(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109461473 B
(45)授权公告日 2019.12.17

(21)申请号 201811162012.9

(22)申请日 2018.09.30

(65)同一申请的已公布的文献号
申请公布号 CN 109461473 A

(43)申请公布日 2019.03.12

(73)专利权人 北京优迅医疗器械有限公司
地址 102629 北京市大兴区中关村科技园区大兴生物医药产业基地永大路38号5幢3层303

(72)发明人 关永涛 党明浩 徐寒黎 张静波 方楠 白灵 王建伟 刘倩 唐宇

(74)专利代理机构 北京康信知识产权代理有限责任公司 11240
代理人 吴贵明 路秀丽

(51)Int.Cl.
G16B 30/00(2019.01)

审查员 郭俊

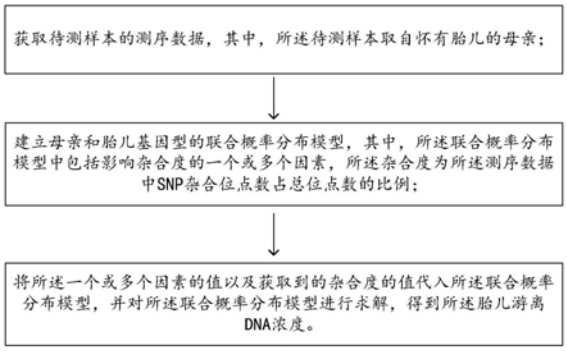
权利要求书2页 说明书9页 附图2页

(54)发明名称

胎儿游离DNA浓度获取方法和装置

(57)摘要

本发明公开了一种胎儿游离DNA浓度获取方法及装置。该方法包括:获取待测样本的测序数据,其中,待测样本取自怀有胎儿的母亲;建立母亲和胎儿基因型的联合概率分布模型,其中,联合概率分布模型中包括影响杂合度的一个或多个因素,杂合度为测序数据中SNP杂合位点数占总位点数的比例;将一个或多个因素的值以及获取到的杂合度的值代入联合概率分布模型,并对联合概率分布模型进行求解,得到胎儿游离DNA浓度。该方法解决了现有技术中胎儿浓度检测成本高的问题。



CN 109461473 B

1.一种胎儿游离DNA浓度获取方法,其特征在于,包括:

获取待测样本的测序数据,其中,所述待测样本取自怀有胎儿的母亲;

建立母亲和胎儿的基因型的联合概率分布模型,其中,所述联合概率分布模型中包括影响杂合度的一个或多个因素,所述一个或多个因素包括以下至少之一:母亲的近交系数、胎儿的近交系数、测序错误率、人群等位基因频率信息,所述杂合度为所述测序数据中SNP杂合位点数占总位点数的比例;

将所述一个或多个因素的值以及获取到的杂合度的值代入所述联合概率分布模型,并对所述联合概率分布模型进行求解,得到所述胎儿游离DNA浓度;

所述联合概率分布模型通过如下公式表示:

MMFF	Prob	f_A
AA + AA	$p^3(1+q/p\ F_1)(1+q/p\ F_2)$	1-e
AB + AB	$pq(1-F_1)(1-F_2)$	1/2
BB + BB	$q^3(1+p/q\ F_1)(1+p/q\ F_2)$	e
AA + AB	$p^2q(1+q/p\ F_1)(1-F_2)$	$(1-h/2)-(1-h)e$
BB + AB	$pq^2(1+p/q\ F_1)(1-F_2)$	$h/2+(1-h)e$
AB + AA	$p^2q(1-F_1)(1+q/p\ F_2)$	$1/2+h/2\ (1-e)$
AB + BB	$pq^2(1-F_1)(q+p/q\ F_2)$	$1/2-h/2\ (1-e)$

,

其中,MMFF列表示的是所述母亲和胎儿的基因型,A和B分别表示一SNP位点上的两种等位基因,Prob列表示的是所述母亲和胎儿的所述基因型的联合概率,p和q分别表示所述等位基因A和B的人群等位基因频率信息,F1表示所述母亲的近交系数,F2表示所述胎儿的近交系数,e表示所述测序错误率,f_A列表示所述测序数据中所述等位基因A的频率,h表示所述胎儿游离DNA浓度。

2.根据权利要求1所述的方法,其特征在于,在将所述一个或多个因素的值和所述杂合度的值代入所述联合概率分布模型之前,获取所述一个或多个因素的值。

3.根据权利要求2所述的方法,其特征在于,在所述一个或多个因素包括所述母亲的近交系数的情况下,所述母亲的近交系数是通过对白细胞低深度测序获取到的。

4.根据权利要求2所述的方法,其特征在于,在所述一个或多个因素包括所述胎儿的近交系数的情况下,通过以下之一得到所述胎儿的近交系数:

将所述胎儿的近交系数设置为0;

通过对所述胎儿的父亲进行白细胞测序得到所述胎儿的近交系数;

将人群近交系数的均值作为所述胎儿的近交系数。

5.根据权利要求2所述的方法,其特征在于,在所述一个或多个因素包括所述人群等位基因频率信息的情况下,通过以下之一得到所述人群等位基因频率信息:

