



(12) 发明专利申请

(10) 申请公布号 CN 113166757 A

(43) 申请公布日 2021. 07. 23

(21) 申请号 202080006350.2

雅各布·梅耶斯

(22) 申请日 2020.02.14

(74) 专利代理机构 北京弘权知识产权代理有限公司 11363

(30) 优先权数据

代理人 郭放 王翠华

62/806,698 2019.02.15 US

(85) PCT国际申请进入国家阶段日

(51) Int.Cl.

2021.05.21

C12N 15/10 (2006.01)

(86) PCT国际申请的申请数据

C12Q 1/686 (2006.01)

PCT/US2020/018360 2020.02.14

G01N 25/04 (2006.01)

(87) PCT国际申请的公布数据

W02020/168239 EN 2020.08.20

(71) 申请人 宝生物工程(美国)有限公司

地址 美国加利福尼亚

(72) 发明人 伊曼纽尔·坎伯罗夫 木村刚隆

朱莉·凯瑟琳·拉贝特

帕特里克·凯文·马丁

权利要求书2页 说明书25页

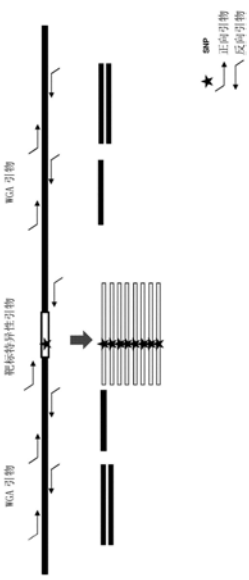
序列表6页 附图17页

(54) 发明名称

核酸文库的制备和分析方法

(57) 摘要

在相同样品中检测不同突变是必不可少的，尤其是在样品数量有限且需要高通量方法以快速检测突变的情况下。本领域常规使用的方法需要单独的测定方法来检测样品中的不同突变或突变类型(例如单核苷酸多态性(SNP)或拷贝数变异(CNV))。本公开提供了用于在相同样品中检测不同突变例如SNP和CNV的方法。本文所述的方法可以用于植入前基因测试、携带者筛查或基因分型。



1. 一种检测样品中单核苷酸多态性 (SNP) 和拷贝数变异 (CNV) 的方法, 所述方法包括:
  - a) 获得包含核酸分子的样品;
  - b) 使所述核酸分子经受用于全基因组扩增或全转录组扩增的引物群以及用于靶向扩增的至少一种靶标特异性引物, 从而生成由所述全基因组扩增或全转录组扩增和所述靶向扩增所产生的扩增子的混合物;
  - c) 在测序仪上使用测序测定法对所述扩增子混合物进行测序, 从而生成测序读数; 以及
  - d) 评估所述测序读数以确定所述样品中的 SNP 和 CNV。
2. 根据权利要求1所述的方法, 其中, 利用聚合酶链式反应对所述核酸分子进行扩增。
3. 根据权利要求1所述的方法, 其中, 使用至少一种嵌套引物对使步骤 (b) 中产生的所述扩增子混合物进行另外的靶向扩增, 以进一步扩增由所述靶向扩增生成的扩增子。
4. 根据权利要求1所述的方法, 所述方法进一步包括: 使用所述测序读数来进行单核苷酸变异 (SNV) 基因分型, 微卫星基因分型, 检测插入和/或缺失, 确定合子型, 确定性别, 检测基因融合, 检测易位, 检测突变或检测染色体异常。
5. 根据权利要求1所述的方法, 其中, 所述引物群是自身不互补的并与所述群中的其他引物不互补的, 并且在5' 至3' 方向上包含恒定区和可变区, 其中, 所述恒定区序列具有在所述群的多个引物之间恒定的已知序列, 并且所述可变区序列在所述群的多个引物之间是简并的, 并且进一步地, 其中, 组成所述恒定区和可变区的序列不会在进行步骤 (a) - (c) 的条件下交叉杂交或自杂交。
6. 根据权利要求1所述的方法, 其中, 多个核酸分子至少为50个碱基对。
7. 根据权利要求1所述的方法, 其中, (b) 中的引物包含至少10个核苷酸。
8. 根据权利要求1所述的方法, 其中, 所述至少一种靶标特异性引物特异于一个或更多个靶序列。
9. 根据权利要求1所述的方法, 其中, 所述至少一种靶标特异性引物不包括接头序列。
10. 根据权利要求1所述的方法, 其中, 所述至少一种靶标特异性引物包括接头序列的至少一部分。
11. 根据权利要求1所述的方法, 其中, (b) 中的引物包含至少一个经修饰的核苷酸。
12. 根据权利要求1所述的方法, 其中, (b) 中的引物的解链温度为至少30摄氏度。
13. 根据权利要求1所述的方法, 其中, 所述核酸分子包括基因组DNA或RNA。
14. 根据权利要求1所述的方法, 其中, 所述样品选自血液, 血清, 血浆, 脑脊液, 面颊刮片, 乳头抽吸液, 活检, 宫颈样品, 精液, 体液, 微生物, 线粒体, 叶绿体, 细胞裂解物, 尿液, 粪便, 毛囊, 唾液, 汗液, 经免疫沉淀或物理分离的染色质, 循环肿瘤细胞, 肿瘤活检样品, 外泌体, 胚胎, 细胞培养基, 培养了细胞、组织、类器官或胚胎的废培养基, 活检胚胎, 滋养层, 羊水, 母体血液, 胎儿细胞, 胎儿DNA, 无细胞DNA, 子宫灌洗液, 子宫内膜液, 卵丘细胞, 颗粒细胞, 福尔马林固定的组织, 石蜡包埋的组织或囊胚腔。
15. 根据权利要求9所述的方法, 其中, 所述至少一种靶标特异性引物包括单个靶标特异性引物对。
16. 根据权利要求15所述的方法, 其中, 所述一个或更多个靶序列包括冗余基因组区域。

17. 根据权利要求16所述的方法, 其中, 所述冗余基因组区域包括重复元件。
18. 根据权利要求17所述的方法, 其中, 所述重复元件包括SVA元件。
19. 一种试剂盒, 包括:
  - a) 用于全基因组扩增或全转录组扩增的引物群;
  - b) 用于靶向扩增的至少一种靶标特异性引物; 和
  - d) 一套说明书, 以使用所述试剂盒来检测拷贝数变异 (CNV), 进行单核苷酸多态性 (SNP) 基因分型, 检测单核苷酸变异 (SNV), 进行微卫星基因分型, 检测插入和/或缺失, 确定合子型, 确定性别, 检测基因融合, 检测易位, 检测突变或检测染色体异常。

## 核酸文库的制备和分析方法

[0001] 相关申请的交叉引用

[0002] 根据35 U.S.C. §119(e), 本申请要求于2019年2月15日提交的序列号62/806,698的美国临时专利申请的提交日期的优先权;该申请的公开内容通过引用并入本文。

### 背景技术

[0003] 在相同样品中检测不同突变是必不可少的,尤其是在样品数量有限以及需要高通量方法快速检测突变的情况下。本领域常规使用的方法需要单独的测定方法来检测样品中的不同突变或突变类型(例如,单核苷酸多态性(SNP)或拷贝数变异(CNV))。使用单独的测定法可能会导致丢失数量有限的样品中具有临床意义的突变的风险。

### 发明内容

[0004] 本公开提供了用于在相同样品中检测不同突变例如SNP和CNV的方法。本文所述的方法可以用于植入前基因测试、携带者筛查或基因分型。

[0005] 一方面,本公开提供了一种检测样品中的单核苷酸多态性(SNP)和拷贝数变异(CNV)的方法。该方法包括:a)获得包含核酸分子的样品;b)使所述核酸分子经受用于全基因组扩增或全转录组扩增的引物群以及用于靶向扩增的至少一种靶标特异性引物,从而生成由所述全基因组扩增或全转录组扩增和所述靶向扩增所产生的扩增子的混合物;c)在测序仪上使用测序测定法对所述扩增子混合物进行测序,从而产生测序读数;以及d)评估所述测序读数以确定所述样品中的SNP和CNV。

[0006] 在一些实施方案中,利用聚合酶链式反应对所述核酸分子进行扩增。在一些实施方案中,多个核酸分子至少为50个碱基对。在一些实施方案中,所述核酸分子包括基因组DNA或RNA。

[0007] 在一些实施方案中,使用至少一种嵌套引物对使步骤(b)中产生的扩增子混合物进行另外的靶向扩增,以进一步扩增由所述靶向扩增生成的扩增子。

[0008] 在一些实施方案中,该方法进一步包括:使用所述测序读数来进行单核苷酸变异(SNV)基因分型,微卫星基因分型,检测插入和/或缺失,确定合子型(zygosity),确定性别,检测基因融合,检测易位,检测突变,或检测染色体异常。

[0009] 在一些实施方案中,所述引物群是自身不互补的并与所述群中的其他引物不互补的,并且在5'至3'方向上包含恒定区和可变区,其中,所述恒定区序列具有在所述群的多个引物之间恒定的已知序列,并且所述可变区序列在所述群的多个引物之间是简并的,并且进一步地,其中,组成所述恒定区和可变区的序列不会在进行步骤(a)-(c)的条件下交叉杂交或自杂交。

[0010] 在一些实施方案中,如(b)中的引物包含至少10个核苷酸。在一些实施方案中,所述至少一种靶标特异性引物特异于一个或更多个靶序列。在一些实施方案中,所述至少一种靶标特异性引物不包括接头序列。在一些实施方案中,所述至少一种靶标特异性引物包括接头序列的至少一部分。在一些实施方案中,如(b)中的引物包含至少一个经修饰的核苷

酸。在一些实施方案中,如(b)中的引物的解链温度为至少30摄氏度。在一些实施方案中,所述至少一种靶标特异性引物包括单个靶标特异性引物对。在一些实施方案中,所述一个或更多个靶序列包括冗余基因组区域。在一些实施方案中,所述冗余基因组区域包括重复元件(repetitive element)。在一些实施例中,所述重复元件包括SVA元件。

[0011] 在一些实施方案中,所述样品选自血液,血清,血浆,脑脊液,面颊刮片,乳头抽吸物,活检,宫颈样品,精液,体液,微生物,线粒体,叶绿体,细胞裂解物,尿液,粪便,毛囊,唾液,汗液,经免疫沉淀或物理分离的染色质,循环肿瘤细胞,肿瘤活检样品,外泌体,胚胎,细胞培养基,培养了细胞、组织、类器官或胚胎的废培养基(spent medium),活检胚胎,滋养层(trophoblast),羊水,母体血液,胎儿细胞,胎儿DNA,无细胞DNA,子宫灌洗液,子宫内膜液,卵丘细胞,颗粒细胞,福尔马林固定的组织,石蜡包埋组织或囊胚腔。

[0012] 在一个方面,本公开提供了一种试剂盒。该试剂盒包括:a)用于全基因组扩增或全转录组扩增的引物群;b)用于靶向扩增的至少一种靶标特异性引物;以及d)一套使用该试剂盒的说明书,其用于检测拷贝数变异(CNV),进行单核苷酸多态性(SNP)基因分型,检测单核苷酸变异(SNV),进行微卫星基因分型,检测插入和/或缺失,确定合子型,确定性别,检测基因融合,检测易位,检测突变或检测染色体异常。

[0013] 附图简要说明

[0014] 图1提供了一种方法的示意图,该方法使用相同的核酸样品,利用WGA引物进行全基因组扩增(WGA)以检测拷贝数变异(CNV),并利用靶标特异性引物进行靶向扩增以检测单核苷酸多态性(SNP)。

[0015] 图2提供了用于制备核酸分子以通过使用相同的核酸分子样品分别进行全基因组扩增(WGA)和靶向扩增来检测拷贝数变异(CNV)和单核苷酸多态性(SNP)的方案示例。

[0016] 图3提供了使用相同的核酸分子样品生成用于检测SNP和CNV的核酸文库分子的步骤的示意图。这些步骤可以包括具有WGA和靶向扩增的预扩增步骤,可选的清洁(clean-up)步骤,一个或更多个文库制备步骤例如利用嵌套PCR的靶向扩增步骤,以及索引PCR(Indexing PCR)步骤以生成用于测序的核酸文库分子。

[0017] 图4提供了冗余基因组元件的实施方案的示意图。

[0018] 图5提供了SINE/VNTR/Alu(SVA)元件的示意图(图5A)和与SVA元件的区域互补的靶标特异性引物的示意图(图5B)。

[0019] 图6提供了用于通过使用相同的核酸分子样品,利用WGA引物进行全基因组扩增(WGA)和利用与冗余基因组元件互补的靶标特异性引物进行靶向扩增来检测SNP和CNV的方法的示意图。

[0020] 图7提供了使用跨越靶序列的多重靶标特异性引物的示意图。

[0021] 图8A和图8B提供了使用三种不同的预扩增条件即在不存在靶标特异性引物、存在30种靶标特异性引物以及存在90种靶标特异性引物的情况下进行的实验中的数据。图8A示出了三个预扩增条件的覆盖度(coverage)。图8B示出了在三个预扩增条件之间的覆盖度的变化,如变化系数所指示的。

[0022] 图9A至图9D提供了在有或没有靶向扩增的情况下进行预扩增的实验中的数据。在任一种情况下(即在预扩增步骤中有或没有靶向扩增),都在预扩增步骤之后进行靶向扩增。图9A显示了利用在预扩增步骤中有或没有靶向扩增的测定法,跨越全基因组和靶序列

即CFTR基因的读段百分比。图9B显示了在预扩增步骤中有或没有靶向扩增的情况下的全基因组和CFTR基因的平均覆盖度。图9C显示了其中预扩增反应包括靶向扩增的测定法中的在CFTR基因中的十五个不同靶标或变体 (variants) 上的测序读段的覆盖度, 而图9D则显示了其中预扩增反应不包括靶向扩增的测定法的覆盖度。

[0023] 图10提供了使用5个细胞 (图10A) 或单个细胞 (图10B) 进行的实验中的测序读段的覆盖度数据。

[0024] 图11显示了使用5个细胞 (图11A) 或单个细胞 (图11B) 进行的实验中的测序读段的分布。

[0025] 图12A至图12C提供了对使用五个细胞的重复 (图12A) 或单个细胞的重复 (图12B) 的重复之间的相关性进行评估的实验中的数据。图12C示出了两次重复中1Mb bin的读段的  $\log_2$  比率的基因组图。

[0026] 图13提供了显示用于进行例如携带者筛查的仅使用靶向扩增而不使用WGA的CFTR基因上15个不同靶标的覆盖度的实验中的数据。

[0027] 图14提供了传统方法 (图14A) 和本方法 (图14B) 之间的比较示意图, 以及使用本方法 (图14C和图14D) 检测滋养外胚层 (trophectoderm) 活检中的CFTR基因的单核苷酸多态性 (SNP) 和染色体非整倍性 ( $n=4$ ) 的数据。使用本发明方法, 图14C提供了与检测CFTR基因中的SNP有关的数据, 并且图14D提供了与检测非整倍性有关的数据。

[0028] 图15提供了在人类基因组 (组装hg38) 上的SVA元件内发现的SNP的可视表示。顶部的条形表示各个染色体1-22, X和Y。底部的图将各个SNP描述为基因组中的点。Y轴表示每一SNP的次要等位基因频率。黑点表示次要等位基因频率大于或等于0.05的SNP。灰色点表示次要等位基因频率低于0.05的SNP。

[0029] 图16提供了靶标特异性引物对和每一引物对的一些预测的PCR产物或扩增子的实施方案。所述序列如下所示: Alu样引物序列从上到下 (SEQ ID NO:1-10); 和SINE-R引物序列从上到下 (SEQ ID NO:11-20)。

## 具体实施方式

[0030] 提供了通过扩增全基因组或转录组扩增 (WGA或WTA) 联合靶向扩增以扩增来自相同的核酸分子样品的全基因组和靶序列来制备和分析核酸分子的方法。该方法可以用于检测相同样品中的各种突变, 例如拷贝数变异 (CNV)、插入和/或缺失 (indel) 和单核苷酸多态性 (SNP)。这些方法可用于临床测试 (例如, 携带者筛查、胚胎筛检、废培养基测试), 法医学分析等。

[0031] 在更详细地描述本发明之前, 应当理解, 本发明不限于所描述的特定实施方案, 因为这些当然可以改变。还应理解, 本文中使用的术语仅出于描述特定实施方案的目的, 而非旨在限制本发明, 因为本发明的范围将仅由所附权利要求书限定。

[0032] 在提供数值范围的情况下, 应理解为, 除非上下文另有明确规定, 否则介于该范围的上限和下限之间的每个中间值直到下限的十分之一单位, 以及在所述范围中的其他说明的数值或中间值被涵盖在本发明之内。这些较小范围的上限和下限可以独立地包括在所述较小范围内, 并且也被涵盖在本发明之内, 但在所述范围内的任何被明确排除的限值除外。当所述范围包括一个或两个限值时, 排除了那些所包括的限值中的一个或两个限值的范围

也包括在本发明中。

[0033] 本文提供的某些范围,在数值之前带有术语“约”。术语“约”在本文中用于为其后的确切数字以及与该术语后的数字接近或近似的数字提供文字支持。在确定数字是否接近或近似于具体列举的数字时,接近或近似的未列举数字可以是这样的数字,在呈现其的上下文中,该数字提供了具体列举的数字的基本等同量。

[0034] 除非另有定义,否则本文中使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常所理解的相同的含义。尽管在本发明的实践或测试中也可以使用与本文描述的那些方法和材料类似或等同的任何方法和材料,但是现在对代表性的说明性方法和材料进行描述。

[0035] 本说明书中引用的所有出版物和专利均通过引用并入本文,就如同每个单独的出版物或专利均被具体地和单独地指出通过引用并入本文一样,并通过引用并入本文以公开和描述与被引用出版物的内容相关的方法和/或材料。任何出版物的引用都是针对其在本申请日之前的公开,并且不应被解释为承认本发明无权凭借在先发明而早于该出版物。此外,提供的公布日期可能与实际公布日期有所不同,实际公布日期可能需要独立确认。

[0036] 注意,如本文和所附权利要求书中所使用的,单数形式“一个”、“一种”和“该”包括复数的指称对象,除非上下文另外明确指出。还应注意的是,权利要求书可以被撰写为排除任何可选要素。这样,该陈述旨在作为与权利要求要素的叙述相关联的诸如“只”、“仅”等排他性术语的应用、或“否定”限制的应用的先行基础。

