

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 7 月 5 日 (05.07.2018)



(10) 国际公布号
WO 2018/121468 A1

(51) 国际专利分类号:
C12Q 1/68 (2018.01)

(21) 国际申请号: PCT/CN2017/118213

(22) 国际申请日: 2017 年 12 月 25 日 (25.12.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201611270836.9 2016 年 12 月 29 日 (29.12.2016) CN

(71) 申请人: 安诺优达基因科技(北京)有限公司 (ANNOROAD GENE TECHNOLOGY (BEIJING) CO., LTD) [CN/CN]; 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。

(72) 发明人: 杜洋(DU, Yang); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。 彭胜斌(PENG, Shengbin); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。 惠峰(HUI, Feng); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。 张晗(ZHANG, Han); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。 玄兆伶(XUAN, Zhaoling); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。 李大为(LI, Dawei); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing 100176 (CN)。 梁峻彬(LIANG, Junbin); 中国北京市经济技术开发区科创六街 88 号院 8 号楼 2 单元 701 室, Beijing

(54) Title: METHOD, DEVICE AND KIT FOR DETECTING FETAL GENETIC MUTATION

(54) 发明名称: 检测胎儿基因突变的方法、装置和试剂盒

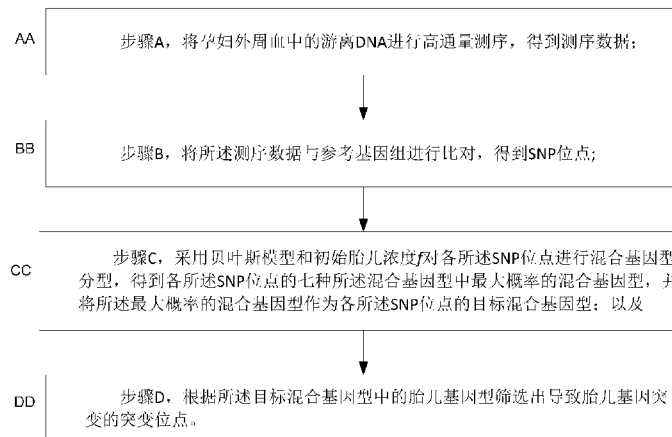


图 1

AA Step A, perform high-throughput sequencing on free DNA in a pregnant woman's peripheral blood to obtain sequencing data;
BB Step B, compare the sequencing data with reference genome to obtain SNP sites
CC Step C, perform mixed genotyping on each SNP site using Bayesian model and an initial fetal concentration f to obtain a mixed genotype with the maximum probability among the seven mixed genotypes for each of the SNP site, and take the mixed genotype with maximum probability as the target mixed genotype for each of the SNP site; and
DD Step D, select a mutation site that causes the gene mutation of the fetus according to the genotype of the fetus in the target mixed genotypes.

(57) Abstract: Provided are a method and a device for detecting a genetic mutation, and a kit for typing genotypes of a pregnant woman and a fetus. The method comprises: performing high-throughput sequencing on free DNA in a pregnant woman's peripheral blood to obtain sequencing data; comparing the sequencing data with reference genome to obtain SNP sites; performing mixed genotyping on each SNP site to obtain target genotypes for each SNP site; and selecting a mutation site that causes the gene mutation from the genotype of the fetus in the target genotypes.

100176 (CN)。 陈重建(CHEN, Chongjian); 中国
北京市经济技术开发区科创六街88号院8号
楼2单元701室, Beijing 100176 (CN)。

(74) 代理人: 北京康信知识产权代理有限公司 (KANGXIN PARTNERS, P.C.); 中国北京
市海淀区知春路甲48号盈都大厦A座
16层, Beijing 100098 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家
保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,
BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,
CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,
JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,
LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 提供了一种检测基因突变的方法、装置和用于对孕妇和胎儿的基因型进行分型的试剂盒。该方法包括: 将孕妇外周血中的游离DNA进行高通量测序得到测序数据; 将测序数据与参考基因组比对得到SNP位点, 对各SNP位点进行混合基因型分型, 得到各SNP位点的目标基因型分型; 从目标基因型分型中胎儿的基因型中筛选出导致基因突变的突变位点。

检测胎儿基因突变的方法、装置和试剂盒

技术领域

本发明涉及生物信息领域，具体而言，涉及一种检测胎儿基因突变的方法、装置和试剂盒。

背景技术

产前诊断又称宫内诊断，是指胎儿出生前采用各种方法预测其是否有先天性疾病（包括畸形和遗传性疾病），为能否继续妊娠提供科学依据。其中，遗传性疾病的产前诊断主要针对染色体病及孟德尔遗传病（Mendelian inheritant disease）。孟德尔遗传病是指按照孟德尔遗传方式进行传递的疾病，通常由一对等位基因控制的单个基因突变引起，涉及单个核苷酸到整个基因改变，故又称单基因病（single-genedefects）。截至到 2013 年 6 月 25 日，OMIM（online mendelian inheritance in man）数据库已收入分子机制明确的单基因病 4912 种，涉及致病基因 2992 个。

从目前产前检测的策略来看，基因诊断可以分为直接基因诊断和间接基因诊断两大类。直接基因诊断即直接检测致病基因本身，此类方法主要适用于先证者基因突变位点、类型、致病性明确的家系。从目前单基因病产前诊断的检测项目来看，根据诊断时间段的不同，分为植入前遗传学诊断（PGD）和妊娠期产前诊断两种。高通量测序技术的发展大大加速了临床检测技术的革新。

妊娠产前诊断包括有创产前诊断和无创产前诊断。无创产前诊断又称非侵入性的产前诊断技术（non-invasive prenatal diagnosis, NIPD）。随着母亲血浆中存在胎儿游离 DNA（cffDNA）的发现，无创产前诊断因其风险低的优势越来越受到欢迎。但由于母源 DNA 与胎儿游离 DNA 差异极小，大量母源 DNA 背景无疑增加了对于胎儿游离 DNA 检测的难度，这一点尤其表现在点突变的检测方面。

最近，Liao 等和 Lo 等通过对妊娠妇女血浆游离 DNA 进行高达 65x 人类基因组覆盖度的测序分析，检测出胎儿携带有不少于父源 95% 的特异性 SNP，通过测序结果推导出胎儿和妊娠妇女基因组遗传图谱，成功检测出 1 例胎儿遗传了父亲的地中海贫血症基因 4bp 的已知碱基突变。

上述方法虽然能够通过对妊娠妇女血浆游离 DNA 的测序分析推导出胎儿的基因组遗传图谱，但需要结合父亲来源的遗传信息。对多个样本进行测序无疑会增加大量测序成本，对父亲来源遗传信息的依赖也可能受到限制。上述方法同时还存在需要全基因组测序、对测序深度要求高和仅能判断与父源相关突变的问题。因此，仍需要对现有的检测方法进行改进。

发明内容

本发明旨在提供一种检测基因突变的方法、装置和用于对孕妇和胎儿的基因型进行分型的试剂盒，以对测序数据范围内的胎儿所有 SNP 进行检测，同时降低检测成本。

为了实现本发明的上述目的,根据本发明的一个方面,提供了一种检测基因突变的方法,该方法包括以下步骤:将孕妇外周血中的游离 DNA 进行高通量测序,得到测序数据;将测序数据与参考基因组进行比对,得到 SNP 位点;采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型,得到各 SNP 位点的七种混合基因型中最大概率的混合基因型,并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型;以及根据目标混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点;其中,混合基因型是指孕妇和胎儿组成的假四倍体的基因型,混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种,将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种, A 表示各 SNP 位点的参考基因型, B 表示各 SNP 位点的突变基因型。

进一步地,当初始胎儿浓度不是真实胎儿浓度时,获取目标混合基因型的步骤包括:步骤 C1,采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型,得到各 SNP 位点的七种混合基因型中最大概率的混合基因型,并将最大概率的混合基因型作为各 SNP 位点的初始混合基因型;步骤 C2,选取方便计算第二胎儿浓度的初始混合基因型作为第二混合基因型;步骤 C3,根据第二混合基因型及测序数据,计算得到第二胎儿浓度 f' ;步骤 C4,比较第二胎儿浓度 f' 与初始胎儿浓度 f ,得到差值 Δf ;步骤 C5,判断差值 Δf 与设定值的关系;步骤 C6,当 Δf 大于设定值时,将 f' 作为 f 重复步骤 C1 至步骤 C5;当 Δf 在设定值以下时,初始胎儿浓度 f 对应的初始混合基因型即为目标混合基因型。

进一步地,采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型,得到各 SNP 位点的七种混合基因型中最大概率的混合基因型的步骤包括:根据七种混合基因型的条件概率和为 1 得到下列公式 (1),

$$\sum P(G_j | S) = 1 \quad (1)$$

其中 G_j 表示七种混合基因型中的任一种, S 表示各 SNP 位点中的某一位点, $P(G_j | S)$ 表示在 S 条件下, SNP 位点的混合基因型是 G_j 的概率;由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中, $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 的概率, j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7;

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准混合基因型,则由公式 (2) 得出以下公式 (3):

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除,得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中, φ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度, 通过负二项分布公式得到; 然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型, 将出现概率最大的混合基因型记为第 i 个 SNP 位点上的最大概率的混合基因型。

进一步地, 公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

中孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, θ 为第 i 个 SNP 位点的群体突变频率。

进一步地, 公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因型的个数, k 代表第 i 个 SNP 位点上出现参考等位基因型的个数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上出现突变等位基因型的概率。

进一步地, 第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上出现突变等位基因的概率 $f(b)$, 根据混合基因型 G_{ij} 的不同, 概率 $f(b)$ 分别按如下方式计算: 当混合基因型 G_{ij} 为 G_{i1} 时, $f(b)$ 的值为 0; 当混合基因型 G_{ij} 为 G_{i2} 时, $f(b)$ 的值为 $f/2$; 当混合基因型 G_{ij} 为 G_{i3} 时, $f(b)$ 的值为 $0.5-f/2$; 当混合基因型 G_{ij} 为 G_{i4} 时, $f(b)$ 的值为 0.5; 当混合基因型 G_{ij} 为 G_{i5} 时, $f(b)$ 的值为 $0.5+f/2$; 当混合基因型 G_{ij} 为 G_{i6} 时, $f(b)$ 的值为 $1-f/2$; 当混合基因型 G_{ij} 为 G_{i7} 时, $f(b)$ 的值为 1; 其中, f 代表初始胎儿浓度。

进一步地, 初始胎儿浓度为预估胎儿浓度, 优选预估胎儿浓度为 10%; 更优选设定值 ≤ 0.001 。

进一步地, 第二混合基因型选自如下四种混合基因型中的任意一种或两种以上: AAAB、ABAA、ABBB 以及 BBAB。

进一步地, 从混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点的步骤包

括：将各 SNP 位点的真实混合基因型中的胎儿基因型中属于人群中高发的多态性位点进行过滤，得到初步候选突变位点；对初步候选突变位点中导致同义突变、无义突变以及发生在非保守区域的 SNP 位点进行过滤，得到候选突变位点；对候选突变位点进行文献筛查和临床资料筛查，得到导致胎儿基因突变的突变位点。

根据本发明的另一个方面，提供了一种检测基因突变的装置，该装置包括：检测模块，用于对孕妇外周血中的游离 DNA 进行高通量测序，得到测序数据；比对模块，用于将测序数据与参考基因组序列进行比对，得到 SNP 位点；目标混合基因型确定模块，用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型，并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型；以及突变位点筛选模块，用于根据各 SNP 位点的目标混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点；其中，混合基因型是指孕妇和胎儿组成的假四倍体的基因型，混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种，将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点的突变基因型。

进一步地，当初始胎儿浓度不是真实胎儿浓度时，目标混合基因型确定模块包括：预估模块，用于采用贝叶斯模型和初始胎儿浓度 f 计算得到各 SNP 位点的七种混合基因型中概率最大的混合基因型，并将概率最大的混合基因型作为初始混合基因型；选取模块，用于选取方便计算第二胎儿浓度的初始混合基因型作为第二混合基因型；计算模块，用于根据第二混合基因型及测序数据，计算得到第二胎儿浓度 f' ；比较模块，用于比较第二胎儿浓度 f' 与初始胎儿浓度 f ，得到差值 Δf ；判断模块，用于判断差值 Δf 与设定值的关系；循环模块，用于当 Δf 大于设定值时，将 f' 作为 f 重复执行预估模块、选取模块、计算模块、比较模块以及判断模块；标记模块，用于当 Δf 在设定值以下时，将初始胎儿浓度 f 对应的初始混合基因型记为目标混合基因型。

进一步地，目标混合基因型确定模块采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型的步骤包括：根据七种混合基因型的条件概率和为 1 得到下列公式 (1)，其中 G_j 表示七种所述混合基因型中的任一种， S 表示各 SNP 位点中的某一位点， $P(G_j | S)$

$$\sum P(G_j | S) = 1 \quad (1)$$

表示某一 SNP 位点如 S 一样时，SNP 位点的混合基因型是 G_j 的概率；由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中， $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 基因型的概率， j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7；

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准, 则由公式 (2) 得出以下公式 (3):

$$P(G_{j*} | S_i) = \frac{P(S_i | G_{j*})P(G_{j*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除, 得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{j*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{j*})P(G_{j*})} \quad (4)$$

其中, φ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度, 通过负二项分布公式得到; 然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型, 将出现概率最大的混合基因型记为第 i 个 SNP 位点上的最大概率的混合基因型。

进一步地, 上述公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

中孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, θ 为第 i 个 SNP 位点的群体突变频率。

进一步地, 上述公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因的个数, k 第 i 个 SNP 位点上出现参考等位基因的个数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上胎儿出现突变等位基因的概率。

进一步地, 第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上胎儿出现突变等位基因的概率 $f(b)$, 根据混合基因型 G_{ij} 的不同, 分别按如下方式计算: 当混合基因型 G_{ij} 为 G_{i1} 时, $f(b)$ 的值为 0; 当混合基因型 G_{ij} 为 G_{i2} 时, $f(b)$ 的值为 $f/2$; 当混合基因型 G_{ij} 为 G_{i3} 时, $f(b)$ 的值为 $0.5-f/2$; 当混合基因型 G_{ij} 为 G_{i4} 时, $f(b)$ 的值为 0.5; 当混合基因型 G_{ij} 为 G_{i5} 时, $f(b)$ 的值为 $0.5+f/2$; 当混合基因型 G_{ij} 为 G_{i6} 时, $f(b)$ 的值为 $1-f/2$; 当混合基因型 G_{ij} 为 G_{i7} 时, $f(b)$ 的值为 1; 其中, f 代表初始胎儿浓度。

进一步地, 预估模块中的初始胎儿浓度为预估胎儿浓度, 优选地, 预估胎儿浓度为 10%, 更优选, 判断模块中的设定值 ≤ 0.001 。

进一步地, 计算模块中的第二混合基因型选自如下四种混合基因型中的任意一种或多种: AAAB、ABAA、ABBB 以及 BBAB。

进一步地, 突变位点筛选模块包括: 高发多态性位点过滤子模块: 用于将各 SNP 位点的目标混合基因型中的胎儿基因型中属于人群高发的多态性位点进行过滤, 得到初步候选突变位点; 基因突变位点筛选子模块: 用于从初步候选突变位点中的导致同义突变、无义突变以及发生在非保守区域的 SNP 位点过滤掉, 得到候选突变位点; 文献及临床资料筛查子模块: 用于对候选突变位点进行文献筛查和临床资料筛查, 得到导致胎儿基因突变的突变位点。

根据本发明的再一个方面, 提供了一种用于对孕妇和胎儿的基因型进行分型的试剂盒, 该试剂盒包括: 用于从孕妇外周血血浆中富集游离 DNA 进行高通量测序的试剂和器械; 用于将高通量测序得到的测序数据与参考基因组比对得到 SNP 位点的器械; 以及用于采用贝叶斯模型和初始胎儿浓度 f 得到各 SNP 位点的七种混合基因型中最大概率的混合基因型, 并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型的器械; 其中, 混合基因型是指孕妇和胎儿组成的假四倍体的基因型, 混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种, 将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种, A 表示各 SNP 位点的参考基因型, B 表示各 SNP 位点突变基因型。

