



(12) 发明专利

(10) 授权公告号 CN 103492589 B

(45) 授权公告日 2016. 06. 15

(21) 申请号 201280016005. 2

(22) 申请日 2012. 02. 24

(30) 优先权数据

61/446, 256 2011. 02. 24 US

(85) PCT国际申请进入国家阶段日

2013. 09. 27

(86) PCT国际申请的申请数据

PCT/IB2012/000344 2012. 02. 24

(87) PCT国际申请的公布数据

W02013/041921 EN 2013. 03. 28

(73) 专利权人 香港中文大学

地址 中国香港新界

(72) 发明人 卢煜明 赵慧君 陈君赐 梁德杨

江培勇

(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 2010/075459 A1, 2010. 07. 01,

PERTL B ET AL. Detection of male and female fetal DNA in maternal plasma by multiplex Euorescent polymerase chain reaction amptification of short tandem repeats. 《HUMAN GENETICS》. 2000, 第 106 卷 (第 1 期), 45-49.

TUNGWIWAT W ET AL. Non-invasive fetal sex determination using a conventional nested PCR analysis of fetal DNA in maternal plasma. 《CLINICA CHIMICA ACTA》. 2003, 第 334 卷 (第 1-2 期), 173-177.

审查员 朱宁

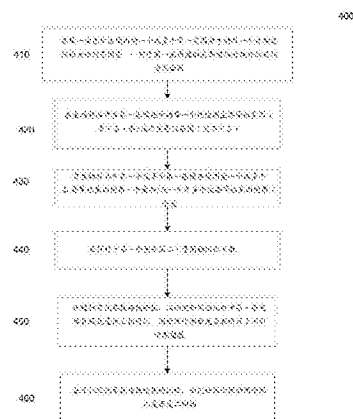
权利要求书13页 说明书33页 附图21页

(54) 发明名称

多胎妊娠的分子测试

(57) 摘要

本发明提供了用于使用取自母亲的生物样品测定多胎妊娠的接合性的方法、系统和仪器。可对于特定染色体区域分析样品(例如血浆)中的胎儿和母体 DNA, 以鉴定胎儿中的基因差异。例如, 当胎儿是二卵的时候, 用于测量主要或次要等位基因的标准化参数对于不同染色体区域可显示出变化。如果胎儿是遗传上相同的, 则此类变化可相对于期望值进行测定。提供统计方法用于分析标准化参数的变化, 以测定胎儿 DNA 浓度和在多个基因座处的母系-胎儿混合基因型。亲本基因型和单元型信息也可用于鉴定不同亲本单元型的继承, 以指示在胎儿中的遗传差异。



1. 一种用于分析具有多个胎儿的妊娠女性的生物样品以测定妊娠女性的至少两个胎儿是否是二卵的方法,所述生物样品包含胎儿和母体DNA的DNA片段,其中所述生物样品是血浆或血清,所述方法包括:

测定第一染色体区域内的一个或多个第一基因座中的每一基因座处的所述妊娠女性的基因型,所述妊娠女性在所述一个或多个第一基因座中的每一基因座处是纯合的,或在所述一个或多个第一基因座中的每一基因座处是杂合的,其中所述第一基因座中的每一个显示出所述生物样品中各自的主要等位基因和各自的次要等位基因,其中对于所述第一基因座中的每一个,所述各自的主要等位基因比所述各自的次要等位基因更丰富;

在所述生物样品的DNA片段中在所述一个或多个第一基因座处测量一个或多个主要等位基因的第一个量和/或一个或多个次要等位基因的第二个量;

获得所述第一个量或所述第二个量的标准化参数;

计算对于一个或多个其他染色体区域的一个或多个另外的标准化参数,

如果所述胎儿对于所述第一染色体区域是遗传上相同的,比较所述标准化参数与截断值,以测定所述标准化参数是否在统计上不同于期望值,所述期望值得自所述生物样品的测量并且对应于所述一个或多个另外的标准化参数;

其中比较所述标准化参数与截断值,以测定所述标准化参数是否在统计上不同于期望值通过以下方式完成:

计算所述标准化参数与所述一个或多个另外的标准化参数之间的差异;并且

将所述差异与截断值比较;和

当所述差异高于所述截断值时,确定所述妊娠女性的至少两个胎儿是二卵的。

2. 根据权利要求1所述的方法,其还包括:

在所述生物样品中在所述第一基因座中的每一基因座处检测所述各自的主要等位基因和所述各自的次要等位基因。

3. 根据权利要求1所述的方法,其中获得所述标准化参数包括:

测量来自不同染色体区域内的一个或多个基因座的一种或多种序列的第三个量;和
使用所述第三个量和所述第一个量或所述第二个量计算所述标准化参数。

4. 根据权利要求1所述的方法,其中获得所述标准化参数包括:

在所述生物样品中在所述一个或多个第一基因座处测量所述一个或多个主要等位基因的所述第一个量和所述一个或多个次要等位基因的所述第二个量;

测定来自所述第一个量和所述第二个量的第一参数,所述第一参数提供在所述第一个量和所述第二个量之间的相对量;和

使用所述第一参数作为所述第一个量的标准化值。

5. 根据权利要求1所述的方法,其中获得所述标准化参数包括:

使用校准过程执行所述测量步骤;和

使用所述第一个量或所述第二个量作为所述标准化参数。

6. 根据权利要求1所述的方法,其中所述期望值是所述生物样品中的胎儿DNA浓度。

7. 根据权利要求6所述的方法,其还包括:

使用一种或多种后生标记测量所述胎儿DNA浓度,所述一种或多种后生标记用于计算对于所述一个或多个其他染色体区域的所述一个或多个另外的标准化参数中的至少一个。

8. 根据权利要求7所述的方法, 其中所述一种或多种后生标记包括一种或多种DNA甲基化标记。

9. 根据权利要求6所述的方法, 其还包括:

通过下述使用遗传标记计算所述胎儿DNA浓度:

测量具有选自一种或多种胎儿特异性序列的胎儿特异性序列的DNA片段的第三个量, 其中所有所述胎儿具有所述胎儿特异性序列;

获得对于所述第三个量的标准化值, 所述标准化值对应于所述一个或多个另外的标准化参数中的一个; 和

使用所述标准化值作为所述胎儿DNA浓度。

10. 根据权利要求9所述的方法, 其中所述胎儿均为男性, 并且其中所述一种或多种胎儿特异性序列在Y染色体上。

11. 根据权利要求9所述的方法, 其中所述胎儿特异性序列是RHD基因, 其中所述母亲是RhD-阴性的, 并且其中所有所述胎儿为RhD-阳性的。

12. 根据权利要求9所述的方法, 其还包括:

鉴定一个或多个第二基因座, 在所述基因座处所述胎儿具有各自的第一等位基因, 并且所述母亲不具有所述各自的第一等位基因, 其中所述胎儿特异性序列是所述各自的第一等位基因, 其中获得对于所述第三个量的标准化值包括:

测量在所述一个或多个第二基因座处的等位基因的总量;

由所述第三个量与所述总量的比计算所述胎儿DNA浓度。

13. 根据权利要求1所述的方法, 其中所述一个或多个基因座包括RHD基因的基因座, 通过所述RHD基因的不存在表示的, 所述母亲对于等位基因是纯合的, 并且其中所述胎儿中的至少一个是RhD-阳性的。

14. 根据权利要求1所述的方法, 其还包括:

测定在第二染色体区域内的一个或多个第二基因座中的每一基因座处的所述妊娠女性的基因型, 所述第二染色体区域对应于所述一个或多个其他染色体区域中的一个, 所述妊娠女性在所述一个或多个第二基因座中的每一基因座处是纯合的, 或在所述一个或多个第二基因座中的每一基因座处是杂合的, 其中所述第二基因座中的每一个显示出所述生物样品中各自的主要等位基因和各自的次要等位基因;

在所述生物样品中在所述一个或多个第二基因座处测量所述一个或多个主要等位基因的第三个量和/或所述一个或多个次要等位基因的第四个量;

获得对于所述第三个量或所述第四个量的第二标准化参数; 和

使用所述第二标准化参数获得所述期望值。

15. 根据权利要求14所述的方法, 其中所述第二标准化参数用作所述期望值。

16. 根据权利要求14所述的方法, 其中计算所述差异包括:

由第一组染色体区域的所述一个或多个另外的标准化参数计算第一统计值; 其中使用所述第一统计值作为所述期望值。

17. 根据权利要求16所述的方法, 其中对于所述第一组的染色体区域的所有基因座, 所述各自的主要等位基因比所述各自的次要等位基因更丰富。

18. 根据权利要求16所述的方法, 其中所述第一统计值是在对于所述第一组染色体区

域的标准化参数的直方图中所述标准化参数的值中的峰的位置。

19. 根据权利要求16所述的方法, 其还包括:

由包括所述第一染色体区域的第二组染色体区域的参数计算第二统计值,

其中比较所述标准化参数与所述截断值包括:

比较所述第二统计值与所述截断值。

20. 根据权利要求19所述的方法, 其还包括:

通过分析所计算参数的直方图鉴定所述第一组和第二组染色体区域, 其中所述第一组染色体区域对应于所述直方图中的第一个峰, 并且所述第二组染色体区域对应于所述直方图中的第二个峰。

21. 根据权利要求1所述的方法, 其中所述差异是所述标准化参数中的标准差。

22. 根据权利要求1所述的方法, 其中使用得自大规模平行测序、PCR、连接、杂交或质谱法的数据执行所述测量。

23. 根据权利要求1所述的方法, 其中使用得自靶标大规模平行测序、定量PCR、实时PCR或数字PCR的数据执行所述测量。

24. 一种用于分析具有多个胎儿的妊娠女性的生物样品以测定妊娠女性的至少两个胎儿是否是二卵的方法, 所述生物样品包含胎儿和母体DNA的DNA片段, 其中所述生物样品是血浆或血清, 所述方法包括:

在所述生物样品的DNA片段中在一个或多个第一基因座处测量一种或多种胎儿特异性序列的第一个量;

获得对于所述第一个量的标准化参数;

计算对于一个或多个其他染色体区域的一个或多个另外的标准化参数,

如果所述胎儿对于所述第一染色体区域是遗传上相同的, 比较所述标准化参数与截断值, 以测定所述标准化参数是否在统计上不同于期望值, 所述期望值得自所述生物样品的测量并且对应于所述一个或多个另外的标准化参数,

其中比较所述标准化参数与截断值, 以测定所述标准化参数是否在统计上不同于期望值通过以下方式完成:

计算所述标准化参数与所述一个或多个另外的标准化参数之间的差异; 并且

将所述差异与截断值比较; 和

当所述差异高于所述截断值时, 确定所述妊娠女性的至少两个胎儿是二卵的。

25. 根据权利要求24所述的方法, 其中所述胎儿特异性序列在Y染色体上。

26. 根据权利要求24所述的方法, 其中所述胎儿特异性序列是RHD基因, 并且其中所述妊娠女性是RhD-阴性的。

27. 根据权利要求24所述的方法, 其中所述期望值是所述生物样品中的胎儿DNA浓度。

28. 根据权利要求24所述的方法, 其中所述一个或多个第一基因座在第一染色体区域内, 所述方法还包括:

在所述生物样品中在一个或多个第二基因座处测量一种或多种另外的胎儿特异性序列的第二个量, 所述第二基因座在与所述第一染色体区域不同的第二染色体区域内, 所述第二染色体区域对应于所述一个或多个其他染色体区域中的一个;

获得对于所述第二个量的另外的标准化参数; 和

使用所述另外的标准化参数获得所述期望值。

29. 一种用于分析具有多个胎儿的妊娠女性的生物样品以测定所述妊娠女性的至少两个胎儿是否是二卵的方法, 所述生物样品包含胎儿和母体DNA的DNA片段, 其中所述生物样品是血浆或血清, 所述方法包括:

对于多个染色体区域中的每一个:

在各自的染色体区域中的一个或多个基因座中的每一基因座处:

测量所述生物样品的DNA片段中的一个或多个等位基因; 和

测定所述生物样品的DNA片段中每个测量的等位基因的各自量;

基于所测量的等位基因的各自量, 测定所述胎儿中的至少两个是否已继承来自第一亲本的所述各自的染色体区域的不同单元型;

测定其中所述胎儿中的至少两个已继承来自所述第一亲本的不同单元型的染色体区域的第一个量; 和

比较所述第一个量与一个或多个截断值, 以测定所述胎儿中的至少两个是否是二卵的。

30. 根据权利要求29所述的方法, 其中所述第一个量是比例。

31. 根据权利要求29所述的方法, 所述方法还包括:

对于第一染色体区域测定在多个基因座处的所述第一亲本的两个单元型,

其中测定所述胎儿中的至少两个已继承来自所述第一亲本的所述第一染色体区域的不同单元型包括:

鉴定所述第一染色体区域中的第一基因座和第二基因座, 在所述基因座处所述第一亲本是杂合的;

在所述生物样品中检测在所述第一基因座处的所述第一亲本的第一单元型; 和

在所述生物样品中检测在所述第二基因座处的所述第一亲本的第二单元型。

32. 根据权利要求31所述的方法, 其中所述第一亲本是母亲, 其中所述父亲对于第一等位基因在所述第一基因座处是纯合的, 并且所述母亲对于在所述第一基因座处的所述第一等位基因和第二等位基因是杂合的, 所述第一等位基因在所述母亲的第一单元型上, 并且所述第二等位基因在所述母亲的第二单元型上, 其中在所述生物样品中检测在所述第一基因座处的所述第一亲本的第一单元型包括:

测定在所述第一基因座处测量的所述第一等位基因的各自量以统计上显著量超过在所述第二基因座处测量的所述第二等位基因的各自量;

其中所述父亲对于第三等位基因在所述第二基因座处是纯合的, 并且所述母亲对于在所述第二基因座处的所述第三等位基因和第四等位基因是杂合的, 所述第三等位基因在所述母亲的第一单元型上, 并且所述第四等位基因在所述母亲的第二单元型上, 其中在所述生物样品中检测在所述第二基因座处的所述第一亲本的第二单元型包括:

测定在所述第二基因座处测量的所述第四等位基因的各自量以统计上显著量超过在所述第二基因座处测量的所述第三等位基因的各自量。

33. 根据权利要求31所述的方法, 其中所述第一亲本是父亲, 其中所述第二亲本对于第一等位基因在所述第一基因座处是纯合的, 并且所述第一亲本对于在所述第一基因座处的所述第一等位基因和第二等位基因是杂合的, 其中在所述生物样品中检测在所述第一基因

座处的所述第一亲本的第一单元型包括：

在所述生物样品中检测在所述第一基因座处的所述第二等位基因。

34. 根据权利要求31所述的方法，其中所述第一基因座和所述第二基因座足够接近，使得在所述第一基因座和所述第二基因座之间的重组概率小于0.1%。

35. 根据权利要求29所述的方法，其中所述第一亲本是父亲，所述方法还包括：

鉴定在第一染色体区域中的第一基因座，其中所述第一亲本对于第一等位基因和第二等位基因是杂合的，两者在所述第一基因座处在所述第二亲本中不存在，

其中测定所述胎儿中的至少两个已继承来自所述第一亲本的所述第一染色体区域的不同单元型包括：

检测在所述第一基因座处的所述第一等位基因和所述第二等位基因。

36. 根据权利要求35所述的方法，其中所述第一等位基因和所述第二等位基因含有不同数目的短串联重复。

37. 一种测定来自具有至少两个胎儿的妊娠女性的生物样品中的胎儿DNA百分比的方法，所述生物样品包含胎儿和母体DNA的DNA片段，其中所述生物样品是血浆或血清，所述方法包括：

通过下述制备直方图：

对于多个染色体区域中的每一个：

鉴定在各自的染色体区域中的一个或多个基因座，在所述生物样品中在所述基因座处检测到各自的第一等位基因和各自的第二等位基因；

在所述生物样品的DNA片段中在所述一个或多个基因座处测量一个或多个第一等位基因的第一个量和/或一个或多个第二等位基因的第二个量；和

获得对于所述第一个量或所述第二个量的标准化参数；和

基于对于所述标准化参数具有指定值的多个染色体区域增大计数器；

使概率分布的线性组合与所述直方图拟合，其中所述胎儿DNA百分比是所述概率分布的线性组合的输入；和

改变所述输入胎儿DNA百分比，以发现最佳化所述概率分布的线性组合与所述直方图的拟合的最佳胎儿DNA百分比。

38. 根据权利要求37所述的方法，其中所述概率分布的线性组合具有两个概率分布，其中两个概率分布依赖所述胎儿DNA百分比。

39. 根据权利要求37所述的方法，其中所述概率分布的线性组合具有四个概率分布，其中所述概率分布中的两个依赖所述胎儿DNA百分比。

40. 根据权利要求37所述的方法，其中由所述第一个量和所述第二个量计算所述标准化参数，所述标准化参数提供在所述第一个量和所述第二个量之间的相对量。

41. 根据权利要求37所述的方法，其中所述概率分布是二项式分布或高斯分布。

42. 根据权利要求37所述的方法，其还包括：

基于所述最佳胎儿DNA百分比，鉴定对应于基因座的概率分布，在所述基因座处所述母亲是纯合的并且所述胎儿中的至少一个是杂合的；

使多组分混合模型与所鉴定的概率分布拟合，所述多组分混合模型包括对于多个组分中的每一个的混合系数；和

当所述混合系数中的至少两个超过阈值时,测定所述胎儿中的至少两个是二卵的。

43.根据权利要求42所述的方法,其中所述阈值依赖所述混合系数中的至少一个。

44.根据权利要求42所述的方法,其中所述混合模型的组分的峰的位置受约束,以具有超过预定值的分离缺口。

45.一种用于分析妊娠女性的生物样品以测定所述妊娠女性的至少两个胎儿是否是二卵的方法,所述生物样品包含胎儿和母体DNA的DNA片段,其中所述生物样品是血浆或血清,所述方法包括:

通过下述制备直方图:

对于多个染色体区域中的每一个:

鉴定在各自的染色体区域中的一个或多个基因座,在所述生物样品中在所述基因座处检测到各自的第一等位基因和各自的第二等位基因;

在所述生物样品的DNA片段中在所述一个或多个基因座处测量一个或多个第一等位基因的第二个量和/或一个或多个第二等位基因的第二个量;和

获得对于所述第一个量或所述第二个量的标准化参数;和

基于对于所述标准化参数具有指定值的多个染色体区域增大计数器;

鉴定对应于基因座的染色体区域,在所述基因座处所述母亲是纯合的并且所述胎儿中的至少一个是杂合的,或在所述基因座处所述母亲是杂合的并且所述胎儿中的至少一个是纯合的;

使多组分混合模型与对应于所鉴定的染色体区域的直方图拟合,所述多组分混合模型包括对于多个组分中的每一个的混合系数;和

使用所述混合系数中的至少两个测定所述胎儿中的至少两个是否是二卵的。

46.根据权利要求45所述的方法,其中所述多组分混合模型具有三个组分。

47.根据权利要求45所述的方法,其中所述多组分混合模型是高斯混合模型。

48.根据权利要求45所述的方法,其还包括:

通过鉴定与对于所述第一染色体区域的相应标准化参数具有最高重叠的所述混合模型的组分,测定所述胎儿对于第一染色体区域的所述一个或多个基因座的基因型。

49.根据权利要求45所述的方法,其中当所述混合系数中的至少两个超过阈值时,所述胎儿中的至少两个测定为二卵的。

50.根据权利要求45所述的方法,其中所述混合模型的组分的峰的位置受约束,以具有超过预定值的分离缺口。

51.根据权利要求45所述的方法,其中鉴定所述染色体区域包括:

对于所述生物样品的给定胎儿DNA百分比,使概率分布的线性组合与所述直方图拟合;和

鉴定对应于基因座的概率分布,在所述基因座处所述母亲是纯合的并且所述胎儿中的至少一个是杂合的,或在所述基因座处所述母亲是杂合的并且所述胎儿中的至少一个是纯合的,其中使所述多组分混合模型与所鉴定的概率分布拟合。

52.根据权利要求45所述的方法,其还包括:

使用一种或多种后生标记测定所述给定胎儿DNA百分比。

53.根据权利要求52所述的方法,其中所述一种或多种后生标记包括一种或多种DNA甲

基化标记。

54. 一种用于分析具有多个胎儿的妊娠女性的生物样品以测定妊娠女性的至少两个胎儿是否是二卵的计算机系统, 所述生物样品包含胎儿和母体DNA的DNA片段, 其中所述生物样品是血浆或血清, 所述计算机系统包括:

测定第一染色体区域内的一个或多个第一基因座中的每一基因座处的所述妊娠女性的基因型的装置, 所述妊娠女性在所述一个或多个第一基因座中的每一基因座处是纯合的, 或在所述一个或多个第一基因座中的每一基因座处是杂合的, 其中所述第一基因座中的每一个显示出所述生物样品中各自的主要等位基因和各自的次要等位基因, 其中对于所述第一基因座中的每一个, 所述各自的主要等位基因比所述各自的次要等位基因更丰富;

在所述生物样品的DNA片段中在所述一个或多个第一基因座处测量一个或多个主要等位基因的第二个量和/或一个或多个次要等位基因的第二个量的装置;

获得所述第一个量或所述第二个量的标准化参数的装置;

计算对于一个或多个其他染色体区域的一个或多个另外的标准化参数的装置,

如果所述胎儿对于所述第一染色体区域是遗传上相同的, 比较所述标准化参数与截断值, 以测定所述标准化参数是否在统计上不同于期望值的装置, 所述期望值得自所述生物样品的测量,

其中比较所述标准化参数与截断值, 以测定所述标准化参数是否在统计上不同于期望值的装置通过以下装置实现:

计算所述标准化参数与所述一个或多个另外的标准化参数之间的差异的装置; 和

将所述差异与截断值比较的装置; 和

当所述差异高于所述截断值时, 确定所述妊娠女性的至少两个胎儿是否是二卵的装置。

55. 根据权利要求54所述的计算机系统, 其还包括:

在所述生物样品中在所述第一基因座中的每一基因座处检测所述各自的主要等位基因和所述各自的次要等位基因的装置。

56. 根据权利要求54所述的计算机系统, 其中获得所述标准化参数的装置包括:

测量来自不同染色体区域内的一个或多个基因座的一种或多种序列的第三个量的装置; 和

使用所述第三个量计算所述标准化参数的装置。

57. 根据权利要求54所述的计算机系统, 其中获得所述标准化参数的装置包括:

在所述生物样品中在所述一个或多个第一基因座处测量所述一个或多个主要等位基因的第二个量和所述一个或多个次要等位基因的第二个量的装置;

测定来自所述第一个量和所述第二个量的第一参数的装置, 所述第一参数提供在所述第一个量和所述第二个量之间的相对量; 和

使用所述第一参数作为所述第一个量的标准化值的装置。

58. 根据权利要求54所述的计算机系统, 其中获得所述标准化参数的装置包括:

使用校准过程执行所述测量步骤的装置; 和

使用所述第一个量或所述第二个量作为所述标准化参数的装置。

59. 根据权利要求54所述的计算机系统, 其中所述期望值是所述生物样品中的胎儿DNA

浓度。

60. 根据权利要求59所述的计算机系统, 其还包括:

使用一种或多种后生标记测量所述胎儿DNA浓度的装置, 所述装置用于计算对于所述一个或多个其他染色体区域的所述一个或多个另外的标准化参数中的至少一个。

61. 根据权利要求60所述的计算机系统, 其中所述一种或多种后生标记包括一种或多种DNA甲基化标记。

62. 根据权利要求59所述的计算机系统, 其还包括:

通过下述使用遗传标记计算所述胎儿DNA浓度的装置:

测量具有选自一种或多种胎儿特异性序列的胎儿特异性序列的DNA片段的第三个量, 其中所有所述胎儿具有所述胎儿特异性序列;

获得对于所述第三个量的标准化值, 所述标准化值对应于所述一个或多个另外的标准化参数中的一个; 和

使用所述标准化值作为所述胎儿DNA浓度。

63. 根据权利要求62所述的计算机系统, 其中所述胎儿均为男性, 并且其中所述一种或多种胎儿特异性序列在Y染色体上。

64. 根据权利要求62所述的计算机系统, 其中所述胎儿特异性序列是RHD基因, 其中所述母亲是RhD-阴性的, 并且其中所有所述胎儿为RhD-阳性的。

65. 根据权利要求62所述的计算机系统, 其还包括:

鉴定一个或多个第二基因座的装置, 在所述基因座处所述胎儿具有各自的第一等位基因, 并且所述母亲不具有所述各自的第一等位基因, 其中所述胎儿特异性序列是所述各自的第一等位基因, 其中获得对于所述第三个量的标准化值包括:

测量在所述一个或多个第二基因座处的等位基因的总量的装置;

由所述第三个量与所述总量的比计算所述胎儿DNA浓度的装置。

66. 根据权利要求54所述的计算机系统, 其中所述一个或多个基因座包括RHD基因的基因座, 通过所述RHD基因的不存在表示的, 所述母亲对于等位基因是纯合的, 并且其中所述胎儿中的至少一个是RhD-阳性的。

67. 根据权利要求54所述的计算机系统, 其还包括:

测定在第二染色体区域内的一个或多个第二基因座中的每一基因座处的所述妊娠女性的基因型的装置, 所述第二染色体区域对应于所述一个或多个其他染色体区域中的一个, 所述妊娠女性在所述一个或多个第二基因座中的每一基因座处是纯合的, 或在所述一个或多个第二基因座中的每一基因座处是杂合的, 其中所述第二基因座中的每一个显示出所述生物样品中各自的主要等位基因和各自的次要等位基因;

在所述生物样品中在所述一个或多个第二基因座处测量所述一个或多个主要等位基因的第三个量和/或所述一个或多个次要等位基因的第四个量的装置;

获得对于所述第三个量或所述第四个量的第二标准化参数的装置; 和

使用所述第二标准化参数获得所述期望值的装置。

68. 根据权利要求67所述的计算机系统, 其中所述第二标准化参数用作所述期望值。

69. 根据权利要求67所述的计算机系统, 其中所述计算所述标准化参数与所述一个或多个另外的标准化参数之间的差异的装置包括:

由第一组染色体区域的所述一个或多个另外的标准化参数计算第一统计值的装置；其中使用所述第一统计值作为所述期望值。

70. 根据权利要求69所述的计算机系统，其中对于所述第一组的染色体区域的所有基因座，所述各自的主要等位基因比所述各自的次要等位基因更丰富。

71. 根据权利要求69所述的计算机系统，其中所述第一统计值是在对于所述第一组染色体区域的标准化参数的直方图中所述标准化参数的值中的峰的位置。

72. 根据权利要求69所述的计算机系统，其还包括：

由包括所述第一染色体区域的第二组染色体区域的参数计算第二统计值的装置，

其中比较所述标准化参数与所述截断值的装置包括：

比较所述第二统计值与所述截断值的装置。

73. 根据权利要求72所述的计算机系统，其还包括：

通过分析所计算参数的直方图鉴定所述第一组和第二组染色体区域的装置，其中所述第一组染色体区域对应于所述直方图中的第一个峰，并且所述第二组染色体区域对应于所述直方图中的第二个峰。

74. 根据权利要求73所述的计算机系统，其中所述差异是所述标准化参数中的标准差。

75. 根据权利要求54所述的计算机系统，其中使用得自大规模平行测序、PCR、连接、杂交或质谱法的数据执行所述测量。

76. 根据权利要求54所述的计算机系统，其中使用得自靶标大规模平行测序、定量PCR、实时PCR或数字PCR的数据执行所述测量。

77. 一种用于分析具有多个胎儿的妊娠女性的生物样品以测定妊娠女性的至少两个胎儿是否是二卵的计算机系统，所述生物样品包含胎儿和母体DNA的DNA片段，其中所述生物样品是血浆或血清，所述计算机系统包括：

在所述生物样品的DNA片段中在一个或多个第一基因座处测量一种或多种胎儿特异性序列的第一个量的装置；

获得对于所述第一个量的标准化参数的装置；

计算对于一个或多个其他染色体区域的一个或多个另外的标准化参数的装置；

如果所述胎儿对于所述第一染色体区域是遗传上相同的，比较所述标准化参数与截断值，以测定所述标准化参数是否在统计上不同于期望值的装置，所述期望值得自所述生物样品的测量，

