



(21) 申请号 201680046621.0

CN 107002130 A, 2017.08.01

(22) 申请日 2016.08.12

CN 103827320 A, 2014.05.28

WO 2012060595 A2, 2012.05.10

(65) 同一申请的已公布的文献号

申请公布号 CN 107849607 A

夏帆. 基于纳米孔在复杂基质中的核酸和小分子双检测.《中国化学会第29届学术年会摘要集——第02分会: 分离分析及微、纳流控新方法》. 2014,

(43) 申请公布日 2018.03.27

(30) 优先权数据

62/204,396 2015.08.12 US

Nannan Liu等. Two-Way Nanopore Sensing of Sequence-Specific Oligonucleotides and Small-Molecule Targets in Complex Matrices Using Integrated DNA Supersandwich Structures.《Angew. Chem. Int.Ed.》. 2013, 第52卷

(85) PCT国际申请进入国家阶段日

2018.02.07

Nannan Liu等. Two-Way Nanopore Sensing of Sequence-Specific Oligonucleotides and Small-Molecule Targets in Complex Matrices Using Integrated DNA Supersandwich Structures.《Angew. Chem. Int.Ed.》. 2013, 第52卷

(86) PCT国际申请的申请数据

PCT/CN2016/094802 2016.08.12

(87) PCT国际申请的公布数据

WO2017/025059 EN 2017.02.16

(73) 专利权人 香港中文大学

地址 中国香港新界

(72) 发明人 卢煜明 赵慧君 郑淑恒

(74) 专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

专利代理师 王达佐 洪欣

Vitali Proutski等. SWAN: sliding window analysis of nucleotide sequence variability.《Bioinformatics》. 1998, 第14卷 (第5期),

(51) Int.Cl.

C12Q 1/6869 (2018.01)

审查员 吴至芳

(56) 对比文件

WO 2016077313 A1, 2016.05.19

权利要求书4页 说明书20页 附图17页

(54) 发明名称

血浆DNA的单分子测序

(57) 摘要

本发明提供了一种确定核酸序列的方法。所述方法可以包括接收多个DNA片段。所述方法还可以包括将第一组DNA片段串联以获得串联体。所述方法可以包括对所述串联体进行单分子测序以获得所述串联体的第一序列。在一些实施例中, 可以使用纳米孔来进行单分子测序, 并且所述方法可以包括使所述串联体穿过纳米孔。随后, 当所述串联体穿过所述纳米孔时, 可以检测

第一电信号。所述第一电信号可以对应于所述串联体的第一序列。另外, 所述方法可以包括分析所述第一电信号以确定所述第一序列。可以将所述第一序列的子序列进行对齐以识别与所述第一组DNA片段中的每一个相对应的序列。

1. 一种方法,其包括:

从获自受试者的生物样品提取多个无细胞DNA片段,其中所述生物样品包括包含所述无细胞DNA片段的体液;

将第一组所述多个无细胞DNA片段串联以获得第一串联体,其中所述第一组的所述无细胞DNA片段具有不同的序列,并来自所述受试者基因组的不同位置;以及

对所述第一串联体进行单分子测序以获得所述第一串联体的第一序列,所述第一序列包括所述第一组多个无细胞DNA片段的不同序列作为所述第一序列内的子序列。

2. 根据权利要求1所述的方法,其还包括:

将所述第一串联体提供给测序装置作为进行所述单分子测序的一部分;以及

使用所述测序装置检测对应于所述第一串联体的多个信号,所述多个信号对应于所述第一串联体的所述第一序列。

3. 根据权利要求2所述的方法,其中所述测序装置包括第一纳米孔,所述方法还包括:

使所述第一串联体穿过所述第一纳米孔;以及

当所述第一串联体穿过所述第一纳米孔时检测第一电信号,所述第一电信号对应于所述第一串联体的所述第一序列,其中所述多个信号包括所述第一电信号。

4. 根据权利要求3所述的方法,其还包括通过计算机系统执行:

分析所述第一电信号以确定所述第一序列;以及

将所述第一序列的子序列与参考基因组对齐,以识别对应于所述第一组多个无细胞DNA片段中的每一个的片段序列。

5. 根据权利要求3所述的方法,其中使所述第一串联体穿过所述第一纳米孔包括:

使所述第一串联体的第一链穿过所述第一纳米孔;以及

随后使所述第一串联体的第二链穿过所述第一纳米孔。

6. 根据权利要求5所述的方法,其还包括:

通过计算机系统分析所述第一链和所述第二链的所述第一电信号以确定所述第一序列。

7. 根据权利要求3所述的方法,其中所述第一纳米孔是基底上的多个纳米孔中的一个。

8. 根据权利要求7所述的方法,其还包括:

将第二组所述多个无细胞DNA片段串联以获得第二串联体;

使所述第二串联体穿过所述多个纳米孔中的第二纳米孔;以及

当所述第二串联体穿过所述第二纳米孔时检测多个第二电信号,所述第二电信号对应于所述第二串联体的第二序列。

9. 根据权利要求2所述的方法,其还包括:

将荧光标记的核苷酸与所述第一串联体杂交,以及

检测荧光信号,所述荧光信号对应于核苷酸。

10. 根据权利要求1所述的方法,其还包括通过计算机系统执行:

将所述第一序列的子序列与参考基因组对齐,以识别对应于所述第一组多个无细胞DNA片段中的每一个的片段序列。

11. 根据权利要求10所述的方法,其还包括:

基于所述子序列的对齐确定所述第一组多个无细胞DNA片段中的每一个的大小。

12. 根据权利要求10所述的方法, 其中:

将所述第一组多个无细胞DNA片段串联包括将所述第一组多个无细胞DNA片段和第二组具有已知序列的DNA片段串联, 其中所述第二组DNA片段散布在所述第一组多个无细胞DNA片段之间, 以及

将所述第一序列的子序列进行对齐包括将子序列与所述已知序列对齐, 以识别所述第一串联体中的所述第二组DNA片段的位置。

13. 根据权利要求12所述的方法, 其中所述串联将所述第二组DNA片段中的一个DNA片段置于所述第一组多个无细胞DNA片段中的两个DNA片段之间。

14. 根据权利要求13所述的方法, 其还包括通过所述计算机系统执行:

基于识别在所述第一子序列之前的所述第二组DNA片段的所述已知序列中的一个, 通过所述计算机系统识别与所述第一组多个无细胞DNA片段中的第一DNA片段相对应的第一子序列的起始碱基; 以及

基于识别在所述第一子序列之后的所述第二组DNA片段的所述已知序列中的一个, 通过所述计算机系统识别与所述第一组多个无细胞DNA片段中的所述第一DNA片段相对应的所述第一子序列的终止碱基。

15. 根据权利要求12所述的方法, 其中所述第二组DNA片段的每个DNA片段包括少于或等于7个核苷酸的已知序列。

16. 根据权利要求12所述的方法, 其中所述第二组DNA片段中的每一个具有相同的已知序列。

17. 根据权利要求10所述的方法, 其中所述子序列的所述对齐包括:

将所述第一序列的滑动窗口与所述参考基因组对齐, 每个滑动窗口对应于与所述参考基因组对齐的子序列; 以及

识别所述滑动窗口的两个何时与所述参考基因组的不同区域对齐, 以区分所述第一组多个无细胞DNA片段的两个DNA片段的序列。

18. 根据权利要求17所述的方法, 其还包括通过所述计算机系统执行:

基于与所述参考基因组的所述不同区域对齐的所述两个滑动窗口确定所述第一组多个无细胞DNA片段的第一DNA片段的第一子序列的终点或起点。

19. 根据权利要求17所述的方法, 其中所述参考基因组的所述不同区域包括不同染色体上的区域。

20. 根据权利要求17所述的方法, 其中所述两个滑动窗口之间的一个或多个窗口被识别为不与所述参考基因组对齐。

21. 根据权利要求20所述的方法, 其还包括确定所述第一组多个无细胞DNA片段的第一DNA片段的大小, 其中确定所述第一DNA片段的大小包括:

确定与所述参考基因组的单个区域对齐的最长子序列的长度。

22. 根据权利要求1所述的方法, 其中所述生物样品是血浆或血清。

23. 一种方法, 其包括通过计算机系统执行:

接收通过串联来自获自受试者的生物样品的第一组多个无细胞DNA片段产生的第一串联体的第一序列, 其中所述生物样品包括包含所述无细胞DNA片段的体液, 以及所述第一组的所述无细胞DNA片段具有不同的序列, 并来自所述受试者基因组的不同位置; 以及

将所述第一序列的子序列进行对齐以识别对应于第一组所述多个无细胞DNA片段的每个DNA片段的片段序列。

24. 根据权利要求23所述的方法, 其中:

所述第一串联体是通过将所述第一组多个无细胞DNA片段和第二组DNA片段串联产生, 以及

将所述第一序列的子序列进行对齐包括将所述子序列与所述第二组DNA片段对齐。

25. 根据权利要求24所述的方法, 其中所述第二组DNA片段的每个DNA片段包括少于或等于7个核苷酸的已知序列。

26. 根据权利要求23所述的方法, 其中将所述第一序列的子序列进行对齐包括将所述第一序列的子序列与参考基因组对齐。

27. 根据权利要求26所述的方法, 其中所述子序列的所述对齐包括:

将所述第一序列的窗口与所述参考基因组对齐; 以及

识别所述窗口的两个何时与所述参考基因组的不同区域对齐。

28. 根据权利要求27所述的方法, 其中所述两个窗口之间的一个或多个窗口被识别为不与所述参考基因组对齐。

29. 一种计算机产品, 其包括计算机可读介质, 所述计算机可读介质存储用于控制计算机系统执行权利要求23到28中任一项所述方法中的步骤的多个指令。

30. 一种系统, 其包括:

根据权利要求29所述的计算机产品; 以及

一个或多个处理器, 其用于执行存储在所述计算机可读介质上的指令。

31. 一种系统, 其包括用于执行权利要求23到28中任一项所述方法中的步骤的装置。

32. 一种系统, 其被配置为执行权利要求23到28中任一项所述方法中的步骤。

33. 一种系统, 其包括分别执行权利要求23到28中任一项所述方法中的步骤的模块。

34. 计算机可读介质, 其存储多个指令, 其中所述指令被处理器执行时控制系统实现权利要求1-22任一项所述方法中的步骤。

35. 计算机可读介质, 其存储多个指令, 其中所述指令被处理器执行时控制系统实现权利要求23-28任一项所述方法中的步骤。

36. 系统, 其包括:

提取装置;

制备装置;

测序装置; 以及

计算机系统, 其中:

所述计算机系统包含存储多个指令的计算机可读介质, 其中所述指令被处理器执行时控制所述系统实现权利要求1-22任一项所述方法中的步骤。

37. 系统, 其包括:

提取装置;

制备装置;

测序装置; 以及

计算机系统, 其中:

所述计算机系统包含存储多个指令的计算机可读介质,其中所述指令被处理器执行时控制所述系统实现权利要求23-28任一项所述方法中的步骤。

血浆DNA的单分子测序

[0001] 相关申请的交叉引用

[0002] 本申请要求2015年8月12日提交的第62/204,396号美国临时申请的优先权,其内容通过引用并入本文用于所有目的。

背景技术

[0003] 通过母体血浆DNA测序进行的无创产前检测 (NIPT) 现在临床上可用于筛选胎儿染色体非整倍体 (1)。与羊膜穿刺术不同,母体血浆DNA测序不会造成任何流产风险。这些测试具有将近99%的灵敏度和99%的特异性 (1)。因此,自NIPT于2011年首次登陆市场以来,其临床需求大幅增加。

[0004] 大规模平行测序是目前大部分染色体非整倍体NIPT实验室方案的核心组成部分 (2)。由于仪器成本高,这些测试目前在参考实验室进行。牛津纳米孔技术公司 (Oxford Nanopore Technologies) 已经开发出一种基于纳米孔的DNA测序平台 (3)。纳米孔测序仪具有相对较低的设备成本,且占地面积小。每个流动池的成本为500-900美元,可以多次使用高达48小时。测序速度也相对较快,每秒可以从每个纳米孔中读取30个碱基。这些特征对于在临床实验室中使用是有利的。但是,目前的纳米孔技术对于通常用于NIPT的样品 (例如血浆) 来说是低效的。

发明内容

[0005] 当分析具有较小的DNA片段的样品时,实施例提高单分子测序技术的效率。例如,可以显著增加样品中DNA片段的浓度,从而允许更多的DNA片段与测序装置 (例如纳米孔) 相互作用。作为另一个实施例,将DNA片段组合成串联体以获得可读取的较长分子,从而允许通过对一个分子的测序有效地读取多个DNA片段 (即,最初在样品中的DNA片段)。生物信息学程序可用于检测属于同一串联体的一部分的不同DNA片段。实施例也可以将这两种技术组合。

[0006] 实施例可以包括一种确定核酸序列的方法。所述方法可以包括接收多个DNA片段。所述方法还可以包括将第一组DNA片段串联以获得第一串联体。所述方法可以包括对第一串联体进行单分子测序以获得第一串联体的第一序列。在一些实施例中,可以使用纳米孔来进行单分子测序,并且所述方法可以包括使第一串联体穿过第一纳米孔。随后,当第一串联体穿过第一纳米孔时,可以检测第一电信号。第一电信号可以对应于第一串联体的第一序列。

[0007] 其他实施例可以包括一种确定核酸序列的方法。所述方法可以包括接收多个DNA片段。第一组DNA片段可以被串联以获得第一串联体。荧光标记的核苷酸可以与串联体杂交。可以检测第一荧光信号,第一荧光信号对应于特定的核苷酸。之后荧光标记可以被切除,并且可以添加另一种荧光标记的核苷酸,所述过程可以重复。

[0008] 实施例可以包括通过计算机系统执行的方法。所述方法可以包括接收通过串联第一组DNA片段产生的第一串联体的第一序列。所述方法还可以包括将第一序列的子序列进

行对齐以识别对应于第一组所述DNA片段的每个DNA片段的片段序列。

[0009] 一些实施例可以包括一种无细胞DNA片段的测序方法。无细胞DNA片段可以包括血浆DNA片段。所述方法可以包括接收包括多个DNA片段的生物样品。生物样品可以具有第一DNA片段浓度。所述方法还可以包括浓缩生物样品以使其具有第二DNA片段浓度。第二DNA片段浓度可以是第一DNA片段浓度的5倍或更多倍。所述方法还可以包括使多个DNA片段穿过基底上的纳米孔。对于多个DNA片段中的每一个，当DNA片段穿过纳米孔时检测电信号。电信号可以对应于DNA片段的序列。

[0010] 实施例可以包括一种无细胞DNA片段的测序方法。所述方法可以包括浓缩生物样品以使其具有第二DNA片段浓度，所述第二DNA片段浓度是初始DNA片段浓度的5倍或更多倍。所述方法还可以包括单分子测序技术。DNA片段可以与荧光标记的核苷酸杂交。所述方法还可以包括检测来自荧光标记的核苷酸的信号，其中信号对应于核苷酸。荧光标记的核苷酸可以被切除，并且可以重复所述过程来识别另外的核苷酸和DNA片段的序列。

[0011] 实施例还可以包括计算机产品，其包括计算机可读介质，所述计算机可读介质存储多个指令以执行本文描述的任何DNA测序方法的操作。一些实施例可以包括计算机产品以及用于执行存储在计算机可读介质上的指令的一个或多个处理器。另外的实施例包括执行任何所述方法的系统。

[0012] 其他实施例涉及与本文描述的方法相关联的系统、便携式消费类设备和计算机可读介质。

[0013] 参考下面的详细描述和附图可以更好地理解本发明的实施例的本质和优点。

附图说明

[0014] 图1A示出了根据本发明实施例的纳米孔装置和核酸的简化图。

[0015] 图1B说明了根据本发明实施例的串联DNA片段的过程。

[0016] 图2描述了根据本发明实施例将DNA片段连同间隔DNA片段串联在一起的过程。

[0017] 图3A示出了根据本发明实施例的通过串联DNA片段和使用单分子测序对DNA片段进行测序的方法的简化方块流程图。

[0018] 图3B示出了根据本发明实施例的通过纳米孔测序对DNA片段进行测序的方法的简化方块流程图。

[0019] 图4示出了根据本发明实施例的分析串联体序列的方法的简化方块流程图。

[0020] 图5示出了根据本发明实施例的通过多次增加DNA片段的浓度来更有效地对DNA片段进行测序的方法的简化方块流程图。

[0021] 图6示出了根据本发明实施例的由纳米孔测序仪测序的血浆DNA库的大小分布图。绘制了从0到500个碱基对的测序的血浆DNA片段的频率分布图。

[0022] 图7示出了根据本发明实施例的由通过纳米孔测序和Illumina测序平台获得的数据得出的怀有女性胎儿的母体血浆的血浆DNA的大小分布图。

[0023] 图8示出了根据本发明实施例的与可映射的人类基因组预期的分布相比的染色体的读取分布。纳米孔序列读数对每个样品池的每个染色体的比例分布。填充的灰色条代表基于参考人类基因组的可映射部分hg19，来源于各个人类染色体的核苷酸的比例。剩余的彩色条代表与血浆DNA样品的各个人类染色体对齐的测序读数的比例。