[0037] 对于本领域技术人员而言,在阅读本公开后将显而易见的是,本文描述和示出的各个实施方案中的每一个具有离散的组件和特征,在不脱离本发明的范围或精神的情况下,这些组件和特征可以容易地与任何一个其他若干实施方案中的特征分开或组合在一起。任何叙述的方法都可以按照叙述的事件的顺序或在逻辑上可能的任何其他顺序进行。

[0038] 尽管为了语法的流畅性,设备和方法已经或将要与功能上的说明一起进行描述,但应明确地理解,除非已遵循35U.S.C.§112被确切阐述,否则这些权利要求不应被解释为必须由于“方式”或“步骤”限制的结构而以任何方式受到限制,而应遵循等同司法原则被授予由权利要求书所提供的限定的含义和等同物的全部范围,并且在权利要求已遵循35U.S.C.§112进行了明确的阐述的情况下这些权利要求根据35U.S.C.§112被授予完全的法定等同。

[0039] 方法

[0040] 如上所述,本公开中描述的方法涉及以相同样品(例如血液、细胞、废培养基或提取的核酸)制备和分析核酸分子以检测各种突变(例如,拷贝数变异和单核苷酸多态性)。广义地,该方法包括使用用于全基因组扩增(WGA)或全转录组扩增(WTA)的引物,结合和/或随后使用靶标特异性引物进行靶序列的靶向扩增,来扩增核酸分子。例如,如图1所示,可以使用相同的核酸样品来进行使用WGA引物进行全基因组扩增(WGA)以检测拷贝数变异(CNV)结合使用靶标特异性引物进行包含SNP的靶向扩增以检测SNP。

[0041] 本文公开的方法可以包括各种步骤。在图2和图3中提供了一种这样的方案的示例。该方案可以包括以下步骤,例如获得包含核酸分子的样品,裂解该样品以从样品中提取核酸分子,使核酸分子经历预扩增步骤以利用WGA/WTA引物进行全基因组扩增结合利用靶标特异性引物扩增靶序列进行靶向扩增,可选地执行清洁步骤,然后使扩增子经历文库制

备程序以制备用于测序的文库分子。文库制备步骤可以包括一个或更多个步骤以连接测序测定所需的序列。文库制备步骤可以包括多于一个的步骤,例如,其中预扩增步骤不包括靶向扩增,或者预扩增步骤包括靶向扩增,但是可以在预扩增步骤之后应用另外的靶向扩增。在预扩增步骤之后进行另外的靶向扩增的情况下,可以进行嵌套PCR以进一步扩增靶序列并连接接头序列(例如,P5或P7)。嵌套PCR可以利用嵌套在预扩增步骤中使用的靶标特异性引物内部的引物来进行。在某些情况下,可以以单一步骤制备文库,从而以单一反应连接接头序列和索引(indices)。例如,预扩增步骤可以包括靶向扩增,并且在预扩增步骤之后的另外的靶向扩增可以是可选的。在这种情况下,可以以单一步骤(例如在索引PCR期间)制备文库。可以在预扩增步骤和/或靶向扩增之后进行索引PCR,以将索引(例如,index 1或index 2)连接至扩增子。该方法的各个步骤在图1至图3以及下面进行了更详细的描述。

#### [0042] 样品

[0043] 本公开中的方法可以用于包含核酸分子例如DNA或RNA的各种样品。在某些情况下,样品可以是血液,血清,血浆,脑脊液,面颊刮片,宫颈液/细胞,乳头抽吸物,活检,精液,尿液,粪便,毛囊,唾液,汗液,经免疫沉淀或物理分离的染色质,循环肿瘤细胞,肿瘤活检,外泌体,胚胎,细胞培养基,培养了细胞、组织、类器官或胚胎的废培养基,活检胚胎(例如来自囊胚内细胞团(ICM)的一种或更多种细胞或来自滋养外胚层(TE)中的一种或更多种细胞,即滋养外胚层细胞),羊水,福尔马林固定的组织,母体血液,胎儿细胞,无细胞DNA,子宫灌洗液,子宫内膜液,卵丘细胞,颗粒细胞,癌细胞,石蜡包埋的组织或囊胚腔。在某些情况下,样品可以是卵母细胞或其极体,微生物,植物细胞,动物细胞,线粒体,叶绿体,法医样品,细胞裂解物,体液,宫颈样品。也可以使用其他类型的包含核酸分子的样品。

#### [0044] 细胞裂解和核酸分子的提取

[0045] 可以将包含核酸分子的样品裂解以释放核酸分子。在某些情况下,样品可以使用本领域已知的任何方法来裂解,例如基于试剂的方法和物理方法。例如,基于试剂的方法可以包括使用酶(例如,溶菌酶)和/或有机溶剂(例如,醇,氯仿,醚,EDTA,triton,碱裂解)。物理方法的实例可以包括超声处理,均质器,冻融循环,研磨等。在某些情况下,可能不需要细胞裂解,并且可以将样品直接利用本文公开的方法来制备核酸分子。例如,样品可以是能够用于本公开内容的方法的无细胞DNA。

[0046] 在一些实施方案中,可以用于本文所述方法的核酸分子的数量/量可以是至少0.5皮克(pg),至少1pg,至少2pg,至少5pg,至少10pg,至少20pg,至少30pg,至少40pg,至少50pg,至少100pg,至少200pg,至少500pg,至少1纳克(ng)或大于1ng。其他量也可以用于本公开中的方法。

[0047] 在一些实施方案中,可以用于本公开中的方法的核酸分子的品质可以是不存在大量抑制剂的高品质核酸分子,例如使用本领域公开的方法提取的DNA。在某些情况下,核酸分子样品可以包括抑制剂,例如经福尔马林固定的样品。

#### [0048] 预扩增

[0049] 核酸分子可以经历预扩增步骤。预扩增步骤可以包括使核酸分子经受用于全基因组扩增(WGA)或全转录组扩增(WTA)的引物。在一些实施方案中,预扩增步骤可以包括靶标特异性引物以进行靶向扩增,从而生成来自WGA/WTA和靶向扩增的扩增子的混合物。在某些情况下,预扩增步骤可以不包括靶标特异性引物,因此,预扩增步骤可以生成仅来自WGA的



扩增子。在这种情况下,可以在预扩增步骤之后进行靶向扩增,以利用靶标特异性引物扩增靶序列。在其中预扩增反应可以包括WGA/ WTA引物与靶标特异性引物的组合以生成扩增子混合物的实施方案中,可以进一步使用嵌套在由所述预扩增步骤中的靶向扩增所产生的扩增子内部的引物对所述扩增子混合物进行靶向扩增。在一些特定的实施方案中,可以不进行预扩增步骤。在这种情况下,使用靶标特异性引物对核酸分子进行靶向扩增以扩增靶序列。

[0050] WGA或WTA基本上可以扩增样品中核酸分子的所有片段。WGA或WTA可以扩增基本上整个基因组或整个转录组,而不会丢失特定位点的表达(representation)。基本上所有或基本上整个可以指基因组或转录组中所有序列中的约30%,约40%,约50%,约60%,约70%,约80%,约85%,约90%,约95%或更多。在某些情况下,WGA或WTA可以包括特定序列相对于其他序列的非等价扩增(non-equivalent amplification),尽管这种扩增的相对差异在某些情况下并不明显。WGA/ WTA可以靶向基因组或转录组中的一个或更多个序列。在大多数情况下,WGA/ WTA可以靶向基因组或转录组中的至少约100,至少约1000,至少约10,000,至少约100,000,至少约1,000,000,至少约10,000,000,至少约100,000,000,至少约1,000,000,000个位点或更多位点。WGA和/或WTA可以利用任何合适的引物进行。合适的WGA/ WTA引物包括但不限于在以下试剂盒中提供的引物:**PicoPLEX®** WGA试剂盒,**SMARTer® PicoPLEX®** Single Cell WGA试剂盒,**SMARTer® PicoPLEX®** DNA-seq试剂盒,**SMARTer® PicoPLEX®** Gold Single Cell DNA-Seq试剂盒,Ion ReproSeq™ PGS试剂盒,**MALBAC®** Single Cell WGA试剂盒,**GenomePlex®** WGA试剂盒,**REPLI-g®** WGA和WTA试剂盒,**Amplii1™** WGA和WTA试剂盒,**Transplex®** WTA试剂盒,**TruePrime®** WGA试剂盒,**Quantitect®** 全转录组试剂盒,**Dopliify®** WGA试剂盒,**GenoMatrix™** WGA试剂盒,**PG-Seq™**试剂盒,**Sureplex™** DNA扩增系统试剂盒,**Illustra GenomiPhi™** DNA扩增试剂盒。合适的WGA/ WTA引物可以描述于,例如,美国专利号7,718,403;8,206,913;9,249,459;9,617,598;5,731,171;6,365,375;10,017,761;8,034,568;6,617,137;6,977,148,10,190,163;9,840,732;9,777,316;8,512,956;8,349,563,其中每一个文件的内容通过引用并入本文,以及美国专利公开号2016/0355879;2018/0030522;2019/0271033;2013/0085083;2007/0054311;2007/0178457;2011/0033862;2016/0312276;2009/0099040;2010/0184152;2015/0072899;2011/0189679;2019/0300933;2016/0289740,其中的每个文件的内容通过引用并入本文。

[0051] 类似地,靶标特异性引物可以在靶标扩增过程中扩增基因组或转录组中的一个或更多个序列。在某些情况下,靶标特异性引物可以扩增一个序列,2个序列,3个序列,10个序列,100个序列,1000个序列,10,000个序列,100,000个序列,1,000,000个序列,10,000,000个序列或更多。在某些情况下,靶向扩增可以使用一种或更多种靶标特异性引物来扩增相同的序列。在其他情况下,靶向扩增可以扩增基因组或转录组中的不同序列。在某些情况下,“靶标特异性引物”是指在合适的杂交条件下选择性地和可预测地与靶序列杂交的引物。在一些情况下,“靶序列”或“感兴趣的靶序列”及其衍生物通常是指可以根据本公开内容扩增的任何单链或双链核酸序列,包括怀疑或预期存在于样品中的任何核酸序列。在一些实施方案中,在添加靶标特异性引物之前,靶序列以双链形式存在,并且包括待扩增或合

成的特定核苷酸序列的至少一部分或其互补序列。靶序列可以包括在通过聚合酶扩展之前可以与靶标特异性引物杂交的核酸。在某些情况下,靶标特异性引物扩增这样的靶序列,其包括一个或更多个突变热点、基因组标志物、感兴趣的单核苷酸多态性、冗余基因组元件(例如SVA元件)、编码区、外显子、基因、内含子、非编码区、启动子区、假基因、内含子-外显子连接和基因间区域。在某些情况下,靶标特异性引物可以扩增这样的靶序列,其包括一个或更多个感兴趣的基因组区域,例如,感兴趣的基因(例如,CFTR基因)或感兴趣的基因的一个或更多个区域。在某些情况下,靶标特异性引物可以扩增包括有一个或更多个感兴趣的SNP的靶序列。在一些情况下,靶标特异性引物可以扩增这样的靶序列,其包括与遗传障碍例如本文公开的任何遗传障碍有关的基因或基因组区域。

[0052] 在某些实施方案中,靶标特异性引物的一个或更多个靶序列包括冗余基因组区域或冗余基因组元件,即在例如人的基因组各处存在的基因组区域。冗余基因组区域可以例如以均匀的方式存在于所有染色体上。在一些情况下,冗余基因组区域存在于基因组中的多个位置,例如基因组中的1000或更多个位置,基因组中的2000或更多个位置,基因组中的3000或更多个位置,基因组中的4000或更多个位置,基因组中的5000或更多个位置,基因组中的6000或更多个位置,基因组中的7000或更多个位置,基因组中的8000或更多个位置,基因组中的9000或更多个位置,基因组中的10,000或更多个位置,基因组中100,000或更多个位置,基因组中1,000,000或更多个位置,基因组中10,000,000或更多个位置或基因组中100,000,000或更多个位置。在某些情况下,冗余基因组区域存在于基因组中的多个位置,其范围为基因组中的1000至10,000,000个位置,基因组中的1000至1,000,000个位置,基因组中的10,000至500,000个位置,或基因组中的50,000至200,000个位置。

[0053] 存在于基因组中多个位置的基因组区域可以在序列上是不同的,例如,使得基因组区域独一无二地映射在基因组上。在某些情况下,冗余基因组区域是多态的(例如,包括SNP)。如本文在其常规意义上所使用的,“多态性”是指其中可以在群体中发现特定基因组序列的两个或更多个变体的情况。在某些情况下,冗余基因组区域包括一个或更多个多态性区域。多态性区域可以包括插入、缺失、结构变体连接、可变长度的串联重复、单核苷酸突变、单核苷酸变异、拷贝数变异或它们的组合。在某些情况下,多态性区域的次等位基因频率的范围为大于或等于0.01,大于或等于0.02,大于或等于0.03,大于或等于0.04,大于或等于0.05,大于或等于0.06,大于或等于0.07,大于或等于0.08,大于或等于0.09,大于或等于0.1,大于或等于0.2,大于或等于0.3,或者大于或等于0.4。在某些情况下,一个或更多个多态性区域提供每个区域一个或更多个SNP,例如每个区域1-5个SNP,每个区域10-20个SNP,每个区域10-40个SNP,每个区域15-35个SNP,每个区域20-60个SNP,或每个区域20-50个SNP。在某些情况下,冗余基因组区域包括一个或更多个保守区域。如本文在常规意义上所使用的,“保守区域”是指异源多核苷酸或多肽序列或者在不同物种中存在的或在基因组内重复的多核苷酸或多肽序列中的其中在不同的序列之间具有相对高度的序列同一性的区域。保守区域之间的序列同一性可以是至少75%,至少80%,至少85%,至少90%,至少95%,至少98%或至少99%。在某些情况下,冗余基因组区域包括在其两端侧翼均为保守区域的多态性区域。在某些情况下,冗余基因组区域包括基因组的非编码区域。感兴趣的基因组区域可以包括:例如,一个或更多个内含子,一个或更多个调控元件,一个或更多个假基因,一个或更多个重复序列或重复元件,一个或更多个病毒元件(例如内源性逆转录病毒序

列),一个或更多个端粒,一个或更多个转座元件,一个或更多个逆转录转座子,一个或更多个短串联重复,它们的一部分或它们的组合。

[0054] 冗余基因组区域可以具有适合于利用本发明方法进行扩增的任何长度。在某些情况下,冗余基因组区域的长度范围为1000至4000个碱基对(bp),1000至3000bp,1000至2000bp或500至1500bp。在某些情况下,基因组区域的长度范围为1至500个碱基对(bp),10至500bp或100至500bp。

[0055] 图4提供了存在于基因组上多个位置的冗余基因组元件的实施方案的示意图(顶部),以及具有在两端侧翼均为保守区域的多态性区域的冗余基因组元件的实施方案的示意图(底部)。冗余基因组元件可以在基因组各处找到,并以相对均匀的方式存在于所有染色体上。基因组可以包括冗余基因组元件的1500-3000个拷贝或更多,3000-30000个拷贝或更多,30000-300000个拷贝或更多,其数量约等于基因组每1-2Mb一个SNP区域。

[0056] 在某些情况下,冗余基因组区域包括重复元件或重复序列。重复元件可以包括一个或更多个串联重复、一个或更多个分散重复(interspersed repeat)或它们的组合。

[0057] 串联重复可以包括一个或更多个卫星DNA、一个或更多个小卫星(长串联重复;10-100bp的重复单元)、一个或更多个微卫星(短串联重复;小于10bp的重复单元)或它们的组合。在某些情况下,冗余基因组区域包括VNTR(可变数目串联重复)。在某些情况下,冗余基因组区域包括大卫星(重复单元长于100bp)。

[0058] 分散重复可以散布在基因组各处的基因序列内部或基因间。分散重复可以包括一个或更多个转座子。转座子可以是可移动的遗传元件。可移动的遗传元件可以改变其在基因组中的位置。转座子可以分为I类可转座元件(I类TE)或II类可转座元件(II类TE)。I类TE(例如逆转录转座子)可以在两个阶段中复制自身,首先是通过转录从DNA复制到RNA,然后通过逆转录从RNA复制回到DNA。然后可以将DNA拷贝插入基因组的新位置中。I类TE可以包括一个或更多个长末端重复(LTR)、一个或更多个长分散核元件(LINE)、一个或更多个短分散核元件(SINE)或它们的组合。LTR的实例包括但不限于人内源性逆转录病毒(HERV)、中等反复重复(medium reiterated repeats)4(MER4)和逆转录转座子。LINE的示例包括但不限于LINE1和LINE2。SINE可以包括一个或更多个Alu序列、一个或更多个哺乳动物范围的分散重复(MIR)或它们的组合。II类TE(例如DNA转座子)通常不涉及RNA中间体。DNA转座子通常被从一个位点切下,然后被插入基因组的另一个位点。或者,DNA转座子被复制并被插入到基因组中的新位置中。DNA转座子的实例包括但不限于MER1、MER2和Mariner。

[0059] 分散重复可以包括一个或更多个逆转录转座元件。逆转录转座元件(RE)包括长分散核元件(LINE)、短分散核元件(SINE)和SVA元件。SINE是一类其长度通常不到500个核苷酸的RE;而LINE的长度通常大于500个核苷酸(A.F.A.Smit,The origin of interspersed repeats in the human genome(人类基因组中的分散重复序列的起源),Current Opinion in Genetics Development,6(6):743-748(1996);Batzer,M.A.等,Alu repeats and human genomic diversity(Alu重复序列和人类基因组多样性),Nature Reviews Genetics,3(5):370-379(2002);Batzer,M.A.等人,African origin of human-specific polymorphic Alu insertions(人类特有的多态性Alu插入的非洲起源),Proceedings of the National Academy of Sciences,91(25):12288(1994);Feng,Q.等,Human L1 retrotransposon encodes a conserved endonuclease required for

retrotransposition(人L1逆转录转座子编码逆转录转座所需的保守的核酸内切酶), Cell, 87(5):905-916(1996); Houck, C.M., 等, A ubiquitous family of repeated DNA sequences in the human genome(人类基因组中无处不在的重复DNA序列家族), Journal of Molecular Biology, 132(3):289-306(1979); Kazazian, H.H. 等人, The impact of L1 retrotransposons on the human genome(L1逆转录转座子对人类基因组的影响), Nature Genetics, 19(1):19-24(1998); Ostertag, E.M. 等, Biology of mammalian L1 retrotransposons(哺乳动物L1逆转录转座子生物学), Annual Review of Genetics, 35(1):501-538(2001)。LINE全长元件的长度约为6kb, 包含一个聚合酶II的内部启动子和两个开放阅读框(ORF), 并以多聚A尾结尾。SINE包括Alu元件, 其是灵长类特异性SINE, 在人类基因组中拷贝数超过一百万。SINE最初是通过其散布的性质和长度(75-500bp)来定义的, 但现在已通过其RNA聚合酶III转录进行了进一步表征。