其中比较所述标准化参数与截断值，以测定所述标准化参数是否在统计上不同于期望值的装置通过以下装置实现：

计算所述标准化参数与所述一个或多个另外的标准化参数之间的差异的装置；和

将所述差异与截断值比较的装置；和

当所述差异高于所述截断值时，确定所述妊娠女性的至少两个胎儿是否是二卵的装置。

78. 根据权利要求77所述的计算机系统，其中所述胎儿特异性序列在Y染色体上。

79. 根据权利要求77所述的计算机系统，其中所述胎儿特异性序列是RHD基因，并且其中所述妊娠女性是RhD-阴性的。

80. 根据权利要求77所述的计算机系统，其中所述期望值是所述生物样品中的胎儿DNA

浓度。

81. 根据权利要求77所述的计算机系统, 其中所述一个或多个第一基因座在第一染色体区域内, 所述计算机系统还包括:

在所述生物样品中在一个或多个第二基因座处测量一种或多种另外的胎儿特异性序列的第二个量的装置, 所述第二染色体区域对应于所述一个或多个其他染色体区域中的一个, 所述第二基因座在与所述第一染色体区域不同的第二染色体区域内;

获得对于所述第二个量的另外的标准化参数的装置; 和

使用所述另外的标准化参数获得所述期望值的装置。

82. 一种用于分析具有多个胎儿的妊娠女性的生物样品以测定所述妊娠女性的至少两个胎儿是否是二卵的计算机系统, 所述生物样品包含胎儿和母体DNA的DNA片段, 其中所述生物样品是血浆或血清, 所述计算机系统包括:

对于多个染色体区域中的每一个:

在各自的染色体区域中的一个或多个基因座中的每一基因座处:

测量所述生物样品的DNA片段中的一个或多个等位基因的装置; 和

测定所述生物样品的DNA片段中每个测量的等位基因的各自量的装置;

基于所测量的等位基因的各自量, 测定所述胎儿中的至少两个是否已继承来自第一亲本的所述各自的染色体区域的不同单元型的装置;

测定其中所述胎儿中的至少两个已继承来自所述第一亲本的不同单元型的染色体区域的第一个量的装置; 和

比较所述第一个量与一个或多个截断值, 以测定所述胎儿中的至少两个是否是二卵的装置。

83. 根据权利要求82所述的计算机系统, 其中所述第一个量是比例。

84. 根据权利要求82所述的计算机系统, 其还包括:

对于第一染色体区域测定在多个基因座处的所述第一亲本的两个单元型的装置,

其中测定所述胎儿中的至少两个已继承来自所述第一亲本的所述第一染色体区域的不同单元型的装置包括:

鉴定所述第一染色体区域中的第一基因座和第二基因座的装置, 在所述基因座处所述第一亲本是杂合的;

在所述生物样品中检测在所述第一基因座处的所述第一亲本的第一单元型的装置; 和

在所述生物样品中检测在所述第二基因座处的所述第一亲本的第二单元型的装置。

85. 根据权利要求84所述的计算机系统, 其中所述第一亲本是母亲, 其中所述父亲对于第一等位基因在所述第一基因座处是纯合的, 并且所述母亲对于在所述第一基因座处的所述第一等位基因和第二等位基因是杂合的, 所述第一等位基因在所述母亲的第一单元型上, 并且所述第二等位基因在所述母亲的第二单元型上, 其中在所述生物样品中检测在所述第一基因座处的所述第一亲本的第一单元型的装置包括:

测定在所述第一基因座处测量的所述第一等位基因的各自量以统计上显著量超过在所述第二基因座处测量的所述第二等位基因的各自量的装置;

其中所述父亲对于第三等位基因在所述第二基因座处是纯合的, 并且所述母亲对于在所述第二基因座处的所述第三等位基因和第四等位基因是杂合的, 所述第三等位基因在所

述母亲的第一单元型上,并且所述第四等位基因在所述母亲的第二单元型上,其中在所述生物样品中检测在所述第二基因座处的所述第一亲本的第二单元型的装置包括:

测定在所述第二基因座处测量的所述第四等位基因的各自量以统计上显著量超过在所述第二基因座处测量的所述第三等位基因的各自量的装置。

86.根据权利要求84所述的计算机系统,其中所述第一亲本是父亲,其中所述第二亲本对于第一等位基因在所述第一基因座处是纯合的,并且所述第一亲本对于在所述第一基因座处的所述第一等位基因和第二等位基因是杂合的,其中在所述生物样品中检测在所述第一基因座处的所述第一亲本的第一单元型的装置包括:

在所述生物样品中检测在所述第一基因座处的所述第二等位基因的装置。

87.根据权利要求84所述的计算机系统,其中所述第一基因座和所述第二基因座足够接近,使得在所述第一基因座和所述第二基因座之间的重组概率小于0.1%。

88.根据权利要求82所述的计算机系统,其中所述第一亲本是父亲,所述计算机系统还包括:

鉴定在第一染色体区域中的第一基因座的装置,其中所述第一亲本对于第一等位基因和第二等位基因是杂合的,两者在所述第一基因座处在所述第二亲本中不存在,

其中测定所述胎儿中的至少两个已继承来自所述第一亲本的所述第一染色体区域的不同单元型的装置包括:

检测在所述第一基因座处的所述第一等位基因和所述第二等位基因的装置。

89.根据权利要求88所述的计算机系统,其中所述第一等位基因和所述第二等位基因含有不同数目的短串联重复。

90.一种测定来自具有至少两个胎儿的妊娠女性的生物样品中的胎儿DNA百分比的计算机系统,所述生物样品包含胎儿和母体DNA的DNA片段,其中所述生物样品是血浆或血清,所述计算机系统包括:

通过下述制备直方图的装置:

对于多个染色体区域中的每一个:

鉴定在各自的染色体区域中的一个或多个基因座,在所述生物样品中在所述基因座处检测到各自的第一等位基因和各自的第二等位基因;

在所述生物样品的DNA片段中在所述一个或多个基因座处测量一个或多个第一等位基因的第二个量和/或一个或多个第二等位基因的第二个量;和

获得对于所述第一个量或所述第二个量的标准化参数;和基于对于所述标准化参数具有指定值的多个染色体区域增大计数器;

使概率分布的线性组合与所述直方图拟合的装置,其中所述胎儿DNA百分比是所述概率分布的线性组合的输入;和

改变所述输入胎儿DNA百分比,以发现最佳化所述概率分布的线性组合与所述直方图的拟合的最佳胎儿DNA百分比的装置。

91.根据权利要求90所述的计算机系统,其中所述概率分布的线性组合具有两个概率分布,其中两个概率分布依赖所述胎儿DNA百分比。

92.根据权利要求90所述的计算机系统,其中所述概率分布的线性组合具有四个概率分布,其中所述概率分布中的两个依赖所述胎儿DNA百分比。

93. 根据权利要求90所述的计算机系统,其中由所述第一个量和所述第二个量计算所述标准化参数,所述标准化参数提供在所述第一个量和所述第二个量之间的相对量。

94. 根据权利要求90所述的计算机系统,其中所述概率分布是二项式分布或高斯分布。

95. 根据权利要求90所述的计算机系统,其还包括:

基于所述最佳胎儿DNA百分比,鉴定对应于基因座的概率分布的装置,在所述基因座处所述母亲是纯合的并且所述胎儿中的至少一个是杂合的;

使多组分混合模型与所鉴定的概率分布拟合的装置,所述多组分混合模型包括对于多个组分中的每一个的混合系数;和

当所述混合系数中的至少两个超过阈值时,测定所述胎儿中的至少两个是二卵的装置。

96. 根据权利要求95所述的计算机系统,其中所述阈值依赖所述混合系数中的至少一个。

97. 根据权利要求95所述的计算机系统,其中所述混合模型的组分的峰的位置受约束,以具有超过预定值的分离缺口。

98. 一种用于分析妊娠女性的生物样品以测定所述妊娠女性的至少两个胎儿是否是二卵的计算机系统,所述生物样品包含胎儿和母体DNA的DNA片段,其中所述生物样品是血浆或血清,所述计算机系统包括:

通过下述制备直方图的装置:

对于多个染色体区域中的每一个:

鉴定在各自的染色体区域中的一个或多个基因座,在所述生物样品中在所述基因座处检测到各自的第一等位基因和各自的第二等位基因;

在所述生物样品的DNA片段中在所述一个或多个基因座处测量一个或多个第一等位基因的第二个量和/或一个或多个第二等位基因的第二个量;和

获得对于所述第一个量或所述第二个量的标准化参数;和

基于对于所述标准化参数具有指定值的多个染色体区域增大计数器;

鉴定对应于基因座的染色体区域的装置,在所述基因座处所述母亲是纯合的并且所述胎儿中的至少一个是杂合的,或在所述基因座处所述母亲是杂合的并且所述胎儿中的至少一个是纯合的;

使多组分混合模型与对应于所鉴定的染色体区域的直方图拟合的装置,所述多组分混合模型包括对于多个组分中的每一个的混合系数;和

使用所述混合系数中的至少两个测定所述胎儿中的至少两个是否是二卵的装置。

99. 根据权利要求98所述的计算机系统,其中所述多组分混合模型具有三个组分。

100. 根据权利要求98所述的计算机系统,其中所述多组分混合模型是高斯混合模型。

101. 根据权利要求98所述的计算机系统,其还包括:

通过鉴定与对于所述第一染色体区域的相应标准化参数具有最高重叠的所述混合模型的组分,测定所述胎儿对于第一染色体区域的所述一个或多个基因座的基因型的装置。

102. 根据权利要求98所述的计算机系统,其中当所述混合系数中的至少两个超过阈值时,所述胎儿中的至少两个测定为二卵的。

103. 根据权利要求98所述的计算机系统,其中所述混合模型的组分的峰的位置受约

束,以具有超过预定值的分离缺口。

104.根据权利要求98所述的计算机系统,其中鉴定所述染色体区域的装置包括:

对于所述生物样品的给定胎儿DNA百分比,使概率分布的线性组合与所述直方图拟合的装置;和

鉴定对应于基因座的概率分布的装置,在所述基因座处所述母亲是纯合的并且所述胎儿中的至少一个是杂合的,或在所述基因座处所述母亲是杂合的并且所述胎儿中的至少一个是纯合的,其中使所述多组分混合模型与所鉴定的概率分布拟合。

105.根据权利要求98所述的计算机系统,其还包括:

使用一种或多种后生标记测定所述给定胎儿DNA百分比的装置。

106.根据权利要求105所述的计算机系统,其中所述一种或多种后生标记包括一种或多种DNA甲基化标记。

多胎妊娠的分子测试

[0001] 与相关申请的交叉参考

[0002] 本申请要求来自2011年2月24日提交的名称为“MOLECULAR TESTING OF TWIN PREGNANCIES”的美国临时申请号61/446,256的优先权,并且是它的非临时申请,所述美国临时申请的完整内容以引用方式并入本文用于所有目的。

[0003] 本申请涉及2010年11月5日提交的、Lo等人共同拥有的名称为“Fetal Genomic Analysis From A Maternal Biological Sample”的美国专利申请号12/940,993(代理人案号80015-006710US),所述美国专利申请的公开内容全文以引用方式并入。

背景技术

[0004] 多胎妊娠指的是其中孕妇怀有超过一个胎儿的妊娠。双胎妊娠是最常见形式的多胎妊娠。单卵双生指的是源自相同受精卵的一对双胞胎。因此,该对双胞胎具有整个全基因组的相同基因构成。二卵双生是源自两个不同受精卵的一对双胞胎。该对双胞胎的基因构成将是不同的。相反,这对双胞胎的基因构成的相似性将类似在不同时间出生的一对同胞。

[0005] 关于双胎妊娠的接合性的信息常规上已通过超声扫描(Chauhan SP等人Am J Obstet Gynecol2010;203:305-315)或侵入性产前诊断(例如羊膜穿刺术)(Chen CP等人Hum Reprod2000;15:929-934)获得。此类接合性信息可用于后续产科管理。例如,在执行羊膜穿刺术以用于非整倍性检测的事件中,涉及二卵双生的妊娠将要求每个羊膜囊的个别取样。对于涉及两个羊膜囊的单卵双生妊娠,理论上仅需要两个羊膜囊之一的取样。然而,超声扫描可以是不准确或受限制的(例如胎儿具有不同性别),并且侵入性产前诊断可导致对胎儿和/或母亲有害。

[0006] 相应地,希望提供关于具有多胎儿的妊娠的接合性信息的新技术。

发明内容

[0007] 本发明的实施例提供了用于使用取自母亲的生物样品测定多胎妊娠的接合性的方法、系统和仪器,其对于胎儿是非侵入性的。可对于特定染色体区域分析样品(例如血浆)中的胎儿和母体DNA,以鉴定胎儿中的基因差异。例如,当胎儿是二卵的时,用于测量主要(primary)或次要(secondary)等位基因的标准化参数可显示关于不同染色体区域的变化。如果胎儿是遗传上相同的,则此类变化可相对于期望值进行测定。提供统计方法以用于分析标准化参数的变化,以测定胎儿DNA浓度和在多个基因座处的母系-胎儿混合的基因型。亲本基因型和单元型信息也可用于鉴定不同亲本单元型的继承,以指示在胎儿中的遗传差异。在其他利益中,多胎妊娠的接合性的测定可帮助例如使用母血完成的非侵入性产前测试程序的使用。

[0008] 根据一个实施例,用于分析具有多个胎儿的妊娠女性的生物样品以测定妊娠女性的至少两个胎儿是否是二卵的方法。生物样品包含胎儿和母体DNA。在第一染色体区域内的一个或多个第一基因座中的每一基因座处测定妊娠女性的基因型。母亲在第一基因座中的每一基因座处是纯合的,或在第一基因座中的每一基因座处是杂合的。第一基因座中的每

一个显示出生物样品中各自的主要等位基因和各自的次要等位基因。对于第一基因座中的每一个,各自的主要等位基因比各自的次要等位基因更丰富。在生物样品中在一个或多个第一基因座处测量一个或多个主要等位基因的第二个量和/或一个或多个次要等位基因的第二个量。获得对于第一个量或第二个量的标准化参数。比较标准化参数与截断值,以测定如果胎儿对于第一染色体区域是遗传上相同的,则标准化参数是否在统计上不同于期望值。期望值得自生物样品的测量。基于标准化参数与截断值的比较,测定妊娠女性的至少两个胎儿是否是二卵的。

[0009] 根据另一个实施例,在生物样品中测量在一个或多个第一基因座处测量的一个或多个胎儿特异性序列的第一个量。获得对于第一个量的标准化参数。比较标准化参数与截断值,以测定如果胎儿对于第一染色体区域是遗传上相同的,则标准化参数是否在统计上不同于期望值。期望值得自生物样品的测量。随后,基于标准化参数与截断值的比较,测定妊娠女性的至少两个胎儿是否是二卵的。

[0010] 根据另一个实施例,对于多个染色体区域中的每一个,在生物样品中在各自的染色体区域中的一个或多个基因座中的每一基因座处测量一个或多个等位基因,并且在每个基因座处测定每个测量的等位基因的各自量。基于测量的等位基因的各自量测定胎儿中的至少两个是否已继承来自第一亲本的各自染色体区域的不同单元型。测定其中胎儿中的至少两个已继承来自第一亲本的不同单元型的染色体区域的第一个量。比较第一个量与一个或多个截断值,以测定胎儿中的至少两个是否是二卵的。

[0011] 根据另一个实施例,如下制备直方图。对于多个染色体区域中的每一个:鉴定在各自的染色体区域中的一个或多个基因座,在生物样品中在所述基因座处检测到各自的第一等位基因和各自的第二等位基因,在生物样品中在一个或多个基因座处测量一个或多个第一等位基因的第二个量和/或一个或多个第二等位基因的第二个量,并且获得对于第一个量或第二个量的标准化参数。基于对于标准化参数具有指定值的多个染色体区域增大直方图的计数器。鉴定对应于基因座的染色体区域,在所述基因座处母亲是纯合的并且胎儿中的至少一个是杂合的,或在所述基因座处母亲是杂合的并且胎儿中的至少一个是纯合的。使多组分混合模型与对应于鉴定的染色体区域的直方图拟合。多组分混合模型包括对于多个组分中的每一个的混合系数。使用混合系数中的至少两个测定胎儿中的至少两个是否是二卵的。

[0012] 根据另一个实施例,提供了测定来自具有至少两个胎儿的妊娠女性的生物样品中的胎儿DNA百分比的方法。对于多个染色体区域中的每一个:鉴定在各自的染色体区域中的一个或多个基因座,在生物样品中在所述基因座处检测到各自的第一等位基因和各自的第二等位基因,在生物样品中在一个或多个基因座处测量一个或多个第一等位基因的第二个量和/或一个或多个第二等位基因的第二个量,并且获得对于第一个量或第二个量的标准化参数。基于对于标准化参数具有指定值的多个染色体区域增大直方图的计数器。使概率分布的线性组合与直方图拟合,其中胎儿DNA百分比是概率分布的线性组合的输入。改变输入胎儿DNA百分比,以发现最佳化概率分布的线性组合与直方图的拟合的最佳胎儿DNA百分比。

[0013] 其他实施例涉及与本文描述的方法相关的系统和计算机可读介质。

[0014] 本发明的性质和优点的更佳理解可关于下述详述和附图获得。

附图说明

[0015] 图1是示出根据本发明的实施例的方法100的流程图,所述方法100用于分析具有多个胎儿的妊娠女性的生物样品,以测定妊娠女性的至少两个胎儿是否是二卵的。

[0016] 图2A和2B显示了当对于给定染色体区域已知在两个基因座处的亲本单元型时,用于测定胎儿的接合性的例子。图2A显示了其中母亲在两个基因座处是纯合的并且父亲是杂合的例子。图2B显示了其中母亲在两个基因座处是杂合的并且父亲是纯合的例子。

[0017] 图3还显示了其中基因座具有四个不同等位基因的微卫星分析的例子。

[0018] 图4是示出根据本发明的实施例的方法400的流程图,所述方法400用于分析具有多个胎儿的妊娠女性的生物样品,以测定妊娠女性的至少两个胎儿对于第一染色体区域是否是遗传上不同的。

[0019] 图5是示出方法500的流程图,所述方法500用于通过测定对于来自第一亲本的第一染色体区域的表观部分参数(fractional parameter)(例如胎儿DNA浓度)测定妊娠女性的至少两个胎儿对于该第一染色体区域是否是遗传上不同的。

[0020] 图6和7示出了根据本发明的实施例,使用母体血浆中的胎儿DNA部分(fraction)中的区域基因组变化揭示双胎妊娠的接合性。

[0021] 图8显示了根据本发明的实施例,对于贡献等量的胎儿DNA的二卵双生的胎儿DNA浓度的示例直方图。

[0022] 图9显示了根据本发明的实施例,当两个胎儿对母体血浆样品贡献不同量的DNA时,基于SNP分析的部分(fractional)胎儿DNA分布的直方图。

[0023] 图10显示了在怀有一对二卵双生的孕妇中重组对表观部分胎儿DNA浓度(fractional fetal DNA concentration)的作用的例子。

[0024] 图11是示出根据本发明的实施例的方法1100的流程图,所述方法1100测定来自具有至少两个胎儿的妊娠女性的生物样品中的胎儿DNA百分比,并且测定胎儿中的至少两个是否是二卵的。

[0025] 图12A-12E是显示根据本发明实施例的推导SNP分析的结果的表。

[0026] 图13显示了关于二卵妊娠的两个不同峰的鉴定。

[0027] 图14显示了关于单卵妊娠的一个峰的鉴定。

[0028] 图15显示了对于怀有单卵和二卵双生的孕妇的不同染色体区域的血浆部分胎儿DNA浓度。

[0029] 图16显示了假定一对单卵双生的存在,测定随机变异的水平的模拟分析。

[0030] 图17显示了根据本发明的实施例,示出对于三个胎儿(胎儿A、B和C)的多个可能峰的直方图。

[0031] 图18显示了可与根据本发明的实施例的系统和方法一起使用的示例计算机系统1800的方框图。

[0032] 定义

[0033] 如本文使用的术语“生物样品”指的是取自受试者(例如人,例如孕妇)并含有一种或多种目的核酸分子的任何样品。例子包括血浆、唾液、胸膜液、汗、腹水液、胆汁、尿、血清、胰液、粪便和宫颈涂片样品。

[0034] 术语“核酸”或“多核苷酸”指的是以单链或双链形式的脱氧核糖核酸(DNA)或核糖核酸(RNA)及其聚合物。除非具体限制,该术语包含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考核酸相似的结合性质并且以与天然存在的核苷酸相似的方式代谢。除非另有说明,特定核酸序列无疑还包含其保守修饰的变体(例如简并密码子置换)、等位基因、直系同源物、SNP和互补序列以及明确指示的序列。具体地,简并密码子置换可通过生成其中一个或多个所选(或全部)密码子的第三个位置由混合碱基和/或脱氧肌苷残基置换的序列来实现(Batzer MA等人,Nucleic Acid Res1991;19:5081;Ohtsuka E等人,J Biol Chem1985;260:2605-2608;和Rossolini GM等人,Mol Cell Probes1994;8:91-98)。术语核酸可与基因、cDNA、mRNA、小非编码RNA、微小RNA(miRNA)、Piwi-相互作用RNA和由基因或基因座编码的小发夹RNA(shRNA)互换使用。

[0035] 术语“基因”意指涉及于产生多肽链或转录的RNA产物的DNA区段。它可包括在编码区前和后的区域(引导区和非转录尾区)以及在各个编码区段(外显子)之间的间插序列(内含子)。

[0036] 如本文使用的术语“反应”指的是涉及化学、酶促或物理反应的任何过程,所述反应指示特定目的多核苷酸序列的存在或不存在。“反应”的例子是扩增反应例如聚合酶链反应(PCR)。“反应”的另一个例子是通过合成、连接、杂交或降解的测序反应。“信息反应”是指示一个或多个特定目的多核苷酸序列的存在的反应,并且在一种情况下仅存在一个目的序列。如本文使用的术语“孔”指的是在局限结构内的预定位置处的反应,所述局限结构例如孔形小瓶、小室、PCR阵列中的室、乳状液中的小滴、颗粒、纳米孔或表面上的区域。

[0037] 如本文使用的术语“过度表现的核酸序列”指的是比生物样品中的其他序列更丰富的在两个目的序列(例如临床有关序列和本底序列)中的核酸序列。

[0038] 如本文使用的术语“基于”意指“至少部分基于”,并且指的是在另一个值的测定中使用的一个值(或结果),例如在方法的输入和该方法的输出的关系中发生的。如本文使用的术语“衍生”还指的是方法的输入和该方法的输出的关系,例如当衍生是公式的计算时发生的。

[0039] 如本文使用的术语“定量数据”意指得自一个或多个反应且提供一个或多个数值的数据。例如,显示用于特定序列的荧光标记的孔数目将是定量数据。

[0040] 如本文使用的术语“参数”意指表征定量数据集和/或定量数据集之间的数字关系的数值。例如,在第一核酸序列的第一个量和第二核酸序列的第二个量之间的比(或比的函数)是参数。

[0041] 如本文使用的,术语“基因座”或其复数形式是在整个基因组上具有变化的任何长度的核苷酸(或碱基对)的位置或地址。术语“等位基因”指的是在相同物理基因组基因座处的备选DNA序列,其可导致或不导致不同的表型性状。在具有每个染色体(除了男性人受试者中的性染色体外)的两个拷贝的任何特定二倍体生物中,关于每个基因的基因型包含在那个基因座处存在的等位基因对,其在纯合子中是相同的并且在杂合子中是不同的。生物群体或物种一般包括在多个个体中的每个基因座处的多个等位基因。其中在群体中发现超过一个等位基因的基因组基因座称为多态位点。在基因座处的等位基因变化可测量为存在的等位基因数目(即多态性程度),或群体中的杂合子比例(即杂合性率)。序列(例如基因)的存在或不存在也视为等位基因变化的类型,因为基因座可包括该序列或不包括该序列。

例如通过连接通常在缺失序列前和后出现的序列,可鉴定序列(例如RHD基因)的这种不存在。如本文使用的,术语“多态性”指的是人基因组中的任何个体间变化,与其频率无关。此类变化的例子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(其可以是引起疾病的)和拷贝数变化。

[0042] 如本文使用的术语“单元型”指的是在相同染色体或染色体区域上一起传递的在多个基因座处的等位基因组合。单元型可指的是少至一对基因座或染色体区域或整个染色体。“染色体区域”指的是关于特定染色体的多个核苷酸位置。染色体区域可以是整个染色体或较小的部分。在正常人中,染色体区域将具有两个单元型,区域在其内的染色体的每个拷贝一个。两个单元型在染色体区域中可以是相同或不同的。

[0043] 如本文使用的术语“截断值”意指其值用于在生物样品分类的两个或更多个状态(例如患病和非患病)之间仲裁的数值。例如,如果参数大于截断值,则作出定量数据的第一分类(例如患病状态);或如果参数小于截断值,则作出定量数据的不同分类(例如非患病状态)。

[0044] 如本文使用的术语“不平衡”意指如通过一些临床有关核酸序列中的至少一个截断值定义的与参考数量的任何显著偏差。例如,参考数量可以是3/5的比,并且因此如果测量的比是1:1,则将发生不平衡。

[0045] 术语“测序标签”指的是得自核酸分子的全部或部分的序列,例如DNA片段。在一个实施例中,仅测序片段的一个末端,例如约30bp。测序标签随后可与参考基因组比对。作为另外一种选择,可测序片段的两个末端,以生成两个测序标签,其可提供比对中的更大准确度,并且还提供片段长度。在另外一个实施例中,线性DNA片段可例如通过连接而环化,并且可测序跨越连接位点的部分。

[0046] 术语“通用测序”指的是其中衔接子加入片段的末端并且用于测序的引物附着至衔接子的测序。因此,任何片段均可用相同引物进行测序,并且因此测序可以是随机的。

[0047] 如本文使用的术语“分类”指的是与样品的特定性质相关的任何数目或其他字符(包括字)。例如,“+”号可表明样品分类为具有缺失或扩增。术语“截断”和“阈值”指的是在操作中使用的预定数目。例如,截断大小可指的是超过其的片段被排除的大小。阈值可以是超过或低于其就应用特定分类的值。这些术语中的任一均可在这些背景中的任一中使用。

[0048] 术语“直方图”指的是贮存在指定范围内的多个数据点的计数的数据结构。例如,染色体区域数目显示出在一组值上的参数(例如胎儿DNA百分比)。

[0049] 术语“最佳的”指的是测定为在数字上优于一个或多个其他值的任何值。例如,最佳值不一定是可能的值,而是可仅仅满足标准(例如来自先前值的成本函数中的变化在容许量内)。

具体实施方式

[0050] 胎儿DNA已显示存在于孕妇的血浆和血清中(Lo等人Lancet1997;350:485-487;和美国专利6,258,540)。母体血浆或血清中的胎儿DNA的分析具有它是相对非侵入性的优点,仅需要母亲血液的样品。与用于产前筛选的常规非侵入性方法例如超声扫描比较,测试母体血浆或血清中的胎儿DNA将允许胎儿遗传信息的直接评估。此处,我们示出了可如何分析母体血浆或血清(或其他生物样品)中的DNA的原理,以区分孕妇是怀有单卵还是二卵胎儿

(例如一对单卵或二卵双生)。

[0051] 首先,跨越多个染色体区域的分析测定在胎儿基因组之间的差异水平,其用于执行关于胎儿接合性的分类。接下来,我们讨论分析特定染色体区域的具体例子,以测定当两个不同父系单元型在多个基因座处是已知的,并且母亲在这些基因座处是纯合的时(还讨论了其中母亲是杂合的并且父亲是纯合的例子),胎儿在该区域中是否在遗传上不同(例如如果双胞胎各自继承不同的父系单元型)。还描述了当两个亲本的基因型信息均是已知时的其他例子,例如当三个或更多个不同等位基因在特定基因座处时。随后,描述了比较和/或鉴定在表观胎儿DNA浓度的量度中的变化或跨越区域的其他参数的变化的技术。此类技术可使用明确的母系基因型信息,或经由含有胎儿和母体DNA的生物样品的测量例如孕妇的血浆推导母系基因型。还解释了用于多胎妊娠的推导技术。

[0052] I. 使用不同染色体区域测定接合性

[0053] 单卵胎儿是遗传上相同的,而二卵胎儿是遗传上不同的。遗传差异的程度将类似于在其他妊娠时相同亲本生育的其他同胞。然而,由于统计机率,二卵胎儿可共享在基因组的部分处的相同遗传序列。

[0054] 胎儿通常对于特定染色体区域具有两个单元型(其可以是相同或不同的),关于染色体的两个拷贝各自的一个单元型。如果胎儿是单卵的,则胎儿将在染色体区域中具有相同的两个单元型。此外,由于统计机率,二卵胎儿可对于给定染色体区域具有相同对的单元型。实施例可分析多个染色体区域,以测定胎儿是否已继承不同单元型,并且随后不同的区域百分比(或其他参数)用于测定胎儿是单卵的还是二卵的。可分析指定数目的染色体区域以获得所需统计显著性。

