



(12)发明专利

(10)授权公告号 CN 103987856 B

(45)授权公告日 2016.08.24

(21)申请号 201180075423.4

(22)申请日 2011.12.17

(85)PCT国际申请进入国家阶段日
2014.06.10

(86)PCT国际申请的申请数据
PCT/CN2011/084165 2011.12.17

(87)PCT国际申请的公布数据
W02013/086744 ZH 2013.06.20

(73)专利权人 深圳华大基因股份有限公司
地址 518083 广东省深圳市盐田区洪安三
街21号华大综合园7栋7层-14层

(72)发明人 邱咏 刘立夫 蒋慧 陈芳
张春雷 汪建 王俊 杨焕明
张秀清

(74)专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201

代理人 李志东

(51)Int.Cl.
C12Q 1/68(2006.01)

(56)对比文件
CN 101675169 A,2010.03.17,
CN 101305087 A,2008.11.12,

审查员 李谦

权利要求书5页 说明书16页 附图6页

(54)发明名称

确定基因组是否存在异常的方法及系统

(57)摘要

本发明提供了用于确定基因组是否存在异常的方法及系统。确定基因组是否存在异常的方法包括以下步骤：从孕妇样本中分离胎儿有核红细胞；对所述有核红细胞基因组的至少一部分进行测序，以便获得测序结果；以及基于所述测序结果，确定所述有核红细胞基因组是否存在异常。

1. 一种确定基因组是否存在异常的非诊断方法,其特征在于,包括以下步骤:
从孕妇样本中分离胎儿有核红细胞;
对所述有核红细胞基因组进行测序,以便获得由多个测序数据构成的测序结果;以及
基于所述测序结果,确定所述有核红细胞基因组是否存在异常,所述异常为染色体非整倍性,包括:
将已知的第一染色体序列划分为多个窗口,所述多个窗口分别独立地具有预定的长度,
将所述测序结果的测序数据与所述已知的第一染色体序列进行比对,以便获得落入每个窗口的测序数据的数目,
基于获得落入每个窗口的测序数据的数目,确定第一参数,
基于所述第一参数,确定所述有核红细胞针对所述第一染色体是否具有非整倍性;
其中,所述确定第一参数包括,
分别对落入每个窗口的测序数据的数目设置预定的加权系数,以及
根据所述加权系数,将所述落入每个窗口的测序数据的数目进行加权平均,以获得所述第一染色体的中位数,所述第一染色体的中位数构成所述第一染色体的第一参数,所述预定的加权系数是通过将落入每个窗口的测序数据的数目分别与各自窗口的GC含量进行关联而获得的。
2. 根据权利要求1所述的方法,其特征在于,所述孕妇样本为孕妇外周血。
3. 根据权利要求1或2所述的方法,其特征在于,所述孕妇的孕周为12-20周。
4. 根据权利要求2所述的方法,其特征在于,从所述孕妇外周血中分离胎儿有核红细胞进一步包括:
利用密度梯度试剂,对所述外周血进行梯度离心,以便获得单核细胞;以及
利用携带抗体的磁珠,从所述单核细胞富集有核红细胞,其中,所述抗体特异性识别有核红细胞表面的抗原。
5. 根据权利要求4所述的方法,其特征在于,所述密度梯度试剂是聚蔗糖。
6. 根据权利要求5所述的方法,在800Xg下进行所述梯度离心30分钟。
7. 根据权利要求4所述的方法,其特征在于,在获得所述单核细胞之后,在利用携带CD71抗体的磁珠富集所述有核红细胞之前,进一步包括利用含有1%BSA的PBS缓冲液对所述单核细胞进行清洗,以便去除残余的密度梯度试剂。
8. 根据权利要求7所述的方法,其特征在于,所述PBS缓冲液不含钙离子和镁离子。
9. 根据权利要求7所述的方法,其特征在于,利用含有1%的PBS缓冲液对所述单核细胞进行清洗进一步包括:
将所述单核细胞与所述含有1%BSA的PBS缓冲液混合,以便获得含有单核细胞的悬浮液;以及
将所述含有单核细胞的悬浮液进行离心,在200Xg下离心5分钟,弃上清,以便获得经过清洗的有核红细胞。
10. 根据权利要求4所述的方法,其特征在于,所述抗体为特异性识别CD71的抗体。
11. 根据权利要求1所述的方法,其特征在于,对一个有核红细胞进行测序。
12. 根据权利要求1所述的方法,其特征在于,对所述有核红细胞的全基因组进行测序。

13. 根据权利要求12所述的方法,其特征在于,对所述有核红细胞的全基因组进行测序进一步包括:

对所述有核红细胞的全基因组进行扩增得到经过扩增的全基因组;

利用所述经过扩增的全基因组构建全基因组测序文库;以及

对所述全基因组测序文库进行测序,以便获得由多个测序数据构成的测序结果。

14. 根据权利要求13所述的方法,其特征在于,通过OmniPlex WGA方法对所述有核红细胞的全基因组进行扩增。

15. 根据权利要求13所述的方法,其特征在于,利用所述经过扩增的全基因组构建全基因组测序文库进一步包括:

将所述经过扩增的全基因组片段化,以便获得DNA片段;

将所述DNA片段进行末端修复,以便获得经过末端修复的DNA片段;

将所述经过末端修复的DNA片段的3'末端添加碱基A,以便获得具有粘性末端A的DNA片段;

将所述具有粘性末端A的DNA片段与接头相连,以便获得连接产物;

将所述连接产物进行PCR扩增,以便获得第二扩增产物;以及

将所述第二扩增产物进行纯化回收,以便获得回收产物,所述回收产物构成所述全基因组测序文库。

16. 根据权利要求15所述的方法,其特征在于,将所述经过扩增的全基因组片段化是通过Covaris打断仪进行的。

17. 根据权利要求15所述的方法,其特征在于,所述DNA片段的长度为350bp。

18. 根据权利要求17所述的方法,其特征在于,将所述DNA片段进行末端修复是利用Klenow片段、T4 DNA聚合酶和T4多核苷酸激酶进行的,所述Klenow片段具有5'→3'聚合酶活性和3'→5'外切酶活性,但缺少5'→3'外切酶活性。

19. 根据权利要求15所述的方法,其特征在于,将所述经过末端修复的DNA片段的3'末端添加碱基A是利用Klenow片段(3'-5' exo-)进行的。

20. 根据权利要求15所述的方法,其特征在于,将所述具有粘性末端A的DNA片段与接头相连是利用T4 DNA连接酶进行的。

21. 根据权利要求1所述的方法,其特征在于,利用选自Hi seq2000、SOLiD、454、和单分子测序装置的至少一种进行所述测序。

22. 根据权利要求1所述的方法,其特征在于,所述第一染色体为选自人类21号染色体、18号染色体、13号染色体、X染色体和Y染色体的至少一种。

23. 根据权利要求1所述的方法,其特征在于,所述多个窗口的预定的长度相同。

24. 根据权利要求23所述的方法,其特征在于,所述多个窗口的预定的长度均为60KB。

25. 根据权利要求1所述的方法,其特征在于,所述落入每个窗口的测序数据为唯一比对的测序数据。

26. 根据权利要求1所述的方法,其特征在于,基于所述第一参数,确定所述有核红细胞针对所述第一染色体是否具有非整倍性进一步包括:

针对第二染色体,进行与所述第一染色体相同的处理,以便获得所述第二染色体的中位数;

将所述第一染色体的中位数与所述第二染色体的中位数进行t值检验,以便获得所述第一染色体与所述第二染色体的差异值;

将所述差异值与预定的第一阈值和第二阈值进行比较,如果所述差异值低于所述预定的阈值,则确定针对第一染色体,所述有核红细胞具有非整倍性,如果所述差异值高于所述预定的阈值,则确定针对第一染色体,所述有核细胞不具有非整倍性。

27. 根据权利要求26所述的方法,其特征在于,所述第二染色体为选自人基因组第1-12号染色体中的任一种。

28. 根据权利要求26所述的方法,其特征在于,利用下列公式对所述第一染色体的中位数与所述第二染色体的中位数进行t值检验,

$$T_{i,j} = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}}$$

其中, $T_{i,j}$ 表示所述第一染色体与所述第二染色体的差异值, μ_i 表示第一染色体的中位数, μ_j 表示第二染色体的中位数, σ_i 表示第一染色体各个窗口中测序数据数目分布的标准差, σ_j 表示第二染色体各个窗口中测序数据数目分布的标准差, n_i 表示第一染色体中窗口的数目, n_j 表示第二染色体中窗口的数目。

29. 根据权利要求26所述的方法,其特征在于,所述预定的第一阈值为-4或者更小,所述第二阈值为-3.5或者更大。

30. 根据权利要求1所述的方法,其特征在于,所述基因组异常还包括预定区域的突变。

31. 根据权利要求30所述的方法,其特征在于,基于所述测序结果,确定所述有核红细胞基因组是否存在异常进一步包括:

基于所述测序结果,确定所述有核红细胞的预定区域的核酸序列;

将所述有核红细胞的预定区域的核酸序列与对照核酸序列进行比对,所述对照核酸序列为正常人的基因组序列;以及

基于所述比对结果,确定所述有核红细胞的预定区域是否存在异常。

32. 根据权利要求31所述的方法,其特征在于,所述,所述预定区域的突变包括选自插入突变、缺失突变、置换突变、异位突变、倒置突变、拷贝数变异以及单核苷酸多态性的至少一种。

33. 一种用于确定基因组是否存在异常的系统,其特征在于,包括:

有核红细胞分离装置,所述有核红细胞分离装置用于从孕妇样本中分离胎儿有核红细胞;

测序装置,所述测序装置用于对所述有核红细胞基因组进行测序,以便获得由多个测序数据构成的测序结果;以及

测序结果分析装置,所述测序结果分析装置与测序装置相连,以便从所述测序装置接收所述测序结果,并且基于所述测序结果,判断所述有核红细胞基因组是否存在异常,所述异常为染色体非整倍性,所述测序结果分析装置适于执行下列操作:

将已知的第一染色体序列划分为多个窗口,所述多个窗口分别独立地具有预定的长度,

将所述测序结果的测序数据与所述已知的第一染色体序列进行比对,以便获得落入每

个窗口的测序数据的数目，

基于获得落入每个窗口的测序数据的数目，确定第一参数，

基于所述第一参数，确定所述有核红细胞针对所述第一染色体是否具有非整倍性；

其中，所述测序结果分析装置进一步包括适于通过下列步骤基于获得落入每个窗口的测序数据的数目，确定第一参数的单元：

分别对落入每个窗口的测序数据的数目设置预定的加权系数，以及

根据所述加权系数，将所述落入每个窗口的测序数据的数目进行加权平均，以获得所述第一染色体的中位数，所述第一染色体的中位数构成所述第一染色体的第一参数，所述预定的加权系数是通过将落入每个窗口的测序数据的数目分别与各自窗口的GC含量进行关联而获得的。