[0060] RE的第三种类型是称为SVA(SINE/VNTR/Alu)元件的复合逆转录转座子(Wang, H., 等, SVA Elements: A Hominid-specific Retroposon Family(SVA元件: 人类特有的逆转座子家族), J. Mol. Biol. 354:994-1007(2005))。SVA在进化上很年轻, 并且可能被LINE-1逆转录酶反式动员。SVA当前活跃, 并可以通过多种机制影响宿主, 包括插入诱变、外显子重排、替代性拼接以及生成差异甲基化区域(DMR)。SVA的每个结构域均来自逆转录转座子或重复序列。典型的SVA平均约为2千个碱基(kb)(例如, ~1,650bp), 但是SVA插入的大小的范围可以为700-4000个碱基对(bp)(Hancks, D.C.和Kazazian, H.H., SVA Retrotransposons: Evolution and Genetic Instability(SVA逆转录转座子: 进化和遗传不稳定性), Semin. Cancer Biol., 20:234-45(2010))。SVA是按照其主要组件(SINE、可变数目串联重复(VNTR)和Alu)来命名的复合元件。SVA元件包含逆转录转座子的特征, 因为它们的侧翼是靶位点重复(TSD), 以多聚(A)尾终止, 并且在它们整合到基因组过程中偶尔会被截短和倒置。典型的SVA通常包含五个不同的区域: 在5'端的(CCCTCT)<sub>n</sub>(SEQ ID NO:25)六聚体重复, Alu样结构域, 可变数目串联重复(VNTR), SINE衍生区域(例如SINE-R, 其中R表示逆转录病毒来源)和多聚(A)尾。由于重复的结构域, 例如VNTR区域, 全长SVA元件的大小可以有很大变化。SVA可以分为六个亚家族, 分别称为SVA\_A, SVA\_B, SVA\_C, SVA\_D, SVA\_E, SVA\_F。利用家族共有序列, 家族的同源性在90-95%的范围内。在第七亚家族SVA-F1中, (CCCTCT)<sub>n</sub>(SEQ ID NO:25)六聚体被MAST2基因第一个外显子的5'转导所替代(Quinn, J., 等, The Role of SINE-VNTR-Alu(SVA) Retrotransposons in Shaping the Human Genome(SINE-VNTR-Alu(SVA)逆转录转座子在塑造人类基因组中的作用), Int. J. Mol. Sci. 20:5977(2019))。

[0061] 在某些情况下, SVA元件是多态的(例如, 包括SNP)。SVA元件的多态性区域可包括本文所述的SVA元件的任何结构域和区域中的一种或更多种。在某些情况下, SVA元件的Alu样结构域是多态的。在某些情况下, SVA元件的SINE-R区域是多态的。在某些情况下, SVA元件的保守区域包括靶位点重复(target site duplication)结构域、六聚体重复、VNTR和多聚-A尾中的一个或更多个。在图5A中提供了SVA元件的实施方案(改编自Wang, H., 等人, SVA Elements: A Hominid-specific Retroposon Family, J. Mol. Biol. 354:994-1007(2005))。该SVA元件包括: 两个侧翼靶位点重复结构域; 六聚体重复(CCCTCT)<sub>n</sub>(SEQ ID NO:25); Alu样结构域, 其包括通过SVA-U(335nt)连接的两个不完全的Alu元件; VNTR区域(长度介于48-2306bp之间; 平均长度: 819bp); SINE-R区域, 其由人内源性逆转录病毒的片段(env, U3, R)

(490nt) 组成;和多聚-A尾。

[0062] 在某些情况下,冗余基因组区域包括假基因。如本文所用,“假基因”是指与已鉴定的基因具有高度序列相似性或序列同一性,但由于无功能启动子、缺失起始密码子或其他缺陷而通常未被转录和未被翻译的序列。大多数假基因是无内含子的,主要代表亲本基因的编码序列。对于某些情况,已经表明在不同的生物或组织中可能发生功能性激活。

[0063] 在一些情况下,如上所述的靶向扩增包括使用一个或更多个靶标特异性引物对来扩增靶序列。在一些情况下,所述一个或更多个靶标特异性引物对包括五十个以下的引物对,十五个以下的引物对,十个以下的引物对,九个以下的引物对,八个以下的引物对,七个以下的引物对,六个以下的引物对,五个以下的引物对,四个以下的引物对,三个以下的引物对,两个以下的引物对或单个引物对。在某些实施方案中,本发明方法包括使用WGA/ WTA引物结合和/或随后使用至少一种靶标特异性引物来扩增核酸分子,其中所述至少一种靶标特异性引物包括单个靶标特异性引物对。

[0064] 在一些情况下,在本发明方法中用于靶向扩增的靶标特异性引物包括用于扩增如上所述的冗余基因组区域的单个引物对。在一些情况下,单个引物对的引物对于冗余基因组区域或冗余基因组区域中的一个或更多个部分(例如,冗余基因组区域的多态性区域)具有特异性或互补。在某些情况下,单个引物对的引物对于重复元件例如SVA元件的一个或更多个区域或结构域是特异性的。在图5B中,提供了与Alu样结构域的部分或SINE-R结构域的部分互补的引物对中的引物。在一些情况下,单个引物对中的一个或更多个引物与SVA元件的Alu样结构域或Alu样结构域的一部分互补。在某些情况下,单个引物对中的一个或更多个引物与SVA元件的SINE-R区域或SINE-R区域的一部分互补。在某些情况下,本发明方法除WGA/ WTA准随机引物(quasi-random primers)以外还包括使用对冗余基因组元件(例如SVA元件)具有特异性的单个引物对进行靶向扩增,这可用于基于SNP的CNV调用,检测单亲二体性,检测染色体镶嵌或进行连锁分析。

[0065] 图6提供了用于通过WGA和冗余基因组元件的靶向扩增来检测各种突变例如SNP和CNV的方法的实施方案。在图6中,准随机WGA引物提供了对基因组的浅且均匀的覆盖,并且用于冗余基因组元件的靶标特异性引物提供了对含SNP区域的强劲的覆盖。

[0066] 在一些情况下,WGA/ WTA引物和/或靶标特异性引物的长度可以是至少约5个碱基对(bp),6bp,7bp,8bp,9bp,10bp,11bp,12bp,13bp,14bp,15bp,16bp,17bp,18bp,19bp,20bp,21bp,22bp,23bp,24bp,25bp,26bp,27bp,28bp,29bp,30bp,31bp,32bp,33bp,34bp,35bp,36bp,37bp,38bp,39bp,40bp,50bp,60bp,70bp,80bp,90bp,100bp或更多。

[0067] 在某些情况下,WGA/ WTA引物和/或靶标特异性引物的解链温度可以为至少约10℃,15℃,20℃,25℃,30℃,35℃,40℃,45℃,50℃,60℃,65℃,70℃或更高。在某些情况下,WGA/ WTA引物可以具有与靶标特异性引物相同的解链温度。在其他情况下,WGA/ WTA引物可以具有与靶标特异性引物不同的解链温度。

[0068] 在一些情况下,WGA/ WTA引物和/或靶标特异性引物的GC含量可以为至少约5%,10%,15%,20%,25%,30%,35%,40%,45%,50%,55%,60%或大于60%。在某些情况下,WGA/ WTA引物可以具有与靶标特异性引物相同的GC含量。在其他情况下,WGA/ WTA引物可以具有与靶标特异性引物不同的GC含量。

[0069] 在某些情况下,WGA/ WTA引物和/或靶标特异性引物的浓度可以为1纳摩尔(nM),

10nM, 20nM, 30nM, 40nM, 50nM, 60nM, 70nM, 80nM, 90nM, 100nM或更高。在一些情况下, WGA/ WTA引物和/或靶标特异性引物的浓度可以是至少5微摩尔( $\mu\text{M}$ ), 10 $\mu\text{M}$ , 15 $\mu\text{M}$ , 20 $\mu\text{M}$ , 25 $\mu\text{M}$ , 30 $\mu\text{M}$ , 40 $\mu\text{M}$ , 50 $\mu\text{M}$ , 100 $\mu\text{M}$ , 200 $\mu\text{M}$ , 300 $\mu\text{M}$ , 400 $\mu\text{M}$ , 500 $\mu\text{M}$ , 600 $\mu\text{M}$ , 700 $\mu\text{M}$ , 800 $\mu\text{M}$ , 900 $\mu\text{M}$ 或更高。在某些情况下, WGA/ WTA引物可以具有与靶标特异性引物相同的引物浓度。在其他情况下, WGA/ WTA引物可以具有与靶标特异性引物不同的引物浓度。

[0070] 在某些情况下, 通过WGA/ WTA引物和/或靶标特异性引物生成的扩增子的大小可以为至少约50bp, 100bp, 150bp, 200bp, 250bp, 300bp, 350bp, 400bp, 450bp, 500bp, 550bp, 600bp, 650bp, 700bp, 750bp, 800bp, 850bp, 900bp或更多。在某些情况下, WGA/ WTA引物与靶标特异性引物可以生成基本上相似大小的扩增子。在其他情况下, WGA/ WTA引物与靶标特异性引物可以生成基本上不同大小的扩增子。在某些情况下, WGA/ WTA引物可以在WGA或WTA期间生成基本上相似大小的扩增子。在某些情况下, WGA/ WTA引物可以在WGA/ WTA期间生成基本上不同大小的扩增子。在一些情况下, 靶标特异性引物可以在一个或更多个靶序列的靶标特异性扩增期间生成基本上相似大小的扩增子。在一些情况下, 靶标特异性引物可以在一个或更多个靶序列的靶标特异性扩增期间生成基本上不同大小的扩增子。在某些情况下, WGA/ WTA引物和靶标特异性引物对基因组的相同或基本相同的区域进行扩增。例如, 靶标特异性引物可以嵌套在WGA/ WTA引物之中, 反之亦然。在一些情况下, WGA/ WTA引物和靶标特异性引物可以生成相同或基本相同的扩增子。例如, WGA/ WTA引物和靶标特异性引物可以共享在核酸分子上相同或基本相同的结合位点。

[0071] 在某些情况下, WGA/ WTA引物和/或靶标特异性引物可以具有不同的核苷酸序列。例如, 群体中所有的或基本上所有的WGA/ WTA引物可以具有不同的核苷酸序列。类似地, 群体中所有的或基本上所有的靶标特异性引物可以具有不同的核苷酸序列, 尤其是当诸如在多重反应中靶向超过一个序列时。

[0072] 在某些情况下, WGA/ WTA引物和/或靶标特异性引物可以包括附加的序列, 诸如接头序列或条形码(barcode), 例如在下文中所述的独一无二的分子条形码: Winzeler等, (1999) Science 285:901; Brenner (2000) Genome Biol. 1:1; Kumar等, (2001) Nature Rev. 2:302; Giaever等, (2004) Proc. Natl. Acad. Sci. USA 101:793; Eason等, (2004) Proc. Natl. Acad. Sci. USA 101:11046; 以及Brenner (2004) Genome Biol. 5:240, 其中每一个文件也都通过引用整体并入本文。例如, WGA/ WTA引物可以包括基本完整的或部分的Illumina接头序列, 例如用于流动池(flow cell) 附接位点的序列(例如, P5、P7), 用于测序引物结合位点的序列(例如, Read引物1, Read引物2), 索引序列等。在某些情况下, WGA/ WTA引物和/或靶标特异性引物不包括任何附加序列。在其他一些情况下, WGA/ WTA引物可以包括附加序列, 而靶标特异性引物不包括任何附加序列。基于靶向扩增实施的步骤以及靶向扩增进行的次数, 靶标特异性引物可包括附加序列。例如, 如果靶向扩增结合和/或在WGA之后实施, 那么在WGA之后实施的靶向扩增中使用的靶标特异性引物可以包括完整的或部分的接头序列。另一方面, 如果在预扩增步骤中而不是在任何后续步骤中将靶标特异性引物与WGA引物组合在一起, 则靶标特异性引物可以包括接头序列。

[0073] 在一些情况下, WGA/ WTA引物和/或靶标特异性引物可以具有一个或更多个经修饰的核苷酸, 例如锁核酸(LNA)、蛋白质核酸(PNA) 和甲基化核酸等。在某些情况下, 修饰可以包括具有一个或更多个硫代磷酸酯键、荧光团、生物素、氨基修饰物、硫醇修饰物、炔烃修饰

物、叠氮类修饰物、间隔基等的核酸。经修饰的核苷酸可以有益于交联、双链稳定或核酸酶抗性。例如，经修饰的核苷酸可以有益于保护核酸分子免受核酸外切酶或具有核酸外切酶活性的聚合酶的活性的影响。在某些情况下，WGA/ WTA引物和/或靶标特异性引物可以在寡核苷酸的一个或两个末端（例如5' 末端，3' 末端）具有经修饰的核苷酸。在某些情况下，WGA/ WTA引物和/或靶标特异性引物可以在寡核苷酸的一个末端（例如5' 末端或3' 末端）具有修饰的核苷酸。

[0074] 在一些情况下，可以将WGA/ WTA引物和/或靶标特异性引物设计为基本上自身不互补的，并且与群体中的其他引物基本上不互补。例如，WGA/ WTA引物可以被设计为包含非互补碱基，例如鸟嘌呤 (G) 和胸腺嘧啶 (T) 或胞嘧啶 (C) 和腺嘌呤 (A)，以限制群体中碱基的相互作用，从而防止过量的引物二聚体形成，减少完全或偶发的基因座缺失，减少非常短的扩增产物的产生，和/或降低无法扩增单链、短的或片段化的DNA和RNA分子的能力。在一些情况下，WGA/ WTA引物和/或靶标特异性引物可以具有一个或多个简并核苷酸，其中，所述识别 (identify) 可以选自多样化的核苷酸选择而不是限定的序列。在整个WGA/ WTA和/或靶标特异性引物中简并核苷酸可以有均匀的间距。通过将简并核苷酸包含在特定位置（例如每隔一个碱基、每隔两个碱基或每隔3个碱基）或通过实验人员发现对其特定应用有益的任何其他排列方式，可以使简并核苷酸均匀地间隔开。在其他情况下，简并核苷酸可以限于引物中的简并或可变区。简并或可变区的实例可以包括一个或多个“N”残基，其中N=任何碱基。这样的简并或可变区可以处于引物序列的5' 末端和/或3' 末端。在某些情况下，5' 末端可以包含除了自身不互补的碱基和非互补碱基以外的一个或多个核苷酸。在一些情况下，WGA引物的可变区或简并区可以包括接头序列，例如Illumina接头序列，如P5或P7。在一些情况下，在恒定区与可变区或简并区或WGA/ WTA引物的任一末端之间可以包括其他序列。

[0075] 在某些情况下，WGA/ WTA引物和/或靶标特异性引物可以与核酸分子上的相邻或重叠的位置互补。例如，如图7所示，可以将正向和反向的靶标特异性引物设计为在核酸分子上彼此相邻。这样的靶标特异性引物可以生成由正向引物与反向引物之间的各种组合产生的多个扩增子。如图7所示，三个正向引物和三个反向引物可以生成九个不同的扩增子。这样的方法可以导致具有突变（例如SNP）的靶序列的更多的扩增，这可以有益于与不那么感兴趣的区域相比，更好地覆盖感兴趣的区域。

[0076] 在一些情况下，在预扩增步骤期间，WGA/ WTA引物和靶标特异性引物可以同时地、基本同时地或相继地（例如，WTA/ WGA随后是靶向扩增，反之亦然）分别对全基因组或全转录组和靶序列进行扩增。

[0077] 在一些情况下，WGA/ WTA和靶向扩增可以在同一试管、孔、空腔、腔室、液滴 (drop)、小滴 (droplet)、溶液、反应等中发生。在一些情况下，可以将用于WGA/ WTA和靶向扩增的试剂混合在一起并分配到反应体积中。在其他一些情况下，可以先将用于WGA/ WTA的试剂分配到反应体积中，然后分配用于靶向扩增的试剂，或者反之亦然。换言之，可以将用于靶向扩增的试剂叠放在用于WGA/ WTA的试剂之上。在某些情况下，靶向扩增和WGA/ WTA扩增在同一反应混合物中同时或基本同时进行。在某些情况下，靶向扩增和WGA/ WTA扩增在同一反应混合物中顺序地发生。例如，靶标特异性引物可以在WGA/ WTA引物扩增其靶序列之前对它们的靶序列进行扩增，反之亦然。在另一实例中，靶标特异性引物和WGA/ WTA引物可以基本同时或同时地对其靶标进行扩增。

[0078] 在一些情况下,靶标特异性引物可以基本上与靶序列互补。例如,靶标特异性引物可以与靶序列互补至少约50%,65%,70%,75%,80%,85%,90%,95%,99%或100%互补。在某些情况下,靶标特异性引物可以扩增可能包含突变例如SNP的靶序列。在一些情况下,靶标特异性引物可以扩增包含一个以上突变(例如两个不同的SNP)的靶序列。在一些情况下,靶标特异性引物可以扩增包含多于一个的不同种类的突变(例如,SNP和SNV)的靶序列。

[0079] 在一些实施方案中,WGA/WT A与靶向扩增结合可以产生包含WGA/WT A扩增子和靶向扩增子的扩增子混合物。在一些情况下,所述扩增子混合物可以包含相等或基本相等的WGA/WT A扩增子部分和靶向扩增子部分。在一些情况下,所述扩增子混合物可以包含比靶向扩增子部分更大或基本上更大的WGA/WT A扩增子部分。例如,WGA/WT A扩增子可以占扩增子混合物的90%或更多,而靶向扩增子可以占扩增子混合物的10%或更少。

[0080] 在一些实施方案中,可以在测序仪上直接对扩增子混合物进行测序。在某些情况下,在测序之前,可以使扩增子混合物经历清洁程序、靶向扩增、索引PCR和/或任何其他扩增程序。例如,可以清洗扩增子混合物以去除引物和其他试剂(例如,扩增试剂、裂解试剂等),然后进行嵌套PCR以扩增靶向扩增子,然后再在测序仪上对WGA扩增子和靶向扩增子进行测序。

