



## (12) 发明专利

(10) 授权公告号 CN 112210834 B

(45) 授权公告日 2025.02.25

(21) 申请号 202010451081.2

(74) 专利代理机构 北京安信方达知识产权代理

(22) 申请日 2015.06.25

有限公司 11262

(65) 同一申请的已公布的文献号

专利代理人 王玮玮 郑霞

申请公布号 CN 112210834 A

(51) Int.CI.

(43) 申请公布日 2021.01.12

C40B 50/06 (2006.01)

(30) 优先权数据

(56) 对比文件

62/017,786 2014.06.26 US

US 2010120098 A1, 2010.05.13

62/027,198 2014.07.21 US

WO 2013131962 A1, 2013.09.12

(62) 分案原申请数据

审查员 纪圆圆

201580046283.6 2015.06.25

(73) 专利权人 伊鲁米那股份有限公司

权利要求书3页 说明书35页 附图26页

地址 美国加利福尼亚州

(72) 发明人 F·凯珀 戈登·卡恩

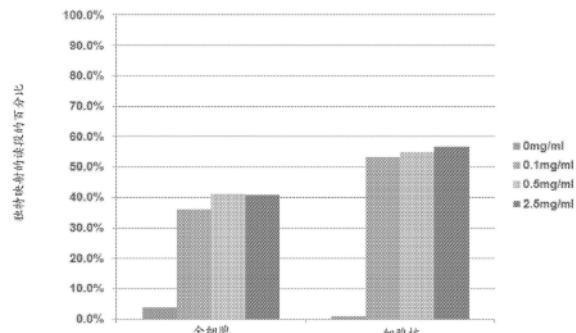
## (54) 发明名称

利用单管添加方案的加标签的核酸的文库

制备

## (57) 摘要

本申请涉及利用单管添加方案的加标签的核酸的文库制备。本申请提供了一种制备加标签的核酸片段的文库的方法，所述方法包括：使细胞群体与具有一种或更多种蛋白酶的裂解试剂直接接触，以生成细胞裂解物；使蛋白酶失活，以生成失活的细胞裂解物，和在其中靶核酸和转座子末端成分经历转座反应的条件下，将转座酶和包含转移链的转座子末端成分应用至失活的细胞裂解物。本申请还提供了一种用于制备加标签的核酸片段的文库的试剂盒，所述试剂盒包括：  
(a) 裂解试剂，所述裂解试剂具有一种或更多种蛋白酶，和(b) 转座反应组合物，所述转座反应组合物具有至少一种转座酶和包含转移链的至少一种转座子末端成分。



1. 一种制备加标签的核酸片段的文库的方法,所述方法包括:

(a) 使细胞群体与裂解试剂直接接触以生成细胞裂解物,所述细胞群体包含至少两个细胞且小于15个细胞,其中所述裂解试剂具有一种或更多种蛋白酶,并且其中所述细胞裂解物包含靶核酸;

(b) 使所述一种或更多种蛋白酶失活,以形成失活的细胞裂解物,以及

(c) 在其中所述靶核酸和转座子末端成分经历转座反应以生成混合体的条件下,将至少一种转座酶和包含转移链的至少一种转座子末端成分直接应用至所述失活的细胞裂解物,其中:

(i) 所述靶核酸被片段化,以生成多个靶核酸片段,

(ii) 所述转座子末端成分的转移链被连接至多个所述靶核酸片段的每一个的5'末端,以生成多个5'加标签的靶核酸片段,

(iii) 所述靶核酸是双链DNA,且

(iv) 所述靶核酸在步骤(c)中应用转座酶和转座子末端成分之前依然是双链DNA,其中步骤(a)、(b)和(c)在单个反应管中进行并且在步骤(a)和(c)之间没有DNA纯化和扩增步骤。

2. 根据权利要求1所述的方法,其中所述细胞群体包含两个、三个、四个或五个细胞。

3. 根据权利要求1所述的方法,其中所述一种或更多种蛋白酶为枯草杆菌蛋白酶及其变体。

4. 根据权利要求1所述的方法,其中所述细胞裂解物中所述一种或更多种蛋白酶的浓度为4.5 mAU/ml至500 mAU/ml。

5. 根据权利要求4所述的方法,其中所述细胞裂解物中所述一种或更多种蛋白酶的浓度为22.5 mAU/ml。

6. 根据权利要求1所述的方法,其中在步骤(a)中所述细胞群体在pH 7.0至pH 10.0与所述裂解试剂接触。

7. 根据权利要求6所述的方法,其中所述细胞群体在pH 7.0至pH 9.0与所述裂解试剂接触。

8. 根据权利要求1所述的方法,其中在步骤(b)中所述一种或更多种蛋白酶通过增加温度失活。

9. 根据权利要求8所述的方法,其中所述一种或更多种蛋白酶通过增加温度至50°C-80°C失活。

10. 根据权利要求9所述的方法,其中所述一种或更多种蛋白酶通过增加温度至70°C失活。

11. 根据权利要求1所述的方法,其中所述一种或更多种蛋白酶通过添加所述一种或更多种蛋白酶的一种或更多种抑制剂失活。

12. 根据权利要求1所述的方法,其中所述裂解试剂包含一种或更多种去垢剂。

13. 根据权利要求12所述的方法,其中所述一种或更多种去垢剂为非离子型去垢剂。

14. 根据权利要求12所述的方法,其中所述一种或更多种去垢剂包括Triton。

15. 根据权利要求1所述的方法,其中所述靶核酸为基因组DNA。

16. 根据权利要求1所述的方法,其中所述靶核酸包含染色体DNA或其片段。

17. 根据权利要求1所述的方法,其中所述靶核酸包括基因组或部分基因组。
18. 根据权利要求1所述的方法,其中所述至少一种转座酶为Tn5转座酶。
19. 根据权利要求1所述的方法,其中所述至少一种转座子末端成分包括Tn5转座子末端。
20. 根据权利要求1所述的方法,其中所述转移链包含标签域,所述标签域包含限制性位点域、捕获标签域、测序标签域、扩增标签域、检测标签域和寻址标签域中的一个或更多个。
21. 根据权利要求1所述的方法,所述方法还包括:
  - (d) 在其中3'标签被连接至所述5'加标签的靶核酸片段以生成多个加双标签的靶核酸片段的条件下,将来自步骤(c)的混合体直接与至少一种核酸修饰酶一起孵育。
22. 根据权利要求21所述的方法,其中步骤(a)、(b)、(c)和(d)在单个反应管中进行。
23. 根据权利要求21所述的方法,其中所述至少一种核酸修饰酶为聚合酶,并且其中所述3'标签通过所述5'加标签的靶核酸片段的3'末端的延伸形成。
24. 根据权利要求21所述的方法,其中所述至少一种核酸修饰酶为连接酶,并且其中所述3'标签通过将寡核苷酸连接至所述5'加标签的靶核酸片段的3'末端形成。
25. 根据权利要求21所述的方法,所述方法还包括:
  - (e) 扩增一种或更多种加双标签的靶核酸片段,以生成在所述加双标签的靶核酸片段的5'末端和/或3'末端具有另外的序列的加标签的核酸片段的文库。
26. 根据权利要求25所述的方法,其中步骤(a)、(b)、(c)、(d)和(e)在单个反应管中进行。
27. 根据权利要求25所述的方法,其中所述扩增包括利用聚合酶链式反应(PCR)、链置换扩增反应、滚环扩增反应、连接酶链式反应、转录介导的扩增反应或环介导的扩增反应中的一个或更多个。
28. 根据权利要求25所述的方法,其中所述扩增包括使用单个引物的聚合酶链式反应,所述单个引物与所述加双标签的靶核酸片段的3'标签互补。
29. 根据权利要求25所述的方法,其中所述扩增包括使用第一引物和第二引物的聚合酶链式反应,其中所述第一引物的至少3'末端部分与所述加双标签的靶核酸片段的3'标签的至少一部分互补,并且其中所述第二引物的至少3'末端部分展现所述加双标签的靶核酸片段的5'标签的至少一部分的序列。
30. 根据权利要求29所述的方法,其中所述第一引物的5'末端部分与所述加双标签的靶核酸片段的3'标签不互补,并且其中所述第二引物的5'末端部分不展现所述加双标签的靶核酸片段的5'标签的至少一部分的序列。
31. 根据权利要求29-30中任一项所述的方法,其中所述第一引物包含第一通用序列,和/或其中所述第二引物包含第二通用序列。
32. 根据权利要求1所述的方法,所述方法还包括对所述加标签的核酸片段测序。
33. 根据权利要求32所述的方法,其中对所述加标签的核酸片段的所述测序包括使用合成测序、桥式聚合酶链式反应、链终止测序、杂交测序、纳米孔测序和连接测序中的一个或更多个。
34. 根据权利要求32所述的方法,其中对所述加标签的核酸片段的所述测序包括使用

下一代测序。

35. 根据权利要求32所述的方法,所述方法还包括分析拷贝数变异。
36. 根据权利要求32所述的方法,所述方法还包括分析单核苷酸变异。
37. 根据权利要求35所述的方法,其中所述分析拷贝数变异用于检测镶嵌现象。

## 利用单管添加方案的加标签的核酸的文库制备

[0001] 本申请是申请日为2015年6月25日,申请号为201580046283.6,发明名称为“利用单管添加方案的加标签的核酸的文库制备”的申请的分案申请。

### 技术领域

[0002] 本公开内容大体涉及用于制备核酸片段的文库的方法,且更特别地涉及用于在单管中利用蛋白酶制备核酸片段的文库用于多种应用包括例如下一代DNA测序的方法。

### 背景技术

[0003] 存在这样的多种方法和应用,对于其来说,期望生成片段化并加标签的核酸的文库,例如用作DNA测序的模板和/或用于分析拷贝数变异。

[0004] 最近开发的“下一代”DNA测序技术,诸如由Illumina, Inc. (San Diego, CA) 开发的那些“下一代”DNA测序技术,利用大规模并行或多重格式使能够在单个序列运行中从数百万个测序模板生成序列数据。“下一代”测序的该大规模并行性质要求生成包含来自靶核酸样品例如基因组DNA的核酸片段的集合或群体的核酸片段文库。更重要地,它要求这些核酸片段的组合展现出为来自靶核酸样品的序列的定性和/或定量代表的序列。当核酸样品来自细胞时,目前的用于生成核酸片段的文库的方法通常要求分离步骤,用于在核酸片段化之前将靶核酸从细胞分离。该核酸提取步骤通常浪费靶核酸样品,并且通常致使制备的核酸不能定性地代表来自样品的靶核酸。当样品的量有限或难以获得时,这成为特别严重的问题。为了解决该问题,目前的一些方法在片段化之前使用核酸扩增。但是,扩增无法确保靶核酸的代表性,因为靶核酸在扩增之前的提取期间仍部分地丢失。

[0005] 因此,对能够快速并有效地制备核酸片段文库的新方法存在需求。本公开内容通过提供用于利用蛋白酶在单个反应混合物中例如在单个管中制备核酸片段的文库的方法,解决了该需求。还提供了相关优势。

### 发明内容

[0006] 在一个方面,本文提供了制备加标签的核酸片段的文库的方法,所述方法包括:(a)使细胞群体直接与裂解试剂接触,以生成细胞裂解物,其中裂解试剂具有一种或更多种蛋白酶,并且其中细胞裂解物包含靶核酸;(b)使一种或更多种蛋白酶失活,以形式失活的细胞裂解物,和(c)在其中靶核酸和转座子末端成分经历转座反应以生成混合体(mixture)的条件下,将至少一种转座酶和包含转移链的至少一种转座子末端成分直接应用于失活的细胞裂解物,其中(i)靶核酸被片段化以生成多个靶核酸片段,并且(ii)转座子末端成分的转移链被连接至多个靶核酸片段的每一个的5'末端,以生成多个5'加标签的靶核酸片段。

[0007] 在一些实施方案中,本文提供的步骤(a)、(b)和(c)在单个反应混合物中例如在一个管中进行。在一些实施方案中,细胞群体为细胞的最小群体。在一些实施方案中,细胞的最小群体包含一个、两个、三个、四个或五个细胞。

[0008] 在一些实施方案中,一种或更多种蛋白酶选自由以下组成的组:丝氨酸蛋白酶、苏

氨酸蛋白酶、半胱氨酸蛋白酶、天冬氨酸蛋白酶、谷氨酸蛋白酶和金属蛋白酶。在一些实施方案中，一种或更多种蛋白酶为枯草杆菌蛋白酶及其变体。在一些实施方案中，一种或更多种蛋白酶在细胞裂解物中的浓度为0.1mg/ml至10mg/ml。在一些实施方案中，一种或更多种蛋白酶在细胞裂解物中的浓度为0.1mg/ml至2.5mg/ml。在一些实施方案中，一种或更多种蛋白酶在细胞裂解物中的浓度为0.5mg/ml。在一些实施方案中，一种或更多种蛋白酶在细胞裂解物中的浓度为4.5mAU/ml至500mAU/ml。在一些实施方案中，一种或更多种蛋白酶在细胞裂解物中的浓度为22.5mAU/ml。

[0009] 在一些实施方案中，在步骤(a)中细胞群体在pH 7.0至pH 10.0与裂解试剂接触。在一些实施方案中，细胞群体在pH 7.0至pH 9.0与裂解试剂接触。

[0010] 在一些实施方案中，在步骤(b)中一种或更多种蛋白酶通过增加温度失活。在一些实施方案中，一种或更多种蛋白酶通过将温度增加至50°C-80°C失活。在一些实施方案中，一种或更多种蛋白酶通过将温度增加至70°C失活。在一些实施方案中，一种或更多种蛋白酶通过添加该一种或更多种蛋白酶的一种或更多种抑制剂失活。

[0011] 在一些实施方案中，裂解试剂包含一种或更多种去垢剂。在一些实施方案中，一种或更多种去垢剂为非离子型去垢剂。在一些实施方案中，一种或更多种去垢剂包括Triton。

[0012] 在一些实施方案中，靶核酸为双链DNA，并且其中靶核酸在步骤(c)中应用转座酶和转座子末端成分之前维持双链DNA。在一些实施方案中，靶核酸是基因组DNA。在一些实施方案中，靶核酸包含染色体DNA或其片段。在一些实施方案中，靶核酸包括基因组或部分基因组。

[0013] 在一些实施方案中，至少一种转座酶为Tn5转座酶。在一些实施方案中，至少一种转座子末端成分包括Tn5转座子末端。

[0014] 在一些实施方案中，转移链包含标签域，所述标签域包含限制性位点域、捕获标签域、测序标签域、扩增标签域、检测标签域和寻址标签域中的一个或更多个。

[0015] 在一些实施方案中，本文提供的方法还包括：(d) 在其中3'标签与5'加标签的靶核酸片段连接以生成多个加双标签的靶核酸片段的条件下，使来自步骤(c)的混合体直接与至少一种核酸修饰酶一起孵育。在一些实施方案中，步骤(a)、(b)、(c)和(d)在单个反应管中进行。

[0016] 在一些实施方案中，核酸修饰酶为聚合酶，并且其中所述3'标签通过5'加标签的靶核酸片段的3'末端的延伸形成。在一些实施方案中，核酸修饰酶为连接酶，并且其中3'标签通过将寡核苷酸连接至5'加标签的靶核酸片段的3'末端形成。

[0017] 在一些实施方案中，本文提供的方法还包括(e)扩增一种或更多种加双标签的靶核酸片段，以生成在加双标签的核酸片段的5'末端和/或3'末端具有另外的序列的加标签的核酸片段的文库。在一些实施方案中，步骤(a)、(b)、(c)、(d)和(e)在单个反应管中进行。

[0018] 在一些实施方案中，扩增包括利用聚合酶链式反应(PCR)、链置换扩增反应、滚环扩增反应、连接酶链式反应、转录介导的扩增反应或环介导的扩增反应中的一个或更多个。在一些实施方案中，扩增包括使用与加双标签的靶DNA片段的3'标签互补的单个引物的PCR。在一些实施方案中，扩增包括使用第一引物和第二引物的PCR，其中第一引物的至少3'末端部分与加双标签的靶核酸片段的3'标签的至少一部分互补，并且其中第二引物的至少3'末端部分展现加双标签的靶核酸片段5'标签的至少一部分的序列。在一些实施方案中，

第一引物的5'末端部分与加双标签的靶核酸片段的3'标签不互补，并且第二引物的5'末端部分不展现加双标签的靶核酸片段的5'标签的至少一部分的序列。在一些实施方案中，第一引物包含第一通用序列，和/或其中第二引物包含第二通用序列。

[0019] 在一些实施方案中，本文提供的方法还包括对加标签的核酸片段测序。在一些实施方案中，加标签的核酸片段的测序包括使用合成测序、桥式PCR、链终止测序、杂交测序、纳米孔测序和连接测序中的一个或更多个。在一些实施方案中，加标签的核酸片段的测序包括使用下一代测序。

[0020] 在一些实施方案中，本文提供的方法还包括分析拷贝数变异。在一些实施方案中，本文提供的方法还包括分析单核苷酸变异。

[0021] 在另一个方面，本公开内容提供了用于制备加标签的核酸片段的文库的试剂盒，所述试剂盒包括：(a)具有一种或更多种蛋白酶的裂解试剂，和(b)具有至少一种蛋白酶和至少一种转座子末端成分的转座反应组合物，所述至少一个转座子末端成分包含转移链。

[0022] 在一些实施方案中，一种或更多种蛋白酶选自由以下组成的组：丝氨酸蛋白酶、苏氨酸蛋白酶、半胱氨酸蛋白酶、天冬氨酸蛋白酶、谷氨酸蛋白酶和金属蛋白酶。在一些实施方案中，一种或更多种蛋白酶为枯草杆菌蛋白酶及其变体。在一些实施方案中，裂解剂包含一种或更多种去垢剂。在一些实施方案中，一种或更多种去垢剂包括Triton。

[0023] 在一些实施方案中，至少一种转座子末端成分包含：标签域和包含转移链的3'部分。在一些实施方案中，标签域包含限制性位点域、捕获标签域、测序标签域、扩增标签域、检测标签域和寻址标签域中的一个或更多个。在一些实施方案中，转座反应组合物包含两种或更多种转座子末端成分，两种或更多种转座子末端成分的每种包含以至少一个核苷酸不同的转移链。在一些实施方案中，转座酶为Tn5转座酶。在一些实施方案中，转座子末端成分包括Tn5转座子末端。

[0024] 在一些实施方案中，本文提供的试剂盒还包括聚合酶。在一些实施方案中，本文提供的试剂盒还包括连接酶。

[0025] 在一些实施方案中，本文提供的试剂盒还包括用于扩增反应的试剂。在一些实施方案中，用于扩增反应的试剂为用于PCR的试剂。在一些实施方案中，用于扩增反应的试剂包括至少一种引物。在一些实施方案中，至少一种引物包含3'部分，所述3'部分展现转移链的至少一部分的序列。在一些实施方案中，至少一种引物包含5'部分，所述5'部分包含通用序列。

[0026] 在一些实施方案中，本文提供的试剂盒还包括尺寸选择试剂。在一些实施方案中，尺寸选择试剂包括AMPure XP珠。在一些实施方案中，本文提供的试剂盒还包括文库标准化试剂。

[0027] 在一些实施方案中，本文提供的试剂盒还包括具有固体表面的装置。在一些实施方案中，装置是流式细胞装置。在一些实施方案中，固体表面包括适于以有序的模式固定分子的模式化表面。

[0028] 具体地，本申请提供了以下内容：

[0029] 1. 一种制备加标签的核酸片段的文库的方法，所述方法包括：

[0030] (a)使细胞群体与裂解试剂直接接触，以生成细胞裂解物，其中所述裂解试剂具有一种或更多种蛋白酶，并且其中所述细胞裂解物包含靶核酸；

- [0031] (b) 使所述一种或更多种蛋白酶失活,以形成失活的细胞裂解物,以及
- [0032] (c) 在其中所述靶核酸和转座子末端成分经历转座反应以生成混合体的条件下,将至少一种转座酶和包含转移链的至少一种转座子末端成分直接应用至所述失活的细胞裂解物,其中:
- [0033] (i) 所述靶核酸被片段化,以生成多个靶核酸片段,且
- [0034] (ii) 所述转座子末端成分的转移链被连接至多个所述靶核酸片段的每一个的5'末端,以生成多个5'加标签的靶核酸片段。
- [0035] 2. 根据项目1所述的方法,其中步骤(a)、(b)和(c)在单个反应管中进行。
- [0036] 3. 根据项目1所述的方法,其中所述细胞群体为最小细胞群体,并且其中所述最小细胞群体包含一个、两个、三个、四个或五个细胞。
- [0037] 4. 根据项目1所述的方法,其中所述一种或更多种蛋白酶为枯草杆菌蛋白酶及其变体。
- [0038] 5. 根据项目1所述的方法,其中所述细胞裂解物中所述一种或更多种蛋白酶的浓度为4.5mAU/ml至500mAU/ml。
- [0039] 6. 根据项目5所述的方法,其中所述细胞裂解物中所述一种或更多种蛋白酶的浓度为22.5mAU/ml。
- [0040] 7. 根据项目1所述的方法,其中在步骤(a)中所述细胞群体在pH 7.0至pH 10.0与所述裂解试剂接触。
- [0041] 8. 根据项目7所述的方法,其中所述细胞群体在pH 7.0至pH 9.0与所述裂解试剂接触。
- [0042] 9. 根据项目1所述的方法,其中在步骤(b)中所述一种或更多种蛋白酶通过增加温度失活。
- [0043] 10. 根据项目9所述的方法,其中所述一种或更多种蛋白酶通过增加温度至50°C-80°C失活。
- [0044] 11. 根据项目10所述的方法,其中所述一种或更多种蛋白酶通过增加温度至70°C失活。
- [0045] 12. 根据项目1所述的方法,其中所述一种或更多种蛋白酶通过添加所述一种或更多种蛋白酶的一种或更多种抑制剂失活。
- [0046] 13. 根据项目1所述的方法,其中所述裂解试剂包含一种或更多种去垢剂。
- [0047] 14. 根据项目13所述的方法,其中所述一种或更多种去垢剂为非离子型去垢剂。
- [0048] 15. 根据项目13所述的方法,其中所述一种或更多种去垢剂包括Triton。
- [0049] 16. 根据项目1所述的方法,其中所述靶核酸为双链DNA,并且其中所述靶核酸在步骤(c)中应用转座酶和转座子末端成分之前维持所述双链DNA。
- [0050] 17. 根据项目16所述的方法,其中所述靶核酸为基因组DNA。
- [0051] 18. 根据项目16所述的方法,其中所述靶核酸包含染色体DNA或其片段。
- [0052] 19. 根据项目16所述的方法,其中所述靶核酸包括基因组或部分基因组。
- [0053] 20. 根据项目1所述的方法,其中所述至少一种转座酶为Tn5转座酶。
- [0054] 21. 根据项目1所述的方法,其中所述至少一种转座子末端成分包括Tn5转座子末端。

[0055] 22. 根据项目1所述的方法,其中所述转移链包含标签域,所述标签域包含限制性位点域、捕获标签域、测序标签域、扩增标签域、检测标签域和寻址标签域中的一个或更多个。

[0056] 23. 根据项目1所述的方法,所述方法还包括:

[0057] (d) 在其中3'标签被连接至所述5'加标签的靶核酸片段以生成多个加双标签的靶核酸片段的条件下,将来自步骤(c)的混合体直接与至少一种核酸修饰酶一起孵育。

[0058] 24. 根据项目23所述的方法,其中步骤(a)、(b)、(c)和(d)在单个反应管中进行。

[0059] 25. 根据项目23所述的方法,其中所述至少一种核酸修饰酶为聚合酶,并且其中所述3'标签通过所述5'加标签的靶核酸片段的3'末端的延伸形成。

[0060] 26. 根据项目23所述的方法,其中所述至少一种核酸修饰酶为连接酶,并且其中所述3'标签通过将寡核苷酸连接至所述5'加标签的靶核酸片段的3'末端形成。

[0061] 27. 根据项目23所述的方法,所述方法还包括:

[0062] (e) 扩增一种或更多种加双标签的靶核酸片段,以生成在所述加双标签的靶核酸片段的5'末端和/或3'末端具有另外的序列的加标签的核酸片段的文库。

[0063] 28. 根据项目27所述的方法,其中步骤(a)、(b)、(c)、(d)和(e)在单个反应管中进行。

[0064] 29. 根据项目27所述的方法,其中所述扩增包括利用聚合酶链式反应(PCR)、链置换扩增反应、滚环扩增反应、连接酶链式反应、转录介导的扩增反应或环介导的扩增反应中的一个或更多个。

[0065] 30. 根据项目27所述的方法,其中所述扩增包括使用单个引物的PCR,所述单个引物与所述加双标签的靶核酸片段的3'标签互补。

[0066] 31. 根据项目27所述的方法,其中所述扩增包括使用第一引物和第二引物的PCR,其中所述第一引物的至少3'末端部分与所述加双标签的靶核酸片段的3'标签的至少一部分互补,并且其中所述第二引物的至少3'末端部分展现所述加双标签的靶核酸片段的5'标签的至少一部分的序列。

[0067] 32. 根据项目31所述的方法,其中所述第一引物的5'末端部分与所述加双标签的靶核酸片段的3'标签不互补,并且其中所述第二引物的5'末端部分不展现所述加双标签的靶核酸片段的5'标签的至少一部分的序列。

[0068] 33. 根据项目31-32中任一项所述的方法,其中所述第一引物包含第一通用序列,和/或其中所述第二引物包含第二通用序列。

[0069] 34. 根据项目1所述的方法,所述方法还包括对所述加标签的核酸片段测序。

[0070] 35. 根据项目34所述的方法,其中对所述加标签的核酸片段的所述测序包括使用合成测序、桥式PCR、链终止测序、杂交测序、纳米孔测序和连接测序中的一个或更多个。

[0071] 36. 根据项目34所述的方法,其中对所述加标签的核酸片段的所述测序包括使用下一代测序。

[0072] 37. 根据项目34所述的方法,所述方法还包括分析拷贝数变异。

[0073] 38. 根据项目34所述的方法,所述方法还包括分析单核苷酸变异。

[0074] 39. 一种用于制备加标签的核酸片段的文库的试剂盒,所述试剂盒包括:

[0075] (a) 裂解试剂,所述裂解试剂具有一种或更多种蛋白酶,和

- [0076] (b) 转座反应组合物,所述转座反应组合物具有至少一种转座酶和包含转移链的至少一种转座子末端成分。
- [0077] 40. 根据项目39所述的试剂盒,其中所述一种或更多种蛋白酶为枯草杆菌蛋白酶及其变体。
- [0078] 41. 根据项目39所述的试剂盒,其中所述裂解试剂包含一种或更多种去垢剂。
- [0079] 42. 根据项目41所述的试剂盒,其中所述一种或更多种去垢剂包括Triton。
- [0080] 43. 根据项目39所述的试剂盒,其中所述至少一种转座子末端成分包含:标签域和包含转移链的3'部分。
- [0081] 44. 根据项目43所述的试剂盒,其中所述标签域包含限制性位点域、捕获标签域、测序标签域、扩增标签域、检测标签域和寻址标签域中的一个或更多个。
- [0082] 45. 根据项目39所述的试剂盒,其中所述转座反应组合物包含两种或更多种转座子末端成分,所述两种或更多种转座子末端成分的每一种包含至少一个核苷酸不同的转移链。
- [0083] 46. 根据项目39所述的试剂盒,其中所述转座酶为Tn5转座酶。
- [0084] 47. 根据项目39所述的试剂盒,其中所述转座子末端成分包括Tn5转座子末端。
- [0085] 48. 根据项目39所述的试剂盒,所述试剂盒还包括聚合酶。
- [0086] 49. 根据项目39所述的试剂盒,所述试剂盒还包括连接酶。
- [0087] 50. 根据项目39所述的试剂盒,所述试剂盒还包括用于扩增反应的试剂。
- [0088] 51. 根据项目50所述的试剂盒,其中所述用于扩增反应的试剂为用于PCR的试剂。
- [0089] 52. 根据项目51所述的试剂盒,其中所述用于扩增反应的试剂包括至少一种引物。
- [0090] 53. 根据项目52所述的试剂盒,其中所述至少一种引物包含3'部分,所述3'部分展现所述转移链的至少一部分的序列。
- [0091] 54. 根据项目52所述的试剂盒,其中所述至少一种引物包含5'部分,所述5'部分包含通用序列。
- [0092] 55. 根据项目39所述的试剂盒,所述试剂盒还包括尺寸选择试剂。
- [0093] 56. 根据项目55所述的试剂盒,其中所述尺寸选择试剂包括AMPure XP珠。
- [0094] 57. 根据项目39所述的试剂盒,所述试剂盒还包括文库标准化试剂。
- [0095] 58. 根据项目39所述的试剂盒,所述试剂盒还包括具有固体表面的装置。
- [0096] 59. 根据项目58所述的试剂盒,其中所述装置为流式细胞装置。
- [0097] 60. 根据项目58所述的试剂盒,其中所述固体表面包括适于以有序的模式固定分子的模式化表面。