[0055] A. 方法

[0056] 图1是示出根据本发明的实施例的方法100的流程图,所述方法100用于分析具有多个胎儿的妊娠女性的生物样品,以测定妊娠女性的至少两个胎儿是否是二卵的。生物样品包括胎儿和母体DNA。例如,可使用来自孕妇的血浆。方法100可使用计算机系统实现,如本文描述的任何方法可以一样。

[0057] 在步骤110中,对于多个染色体区域中的每一个,在各自的染色体区域中的一个或多个基因座中的每一基因座处测量生物样品中的一个或多个等位基因。在生物样品中的DNA可通过多种技术进行分析,所述多种技术包括定量聚合酶链反应(PCR)、数字PCR、测序(例如Sanger测序和大规模平行测序)、连接、杂交和质谱法(例如Sequenom MassARRAY平台),以测量在该基因座处的特定等位基因。对于测序,在测序前可执行富集步骤,以增加来自特定组的染色体区域的DNA片段的百分比。在一个实施例中,可使用溶液相(例如使用Agilent SureSelect平台)或固体相(例如使用Roche NimbleGen平台)杂交执行此类富集步骤。测量步骤自身可使用得自任何一种或多种上述技术的数据实现。例如,测序标签可与参考基因组比对,以鉴定由其获得测序标签的相应DNA片段的位置和等位基因。可用于分析生物样品中的DNA的一种方法是称为选择区域数字分析(DANSR)的技术,其涉及杂交、连接、扩增和大规模平行测序的步骤(Sparks AB等人Am J Obstet Gynecol 2012;doi:10.1016/j.ajog.2012.01.030)。

[0058] 可使用的大规模平行测序平台的例子包括Illumina Genome Analyzer平台、Life Technologies SOLiD、Ion Torrent and Ion Proton系统、Roche454系统、来自Helicos的

单分子测序系统、Pacific Biosciences、或基于纳米孔的系统(例如来自Oxford Nanopore Technologies的系统)。在另一个实施例中,执行靶标测序,其中捕获或扩增所选基因组区域(例如含有SNP或其他类型的变化例如微卫星多态性的那些),并且随后优先对于此类捕获或扩增区域执行大规模平行测序。在一个实施例中,使用Agilent SureSelect系统执行靶标测序(Liao GJ等人Clin Chem2011;57:92-101)。靶标测序还可使用Roche NimbleGen系统执行。

[0059] 数字PCR可用于分析母体血浆中的单个DNA分子(Vogelstein B和Kinzler KW.Proc Natl Acad Sci USA1999;96:9236-9241;Lo YMD等人Proc Natl Acad Sci USA2007;104:13116-13121)。数字PCR可使用多种平台执行,包括但不限于微流体(Lun FMF等人Clin Chem2008;54:1664-1672),乳滴PCR(Dressman D等人Proc Natl Acad Sci USA2003;100:8817-8822),包括RainDance平台(Kiss MM等人Anal Chem2008;80:8975-8981)。

[0060] 在步骤120中,测定每个测量的等位基因的各自量。例如,可测序(例如使用通用测序)样品中的DNA片段,以获得测序标签(其可以是配对末端标签),并且测序标签可与参考基因组比对,以鉴定片段的基因组位置。如果母亲和/或胎儿的基因组在基因座处均具有变化,则对于该基因座测量不同等位基因。可跟踪对应于在基因座处的每个等位基因的片段的各自量。在特定基因座处的特定等位基因的各自量可以多种方式进行测量,例如通过片段的数目或比例、在相同多态位点处的不同等位基因之间的比、在微阵列杂交上的信号强度、在实时PCR中的阈值循环或阈值循环中的差异、如通过数字PCR检测的对于等位基因阳性的反应的比例或数目、和质谱法分析中的峰高。

[0061] 在步骤130中,对于多个染色体区域中的每一个,基于测量的等位基因的各自量测定胎儿中的两个是否已继承来自第一亲本的各自染色体区域的不同单元型。如果一个胎儿继承第一单元型并且另一个胎儿继承不同单元型,则这指示二卵。如果存在超过两个胎儿,则一对可继承相同单元型并且不同对可继承不同单元型。在一个实施例中,不同单元型的继承可由在一个或多个基因座处测量的等位基因数据推断。例如,推导方法可用于鉴定胎儿基因组中的差异,如下文描述的。

[0062] 在另一个实施例中,来自亲本中的一个或多个的基因型信息可以是已知的。此类信息可允许来自仅一个胎儿的测量用于测定不同单元型是否已从第一亲本继承。例如,如果在第一基因座处在亲本中存在三个不同基因型,则随后可使用仅第一基因座作出决定。然而,如果在基因座处仅存在两个基因型,则随后可能需要在另一基因座处的测量。一些例子在下文提供。

[0063] 如果多个基因座用于染色体区域,则来自基因座的数据可以不同方式组合。例如,如果已知等位基因与特定单元型相关,则随后在特定基因座处具有特定等位基因的片段数目的计数有效成为关于特定单元型的片段数目的计数。例如,通过总计具有第一单元型的等位基因和基因座的片段计数,可测定对应于第一亲本(例如父亲)的第一单元型的片段数目。作为另外一种选择,可对于每个基因座独立地作出测定,并且关于每个基因座的测定可就一致性进行比较。

[0064] 在步骤140中,测定其中胎儿中的至少两个已继承来自第一亲本的不同单元型的染色体区域的第一个量。第一个量可仅仅是已鉴定为在胎儿之间具有差异的染色体区域数

目。作为另一个例子,第一个量可以是鉴定为在胎儿之间具有差异的染色体区域百分比。

[0065] 在步骤150中,比较第一个量与一个或多个截断值,以测定两个胎儿是单卵的还是二卵的。例如,第一个量可以是百分比(或其他比例),例如10%,并且这个量可与截断值比较,其中超过5%分类为二卵的。可基于所需准确度、步骤130中的测定准确度、使用的染色体区域数目、和群体中的不同染色体区域的连锁不平衡和分析的不同染色体区域之间的重组概率,测定截断值,其在下一个部分中描述。在一个方面,如果存在超过两个胎儿,则测定可仅是一对是二卵的,从而留下另一对是单卵的可能性。

[0066] B. 关于单元型检测的统计分析

[0067] 如上所述,来自亲本的染色体区域的两个单元型(即当不同时)的继承指示该对双胞胎是二卵而不是单卵的。例如,在染色体区域的第一基因座处的两个父系单元型的检测将指示该对双胞胎是二卵的。然而,存在用于检测母体血浆样品中的染色体区域的仅一个父系单元型的几种可能解释。

[0068] 首先,两个胎儿可仅偶然地已继承来自父亲的染色体区域中的相同父系单元型。当它们是一对单卵双生时,它们将跨越全基因组一直继承来自父亲的相同父系单元型。然而,即使它们是二卵双生,对于任何特定区域也存在它们将继承来自父亲的相同父系单元型的50%机率。然而,该对二卵双生将跨越全基因组继承相同父系单元型将是非常不可能的。

[0069] 在另一种情况下,两个二卵双生胎儿可继承不同的父系单元型,但由于取样不足,仅一个父系单元型在特定分析中检测到。这些多种情况发生的概率依赖对于特定染色体区域分析的在母体血浆样品中的部分胎儿DNA浓度和母体血浆DNA分子数目。下文,我们提供了关于对应于染色体区域的多少分子以及多少染色体区域可用于取得具有足够统计力的强分类的计算,以使由于取样不足的假单元型解读的机率降到最低。

[0070] 需要分析的对应于特定染色体区域的分子数目可以下述方式进行测定。当父系单元型存在于母体血浆中时,在特定母体血浆DNA样品中检测到它的概率依赖携带那个父系单元型的胎儿DNA的部分浓度和分析的分子总数目,并且受泊松分布控制。

[0071] 表1显示了对应于需要分析的目的染色体区域的分子数目,使得具有在母体血浆中存在但在特定样品中未检测到的父系单元型的概率小于1%。该数字基于下式进行计算: $0.01 > \exp(-N \times f/2)$,其中N是需要分析的分子数目;f是由单个双胞胎贡献的部分胎儿DNA浓度;并且exp是指数函数。分子数目是用于分析染色体区域的任何基因座处的DNA片段数目。需要分析以实现检测父系单元型的所需确定性的分子数目可通过测量染色体区域中的一个基因座直到表1中列出的数目来取得。作为另外一种选择,如果分析相同染色体区域中的几个基因座,则需要分析的分子数目/基因座可减少至基因座数目乘以分析的平均分子数目/基因座达到表1中列出的分子数目的程度。

[0072]

由单个双胞胎儿贡献的部 分胎儿 DNA 浓度(%)	待分析的分子数目
20	46
15	61
10	92
8	115
6	154
4	230
2	461

[0073] 表1:待分析以对于各种胎儿DNA浓度达到检测父系单元型的小于1%概率的DNA片段数目

[0074] 还可测定需要分析的染色体区域数目。假定染色体区域不处于连锁不平衡,两个二卵双生继承不同父系单元型的机率对于不同染色体区域中的每一个将是50%。如果染色体区域的数目是n,则对于这n个染色体区域中的每一个已继承相同父系单元型的一对二卵双生的概率将是 2^{-n} 。因此,当分析七个独立染色体区域时,对于7个区域中的每一个已继承相同父系单元型的一对二卵双生的概率将小于1%。在这种情况下,在步骤140中的截断可以是14%,其中在七个中显示不同继承单元型的一个区域将提供二卵的分类。还可使用一个区域的绝对值。如果使用大量区域(例如50或100个),则可允许一个或多个区域指示不同的继承单元型,同时仍提供单卵的分类。

[0075] II. 使用亲本基因型鉴定胎儿单元型

[0076] 如上所述,在染色体区域的一个或多个基因座处的亲本基因型可用于帮助测定两个胎儿是否已继承来自亲本的不同单元型。例如,可使用对应于母体血浆样品中的相同基因组区域的两个不同父系单元型的检测,所述母体血浆样品得自具有双胞胎妊娠的女性。尽管下文分析集中于基于两个不同父系单元型的检测的例子,但技术的变动也可应用于两个不同的母系单元型。

[0077] a. 在两个基因座处的SNP分析

[0078] 图2A显示了当对于给定染色体区域已知在两个不同基因座处的母亲基因型和父亲单元型时,用于测定接合性的例子。这个实施例集中于双胞胎的妊娠母亲是纯合的并且双胞胎的父亲是杂合的SNP基因座。在图2A中所示的例子中,母亲在SNP基因座1和2处是纯合的,分别具有基因型AA和TT。父亲在SNP基因座1和2处是杂合的,分别具有基因型AC和GT。

[0079] 假定基因座1和2是紧密的,即重组不太可能(例如在两个基因座之间发生重组的概率 $<0.1\%$)。在两个基因座之间发生,在两个基因座处的等位基因将由胎儿一起继承且形成单元型。因为母亲对于基因座1和基因座2是纯合的,所以她具有两个相同的单元型。我们将这两个相同的母系单元型定义为Hap I和Hap II。另一方面,父亲具有两个不同的单元型,并且如图2A中示出的,我们将它们定义为Hap III和IV。

[0080] 当孕妇携带一对单卵双生时,两个胎儿的基因构成将是相同的。换言之,两个父系

单元型中的仅一个将由这两个双胞胎儿继承。在示出例子中,两个胎儿均继承来自父亲的Hap III。

[0081] 当分析母体血浆样品时,仅等位基因A将对于基因座1检测到,因为母亲和两个胎儿对于等位基因A均是纯合的。当待分析的对应于基因座1的分子数目足够大时,母体血浆中的C等位基因的不存在将指示胎儿无一已继承来自父亲的Hap IV。需要分析的分子数目将依赖母体血浆DNA样品中的部分胎儿DNA浓度,和排除母体血浆样品中的C等位基因的存在所需的统计力,例如如表1中所示。

[0082] 另一方面,T和G等位基因将在关于基因座2的母体血浆样品中检测到。因为母亲对于T等位基因是纯合的,这指示胎儿中的至少一个继承来自父亲的Hap III。来自基因座1和基因座2的信息一并考虑,两个胎儿将均已继承来自父亲的Hap III。

[0083] 在二卵双生的情况下,两个双胞胎儿可继承来自父亲的不同单元型。在图2A的例子中,双胞胎1已继承来自父亲的Hap III,并且双胞胎2已继承来自父亲的Hap IV。因此,在母体血浆中,A和C等位基因可对于基因座1检测到,并且G和T等位基因可对于基因座2检测到。可定量在基因座处的等位基因检测,以确保检测不是假的(例如由于分析误差,测量特定类型的仅一个或两个等位基因)。例如,可测量特定类型的等位基因数目且与阈值比较,这确保已测量统计上显著量的特定等位基因。截断可基于对于样品作出的测量数目(例如对于特定基因座测量的等位基因数目)改变。例如,如果对于基因座测量1,000个等位基因,则阈值可大于仅在该基因座处测量100个等位基因的情况。因此,如果测量的量超过阈值,则等位基因可视为检测到。

[0084] 这些发现指示父系Hap III和Hap IV均存在于母体血浆中。因为每个胎儿可仅继承来自父亲的一个单元型,这些发现还指示两个胎儿已继承来自父亲的不同单元型,并且因此它们在遗传上不同。因此,两个胎儿将鉴定为已继承关于包括基因座1和2的染色体区域的不同单元型,并且像这样两个胎儿可测定为一对二卵双生,例如仅使用这个染色体区域或与来自其他染色体区域的数据组合。

[0085] 相应地,当第一亲本是父亲时,第一亲本的两个单元型可对于第一染色体区域在多个基因座处进行测定。例如,对于特定染色体区域可测定Hap III和Hap IV。测定胎儿中的两个已继承来自第一亲本的第一染色体区域的不同单元型可如下进行。可在第一染色体区域中鉴定在该处第一亲本是杂合的第一基因座和第二基因座(例如基因座1和基因座2),并且在基因座1和基因座2处的父系独特的等位基因(即在母体基因组中不表示)不存在于相同父系单元型上。

[0086] 可在生物样品中检测在第一基因座处统计上显著量的第一亲本的第一单元型。如上所述,当母亲对于第一单元型上的第一等位基因(对于基因座1,A)是纯合的,并且父亲对于第二等位基因是杂合的时(其对于基因座1是Hap IV),可实现这点。可检测具有第二等位基因(对于基因座1,C)的DNA片段数目,并且与截断(阈值)值比较,以测定是否已检测到统计上显著量的第一单元型。含有第二等位基因的DNA片段数目(第二等位基因的测量量的例子)可单独使用(例如截断可以是绝对数)或标准化(例如截断可以是比例)。

[0087] 随后可在生物样品中检测在第二基因座处统计上显著量的第一亲本的第二单元型。如上所述,当母亲对于第二单元型上的第三等位基因(对于基因座2,T)是纯合的,并且父亲对于第四等位基因是杂合的时(其对于基因座2是Hap III),可实现这点。可检测具有

第四等位基因(对于基因座1,G)的DNA片段数目,并且与截断(阈值)值比较,以测定是否已检测到统计上显著量的第一单元型。应注意到第三和第四等位基因可再次是A和C,但在Hap III上具有C。

[0088] 相应地,实施例可测定对于染色体区域的第一亲本的第一单元型是否由胎儿中的任一个继承。如果第一单元型已继承,则随后测定对于染色体区域的第一亲本的第二单元型是否由胎儿中的任一个继承。如果第二单元型也已对于染色体区域继承,则随后将胎儿分类为二卵的。上文讨论提供其中第一亲本是父亲的例子,现在提供其中第一亲本是母亲的例子。

[0089] 两个母系单元型的定量分析也可用于测定两个胎儿是单卵的还是二卵的。图2B显示了当对于给定染色体区域已知在两个不同基因座处的父亲基因型和母亲单元型时,用于测定接合性的例子。在其中两个胎儿是单卵的情况下,两个胎儿将对于所有染色体区域继承相同的母系单元型。由胎儿继承的单元型将以更高浓度存在于母体血浆中。两个母系单元型的浓度中的差异与母体血浆中的胎儿DNA的部分浓度成比例。这显示于图2B中,对于基因座1,其中A等位基因通过与20%胎儿DNA浓度成比例的差异比B等位基因更频繁出现。随后,对于基因座2,存在等量的等位基因A和等位基因B,因为父亲对于等位基因B是纯合的。如果胎儿贡献不等百分比的胎儿DNA,则关于基因座1的等位基因A与B的比将依赖总胎儿DNA百分比(即各个胎儿DNA百分比的总和)。如果胎儿贡献不等百分比的胎儿DNA,则关于基因座1的等位基因A与B的比将与总部分胎儿DNA浓度(即各个部分胎儿DNA浓度或胎儿DNA百分比的总和)正相关。在本发明中,术语部分胎儿DNA浓度和胎儿DNA百分比可互换使用。

[0090] 另一方面,当两个胎儿是二卵的时,两个胎儿可继承在任何染色体区域处的不同母系单元型。当两个胎儿继承在特定染色体区域处的不同母系单元型时,两个母系单元型将以等量存在于母体血浆中。因此,在一个或多个染色体区域处的两个母系单元型的等量表示的存在可潜在地用于指示一对二卵双生的存在。这个相等表示将其自身表现为对于基因座1和基因座2在等位基因A和B之间的相同差异,但过度表现的等位基因是不同的。过度表现的等位基因来自不同单元型的事实可用于鉴定胎儿已鉴定母亲的不同单元型。应注意到过度表现的程度是总胎儿DNA百分比的一半(如所示的10%)。此类现象在以后更详细地讨论。如果胎儿贡献不等百分比的胎儿DNA,则关于基因座1的等位基因A与B的比将依赖由双胞胎1贡献的胎儿DNA百分比;并且关于基因座2的等位基因A与B的比将依赖由双胞胎2贡献的胎儿DNA百分比。应注意到在Hap I上的等位基因对于两个基因座无需是相同的。

[0091] 相应地,当母系单元型在第一基因座和第二基因座处是已知的时,方法可检测被检测母亲的不同单元型。如图2B中,父亲对于第一等位基因在第一基因座处是纯合的(如所示的A),并且母亲对于在第一基因座处的第一等位基因和第二等位基因(如所示的B)是杂合的。第一等位基因在母亲的第一单元型上(如所示的Hap I),并且第二等位基因在第二单元型上(如所示的Hap II)。检测在第一基因座处的第一亲本的第一单元型可包括测定在第一基因座处测量的第一等位基因的各自量以统计上显著量超过在第二基因座处测量的第二等位基因的各自量。这由110:90的比显示。在一个实现中,截断值可用于确保在值之间的差异是统计上显著的。

[0092] 对于第二基因座,父亲对于第三等位基因在第二基因座处是纯合的(如所示的B,但可以是任何等位基因包括A),并且母亲对于在第一基因座处的第三等位基因和第四等位

基因(如所示的A)是杂合的。第三等位基因在母亲的第一单元型上,并且第四等位基因在第二单元型上。在生物样品中检测在第二基因座处的第一亲本的第二单元型可包括测定在第二基因座处测量的第四等位基因的各自量以统计上显著量超过在第二基因座处测量的第三等位基因的各自量。

[0093] b.关于其他类型的多态性的分析

[0094] 上文例子涉及在基因座处的不同核苷酸,所述基因座用于测定对于染色体区域不同单元型是否已从第一亲本继承。然而,SNP以外的多态性也可用作不同父系单元型的标记。其他类型的多态性的例子包括但不限于微卫星、限制性片段长度多态性、插入/缺失多态性和拷贝数变异(CNV)。此类其他多态性可导致图2A和2B的构型,但其中第一等位基因和第二等位基因起因于这些其他多态性之一。对于任何多态性,还可存在超过两个等位基因。下文是例子。

[0095] 图3还显示了其中基因座具有四个不同等位基因的微卫星分析的例子。微卫星是包含可变数目的短串联重复(STR)的基因组中的多态性区域。在这个例子中,母亲对于这个微卫星区域具有两个不同等位基因,即等位基因I和等位基因II,分别包含4和7个重复。父亲也具有两个不同等位基因,即等位基因III和等位基因IV,分别包含5和8个重复。因此,基因座实际上在亲本中具有四个不同等位基因。

[0096] 因为一对单卵双生的基因构成将是相同的,所以它们将已继承相同父系等位基因。因此,仅一个父系等位基因可在母体血浆中检测到。在这个例子中,仅具有5个重复的父系等位基因III连同具有4个重复的母系等位基因I一起可在母体血浆样品中检测到。当对于给定染色体区域分析多个基因座并且对于基因座中的每一个仅一个父系等位基因可在母体血浆样品中检测到时,随后将如上所述在统计上测定双胞胎是单卵的概率。

[0097] 另一方面,当该对双胞胎是二卵的时,两个双胞胎儿可继承不同父系等位基因。如图3中示出的,二卵双生之一已继承来自父亲的等位基因III,并且另一个已继承等位基因IV。因此,两个父系等位基因均可例如经由上文描述的检测方法在母体血浆样品中检测到。换言之,如果两个父系等位基因均可同时在母体血浆样品中检测到,则双胞胎儿将是二卵的,除非胎儿具有染色体异常(例如胎儿在基因座处是三体的)。如果母亲在该基因座处是纯合的,例如对于为4个的重复(或多态性的任何其他等位基因),则可执行相似分析。

[0098] 因此,在其中多态性具有三个或更多个等位基因的情况下,如果两个父系继承的等位基因(两者在孕妇的基因组中均不存在)在母体血浆中检测到,则可检测二卵双生妊娠。在一个实施例中,母体血浆中的此类基因型模式由双胞胎妊娠的存在的超声证据支持。在不存在双胞胎妊娠的此类超声证据的情况下,此类血浆基因型模式将指示三体胎儿的存在(Ghanta S等人PLoS ONE2010;5:e13184)。

[0099] III. 鉴定基因组区域中的变化

[0100] 先前部分描述了用于测定染色体区域的不同单元型是否从第一亲本继承的示例技术。在此类例子中,亲本基因型信息对于两个亲本是已知的并且在分析中使用。在下述说明中,不需要亲本的基因型,尽管它可使用。例如,胎儿DNA浓度(或其他参数)将显示对于二卵胎儿在多个基因座处的不同值。

[0101] A. 一般方法

[0102] 图4是示出根据本发明的实施例的方法400的流程图,所述方法400用于分析具有

多个胎儿的妊娠女性的生物样品,以测定妊娠女性的至少两个胎儿对于第一染色体区域是否是遗传上不同的。与其他方法一样,生物样品包括胎儿和母体DNA。方法400可用于执行方法100的步骤130。

[0103] 在步骤410中,妊娠女性的基因型在第一染色体区域内的一个或多个第一基因座中的每一基因座处进行测定。第一基因座是这样的:妊娠女性在一个或多个第一基因座中的每一基因座处是纯合的,或在一个或多个第一基因座中的每一基因座处是杂合的。相应地,第一基因座中的每一个对于母系基因型具有相同种类(即纯合或杂合的)。基因型可以多种方法进行测定。例如,可分析来自妊娠女性的全血的血沉棕黄层或细胞团块层,其中占优势地仅存在母体DNA或其他仅母体样品。可使用标准基因型技术。作为另一个例子,可由包括胎儿和母体DNA的生物样品的分析推导基因型(此类技术在下文更详细地描述)。

[0104] 第一基因座中的每一个显示出主要等位基因和次要等位基因,即生物样品含有关于基因座中的每一个的主要等位基因和次要等位基因。当母亲在第一基因座处是纯合的时,次要等位基因由至少一个胎儿贡献。在此类情况下,每个基因座将具有主要等位基因和次要等位基因,其中主要等位基因比次要等位基因更丰富。当母亲是杂合的并且至少一个胎儿是纯合的时,主要等位基因也比次要等位基因更丰富。然而,当母亲是杂合的并且所有胎儿也是杂合的时(对于相等等位基因),主要等位基因和次要等位基因均不是更丰富的。基因座各自具有主要等位基因和次要等位基因的事实可以多个方法进行测定,例如通过检测等位基因(参见步骤420)或亲本基因型的了解(推导或明确的)。

[0105] 在步骤420中,可在生物样品中检测在第一基因座中的每一基因座处各自的主要等位基因和/或各自的次要等位基因。对于其中仅主要等位基因或仅次要等位基因检测到的实施例,其他等位基因的存在了解可通过关于亲本基因型的一些信息获得。例如,如果检测到次要等位基因B并且母亲已知在基因座处对于主要等位基因A是纯合的,则随后可测定生物样品在该基因座处具有两个等位基因。此类例子是下述情况:其中母亲是RhD阴性的(并且因此关于等位基因纯合通过RHD基因的不存在表示),并且RHD基因在生物样品中检测到。还可检测在每个基因座处的两个等位基因。

[0106] 如果主要和次要等位基因已知在基因座处存在,则随后可鉴定胎儿基因型,以有效地具有相似种类。例如,如果母亲对于在基因座之一处的等位基因A是纯合的,则随后等位基因A将是主要等位基因。另外,因为将检测到次要等位基因B,所以已知胎儿中的至少一个是杂合的。因此,所有第一基因座将具有种类 AA_{AB} ,其中下标指示胎儿中的至少一个是杂合的。对于其中母亲是杂合AB并且胎儿中的至少一个是纯合AA的例子,随后B再次将是次要等位基因,并且A将是主要等位基因。在这种情况下,所有第一基因座将具有种类 AB_{AA} ,其中下标指示胎儿中的至少一个是纯合的。

[0107] 在步骤430中,在生物样品中在一个或多个基因座处测量一个或多个主要等位基因的第一个量和/或一个或多个次要等位基因的第二个量。每个基因座可具有不同的主要等位基因,但可合并(例如总计)每个主要等位基因的量,以获得第一个量。这可同样对于第二个量完成。在一个实施例中,仅测定第一个量。在另一个实施例中,仅测定第二个量。在另外一个实施例中,测定第一个和第二个量。步骤420和430可同时完成,并且因此有效地是相同步骤。

[0108] 在步骤440中,获得对于第一个量或第二个量的标准化参数。在一个实施例中,通

过计算对于第一区域的部分参数值(例如部分胎儿DNA浓度)获得标准化参数,其中部分(fraction)是相对于第二个量的第一个量。在一个方面,部分胎儿DNA浓度是表观值,因为当两个胎儿在第一染色体区域中是遗传上不同的时(例如仅一个是杂合的),它可不同于实际胎儿DNA浓度。在另一个实施例中,标准化参数以校准方式即以相同方式或相关方式获得作为期望值,其在下文描述。相关方式可以是重现性地提供具有固定差异或比的值的任何程序(例如一种技术有规则地提供另一种技术的值的1.2倍,或存在关于将得自一种技术的结果转换为另一种技术的期望值的转换曲线。因此,标准化参数可仅仅是第一个量(或第二个量),如果它以校准方式获得。

[0109] 标准化参数也可通过计算来自另一个染色体区域内的一个或多个基因座的一种或多种其他序列(例如等位基因或纯合序列)的第三个量进行测定,并且随后使用第三个量标准化第一个量或第二个量。此类标准化(以及其他类型的标准化)可允许使用不同技术或未校准的技术对于其他区域计算的量的比较。对于校准技术,第一个量或第二个量可与来自其他区域的测量比较。使用第一个量或第二个量可传递相同信息。

[0110] 在步骤450中,比较标准化参数与截断值,以测定如果胎儿对于第一染色体区域是遗传上相同的,则标准化参数是否统计上不同于期望值。例如,当所有胎儿均为杂合或纯合的时,如果仅胎儿中的一个与数目比较是杂合的,则次要等位基因的数目将是不同的。期望值(例如胎儿DNA浓度)可得自生物样品的测量,例如来自其他染色体区域,定量染色体Y序列的量或使用一种或多种后生标记,如下文更详细地描述。其他期望值可源自胎儿DNA浓度的此类测量或可直接测量,并且因此期望值并不限于预期的胎儿DNA浓度。

[0111] 截断值可基于所需准确度进行选择。例如,可已知期望值测量的标准差。随后,截断可选择为期望值减去(或当情况允许时加上)三倍标准差(SD)。以这种方式,实施例可基于与截断值的相关测定标准化参数与期望值的差异(或比与1的偏差)是统计上显著的。在其他实施例中,截断可以是2.0、2.5、3.5或4个SD。

