核苷酸、锁核酸(LNA)、反向核酸、2'氟核苷酸、3'磷酸化核苷酸、2'-O-甲基修饰核苷酸和反式核酸。在上述任何方法的一个实施方案中,核酸聚合酶不具有3'→5'核酸外切酶活性。在一个特定的实施方案中,聚合酶选自Bst DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent(exo-)DNA聚合酶、Klenow片段(exo-)DNA聚合酶和Therminator DNA聚合酶。在一个特定的实施方案中,终止子核苷酸包括脱氧核糖的3'碳的r基团的修饰。在一个特定的实施方案中,终止子核苷酸选自包括核苷酸的3'封闭的可逆终止子、包括核苷酸的3'未封闭的可逆终止子、包括脱氧核苷酸的2'修饰的终止子、包括对脱氧核苷酸的含氮碱基的修饰的终止子及其组合。在一个特定的实施方案中,终止子核苷酸选自双脱氧核苷酸、反向双脱氧核苷酸、3'生物素化核苷酸、3'氨基核苷酸、3'-磷酸化核苷酸、3'-O-甲基核苷酸、包括3' C3间隔子核苷酸、3' C18核苷酸、3'己二醇间隔子核苷酸的3'碳间隔子核苷酸、无环核苷酸,及其组合。在上述任何方法的一个实施方案中,扩增引物的长度为4至70个核苷酸。在上述任何方法的一个实施方案中,扩增产物的长度为约50至约2000个核苷酸。在上述任何方法的一个实施方案中,靶核酸是DNA(例如,cDNA或基因组DNA)。在上述任何方法的一个实施方案中,扩增引物是随机引物。在上述任何方法的一个实施方案中,扩增引物包括条形码。在一个特定实施方案中,条形码包括细胞条形码。在一个特定实施方案中,条形码包括样品条形码。在上述任何方法的一个实施方案中,扩增引物包括独特分子标识码(UMI)。在上述任何方法的一个实施方案中,该方法包括在初始引物退火之前使靶核酸或基因组DNA变性。在一个特定的实施方案中,变性在碱性条件下进行,然后中和。在上述任何方法的一个实施方案中,样品、扩增引物、核酸聚合酶和核苷酸混合物包含在微流体装置中。在上述任何方法的一个实施方案中,样品、扩增引物、核酸聚合酶和核苷酸混合物包含在液滴中。在上述任何方法的一个实施方案中,样品选自组织样品、细胞、生物液样品(例如,血液、尿液、唾液、淋巴液、脑脊液(CSF)、羊水、胸膜液、心包液、腹水、房水)、骨髓样品、精液样品、活检样品、癌症样品、肿瘤样品、细胞裂解物样品、法医样品、考古样品、古生物样品、感染样品、生产样品、整株植物、植物部分、微生物群样品、病毒制剂、土壤样品、海洋样品、淡水样品、家庭或工业样品,及其组合和分离物。在上述任何方法的一个实施方案中,样品是细胞(例如,动物细胞[例如,人类细胞]、植物细胞、真菌细胞、细菌细胞和原生动物细胞)。在一个特定的实施方案中,细胞在复制之前裂解。在一个特定的实施方案中,细胞裂解伴随

蛋白水解。在一个特定的实施方案中,该细胞选自来自植入前胚胎的细胞、干细胞、胎儿细胞、肿瘤细胞、疑似癌细胞、癌细胞、经过基因编辑程序的细胞、来自致病生物的细胞、从法医样品获得的细胞、从考古样品获得的细胞和从古生物样品获得的细胞。在上述任何方法的一个实施方案中,样品是来自来自植入前胚胎(例如卵裂球[例如,通过体外受精产生的八细胞期胚胎获得的卵裂球])的细胞。在一个特定的实施方案中,该方法还包括确定在胚胎细胞中疾病易感种系或体细胞变体的存在。在上述任何方法的一个实施方案中,样品是来自致病生物(例如,细菌、真菌、原生动物)的细胞。在一个特定的实施方案中,致病生物细胞从取自患者的体液、微生物群样品(例如GI微生物群样品、阴道微生物群样品、皮肤微生物群样品等)或留置医疗装置(例如,静脉导管、导尿管、脑脊髓分流器、假体瓣膜、人工关节、气管导管等)获得的。在一个特定的实施方案中,该方法还包括确定致病生物的身份的步骤。在一个特定的实施方案中,该方法还包括确定导致致病生物对治疗有抗性的遗传变体的存在。在上述任何方法的一个实施方案中,样品是肿瘤细胞、疑似癌细胞或癌细胞。在一个特定的实施方案中,该方法还包括确定一种或多种诊断或预后突变的存在。在一个特定的实施方案中,该方法还包括确定导致对治疗有抗性的种系或体细胞变体的存在。在上述任何方法的一个实施方案中,样品是经过基因编辑程序的细胞。在一个特定的实施方案中,该方法还包括确定由基因编辑过程引起的计划外突变的存在。在上述任何方法的一个实施方案中,该方法还包括确定细胞谱系历史。在相关方面,本发明提供了上述任何方法在鉴定低频序列变体(例如,构成总序列的 $\geq 0.01\%$ 的变体)中的用途。

[0144] 在相关方面,本发明提供了包含逆转录酶、核酸聚合酶、一种或多种扩增引物、包含一种或多种终止子核苷酸的核苷酸混合物以及任选的使用说明的试剂盒。在本发明试剂盒的一个实施方案中,核酸聚合酶是链置换DNA聚合酶。在一些情况下,逆转录酶进行模板切换。在一些情况下,逆转录酶是MMLV(莫洛尼鼠类白血病病毒)、HIV-1、AMV(禽成髓细胞瘤病毒)、端粒酶RT、FIV(猫类免疫缺陷病毒)或XMRV(嗜异性鼠类白血病病毒相关病毒)的变体。逆转录酶的非限制性示例包括SuperScript I (Thermo)、SuperScript II (Thermo)、SuperScript III (Thermo)、SuperScript IV (Thermo)、OmniScript (Qiagen)、SensiScript (Qiagen)、PrimeScript (Takara)、Maxima H- (Thermo)、AcuuScript Hi-Fi (Agilent)、iScript (Bio-Rad)、eAMV (Merck KGaA)、qScript (Quanta Biosciences)、SmartScribe (Clontech)或GoScript (Promega)。在本发明试剂盒的一个实施方案中,核酸聚合酶选自噬菌体phi29($\Phi 29$)聚合酶、基因修饰的phi29($\Phi 29$)DNA聚合酶、DNA聚合酶I的Klenow片段、噬菌体M2 DNA聚合酶、噬菌体phiPRD1 DNA聚合酶、Bst DNA聚合酶、Bst大片段DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent DNA聚合酶、Deep Vent(exo-)DNA聚合酶、IsoPol DNA聚合酶、DNA聚合酶I、Therminator DNA聚合酶、T5 DNA聚合酶、测序酶、T7 DNA聚合酶、T7-测序酶和T4 DNA聚合酶。在本发明试剂盒的一个实施方案中,核酸聚合酶具有3'→5'核酸外切酶活性,并且终止子核苷酸抑制这种3'→5'核酸外切酶活性(例如,带有 α 基团修饰的核苷酸[例如, α -硫代双脱氧核苷酸]、C3间隔子核苷酸、锁核酸(LNA)、反向核酸、2'氟核苷酸、3'磷酸化核苷酸、2'-O-甲基修饰的核苷酸、反式核酸)。在本发明试剂盒的一个实施方案中,核酸聚合酶不具有3'→5'核酸外切酶活性(例如,Bst DNA聚合酶、exo(-)Bst聚合酶、exo(-)Bca DNA聚合酶、Bsu DNA聚合酶、Vent_R(exo-)DNA聚合酶、Deep Vent(exo-)DNA聚合酶、Klenow片段

(exo-)DNA聚合酶、Therminator DNA聚合酶)。在一个特定的实施方案中,终止子核苷酸包括脱氧核糖的3' 碳的r基团的修饰。在一个特定的实施方案中,终止子核苷酸选自包括核苷酸的3' 封闭的可逆终止子、包括核苷酸的3' 未封闭的可逆终止子、包括脱氧核苷酸的2' 修饰的终止子、包括对脱氧核苷酸的含氮碱基的修饰的终止子及其组合。在一个特定的实施方案中,终止子核苷酸选自双脱氧核苷酸、反向双脱氧核苷酸、3' 生物素化核苷酸、3' 氨基核苷酸、3' -磷酸化核苷酸、3' -O-甲基核苷酸、包括3' C3间隔子核苷酸、3' C18核苷酸、3' 己二醇间隔子核苷酸的3' 碳间隔子核苷酸、无环核苷酸,及其组合。在一些情况下,试剂盒包含至少一种酶稳定剂、中和缓冲剂、变性缓冲剂或其组合。在一些情况下,试剂盒包含一个或多个模块。在一些情况下,试剂盒包含基因组模块和转录物组模块。

[0145] 编号实施方案

[0146] 本文描述了以下编号实施方案1-46。1. 本文描述了实施方案,所述实施方案包括一种多组学单细胞分析的方法,所述方法包括:a. 从细胞群体中分离单细胞;b. 对包含来自所述细胞的mRNA转录物扩增的多核苷酸的cDNA文库进行测序;以及c. 对所述细胞的基因组进行测序,其中对所述细胞的所述基因组进行测序包括:i. 提供来自单细胞的基因组;ii. 使所述基因组与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制;和iii. 扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行;iv. 将在步骤(ii)中获得的分子连接至衔接子,从而产生基因组DNA文库;以及v. 对所述基因组DNA文库进行测序。2. 本文还提供了如实施方案1所述的方法,其中所述方法还包括鉴定细胞表面上的至少一种蛋白质。3. 本文还提供了如实施方案1所述的方法,其中所述mRNA转录物包括聚腺苷酸化mRNA转录物。4. 本文还提供了如实施方案1所述的方法,其中所述mRNA转录物不包括聚腺苷酸化mRNA转录物。5. 本文还提供了如实施方案1-4中任一项所述的方法,其中对cDNA文库进行测序包括用模板切换引物扩增mRNA转录物。6. 本文还提供了如实施方案1-4中任一项所述的方法,其中所述cDNA文库中所述多核苷酸中的至少一些包含条形码。7. 本文还提供了如实施方案1-4中任一项所述的方法,其中所述cDNA文库中所述多核苷酸中的至少一些包含至少两种条形码。8. 本文还提供了如实施方案6或7所述的方法,其中所述条形码包括细胞条形码。9. 本文还提供了如实施方案6或7所述的方法,其中所述条形码包括样品条形码。10. 一种多组学单细胞分析的方法,所述方法包括:a. 从细胞群体中分离单细胞;b. 鉴定细胞表面上的至少一种蛋白质;以及c. 对所述细胞的基因组进行测序,其中对所述细胞的所述基因组进行测序包括:i. 提供来自单细胞的基因组;ii. 使所述基因组与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制;iii. 扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行;iv. 将在步骤(ii)中获得的分子连接至衔接子,从而产生基因组DNA文库;以及v. 对所述基因组DNA文库进行测序。11. 本文还提供了如实施方案10所述的方法,其中鉴定所述细胞表面上的至少一种蛋白质包括使所述细胞与结合至所述至少一种蛋白质的经标记的抗体接触。12. 本文还提供了如实施方案11所述的方法,其中所述经标记的抗体包含至少一种荧光标记。13. 本文还提供了如实施方案11所述的方法,其中所述经标记的抗体包含至少一种质量标签。14. 本文还提供了如实施方案11所述的方法,其中所

述经标记的抗体包含至少一种核酸条形码。15. 一种多组学单细胞分析的方法,所述方法包括:a.从细胞群体中分离单细胞;b.对所述细胞的基因组进行测序,其中对所述细胞的所述基因组进行测序包括:i.提供来自单细胞的基因组;ii.用甲基化敏感性限制性酶消化所述基因组以产生基因组片段;iii.使所述基因组片段的至少一些与至少一种扩增引物、至少一种核酸聚合酶和核苷酸混合物接触,其中所述核苷酸混合物包含至少一种终止子核苷酸,所述至少一种终止子核苷酸终止所述聚合酶的核酸复制;iv.扩增所述基因组的至少一些以产生多个终止的扩增产物,其中所述复制通过链置换复制进行;v.用甲基化特异性PCR扩增所述基因组片段的至少一些;vi.将在步骤(iv和v)中获得的分子连接至衔接子,从而产生基因组DNA文库和甲基化组DNA文库;以及vii.对所述基因组DNA文库和所述甲基化组文库进行测序。16. 本文还提供了如实施方案15所述的方法,其中鉴定所述细胞表面上的至少一种蛋白质包括使所述细胞与结合至所述至少一种蛋白质的经标记的抗体接触。17. 本文还提供了如实施方案16所述的方法,其中所述经标记的抗体包含至少一种荧光标记。18. 本文还提供了如实施方案16所述的方法,其中所述经标记的抗体包含至少一种质量标签。19. 本文还提供了如实施方案16所述的方法,其中所述经标记的抗体包含至少一种核酸条形码。20. 本文还提供了如实施方案1-19中任一项所述的方法,其中所述单细胞是哺乳动物细胞。21. 本文还提供了如实施方案1-19中任一项所述的方法,其中所述单细胞是人细胞。22. 本文还提供了如实施方案1-19中任一项所述的方法,其中所述单细胞源自肝脏、皮肤、肾脏、血液或肺。23. 本文还提供了如实施方案1-19中任一项所述的方法,其中所述单细胞是原代细胞。24. 本文还提供了如实施方案1-23中任一项所述的方法,其中所述方法还包括从所述终止的扩增产物中除去至少一种终止子核苷酸。25. 本文还提供了如实施方案1-23中任一项所述的方法,其中所述扩增产物中的至少一些包含条形码。26. 本文还提供了如实施方案1-23中任一项所述的方法,其中所述扩增产物中的至少一些包含至少两种条形码。27. 本文还提供了如实施方案24或26所述的方法,其中所述条形码包括细胞条形码。28. 本文还提供了如实施方案24或26所述的方法,其中所述条形码包括样品条形码。29. 本文还提供了如实施方案1-28中任一项所述的方法,其中所述扩增引物的至少一些包含独特的分子标识码(UMI)。30. 本文还提供了如实施方案1-28中任一项所述的方法,其中所述扩增引物的至少一些包含至少两种独特的分子标识码(UMI)。31. 本文还提供了如实施方案1-30中任一项所述的方法,其中所述方法还包括使用PCR的另外的扩增步骤。32. 本文还提供了如实施方案1-30中任一项所述的方法,其中在所述细胞的所述基因组中鉴定出至少一个突变,其中所述突变不同于参考序列中的相应位置。33. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的少于50%中发生。34. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的少于25%中发生。35. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的少于1%中发生。36. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的不超过0.1%中发生。37. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的不超过0.01%中发生。38. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的不超过0.001%中发生。39. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述细胞群体的不超过0.0001%中发生。40. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过

50%中发生。41. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过25%中发生。42. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过1%中发生。43. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过0.1%中发生。44. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过0.01%中发生。45. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过0.001%中发生。46. 本文还提供了如实施方案32所述的方法,其中所述至少一个突变在所述扩增产物序列的不超过0.0001%中发生。

[0147] 实施例

[0148] 提出以下实施例以更清楚地向本领域技术人员说明本文公开的实施方案的原理和实践,并且不应被解释为限制任何要求保护的实施方案的范围。除非另有说明,否则所有份数和百分比均以重量计。

[0149] 实施例1:原代模板定向扩增(PTA)

[0150] 尽管PTA可以用于任何核酸扩增,但它对于全基因组扩增特别有用,因为与目前使用的方法如多重置换扩增(MDA)相比,它可以以更均匀且可再现的方式捕获更大百分比的细胞基因组,并且错误率更低,避免了目前使用的方法的缺点,如在聚合酶首先延伸随机引物的位置处进行的指数扩增,而该指数扩增会导致基因座和等位基因的随机过度表达和突变传播(参见图1G)。还将PTA与其他分析技术,如转录物组分析一起使用。

[0151] 细胞培养

[0152] 将人NA12878(Coriell Institute)细胞维持在RPMI培养基中,该培养基补充有15% FBS和2mM的L-谷氨酰胺、100单位/mL的青霉素、100 μ g/mL的链霉素和0.25 μ g/mL的两性霉素B(Gibco,Life Technologies)。细胞以 3.5×10^5 个细胞/mL的密度接种。培养物每3天分离一次,并在37 $^{\circ}$ C,5% CO₂的潮湿培养箱中维持。

[0153] 单细胞分离和WTA

[0154] 将WTA(全转录物组分析)的一般方案示出于图2F中。将细胞以150-500个细胞/ μ L的浓度重悬。将该细胞悬浮液用20 μ L新鲜制备的染色缓冲剂(添加至含有1X PBS和0.05% tween-20的1.25mL细胞缓冲剂的来自Life Technology的LIVE/DEAD[®]活力/细胞毒性试剂盒的2.5 μ L乙锭同型二聚体-1和0.625 μ L钙黄绿素AM)染色。然后使用FACS Aria III分选仪分选细胞以在96个孔中的每一个中沉积细胞。将含有以下的反应混合物添加至每个孔:5x RT缓冲剂、PEG4000、RT引物(100 μ M)、TS寡核苷酸(20 μ M)、逆转录酶、RNA酶抑制剂、明胶、Tween-20、Triton-X、dNTP混合物、TMAC(1M)、甜菜碱(5M)、MgCl₂(50mM)、ERCC掺入物。然后将样品放置于热循环仪上在42 $^{\circ}$ C下持续90min,在50 $^{\circ}$ C下持续30min,且然后在4 $^{\circ}$ C下保持直到样品可处理用于预扩增。在针对RT热循环之后,将样品处理用于DNA扩增或预扩增由RT反应产生的第一链cDNA。使用单一引物(半抑制性PCR)用扩增cDNA产物的以下方案来完成样品的预扩增。简而言之,使用以下热循环条件将5 μ L RT反应物添加至含有2X主混合物、1微摩尔引物和5X预扩增缓冲剂的30微升反应物:95 $^{\circ}$ C-1min,21个循环的95 $^{\circ}$ C-15s、60 $^{\circ}$ C-30s、68 $^{\circ}$ C-4min,然后在72 $^{\circ}$ C下保持十分钟的时段。然后使用制造商的说明使用Nextera XT文库制备试剂盒将样品转化为测序文库(图2G)。RT实验的结果针对六个样品示出于表1中。

[0155] 表1

样品#	映射至基因组	映射至转录物组	所检测的基因	读段/样品	细胞
1	99.46%	94.24%	9298	2,530,484	~350
2	99.48%	94.17%	9692	2,239,730	~350
3	99.45%	94.77%	3630	2,466,899	~35
4	98.91%	94.45%	4351	2,235,999	~35
5	98.72%	92.70%	2189	1,788,750	~3
6	96.13%	89.92%	4272	1,565,289	~3

[0157] 单细胞分离和WGA

[0158] 以 3.5×10^5 细胞/ml 的密度接种后, 将 NA12878 细胞培养至少三天, 在这之后将 3mL 的细胞悬浮液以 300xg 沉淀 10 分钟。然后弃去培养基, 并且用 1mL 的细胞洗涤缓冲剂 (含有 2% FBS, 不含 Mg^{2+} 或 Ca^{2+} 的 1X PBS) 洗涤 3 次, 以 300xg、200xg 和最后 100xg 离心 5 分钟。然后将细胞重悬于 500 μ L 细胞洗涤缓冲剂中。随后用 100nM 钙黄绿素 AM (分子探针) 和 100ng/ml 碘化丙啶 (PI; Sigma-Aldrich) 染色, 以区分活细胞群。将细胞加载在已被 ELIMINase (Decon Labs) 彻底清洗过的 BD FACScan 流式细胞仪 (FACSAria II) (BD Biosciences) 上, 并且使用 Accudrop 荧光珠 (BD Biosciences) 校准, 以进行细胞分选。在要经历 PTA (Sigma-Aldrich) 的细胞中将来自钙黄绿素 AM 阳性且 PI 阴性的级分中的单细胞分入 96 孔板的每个孔中, 孔中含有 3 μ L 具有 0.2% Tween 20 的 PBS (Qiagen, REPLI-g SC 试剂盒)。有意将多个孔留空, 以用作无模板对照 (NTC)。分选后, 立即将板短暂离心并置于冰上。然后, 将细胞在 -20 $^{\circ}$ C 冷冻至少过夜。第二天, 在预 PCR 工作站上进行 WGA 反应, 该工作站提供恒定正压的 HEPA 过滤的空气, 并在每次实验前用紫外线消毒 30 分钟。

[0159] 在伴有修改的情况下进行 MDA, 所述修改先前已显示出改善扩增均匀性。具体地, 将抗核酸外切酶的随机引物 (ThermoFisher) 添加至裂解缓冲剂/混合物, 至最终浓度为 125 μ M。将 4 μ L 所得到的裂解/变性混合物添加至含有单细胞的管, 涡旋振荡, 短暂离心并在冰上温育 10 分钟。将细胞裂解液通过添加 3 μ L 淬灭缓冲剂来中和, 涡旋混合, 短暂离心, 并置于室温下。随后添加 40 μ L 扩增混合物, 然后在 30 $^{\circ}$ C 温育 8 小时, 然后通过加热至 65 $^{\circ}$ C 持续 3 分钟来终止扩增。

[0160] PTA 通过在冻融后首先进一步裂解细胞来进行, 该裂解是通过添加 5% Triton X-100 (Sigma-Aldrich) 和 20mg/ml 蛋白酶 K (Promega) 的 1:1 混合物的 2 μ L 预冷溶液。然后将细胞涡旋并短暂离心, 然后在 40 度放置 10 分钟。然后将 4 μ L 裂解缓冲剂/混合物和 1 μ L 500 μ M 抗核酸外切酶的随机引物添加至裂解的细胞, 以使 DNA 变性, 然后涡旋, 离心并在 65 度下放置 15 分钟。然后添加 4 μ L 室温淬灭缓冲剂, 并且将样品涡旋并离心分离。56 μ L 扩增混合物 (引物、dNTP、聚合酶、缓冲剂) 含有等比率的 α -硫代-ddNTP, 其在最终的扩增反应物中浓度为 1200 μ M。然后将样品在 30 $^{\circ}$ C 放置 8 小时, 然后加热至 65 $^{\circ}$ C 持续 3 分钟来终止扩增。

[0161] 扩增步骤后, 使用 AMPure XP 磁珠 (Beckman Coulter) 以 2:1 的磁珠样品比纯化来自 MDA 和 PTA 反应的 DNA, 并使用 Qubit dsDNA HS 测定试剂盒和 Qubit 3.0 荧光计根据制造商的说明 (Life Technologies) 来测量产量。

[0162] 文库制备

[0163] MDA反应导致产生40 μ g的扩增的DNA。根据标准方案,将1 μ g产物片段化30分钟。然后用15 μ M双索引衔接子(T4聚合酶的末端修复、T4多核苷酸激酶和用于加A尾的Taq聚合酶)和4个PCR循环,对样品进行标准文库制备。每个PTA反应产生40-60ng的物质,这些物质在不进行片段化的情况下全部用于标准DNA测序文库制备。在连接时使用具有UMI和双索引的2.5 μ M衔接子,并且在最终扩增中使用15个PCR(热启动聚合酶)循环。然后使用双侧SPRI来清理文库,右侧选择和左侧选择所采用的比例分别为0.65X和0.55X。在Illumina NextSeq平台上测序之前,使用Qubit dsDNA BR测定试剂盒和2100生物分析仪(Agilent Technologies)对最终文库进行量化。包括NovaSeq在内的所有Illumina测序平台也与该方案兼容。

[0164] 数据分析

[0165] 使用Bcl2fastq基于细胞条形码对测序读段进行多路解编。然后使用trimmomatic修剪读段,随后使用BWA将其与hg19比对。读段由Picard进行重复标记,然后使用GATK 4.0进行局部重新比对和碱基重新校准。所有用于计算质量指标的文件都使用Picard DownSampleSam下取样至2000万个读段。质量指标是使用qualimap以及Picard AlignmentSummaryMetrics和CollectWgsMetrics从最终bam文件获取的。总基因组覆盖也使用Preseq估算。

[0166] 变体判定

[0167] 使用来自GATK 4.0的GATK UnifiedGenotyper判定单核苷酸变体和插入/缺失。将使用GATK最佳实践的标准过滤条件用于过程中的所有步骤(<https://software.broadinstitute.org/gatk/best-practices/>)。使用Control-FREEC(Boeva等人,Bioinformatics,2012,28(3):423-5)判定拷贝数变体。结构变体还使用CREST检测(Wang等人,Nat Methods,2011,8(8):652-4)。

[0168] 结果

[0169] 如图3A和图3B所示,仅用双脱氧核苷酸(“可逆”)扩增的映射率和映射质量得分分别为15.0+/-2.2和0.8+/-0.08,而掺入抗核酸外切酶的 α -硫代双脱氧核苷酸终止子(“不可逆”)的映射率和映射质量得分分别为97.9+/-0.62和46.3+/-3.18。实验还使用可逆的ddNTP和不同浓度的终止子进行。(图2A,底部)。

[0170] 图2B-图2E展示了经过MDA(按照Dong,X.等人,Nat Methods.2017,14(5):491-493的方法)或PTA的NA12878人单细胞所产生的比较数据。虽然两种方案均产生了相当的低PCR重复率(MDA为1.26%+/-0.52,而PTA为1.84%+/-0.99)以及GC%(MDA为42.0+/-1.47,而PTA为40.33+/-0.45),但PTA产生的扩增子更小。与MDA相比,PTA的映射读段百分比和映射质量得分也明显更高(分别为PTA 97.9+/-0.62与MDA 82.13+/-0.62,以及PTA 46.3+/-3.18与MDA 43.2+/-4.21)。总体而言,与MDA相比,PTA生成更多可用的映射数据。图4A示出,与MDA相比,PTA显著提高了扩增的均匀性,覆盖范围更广且覆盖接近0的区域更少。使用PTA可以识别核酸群体中的低频序列变体,包括占总序列的 $\geq 0.01\%$ 的变体。PTA可以成功用于单细胞基因组扩增。

[0171] 实施例2:PTA的比较分析

[0172] 基准化PTA和SCMDA细胞维持和分离

[0173] 将来自1000个基因组计划对象NA12878(Coriell Institute, Camden, NJ, USA)的类淋巴母细胞维持在RPMI培养基中,该培养基中补充了15% FBS、2mM的L-谷氨酰胺、100单位/mL的青霉素、100 μ g/mL的链霉素和0.25 μ g/mL的两性霉素B)。细胞以 3.5×10^5 个细胞/ml的密度接种,并且每3天拆分一次。将它们维持在37 $^{\circ}$ C, 5% CO₂的潮湿培养箱中。在单细胞分离前,将3mL在过去3天中已扩增的细胞的悬浮液以300xg离心10分钟。将沉淀的细胞用1mL细胞洗涤缓冲剂(含有2% FBS, 不含Mg²⁺或Ca²⁺的1X PBS)洗涤三次,然后依次以300xg、200xg和最后100xg离心5分钟,以除去死细胞。然后将细胞重悬于500 μ L细胞洗涤缓冲剂中,然后用100nM钙黄绿素AM和100ng/ml碘化丙啶(PI)染色,以区分活细胞群。将细胞加载至已被ELIMINase彻底清洗并使用Accudrop荧光珠校准的BD FACScan流式细胞仪(FACSAria II)上。将来自钙黄绿素AM阳性且PI阴性级分的单细胞分选入96孔板的每个孔中,孔中含有3 μ L具有0.2% Tween20的PBS。有意将多个孔留空,以用作无模板对照。分选后,立即将板短暂离心并置于冰上。然后,将细胞在-80 $^{\circ}$ C冷冻至少过夜。

