



# (12) 发明专利

(10) 授权公告号 CN 102770558 B

(45) 授权公告日 2016. 04. 06

(21) 申请号 201080059486. 6

(22) 申请日 2010. 11. 05

(30) 优先权数据

61/258, 567 2009. 11. 05 US

61/259, 075 2009. 11. 06 US

61/381, 854 2010. 09. 10 US

(85) PCT国际申请进入国家阶段日

2012. 06. 26

(86) PCT国际申请的申请数据

PCT/US2010/055655 2010. 11. 05

(87) PCT国际申请的公布数据

W02011/057094 EN 2011. 05. 12

(73) 专利权人 香港中文大学

地址 中国香港新界

专利权人 塞昆纳姆股份有限公司

(72) 发明人 卢煜明 陈君赐 赵慧君

查尔斯·坎特

(74) 专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

代理人 王达佐 洪欣

(51) Int. Cl.

C12Q 1/68(2006. 01)

(56) 对比文件

WO 2007100911 A3, 2007. 11. 15,

WO 2004078999 A1, 2004. 09. 16,

CN 1469932 A, 2004. 01. 21,

Chunming Ding et al. MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis. 《PNAS》. 2007, 第 101 卷 (第 29 期), 10764 页第 2 栏第 3 段, 第 10765 页第 1 栏第 3 段, 表 3.

周菁琦 等. 孕妇血浆中的游离胎儿核酸. 《临床检验杂志》. 2003, 17-18.

审查员 李有朝

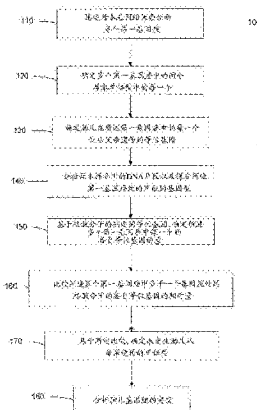
权利要求书7页 说明书34页 附图47页

(54) 发明名称

由母本生物样品进行胎儿基因组的分析

(57) 摘要

提供了确定胎儿基因组的至少一部分的系统、方法和装置。可以分析来自母本样品 (母本和胎儿 DNA) 的 DNA 片段以鉴定某些基因座处的等位基因。综合分析这些基因座处各自等位基因的 DNA 片段的量, 以确定这些基因座的单倍型的相对量, 并确定从亲本基因组遗传了哪种单倍型。可以分析亲本为纯合和杂合的特定组合的基因座, 以确定胎儿基因组的区域。可以使用群体中共有的参照单倍型, 同时分析母体样品的 DNA 片段, 以确定母本和父本基因组。还提供了突变、母本样品中部分胎儿 DNA 浓度以及母本样品测序覆盖度比例确定。



1. 确定怀孕女性的一个未出生胎儿的胎儿基因组的至少一部分的计算机系统,所述胎儿有父亲和身为所述怀孕女性的母亲,且所述父亲的父本基因组具有父本单倍型,所述母亲的母本基因组具有母本单倍型,所述计算机系统包括:

分析获自所述怀孕女性的生物样品的多个核酸分子的装置,其中所述生物样品含有母本核酸和胎儿核酸的混合物,且其中所述分析核酸分子的装置包括:

鉴定所述核酸分子在人类基因组中的位置的装置;以及

确定所述核酸分子的各自的等位基因的装置;

确定在多个第一基因座中的每一个处所述胎儿从所述父亲遗传的父本等位基因的装置,其中所述母本基因组在所述多个第一基因座处为杂合的;

确定所述多个第一基因座的两个母本单倍型中的每一个的装置;

基于所述核酸分子的确定的等位基因,确定在所述多个第一基因座中的每一个处的各自等位基因的量的装置;

比较在所述多个第一基因座中的多于一个基因座处的所述核酸分子的各自等位基因的相对量的装置;以及

基于所述比较,确定在由所述多个第一基因座所覆盖的胎儿基因组的一部分处所述未出生胎儿从所述母亲遗传了两个母本单倍型中的哪一个的装置。

2. 如权利要求 1 所述的计算机系统,其中所述相对量包括所述核酸分子的大小分布。

3. 如权利要求 1 所述的计算机系统,其中基于所述来自生物样品的多个核酸分子的分析,确定所述多个第一基因座的两个母本单倍型中的每一个。

4. 如权利要求 1 所述的计算机系统,其中确定在所述多个第一基因座中的每一个处从所述父亲遗传的等位基因的装置包括:

确定所述父本基因组的、为杂合的多个第二基因座的装置,其中所述母本基因组在所述多个第二基因座处为纯合的;

鉴定所述多个核酸分子中在所述父本基因组中于所述多个第二基因座中的各个基因座处存在而在所述母本基因组中不存在的等位基因的装置;

将遗传的父本单倍型鉴定为具有鉴定的等位基因的单倍型的装置;以及

利用所述遗传的父本单倍型确定在所述多个第一基因座处从所述父亲遗传的等位基因的装置。

5. 如权利要求 1 所述的计算机系统,其中确定所述多个第一基因座的两个母本单倍型中的每一个的装置包括:

基于各自基因座处所述核酸分子的确定的各自等位基因的量的量,鉴定在所述多个第一基因座中的一个或多个处所述母本基因组的等位基因的装置;

鉴定多个参照单倍型的装置;以及

将所鉴定的所述母本基因组的等位基因与所述多个参照单倍型的相应基因座中的等位基因进行比较的装置,以鉴定所述两个母本单倍型。

6. 如权利要求 5 所述的计算机系统,其中确定所述多个第一基因座的两个母本单倍型中的每一个的装置还包括:

将所鉴定的所述母本基因组的等位基因与所述多个参照单倍型反复进行比较的装置,直到唯一鉴定出所述两个母本单倍型中的每一个。

7. 如权利要求 1 所述的计算机系统,其中基于所述来自生物样品的多个核酸分子的分析,确定在所述多个第一基因座中的每一个处从所述父亲遗传的等位基因,并且其中确定在所述多个第一基因座中的每一个处从所述父亲遗传的等位基因的装置包括:

确定所述胎儿基因组为杂合的而所述母本基因组为纯合的多个第二基因座的装置;

通过以下方式确定在所述多个第二基因座中的每一个处从所述父亲遗传的等位基因的装置:

确定在所述多个第二基因座的各个基因座处所述核酸分子的确定的各自等位基因的相对量;以及

将具有最小相对量的等位基因鉴定为在所述各个基因座处遗传的等位基因;

鉴定多个参照单倍型的装置;

利用在所述多个第二基因座中的每一个处从所述父亲遗传的等位基因来确定从所述父亲遗传了哪种参照单倍型的装置,确定的单倍型包括所述多个第一基因座;以及

根据确定为从所述父亲遗传的单倍型确定在所述多个第一基因座处从所述父亲遗传的等位基因的装置。

8. 如权利要求 7 所述的计算机系统,其中确定从所述父亲遗传了哪种参照单倍型的装置包括:

将确定为在所述多个第二基因座中的每一个处从所述父亲遗传的等位基因与所述多个参照单倍型的相应基因座中的等位基因反复进行比较的装置,直到唯一鉴定出从所述父亲遗传的参照单倍型。

9. 如权利要求 7 所述的计算机系统,其中将具体基因座确定为所述多个第二基因座中的一个基因座的装置,在所述一个基因座处,所述胎儿基因组为杂合的而所述母本基因组为纯合的,所述装置包括:

确定所述具体基因座处等位基因的预测计数数量的截止值的装置,所述截止值预测所述母本基因组是否为纯合的以及所述胎儿基因组是否为杂合的,其中基于所述具体基因座处纯合性和杂合性的不同组合的计数数量的统计学分布确定所述截止值;

基于所述来自生物样品的核酸分子的分析,检测所述具体基因座处的第一等位基因和第二等位基因的装置;

基于所述来自生物样品的多个核酸分子的分析,确定所述第一等位基因的实际计数的数量的装置;以及

当所述实际计数的数量小于所述截止值时,确定所述胎儿基因组对于所述第一等位基因和第二等位基因是杂合的,并且所述母本基因组对于所述第二等位基因是纯合的装置。

10. 如权利要求 9 所述的计算机系统,其中所述统计学分布依赖于来源于所述胎儿的生物样品的核酸分子的部分浓度。

11. 如权利要求 10 所述的计算机系统,其中所述统计学分布还依赖于对应于所述具体基因座的所述多个核酸分子的数量。

12. 如权利要求 1 所述的计算机系统,其中确定在所述多个第一基因座中的每一个处从所述父亲遗传的等位基因的装置包括:

通过分析所述父本基因组确定所述父本基因组的、为纯合的多个第二基因座的装置,其中所述多个第一基因座是所述多个第二基因座;

确定在所述多个第一基因座中的每一个处所述父本基因组的等位基因的装置；以及  
将所述多个第一基因座处的各自等位基因指定为从所述父亲遗传的等位基因的装置。

13. 如权利要求 1 所述的计算机系统,其中分析核酸分子的装置包括对所述核酸分子的至少一部分实施至少一项选自以下的技术的装置:大规模平行测序、微阵列、杂交、实时 PCR、数字 PCR 和质谱。

14. 如权利要求 1 所述的计算机系统,还包括:

对于所述多个第一基因座的邻近基因座第一子集中的每一个邻近基因座,确定所述未出生胎儿从所述母亲遗传了包括所述邻近基因座第一子集的第一基因组部分的哪种单倍型的装置,其中确定哪种单倍型的装置包括:

(a) 确定与所述连续基因座第一子集的两个母本单倍型之一匹配的所述核酸分子的确定的各自等位基因的第一量的装置;

(b) 确定与所述连续基因座第一子集的两个母本单倍型中的另一个匹配的所述核酸分子的确定的各自等位基因的第二量的装置;以及

(c) 基于所述第一量与所述第二量的比较确定遗传的所述第一基因组部分的单倍型的装置。

15. 如权利要求 14 所述的计算机系统,其中所述第一量与所述第二量的比较利用序贯概率比检验。

16. 如权利要求 14 所述的计算机系统,其中确定所述第一量与所述第二量均相对于所述邻近基因座第一子集的位置相继进行。

17. 如权利要求 14 所述的计算机系统,其中所述邻近基因座第一子集被进一步分成两个亚组,其中第一亚组由所述父亲的基因型与所述母亲的第一母本单倍型的基因型匹配的基因座组成,且第二亚组由所述父亲的基因型与所述母亲的第二母本单倍型的基因型匹配的基因座组成,并且其中对所述两个亚组单独进行(a)-(c),所述计算机系统还包括:

基于这两个亚组的(c)结果确定遗传的所述第一基因组部分的单倍型的装置。

18. 确定怀孕女性的一个未出生胎儿的胎儿基因组的至少一部分的计算机系统,所述胎儿有父亲和身为所述怀孕女性的母亲,且所述父亲的父本基因组具有父本单倍型,所述母亲的母本基因组具有母本单倍型,所述计算机系统包括:

分析获自所述怀孕女性的生物样品的多个核酸分子的装置,其中所述生物样品含有母本核酸和胎儿核酸的混合物,并且其中所述分析核酸分子的装置包括:

鉴定所述核酸分子在人类基因组中的位置的装置;以及

确定所述核酸分子的各自的等位基因的装置;

确定所述父本基因组的、为杂合的多个第一基因座的装置,其中所述父本基因组获自所述未出生胎儿的所述父亲,并且其中所述母本基因组在所述多个第一基因座处为纯合的;以及

基于在所述多个第一基因座处的确定的各自的等位基因,确定在由所述多个第一基因座所覆盖的胎儿基因组的一部分处所述未出生胎儿从所述父亲遗传的单倍型的装置。

19. 如权利要求 18 所述的计算机系统,其中确定所述未出生胎儿从所述父亲遗传的单倍型的装置包括:

鉴定所述多个核酸分子中在所述父本基因组中于所述多个第一基因座中的各个基因



座处存在而在所述母本基因组中不存在的等位基因的装置 ; 以及

将遗传的父本单倍型鉴定为具有鉴定的等位基因的单倍型的装置。

20. 确定怀孕女性的一个未出生胎儿的胎儿基因组的至少一部分的计算机系统, 所述胎儿有父亲和身为所述怀孕女性的母亲, 且所述父亲的父本基因组具有父本单倍型, 所述母亲的母本基因组具有母本单倍型, 所述计算机系统包括 :

确定所述父本基因组的、为杂合的多个第一基因座的装置, 其中所述父本基因组获自所述未出生胎儿的所述父亲, 并且其中所述母本基因组获自所述未出生胎儿的母亲, 并在所述多个第一基因座处也为杂合的, 并且其中在所述多个第一基因座处两个父本单倍型中的每一个和两个母本单倍型中的每一个都是已知的 ;

确定父本基因组的、为杂合的一个或多个第二基因座的装置, 其中所述母本基因组在所述第二基因座处为纯合的, 并且其中所述多个第一基因座和所述第二基因座位于同一染色体上 ;

分析获自所述怀孕女性的生物样品的多个核酸分子的装置, 其中所述生物样品含有母本核酸和胎儿核酸的混合物, 并且其中所述分析核酸分子的装置包括 :

鉴定所述核酸分子在人类基因组中的位置的装置 ; 以及

确定所述核酸分子的各自的等位基因的装置 ;

通过分析在所述第二基因座中的至少一个处的所述来自生物样品的多个核酸分子的确定的各自等位基因, 来确定所述胎儿遗传了两个父本单倍型的哪一个的装置 ;

比较在所述多个第一基因座中多于一个基因座处的所述核酸分子的确定的各自等位基因的相对量的装置 ; 以及

基于确定为由所述胎儿遗传的父本单倍型并基于所述相对量的比较, 确定在由所述多个第一基因座所覆盖的胎儿基因组的一部分处所述未出生胎儿从所母亲遗传的单倍型的装置。

21. 确定怀孕女性的一个未出生胎儿的胎儿基因组的至少一部分的计算机系统, 所述胎儿有父亲和身为所述怀孕女性的母亲, 且所述父亲的父本基因组具有父本单倍型, 所述母亲的母本基因组具有母本单倍型, 所述计算机系统包括 :

分析获自所述怀孕女性的生物样品的多个核酸分子的装置, 其中所述生物样品含有母本核酸和胎儿核酸的混合物, 并且其中所述分析核酸分子的装置包括 :

鉴定所述核酸分子在人类基因组中的位置的装置 ; 以及

确定所述核酸分子的各自的等位基因的装置 ;

确定所述胎儿基因组为杂合的而所述母本基因组为纯合的多个第一基因座的装置 ;

通过以下方式, 确定在所述多个第一基因座中的每一个处从所述父亲遗传的等位基因的装置 :

确定所述第二基因座的各个基因座处所述核酸分子的确定的各自等位基因的相对量 ; 以及

将具有最小相对量的等位基因鉴定为在所述各个基因座处遗传的等位基因 ;

鉴定多个参照单倍型的装置 ; 以及

利用在所述多个第一基因座中的每一个处从所述父亲遗传的等位基因来确定在由所述多个第一基因座所覆盖的胎儿基因组的一部分处从所述父亲遗传了哪种参照单倍型的

装置。

22. 如权利要求 21 所述的计算机系统,其中确定从所述父亲遗传了哪种参照单倍型的装置包括:

将确定为在所述多个第二基因座中的每一个处从所述父亲遗传的等位基因与所述多个参照单倍型的相应基因座中的等位基因反复进行比较的装置,直到唯一鉴定出从父亲遗传的参照单倍型。

23. 如权利要求 21 所述的计算机系统,其中将具体基因座确定为所述胎儿基因组为杂合的而所述母本基因组为纯合的所述多个第一基因座之一的装置包括:

确定所述具体基因座处等位基因的预测计数数量的截止值的装置,所述截止值预测所述母本基因组是否为纯合的以及所述胎儿基因组是否为杂合的,其中基于所述具体基因座处纯合性和杂合性的不同组合的计数数量的统计学分布确定所述截止值;

基于来自所述生物样品的所述核酸分子的分析,检测所述具体基因座处的第一等位基因和第二等位基因的装置;

基于来自所述生物样品的所述多个核酸分子的测序,确定第一等位基因的实际计数的数量的装置;以及

当所述实际计数的数量小于所述截止值时,确定所述胎儿基因组对于所述第一等位基因和第二等位基因是杂合的,并且所述母本基因组对于所述第二等位基因是纯合的装置。

24. 确定采自怀孕女性的生物样品中胎儿 DNA 的部分浓度的计算机系统,所述胎儿有父亲和身为所述怀孕女性的母亲,其中所述生物样品含有母本核酸和胎儿核酸的混合物,所述计算机系统包括:

分析来自所述生物样品的多个核酸分子的装置,其中所述分析核酸分子的装置包括:

鉴定所述核酸分子在人类基因组中的位置的装置;以及

确定所述核酸分子的各自的等位基因的装置;

确定一个或多个第一基因座的装置,其中所述胎儿基因组在每一个第一基因座处均为杂合的,使得所述胎儿基因组在所述第一基因座处具有各自的第一和第二等位基因,并且其中母本基因组在每一个第一基因座处均为纯合的,使得所述母本基因组在所述第一基因座处具有两个各自的第二等位基因,所述第一等位基因不同于所述第二等位基因,其中将具体基因座确定为一个或多个第一基因座之一包括:

确定所述具体基因座处各自第一等位基因的预测计数数量的截止值的装置,所述截止值预测所述母本基因组是否为纯合的以及所述胎儿基因组是否为杂合的,其中基于所述具体基因座处纯合性和杂合性的不同组合的计数数量的统计学分布确定所述截止值;

基于所述多个核酸分子的分析,检测所述具体基因座处的各自的第一等位基因和各自的第二等位基因的装置;

基于来自所述生物样品的所述多个核酸分子的分析,确定所述各自的第一等位基因的实际计数的数量的装置;以及

当所述实际计数的数量小于所述截止值时,确定所述具体基因座为所述第一基因座之一的装置;

对于所述第一基因座中的至少一个:

确定所述各自的第一等位基因计数的第一数量  $p$  和所述各自的第二等位基因计数的

第二数量  $q$  的装置 ; 以及

基于所述第一和第二数量计算所述部分浓度的装置。

25. 如权利要求 24 所述的计算机系统, 其中所述部分浓度被确定为  $2xp/(p+q)$ 。

26. 如权利要求 24 所述的计算机系统, 其中为多个第一基因座确定  $p$  和  $q$ , 并且其中所

述部分浓度  $f$  被确定为  $f = \frac{\sum_{i=1}^n 2p_i}{\sum_{i=1}^n (p_i + q_i)}$  其中  $p_i$  是第  $i$  个第一基因座的第一数量,  $q_i$

是第  $i$  个第一基因座的第二数量。

27. 如权利要求 24 所述的计算机系统, 其中确定所述截止值的装置包括确定最大和最部分浓度的统计学分布的装置。

28. 确定已从采自怀孕女性的生物样品测序的胎儿基因组比例的计算机系统, 所述胎儿有父亲和身为所述怀孕女性的母亲, 所述父亲的父本基因组具有父本单倍型, 其中所述样品含有母本核酸和胎儿核酸的混合物, 所述计算机系统包括:

接受获自所述怀孕女性的所述生物样品的多个核酸分子测序的测序结果的装置;

分析所述测序结果的装置, 所述分析核酸分子的装置包括:

鉴定所述核酸分子在人类基因组中的位置的装置; 以及

确定所述核酸分子的各自的等位基因的装置;

确定多个第一基因座的装置, 其中所述胎儿基因组在所述多个第一基因座中的每一个处均为杂合的, 使得所述胎儿基因组在所述基因座处具有各自的第一和第二等位基因, 并且其中母本基因组在所述多个第一基因座中的每一个处均为纯合的, 使得所述母本基因组在所述基因座处具有两个各自的第二等位基因, 所述第一等位基因不同于所述第二等位基因;

确定所述多个第一基因座中从所述测序结果检测到各自第一等位基因的基因座的比例的装置, 以及

基于该比例, 确定已从所述生物样品测序的所述胎儿基因组的比例的装置。

29. 如权利要求 28 所述的计算机系统, 其中确定所述多个第一基因座的装置包括:

确定所述父本基因组对于在所述多个第一基因座中的每一个处的所述各自的第一等位基因为纯合的, 以及确定所述母本基因组对于在同一基因座处的所述各自的第二等位基因为纯合的装置。

30. 如权利要求 28 所述的计算机系统, 其中将具体基因座确定为所述多个第一基因座之一的装置包括:

确定所述具体基因座处各自的第一等位基因的预测计数数量的截止值的装置, 所述截止值预测所述母本基因组是否为纯合的以及所述胎儿基因组是否为杂合的, 其中基于所述具体基因座处纯合性和杂合性的不同组合的计数数量的预期分布确定所述截止值;

基于所述测序结果的分析, 检测所述具体基因座处的各自的第一和第二等位基因的装置;

基于来自所述生物样品的所述多个核酸分子的测序, 确定所述各自的第一等位基因的实际计数的数量的装置; 以及

当所述实际计数的数量小于所述截止值时,确定所述胎儿基因组对于所述各自的第二和第一等位基因是杂合的,并且所述母本基因组对于所述各自的第二等位基因是纯合的装置。

31. 如权利要求 28 所述的计算机系统,其中至少两个基因座的所述第一等位基因彼此不同。

32. 确定采自怀孕女性的生物样品中胎儿 DNA 的部分浓度的计算机系统,所述胎儿有父亲和身为所述怀孕女性的母亲,所述父亲的父本基因组具有父本单倍型,其中所述生物样品含有母本核酸和胎儿核酸的混合物,所述计算机系统包括:

富集获自所述怀孕女性的所述生物样品的靶区域内的核酸分子的装置;

对所述富集的生物样品的多个核酸分子进行测序的装置,所述测序对所述靶区域是特异性的,

分析所述测序结果的装置,以

鉴定所述核酸分子在人类基因组靶区域中的位置;以及

确定所述核酸分子的各自的等位基因;

确定一个或多个第一基因座的装置,其中所述胎儿基因组在每一个第一基因座处均为杂合的,使得所述胎儿基因组在所述第一基因座处具有各自的第一和第二等位基因,并且其中母本基因组在每一个第一基因座处均为纯合的,使得所述母本基因组在所述第一基因座处具有两个各自的第二等位基因,所述第一等位基因不同于所述第二等位基因;

对于所述第一基因座中的至少一个:

确定所述各自的第一等位基因计数的第一数量  $p$  和所述各自的第二等位基因计数的第二数量  $q$  的装置;以及

基于所述第一和第二数量确定所述部分浓度的装置。

33. 如权利要求 32 所述的计算机系统,其中确定所述一个或多个第一基因座的装置包括:

确定所述父本基因组对于在所述多个第一基因座中的每一个处的所述各自的第一等位基因为纯合的,以及确定所述母本基因组对于在同一基因座处的所述各自的第二等位基因为纯合的装置。

34. 如权利要求 32 所述的计算机系统,其中将具体基因座确定为一个或多个第一基因座的装置包括:

确定所述具体基因座处所述各自的第一等位基因的预测计数数量的截止值的装置,所述截止值预测所述母本基因组是否为纯合的以及所述胎儿基因组是否为杂合的,其中基于所述具体基因座处纯合性和杂合性的不同组合的计数数量的统计学分布确定所述截止值;

基于所述测序结果的分析,检测所述具体基因座处的所述各自的第一和第二等位基因的装置;

基于来自所述生物样品的所述多个核酸分子的测序,确定所述各自的第一等位基因的实际计数的数量的装置;以及

当所述实际计数的数量小于所述截止值时,确定所述具体基因座为所述第一基因座之一的装置。

## 由母本生物样品进行胎儿基因组的分析

[0001] 相关申请的交叉引用

[0002] 本申请要求 2009 年 11 月 5 日提交的题目为“Fetal Genomic Analysis(胎儿基因组分析)”的美国临时申请第 61/258567 号、2009 年 11 月 6 日提交的题目为“Fetal Genomic Analysis from a Maternal Biological Sample(由母本生物样品进行胎儿基因组的分析)”的美国临时申请第 61/259075 号和 2010 年 9 月 10 日提交的题目为“Fetal Genomic Analysis from a Maternal Biological Sample(由母本生物样品进行胎儿基因组的分析)”的美国临时申请第 61/381854 号的权益,并且是这些临时申请的非临时申请,通过引用将这些临时申请的全部内容并入本文,用于所有目的。

[0003] 本申请还涉及 2008 年 7 月 23 日提交的题目为“Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing(利用大规模平行基因组测序诊断胎儿染色体的非整倍性)”的美国申请第 12/178,181 号(代理人案号 016285-005220US)、题目为“Diagnosing Fetal Chromosomal Aneuploidy Using Genomic Sequencing With Enrichment(利用基因组测序与富集诊断胎儿染色体的非整倍性)”的美国申请第 12/614350 号(代理人案号 016285-005221US)和同时提交的题目为“Size-Based Genomic Analysis(基于大小的基因组分析)”的美国申请(代理人案号 016285-006610US),通过引用将这些申请的全部内容并入本文,用于所有目的。

[0004] 发明背景

[0005] 本发明通常涉及基于母本样品分析胎儿基因组,更具体而言,涉及基于母本样品中基因片段的分析确定全部或部分胎儿基因组。

[0006] 1997 年,母本血浆中无细胞胎儿核酸的发现已为非侵入性产前诊断开启了新的可能性(Lo YMD et al Lancet 1997;350:485-487;和美国专利 6,258,540)。这项技术被快速转化于临床应用,用于检测胎儿来源的、从父系遗传的基因或序列,例如用于胎儿性别确定、胎儿 RhD 状态确定和确定胎儿是否遗传了父系遗传性突变(Amicucci P et al Clin Chem 2000;46:301-302;Saito H et al Lancet 2000;356:1170;以及 Chiu RWK et al Lancet 2002;360:998-1000)。最近在该领域内的进展已经能够由母本血浆核酸分析进行胎儿染色体非整倍性如三体性 21 的产前诊断(Lo YMD et al Nat Med 2007;13:218-223;Tong YK et al Clin Chem 2006;52:2194-2202;美国专利申请 2006/0252071;Lo YMD et al Proc Natl Acad Sci USA 2007;104:13116-13121;Chiu RWK et al Proc Natl Acad Sci USA 2008;105:20458-20463;Fan HC et al Proc Natl Acad Sci 2008;105:16266-16271;美国专利申请 2007/0202525;以及美国专利申请 2009/0029377)。

[0007] 有意义的近期进展的另一方面是利用单分子计数法如数字 PCR 进行单基因疾病的非侵入性产前诊断,在所述单基因疾病中,母亲和父亲二人均携带同一突变。这已经通过母本血浆中的相对突变剂量(RMD)分析得到实现(美国专利申请 2009/0087847;Lun FMF et al Proc Natl Acad Sci USA 2008;105:19920-19925;以及 Chiu RWK et al. Trends 基因 t 2009;25:324-331)。

[0008] 然而,这些方法利用可能突变的现有知识来分析基因组的特定部分,因而不可能鉴定出潜在的或罕见的突变或遗传疾病。因此,期望提供能够利用非侵入性技术鉴定全部或部分胎儿基因组的新方法、系统和装置。

#### [0009] 发明概述

[0010] 本发明的某些实施方案提供了确定怀孕女性的未出生胎儿的基因组的至少一部分的方法、系统和装置。利用含有胎儿和母本遗传材料的样品(例如,来自怀孕母亲的血液样品)可以在产前构建胎儿的全基因组或所选基因组区域的基因图谱。基因图谱可以是胎儿从其父母二人或者仅从其父母之一遗传来的序列的图谱。基于一个或数个这样的基因图谱,可以确定胎儿患遗传疾病或易患遗传疾病或其他疾病或遗传性状的风险。本文还描述了实施方案的其他应用。

[0011] 在一个实施方案中,可以分析来自母本样品(含有母本和胎儿 DNA)的 DNA 片段,以鉴定某些指定基因座(界标)处的等位基因。然后,可以综合分析这些基因座处各等位基因的 DNA 片段的量,以确定这些基因座的单倍型的相对量,进而确定胎儿从母本和/或父本基因组遗传了哪种单倍型。通过鉴定胎儿单倍型,可以确定包括指定基因座的相应基因组区域内的单个基因座处的胎儿基因型。在各种实施方案中,以确定胎儿基因组的区域的方式分析亲本为纯合和杂合的特定组合的基因座。在一项应用中,使用代表群体中常见单倍型的参照单倍型,同时分析母本样品的 DNA 片段,以确定母本和父本基因组。还提供了其他实施方案,诸如确定突变,确定母本样品中的部分胎儿浓度和确定母本样品测序覆盖度的比例。

[0012] 本发明的其他实施方案涉及与本文所述方法相关的系统、装置和计算机可读介质。在一个实施方案中,计算机可读介质含有接收数据和分析数据的指令,而不是指导仪器产生数据(例如,测序核酸分子)的指令。在另一个实施方案中,计算机可读介质含有指导仪器产生数据的指令。在一个实施方案中,计算机程序产品包含存储多个指令的计算机可读介质,所述指令用于控制处理器执行本文所描述的方法的操作。实施方案还涉及经设置执行本文所述任何方法的步骤的计算机系统,可能有不同的组件执行各个步骤或各组步骤。

[0013] 参照说明书的其余部分,包括附图和权利要求书,将会理解本发明实施方案的其他特征和优点。下面结合附图详述本发明各实施方案的其他特征和优点以及结构和操作。在附图中,相似的参考编号可以表示相同的或功能上相似的元件。

#### 附图说明

[0014] 图 1 是按照本发明的实施方案确定怀孕女性的未出生胎儿的基因组的至少一部分的方法 100 的流程图。

[0015] 图 2 示出了按照本发明的实施方案,针对父母各自的基因组密码的具体节段的父亲的两个单倍型和母亲的两个单倍型。

[0016] 图 3 示出了按照本发明的实施方案,图 2 的亲本单倍型中的两种类型的 SNP。

[0017] 图 4A 和 4B 示出了按照本发明的实施方案,确定两种类型 SNP 的胎儿单倍型的分析。

[0018] 图 5A 和 5B 示出了按照本发明的实施方案比较每个基因座片段的相对量(例如,

计数)的分析以及所述比较的结果是否能将特定的单倍型归为遗传的或非遗传的。

[0019] 图 6 示出了按照本发明的实施方案改变 SPRT 分类的似然比 (likelihood ratio) 的影响。

[0020] 图 7 是按照本发明的实施方案确定怀孕女性的未出生胎儿从父亲遗传的基因组的至少一部分的方法 700 的流程图。

[0021] 图 8 是按照本发明的实施方案确定母亲和父亲均为杂合的区域内未出生胎儿的基因组的至少一部分的方法 800 的流程图。

[0022] 图 9 示出了按照本发明的实施方案在特定基因组区域内均为杂合的父亲和母亲的单倍型。

[0023] 图 10 是按照本发明的实施方案显示确定母本样品中胎儿材料的部分浓度的方法 1000 的流程图。

[0024] 图 11 是按照本发明的实施方案确定基因座是否为信息性的方法的流程图。

[0025] 图 12A 和 12B 分别示出了,按照本发明的实施方案,假设部分胎儿 DNA 浓度分别为 20% 和 5%,三种情况下等位基因 T(情况(a)和(c)下等位基因丰度较低)计数的预测分布。

[0026] 图 13A、13B 和 14 示出了,按照本发明的实施方案,对于部分胎儿 DNA 浓度为 20%,较低丰度的等位基因计数的预测分布,不同的分子总计数的每一分布都对应于 SNP。

[0027] 图 15A 和 15B 示出了按照本发明的实施方案参照单倍型、采自参照单倍型的亲本单倍型以及得到的胎儿单倍型的实例。

[0028] 图 16 是按照本发明的实施方案,一组参照单倍型已知,但亲本单倍型未知时,确定胎儿基因组的至少一部分的方法 1600 的流程图。

[0029] 图 17 示出了按照本发明的实施方案根据母本样品的 DNA 片段分析确定信息性基因座的实例。

[0030] 图 18 示出了三个参照单倍型 (Hap A、Hap B 和 Hap C) 和父本等位基因。

[0031] 图 19 示出了按照本发明的实施方案由父本等位基因确定亲本单倍型。

[0032] 图 20 示出了按照本发明的实施方案根据母本样品分析推导母本基因型。

[0033] 图 21 示出了按照本发明的实施方案由母本基因型和参照单倍型确定母本单倍型的一个实施方案。

[0034] 图 22 示出了按照本发明的实施方案确定的母本单倍型和父系遗传的单倍型。

[0035] 图 23 示出了按照本发明的实施方案相对于父本单倍型的母本单倍型的不同类型的基因座 ( $\alpha$  (A) 和  $\beta$  (B))。

[0036] 图 24 是显示鉴定怀孕女性的未出生胎儿的基因组中新生突变的方法 2400 的流程图。

[0037] 图 25A 示出了按照本发明的实施方案表示父亲、母亲和胎儿 (CVS) 的不同基因型组合的 SNP 的绝对数量和百分比。

[0038] 图 25B 示出了列出前 20 个流通池的比对统计数据的表格。

[0039] 图 26 是示出按照本发明的实施方案通过两个方法计算出的 SNP 的胎儿 DNA 的部分浓度的表格。

[0040] 图 27A 的图示出了在该子集中观察到的 SNP 百分比,其中对于所分析的前 20 个流

通池可以从测序数据看到胎儿等位基因。图 27B 的图示出了按照本发明的实施方案覆盖度对比读取 (read) 的数量。

[0041] 图 28A 和 28B 分别示出了按照本发明的实施方案父系遗传的等位基因的覆盖度与可作图的序列读取的数量和流通池序列的数量之间的相互关系的图。

[0042] 图 29A 示出了假阳性率与测序的流通池数量之间的相互关系,图 29B 示出了按照本发明的实施方案假阳性率与测序的流通池数量之间的相互关系。

[0043] 图 30 示出了按照本发明的实施方案所分析的不同数量的流通池的胎儿特异性 SNP 的覆盖度。

[0044] 图 31 示出了按照本发明的实施方案,当使用来自 10 个流通池的数据时 A 型分析的精度。

[0045] 图 32 示出了按照本发明的实施方案,当使用来自 10 个流通池的数据时 B 型分析的精度。

[0046] 图 33 示出了按照本发明的实施方案,当使用来自 20 个流通池的数据时 A 型分析的精度。

[0047] 图 34 示出了按照本发明的实施方案,当使用来自 20 个流通池的数据时 B 型分析的精度。

[0048] 图 35A 和 35B 示出了按照本发明的实施方案,在密码子 41/42 处具有突变和野生型序列的读取。

[0049] 图 36 示出了按照本发明的实施方案 A 型 RHD0 分析的表格,而 B 型 RHD0 分析的表如图 37 所示。

[0050] 图 38A 和 38B 示出了以病例 PW226 为实例的 SPRT 分类结果。

[0051] 图 39 示出了按照本发明的实施方案总结了 5 个病例的 RHD0 分析结果的表格。

[0052] 图 40 示出了按照本发明的实施方案测序深度与测序的流通池数量相比的图。

[0053] 图 41 示出了全基因组的胎儿序列和总序列的大小,图 42A-42C 示出了按照本发明的实施方案每条染色体单独的相似的图。

[0054] 图 43 示出了可与本发明实施方案的系统和方法一起使用的示例性计算机系统 4300 的框图。

[0055] 定义

[0056] 本文所用术语“生物样品”指从个体(如诸如孕妇的人)采集的含有一种或多种目的核酸分子的任何样品。

[0057] 术语“核酸”或“多核苷酸”指单链或双链形式的脱氧核糖核酸(DNA)或核糖核酸(RNA)和其多聚体。除非另有限制,该术语包括含有天然核苷酸的已知类似物的核酸,所述类似物具有与参照核酸类似的结合特性,并且以与天然存在的核苷酸类似的方式代谢。除非另有说明,特定的核酸序列还隐含地包括其保守修饰的变体(如简并密码子取代)、等位基因、直系同源物(orthologs)、SNP 和互补序列以及明确表示的序列。具体来说,简并密码子的取代可以通过产生如下的序列实现:其中一个或多个选择的(或全部)密码子的第三位被混合碱基和/或脱氧次黄苷残基取代(Batzer et al., Nucleic Acid Res. 19:5081(1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608(1985); 以及 Rossolini et al., Mol. Cell. Probes 8:91-98(1994))。术语核酸与基因、cDNA、mRNA、小



非编码 RNA、微 RNA (miRNA)、Piwi- 相互作用 RNA 和基因或基因座编码的短发夹 RNA (shRNA) 交换地使用。

[0058] 术语“基因”意指参与产生多肽链或转录的 RNA 产物的 DNA 的节段。其可以包括编码区之前和之后的区域（前导区和非转录尾区），以及单独的编码节段（外显子）间的间插序列（内含子）。

[0059] 本文所用术语“临床相关核酸序列”（也称为靶序列或染色体）可以指对应于潜在的失衡正被检测的更大的基因组序列节段的多核苷酸序列，或指更大的基因组序列本身。一实例是 21 号染色体的序列。其他的实例包括 18 号、13 号、X 和 Y 染色体。除此以外的其他实例包括，胎儿从其父母之一或两者遗传的突变的基因序列或基因多态性或拷贝数变异，或者作为胎儿中新生突变。在某些实施方案中，多种临床相关核酸序列，或临床相关核酸序列等同的多种标记，可用于提供用来检测失衡的数据。例如，来自 21 号染色体的 5 个不连续序列的数据，能够以加成方式 (additive fashion) 用于确定可能的 21 号染色体失衡，从而将所需的样品体积有效地减少至 1/5。

[0060] 本文所用术语“基于”意指“至少部分地基于”，并指确定另一值所用的一个值（或结果），如存在于方法的输入和该方法的输出的关系中的值。本文所用术语“获得”也指方法的输入和该方法的输出的关系，如该当获得是公式的计算时存在的关系。

[0061] 本文所用术语“参数”意指，表征定量数据集和 / 或定量数据集之间数值关系的数值。例如，第一核酸序列的第一量和第二核酸序列的第二量之间的比值（或比值函数）是参数。

[0062] 本文所用术语“基因座 (locus)”或其复数形式“基因座 (loci)”是在基因组间有变化的任何长度的核苷酸（或碱基对）的位置或地址。

[0063] 本文所用术语“序列失衡”意指，与参考量的任何显著偏差，其是由临床相关核酸序列的量中的至少一个截止值所限定的。序列失衡可以包括染色体剂量失衡、等位基因失衡、突变剂量失衡、单倍型剂量失衡和其他相似的失衡。举例来说，当胎儿具有不同于母亲的基因型时，能够发生等位基因或突变剂量失衡，从而在样品的具体基因座处产生失衡。

[0064] 本文所用术语“染色体非整倍性”意指，染色体的定量数量与二倍体基因组的染色体数量的变化。这种变化可以是增加或丢失。该变化可以包括一个染色体的全部或染色体的区域。

