



(12) 发明专利申请

(10) 申请公布号 CN 104350158 A

(43) 申请公布日 2015.02.11

---

(21) 申请号 201380026625.9 (51) Int. Cl.  
(22) 申请日 2013.03.22 *C12Q 1/68* (2006.01)  
(30) 优先权数据 *C12N 15/11* (2006.01)  
61/615, 535 2012.03.26 US *G01N 33/53* (2006.01)  
61/659, 695 2012.06.14 US *G01N 33/48* (2006.01)

(85) PCT国际申请进入国家阶段日  
2014.11.21

(86) PCT国际申请的申请数据  
PCT/US2013/033451 2013.03.22

(87) PCT国际申请的公布数据  
W02013/148496 EN 2013.10.03

(71) 申请人 约翰霍普金斯大学  
地址 美国马里兰州

(72) 发明人 B·沃戈尔斯滕 K·W·金泽勒  
N·派帕多普勒斯 伊萨克·G·金德

(74) 专利代理机构 北京安信方达知识产权代理  
有限公司 11262  
代理人 王思琪 郑霞

权利要求书2页 说明书10页  
序列表4页 附图10页

---

(54) 发明名称

快速非整倍性检测

(57) 摘要

最近证明,对无细胞的母体血浆 DNA 的大规模并行测序是胎儿染色体非整倍性的安全有效的筛选方法。在这里,我们报告了一种通过以采用单引物对的 PCR 替换费力的测序文库制备步骤,实现显著提高通量和降低成本的改进的测序方法。使用这种方法,可容易区分含有少至 4% 的 21 三体 DNA 的样品与整倍体样品。

1. 一种检测人类非整倍性的方法,所述方法包括:

将来自人类的 DNA 样品中的多个染色体序列以与所述染色体序列互补的单引物对扩增以形成多个扩增子,其中所述多个扩增子不相同,并且其中所述多个扩增子包括查询染色体上的序列和多个参考染色体上的序列;

进行反应以确定所述多个扩增子的至少 3nt 的核苷酸序列;

经电脑模拟,将扩增子核苷酸序列与基因组基因座上的基因组序列匹配;

计数单独基因组基因座上匹配的扩增子数目;

比较匹配到所述查询染色体上的基因组基因座的扩增子数目与匹配到所述参考染色体上的基因组基因座的扩增子数目。

2. 如权利要求 1 所述的方法,其中所述查询染色体是染色体 21。

3. 如权利要求 1 所述的方法,其中所述查询染色体是染色体 13。

4. 如权利要求 1 所述的方法,其中所述查询染色体是染色体 18。

5. 如权利要求 1 所述的方法,其中所述查询染色体是 X 染色体。

6. 如权利要求 1 所述的方法,其中所述查询染色体是 Y 染色体。

7. 如权利要求 2 所述的方法,其中所述扩增子的至少一个与染色体 21 上唐氏综合征关键区域内的基因组序列匹配。

8. 如权利要求 2 所述的方法,其中所述扩增子的至少三个与染色体 21 上的基因组序列相匹配。

9. 如权利要求 1 所述的方法,其中所述扩增子是 180 个碱基对或更少。

10. 如权利要求 1 所述的方法,其中所述引物对包括第一引物和第二引物,其中所述第一引物包含 SEQ ID NO:1 且所述第二引物包含 SEQ ID NO:2。

11. 如权利要求 1 所述的方法,其中所述 DNA 样品来自妊娠女性并且所述 DNA 样品包括母体 DNA 和胎儿 DNA。

12. 如权利要求 11 所述的方法,其中所述 DNA 样品来自血浆。

13. 如权利要求 11 所述的方法,其中所述 DNA 样品来自血清。

14. 如权利要求 1 所述的方法,其中所述 DNA 样品来自血浆。

15. 如权利要求 1 所述的方法,其中所述 DNA 样品来自血清。

16. 如权利要求 11 所述的方法,其中计数足够的扩增子以检测胎儿非整倍性是否存在。

17. 如权利要求 1 所述的方法,其中确定所述多个扩增子的 4 至 30 之间含 4 和 30 的任何整数 nt 的核苷酸序列。

18. 如权利要求 1 所述的方法,其中所述查询染色体与所述参考染色体是细胞核的。

19. 一种检测人类 DNA 拷贝数变化的方法,包括:

将来自人类的 DNA 样品中的多个染色体序列以与所述染色体序列互补的单引物对扩增以形成多个扩增子,其中所述多个扩增子不相同,并且其中所述多个扩增子包括查询染色体区域中的序列和多个参考染色体区域中的序列;

进行反应以确定所述多个扩增子的至少 3nt 的核苷酸序列;

经电脑模拟,将扩增子的核苷酸序列与基因组基因座上的基因组序列匹配;

计数单独基因组基因座上匹配的扩增子数目;

比较匹配到所述查询染色体区域中的基因组基因座的扩增子数目与匹配到所述参考染色体区域中的基因组基因座的扩增子数目。

20. 如权利要求 19 所述的方法,其中所述查询染色体区域在染色体 21 上的唐氏综合征关键区域内。

21. 如权利要求 19 所述的方法,其中所述查询染色体区域是染色体 21。

22. 如权利要求 19 所述的方法,其中所述查询染色体区域是染色体 13。

23. 如权利要求 19 所述的方法,其中所述查询染色体区域是染色体 18。

24. 如权利要求 19 所述的方法,其中所述查询染色体区域是 X 染色体。

25. 如权利要求 19 所述的方法,其中所述查询染色体区域是 Y 染色体。

26. 如权利要求 19 所述的方法,其中所述扩增子是 180 个碱基对或更少。

27. 如权利要求 19 所述的方法,其中所述引物对包括第一引物和第二引物,其中所述第一引物包含 SEQ ID NO :1 且所述第二引物包含 SEQ ID NO :2。

28. 如权利要求 19 所述的方法,其中所述 DNA 样品来自妊娠女性并且所述 DNA 样品包含母体 DNA 和胎儿 DNA。

29. 如权利要求 28 所述的方法,其中所述 DNA 样品来自血浆。

30. 如权利要求 28 所述的方法,其中所述 DNA 样品来自血清。

31. 如权利要求 19 所述的方法,其中所述 DNA 样品来自血浆。

32. 如权利要求 19 所述的方法,其中所述 DNA 样品来自血清。

33. 如权利要求 19 所述的方法,其中计数足够的扩增子以检测胎儿的重复或缺失是否存在。

34. 如权利要求 19 所述的方法,其中所述查询染色体区域与所述参考染色体区域是细胞核的。

35. 如权利要求 19 所述的方法,其中确定所述多个扩增子的 4 至 30 之间含 4 和 30 的任何整数 nt 的核苷酸序列。

36. 可用于分析人类非整倍性的引物对,其中第一引物包含 SEQ ID NO :1 且第二引物包含 SEQ ID NO :2。

37. 一种选择用于检测人类非整倍性的引物的方法,包括:

鉴定与多个染色体区段对互补的引物对,所述多个染色体区段对在人类基因组中被 3 至 10,000nt 的间插序列间隔,其中所述间插序列分布在至少三条人类染色体上;和

扩增所述染色体区段对及其间插序列以形成扩增子分子的群体,所述扩增子分子在群体中的代表性反映染色体区段及其间插序列在人类基因组中的代表性。

38. 如权利要求 37 所述的方法,其中被鉴定的所述引物对分布在 4 和 23 之间含 4 和 23 任何整数的人类染色体上。

39. 如权利要求 37 所述的方法,其中所述染色体区段被 3 至 18nt 的间插序列间隔。

## 快速非整倍性检测

[0001] 本发明在美国国家卫生研究院 (U. S. National Institutes of Health) 项目 CA62924、CA43460 和 CA57345 资助下完成。因此美国政府保留本发明中一定的权利。