## 附图说明

- [0098] 图1是显示在使用0mg/ml、0.1mg/ml、0.5mg/ml或2.5mg/ml蛋白酶处理的全细胞或细胞核的测序中,独特映射的读段(unique mapped read)的百分比的柱状图。
- [0099] 图2显示了使用总DNA、用足够的蛋白酶活性处理的单个细胞和用不足的蛋白酶活性处理的单个细胞的拷贝数分析结果的柱状图。
- [0100] 图3A显示了用0.5mg/ml活性蛋白酶、2mg/ml活性蛋白酶或2mg/ml预先热失活的蛋白酶处理的单个细胞中的拷贝数分析结果的柱状图。图3B显示了在用0.5mg/ml活性蛋白

酶、1mg/ml活性蛋白酶、2mg/ml在反应温度的蛋白酶、或2mg/ml预先热失活的蛋白酶处理的单个细胞和无细胞的对照样品的测序中,独特映射的读段的百分比的柱状图。图3C显示了在用0.5mg/ml活性蛋白酶、1mg/ml活性蛋白酶、2mg/ml活性蛋白酶、或2mg/ml预先热失活的蛋白酶处理的单个细胞和无细胞的对照样品的测序中,邻近箱之间的读段计数差异(邻近箱之间的读段计数差异的四分位数范围)的柱状图。

[0101] 图4A是显示在pH 7.0、pH 7.5、pH 8.0、pH 8.5、pH 9.0或pH 10.0蛋白酶的相对活性的柱状图。图4B显示了在用蛋白酶在pH 7.0、pH 8.0、pH 9.0或pH 10.0处理的单个细胞的测序实验中,独特映射的读段的百分比的柱状图。图4C显示了在用蛋白酶在pH 7.0、pH 8.0、pH 9.0或pH 10.0处理的单个细胞的测序实验中,邻近箱之间的读段计数差异(邻近箱之间的读段计数差异的四分位数范围)的柱状图。

[0102] 图5A是显示在室温、50°C、60°C或70°C相对蛋白酶活性的柱状图。图5B显示了在用在室温、50°C、60°C或70°C预热的蛋白酶处理的单个细胞、三个细胞或15pg基因组DNA的测序实验中,独特映射的读段的百分比的柱状图。图5C显示了在用在室温、50°C、60°C或70°C预热的蛋白酶处理的单个细胞、三个细胞或15pg基因组DNA的测序实验中,相邻箱之间的读段计数差异(相邻箱之间的读段计数差异的四分位数范围)的柱状图。

[0103] 图6A显示了用1μl Tn5或2μl Tn5处理生成的文库的插入尺寸。图6B显示了用1μl Tn5或2μl Tn5处理生成的文库的插入尺寸。图6C显示了用1μl Tn5或2μl Tn5处理生成的文库的多样性。

[0104] 图7显示了在根据本文提供的方法使用具有16个循环、18个循环或20个循环的PCR的单个细胞的测序实验中,计数和拷贝数分析结果的柱状图。

[0105] 图8A显示了三次单细胞测序实验的读段分布。图8B显示了单细胞测序、三细胞测序或五细胞测序的读段分布。图8C显示了使用单个细胞、三个细胞或五个细胞的平均文库多样性和评估的基因组覆盖率的柱状图。图8D显示了整体方案成功率。

[0106] 图9A显示了使用REPLIg单细胞(REPLIg Single Cell,MDA)与Nexterol XT文库制备(Nexterol XT library preparation)的拷贝数分析。图9B显示了使用SurePlex与Nexterol XT文库制备的拷贝数分析。图9C显示了使用本文提供的Nextera单细胞(Nextera Single Cell)的拷贝数分析。

[0107] 图10A显示了使用一式三份的单个GM50121细胞,18号染色体的拷贝数分析数据。图10B显示了使用一式三份的单个GM20916细胞的计数数目数据。图10C显示了使用一式三份的单个GM20916细胞,15号染色体、X染色体和10号染色体的拷贝数分析数据。图10D显示了使用一式三份的单个GM1021细胞,1号染色体和11号染色体的拷贝数分析数据。

## 具体实施方式

[0108] 本公开内容大体涉及用于制备核酸片段的文库的方法,且更特别地涉及用于在单个反应混合物例如单个管中使用蛋白酶制备核酸片段的文库用于多种应用包括例如下一代DNA测序的方法。

[0109] 定义

[0110] 如本文使用的,术语“包括(includes)”、“包括(including)”、“包括/includes)”、“包括(including)”、“含有(contains)”、“含有(containing)”、“具有(have)”、“具有

(having)”及其任何变形被意图覆盖非排除性包容物,以使得包括(includes)、包括(includes)或含有(contains)元素或一列元素的方法(process)、方法(method)、方法限定的产品或物质的组合物不仅包括那些元素,还可包括此类方法(process)、方法(method)、方法限定的产品或物质的组合物未明确列出或非固有的其它元素。

[0111] 如本文使用的,除非本文另外指示或与上下文明显矛盾,否则在描述本发明的上下文中(尤其是在以下权利要求书的上下文中)的术语“a(—)”和“an(—)”和“the(该)”以及相似的指示物,要被解释为覆盖单数和复数两者。

[0112] 如本文使用的,术语“约”或“大约”意指在给定的值或范围的5%以内。

[0113] 如本文使用的,术语“最小细胞群体”意指包含低于在标签化之前缺少分离步骤诸如DNA提取的情况下核酸测序能力的DNA拷贝量的细胞群体。示例性分离步骤包括从细胞裂解物提取DNA内容物和/或DNA扩增。最小细胞群体可以包含一个、两个、三个、四个或五个细胞。最小细胞群体可以是单个细胞。如本文使用的,“核酸测序能力”意指能产生清晰的基因组拷贝数变异数据的测序能力。

[0114] 如本文使用的,术语“核酸”意指核苷酸单体的单链聚合物和双链聚合物,包括通过核苷酸间磷酸二酯键键合或核苷酸间类似物连接的2'-脱氧核糖核苷酸(DNA)和核糖核苷酸(RNA),以及缔合的抗衡离子,例如H<sup>+</sup>、NH<sub>4</sub><sup>+</sup>、三烷基铵、四烷基铵、Mg<sup>2+</sup>、Na<sup>+</sup>等。核酸包括多核苷酸和寡核苷酸。核酸可完全由脱氧核糖核苷酸,完全由核糖核苷酸构成,或包括其镶嵌的混合物。核苷酸单体单位可包括本文描述的任何核苷酸,包括但不限于天然存在的核苷酸和核苷酸类似物。核酸的尺寸范围通常从几个单体单元例如5-40个到几千个单体核苷酸单元。核酸包括但不限于基因组DNA、cDNA、hnRNA、mRNA、rRNA、tRNA、片段化的核酸、从亚细胞细胞器诸如线粒体或叶绿体获得的核酸、以及从可存在于生物样品上或生物样品中的微生物或DNA或RNA病毒获得的核酸。

[0115] 如本文使用的,术语“靶核酸”被意图意指为分析或作用的目标物的核酸。分析或作用包括使核酸经历拷贝、扩增、测序和/或用于核酸询问(interrogation)的其他程序。靶核酸可包括除了待被分析的靶序列之外的核苷酸序列。例如,靶核酸可包含一个或更多个衔接子,所述衔接子包括在待被分析的靶核酸序列的侧翼作为引物结合位点发挥作用的衔接子。与捕获寡核苷酸或捕获引物杂交的靶核酸可包含延伸超过捕获寡核苷酸的5'或3'末端的核苷酸,以此方式使得并不是所有的靶核酸都适于延伸。

[0116] 如本文使用的,本文使用的术语“分离”和“纯化”指至少一种污染物(诸如蛋白和/或核酸序列)的量从样品或从将物质从其分离或纯化的来源的减少。

[0117] 如本文使用的,术语“尺寸选择”意指如下程序:在所述程序期间,从核酸片段的群体选择这样的核酸片段的子群体:其大部分核酸片段具有落入限定范围的核苷酸数目的,并因此具有落入限定范围的核苷酸数目的核酸片段的百分比增加。

[0118] 如本文使用的,术语“蛋白酶”指表现出水解多肽或具有多肽部分的底物的能力的蛋白、多肽或肽。在本方法中提供的蛋白酶可以是具有宽的特异性的单一蛋白酶。本方法可使用多种蛋白酶的混合物。本文提供的蛋白酶可以是热不稳定的并因此可通过热失活。在某些实施方案中,本文提供的蛋白酶可以在高于约25°C、30°C、35°C、40°C、45°C、50°C、55°C、60°C、65°C、70°C、75°C、80°C或高于约85°C的温度失活。本文提供的蛋白酶可消化染色质蛋白和其他DNA结合蛋白以释放裸露的基因组DNA,并且还可以消化内源性DNA酶以保护

DNA免于降解。本文提供的蛋白酶包括但不限于丝氨酸蛋白酶、苏氨酸蛋白酶、半胱氨酸蛋白酶、天冬氨酸蛋白酶、谷氨酸蛋白酶和金属蛋白酶。通常，天冬氨酸蛋白酶、谷氨酸蛋白酶和金属蛋白酶活化水分子，所述水分子对肽键进行亲核攻击以水解它。丝氨酸蛋白酶、苏氨酸蛋白酶和半胱氨酸蛋白酶通常使用亲和残基进行亲核攻击，以将蛋白酶共价地连接至底物蛋白，释放产物的第一半。然后该共价的酰基-酶中间体通过活化的水被水解，以通过释放产物的第二半并再生游离的酶而完成催化。本文使用的示例性蛋白酶包括从重组杆菌菌株分离的丝氨酸蛋白酶。本文使用的示例性蛋白酶包括枯草杆菌蛋白酶及其变体，包括枯草杆菌蛋白酶Carlsberg、alcalase和枯草杆菌蛋白酶S41。枯草杆菌蛋白酶及其变体是本领域技术人员已知的，并且包括例如alcalase、alcalase 0.6L、alcalase 2.5L、ALK酶、杆菌肽酶A、杆菌肽酶B、枯草杆菌碱性蛋白酶bioprase、bioprase AL 15、bioprase APL 30、colistinase、枯草杆菌蛋白酶J、枯草杆菌蛋白酶S41、枯草杆菌蛋白酶Sendai、枯草杆菌蛋白酶GX、枯草杆菌蛋白酶E、枯草杆菌蛋白酶BL、genenase I、esperase、maxatase、耐热蛋白酶(thermoase)PC 10、蛋白酶XXVII、耐热酶、superase、枯草杆菌蛋白酶Carlsberg、枯草杆菌蛋白酶DY、枯草杆菌肽酶、SP 266、savinase 8.0L、savinase 4.0T、kazusase、蛋白酶VIII、opticlean、肌球吸附蛋白(protin)A 3L、savinase、savinase 16.0L、savinase 32.0L EX、orientase 10B、蛋白酶S、丝氨酸内肽酶。在本文展示的方法和组合物的特定实施方案中，可使用热不稳定性蛋白酶诸如枯草杆菌蛋白酶和枯草杆菌蛋白酶的热不稳定性变体，如由Davail等，1994, J.Biol.Chem., 26:17448-17453的示例性公开内容展示的，其通过引用以其整体并入本文。

[0119] 如本文使用的，术语“蛋白酶抑制剂”指能够至少部分地降低蛋白酶水解肽的能力的物质，例如化合物。

[0120] 如本文使用的，术语“连接酶”指催化核酸链的5'-磷酸末端和3'-羟基末端之间的磷酸二酯键的分子内和分子间形成的核酸修饰酶。连接酶包括例如非模板依赖性连接酶，诸如CIRCLIGASE<sup>TM</sup> ssDNA连接酶，其可连接单链RNA和DNA的末端；以及模板依赖性的，其封闭双链DNA中的切口。如本文使用的，“模板依赖性连接酶”意指当退火至互补的多核苷酸时，催化DNA链彼此邻近的5'-磷酸末端和3'-羟基末端之间的磷酸二酯键的分子内和分子间形成的DNA连接酶。待被连接的两个DNA末端邻近地退火至的多核苷酸，在本文中被称为“连接模板”，并且连接被称为“模板依赖性连接”。连接模板可以是生物样品中的基因组或其他DNA中的互补DNA序列，或者连接模板可以是合成的和/或被特别地提供用于特定的测定或方法的“桥式寡脱氧核糖核苷酸”或“连接夹板寡脱氧核糖核苷酸(ligation splint oligodeoxyribonucleotide)”(或“连接夹板(ligation splint)”)。模板依赖性DNA连接酶的实例包括：NAD型DNA连接酶，诸如大肠杆菌(E.coli)DNA连接酶、Tth DNA连接酶、TfI DNA连接酶和AMPLIGASE<sup>®</sup> DNA连接酶(EPICENTRE Biotechnologies, Madison, WI, USA)，NAD型DNA连接酶只有在连接模板的存在下催化ssDNA分子的分子内连接；以及ATP型DNA连接酶，诸如T4 DNA连接酶或FASTLINK<sup>TM</sup> DNA连接酶(EPICENTRE Biotechnologies)。

[0121] 如本文使用的，术语“标签化”指通过转座体复合物修饰DNA，所述转座体复合物包含与包含转座子末端序列的衔接子复合的转座酶。标签化导致同时的DNA片段化和衔接子与双链体片段的两条链的5'末端的连接。可例如通过PCR、连接或本领域技术人员已知的任何其他合适的方法向适合的片段的末端添加另外的序列。如本文使用的，术语“转座体复合

物”指与双链核酸非共价地结合的转座酶。例如,复合物可以是在支持非共价复合物形成的条件下与双链转座子DNA预孵育的转座酶。双链转座子DNA可包括但不限于Tn5 DNA、Tn5 DNA的部分、转座子末端成分、转座子末端成分的混合物或能与转座酶诸如超活性Tn5转座酶相互作用的其他双链DNA。

[0122] 如本文使用的,术语“转座反应”指其中一个或更多个转座子在例如随机位点或几乎随机位点处被插入到靶核酸中的反应。转座反应中的重要组分是转座酶和展现转座子的核苷酸序列的DNA寡核苷酸,所述DNA寡核苷酸包含转移的转座子序列和其互补物(非转移的转座子末端序列)以及形成功能性转座或转座体复合物所需要的其它组分。DNA寡核苷酸还可根据需要或期望包含另外的序列(例如衔接子或引物序列)。在一些实施方案中,本文提供的方法通过采用由超活性Tn5转座酶和Tn5型转座子末端(Goryshin and Reznikoff, 1998, J.Biol.Chem., 273: 7367)或由MuA转座酶和包含R1和R2末端序列的Mu转座子末端(Mizuuchi, 1983, Cell, 35: 785; Savilahti等, 1995, EMBO J., 14: 4893)形成的转座复合物来示例。然而,能够将转座子末端以随机或几乎随机的方式以足够的效率插入以将靶DNA加5'标签和片段化用于其意图的目的的任何转座系统可用于本发明。可用于本方法的本领域已知的转座系统的实例包括但不限于金黄色葡萄球菌Tn552(Colegio等人, 2001, J Bacterid., 183: 2384-8; Kirby等, 2002, Mol Microbiol, 43: 173-86)、TyI(Devine和Boeke, 1994, Nucleic Acids Res., 22: 3765-72以及国际专利申请第W0 95/23875号)、转座子Tn7(Craig, 1996, Science, 271: 1512; Craig, 1996, Review in: Curr Top Microbiol Immunol, 204: 27-48)、TnI0和IS10(Kleckner等, 1996, Curr Top Microbiol Immunol, 204: 49-82)、Mariner转座酶(Lampe等, 1996, EMBO J., 15: 5470-9)、Tci(Plasterk, 1996, Curr Top Microbiol Immunol, 204: 125-43)、P因子(Gloor, 2004, Methods Mol Biol, 260: 97-114)、TnJ(Ichikawa和Ohtsubo, 1990, J Biol Chem, 265: 18829-32)、细菌插入序列(Ohtsubo和Sekine, 1996, Curr.Top.Microbiol.Immunol. 204: 1-26)、逆转录病毒(Brown等, 1989, Proc Natl Acad Sci USA, 86: 2525-9)和酵母的逆转录转座子(Boeke和Corces, 1989, Annu Rev Microbiol, 43: 403-34)。用于将转座子末端插入到靶序列中的方法可以利用任何合适的转座子系统在体外进行,对于所述合适的转座子系统合适的体外转座系统是可用的或所述合适的转座子系统可基于本领域的知识开发。通常,用于本文提供的方法的合适的体外转座系统至少需要:足够纯度、足够浓度和足够的体外转座活性的转座酶;转座酶与其一起形成功能性复合物的转座子末端,所述功能性复合物具有各自的能够催化转座反应的转座酶。可用于本发明的合适的转座酶转座子末端序列包括但不限于与转座酶一起形成复合物的野生型转座子末端序列、衍生的转座子末端序列或突变的转座子末端序列,所述转座酶选自野生型转座酶、衍生形式的转座酶、或突变形式的转座酶。

[0123] 如本文使用的,术语“转座酶”指能够与含有转座子末端的成分(例如,转座子、转座子末端、转座子末端成分)一起形成功能性复合物并能够在例如体外转座反应中催化含有转座子末端的成分插入或转座至所述含有转座子末端的成分与其一起孵育的双链靶核酸的酶。如本文所示的转座酶还可以包括来自逆转录转座子和逆转录病毒的整合酶。转座酶、转座体和转座体复合物通常是本领域技术人员已知的,如由US 2010/0120098的公开内容所示例的,其内容通过引用以其整体并入本文。尽管本文描述的很多实施方案涉及Tn5转座酶和/或超活性Tn5转座酶,但是将领体会到,能够为了其意图的目的以足够的效率将转座

子末端插入以将靶核酸加5'-标签和片段化的任何转座系统可用于本发明。在特定的实施方案中，转座系统能够以随机的方式或以几乎随机的方式将转座子末端插入以将靶核酸加5'-标签和片段化。

[0124] 如本文使用的，术语“转座子末端”指仅展现对于与在体外转座反应中是功能性的转座酶或整合酶一起形成复合物所必要的核苷酸序列（“转座子末端序列”）的双链核酸DNA。转座子末端与识别并结合转座子末端的转座酶或整合酶一起形成“复合物”或“突触复合物”或“转座体复合物”或“转座体组合物”，并且所述复合物能够将转座子末端插入至所述复合物在体外转座反应中与其一起孵育的靶DNA。转座子末端展现由“转移链”和“非转移链”组成的两个互补序列。例如，与在体外转座反应中有活性的高活性Tn5转座酶（例如，EZ-Tn5<sup>TM</sup>转座酶，EPICENTRE Biotechnologies, Madison, WI, USA）一起形成复合物的一个转座子末端包含展现“转移的转座子末端序列”的转移链和展现“非转移的转座子末端序列”的非转移链，所述“转移的转座子末端序列”如下：5'AGATGTGTATAAGAGACAG 3' (SEQ ID NO: 1)，且所述“非转移的转座子末端序列”如下：5'CTGTCT CTTATACACATCT 3' (SEQ ID NO: 2)。转移链的3'末端在体外转座反应中连接或转移至靶核酸。展现与转移的转座子末端序列互补的转座子末端序列的非转移链在体外转座反应中不被连接或转移至靶核酸。

[0125] 如本文使用的，术语“转座子末端成分”指包含转座子末端（能够与转座酶作用以经历转座反应的最小双链DNA区段）、任选地加上转移的转座子末端序列的5'-和/或非转移的转座子末端序列的3'的另外的一个或更多个序列的成分。例如，附接至标签的转座子末端是“转座子末端成分”。

[0126] 如本文使用的，术语“转移链”指“转座子末端”和“转座子末端成分”二者的转移部分（不管转座子末端是否附接至标签或其他部分）。相似地，术语“非转移链”指“转座子末端”和“转座子末端成分”二者的非转移部分。

[0127] 如本文使用的，术语“标签”指非靶核酸组分，通常是DNA，所述非靶核酸组分提供搜寻(addressing)所述标签所连接的核酸片段的手段。例如，在一些实施方案中，标签包括允许标签所附接至的DNA的鉴定、识别和/或分子或生物化学操作（例如，通过提供位点用于使寡核苷酸退火，所述寡核苷酸诸如用于通过DNA聚合酶延伸的引物，或用于捕获或用于连接反应的寡核苷酸）的核苷酸序列。将标签连接至核酸分子的方法在本文中有时候被称为“加标签”并且经历加标签或包含标签的核酸被称为“加标签的”（例如，“加标签的DNA”）。

[0128] 如本文使用的，术语“标签域”指展现用于期望的意图的目的或应用的序列的标签的部分或域。一种标签域是“转座子末端域”，该标签域展现转移的转座子末端序列。在一些实施方案中，转移链还展现出在转移的转座子末端序列的5'-的一个或更多个其他的核苷酸序列，标签还在5'部分具有一个或更多个其他的“标签域”，所述其他的“标签域”中的每个被提供用于任何期望的目的。例如，一些实施方案包含转座子末端成分，所述转座子末端成分包含标签域，所述标签域选自限制性位点标签域、捕获标签域、测序标签域、扩增标签域、检测标签域、寻址标签域和转录启动子域中的一个或更多个。

[0129] 如本文使用的，术语“限制性位点域”指展现出用于便于使用限制性内切核酸酶裂解的目的的序列的标签域。例如，限制性位点域可被用于生成加双标签的线性ssDNA片段。限制性位点域还可被用于生成标签域中的相容性双链5'末端，以使得可使用模板依赖性DNA连接酶将该末端连接至另一个DNA分子。

[0130] 如本文使用的,术语“捕获标签域”指展现出用于便于捕获标签域所连接的核酸片段的目的(例如,提供退火位点或亲和标签用于捕获珠或其他表面上的加双标签的线性ssDNA片段,例如其中标签域序列的退火位点允许通过退火至表面上的特定序列而捕获,所述特定序列诸如珠上、或微芯片或微阵列上、或测序珠上的探针)的序列的标签域。在一些实施方案中,捕获标签域包含转移链的5'部分,所述转移链的5'部分被连接至化学基团或部分,所述化学基团或部分包括亲和结合分子(例如生物素、链霉亲和素、抗原或结合抗原的抗体,它们允许在表面上捕获加双标签的线性ssDNA片段,第二亲和结合分子附接至所述表面以与第一亲和结合分子一起形成特定的结合对)。

[0131] 如本文使用的,术语“测序标签域”指展现用于便于标签所连接的核酸片段的测序的目的(例如,提供用于合成测序的引发位点,或提供用于连接测序的退火位点,或提供用于杂交测序的退火位点)的序列的标签域。

[0132] 如本文使用的,术语“扩增标签域”指展现用于便于所述标签所附加至的核酸的扩增的目的的序列的标签域。例如,在一些实施方案中,扩增标签域提供用于使用DNA聚合酶的核酸扩增反应(例如,PCR扩增反应或链置换扩增反应或滚环扩增反应)的引发位点,或用于在核酸扩增反应(例如,连接链式反应)中使用模板依赖性连接酶连接探针的连接模板。

[0133] 如本文使用的,术语“检测标签域”指展现用于便于检测加标签的核酸片段的目的的序列或可检测的化学或生物化学部分(例如,可视染料、荧光染料、化学发光染料或其他可检测染料;在底物的存在下可检测的酶,例如具有NBT加BCIP的碱性磷酸酶,或具有合适底物的过氧化物酶;可检测蛋白,例如绿色荧光蛋白;以及与可检测部分结合或可与另一种可检测的亲和结合分子一起形成亲和结合对或特异性结合对的亲和结合分子;或任何本领域已知的很多其他可检测的分子或系统)的标签域。

[0134] 如本文使用的,术语“寻址标签域”意指展现允许鉴定特定样品的序列的标签域(例如,其中转移链具有不同的寻址标签域,所述不同的寻址标签域针对每个样品展现出不同的序列)。

[0135] 如本文使用的,如涉及核酸或核酸反应使用的术语“扩增(amplify)”或“扩增的(amplified)”“扩增(amplifying)”指制备特定核酸的拷贝的体外方法,所述特定核酸诸如靶核酸或加标签的核酸。扩增核酸的很多方法是本领域已知的,并且扩增反应包括但不限于聚合酶链式反应、连接酶链式反应、链置换扩增反应、滚环扩增反应。扩增的核酸可以是DNA。不管起始核酸是DNA、RNA或两者,从一个核酸分子或更多个核酸分子(“扩增产物”)的扩增得到的产物都可以是DNA或RNA、或DNA和RNA核苷或核苷酸两者的混合物,或者它们可包括修饰的DNA或RNA核苷或核苷酸。“拷贝”并不必然意指与靶序列完美互补的序列或与靶序列一致。例如,拷贝可以包含:核苷酸类似物,诸如脱氧肌苷或脱氧尿苷;有意的序列改变(诸如通过包含与靶序列可杂交但不互补的序列的引物引入的序列改变);和/或在扩增期间出现的序列错误。

[0136] 如本文使用的,术语“加标签的核酸片段的文库”指从来源例如全基因组生成的加标签的核酸片段(例如,加双标签的核酸片段)的集合或群体,其中集合或群体中的加标签的核酸片段的组合展现了定性和/或定量代表生成加标签的核酸片段的来源例如全基因组的序列的序列。加标签的核酸片段的文库不包含代表来源展现的每个序列的加标签的核酸片段是可能的。

[0137] 如本文使用的,术语“核酸修饰酶”指作用于核酸例如DNA以影响修饰例如裂解、连接、聚合、磷酸化等的酶。核酸修饰酶包括例如聚合酶、核酸酶、转移酶、连接酶、磷酸化酶、磷酸酶、甲基化酶、转座酶等。“DNA修饰酶”包括作用于DNA的任何酶,包括也作用于其他底物诸如RNA的酶。

[0138] 如本文使用的,术语“DNA聚合酶”指催化脱氧核糖核苷酸聚合成DNA链的修饰酶。DNA聚合酶包括“模板依赖性DNA聚合酶”,所述“模板依赖性DNA聚合酶”需要模板核酸以确定聚合物中脱氧核糖核苷酸添加的顺利,或它们可以是“非模板依赖性的”以使得它们催化聚合而不参考模板序列。除了合成DNA聚合物之外,DNA聚合酶可包含其他特征或活性。例如,DNA聚合酶可表征为具有或缺乏5'到3'外切核酸酶活性(也被称为5'外切核酸酶或5'核酸酶活性)、3'到5'外切核酸酶活性和链置换活性。