[0174] PTA和SCMDA实验

[0175] WGA反应在预PCR工作站上组装,该工作站使用HEPA过滤后的空气提供恒定正压,并在每次实验前用紫外线消毒30分钟。MDA是根据SCMDA方法根据已公开的方案进行的(Dong等人, Nat. Meth. 2017, 14, 491-493)。具体而言,将抗核酸外切酶的随机引物以最终浓度12.5 μ M添加至裂解缓冲剂。将4 μ L所得的裂解混合物添加至含有单细胞的管中,移液三次以混合,短暂离心并在冰上温育10分钟。细胞裂解液通过添加3 μ L淬灭缓冲剂来中和,移液3次混合,短暂离心,并置于冰上。随后添加40 μ L扩增混合物,然后在30 $^{\circ}$ C温育8小时,然后通过加热至65 $^{\circ}$ C持续3分钟来终止扩增。PTA通过在冻融后首先进一步裂解细胞来进行,该裂解是通过添加5% Triton X-100和20mg/ml蛋白酶K的1:1混合物的2 μ L预冷溶液。然后将细胞涡旋并短暂离心,然后在40度放置10分钟。然后将4 μ L变性缓冲剂和1 μ L 500 μ M抗核酸外切酶的随机引物添加至裂解的细胞,以使DNA变性,然后涡旋,离心并在65 $^{\circ}$ C放置15分钟。然后添加4 μ L室温淬灭溶液,并将样品涡旋并离心分离。56 μ L扩增混合物含有等比例的 α -硫代-ddNTP,其在最终的扩增反应中浓度为1200 μ M。然后将样品在30 $^{\circ}$ C放置8小时,然后加热至65 $^{\circ}$ C持续3分钟来终止扩增。在SCMDA或PTA扩增后,使用AMPure XP磁珠以2:1的磁珠样品比纯化DNA,并使用Qubit dsDNA HS测定试剂盒和Qubit 3.0荧光计根据制造商的说明来测量产量。

[0176] 文库制备

[0177] 在添加调节溶液后,根据HyperPlus方案,将1 μ g SCMDA产物片段化30分钟。然后用15 μ M独特双索引衔接子和4个PCR循环,对样品进行标准文库制备。将每个PTA反应的全部产物在不进行片段化的情况下用于使用标准扩增方案的DNA测序文库制备。在连接时使用2.5 μ M独特双索引衔接子,并且在最终扩增中使用15个PCR循环。然后在1%琼脂糖E-Ge1上将来自SCMDA和PTA的文库可视化。从凝胶切下400-700bp之间的片段,并使用Ge1 DNA回收试剂盒回收。在NovaSeq6000上测序之前,使用Qubit dsDNA BR测定试剂盒和Agilent 2100生物分析仪对最终文库进行量化。

[0178] 数据分析

[0179] 用trimmomatic修剪数据,随后使用BWA将其与hg19比对。读段由Picard进行重复标记,然后使用GATK 3.5最佳实践进行局部重新比对和碱基重新校准。使用Picard

DownSampleSam将所有文件下取样至指定的读段数。质量指标是使用qualimap以及Picard AlignmentMetricsSummary和CollectWgsMetrics从最终bam文件获取的。绘制洛伦兹曲线，并使用htSeqTools计算基尼指数。使用UnifiedGenotyper进行SNV判定，然后使用标准推荐条件(QD<2.0||FS>60.0||MQ<40.0||SOR>4.0||MQRankSum<-12.5||ReadPosRankSum<-8.0)进行过滤。分析中没有排除任何区域，也没有进行其他数据归一化或操作。表2中列出了所测试方法的测序指标。

[0180] 表2:所测试的方法之间的测序指标的比较。

	PTA	MDA 试剂 盒 2	PicoPlex	MALBA C	LIANT I	MDA 试剂盒 1	DOP PCR
[0181] % 基因组映射	97	88	55	79	92	65	52
% 基因组回收 (300 M 读 段)	95	75	43	60	82	73	23
覆盖的 CV (300 M 读 段)	0.8	1.8	3	2.5	1.1	2	3.5
[0182] SNV 灵敏度 % (300 M 读 段)	76	50	15	34	49	46	5
SNV 特异性 % (300 M 读 段)	93	91	56	47	88	90	35

[0183] CV=变异系数;SNV=单核苷酸变异;值是指15X覆盖。

[0184] 基因组覆盖范围和均匀性

[0185] 将PTA与所有常见的单细胞WGA方法进行综合比较。为实现此目的,分别对10个NA12878细胞进行了PTA和改进版本的MDA,称为单细胞MDA(Dong等人,Nat.Meth.2017,14,491-493)(SCMDA)。此外,使用作为LIANTI研究的一部分产生的数据来比较通过DOP-PCR(Zhang等人,PNAS1992,89,5847-5851)、MDA试剂盒1(Dean等人,PNAS2002,99,5261-5266)、MDA试剂盒2、MALBAC(Zong等人,Science 2012,338,1622-1626)、LIANTI(Chen等人,Science 2017,356,189-194)或PicoPlex(Langmore,Pharmacogenomics 3,557-560(2002))扩增的细胞的结果。

[0186] 为了在样品之间进行归一化,将来自所有样品的原始数据进行比对并进行预处理,以使用相同的管线进行变体判定。然后,在进行比较之前,将bam文件下取样至3亿个读段。重要的是,在进行进一步分析之前未筛选PTA和SCMDA产物,而所有其他方法均在选择后续分析中使用的最高质量细胞之前进行了基因组覆盖和均匀性的筛选。值得注意的是,将SCMDA和PTA与大量二倍体NA12878样品进行比较,而所有其他方法均与LIANTI研究中使用的大量BJ1二倍体成纤维细胞进行比较。如图3C-图3F所示,PTA具有最高百分比的与基因组对齐的读段,以及最高的映射质量。PTA、LIANTI和SCMDA具有相似的GC含量,均低于其他方法。在所有方法中,PCR复制率相似。此外,相对于其他经测试方法,PTA方法使较小的模板(如线粒体基因组)能够提供更高的覆盖率(类似于较大的典型染色体)(图3G)。

[0187] 然后比较所有方法的覆盖范围和均匀性。展示了SCMDA和PTA在染色体1上的覆盖图示例,其中示出,PTA具有显著改进的覆盖均匀性和等位基因频率(图4B)。然后,使用增加的读段数来计算所有方法的覆盖率。PTA在每个深度下都接近两种大量样品,这是相对于所有其他方法的显著改进(图5A)。然后,我们使用两种策略来衡量覆盖均匀性。第一种方法是在增加测序深度时计算覆盖变异系数,其中发现PTA比所有其他方法更均匀(图5B)。第二种策略是计算每个下取样bam文件的洛伦兹曲线,其中再次发现PTA具有最大的均匀性(图5C)。为了测量扩增均匀性的可再现性,计算了基尼指数以估算每个扩增反应与完全均匀性的差异(de Bourcy等人,PloS one 9,e105585(2014))。PTA再次显示比其他方法具有更高的可再现性(图5D)。

[0188] SNV灵敏度

[0189] 为了确定扩增方法性能的这些差异对SNV判定的影响,在增加的测序深度下比较了每种方法与相应大量样品的变体判定比率。为了估计灵敏度,比较了在每个测序深度下,在每个细胞中发现的在被下取样至6.5亿个读段的相应大量样品中判定的变体百分比(图5E)。PTA的覆盖和均匀性的改善使得检测出的变体比MDA试剂盒2多45.6%,而MDA试剂盒2是第二高灵敏度的方法。在大量样品中判定为杂合的位点的检查显示,PTA显著减少了那些杂合位点的等位基因倾斜(图5F)。这一发现支持了PTA不仅在整個基因组中具有更均匀的扩增,而且在同一细胞中还更均匀地扩增两个等位基因的观点。

[0190] SNV特异性

[0191] 为了估计突变判定的特异性,在每个单细胞中判定的未在相应的大量样品中发现的变体被认为是假阳性。SCMDA的较低温度裂解显著减少了假阳性变体判定的数目(图5G)。使用热稳定聚合酶的方法(MALBAC、PicoPlex和DOP-PCR)显示,随着测序深度的增加,SNV判

定的特异性进一步降低。不受理论约束,这很可能是这些聚合酶的错误率与phi29 DNA聚合酶相比显著提高的结果。此外,在假阳性判定中看到的碱基改变模式也似乎是聚合酶依赖性的(图5H)。如图5G所示,与标准MDA方案相比,PTA中较低的假阳性SNV判定率支持了PTA中抑制错误传播的模型。此外,PTA的假阳性变体判定的等位基因频率最低,这再次与PTA抑制错误传播的模型一致(图5I)。

[0192] 实施例3:大规模并行单细胞DNA测序

[0193] 使用PTA,建立了用于大规模并行DNA测序的方案。首先,将细胞条形码添加至随机引物。采用了两种策略来最大程度地减少细胞条形码引入的扩增中的任何偏差:1) 延长随机引物的大小和/或2) 创建引物,使其自身环回,以防止细胞条形码与模板结合(图10B)。一旦建立了最佳引物策略,则可以使用,例如,Mosquito HTS液体处理器对多达384种分选的细胞进行扩展,该处理器可以高精度地将粘性液体移至25nL的体积。通过使用1 μ L PTA反应代替标准的50 μ L反应体积,该液体处理器还可将试剂成本降低约50倍。

[0194] 通过将带有细胞条形码的引物递送至液滴,将扩增方案转化到液滴中。任选地使用固体支持物,如使用裂池策略创建的珠。例如,合适的珠可以从ChemGenes获得。在一些情况下,寡核苷酸含有随机引物、细胞条形码、独特分子标识码以及可切割的序列或间隔子,以在将珠和细胞封装在同一液滴中后释放寡核苷酸。在此过程中,优化液滴中低纳升体积的模板、引物、dNTP、 α -硫代-ddNTP和聚合酶浓度。在一些情况下,优化包括使用较大的液滴以增加反应体积。如图9所示,此过程需要两个连续的反应来裂解细胞,然后进行WGA。含有裂解细胞和珠的第一液滴与含有扩增混合物的第二液滴结合。替代地或组合地,细胞在裂解前封装在水凝胶珠中,然后将两种珠添加至油滴中。参见Lan, F. 等人, *Nature Biotechnol.*, 2017, 35:640-646)。

[0195] 其他的方法包括使用微孔,在一些情况下,微孔在3"×2"的显微镜载玻片大小的装置上的20皮升反应室中捕获140,000个单细胞。与基于液滴的方法类似,这些孔将细胞与含有细胞条形码的珠结合,从而允许进行大规模并行处理。参见Gole等人, *Nature Biotechnol.*, 2013, 31:1126-1132。

[0196] 实施例4:单细胞中基因组和转录物组的平行分析

[0197] 分选来自细胞群体的单细胞,每孔放置一个细胞。每个孔包括固定至表面区域的抗体,其中抗体与细胞核结合。将细胞的外膜裂解,从而将mRNA释放至孔中的溶液中,而核酸酶保持完整并结合至孔的区域。使用溶液中的mRNA作为模板进行RT以使用图8A中的引物产生cDNA。任选地,进行rRNA(核糖体RNA)耗尽步骤。将从5'至3'包含TSS区域(转录起始位点)、锚定区域、RNA BC区域和聚dT尾的第一模板;和从5'至3'包含TSS区域、锚定区域和聚G区域的第二模板用于RT PCR。在取出RT PCR产物(cDNA文库)以进行后续测序之后,通过UNG除去细胞中的任何剩余RNA。使用基于Nextera/转座子的测序方法和试剂制备RNA文库(图8B)。cDNA文库包含具有大约1000倍扩增的短cDNA。然后将核裂解,并且使用随机引物和等温聚合酶对释放的基因组DNA进行PTA方法,其中随机引物的长度为6-9个碱基。选择用于PTA的扩增条件以产生长度为250-1500个碱基的扩增子。任选地对PTA产物进行另外的扩增,并且进行测序。将RNA测序数据和DNA测序数据编译至数据库中以进行分析。

[0198] 实施例5:单细胞多组学分析

[0199] 将细胞群体与抗体文库接触,其中将抗体标记。将抗体用荧光标记、核酸条形码或

两者标记。经标记的抗体与群体中的至少一种细胞结合,并且将此类细胞分选,每孔放置一个细胞。一些经标记的抗体提供有关结合之后的细胞表面蛋白标记物的特定信息,所述特定信息通过荧光显微镜检查或标记至抗体的条形码的读取来获得。每个孔包括固定至表面区域的抗体,其中抗体与细胞核结合。将细胞的外膜裂解,从而将mRNA释放至孔中的溶液中,而核酸酶保持完整并结合至孔的区域。任选地,进行rRNA(核糖体RNA)耗尽步骤。接下来,使用溶液中的mRNA作为模板来进行RT以产生cDNA。将从5'至3'包含TSS区域(转录起始位点)、锚定区域、RNA BC区域和聚dT尾的第一模板;和从5'至3'包含TSS区域、锚定区域和聚G区域的第二模板用于RT PCR。在取出RT PCR产物(cDNA文库)以进行后续测序之后,通过UNG除去细胞中的任何剩余RNA。cDNA文库包含具有大约1000倍扩增的短cDNA。然后将核裂解,并且使用随机引物和等温聚合酶对释放的基因组DNA进行PTA方法,其中随机引物的长度为6-9个碱基。选择用于PTA的扩增条件以产生长度为250-1500个碱基的扩增子。任选地对PTA产物进行另外的扩增,并且进行测序。将蛋白质数据、RNA测序数据和DNA测序数据编译至数据库中以进行分析。

[0200] 实施例6:甲基化组和转录物组的单细胞分析

[0201] 分选来自细胞群体的单细胞,每孔放置一个细胞。每个孔包括固定至表面区域的抗体,其中抗体与细胞核结合。将细胞的外膜裂解,从而将mRNA释放至孔中的溶液中,而核酸酶保持完整并结合至孔的区域。使mRNA转录物与末端转移酶接触以将核糖鸟嘌呤添加至mRNA链的5'端。接下来,使用溶液中的mRNA作为模板来进行RT以产生cDNA。任选地,进行rRNA(核糖体RNA)耗尽步骤。将从5'至3'包含TSS区域(转录起始位点)、锚定区域、RNA BC区域和聚dT尾的第一模板;和从5'至3'包含TSS区域、锚定区域和聚G区域的第二模板用于RT PCR。在取出RT PCR产物(cDNA文库)以进行后续测序之后,通过UNG除去细胞中的任何剩余RNA。cDNA文库包含具有大约1000倍扩增的短cDNA。然后将核裂解,并且使用甲基化敏感性核酸内切酶将释放的基因组DNA片段化。使用随机引物和等温聚合酶对基因组片段进行PTA方法,其中随机引物的长度为6-9个碱基。选择用于PTA的扩增条件以产生长度为250-1500个碱基的扩增子。任选地对PTA产物进行另外的扩增,并且进行测序。将RNA测序数据和DNA测序数据编译至数据库中以进行分析,并且鉴定出甲基化敏感性核酸内切酶切割位点。使用这些位点来映射原始基因组DNA上甲基化的位置。

[0202] 实施例7:甲基化组和基因组的单细胞分析

[0203] 分选来自细胞群体的单细胞,每孔放置一个细胞。每个孔包括固定至表面区域的抗体,其中抗体与细胞核结合。将细胞用甲基化敏感性酶裂解,并且使用随机引物和等温聚合酶对基因组进行PTA方法,其中随机引物的长度为6-9个碱基。选择用于PTA的扩增条件以产生长度为250-1500个碱基的扩增子。将反应混合物拆分,其中对一半混合物进行外显子组富集、全基因组测序或其他靶向测序方法。对另一半反应混合物进行甲基化敏感性PCR条件。将甲基化和DNA测序数据编译至数据库中以进行分析。

[0204] 实施例8:表面蛋白质组和基因组的单细胞分析

[0205] 将来自包含细胞群体的样品的细胞与诱饵的文库,如抗体、多核苷酸或其他小分子的文库接触。在一些情况下,将诱饵条形码化(如条形码化抗体),以允许下拉和鉴定诱饵与细胞表面上的蛋白质的结合。可替代地或组合地,将诱饵用其他标记(如荧光标记或质量标签)进行标记。分选来自细胞群体的单细胞,每孔放置一个细胞。任选地,取出与细胞表面

结合的诱饵以在基因组文库制备之前进行测序或鉴定。将细胞裂解,将基因组释放至溶液中,并且产生片段。使用随机引物和等温聚合酶对基因组片段进行PTA方法,其中随机引物的长度为6-9个碱基。可替代地,在用PTA扩增前未将基因组片段化。选择用于PTA的扩增条件以产生长度为250-1500个碱基的扩增子。任选地对PTA产物进行另外的扩增,并且进行测序。将细胞表面蛋白和DNA测序数据编译至数据库中以进行分析。

[0206] 实施例9:用于测量药物抗性的多组学

[0207] 在AML(急性髓性白血病)中用靶向FLT3的小分子抑制剂的单一疗法已显示出临床益处,但抗性总是发生。FLT3抑制剂奎扎替尼(AC220)是一种这样的抑制剂,其中药物在复发性或难治性AML患者中产生大约50%的复合完全缓解。尽管这一成功,在接受奎扎替尼疗法时复发的FLT3-ITD患者中已鉴定出激活环(D835)和中和门卫残基F691处的继发性FLT3突变。确定对多激酶抑制剂PKC412的临床抗性是FLT3激酶结构域中的继发性突变的结果。在FLT3-ITD AML中已鉴定出针对靶向疗法的另外FLT3独立性抗性模式,包括AXL的旁路途径激活,以及NRAS、TET2和IDH1/2突变。还观察到表观遗传修饰酶和转录因子中的突变,突出了对FLT3抑制的抗性的机制复杂性和多样性。

[0208] 产生了奎扎替尼抗性和匹配的亲本MOLM-13AML细胞系,以及具有杂合FLT3-ITD突变的细胞系。PTA方法是组合的RNAseq化学,并且用于以基因组和转录方式探测这些药物抗性单细胞,以便深入了解AML中FLT3抑制后的抗性机制。简而言之, workflow包括(1)产生抗性细胞,(2)分离抗性细胞,(3)胞质裂解以释放mRNA,(4)逆转录以从mRNA产生cDNA,(5)核裂解以释放基因组DNA,(6)PTA扩增,(7)分离DNA/RNA富集,(8)cDNA预扩增富集的mRNA,(9)文库制备、QC和池化,(10)下一代产生测序,以及(11)数据分析。

[0209] 细胞培养。具有杂合FLT3内部串联重复(ITD)1的MOLM-13急性髓性白血病细胞从DSMZ-德国微生物和细胞培养物保藏中心(ACC 554)获得。将细胞维持在补充有10% FBS和青霉素/链霉素的RPMI 1640(Gibco 11875-093)中,并且每2-3天继代培养,同时保持 $2.5E5-1.5E6$ 个细胞/ml的密度范围。为了产生奎扎替尼抗性MOLM-13系,将细胞用2nM奎扎替尼连续处理细胞,并且在每次继代培养下补充药物,直到在培养的5周持续时间出现抗性克隆(图9A)。在FACS分选时从奎扎替尼抗性和匹配的亲本MOLM-13细胞中分离出基因组DNA或总RNA,以产生用于与单细胞数据集比较的大量测序对照文库。

[0210] FACS。对于单细胞分析,将约 $2.0E6$ 个MOLM-13奎扎替尼抗性或匹配的亲本细胞在缺乏钙和酶、补充有2% FBS的杜尔贝科氏磷酸盐缓冲盐水(Gibco)冲洗两次,并且保持在冰上,直到BD FACSAria III FACS分选。在钙黄绿素AM、碘化丙啶和DAPI染色后,建立活细胞门控(DAPI/PI阴性,前70%钙黄绿素-AM阳性)并且将单细胞分选(130微米喷嘴组件)分选至含有细胞缓冲剂的低结合96孔PCR板(半裙边)中并且在短暂的涡旋和离心后立即在干冰上冷冻。

[0211] 组合的基因组/转录物组分析。首先,在模板切换逆转录反应中利用生物素缀合的寡dT引物,以产生来自单MOLM-13亲本或奎扎替尼抗性细胞的第一链cDNA。在逆转录后连续进行原代模板定向扩增(PTA)。然后将第一链cDNA使用链霉亲和素M-280珠亲和纯化,并进行两次高盐洗涤,然后进行一次低盐洗涤。进行20个循环的预扩增以产生第2链cDNA,并且使用Nextera DNA Flex文库制备试剂盒制备RNA测序文库。为了制备PTA文库,使用珠纯化未结合至链霉亲和素珠的PTA产物并将其连接至TruSeq衔接子。将来自PTA反应的扩增产物

首先通过珠净化进行纯化,通过Qubit测量并且通过电泳分析。哺乳动物细胞(约6pg DNA)的典型产量为1-3ug,其中产生高达50ng的单细菌基因组(2-4fg)。通过PTA扩增的样品的扩增子产物大小在0.2-4kb之间(平均为1.5Kb)。在不进行WGS方法的片段化的情况下制备PTA文库,并且产生大约500ng的产量,其中大小范围为300-550个碱基。通过NovaSeq分析来自哺乳动物细胞的全基因组,目标为约5.5亿个读段。然后将测序文件传输用于修剪比对和VCF文件创建,并且通过基于Trailblazer™云端的生物信息学平台解决方案分析。QC和文库制备时间为4-6小时。使用单独RNASeq进行平行实验以进行比较。

[0212] 结果。来自亲本培养物和抗性培养物两者的RNA表达证明了使用单锅RNAseq化学产生cDNA池(图9B)的能力,并且这些细胞中表达的基因产生能够通过每细胞检测的平均约10K个基因上的基因表达可视化细胞群的不同图案。在单独的工作流中,使用PTA方法扩增单细胞基因组。然后将两种方案组合(产量在图9D中)以产生来自每个细胞的组合转录组和基因组cDNA池。低通量(约500万个读段/细胞)展示了抗性系和亲本系两者的有效扩增和文库制备,具有低的线粒体染色体量和高完全PreSeq基因组估计(图10A-10C)。数据证明了相比于DNA,在RT步骤期间产生的转录物未通过PTA反应有效扩增,并且相比于来自单细胞的标准PTA扩增基因组,使用组合方案有效扩增单细胞中的DNA(图9D)。组合的RNASeq/PTA方法产生与标准PTA方案类似的结果(图10A),其中ChrM和重复百分比通常小于2%,并且估计基因组大小大于30亿个碱基(图10A-10C)。基因组的评价揭示出超过90%的映射率和覆盖率,并且对每个细胞中的单一核苷酸变体的特定判定超过75%。与标准PTA基因组化学相比,在双重方案中观察到更多变异。对于转录物组,原型化学显现为检测含有外显子-外显子连接点的约3000-5000个基因。与仅RNAseq方案(图9C)相比,在双重方案中检测到基因中的约30%(图10D)。另外,将双重/组合的RNASeq/PTA方案与第二抗性细胞系SUM159(三阴性乳腺癌细胞系)一起使用。两种方案中运行的RNAseq数据产生了类似的PCA分布,这表明组合化学能够检测不限于亲本细胞和抗性细胞的单细胞类型的差异基因表达。

[0213] (图10E-10F)。

[0214] 进行7个亲本和5个抗性molm13细胞的深度测序至25x的近似深度(图11)。使用bwa mem将读段与Hg38比对。使用GATK4最佳操作进行质量控制和SNV判定。仅在如果SNV限于至少2个抗性细胞情况下考虑SNV,在任何亲本细胞中未判定替代性等位基因,并且对至少6个亲本细胞基因型化。所有细胞在1x覆盖率下覆盖了至少96%的基因组,并且在10x下覆盖至少76%。插图展示了在所有细胞(为清楚起见展示了4个)中检测出molm13细胞中的已知Flt3插入缺失。

[0215] RNAseq和PTA方法通常是可比的,其中映射率和覆盖率两者均超过95%,并且ChrM和PCR重复大体上低于2.0%。另外,在sum 159亲本和抗性细胞系的选择样品中超过95%的基因组被回收。对于Molm13细胞系,鉴定出过表达基因GAS6(L),其是奎扎替尼抗性的已知机制。Gas6是AXL的配体,AXL是经历奎扎替尼治疗失败的复发患者中的临床相关抗性机制(图11B)。由双重方案对亲本和抗性MOLM13细胞系的深度基因组测序检测出分布在所有染色体上的突变。总起来说,在所有单一细胞中,鉴定出奎扎替尼抗性群体所特有的5675个SNV。检测到编码序列变异,然而,大多数观察到的变体在基因间空间中。不受理论的约束,虽然随突变无疑存在于该变体队列中,但这表明增强子或启动子水平下的基因表达的调控是有助于抗性并且潜在地有助于对非编码RNA的调控。双重mRNAseq转录物组化学/PTA具

有检测单细胞中超过10K个基因的能力,所述单细胞可以通过FACS富集。PTA方法具有回收单独细胞的完全基因组的超过97%的能力。回收转录物组和基因组两者的能力未显著影响回收大部分基因组的能力的灵敏度。当比较仅转录物组或组合的转录物组/基因组扩增化学时,可以在许多细胞中检测到超过70%的表达基因。

[0216] 实施例10:用外显子组捕获的PTA单细胞分析

[0217] 使用实施例3的一般PTA方法并进行以下修改:利用另外的外显子组捕获步骤来富集PTA产生的扩增子。对于单细胞样品(27个样品)和大量样品(112个样品)两者获得6000万个读段。将来自单细胞的外显子组捕获测序结果与大量样品的那些进行比较(图12A-12D、13A、14A和14B)。测序结果在多个样品之间是一致的(图13A),并且捕获扩增子的平均大小是623个碱基(图13B)。

[0218] 实施例11:外显子组捕获+多组学

[0219] 使用实施例5-8中任一个的一般方法并进行以下修改:利用另外的捕获步骤来富集从基因组DNA产生的PTA生成扩增子。捕获步骤包括外显子组组或靶向特定基因的其他组。在一些情况下,此类组涉及癌症热点、病毒基因组或线粒体DNA。

[0220] 本文中描述的实施例,对于本领域技术人员显而易见的是,这些实施方案仅以示例的方式提供。本领域技术人员在不脱离本发明的情况下现将想到多种变化、改变和替代。应当理解,本文中所述的本发明实施方案的各种替代方案可用于实施本发明。以下权利要求旨在限定本发明的范围,并由此涵盖这些权利要求范围内的方法和结构及其等同项。

- <110> 铂赛基因组学公司
- <120> 单细胞分析
- <130> 55461-703.601
- <140> PCT/US2020/044338
- <141> 2020-07-30
- <150> 62/881,183
- <151> 2019-07-31
- <160> 20
- <170> PatentIn 版本 3.5
- <210> 1
- <211> 100
- <212> DNA
- <213> 人工序列
- [0001] <220>
- <223> 对人工序列的描述：合成的多核苷酸
- <220>
- <221> 经修饰的_碱基
- <222> (63)..(68)
- <223> a、c、t、g、未知或其他
- <220>
- <221> 经修饰的_碱基
- <222> (100)..(100)
- <223> a、c、t、g、未知或其他
- <400> 1
- tgtcgttggtt gtcggccttc tctggaagca gtggtatcaa cgcagagtac ggacgttcag 60
- agnnnnnntt tttttttttt tttttttttt ttttttttvn 100
- <210> 2