34. 根据权利要求33所述的系统，其特征在于，所述有核红细胞分离装置进一步包括：

单核细胞分离单元，所述单核细胞分离单元适于利用密度梯度试剂，对所述孕妇样本进行梯度离心，以便获得单核细胞；以及

磁性富集单元，所述磁性富集单元与所述单核细胞分离单元相连，并且适于利用携带抗体的磁珠，从所述单核细胞富集有核红细胞，其中，所述抗体特异性识别有核红细胞表面的抗原。

35. 根据权利要求34所述的系统，其特征在于，所述孕妇样本为孕妇外周血。

36. 根据权利要求33所述的系统，其特征在于，所述测序装置包括选自HiSeq2000、SOLiD、454、和单分子测序装置的至少一种。

37. 根据权利要求33所述的系统，其特征在于，进一步包括全基因组测序文库制备装置，所述全基因组测序文库装置与测序装置相连，并为所述测序装置提供用于测序的全基因组测序文库，

其中，

所述全基因组测序文库制备装置进一步包括：

有核红细胞裂解单元，所述有核红细胞裂解单元与所述有核红细胞分离装置相连，接收并且裂解有核红细胞，以便释放所述有核红细胞的全基因组；

全基因组扩增单元，所述全基因组扩增单元与所述有核红细胞裂解单元相连，用于对所述有核红细胞的全基因组进行扩增，以便得到经过扩增的全基因组；以及

测序文库构建单元，所述测序文库构建单元用于接收所述经过扩增的全基因组，并且利用所述经过扩增的全基因组构建所述全基因组测序文库。

38. 根据权利要求33所述的系统，其特征在于，所述测序结果分析装置进一步包括适于通过下列进行基于所述第一参数，确定所述有核红细胞针对所述第一染色体是否具有非整倍性的单元：

针对第二染色体，进行与所述第一染色体相同的处理，以便获得所述第二染色体的中位数；

将所述第一染色体的中位数与所述第二染色体的中位数进行t值检验，以便获得所述第一染色体与所述第二染色体的差异值；

将所述差异值与预定的阈值进行比较，如果所述差异值低于所述预定的阈值，则确定针对第一染色体，所述有核红细胞具有非整倍性。

39. 根据权利要求38所述的系统,其特征在于,所述测序结果分析装置进一步包括适于利用下列公式对所述第一染色体的中位数与所述第二染色体的中位数进行t值检验的单元,

$$T_{i,j} = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}}$$

其中, $T_{i,j}$ 表示所述第一染色体与所述第二染色体的差异值, μ_i 表示第一染色体的中位数, μ_j 表示第二染色体的中位数, σ_i 表示第一染色体各个窗口中测序数据数目分布的标准差, σ_j 表示第二染色体各个窗口中测序数据数目分布的标准差, n_i 表示第一染色体中窗口的数目, n_j 表示第二染色体中窗口的数目。

40. 根据权利要求33所述的系统,其特征在于,所述测序结果分析装置进一步包括:

核酸序列确定单元,所述预定区域核酸序列确定单元适于基于所述测序结果,确定所述有核红细胞的预定区域的核酸序列;

比对单元,所述比对单元与所述核酸序列确定单元相连,并且适于将所述有核红细胞的预定区域的核酸序列与对照核酸序列进行比对;以及

异常确定单元,所述异常确定单元与所述比对单元相连,并且适于基于所述比对结果,确定所述有核红细胞的预定区域是否存在异常。

41. 根据权利要求40所述的系统,其特征在于,所述比对单元中存储有对照核酸序列,所述对照核酸序列为正常人的基因组序列。

确定基因组是否存在异常的方法及系统

技术领域

[0001] 本发明涉及生物医学领域。具体而言,涉及确定基因组是否存在异常的方法及系统,更具体的,本发明涉及一种确定胎儿有核红细胞基因组序列的方法、一种确定基因组是否存在异常的方法以及一种确定基因组是否存在异常的系统。

背景技术

[0002] 产前诊断又称出生前诊断,是指结合遗传学检测和影像检查结果,对出生前的胎儿是否患有某些遗传病或先天畸形做出高准确度的诊断。目前所用的产前诊断方法按照取材方法的不同主要分为有创性诊断和无创性诊断。有创诊断主要有羊膜腔穿刺(羊水检测)、绒毛膜穿刺、脐血取样、胎儿镜和胚胎活检等,目前应用较普遍的是羊膜腔穿刺和绒毛膜穿刺。有创诊断因为可以直接对胎儿细胞或组织进行取材,在经过遗传学检测后可以得到准确可靠的结果,但是由于有创取样过程会给孕妇和胎儿带来潜在的危害,严重的甚至引起胎儿流产或宫内感染等。据统计羊膜腔穿刺和绒毛膜穿刺可分别导致约1%和1-2%的流产率。

[0003] 目前产前诊断的方法,仍有待改进。

发明内容

[0004] 本发明旨在至少解决现有技术中存在的技术问题之一。为此,本发明提出了能够有效确定基因组是否存在异常的方法和系统。

[0005] 根据本发明的一个方面,本发明提出了一种确定基因组是否存在异常的方法。根据本发明的实施例,该确定基因组是否存在异常的方法包括以下步骤:从孕妇样本中分离胎儿有核红细胞;对所述有核红细胞基因组的至少一部分进行测序,以便获得测序结果;以及基于所述测序结果,确定所述有核红细胞基因组是否存在异常。发明人发现,利用根据本发明实施例的方法,能够有效地确定从孕妇样本中分离的胎儿有核红细胞基因组是否存在异常。该方法可以是非医疗目的的。

[0006] 根据本发明的第二方面,本发明提出了一种用于确定基因组是否存在异常的系统。根据本发明的实施例,该系统包括:有核红细胞分离装置,所述有核红细胞分离装置用于从孕妇样本中分离胎儿有核红细胞;测序装置,所述测序装置用于对所述有核红细胞基因组的至少一部分进行测序,以便获得测序结果;以及测序结果分析装置,所述测序结果分析装置与测序装置相连,以便从所述测序装置接收所述测序结果,并且基于所述测序结果,判断所述有核红细胞基因组是否存在异常。利用该用于确定有核红细胞染色体非整倍性的系统,能够有效地实施根据本发明实施例的确定基因组是否存在异常的方法,由此,能够有效地确定有核红细胞的基因组是否存在异常。

[0007] 根据本发明的又一方面,本发明提出了一种确定胎儿有核红细胞基因组序列的方法。根据本发明的实施例,包括以下步骤:从孕妇样本中分离胎儿有核红细胞;以及对所述有核红细胞基因组的至少一部分进行测序,以便获得测序结果。利用该方法,能够有效地确

定有核红细胞的基因组序列信息,从而可以确定胎儿基因组的序列信息。

[0008] 本发明的附加方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

附图说明

[0009] 本发明的上述和/或附加的方面和优点从结合下面附图对实施例的描述中将变得明显和容易理解,其中:

[0010] 图1显示了根据本发明一个实施例的确定有核细胞基因组是否存在异常的方法的流程示意图。

[0011] 图2显示了根据本发明又一个实施例的确定有核细胞基因组是否存在异常的方法的流程示意图。

[0012] 图3显示了根据本发明一个实施例的用于有核细胞基因组是否存在异常的系统的示意图。

[0013] 图4显示了根据本发明一个实施例的有核红细胞分离装置的示意图。

[0014] 图5显示了根据本发明又一个的用于有核细胞基因组是否存在异常的系统的示意图。

[0015] 图6显示了根据本发明一个实施例的用于全基因组测序文库制备装置的示意图。

[0016] 图7显示了根据本发明一个实施例,构建的DNA文库,经过 Agilent®Bioanalyzer2100 的检测结果。简言之,将分离得到的阳性细胞(有核红细胞)全基因组扩增后产物进行超声波打断,打断主带为350bp左右,在连接接头以后片段长度增加约120bp左右,切胶回收430-450bp的片段,由图7可以看出四个文库片段范围符合要求,而且文库质量符合测序要求,GP9为检测样品,YH6为对样品(正常人样品)。

[0017] 图8显示了根据本发明一个实施例的对测序数据进行分析的结果。其中,(A)为待测样本各个窗口的GC值与唯一比对的测序数据数目的分布;(B)为经过smooth spline拟合后的GC含量与唯一比对的测序数据数目关系的曲线;(C)为待测样品各个窗口数据进行修正对应的加权系数的分布,每一个GC含量的窗口对应一个UR值作为修正权重;(D)为显示各条染色体测序数据的箱式图。

具体实施方式

[0018] 下面详细描述本发明的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,仅用于解释本发明,而不能理解为对本发明的限制。

[0019] 需要说明的是,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。进一步地,在本发明的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0020] 1、确定基因组是否存在异常的方法

[0021] 本发明的一个方面提出了一种确定基因组是否存在异常的方法。参考图1,根据本发明的实施例,该方法包括以下步骤:

[0022] S10:从孕妇样本中分离胎儿有核红细胞。

[0023] 在本发明中选择从孕妇样本中分离胎儿的有核红细胞是基于发明人的下述发现而完成的:目前,针对胎儿遗传异常进行研究,主要基于从孕妇体内分离胎儿游离DNA。然而,由于孕妇外周血中除了胎儿来源的游离DNA,还有存在大量母源DNA,而且胎儿基因组DNA的一半是来源于母亲,因此准确的区分DNA的来源在目前来说比较困难。同时,母体外周血中胎儿游离DNA是以不完整的基因组存在的,因而在对特定基因位点检测过程中会因模板丢失而导致假阴性的几率大大增加。因而,利用胎儿游离DNA进行胎儿遗传异常的检测有其自身的缺陷,另外,发明人进一步发现,在母体的外周血中除了存在游离胎儿核酸外,还存在着完整的胎儿细胞。游离于孕妇外周血的胎儿细胞主要包括:滋养层细胞,白细胞及胎儿有核红细胞。其中,发明人发现,滋养层细胞由于存在多核和单核两种形式,容易导致误诊。白细胞在胎儿出生以后会持久的存在于母体血液中,会干扰下一次怀孕的检测。并且发明人发现胎儿有核红细胞有较短的生命周期,在胎儿出生后90天内就会消失,不会干扰第二胎的检测,同时,有核红细胞表面的抗原相对稳定,易于识别和分离。因此,利用胎儿有核红细胞能够有效地对胎儿基因组的异常进行确定。发明人发现,利用孕妇外周血中胎儿有核红细胞结合高通量测序进行无创产前诊断的方法的效果,要比现在用孕妇血浆进行无创产前诊断要优越许多。