[0002] 发明的技术领域

[0003] 本发明涉及遗传分析领域。特别是,它涉及对染色体序列的相对拷贝数的评价。

[0004] 发明背景

[0005] 在大约每 140 个活产新生儿中就有 1 例可检测到主要的染色体异常<sup>1</sup>,并在没有达到足月或死产胎儿中比例更高的多<sup>2</sup>。最常见的非整倍性是 21 三体,目前每 730 个新生儿中就有 1 例<sup>1</sup>。虽然比 21 三体少见,18 三体(爱德华兹综合征)和 13 三体(Patau 综合征)分别以 5,500 分之一和 17,200 分之一发生在新生儿中<sup>1</sup>。种类繁多的先天性缺陷、发育缺陷和智力残疾被发现于染色体非整倍性儿童中,且这些为家庭和社会带来终生挑战<sup>3</sup>。由于这些原因,许多努力致力于在胎儿生命早期检测染色体异常,此时治疗性流产可作为提供给准父母的选项。

[0006] 有各种各样的产前诊断检查可以指示增加的胎儿非整倍性的风险,虽然侵入性试验,诸如羊膜穿刺术或绒毛膜取样是目前的黄金标准<sup>4</sup>,并与不可忽视的胎儿流产的风险相关。因此,长期以来一直在探索更可靠的、非侵入性的胎儿非整倍性检查。其中最具有前途的是基于检测母体血浆中的胎儿 DNA,如 Lo 课题组的先驱工作<sup>5</sup>。已经证明,对由母体血浆产生的文库进行大规模并行测序能够可靠地检测 21 号染色体异常<sup>6,7</sup>。在迄今为止最全面的研究中<sup>8</sup>,98.6% 的 21 三体胎儿可以在母体血浆中被检测到,其假阳性率为 0.2%。在另外的 0.8% 样品中,该检查不能给出结果。这些令人振奋的研究有望进入非侵入性产前检查的新时代。

[0007] 目前,几乎一半 21 三体的患儿的母亲不到 35 岁,因为超过 80% 的孕妇都在 35 岁以下<sup>9,10</sup>。虽然对于合格的年轻母亲来说,侵入性操作的风险被认为超过了侵入性检查带来的好处,很显然,通过可安全施用于所有孕妇的可靠的非侵入性检查可以安全地预防绝大多数与染色体非整倍性相关的新生儿。产前检查对孕妇和家属是非常紧张的事情,且过程越快越好。

[0008] 为了通过循环胎儿 DNA 检查实现这一目标,降低成本并提高通量将是必要的。血浆 DNA 检查有三个主要部分:制备用于加载到测序仪器上的 DNA 文库,对这些文库测序,以及它们的分析。第二部分正在由仪器制造商解决,他们在过去的几年中取得了显著的进步。当前研究的主题是第一和第三部分的潜在改进。

[0009] 唯一商业化的用于循环胎儿 DNA 非整倍性的检查包括制备全基因组文库并分析有关染色体上的足够数量的序列,以可靠地检测拷贝数的细小差异。制备全基因组文库包括几个连续步骤,包括末端修复、5'-磷酸化、添加末端 dA 核苷酸到片段的 3' 末端、片段连接到接头上和 PCR 扩增连接产物,其中许多需要介入纯化。然后将 PCR 产物定量并加载到测序仪上。测序运行结束后,将标签与人类基因组比对并用数字核型评估<sup>11</sup>,即,使用每个基因组基因座的标签数目来构建虚拟核型。另一个最近描述的检查涉及较少的、但仍然大量的步骤来制备测序文库<sup>12</sup>。

[0010] 在本领域中,对迅速和非侵入性地检测遗传异常存在持续的需求。

[0011] 发明概述

[0012] 根据本发明的一个实施方案,提供了一种用于检查人类非整倍性的方法。来自人类 DNA 样品中的多个染色体序列用与所述染色体序列互补的单引物对扩增形成多个扩增子,其中多个扩增子是不相同的,并且其中多个扩增子包含查询染色体上的序列和多个参考染色体上的序列。进行反应以确定多个扩增子的至少 3nt 的核苷酸序列。经电脑模拟将扩增子核苷酸序列与基因组基因座的基因组序列进行匹配。计数匹配到单个基因组基因座的扩增子的数量。比较匹配到查询染色体的基因组基因座的扩增子数量与匹配到参考染色体的基因组基因座的扩增子数量。

[0013] 根据另一个实施方案,提供一种用于检查人类 DNA 拷贝数变化的方法。来自人类 DNA 样品中的多个染色体序列用与所述染色体序列互补的单引物对扩增形成多个扩增子,其中所述多个扩增子是不相同的,并且其中所述多个扩增子包含查询染色体区域中的序列和多个参考染色体区域中的序列。进行反应以确定多个扩增子的至少 3nt 的核苷酸序列。经电脑模拟将扩增子核苷酸序列与基因组基因座的基因组序列进行匹配。计数匹配到单个基因组基因座的扩增子的数量。比较匹配到查询染色体区域中的基因组基因座的扩增子数量与匹配到参考染色体区域中的基因组基因座的扩增子数量。

[0014] 提供了对于分析人类非整倍性有用的引物对。第一引物包含 SEQ ID NO :1 且第二引物包含 SEQ ID NO :2。

[0015] 提供了一种选择用于检测人类非整倍性的引物的方法。鉴定其中引物与人类基因组中被 3-10,000nt 的间插序列间隔的多个染色体区段对互补的引物对。间插序列分布在至少三条人类染色体上。扩增染色体区段的对及其间插序列,以形成扩增子分子的群体,扩增子分子在群体中的代表性 (representation) 反映了染色体片段和它们的间插序列在人类基因组中的代表性。

[0016] 在阅读了本说明书后,对本领域技术人员将是明显的这些和其它实施方案为本领域提供灵敏并快速评估基因组拷贝数的新工具和方法。

[0017] 附图简述

[0018] 图 1a-1d :观察的和预测的 Fast-Seq 扩增产物分布的比较。(图 1a) 预期片段长度分布密度图,在 124 和 142bp 处有峰。(图 1b) 从 8 个正常血浆 DNA 中获得的实际标签计数的密度图。与 142bp 片段相比,124bp 片段被优先扩增。插图 :代表性 FAST-SeqS 测序文库的聚丙烯酰胺凝胶。注意扩增产物含有另外的 121bp 侧翼序列以辅助测序 (补充表 3)。(图 1c) 8 个正常血浆样品中重复 DNA 内的 L1 逆转录转座子的平均代表性。大致 97% 的标签比对到代表仅 7 个 L1 逆转录转座子位置上。(图 1d) 观察到的 8 个正常血浆 DNA 中的分布与 RepeatMasker 所预测的 7 个 L1 逆转录转座子中的每一个的分布相比的详细检查。每个图中的误差棒描绘范围。

[0019] 图 2a-2c :Fast-SeqS 的重复性的证明。(图 2a) 对 8 个正常血浆 DNA 样品的每条常染色体计算的 z 值。没有染色体的 z 值 >3.0 (范围 : -2.1 至 1.9)。(图 2b) 比较 21 三体 (n = 4)、18 三体 (n = 2) 和 13 三体 (n = 1) 患者与 8 个正常脾或 WBC DNA 的 z 值。显示的 z 值表示用于比较的相关染色体。对任何比较的正常染色体观察到的最大 z 值为 1.9 (chr13)。(图 2c) 分析单独的对照 WBC DNA (z 值范围 : -0.8 至 1.3),或与 4% (z 值范