进一步地, 当初始胎儿浓度不是真实胎儿浓度时, 获取各 SNP 位点的目标混合基因型的器械包括: 第一计算元件, 用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型, 得到各 SNP 位点的 7 种混合基因型中概率最大的混合基因型, 并将概率最大的混合基因型作为初始混合基因型; 选取元件, 用于选取方便计算第二胎儿浓度的初始混合基因型, 记为第二混合基因型; 第二计算元件, 用于根据第二混合基因型及测序数据, 计算得到第二胎儿浓度 f' ; 比较元件, 用于比较第二胎儿浓度 f' 与初始胎儿浓度 f , 得到差值 Δf ; 判断元件, 用于判断 Δf 是否大于设定值; 以及循环元件, 用于当 Δf 大于设定值时, 将 f' 作为 f 重复操作第一计算元件、选取元件、第二计算元件、比较元件以及判断元件; 标记元件, 用于当 Δf 在设定值以下时, 将初始胎儿浓度 f 对应的初始混合基因型标记为目标混合基因型。

进一步地, 获取目标混合基因型的器械中采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型, 得到各 SNP 位点的七种混合基因型中最大概率的混合基因型的步骤包括: 根据七种混合基因型的条件概率和为 1 得到下列公式 (1), 其中 G_j 表示七种混合基因型中的任一种, S 表示各 SNP 位点中的某一位点,

$$\sum P(G_j | S) = 1 \quad (1)$$

$P(G_j | S)$ 表示某一 SNP 位点如 S 一样时, SNP 位点的混合基因型是 G_j 的概率; 由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中, $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_{ij} 基因型的概率, j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7;

从 G_j 中选取任意一种混合基因型 G_{ij^*} 为基准, 则由公式 (2) 得出以下公式 (3):

$$P(G_{ij^*} | S_i) = \frac{P(S_i | G_{ij^*})P(G_{ij^*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除, 得到下列公式 (4)

$$\phi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij^*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij^*})P(G_{ij^*})} \quad (4)$$

其中, ϕ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j^*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度, 通过负二项分布公式得到; 然后由下列公式 (5)

$$G = \arg \max(\phi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型, 将出现概率最大的混合基因型记为第 i 个 SNP 位点上的初始混合基因型。

进一步地, 公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

中孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, θ 为第 i 个 SNP 位点的群体突变频率。

进一步地, 公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因型的个数, k 代表第 i 个 SNP 位点上出现参考等位基因型的个数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上出现突变等位基因型的概率。

进一步地, 公式 (7) 中的 $f(b)$, 根据混合基因型 G_{ij} 的不同, 分别按如下公式计算: 当混合基因型 G_{ij} 为 G_{i1} 时, $f(b)$ 的值为 0; 当混合基因型 G_{ij} 为 G_{i2} 时, $f(b)$ 的值为 $f/2$; 当混合基因型 G_{ij} 为 G_{i3} 时, $f(b)$ 的值为 $0.5 \cdot f/2$; 当混合基因型 G_{ij} 为 G_{i4} 时, $f(b)$ 的值为 0.5; 当混合基因型 G_{ij} 为 G_{i5} 时, $f(b)$ 的值为 $0.5 + f/2$; 当混合基因型 G_{ij} 为 G_{i6} 时, $f(b)$ 的值为 $1 - f/2$; 当混合基因型 G_{ij}

为 G_7 时, $f(b)$ 的值为 1; 其中, f 代表初始胎儿浓度。

进一步地, 预估元件中的初始胎儿浓度为预估胎儿浓度, 优选地预估胎儿浓度为 10%, 判断元件中的设定值 ≤ 0.001 。

进一步地, 第二计算元件中的第二混合基因型选自如下四种混合基因型中的任意一种或多种: AAAB、ABAA、ABBB 以及 BBAB。

应用本发明的技术方案, 通过高通量测序及与参考基因组序列比对, 得到混合有母亲和胎儿基因组信息的 SNP 位点, 并通过利用本发明所提出的假四分体分型模型对孕妇外周血中游离的胎儿 DNA 和母亲自身 DNA 的基因型进行分型, 从而实现了仅利用孕妇外周血便能检测出胎儿所有可能的基因突变的情况。本发明的方法减少了对父亲来源或/和母亲来源的样本的单独测序; 并且对测序技术没有特别限制, 可以利用目标区域捕获测序技术获得测序数据, 降低测序成本。同时, 能够对测序数据范围内的所有 SNP 位点的胎儿基因突变情况进行检测, 为产前诊断提供了便利和多样化的服务。

附图说明

构成本申请的一部分的说明书附图用来提供对本发明的进一步理解, 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:

图 1 示出了本申请一种优选实施例中一种检测基因突变的方法的流程图;

图 2 示出了本申请的实施例 1 中的操作流程图; 以及

图 3 示出了本申请的实施例 2 采用本申请的检测基因突变的方法所检测到的突变通过现有方法进行验证的结果图。

具体实施方式

需要说明的是, 在不冲突的情况下, 本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本发明。

名词解释:

群体突变频率是指某种基因在某个特定种群内出现突变的比例, 如亚洲人千人突变频率。

设定值反应的是检测分辨率的高低, 可以根据测序的实际情况进行合理设定, 例如, 在测序深度 $\geq 1000\times$ 时, 优选的设定值 ≤ 0.001 。

胎儿浓度即孕妇血浆中胎儿游离 DNA 占血浆总游离 DNA 的比例。胎儿浓度 f 可以通过本领域技术人员公知的实验方法获得, 也可以根据本领域公知常识初步预估, 例如为 5~20%。

在本发明中, 对孕妇外周血中游离 DNA 的高通量测序可以是全测序, 或者对所关注基因的目标区域的捕获测序。

在本发明中, 混合基因型是指孕妇和胎儿组成的假四倍体的基因型, A 和 B 均为单倍型。混合基因型的前两个单倍型表示母亲的二倍体基因型, 后两个单倍型表示胎儿的二倍体基因

型。其中，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点的突变基因型。对于测序数据的某个位点，若其与在参考基因组的对应位点的碱基一致则为参考基因型，反之则为突变基因型。混合基因型，例如可以是 AAAB，其表示：母亲的二倍体基因型是 AA 且为纯合参考型，胎儿的二倍体基因型是 AB 且为携带突变型。

在本发明中，群体突变频率是指在特定群体内发生某种突变的细胞或个体数的比例，如亚洲人千人突变频率。

正如背景技术部分提到的，现有技术中利用高通量测序检测胎儿基因突变的方法，通常需要借助于额外的父源和母源的样品信息，且仅能检测与 Y 染色连锁的单基因遗传病。为了降低检测成本，并提供给多样化的产前检测服务，在本发明一种典型的实施方式中，提供了一种检测基因突变的方法，如图 1 所示，该方法包括以下步骤：将孕妇外周血中的游离 DNA 进行高通量测序，得到测序数据；将测序数据与参考基因组进行比对得到 SNP 位点；采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型，并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型；根据目标混合基因型中胎儿的基因型筛选出导致胎儿基因突变的突变位点；其中，混合基因型是指孕妇和胎儿形成的假四倍体的基因型，混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 所组成的七种中的任一种，且将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点突变基因型。

基于孕妇外周血中游离 DNA 既包含母亲的 DNA 又包含胎儿的 DNA，而且利用目前的技术手段难以将两种来源的 DNA 进行彻底分离这一事实，发明人提出的假四倍体的概念，将孕妇和胎儿混合得到的四倍体称为假四倍体，并在基因组的各位点上，由孕妇的基因型和胎儿的基因型混合得到的该位点的基因型称为混合基因型。为了检测各位点上出现突变基因型的概率，用 A 表示该位点上的正常的参考基因型；用 B 表示该位点的突变基因型。

通过将孕妇在各位点上的二倍体基因型放在前面，胎儿在对应位点上的二倍体基因型放在后面，来表示假四倍体在某一位置上的混合基因型，就得到 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种可能的混合基因型，通过测序数据推算出各 SNP 位点上最大概率的混合基因型即可得到在该位点的目标混合基因型，进而从目标混合基因型中获得胎儿的基因型。

进一步，发明人提出了利用条件概率和贝叶斯模型对上述假四倍体的混合基因型进行分型的思想。

本发明的上述方法，通过仅对母亲外周血中的游离 DNA 进行高通量测序和序列比对，得到混有母本和胎儿基因组信息的 SNP 位点，并通过利用本发明所提出的混合基因型分型的概念对孕妇外周血中游离的胎儿在各 SNP 位点上的基因型和母本在各 SNP 位点上的基因型进行分型，从而实现仅利用孕妇外周血便能检测出胎儿所有可能的基因突变的情况。一方面减少了对父源和母源样本的测序，降低测序成本；另一方面也为某些特殊条件下，如父源样本无法获取的情况下，对胎儿基因突变情况的检测提供了便利，为产前诊断提供了多样化的服务。

在本发明的上述方法中，本领域技术人员可以在本发明所提出的假四倍体及假四倍体的混合基因型分型概念的基础上，能够利用条件概率和贝叶斯模型对假四倍体的混合基因型进

行分型,进而得到胎儿在 SNP 位点上的基因型,为筛选导致胎儿基因突变的突变位点奠定基础。根据孕妇外周血样本来源的不同,其中的初始胎儿浓度分为已知或未知两种情况,当已知胎儿浓度时,初始胎儿浓度 f 即为真实的胎儿浓度。利用初始胎儿浓度 f 和贝叶斯模型即可计算得到各 SNP 位点的最大概率的混合基因型。而当胎儿浓度未知时,需要进行真实胎儿浓度的推定过程。

在本发明一种优选的实施例中,当初始胎儿浓度不是真实胎儿浓度时,获取目标混合基因型的步骤包括:步骤 C1,采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型,得到各 SNP 位点的七种混合基因型中最大概率的混合基因型,并将最大概率的混合基因型作为各 SNP 位点的初始混合基因型;步骤 C2,选取方便计算第二胎儿浓度的初始混合基因型作为第二混合基因型;步骤 C3,根据第二混合基因型及测序数据,计算得到第二胎儿浓度 f' ;步骤 C4,比较第二胎儿浓度 f' 与初始胎儿浓度 f ,得到差值 Δf ;步骤 C5,判断差值 Δf 与设定值的关系;步骤 C6,当 Δf 大于设定值时,将 f' 作为 f 重复步骤 C1 至步骤 C5;当 Δf 在设定值以下时,初始胎儿浓度 f 对应的初始混合基因型即为目标混合基因型。

上述假四倍体的混合基因型分型步骤中,由于胎儿浓度未知,通过根据常识初步预估的胎儿浓度,例如 5~15%,作为初始胎儿浓度计算,在该胎儿浓度下七种可能的混合基因型出现的概率,从而得到各个 SNP 位点上概率最大的混合基因型。结合实际测序数据,通过突变来自母体或胎儿一方的基因突变来计算实际的胎儿浓度,然后将此时计算得到的胎儿浓度与预估的初始胎儿浓度进行比较,若差值小于某一设定值,则需要将此时计算得到的胎儿浓度作为步骤 C1 中的初始浓度,重新经过步骤 C1 到步骤 C5 的步骤,直到某次计算得到的胎儿浓度与该循环中步骤 C1 中的初始浓度差值小于设定值时,终止循环,并将该循环中步骤 C1 中的初始浓度下的各个 SNP 位点上概率最大的混合基因型记为各个 SNP 位点的目标混合基因型。上述设定值反应的是检测分辨率的高低,可以根据实际情况进行合理设定。例如,在测序深度 $\geq 1000\times$ 时,优选的设定值 ≤ 0.001 。

上述优选实施例中,便于计算胎儿浓度的混合基因型的选取原则可以根据计算方法的不同进行合理选择。在本发明一种优选的实施例中,上述胎儿浓度计算混合基因型包括但不限于 AAAB、ABAA、ABBB 以及 BBAB 中的任意一种或多种。这些混合基因型中的突变等位基因或者参考等位基因仅来自孕妇或胎儿一方,根据测序数据中突变等位基因和参考等位基因被检测到的次数,即可计算其中一方的浓度,从而很容易得到胎儿的浓度。

在本发明一种优选的实施例中,上述采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型,得到各 SNP 位点的七种混合基因型中最大概率的混合基因型的步骤包括:根据七种混合基因型的条件概率和为 1 得到下列公式 (1),其中 G_j 表示七种混合基因型中的任一种, S 表示各 SNP 位点中的

$$\sum P(G_j | S) = 1 \quad (1)$$

某一位点, $P(G_j | S)$ 表示某一 SNP 位点如 S 一样时, SNP 位点的混合基因型是 G_j 的概率;由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中, $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 基因型的概率, j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7; 从 G_i 中选取任意一种混合基因型 G_{j^*} 为基准, 则由公式 (2) 得出以下公式 (3):

$$P(G_{ij^*} | S_i) = \frac{P(S_i | G_{ij^*})P(G_{ij^*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除, 得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij^*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij^*})P(G_{ij^*})} \quad (4)$$

其中, φ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j^*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度 f , 通过负二项分布公式得到; 然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型, 将出现概率最大的混合基因型记为第 i 个 SNP 位点的最大概率的混合基因型。

在本发明的上述优选实施例中, 通过将某一 SNP 位点上的最大概率的混合基因型的计算方法转化为计算在该 SNP 位点出现上述七种混合基因型的概率与出现其中某一种混合基因型的概率的比值, 间接得到在该 SNP 位点上出现概率最大的混合基因型, 进而推定为该位点的初始混合基因型。

在本发明的上述方法中, 在已知某一位点上出现某种突变基因型的情况下, 本领域技术人员能够利用千人基因组的数据库中的群体突变频率计算出公式 (4) 中的 $P(G_{ij})$ 。在本发明一种优选的实施例中, $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

计算得到。在上述公式 (6) 中, G' 代表孕妇或胎儿在某一位点分别出现上述三种可能的基因型, 则该特定位点上混合基因型出现的概率即为孕妇在该位点出现基因型 G' 的概率与胎儿在该位点出现基因型 G' 的概率的乘积, 其中, θ 为第 i 个 SNP 位点的群体突变频率。该参数 θ 通过千人基因组的数据库的群体频率得到。

在本发明的上述方法中， $P(S_i | G_{ij})$ 的数值根据实际测序数据中，某一位点为特定的混合基因型 G_{ij} 时，出现突变等位基因的次数与出现参考等位基因的次数不同以及初始胎儿浓度的差异利用负二项分布公式得到。在本发明一种具体的实施例中，公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到：

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k$$

其中， r 代表第 i 个 SNP 位点上出现等位基因的次数， k 第 i 个 SNP 位点上出现参考等位基因的次数， $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上胎儿出现突变等位基因的概率。

在上述实施例中， $f(b)$ 与孕妇外周血中胎儿游离 DNA 的浓度有关，利用常规的胎儿浓度的计算方法，如 Fetal Quant 进行计算即可（参见文献 Lench N, Barrett A, Fielding S, *et al.* The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made[J]. Prenatal diagnosis, 2013, 33(6): 555-562.）。在获得胎儿浓度之后，根据某位点为混合基因型中的某一种时，推测得到该混合基因型的父本可能的基因型，从而推算出某一特定混合基因型出现时，理论上出现突变等位基因的概率。