[0024] 图9示出了根据本发明实施例的使用纳米孔测序和大规模平行测序对来自癌症患者的血浆DNA进行测序的结果的Circos图。

[0025] 图10示出了根据本发明实施例的串联DNA分子的大小分布图。

[0026] 图11示出了根据本发明实施例的来自非怀孕女性的串联片段的血浆DNA分子的大小分布图。

[0027] 图12示出了根据本发明实施例的来自非怀孕女性的串联血浆DNA的对齐片段的基因组代表。

[0028] 图13示出了根据本发明实施例的来自怀有男性胎儿的女性的串联片段的血浆DNA分子的大小分布图。

[0029] 图14示出了根据本发明实施例的来自怀有男性胎儿的女性的串联血浆DNA的对齐片段的基因组代表。

[0030] 图15示出了根据本发明实施例的执行系统的方块图。

[0031] 图16示出了根据本发明实施例的可用于系统和方法的示例计算机系统的方块图。

[0032] 术语

[0033] “组织”对应于一组作为功能单元的细胞。在一个组织中可以找到不止一种细胞。不同类型的组织可以由不同类型的细胞(例如肝细胞、肺泡细胞或血细胞)组成,但也可以对应于来自不同生物体(母体与胎儿)或健康细胞与肿瘤细胞的组织。

[0034] “生物样品”是指取自受试者的任何样品,并且含有一种或多种相关的核酸分子,其中受试者可包括例如人,如孕妇、癌症患者或怀疑患有癌症的人、器官移植接受者或怀疑患有涉及器官的病情(例如,心肌梗塞中的心脏,或中风中的脑,或贫血中的造血系统)。生物样品可以是体液,例如血液、血浆、血清、尿液、阴道液、来自鞘膜积液(例如睾丸)的液体、阴道冲洗液、胸膜液、腹水、脑脊液、唾液、汗液、泪液、痰液、支气管肺泡灌洗液、乳头排出液体、不同身体部位(如甲状腺、乳房)的吸引液等。也可以使用粪便样品。在各种实施例中,富含无细胞DNA的生物样品(例如通过离心方案获得的血浆样品)中大部分DNA可以是无细胞的,例如大于50%、60%、70%、80%、90%、95%或99%的DNA可以是无细胞的。离心方案可以包括,例如,在3,000g条件下离心10分钟以获得流体部分,并在例如30,000g条件下再次离心10分钟以除去残余细胞。样品中的无细胞DNA可以来源于各种组织的细胞,因此样品可以包括无细胞DNA的混合物。

[0035] “核酸”可以指单链或双链形式的脱氧核糖核苷酸或核糖核苷酸及其聚合物。所述术语可以涵盖含有已知核苷酸类似物或修饰的骨架残基或键的核酸,其是合成的、天然存在的和非天然存在的,其具有与参考核酸相似的结合特性,并且以与参考核苷酸类似的方式代谢。这样的类似物的示例可以包括,但不限于,硫代磷酸酯、亚磷酰胺、甲基膦酸酯、手性甲基膦酸酯、2-O-甲基核糖核苷酸、肽核酸(PNA)。

[0036] 除非另外指出,否则特定的核酸序列也隐含地包括其保守修饰变体(例如简并密码子取代)和互补序列,以及明确指出的序列。具体地,可通过产生一个或多个选定(或所有)密码子的第三位置被混合碱基和/或脱氧肌苷残基取代的序列来实现简并密码子取代(Batzer等Nucleic Acid Res.19:5081(1991);Ohtsuka等J.Biol.Chem.260:2605-2608(1985);Rossolini等Mol.Cell.Probes8:91-98(1994))。术语核酸可与基因、互补DNA(cDNA)、信使核糖核酸(mRNA)、寡核苷酸和多核苷酸互换使用。

[0037] 除了指天然存在的核糖核苷酸或脱氧核糖核苷酸单体之外,术语“核苷酸”可以理解为是指其相关的结构变体,包括衍生物和类似物,其就其中使用核苷酸的特定上下文而言在功能上是等同的(例如与互补碱基杂交),除非上下文另外明确指出。

[0038] “串联体”是由单独的DNA片段组成的连续的DNA分子,单独的DNA片段被合并到单个分子中。串联体的各个单独的DNA片段可以具有也可以不具有相同的序列。串联体中的至少一些DNA片段可以具有不同的序列。用于制造串联体的单独的DNA片段可以来源于生物样品中存在的各种组织,例如当DNA片段是无细胞DNA片段时,可能存在于血浆和其他无细胞DNA混合物中。

[0039] “序列读数”是指从核酸分子的任何部分或全部测序的一串核苷酸。例如,序列读数可以是生物样品中存在的整个核酸片段。序列读数可以由单分子测序获得。“单分子测序”是指对单个模板DNA分子进行测序以获得序列读数,而不需要从模板DNA分子的克隆拷贝中解读碱基序列信息。单分子测序可以对DNA分子的整个分子或一部分进行测序。DNA分子的大部分都可以被测序,例如大于50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%。

[0040] 对于测序来说,“信号”可以指瞬时测量值。这种信号的示例包括光信号或电信号。例如,当发生杂交时,光信号可以提供对应于与特定碱基对应的特定颜色的图像。单个图像可以是测序装置(例如,纳米孔或其他测序仪器)的阵列,由此具有许多光学信号的单个图像。电信号可以是瞬时电极间的测量值。一个时间段内的电信号可以提供DNA分子序列中的一个或多个碱基。电信号可以包括电流或电压信号。

[0041] “纳米孔”是指其中可以放置分子或分子的一部分的开口,其中可以基于纳米孔中的分子的部分的一个或多个特性来检测信号。纳米孔可以由各种材料组成,例如聚合物、金属或其他固态材料、蛋白质或其组合物。

[0042] “类别”是指与样品的特定属性相关联的任何数字(一个或多个)或其他字符(一个或多个)。例如,符号“+”(或文字“正”)可以表示样品被分类为具有缺失或扩增。类别可以是二元的(例如,正或负),或者具有更多级别的类别(例如,从1到10或从0到1的数值范围)。术语“截断”和“阈值”是指在操作中使用的预定数字。例如,截断大小可以指高于片段被去除的大小。阈值可以是高于或低于适用特定类别的值。这两个术语中的任何一个都可以在这两种情况下使用。

[0043] 本文使用的术语“染色体非整倍体”是指与二倍体基因组的定量数量不同的染色体的定量数量的变异。变异可以是增加或减少。它可能涉及整个染色体或染色体的一个区域。

[0044] 本文使用的术语“序列不平衡”或“畸变”意指来自参考量的临床相关染色体区域的量中的至少一个截断值定义的任何显著偏差。序列不平衡可能包括染色体剂量不平衡、等位基因不平衡、突变剂量不平衡、拷贝数量不平衡、单元型剂量不平衡以及其他类似的不平衡。举例来说,当肿瘤基因的一个等位基因缺失,或基因的一个等位基因扩增,或其基因组内的两个等位基因差异扩增,从而在样品中的特定基因座处产生不平衡时,可发生等位基因不平衡。又如,患者可能在肿瘤抑制基因中具有遗传突变。然后患者可能继续患上其中肿瘤抑制基因的未突变的等位基因缺失的肿瘤。因此,在肿瘤内,存在突变剂量不平衡。当肿瘤将DNA释放到患者的血浆中时,肿瘤DNA将与血浆中患者的组成性DNA(来自正常细胞)

混合。通过使用本文所述的方法,可以检测血浆中所述DNA混合物的突变剂量不平衡。畸变可以包括染色体区域的缺失或扩增。

[0045] 术语“大小分布图”通常涉及生物样品中DNA片段的大小。大小分布图可以是提供各种大小的DNA片段的量分布的直方图。可以使用各种统计参数(也称为大小参数或参数)来区分一个大小分布图与另一个大小分布图。一个参数是相对于所有DNA片段或相对于具有另一个大小或范围的DNA片段的特定大小或大小范围的DNA片段的百分比。

具体实施方式

[0046] 当应用于生物样品(例如血浆或血清)中的无细胞DNA片段时,实施例可以提高单分子测序的效率,所述无细胞DNA片段是约200个碱基的相对较短的片段。由于无细胞DNA片段(如血浆DNA片段)较小或较短,且通常以低浓度存在于生物样品中,因此使用纳米孔对无细胞DNA片段进行测序可能不会提供准确的结果。由于小片段穿过小孔,使得当片段在纳米孔中时很难对片段进行测序。另外,考虑到预期的测序误差,对于小片段,与参考基因组对齐可能更困难。纳米孔测序可能具有约10-15%的测序误差。在实施例中,通过DNA片段的串联以产生更长的待测序分子可以提高效率。在其他实施例中,通过增加样品中DNA片段的浓度可以提高效率。

[0047] 没有这些发展,对血浆DNA的单分子测序可能是无法实施的。采用这些方法,通过单分子测序可检测出一系列血浆DNA畸变。实施例可应用于医学领域的许多领域,包括产前诊断、癌症评估、炎症性疾病管理、自身免疫疾病评估以及急性医学,如创伤。此外,据报道,甲基化的胞嘧啶可以通过单分子测序与未甲基化的胞嘧啶区分开来(12)。我们之前曾报告通过检测血浆DNA的甲基化谱,我们能够确定胎儿DNA分数,检测胎盘甲基化异常的妊娠相关疾病,以及检测与癌症和系统性红斑狼疮有关的异常甲基化(7,9,11)。因此,实施例也可以用于检测甲基化的应用中。

[0048] I. DNA片段的单分子测序

[0049] 血浆或血清DNA是在人类受试者循环中发现的无细胞核酸分子,在细胞死亡过程中释放,并作为自然转换或病理过程的一部分。因为DNA分子作为细胞降解过程的一部分被释放到循环中,所以它们以短片段(<200bp)的形式循环并且以低浓度存在。已知健康受试者中的大部分血浆DNA来源于血液细胞(6)。

[0050] 在怀孕期间,胎盘将血浆DNA分子送入母体循环,并提供一种获取胎儿DNA的手段,用于无创性产前诊断(7)。肿瘤和癌症具有高细胞更新率,并将DNA提供给血浆,在血浆中染色体和遗传异常可以被无创检测,并用作癌症诊断、监测、筛选和预测的液体活检(8,9)。诸如心肌梗塞、中风、肝炎等炎症性疾病也会导致发炎器官的血浆DNA增加(10)。自身免疫性疾病,如系统性红斑狼疮,也与血浆DNA异常有关(11)。在上述疾病中,血浆DNA图谱可显示与疾病相关的单核苷酸变体、染色体或亚染色体拷贝数畸变(拷贝数增加或减少、非整倍体)、异常甲基化(过甲基化或低甲基化)和大小分布异常(过量的短或长DNA分子)。总之,循环无细胞DNA分析已经成为开发针对多种疾病的分子诊断试验的重要手段。

[0051] 由于其可以高速识别每个DNA碱基,单分子测序(不限于由牛津纳米孔技术公司制造的格式)是DNA测序分析的一个有吸引力的平台。由于省略了扩增步骤,文库的构建流程被简化。它可以对连续的DNA片段进行测序,最多可达几百个千碱基。但是,所述优点不容易

应用于血浆DNA分析。首先,血浆DNA分子是短片段,长度大多小于200bp。血浆和血清中的DNA浓度大大低于组织活检中的浓度。因此,当将血浆DNA样品应用于纳米孔测序仪时,测序效率低。换句话说,样品中的DNA太稀,且血浆DNA分子也不容易进入纳米孔。即使当血浆DNA分子进入了纳米孔并被测序,由于其长度较短,测序信息也只能提供关于人类基因组的极少量的信息。单倍体人类基因组大小为 3.3×10^9 个碱基。此外,人们会认为使用纳米孔对短DNA片段进行测序的测序误差率比对长DNA片段进行测序时更高。因此,我们的目标是开发方法来增加纳米孔和其他单分子测序装置处理血浆DNA分子的频率或机会。

[0052] A. 纳米孔测序

[0053] 纳米孔测序是单分子测序仪的一种形式,其中通过使用纳米孔来检测DNA碱基。牛津纳米孔技术公司使用蛋白孔和 α -溶血素。这种孔的基体被制造成坐落在膜上(5)。除了蛋白孔之外,纳米孔可以是固态纳米孔,其中纳米孔在半导体材料(包括硅化合物,例如氮化硅和石墨烯)中制造。每个孔可以连接到电路。

[0054] 图1A示出了根据本技术实施例的纳米孔102和DNA分子104的简化图。电极106和电极108可以限定纳米孔的一部分或者可以位于纳米孔附近。电极106和电极108被示出为具有可以是圆锥形或三角形的端部。在一些实施例中,电极可以具有不同的形状。例如,电极可以具有平坦的端部、半球形的端部或圆形的端部。两个电极可能不具有相同的形状。例如,一个电极可以具有圆锥形的端部,而另一个电极可以具有平坦的端部。电极之间的距离可以等于、小于或大于纳米孔的直径或宽度。电极106和电极108可以连接到电源110。电流112可以从电极106隧穿到电极108。电源110可以与多个纳米孔和各自的一对电极电连通。

[0055] 当DNA分子104穿过纳米孔102时,电流会有变化,这可以通过仪表114来测量。不同的DNA碱基,即A、C、G、T会引起不同幅度的电流变化。通过观察每个纳米孔的电压或电流模式,可以确定已经穿过纳米孔102的DNA分子104的序列。因为测序足够敏感以检测单个DNA分子上的DNA碱基,即不需要扩增DNA测序文库,所以序列检测的速度大大提高。

[0056] 然而,缺点是DNA碱基识别的准确性比检测来自每个DNA分子的扩增克隆的共有序列的测序技术(如Illumina测序)相对较差。但是,已经显示,当解读2D读数时,碱基检测精度显著提高(3)。当制备用于纳米孔测序的DNA样品时,将发夹接头加装到每个双链DNA分子的一端。当这样的DNA分子接近纳米孔时,双链将剥离,并且当检测到电流变化时,现在的单链DNA分子的一端将穿过纳米孔。当测序接近所述单端时,由发夹连接的互补链将继续通过纳米孔并被测序。当DNA分子的两条链被测序时,可以得到共有序列,这被称为2D读数。正如在文献中所报道的,2D读数与1D读数相比具有更高的碱基调用精度(仅从单链解读的序列)。

[0057] B. 其他单分子测序

[0058] 除了使用纳米孔的技术之外,实施例还可以包括单分子测序技术。例如,使用Helicos单分子测序仪(SeqLL),可以将DNA分子杂交到玻璃表面上。然后可以将荧光标记的核苷酸添加到每个DNA分子,并且可以捕获图像。之后荧光分子可以被切除冲走,还可以添加另一种荧光标记的核苷酸并重复所述过程。每个核苷酸可以具有不同的荧光标记,使得DNA被测序。

[0059] 另一个示例可以包括太平洋生物科学公司的单分子实时(SMRT)测序方法。在所述方法中,可以将DNA聚合酶附加到零模波导(ZWM)的底部。然后聚合酶可以捕获单个DNA分

子。然后可以将荧光标记的核苷酸并入DNA-酶复合物中。检测器然后可以检测荧光信号,并且可以进行碱基调用。荧光标签然后可以被切割并且可以扩散出ZWM。所述过程可以重复进行,每种类型的核苷酸具有不同的荧光标记,使得DNA被测序。

[0060] II. 提高效率

[0061] 为了提高效率,一种选择是扩大血浆DNA库以在有限体积内增加样品中的遗传物质的量,以用于测序室或流动池。然而,采用这样的方案意味着不再进行单分子测序,因为现在人们将对原始DNA模板和所有复制DNA片段进行测序。当试图获得关于原始DNA片段的定量信息,例如在给定的基因组区域(例如整个染色体)中的原始DNA片段的拷贝数时,这种复制DNA片段的使用可能引起误差。

[0062] 在下面的部分中,描述了两种提高效率的方法。在一种方法中,可以在进行单分子测序之前将DNA片段串联以形成串联体。在另一种方法中,在单分子测序之前(例如在纳米孔测序之前),将DNA片段高度集中在样品中。

[0063] A. 串联体

[0064] 在一些实施例中,DNA片段(例如血浆DNA片段)可以被串联,如图1B所示。串联可以将一系列DNA片段彼此连接。例如,DNA片段152和DNA片段154可以存在于样品中。作为将DNA片段彼此连接的初始步骤,可以在末端添加磷酸基团来制备DNA片段152和DNA片段154的末端,从而产生DNA片段156和DNA片段158。然后通过平端接合的连接酶将DNA片段连接在一起以形成长DNA分子,例如串联体160。串联体160可被认为是由DNA片段152和154组合而成的新分子。串联体160将具有对应于DNA片段152和154的子序列的序列。

[0065] 图2示出了形成串联体的类似方法,在DNA片段之间具有间隔片段(也简称为“间隔子”)。可以对DNA片段202和DNA片段204进行“加A尾”,其中可以将A核苷酸添加到DNA片段的每条链的一端,形成DNA片段206和DNA片段208。间隔DNA片段210可以具有在间隔DNA片段的每条链的一端具有T核苷酸的已知序列。间隔DNA片段的已知序列可以低于20个碱基对,包括4至10个碱基对和10至20个碱基对,不包括添加至间隔片段末端的A或T核苷酸。其他互补核苷酸也可以在间隔DNA片段和血浆DNA片段的末端使用。然后可通过连接酶将间隔DNA片段210连接至DNA片段206和DNA片段208以形成显示为串联体212的长DNA分子。间隔DNA片段可以放置在两个不是间隔子的DNA片段之间。例如,一个间隔DNA片段可以位于从生物样品中提取的两个DNA片段之间。串联体212可以具有由对应于DNA片段206和208以及一个或多个间隔DNA片段210的序列组合而成的序列。