围:4.5至7.2)或8%(z值范围:8.9至10)水平的来自21三体患者的DNA混合的对照WBC DNA。(图2c)中的每个实验重复4次进行。

[0020] 图3(补充图1)使用不同的仪器、样品和测序深度,对FAST-SeqS重复性的进一步证明。(图3a)在图2a中血浆DNA已被测序的8个样品的匹配的外周血白细胞(WBC)DNA在Illumina HiSeq 2000的四分之一通道上测序。标绘了对每个常染色体计算的z值。没有染色体的z值>3.0(范围:-2.2至1.9)。(图3b)8个脾或WBC DNA样品用Illumina GA IIx的二分之一通道进行测序,设计比前面提及的血浆和WBC样品(图2a和(a))产生更少的标签。显示了对每个常染色体计算的z值。尽管测序少3倍,没有染色体的z值>3.0(范围:-2.2至2.1)。

[0021] 图4(补充图2)21三体和整倍体DNA的预混合实验。分析单独的对照外周血白细胞(WBC)DNA( $n = 2$ ),或与5%( $n = 2$ )、10%( $n = 1$ )、或25%( $n = 1$ )水平的21三体患者的DNA混合的对照外周血白细胞(WBC)DNA。预期的和观察的额外的21号染色体比例之间存在紧密的相关性(根据Pearson相关性检验 $r = 0.997$ ,  $n = 6$ )。

[0022] 图5。(补充表1。)本FAST-SeqS研究中分析的样品。

[0023] 图6A-6B。(补充表2)FAST-SeqS的测序特征。

[0024] 图7。(补充表3。)用于制备和测序FAST-SeqS样品的寡核苷酸(SEQ ID NO: 4-14连续地)。

[0025] 图8。(表1)本FAST-SeqS研究中分析的样品。

[0026] 发明详述

[0027] 本发明人开发了一种快速和非侵入性地检测遗传异常,特别是拷贝数异常的方法。我们推测,如果来自遍及基因组的确定数量的片段可以使用单引物对进行扩增,而无需末端修复、末端3'-dA添加、或连接到接头上,那么商业化可获得的该检查过程可以被简化。此外,对较小数目片段进行评估(与全基因组相比)将简化基因组匹配和分析过程。开发的方法能够在预实验中以可重复的方式检测三体性。它在易于实现、成本、分析时间和通量方面优于无偏好的全基因组测序。

[0028] 我们实现这些目标的方法是基于使用退火到散布在整个基因组的人类序列子集的特定引物。我们将这种方法称作“快速非整倍性筛选检测-测序系统(Fast Aneuploidy Screening Test-Sequencing System)”(以下简称FAST-SeqS)。为了获得最大的效用,我们试图鉴定出具有足够相似性的区域,使它们能被单引物对扩增,但又足够独特,允许区分大多数扩增的基因座。为了与见于血浆中的降解DNA相容<sup>8</sup>,我们进一步要求扩增的序列 $\leq 150$ bp。使用BLAST-Like Alignment Tool(BLAT)算法<sup>13</sup>,我们迭代搜索21号染色体上含有唐氏综合征关键区域的一小部分的片段( $\sim 6.8$ Mb)以鉴定合适的引物对,即,将扩增来自遍及基因组以及遍及唐氏综合征关键区域的DNA的许多不同片段的引物对。鉴定了3个这样的引物对,并且经电脑模拟测试(使用电脑模拟PCR<sup>14</sup>)这些引物后以及在预测序实验中,我们发现三个引物对之一(以下FAST-1)是最佳的(Online Methods)。FAST-1引物对预计扩增长的散布的核苷酸元件(L1逆转录转座子)的亚家族,其如同其他人类重复序列,已经通过逆转录转座扩散到整个基因组,特别是AT-rich区域<sup>15</sup>。由于均一地扩增和测序其GC含量差异很大的区域通常更困难<sup>8,16</sup>,我们期望这一差异定位将有利于我们地工作。

[0029] 任何染色体或染色体区域可以被查询。有用的查询染色体和区域的非限制性实例

是 21 号染色体、13 号染色体、18 号染色体、X 染色体、Y 染色体、21 号染色体上唐氏综合征关键区域。查询和参考染色体可以是细胞核的。在一些实施方案中,至少两个、三个、四个或五个检测的扩增子匹配查询染色体或查询区域内的基因组序列。

[0030] 扩增子可以是相对小的,使得一些样品的降解性质不妨碍检测。扩增子可以小于 1kbp、小于 500bp、小于 250bp、小于 180bp、小于 150bp、小于 120bp、小于 100bp、或小于 50bp。在一个实施方案中,引物对包括第一和第二引物,其中第一引物包含 SEQ ID NO :1 和第二引物包含 SEQ ID NO :2。

[0031] DNA 样品可以来自任何来源,但优选地来自哺乳动物,更优选地来自人类。在某些实施方案中,样品来自妊娠女性,且 DNA 样品包含母体和胎儿 DNA。所述 DNA 样品可以例如来自血浆或血清。其它体液如唾液、泪液、汗液、淋巴、尿液,都可以使用。组织样品可以被用作 DNA 来源。脐带血可以用作 DNA 来源。由于胎儿 DNA 是母体血浆 DNA 的仅一小部分,母体 DNA 可掩盖胎儿畸形。因此必须计数足够数量的扩增子以检测胎儿非整倍性是否存在。

[0032] 虽然扩增子可能较大,不必对整个扩增子进行测序以鉴定扩增子。在一些实施方案中,确定多个扩增子的少至 4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、或 30nt 的核苷酸序列。如果需要,更多的可以被确定。

[0033] 有用的引物对可以与分布在至少 4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、或 23 个人类染色体上的序列互补。互补的序列可以由至少 3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18 或更多个 nt 间隔。

[0034] 上述公开内容总体描述了本发明。本文所公开的所有参考文献以引用的方式明确地并入。更完整的理解可以通过参考以下具体实施例来获得,本文中提供以下具体实施例的目的仅在于说明,而不是为了限制本发明的范围。

[0035] 实施例 1

[0036] 方法

[0037] 引物选择

[0038] 我们开始在 21 号染色体上含有唐氏综合征关键区域的~ 6.8MB 区域 (hg19<sup>17</sup> 坐标 35,888,786 至 42,648,523) 搜索~ 150bp 的序列模块,其与所有染色体上存在的那些相似但不完全相同。使用 150bp 滑动窗口、50bp 递增 (135,186 条长 150bp 的序列),我们用 BLAST-Like Alignment Tool (BLAT) 算法<sup>13</sup> 查询序列以分离这样的模块。除了一个如上所述在~ 6.8MB 区域内,我们还要求至少三个模块存在于 21 号染色体上。

[0039] 在查询的 135,186 个模块中,我们仅找到 56 个符合以下标准:

- [0040] • 包含至少 11 个与查询序列不同的碱基,以辅助区分扩增的基因座;
- [0041] • 包含不超过 30 个与查询序列不同的碱基,以提高均一扩增的几率;和
- [0042] • 跨度不超过总共 180 个碱基,以与对降解的 DNA 的分析相容。

[0043] 然后,我们手动扫描这 56 个模块的 BLAT 比对结果,以寻找具有高度相似的 5' 和 3' 末端的那些。56 个序列中的至少 3 个符合我们的标准,我们为它们设计了引物。电脑模拟 PCR<sup>14</sup> 验证,每个理论引物对将从每一条核染色体扩增许多不同的序列。