如前述，本发明中的初始胎儿浓度利用最大期望算法进行迭代计算后得到最终的第二胎儿浓度与初始胎儿浓度的差值与设定值无明显区别时，认为此时的初始胎儿浓度或者第二胎儿浓度为该样本中胎儿的真实浓度。假设初始胎儿浓度（预估胎儿浓度） f 为 10%，计算当 $f=10\%$ 时所有 SNP 位点的混合基因型，利用混合基因型实际检测的参考等位基因与突变等位基因的频率计算实际的胎儿浓度 f' （第二胎儿浓度 f' ），如果 f' 与 f 之间的差值小于设定的值，则迭代结束，迭代结束时对应的 f' 即为真实胎儿浓度。更优选，上述设定值小于等于 0.001。

具体来说，采用以下算法对真实胎儿浓度进行计算：

第 0 步：预估胎儿浓度 f 为 10%；

迭代：

第 1 步：根据混合基因型分型推断胎儿基因型，并根据基因型中的 $f(b)$ 计算胎儿浓度 f' ；

第 2 步：计算 Δf ，其中 $\Delta f = D(f - f')$ ；

第 3 步：如果 $\Delta f < \varepsilon$ ，迭代停止，其中 ε 代表任意小的正数；

第 4 步：胎儿浓度 $f = f(b)$ ；

其中，函数 $D()$ 代表距离函数，衡量的是两个变量之间的差异。

下面举例说明此处真实胎儿浓度的计算过程：

假如，选择了 3 个 SNP 位点进行计算，假设都为 AAAB 型， f 为初始预估的胎儿浓度， f' 为根据检测出的 A 和 B 的频率推算的第二胎儿浓度。

第一个 SNP，检测到 A 19 次，B 1 次，假设 $f=10\%$ ，计算 7 种基因型的概率值，比较获得该 SNP 的混合基因型为 AAAA，其不符合 AAAB 的假设，应当去除；

第二个 SNP，检测到 A 16 次，B 4 次，假设 $f=10\%$ ，计算 7 种基因型的概率值，比较获得该 SNP 的混合基因型为 AAAB，符合 AAAB 的假设；此时 $f'=40\%$ ；

第三个 SNP，检测到 A 18 次，B 2 次，假设 $f=10\%$ ，计算 7 种基因型的概率值，比较获得该 SNP 的混合基因型 AAAB，符合 AAAB 的假设；此时 $f'=20\%$ 。

综合 3 个 SNP 的情况，第二三个采信，第一个排除，则 f' 的平均值为 30%，假设的 f 与检测推算的 f' 之间的差值大于 0.001，需要进行迭代计算直到之间的差值小于 0.001 为止。

本发明的上述通过迭代算法计算得到胎儿浓度 f 的方法相比现有技术中男胎用 X 染色体，女胎用甲基化的方法计算胎儿浓度 f 的方法相比，具有不仅准确性高、而且不受性别限定的优点。

利用上述方法得到胎儿浓度后，便可计算上述第 i 个 SNP 位点的混合基因型为 G_{ij} 时，理论上出现等位基因的概率 $f(b)$ ，根据混合基因型 G_{ij} 的不同，分别按如下公式计算：当混合基因型 G_{ij} 为 G_{i1} 时， $f(b)$ 的值为 0；当混合基因型 G_{ij} 为 G_{i2} 时， $f(b)$ 的值为 $f/2$ ；当混合基因型 G_{ij} 为 G_{i3} 时， $f(b)$ 的值为 $0.5-f/2$ ；当混合基因型 G_{ij} 为 G_{i4} 时， $f(b)$ 的值为 0.5；当混合基因型 G_{ij} 为 G_{i5} 时， $f(b)$ 的值为 $0.5+f/2$ ；当混合基因型 G_{ij} 为 G_{i6} 时， $f(b)$ 的值为 $1-f/2$ ；当混合基因型 G_{ij} 为 G_{i7} 时， $f(b)$ 的值为 1；其中， f 代表初始胎儿浓度。

$f(b)$ 是指假四倍体的混合基因型中出现突变等位基因的概率，所以只要计算假四倍体的混合基因型中 B 出现的概率即可，具体计算如下表 1：(假设假四倍体的混合基因型概率为 1)

表 1:

混合基因型	母亲		胎儿		出现 B 基因型的概率 $f(b)$
	基因型	浓度	基因型	浓度	
1--AAAA	AA	$1-f$	AA	f	未出现 B，所以是 0
2--AAAB	AA	$1-f$	AB	f	$f/2$
3--ABAA	AB	$1-f$	AA	f	$(1-f)/2=0.5-f/2$
4--ABAB	AB	$1-f$	AB	f	$1/2$
5--ABBB	AB	$1-f$	BB	f	$(1-f)/2+f=0.5+f/2$
6--BBAB	BB	$1-f$	AB	f	$1-f/2$
7--BBBB	BB	$1-f$	BB	f	1

在本发明的上述方法中，通过本发明的上述混合基因型分型便能够推算得到各 SNP 位点的目标混合基因型，进而得到胎儿的基因型。在得到胎儿的基因型之后便可以找到导致基因突变的致病突变位点。在本发明一种具体的实施例，根据各 SNP 位点的目标混合基因型中的胎儿基因型的不同，从 SNP 位点中筛选出突变位点的步骤包括：将推算出胎儿基因型的各 SNP 位点中属于人群中高发的多态性位点进行过滤，得到初步候选突变位点；从初步候选位点中过滤能够导致同义突变和无义突变的 SNP 位点和发生在非保守区域的 SNP 位点，得到候选突变位点；对候选突变位点进行文献筛查和临床资料筛查，得到导致基因突变的突变位点。

上述实施例中，对胎儿的各 SNP 进行分析找到致病突变位点的过程中，对导致人类群体中不同个体之间存在差异的高频的 SNP 位点进行删除，因为这些位点明显不属于致病突变位

点。在本发明中,利用目前医学界已经整理的 dnSNP135 公共数据库和 Freq_1000g2012feb(千人基因组)数据库,去除在人群中的高发的多态性位点。在去除个体差异导致的 SNP 位点之后,得到初步的候选突变位点,然后利用本领域常用的突变预测软件筛选有害突变,例如,ANNOVAR 软件既可以筛选突变是否导致氨基酸改变,即筛选突变是否导致有义突变,还可以筛选突变是否发生在保守序列区域。

在进行上述软件筛选之后,还需要对筛到的可能的致病突变位点进行人工解读,所谓“人工解读”就是通过已有的数据库和文献的检索,从可能的致病突变位点中找到与单基因遗传病相关的位点信息并进行相应解读。此外,本发明的方法不仅局限于检测是否存在导致已知单基因遗传病的热点突变位点,还可以检测已知单基因遗传病的非热点突变位点以及未报道过的潜在的致病基因及其突变位点。因而,可以根据客户的不同需求,为客户提供多样化的服务。

在本发明的上述方法中,对孕妇外周血中的游离 DNA 进行高通量测序,得到测序数据的步骤通过利用本领域常用的高通量测序方法先进行样本 DNA 文库构建,然后利用现有高通量测序平台进行测序即可。在本发明一种优选的实施例中,对孕妇外周血中的游离 DNA 进行高通量测序,得到测序数据的步骤包括:获取孕妇外周血血浆 DNA,对血浆 DNA 中的游离 DNA 进行富集,得到富集 DNA;对富集 DNA 进行文库构建,得到测序文库;对测序文库进行高通量测序,得到测序数据。

上述优选实施例中,由于孕妇外周血中的游离 DNA 在母源血浆中的含量比较低,基本在 10%以下,因此,在高通量测序之前需要对游离 DNA 进行提取和富集的步骤。该提取和富集步骤本领域技术人员可以根据各自样品的差异和对数据需求的不同选择合适的提取和富集方法。比如,可以采用德国 Qiagen 的 QIAmp DNA Blood Mini Kit 进行提取富集或者市售的其他公司的同类试剂进行提取富集,或者自己配制的能够对孕妇外周血进行提取富集的相关试剂。

在将上述富集 DNA 进行文库构建,得到测序文库的步骤之后,根据不同样本检测目的的不同,选择不同的目标区域进行测序检测。在本发明的实际操作过程中,在得到测序文库即将进行高通量测序之前,还包括将测序文库进行目标区域捕获,得到含有目标区域的测序文库的步骤。在本发明一种更优选的实施例中,通过增加将测序文库进行外显子捕获杂交的步骤,这样使得后续的高通量测序只针对外显子进行测序。由于基因在转录的过程中,内含子通常会被剪切掉,而外显子是最终编码蛋白的区域,因而,只进行外显子测序能够有效提高数据量,提高测序的效率。得到针对外显子的测序文库后,还可以根据检测目的不同和/或检测样本的不同,在特定的目标区域内对已知的单基因遗传病的突变位点进行检测或对所有测序数据中的突变位点进行整体检测。

在本发明的上述优选实施例中,根据所要捕获的目标区域的不同,可以选择不同的方法或试剂进行捕获。比如可以采用美国 Roche NimbleGen 公司的捕获试剂盒进行目标区域捕获、自制的或市售的其他公司的同类的功能的试剂盒。

在本发明另一种典型的实施方式中,提供了一种检测基因突变的装置,该装置包括:检测模块,用于对孕妇外周血中的游离 DNA 进行高通量测序,得到测序数据;比对模块,用于将测序数据与参考基因组序列进行比对得到 SNP 位点;目标混合基因型确定模块,用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型,得到各 SNP 位点的七种混

合基因型中最大概率的混合基因型，并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型；突变位点筛选模块，根据各 SNP 位点的目标混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点；其中，混合基因型是指孕妇和胎儿组成的染色体组的假四倍体基因型，是 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种，将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点的突变基因型。

在本发明的上述装置，通过检测模块及与比对模块，得到母本和胎儿基因组信息中与参考基因组不同的 SNP 位点，并通过利用目标混合基因型确定模块对孕妇和胎儿组成的假四倍体的混合基因型进行分型，从而得到各 SNP 位点上母本和胎儿的基因型，从而实现了仅利用孕妇外周血样本便能检测出胎儿所有可能的基因突变的情况。本发明的该装置不仅减少了对父源、母源（外周血中细胞基因组样本）单独样本的测序，降低测序成本；而且，为在某些特殊条件下，如父源样本无法获取的情况下，对胎儿基因突变情况的检测提供了便利，为产前诊断提供了多样化的服务。

在本发明的上述目标混合基因型确定模块中，在初始胎儿浓度即为真实胎儿浓度时，本领域技术人员可以在本发明所提出的假四倍体及假四倍体的混合基因型的概念的基础上，通过对条件概率和贝叶斯模型进行变形，得到各 SNP 位点的目标混合基因型。在本发明一种优选的实施例中，当初始胎儿浓度不是真实胎儿浓度时，目标混合基因型确定模块包括：预估模块，用于采用贝叶斯模型和初始胎儿浓度 f 计算得到各 SNP 位点的 7 种混合基因型中概率最大的混合基因型，并将该概率最大的混合基因型作为初始混合基因型；选取模块，用于选取方便计算第二胎儿浓度的初始混合基因型，记为第二混合基因型；计算模块，用于根据第二混合基因型及测序数据，计算得到第二胎儿浓度 f' ；比较模块，用于比较第二胎儿浓度 f' 与初始胎儿浓度 f ，得到差值 Δf ；判断模块，用于判断 Δf 是否大于设定值；循环模块，用于当 Δf 大于设定值时，将 f' 作为 f 重复执行预估模块、选取模块、计算模块、比较模块以及判断模块；标记模块，用于当 Δf 在设定值以下时，将初始胎儿浓度 f 对应的初始混合基因型记为目标混合基因型。

上述利用初始胎儿浓度 f 计算各 SNP 位点的 7 种混合基因型中概率最大的混合基因型的具体计算方法可以根据条件概率和贝叶斯模型进行各种变形得到。优选地，在计算七种混合基因型的概率时，将七种混合基因型的概率计算公式都分别与其中某一特定的混合基因型的概率计算公式相除，进而得到每种混合基因型与该特定混合基因型的概率比值，比值最大的即为该 SNP 位点的概率最大的混合基因型，即为该 SNP 位点的初始混合基因型。其中，上述特定的混合基因型可以是七种混合基因型中的任意一种，可以根据计算方便程度合理选择。

上述优选实施例中的计算模块中的胎儿浓度计算混合基因型的选取原则可以根据计算方法的不同进行合理选择。在本发明一种优选的实施例中，上述胎儿浓度计算混合基因型包括但不限于 AAAB、ABAA、ABBB 以及 BBAB 中的任意一种或多种。这些混合基因型中的突变等位基因或者参考等位基因仅来自孕妇或胎儿一方，根据测序数据中突变等位基因和参考等位基因被测到的次数，即可计算其中一方的浓度，从而很容易得到胎儿的浓度。

在本发明另一种优选的实施例中，目标混合基因型确定模块采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的

混合基因型的步骤包括包括：根据七种混合基因型的条件概率和为 1 得到下列公式 (1)，其中 G_j 表示七种混合基因型中的任一种， S 表示各 SNP 位点中的某一位

$$\sum P(G_j | S) = 1 \quad (1)$$

点， $P(G_j | S)$ 表示某一 SNP 位点如 S 一样时，SNP 位点的混合基因型是 G_j 的概率；由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中， $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 基因型的概率， j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7；

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准，则由公式 (2) 得出以下公式 (3)：

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除，得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中， φ_j 表示在 S_i 条件下，第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值； $P(G_{ij})$ 通过群体突变频率计算得到， $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度，通过负二项分布公式得到；

然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型，将出现概率最大的混合基因型记为第 i 个 SNP 位点的最大概率的混合基因型。

在本发明的上述优选实施例中，通过利用条件概率和贝叶斯模型，将某一 SNP 位点上的混合基因型的计算方法转化为计算在该 SNP 位点出现上述七种混合基因型的概率与出现其中某一种混合基因型的概率的比值，间接得到在该 SNP 位点上出现概率最大的混合基因型，进而推定为该位点的最大概率的混合基因型。

在本发明的上述装置中，在已知某一位点上出现某种突变基因型的情况下，本领域技术人员能够利用千人基因组的数据库中的群体突变频率计算出公式 (4) 中的 $P(G_{ij})$ 。在本发明一种优选的实施例中，上述胎儿基因型确定模块中的公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1-\theta)^2 \\ P(G' = AB) = 2\theta(1-\theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

计算得到。在上述公式(6)中, G' 代表孕妇或胎儿在某一位点分别出现上述三种可能的基因型, 则该特定位点上混合基因型出现的概率即为孕妇在该位点出现基因型 G' 的概率与胎儿在该位点出现基因型 G' 的概率的乘积, 其中, θ 为第 i 个 SNP 位点的群体突变频率。该参数 θ 通过千人基因组的数据库的群体频率得到。

在本发明的上述装置中, 胎儿基因型确定模块中的公式(4)中的 $P(S_i | G_{ij})$ 的数值根据实际测序数据中, 某一位点为特定的混合基因型 G_{ij} 时, 出现突变等位基因的次数的不同以及初始胎儿浓度的差异利用负二项分布公式得到。在本发明一种具体的实施例中, 公式(4)中的 $P(S_i | G_{ij})$ 通过以下公式(7)计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现等位基因的次数的次数, k 第 i 个 SNP 位点上出现参考等位基因的次数的次数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上出现突变等位基因的概率。

在上述实施例中, $f(b)$ 与孕妇外周血中胎儿游离 DNA 的浓度有关, 利用常规的胎儿浓度的计算方法, 如 Fetal Quant 进行计算即可 (参见文献 Lench N, Barrett A, Fielding S, *et al.* The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made[J]. Prenatal diagnosis, 2013, 33(6): 555-562.)。在获得胎儿浓度之后, 根据某位点为混合基因型中的某一种时, 推测得到该混合基因型的父本可能的基因型, 从而推算出某一特定混合基因型出现时, 理论上胎儿出现突变等位基因的概率。