[0139] 如本文使用的,术语“引物”为寡核苷酸(“寡核苷酸(oligo)”),通常具有可被核酸聚合酶延伸的游离3'-OH基团。对于模板依赖性聚合酶,通常引物寡核苷酸的至少3'部分与模板核酸的部分互补,寡核苷酸通过氢键和其他分子力“结合”(或“复合”、“退火”或“杂交”)至所述模板核酸的部分与模板结合,以给出用于起始通过DNA聚合酶的合成的引物/模板复合物,并且所述引物寡核苷酸的至少3'部分在DNA合成的过程中通过在其3'末端添加与模板互补的共价结合的碱基而被延伸。结果得到引物延伸产物。

[0140] 如本文使用的,术语“通用序列”指两个或更多个核酸分子共同或共有的核苷酸序列的区域。任选地,两个或更多个核酸分子还具有序列不同的区域。因此,例如,5'标签可包含相同或通用的核酸序列并且3'标签可包含相同或通用的序列。可存在于多个核酸分子的不同成员中的通用序列可允许使用与通用序列互补的单个通用引物来复制或扩增多个不同序列。

[0141] 如本文使用的,本文中的术语“固体表面”、“固体支持物”和其他语法等同物指适合于或可被改性以适合于附接多核苷酸的任何材料。可能的基底包括但不限于:玻璃和改性玻璃或功能化玻璃、塑料(包括丙烯酸树脂、聚苯乙烯以及苯乙烯和其他材料的共聚物、聚丙烯、聚乙烯、聚丁烯、聚氨酯、Teflon<sup>TM</sup>等)、多糖、尼龙或硝酸纤维素、陶瓷、树脂、二氧化硅或基于二氧化硅的材料包括硅和改性硅、碳、金属、无机玻璃、塑料、光学纤维束、和很多其他聚合物。在一些实施方案中,固体支持物和固体表面位于流通池装置内。在一些实施方案中,固体支持物包括适于以有序的模式固定转座体复合物的模式化表面。“模式化表面”指在固体支持物的暴露层中或暴露层上不同区域的排列。在一些实施方案中,固体支持物包括表面中的孔或凹的阵列。固体支持物的组成和几何结构可随其使用而变化。在一些实施方案中,固体支持物是平面结构,诸如玻片、芯片、微芯片和/或阵列。因此,基底的表面可以是平面层的形式。在一些实施方案中,固体支持物包括流通池的一个或更多个表面。如本文使用的术语“流通池”指包含固体表面的室,一种或更多种流体试剂可流动穿过所述室。可在本公开内容的方法中容易地使用的流通池和相关流体系统和检测平台的实例被描述于例如Bentley等人,Nature 456:53-59 (2008); WO 04/018497; US 7,057,026; WO 91/06678; WO 07/123744; US 7,329,492; US 7,211,414; US 7,315,019; US 7,405,281和US 2008/0108082中,通过引用将其每个并入本文。在一些实施方案中,固体支持物或其表面是非平面的,诸如管或容器的内表面或外表面。在一些实施方案中,固体支持物包括微球或珠。在本文中“微球”、“珠”、“颗粒”或语法等同物被意图意指由包括但不限于塑料、陶瓷、玻

璃和聚苯乙烯的多种材料制成的小的离散颗粒。在某些实施方案中，微球为磁性微球或珠。可选地或另外的，珠可以是多孔的。珠的尺寸的范围为从纳米例如100nm到毫米例如1mm。

[0142] 用于制备加标签的核酸片段的文库的方法

[0143] 本公开内容大体涉及用于制备核酸片段的文库的方法，且更特别地涉及用于在单个反应混合物例如单个反应管或其他容器中使用蛋白酶制备核酸片段的文库用于多种应用的方法，所述多种应用包括例如下一代DNA测序、分析拷贝数变异和分析单核苷酸变异。

[0144] 存在多种方法和应用，对于所述方法或应用，从最小细胞群体例如单个细胞制备核酸片段的文库用于多种应用诸如对基因组测序是期望的。目前用于制备核酸片段的文库的方法在DNA片段化之前需要分离核酸提取物和/或扩增步骤。通常，首先处理细胞以生成细胞裂解物，从所述细胞裂解提取并纯化靶核酸内容物。然后在分离步骤中，使用例如从 Illumina, Inc (San Diego, CA) 可得的Nextera转座体使纯化的靶核酸经历片段化。该分离核酸提取物步骤和在反应管或容器之间转移样品通常浪费靶核酸样品，并因此致使制备的核酸片段较不可能成功地充分代表来自样品的整个靶核酸。当细胞样品的量有限或难以获得时，这种不充分代表变的特别有挑战性。在通过预扩增步骤输入单个或少许细胞的情况下，已开发出一些方法以解决该问题。但是，这些方法无法有效解决不充分代表的问题，并通常引入高的噪声。本公开内容通过在例如单个管中以添加方案使用单反应混合物以生成核酸片段的文库提供了该问题的解决方案。本文提供的方法整合了多个步骤，包括在单个反应管中生成细胞裂解物、标签化等，任选地使用一种或更多种添加方案进行。在此类单管添加方法中，来自细胞的起始核酸材料的量被保持，并且从其生成的文库因此可更好地代表靶核酸例如基因组。

[0145] 在一个方面，本公开内容提供了制备加标签的核酸片段的文库的方法，所述方法包括：(a)使细胞群体直接与裂解试剂接触，以生成细胞裂解物，其中裂解试剂具有一种或更多种蛋白酶，并且其中细胞裂解物包含靶核酸；(b)使一种或更多种蛋白酶失活，以形式失活的细胞裂解物，和(c)在其中靶核酸和转座子末端成分经历转座反应以生成混合体的条件下，将至少一种转座酶和包含转移链的至少一种转座子末端成分直接应用至失活的细胞裂解物，其中(i)靶核酸被片段化以生成多个靶核酸片段，并且(ii)转座子末端成分的转移链被连接至多个靶核酸片段的每一个的5'末端，以生成多个5'加标签的靶核酸片段。

[0146] 在一些实施方案中，将细胞样品直接与包含一种或更多种蛋白酶的组合裂解试剂接触，并因此本文提供的蛋白酶可直接与完整的细胞接触。在一些实施方案中，使细胞样品与包含去垢剂的第一裂解试剂接触以生成第一细胞裂解物，并且然后将包含一种或更多种蛋白酶的第二裂解试剂添加至包含第一细胞裂解物的反应管。在该替代方案中，本文提供的蛋白酶与细胞裂解物接触。以下提供的实施例1例证了生成包含靶核酸的细胞裂解物的方法。在实施例1和表1-3中例证了包含去垢剂和QIAGEN (San Diego, CA) 蛋白酶(产品编号19155)的示例性裂解总混合物(master mixture)。

[0147] 根据本文提供的方法的起始材料可以是最小细胞群体，利用所述最小细胞群体，传统的测序方案通常可能由于不充分代表整个靶核酸例如基因组，只能产生噪声测序数据和拷贝数变异数据。在一些实施方案中，最小细胞群体可以包含一个、两个、三个、四个或五个细胞。在一些实施方案中，最小细胞群体可以是少于10个细胞、少于15个细胞、少于20个细胞、少于25个细胞、少于30个细胞、少于35个细胞、少于40个细胞、少于45个细胞、少于50

个细胞、少于60个细胞、少于70个细胞、少于80个细胞、少于90个细胞或少于100个细胞。在一些实施方案中，本方法中使用的起始材料只包含单个细胞。在一些实施方案中，靶核酸是基因组DNA。在一些实施方案中，靶核酸包含染色体DNA或其片段。在一些实施方案中，靶核酸包含基因组或部分基因组。

[0148] 本文使用的蛋白酶可消化染色质蛋白，例如组蛋白和其他DNA结合蛋白，以释放裸露的基因组DNA。另外，本文提供的蛋白酶可消化内源DNA酶以保护基因组免于降解。在一些实施方案中，本文的方法只使用拥有广谱特异性的一种蛋白酶，并因此蛋白酶可消化多种不同的蛋白和多肽，包括细胞中的一些或很多蛋白。在一些其他的实施方案中，广谱特异性可通过使用多种蛋白酶的混合物来实现，并且多种蛋白酶的组合可消化多种不同的蛋白和多肽，包括细胞中的一些或很多蛋白。示例性蛋白酶包括枯草杆菌蛋白酶，诸如alcalase、枯草杆菌蛋白酶carlsberg、枯草杆菌蛋白酶S41、热不稳定蛋白酶K和Qiagen蛋白酶。实施例4例证了蛋白酶活性对于均一地接近基因组DNA是有用的。应领会到，根据不同条件，例如细胞类型和样品量，可使用不同的蛋白酶和/或蛋白酶的混合物。

[0149] 在本文提供的每个反应中使用的蛋白酶的量和浓度可根据染色体DNA的量和/或使用的细胞数目以及蛋白酶的活性而变化。在一些实施方案中，在细胞裂解物中一种或更多种蛋白酶的浓度为0.1mg/ml至10mg/ml。在一些实施方案中，在细胞裂解物中一种或更多种蛋白酶的浓度为0.1mg/ml至2.5mg/ml。在一些实施方案中，在细胞裂解物中一种或更多种蛋白酶的浓度为2mAU/ml至500mAU/ml。在一些实施方案中，在细胞裂解物中一种或更多种蛋白酶的浓度为4.5mAU/ml至500mAU/ml。在一些实施方案中，在细胞裂解物中一种或更多种蛋白酶的浓度为10mAU/ml至100mAU/ml。本公开内容示例了使用蛋白酶例如QIAGEN蛋白酶(产品编号19155)测试并优化蛋白酶浓度，如实施例5中显示的。如在该实施例中显示的，当在正常反应温度(例如，室温)用0.5mg/ml(等于22.5mAU/ml)或2mg/ml(等于90mAU/ml)蛋白酶处理单个细胞时，相似地得到了清晰的拷贝数分析结果，如在图3A的上两幅柱状图中显示的。因此，在一些实施方案中，在细胞裂解物中蛋白酶的浓度为0.5mg/ml至2mg/ml。细胞裂解物中的蛋白酶的示例性浓度包括0.5mg/ml、0.6mg/ml、0.7mg/ml、0.8mg/ml、0.9mg/ml、1.0mg/ml、1.1mg/ml、1.2mg/ml、1.3mg/ml、1.4mg/ml、1.5mg/ml、1.6mg/ml、1.7mg/ml、1.8mg/ml、1.9mg/ml和2.0mg/ml。在一些实施方案中，在细胞裂解物中一种或更多种蛋白酶的浓度为20mAU/ml至90mAU/ml。在细胞裂解物中一种或更多种蛋白酶的示例性浓度包括20mAU/ml、30mAU/ml、40mAU/ml、50mAU/ml、60mAU/ml、70mAU/ml、80mAU/ml、90mAU/ml。

[0150] 包括PH值的多种条件可影响被反应管中的蛋白酶的消化和其他酶的活性两者，并因此可优化这些条件例如PH值。实施例6例证了优化蛋白酶消化反应的pH条件，平衡蛋白酶活性和测序结果。如显示的，在不同的pH条件下分析QIAGEN蛋白酶活性，并且蛋白酶的活性在pH 7.0至pH 10.0的范围内随着pH值增加而增加，蛋白酶在pH 7.0具有最低的活性且在pH10.0具有最高的活性。然后，还在不同的pH条件下分析拷贝数数据中独特映射的读段和噪声的百分比。如显示的，当pH为7、8或9时，可得到约70%的清晰的独特映射的读段。但是，当pH为10时，可得到较小百分比的独特映射的读段，并且数据变异显著增加。相似地，当pH为7、8或9时，相邻箱(bins)之间的计数差异是相对小的(约20%)，具有小的变异；而在pH 10.0，相邻箱之间的计数差异显著增加，具有巨大的变异。因此，在一些实施方案中，细胞群

体在pH 7.0至pH 10.0与裂解试剂接触。在一些实施方案中,细胞群体在pH 7.0至pH 9.0与裂解试剂接触。示例性pH条件包括pH 7.0、pH 7.5、pH 8.0、pH 8.5、pH 9.0和pH 9.5。

[0151] 因为核酸制备和片段化步骤在同一反应管中进行,所以根据本方法的蛋白酶可有效地失活而不干扰通常需要双链DNA的下一标签化步骤可以是有益的。在一些实施方案中,蛋白酶可通过在标签化步骤之前增加温度而失活。高的温度可使双链DNA构象变性。因此,在一些实施方案中,本文提供的蛋白酶可在相对低的温度失活而不使双链DNA变性。实施例7例证了测试蛋白酶的热失活。如显示的,在不同的温度中测试蛋白酶活性,并且蛋白酶活性随着温度增加而逐渐降低,并在70°C完全失活。因此,在一些实施方案中,一种或更多种蛋白酶通过将温度增加至50°C-80°C而失活。在一些实施方案中,一种或更多种蛋白酶通过将温度增加至70°C而失活。

[0152] 在一些实施方案中,本文提供的蛋白酶还可通过向反应管添加蛋白酶抑制剂而失活。本文提供的蛋白酶抑制剂并不干扰稍后在同一反应管中将进行的标签化和扩增步骤。示例性蛋白酶抑制剂包括例如AEBSF、苯丁抑制素、E-64、胃蛋白酶抑制剂A、磷酸阿米酮、亮肽素、抑肽酶、盐酸苯丁抑制素、亮肽素、磷酸阿米酮二钠盐、弹性酶抑制剂(elastatinal)、抑肽酶、甲磺酸萘莫司他(nafamostat mesylate)、抗痛素(antipain)、PMSF(苯基甲基磺酰氟(phenylmethanesulfonyl fluoride))、Pefabloc、二异丙基氟磷酸和链霉菌属枯草杆菌蛋白酶抑制剂。

[0153] 如以上讨论的,还可向细胞添加一种或更多种去垢剂。在一些实施方案中,连同蛋白酶一起向细胞添加去垢剂。在其他实施方案中,先向细胞添加去垢剂,然后向反应管添加蛋白酶。本文使用的去垢剂的功能包括破坏细胞膜并释放可溶形式的细胞内物质。在一些实施方案中,本文使用的去垢剂不干扰下游的酶活性。因此,在一些实施方案中,使用非离子型去垢剂。这些去垢剂破坏蛋白-脂质缔合和脂质-脂质缔合,但不破坏蛋白-蛋白相互作用,并因此较不可能干扰其他的下游酶。通常,非离子型去垢剂包含不带电荷的、亲水性头基。通常非离子型去垢剂基于聚氧乙烯或糖苷。示例性非离子型去垢剂包括Tween® 80、Tween® 20 Tween、Triton® X-100、Triton® X-100-R、Triton® X-114、NP-40、Genapol® C-100、Genapol® X-100、Igepal® CA 630、Arlasolve® 200、Brij® 96/97 Triton、Brij® 98、Brij® 58、Brij® 35 Brij系列、Pluronic® L64、Pluronic® P84、非去垢剂磺基甜菜碱(NDSB 201)、amphipols(PMAL-C8)、CHAPS、辛基β-D-吡喃葡萄糖苷、皂苷、单十二烷基九乙二醇醚(C12E9,聚多卡醇)、十二烷基硫酸钠、N-月桂基肌氨酸、脱氧胆酸钠、胆汁盐、十六烷基三甲基溴化铵、SB3-10、SB3-12、脒基磺基甜菜碱-14、辛基硫代葡糖苷、麦芽糖苷、HEGA和MEGA系列。

[0154] 蛋白酶失活后,可在同一反应混合物中例如同一反应管中,通过添加包含在转座酶和转座子末端成分之间形成的稳定复合物的转座体组合物或使用分离的转座酶和转座子末端成分来进行体外转座反应。由转座酶催化的体外转座反应导致同时地将靶核酸破坏成片段和将标签连接至每个片段的5'末端。应理解,描述使用转座酶和转座子末端成分的任何方法还可使用从转座酶和转座子末端成分制备的转座体组合物,并且描述使用转座体组合物的任何方法还可使用转座体组合物所包含的分离的转座酶和转座子末端成分。

[0155] 在一些实施方案中,本文提供的方法包括在体外转座反应中在其中发生多次插入

至靶核酸的条件下,将包含靶核酸的失活细胞裂解物与至少一种转座酶和转座酶与其一起形成转座复合物的转座子末端成分一起孵育,并持续足以发生多次插入至靶核酸的时间,转座子末端成分包含:(i)转移链,所述转移链展现出转移的转座子末端序列和任选地在转移的转座子末端序列的5'-的另外的序列,以及(ii)非转移链,所述非转移链展现出与转移的转座子末端序列互补的序列,所述多次插入的每次导致包含转移链的第一标签连接至靶核酸中的核苷酸的5'末端,从而使靶核酸片段化并生成退火的5'加标签的DNA片段的群体,群体中的每个在靶核酸片段的5'末端上具有第一标签。

[0156] 在一些实施方案中,以上描述的方法使用分离的转座酶和转座子末端成分进行。在其他的实施方案中,以上描述的方法使用包含在转座酶和转座子末端成分之间形成的复合物的转座体组合物进行。

[0157] 在一些特定的实施方案中,本文提供的方法使用从Illumina Inc (San Diego, CA) 可得的Nextera转座体进行,如在US 2010/0120098的公开内容中大体描述的,其内容通过引用以其整体并入本文。

[0158] 转座酶和转座体组合物通常是本领域技术人员已知的,如由US2010/0120098的公开内容所示例的,其内容通过引用以其整体并入本文。在一些实施方案中,本文提供的方法采用由超活性Tn5转座酶和Tn5型转座子末端形成的转座体组合物(Goryshin and Reznikoff, 1998, J. Biol. Chem., 273:7367)。在一些实施方案中,本文提供的方法采用由MuA转座酶和包含R1和R2末端序列的Mu转座子末端形成的转座体组合物(Mizuuchi, 1983, Cell, 35:785; Savilahti等, 1995, EMBO J., 14:4893)。能够将转座子末端以随机的方式或几乎随机的方式以足够的效率插入以将靶核酸加5'标签和片段化用于其意图的目的的任何转座系统可用于本公开内容。示例性转座体组合物系统包括但不限于金黄色葡萄球菌Tn552(Colegio等, 2001, J Bacteriol., 183:2384-8; Kirby等, 2002, MoI Microbiol, 43: 173-86)、TyI(Devine和Boeke, 1994, Nucleic Acids Res., 22:3765-72以及国际专利申请第W095/23875号)、转座子Tn7(Craig, 1996, Science. 271:1512; Craig, 1996, Review in: Curr Top Microbiol Immunol, 204:27-48)、TnI0和IS10(Kleckner等, 1996, Curr Top Microbiol Immunol, 204:49-82)、Mariner转座酶(Lampe等, 1996, EMBO J., 15:5470-9)、Tci(Plasterk, 1996, Curr Top Microbiol Immunol, 204:125-43)、P因子(Gloor, 2004, Methods MoI Biol, 260:97-114)、TnJ(Ichikawa和Ohtsubo, 1990, J Biol Chem. 265: 18829-32)、细菌插入序列(Ohtsubo和Sekine, 1996, Curr.Top.Microbiol.Immunol. 204:1-26)、逆转录病毒(Brown等, 1989, Proc Natl Acad Sci USA, 86:2525-9)和酵母的逆转录转座子(Boeke和Corces, 1989, Annu Rev Microbiol. 43:403-34)。

[0159] 作为非限制性实例,转座子末端可包括19-bp外末端(“OE”)转座子末端、内末端(“IE”)转座子末端、或被野生型或突变Tn5转座酶识别的“镶嵌(mosaic)末端”(“ME”)转座子末端、或如在US 2010/0120098的公开内容中描述的R1和R2转座子末端,其内容通过引用以其整体并入本文。转座子末端可包括适于在体外转座反应中与转座酶或整合酶一起形成功能性复合物的核酸或核酸类似物。例如,转座子末端可包括DNA、RNA、修饰的碱基、非天然碱基、修饰的骨架,并且可在一条或两条链中包含缺口。

[0160] 在一些实施方案中,其中转移链包含3'部分和5'部分,其中3'部分展现转移的转座子末端序列,并且转移链的5'部分展现包含用于特定目的的一个或更多个标签域(例如,

用于下一代测序或扩增的测序标签域或扩增标签域,和任选地寻址标签域)的序列。示例性标签域包括限制性位点标签域、捕获标签域、测序标签域、扩增标签域、检测标签域、寻址标签域和转录启动子域。

[0161] 在一些实施方案中,在体外转座反应中使用两种不同的转座体,并且两种转座体中的每种包含相同的转座酶但不同的转座子末端成分。在一些实施方案中,使用两种不同的转座体,并且两种不同的转座体各自包含相同的转座酶并且转座子末端成分包含不同的转移链。在一些实施方案中,使用两种不同的转座体,并且两种转座体中的每种包含不同的转座酶和不同的转座子末端成分,所述不同的转座子末端成分的每种与各自的转座酶一起形成功能性复合物。

[0162] 在一些实施方案中,在体外转座反应中使用的转座酶和转座子末端成分的量或转座体组合物的量在每50微升反应每50纳克靶核酸约1皮摩尔和约25皮摩尔之间。在一些实施方案中,在体外转座反应中使用的转座酶和转座子末端成分的量或转座体组合物的量在每50微升反应每50纳克靶核酸约5皮摩尔和约50皮摩尔之间。在一些实施方案中,转座酶的浓度为0.5-1nM。在一些实施方案中,转座酶的浓度为0.01-0.02皮摩尔/20 $\mu$ l反应。

[0163] 实施例2例证了用于使用本文提供的方法的标签化步骤的方案。在其中使用单个细胞来制备用于测序的文库的实施方案中,只存在两个拷贝的基因组,并且因此较小的插入尺寸倾向于增加文库多样性。如在实施例8中显示的,计数随着插入尺寸减小而增加,并因此文库代表的多样性随着插入尺寸减小而增加。因此,在一些实施方案中,本文的方法在标签化步骤中使用较高量的转座酶,以增加片段化并减小加标签的核酸片段的插入尺寸。如显示的,当在标签化反应中使用1 $\mu$ l Tn5时,平均片段尺寸为约550bp;而当在标签化反应中使用2 $\mu$ l Tn5时,平均片段尺寸为约400bp。与较小的插入尺寸一致,当用2 $\mu$ l Tn5处理时的文库多样性与用1 $\mu$ l Tn5处理的文库多样性相比增加。使用Tn5来例证转座酶的调整。应领体会到,在本方法中还可使用其他的转座酶,并且它们的量可使用本文提供的方法和本领域技术人员已知的方法来调整和优化。

[0164] 在一些实施方案中,用于体外转座反应的反应时间为2个小时或更少、1小时或更少、30分钟或更少、15分钟或更少、或10分钟或更少。在一些实施方案中,用于体外转座反应的反应时间为5分钟或更少。

[0165] 在一些实施方案中,用于体外转座反应的反应温度为从约40°C至约70°C、从约45°C至约65°C或从约50°C至约60°C。在一些实施方案中,用于体外转座反应的反应温度为约55°C。

[0166] 在一些实施方案中,体外转座反应可通过将例如管中的样品保持在4°C来终止。在一些实施方案中,使标签化缓冲液与标签化产物中和,并使样品在室温孵育持续5分钟。

[0167] 通过体外转座反应,靶核酸片段被在5'末端加标签。在一些实施方案中,本文提供的方法还包括将3'末端标签并入5'加标签的核酸片段以制备加双标签的核酸片段的文库的步骤。在一些实施方案中,加双标签的核酸片段的文库在单个管中从5'加标签的靶核酸生成,而不进行任何中间的纯化步骤。添加3'末端标签可通过多种方法进行,例如如在W0 2010/048605中描述的通过使用DNA聚合酶、末端转移酶和/或连接酶进行,其内容通过引用以其整体并入。

[0168] 因此,在一些实施方案中,本文提供的方法还包括:(d)在其中3'标签连接至5'加

标签的靶核酸片段以生成多个加双标签的靶核酸片段的条件下,使来自步骤(c)的混合体直接与至少一种核酸修饰酶一起孵育。在一些实施方案中,步骤(a)、(b)、(c)和(d)在单个反应管中进行。以下讨论了例证生成加双标签的核酸片段的文库的实施方案。

[0169] 在一些实施方案中,加双标签的核酸片段通过使用具有链置换或5'核酸酶活性的聚合酶例如DNA聚合酶来生成。在一些实施方案中,本文提供的方法包括:在无热循环且其中退火的5'加标签的核酸片段不变性的条件下,使退火的5'加标签的核酸片段的群体与具有链置换或5'核酸酶活性的DNA聚合酶一起孵育,其中DNA聚合酶利用互补链作为模板延伸退火的5'加标签的核酸片段的每条链的3'末端,并置换或消化非转移链,从而生成加双标签的双链DNA片段的文库。在一个实施方案中,延伸步骤在72°C利用相对链上的5'标签作为模板来进行。

[0170] 在一些实施方案中,通过以上提供的方法使生成的加双标签的双链DNA片段变性,以生成包含加双标签的单链DNA片段的加标签的DNA片段的文库(例如,通过加热至95°C并快速冷却)。

[0171] 在其他的实施方案中,加双标签的核酸片段通过使用末端转移酶来生成。在一些实施方案中,使5'加标签的双链核酸片段变性,以生成5'加标签的单链核酸片段。将5'加标签的单链核酸片段与由末端转移酶组成的DNA聚合酶和末端转移酶的至少一种底物一起孵育,在所述孵育期间,末端转移酶将第二标签连接至5'加标签的核酸片段的3'末端,从而生成包含加双标签的核酸片段的加标签的核酸片段的文库。在一些实施方案中,包含转座子末端成分的非转移的转座子末端的3'末端被封闭(例如,通过使用具有双脱氧核苷酸或3'-0-甲基核苷酸作为3'末端核苷酸的非转移的转座子末端),所述封闭将3'核苷酸封闭并阻止通过末端转移酶的添加,从而阻止非转移的转座子末端的背景加标签。

[0172] 在其他的实施方案中,不使5'加标签的双链核酸片段变性生成5'加标签的单链核酸片段。而是,在其中末端转移酶将第二标签连接至5'加标签的核酸片段的3'末端的条件下,将5'加标签的核酸片段与由末端转移酶组成的DNA聚合酶和末端转移酶的至少一种底物一起孵育,持续足以使末端转移酶将第二标签连接至5'加标签的核酸片段的3'末端的时间,而无之前的变性步骤,从而生成加双标签的核酸片段的文库。在一些实施方案中,包含转座子末端成分的非转移的转座子末端的3'末端被封闭(例如,通过使用具有双脱氧核苷酸或3'-0-甲基核苷酸作为3'末端核苷酸的非转移的转座子末端)。

[0173] 在其他的实施方案中,加双标签的核酸片段通过使用DNA聚合酶和末端加标签寡核苷酸来生成。在一些实施方案中,使5'加标签的双链核酸片段变性,以生成5'加标签的单链核酸片段(例如,通过加热至95°C并快速冷却),并使用DNA聚合酶和末端加标签寡核苷酸将第二标签连接至5'加标签的单链核酸片段的3'末端,从而生成加双标签的核酸片段的文库。在一些实施方案中,使用DNA聚合酶和末端加标签寡核苷酸将第二标签连接至5'加标签的核酸片段的3'末端的步骤包括:(1)提供具有5'部分和3'部分的末端加标签寡核苷酸,5'部分展现与期望被连接至5'加标签的单链核酸片段的3'末端的第二标签的序列互补的序列,并且3'部分展现包含3个随机核苷酸和8个随机核苷酸之间的随机序列,所述末端加标签的寡核苷酸的3'末端核苷酸被封闭以使得其不能够被DNA聚合酶延伸;(2)在其中末端加标签寡核苷酸退火至5'加标签的单链核酸片段的条件下,使5'加标签的单链核酸片段与末端加标签核苷酸接触,并持续足以使末端加标签寡核苷酸退火至5'加标签的单链核酸片段