[0065] 本文所用的术语“单倍型”指在同一染色体或染色体区域上共同传递的多个基因座处的等位基因的组合。单倍型可以指少至一对基因座或者是指染色体区域，或者是指整个染色体。术语“等位基因”指处于同一物理基因组基因座处的两个 DNA 序列之一，它们可能导致不同的表型性状或可能不导致不同的表型性状。在任何具体二倍体生物中，每条染色体具有两个拷贝（除了雄性人个体中的性染色体），每个基因的基因型包括该基因座处存在的等位基因对，所述等位基因对在纯合子中是相同的，在杂合子中是不同的。生物群体或生物物种通常包括处于不同个体中的每个基因座处的多个等位基因。在群体中存在多于一个等位基因的基因组基因座被称为多态位点。可以将一个基因座处的等位基因变化测量为所存在的等位基因的数量（即多态性的程度）、群体中杂合子的比例（即杂合率）。本文所用的术语“多态性”指人类基因组中任何个体间的变化，与其频率无关。这类变化的实例包括但不限于，单核苷酸多态性、简单串联重复多态性、插入 - 缺失多态性、突变（其可能

会引起疾病)和拷贝数变化。

#### [0066] 详细描述

[0067] 未出生胎儿的部分基因图谱或全基因组序列的构建可以基于其父母的多态性序列的单倍型而提供。本文所用的术语“单倍型”指在同一染色体或染色体区域上共同传递的多个基因座处的等位基因的组合。例如,实施方案可以分析来自母本样品(含有母本和胎儿 DNA)的 DNA 片段,以鉴定某些指定基因座(界标)处的等位基因。然后,可以综合分析这些基因座处的各等位基因的 DNA 片段的量,以确定这些基因座的单倍型的相对量,进而确定胎儿从母本和/或父本基因组遗传了哪种单倍型。通过鉴定 胎儿单倍型,可以确定包括指定的基因座的相应基因组区域内的单个基因座处的胎儿基因型。在各种实施方案中,可以确定胎儿基因组的区域的方式,分析父母为纯合和杂合的特定组合的基因座。在一种实施中,使用代表群体中常见单倍型的参考单倍型,同时分析母本样品的 DNA 片段,以确定母本和父本基因组。

[0068] 用于确定胎儿基因组的至少一部分的实施方案的应用实例为通过将推断的胎儿基因型或单倍型与疑父的基因型或单倍型进行比较而用于亲子鉴定(paternity testing)。另一实例是检测胎儿已经获得的一种或多种新生突变或检测从其父母产生配子期间就已经发生的减数分裂重组事件。这些是已经受精的配子,并且得到的受精卵已发育成胎儿。

[0069] 此外,一些实施方案还允许以任何所需的分辨率确定未出生胎儿的基因组序列。例如,在某些应用中,实施方案可允许确定胎儿的完整或接近完整的基因组序列。在一个实施方案中,可被确定的胎儿基因组序列的分辨率取决于所了解的父亲和母亲基因组的分辨率,以及来自含有胎儿核酸的母本生物样品的测序信息。如果已知父亲和母亲的完整或接近完整的基因组序列,则可以推导出未出生胎儿的完整或接近完整的基因组序列。

[0070] 在其他实施方案中,只阐述基因组内所选区域的基因组序列,例如,用于所选的遗传、后生(诸如印记失误)或染色体病症的产前诊断。实施方案可应用的遗传病症的实例包括血红蛋白病(诸如 $\beta$ -地中海贫血、 $\alpha$ -地中海贫血、镰状细胞贫血、血红蛋白 E 病)、囊性纤维化和性连锁病症(诸如血友病和杜兴氏肌营养不良)。可利用实施方案检测的突变的其他实例可参见在线人类孟德尔遗传(Online Mendelian Inheritance in Man, [www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi](http://www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi))。

[0071] 一些实施方案还可以用于确定胎儿 DNA 的部分浓度(fractional concentration),这可以在没有父母具体基因组的任何现有知识的情况下进行。相似的分析还可以用于确定对于胎儿基因组的精确确定所需的覆盖深度。因此,该覆盖度确定可用于估计为获得精确结果需要分析多少数据。

#### [0072] I. 引言

[0073] 当将母本样品(例如,血浆或血清)用作阐述胎儿单倍型的材料时,可能面临两个主要的挑战。第一个挑战是母本血浆或血清由胎儿 DNA 和母本 DNA 的混合物组成,并且胎儿 DNA 占很小的比例。已经确定的是,在妊娠的前两三个月,胎儿 DNA 所代表的平均/中值浓度为母本血浆总 DNA 的约 5%至 10%(Lo YMD et al Am J Hum Genet 1998;62:768-775;Lun FMF et al Clin Chem 2008;54:1664-1672)。由于 DNA 是在血液凝固过程中由母本血细胞释放的,所以母本血清中胎儿 DNA 的部分浓度甚至比母本血浆中的浓度低。因此,在一些实

施方案中,母本血浆优于母本血清。

[0074] 第二个挑战是母本血浆中的胎儿 DNA 和母本 DNA 由短片段组成 (Chan KCA et al Clin Chem 2004;50:88-92)。的确,在母本血浆中,胎儿来源的 DNA 通常短于母本来源的 DNA。母本血浆中大多数胎儿 DNA 的长度小于 200bp。单独利用如此短的血浆 DNA 片段,来构建长基因组距离上的基因多态性的单倍型是有挑战性的。母本血浆和血清的上述挑战同样适用于母本尿液中胎儿 DNA 的检测 (Botezatu I et al Clin Chem 2000;46:1078-1084)。胎儿 DNA 仅代表孕妇尿液 DNA 的较小一部分,并且母本尿液中的胎儿 DNA 同样由短 DNA 片段组成。

#### [0075] A. 母本样品的测序与分析

[0076] 一些实施方案所采用的用来解决第一个挑战的方法是,利用允许对获自母本生物样品的核酸进行高精度的定量基因分型的方法。在该方法的一个实施方案中,通过分析大量的(例如,数百万或数亿计的)核酸分子来实现该精度。此外,通过分析单个核酸分子或单个核酸分子的克隆扩增可以提高该精度。一个实施方案利用大规模平行 DNA 测序,例如但不限于,通过以下所进行的测序: Illumina 基因组分析仪平台 (Bentley DR et al. Nature 2008;456:53-59)、Roche 454 平台 (Margulies M et al. Nature 2005;437:376-380)、ABI SOLiD 平台 (McKernan KJ et al. Genome Res 2009;19:1527-1541)、Helicos 单分子测序平台 (Harris TD et al. Science 2008;320:106-109)、利用单个聚合酶分子的实时测序 (Science 2009;323:133-138) 和纳米孔测序 (Clarke J et al. Nat Nanotechnol. 2009;4:265-70)。在一个实施方案中,对生物样品的核酸分子的随机子集进行大规模平行测序。

[0077] 在一些实施方案中,从每个分子获得尽可能长的序列读取可能是有益的。对可以实现的测序读取长度的一个限制是母本生物样品中核酸分子的性质。例如,已知母本血浆中的大多数 DNA 分子由短片段组成 (Chan CA et al Clin Chem 2004;50:88-92)。此外,在长读取长度时,读取长度必须与测序系统的保真度相平衡。对于上述系统中的一些系统,从分子两端获得序列即所谓的双末端测序可能是更好。举例来说,一个方法是从 DNA 分子的每端进行 50bp 的测序,因而产生每分子总共 100bp 的序列。在另一个实施方案中,可以从 DNA 分子的每端进行 75bp 的测序,因而产生每分子总共 150bp 的序列。

[0078] 测序之后,然后将序列与参照人基因组进行比对。由于实施方案阐明未出生胎儿从其父母遗传的基因组变异,因此比对算法能够处理序列变异。这种软件包的一个实例是 Illumina 开发的核苷酸数据库的高效大规模比对 (Efficient Large-Scale Alignment of Nucleotide Databases) (ELAND)。这种软件包的另一个实例是 SOAP (短寡核苷酸比对程序) 和 SOAP2 软件 (Li R et al. Bioinformatics 2008;24:713-714; Li R et al. Bioinformatics 2009;25:1966-1967)。

[0079] 可能需要进行的 DNA 测序的量可以取决于需要构建的胎儿基因图谱或胎儿基因组序列的分辨率。通常,测序的分子越多,分辨率越高。在给定水平或深度的 DNA 测序的胎儿基因图谱或胎儿基因组序列分辨率的另一决定因素是母本生物样品中胎儿 DNA 的部分浓度。通常,部分胎儿 DNA 浓度越高,在给定水平的 DNA 测序可以阐述的胎儿基因图谱或胎儿基因组序列的分辨率越高。由于母本血浆中胎儿 DNA 的部分浓度高于母本血清中部分浓度,所以对于某些实施方案而言,相对于母本血清,母本血浆是更优选的母本生物样品类

型。

[0080] 在使用索引 (indexing) 或条形码 (barcoding) 的情况下,可以增加基于上述测序的方法的通量。因此,可以将样品或患者特异性索引或条形码添加到特定核酸测序库的核酸片段中。然后,将许多这样的库混合起来并一起测序,每个库都具有样品或患者特异性索引或条形码。测序反应之后,可以基于索引或条形码由每个样品或患者收集测序数据。这种策略 可以增加本发明实施方案的通量以及因此增加成本效益。

[0081] 在一个实施方案中,可以在进行定量基因分型 (例如,测序) 之前选择生物样品中的核酸分子或对其分级。在一个变型中,用与来自基因组中选择的基因座 (例如,染色体 7 上含有 CFTR 基因的区域) 的核酸分子优先结合的装置 (例如,微阵列) 处理核酸分子。然后,对由所述装置捕获的核酸分子优先进行测序。该方案允许将测序靶向目的基因组区域。在该方案的一个实施方案中,可以使用 Nimblegen 序列捕获系统 ([www.nimblegen.com/products/seqcap/index.html](http://www.nimblegen.com/products/seqcap/index.html)) 或 Agilent SureSelect 靶标富集系统 (Agilent Sure Select Target Enrichment System, [www.opengenomics.com/SureSelect Target Enrichment System](http://www.opengenomics.com/SureSelect Target Enrichment System)) 或相似的平台。在一些实施方案中,对来自基因组所选区域的核酸分子进行随机测序。

[0082] 在另一个实施方案中,可以首先通过一组或多组扩增引物扩增生物样品中的目的基因组区域。然后,对所扩增的产物进行定量基因分型,例如测序。在该方案的一项实施中,可以使用 RainDance ([www.raindancetech.com/technology/pcr-genomics-research.asp](http://www.raindancetech.com/technology/pcr-genomics-research.asp)) 系统。在一些实施方案中,对所扩增的核酸分子进行随机测序。

[0083] 还可以在生物样品中的核酸分子上进行大小分级步骤。因为已知母本血浆中胎儿 DNA 短于母本 DNA (Li et al Clin Chem 2004;50:1002-1011; 美国专利申请 20050164241; 美国专利申请 20070202525), 因此可以收集分子大小较小的部分,然后用于定量基因分型,例如测序。这一部分将比原始生物样品含有较高的胎儿 DNA 部分浓度。因此,胎儿 DNA 富集的部分的测序可允许在特定的分析水平 (例如,测序深度) 上以比使用非富集的样品时更高的分辨率构建胎儿基因图谱或推导出胎儿基因组序列。因此,这使得该项技术成本效益更高。作为大小分级方法的实例,可以使用 (i) 凝胶电泳,随后从具体凝胶部分提取核酸分子; (ii) 对不同大小的核酸分子具有差别亲和力的核酸结合基质; 或 iii) 对不同大小的核酸分子具有差别截留的过滤系统。

[0084] 在另一个实施方案中,在核酸测序之后可以优先分析特定大小或大小范围的分子。例如,可以进行对 DNA 分子的两端进行测序的双末端测序。然后,可以将这些两端的基因组坐标定位回参照人类基因组。然后通过减去两端的基因组坐标可以推导出分子的大小。进行这种双末端测序的一种方法是使用 Illumina Genome Analyzer 的双末端测序方案。推导 DNA 分子大小的另一方法是对整个 DNA 分子进行测序。这最容易通过读取长度相对长的测序平台进行,例如 Roche 454 平台 (Marguelis et al Nature 2005;437:376-380) 和 Pacific Biosciences 单分子、实时 (SMRT™) 技术 (Eid et al Science 2009;323:133-138)。推导出核酸分子的大小之后,可以选择将随后的分析集中在小于特定截止大小的分子上,进而富集胎儿 DNA 的部分浓度。与没有进行该方案相比,大小选择之后,这种分子子集的分析可以允许用较少的分析分子推导出胎儿基因图谱或胎儿基因组序列。在一个实施方案中,使用 300bp 的截止大小。在其他实施方案中,可以使用

250bp、200bp、180bp、150bp、125bp、100bp 或 75bp 的截止大小。

**[0085] B. 利用亲本基因组作为框架 (Scaffold)**

**[0086]** 为了解决第二个挑战,一些实施方案可以利用母亲染色体的单倍型作为‘框架’。还可以使用父亲染色体的单倍型作为另一个‘框架’。该框架可以与获自含有胎儿 DNA 的母本样品的胎儿的遗传信息进行比较。该胎儿遗传信息可以用于确定母亲和/或父亲的框架是如何在胎儿基因组中建立的,进而利用框架的组成部分确定得到的胎儿基因组。

**[0087]** 可以由父亲和母亲的基因组 DNA 以及家族其他成员(例如,当前处于孕期的胎儿的兄弟姐妹)的基因组 DNA 构建亲本单倍型,鉴于基因组测序成本的降低,亲本单倍型的有效性可能正变得越来越平常。在一种情况中,如果父母一方或双方的基因组已被测序,且他们在一个或多个染色体区域上的单倍型已被确定,那么该信息可以用作上文提到的框架。

**[0088]** 可以使用能够探寻基因组中的序列变异的本领域技术人员已知的任何基因分型平台,包括 DNA 测序、微阵列、杂交探针、基于荧光的技术、光学技术、分子条形码和单分子成像(Geiss GK et al. *Nat Biotechnol* 2008;26:317-325)、单分子分析、PCR、数字 PCR、质谱技术(例如 Sequenom MassARRAY 平台)等。作为更极端的实例,可以通过利用大规模平行测序方法的全基因组 DNA 测序(例如, Bentley DR et al. *Nature* 2008;456:53-59; McKernan KJ et al. *Genome Res* 2009;19:1527-1541)确定父亲和母亲的 DNA 序列。可能感兴趣的序列变异的一个实例是单核苷酸多态性(SNP)。用于确定亲本基因型的特别优选的方法是,通过对基因组范围上或所选的基因组区域的 SNPs 的微阵列分析,所选的基因组区域例如含有突变能导致遗传疾病(例如  $\beta$  珠蛋白簇中的基因或囊性纤维化跨膜传导调节因子(CFTR)基因)的基因的那些区域。除了序列变异之外,还可以使用拷贝数变化。序列变异和拷贝数变化均被称为多态遗传特征(PMF)。

**[0089]** 一方面,可以将目的染色体或染色体区域上的母本基因型构建成单倍型。构建单倍型的一个方式是,通过分析与母亲相关的其他家族成员,例如,母亲的儿子或女儿、父母、兄弟姐妹等。能够构建单倍型的另一方式是,通过上文提到的本领域技术人员公知的其他方法。

**[0090]** 然后,通过与来自其他家族成员例如当前处于孕期的胎儿的兄弟姐妹或来自祖父母的基因型等的基因型信息进行比较,基因型信息可以被扩展成父母的单倍型信息。还可以通过本领域技术人员公知的其他方法构建父母的单倍型。这类方法的实例包括基于单分子分析的方法如数字 PCR(Ding C and Cantor CR. *Proc Natl Acad Sci USA* 2003;100:7449-7453; Ruano G et al. *Proc Natl Acad Sci USA* 1990;87:6296-6300)、精子单倍型分析(Lien S et al. *Curr Protoc Hum Genet* 2002;Chapter 1:Unit 1.6)和成像技术(Xiao M et al. *Hum Mutat* 2007;28:913-921)。其他方法包括基于等位基因特异性 PCR 的方法(Michalatos-Beloin S et al. *Nucleic Acids Res* 1996;24:4841-4843; Lo YMD et al. *Nucleic Acids Res* 1991;Nucleic Acids Res 19:3561-3567)、克隆和限制性酶消化(Smirnova AS et al. *Immunogenetics* 2007;59:93-8)等。其他方法基于群体中单倍型域的分布和连锁不平衡结构,其允许根据统计学评价推知母本单倍型(Clark AG. *Mol Biol Evol* 1990;7:111-22;10:13-9; Salem RM et al. *Hum Genomics* 2005;2:39-66)。

**[0091] C. 利用母本样品的基因组信息组建框架**

**[0092]** 在一个实施方案中,为了研究母本染色体的哪条染色体已传递给胎儿,使用相对

单倍型剂量 (RHDO) 方法。该方法的一般原理如下,以母亲对于基因多态性的每一个都是杂合的为例。因此,有两个单倍型,这些单倍型的相对剂量为 1:1。然而,在母本样品中,小比例胎儿 DNA 的存在可能改变相对单倍型剂量。这是因为胎儿的单倍型互补物的一半遗传自母亲,而另一半遗传自父亲。此外,对于每条染色体,胎儿可能已经遗传了源自父母一方的一个或另一个同源染色体的单倍型的‘拼凑物 (patchwork)’,这取决于减数分裂重组的存在。所有这些因素都可能使母本组成型 DNA 中的相对单倍型剂量偏离 1:1 的比值。因此,对于给定的染色体或染色体区域,这些单倍型的组成型等位基因可以根据母本样品所产生的分析数据 (例如,测序数据) 寻找。

[0093] 然后,可以进行统计学程序以确定相对单倍型剂量,或者确定这些单倍型中的一个是否相对于另一个单倍型被过多表现。可以根据部分胎儿 DNA 浓度调整该统计学程序的分类阈值。通常,较高的部分胎儿 DNA 浓度可以允许所述阈值以较少的分子获得。还可以根据希望在目的基因组或基因组区域上实现的成功分类的片段的数量调整所述分类阈值。在一个实施方案中,可以使用序贯概率比检验 (SPRT)。

[0094] 在一个实施方案中,可以使用美国专利申请 2009/0087847 中所描述的相对突变剂量 (RMD) 来确定母亲具体多态性处等位基因的相对量。这些相对量可用于确定胎儿的单倍型 (例如,当多态性位于连续或连锁的基因座时)。该靶向方法的一项实施是使用聚合酶链式反应 (PCR) 从基因组的所选部分扩增具体序列,用于 RMD 分析。为了使该 RMD 方法扩展至能确定在大基因组区域或全基因组上的胎儿遗传,需要大量母本样品。

[0095] 在利用随机测序的一个实施方案中,没有特异性地靶向目的基因组区域。因此,在目的基因组区域中所获得的序列的数量可能不会像靶向方法中一样多 (除非进行很深的测序)。然而,可以将计数汇总成跨越很多连锁多态性的计数,以实现诊断目的必要的统计学功效。利用该测序实施方案的实际含义在于,其可以通过避免需要过深测序而节约成本。与基于数字 PCR 的方法相比,其也需要输入较少量的母本样品。

[0096] 此外,在区段中适合进行这种 RHDO 分析。换言之,可以分析每条染色体的一个区段,优选多于一个区段。一方面,后者可能允许观察减数分裂重组。例如,胎儿具体染色体节段的单倍型似乎来自母本同源染色体之一,而同一胎儿染色体的另一节段似乎具有来自另一母本同源染色体的单倍型。SPRT 分析可以允许进行这种节段化。

[0097] 例如,可以在邻近的 SNP 上进行 SPRT 分析,所述邻近的 SNP 显示从染色体的一端开始的所需的亲本基因型构型 (即父亲是纯合的而母亲是杂合的)。这将继续,直到 SPRT 分析已经表明,母本单倍型之一在母本血浆分析数据 (例如,测序数据) 中是主要的。然后,SPRT 分析可以‘重置 (reset)’,并从显示所需的亲本基因型构型的下一个邻近的 SNP 重新开始。再次继续,直到 SPRT 分析再一次表明,母本单倍型之一在母本血浆分析数据 (例如,测序数据) 中是主要的。继续该过程直到所述染色体上最后选择的 SNP。然后,可以将染色体上这些多种 SPRT 决定的单倍型节段与母亲基因组的两条同源染色体的单倍型进行比较。当胎儿的单倍型节段似乎已经从一个母本同源染色体转换到另一个时,观察到减数分裂重组。即使每条染色体有多于一处减数分裂重组,该系统也能工作。

[0098] 如后文所描述的,对于组成型基因多态性父亲和母亲均为杂合的基因组区域也可以实施 RHDO 分析。该方案尤其适用于父亲和母亲共有同一祖先来源的疾病基因的突变拷贝的情况,例如当他们有血缘时或当疾病的显性突变是由于大的奠基者效应 (即具有突变

的大多数个体从群体的共同祖先奠基者遗传了相同的单倍型)。因此,父亲和母亲在该区域的单倍型可用于推导胎儿单倍型。

#### [0099] II. 由母本基因组构建胎儿基因组

[0100] 以下描述利用对亲本基因组的明确了解来构建胎儿基因图谱或阐明胎儿基因组序列。

##### [0101] A. 方法

[0102] 图 1 是确定怀孕女性的未出生胎儿的基因组的至少一部分的方法 100 的流程图。胎儿有父亲和身为怀孕女性的母亲。父亲的父本基因组具有两个单倍型,母亲的母本基因组具有两个单倍型。方法 100 分析获自怀孕女性的生物样品的核酸分子(片段),以确定胎儿的基因组。方法 100 主要描述在多个基因座处父亲为纯合的而母亲为杂合的时的情况,而其他实例描述其他实施方案。

[0103] 方法 100 和本文所述的任何方法可以完全或部分地用包括处理器的计算机系统,所述计算机系统被设置成能够进行这些步骤。因此,实施方案涉及被设置成能进行任何本文所述方法的步骤的计算机系统,可能有不同组件执行各个步骤或各组步骤。尽管以编号的步骤呈现,但是本文方法的步骤可以同时或以不同的顺序进行。此外,这些步骤的一部分可以其他方法的其他步骤的一部分一起使用。同样,步骤的全部或部分可以是任选的。此外,任何方法的任何步骤可以用进行这些步骤的模块、电路或其他手段进行。

[0104] 在步骤 110 中,鉴定母本基因组为杂合的多个第一基因座。在一个实施方案中,在基因组范围水平或在选择的基因座处在父亲和母亲的一部分基因分型上进行该确定。在其他实施方案中,可以在分析母本样品的过程中进行多个第一基因座的确定,这将在后面的章节描述。

[0105] 在步骤 120 中,确定覆盖所述多个第一基因座的两个母本单倍型中的每一个。如上文提到的,可以由直接测序获得母本基因组。在其他实施方案中,可以在多个基因座处进行基因分型,然后使用预期具有相似基因组的某人的已知图谱的基因组,例如,来自家族成员或来自相同或相似群体中共有的参照基因组。在一个实施方案中,可以首先对母本基因组的全部或部分进行步骤 120,然后可以研究母本基因组以找到母亲为杂合的基因座。

[0106] 一方面,没有必要构建父亲染色体的单倍型。然而,如果构建了父本单倍型,则可以根据测序结果获得其他信息。一种这类其他信息包括这样的事实:可以对父母均为杂合的区域进行相对单倍型剂量分析。如果可获得父本单倍型,则可获得的另一种其他信息是涉及参与一条或多条父本染色体的减数分裂重组的信息,以及确定与这类多态性相关的疾病等位基因是否已传递给胎儿。

[0107] 在步骤 130 中,确定胎儿在所述多个第一基因座处从父亲遗传的等位基因。一些实施方案使用对父亲为纯合而对母亲为杂合的基因组基因座(如步骤 110 中所提到的)。因此如果父亲在该基因座处为纯合的,则从父亲遗传的等位基因是已知的。确定父亲为纯合的基因座的父亲的基因分型可以以本文所述的任何方式来确定。在一个实施方案中,多个第一基因座的确定可以基于父亲和母亲的基因分型来确定,以便找到父亲为纯合而母亲为杂合的基因座。

[0108] 在另一个实施方案中,可以使用父本基因组的、为杂合的多个第二基因座,以确定在父亲为纯合的多个第一基因座处由胎儿遗传的父本单倍型。例如,如果母本基因组在所

述多个第二基因座处为纯合的,则可以鉴定出在父本基因组中于所述多个第二基因座中的每一个处存在而在母本基因组中不存在的等位基因。然后,遗传的父本单倍型可以被鉴定为具有所鉴定的等位基因的单倍型,并用于确定在所述多个第一基因座处从父亲遗传的等位基因。这些确定父亲单倍型的方面将在下文进行更为详细的讨论。

[0109] 在步骤 140 中,分析获自怀孕女性的生物样品的多个核酸分子。所述样品含有母本核酸和胎儿核酸的混合物。可以采集和接收母本生物样品,用于分析。在一个实施方案中,使用母本血浆和血清。在其他实施方案中,可以使用母本血液、母本尿液、母本唾液、子宫灌洗液或从母本血液获得的胎儿细胞。

[0110] 在一个实施方案中,分析核酸分子包括鉴定所述核酸分子在人类基因组中的位置,以及确定所述核酸分子在各个基因座处的等位基因。因此,一个实施方案可以利用从同一基因座确定的核酸分子的等位基因进行定量基因分型。可以使用允许确定核酸分子在母本生物样品中的基因组位置和等位基因(用于基因型分型的信息)的任何方法。这类方法有一些描述于美国申请 12/178,181 和 12/614350 以及题目为“Size-Based Genomic Analysis(基于大小的基因组分析)”的申请中。

[0111] 在步骤 150 中,基于核酸分子的确定的等位基因,确定所述多个第一基因座中每一个的各自等位基因的量。在一个实施方案中,所述量可以是第一基因座处每种类型的等位基因的数量。例如,6 个 A 和 4 个 T。在另一个实施方案中,量可以是具有特定等位基因的核酸分子的大小分布。例如,相对量还可以包括具有特定基因型的片段的大小分布,所述基因型能够传达相对量的某长度的片段。这种相对量还可以提供关于胎儿基因组中存在哪种基因型的信息,因为胎儿片段倾向小于母本片段。量和方法的一些实例描述于美国申请 12/178,181 和 12/614350 以及申请名称为“Size-Based Genomic Analysis(基于大小的基因组分析)”的申请中。

[0112] 在一个实施方案中,某基因座的等位基因的相对量能够提供关于胎儿遗传了哪种基因型的信息(例如,在资料集达到足够的统计学强度后)。例如,相对量可用于确定相对于某基因座处母亲的基因型是否发生了序列失衡。上文引用的相关专利申请提供了检测具体基因座或区域处的序列失衡的实施方案的实例。

[0113] 在步骤 160 中,比较所述多个第一基因座中多于一个基因座处的核酸分子的各自等位基因的相对量。在一些实施方案中,在比较之前,汇总在所述多个第一基因座的每个基因座处包含单倍型的每个等位基因的量。然后,可以将汇总的亲本单倍型的量进行比较,以确定单倍型是过多表现、均等表现还是表现不足。在其他实施方案中,比较某基因座处等位基因的量,并使用多个基因座处的比较。例如,可以汇总分离值(例如,差异或比值),其可用于与截止值进行比较。这些实施方案的每一个都适用于本文所描述的任何比较步骤,

[0114] 在各种实施方案中,相对量可以是在具体基因座处具有具体等位基因的每个片段的数量的计数,来自具体单倍型上的任何基因座(或区域内的任何基因座)的片段的数量的计数,以及具体基因座或具体单倍型的计数(例如平均值)的统计值。因此,在一个实施方案中,所述比较可以是确定每个基因座处一个等位基因与另一个等位基因相比较的分离值(例如,不同或比值)。

[0115] 在步骤 170 中,基于所述比较,可以确定在由所述多个第一基因座覆盖的基因组的一部分处未出生胎儿从母亲遗传的单倍型。在一个实施方案中,为研究哪条母本染色体



被传递给胎儿,使用例如如上文提到的相对单倍型剂量 (RHDO) 方法。因为母亲对所述第一基因座的每一个都是杂合的,所以所述第一基因座对应于第一基因座的基因组区域的一个单倍型。如果样品仅来自母亲,则这些单倍型的相对剂量为 1:1。偏离该比值或缺少从该比值的偏离可用于确定胎儿从母亲(以及父亲,这在下文有详述)遗传的单倍型。因此,对于给定的染色体或染色体区域,可以根据步骤 130 中产生的分析数据(例如,测序数据)寻找这些单倍型的组成型等位基因。

[0116] 由于分析了多个基因座并将其与母亲的单倍型进行比较,所以基因座之间的序列可被归于具体的单倍型。在一个实施方案中,如果数个基因座匹配一个具体的单倍型,则基因座之间的序列节段可被假定为与母本单倍型的序列节段相同。由于减数分裂重组的发生,由胎儿遗传的最终单倍型可能由来源于这两个同源染色体之一的‘单倍型节段’的拼凑物组成。实施方案能够检测这种重组。

[0117] 可以检测这类重组的分辨率依赖已在父亲和母亲的组成型 DNA 中确定的遗传标志物的数量和分布,以及在随后的生物信息学分析(利用例如 SPRT)中使用的阈值。例如,如果所述比较提示在第一组连续的基因座的每一个处从母亲遗传的等位基因对应于第一单倍型,则所述第一单倍型被确定为是针对对应于第一组基因座的基因组位置而遗传的。如果第二组连续基因座提示遗传了第二单倍型,则第二单倍型被确定为是针对对应于第二组基因座的基因组位置而遗传的。

[0118] 在一个实施方案中,当分析多个基因座时,可以以更高的精度确定单倍型。例如,一个基因座的统计学数据可能无法确定,但是当与其他基因座的统计学数据组合时,可以确定遗传了哪种单倍型。在另一个实施方案中,可以独立地分析每个基因座,以进行分类,然后可以分析分类以提供针对给定区域遗传了哪种单倍型的确定。

[0119] 在一个实施方案中,可以进行统计学步骤以确定相对单倍型剂量(例如,如果这些单倍型中的一个相对于另一个单倍型过多表现)。可以根据部分胎儿 DNA 浓度调整该统计学步骤的分类阈值。通常,较高的部分胎儿 DNA 浓度可以允许用较少的分子达到所述阈值。还可以根据希望在目的基因组或基因组区域上实现的成功分类的节段的数量调整所述分类阈值。

[0120] 返回参见图 1,在步骤 180 中,可以分析突变的胎儿基因组。例如,实施方案可用于搜索在具体群体中引起遗传疾病的一组突变。可利用实施方案检测的突变的实例可参见在线人类孟德尔遗传 ([www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi](http://www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi)) 这些突变可以在步骤 140-160 中搜索或作为本文所述的单独的步骤。例如,在父亲是母亲中不存在的一个或多个突变的携带者的家族中,那么能够根据母本生物样品的分析数据(例如,测序数据)搜索到所述突变。

[0121] 除了检测实际的突变之外,还可以寻找与父亲或母亲的突变体或野生型等位基因相关的多态性遗传标志物。例如, RHDO 分析可以揭示,胎儿已从母亲遗传了已知携带疾病突变的单倍型。本发明的实施方案还可以用于非侵入性产前诊断由染色体区域缺失所引起的疾病,例如,东南亚 (Southeast Asian) 缺失引起的  $\alpha$ -地中海贫血。在父亲和母亲都是缺失携带者的情况下,如果胎儿对于所述缺失是纯合的,且如果对母本血浆 DNA 进行大规模平行测序,那么母本血浆中来源于所述缺失区域的 DNA 序列的频率应该有所下降。

[0122] B. 实例

[0123] 这部分描述应用于母亲为杂合的单核苷酸多态性 (SNP) 的 (例如, 方法 100 的) 实施方案的实例。同一染色体上的 SNP 等位基因形成单倍型, 由于母亲的每条染色体都具有同源的一对, 因此有两个单倍型。为了说明如何进行这样一种确定, 考虑例如如图 2 所显示的 3 号染色体的一个节段。

[0124] 图 2 示出了针对父母各自基因组密码的具体节段的父亲的两个单倍型和母亲的两个单倍型。在该节段内发现 5 个 SNP, 其中对于这些 SNP 中的所有 5 个, 父亲和母亲分别为纯合的和杂合的。父亲的两个同源染色体具有相同的单倍型 (Hap), 即 A-G-A-A-G (图 2 中从上到下)。为了简单起见, 将父本单倍型称为 Hap I 和 Hap II, 记住这两个单倍型对于该 5 个 SNP 组是相同的。对母亲而言, 观察到两个单倍型, 即 Hap III, A-A-A-G-G 和 Hap IV, G-G-G-A-A。

[0125] 该实例中的 SNP 可进一步分成两种类型。图 3 示出了本发明实施方案的两种类型的 SNP。A 型由这样的 SNP 组成: 其中父本等位基因与母本单倍型 III 上的等位基因相同。B 型由这样的 SNP 组成: 其中父本等位基因与母本单倍型 IV 上的等位基因相同。

[0126] 这两种类型的 SNP 可能需要略微不同的数学处理。因此, 在 A 型情况下, 胎儿遗传单倍型 III 将导致母本血浆中单倍型 III 相对于单倍型 IV 的过多表现 (图 4A)。例如, 为了方便讨论, 仅查看一种 SNP 410, 从父亲遗传了等位基因 A, 并且如果从母亲遗传了 Hap III, 则胎儿将向样品贡献两个 A 等位基因, 这将会导致 A 的过多表现。如果胎儿遗传了单倍型 IV, 则不会观察到过多表现, 因为在 A 和 G 位于该基因座的情况下, 胎儿在该基因座处也是杂合的。

[0127] 另一方面, 在 B 型情况下, 胎儿遗传单倍型 III 将导致母本血浆中单倍型 III 和单倍型 IV 的均等表现 (图 4B)。例如, 查看 SNP 420, 从父亲遗传了 G 且 A 作为 Hap III 的一部分将导致胎儿在 SNP 420 处贡献等量的 A 和 G, 如同母亲一样。如果胎儿遗传了单倍型 IV, 则如同上文的讨论所表明的, 将观察到过多表现。

[0128] 图 5A 和 5B 示出了比较每个基因座片段的相对量 (例如, 计数) 的分析, 以及比较的结果是否将具体单倍型归为遗传的还是非遗传的。其中有匹配父亲和母亲的这些基因型构型 (例如, A 型或 B 型情况) 之一的 SNP 的任何基因组位置可用于该实例。根据母本血浆测序数据, 可以集中于对应于 SNP 具体等位基因的测序分子的数量。SPRT 分析 (或其他比较法) 可用于确定在这些等位基因间是否有任何等位基因失衡 (Lo Y D et al Proc Natl Acad Sci USA 2007;104:13116-13121)。

[0129] 图 5A 示出了对 A 型 SNP 的分析。如图所示, 对于每个 SNP, 相对量 (例如, 如分离值所限定的) 与截止值的 SPRT 比较提供了分类。在一个实施方案中, 如果达到 SPRT 的分类阈值, 则可断定胎儿遗传了具体母本单倍型。然后可以重置 SPRT 分析的计数。接着, 分析可以从端粒至着丝粒的方向或着丝粒至端粒的方向移动至与所需基因型构型匹配的邻近 SNP 上。并且新的 SPRT 分析可从该下一个 SNP 开始。

[0130] 另一方面, 在一个实施方案中, 如果 SPRT 分类没有用 SNP 到达, 那么我们也可以以相似的方式移动至邻近的 SNP, 然后再次进行 SPRT, 除了下一个 SNP 的计数被加到之前的 SNP 之外。该过程可以继续直到达到分类阈值。图 5A 和图 5B 示出了对 A 型和 B 型分析运行该过程。在一个实施方案中, 将分类汇总分析以组成区域的总分类。例如, 如果获得了第一组 SNP 和下一组 SNP 的分类, 则可以比较两组的分类, 以查看分类是否一致。

[0131] 图 6 示出了改变 SPRT 分类似然比的影响 (Zhou W et al. Nat Biotechnol 2001;19:78-81; Karoui NE et al. Statist Med 2006;25:3124-33)。通常,较低的分类似然比,比如 8,可以允许更容易地进行分类。这能在基因组内产生更大数量的分类区域。然而,预期很多这样的区域可能被错分。另一方面,较高的分类似然比,比如 1200,可能仅允许当已对较多的 SNP 进行评分时进行分类。这能在基因组内产生较小数量的分类区域。当与使用较低分类阈值的情况相比时,预期错分区域的数量和比例会更低。

[0132] 在一个实施方案中,只有当两个连续的 SPRT 分类导致相同的单倍型(称为“二连续域(two consecutive blocks)”算法)时才进行分类。一方面,“二连续域”算法可以增加分类的精度。在一些实施方案中,对于任何一段序列,实施方案可以首先进行 A 型 SNP 的 SPRT 分析,然后进行 B 型 SNP 的另一 SPRT 分析。在一个实施方案中,可以考虑这样一段序列的情况,对该段序列而言,A 型和 B 型 SNP 形成两交错的基因界标组(例如,SNP)。在使用“二连续域”算法的实施方案中,两域可以为不同的类型。

[0133] 来自 A 型和 B 型分析的 SPRT 结果可以允许核对其分类结果的一致性或不一致性。为了提高分类精度,一个实施方案(“交错法”)只有当给定的基因组区域的 A 型和 B 型分析能够产生一致的结果时才进行分类。如果两种类型分析产生不一致的结果,我们可以查看紧邻该区域的两个连续分类区域的分类结果,一个区域位于端粒端,另一个区域位于着丝粒端。如果这两个连续的区域产生一致的结果,那么我们可以将第一区域分类为具有这两个区域的连续单倍型。如果这两个连续的区域没有产生一致的结果,那么我们可以移动到下一个或两个或甚至更多个连续区域的分类结果作为所关注的最初区域的结果。一般原则是使用邻近基因组区域的分类结果来证实具体区域的分类结果。

### [0134] III. 胎儿遗传的父本等位基因的确定

[0135] 图 7 是确定怀孕女性的未出生胎儿从父亲遗传的基因组的至少一部分的方法 700 的流程图。方法 700 分析获自怀孕女性的生物样品的核酸分子(片段),以确定胎儿的基因组。所述样品含有母本核酸和胎儿核酸的混合物。

[0136] 在步骤 710 中,分析来自生物样品的多个核酸分子的每一个,以鉴定所述核酸分子在人类基因组中的位置,并确定所述核酸分子的等位基因类型。因此,在一个实施方案中,可以确定具体位置(基因座)处的核酸分子的基因型。上文和其他地方所描述任何一种方法都可以用于该分析。