	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的寡核苷酸	
	<400> 2	
	aagcagtggg atcaacgcag agt	23
	<210> 3	
	<211> 30	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的寡核苷酸	
[0002]	<220>	
	<223> 对合并的 DNA/RNA 分子的描述: 合成的寡核苷酸	
	<400> 3	
	aagcagtggg atcaacgcag agtacctggg	30
	<210> 4	
	<211> 15	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的寡核苷酸	
	<400> 4	
	gtctcgtggg ctcgg	15
	<210> 5	
	<211> 47	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的寡核苷酸	
	<400> 5	
	caagcagaag acggcatacg agattcgct tagtctcgtg ggctcgg	47
	<210> 6	
	<211> 66	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的寡核苷酸	
	<400> 6	
	aatgatacgg cgaccaccga gatctacacg cctgtccgcg gaagcagtgg tatcaacgca	60
[0003]	gagtac	66
	<210> 7	
	<211> 68	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的寡核苷酸	
	<400> 7	
	aatgatacgg cgaccaccga gatctacacg cctgtccgcg gaagcagtgg tatcaacgca	60
	gagtacgg	68
	<210> 8	
	<211> 28	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	对人工序列的描述：合成的寡核苷酸	
	<220>		
	<221>	经修饰的_碱基	
	<222>	(1)..(1)	
	<223>	a、c、t、g、未知或其他	
	<400>	8	
		nbaaaaaaaa aaaaaaaaaa aaaaaaaaa	28
	<210>	9	
	<211>	30	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	对人工序列的描述：合成的寡核苷酸	
[0004]	<220>		
	<223>	对合并的 DNA/RNA 分子的描述：合成的寡核苷酸	
	<400>	9	
		cccatgtact ctgcgttgat accactgctt	30
	<210>	10	
	<211>	100	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	对人工序列的描述：合成的多核苷酸	
	<220>		
	<221>	经修饰的_碱基	
	<222>	(1)..(1)	
	<223>	a、c、t、g、未知或其他	

	<220>		
	<221>	经修饰的_碱基	
	<222>	(33).. (38)	
	<223>	a、c、t、g、未知或其他	
	<400>	10	
		nvaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aannnnnct ctgaacgtcc gtactctgcg	60
		ttgataccac tgcttcaga gaagccgac aacaacgaca	100
	<210>	11	
	<211>	75	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	对人工序列的描述：合成的寡核苷酸	
[0005]	<220>		
	<221>	经修饰的_碱基	
	<222>	(1).. (1)	
	<223>	a、c、t、g、未知或其他	
	<220>		
	<221>	经修饰的_碱基	
	<222>	(33).. (38)	
	<223>	a、c、t、g、未知或其他	
	<400>	11	
		nvaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aannnnnct ctgaacgtcc gtactctgcg	60
		ttgataccac tgctt	75
	<210>	12	
	<211>	75	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	对人工序列的描述：合成的	

寡核苷酸

<220>

<221> 经修饰的_碱基

<222> (38).. (43)

<223> a、c、t、g、未知或其他

<220>

<221> 经修饰的_碱基

<222> (75).. (75)

<223> a、c、t、g、未知或其他

<400> 12

aagcagtggt atcaacgcag agtacggacg ttcagagnnn mnnttttttt tttttttttt 60

tttttttttt tttvn

75

<210> 13

<211> 19

[0006]

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的
寡核苷酸

<400> 13

ctgtctctta tacacatct 19

<210> 14

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的
寡核苷酸

<400> 14

gtctcgtggg ctcggagatg tgtataagag acag 34

	<210> 15	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的 多核苷酸	
	<220>	
	<221> 经修饰的_碱基	
	<222> (79).. (84)	
	<223> a、c、t、g、未知或其他	
	<220>	
	<221> 经修饰的_碱基	
	<222> (116).. (116)	
	<223> a、c、t、g、未知或其他	
	<400> 15	
	aatgatacgg cgaccaccga gatctacacg cctgtccgcg gaagcagtgg tatcaacgca	60
[0007]		
	gagtiacggac gttcagagnn nnnntttttt tttttttttt tttttttttt ttttvn	116
	<210> 16	
	<211> 34	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述: 合成的 寡核苷酸	
	<400> 16	
	ctgtctctta tacacatctc cgagcccacg agac	34
	<210> 17	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	

	<223> 对人工序列的描述：合成的 多核苷酸	
	<220>	
	<221> 经修饰的_碱基	
	<222> (1).. (1)	
	<223> a、c、t、g、未知或其他	
	<220>	
	<221> 经修饰的_碱基	
	<222> (33).. (38)	
	<223> a、c、t、g、未知或其他	
	<400> 17	
	nvaaaaaaaaa aaaaaaaaa aaaaaaaaa aannnnnnct ctgaacgtcc gtactctgcg	60
	ttgataccac tgcttccgcg gacaggcgtg tagatctcgg tggtcgccgt atcatt	116
	<210> 18	
	<211> 66	
[0008]	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述：合成的 寡核苷酸	
	<400> 18	
	caagcagaag acggcatacg agattcgcct tagtctcgtg ggctcggaga tgtgtataag	60
	agacag	66
	<210> 19	
	<211> 66	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述：合成的 寡核苷酸	
	<400> 19	

	ctgtctctta tacacatctc cgagcccacg agactaagge gaatctcgta tgccgtcttc	60
	tgcttg	66
	<210> 20	
	<211> 37	
	<212> DNA	
[0009]	<213> 人工序列	
	<220>	
	<223> 对人工序列的描述：合成的寡核苷酸	
	<400> 20	
	gcctgtccgc ggaagcagtg gtatcaacgc agagtac	37