的时间；以及(3)在反应混合物中，并且在其中5'加标签的单链核酸片段的3'末端使用末端加标签寡核苷酸作为模板而被延伸的DNA聚合条件下，使末端加标签寡核苷酸所退火至的5'加标签的单链核酸片段与DNA聚合酶接触，并持续足以使5'加标签的单链核酸片段的3'末端使用末端加标签寡核苷酸作为模板而被延伸的时间，从而使第二标签连接至它们的3'末端，并生成5'和3'加标签的单链核酸片段。

[0174] 在仍然其他的实施方案中，加双标签的核酸片段通过使用模板依赖性连接酶和连接加标签寡核苷酸来生成。在一些实施方案中，在其中第二标签被连接至退火的5'加标签的DNA片段的条件下，使5'加标签的核酸片段与模板依赖性DNA连接酶和具有3'部分和5'部分的连接加标签寡脱氧核苷酸一起孵育，并持续足以使第二标签被连接至退火的5'加标签的DNA片段的时间，其中3'部分展现第二标签，所述第二标签展现被期望连接至5'加标签的DNA片段的3'末端的任何序列，并且5'部分具有5'单磷酸基团并展现随机序列，从而生成包括退火的加双标签的DNA片段的DNA片段的文库。在一些实施方案中，方法还包括使包括退火的加双标签的DNA片段的DNA片段的文库变性(例如，通过加热至95°C并快速冷却)的步骤，从而生成加双标签的单链DNA片段的文库。

[0175] 在生成加标签的核酸片段的文库之后，可利用例如有限循环聚合酶链式反应(PCR)扩增加标签的核酸片段，以引入其他的末端序列或衔接子，例如索引物(index)、通用引物和簇形成和测序所需的其它序列。在一些实施方案中，对5'加标签的核酸片段的文库进行此类扩增。在一些实施方案中，对加双标签的核酸片段的文库进行此类扩增。在一些实施方案中，在生成加标签的核酸片段的文库的同一反应管中进行扩增，并将用于扩增的剂直接加入至同一反应管。

[0176] 因此，本文提供的方法还包括(e)扩增一种或更多种加双标签的靶核酸片段，以生成在加双标签的核酸片段的5'末端和/或3'末端具有另外的序列的加标签的核酸片段的文库。在一些实施方案中，步骤(a)、(b)、(c)、(d)和(e)在单个反应管中进行。示例性扩增方法包括：聚合酶链式反应(PCR)、链置换扩增反应、滚环扩增反应、连接酶链式反应、转录介导的扩增反应或环介导的扩增反应。

[0177] 在一些实施方案中，本文提供的方法包括利用PCR扩增加双标签的单链核酸片段的文库。在一些实施方案中，本文提供的方法利用加标签的DNA片段的文库的单引物PCR扩增。在一些实施方案中，扩增加双标签的DNA片段的步骤包括使用DNA聚合酶和与第二标签互补的至少一种引物。在一些实施方案中，扩增加双标签的DNA片段的文库的步骤包括通过PCR使用展现转移链的至少部分的序列的仅一种寡脱氧核糖核苷酸作为PCR引物并使用加双标签的DNA片段作为模板，来扩增加双标签的DNA片段的文库。在一些实施方案中，引物包含5'部分，所述5'部分包含另外的序列，例如衔接子序列。

[0178] 在一些实施方案中，使用两种不同的PCR引物，每一种所述PCR引物展现包含转座子末端成分的转移的转座子末端的至少一部分的序列。在一些实施方案中，每种PCR引物包含3'部分和5'部分，其中3'部分展现各自的转移的转座子末端序列，并且5'部分展现用于特定目的的各自的标签域或衔接子(例如，用于下一代测序或扩增的测序标签域/衔接子或扩增标签域/衔接子，和任选地寻址标签域/衔接子)的序列。例如，当在体外转座反应中使用单个转座子末端成分以使用具有链置换或5'核酸酶活性的DNA聚合酶生成加双标签的DNA片段的文库时，加双标签的DNA片段可通过PCR使用两种不同的PCR引物来扩增。每种PCR

引物包含3'部分和5'部分,其中3'部分展现各自的转移的转座子末端序列,并且5'部分展现用于特定目的各自的标签域/衔接子(例如,用于下一代测序或扩增的测序标签域/衔接子或扩增标签域/衔接子,和任选地寻址标签域/衔接子)的序列。在一些实施方案中,每种PCR引物的5'部分与另一种引物的5'部分不同,并且因此PCR产物的两个末端的序列不同。例如,一个末端包含一个索引物和/或通用引物序列,并且另一个末端包含不同的索引物和/或通用引物序列。

[0179] 在一些实施方案中,加双标签的核酸片段的两个末端源自两个不同的转移链序列。例如,在一些实施方案中,在体外转座反应中可使用两种不同的转座体,并且两种转座体中的每种包含相同的转座酶但不同的转座子末端成分。在一些实施方案中,使用两种不同的转座体,并且两种不同的转座体各自包含相同的转座酶并且转座子末端成分包含不同的转移链。在一些实施方案中,使用两种不同的转座体,并且两种转座体中的每种包含不同的转座酶和不同的转座子末端成分,所述不同的转座子末端成分的每种与各自的转座酶一起形成功能性复合物。在一些实施方案中,其中在体外转座反应中使用两种不同的转座子末端成分,并且加双标签的单链核酸片段的文库使用具有链置换或5'核酸活性的DNA聚合酶来生成,第一标签展现一种转座子末端成分的转移链的序列并且第二标签展现另一种转座子末端成分的非转移链的序列。

[0180] 在以上提及的实施方案和其中两种不同的转移链被连接至双链核酸的每条相对链的5'末端的其他实施方案中,本文提供的方法还可包括通过PCR使用两种不同的PCR引物扩增加双标签的核酸片段的步骤。PCR引物中的一种展现包含一种转座子末端成分的一种转移链的至少一部分的序列,并且PCR引物中的另一种展现包含另一种转座子末端成分的另一种转移链的至少一部分的序列。

[0181] 在其中使用两种引物的一些实施方案中,每种PCR引物包含3'部分和5'部分,其中3'部分展现各自的转移的转座子末端序列,并且5'部分展现用于特定目的各自的标签域/衔接子(例如,用于下一代测序或扩增的测序标签域或扩增标签域,和任选地寻址标签域)的序列。在一些实施方案中,每种PCR引物的5'部分与另一种引物的5'部分不同,并且因此将不同的序列引入至PCR产物的两个末端。在一些实施方案中,第一PCR引物的5'部分或第二PCR引物的5'部分,或第一PCR引物和第二PCR引物两者的5'部分分别包含第一测序标签/衔接子或第二测序标签/衔接子,用于针对特定的测序平台生成用于下一代测序的模板(例如,用于Illumina Nextera测序平台的测序标签)。在一些实施方案中,第一PCR引物的5'部分或第二PCR引物的5'部分另外包含寻址标签域/衔接子或用于特定目的的另一种标签域/衔接子。

[0182] 实施例3例证了可在加标签的核酸片段的两个末端处添加其他序列的有限循环PCR扩增,所述其他序列例如索引物1(i7)和索引物2(i5)(来自Illumina, Inc, San Diego, CA)以及其他目的例如簇形成所需的序列。在单细胞测序中,输入DNA是相对小的,并因此可调整PCR的循环数以实现更好的测序结果。在实施例9中,使用单个细胞作为起始材料测试并优化了PCR的循环数。如显示的,当在拷贝数分析中使用具有16个循环的PCR时,噪声大,并且当使用具有18个循环或20个循环的PCR时,噪声显著减小。因此,在一些实施方案中,PCR循环数为18、19或20。

[0183] 对于通过本领域技术人员已知的PCR进行扩增反应,很多种酶和试剂盒是可用的。

例如,在一些实施方案中,使用来自EPICENTRE Biotechnologies, Madison, WI的FAILSAFE™PCR系统或MASTERAMP™Extra-Long PCR系统,如制造商描述的进行PCR扩增。但是,本公开内容不限于使用用于扩增反应的那些产品或条件,并且可使用允许在退火至靶序列的引物和退火至转座子的引物之间的序列扩增的任何合适的热稳定DNA聚合酶和反应混合物。

[0184] 本文提供的方法不限于使用PCR来扩增加标签的核酸片段的文库。扩增同一序列并生成用于意图目的的合适成分和量的扩增产物的任何合适的扩增方法(例如,滚环扩增、ribopriimer扩增(例如美国专利第7,413,857号)、ICAN、UCAN、ribospia、末端加标签(美国专利申请第20050153333号)、Eberwine-type aRNA扩增或链置换扩增)可用于本发明的实施方案。例如,可使用的一些链置换方法被描述于Takara Shuzo Company, Kyoto, Japan的PCT专利公布第W0 02/1661号、第W0 00/56877号、和第AU00/29742号;Becton Dickinson and Company的美国专利第5,523,204号、第5,536,649号、第5,624,825号、第5,631,147号、第5,648,211号、第5,733,752号、第5,744,311号、第5,756,702号和第5,916,779号;Nanogen/Becton Dickinson Partnership的美国专利第6,238,868号、第6,309,833号和第6,326,173号;Bio Merieux的美国专利第5,849,547号、第5,874,260号和第6,218,151号;Gen-Probe, Inc.的美国专利第5,786,183号、第6,087,133号和第6,214,587号;Wick等的美国专利第6,063,604号;Kurn的美国专利第6,251,61号;Eiken Kagaku Kabushiki Kaishi, Tokyo, Japan的美国专利第6,410,278号和PCT公布第W0 00/28082号;Auerbach的美国专利第5,591,609号、第5,614,389号、第5,773,733号、第5,834,202号和第6,448,017号;以及Lizardi的美国专利第6,124,120号和第6,280,949号。

[0185] 在一些实施方案中,通过本公开内容的任何方法制备的加标签的核酸片段的文库然后可经历用于纯化文库核酸和任选地用于提供尺寸选择的步骤。这些步骤可有助于清理PCR产物并去除具有不期望的尺寸的核酸。本领域的很多方法可被用于清理在本方法中生成的核酸片段,包括但不限于:利用柱来清理片段,例如使用Qiagen QIAquick PCR纯化试剂盒;以及利用凝胶尺寸选择,例如利用Pippin Prep电泳平台。本领域已知的用于清理核酸片段和/或用于选择核酸尺寸的其他方法也可用于本文提供的方法。

[0186] 例如,在一些实施方法中,AMPure XP珠(来自Beckman Coulter Genomics)被用于纯化加标签的核酸片段。核酸片段可与固相可逆固定(SPRI)珠结合,并且可通过改变PEG/NaCl浓度来控制具有不同长度的核酸片段与珠的亲和力。因此,通过改变PEG/NaCl浓度,具有不同尺寸的核酸可被选择性地纯化。在一些实施方案中,本文提供的方法使用单次AMPure XP处理来去除小于某一尺寸(例如150-200bp)的核酸片段。在一些实施方案中,可通过两次连续的AMPure XP步骤进行双(上和下)尺寸选择。在第一选择步骤中,将低浓度的AMPure XP珠添加至样品以结合较大的DNA片段。在该步骤中,包含较大片段的珠被丢弃。然后,在第二选择步骤中,然后将更多的珠添加至上清液。在该第二步骤中,增加PEG和NaCl的量以使得较小片段尺寸将被结合。接下来,丢弃包含很短的文库片段的上清液,并洗涤珠,并洗脱介于中间的片段。本领域技术人员将理解,根据第一SPRI步骤和最终SPRI步骤中PEG和NaCl的浓度,可生成不同的尺寸范围,如在Bronner等,2009,Curr Protoc Hum Genet.18:10中例证的。

[0187] 用于使用AMPure XP珠清理核酸片段的文库的通常程序包括:(1)涡旋AMPure XP

珠以确保珠被均匀地分散；(2)将某一量的AMPure XP珠添加至生成的每个PCR产物，并在室温孵育；(3)将管放入磁力架上的管座中，直到上清液已清澈；(5)移除并丢弃上清液；(6)洗涤珠一次或多次，而不将管从磁力架移除；(7)使管仍在磁力架上，使珠空气干燥；(8)将管从磁力架移除并添加重悬缓冲液并在室温孵育；以及(9)将上清液转移至新管。

[0188] 在核酸片段的文库被清理并尺寸选择后，可使其进一步经历文库标准化步骤，以使每个文库的量标准化并确保每个混合样品中粗略相等的文库代表。在一些实施方案中，在本文提供的方法中，使用基于珠的文库标准化方法。在基于珠的文库标准化方法中，将粗略相等的量的珠添加至包含核酸片段的样品的每个孔。由于被添加至每个孔的珠的量粗略地相等，因此在每个孔中附接至珠的核酸片段的量也粗略地相等。因此，在移除上清液后，在每个孔中从珠洗脱的核酸片段的量可以粗略相等。

[0189] 通常的基于珠的文库标准化方法包括：(1)将粗略相等的量的珠（例如，在珠缓冲液中的）添加至包含在以上提供的方法中生成的核酸片段的每个孔；(2)孵育和/或震荡以允许珠与核酸片段结合；(3)将孔（可以是板上的）置于磁力架上并允许上清液变得清澈；(4)使孔在磁力架上，小心地移除并丢弃上清液；(5)洗涤珠一次或多次；以及(6)洗脱附接至珠的核酸片段。

[0190] 在一些实施方案中，通过本文提供的方法生成的加标签的核酸片段的文库可被用作用于核酸测序的模板。

[0191] 在一些实施方案中，在测序之前，扩增文库中的加标签的核酸片段，以在测序期间例如在合成测序中相对于噪声增强信号。在一些实施方案中，加标签的核酸片段的文库被用作扩增反应（例如，使用与加标签的核酸片段的末端序列互补的PCR引物的PCR扩增反应）的模板。在一些实施方案中，扩增的加标签的核酸片段的文库包含靶核酸展现的大部分或大约全部的序列。在一些实施方案中，其中靶核酸包含生物体的基因组DNA，扩增反应是全基因组扩增反应。

[0192] 在一些实施方案中，加标签的核酸片段可被固定至固体表面上。例如，固体表面可与互补于加标签的核酸片段的末端序列的多核苷酸附接，并因此加标签的核酸片段可被固定在固体表面上。然后，在表面上扩增固定的核酸片段。例如，在一些实施方案中，固定的核酸片段利用簇扩增方法来扩增，如由美国专利第7,985,565号和第7,115,400号的公开内容所示例的，其每个的内容通过引用以整体并入本文。美国专利第7,985,565号和第7,115,400号的并入的材料描述了固相核酸扩增的方法，其允许扩增产物被固定在固体支持物上，以形成包括固定的核酸分子的簇或“集群”的阵列。此类阵列上的每个簇或集群由多个相同的固定的多核苷酸链和多个相同的固定的互补多核苷酸链形成。如此形成的阵列在本文中通常地被称作“成簇的阵列”。固相扩增反应的产物，诸如在美国专利第7,985,565号和第7,115,400号中描述的那些，是通过固定的多核苷酸链对和固定的互补链的对的退火形成的所谓的“桥式”结构，两种链在5'末端经由例如共价附接被固定在固体支持物上。簇扩增方法是其中固定的核酸模板被用来产生固定的扩增子的方法的实例。本领域已知的其他合适的方法也可被用来从根据本文提供的方法产生的固定的加标签的核酸片段产生固定的扩增子。

[0193] 根据本文提供的方法制备的加标签的核酸片段的文库可根据任何合适的测序方法来测序，诸如直接测序，包括合成测序、连接测序、杂交测序、纳米孔测序等。在一些实施

方案中,在固体支持物上对固定的DNA片段测序。在一些实施方案中,用于测序的固体支持物是扩增于其上发生的固体支持物同一固体支持物。

[0194] 在一些实施方案中,在本文提供的方法中使用的测序方法是合成测序(SBS)。在SBS中,监测核酸引物沿着核酸模板(例如,靶核酸或其扩增子)的延伸,以确定模板中的核苷酸的序列。潜在的化学方法可以是聚合作用(例如,如被聚合酶催化的)。在特定的基于聚合酶的SBS实施方案中,荧光标记的核苷酸以模板依赖性方式被添加至引物(从而延伸引物),以使得检测被添加至引物的核苷酸的顺序和类型可被用来确定模板的序列。

[0195] 使用循环反应的其他测序程序可被使用,诸如焦磷酸测序。焦磷酸测序随着特定核苷酸被并入至新生的核酸链,检测无机焦磷酸(PPi)的释放(Ronaghi等,1996,Analytical Biochemistry 242(1),84-9;Ronaghi,2001,Genome Res.11(1),3-11;Ronaghi等,1998,Science 281(5375),363;US 6,210,891;US 6,258,568和US 6,274,320,其每个通过引用并入本文)。在焦磷酸测序中,释放的PPi可通过立即被ATP硫酸化酶转化成腺苷三磷酸(ATP)来检测,并且产生的ATP的水平可经由荧光素酶产生的质子来检测。因此,测序反应可经由发光检测系统来监测。用于基于荧光的检测系统的激发放射源对于焦磷酸测序程序不是必需的。有用的流体系统、检测器以及可适用于对根据本公开内容产生的扩增子应用焦磷酸测序的程序被描述于例如WIPO专利申请系列第PCT/US11/57111号、US 2005/0191698 A1、US 7,595,883和US 7,244,559,其每个通过引用并入本文。

[0196] 一些实施方案可利用包括DNA聚合酶活性的实时监测的方法。例如,核苷酸并入可通过载有荧光团的聚合酶和 $\gamma$ -磷酸标记的核苷酸之间的荧光共振能量转移(FRET)相互作用,或用零模波导(ZMW)来检测。用于基于FRET的测序的技术和试剂被描述于例如Levene等2003,Science 299,682-686;Lundquist等,2008,Opt.Lett.33,1026-1028;Korlach等,2008,Proc.Natl.Acad.Sci.USA 105,1176-1181,其公开内容通过引用并入本文。

[0197] 一些SBS实施方案包括检测将核苷酸并入延伸产物后释放的质子。例如,基于检测释放的质子的测序可使用电子检测器和从Ion Torrent(Guilford,CT,a Life Technologies subsidiary)商购可得的相关技术,或在US 2009/0026082 A1、US 2009/0127589 A1、US 2010/0137143 A1或US 2010/0282617 A1中描述的测序方法和系统,其每个通过引用并入本文。本文描述的用于利用动力学排除来扩增靶核酸的方法可容易地应用于被用来检测质子的底物。更具体地,本文描述的方法可被用来产生用于检测质子的扩增子的克隆群体。

[0198] 另一种有用的测序技术是纳米孔测序(参见例如Deamer等,2000,Trends Biotechnol.,18,147-151;Deamer等,2002,Acc.Chem.Res.35:817-825;Li等,2003,Nat.Mater.2:611-615),其公开内容通过引用并入本文)。在一些纳米孔实施方案中,靶核酸或从靶核酸移除的个体核苷酸穿过纳米孔。随着核酸或核苷酸通过纳米孔,每个核苷酸类型可通过测量孔的电导率波动来鉴定。(美国专利第7,001,792号;Soni等,2007,Clin.Chem.,53,1996-200;Healy,2007,Nanomed.2,459-481;Cockroft等,2008,J.Am.Chem.Soc.,130,818-820,其公开内容通过引用并入本文)。

[0199] 在一些实施方案中,本文提供的方法还包括分析细胞的拷贝数变异。拷贝数分析测试样品中的DNA拷贝数变异。此类分析有助于检测染色体拷贝数变异,所述染色体拷贝数变异可导致或可增加多种严重紊乱的风险。例如,已报道孤独症与拷贝数突变有关(Sebat

等,2007,Strong association of de novo copy number mutations with autism, Science 316 (5823) :445-9。还已报道,精神分裂症与拷贝数变异有关(St Clair,2008, Copy number variation and schizophrenia,Schizophr Bull 35(1):9-12)。已开发出多种方法用于检测拷贝数变异。但是,当起始材料有限并来自最小细胞群体时,噪声显著并且结果被打折扣。本方法提供了用于在此类情况中检测拷贝数变异的方法。以下提供的实施例展示了使用本方法的拷贝数变异分析,并且对于拷贝数变异分析,若干参数被优化。在一些实施方案中,在拷贝数变异分析中使用的最小细胞群体包含一个、两个、三个、四个或五个细胞。通常,随着细胞数增加,可实现更完整的读段分布,并因此在数据中存在更少的噪声,如实施例10中显示的。在该实例中,该实例中分析了使用一个、三个或五个细胞的读段分布。如显示的,基因组覆盖率随着细胞数增加而增加,经评估,一个细胞可覆盖基因组的约40%,并且三个细胞可覆盖基因组的50%以上,并且五个细胞可覆盖基因组的约60%。使用一个细胞、三个细胞和五个细胞的平均文库计数分别为约5百万、1千5百万和2千万。在该实例中还显示,当使用单个细胞时,整体成功率相对高94% (N=187)。一个细胞测定的失败可能是由细胞自身的质量,例如选择一个正在复制的细胞或凋亡/坏死的细胞而造成。

[0200] 实施例11将本方法与目前的一些单细胞制备方法比较。当使用由QIAGEN (San Diego, CA) 开发的REPLI-g单细胞试剂盒来制备核酸时,拷贝数变异数据当源自单个细胞、三个细胞或五个细胞时很噪杂。当使用由Illumina, Inc (San Diego, CA) 开发的SurePlex (PicoPlex) 来制备核酸时,它与REPLI-g单细胞试剂盒相比减少噪声。如显示的,本方法 (Nextera SC) 与使用SurePlex扩增系统相比进一步减少噪声。因此,本方法提供了用于分析拷贝数变异的先进方法。

[0201] 拷贝数变异分析的一个方面是要检测镶嵌现象 (mosicism)。镶嵌体或镶嵌现象表明了在一个个体中两种或更多种基因型的存在。存在两大类镶嵌现象:体细胞镶嵌现象和种系镶嵌现象。当体细胞例如由于在第一次分裂或之后的分裂时的有丝分裂错误而包含多于一个基因型时,体细胞镶嵌现象发生。研究人员已显示,存在的体细胞突变贯穿一生增加,并且是造成很多白血病、淋巴瘤和实体肿瘤的原因 (Jacobs等,2012, Detectable Clonal Mosaicism and Its Relationship to Aging and Cancer, Nature Genetics 44 (6) :651-U668)。在种系镶嵌现象中,一些配子(精子或卵母细胞)携带突变但其他配子正常,这同样导致很多疾病。因此,检测镶嵌现象可提供有价值的诊断信息。本公开内容提供了用于检测镶嵌现象的方法。在实施例12中,示例了使用本文提供的方法检测镶嵌现象。如显示的,在单个GM50121细胞的18号染色体的拷贝数分析中,在每次单细胞测序中检测了代表15.4MB DNA的群体。相似地,使用单个GM20916的15号染色体、X染色体和10号染色体的拷贝数分析数据,以及使用单个GM1021细胞的1号染色体和11号染色体的拷贝数分析数据二者均检测了代表其他染色体的另外的群体。

[0202] 本方法还可被用于其他应用,例如移植前基因筛选、单细胞研究、循环肿瘤细胞的分析、细针抽吸活检、血沉棕黄层和羊水细胞的分析。在这些应用中,起始核酸材料通常是有限的,并因此本方法可改进这些应用的分析。除了拷贝数变异分析之外,本方法还可在以上提及的应用中被用于检测存在于最小细胞群体中的单核苷酸变体 (single nucleotide variant)。单核苷酸变体包括单核苷酸多态性 (SNP) 和点突变。单核苷酸多态性 (SNP) 是常见类型的遗传变异,其包括DNA位置中的多态性,在所述DNA位置处在人口群体中两种或更

多种替代碱基以可观的频率存在(通常大于或等于1%)。点突变为频率小于1%的碱基变异。单核苷酸多态性(SNP)和点突变代表了在人基因组中多样性的最大来源。这些单核苷酸多态性(SNP)和点突变可充当生物标志物用于将疾病定位到人基因组图谱上,因为它们通常位于靠近与某种疾病有关的基因处。因此,检测单核苷酸多态性(SNP)、点突变和相似突变对于临床活动性、人类健康和控制遗传疾病是很重要的。本方法提供了均一地接近基因组DNA的优势,并有助于保持靶核酸材料。因此,它可改进使用最小细胞群体的单核苷酸变异检测。

[0203] 在以上多种方法的一些实施方案的说明中,使用“反应管”或“管”。应领会,在本方法中还可使用其他的反应介质和/或容器。

[0204] 用于制备加标签的核酸片段的文库的试剂盒

[0205] 在另一方面,本公开内容提供了用于制备加标签的核酸片段的文库的试剂盒,所述试剂盒包括:(a)裂解试剂,所述裂解试剂具有一种或更多种蛋白酶,和(b)转座反应组合物,所述转座反应组合物具有至少一种转座酶和包含转移链的至少一种转座子末端成分。

[0206] 在一些实施方案中,提供的裂解试剂包含拥有广谱特异性的仅一种蛋白酶,并因此蛋白酶可消化多种蛋白和多肽。在一些其他的实施方案中,本文提供的裂解试剂包含多种蛋白酶的混合物,并且多种蛋白酶的组合可消化多种蛋白和多肽。本文提供的示例性蛋白酶包括丝氨酸蛋白酶、苏氨酸蛋白酶、半胱氨酸蛋白酶、天冬氨酸蛋白酶、谷氨酸蛋白酶和金属蛋白酶。本文使用的示例性蛋白酶包括从重组杆菌菌株分离的丝氨酸蛋白酶。本文使用的示例性蛋白酶包括枯草杆菌蛋白酶及其变体,包括枯草杆菌蛋白酶Carlsberg、alcalase和枯草杆菌蛋白酶S41。枯草杆菌蛋白酶及其变体是本领域技术人员已知的,并且包括例如alcalase、alcalase 0.6L、alcalase 2.5L、ALK酶、杆菌肽酶A、杆菌肽酶B、枯草杆菌碱性蛋白酶bioprase、bioprase AL 15、bioprase APL 30、colistinase、枯草杆菌蛋白酶J、枯草杆菌蛋白酶S41、枯草杆菌蛋白酶Sendai、枯草杆菌蛋白酶GX、枯草杆菌蛋白酶E、枯草杆菌蛋白酶BL、genenase I、esperase、maxatase、耐热蛋白酶PC 10、蛋白酶XXVII、耐热酶、superase、枯草杆菌蛋白酶Carlsberg、枯草杆菌蛋白酶DY、枯草杆菌肽酶、SP 266、savinase 8.0L、savinase 4.0T、kazusase、蛋白酶VIII、opticlean、肌球吸附蛋白A 3L、savinase、savinase 16.0L、savinase 32.0 L EX、orientase 10B、蛋白酶S、丝氨酸内肽酶。在本文展示的方法和组合物的特定实施方案中,可使用热不稳定蛋白酶诸如枯草杆菌蛋白酶和枯草杆菌蛋白酶的热不稳定性变体,如由Davaill等,1994,J.Biol.Chem.,26:17448-17453的示例性公开内容展示的,其通过引用以其整体并入本文。