从所述母亲所属人群的数据中获取;

从纳入预定数量的NIPT样本中计算得到。

6.根据权利要求1至5中任一项所述的方法,其特征在于,获取所述待测样本的测序数据包括:

对所述待测样本提取游离DNA并进行测序之后得到原始测序数据;

对所述原始测序数据进行加工得到所述测序数据，所述加工用于将所述原始测序数据处理成适用于得到所述杂合度的测序数据。

7.根据权利要求6所述的方法，其特征在于，对所述原始测序数据进行加工得到所述测序数据包括：

删除低质量的reads；

将删除后被保留的reads对比到参考基因组，得到满足比对策略的reads作为所述测序数据。

8.根据权利要求7所述的方法，其特征在于，

所述低质量的reads包括以下至少之一：PCR扩增引入的重复片段的reads、包含一个以上碱基N的reads、连续5个核苷酸的平均测序质量低于20的reads；和/或，

所述比对策略包括以下之一：允许最多一个错配及只保留唯一比对上的reads。

9.根据权利要求6所述的方法，其特征在于，对所述待测样本提取游离DNA并进行测序包括：

对所述待测样本提取游离DNA并进行全基因组低深度测序。

10.一种胎儿游离DNA浓度获取装置，其特征在于，包括：

所述装置用于存储或者运行模块，或者所述模块为所述装置的组成部分；其中，所述模块为软件模块，所述软件模块为一个或多个，所述软件模块用于执行上述权利要求1至9中任一项所述的方法。

胎儿游离DNA浓度获取方法和装置

技术领域

[0001] 本发明涉及声音领域，具体而言，涉及一种胎儿游离DNA浓度获取方法及装置。

背景技术

[0002] 胎儿游离核酸浓度的定量在无创产前筛查中有重要价值，它决定了NIPT是否有效检出。胎儿核酸浓度定量的重要性体现在：第一，在已知胎儿浓度的情况下，对于胎儿浓度极低的样本(譬如低于3％)，就需要重取样。这能够在很大程度上避免NIPT的假阴性，毕竟胎儿浓度过低是假阴性的主要原因。第二，在已知胎儿浓度的情况下，就可知染色体含量变化的期望值，NIPT筛查的统计功效能得到很大提升。第三，在已知胎儿浓度的情况下，性染色体异常，双胎、嵌合等特殊样本的NIPT也变得更加简单，准确性更高。但是如何对胎儿浓度精准定量仍是待解难题。

[0003] 当前已有的胎儿游离DNA的定量方法有以下几种：

[0004] (1)实时定量PCR技术

[0005] 1998年，香港中文大学的Dennis Lo等用实时定量PCR技术定量分析了孕妇血浆中的胎儿游离DNA，发现它早在妊娠7周可以测得，浓度随着妊娠周数的增加而增加。以实时荧光定量PCR方法为例，设计引物扩增并检测孕妇外周血浆样本中Y性别决定区(SRY)基因。这类方法的依据是SRY基因是男胎的标志基因，母体的cfDNA中不存在该基因。根据标准曲线的绘制，推算每ml样本中SRY基因的拷贝数，从而推断男胎的胎儿浓度。

[0006] (2)全基因组NGS测序，基于性染色体推断胎儿浓度

[0007] 基于新一代高通量测序，NIPT的检测能得到孕妇外周血全基因组的低深度测序数据。通过将测序数据比对到参考基因组上，比对结果进行GC校正等，得到每条染色体的含量的估计值。这类方法的依据是Y染色体的片段只能来源于男胎，胎儿浓度越高则Y染色体的含量越高；同理，男胎少一条X染色体，胎儿浓度越高则X染色体的含量会越低。因此，可通过性染色体的含量来推断男胎的胎儿浓度。

[0008] (3)全基因组NGS测序(PE测序)，基于游离DNA片段长度分布推断胎儿浓度

[0009] 这类方法在测序时必须采用双末端测序法(paired-end sequencing)，从而根据Read1和Read2的比对位置来推断cfDNA片段的长度。这类方法的依据是胎儿cfDNA长度分布与母体cfDNA有所不同，研究显示，血浆内主要的cfDNA长度为166bp，存在以10bp为单位的递减规律，并在143bp处也有明显存在。胎儿浓度越高，孕母外周血中：以143bp为峰值的cfDNA显著增加，同时以166bp为峰值的cfDNA的则显著降低。因而可根据孕母外周血浆中cfDNA片段长度的分布来推断胎儿浓度。

[0010] (4)深度靶向的NGS测序法，对若干个SNP位点进行高深度测序

[0011] 这类方法可以采用深度靶向的NGS测序法，对孕妇外周血全基因组的若干SNP位点进行高深度测序，将该位点的孕妇外周血中的cfDNA看成复合基因型(AAAA，AAAB，ABAA，ABAB，每组前两个字母代表母亲基因型，后两个代表胎儿基因型)，直接根据测序数据中杂合比的数值来估算胎儿cfDNA浓度。