#### [0081] 清洁步骤

[0082] 可以在细胞裂解,或者一个或更多个扩增步骤后执行清洁步骤。在可选的清洁步骤中,清洁步骤可以有益于去除聚合酶、裂解试剂、扩增试剂、引物、未掺入的dNTP等,它们可能潜在地干扰和/或抑制下游过程,诸如靶向扩增、索引PCR、测序测定等。清洁步骤可以通过使用本领域已知的任何一种方法来进行。例如,可以通过基于柱的、基于凝胶的、基于酶的和/或基于微珠(bead-based)的纯化技术来清洗由WGA结合靶向扩增所生成的扩增子混合物,以除去未掺入的dNTP、扩增试剂等。

#### [0083] 靶向扩增

[0084] 靶向扩增可以在预扩增步骤中与WGA/WT A结合进行并随后进行。在其他情况下,预扩增步骤可以仅包括WGA/WT A,并且靶向扩增可以在预扩增步骤之后。在一些其他情况下,可以不执行预扩增步骤,并且使用靶标特异性引物对核酸分子进行靶向扩增以扩增靶序列。

[0085] 在预扩增步骤中靶向扩增与WGA/WT A结合进行可以生成扩增子混合物。这种扩增子混合物可以使用嵌套在预扩增步骤中使用的靶标特异性引物内部的引物以嵌套PCR进一步扩增。嵌套PCR可以获得靶序列的充分表达以在测序测定中进行测序。例如,以低频率出现的靶序列可以首先在预扩增步骤中使用靶标特异性引物进行扩增,然后再在嵌套PCR中使用嵌套引物以另外的靶向扩增进行扩增。这将确保靶序列的充分表达,其通过足够的覆盖度表示,由测序测定法中的唯一读段(unique reads)的数量确定。嵌套引物可以与WGA/WT A引物或靶标特异性引物共有一个或更多个特征。例如,嵌套引物可以具有与WGA/WT A引物或靶标特异性引物相比基本相似的GC含量。嵌套引物还可以包括如WGA/WT A引物中的接头序列(例如P5或P7),因此可以通过索引引物进一步对生成的嵌套扩增子进行扩增,从而能够在测序平台(例如Illumina)上进行测序。WGA/WT A或靶标特异性引物(例如嵌套引物)中存在的接头序列可以包括不完全的Illumina序列(例如GCTCTTCCGATCT)(SEQ ID NO:21)



或完整的序列(例如AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXACACTCTTCCCTACACGACGCTCTTCCGATCT)(SEQ ID NO:22)其中X=A、C、G或C作为条形码索引(例如样品索引)的一部分,这取决于用户是希望通过索引PCR步骤间接添加测序索引还是在另外的靶向扩增步骤期间直接添加测序索引。接头不必仅特定于Illumina测序平台;用户可以修改接头序列以使任何合适地序列匹配其选择的测序平台。

[0086] 在某些情况下,靶向扩增中使用的嵌套引物的长度(其全长或者其靶标特异性区域)可以为至少约5个碱基对(bp),6bp,7bp,8bp,9bp,10bp,11bp,12bp,13bp,14bp,15bp,16bp,17bp,18bp,19bp,20bp,21bp,22bp,23bp,24bp,25bp,26bp,27bp,28bp,29bp,30bp,31bp,32bp,33bp,34bp,35bp,36bp,37bp,38bp,39bp,40bp,50bp,60bp,70bp,80bp,90bp,100bp或更高。

[0087] 在一些情况下,具有或不具有接头序列的嵌套引物的解链温度可以为至少约40℃,45℃,50℃,60℃,65℃,70℃或更高。在某些情况下,嵌套引物可以具有与靶标特异性引物相同的解链温度。在其他情况下,嵌套引物可以具有与靶标特异性引物不同的解链温度。

[0088] 在某些情况下,嵌套引物的GC含量可以为至少约5%,10%,15%,20%,25%,30%,35%,40%,45%,50%,55%,60%,或大于60%。在某些情况下,嵌套引物可以具有与靶标特异性引物和/或WGA/WT A引物相同的GC含量。在其他情况下,嵌套引物可以具有与靶标特异性引物和/或WGA/WT A引物不同的GC含量。

[0089] 在某些情况下,嵌套PCR中嵌套引物的浓度可以为至少1纳摩尔(nM),10nM,20nM,30nM,40nM,50nM,60nM,70nM,80nM,90nM,100nM,5微摩尔(μM),10μM,15μM,20μM,25μM,30μM,40μM,50μM,100μM,200μM,300μM,400μM,500μM,600μM,700μM,800μM,900μM或更高。

[0090] 在一些情况下,在嵌套PCR中由嵌套引物生成的扩增子可以为至少约50bp,100bp,150bp,200bp,250bp,300bp,350bp,400bp,450bp,500bp,550bp,600bp,650bp,700bp,750bp,800bp,850bp,900bp或更高。

[0091] 在一些情况下,嵌套引物可以具有一个或更多个经修饰的核苷酸,诸如锁核酸(LNA)、蛋白质核酸(PNA)、甲基化核酸等。在某些情况下,所述修饰可以包括具有一个或更多个硫代磷酸酯键、荧光团、生物素、氨基修饰物、硫醇修饰物、炔烃修饰物、叠氮类修饰物、间隔基的核酸。经修饰的核苷酸可以有益于保护该核酸分子免受核酸外切酶或具有核酸外切酶活性的聚合酶的活性的影响。在某些情况下,嵌套引物可以在寡核苷酸的一个或两个末端(例如5'末端,3'末端)具有经修饰的核苷酸。在某些情况下,嵌套引物可以在寡核苷酸的一个末端(例如5'末端或3'末端)具有经修饰的核苷酸。

[0092] 索引PCR

[0093] 在预扩增反应之后或在靶向扩增(例如,嵌套PCR)之后,可以对扩增子混合物或靶向扩增子进行索引PCR测定,以添加在测序仪上执行测序测定所需的附加的核酸序列,诸如index 1、index 2、P5、P7等。例如,可以将包含有针对不同的Illumina测序仪(诸如MiSeq,NextSeq,MiniSeq,HiSeq,iSeq,NovaSeq)对于兼容性和文库聚类所需要的Illumina接头序列的索引引物添加到所述扩增子中,以生成用于进一步测序的核酸文库。包含条形码的索引引物可以用于在单一运行或通道的混合(pooling)后拆分(demultiplex)文库。

[0094] 分析

[0095] 测序后,可以使用自定义管道(custom pipeline)分析数据以检测变体,例如非整

倍性、拷贝数变异等。在某些情况下,管道可以包含一些功能,诸如修剪额外的碱基(例如接头序列),与参考序列(例如hg19)比对,对重复读段进行分类和标记,和/或调用变体。在某些情况下,可以自定义管道以适应不同的索引序列。在某些情况下,基因组的浅且均匀的覆盖(如唯一读段的数量所表明的)可能就足够了(例如,约0.025x)。在某些情况下,可能需要强劲而深入的覆盖(例如,>30x)来检测变体,例如SNP或小的indel等。在某些情况下,例如,通过使用本文所述的预扩增步骤,可以利用浅的覆盖来检测SNP或小的indel。测序读段可能需要基于应用(例如CNV、SNP或这两者的检测)来分配。

#### [0096] 试剂盒

[0097] 本公开的一些方面还包括试剂盒。所述试剂盒可以包括例如用于WGA/WTAs的引物群、至少一种用于靶向扩增的靶标特异性引物等。试剂盒可以包括一套说明书,其用来说明使用该试剂盒来检测CNV,进行SNP、SNV基因分型,微卫星基因分型,检测插入和/或缺失,确定合子型,检测基因融合,检测易位或检测任何其他突变。在某些情况下,试剂盒可以包括一种或更多种选自以下的试剂:蛋白酶如嗜热菌蛋白酶,碱性裂解剂(NaOH),十二烷基硫酸钠(SDS),Triton X-100,毛地黄皂苷,胍,3-[(3-胆酰胺丙基)二甲基铵]-1-丙磺酸内盐,激光脉冲,电脉冲,超声,甘油,1,2丙二醇,甜菜碱一水合物,吐温20,甲酰胺,四甲基氯化铵(AC),7-脱氮-2'-脱氧鸟苷,二甲亚砜(DMSO),Triton X-100,NP-40,镁,牛血清白蛋白(BSA),乙二醇,二硫苏糖醇(DTT),KAPA HiFi和KAPA HiFi Uracil+,VeraSeq Ultra DNA聚合酶,VeraSeq 2.0High Fidelity DNA聚合酶,Takara PrimeSTAR DNA聚合酶,Agilent Pfu Turbo CX聚合酶,Phusion U DNA聚合酶,Deep VentR DNA聚合酶,LongAmp Tag DNA聚合酶,Phusion High-Fidelity DNA聚合酶,Phusion Hot Start High-Fidelity DNA聚合酶,Kapa High-Fidelity DNA聚合酶,Q5 High-Fidelity DNA聚合酶,Platinum Pfx High-Fidelity聚合酶,Pfu High-Fidelity DNA聚合酶,Pfu Ultra High-Fidelity DNA聚合酶,KOD High-Fidelity DNA聚合酶,iProof High-Fidelity聚合酶,High-Fidelity 2DNA聚合酶,Velocity High-Fidelity DNA聚合酶,ProofStart High-Fidelity DNA聚合酶,Tigo High-Fidelity DNA聚合酶,Accuzyme High-Fidelity DNA聚合酶,VentR DNA聚合酶,DyNAzyme II Hot Start DNA聚合酶,Phire Hot Start DNA聚合酶,Phusion Hot Start High-Fidelity DNA聚合酶,Crimson LongAmp Tag DNA聚合酶,DyNAzyme EXT DNA聚合酶,LongAmp Tag DNA聚合酶,Phusion High-Fidelity DNA聚合酶,带有标准Taq(无镁)缓冲液的Tag DNA聚合酶,带有标准Tag缓冲液的Tag DNA聚合酶,带有ThermoPol II(无镁)缓冲液的Tag DNA聚合酶,带有ThermoPol缓冲液的Tag DNA聚合酶,Crimson Taq DNA聚合酶,带(无镁)缓冲液的Crimson Taq DNA聚合酶,Phire Hot Start DNA聚合酶,VentR(exo-)DNA聚合酶,Hemo KlenTaq,Deep VentR(exo-)DNA聚合酶,Deep VentR DNA聚合酶,DyNAzyme EXT DNA聚合酶,Hemo KlenTaq,LongAmp Tag DNA聚合酶,Prot Script AMV First Strand cDNA合成试剂盒,Prot Script M-MuLV First Strand cDNA合成试剂盒,Bst DNA聚合酶,全长,Bst DNA聚合酶,大片段,9Nm DNA聚合酶,DyNAzyme II Hot Start DNA聚合酶,Hemo KlenTaq,Sulfolobus DNA聚合酶IV,Therminator y DNA聚合酶,Therminator DNA聚合酶,Therminator II DNA聚合酶,Therminator III DNA聚合酶,Bsu DNA聚合酶,大片段,DNA聚合酶I(E.coli),DNA聚合酶I,大(Klenow)片段,Klenow片段(3'→5' exo"),phi29 DNA聚合酶,T4 DNA聚合酶,T7 DNA聚合酶(未修饰),末端转移酶,逆转录酶和RNA聚合酶,大肠杆菌

Poly (A) 聚合酶, AMV 逆转录酶, M-MuLV 逆转录酶, phi6 RNA 聚合酶 (RdRP), Poly (U) 聚合酶, 5P6 RNA 聚合酶和 T7 RNA 聚合酶, 镁盐, 三磷酸核苷酸 (dNTP) 及其衍生物, 氯化钠, 氯化钾, 带负电荷的羧基包覆的磁性 (聚苯乙烯) 微珠诸如 AMPure-Beckman Coulter, NucleoMag-MACHEREY-NAGEL, MagJet-ThermoFisher, Mag-Bind-Omega Biotek, ProNex beads-Promega, Kapa Pure Beads-Kapa Biosystems, 硅胶柱诸如 QIAquick PCR 纯化试剂盒和 MinElute PCR 纯化试剂盒-Qiagen, PureLink-Thermo Fisher Scientific, GenElute PCR 清洁试剂盒-Sigma, **NucleoSpin®** 凝胶和 PCR 清洁-MACHEREY-NAGEL, 琼脂糖或丙烯酰胺凝胶, 乙醇或异丙醇沉淀, 酚氯仿萃取, Tris 缓冲液, 吐温 20, SDS, 三磷酸核苷酸 (dNTP), 二甲基亚砜, 二甲基甲酰胺, Tris-HCl pH8.4, 硫酸铵, 硝酸铵, 硝酸钾, TMA-SO<sub>4</sub> (四甲基硫酸铵), TMA-Cl (四甲基氯化铵), 甘油, 测序所需的试剂 (例如 MiSeq 试剂, NextSeq 试剂), 具有或没有修饰的引物寡核苷酸 (例如 LNA, 具有硫代磷酸化碱基), AMPureXP 微珠, 硅胶膜柱, 乙醇, 酚-氯仿萃取, PEG 萃取或琼脂糖凝胶。

#### [0098] 应用

[0099] 本主题方法可以用于检测与遗传障碍相关的各种突变, 诸如 SNP、SNV、CNV、非整倍性、易位、基因融合等。在某些实施方案中, 本主题方法可以用于检测染色体异常和非整倍性例如单亲二体性, 检测子宫灌洗液、子宫内膜液中的体细胞变异体, 以了解植入失败的原因或了解流产的原因, 以及检测临床样品等。在某些实施方案中, 本主题方法可以用于基因组作图 (genomic mapping) 和全基因组关联分析, 例如, 进行基于 SNP 的 CNV 调用, 通过使用 SNP 确定 CNV 分析的准确性, 检测染色体镶嵌, 以及进行连锁分析。本主题方法可以用于携带者筛查, 以筛选怀疑携带潜在突变或已知携带那些突变的个体。该方法可以用于植入前胚胎筛检 (例如, 使用一个或多个胚胎细胞, 使用其中曾培养了胚胎的培养基等), 以检测与遗传障碍相关的突变。该方法可以用于筛检胎儿 DNA 或母体样本 (例如血液, 子宫颈) 中的无细胞 DNA。该方法还可以用于确定在胚胎活检或培养基 (例如在其中生长过胚胎、细胞、组织或类器官的废培养基) 中的污染, 例如母本或父本 DNA 或 RNA 的污染。本主题方法可以用于确定样品中的杂合性或克隆性。例如, 该方法可以用于筛检样品例如肿瘤活检、血液样品、循环肿瘤细胞、无细胞 DNA 或外泌体, 以检测遗传变化, 例如 CNV 和 SNP。这样的筛检可以帮助鉴定肿瘤细胞群内的异质性/克隆性。这可以帮助临床医生确定治疗方案。在某些情况下, 本主题方法可以用于人类识别应用, 法医学应用, DNA 指纹识别, DNA 分析, DNA 分型 (例如在移植或植入监测期间) 或性别确定。在某些情况下, 本主题方法可以用于生物祖先 (bio-ancestry) 或谱系应用, 亲属关系分析, 亲子鉴定, 系统发育分析或进化研究。在某些情况下, 本主题方法可以用于药物遗传学并确定响应于药物治疗的变异性。

[0100] 遗传障碍的实例包括但不限于软骨发育不全, 肾上腺脑白质营养不良,  $\alpha$  地中海贫血,  $\alpha$ -1-抗胰蛋白酶缺乏症, Alport 综合征, 肌萎缩性侧索硬化症,  $\beta$  地中海贫血, 腓骨肌萎缩症, 1a 型先天性糖基化障碍, Crouzon 综合征, 囊性纤维化, 杜兴氏和贝克氏 (Duchenne and Becker) 肌营养不良症, 1 型肌张力障碍, 扭转症 (Torsion), Emery-Dreifuss 肌营养不良症, 面肩肱型肌营养不良 (facioscapulohumeral dystrophy), 家族性腺瘤性息肉病 (familial adenomatous polyposis), 家族性淀粉样变多发性神经病 (familial amyloidotic polyneuropathy), 家族性自主神经异常 (familial dysautonomia), 范可尼贫血 (fanconi anaemia), 脆性 X 综合征 (Fragile X), 戊二酸血症 1 型 (glutaric aciduria

type 1), 血友病A和B, 噬血细胞性淋巴组织细胞增生症(hemophagocytic lymphohistiocytosis), Holt-Oram综合征, 亨廷顿氏病, 高胰岛素性低血糖症(hyperinsulinemic hypoglycemia), 低钾性周期性麻痹(hypokalaemic periodic paralysis), 色素失调症(Incontinentia pigmenti), Lynch综合征, Marfan综合征, Menkes病, 异染性脑白质营养不良(metachromatic leukodystrophy), II型粘多糖贮积症(Hunter综合征), 多发性内分泌腺瘤(multiple endocrine neoplasia, MEN2), 多发性外生性骨疣(multiple exostosis), 肌强直性营养不良, 神经纤维瘤病I型和II型, 非综合征型感觉神经性耳聋(non-syndromic Sensorineural Deafness), Norrie综合征, 成骨不全症(脆骨病), 常染色体显性遗传多囊肾, 常染色体隐性遗传多囊肾, 庞贝氏综合症(Pompe's syndrome), 镰状细胞性贫血, Smith-Lemli-Opitz综合症, 4型痉挛性截瘫(spastic paraplegia 4), 脊髓延髓肌萎缩症, 脊髓性肌萎缩症, 脊髓小脑性共济失调1、2和3型, 脊柱干骺端发育不良(Spondylometaphyseal dysplasia) (Schmidt), Tay-Sachs病, Treacher Collins综合征, 结节性硬化症, Von Hippel-Lindau综合征, X连锁肌张力障碍帕金森综合征(XDP), X连锁无丙种球蛋白血症(X-linked agammaglobulinemia), 白血病, 遗传性椭圆形红细胞增多症和热异形细胞增多症(hereditary elliptocytosis and pyropoikilocytosis), 常染色体隐性高胆固醇血症, 福山型肌营养不良。通过举例说明而非限制的方式提供以下示例。

#### [0101] 实施例

##### [0102] 示例1:检测拷贝数变异(CNV)和单核苷酸多态性(SNP)

[0103] 使用本公开在相同样品中检测CNV和SNP。简而言之, 利用WGA/WT A引物结合靶标特异性引物进行预扩增程序, 随后利用嵌套引物进行嵌套PCR测定来进行靶向扩增, 并使用索引PCR来添加在测序仪上进行测序测定所需的序列, 在具有数量有限的细胞(例如, 单个细胞或五个细胞)或基因组DNA(例如30pg基因组DNA)的样品中检测CNV和SNP。进行了下一代测序(NGS)测定, 以生成序列读段, 并通过自定义的生物信息学管道对其进行分析, 以检测CNV和SNP。该方法允许以约一百万个读段的低测序深度检测不同的突变。