在本发明另一种优选的实施例中, 初始胎儿浓度 f 采用最大期望算法进行迭代计算, 假设初始预估胎儿浓度 f 为 10%, 计算当 $f=10\%$ 时所有位点的混合基因型, 利用某些混合基因型实际检测的参考基因型与突变等位基因的频率计算实际的胎儿浓度 f' , 如果 f' 与 f 之间的差值小于设定的值, 则迭代结束, 迭代结束时对应的 f' 即为胎儿浓度 f 。更优选, 上述设定值小于等于 0.001 时, 迭代结束。具体采用的算法与前述检测方法中相同, 在此不再赘述。同样地, 上述第 i 个 SNP 位点的混合基因型 G_{ij} 为 7 种中的任意一种时, 理论上出现等位基因的概率 $f(b)$ 同表 1。

在本发明的上述装置中, 通过上述目标混合基因型确定模块中对各 SNP 位点的假四倍体进行分型便能够推算得到目标混合基因型, 进而从目标混合基因型中获得胎儿在各 SNP 位点的基因型, 在得到胎儿的基因型之后便可以找到致病突变位点。在本发明一种典型的实施例中, 上述装置中的突变位点筛选模块包括高发多态性位点过滤子模块、基因突变位点筛选子模块以及文献及临床资料筛查子模块, 高发多态性位点过滤子模块用于将各 SNP 位点的目标混合基因型中的胎儿基因型中属于人群中高发的多态性位点进行过滤, 得到初步突变位点; 基因突变位点筛选子模块, 用于从初步突变位点中过滤掉导致同义突变和无义突变的 SNP 位点以及发生在非保守区域的 SNP 位点, 得到候选突变位点; 文献及临床资料筛查子模块, 用

于从候选突变位点中筛查出文献及临床资料里记载过的导致致病基因突变的突变位点。

上述实施例中，对胎儿的各 SNP 进行分析找到致病突变位点的过程中，通过高发多态性位点过滤子模块对导致人类群体存在个体差异的高频 SNP 位点进行删除，因为这些位点明显不属于致病突变位点。在本发明中，上述高发多态性位点过滤子模块通过利用目前医学界已经整理的 dnSNP135 公共数据库和 Freq_1000g2012feb（千人基因组）数据库，去除在人群中的高发的多态性位点。在去除个体差异导致的 SNP 位点之后，得到的是胎儿所特有的 SNP 位点，然后利用基因突变位点筛选子模块筛选真正导致基因突变的位点。该模块可以利用本领域常用的突变预测模块进行有害突变筛选。ANNOVAR 模块，既可以筛选突变是否导致氨基酸改变，即筛选突变是否导致有义突变，也可以筛选突变是否发生在保守序列区域。

上述基因突变位点筛选子模块在经过上述突变预测模块筛选之后，还包括对筛到的可能的致病突变位点进行人工解读的人工解读子模块，所谓“人工解读子模块”就是将突变预测模块筛选之后的 SNP 位点与已有的数据库和文献的检索到的致病位点进行比对，从而找到与单基因遗传病相关的位点信息并进行相应解读。本发明的上述装置并非局限于检测是否存在导致已知单基因遗传病的热点突变位点，还可以检测已知单基因遗传病的非热点突变位点以及未报道过的潜在的致病基因及其突变位点。因而，可以根据客户的不同需求，为客户提供多样化的服务。

在本发明的上述装置中，检测模块是将从孕妇外周血血浆中富集的自由 DNA 制成测序文库并进行高通量测序，得到测序数据的过程。对孕妇外周血中的自由 DNA 进行高通量测序，得到测序数据的步骤通过本领域常用的高通量测序方法先进行样本 DNA 文库构建，然后利用现有高通量测序平台进行测序即可。在本发明一种优选的实施例中，上述检测装置包括获取孕妇外周血血浆 DNA，对血浆 DNA 中的自由 DNA 进行富集，得到富集 DNA；对富集 DNA 进行文库构建，得到测序文库；以及对测序文库进行高通量测序，得到测序数据的过程。

上述优选实施例中，由于孕妇外周血中的自由 DNA 在母源血浆中的含量比较低，基本在 10% 以下，因此，上述检测装置还包括了在高通量测序之前对自由 DNA 进行提取和富集的步骤，可以根据各自样品的差异和对数据需求的不同选择合适的提取和富集方法进行。比如可以采用德国 Qiagen 的 QIAmp DNA Blood Mini Kit 进行提取富集或者市售的其他公司的同类试剂进行提取富集，或者自己配制的能够对孕妇外周血进行提取富集的相关试剂。

在将上述检测装置中，对富集 DNA 进行文库构建，得到测序文库的步骤之后，根据不同样本检测目的的不同，还可以选择不同的目标区域进行测序检测。在本发明的实际操作过程中，上述检测装置在得到测序文库即将进行高通量测序之前，还包括将测序文库进行目标区域捕获，得到含有目标区域的测序文库的步骤。在本发明一种更优选的实施例中，通过增加将测序文库进行外显子捕获杂交的步骤，这样使得后续的高通量测序只针对外显子进行测序。由于基因在转录的过程中，内含子通常会被剪切掉，而外显子是最终编码蛋白的区域，因而，只进行外显子测序能够有效提高有效数据量，提高测序的效率。得到针对外显子的测序文库后，还可以根据检测目的不同和/或检测样本的不同，在特定的目标区域内对已知的单基因遗传病的突变位点进行检测或对所有测序数据中的突变位点进行整体检测。

在本发明的上述优选实施例中，根据所要捕获的目标区域的不同，可以选择不同的方法或试剂进行捕获。比如可以采用美国 Roche NimbleGen 公司的捕获试剂盒进行目标区域捕获、

自制的或市售的其他公司的同类的功能的试剂盒。

在本发明又一种典型的实施方式中，还提供了一种用于对孕妇和胎儿的基因型进行分型的试剂盒，该试剂盒包括：用于从孕妇外周血血浆中富集游离 DNA 进行高通量测序的试剂和器械；用于将高通量测序得到的测序数据与参考基因组比对得到 SNP 位点的器械；以及用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型，并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型的器械；其中，混合基因型是指孕妇和胎儿形成的假四倍体的基因型，是 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种，将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点突变的等位基因型；。

本发明的上述试剂盒，通过将孕妇与胎儿形成的假四倍体的混合基因型，并利用条件概率和贝叶斯模型，推算出各 SNP 位点上的混合基因型，进而获得母本和胎儿的基因型，从而实现了仅利用孕妇外周血样本便能检测出胎儿所有可能基因型。本发明的试剂盒不仅减少了对父源和/或母源样本的测序，降低测序成本；而且，为在某些特殊条件下，如父源样本无法获取的情况下，对胎儿基因型的检测提供了便利和多样化的服务。

在本发明的上述试剂盒中，当初始胎儿浓度即为胎儿真实浓度时，本领域技术人员可以在本发明所提出的假四倍体及假四倍体的混合基因型分型的基础上，通过对条件概率和贝叶斯模型进行变形，能够得到本发明的混合基因型。在本发明一种优选的实施例中，当初始胎儿浓度不是真实胎儿浓度时，上述试剂盒中，用于获取目标混合基因型的器械包括：第一计算元件，用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的 7 种混合基因型中概率最大的混合基因型，并将概率最大的混合基因型作为各 SNP 位点的初始混合基因型；选取元件，用于选取方便计算第二胎儿浓度的初始混合基因型，记为胎儿浓度计算混合基因型；第二计算元件，用于根据胎儿浓度计算混合基因型及测序数据，计算得到计算胎儿浓度 f' ；比较元件，用于比较计算胎儿浓度 f' 与初始胎儿浓度 f ，得到差值 Δf ；以及判断元件，用于判断 Δf 是否大于设定值；循环元件，用于当 Δf 大于设定值时，将 f' 作为 f 重复操作第一计算元件、选取元件、第二计算元件、比较元件以及判断元件；标记元件，用于当 Δf 在设定值以下时，将初始胎儿浓度所对应的初始混合基因型记为目标混合基因型。

用于对各 SNP 位点进行混合基因型分型获得各 SNP 位点的目标混合基因型的器械，通过利用预估的初始胎儿浓度计算各 SNP 位点的 7 种混合基因型的概率，得到各 SNP 位点的最大概率的混合基因型，将该概率最大的混合基因型作为初始混合基因型，然后利用选取元件选取便于计算胎儿浓度的初始混合基因型作为第二混合基因型，接着用第二计算元件根据第二混合基因型及测序数据，计算得到第二胎儿浓度 f' ，然后利用比较元件和判断元件，通过比较初始胎儿浓度与计算胎儿浓度之间的差值 Δf ，根据该差值 Δf 来评判初始胎儿浓度与计算胎儿浓度之间的差别，当该差值 Δf 大于设定值时，则利用循环元件将第二胎儿浓度 f' 记为初始胎儿浓度 f ，循环执行进行上述预估元件、选取元件、计算元件、比较元件以及判断元件，直到当该差值 Δf 小于设定值时，则认为初始胎儿浓度与计算的胎儿浓度无显著差别，即此时的初始胎儿浓度是真实的胎儿浓度，因而利用标记元件将该初始胎儿浓度计算得到的最大概率混合基因型作为目标混合基因型。

上述优选实施例中的选取元件中选取的第二混合基因型的选取原则可以根据计算方法的不同进行合理选择。在本发明一种优选的实施例中，上述第二混合基因型包括但不限于 AAAB、ABAA、ABBB 以及 BBAB 中的任意一种或多种。这些混合基因型中的突变等位基因或者参考等位基因仅来自孕妇或胎儿一方，根据测序数据中突变等位基因和参考等位基因被测到的次数，即可计算其中一方的浓度，从而很容易得到胎儿的浓度。

在本发明另一种优选的实施例中，上述获取目标混合基因型的器械中采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型的步骤包括：根据七种混合基因型的条件概率和为 1 得到下列公式 (1)，其中 G_j 表示七种混合基因型中的任一种， S

$$\sum P(G_j | S) = 1 \quad (1)$$

表示各 SNP 位点中的某一位点， $P(G_j | S)$ 表示某一 SNP 位点如 S 一样时，SNP 位点的混合基因型是 G_j 的概率；由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中， $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 基因型的概率， j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7；

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准，则由公式 (2) 得出以下公式 (3)：

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除，得到下列公式 (4)

$$\phi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中， ϕ_j 表示在 S_i 条件下，第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值； $P(G_{ij})$ 通过群体突变频率计算得到， $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度，通过负二项分布公式得到；然后由下列公式 (5) 找到七种混合基因型中出现概率最大的混合基因型，将出现概率最大的混合基因型记为第 i 个 SNP 位点上的初始混合基因型

$$G = \arg \max(\phi_j) \quad (5).$$

在本发明的上述优选的实施例中，该试剂盒中，上述第一计算元件通过利用条件概率和贝叶斯模型，将某一 SNP 位点上的混合基因型的计算方法转化为计算在该 SNP 位点出现上述七种混合基因型的概率与出现其中某一种混合基因型的概率的比值，间接得到在该 SNP 位点

上出现概率最大的混合基因型，进而记为该位点的初始混合基因型。

在本发明的上述试剂盒中，在已知某一位点上出现某种突变基因型的情况下，本领域技术人员能够利用千人基因组的数据库中的群体突变频率计算出公式（4）中的 $P(G_{ij})$ 。在本发明一种优选的实施例中，上述推算胎儿各 SNP 位点基因型的器械中的公式（4）中的 $P(G_{ij})$ 通过以下公式（6）

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

计算得到。在上述公式（6）中， G' 代表孕妇或胎儿在某一位点分别出现上述三种可能的基因型，则该特定位点上混合基因型出现的概率即为孕妇在该位点出现基因型 G' 的概率与胎儿在该位点出现基因型 G' 的概率的乘积，其中， θ 为第 i 个 SNP 位点的群体突变频率。该参数 θ 通过千人基因组的数据库的群体频率得到。

在本发明的上述试剂盒中，推算胎儿各 SNP 位点基因型的器械中的公式（4）中的 $P(S_i | G_{ij})$ 的数值根据实际测序数据中，某一位点为特定的混合基因型 G_{ij} 时，出现突变等位基因的次数与出现参考等位基因的次数的不同以及胎儿浓度的差异利用负二项分布公式得到。在本发明一种优选的实施例中，公式（4）中的 $P(S_i | G_{ij})$ 通过以下公式（7）计算得到：

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中， r 代表第 i 个 SNP 位点上出现等位基因的次数， k 第 i 个 SNP 位点上出现参考等位基因的次数， $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上出现突变等位基因的概率。

在上述实施例中， $f(b)$ 与孕妇外周血中胎儿游离 DNA 的浓度有关，利用常规的胎儿浓度的计算方法，如 Fetal Quant 进行计算即可（参见文献 Lench N, Barrett A, Fielding S, *et al.* The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made[J]. Prenatal diagnosis, 2013, 33(6): 555-562.）。在获得胎儿浓度之后，根据某位点为混合基因型中的某一种时，推测得到该混合基因型的父本可能的基因型，从而推算出某一特定混合基因型出现时，理论上胎儿出现突变等位基因的概率。

在本发明另一种优选的实施例中，初始胎儿浓度 f 采用最大期望算法进行迭代计算，假设初始胎儿浓度 f 为 10%，计算当 $f=10\%$ 时所有位点的混合基因型，利用混合基因型实际检测的参考基因型与突变等位基因的频率计算实际的胎儿浓度 f' （即第二胎儿浓度 f' ），如果 f' 与 f 之间的差值小于设定的值，则迭代结束，迭代结束时对应的 f' 即为胎儿浓度 f 。更优选，上述设定值为 0.001 时，迭代结束。具体采用的算法与前述检测方法中相同，在此不再赘述。

利用上述方法得到胎儿浓度后，便可计算上述第 i 个 SNP 位点的混合基因型为 G_{ij} 时，理论上出现等位基因的概率 $f(b)$ ，根据混合基因型 G_{ij} 的不同，分别按如下公式计算：当混合基因型 G_{ij} 为 G_{i1} 时， $f(b)$ 的值为 0；当混合基因型 G_{ij} 为 G_{i2} 时， $f(b)$ 的值为 $f/2$ ；当混合基因型 G_{ij} 为 G_{i3} 时， $f(b)$ 的值为 $0.5-f/2$ ；当混合基因型 G_{ij} 为 G_{i4} 时， $f(b)$ 的值为 0.5；当混合基因型 G_{ij} 为

G_{i5} 时, $f(b)$ 的值为 $0.5+f/2$; 当混合基因型 G_{ij} 为 G_{i6} 时, $f(b)$ 的值为 $1-f/2$; 当混合基因型 G_{ij} 为 G_{i7} 时, $f(b)$ 的值为 1; 其中, f 代表胎儿浓度, 具体计算方法同表 1 所示的方法, 此处不再赘述。

下面将结合实施例来进一步说明本发明的有益效果。

需要说明的是, 实施例 1 是按照图 2 所示的流程进行的。下列实施例中用的所有试剂除有特殊说明外, 其余均来自 NEB 公司; 所用方法如无特殊说明, 则本领域的常规方法均可实现。

实施例 1 一种检测基因突变的方法

实验一: 样品获取及游离 DNA 提取

- (1) 将孕妇离体外周血液置于离心机中, 1600g 速度下, 离心 10min, 之后吸取血浆。
- (2) 获取孕妇的离体外周血浆, 利用试剂盒 QIAamp DNA Blood Mini Kit (Qiagen, 德国, catlog # 51106) 提取血浆中的游离 DNA, 方法是该试剂盒中的血液和体液处理方法。

实验二: 捕获富集及文库构建

2.1 孕妇血浆中游离 DNA 的末端修复

实验目的: 由于从孕妇血浆中提取的游离 DNA 是双链 DNA 片段, 这些片段或是平末端的, 或是含有 3' 或 5' 突出末端。这一步通过 T4 DNA 聚合酶、大肠杆菌 DNA 多聚酶 I 大片段(Klenow 片段)和多核苷酸激酶 T4, 使突出端磷酸化成平末端。DNA 多聚酶 I 大片段的 3' 到 5' 外切酶活性去除 3' 突出末端, T4 DNA 聚合酶活性填充 5' 突出端。最终使游离 DNA 具有平末端。