[0024] 根据本发明的实施例,作为有核红细胞来源的孕妇样本不受特别限制。根据本发明的一些实施例,根据本发明的一个实施例,所述孕妇样本优选为孕妇外周血。由此,可以方便地从孕妇获取这些样本,并且可以在不影响胎儿发育的前提下,从孕妇的外周血中获得胎儿的有核红细胞进行全基因组测序,实现无创产前检验。另外,对于作为胎儿有核红细胞来源的孕妇的怀孕阶段并不受特别限制。根据本发明的实施例,可以采用孕周为20周以下的孕妇样本提取胎儿有核红细胞进行,例如可以采用孕周为12-20周的孕妇样本作为研究对象。由此,能够更有效地提取胎儿有核红细胞进行研究。另外,发明人惊奇地发现,利用本发明的方法,即使仅获得1个胎儿有核红细胞,也能够有效地进行分析。

[0025] 根据本发明的实施例,从生物样本例如外周血中分离有核红细胞的方法,并不受特别限制。根据本发明的具体实施例,从外周血分离所述有核红细胞进一步包括下列步骤:

[0026] 首先,利用密度梯度试剂,对所述外周血进行梯度离心,以便获得单核细胞。根据本发明的实施例,密度梯度试剂的类型并不受特别限制,根据具体的实例,可以采用聚蔗糖,例如Ficoll形成密度梯度。优选地,可以在800X g下进行所述梯度离心30分钟。

[0027] 在获得单核细胞后,利用携带抗体的磁珠,从所得到的单核细胞富集有核红细胞,其中,磁珠上所携带的抗体特异性识别有核红细胞表面的抗原,从而有核红细胞会通过抗体与磁珠结合,接下来可以通过磁性筛选,获得有核红细胞。

[0028] 根据本发明的实施例,在获得所述单核细胞之后,在利用携带CD71抗体的磁珠富集所述有核红细胞之前,进一步包括利用含有1%BSA的PBS缓冲液对所述单核细胞进行清洗,以便去除残余的密度梯度试剂,优选所述PBS缓冲液含有磷酸二氢钾和磷酸氢二钠,但不含钙离子和镁离子,由此,能够显著提高富集有核红细胞的效率。根据本发明的具体实例,利用含有1%的PBS缓冲液对所述单核细胞进行清洗进一步包括:将所述单核细胞与所述含有1%BSA的PBS缓冲液混合,以便获得含有单核细胞的悬浮液;以及将所述含有单核细胞的悬浮液进行离心,优选在200X g下离心5分钟,弃上清,以便获得经过清洗的有核红细

胞。优选地,所述抗体为特异性识别CD71的抗体。发明人发现,通过根据本发明实施例的分离有核红细胞的方法,能够有效地从外周血中分离有核红细胞,尤其是能够分离到胎儿的有核红细胞。由此,本发明提出了一种简单容易操作的从外周血分离胎儿有核红细胞的方法。当然,本领域技术人员可以理解,在分离有核红细胞的过程中还可以具有其他步骤。例如,根据本发明的具体示例,分离有核红细胞的方法包括:取适量孕妇的外周血,抗凝剂抗凝,将血液样本用不含钙离子和镁离子的0.1M磷酸盐缓冲液(PBS)按比例进行稀释,将稀释后的样本缓置于密度梯度离心用试剂上,在室温条件下,进行密度梯度离心。离心之后可见单核细胞层,小心吸出该层细胞,并转移至新的离心管中,用3倍体积的含1%BSA的PBS缓冲液重悬,在室温下再离心,去上清,将所得到的细胞沉淀再用相同的方法洗2次以去除残留的密度梯度液,最后将细胞沉淀重悬于含0.1%BSA的PBS中,吹匀。然后按20微升/ 10^6 个细胞的比例加入携带抗体的磁珠,4℃静置后离心,弃上清,将沉淀重悬于含0.1%BSA的PBS中;组装磁珠分选系统。以500微升含0.1%BSA的PBS缓冲液润洗分选柱,待液体流空后,将分选的细胞上柱,收集流出液,标记为阴性细胞。用含0.1%BSA的PBS润洗管子,待液体流空后再重复2次,用PBS/EDTA/BSA加入分选柱,待液体流空后再重复一次。最后,加入含0.1%BSA的PBS到分选柱中,离开磁场,将液体冲洗入新的离心管内,即得到有核红细胞。

[0029] S20:在从孕妇样本中分离胎儿有核红细胞之后,可以对有核红细胞基因组的至少一部分进行测序,从而可以获得对应测序对象的测序结果。

[0030] 根据本发明的实施例,在分离胎儿有核红细胞之后,可以对有核红细胞基因组的至少一部分进行测序。本领域技术人员可以根据所感兴趣的基因来选择有核红细胞基因组的测序对象,从而获得对应这些测序对象的测序结果。根据本发明的实施例,本领域技术人员可以采取任何已知的方法,选择测序对象,例如,可以仅选择其中的几条染色体。当然,本领域技术人员能够理解的是,也可以直接对有核红细胞的全基因组进行测序,从而在得到测序结果之后,再从测序结果中选择来自特定位点的测序数据进行研究(详见后)。为了方便起见,下面以对有核红细胞的全基因组进行测序为例进行说明。

[0031] 根据本发明的实施例,在获得有核红细胞后,对有核红细胞的全基因组进行测序的方法不受特别限制。根据本发明的一个实施例,对有核红细胞的全基因组进行测序进一步包括:首先,对有核红细胞的全基因组进行扩增得到经过扩增的全基因组;接下来,利用经过扩增的全基因组构建全基因组测序文库;最后,对全基因组测序文库进行测序,以便获得由多个测序数据构成的测序结果。由此,能够有效地获取有核红细胞的全基因组信息,从而进一步提高了确定有核红细胞基因组是否存在异常的效率。本领域技术人员可以根据采用的基因组测序技术的具体方案选择不同的构建全基因组测序文库的方法,关于构建全基因组测序文库的细节,可以参见测序仪器的厂商例如Illumina公司所提供的规程,例如参见Illumina公司Multiplexing Sample Preparation Guide(Part#1005361;Feb2010)或Paired-End SamplePrep Guide(Part#1005063;Feb2010),通过参照将其并入本文。

[0032] 任选地,根据本发明的实施例,可以进一步包括对所述有核红细胞进行裂解,以便释放所述有核红细胞的全基因组的步骤。根据本发明的一些示例,可以用于裂解有核红细胞并释放全基因组的方法不受特别限制,只要能够将有核红细胞裂解优选充分裂解即可。根据本发明的具体示例,可以利用碱性裂解液将所述有核红细胞裂解并释放所述有核红细胞的全基因组。发明人发现,这样能够有效地裂解有核红细胞并释放出全基因组,并且所释

放的全基因组在进行测序时,能够提高准确率,从而进一步提高了确定有核红细胞染色体非整倍性的效率。根据本发明的实施例,有核红细胞全基因组扩增的方法不受特别限制,可以采用基于PCR的方法例如可以采用PEP-PCR、DOP-PCR、和OmniPlex WGA,也可以采用非基于PCR的方法例如MDA(多重链置换扩增)。根据本发明的具体示例,优选采用基于PCR的方法,例如OmniPlex WGA方法。可选用的商业化试剂盒包括但不限于Sigma Aldrich的GenomePlex,Rubicon Genomics的PicoPlex,Qiagen的REPLI-g,GE Healthcare的illustra GenomiPhi等。因而,根据本发明的具体示例,在构建测序文库之前,可以采用OmniPlex WGA对有核红细胞全基因组进行扩增。由此,能够有效地对全基因组进行扩增,从而进一步提高了确定有核红细胞染色体非整倍性的效率。

[0033] 根据本发明的实施例,利用所述经过扩增的全基因组构建全基因组测序文库(在本文中有时也称为“核酸文库”或“测序文库”)进一步包括:

[0034] 首先,将经过扩增的全基因组片段化,以便获得DNA片段;根据本发明的实施例,将所得的DNA进行片段化的方法不受特别限制,根据一些具体示例,可以通过选自雾化、超声打断法、HydroShear以及酶切处理的至少一种进行片段化,优选使用covaris超声打断仪将经过扩增的全基因组进行片段化。根据本发明的实施例,片段化处理后所得的DNA片段的长度为200-400bp,优选350bp。发明人发现,通过采用所获得的该长度的DNA片段能够有效地用于核酸文库的构建及后续处理。

[0035] 在获得DNA片段后,可以将所得到的DNA片段进行末端修复,以便获得经过末端修复的DNA片段。根据本发明的实施例,可以利用Klenow片段、T4DNA聚合酶和T4多核苷酸激酶将DNA片段进行末端修复,其中,该Klenow片段具有5'→3'聚合酶活性和3'→5'外切酶活性,但缺少5'→3'外切酶活性,由此,能够有效地将DNA片段进行末端修复。

[0036] 在对DNA片段进行末端修复后,可以将经过末端修复的DNA片段的3'末端添加碱基A,以便获得具有粘性末端A的DNA片段。根据本发明的一些具体示例,可以利用Klenow片段(3'-5' exo-),即缺失了3'→5'外切酶活性的Klenow片段,将经过末端修复的DNA片段的3'末端添加碱基A,由此,能够有效地获得具有粘性末端A的DNA片段。

[0037] 在末端添加碱基A后,可以将具有粘性末端A的DNA片段与接头相连,以便获得连接产物。根据本发明的实施例,可以利用T4DNA连接酶将具有粘性末端A的DNA片段与接头相连,由此,能够有效地获得连接产物。根据本发明的实施例,该接头中可以进一步包含标签,由此可以方便地同时构建多个有核红细胞样本的全基因组测序文库,对多个样本的测序文库进行组合,同时进行测序。由此,能够充分地利用高通量测序平台,有效地节省时间、降低测序成本。

[0038] 在获得连接产物后,将所述连接产物进行PCR扩增,以便获得第二扩增产物;以及将所述第二扩增产物进行纯化回收,以便获得回收产物,所述回收产物构成所述全基因组测序文库。

[0039] 在构建了全基因组测序文库后,根据本发明的实施例,可以对所述全基因组测序文库进行测序。当然,本领域技术人员能够理解的是本发明的测序步骤可通过任何测序方法进行,包括但不限于双脱氧链终止法;优选高通量的测序方法,由此,能够利用这些测序装置的高通量、深度测序的特点,进一步提高了确定有核红细胞染色体非整倍性的效率。所述高通量的测序方法包括但不限于第二代测序技术或者是单分子测序技术。

