



(12) 发明专利

(10) 授权公告号 CN 101971178 B

(45) 授权公告日 2014. 03. 26

(21) 申请号 200880108126. 3

(22) 申请日 2008. 07. 23

(30) 优先权数据

60/951, 438 2007. 07. 23 US

(85) PCT国际申请进入国家阶段日

2010. 03. 22

(86) PCT国际申请的申请数据

PCT/GB2008/002524 2008. 07. 23

(87) PCT国际申请的公布数据

W02009/013492 EN 2009. 01. 29

(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 徐仲镁

庄家俊

(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.

G06F 19/18 (2011. 01)

C12Q 1/68 (2006. 01)

(56) 对比文件

US 2007/0202525 A1, 2007. 08. 30, 全文 .

CN 1997757 A, 2007. 07. 11, 全文 .

CN 101068937 A, 2007. 11. 07, 全文 .

审查员 陈龙

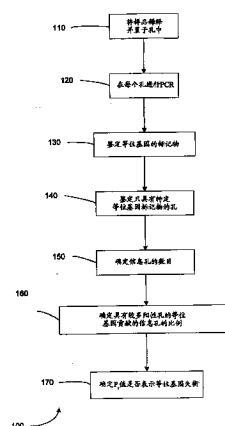
权利要求书3页 说明书41页 附图31页

(54) 发明名称

核酸序列失衡的确定

(57) 摘要

本发明提供了用于确定生物样品中是否存在核酸序列失衡的方法、系统和装置。选取了用于确定,例如,两个序列(或两组序列)的比率的失衡的一个或多个截止值。可以至少部分地基于含有母体核酸序列背景的诸如母体血浆的样品中胎儿DNA的百分比来确定所述截止值。还可以基于每一反应的序列的平均浓度来确定该截止值。在一方面,从估计含有特定核酸序列的信息孔的比例来确定该截止值,其中该比例基于上文所述的百分比和/或平均浓度来确定。可以利用诸如序贯概率比检验(SPRT)的许多不同类型的方法来确定该截止值。



1. 用于确定含核酸分子的生物样品的概率比的设备,所述设备包括:

用于接收来自多个反应的数据的装置,所述多个反应的实施用来指示来自生物样品的核酸分子中感兴趣的多核苷酸序列的存在或缺失,其中每一反应指示临床相关的核酸序列的存在或缺失,以及不同于所述临床相关的核酸序列的背景核酸序列,所述数据包括:

(1) 表示所述临床相关的核酸序列存在的指示的反應的第一量的第一组定量数据;和

(2) 表示所述背景核酸序列存在的指示的反應的第二量的第二组定量数据;

用于测量每一个反应的参考核酸序列的平均浓度的装置,其中所述参考核酸序列是所述临床相关的核酸序列或所述背景核酸序列,其中所述用于测量的装置包括:

用于利用将从所述参考核酸序列的数据导出的值作为输入的概率分布的逆函数来确定所述多个反应的每一个中的所述参考核酸序列的平均浓度的装置;

用于由所述测量的所述多个反应的每一个中的参考核酸序列的平均浓度,以及临床相关核酸序列与背景核酸序列之间预期的比率,来确定概率比的装置,其中所述用于确定概率比的装置包括:

用于乘以所测量的平均浓度和所述预期的比率以获得第二平均浓度的装置;

用于利用所述测量的平均浓度作为输入的概率分布来确定含有至少一种参考核酸序列的任何反应的第一概率的装置,

用于利用所述第二平均浓度作为输入的概率分布来确定含有至少一种非参考核酸序列的任何反应的第二概率的装置,以及

用于计算所述第一概率和所述第二概率的比率的装置。

2. 如权利要求 1 所述的设备,其中从一个或多个第一标记物获得所述第一组数据,并且其中从一个或多个第二标记物获得所述第二组数据,所述第一标记物的每个检测一部分所述临床相关的核酸序列在反应中存在,所述第二标记物的每个检测一部分所述背景核酸序列在反应中存在。

3. 如权利要求 1 所述的设备,其中所述临床相关的核酸序列来自染色体 21,并且所述背景核酸序列来自染色体 21 以外的染色体。

4. 如权利要求 1 所述的设备,其中所述临床相关的核酸序列来自染色体 18 或染色体 13,并且所述背景核酸序列分别来自染色体 18 以外的或染色体 13 以外的染色体。

5. 如权利要求 1 所述的设备,其中所述临床相关的核酸序列是遗传多态性的等位基因,并且所述背景核酸序列是所述遗传多态性的另一等位基因。

6. 如权利要求 1 所述的设备,其中所述临床相关的核酸序列是囊性纤维化跨膜传导调节因子基因、 β -球蛋白基因或 α -球蛋白基因的突变拷贝,并且所述背景核酸序列来自相应基因的野生型拷贝。

7. 如权利要求 1 所述的设备,其中所述生物样品是来自孕妇的血浆或血清。

8. 如权利要求 1 所述的设备,其中反应是扩增反应。

9. 如权利要求 8 所述的设备,其中反应是数字 PCR 方法的一部分。

10. 如权利要求 1 所述的设备,其中反应是测序反应。

11. 如权利要求 1 所述的设备,其中所述临床相关的核酸序列的第一部分和所述背景核酸序列的第一部分来自第一个体,并且所述临床相关的核酸序列的第二部分和所述背景核酸序列的第二部分来自第二个体。

12. 如权利要求 11 所述的设备,其中所述预期的比率取决于所述第一部分的一个的测量值或基于所述第二部分的一个的测量值。

13. 如权利要求 1 所述的设备,其中所述用于确定所述概率比的装置包括包括用于使用序贯概率比检验、假发现率、置信区间和接收器工作特性曲线中的至少一种的装置。

14. 如权利要求 1 所述的设备,其中用于确定所述概率比的装置包括:

用于利用所测量的平均浓度和所预期的比率来确定预期含有过度表现的核酸序列的信息反应的比例 P1 的装置,所述过度表现的核酸序列是所述参考核酸序列或非参考核酸序列;以及

用于从所述比例 P1 计算所述概率比的装置。

15. 如权利要求 14 所述的设备,其中用于确定所述比例 P1 的装置包括:

用于确定含有至少一种所述过度表现的核酸序列的反应的第三概率的装置,所述第三概率是所述第一概率或所述第二概率;

用于计算反应是信息反应的第四概率的装置;以及

用于利用所述第三概率和所述第四概率来计算所述比例 P1 的装置。

16. 如权利要求 15 所述的设备,其中所述第一概率是通过将所述参考核酸序列的平均浓度乘以相对于所述非参考核酸序列的预期比率来确定的。

17. 如权利要求 15 所述的设备,其中所述反应是信息反应的第四概率的计算包括假定所述第一概率和所述第二概率是独立的。

18. 用于确定含核酸分子的生物样品的概率比的设备,所述设备包括:

用于接收来自多个反应的数据的装置,所述多个反应的实施用来指示来自生物样品的核酸分子中感兴趣的多核苷酸序列的存在或缺失,其中每一反应指示临床相关的核酸序列的存在或缺失,以及不同于所述临床相关的核酸序列的背景核酸序列,所述数据包括:

(1) 表示所述临床相关的核酸序列存在的指示的反应的第一量的第一组定量数据;和

(2) 表示所述背景核酸序列的第二量的第二组定量数据;其中所述临床相关的核酸序列和所述背景核酸序列来自第一类型的细胞和来自一种或多种第二类型的细胞;

用于确定来自核酸序列的量的测量产生的第一百分比的临床相关核酸序列与背景核酸序列之间的概率比的装置,所述核酸序列来自所述生物样品中的所述第一类型的细胞。

19. 如权利要求 18 所述的设备,其中所述第一类型的细胞来自第一有机体,并且所述第二类型的细胞来自第二有机体。

20. 如权利要求 18 所述的设备,其中用于确定所述概率的装置包括:

用于确定每个反应的参考核酸序列的第一平均浓度的装置,其中所述参考核酸序列是过少表现的所述临床相关的核酸序列或所述背景核酸序列;以及

用于将所述第一平均浓度乘以从所述第一百分比导出的因子来获得不是所述参考核酸序列的所述核酸序列的第二平均浓度的装置。

21. 如权利要求 20 所述的设备,其还包括:

用于利用将从所述参考核酸序列的数据导出的值作为输入的概率分布的逆函数来确定所述多个反应的每一个中的所述参考核酸序列的平均浓度的装置。

22. 如权利要求 21 所述的设备,其中所述概率分布是泊松分布。

23. 如权利要求 18 所述的设备,其中利用定量实时 PCR、数字 PCR、半定量竞争 PCR、实

时竞争 PCR 或质谱,通过测定胎儿特异性标记物的量来测量所述百分比。

核酸序列失衡的确定

[0001] 优先权的要求

[0002] 本申请要求于 2007 年 7 月 23 日提交的、题目为“核酸序列失衡的测定”的第 60/951438 号美国临时申请（代理公司案卷号 016285-005200US）的优先权，并且是所述临时申请的正式申请，该临时申请的全部内容通过引用的方式并入本文用于所有目的。

[0003] 相关申请的交叉引用

[0004] 本申请还涉及同时提交的、题目为“利用基因组测序诊断胎儿染色体非整倍性”的正式申请（代理公司案卷号 016285-005220US），该正式申请的全部内容通过引用的方式并入本文用于所有目的。

发明领域

[0005] 本发明一般地涉及通过确定两个不同核酸序列之间的失衡对基因型和疾病进行的诊断检测，更具体地，涉及通过检测母体血液样品对胎儿的唐氏综合征、其它染色体非整倍性、突变和基因型的鉴定。本发明还涉及癌症的检测、移植的监测和传染病监测。

[0006] 发明背景

[0007] 遗传疾病、癌症和其它病况通常由两个对应的染色体或等位基因或其它核酸序列中的失衡导致或产生两个对应的染色体或等位基因或其它核酸序列中的失衡。也就是说，一个序列相对于另一序列的量大于或小于正常值。通常地，正常比率恰好是 50/50 的比率。唐氏综合征（21 三体性）是具有额外的染色体 21 失衡的这类疾病。

[0008] 21 三体性的常规产前诊断方法包括通过诸如羊膜穿刺取样或绒毛膜绒毛取样的侵入性操作的胎儿物质的取样，这引起胎儿丢失的有限风险。诸如通过超声波扫描术和生化标记物的筛查的无创性方法已经用于在确定性的侵入性诊断方法前对孕妇进行风险分级（risk-stratify）。然而，这些筛查方法通常测量与 21 三体性有关的附带现象，而不是核心染色体异常，因此该筛查方法的诊断准确性不是最佳的，并且具有其它劣势，例如受孕龄影响大。

[0009] 1997 年发现的母体血浆中循环的无细胞胎儿 DNA 为无创产前诊断提供了新的可能性（Lo, YMD and Chiu, RWK 2007 Nat Rev Genet 8, 71-77）。尽管这种方法已经容易地应用于性连锁（Costa, JM et al. 2002 N Engl J Med 346, 1502）和某些单基因病症（Lo, YMD et al. 1998 N Engl J Med 339, 1734-1738）的产前诊断，但是该方法在胎儿染色体非整倍性的产前检测的应用表现出相当的挑战（Lo, YMD and Chiu, RWK 2007, 见上文）。首先，胎儿核酸与经常能够干扰分析的母体来源的核酸的高背景共同存在于母体血浆中（Lo, YMD et al. 1998 Am J Hum Genet 62, 768-775）。其次，胎儿核酸主要以无细胞形式在母体血浆中循环，这使得难以获得胎儿基因组中的基因或染色体的剂量信息。

[0010] 最近实现了克服这些挑战的明显发展（Benachi, A & Costa, JM 2007 Lancet 369, 440-442）。一种方法检测母体血浆中的胎儿特异性核酸，从而克服了母体背景干扰的问题（Lo, YMD and Chiu, RWK 2007, 见上文）。从源自胎盘的 DNA/RNA 分子中的多态性等位基因的比率来推断染色体 21 的剂量。然而，当样品含有较低量的靶向的基因时，这种方法较不

准确,并且只能应用于对靶向的多态性是杂合的胎儿,如果使用了一种多态性,则该靶向的多态性只是群体的子集。

[0011] Dhallan 等人 (Dhallan, R, et al. 2007, 见上文, Dhallan, R, et al. 2007 Lancet 369, 474-481) 描述了通过向母体血浆中添加甲醛来富集循环的胎儿 DNA 比例的替代策略。通过评价对于染色体 21 上的单核苷酸多态性 (SNP) 遗传自父亲的胎儿特异性等位基因比非胎儿特异性等位基因的比率,来确定由母体血浆中胎儿贡献的染色体 21 序列的比例。类似地计算参考染色体的 SNP 比率。然后通过检测染色体 21 的 SNP 比率与参考染色体的 SNP 比率之间统计学的显著差异来推断胎儿染色体 21 的失衡,其中使用确定的小于 0.05 的 p 值来定义显著。为了保证高群体覆盖,靶向每个染色体多于 500 个的 SNP。然而,对富集高比例的甲醛的有效性仍有争议 (Chung, GTY, et al. 2005 Clin Chem 51, 655-658), 因此,该方法的可重复性需要进一步的评价。此外,由于每个胎儿和母亲将提供每个染色体的不同数目的 SNP 的信息,所以 SNP 比率比较的统计学检验的效能在不同个例之间是可变的 (Lo, YMD & Chiu, RWK. 2007 Lancet 369, 1997)。而且,由于这些方法依赖于遗传多态性的检测,所以它们局限于对这些多态性是杂合的胎儿。

[0012] 利用聚合酶链式反应 (PCR) 以及从 21 三体性胎儿和整倍体胎儿获得的羊膜细胞 (amniocyte) 培养物中的染色体 21 基因座和参考基因座的 DNA 定量,基于 21 三体性胎儿中染色体 21 的 DNA 序列的 1.5 倍的增加, Zimmermann 等人 (2002 Clin Chem 48, 362-363) 能够区分这两组胎儿。由于 DNA 模板浓度的 2 倍差异组成了只有一个阈值循环的差别 (Ct), 所以 1.5 倍差异的鉴别已经是常规实时 PCR 的极限。为了实现更精细程度的定量鉴别,亟需替代的策略。因此,出于这一目的,本发明的某些实施方案使用数字 PCR (Vogelstein, B et al. 1999 Proc Natl Acad Sci USA 96, 9236-9241)。

[0013] 已经开发了数字 PCR 来检测核酸样品中偏移的等位基因比率 (Chang, HW et al. 2002 J Natl Cancer Inst 94, 1697-1703)。数字 PCR 在临床上已经被证实对于检测肿瘤 DNA 样品中的杂合性丢失 (LOH) 是有用的 (Zhou, W. et al. 2002 Lancet 359, 219-225)。对于数字 PCR 结果分析,以前的研究采用了序贯概率比检验 (SPRT) 来将实验结果分类为提示样品中存在 LOH 或不存在 LOH (El Karoui et al. 2006 Stat Med 25, 3124-3133)。在以前的研究所用的方法中,确定 LOH 的截止值 (cutoff value) 使用了 DNA 中两个等位基因的固定参考比率,该比率为 2/3。由于母体血浆中胎儿核酸的量、比例和浓度是可变的,所以这些方法对于使用母体血浆中的母体核酸背景中的胎儿核酸来检测 21 三体性是不合适的。

[0014] 期望具有基于循环的胎儿核酸分析的胎儿 21 三体性 (和其它失衡) 检测的无创检测,特别是不依赖于遗传多态性和 / 或胎儿特异性标记物的使用的无创检测。还期望具有截止值和序列记数的准确测定,这能够减少准确性所需的数据孔的数目和 / 或母体血浆核酸分子的量,从而提供了增加的效率和成本效益。还期望该无创检测具有高灵敏度和特异性以将误诊断降至最低。

[0015] 母体血浆中胎儿 DNA 检测的另一应用是单基因病症的产前诊断,例如 β -地中海贫血症。然而,由于胎儿 DNA 只组成了母体血浆 DNA 的一小部分,所以这种方法被认为只能够检测胎儿从其父亲遗传但是其母亲没有的突变。这种突变的实例包括导致 β -地中海贫血症的 β -球蛋白基因的密码子 41/42 的 4bp 缺失 (Chiu RWK et al. 2002 Lancet, 360, 998-1000) 和导致囊性纤维化的囊性纤维化跨膜传导调节因子基因的 Q890X 突变

(Gonzalez-Gonzalez et al. 2002 Prenat Diagn, 22, 946-8)。然而, 由于 β -地中海贫血症和囊性纤维化都是常染色体隐性条件的, 其中在该疾病自身显现前, 胎儿需要继承来自双亲中每个的突变, 所以只检测遗传自父亲的突变只会使得胎儿患有该疾病的风险从 25% 增加至 50%。这在诊断上不是理想的。因此, 当胎儿能够被排除具有纯合疾病状态时, 现有方法的主要诊断应用是用于在母体血浆中不能检测到遗传自父亲的胎儿突变的情况。然而, 这种方法在诊断上的劣势是, 结论是基于父亲突变的阴性检测做出的。因此, 允许从母体血浆中确定完整的胎儿基因型 (纯合正常、纯合突变体或杂合) 而没有上文的限制的方法是非常理想的。

[0016] 发明简述

[0017] 本发明的实施方案提供了用于确定在生物样品中是否存在核酸序列失衡 (例如, 等位基因失衡、突变失衡或染色体失衡) 的方法、系统和装置。例如, 选择了用于确定两个序列 (或两组序列) 的量的比率的失衡的一个或多个截止值。

[0018] 在一实施方案中, 至少部分地基于诸如母体血浆或血清或尿的含有母体核酸序列背景的生物样品中的胎儿 (临床相关的核酸) 序列的百分比来确定所述截止值。在另一实施方案中, 基于多个反应中的序列的平均浓度来确定所述截止值。在一方面, 从估计含有特定核酸序列的信息孔的比例来确定所述截止值, 其中该比例是基于上文所述的百分比和 / 或平均浓度来确定的。

[0019] 可以使用许多不同类型的方法来确定所述截止值, 例如 SPRT、假发现 (false discovery)、置信区间、接收器工作特性 (receiver operating characteristic) (ROC)。这种策略还在能够做出置信分类 (confident classification) 前将检测所要求的量降至最少。这种策略与模板的量通常是有限的血浆核酸分析是特别相关的。

[0020] 根据一示例性实施方案, 提供了用于确定生物样品中是否存在核酸序列失衡的方法, 该方法包括: 接收来自多个反应的数据, 其中该数据包括: (1) 表明临床相关的核酸序列的第一量的第一组定量数据; 和 (2) 表明不同于所述临床相关的核酸序列的背景核酸序列的第二量的第二组定量数据; 从这两个数据组来确定参数; 从多个反应的每一个中的参考核酸序列的平均浓度导出第一截止值, 其中该参考核酸序列是所述临床相关的核酸序列或所述背景核酸序列; 将所述参数与所述第一截止值比较; 并且, 基于该比较来确定是否存在核酸序列失衡的分类。

[0021] 根据另一示例性实施方案, 提供了用于确定生物样品中是否存在核酸序列失衡的方法, 该方法包括: 接收来自多个反应的数据, 其中该数据包括: (1) 表明临床相关的核酸序列的第一量的第一组定量数据; 和 (2) 表明不同于所述临床相关的核酸序列的背景核酸序列的第二量的第二组定量数据, 其中, 所述临床相关的核酸序列和所述背景核酸序列来自第一类型的细胞和来自一种或多种第二类型的细胞; 从这两个数据集来确定参数; 从得自核酸序列的量的测量的第一百分比导出第一截止值, 该核酸序列来自生物样品中所述第一类型的细胞; 将所述参数与所述截止值比较; 并且, 基于该比较来确定是否存在核酸序列失衡的分类。

[0022] 本发明的其它实施方案涉及与本文所述的方法相关的系统和计算机可读取的介质。

[0023] 参照下文的发明详述和附图将更好地理解本发明的特性和优势。

[0024] 附图简述

[0025] 图 1 是示出数字 PCR 实验的流程图。

[0026] 图 2A 示出本发明实施方案的数字 RNA-SNP 和 RCD 方法。

[0027] 图 2B 显示了在癌症中可频繁检测到的染色体畸变的实例的表格。

[0028] 图 3 示出按照本发明的实施方案用于确定唐氏综合征的具有 SPRT 曲线的图。

[0029] 图 4 显示了按照本发明的实施方案利用胎儿细胞百分比来确定疾病状态的方法。

[0030] 图 5 显示了按照本发明的实施方案利用平均浓度来确定疾病状态的方法。

[0031] 图 6 显示的表格的列出了按照本发明的实施方案对于表示为每孔的平均参考模板浓度 (m_r) 的一系列模板浓度而言, 21 三体性样品的预期数字 RNA-SNP 等位基因比率和 P_r 。

[0032] 图 7 显示的表格列出了按照本发明的实施方案对于表示为每孔的平均参考模板浓度 (m_r) 的一系列模板浓度而言, 21 三体性样品中的 10%、25%、50% 和 100% 的部分胎儿 DNA 浓度的预期 P_r 。

[0033] 图 8 显示的图示出了按照本发明的实施方案, 数字 RNA-SNP 分析的 0.1、0.5 和 1.0 的 m_r 值的 SPRT 曲线的差异程度。

[0034] 图 9A 显示了按照本发明的实施方案在 96 孔数字 RNA-SNP 分析中比较用于分类整倍体和 21 三体性实例的新和旧 SPRT 算法的有效性的表格。

[0035] 图 9B 显示了按照本发明的实施方案在 384 孔数字 RNA-SNP 分析中比较用于分类整倍体和 21 三体性实例的新和旧 SPRT 算法的有效性的表格。

[0036] 图 10 的表格显示了按照本发明的实施方案, 对于给定的信息计数, 被正确或错误分类为整倍体或非整倍体以及那些不可分类的胎儿的百分比。

[0037] 图 11 是表格 1100, 显示了按照本发明的实施方案, 纯 (100%) 胎儿 DNA 样品的数字 RCD 分析的计算机模拟。

[0038] 图 12 是表格 1200, 显示了按照本发明的实施方案, $m_r = 0.5$ 的数字 RCD 分析的准确性的计算机模拟的结果, 该数字 RCD 分析用于对来自具有不同部分浓度的胎儿 DNA 的整倍体或 21 三体性胎儿的样品 进行分类。

[0039] 图 13A 显示了按照本发明的实施方案, 整倍体妊娠和 21 三体性妊娠的胎盘组织的数字 RNA-SNP 分析的表格 1300。

[0040] 图 13B 显示了按照本发明的实施方案, 来自整倍体妊娠和 21 三体性妊娠的母体血浆的数字 RNA-SNP 分析的表格 1350。

[0041] 图 14A-14C 显示的图示例了按照本发明实施方案得自 RCD 分析的截止曲线。

[0042] 图 15A 显示了按照本发明的实施方案, 整倍体妊娠和 21 三体性妊娠的胎盘组织中的数字 RNA-SNP 分析的表格。

[0043] 图 15B 显示了按照本发明的实施方案, 来自一个母体血浆样品的 12 个反应板的数字 RNA-SNP 数据的表格。

[0044] 图 15C 显示了按照本发明的实施方案, 来自整倍体妊娠和 21 三体性妊娠的母体血浆的数字 RNA-SNP 分析的表格。

[0045] 图 16A 显示了按照本发明的实施方案, 整倍体胎盘和 18 三体性胎盘的数字 RNA-SNP 分析的表格。

[0046] 图 16B 显示了按照本发明的实施方案,整倍体胎盘和 18 三体性胎盘的数字 RNA-SNP 数据的 SPRT 解释。

[0047] 图 17 显示了按照本发明的实施方案,整倍体妊娠和 21 三体性妊娠的 50%胎盘 / 母体血液细胞 DNA 混合物的数字 RCD 分析的表格。

[0048] 图 18 显示的 SPRT 曲线示例了按照本发明的实施方案,用于正确分类的判定边界 (decision boundary)。

[0049] 图 19 显示了按照本发明的实施方案,来自整倍体妊娠和 21 三体性妊娠的羊水样品的数字 RCD 分析的表格。

[0050] 图 20 显示了按照本发明的实施方案,来自整倍体妊娠和 18 三体性妊娠的胎盘 DNA 样品的数字 RCD 分析的表格 (E = 整倍体 ; T18 = 18 三体性)。

[0051] 图 21 显示了按照本发明的实施方案,整倍体妊娠和 21 三体性妊娠的 50%胎盘 / 母体血液细胞 DNA 混合物的多重数字 RCD 分析的表格 (E = 整倍体 ; T21 = 21 三体性 ; U = 未分类的)。

[0052] 图 22A 和 22B 显示按照本发明的实施方案,50%整倍体或 21 三体性胎盘基因组 DNA/50%母体血沉棕黄色层 (buffy coat) DNA 混合物的多重数字 RCD 分析的表格。Unclass 表示不可分类的并且 T21 表示 21 三体性。

[0053] 图 23 显示了雄性和雌性配偶都携带相同突变的情况。

[0054] 图 24A 显示按照本发明的实施方案,雌性 / 雄性和雄性 / 雄性 DNA 混合物的数字 RMD 分析的表格。

[0055] 图 24B 显示了按照本发明的实施方案,25%雌性与 75%雄性 DNA 混合物的数字 RMD 分析的表格。

[0056] 图 25 显示了按照本发明的实施方案,模拟母体血浆样品 HbE 突变的 15% -50% DNA 混合物的数字 RMD 分析的表格。

[0057] 图 26A 显示了按照本发明的实施方案,模拟母体血浆样品 CD41/42 突变的 5% -50%的 DNA 混合物的数字 RMD 分析的表格。

[0058] 图 26B 显示了按照本发明的实施方案,模拟母体血浆样品 CD41/42 突变的 20%的 DNA 混合物的数字 RMD 分析的表格。

[0059] 图 27 显示了可用于本发明的实施方案的系统和方法的示例性计算机装置的方框图。

[0060] 定义

[0061] 本文所用的术语“生物样品”意指取自个体 (例如,诸如孕妇的人) 并含有一种或多种感兴趣的核酸分子的任何样品。

[0062] 术语“核酸”或“多核苷酸”意指脱氧核糖核酸 (DNA) 或核糖核酸 (RNA) 及其单链或双链形式的聚合物。除非特别地限定,该术语包括含有天然核苷酸的已知类似物的核酸,其具有与参考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进行代谢。除非另外指明,特定的核酸序列还隐含地包括其保守地修饰的变体 (例如,简并密码子取代)、等位基因、直向同源物、SNP 和互补序列以及明确地指出的序列。具体地,简并密码子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) 和 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 以及基因或基因座编码的短发夹 RNA(shRNA) 可交换使用。

[0063] 术语“基因”表示与产生多肽链有关的 DNA 的片段。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）以及单独的编码片段（外显子）之间的间插序列（内含子）。

[0064] 本文所用的术语“反应”意指与表示感兴趣的特定多核苷酸序列的存在或不存在的化学、酶或物理作用有关的任何过程。“反应”的实例是诸如聚合酶链式反应 (PCR) 的扩增反应。“反应”的另一实例是通过合成或通过连接的测序反应。“信息反应”是表明一种或多种感兴趣的特定多核苷酸序列的存在的反应,并且在一种情况下,只存在一种感兴趣的序列。本文所用的术语“孔”意指在有限的结构内的预定位置的反应,例如,PCR 阵列中的孔状小管、单元或室。

[0065] 本文所用的术语“临床相关的核酸序列”能够指对应于更大的基因组序列的片段的多核苷酸序列或者指更大的基因组序列自身,该多核苷酸序列的潜在失衡被检测。一个实例是染色体 21 的序列。其它实例包括染色体 18、13、X 和 Y。仍然其它的实例包括胎儿遗传自其双亲中一个或两个的突变的遗传序列或遗传多态性或拷贝数变异。仍然其它的实例包括在恶性肿瘤中突变、缺失或扩增的序列,例如,发生了杂合性丢失或基因重复的序列。在某些实施方案中,多个临床相关的核酸序列或该临床相关的核酸序列等同的多个标记物能够用于提供检测失衡的数据。例如,来自染色体 21 上的 5 个不连续序列的数据能够以累加的方式用于确定可能的染色体 21 失衡,从而将所需的样品体积有效地减少至 1/5。

[0066] 本文所用的术语“背景核酸序列”意指与所述临床相关的核酸序列的正常比率是已知的核酸序列,例如,1 比 1 的比率。作为一个实例,所述背景核酸序列和所述临床相关的核酸序列是来自相同的染色体并且由于杂合性而不同的两个等位基因。在另一实例中,所述背景核酸序列是与另一等位基因杂合的一个等位基因,所述另一等位基因是所述临床相关的核酸序列。而且,某些背景核酸序列和临床相关的核酸序列的每一个可以来自不同的个体。

[0067] 本文所用的术语“参考核酸序列”意指每个反应的平均浓度是已知的或者已经被等同地测量过的核酸序列。

[0068] 本文所用的术语“过度表现的 (overrepresented) 核酸序列”意指在生物样品中的两个感兴趣的序列（例如,临床相关的序列和背景序列）之中丰度比另一序列更高的核酸序列。

[0069] 本文所用的术语“基于”表示“至少部分地基于”,并且意指在确定另一值时所用的一个值（或结果）,例如,发生在方法的输入和该方法的输出的联系中。本文所用的术语“导出”也意指方法的输入和该方法的输出的联系,例如,当导出是公式的计算时发生。

[0070] 本文所用的术语“定量数据”表示从一个或多个反应获得并且提供一个或多个数值的数据。例如,显示特定序列的荧光标记物的孔的数目是定量数据。

[0071] 本文所用的术语“参数”表示表征定量数据组和 / 或定量数据组之间的数值联系的数值。例如,第一核酸序列的第一量与第二核酸序列的第二量之间的比率（或比率的函数）是参数。

[0072] 本文所用的术语“截止值”表示用于在生物样品的两个或更多个类别状态（例如，患病和未患病）之间进行裁定（arbitrate）的数值。例如，如果参数大于截止值，将定量数据分为第一类（例如，患病状态），或者如果该参数小于该截止值，则将定量数据分为另一类（例如，未患病状态）。

[0073] 本文所用的术语“失衡”表示由临床相关的核酸序列的量中至少一个截止值所定义的与参考量的任何显著偏差。例如，该参考量能够是 3/5 的比率，因此如果测量的比率是 1 : 1，则发生了失衡。

[0074] 发明详述

[0075] 本发明提供了方法、系统和装置，用于确定在生物样品中，与临床相关的核酸序列相对于其它非临床相关的序列的参考（例如，未患病）量比较，是否存在增加或减少（例如，染色体或等位基因失衡）。选择一个或多个截止值来确定与参考量相比是否存在变化（即，失衡），例如，关于两个序列（或两组序列）的量的比率。检测到的参考量变化可以是临床相关的核酸序列与其它非临床相关的序列的关系的任何偏差（上升或下降）。因此，参考状态可以是任何比率或其它量（例如，除了 1-1 的对应），并且表示变化的测量状态可以是任何比率或不同于由一个或多个截止值所确定的参考量的其它量。

[0076] 所述临床相关的核酸序列和所述背景核酸序列可以来自第一类型的细胞和来自一种或多种第二类型的细胞。例如，源自胎儿 / 胎盘细胞的胎儿核酸序列存在于诸如母体血浆的生物样品中，该生物样品包含源自母体细胞的母体核酸序列的背景。因此，在一实施方案中，至少部分地基于生物样品中所述第一类型的细胞的百分比来确定截止值。注意，可以通过任何源自胎儿的基因座来测定样品中胎儿序列的百分比，并且不限于测量所述临床相关的核酸序列。在另一实施方案中，至少部分地基于诸如血浆、血清、唾液或尿的生物样品中肿瘤序列的百分比来确定截止值，该生物样品包含源自体内的非恶性细胞的核酸序列的背景。