[0104] 使用具有一些修改的**SMARTer® PicoPLEX® Gold Single Cell DNA-Seq**试剂盒(Takara Bio USA, R300669)进行测定。该试剂盒包括以下步骤: 细胞裂解, 全基因组扩增(WGA), DNA纯化和添加Illumina接头用于测序兼容性。修改了试剂盒, 以结合WGA使用靶标特异性引物扩增CFTR基因的某些区域。这样, 在预扩增步骤中添加了靶标特异性引物, 并且在预扩增步骤后添加了嵌套引物。

[0105] 靶标特异性引物被设计用于扩增基因组的包括变体(诸如感兴趣的SNP或indel)的特定区域, 并且被设计为与基因组的其余部分相比对靶序列具有更大的特异性。在设计靶标特异性引物时, 考虑了可能影响引物对靶序列的特异性的其他高频SNP的染色体位置。选定了生成约600个碱基对(bp)的扩增子的靶标特异性引物。设计了多种靶标特异性引物并混合在一起以靶向多个序列, 以增大覆盖所需靶序列的几率。使用诸如ThermoBLAST(dnasoftware)之类的工具来设计靶标特异性引物。总共设计了90个靶标特异性引物来靶向CFTR基因中的15个区域, 这样3个引物对扩增一个靶标区域。引物购自Integrated DNA Technology(美国爱荷华州科拉维尔)。

[0106] 首先, 确定与WGA引物结合使用的靶标特异性引物的数量, 例如使用90个靶标特异

性引物、30个靶标特异性引物或无靶标特异性引物,对CFTR基因的覆盖度产生的影响。90个靶标特异性引物包括每个靶标区域3个引物对,而30个靶标特异性引物包括每个靶标区域一个引物对。使用购自Coriell研究所(美国新泽西州卡姆登市)的30皮克(pg)的gDNA,进行与所述数量的靶标特异性引物和WGA引物的预扩增反应。将靶标特异性引物以每一靶标特异性引物约20nM的浓度包含在**SMARTer® PicoPLEX®** Gold Single Cell DNA-Seq试剂盒中的PreAmp缓冲液和PreAmp酶中。

[0107] 预扩增反应使用以下循环条件进行:

[0108] 热启动:95℃持续3分钟-1个循环

[0109] 靶标特异性扩增:95℃持续15秒,55℃持续90秒,68℃持续90秒-0至6个循环

[0110] WGA:95℃持续15秒,15℃持续50秒,25℃持续40秒,35℃持续30秒,65℃持续40秒,75℃持续40秒-14至18个循环。

[0111] 调整扩增循环的数量以获得足够量的扩增子(例如0.5至5纳克)用于进一步分析。如果需要,可以在与WGA分开的反应中进行靶标特异性扩增。在这种情况下,由于最优的循环条件,靶标特异性引物可以有效地扩增靶序列。然后,例如使用AMPure XP微珠(Beckman Coulter,目录号A63882)清洗已扩增的DNA以去除引物。

[0112] 如图8A中所示,比较了三种不同引物组合-0个靶标特异性引物(0个增效引物(booster primer)),30个靶标特异性引物(15个正向引物和15个反向引物;30个增效引物)以及90个靶标特异性引物(45个正向引物和45个反向引物;90个增效引物)-对CFTR基因的十五个不同靶区域的覆盖度,比较了在所述基因上的覆盖度和覆盖度变化。如图8A所示,靶标特异性引物的数量与在CFTR靶序列上的覆盖度直接相关。X轴显示CFTR靶序列中的15个靶标区域。Y轴显示了在CFTR靶标基因上的测序读段数量或覆盖度,如唯一读段的数量所示。例如,相比于30个或无靶标特异性引物,当使用90个靶标特异性引物时,观察到更大的靶序列覆盖度。接下来,如图8B所示,当与30个或无靶标特异性引物相比时,90个靶标特异性引物减少了在CFTR靶序列上的覆盖度变化。换言之,与使用30个靶标特异性引物时的覆盖度相比,当使用90个靶标特异性引物时观察到更均匀的覆盖度。X轴显示变化系数,而Y轴显示每一反应中引物的数量。当使用90个靶标特异性引物时,覆盖度的变化系数低于0.5,但是当使用30个或无靶标特异性引物时,覆盖度变化的置信度接近1。

[0113] 进一步地,在嵌套PCR测定中使用嵌套引物进行了CFTR区域的靶向扩增。共设计了15个嵌套引物对,每个引物包括2个功能部分,每个末端(即5'末端和3'末端)一个。引物的5'末端部分包括Illumina接头序列。更具体地,正向和反向引物包括P5和P7 Illumina接头的13个共用碱基(common base)。正向引物包括用下划线标出的特定于P5的6个额外碱基(read 1):CACGACGCTCTTCCGATCT(SEQ ID NO:23),而反向引物包括用下划线标出的特定于P7的7个额外碱基(read 2):GACGTGTGCTCTTCCGATCT(SEQ ID NO:24)。嵌套引物的3'末端部分被设计为用于扩增在预扩增步骤中由靶标特异性引物生成的扩增子的片段。在选择和设计嵌套引物的过程中,考虑了引物的特异性。像在预扩增步骤中使用的靶标特异性引物一样,使用ThermoBLAST(dnasoftware)等工具设计嵌套式靶标特异性引物,并且选择了与感兴趣的区域相比对基因组其他区域的亲和力有限的引物。在设计嵌套引物时,也考虑了可能潜在影响引物对靶序列的特异性的其他高频SNP的染色体位置。选定了产生约150个碱基对(bp)的扩增子的嵌套引物。考虑了由嵌套PCR生成的扩增子中的变体(感兴趣的SNP或

indel)的位置,以确保所述变体被包括在测序仪产生的测序读段中。例如,需要 $2 \times 75$ 个碱基对配对的末端读段,因此执行嵌套PCR以使目标SNP或突变被包括在从任一用于生成扩增子的嵌套引物的3'末端起的前75个碱基之内,例如在15-60个碱基之间,或30至40个碱基之间。将多种嵌套引物混合在一起以使靶标的数量多重扩增。将30个嵌套引物以25nM的最终浓度与来自**SMARTer® PicoPLEX®** Gold Single Cell DNA-Seq试剂盒之中的扩增缓冲液(少镁的版本)和扩增酶混合。嵌套PCR测定使用以下循环条件进行:

[0114] 95°C持续3分钟-1个循环

[0115] 95°C持续30秒,56°C持续2分钟,68°C持续30秒-14个循环

[0116] 将嵌套PCR步骤的全部内容物加入来自**SMARTer® PicoPLEX®** Gold Single Cell DNA-Seq试剂盒之中的扩增缓冲液和扩增酶以及索引引物SMARTer DNA HT Dual Index Kit-24N (Takara Bio,目录号R400664)或SMARTer DNA Unique Dual Index Kit-24U A至D系列 (Takara Bio,目录号R400665-R400668)或SMARTer DNA HT Dual Index Kit-96N A至D系列 (Takara Bio,目录号R400660-R400663)。所有索引引物均包含针对不同的Illumina测序仪(例如Miseq,NextSeq,Miniseq,HiSeq,iSeq或NovaSeq)对于兼容性和文库聚类所需的必要Illumina接头序列。索引引物还包含条形码,以能够拆分从多个不同样品生成并在同一测序运行或通道上同时测序的文库。

[0117] 索引PCR使用以下循环条件进行:

[0118] 95°C持续3分钟-1个循环

[0119] 95°C持续30秒,63°C持续30秒,68°C持续60秒-4个循环

[0120] 95°C持续30秒,68°C持续60秒-6至10个循环

[0121] 调整循环数以在索引PCR期间获得足够的产物产量(例如100至500纳克)。根据制造商的说明,使用AMPure XP微珠(Beckman Coulter,目录号A63882)清洗所扩增的文库,以除去扩增试剂、引物和DNA聚合酶等。将文库进一步在MiSeq或NextSeq上进行 $2 \times 75$ 个循环。

[0122] 测序后,使用自定义管道分析数据。首先,将fastq文件下采样(down-sampled)为总共100万个读段。修剪接头序列和读段的前14个碱基,并使用Trimmomatic (Bolger AM, Lohse M和Usadel B.,Trimmomatic:a flexible trimmer for Illumina sequence data (Trimmomatic:Illumina序列数据的灵活修剪器),Bioinformatics.,2014年8月1日;30(15):2114-2120)来过滤低质量读段。然后使用Bowtie2 (Langmead B,Salzberg S.,Fast gapped-read alignment with Bowtie 2 (利用Bowtie 2进行的快速空位读段比对),Nature Methods.,2012,9:357-359)进行与人类基因组组装序列GRCh37的比对 (Church DM等,Modernizing reference genome assemblies (现代化的参考基因组组装),PLoS Biol.,2011年7月;9(7):e1001091)。使用Vardict (Lai Z,Markovets A,Ahdesmaki M,Chapman B,Hofmann O,McEwen R,Johnson J,Dougherty B,Barrett JC和Dry JR,VarDict:a novel and versatile variant caller for next-generation sequencing in cancer research (VarDict:用于癌症研究中的下一代测序的新型多功能变体调用器),Nucleic Acids Res.,2016,pii:gkw227)执行变体调用。

[0123] 可以以浅但均匀的覆盖来检测CNV,而诸如SNP、SNV或小的indel等的变体可能需要较深的覆盖。因此,为了检测CNV以及SNP、SNV,优化了对于整个基因组的覆盖和CFTR基因中靶标区域的覆盖所分配的测序读段数。为此,比较了以下两个条件下的基因组和CFTR基

因的覆盖度:具有WGA和靶向扩增的预扩增,以及具有WGA而没有靶向扩增的预扩增。在这两种情况下,均在预扩增步骤之后进行靶向扩增。该测定使用了三十皮克的基因组DNA。

[0124] 如图9A所示,与没有靶向扩增的预扩增步骤相比,当预扩增步骤包括靶向扩增时,获得了更大百分比的CFTR基因读段。X轴显示在预扩增步骤中有和没有靶向扩增CFTR基因的两种测定的结果。Y轴显示读段百分比。与预扩增步骤不包括靶向扩增情况下的读段百分比(4.3%)相比,预扩增步骤包括靶向扩增的情况获得了更高百分比的读段(12%)。相反,当在预扩增步骤中不包括靶向扩增时观察到的WGA的读段百分比(95.7%)高于在预扩增步骤中包括靶向扩增时的测定中的读段(88%)。如图9B所示,在预扩增步骤中具有靶向扩增的测定下观察到的CFTR基因的覆盖度(8633x)(如唯一读段数所示的)大于在在预扩增步骤中没有靶向扩增的测定下获得的覆盖度(3184x)。X轴显示了两种不同的测定方法-在预扩增步骤中有和没有CFTR基因的靶向扩增。Y轴显示了平均覆盖度。此外,在预扩增步骤中包括靶向扩增的测定中,CFTR基因的十五个不同区域的覆盖均一性(图9C)与在预扩增步骤中不包括靶向扩增的覆盖(图9D)相比得到了改善。X轴显示CFTR基因中的15个不同的靶标区域,Y轴显示了每个靶标区域的覆盖度或唯一读段数。

[0125] 对CFTR基因的15个扩增子的覆盖均一性进行了评估:使用单个细胞( $n=4$ )和五个( $n=4$ )分选的细胞,并在预扩增步骤中使用90个靶标特异性引物,然后进行靶向扩增和索引PCR,以生成用于测序的文库分子。如图10A和图10B所示,与单个细胞的样品相比,在五细胞样品中观察到在CFTR基因的十五个靶标区域上的更均匀的覆盖。但是,对于进一步的分析,两种样品类型的覆盖均一性都完全可以接受。

[0126] 接下来,对测序读段进行分析以检测CFTR基因中的变体。如表1所示,对于GM07552和GM012785,使用单个细胞或五个分选的细胞对不同的变体及其等位基因频率进行了检测。GM07552细胞包含已知的变体-Phe508DEL、Arg553TER,并且在CFTR基因中具有等位基因7T/9T。GM12785细胞在CFTR基因中包含已知的变体ARG347PRO、GLY551ASP、7T/7T。对于使用GM07552或GM12785针对五个分选的细胞进行的实验,正确鉴定了所有杂合变体等位基因频率在0.2与0.8之间。当探查被分组(panel)覆盖的所有碱基(2,250个碱基)时,没有报告等位基因频率高于0.1的其他变体。假阳性率实际上为0%。类似地,使用单个细胞鉴定了杂合变体。

[0127] 表1

GM07552					GM12785				
来自五个细胞					来自五个细胞				
		delF508	Arg553ter	7T/9T			Arg347Pro	Arg551Asp	7T/7T
重复	01	0.52	0.58	0.43/0.57	重复	01	0.46	0.52	1
重复	02	0.43	0.58	0.30/0.70	重复	02	0.54	0.58	1
重复	03	0.49	0.63	0.29/0.71	重复	03	0.68	0.51	1
重复	04	0.44	0.63	0.30/0.70	重复	04	0.64	0.58	1
来自单一细胞					来自单一细胞				
		delF508	Arg553ter	7T/9T			Arg347Pro	Arg551Asp	7T/7T
重复	01	0.64	0.64	0.43/0.57	重复	01	0.19	0.49	1
重复	02	0.21	0.98	0.31/0.69	重复	02	0.97	0.30	1
重复	03	0.53	0.27	0.45/0.55	重复	03	0.97	0.59	1
重复	04	0.51	低覆盖率	0.30/0.70	重复	04	0.74	0.76	1

[0129] 分别如图11A和图11B所示,使用GM12785(五个细胞或单个细胞)确定了1Mb bin的测序读段的分布。如图11A和图11B所示,每bin的读段数显示了在各个bin上的五细胞样品

和单细胞样品之间的相似模式,证明了测定的灵敏度和可重复性。

[0130] 通过计算Pearson和Spearman相关性评估了GM12785的五个分选的细胞(N=4)的重复(如图12A所示)和单个细胞(N=4)的重复(如图12B所示)之间的读段分布的可重复性。观察到五个细胞和单个细胞的重复都有强的相关性,证明了即使在存在靶向引物的情况下的全基因组扩增的稳健性。如图12C所示,计算了在五个GM12785分选细胞或单个细胞的两种重复之间的每个bin的log2比率,并使用IGV(Broad Institute)作图。如图所示,在GM05067和GM22601中,在两种重复之间这些bin是保守的,并且在Chr.9和染色体4中分别观察到了预期的拷贝数变异。

[0131] 总而言之,在**SMARTer® PicoPLEX® Gold Single Cell DNA-Seq**中添加靶标特异性引物和嵌套引物实现了基因组的稳健而均匀的覆盖,以及单一试管工作流程中的单个或五个细胞中对CFTR基因的15个关键区域的深度覆盖。当使用总共100万个读段时,该测定表现良好。当使用五个分选的细胞时,对五个不同特征的杂合突变的检测差不多为100%。在2,250个碱基组中未检测到假阳性。

[0132] 示例2:用于SNP检测的靶向扩增

[0133] 在此示例中,我们证明了靶标特异性引物在例如携带者筛查中用于检测SNP的用途。分别从GM07552或GM12785细胞中提取了15ng的基因组DNA(NA07552或NA012785)。GM07552细胞包含以下CFTR的已知变体:Phe508DEL,Arg553TER,并具有等位基因7T/9T。GM12785包含以下CFTR基因中的已知变体:Arg347Pro,Gly551Asp,并具有等位基因7T/7T。使用15对靶标特异性引物对所提取的基因组DNA(NA07552或NA012785)进行靶向扩增,以扩增CFTR基因中的15个不同变体。将靶标特异性引物以25nM的最终浓度与来自**SMARTer® PicoPLEX® Gold Single Cell DNA-Seq**试剂盒中的扩增缓冲液(少镁的版本)和扩增酶混合。靶向扩增PCR如下进行:

[0134] 95℃持续3分钟-1个循环

[0135] 95℃持续30秒,56℃持续2分钟,68℃持续30秒-14个循环

[0136] 将所述靶向扩增的内容物加入来自**SMARTer® PicoPLEX® Gold Single Cell DNA-Seq**试剂盒中的扩增缓冲液和扩增酶以及索引引物SMARTer DNA HT Dual Index Kit-24N(Takara Bio,目录号R400664)或SMARTer DNA Unique Dual Index Kit-24U A至D系列(Takara Bio,目录号R400665-R400668)或SMARTer DNA HT Dual Index Kit-96N A至D系列(Takara Bio,目录号R400660-R400663)。所有索引引物均包括针对不同的Illumina测序仪(如Miseq,NextSeq,Miniseq,HiSeq,iSeq,NovaSeq)对于兼容性和文库聚类所需的Illumina接头序列。索引引物还包含条形码,其用于在单次运行的混合之后拆分文库。

[0137] 索引PCR如下进行:

[0138] 95℃持续3分钟-1个循环

[0139] 95℃持续30秒,63℃持续30秒,68℃持续60秒-4个循环

[0140] 95℃持续30秒,68℃持续60秒-6个循环

[0141] 使用AMPure XP微珠(Beckman Coulter,目录号A63882)清洗所扩增的文库,以除去扩增试剂、引物和DNA聚合酶等。文库在MiSeq上进一步处理2×75个循环。

[0142] 测序后,使用自定义的生物信息学管道分析数据。首先,将fastq文件下采样为总共100万个读段。修剪了接头序列和读段的前14个碱基,并使用Trimmomatic(Bolger AM,



Lohse M和Usadel B., Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics., 2014年8月1日; 30 (15): 2114-2120) 过滤低质量读段。随后使用Bowtie2 (Langmead B, Salzberg S., Fast gapped-read alignment with Bowtie 2, Nature Methods., 2012, 9: 357-359) 进行与人类基因组组装序列GRCh37的比对 (Church DM等, Modernizing reference genome assemblies, PLoS Biol., 2011年7月; 9 (7): e1001091)。使用Vardict (Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, Johnson J, Dougherty B, Barrett JC和Dry JR, VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research, Nucleic Acids Res., 2016, pii: gkw227) 执行变体调用。

[0143] 如图13所示, 观察到了单独使用本公开内容中描述的靶标特异性引物的靶向扩增在CFTR基因的十五个靶标区域上的均匀的覆盖。X轴显示了CFTR基因中的十五个靶标区域或变体。Y轴显示了每个靶标区域的覆盖度, 如唯一读段数所表明的。此外, 如表2所示, 仅使用靶向扩增, 我们就能够正确鉴定所有5个杂合变体等位基因频率在0.4至0.6之间。当探查被分组覆盖的所有碱基 (2,250个碱基) 时, 未报告等位基因频率高于0.05的其他变体。假阳性率实际上为0%。