[0112] 在步骤460中,基于标准化参数与截断值的比较,测定妊娠女性的两个胎儿是否是二卵的。例如,可测定当标准化参数统计上不同于期望值时,妊娠女性的两个胎儿对于第一染色体区域是遗传上不同的。如果标准化参数与期望值并无统计上不同,则随后胎儿可测定为单卵的。染色体区域的此类测定可与来自其他染色体区域的用于测定接合性的测量(例如如上文对于步骤150描述的)组合。作为另外一种选择,对于第一染色体区域不是相同的两个胎儿的测定可提供二卵的分类,而无需其他染色体区域的测量。

[0113] B. 表观部分参数(胎儿DNA浓度)

[0114] 如上所述,步骤440中的标准化参数可以是部分值(fractional value),包括第一个量和第二个量。图5是示出方法500的流程图,所述方法500用于通过测定关于来自第一亲本的第一染色体区域的表观部分参数(例如胎儿DNA浓度)测定妊娠女性的至少两个胎儿对于该第一染色体区域是否是遗传上不同的。方法500分析具有多个胎儿的妊娠女性的生物样品,如方法400中。尽管方法500描述的是其中母亲在第一基因座处是纯合的情况,但该方法可同样对于其中母亲是杂合的基因座执行。

[0115] 在步骤510中,鉴定在第一染色体区域内的一个或多个基因座,在所述基因座处妊娠女性是纯合的并且胎儿中的至少一个是杂合的。在此类情况下,每个基因座将具有主要等位基因和次要等位基因,其中主要等位基因比次要等位基因更丰富。例如,如果母亲对于

等位基因A是纯合的,那么则随后等位基因A将是主要等位基因。作为另外一种选择,可鉴定一个或多个基因座,在所述基因座处妊娠女性是杂合的并且胎儿中的至少一个是纯合的,其中对于胎儿纯合的等位基因将是主要等位基因。

[0116] 此类基因座可通过分析起因于测量步骤例如步骤110的等位基因进行鉴定。例如,在特定基因座处的等位基因的测量可显示大多数例如大于70%是等位基因A(主要等位基因),并且仅一个另一等位基因(例如T)以显著量计数。此类测量可指示母亲对于等位基因A是纯合的,并且胎儿中的至少一个对于等位基因T是杂合的。不同截断(例如更高部分的次要等位基因)可用于测定母亲是否是杂合的并且至少一个胎儿是否是纯合的。可基于假定或测量的胎儿DNA浓度测定用于测定基因座是否满足上文条件的多个截断值。美国专利申请号12/940,993对于具有一个胎儿的妊娠更详细地描述了此类技术。下文的其他部分提供了用于获得在多胎妊娠中的基因座的基因型信息的更复杂程序。

[0117] 在步骤520中,在生物样品中在一个或多个基因座处测量一个或多个主要等位基因的第二个量q。每个基因座可具有不同的主要等位基因,但可合并(例如总计)每个主要等位基因的量,以获得第二个量。以这种方式,关于染色体区域的分析的覆盖深度是基本上增加的,因为使用多个基因座。例如,可对于每个基因座测量有限量的等位基因,但一旦基因座以聚集体显现,就可测量足够数目的等位基因,以获得统计坚固性。可选择在染色体区域内的基因座,以使得在基因座之间的重组概率很低,例如小于1%。

[0118] 在步骤530中,在生物样品中在一个或多个基因座处测量一个或多个次要等位基因的第二个量p。以相似方式,每个基因座可具有不同的次要等位基因,但可合并(例如总计)每个次要等位基因的量,以获得第二个量。次要等位基因可起于基因座中的不同单元型。

[0119] 在步骤540中,由第一个量和第二个量测定第一参数。第一参数提供在第一个量和第二个量之间的相对量。例如,第一参数可以是如下计算的部分胎儿DNA浓度F的度量: $F = 2p/(p+q)$ 。其他例子包括两个量的任何比,第一个量的函数比,或此类比的函数。另一个例子是可以是可标准化的两个量之间的差异。

[0120] 在步骤550中,比较第一参数与截断值,以测定如果两个胎儿在第一染色体区域中是遗传上相同的,则第一参数是否在统计上不同于期望值。例如,如果两个胎儿对于第一染色体区域中的基因座处的主要和次要等位基因是杂合的,则随后当使用上式时,第一参数应等于总胎儿DNA浓度。然而,如果胎儿之一在基因座处也是纯合的,则随后胎儿DNA浓度将小于总胎儿DNA浓度。预期的胎儿DNA浓度可得自生物样品的测量,例如来自其他染色体区域,如果两个胎儿均已知是男性(例如通过超声扫描)则测量染色体Y序列的量,或使用一种或多种后生标记,两者均在下文更详细地描述。其他期望值可源自胎儿DNA浓度的此类测量或可直接测量,并且因此期望值并不限于预期的胎儿DNA浓度。

[0121] 在步骤560中,基于第一参数与截断值的比较,测定妊娠女性的两个胎儿是否是二卵的。例如,可测定当第一参数统计上不同于期望值时,妊娠女性的两个胎儿在第一染色体区域中是遗传上不同的。如果标准化参数与期望值并无统计上不同,则随后胎儿可测定为单卵的。当第一参数是部分胎儿DNA浓度时,第一参数可以是表观部分胎儿DNA浓度,并且不是实际胎儿DNA浓度。使用术语表观是因为当两个胎儿是二卵的时,部分胎儿DNA浓度的计算可不同,并且因此计算的值不是实际部分胎儿DNA浓度。

[0122] 相应地,根据下式可测定在多态基因座(例如SNP基因座)处的母体血浆样品中的表观部分胎儿DNA浓度(F),所述多态基因座在母亲中是纯合的并且对于双胞胎中的至少一个是杂合的:

[0123] $F=2p/(p+q)$,

[0124] 其中p是携带胎儿特异性等位基因的DNA分子的计数,并且q是携带由胎儿和母亲共享的等位基因的DNA分子的计数。适合于表观部分胎儿DNA浓度分析的SNP基因座是其中两个不同等位基因对于个别SNP基因座在母体血浆中检测到的那些,如本申请的先前部分中描述的。此类基因座可具有其中母亲是纯合或杂合的类型,但当比较时区域应具有相同类型。

[0125] 对于部分胎儿DNA浓度的计算,可使用来自一个SNP基因座或多个SNP基因座的信息。来自多个SNP基因座的红色计数可总计在一起。换言之,将整个基因组的SNP基因座分成“组”。一组的基因座应优选在同一邻接DNA段中。在一个实施例中,此类组涉及在相同染色体臂上的SNP。另一个实施例中,此类组涉及在1kb、5kb或10kb DNA段内的SNP。在每组内的SNP基因座数目将依赖部分胎儿DNA浓度的测量和每个SNP基因座的覆盖深度(即检测且定量每个SNP基因座的数目)的所需精确度。在一个方面,部分浓度的测量的所需精确度将是允许测定两个不同组的SNP基因座是否具有相同或至少两个不同的表观部分胎儿DNA浓度的最小精确度。

[0126] 对于每个SNP基因座的更高覆盖深度,可减少精确地测量部分胎儿DNA浓度所需的SNP数目。在一个实施例中,关于目的SNP基因座的覆盖深度可通过靶标测序方法增加(Liao GJ等人Clin Chem2011;57:92-101.)。在这种方法中,血浆DNA分子首先与探针杂交,所述探针与目的区域中的序列互补,并且使捕获分子进行大规模平行测序。关于所需分子数目的计算将在以后部分中讨论。

[0127] C.使用关于不同区域的量作为期望值

[0128] 如上所述,可由另一个染色体区域的分析测定期望值。例如,可对于第二染色体区域测定第二参数(例如以与上文描述相似的方式),并且此类值可用作期望值。在对于第二染色体区域的基因座处(例如当母亲是纯合的A:A并且父亲已知是T:T时)所有胎儿均为杂合例如A:T的情况下,随后得自两个等位基因的两个量的第二参数可提供关于总胎儿DNA浓度的期望值。随后,当关于第一区域的第一参数统计上不同于期望值时,如对于第二区域测定的,数据暗示仅一个胎儿在第一染色体区域中的基因座处是杂合的,并且胎儿是遗传上不同的。其他技术可用于测定期望值,如下一个部分中解释的。

[0129] 还可测定其他染色体区域的参数,并且每一个均可用于方法400的多次迭代中。多个实施例可将测定参数分组成簇并且测定簇是否彼此不同,如以后更详细地描述的。此类簇可首先滤过,以确保使用母亲的相似区域,例如以确保母亲在参数值被聚簇的区域处是纯合的。随后,如果参数值的两个或更多个簇在数据中出现,则可推测一些区域在胎儿之间是相同的,但其他区域不同,从而指示二卵。相应地,在母体血浆中的胎儿DNA部分中的区域基因组变化可揭示双胎妊娠的接合性。在下文例子中,使用表观部分胎儿DNA浓度分析,但还可使用其他参数。

[0130] 一对双胞胎是单卵还是二卵的测定可通过使用多个SNP基因座分析表观部分胎儿DNA浓度来实现。表观胎儿DNA浓度是使用胎儿特异性遗传标记在特定区域处测量的胎儿

DNA浓度。当胎儿中的至少两个是二卵的时,在特定区域处的表观胎儿DNA浓度可不同于实际胎儿DNA浓度。对于单卵双生,表观部分胎儿DNA浓度将在整个全基因组的不同SNP基因座处是相似的。然而,对于二卵双生,由于两个胎儿的基因型中的差异,表观部分胎儿DNA浓度将显示双峰或三峰分布。在此类方法中,不需要关于父亲和/或母亲的单元型信息,尽管也可使用。

[0131] 使用表观胎儿DNA浓度的例子

[0132] 图6和7示出了根据本发明的实施例,使用母体血浆中的胎儿DNA部分中的区域基因组变化揭示双胎妊娠的接合性。图6显示了其中双胞胎是单卵的例子,并且图7显示了其中双胞胎是二卵的例子。显示了在两个SNP基因座(基因座1和基因座2)处的父亲、母亲和该对双胞胎(单卵或二卵)的基因型。这两个基因座均来自不同染色体区域,并且因此是两个不同组的部分,其中使用上文术语。

[0133] 为了示出目的,我们假定两个双胞胎各自将贡献10%的母体血浆DNA,并且在母体血浆样品中存在总共100个基因组当量的DNA。一个基因组当量是在整倍体人细胞中含有的DNA量。在其他情况下,两个胎儿各自对生物样品(例如母体血浆)贡献不同量的DNA。

[0134] 对于单卵双生(如图6中所示),它们的遗传组成是相同的。因此,两个双胞胎在基因座1和基因座2处均为杂合的,并且适合于通过定量母体血浆样品中的共享和胎儿特异性等位基因测定表观部分胎儿DNA浓度的任何其他基因座。对于基因座1,携带A等位基因的180个分子和携带T等位基因的20个分子存在于母体血浆中,因此产生20%的表观部分胎儿DNA浓度 $[20 \times 2 / (180 + 20)]$ 。对于基因座2,还通过定量G和C等位基因检测到20%的表观部分胎儿DNA浓度。因为两个双胞胎的遗传构成是相同的,当双胞胎1在特定SNP基因座处是杂合的时,双胞胎2也将在相同SNP基因座处是杂合的。因此,在整个全基因组任何SNP基因座处测量的表观部分胎儿DNA浓度将是相似的。

[0135] 对于二卵双生,关于至少部分基因组的基因型将是不同的。如图7中示出的,双胞胎1和双胞胎2在基因座1处均为杂合的,因此给出20%的表观部分胎儿DNA浓度。另一方面,双胞胎1对于在基因座2处的G等位基因是纯合的,而双胞胎2是杂合的。因此,基于G和C等位基因的数量,在基因座2处的表观部分胎儿DNA浓度是10% $[10 \times 2 / (10 + 190)]$ 。换言之,当两个胎儿在基因座处均为杂合的时,表观部分胎儿DNA浓度将是20%。然而,当仅一个胎儿在基因座处是杂合的时,表观部分胎儿DNA浓度将是减少的(例如如果通过双胞胎的每个成员释放的胎儿DNA的量是相等的,则为10%)。因此,当表观部分胎儿DNA浓度在多个SNP基因座处进行测量时,当两个胎儿贡献等量的胎儿DNA时,将存在双峰分布。双胞胎的每个成员将不同量的DNA释放到母体血浆内是可能的。在这种情况下,当分析母体血浆DNA时,将可见部分胎儿DNA浓度的三峰分布,这在下文讨论。

[0136] D. 用于测量期望值的其他技术

[0137] 如上所述,可由在其他染色体区域处的遗传测量测定期望值(例如实际部分胎儿DNA浓度)。用于测量这个变化的另一种方法是比较使用遗传标记测定的胎儿DNA的部分浓度(或一些其他参数)与使用另一种方法测量的那种,例如基于性质不是遗传性的标记的那种。因此,还可通过其他测量来测定期望值(例如来自所有胎儿的总部分胎儿DNA浓度)。一个实施例涉及生物样品中的胎盘特异性后生标记例如DNA甲基化的量的测量。

[0138] 在一个实施例中,可使用后生标记测量胎儿DNA的部分浓度。在一种实现中,后生

标记可以是DNA甲基化标记。胎儿DNA甲基化标记的一个例子是在血浆中显示出源自胎儿的和源自母体的DNA之间的差异DNA甲基化模式的标记(美国专利6,927,028)。此类标记的一个例子是编码乳腺丝抑蛋白(maspin)的SERPINB5基因,其在母体血浆中存在的胎儿DNA中是低甲基化的,但对于母体血浆中的母体DNA是超甲基化的(美国专利8,026,067)。另一个例子是RASSF1A基因,其在母体血浆中存在的胎儿DNA中是超甲基化的,但对于母体血浆中的母体DNA是低甲基化的(美国专利7,754,428)。其他DNA甲基化标记在其他地方描述(Papageorgiou EA等人Am J Pathol2009;174:1609-1618)。

[0139] 此类DNA甲基化标记可使用本领域技术人员众所周知的许多技术在母体血浆中进行检测,包括甲基化特异性PCR(Herman JG等人Proc Natl Acad Sci USA1996;93:9821-9826)、实时甲基化特异性PCR(Lo YMD等人Cancer Res1999;59:3899-3903)或MethyLight(Eads C等人Nucleic Acids Res2000;28:E32)、亚硫酸氢盐测序(Frommer M.Proc Natl Acad SciUSA1992;89:1827-1831)、甲基化敏感的限制性酶消化(Chan KCA等人Clin Chem2006;52:2211-2218)、甲基-BEAMing(Li M等人Nat Biotechnol2009;27:858-863)、和大规模平行测序(Komori HK等人Genome Res2011;21:1738-1745)。

[0140] 作为例子,使用RASSF1A作为分子标记,可通过测量母体血浆中的超甲基化RASSF1A序列在总体(即超甲基化的加上低甲基化的)RASSF1A序列上的比例,计算母体血浆中的部分胎儿DNA浓度。对于单卵双生妊娠,使用跨越基因组不同部分的一个或一系列胎儿遗传标记测量的母体血浆中的部分胎儿DNA浓度应与使用如这段第一句中所述的RASSF1A DNA甲基化标记系统测量的浓度具有紧密关联。然而,对于二卵双生妊娠,使用胎儿遗传标记测量的部分胎儿DNA浓度之间的关联应显示出与RASSF1A DNA甲基化标记系统的更弱关联。在一个实施例中,可使用一系列这些类型的标记中的每一个分析在遗传和DNA甲基化标记系统之间的关联。例如,关联可使用皮尔森相关或线性回归进行检查。其他类型的后生标记包括基于组蛋白修饰例如甲基化和乙酰化的那些。

[0141] 另一个实施例是测量在胎儿基因组中存在但在母体基因组中不存在的遗传序列的量。此类遗传序列的例子包括关于妊娠的RHD基因,其中母亲是RhD-阴性的,并且父亲对于RHD是纯合阳性的。因此,如果所有胎儿均为RhD-阳性的,并且母亲是RhD-阴性的,RHD基因可用于测定实际胎儿DNA浓度。另一个例子是仅涉及男性胎儿的多胎妊娠中的染色体Y序列的量的测量。例如,如果所有胎儿均为男性,则随后可使用在Y染色体上的基因座测量实际胎儿DNA浓度。

[0142] 因此,通过测量具有选自一种或多种胎儿特异性序列的胎儿特异性序列的DNA片段的第三个量,一个实施例可使用遗传标记计算期望值(例如胎儿DNA浓度)。因为计算具有期望值(例如胎儿DNA浓度),如果胎儿是遗传上相同的,则所有胎儿均具有胎儿特异性序列(例如上文例子之一)。例如经由本文描述的方法获得关于第三个量的标准化值,以测定标准化参数。标准化值随后可用作期望值(例如胎儿DNA浓度)。一种实现鉴定一个或多个第二基因座,在所述基因座处胎儿具有各自的第一等位基因,并且母亲不具有各自的第一等位基因。胎儿特异性序列随后是各自的第一等位基因。可通过测量在一个或多个第二基因座处的等位基因总量,并且由第三个量与总量的比计算胎儿DNA浓度,获得关于第三个量的标准化值。

[0143] E. 非多态性胎儿特异性序列

[0144] 在另一个实施例中,非多态性胎儿特异性序列可用于测量表观部分胎儿DNA浓度。例如,可就非胎儿特异性序列例如编码瘦素的LEP基因而言测量母体血浆样品中的染色体Y序列的量。染色体Y与LEP DNA比可与期望值比较,所述期望值例如如使用胎盘特异性DNA甲基化标记测量的总部分胎儿DNA浓度。如果部分染色体Y浓度不同于期望值,则它暗示妊娠涉及至少一组二卵胎儿,并且至少一个胎儿是男性,以及至少一个胎儿是女性。除了部分染色体Y量外,在母体血浆样品中的染色体Y序列的量的绝对值也可用作测定接合性的量度,如以后在下文描述的。

[0145] 相应地,可在生物样品中在一个或多个第一基因座处测量一种或多种胎儿特异性序列(其可包括非多态序列)的第一个量。可获得对于第一个量的标准化参数,并且随后如本文描述的用于测定胎儿中的至少两个是否是二卵的。

[0146] F. 绝对胎儿DNA浓度

[0147] 作为部分胎儿DNA浓度的替代,可使用关于染色体区域的次要等位基因的测量(例如计数)的绝对值,只要执行一些校准(绝对标准化)。此类用途在不计算明确部分的意义上称为“绝对的”。例如,如果可校准测量步骤以使得来自区域的一定量的DNA片段可得到控制或从一个实验关联到另一个,主要等位基因的量无疑可测定为总数减去关于次要等位基因的第二个量。除了了解关于区域(即对于该区域的特定基因座)的DNA片段的量(例如数目)的绝对值外,DNA片段从一个区域到另一个的固定比可提供校准。例如,可调用以特定方式制备样品的方案,使得来自第一区域的DNA片段总数目是关于第二区域的DNA片段总数目的数目的1.4倍。这个定标因子随后可用作比较的部分。在一个实施例中,已知量的DNA或其他类型的校准物可加入生物样品中。

[0148] 以这种方式,来自不同染色体区域的次要等位基因(或其他胎儿特异性遗传标记)的量可直接彼此比较。在此类情况下,参数仍使用主要等位基因的第一个量有效计算,但此类值由于校准而消失(即第一个量,或第一个量加上第二个量是相同的,并且因此它们从等式中抵消)。例如,除了计数数字PCR结果或分析测序数据外,标记也可通过适当构建的校准曲线进行测量,使得如果靶标浓度是相同的,则标记将进行校准以给出相同的定量读数。以相似方式,来自不同染色体区域的主要等位基因的第一个量可彼此比较,并且此类变化可对于次要等位基因的量提供本文描述的相同结果,因为第一个量将与第二个量中的方差相反改变(例如对于二卵胎儿将可见第一个量中的两个峰,与图8比较)。

[0149] 相应地,对于给定组的此类胎儿特异性遗传标记,它们在单卵和二卵双生妊娠中的关联将是不同的。在一个实施例中,标记使用数字PCR进行定量,在所述情况下它们将在单卵双生妊娠中都给出相同或相似的测量量。相反地,在二卵双生妊娠中,此类标记的比例将给出更分散的结果。为了示出起见,假定二卵妊娠中的两个胎儿均将相同浓度的胎儿DNA释放到母体血浆内(尽管不一定是这种情况)。随后对于其中两个胎儿均已继承在母亲基因组中不存在的相同父系继承的等位基因的第一标记,测量的第一标记(例如当母亲是纯合的时,次要等位基因)的量应是通过第二标记测量的量的两倍,在第二标记中仅胎儿中的一个已继承在母亲基因组中不存在的父系继承的等位基因。对于各自释放不同量的胎儿DNA的二卵双生,测量的第一标记的量将超过第二标记的量,但两个值将不相差2倍。

[0150] 作为校准程序的例子,使用数字PCR分析,可使用相同的母体血浆DNA样品和以相同平均模板分子浓度/数字PCR反应,执行关于在基因座1处的胎儿特异性等位基因的定量

和关于在基因座2处的另一种胎儿特异性等位基因的定量。在此类情况下,非胎儿DNA的本底量对于基因座1和基因座2是相同的。因此,可仅仅计数对于在基因座1处的胎儿特异性等位基因阳性的数字PCR孔数目,以测定在基因座1处的胎儿DNA的表观绝对量。可在基因座2处执行相同过程,以测定胎儿DNA的表观绝对量。两个绝对值随后可彼此比较(例如其中一个用作期望值)或与以另一种方式(例如如上所述)测定的期望值比较,以测定统计上显著的差异是否存在。

[0151] G.检测胎儿DNA浓度中的差异

[0152] 图8显示了根据本发明的实施例,对于贡献等量的胎儿DNA的二卵双生的胎儿DNA浓度的示例直方图800。在直方图中,横(X)轴是胎儿DNA浓度。对于染色体区域测量的胎儿DNA浓度(绝对或部分)可用于增大关于包括测量值的范围的计数器。因为当仅胎儿中的一个具有胎儿特异性等位基因时,可见第一个峰810对应于对于基因座测量的胎儿DNA浓度,而当两个胎儿均具有胎儿特异性等位基因时,第二个峰820对应于对于基因座测量的胎儿DNA浓度。峰820对应于实际胎儿DNA浓度,而峰810对应于表观胎儿DNA浓度。因为可见两个峰,所以可推测胎儿是二卵的。在一个方面,峰的相对高度可用于测定接合性,例如作为步骤150的部分。

[0153] 直方图800还可帮助示出其中实际胎儿DNA浓度使用除了由多个染色体区域处的参数值制备直方图外的技术测量的方法。例如,实际胎儿DNA浓度可测量为2F%(例如使用后生方法)。随后,如果分析属于峰810的区域,则关于表观DNA浓度的测量参数应出现在峰810附近,这将是远离2F%值的统计上显著距离。

[0154] 峰的位置是一组标准化参数(这个例子中为胎儿DNA浓度)的统计值的一个例子,其中所述组是关于曲线815和820之一的数据点(即标准化参数的计数)。例如,峰820是关于由曲线825表示的标准化参数组的统计值,并且峰820可用作期望值。应注意到曲线815和825(当与潜在(underlying)数据拟合时)是概率分布和混合模型的组分的例子,如以后将讨论的。

[0155] 测定从两个不同部分胎儿DNA浓度区分出一个需要多少读数计数可如下进行。假定两个二卵双生各自将贡献F%的母体DNA。将存在SNP基因座的两个群体(簇),显示在母体血浆中的两个不同部分胎儿DNA浓度。在一个SNP群体中,表观部分浓度是F%,而在另一个SNP群体中,表观部分浓度是2F%。先前组的SNP基因座的分布标准差(SD)将是

$\sqrt{\frac{F\%(1-F\%)}{N}}$, 其中N是与这组SNP基因座比对的读数总数目。后面一组SNP基因座的SD也近似等于 $\sqrt{\frac{F\%(1-F\%)}{N}}$ 。

[0156] 如果 $2F\% - F\% > 4x\sqrt{\frac{F\%(1-F\%)}{N}}$, 则两个SNP基因座群体可以小于5%重叠区分。

因此, $N > 16(1-F\%)/F\%$ 。

[0157] 表2显示了对于F%的多个值,区分起因于两个SNP基因座群体的两个表观部分胎儿DNA浓度(F%与2F%相比较)所需的序列读数数目。

[0158]

部分胎儿DNA浓度(F)(%)	所需的序列读数数目
1	1584
2	784
3	517
4	384
5	304
6	251

[0159]

7	213
8	184
9	162
10	144
15	91
20	64
25	48
30	37

[0160] 表2

[0161] H.关于每个胎儿的不同峰

[0162] 可能双胞胎的每个成员将不同量的DNA释放到母体血浆内。在这种情况下,当分析母体血浆DNA时,可见部分胎儿DNA浓度的三峰分布。三个峰中的两个将代表由两个个体双胞胎各自贡献的部分胎儿DNA浓度,而第三个峰将代表总计的两个胎儿的部分胎儿DNA总和。因此,实施例还可提供由此可推导由双胞胎的每个成员释放的DNA相对量的方法。

[0163] 由每个双胞胎释放的DNA量中的大偏差可能与不利结果例如双胞胎之一的危急死亡相关,这是可能的。可随着时间过去跟踪每个胎儿的胎儿DNA的不同贡献,以监控胎儿的健康。由每个双胞胎释放的相对DNA量的另一个效用将是使用母体血浆DNA的大规模平行测序用于检测胎儿染色体非整倍性,例如21三体(Chiu RWK等人Proc Natl Acad Sci USA2008;105:20458-20463;Fan HC等人Proc Natl Acad Sci USA2008;105:16266-16271;Sehnert AJ等人Clin Chem2011;57:1042-1049;Sparks AB等人Am J Obstet Gynecol2012;doi:10.1016/j.ajog.2012.01.030)。部分胎儿DNA浓度是此类方法的诊断灵敏度中的重要参数。

[0164] 因此,如果双胞胎可显示为单卵的,则随后可基本上仅使用相同算法用于处理关于胎儿21三体的非侵入性检测的大规模平行测序数据。另一方面,如果情况涉及二卵双生,则随后可首先测量由每个双胞胎贡献的胎儿DNA的相对比例;并且随后察看已将更少量的DNA释放到母体血浆内的双胞胎的部分胎儿DNA浓度是否可在使用的测序深度处检测到。需要时,可增加测序深度。换言之,实施例允许甚至对于双胎妊娠执行胎儿染色体非整倍性筛选。

[0165] 作为示出,假定在特定二卵双生妊娠中,胎儿1和胎儿2分别贡献妊娠母亲的血浆中的3%和2%DNA。假定希望使用母亲血浆的大规模平行测序执行21三体检测(Chiu RWK等人Proc Natl Acad Sci USA2008;105:20458-20463)。使用实施例将允许我们测定3%和2%的

部分胎儿DNA浓度,以及由两个胎儿一起贡献的部分胎儿DNA浓度,即5%。如果部分胎儿DNA浓度是2%,则允许强21三体检测需要完成的测序深度将是足以检测21三体胎儿的那种。在所需的测序深度和胎儿DNA的部分浓度之间的关系先前已得到报道(Fan HC等人PLoS ONE2010;5:e10439;Chiu RWK等人BMJ2011;342:c7401)。相似考虑还可应用于13三体、18三体(Chen EZ等人PLoS ONE2011;6:e21791;Palomaki GE等人Genet Med2012;doi:10.1038/gim.2011.73)、性染色体非整倍性(Lau TK等人J Matern Fetal Neonatal Med2011;doi:10.3109/14767058.2011.635730)、染色体易位(Lun FM等人Clin Chem2011;57:917-919)和染色体少量缺失(Peters D等人N Engl J Med2011;365:1847-1848)的产前检测。

[0166] 相应地,部分胎儿DNA浓度的计算对于产前诊断的其他应用可以是必需的。例如,通过母体血浆分析的非侵入性产前诊断的准确度可依赖在母体血浆样品中预期评估的由胎儿贡献的DNA部分浓度。对于染色体非整倍性的产前检测,母体血浆中的受累染色体的另外染色体剂量与部分胎儿DNA浓度成比例。由两个胎儿各自贡献的部分胎儿DNA浓度可使用上文描述的技术如下进行测定。