[0077] 仍然在另一实施方案中，基于多个反应中序列的平均浓度来确定截止值。在一方面，从估计含有特定核酸序列的信息孔的比例来确定所述截止值，其中该比例是基于上文所述的百分比和 / 或平均浓度来确定的。可以使用许多不同类型的方法来确定截止值，例如，SPRT、假发现、置信区间、接收器工作特性 (ROC)。这种策略还能够在做出确信的分类前将检测所要求的量降至最少。这与模板的量通常有限的血浆核酸分析是特别相关的。尽管通过数字 PCR 来表现这种策略，但是也可以使用其它方法。

[0078] 数字 PCR 包括极端稀释的核酸的多个 PCR 分析，从而大部分阳性扩增反映了来自单个模板分子的信号。由此数字 PCR 允许计数单独的模板分子。分析的 PCR 总数中的阳性扩增的比例允许估计原始或未稀释的样品中的模板浓度。这种技术被认为允许检测各种遗传现象 (Vogelstein, B et al. 1999, 见上文)，并且最近被用于检测肿瘤样品 (Zhou, W. et al. 2002, 见上文) 和癌症患者血浆 (Chang, HW et al. 2002, 见上文) 中的杂合性丢失。由于通过数字 PCR 的模板分子定量不依赖于报道染料与核酸浓度之间的剂量反应关系，所以理论上数字 PCR 分析的精度应当高于实时 PCR 的精度。因此，数字 PCR 潜在地能够允许鉴别靶基因座与参考基因座之间更精细程度的定量差异。

[0079] 为了对此进行检测，我们首先评价数字 PCR 是否能够测定母体血浆中来自染色体 21 的胎盘转录物，PLAC4mRNA 的等位基因比率 (Lo, YMD, et al. 2007 Nat Med 13, 218-223)，

从而区分 21 三体性胎儿和整倍体胎儿。这种方法被称为数字 RNA-SNP 方法。我们然后评价数字 PCR 增加的精度是否能够允许检测胎儿的染色体非整倍性而不依赖于遗传多态性。我们将这种方法称为数字相关的染色体剂量 (RCD) 分析。数字 RNA-SNP 方法依赖于多态性,但是在定量鉴别中要求较低的精度,而数字相关的染色体剂量 (RCD) 分析不依赖于多态性,但是对于定量鉴别要求较高的精度。

[0080] I. 数字 RNA-SNP

[0081] A. 概述

[0082] 数字 PCR 能够检测 DNA 样品中两个等位基因的等位基因比率偏移的存在。例如,数字 PCR 已经用于检测肿瘤 DNA 样品中的杂合性丢失 (LOH)。假定在 DNA 样品中有两个等位基因,即 A 和 G,并且 A 等位基因将在细胞中随着 LOH 而丢失。当在肿瘤样品的 50% 的细胞中存在 LOH 时,该 DNA 样品中 G : A 的等位基因比率将是 2 : 1。然而,如果在该肿瘤样品中不存在 LOH,则 G : A 的等位基因比率的比率将是 1 : 1。

[0083] 图 1 是示出数字 PCR 实验的流程图。在步骤 110 中,将 DNA 样品稀释,然后分配至单独的孔中。注意,发明人已经确定在原始样品中,某些血浆核酸种类已经被充分地稀释。因此,如果某些模板已经以需要的浓度存在,则不需将它们稀释。在以前的研究中 (例如, Zhou et al. 2002, 见上文),将 DNA 样品稀释至特定的“模板 DNA”的平均浓度约是每孔的两个模板中的一个模板 0.5 分子的程度。注意,术语“模板 DNA”看起来意指 A 等位基因或 G 等位基因,并且没有为这种具体的浓度提供原理阐述。

[0084] 在步骤 120 中,在每个孔中进行 PCR 过程来同时检测 A 等位基因和 / 或 G 等位基因。在步骤 130 中,在每个孔中鉴定了标记物 (例如,通过荧光),例如, A、G、A 和 G 或者 A 和 G 都不是。在没有 LOH 的情况下, DNA 样品中的 A 等位基因与 G 等位基因的丰度将是相同的 (每孔一个拷贝)。因此,孔对该 A 等位基因与对该 G 等位基因是阳性的概率是相同的。这通过对该 A 等位基因或对该 G 等位基因是阳性的孔的数目相似反映出。然而,当在肿瘤样品的 50% 或更多的细胞中存在 LOH 时, G 等位基因和 A 等位基因的等位基因比率将至少是 2 : 1。以前的方法简单地假定,样品是至少 50% 癌性的。因此,孔对 G 等位基因是阳性的概率将高于对 A 等位基因是阳性的概率。因此,对 G 等位基因是阳性的孔的数目将大于对 A 等位基因是阳性的孔的数目。

[0085] 在步骤 140 中,为了分类数字 PCR 的结果,计数对每个等位基因是阳性的,但是对另一等位基因不是阳性的孔。在上文的实例中,计数了对 A 等位基因是阳性,但对 G 等位基因是阴性的孔的数目和对 G 等位基因是阳性,但对 A 等位基因是阴性的孔的数目。在一实施方案中,表现出较少的阳性孔的等位基因被视为参考等位基因。

[0086] 在步骤 150 中,信息孔的总数被确定为对所述两个等位基因的任一个是阳性的孔的数目的总和。在步骤 160 中,计算了由具有较多的阳性孔的等位基因贡献的信息孔的比例 (P_r) (参数的实例)。 $P_r = \text{只对具有较多阳性孔的等位基因是阳性的孔的数目} / \text{只对一个等位基因 (A 或 G) 是阳性的孔的总数}$ 。其它实施方案能够使用具有一个等位基因的全部孔除以具有至少一个等位基因的全部孔。

[0087] 在步骤 170 中,确定 P_r 的值是否表示等位基因失衡。由于期望准确度和效能,所以这一任务并非简单的。确定失衡的一种方法使用了 Bayesian 类似然方法,序贯概率比检验 (SPRT)。SPRT 是允许随着数据的积累比较两种概率假设的方法。换言之,SPRT 是将数

字 PCR 结果分类为表示等位基因偏移存在或不存在的统计学方法。该方法具有将获得特定统计功效和准确度所需要分析的孔的数目减至最小的优势。

[0088] 在示例性的 SPRT 分析中,将针对无效假设和备选假设来检验实验结果。当在样品中有等位基因比率偏移时,则接受备选假设。当在样品中没有等位基因比率偏移时,则接受无效假设。将该 P_r 值与两个截止值比较以接受无效假设或备选假设。如果没有接受任何一个假设,则将该样品标记为未分类的,这表示观察到的数字 PCR 结果不足以以期望的统计学可信度将该样品进行分类。

[0089] 通常基于在假设中给出的假定下的 P_r 固定值来计算接受无效假设或备选假设的截止值。在所述无效假设中,假定样品没有表现出等位基因比率偏移。因此,对 A 等位基因和 G 等位基因是阳性的每个孔的概率将是相同的,因此, P_r 的预期值将是 $1/2$ 。在所述备选假设中, P_r 的预期值是 $2/3$,或者大约是 0.5 与 $2/3$ 的中间值,例如 0.585 。并且,由于有限的实验数目,能够选择上限 $(.585+3/N)$ 和表示为 $(.585-3/N)$ 的下限。

[0090] B. 唐氏综合征的检测

[0091] 在本发明的一实施方案中,数字 SNP 用于从孕妇血浆中检测胎儿唐氏综合征。使用对胎儿/胎盘细胞特异性的标记物可以测量染色体 21 中的等位基因比率。例如,为了确定观察到的 PLAC4 等位基因的过度表现的程度是否是统计学显著的,使用 SPRT。

[0092] 根据一示例性的实施方案,数字 RNA-SNP 确定了位于 PLAC4mRNA 的 A/G SNP, rs8130833 的多态性等位基因比率的失衡,该 mRNA 是从染色体 21 转录并被胎盘表达的。对于杂合的整倍体胎儿, A 等位基因和 G 等位基因应当在胎儿基因组中被相等地表现 ($1:1$ 基因组比率);而在 21 三体性中,三体的染色体 21 将与胎儿基因组中的一个 SNP 等位基因的额外拷贝有关,从而获得 $2:1$ 的比率。数字 PCR 的目的是确定分析的样品中的两个 PLAC4 等位基因的量是否相等。因此, A PLAC4 等位基因和 G PLAC4 等位基因都是靶模板。设计了实时 PCR 测定来扩增 PLAC4 mRNA,并且通过 TaqMan 荧光探针来鉴别这两个 SNP 等位基因。分析步骤的示意图示于图 2A 中。

[0093] 图 2A 示出本发明实施方案的数字 RNA-SNP 方法 200。在步骤 210 中,接收样品。在步骤 220 中,在提取的 RNA 样品中将诸如 PLAC4 mRNA 的核酸序列定量。在一实施方案中,通过 PLAC4 mRNA 的实时 PCR 来进行这种定量。在一方面,这个步骤为操作者提供在靶标达到数字 PCR 分析的“范围”前所需的稀释程度的概念。

[0094] 在步骤 230 中,将样品稀释。在步骤 240 中,测量稀释的样品的浓度。稀释的样品浓度可以被证实为约 1 个模板/孔(即,参考序列或非参考序列或任何一个等位基因)。某些实施方案使用第 IV 部分所述的技术来进行这一测量。例如,我们将稀释的样品分配至实时 PCR 分析的 96 个孔中来保证实现了可用的稀释。如在后文中将解释的,稀释浓度也可以是未知的,从而省略这一步骤。

[0095] 在步骤 250 中,在阵列的每个孔中进行数字 PCR。例如,将相同的稀释的样品分配至实时 PCR 分析的 384 个孔中。从 PCR 结果中鉴定了每个核酸序列的标记物的量和信息孔的数目。信息孔被定义为仅对 A 等位基因或 G 等位基因是阳性,而不是对两个等位基因都是阳性的孔。在步骤 260 中,计算 P_r 的预期值。在后文中将更详细地讨论这些步骤。所述计算包括从步骤 250 所测定的值来确定参数。例如,可以计算每孔的实际平均模板浓度。

[0096] 在步骤 270 中,可以进行 SPRT 或其它似然比率检验来确定是否存在失衡。对于整

倍体情况,我们预期相等数目的 A 阳性孔和 G 阳性孔。然而,当分析来自 21 三体性胎儿的模板分子时,只含有一个等位基因的孔的数目将大于只含有另一等位基因的孔的数目。简而言之,等位基因失衡对 21 三体性是预期的。

[0097] 如上文所述的,SPRT 是 Bayesian 类似然方法 (Bayesian-type likelihood method),该方法允许随数据的积累比较两个概率假设。在 21 三体性检测的数字 PCR 分析中,当存在等位基因失衡时(即,检测到 21 三体性),则接受备选假设;当没有等位基因失衡时(即,没有检测到 21 三体性),则接受无效假设。更多数目计数的等位基因被称为潜在地过度表现的等位基因,并且将计算该等位基因在全部信息孔中的比例 (P_r)。如果该 P_r 表明了足够程度的对 21 三体性样品预期的等位基因失衡,则应用 SPRT 来进行确定。

[0098] 可操作地,能够通过使用具有一对 SPRT 曲线的图来应用和解释 SPRT,构建该 SPRT 曲线来定义接受或拒绝任何一个假设的概率边界。图 3 示出按照本发明的实施方案用于确定唐氏综合征的 SPRT 曲线的图。当能做出确信的分类时,SPRT 曲线将对潜在过度表现的等位基因是阳性的信息孔的所需比例 P_r (y-轴)对信息孔的给定的总数 (x-轴)作图。如图 3 所示,上部曲线设定接受备选假设的概率边界,而下部曲线设定接受无效假设的概率边界。

[0099] 将实验推导出的 P_r 值与预期 P_r 值相比较以便接受或拒绝任一假设。如果接受无效假设,则将该样品分类为从怀有整倍体胎儿的孕妇获得的样品。如果接受备选假设,则将该样品分类为从怀有 21 三体性胎儿的孕妇获得的样品。可选择地,如果给定数目的信息计数的 P_r 没有达到疾病分类所要求的统计学可信度,则不能接受任何一个假设。在有更多的可用数据以前,这些情况被视为不可分类的。如果疾病分类是不可能的,则可以进行额外的 384 孔板直到累积的数据可以通过 SPRT 来分类。

[0100] 因此,对于给定水平的可信度,SPRT 比其它统计学方法提供了更少的所需检测量的优势。在实践中,只要积累了所需量的数据,SPRT 就允许接受或拒绝任何一个假设,从而将不需要的额外分析降至最低。这种特性与通常以低浓度存在的血浆核酸的分析特别相关,其中可用的模板的数目是有限的。除了严格的分类以外,所述分类还可以包括百分比准确度。例如,来自与截止值比较的分类可以提供表现出具有某一百分比的核酸序列失衡的可能性的样品,或者,等效地提供准确至某一百分比或其它值的确定失衡。

[0101] 利用母体血浆或血清中的胎儿核酸,可以应用类似的方法来确定关于突变或遗传多态性的胎儿基因型。应当记得的是,胎儿将从其母亲遗传胎儿一半的基因组。作为示例,考虑具有两个等位基因 A 和 B 的特定遗传基因座。如果母亲是基因型为 AB 的杂合子,则胎儿理论上能够具有 AA、BB 或 AB 的基因型。如果胎儿的基因型为 AB,即,与母亲相同,则母体血浆中将只有 AB 基因型的核酸(既来自母亲又来自胎儿)。因此,在母体血浆中观察到了核酸或等位基因的平衡。在另一方面,如果胎儿的基因型为 AA 或 BB,则在母体血浆中将分别有过度表现的 A 等位基因或 B 等位基因的等位基因失衡。这种考虑还适用于导致疾病的突变(例如,导致囊性纤维化、 β -地中海贫血症或脊髓型肌萎缩的那些突变),在这种情况下,A 能够被考虑为野生型等位基因,而 B 能够被考虑为突变体等位基因。

[0102] II. 数字 RCD

[0103] 数字 RNA-SNP 的劣势是,其只能应用于被分析的 SNP 是杂合的个例。一个改进是,基于循环的胎儿核酸分析的检测胎儿 21 三体性或其它胎儿染色体非整倍性(例如,18 三体

性、13 三体性和性染色体非整倍性) 的无创检测与遗传多态性的使用无关将是理想的。因此, 在一实施方案中, 通过相对于位于参考染色体, 即本研究中的染色体 1 上的基因座的非多态性的染色体 21 基因座的数字 PCR 分析来测定染色体剂量。从 21 三体性个例中区分整倍体胎儿基因组中染色体 21 比染色体 1 的比率偏离 2 : 2 的变化。在 21 三体性检测的数字 PCR 分析中, 要比较的两个假设将是没有染色体失衡 (即, 没有检测到 21 三体性) 的无效假设和存在染色体失衡 (即, 检测到了 21 三体性) 的备选假设。

[0104] 这种方法能够被推广至与其它染色体非整倍性有关的其它染色体, 例如, 18 三体性中的染色体 18、13 三体性中的染色体 13、特纳综合征中的染色体 X。另外, 除了染色体 1, 与非整倍性无关的其它染色体也能够用作参考染色体。通过分析在癌症中通常部分地缺失的染色体比参考染色体的比率的变化, 能够将类似的方法应用于检测癌症。通常部分地缺失的染色体的实例包括直结肠癌中的染色体 5q、肺癌中的染色体 3p 和鼻咽癌中的染色体 9p。图 2B 列出了某些导致序列失衡的某些常见的与癌症有关的染色体畸变。

[0105] 图 2A 还示出本发明实施方案的数字 RCD 方法 205。在步骤 220-230 的一实施方案中, 例如, 通过 Nanodrop 技术, 将提取的 DNA 定量, 并稀释至每孔大约一个靶模板的浓度, 所述靶模板来自染色体 21 或标准化的染色体 (例如, 染色体 1) 的。在步骤 240 的一实施方案中, 在 384 孔板中使用两个 TaqMan 探针进行数字 RCD 分析前, 可以进行如下证实: 通过分析稀释的 DNA 样品来证实约 37% 的水平 的孔 是否是阴性的, 该分析只通过使用 96 孔格式的染色体 1 探针的测定来进行。37% 的显著性将在后面的第 IV 部分中进行讨论。

[0106] 步骤 240 的检测和步骤 250 的结果可以用设计成扩增存在于两条染色体上的种内同源序列 (paralogous sequence) (Deutsch, S. et al. 2004 J Med Genet 41, 908-915) 的实时 PCR 测定来完成, 所述染色体被通过一对 TaqMan 探针鉴别的平行同源序列变化所区分。在本文中, 信息孔被定义为对任一染色体 21 或染色体 1 基因座是阳性的, 而对这两条染色体不都是阳性的孔。对于整倍体胎儿, 对任一基因座是阳性的信息孔的数目应当大致相等。对于 21 三体性胎儿, 应当有与染色体 1 阳性孔相比, 染色体 21 阳性孔的过度表现。在下文的部分中描述了过度表现的确切比例。

[0107] III. 并入胎儿序列的百分比

[0108] 上文所述的方法 200 和 205 的实施方案的劣势在于胎儿特异性的标记物是必需的。因此, 在本发明的一实施方案中使用了非胎儿特异性的标记物。为了使用这种非胎儿特异性的标记物, 本发明的实施方案测量了母体血浆 (即, 生物样品) 中胎儿 DNA 的部分浓度 (fractional concentration)。通过这些信息, 可以按照如下步骤来计算更有用的 P_r 值。

[0109] 即便对于母体血浆中胎儿 DNA 的小的部分百分比, 21 三体性胎儿将通过释放至母体血浆中的胎儿 DNA 的基因组当量 (genome-equivalent) (GE) 贡献额外剂量的染色体 21 序列。例如, 含有 50GE/ml 总 DNA 和 5GE/ml 胎儿贡献的 DNA (即, 10% 胎儿 DNA 部分浓度) 的来自整倍体妊娠的母体血浆样品将会含有每毫升母体血浆总共 100 个拷贝 (90 个母体拷贝 + 10 个胎儿拷贝) 的染色体 21 序列。对于 21 三体性妊娠, 每个胎儿 GE 将贡献 3 个拷贝的染色体 21, 这导致母体血浆中总共 105 个拷贝 /ml (90 个母体拷贝 + 15 个胎儿拷贝) 的染色体 21 序列。因此, 在 10% 的胎儿 DNA 浓度时, 三体妊娠母体血浆中源自染色体 21 的序列的量将是整倍体情况的 1.05 倍。因此, 如果能够开发测定这种小程度的定量差异的分析方法, 将实现不依赖于多态性的胎儿 21 三体性的无创产前诊断检测。

[0110] 因此,过度表现的程度将取决于分析的 DNA 样品中部分胎儿 DNA 浓度。例如,当分析胎盘 DNA 时,胎儿基因组中的理论 RCD 比率应当是 3 : 2,即,1.5 倍的差异。然而,如上文所述的,当分析含有 10% 的胎儿母体血浆时,该理论 RCD 比率将降至 1.05。通过将只对染色体 21 基因座是阳性的孔的数目除以信息孔的总数来计算实验导出的 P_r 。用计算的 P_r 和理论 RCD 比率来对实验导出的 P_r 进行 SPRT 分析。

[0111] 图 4 表示按照本发明的实施方案,利用胎儿核酸百分比来确定疾病状态的方法 400。在步骤 410 中,测量了胎儿物质的部分百分比。在一实施方案中,通过测量相对于非胎儿特异性标记物(即,在母亲和胎儿中都存在的基因序列)的胎儿特异性标记物(例如,Y 染色体,遗传多态性标记物(例如,SNP)、胎盘外遗传特征(epigenetic signature))的量来确定所述部分百分比。通过实时 PCR、数字 PCR、测序反应(包括大规模平行基因组测序)或任何其它定量方法来进行实际的测量。在一方面,优选地不使用对于本测量能够潜在地处于等位基因失衡的基因靶标。

[0112] 在步骤 420 中,进行了数字 PCR 或其它测量方法,包括将样品稀释,将该稀释的样品置于孔中并测量每孔中的反应。在步骤 430 中,将 PCR 结果用于鉴定不同参考核酸序列(例如染色体或等位基因)的标记物。在步骤 440 中,计算了过度表现的序列的实际比率(P_r)。在步骤 450 中,利用样品中胎儿物质的百分比来计算用于确定疾病状态的截止值。在步骤 460 中,从该实际 P_r 和该截止值来确定是否存在失衡。

[0113] 在一实施方案中,将参考核酸序列的部分百分比并入数字 RNA-SNP 方法中。因此,当研究由于癌细胞的 LOH 时,能够用少于 50% 癌细胞的肿瘤样品来进行这一步骤。还可以将这一步骤用于多于 50% 的癌细胞的样品以获得更准确的 P_r ,并因此减少将导致错误诊断的假阳性的数目。在另一实施方案中,将胎儿核酸百分比并入数字 PCR 方法中以确定胎儿是否已遗传了父母的基因突变(例如,导致囊性纤维化或 β -地中海贫血症或脊髓型肌萎缩的突变)或确定来自母体血浆核酸分析的多态性。

[0114] IV. 并入每孔的平均浓度

[0115] 以前的方法(例如,Zhou, W. et al. 2002,见上文)的另一个劣势是要求每孔的平均模板浓度(m)是每孔 1 个。考虑到难以获得确切的浓度,这能够导致误差。而且,甚至对于每孔 1 个模板的确切浓度,以前的方法忽略了孔中的模板的统计学分布。在以前的方法,即,老的算法中,假定接受备选假设的 P_r 的预期值是等位基因比率,因此,该 P_r 的预期值与每孔中的模板 DNA 的平均浓度无关。

[0116] 然而,由于稀释样品中模板的天然统计变异(statistical variation),将不会有确切的每孔 1 个模板。本发明的实施方案测量至少一种序列的平均浓度,然后将该平均浓度用于计算截止值,即预期的 P_r 。在一方面,这种计算包括了统计学分布以确定含有不同核酸序列的孔的概率,然后将该概率用于确定预期的 P_r 。

[0117] 在一实施方案中,获取了一种参考核酸序列的平均浓度,其在一实例中是 DNA 样品中较低浓度的核酸序列。在样品不具有失衡的情况下,样品中两种序列的浓度将是相同的,并且任何一种都能够被视为参考等位基因。在样品具有,例如,LOH 的情况下,在癌细胞中缺失的等位基因将被视为参考等位基因。将该参考等位基因的平均浓度表示为 m_r 。在另一实施方案中,浓度较高的序列可以被视作参考序列。

[0118] A. 数字 SNP :使用 SPRT 和数字 PCR 的实例

[0119] 图 5 显示了按照本发明的实施方案,使用平均模板浓度来确定疾病状态的方法 500。在步骤 510 中,测量了不同序列的量。例如,可以通过计数如上文所解释的数字 PCR 实验中的标记物来进行这一步骤。然而,可以通过其它方法来进行这一步骤,该方法不包括扩增步骤或者不使用荧光标记物,但是能够使用其它属性,例如如同质量的物理属性、比旋光属性或碱基配对属性。

[0120] 在步骤 520 中,测定了过度表现的序列的实际比例。如上文所述的,可以通过获取只表现出过度表现的序列的孔的数目,然后将该数目除以信息孔的数目来完成这个步骤。在步骤 530 中,测量了至少一种序列(参考序列)的平均浓度。在一实施方案中,所述参考序列是过度表现的序列。在另一实施方案中,所述参考序列是过少表现(underrepresented)的序列。可以通过计数在数字 PCR 实验中对参考序列是阴性的孔的数目来进行测量。如在下个分段中所述的,通过泊松分布(Poisson distribution)来描述阴性孔的比例与平均目标浓度之间的关系。

[0121] 在步骤 540 中,例如,使用泊松分布来计算对不同的序列是阳性的孔的预期量。该预期量可以是每孔的序列的概率、每孔的平均序列、含有序列的孔的数目或其它合适的量。在步骤 550 中,从该预期的量计算预期的 P_r 。在步骤 560 中,例如,通过使用 SPRT,从预期的 P_r 计算截止值。在步骤 570 中,确定了核酸序列失衡的分类。现在将描述方法 500 的具体方面。

[0122] 1. 确定序列的预期量

[0123] 一旦从步骤 530 知道了每孔的平均浓度(反应或反应混合物),就可以在步骤 540 中计算表现出该序列的孔的预期数目。这种量可以表示为%、分数值或整数值。利用具体的实例进行说明,假定每孔的参考模板的平均浓度(m_r)是每孔 0.5 个,并且 21 三体性胎儿在 PLAC4SNP, rs8130833 的基因型是 AGG。因此,参考模板是 A 等位基因,并且过度表现的模板是 G 等位基因。

[0124] 在一实施方案中,假定 A 等位基因在诸如数字 PCR 的测量方法的孔的反应混合物中的分布是泊松分布。在其它实施方案中,使用了其它分布函数,例如二项分布。

[0125] 泊松方程式是: $P(n) = \frac{m^n e^{-m}}{n!}$, 其中, n = 每孔的模板分子的数目; $P(n)$ = n 个模板分子在特定的孔中的概率; 并且 m = 特定的数字 PCR 实验中一个孔中的模板分子的平均数目。

[0126] 因此,在 0.5 的平均 A 等位基因的浓度下,不含 A 等位基因的任何分子的任何孔的概率是:

$$[0127] \quad P(0) = \frac{0.5^0 e^{-0.5}}{0!} = e^{-0.5} = 0.6065$$

[0128] 因此,含有 A 等位基因的至少一个分子的任何孔的概率是: $1 - 0.6065 = 0.3935$ 。因此,预期约 39% 的孔将含有 A 等位基因的至少一个分子。

[0129] 关于非参考核酸序列,对于 21 三体性胎儿的每个细胞, A 比 G 的基因组比率将是 1 : 2。假定在提取的 RNA 或 DNA 样品中的 A 比 G 的比率保持不变,则每孔的 G 等位基因的平均浓度将是 A 等位基因的平均浓度的 2 倍,即, $2 \times 0.5 = 1$ 。

[0130] 因此,在平均的 G 等位基因浓度为 1 的情况下,不含 G 等位基因的任何分子的任何孔的概率是:

$$[0131] \quad P(0) = \frac{1^0 e^{-1}}{0!} = e^{-1} = 0.3679$$

[0132] 因此,含有 G 等位基因的至少一个分子的任何孔的概率是 : $1-0.3679 = 0.6321$ 。因此预期约 63% 的孔会含有 G 等位基因的至少一个分子。

[0133] 2. 确定过度表现的序列的比例

[0134] 计算了预期量后,可以确定过度表现的核酸序列的比例。假定用 A 等位基因和 G 等位基因填充孔是独立的,则含有两个等位基因的孔的概率是 $0.3935 \times 0.6321 = 0.2487$ 。因此,预期约 25% 的孔将含有两个等位基因。

[0135] 预期含有 A 等位基因,但是不含有 G 等位基因的孔的比例将是含有至少一个 A 等位基因的孔的数目减去既含有 A 等位基因又含有 G 等位基因的孔的数目 : $0.3935-0.2487 = 0.1448$ 。类似地,预期含有 G 等位基因,但是不含有 A 等位基因的孔的比例将是 $0.6321-0.2487 = 0.3834$ 。信息孔被定义为对 A 等位基因或 G 等位基因是阳性,但是不对两个等位基因都是阳性的孔。

[0136] 因此,在数字 RNA-SNP 分析中,含有 A 等位基因的孔相对于 G 等位基因的孔的预期比率是 $0.1448/0.3834$ 。换言之,只对 G 等位基因是阳性的孔的比例是只对 A 等位基因是阳性的孔的比例的 2.65 倍。这与胎儿基因组比率形成对比,其中过度表现的等位基因是另一等位基因的 2 倍。

[0137] 对于 SPRT 分析,计算了对过度表现的等位基因是阳性的信息孔的比例 (P_r),并利用 SPRT 曲线对该比例进行解释。在本实例中,信息孔的比例是 : $0.1448+0.3834 = 0.5282$ 。因此,在 m_r 为 0.5 时,21 三体性病例的预期 P_r 是 $0.3834/0.5282 = 0.73$ 。

[0138] 由于平均模板浓度 (m) 是泊松方程式中的关键参数,所以 P_r 将随 m 而变化。图 6 显示了本发明实施方案的表 600,该表列出了对于表示为每孔的平均参考模板浓度 (m_r) 的一系列模板浓度,21 三体性样品的预期的数字 RNA-SNP 等位基因比率和 P_r 。表 600 显示了一系列每孔的平均参考模板浓度 (m_r),预期的等位基因比率和对过度表现的等位基因是阳性的信息孔的比例 (P_r)。

[0139] P_r 的预期值以非线性的方式随每孔的参考等位基因的平均浓度 (m_r) 而变化。如表 600 所示,接受备选假设的 P_r 的预期值随 m_r 而增加。由于接受无效假设的 P_r 的预期值固定为 0.5,所以当 m_r 增加时,具有或不具有等位基因失衡的样品就 P_r 值而言将分得更开。注意,在其它实施方案中,接受无效假设的值可以不同于 0.5。当正常比率不同于 1 : 1,例如 5 : 3 时,可能发生这种情况,因此,当比率偏离 5 : 3 时,将发生失衡。将基于具体情况来确定两种不同核酸序列的量的差异。

[0140] 然而,由于以前的方法(例如,Zhou,W. et al. 2002,见上文)使用了 LOH 样品的固定的 P_r 预期值,所以这些方法低估了具有 LOH 的那些样品的 P_r 值(接受备选假设)。低估的程度将随 m_r 而增加。换言之,DNA 样品中参考等位基因的平均浓度越高,旧方法就越不准确。这种接受备选假设的 P_r 的低估将导致既接受无效假设又接受备选假设的截止值的不准确的计算。

[0141] 3. 基于预期 P_r 计算截止值

[0142] 对于使用 SPRT 的实施方案,可以使用来自 El Karoui 等人(2006)的计算 SPRT 曲线的上限和下限的方程式。而且,优选接受无效假设或备选假设的统计学可信度能够通过

调整方程式中的阈值似然比率而变化。在这个研究中,使用的阈值似然比率是 8,因为这个值已经被证实在癌症检测的环境下提供了鉴别包含或不包含等位基因失衡的样品的令人满意的性能。因此,在一实施方案中,计算 SPRT 曲线的上限和下限的方程式是:

[0143] 上限 = $[(\ln 8) / (N - \ln \delta)] / \ln \gamma$

[0144] 下限 = $(\ln 1/8) / (N - \ln \delta)] / \ln \gamma$

[0145] 其中,

[0146] $\delta = (1 - \theta_1) / (1 - \theta_0)$

[0147] $\gamma = -(\theta_1(1 - \theta_0) / \theta_0(1 - \theta_1))$

[0148] θ_0 = 如果无效假设是真实的,含有非参考等位基因的信息孔的比例

[0149] = 0.5 (参见下文)

[0150] θ_1 = 如果备选假设是真实的,含有非参考(即,过度表现的)等位基因的信息孔的比例

[0151] N = 信息孔的数目

[0152] = 只对任何一个等位基因是阳性的孔的数目

[0153] ($\ln$ 是表示自然对数的数学符号,即 $\log_e$)

[0154] 对于接受无效假设的 θ_0 的测定,假定从怀有整倍体胎儿的孕妇获得样品。在这一假定下,对任一个模板是阳性的孔的预期数目将是 1 : 1,因此含有非参考等位基因的信息孔的预期比例将是 0.5。