[0144] 表2

GM07552					GM12785				
来自 15ng gDNA					来自 15ng gDNA				
		delF508	Arg553ter	7T/9T			Arg347Pro	Arg551Asp	7T/7T
重复	01	0.45	0.51	0.43/0.57	重复	01	0.49	0.48	1
重复	02	0.52	0.50	0.42/0.58	重复	02	0.48	0.48	1
重复	03	0.50	0.50	0.40/0.60	重复	03	0.49	0.49	1

[0146] 基于该实验, 我们得出结论, 靶向扩增可用于检测SNP, 特别是在不需要WGA或有大量的输入DNA可用的情况下。一个这样的实例可以包括在针对父母的携带者筛查中的SNP检测。

[0147] 示例3: 临床样品中的CFTR突变的检测

[0148] 这项研究是使用滋养外胚层活检样品完成的, 该样品是从之前已经历传统SNP和CNV分析的胚胎中收集的, 所述传统SNP和CNV分析采用两步法, 其中第一次活检用于SNP测定, 然后第二次活检用于测定拷贝数。这在图14A中进行了示意性地概述。4个胚胎来自确定为病理性CFTR变体SNP F1052V携带者的母亲, 以及确定为R117H变体携带者的父亲。如图14A所示, 第一次活检显示胚胎3和胚胎4为复合杂合子, 携带来自母亲和父亲的两种病理性变体。因此, 没有使用第二次活检进一步筛查这两个胚胎的潜在拷贝数变异 (CNV)。将胚胎1和胚胎2进行第二次活检, 并在胚胎1和胚胎2中鉴定出潜在的CNV非整倍性。

[0149] 在该示例中, 使用本公开中描述的方法, 从同样这4个胚胎中取下了第三次活检切片, 并用于说明当前公开的方法如何能够从单次活检测试中鉴别出SNP和CNV异常。这在图14B中示意性地示出。取下了每个含有大约5个细胞的滋养外胚层活检样品 (n=4), 并使其经历使用WGA结合靶向扩增的本公开的方法。对于相同的活检样品重复两次实验, 并且结果如图14C所示。如上图所示 (图14C), 评估了4个胚胎中每个胚胎的母本和父本的病理性CFTR变体的合子型, 并发现两次均与传统方法一致, 表明胚胎3和胚胎4是复合杂合子, 胚胎1是母本CFTR变异体的携带者, 而胚胎2是野生型。在下图 (图14D) 中还显示了来自所进行的一对测定中之一的4个胚胎的CNV分析。这表明胚胎1、3和4具有正常的核型, 而胚胎2则显示了

染色体19q的部分丢失,这证实了通过传统的两步法获得的结果。发现胚1是野生型。这可能反映了胚胎中的镶嵌现象。总之,我们证明了WGA/靶向测序联合的方法可以从单次胚胎活检中确定SNP和CNV的变化;因此,相比于分别评估SNP和CNV的传统两步法,提高了实用性。

[0150] 示例4:SVA元件中的变体的检测

[0151] 使用人类基因组DNA样本并使用包括用于扩增冗余基因组元件的单个靶标特异性引物对与用于全基因组扩增的引物对的组合的预扩增程序检测了SNP和CNV。选择SVA元件作为候选的冗余基因组元件,是由于发现它们以一定密度存在于所有常染色体和性染色体上使得可以针对所有染色体进行基于SNP的分析(表3)。

[0152]

染色体	SVA元件的数量	SVA元件的密度 (bp)
1	13,944	16,529
2	8,639	27,844
3	7,222	27,430
4	4,687	40,485
5	6,148	29,484
6	5,854	29,053
7	9,182	17,313
8	4,614	31,376
9	5,861	20,780
10	6,207	21,470
11	5,972	22,527
12	8,100	16,437
13	2,443	40,108
14	4,348	20,830
15	4,630	18,281
16	8,012	10,210
17	10,004	8,289
18	2,243	35,706
19	12,884	4,536
20	4,295	14,888
21	1,320	30,370
22	4,336	9,031
X	5,094	30,407
Y	1,058	24,967
全基因组	147,097	19,971

[0153] 表3.SVA元件的数量及其在全基因组中的平均出现率

[0154] SVA元件的数量及其位置是从使用hg38人类基因组组装序列的Dfam的重复DNA家族数据库中获取的。SVA元件的密度是基于使用hg38人类基因组组装序列的每个染色体的可作图部分。

[0155] 为了确定这些SVA元件中包含的SNP的数量,使用国家生物技术信息中心最新发布

的人类SNP数据库作为参考来确定在SVA元件内发现的SNP总数和信息SNP (次等位基因频率  $\geq 0.05$ ) 数 (表4)。据估计,在SVA元件中发现的146,856个信息SNP平均每67,109bp出现一次。信息SNP出现在所有染色体的SVA元件内 (图15)。

[0156]

染色体	SNP的数量		SNP 的密度(SNP/bp)	
	全体	信息 (AF > 0.05)	全体	信息 (AF > 0.05)
1	46,111	13,177	4,998	17,491
2	28,315	8,110	8,495	29,661
3	23,688	6,934	8,363	28,569
4	17,909	5,457	10,595	34,772
5	21,266	6,134	8,524	29,551
6	21,908	6,959	7,763	24,440
7	30,237	8,960	5,257	17,742
8	16,611	4,732	8,715	30,593
9	20,015	5,870	6,085	20,748
10	21,085	6,351	6,320	20,983
11	20,110	5,844	6,690	23,021
12	27,235	8,259	4,888	16,120
13	9,113	2,915	10,752	33,613
14	15,321	4,483	5,911	20,203
15	14,988	4,293	5,647	19,716
16	25,330	7,178	3,230	11,397
17	32,408	9,506	2,559	8,723
18	8,381	2,507	9,556	31,946
19	46,147	14,685	1,266	3,980
20	13,240	3,961	4,830	16,143
21	4,219	1,399	9,502	28,655
22	14,010	4,436	2,795	8,828
X	15,999	4,682	9,681	33,083
Y	174	24	151,811	1,100,627
总计	493,820	146,856		
每个染色体的平均值	20,576	6,119	12,676	67,109

[0157]

[0158] 表4.SVA元件内SNP的数量及其在基因组中的平均出现率

[0159] SVA元件包含七个不同的区域 (图5)。设计靶标特异性引物对以扩增SVA元件的诸如Alu样区域或SINE-R区域之类的区域。对五十种候选的靶标特异性引物扩增靶标SVA元件的能力进行筛选。靶标特异性引物使用诸如BiSearch Primer Design和Search Tool (Aranyi等, (2006)) 的工具来设计 (图16)。对于每个区域,即Alu样或SINE-R区域,测试了图16中公开的25种不同的正向引物与反向引物的引物对组合。共测试了50种引物对组合,并选择了成功扩增了其靶标区域并产生了接近其预期大小的扩增子产物的靶标特异性引物,

以掺入到如本公开所提供的WGA/WTa方法的预扩增步骤中。在五十个引物对中,总共选择了37个SVA特异性引物对。

[0160] 总而言之,将SVA特异性引物对掺入全基因组扩增过程的预扩增步骤中,来扩增在人类全基因组中具有一定的密度和分布的SVA元件中包含SNP的区域,以进行本专利申请的方法部分详细介绍的基于SNP的分析。

[0161] 尽管出于清楚理解的目的已经通过图示和示例的方式详细地描述了前述发明,但是根据本发明的教导,对于本领域普通技术人员而言显而易见的是,可以在不脱离所附权利要求书的精神或范围的情况下对其进行某些改变和修改。

[0162] 因此,前面仅说明了本发明的原理。应当理解,本领域技术人员将能够设计出各种布置,尽管在本文中并没有明确描述或示出,但是它们体现了本发明的原理并且被包括在本发明的精神和范围之内。此外,本文陈述的所有示例和条件语言主要旨在帮助读者理解本发明的原理和发明人为进一步发展本领域所做出的构思,并且应解释为不限于这些具体陈述的示例和条件。此外,本文中陈述本发明的原理、方面和实施方案及其特定示例的所有陈述旨在涵盖其结构和功能上的等同物。另外,这些等同物旨在包括当前已知的等同物和将来开发的等同物,即,不管结构如何,被开发执行相同功能的任何元件。而且,无论在权利要求书中是否明确叙述了这种公开,本文所公开的任何内容都不旨在献给公众。

[0163] 因此,本发明的范围不旨在限于本文示出和描述的示例性实施方案。相反,本发明的范围和精神由所附权利要求体现。在权利要求中,35U.S.C. §112 (f) 或35U.S.C. §112 (6) 被明确定义为仅在这种权利要求的限制的开头叙述了确切的短语“用于……的方式”或确切的短语“用于……的步骤”时被援引用于该权利要求中的限制;如果在权利要求的限制中没有使用这种确切的短语,则不援引以35U.S.C. §112 (f) 或35U.S.C. §112 (6) 不被调用。

## 序列表

<110> 宝生物工程(美国)有限公司

伊曼纽尔·坎伯罗夫

木村刚隆

朱莉·凯瑟琳·拉贝特

<120> 核酸文库的制备和分析方法

<130> CLON-178W0

<150> US 62/806698

<151> 2019-02-15

<160> 25

<170> PatentIn version 3.5

<210> 1

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 1

tccacggtct ccctctcatg c 21

<210> 2

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 2

cggctcactg caacctccct g 21

<210> 3

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 3

gatacctcctg cctcggcctc c 21

<210> 4

<211> 21

<212> DNA

<213> 人工序列  
<220>  
<223> 合成序列  
<400> 4  
cgagtgcctg ggattgcagg c 21  
<210> 5  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 5  
gtggagacgg ggtctcgccg t 21  
<210> 6  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 6  
agaggctgca atctcggcac t 21  
<210> 7  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 7  
acgccactgc actccagcct g 21  
<210> 8  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 8  
ggcaccattg agcactgagt g 21  
<210> 9  
<211> 21

<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 9  
cgagactccg tctgcaatcc c 21  
<210> 10  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 10  
ccggcacctc gggaggccga g 21  
<210> 11  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 11  
ggagcccctc tgcccggcca c 21  
<210> 12  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 12  
gggccatgat gacgatggcg g 21  
<210> 13  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 13  
agagatcaga ttgttactgt g 21  
<210> 14

<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 14  
gtgcaagatg tgctttgtta a 21  
<210> 15  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 15  
tgaaggcagc atgctcgta a 21  
<210> 16  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 16  
gataattcctt gggtgtttct c 21  
<210> 17  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 17  
agagggggat ttggcagggt c 21  
<210> 18  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 18  
aatagtggag ggaaggtcag c 21



<210> 19  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 19  
agggaagggtc agcagataaa c 21  
<210> 20  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 20  
gtctctgggtt ttcctaggca g 21  
<210> 21  
<211> 13  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 21  
gctcttccga tct 13  
<210> 22  
<211> 70  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<220>  
<221> misc\_feature  
<222> (30) .. (37)  
<223> n is a, c, g, or t  
<400> 22  
aatgatacgg cgaccaccga gatctacacn nnnnnnnaca ctcttttcct acacgacgct 60  
cttccgatct 70  
<210> 23  
<211> 19

<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 23  
cacgacgctc ttccgatct 19  
<210> 24  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<400> 24  
gacgtgtgct cttccgatct 20  
<210> 25  
<211> 6  
<212> DNA  
<213> 人工序列  
<220>  
<223> 合成序列  
<220>  
<221> misc\_feature  
<222> (1) .. (6)  
<223> 位置1至位置6中的核苷酸可以重复n次。  
<400> 25  
ccctct 6