[0066] 在一些实施例中,间隔片段可以指示一个DNA片段的末端和另一个DNA片段的起点,例如,当间隔片段的已知序列不出现在对应于受试者的参考基因组中(例如人类基因组)或出现少于特定次数(例如少于2或3次)时。计算机系统可以通过将子序列与用于间隔片段的一个或多个已知序列的预期组进行比较来识别间隔片段的序列。在一些实施方式中,间隔片段可以是单个位置处的A-T组合。在这样的实施方式中,A和T可以主要用于附着,而不是用于识别DNA片段何时开始和结束。

[0067] 具有间隔子的方法可以包括识别对应于生物样品的DNA片段的子序列的起始碱基。识别起始碱基可以基于识别间隔DNA片段的已知序列之一。基于子序列的结束可以基于在子序列之后识别间隔DNA片段的已知序列之一而被识别。

[0068] 利用这些方法,当长分子(串联体)到达纳米孔时,可以对多个血浆DNA片段进行测

序,因为串联体可以在单个长分子中包括几个血浆DNA片段。由于血浆DNA分子可以是天然的双链,发夹接头可以应用在串联体的末端,从而可以产生2D读数,即读出两条链。测序后,可以识别并入串联体的血浆DNA分子的原始单位。由几个DNA片段创建长分子并对纳米孔中的长分子进行测序比等待相同的几个分子分别移动到纳米孔中以进行测序更有效。

[0069] 1. 单分子测序中的应用

[0070] 图3A示出了根据本发明实施例的通过串联DNA片段和使用单分子测序对DNA片段进行测序的方法300的简化方块流程图。

[0071] 在方块302处,方法300可以包括接收多个DNA片段。所述多个DNA片段可以是来自生物样品的无细胞DNA片段。生物样品可以是血浆或血清。DNA片段可以与生物样品的其他组分分离,例如,血浆与其他血液成分分离。DNA片段可以是本文所述的任何DNA片段。

[0072] 在方块304处,方法300可以包括将第一组DNA片段串联以获得第一串联体。串联可以通过本文描述的任何方法进行。第一串联体可以包括除第一组以外的DNA片段。因此,第一组DNA片段可能不是构成第一串联体的全部DNA片段。第一串联体中的其他DNA片段可以包括间隔DNA片段。

[0073] 方块306示出方法300还可以包括对第一串联体进行单分子测序以获得第一串联体的第一序列。单分子测序可以包括纳米孔测序。第一串联体可作为进行单分子测序的一部分提供给测序装置。测序装置可以是纳米孔装置、光波导、配置成使得串联体杂交在其上的流动池、或者发生单个DNA分子的序列检测的任何反应池或位置。流动池的示例可以包括具有粘附到流动池表面的寡核苷酸的流动池,其中寡核苷酸可以与第一串联体杂交。方法300还可以包括使用测序装置检测对应于第一串联体的多个信号。所述多个信号可以对应于第一串联体的第一序列。信号可以包括来自荧光标记或具有结合到第一串联体的光学可检测信号的其他标记的信号。来自荧光标记的信号可以由光学检测器、激光器、电荷耦合器件、零模波导、可以确定是否存在光学事件的其它光敏器件或其组合来检测。

[0074] 2. 检测纳米孔中的电信号

[0075] 图3B示出了根据本发明实施例的通过纳米孔测序对DNA片段进行测序的方法350的简化方块流程图。在方法350中,测序装置是纳米孔,其可以存在于基底上的纳米孔阵列中。方法350的各个方面可以在实施方法300时执行。

[0076] 在方块352处,方法350可以包括接收多个DNA片段。所述多个DNA片段可以是来自生物样品的无细胞DNA片段。生物样品可以是血浆或血清。DNA片段可以与生物样品的其他组分分离,例如,血浆与其他血液成分分离。

[0077] DNA片段可以在用于串联的容器中接收。容器可以是小瓶或管,例如Eppendorf管。DNA片段可以与连接酶和缓冲液在容器中混合。

[0078] 方块354示出了方法350还可以包括将第一组DNA片段串联以获得第一串联体。串联可以通过本文描述的任何方法进行。第一串联体可以包括除第一组以外的DNA片段。因此,第一组DNA片段可能不是构成第一串联体的全部DNA片段。

[0079] 可以并行执行多个串联过程,每个过程形成单独的串联体。多个串联体可以具有不同的长度,例如,因为不同的DNA片段可以具有不同的长度,并且因为不同数量的DNA片段可以被并入串联体中。例如,一个串联体可以由3个DNA片段组成,而另一个串联体可以由100个DNA片段组成。

[0080] 在方块356中,方法350可以进一步包括使第一串联体穿过第一纳米孔。第一纳米孔可以是基底上的多个纳米孔中的一个。使第一串联体穿过第一纳米孔可以包括使第一串联体的第一链穿过纳米孔。在第一链穿过纳米孔之后,第一串联体的第二链可以穿过纳米孔。第一链和第二链可以通过发夹接头连接。以这种方式,两条链均可以被测序,这可以为碱基调用提供更高的精度。两条链的电信号可以作为碱基调用的一部分进行比较。

[0081] 在方块358中,当第一串联体穿过第一纳米孔时,可以检测第一电信号。第一电信号可以对应于第一串联体的第一序列。第一电信号可以包括电流或电压。在没有第一串联体穿过第一纳米孔的情况下,离子电流可以在电极之间通过。当诸如串联体等生物分子通过电极之间时,生物分子可能会影响电极之间离子或电子的通过。结果,电流或电压可能下降。变化的大小可能与电极之间的生物分子部分有关。例如,当特定的核苷酸或官能团在电极之间通过时,电流或电压可能具有特定的电信号特征。

[0082] 在一些实施例中,纳米孔可以是包括两个电极的电路的一部分。两个电极之间的电流可以基于哪个核苷酸(碱基)或相应的标签在纳米孔中而变化。可以使用用于测量电路中的电压或电流的任何合适的技术来检测第一电信号。

[0083] 在方块360中,方法350可以包括分析第一电信号以确定第一序列。分析可以包括将电信号的模式与对应于特定碱基的已知模式进行比较。如Schreiber J.和Karplus K.在Bioinformatics 2015 31:1897-1903发表的“使用隐性马尔可夫模型分析纳米孔数据(Analysis of nanopore data using hidden Markov models)”中所讨论的,纳米孔碱基调用可以包括使用不同的模型,包括隐性马尔可夫模型,其通过引用并入本文以用于各种目的。分析可以包括通过计算机系统分析第一链和第二链的第一电信号以确定第一序列。例如,可以确定第一链和第二链的序列,并且可以将两个序列彼此比较。序列应该是互补的。例如,不能互补的位置可以被忽略或重新分析。

[0084] 在一些实施例中,可以使用第一电信号的分析来确定第一串联体的各个位点(例如CpG位点)的甲基化分类。甲基化分类可以包括碱基是否甲基化、异常甲基化(过甲基化或低甲基化)是否存在(例如,诸如CpG岛的区域是否具有异常甲基化)以及所述串联体是否被羟基甲基化。

[0085] 在方块362处,方法350可以包括将第一序列的子序列进行对齐以识别对应于第一组DNA片段的每个DNA片段的片段序列。子序列可以是第一序列中的任何连续碱基的集合,例如,如滑动窗口所指定的。可以针对参考基因组来进行对齐,其可以允许与参照基因组对齐中的错配。下面更详细地描述子序列的对齐。

[0086] 可以对第二组DNA片段进行串联,以获得第二串联体,以及串联其他组DNA片段以获得其他串联体。每个串联体可以具有不同于其他串联体的DNA片段的组合或排列。每个串联体可以穿过可以包括测序装置的多个纳米孔的第一纳米孔或其他纳米孔。其他电信号可以在每个其他串联体穿过纳米孔时获得。其他电信号可以对应于其他串联体的各自的序列。涉及其他串联体的细节可能与涉及第一个串联体的方法类似。

[0087] 方法350还可以包括确定第一串联体的第一组DNA片段中的每一个的大小。也可以确定其他串联体的其他组DNA片段的DNA片段的大小。作为一个示例,DNA片段的大小可以通过将子序列与参照基因组或者已知的间隔序列对齐来确定。例如,如果可以识别间隔序列,则可以将DNA片段的长度识别为两个间隔序列之间的碱基数。在使用间隔序列的这样的实

施例中,DNA片段的识别序列不需要与参考基因组对齐以进行识别,因为间隔序列可以提供这样的信息。此外,DNA片段的序列可以在与参照基因组对齐后或代替与参考基因组对齐来进行组装。当与参考基因组对齐时,确定DNA片段的大小可以包括确定与参考基因组的单个区域对齐的最长子序列的长度。

[0088] 3. 对齐子序列

[0089] 如图4所示,实施例可以包括通过计算机系统执行的方法400。图4示出了根据本发明实施例的分析串联体序列的方法的简化方块流程图。

[0090] 在方块402处,方法400可以包括接收通过串联第一组DNA片段产生的第一串联体的第一序列。在一些实施例中,第一串联体可以通过将第一组DNA片段和第二组DNA片段串联而产生。第一序列可以从可以驻留在测序装置上的碱基调用程序接收,计算机系统也可以驻留在所述测序装置上。作为另一个示例,计算机系统可以与测序装置分开,并且第一序列可以通过网络连接或通过可移动存储装置来接收。所述串联体可以是任何串联体,例如,如本文所述的串联体。

[0091] 在方块404处,方法400还可以包括将第一序列的子序列进行对齐以识别对应于第一组DNA片段的每个DNA片段的片段序列。在一些实施例中,将子序列进行对齐可以包括将子序列与第二组DNA片段对齐。如本文所述,第二组DNA片段可以是间隔DNA。间隔DNA可以是已知序列,其可以是小于或等于20个核苷酸,包括15至20、10至15以及5至10个核苷酸。

[0092] 在一些实施例中,方法400可以包括将第一序列的子序列与参考基因组对齐。参考基因组可以是人类基因组。为了识别血浆DNA分子的原始单位,实施例可以通过窗口将长DNA序列与人类基因组对齐。例如,可以从串联体的长序列中选择滑动窗口(例如100-300个碱基),并且可以将窗口序列(子序列)与参照基因组对齐。参考基因组可以是参考人类基因组的衍生物,例如,但不限于人类基因组序列的亚单位、重复掩蔽的基因组、外显子组或具有中等或均衡GC含量的基因组部分。

[0093] 窗口可以向前或向后移动(滑动)的量小于窗口的长度(例如20-50个碱基)。窗口在新的位置可能被认为是第二个窗口。第二个窗口的子序列也可以与参考基因组对齐。如果第二个窗口的子序列与参考基因组的子序列(其与参考基因组的之前的子序列重叠)对齐,则这两个子序列可被认为是同一DNA片段的一部分。如果两个滑动窗口与参考基因组的不同区域、非连续区域或非重叠区域对齐,则可以区分两个DNA片段的序列。

[0094] 如果子序列不与基因组对齐,而是之前或之后的子序列与基因组对齐,则可以识别两个DNA片段之间的交线(边缘)。可以分析两个DNA片段对齐的区域以确定交线的特定点(例如,一个DNA片段的起始碱基和另一个DNA片段的末端碱基)。所述交线可以是DNA片段的子序列的末端或起点。当串联体的结构中不使用间隔子时,这种方法可能是特别有用的。在其他实施例中,可以将特定的序列(例如特定的条码或间隔子)添加到原始分子的末端,那些特定的序列可以指示一个分子的末端和另一个分子的起点。

[0095] 因此,实施例可以找到属于人类基因组上不同区域的DNA碱基的片段或区段(通常长达数百个碱基)。DNA碱基的每个连续的片段或区段可以代表一个原始的血浆DNA分子。与人类基因组的不同远端部分对齐的相邻并置DNA片段或区段可能属于组装成串联体的其他血浆DNA分子。

[0096] 窗口的大小和窗口移动或滑动的步长的大小可基于DNA片段的大小的期望分辨率

来调整。较小的步长可增加确定的DNA片段大小的分辨率,同时增加计算强度。较大的窗口大小可能不识别小于窗口大小的DNA片段,而较小的窗口大小可能不会实现与基因组的唯一对齐。

[0097] 窗口大小和步长可以动态调整。例如,可以使用大步长来缩小潜在匹配的区域,然后可以减小步长以更精确地识别匹配。可以将子序列与参考基因组的染色体或染色体区域对齐。将子序列进行对齐可以包括使用错配的数量和频率来解释测序误差。例如,对齐序列可以允许小于或等于约10-15%的错配。

[0098] 如下所述,这种方法在对血浆DNA分子进行测序、识别人类染色体、检测每个染色体的核苷酸含量之间的比例差异以及在串联之前确定原始血浆DNA样品的大小分布方面是有效的。串联体可以包括来自整个基因组的DNA片段。串联体中的DNA片段可以从全部或几乎全部染色体中随机分布。

[0099] B. 增加浓度

[0100] 另一个实施例可以包括增加加载到流动池的样品室上的血浆DNA文库的浓度。通常不会期望增加血浆DNA的浓度会提高纳米孔测序的效率。纳米孔具有相对较高的测序误差,并且,考虑到从血浆和其他无细胞样品分析小DNA片段以及在这样的无细胞样品中的低DNA浓度,并未期望纳米孔可以有效地对片段进行测序。DNA片段浓度的增加并未期望可以解决这个问题。然而,通过浓缩提取的DNA或输入测序文库,DNA分子到达纳米孔或其他单分子分析技术的几率得到增加。浓缩提取的DNA涉及浓缩超出获得与测序技术相容的体积所需的水平。在一些情况下,提取的DNA的浓度可能增加超过10倍。换句话说,体积可能会降低到原来的10%以下。

[0101] 图5示出了根据本发明实施例的通过多次增加DNA片段的浓度来更有效地对DNA片段进行测序的方法500的简化方块流程图。方法500可以用于各种单分子测序平台。在提供的示例中,描述了一种纳米孔测序平台。

[0102] 在方块502中,方法500可以包括接收包括多个DNA片段的生物样品。生物样品可以在起始体积中具有第一DNA片段浓度。生物样品可以是各种类型的,例如,如本文所述的类型。例如,生物样品可以是血浆或血清。

[0103] 在方块504中,方法500还可以包括浓缩生物样品以使其具有第二DNA片段浓度。作为各种示例,第二DNA片段浓度可以是第一DNA片段浓度的5倍以上、6倍以上、7倍以上、8倍以上、9倍以上、10倍以上、50倍以上、100倍以上、500倍以上或1000倍以上。浓度可以按体积或质量测量。

[0104] 正如本领域技术人员所理解的那样,浓缩可以以各种方式完成。例如,浓缩生物样品可以通过真空干燥、通过渗滤或过滤除去流体、或通过本领域技术人员已知的其他浓缩技术。过滤或渗滤可以与离心结合并使流体通过尺寸过滤器或分子筛。可以允许流体单向流动的半透膜也可以用于浓缩。浓缩后的生物样品的体积可以与浓度的增加成反比降低。例如,如果浓度增加了5倍,则体积可能减少为1/5。

[0105] 浓度可能比使用传统方法增加更明显。在一些常规方法中,从一定体积的血浆中提取较小体积的血浆DNA。为了减小体积以满足分析仪器中的反应体积或其他要求,体积可以进一步浓缩。例如,在常规方法中,可从4mL血浆中提取210 μ L的血浆DNA。可以将210 μ L的血浆DNA浓缩至85 μ L以提供100 μ L的总反应体积。使用常规方法,浓度的增加小于3倍,并且

血浆DNA没有被浓缩以提高测序准确度和精度。在本发明的方法中,增加的浓度可能导致DNA片段更频繁地穿过纳米孔或其他测序装置,因此改进了检测和分析。

[0106] 在方块506中,方法500还可以包括使多个DNA片段穿过基底上的纳米孔。在一些实施例中,方法可以包括除了纳米孔之外的单分子测序技术。例如,测序技术可以包括太平洋生物科学公司(Pacific Biosciences)的SMRT技术或SeqLL的Helicos测序。单分子测序技术可以包括Eid J.等在Science 2009 323:133-138发表的“单个聚合酶分子的实时DNA测序(Real-time DNA sequencing from single polymerase molecules)”中描述的技术,所述技术的内容通过引用并入本文以用于各种目的。

[0107] 在方块508中,对于多个DNA片段中的每一个,当DNA片段穿过纳米孔时检测电信号。电信号可以对应于DNA片段的序列或子序列。电信号可以包括电流或电压信号或本文所述的任何电信号。当使用其他测序技术时,可以使用荧光信号代替电信号。