实验材料、试剂及仪器: 实验一的游离 DNA; dNTPs 的混合物 (10mM); T4 DNA 聚合酶 (3 单位/ μ L); Klenow 片段 (5 单位/ μ L); T4 PNK (T4 多核苷酸激酶, 10 单位/ μ L) 及 PNK 缓冲液; DNA 纯化用磁珠; PCR 仪。

实验程序:

A. 配制如下反应体系:

孕妇血浆 DNA	40 μ L
dNTPs 的混合物(10mM)	2 μ L
T4 DNA 聚合酶	1 μ L
Klenow 片段	1 μ L
T4 PNK (T4 多核苷酸激酶)	1 μ L
PNK 缓冲液	5 μ L
总体积	160 μ L

B. 在 PCR 仪中孵育 30 分钟, 温度 20°C;

C. 用磁珠对反应后产物进行纯化, 用 19.5 μ L 的洗脱缓冲液 EB (Elution Buffer) 洗脱。

2.2 对游离 DNA 进行 3' 端加“A”碱基处理

实施目的: 由于后续的接头序列是需要与 3' 端具有单个“T”碱基突出端的磷酸化的 DNA 片段进行连接的, 因此, 需要先对末端修复后的片段的 3' 端添加互补的“A”碱基。该步骤由

缺失了 3' 到 5' 外切酶活性的 Klenow 片段来完成。

实验材料、试剂及仪器：末端修复后的 DNA；Klenow 缓冲液(10×)；dATP(1mM)；Klenow 片段(缺失 3' 到 5' 外切酶活性，5U/μl)；PCR 仪。

实验程序：

A.配制如下反应体系：

末端修复后的 DNA	19.5μL
Klenow 缓冲液（10×）	2.5μL
dATP（1mM）	2.5μL
Klenow 片段	0.5μL
总体积	25μL

B.在 PCR 中仪孵育 30 分钟，温度 37℃。

2.3 接头（Adapter）与 DNA 片段的两端进行连接

实验目的：为了使末端加 A 之后的 DNA 在后续 PCR 步骤中能够特异性扩增，需要用 DNA 连接酶在其两端各连接一个特异性接头（Adapter，即接头序列 1 与接头序列 2 的退火产物）。

实验材料、试剂及仪器：加 A 之后的 DNA；DNA 连接酶缓冲液（2×）；DNA 连接酶（1 U/μL）；接头序列 1 和接头序列 2；PCR 仪；DNA 纯化用磁珠。

其中，接头序列 1 的序列为：

SEQ ID NO:1: 5'-GATCGGAAGAGCACACGTCT-3'；

接头序列 2 的序列为：SEQ ID NO:2:

5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'（Illumina 公司）。

实验程序：

A. 配制如下反应体系：

2.2 步获得的 DNA	23μL
DNA 连接酶缓冲液（2×）	25μL
DNA 连接酶（1 单位/μL）	1μL
Adapter（20 pmol/μL）	1μL
总体积	50μL

B.在 PCR 中仪孵育 15 分钟，温度 20℃。

C.用磁珠对反应后产物进行纯化，用 38.2μL 的洗脱缓冲液 EB 洗脱。

2.4 对两端连接接头的 DNA 片段进行 PCR 扩增。

实验目的：对两端连接接头修饰过的 DNA 片段进行 PCR 扩增，一方面在 PCR 的过程中将 DNA 片段两端添加上与测序引物匹配的碱基序列，另一方面保证 DNA 片段有足够的总量用来进行接下来的测序。

实验材料、试剂及仪器：两端带有接头序列的 DNA；10×Pfx DNA 聚合酶扩增缓冲液；dNTP 的混合物（10 mM）；MgSO₄（50 mM）；PCR 引物 1（10 pmol/μL）；PCR 引物 2（10 pmol/μL）；Pfx DNA 聚合酶（2.5U/μL）。

其中，PCR 引物 1 的序列为 SEQ ID NO:3：

5'-AATGATACGGCGACCAACGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCT-3'；

PCR 引物 2 的序列为 SEQ ID NO:4：

5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'。

实验程序：

A. 配制如下反应体系：

2.3 步获得的 DNA	38.2μL
10×Pfx DNA 聚合酶扩增缓冲液	5μL
dNTP 的混合物（10mM）	2μL
MgSO ₄ （50mM）	2μL
PCR 引物 1（10pmol/μL）	1μL
PCR 引物 2（10pmol/μL）	1μL
Pfx DNA 聚合酶	0.8μL
总体积	50μL

B.按以下 PCR 程序进行扩增：第一步：94℃保持 2 分钟；第二步：94℃变性 15 秒；62℃退火 30 秒；72℃延伸 30 秒，进行 15 个循环；第三步：72℃保持 10 分钟；第四步，反应结束，4℃保存。

C.用磁珠对反应后产物进行纯化，用 ddH₂O 洗脱。

D.文库制备完成，用安捷伦的 DNA 检测仪 2100 进行浓度测定，测得浓度为 21.34ng/μL。

实验三 目标区域捕获

3.1 文库杂交

文库定量完成后，采用美国 Roche NimbleGen 公司的捕获试剂盒 SeqCap EZ Human Exome+UTR Kit（货号:06740308001）进行外显子捕获杂交。

实验材料、试剂：DNA 文库；SeqCap.EZ Exome+UTR.Library；Cot DNA；SeqCap.EZ.Hyb.and.Wash.Kit；HE oligo 序列和 TS-INV-HE index oligo 序列；

其中，HE oligo 序列为：SEQ ID NO:5：

5'-AATGATACGGCGACCAACGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCT-3'

TS-INV-HE index oligo 序列为：SEQ ID NO:6

5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'。

实验程序：

A.配制以下反应体系：

试剂	用量
2.4 步获得的 DNA 文库	1 μ g
Cot DNA	5 μ g
HE oligo	1000pmol (1 μ L of 1000 μ M)
TS-INV-HE index oligo	1000pmol (1 μ L of 1000 μ M)

B.上述完成后,使用真空浓缩仪 56℃干燥。

样本蒸干后,加入 7.5 μ L 2 $\times$ Hybridization Buffer 和 3 μ L Hybridization Component A,混匀后,95℃变性 10 分钟。

C.变性完成后,将上述混合物转移至 0.2mL 凸盖 PCR 管中,加入 4.5 μ L SeqCap.EZ Exome+UTR.Library。涡旋震荡 3 秒,充分混匀后最大转数离心 10 秒。

D.将杂交样品混合物置于 47℃的加热 64 至 72 小时,进行杂交。

3.2 捕获 DNA 的洗脱回收

实验试剂:链霉亲和素 Dynabeads; 1 $\times$ Stringent Wash Buffer; 1 $\times$ Wash Buffer I; 1 $\times$ Wash Buffer II; 1 $\times$ Wash Buffer III和 ddH₂O。

实验程序:

A.将杂交的样本文库转入盛有链霉亲和素 Dynabeads 的 0.2mL PCR 管中,吸打 10 次,将二者混匀;然后将 0.2mL PCR 管置于 47℃的加热模块中加热 45 分钟,每隔 15 分钟涡旋混匀一次,使 DNA 与磁珠结合。

B.孵育 45min 后,向 15 μ L 捕获的 DNA 样本中加入 47℃的 1 $\times$ Wash Buffer I 100 μ L。涡旋混匀 10sec。将 0.2mL PCR 管中的全部组分转入 1.5mL 离心管中。置于磁力架上吸附磁珠,弃上清。

然后将 1.5mL 离心管从磁力架上取下,加入 200 μ L 预热 47℃的 1 $\times$ Stringent Wash Buffer。吸打混匀 10 次。混匀后将样本置于 47℃加热模块上 5 分钟,用 47℃的 1 $\times$ Stringent Wash Buffer 共洗涤两次。将 1.5mL 的离心管置于磁力架上,吸附磁珠,弃上清。

向上述 1.5mL 离心管中加入 200 μ L 室温的 1 $\times$ Wash Buffer I, 涡旋混匀 2min。将离心管置于磁力架上,吸附磁珠,弃上清。

向上述 1.5mL 离心管中加入 200 μ L 室温的 1 $\times$ Wash Buffer II, 涡旋混匀 1min。将离心管置于磁力架上,吸附磁珠,弃上清。

向上述 1.5mL 离心管中加入 200 μ L 室温的 1 $\times$ Wash Buffer III, 涡旋混匀 30sec。将离心管置于磁力架上,吸附磁珠,弃上清。

将 1.5mL 离心管从磁力架上取下,加入 50 μ L ddH₂O,溶解洗脱磁珠捕获样本。将磁珠-样本混合物放在-20℃保存。

3.3 捕获 DNA 的 PCR 扩增

实验目的:由于捕获得到的 DNA 样本浓度很低,需要进行 PCR 扩增才能达到后续实验要求。

实验材料及试剂：杂交捕获的 DNA；10×Pfx DNA 聚合酶扩增缓冲液；dNTP 的混合物（10 mM）；MgSO₄（50mM）；PCR 引物 3（10 pmol/μL）（英潍捷基公司）；PCR 引物 4（10 pmol/μL）；Pfx DNA 聚合酶（2.5U/μL）。

其中，PCR 引物 3 的序列为：SEQ ID NO:7:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'；

PCR 引物 4 的序列为：SEQ ID NO:8:

5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

实验程序：

A. 配制以下反应体系：

捕获得到的 DNA	38.2μL
10×PFX DNA 聚合酶扩增缓冲液	5μL
dNTP 的混合物（10mM）	2μL
MgSO ₄ （50mM）	2μL
PCR 引物 1（10pmol/μL）	1μL
PCR 引物 2（10pmol/μL）	1μL
PFX DNA 聚合酶	0.8μL
总体积	50μL

B. 按以下 PCR 程序进行扩增：第一步：94℃保持 2 分钟；第二步：94℃变性 15 秒；62℃退火 30 秒；72℃延伸 30 秒，进行 13 个循环；第三步：72℃保持 10 分钟；第四步，反应结束，4℃保存。

C. 用磁珠对反应后产物进行纯化，用 30μL 的洗脱缓冲液 ddH₂O 洗脱。

D. 文库制备完成，用安捷伦 DNA 检测仪 2100 进行浓度测定，测得的浓度为 13.60ng/uL。

实验四 上样与测序

利用 Illumina 的 cBot 仪器将测序文库中 DNA 分子制成 DNA 簇，生成的 DNA 簇在 Illumina HiSeq 2000（或 Illumina HiSeq 2500）测序仪上进行 100 个循环的双端测序。

高通量测序（Illumina HiSeqTM2000）得到的原始图像数据文件经 CASAVA 碱基识别（Base Calling）分析转化为原始测序序列（Sequenced Reads），我们称之为 Raw Data 或 Raw Reads，结果以 FASTQ（简称为 fq）文件格式存储，其中包含测序序列（reads）的序列信息以及其对应的测序质量信息。

实验五 原始测序数据处理

5.1 不合格序列的过滤

经测序得到的原始序列中含有带接头的序列以及低质量的序列（质量值 Q≤5 的碱基数占整个 read 的 50% 以上的序列），为了保证分析结果的质量，必须对原始序列进行过滤，得到去除接头的测序质量合格的序列（又称为 clean reads），后续的分析都是基于过滤后的序列进

行的。其中，对如下序列进行过滤：(1)含 N 比例大于 5%的 reads；(2)低质量 reads（质量值 $Q \leq 5$ 的碱基数占整个 read 的 50%以上）；(3)接头污染的 reads。

对样本的原始数据统计结果如表 2 所示。其中 Modified Q30 bases rate (%) 指示过滤后总序列中质量值大于 30（错误率小于 0.1%）的碱基数的比例。该值越大说明测序质量越好。一般该指标大于 85%即认为测序质量合格。若小于 85%则重新测序。

表 2:

原始序列 (Raw reads)	有效序列 (Clean reads)	影响率 (Effect rate) (%)	Q30 (%)
140,070,440	136,958,780	97.78	86.76

Effect rate(%): 1.质量值 ≤ 5 的碱基数占整个 read 的 50%以上的序列；2.含 N 比例大于 5%的 reads；3.去掉接头污染的 reads。

5.2 比对质控

通过比对软件 BWA (bwa-0.7.5a) 将过滤后的 Clean Reads 比对参考基因组(HG19,NCBI built 37)上，对于比对结果进行比对质控，得到比对文件。

质控点包括数据的比对率，捕获特异性，目标区域测序深度，目标区域测序深度分布，以及 PCR 重复率等，比对质控的结果如表 3 所示。

表 3:

重复率 (%) Duplication rate)	比对率 (%) (Mapping rate)	特异捕获 (Specificity capture)	目标区域测序深度 (X) (Target Average depth)
5.54	97.44	72.38	139.78

上表 3 中，Specificity capture: reads 完全比对到设定区域以及 reads 部分在设定区域部分在区域外；Target Average depth (X): 目标区域的深度；Duplication rate (%): 由于 duplication 扩增出的 reads；Mapping rate (%): 利用 BWA 比对 hg19 参考基因组的数目与原始数据的比例；一般在 90%以上才能算正常结果。

然后，以上述比对文件为基础，去除无用数据（比如重复序列等），获得与参考基因组存在差异的碱基位点的点集；

最后，统计上述存在差异的碱基位点的测序结果，例如某个位点测到 40 个 A 和 60 个 T，以及其他的一些位点信息，比如位点所在的位置等。

实验六 假四倍体的混合基因型分型

下仅通过来源于上述孕妇外周血的遗传信息，直接对孕妇及胎儿的目标区域进行基因分型。根据本发明的假四分体分型模型，对由孕妇的基因型和胎儿的基因型所组成的假四分体的混合基因型进行推算，从而得到各个相应位点上孕妇与胎儿的基因型。下面以某位点的具体情况为例，说明推算出孕妇与胎儿基因型的过程。

假设该位点测到参考基因 50 次，等位基因 8 次，胎儿浓度 8%，群体突变频率 0.03。令公式 (4) 中的 G_{j*} 为 G_4 ，即得到公式 (4')

$$\varphi_j = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{i4})P(G_{i4})} \quad (4')$$

结合公式 (4')，至公式 (7) 和表 1，计算 7 种混合基因型与第四种混合基因型的概率比 φ ，从而得到概率比 φ 最高时所对应的混合基因，具体计算过程如下：

当混合基因型为 AAAA 时，

$$\varphi_1 = \frac{P(S_i | G_{i1})P(G_{i1})}{P(S_i | G_{i4})P(G_{i4})} = \frac{\binom{50+8-1}{8-1} 0^8 (1-0)^{50} \times (1-0.03)^2 (1-0.03)^2}{\binom{50+8-1}{8-1} 0.5^8 (1-0.5)^{50} \times (2 \times 0.03 \times (1-0.03))^2} ;$$

当混合基因型为 AAAB 时，

$$\varphi_2 = \frac{P(S_i | G_{i2})P(G_{i2})}{P(S_i | G_{i4})P(G_{i4})} = \frac{0.04^8 (1-0.04)^{50} \times (1-0.03)^2 \times 2 \times 0.03 \times (1-0.03)}{0.5^8 (1-0.5)^{50} \times (2 \times 0.03 \times (1-0.03))^2} ;$$

当混合基因型为 ABAA 时，

$$\varphi_3 = \frac{P(S_i | G_{i3})P(G_{i3})}{P(S_i | G_{i4})P(G_{i4})} = \frac{0.46^8 (1-0.46)^{50} \times (1-0.03)^2 \times 2 \times 0.03 \times (1-0.03)}{0.5^8 (1-0.5)^{50} \times (2 \times 0.03 \times (1-0.03))^2} ;$$