[0155] 对于接受备选假设的 θ_1 的测定,假定从怀有 21 三体性胎儿的孕妇获得样品。数字 RNA-SNP 分析的 21 三体性情况的预期 P_r 的计算细节列于表 600 中。因此,数字 RNA-SNP 分析的 θ_1 意指表 600 最后一列所示的数据。

[0156] 4. 平均浓度的测量

[0157] 可以通过对本领域技术人员公知或将公知的多种方法来测量 m_r 。在一实施方案中,在数字 PCR 分析的实验过程期间确定 m_r 的值。由于 m_r 值与对参考等位基因是阳性的孔的总数的关系能够被分布所控制(例如,泊松分布),所以能够利用下述公式,从对参考等位基因是阳性的孔的数目来计算 m_r :

[0158] $m_r = -\ln(1 - \text{对参考等位基因是阳性的孔的比例})$ 。

[0159] 注意, $\ln$ 是自然对数,即 $\log_e$ 。这种方法提供了用于数字 PCR 实验的 DNA 样品中 m_r 直接和精确的估计。

[0160] 这种方法可以用于获得期望的浓度。例如,如方法 200 的步骤 240 中所做的,可以将样品中提取的核酸稀释至具体的浓度,例如,每一反应孔一个模板分子。在使用泊松分布的实施方案中,不含模板的孔的预期比例可以计算为 e^{-m} ,其中 m 是每孔的模板分子的平均浓度。例如,在每孔一个模板分子的平均浓度下,不含模板分子的孔的预期比例是 e^{-1} ,即 0.37 (37%)。剩下的 63% 的孔将含有一个或多个模板分子。通常,然后将计数数字 PCR 运行(run)中阳性孔和信息孔的数目。信息孔的定义和解释数字 PCR 数据的方式取决于应用。

[0161] 在其它实施方案中,每孔的平均浓度 m_r 通过其它定量方法来测量,例如,定量实时 PCR、半定量竞争 PCR、利用质谱方法的实时竞争 PCR(real-competitive PCR) 等。

[0162] B. 数字 RCD

[0163] 可以以与上文所述的数字 SNP 方法类似的方式进行使用平均浓度的数字 RCD。能够通过数字 PCR 来确定对参考染色体（非染色体 21）标记物是阳性、对染色体 21 标记物是阳性以及对两种标记物都是阳性的孔的数目。根据数字 SNP 分析的 m_t 计算中的泊松概率函数，能够从对参考标记物是阴性的孔的总数来计算每孔的参考标记物的平均浓度（ m_t ），而不管染色体 21 标记物的阳性。

[0164] 然后可以将 SPRT 分析用于将血浆样品分类为获自怀有整倍体胎儿的孕妇或者获自怀有 21 三体性胎儿的孕妇。当胎儿是整倍体时，接受无效假设。在这种情况下，对参考标记物和染色体 21 标记物是阳性的孔的预期比率将是 1 : 1，因此，具有对染色体 21 标记物阳性信号的信息孔的预期比率将是 0.5。当胎儿是染色体 21 三体时，将接受备选假设。在这种情况下，如果样品 DNA 只来自胎儿，则每个孔中的染色体 21 标记物的平均浓度将是参考标记物的平均浓度（ m_t ）的 3/2 倍。

[0165] 当将数字 RCD 用于通过检测诸如胎盘的外遗传特征的胎儿特异性标记物来测定染色体剂量时（Chim, SSC. et al. 2005 Proc Natl AcadSci USA 102,14753-14758），数字 RCD 分析的实施方案使用非胎儿特异性的标记物。因此，当使用非胎儿特异性标记物时，将进行测量胎儿物质百分比的额外步骤。因此，每孔的染色体 21 标记物的平均浓度将取决于样品中胎儿 DNA 的比例，并且能够使用下式进行计算： $m_t[(200\% + \text{胎儿 DNA 百分比})/200\%]$ 。

[0166] 再次利用具体实例进行说明，假定每孔的参考模板、染色体 1 的平均浓度（ m_t ）是 0.5，并且假定 50% 的 DNA 源自胎儿，样品中 50% 的 DNA 源自母亲。

[0167] 因此，利用泊松分布，当染色体 1 的平均浓度是每孔 0.5 时，不含染色体 1 基因座的任何分子的任何孔的概率将是：

$$[0168] \quad P(0) = \frac{0.5^0 e^{-0.5}}{0!} = e^{-0.5} = 0.6065$$

[0169] 因此，含有染色体 1 基因座的至少一个分子的任何孔的概率将是： $1 - 0.6065 = 0.3935$ 。因此，预期约 39% 的孔将含有该基因座的至少一个分子。

[0170] 对于这种三体性胎儿的每个细胞，染色体 21 比染色体 1 的基因组比率将是 3 : 2。DNA 样品中染色体 21 与染色体 1 之间的比率将取决于部分胎儿 DNA 浓度（胎儿 DNA %），并且是： $3 \times \text{胎儿 DNA \%} + 2(1 - \text{胎儿 DNA \%}) : 2 \times \text{胎儿 DNA \%} + 2 \times (1 - \text{胎儿 DNA \%})$ 。因此，在这种情况下，当部分胎儿 DNA 浓度是 50% 时，该比率是： $(3 \times 50\% + 2 \times 50\%) / (2 \times 50\% + 2 \times 50\%) = 1.25$ 。如果数字 SNP 方法没有使用胎儿特异性标记物，则这种计算还能够用于计算非参考序列的平均浓度。

[0171] 因此，当每孔的染色体 1 基因座的平均浓度是 0.5 时，每孔的染色体 21 基因座的平均浓度是： $1.25 \times 0.5 = 0.625$ 。因此，不含染色体 21 基因座的任何分子的任何孔的概率在染色体 21 基因座平均浓度为每孔 0.625 时将是：

$$[0172] \quad P(0) = \frac{0.625^0 e^{-0.625}}{0!} = e^{-0.625} = 0.5353$$

[0173] 因此，含有染色体 21 基因座的至少一个分子的任何孔的概率将是： $1 - 0.5353 = 0.4647$ 。因此，预期约 46% 的孔将含有该基因座的至少一个分子。假定用任一个基因座填

充孔是独立的,则含有两个基因座的孔的概率将是: $0.3935 \times 0.4647 = 0.1829$ 。因此,预期约 18% 的孔将含有两个基因座。

[0174] 预期含有染色体 1 基因座,但是不含有染色体 21 基因座的孔的比例将是含有至少一个染色体 1 基因座的孔的数目减去含有两个基因座的孔的数目: $0.3935 - 0.1829 = 0.2106$ 。类似地,预期含有染色体 21 基因座而不是两个基因座都含有的孔的比例将是: $0.4647 - 0.1829 = 0.2818$ 。信息孔被定义为对染色体 1 基因座或染色体 21 基因座是阳性,但是不对两个基因座都是阳性的孔。

[0175] 因此,数字 RCD 分析中预期的染色体 21 比染色体 1 的比率是 $0.2818 / 0.2106 = 1.34$ 。换言之,只对染色体 21 基因座是阳性的孔的比例是只对染色体 1 基因座是阳性的孔的比例的 1.34 倍。这与 DNA 样品中 1.25 的比率形成对比。

[0176] 对于 SPRT 分析,需要计算对染色体 21 基因座是阳性的信息孔的比例 (P_r),并使用 SPRT 曲线来解释该比例。在本实例中,信息孔的比例将是: $0.2106 + 0.2818 = 0.4924$ 。因此,在 0.5 的 m_r 下,具有 50% 胎儿 DNA 的 21 三体性个例的预期 P_r 是 $0.2818 / 0.4924 = 0.57$ 。

[0177] 由于平均模板浓度 (m) 是泊松方程式中的关键参数,所以 P_r 将随 m 而变化。图 7 显示了按照本发明实施方案的表 700,其列出了在表示为每孔的平均参考模板浓度 (m_r) 的一系列模板浓度下,21 三体性样品中 10%、25%、50% 和 100% 的部分胎儿 DNA 浓度的预期 P_r 。对于数字 RCD 分析,21 三体性个例的预期 P_r 的运算详细列于表 700 中。因此,能够从表 700 中表示对应的预期 P_r 值的列获得胎儿 DNA 部分浓度变化的样品的数字 RCD 分析的 θ_1 。

[0178] C. 结果

[0179] 1. 不同 m_r 的比较

[0180] 等位基因或染色体失衡的理论程度(胎儿基因组中)和实验预期程度之间的差异的基础以及确定针对一系列 m_r 值的实验预期程度的计算结果示于表 600 和 700 中。在 21 三体性样品的数字 RNA-SNP 分析中,当 $m_r = 0.5$ 时,只含有过度表现的等位基因的孔比只含有参考等位基因的孔,即数字 RNA-SNP 比率是 2.65(表 600)。在由 100% 的胎儿 DNA 组成的样品的数字 RCD 分析中,当 $m_r = 0.5$ 时,只对染色体 21 基因座是阳性的孔比只对染色体 1 基因座是阳性的孔,即数字 RCD 比率,是 1.7(表 700) ($P_r = 0.63$, 因此,数字 RCD 比率是 $0.63 / (1 - 0.63) = 1.7$)。随着部分胎儿 DNA 浓度下降,数字 RCD 对相同的 m_r 下降。

[0181] 如表 600 和 700 所示,等位基因或染色体的过度表现的程度随着 m_r 而增加。然而,信息孔的百分比在 $m_r = 0.5$ 附近时达到其最大值,并且随着 m_r 的进一步增加而逐渐下降。在实践中,如果不限制样品的模板分子的量,则信息孔比例的下降能够通过增加分析的孔的总数来补偿,但是额外的孔需要增加试剂成本。因此,最优的数字 PCR 性能是模板浓度和每个样品测试的孔的总数之间的权衡。

[0182] 2. 使用 SPRT 曲线的实例

[0183] 如上文所讨论的,数字 PCR 实验的等位基因或染色体失衡的预期程度取决于每个反应混合物(例如,孔)的实际模板浓度。我们描述了基于参考等位基因的模板浓度,即每孔的平均参考模板浓度 (m_r)。如在上文的方程式中所示,预期的 P_r 能够用于确定上部 SPRT 曲线和下部 SPRT 曲线的作图。由于预期的 P_r 相应地依赖于 m_r 的值,所以 SPRT 曲线的作图基本上取决于 m_r 的值。因此,在实践中,需要使用与数字 PCR 数据集的实际 m_r 相关的一组

SPRT 曲线来解释来自该特定运行 (run) 的 P_r 。

[0184] 图 8 显示了图 800, 示例了按照本发明的实施方案, 对于用于数字 RNA-SNP 分析的 0.1、0.5 和 1.0 的 m_r 值的 SPRT 曲线中的差异程度。每组数字 PCR 结果应当用与该特定运行的确切 m_r 值相关的具体曲线来解释。注意, 由于对于数字 RNA-SNP 和数字 RCD 方法而言, 等位基因或染色体失衡的预期程度是不同的 (对于数字 RNA-SNP 为 2 : 1, 对于数字 RCD 为 3 : 2), 所以两个数字 PCR 系统需要不同组的 SPRT 曲线。用通过数字 PCR 运行的对应 m_r 选择的相关 SPRT 曲线来解释实验导出的 P_r 。这与以前报道的将 SPRT 用于通过数字 PCR 的 LOH 的分子检测形成对比, 该以前报道中使用了一组固定的曲线。

[0185] 利用假设的数字 RNA-SNP 运行, 下文示例了使用 SPRT 解释数字 PCR 数据的实施方式。每个实例的数字 RNA-SNP 分析后, 计数了只对 A 等位基因是阳性的孔的数目、只对 G 等位基因是阳性的孔的数目或者对这两个等位基因都是阳性的孔的数目。参考等位基因被定义为具有较少数目的阳性孔的等位基因。根据泊松概率密度函数, 使用对参考等位基因是阴性的孔的总数来计算 m_r 值, 而不管其它等位基因是否是阳性。我们假设的实例的数据如下所述:

[0186] 在 96 孔反应中, 20 个孔只对 A 等位基因是阳性的, 24 个孔只对 G 等位基因是阳性的, 并且 33 个孔对这两个等位基因都是阳性的。将 A 等位基因视为参考等位基因, 因为 A 阳性的孔少于 G 阳性的孔。对参考等位基因是阴性的孔的数目是 $96 - 20 - 33 = 43$ 。因此, 能够使用泊松方程式来计算 m_r : $-\ln(43/96) = 0.80$ 。这种情况的实验确定的 P_r 是 $24/(20+24) = 0.55$ 。

[0187] 根据表 600, $m_r = 0.8$ 时的 21 三体性样品的预期 P_r 是 0.76。因此, 这种情况的 θ_1 是 0.76。将基于 $\theta_1 = 0.76$ 的 SPRT 曲线用于解释这种情况的实验导出的 P_r , 0.55。当将 $P_r = 0.55$ 拟合至相关的 SPRT 曲线时, 数据点落在下部曲线下。因此, 将这种情况分类为整倍体, 参见图 3。

[0188] 3. 与旧方法的比较

[0189] 图 9A 显示了表 900, 其比较了用于分类 96 孔数字 RNA-SNP 分析中整倍体和 21 三体性实例的新和旧 SPRT 算法的有效性。图 9B 显示了表 950, 其比较了用于分类 384 孔数字 RNA-SNP 分析中整倍体和 21 三体性实例的新和旧 SPRT 算法的有效性。新的算法指选择对源自数字 PCR 数据的 m_r 特异性的 SPRT 曲线。旧的算法指对全部数字 PCR 运行使用固定组的 SPRT 曲线。通过如表 900 所示的模拟分析揭示了截止值的错误计算对分类准确度的影响。

[0190] 如表 900 和 950 所示, 与以前的研究中使用固定组的 SPRT 曲线相比, 我们的研究中不可分类的数据的比例低得多。例如, 使用我们的方法, 在 $m_r = 0.5$ 时, 14% 和 0% 的三体性样品分别对 96 孔和 384 孔数字 RNA-SNP 分析是不可分类的, 但是使用固定曲线时, 62% 和 10% 的样品分别是不可分类的 (表 900)。因此, 我们的方法允许用较少数目的信息孔对疾病分类。

[0191] 如表 900 所示, 对于 0.1 至 2.0 的全部 m_r 值, 新的算法在将样品分类为具有或不具有等位基因比率偏移上更准确。例如, 当 m_r 等于 1.0 并且进行了 96 孔数字 RNA-SNP 运行时, 新的算法分别正确地分类了 88% 和 92% 的具有或不具有等位基因比率偏移的样品, 而使用旧的算法, 具有或不具有等位基因比率偏移的样品的正确分类的百分比分别是 19%

和 36%。

[0192] 使用新的算法,具有或不具有等位基因比率偏移的样品的分离将随 m_t 而增加。因此,分类准确度将随 m_t 而增加。当 m_t 增加至大于 2.0 时,由于信息孔百分比的下降,两组样品分离的增加对分类准确度影响将降低。相比之下,使用旧的算法,当 m_t 增加时,由于预期的 P 值与其真实值偏离的增加,分类准确度明显地下降。

[0193] 我们的实验和模拟数据表明对于 21 三体性检测,数字 RNA-SNP 是有效和准确的方法。由于母体血浆中的 PLAC4mRNA 完全源自胎儿,所以对于 13 个检测的母体血浆样品中的 12 个,只需要一个 384 孔数字 PCR 来进行正确的分类(图 13B 的表 1350)。因此,这种均一的基于实时数字 PCR 的方法为 RNA-SNP 分析的基于质谱的方法提供了替代选择(Lo, YMD, et al. 2007 Nat Med, 见上文)。除了胎盘特异性转录物以外,我们还展望,母体血浆中其它类型的胎儿特异性核酸种类能够用于基于数字 PCR 的胎儿染色体非整倍性检测。一个实例是胎儿外遗传标记物(Chim, SSC et al. (2005) Proc Natl Acad Sci USA 102, 14753-14758; Chan, KCA et al. (2006) Clin Chem 52, 2211-2218),其最近被用于使用外遗传等位基因比率(EAR)方法的 18 三体性的无创产前检测(Tong, YK et al. (2006) Clin Chem 52, 2194-2202)。因此,我们,预测数字 EAR 将是可能的分析技术。

[0194] V. 渐增的百分比、多个标记物和 PCR 替代选择

[0195] 如上文所述,当胎儿 DNA 只组成母体血浆 DNA 的一小部分,并且妊娠的 11 周至 17 周之间的平均部分浓度为约 3% 时,本发明的实施方案应用于从母体血浆提取的 DNA 是复杂的。然而,如本文所示,甚至当非整倍体 DNA 以较少的群体存在时,数字 RCD 允许非整倍性检测。随着胎儿 DNA 部分浓度的下降,例如可以存在于早期妊娠期间,数字 RCD 需要较大数目的信息计数。如图 12 的表 1200 所总结的,本工作的意义是我们提供了可以在其上建立诊断测定的一组基准参数,例如所需的部分胎儿 DNA 和总模板分子。我们认为,对于 25% 的部分胎儿 DNA 浓度,7680 的反应总数是特别有吸引力的一组基准参数。如表 1200 所示,这些参数应当允许正确地分类当前 97% 的整倍体样品和 21 三体性样品。

[0196] 存在于单位体积的母体血浆中的血浆 DNA 分子的数目是有限的(Lo, YMD. et al. 1998 Am J Hum Genet 62, 768-7758)。例如,在早期妊娠中,常染色体基因座, β -球蛋白基因的中间母体血浆浓度被证实是 986 个拷贝/毫升,这既有来自胎儿的贡献又有来自母亲的贡献(Lo, YMD. et al. 1998 Am J Hum Genet 62, 768-7758)。为了捕获 7,680 个分子,需要提取约 8mL 的母体血浆的 DNA。可以从约 15mL 母体血液获得的这个体积的血浆是常规实践的极限。然而,我们预测对于数字 RCD 分析可以组合多组染色体 21 和参考染色体靶标。对于 5 对染色体 21 和参考染色体靶标,只需要 1.6mL 的母体血浆来提供分析所需的模板分子的数目。能够进行多重单分子 PCR。对于单分子单体型分析,以前已经证明了这种多重单分子分析的稳健度(Ding, C. and Cantor, CR. 2003 Proc Natl Acad Sci USA 100, 7449-7453)。

[0197] 可选择地,为了实现 25% 的部分胎儿 DNA 浓度,方法可以允许选择性地富集母体血浆中的胎儿的 DNA(Li, Y. et al. 2004 Clin Chem 50, 1002-1011) 或抑制母体血浆中的母体 DNA 背景(Dhallan, R et al. 2004 JAMA 291, 1114-1119) 或者既富集母体血浆中的胎儿 DNA 又抑制母体血浆中的母体 DNA 背景。除了富集胎儿 DNA 和抑制母体 DNA 的物理方法以外,还可能使用分子富集策略,例如通过靶向表现出特定 DNA 甲基化方式的胎儿 DNA 分

子 (Chim, SSC et al. 2005 Proc Natl Acad Sci USA 102, 14753-14758, Chan, KCA et al. 2006 Clin Chem 52, 2211-2218; Chiu, RWK et al. 2007 Am J Pathol 170, 941-950)。

[0198] 另外, 如进行数字 PCR 的目前的研究所用的, 现在有许多人工进行数字实时 PCR 分析的替代方法。这些替代方法包括微流体数字 PCR 芯片 (Warren, L et al. 2006 Proc Natl Acad Sci USA 103, 17807-17812; Ottesen, EA et al. 2006 Science 314, 1464-1467)、乳液 PCR (Dressman, D et al. 2003 Proc Natl Acad Sci USA 100, 8817-8822) 和使用例如 Roche454 平台、Illumina Solexa 平台和 Applied Biosystems 的 SOLiD™ 系统的大规模平行基因组测序 (Margulies, M. et al. 2005 Nature 437, 376-380)。对于后者, 我们的方法也适用于单个 DNA 分子的大规模平行测序方法, 该方法不要求扩增步骤, 例如 Helicos True 单分子 DNA 测序技术 (Harris TD et al. 2008 Science, 320, 106-109)、Pacific Biosciences 的单分子实时 (SMRT™) 技术和纳米孔测序 (Soni GV and Meller A. 2007 Clin Chem 53, 1996-2001)。通过使用这些方法, 能够在大量样品上快速地进行数字 RNA-SNP 和数字 RCD, 从而增强了本文提出的方法用于无创产前诊断的临床可行性。

实施例

[0199] 提供下文的实施例来示例而不是不限制要求保护的发明。

[0200] I. 计算机模拟

[0201] 进行计算机模拟来估计使用 SPRT 方法的 21 三体性诊断的准确性。使用 Microsoft Excel 2003 软件 (Microsoft Corp., USA) 和 Windows 软件 SAS 9.1 (SAS Institute Inc., NC, USA) 来进行计算机模拟。数字 PCR 的性能是参考模板浓度 (m_t)、信息计数的数目和等位基因或染色体失衡的预期程度 (P_t) 之间的相互作用。对一系列这些变量的每个进行了单独的模拟。由于数字 RNA-SNP 和数字 RCD 的 SPRT 曲线的判定边界是不同的, 所以分别进行这两个系统的模拟分析。

[0202] 对模拟的每个数字 PCR 条件 (即, m_t 、胎儿 DNA 部分浓度、总孔数), 进行了两轮模拟。第一轮模拟的情境是检测的样品获得自怀有整倍体胎儿的孕妇。第二轮模拟的情境是检测的样品获得自怀有 21 三体性胎儿的孕妇。每轮模拟检测了 5000 个胎儿。

[0203] A. RNA-SNP

[0204] 对于数字 RNA-SNP, 进行了 $m_t = 0.1$ 至 $m_t = 2.0$ 的 384 孔实验的模拟。在每个 m_t 值, 我们模拟了检测 5000 个整倍体胎儿和 5000 个 21 三体性胎儿的情境。使用对给定的 m_t 合适的 SPRT 曲线来分类所述 10,000 个胎儿。图 10 是本发明实施方案的表 1000, 该表 1000 表示正确或错误地分类为整倍体或非整倍体的胎儿, 以及对于给定的信息计数不可分类的胎儿的百分比。对于 0.5 至 2.0 的 m_t , 诊断整倍体和非整倍体情况的准确度都是 100%。当 $m_t = 0.1$ 时, 384 孔分析后, 只有 57% 的整倍体胎儿和 88% 的 21 三体性胎儿能够被准确地分类。

[0205] 按下述步骤生成模拟数据:

[0206] 在步骤 1 中, 对每个孔, 利用 SAS 程序的随机 (泊松) 函数 (www.sas.com/technologies/analytics/statistics/index.html) 生成两个随机数来分别表示 A 等位基因和 G 等位基因。该随机 (泊松) 函数将生成从 0 开始的正整数 (即, 0、1、2、3...), 对于表示每孔的等位基因的平均浓度的给定的平均值, 生成每个整数的概率取决于这个数根据

泊松概率密度函数的概率。如果表示 A 等位基因的随机数大于 0, 则孔被视为对 A 等位基因是阳性的, 即含有 A 等位基因的一个或多个分子。类似地, 如果表示 G 等位基因的随机数大于 0, 则孔被视为对 G 等位基因是阳性的。

[0207] 为了模拟怀有整倍体胎儿的孕妇的情境, 使用相同的平均值来生成对于 A 等位基因和 G 等位基因的随机数。例如, 在模拟 $m_t = 0.5$ 的数字 RNA-SNP 分析的分析中, 将对于 A 等位基因或 G 等位基因的平均值相同地设为 0.5, 这表示任何一个等位基因的平均浓度是每孔 0.5 个分子。利用泊松方程式, 在 0.5 的平均浓度下, 对 A 等位基因或 G 等位基因是阳性的孔的比例将是相同的, 并且是 0.3935, 参见表 600。

[0208] 当模拟 $m_t = 0.5$ 的怀有 21 三体性胎儿的孕妇的数字 RNA-SNP 分析时, 预期每孔的过度表现的等位基因的平均浓度将是参考等位基因的平均浓度即 1 的 2 倍。在这种情况下, 对过度表现的等位基因是阳性的孔的概率是 0.6321, 参见表 600。

[0209] 生成数字 PCR 孔的随机数后, 该孔能够被分类为以下情况的一种:

[0210] a. 对 A 等位基因和 G 等位基因都是阴性

[0211] b. 对 A 等位基因和 G 等位基因都是阳性

[0212] c. 对 A 等位基因是阳性, 但是对 G 等位基因是阴性

[0213] d. 对 G 等位基因是阳性, 但是对 A 等位基因是阴性

[0214] 在步骤 2 中, 重复步骤 1 至生成了期望数目的孔, 对当前的模拟是 384 个孔。计数了只对 A 等位基因是阳性的孔的数目和只对 G 等位基因是阳性的孔的数目。将具有较少阳性孔的等位基因视为参考等位基因, 并且将具有较多阳性孔的等位基因视为潜在地过度表现的等位基因。信息孔的数目是对任何一个等位基因是阳性, 但是不对两个等位基因都是阳性的孔的总数。然后计算了含有潜在地过度表现的等位基因的信息孔的比例 (P_t)。根据本发明的实施方案, 计算了接受无效假设或备选假设的相关 SPRT 曲线的上限和下限。

[0215] 在步骤 3 中, 对怀有整倍体胎儿或 21 三体性胎儿的孕妇的两种情境的每一种进行了 5000 个模拟。每个模拟能够被视为从孕妇获得的独立生物样品。在表 1000 中, 正确分类的整倍体个例指接受了无效假设的那些整倍体个例, 而错误分类的整倍体个例指接受了备选假设的那些整倍体个例。类似地, 接受备选假设的那些 21 三体性个例被视为正确的分类, 而接受无效假设的那些 21 三体性个例被视为错误的分类。对于两个组, 在模拟了预先指定的总数的孔后, 那些既没有接受无效假设又没有接受备选假设的个例被视为不可分类的。

[0216] 在步骤 4 中, 对以 0.1 的增量增加的 0.1 至 2.0 的 m_t , 进行步骤 1 至步骤 3。

[0217] B. RCD

[0218] 图 11 是本发明实施方案的表 1100, 该表 1000 显示了对于从 0.1 至 2.0 的 m_t , 纯 (100%) 胎儿 DNA 样品的数字 RCD 分析的计算机模拟。随着部分胎儿 DNA 浓度变小, 过度表现的染色体 21 的程度降低, 因此需要用于准确的疾病分类的更多数目的信息孔。因此, 还对在 $m_t = 0.5$ 时, 384 孔至 7680 孔的总孔数的 50%、25% 和 10% 的胎儿 DNA 浓度进行了模拟。

[0219] 图 12 是本发明实施方案的表 1200, 该表 1200 显示了对于来自具有不同胎儿 DNA 部分浓度的整倍体胎儿或 21 三体性胎儿的样品的分类, 在 $m_t = 0.5$ 时, 数字 RCD 分析的准确度的计算机模拟的结果。数字 RCD 的效能对于胎儿 DNA 部分浓度较高的个例更好。在胎儿 DNA 浓度为 25% 并且 PCR 分析的总数为 7680 时, 97% 的整倍体和非整倍体个例都是可以

分类的,并且没有错误的分类。剩下的 3%个例在能够实现分类前,需要进一步的分析。

[0220] 模拟数字 RCD 分析的过程类似于对数字 RNA-SNP 所述的那些过程。模拟的步骤如下所述:

[0221] 在步骤 1 中,在泊松概率密度函数下产生两个随机数来表示参考等位基因座,染色体 1 基因座和染色体 21 基因座。对于怀有整倍体胎儿的个体,染色体 1 基因座和染色体 21 基因座的平均浓度是相同的。在一模拟分析中,使用了每个基因座每孔 0.5 的平均模板浓度。如表 700 所示,对于怀有 21 三体性胎儿的个体,这一模拟中的 m_r 是 0.5,但是每孔的染色体 21 基因座的平均浓度将取决于检测样品中部分胎儿 DNA 浓度。通过代表各自的基因座的随机数来确定参考基因座和 / 或染色体 21 基因座向孔的分布,所述随机数是根据泊松概率密度函数,用每孔的基因座的合适的平均浓度产生的。

[0222] 在步骤 2 中,重复步骤 1 至生成了期望数目的孔,例如,384 孔板实验的 384 个孔。计数了只对染色体 1 是阳性的孔的数目和只对染色体 21 是阳性的孔的数目。信息孔的总数是对上述染色体任一条是阳性而不对两条染色体都是阳性的孔的总数。然后计算了对染色体 21 是阳性的信息孔的比例 (P_r)。如上文关于 SPRT 分析的部分所述,计算了接受无效假设或备选假设的相关 SPRT 曲线的上限和下限。

[0223] 在步骤 3 中,对怀有整倍体胎儿或 21 三体性胎儿的孕妇的两种情境的每一种进行了 5000 个模拟。每个模拟能够被视为从孕妇获得的独立生物样品。在表 1100 中,正确分类的整倍体个例指接受了无效假设的那些整倍体个例,而错误分类的整倍体个例指接受了备选假设的那些整倍体个例。类似地,接受备选假设的那些 21 三体性个例被视为正确分类的,而接受无效假设的那些 21 三体性个例被视为错误分类的。对于两个组,在模拟了预先指定的总数的孔后,那些既没接受无效假设又没接收备选假设的个例被视为不可分类的。

[0224] 在步骤 4 中,在 384 至 7680 的总孔数下,对 10%、25%、50% 和 100% 的胎儿 DNA 的样品重复步骤 1 至步骤 3。

[0225] II. 21 三体性检测的验证

[0226] A. PLAC4 的 RNA-SNP

[0227] 利用染色体 21 上的 PLAC4 基因的 rs8130833 SNP 来证明数字 RNA-SNP 的实际可行性 (Lo, YMD et al. 2007 Nat Med 13, 218-223)。分析了来自两个整倍体杂合胎盘和两个 21 三体性杂合胎盘的胎盘 DNA 和 RNA 样品。用数字 RNA-SNP 方法来分析胎盘 DNA 样品,但是省略了逆转录步骤,因此基本上将该方法转变成数字 DNA-SNP 分析。为了达到正确的个例分类的可能性与信息孔的比例之间的平衡,我们将样品稀释旨在实现每孔一个任何类型的等位基因,并且通过 96 孔数字 PCR 分析来证实。然后,进行 384 孔数字 RNA-SNP 实验。计算了 P_r 和 m_r ,并将这个 m_r 值的 SPRT 曲线用于数据解释。

[0228] 图 13A 显示了按照本发明的实施方案,整倍体妊娠和 21 三体性妊娠的胎盘组织的数字 RNA-SNP 分析的表 1300。通过质谱测定来确定基因型。当实验获得的 P_r 低于不可分类的区域时,指定为“整倍体”;当实验获得的 P_r 高于不可分类的区域时,指定为表示 21 三体性的“T21”。T21 即 21 三体性。通过一个 384 孔实验,既利用 DNA 样品又利用 RNA 样品正确地分类了这些个例每一个。

[0229] 我们还检测了来自 9 个怀有整倍体胎儿的妇女和 4 个怀有 21 三体性胎儿的妇女的血浆 RNA 样品。图 13B 显示了按照本发明的实施方案,来自整倍体妊娠和 21 三体性妊

娠的母体血浆的数字 RNA-SNP 分析的表 1350。正确分类了全部个例。一个 21 三体性实例 (M2272P) 在一个 384 孔实验后, 的初始结果落在 SPRT 曲线之间的不可分类区域中。因此, 进行了另外的 384 孔实验。从总共 768 个孔汇集的数据计算了新的 m_t 和 P_t 值, 并且使用基于这个 m_t 值选择的新的一组 SPRT 曲线进行了分类。然后, 该个例被正确地评分为非整倍体。