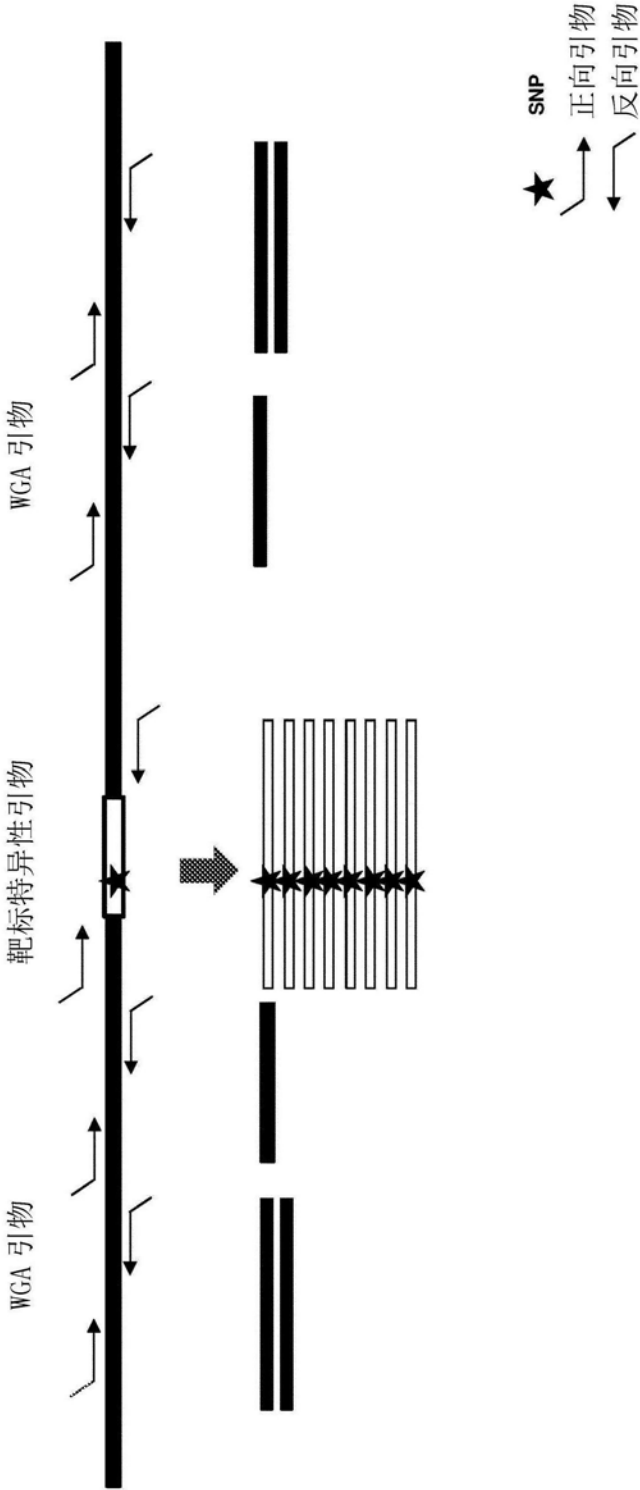


图1

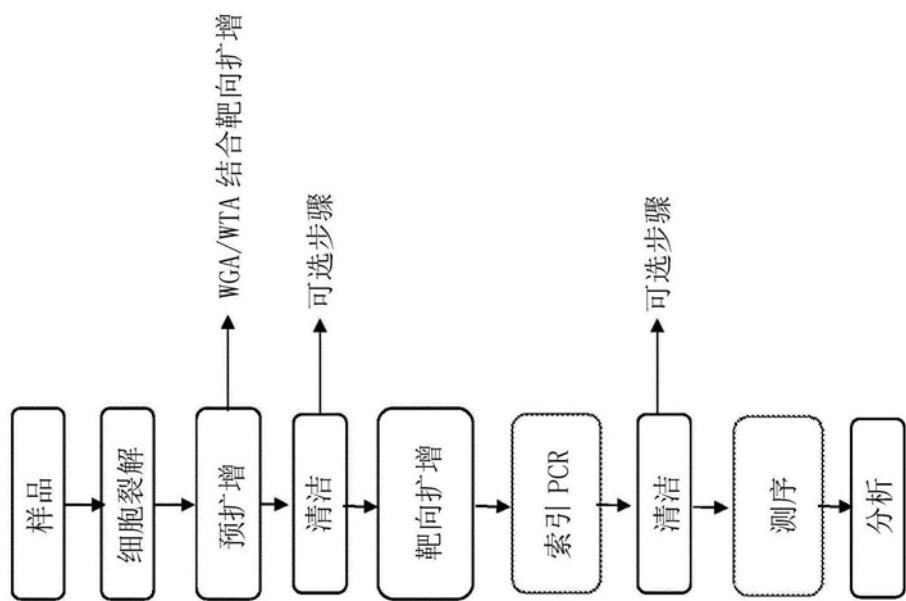


图2

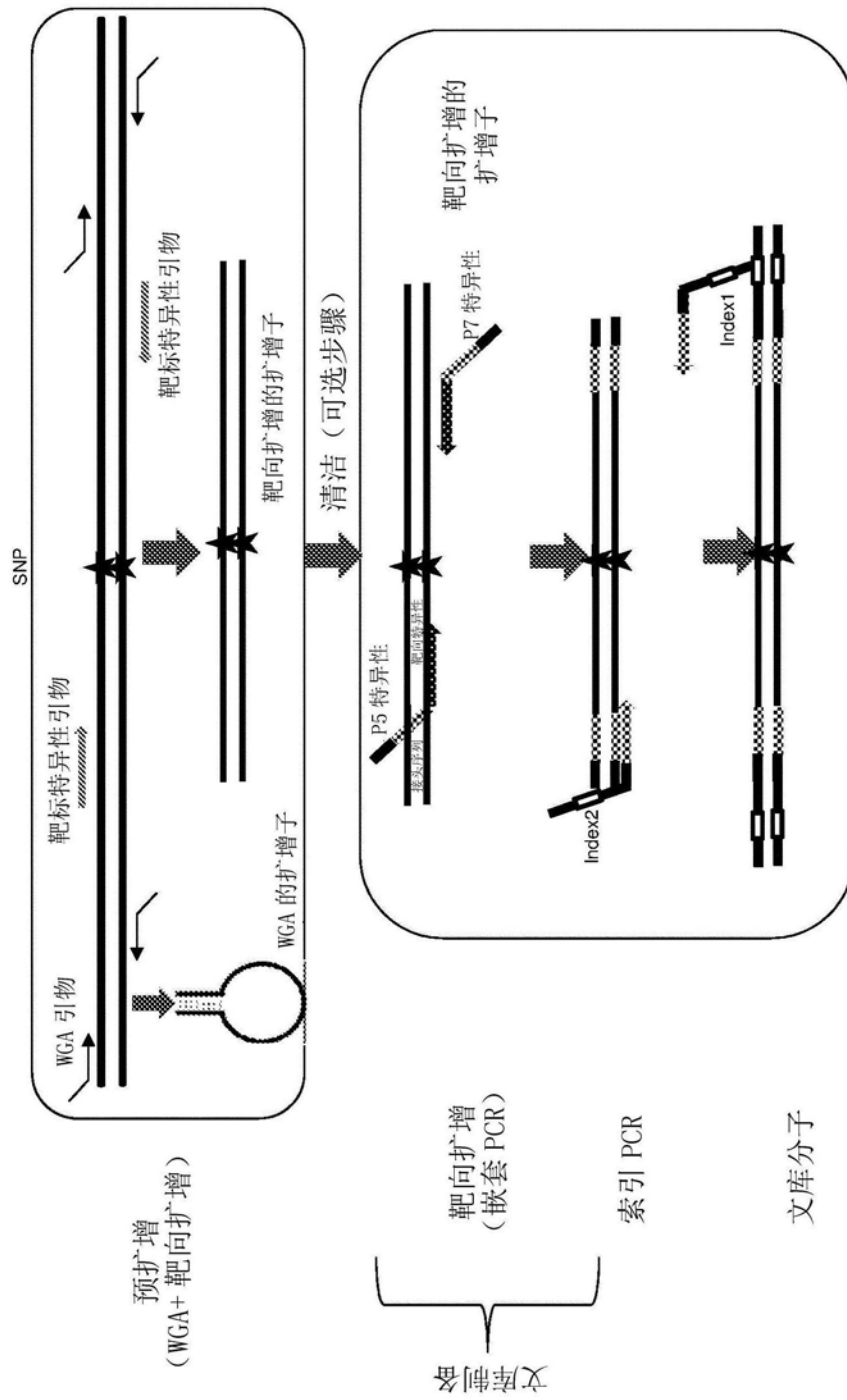


图3

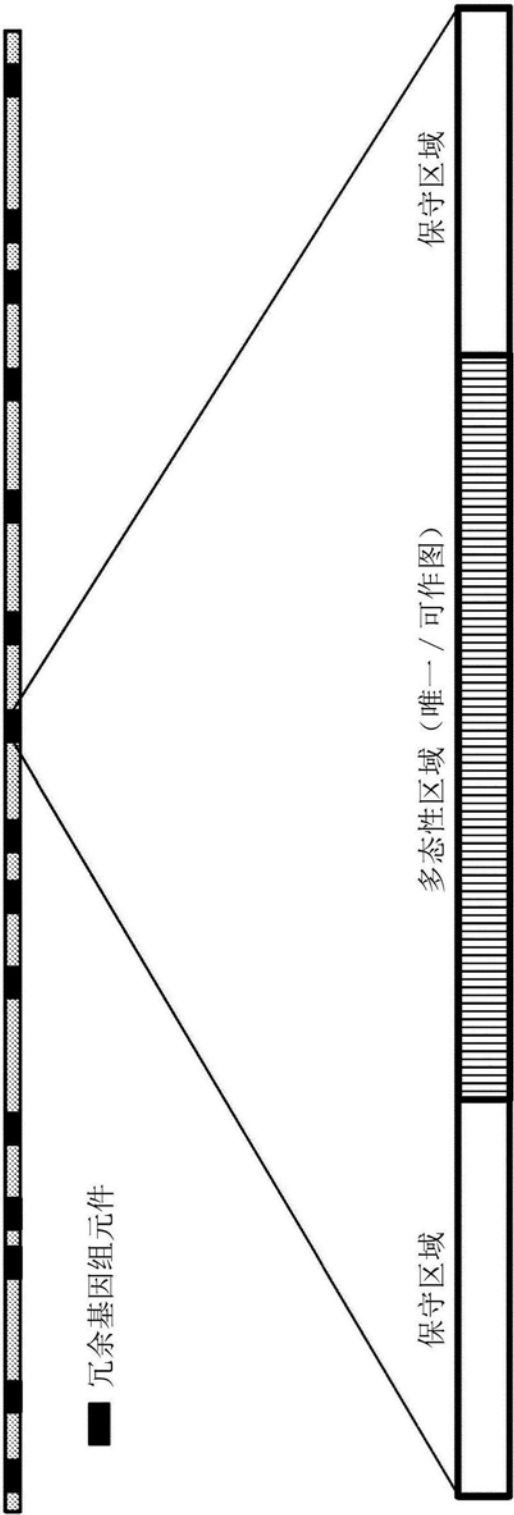


图4

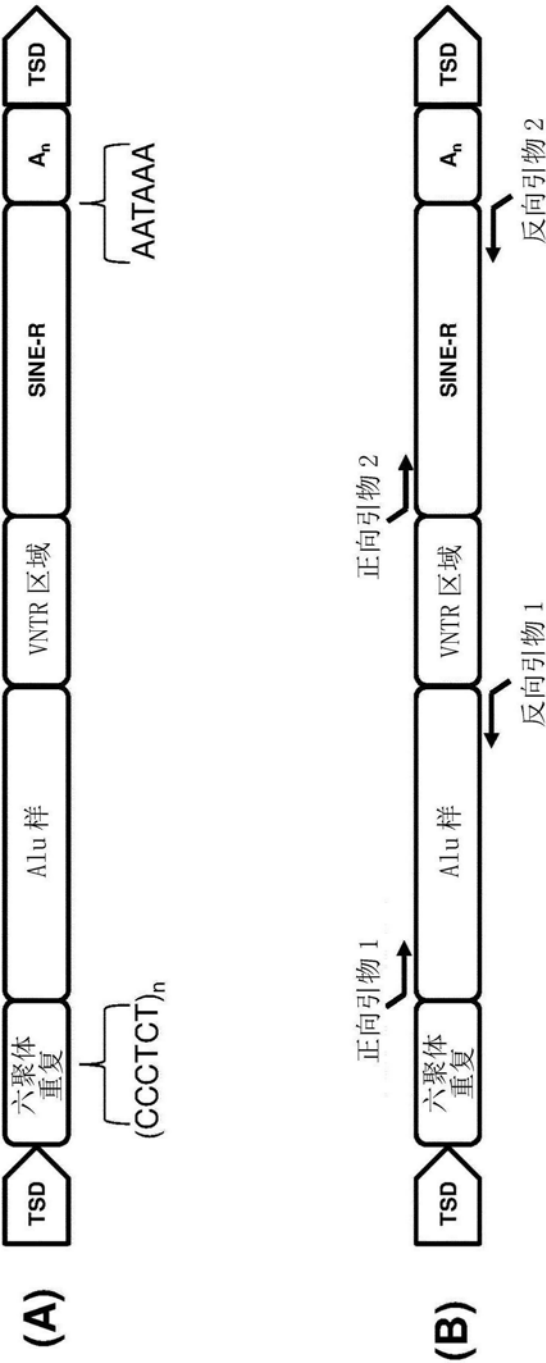


图5

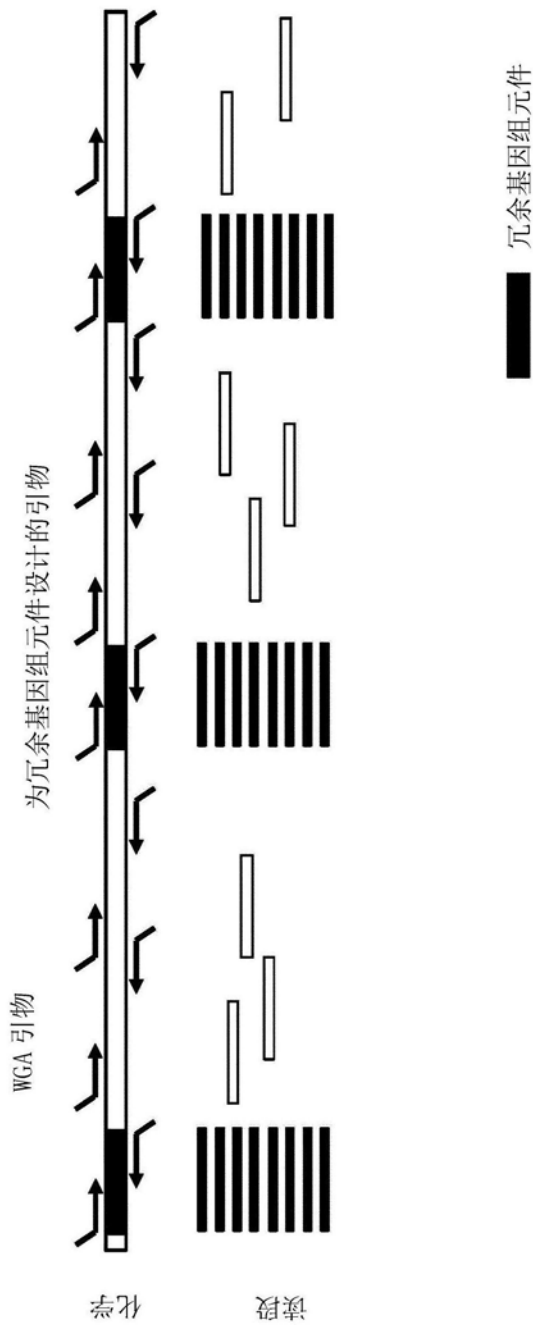


图6



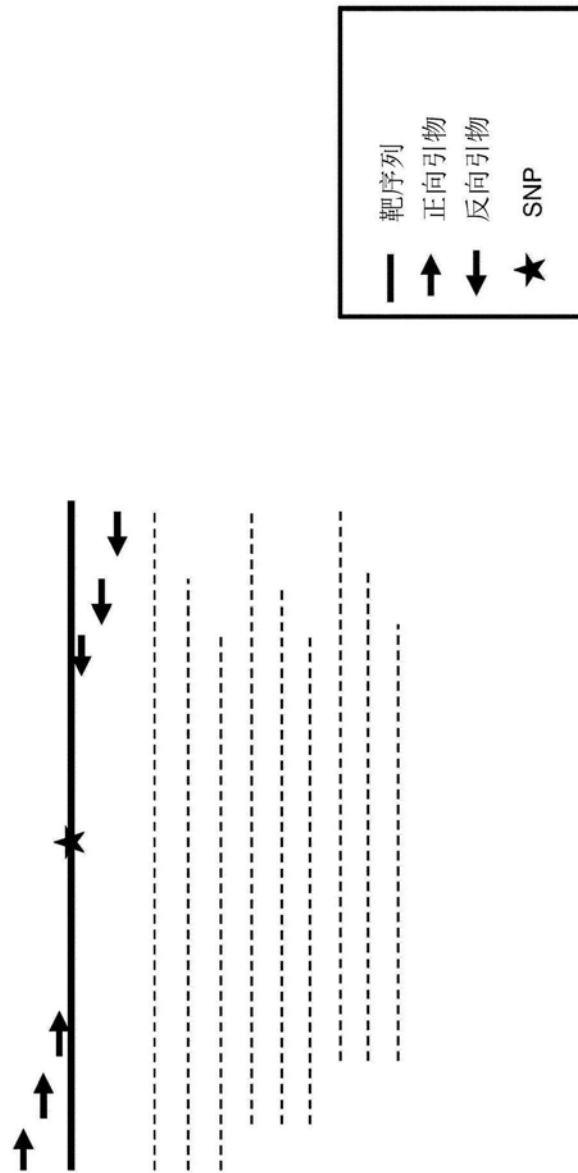


图7

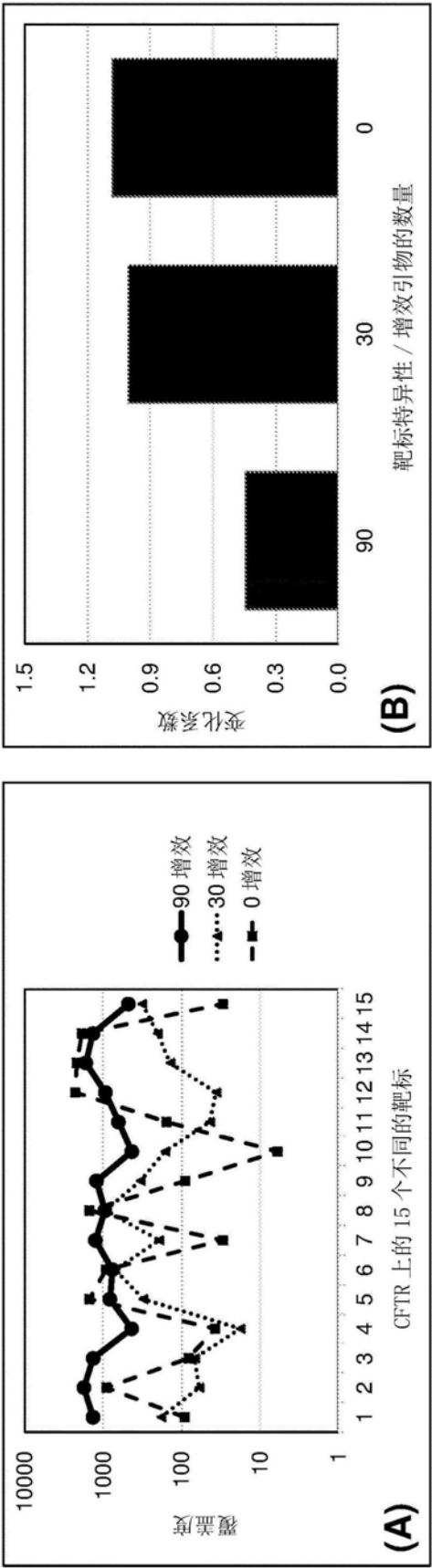


图8

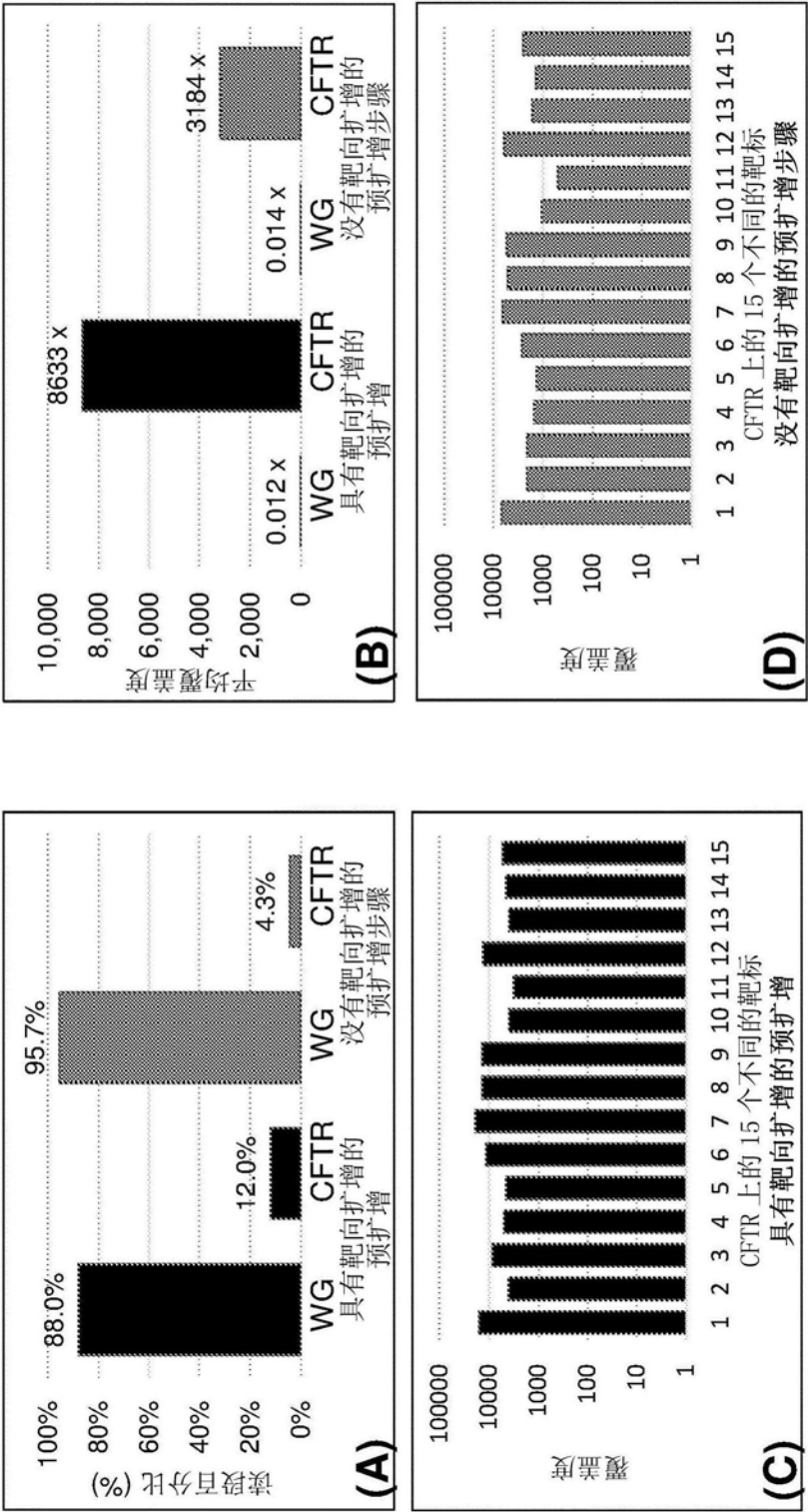


图9

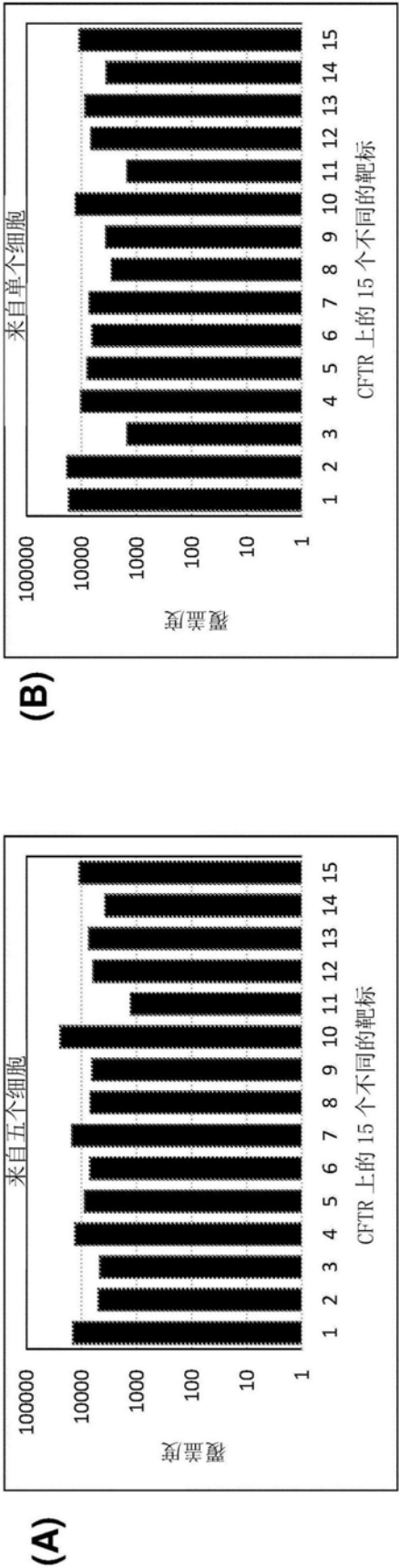
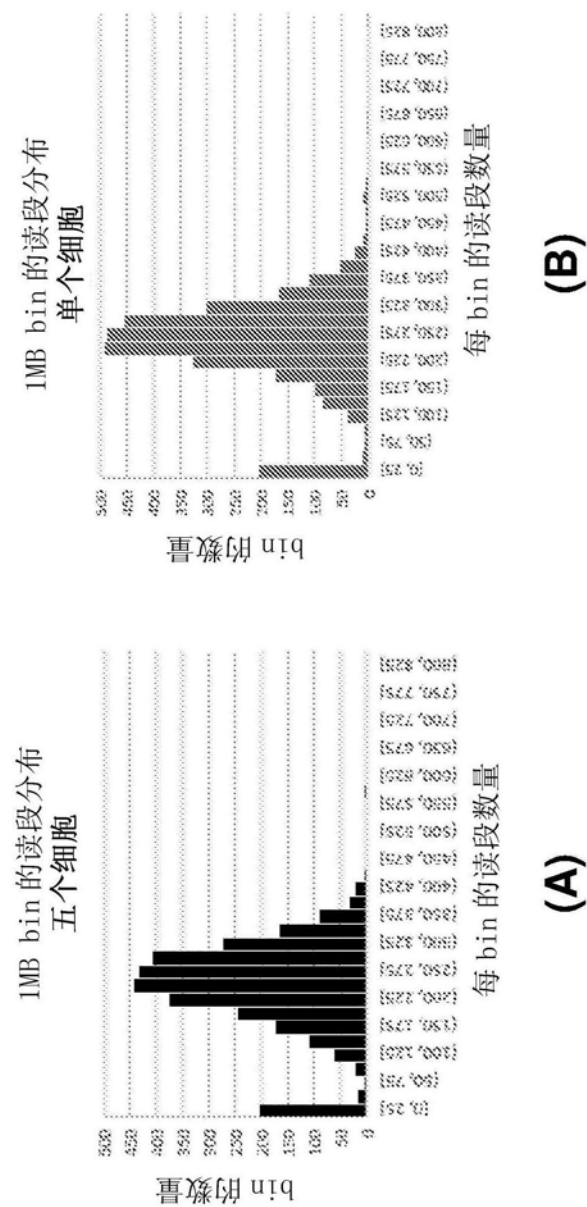


图10



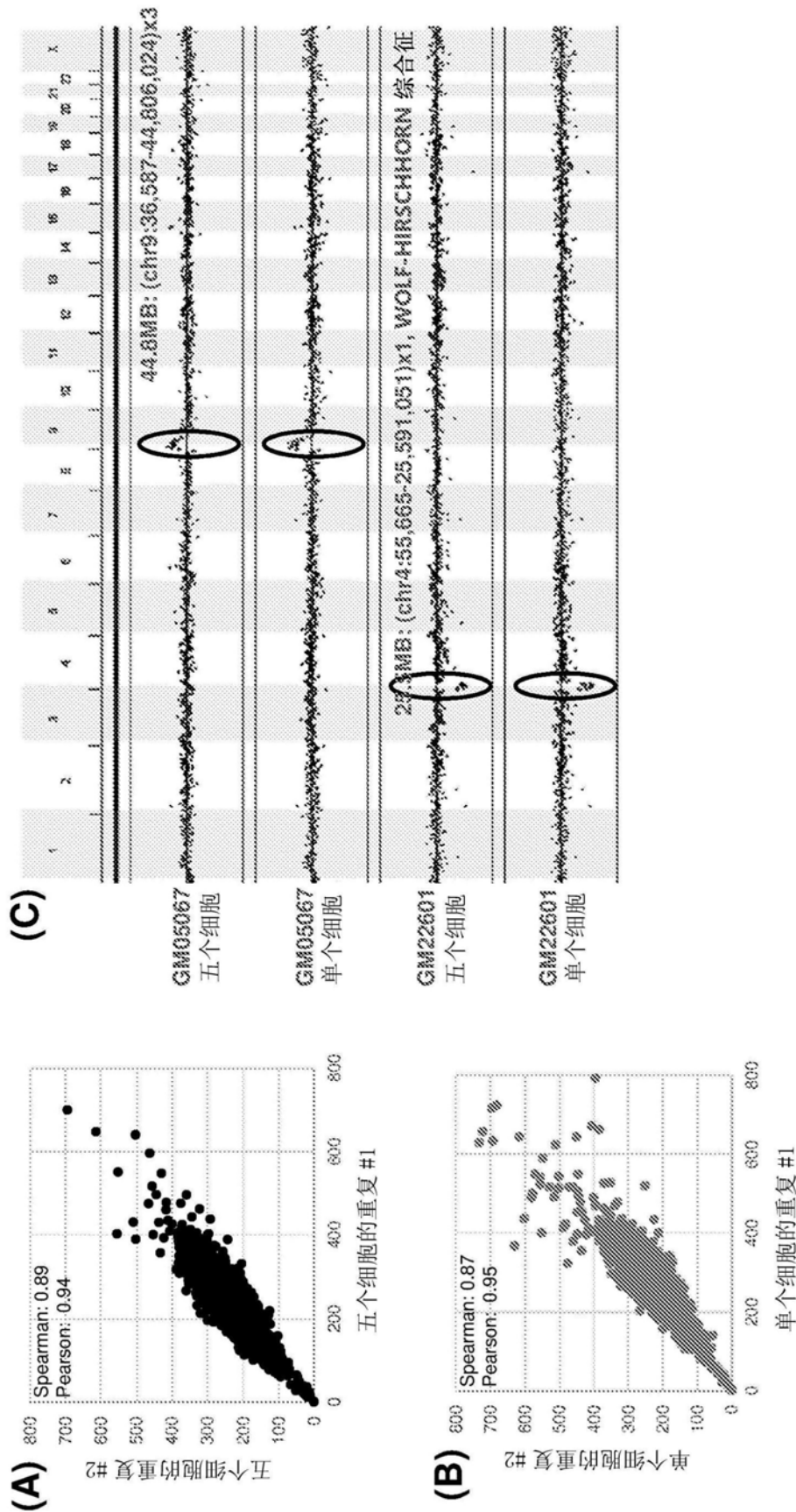


图12

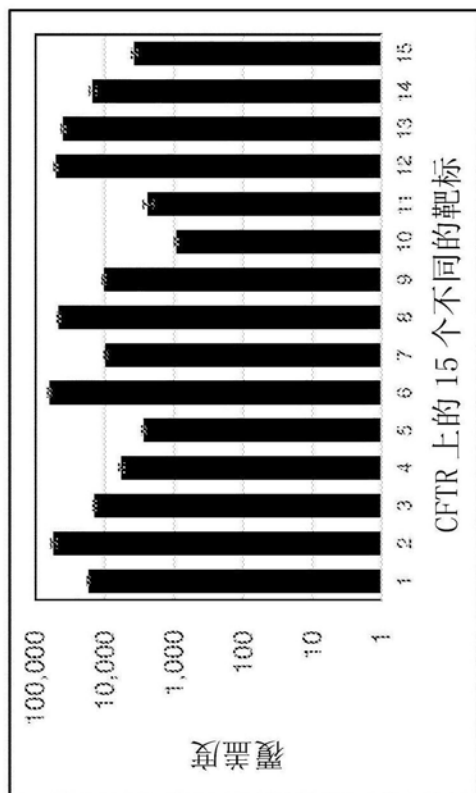


图13

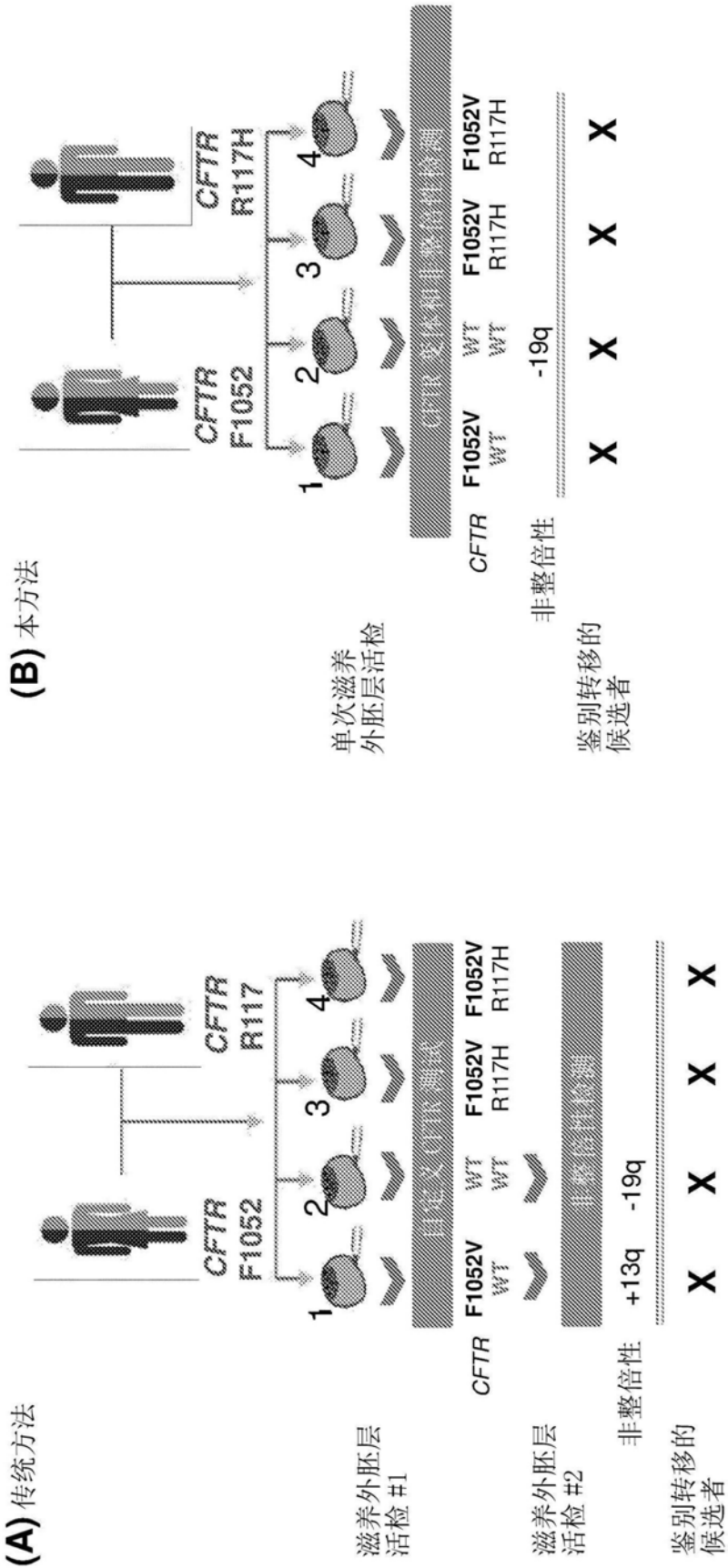


图14



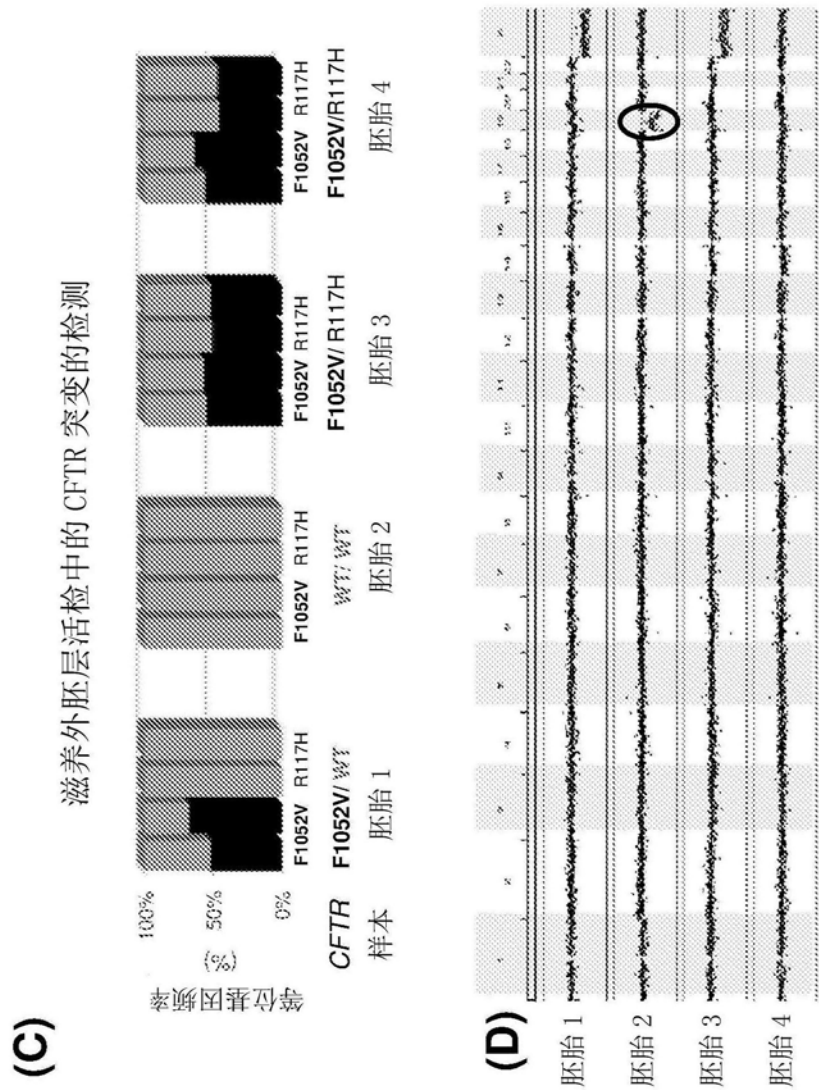


图14 (续)

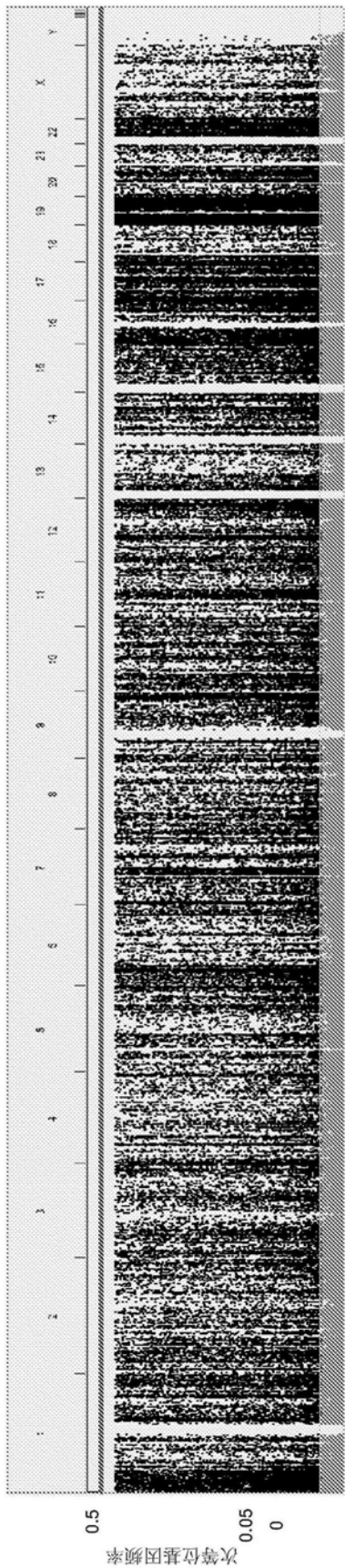


图15

Alu 样引物序列	
FWD-1	TCCACGGTCTCCCTCTCATGC
FWD-2	CGGCTCACTGCAACCTCCCTG
FWD-3	GATCCTCCTGCCTCGGCCTCC
FWD-4	CGAGTGCCCTGGGATTGCAGGC