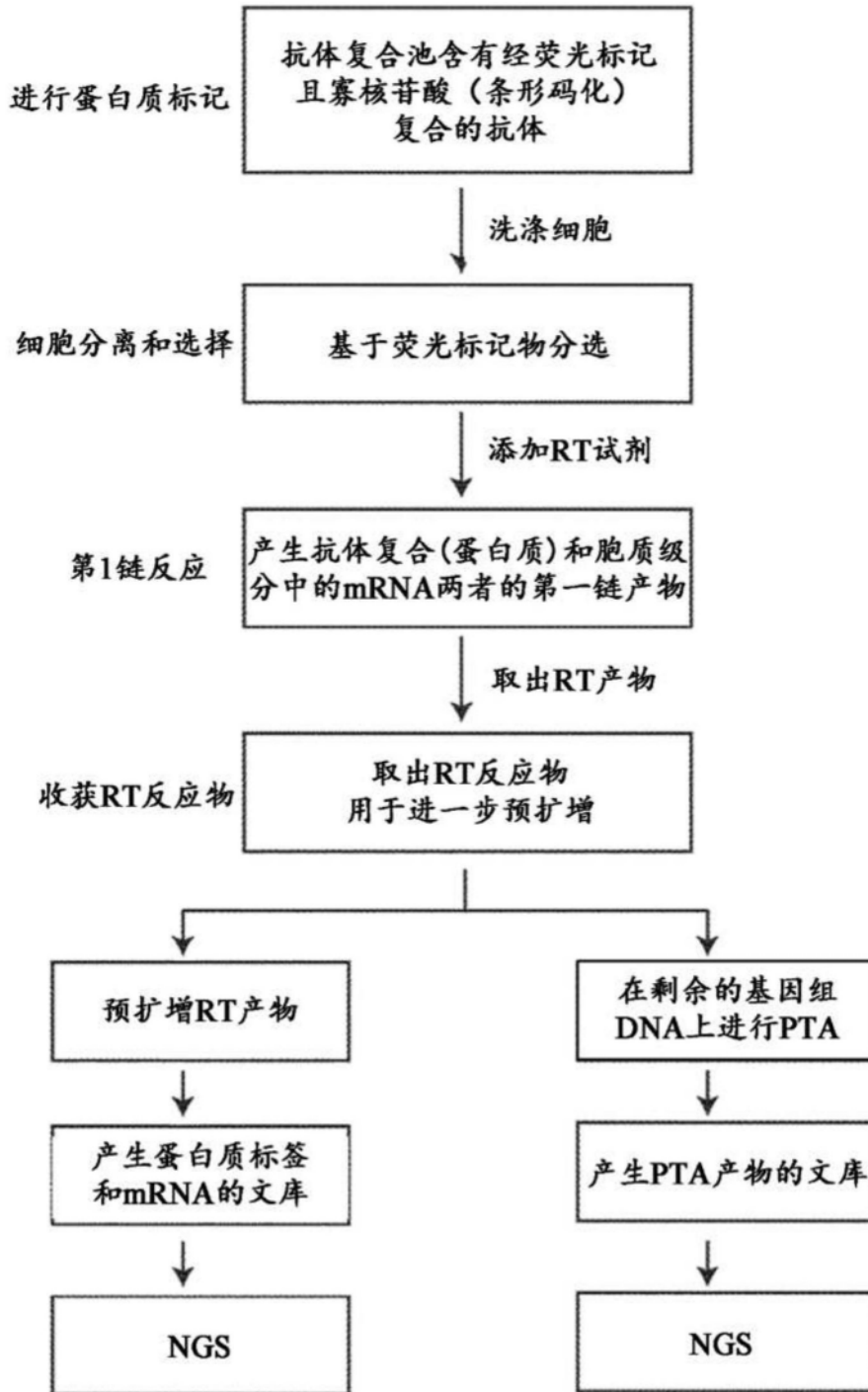


图1A

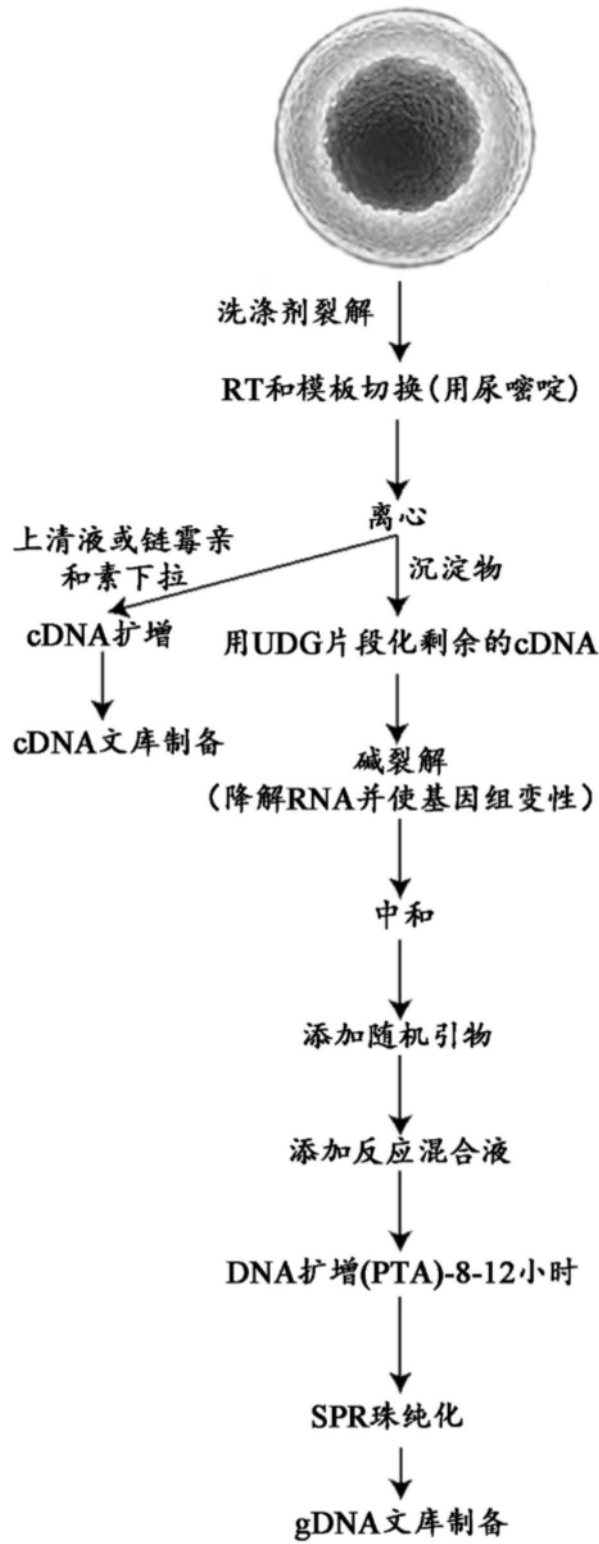


图1B

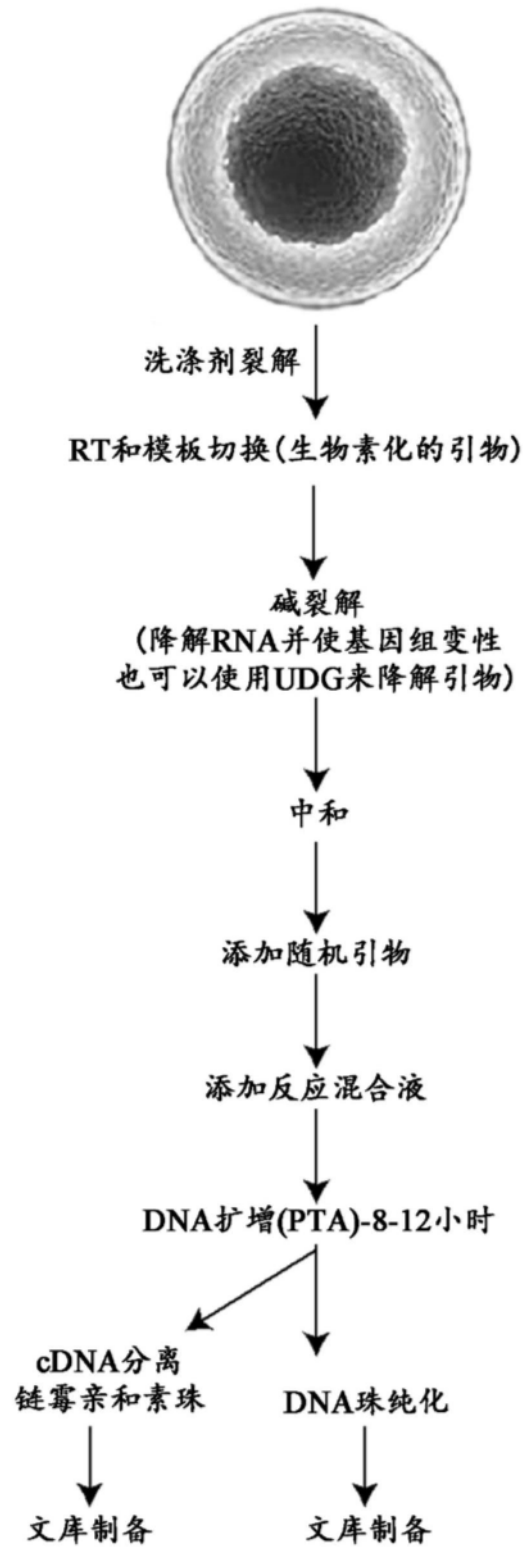


图1C

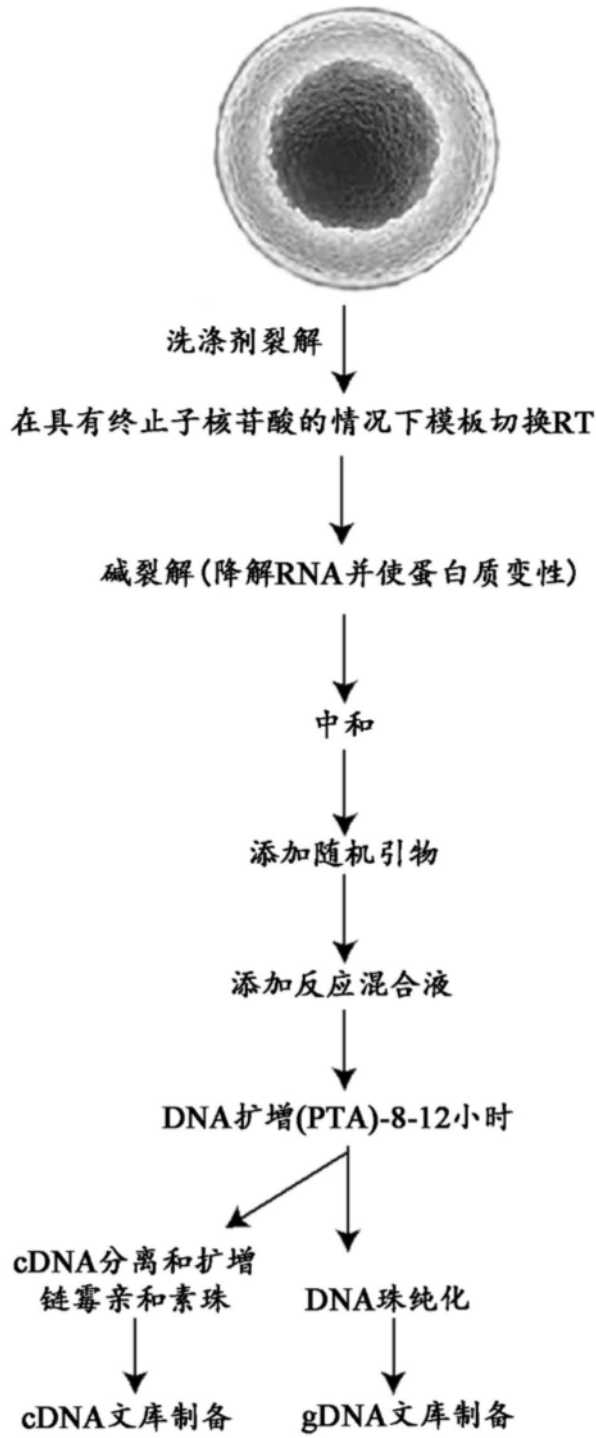


图1D

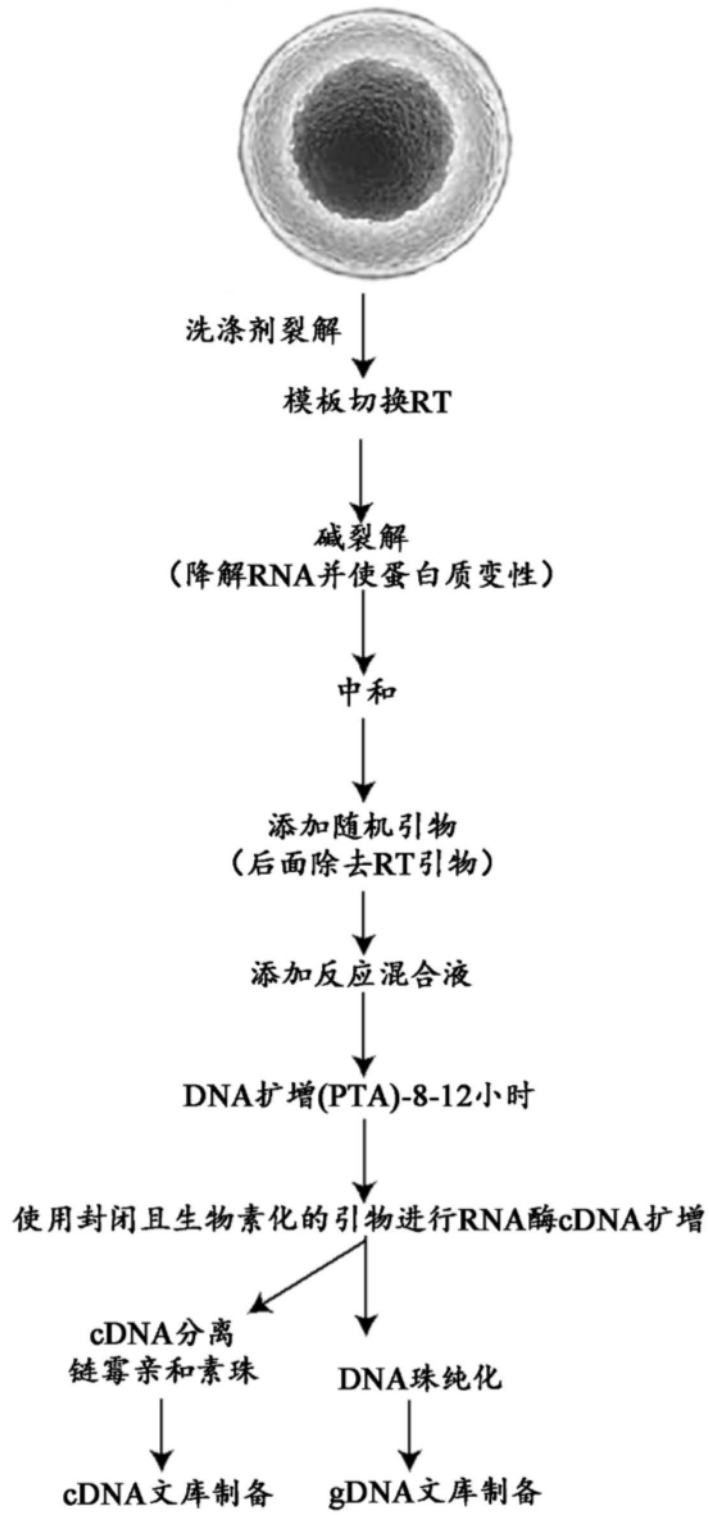


图1E

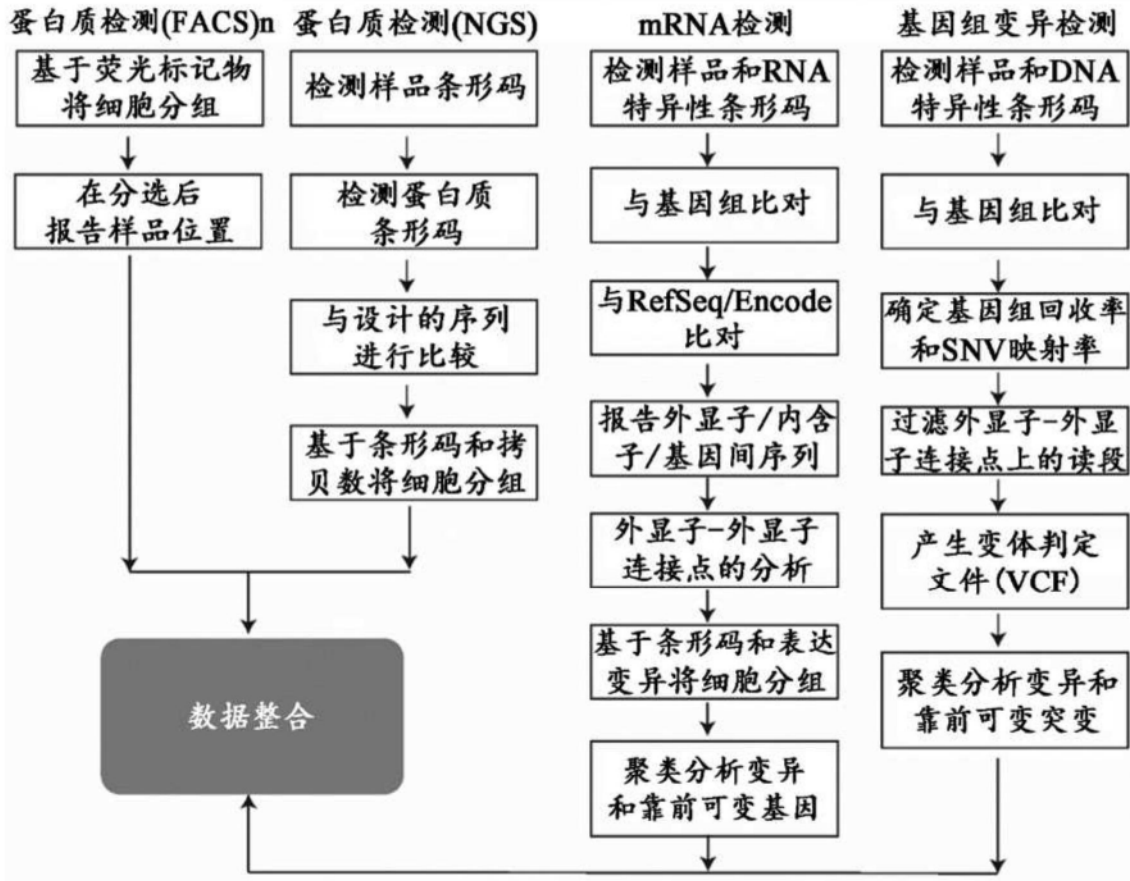


图1F

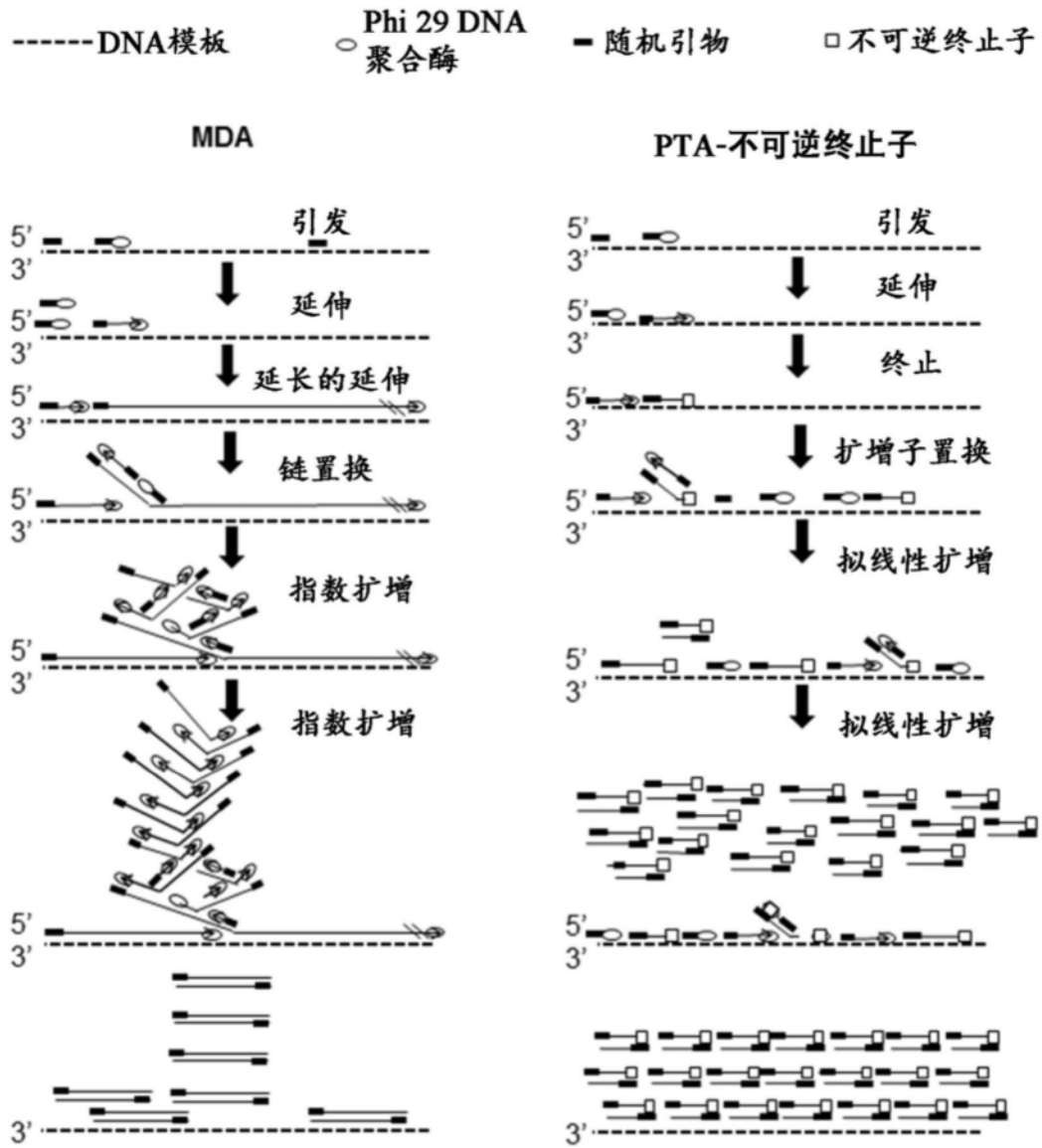


图1G

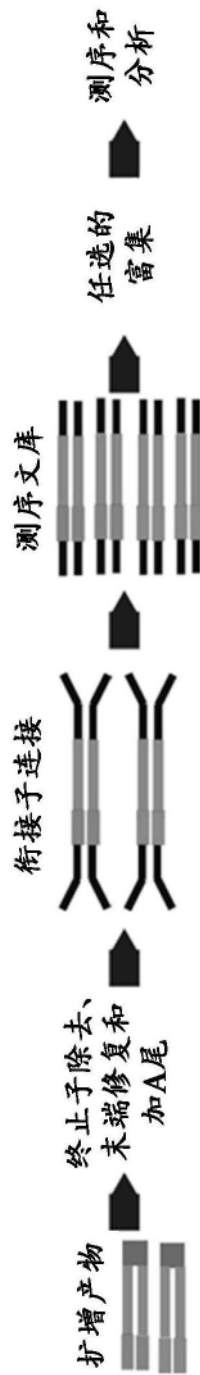


图2A

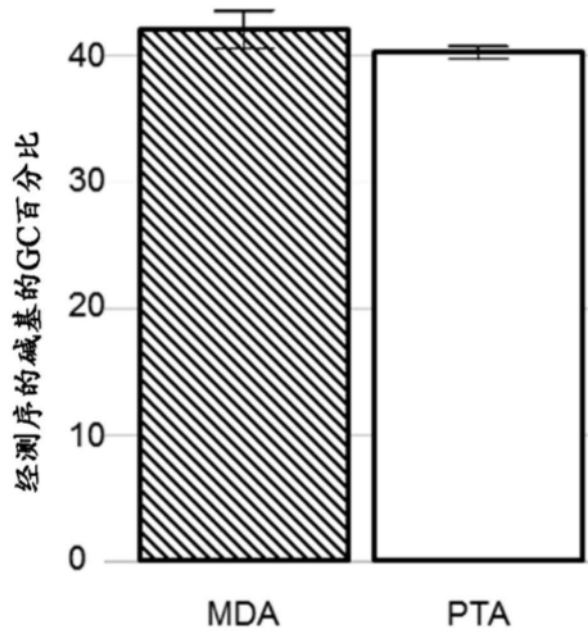


图2B

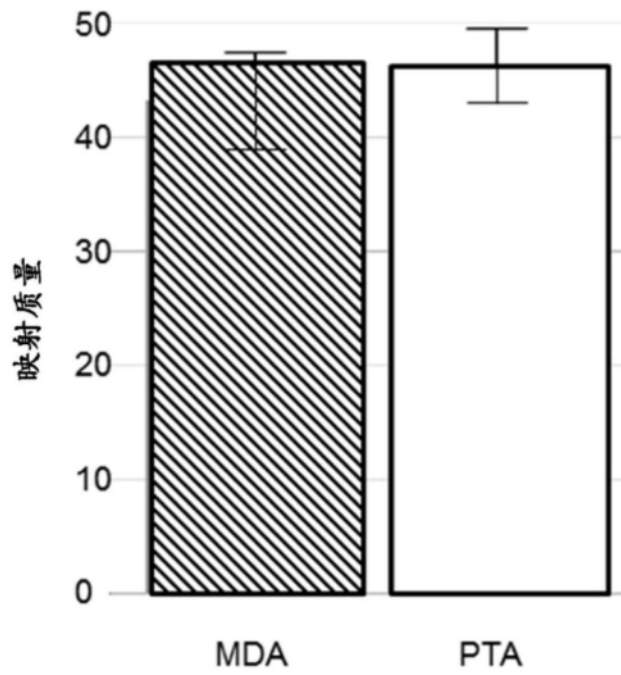


图2C

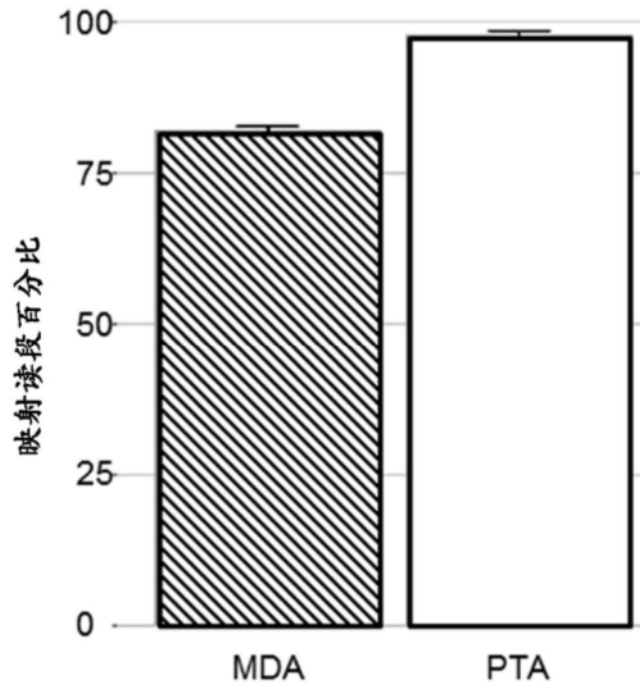


图2D