[0230] 我们的实验和模拟数据表明数字 RNA-SNP 对 21 三体性检测是有效且准确的方法。由于母体血浆中的 PLAC4 mRNA 全部源自胎儿, 所以对于 13 个检测的母体血浆样品中的 12 个, 正确的分类只要求一个 384 孔数字 PCR 实验。因此, 这种均一的基于实时数字 PCR 的方法提供了用于 RNA-SNP 分析的基于质谱的方法的替代选择。除了胎盘特异性 mRNA 转录物以外, 我们还预测母体血浆中的其它类型的胎儿特异性核酸种类能够用于胎儿染色体非整倍性的基于数字 PCR 的检测。一个实例是胎儿外遗传标记物, 其最近被用于利用外遗传等位基因比率 (EAR) 方法的 18 三体性的无创产前检测 (Tong YK et al. 2006 ClinChern, 52, 2194-2202)。因此, 我们预测数字 EAR 将是可能的分析技术。

[0231] B. RCD

[0232] 还利用靶向染色体 21 和染色体 1 上的种内同源序列的 PCR 测定来研究用于检测 21 三体性的数字 RCD 的实际可行性。作为示例, 本文使用了种内同源基因座 (paralogous loci)。染色体 21 和其它参考染色体上的非种内同源序列也能够用于 RCD。将来自两个整倍体胎盘和两个 21 三体性胎盘的胎盘 DNA 样品稀释至约每孔来自任何一条染色体的一个靶模板, 并且通过 96 孔数字 PCR 分析来证实。通过 384 孔数字 RCD 实验来分析每个证实的样品, 并且计算了 P_t 和 m_t 的值。对于数字 RCD, 染色体 1 种内同源基因 (paralog) 是参考模板。将这个 m_t 值用于选择解释数据的一组对应的 SPRT 曲线。如图 14A 所示, 正确地分类了全部胎盘样品。

[0233] 为了证明数字 RCD 方法能够用于检测与过量的整倍体 DNA 混合的 21 三体性 DNA, 例如母体血浆中的胎儿 DNA 的情境, 在整倍体母体血液细胞 DNA 的背景下, 分析了含有 50% 和 25% 的 21 三体性胎盘 DNA 的混合物。将来自 10 个 21 三体性个例和 10 个整倍体个例的胎盘 DNA 分别与等量的整倍体母体血液细胞 DNA 混合, 从而获得 20 个 50% 的 DNA 混合物。图 14B 显示按照本发明的实施方案, 示出 50% 胎儿 DNA 混合物的 RCD 分析的 SPRT 解释的图 1440。类似地, 将来自 5 个 21 三体性个例和 5 个整倍体个例的胎盘 DNA 分别与超出 3 倍的整倍体母体血液细胞 DNA 混合, 从而获得 10 个 25% 的 DNA 混合物。图 14C 显示了示出 25% 胎儿 DNA 混合物的 RCD 分析的 SPRT 解释的图 1470。如图 14B 和 14C 所示, 正确地分类了全部的整倍体和非整倍体 DNA 混合物。

[0234] 如图 14B 和 14C 所表示的, 多个 384 孔数字 PCR 分析后, 每个样品达到了可以分类的点。50% 的 DNA 混合物所要求的 384 孔板的数目是 1 至 5。25% 的 DNA 混合物所要求的 384 孔板的数目是 1 至 7。通过如表 1200 所示的计算机模拟, 预测了随每个 384 数字 PCR 分析的增加, 正确分类的个例的累积比例。

[0235] III. 数字 PCR 方法

[0236] A. 数字 RNA-SNP

[0237] 首先使用 ThermoScript 逆转录酶 (Invitrogen), 利用基因特异性的逆转录引物, 将全部 RNA 样品进行逆转录。逆转录引物的序列是 5' -AGTATATAGAACCATGTTTAGGCCA

GA-3' (Integrated DNA Technologies, Coralville, IA)。用于数字 RNA-SNP 的逆转录 RNA 样品 (即, cDNA) 随后的处理与 DNA 样品 (例如, 胎盘 DNA) 基本上相同。在数字 PCR 分析前, 首先利用针对 PLAC4 的实时 PCR 测定, 将 DNA 样品和 cDNA 样品定量, 该测定由引物 5'-CCGCTAGGGTGTCTTTTAAGC-3'、5'-GTGTTGCAATACAAAATGAGTTTCT-3' 和荧光探针 5'-(FAM)ATTGGAGCAAATTC(MGBNFQ)-3' (Applied Biosystems, Foster City, CA) 组成, 其中 FAM 是 6-羧基荧光素并且 MGBNFQ 是小沟结合非荧光猝灭剂。

[0238] 通过指定扩增子的 HPLC 纯化的单链合成的 DNA 寡核苷酸 (Proligo, Singapore) 的连续稀释来制备校准曲线。序列是 5'-CGCCGCTAGGGTGTCTTTTAAGCTATTGGAGCAAATTCAAATTGGCTTAAAGAAAAAGAACTCATTTTGTATTGCAACACCAGGAGTATCCCAAGGGACTCG-3'。使用 2X TaqMan Universal PCR Master Mix (Applied Biosystems) 进行反应, 反应体积为 25 μ L。在每个反应中使用 400nM 的每种引物和 80nM 的探针。在 50°C 下开始反应持续 2min, 然后 95°C、10min 以及 95°C、15s 与 60°C、1min 的 45 个循环, 在 ABI PRISM 7900HT 序列检测系统 (Applied Biosystems) 中进行上述反应。然后将 DNA 或 cDNA 样品进行连续稀释, 从而随后的数字 PCR 扩增能够在约每孔 1 个模板分子下进行。在这种浓度下, 预期约 37% 的反应孔将表现出阴性扩增, 并且首先通过进行 96 孔数字实时 PCR 分析来证实。然后使用一组非内含子跨越引物 (non-intron spanning primer) 在 384 孔板中进行数字 RNA-SNP 分析: 正向引物是 5'-TTTGTATTGCAACACCATTGG-3', 基因特异性逆转录引物如上文所述。

[0239] 设计了靶向 PLAC4 序列上的 rs8130833SNP 的两个等位基因的每一个的两个等位基因特异性 TaqMan 探针。它们对于 G 等位基因和 A 等位基因的序列分别是 5'-(FAM)TCGTCGTCTAACTTG(MGBNFQ)-3' 和 5'-(VIC)ATTCGTCATCTAACTTG(MGBNFQ)。使用 2X TaqMan Universal PCR Master Mix (Applied Biosystems) 进行反应, 反应体积为 5 μ L。每个反应含有 1X TaqMan Universal PCR Master Mix、572nM 的每种引物、107nM 的等位基因 -G- 特异性探针和 357nM 等位基因 -A- 特异性探针。在 ABI PRISM 7900HT 序列检测系统中进行反应。在 50°C 下开始反应持续 2min, 然后 95°C、10min 以及 95°C、15s 与 57°C、1min 的 45 个循环。在反应期间, 通过 SDS 2.2.2 软件 (Applied Biosystems) 的“绝对定量”应用来收集荧光数据。该软件自动地计算基线和阈值。记录了对 A 等位基因或 G 等位基因是阳性的孔的数目, 并对其进行 SPRT 分析。

[0240] B. 数字 RCD 分析

[0241] 首先通过 NanoDrop 分光光度计 (NanoDrop Technology, Wilmington, DE), 将在本研究中所用的全部胎盘和母体血沉棕黄色层 DNA 样品进行定量。利用 6.6pg/细胞的换算, 将 DNA 浓度转换成拷贝 / μ L。通过将 DNA 样品连续稀释来确定对应于约每孔一个模板的 DNA 的量, 并通过 96 孔格式的实时 PCR 测定来证实, 其中我们预期约 37% 的孔表现出阴性扩增。除了只添加了参考染色体的探针以外, 确认板 (confirmatory plate) 的 PCR 设置和下文所述的相同。在数字 RCD 分析中, 首先通过正向引物 5'-GTTGTTCTGCAAAAAACCTTCGA-3' 和反向引物 5'-CTTGGCCAGAAATACTTCATTACCATAT-3', 将染色体 21 和染色体 1 上的种内同源基因座 (Deutsch, S. et al. 2004 J Med Genet 41, 908-915) 进行共扩增。设计了两个染色体特异性 TaqMan 探针来靶向染色体 21 和染色体 1 种内同源基因, 并且它们的序列分别是 5'-(FAM)TACCTCCATAATGAGTAAA(MGBNFQ)-3' 和 5'-(VIC)CGTACCTCTGTAATGTGTAA(MGBNFQ)-3'。每个反应含有 1X TaqMan Universal PCR Master

Mix、450nM 的每种引物和 125nM 的每种探针。总反应体积是 5 μ L/ 孔。在 50°C 下开始反应持续 2min, 然后 95°C、10min 以及 95°C、15s 与 60°C、1min 的 50 个循环。在 ABI PRISM 7900HT 序列检测系统 (Applied Biosystems) 中进行全部实时 PCR 实验, 并通过 SDS 2.2.2 软件 (Applied Biosystems) 的“绝对定量”应用来收集荧光数据。使用了默认的基线和手动阈值 (manual threshold)。记录了对染色体 21 或染色体 1 是阳性的孔的数目, 并对其进行 SPRT 分析。将分析一个或多个 384 孔板直至疾病通过 SPRT 进行分类是可能的。

[0242] IV. 使用基于微流体的数字 PCR

[0243] A. 数字 RNA-SNP

[0244] 本实施例证明了使用基于微流体的数字 PCR 的数字 PCR 分析的性能。本文示例了这种方法的一个变体, 为了示例而不是为了限制, 使用 Fluidigm BioMark™ 系统。这个系统每个运行能够进行超过 9000 个数字 PCR。

[0245] 从怀有整倍体胎儿或 21 三体性胎儿的孕妇获得胎盘组织样品和母体外周血液样品。通过引物延伸及随后的质谱来进行胎盘 DNA 样品中的 PLAC4 基因上的 rs8130833 SNP 的基因型分型。从胎盘样品和母体血浆样品中提取 RNA。

[0246] 使用 ThermoScript 逆转录酶 (Invitrogen), 利用基因特异性逆转录引物 (5'-AGT ATATAGAACCATGTTTAGGCCAGA-3'), 将全部 RNA 样品逆转录。对胎盘 cDNA 样品, 进行连续稀释, 从而随后的数字 PCR 扩增能够在约每孔一个模板分子下进行。

[0247] 在具有 12.765 数字阵列 (Fluidigm) 的 BioMark System™ (Fluidigm) 上进行数字 PCR。每个数字阵列由用于容纳 12 个样品 - 测定混合物的 12 个板组成。每个板还分成进行 7nL 反应 / 孔的 765 个孔。通过正向引物 (5'-TTTGTATTGCAACACCATTGG-3') 和上文所述的基因特异性逆转录引物来扩增 PLAC4 基因上的 rs8130833 SNP 区域。设计了靶向 rs8130833 SNP 的两个等位基因的每一个的两个等位基因特异性 TaqMan 探针。它们对 G 等位基因和 A 等位基因的序列分别是 5'-(FAM) TCGTCGTCTAACTTG (MGBNFQ)-3' 和 5'-(VIC) ATTCGTCATCTAACTTG (MGBNFQ)-3'。利用 2X TaqMan Universal PCR Master Mix 来进行一个阵列板的反应, 反应体积为 10 μ L。每个反应含有 1X TaqMan Universal PCR Master Mix、572nM 的每种引物、53.5nM 的等位基因 -G- 特异性探针、178.5nM 的等位基因 -A- 特异性探针和 3.5 μ L 的 cDNA 样品。对每个胎盘 cDNA 样品使用一个反应板, 而对每个母体血浆样品使用 12 个板。通过 NanoFlex™ IFC 控制器 (Fluidigm), 将样品 - 测定混合物加载至数字阵列中。在 BioMark™ 系统中进行反应。在 50°C 下开始反应持续 2min, 然后 95°C、10min 以及 95°C、15s 与 57°C、1min 的 40 个循环。

[0248] 在 765 孔反应板中分析来自一个整倍体杂合胎盘和两个 T21 杂合胎盘的胎盘 RNA 样品。对于每个样品, 计数了包括对 A 等位基因或 G 等位基因是阳性 (但是不对这两个等位基因都是阳性) 的孔的信息孔的数目。确定了在全部信息孔中过度表现的等位基因的比例 (P_r)。应用对这些运行的每孔确切的平均参考模板浓度 (m_r) 合适的 SPRT 曲线来确定实验获得的 P_r 是否表示了整倍体样品或 T21 样品。如图 15A 所示, 利用这种方法, 正确地分类了全部 RNA 样品。

[0249] 我们还检测了来自 4 个怀有整倍体胎儿的妇女和 1 个怀有 21 三体性胎儿的妇女的血浆 RNA 样品。在 12 个 765 孔反应板中分析每个样品, 即每个血浆 RNA 样品 9180 个反应。图 15B 表示这个血浆 RNA 样品的 12 个板的每一个的信息孔的数目。如该表中所示, 血

浆样品中的模板浓度被稀释至任一反应板中的信息孔的数目不足以进行 SPRT 分类。在将样品分类为整倍体样品前,必须将来自三个反应板的信息孔合并(图 15C)。图 15C 表明,利用从 2 至 12 个板汇集的数据,全部血浆个例能够被正确地分类。

[0250] 与进行数字 PCR 的手动方法相比,这种基于微流体的方法快的多,并且劳动强度小得多。能够在 2 个半小时内完成全部过程。

[0251] B. 用于 18 三体性产前检测的数字 RNA-SNP

[0252] 在本实施例中,我们在染色体 18 上的胎盘表达的转录物,即丝氨酸蛋白酶抑制剂肽酶抑制剂进化枝 (clade)B(卵清蛋白)成员 2(SERPINB2)mRNA 上进行基于数字 PCR 的等位基因鉴别测定,来检测 18 三体性胎儿的多态性等位基因比率的失衡。如生产商的说明书所述,分别使用 QIAamp DNA Mini 试剂盒 (Qiagen, Hilden, Germany) 和 TRIzol 试剂 (Invitrogen, Carlsbad, CA) 从胎盘组织样品中提取 DNA 和 RNA。提取的胎盘 RNA 样品进行 DNase I (Invitrogen) 处理以除去污染的基因组 DNA。如上文所述,利用 MassARRAY Compact (Sequenom, San Diego), 使用同类 MassEXTEND (hME) 测定 (homogenous MassEXTEND assay), 在胎盘组织 DNA 样品中进行 SERPINB2 基因上的 rs6098 SNP 的基因型分型。

[0253] 使用基因特异性引物 5'-CGCAGACTTCTCACCAACA-3' (Integrated DNA Technologies, Coralville, IA), 利用 ThermoScript 逆转录酶 (Invitrogen), 在胎盘组织 RNA 样品上进行 SERPINB2 转录物的逆转录。全部 cDNA 样品稀释的浓度使得随后的数字 PCR 扩增能够在每个反应孔一个模板分子的平均浓度下进行。使用 TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA) 和 Biomark™ PCR 试剂 (Fluidigm, San Francisco) 进行数字 PCR。所用的正向引物 5'-CTCAGCTCTGCAATCAATGC-3' (Integrated DNA Technologies) 和反向引物 (与用于逆转录的基因特异性引物相同) 的浓度是 600nM。靶向 SERPINB2 序列上的 rs6098 SNP 的 A 等位基因或 G 等位基因的两个 TaqMan 探针是 5'-(FAM)CCACAGGGAATTATTT (MGBNFQ)-3' 和 5'-(VIC)CCACAGGGGATTATTT (MGBNFQ)-3' (Applied Biosystems)。FAM 是 6-羧基荧光素, MGBNFQ 是小沟结合非荧光猝灭剂, 并且 FAM 和 MGBNFQ 的使用浓度分别是 300nM 和 500nM。利用 Nanoflex™ IFC 控制器 (Fluidigm), 将每个样品-试剂混合物分配至 Biomark™12.765 数字阵列上的 765 个反应孔。分配后, 将该阵列置于 Biomark™ 实时 PCR 系统 (Fluidigm) 中进行热扩增和荧光检测。在 50℃ 下开始反应持续 2min, 在 95℃ 下继续 5min, 然后进行 95℃、15sec 与 59℃、1min 的 45 个循环。扩增后, 计数信息孔的数目 (只对 A 等位基因或 G 等位基因是阳性的孔) 和对两个等位基因都是阳性的孔的数目, 并对它们进行序贯概率比检验 (SPRT) 分析。

[0254] 对于杂合的整倍体胎儿, A 等位基因和 G 等位基因应当在胎儿基因组中相等地表现 (1 : 1), 而对于 18 三体性, 将有一个等位基因的额外拷贝, 从而在胎儿基因组中表现出 2 : 1 的比率。对不同样品的解释产生一系列 SPRT 曲线。这些曲线示出了对于分类所需的给定的信息孔总数 (x 轴), 对过度表现的等位基因是阳性的信息孔的预期比例 P_r (y 轴)。对于每个样品, 将实验导出的 P_r 与预期的 P_r 值进行比较。将高于上部曲线的样品分类为 18 三体性, 而将低于底部曲线的样品分类为整倍体。两个曲线之间的面积是不可分类的区域。

[0255] 通过利用 SERPINB2 基因上的 rs6098 SNP 来证明用于检测胎儿 18 三体性的数字 RNA-SNP 分析的可行性。首先通过质谱对来自具有整倍体胎儿和 18 三体性胎儿的个体的

胎盘组织 DNA 样品进行基因型分型,以鉴定杂合个例。发现了 9 个整倍体杂合胎盘和 3 个 18 三体性 杂合胎盘,并对它们进行数字 RNA-SNP 分析。对于每个样品,计算了 P_r 和 m_r ,并且将这个 m_r 值的 SPRT 曲线用于疾病分类。如图 16A 所示,正确地分类了全部样品。18 三体性胎盘的 P_r 值高于不可分类的区域,而整倍体胎盘的 P_r 值则低于这个区域。

[0256] 具有基于 $m_r = 0.1$ 、 0.2 和 0.3 的 SPRT 曲线的样品示于图 16B 中。这些数据表明数字 RNA-SNP 方法对 18 三体性妊娠是有价值的诊断工具。两个曲线描绘出不可分类区域的界限。将数据点高于上部曲线的样品分类为非整倍体,而将数据点低于底部曲线的样品分类为整倍体。

[0257] C. 数字 RCD 分析

[0258] 本实施例证明了利用基于微流体的数字 PCR 的数字 RCD 分析的效能。作为示例但不作为限制,利用 Fluidigm BioMark™ 系统,在此处示例了这种方法的一个变体。这种系统的每个运行能进行超过 9000 个数字 PCR。

[0259] 从怀有整倍体胎儿或 21 三体性 (T21) 胎儿的孕妇获得胎盘组织、母体血液细胞和羊水样品。将来个 10 个 T21 个例和 10 个整倍体个例的胎盘 DNA 分别与等量的整倍体母体血液细胞 DNA 混合,从而获得 20 个 50% 的 DNA 混合物。为了保证该混合物样品中准确的胎儿比例,首先通过 260nm 下的光密度 (OD) 测量来定量提取的 DNA。然后利用 12.765 数字阵列 (Fluidigm),通过 BioMark™ 系统 (Fluidigm) 来数字地定量该提取的 DNA。除了只使用参考染色体的探针以外,用于定量样品的测定与下文所述的相同。

[0260] 通过数字 PCR 对相对于位于染色体 1 的基因座的非多态性染色体 21 基因座进行分析来测定 50% 的 DNA 混合物和羊水样品中的染色体剂量。首先通过正向引物 5' -GTTGTTCTGCAAAAAACCTTCGA-3' 和反向引物 5' -CTTGGCCAGAAATACTTCATTACC ATAT-3' 来共扩增染色体 21 和染色体 1 上的一对种内同源基因座的 101-bp 的扩增子。设计了两个染色体特异性 TaqMan 探针来区分染色体 21 和染色体 1 的种内同源基因,且探针的序列分别是 5' -(FAM) TACCTCCATAATGAGTAAA (MGBNFQ) -3' 和 5' -(VIC) CGTACCTCTGTAATGTGTAA (MGBNFQ) -3'。仅作为示例在 此处使用种内同源基因座。换言之,非种内同源基因座也能够用于这种分析。

[0261] 为了证明数字 RCD 方法用于检测 18 三体性 (T18) 的,设计了靶向染色体 21 和染色体 18 上的种内同源序列的另一测定。首先通过正向引物 5' -GTACAGAAACCACAAACTGATCGG-3' 和反向引物 5' -GTCCAGGCTGTGGCCT-3' 来共扩增染色体 21 和染色体 18 上的种内同源基因座的 128-bp 的扩增子。设计了两个染色体特异性 TaqMan 探针来区分染色体 21 和染色体 18 的种内同源基因,且探针的序列分别是 5' -(FAM) AAGAGGCGAGGCAA (MGBNFQ) -3' 和 5' -(VIC) AAGAGGACAGGCAAC (MGBNFQ) -3'。仅作为示例在此处使用种内同源基因座。换言之,非种内同源基因座也能够用于这种分析。

[0262] 利用 12.765 数字阵列 (Fluidigm),在 BioMark™ 系统 (Fluidigm) 上进行全部实验。利用 2X TaqMan Universal PCR Master Mix (Applied Biosystems) 在 10 μ L 的反应体积下进行一个板的反应。每个反应含有 1X TaqMan Universal PCR Master Mix、900nM 的每种引物、125nM 的每种探针和 3.5 μ L 的 50% 胎盘 / 母体血液细胞 DNA 样品。通过 NanoFlex™ IFC 控制器 (Fluidigm) 将样品 / 测定混合物加载至数字阵列。在 50°C 下开始反应持续 2min,然后 95°C、10min 以及 95°C、15s 和 57°C、1min 的 40 个循环。

[0263] 通过染色体 21/ 染色体 1(chr21/chr1) 测定,在数字阵列上分析整倍体和 T21 50%胎盘 / 母体血液细胞 DNA 样品。对于每个样品,计数了包括只对染色体 21 标记物或染色体 1 标记物是阳性(但不对这两个标记物都是阳性)的孔的信息孔的数目。确定了在全部信息孔中过度表现的标记物的比例(P_r)。将对任一数字 PCR 板的每孔的确切平均参考模板浓度(m_r)合适的 SPRT 曲线用于确定实验获得的 P_r 是否表示整倍体样品或 T21 样品。对尚未分类的样品,从额外的板汇集数据直至能够做出判定。如图 17 所示,利用这种方法和所需的 1 个到 4 个板的数据,正确地分类了全部 50%胎盘 / 母体血液细胞 DNA 样品。如图 18 所示,还将 SPRT 曲线作图来显示用于正确分类的判定边界(decision boundaries)。

[0264] 我们还将 RCD 分析应用于从 23 个怀有整倍体胎儿的妇女和 6 个 怀有 T21 胎儿的妇女获得的羊水样品。使用染色体 21/ 染色体 1 测定,在单个 765 孔反应板中分析了每个样品。图 19 显示了 SPRT 分类的汇总。如图 19 所示,正确地分类了全部 29 个样品。因此,数字 RCD 方法是在诸如羊水和绒毛膜绒毛活组织检查的用于产前诊断的多种样品类型中,利用微卫星标记物(Levett LJ,et al.A large-scale evaluation of amnio-PCR for the rapid prenatal diagnosis of fetal trisomy(用于胎儿三体性快速产前诊断的羊膜 PCR 的大规模评价).Ultrasound Obstet Gynecol 2001;17:115-8)或单核苷酸多态性(SNP)(Tsui NB,et al.Detection of 21 trisomy by quantitative mass spectrometric analysis of single-nucleotide polymorphisms(通过单核苷酸多态性的定量质谱分析检测 21 三体性).Clin Chem 2005;51:2358-62)标记物或实时非数字 PCR(Zimmermann B, et al.Novel realtime quantitative PCR test for 21trisomy(21 三体性的新实时定量 PCR 检测).Clin Chem 2002;48:362-3)检测三体性的替代方法。

[0265] 在检测 T18 个例的尝试中,我们将染色体 21/ 染色体 18(chr21/chr18) 测定应用于 3 个整倍体胎盘 DNA 样品和 5 个 T18 胎盘 DNA 样品。计算了全部信息孔中过度表现的标记物的比例(P_r)。除了一个 T18 个例被错误地分类为整倍体以外,正确地分类了全部样品。结果汇总于图 20 中。

[0266] V. 在质谱平台上使用多重数字 RCD 测定

[0267] 每单位体积的母体血浆存在的血浆 DNA 分子的数目是有限的(Lo YMD. et al. 1998 Am J Hum Genet 62,768-7758)。例如,在早期妊娠中,常染色体基因座,即 β - 球蛋白基因的母体血浆浓度中位数被证实是 986 个拷贝 /mL,这既有来自胎儿的贡献又有来自母亲的贡献(Lo YMD. et al. 1998 Am J Hum Genet 62,768-7758)。为了捕获 7,680 个分子,需要从约 8mL 母体血浆提取的 DNA。可以从约 15mL 母体血液获得的这个体积的血浆是常规实践的极限。然而,我们预测对于数字 RCD 分析,可以组合多组染色体 21 和参考染色体靶标。对于 5 对染色体 21 和参考染色体靶标,只需要 1.6mL 的母体血浆来提供分析所需的模板分子的数目。能够进行多重单分子 PCR。以前已经证明了这种多重单分子分析对于单分子单体型分析的稳健度(Ding, C. and Cantor, CR. 2003 Proc Natl Acad Sci USA 100, 7449-7453)。

[0268] 在一实施例中,从怀有整倍体胎儿或 21 三体性(T21) 胎儿的孕妇获得胎盘组织和母体血液细胞样品。将 5 个整倍体胎盘 DNA 样品和 5 个 T21 胎盘 DNA 样品分别与等比例的母体血液细胞 DNA 混合,以获得 10 个模拟 50%胎儿 DNA 的血浆样品的 DNA 混合物。为了保证混合物样品中准确的胎儿比例,首先通过 260nm 下的光密度(OD) 测量来定量提取的 DNA。

然后通过 384 孔格式的实时 PCR 对提取的 DNA 进行数字定量。用于定量样品的测定与上文的数字 RCD 分析的实施例中所述的相同。

[0269] 通过相对于位于染色体 1 上的基因座的非多态性染色体 21 基因座的数字 PCR 分析来确定 50% 的混合物中的染色体剂量。这种方法被称为数字相对染色体剂量 (Digital Relative Chromosome Dosage, RCD) 分析。通过正向引物 5' -ACGTTGGATGGTTGTTCTGCAAA AACCTTCGA-3' 和反向引物 5' -ACGTTGGATGCTTGGCCAGAAATACTTCATTACCATAT-3' 来共扩增染色体 21 和染色体 1 上的一对种内同源基因座的 121-bp 的扩增子 (包括每种引物上的 10-mer)。设计了靶向染色体 21 和染色体 1 之间的碱基差异的延伸引物, 该延伸引物的序列是 5' -CTCATCCTCACTTCGTACCTC-3'。

[0270] 为了证明多重数字 PCR 测定检测 T21 个例 (case) 的用途, 设计了靶向染色体 21 和染色体 18 上的种内同源序列的另一数字 RCD 测定。通过正向引物 5' -ACGTTGGATGGTACAGAAACCACAAACTGATCGG-3' 和反向引物 5' -ACGTTGGATGGTCCAGGCTGTGGGCCT-3' 来共扩增染色体 21 和染色体 18 上的种内同源基因座的 148-bp 的扩增子 (包括每种引物上的 10-mer)。设计了靶向染色体 21 和染色体 18 之间的碱基差异的延伸引物, 该延伸引物的序列是 5' -ACAAAAGGGGAAGAGG-3'。

[0271] 利用延伸引物方案来进行多重数字 RCD 分析。利用 GeneAmp PCR Core 试剂盒 (Applied Biosystems) 来进行 PCR 反应, 并且反应体积为 5 μ L。每个反应含有 1X Buffer II、2mM 的 $MgCl_2$ 、200 μ M 的 dNTP 混合物、0.2U 的 AmpliTaq Gold、4 种引物各 200nM 以及 50% DNA 混合物。将测定 / 样品混合物分配至 384 孔 PCR 板, 在 50°C 下开始反应持续 2min, 然后 95°C 进行 10min 以及进行 95°C、15s 和 57°C、1min 的 40 个循环。

[0272] 将 PCR 产物用虾碱性磷酸酶 (SAP) 处理以除去未并入的 dNTP。将该混合物在 37°C 下孵育 40min, 然后在 85°C 下孵育 5min。然后进行引物延伸反应。简而言之, 向 SAP 处理的 PCR 产物中添加 771nm 来自染色体 21 / 染色体 1 测定的延伸引物、1.54 μ M 来自染色体 21 / 染色体 18 测定的延伸引物、0.67U 热测序酶 (Thermosequenase) (Sequenom) 以及延伸混合物 (extension cocktail) 中的 ddCTP、ddGTP、dATP 和 dTTP 各 64 μ M。反应条件是 94°C、2min, 然后 94°C、5s, 50°C、5s 以及 72°C、5s 进行 80 个循环。向该延伸产物中添加 16 μ L 水和 3mg 纯净树脂 (Clean Resin) (Sequenom) 来进行最后的清理。将该混合物在旋转器中混合 20min 至 30min, 然后在 361g 下离心 5min。通过 MassARRAY 纳米分配器 (Nanodispenser) S (Sequenom) 将 15nL 至 25nL 终产物分配至 SpectroCHIP 中。在 MassARRAY Analyzer 紧凑型质谱仪 (MassARRAY Analyzer Compact Mass Spectrometer) (Sequenom) 进行来自 SpectroCHIP 的数据获取。将质谱数据输入 MassARRAY Typer (Sequenom) 软件进行分析。

[0273] 用双重 RCD 测定来分析 5 个整倍体和 5 个 T21 50% 胎盘 / 母体 DNA 样品。对于每个样品, 计数了来自单独测定的信息孔的数目, 该信息孔包括只对染色体 21 标记物或染色体 1 标记物或染色体 18 标记物是阳性的孔。对每个 RCD 测定, 单独计算了全部信息孔中染色体 21 标记物的比例 (P_i)。然后应用序贯概率比检验 (SPRT) 来确定该 P_i 是否表示整倍体样品或 T21 样品。通过进行这个步骤, 由于将每个板计数了两次, 所以减少了所需的孔的数目。

[0274] 通常首先应用染色体 21 / 染色体 1 测定。如果样品尚未分类, 则加入来自染色体 21 / 染色体 18 测定的值来进行进一步的计算。对尚未分类的样品使用额外的板直至能够

进行判定。如图 21 所示,利用单个 384 孔板正确地分类了全部整倍体 50%混合物样品。某些 T21 个例需要 2 个或更多个板来进行正确的分类。如果只使用一个测定,则需要更多数目的板来获得实现分类所要求的信息孔的数目。例如,当只使用任何一个 RCR 测定时,个例 N0230 的数据是不可分类的。然而,当合并来自两个测定的数据时,实现了正确的分类。如果不使用二重 RCD 测定,则需要额外板的分析。我们预期更高水平的多重测定将进一步减少孔的数目。

[0275] 在另一实施例中,我们开发了靶向染色体 21 上的 4 个不同扩增子以及和它们位于其它非染色体 21 的常染色体上的对应种内同源伴侣的 4 重 (4-plex) 测定。在数字 RCD 分析中使用这种 4 重测定,然后进行来自整倍体妊娠和 21 三体性妊娠的样品的 SPRT 分类。利用 QIAamp 组织试剂盒 (Qiagen, Hilden, Germany) 从胎盘样品提取 DNA。