[0137] 在步骤 720 中,确定多个第一基因座,在所述基因座处,父本基因组为杂合的而母本基因组为纯合的。在一个实施方案中,通过确定父本和母本基因组获得所述多个第一基因座。可采集基因组中父亲为杂合的而母亲为纯合的基因组基因座。

[0138] 在步骤 730 中,基于所述多个第一基因座处的确定的基因型,确定在由所述多个第一基因座覆盖的基因组的一部分处未出生胎儿从父亲遗传的单倍型。在一个实施方案中,在分析数据(例如,测序数据)中寻找父亲具有而母亲的基因组中却没有的这些基因座中每一个基因座的等位基因。这些等位基因的组合可以指示胎儿从父亲遗传的染色体的单倍型。

[0139] 在另一个实施方案中,如果父亲基因组中的每个目的染色体或染色体区域的单倍型是已知的,则还可以确定父亲精子发生过程中减数分裂重组发生的地方。因此,在胎儿

和父亲之间,当父系遗传的染色体中的一段 DNA 的单倍型不同时,则可以观察到父本减数分裂重组。当分析数据(例如,测序数据)用于通过与基因多态性的连锁分析(linkage analysis)进行遗传疾病的产前诊断时,包含这种重组信息可能是有用的。

[0140] IV. 父亲和母亲对于某一基因组区域均为杂合的

[0141] 实施方案可以解决这样的情况,其中父亲和母亲对于某一基因组区域均为杂合的。这种情况在父亲和母亲有血缘的家族中尤其相关。当疾病与已产生的显性突变相关时,大的奠基者效应可能也是相关的。在这种情况下,预期如果未出生胎儿的父亲和母亲均是突变基因的携带者,那么携带突变拷贝的基因的染色体的单倍型必然是相同的,除非发生减数分裂重组事件。这种类型的分析尤其适用于常染色体隐性疾病,诸如囊性纤维化、 $\beta$ -地中海贫血、镰状细胞贫血和血红蛋白 E 病。

[0142] 图 8 是按照本发明的实施方案确定母亲和父亲均为杂合的区域内未出生胎儿的基因组的至少一部分的方法 800 的流程图。

[0143] 在步骤 810 中,确定多个第一基因座,在所述基因座处,父亲和母亲均为杂合的。在一个实施方案中,可以通过本文提到的任何方法确定所述第一基因座。例如,可以对亲本基因组的全部或区域进行测序,或者对不同的部分进行基因分型以找到第一基因座。因此,在所述第一多个基因座处的两个父本单倍型的每一个和两个母本单倍型的每一个都是已知的。

[0144] 举例来说,图 9 示出了父亲和母亲的、在具体基因组区域均为杂合的单倍型。如图所示,父母在区域 1 内均具有突变基因(等位基因)。具体而言,父亲的 Hap I 和母亲的 Hap III 具有突变基因。还如图所示,父亲和母亲每人都具有携带野生型基因拷贝的其他染色体拷贝。具体而言,父亲的 Hap II 和母亲的 Hap IV 具有野生型基因。因此,本实例在确定胎儿是否遗传了突变基因中具有相关性。父亲和母亲的携带野生型基因的染色体在紧邻该基因处具有相同的单倍型,但是在离该基因更远一点的位置可能具有不同的单倍型。由于该染色体可能具有不同的祖先来源,所以该染色体在父亲和母亲之间的整个染色体中不大可能具有相同的单倍型。

[0145] 在步骤 820 中,确定多个第二基因座,在所述基因座处,父亲为杂合的而母亲为纯合的。如图所示,所述多个第一和第二基因座位于一染色体上。区域 2 示出了这样的第二基因座。可以选择区域 2,使得对于该区域内的一个或多个 SNP 而言,父亲是杂合的,而在该区域内母亲为纯合的。

[0146] 在步骤 830 中,可以分析来自怀孕女性样品的片段,以鉴定在人类基因组中的位置和基因型。所述位置可用于确定片段(核酸分子)是否包括所述第一基因座中的一个或多个或所述第二基因座中的一个或多个。然后,该信息可用于确定从父亲遗传的单倍型和从母亲遗传的单倍型。

[0147] 在步骤 840 中,通过分析所述第二基因座中的至少一个基因座处的来自生物样品的多个核酸分子的确定的基因型,来确定胎儿遗传了两个父本单倍型的哪一个。例如,可以根据母本生物样品的分析数据(例如,从步骤 710 产生的位置和基因型)找到仅存在于父亲基因组而不存在于母亲基因组中的 SNP 等位基因,诸如图 9 中由 \* 标记的等位基因 T 和由 + 标记的等位基因 A。如同方法 700 可以进行的,如果从母本血浆检测到由 \* 标记的等位基因 T,那么这表示,胎儿从父亲遗传了单倍型 II (Hap II)。相反,如果从母本血浆检测到

由 + 标记的等位基因 A, 那么这表示, 胎儿从父亲遗传了 Hap I。

[0148] 在步骤 850 中, 比较所述多个第一基因座中的多于一个基因座处的确定的核酸分子基因型的相对量。在一个实施方案中, 汇总每一基因座处的量, 并比较母本单倍型的相对量。相对量可以指计算的数量、大小分布以及可以传达有关哪个基因型位于胎儿基因组中具体基因座处信息的任何其他参数。

[0149] 在步骤 860 中, 基于确定为由胎儿遗传的父亲单倍型, 且基于相对量的比较, 确定在由所述多个第一基因座所覆盖的基因组的一部分处, 未出生胎儿从母亲遗传的单倍型。因此, 考虑到区域 2 中由胎儿遗传的父本单倍型, 可以根据母本生物样品的分析数据对区域 1 中的 SNP 进行 RHDO 分析 (例如, 如上文所描述的), 以确定胎儿遗传了两个母本单倍型中的哪一个。在一个实施方案中, 假设当这些区域从父母传递给胎儿时, 区域 1 和 2 之间没有重组。

[0150] 例如, 考虑当通过区域 2 分析胎儿已被确定已经从父亲遗传了 Hap I 的情况。则胎儿从母亲遗传了 Hap III (其在区域 1 与 Hap I 相同) 将导致母本血浆中 Hap III 相对于 Hap IV 过多表现。相反地, 如果胎儿从母亲遗传了 Hap IV, 则在母本血浆中将观察到 Hap III 和 Hap IV 的均等表现。

[0151] 作为另一个实例, 考虑当通过区域 2 分析胎儿已被确定已经从父亲遗传了 Hap II 的情况。则胎儿从母亲遗传了 Hap IV (其在区域 1 与 Hap II 相同) 将导致母本血浆中 Hap IV 相对于 Hap III 过多表现。相反地, 如果胎儿从母亲遗传了 Hap III, 则在母本血浆中将观察到 Hap III 和 Hap IV 的均等表现。

[0152] 在以前的部分中, 我们利用从母本血浆 DNA 测序获得的数据以及胎儿父母的基因型信息推导出胎儿基因组和部分胎儿 DNA 浓度。在下面的部分中, 我们描述在没有母本和父本基因型 / 单倍型现有信息的情况下推导部分胎儿 DNA 浓度和胎儿基因型的实施方案。

[0153] V. 部分胎儿 DNA 浓度的确定

[0154] 在一些实施方案中, 任选的步骤是确定部分胎儿 DNA 浓度。在各种方面中, 该部分浓度可以指导分析的量 (例如, 所需的测序量) 或允许对给定的数据量 (例如, 基因组测序覆盖深度) 估计分析的精度。部分胎儿 DNA 浓度的确定还可以用于确定截止值, 从而确定遗传的单倍型和 / 或基因型的分类。

[0155] 在一个实施方案中, 可以通过采集对父亲和母亲而言均为纯合但具有不同的等位基因的基因座的分析数据 (例如, 如在步骤 140 和 710 中可以获得的) 来确定部分胎儿 DNA 浓度。例如, 对于具有两个等位基因即 A 和 G 的 SNP, 父亲可以为 AA, 母亲可以为 GG, 反之亦然。对于这样的基因座, 胎儿应是肯定杂合子。在上文的一个实例中, 胎儿基因型应是 AG, 母本样品中等位基因 A 的比例可用于确定部分胎儿 DNA 浓度。在另一个实施方案中, 可以进行统计学分析, 以确定母亲为纯合的而胎儿为杂合的基因座。以这种方式, 无需关于母本基因组或父本基因组的现有信息。

[0156] 作为采集分析数据的备选方案, 还可以在一组多态性遗传标志物上通过另一种方法如 PCR 测定的使用、数字 PCR 测定或基于质谱的测定来确定部分胎儿 DNA 浓度 (Lun FMF et al Clin Chem 2008;54:1664-1672)。另一备选方案是利用在胎儿和母亲之间表现出不同 DNA 甲基化的一个或多个基因组基因座 (Poon LLM et al.Clin Chem 2002;48:35-41;Chan KCA et al.Clin Chem 2006;52:2211-2218; 美国专利 6,927,028)。

另一备选方案是利用从参照群体如相似妊娠期确定的近似的部分胎儿 DNA 浓度。然而，由于部分胎儿 DNA 浓度在样品与样品之间不同，所以该后面的方法的精度预期比具体测量受试样品的浓度低。

[0157] A. 确定肯定杂合子的部分浓度

[0158] 在胎儿是肯定杂合子的实施方案中，可以利用一系列下述计算（例如，利用大规模平行测序）确定部分胎儿 DNA 浓度。p 为母本基因组缺乏的胎儿等位基因的计数。q 为其他等位基因即母本和胎儿基因组共有的等位基因的计数。通过以下方程给出部分胎儿 DNA 浓度：

$$\frac{2p}{p+q}$$

[0159]

[0160] 在一项应用中，对跨越满足亲本基因型构型（例如父母双方均为纯合的，但是是等位基因不同）的不同多态性基因座或多态性遗传特征的累积数据进行该计算。

[0161] B. 基于信息性 SNP 的确定

[0162] 对于母亲为纯合而胎儿为杂合的任何基因座，也可以确定胎儿 DNA 的部分浓度，并且不只是当母亲对一个等位基因为纯合的而父亲对不同的等位基因为纯合的时。两种方法均提供了基因座是否是信息性的。术语“信息性 SNP”取决于所需的信息可用于不同的背景下。在一种背景下，信息是胎儿基因组中具体基因座中的等位基因，等位基因不存在于母本基因组中的该基因座处。因此，对于确定胎儿 DNA 浓度的背景，母亲为纯合而胎儿为杂合的 SNP 子集可以称为“信息性 SNP”。母亲和胎儿均为杂合但是对于至少一个不同的等位基因的情况也可以用作信息性 SNP。然而，基因组中三等位基因 SNP 是相对不常见的。

[0163] 图 10 是显示按照本发明的实施方案确定母本样品中胎儿材料的部分浓度的方法 1000 的流程图。在步骤 1010 中，可以分析来自怀孕女性样品的片段，以鉴定在人类基因组中的位置及等位基因类型（这能够导致该位置的基因型确定）。在一个实施方案中，通过对来自怀孕女性获得的生物样品的多个核酸分子进行测序来分析片段。在其他实施方案中，可以使用实时 PCR 或数字 PCR。

[0164] 在步骤 1020 中，确定一个或多个信息性的第一基因座。在一些实施方案中，母本基因组是纯合的，但在样品的信息性基因座处检测到非母本等位基因。在一个实施方案中，胎儿基因组在每个第一基因座处为杂合的，而母本基因组在每个第一基因座处为纯合的。例如，胎儿基因组可以在第一基因座处具有各自的第一和第二等位基因（例如，TA），且母本基因组可以在所述第一基因座处具有两个各自的第二等位基因（例如，AA）。然而，这种基因座可以不是之前已知的，例如在胎儿不是肯定杂合子的情况下。

[0165] 在确定信息性基因座的一个实施方案中，考虑母亲为纯合的 SNP。对于母亲为纯合的 SNP，胎儿对于同一等位基因或者为纯合的或者为杂合的。例如，如果 SNP 对于 A 和 T 是多态性的，且母亲的基因型为 AA，那么胎儿的基因型为 AA 或 TA。在这种情况下，母本血浆样品中等位基因 T 的存在将指示，胎儿基因型是 TA 而不是 AA。某些实施方案可以通过计算必需的截止值，解决存在多少等位基因 T 能指示 TA 的基因型，如下文描述。

[0166] 在步骤 1030 中，对于第一基因座中的至少一个，确定各自的第一等位基因计数的

第一数量  $p$  和各自的第二等位基因计数的第二数量  $q$ 。在一个实施方案中,可以通过多种方法确定母本血浆中胎儿特异性等位基因(等位基因 T)和共有(等位基因 A)等位基因的计数,例如但不限于实时 PCR、数字 PCR 和大规模平行测序。

[0167] 在步骤 1040 中,基于第一和第二数量计算部分浓度。在一个实施方案中,在基因型为 AA 以及其胎儿基因型为 TA 的孕妇中,可以利用以下方程计算部分胎儿 DNA 浓度 ( $f$ ),其中  $p$  表示胎儿特异性等位基因(等位基因 T)的计数, $q$  表示母亲和胎儿共有的等位基因(等位基因 A)的计数。

[0168] 在另一个实施方案中,通过利用多个信息性 SNP,母本血浆中胎儿 DNA 的部分浓度可以提高的精度来估计。对于利用多个 SNP(总共  $n$  个 SNP)的等位基因计数,可以利用以下方程计算胎儿 DNA 的部分浓度 ( $f$ )

$$f = \frac{\sum_{i=1}^n 2p_i}{\sum_{i=1}^n (p_i + q_i)}$$

[0170] 其中  $p_i$  表示信息性 SNP 的胎儿特异性等位基因的计数; $q_i$  表示信息性 SNP 的母亲和胎儿共有的等位基因的计数; $n$  表示信息性 SNP 的总数。多个 SNP 的等位基因计数的使用可以增加部分胎儿 DNA 浓度的估算精度。

[0171] C. 没有父母明确的遗传信息的部分浓度

[0172] 下面描述确定母本血浆样品中部分胎儿 DNA 浓度的方法,所述方法不需要关于胎儿和母亲基因型的现有信息。在一个实施方案中,从母本血浆中 SNP 基因座处不同等位基因的计数,进行信息性 SNP 的鉴定。因此,可以使用方法 1000,同时基于以下描述的实施方案并联合确定信息性 SNP。首先,提供概率的描述以辅助理解用于鉴定信息性 SNP 的截止值的计算。

[0173] 在一个实施方案中,检测胎儿特异性等位基因的概率遵循泊松分布。可以利用以下方程计算检测胎儿特异性等位基因的概率 ( $P$ ):  $P=1-\exp(-fx N/2)$ ,其中  $f$  表示母本血浆样品中胎儿 DNA 的部分浓度, $N$  表示对应于被分析的该具体 SNP 基因座的分子的总数, $\exp()$  表示指数函数。一方面, $P$  可被认为是预期分布,因为其不是测量多种样品的分子数所产生的分布。在其他实施方案中,可以使用其他分布。

[0174] 假设胎儿 DNA 的部分浓度是 5%(前三个月妊娠的典型数值),并分析对应于该 SNP 基因座的 100 个分子(母本+胎儿)(相当于 50 个二倍体基因组中所含的量),检测胎儿特异性等位基因(等位基因 T)的概率是  $1-\exp(-0.05 \times 100/2)=0.92$ 。检测胎儿特异性等位基因的概率将随着部分胎儿 DNA 浓度和针对 SNP 基因座而分析的分子数量而增加。例如,如果胎儿 DNA 浓度是 10% 且分析 100 个分子,则检测胎儿特异性等位基因的概率是 0.99。

[0175] 因此,在母亲为纯合的 SNP 基因座处,母本血浆中与母本等位基因不同的等位基因的存在可以表示,该 SNP 对于部分胎儿 DNA 浓度的计算是“信息性的”。遗失任何信息性 SNP 的概率可以取决于所分析的分子数量。换言之,对于检测信息性 SNP 的任何所需的置信水平,可以根据泊松概率函数计算需要被分析以获得所需精度的分子数量。

[0176] 利用上述分析,一些实施方案可以确定,当母亲基因型未知时基因座是否为信息性的。在一个实施方案中,鉴定了在母本血浆样品中检测到两个不同的等位基因的基因座。例如,对于具有两个可能的等位基因 A 和 T 的 SNP 基因座,在母本血浆中检测到等位基因 A

和 T。

[0177] 图 11 是按照本发明的实施方案确定基因座是否为信息性的方法 1100 的流程图。在一个实施方案中,可以使用方法 1100 来实施方法 1000 的步骤 1020。在另一个实施方案中,方法 1100 的一个步骤是基于统计学分布确定截止值,另一步骤使用截止值确定基因座 (SNP) 是否为信息性的。

[0178] 在步骤 1110 中,确定具体基因座处各自的第一等位基因的多个预测计数的截止值。在一项应用中,截止值预测母本基因组是否为纯合的,以及胎儿基因组是否为杂合的。在一个实施方案中,基于具体基因座处纯合性和杂合性不同组合计数数量的统计分布,确定截止值。例如,可以利用泊松分布函数预测等位基因频率分布。

[0179] 在步骤 1120 中,基于母本样品核酸分子的分析(例如,根据步骤 1010),检测基因座处的第一等位基因和第二等位基因。例如,将一组片段定位到被分析的基因座,并检测第一等位基因或第二等位基因。第一等位基因可以对应于步骤 1020 的各自的第一等位基因之一,第二等位基因可以对应于各自的第二等位基因之一。在一个实施方案中,如果没有检测到两个不同的等位基因,则已知该基因座不是信息性的。

[0180] 在步骤 1130 中,基于核酸分子的分析确定基因座处各自的第一等位基因的实际计数的数量,可以对多个核酸分子的测序结果进行计数,以确定将具有第一等位基因的基因型的片段定位到基因座的次数。

[0181] 在步骤 1140 中,基于对实际计数数量与截止值的比较,将基因座鉴定为第一基因座之一。一方面,截止值可用于区分三种概率:(a) 母亲为纯合的 (AA) 而胎儿为杂合的 (AT);(b) 母亲为杂合的 (AT) 而胎儿为杂合的 (AT);以及 (c) 母亲为杂合的 (AT) 而胎儿为纯合的 (AA) 或 (TT)。为了方便说明,以下的实例假设胎儿基因型为 (c) 情况下的 AA。然而,如果胎儿基因型是 TT,计算是一样的。信息性基因座将具有概率 (a)。

[0182] 在一个实施方案中,当实际计数数量小于截止值,基因座被鉴定为第一基因座之一。在另一个实施方案中,较低的阈值还可以用于确保没有发生假图谱。

[0183] 下面描述确定截止值的实施方案。基于生理上可能的部分胎儿 DNA 浓度(该信息可从以前的研究获得)和对应于 SNP 基因座的分子总数,可以预测以上三种可能情况的等位基因计数的分布。基于预测的分布,可以确定解释母本血浆中所观察到的等位基因计数的截止值,从而确定 SNP 是否为“信息性的”(即情况 (a))。

[0184] 胎儿 DNA 的部分浓度范围在早期妊娠中通常为 5% 至 20%,在晚期妊娠中为 10% 至 35%(Lun et al., Micro fluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma(微流体数字 PCR 表明高于母本血浆中胎儿 DNA 的预期部分). Clin Chem 2008;54:1664-72)。因此,在一个实施方案中,确定 5% 和 20% 的胎儿 DNA 的部分浓度的等位基因计数的预测分布。

[0185] 图 12A 示出了对假设部分胎儿 DNA 浓度为 20% 的三种情况下等位基因 T(情况 (a) 和 (c) 下等位基因丰度较低)计数的预测分布。图 12B 示出了对假设部分胎儿 DNA 浓度为 5% 的三种情况下等位基因 T(情况 (a) 和 (c) 下等位基因丰度较低)计数的预测分布。在两种预测模型中,对 SNP 基因座假设分析总共 200 个分子。

[0186] 利用存在较低丰度的等位基因(等位基因 T)的 40 次计数作为截止值,可以在统计学上辨别三种概率。换言之,对于在母本血浆中检测到的两个等位基因且分析总共 200



个分子的任何 SNP 基因座,如果较少等位基因(较低丰度的等位基因)的等位基因频率小于 40,则 SNP 基因座可被认为是“信息性的”。对于 5% 和 20% 的部分胎儿 DNA 浓度,区分“信息性”SNP(情况(a))与母亲为杂合的 SNP(情况(b)和(c))的准确性 100% 精确的。

[0187] 在实践中,对于不同的 SNP,所检测的分子总数可以是不同的。对于每个 SNP 基因座,可以通过考虑母本血浆样品中所检测到的覆盖 SNP 基因座的分子总数来构建具体的预测分布曲线。换言之,确定 SNP 是否为信息性的计数截止值在 SNP 之间可以不同,并且取决于 SNP 基因座被计数的次数。

[0188] 以下表格示出了对于测序的母本血浆样品,母本血浆中三个 SNP 基因座的等位基因计数。对于三个 SNP 中的每一个,在母本血浆样品中检测到两个不同的等位基因。在母本血浆中检测的对应于这三个 SNP 的计数总数是不同的。

[0189]

| SNP 基因座 | SNP 标识<br>(SNI id) | 等位基因(计<br>数) | 等位基因(计<br>数) | 计数的总数 |
|---------|--------------------|--------------|--------------|-------|
| 1.      | rs3107146          | A(10)        | G(163)       | 173   |
| 2.      | rs7522344          | G(9)         | T(112)       | 121   |
| 3.      | rs2272908          | A(72)        | G(62)        | 134   |

[0190] 图 13A、13B 和 14 示出了部分胎儿 DNA 浓度为 20% 的较低丰度的等位基因计数以及对应于 SNP 的分子的不同总计数的预测分布。利用假设的 20% 的胎儿 DNA 浓度来绘制预测分布,因为这表示前三个月中胎儿 DNA 浓度的上限。胎儿 DNA 浓度越高,预期母亲为纯合的较少等位基因的分布曲线间的重叠越多,当母亲为杂合的时,对于与较少等位基因相对的主要的等位基因。因此,利用较高胎儿 DNA 浓度预测信息性 SNP 得到较少等位基因计数的截止值是更特异的。

[0191] 图 13A 示出了在 173 个分子的总数以及部分胎儿 DNA 浓度为 20% 情况下,较低丰度等位基因计数的预测分布。在一个实施方案中,基于该分布,较低丰度等位基因的计数小于 40 的截止值标准可适用于鉴定信息性 SNP。当等位基因 A 的计数是 10 时,1 号 SNP 基因座被认为对于部分胎儿 DNA 浓度的计算是“信息性的”。

[0192] 图 13B 示出了在 121 个分子的总数以及部分胎儿 DNA 浓度为 20% 情况下,较低丰度等位基因计数的预测分布。在一个实施方案中,基于该分布,较低丰度等位基因的计数小于 26 的截止值可适用于鉴定信息性 SNP。当等位基因 T 计数的数量是 10 时,2 号 SNP 基因座被认为对于部分胎儿 DNA 浓度的计算是“信息性的”。

[0193] 图 12 示出了在 134 个分子的总数以及部分胎儿 DNA 浓度为 20% 情况下,较低丰度等位基因计数的预测分布。在一个实施方案中,基于该分布,较低丰度等位基因的计数小于 25 的截止值可适用于鉴定信息性 SNP。当等位基因 T 计数的数量是 62 时,SNP 基因座 3 号被认为对于部分胎儿 DNA 浓度的计算是“非信息性的”,并且不能用于计算部分胎儿 DNA 浓度。

[0194] 在一些实施方案中,利用方程  $f=2 \times p/(p+q)$ ,利用 SNP 1 和 2 以及两者结合的等位基因计数可以计算胎儿 DNA 的部分浓度。结果如下所示。

[0195]

| 基于 SNP 基因座计算 | 胎儿 DNA 的部分浓度                                        |
|--------------|-----------------------------------------------------|
| 1.           | $10 \times 2 / (10 + 163) = 11.6\%$                 |
| 2.           | $9 \times 2 / (9 + 112) = 14.9\%$                   |
| 1. 和 2.      | $(10 + 9) \times 2 / (10 + 9 + 163 + 112) = 12.9\%$ |

[0196] D. 胎儿基因组深度覆盖度的确定

[0197] 除了获得部分浓度之外,实施方案能够确定在步骤 1010 中的分析程序(例如,测序)已经完成的胎儿基因组的百分比覆盖度。在一些实施方案中,信息性基因座可用于确定覆盖度的百分比。例如,可以使用上文的任何实例。在一个实施方案中,可以使用胎儿为肯定杂合子的基因座。在另一个实施方案中,可以使用经确定胎儿为杂合的而母亲为纯合的基因座(例如,利用方法 1100)。

[0198] 已经定位到信息性基因座的片段可用于确定覆盖度的百分比。在一个实施方案中,确定多个第一基因座中从测序结果检测到各自的第一等位基因的基因座的比例。例如,如果胎儿在某一基因座为 TA,而母亲在该基因座为 AA,则如果所述基因座已被测序,在测序结果中应当检测到等位基因 T。因此,可以基于该比例计算已从生物样品测序的胎儿基因组的比例。在一个实施方案中,观察到胎儿特异性等位基因的第一基因座的比例可以视为胎儿基因组的百分比覆盖度。在其他实施方案中,可以基于基因座的位置修改比例。例如,可以确定每条染色体的百分比覆盖度。作为另一实例,如果第一基因座没有形成基因组的良好表现,那么可以估计百分比不超过该比例。作为另一实例,如果比例是范围的一端,则可以提供该范围。当高百分比即接近 100% 表示接近胎儿基因组的完全覆盖度时,可以以比 100% 少的多的覆盖度,例如,80% 或 50% 或更低,诊断大多数遗传疾病。

[0199] VI. 没有母亲和父亲基因组的现有信息

[0200] 在之前的部分中,一些实施方案已经确定了当母亲的单倍型和父亲的基因型为已知时胎儿(或胎儿基因组的一部分)的基因图谱。其他实施方案已经表明,在没有关于母本、父本或胎儿基因型的现有知识的情况下,可以通过分析母本血浆 DNA 来确定部分胎儿 DNA 浓度。在其他实施方案中,我们现在进一步描述在没有母本和父本基因型/单倍型的现有信息的情况下利用 RHD0 分析确定胎儿(或胎儿基因组的一部分)的基因图谱的方法。

[0201] 在一个实施方案中,使用父母所属的群体的参照(例如,共有的或已知的)单倍型的信息。该信息可用于推导母本和父本单倍型。一个实例用于说明该方法的原理。可以从例如国际人类基因组单体型图计划(hapmap.ncbi.nlm.nih.gov/)网站获得涉及这类参照单倍型的信息。

[0202] 作为示例性实例的一部分,假设群体中存在三种参照单倍型(图 15A 所示的 Hap A、Hap B 和 Hap C)。这三种单倍型中的每一种都由 14 个 SNP 基因座组成,并且对于每一个基因座,有两个可能的等位基因。在该实例中,父亲具有 Hap B 和 Hap C,而母亲具有 Hap A 和 Hap B,如图 15B 所示。该实例假设胎儿从母亲遗传了 Hap A 且从父亲遗传了 Hap C,因此,胎儿具有 Hap A 和 Hap C,如图 15B 所示。

[0203] 图 16 是按照本发明的实施方案,一组参照单倍型已知,但亲本单倍型未知时,确定胎儿基因组的至少一部分的方法 1600 的流程图。

[0204] 在步骤 1610 中,可以分析母本样品以鉴定母亲为纯合的而胎儿为杂合的 SNP。该分析可以以与上文所描述的确定基因座是否为信息性的相似的方式进行。因此,在一个实

施方案中,可以使用方法 1000 和 / 或 1100。在上文所描述的其他实施方案中,可以分析母本和父本基因组以确定进行胎儿基因组作图的信息。

[0205] 图 17 示出了根据母本样品的 DNA 片段分析确定信息性基因座的实例。对于 14 个基因座中的每一个,确定每个基因座的两个等位基因的计数。可以例如但不限于利用实时 PCR、数字 PCR 和大规模平行测序确定这些等位基因的计数。对于这些基因座中的每一个,将在母本血浆中检测到两个不同的等位基因。相对于母亲为杂合的那些 SNP,两个等位基因的比例将显著不同。与母本等位基因相比,胎儿特异性等位基因(胎儿从父亲遗传的等位基因)具有很低的丰度。在图 17 中标记了信息性基因座 1710。

[0206] 在步骤 1620 中,推导胎儿遗传的父本单倍型的一个或多个等位基因。在一个实施方案中,基因座 1710 中的每一个都可用于确定遗传的父本单倍型。例如,胎儿遗传的父本等位基因可被鉴定为基因座 1720 的胎儿特异性等位基因,因为胎儿特异性等位基因是母本样品中丰度远低于母本等位基因的等位基因。

[0207] 在步骤 1630 中,将父本等位基因与参照单倍型相比,以确定从父亲遗传的单倍型。在某些实施方案中,可以推导很多可能的胎儿单倍型,每一单倍型都具有其自身的概率。然后,一种或多种最可能的胎儿单倍型可用于随后的分析或用于临床诊断。

[0208] 在图 18 所示的实例中,群体中有三种可能的单倍型(Hap A、Hap B 和 Hap C)。根据母本血浆分析,已将四个 SNP 鉴定为对母亲为纯合的而对胎儿为杂合的,因此代表了胎儿遗传的父本等位基因。这四个 SNP 的基因型符合 Hap C 的模式。因此,胎儿从父亲遗传了 Hap C,如图 19 所示。换言之,对于同一单倍型域内的所有 SNP,可以推导出胎儿遗传的父本等位基因。

[0209] 在步骤 1640 中,可以确定母亲为杂合的基因座(例如,SNP)。在一个实施方案中,母本样品的分析可以提供母亲为杂合的 SNP。例如,在这些 SNP 中的每一个,可以在母本血浆中检测到两个不同的等位基因。与母亲为纯合的而胎儿为杂合(胎儿特异性等位基因仅在母本血浆中贡献了总等位基因)的 SNP 相比,两种等位基因的计数与母亲为杂合的 SNP 相似。因此,可以根据母本血浆分析,例如如图 20 所示的分析,确定单倍型域内所有 SNP 基因座的完整母本基因型。

[0210] 在步骤 1650 中,通过将基因座的基因型与相关群体的单倍型信息进行比较而根据步骤 1640 的母本基因型推导母本单倍型。图 21 示出了根据母本基因型和参照单倍型确定母本单倍型的实施方案。在所用的实例中,母亲对于第三 SNP 基因座的等位基因 G 为纯合的。由于只有 Hap A 和 Hap B 满足该标准,这表明母亲具有三种单倍型组合即 Hap A/HapA、Hap A/Hap B 或 Hap B/HapB 之一。此外,由于母亲对于第一 SNP 的 A 和 C 为杂合的,所以我们可以推导出母亲具有 Hap A/Hap B 的单倍型组合。在一个实施方案中,可以产生多于一种可能性,且每种可能性可以在下一步骤中检验。根据上述分析,已经确定母亲的单倍型和胎儿从父亲遗传的单倍型。图 22 示出了确定的母本单倍型和父系遗传的单倍型。

[0211] 在步骤 1660 中,根据在步骤 1650 中鉴定的母本单倍型和步骤 1630 中鉴定的父系遗传的单倍型确定胎儿遗传的母本单倍型。利用该信息,一个实施方案可以使用 RHDO 分析来确定哪种母本单倍型被传递给胎儿。可以按照本文所述的任何实施方案进行 RHDO 分析。

[0212] 在一个实施方案中,对于 RHDO 分析,可以将母亲为杂合的 SNP 分成两种类型,即  $\alpha$  型和  $\beta$  型(例如,如图 23 所示和上文所描述的)。 $\alpha$  型 SNP 是指传递给胎儿的父本等位

基因与位于 Hap A 上的母本等位基因相同的那些基因座。对于  $\alpha$  型 SNP, 如果胎儿从母亲遗传了 Hap A, 则母本血浆中 Hap A 上的等位基因将过多表现。另一方面, 如果胎儿从母亲遗传了 Hap B, 则母本血浆中两种母本等位基因将均等表现。

[0213]  $\beta$  型 SNP 是指传递给胎儿的父本等位基因与位于 Hap B 上的母本等位基因相同那些基因座。对于  $\beta$  型 SNP, 如果胎儿从母亲遗传了 Hap B, 则母本血浆中 Hap B 上的等位基因将过多表现。然而, 如果胎儿从母亲遗传了 Hap A, 则母本血浆中两种母本等位基因将均等表现。可以利用 RHD0 分析确定 Hap A 或 Hap B 等位基因可能的过多表现。

[0214] 在一些实施方案中, 为了在没有母本单倍型和父本基因型的现有信息的具体区域上应用 RHD0 分析, 可能需要单倍型域内的相对高倍的 SNP 覆盖度, 例如, 一个实施方案中可能需要分析对应于 SNP 基因座的 200 个分子。可以通过例如但不限于实时 PCR、数字 PCR 和大规模平行测序获得该信息。在一个实施方案中, 靶向测序 (例如, 通过靶标富集和大规模平行测序的组合) 可用于获得靶向区域内不同等位基因的代表性的无偏差的定量信息。以下的实例描述了靶向测序。因此, 该 RHD0 分析能被应用于母本血浆 DNA 的靶向测序数据, 从而在没有关于亲本基因型 / 单倍型的现有信息的情况下确定哪种母本等位基因 / 单倍型被传递给胎儿。

#### [0215] VII. 新生突变的检测

[0216] 一些实施方案可以检测胎儿获得的突变。新生突变是并非由父亲或母亲携带的而是例如在从父亲或母亲或双亲的配子发生过程中产生的突变。这种检测具有临床功效, 因为新生突变在引起很多遗传疾病例如血友病 A 和软骨发育不全中起到重要的作用。

[0217] 图 24 是示出了鉴定怀孕女性的未出生胎儿的基因组中新生突变的方法 2400 的流程图。胎儿具有父亲和身为怀孕女性的母亲, 且父亲的父本基因组具有两种单倍型, 而母亲的母本基因组具有两种单倍型, 所述方法包括:

[0218] 在步骤 2410 中, 对获自怀孕女性的生物样品的多个核酸分子进行测序。注意所述样品含有母本核酸和胎儿核酸的混合物。

[0219] 在步骤 2420 中, 鉴定每一个测序的核酸分子在人类基因组中的位置。在一个实施方案中, 可以通过单末端或双末端测序进行序列的作图。一方面, 为找到位置而定位人类基因组不需要每一个核苷酸对待找的位置精确匹配。

[0220] 在步骤 2430 中, 对于所述位置的至少一部分中的每一个, 确定所考虑的位置上的母本序列和父本序列。例如, 如果在步骤 2420 中确定了 100 个位置, 那么可以确定这 100 个位置处的母本和父本基因组。在一个实施方案中, 与上文所述的利用参照单倍型的相反, 由父本样品确定父本序列。因此, 仍能检测到不在参照基因组中的突变。在各种实施方案中, 可以由仅包括母本 DNA 的样品获得母本序列, 或者还可以例如利用本文所述的方法由生物样品获得母本序列。

[0221] 在步骤 2440 中, 鉴定不存在于确定的母本或父本序列中的多个核酸分子的第一序列。在一个实施方案中, 所述第一序列与确定的母本或父本序列的比较需要精确匹配。因此, 如果匹配不精确, 则认为第一序列不存在于确定的母本或父本序列中。以这种方式, 即使微小的新生突变也可以被鉴定出, 因为新生突变可能仅是单个核苷酸的改变。在另一个实施方案中, 对于被认为是新生突变的序列, 需要显示非母本和非父本序列的某一数量的 DNA 片段。例如, 3 个 DNA 片段的截止值可用于确定序列即新生突变是否存在。

[0222] 在步骤 2450 中,确定生物样品中所述第一序列的第一部分浓度。例如,显示所述第一序列的 DNA 片段的数量可以表示为从该基因座检测到的所有 DNA 片段的比例。

[0223] 在步骤 2460 中,利用胎儿从其父亲遗传的并存在于父本基因组中但不存在于母本基因组中的核酸分子确定生物样品中胎儿核酸的第二部分浓度。这样的核酸分子可能含有在这样的位置处的第一等位基因,在该位置中,父亲为纯合的而母亲也为纯合,但等位基因不同,并且因而胎儿是肯定杂合子。上文所述的信息性基因座可用于确定核酸分子,从而用于确定所述第二部分浓度。

[0224] 在其他实施方案中,可以利用其他方法,例如使用 PCR 测定、数字 PCR 测定或基于质谱的测定,在 Y 染色体、一组基因多态性即单核苷酸多态性或插入-缺失多态性上确定所述第二部分浓度 (Lun FMF et al Clin Chem 2008;54:1664-1672)。另一备选方案是利用在胎儿和母亲之间呈现不同 DNA 甲基化的一个或多个基因组基因座 (Poon LLM et al.Clin Chem2002;48:35-41;Chan KCA et al.Clin Chem 2006;52:221 1-2218;US Patent 6,927,028)。

[0225] 在一个实施方案中,通过不同的 DNA 甲基化模式反映不同的后生状态。不同的 DNA 甲基化模式可以包括 RAS 相关结构域家族 1A (RASSF1A) 或羧化全酶合成酶 (生物素 (丙酰辅酶 A 羧化酶 (ATP 水解)) 连接酶 (HLCS) 基因。具有胎儿特异性 DNA 甲基化特征的 DNA 片段的量可以表示为源自差异甲基化基因座的所有 DNA 片段的比例。

[0226] 在步骤 2470 中,如果所述第一和第二部分浓度大致相同,则所述第一序列被归为新生突变。源自分析过程中的错误例如测序错误的非母本和非父本序列是随机事件,且复发的概率低。因此,在与样品的所测量的部分胎儿 DNA 浓度相似的量上呈现相同的非母本和非父本序列的多个 DNA 片段很可能是存在于胎儿基因组中而不是来源于测序错误的新生突变。在一个实施方案中,截止值可用于确定部分浓度是否相同。例如,如果浓度在彼此的指定值内,那么所述第一序列被归为新生突变。在各种实施方案中,所述指定值可以为 5%、10% 或 15%。

## 实施例

### [0227] I. 实施例 1

[0228] 为了说明本发明的实施方案,分析了以下病例。招募了去产科诊所进行  $\beta$ -地中海贫血产前诊断的一对夫妻。父亲是人  $\beta$  珠蛋白基因的密码子 41/42 的 -CTTT 4 碱基对缺失的携带者。妊娠母亲是人  $\beta$  珠蛋白基因的核苷酸 -28 位处 A  $\rightarrow$  G 突变的携带者。从父亲和母亲采集血液样品。对于母亲,在怀孕期第 12 周的绒毛膜绒毛取样 (CVS) 之前采集血液样品。CVS 之后,一部分储存用于实验。实验的目的是通过母本血浆 DNA 的大规模平行测序构建基因组范围的基因图谱或确定胎儿的部分或完全基因组序列。