[0044] 扩增产物的唯一性的确认

[0045] 由于文库制备或测序过程中引入的不可避免的错误,过度相似的序列在比对时可能会出现。因此,我们写了一个自定义脚本,以评估在允许每~ 150bp 序列中的一个、

两个或三个错误后,多少不同的序列将被保留。一个引物对 (FAST-1) 的理论扩增产物大大优于其它两个,且 FAST-1 的优越性在预测序实验中得到证实。电脑模拟 PCR<sup>14</sup> 预测 PCR 片段双峰分布,这由扩增的 PCR 产物在聚丙烯酰胺凝胶上的大小分离,和通过观察到的每个位置的计数的分布 (图 1b) 证实。

#### [0046] 模板

[0047] 从正常脾、外周血白细胞 (WBC) 或血浆 (补充表 1) 中得到对照 DNA。来自 5 个 21 三体 (NA02767、NA04616、NG05120、NG05397 和 NG07438)、2 个 18 三体 (NA03623 和 NG12614)、和 1 个 13 三体 (NA03330) 个体的成纤维细胞 DNA 购自 the Coriell Institute for Medical Research (Camden, New Jersey)。总共 33ng DNA 被用于每个实验。所有的模板由 OD 定量,除了混合实验,其中模板由数字 PCR<sup>20</sup> 定量以达到更精确的浓度估计。

#### [0048] 测序文库制备

[0049] FAST-SeqS 显著节省时间是由于用使用单一引物对的扩增替换了传统的全基因组扩增文库制备。使用 FAST-1 引物扩增模板,按 Kinde 等<sup>21</sup> 所描述的,其中单独模板分子被用唯一的标识符 DNA 序列加标签。虽然该唯一的标识符序列被证明对于 FAST-SeqS 是没有必要的 (见下面的准确模板计数部分),我们把它们包含在最初的实验设计中,并一直使用它们,只要它们被观察到在我们最初的实验中提供牢靠的 PCR 产物。简言之,FAST-1 扩增引物各包含不同的 5' 通用引物序列 (UPS),随后是允许扩增上述重复元件的序列 (正向: ACACAGGGAGGGGAACAT; SEQ ID NO:1; 反向: TGCCATGGTGGTTTGCT; SEQ ID NO:2) (补充表 3)。此外,该正向引物中包含紧随其 UPS 的 3' 的一段 16-20 简并碱基 (补充表 3)。使用 Phusion Hot Start II Polymerase (Thermo Scientific, 目录号 F-549L),在 1X Phusion HF 缓冲液、0.5 μM 每种引物、和 2 单位聚合酶中在共 50 μL,在以下循环条件下进行 PCR: 98°C 120s, 随后 2 个循环的 98°C 10s、57°C 120s、和 72°C 120s。最初的扩增引物用 AMPure XP 珠 (Beckman Coulter Genomics, 目录号 A63880) 去除,根据制造商,除了仅加入 1.4 倍 PCR 体积的珠子和洗脱体积减少到 10 μL TE。洗脱液直接用于第二轮扩增,使用退火到经第一轮引物引入的 UPS 位点、且另外含有杂交到 Illumina 流动池所必需的 5' 接枝序列的引物 (补充表 3)。进一步,我们通过反向引物引入 5 个索引中的一个 (“条形码”) (补充表 3) 到每一个样品中,以后续允许多重测序。第二轮 PCR 使用 Phusion Hot Start II Polymerase,在 1X Phusion HF 缓冲液、0.5 μM 每种引物、和 2 个单位聚合酶中在共 50 μL 中,在下述循环条件下进行: 98°C 120s,接着是 13 个循环的 98°C 10s、65°C 15s、和 72°C 15s。扩增产物再次使用 AMPure XP 珠纯化,并通过分光光度法、实时 PCR 或在 Agilent 2100 Bioanalyzer 上进行定量;所有定量方法得到类似的结果。寡核苷酸购自 IDT (Coralville, Iowa)。

#### [0050] 数据分析概述

[0051] 相对于传统的全基因组扩增文库,其中的标签绝大多数比对到基因组中唯一的位置,因此每个标签需要独立的比对,FAST-SeqS 产生的序列平均只比对到 21,676 个位置;补充表 2)。与跨所有实验获得的序列数据范围相比,序列比对的位置数量几乎不变。虽然每个实验唯一比对的标签的数量跨越了 12 倍的范围 (1,343,382 至 16,015,347),位置的数量仅变化了 0.25 倍 (范围: 18,484 至 24,562 个位置;补充表 2)。来自所有实验的原始读数 (补充表 2) 可以从域 [sagenet.org](http://sagenet.org),子域 fast,文件 fast.htm 下载。

#### [0052] 序列标签过滤和比对

[0053] 将通过 Illumina 纯洁过滤器且包含扩增引物的至少 3 个正确的末端碱基的标签的 37 碱基序列标签对质量过滤,使用自定义脚本用 N 屏蔽质量分数 <20 的任何碱基。因此,为具有低质量碱基的标签提供机会比对,提供只考虑其最可靠的碱基。质量屏蔽后,只有不同的序列由 Bowtie 0.12.7<sup>18</sup> 比对到人类基因组 (hg19<sup>17</sup>)。当建立 Bowtie 的参考索引时,我们包含了尚未解决的或者未放置的重叠群<sup>22</sup>,以确保最准确的比对。唯一地比对具有最多 1 个错配(分别使用标志 -m 1 和 -v 1)的序列被保留,并且它们的比对与原始数据返回匹配。跨所有样品平均 38% 的标签可以唯一地指派到基因组的位置(范围:31%至 45%;补充表 2)。

[0054] 估算测序片段的分布

[0055] 通过凝胶电泳证实了电脑模拟预测的 FAST-SeqS 扩增产物主要是双峰分布后,我们研究了比对到唯一位置的测序片段的计数是否类似分布。虽然我们只测序了 37 个碱基,我们可以根据电脑模拟 PCR<sup>14</sup> 来估算每个标签的相对大小及其在基因组中的唯一位置。这个练习可以为实际的扩增产物与预测的相匹配的另外的证据提供,并能提醒我们任何扩增偏好(见下文归一化部分)。

[0056] 首先,我们将每个唯一比对的位置的标签计数转换成对数刻度,一种对这一类数据常用的转换以引入对称性<sup>23</sup>。我们对实验的每个组进行这种转换(例如,在同一仪器运行中分析的 8 个正常血浆样品;补充表 2)。接下来,我们使用了非参数方法来估计平滑分布(一种内核密度估计,使用密度函数在  $R^{24}$  中实现),这使我们的数据形式直接可视化。使用 ggplot2<sup>25</sup>(一种  $R^{24}$  软件包)对分布绘图后,我们观察到,各组实验显示出与每个位置的标签计数相似的聚类,与主要的双峰分布以及负偏斜一致。一个代表性的绘图显示在图 1b。

[0057] 归一化

[0058] 由于随机的和实验性的变化,大规模并行测序将从每个样品中,以及同一样品在不同的测序运行时,产生不同数量的序列标签。因此,有必要将数据归一化,以进行本文的类型的有意义的比较。虽然简单地将标签数表示为实验中测序的总标签数的分数是最直接的,这种归一化方法过于简单,并且对于系统性偏差高度敏感,所述系统性偏差经常困扰使用 DNA 和 RNA 模板的新一代测序,且这些常规地用于诸如用于 21 三体的诊断的数字核型分析<sup>8,16</sup>。