[0012] (5)基于甲基化标记的方法

[0013] 这类方法的依据是胎儿DNA甲基化与母亲DNA甲基化程度不同,利用甲基化测序区分胎儿和母亲来源的cfDNA,从而推断胎儿游离核酸浓度。

[0014] 胎儿浓度的准确定量一直是技术难点,存在多方面的困难。传统的基于性染色体的胎儿浓度定量方法,弊端在于无法对女胎的胎儿浓度进行定量。基于胎儿和母体cfDNA片段长度差异的胎儿浓度定量方法,需要双端测序,增加测序成本且准确性不高。基于SNP位点的等位基因频率的胎儿浓度定量方法,需要高深度测序,目前NIPT的0.1X低深度测序无法满足要求。基于甲基化的胎儿浓度定量的实验处理步骤繁琐,测序成本较高。

[0015] 由此可见,现有方法均存在一定的缺陷,主要有以下几方面:增加额外的实验工作;对仪器和设备有额外需求;受限于男胎的检测;检测准确性不够理想;检测成本较高。

[0016] 对于现有技术中的问题,目前没有提出相应的解决方案。

发明内容

[0017] 本发明实施例提供了一种胎儿游离DNA浓度获取方法及装置,以解决现有技术中胎儿浓度检测成本高的问题。

[0018] 根据本发明实施例的一个方面,提供了一种胎儿游离DNA浓度获取方法,该方法包括:获取待测样本的测序数据,其中,待测样本取自怀有胎儿的母亲;建立母亲和胎儿基因型的联合概率分布模型,其中,联合概率分布模型中包括影响杂合度的一个或多个因素,杂合度为测序数据中SNP杂合位点数占总位点数的比例;将一个或多个因素的值以及获取到的杂合度的值代入联合概率分布模型,并对联合概率分布模型进行求解,得到胎儿游离DNA浓度。

[0019] 进一步地,在一个或多个因素包括以下至少之一的情况下:母亲的近交系数、胎儿的近交系数、测序错误率、人群等位基因频率信息,其中,在将一个或多个因素的值和杂合度的值代入联合概率分布模型之前,获取一个或多个因素的值。

[0020] 进一步地,在一个或多个因素包括母亲的近交系数的情况下,母亲的近交系数是通过对白细胞低深度测序获取到的。

[0021] 进一步地,在一个或多个因素包括胎儿的近交系数的情况下,通过以下之一得到胎儿的近交系数:将胎儿的近交系数设置为0;通过对胎儿的父亲进行白细胞测序得到胎儿的近交系数;将人群近交系数的均值作为胎儿的近交系数。

[0022] 进一步地,在一个或多个因素包括人群等位基因频率信息的情况下,通过以下之一得到人群等位基因频率信息:从母亲所属人群的数据中获取;从纳入预定数量的NIPT样本中计算得到。

[0023] 进一步地,获取待测样本的测序数据包括:对待测样本提取游离DNA并进行测序之后得到原始测序数据;对原始测序数据进行加工得到测序数据,加工用于将原始测序数据处理成适用于得到杂合度的测序数据。

[0024] 进一步地,对原始测序数据进行加工得到测序数据包括:删除低质量的reads;将删除后被保留的reads对比到参考基因组,得到满足比对策略的reads作为测序数据。

[0025] 进一步地,低质量的reads包括以下至少之一:PCR扩增引入的重复片段的reads、包含一个以上碱基N的reads、连续5个核苷酸的平均测序质量低于20的reads;和/或,比对

策略包括以下之一:允许最多一个错配及只保留唯一比对上的reads。

[0026] 进一步地,对待测样本提取游离DNA并进行测序包括:对待测样本提取游离DNA并进行全基因组低深度测序。

[0027] 进一步地,通过如下公式表示联合概率分布模型:

[0028]

MMFF	Prob	f_A
AA + AA	$p^3(1+q/p\,F_1)(1+q/p\,F_2)$	1-e
AB + AB	$pq(1-F_1)(1-F_2)$	1/2
BB + BB	$q^3(1+p/q\,F_1)(1+p/q\,F_2)$	e
AA + AB	$p^2q(1+q/p\,F_1)(1-F_2)$	$(1-h/2)-(1-h)e$
BB + AB	$pq^2(1+p/q\,F_1)(1-F_2)$	$h/2+(1-h)e$
AB + AA	$p^2q(1-F_1)(1+q/p\,F_2)$	$1/2+h/2\,(1-e)$
AB + BB	$pq^2(1-F_1)(q+p/q\,F_2)$	$1/2-h/2\,(1-e)$

其中,MMFF列表示的是母亲和胎儿的基因型,A和B分别表示一SNP位点上的两种等位基因,Prob列表示的是母亲和胎儿的基因型的联合概率,p和q分别表示等位基因A和B的人群等位基因频率信息,F1表示母亲的近交系数,F2表示胎儿的近交系数,e表示测序错误率,f_A列表示测序数据中等位基因A的频率,h表示胎儿游离DNA浓度。

[0029] 根据本发明实施例的另一个方面,还提供了一种胎儿游离DNA浓度获取装置,包括:装置用于存储或者运行模块,或者模块为装置的组成部分;其中,模块为软件模块,软件模块为一个或多个,软件模块用于执行上述任一种方法。