[0108] 在方块510中,方法500可以包括分析电信号以确定DNA片段的序列或子序列。方法500可以包括例如通过使用对齐信息来确定所述DNA片段的大小和多个DNA片段的大小。因此也可以确定DNA片段的大小分布。基于DNA片段的大小分布,也可以确定染色体差异,例如,如2010年11月5日提交的标题为“基于尺寸的基因组学(Size-based genomics)”的美国申请序列号12/940,992;2011年11月30日提交的标题为“与癌症有关的遗传或分子畸变的检测(Detection of genetic or molecular aberrations associated with cancer)”的美国申请序列号13/308,473;以及2013年3月7日提交的标题为“母体血浆中胎儿DNA分数的基于尺寸的分析(Size-based analysis of fetal DNA fraction in maternal plasma)”的美国申请序列号13/789,553中所描述的。

[0109] 在一些实施例中,电信号可以对应于第一串联体的甲基化分类。甲基化分类可以包括碱基是否甲基化、异常甲基化(过甲基化或低甲基化)是否存在以及所述串联体是否被羟基甲基化。

[0110] 方法可以包括将DNA片段的序列或子序列与参考基因组对齐。具体而言,可以与参考基因组的特定染色体或染色体区域对齐。

[0111] 相同的DNA片段可以多次穿过相同的纳米孔。每次穿过,都可以检测到电信号。来自不同通路的电信号可以进行比较以帮助识别序列。增加DNA片段的浓度可以与除了纳米孔以外的单分子测序技术一起使用。

[0112] III. 使用增加浓度的示例

[0113] 示例显示血浆DNA可以浓缩以增加纳米孔测序的效率,同时提供准确的结果。Cheng S.H.等在“通过母体血浆DNA的纳米孔测序进行的无创产前测试:可行性评估(Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: feasibility assessment)”中进一步描述了使用增加浓度的测序,该文发表在Chem.61:10(2015)。

[0114] A. 材料和方法

[0115] 从基于知情同意和机构批准招募的四组个体中获得血浆样品,即怀有男性胎儿的晚期妊娠孕妇、怀有女性胎儿的晚期妊娠孕妇、成年男性和未怀孕女性。将EDTA血浆样品合并并在每组内以提供每组至少20mL的血浆。合并的血浆样品使用QIAamp DSP DNA血液微型试剂盒(Qiagen,Germany)提取(2)。用美国马萨诸塞州沃尔瑟姆赛默飞世尔科技有限公司

(Thermo Fisher Scientific, Waltham, MA) 的Speedvac浓缩器将洗脱的每池1050L的血浆DNA浓缩至85L。每个浓缩的血浆DNA池被完全消耗,用于使用美国马萨诸塞州伊普斯威奇纽英伦生物技术公司(New England Biolabs, Ipswich, MA) 的末端修复和加A尾模块和基因组DNA测序试剂盒(SQK-MAP-005, 牛津纳米孔技术公司) 制备DNA文库。将每个文库(150L) 完全加载到MinION流动池(v7.3) (纳米孔) 上并测序。使用METRICOR™软件(纳米孔) 对输出数据文件进行碱基调用。提取2D读数并使用日本计算生物学研究财团(Computational Biology Research Consortium, Japan) 的LAST基因组规模序列对齐软件与参考基因组hg19对齐。

[0116] 测序运行直到每个文库被消耗,持续6至24小时。26.9%-32.5%的读数通过碱基调用器。怀有男性胎儿的孕妇、怀有女性胎儿的孕妇、成年男性和未怀孕女性的血浆池的2D读数分别为56,844、50,268、358,78和36,167。平均观察到的同一性,与参考序列中的匹配碱基对齐的读数中的碱基比例(3) 为82.7% (81.4-84.5%)。在2D读数中,16.9% (15.6-23.9%) 与唯一的基因组位置对齐,并进一步分析。

[0117] 与人类基因组对齐的测序的血浆DNA片段的长度为76至5,776bp,并在162bp (155-168bp) 处达到峰值(图6)。图6中的每个图表显示在x轴上碱基对中测序的血浆DNA片段的大小,以及血浆DNA片段大小的频率占总测序血浆DNA片段的百分比。所述图是从怀有男性胎儿的母体血浆、怀有女性胎儿的母体血浆、男性血浆和未怀孕女性血浆中测序的DNA的结果。四种血浆中最大的血浆DNA大小与我们以前的基于Illumina测序平台的研究结果一致(4)。从纳米孔测序数据观察到极少量(0.06-0.3%) 的长血浆DNA片段(>1,000bp),但是从其他测序平台的先前的数据分析并没有观察到(4)。

[0118] 图7示出了由通过纳米孔测序和Illumina测序平台获得的数据得出的怀有女性胎儿的母体血浆的血浆DNA的大小分布图。纳米孔测序数据与来自图6的怀有女性胎儿的母体血浆数据相同。图7中的大小分布图具有相似的形状,峰所在的尺寸大致相同。例如,对于纳米孔测序,小于或等于150bp的片段与161-170bp的片段的大小比率为1.21,而Illumina测序为1.10。这些结果表明,使用纳米孔和浓缩血浆DNA可以精确确定血浆DNA的大小分布。在纳米孔测序获得的数据中,250-400bp范围内的峰比Illumina测序更为突出。所述峰对应于来源于二核体的无细胞DNA,并且峰的出现随个体而变化。此外,Illumina测序可能在这个尺寸范围内测序片段的效率较低。

[0119] B. 读数分析

[0120] 图8示出了与从可映射的人类基因组预期的分布相比,染色体的读数分布,标记为hg19。染色体被列在x轴上。在y轴上,通过计算从所述样品测序的唯一对齐的读数的总数中与每个染色体对齐的读数的数目,其表示为频率和百分比,来计算每个样品的读数对每个染色体(基因组代表) 的比例分布。用hg19绘制的是来自怀有男性胎儿的母体血浆、怀有女性胎儿的母体血浆、男性血浆和非怀孕的女性血浆的结果。所有四种血浆DNA库的常染色体的读数分布与可映射的人类基因组所预期的相当。

[0121] 在读数分布中观察到X染色体和Y染色体的差异。与女性血浆(5.22%) 相比,男性血浆中映射到X染色体的读数比例较低(2.70%)。在成年男性(0.30%) 血浆DNA库中检测到Y染色体序列,但在未怀孕的女性血浆DNA库中并未检测到。

[0122] 怀有男性胎儿的女性血浆DNA库有0.11%的读数与Y染色体对齐。与之前的数据(2) 一致,怀有女性胎儿的女性血浆DNA库有0.018%的读数与Y染色体序列对齐。在怀有女

性胎儿的女性中与Y染色体对齐的读数的存在可能是由与男性基因组对齐导致的已知误差。怀有男性胎儿的母体血浆DNA库的X染色体序列比怀有女性胎儿的女性小约1%。

[0123] 观察到X染色体和Y染色体的读数的相对分布与预期的结果类似。男性血浆比怀孕和未怀孕女性的血浆具有更多Y染色体。男性血浆的X染色体少于怀孕和未怀孕女性的血浆。男性血浆的X染色体数量约为非怀孕女性的一半,这是因为男性只有一个X染色体,而女性有两个X染色体。怀有男性胎儿的母体血浆Y染色体的读数分布高于怀有女性胎儿的母体血浆和女性血浆。

[0124] 因此,通过纳米孔测序和浓缩血浆DNA可以检测到男性和女性胎儿之间的胎儿DNA序列和X染色体剂量差异。男性胎儿的X染色体剂量相当于患有X单体综合征或特纳综合征的女性胎儿的X染色体剂量。因此,这个观察结果表明了使用纳米孔测序法进行胎儿染色体非整倍体的无创检测,如X单体或拷贝数畸变的潜在可行性。由于X单体代表基因组中一个染色体拷贝的减少,拷贝数变化的程度相当于三染色体性,即在基因组中增加一个染色体拷贝。因此,这些数据也反映了我们的方案可以应用于无创检测胎儿21三体综合征、18三体综合征、13三体综合征和其他胎儿染色体非整倍体。这些数据表明基于纳米孔测序的NIPT和即时治疗NIPT的可行性。

[0125] IV. 使用增加浓度的癌症测序

[0126] 循环无细胞DNA可用作实时监测癌症的“液体活检”。无细胞DNA显示潜在肿瘤中发现的遗传异常,其可以通过大规模并行测序技术检测。这些癌症患者的血浆DNA中的染色体畸变也可以使用纳米孔测序来检测。在Illumina平台上通过纳米孔测序和大规模并行测序对来自两名患有肝细胞癌(HCC)的患者的血浆DNA样品进行分析。

[0127] A. 材料和方法

[0128] 手术前从两名诊断为HCC的患者各收集外周血20ml。通过在1600g的条件下离心10分钟,然后在16000g的条件下离心10分钟分离血浆。使用QIAamp DSP DNA血液微型试剂盒(Qiagen)从8mL血浆中提取DNA。对四分之三的血浆DNA进行纳米孔测序,而剩余的DNA用NextSeq 500(Illumina)测序。

[0129] 纳米孔测序文库通过末端修复/加A尾模块(NEB)和纳米孔测序试剂盒(SQK-NSK007,牛津纳米孔技术公司)制备。将所述文库完全加载到MinION流动池(R9版本)上,并在MinION Mk1B测序仪(纳米孔)上测序。使用METRICHORTM软件(纳米孔)对输出数据文件进行碱基调用。提取2D读数并使用LAST软件与参照基因组hg19对齐。如前所述进行血浆DNA的Illumina测序(8)。

[0130] B. 结果

[0131] 计算每个样品的对齐读数对每个染色体臂(基因组代表,GR)的比例分布。换句话说,与染色体臂p或q对齐的高通滤波器读数的数量表示为占从样品测序的所有高通滤波器读数的比例。然后计算GR中相对于正常个体的血浆DNA样品的差异。如果染色体臂的GR比对照组的平均值高3个标准差,则认为所述区域表现出拷贝数增加。如果染色体组的GR比对照组的平均值低3个标准差,则认为所述区域表现出拷贝数降低。

[0132] 图9通过纳米孔测序(外环)和Illumina平台(内环)示出了HCC的两种情况(HOT530和HOT536)在GR中的差异。不同的染色体在外面被标记。分析的区域是染色体臂。染色体增加以绿色条表示,并从相应环的中心向外延伸。染色体缺失以红色条表示,并从相应环的中

心向内延伸。如图9所示,纳米孔测序的结果与Illumina平台的结果基本一致。与来自非癌症受试者的血浆DNA相比,测序的样品还显示具有更长的DNA的趋势。所述示例显示使用增加浓度法的纳米孔测序可用于分析来自癌症患者的血浆DNA。

[0133] V. 使用串联体的示例

[0134] 血浆DNA分子通常较短(<200bp),而纳米孔测序仪通常可用于测序长DNA分子。短血浆DNA分子测序效率可以通过连接单独的分子以构建长DNA分子(称为串联体)来提高。

[0135] A. 材料和方法

[0136] 1. 血浆DNA串联体的生成

[0137] 测试来自两个不同受试者的样品。从未怀孕的女性受试者和怀有男性胎儿的女性受试者各收集20ml的外周静脉血。通过在1600g的条件下离心10分钟,然后在16000g的条件下离心10分钟收集血浆。然后使用QIAamp DSP DNA血液微型试剂盒(Qiagen)从8mL血浆中提取DNA,产生420μL的血浆DNA。提取的DNA通过SpeedVac浓缩器(赛默飞世尔科技有限公司)浓缩至85μL,并通过NEBNext末端修复模块(纽英伦生物技术公司,NEB)进行末端修复。末端修复的DNA使用MinElute反应清洗试剂盒(Qiagen)纯化,并用20μL缓冲液EB洗脱。然后,通过添加20μL的Blunt/TA连接酶主混合物(NEB)并在25℃下温育4小时浓缩血浆DNA,并在温育后用MinElute反应清洗试剂盒纯化血浆DNA。

[0138] 2. 纳米孔测序

[0139] 然后将串联的DNA用于通过末端修复和加A尾模块(NEB)和基因组DNA测序试剂盒(SQK-MAP-005,牛津纳米孔技术公司)制备纳米孔测序文库。将所述文库完全加载到MinION流动池(v7.3)(纳米孔)并测序。使用METRICOR™软件(纳米孔)对输出数据文件进行碱基调用。提取2D读数。

[0140] 3. 对齐

[0141] 使用LAST软件进行对齐。为每个可能的起始位置找到初始匹配。匹配被限制为与在参考基因组中出现最多次数的最小长度匹配,或者匹配限于特定的预定长度。从这些初始匹配中,进行比这些初始匹配更长的序列的额外对齐,并且基于对测序误差的期望容忍保留具有特定空位评分的对齐。如果多个对齐共享相同的端点,则保留最高评分的对齐。以这种方式,识别了与人类基因组对齐的片段。在LAST中使用的参数和规则可能会基于准确度和精度考虑而改变。

[0142] B. 结果

[0143] 将串联的血浆DNA在MinION上测序6小时,直到文库被消耗。

[0144] 1. 非怀孕女性

[0145] 碱基调用产生了2,234个2D读数,读数长度范围为从86bp到8,672bp。图10示出了从非怀孕女性测序的串联体大小的频率分布曲线。碱基对中的串联体的大小在x轴上。以百分比表示的以给定的串联体大小存在于测序样品中的频率在y轴上绘制。为了提高图形可读性,在8,672bp异常数据点上的异常数据点未被显示。对于每个测序样品,检测到约20至50个长DNA分子。

[0146] 测序分子的大小比血浆DNA片段的典型长度长得多,一般小于200bp。这些数据表明血浆DNA片段已成功组装为串联体。然后将读数与人类基因组(hg19)对齐。然后分离属于人类基因组上不连续区域的碱基片段或区段,并将其视为一个血浆DNA片段。总共获得了3,

801个唯一映射的片段,即测序的血浆DNA片段,80.6%的唯一映射片段显示与参照基因组序列相同的序列。

[0147] 图11示出了非怀孕女性的对齐片段的大小分布图。这些对齐片段的大小从78bp至560bp不等,大部分低于200bp。峰值大小为162bp。平均大小为173bp。中值大小为162bp,这与之前在Illumina平台上通过大规模并行测序和使用DNA片段浓度增加的纳米孔测序对血浆DNA的大小的观察结果是一致的。

[0148] 图12示出了与男性的参考基因组相比的对齐区段的计算分布图。计算对齐区段对每个染色体(基因组代表)的比例分布。计算的分布与所有常染色体中hg19的分布相似。对于性染色体,X染色体的基因组代表为5.79%,这对于女性样品来说是预料之中的。没有发现未与Y染色体对齐的情况。因此,使用具有纳米孔的串联体的方法可以将女性血浆与男性血浆区分开来。

[0149] 2. 怀有男性胎儿的女性

[0150] 怀孕女性携带的男性胎儿在采血时具有38周零4天的胎龄。怀孕被认为是正常妊娠。在处理纳米孔测序之前,将血浆DNA片段串联。串联体的大小范围为100bp至13,466bp,其中中值大小为676bp,平均大小为965bp。

[0151] 图13示出了怀有男性胎儿的女性的串联体的对齐区段的大小分布图。大小分布图是血浆DNA的典型特征。片段的中值大小是196bp。峰值大小为174bp,大小在92bp至2,934bp之间。只有大约0.4%的片段的大小大于2,000bp。这种分布类似于通过使用串联体、增加浓度或大规模并行测序的方法获得的其他大小分布。例如,所述分布类似于使用将怀有男性胎儿的母体血浆的浓度增加的方法发现的大小分布,如图6所示。

[0152] 图14示出了与男性和非怀孕女性的参考基因组相比,怀有男性胎儿的女性的对齐区段的计算染色体分布。串联体的分布显示X染色体水平介于男性和非怀孕女性水平之间。这表明X染色体数据反映了正常男性胎儿X染色体的单体性。此外,串联体显示胎儿来源的Y染色体的证据。通过从参考基因组测序获得的读数分布的一些偏差可能是限制分析仅参考基因组的可对齐部分的结果。然而,如这些结果所示,使用串联体的方法可以将怀有男性胎儿的母体血浆与男性或非怀孕女性的血浆区分开来。这些方法可能可以用来区分怀有男性胎儿的母体血浆与怀有女性胎儿的母体血浆。

[0153] VI. 其他实施例

[0154] 实施例1包括一种方法,其包括:接收多个DNA片段;将第一组DNA片段串联以获得第一串联体;使第一串联体通过第一纳米孔;以及在所述第一串联体穿过所述第一纳米孔时检测第一电信号,所述电信号对应于所述第一串联体的第一序列。

[0155] 实施例2包括实施例1的方法,还包括通过计算机系统执行:分析第一电信号以确定第一序列;以及将第一序列的子序列与参考基因组对齐以识别对应于第一组DNA片段中的每一个的片段序列。

[0156] 实施例3包括实施例1的方法,其中使第一串联体穿过第一纳米孔包括:使第一串联体的第一链穿过纳米孔;随后使第一串联体的第二链穿过纳米孔。实施例4包括实施例3的方法,还包括通过计算机系统分析第一链和第二链的第一电信号以确定第一序列。

[0157] 实施例5包括实施例1的方法,其中所述多个DNA片段是来自生物样品的无细胞DNA片段。

[0158] 实施例6包括实施例2的方法,其中生物样品是血浆或血清。

[0159] 实施例7包括实施例1的方法,其中第一纳米孔是基底上的多个纳米孔中的一个。实施例8包括实施例7的方法,进一步包括:将第二组DNA片段串联以获得第二串联体;使第二串联体穿过多个纳米孔中的第二纳米孔;以及当所述第二串联体穿过所述第二纳米孔时检测电信号,所述电信号对应于所述第二串联体的第二序列。实施例9包括实施例7的方法,进一步包括:将第二组DNA片段串联以获得第二串联体;使第二串联体穿过第一纳米孔;以及当所述第二串联体穿过所述第一纳米孔时检测电信号,所述电信号对应于所述第二串联体的第二序列。