### [0229] 1. 亲本基因型的确定

[0230] 从父亲和母亲的白细胞层和 CVS 样品提取 DNA。通过 Affymetrix Genome-Wide Human SNP Array 6.0 系统分析这些 DNA 样品。该系统的特征为具有 1800 万种遗传标志物,包括约 900,000 个单核苷酸多态性 (SNPs) 和用于检测拷贝数变化的多于约 950,000 个探针。显示出父亲、母亲和胎儿 (CVS) 的不同基因型组合的 SNP 的绝对数量和百分比显示于图 25A 的表格中。

[0231] 尽管 Affymetrix 系统用于本实施例中,但是在实践中,可以使用本关于技术人员已知的任何基因分型平台。的确,除了基因分型外,还可 以在全基因组基础上或对所选的基因组区域对父亲和母亲的白细胞层 DNA 进行测序。此外,来自父亲和母亲的任何来源的组成型 DNA(例如,口腔细胞 DNA、毛囊 DNA 等)都可以用于建立亲本基因型。

[0232] 分析 CVS 样品以提供用于与根据母本血浆分析推导出的胎儿基因图谱进行比较的标准。此外,对于该实验,CVS 样品的基因型还可以用于构建母亲的单倍型,用于 RHDO 分析。在这种情况下,用于这类单倍型构建目的的 CVS 基因型的使用仅用于说明目的。在实施方案的临床应用中,可以通过分析家族的其他个体来构建母本单倍型,例如,母亲以前的子女、兄弟姐妹、父母或其他亲戚。还可以通过本领域技术人员公知的其他方法构建感兴趣的染色体区域的母本单倍型,本文提到其中的一些方法。

[0233] 对于选择的实施方案,还可以确定待分析的未出生胎儿的父亲单倍型。该信息尤其可用于其中父亲和母亲均为杂合的染色体区域的相对单倍型剂量。

[0234] 2. 母本血浆 DNA 的大规模平行测序

[0235] 利用 Illumina Genome Analyzer 平台对获自母亲的血浆 DNA 进行大规模平行测序。进行血浆 DNA 分子的双末端测序。每个分子在每端测序 50bp,因而每个分子共计 100bp。利用来自深圳的北京基因组研究所的 SOAP2 程序([soap.genomics.org.cn](http://soap.genomics.org.cn)) (Li R et al. *Bioinformatics* 2009, 25(15):1966-7) 将每个序列的两端与重复 - 揭示的人类基因组(从 UCSC <http://genome.ucsc.edu> 下载的 Hg18 NCBI.36) 进行比对。图 25B 的表格列出了前 20 个流通池的比对统计数据。因此,用 20 个流通池,将超过 39.32 亿个读取与参照人类基因组比对。

[0236] 3. 部分胎儿 DNA 浓度的计算

[0237] 所上文所提到的,可以根据测序数据计算母本血浆样品中胎儿 DNA 的部分浓度。一个方法是分析父亲和母亲均为纯合的但彼此等位基因不同的 SNP。对于这类 SNP,胎儿对于一个父系遗传的和一个母系遗传的等位基因为肯定杂合子。在一个实施方案中,可以使用 V 部分中所描述的任何计算方法。在本实施例中,对满足不同染色体上亲本基因型构型(即父母均为纯合,但为不同的等位基因)的不同多态性基因座范围内的累积数据进行计算。对位于不同染色体上的 SNP 计算的部分胎儿 DNA 浓度列在图 26 的最右栏中。从表格中可以看出,对位于不同染色体上的 SNP 确定的部分浓度彼此密切相关。

[0238] 作为质量对照实验,还根据白细胞层样品的 Affymetrix SNP 6.0 分析研究母亲为纯合的而父亲为杂合的 SNP(图 26 的中间栏)。可以看出,在足够深的 DNA 测序时,根据该分析测量的部分胎儿 DNA 浓度与父亲和母亲均为纯合但等位基因不同的 SNP 的测量非常相似。

[0239] 在一项应用中,当从这两种类型的 SNP 观察到部分胎儿 DNA 浓度的接近一致性,可以推断一类接近胎儿基因组的完全测序覆盖度。一方面,在较小的覆盖深度,对母亲为纯合的而父亲为杂合的 SNP 测量的部分胎儿 DNA 浓度高于对父亲和母亲均为纯合的等位基因不同的 SNP 的测量结果。在这样较小的覆盖深度,来自测序结果的父系特有的等位基因的缺失具有两个可能的原因:(i) 胎儿没有从父亲遗传该等位基因;和/或(ii) 胎儿从父亲遗传了该等位基因,但是然后因为测序深度不够,该等位基因从测序结果丢失。

[0240] 4a. 胎儿基因组百分比覆盖度的计算

[0241] 同样如上文所提到的,可以通过观察其中父亲和母亲均为纯合的但等位基因不同的 SNP 子集来确定已通过母本血浆 DNA 测序进行分析的胎儿基因组的百分比。在该家族中, Affymetrix SNP 6.0 阵列上的 45,900 个 SNP 属于该子集。可以通过分析血浆 DNA 测序数据观察通过测序胎儿等位基因可以以该 SNP 子集的多少百分比被检测到来推导胎儿基因组的百分比覆盖度。

[0242] 图 27A 的图表示出了该子集中观察到的 SNP 百分比,其中对于所分析的前 20 个流通池可以根据测序数据观察到胎儿等位基因。因此,可以在 94% 的这些 SNP 中观察到胎儿等位基因。该测序的程度相当于超过 39.32 亿个读取,每个具有 100bp 的序列。图 27B 的图示出了覆盖度对比读取的数量而不是流通池的数量。随着不同测序平台的通量的增加,预期用于或需要产生这些数量的序列读取或序列长度的流通池或运行的数量在将来会降低。

[0243] 在一些实施方案中,由于在每个染色体区域或染色体中检测到多个 SNP,所以胎儿基因组的覆盖度远低于 94%,尽管仍能提供精确的基因组图谱。例如,假设染色体区域中存在 30 个信息性 SNP,但是仅检测出 30 个 SNP 中的 20 个 SNP 的胎儿等位基因。然而,用 20 个 SNP 仍可以精确地鉴定出染色体区域。因此,在一个实施方案中,可以用小于 94% 的覆盖度获得相当的精度。

[0244] 4b. 胎儿从其父亲遗传的等位基因的基因图谱的覆盖度

[0245] 该示例性的分析集中在父亲为杂合的而母亲为纯合的 SNP 等位基因上。在该家族中, Affymetrix SNP 6.0 平台上的 131,037 个 SNP 属于这一类别。这些 SNP 子集由母亲为纯合的而父亲和胎儿均为杂合的 65,875 个 SNP 组成。在使用 20 个流通池的情况下,可以在这些 SNP 的 61,875 中观察到父系遗传的等位基因,这表明 93.9% 的覆盖度。这后面的百分比与之前的段落中推导的百分比覆盖度数据完全匹配。父系遗传的等位基因的覆盖度与可作图的序列取的数量和流通池序列的数量之间的相互关系分别如图 28A 和图 28B 所示。

[0246] 为了阐明该方法检测真正的父系遗传的胎儿等位基因的特异性,分析了 65,162 (即 131,037-65,875) 个这样的 SNP,其中胎儿遗传了与母亲所具有的等位基因相同的等位基因。对于这类 SNP,明显地检测到不同于母亲的等位基因表示假阳性。因此,当分析 20 个流通池时,在 65,162 个 SNP 中,只观察到 3,225 个假阳性 (4.95%)。这些假阳性可能是父亲或母亲 DNA 的测序错误或基因分型错误或胎儿新生突变的结果。假阳性率与测序的流通池数量之间的相互关系如图 29A 所示。

[0247] 还可以通过考虑父亲和母亲均为纯合并具有等位基因相同的 SNP 子集来估算假阳性率。在具体基因座的任何可选等位基因的存在被认为是假阳性的。这些假阳性可能是父亲或母亲 DNA 的测序错误或基因分型错误或胎儿新生突变的结果。在该子集中存在 500,673 个 SNP。利用来自 20 个流通池的序列数据,在 48,396 个 SNP 中检测到假阳性结果 (9.67%)。假阳性率与测序的流通池数量之间的相互关系如图 29B 所示。该假阳性率高于利用母亲和胎儿为纯合的而父亲为杂合的 SNP 子集的估算。这是因为在后面的 SNP 子集中,只有在母本血浆中存在父系遗传的等位基因被认为是假阳性的,而在前面的子集中,除了由父亲和母亲共有的共有等位基因之外的任何等位基因都被认为是假阳性结果。

[0248] 图 30 示出了所分析的不同数量的流通池的胎儿特异性 SNP 的覆盖度。在该分析中包括父亲和母亲均为纯合的但等位基因不同的 SNP。X 轴是胎儿特异性 SNP 的倍数覆盖度, Y 轴是具有指定倍数覆盖度 SNP 的百分比。随着分析的流通池数量的增加,胎儿特异性

SNP 的倍数覆盖度的平均数增加。例如,当分析一个流通池时,SNP 的平均覆盖度为 0.23 倍。当分析 20 个流通池时,平均覆盖度增加至 4.52 倍。

#### [0249] 5. 从母亲遗传的基因图谱的精度

[0250] 图 31 示出了使用来自 10 个流通池的数据时 A 型分析的精度。II. B 部分描述了 A 型和 B 型分析(也称为  $\alpha$  和  $\beta$ )的实施方案。该精度用于正确确定从母亲遗传的单倍型。为每条染色体单独提供精度。

[0251] 利用 SPRT 分析的 1,200 的似然比(Zhou W et al.Nat Biotechnol 2001;19:78-81;Karoui NE et al.Statist Med 2006;25:3124-33),精度范围为 96% 至 100%。如图所示,对于 SPRT 分类即使具有如此高的似然比,仍可以对基因组范围内的共计 2,760 个节段进行分类。这种分辨率程度对于大多数目的是足够的,当考虑减数分裂重组以每条染色体臂每代一至低单数数量的频率发生时。此外,当使用交错法时,可以看出能防止所有的错误分类(图 31 的右侧)。如上文所描述的,交错法使用 A 型和 B 型分析。

[0252] 图 32 示出了使用来自 10 个流通池的数据时 B 型分析的精度。利用 SPRT 分析的 1,200 的似然比,精度范围为 94.1% 至 100%。如同图 31 所见的,当使用交错法时,能防止所有的错误分类(图 32 的右侧)。

[0253] 图 33 示出了使用来自 20 个流通池的数据时 A 型分析的精度。利用 SPRT 分析的 1,200 的似然比和“二连续域”算法,共计进行了 3,780 次分类,且只有 3(0.1%)次分类是错误的。图 34 示出了使用来自 20 个流通池的数据时 B 型分析的精度。利用 SPRT 分析的 1,200 的似然比和“二连续域”算法,共计进行了 3,355 次分类,且只有 6(0.2%)次分类是错误的。在这些实例中,在很多遗传标志物如 SNP 内进行 SPRT。

#### [0254] II. B-地中海贫血风险的产前确定

[0255] 在一个实施方案中,为了确定胎儿患  $\beta$ -地中海贫血(一种常染色体隐性疾病)的风险,可以确定胎儿是否遗传了其父亲和母亲所携带的突变等位基因。在上面提到的这种病例下,父亲是人  $\beta$  珠蛋白基因的密码子 41/42 的 -CTTT 4 碱基对缺失的携带者。妊娠母亲是人  $\beta$  珠蛋白基因的核苷酸 -28 位处 A  $\rightarrow$  G 突变的携带者。

[0256] 为了确定胎儿是否遗传了父本密码子 41/42 突变,为该突变搜索利用前 10 个流通池的母本血浆 DNA 的测序数据。发现具有该突变的共计 10 个读取(图 35A)。因此胎儿遗传了父本突变。此外,发现 62 个读取含有密码子 41/42 处的野生型序列(图 35B)。因此,在含有突变的该区域内的读取的百分比是 0.1389。该图非常接近图 26 中所确定的部分胎儿 DNA 浓度。在一个实施方案中,还可以通过阐明其遗传与父本突变连锁的基因多态性来确定胎儿遗传父本突变的风险。

[0257] 在一个实施方案中,为了确定胎儿遗传母本 -28 位突变的风险,进行了 RHD0 分析。在该家族中,-28 位突变位于单倍型 IV 上,而野生型等位基因位于单倍型 III 上。A 型 RHD0 分析的结果如图 36 所示,而 B 型 RHD0 分析的结果如图 37 所示。在两种类型的分析中,推导出胎儿从母亲遗传了单倍型 III。换言之,胎儿从母亲遗传了野生型等位基因。胎儿的最终诊断结论是其从父亲遗传了密码子 41/42 突变并从母亲遗传了野生型等位基因。因此,胎儿是  $\beta$ -地中海贫血的杂合携带者,因而在临床上应当是健康的。

#### [0258] III. 靶标富集和靶向测序

[0259] 如之前的部分所讨论的,估算部分胎儿 DNA 浓度的精度和根据母本血浆 DNA 分析



推导出基因图谱的分辨率可以取决于目的基因座的覆盖深度。例如,我们已经证实,在没有母本基因型的现有信息的情况下,为了高精度地确定部分胎儿 DNA 浓度,可能需要对应于 SNP 基因座的总计 200 个分子。可以通过例如但不限于实时 PCR、数字 PCR 和大规模平行测序获得母本血浆中 SNP 的等位基因计数。

[0260] 由于母本血浆 DNA 的大规模平行测序可以同时确定整个基因组上的数百万 SNP 的等位基因计数,因而其是跨越不同基因座的基因组范围分析的理想平台。大规模平行测序的基本模式允许以相似的深度覆盖基因组内的不同区域。然而,为了利用随机大规模平行测序以高测序深度 对具体目的区域进行测序,基因组的其余部分(并非意欲分析的)必须被测序至相同的程度。因此,该方法将是昂贵的。为了提高大规模平行测序方法的成本效益,一个方法是在进行测序之前富集靶区域。可以通过液相捕获(Gnirke A, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing(具有超长寡核苷酸的溶液杂交选择用于大规模平行靶向测序). Nat Biotechnol 2009; 27:182-9)、微阵列捕获(例如,利用 NimbleGen 平台)或靶向扩增(Tewhey R, et al. Microdroplet-based PCR enrichment for large-scale targeted sequencing(基于微滴的 PCR 富集用于大规模靶向测序). Nat Biotechnol 2009; 27:1025-31)进行靶向测序。

[0261] 靶向测序最初应用于检测群体遗传变异,例如,用于遗传相关研究。因此,其目前在基因组研究中的应用旨在解决定性问题(例如,基因分型或突变检测)。然而,将靶向测序应用于母本血浆 DNA 中,用于非侵入性产前诊断,涉及定量考虑,其可行性还不清楚。例如,靶向测序的使用可能在母本血浆中胎儿和母本 DNA 的检测中引入定量偏差。此外,之前的工作已经显示,胎儿 DNA 短于母本 DNA (Chan KCA et al. Size distributions of maternal and fetal DNA in maternal plasma (母本血浆中母本和胎儿 DNA 的大小分布). Clin Chem 2004; 50:88-92)。这种大小差异还可能在母本血浆的胎儿和母本 DNA 捕获中引入定量偏差或差别效率。还无法确定这类片段化的 DNA 分子可能被捕获的效率。在以下描述中,我们证实,可以通过靶标富集随后进行大规模平行测序来实现靶向测序。我们还表明,与全基因组测序相比,靶标富集是估算部分胎儿 DNA 浓度的有效方法。

[0262] A. 利用靶标富集确定部分浓度

[0263] 1. 材料和方法

[0264] 招募了四名(M6011、M6028、M6029 和 M6043)具有一个女性胎儿的孕妇。在第三个三个月的自选式剖腹产之前,将母本外周血样品收集到 EDTA 血液管中,同时在自选式剖腹产之后收集胎盘样品。离心后,利用 Blood Mini Kit(Qiagen) 从外周血细胞提取 DNA。通过 DSP DNA Blood Mini Kit(Qiagen) 从 2.4mL 血浆提取 DNA。从白细胞层提取母本基因组 DNA 并从胎盘组织提取胎儿基因组 DNA。用于本实施例中的第三个三个月的样品仅用于说明的目的。可以等同使用第一和第二个三个月的样品。

[0265] 通过 Genome-Wide Human SNP Array 6.0(Affymetrix) 确定母本和胎儿基因型。每个病例的 5-30ng 血浆 DNA 用于通过双末端样品制备试剂盒(Illumina) 按照生产商的染色质免疫沉淀测序样品制备的方案构建 DNA 文库。直接利用 QIAquick PCR 纯化试剂盒(Qiagen) 中提供的吸附柱纯化连接体连接的 DNA,而不用进一步的大小选择。然后,利用 15 个循环的 PCR 以标准引物扩增连接体连接的 DNA。所述引物是来自 Illumina 的 PCR Primer

PE 1.0 和 2.0。利用 NanoDrop ND-1000 分光光度计 (NanoDrop Technologies) 对 DNA 文库进行定量, 并利用 DNA 1000 试剂盒 (Agilent) 在 2100 生物分析仪上运行, 以检测大小分布。为每个样品产生平均大小为大约 290bp 的 0.6-1  $\mu$ g 的扩增血浆 DNA 文库。捕获文库获自 Agilent 并覆盖人 chrX(目录号: 5190-1993) 上 85% 的外显子。对于本研究的所有四个病例, 按照生产商的说明书, 将每个病例的 500ng 扩增的血浆 DNA 文库与捕获探针在 65° C 下孵育 24 小时。杂交后, 通过利用链霉亲和素包被的磁珠 (DynaMag-2 Invitrogen) 拉下生物素化的探针/靶标杂交体来选择捕获的靶标, 并用 MinElute PCR 纯化试剂盒 (Qiagen) 纯化。最后, 用来自 Agilent 的 SureSelect GA PE 引物、通过 12 个循环的 PCR 扩增富集所靶向的 DNA 文库。通过 QIAquick PCR 纯化试剂盒 (Qiagen) 纯化 PCR 产物。然后, 利用 Illumina Genome Analyzer IIx 将在有或无靶标富集的情况下制备的 DNA 文库进行随机大规模平行测序。标准流通池上的一个测序泳道用于对每个 DNA 文库进行测序。

[0266] 2. 在无靶标富集的情况下胎儿 DNA 的部分浓度

[0267] 可以基于信息性 SNP (即母亲为纯合的而胎儿为杂合的 SNP) 的等位基因计数计算部分胎儿 DNA 浓度。以下的表格显示, 在所有四个病例的全基因组上鉴定了 120184、110730、107362 和 110321 个信息性 SNP, 而 63、61、69 和 65 (分别以相同的病例顺序) 个落入 X 染色体上的靶区域内。在无靶标富集的情况下, 基于基因组中所有信息性 SNP 的数据, 部分胎儿 DNA 浓度为 33.4%、31.3%、29.2% 和 34.4%。

[0268]

| 样品 | 靶标富集 | 全基因组信息性 SNP 的数量 | 共有的等位基因计数 | 胎儿特异性等位基因计数 | 部分胎儿 DNA 浓度 |
|----|------|-----------------|-----------|-------------|-------------|
|----|------|-----------------|-----------|-------------|-------------|

[0269]

|       |   |         |        |       |       |
|-------|---|---------|--------|-------|-------|
| M6011 | 无 | 120,184 | 15,309 | 3,064 | 33.4% |
| M6028 | 无 | 110,730 | 16,778 | 3,114 | 31.3% |
| M6029 | 无 | 107,362 | 19,889 | 3,404 | 29.2% |
| M6043 | 无 | 110,321 | 21,070 | 4,369 | 34.4% |

[0270] 3. 在有或无靶标富集的情况下样品的比较

[0271] 在一些实施方案中, 序列覆盖深度代表特定区域内已经测序的每个碱基的次数的平均数。在本实施方案中, 我们通过将靶区域内测序碱基的总数除以靶区域长度 (3.05Mb) 来计算所述靶区域的序列深度。对于富集试剂盒覆盖的区域, 平均序列覆盖度对于非富集的样品为 0.19 倍, 对于富集的样品为 54.9 倍, 这表明平均富集 289 倍。以该测序深度, 在靶标富集之前, 仅检测到靶区域内 4.0% 的胎儿特异性等位基因 (参见以下的表格)。相比之下, 在靶标富集之后, 它们当中的 95.8% 可被检测 (参见以下的表格)。因此, 靶标富集极大地提高了靶区域内胎儿特异性等位基因的检测率。

[0272] 然后, 我们基于每个样品的靶区域内的所有信息性 SNP 的读取计数, 对有和无靶标富集情况下的部分胎儿 DNA 浓度进行比较。在无靶标富集的情况下, 四种样品的胎儿特异性读数的数量范围为 0 至 6 (参见以下的表格)。由于序列覆盖度低, 不充足的胎儿 DNA 分子取样会阻止精确估算部分胎儿 DNA 浓度。在有靶标富集的情况下, 在靶区域内观察到数

量大得多的胎儿特异性等位基因计数 (511-776) 和共有的等位基因计数 (2570-3922) (参见以下的表格)。经计算, 胎儿 DNA 百分比为 35.4%、33.2%、26.1% 和 33.0%, 这与非富集样品内的基因组范围的数据所估算的胎儿 DNA 百分比一致 (参见以下的表格)。这些结果表明母本和胎儿 DNA 分子在靶区域内被富集至相似的程度。

[0273]

| 样品    | 靶标富集 | 靶向区域内信息性 SNP 的数量 | 可检测的胎儿特异性等位基因的数量 | 胎儿特异性等位基因检测率 | 共有的等位基因计数 | 胎儿特异性等位基因计数 | 部分胎儿 DNA 浓度 |
|-------|------|------------------|------------------|--------------|-----------|-------------|-------------|
| M6011 | 无    | 63               | 6                | 9.5%         | 13        | 6           | 63.2%       |
| M6028 | 无    | 61               | 2                | 3.3%         | 6         | 2           | 50%         |
| M6029 | 无    | 69               | 2                | 2.9%         | 11        | 2           | 30.8%       |

[0274]

|       |   |    |    |       |      |     |       |
|-------|---|----|----|-------|------|-----|-------|
| M6043 | 无 | 65 | 0  | 0.0%  | 15   | 0   | 0.0%  |
| M6011 | 有 | 63 | 60 | 95.2% | 3072 | 661 | 35.4% |
| M6028 | 有 | 61 | 60 | 98.4% | 2570 | 511 | 33.2% |
| M6029 | 有 | 69 | 66 | 95.7% | 3835 | 575 | 26.1% |
| M6034 | 有 | 65 | 61 | 9.39% | 3922 | 776 | 33.0% |

[0275] B. 利用靶标富集确定胎儿基因组

[0276] RHDO 方法的一个应用是, 用于母系遗传的遗传疾病的非侵入性产前检测。当母本血浆 DNA 的测序深度大约为 65 倍人类基因组覆盖度时, 在无靶标富集的情况下利用母本血浆的大规模平行测序, RHDO 分析能够精确地确定哪种母本单倍型被传递给平均具有 17 个 SNP 的胎儿。为了提高该方法的成本效益, 可以进行选择性引导对基因组内的特定目的区域进行测序, 然后对测序数据应用 RHDO 分析。举例来说, 我们说明了利用 X 染色体的靶向测序和 RHDO 分析的概念。然而, 也可以将靶向测序和 RHDO 分析应用于所有染色体, 例如常染色体。在一个实施方案中, 上文所述的 RHDO 分析可用于靶向实施方案。

[0277] 招募了招募了五名 (PW226、PW263、PW316、PW370 和 PW421) 具有一个男性胎儿的孕妇。在前三个月的绒毛膜绒毛取样之前, 将母本外周血样品收集到 EDTA 血液管中, 离心后, 利用 Blood Mini Kit (Qiagen) 从外周血细胞提取 DNA。通过 DSP DNA Blood Mini Kit (Qiagen) 从 3.2mL 血浆提取 DNA。从白细胞层提取母本基因组 DNA, 并从绒毛膜绒毛提取胎儿基因组 DNA。按上文所述制备并分析样品。然后, 每个样品用 Illumina 流通池上的一个泳道进行随机测序。

[0278] 在本实施例中, 我们利用胎儿基因型以及来自母亲核酸的测序信息来推导 X 染色体的母本单倍型并推导从母亲遗传了哪种单倍型。对于母亲为杂合的 X 染色体上的每一个 SNP (即信息性 SNP), 由胎儿遗传的等位基因被限定为来源于母本单倍型 1 (Hap I), 而没有传递给胎儿的母本等位基因被限定为来源于母本单倍型 2 (Hap II)。在一些实施方案中, 对于临床应用, 胎儿基因型可能无法事先获得的, 并且母本单倍型可以通过本领域技术人员公知的方法和本文所述的方法来确定或推知。此处所用的 X 染色体仅用于说明的目的。

其他染色体,例如常染色体也可以用于这样的分析中。

[0279] 对于此处所描述的五病例,她们所有人都携带一个雄性胎儿。由于雄性胎儿仅从母亲遗传一个 X 染色体,而不是从父亲遗传 X 染色体,所以传递给胎儿的母本 X 染色体在母本血浆中会过多表现。从 X 染色体的 pter 到 qter 实施 RHD0 分析。从最接近 X 染色体 pter 的 SNP 开始,SPRT 分析可以确定母本血浆中 Hap I 或 Hap II 上的等位基因是否在统计学上显著过多表现。如果两个单倍型中没有一个在统计学上被显著过多表现,那么可以结合下一个 SNP 的等位基因计数用于进一步的 SPRT 分析。可以结合其他 SNP 用于分析直到 SPRT 过程将一个单倍型鉴定为统计学上显著过多表现。然后在下一个 SNP 处重新开始分类过程。

[0280] 图 38A 和 38B 示出了以病例 PW226 为例的 SPRT 分类结果。在该病例中,对于 X 染色体总共有 9 个成功的 SPRT 分类。对于每个 SPRT 分类,Hap I 上的等位基因被显示为在母本血浆样品中过多表现,这表明胎儿从母亲遗传了 Hap I。由于我们将 Hap I 限定为含有传递给胎儿的等位基因的单倍型,所以所有这些 SPRT 分类的结果都是正确的。

[0281] 五个病例的 RHD0 分析结果总结于图 39 中。成功 SPRT 分类的数量范围为 1 至 9。所有 SPRT 分类都是正确的。较高的部分胎儿 DNA 浓度与较高的分类数量相关。这是因为当胎儿 DNA 的部分浓度较高时,由于存在胎儿 DNA 导致的等位基因失衡更容易被检测到。因此,实现成功的 RHD0 分类可以需要较少的 SNP。限定的染色体区域因而可以被分成更多的 RHD0 域。我们的结果证实,可以在靶标富集之后对大规模测序数据进行 RHD0 分析。

[0282] 我们的数据进一步表明,靶向方法是进行 RHD0 分析的成本效益更高的方式。在无靶标富集的情况下,对于具有相似胎儿 DNA 浓度的样品,实现图 39 所示样品获得的平均深度需要大约 5 个流通池(即 40 个测序泳道)的测序(图 40)。在此我们证明,在靶标富集的情况下,仅通过一个泳道的测序就已经实现成功 RHD0 分类的 15 至 19 倍的平均测序深度。可选择地,当使用靶标富集时,用相对较少的附加成本就能够实现更高倍水平的测序覆盖度。更高水平的测序覆盖度能有效降低成功 RHD0 分类所需的基因组区域的大小,并因此提高分析的分辨率。

#### [0283] IV. 靶标富集

[0284] 自从 2004 年就已经了解到,母本血浆中循环的胎儿 DNA 分子通常短于母本 DNA (Chan KCA et al Clin Chem 2004;50:88-92;Li et al Clin Chem 2004)。然而,这种观察的分子基础仍未解决。在我们目前的研究中,我们在研究的血浆样品中产生  $3.931 \times 10^9$  个读取并在我们的生物信息学分析中使用 1-bp bins。由双末端读数的末端基因组坐标推导出每个测序的血浆 DNA 分子的大小。

[0285] 对于该分析,我们集中在其中父亲和母亲均为纯合的但等位基因不同的单核苷酸多态性(SNP)上。对于这类 SNP,胎儿是肯定杂合子。胎儿从父亲遗传的每个 SNP 的等位基因可以用作胎儿特异性标志物。为全基因组(图 41)和单独为每条染色体(图 42A-42C)确定胎儿序列(利用父系遗传的胎儿特异性等位基因)和总序列的大小。

[0286] 我们观察到,母本血浆中胎儿和母本 DNA 之间最显著的差异是,相对于 143bp 峰,166bp 峰降低(图 41)。最丰富的总序列(主要为母本的)的长度为 166bp。胎儿 DNA 和总 DNA 的大小分布中最显著的差异是胎儿 DNA 表现出 166bp 的峰降低(图 41)和 143bp 的峰相对突出。后者可能对应于剪切从核小体到其约 146bp 的核心颗粒的约 20-bp 连接体片段

(Lewin B, in Gene IX, Jones and Bartlett, Sudbury, 2008, pp. 757-795)。

[0287] 从大约 143bp 及向下, 胎儿 DNA 和总 DNA 的分布显示了让人联想到核酸酶切割的核小体的 10bp 周期性。这些数据提示, 血浆 DNA 片段来源于凋亡酶促加工。相比之下, 定位到非组蛋白连接的线粒体基因组的读取的大小分析没有显示出该核小体模式 (图 41)。利用 Y 染色体和选择的多态性遗传标志物, 这些结果为胎儿和母本 DNA 之间的已知的大小差异提供了以前不了解的分子解释 (Chan KCA et al Clin Chem 2004;50:88-92; Li et al Clin Chem 2004;50:1002-1011; 美国专利申请 20050164241; 美国专利申请 20070202525), 并且表明, 在整个基因组上都存在这类大小差异。这种差异最可能的解释是循环的胎儿 DNA 分子由更多的分子组成, 在所述分子中, 已经从核小体剪切了约 20bp 的连接体片段。

[0288] 考虑到这些观察结果, 有很多能够富集样品的胎儿 DNA 的方式。在一个实施方案中, 可以使用优先与连接体片段结合的试剂。与母本血浆中胎儿来源的 DNA 相比, 这类试剂预期优先与母系来源的 DNA 结合。这类试剂的一个实例是抗体。这类抗体的一个靶标是与组蛋白 HI 结合的靶标。已知组蛋白 HI 与连接体片段结合。这类抗体的一个应用是通过阴性选择进行胎儿 DNA 的富集, 即通过优先免疫沉淀母本血浆中含有含连接体、组蛋白 HI 的片段的母系来源的 DNA。此外, 已知 HI 具有很多变体, 它们当中的一些在表达时表现出组织特异性变异 (Sancho M et al PLoS Genet 2008;4:e1000227)。这些变体可被进一步开发以区别胎儿 (主要为胎盘的) 和母本 (主要为造血的) (Lui YYN et al Clin Chem 2002;48:421-427) DNA。例如, 可以靶向主要由滋养层细胞表达的组蛋白 HI 变体, 从而优先并阳性选择母本血浆中胎儿来源的 DNA。该策略还可以应用于表现出组织特异性、尤其是滋养层特异性表达模式的其他组蛋白或其他核小体蛋白。

[0289] 考虑到母本 DNA 的尖锐的 166bp 的峰, 富集胎儿 DNA 的另一种可能性是, 为长度为  $166 \pm 2$ bp 的 DNA 片段的阴性选择设计系统。例如, 基于毛细管电泳或高效液相色谱的系统能够允许 DNA 分子的精确大小测量和分离。阴性选择的另一种方法是, 在测序数据的生物信息分析期间在电脑 (in silico) 中进行。

[0290] 由于血浆中的其他 DNA 种类, 例如, 肿瘤 DNA (Vlassov VV et al. Curr Mol Med 2010;10:142-165) 和移植器官 DNA (Lo YMD et al Lancet 1998;351:1329-1330), 预期也与母本血浆中的胎儿 DNA 共有这些特征, 因此上文在 (1) 和 (2) 中列出的策略也能用于富集这些 DNA 种类。

[0291] 按照一个实施方案, 提供了通过靶向核小体的连接体片段来差别富集人血浆或血清中 DNA 种类的方法。在一个实施方案中, 通过去除以下之一进行富集: 母系来源的 DNA 或来源于造血细胞的 DNA。在另一个实施方案中, 所述靶向涉及优先与核小体的连接体片段的蛋白或核酸组分结合的试剂 (诸如抗体或另一类蛋白)。在另一个实施方案中, 靶向试剂与组蛋白 HI 或结合核小体的连接体片段的另一蛋白选择性结合。在另一个实施方案中, 靶向试剂与组蛋白 HI 的母本或血液学变体或结合核小体的连接体片段的另一蛋白结合。在一个实施方案中, 通过免疫沉淀或与固体表面的结合来实施 DNA 的去除。

[0292] 按照另一实施方案, 差别富集母本血浆或血清中的胎儿 DNA 的方法包括: (a) 使用与核小体的连接体片段的一个或多个组分结合的抗体; (b) 通过免疫沉淀或捕获至固体表面来去除结合的部分; 以及 (c) 收获含有部分浓度增加的胎儿 DNA 的未结合的部分。

[0293] 利用任何合适的计算机语言,如 Java、C++ 或使用例如常规或面向对象技术的 Perl,本申请所述的任何软件组件或函数可以作为由处理器运行的软件代码来执行。软件代码可存储在用于存储和 / 或传输的计算机可读介质上作为一系列指令或命令,合适的介质包括随机存取存储器 (RAM)、只读存储器 (ROM)、诸如硬盘或软盘的磁性介质或诸如光盘 (CD) 或 DVD (多功能数码光盘) 的光学介质、闪存等。计算机可读介质可以是此类存储或传输装置的任何组合。

[0294] 此类程序也可以利用适合通过有线、光学和 / 无线网络传播的载波信号来编码和传输,该网络符合包括国际互联网在内的各种协议。因此,本发明实施方案的计算机可读介质,可以利用此类程序编码的数据信号产生。用程序代码编码的计算机可读介质可以与兼容的装置组装,或由其他装置 (如经由互联网下载) 独立地提供。任何此类计算机可读介质可以位于一个计算机程序产品上或在该产品内 (例如,硬盘或整个计算机系统),并且可以存在于系统或网络内不同计算机程序产品上或在该产品内。计算机系统可以包括显示屏、打印机或向用户提供本文所提到的任何结果的其他合适的显示器。

[0295] 计算机系统的实例显示于图 43 中。图 43 中所示的子系统经由系统总线 4375 相互连接。图 43 显示了其他子系统,如打印机 4374、键盘 4378、硬盘 4379、与显示适配器 4382 连接的显示屏 4376 等。与 I/O 控制器 4371 连接的外围装置和输入 / 输出 (I/O) 装置,可以通过本领域已知的任何数量的方式连接至计算机系统,如串行端口 4377。例如,串行端口 4377 或外部界面 4381 可用于将计算机装置连接至诸如互联网的广域网、鼠标输入装置或扫描仪。经由系统总线互联允许中央处理器 4373 与每个子系统通讯,并控制系统内存 4372 或硬盘 4379 的指令的执行以及子系统间信息的交换。系统内存 4372 和 / 或硬盘 4379 是计算机可读介质的具体表现。本文提到的任何值都可以从一个组件向另一个组件输出,且可以向用户输出。

[0296] 计算机系统可以包括多个相同的组件或子系统,例如通过外部界面 4381 或通过内部界面连接在一起。在一些实施方案中,计算机系统、子系统或装置在网络中通讯。在这种情况下,一台计算机可以被认为是一个客户而另一个计算机被认为是服务器,其中每个计算机都是同一计算机系统的一部分。客户和服务器各自都可以包括多个系统、子系统或组件。

[0297] 具体实施方案的详细细节可以以任何合适的方式组合或根据本文所示和描述的那些而改变,这不脱离本发明实施方案的实质和范围。

[0298] 出于示例和描述的目的,上文呈现了本发明示例性实施方案的描述。其并非意图穷举或将本发明限制为所述的准确形式,并且根据上文的教导,可以做出许多修饰和变化。为了最好地解释本发明的原理及其实际应用而选择和描述了实施方案,由此使本领域技术人员在各种实施方案中,并且通过适于所考虑的具体用途的各种修饰来最佳地利用本发明。