[0030] 在本发明实施例中,提供的胎儿游离DNA浓度获取方法,通过建立母亲和胎儿基因型的联合概率分布模型,并利用该模型中的各因素的值及这些因素所影响的杂合度的值进行求解,即可获得胎儿游离DNA浓度。该方法可以利用NIPT常规的NGS低深度测序数据,在不增加任何额外的实验和测序的成本的基础上,不仅能够实现对胎儿浓度的定量检测,而且该方法成本低,准确性高,还适用于女胎胎儿浓度检测。

附图说明

[0031] 此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。在附图中:

[0032] 图1是根据本发明实施例的胎儿游离DNA浓度获取方法的流程图;

[0033] 图2是根据本发明实施例1的基于模拟混样数据实际获得的胎儿浓度与预期相比较的结果图;

[0034] 图3是根据本发明实施例2的基于真实混样样本获得的胎儿浓度与混样浓度相比较的结果图;

[0035] 图4是根据本发明实施例3的基于真实男胚NIPT样本获得的胎儿浓度与性染色体推断出的浓度相比较的结果图。

具体实施方式

[0036] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是

本发明一部分的实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本发明保护的范围。

[0037] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0038] 在本实施例中，提供了一种胎儿游离DNA浓度获取方法，如图1所示，该方法包括：获取待测样本的测序数据，其中，待测样本取自怀有胎儿的母亲；建立母亲和胎儿基因型的联合概率分布模型，其中，联合概率分布模型中包括影响杂合度的一个或多个因素，杂合度为测序数据中SNP杂合位点数占总位点数的比例；将一个或多个因素的值以及获取到的杂合度的值代入联合概率分布模型，并对联合概率分布模型进行求解，得到胎儿游离DNA浓度。

[0039] 上述胎儿游离DNA浓度获取方法，通过建立母亲和胎儿基因型的联合概率分布模型，并利用该模型中的各因素的值及这些因素所影响的杂合度的值进行求解，即可获得胎儿游离DNA浓度。该方法可以利用NIPT常规的NGS低深度测序数据，在不增加任何额外的实验和测序的成本的基础上，不仅能够实现对胎儿浓度的定量检测，而且该方法成本低，准确性高，还适用于女胎胎儿浓度检测。

[0040] 在一种优选实施例中，在一个或多个因素包括以下至少之一的情况下：母亲的近交系数F1、胎儿的近交系数F2、测序错误率e、人群等位基因频率信息，将一个或多个因素的值和杂合度的值代入联合概率分布模型之前，上述方法还包括：获取一个或多个因素的值。

[0041] 实际应用中，根据测序数据来源的不同，影响杂合度的上述因素的数量不等，其各因素的值也不相同。比如，测序质量很高的情况下，测序错误率e通常在0.001左右。人群等位基因频率信息根据人群的不同而不同，比如，从东亚人群中获得的人群等位基因频率信息与从欧美人群中获得的人群等位基因频率信息是不同的。母亲的近交系数F1和胎儿的近交系数F2都对测序数据中SNP杂合位点的统计有影响。近交系数越高，胎儿出现杂合位点的概率就高，近交系数越低，胎儿出现杂合位点的概率就低。

[0042] 在一种优选实施例中，在一个或多个因素包括母亲的近交系数F1的情况下，母亲的近交系数F1是通过对白细胞低深度(0.1x～0.5x)测序获取到的。具体地，通过对白细胞低深度测序建立类似于本申请的模型，令其中胎儿浓度h为0即可获取到。

[0043] 在一种优选实施例中，在一个或多个因素包括胎儿的近交系数F2的情况下，通过以下之一得到胎儿的近交系数F2：将胎儿的近交系数F2设置为0；通过对胎儿的父亲进行白细胞测序得到胎儿的近交系数F2；将人群近交系数的均值作为胎儿的近交系数F2。

[0044] 胎儿的近交系数F2理论上受母亲和父亲的影响，因而理论上需要对父亲的白细胞进行测序得到，但本申请的发明人发现，将胎儿的近交系数F2设置为0或者取人群近交系数的均值就足以获得胎儿游离DNA浓度了，因为胎儿游离DNA浓度一般在10％左右。

[0045] 在一种优选实施例中，在一个或多个因素包括人群等位基因频率信息的情况下，通过以下之一得到人群等位基因频率信息：从母亲所属人群的数据中获取；从纳入预定数量的NIPT样本中计算得到。

[0046] 从母亲所属人群的数据中获取，比如母亲属于东亚人，则可以从1000genome(千人基因组)的东亚人群数据中获取。从纳入预定数量的NIPT样本中计算得到，比如可以从大量真实的NIPT样本计算得到，该样本的具体数量可以是几千或几万。

[0047] 上述方法中，获得待测样本的测序数据的步骤采用现有的步骤即可。在一种优选实施例中，获取待测样本的测序数据包括：对待测样本提取游离DNA并进行测序之后得到原始测序数据；对原始测序数据进行加工得到测序数据，加工用于将原始测序数据处理成适用于得到杂合度的测序数据。