[0040] 所述第二代测序平台(Metzker ML.Sequencing technologies—the next generation.Nat Rev Genet.2010Jan;11(1):31-46)包括但不限于Illumina-Solexa(GA™, HiSeq2000™等)、ABI-Solid和Roche-454(焦磷酸测序)测序平台;单分子测序平台(技术)包括但不限于Helicos公司的真实单分子测序技术(True Single Molecule DNA sequencing),Pacific Biosciences公司单分子实时测序(single molecule real-time (SMRT™)),以及Oxford Nanopore Technologies公司的纳米孔测序技术等(Rusk,Nicole (2009-04-01).Cheap Third-Generation Sequencing.Nature Methods6(4):244-245)。

[0041] 随着测序技术的不断进化,本领域技术人员能够理解的是还可以采用其他的测序方法和装置进行全基因组测序。根据本发明的实施例,通过全基因组测序所得到的测序数据的长度不受特别限制。根据本发明的一个具体示例,所述多个测序数据的平均长度为约50bp。发明人发现,当测序数据的平均长度为约50bp时,能够极大地方便对测序数据进行分析,提高分析效率,同时能够显著降低分析的成本。进一步提高了确定有核红细胞染色体非整倍性的效率,并且降低了确定有核红细胞染色体非整倍性的成本。这里所使用的术语“平均长度”是指各个测序数据长度数值的平均值。

[0042] S30:在获得测序结果之后,基于所得到的测序结果,确定有核红细胞基因组是否存在异常。

[0043] 在本文中所使用的术语“基因组异常”应做广义理解,其可以是指的是基因组序列的任何变化,例如染色体非整倍性,结构变异,单核苷酸突变等遗传变异(http://en.wikipedia.org/wiki/Genetic_variation);也可以是基因组修饰位点的变化例如甲基化水平等。根据本发明的实施例,所研究的基因组异常为选自染色体的非整倍型、和预定区域的突变至少一种。在本发明的实施例中预定区域的突变是指结构变异(http://en.wikipedia.org/wiki/Structural_variation)或单核苷酸突变(SNP,http://en.wikipedia.org/wiki/Single-nucleotide_polymorphism)。

[0044] 根据本发明的实施例,确定基因组中预定区域的突变可以进一步包括:

[0045] 首先,基于测序结果,确定所述有核红细胞的预定区域的核酸序列。本领域技术人员可以采用任何已知的工具确定预定区域的核酸序列。例如,可以采用已知的手段,对测序结果中来自特定区域的测序数据进行组装,从而得到来自预定序列的核酸序列。

[0046] 在得到来自该预定序列的核酸序列后,将所述有核红细胞的预定区域的核酸序列与对照核酸序列进行比对,优选所述对照核酸序列为正常人的基因组序列。接下来,可以基于该比对结果,确定有核红细胞的预定区域是否存在异常。根据本发明的实施例,通过该方法可以检测的预定区域的突变包括选自插入突变、缺失突变、置换突变、倒置突变、拷贝数变异、异位突变以及单核苷酸多态性的至少一种。

[0047] 参考图2,根据本发明的实施例,确定染色体的非整倍性的方法包括下列步骤:

[0048] S100:首先,对有核红细胞的全基因组进行测序,以便获得第一测序结果。前面已经就对全基因组进行测序进行了详细描述,不再赘述。

[0049] S200:将第一染色体的已知序列划分窗口

[0050] 为了对有核红细胞的测序数据进行分析,首先将已知的第一染色体的序列进行划分窗口,这些窗口分别独立地具有预定的长度。根据本发明的实施例,这些窗口的长度可以相同也可以不同,并且不受特别限制。根据本发明的一个实施例,所述多个窗口的预定的长

度相同。优选,所述多个窗口的预定的长度均为60KB。由此,可以提高确定有核红细胞中染色体非整倍性的效率。另外,本领域技术人员可以根据需要选择窗口所覆盖的第一染色体的序列范围。

[0051] S300确定落入每个窗口的测序数据的数目

[0052] 在获得测序数据后,将所得到的测序数据与已知的第一染色体序列进行比对第一染色体进行比对,从而可以将所得到的测序数据分别划分到具有预定长度的窗口中。本领域技术人员能够理解,可以采用任何已知的方法和手段进行比对和对这些测序数据的总数目进行计算。例如,可以采用测序仪器的制造商所提供的软件进行分析,例如SOAP v2.20。根据本发明的一个实施例,所述落入每个窗口的测序数据为唯一比对的测序数据。由此,可以通过对测序数据进行筛选,而进一步提高确定有核红细胞中染色体非整倍性的效率。这里所使用的术语“唯一比对的测序数据”指的是,在将测序数据与已知的染色体序列,例如人类基因组Hg19进行比对,能够与参考基因组完全匹配且仅比对成功一次的测序数据,在本文中有时也称为unique read。

[0053] 这里所使用的术语“第一染色体”应做广义理解,其可以是指任何期望研究的目的染色体,其数目并不仅限于一条染色体,甚至可以同时将全部染色体进行分析。根据本发明的实施例,第一染色体可以为人类染色中的任意染色体,可以为选自人类1~23号染色体的任意染色体。根据本发明的实施例,优选为选自人类21号染色体、18号染色体、13号染色体、X染色体和Y染色体的至少一种。由此,能够有效地确定常见的人类染色体疾病,例如可以预测胎儿的遗传性疾病。因而,根据本发明的实施例的确定有核红细胞染色体非整倍性的方法,能够非常有效地应用于体外生殖领域中的植入前筛查(PGS)和植入前诊断(PGD),以及胎儿有核细胞的产前检测等,由此,可以通过简单地提取有核红细胞来快速预测胎儿的染色体是否存在异常,避免胎儿患有严重的遗传性疾病。在本文中所使用的术语“可比对到第一染色体”是指,通过将测序数据与参考基因组第一染色体的已知序列进行比对,能够与第一染色体的序列比对上的,从而确定这些测序数据来源于第一染色体的测序结果。

[0054] S400基于获得落入每个窗口的测序数据的数目,确定第一参数

[0055] 基于有核红细胞的全基因组测序的测序结果中,针对某一特定染色体的测序数据的数目,是与全基因组中该染色体的含量呈正相关的,因而,通过对测序结果中来源于某一特定染色体的测序数据的数目以及全基因组测序的总数目进行分析,能够有效地确定关于该染色体。为此,可以通过对落入第一染色体的每个窗口的测序数据的数目,进行分析,确定第一参数。为了能够使得该第一参数能够真实地反映第一染色体是否存在非整倍性,根据本发明的一个实施例,基于获得落入每个窗口的测序数据的数目,确定第一参数进一步包括:分别对落入每个窗口的测序数据的数目设置预定的加权系数(在本文有时也成为权重系数);以及根据该加权系数,将落入每个窗口的测序数据的数目进行加权平均,以获得所述第一染色体的中位数,所得到的中位数构成该第一染色体的第一参数。发明人发现,如此获得的第一参数,能够有效地体现测序数据中来自第一染色体的测序数据的数目,从而可以进一步提高确定染色体非整倍性的效率。根据本发明的实施例,设置加权系数,是为了能够消除在测序过程中可能存在的某些误差所引起的数据失真。根据本发明的一个示例,预定的加权系数是通过将落入每个窗口的测序数据的数目分别与各自窗口的GC含量进行关联而获得的。由此,可以有效地消除由于测序技术对GC含量高的区域的偏向所造成的误

差。例如,在本发明的实施例中,可以通过下列方法获得加权系数:

[0056] 首先,分别将人类全基因组各条染色体划分成固定长度60Kb的窗口,记录各个窗口的起始终止位置并统计其GC均值(记为 GC_{ref})。得到待测样本的测序数据后,将测序结果序列比对至基因组上,取出完全匹配且仅比对成功一次的测序数据,即唯一比对的测序数据,并获得所有唯一比对的测序数据的位点信息。统计待测样本比对结果对应在基因组每条染色体各个窗口中的唯一比对的测序数据的数目(记为 UR_{sample}),将所得到的 UR_{sample} 和 GC_{ref} 进行制图,得到图8A。通过图8A,可以看出,由于测序仪引入的GC偏好导致GC大致在[0.35,0.55]区域的测序数据分布较多。用smooth spline方法,将图8A中离散的点拟合成平滑的曲线,反应GC含量与测序数据数目关系的图,得到图8B,其中图8B中,所有的窗口按照GC均值步长为1%划分。由拟合后的数据可得出每个GC均值对应窗口中的唯一比对的测序数据的数目,即 M_{fit} 。根据公式 $W_{GC} = M/M_{fit}$,得到预定的加权系数 W_{GC} ,该加权系数的分布见图8C,其中M是待测样本落在同等GC均值窗口中唯一比对的测序数据的数目。

[0057] 接下来,在获得第一参数后,可以通过将第一参数与预定的对照参数进行比较,确定有核红细胞针对所述第一染色体是否具有非整倍性。本文中所使用的术语“预定的”,应做广义理解,可以是预先通过实验确定的,也可以是在进行生物样本分析时,采用平行实验获得的。这里所使用的术语“平行实验”应作广义理解,既可以指的是同时进行未知样品和已知样品的测序和分析,也可以是先后进行在相同条件下的测序和分析。例如,可以通过对已知具有非整倍性的有核红细胞样本或者没有非整倍性的有核红细胞进行试验所得到的针对该样本的第一参数作为对照参数。

[0058] 另外,发明人发现,可以通过将同一次测序中的不同染色体的测序数据的数目进行比较和统计分析,确定第一染色体是否具有非整倍性。为此,根据本发明的一个实施例,基于所述第一参数,确定所述有核红细胞针对所述第一染色体是否具有非整倍性进一步包括:对第二染色体,进行与第一染色体相同的处理,以便获得所述第二染色体的中位数;将第一染色体的中位数与第二染色体的中位数进行t值检验,以便获得所述第一染色体与所述第二染色体的差异值;将所得到的差异值与预定的第一阈值和第二阈值进行比较,如果所得到的差异值低于预定的阈值,则确定针对第一染色体,有核红细胞具有非整倍性,如果所得到的差异值高于所述预定的阈值,则确定针对第一染色体,有核细胞不具有非整倍性。由此,可以进一步体现来自第一染色体与其他染色体的差异,从而提高确定染色体非整倍性的效率。