[0207] 在一些实施方案中,裂解试剂包含一种或更多种去垢剂。在一些实施方案中,本文提供的去垢剂不干扰下游的酶活性。因此,在一些实施方案中,裂解试剂包含非离子型去垢剂。通常,非离子型去垢剂包含不带电荷、亲水性头基。通常非离子型去垢剂基于聚氧乙烯或糖苷。示例性非离子型去垢剂包括Tween® 80、Tween® 20、Triton® X-100、Triton® X-100-R、Triton® X-114、NP-40、Genapol® C-100、Genapol® X-100、Igepal® CA 630、Arlasolve® 200、Brij® 96/97、Triton、Brij® 98、Brij® 58、Brij® 35、Brij系列、Pluronic® L64、Pluronic® P84、非去垢剂碘基甜菜碱(NDSB 201)、amphipols(PMAL-C8)、CHAPS、辛基β-D-吡喃葡萄糖苷、皂苷、单十二烷基九乙二醇(C12E9,聚多卡醇)醚、十二烷基

硫酸钠、N-月桂基肌氨酸、脱氧胆酸钠、胆汁盐、十六烷基三甲基溴化铵、SB3-10、SB3-12、脒基磺基甜菜碱-14、辛基硫代葡萄糖苷、麦芽糖苷、HEGA和MEGA系列。在一个实施方案中，裂解试剂包括表1-3中提供的组分。

[0208] 在一些实施方案中，转座组合物包含至少一种转座酶和至少一种转座子末端成分，所述至少一种转座子末端成分包含：(i) 转移链，所述转移链具有3'部分和5'部分，所述3'部分展现转移的转座子末端序列，所述5'部分展现用于下一代测序或扩增反应的标签域的序列；以及(ii) 包含5'磷酸的非转移链，所述包含5'磷酸的非转移链仅展现非转移的转座子末端序列，其中转座酶与在体外转座反应中有活性的转座子末端成分一起形成复合物。在一些实施方案中，试剂盒还包括反应缓冲液，所述反应缓冲液以导致二甲基甲酰胺以10%的终浓度存在于体外转座反应中的量包含二甲基甲酰胺。在一些实施方案中，标签域包含限制性位点域、捕获标签域、测序标签域、扩增标签域、检测标签域和寻址标签域中的一个或更多个。

[0209] 在一些实施方案中，转座反应组合物包含两种或更多种转座子末端成分，两种或更多种转座子末端成分的每种包含以至少一个核苷酸不同的转移链。

[0210] 在一些实施方案中，转座酶为Tn5转座酶。在一些实施方案中，转座子末端成分包括Tn5转座子末端。在试剂盒的一个实施方案中，转座体包含以如下浓度被提供的野生型Tn5转座酶或超活性Tn5转座酶或MuA转座酶：其中在体外转座反应中转座体的终浓度为至少250nM。在一些其他的实施方案中，野生型Tn5转座体或超活性Tn5转座体或MuA转座体的终浓度为至少500nM。

[0211] 在一个实施方案中，试剂盒中的转座酶为大于或等于约5单位/微升、约10-20单位/微升、约20-40单位/微升、约40-60单位/微升、约60-80单位/微升、或约80-100单位/微升的浓度的野生型或突变形式的Tn5转座酶（例如EZ-Tn5<sup>TM</sup>转座酶）。在一些实施方案中，本文提供的试剂盒包含表6中提供的组分。

[0212] 在一些实施方案中，试剂盒另外包含修饰酶。在一些实施方案中，修饰酶为聚合酶或连接酶。在一些实施方案中，试剂盒包含选自以下的至少一种其他的酶组分：具有5'核酸酶或链置换活性的DNA聚合酶、缺乏5'核酸酶活性的DNA聚合酶、模板依赖性NAD连接酶和非模板依赖性连接酶。在一些实施方案中，至少一种其他的酶组分选自以下：FAILSAFE<sup>TM</sup>DNA聚合酶混合物、Taq DNA聚合酶、TfI DNA聚合酶、T4 DNA聚合酶、大肠杆菌DNA连接酶、噬菌体TS2126热稳定RNA连接酶、Mth Rn 1热稳定RNA连接酶、和CIRCLIGASE<sup>TM</sup>热稳定ssDNA连接酶。

[0213] 在其中试剂盒中的至少一种酶为模板依赖性连接酶（例如，大肠杆菌DNA连接酶）的一些实施方案中，高比例的连接酶分子被腺苷酸化并且在试剂盒中不提供ATP。在其中试剂盒中的至少一种酶是模板依赖性连接酶（例如大肠杆菌DNA连接酶）的一些实施方案中，试剂盒另外包含连接加标签寡核苷酸，所述连接加标签寡核苷酸包含3'部分和5'部分，其中所述3'部分展现标签域的序列并且5'部分展现由约3个核苷酸至约8个核苷酸组成的随机序列。在一些实施方案中，连接加标签寡核苷酸包含5'部分，所述5'部分展现由4个核苷酸组成的随机序列。

[0214] 在其中试剂盒中的至少一种酶是非模板依赖性连接酶的一些实施方案中，所述非模板依赖性连接酶选自噬菌体TS2126热稳定RNA连接酶、Mth Rn 1热稳定RNA连接酶和CIRCLIGASE<sup>TM</sup>热稳定ssDNA连接酶，非模板依赖性连接酶以高度腺苷酸化的形式提供并且在

试剂盒中不提供ATP。在一个实施方案中，试剂盒包含EZ-Tn5<sup>TM</sup>转座酶和非模板依赖性核酸连接酶，EZ-Tn5 pMEDS转座子末端成分包含具有5'单磷酸基团的EZ-Tn5METS转移链和具有5'单磷酸基团的EZ-Tn5 pMENTS非转移链两者。

[0215] 在一些实施方案中，试剂盒还包括用于扩增反应的试剂。在一些实施方案中，用于扩增反应的试剂为用于PCR的试剂。在一些实施方案中，用于扩增反应的试剂包括至少一种引物。在一些实施方案中，至少一种引物包含3'部分，所述3'部分展现转移链的至少一部分的序列。在一些实施方案中，至少一种引物包含5'部分，所述5'部分包含通用序列。

[0216] 在一些实施方案中，试剂盒包括两种引物，每种PCR引物包含3'部分和5'部分，其中3'部分展现各自的转移的转座子末端序列，并且5'部分展现用于特定目的的各自的标签域/衔接子(例如，用于下一代测序或扩增的测序标签域或扩增标签域，和任选地寻址标签域)的序列。在一些实施方案中，每种PCR引物的5'部分与另一种引物的5'部分不同。在一些实施方案中，第一PCR引物的5'部分或第二PCR引物的5'部分，或第一PCR引物和第二PCR引物两者的5'部分分别包含第一测序标签/衔接子或第二测序标签/衔接子。在一个实施方案中，本文提供的试剂盒包含表7中提供的组分。

[0217] 在一些实施方案中，试剂盒还包括尺寸选择试剂。在一些实施方案中，尺寸选择试剂包括AMPure XP珠(来自Beckman Coulter Genomics)。核酸片段可与固相可逆固定(SPRI)珠结合。在一些实施方案中，尺寸选择试剂还包括PEG和NaCl。

[0218] 在一些实施方案中，本文提供的试剂盒还包括文库标准化试剂。在一些实施方案中，文库标准化试剂包括由Illumina, Inc (San Diego, CA, 产品编号1502511)提供的文库标准化添加物(Library Normalization Additives)和由Illumina, Inc (产品编号15022566)提供的文库标准化珠(Library Normalization Beads)。在一些实施方案中，文库标准化试剂还包括由Illumina, Inc (产品编号15022565)提供的文库标准化洗涤(Library Normalization Wash)。在一些实施方案中，文库标准化试剂还包括由Illumina, Inc (San Diego, CA, 产品编号1502511)提供的文库标准化储存缓冲液。

[0219] 在一些实施方案中，试剂盒还包括具有固体表面的装置。在一些实施方案中，固体表面与寡核苷酸的群体附接。在一些实施方案中，装置是流式细胞装置。在一些实施方案中，固体表面包括适于以有序的模式固定分子的模式化表面。

[0220] 从前面的描述将明显的是，可对本文描述的发明进行变化和改变，以使其适应多种用途和条件。此类实施方案也在以下权利要求书的范围内。

[0221] 在本文的任何的变量的定义中一列元素的引述包括做为所列元素的任何单个元素或所列元素的组合(或子组合)的该变量的定义。本文的实施方案的引述包括作为任何单个实施方案的实施方案或与任何其他实施方案组合的实施方案或其部分。

[0222] 在该说明书中提到的所有专利和出版物通过引用并入本文，至如同每个单独专利和出版物被特别地和单独地指明通过引用并入的相同程度。

[0223] 以下实施例通过例证的方式而非限制性的方式来提供。

[0224] 实施例

[0225] 实施例1生成包含靶核酸的细胞裂解物

[0226] 在一些实施方案中，在生成细胞裂解物的步骤期间，细胞膜被去垢剂破坏，期间蛋白-脂质缔合和脂质-脂质缔合被破坏，并从而释放可溶形式的细胞内物质。广谱特异性的

蛋白酶的主要功能是去除DNA结合蛋白,诸如来自DNA的组蛋白,以允许转座酶均一地接近DNA。在一些实施方案中,如该实施例中例证的,在单个裂解试剂混合物中提供去垢剂和蛋白酶。将混合物直接应用于细胞,用于生成包含靶核酸的细胞裂解物。如以上讨论的,在一些实施方案中,当利用热使蛋白酶失活时,热不使双链核酸变性和确保标签化步骤不被干扰是重要的。

[0227] 在该实施例中,可使蛋白酶在70°C热失活,并且在该温度,DNA的双链构象被保持。用于生成包含靶核酸的细胞裂解物的方案在实施例1中被例证如下:

[0228] (1)通过温和地颠倒和轻弹管3-5次来充分混合试剂,然后在微型离心机中短暂旋转。

[0229] (2)在干净的微型离心管中,混合表1中的组分以制备5X裂解总混合物。可根据样品数按比例扩大裂解总混合物,例如可包括额外10%以补偿移液期间的损失。

[0230] 表1裂解总混合物的组分

[0231]	裂解总混合物的组分	体积(μl)
	5X裂解缓冲液	1.1
	5X蛋白酶贮存溶液	1.1
	总计	2.2

[0232] 以上表1中的5X裂解缓冲液可根据以下表2来制备:

[0233] 表2 5X裂解缓冲液的组分

[0234]	组分	贮存浓度	5X 总混合物浓度	体积(μl)
	Tris-HCl (pH 8.0)	1 M	250 mM	250
	EDTA	0.5 M	5 mM	10
	Triton X-100	10 %	2.5 %	250
	Super Q H <sub>2</sub> O			490
	总计			1000

[0235] 可通过温和地将管涡旋数次使所有试剂充分混合,然后在微型离心机中短暂旋转。该步骤可重复3-5次。可在室温贮存5X裂解缓冲液,以防止去垢剂沉淀。

[0236] 5X蛋白酶贮存溶液可如下制备:(1)通过直接在玻璃小瓶中重悬蛋白酶例如QIAGEN蛋白酶来制备单次使用贮存等分试样,所述重悬通过添加2.38ml Super Q H<sub>2</sub>O至3150mAU/ml的终浓度进行。通过温和地涡旋小瓶数次,确保蛋白酶被充分溶解。将溶液等分成25μl等分试样并立即在-80°C冷冻,并(ii)将单次使用贮存等分试样从冷冻仪移除并解冻,并根据以下表3制备5X蛋白酶贮存溶液:

[0237] 表3 5X蛋白酶贮存溶液的组分

组分	贮存浓度	5X 总混合物浓度	体积(μl)
QIAGEN 蛋白酶	3150 mA/ml	450 mA/ml	15
Super Q H <sub>2</sub> O			90
总计			105

[0239] 相应地,5X蛋白酶贮存溶液的终浓度为450mA/ml。

[0240] (3) 将2μl以上制备的裂解总混合物添加至包含细胞、阳性对照基因组DNA或阴性对照的每个管。根据以下程序在热循环仪中孵育样品:50°C30min、70°C20min和4°C保持。

[0241] 在一些实施方案中,在每个实验中包括阳性对照基因组DNA(约30pg)。可从如在以下表4和5中制备的10ng/μl贮存溶液以两步系列稀释来制备阳性对照基因组DNA:

[0242] 表4中间基因组DNA稀释物的组分

组分	贮存浓度	中间浓度	体积(μl)
DNA	10ng/μl	100pg/μl	2
1X RS1			198
总计			200

[0244] 然后,可根据以下表5随后稀释根据上表制备的中间DNA稀释物:

[0245] 表5最终基因组DNA稀释物的组分

组分	贮存浓度	中间浓度	体积(μl)
DNA	100pg/μl	10pg/μl	10
1X PBS			90
总计			100

[0247] 可将3μl在以上表中制备的最终稀释物用作阳性对照基因组DNA的输入。这对应于30pg或5个细胞的基因组当量。根据本文提供的方法,还可使用更多或更少的基因组DNA。

[0248] 实施例2在细胞裂解物中直接片段化靶核酸

[0249] 在一些实施方案中,细胞裂解物(例如,如在实施例1中制备的)中的基因组DNA可通过Nextera转座体(从Illumina, Inc, San Diego, CA可得)来标签化(加标签和片段化)。Nextera转座体可同时片段化输入DNA并将标签/衔接子序列添加至末端。可将标签化总混合物直接添加至在实施例1中制备的细胞裂解物而无任何先前的DNA纯化或扩增步骤。可如下表6中显示的制备标签化总混合物,并且可根据样品数扩大总混合物,例如扩大额外10%以补偿移液期间的损失。

[0250] 表6标签化总混合物的组分

组分	体积(μl)
标签化DNA缓冲液	11
Nextera扩增子标签化混合物	2.2

Super Q H <sub>2</sub> O	3.3
总计	16.5

[0252] 标签化DNA缓冲液和Nextera扩增子标签化混合物从Illumina, Inc (San Diego, CA; 产品编号15027866和15031561) 可得。标签化DNA缓冲液包含Tris(羟甲基)氨基甲烷、MgCl<sub>2</sub>和二甲基甲酰胺。Nextera扩增子标签化混合物包含转座体酶。然后可将15μl标签化总混合物添加至从例如实施例1生成的每个细胞裂解物，并在55°C与细胞裂解物一起孵育持续5min，并且然后在4°C终止反应。然后可将包含SDS的中和标签化缓冲液(从Illumina, Inc, San Diego, CA可得)添加至管并在室温孵育持续5分钟。

[0253] 实施例3有限循环PCR扩增(Limited-Cycle PCR Amplification)

[0254] 标签化的DNA片段(例如，如在实施例2中制备的)可通过有限循环PCR程序来扩增。该PCR步骤还可在加标签的核酸片段的两个末端处添加其他的序列，例如索引物1(i7)和索引物2(i5)(从Illumina, Inc, San Diego, CA可得)以及其他目的例如簇形成所需的序列。例如，可将以下表7中的组分(从Illumina, Inc, San Diego, CA可得)添加至从实施例2产生的中和的标签化产物。

[0255] 表7用于有限循环PCR的组分

组分	体积(μl)
PCR总混合物	15
索引物1引物(P5引物)	5
索引物2引物(P7引物)	5

[0257] 表7中的PCR总混合物可如以下表8来制备：

[0258] 表8PCR总混合物的组分

组分	贮存浓度	总混合物浓度	体积(μl)
KAPA HiFi保真缓冲液	5X	3.33X	999
dNTP混合物	各25mM	各1.00mM	59.94
KAPA HiFi DNA聚合酶	1U/μl	0.033U/μl	49.95
Super Q H <sub>2</sub> O			11.11
总计			1500

[0260] 示例性PCR程序如下：72°C3min、98°C30sec, 和然后20个循环的98°C10秒、60°C30秒、和72°C30秒，以及最后使样品保持在4°C。

[0261] 实施例4蛋白酶活性对于均一地接近DNA是有用的

[0262] 在该实施例中分析了蛋白酶活性对均一地接近DNA的影响。特别地，使用0mg/ml、0.1mg/ml(4.5mAU/ml)、0.5mg/ml(22.5mAU/ml)或2.5mg/ml(112.5mAU/ml)蛋白酶来处理全细胞和细胞核。对每次测序分析独特映射的读段的百分比。图1是显示在使用0mg/ml、0.1mg/ml、0.5mg/ml或2.5mg/ml蛋白酶处理的全细胞或细胞核的测序中，独特映射的读段的百分比的柱状图。如显示的，独特映射的读段随着蛋白酶浓度增加而增加，并且使用全细胞和细胞核二者作为起始材料，这均是真实的。还注意到，使用0.5mg/ml蛋白酶的独特映射的读段的百分比与使用2.5mg/ml的独特映射的读段的百分比相似。

[0263] 还通过在使用具有Nextera XT文库制备物的总基因组DNA对照、使用具有足够的蛋白酶活性的单个细胞、和使用具有不足的蛋白酶活性的单个细胞之间比较计数和拷贝数

分析结果来分析蛋白酶活性对均一地接近DNA的影响。图2显示了使用总DNA、用足够的蛋白酶活性处理的单个细胞和用不足的蛋白酶活性处理的单个细胞的计数和拷贝数分析结果的柱状图。如显示的,当以目前的Nextera XT文库制备方法使用相对大的量的基因组DNA时,如图2的上图中显示的,可以以不显著的噪声得到相对清晰的拷贝数分析结果。当仅单个细胞被用于测序时,噪声是显著的,并且拷贝数分析数据显示出离散的分布模式,如图2的下图中显示的。出乎意料地,当用足够的蛋白酶(0.5mg/ml)处理单个细胞时,拷贝数分析结果恢复至与使用总基因组DNA的拷贝数分析结果相当,以不显著的噪声显示出清晰的数据,如图2的中图中显示的。这表明蛋白酶可增加转座酶对基因组DNA的可接近性,因为DNA结合蛋白可被均一地去除。

[0264] 这些结果显示,蛋白酶活性对于在测序中均一地接近DNA是有用的。

[0265] 实施例5优化蛋白酶浓度

[0266] 在该实施例中,分析了在本方法中使用的蛋白酶的浓度。图3A显示了用0.5mg/ml活性蛋白酶、2mg/ml活性蛋白酶或2mg/ml HI蛋白酶处理的单个细胞中的拷贝数分析结果的柱状图。如显示的,当用0.5mg/ml或2mg/ml活性蛋白酶处理单个细胞时,相似地得到清晰的拷贝数分析结果,如图3A的上两个柱状图中显示的。相反,当以在70°C预先热失活的蛋白酶进行反应时,未能得到清晰的拷贝数结果,如图3A的底部的柱状图中显示的。该结果显示,0.5mg/ml或2mg/ml两种浓度的蛋白酶均是有效且充分的。

[0267] 在用0.5mg/ml活性蛋白酶、1mg/ml活性蛋白酶、2mg/ml活性蛋白酶、或2mg/ml预先热失活的(70°C)蛋白酶处理的单个细胞的测序中,还分析了独特映射的读段的百分比。图3B显示了在用0.5mg/ml活性蛋白酶、1mg/ml活性蛋白酶、2mg/ml活性蛋白酶、或2mg/ml预先热失活的蛋白酶处理的单个细胞和无细胞的对照样品的测序中,独特映射的读段的百分比的柱状图。如显示的,在使用以0.5mg/ml活性蛋白酶、1mg/ml活性蛋白酶、和2mg/ml活性蛋白酶处理的单个细胞的测序中独特映射的读段的百分比全部为约65%,具有小的变异。相反,当蛋白酶在70°C失活时,即使使用较高的量的蛋白酶,独特映射的读段的百分比低得多,具有巨大的变异。

[0268] 另外,通过分析邻近箱计数之间的计数差异,分析了拷贝数数据中的噪声。图3C显示了在用0.5mg/ml活性蛋白酶、1mg/ml活性蛋白酶、2mg/ml活性蛋白酶、或2mg/ml预先热失活的蛋白酶处理的单个细胞和无细胞的对照样品的测序中,邻近箱之间的读段计数差异(邻近箱之间的读段计数差异的四分位数范围)的柱状图。如显示的,在使用以0.5mg/ml活性蛋白酶、1mg/ml活性蛋白酶、和2mg/ml活性蛋白酶处理的单个细胞的测序中,邻近箱计数之间的计数差异全部是相对小的(约20%),具有小的变异。相反,当蛋白酶在70°C失活时,即使使用较高的量的蛋白酶(2mg/ml),邻近箱计数之间的计数差异大得多,具有巨大的变异。

[0269] 总而言之,这些结果显示,在本文提供的方法中,具有从0.5mg/ml到2.0mg/ml(22.5mAU/ml到90mAU/ml)的浓度范围的蛋白酶是足够且有效的。

[0270] 实施例6优化蛋白酶消化反应的PH条件

[0271] 在该实施例中,蛋白酶消化反应的pH条件被优化,平衡了蛋白酶活性和测序结果。

[0272] 在不同的pH条件下分析了蛋白酶活性。结果被示于图4A中。图4A是显示在pH 7.0、pH 7.5、pH 8.0、pH 8.5、pH 9.0或pH 10.0蛋白酶的相对活性(相对于在pH 8.0的蛋白酶活

性)的柱状图。如显示的,蛋白酶的活性随着pH值增加而增加,蛋白酶在pH7.0具有最低的活性,并且在pH 10.0具有最高的活性。

[0273] 然后,在不同pH条件下分析独特映射的读段的百分比。图4B显示了在用蛋白酶在pH 7.0、pH 8.0、pH 9.0或pH 10.0处理的单个细胞的测序中,独特映射的读段的百分比的柱状图。如显示的,当pH为7.8或9时,可得到约70%的清晰的独特映射的读段。但是,当pH为10时,可得到较小百分比的独特映射的读段,并且数据变异显著增加。

[0274] 还通过比较邻近箱之间的计数差异分析了拷贝数数据中的噪声。图4C显示了在用0.5mg/ml蛋白酶在pH 7.0、pH 8.0、pH 9.0或pH 10.0处理的单个细胞的测序中,邻近箱之间的读段计数差异(邻近箱之间的读段计数差异的四分位数范围)的柱状图。如显示的,与独特映射的读段结果一致,相邻箱之间的计数差异是相对小的(约20%),具有小的变异;而在pH 10.0,相邻箱之间的计数差异显著增加,具有巨大的变异。

[0275] 在一些实施方案中,消化反应的pH值在pH 7.0至pH 9.0之间。

#### [0276] 实施例7测试蛋白酶的热失活

[0277] 在一些实施方案中,本文提供的蛋白酶可以是热失活的。如以上讨论的,在制备性实施方案中,蛋白酶可在相对低的温度(例如70°C)失活,以使得双链DNA构象可被保持用于标签化反应。在该实施例中,针对热失活以及其对测序结果的影响分析蛋白酶(来自QIAGEN)。

[0278] 在不同的温度预热蛋白酶,并测试蛋白酶的活性。结果被显示于图5A,其显示了当在室温、50°C、60°C或70°C预热时,相对蛋白酶活性的柱状图。如显示的,蛋白酶活性随着温度增加而逐渐降低,并在70°C完全失活。该结果与以上实施例5中显示的结果一致。

[0279] 分析了在不同温度单个细胞、三个细胞、和15pg基因组DNA的测序中独特映射的读段的百分比。图5B显示了在用2.0mg/ml蛋白酶在室温、50°C、60°C或70°C处理单个细胞、三个细胞或15pg基因组DNA的测序中,独特映射的读段的百分比的柱状图。如显示的,独特映射的读段的百分比随着温度增加而降低。但是,因为在该实施例中使用相对较高的蛋白酶浓度(2.0mg/ml),所以在70°C降低的蛋白酶活性更有耐受性。因此,在70°C的独特映射的读段的百分比仍相对高,尽管比在较低温度处理的那些独特映射的读段的百分比低。

[0280] 还分析了在不同温度单个细胞、三个细胞、和15pg基因组DNA的测序中相邻箱之间的计数差异。图5C显示了在用2mg/ml蛋白酶在室温、50°C、60°C或70°C处理的单个细胞、三个细胞或15pg基因组DNA的测序中,相邻箱之间的读段计数差异(相邻箱之间的读段计数差异的四分位数范围)的柱状图。如显示的,在较低的温度(例如,在室温和50-60°C),相邻箱之间的计数差异是相对小的,具有小的变异;而在70°C,相邻箱之间的计数差异显著增大,具有较大的变异。

#### [0281] 实施例8文库的多样性随着较小的插入尺寸增加

[0282] 在单细胞测序中,只存在两个拷贝的基因组,并因此较小的插入尺寸倾向于增加文库多样性。如图6A中显示的,计数随着插入尺寸减小而增加,并因此文库代表的多样性随着插入尺寸减小而增加。因此,在一些实施方案中,本文的方法在标签化步骤中使用较高量的转座酶,以增加片段化并减小加标签的核酸片段的插入尺寸。图6B显示了用1μl Tn5或2μl Tn5处理的文库的插入尺寸。如显示的,当在标签化反应中使用1μl Tn5时,平均片段尺寸为约550bp;而当在标签化反应中使用2μl Tn5时,平均片段尺寸为约400bp。与较小的插入

尺寸一致,当用 $2\mu\text{l}$  Tn5处理时的文库多样性与用 $1\mu\text{l}$  Tn5处理的文库多样性相比增加,如图6C中显示的。

[0283] 实施例9优化PCR循环

[0284] 在使用最小细胞群体的测序中,输入DNA是相对小的,并因此可调整PCR的循环数以实现更好的测序结果。在该实施例中,使用单个细胞作为起始材料测试并优化了PCR的循环数。图7显示了在根据本文提供的方法使用具有16个循环、18个循环或20个循环的PCR的单细胞测序中,计数和拷贝数分析结果的柱状图。如显示的,当使用具有16个循环的PCR时,噪声大,并且当使用具有18个循环或20个循环的PCR时,噪声显著减小。

[0285] 实施例10使用一个、三个或五个细胞的读段分布

[0286] 在该实施例中,分析了使用一个、三个或五个细胞的读段分布。图8A显示了三次单细胞测序的读段分布。如显示的,在三次单细胞测序间,读段区域并不完全重叠。因此,增加细胞数目可有助于拓宽覆盖率。图8B显示了单细胞测序、三细胞测序或五细胞测序的读段分布。如显示的,基因组覆盖率随着细胞数目增多而增大。图8C显示了使用单个细胞、三个细胞或五个细胞的平均文库多样性和评估的基因组覆盖率的柱状图。如显示的,经评估,一个细胞可覆盖基因组的约40%,并且三个细胞可覆盖超过基因组的50%,并且五个细胞可覆盖基因组的约60%。使用一个细胞、三个细胞和五个细胞的平均文库计数分别为约5百万、1千5百万和2千万。

[0287] 图8D显示了整体成功率。如显示的,当使用多于一个细胞时,整体成功率为99% ( $N=81$ )。当使用单个细胞时,整体成功率仍然相对高94% ( $N=187$ )。

[0288] 实施例11比较不同文库制备方法间的计数和拷贝数数据

[0289] 在该实施例中,将本文提供的方法与目前的一些单细胞制备方法比较。

[0290] 图9A显示了使用REPLIg单细胞(REPLIg Single Cell, MDA)与Nexteral XT文库制备的拷贝数分析。由QIAGEN开发的REPLI-g单细胞试剂盒是被特别地设计以扩增来自单细胞(1至<1000个细胞)的基因组DNA或具有基因组覆盖率的纯化的基因组DNA。由QIAGEN开发的REPLI-g单细胞试剂盒利用多重置换扩增(MDA)技术。参见Spits等,2006,Whole-genome multiple displacement amplification from single cells,Nature protocols 1 (4): 1965-70。但是,由于MDA引入过度扩增偏倚,当源自单个细胞、三个细胞或五个细胞时,拷贝数变异数据很嘈杂,如图9A中显示的。

[0291] 图9B显示了使用SurePlex(PicoPlex)与Nexteral XT文库制备的拷贝数分析。由Illumina, Inc (San Diego, CA)开发的SurePlex扩增系统是用于提取并扩增来自单细胞或很少的单细胞的DNA的溶液。如显示的,SurePlex扩增系统与MDA相比显著减少噪声。