[0276] 首先通过 NanoDrop 分光光度计 (NanoDrop Technology, Wilmington, DE) 来定量本研究中所用的全部胎盘和母体血沉棕黄色层 DNA 样品。利用 6.6pg/ 细胞的换算将 DNA 浓度转换成基因组当量 (GE)/ μ L。通过将 DNA 样品连续稀释来确定对应于约每孔一个模板的 DNA 的量。在这种条件下,我们预期约 37% 的孔将表现出阴性扩增。在多重数字 RCD 分析中,选择了 4 组种内同源序列靶标:通过正向引物 5' -ACGTTGGATGTTGATGAAGTCTCATCTCTACTTCG-3' 和反向引物 5' -ACGTTGGATGCAATAAGCTTGGCCAGAAATACT-3' 来共扩增染色体 21 和染色体 1 上的种内同源基因座,从而获得 81bp 的扩增子。通过正向引物 5' -ACGTTGGATGGAATTTAAGCTAAATCAGCCTGAACTG-3' 和反向引物 5' -ACGTTGGATGGTTTCTCATAGTTCATCGTAGGCTTAT-3' 来共扩增染色体 21 和染色体 7 上的种内同源基因座,从而获得 82bp 的扩增子。通过正向引物 5' -ACGTTGGATGTCAGGCAGGGTTCTATGCAG-3' 和反向引物 5' -ACGTTGGATGAGGCGCTTCCTGGCTCTTG-3' 来共扩增染色体 21 和染色体 2 上的种内同源基因座,从而获得 101bp 的扩增子。通过正向引物 5' -ACGTTGGATGGCTCGTCTCAGGCTCGTAGTT-3' 和反向引物 5' -ACGTTGGATGTTTCTTCGAGCCCTTCTTGG-3' 来共扩增染色体 21 和染色体 6 上的种内同源基因座,从而获得 102bp 的扩增子。每个反应含有 10X 缓冲液 II (Applied Biosystems)、 $MgCl_2$ 和 100 nM 的每种引物。总反应体积是 5 μ L/ 孔。在 95°C 下开始反应持续 5min, 然后 95°C、30sec, 62°C、30sec 和 72°C、30sec 进行 45 个循环,最后在 72°C 下最终延伸 7min。在 GeneAmp PCR 系统 9700 (Applied Biosystems) 上进行全部常规 PCR 扩增。通过虾碱性磷酸酶 (SAP) 处理来灭活未被并入的核苷酸。每个反应含有 10X SAP 缓冲液 (Sequenom) 和 SAP 酶 (Sequenom)。向每一 PCR 中添加 2 μ L 的 SAP 混合物。将该 SAP 反应在 37°C 下孵育 40min, 以及在 85°C 下孵育 5min。SAP 处理后,利用 iPLEX Gold 试剂盒 (Sequenom) 在 PCR 产物上进行引物延伸反应。通过延伸引物 5' -GTCTCATCTCTACTTCGTACCTC-3' 来询问染色体 21 和染色体 1 上的种内同源基因座上的种内同源序列错配 (PSM)。通过延伸引物 5' -TTTTACGCTGTCCCCATTT-3' 来询问染色体 21 和染色体 7 上的种内同源基因座上的 PSM。通过延伸引物 5' -GGTCTATGCAGGAGCCGAC-3' 来询问染色体 21 和染色体 2 上的种内同源基因座上的 PSM。通过延伸引物 5' -TGGGCGCGGGAGCGGACTTCGCTGG-3' 来询问染色体 21 和染色体 6 上的种内同源基因座上的 PSM。除了用于染色体 21 和染色体 6 上的 PSM 的延伸引物是 1.03 μ M 以外,每个反应含有 10X iPLEX 缓冲液 (Sequenom)、iPLEX 终止混合物 (Sequenom)、iPLEX 酶 (Sequenom) 和 343nM 的每种延伸引物。向 5 μ L 的 PCR 产物中添加 2 μ L 的 iPLEX 混合物。根据 200- 短- 循环程序来循环 iPLEX 反应。简而言之,将样品首先在 94°C 下变性

35sec,然后在 52℃下退火 5sec,并在 80℃下延伸 5sec。将退火和延伸循环再重复 4 次,总共 5 个循环,然后返回 94℃的变性步骤保持 5sec,然后再进行 5 个循环的退火和延伸循环。将 5 个退火和延伸循环与单次变性步骤重复 39 次,总共是 40 次。问在 72℃下进行最终延伸 3min。然后将每个 PCR 的 iPLEX 反应产物用 16 μ L 水稀释,并用 6mg 树脂脱盐。在分配至 SpectroCHIP (Sequenom) 和基质辅助激光解吸 / 电离飞行时间质谱法 MS 分析 (Sequenom) 前,将 384 孔板在 1600g 下离心 3min。

[0277] 独立地记录了 4 个测定的每一个的只对染色体 21 是阳性或只对参考染色体是阳性的孔的数目。对于每个测定,计算了染色体 21 和参考染色体的分子的泊松 (Poisson) 修正数目。计算来自全部 4 个测定的染色体 21 的泊松修正分子数目的总和及参考染色体的泊松修正分子数目的总和,并将这两个总和视为 4 重测定的信息计数。 P_r 值是 4 重测定的染色体 21 计数除以 4 重测定的染色体 21 计数与参考染色体计数的总和。将实验导出的 P_r 进行 SPRT 分析。分析一个或多个 384 孔板直至通过 SPRT 进行疾病分类是可能的。分析了总共两个 50% 整倍体胎盘基因组 DNA/50% 母体血沉棕黄色层 DNA 混合物和两个 50% 21 三体性胎盘基因组 DNA/50% 母体血沉棕黄色层 DNA 混合物。

[0278] 将实验导出的 P_r 值与 P_r 的预期值进行比较来检验无效假设或备选假设。可选择地,如果给定数目的信息计数的 P_r 仍然未能达到疾病分类的统计置信度,则既不接受无效假设也不接受备选假设。在有更多的可用数据前,将这些个例视为不可分类的。

[0279] 每个样品的结果和 SPRT 分类列于表 22A 和 22B 中。在实现 SPRT 分类前,两个整倍体样品需要 2 个和 5 个 384 孔多重数字 RCD 分析。没有来自 4 重测定的单独成员的数据允许通过 SPRT 进行疾病分类。只通过一个 384 孔多重数字 RCD 分析分别正确地分类了两个 21 三体性样品。类似地,没有来自 4 重测定的单独成员的数据允许通过 SPRT 进行疾病分类。然而,来自 4 重测定的综合计数允许正确的 SPRT 分类。这些数据表明,通过使用多重数字 RCD,与使用单重 (single-plex) 数字 RCD 测定相比,对于给定数目的数字 PCR 分析,信息计数的有效数目明显增加。

[0280] VI. 使用数字外遗传相对染色体剂量

[0281] 在此处,我们概述了称为数字外遗传相对染色体剂量 (digital epigenetic relative chromosome dosage) (数字 ERCD) 的方法,其中对在染色体非整倍性有关的染色体 (例如,21 三体性中的染色体 21) 和参考染色体上表现出胎儿特异性的 DNA 甲基化方式或其它外遗传变化的外遗传标记物进行数字 PCR 分析。在从怀有正常胎儿的孕妇提取的血浆 DNA 中,对染色体 21 外遗传标记物是阳性的孔的数目与对参考染色体外遗传标记物是阳性的孔的数目的比率将为我们提供参考范围。如果胎儿具有 21 三体性,则预期该比率将增加。对本领域的技术人员显而易见的是,在这个分析中能够使用多于一个的染色体 21 标记物和多于一个的参考染色体标记物。

[0282] 表现出胎儿 (胎盘) 特异性甲基化方式的染色体 21 上的基因的一个实例是羧化全酶合成酶 (HLCS) 基因。HLCS 在胎盘中是高度甲基化的,但是在母体血液细胞中是低甲基化的,并且被第 11/784499 号美国专利申请所包括,该申请在此处通过引用的方式被并入。表现出胎儿 (胎盘) 特异性甲基化方式的参考染色体上的基因的一个实例是染色体 3 上的 RASSF1A 基因 [10]。RASSF1A 在胎盘中是高度甲基化的,但是在母体血液细胞中是低甲基化的,参见第 11/784501 号美国专利申请,该申请在此处通过引用的方式被并入。

[0283] 在高度甲基化的 HLCS 和高度甲基化的 RASSF1A 在利用母体血浆来检测胎儿 21 三体性的数字 PCR 的应用中,首先收集母体外周血液。然后将该血液离心并收集血浆。然后利用本领域技术人员公知的技术来从该血浆中提取 DNA,例如使用 QIAamp 血液试剂盒 (Qiagen)。然后用一种或多种甲基化敏感的限制性内切酶来消化该血浆 DNA,例如 HpaII 和 BstUI。这些甲基化敏感的限制性内切酶将切割这些基因的母体的未甲基化的形式,而留下完整的胎儿高度甲基化的序列。然后将该消化的血浆 DNA 样品稀释至某种程度,即每个反应孔将检测到平均约 0.2 至 1 个分子的任何一种经限制性内切酶处理但完整的 HLCS 或 RASSF1A 序列。使用两种实时 PCR 系统来扩增稀释的 DNA,一种系统具有两种引物和一种对 HLCS 基因特异性的 TaqMan 探针,HLCS 基因包含如果序列是未甲基化的,则被限制性内切酶切割的区域;另一种系统针对 RASSF1A 基因,该系统类似地具有两种引物和一种 TaqMan 探针。对于后一种 RASSF1A 引物/探针组,Chan et al. 2006, Clin Chem 52, 2211-2218 已经描述了一个实例。针对 HLCS 和 RASSF1A 靶标的 TaqMan 探针具有不同的荧光报道分子,例如分别为 FAM 和 VIC。然后使用 384 孔板来进行数字 PCR 实验。计数只对 HLCS 是阳性的孔的数目和只对 RASSF1A 是阳性的孔的数目,并计算这些计数的比率。与怀有正常的整倍体胎儿的孕妇相比,预期取自怀有 21 三体性胎儿的孕妇的母体血浆的 HLCS : RASSF1A 比率将更高。过度表现的程度将取决于数字 PCR 运行中每孔的平均参考模板浓度。

[0284] 评分这些结果的其它方法将是可能的,例如,计数对 HLCS 是阳性的孔的数目,而不论同时存在的对 RASSF1A 的阳性;反之对 RASSF1A 也是这样,而不论同时存在的对 HLCS 的阳性。而且,代替计算比率,HLCS 和 RASSF1A 计数的总数或差异能够用于表示胎儿的 21 三体性状态。

[0285] 除了在板中进行数字 PCR 以外,对本领域技术人员显而易见的是,能够使用数字 PCR 的其它变体,例如,微流体芯片、纳升 PCR 微板系统、乳液 PCR、polony PCR 和滚动循环扩增、引物延伸以及质谱等。作为示例而不作为限制来描述这些数字 PCR 的变体。

[0286] 除了实时 PCR 以外,对本领域技术人员还显而易见的是,诸如质谱的方法能够用于评分数字 PCR 的结果。

[0287] 除了利用甲基化敏感性限制性内切酶来区分 HLCS 和 RASSF1A 的胎儿与母体形式以外,对本领域技术人员显而易见的是,确定甲基化状态的其它方法也是可用的,例如,亚硫酸氢盐修饰、甲基化特异性 PCR、利用甲基化的胞嘧啶的抗体的免疫沉淀、质谱等。

[0288] 对本领域技术人员还显而易见的是,本实施例和本专利申请中的其它实施例中所示的方法能够用于可以发现胎儿 DNA 的其它体液,包括母体尿、羊水、宫颈灌洗、绒毛膜绒毛、母体唾液等。

[0289] VII. 利用乳液 PCR 和其它策略的大规模平行基因组测序

[0290] 此处,我们将描述能够将核酸分子的数字读取用于检测母体血浆中的诸如 21 三体性的胎儿染色体非整倍性的另一实施例。胎儿染色体非整倍性由染色体或染色体区域的异常剂量导致。无创检测具有高灵敏度和特异性以将误诊降至最低是理想的。然而,在母体血浆和血清中,胎儿 DNA 以低的绝对浓度存在,并表现出全部 DNA 序列的一小部分。因此,靶向具体基因座的数字 PCR 取样的数目不能在相同的生物样品中无限地增加。因此,多组具体靶基因座的分析可以用于增加从样品中获得的数据的量,而不增加进行的数字 PCR 取样的数目。

[0291] 因此,实施方案允许通过将遗传信息的量最大化来进行胎儿染色体非整倍性的无创检测,所述遗传信息能够从含有母体背景核酸的生物样品中以小群体存在的有限量的胎儿核酸推断出。在一方面,获得的遗传信息的量足以进行准确的诊断,但是不过度地过量,从而控制所需的输入生物样品的成本和量。

[0292] 诸如可在 454 平台 (Roche) (Margulies, M. et al. 2005 Nature 437, 376-380)、Illumina 基因组分析仪 (或 Solexa 平台) 或 SOLiD 系统 (Applied Biosystems) 或 Helicos True 单分子 DNA 测序技术 (Harris TD et al. 2008 Science, 320, 106-109)、Pacific Biosciences 的单分子实时 (SMRT™) 技术以及纳米孔测序 (Soni GV and Meller A. 2007 Clin Chem 53:1996-2001) 上实现的大规模平行测序允许在平行模式下,以高阶的倍增来测序分离自样品的许多核酸分子 (Dear Brief Funct Genomic Proteomic 2003;1:397-416)。这些平台的每个测序了克隆扩增的或者甚至未扩增的单分子核酸片段。

[0293] 由于在每个运行中,从每个样品产生了十万至百万或者甚至可能亿至十亿级的大量测序读数,得到的测序的读数形成了原始样品中核酸种类的混合物的代表性谱。例如,测序读数的单元型、转录物组、甲基化谱与原始样品的那些 (Brenner et al. Nat Biotech 2000;18:630-634; Taylor et al. Cancer Res 2007;67:8511-8518) 类似。由于每个样品的序列的大量取样,相同序列的数目,诸如以几倍的覆盖率或高冗余度从核酸池的测序产生的那些,也是原始样品中特定核酸种类或基因座的计数的好的定量表现方式。

[0294] 在一实施方案中,对存在于孕妇血浆中的 DNA 片段进行随机测序,并获得最初来自胎儿或母亲的基因组序列。随机测序包括存在于生物样品中的核酸分子的随机部分的取样 (测序)。由于测序是随机的,所以可以在每个分析中测序核酸分子的不同子集 (部分) (并且因此基因组)。甚至当这个子集随样品变化和随分析变化时,这甚至可以在使用相同的样品时发生,实施方案也起作用。该部分的实例是基因组的约 0.1%、0.5% 或 1%。在其它实施方案中,所述部分是这些值的至少任意一个。

[0295] 然后可以用生物信息学方法将这些 DNA 序列的每一个定位于人基因组上。可能地,将从随后的分析中排除这些序列的比例,因为它们存在于人基因组的重复区域中,或存在于发生了个体间变异 (inter-individual variation) 的区域中,例如拷贝数变异。因此,可以测定感兴趣的染色体的量和一个或多个其它染色体的量。

[0296] 在一实施方案中,则能够由生物信息学方法的结果来计算潜在地涉及诸如染色体 21 或染色体 18 或染色体 13 的染色体非整倍性的染色体参数 (例如,部分表现 (fractional representation))。可以基于全部序列 (例如,全部染色体的某些测量) 或染色体的特定子集 (例如,除了被检测的染色体之外的仅一个其它染色体) 的量来获得部分表现。

[0297] 然后将该部分表现与在涉及正常 (即,整倍体) 胎儿的妊娠中建立的参考范围进行比较。可能地,在所述方法某些变体中,根据特定的母体血浆样本中的胎儿 DNA 的部分浓度 (f) 来调整该参考范围。能够从测序数据集来确定 f 的值,例如,如果胎儿是男性,则使用可被定位到 Y 染色体的序列。还可以在单独的分析中确定 f 的值,例如,使用胎儿外遗传标记物 (Chan KCA et al. 2006 Clin Chem 52, 2211-8) 的分析或通过单核苷酸多态性的分析。

[0298] 在一方面,甚至当样品中的核酸池以小于 100% 的基因组覆盖率进行测序,并且捕获的核酸分子的比例中,大部分的每个核酸种类只测序了一次时,也能够定量地确定特定

基因座或染色体的剂量失衡。换言之,从所述基因座在样品的全部可被定位的 (mappable) 测序的标签中的表现百分比来推断该基因座或染色体的剂量失衡。

[0299] 在大规模平行基因组测序方法的一方面,可以同时生成全部染色体的代表性数据。不提前选择特定片段的来源。随机地进行测序,然后进行数据库检索来发现特定片段的出处。这与扩增来自染色体 21 的一个具体片段和来自染色体 1 的另一具体片段的情况形成对比。

[0300] 在一实施例中,这种序列的比例将来自与非整倍性有关的染色体,例如示例性实施例中的染色体 21。这种测序应用的其它的序列将源自其它染色体。考虑到染色体 21 与其它染色体的相对大小,能够获得来自这种测序应用的染色体 21 特异性序列的参考范围内的标准化频率 (normalized frequency)。如果胎儿具有 21 三体性,则来自这种测序应用的源自染色体 21 的序列的标准化频率将增加,从而能够检测出 21 三体性。标准化化频率的变化程度将取决于分析的样品中胎儿核酸的 部分浓度。

[0301] 在一实施方案中,我们使用了用于人基因组 DNA 样品和人血浆 DNA 样品的单末端测序的 Illumina 基因组分析仪。该 Illumina 基因组分析仪测序了被捕捉在称为流动池 (flow cell) 的固体表面的克隆扩增的单个 DNA 分子。每个流动池具有用于测序 8 个单独样品或样品池的 8 条道。每条道能够生成约 200Mb 的序列,这只是人基因组序列的 30 亿个碱基对的一部分。利用流动池的一条道来测序每个基因组 DNA 样品或血浆 DNA 样品。将生成的短序列标签与人参考基因组进行比对,并且记录染色体来源。将与每条染色体比对的单独测序的标签的总数列成表,并将其与预期来自参考人基因组或非疾病代表性样品的每条染色体的相对大小进行比较。然后鉴定了染色体增加或减少。

[0302] 所述方法只是目前描述的基因 / 染色体剂量策略的一个示例。可选择地,能够进行配对末端测序。不是如 Campbell 等人所述 (Nat Genet 2008 ;40 :722-729) 将测序的片段长度与参考基因组中预期的片段长度进行比较,而是根据染色体位置计数和分类比对的测序的标签的数目。通过将标签计数与参考基因组中的预期染色体大小或与非疾病代表性样品中的预期染色体大小进行比较来确定染色体区域或完整染色体的增加或减少。

[0303] 在另一实施方案中,将运行中测序的核酸池部分进行另外的子选择 (subselect), 然后进行测序。例如,能够将诸如寡核苷酸阵列的基于杂交的技术用于首先从某些染色体对核酸序列进行子选择,例如,潜在的非整倍体染色体和与检测的非整倍性无关的其它染色体。另一实例是,在测序前,将来自样品池的核酸序列的某些亚群进行子选择或富集。例如,据报道母体血浆中的胎儿 DNA 分子由比母体背景 DNA 分子短的片段组成 (Chan et al. Clin Chem 2004 ;50 :88-92)。因此,根据分子大小,可以使用对本领域技术人员公知的一种或多种方法来将样品中的核酸序列分级,例如,通过凝胶电泳或尺寸排除柱,或者通过基于微流体的方法。仍然,可选择地,在分析母体血浆中无细胞胎儿 DNA 的实例中,能够通过抑制母体背景的方法来富集胎儿核酸部分,例如,通过添加甲醛 (Dhallan et al. JAMA 2004 ;291 :1114-9)。

[0304] 能够在这种应用中类似地使用其它单分子测序策略,例如, Roche454 平台、Applied Biosystems SOLiD 平台、Helicos True 单分子 DNA 测序技术、Pacific Biosciences 的单分子实时 (SMRT™) 技术以及纳米孔测序。

[0305] 可以在同时提交的申请“DIAGNOSING FETAL CHROMOSOMAL ANEUPLOIDY USING

GENOMIC SEQUENCING(利用基因组测序来诊断胎儿染色体非整倍性)”中找到结果的实例和进一步的讨论(例如,对于测序和计算参数)(代理公司案卷号 016285-005220US),该申请通过引用的方式被并入。注意,当反应是例如本部分所述的测序时,可以实施用于确定截止值的本文所述的方法。

[0306] 还能够与测序运行分开地进行母体血浆中胎儿 DNA 的部分浓度的确定。例如,能够利用实时 PCR、微流体 PCR 或质谱来预确定 Y 染色体 DNA 浓度。实际上,能够利用 Y 染色体以外的基因座来测定胎儿 DNA 浓度,并且适用于女性胎儿。例如,Chan 等人表明源自胎儿的甲基化的 RASSF1A 序列会在源自母体的未甲基化的 RASSF1A 序列的背景下的孕妇血浆中检测出(Chan et al.Clin Chem2006 ;52 :2211-8)。因此,能够通过用甲基化的 RASSF1A 序列的量除以总 RASSF1A(甲基化的和未甲基化的)序列的量来确定部分胎儿 DNA 浓度。

[0307] 对于实施我们的发明,预期母体血浆比母体血清优选,因为在血液凝固中,从母体血液细胞中释放了 DNA。因此,如果使用血清,预期母体血浆中的胎儿 DNA 的部分浓度将低于母体血清中的浓度。换言之,如果使用母体血清,与同时从相同孕妇获得的血浆样品相比,预期将需要生成更多的序列来诊断胎儿染色体非整倍性。

[0308] 测定胎儿 DNA 的部分浓度的仍然另一替代方法是通过孕妇与胎儿之间的多态性差异的定量(Dhallan R,et al.2007 Lancet,369,474-481)。这种方法的实例是靶向孕妇是纯合而胎儿是杂合的多态性位点。能够将胎儿特异性等位基因的量与共同等位基因的量进行比较以确定胎儿 DNA 的部分浓度。

[0309] 与检测染色体畸变的现有技术相比,大规模平行测序不依赖于预定或预定义的组的 DNA 序列的检测或分析,所述现有技术包括比较基因组杂交、微阵列比较基因组杂交、检测并定量一种或多种特异性序列的定量实时聚合酶链式反应,。测序了来自样品池的 DNA 分子的随机代表性部分。在含有或不含有肿瘤 DNA 的样品之间比较与各种染色体区域比对的不同序列标签的数目。通过与样品中任何给定的染色体区域比对的序列的数目(或百分比)的差异来揭示染色体畸变。

[0310] 在另一实施例中,可以使用血浆无细胞 DNA 的测序技术来检测血浆 DNA 中的染色体畸变,以检测具体的癌症。不同的癌症具有一组典型的染色体畸变。可以使用多个染色体区域的变化(扩增和缺失)。因此,将有增加比例的与扩增区域比对的序列,和降低比例的与减少的区域比对的序列。能够将每条染色体的百分比表现与表示为相对于全基因组的任何给定的染色体的基因组表现的百分比的参考基因组中每条对应染色体的大小进行比较。还可以使用与参考染色体的直接比较或比较。

[0311] VIII. 突变检测

[0312] 母体血浆中的胎儿 DNA 以较小的群体存在,并且母体血浆 DNA 由胎儿贡献的平均为 3%至 6%。由于这一原因,本领域中以前的大部分工作集中于检测胎儿从父亲遗传的 DNA 靶标,其可以与母体血浆中的占多数的母体 DNA 背景区分开。这种以前检测的靶标的实例包括 Y 染色体上的 SRY 基因(Lo YMD et al.1998 Am J Hum Genet,62,768-775)和当母亲是 RhD 阴性时的 RHD 基因(Lo YMD et al.1998 N Engl J Med,339,1734-1738)。

[0313] 对于胎儿突变检测,利用母体血浆的以前的策略局限于父亲是携带者的常染色体显性状态,当父亲和母亲携带不同的突变时,通过直接突变检测来排除常染色体隐性疾病,或者通过连锁分析来排除常染色体隐性疾病(Ding C.et al.2004 Proc Natl Acad Sci

USA 101,10762-10767)。这些以前的策略具有明显的局限性。例如,对于雄性和雌性配偶都携带相同突变的情况,则不可能通过母体血浆中的直接突变检测来进行有意义的产前诊断。

[0314] 这种情境示于图 23 中。在该情境中,有 3 个可能的胎儿基因型:NN、NM 和 MM,其中 N 表示正常等位基因,M 表示突变体等位基因。突变体等位基因的实例包括那些导致囊性纤维化、 β -地中海贫血症、 α -地中海贫血症、镰形细胞贫血、脊髓型肌萎缩、先天性肾上腺增生等的突变体等位基因。能够在线人类孟德尔遗传(OMIM) www.ncbi.nlm.nih.gov/sites/entrez?dh=OMIM&itool=toolbar 找到这些病症的其它实例。在母体血浆中,大部分 DNA 来自母亲,并且是 NM 的。对于 3 种胎儿基因型的任一种,将不会有允许在母体血浆中被独特地检测出的任何独特的胎儿等位基因。因此,不能在此处应用常规策略。

[0315] 本文所述的实施方案允许处理这种情境。在母亲和胎儿都是 NM 的情境下,N 等位基因和 M 等位基因将是等位基因平衡的。然而,如果母亲是 NM 而胎儿是 NN,则将在母体血浆中有等位基因失衡,并且 N 等位基因是过度表现的。在另一方面,如果母亲是 NM 而胎儿是 MM,则将在母体血浆中有等位基因失衡,并且 M 等位基因是过度表现的。因此,对于胎儿突变检测,无效假设指当胎儿的基因型是 NM 时,没有等位基因失衡。备选假设指存在等位基因失衡,并且取决于 N 等位基因或 M 等位基因是否是过度表现的,胎儿的基因型能够是 NN 或 MM。

[0316] 能够通过利用本文所述的实施方案的数字 PCR 来确定存在或不存在等位基因失衡。在第一种情境下,特定体积的母体血浆含有从 100 个细胞释放的 DNA,其中 50 个细胞来自母亲,50 个细胞来自胎儿。因此,在这个体积的血浆中的胎儿 DNA 的部分浓度是 50%。当母亲的基因型是 NM 时,则母亲将贡献 50 个 N 等位基因和 50 个 M 等位基因。如果胎儿的基因型是 NM,则胎儿将贡献 50 个 N 等位基因和 50 个 M 等位基因。因此,N 等位基因和 M 等位基因之间没有等位基因失衡,并且每个等位基因总共各 100 个拷贝。在另一方面,如果胎儿的基因型是 NN,则在这个体积的血浆中将有 100 个源自胎儿的 N 等位基因。因此,总共有 150 个 N 等位基因和 50 个 M 等位基因。换言之,在 N 与 M 之间存在等位基因失衡,并且 N 是过度表现的,N 与 M 的比率为 3 : 1。

[0317] 在相反的情况下,如果胎儿的基因型是 MM,则在这个体积的血浆中将有 100 个源自胎儿的 M 等位基因。因此,将有 150 个 M 等位基因和 50 个 N 等位基因。换言之,在 N 等位基因与 M 等位基因之间存在等位基因失衡,并且 M 是过度表现的,M 与 N 的比率是 3 : 1。能够通过数字 PCR 来测量这种等位基因失衡。将具有较少数目的阳性孔的等位基因视为参考模板。与数字 RNA-SNP 和数字 RCD 分析类似,数字 PCR 实验中等位基因的实际分布将受泊松概率密度函数控制。因此,由于本情境中等位基因失衡的理论程度是 3 : 1,所以等位基因失衡的预期程度将取决于数字 PCR 分析中每孔的平均模板浓度。因此,需要将对每孔的平均参考模板浓度 (m_r) 适合的诸如用于 SPRT 分析的截止值用于对个例进行分类。

[0318] 而且,需要测量的等位基因失衡的程度取决于部分胎儿 DNA 浓度。与上文的实施例对比,让我们考虑含有从 100 个细胞中释放的 DNA 的特定体积的母体血浆,其中 90 个细胞来自母亲而 10 个细胞来自胎儿。因此,这个体积的血浆中的胎儿 DNA 的部分浓度是 10%。当母亲的基因型是 NM 时,则母亲将贡献 90 个 N 等位基因和 90 个 M 等位基因。如果胎儿的基因型是 NM,则胎儿将贡献 10 个 N 等位基因和 10 个 M 等位基因。因此,在 N 等位基因与 M

等位基因之间不存在等位基因失衡,并且每个等位基因总共各 100 个拷贝。在另一方面,如果胎儿的基因型是 NN,则在这个体积的血浆中将有 20 个源自胎儿的 N 等位基因。因此,总共有 110 个 N 等位基因和 90 个 M 等位基因。

[0319] 换言之,N 等位基因与 M 等位基因之间存在等位基因失衡,并且 N 等位基因是过度表现的。在相反的情况下,如果胎儿的基因是 MM,则在这个体积的血浆中将有 20 个源自胎儿的 M 等位基因。因此,将有 110 个 M 等位基因和 90 个 N 等位基因。换言之,N 等位基因与 M 等位基因之间存在等位基因失衡,并且 M 是过度表现的。当胎儿 DNA 部分浓度是 10% 时,等位基因失衡的理论程度是 110 : 90,这不同于上文的实施例所示的,有 50% 的胎儿 DNA 时的比率 3 : 1。因此,需要将 对胎儿 DNA 部分浓度适合的解释截止值用于对个例进行分类,例如,用于 SPRT 分析的解释截止值。

[0320] 因此,提取血浆 DNA。定量血浆样品中母体 DNA 和胎儿 DNA 的量,例如,通过以前建立的实时 PCR 测定 (Lo, et al. 1998 Am J Hum Genet 62, 768-775) 或对本领域技术人员公知的其它类型的量标 (quantifier),例如 SNP 标记物 (Dhallan R et al. 2007 Lancet, 369, 474-481) 和胎儿外遗传标记物 (Chan KCA et al. 2006 Clin Chem, 52, 2211-2218)。将计算胎儿 DNA 百分比。然后制备定量的血浆 DNA 样品 (例如,稀释的或浓缩的),从而在数字 PCR 分析中,每个反应孔平均含有 1 个模板分子 (能够是 N 等位基因或 M 等位基因)。利用一对引物加两个 TaqMan 探针来进行数字 PCR 分析,其中一个 TaqMan 探针对 N 等位基因是特异性的,而另一个 TaqMan 探针对 M 等位基因是特异性的。计数只对 M 等位基因是阳性的孔的数目和只对 N 等位基因是阳性的孔的数目。将这些孔的比率用于确定是否有等位基因失衡的证据。能够通过对本领域技术人员公知的方法来寻找等位基因失衡的统计证据,例如利用 SPRT。在该分析的一个变体中,还可能计数只对 M 等位基因是阳性的孔的数目或对 M 等位基因和 N 等位基因是阳性的孔的数目;并计数只对 N 等位基因是阳性的孔的数目或对 M 等位基因和 N 等位基因是阳性的孔的数目,并导出这些计数的比率。再次地,能够通过对本领域技术人员公知的方法来寻找等位基因失衡的统计证据,例如利用 SPRT。

[0321] 利用雌性 / 雄性 (XX/XY) DNA 混合物来验证称为数字相对突变体剂量 (RMD) 的胎儿基因突变的剂量测定。如图 24A 所示,将来自雄性和雌性的血液细胞 DNA 分别与雄性 DNA 混合,从而分别在部分浓度为 25% 和 50% 时获得 XY 的背景下的 XX 基因型或 XY 基因型的样品,部分浓度分别为 25% 和 50% 的。

[0322] 另外,还从 12 个雄性个体和 12 个雌性个体获得血液细胞样品。将雌性血液细胞 DNA (基因型 XX) 分别与 3 倍的过量雄性血液细胞 DNA (基因型 XY) 混合,从而获得 75% XY 基因型的 DNA 背景下的 25% XX 基因型的 DNA 的 12 个 DNA 混合物,并且结果示于图 24B 中。