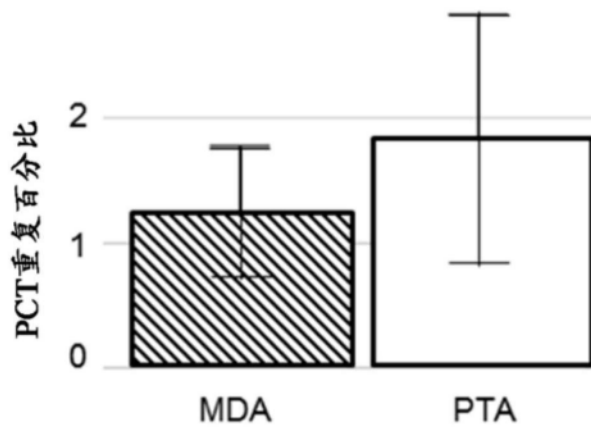


图2E

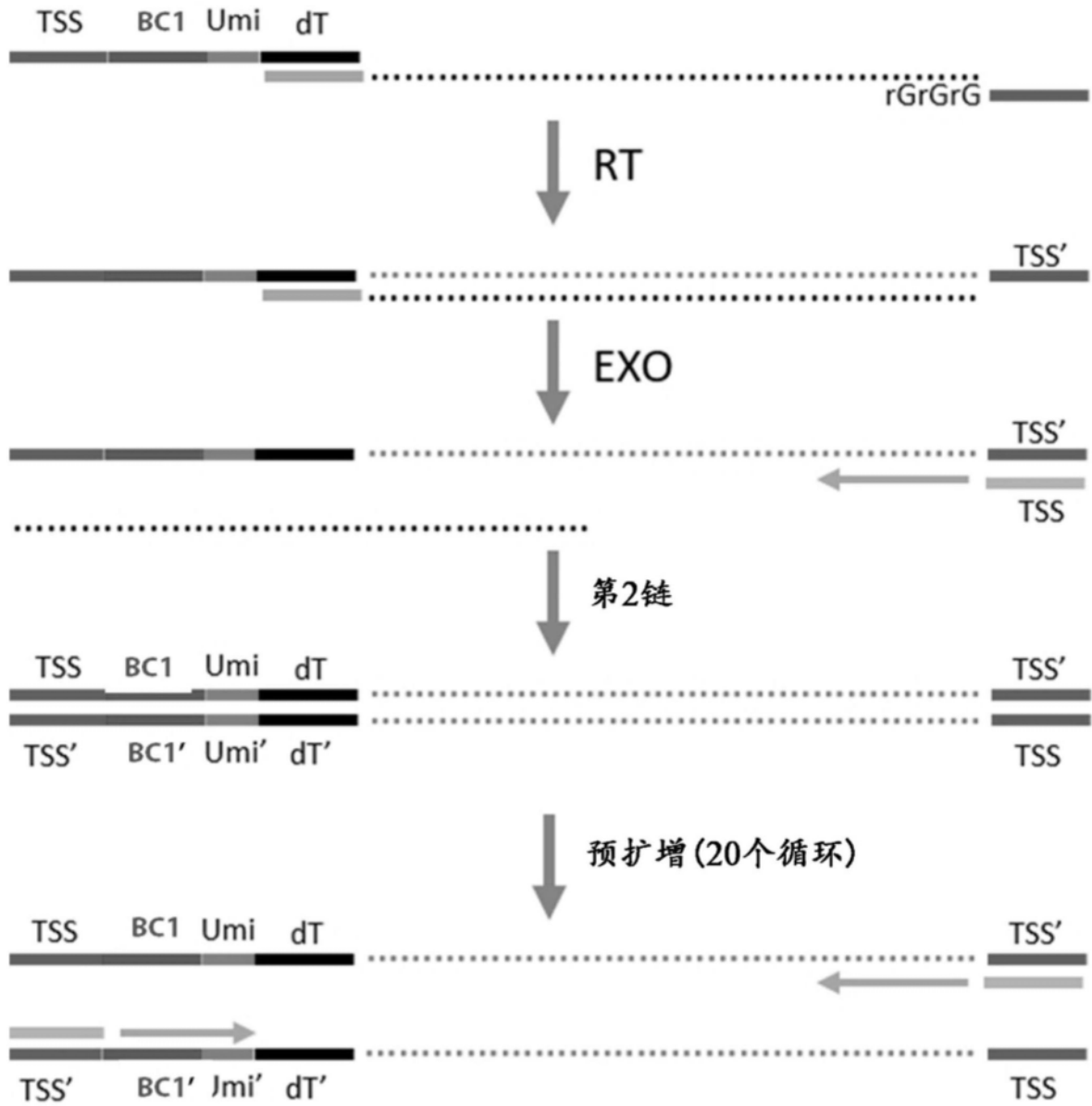


图2F

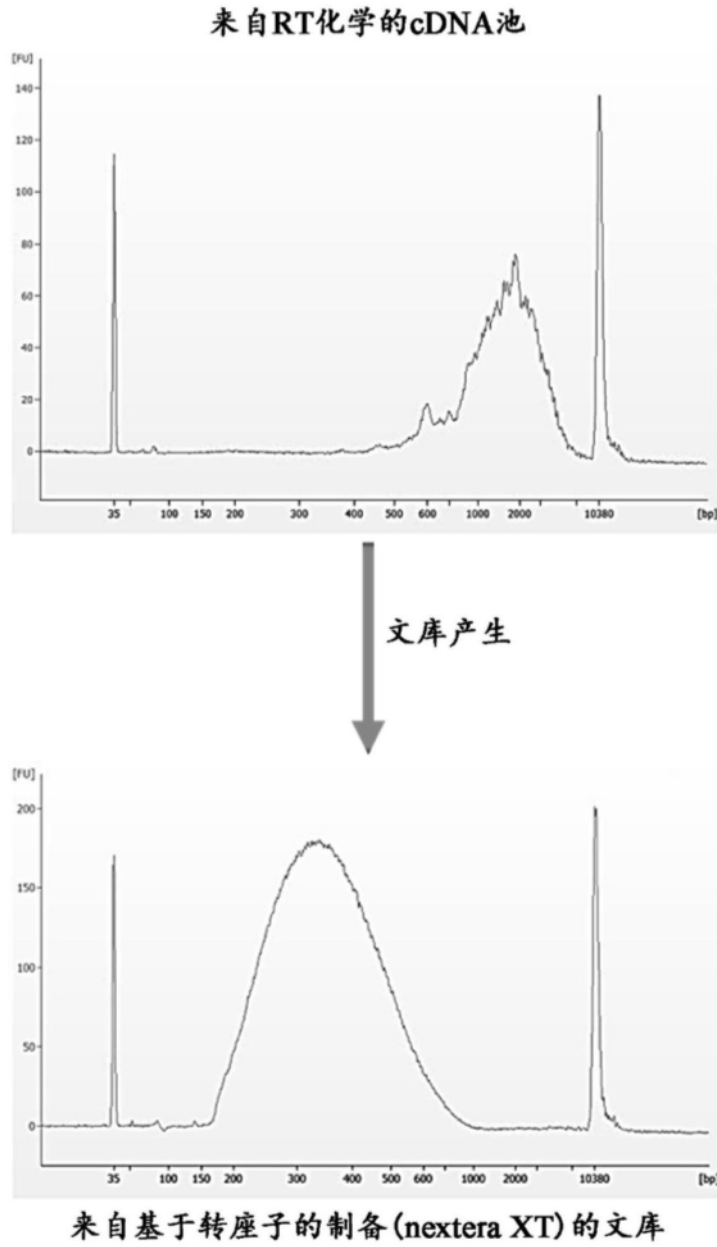


图2G

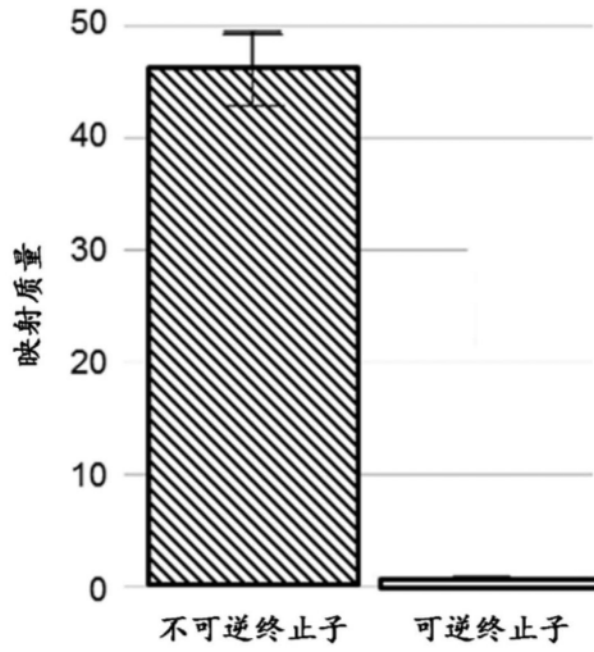


图3A

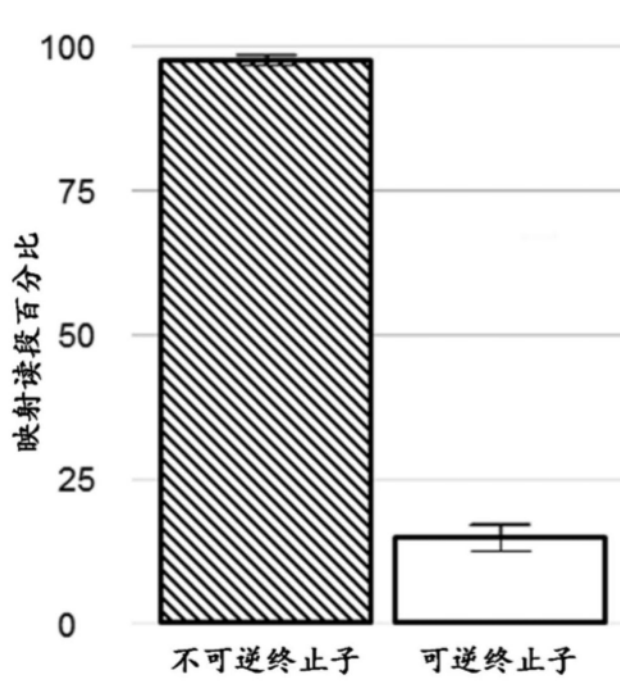


图3B

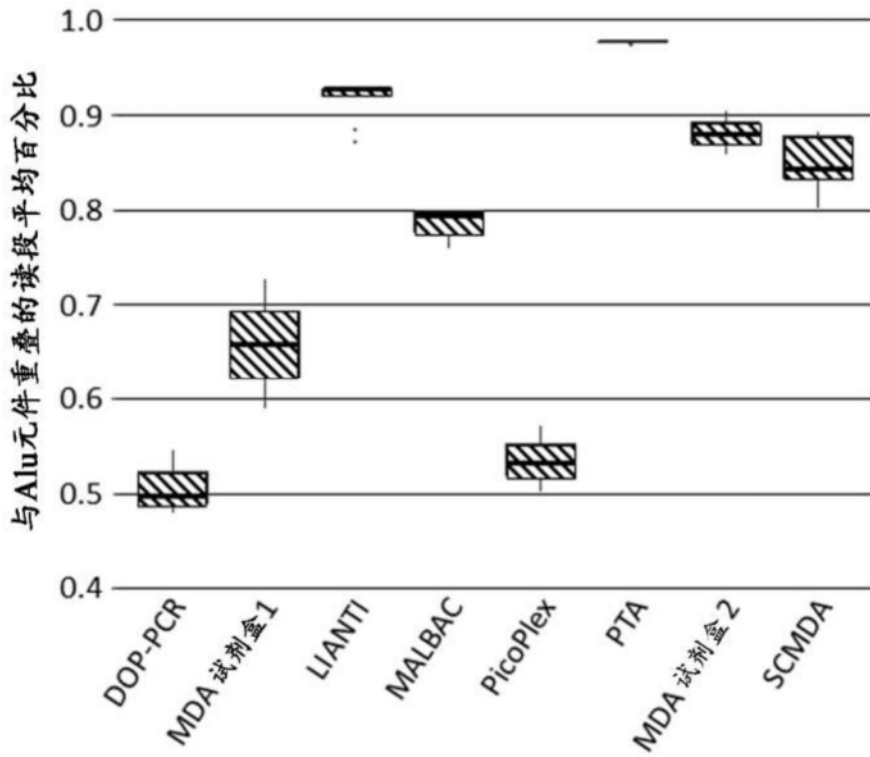


图3C

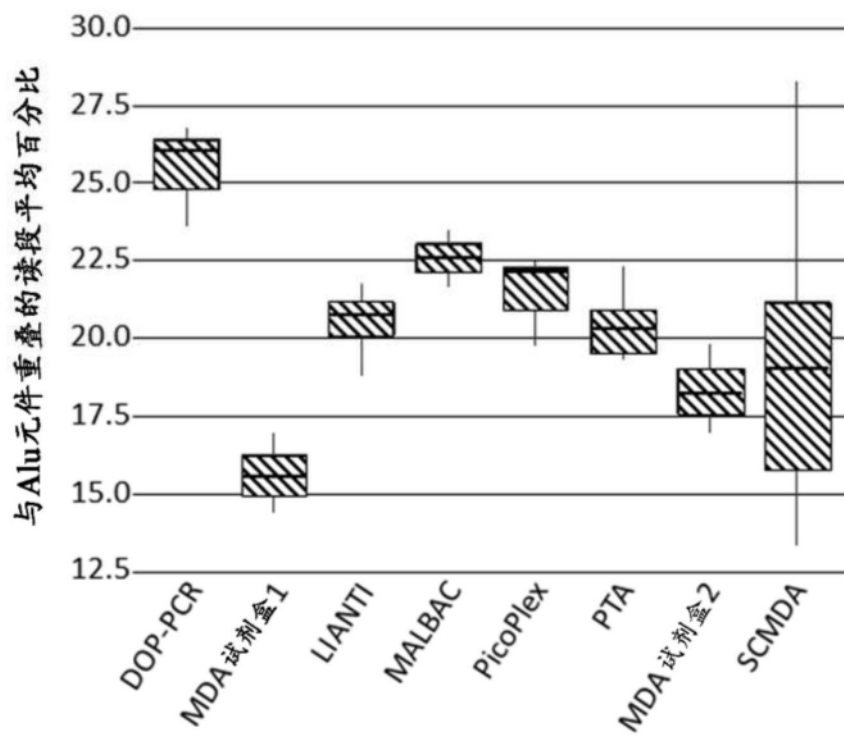


图3D

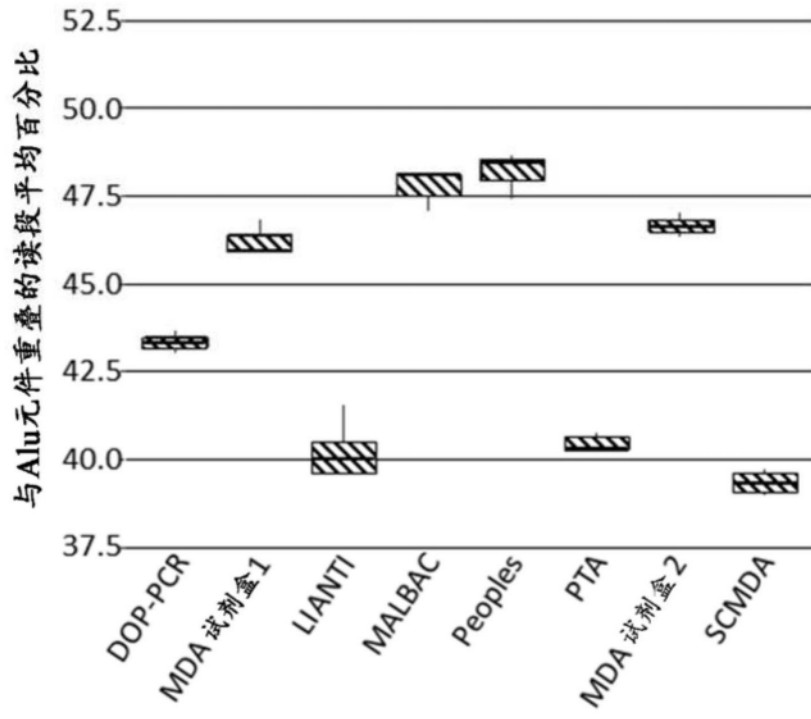


图3E

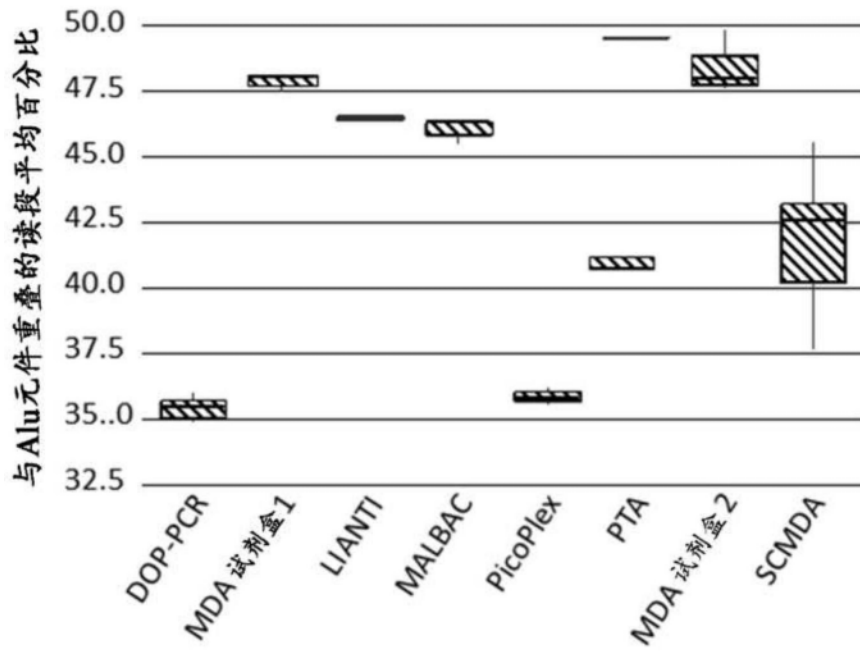


图3F

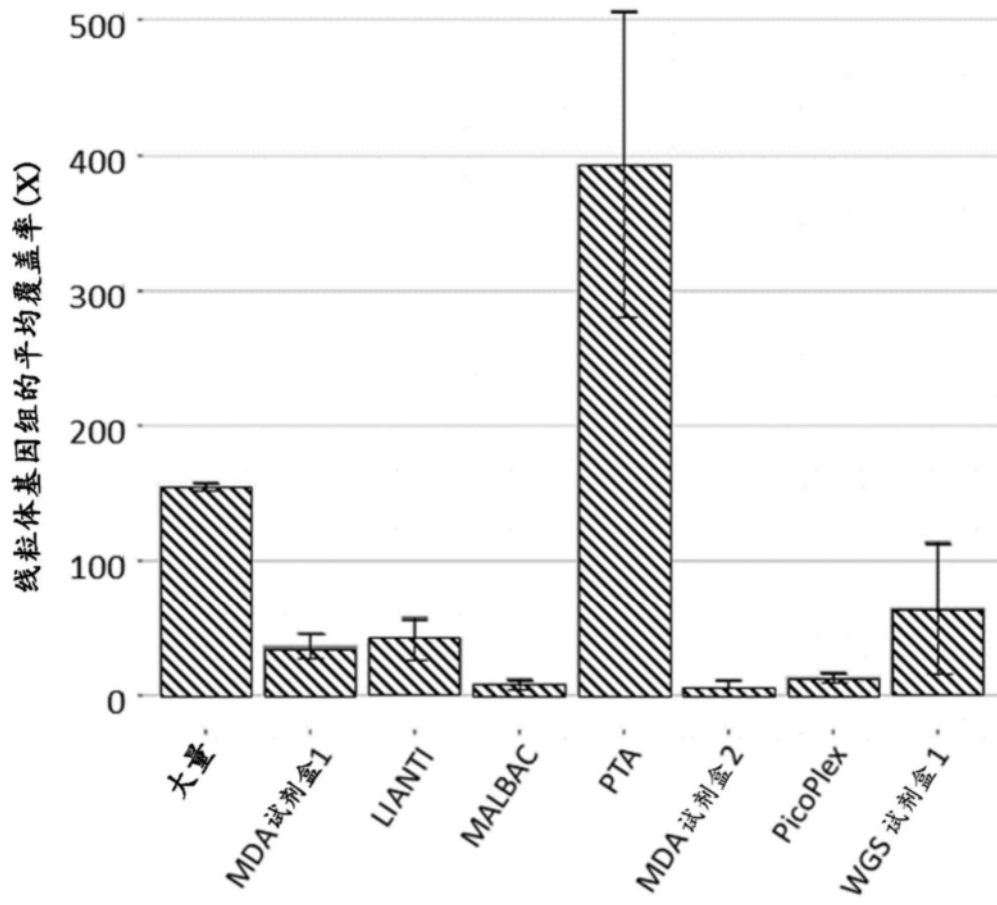


图3G

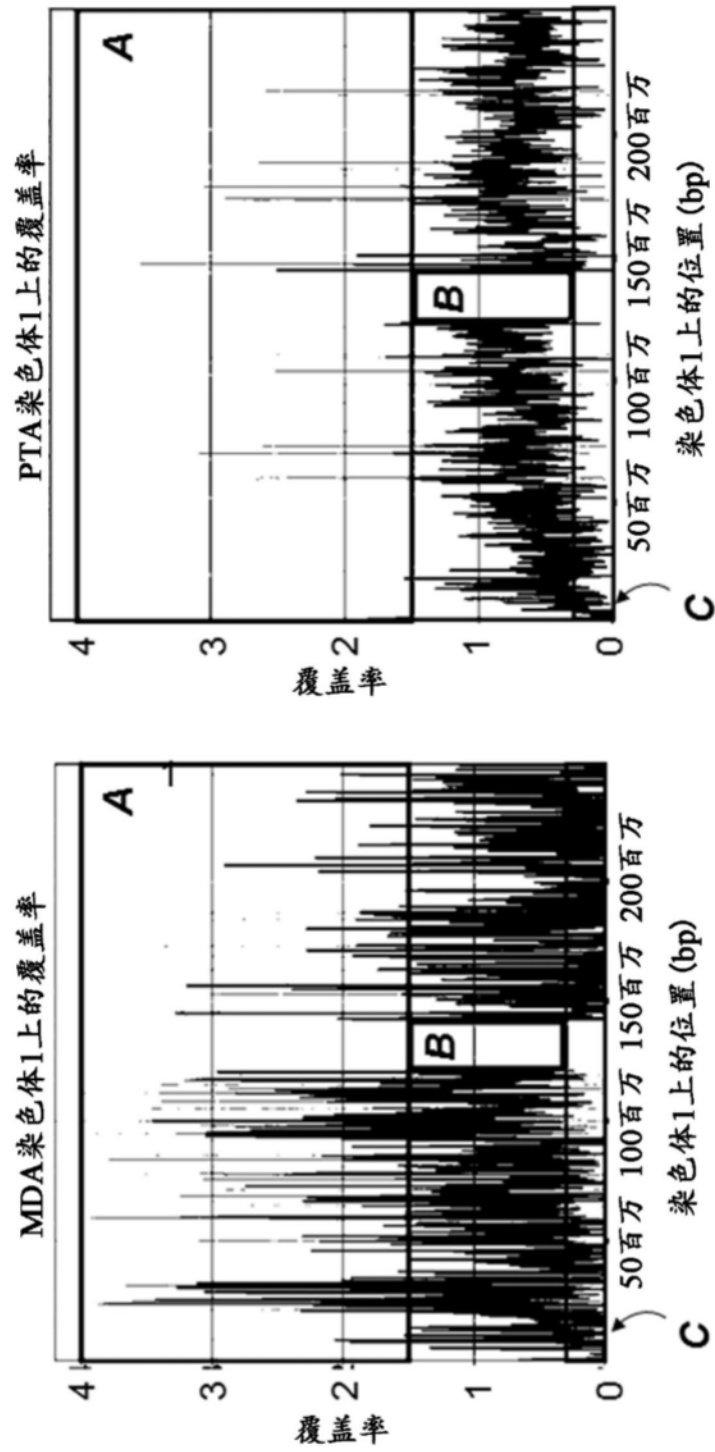


图4A