[0292] 图9C显示了使用本文提供的方法(Nextera SC)的拷贝数分析。如显示的,与使用SurePlex扩增系统相比,噪声被进一步减少。

[0293] 实施例12检测镶嵌现象

[0294] 在该实施例中,示例了使用本文提供的方法检测镶嵌现象。图10A显示了使用单个GM50121细胞,18号染色体的拷贝数分析数据。显示了来自三次单细胞测序的拷贝数数据。在每次单细胞测序中,检测代表15.4MB DNA的群体。图10B显示了使用单个GM20916细胞的计数数目数据。如显示的,箭头指示源于镶嵌现象的计数。图10C显示了使用单个GM20916细胞,15号染色体、X染色体和10号染色体的拷贝数分析数据。所分析的每条染色体的拷贝数

数据检测到代表另一条染色体的另外的群体。相似地,图10D显示了使用单个GM1021细胞,1号染色体和11染色体的拷贝数分析数据。如这些图中显示的,在图10D中分析的每条染色体的拷贝数数据也检测到代表另一条染色体的另外的群体。

**[0295]** 已描述了很多实施方案。但是,将理解可进行多种改变。相应地,其他的实施方案在以下的权利要求书的范围内。

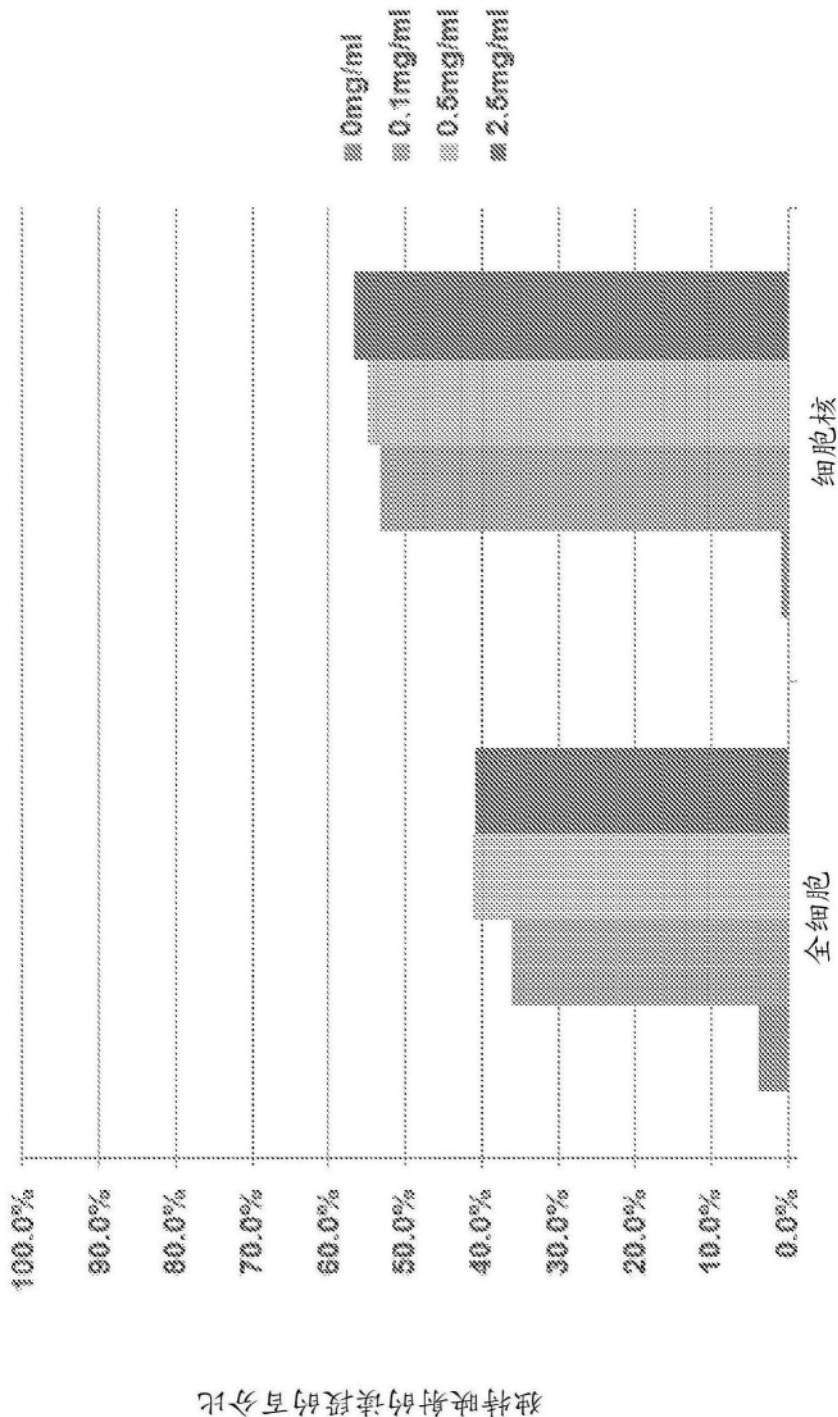


图1

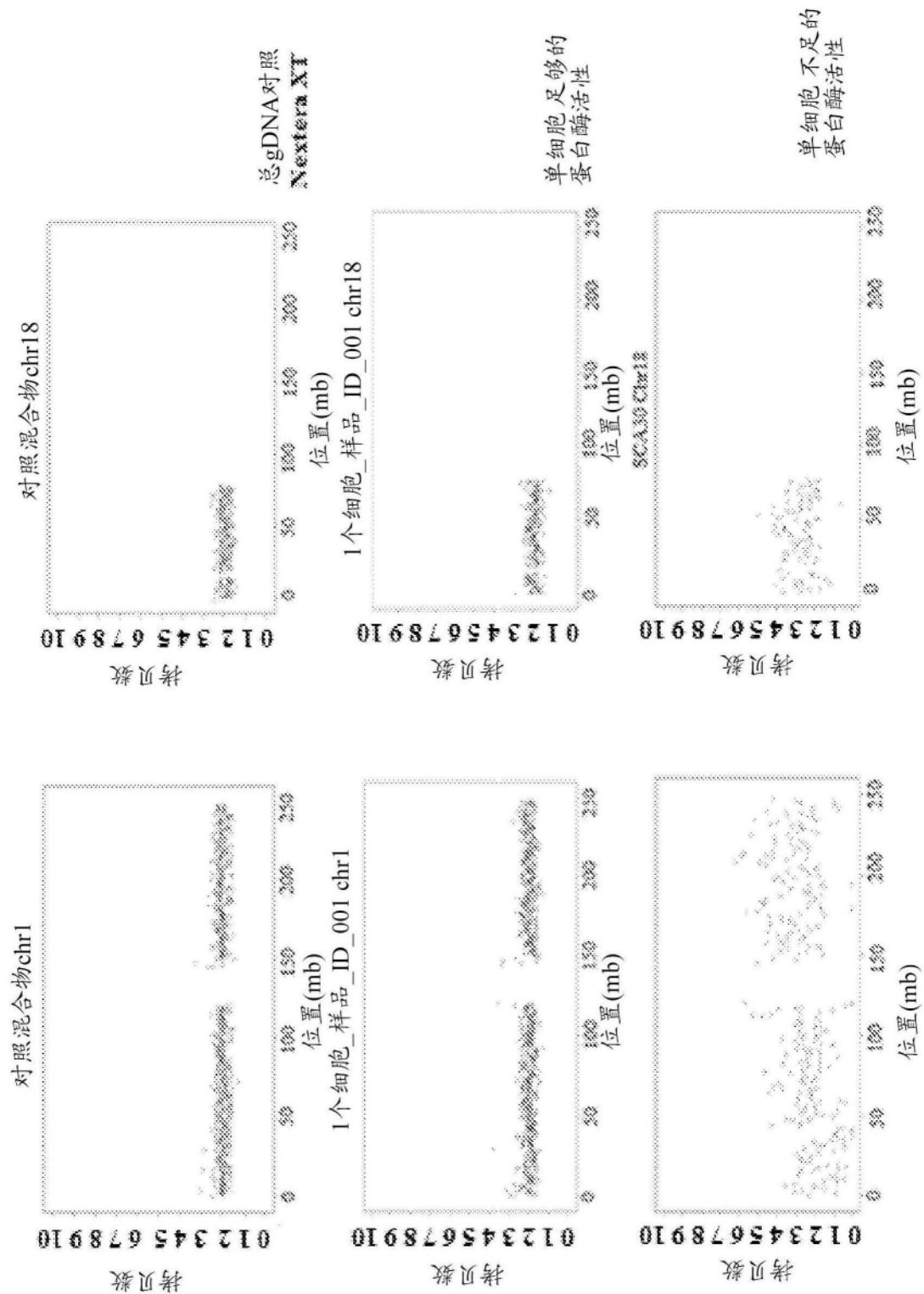


图2

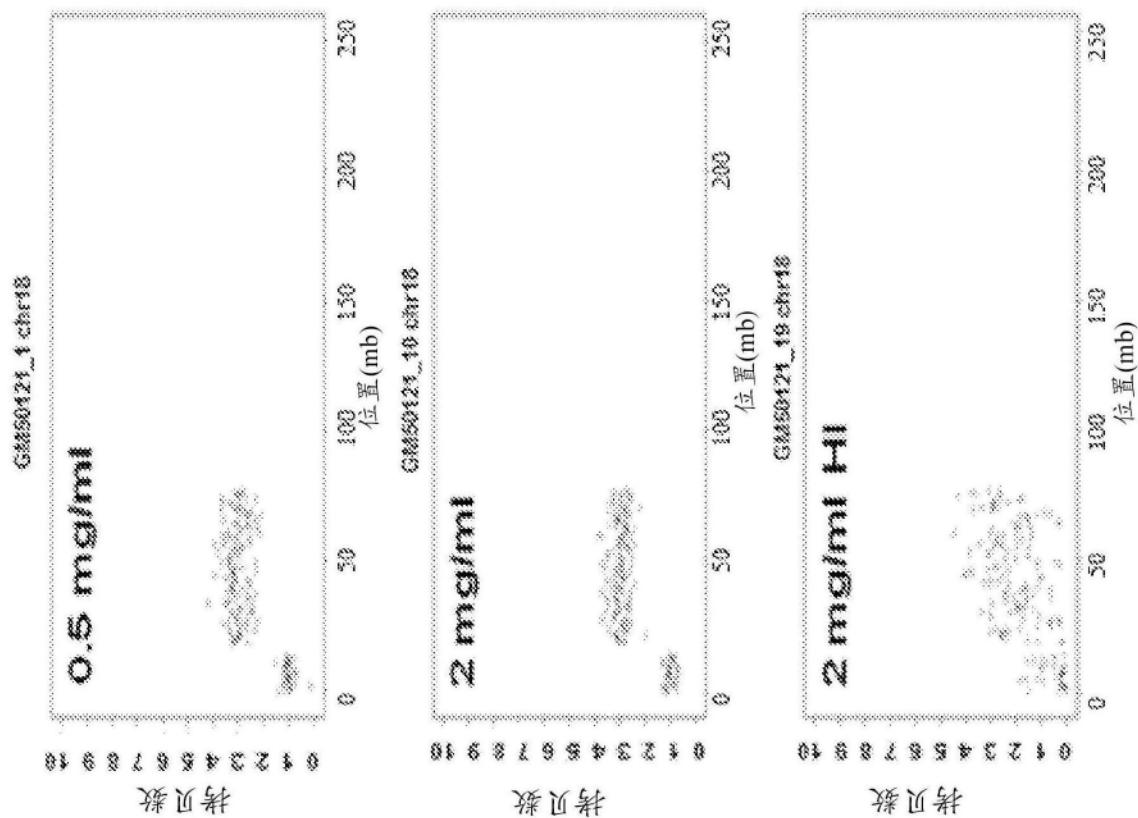
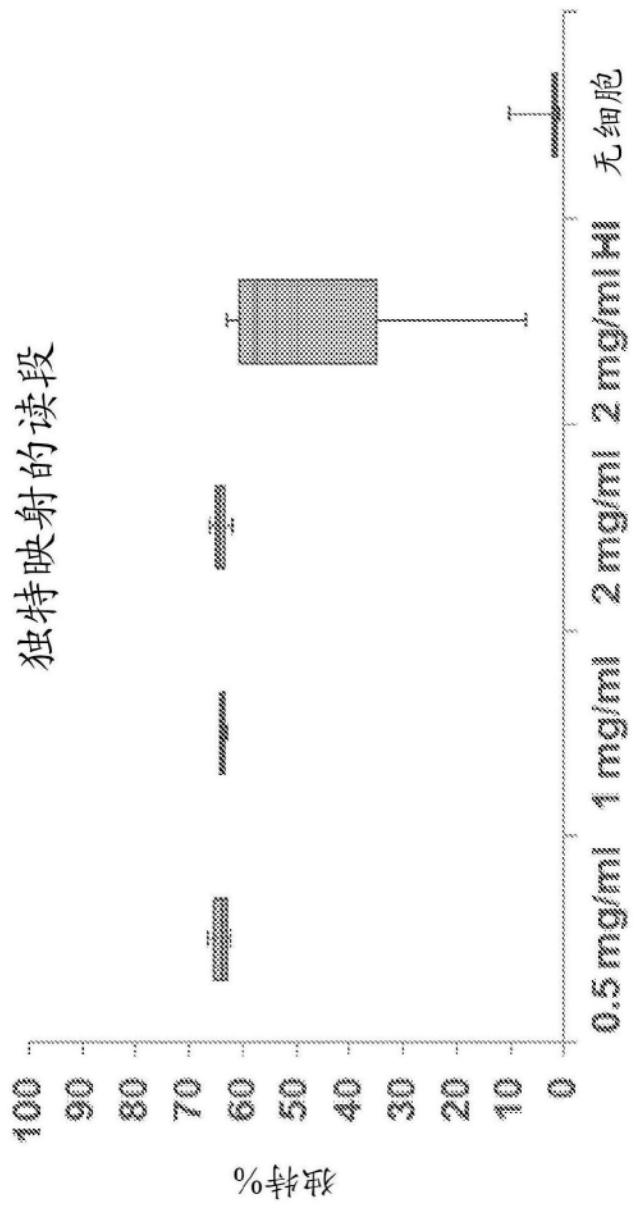


图3A



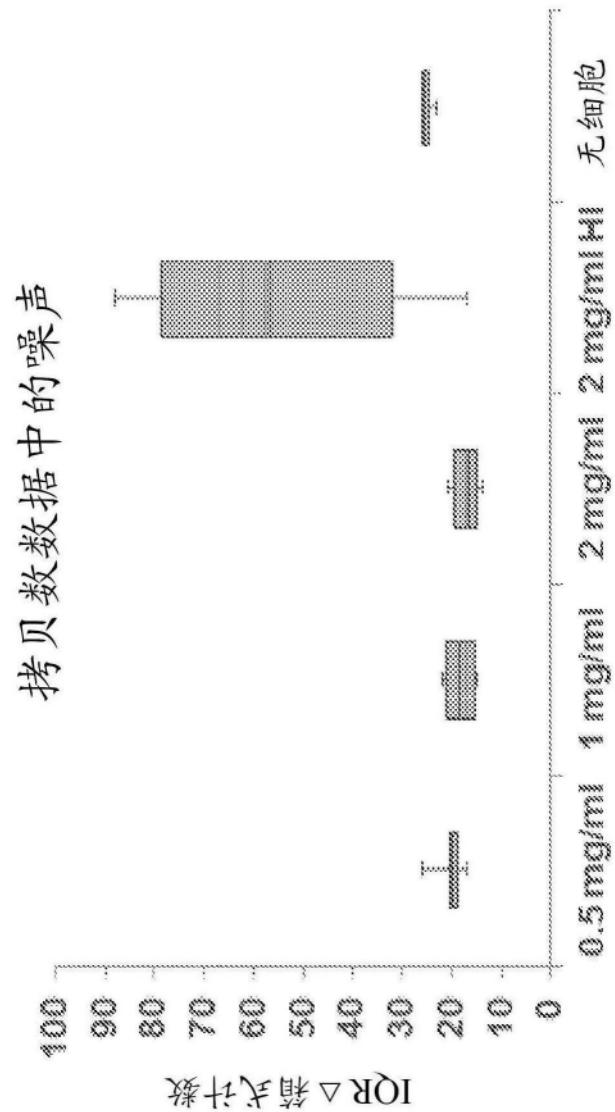


图3C

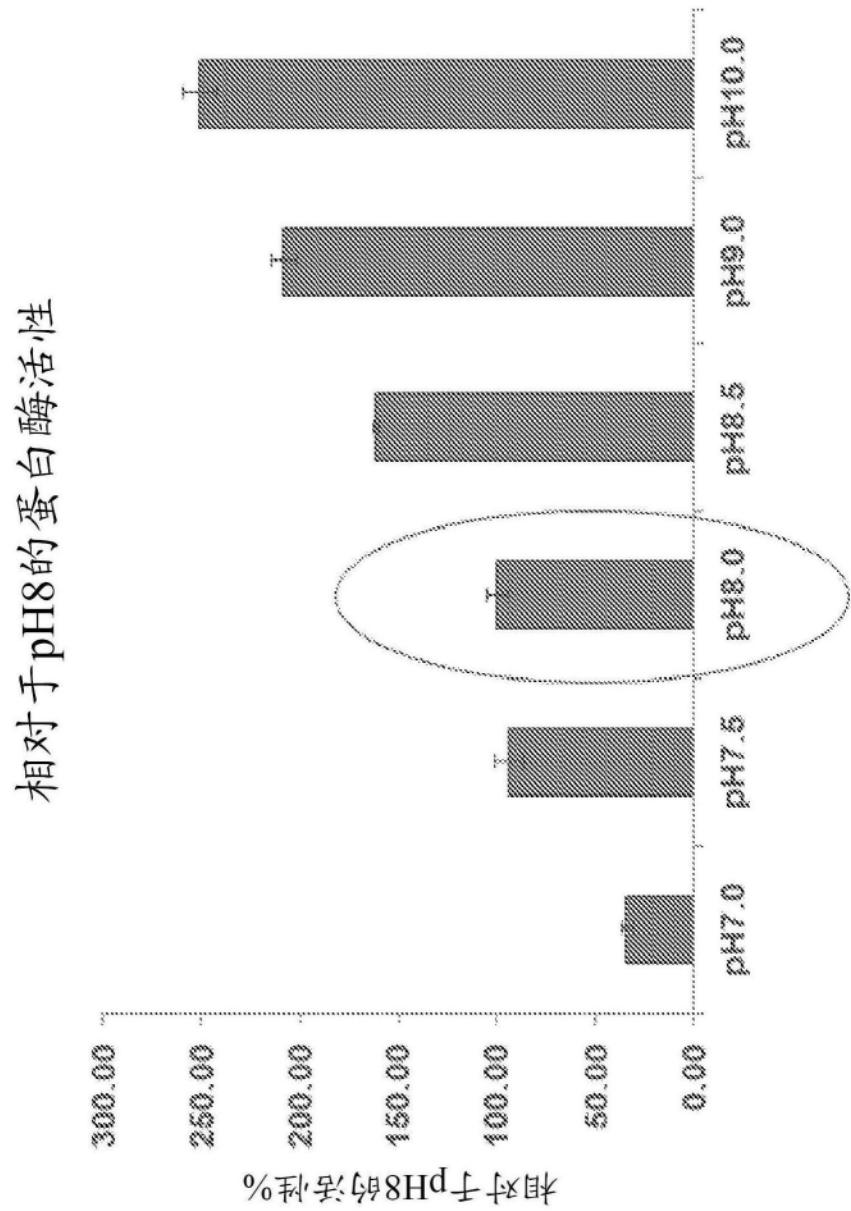


图4A

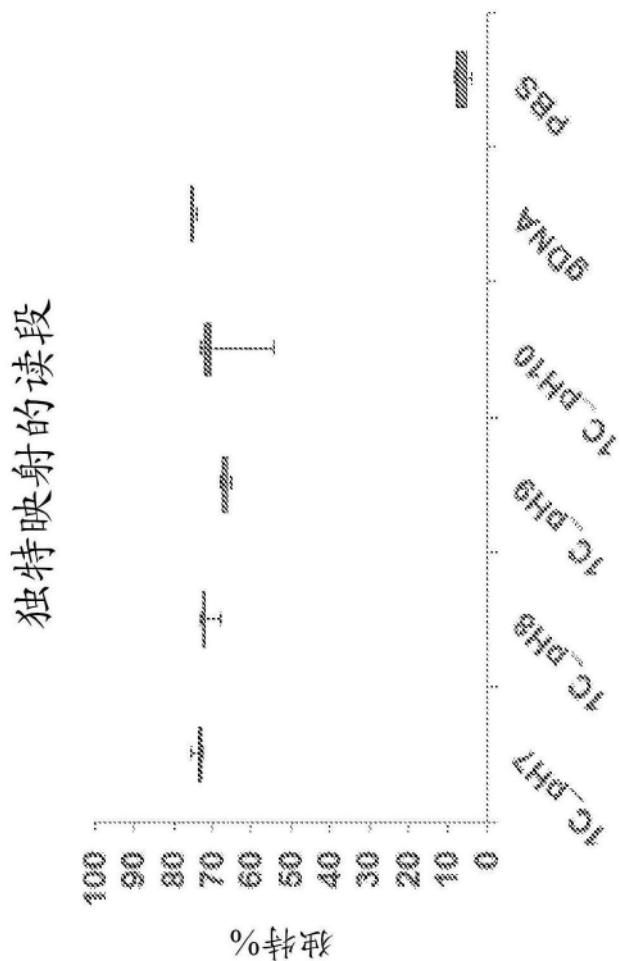


图4B

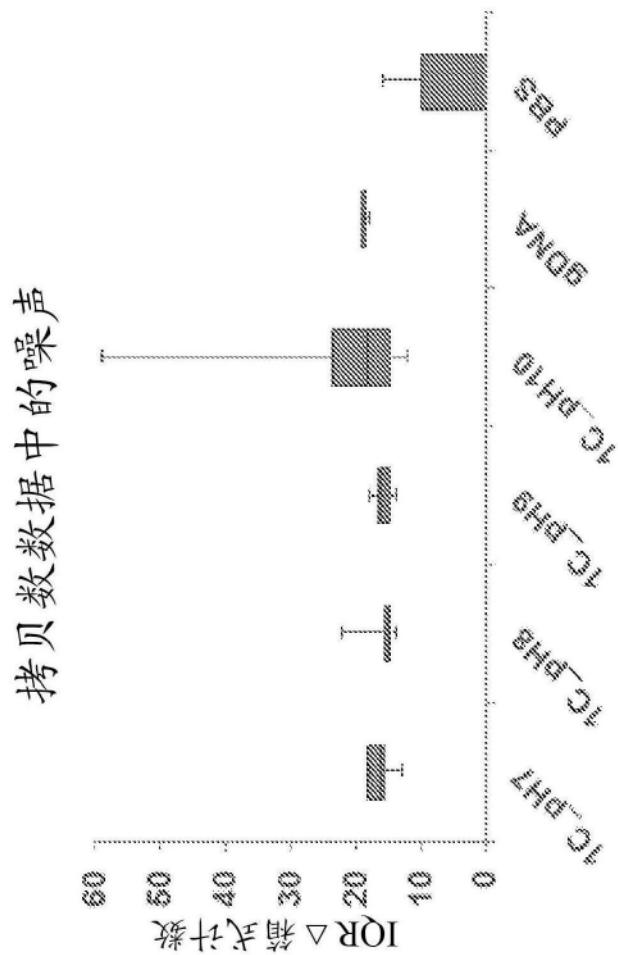


图4C

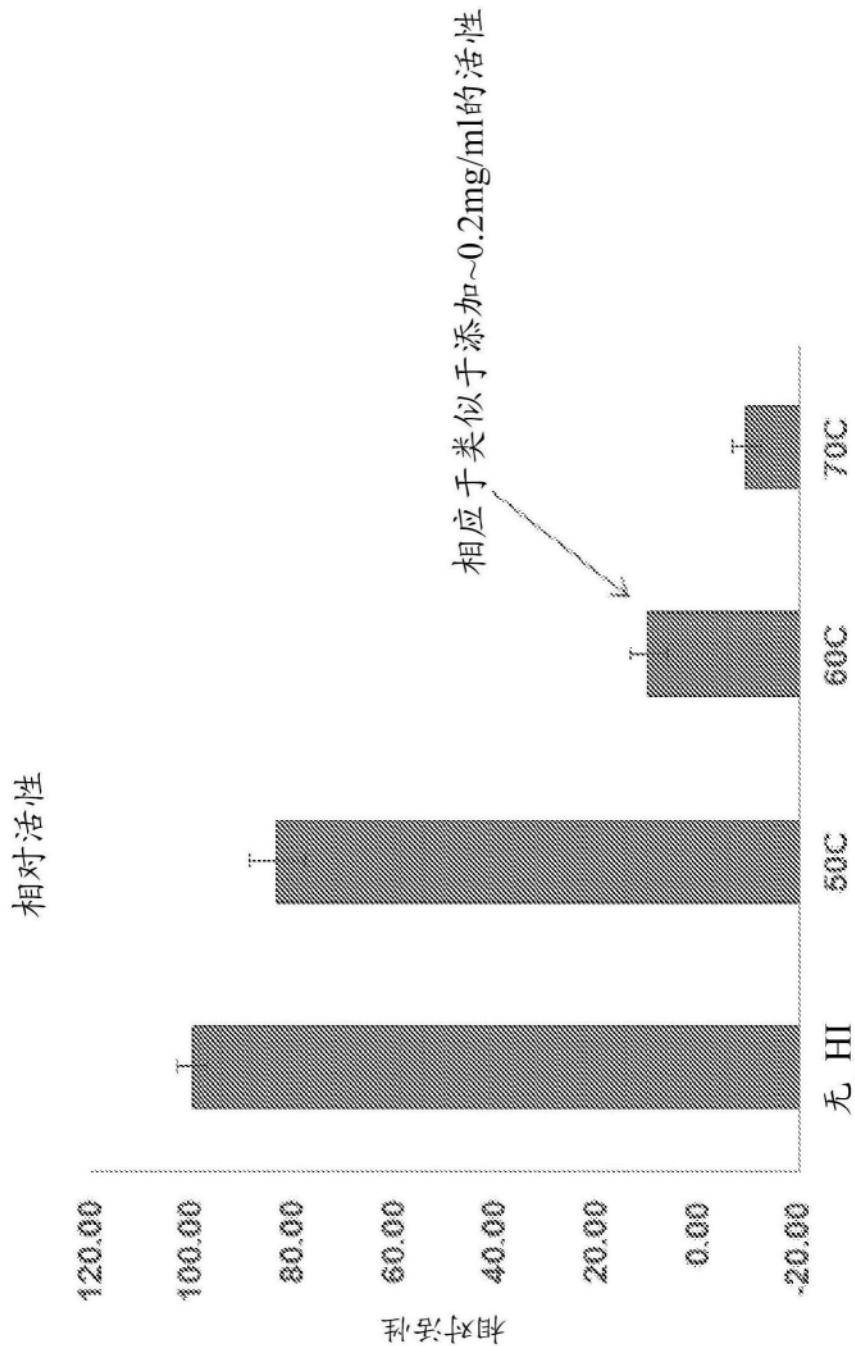
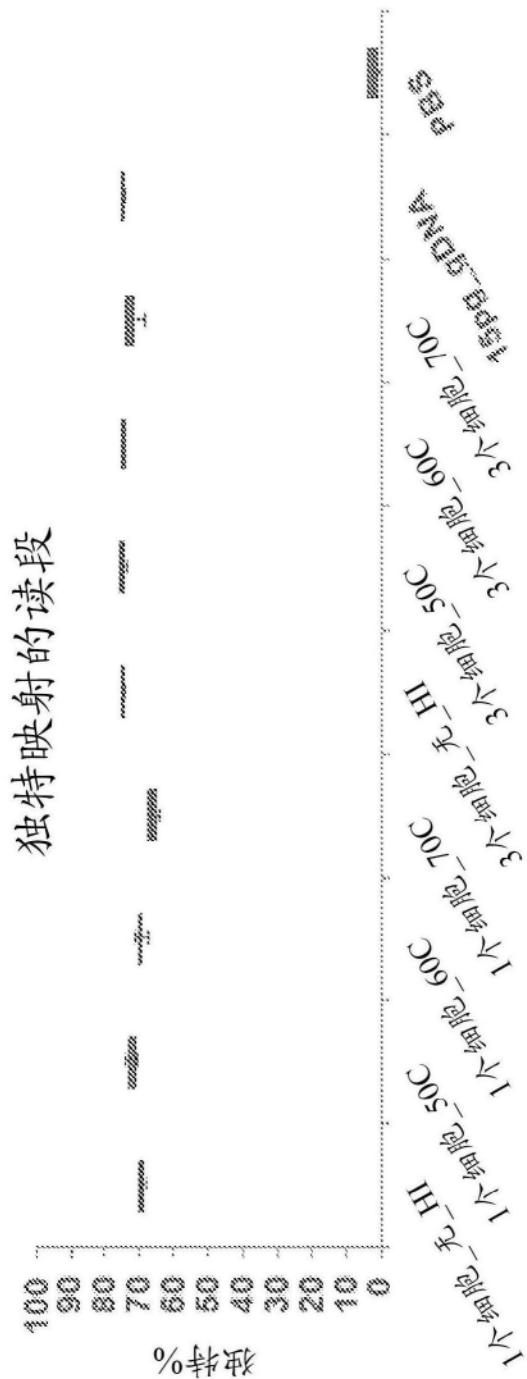