[0323] SPRT 的目的在于确定背景 DNA 中存在的少数基因型。在 75% 的 XY DNA 的背景下的 25% 的 XX DNA 的 DNA 混合物中,少数等位基因是源自 75% 的 DNA 的 Y。由于该样品中 25% 的 DNA 的基因型是 XX,因此如果在该样品中共有 200 个分子的 DNA,则 150 个分子将源自 XY 个体。因此,预期 Y 等位基因的数目是 75。雄性部分 DNA (基因型 XY) 贡献的 X 等位基因的数目也是 75。雌性 (基因型 XX) 贡献的 X 等位基因的数目是 50 (2 乘以 25)。因此, X 比 Y 的比率是 $125/75 = (1+25\%)/(1-25\%) = 5/3$ 。

[0324] 对于这个研究的第二部分,从在 β -球蛋白基因,即血红蛋白, β (HBB) 基因上携带 HbE (G $\rightarrow$ A) 和 CD41/42 (CTTT/-) 突变的雄性和雌性个体获得血液细胞样品。为了模拟从

怀有具有全部可能的基因型 (MM、MN 或 NN) 的雄性胎儿的杂合母亲 (MN, 其中 M = 突变体并且 N = 野生型) 获得的母体血浆样品, 将来自对野生型等位基因是纯合 (NN), 或对两个突变中的一个为杂合 (MN) 的雄性的血液细胞 DNA 分别与从对相同的突变是杂合 (MN) 的雌性收集的血液细胞 DNA 样品混合。从而获得各种部分雄性 / 突变体 DNA 浓度的 DNA 混合物。还将来自对 CD41/42 缺失是纯合 (MM) 的雌性的血液细胞 DNA 样品用于制备 DNA 混合物。为了保证用于 SPRT 分类的准确雄性比例, 利用 ZFY/X 测定来确定每种 DNA 混合物的部分雄性 DNA 浓度。

[0325] 将数字 ZFY/X 测定用于验证 SPRT, 以及用于确定 DNA 混合物中部分雄性 DNA 浓度。通过数字 PCR 分析来测定染色体 X (ZFX) 和染色体 Y (ZFY) 上的锌指蛋白序列的剂量。首先通过正向引物 5'-CAAGTGCTGGACTCAGATGTAAGT-3' 和反向引物 5'-TGAAGTAATGTCAGAAGCTAAACATCA-3' 来共扩增 ZFX 基因座和 ZFY 基因座的 87-bp 的扩增子。设计了两个染色体特异性 TaqMan 探针来区分染色体 X 和染色体 Y 种内同源基因, 且探针的序列分别是 5'-(VIC) TCTTTAGCACATTGCA (MGBNFQ)-3' 和 5'-(FAM) TCTTTACCACACTGCAC (MGBNFQ)-3'。

[0326] 通过正常等位基因相对于突变体等位基因的数字 PCR 分析来确定 DNA 混合物中的突变体剂量。对于 HbE 突变, 首先通过正向引物 5'-GGGCAAGGTGAACGTGGAT-3' 和反向引物 5'-CTATTGGTCTCCTTAAACCTGTCTTGTA-3' 来共扩增正常等位基因和突变体等位基因的 87-bp 的扩增子。设计了两个等位基因特异性 TaqMan 探针来区分正常 (G) 等位基因和突变体 (A) 等位基因, 且探针的序列分别是 5'-(VIC) TTGGTGGTGAGGCC (MGBNFQ)-3' 和 5'-(FAM) TTGGTGGTAAGGCC (MGBNFQ)-3'。HbE 突变的结果示于图 25 中。

[0327] 对于 CD41/42 缺失突变, 首先通过正向引物 5'-TTTTCCCACCTTAGGCTGC-3' 和反向引物 5'-ACAGCATCAGGAGTGGACAGATC-3' 来分别共扩增正常等位基因和突变体等位基因的 87-bp 和 83-bp 的扩增子。设计了两个等位基因特异性 TaqMan 探针来区分正常 (无缺失) 等位基因和突变体 (有缺失) 等位基因, 且探针的序列分别是 5'-(VIC) CAGAGTTCTTTGAGTCCT (MGBNFQ)-3' 和 5'-(FAM) AGAGGTTGAGTCCTT (MGBNFQ)-3'。HbE 突变的结果示于图 26A 和 26B 中。

[0328] 利用 12.765 数字阵列 (Fluidigm), 在 BioMark™ 系统 (Fluidigm) 上进行这些实验。利用 2X TaqMan Universal PCR Master Mix (Applied Biosystems) 进行一个板的反应, 并且反应体积是 10 μ L。对于 CD41/42 和 ZFY/X 测定, 每个反应含有 1X TaqMan Universal PCR Master Mix、900nM 的每种引物、125nM 的每种探针和 3.5 μ L 1ng/ μ L 的 DNA 混合物。对于 HbE 测定, 添加分别靶向正常 (G) 等位基因和突变体 (A) 等位基因的 250nM 和 125nM 的探针。通过 NanoFlex™ IFC 控制器 (Fluidigm) 将样品 / 测定混合物加载至数字阵列。在用于信号检测的 BioMark™ 系统上进行反应。在 50°C 下开始反应持续 2min, 然后 95°C、10min, 以及 95°C、15s 和 57°C (对于 ZFY/X 和 CD41/42) 或 56°C (对 HbE)、1min 的进行 50 个循环。每一个例至少使用一个反应板, 对尚未可分类的样品从额外的板汇集数据直至能够做出判定。

[0329] 对本领域技术人员还显而易见的是, 能够利用对本领域技术人员公知的方法来进行数字 PCR, 例如, 微流体芯片、纳升 PCR 微板系统、乳液 PCR、polony PCR 和滚动循环扩增、引物延伸以及质谱。

[0330] IX. 癌症的实施例

[0331] 在一实施方案中,可以进行本发明来将样品分类为具有或不具有在肿瘤中可能发生的等位基因比率偏移。在一方面,对于每一个例,通过数字 PCR 来确定只对 A 等位基因、只对 G 等位基因和对两个等位基因都是阳性信号的孔的数目。参考等位基因定义为具有较少数目阳性孔的等位基因。(在不大可能的情况下,即两个等位基因都具有相同数目的阳性孔,则任何一个等位基因都能够用作参考等位基因)。根据泊松概率密度函数,利用对参考等位基因是阴性的孔的数目来计算每孔的参考等位基因的推断的平均浓度 (m_r),而不论其它等位基因是否是阳性的。我们使用假设的实例来对计算进行说明。

[0332] 在 96 孔反应中,20 个孔对 A 等位基因是阳性的,24 个孔对 G 等位基因是阳性的,并且 28 个孔对两个等位基因都是阳性的。将 A 等位基因视为参考等位基因,因为对这个等位基因是阳性的孔较少。对参考等位基因是阴性的孔的数目是 $96-20-28=48$ 。因此,能够利用泊松分布来计算 m_r ,并且 m_r 是 $-\ln(48/96)=0.693$ 。

[0333] 在 LOH 检测的情况下,无效假设指假定没有由存在一个等位基因缺失导致的等位基因比率偏移的样品。在这个假定下,两个等位基因的阳性孔数目的预期比率将是 1 : 1,因此,含有潜在地过度表现的等位基因的信息孔(只对一个等位基因是阳性的孔)的预期比例将是 0.5。

[0334] 在 LOH 检测的情况下,备选假设指假定具有由样品的 50% 的细胞中存在一个等位基因缺失而导致的等位基因比率偏移的样品。由于过度表现的等位基因与参考等位基因之间的等位基因比率是 2 : 1,所以每孔过度表现的等位基因的平均浓度将是参考等位基因的平均浓度的 2 倍。然而,对该过度表现的等位基因是阳性的孔的数目并不简单地是对该参考等位基因是阳性的孔的数目的 2 倍,而是服从泊松分布。

[0335] 信息孔被定为对 A 或 G 等位基因是阳性的,但对 A 和 G 不都是阳性的孔。对于具有等位基因比率偏移的样品,含有过度表现的等位基因的孔的数目的预期比例的计算与表 600 所示的相同。在上文的实施例中,如果在 50% 的肿瘤细胞中存在 LOH,则每孔的 G 等位基因的平均浓度将是 $2 \times 0.693 = 1.386$ 。如果在多于 50% 的肿瘤细胞中存在 LOH,则每孔的 G 等位基因的平均浓度将根据下述公式来计算: $1/[1-(\text{LOH 的比例})] \times m_r$ 。

[0336] 对 G 等位基因是阳性的孔的预期比例将是 $1-e^{-1.386}=0.75$ (即 75% 或 72 个孔)。假定对于 A 等位基因或 G 等位基因,孔的阳性是独立的,则 $0.5 \times 0.75 = 0.375$ 的孔将对 A 等位基因和 G 等位基因都是阳性的。因此, $0.5-0.375 = 0.125$ 的孔只对 A 等位基因是阳性的,而 $0.75-0.375 = 0.375$ 的孔只对 G 等位基因是阳性的。因此,信息孔的比例是 $0.125+0.375 = 0.5$ 。携带 G 等位基因的信息孔的预期比例是 $0.375/0.5 = 0.75$ 。然后将 P_r 的这个预期值用于构建合适的 SPRT 曲线来确定样品中是否存在等位基因比率偏移(即,这种情况下的 LOH)。

[0337] 然后将通过数字 PCR 分析实验确定的携带非参考等位基因的信息孔的实际比例用于确定是否接受无效假设或备选假设,或者确定是否需要更多孔的进一步分析。基于阈值似然比率 8 来计算接受无效假设或备选假设的 P_r 的判定边界,因为这个值已经被证实在癌症检测的情况下提供了鉴别具有或不具有等位基因失衡的样品的令人满意的性能(Zhou, W, et al. (2001) Nat Biotechnol 19, 78-81; Zhou et al. 2002, 见上文)。在上文的实施例中,信息孔的数目是 $20+24 = 44$,而实验获得的 P_r 是 $24/44 = 0.5455$ 。接受无效假设的判定边界是 ≤ 0.5879 而接受备选假设的判定边界是 ≥ 0.6739 。因此,将这个实施例中

的样品分类为不具有等位基因比率偏移。

[0338] 综上所述,我们描述了在样品中检测序列失衡的方法。在一实施方案中,本发明能够通过分析母体血浆中的胎儿核酸用于诸如 21 三体性的胎儿染色体非整倍性的无创检测。这种方法还能够应用于其它含有胎儿核酸的生物材料,包括羊水、绒毛膜绒毛样品、母体尿、宫颈样品、母体唾液等。首先,我们证明了本发明在怀有 21 三体性胎儿的妇女的母体血浆中,用于确定 PLAC4 mRNA 即染色体 21 上的胎盘表达的转录物上的 SNP 的等位基因失衡的用途。其次,我们证明,通过相对染色体剂量 (RCD) 分析,我们的发明能够用作用于 21 三体性的无创产前检测的基于非多态性的方法。这种基于数字 RCD 的方法包括直接评价含有胎儿 DNA 样品中的染色体 21 的总拷贝数相对于参考染色体是否是过度表达的。甚至不需要复杂的仪器,数字 RCD 允许在含有 25% 胎儿 DNA 的样品中检测 21 三体性。我们应用序贯概率比检验 (SPRT) 来解释数字 PCR 数据。计算机模拟分析证实了疾病分类算法的高准确性。

[0339] 我们还描述了能够用于确定染色体非整倍性以外的其它形式的核酸序列失衡的方法,例如,用于胎儿突变检测的方法,或用于母体血浆中多态性检测的方法,以及通过血浆中源自肿瘤的核酸的分析来检测恶性肿瘤细胞基因组中的区域增加或减少的方法。

[0340] 本申请中所述的任何软件组分或函数可以作为通过使用任何合适的计算机语言的处理器运行的软件代码来执行,所述计算机语言诸如例如 Java、C++ 或使用例如常规或面向对象技术的 Perl。所述软件代码可以保存为用于储存和 / 或传送的计算机可读的介质上的一系列指令或命令,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、诸如光盘 (CD) 或 DVD (数字多用光盘) 的光学介质、闪存存储器等。计算机可读的介质可以是这些储存或传送装置的任何组合。

[0341] 还可以利用适合通过有线、光和 / 无线网络传送的载波信号来编码和传送这些程序,所述网络符合包括因特网在内的各种协议。这样,可以利用这些程序编码的数据信号来产生本发明实施方案的计算机可读的介质。用所述程序代码编码的计算机可读的介质可以用兼容的装置来包装或与其它装置分开提供 (例如,通过因特网下载)。任何这种计算机可读的介质可以位于单个计算机程序产品 (例如,硬盘或整个计算机系统) 上或在产品内,并且可以存在于系统或网络内的不同计算机程序产品上或在产品内。计算机系统可以包括监视器、打印机或用于向使用者提供本文所述的任何结果的其他合适的显示器。

[0342] 计算机系统的实例示于图 27 中。图 27 所示的子系统通过系统总线 2775 互连接。显示了另外的子系统,例如打印机 2774、键盘 2778、固定磁盘 2779、与显示适配器 2782 连接的监视器 2776 等。与 I/O 控制器 2771 连接的外设和输入 / 输出 (I/O) 装置能够通过本领域已知的多种方式来与该计算机系统连接,诸如串行端口 2777。例如,能够使用串行端口 2777 或外部界面 2781 来将所述计算机设备与诸如因特网的广域网络、鼠标输入装置或扫描仪相连接。通过系统总线的相互连接允许中央处理器 2773 与每个子系统通讯并控制来自系统内存 2772 或固定磁盘 2779 的指令的执行,以及子系统之间的信息交换。系统内存 2772 和 / 或固定磁盘 2779 可以包括计算机可读的介质。

[0343] 出于示例和说明的目的,上文描述了本发明的示例性的实施方案。并不意图穷尽本发明或将本发明局限于所述的精确形式,并且根据上文的教导,许多修饰和变化是可能的。选择并描述实施方案以最好地解释本发明的原理和及其实际应用,从而允许本领域其

他技术人员在各种实施方案中最好地利用本发明,并且可以作出适用于所包括的特定用途的各种修饰。

[0344] 本文引用的全部出版物、专利和专利申请的全部内容通过引用的方式并入本文用于所有目的。

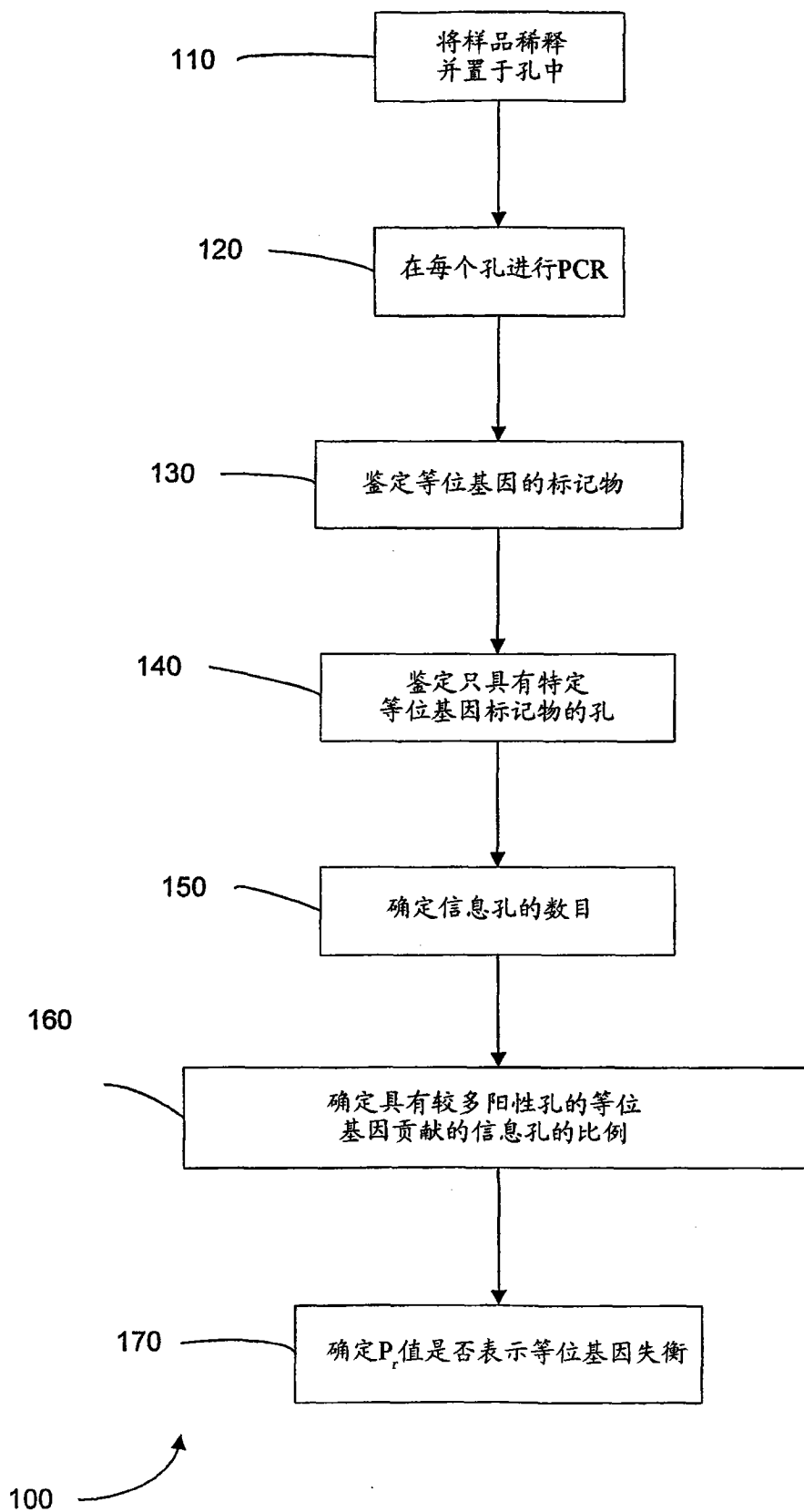


图 1

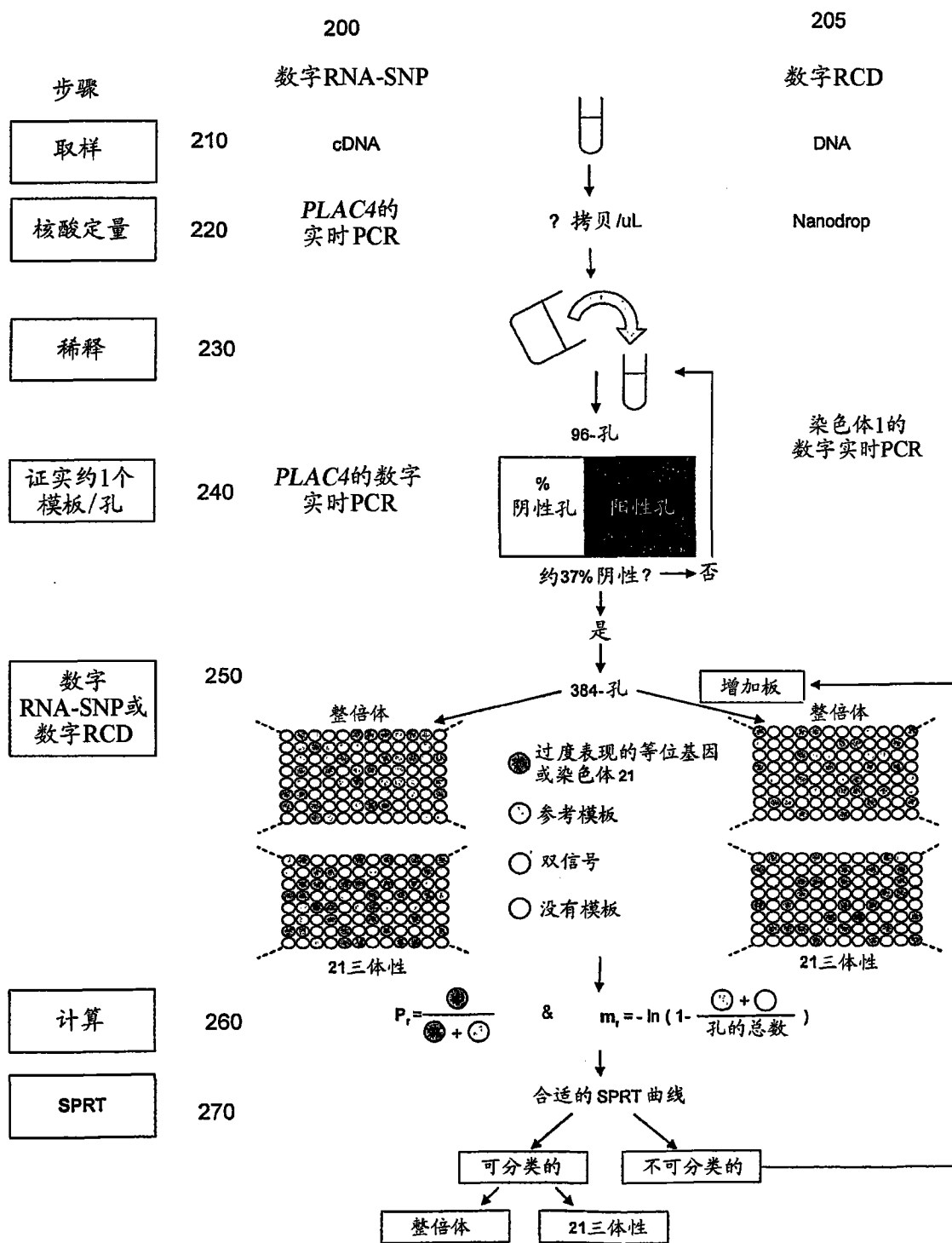


图 2A

癌症	常见染色体畸变	
	增加	减少
膀胱癌	7	9p; 11q; 17p
乳腺癌	1q; 8q; 17q; 20q	8p; 13q; 16q; 18q
结肠直肠癌	7; 8q; 13q; 20q	1p; 5q; 8p; 15q; 17p; 18q
肝细胞癌	1q; 8q; 20q	4q; 8p; 13q; 16q; 17p
非小细胞肺癌	1q; 3q; 5p; 8q	3p; 8p; 13q; 17p
小细胞肺癌	3q; 5p; 8q; 19q	3p; 4q; 5q; 10q; 13q; 16q
鼻咽癌	1q; 3q; 8; 12; 19	1p; 3p; 9; 11q; 13q; 14q; 16q
前列腺癌	7; 8q; Xq	8p; 13q; 16q

图 2B

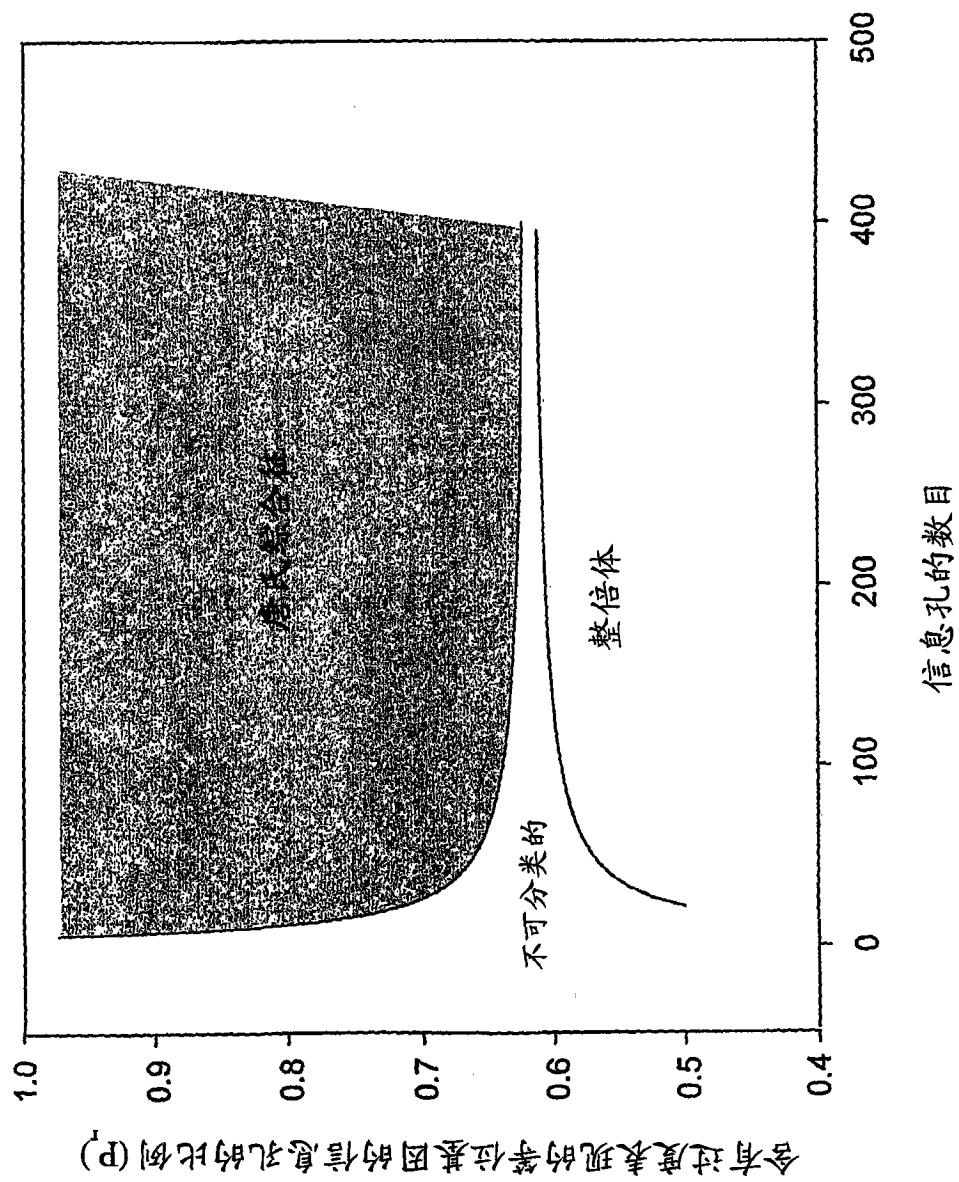


图 3

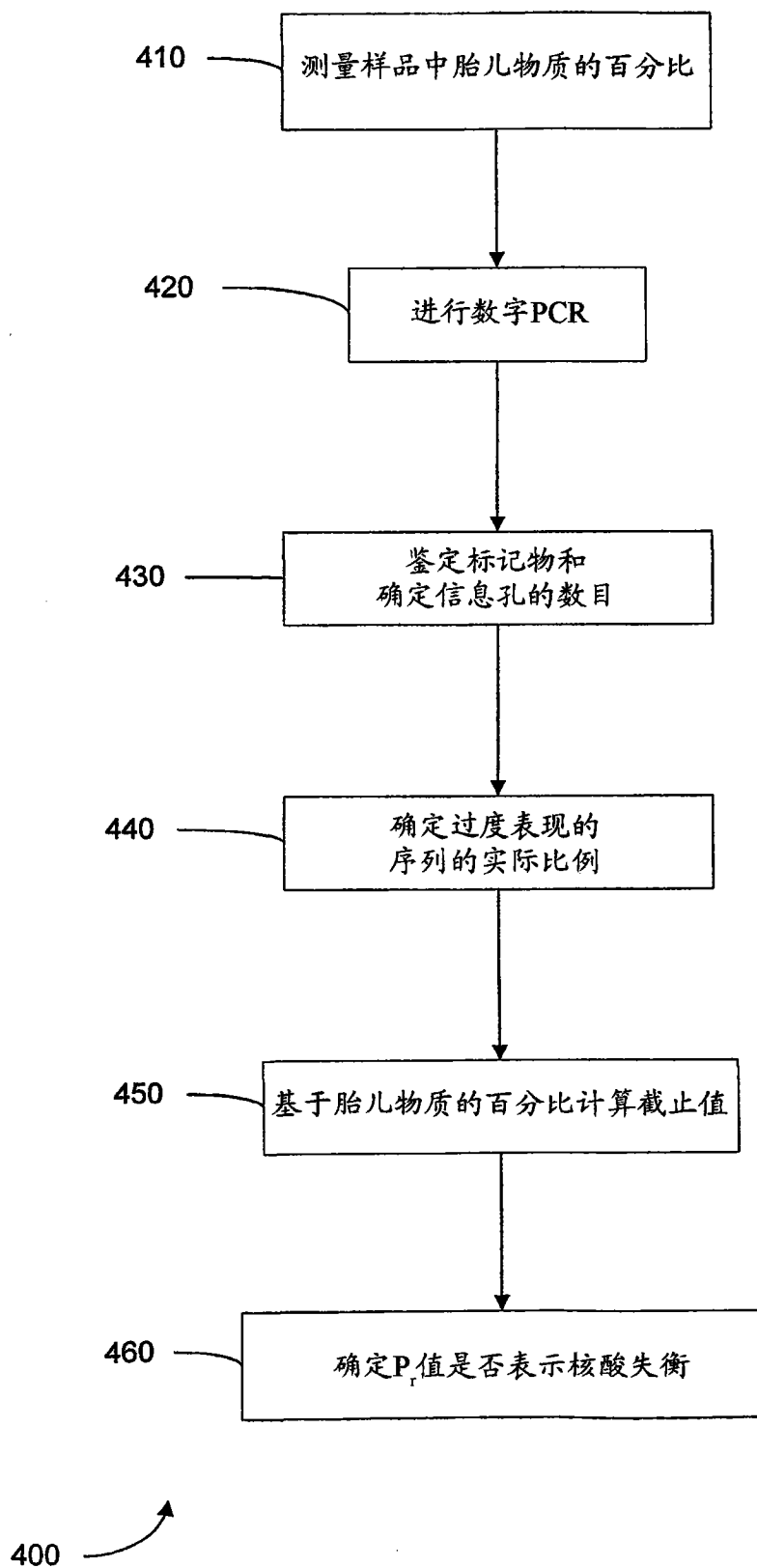


图 4

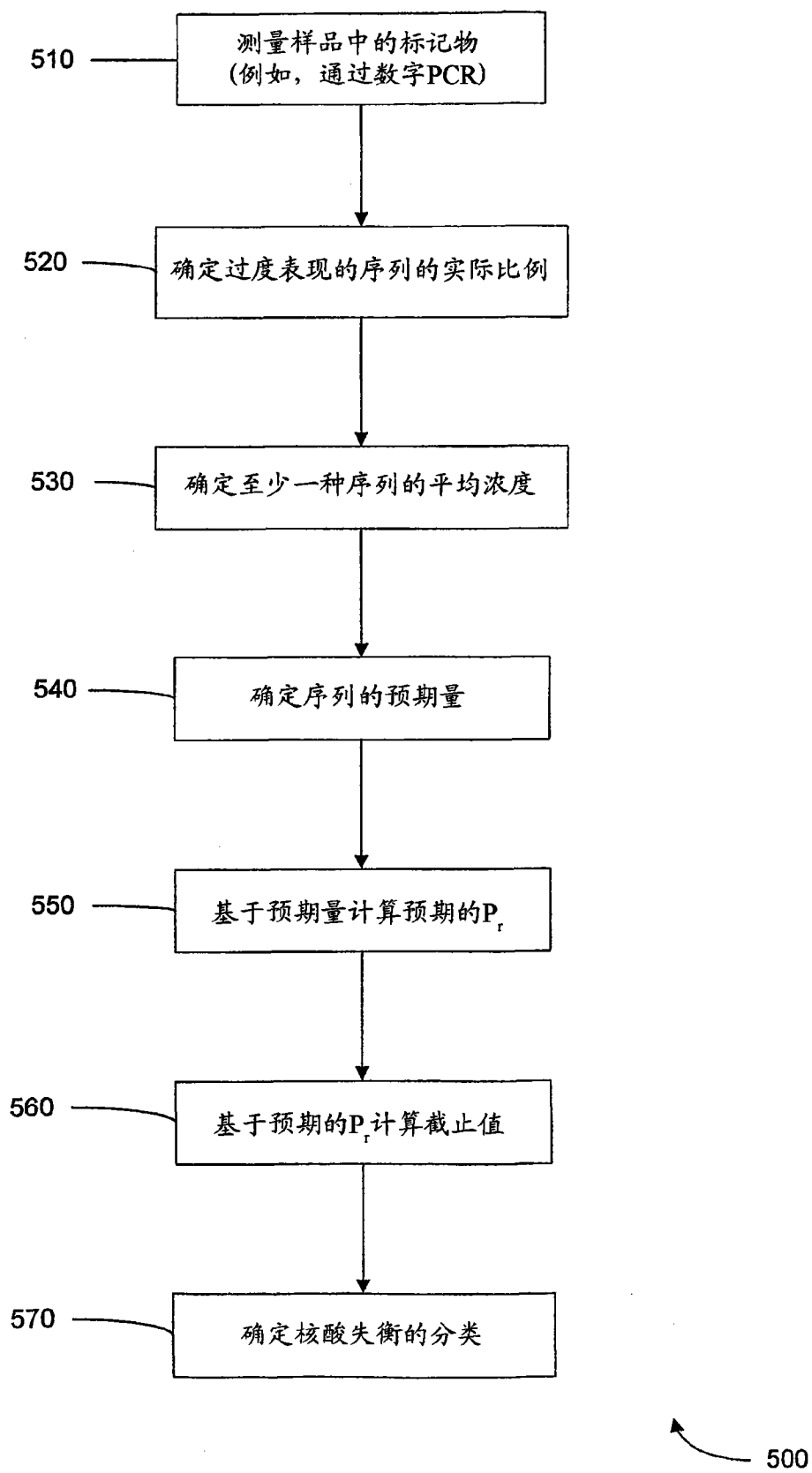


图 5

m	对参考等位基因是阳性的孔的比例	对过度表现的等位基因是阳性的孔的比例	对两个等位基因都是阳性的孔的比例	只对参考等位基因是阳性的孔的比例	只对过度表现的等位基因是阳性的孔的比例	数字 RNA-SNP 等位基因比率	信息孔的比例	P_r
0.1	0.0952	0.1813	0.0173	0.0779	0.164	2.11	0.2419	0.68
0.2	0.1813	0.3297	0.0598	0.1215	0.2699	2.22	0.3914	0.69
0.3	0.2592	0.4512	0.1169	0.1422	0.3342	2.35	0.4765	0.7
0.4	0.3297	0.5507	0.1815	0.1481	0.3691	2.49	0.5173	0.71
0.5	0.3935	0.6321	0.2487	0.1448	0.3834	2.65	0.5282	0.73
0.6	0.4512	0.6988	0.3153	0.1359	0.3835	2.82	0.5194	0.74
0.7	0.5034	0.7534	0.3793	0.1241	0.3741	3.01	0.4983	0.75
0.8	0.5507	0.7981	0.4395	0.1112	0.3586	3.23	0.4698	0.76
0.9	0.5934	0.8347	0.4953	0.0981	0.3394	3.46	0.4375	0.78
1	0.6321	0.8647	0.5466	0.0855	0.3181	3.72	0.4036	0.79
1.1	0.6671	0.8892	0.5932	0.0739	0.296	4	0.3699	0.8
1.2	0.6988	0.9093	0.6354	0.0634	0.2739	4.32	0.3373	0.81
1.3	0.7275	0.9257	0.6734	0.054	0.2523	4.67	0.3063	0.82
1.4	0.7534	0.9392	0.7076	0.0458	0.2316	5.06	0.2774	0.83
1.5	0.7769	0.9502	0.7382	0.0387	0.212	5.48	0.2507	0.85
1.6	0.7981	0.9592	0.7656	0.0325	0.1937	5.95	0.2262	0.86
1.7	0.8173	0.9666	0.79	0.0273	0.1766	6.47	0.2039	0.87
1.8	0.8347	0.9727	0.8119	0.0228	0.1608	7.05	0.1836	0.88
1.9	0.8504	0.9776	0.8314	0.019	0.1462	7.69	0.1652	0.88
2	0.8647	0.9817	0.8488	0.0158	0.1329	8.39	0.1487	0.89