[0048] 具体加工的方式与现有的原始测序数据的加工方式类似，都包括对原始数据进行过滤得到测序数据的步骤。即从raw data处理为clean data。在一种优选实施例中，对原始测序数据进行加工得到测序数据包括：删除低质量的reads；将删除后被保留的reads对比到参考基因组，得到满足比对策略的reads作为测序数据。

[0049] 此处的低质量与常规高通量测序领域的低质量的涵义相同，广义上指无法进行有效的数据处理或者明显对处理结果有不利影响的数据。在一种优选实施例中，低质量的reads包括以下至少之一：PCR扩增引入的重复片段的reads、包含一个以上碱基N的reads、连续5个核苷酸的平均测序质量低于20的reads；和/或，比对策略包括以下之一：允许最多一个错配及只保留唯一比对上的reads。

[0050] 上述优选实施例中，碱基N表示测序的原始数据中会有无法测出来的碱基，用N来表示。现有多种软件可以检测测序中碱基的测序质量，因而能够很方便地将连续5个核苷酸的平均测序质量低于20的reads筛选出来。

[0051] 比对策略中，仅允许最多一个错配以确保用于后续处理的测序数据的质量较高，更倾向于是真实的碱基类型，而非测序错误导致，进而有助于使胎儿游离DNA浓度更准确。只保留唯一比对上的reads是指最终用于后续分析的数据是能够完全与参考基因组比对上的reads，以确保所检测到各SNP位点的碱基类型是真实的。具体比对后的数据的量不限，可根据样本来源的不同进行合理设置。优选加工后得到的测序数据至少有4M的reads数。

[0052] 上述对待测样本提取游离DNA并进行测序采用现有常规的测序即可，无需高深度测序，也无需进行双端测序，只需按照目前NIPT的0.1x的低深度测序即可满足要求。当然，如果测序是进行高深度测序，同样可以满足要求。在一种优选实施例中，对待测样本提取游离DNA并进行测序包括：对待测样本提取游离DNA并进行全基因组低深度测序。此处的低深度测序使目标覆盖度在0.1x～0.5x即可。

[0053] 上述方法中，建立母亲和胎儿基因型的联合概率分布模型的理论基础在于：即便是对于NIPT这样低深度测序的数据，存在足够多的1000genome SNP位点被1条以上的read覆盖，并且这些1000genome SNP位点的覆盖度服从Poisson分布。

[0054] 对于任何覆盖度大于1的SNP位点，都可以定义该位点为纯合或杂合。

[0055] 杂合位点占总位点的百分比与胎儿浓度h之间存在函数关系。因为胎儿的存在会引入父源DNA，使得样本中某些纯合位点变成了杂合位点。由于是低深度测序，杂合能够被测到的概率与胎儿浓度有关。对于同一个母体背景而言，胎儿浓度越大，测得的杂合位点的比例就越高。因此可用杂合位点占总位点的百分比来推断胎儿浓度h。

[0056] 在最理想的条件下，假定母亲和胎儿的近交系数(inbreeding coefficient)都为0，测序平台的测序错误率也为0，群体等位基因频率服从均一分布，则能够得到母亲和胎儿

基因型的联合概率模型，如下表1。

[0057] 表1：

[0058]

MMFF	Prob	f_A
AA + AA	p^3	1
AB + AB	$p(1-p)$	1/2
BB + BB	$(1-p)^3$	0
AA + AB	$p^2(1-p)$	$1-h/2$
BB + AB	$p(1-p)^2$	$h/2$
AB + AA	$p^2(1-p)$	$1/2+h/2$
AB + BB	$p(1-p)^2$	$1/2-h/2$

[0059] 上表1中，MMFF表示母亲和胎儿的基因型，A和B表示某一SNP位点的等位基因，Prob列表示为对应的母亲和胎儿的基因型的概率，f_A表示测序数据中等位基因A的频率。

[0060] 如果某些测序位点的覆盖度为2，且群体等位基因频率为p的一类位点上，杂合位点占该类位点的百分为：

[0061] $P_H=(1+h-h^2)p(1-p)$

[0062] 根据p～uniform(0,1)，对P_H做积分运算。在测序数据中所有等位基因频率下，杂合位点占总位点的百分比为：$\frac{1}{6}(1+h-h^2)$。

[0063] 而在实际应用中，有三个因素会影响杂合程度：胎儿的近交系数F2，母亲的近交系数F1，测序错误率e。

[0064] 对于两等位基因的SNP，近交系数F会直接影响纯合AA，BB，以及杂合AB的频率，如下：

[0065] AA～p^2+pqF，AB～$2pq(1-F)$ BB～q^2+pqF

[0066] 因此，在一种优选实施例中，联合概率分布模型为下表2。

[0067] 表2：

[0068]

MMFF	Prob	f_A
AA + AA	$p^3(1+q/p\,F_1)(1+q/p\,F_2)$	1-e
AB + AB	$pq(1-F_1)(1-F_2)$	1/2
BB + BB	$q^3(1+p/q\,F_1)(1+p/q\,F_2)$	e
AA + AB	$p^2q(1+q/p\,F_1)(1-F_2)$	$(1-h/2)-(1-h)e$
BB + AB	$pq^2(1+p/q\,F_1)(1-F_2)$	$h/2+(1-h)e$
AB + AA	$p^2q(1-F_1)(1+q/p\,F_2)$	$1/2+h/2\,(1-e)$
AB + BB	$pq^2(1-F_1)(q+p/q\,F_2)$	$1/2-h/2\,(1-e)$