图5A



冬 5B

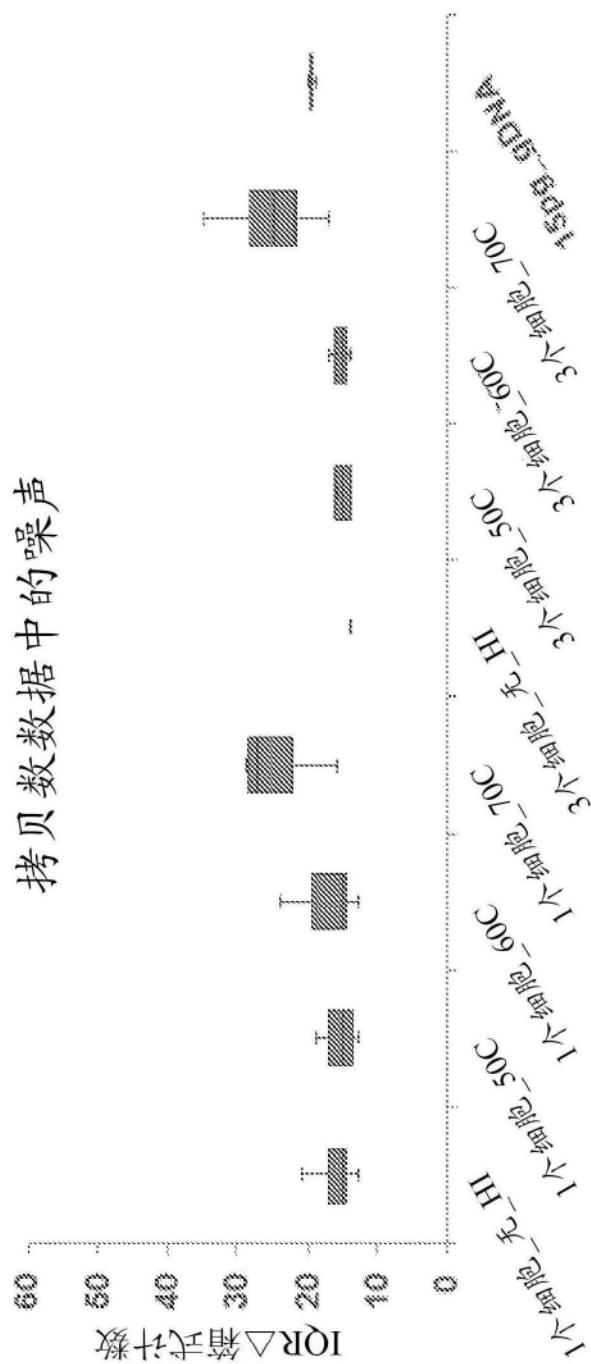


图5C

GM50121的PE插入尺寸

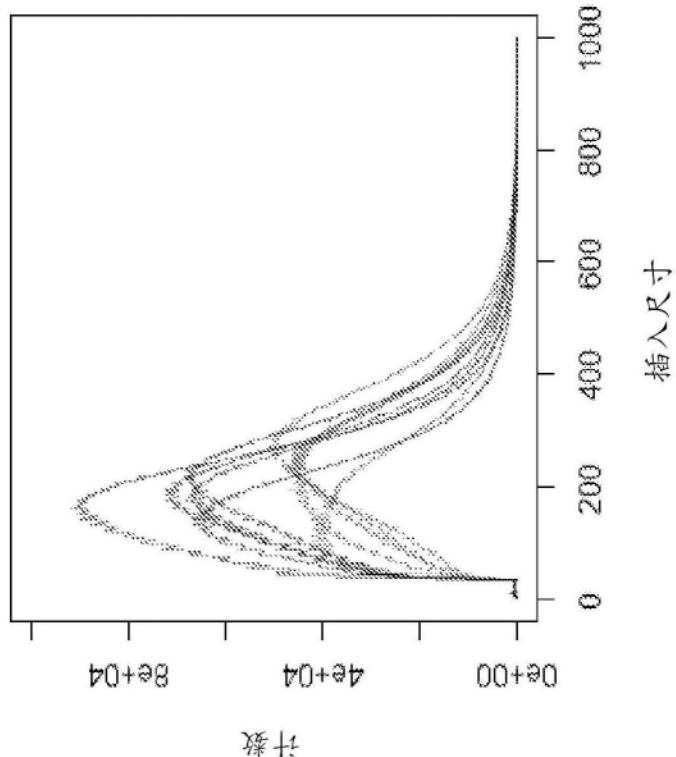


图6A

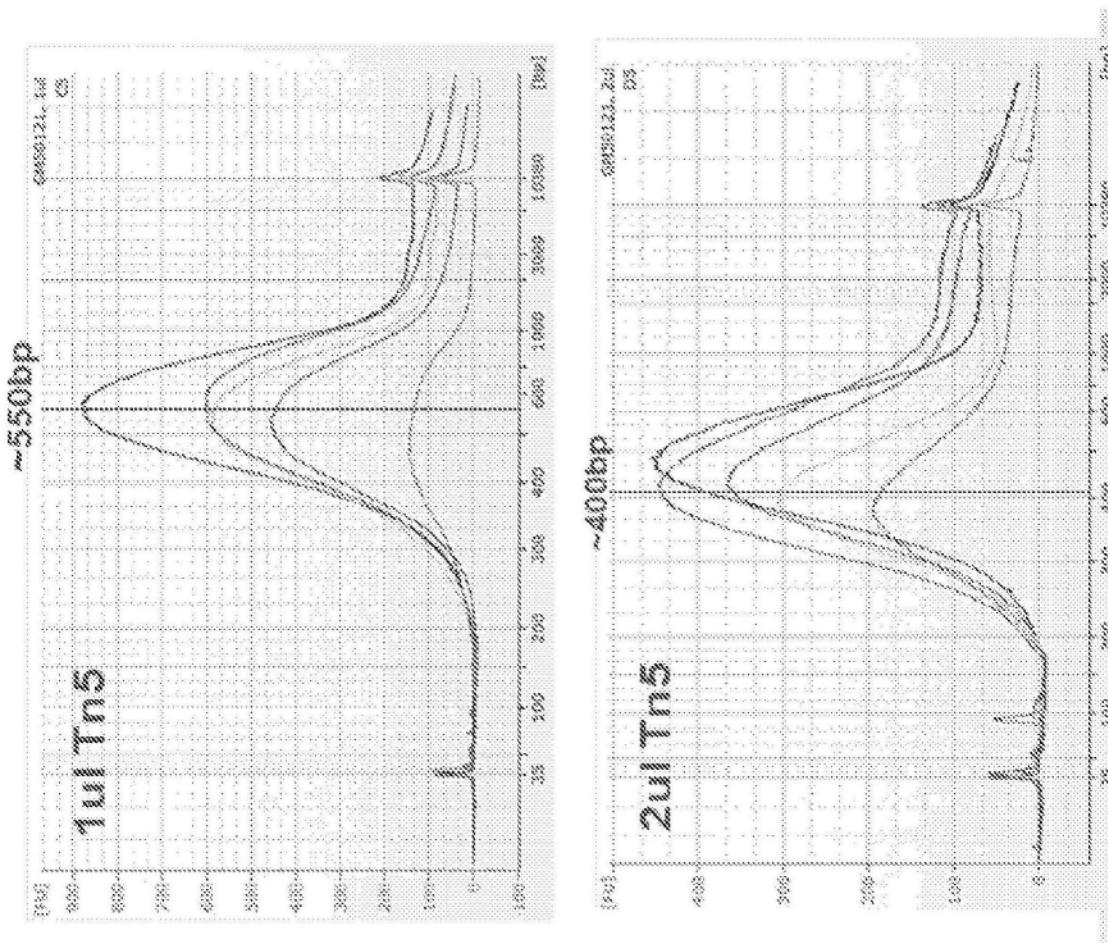


图6B

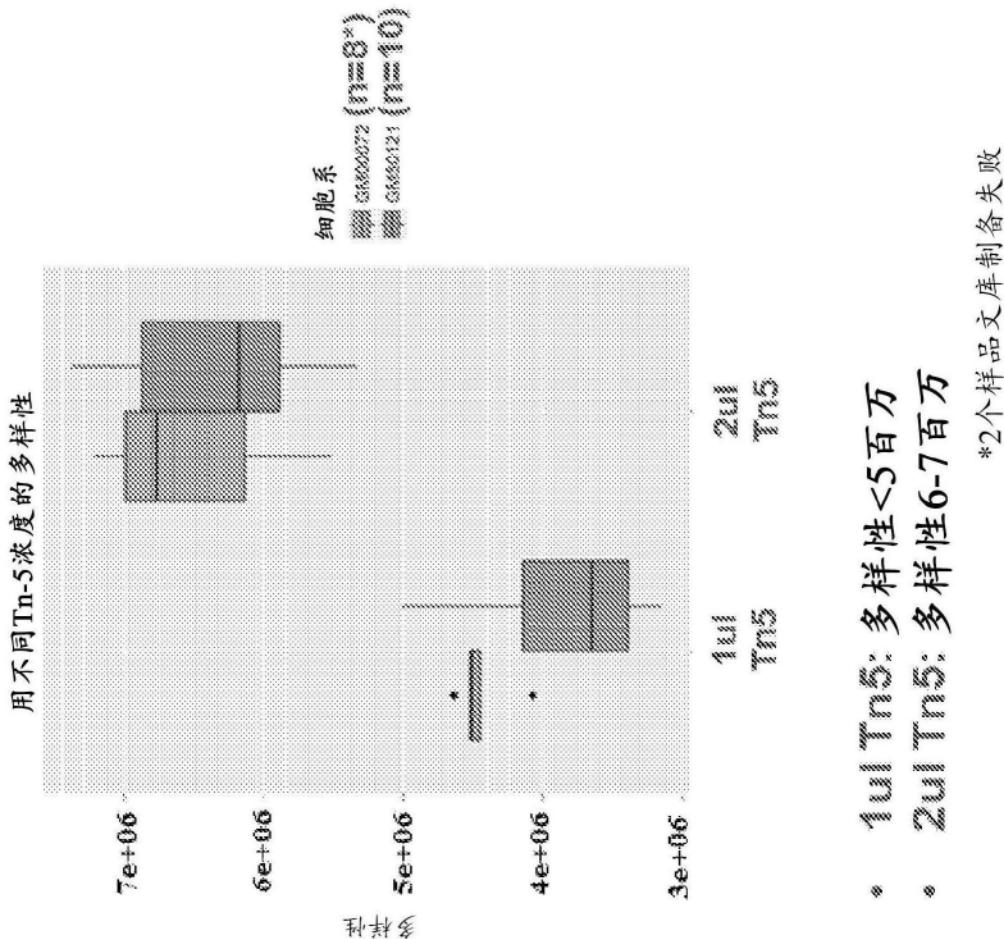


图6C

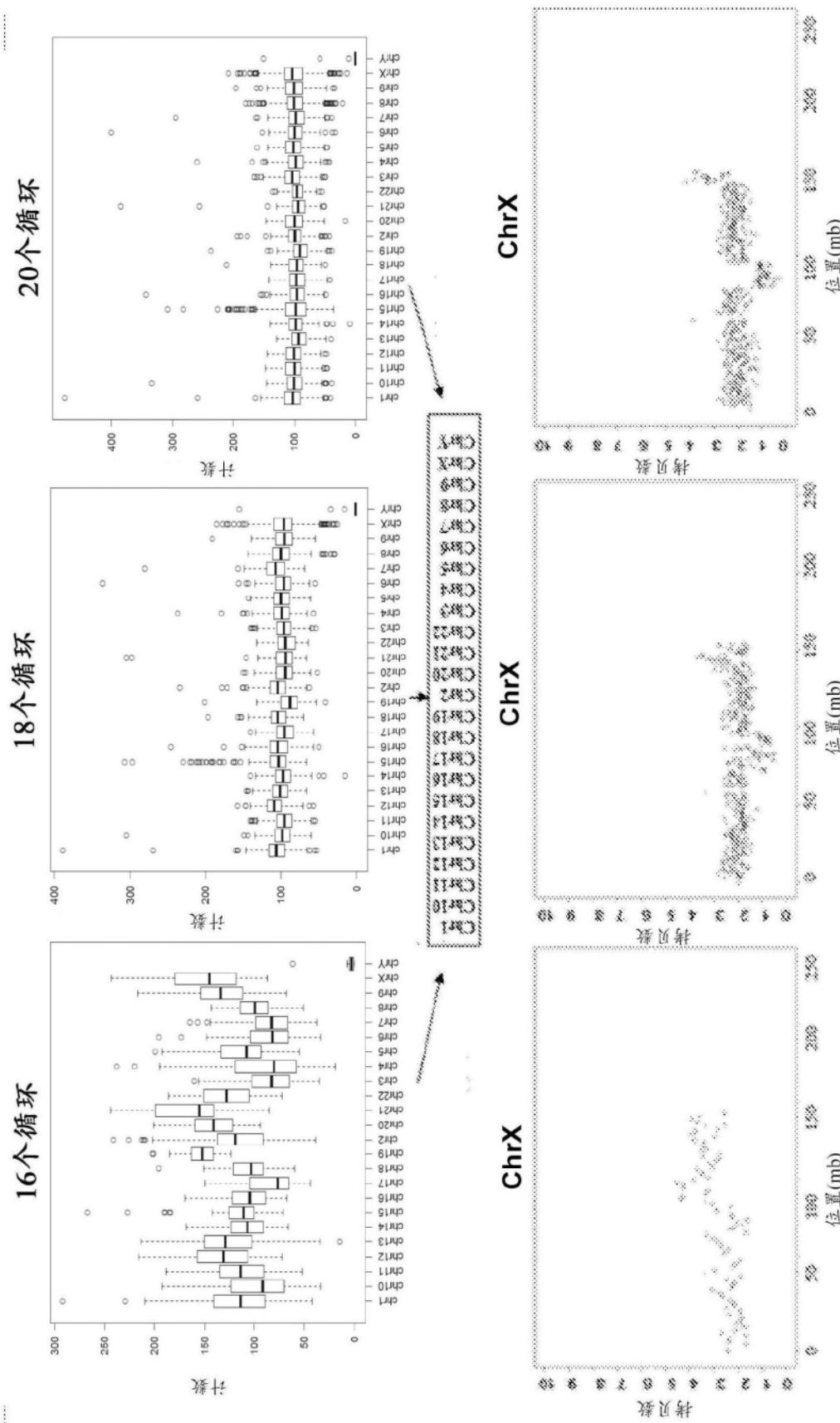


图7

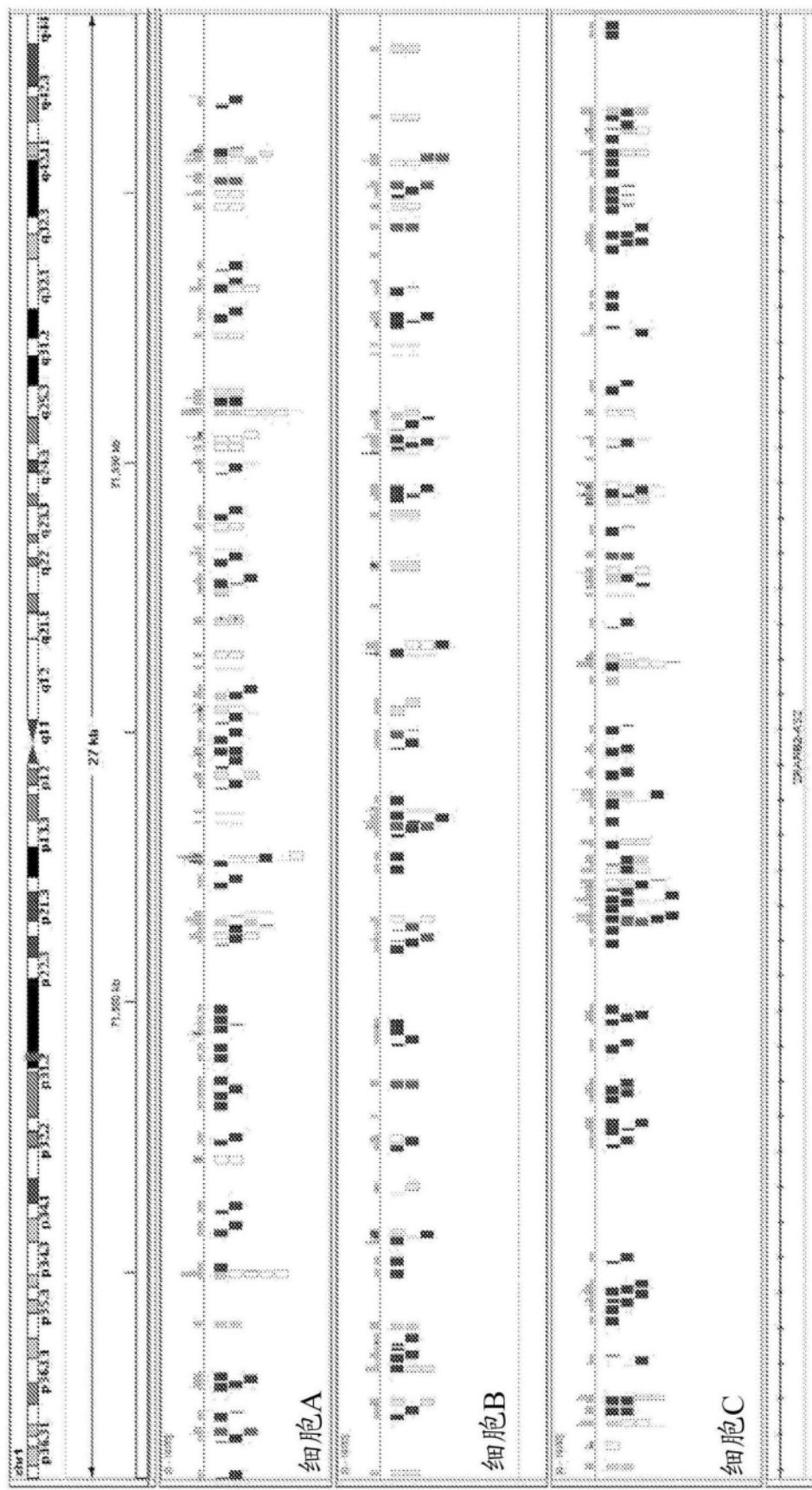


图8A

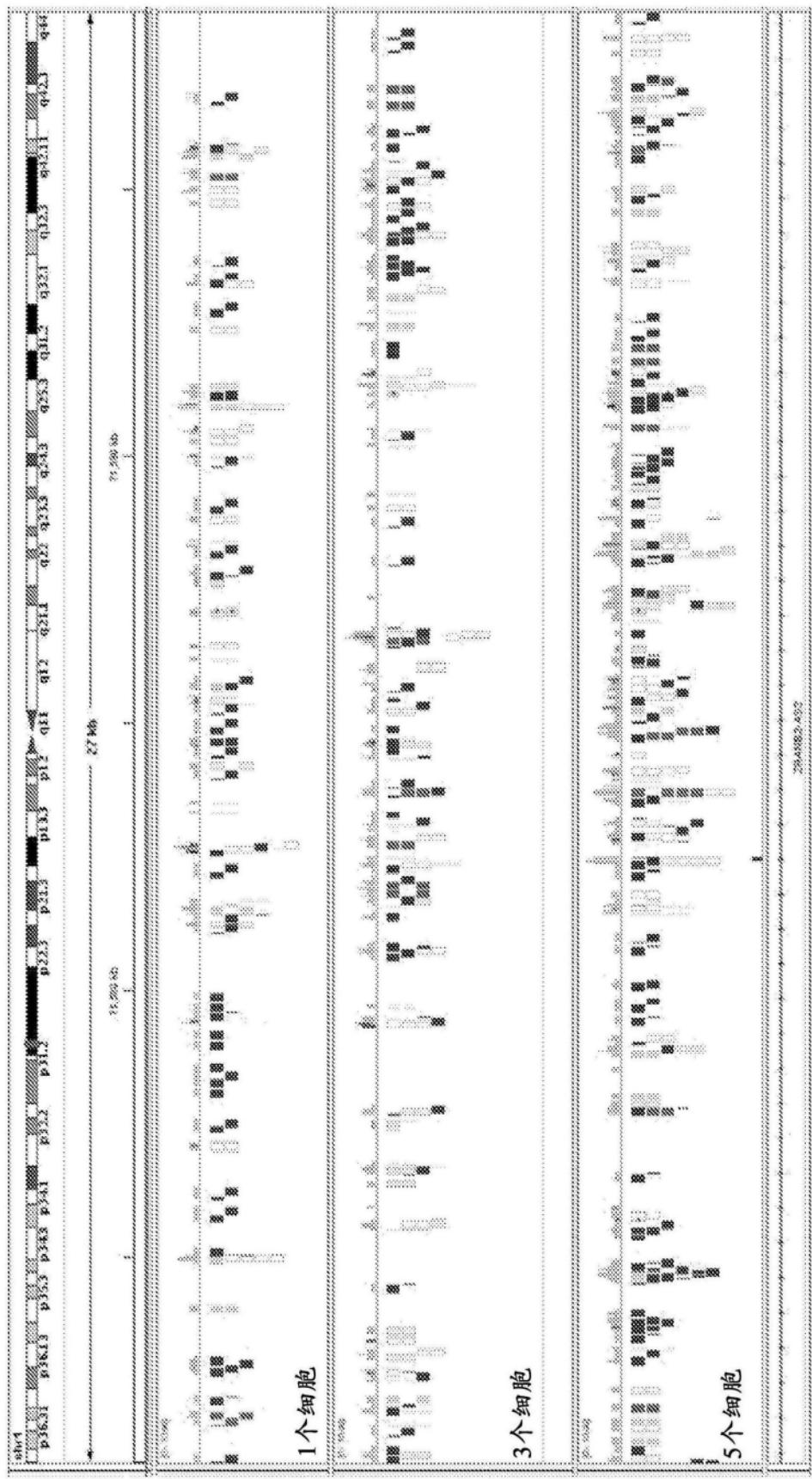


图8B

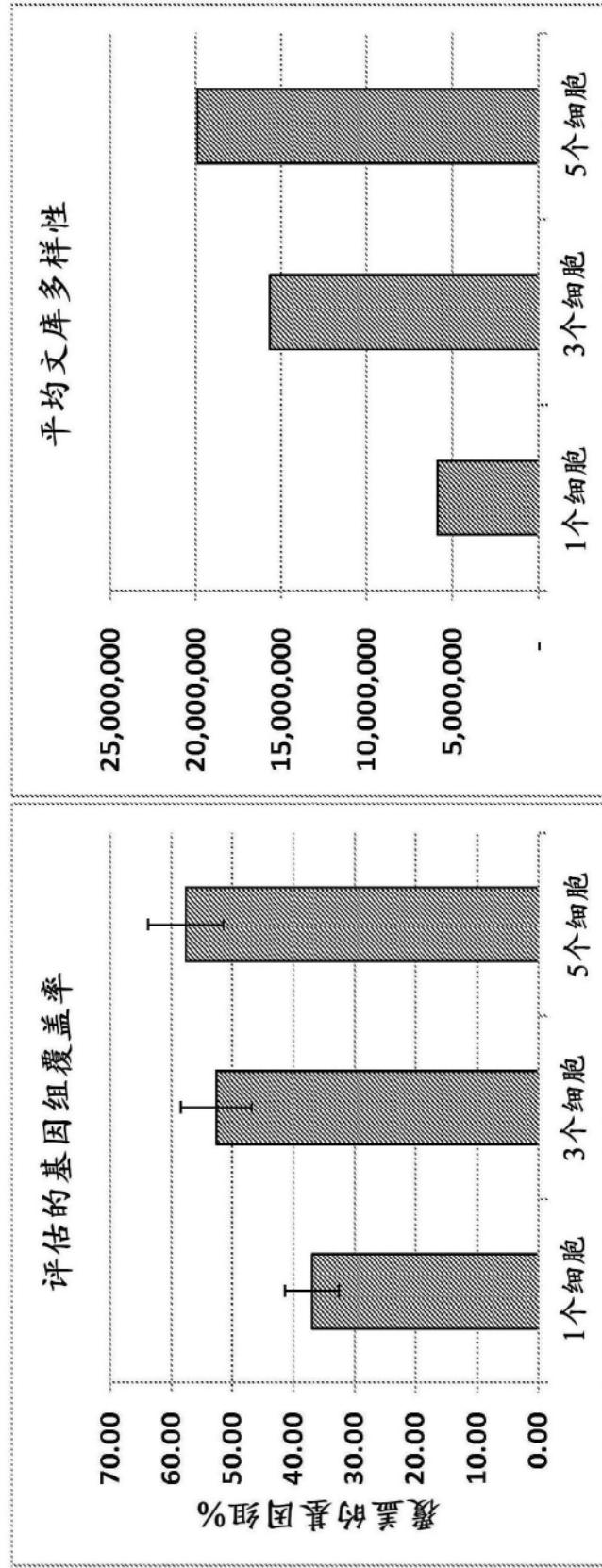


图8C

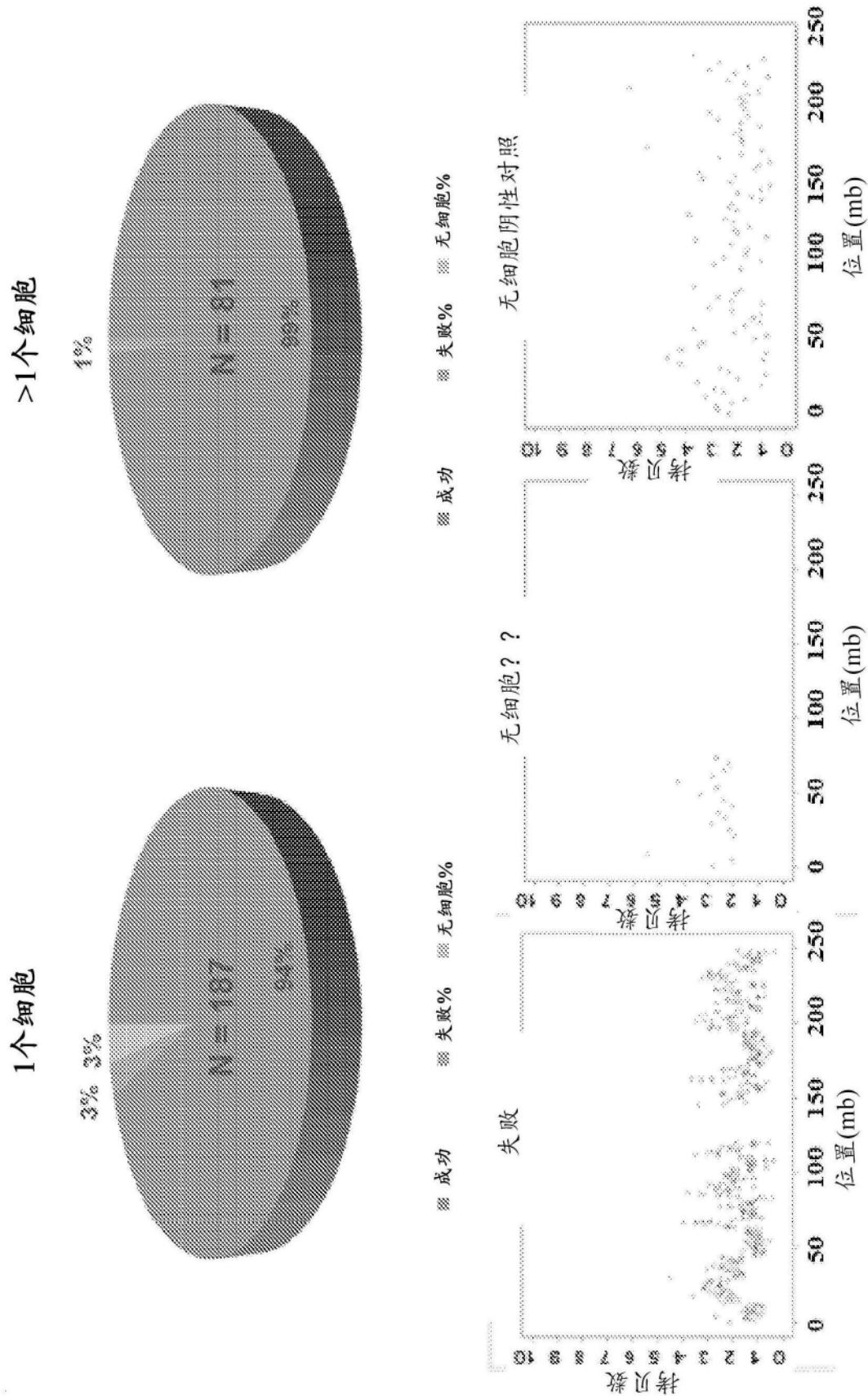
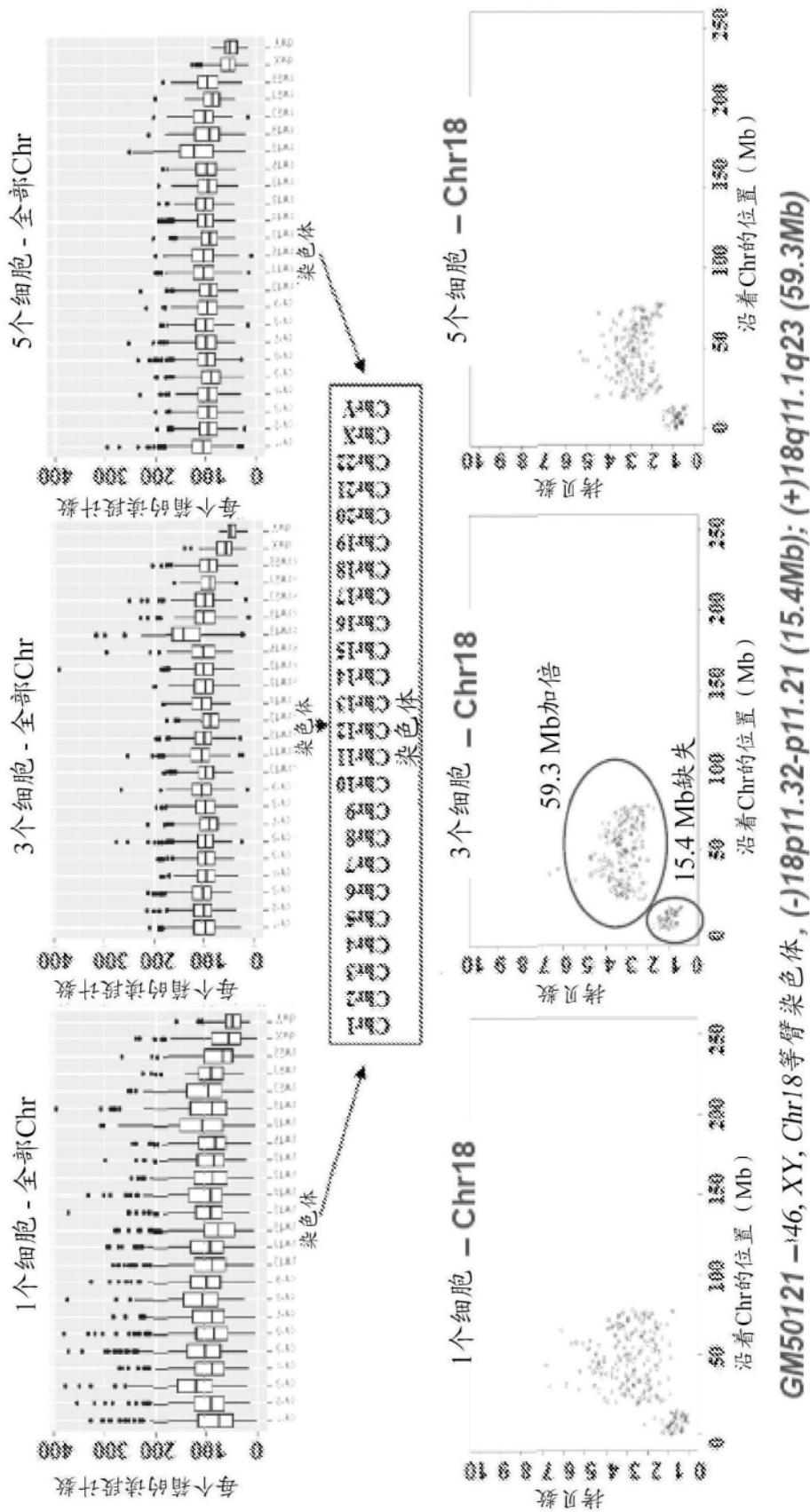


图8D

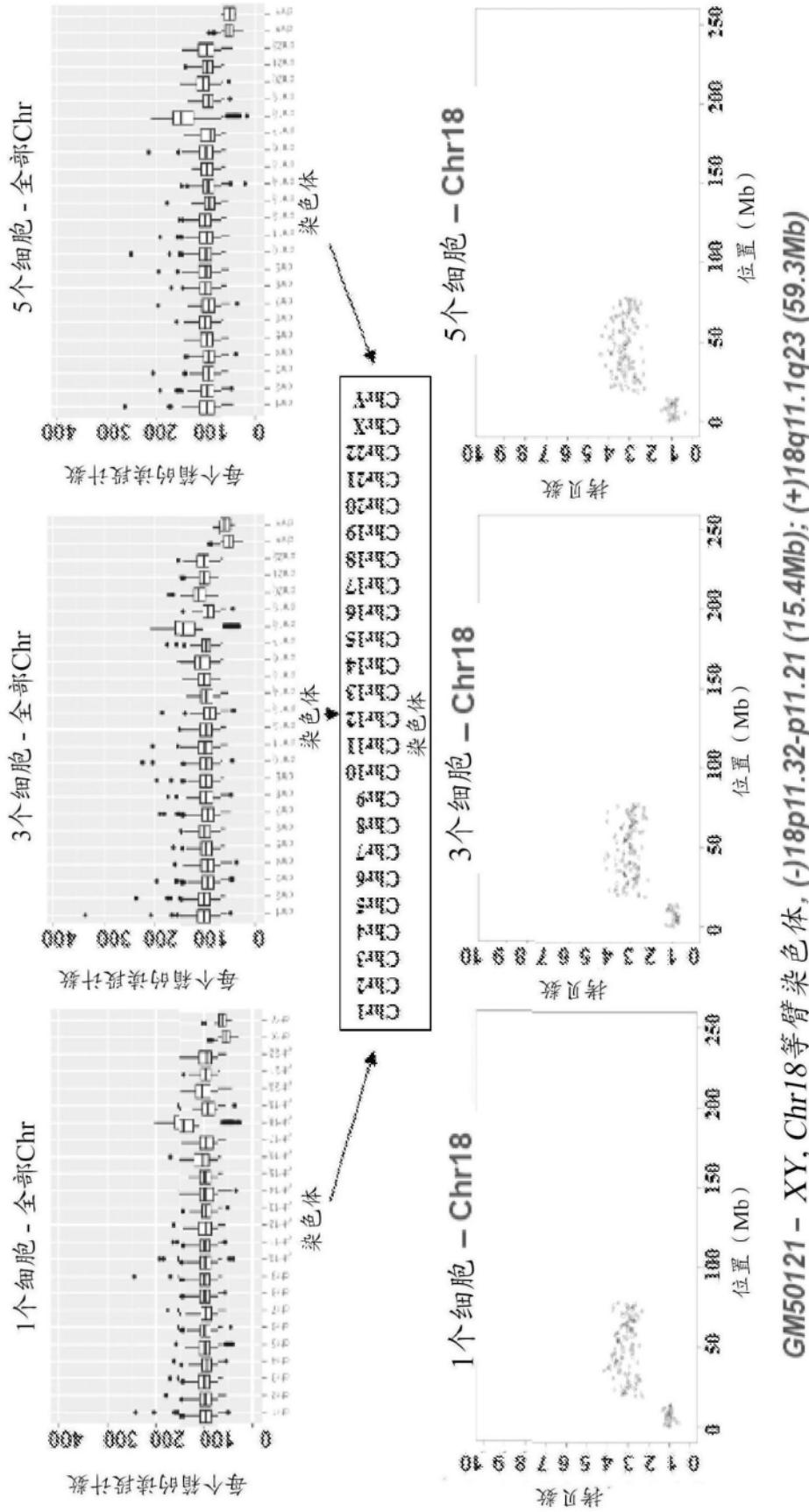