600

图 6

m	样品中胎儿DNA的百分比							
	10%		25%		50%		100%	
	信息孔的百分比	P _r	信息孔的百分比	P _r	信息孔的百分比	P _r	信息孔的百分比	P _r
0.1	18%	0.51	18%	0.53	19%	0.56	21%	0.61
0.2	30%	0.51	31%	0.53	32%	0.56	35%	0.61
0.3	39%	0.51	40%	0.53	41%	0.57	43%	0.62
0.4	45%	0.51	45%	0.54	46%	0.57	48%	0.63
0.5	48%	0.52	49%	0.54	49%	0.57	51%	0.63
0.6	50%	0.52	50%	0.54	50%	0.58	51%	0.64
0.7	50%	0.52	50%	0.54	50%	0.58	50%	0.65
0.8	49%	0.52	49%	0.54	49%	0.58	48%	0.65
0.9	48%	0.52	47%	0.55	47%	0.59	46%	0.66
1	46%	0.52	45%	0.55	44%	0.59	43%	0.67
1.1	44%	0.52	43%	0.55	42%	0.6	40%	0.68
1.2	41%	0.52	40%	0.55	39%	0.6	37%	0.69
1.3	39%	0.52	38%	0.55	36%	0.6	34%	0.69
1.4	36%	0.52	35%	0.56	33%	0.61	31%	0.7
1.5	34%	0.52	33%	0.56	31%	0.61	28%	0.71
1.6	31%	0.52	30%	0.56	28%	0.62	26%	0.72
1.7	29%	0.53	28%	0.56	26%	0.62	23%	0.73
1.8	27%	0.53	25%	0.57	24%	0.63	21%	0.73
1.9	24%	0.53	23%	0.57	21%	0.63	19%	0.74
2	22%	0.53	21%	0.57	20%	0.64	17%	0.75

700

图 7

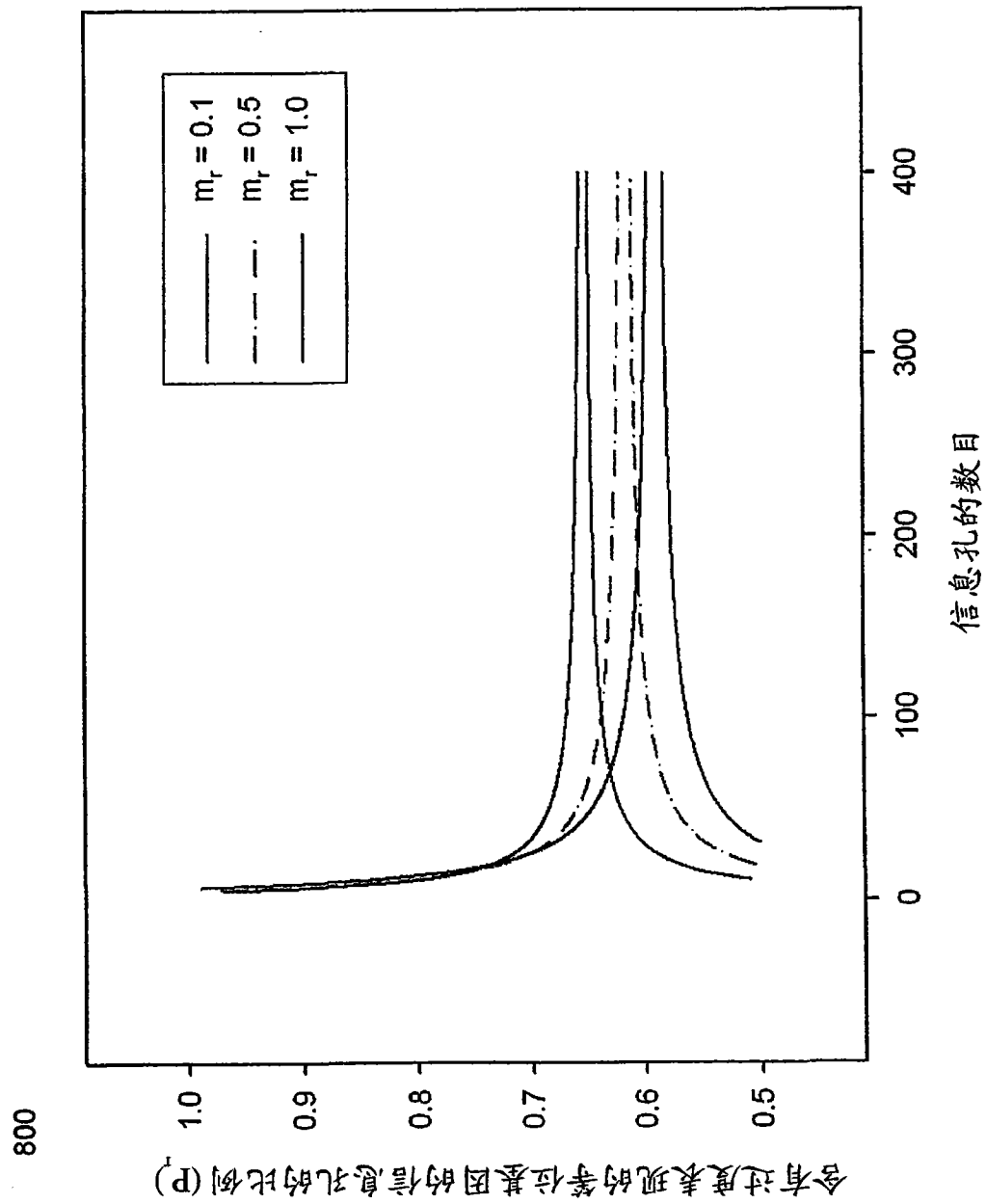


图 8

mr	分类	新算法			旧算法			mr	分类	新算法		旧算法	
		整倍体	21三体性	21三体性	整倍体	21三体性	21三体性			整倍体	21三体性	整倍体	21三体性
0.1	正确	0%	37%		0%	35%		1.1	正确	89%	92%	14%	35%
	错误	3%	2%		3%	1%			错误	1%	0%	4%	0%
	未分类的	97%	61%		97%	63%			未分类的	10%	7%	81%	65%
0.2	正确	6%	64%		0%	40%		1.2	正确	90%	92%	11%	35%
	错误	4%	2%		4%	0%			错误	1%	0%	4%	0%
	未分类的	90%	34%		96%	60%			未分类的	9%	8%	84%	65%
0.3	正确	34%	75%		6%	39%		1.3	正确	91%	92%	7%	34%
	错误	4%	2%		4%	0%			错误	1%	1%	4%	0%
	未分类的	62%	24%		90%	61%			未分类的	8%	7%	89%	66%
0.4	正确	54%	81%		14%	39%		1.4	正确	92%	91%	5%	34%
	错误	3%	1%		4%	0%			错误	1%	1%	4%	0%
	未分类的	43%	19%		82%	61%			未分类的	8%	9%	91%	66%
0.5	正确	64%	85%		20%	38%		1.5	正确	92%	90%	2%	32%
	错误	3%	1%		4%	0%			错误	1%	1%	4%	0%
	未分类的	33%	14%		76%	62%			未分类的	8%	9%	93%	68%
0.6	正确	74%	88%		23%	38%		1.6	正确	92%	89%	1%	31%
	错误	2%	1%		4%	0%			错误	1%	1%	4%	0%
	未分类的	24%	11%		72%	62%			未分类的	8%	10%	94%	69%
0.7	正确	80%	89%		25%	37%		1.7	正确	92%	89%	1%	30%
	错误	2%	1%		4%	0%			错误	1%	1%	4%	0%
	未分类的	18%	10%		71%	63%			未分类的	7%	10%	96%	70%
0.8	正确	83%	91%		23%	37%		1.8	正确	92%	88%	0%	28%
	错误	2%	1%		4%	0%			错误	1%	1%	4%	0%
	未分类的	15%	9%		73%	63%			未分类的	7%	11%	96%	72%
0.9	正确	86%	91%		21%	36%		1.9	正确	92%	87%	0%	26%
	错误	1%	1%		4%	0%			错误	1%	1%	3%	0%
	未分类的	12%	8%		75%	64%			未分类的	8%	12%	97%	74%
1	正确	88%	92%		19%	36%		2	正确	91%	86%	0%	24%
	错误	1%	0%		4%	0%			错误	1%	1%	4%	0%
	未分类的	11%	8%		77%	64%			未分类的	9%	13%	96%	76%

900

图 9A

m _r	分类	新算法			m _r	分类	新算法			旧算法	
		整倍体	21三体性	整倍体			21三体性	整倍体	21三体性		整倍体
0.1	正确	57%	88%	48%	1.1	正确	100%	100%	92%		90%
	错误	3%	1%	3%		错误	0%	0%	1%		0%
	未分类的	40%	11%	49%		未分类的	0%	0%	7%		10%
0.2	正确	89%	98%	79%	1.2	正确	100%	100%	90%		90%
	错误	1%	0%	2%		错误	0%	0%	1%		0%
	未分类的	10%	2%	19%		未分类的	0%	0%	9%		10%
0.3	正确	96%	99%	88%	1.3	正确	100%	100%	89%		90%
	错误	0%	0%	1%		错误	0%	0%	1%		0%
	未分类的	4%	1%	11%		未分类的	0%	0%	10%		10%
0.4	正确	99%	100%	92%	1.4	正确	100%	100%	87%		90%
	错误	0%	0%	1%		错误	0%	0%	1%		0%
	未分类的	1%	0%	7%		未分类的	0%	0%	12%		10%
0.5	正确	100%	100%	94%	1.5	正确	100%	100%	85%		90%
	错误	0%	0%	1%		错误	0%	0%	1%		0%
	未分类的	0%	0%	6%		未分类的	0%	0%	13%		10%
0.6	正确	100%	100%	94%	1.6	正确	100%	100%	82%		90%
	错误	0%	0%	1%		错误	0%	0%	1%		0%
	未分类的	0%	0%	5%		未分类的	0%	0%	17%		10%
0.7	正确	100%	100%	95%	1.7	正确	100%	100%	79%		90%
	错误	0%	0%	1%		错误	0%	0%	2%		0%
	未分类的	0%	0%	5%		未分类的	0%	0%	19%		10%
0.8	正确	100%	100%	95%	1.8	正确	100%	100%	74%		90%
	错误	0%	0%	1%		错误	0%	0%	2%		0%
	未分类的	0%	0%	5%		未分类的	0%	0%	24%		10%
0.9	正确	100%	100%	94%	1.9	正确	100%	100%	70%		90%
	错误	0%	0%	1%		错误	0%	0%	2%		0%
	未分类的	0%	0%	6%		未分类的	0%	0%	28%		10%
1	正确	100%	100%	94%	2	正确	100%	100%	67%		90%
	错误	0%	0%	1%		错误	0%	0%	3%		0%
	未分类的	0%	0%	6%		未分类的	0%	0%	30%		10%

950

图 9B

m_r	分类结果	整倍体	T21	m_r	分类结果	整倍体	T21
0.1	正确	57%	88%	0.9	正确	100%	100%
	错误	3%	1%		错误	0%	0%
	未分类的	40%	11%		未分类的	0%	0%
0.2	正确	89%	98%	1	正确	100%	100%
	错误	1%	0%		错误	0%	0%
	未分类的	10%	2%		未分类的	0%	0%
0.3	正确	96%	99%	1.1	正确	100%	100%
	错误	0%	0%		错误	0%	0%
	未分类的	4%	1%		未分类的	0%	0%
0.4	正确	99%	100%	1.2	正确	100%	100%
	错误	0%	0%		错误	0%	0%
	未分类的	1%	0%		未分类的	0%	0%
0.5	正确	100%	100%	1.3	正确	100%	100%
	错误	0%	0%		错误	0%	0%
	未分类的	0%	0%		未分类的	0%	0%
0.6	正确	100%	100%	1.4	正确	100%	100%
	错误	0%	0%		错误	0%	0%
	未分类的	0%	0%		未分类的	0%	0%
0.7	正确	100%	100%	1.5	正确	100%	100%
	错误	0%	0%		错误	0%	0%
	未分类的	0%	0%		未分类的	0%	0%
0.8	正确	100%	100%	2	正确	100%	100%
	错误	0%	0%		错误	0%	0%
	未分类的	0%	0%		未分类的	0%	0%

1000

图 10

m_T	分类结果	整倍体	21三体性	m_T	分类结果	整倍体	21三体性
0.1	正确	38%	44%	0.7	正确	95%	94%
	错误	2%	2%		错误	0%	1%
	未分类的	60%	54%		未分类的	5%	5%
0.2	正确	64%	69%	0.8	正确	95%	95%
	错误	2%	2%		错误	1%	0%
	未分类的	34%	29%		未分类的	4%	5%
0.3	正确	78%	82%	0.9	正确	96%	96%
	错误	2%	1%		错误	0%	0%
	未分类的	20%	17%		未分类的	4%	4%
0.4	正确	85%	87%	1	正确	97%	96%
	错误	1%	1%		错误	0%	0%
	未分类的	14%	12%		未分类的	3%	4%
0.5	正确	89%	90%	1.5	正确	97%	96%
	错误	1%	1%		错误	0%	0%
	未分类的	10%	9%		未分类的	3%	4%
0.6	正确	92%	93%	2	正确	96%	95%
	错误	1%	1%		错误	0%	0%
	未分类的	7%	7%		未分类的	4%	5%

1100

图 11

孔的 数目	分类结果	样品中胎儿DNA的百分比										
		10%			25%			50%			100%	
		整倍体	21三体性	整倍体	21三体性	整倍体	21三体性	整倍体	21三体性	整倍体	21三体性	
384	正确	0%	0%	6%	7%	47%	49%	89%	90%			
	错误	0%	0%	0%	1%	2%	2%	1%	1%			
	未分类的	100%	100%	93%	92%	51%	49%	10%	9%			
768	正确	0%	0%	23%	24%	73%	76%	98%	99%			
	错误	0%	0%	2%	2%	2%	1%	0%	0%			
	未分类的	100%	100%	75%	74%	25%	23%	2%	1%			
1152	正确	1%	1%	39%	40%	86%	87%	100%	100%			
	错误	0%	0%	2%	2%	1%	1%	0%	0%			
	未分类的	99%	99%	59%	58%	13%	12%	0%	0%			
1536	正确	2%	2%	50%	51%	93%	93%	100%	100%			
	错误	0%	0%	2%	2%	1%	1%	0%	0%			
	未分类的	98%	97%	48%	47%	7%	6%	0%	0%			
1920	正确	4%	4%	60%	60%	96%	96%	100%	100%			
	错误	0%	1%	2%	2%	0%	0%	0%	0%			
	未分类的	96%	95%	38%	37%	4%	3%	0%	0%			
3840	正确	18%	19%	84%	83%	100%	100%	100%	100%			
	错误	1%	1%	1%	1%	0%	0%	0%	0%			
	未分类的	81%	80%	15%	15%	18%	22%	0%	0%			
7680	正确	43%	45%	97%	97%	100%	100%	100%	100%			
	错误	2%	2%	0%	0%	0%	0%	0%	0%			
	未分类的	55%	53%	3%	3%	0%	0%	0%	0%			

1200

图 12

样品	基因型 ^a	孔的数目	对单独的等位基因是阳性的孔的数目				m _r ^b	SPRT结果			
			仅A	仅G	AG	全部阴性		P _r ^c	不可分类的区域 ^d	分类 ^e	
胎盘DNA											
N677	AG	384	85	83	126	90	0.79	0.51	0.63 - 0.65	整倍体	
N710	AG	384	102	83	73	126	0.52	0.55	0.61 - 0.63	整倍体	
N435	AGG	384	49	157	130	48	0.63	0.76	0.62 - 0.64	T21	
N981	AAG	384	135	69	82	98	0.5	0.66	0.61 - 0.63	T21	
胎盘RNA											
V533	AG	384	103	93	71	117	0.56	0.53	0.61 - 0.63	整倍体	
V943	AG	384	89	100	74	121	0.55	0.53	0.61 - 0.63	整倍体	
N435	AGG	384	52	138	95	99	0.48	0.73	0.61 - 0.63	T21	
T215	AAG	384	146	58	138	42	0.71	0.72	0.62 - 0.64	T21	

图 13A

1300

样品	基因型	孔的数目	对单独的等位基因是阳性的孔的数目				m	SPRT结果		
			仅A	仅G	AG	全部阴性		比率	不可分类的区域	分类
M2390P	AG	384	90	100	97	97	0.67	0.526	0.62 - 0.64	整倍体
M2391P	AG	384	97	105	65	117	0.55	0.52	0.61 - 0.63	整倍体
M2473P	AG	384	66	92	34	192	0.3	0.582	0.59 - 0.62	整倍体
M2524P	AG	384	29	28	3	324	0.08	0.509	0.54 - 0.64	整倍体
M2528P	AG	384	112	85	44	143	0.41	0.569	0.60 - 0.62	整倍体
M2601P	AG	384	90	101	72	121	0.55	0.529	0.61 - 0.63	整倍体
M2607P	AG	384	73	91	57	163	0.41	0.555	0.60 - 0.63	整倍体
M2638P	AG	384	66	90	52	176	0.37	0.577	0.59 - 0.62	整倍体
M2639P	AG	384	71	56	17	240	0.21	0.559	0.58 - 0.62	整倍体
M2525P	AAG	384	110	53	21	200	0.21	0.675	0.58 - 0.61	T21
M2272P	AAG	768	246	127	112	283	0.37	0.66	0.60 - 0.61	T21
M2718P	AGG	384	66	114	66	138	0.42	0.633	0.60 - 0.62	T21
M1519P	AGG	384	58	130	54	142	0.34	0.691	0.59 - 0.62	T21

1350

图 13B

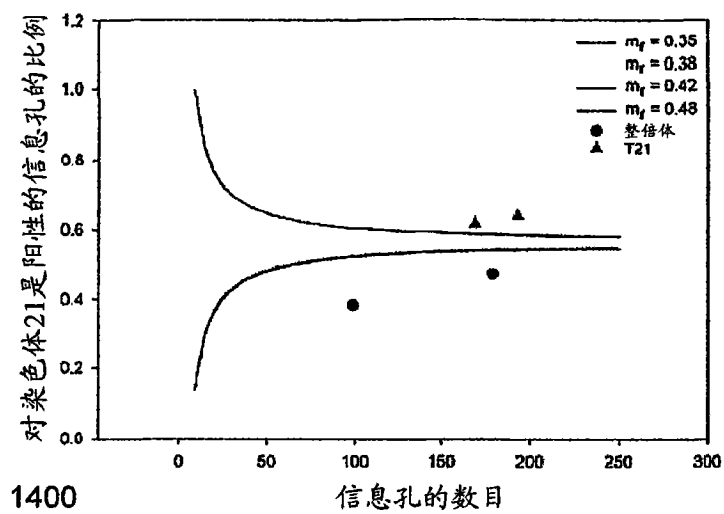


图 14A

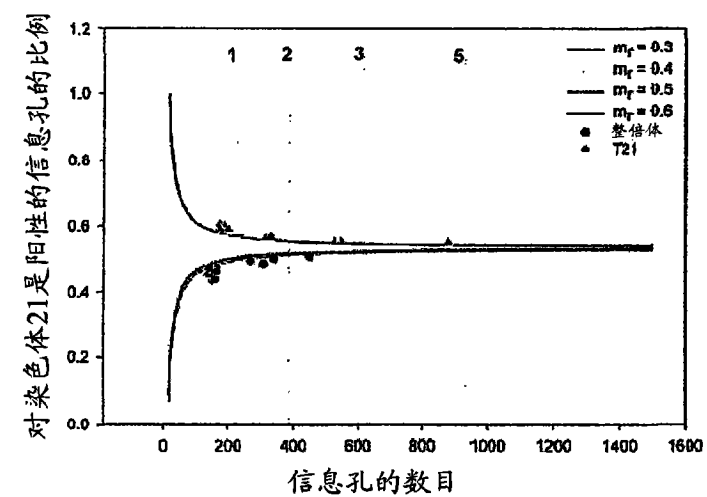


图 14B

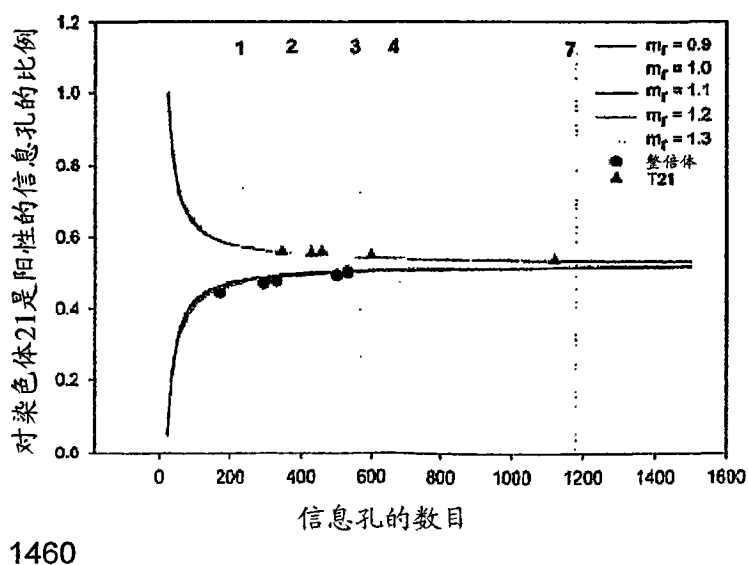


图 14C

样品	基因型	孔的数目	对单独的等位基因是阳性的孔的数目				m_r^a	SPRT结果		
			仅A	仅G	GA	全部阴性		P_r^b	不可分类的区域	分类 ^c
V943	AG	765	168	209	117	271	0.47	0.554	0.609- 0.621	整倍体
N435	AGG	765	92	281	105	287	0.30	0.753	0.597- 0.610	T21
T21-5	AAG	765	267	126	141	231	0.43	0.679	0.607- 0.618	T21

^a 每孔参考分子的数目

^b 通过下述方程式计算 P_r 值:

$$P_r = \frac{\text{对过度表现的等位基因是阳性的孔的数目}}{\text{只对A是阳性的孔的数目} + \text{只对G是阳性的孔的数目}}$$

^c 当 P_r 低于不可分类的区域时, 指定为“整倍体”; 当 P_r 高于不可分类的区域时, 指定为“T21”。T21即21三体性

图 15A

板#	孔的数目	对单独的等位基因是阳性的孔的数目			
		仅A	仅G	GA	全部阴性
S1	765	14	15	0	736
S2	765	7	10	0	748
S3	765	13	9	1	744
S4	765	7	8	0	750
S5	765	12	8	0	745
S6	765	8	14	0	743
S7	765	13	9	0	743
S8	765	9	8	0	748
S9	765	14	15	0	736
S10	765	9	12	0	744
S11	765	8	17	0	740
S12	765	11	9	0	745

图 15B

样品	基因型	板 ^a	对单独的等位基因是阳性的孔的数目					SPRT结果		
			孔的数目	仅A	仅G	GA	全部阴性	P _r	不可分类的区域	分类
M2879P	AG	5	3825	82	103	5	3635	0.557	0.570- 0.602	整倍体
M2875P	AG	3	2295	35	40	0	2220	0.533	0.546- 0.625	整倍体
M3078P	AG	3	2295	33	33	1	2228	0.500	0.541- 0.631	整倍体
M2976P	AG	12	9180	56	46	0	9078	0.549	0.556- 0.615	整倍体
M2831P	AAG	2	1530	42	13	0	1475	0.764	0.531- 0.640	T21

^a 分类所需的板的数目

图 15C

样品	基因型 ^a	对A或G等位基因是阳性的孔的数目			m_r^b	SPRT分析		
		仅A	仅G	两者		P_r^c	不可分类的区域 ^d	分类 ^e
N3505	AG	157	131	48	0.27	0.545	0.59-0.61	整倍体
N3530	AG	174	132	52	0.28	0.569	0.59-0.61	整倍体
N3531	AG	146	129	28	0.23	0.531	0.58-0.60	整倍体
N3533	AG	149	135	46	0.27	0.525	0.59-0.61	整倍体
N2772	AG	159	149	79	0.35	0.516	0.60-0.61	整倍体
N3506	AG	154	145	66	0.32	0.515	0.60-0.61	整倍体
N3555	AG	94	73	7	0.11	0.563	0.58-0.61	整倍体
N3560	AG	101	109	28	0.18	0.519	0.58-0.61	整倍体
N3561	AG	91	70	17	0.12	0.565	0.58-0.61	整倍体
N828	AAG	153	89	22	0.16	0.632	0.58-0.60	T18
CA2	AGG	12	28	0	0.016	0.700	0.51-0.66	T18
CA7	AGG	97	171	56	0.22	0.638	0.59-0.61	T18

^a 通过质谱来确定基因型

^b m_r 值表示每孔的参考模板分子的平均数目

^c P_r 值表示对过度表现的等位基因是阳性的信息孔的比例

^d 不可分类的区域随 m_r 值而变化

^e 当 P_r 低于不可分类的区域时, 指定为“整倍体”; 当 P_r 高于不可分类的区域时, 指定为“T18”