[0167] 图9显示了根据本发明的实施例,当两个胎儿对母体血浆样品贡献不同量的DNA时,基于SNP分析的部分胎儿DNA分布的直方图900。观察到部分胎儿DNA浓度的三个峰且对应于4%、7%和11%的部分浓度。当仅胎儿中的一个在对于其母亲是纯合的SNP基因座处是杂合的时,前两个峰对应于由两个胎儿各自贡献的部分浓度,由胎儿特异性等位基因贡献的部分将用于反映那个特定双胞胎的部分胎儿DNA浓度。最后一个峰对应于两个胎儿是杂合的但母亲是纯合的SNP基因座。使用这些SNP基因座计算的部分胎儿DNA浓度将是由两个胎儿贡献的总部分胎儿DNA浓度。对于需要最小部分胎儿DNA浓度以便作出准确诊断的任何应用,三个数字中最低的将是有关的,因为这将反映来自对母体血浆样品具有最小DNA贡献的胎儿的部分胎儿DNA浓度。考虑到后面一个胎儿的部分胎儿DNA浓度,随后可调整测序深度,以提供统计上强的诊断结果。

[0168] I. 使用关于整个群体的统计变化

[0169] 在一个实施例中,关于两个或更多峰的合并群体的SD值可用于测定接合性。对于单卵胎儿,SD(在参数值扩展中的量度例子)将小于关于二卵双生的SD。这是因为潜在数据实际上来自两个峰,并且因此参数值将比存在一个峰更不同,并且测量的参数值中的扩展仅仅是由于统计变化。此类技术将不要求在特定母系基因型内的分开基因座群体的鉴定。例如,对于其中检测到两个等位基因的基因座,仍可能需要区分其中母亲是杂合的区域与其中母亲是纯合的区域。

[0170] 因此,在二卵双生妊娠中,如通过与母体血浆中的总DNA比较,一系列胎儿特异性遗传标记的部分浓度中的每一个或所选组合测量的,母体血浆或血清中的胎儿DNA的部分浓度将显示出比单卵双生妊娠中的部分浓度更大的变化。变化可通过本领域技术人员众所周知的统计方法进行测量,所述统计方法例如SD、范围、四分位数间距等。

[0171] 相应地,标准化参数中的标准差或其他方差(例如,如步骤440中计算的)可用于测定接合性。方差可与阈值比较,并且如果方差超过阈值,则随后可测定胎儿中的至少两个是二卵的。比较标准化参数与截断值以测定标准化参数是否在统计上不同于期望值的动作通过计算方差和与阈值比较有效地实现。

[0172] J. 关于减数分裂重组的考虑

[0173] 如上所述,可选择在染色体区域内的基因座,以使得在基因座之间的重组概率很低,例如小于1%。下文例子解决了当重组在染色体区域的基因座之间发生时出现的问题。

[0174] 如上所述,对于任何染色体区域,当两个胎儿继承相同父系单元型时,基于胎儿特异性等位基因计算的表观部分胎儿DNA浓度将代表由两个胎儿贡献的总DNA。另一方面,当在父亲是杂合(AB)并且母亲是纯合(AA)的任何SNP基因座处,两个胎儿继承来自父亲的不同父系单元型时,仅胎儿中的一个将对母体血浆贡献胎儿特异性等位基因。因此,在这些基因座处测量的表观部分胎儿DNA浓度将低于在其中两个胎儿均继承相同父系单元型的基因座处测量的值。

[0175] 图10显示了在怀有一对二卵双生的孕妇中重组对表观部分胎儿DNA浓度的作用的例子。在这个例子中,当父系单元型传递给双胞胎1时,在父系Hap III和Hap IV之间存在重组。重组在SNP基因座3和4之间发生。有效地,双胞胎1继承关于基因座1至3的Hap III,和关于基因座4至6的Hap IV。双胞胎2继承来自父亲的Hap IV,而无任何重组。

[0176] 在关于这个孕妇的部分胎儿DNA浓度的分析过程中,仅1、2、3和5是提供信息的,因为胎儿特异性等位基因(B等位基因)将存在于母体血浆中。基因座4和6将变成不提供信息的,因为两个胎儿均继承来自父亲的A等位基因,其等于母系等位基因。另一方面,在基因座5处,两个胎儿均继承来自父亲的B等位基因,导致与对于基因座1至3测定的值比较的更高表观胎儿DNA。如果整个区域(涉及基因座1至6)用于分析表观胎儿DNA浓度,则估计浓度将在关于两个胎儿均继承相同父系单元型的区域和两个胎儿继承不同父系单元型的区域的表观部分胎儿DNA浓度之间。

[0177] IV. 推导的SNP调用

[0178] 当关于两个亲本的基因型信息是已知的时,并且当仅母亲的基因型是已知的时,上文部分描述了用于鉴定接合性的例子。然而,当不了解关于亲本的基因型信息时,可应用实施例(例如方法100和400)。在此类情况下,测量的参数值可进行分组,以便鉴定母系基因型,或至少最可能鉴定母系基因型。以这种方式,在分析取自母亲的生物样品前,无需了解先验知识。因此,实施例可推导母系基因型和胎儿基因型。应注意到一些区域可保持未分类。

[0179] 一旦至少一些染色体区域被分类(例如分为其中母亲是纯合的并且胎儿中的至少一个是杂合的组),本文描述的方法就可用于测定胎儿的接合性。例如,一对双胞胎是单卵还是二卵的测定可通过使用多个SNP基因座分析表观部分胎儿DNA浓度(或其他参数)来实现。对于单卵双生,表观部分胎儿DNA浓度在跨越全基因组的不同SNP基因座处将是相似的。然而,对于二卵双生,由于两个胎儿的基因型中的差异,表观部分胎儿DNA浓度将显示双峰分布。

[0180] 除了测定接合性外,胎儿DNA浓度也可由遗传分析进行测定,其如上所述由于多胎妊娠而是复杂的。如上所述,除了测定接合性外,胎儿DNA浓度对于其他非侵入性胎儿诊断技术是有用的。

[0181] A. 方法

[0182] 为了测定直接来自测序数据(例如具有高倍覆盖)或PCR数据的部分胎儿DNA浓度,我们定义在SNP位点*i*处的可变比 $a_i: b_i = \frac{b_i}{a_i + b_i}$, 其中 a_i 是在SNP基因座(位点)*i*处的特定等

位基因(即主要等位基因)的最大计数, b_i 是在SNP基因座 i 上的另一等位基因(即次要等位基因)的次最大计数。这个值是针对其中母亲是纯合的基因座使用上式的表观胎儿DNA浓度的一半。这种技术也可与其他标准化参数一起使用,包括如上所述的绝对胎儿DNA浓度。

[0183] 考虑单胎妊娠的母体血浆中的SNP位点(或其他多态位点),基于母系基因型和胎儿基因型组合,即所谓的母系-胎儿混合基因型,这些SNP位点可分类成四个种类,无疑为 AA_{AA} 、 AA_{AB} 、 AB_{AA} 和 AB_{AB} 的形式,其中上标代表母系基因型,下标代表胎儿基因型。 AA 指示纯合性并且 AB 指示杂合性。因此,在每个SNP基因座 i 处的测量由等位基因A出现数目(a_i)(其对应于最大计数)和等位基因B出现数目(b_i)(其对应于来自测序数据的次最大计数)组成。对于种类 AA_{AB} 和 AB_{AA} ,部分胎儿DNA浓度将影响在每个SNP位点处的 $\frac{a_i}{a_i+b_i}$ 。理论上,关于种类 AA_{AA} 的 $\frac{a_i}{a_i+b_i}$ 是0,关于种类 AB_{AB} 的 $\frac{a_i}{a_i+b_i}$ 是0.5。因此,部分胎儿DNA浓度可测定母体血浆中的母系-胎儿混合基因型的分布。原则上,部分胎儿DNA浓度的最佳估计将产生生成具有最高概率的观察到的 $\frac{a_i}{a_i+b_i}$ 概况的分布。

[0184] 实施例可区分四个种类且执行种类 AA_{AB} 和 AB_{AA} 的更多分析,如上文例如在方法400中所述。作为区分不同种类的部分(即,哪个染色体区域属于哪个种类),实施例可通过发现将数据与概率分布的线性组合最佳拟合的最佳浓度来测定胎儿DNA浓度。

[0185] 图11是示出根据本发明的实施例的方法1100的流程图,所述方法1100测定来自具有至少两个胎儿的妊娠女性的生物样品中的胎儿DNA百分比,并且测定至少两个胎儿是否是二卵的。生物样品包含胎儿和母体DNA。

[0186] 在步骤1110中,通过测定对于多个染色体区域中的每一个的标准化参数(例如如上文对于方法400描述的)制备直方图。对于多个染色体区域中的每一个,鉴定在各自的染色体区域中的一个或多个基因座,在生物样品中检测在所述染色体区域处各自的第一等位基因和各自的第二等位基因。此类区域对应于上文描述的四个不同种类。

[0187] 等位基因可以任何合适的方式进行检测,并且可确保统计上显著量的每个等位基因。在生物样品中在一个或多个基因座处测量一个或多个第一等位基因的第一个量和/或一个或多个第二等位基因的第二个量。在基因座中的第一等位基因可以是不同等位基因,如第二等位基因可以是一样。在一种实现中,如果在染色体区域中的特定基因座处的第一等位基因比第二等位基因更丰富,则随后在各自的基因座处的所有第一等位基因均比第二等位基因更丰富。以此类方式,可合并来自多个基因座的等位基因的计数。

[0188] 获得对于第一个量或第二个量的标准化参数。标准化参数可以是如本文描述的任何类型。例如,标准化参数可由第一个量和第二个量进行测定,并且提供在第一个量和第二个量之间的相对量度。例子包括部分胎儿DNA浓度。标准化参数还可以是绝对部分胎儿DNA浓度,如可用合适校准测定的。在一个实施例中,使用如上定义的 $\frac{a_i}{a_i+b_i}$ 。

[0189] 一旦对于每个染色体区域测定标准化参数,这些数据点就可用于制备直方图。可通过将标准化参数的可能值分离成多个子范围制备直方图的数据结构。例如,如果范围在0-0.5之间,则子范围可具有各0.01的大小。计数器可与每个子范围相关。随后可基于对于标准化参数具有指定值(即对于在相应子范围内的值)的多个染色体区域增大计数器。

[0190] 在步骤1120中,使概率分布的线性组合与直方图拟合。例如,两个概率分布可用于拟合关于种类 AA_{AB} 和 AB_{AA} 的数据。如下所述,关于这些种类中的每一个的数据点位置依赖胎

儿DNA百分比。因此,胎儿DNA百分比是对概率分布的线性组合的输入。还可使用关于其他种类分布,例如如下所述的。

[0191] 在另一个实施例中,数据点可已限制于特定种类 AA_{AB} 和 AB_{AA} 。对于双胎妊娠,如果两个胎儿对生物学血浆贡献等量的DNA,则分布数目将是二,或如果两个胎儿对母体血浆贡献不同量的DNA,则分布数目将是三。对于涉及3个或更多个胎儿的妊娠,如果每个胎儿对母体血浆贡献等量的DNA,则分布数目将等于胎儿数目。在此类实施例中,步骤1150可作为冗余省略,并且步骤1160将使用概率分布的系数。

[0192] 在步骤1130中,改变输入胎儿DNA百分比 θ ,以发现最佳化概率分布的线性组合与直方图的拟合的最佳胎儿DNA百分比。术语“最佳的”在上文定义。作为最佳化过程的部分,可在概率分布的数据和线性组合之间测定误差。这个误差项可用于测定 θ 值何时提供与数据的足够拟合(即误差足够小,使得 θ 值称为最佳的)。

[0193] 在步骤1140中,一旦已知胎儿DNA百分比 θ ,就鉴定对应于基因座的概率分布,在所述基因座处母亲是纯合的并且胎儿中的至少一个是杂合的,或在所述基因座处母亲是杂合的并且胎儿中的至少一个是纯合的。例如,如果使用两个线性组合,则胎儿DNA百分比 θ 可提供关于分布峰的近似位置。因为胎儿可能是二卵的(或其中至少两个),所以峰将不在由 θ 预测的确切位置处,但两个分布可以被区分。例如,对于种类 AA_{AB} 将是具有最接近于值 $\theta/2$ 的峰的分布,并且对于种类 AB_{AA} 将是具有最接近于 $0.5-\theta/2$ 的峰的分布。

[0194] 在步骤1150中,使多组分混合模型与鉴定的概率分布拟合。多组分混合模型包括关于多个组分中的每一个的混合系数。在一个方面,组分对应于图8或9中的峰。组分可使用任何合适的函数形式例如高斯进行限定,其中每个组分将对应于不同高斯函数。

[0195] 如下所述,附加峰将由三胞胎导致。如果胎儿对母体血浆贡献等量的DNA(例如以解释显示实际胎儿DNA浓度的区域),则测定混合模型中的组分数目的一种方法是使组分数目等于胎儿数目。作为另外一种选择,贝叶斯信息判据(BIC)或阿开克信息判据(AIC)可用于测定混合模型的组分数目。因此,混合模型可具有关于双胞胎的两个组分(其贡献相同量的胎儿DNA)或当双胞胎各自贡献不同量的DNA时,或对于涉及超过两个胎儿的妊娠,具有更高数目的组分。

[0196] 在一个实施例中,可通过将母亲基因分型来发现鉴定的分布,以便以那种方式测定种类 AA_{AB} 或 AB_{AA} 。在此类实施例中,将不完成步骤1130和1140。此外, θ 值可通过其他方法例如使用后生标记进行测定。相反,可通过测定对应于特定染色体区域的概率分布来鉴定在特定染色体区域处的母亲基因型。例如,可测定对于特定染色体区域的标准化参数具有最高值的概率分布,并且使用那个概率分布测定母系基因型。因为概率分布不同地依赖胎儿DNA百分比,所以相应母系基因型是直接测定的。

[0197] 类似地,可通过测定概率分布(其可以是混合模型的组分)鉴定在特定染色体区域处的胎儿基因型,所述概率分布对应于特定染色体区域。例如,可测定对于特定染色体区域的标准化参数具有最高值的概率分布,并且使用那个概率分布测定在各自的SNP基因座处是杂合的胎儿的数目。因为概率分布依赖在各自的SNP基因座处是杂合的胎儿的数目以及胎儿DNA百分比,所以可推导胎儿的基因型信息。

[0198] 在步骤1160中,使用混合系数中的至少两个测定胎儿中的至少两个是否是二卵的。例如,混合系数中的至少两个可与阈值比较。系数的值可用于测定峰实际上是否存在。

如果峰很小,则随后混合系数将是很小的,并且可假定附加峰实际上不存在(例如如果2F%是实际胎儿DNA百分比,则峰将在F%处出现)。在那种情况下,混合模型将传递实际上存在仅一个峰,并且胎儿是单卵的。如果混合系数中的两个超过阈值,则随后实际上存在至少两个峰,并且胎儿中的至少两个是二卵的。阈值可以是绝对值或与其他组分的系数比较的相对值(例如阈值可依赖系数中的一个或多个的值,例如最大值的百分比或系数的超额)。应注意到术语“系数”包含乘以系数或在组分中牵涉的任何定标因子。如果使用三个组分(例如因为每个胎儿对总胎儿DNA百分比的各个贡献是期望的或因为测试的是三胞胎),则二卵的测定可在仅测试两个系数后停止,因为随后已鉴定二卵。

[0199] 在一个实施例中,混合模型的组分的峰的位置受约束,以具有超过预定值的分离缺口。分离缺口是从一个峰到最近峰的距离(如以直方图的参数值单位测量的)。预定值可依赖所需准确度和期望的结果。例如,如果希望仅仅鉴定接合性,则随后可使用模型的两个组分(例如因为来自每个个体胎儿的贡献的峰之间的区分不是期望的),并且预定值可依赖测定的胎儿DNA浓度(例如更大的胎儿DNA浓度可导致更大的分离距离)。预定值也可依赖胎儿数目,例如对于三胞胎,不同峰的分离在相对基础上可以是更少的,因为峰可大致对应于实际胎儿DNA浓度的三分之一而不是一半。

[0200] 可使用峰的相对大小。例如,相对大小可指示是否存在其中胎儿是遗传上相同或胎儿是遗传上不同的更多个区域。这将依赖染色体区域的选择。然而,在胎儿的情况下,如果染色体区域的选择是随机的,则两个二卵双生继承相同父系单元型和不同父系单元型的区域数目应是相似的。此处,使用峰的相对大小可使得数据与不同分布曲线的拟合更准确。

[0201] B. 通过推导SNP调用分析的部分胎儿DNA测定

[0202] 为了示出上文原理的操作,我们执行靶标测序且应用二项式混合模型,以估计胎儿DNA浓度。二项式混合建模已成功应用于鉴定肿瘤基因组中的单核苷酸变体(Goya R等人Bioinformatics2010;26:730-736;Shah SP等人Nature2009;461:809-813)。我们采用模型估计妊娠母体血浆中的胎儿DNA浓度。我们假定在血浆中在SNP i 处的母系-胎儿混合基因型 $G_i=k$, $k \in \{AAAA, AAAB, ABAA, ABAB\}$ 是多项式随机变量。我们使 $\mathbf{x}_i = \begin{bmatrix} a_i \\ b_i \end{bmatrix}$ 代表在SNP位置 i 处的A等位基因和B等位基因的等位基因计数,其中 $N_i=a_i+b_i$ 是观察到的读数深度。我们假定在SNP i 处的计数由二项式分布产生,所述二项式分布在 $G_i=k$, $X_i \sim \text{Binom}(b_i | \mu_k, N_i)$ 上是条件化的,其中 $\mu_k \in \{\mu_{AAAA}, \mu_{AAAB}, \mu_{ABAA}, \mu_{ABAB}\}$ 是在母体血浆中的B等位基因频率(b_i)的预期分布,其由下式给出:

$$[0203] \quad \mathbf{x}_i \sim \binom{N_i}{b_i} \mu_k^{b_i} (1 - \mu_k)^{a_i} \quad (1)$$

[0204] 理论上 μ_{AAAA} 接近于常数0, μ_{AAAB} 接近于胎儿DNA浓度的一半 $\frac{f}{2}$, μ_{ABAA} 接近于 $0.5 - \frac{f}{2}$, 并且 μ_{ABAB} 接近于常数0.5。对于 μ_{AAAA} 和 μ_{ABAB} ,与期望常数值0和0.5的偏差分别主要受测序误差和分析偏差影响,其中对于 μ_{AAAB} 和 μ_{ABAA} ,来自0和0.5的干扰在很大程度上由部分胎儿DNA浓度决定。分析偏差的例子包括在大规模平行测序中的GC偏差(Chen EZ等人PLoS One2011;6:e21791)和比对偏差(Degner JF等人Bioinformatics2009.25:3207-3212.)。

[0205] 随后,我们应用二项式混合模型,以解释观察到的等位基因计数。对于给定SNP i , X_i 的混合分布 $p(X_i)$ 衍生自二项式分布的线性组合,其通过多项式 $\pi_k, 0 \leq \pi_k \leq 1$ 和 $\sum_k \pi_k = 1$ 加权:

$$[0206] \quad p(X_i) = \sum_{c_k} \pi_k \text{Binom}(X_i | \mu_k, N_i) \quad (2)$$

[0207] 其中 π_k 是在母系-胎儿混合基因型上的先验概率。换言之, π_k 代表随机选择的位置将在基因型中的每一个上发生的先前信念。

[0208] 完全数据对数似然性(log-likelihood)通过下式给出:

$$[0209] \quad \log p(X_{1:T} | \mu, \pi) = \sum_{i=1}^T \log \sum_{c_k} \pi_k \text{Binom}(X_i | \mu_k, N_i) \quad (3)$$

[0210] 其中 T 是在母体血浆中观察到的SNP位置总数目, $\pi_k \in \{\pi_{AAAA}, \pi_{AAAB}, \pi_{AABAA}, \pi_{AABAB}\}$ 。此外,似然性还可通过作图质量和碱基质量建模(Goya R等人 Bioinformatics 2010; 26:730-736),其可通过下式潜在地改善准确度:

$$[0211] \quad \log p(X_{1:T} | \mu, \pi) \propto \sum_{i=1}^T \log \sum_{c_k} \pi_k \prod_{j=1}^{N_i} (0.5(1-r_j^i) + r_j^i[(1-q_j^i)(1-\mu_k) + q_j^i\mu_k]) \quad (4)$$

[0211] 其中 r_j^i 是作图质量,并且 q_j^i 是在位置 i 处关于第 j 个比对碱基的测序质量。在我们的缺省分析中,我们建模测序质量,因为大多数目前的作图软件不获得作图质量。

[0212] 一般而言,根据不同母系-胎儿混合基因型的群体频率, π_k 分别通过0.7、0.1、0.1和0.1指定,并且可例如使用Affymetrix Genome-Wide Human SNP Array6.0进行估计。对于 AA_{AB} 和 AB_{AA} , μ_k 依赖部分胎儿DNA浓度 ϕ , $0 \leq \phi \leq 1.0$ 。 ϕ 在妊娠母体血浆中大于0.4是不太可能的(Chiu RW等人BMJ 2011; 342:c7401)。因此,我们计算从0到0.5迭代的部分胎儿DNA浓度 ϕ ,处理0.001或更少的增量/迭代,直至对数似然性达到最大值。当 ϕ 接近实际部分胎儿DNA浓度时,对数似然性预期达到最大值。换言之,使对数似然性达到最大的典型 ϕ 可解释观察到的具有最高概率的等位基因计数,否则我们需要重复更新 π_k 和 μ_k ,直至对数似然性已达到其如上文定义的最大值。

[0213] 当考虑 AA_{AA} 和 AB_{AB} 时,更新原则如下述等式定义:

$$[0214] \quad \pi^{new}(k) = \frac{\sum_{i=1}^T I(G_i=k) + \delta_k}{\sum_j \sum_{i=1}^T I(G_i=j) + \delta_j} \quad (5)$$

[0215] 其中 $I(G_i=k)$ 是指示函数,以表明根据不同母系-胎儿混合基因型的加权概率,母系-胎儿混合基因型 kG_i 在SNP i 处是否指定至 G_i ,并且:

$$[0216] \quad I = \begin{cases} 1 & G_i=k \\ 0 & G_i \neq k \end{cases} \quad (6)$$

$$[0217] \quad \mu^{new}(k) = \frac{\sum_{i \in \{AA_{AA}, AA_{AB}, AB_{AA}, AB_{AB}\}} \sum_{j=1}^T I(G_i=k) + \mu_k - 1}{\sum_{i \in \{AA_{AA}, AA_{AB}, AB_{AA}, AB_{AB}\}} \sum_{j=1}^T I(G_i=j) + \mu_k + \mu_k - 2} \quad (7)$$

[0218] 而 AA_{AB} 和 AB_{AA} 处于血浆中的部分胎儿DNA浓度的约束下,从而:

$$[0219] \quad \mu^{new}(AA_{AB}) = \frac{\phi}{2} \quad (8)$$

$$[0220] \quad \mu^{new}(AB_{AA}) = \mu^{new}(AB_{AB}) - \frac{\phi}{2} \quad (9)$$

[0221] 此外,对于上式(6)~(9), μ_k 根据 β 分布 $\mu_k\beta(\mu_k|\alpha_k,\beta_k)$ 进行分布。我们分别设置 $\alpha_k=\{10000,9500,5500,5000\}$ 和 $\beta_k=\{1,500,4500,5000\}$ 对应于 $G_k\in\{AA_{AA},AA_{AB},AB_{AA},AB_{AB}\}$,以通过 $\frac{\mu_k}{\alpha_k+\beta_k}$ 初始化 μ_k ,因为我们推理 μ_k 预期围绕 $\{0,0.05,0.45,0.5\}$ 波动,并且通过实际观察到的

的 μ_k 分布进行调整。 π_k 根据狄利克雷分布进行分布: $\pi_k\beta(\pi_k|\delta_k)$ 。 π_k 通过 $\frac{\pi_k}{\sum_k \pi_k}$ 初始化,其中 δ_k 通过权向量 $\{7,1,1,1\}$ 缺省设置,其代表在妊娠母体血浆中的 G_k 比例,并且因此指示随机选择的位置将指定至每个基因型的先前信念。

[0222] 一旦参数 π_k 和 μ_k 通过观察到的等位基因计数拟合,我们就可应用贝叶斯定理以计算在母系-胎儿混合基因型上的后验概率, $\gamma_k=\Pr(G_k|a_i,N_t,\pi_k,\mu_k)$,其中:

$$[0223] \quad \gamma_k = \frac{\pi_k \text{Binom}(N_i | \mu_k, N_i)}{\sum_j \pi_j \text{Binom}(N_i | \mu_j, N_i)} \quad (9)$$

[0224] 此外,基于 γ_k ,我们可鉴定以形式 AA_{AB} 的提供信息的SNP位点。

[0225] C. 来自提供信息的SNP位点的接合性状态的测定

[0226] 在双胎妊娠中,我们可应用上述算法以估计表观部分胎儿DNA浓度且鉴定属于母系-胎儿混合基因型 AA_{AB} 的SNP位点。在这种情况下, AA_{AB} 具有与单胎妊娠中的那种不同的含义。对于单卵双生, AA_{AB} 的母系-胎儿混合基因型的推导表明在两个胎儿中的胎儿基因型AB。对于二卵双生, AA_{AB} 包括起因于不同遗传背景的三个潜在组,分别为AA/AA/AB(母亲/胎儿I/胎儿II)、AA/AB/AA和AA/AB/AB。因为仅基于测序数据我们不能区分AA/AA/AB与AA/AB/AA,我们将前两个种类算术上合并成一个,并且将仅存在两个有效的潜在种类,分别为AA/AA/AB和AA/AB/AB,以便执行下述分析。

[0227] 通过双胞胎的接合性状态测定在母体血浆中关于全基因组(或目的靶标区域)的部分胎儿DNA浓度(f_i /血浆中的SNP位点或块或染色体)的模式(即是存在一种还是两种模式)。对于双胞胎,当胎儿贡献不同百分比的胎儿DNA时,对于二卵状态可出现两个模式,其中这些模式之一实际上可包含两个子节点(例如如图9中所示)。

[0228] 为了阐明基因构成的模式,我们拟合双组分高斯混合模型(GMM)与 f_i 的分布:

$$[0229] \quad f_i = 2 \times \mu_i \quad (10)$$

$$[0230] \quad p(f_i) = \sum_{m=1}^M \lambda_m N(f_i | \theta_m, \sigma_m^2) \quad (11)$$

[0231] 其中 λ_m 是对于 $m=1, \dots, M$ (其中 M 是可能峰的最大数目)满足 $\sum_{m=1}^M \lambda_m = 1$ 和 $0 \leq \lambda_m \leq 1$ 的混合比例。当胎儿将等量的DNA贡献到母体血浆内时, M 等于胎儿数目,即对于双胎妊娠是2。 θ_m 和 σ_m^2 是正态分布的平均值和方差。

[0232] 我们使用标准期望-最大化(EM)算法(McLachlan GJ和Krishnan T.The EM Algorithm and Extensions.New York:Wiley;1997)估计模型(11)中的混合组分。在该算法中,当我们鉴定两个组分时,我们设置两个另外约束:

[0233] 1) $\lambda_m \geq 0.2$,其中 λ_m 代表峰 m 的大小;

[0234] 2) $f_1 - f_2 \geq \sqrt{\sum_{m=1}^M \lambda_m \sigma_m^2}$,其中 $f_1 - f_2$ 指示两个峰之间的距离。

[0235] D.结果

[0236] 我们招募一个双绒毛膜双羊膜腔(DCDA)双胞胎妊娠和一个单绒毛膜双羊膜腔(MCDA)双胞胎妊娠用于我们的研究。对于DCDA病例,从处于怀孕17周的母亲中取得血样。来自每个双胞胎的脐带血贮存用于研究。对于MCDA病例,从处于怀孕12周的母亲中取得血样。绒毛膜绒毛取样(CVS)DNA的部分贮存用于研究。

[0237] 为了执行SNP基因分型,用Affymetrix Genome-Wide Human SNP Array6.0系统基因分型从双胞胎的母体血沉棕黄层和脐带血或CVS样品中提取的DNA。基于母亲和每个胎儿中的纯合性/杂合性状态,将SNP分类成不同种类(图12A的表1200和图12B的表1220)。关于双胞胎I和双胞胎II的基因型的一致性指示DCDA病例是二卵的(图12C的表1240)。表1200显示了关于DCDA病例的基因分型结果。表1220显示了关于MCDA病例的基因分型结果。表3显示了关于DCDA病例中的双胞胎I和双胞胎II的一致性。

[0238] 如下执行血浆DNA的测序。为了示出目的,使用Agilent SureSelect技术靶标富集从母体血浆中提取的DNA,并且随后通过Illumina Hi-Seq标准配对末端方案对提取的DNA片段(203-209,000,000)测序,对于每个末端50bp,所述DNA片段等价于关于大小总计5.5Mb的靶标区域的平均138-143倍覆盖。靶标区域包括chr1(0.33Mb)、chr2(0.30Mb)、chr3(0.62Mb)、chr4(0.32Mb)、chr5(0.33Mb)、chr7(0.31Mb)、chr8(0.62Mb)、chr9(0.31Mb)、chr12(0.05Mb)、chr13(0.30Mb)、chr15(0.33Mb)、chr17(0.66Mb)、chr19(0.35Mb)、chr20(0.34Mb)和chr22(0.30Mb)。还可使用其他靶标富集技术例如Roche NimbleGen平台和基于PCR的技术(例如使用RainDance平台)。