[0299] 出于各种目的,将本文所引用的所有出版物、专利和专利申请通过引用全文并入。

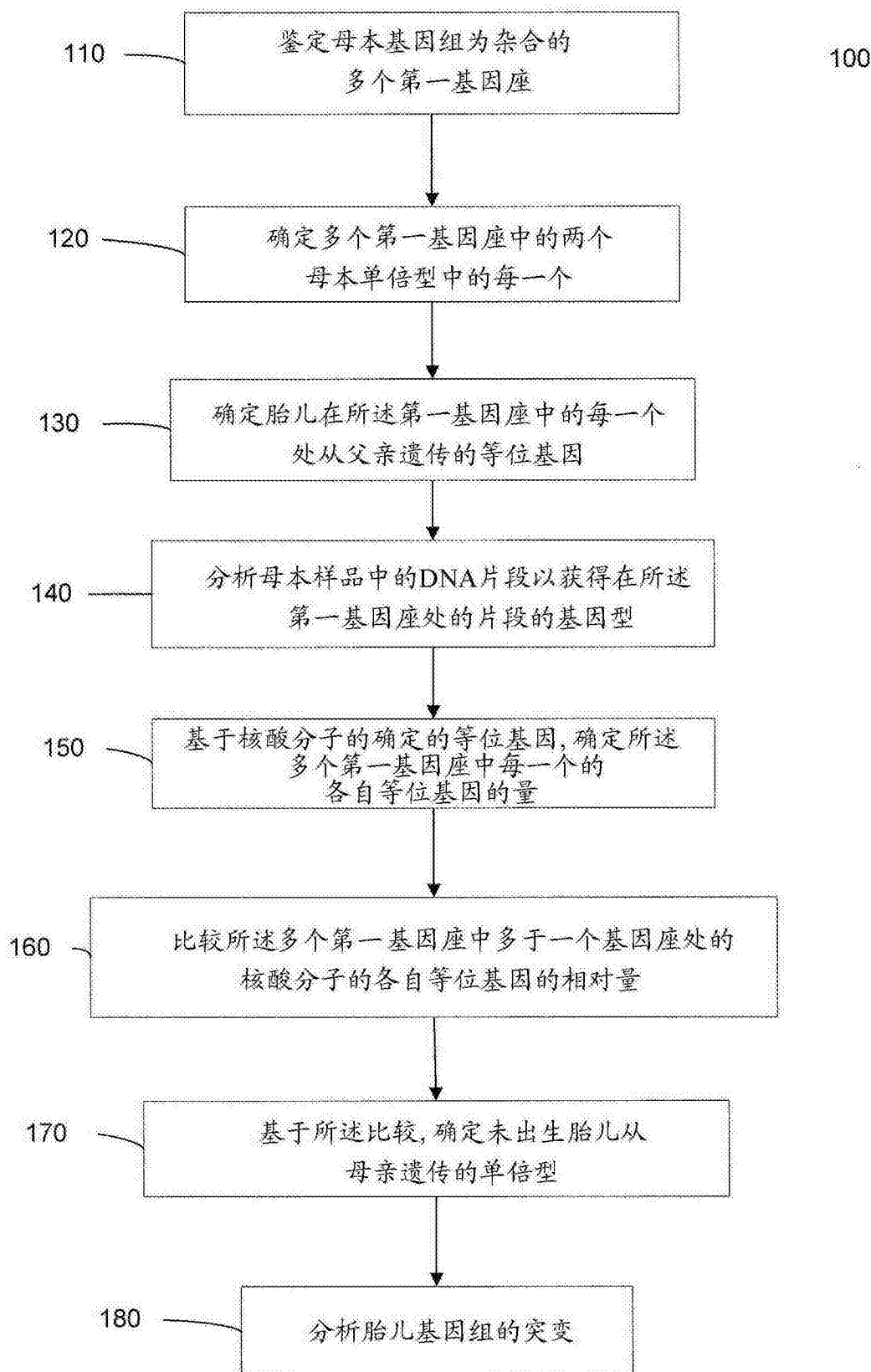


图 1

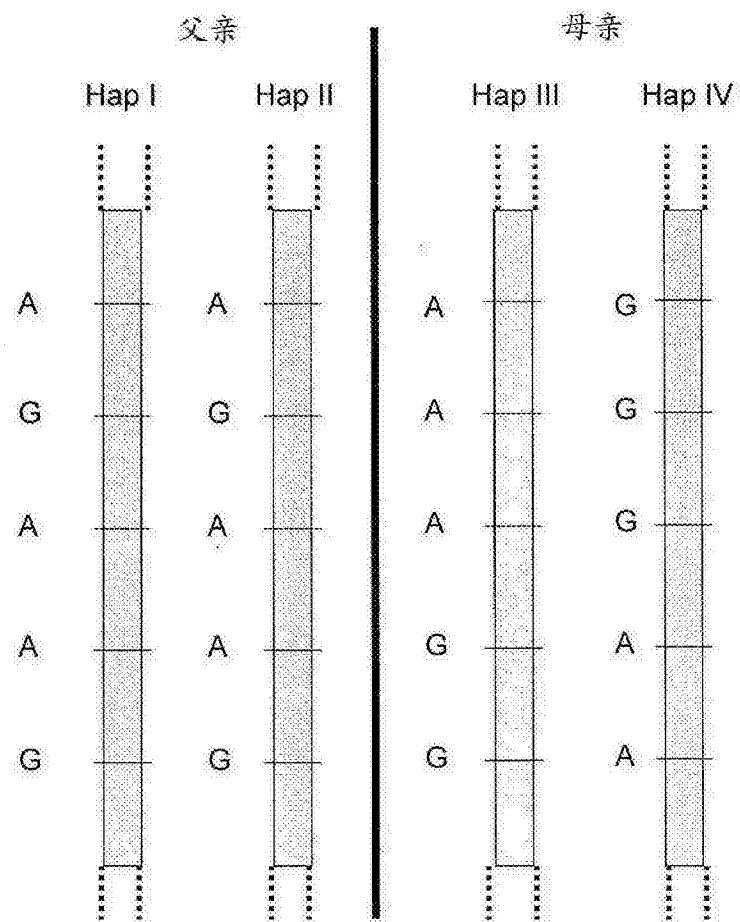


图 2



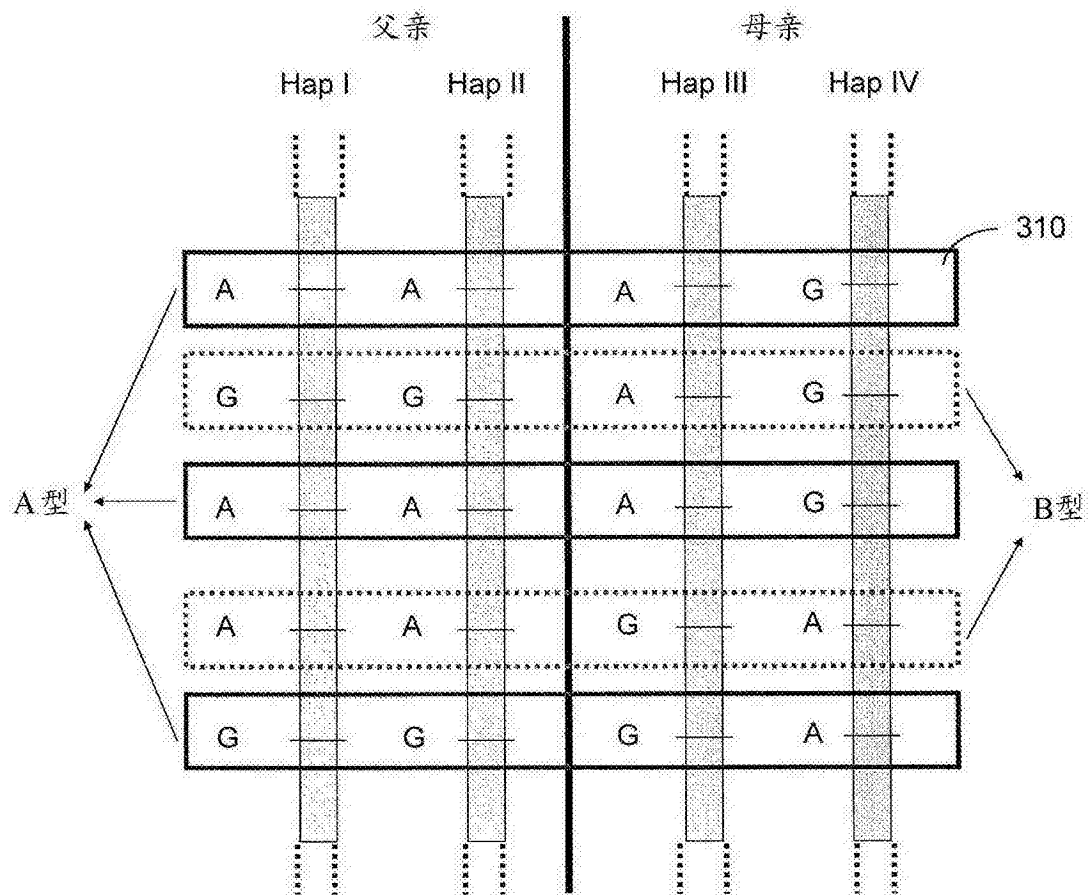


图 3

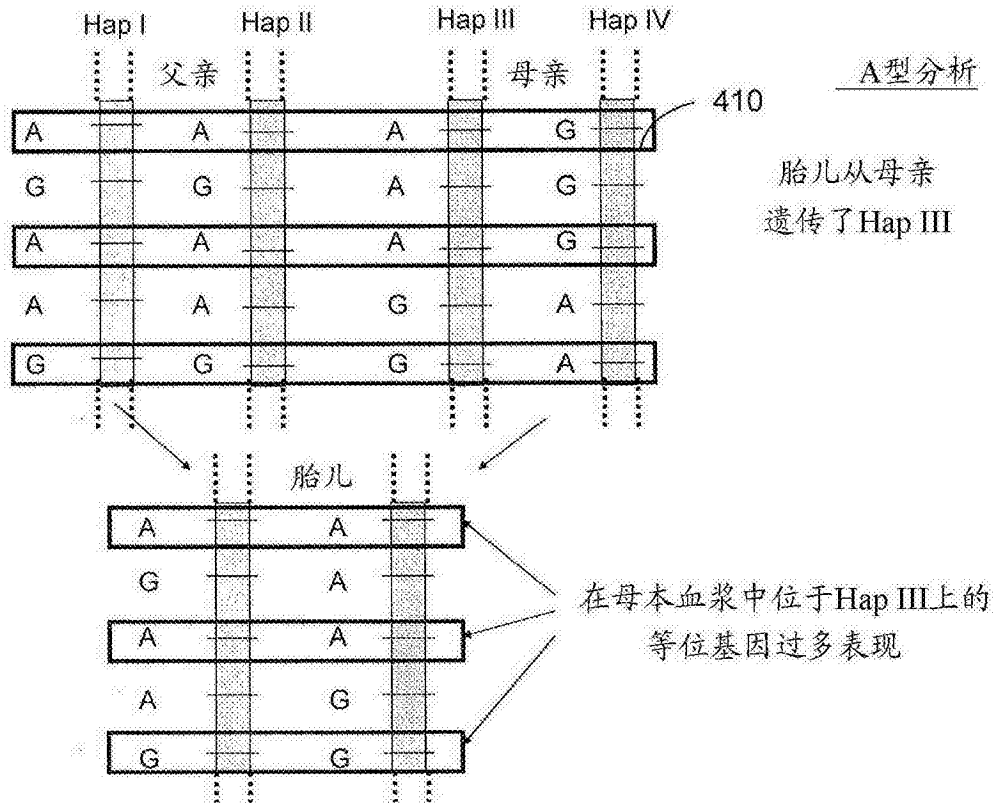


图 4A

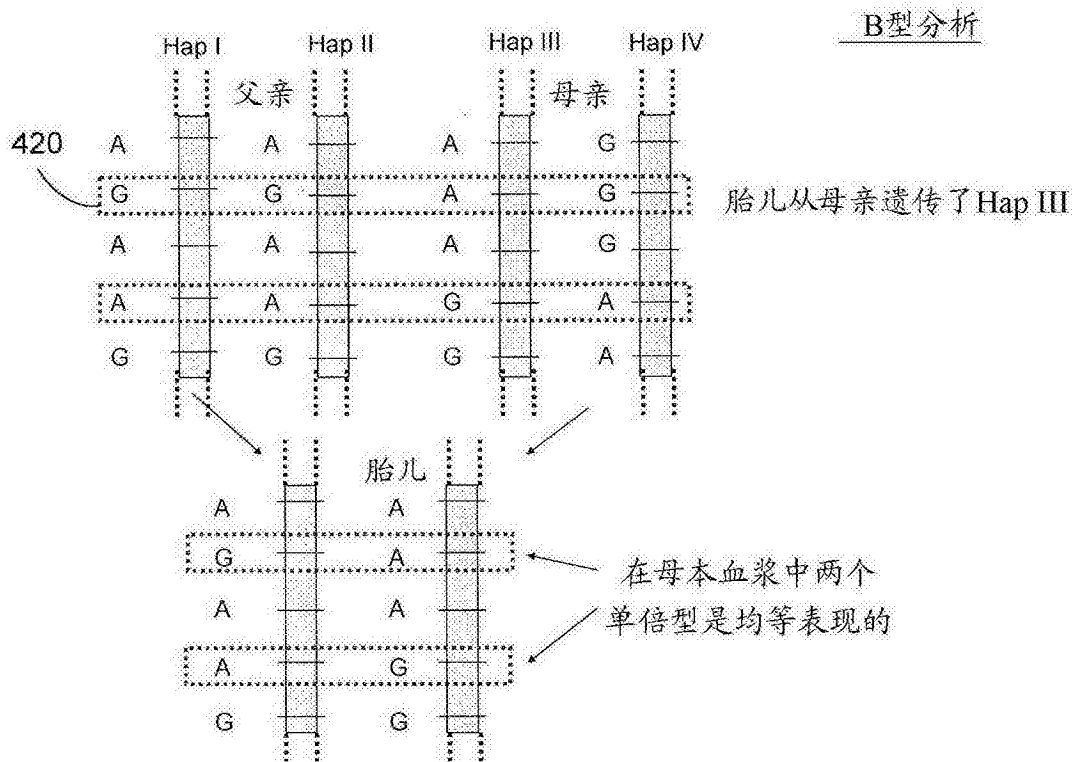


图 4B

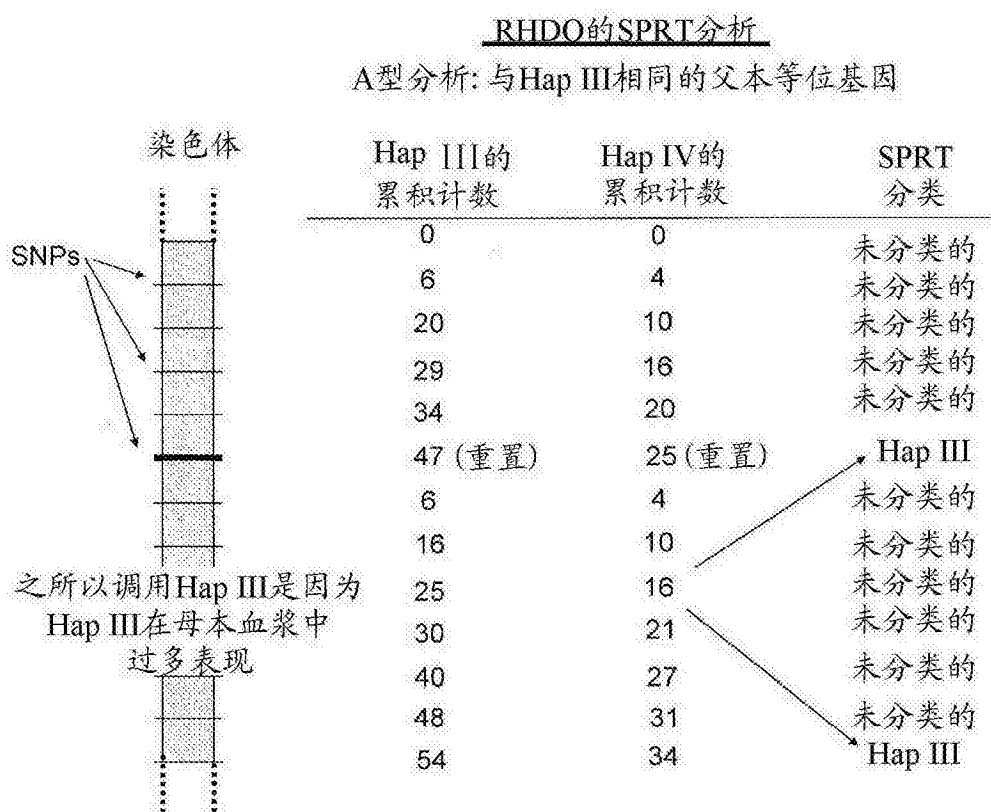


图 5A

RHDO的SPRT分析

A型分析: 与Hap IV相同的父本等位基因

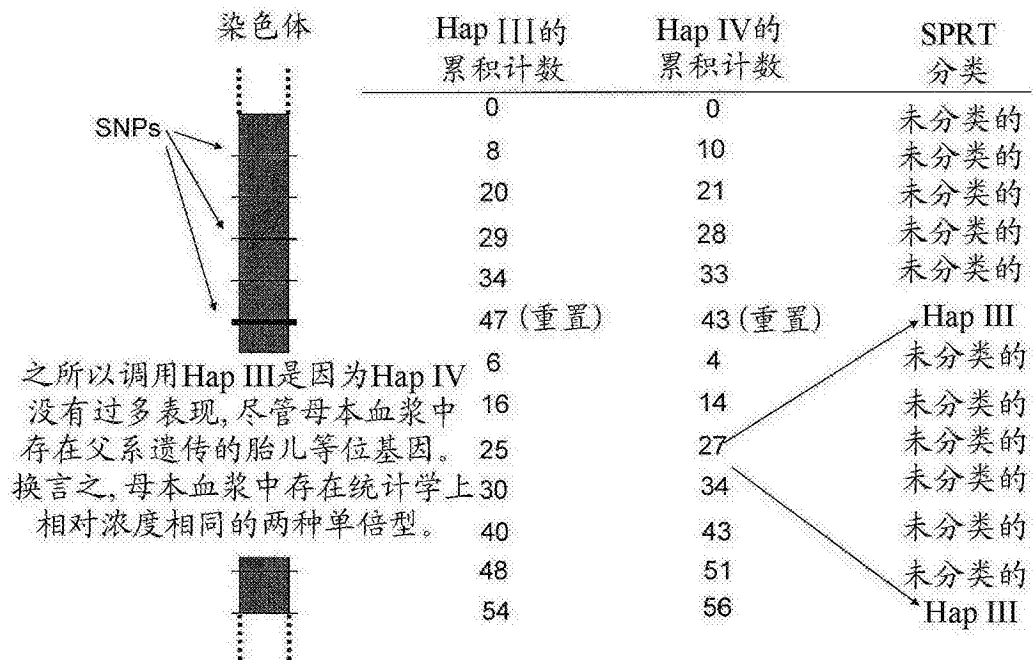


图 5B

改变分类阈值对SPRT曲线的影响  
(与Hap III相同的父本等位基因)