当混合基因型为 ABAB 时，

$$\varphi_4 = \frac{P(S_i | G_{i4})P(G_{i4})}{P(S_i | G_{i4})P(G_{i4})} = 1 ;$$

当混合基因型为 ABBB 时，

$$\varphi_5 = \frac{P(S_i | G_{i5})P(G_{i5})}{P(S_i | G_{i4})P(G_{i4})} = \frac{0.54^8 (1-0.54)^{50} \times (0.03)^2 \times 2 \times 0.03 \times (1-0.03)}{0.5^8 (1-0.5)^{50} \times (2 \times 0.03 \times (1-0.03))^2} ;$$

当混合基因型为 BBAB 时，

$$\varphi_6 = \frac{P(S_i | G_{i6})P(G_{i6})}{P(S_i | G_{i4})P(G_{i4})} = \frac{0.96^8 (1-0.96)^{50} \times (0.03)^2 \times 2 \times 0.03 \times (1-0.03)}{0.5^8 (1-0.5)^{50} \times (2 \times 0.03 \times (1-0.03))^2} ;$$

当混合基因型为 BBBB 时，

$$\varphi_7 = \frac{P(S_i | G_{i7})P(G_{i7})}{P(S_i | G_{i4})P(G_{i4})} = \frac{1^8 (1-1)^{50} \times (0.03)^2 (0.03)^2}{0.5^8 (1-0.5)^{50} \times (2 \times 0.03 \times (1-0.03))^2} ;$$

比较所有的 φ 值，有 $\varphi_2 > \varphi_3 > \varphi_4 > \varphi_5 > \varphi_6 > \varphi_7 = \varphi_1$ ，所以该位点最有可能的混合基因型为 AAAB 型。进而，推算出孕妇的基因型为 AA，胎儿的基因型为 AB。按照上述原理，得到比对结果文件中的所有存在差异的碱基位点 (SNP 位点) 的混合基因型，进而得到胎儿的基因型。

实验七 致病突变位点的确定

7.1 对非致病突变位点进行过滤

根据假四倍体分型结果得到的胎儿在各差异性碱基位点处的基因型，分别与如下数据库进行比对，并将属于下列数据库中的差异性碱基位点进行过滤：（1）dbSNP135 公共数据库中的高频突变；（2）Freq 1000g2012feb（千人基因组）数据库的多态性位点。（3）根据突变预测软件对同义突变、无义突变以及非保守区域的突变位点进行过滤，取上述三个筛选条件共有的位点，得到胎儿所特有的差异性碱基位点，如表 4 所示。

表 4:

总 SNP 数目	非同义突变位点	ASN_freq<0.05	dbSNP	差异性碱基位点
111407	100046	8622	100501	1049

7.2 根据公开文献以及临床资料，从上述过滤后的突变位点中确定导致致病性的突变位点。

实施例 2 是成骨发育不全胎儿的无创单基因病诊断的实例

样本信息：孕妇 28 岁，妊娠 3 次生产 0 次，平素月经规律，经期为 5-6 天，月经周期为 29 天。末次月经 2012-03-27，预产期 2013-01-04。自然受孕，早孕期无发热、皮疹等病史，无放射线毒物等接触史。孕 13 周查弓形虫、风疹、巨细胞、单纯疱疹病毒均阴性，孕 14+周 B 超测胎儿颈厚透明带宽度（NT）0.14cm，孕 17+周血清学筛查 21 三体风险<1:50000，18 三体风险<1:50000。孕 26+周保健医院 B 超提示胎儿股骨、胫骨发育异常，孕 26+周复查 B 超提示：胎儿颅骨薄、回声减低，双侧股骨长 3.3cm，弯曲成角，胫骨、腓骨弯曲成角。考虑胎儿股骨、胫腓骨成角。孕 30+周经产前诊断门诊鉴定考虑胎儿成骨发育不全，建议终止妊娠。

对孕妇抽取 10ml 静脉血，按照实施例 1 的方法提取 DNA，并捕获富集后，通过 HiSeq 平台，对富集后 DNA 进行高深度测序。

假四倍体分型：将测序结果进行质控，过滤掉低质量值的数据，将其余的数据比对到基因组上，并依据比对结果，通过本发明的假四倍体分型模型，推算得到胎儿的基因分型信息，并且筛选是否和成骨发育相关。

突变位点筛选：原始突变 111407 个，按照实施例 1 中的步骤对其中的突变位点进行过滤，最终筛选得到 7 个突变。对筛选到的 7 个突变进行文献筛查以及临床资料筛查，最终确定 1 个突变（COL1A1:NM_000088:c.G2596A:p.G866S）为导致成骨发育不全的致病突变。

结果验证：利用胎儿脐带血样本、孕妇外周血样本和父亲血液样本对本实施例的样本得到的致病位点进行验证，结果见图 3（图 3 中 M-F、F-F 或 C-F 分别表示母亲正链、父亲正链、胎儿正链）。从图 2 可以看出，胎儿在该位点确实含有致病位点。

实施例 3 是对本发明的假四倍体分型模型的验证

对上述孕妇静脉血样本，利用本发明的假四分体分型模型进行检测和分析，同时，用“胎儿脐血样本+孕妇外周血样本”的体细胞检测方式进行验证，来评估本发明的假四倍体分型模型的有效性。

结合孕妇体细胞测序结果和脐带血样本的体细胞测序结果，与假四倍体分型结果进行比较，假四倍体分型检测的结果如下表 5，与现有技术中的孕妇体细胞测序结果和脐带血样本的

体细胞测序结果相比的准确率情况如表 6。

表 5：假四倍体的混合基因型分型及检出率。

混合基因型	匹配的位点数/个	不匹配的位点数 ^A /个	总位点数 ^A /个	检出率/%
AAAA	89078219	2056	89080275	99.99
AAAB	10157	5654	15811	64.24
ABAA	9968	6051	16019	62.23
ABAB	16203	11711	27914	58.05
ABBB	4936	4077	9013	54.77
BBAB	7274	1646	8920	81.54
BBBB	28284	821	29105	97.18
总计 (无 AAAA)	76822	29960	106782	71.94

附：由于AAAA不含有突变基因型B，因此不计算在总位点数内。上表中的总位点数是通过分别对母亲和脐带血进行体细胞测序，获得的母亲比对上的碱基位数+胎儿比对上的碱基位点数的总和（母亲和胎儿重复的部分，只保留一份进行计算）。

总位点数^A：代表采用本发明的假四倍体分型模型检测出的总位点数。

匹配的位点数：即真阳性位点数，代表上述总位点数^A中通过本发明的假四倍体分型模型确定为真实突变的位点数。（采用母亲和脐带血进行体细胞测序是检测母亲和胎儿基因型的金标准，假四分体分型模型确定母亲和胎儿的基因型的方法为受试方法，本发明的受试方法与金标准相比，一致的位点为真阴性或真阳性位点；不一致的位点为假阳性或假阴性位点。）

不匹配的位点数：假阴性位点数，代表总位点数^A中通过本发明的假四倍体分型未确定的位点数，即假阴性位点数。

检出率：是临床试验上评价某一检测方法的标准指标，又称为阳性检出率，即真阳性/（真阳性+假阴性）×100%。检出率是针对受试方法而言的，即真阳性位点数与实际检测出的总位点数^A的比例。

表6：体细胞检测检测结果及准确率

混合基因型	匹配的位点数/个	不匹配的位点数 ^B /个	总位点数 ^B /个	准确率/%
AAAA	89078219	10897	89089116	99.99
AAAB	10157	1808	11965	84.89
ABAA	9968	5820	15788	63.14
ABAB	16203	8872	25075	64.62
ABBB	4936	2897	7833	63.02
BBAB	7274	575	7849	92.67
BBBB	28284	1384	29668	95.34

总计 (无 AAAA)	76822	21356	98178	78.25
----------------	-------	-------	-------	-------

准确率是检测某项检测方法实际检测到的阳性位点数与真正阳性位点数的比例。上表 6 中，总位点数^B：是按照孕妇体细胞测序结果和脐带血样本的体细胞测序结果所确定的真正具有阳性突变的位点数。匹配的位点数：即采用假四倍体分型模型实际确定的真阳性位点数。未匹配位点数：即采用假四倍体分型模型实际未能确定的真阳性位点数（即假阴性位点数）。因而，表 6 中的真阳性/（假阴性+真阳性）×100%代表的是准确率。

实施例 4

一种检测检测染色体非整倍性的装置，该装置包括：

检测模块，用于对孕妇外周血中的游离 DNA 进行高通量测序，得到测序数据；其中包括了对孕妇外周血中的游离 DNA 进行测序的仪器，可以包括 Illumina 的 cBot 仪器和 Illumina 的 Genome AnalXzer、HiSeq2000 测序仪或 HiSeq2500 测序仪或 ABI 公司的 SOLiD 系列的测序仪。

优选地，检测模块还包括区域捕获子模块，区域捕获子模块用于对富集的游离 DNA 制成的测序文库进行目标区域捕获，得到用于高通量测序的测序文库。

比对模块，用于将测序数据与参考基因组序列进行比对得到 SNP 位点；

目标混合基因型确定模块，用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型，并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型；其中，混合基因型是指孕妇和胎儿形成的假四倍体的基因型，混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 所组成的七种中的任一种，且将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点突变的等位基因型；混合基因型分型根据条件概率和贝叶斯模型，推算得到胎儿在各 SNP 位点的目标混合基因型。

优选地，当初始胎儿浓度不是真实胎儿浓度时，目标混合基因型确定模块包括：预估模块、选取模块、计算模块、比较模块以及判断模块，其中，预估模块用于采用贝叶斯模型和初始胎儿浓度 f 分别计算各 SNP 位点的 7 种混合基因型的概率，得到各 SNP 位点的最大概率的混合基因型，并将该最大概率的混合基因型作为初始混合基因型；选取模块用于选取方便计算第二胎儿浓度的初始混合基因型，记为第二混合基因型；计算模块，用于根据第二混合基因型及测序数据，计算得到第二胎儿浓度 f' ；比较模块，用于比较第二胎儿浓度 f' 与初始胎儿浓度 f ，得到差值 Δf ；判断模块，用于判断 Δf 是否大于设定值，并且当 Δf 大于设定值时，利用循环模块将 f' 作为 f 重复执行预估模块、选取模块、计算模块、比较模块以及判断模块；当 Δf 在设定值以下时，利用标记模块将此时各 SNP 位点的初始混合基因型记为目标混合基因型。

优选地，初始胎儿浓度为 10%；更优选地，设定值为 0.001；进一步优选地，胎儿浓度计算混合基因型选自如下四种混合基因型中的任意一种或多种：AAAB、ABAA、ABBB 以及 BBAB。

优选地，目标混合基因型确定模块采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型的

步骤块包括：根据七种混合基因型的条件概率和为 1 得到下列公式 (1)，其

$$\sum P(G_j | S) = 1 \quad (1)$$

中 G_j 表示七种混合基因型中的任一种， S 表示各 SNP 位点中的某一位点， $P(G_j | S)$ 表示某一 SNP 位点如 S 一样时，SNP 位点的混合基因型是 G_j 的概率；由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中， $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 基因型的概率， j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7；从 G_j 中选取任意一种混合基因型 G_{j*} 为基准，则由公式 (2) 得出以下公式 (3)：

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除，得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中， φ_j 表示在 S_i 条件下，第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值； $P(G_{ij})$ 通过群体突变频率计算得到， $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度，通过负二项分布公式得到；然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型，将出现概率最大的混合基因型记为第 i 个 SNP 位点上的最大概率的混合基因型。

优选地，上述公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到，其中， θ 为第 i 个 SNP 位点的群体突变频率。

优选地，胎儿基因型确定模块中的公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到：

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因的次数, k 第 i 个 SNP 位点上出现参考等位基因的次数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上胎儿出现突变等位基因的概率。

优选地, 第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上胎儿出现突变等位基因的概率 $f(b)$, 根据混合基因型 G_{ij} 的不同, 分别按如下方式计算: 当混合基因型 G_{ij} 为 G_{i1} 时, $f(b)$ 的值为 0; 当混合基因型 G_{ij} 为 G_{i2} 时, $f(b)$ 的值为 $f/2$; 当混合基因型 G_{ij} 为 G_{i3} 时, $f(b)$ 的值为 $0.5-f/2$; 当混合基因型 G_{ij} 为 G_{i4} 时, $f(b)$ 的值为 0.5; 当混合基因型 G_{ij} 为 G_{i5} 时, $f(b)$ 的值为 $0.5+f/2$; 当混合基因型 G_{ij} 为 G_{i6} 时, $f(b)$ 的值为 $1-f/2$; 当混合基因型 G_{ij} 为 G_{i7} 时, $f(b)$ 的值为 1; 其中, f 代表胎儿浓度, 胎儿浓度采用最大期望算法进行迭代计算得到, 假设初始 f 为 10%, 计算当 $f=10\%$ 时所有位点的混合基因型, 利用混合基因型实际检测的参考基因型与突变等位基因的频率计算的胎儿浓度 f' , 如果 f' 与 f 之间的差值小于设定值, 则迭代结束, 迭代结束时对应的 f' 即为胎儿浓度 f 。更优选地, 当采用最大期望算法对胎儿浓度进行迭代计算时, 当设定值 ≤ 0.001 时, 则迭代结束。

突变位点筛选模块, 用于根据胎儿的各 SNP 位点的基因型, 从各 SNP 位点中筛选出突变位点。

优选地, 突变位点筛选模块包括: 高发多态性位点过滤子模块: 用于将推算出胎儿基因型的各 SNP 位点中属于人群高发的多态性位点进行过滤, 得到初步候选突变位点; 例如通过利用目前医学界已经整理的 **dnSNP135** 公共数据库和 **Freq_1000g2012feb** (千人基因组) 数据库, 去除在人群中的高发的多态性位点, 得到胎儿所特有的 SNP 位点, 然后利用基因突变位点筛选子模块筛选真正导致基因突变的位点。

基因突变位点筛选子模块: 用于从初步候选突变位点中的导致同义突变、无义突变以及发生在非保守区域的 SNP 位点过滤掉, 得到候选突变位点。该模块可以利用本领域常用的突变预测模块进行有害突变筛选。比如 **ANNOVAR** 模块, 既可以筛选突变是否导致氨基酸改变, 即筛选突变是否导致有义突变, 也可以筛选突变是否发生在保守序列区域。

文献及临床资料筛查子模块: 用于从候选突变位点中筛查出文献及临床资料里记载过的导致致病基因突变的突变位点。将突变预测模块筛选之后的 SNP 位点与已有的数据库和文献的检索到的致病位点进行比对, 从而找到与单基因遗传病相关的位点信息并进行相应解读。不仅检测是否存在导致已知单基因遗传病的热点突变位点, 还可以检测已知单基因遗传病的非热点突变位点以及未报道过的潜在的致病基因及其突变位点。

实施例 5

一种用于对孕妇和胎儿的基因型进行分型的试剂盒, 该试剂盒包括:

从孕妇外周血血浆中富集游离 DNA 进行高通量测序的试剂和器械; 检测试剂可以包括游离 DNA 提取、分离、富集、检测、文库构建等各步骤所用到的各种试剂或化学药品; 检测器械可以包括 1.5ml EPP 管、PCR 管、移液枪、上机所用的 96 孔板以及上机所用高通量测序仪等;

用于将高通量测序得到的测序数据与参考基因组比对得到 SNP 位点的器械；包括各种存储在特定存储介质上，并利用计算机终端或移动终端来执行上述比对功能的硬件模块。以及

用于采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型，并将最大概率的混合基因型作为各 SNP 位点的目标混合基因型的器械；其中，假四倍体是指孕妇和胎儿所组成的假四倍体，假四倍体的基因型记为混合基因型，混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 所组成的七种中的任一种，且将 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各 SNP 位点的参考基因型，B 表示各 SNP 位点突变的等位基因型。