[0160] 实施例10包括一种方法,包括:接收包括多个DNA片段的生物样品;浓缩生物样品以获得更高浓度的DNA片段;使多个DNA片段穿过基底上的纳米孔;以及对于所述多个DNA片段中的每一个:当所述DNA片段穿过纳米孔时检测电信号,所述电信号对应于所述DNA片段的序列。

[0161] 实施例11包括一种方法,包括通过计算机系统执行:接收通过串联第一组DNA片段而产生的第一串联体的第一序列;以及将第一序列的子序列与参考基因组对齐以识别对应于第一组DNA序列中的每一个的片段序列。实施例12包括实施例11的方法,其中子序列的对齐包括:将第一序列的窗口与参考基因组对齐;以及识别两个窗口何时与参考基因组的不同区域对齐。实施例13包括实施例12的方法,其中两个窗口之间的一个或多个窗口被识别为未与参考基因组对齐。

[0162] 实施例14包括一种包括计算机可读介质的计算机产品,所述计算机可读介质存储用于控制计算机系统执行实施例11-13中的任何一个的操作的多个指令。实施例15包括一种系统,包括:实施例14的计算机产品;以及用于执行存储在计算机可读介质上的指令的一个或多个处理器。实施例16包括一种系统,其包括用于执行实施例11-13中的任何一个的装置。实施例17包括被配置为执行实施例11-13中的任何一个的系统。实施例18包括一种包括分别执行实施例11-13中的任何一个的步骤的模块的系统。

[0163] VII. 示例性测序系统

[0164] 图15示出了根据本发明实施例的执行单分子测序的系统1500的方块图。生物样品可以通过提取装置1504从患者1502获得。生物样品可以是本文所述的任何体液或生物样品。提取装置1504可以包括注射器、刺血针、拭子或用于收集诸如尿液的样品的容器或小瓶。提取装置1504可以包括德国凯杰公司(Qiagen, Germany)的QIAamp DSP DNA血液微型试剂盒(2)。生物样品可以包括无细胞DNA片段,其被输送到制备装置1506。制备装置1506可以包括产生一种对于单分子测序更有效的无细胞DNA片段的装置。例如,制备装置1506的输出可以是无细胞DNA片段的串联体或浓缩样品。

[0165] 当制备串联体时,制备装置1506可以包括用于增加浓度的真空干燥器,例如SpeedVac浓缩器(赛默飞世尔科技有限公司)或通过渗滤或过滤除去流体的超浓缩器。对于串联体,浓度可能不会增加五倍或更多。制备装置1506可以包括修复DNA片段末端的模块,例如NEBNext末端修复模块(NEB)。另外,制备装置1506可以包括纯化末端修复的DNA的试剂盒,例如MinElute反应清洗试剂盒(凯杰公司)。制备装置1506还可以包括孵育器,其可以在一定温度下和一定时间内(例如25℃, 4小时)孵育连接酶(例如,NEB的Blunt/TA连接酶主混合物)。制备装置1506可以包括用于在反应后纯化串联体的另外的试剂盒,其可以包括

MinElute反应清洗试剂盒(凯杰公司)。制备装置1506还可以包括可以混合和传输流体的机器人液体处理器。

[0166] 当产生增加浓度的无细胞DNA片段时,制备装置1506可以包括真空干燥器(例如,赛默飞世尔科技有限公司的Speedvac浓缩器)或通过渗滤或过滤去除流体的超浓缩器。制备装置1506还可以包括末端修复和加A尾模块(纽英伦生物技术公司)。制备装置1506还可以包括可以混合和传输流体的机器人液体处理器。

[0167] 可将来自制备装置1506的串联体或浓缩的DNA片段输送至单分子测序装置1508以获得序列读数。单分子测序装置1508还可以包括诸如MinION流动池(v7.3)(纳米孔)、SMRT技术(太平洋生物科学公司)或Helicos单分子测序仪(SeqLL)的装置。单分子测序装置1508还可以包括与纳米孔相关的试剂盒,例如基因组DNA测序试剂盒(SQK-MAP-005,牛津纳米孔技术公司)或纳米孔测序试剂盒(SQK-NSK007,牛津纳米孔技术公司)。

[0168] 单分子测序装置1508可输出来自串联体或浓缩DNA片段的序列读数。序列读数可以作为数据文件输出,并且可以由计算机系统1510分析。计算机系统可以是具有用来分析序列读数的软件的专用计算机系统。输出数据文件可以使用METRICOR™软件(纳米孔)进行碱基调用。2D读数可以使用LAST软件来提取和对齐。所述计算机系统可以是图16中的计算机系统10,如下文所述。

[0169] VIII. 计算机系统

[0170] 本文提到的任何计算机系统可以使用任何合适数量的子系统。这样的子系统的示例在图16中的计算机系统10示出。在一些实施例中,计算机系统包括单个计算机装置,其中子系统可以是计算机装置的组件。在其他实施例中,计算机系统可以包括具有内部组件的多个计算机装置,每个计算机装置都是子系统。计算机系统可以包括台式和膝上型计算机、平板电脑、移动电话和其他移动设备。

[0171] 子系统在图6中通过系统总线75互连。还示出了诸如打印机74、键盘78、存储设备79(一个或多个)、耦合到显示适配器82的监视器76等的附加子系统。耦合到I/O控制器71的外围设备和输入/输出(I/O)设备可以通过本领域中已知的任何数量的装置,例如输入/输出(I/O)端口77(例如USB、FireWire®),连接到计算机系统。例如,可使用I/O端口77或外部接口81(例如,以太网、Wi-Fi等)将计算机系统10连接到诸如互联网、鼠标输入设备或扫描仪之类的广域网。经由系统总线75的互连允许中央处理器73与每个子系统通信并且控制来自系统存储器72或存储设备79(一个或多个)(例如,固定硬盘,诸如硬盘驱动器或光盘)的多个指令的执行,以及子系统之间的信息交换。系统存储器72和/或存储设备79(一个或多个)可以体现为计算机可读介质。另一个子系统是数据收集设备85,诸如照相机、麦克风、加速度计等。这里提到的任何数据都可以从一个组件输出到另一个组件,并且可以输出给用户。

[0172] 计算机系统可以包括多个相同的组件或子系统,例如通过外部接口81、内部接口或者可移动的存储设备连接在一起,可移动的存储设备可以从一个组件连接到另一个组件或从其中移除。在一些实施例中,计算机系统、子系统或装置可以通过网络进行通信。在这种情况下,一台计算机可以被认为是一个客户端,另一台计算机是一台服务器,其中每台计算机可以是同一个计算机系统的一部分。客户端和服务器可以分别包含多个系统、子系统或组件。

[0173] 实施例的各方面可以以使用硬件(例如专用集成电路或现场可编程门阵列)的控制逻辑的形式和/或以模块化或集成的方式使用具有通常可编程处理器的计算机软件来实现。如本文所使用的,处理器包括单核处理器、同一集成芯片上的多核处理器或单个电路板上的多个处理单元或联网的多个处理单元。基于本文提供的公开和教导,本领域的普通技术人员将会了解并且意识到使用硬件以及硬件和软件的组合来实现本发明的实施例的其他方式和/或方法。

[0174] 本申请中描述的任何软件组件或功能可以实现为由处理器通过,例如常规或面向对象的技术,使用任何合适的计算机语言(例如Java、C、C++、C#、Objective-C、Swift或诸如Perl和Python的脚本语言)执行的软件代码。软件代码可以被存储为用于存储和/或传输的计算机可读介质上的一系列指令或命令。合适的非暂态计算机可读介质可以包括随机存取存储器(RAM)、只读存储器(ROM)、诸如硬盘驱动器或软盘的磁介质、诸如光盘(CD)或DVD(数字多功能盘)的光学介质、闪存等。计算机可读介质可以是这种存储或传输设备的任何组合。

[0175] 也可以使用适合于经由符合各种协议(包括因特网)的有线、光学和/或无线网络进行传输的载波信号来编码和传输这样的程序。这样,可以使用用这种程序编码的数据信号来创建计算机可读介质。用程序代码编码的计算机可读介质可以与兼容设备封装在一起,或者与其他设备分开提供(例如,经由因特网下载)。任何这样的计算机可读介质可以驻留在单个计算机产品(例如,硬盘驱动器、CD或整个计算机系统)上或内部,并且可以存在于系统或网络内的不同计算机产品之上或之内。计算机系统可以包括监视器、打印机或其他合适的显示器,用于向用户提供本文提到的任何结果。

[0176] 这里描述的任何方法可以完全地或部分地用包括一个或多个处理器的计算机系统执行,其可以被配置为执行这些步骤。因此,实施例可以针对被配置为潜在地利用执行相应步骤或相应步骤组的不同部件的执行本文描述的任何方法的步骤的计算机系统。虽然以编号的步骤呈现,但是本文的方法的步骤可以同时或以不同的顺序执行。另外,这些步骤的部分可以与其他方法的其他步骤的部分一起使用。而且,步骤的全部或部分可以是可选的。另外,可以用模块、单元、电路或用于执行这些步骤的其他装置来执行任何方法的任何步骤。

[0177] 在不脱离本发明的实施例的精神和范围的情况下,可以以任何合适的方式来组合特定实施例的具体细节。然而,本发明的其他实施例可以针对与每个单独方面有关的具体实施例,或者这些单独方面的具体组合。

[0178] 出于说明和描述的目的给出了本发明的示例性实施例的以上描述。这并不意味着穷举或将本发明限制于所描述的精确形式,并且鉴于上述教导,许多修改和变化是可能的。

[0179] 除非特别指出,否则对“一”、“一个”或“所述”的引用旨在表示“一个或多个”。除非特别指出,否则“或”的使用旨在表示“包含性的”,而不是“排除性的”。提到“第一”部件不一定要求提供第二部件。此外,除非明确说明,对“第一”或“第二”部件的引用不将所引用的部件限制到特定位置。

[0180] 本文提及的所有专利、专利申请、出版物和描述通过引用整体并入本文以用于各种目的。没有被承认是现有技术。

[0181] IX. 参考文献

- [0182] 1.Dondorp W、de Wert G、Bombard Y、Bianchi DW、Bergmann C、Borry P等人“非整倍体等的无创产前检测:产前筛查的负责任创新挑战”《欧洲人类遗传学杂志 (Eur J Hum Genet)》2015年;
- [0183] 2.Chiu RWK、Chan KCA、Gao Y、Lau VYM、Zheng W、Leung TY等人“通过母体血浆DNA大规模并行基因组测序进行的无创产前胎儿染色体非整倍体诊断”《美国科学院院报 (Proc Natl Acad Sci)》2008年;105:20458-63;
- [0184] 3.Jain M、Fiddes IT、Miga KH、Olsen HE、Paten B、Akeson M“MinION纳米孔测序仪的改进数据分析”《自然方法 (Nat Methods)》2015年;12:351-6;
- [0185] 4.Lo YMD、Chan KCA、Sun H、Chen EZ、Jiang P、Lun FMF等人“母体血浆DNA测序揭示胎儿的全基因组遗传和突变特征”《科学转化医学 (Sci Transl Med)》2010年;2:61ra91-61ra91;
- [0186] 5.Bayley H“纳米孔测序:从想象到现实”《临床化学 (Clin Chem)》2015年;61:25-31;
- [0187] 6.Zheng YW、Chan KCA、Sun H、Jiang P、Su X、Chen EZ等人“血浆中非造血来源的DNA短于造血来源的DNA:移植模型”《临床化学 (Clin Chem)》2012年;58:549-58;
- [0188] 7.Lun FMF、Chiu RWK、Sun K、Leung TY、Jiang P、Chan KCA等人“通过母体血浆DNA全基因组亚硫酸氢盐测序进行的无创产前甲基化分析”《临床化学 (Clin Chem)》2013年;59:1583-94;
- [0189] 8.Chan KCA、Jiang P、Zheng YW、Liao GJW、Sun H、Wong J等人“血浆中癌症基因组扫描:通过大规模并行测序检测肿瘤相关拷贝数畸变、单核苷酸变体和肿瘤异质性”《临床化学 (Clin Chem)》2013年;59:211-24;
- [0190] 9.Chan KCA、Jiang P、Chan CWM、Sun K、Wong J、Hui EP等人“通过血浆DNA亚硫酸氢盐测序无创检测癌症相关的全基因组低甲基化和拷贝数异常”《美国科学院院报 (Proc Natl Acad Sci USA)》2013年;110:18761-8;
- [0191] 10.Chan RWY、Wong J、Chan HLY、Mok TSK、Lo WYW、Lee V等人“肝脏病理学中肝脏来源血浆白蛋白MRNA的异常浓度”《临床化学 (Clin Chem)》2010年;56:82-9;
- [0192] 11.Chan RWY、Jiang P、Peng X、Tam LS、Liao GJW、Li EK等人“通过基因组和甲基化测序揭示的系统性红斑狼疮血浆DNA畸变”《美国科学院院报 (Proc Natl Acad Sci USA)》2014年;111:E5302-11;
- [0193] 12.Schreiber J、Wescoe ZL、Abu-Shumays R、Vivian JT、Baatar B、Karplus K、Akeson M“单个DNA链中胞嘧啶、甲基胞嘧啶和羟甲基胞嘧啶的纳米孔识别误差率”《美国科学院院报 (Proc Natl Acad Sci USA)》2013年;110:18910-5。