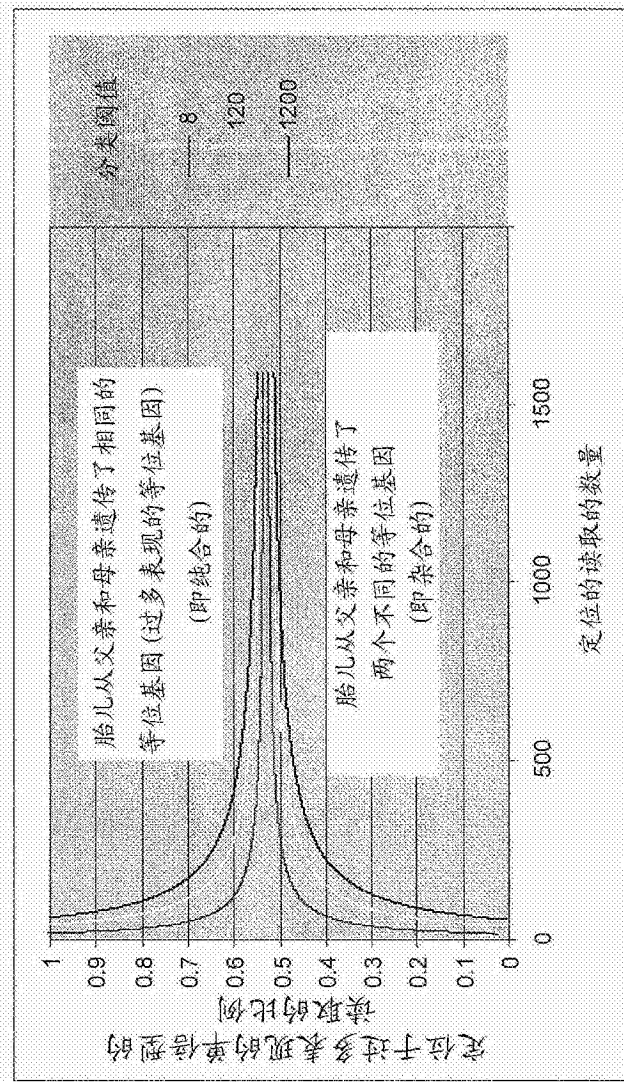


图 6

700

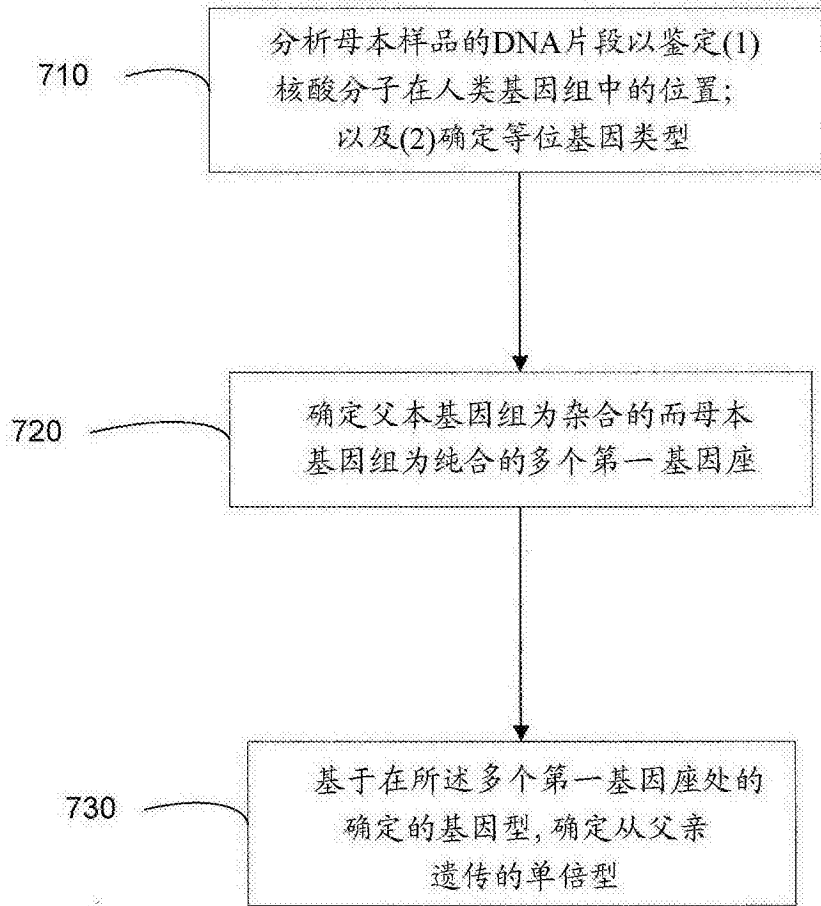


图 7

800

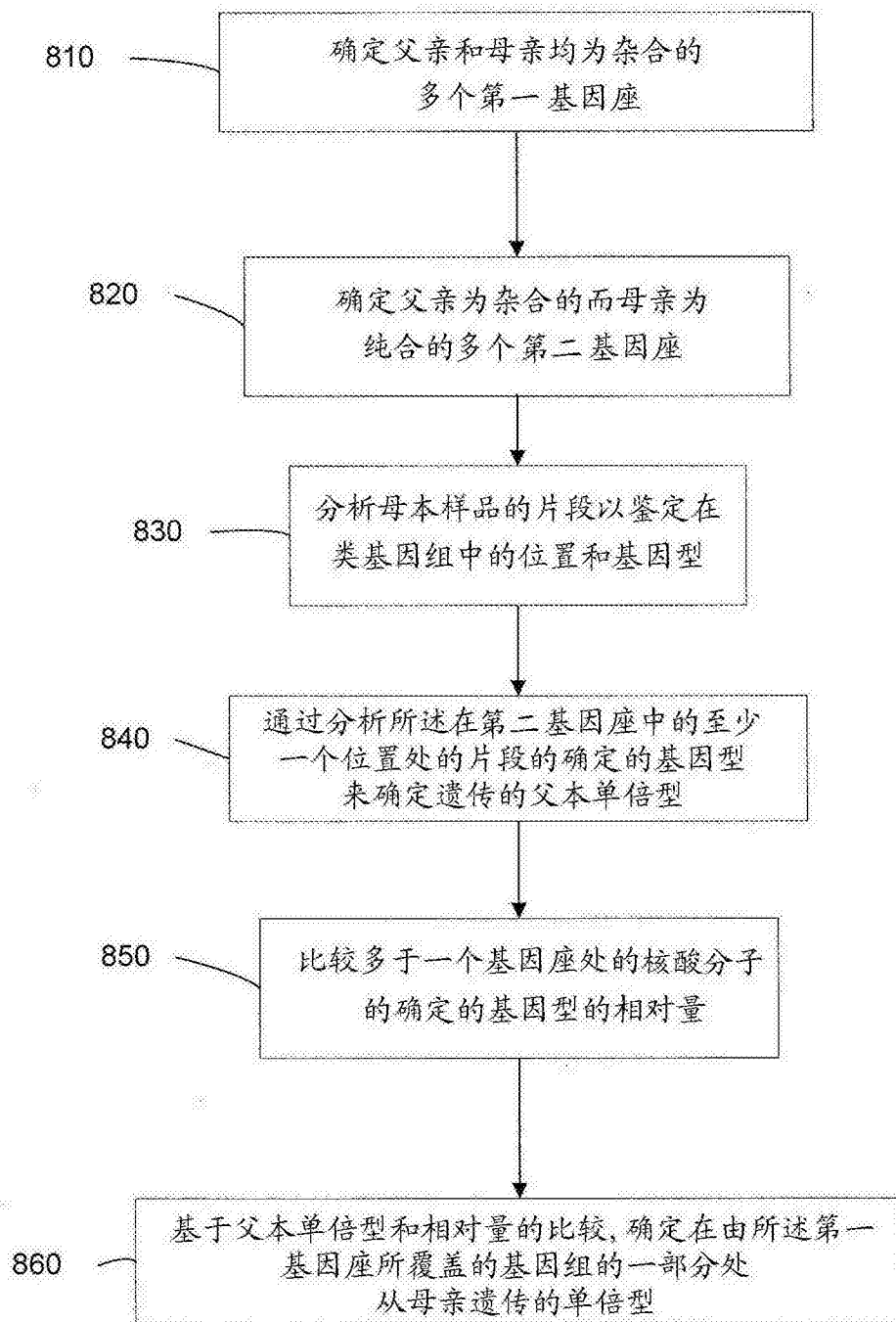


图 8

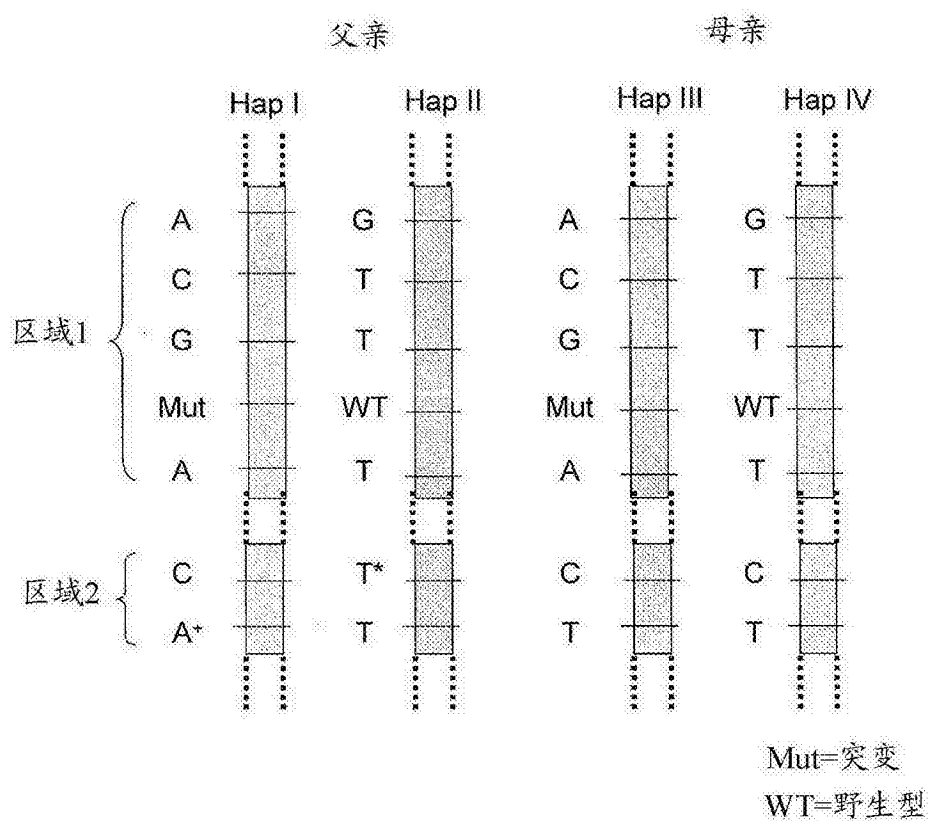


图 9



1000

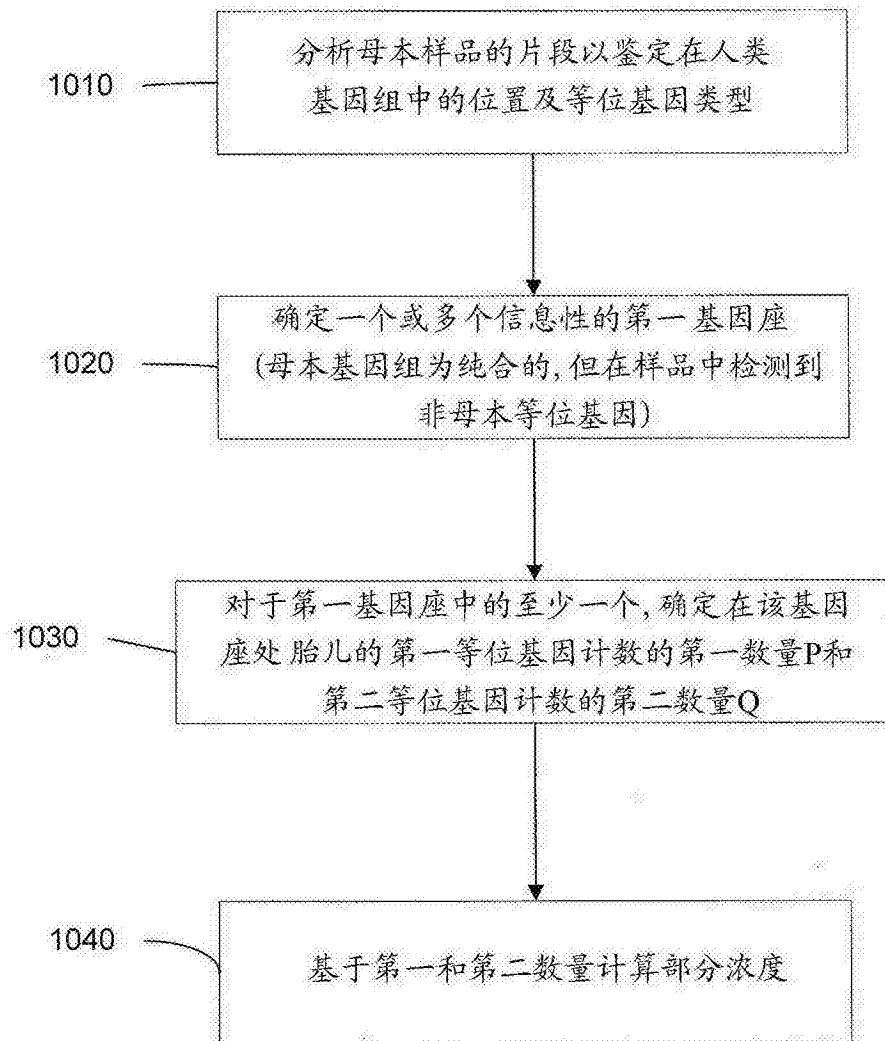


图 10

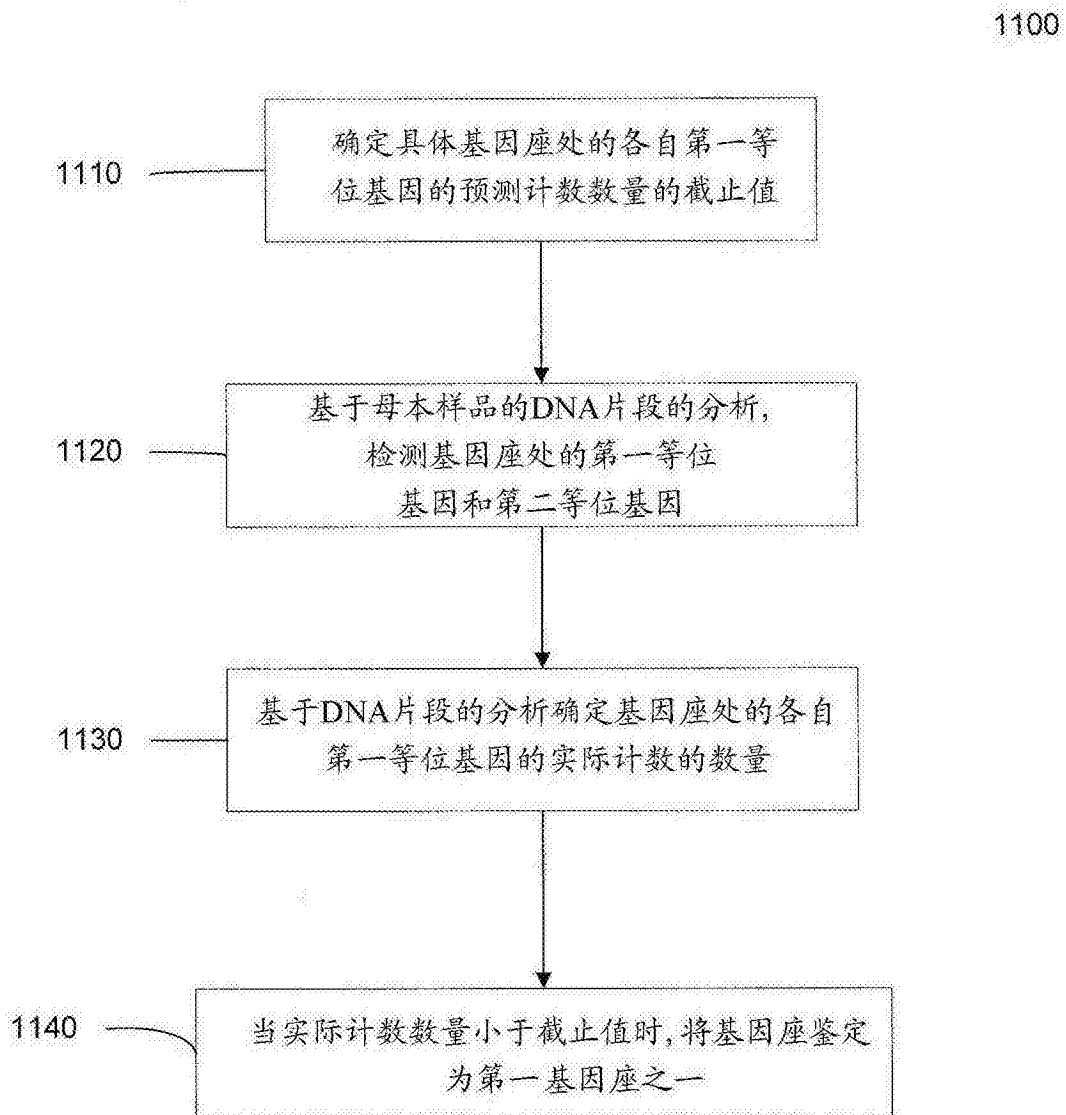


图 11

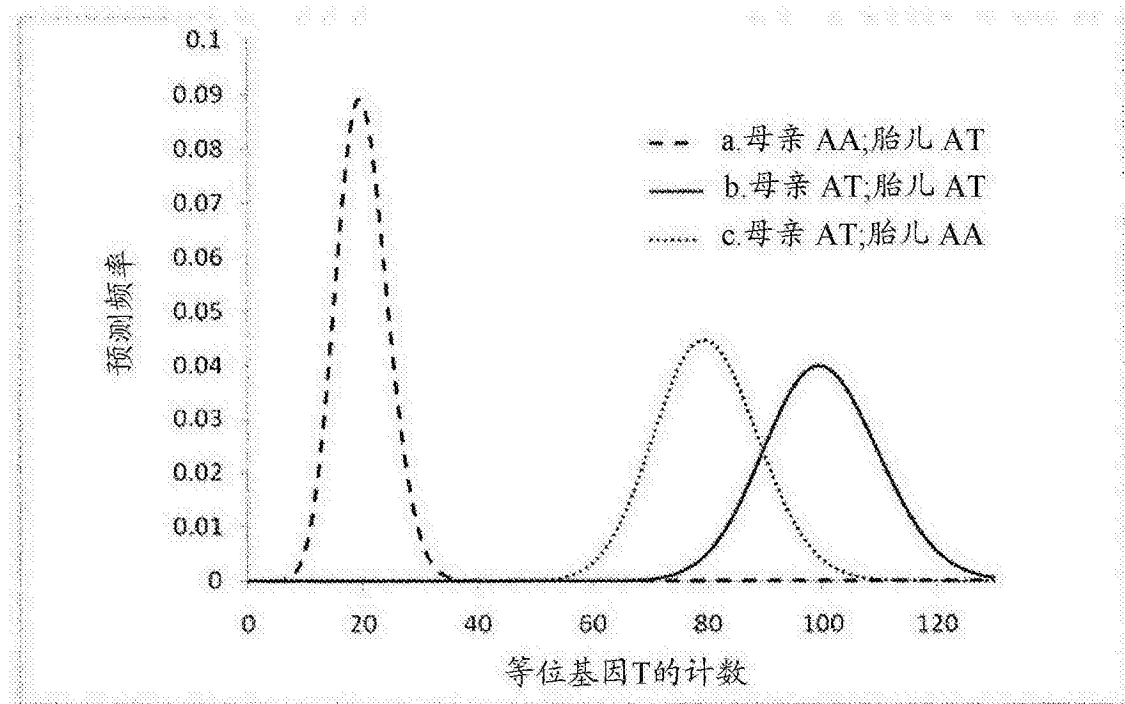


图 12A

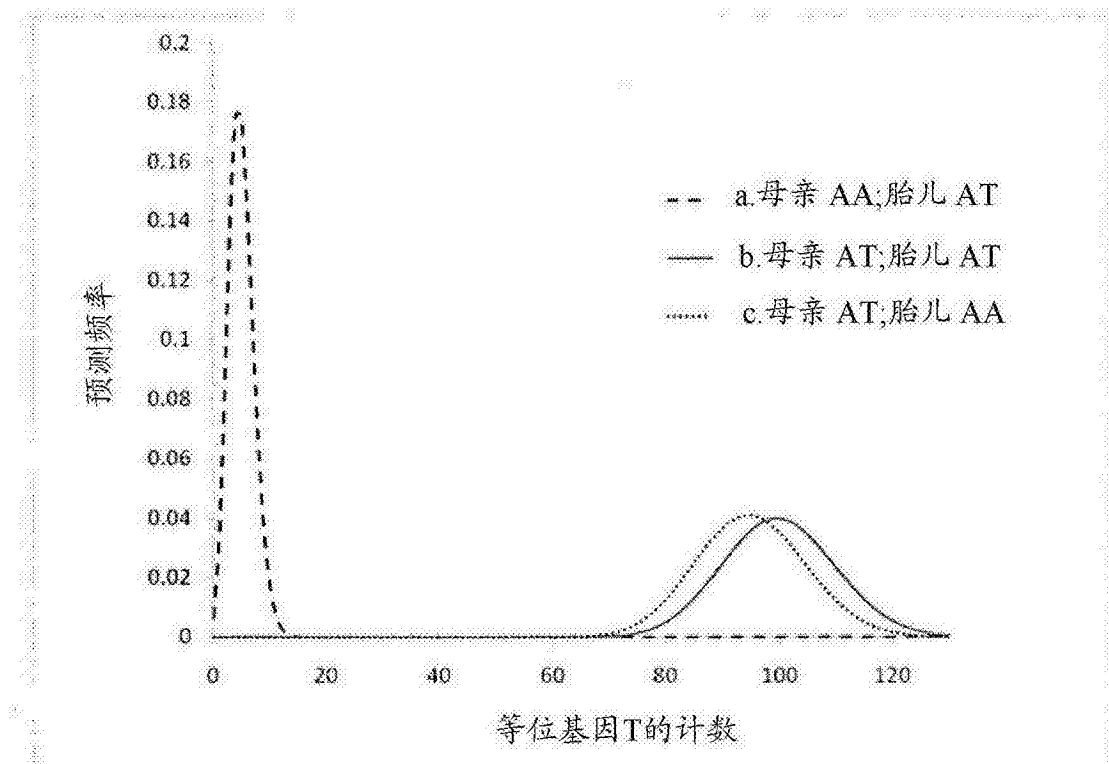


图 12B

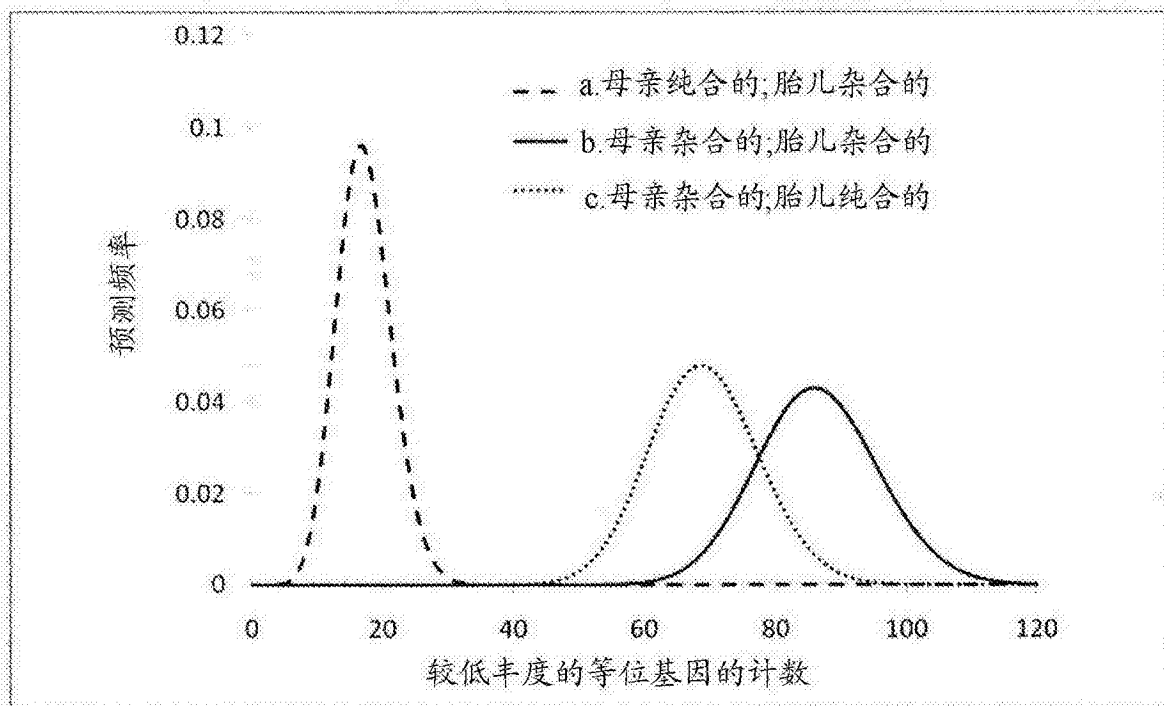


图 13A

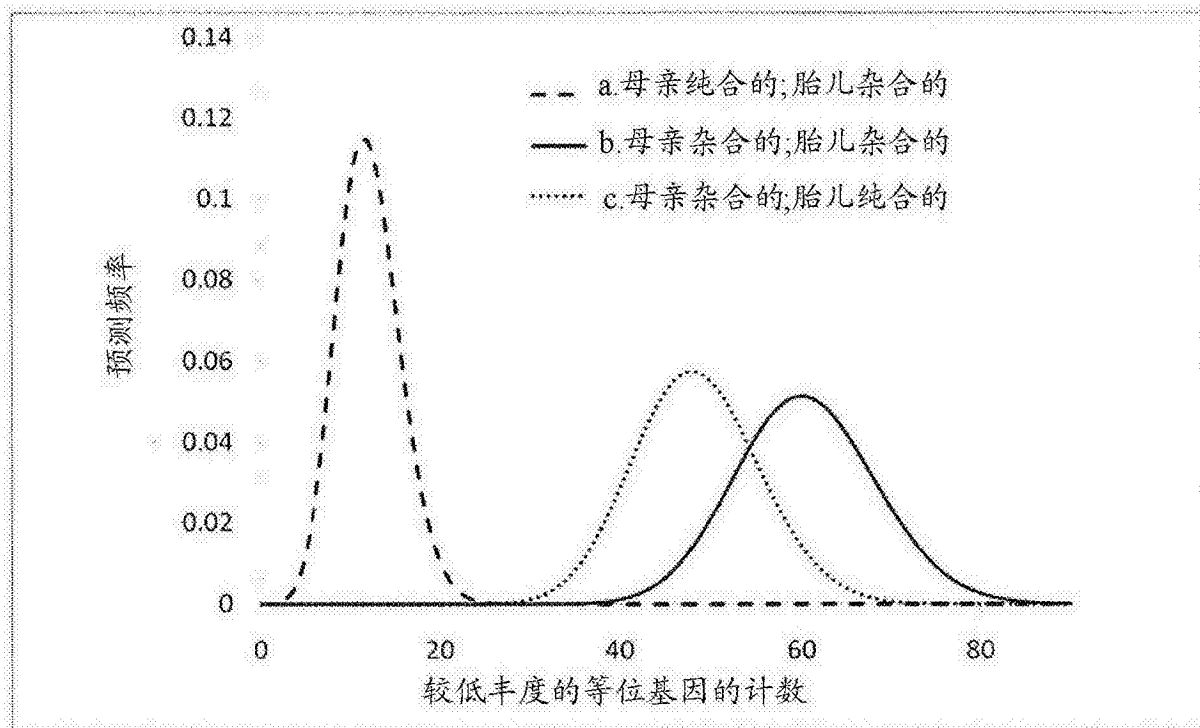


图 13B

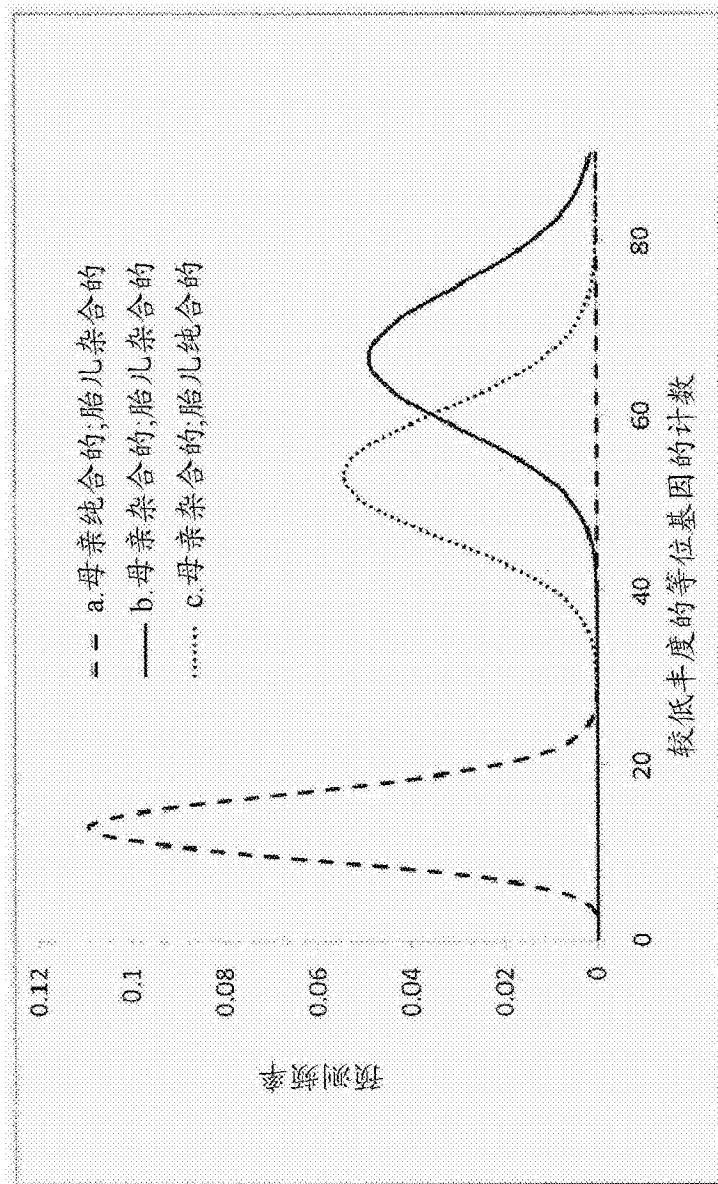
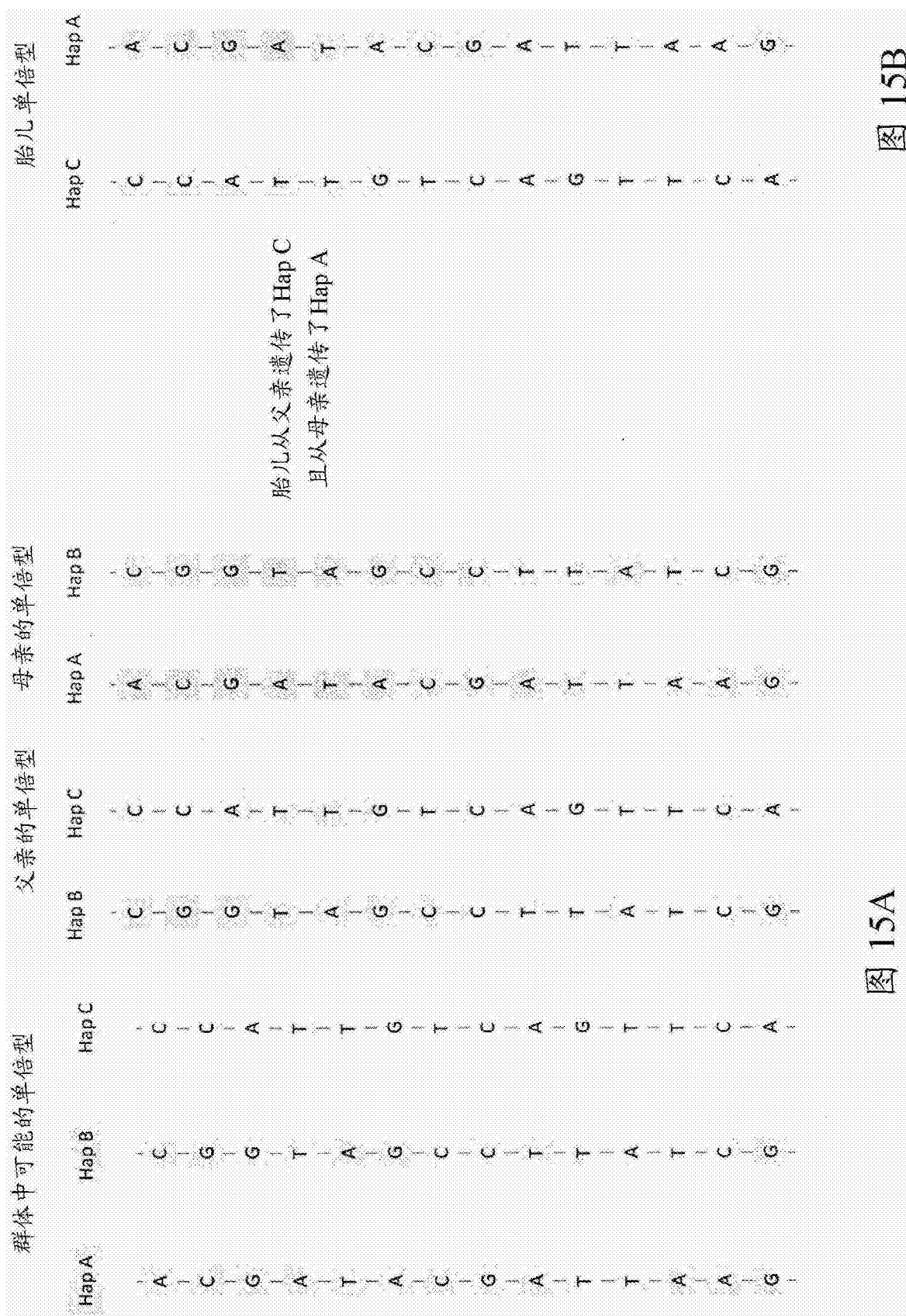