[0059] 这里所使用的术语“第二染色体”应做广义理解,其可以是指任何期望研究的目的染色体,其数目并不仅限于一条染色体,甚至可以同时将除第一染色体外的其他全部染色体进行分析。根据本发明的实施例,第二染色体可以为人类染色中的任意染色体,针对胎儿有核红细胞,第二染色体优选为选自人类1~23号染色体的任意染色体。根据本发明的实施例,优选第二染色体为选自人基因组第1-12号染色体中的任一种。因为这些染色体通常不存在染色体非整倍性,因而可以有效地作为参考,对第一染色体进行检验,提高检验效率。

[0060] 根据本发明的一个实施例,利用下列公式对第一染色体的中位数与第二染色体的中位数进行t值检验,

[0061]
$$T_{i,j} = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}} \quad (\text{公式 1})$$

[0062] 其中, $T_{i,j}$ 表示述第一染色体与第二染色体的差异值, μ_i 表示第一染色体的中位数, μ_j 表示第二染色体的中位数, σ_i 表示第一染色体各个窗口中测序数据数目分布的标准差, σ_j 表示第二染色体各个窗口中测序数据数目分布的标准差, n_i 表示第一染色体中窗口的数目, n_j 表示第二染色体中窗口的数目。

[0063] 根据本发明的实施例, 预定的阈值的大小, 可以通过经验获得, 或者通过对已知具有非整倍性的有核红细胞样本或者没有非整倍性的有核红细胞进行预先试验所得到的相应 t 检验值作为阈值。优选, 所述预定的第一阈值为 -4 或者更小, 所述第二阈值为 -3.5 或者更大。

[0064] 在本文中所使用的术语“非整倍性”是与染色体的整倍性相对而言的, 其是指在其基因组中缺少或额外增加一条或若干条染色体。通常而言, 正常的细胞中每种染色体会会有两条, 但由于在减数分裂时一对同源染色体不分离或提前分离而形成染色体数目异常的配子, 这类配子彼此结合或同正常配子结合, 会产生各种非整倍体细胞。另外在体细胞分裂时也会产生非整倍体细胞, 如变异率非常高的肿瘤细胞等。

[0065] 2、确定有核红细胞基因组是否存在异常的系统

[0066] 根据本发明又一方面, 本发明提出了一种用于确定基因组是否存在异常的系统 1000。参考图 3, 根据本发明的实施例, 该系统 1000 包括: 有核红细胞分离装置 100、测序装置 200 以及测序结果分析装置 300。根据本发明的实施例, 有核红细胞分离装置 100 用于从孕妇样本中分离胎儿有核红细胞。测序装置 200 用于对有核红细胞基因组的至少一部分进行测序, 以便获得测序结果。测序结果分析装置 300 与测序装置 200 相连, 以便从测序装置 200 接收测序结果, 并且基于所得到的测序结果, 判断所分离的有核红细胞基因组是否存在异常。

[0067] 根据本发明的实施例, 有核红细胞分离装置 100 可以进一步包括单核细胞分离单元 101 和磁性富集单元 102。其中, 根据本发明的实施例, 单核细胞分离单元 101 适于利用密度梯度试剂, 对孕妇样本进行梯度离心, 以便获得单核细胞, 其中, 孕妇样本为孕妇外周血。磁性富集单元 102 与单核细胞分离单元 101 相连, 并且适于利用携带抗体的磁珠, 从单核细胞分离有核红细胞, 其中, 所述抗体特异性识别有核红细胞表面的抗原。由此, 借助该有核红细胞分离装置 100, 可以有效地通过下面的方法分离有核红细胞:

[0068] 首先, 利用密度梯度试剂, 对所述外周血进行梯度离心, 以便获得单核细胞。在获得单核细胞后, 利用携带抗体的磁珠, 从所得到的单核细胞富集有核红细胞, 其中, 磁珠上所携带的抗体特异性识别有核红细胞表面的抗原, 从而有核红细胞会通过抗体与磁珠结合, 接下来可以通过磁性筛选, 获得有核红细胞。根据本发明的具体示例, 上述分离有核红细胞的装置 100 可以适合进行下述操作: 取适量孕妇的外周血, 抗凝剂抗凝, 将血液样本用不含钙离子和镁离子的 0.1M 磷酸盐缓冲液 (PBS) 按比例进行稀释, 将稀释后的样本缓置于密度梯度离心用试剂上, 在室温条件下, 进行密度梯度离心。离心之后可见单核细胞层, 小心吸出该层细胞, 并转移至新的离心管中, 用 3 倍体积的含 1% BSA 的 PBS 缓冲液重悬, 在室温下再离心, 去上清, 将所得到的细胞沉淀再用相同的方法洗 2 次以去除残留的密度梯度液, 最后将细胞沉淀重悬于含 0.1% BSA 的 PBS 中, 吹匀。然后按 20 微升 / 10^6 个细胞的比例加入携

带抗体的磁珠,4℃静置后离心,弃上清,将沉淀重悬于含0.1%BSA的PBS中;组装磁珠分选系统。以500微升含0.1%BSA的PBS缓冲液润洗分选柱,待液体流空后,将分选的细胞上柱,收集流出液,标记为阴性细胞。用含0.1%BSA的PBS润洗管子,待液体流空后再重复2次,用PBS/EDTA/BSA加入分选柱,待液体流空后再重复一次。最后,加入含0.1%BSA的PBS到分选柱中,离开磁场,将液体冲洗入新的离心管内,即得到有核红细胞。关于该分离有核红细胞的方法,前面已经进行了详细描述,不再赘述。

[0069] 另外,根据本发明的实施例,参考图5,该系统还可以进一步包括全基因组测序文库制备装置400。根据本发明的实施例,全基因组测序文库制备装置400与测序装置200相连,并为该测序装置200提供用于测序的全基因组测序文库。其中,参考图6,根据本发明的实例,全基因组测序文库制备装置400可以进一步包括有核红细胞裂解单元401、全基因组扩增单元402和测序文库构建单元403。其中,根据本发明的实施例,有核红细胞裂解单元401与有核红细胞分离装置100相连,接收并且裂解所分离的有核红细胞,以便释放有核红细胞的全基因组。全基因组扩增单元402与有核红细胞裂解单元401相连,用于对有核红细胞的全基因组进行扩增,以便得到经过扩增的全基因组。测序文库构建单元403用于接收经过扩增的全基因组,并且利用经过扩增的全基因组构建全基因组测序文库。该全基因组测序文库制备装置,能够有效地构建有核红细胞的测序文库。这里所使用的术语“相连”应作广义理解,既可以是直接相连,也可以是间接相连,甚至可以使用相同的容器或设备,只要能够实现功能上的衔接即可,例如有核红细胞裂解单元302与全基因组扩增单元303可以在相同的设备中进行,即在实现对有核红细胞裂解之后,在相同的设备或者容器中即可进行全基因组扩增处理,不需要将所释放的全基因组输送至其他的设备或者容器,只要将设备内的条件(包括反应条件和反应体系的组成)转换为适于进行全基因组扩增反应即可,这样即实现了有核红细胞裂解单元302与全基因组扩增单元303在功能上的衔接,也可以认为被术语“相连”所涵盖。本领域技术人员可以根据采用的基因组测序技术的具体方案选择不同的构建全基因组测序文库的方法和设备,关于构建全基因组测序文库的细节,可以参见测序仪器的厂商例如Illumina公司所提供的规程。根据本发明的一个实施例,全基因组扩增单元303包括适于利用OmniPlex WGA方法对所述全基因组进行扩增的装置。由此,能够有效地对全基因组进行扩增,从而进一步提高了确定有核红细胞基因组异常的效率。

[0070] 根据本发明的一个实施例,全基因组测序装置100包括选自illumina-Solexa、ABI-Solid、Roche-454、和单分子测序装置的至少一种。由此,能够利用这些测序装置的高通量、深度测序的特点,进一步提高了确定有核红细胞染色体非整倍性的效率。当然,本领域技术人员能够理解的是,还可以采用其他的测序方法和装置进行全基因组测序,例如第三代测序技术,以及以后可能开发出来的更先进的测序技术。根据本发明的实施例,通过全基因组测序所得到的测序数据的长度不受特别限制。

[0071] 如前所述,根据本发明的实施例,在获得有核红细胞的基因组测序结果之后,可以对染色体的非整倍性进行分析。因而,根据本发明的实施例,测序结果分析装置300可以适于执行下列操作:首先,将已知的第一染色体序列划分为多个窗口,这些多个窗口分别独立地具有预定的长度;接下来,将所述测序结果的测序数据与该已知的第一染色体序列进行比对,以便获得落入每个窗口的测序数据的数目;最后,基于获得落入每个窗口的测序数据的数目,确定第一参数;并且基于所述第一参数,确定所述有核红细胞针对所述第一染色体

是否具有非整倍性。利用该系统1000,能够有效地确定有核红细胞是否具有染色非整倍性。根据本发明的一个实施例,测序结果分析装置300进一步包括序列比对单元(图中未示出)。序列比对单元用于将测序结果与已知基因组序列信息进行比对以便获得所有可比对上参考基因组的测序数据以及获得来自于第一染色体的测序数据。由此,能够有效地确定来自特定染色体的测序数据,从而进一步提高了确定有核红细胞染色体非整倍性的效率。这里所使用的术语“第一染色体”应做广义理解,其可以是指任何期望研究的目的染色体,其数目并不仅限于一条染色体,甚至可以同时将全部染色体进行分析。根据本发明的实施例,第一染色体可以为人类染色中的任意染色体,例如可以为选自人类21号染色体、18号染色体、13号染色体、X染色体和Y染色体的至少一种。由此,能够有效地确定常见的人类染色体疾病,例如可以预测胎儿的遗传性疾病。因而,根据本发明的实施例的确定有核红细胞染色体非整倍性的方法,能够非常有效地应用于体外生殖领域中的植入前筛查(PGS)和植入前诊断(PGD),以及胎儿有核细胞的产前检测等。由此,可以通过简单地提取有核红细胞来快速预测胎儿的染色体是否存在异常,避免胎儿患有严重的遗传性疾病。

[0072] 前面已经详细描述了第一参数等特征,此处不再赘述。需要说明的是:

[0073] 根据本发明的一个实施例,测序结果分析装置300可以进一步包括适于通过下列步骤基于获得落入每个窗口的测序数据的数目,确定第一参数的单元:分别对落入每个窗口的测序数据的数目设置预定的加权系数;以及根据所述加权系数,将所述落入每个窗口的测序数据的数目进行加权平均,以获得所述第一染色体的中位数,所述第一染色体的中位数构成所述第一染色体的第一参数。