[0059] 由于通过 FAST-1 引物对获得的扩增子的双峰尺寸分布,我们预测大多数 FAST-1 扩增偏移是由于小尺寸片段潜在的过度代表(over-representation)。这种偏移可发生在文库制备期间或在 Illumina 流动池上进行的固相桥式 PCR 期间。我们发现可以利用微阵列群落内广泛使用的分位数法<sup>19</sup>对这种分布获得适当的归一化。通过组织我们的数据为位置列表(相当于微阵列数据中的探针),每一个与标签计数(相当于微阵列数据中的强度)相关联,我们可以对 FAST-SeqS 数据适用标准分位数均一化。为了最好地接近微阵列数据格式,我们选择仅分析每个实验组中共有的位置(例如,8 个正常血浆样品的数据)。由于 FAST-1 引物扩增位置的高度可重复组,这通常只消除了 <1% 的数据。为了最大限度地提高可重复性,我们排除了比对到未解决或者未定位的重叠群和那些比对到性染色体的位置,虽然包含这些染色体只轻微提高实验差异(例如,在 8 个正常的血浆样品中,任何染色体的最大 z 值从 1.9 升高至 2.3)。包含性染色体可能对于其他应用有用,例如检测包括 X 染色体的非整倍性或确定样品的性别(即,通过存在或不存在比对到 Y 染色体的序列)。

[0060] 我们通过执行以下步骤,实现对每个实验组的分位数归一化<sup>19</sup>(每一组包含多个样品;补充表 2):

[0061] • 产生代表每个样品的每个位置的标签计数的排序阵列(所有具有同样长度,因为只分析在每个实验中共有的位置);

[0062] • 组合这些排序阵列成为  $2 \times 2$  矩阵,其中每个实验以其自己的列代表,而共有的位置构成行;

[0063] • 在特定行用所有样品的平均计数替换单个样品的计数;和

[0064] • 将计数重排回其最初的顺序。

[0065] 我们的数据分布总是负偏态(参见图 1b 作为一个代表实例)。我们通过以下将落在每次实验分布的左尾的位置(包含最小数量标签的位置)排除在我们的分析以外:

[0066] • 如上所述估算归一化值的分布;

[0067] • 确定双峰分布的两个峰值之间的拐点;和

[0068] • 考虑相对密度低于拐点的位置作为左尾。

[0069] 确定左尾并丢弃其中的位置后,重复分位数归一化。通过这个过程,一个实验组中的每个样品具有相同的总标签数和相同的计数分布,所以可以进行直接比较。我们在  $R^{24}$  中使分位数归一化自动化。整个归一化程序通常只需要不到几分钟就能完成。

[0070] 非整倍性状态的定量确定

[0071] 一种在基于数字核型<sup>11</sup>分析中确定特定样品非整倍性状态的常用方法是通过比较  $z$  值<sup>6,8,26</sup>。通过这种方法,人们确定一组参考样品(例如具有已知的整倍体含量的样品)中落在感兴趣的染色体内的标签计数的均值和标准差,然后为每个样品的感兴趣的染色体创建标准化的值(即  $z$  值),如下: $z$  值  $i, \text{chrN} = (\text{chrN}_i - \mu_{\text{chrN}}) / \text{sd}_{\text{chrN}}$ ,其中  $i$  表示将被标准化的样品,  $\text{chrN}$  表示样品染色体的归一化标签计数,且  $\mu_{\text{chrN}}$  和  $\text{sd}_{\text{chrN}}$  分别代表参考组中  $\text{chrN}$  的归一化标签计数的平均值和标准偏差。当所有样品以这种方式标准化,极端值很容易被检测到,因为它们的  $z$  值  $> 3.0$ 。这表明,极端值的归一化标签计数超过参考组的平均数至少 3 个标准偏差。

[0072] 准确模板计数

[0073] 最后,我们评估准确计数模板分子是否可以进一步提高重复性。通过在两条 FAST-1 引物中的一条的 5' 端掺入 16-20 简并碱基(补充表 3),独特地识别产生 PCR 产物的每个模板分子是可能的<sup>21</sup>。通过最小化每条染色体的最终计数中同样的模板分子被不止一次计数的可能性,这可潜在地提高准确度。我们发现这种强化不会显著改变在 8 个正常血浆样品中对每一条染色体归一化计数的一致性:任何染色体的最大  $z$  值从 1.9 略微增加至 2.0。通过对所有常染色体的  $z$  值的绝对值进行双尾  $t$  检验比较分析方法,我们发现这两种方法之间没有统计上的显著差异( $p = 0.759$ , 每组  $n = 22 \times 8$ )。

[0074] 实施例 2

[0075] 作为 FAST-SeqS 性能的初始测试,我们检查了 7 名正常女性血浆 DNA 中的每一条常染色体的代表,包括一个生物学重复(补充表 1)。在 Illumina HiSeq 2000 仪器的通道的四分之一中使用仅 37 个测序循环,我们回收了每个个体平均 31,547,988 条高质量标签(范围:27,179,424 至 36,0418,017 标签;补充表 2)。当使用 Bowtie<sup>18</sup> 比对到 hg19<sup>17</sup>, 允许最多一个错配时,平均 35% 的这些标签(范围:31 至 37%)可以唯一地映射到平均

23,681 个唯一染色体位置（范围：22,589 至 24,562 位置）中的一个。理论电脑模拟（图 1a）和观察到的标签计数分布（图 1b）均显示大小的双峰分布。在这些唯一比对的标签中，99.1% 比对到被 RepeatMasker (<http://www.repeatmasker.org>) 预测是重复 DNA 的位置，其中 97.5% 落入仅有的 7 个 L1 逆转录转座子亚家族（图 1c）。另外，每个亚家族的分布与 RepeatMasker 的预测一致（图 1d）。由于标签比对到一组离散的染色体位置比比对到整个基因组简单，测序后分析过程非常迅速。事实上，这种映射加上后续的统计分析可以使用一台装有 2 个 6 核 CPU 的计算机 (Intel Xeon X5680) 在不到 30 分钟完成每个样品的分析。

[0076] 最重要的是，归一化以比较不同样品的染色体标签计数后<sup>19</sup>，个体样品间映射到每条染色体的标签的相对比例是明显相似的 (Online Methods)。特别是，在研究的 8 个样品的任一个中，与任何常染色体匹配的标签比例从未偏离平均值，即  $z$  值  $>3.0$  (图 2a)。特别值得注意的是，8 个样品中观察到 21、18 和 13 号染色体的最大  $z$  值分别是 1.3、1.4 和 1.0。

#### [0077] 实施例 3

[0078] 接下来，我们在两个另外实验中使用不同类型的样品、不同的仪器、和不同测序深度，研究了染色体代表的重复性。在第一个实验中，我们分析了来自献血浆的相同 7 个个体的外周血白细胞 (WBC) 的 DNA。四个样品使用 Illumina HiSeq 2000 的单个通道进行测序，产生每个样品平均 10,835,559 条唯一比对的标签（范围：4,905,067 至 16,015,347 标签）。任意样品的 21、18 和 13 号染色体的最大  $z$  值分别为 1.0、1.2 和 1.6 (补充图 1a)。

#### [0079] 实施例 4

[0080] 在下一个实验中，我们分析另外 8 个个体的脾或 WBC DNA，每个样品使用 Illumina GA IIx 仪器的二分之一通道（补充表 2）。我们得到每个样品平均 4,013,951 条唯一比对标签（范围：2,847,661 至 4,645,608 标签）。尽管几乎少 3 倍测序，样品的 21、18 和 13 号染色体的最大  $z$  值仍分别只有 1.3、1.5 和 1.9，远低于广泛使用的截断值 3.0 (补充图 1b)。