图 14



1600

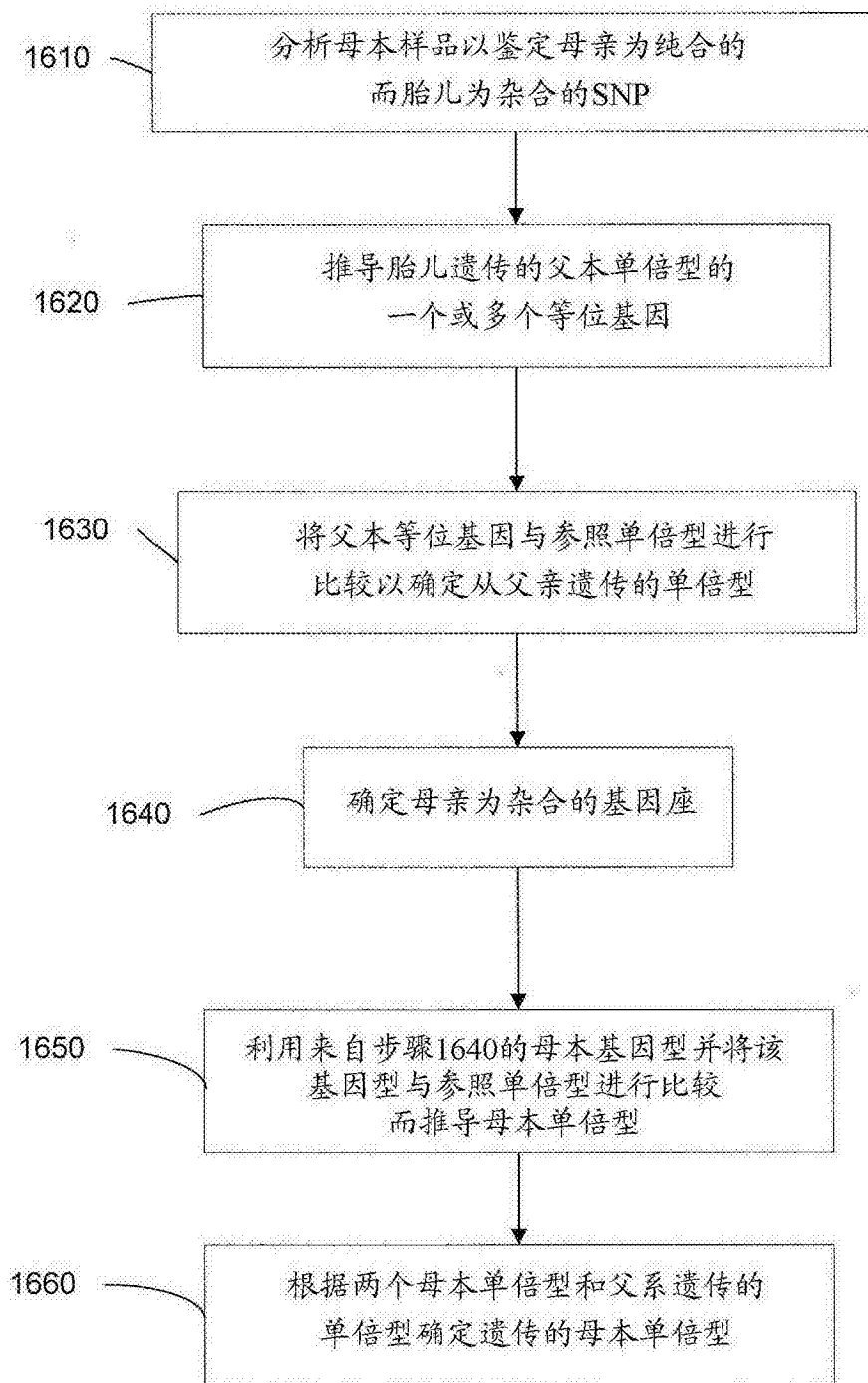


图 16

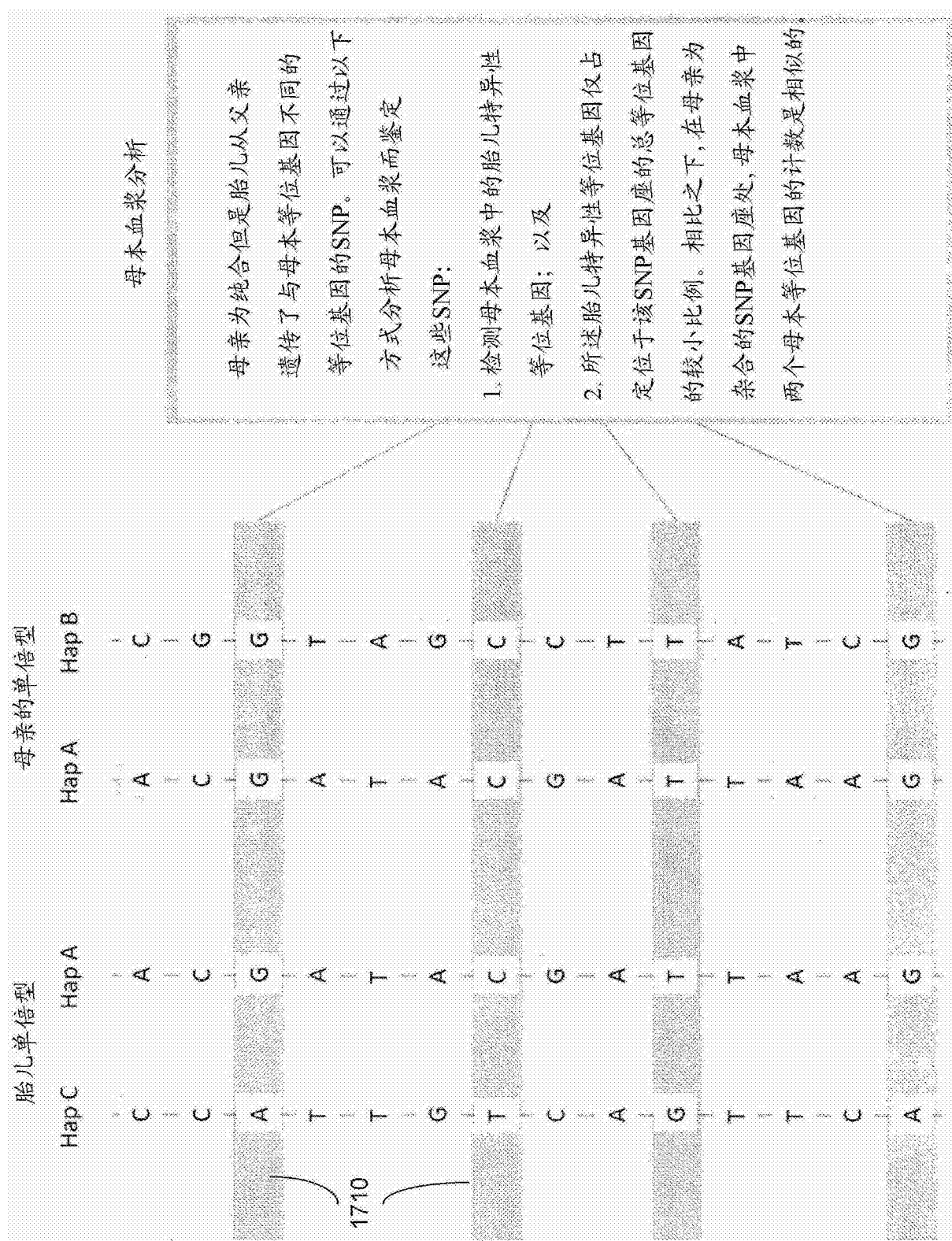


图 17



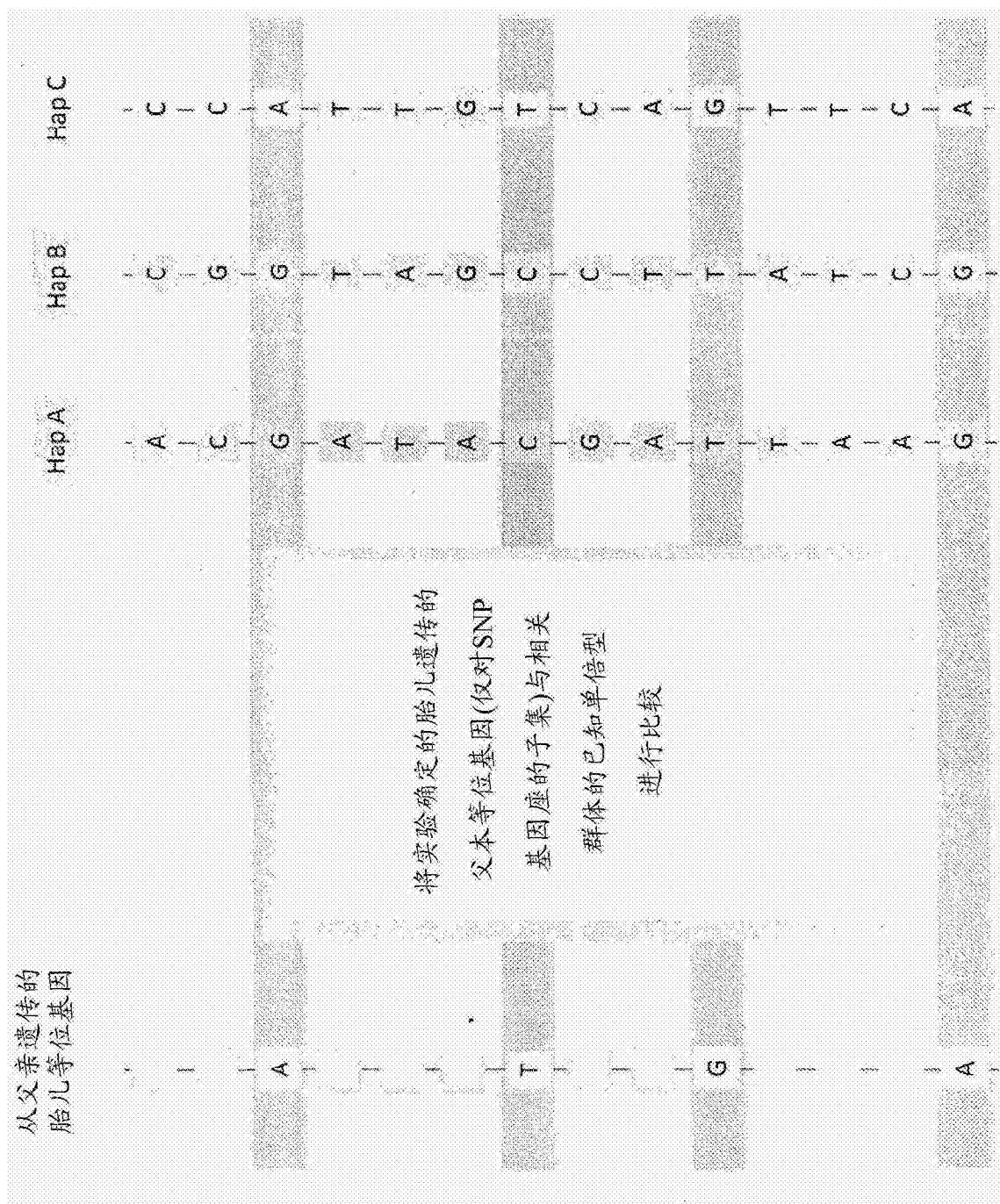


图 18

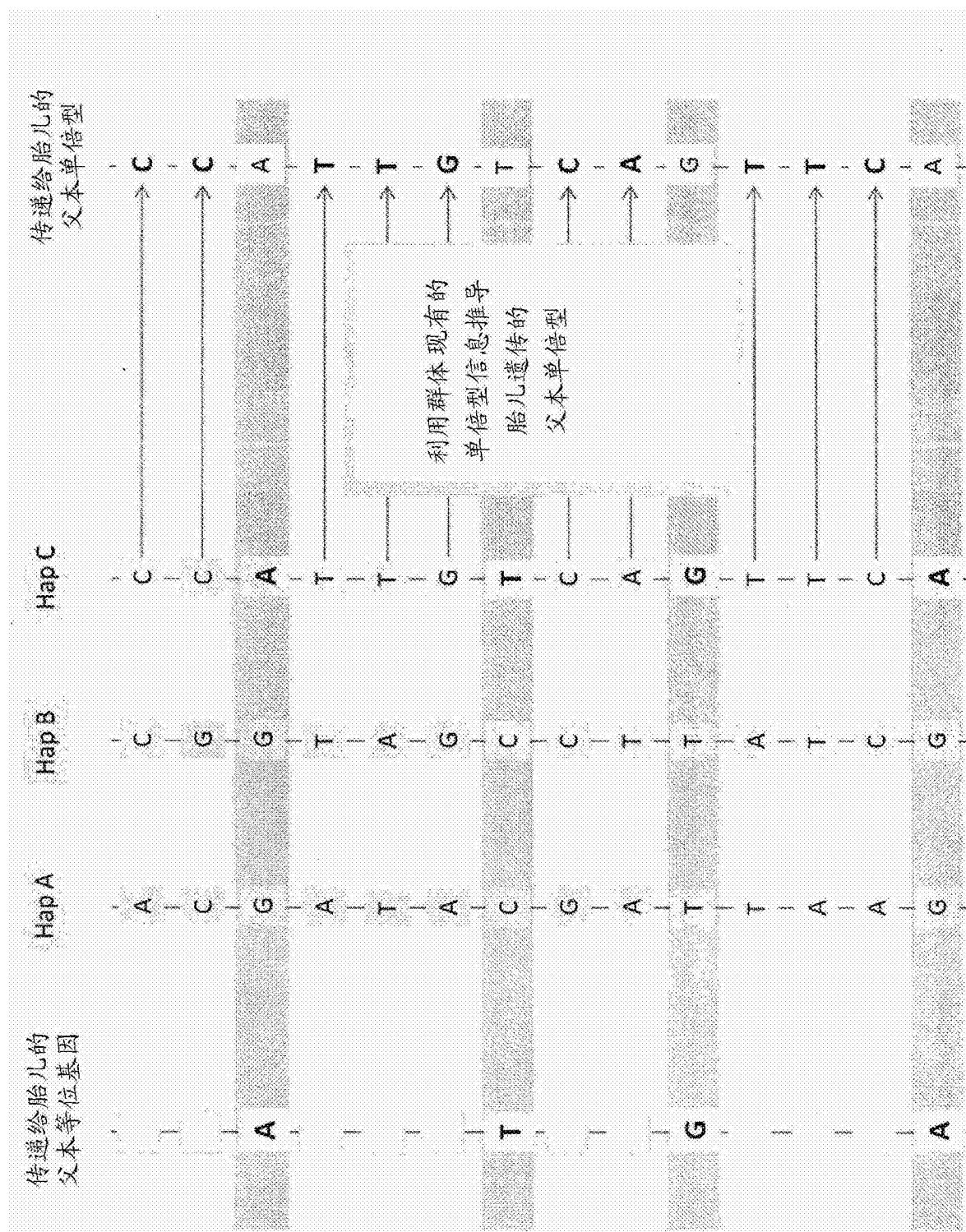


图 19

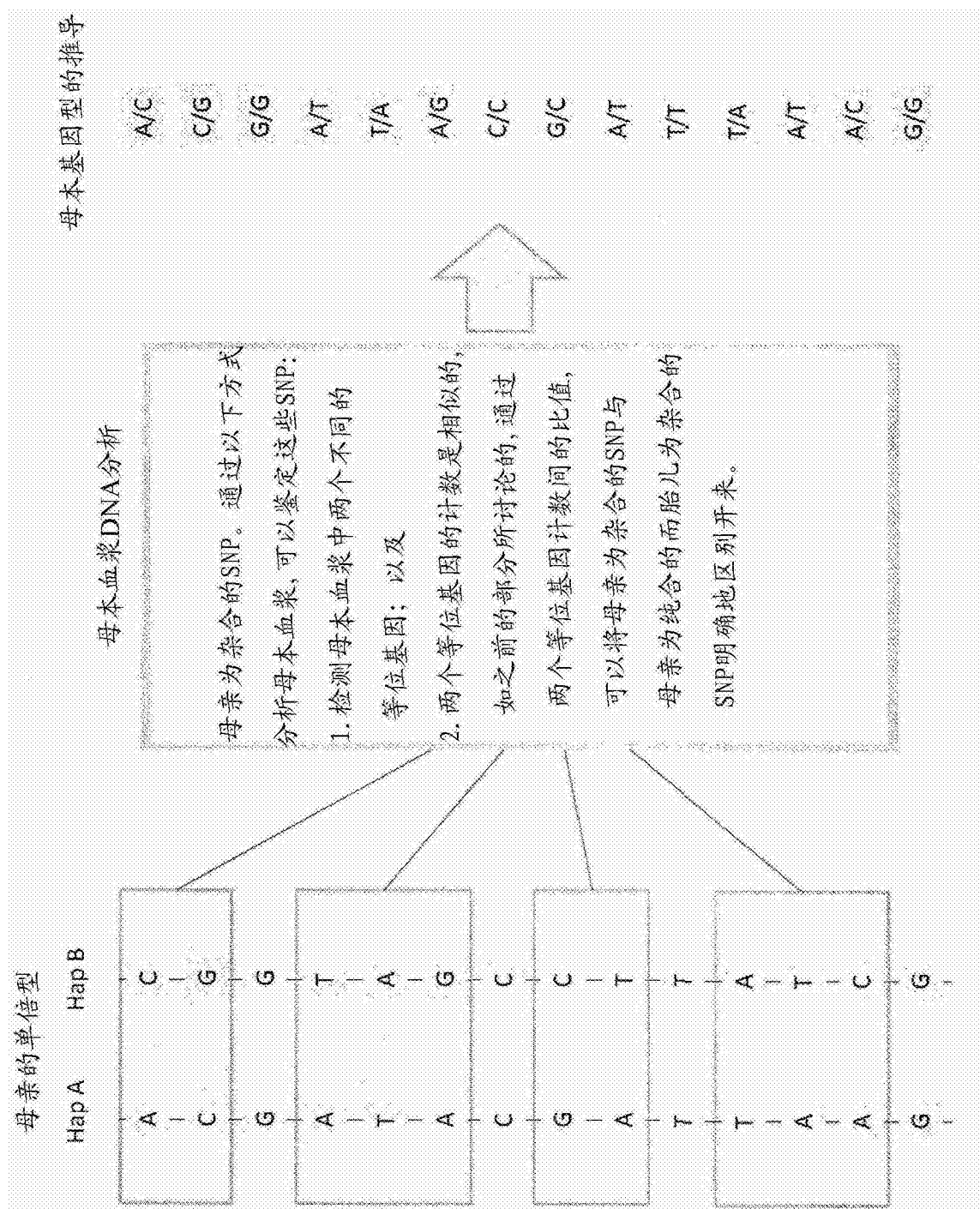


图 20

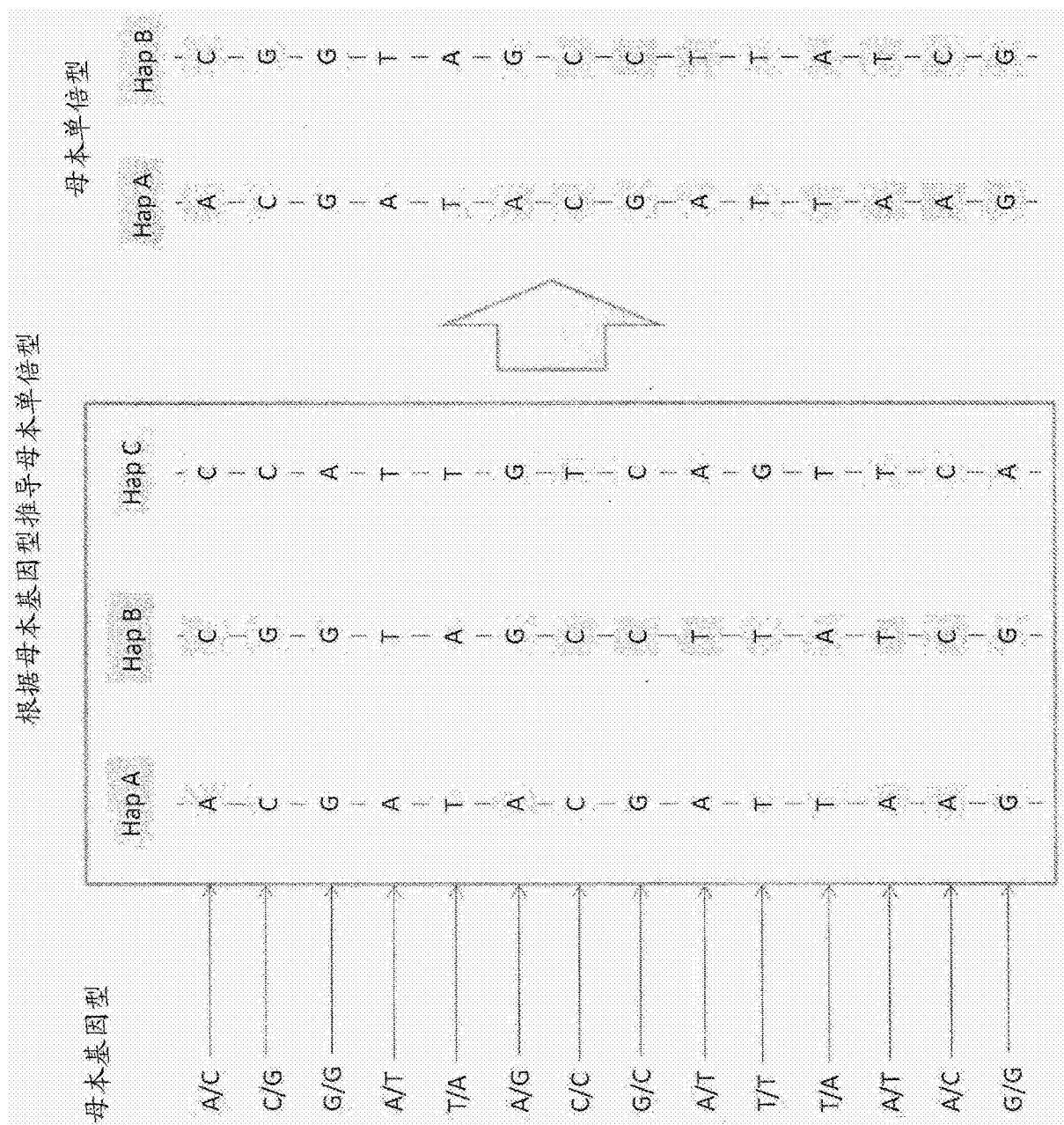


图 21

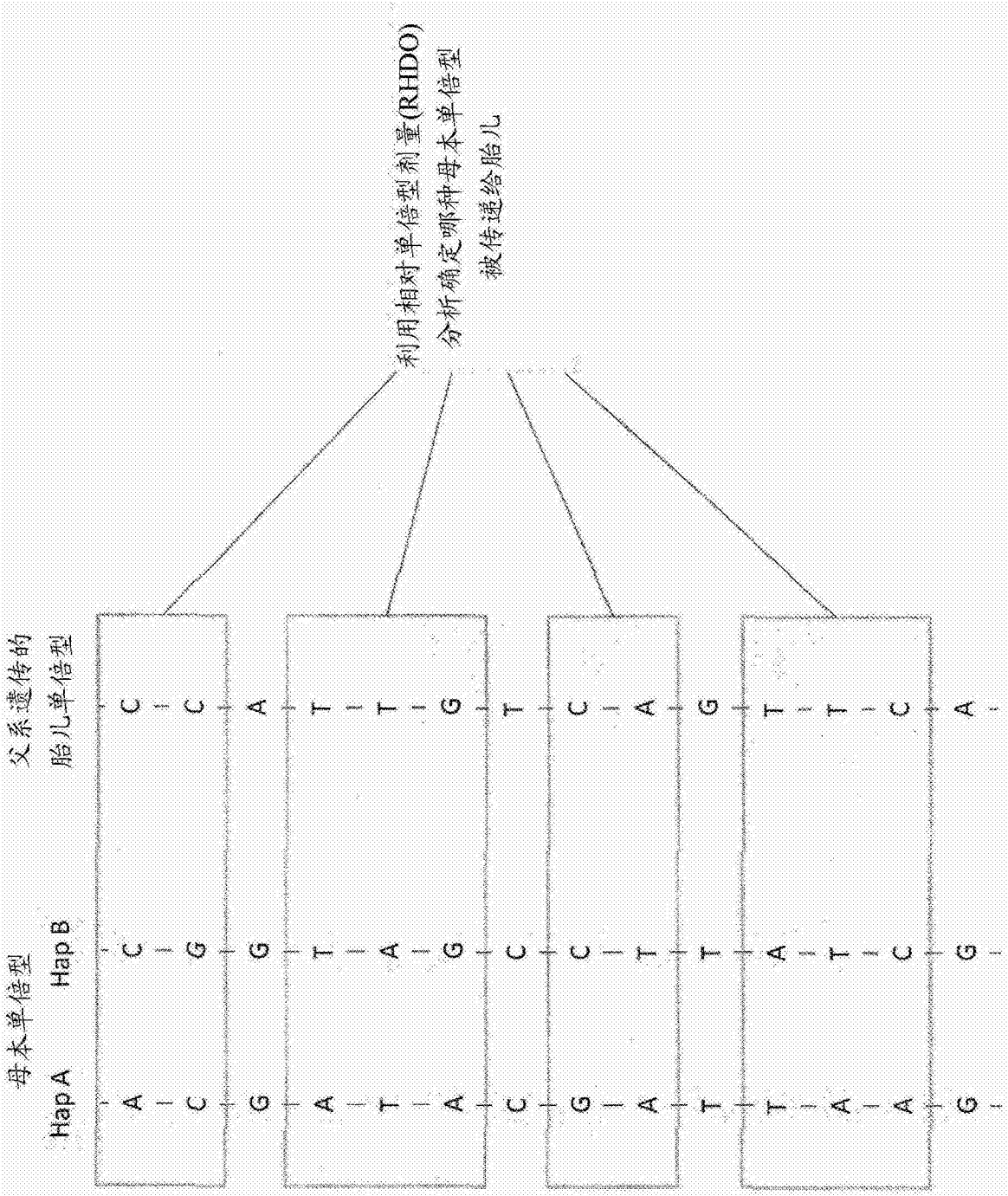


图 22

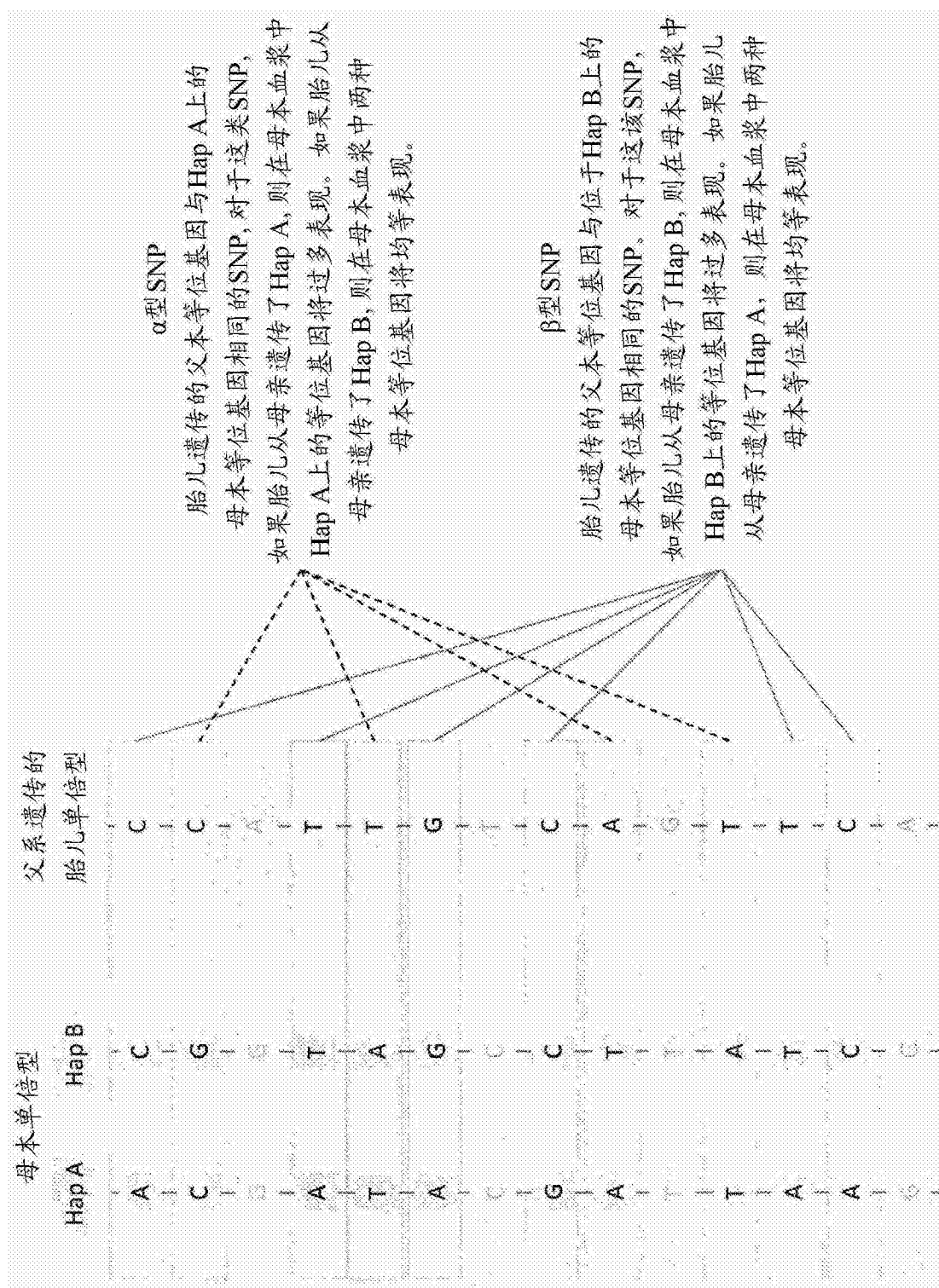


图 23

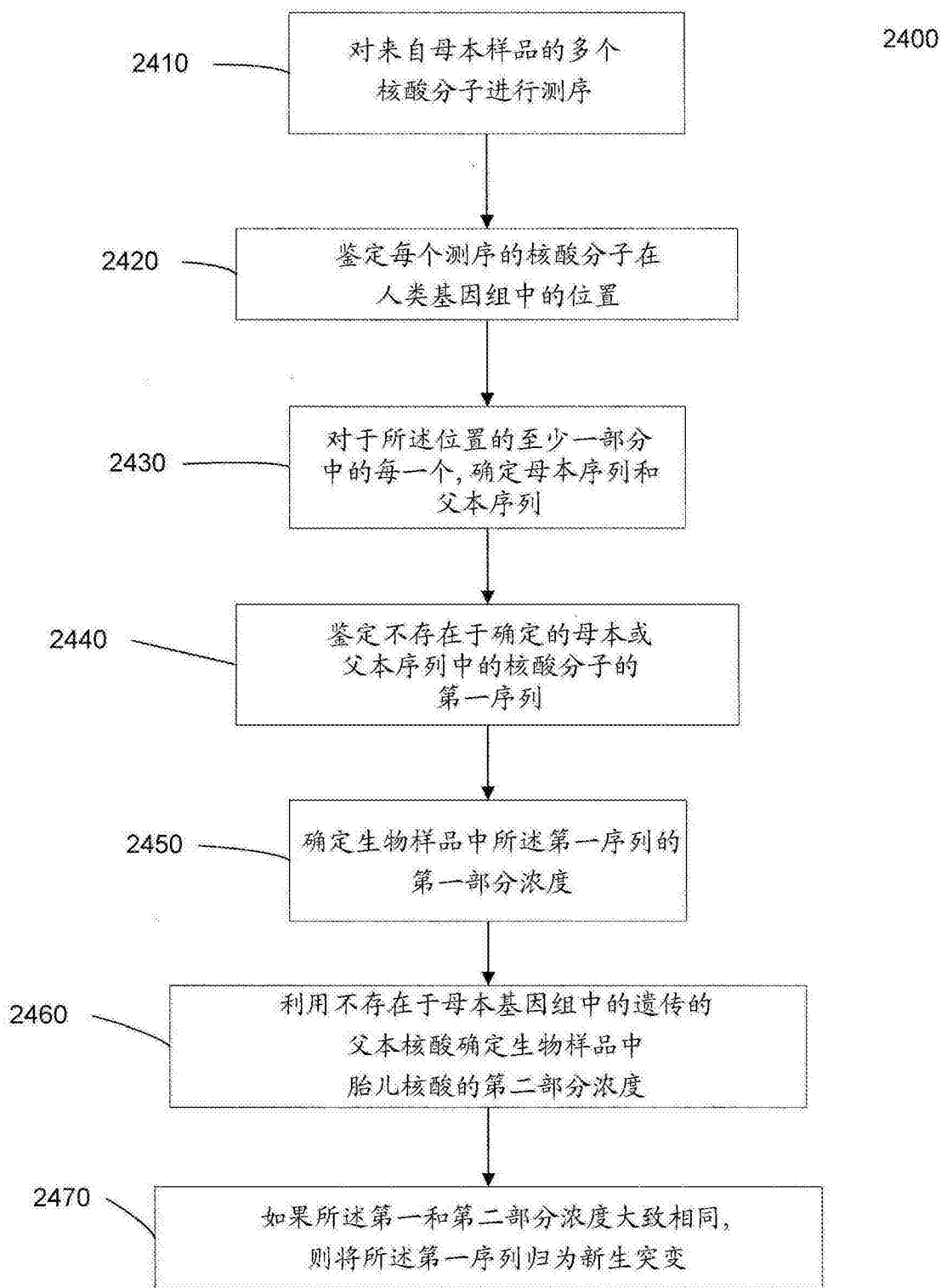


图 24

| 基因型组合                 | SNP的数量  | 百分比(%) |
|-----------------------|---------|--------|
| 胎儿杂合的                 | 220,296 | 24.57  |
| 母亲纯合的; 父亲杂合的          | 131,037 | 14.61  |
| 母亲纯合的; 胎儿杂合的          | 111,775 | 12.46  |
| 母亲纯合的; 父亲杂合的; 胎儿杂合的   | 65,875  | 7.35   |
| 母亲纯合的; 父亲杂合的; 胎儿纯合的   | 65,162  | 7.27   |
| 母亲纯合的; 父亲纯合的, 但等位基因不同 | 45,900  | 5.12   |
| 总计                    | 896,756 | 100    |

图 25A

| 流通池 | 总的原始读取<br>(10 <sup>6</sup> ) | 总数的比对读取<br>(10 <sup>6</sup> ) | 比对率   |
|-----|------------------------------|-------------------------------|-------|
| 1   | 265                          | 200                           | 75.29 |
| 2   | 274                          | 206                           | 75.22 |
| 3   | 268                          | 204                           | 75.96 |
| 4   | 254                          | 193                           | 76.10 |
| 5   | 258                          | 196                           | 76.02 |
| 6   | 280                          | 215                           | 76.80 |
| 7   | 272                          | 207                           | 75.97 |
| 8   | 253                          | 192                           | 75.80 |
| 9   | 260                          | 195                           | 75.17 |
| 10  | 266                          | 208                           | 78.11 |
| 11  | 245                          | 188                           | 76.68 |
| 12  | 249                          | 190                           | 76.22 |
| 13  | 250                          | 189                           | 75.32 |
| 14  | 257                          | 192                           | 74.79 |
| 15  | 250                          | 187                           | 75.14 |
| 16  | 243                          | 187                           | 76.93 |
| 17  | 270                          | 200                           | 74.25 |
| 18  | 258                          | 197                           | 76.48 |
| 19  | 258                          | 193                           | 74.98 |
| 20  | 253                          | 193                           | 76.31 |

图 25B



由20个FC估算的胎儿DNA浓度

| 染色体<br>(Chr.) | 父亲 hetero,<br>母亲 homo,<br>胎儿 hetero.(%) | 父亲 homo,<br>母亲 homo,<br>等位基因不同 |
|---------------|-----------------------------------------|--------------------------------|
| chr1          | 11.83                                   | 11.92                          |
| chr2          | 12.15                                   | 11.99                          |
| chr3          | 11.80                                   | 11.92                          |
| chr4          | 11.89                                   | 11.84                          |
| chr5          | 12.01                                   | 12.10                          |
| chr6          | 12.00                                   | 11.80                          |
| chr7          | 12.14                                   | 11.75                          |
| chr8          | 12.04                                   | 11.95                          |
| chr9          | 11.70                                   | 11.82                          |
| chr10         | 11.57                                   | 11.75                          |
| chr11         | 11.88                                   | 11.90                          |
| chr12         | 12.09                                   | 11.80                          |
| chr13         | 12.39                                   | 11.94                          |
| chr14         | 11.90                                   | 11.84                          |
| chr15         | 11.75                                   | 11.61                          |
| chr16         | 11.93                                   | 11.36                          |
| chr17         | 11.77                                   | 11.49                          |
| chr18         | 11.99                                   | 11.74                          |
| chr19         | 11.77                                   | 12.02                          |
| chr20         | 11.74                                   | 11.56                          |
| chr21         | 11.74                                   | 11.37                          |
| chr22         | 11.81                                   | 11.34                          |
| chrX          |                                         | 12.03                          |

Chr: 染色体  
homo: 纯合的  
hetero: 杂合的

图 26

由胎儿特异性读取覆盖的SNP的比例  
(母亲和父亲均为纯合的,但等位基因不同)

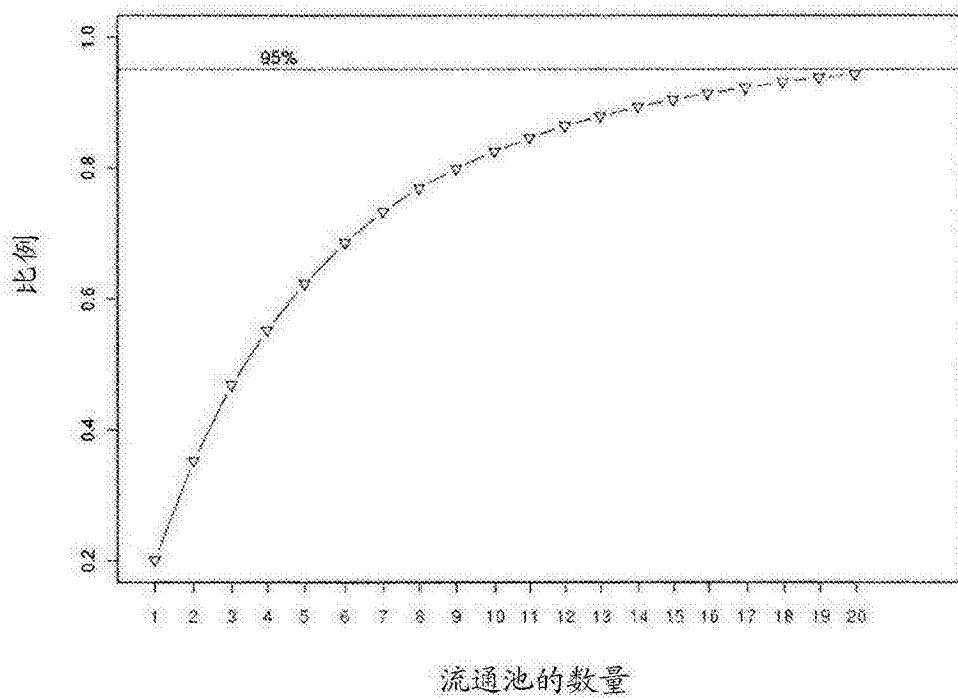


图 27A

由胎儿特异性读取覆盖的SNP的比例  
(母亲和父亲均为纯合的,但等位基因不同)

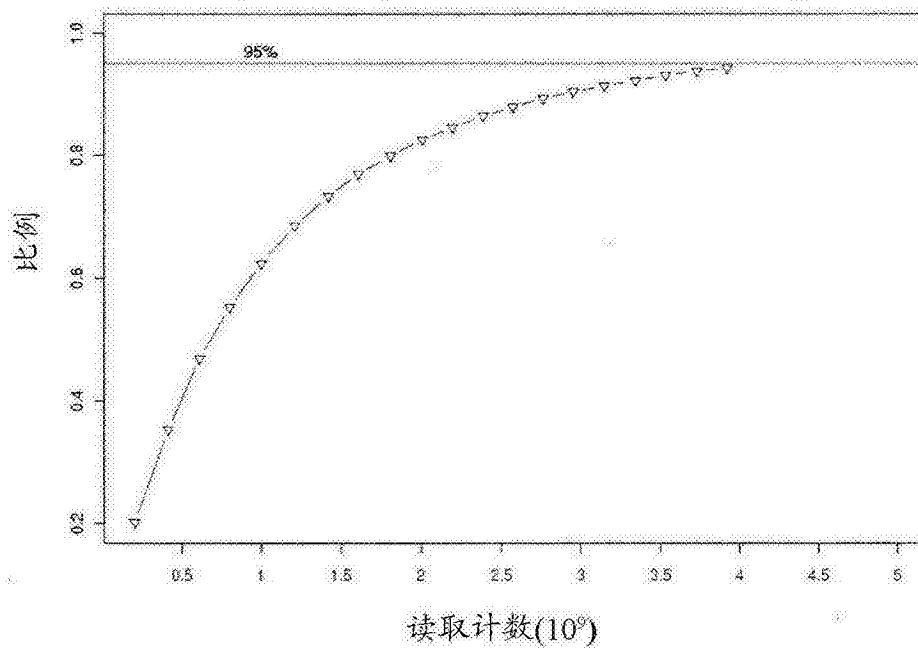


图 27B

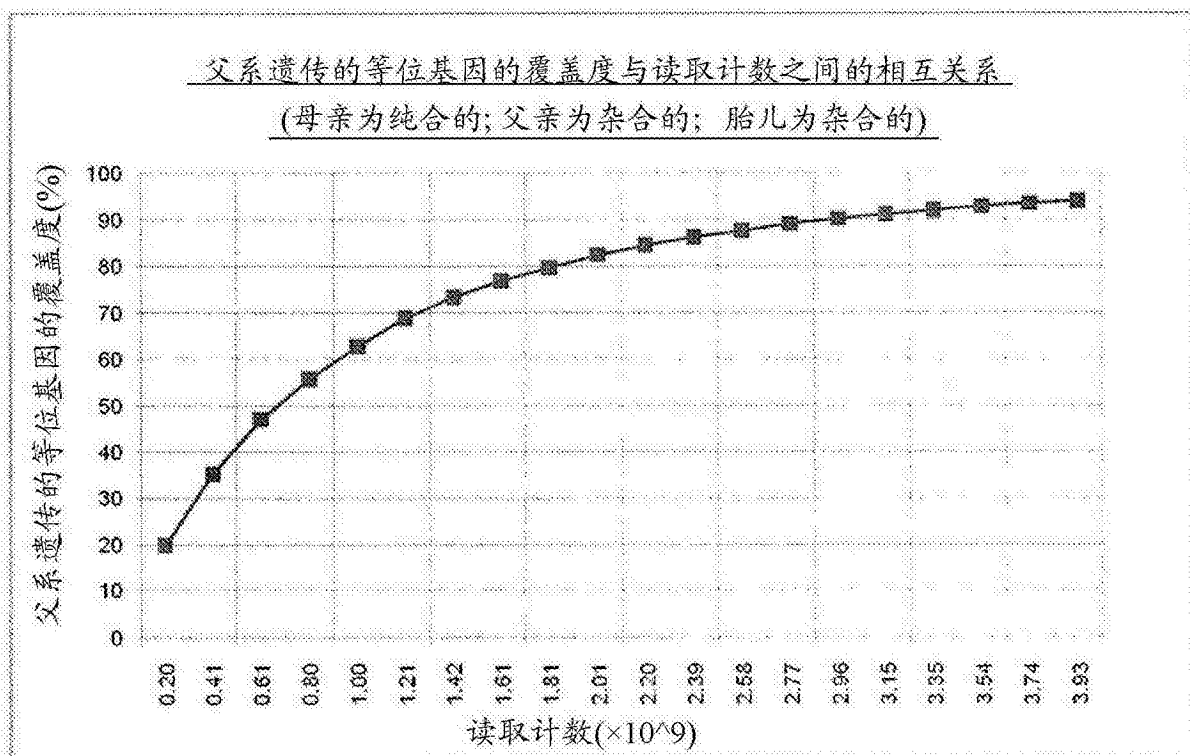


图 28A

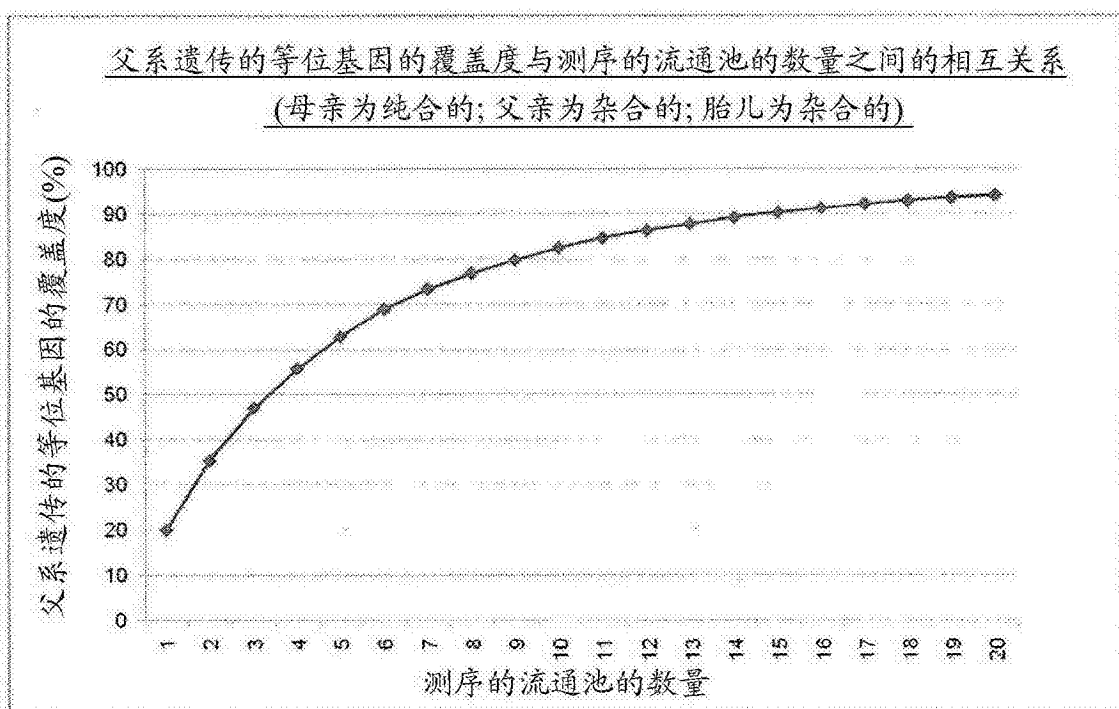


图 28B

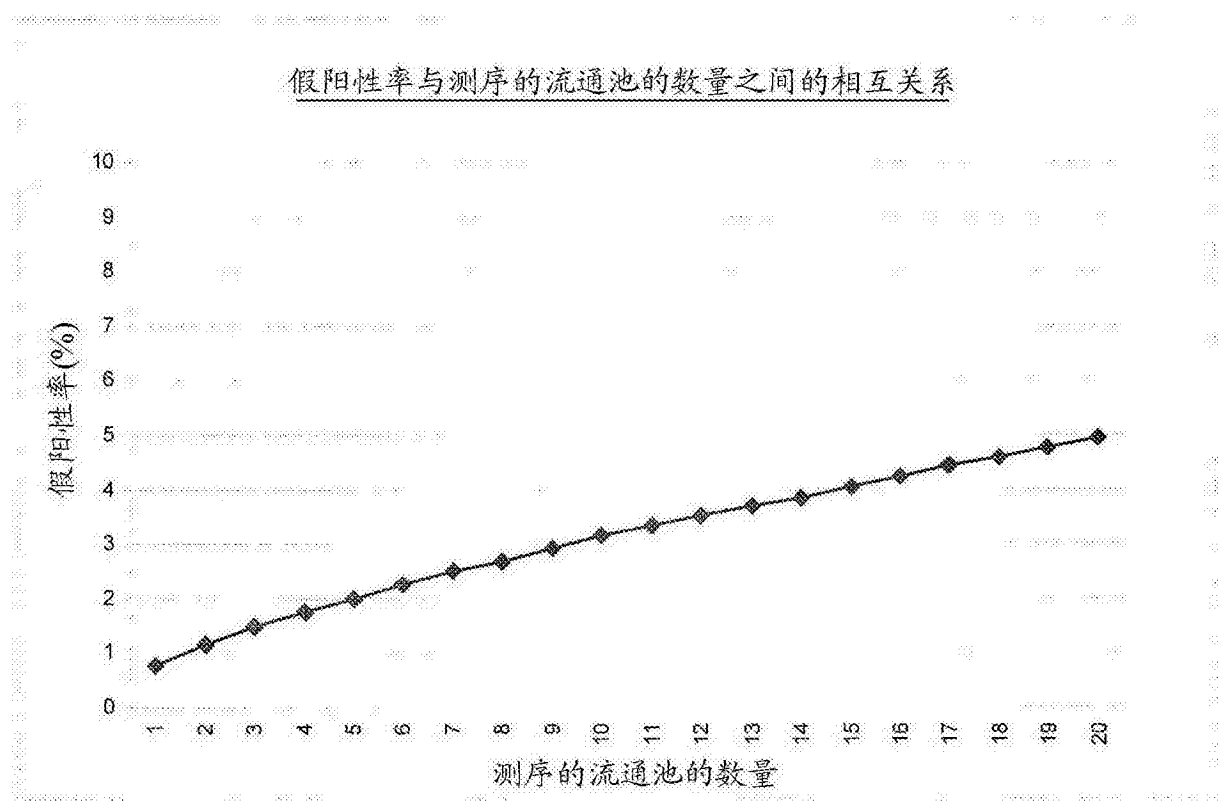


图 29A

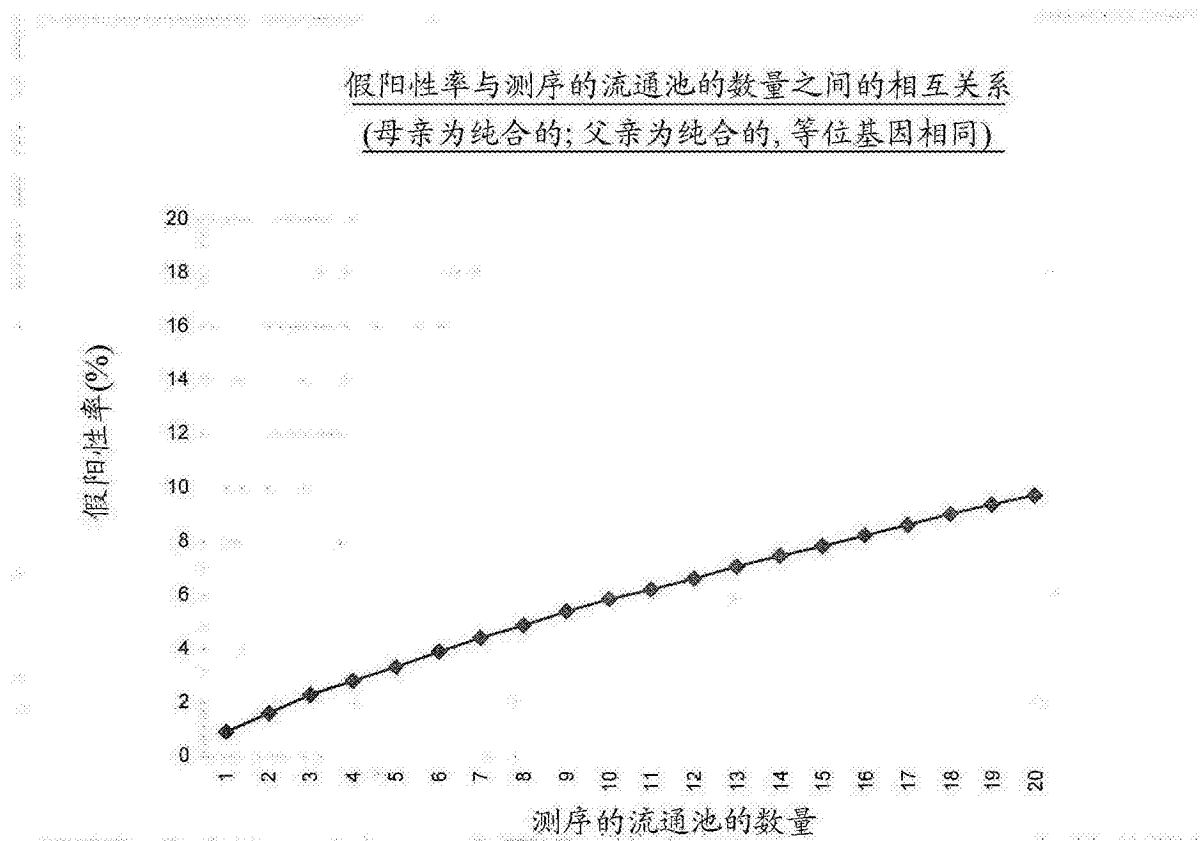


图 29B

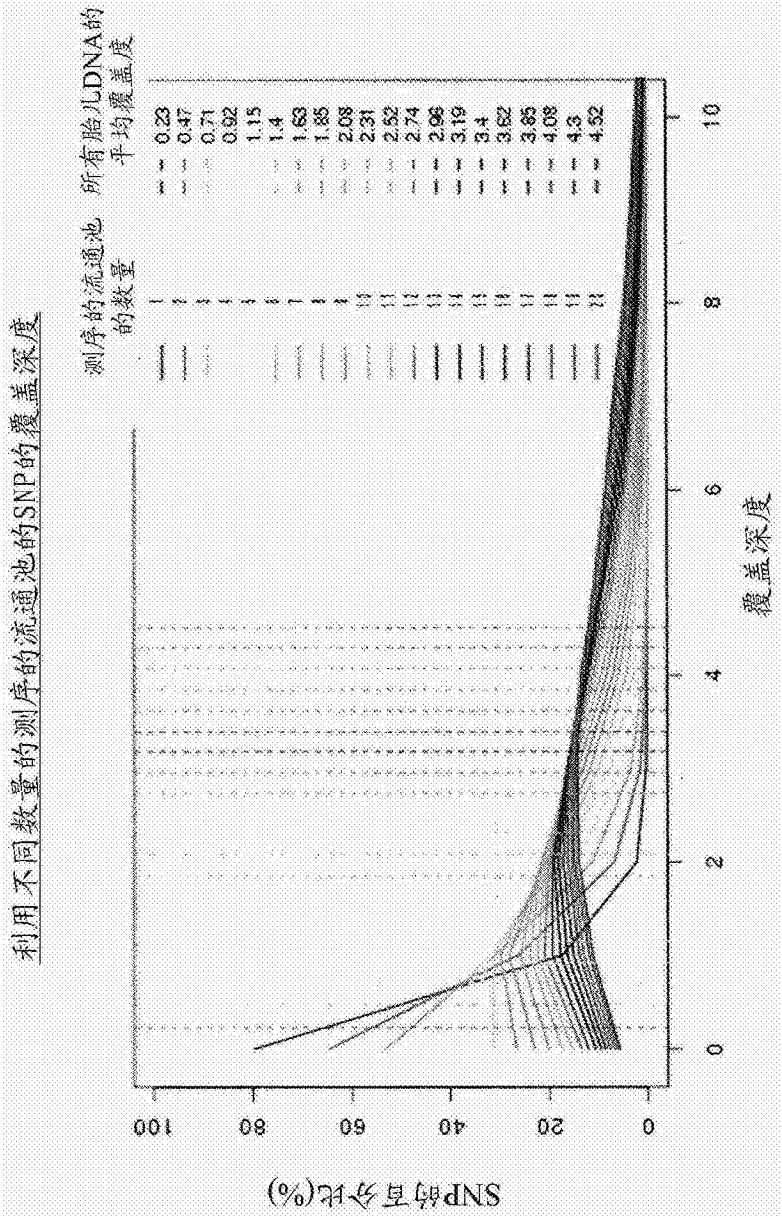


图 30

| A型分类的精度(10个流通池) |       |        |       |      |                 |      |      |      |  |
|-----------------|-------|--------|-------|------|-----------------|------|------|------|--|
| Chr             | 正确的分类 |        | 错误的分类 |      | 两种交错分析<br>均为错误的 |      | 总计   |      |  |
|                 | 数量    | 百分比    | 数量    | 百分比  | 数量              | 百分比  | 数量   | 百分比  |  |
| Chr 1           | 224   | 99.1%  | 2     | 0.9% | 0               | 0.0% | 226  | 0.0% |  |
| Chr 2           | 215   | 96.8%  | 7     | 3.2% | 0               | 0.0% | 222  | 0.0% |  |
| Chr 3           | 176   | 97.2%  | 5     | 2.8% | 0               | 0.0% | 181  | 0.0% |  |
| Chr 4           | 162   | 98.8%  | 2     | 1.2% | 0               | 0.0% | 164  | 0.0% |  |
| Chr 5           | 139   | 99.3%  | 1     | 0.7% | 0               | 0.0% | 140  | 0.0% |  |
| Chr 6           | 170   | 99.4%  | 1     | 0.6% | 0               | 0.0% | 171  | 0.0% |  |
| Chr 7           | 134   | 99.3%  | 1     | 0.7% | 0               | 0.0% | 135  | 0.0% |  |
| Chr 8           | 154   | 98.7%  | 2     | 1.3% | 0               | 0.0% | 156  | 0.0% |  |
| Chr 9           | 134   | 99.3%  | 1     | 0.7% | 0               | 0.0% | 135  | 0.0% |  |
| Chr 10          | 154   | 99.4%  | 1     | 0.6% | 0               | 0.0% | 155  | 0.0% |  |
| Chr 11          | 151   | 98.7%  | 2     | 1.3% | 0               | 0.0% | 153  | 0.0% |  |
| Chr 12          | 136   | 99.3%  | 1     | 0.7% | 0               | 0.0% | 137  | 0.0% |  |
| Chr 13          | 95    | 100.0% | 0     | 0.0% | 0               | 0.0% | 95   | 0.0% |  |
| Chr 14          | 91    | 100.0% | 0     | 0.0% | 0               | 0.0% | 91   | 0.0% |  |
| Chr 15          | 76    | 100.0% | 0     | 0.0% | 0               | 0.0% | 76   | 0.0% |  |
| Chr 16          | 95    | 99.0%  | 1     | 1.0% | 0               | 0.0% | 96   | 0.0% |  |
| Chr 17          | 82    | 100.0% | 0     | 0.0% | 0               | 0.0% | 82   | 0.0% |  |
| Chr 18          | 67    | 100.0% | 0     | 0.0% | 0               | 0.0% | 67   | 0.0% |  |
| Chr 19          | 46    | 100.0% | 0     | 0.0% | 0               | 0.0% | 46   | 0.0% |  |
| Chr 20          | 79    | 97.5%  | 2     | 2.5% | 0               | 0.0% | 81   | 0.0% |  |
| Chr 21          | 45    | 100.0% | 0     | 0.0% | 0               | 0.0% | 45   | 0.0% |  |
| Chr 22          | 32    | 100.0% | 0     | 0.0% | 0               | 0.0% | 32   | 0.0% |  |
| Chr X           | 103   | 96.3%  | 4     | 3.7% | 0               | 0.0% | 107  | 0.0% |  |
| 总计              | 2760  | 98.8%  | 33    | 1.2% | 0               | 0.0% | 2793 | 0.0% |  |

图 31

| B型分类的精度(10个流通池) |       |        |       |      |                 |      |    |      |      |
|-----------------|-------|--------|-------|------|-----------------|------|----|------|------|
| Chr             | 正确的分类 |        | 错误的分类 |      | 两种交错分析<br>均为错误的 |      |    |      |      |
|                 | 数量    | 百分比    | 数量    | 百分比  | 数量              | 百分比  | 数量 | 百分比  | 总计   |
| Chr 1           | 210   | 97.2%  | 6     | 2.8% | 0               | 0.0% | 0  | 0.0% | 216  |
| Chr 2           | 230   | 97.9%  | 5     | 2.1% | 0               | 0.0% | 0  | 0.0% | 235  |
| Chr 3           | 163   | 99.4%  | 1     | 0.6% | 0               | 0.0% | 0  | 0.0% | 164  |
| Chr 4           | 127   | 100.0% | 0     | 0.0% | 0               | 0.0% | 0  | 0.0% | 127  |
| Chr 5           | 182   | 98.9%  | 2     | 1.1% | 0               | 0.0% | 0  | 0.0% | 184  |
| Chr 6           | 155   | 96.9%  | 5     | 3.1% | 0               | 0.0% | 0  | 0.0% | 160  |
| Chr 7           | 115   | 97.5%  | 3     | 2.5% | 0               | 0.0% | 0  | 0.0% | 118  |
| Chr 8           | 160   | 98.8%  | 2     | 1.2% | 0               | 0.0% | 0  | 0.0% | 162  |
| Chr 9           | 123   | 98.4%  | 2     | 1.6% | 0               | 0.0% | 0  | 0.0% | 125  |
| Chr 10          | 130   | 97.0%  | 4     | 3.0% | 0               | 0.0% | 0  | 0.0% | 134  |
| Chr 11          | 129   | 97.7%  | 3     | 2.3% | 0               | 0.0% | 0  | 0.0% | 132  |
| Chr 12          | 118   | 100.0% | 0     | 0.0% | 0               | 0.0% | 0  | 0.0% | 118  |
| Chr 13          | 90    | 97.8%  | 2     | 2.2% | 0               | 0.0% | 0  | 0.0% | 92   |
| Chr 14          | 83    | 97.6%  | 2     | 2.4% | 0               | 0.0% | 0  | 0.0% | 85   |
| Chr 15          | 75    | 98.7%  | 1     | 1.3% | 0               | 0.0% | 0  | 0.0% | 76   |
| Chr 16          | 80    | 94.1%  | 5     | 5.9% | 0               | 0.0% | 0  | 0.0% | 85   |
| Chr 17          | 50    | 100.0% | 0     | 0.0% | 0               | 0.0% | 0  | 0.0% | 50   |
| Chr 18          | 71    | 97.3%  | 2     | 2.7% | 0               | 0.0% | 0  | 0.0% | 73   |
| Chr 19          | 48    | 100.0% | 0     | 0.0% | 0               | 0.0% | 0  | 0.0% | 48   |
| Chr 20          | 80    | 100.0% | 0     | 0.0% | 0               | 0.0% | 0  | 0.0% | 80   |
| Chr 21          | 27    | 100.0% | 0     | 0.0% | 0               | 0.0% | 0  | 0.0% | 27   |
| Chr 22          | 44    | 97.8%  | 1     | 2.2% | 0               | 0.0% | 0  | 0.0% | 45   |
| Chr X           | 118   | 97.5%  | 3     | 2.5% | 0               | 0.0% | 0  | 0.0% | 121  |
| 总计              | 2608  | 98.2%  | 49    | 1.8% | 0               | 0.0% | 0  | 0.0% | 2657 |