[0074] 根据本发明的一个实施例,测序结果分析装置300可以进一步包括适于通过下列进行基于所述第一参数,确定所述有核红细胞针对所述第一染色体是否具有非整倍性的单元:针对第二染色体,进行与所述第一染色体相同的处理,以便获得所述第二染色体的中位数;将所述第一染色体的中位数与所述第二染色体的中位数进行t值检验,以便获得所述第一染色体与所述第二染色体的差异值;将所述差异值与预定的阈值进行比较,如果所述差异值低于所述预定的阈值,则确定针对第一染色体,所述有核红细胞具有非整倍性。

[0075] 根据本发明的一个实施例,测序结果分析装置300可以进一步包括适于利用下列公式对所述第一染色体的中位数与所述第二染色体的中位数进行t值检验的单元,

[0076]
$$T_{i,j} = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}}$$

[0077] 其中, $T_{i,j}$ 表示所述第一染色体与所述第二染色体的差异值, μ_i 表示第一染色体的中位数, μ_j 表示第二染色体的中位数, σ_i 表示第一染色体各个窗口中测序数据数目分布的标准差, σ_j 表示第二染色体各个窗口中测序数据数目分布的标准差, n_i 表示第一染色体中窗口的数目, n_j 表示第二染色体中窗口的数目。

[0078] 另外,根据本发明的实施例,测序结果分析装置300适于确定基因组中是否在预定的区域存在突变。为此,根据本发明的一个实施例,测序结果分析装置300可以进一步包括:核酸序列确定单元、比对单元以及异常确定单元。其中,预定区域核酸序列确定单元适于基于所述测序结果,确定所述有核红细胞的预定区域的核酸序列。比对单元与所述核酸序列确定单元相连,并且适于将所述有核红细胞的预定区域的核酸序列与对照核酸序列进行比

对,优选地,比对单元中存储有对照核酸序列,更优选对照核酸序列为正常人的基因组序列。异常确定单元与所述比对单元相连,并且适于基于所述比对结果,确定所述有核红细胞的预定区域是否存在异常。关于确定预定区域的突变的方法和细节,前面已经进行了详细描述,此处不再赘述。

[0079] 前面在相关方法中所描述的其他特征和优点,也同样适用于该确定有核红细胞是否存在异常的系统,在此不再赘述。

[0080] 3、确定胎儿有核红细胞基因组序列的方法

[0081] 本发明的又一方面,提出了一种确定胎儿有核红细胞基因组序列的方法,其包括以下步骤:

[0082] 首先,从孕妇样本中分离胎儿有核红细胞。在获得胎儿有核红细胞之后,对所分离的有核红细胞基因组的至少一部分进行测序,以便获得测序结果。最后,基于所得到的测序结果,确定胎儿有核红细胞基因组序列。

[0083] 在本文中,所使用的表达方式“基因组序列”应做广义理解,即可以是全基因组的序列,也可以是一部分基因组的序列。前面已经就从孕妇样本中分离胎儿有核红细胞和对有核红细胞基因组的至少一部分进行测序进行了详细描述,在此不再赘述。需要说明的是:

[0084] 根据本发明的一个实施例,所述孕妇样本为孕妇外周血。根据本发明的一个实施例,所述孕妇的孕周为12-20周。根据本发明的一个实施例,从所述孕妇外周血中分离胎儿有核红细胞进一步包括:利用密度梯度试剂,对所述外周血进行梯度离心,以便获得单核细胞;以及利用携带抗体的磁珠,从所述单核细胞分离有核红细胞,其中,所述抗体特异性识别有核红细胞表面的抗原。根据本发明的一个实施例,所述密度梯度试剂是聚蔗糖,任选地,在800X g下进行所述梯度离心30分钟。根据本发明的一个实施例,在获得所述单核细胞之后,在利用携带CD71抗体的磁珠分离所述有核红细胞之前,进一步包括利用含有1%BSA的PBS缓冲液对所述单核细胞进行清洗,以便去除残余的密度梯度试剂,优选所述PBS缓冲液不含钙离子和镁离子。根据本发明的一个实施例,利用含有1%的PBS缓冲液对所述单核细胞进行清洗进一步包括:将所述单核细胞与所述含有1%BSA的PBS缓冲液混合,以便获得含有单核细胞的悬浮液;以及将所述含有单核细胞的悬浮液进行离心,优选在200X g下离心5分钟,弃上清,以便获得经过清洗的有核红细胞。根据本发明的一个实施例,所述抗体为特异性识别CD71的抗体。根据本发明的一个实施例,对一个有核红细胞进行测序。根据本发明的一个实施例,对所述有核红细胞的全基因组进行测序。根据本发明的一个实施例,对所述有核红细胞的全基因组进行测序进一步包括:对所述有核红细胞的全基因组进行扩增得到经过扩增的全基因组;利用所述经过扩增的全基因组构建全基因组测序文库;以及对所述全基因组测序文库进行测序,以便获得由多个测序数据构成的测序结果。根据本发明的一个实施例,通过OmniPlex WGA方法对所述有核红细胞的全基因组进行扩增。根据本发明的一个实施例,利用所述经过扩增的全基因组构建全基因组测序文库进一步包括:将所述经过扩增的全基因组片段化,以便获得DNA片段;将所述DNA片段进行末端修复,以便获得经过末端修复的DNA片段;将所述经过末端修复的DNA片段的3'末端添加碱基A,以便获得具有粘性末端A的DNA片段;将所述具有粘性末端A的DNA片段与接头相连,以便获得连接产物;将所述连接产物进行PCR扩增,以便获得第二扩增产物;以及将所述第二扩增产物进行纯化回收,以便获得回收产物,所述回收产物构成所述全基因组测序文库。根据本发明的一个实施

例,将所述经过扩增的全基因组片段化是通过Covaris打断仪进行的。根据本发明的一个实施例,所述DNA片段的长度为约350bp。根据本发明的一个实施例,将所述DNA片段进行末端修复是利用Klenow片段、T4DNA聚合酶和T4多核苷酸激酶进行的,所述Klenow片段具有5'→3'聚合酶活性和3'→5'外切酶活性,但缺少5'→3'外切酶活性。根据本发明的一个实施例,将所述经过末端修复的DNA片段的3'末端添加碱基A是利用Klenow片段(3'-5' exo-)进行的。根据本发明的一个实施例,将所述具有粘性末端A的DNA片段与接头相连是利用T4DNA连接酶进行的。根据本发明的一个实施例,利用选自Hi seq2000、SOLiD、454、和单分子测序装置的至少一种进行所述测序。关于这些特征的优点,前面已经进行了详细描述,不再赘述。

[0085] 前面在确定有核红细胞染色体基因组是否存在异常的方法中所描述的其他特征和优点,也同样适用于该确定胎儿有核红细胞基因组序列的方法,在此不再赘述。

[0086] 下面通过具体的实施例,对本发明进行说明,需要说明的是这些实施例仅仅是为了说明目的,而不能以任何方式解释成对本发明的限制。

[0087] 实施例1:

[0088] 实验材料

[0089] 外周血样品来自均已具有临床结果的唐氏儿高危孕妇。其他试验材料如未特别说明,均为本领域中常规的方法配置的试剂或者市售可得的试剂。

[0090] 实验流程

[0091] 1、提取有核红细胞

[0092] 取3ml孕妇外周血,选用EDTA作为抗凝剂,将血样本用不含Ca²⁺和Mg²⁺的0.1M磷酸盐缓冲液(PBS)按1:1比例进行稀释,将稀释后的样本缓置于3ml密度为1.077的Ficoll试剂(美国Sigma公司产品)上,在室温条件下,进行密度梯度离心,800g×30min。离心之后可见单核细胞层,小心吸出该层细胞,并转移至新的1.5ml离心管中,用3倍体积的含1%BSA的PBS重悬所得到的单核细胞,并室温下离心200g×5min,去上清,细胞沉淀再用相同的方法清洗2次以去除残留的密度梯度液。最后,将细胞沉淀重悬于300微升含0.1%BSA的PBS中,吹匀。进行细胞计数,然后按20微升/10⁶个细胞的比例加入携带CD71抗体的磁珠(德国美天旎公司),4℃静置15min。300g×10min,离心,弃上清,沉淀重悬于500微升含0.1%BSA的PBS中;组装磁珠分选系统(德国美天旎公司)。用500微升含0.1%BSA的PBS润洗分选柱,待流空后,将分选的细胞上柱,收集流出液,标记为阴性细胞。取500微升含0.1%BSA的PBS润洗管子,待液体流空后再重复2次,取500微升PBS/EDTA/BSA加入分选柱,待液体流空后再重复一次。最后取1ml含0.1%BSA的PBS加入分选柱,将分选柱拿开磁场,冲洗入15ml管内,标记为阳性细胞。将阳性细胞进行离心浓缩,只留最下层约100微升,备用,并标记所得到的有核红细胞为GP9。

[0093] 2、全基因组测序文库的构建

[0094] 将所有的有核红细胞进行全基因组扩增,本实验中选用GenomePlex Single Cell Whole Genome Amplification Kit试剂盒进行全基因组扩增,操作流程按照制造商Sigma公司所提供的说明书进行。将扩增产物进行打断,Covaris打断仪进行,严格按照打断仪所附带的说明书进行,打断主带集中在350bp左右即可,得到打断产物DNA片段。

[0095] 对打断产物进行末端修复和末端添加碱基A,具体过程如下:

[0096] 末端修复反应按照如下体系进行:

[0097]	10×T4 多核苷酸激酶缓冲液	10 μl
	dNTPs(10mM)	4 μl
	T4 DNA 聚合酶	5 μl
	Klenow 片段	1 μl
[0098]	T4 多核苷酸激酶	5 μl
	打断产物 DNA 片段	30μl
	ddH ₂ O 补至 100 μl	

[0099] 在20℃下,反应30分钟后,使用PCR纯化试剂盒(QIAGEN)回收末端修复产物。将所得到的产物最后溶于34μl的EB缓冲液中。

[0100] 末端添加碱基A的反应按照以下体系完成:

	10×Klenow 缓冲液	5μl
	dATP(1mM)	10μl
[0101]	Klenow (3'-5' exo ⁻)	3μl
	DNA	32μl

[0102] 在37℃下,温育30分钟后,经 MinElute[®]PCR纯化试剂盒(QIAGEN)纯化,并将产物溶于12μl的EB中。

[0103] 接头的连接反应如下:

	2x Rapid DNA 连接缓冲液	25μl
	PEI A 接头 oligomix(20uM)	10μl
[0104]	T4 DNA 连接酶	5μl
	末端添加碱基 A 的 DNA	10μl

[0105] 在20℃下反应15分钟后,使用PCR纯化试剂盒(QIAGEN)回收连接产物。将产物最后溶于32μl的EB缓冲液中。

[0106] 在0.2ml的PCR管中配制PCR反应体系:

	样品	10 μl
	Phusion DNA 聚合酶 Mix	25 μl
[0107]	PCR 引物* (10 pmol/μl)	1 μl
	标签 N 引物** (10 pmol/μl)	1 μl
	UltraPure TM 水	13 μl

[0108] *PCR引物序列为

[0109] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

[0110] **标签N引物序列为

[0111] CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

[0112] 反应程序如下:

[0113] 98℃ 30 s

	98°C	10 s	} 10 个循环
	65°C	30 s	
[0114]	72°C	30 s	
	72°C	5 min	
	4°C	保持	

[0115] 使用PCR纯化试剂盒(QIAGEN)回收PCR产物。将样品最后溶于22 μ l的EB缓冲液中。

[0116] 其中标签(index)N为每个文库分别带有唯一的8bp标签序列,将构建好的文库经Agilent[®]Bioanalyzer2100检测,结果示于图8中。如图8所示,所构建的文库片段分布范围符合要求,再经过Q-PCR方法分别对文库进行定量,合格后,将文库混合在一个流动池(flow cell)的同一个泳道(lane)进行上机测序,为节约成本选用单边测序。本实例中测序选择了illumina[®]HiSeq2000[™]测序仪器及方法,其中仪器的参数设置及操作方法都按照illumina[®]制造商所提供的操作手册(可由<http://www.illumina.com/support/documentation.ilmn>获取)严格进行。本实例中所用为HiSeq2000[™]测序仪,测序循环数为PE91index(即双向91bp index测序)。

[0117] 3、数据分析

[0118] 测序得到的测序结果如表1所示,样品GP9测序得到的测序数据(read)总数为13,407,381,可以比对到参考基因组(HG19)中的条数为9,217,701,比率为68.70%,能唯一比对到参考序列上的测序数据条数为7,341,230,唯一比率为80%。

[0119] 表1测序数据

[0120]

样品名称	测序数据总数	可比对的测序数据数目	比率	能唯一比对的测序数据的数目	唯一比率
GP9	13407381	9217701	68.70%	7341230	80%

[0121] 经过对数据的初步分析后,使用SOAP v2.20将测序读长比对到参考序列hg19上,比对方法如下:

[0122] 通过下列方法获得加权系数:

[0123] 首先,分别将人类全基因组各条染色体划分成固定长度60Kb的窗口,记录各个窗口的起始终止位置并统计其GC均值(记为GC_{ref})。得到待测样本的测序数据后,将测序结果序列比对至基因组上,取出完全匹配且仅比对成功一次的测序数据,即唯一比对的测序数据,并获得所有唯一比对的测序数据的位点信息。统计待测样本比对结果对应应在基因组每条染色体各个窗口中的唯一比对的测序数据的数目(记为UR_{sample}),将所得到的UR_{sample}和GC_{ref}进行制图,得到图8A。通过图8A,可以看出,由于测序仪引入的GC偏好导致GC大致在[0.35,0.55]区域的测序数据分布较多。用smooth spline方法,将图8A中离散的点拟合成平滑的曲线,反应GC含量与测序数据数目关系的图,得到图8B,在图8B中,所有的窗口按照GC均值步长为1%划分。由拟合后的数据可得出每个GC均值对应窗口中的唯一比对的测序数据的数目,即M_{fit}。根据公式W_{GC}=M/M_{fit},得到预定的加权系数W_{GC},该加权系数的分布见图8C,其中M是待测样本落在同等GC均值窗口中唯一比对的测序数据的数目。

[0124] 利用所得到的加权系数,对对待测样本落在基因组各条染色体窗口中的唯一比对的测序数据数目(记为 UR_{bin})进行修正,每个窗口对应的 $UR_{fit}=UR_{bin}*W_{GC}$,计算每条染色体修正后唯一比对的测序数据数目的中位数 $UR_i(i=1,2,\dots,22)$,绘制得到箱式图,即图8D。由图8D可以看出,样品GP9中21号染色体对应的唯一比对的测序数据的数目明显高于其它染色体,可以确定21号染色体存在非整倍性。

[0125] 另外,针对常见非整倍性的第13、18和21进行t值检验。这三种比较常见的非整倍体,我们对这三个染色体的数据量按照下面的公式进行t值检验,

[0126]
$$T_{i,j} = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}} \quad (\text{公式 1})$$

[0127] 其中 μ 是待测样本某条染色体 $UR_i(i=13,18,21;j=1,2,\dots,22)$;i和j代表染色体i、j; σ 代表某条染色体UR分布的标准差;n代表的是某条染色体窗口的个数。

[0128] 将染色体i与第1-12号染色体分别进行上述比较,并按照下列公式得到平均差异值(t-value),

[0129]
$$T_i = \frac{1}{12} * \sum_{j=1}^{12} T_{i,j}$$

[0130] 选取第1-12号染色体是因为第1-12号染色体数据波动较小,尽量排除掉我们的关注染色体(包括chr13,chr18和chr21)。本方法设定的阈值(针对常染色体)为 $t\text{-value} \leq -4$,即对应染色体为三倍体高风险; $-4 < t\text{-value} \leq -3.5$,即检测结果不确定,需要重新取样上机或重新建库上机确定检测结果; $t\text{-value} > -3.5$,即检测结果为体低风险。由表2中看出,样品GP9的13、18和21号染色体对应的t-value,21号染色体的t-value已经超过了21-三体高风险标准。

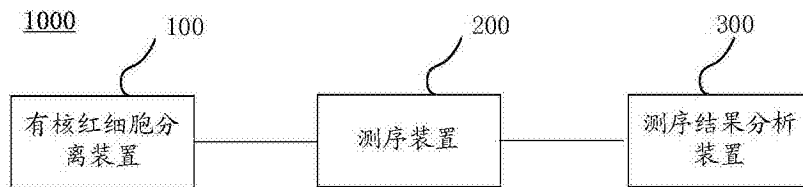
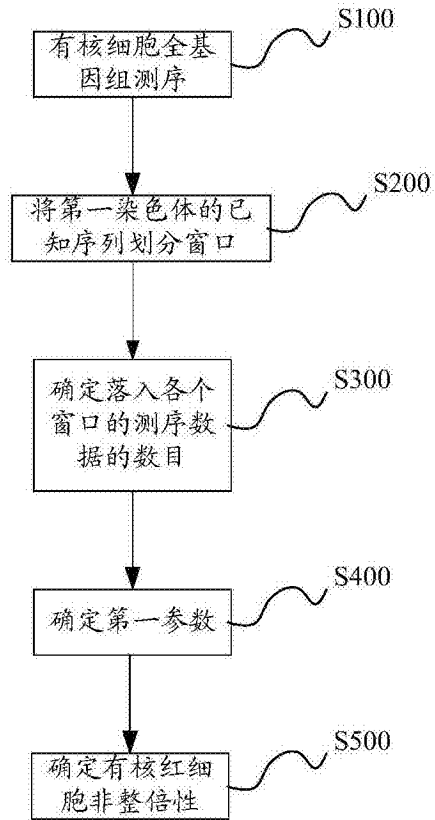
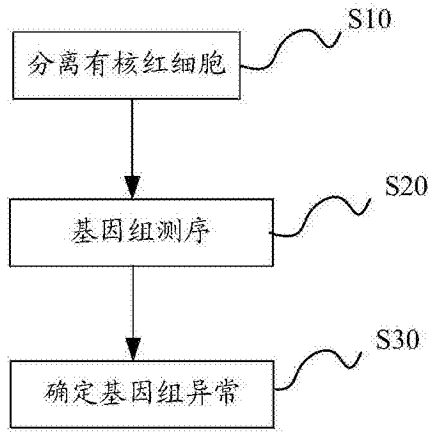
[0131] 表2:t值检验结果

[0132]

样品名称	13号染色体t值	18号染色体t值	21号染色体t值	结果
GP9	0.96287409	-2.206632741	-34.59904702	21-三体

[0133] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

[0134] 尽管已经示出和描述了本发明的实施例,本领域的普通技术人员可以理解:在不脱离本发明的原理和宗旨的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由权利要求及其等同物限定。



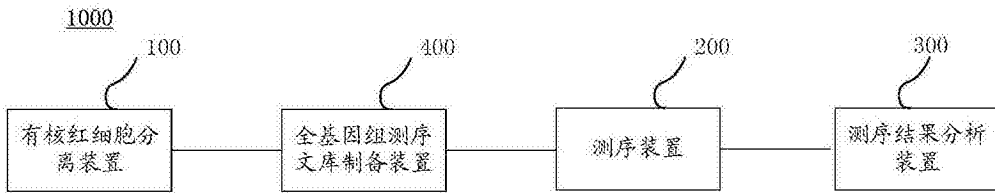


图5

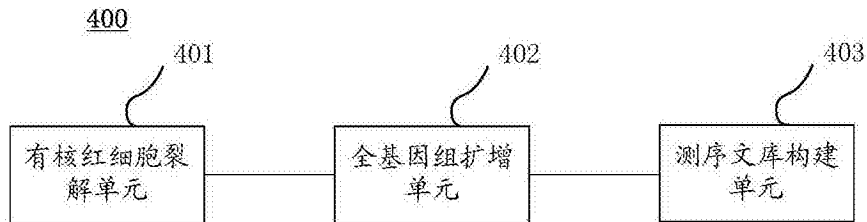
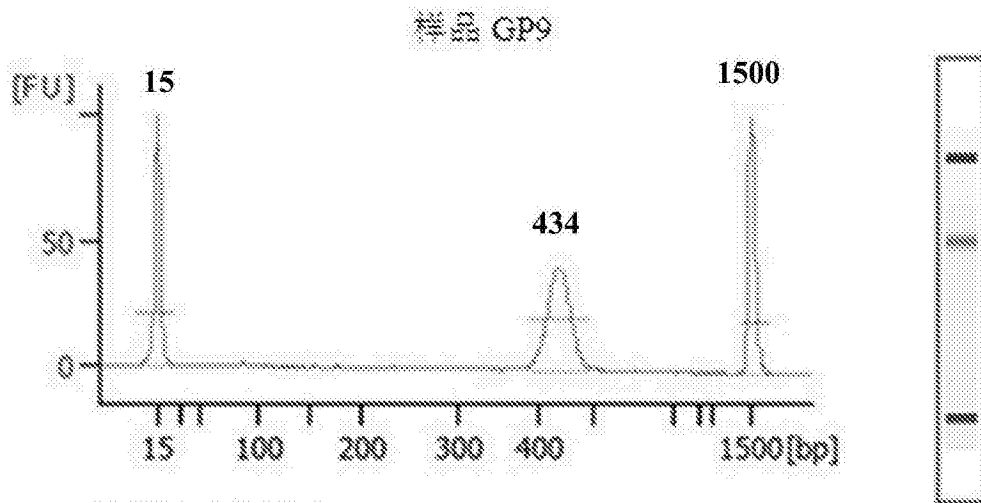