### REPLIG单细胞(MDA)+Nextera XT文库制备



**GM50121 - 46, XY, Chr18等臂染色体, (-)18p11.32-p11.21 (15.4Mb); (+)18q11.1q23 (59.3Mb)**

图9A

SurePlex (基于PEP/DOP-PCR的WGA)+Nextera XT文库制备



**GM50121 - XY, Chr18等臂染色体, (-)18p11.32-p11.21 (15.4Mb); (+)18q11.1q23 (59.3Mb)**

图9B

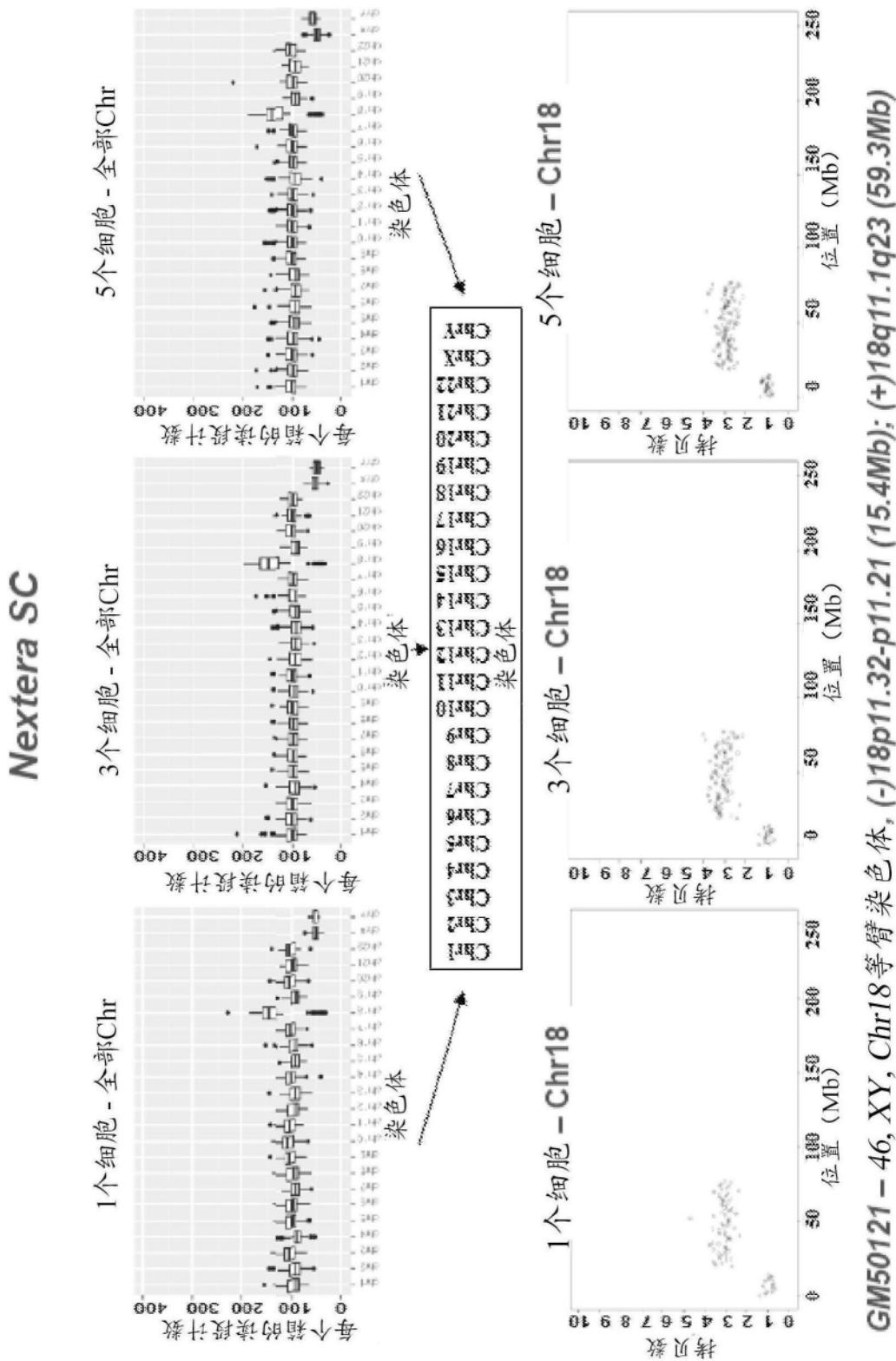


图9C

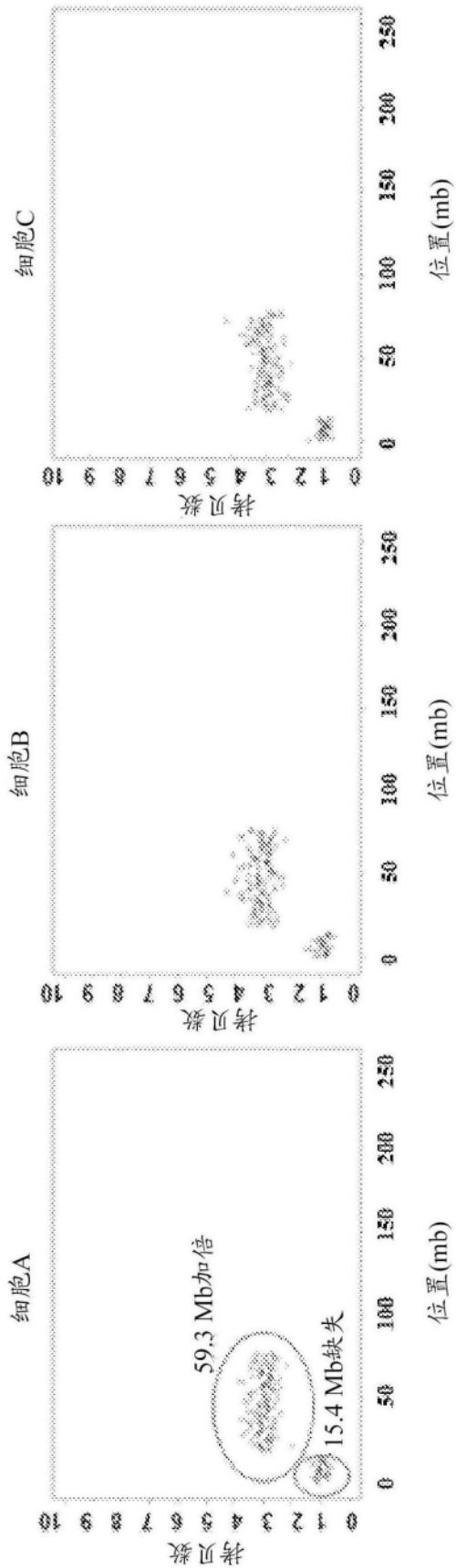


图10A

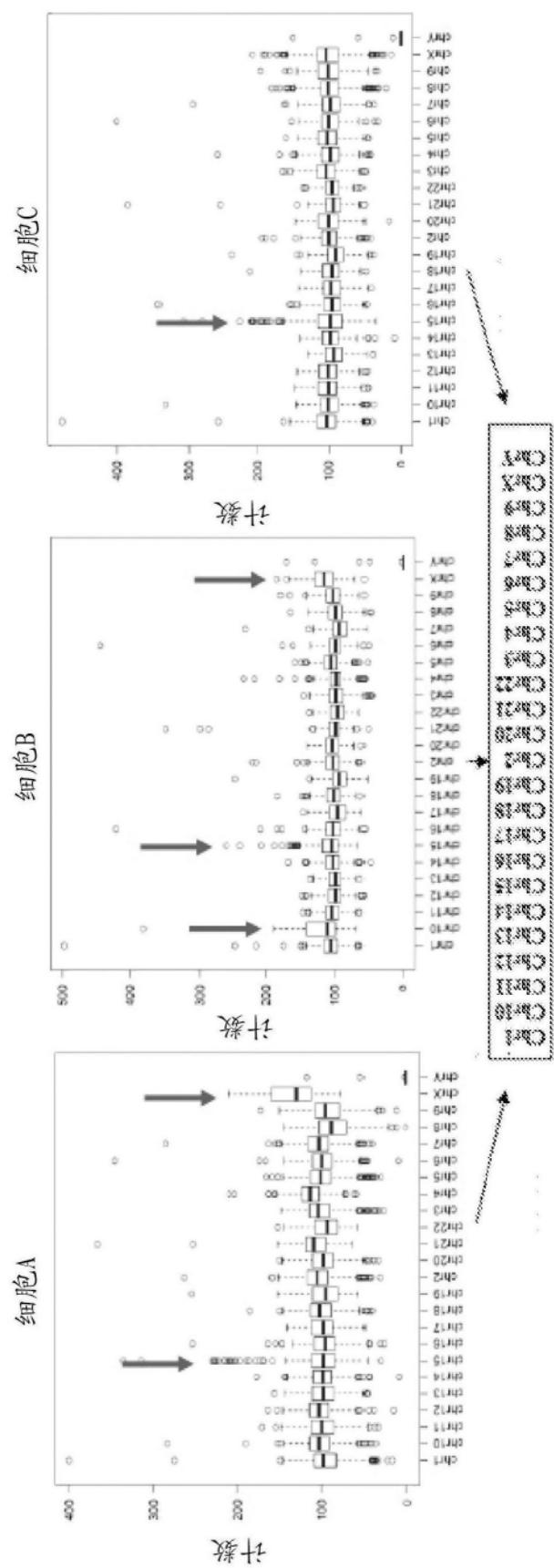


图10B

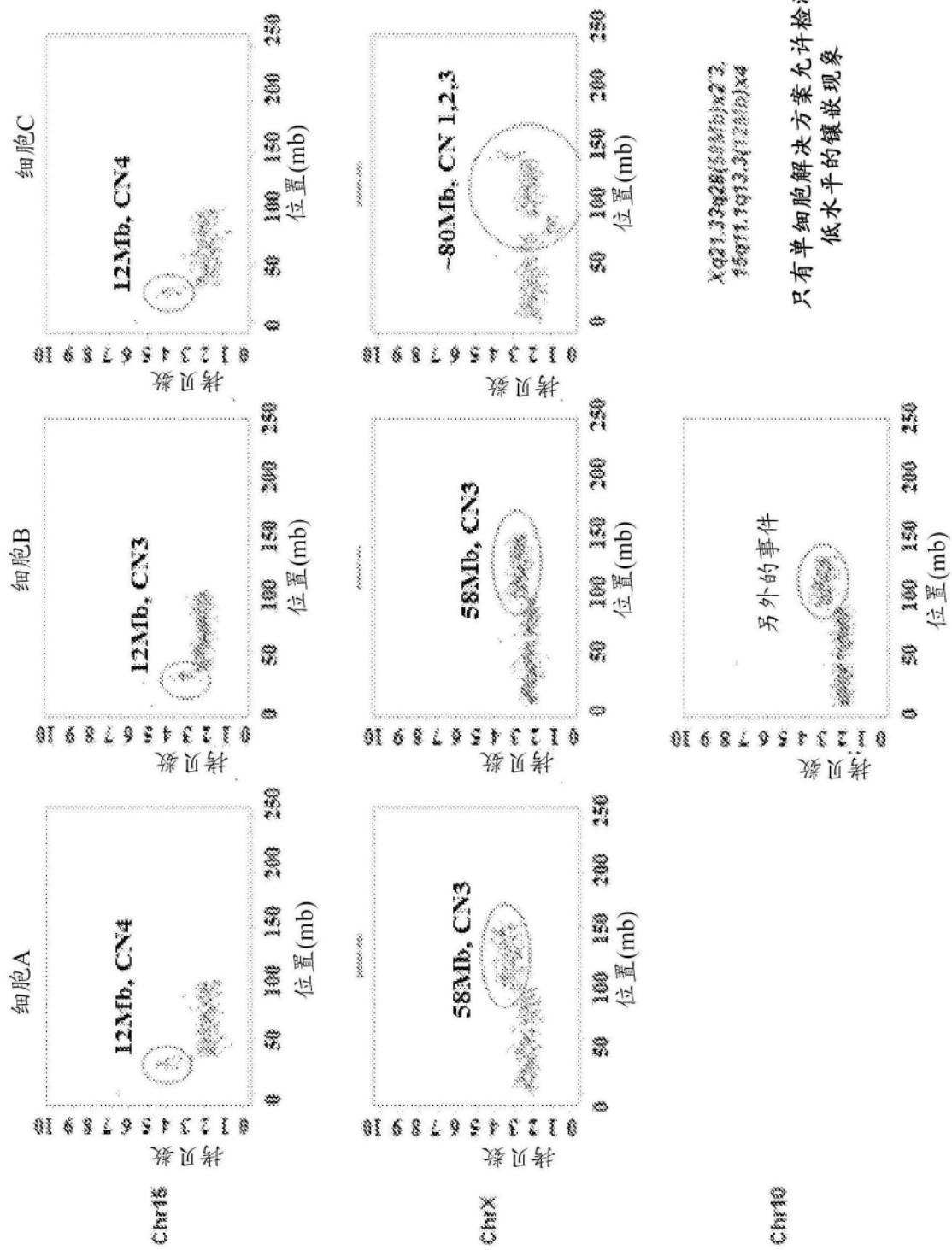


图10C

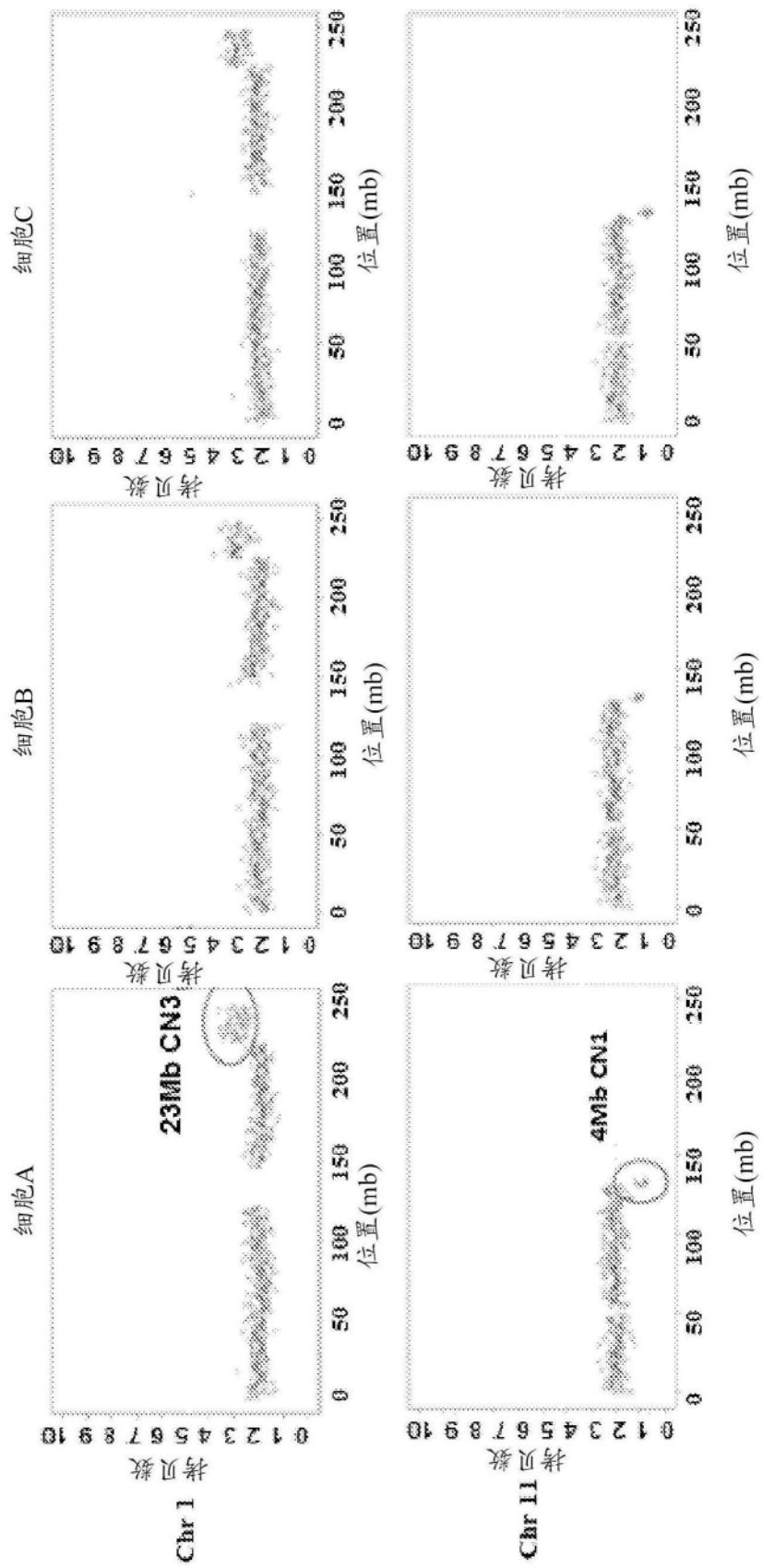


图10D