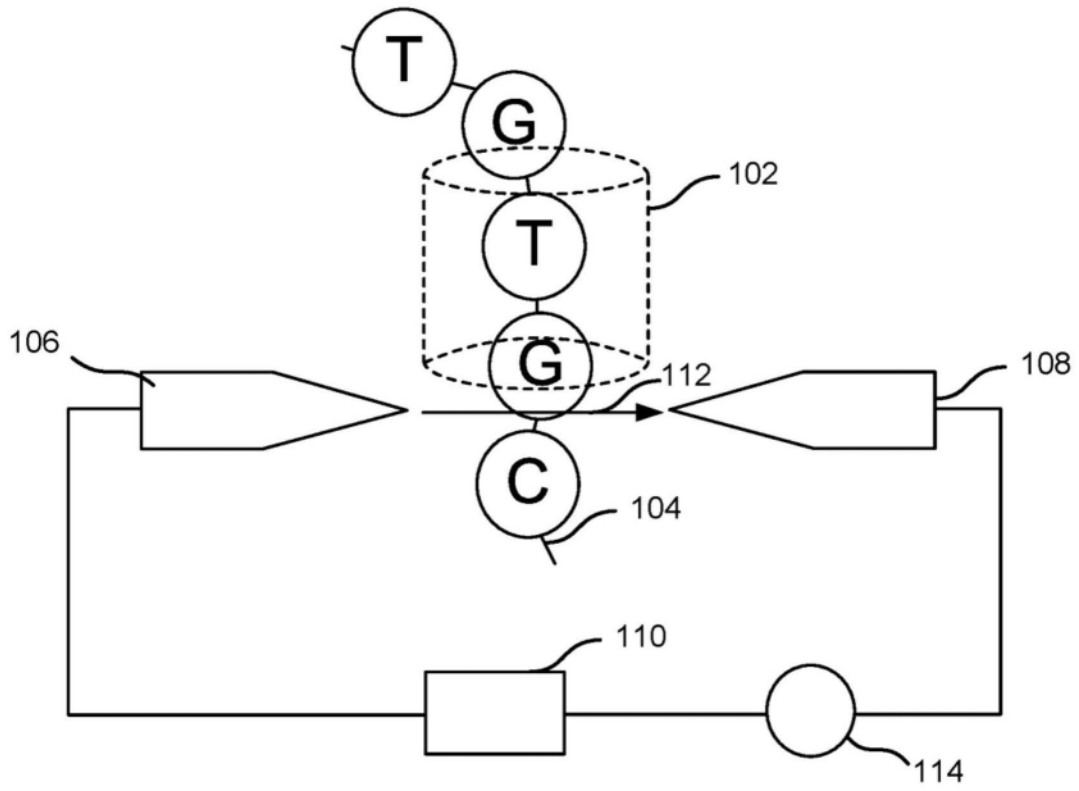


图1A

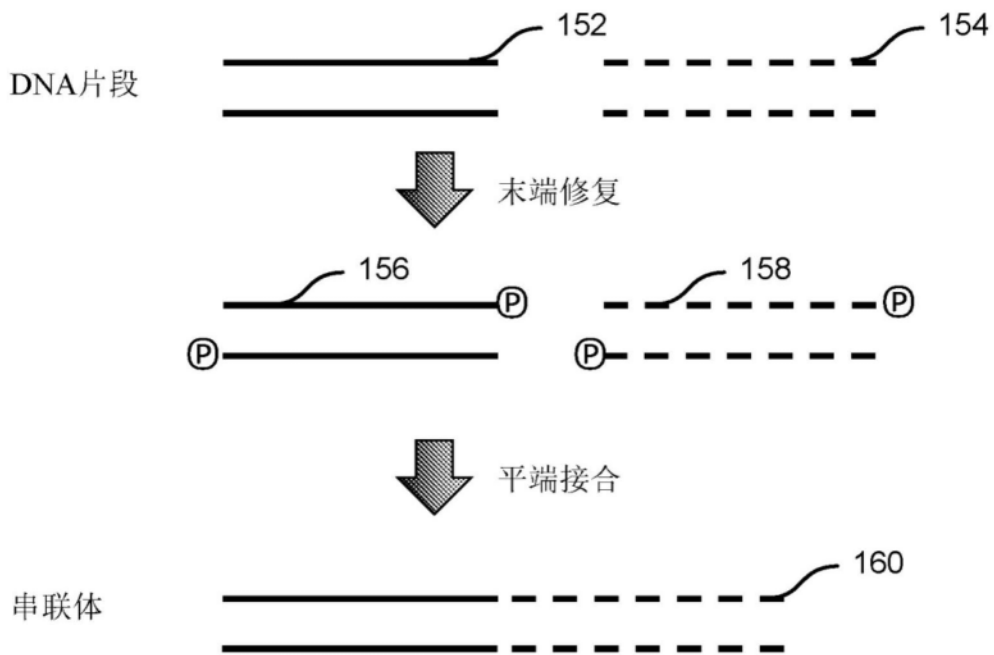


图1B

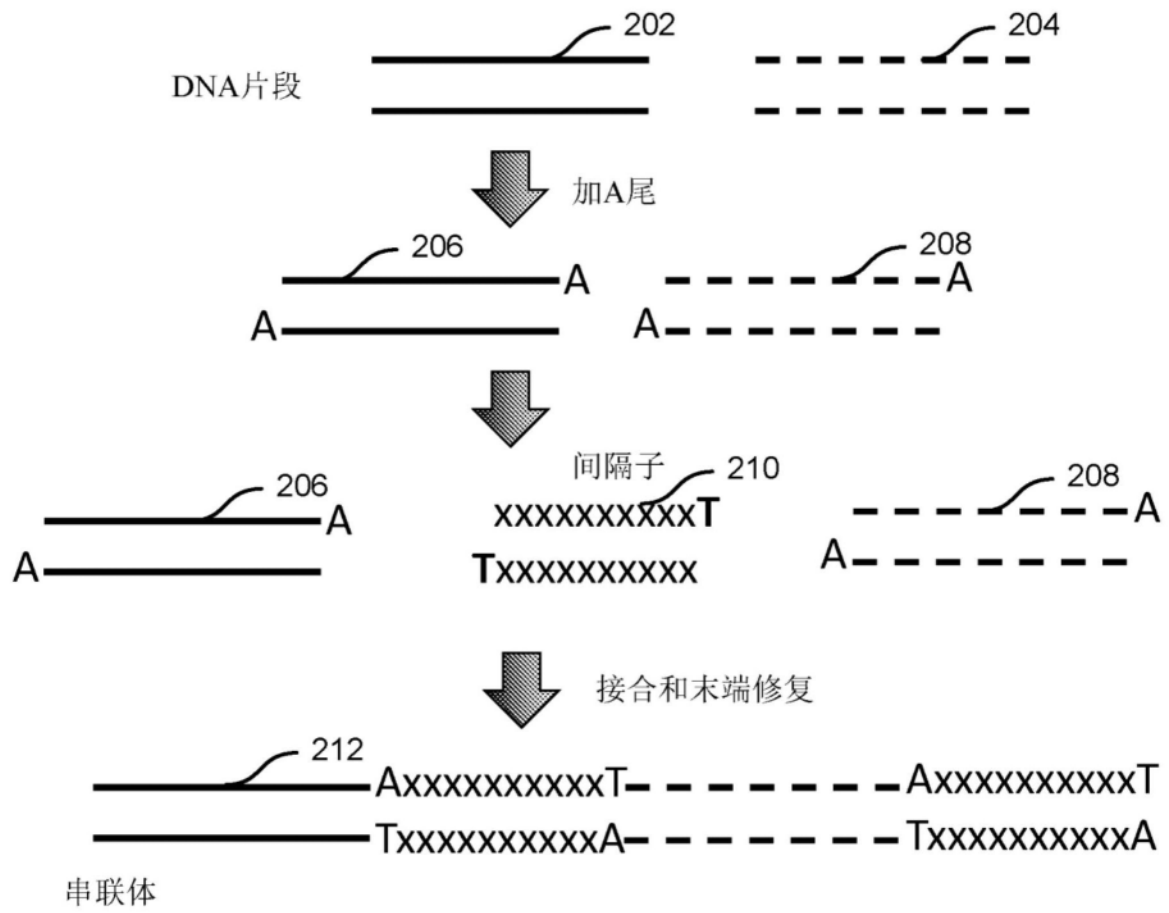


图2

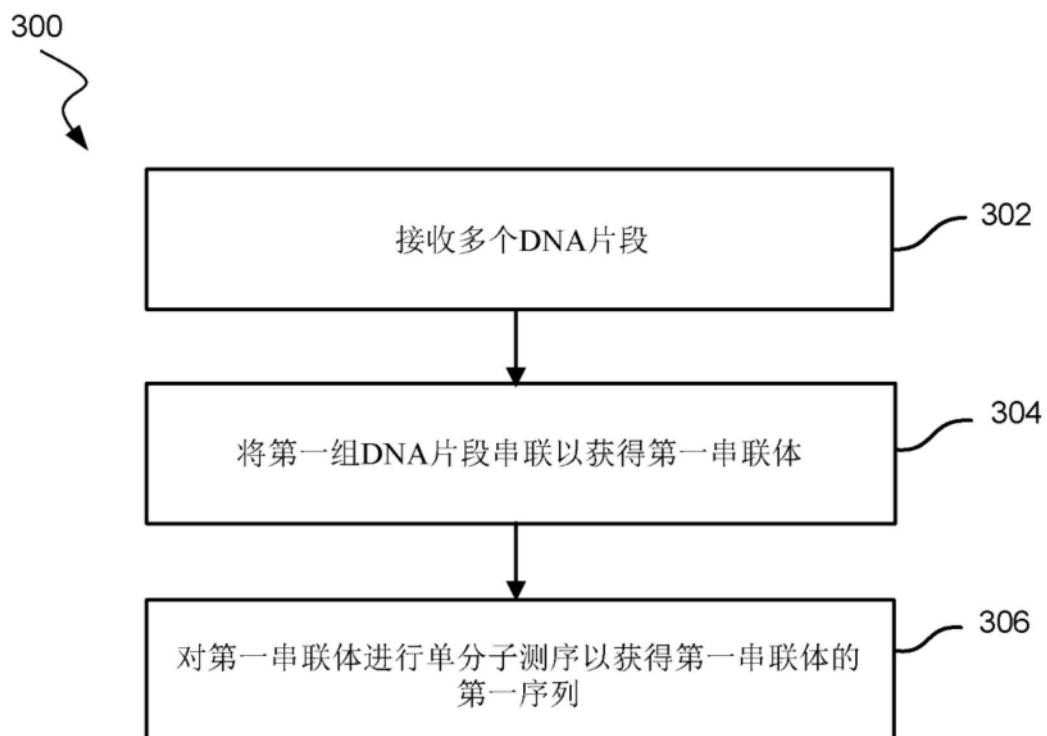


图3A

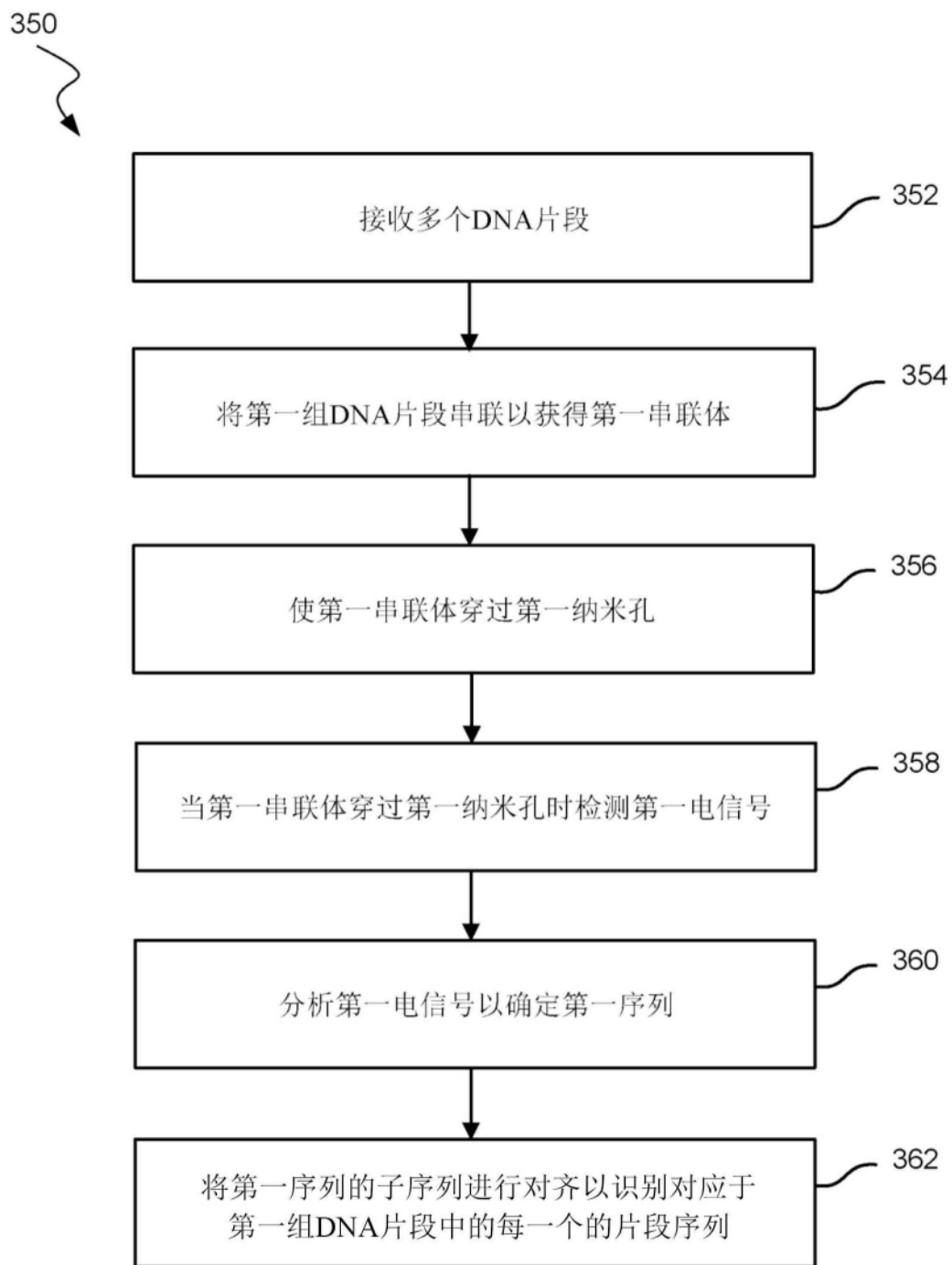


图3B

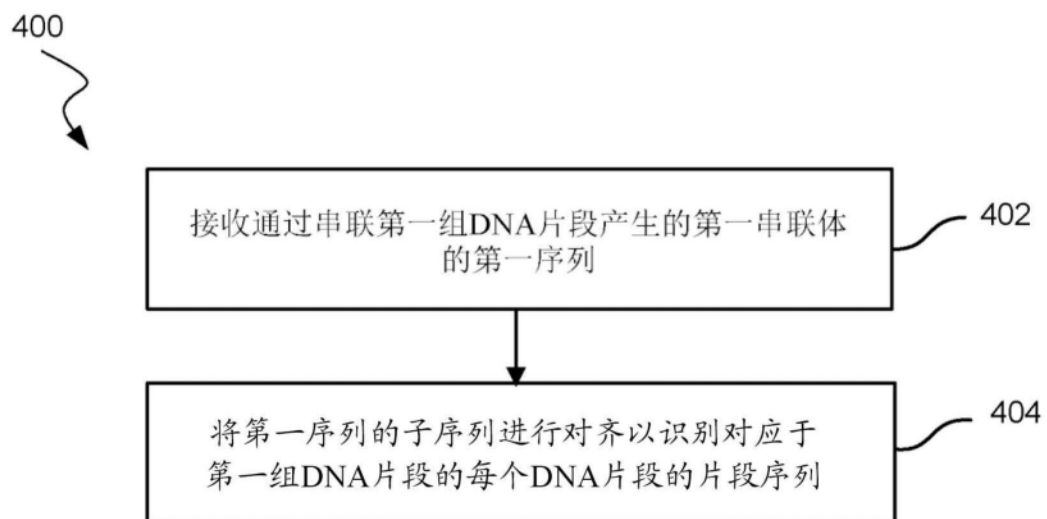


图4