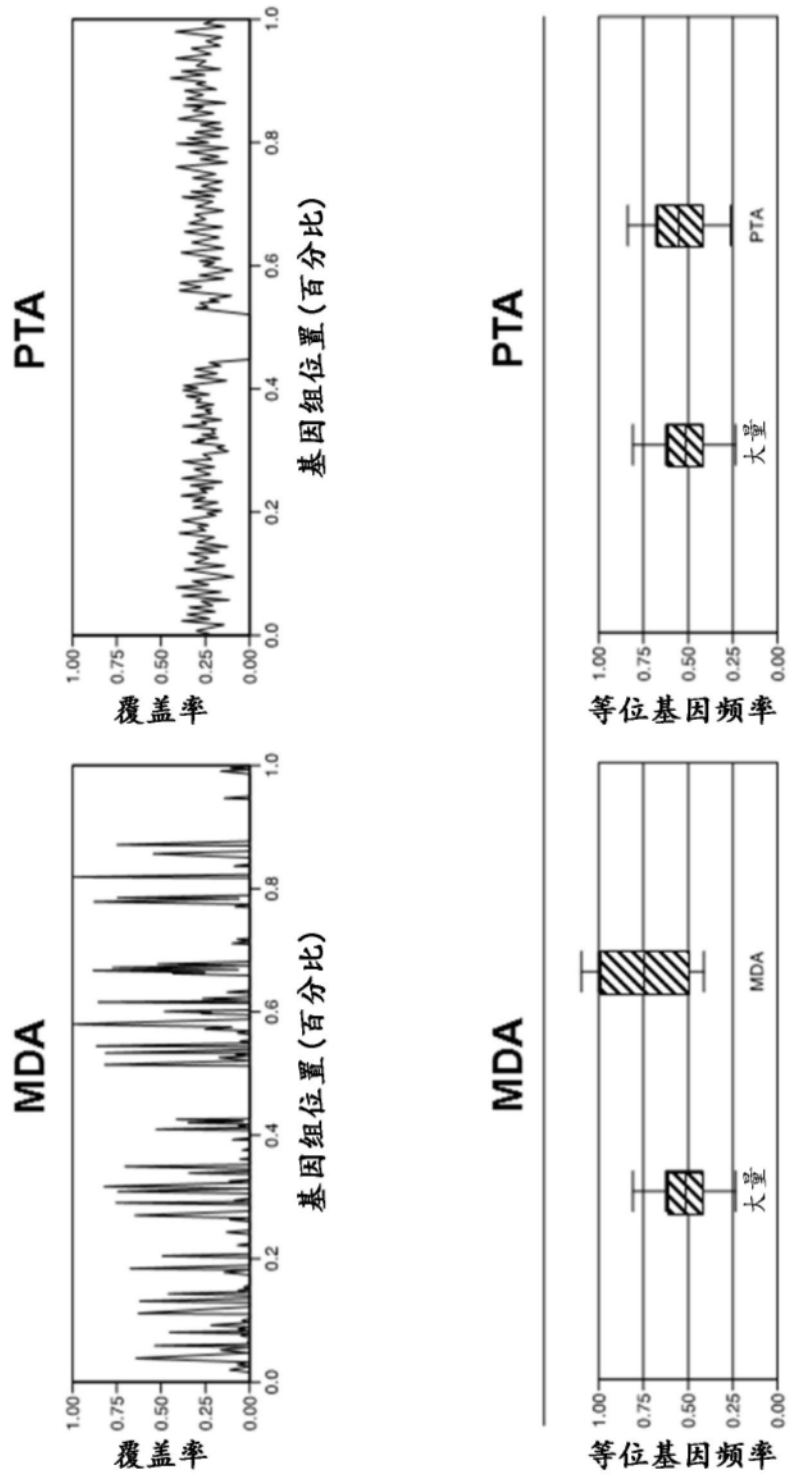


图4B

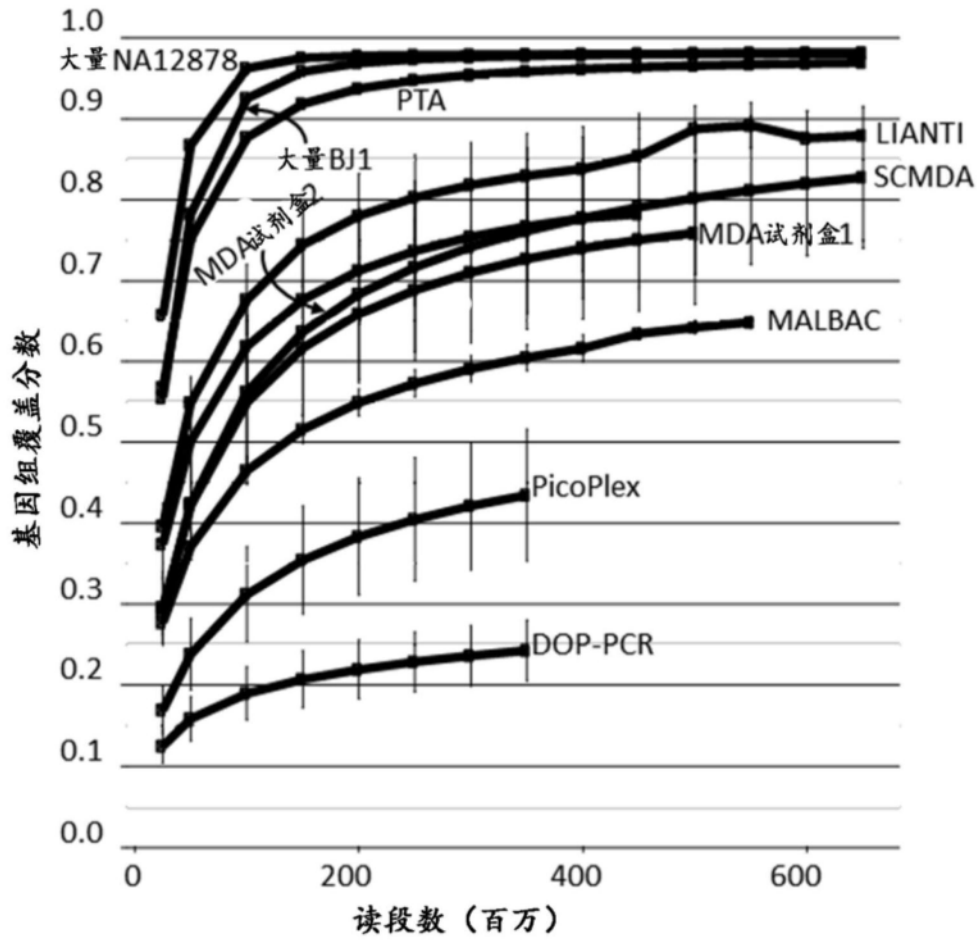


图5A

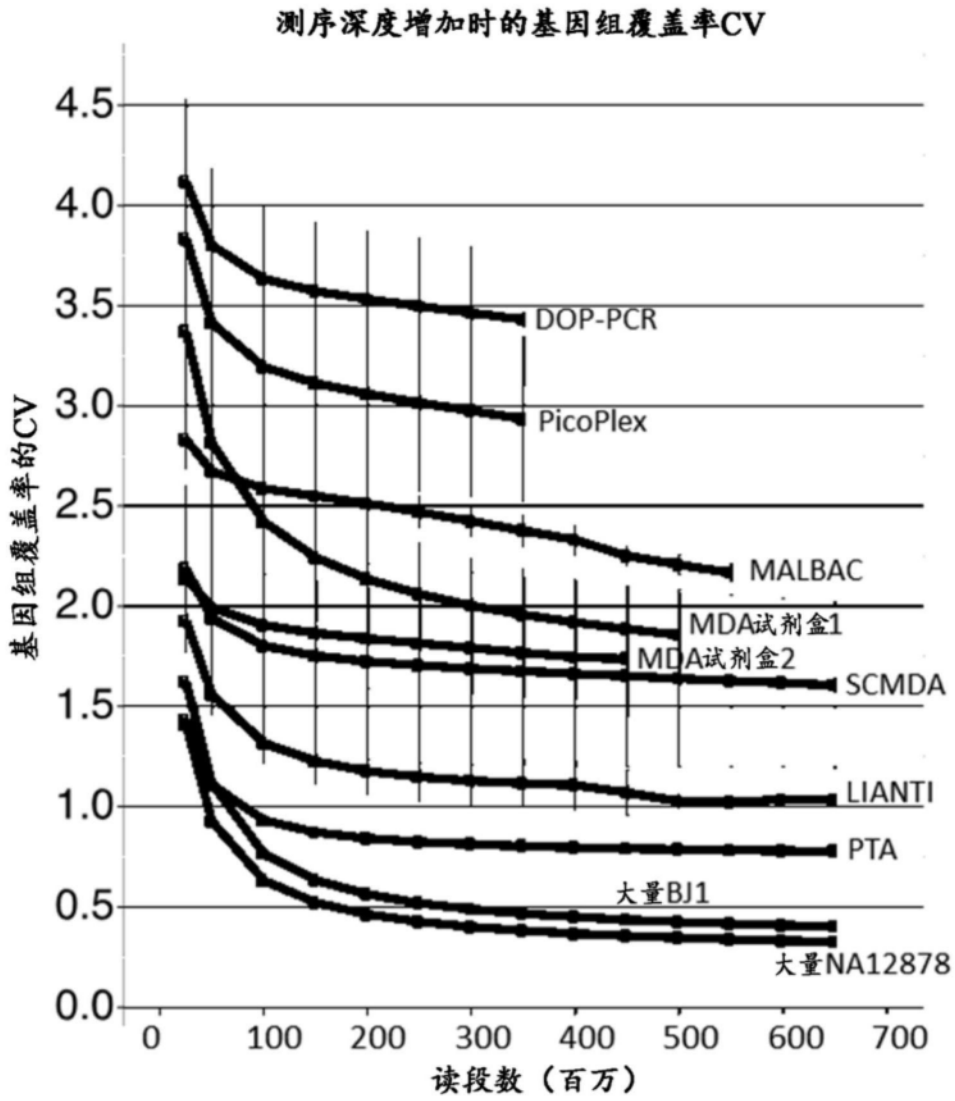


图5B

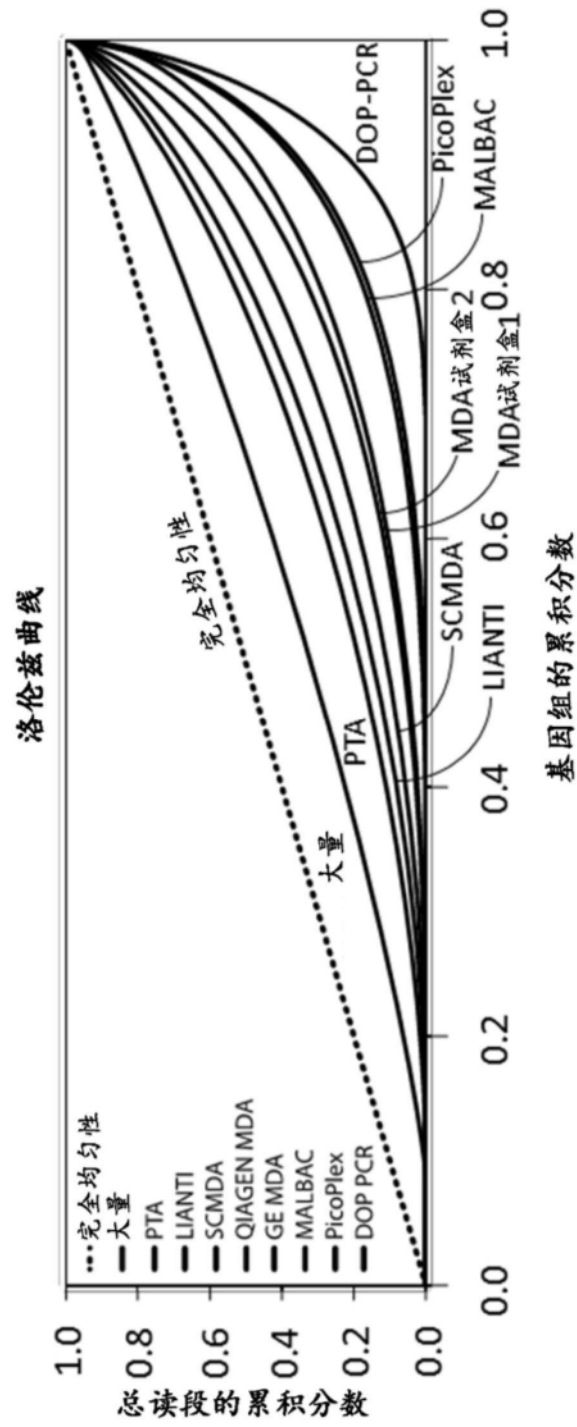


图5C

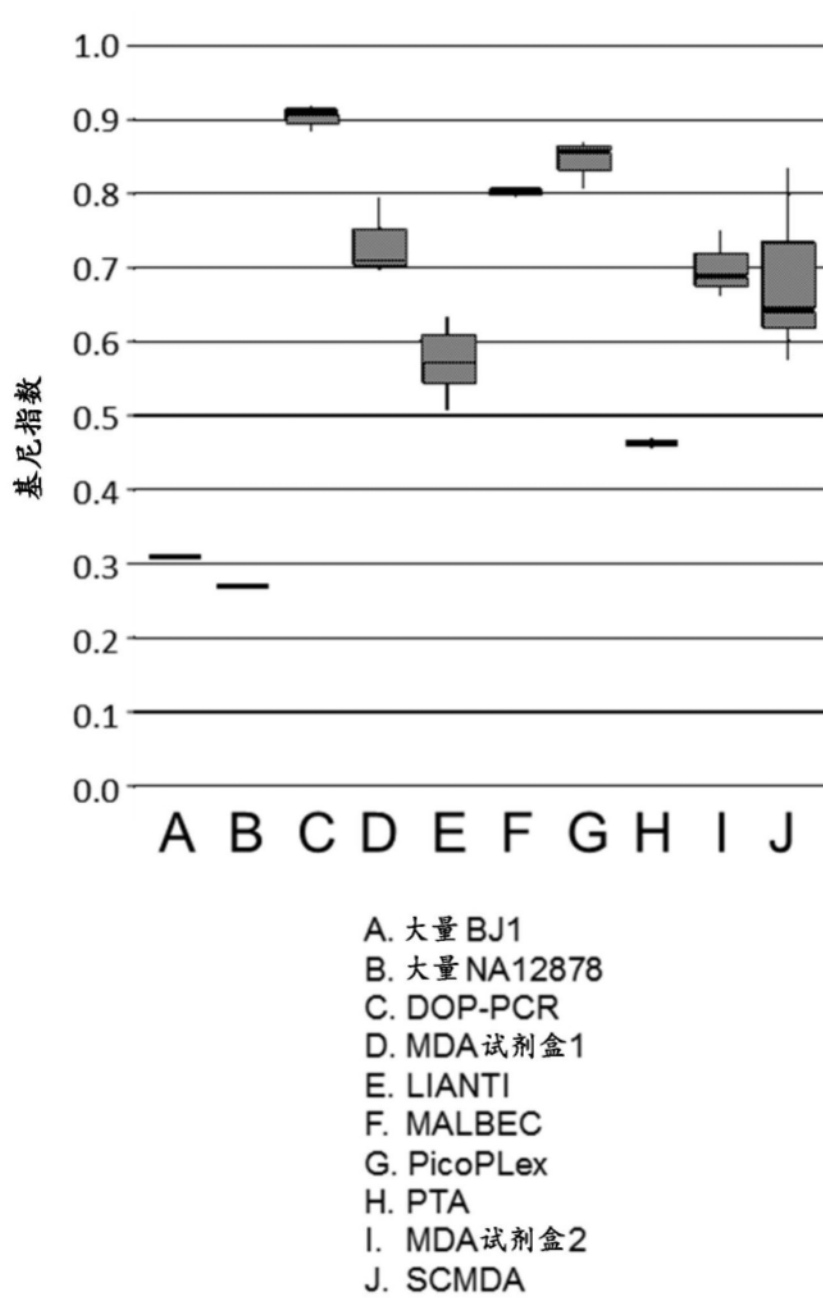


图5D

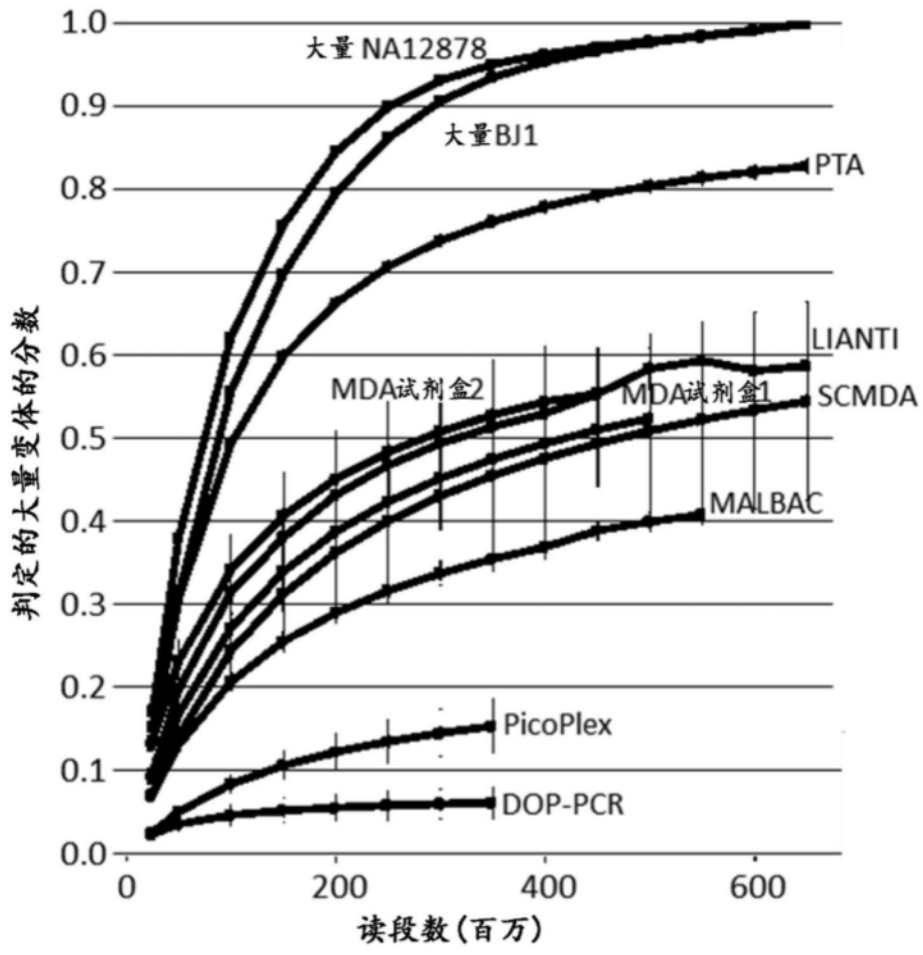


图5E

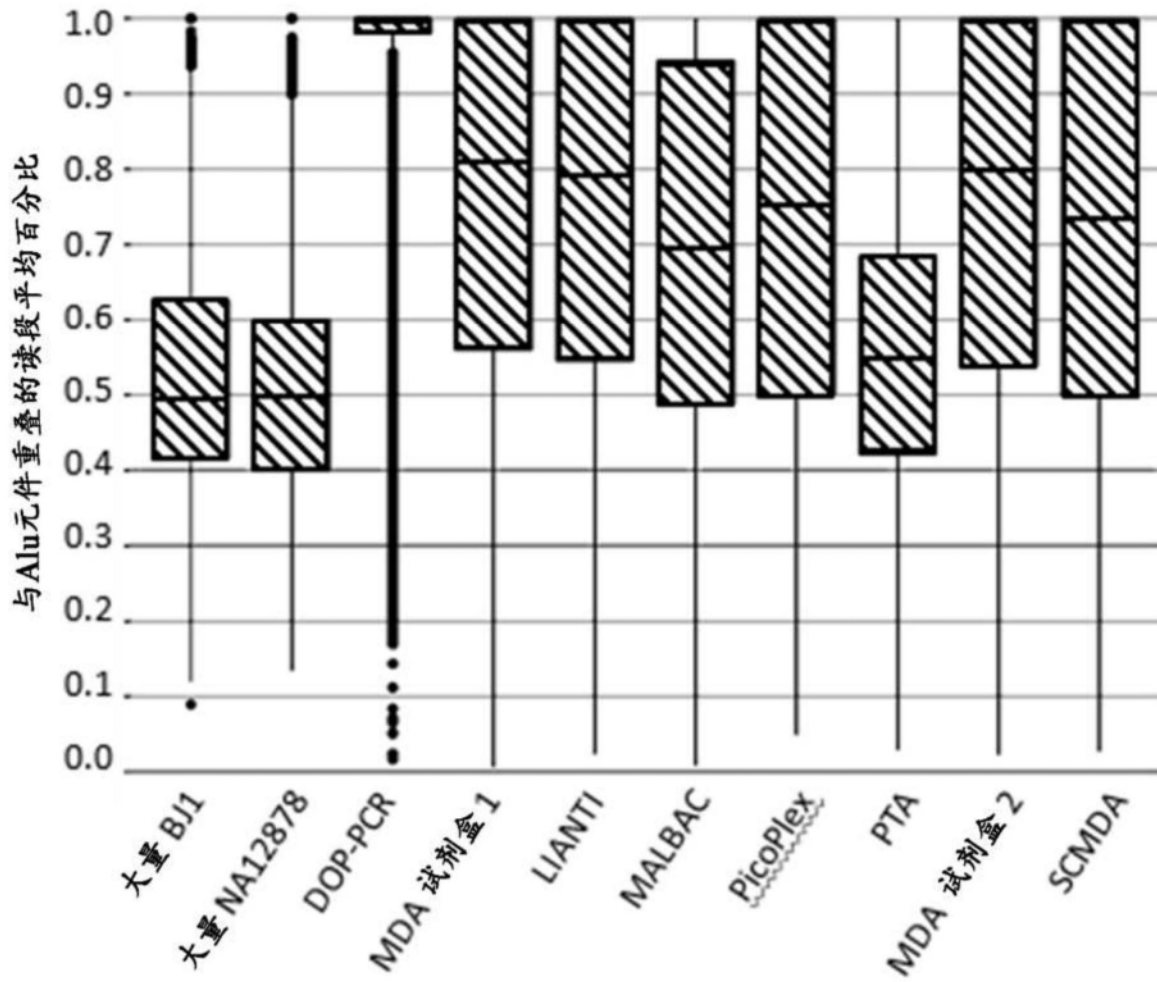


图5F

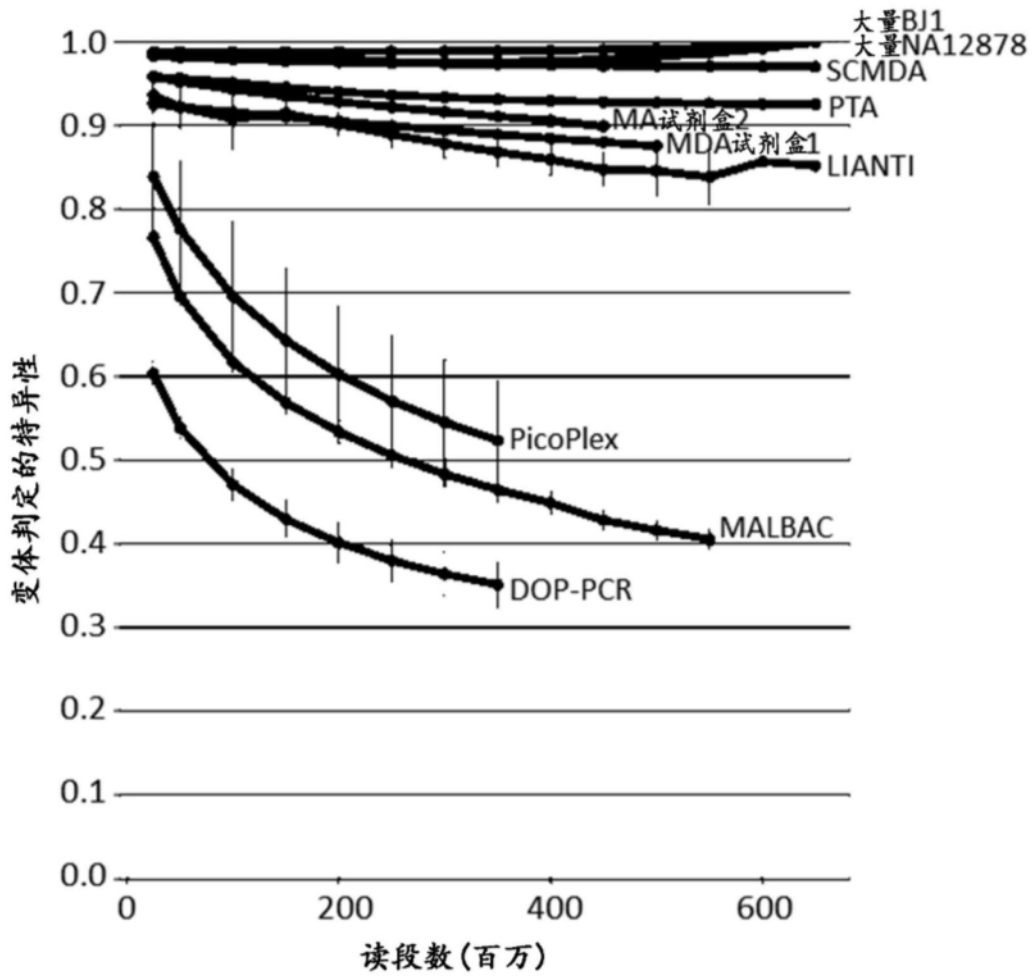


图5G

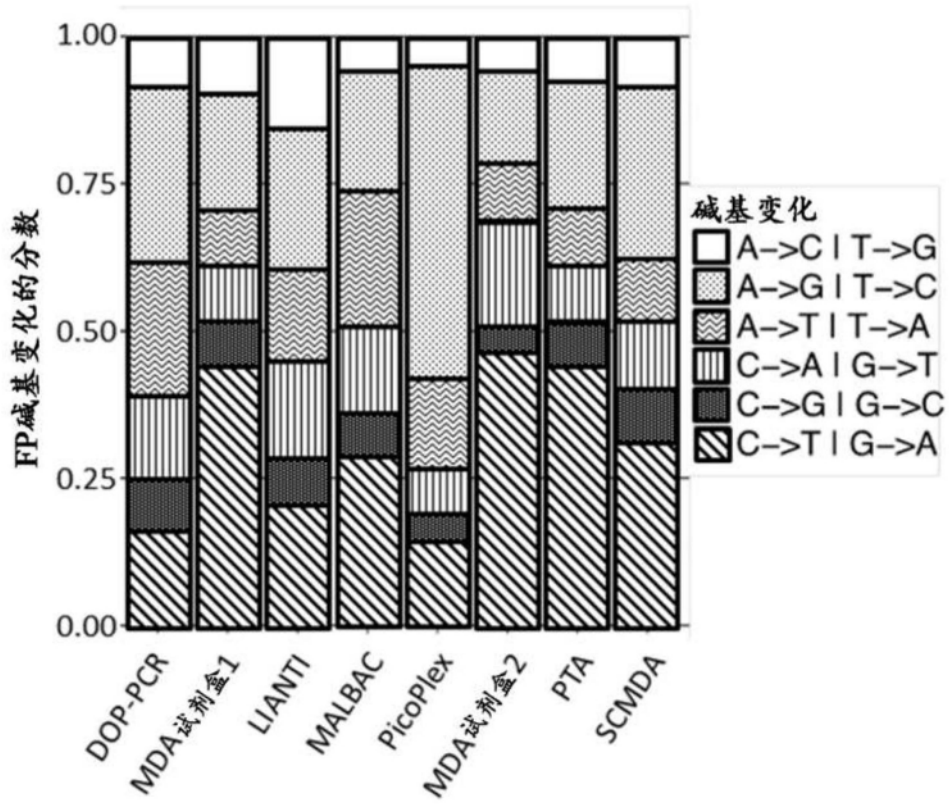


图5H

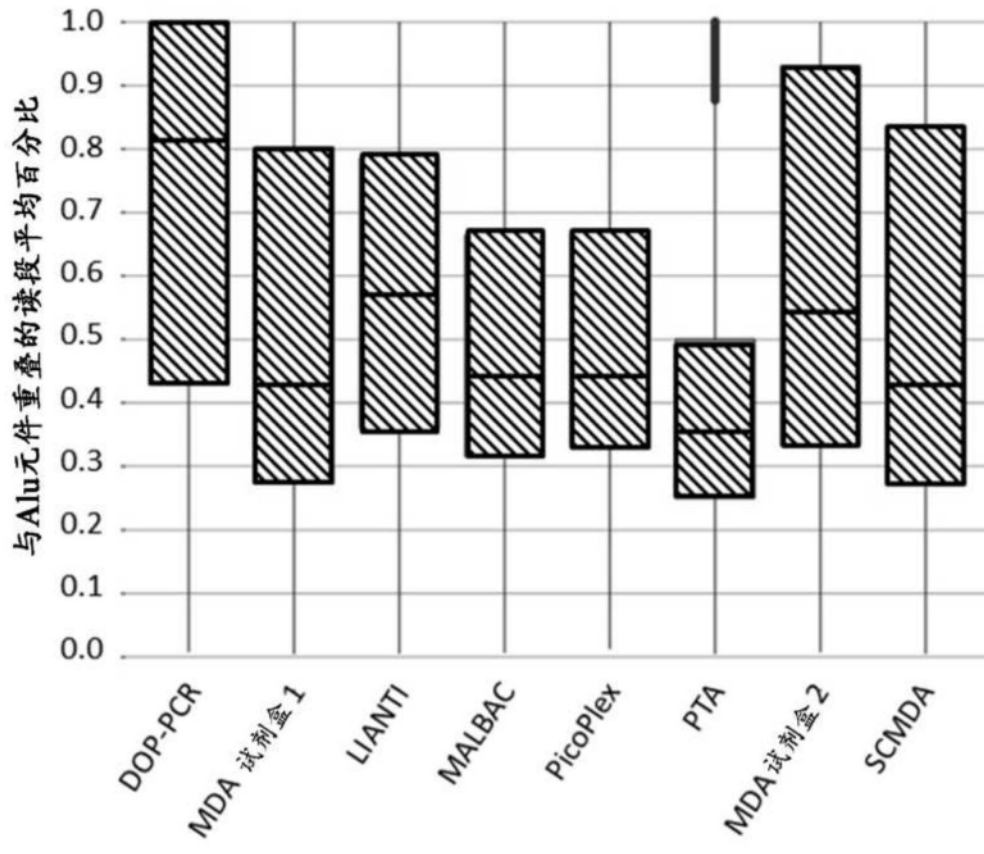


图5I

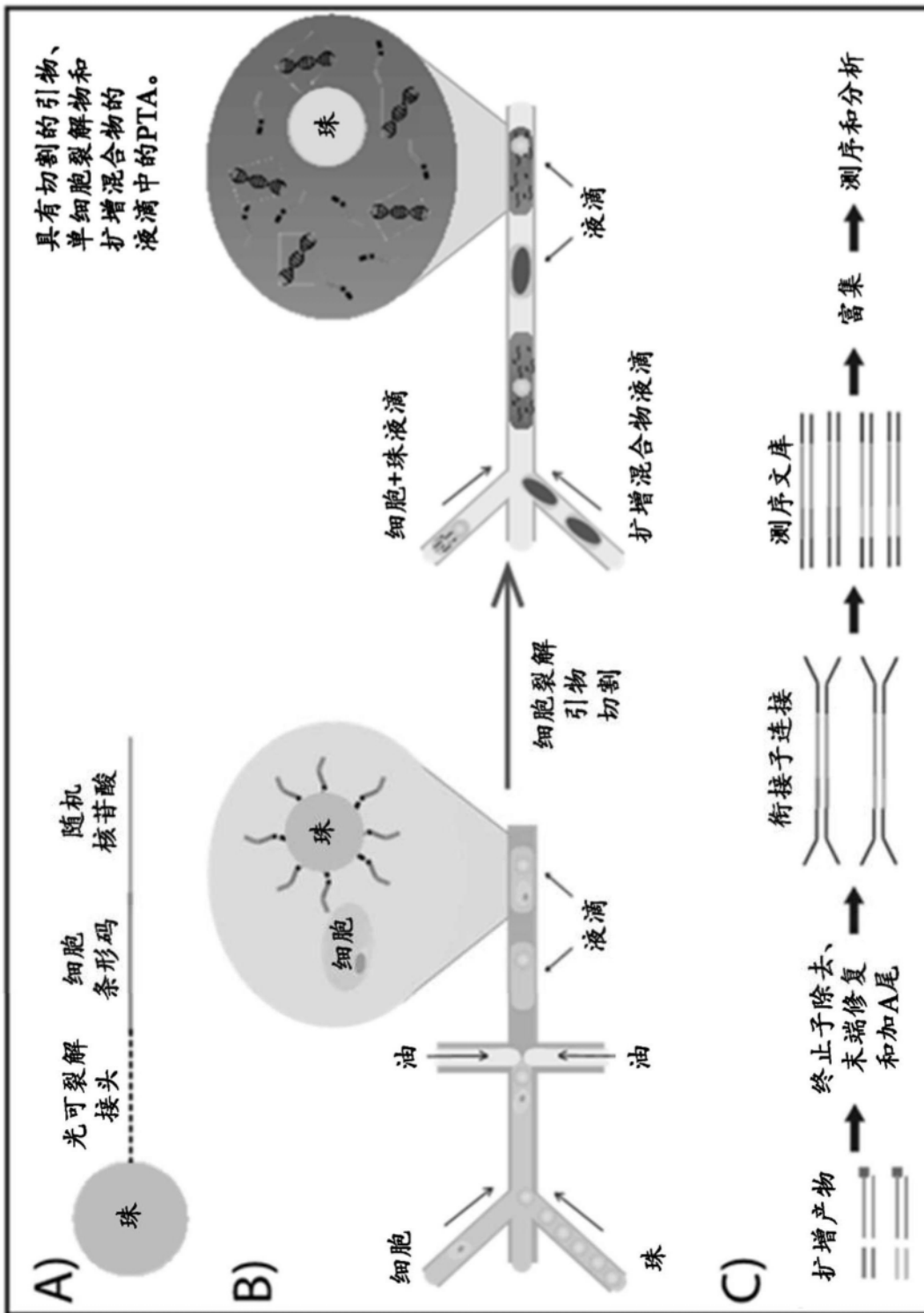


图6

用于多组学分析单细胞的方法-使用PTA