[0081] 鉴于图 2a 中明显的标签的紧密分布，我们预计容易区分三体患者和那些具有整倍体染色体体格的正常个体的 DNA。在图 2b 中所示的数据表明，这种期望在 4 名 21 三体患者的每一个中实现了。这些 21 三体患者的  $z$  值范围从 32 到 36，而 8 个正常个体的最大  $z$  值为 1.3。类似地，2 名 18 三体患者和 1 名 13 三体患者的 DNA 的  $z$  值分别是 51、56 和 36，远超过正常个体的这些染色体的最大  $z$  值（图 2b）。

#### [0082] 实施例 5

[0083] 胎儿 DNA 占母体 DNA 的几何平均数为 13.4%，很大程度上取决于母亲的体重，而不是孕周<sup>8</sup>。为了研究 FAST-SeqS 是否能区分含有二体和三体 DNA 的混合物的样品，我们使用来自 21 三体患者和正常个体的 DNA 进行混合实验。在该类型的第一个实验中，我们向正常 WBC DNA 中混入 5% ( $n = 2$ )、10% ( $n = 1$ )、和 25% ( $N = 1$ ) 的 21 三体 DNA 以及两个对照（补充图 2），并发现预期的和观察到的额外的 21 号染色体分数之间的紧密相关性（根据 Pearson 相关性检验  $r = 0.997$ ,  $n = 6$ ）。在第二个实验中，我们评价了含有 4% 或 8% 的 21 三体 DNA 的混合物。如图 2c 所示，含有 4% 的 21 三体 DNA 的样品与来自正常个体的样品之间存在明显区别 ( $p = 2 \times 10^{-4}$ ，根据双尾  $t$  检验确定的，各组  $n = 4$ )。含有 8% 的 21 三体 DNA 的样品当然甚至更容易被区分（与整倍体组相比  $p = 4 \times 10^{-6}$ ，和与 4% 的 21 三体样品相比  $p = 1 \times 10^{-3}$ ，两者都通过双尾  $t$  检验，每组  $n = 4$ ）。

#### [0084] 参考文献

- [0085] 引用的每个参考文献的公开内容清楚地并入本文。
- [0086] 1. LYF Hsu, in *Genetic Disorders and the Fetus*, edited by A Milunsky (The Johns Hopkins University Press, Baltimore, 1998), pp. 179.
- [0087] 2. M. Staebler, C. Donner, N. Van Regemorter et al., *Prenat Diagn* 25(7), 567 (2005).
- [0088] 3. K. L. Jones, *Smith's recognizable patterns of human malformation*, 6ed. (Elsevier Saunders, Philadelphia, 2006).
- [0089] 4. American College of Obstetricians and Gynecologists, *Obstet Gynecol* 110(6), 1459 (2007).
- [0090] 5. Y. M. Lo, M. S. Tein, T. K. Lau et al., *Am J Hum Genet* 62(4), 768 (1998).
- [0091] 6. R. W. Chiu, K. C. Chan, Y. Gao et al., *Proc Natl Acad Sci U S A* 105(51), 20458 (2008).
- [0092] 7. H. C. Fan, Y. J. Blumenfeld, U. Chitkara et al., *Proc Natl Acad Sci U S A* 105(42), 16266 (2008).
- [0093] 8. G. E. Palomaki, E. M. Kloza, G. M. Lambert-Messerlian et al., *Genet Med* 13(11), 913 (2011).
- [0094] 9. J. Cleary-Goldman, F. D. Malone, J. Vidaver et al., *Obstet Gynecol* 105(5Pt 1), 983 (2005).
- [0095] 10. R. G. Resta, *Am J Med Genet A* 133A(1), 31 (2005).
- [0096] 11. T. L. Wang, C. Maierhofer, M. R. Speicher et al., *Proc Natl Acad Sci U S A* 99(25), 16156 (2002).
- [0097] 12. A. B. Sparks, E. T. Wang, C. A. Struble et al., *Prenat Diagn* (2012).
- [0098] 13. W. J. Kent, *Genome Res* 12(4), 656 (2002).
- [0099] 14. R. M. Kuhn, D. Karolchik, A. S. Zweig et al., *Nucleic Acids Res* 35(Database issue), D668 (2007).
- [0100] 15. A. F. Smit, *Curr Opin Genet Dev* 9(6), 657 (1999).
- [0101] 16. H. C. Fan and S. R. Quake, *PLoS One* 5(5), e10439 (2010).
- [0102] 17. P. A. Fujita, B. Rhead, A. S. Zweig et al., *Nucleic Acids Res* 39(Database issue), D876 (2011).
- [0103] 18. B. Langmead, C. Trapnell, M. Pop et al., *Genome Biol* 10(3), R25 (2009).
- [0104] 19. B. M. Bolstad, R. A. Irizarry, M. Astrand et al., *Bioinformatics* 19(2), 185 (2003).
- [0105] 20. B. Vogelstein and K. W. Kinzler, *Proc Natl Acad Sci U S A* 96(16), 9236 (1999).
- [0106] 21. I. Kinde, J. Wu, N. Papadopoulos et al., *Proc Natl Acad Sci U S A* 108(23), 9530 (2011).
- [0107] 22. E. S. Lander, L. M. Linton, B. Birren et al., *Nature* 409(6822), 860 (2001).
- [0108] 23. J. Tukey, *Exploratory Data Analysis*. (Addison-Wesley, Reading, Massachusetts, 1977).

- [0109] 24. R Development Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2011).
- [0110] 25. H. Wickham, ggplot2: elegant graphics for data analysis. (Springer, New York, 2009).
- [0111] 26. M. Ehrlich, C. Deciu, T. Zwiefelhofer et al., Am J Obstet Gynecol 204(3), 205e1 (2011).

[0001]

## 序列表

- <110> B·沃戈尔斯滕  
 K·W·金泽勒  
 N·派帕多普勒斯  
 伊萨克·G·金德
- <120> 快速非整倍性检测
- <130> 001107.00970
- <150> 61/615535  
 <151> 2012-03-26
- <150> 61/659695  
 <151> 2012-06-14
- <160> 14
- <170> FastSEQ for Windows Version 4.0
- <210> 1  
 <211> 18  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物
- <400> 1  
 acacagggag gggaacat 18
- <210> 2  
 <211> 17  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物
- <400> 2  
 tgccatggtg gtttgct 17
- <210> 3  
 <211> 18  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物

[0002]

<400> 3	
acacagggag gggaacat	18
<210> 4	
<211> 55	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸引物	
<221> misc_feature	
<222> (1)... (55)	
<223> n = A、T、C 或 G	
<400> 4	
cgacgtaaaa cgacggccag thnnnnnnnn nnnnnnaca caggagggg aacat	55
<210> 5	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸引物	
<400> 5	
cgacgtaaaa cgacggccag thbvdhbvdh bvdhbvdhvb dacacagga ggggaacat	59
<210> 6	
<211> 42	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸引物	
<400> 6	
cacacaggaa acagctatga ccatgtgcca tgggtgtttg ct	42
<210> 7	
<211> 50	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸引物	
<400> 7	
aatgatacgg cgaccaccga gatetacacc gacgtaaac gacggcagt	50

[0003]