其中，当初始胎儿浓度不是真实胎儿浓度时，获取各 SNP 位点的目标混合基因型的器械包括：第一计算元件，用于采用贝叶斯模型和初始胎儿浓度 f 分别计算各 SNP 位点的 7 种混合基因型的概率，得到概率最大的混合基因型，并将概率最大的混合基因型记为各 SNP 位点的初始混合基因型；选取元件，用于选取方便计算第二胎儿浓度的初始混合基因型，记为胎儿浓度计算混合基因型；第二计算元件，用于根据胎儿浓度计算混合基因型及测序数据，计算得到第二胎儿浓度 f' ；比较元件，用于比较第二胎儿浓度 f' 与初始胎儿浓度 f ，得到差值 Δf ；以及判断元件，用于判断 Δf 是否大于设定值，并且当 Δf 大于设定值时，利用循环元件将 f' 作为 f 重复操作第一计算元件、选取元件、第二计算元件、比较元件以及判断元件；当 Δf 在设定值以下时，利用标记元件将此初始胎儿浓度 f 所对应的初始混合基因型记为目标混合基因型。

优选地，获取目标混合基因型的器械中采用贝叶斯模型和初始胎儿浓度 f 对各 SNP 位点进行混合基因型分型，得到各 SNP 位点的七种混合基因型中最大概率的混合基因型的步骤包括：根据七种混合基因型的条件概率和为 1 得到下列公式 (1)，其中 G_j 表示七种混合基因型中的任一种，S 表示各 SNP 位点中的某一位点，

$$\sum P(G_j | S) = 1 \quad (1)$$

$P(G_j | S)$ 表示某一 SNP 位点如 S 一样时，SNP 位点的混合基因型是 G_j 的概率；由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在公式 (2) 中， $P(G_{ij})$ 表示在第 i 个 SNP 位点上出现 G_j 基因型的概率， j 值与按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7；从 G_j 中选取任意一种混合基因型 G_{j*} 为基准，则由公式 (2) 得出以下公式 (3)：

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除, 得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中, φ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各 SNP 位点上出现的突变等位基因的个数、与突变等位基因相应的参考等位基因的个数以及初始胎儿浓度, 通过负二项分布公式得到; 然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种混合基因型中出现概率最大的混合基因型, 将出现概率最大的混合基因型记为第 i 个 SNP 位点上的最大概率的混合基因型。

优选地, 公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, θ 为第 i 个 SNP 位点的群体突变频率。

优选地, 公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因的个数, k 第 i 个 SNP 位点上出现参考等位基因的个数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上胎儿出现突变等位基因的概率。

优选地, 第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上胎儿出现突变等位基因的概率 $f(b)$, 根据混合基因型 G_{ij} 的不同, 分别按如下方式计算: 当混合基因型 G_{ij} 为 G_{i1} 时, $f(b)$ 的值为 0; 当混合基因型 G_{ij} 为 G_{i2} 时, $f(b)$ 的值为 $f/2$; 当混合基因型 G_{ij} 为 G_{i3} 时, $f(b)$ 的值为 $0.5-f/2$; 当混合基因型 G_{ij} 为 G_{i4} 时, $f(b)$ 的值为 0.5; 当混合基因型 G_{ij} 为 G_{i5} 时, $f(b)$ 的值为 $0.5+f/2$; 当混合基因型 G_{ij} 为 G_{i6} 时, $f(b)$ 的值为 $1-f/2$; 当混合基因型 G_{ij} 为 G_{i7} 时, $f(b)$ 的值为 1; 其中, f 代表胎儿浓度, 胎儿浓度采用最大期望算法进行迭代计算得到, 假设初始 f 为 10%, 计算当 $f=10\%$ 时所有位点的混合基因型, 利用混合基因型实际检测的参考基因型与突变等位基因的频率计算的胎儿浓度 f' , 如果 f' 与 f 之间的差值小于设定值, 则迭代结束, 迭代结束时对应的 f' 即为胎儿浓度 f 。更优选地, 当采用最大期望算法对胎儿浓度进行迭代计算时, 当设定值 ≤ 0.001 时, 则迭代结束。

上述利用假四倍体的混合基因型分型推算出各 SNP 位点的胎儿基因型的器械, 包括各种

存储在特定存储介质上，并利用计算机终端或移动终端来执行上述计算、判断或确认功能的硬件模块。上述各种计算方式作为器械的一部分，能够单独执行或组装成器械执行上述计算功能，因此负载或存储上述计算方式的部件也是器械的一个组成。

从以上的描述中，可以看出，本发明的无创产前诊断基因突变的方法、装置及试剂盒，无需借助胎儿父亲或者母亲自身的遗传信息，仅通过孕妇静脉血遗传信息即可推断胎儿的基因型，并判断是否导致相应疾病。简化了无创单基因病的检测程序，降低了检测成本。此外，本发明不仅可以对特定的单基因病进行检测，而且可以同时多种单基因病进行检测。

显然，本领域的技术人员应该明白，上述的本申请的一些模块、元件或一些步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本申请不限制于任何特定的硬件和软件结合。

以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

1. 一种检测基因突变的方法，其特征在于，所述方法包括以下步骤：

步骤 A，将孕妇外周血中的游离 DNA 进行高通量测序，得到测序数据；

步骤 B，将所述测序数据与参考基因组进行比对，得到 SNP 位点；

步骤 C，采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型，并将所述最大概率的混合基因型作为各所述 SNP 位点的目标混合基因型；以及

步骤 D，根据所述目标混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点；

其中，所述混合基因型是指孕妇和胎儿组成的假四倍体的基因型，所述混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种，将所述 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各所述 SNP 位点的参考基因型，B 表示各所述 SNP 位点的突变基因型。

2. 根据权利要求 1 所述的方法，其特征在于，当所述初始胎儿浓度不是真实胎儿浓度时，获取所述目标混合基因型的步骤包括：

步骤 C1，采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型，并将所述最大概率的混合基因型作为各所述 SNP 位点的初始混合基因型；

步骤 C2，选取方便计算第二胎儿浓度的所述初始混合基因型作为第二混合基因型；

步骤 C3，根据所述第二混合基因型及所述测序数据，计算得到第二胎儿浓度 f' ；

步骤 C4，比较所述第二胎儿浓度 f' 与所述初始胎儿浓度 f ，得到差值 Δf ；

步骤 C5，判断所述差值 Δf 与设定值的关系；

步骤 C6，当 Δf 大于设定值时，将所述 f' 作为 f 重复步骤 C1 至步骤 C5；当 Δf 在设定值以下时，所述初始胎儿浓度 f 对应的所述初始混合基因型即为所述目标混合基因型。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，所述采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型的步骤包括：

根据七种所述混合基因型的条件概率和为 1 得到下列公式 (1)，

$$\sum P(G_j | S) = 1 \quad (1)$$

其中, G_j 表示七种所述混合基因型中的任一种, S 表示各所述 SNP 位点中的某一位点, $P(G_j | S)$ 表示在 S 条件下, 所述 SNP 位点的混合基因型是 G_j 的概率;

由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在所述公式 (2) 中, $P(G_{ij})$ 表示在第 i 个所述 SNP 位点上出现 G_j 的概率, j 值与所述按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7;

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准混合基因型, 则由所述公式 (2) 得出以下公式 (3):

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除, 得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中, φ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各所述 SNP 位点上出现的突变等位基因的个数、与所述突变等位基因相应的参考等位基因的个数以及所述初始胎儿浓度 f , 通过负二项分布公式得到;

然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种所述混合基因型中出现概率最大的混合基因型, 将所述出现概率最大的混合基因型记为第 i 个 SNP 位点上的最大概率的混合基因型。

4. 根据权利要求 3 所述的方法, 其特征在于, 所述公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

中孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, ϑ 为第 i 个 SNP 位点的群体突变频率。

5. 根据权利要求 3 所述的方法, 其特征在于, 所述公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因型的次数, k 代表第 i 个 SNP 位点上出现参考等位基因型的次数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上出现突变等位基因型的概率。

6. 根据权利要求 5 所述的方法, 其特征在于, 所述第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上出现突变等位基因的概率 $f(b)$, 根据所述混合基因型 G_{ij} 的不同, 所述概率 $f(b)$ 分别按如下方式计算:

当所述混合基因型 G_{ij} 为 G_{i1} 时, 所述 $f(b)$ 的值为 0;

当所述混合基因型 G_{ij} 为 G_{i2} 时, 所述 $f(b)$ 的值为 $f/2$;

当所述混合基因型 G_{ij} 为 G_{i3} 时, 所述 $f(b)$ 的值为 $0.5-f/2$;

当所述混合基因型 G_{ij} 为 G_{i4} 时, 所述 $f(b)$ 的值为 0.5;

当所述混合基因型 G_{ij} 为 G_{i5} 时, 所述 $f(b)$ 的值为 $0.5+f/2$;

当所述混合基因型 G_{ij} 为 G_{i6} 时, 所述 $f(b)$ 的值为 $1-f/2$;

当所述混合基因型 G_{ij} 为 G_{i7} 时, 所述 $f(b)$ 的值为 1;

其中, 所述 f 代表所述初始胎儿浓度。

7. 根据权利要求 2 所述的方法, 其特征在于, 所述初始胎儿浓度为预估胎儿浓度, 优选所述预估胎儿浓度为 10%; 优选所述设定值 ≤ 0.001 。
8. 根据权利要求 2 所述的方法, 其特征在于, 所述第二混合基因型选自如下四种混合基因型中的任意一种或两种以上: AAAB、ABAA、ABBB 以及 BBAB。
9. 根据权利要求 1 所述的方法, 其特征在于, 从所述混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点的步骤包括:

将各 SNP 位点的所述真实混合基因型中的胎儿基因型中属于人群中高发的多态性位点进行过滤, 得到初步候选突变位点;

对所述初步候选突变位点中导致同义突变、无义突变以及发生在非保守区域的 SNP 位点进行过滤, 得到候选突变位点;

对所述候选突变位点进行文献筛查和临床资料筛查，得到导致所述胎儿基因突变的所述突变位点。

10. 一种检测基因突变的装置，其特征在于，所述装置包括：

检测模块，用于对孕妇外周血中的游离 DNA 进行高通量测序，得到测序数据；

比对模块，用于将所述测序数据与参考基因组序列进行比对，得到 SNP 位点；

目标混合基因型确定模块，用于采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型，并将所述最大概率的混合基因型作为各所述 SNP 位点的目标混合基因型；以及

突变位点筛选模块，用于根据各所述 SNP 位点的所述目标混合基因型中的胎儿基因型筛选出导致胎儿基因突变的突变位点；

其中，所述混合基因型是指孕妇和胎儿组成的假四倍体的基因型，所述混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种，将所述 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各所述 SNP 位点的参考基因型，B 表示各所述 SNP 位点的突变基因型。

11. 根据权利要求 10 所述的装置，其特征在于，当所述初始胎儿浓度不是真实胎儿浓度时，所述目标混合基因型确定模块包括：

预估模块，用于采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型，并将所述概率最大的混合基因型作为初始混合基因型；

选取模块，用于选取方便计算第二胎儿浓度的所述初始混合基因型作为第二混合基因型；

计算模块，用于根据所述第二混合基因型及所述测序数据，计算得到第二胎儿浓度 f' ；

比较模块，用于比较所述第二胎儿浓度 f' 与所述初始胎儿浓度 f ，得到差值 Δf ；

判断模块，用于判断所述差值 Δf 与设定值的关系；

循环模块，用于当 Δf 大于设定值时，将所述 f' 作为 f 重复执行所述预估模块、所述选取模块、所述计算模块、所述比较模块以及所述判断模块；

标记模块，用于当 Δf 在设定值以下时，将所述初始胎儿浓度 f 对应的所述初始混合基因型记为所述目标混合基因型。

12. 根据权利要求 10 或 11 所述的装置，其特征在于，所述目标混合基因型确定模块采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型的步骤包括：

根据七种所述混合基因型的条件概率和为 1 得到下列公式 (1)，其中 G_j 表示七种所述混合基因型中的任一种，

$$\sum P(G_j | S) = 1 \quad (1)$$

S 表示各所述 SNP 位点中的某一位点， $P(G_j | S)$ 表示某一 SNP 位点如 S 一样时，所述 SNP 位点的混合基因型是 G_j 的概率；

由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在所述公式 (2) 中， $P(G_{ij})$ 表示在第 i 个所述 SNP 位点上出现 G_j 基因型的概率， j 值与所述按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7；

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准，则由所述公式 (2) 得出以下公式 (3)：

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除，得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中， φ_j 表示在 S_i 条件下，第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值； $P(G_{ij})$ 通过群体突变频率计算得到， $P(S_i | G_{ij})$ 通过利用各所述 SNP 位点上出现的突变等位基因的个数、与所述突变等位基因相应的参考等位基因的个数以及所述初始胎儿浓度 f ，通过负二项分布公式得到；

然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种所述混合基因型中出现概率最大的混合基因型，将所述出现概率最大的混合基因型记为第 i 个 SNP 位点上的所述最大概率的混合基因型。

13. 根据权利要求 12 所述的装置，其特征在于，所述公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1-\theta)^2 \\ P(G' = AB) = 2\theta(1-\theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

中孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, θ 为第 i 个 SNP 位点的群体突变频率。

14. 根据权利要求 12 所述的装置, 其特征在于, 所述公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因的个数, k 第 i 个 SNP 位点上出现参考等位基因的个数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上胎儿出现突变等位基因的概率。

15. 根据权利要求 14 所述的装置, 其特征在于, 所述第 i 个 SNP 位点的混合基因型为 G_{ij} 时理论上胎儿出现突变等位基因的概率 $f(b)$, 根据所述混合基因型 G_{ij} 的不同, 分别按如下方式计算:

当所述混合基因型 G_{ij} 为 G_{i1} 时, 所述 $f(b)$ 的值为 0;

当所述混合基因型 G_{ij} 为 G_{i2} 时, 所述 $f(b)$ 的值为 $f/2$;

当所述混合基因型 G_{ij} 为 G_{i3} 时, 所述 $f(b)$ 的值为 $0.5 \cdot f/2$;

当所述混合基因型 G_{ij} 为 G_{i4} 时, 所述 $f(b)$ 的值为 0.5;

当所述混合基因型 G_{ij} 为 G_{i5} 时, 所述 $f(b)$ 的值为 $0.5 + f/2$;

当所述混合基因型 G_{ij} 为 G_{i6} 时, 所述 $f(b)$ 的值为 $1 - f/2$;

当所述混合基因型 G_{ij} 为 G_{i7} 时, 所述 $f(b)$ 的值为 1;

其中, f 代表所述初始胎儿浓度。

16. 根据权利要求 11 所述的装置, 其特征在于, 所述预估模块中的所述初始胎儿浓度为预估胎儿浓度, 优选地, 所述预估胎儿浓度为 10%, 更优选, 所述判断模块中的所述设定值 ≤ 0.001 。
17. 根据权利要求 11 所述的装置, 其特征在于, 所述计算模块中的所述第二混合基因型选自如下四种混合基因型中的任意一种或多种: AAAB、ABAA、ABBB 以及 BBAB。
18. 根据权利要求 11 所述的装置, 其特征在于, 所述突变位点筛选模块包括:

高发多态性位点过滤子模块, 用于将各 SNP 位点的所述目标混合基因型中的胎儿基

因型中属于人群高发的多态性位点进行过滤，得到初步候选突变位点；

基因突变位点筛选子模块，用于从所述初步候选突变位点中的导致同义突变、无义突变以及发生在非保守区域的 SNP 位点过滤掉，得到候选突变位点；

文献及临床资料筛查子模块，用于对所述候选突变位点进行文献筛查和临床资料筛查，得到导致所述胎儿基因突变的所述突变位点。

19. 一种用于对孕妇和胎儿的基因型进行分型的试剂盒，其特征在于，所述试剂盒包括：

用于从孕妇外周血血浆中富集游离 DNA 进行高通量测序的试剂和器械；

用于将所述高通量测序得到的测序数据与参考基因组比对得到 SNP 位点的器械；以及

用于采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型，并将所述最大概率的混合基因型作为各所述 SNP 位点的目标混合基因型的器械；