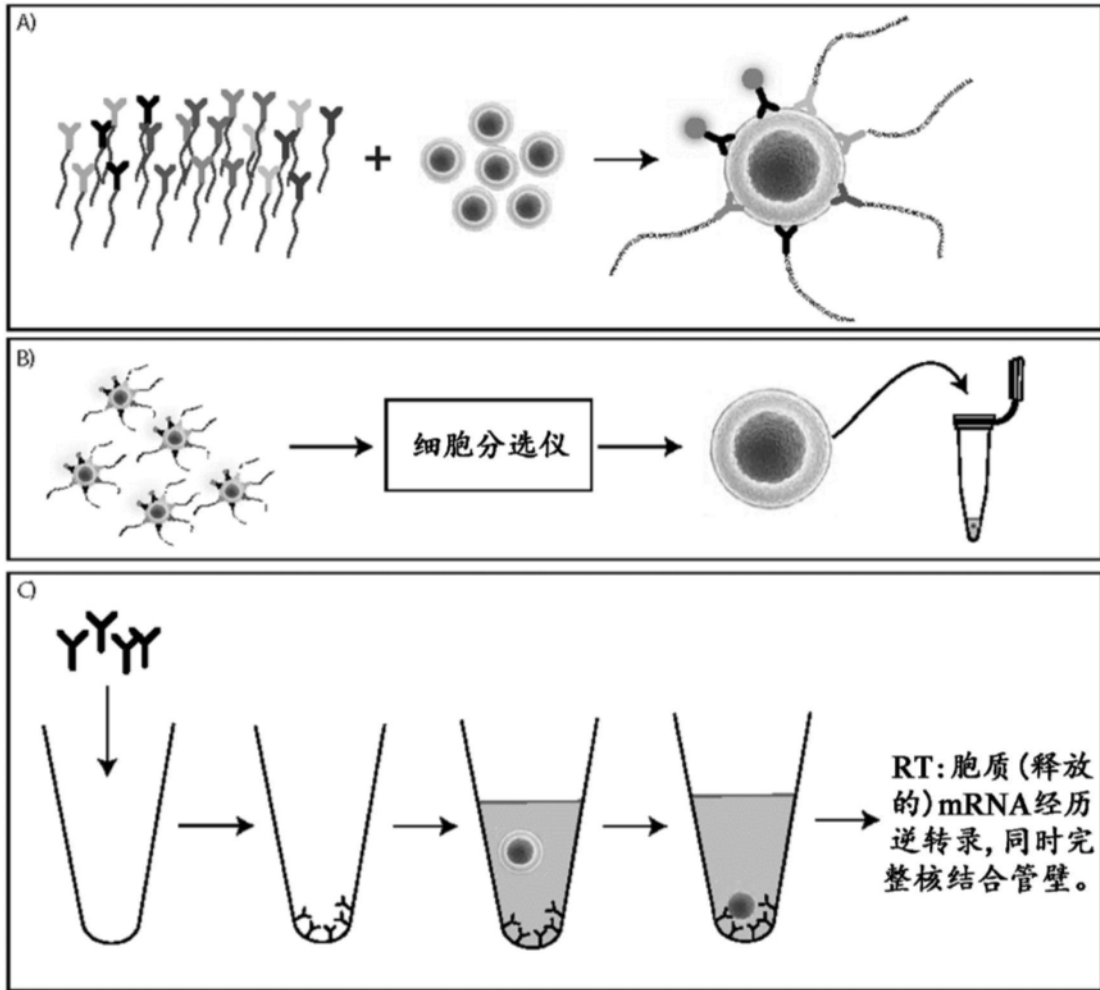


图7A

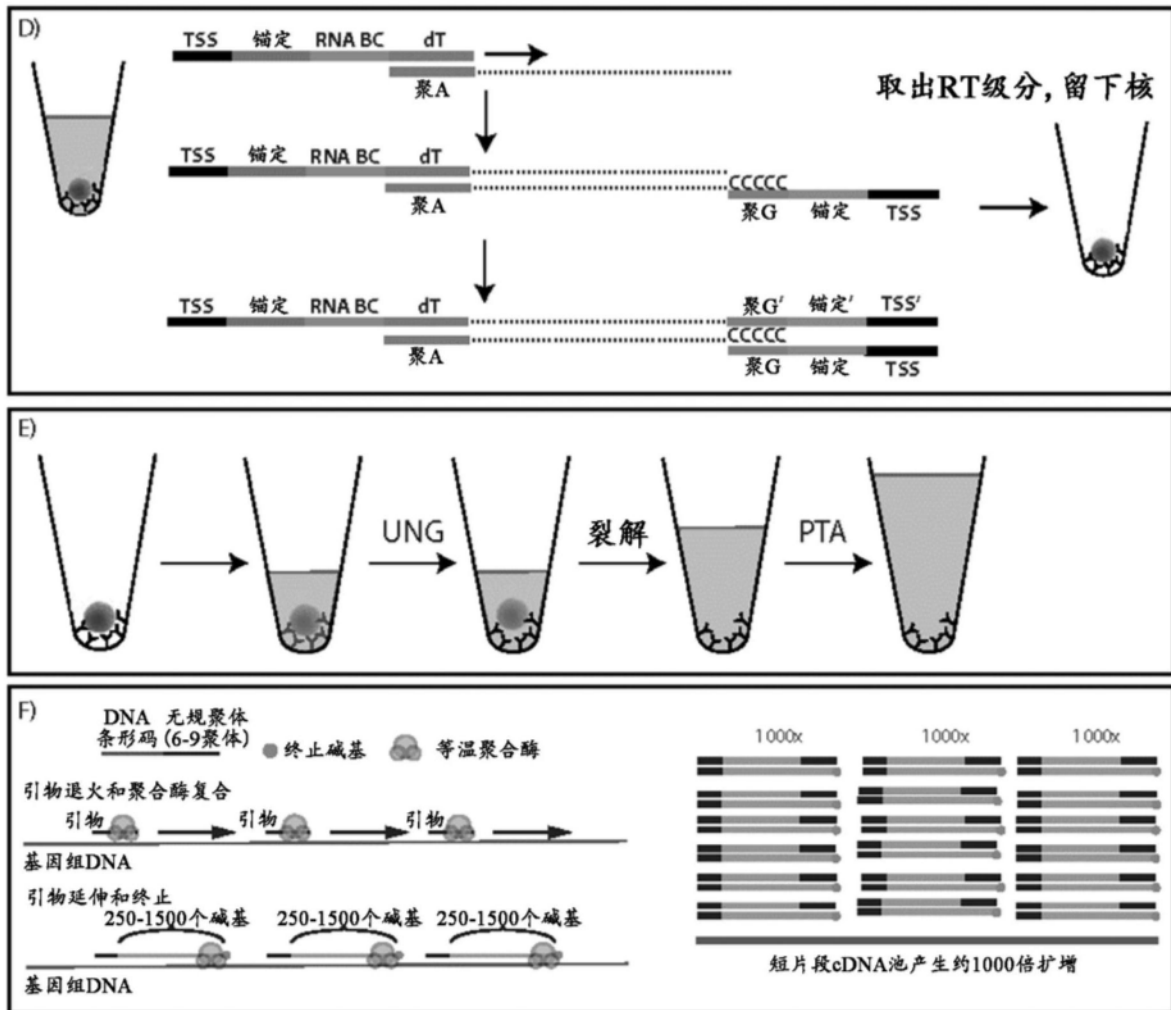


图7B

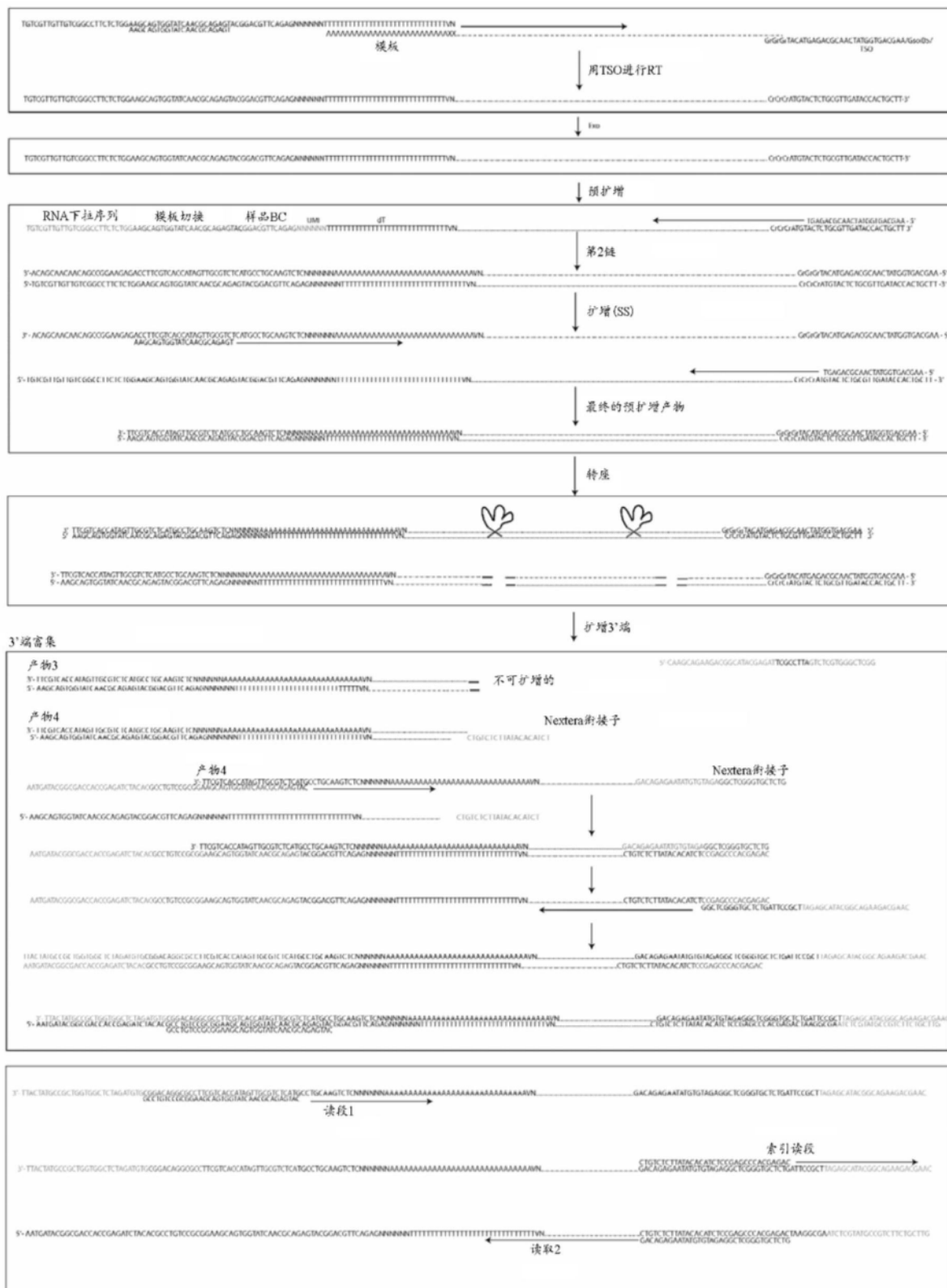


图8B

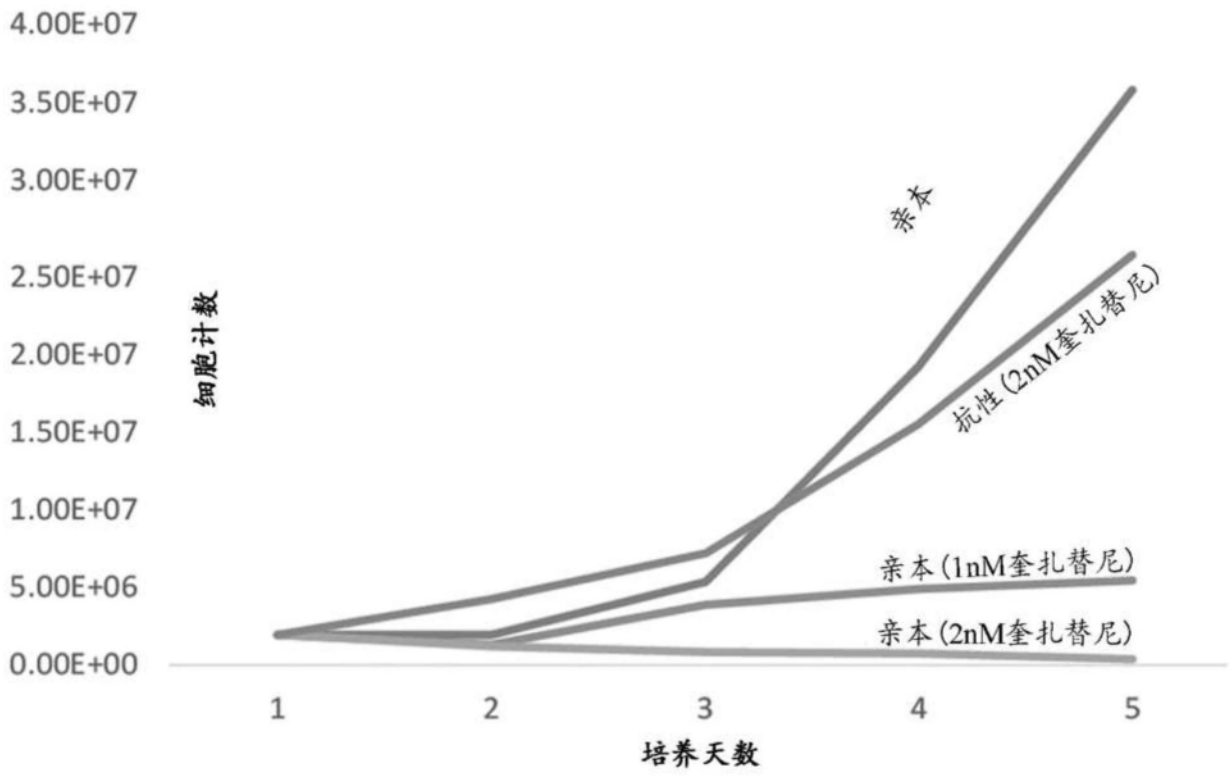


图9A

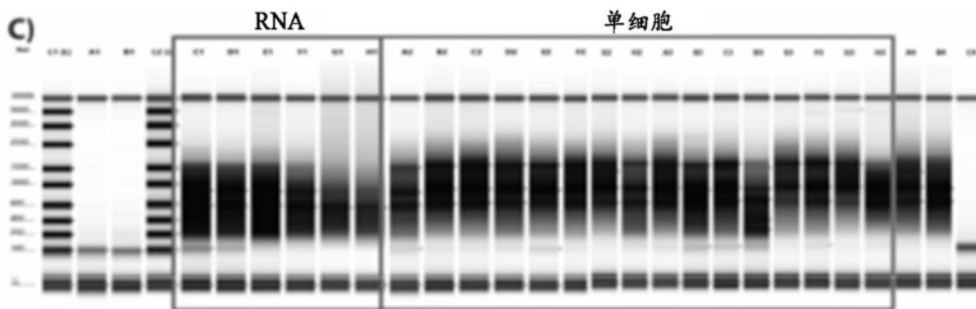


图9B

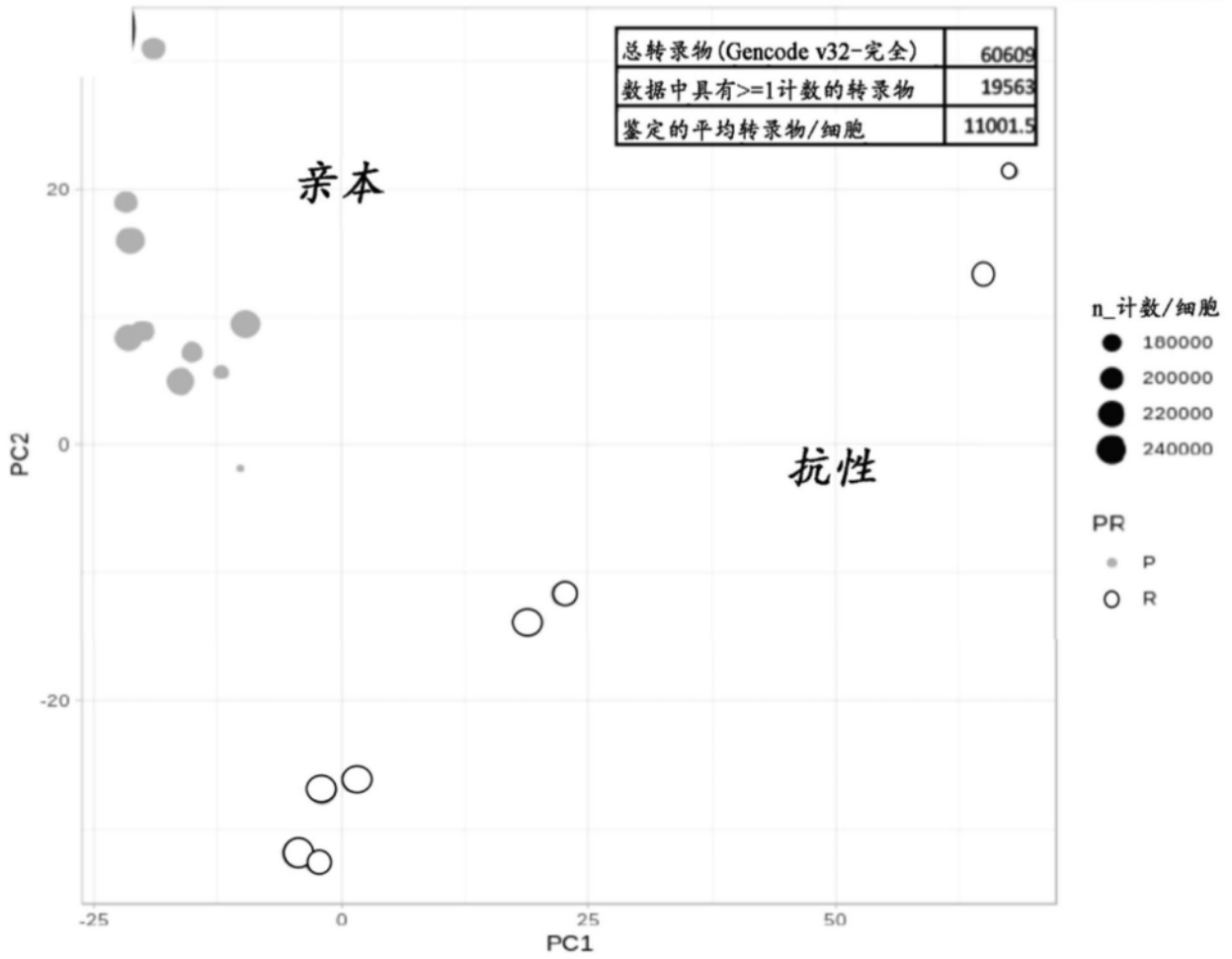


图9C

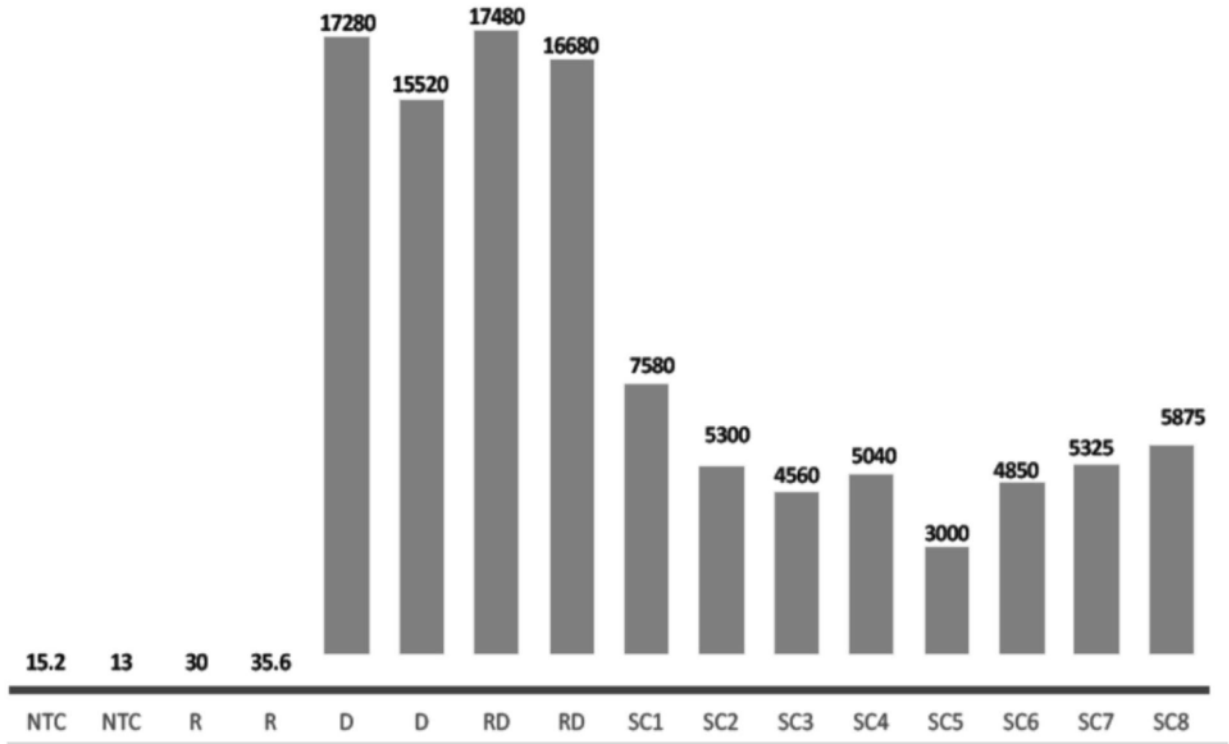


图9D

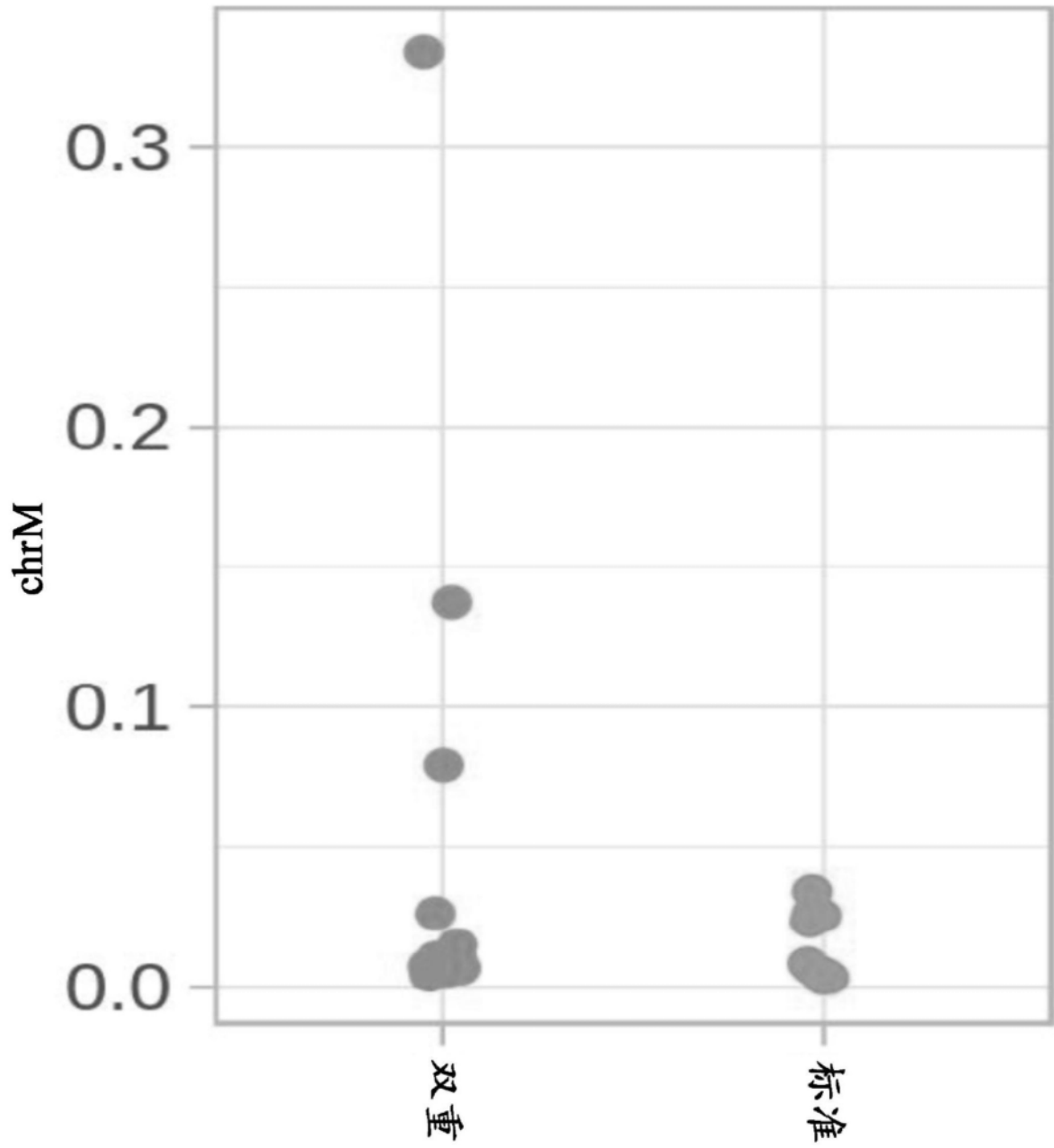


图10A

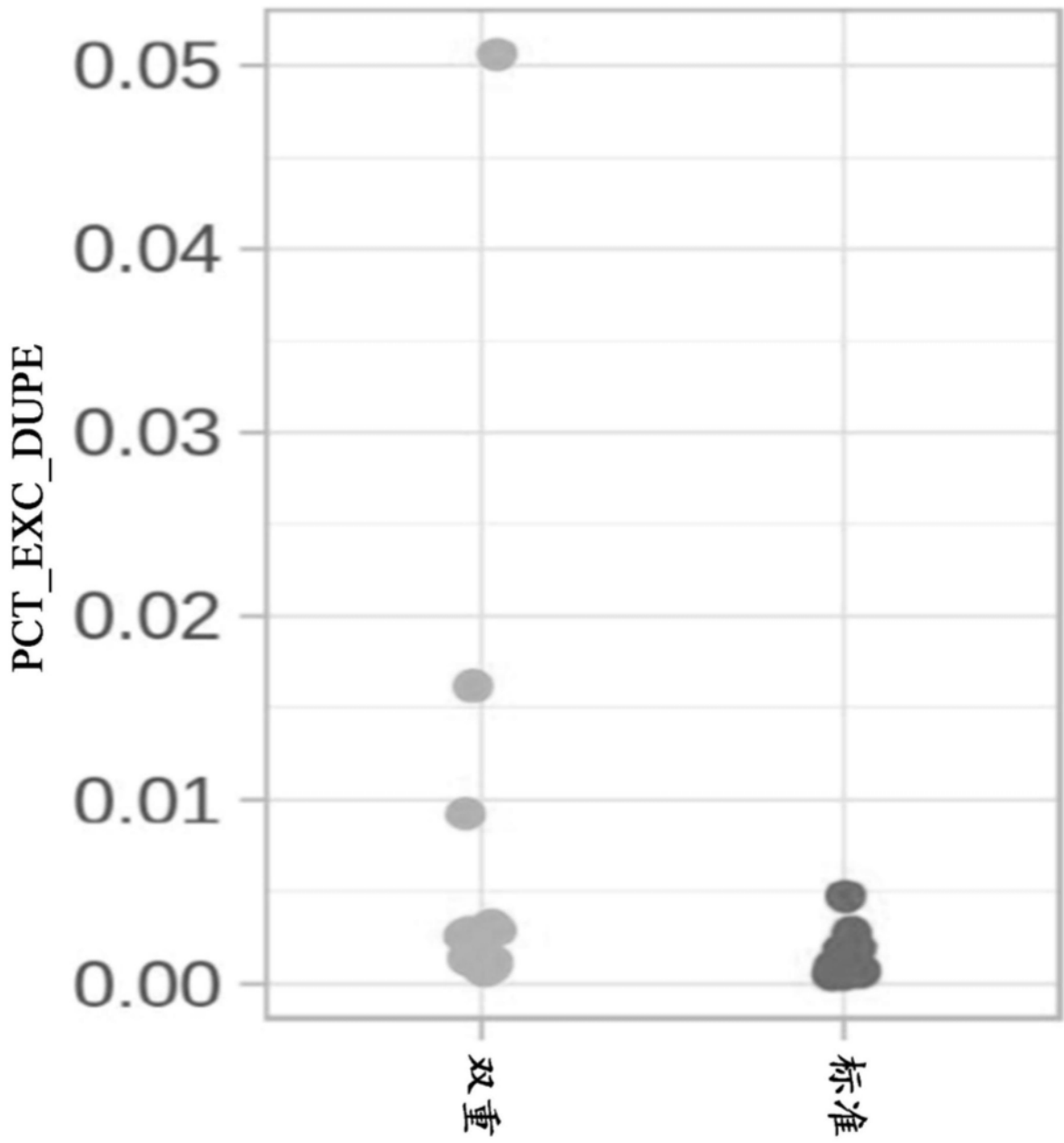


图10B

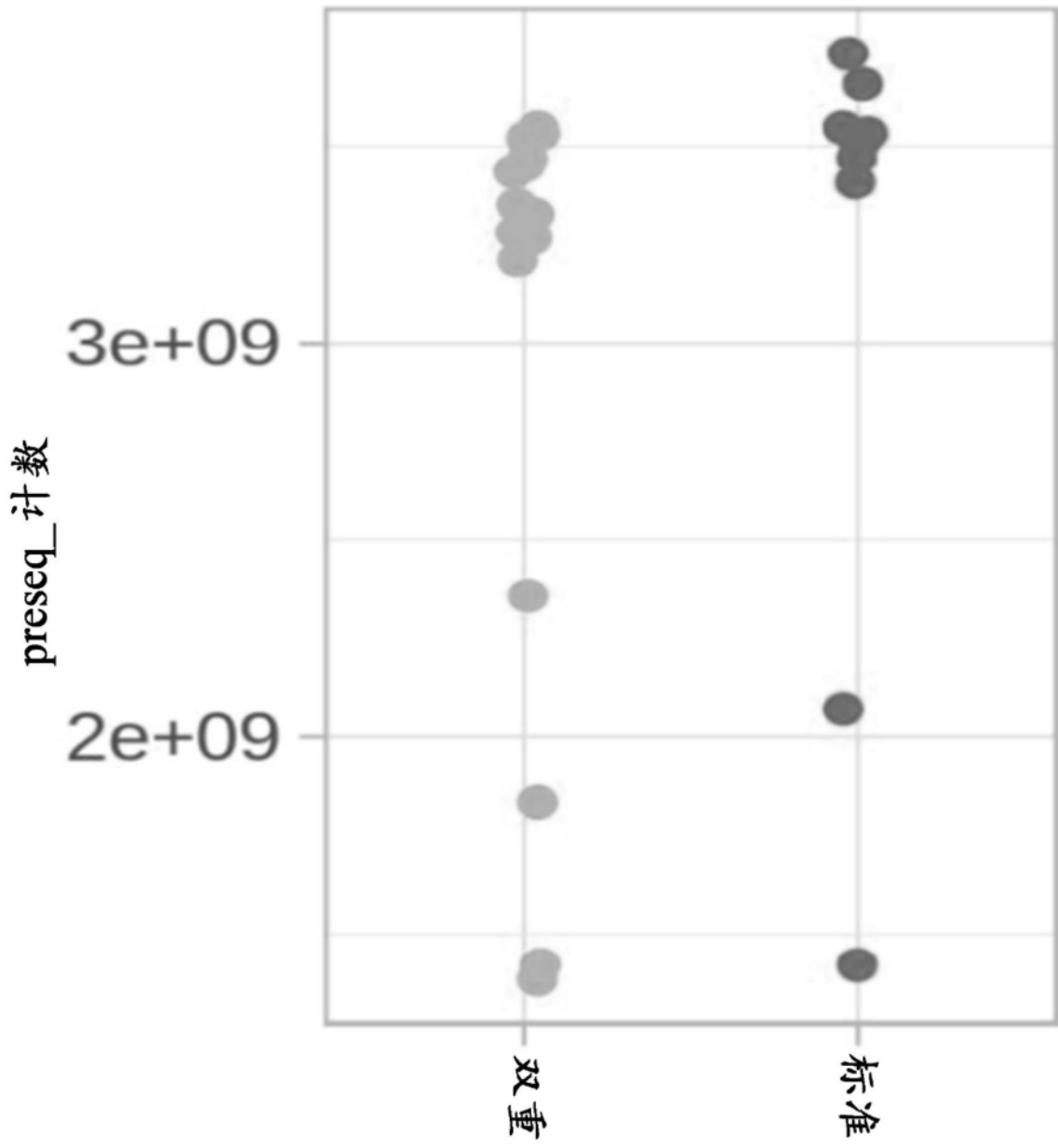


图10C

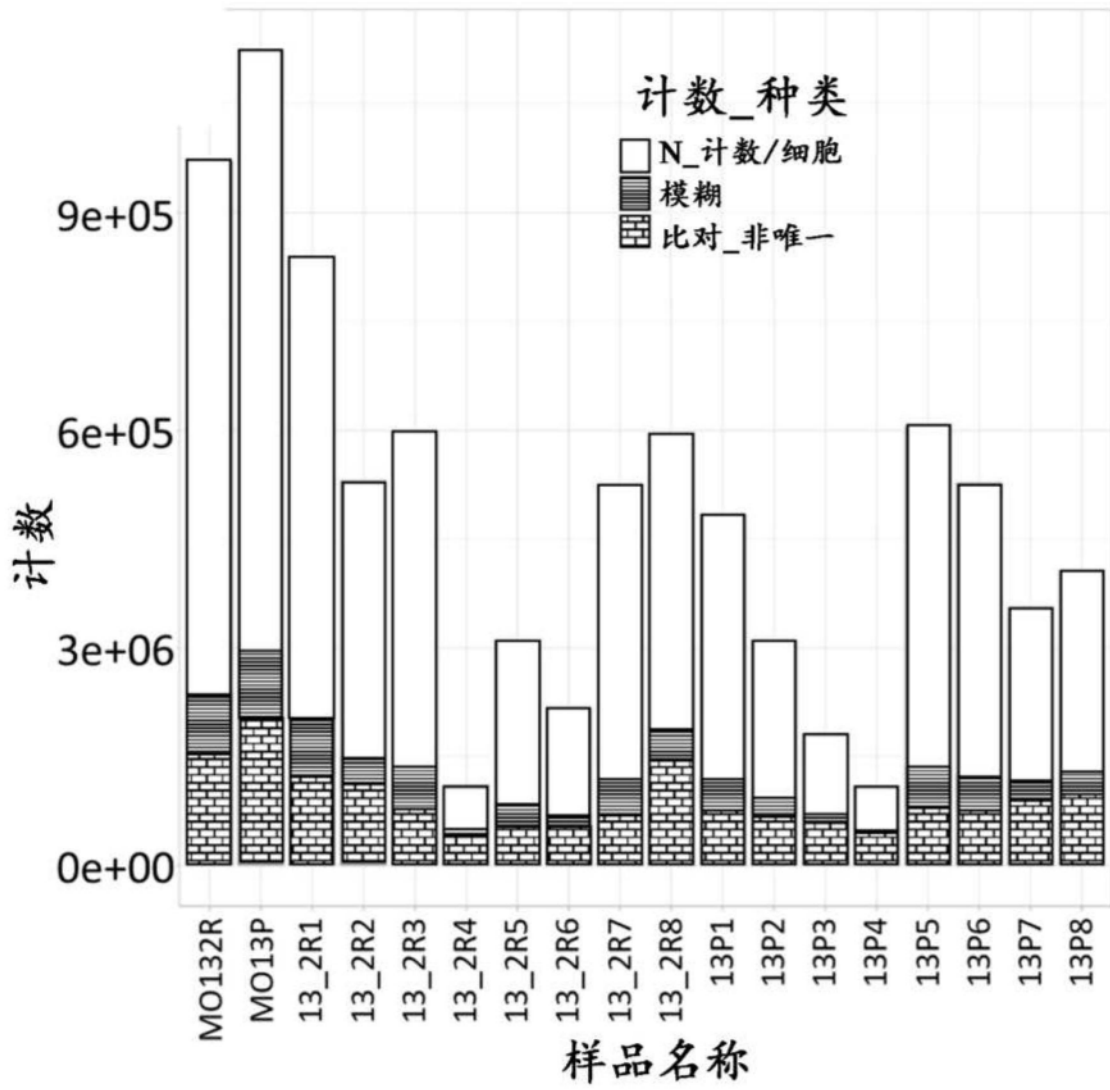


图10D