，

[0069] 其中，MMFF列表示的是母亲和胎儿的基因型，A和B分别表示一SNP位点上的两种等位基因，Prob列表示的是母亲和胎儿的基因型的联合概率，p和q分别表示等位基因A和B的人群等位基因频率信息，F1表示母亲的近交系数，F2表示胎儿的近交系数，e表示测序错误率，f_A列表示测序数据中等位基因A的频率，h表示胎儿游离DNA浓度。

[0070] 该模型可用极大似然法求解h。其求解的前提是需要知道F1、F2、e以及人群等位基因频率信息，其中，母亲的近交系数F1，可以通过白细胞低深度测序得到，该模型可以看作

是常规模型在h＝0时的特殊情况。平台的测序错误率e可以直接从数据中得到。胎儿的近交系数F2，虽然理论上需要对父亲的白细胞测序，但是实际操作中令F2＝0或者取人群近交系数的均值就已经足够满足要求，因为胎儿浓度一般在10％左右。人群等位基因频率信息，可以直接从1000genome的东亚人群数据获取，也可以纳入大量真实NIPT样本来计算得到。

[0071] 基于比对后的数据，通过统计常染色体上大量SNP位点上(深度为2或者3)的杂合和纯合的情况，结合母体自身的近交系数，从千人基因组数据得到的大量SNP位点的人群频率，代入实际模型中，即可求解出胎儿游离核酸浓度h。

[0072] 在一实施例中，对应于上述方式，还提供了一种胎儿游离DNA浓度获取装置，包括：装置用于存储或者运行模块，或者模块为装置的组成部分；其中，模块为软件模块，软件模块为一个或多个，软件模块用于执行上述任一种胎儿游离DNA浓度获取方法。

[0073] 通过上述胎儿游离DNA浓度获取装置，在不增加任何额外的实验和测序成本的基础上，实现了对胎儿游离DNA浓度的定量，且该方法成本低、准确性高，且适用于女胎胎儿浓度检测。

[0074] 本申请中所说的低深度测序是指整个样本的覆盖度的0.1x～0.5x。而覆盖度为2或3是指其中某些位点的深度。比如，1个样本中有30亿个位点，有些位点的深度为0，有些位点的深度为1，有些位点的深度为2，其他位点类似深度也可能存在一定差异，但平均起来，整体样本的深度是0.1x～0.5x。

[0075] 下面结合可选的实施例进行说明。

[0076] 实施例1模拟混样数据验证

[0077] 选取来自1000genome中NA12892(母亲)和NA12878(女儿)的全基因组测序数据，按照不同梯度(分别是2％，4％，6％，8％，10％，12％，14％，16％，18％，20％)的胎儿浓度来混reads，覆盖度最高达到0.5X。

[0078] 母亲和女儿的近交系数通过母本和女儿各自的全基因组测序reads获得，测序错误率通过混合后得到的样本reads计算得到，各SNP位点的人群等位基因频率通过东亚1000genome的东亚人群数据获取，杂合位点占总位点的百分比通过统计混合后得到的样本的reads得到，然后将上述各参数代入前述联合概率分布模型中进行求解，即可获得胎儿游离DNA浓度h。

[0079] 将推断出的胎儿浓度与预期相比较，比较结果如下图2。从图2中可以看出：采用本申请的方法获取的胎儿浓度与预期的胎儿浓度(混reads的比例)一致。

[0080] 实施例2真实混样样本

[0081] 将分别来源于母亲和胎儿的DNA按照不同的胎儿浓度进行混合(胎儿浓度分别为3％、5％、8％和12％)，然后上机测序，该测序为低深度全基因组测序，进而利用本申请所提出的方法推断胎儿浓度。

[0082] 具体的测序深度是0.1x，测序错误率为1/1000，母亲和胎儿的近交系数分别通过各自的DNA测序数据计算得到，各位点的人群等位基因频率通过东亚1000genome的东亚人群数据获取，各混样浓度的测序数据中杂合位点数占总位点数的百分比分通过测序数据获得。

[0083] 将推断出的胎儿浓度与混样浓度相比较，比较结果见图3。从图3可以看出：该方法获得的胎儿浓度与混样的胎儿浓度一致。

[0084] 实施例3真实NIPT男胎样本验证

[0085] 选取怀有男胎的NIPT真实样本40例，采用本申请的方法获取的胎儿浓度。将推断出的胎儿浓度与性染色体推断出的相比较。比较结果见图4，从图4中可以看出：该方法与基于性染色体的推断方法得到的胎儿浓度高度一致。