其中，所述混合基因型是指孕妇和胎儿组成的假四倍体的基因型，所述混合基因型为 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 七种中的任意一种，将所述 AAAA、AAAB、ABAA、ABAB、ABBB、BBAB 和 BBBB 按顺序依次编号为第 1 种、第 2 种、第 3 种、第 4 种、第 5 种、第 6 种和第 7 种，A 表示各所述 SNP 位点的参考基因型，B 表示各所述 SNP 位点突变基因型。

20. 根据权利要求 19 所述的试剂盒，其特征在于，当所述初始胎儿浓度不是真实胎儿浓度时，所述获取各所述 SNP 位点的目标混合基因型的器械包括：

第一计算元件，用于采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型，得到各所述 SNP 位点的 7 种所述混合基因型中概率最大的混合基因型，并将所述概率最大的混合基因型作为初始混合基因型；

选取元件，用于选取方便计算第二胎儿浓度的所述初始混合基因型，记为第二混合基因型；

第二计算元件，用于根据所述第二混合基因型及所述测序数据，计算得到第二胎儿浓度 f' ；

比较元件，用于比较所述第二胎儿浓度 f' 与所述初始胎儿浓度 f ，得到差值 Δf ；

判断元件，用于判断所述 Δf 是否大于设定值；

循环元件，用于当 Δf 大于设定值时，将所述 f' 作为 f 重复操作所述第一计算元件、所述选取元件、所述第二计算元件、所述比较元件以及所述判断元件；以及

标记元件，用于当 Δf 在设定值以下时，将所述初始胎儿浓度 f 对应的所述初始混合基因型标记为所述目标混合基因型。

21. 根据权利要求 19 或 20 所述的试剂盒, 其特征在于, 所述获取目标混合基因型的器械中采用贝叶斯模型和初始胎儿浓度 f 对各所述 SNP 位点进行混合基因型分型, 得到各所述 SNP 位点的七种所述混合基因型中最大概率的混合基因型的步骤包括:

根据七种所述混合基因型的条件概率和为 1 得到下列公式 (1), 其中 G_j 表示七种所

$$\sum P(G_j | S) = 1 \quad (1)$$

述混合基因型中的任一种, S 表示各所述 SNP 位点中的某一位点, $P(G_j | S)$ 表示某一 SNP 位点如 S 一样时, 所述 SNP 位点的混合基因型是 G_j 的概率;

由贝叶斯模型得到下列公式 (2)

$$P(G_{ij} | S_i) = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i)} \quad (2)$$

在所述公式 (2) 中, $P(G_{ij})$ 表示在第 i 个所述 SNP 位点上出现 G_j 基因型的概率, j 值与所述按顺序依次编号的混合基因型相对应分别为 1、2、3、4、5、6 或 7;

从 G_j 中选取任意一种混合基因型 G_{j*} 为基准, 则由所述公式 (2) 得出以下公式 (3):

$$P(G_{ij*} | S_i) = \frac{P(S_i | G_{ij*})P(G_{ij*})}{P(S_i)} \quad (3)$$

将公式 (2) 与公式 (3) 两边相除, 得到下列公式 (4)

$$\varphi_j = \frac{P(G_{ij} | S_i)}{P(G_{ij*} | S_i)} = \frac{P(S_i | G_{ij})P(G_{ij})}{P(S_i | G_{ij*})P(G_{ij*})} \quad (4)$$

其中, φ_j 表示在 S_i 条件下, 第 i 个 SNP 位点的混合基因型为 G_j 的概率与第 i 个 SNP 位点的混合基因型为 G_{j*} 的概率的比值; $P(G_{ij})$ 通过群体突变频率计算得到, $P(S_i | G_{ij})$ 通过利用各所述 SNP 位点上出现的突变等位基因的个数、与所述突变等位基因相应的参考等位基因的个数以及所述初始胎儿浓度, 通过负二项分布公式得到;

然后由下列公式 (5)

$$G = \arg \max(\varphi_j) \quad (5)$$

找到七种所述混合基因型中出现概率最大的混合基因型, 将所述出现概率最大的混合基因型记为第 i 个 SNP 位点上的所述最大概率的混合基因型。

22. 根据权利要求 21 所述的试剂盒, 其特征在于, 所述公式 (4) 中的 $P(G_{ij})$ 通过以下公式 (6)

$$\begin{cases} P(G' = AA) = (1 - \theta)^2 \\ P(G' = AB) = 2\theta(1 - \theta) \\ P(G' = BB) = \theta^2 \end{cases} \quad (6)$$

中孕妇的基因型 G' 的概率与胎儿的基因型 G' 的概率相乘得到, 其中, θ 为第 i 个 SNP 位点的群体突变频率。

23. 根据权利要求 20 所述的试剂盒, 其特征在于, 所述公式 (4) 中的 $P(S_i | G_{ij})$ 通过以下公式 (7) 计算得到:

$$P(S_i | G_{ij}) = \binom{k+r-1}{r-1} \times f(b)^r \times (1-f(b))^k \quad (7)$$

其中, r 代表第 i 个 SNP 位点上出现突变等位基因型的次数, k 代表第 i 个 SNP 位点上出现参考等位基因型的次数, $f(b)$ 代表第 i 个 SNP 位点的混合基因型为 G_{ij} 时, 理论上出现突变等位基因型的概率。

24. 根据权利要求 23 所述的试剂盒, 其特征在于, 所述公式 (7) 中的 $f(b)$, 根据所述混合基因型 G_{ij} 的不同, 分别按如下公式计算:

当所述混合基因型 G_{ij} 为 G_{i1} 时, 所述 $f(b)$ 的值为 0;

当所述混合基因型 G_{ij} 为 G_{i2} 时, 所述 $f(b)$ 的值为 $f/2$;

当所述混合基因型 G_{ij} 为 G_{i3} 时, 所述 $f(b)$ 的值为 $0.5 \cdot f/2$;

当所述混合基因型 G_{ij} 为 G_{i4} 时, 所述 $f(b)$ 的值为 0.5;

当所述混合基因型 G_{ij} 为 G_{i5} 时, 所述 $f(b)$ 的值为 $0.5 + f/2$;

当所述混合基因型 G_{ij} 为 G_{i6} 时, 所述 $f(b)$ 的值为 $1 - f/2$;

当所述混合基因型 G_{ij} 为 G_{i7} 时, 所述 $f(b)$ 的值为 1;

其中, 所述 f 代表所述初始胎儿浓度。

25. 根据权利要求 20 所述的试剂盒, 其特征在于, 所述预估元件中的所述初始胎儿浓度为预估胎儿浓度, 优选地所述预估胎儿浓度为 10%, 优选所述判断元件中的所述设定值 ≤ 0.001 。
26. 根据权利要求 20 所述的试剂盒, 其特征在于, 所述第二计算元件中的所述第二混合基因型选自如下四种混合基因型中的任意一种或多种: AAAB、ABAA、ABBB 以及 BBAB。

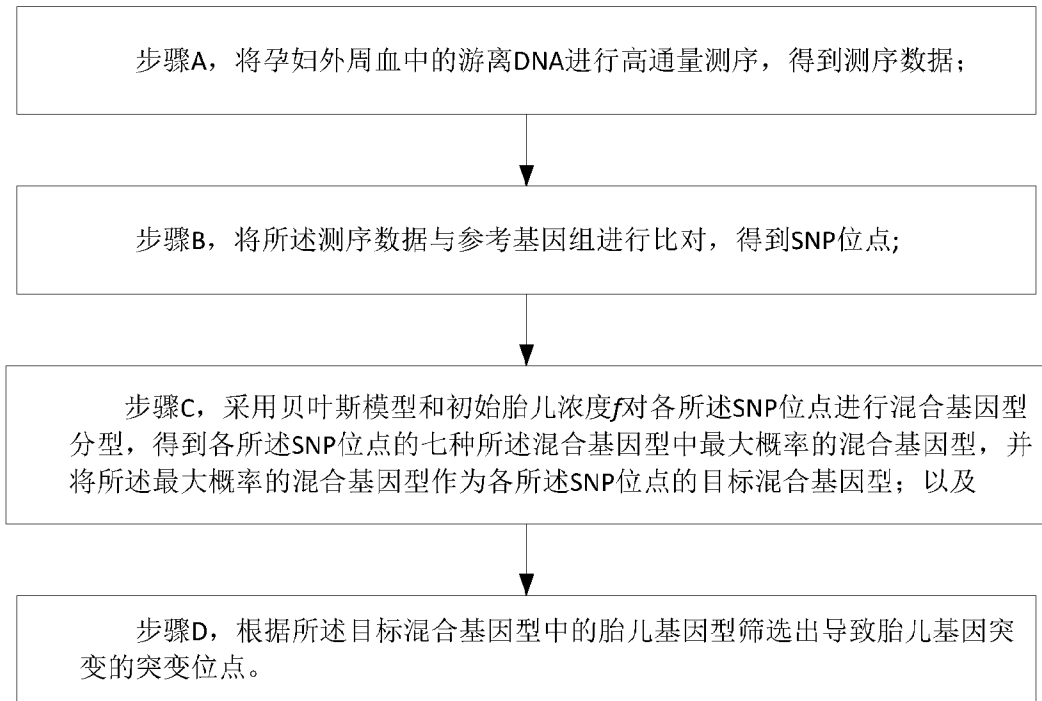


图 1

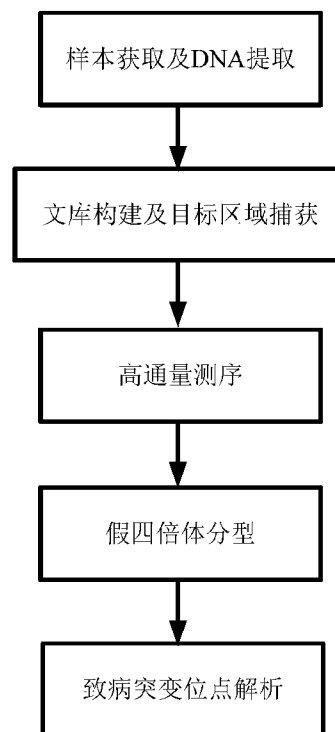


图 2

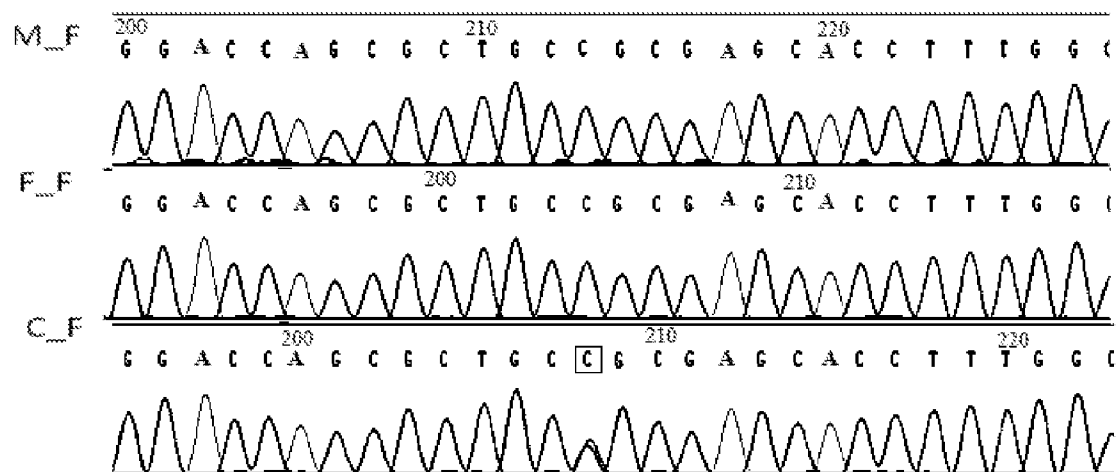


图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/118213

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2018.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNMED, CPRSABS, CNABS, CPEA, DWPI, SIPOABS, AUABS, TWMED, ILABS, TWABS, HKABS, MOABS, SGABS, CNKI, NCBI, EMBL, GoogleScholar: 无创产前诊断, 贝叶斯模型, 基因, SNP, non-invasive prenatal diagnosis (NIPD), Bayesian, gene

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103874767 A (SHENZHEN BGI COLLEGE, SHENZHEN BGI DIAGNOSIS CO., LTD.) 18 June 2014 (18.06.2014), entire document	1-26
A	WO 2008088804 A1 (BRIGHAM & WOMENS HOSPITAL et al.) 24 July 2008 (24.07.2008), entire document	1-26
A	WO 2016183106 A1 (NATERRA INC.) 17 November 2016 (17.11.2016), entire document	1-26

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	
“A” document defining the general state of the art which is not considered to be of particular relevance	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“E” earlier application or patent but published on or after the international filing date	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	“&” document member of the same patent family

Date of the actual completion of the international search
01 February 2018

Date of mailing of the international search report
22 March 2018

Name and mailing address of the ISA
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No. (86-10) 62019451

Authorized officer
ZHANG, Yanxia
Telephone No. (86-10) 62089438

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/118213

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103874767 A	18 June 2014	CN 105392893 A	09 March 2016
		WO 2013053207 A1	18 April 2013
		WO 2013053180 A1	18 April 2013
		HK 1215812 A1	15 September 2016
		US 2014249038 A1	04 September 2014
		CN 103874767 B	17 August 2016
		HK 1193845 A1	12 May 2017
		CN 102329876 A	25 January 2012
		CN 102329876 B	02 April 2014
		CN 103890189 B	07 July 2017
		WO 2013053183 A1	18 April 2013
		TW 201315813 A	16 April 2013
		CN 103890189 A	25 June 2014
		WO 2013053182 A1	18 April 2013
WO 2008088804 A1	24 July 2008	US 8357517 B2	22 January 2013
		US 2009325811 A1	31 December 2009
WO 2016183106 A1	17 November 2016	None	

国际检索报告

国际申请号

PCT/CN2017/118213

A. 主题的分类

C12Q 1/68(2018.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C12Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNMED, CPRSABS, CNABS, CPEA, DWPI, SIPOABS, AUABS, TWMED, ILABS, TWABS, HKABS, MOABS, SGABS, CNKI, NCBI, EMBL, GoogleScholar:无创产前诊断, 贝叶斯模型, 基因, SNP, non-invasive prenatal diagnosis (NIPD), Bayesian, gene

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 103874767 A (深圳华大基因研究院, 深圳华大基因科技有限公司) 2014年 6月 18日 (2014 - 06 - 18) 全文	1-26
A	WO 2008088804 A1 (BRIGHAM & WOMENS HOSPITAL等) 2008年 7月 24日 (2008 - 07 - 24) 全文	1-26
A	WO 2016183106 A1 (NATERA INC) 2016年 11月 17日 (2016 - 11 - 17) 全文	1-26

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2018年 2月 1日

国际检索报告邮寄日期

2018年 3月 22日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

张艳霞

传真号 (86-10)62019451

电话号码 (86-10)62089438

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/118213

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103874767	A	2014年 6月 18日	CN	105392893	A	2016年 3月 9日
				WO	2013053207	A1	2013年 4月 18日
				WO	2013053180	A1	2013年 4月 18日
				HK	1215812	A1	2016年 9月 15日
				US	2014249038	A1	2014年 9月 4日
				CN	103874767	B	2016年 8月 17日
				HK	1193845	A1	2017年 5月 12日
				CN	102329876	A	2012年 1月 25日
				CN	102329876	B	2014年 4月 2日
				CN	103890189	B	2017年 7月 7日
				WO	2013053183	A1	2013年 4月 18日
				TW	201315813	A	2013年 4月 16日
				CN	103890189	A	2014年 6月 25日
				WO	2013053182	A1	2013年 4月 18日
WO	2008088804	A1	2008年 7月 24日	US	8357517	B2	2013年 1月 22日
				US	2009325811	A1	2009年 12月 31日
WO	2016183106	A1	2016年 11月 17日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)