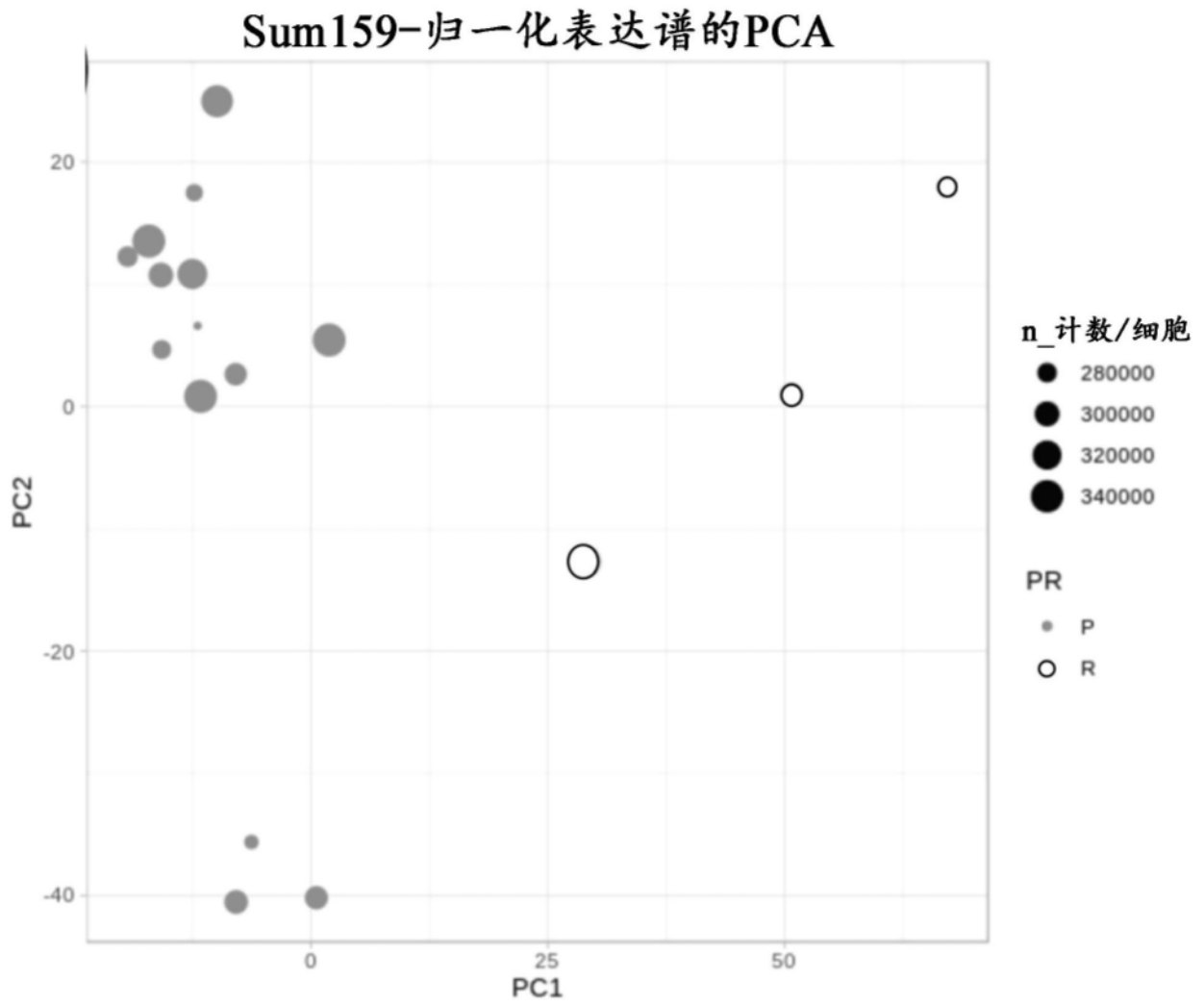


图10E

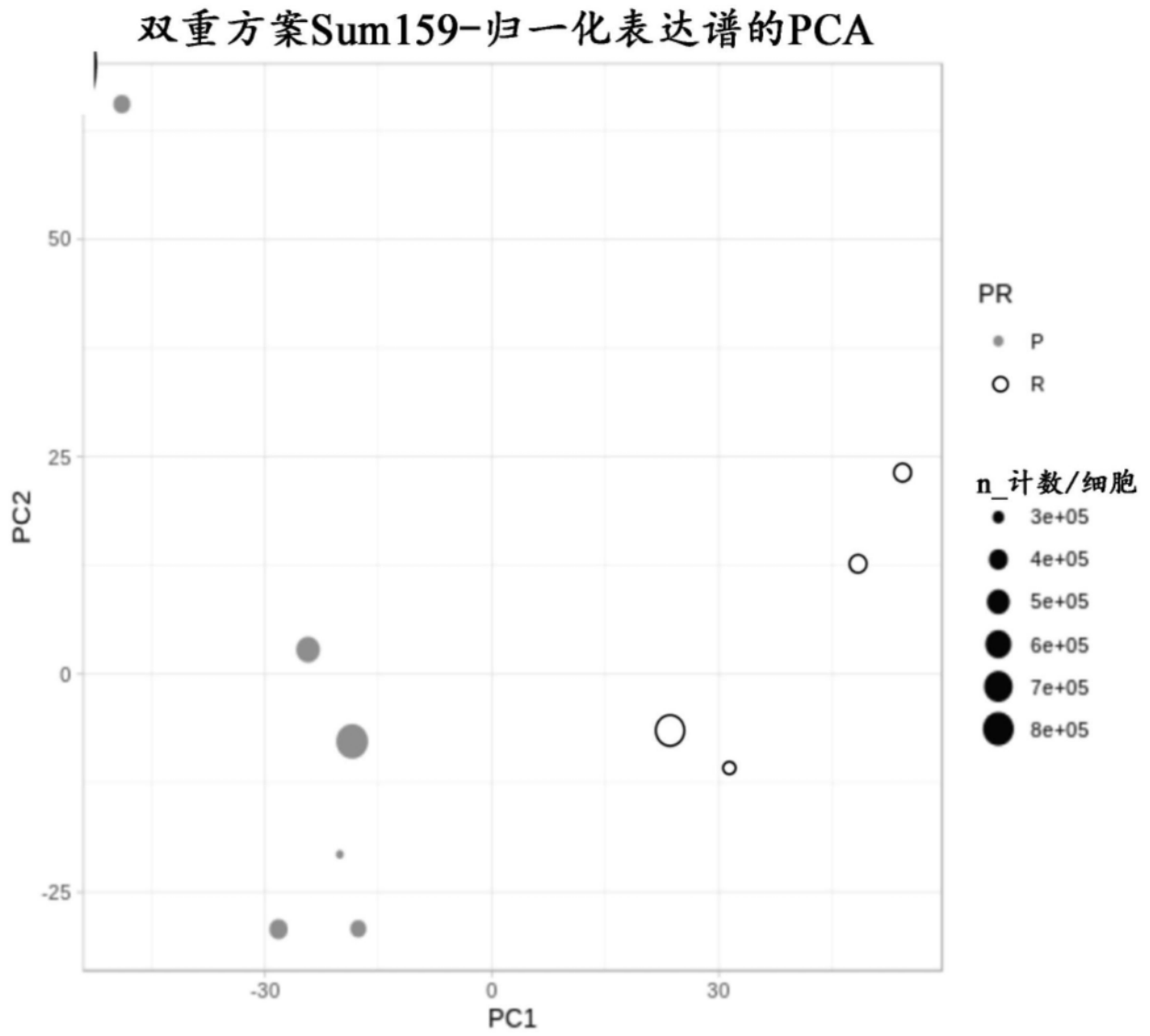


图10F

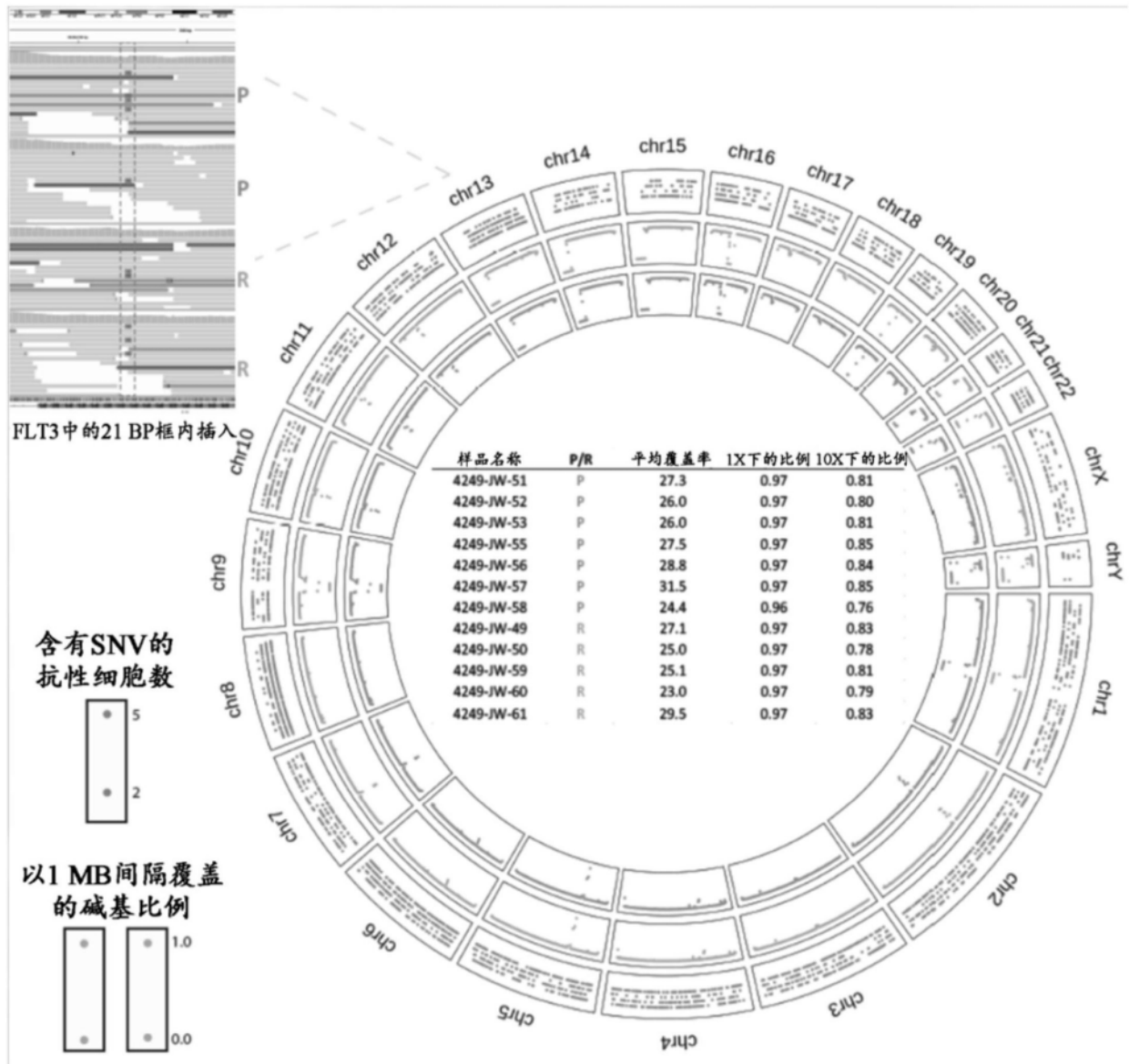


图11A

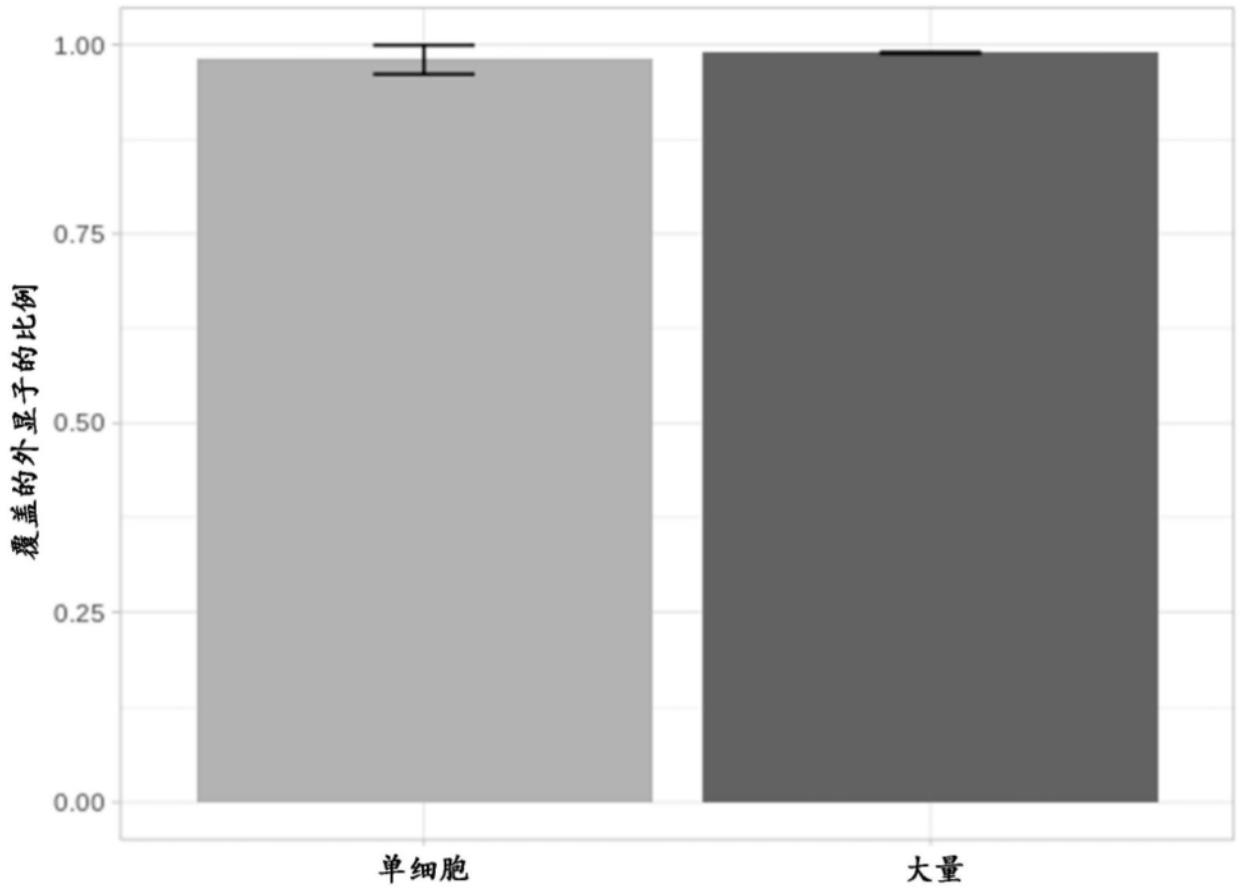


图12A

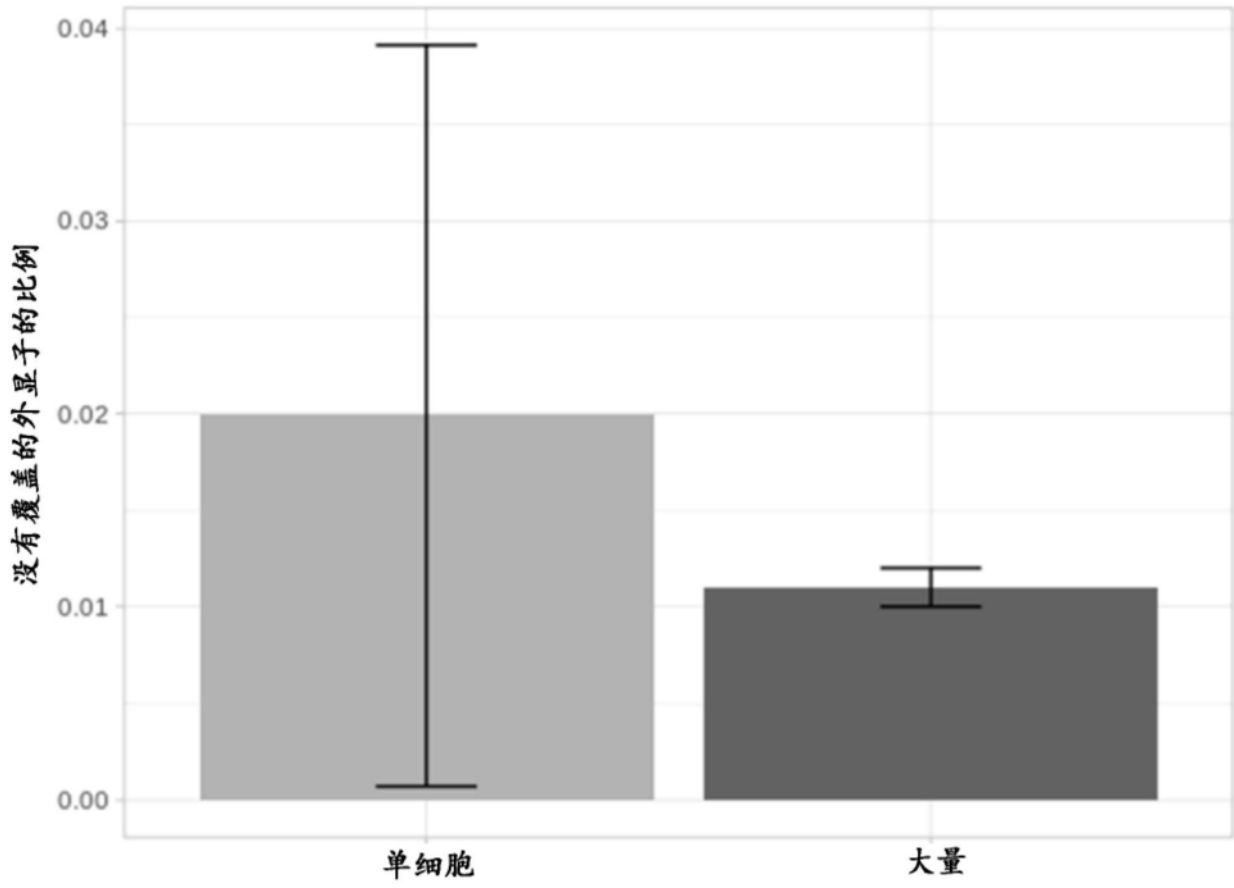


图12B

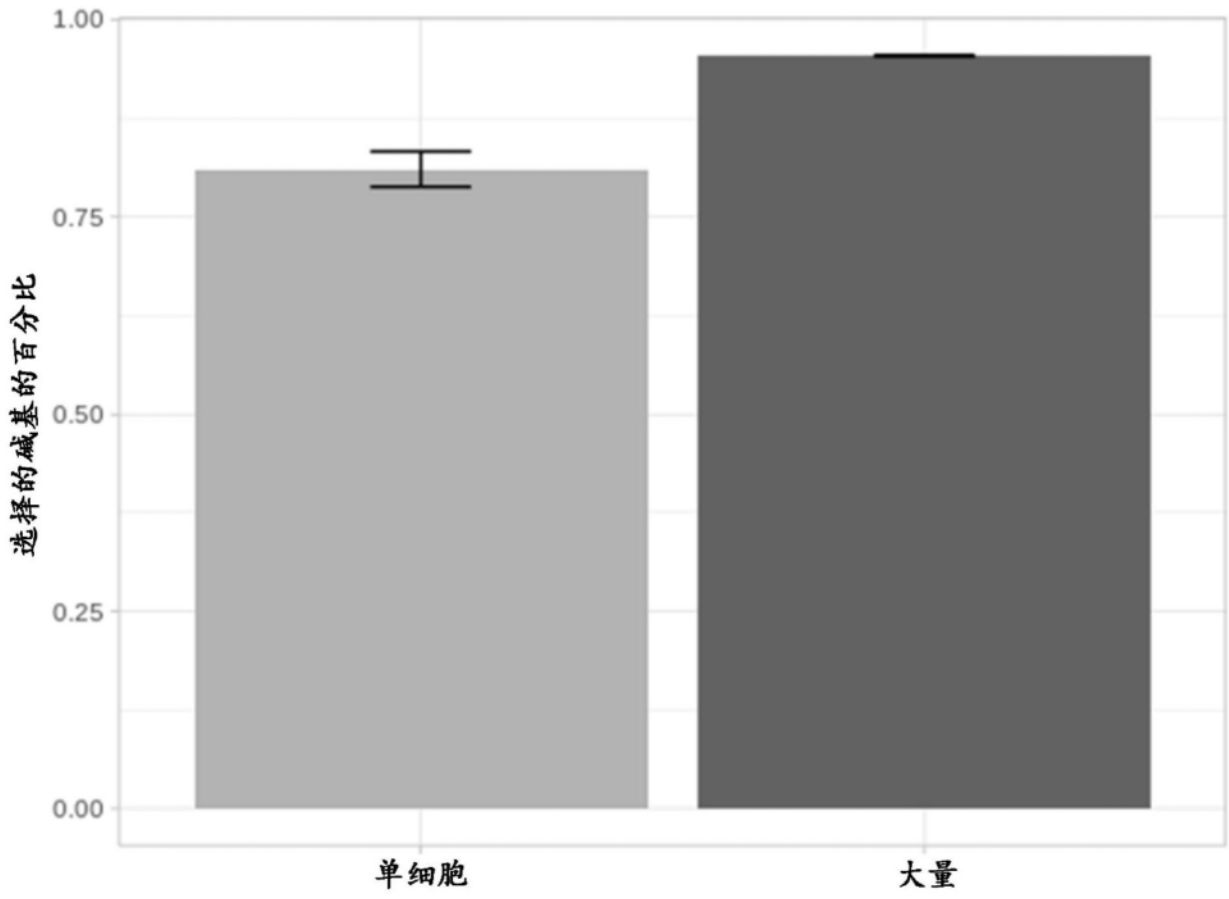


图12C

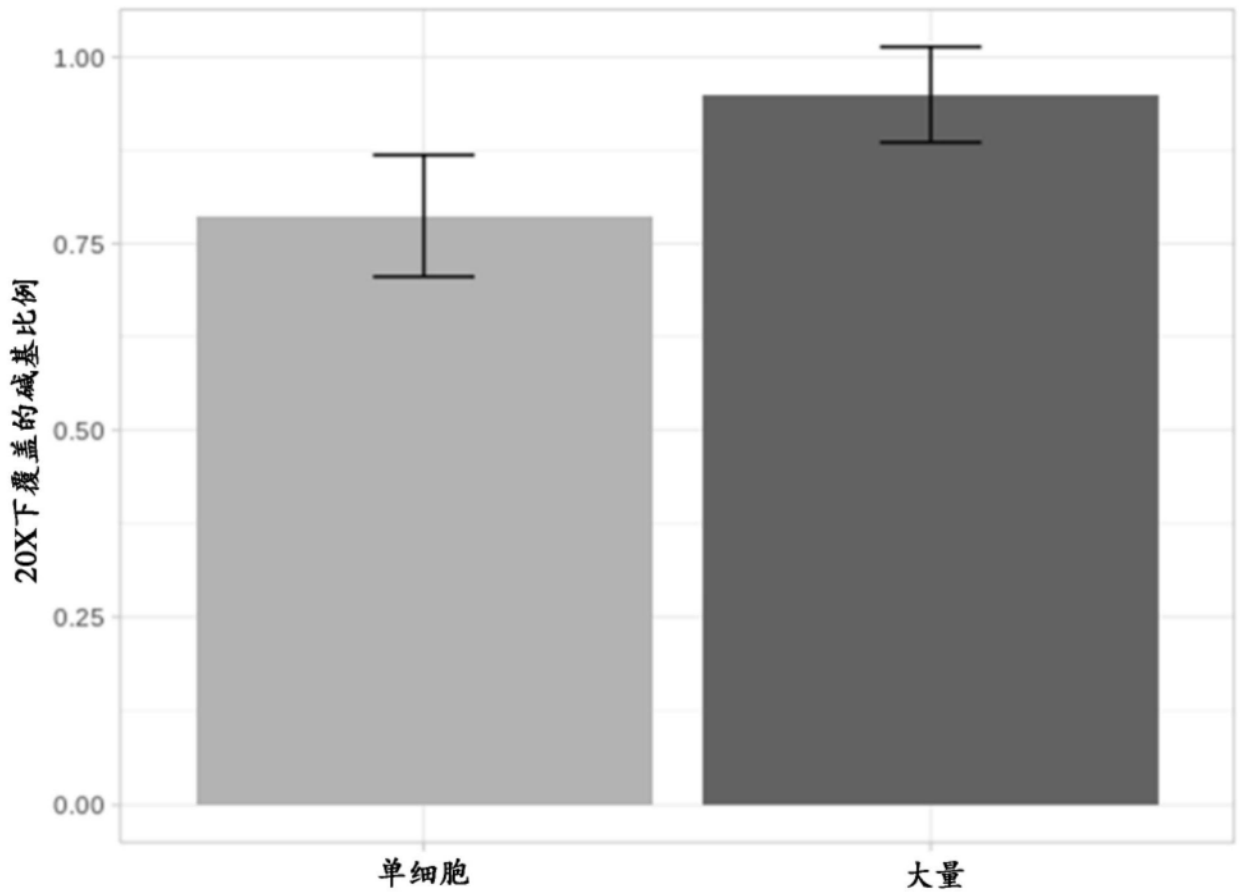


图12D

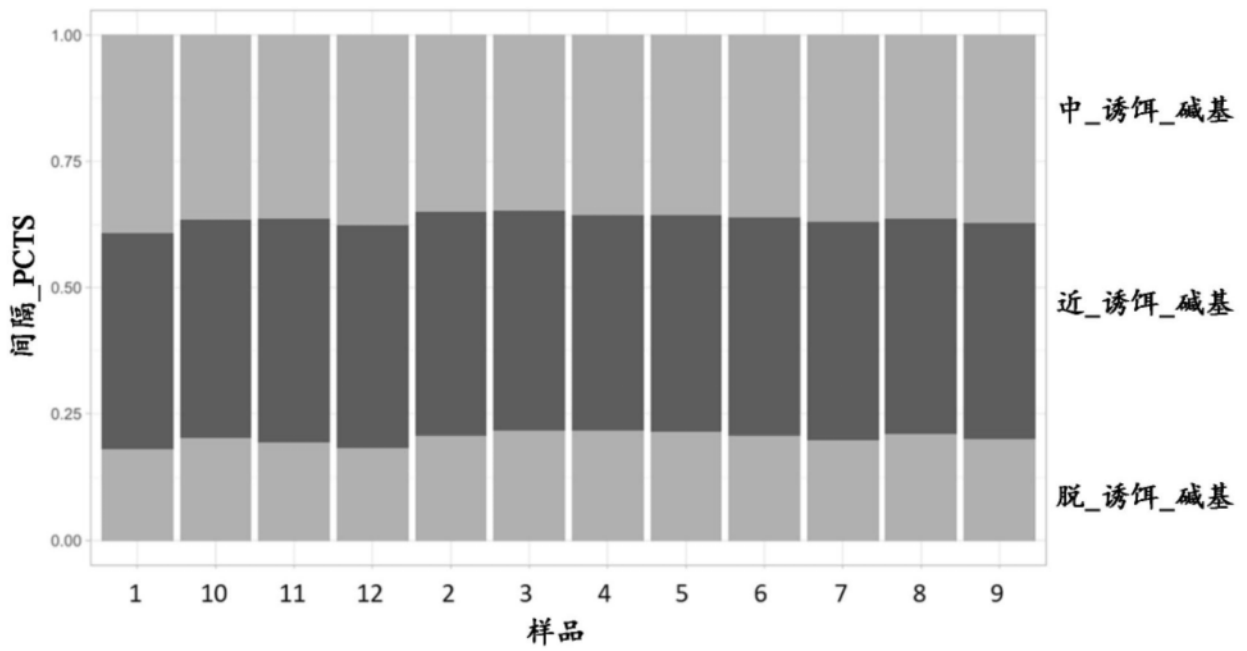


图13A

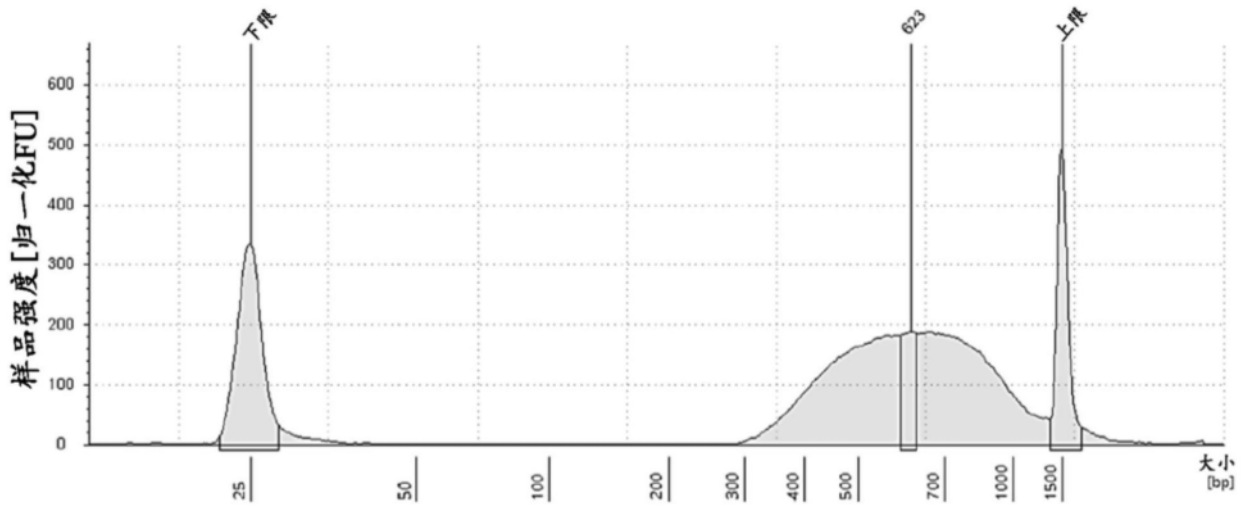


图13B

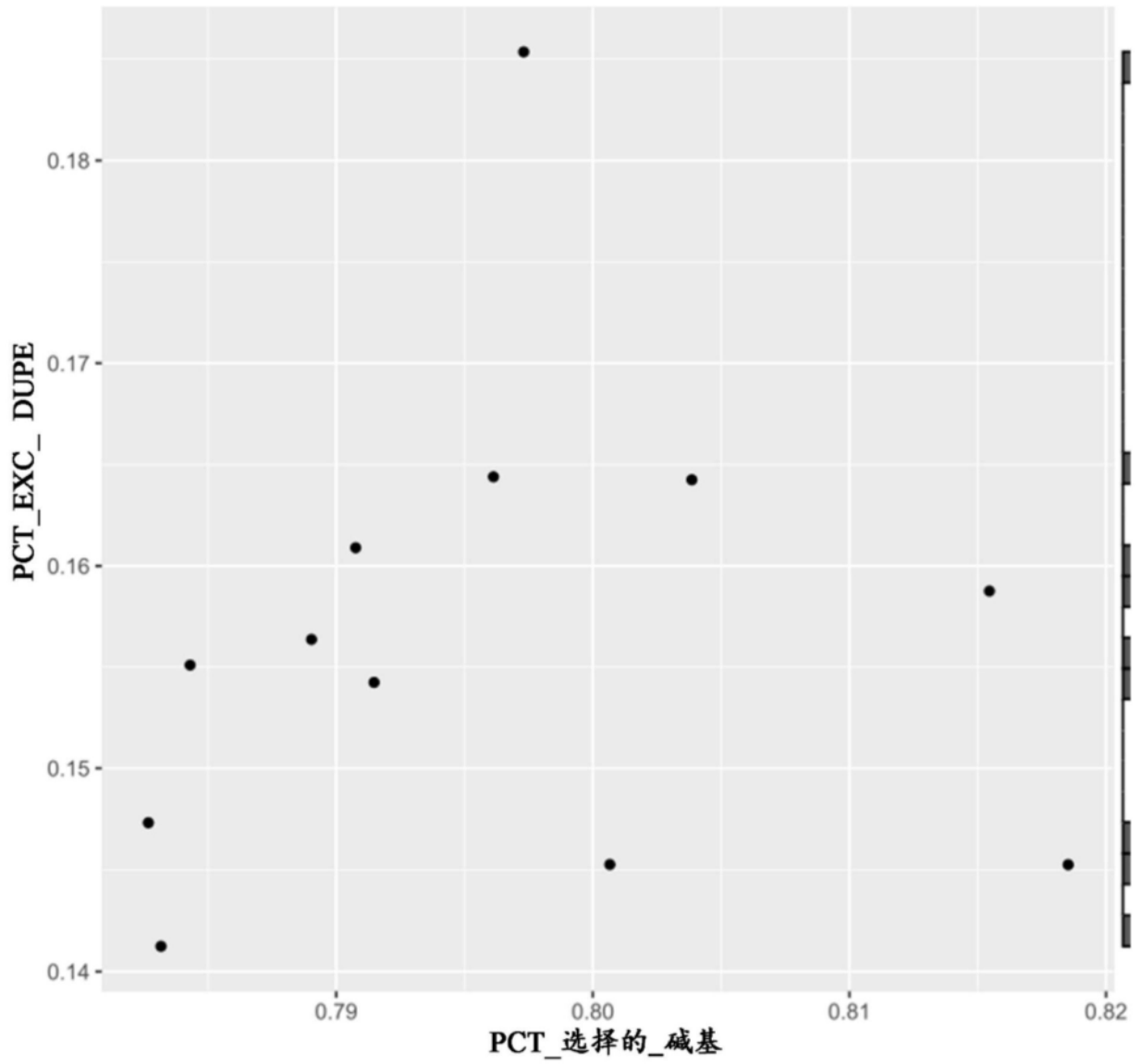


图14A

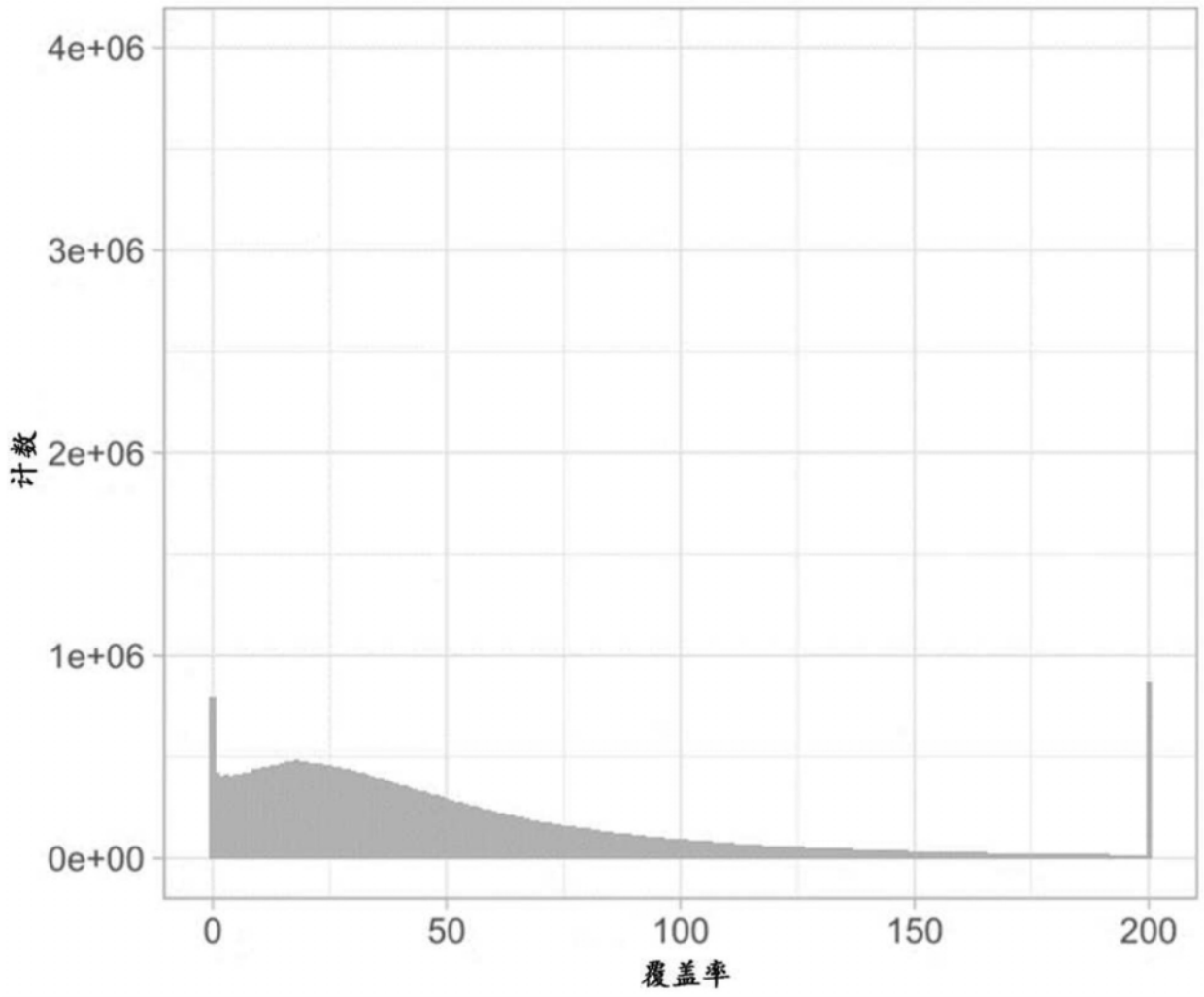


图14B