[0086] 从上述实施例可以看出，本申请的方案具有以下优点：

[0087] 1)准确性高，经用3万多例男胎NIPT样本验证，该方法与基于性染色体的推断方法得到的胎儿浓度高度一致，R^2达到99％。

[0088] 2)适用于女胎，克服了女胎的胎儿浓度难以准确定量的难题。

[0089] 3)不依赖额外的实验步骤和仪器，不需要定制Panel，不需要甲基化测序，不增加任何额外的实验工作，也不依赖额外的实验仪器或平台。

[0090] 4)成本低廉，临床推广价值大。本申请方法基于全基因组低深度测序，可直接使用现有的NIPT样本数据。不需要双端测序，不需要高深度测序(此方法的胎儿浓度获得直接依赖于深度测序得到的某些杂合SNP点的两个等位基因的测序深度的微小差异，需要对每个杂合位点作定量分析；而本申请是统计所有杂合SNP位点占总位点数的比例，只需要粗略对位点作杂合和纯合的定性)，不增加额外测序成本。

[0091] 5)可直接整合入NIPT流程，基于NIPT的数据，因此可以方便地整合进NIPT的分析流程中，提高NIPT筛查的统计功效。

[0092] 需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明并不受所描述的动作顺序的限制，因为依据本发明，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本发明所必须的。

[0093] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中，包括若干指令用以使得计算设备执行本发明各个实施例所述的方法，或者是使得处理器来执行本发明各个实施例所述的方法。

[0094] 上述本发明实施例序号仅仅为了描述，不代表实施例的优劣。

[0095] 在本发明的上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。

[0096] 在本申请所提供的几个实施例中，应该理解到，所揭露的技术内容，可通过其它的方式实现。其中，以上所描述的装置实施例仅仅是示意性的，例如所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，单元或模块的间接耦合或通信连接，可以是电性或其它的形式。

[0097] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个

网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0098] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0099] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算设备(可为个人计算机、服务器或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0100] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

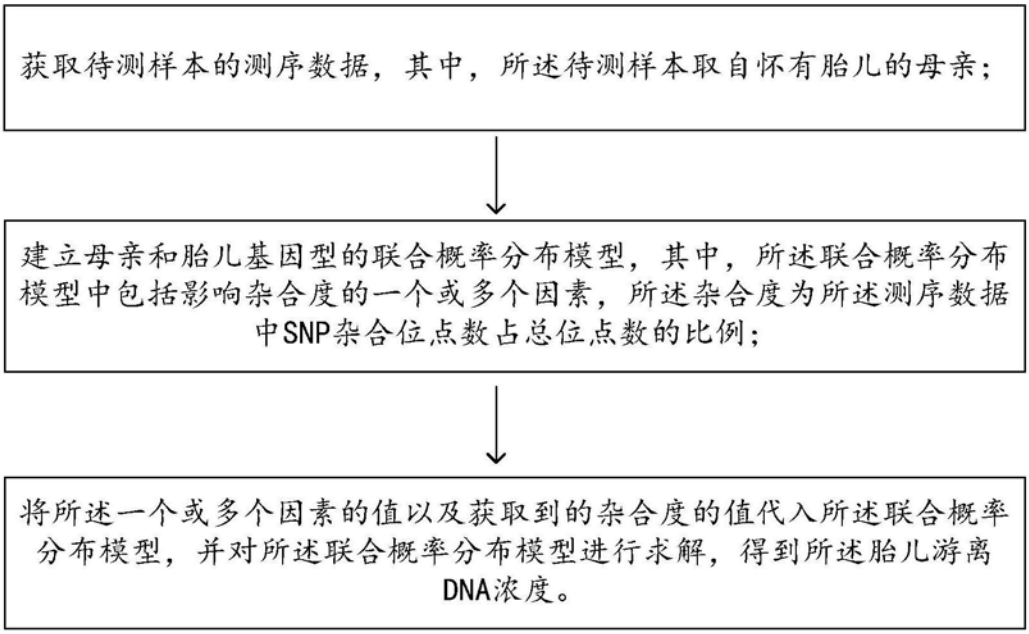

获取待测样本的测序数据，其中，所述待测样本取自怀有胎儿的母亲；
建立母亲和胎儿基因型的联合概率分布模型，其中，所述联合概率分布模型中包括影响杂合度的一个或多个因素，所述杂合度为所述测序数据中SNP杂合位点数占总位点数的比例；
将所述一个或多个因素的值以及获取到的杂合度的值代入所述联合概率分布模型，并对所述联合概率分布模型进行求解，得到所述胎儿游离DNA浓度。