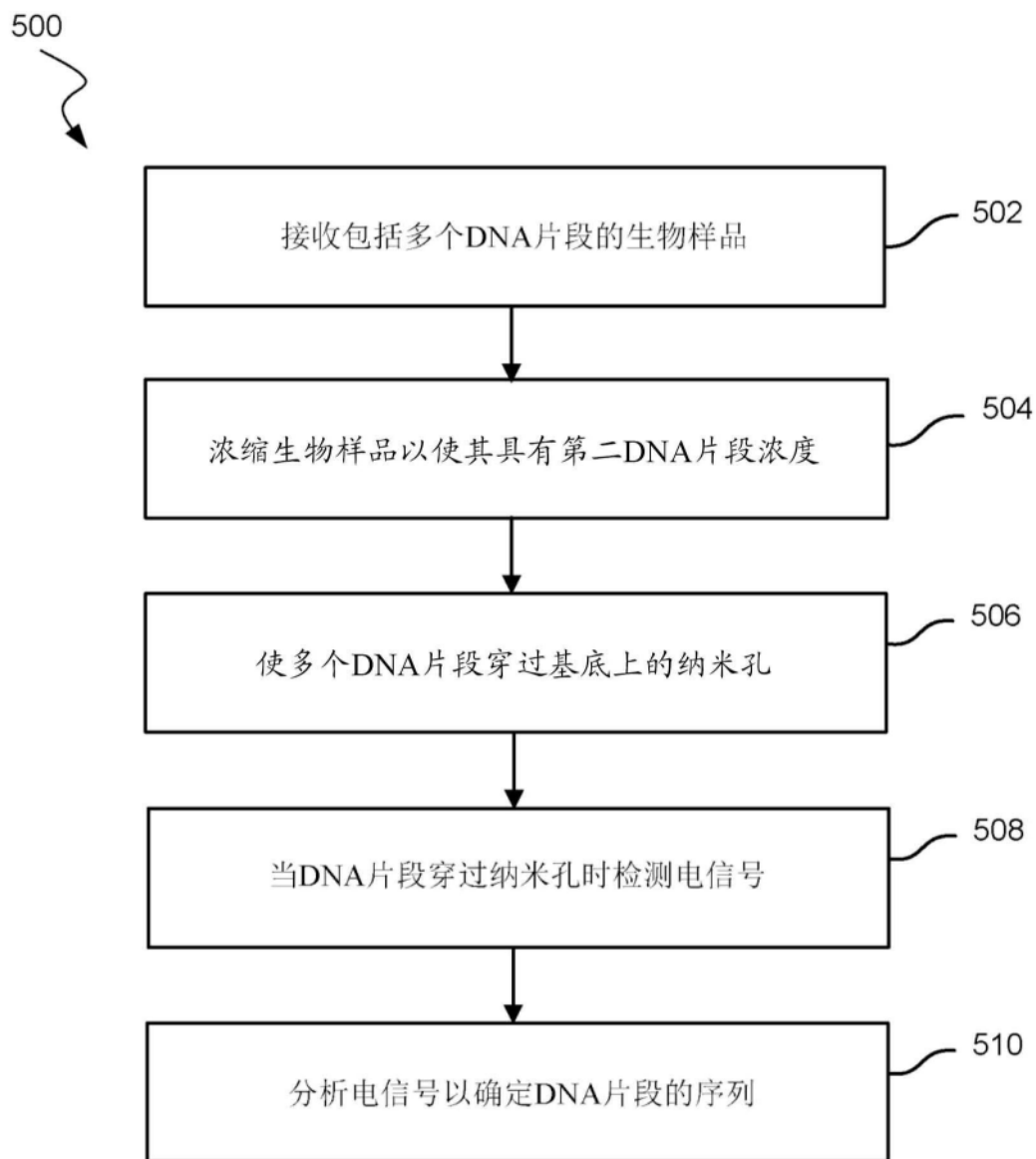


图5

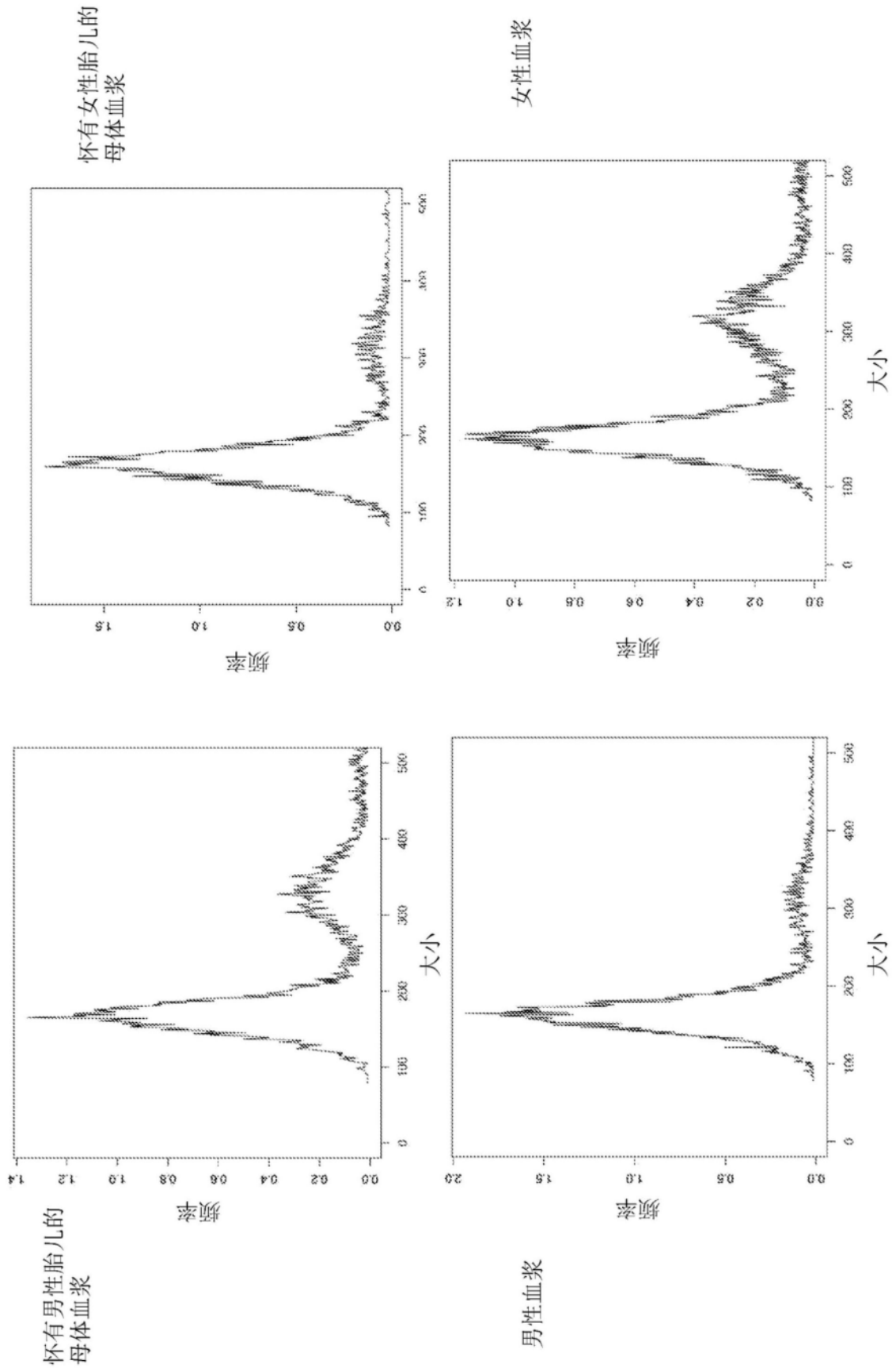


图6

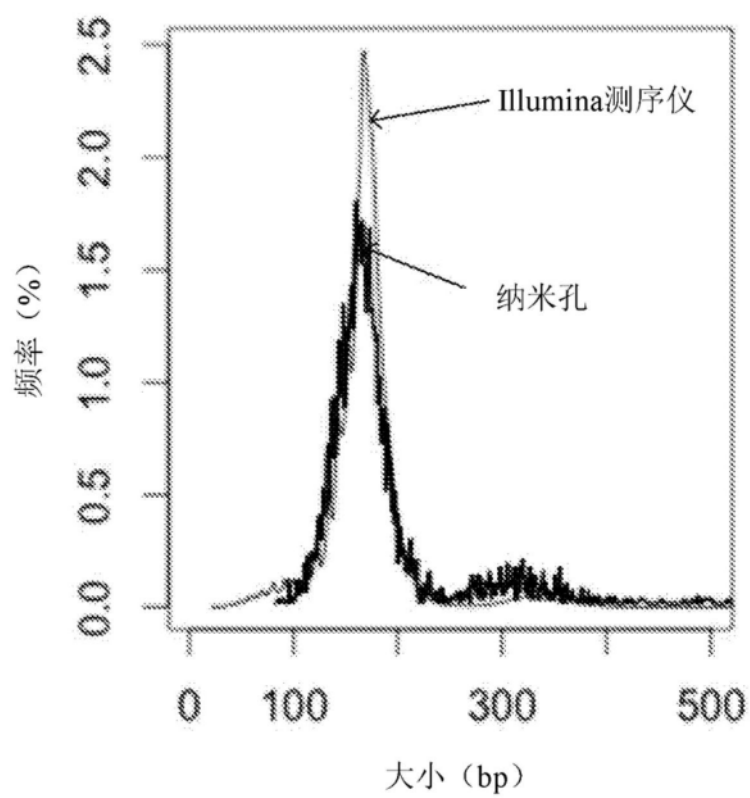


图7

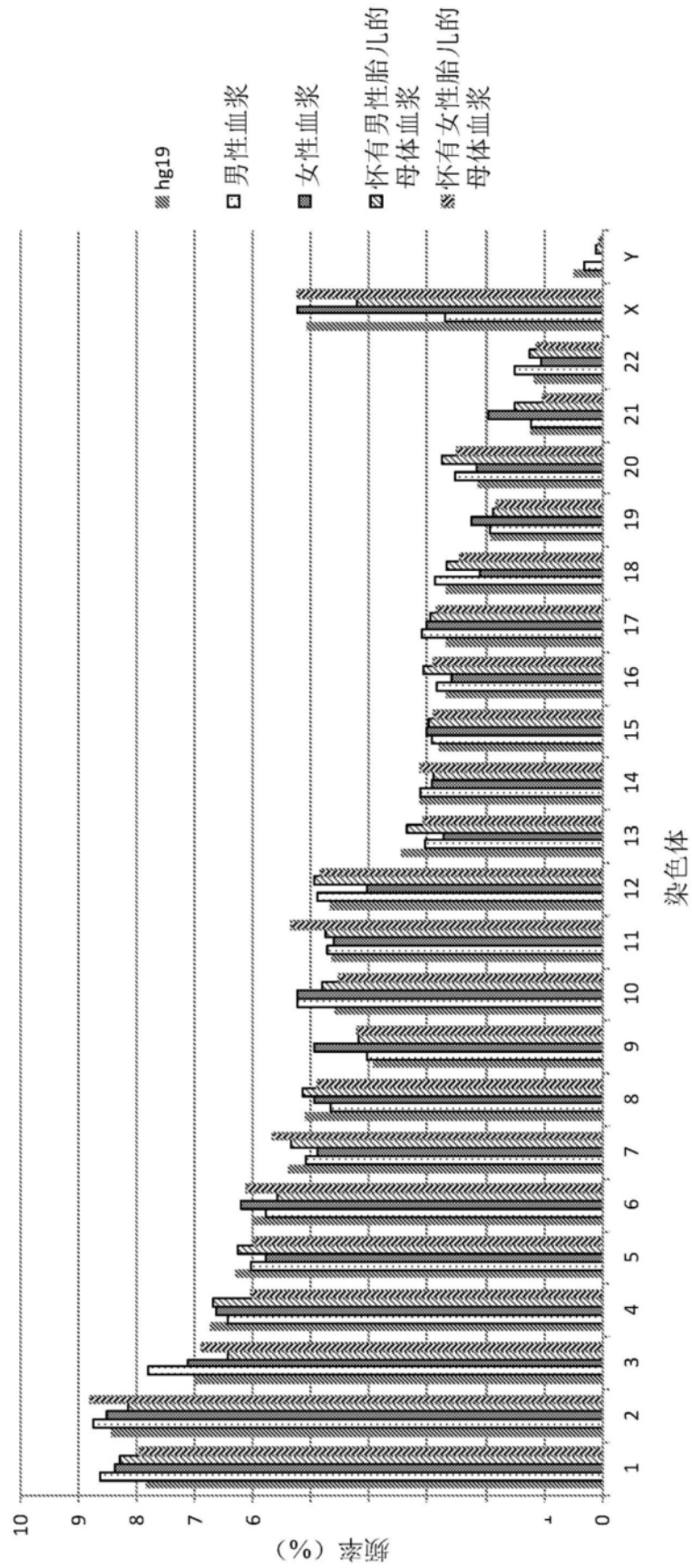


图8

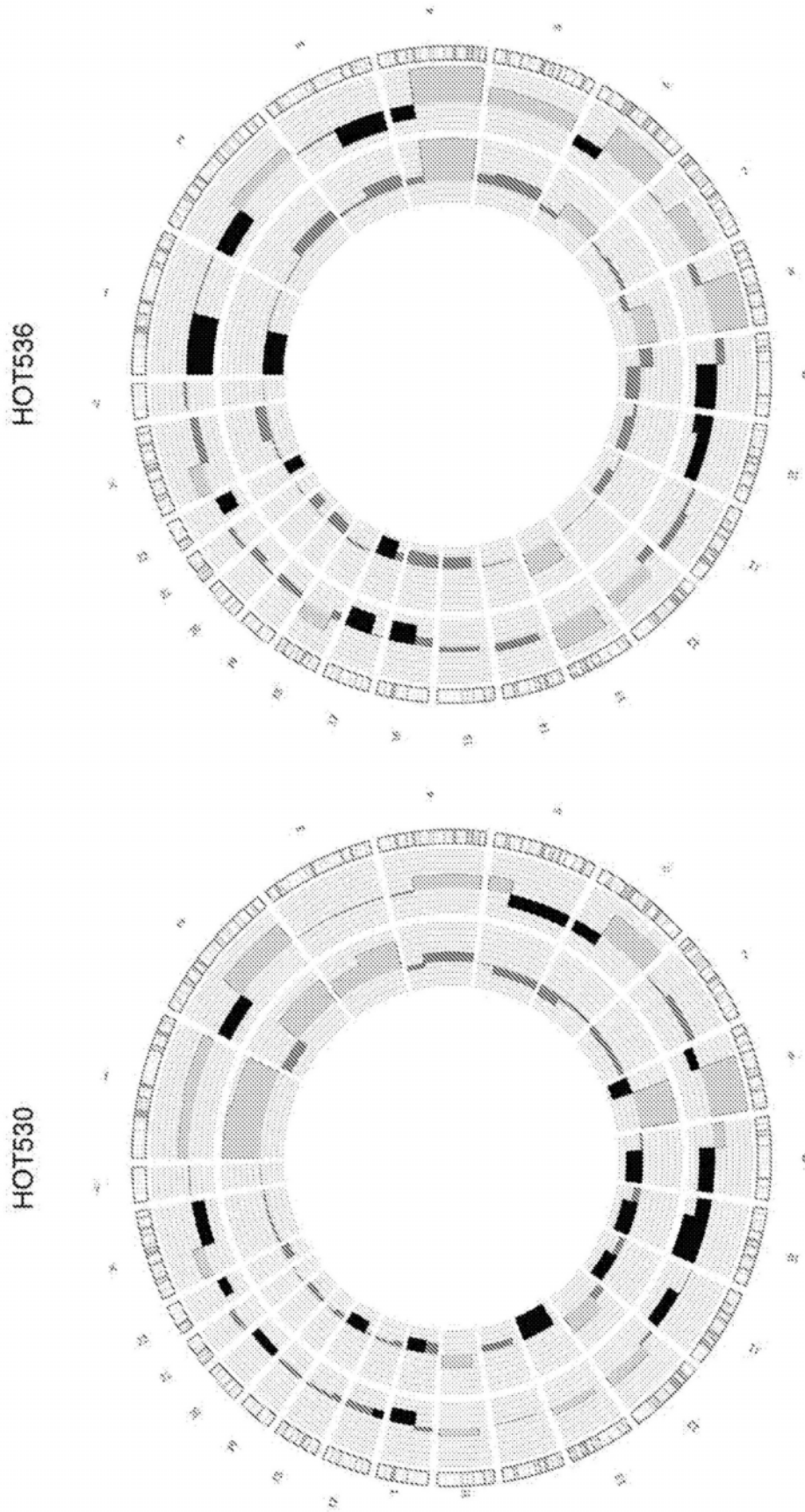


图9

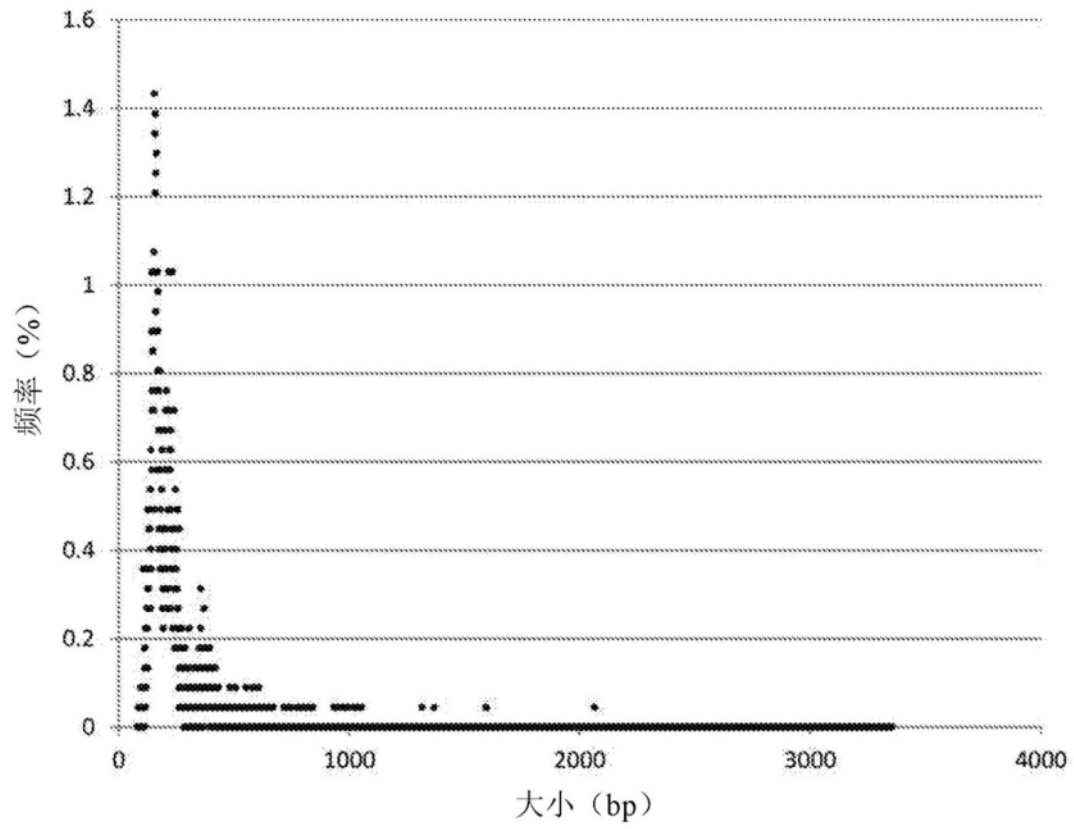


图10