图 32



A型SPRT分析的精度

| chr | 单域分类         |           | 总计   | 二连续域算法       |          | 总计   |
|-----|--------------|-----------|------|--------------|----------|------|
|     | 正确的(%)       | 错误的(%)    |      | 正确的(%)       | 错误的(%)   |      |
| 1   | 314 (99.1%)  | 3 (0.9%)  | 317  | 309 (100%)   | 0 (0.0%) | 309  |
| 2   | 305 (99.3%)  | 2 (0.7%)  | 307  | 301 (100%)   | 0 (0.0%) | 301  |
| 3   | 240 (99.6%)  | 1 (0.4%)  | 241  | 237 (100%)   | 0 (0.0%) | 237  |
| 4   | 218 (99.5%)  | 1 (0.5%)  | 219  | 215 (100%)   | 0 (0.0%) | 215  |
| 5   | 203 (100%)   | 0 (0.0%)  | 203  | 201 (100%)   | 0 (0.0%) | 201  |
| 6   | 240 (98.0%)  | 5 (2.0%)  | 245  | 237 (98.8%)  | 3 (1.3%) | 240  |
| 7   | 182 (98.9%)  | 2 (1.1%)  | 184  | 179 (100%)   | 0 (0.0%) | 179  |
| 8   | 217 (99.5%)  | 1 (0.5%)  | 218  | 214 (100%)   | 0 (0.0%) | 214  |
| 9   | 177 (99.4%)  | 1 (0.6%)  | 178  | 174 (100%)   | 0 (0.0%) | 174  |
| 10  | 217 (99.1%)  | 2 (0.9%)  | 219  | 213 (100%)   | 0 (0.0%) | 213  |
| 11  | 205 (100%)   | 0 (0.0%)  | 205  | 203 (100%)   | 0 (0.0%) | 203  |
| 12  | 193 (99.5%)  | 1 (0.5%)  | 194  | 190 (100%)   | 0 (0.0%) | 190  |
| 13  | 133 (100%)   | 0 (0.0%)  | 133  | 132 (100%)   | 0 (0.0%) | 132  |
| 14  | 125 (100%)   | 0 (0.0%)  | 125  | 124 (100%)   | 0 (0.0%) | 124  |
| 15  | 103 (100%)   | 0 (0.0%)  | 103  | 102 (100%)   | 0 (0.0%) | 102  |
| 16  | 133 (98.5%)  | 2 (1.5%)  | 135  | 129 (100%)   | 0 (0.0%) | 129  |
| 17  | 104 (100%)   | 0 (0.0%)  | 104  | 102 (100%)   | 0 (0.0%) | 102  |
| 18  | 95 (97.9%)   | 2 (2.1%)  | 97   | 91 (100%)    | 0 (0.0%) | 91   |
| 19  | 67 (100%)    | 0 (0.0%)  | 67   | 65 (100%)    | 0 (0.0%) | 65   |
| 20  | 112 (100%)   | 0 (0.0%)  | 112  | 110 (100%)   | 0 (0.0%) | 110  |
| 21  | 60 (100%)    | 0 (0.0%)  | 60   | 59 (100%)    | 0 (0.0%) | 59   |
| 22  | 39 (100%)    | 0 (0.0%)  | 39   | 38 (100%)    | 0 (0.0%) | 38   |
| X   | 156 (98.7%)  | 2 (1.3%)  | 158  | 152 (100%)   | 0 (0.0%) | 152  |
| 总计  | 3838 (99.4%) | 25 (0.6%) | 3863 | 3777 (99.9%) | 3 (0.1%) | 3780 |

图 33

B型SPRT分析的精度

| chr | 单域分类         |           |      | 二连续域算法       |          |      |
|-----|--------------|-----------|------|--------------|----------|------|
|     | 正确的(%)       | 错误的(%)    | 总计   | 正确的(%)       | 错误的(%)   | 总计   |
| 1   | 267 (98.2%)  | 5 (1.8%)  | 272  | 262 (100%)   | 0 (0.0%) | 262  |
| 2   | 283 (99.0%)  | 3 (1.0%)  | 286  | 278 (100%)   | 0 (0.0%) | 278  |
| 3   | 222 (98.2%)  | 4 (1.8%)  | 226  | 216 (100%)   | 0 (0.0%) | 216  |
| 4   | 175 (98.9%)  | 2 (1.1%)  | 177  | 171 (100%)   | 0 (0.0%) | 171  |
| 5   | 255 (99.6%)  | 1 (0.4%)  | 256  | 252 (100%)   | 0 (0.0%) | 252  |
| 6   | 195 (99.0%)  | 2 (1.0%)  | 197  | 192 (100%)   | 0 (0.0%) | 192  |
| 7   | 155 (97.5%)  | 4 (2.5%)  | 159  | 150 (100%)   | 0 (0.0%) | 150  |
| 8   | 211 (100%)   | 0 (0.0%)  | 211  | 209 (100%)   | 0 (0.0%) | 209  |
| 9   | 162 (99.4%)  | 1 (0.6%)  | 163  | 160 (100%)   | 0 (0.0%) | 160  |
| 10  | 172 (96.6%)  | 6 (3.4%)  | 178  | 167 (98.2%)  | 3 (1.8%) | 170  |
| 11  | 171 (99.4%)  | 1 (0.6%)  | 172  | 168 (100%)   | 0 (0.0%) | 168  |
| 12  | 156 (99.4%)  | 1 (0.6%)  | 157  | 153 (100%)   | 0 (0.0%) | 153  |
| 13  | 119 (98.3%)  | 2 (1.7%)  | 121  | 117 (99.2%)  | 1 (0.8%) | 118  |
| 14  | 116 (98.3%)  | 2 (1.7%)  | 118  | 115 (99.1%)  | 1 (0.9%) | 116  |
| 15  | 101 (100%)   | 0 (0.0%)  | 101  | 100 (100%)   | 0 (0.0%) | 100  |
| 16  | 115 (94.3%)  | 7 (5.7%)  | 122  | 107 (99.1%)  | 1 (0.9%) | 108  |
| 17  | 67 (98.5%)   | 1 (1.5%)  | 68   | 65 (100%)    | 0 (0.0%) | 65   |
| 18  | 92 (100%)    | 0 (0.0%)  | 92   | 90 (100%)    | 0 (0.0%) | 90   |
| 19  | 57 (100%)    | 0 (0.0%)  | 57   | 55 (100%)    | 0 (0.0%) | 55   |
| 20  | 95 (100%)    | 0 (0.0%)  | 95   | 93 (100%)    | 0 (0.0%) | 93   |
| 21  | 39 (100%)    | 0 (0.0%)  | 39   | 38 (100%)    | 0 (0.0%) | 38   |
| 22  | 59 (100%)    | 0 (0.0%)  | 59   | 58 (100%)    | 0 (0.0%) | 58   |
| X   | 142 (99.3%)  | 1 (0.7%)  | 143  | 139 (100%)   | 0 (0.0%) | 139  |
| 总计  | 3426 (98.8%) | 43 (1.2%) | 3469 | 3355 (99.8%) | 6 (0.2%) | 3361 |

图 34



对11号染色体上的β-珠蛋白簇进行A型分析

| 基因型    |            |         |    |    |    |                    |        |              |             |
|--------|------------|---------|----|----|----|--------------------|--------|--------------|-------------|
| Chr    | dbSNP标识    | 位置      | 父亲 | 母亲 | 胎儿 | Hap III            | Hap IV | HapIII<br>计数 | HapIV<br>计数 |
| chr 11 |            | 5204905 |    |    |    | β地中海贫血-珠蛋白nt-28位突变 |        |              |             |
| chr11  | rs10742583 | 5205217 | TT | CT | TT | T                  | C      | 22           | 26          |
| chr11  | rs10742584 | 5205346 | TT | CT | TT | T                  | C      | 43           | 38          |
| chr11  | rs1003586  | 5205946 | GG | AG | GG | G                  | A      | 67           | 49          |
| chr11  | rs6578588  | 5208827 | GG | AG | GG | G                  | A      | 110          | 87          |
| chr11  | rs3759076  | 5213765 | CC | CT | CC | C                  | T      | 139          | 87          |
| chr11  | rs10837643 | 5214614 | AA | AT | AA | A                  | T      | 162          | 98          |
| chr11  | rs4283007  | 5215066 | TT | AT | TT | T                  | A      | 215          | 134         |
|        |            |         |    |    |    |                    |        |              | Hap III     |
|        |            |         |    |    |    |                    |        |              | 未分类的        |
|        |            |         |    |    |    |                    |        |              | 未分类的        |
|        |            |         |    |    |    |                    |        |              | 未分类的        |
|        |            |         |    |    |    |                    |        |              | 未分类的        |
|        |            |         |    |    |    |                    |        |              | 未分类的        |
|        |            |         |    |    |    |                    |        |              | 未分类的        |

图 36

对11号染色体上的β-珠蛋白簇进行B型分析

| Chr                | dbSNP标识    | 位置      | 父亲 | 母亲 | 胎儿 | Hap III | Hap IV | Hap III 计数 | Hap IV 计数 | Hap III 累积计数 | Hap IV 累积计数 | 分类      |
|--------------------|------------|---------|----|----|----|---------|--------|------------|-----------|--------------|-------------|---------|
| β地中海贫血-珠蛋白nt-28位突变 |            |         |    |    |    |         |        |            |           |              |             |         |
| chr11              | rs12275403 | 5286913 | CC | CG | CG | G       | C      | 19         | 23        | 19           | 23          | 未分类的    |
| chr11              | rs7926089  | 5289019 | CC | CT | CT | T       | C      | 13         | 21        | 32           | 44          | 未分类的    |
| chr11              | rs12365420 | 5289330 | GG | GT | GT | T       | G      | 5          | 12        | 37           | 56          | 未分类的    |
| chr11              | rs4910747  | 5294746 | CC | AC | AC | A       | C      | 27         | 36        | 64           | 92          | 未分类的    |
| chr11              | rs6578604  | 5299185 | AA | AG | AG | G       | A      | 15         | 18        | 79           | 110         | 未分类的    |
| chr11              | rs10837814 | 5301704 | GG | AG | AG | A       | G      | 39         | 45        | 118          | 155         | 未分类的    |
| chr11              | rs4910550  | 5302436 | GG | AG | AG | A       | G      | 15         | 22        | 133          | 177         | 未分类的    |
| chr11              | rs2723381  | 5304451 | TT | CT | CT | C       | T      | 38         | 34        | 171          | 211         | 未分类的    |
| chr11              | rs6578606  | 5306671 | GG | CG | CG | C       | G      | 40         | 31        | 211          | 242         | 未分类的    |
| chr11              | rs3824950  | 5588548 | CC | CT | CT | T       | C      | 59         | 66        | 270          | 308         | 未分类的    |
| chr11              | rs10838441 | 5601579 | GG | CG | CG | C       | G      | 20         | 10        | 290          | 318         | 未分类的    |
| chr11              | rs11038628 | 5645516 | CC | CT | CT | T       | C      | 65         | 5         | 355          | 323         | Hap III |

图 37

| SNP的<br>基因组<br>坐标 | Hap I、Hap II 上的<br>母本等位基因 |   | Hap I、Hap II 上的<br>等位基因计数 |    | Hap I、Hap II 上的<br>等位基因的累积计数 |     | SPRT分类 |
|-------------------|---------------------------|---|---------------------------|----|------------------------------|-----|--------|
| 3251050           | A                         | G | 22                        | 15 | 22                           | 15  | 未分类的   |
| 3251284           | G                         | A | 14                        | 10 | 36                           | 25  | 未分类的   |
| 3251791           | A                         | G | 18                        | 11 | 54                           | 36  | Hap I  |
| 9865342           | A                         | G | 5                         | 3  | 5                            | 3   | 未分类的   |
| 9872792           | A                         | G | 7                         | 5  | 12                           | 8   | 未分类的   |
| 9874947           | G                         | C | 1                         | 3  | 13                           | 11  | 未分类的   |
| 15242479          | G                         | A | 17                        | 5  | 30                           | 16  | Hap I  |
| 15710521          | G                         | A | 1                         | 2  | 1                            | 2   | 未分类的   |
| 15710672          | A                         | G | 13                        | 13 | 14                           | 15  | 未分类的   |
| 21895385          | C                         | T | 15                        | 3  | 29                           | 18  | 未分类的   |
| 22149641          | T                         | C | 5                         | 3  | 34                           | 21  | Hap I  |
| 22201527          | A                         | C | 11                        | 10 | 11                           | 10  | 未分类的   |
| 23806924          | C                         | T | 20                        | 9  | 31                           | 19  | 未分类的   |
| 31134605          | A                         | G | 2                         | 1  | 33                           | 20  | Hap I  |
| 32413115          | C                         | T | 6                         | 8  | 6                            | 8   | 未分类的   |
| 32473184          | G                         | A | 8                         | 9  | 14                           | 17  | 未分类的   |
| 35730716          | A                         | G | 17                        | 14 | 31                           | 31  | 未分类的   |
| 35903375          | C                         | T | 7                         | 9  | 38                           | 40  | 未分类的   |
| 35903407          | G                         | A | 6                         | 2  | 44                           | 42  | 未分类的   |
| 35917535          | T                         | G | 6                         | 7  | 50                           | 49  | 未分类的   |
| 43475980          | T                         | G | 7                         | 16 | 57                           | 65  | 未分类的   |
| 47351305          | A                         | G | 7                         | 6  | 64                           | 71  | 未分类的   |
| 48561783          | A                         | G | 5                         | 1  | 69                           | 72  | 未分类的   |
| 50676020          | C                         | T | 19                        | 15 | 88                           | 87  | 未分类的   |
| 53601143          | A                         | G | 8                         | 10 | 96                           | 97  | 未分类的   |
| 54854802          | T                         | C | 9                         | 3  | 105                          | 100 | 未分类的   |
| 74921254          | C                         | A | 3                         | 3  | 108                          | 103 | 未分类的   |
| 78102680          | A                         | C | 5                         | 3  | 113                          | 106 | 未分类的   |
| 78313644          | C                         | T | 40                        | 24 | 153                          | 130 | Hap I  |

图 38A

| SNP的<br>基因组<br>坐标 | Hap I、Hap II上的<br>母本等位基因 |   | Hap I、Hap II上的<br>等位基因计数 |    | Hap I、Hap II上的<br>等位基因的累积计数 |     | SPRT分类 |
|-------------------|--------------------------|---|--------------------------|----|-----------------------------|-----|--------|
| 84236666          | A                        | G | 3                        | 0  | 3                           | 0   | 未分类的   |
| 84249796          | A                        | G | 29                       | 12 | 32                          | 12  | Hap I  |
| 100273093         | T                        | C | 16                       | 10 | 16                          | 10  | 未分类的   |
| 100494847         | A                        | G | 13                       | 5  | 29                          | 15  | Hap I  |
| 106283058         | G                        | A | 5                        | 8  | 5                           | 8   | 未分类的   |
| 106283111         | G                        | A | 13                       | 12 | 18                          | 20  | 未分类的   |
| 107283614         | G                        | A | 10                       | 4  | 28                          | 24  | 未分类的   |
| 107335295         | T                        | G | 5                        | 1  | 33                          | 25  | 未分类的   |
| 114785655         | C                        | T | 11                       | 2  | 44                          | 32  | 未分类的   |
| 112584169         | G                        | A | 8                        | 2  | 52                          | 34  | Hap I  |
| 118471031         | T                        | C | 3                        | 5  | 3                           | 5   | 未分类的   |
| 118488464         | C                        | T | 5                        | 15 | 8                           | 20  | 未分类的   |
| 119474561         | A                        | T | 5                        | 4  | 15                          | 24  | 未分类的   |
| 123022716         | C                        | A | 17                       | 12 | 30                          | 36  | 未分类的   |
| 128506711         | C                        | G | 4                        | 3  | 34                          | 39  | 未分类的   |
| 130047404         | C                        | T | 4                        | 1  | 38                          | 40  | 未分类的   |
| 130243499         | C                        | T | 13                       | 19 | 51                          | 59  | 未分类的   |
| 134814366         | C                        | T | 12                       | 8  | 63                          | 67  | 未分类的   |
| 134818744         | G                        | A | 22                       | 18 | 90                          | 85  | 未分类的   |
| 134821671         | C                        | T | 5                        | 9  | 95                          | 94  | 未分类的   |
| 138507378         | A                        | T | 7                        | 3  | 102                         | 97  | 未分类的   |
| 138695073         | A                        | G | 10                       | 3  | 112                         | 100 | 未分类的   |
| 144713436         | C                        | T | 10                       | 16 | 122                         | 116 | 未分类的   |
| 149577161         | A                        | G | 11                       | 5  | 133                         | 121 | 未分类的   |
| 150888835         | C                        | A | 11                       | 11 | 144                         | 132 | 未分类的   |
| 152381349         | T                        | C | 8                        | 3  | 152                         | 135 | Hap I  |
| 152694738         | C                        | T | 1                        | 0  | 1                           | 0   | 未分类的   |
| 152848587         | G                        | A | 4                        | 6  | 5                           | 6   | 未分类的   |

图 38B

| 病例    | 信息性SNP的<br>总数 | 每个SNP的<br>平均测序深度<br>(倍数) | 部分胎儿DNA<br>浓度(%) | SPRT分类的<br>总数 | 正确分类的<br>数量 | 错误分类的<br>数量 |
|-------|---------------|--------------------------|------------------|---------------|-------------|-------------|
| PW226 | 57            | 17.6                     | 19.4             | 9             | 9           | 0           |
| PW263 | 62            | 15.8                     | 19.4             | 4             | 4           | 0           |
| PW316 | 50            | 18.5                     | 17.4             | 2             | 2           | 0           |
| PW370 | 54            | 18.9                     | 11.5             | 1             | 1           | 0           |
| PW421 | 46            | 19.6                     | 15.3             | 2             | 2           | 0           |

图 39

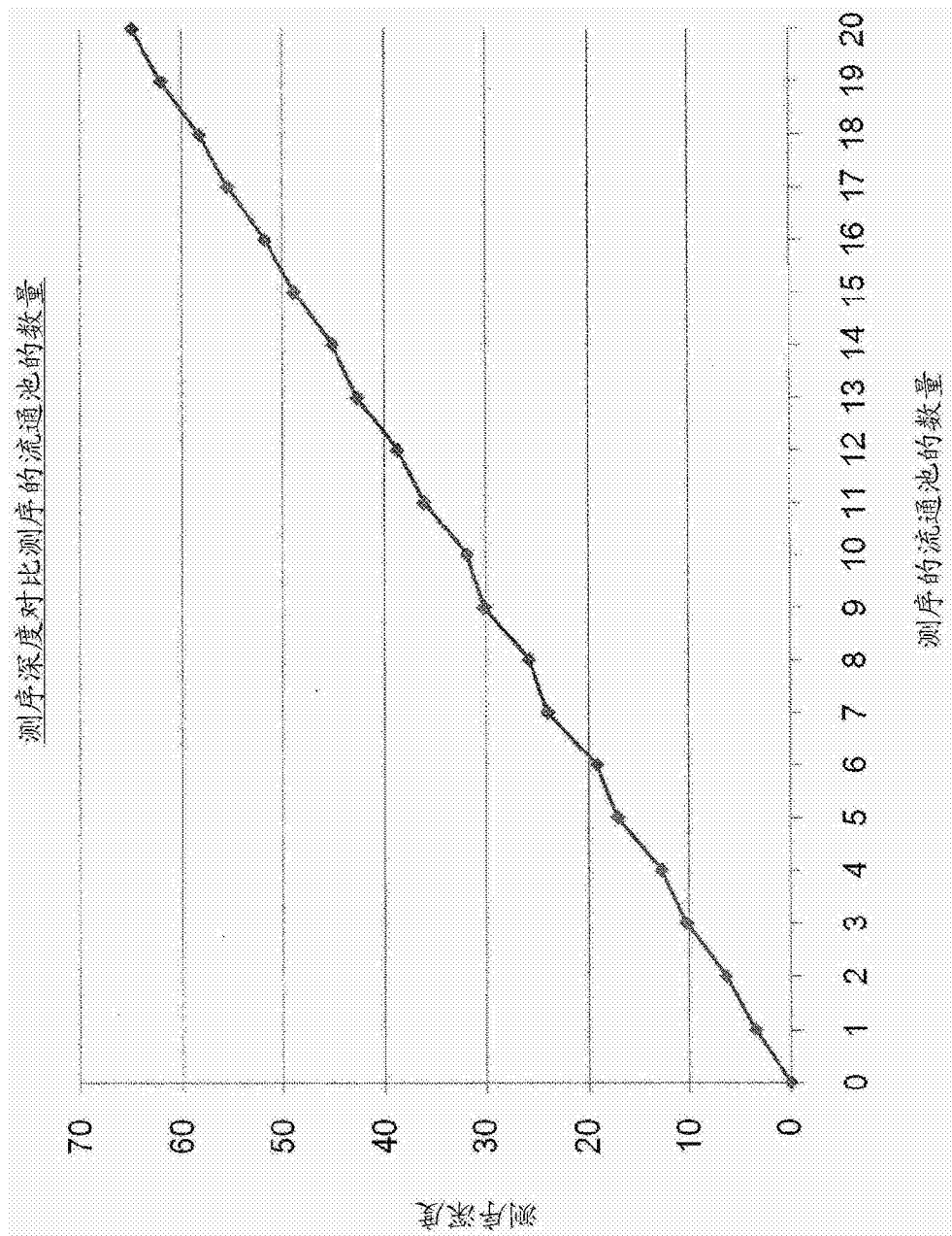


图 40



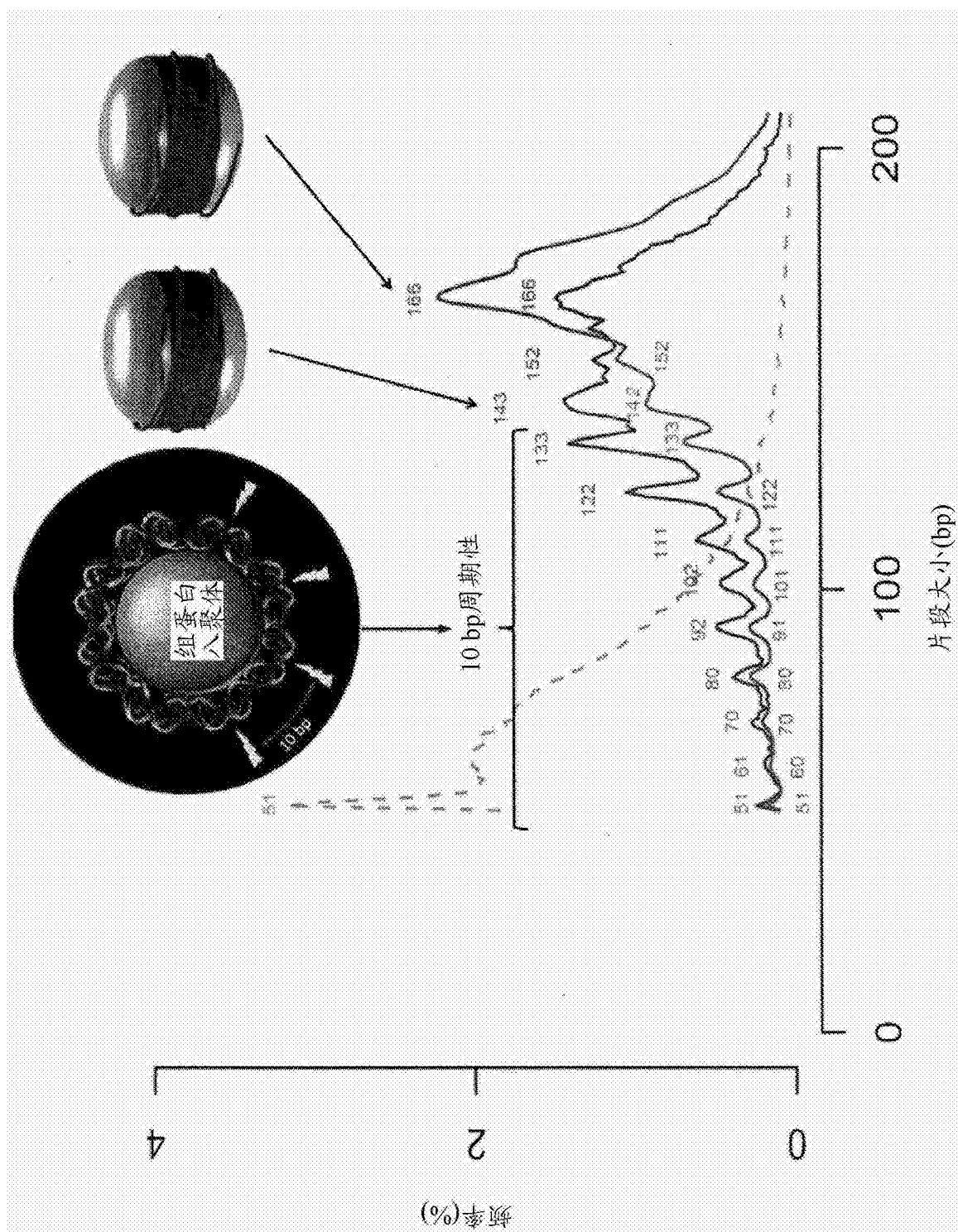


图 41

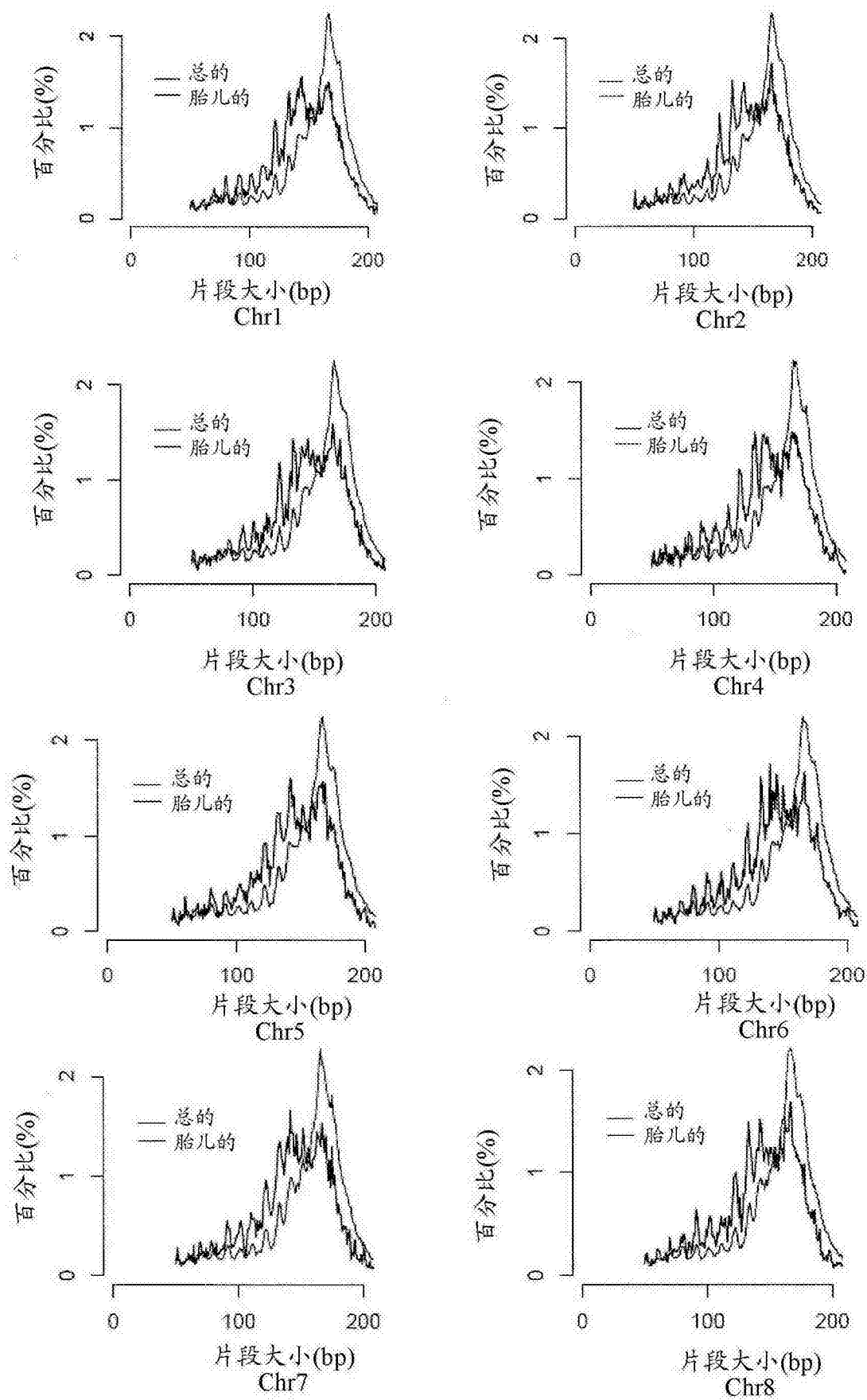


图 42A

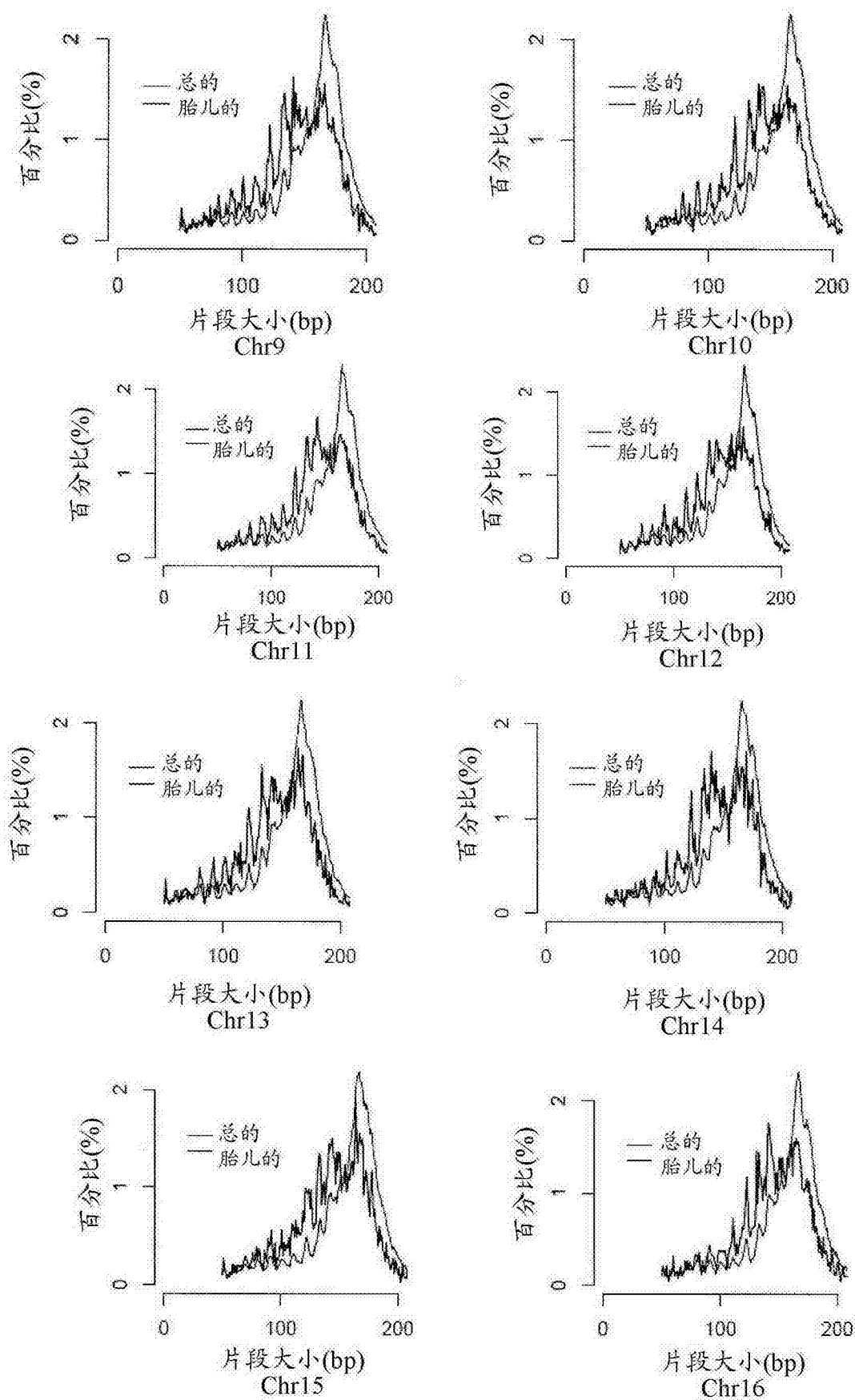


图 42B

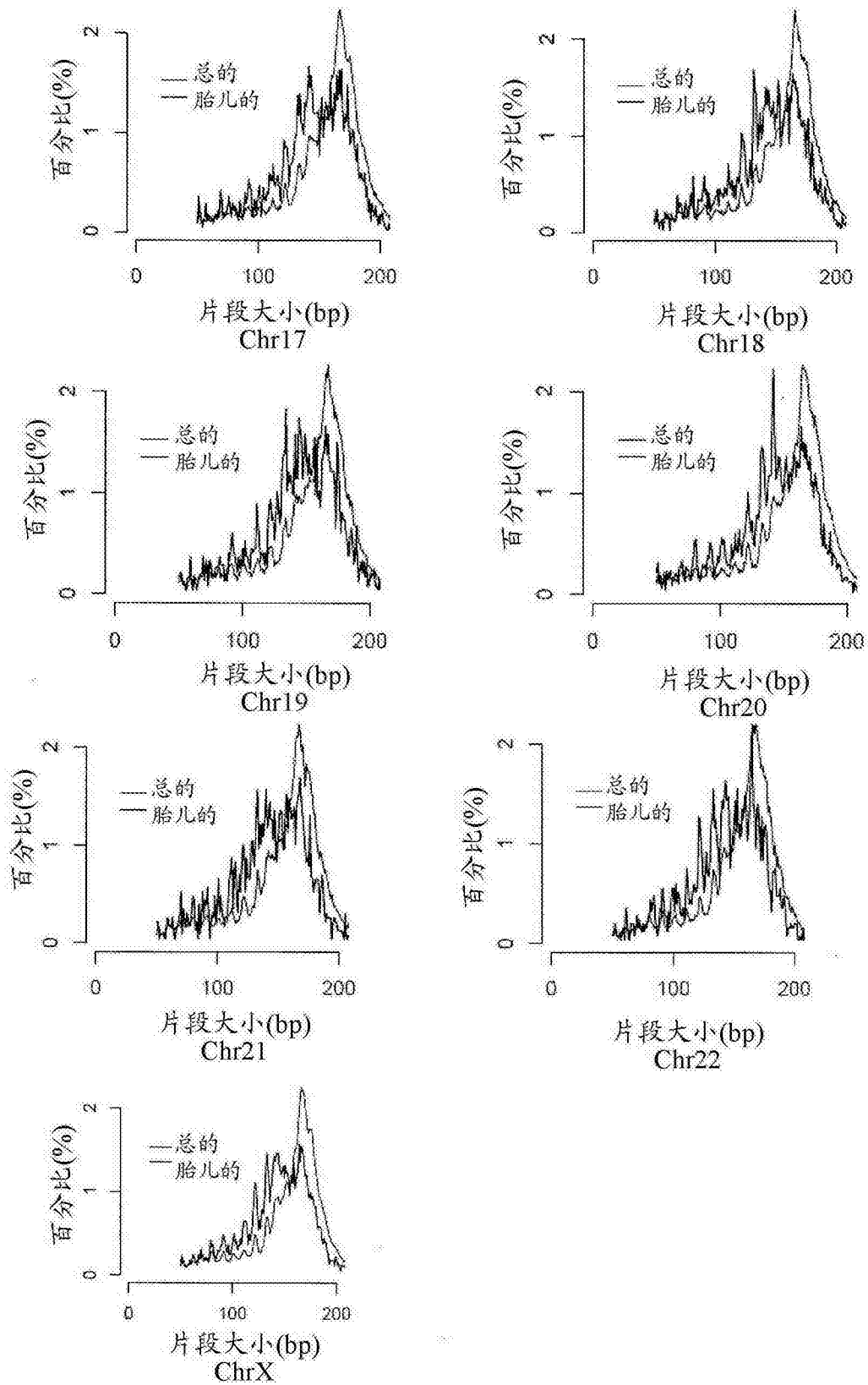


图 42C

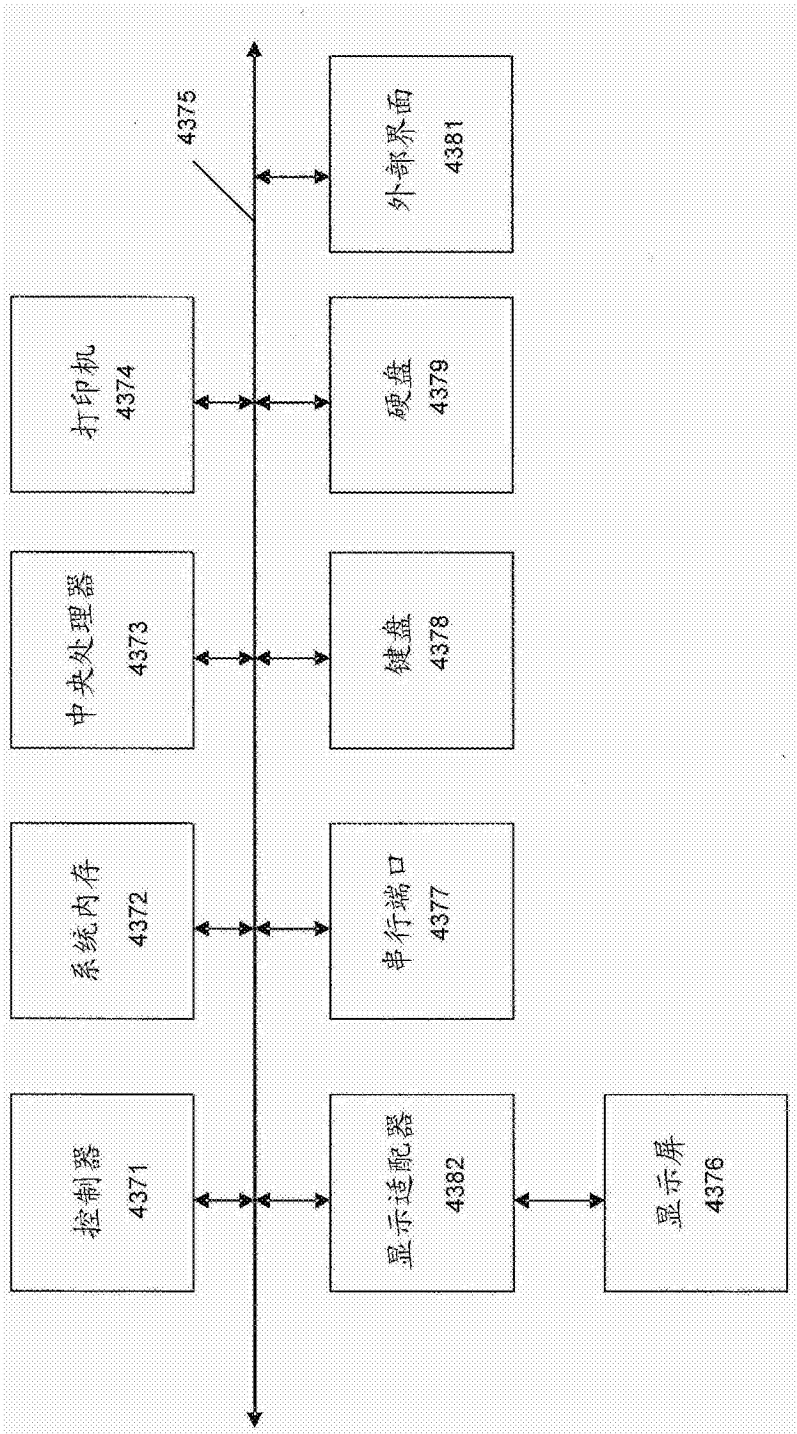


图 43