图1

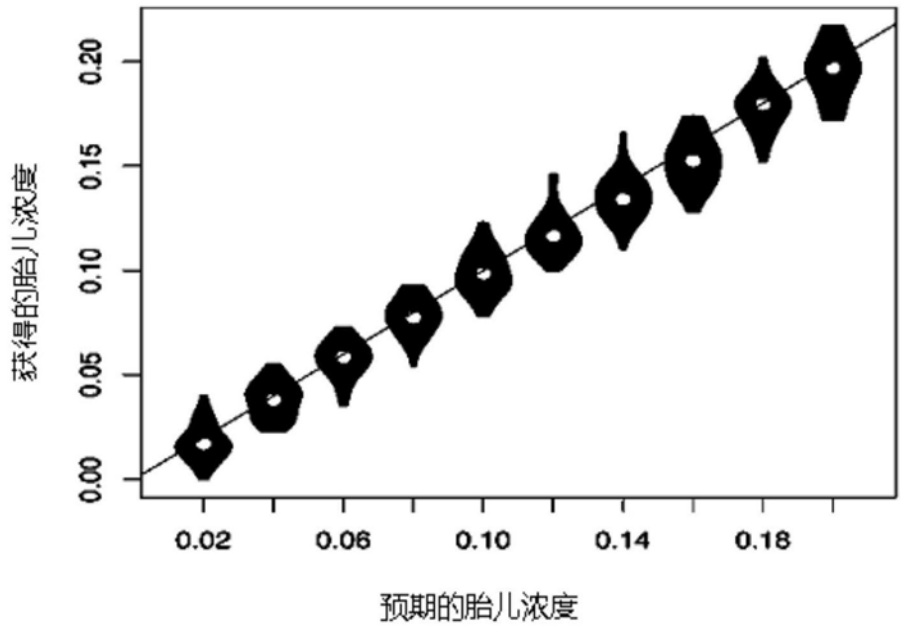

获得的胎儿浓度
0.00
0.05
0.10
0.15
0.20
0.02
0.06
0.10
0.14
0.18
预期的胎儿浓度

图2

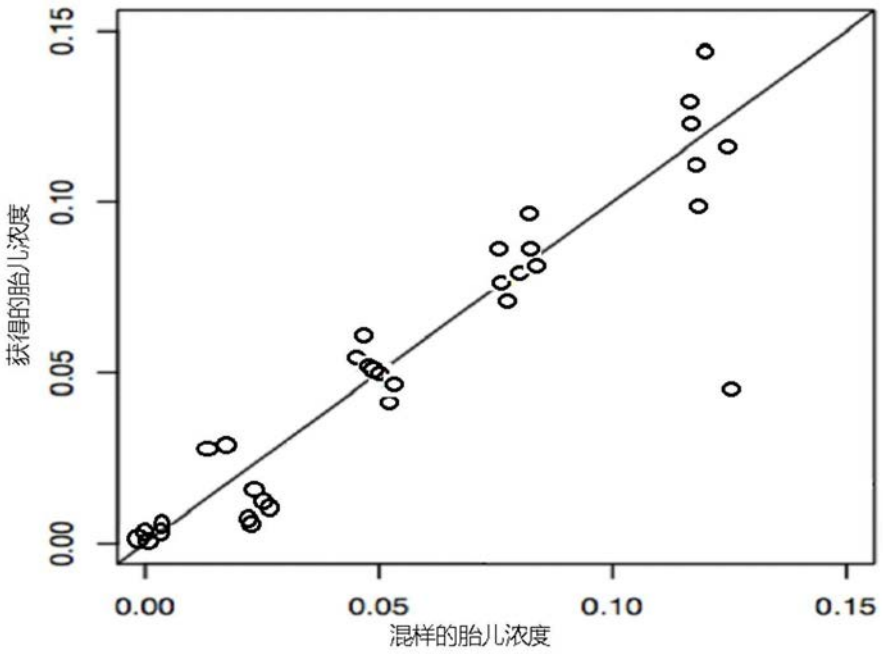

获得的胎儿浓度
0.00
0.05
0.10
0.15
0.00
0.05
0.10
0.15
混样的胎儿浓度

图3

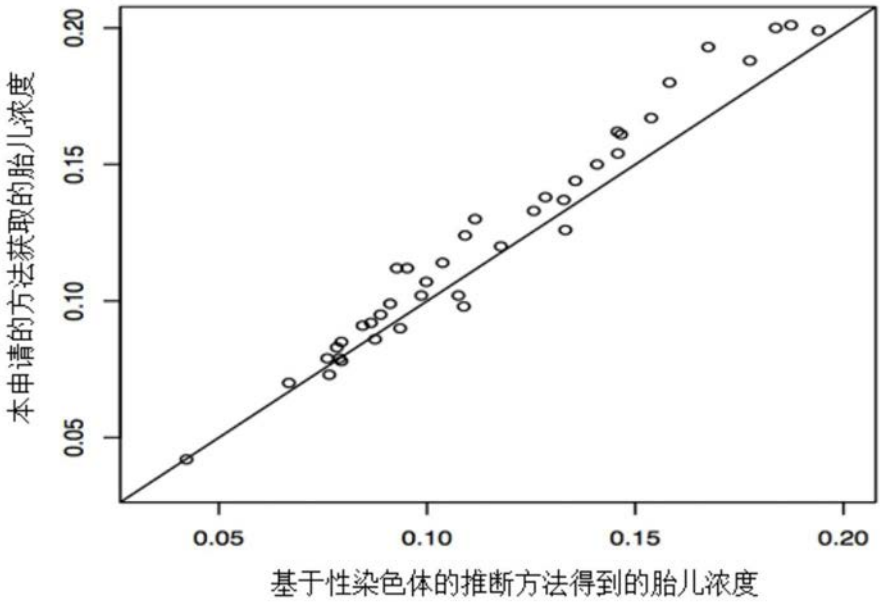

本申请的方法获取的胎儿浓度
0.05
0.10
0.15
0.20
0.05
0.10
0.15
0.20
基于性染色体的推断方法得到的胎儿浓度

图4