[0239] 跨越每个染色体测定表观部分胎儿DNA浓度。通过与测序数据组合的基因型计算在母体血浆中的表观部分胎儿DNA浓度。图12D显示了基于与测序数据结合的基因型计算的表观部分胎儿DNA浓度(母亲是AA并且胎儿中的至少一个是AB)。关于每个染色体的结果对于MCDA病例是相对恒定的,具有SD1.52,而DCDA病例显示出更多波动,其中SD几乎倍增至3.36(图12D的表1260)。对于MCDA,部分胎儿DNA浓度是16.35%。对于DCDA,胎儿I的部分胎儿DNA浓度是12.35%(AA/AB/AA),并且胎儿II的部分胎儿DNA浓度是13.60%(AA/AB/AA)。合并的部分胎儿DNA浓度是22.45%(AA/AB/AB),并且表观部分胎儿DNA浓度是18.82%。

[0240] 为了研究我们是否可直接由靶标测序数据区分接合性,我们执行推导SNP调用分析且随后计算胎儿DNA浓度。图12E显示了通过推导SNP调用分析计算的表观部分胎儿DNA浓度。结果显示和与测序数据结合的基于基因型的估计值的高一致性(关于DCDA和MCDA的皮尔森相关系数分别为0.8和0.8)(图12E的表1280)。此外,SD允许我们直接区分二卵双生与单卵双生(关于二卵双生和单卵双生分别为2.14和1.10)。

[0241] 对于表观部分胎儿DNA浓度的总体分布,我们将靶标基因组区域分成多个块(对于DCDA和MCDA分别为432和445),而每个块通过包括超过距离小于50kb的5个邻近SNP进行构建。随后我们计算关于每个块的表观部分胎儿DNA浓度,并且描绘胎儿DNA浓度的分布。我们使用双组分高斯混合模型以拟合部分胎儿DNA浓度/块的分布。

[0242] 因此,对于DCDA病例,我们可鉴定来自分布的两个不同峰(图13),而对于MCDA病例,我们仅获得1个峰(图14)。因此,二卵双生可通过高斯混合模型与单卵双生区分。

[0243] 在另一个实施例中,我们在计算机分析上分别模拟关于单卵双生和二卵双生的部分胎儿DNA浓度的两个分布。随后,比较实际部分胎儿DNA分布与模拟的部分胎儿DNA分布,

以通过测定模拟分布中的哪个更接近于实际部分胎儿DNA分布来推导接合性状态。通过改变等式(11)中的M值,实施例还可用于涉及三个或更多个胎儿的妊娠。

[0244] 总之,通过母体血浆DNA的大规模平行测序,就表观部分胎儿DNA浓度/SNP或块或染色体而言,我们发现在单卵双生和二卵双生之间的显著性差异。这种技术具有它是非侵入性并且还比通过超声扫描的形态学观察更可靠的优点。在其他实施例中,其他统计学模型例如隐马尔可夫模型也可用于测定是否存在胎儿DNA浓度的一个峰还是多个峰。

[0245] V. 部分胎儿DNA浓度的变化

[0246] 通过测定在不同染色体区域处的部分胎儿DNA浓度,我们可测定是否存在仅一个还是超过一个胎儿DNA浓度。关于怀有单卵和二卵双生的孕妇的不同染色体区域的血浆部分胎儿DNA浓度显示于图15中。对于怀有一对单卵双生的孕妇,部分胎儿DNA浓度跨越不同染色体区域是一致的。相比之下,在怀有一对二卵双生的孕妇中,在跨越不同染色体的部分胎儿DNA浓度之间存在增加的变化。此类变化可与上文对于二卵胎儿显示的增加的标准差(SD)比较。

[0247] 在另一个实施例中,我们使用固定数目的SNP用于计算跨越不同染色体区域的表观部分胎儿DNA浓度。原则上,表观部分胎儿DNA浓度在怀有二卵双生的孕妇中显示跨越不同基因组区域的波动,而表观浓度在怀有单卵双生的孕妇中跨越不同基因组区域是稳定的。为了测定部分浓度的波动是由于具有单卵双生的随机变化还是具有二卵双生的不同部分胎儿DNA浓度的存在,我们已执行模拟分析以测定随机变化的水平,其中假定一对单卵双生的存在(图16)。

[0248] 基于下述假定执行模拟:(a)存在一对单卵双生;(b)所有区域的总体部分浓度用作每个测试区域的部分浓度;(c)每个测试区域中的SNP的测序深度等于所有区域的平均测序深度;和(d)SNP等位基因的分布遵循二项式分布。

[0249] 在图16中,相对表观部分胎儿DNA浓度计算为基于1000个连续SNP的等位基因计数的运行总数。通过将关于1000个SNP的局部区域浓度除以所有区域的平均部分胎儿DNA浓度,计算相对表观部分浓度。绿线代表关于怀有一对单卵双生的母亲的结果,并且红线代表关于怀有一对二卵双生的母亲的结果。灰色的阴影面积代表如上所述的1000次不同模拟的结果。我们可见绿线在阴影面积内波动,指示部分浓度的波动在关于一对单卵双生的预期变化内。另一方面,红线波动超过阴影面积,指示部分浓度的波动不能通过单独的随机变化加以解释,并且因此,暗示该对双胞胎是二卵的。

[0250] VI. 三胎及更多

[0251] 上文描述的方法还可用于测定三胎或更多胎妊娠中的所有胎儿是否是遗传上相同的或胎儿中的至少一个在遗传上不同于其他。在三胎或更多胎妊娠中,如例如通过超声证明的,如果两个父系单元型在母体血浆样品中可检测,则它将指示胎儿中的至少一个不同于其他胎儿。对于部分胎儿DNA浓度的分析,推导提供信息的SNP的方法将不改变。在鉴定提供信息的SNP和计算关于不同基因组区域的部分胎儿DNA浓度后,胎儿的接合性状态可通过上文给出的公式进行测定,伴随公式中的参数相应改变。

[0252] 对于三胎及更多,关于多组分混合模型中的可能峰M的数目可如下测定。当存在3个或更多个胎儿时,M的计算将复杂得多。图17显示了根据本发明的实施例,示出对于三个胎儿(胎儿A、B和C)的多个可能峰的直方图。这三个胎儿是遗传上不同的,并且它们对母体

血浆贡献不同量的DNA。未填充的峰代表其中仅一个胎儿是杂合的染色体区域。具有一种颜色的峰代表其中两个胎儿是杂合的染色体区域。具有不同图案的填充的峰代表其中所有三个胎儿均为杂合的染色体区域。因此,将存在总共6个峰。

[0253] 在胎儿数目(N)和峰数目(M)之间的关系将是 $M = \sum_{i=1}^N C_i^N$ 。实际上,当所有胎儿均对母体血浆贡献不同量的DNA时,将不需要鉴定所有峰。然而,如果所有胎儿均对母体血浆贡献等量的DNA时,M将等于N。

[0254] VII. 实例

[0255] 下述是用于测定由胎儿或多个单卵胎儿贡献的最低部分胎儿DNA浓度的假设实例。关于第一个实例,多胎妊娠中的不同胎儿的部分胎儿DNA浓度的测定用于调整基于母体血浆分析的产前诊断测试的灵敏度(还参见段落134)。在此类应用中,由任一个遗传上不同的胎儿贡献的最低部分胎儿DNA浓度或来自两个或更多个遗传上相同的胎儿的合并部分胎儿DNA浓度(无论哪个更低),可用于指导诊断测试的灵敏度是否足以检测胎儿的遗传异常,例如胎儿非整倍性。在图17中所示的实例中,由胎儿A贡献的峰代表最低表观胎儿DNA浓度。

[0256] 在一个实施例中,最低部分胎儿DNA浓度可通过分析多个提供信息的SNP进行测定,在所述提供信息的SNP处母亲是纯合的并且父亲是杂合的。这些SNP优选位于多个不同染色体或染色体区域上。通过携带胎儿特异性等位基因和在母亲与胎儿之间共享的等位基因的DNA片段的数目,计算在这些SNP中的每一SNP处的表观部分胎儿DNA浓度。在其中仅将最低量的DNA贡献到母体血浆内的胎儿是杂合的而其他胎儿是纯合的任何SNP基因座处,表观部分胎儿DNA浓度将在分析的所有SNP基因座中变成最低的。因此,在这些SNP基因座中测定的最低部分胎儿DNA浓度可用作由任何遗传上不同的胎儿贡献的最低胎儿DNA量的估计量。

[0257] 在另一个实施例中,可使用在SNP基因座上的选择区域数字分析(DANSR)执行在SNP基因座处的DNA部分浓度的测定(Sparks AB等人Am J ObstetGynecol2012;doi:10.1016/j.ajog.2012.01.030)。位于一个或多个染色体上的胎儿特异性和共享等位基因的数字计数可用于估计在SNP基因座处中的每一基因座处的表观部分浓度。作为另外一种选择,在位于相同染色体区域上的多个SNP基因座上的等位基因计数信息可一起分析,以指示在各自的染色体区域处的表观部分胎儿DNA浓度。显示出最低表观胎儿DNA浓度的染色体区域可用于指示由胎儿中任一个贡献的最低胎儿DNA。在其他实施例中,可使用实时PCR、质谱法分析(例如通过Sequenom MassARRAY系统)和数字PCR分析执行在不同SNP基因座处的部分浓度的测定。

[0258] 待分析的SNP基因座数目应足够大,以确保对于涉及至少一对二卵双生的妊娠,分析至少一个基因座,仅对母体血浆贡献最低量的DNA的胎儿(或单卵胎儿的组合)将是杂合的,而其他胎儿对于母体等位基因将是纯合的。例如,应分析超过100个潜在提供信息的SNP。此外,对于每个潜在提供信息的SNP基因座应分析足够大量的DNA片段,以确保检测到胎儿特异性等位基因。

[0259] 第二个实例提供根据实施例的过程的描述。从通过超声已知怀有双胞胎胎儿的孕妇中获得10mL母体外周血样品。将血样分馏成血浆和血细胞。收获血浆且提取DNA。随后在一系列10次多重PCR中扩增母体血浆DNA。每个PCR允许扩增在不同基因组区域中分布的20个

SNP基因座。因此,在这个实例中将分析总共200个SNP基因座。

[0260] 通过引物延伸测定分析来自多重PCR各自的PCR产物,并且例如通过Sequenom MassARRAY系统分析延伸产物。每个引物延伸测定随后为Sequenom MassARRAY分析将揭示对应于SNP基因座中每一个的等位基因的质谱法峰。峰的相对高度将指示SNP等位基因中每一个的相对量。对于妊娠母亲及其胎儿对于相同等位基因是纯合的SNP,在质谱法读数上可见对应于那个等位基因的仅一个峰。对于妊娠母亲是杂合的SNP,在质谱法读数上可见对应于SNP的两个等位基因的具有近似相等高度的两个峰。对于妊娠母亲是纯合的并且胎儿中的至少一个是杂合的SNP,在质谱法读数上可见一个大峰(对应于母亲的等位基因)和一个小峰(对应于母亲的基因组中不存在的胎儿等位基因)。后面两个峰的相对大小将允许测量胎儿DNA百分比(即部分胎儿DNA浓度)。

[0261] 在使用SNP基因座的特定组合的一个实施例中,显示出一个大峰(对应于母系等位基因)和一个小峰(对应于胎儿特异性等位基因)的SNP比例将是例如6%。如果双胞胎儿是单卵的,则随后通过这6%SNP测量的部分胎儿DNA浓度应彼此相对接近。紧密度可例如通过标准差(SD)进行测量。如果双胞胎儿是二卵的,则随后6%SNP可分成两组。对于第一组,两个胎儿仅偶然地均为杂合的。对于第二组,同样仅偶然地,胎儿之一将是纯合的并且另一个将是杂合的。通过第一组中的一个或多个SNP测量的部分胎儿DNA浓度将大于通过第二组中的一个或多个SNP测量的那种。对于此类分析选择的SNP可选择为关于特定群体最提供信息的。

[0262] 作为质谱法的替代,可使用本领域技术人员已知的可比较方法执行这个实例中所述的分析。一个例子是执行在SNP基因座上的扩增测序,随后为大规模平行测序。这种策略的一个变体是选择区域数字分析(DANSR)(Sparks AB等人Am J ObstetGynecol2012;doi:10.1016/j.ajog.2012.01.030)。另一个例子是微滴数字PCR,例如使用QuantaLife平台或RainDance平台(Zhong Q等人Lab Chip2011;11:2167-2174)。另外一个例子是使用微流体数字PCR。此类测定的多路化增加此类分析的通量。

[0263] VIII.计算机系统

[0264] 本文提及的任何计算机系统均可利用任何合适数目的子系统。此类子系统的例子显示于图18中的计算机仪器1800中。在一些实施例中,计算机系统包括单个计算机仪器,其中子系统可以是计算机仪器的组件。在其他实施例中,计算机系统可包括多个计算机仪器,各自为子系统,具有内部组件。

[0265] 图18中所示的子系统经由系统总线1875互联。显示了另外的子系统例如打印机1874、键盘1878、固定盘1879、偶联至显示器适配器1882的监视器1876及其他。偶联至I/O控制器1871的外周和输入/输出(I/O)设备可通过本领域已知的任何数目的装置例如串行端口1877连接至计算机系统。例如,串行端口1877或外部界面1881可用于将计算机系统1800连接至广域网例如因特网、鼠标输入设备或扫描仪。经由系统总线1875的互联允许中央处理器1873与每个子系统通信,并控制来自系统存储器1872或固定盘1879的指令的执行以及在子系统之间的信息交换。系统存储器1872和/或固定盘1879可体现计算机可读介质。本文提及的任何值均可从一个组件输出到另一个组件且可输出至用户。

[0266] 计算机系统可包括例如通过外部界面1881或通过内部界面连接在一起的多个相同组件或子系统。在一些实施例中,计算机系统、子系统或仪器可经过网络通信。在此类情况下,一个计算机可视为客户机,并且另一个计算机视为服务器,其中各自可以是相同计算

机系统的部件。客户机和服务器可各自包括多个系统、子系统或组件。

[0267] 应当理解本发明的任何实施例可以控制逻辑的形式实现,其中使用硬件和/或使用以模块或整合方式的计算机软件。基于本文提供的公开内容和教导,本领域普通技术人员将了解且理解实现本发明的实施例的其他方式和/或方法,其中使用硬件和硬件与软件的组合。

[0268] 本申请中所述的任何软件组件或函数可作为通过处理器执行的软件代码实现,所述处理器使用任何合适的计算机语言例如Java、C++或Perl,使用例如常规或面向对象的技术。软件代码可贮存为在计算机可读介质上的一系列指令或命令以用于贮存和/或传递,合适的介质包括随机存取存储器(RAM),只读存储器(ROM),磁性介质例如硬驱或软盘,或光学介质例如光盘(CD)或DVD(数字多用盘)、闪速存储器等。计算机可读介质可以是此类储存或传递设备的任何组合。

[0269] 此类程序也可使用适于经由遵照多个方案的有线、光学和/或无线网络包括因特网传递的载波信号进行编码且传递。像这样,根据本发明的实施例的计算机可读介质可使用由此类程序编码的数据信号制备。由程序代码编码的计算机可读介质可与相容设备一起包装或与其他设备分开提供(例如经由因特网下载)。任何此类计算机可读介质可位于单个计算机程序产品(例如硬驱、CD或整个计算机系统)上或内,并且可存在于系统或网络内的不同计算机程序产品上或内。计算机系统可包括监视器、打印机或其他合适的显示器,用于给用户提供本文提及的任何结果。

[0270] 本文描述的任何方法可用包括处理器的计算机系统全部或部分执行,所述计算机系统可配置为执行步骤。因此,实施例可涉及配置为执行本文描述的任何方法的步骤的计算机系统,潜在地具有执行各自步骤或各自步骤组的不同组件。尽管作为编号步骤呈现,但本文方法的步骤可同时或以不同次序执行。另外,这些步骤的部分可与来自其他方法的其他步骤的部分一起使用。此外,步骤的全部或部分可以是任选的。另外,任何方法的任何步骤均可用执行这些步骤的模块、线路或其他装置执行。

[0271] 特定实施例的具体细节可以任何合适方式合并,而不背离本发明的实施例的精神和范围。然而,本发明的其他实施例可涉及与每个个别方面有关的具体实施例、或这些个别方面的具体组合。

[0272] 本发明的示例实施例的上文说明已呈现用于示出和描述的目的。它不预期是详尽的或将本发明限制于所述的精确形式,并且许多修饰和变动根据上文教导是可能的。选择且描述实施例,以便最佳解释本发明的原理及其实际应用,从而致使本领域技术人员在多个实施例中和伴随多个修饰最佳地利用本发明,如适合于考虑的具体用途。