图6



样品的总体结果

找到的峰:

峰值大小	质量浓度	摩尔浓度
434bp	3.6ng μ L	12.5nmol/L

图7

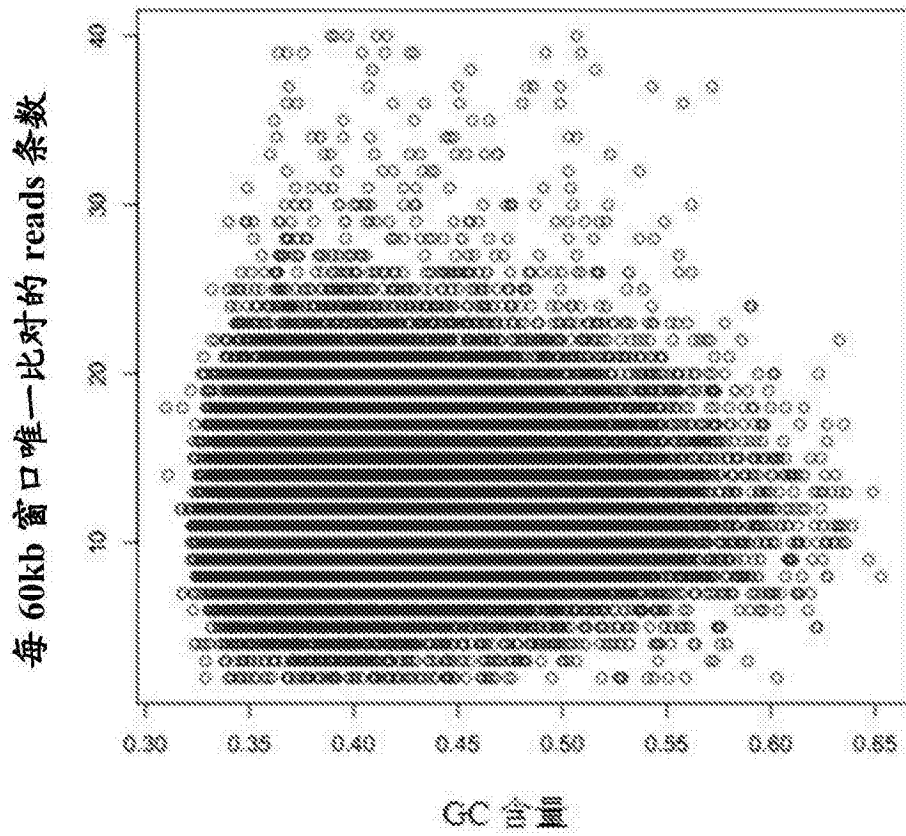


图8A

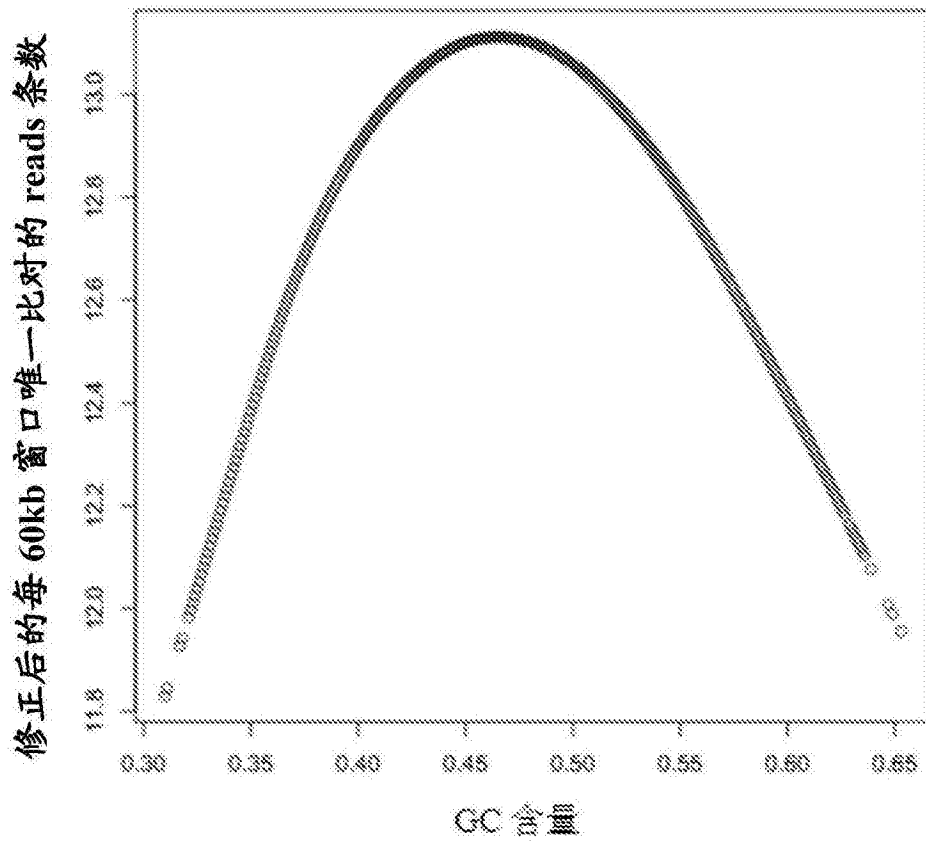


图8B

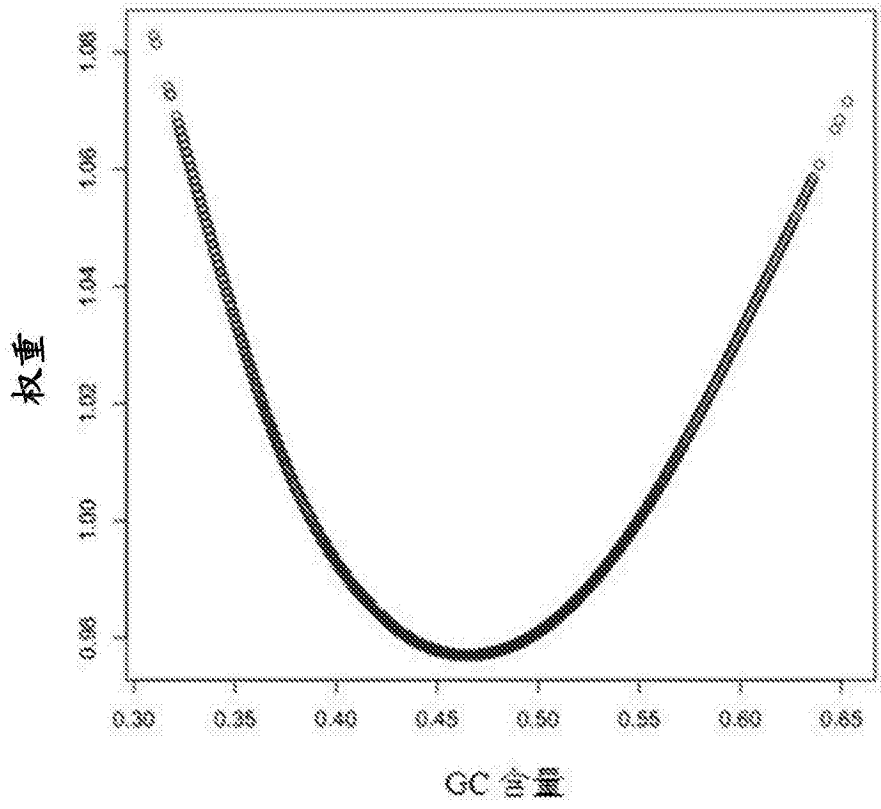


图8C

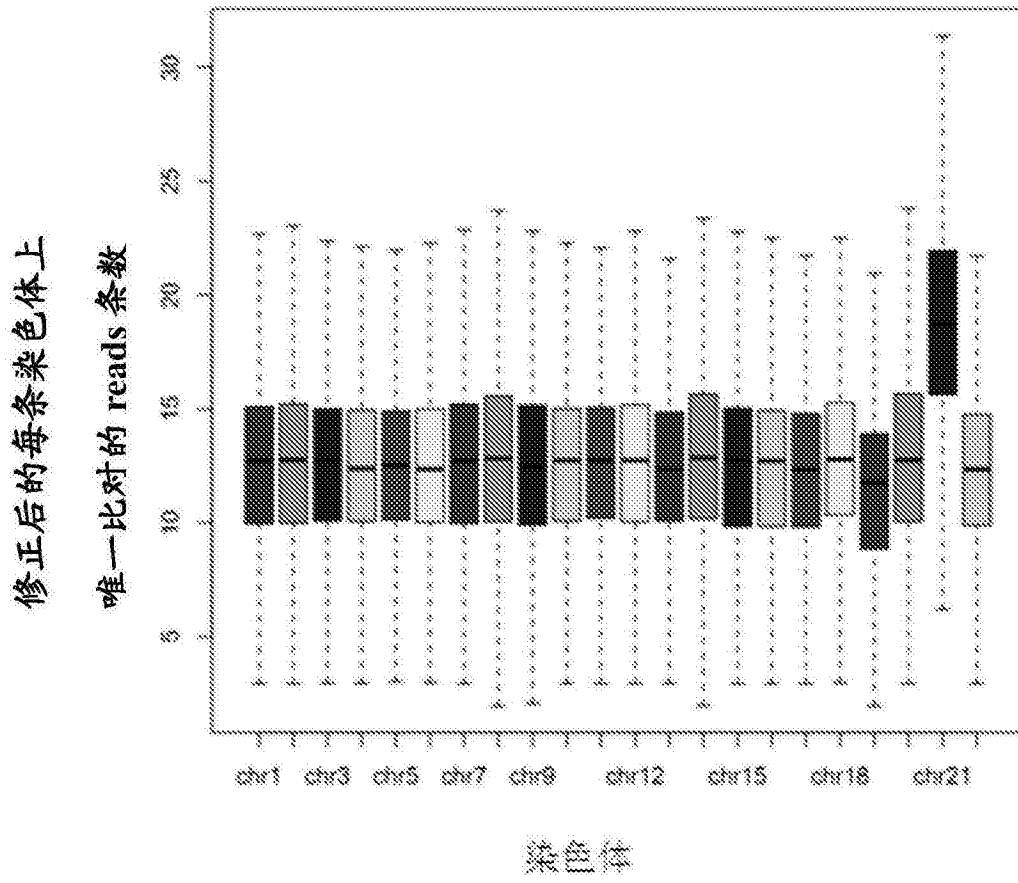


图8D