FWD-5	GTGGAGACGGGGTCTCGCCGT
REV-1	AGAGGCTGCAATCTCGGCAC
REV-2	ACGCCACTGCACCTCCAGCCTG
REV-3	GGCACCATTGAGCACTGAGTG
REV-4	CGAGACTCCGTCTGCAATCCC
REV-5	CCGGCACCTCGGGAGGCCCGAG

SINE-R 引物序列	
FWD-1	GGAGCCCCCTCTGCCCGGCCAC
FWD-2	GGGCCATGATGACGATGGCGG
FWD-3	AGAGATCAGATTGTTACTGTG
FWD-4	GTGCAAGATGTGCTTTGTTAA
FWD-5	TGAAGGCAGCATGCTCGTTAA
REV-1	GATAATCTTGGGTGTTCTC
REV-2	AGAGGGGATTTGGCAGGGTC
REV-3	AATAGTGGAGGGAAGGTCAGC
REV-4	AGGGAAGGTCAGCAGATAAAC
REV-5	GTCTCTGGTTTTCTTAGGCAG

Alu 样引物：预测的 PCR 产物的数量					
	REV-1	REV-2	REV-3	REV-4	REV-5
FWD-1	222	6,928	249	267	248
FWD-2	1,050	7,197	1,413	1,344	1,245
FWD-3	613	8,245	632	641	737
FWD-4	806	7,109	1,092	1,047	972
FWD-5	93	6,881	93	93	6

SINE 引物 - 预测的 PCR 产物的数量					
	REV-1	REV-2	REV-3	REV-4	REV-5
FWD-1	1,541	1,560	1,524	1,494	1,609
FWD-2	480	502	480	487	513
FWD-3	500	419	437	450	746
FWD-4	2,296	2,216	2,151	2,156	2,551
FWD-5	1,466	1,447	1,374	1,402	1,456

图16