[0273] 除非另有相反说明,“一个”、“一种”或“该”的叙述意指“一个或多个”。

[0274] 上文提及的所有专利、专利申请、出版物和说明全文以引用方式并入本文用于所有目的。无一承认为现有技术。

100

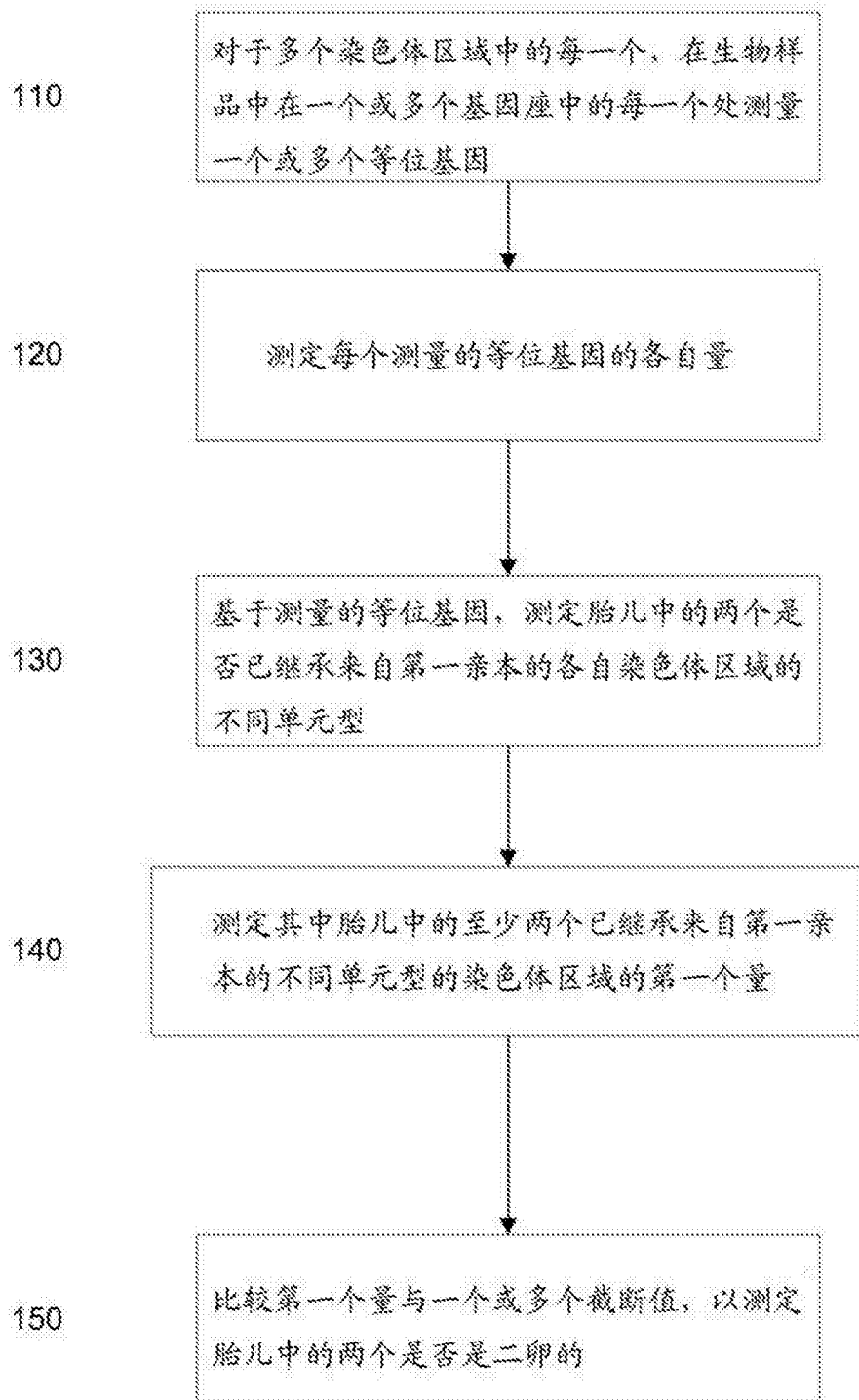


图1

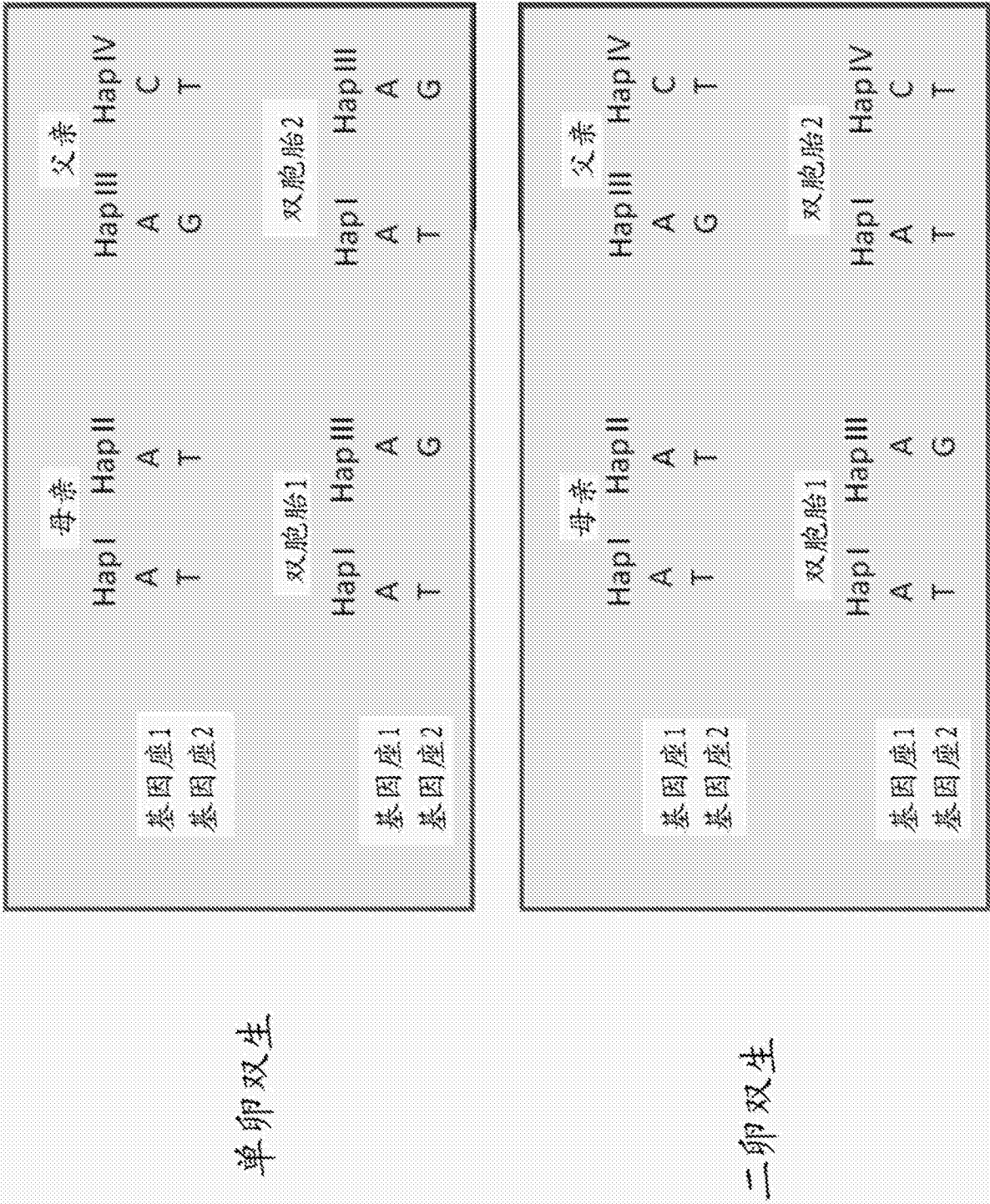


图2A

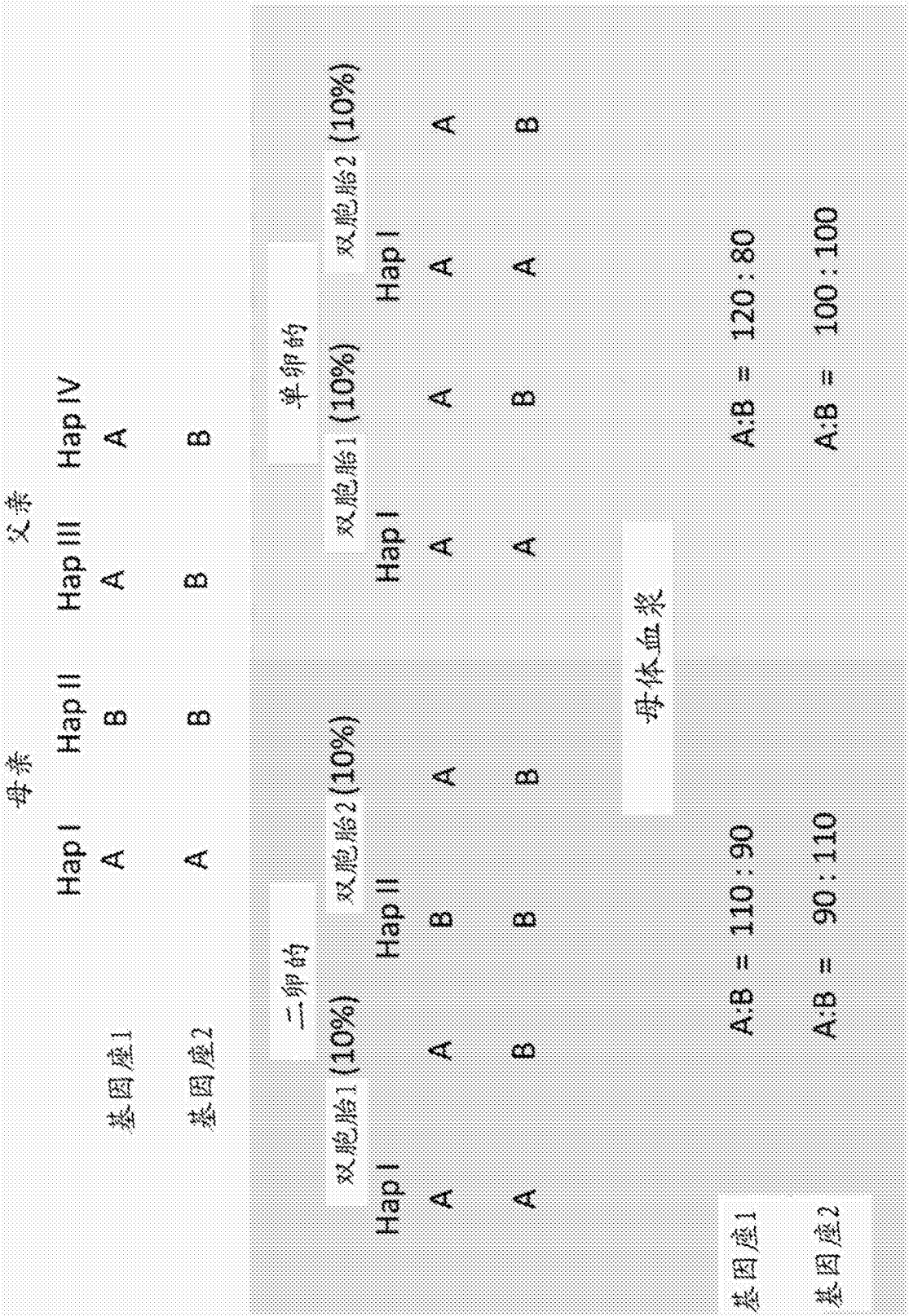


图2B

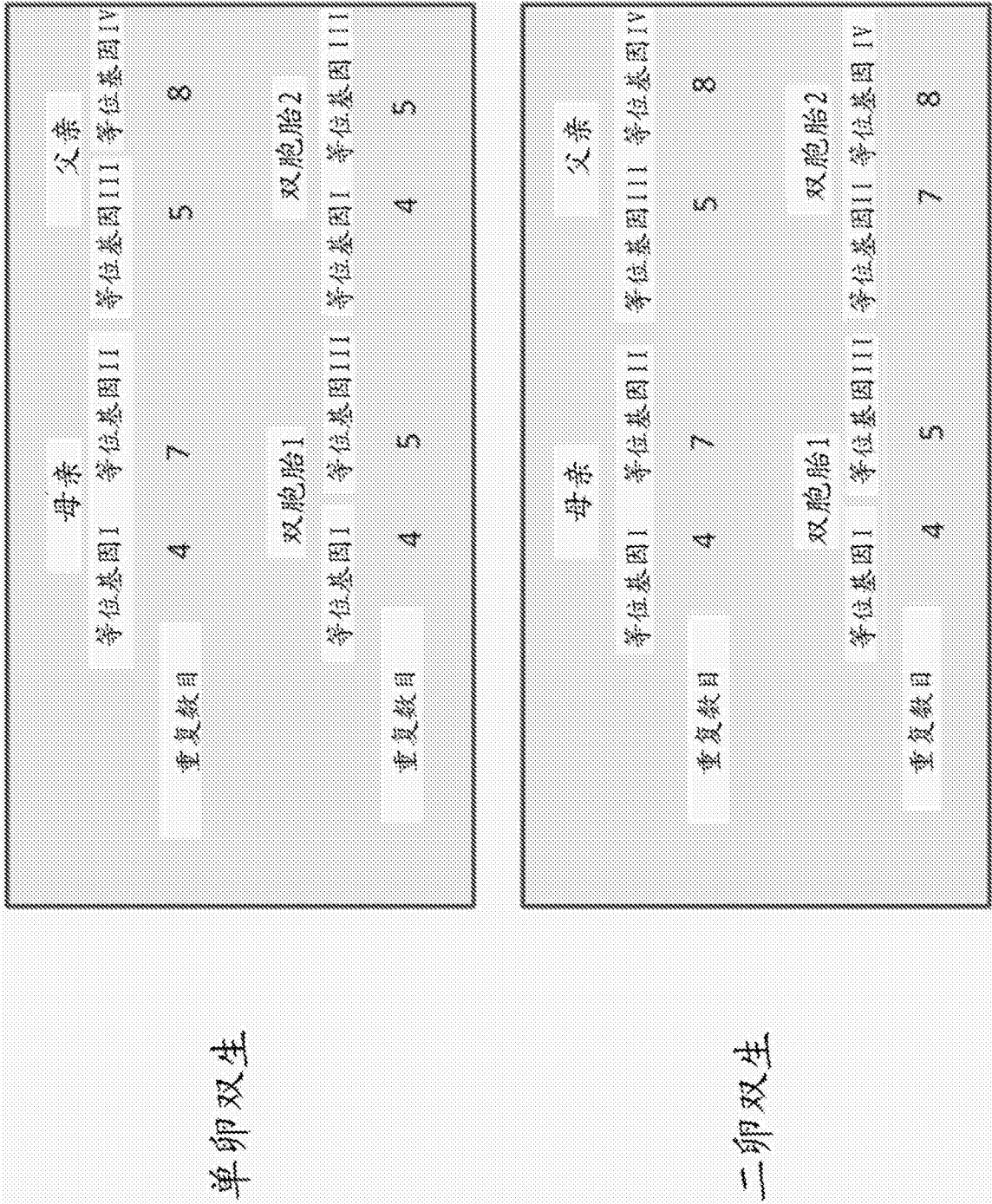


图3

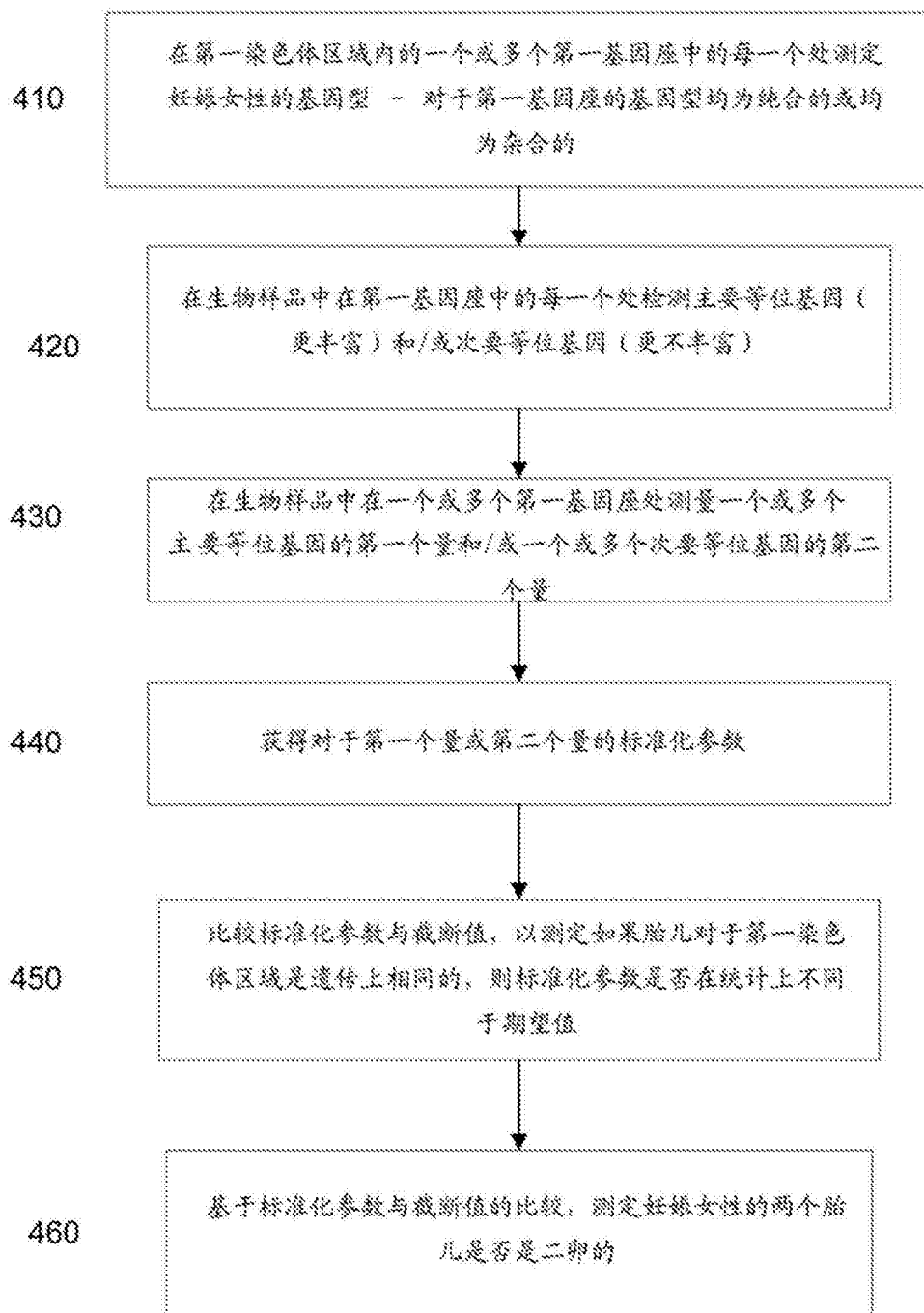


图4

500

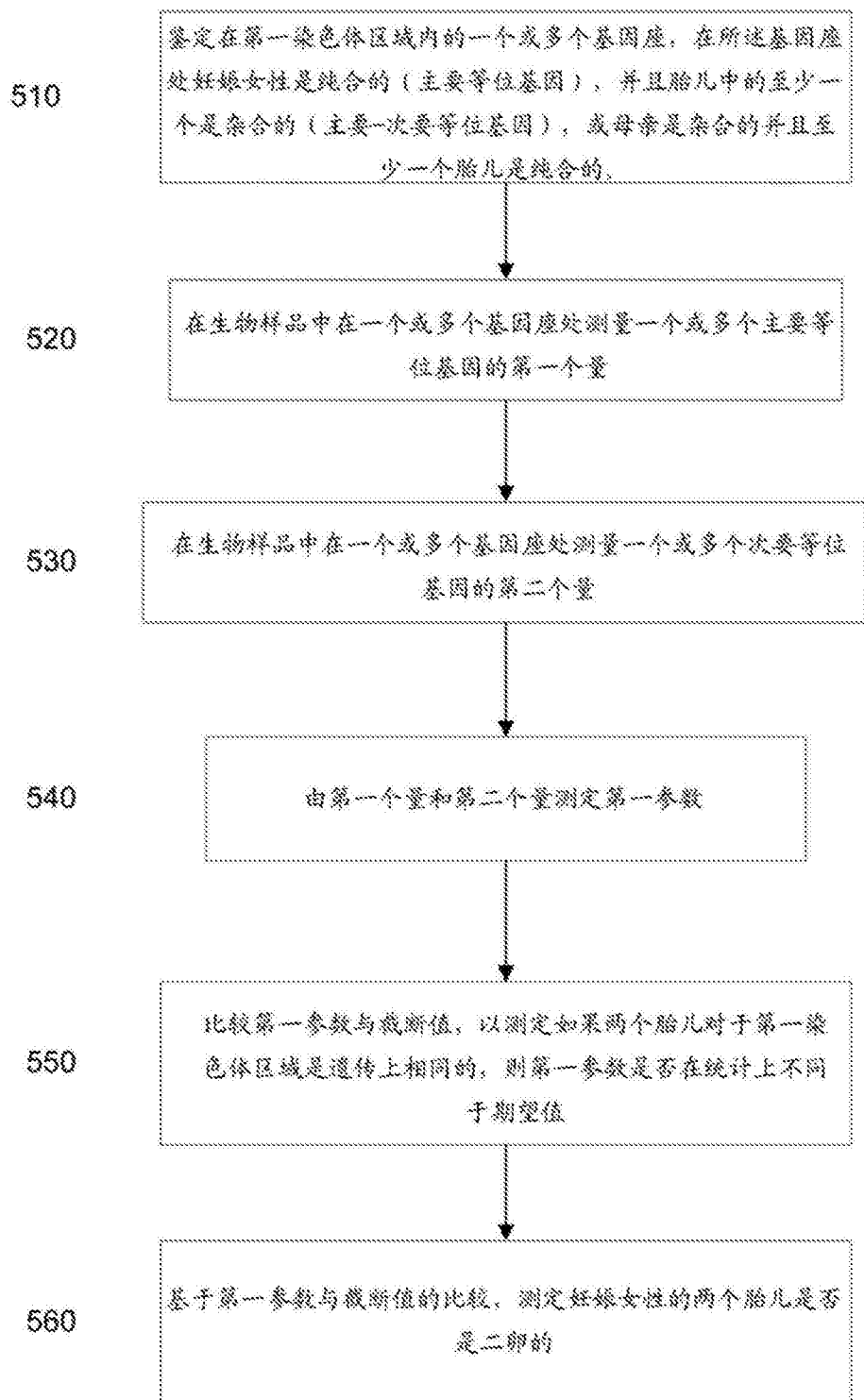


图5

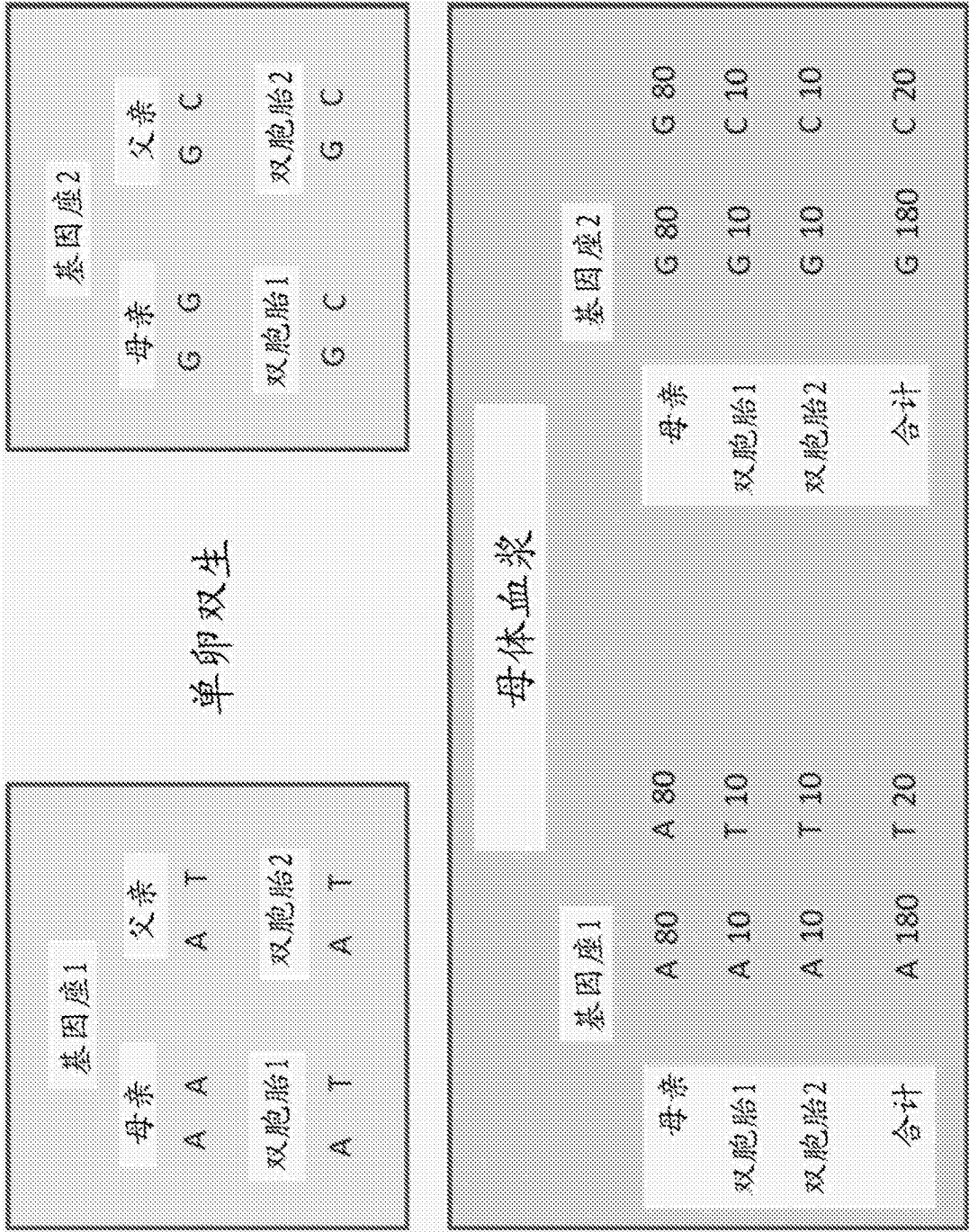


图6

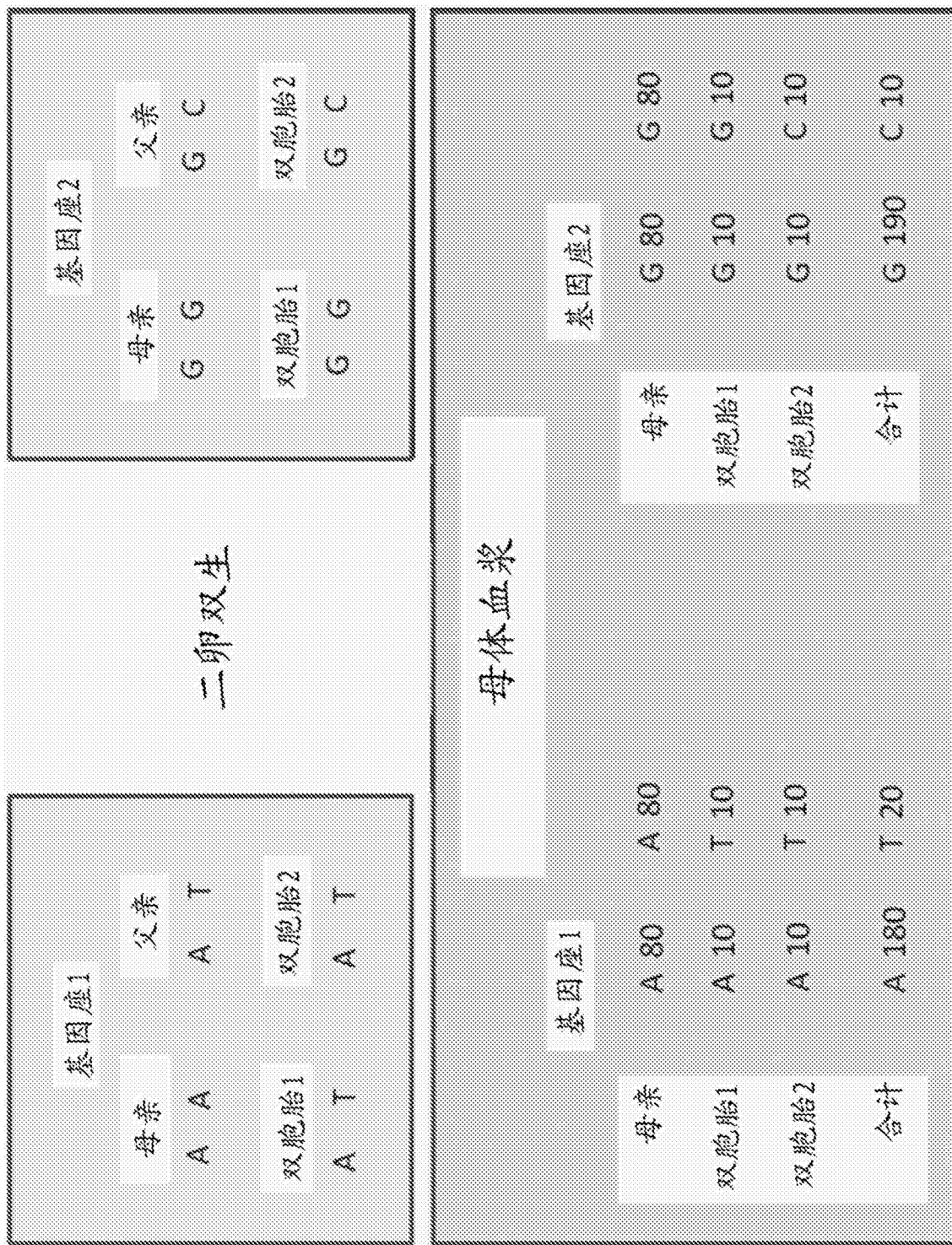


图7

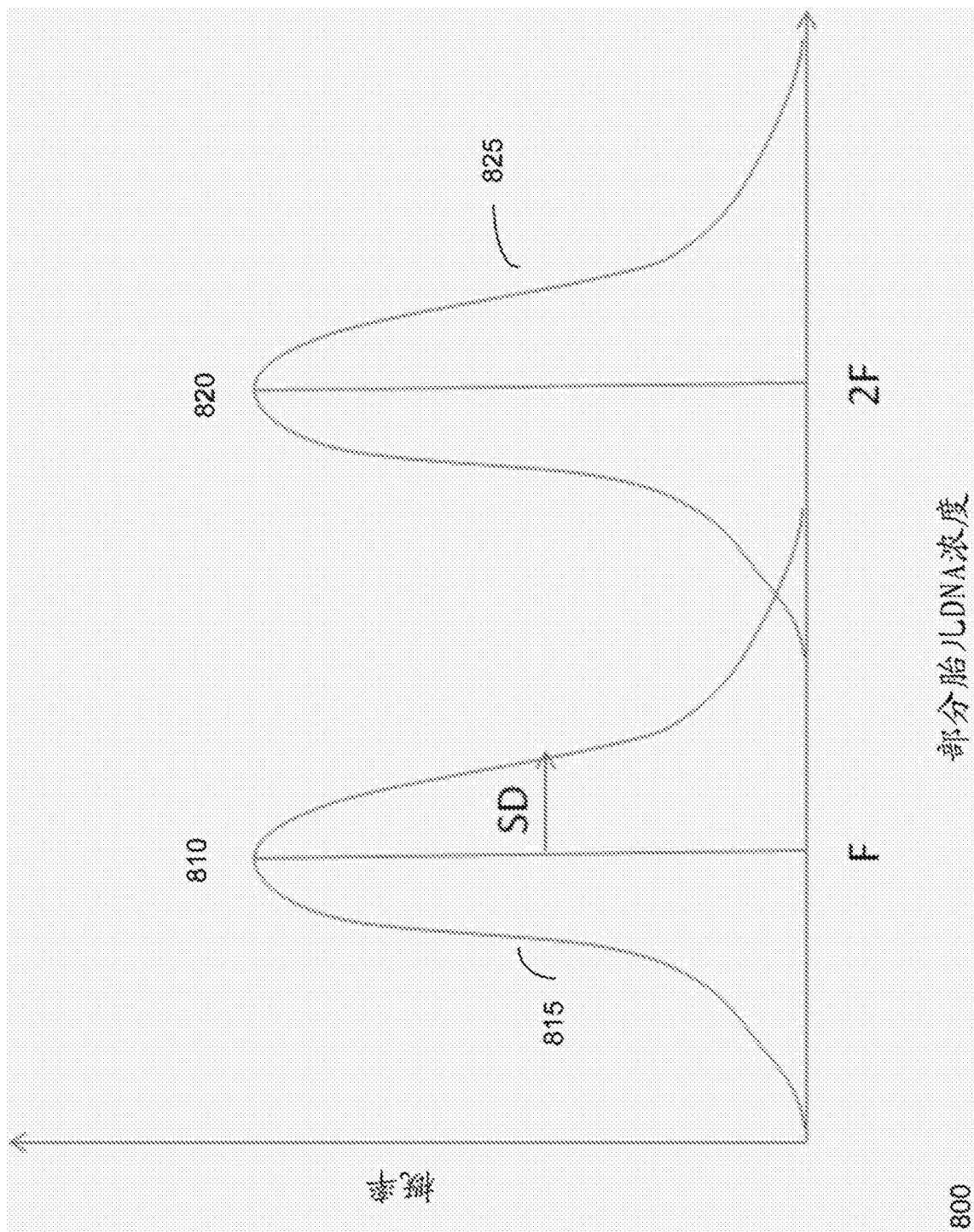


图8

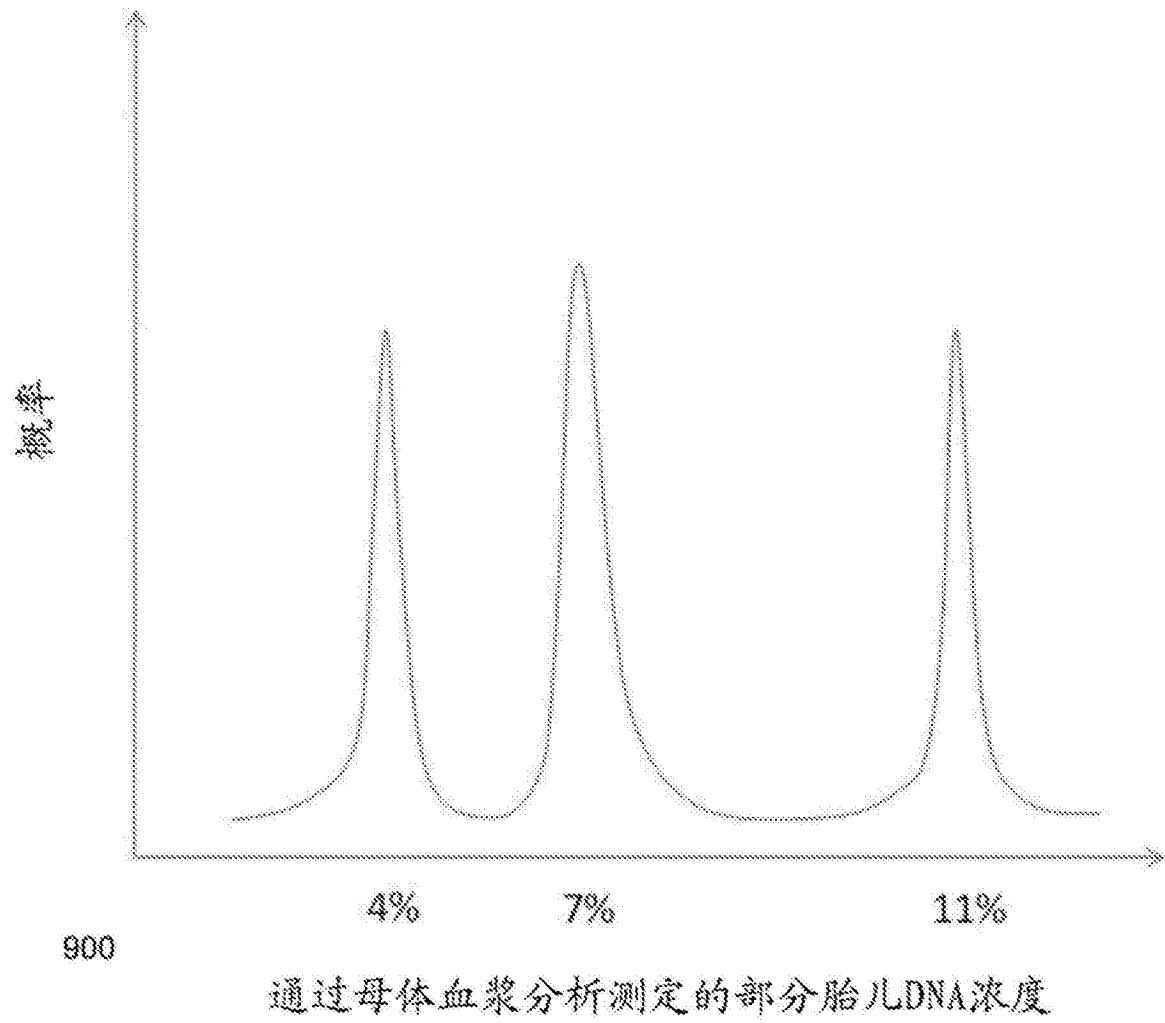


图9

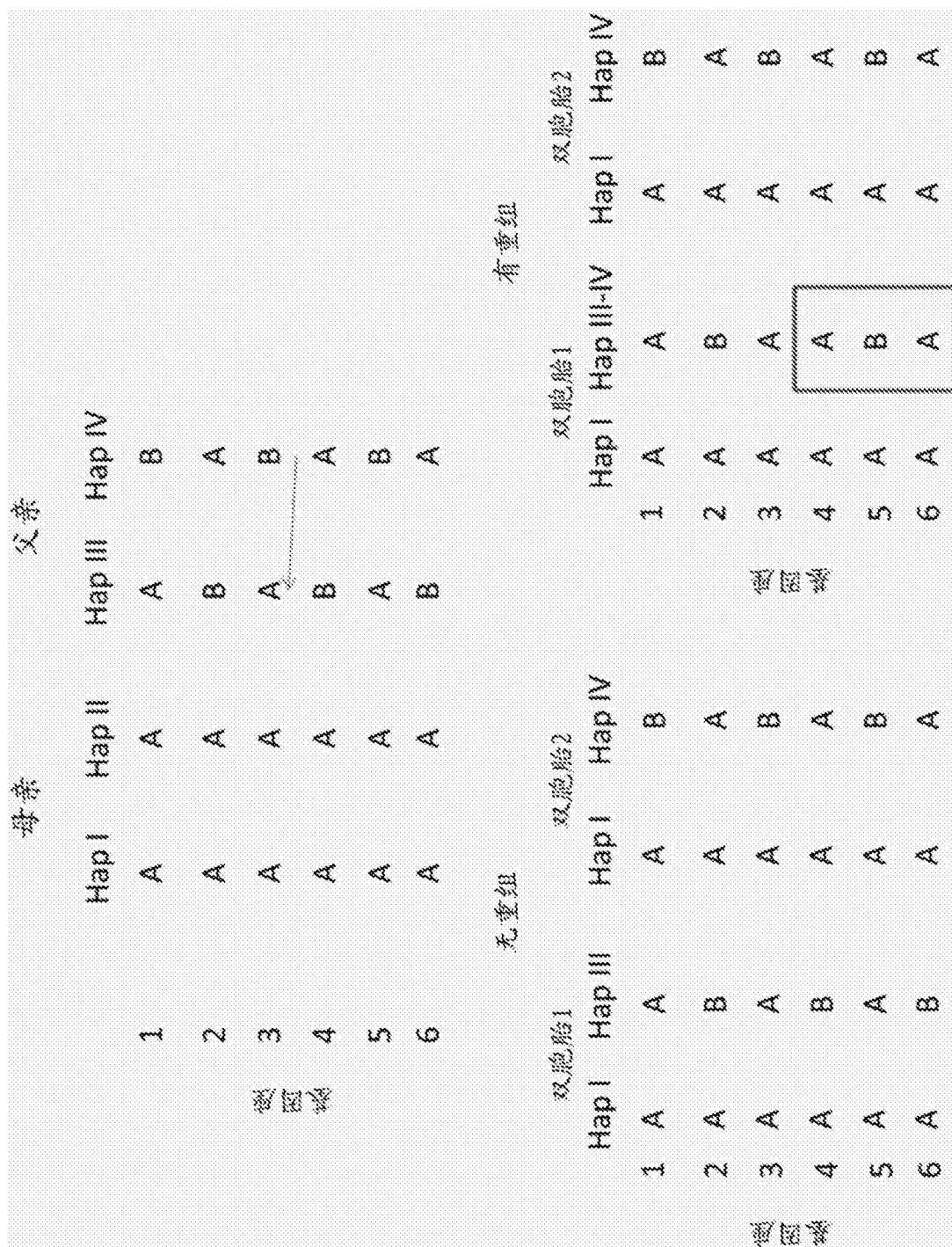


图10

1100

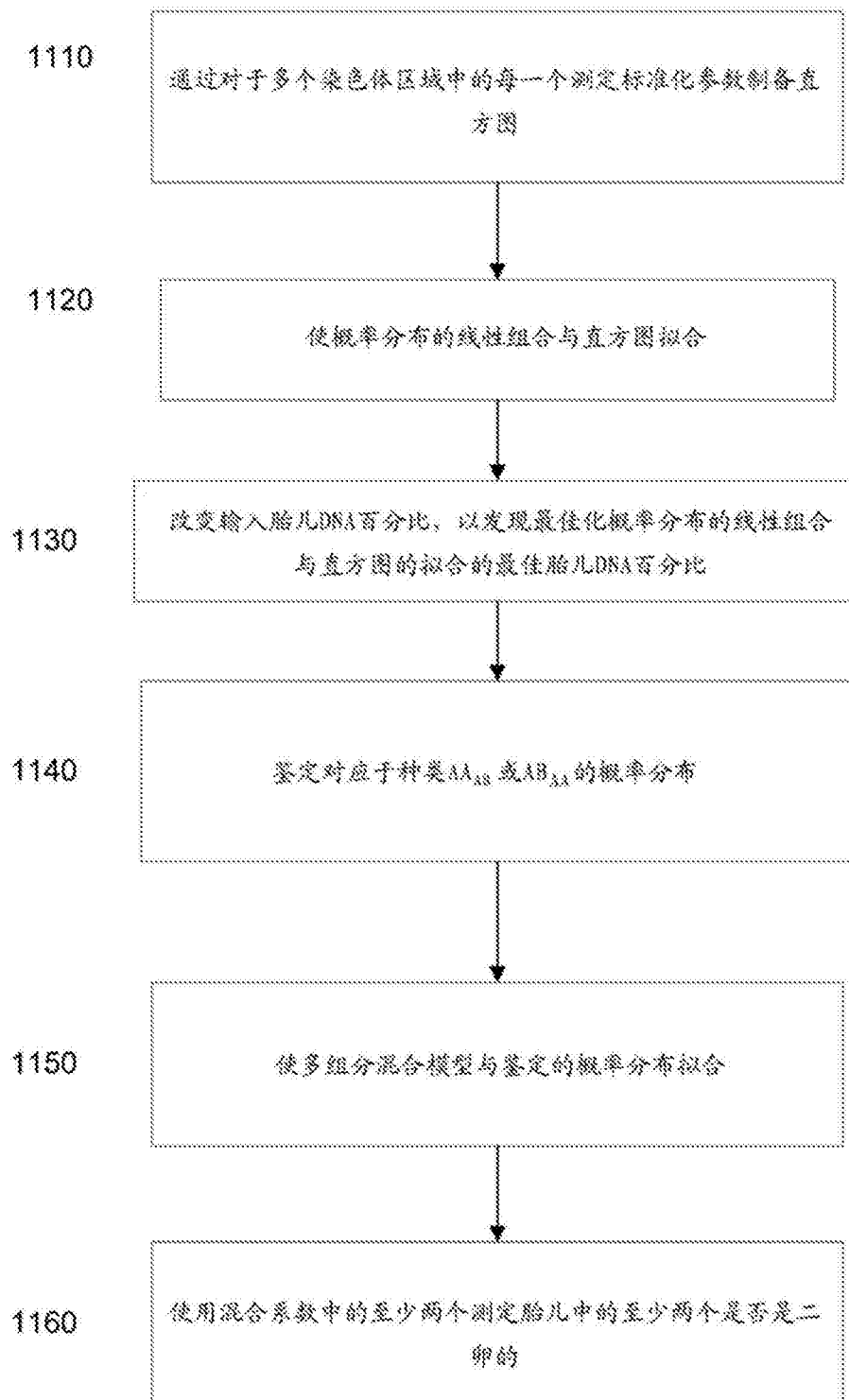


图11

	全基因组		靶标区域	
基因型: 母亲/双胞胎I -双胞胎II	SNP数目	百分比 (%)	SNP数目	百分比 (%)
AA/AA-AA	529473	59.72	9248	34.71
AA/AA-AB	33180	3.74	1098	4.12
AA/AB-AA	33766	3.81	1130	4.24
AA/AB-AB	71528	8.07	4312	16.18
AB/AA-AA	34299	3.87	1739	6.53
AB/AA-AB	61030	6.88	3243	12.17
AB/AA-BB	17067	1.93	375	1.41
AB/AB-AA	61374	6.92	3174	11.91
AB/AB-AB	43967	4.96	2291	8.60

1200

图12A

	全基因组		靶标区域	
基因型: 母亲/胎儿	SNP数目	百分比 (%)	SNP数目	百分比 (%)