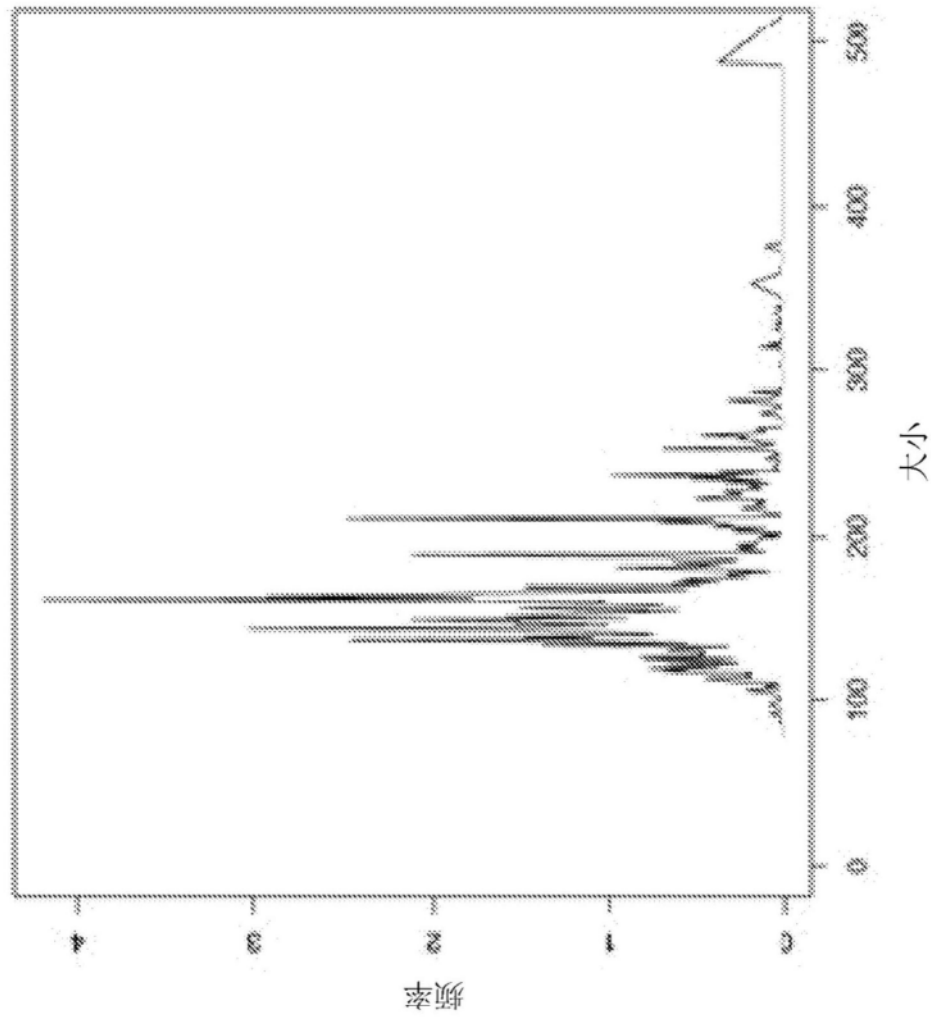


图11

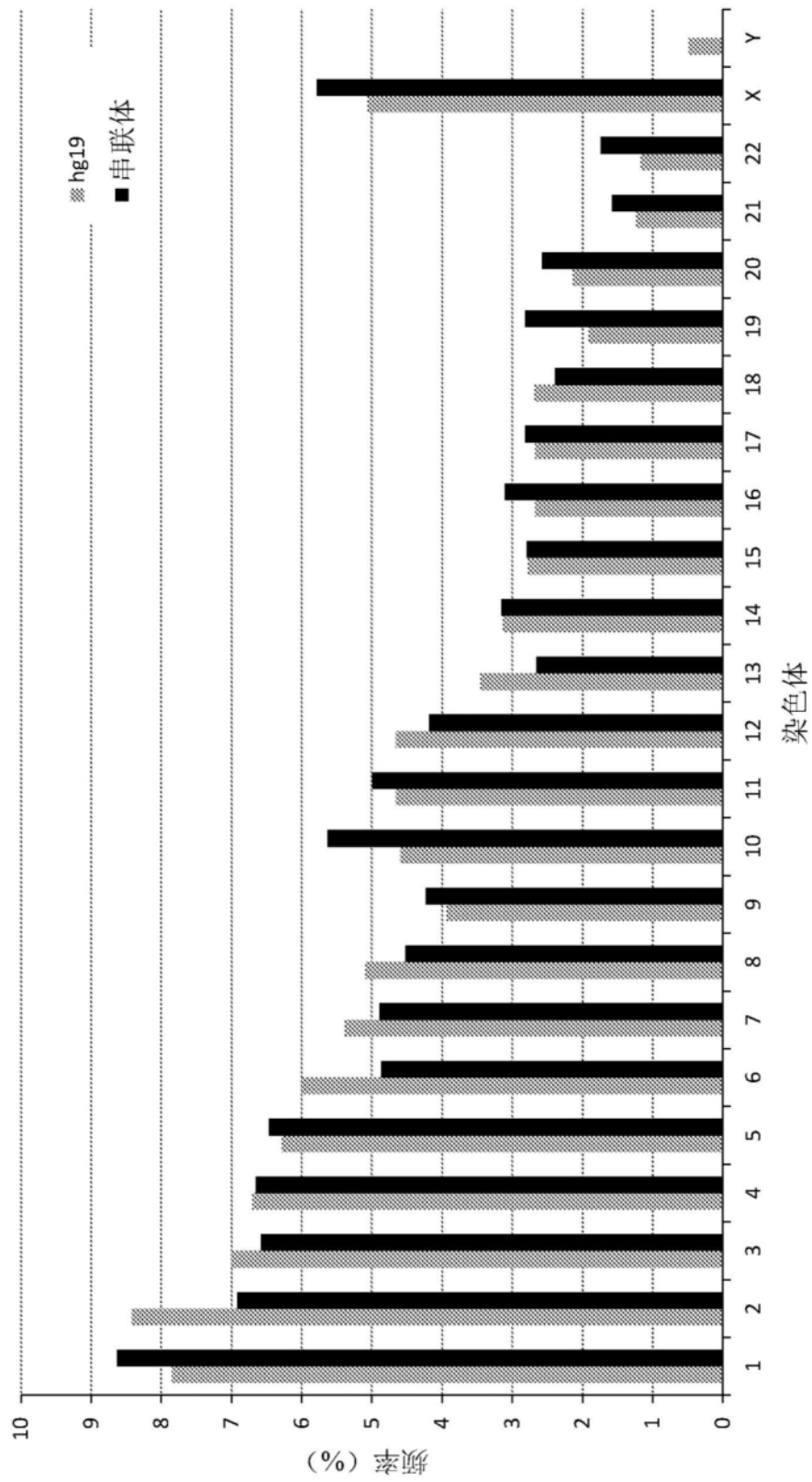


图12

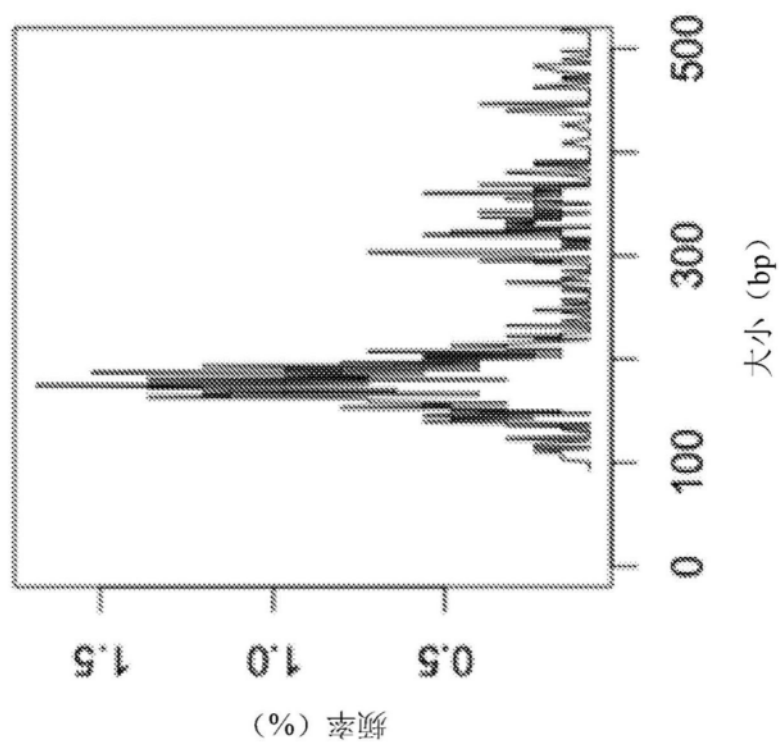


图13

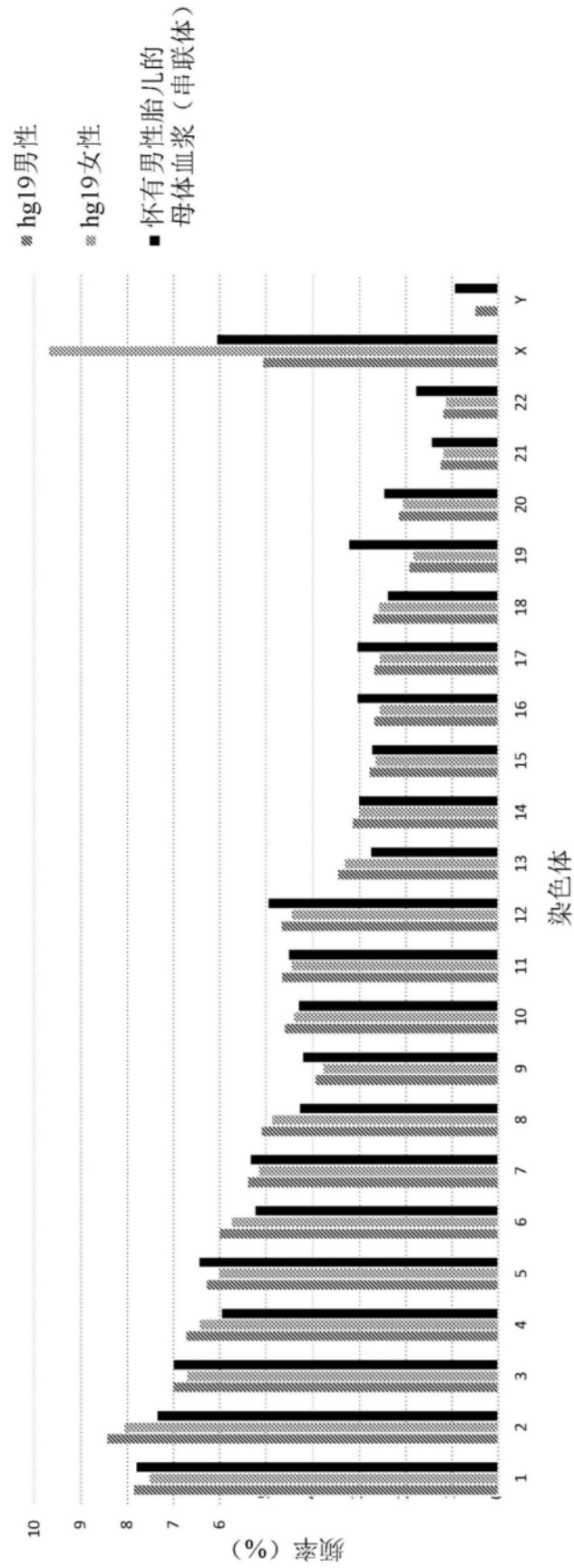


图14

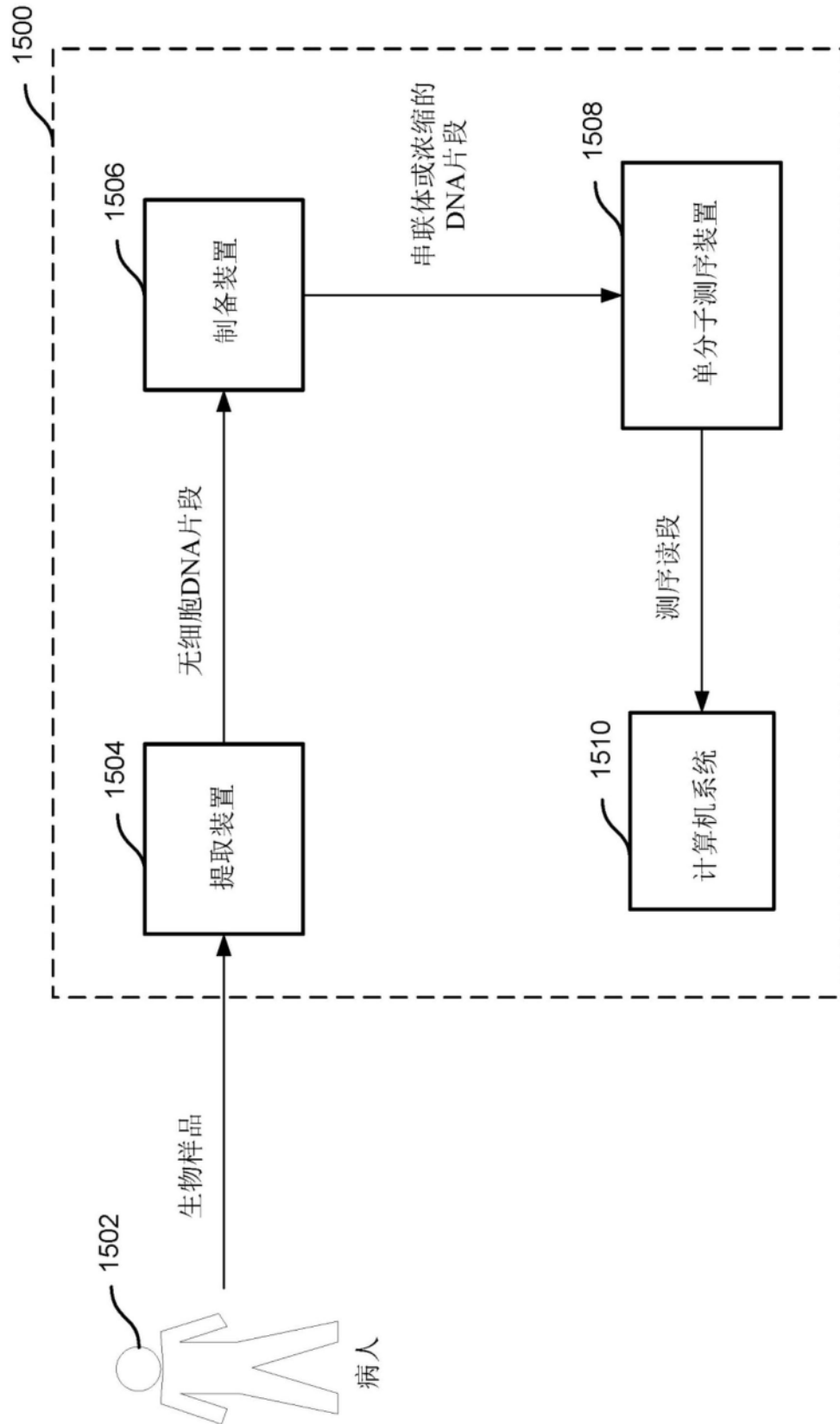


图15

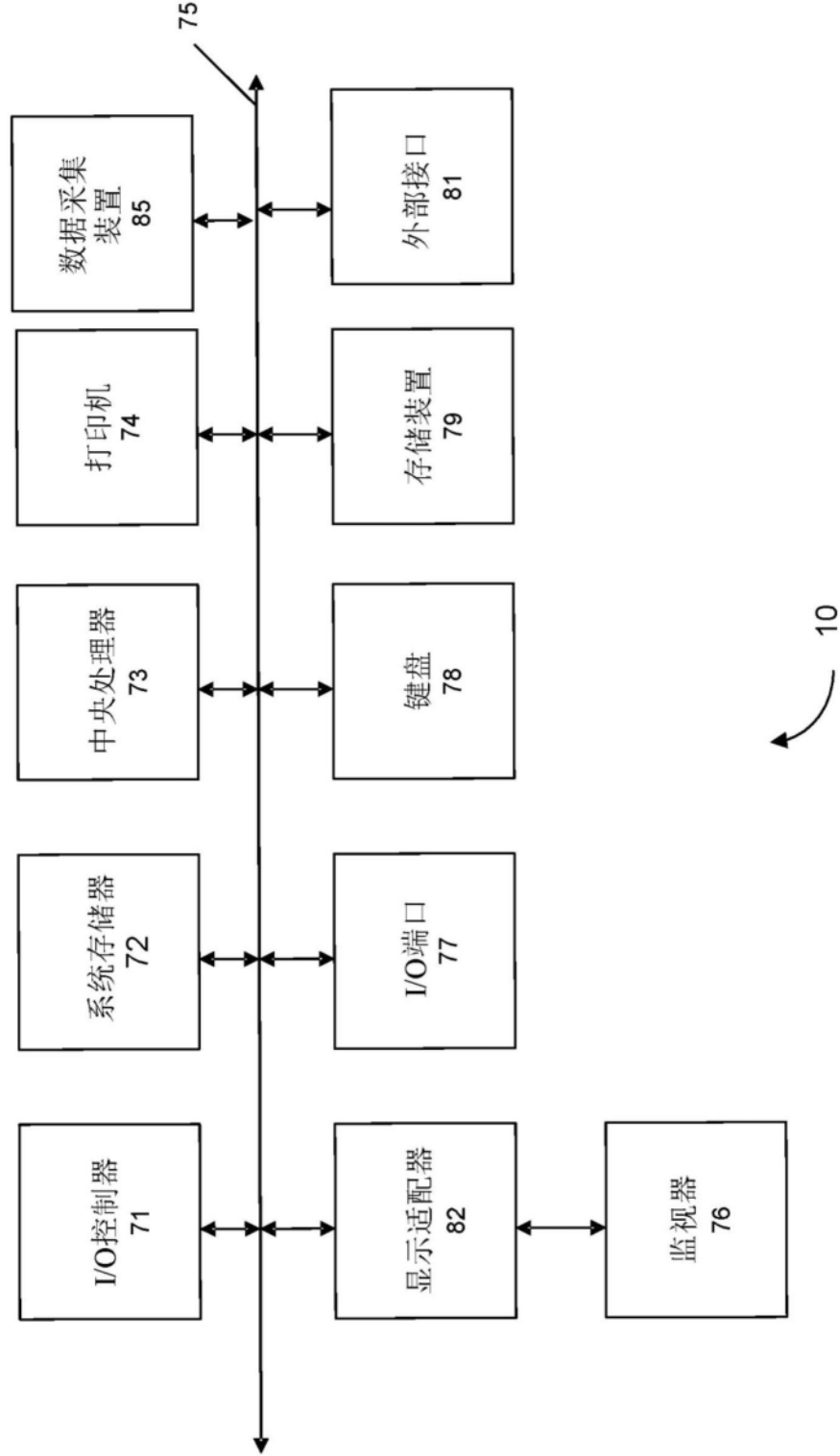


图16