图 16A

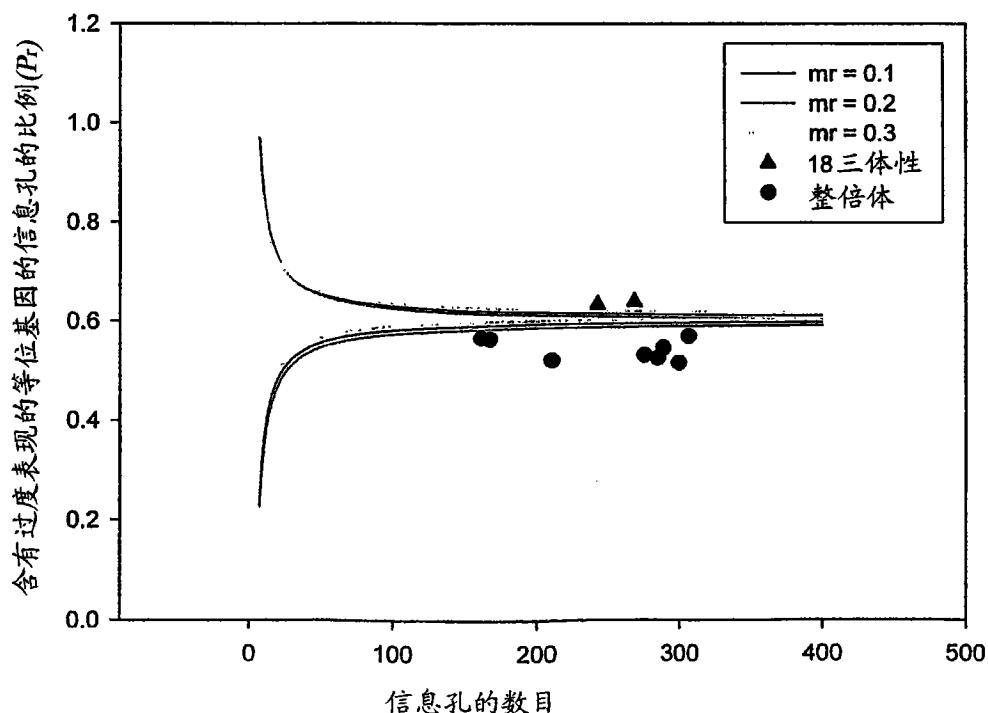


图 16B

类型	个例	所用的板 的总数	信息孔 的总数	孔计数的数目			SPRT分析			
				两者均 阳性	仅染色 体21	仅染色 体1	m_r^a	P_r^b	不可分类的区域 分类 ^c	
T21	N0234	1	303	68	172	131	0.30	0.568	0.507 - 0.568	T21
	N0274	2	558	118	307	251	0.28	0.550	0.518 - 0.547	T21
	N0332	1	293	73	188	125	0.30	0.573	0.508 - 0.580	T21
	N0348	2	574	107	322	252	0.27	0.581	0.518 - 0.548	T21
	N0435	1	299	87	178	123	0.29	0.589	0.508 - 0.559	T21
	N0458	1	318	88	189	129	0.30	0.594	0.508 - 0.558	T21
	N0687	2	602	134	345	257	0.30	0.573	0.519 - 0.548	T21
	N0764	1	288	62	167	122	0.28	0.578	0.506 - 0.560	T21
	N0890	4	1208	272	657	549	0.31	0.545	0.528 - 0.539	T21
	N0891	1	317	103	180	137	0.38	0.588	0.510 - 0.558	T21
正常	N0186	1	262	68	111	151	0.34	0.424	0.504 - 0.583	整倍体
	N0188	1	285	68	132	153	0.34	0.463	0.506 - 0.581	整倍体
	N0230	2	527	109	272	255	0.27	0.516	0.517 - 0.548	整倍体
	N0231	2	542	98	275	287	0.27	0.507	0.517 - 0.547	整倍体
	N0282	1	285	57	144	151	0.32	0.488	0.508 - 0.580	整倍体
	N0463	1	263	70	120	143	0.33	0.466	0.504 - 0.563	整倍体
	N0527	1	292	66	137	155	0.32	0.489	0.508 - 0.580	整倍体
	N0819	1	281	80	138	143	0.31	0.491	0.505 - 0.581	整倍体
	N0825	1	291	80	134	157	0.33	0.480	0.507 - 0.580	整倍体
	N0851	2	566	118	287	279	0.30	0.507	0.518 - 0.547	整倍体

^a 每孔的参考分子的数目

^b 通过下述方程式计算 P_r 值:

对过度表现的等位基因(染色体21)是阳性的孔的数目

只对染色体21是阳性的孔的数目+只对染色体1是阳性的孔的数目

^c 当 P_r 低于不可分类的区域时, 指定为“整倍体”; 当 P_r 高于不可分类的区域时, 指定为“T21”。

图 17

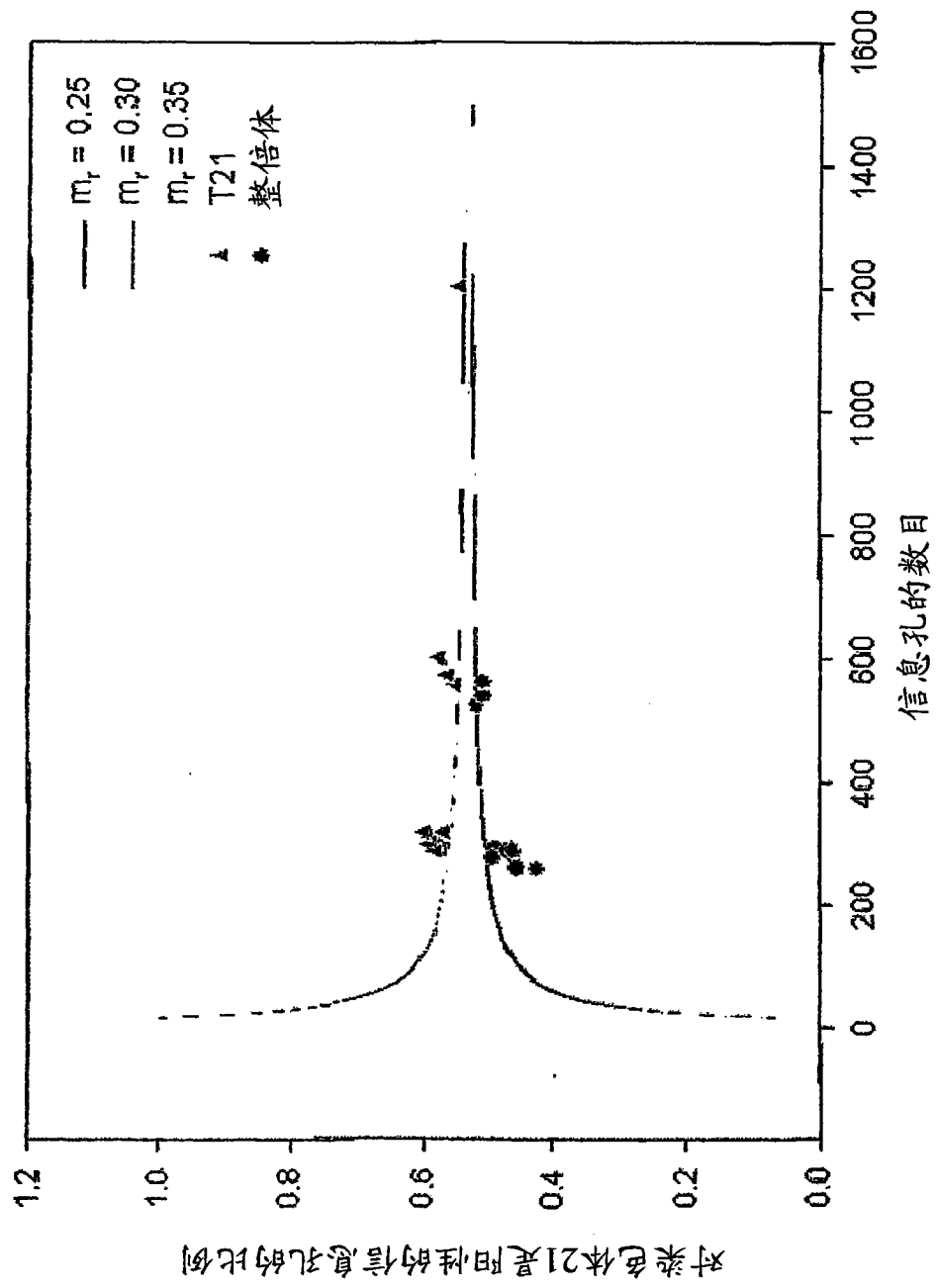


图 18

类型	个例	所用板的数目	信息孔的数目	孔计数的数目			SPRT分析	
				两者均阳性	仅染色体21	仅染色体1	P^b	不可分类的区域分类 ^a
T21	A0274	1	380	268	248	131	0.855	0.567-0.585
	A2008	1	348	278	233	115	0.870	0.579-0.596
	A2178	1	384	288	238	145	0.822	0.571-0.598
	AS363A	1	364	109	244	120	0.870	0.551-0.573
	AS378A	1	258	482	188	88	0.734	0.590-0.610
	AS3457A	1	376	83	258	120	0.881	0.548-0.572
正常	A0236	1	387	275	185	182	0.504	0.576-0.592
	A0237	1	382	228	184	208	0.488	0.572-0.588
	A0255	1	378	222	195	183	0.518	0.588-0.585
	A0283	1	264	43	127	127	0.500	0.541-0.578
	AS365A	1	370	135	182	178	0.518	0.558-0.578
	AS483A	1	354	84	188	188	0.525	0.550-0.574
	AS484A	1	372	271	184	178	0.522	0.573-0.590
	AS780A	1	413	188	200	213	0.484	0.588-0.584
	AS780A	1	316	87	159	157	0.503	0.548-0.575
	AS857A	1	198	8	87	111	0.438	0.532-0.579
	AS821A	1	152	8	89	83	0.454	0.523-0.588
	AS980A	1	394	137	191	203	0.485	0.581-0.580
	AS2282A	1	334	342	177	157	0.530	0.580-0.587
	AS2425A	1	128	8	84	84	0.500	0.518-0.591
	AS2434A	1	378	97	203	176	0.538	0.555-0.575
	A1118	1	354	123	185	188	0.523	0.555-0.577
	A1122	1	389	100	195	174	0.528	0.554-0.576
	A1152	1	258	48	128	128	0.500	0.542-0.576
	A1180	1	246	488	133	113	0.541	0.604-0.622
	A1275	1	373	85	201	172	0.538	0.554-0.575
	A1375	1	359	55	190	188	0.528	0.550-0.573
	A1388	1	367	87	165	182	0.482	0.555-0.577
	A1398	1	371	114	187	174	0.531	0.556-0.577
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体
								整倍体

^a 每孔的参考分子的数目

^b 通过下述方程式计算 P^b 值:

对过度表现的等位基因(染色体21)是阳性的孔的数目
只对染色体21是阳性的孔的数目+只对染色体1是阳性的孔的数目

^c 当 P^b 低于不可分类的区域时, 指定为“整倍体”; 当 P^b 高于不可分类的区域时, 指定为“T21”

图 19

类型	个例	所用板的 的总数	信息孔 的总数	孔计数的数目			SPRT分析			
				两者均 阳性	仅染色 体18	仅染色 体21	m^a	P_r^b	不可分类的区域	分类 ^c
整倍体	N0230	1	380	243	181	199	0.86	0.476	0.573 - 0.589	E
	N0625	1	370	225	179	191	0.78	0.484	0.569 - 0.586	E
	N0651	1	358	206	192	166	0.67	0.536	0.564 - 0.583	E
T18	N0828	1	543	256	329	214	0.95	0.606	0.579 - 0.590	T18
	N0830	1	346	353	240	106	0.92	0.694	0.574 - 0.592	T18
	N0936	1	372	285	209	163	0.88	0.562	0.573 - 0.590	E
	N1678	1	343	115	221	122	0.37	0.644	0.551 - 0.574	T18
	N1701	1	303	94	191	112	0.31	0.630	0.546 - 0.575	T18

^a 每孔的参考分子的数目

^b 通过下述方程式计算 P_r 值:

$$P_r = \frac{\text{对过度表现的等位基因(染色体18)是阳性的孔的数目}}{\text{只对染色体21是阳性的孔的数目} + \text{只对染色体18是阳性的孔的数目}}$$

^c 当 P_r 低于不可分类的区域时, 指定为“整倍体”; 当 P_r 高于不可分类的区域时, 指定为“T18”。

图 20

类型	个例	板的数目	测定	信息孔的总数	孔计数的数目			m_i^a	P_i^b	SPRT分析 不可分类的区域	分类 ^c
					两者均 阳性	仅染色 体21	仅参考 染色体				
正常	N0196	1	chr21/chr1	148	78	71	77	0.52	0.480	0.489 - 0.584	E
		1	chr21/chr18	152	83	79	73	0.44	0.520	0.487 - 0.584	U
			总计	300	141	150	150	0.48	0.500	0.512 - 0.560	E
	N0198	1	chr21/chr1	162	73	70	92	0.56	0.432	0.490 - 0.581	E
		1	chr21/chr18	183	68	93	90	0.53	0.508	0.498 - 0.575	U
			总计	345	141	163	182	0.55	0.472	0.517 - 0.557	E
	N0230	1	chr21/chr1	162	58	81	81	0.45	0.500	0.490 - 0.581	U
		1	chr21/chr18	165	45	81	84	0.41	0.491	0.489 - 0.580	U
			总计	327	103	162	165	0.43	0.495	0.512 - 0.558	E
	N0231	1	chr21/chr1	173	53	81	92	0.47	0.468	0.494 - 0.578	E
		1	chr21/chr18	166	51	85	81	0.42	0.512	0.490 - 0.580	U
			总计	339	104	166	173	0.45	0.480	0.514 - 0.557	E
	N0262	1	chr21/chr1	186	97	101	85	0.64	0.543	0.503 - 0.575	U
		1	chr21/chr18	176	111	91	85	0.71	0.517	0.504 - 0.557	U
			总计	362	208	192	170	0.68	0.530	0.522 - 0.557	U
		2	chr21/chr1	174	103	85	89	0.69	0.489	0.503 - 0.577	E
			总计	536	311	277	259	0.68	0.517	0.523 - 0.552	E
T21	N0234	1	chr21/chr1	193	68	108	85	0.51	0.580	0.500 - 0.573	U
		1	chr21/chr18	178	65	82	86	0.50	0.517	0.496 - 0.576	U
			总计	371	133	200	171	0.50	0.539	0.517 - 0.556	U
		2	chr21/chr1	178	73	92	86	0.53	0.517	0.498 - 0.576	U
			总计	549	206	292	257	0.51	0.532	0.524 - 0.549	U
		2	chr21/chr18	178	62	95	83	0.47	0.534	0.495 - 0.576	U
			总计	727	268	387	340	0.50	0.532	0.527 - 0.546	U
		3	chr21/chr1	184	108	103	81	0.88	0.560	0.504 - 0.575	U
			总计	911	376	490	421	0.54	0.538	0.529 - 0.545	U
		3	chr21/chr18	198	102	120	78	0.63	0.606	0.505 - 0.572	T21
			总计	1109	478	610	499	0.55	0.550	0.531 - 0.544	T21
	N0332	1	chr21/chr1	173	109	103	70	0.63	0.595	0.500 - 0.577	T21
		1	chr21/chr18	191	82	110	81	0.55	0.576	0.501 - 0.574	T21
			总计	364	191	213	151	0.59	0.585	0.519 - 0.557	T21
	N0349	1	chr21/chr1	158	62	83	75	0.44	0.525	0.489 - 0.582	U
		1	chr21/chr18	166	46	87	79	0.39	0.524	0.489 - 0.580	U
			总计	324	108	170	154	0.42	0.525	0.512 - 0.558	U
		2	chr21/chr1	149	66	88	63	0.41	0.577	0.485 - 0.585	U
			总计	473	174	256	217	0.41	0.541	0.519 - 0.551	U
		2	chr21/chr18	168	43	100	68	0.34	0.595	0.487 - 0.580	T21
			总计	641	217	356	285	0.40	0.555	0.523 - 0.546	T21
	N0764	1	chr21/chr1	182	82	106	76	0.53	0.582	0.498 - 0.575	T21
		1	chr21/chr18	178	83	97	81	0.56	0.545	0.498 - 0.576	U
			总计	360	165	203	157	0.54	0.564	0.518 - 0.556	T21
	N0890	1	chr21/chr1	180	103	108	72	0.61	0.600	0.501 - 0.576	T21
		1	chr21/chr18	188	68	104	84	0.50	0.553	0.499 - 0.574	U
			总计	368	171	212	156	0.55	0.576	0.518 - 0.556	T21

^a 每孔的参考分子的数目^b 通过下述方程式计算 P_i 值:

对过度表现的等位基因(染色体21)是阳性的孔的数目

只对染色体21是阳性的孔的数目+只对参考染色体是阳性的孔的数目

^c 当 P_i 低于不可分类的区域时,指定为“整倍体”;当 P_i 高于不可分类的区域时,指定为“T21”。

图 21

		泊松修正的总数		P _r	分类	
		染色体21	参考染色体			
整倍体	个例 1	测定 1	37	33	0.52345751	不可分类的
		测定 2	31	42	0.42506276	不可分类的
		测定 3	29	37	0.44198785	不可分类的
		测定 4	39	26	0.6008691	不可分类的
		板 1 小记	136	138	0.495821	不可分类的
		测定 1	32	25	0.56605316	不可分类的
		测定 2	27	24	0.53165424	不可分类的
		测定 3	10	17	0.3681535	不可分类的
		测定 4	24	18	0.5626764	不可分类的
	板 2 小记	93	84	0.52474	不可分类的	
	总计	229	223	0.507182	整倍体	
	个例 2	测定 1	45	35	0.56321585	不可分类的
		测定 2	31	45	0.41245525	不可分类的
		测定 3	32	22	0.59954763	不可分类的
		测定 4	38	35	0.5228175	不可分类的
		板 1 小记	146	135	0.519153	不可分类的
		测定 1	28	23	0.55272135	不可分类的
		测定 2	23	22	0.51196923	不可分类的
		测定 3	23	15	0.59688187	不可分类的
		测定 4	29	26	0.5293349	不可分类的
		板 2 小记	102	85	0.545204	不可分类的
		板 1+2 小记	248	221	0.529583	不可分类的
		测定 1	17	25	0.4122149	不可分类的
		测定 2	25	16	0.6027299	不可分类的
		测定 3	13	26	0.3382802	不可分类的
		测定 4	27	24	0.5316872	不可分类的
		板 3 小记	82	91	0.475775	不可分类的
		板 1-3 小记	331	311	0.515069	不可分类的
		测定 1	15	9	0.6269193	不可分类的
测定 2		14	15	0.4824155	不可分类的	
测定 3		18	17	0.5146313	不可分类的	
测定 4		21	18	0.5270080	不可分类的	
板 4 小记		69	60	0.532264	不可分类的	
板 1-4 小记		399	372	0.517942	不可分类的	

图 22A

	测定 1	10	13	0.4337957	不可分类的
	测定 2	11	14	0.4390082	不可分类的
	测定 3	7	16	0.3018022	不可分类的
	测定 4	19	10	0.6579085	不可分类的
	板 5 小记	48	54	0.469951	不可分类的
	总计	447	426	0.5123444	整倍体
T21					
个例 3	测定 1	50	33	0.60012253	不可分类的
	测定 2	56	47	0.54458055	不可分类的
	测定 3	46	28	0.6196626	不可分类的
	测定 4	56	36	0.6108731	不可分类的
	总计	208	144	0.59079	T21
个例 4	测定 1	56	36	0.61085513	不可分类的
	测定 2	38	32	0.53906912	不可分类的
	测定 3	40	24	0.62788068	不可分类的
	测定 4	47	27	0.634511	不可分类的
	总计	180	119	0.603476	T21

图 22B

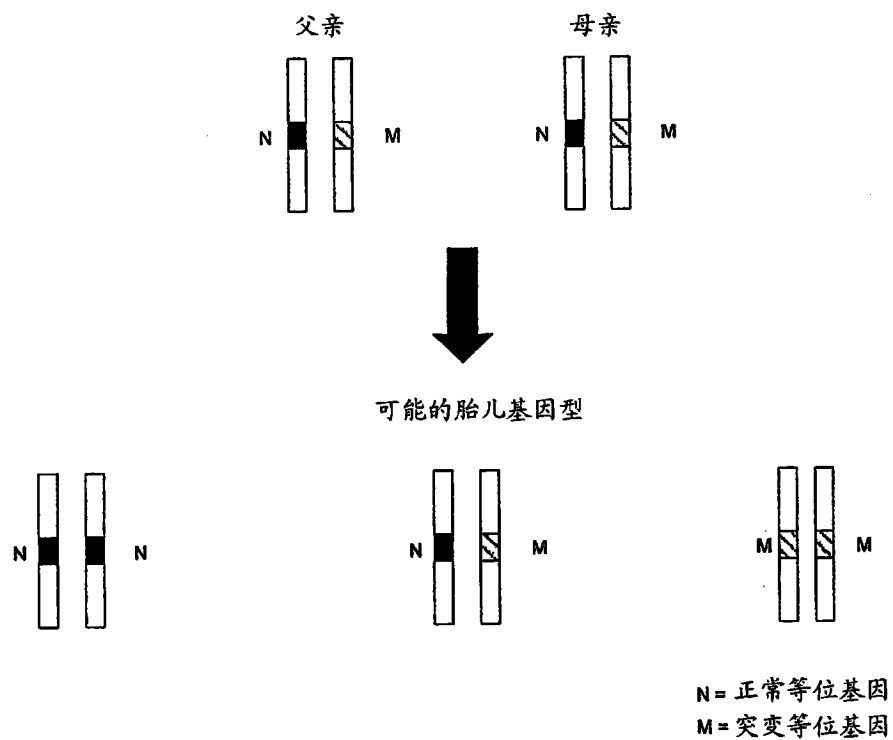


图 23

个例	样品	预期 %	板的数目	ZFY	ZFX	全部 阳性	净(NET) ZFY	净 ZFX	参考等 位基因	m_r	SPRT ^a
1	XY + XX (XX %)	50%	1	239	465	155	84	310	Y	0.37	XX
		25%	1	328	447	205	123	242	Y	0.56	XX
2	XY ₁ + XY ₂ (XY ₂ %)	50%	1	414	405	220	194	185	X	0.85	XY
		25%	1	432	426	245	187	181	X	0.81	XY

图 24A

个例	板的数目	ZFY	ZFX	全部 阳性	净(NET) ZFY	净 ZFX	参考等 位基因	m _r	SPRT
1	1	81	171	18	63	153	Y	0.11	XX
2	1	103	175	23	80	152	Y	0.14	XX
3	1	110	175	28	82	147	Y	0.16	XX
4	1	111	156	24	87	132	Y	0.16	XX
5	2	236	322	50	186	272	Y	0.17	XX
6	1	116	184	29	87	155	Y	0.16	XX
7	1	94	169	19	75	150	Y	0.13	XX
8	1	91	155	24	67	131	Y	0.13	XX
9	3	333	447	74	259	373	Y	0.16	XX
10	1	110	154	24	86	130	Y	0.16	XX
11	1	83	170	26	57	144	Y	0.11	XX
12	1	85	159	16	69	143	Y	0.12	XX

图 24B

个例	板的数目	母体 基因型 ^a	胎儿 基因型 ^a	预期胎儿 %	A	G	全部 阳性	净A	净G	实际 胎儿% ^b	参考等 位基因	m	SPRT ^d
1	1	AG	GG	15%	264	318	116	148	202	9.6	A	0.42	GG
2	1	AG	GG	15%	208	276	81	127	195	14.3	A	0.32	GG
3	1	AG	AG	15%	248	251	87	161	164	15.4	A	0.39	AG
4	1	AG	GG	15%	158	240	51	107	189	13.7	A	0.23	GG
5	1	AG	AG	15%	226	245	70	156	175	16.8	A	0.35	AG
6	1	AG	AG	15%	319	319	130	189	189	18.1	A	0.54	AG
7	1	AG	AG	15%	254	279	94	160	185	19.8	A	0.40	AG
8	1	AG	GG	15%	296	363	144	152	219	15.7	A	0.49	GG
9	1	AG	AG	15%	279	285	100	179	185	13.0	A	0.45	AG
10	1	AG	AA	15%	337	278	120	217	158	NA ^c	G	0.45	AA
1	1	AG	GG	20%	248	295	98	152	199	13.0	A	0.39	GG
2	1	AG	GG	20%	191	289	79	112	210	19.5	A	0.29	GG
3	1	AG	AG	20%	248	218	75	173	143	18.2	G	0.34	U
4	1	AG	GG	20%	158	257	59	99	198	25.9	A	0.23	GG
5	1	AG	AG	20%	231	215	71	160	144	20.3	G	0.33	AG
6	1	AG	AG	20%	281	283	98	183	185	26.3	A	0.46	AG
7	1	AG	AG	20%	256	231	88	168	143	22.7	G	0.36	AG
8	1	AG	GG	20%	246	311	104	142	207	23.0	A	0.39	GG
9	1	AG	AG	20%	267	249	83	184	166	21.9	G	0.39	AG
10	1	AG	AA	20%	303	203	77	226	126	NA ^c	G	0.31	AA
1	1	AG	GG	25%	244	304	99	145	205	15.7	A	0.38	GG
2	1	AG	GG	25%	209	314	88	121	226	24.9	A	0.32	GG
3	1	AG	AG	25%	204	212	55	149	157	28.0	A	0.31	AG
4	1	AG	GG	25%	185	290	76	109	214	25.9	A	0.28	GG
5	1	AG	AG	25%	233	243	77	156	166	24.5	A	0.36	AG
6	1	AG	AG	25%	292	305	110	182	195	27.3	A	0.48	AG
7	1	AG	AG	25%	261	275	95	168	180	22.5	A	0.42	AG
8	1	AG	GG	25%	236	327	98	140	231	26.7	A	0.37	GG
9	1	AG	AG	25%	271	280	104	167	176	25.2	A	0.44	AG
10	1	AG	AA	25%	243	154	46	197	108	NA ^c	G	0.22	AA
1	1	AG	GG	50%	136	293	57	79	236	33.6	A	0.20	GG
2	1	AG	GG	50%	163	358	78	85	280	45.8	A	0.24	GG
3	1	AG	AG	50%	200	208	53	147	155	55.7	A	0.30	AG
4	1	AG	GG	50%	137	356	66	71	290	53.0	A	0.18	GG
5	1	AG	AG	50%	247	246	91	158	155	57.7	G	0.39	AG
6	1	AG	AG	50%	218	245	68	150	177	53.8	A	0.34	AG
7	1	AG	AG	50%	255	282	103	152	179	51.3	A	0.41	AG
8	1	AG	GG	50%	158	398	83	75	315	57.6	A	0.23	GG
9	1	AG	AG	50%	235	225	74	161	151	55.7	G	0.35	AG
10	1	AG	AA	50%	317	106	35	282	71	NA ^c	G	0.15	AA

^a A=突变的等位基因, G=野生型等位基因

^b 通过数字ZFY/X测定来确定实际胎儿DNA%

^c 胎儿是雌性的, 不能应用数字ZFY/X测定

^d 除了个例10以外, 用通过数字ZFY/X测定确定的胎儿%来进行SPRT, U: 未分类的

图 25

个例	母体 基因型 ^a	胎儿 基因型 ^a	板的 数目	M	N	全部 阳性	净M	净N	1 ^b		2 ^c			
									预期胎儿 %	SPRT	通过数字 ZFY/X测定 的胎儿%	SPRT	参考等 位基因	m _t
1	MN	MM	3	848	763	273	573	490	5%	MM	7.8%	MM	N	0.40
	MN	MM	3	863	748	289	574	459	10%	MM	9.6%	MM	N	0.39
	MN	MM	1	323	219	79	244	140	25%	MM	24.1%	MM	N	0.34
	MN	MM	1	418	160	92	326	68	50%	MM	57.2%	MM	N	0.23
2	MN	MN	9	2559	2476	887	1672	1589	5%	U	3.4%	U	N	0.45
	MN	MN	3	895	907	360	535	547	10%	MN	9.7%	MN	M	0.49
	MN	MN	1	277	263	97	180	166	25%	MN	24.2%	MN	N	0.42
	MN	MN	1	273	295	101	172	194	50%	MN	58.0%	MN	M	0.44

^a M=具有CTTT缺失, N=没有CTTT缺失^b 用预期的%来进行SPRT^c 用数字ZFY/X测定确定的胎儿%来进行SPRT, U: 未分类的

个例	板的数目	母体 基因型 ^a	胎儿 基因型 ^a	预期的 胎儿%	M	N	全部 阳性	净M	净N	实际 胎儿%	参考等 位基因	m _t	SPRT ^c
1	1	MN	MN	20%	203	200	58	145	142	29.0	N	0.30	MN
2	1	MN	NN	20%	149	204	39	110	165	29.7	M	0.22	NN
3	1	MN	NN	20%	128	212	31	97	181	30.7	M	0.18	NN
4	1	MN	MN	20%	173	196	45	128	151	27.0	M	0.26	MN
5	1	MN	NN	20%	147	230	47	100	183	24.3	M	0.21	NN
6	1	MN	MN	20%	199	206	45	154	161	19.2	M	0.30	MN
7	1	MN	NN	20%	165	237	58	107	179	18.7	M	0.24	NN
8	1	MN	MN	20%	203	180	43	160	137	23.4	N	0.27	MN
9	1	MN	NN	20%	173	241	61	112	180	16.9	M	0.26	NN
10	1	MN	MN	20%	174	190	52	122	138	16.8	M	0.26	U
11	1	MN	NN	20%	167	242	49	118	193	17.7	M	0.25	NN
12	1	MN	MN	20%	213	199	47	166	152	21.6	N	0.30	MN

^a M=具有CTTT缺失, N=没有CTTT缺失^b 通过数字ZFY/X测定来确定实际的胎儿DNA%^c 用通过数字ZFY/X测定确定的胎儿%来进行SPRT, U: 未分类的

图 26B

图 26A

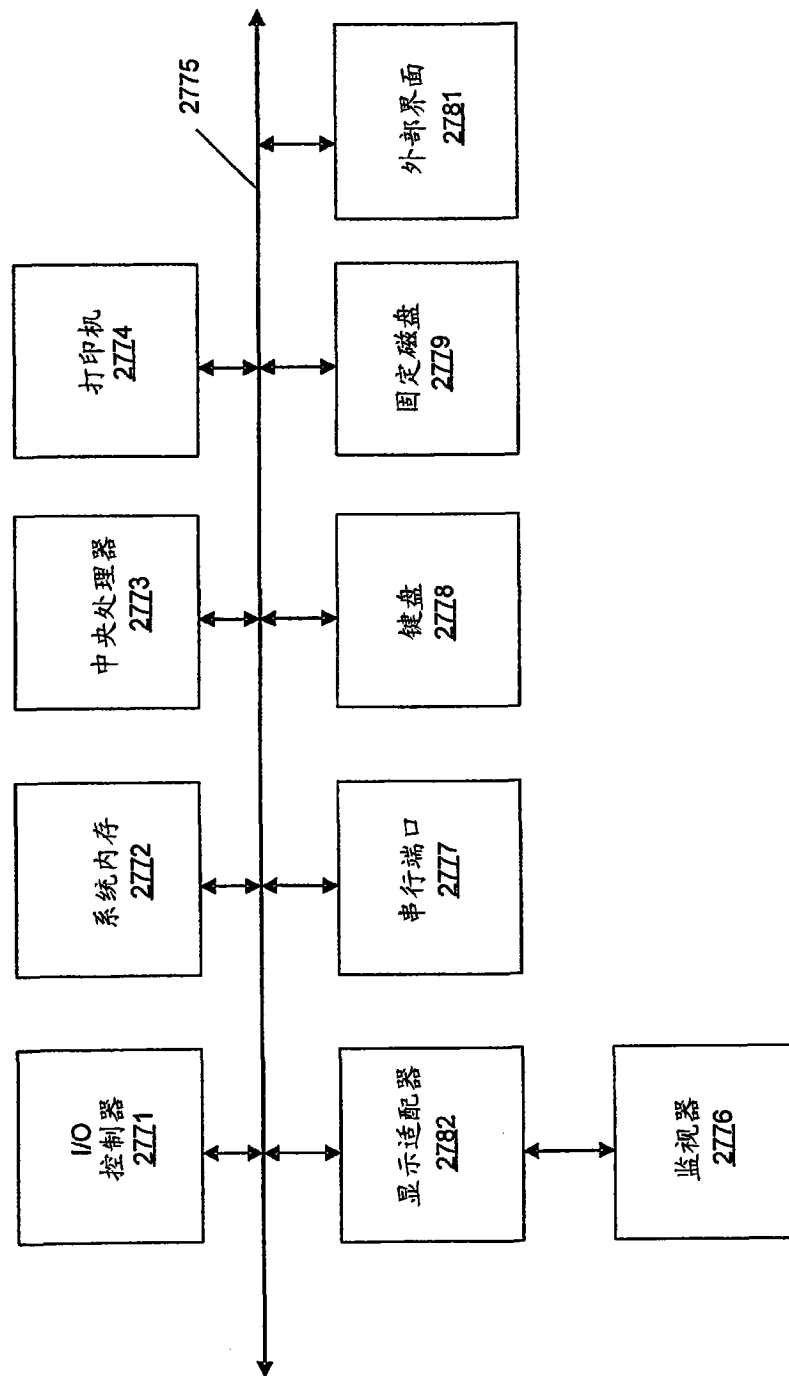


图 27