AA/AA	573422	63.92	10437	38.96
AA/AB	105511	11.76	5812	21.70
AB/AA	114057	12.72	5408	20.19
AB/AB	103518	11.54	5111	19.08

1220

图12B

总SNP	相同基因型	不同基因型	相同基因型 (%)	不同基因型 (%)
890062	682107	207955	76.64	23.36

1240

图12C

1260		
染色体	DCDA (%)	MCDA (%)
chr1	23.79	17.48
chr2	15.80	15.69
chr3	14.88	16.25
chr4	23.30	17.70
chr5	22.25	15.85
chr7	24.50	11.83
chr8	17.53	15.86
chr9	22.29	18.31
chr13	22.05	16.54
chr15	22.47	16.54
chr17	22.26	16.02
chr19	15.15	17.03
chr20	18.65	15.17
chr22	21.90	16.54
SD	3.36	1.52

图12D

1280

染色体	DCDA (%)	MCDA (%)
chr1	21.42	16.37
chr2	15.37	13.60
chr3	14.45	14.06
chr4	20.07	14.95
chr5	18.63	14.17
chr7	15.82	12.05
chr8	16.91	14.37
chr9	19.27	14.81
chr12	19.28	12.53
chr13	19.88	14.54
chr15	19.76	13.46
chr17	19.64	14.81
chr19	15.48	14.75
chr20	16.33	14.13
chr22	18.73	15.71
SD	2.14	1.10

图12E

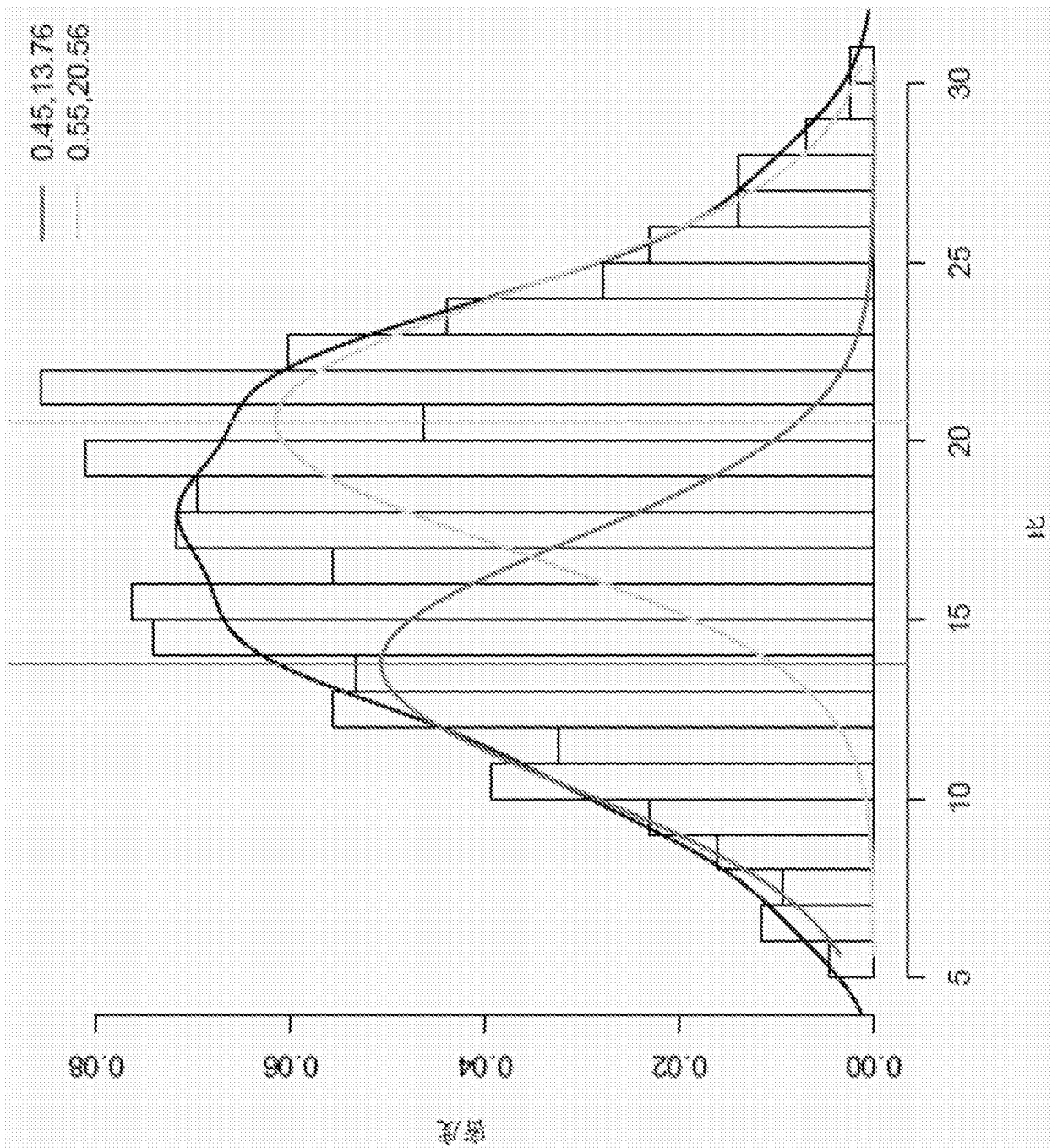


图13

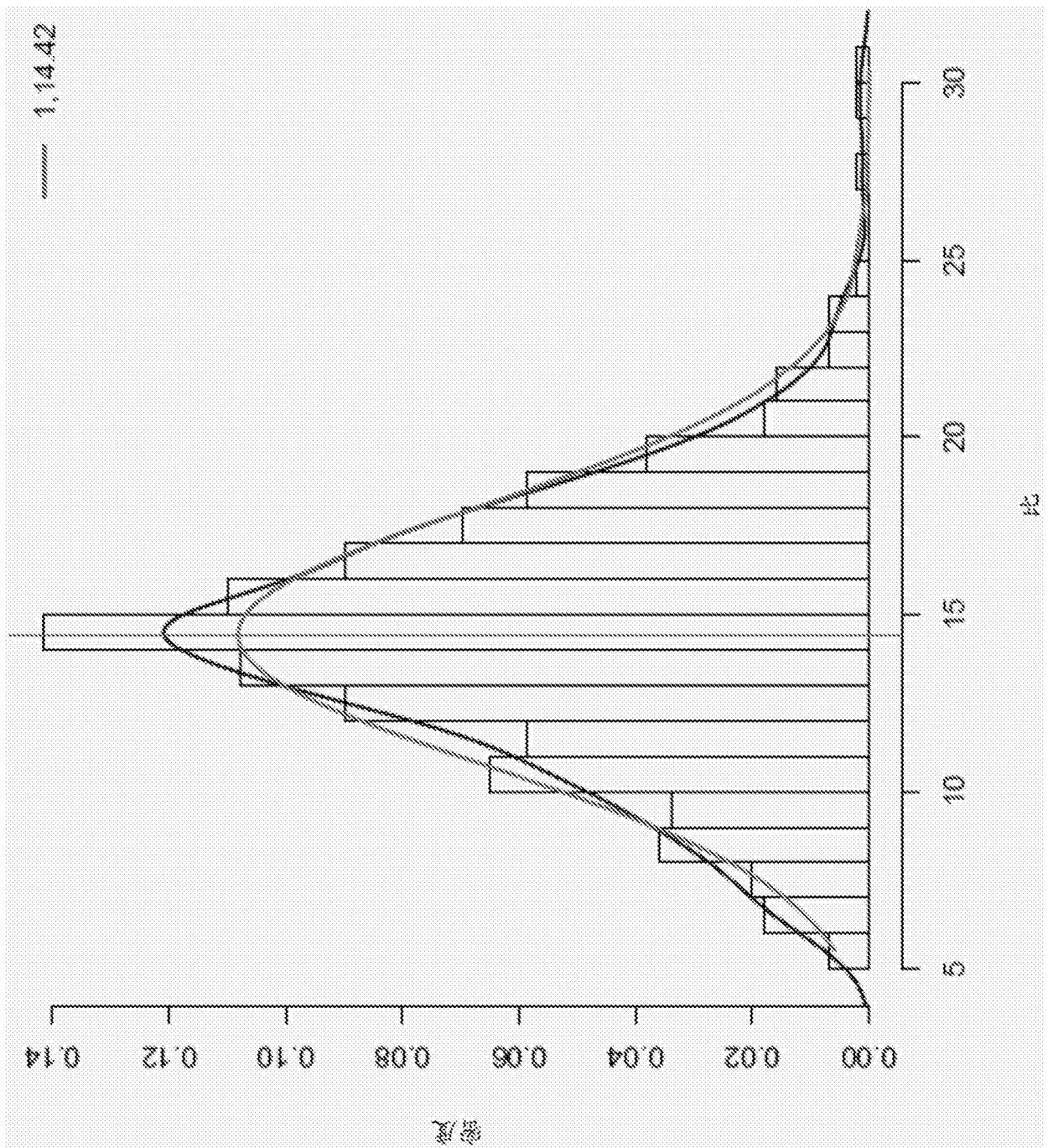


图14

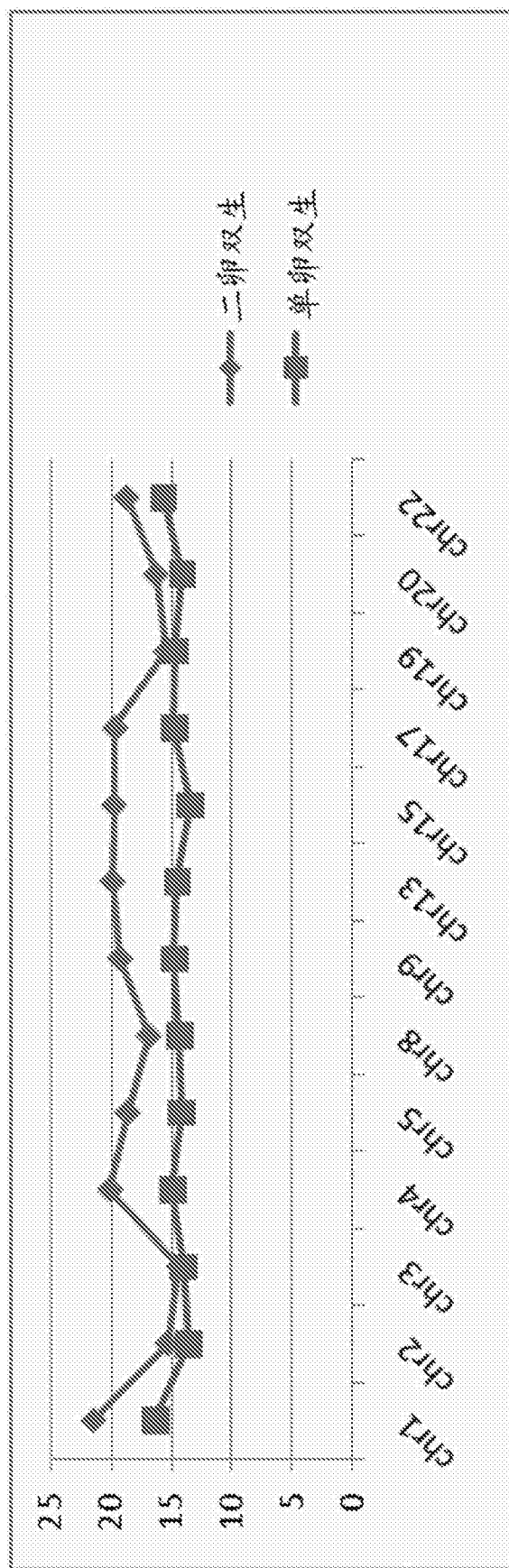


图15

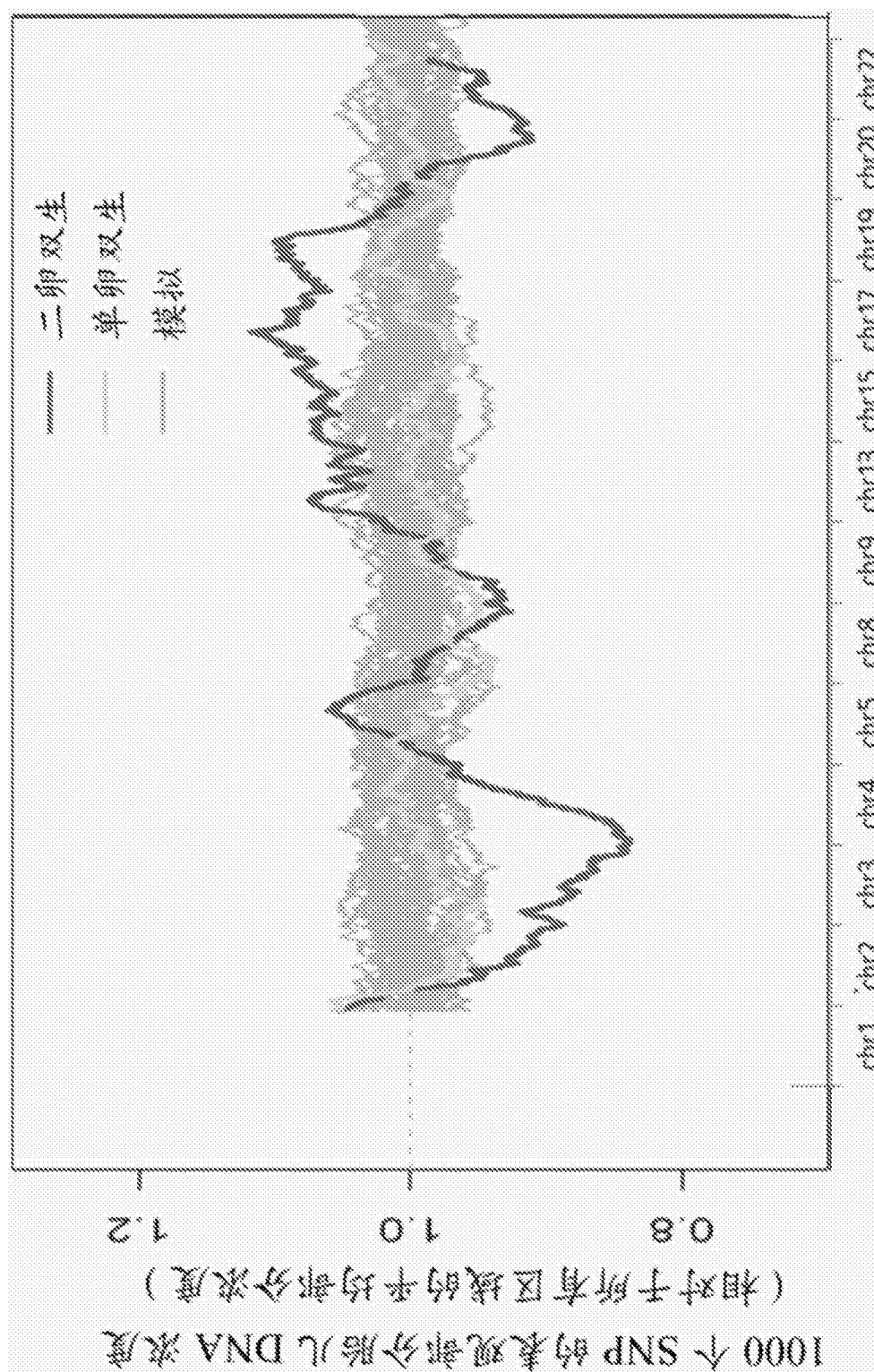


图16

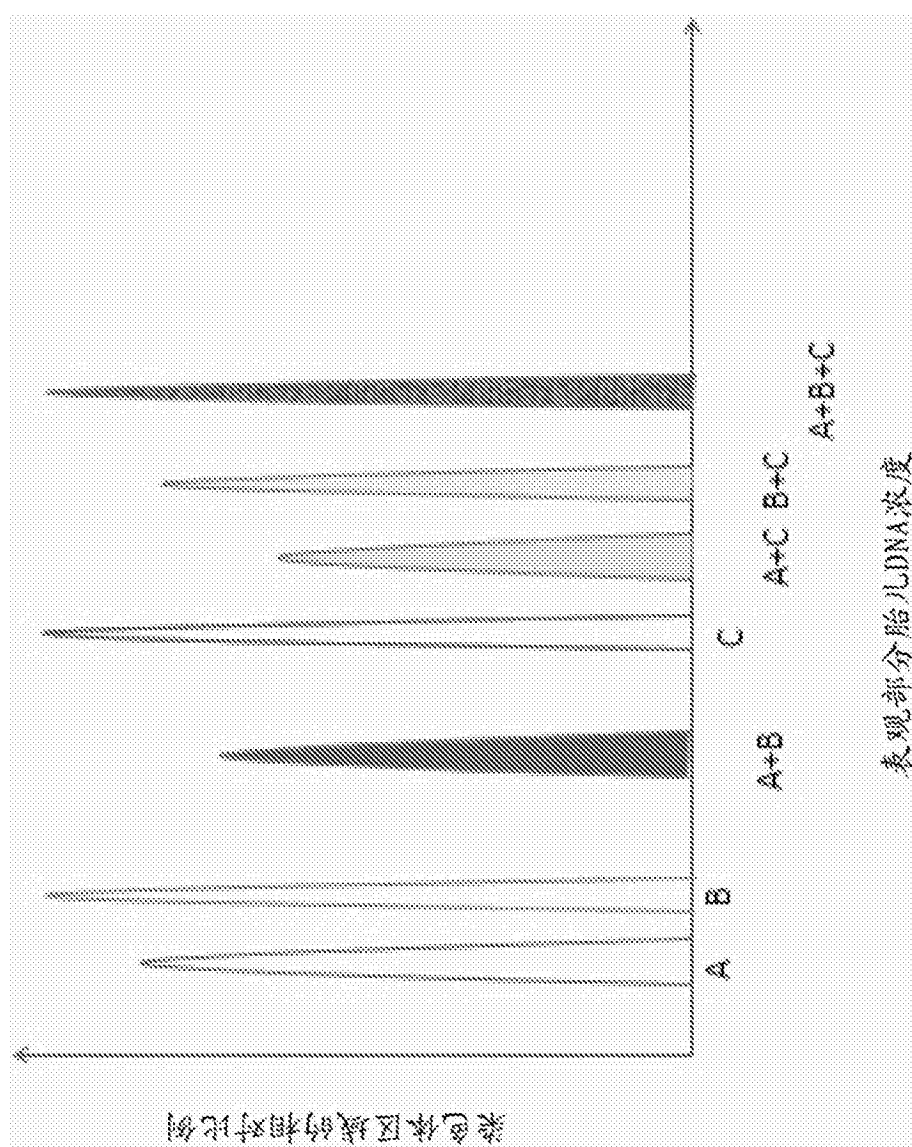


图17

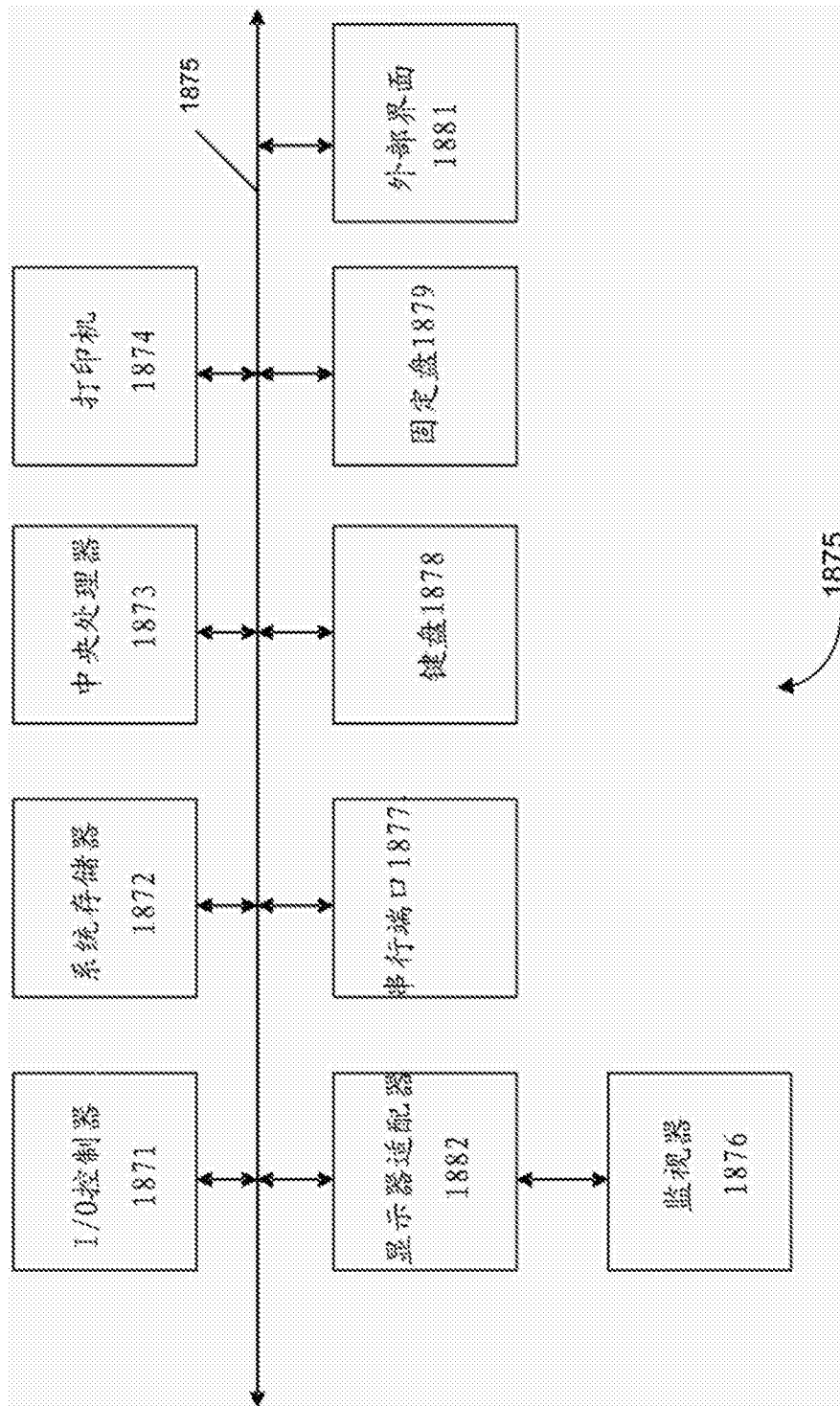


图18