- <210> 8  
 <211> 55  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物
- <400> 8  
 caagcagaag acggcatacg agatgatcag cacacaggaa acagctatga ccatg 55
- <210> 9  
 <211> 55  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物
- <400> 9  
 caagcagaag acggcatacg agatgccaat cacacaggaa acagctatga ccatg 55
- <210> 10  
 <211> 55  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物
- <400> 10  
 caagcagaag acggcatacg agattgacca cacacaggaa acagctatga ccatg 55
- <210> 11  
 <211> 55  
 <212> DNA  
 <213> 人工序列
- <220>  
 <223> 寡核苷酸引物
- <400> 11  
 caagcagaag acggcatacg agatttaggc cacacaggaa acagctatga ccatg 55
- <210> 12  
 <211> 55  
 <212> DNA  
 <213> 人工序列
- <220>

[0004]

<223> 寡核苷酸引物

<400> 12

caagcagaag acggcatac agatcttgta cacacaggaa acagctatga ccatg 55

<210> 13

<211> 32

<212> DNA

<213>人工序列

<220>

<223> 寡核苷酸引物

<400> 13

gagatetaca cegacgtaaa acgacggcca gt 32

<210> 14

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> 寡核苷酸引物

<400> 14

catggteata gctgtttect gtgtg 25

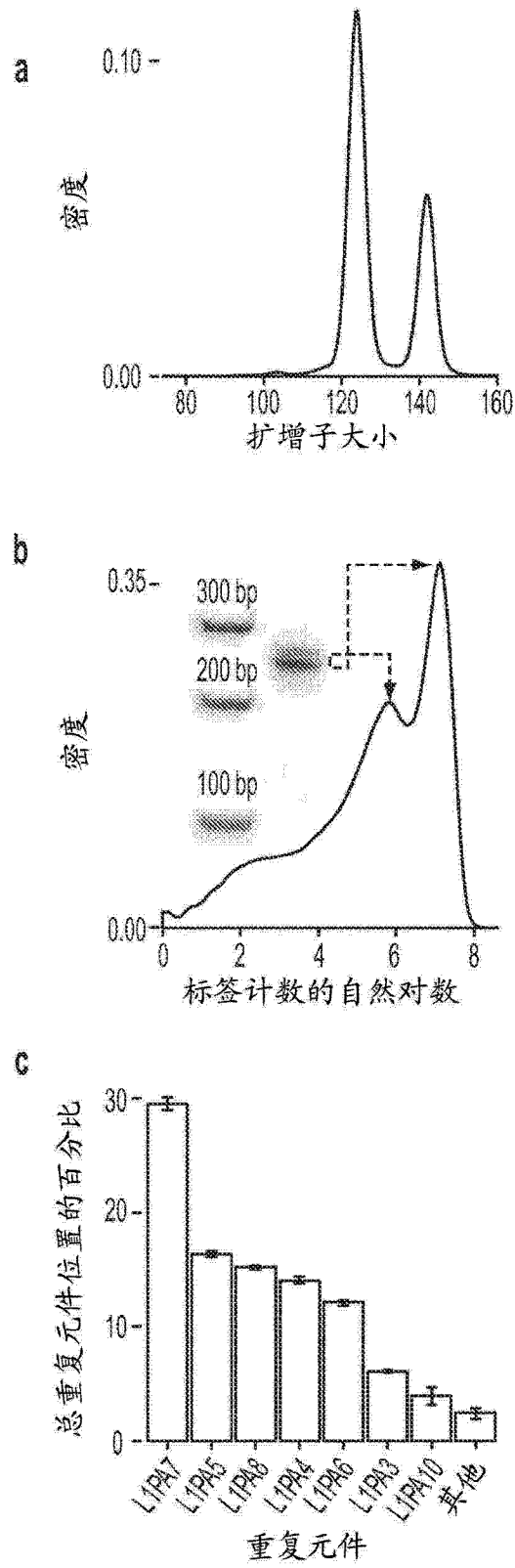


图 1

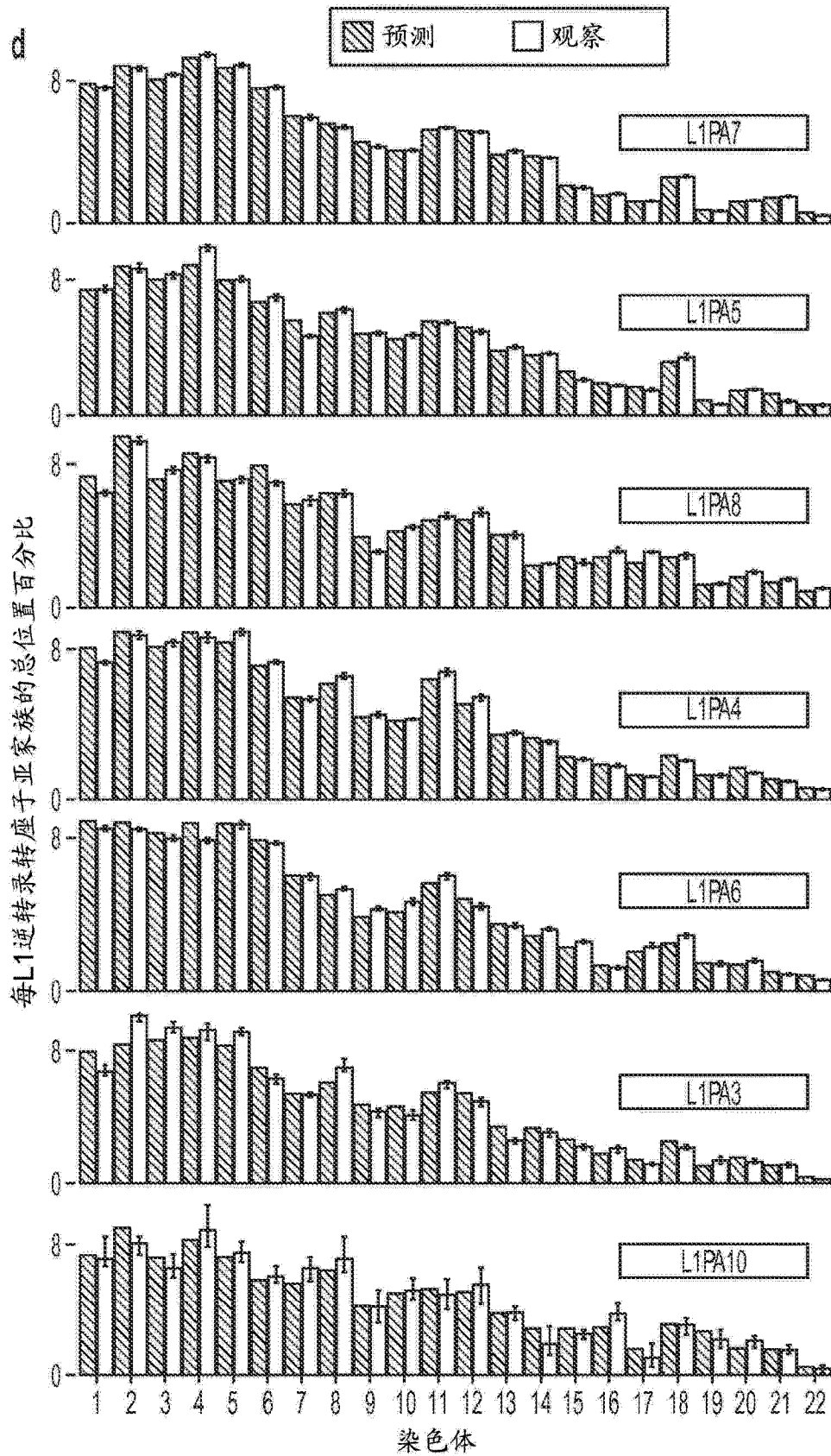


图 1 续

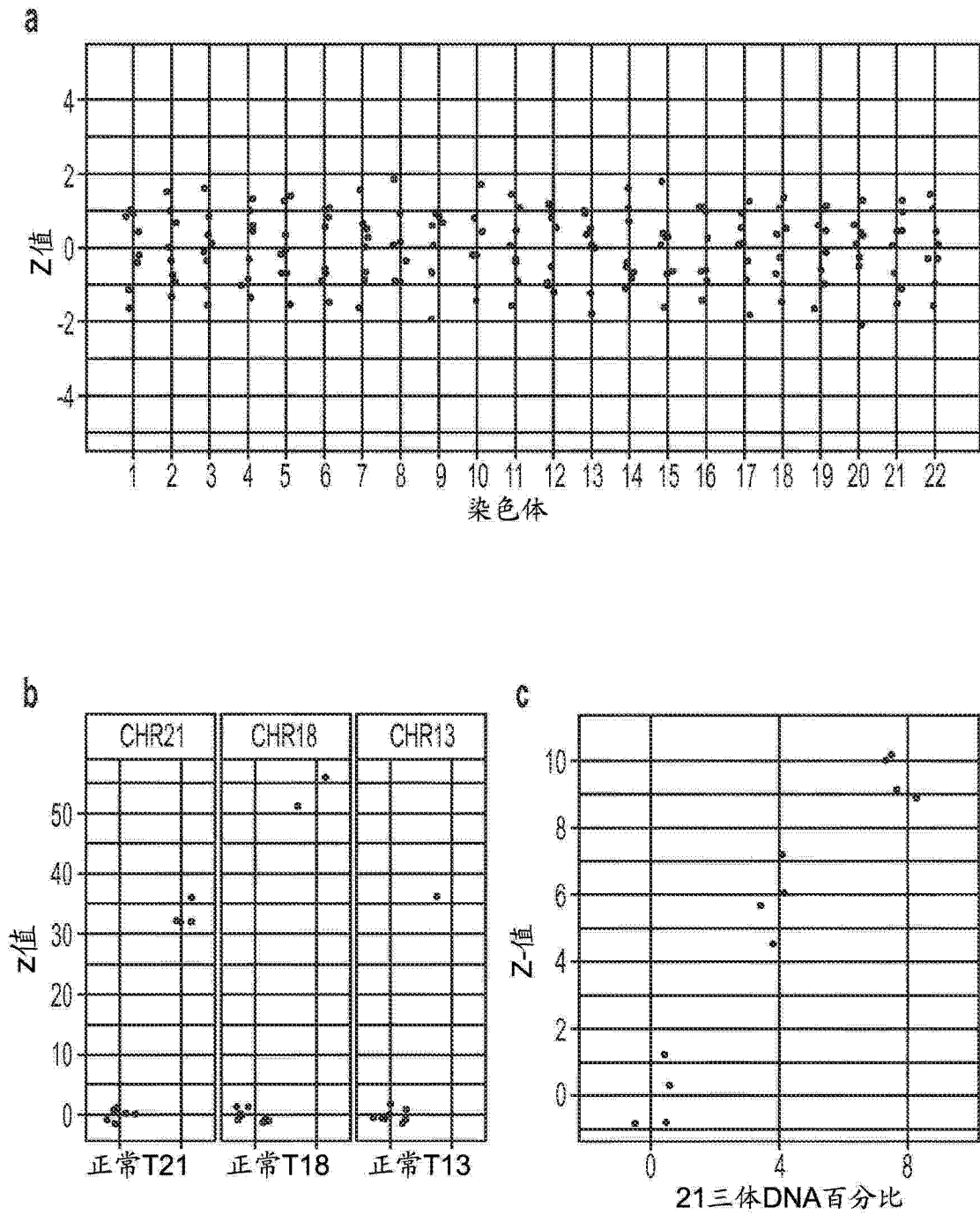


图 2

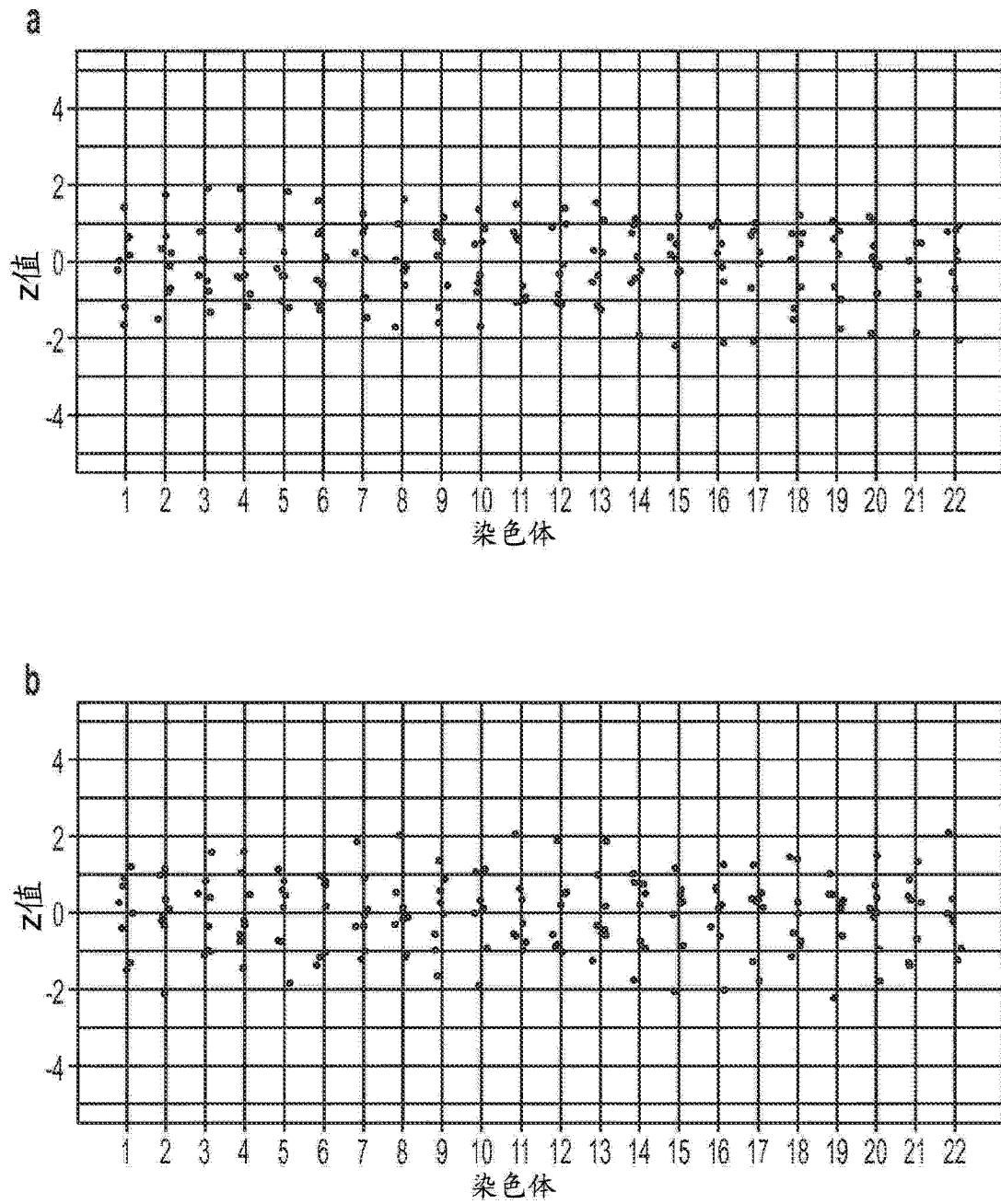


图 3

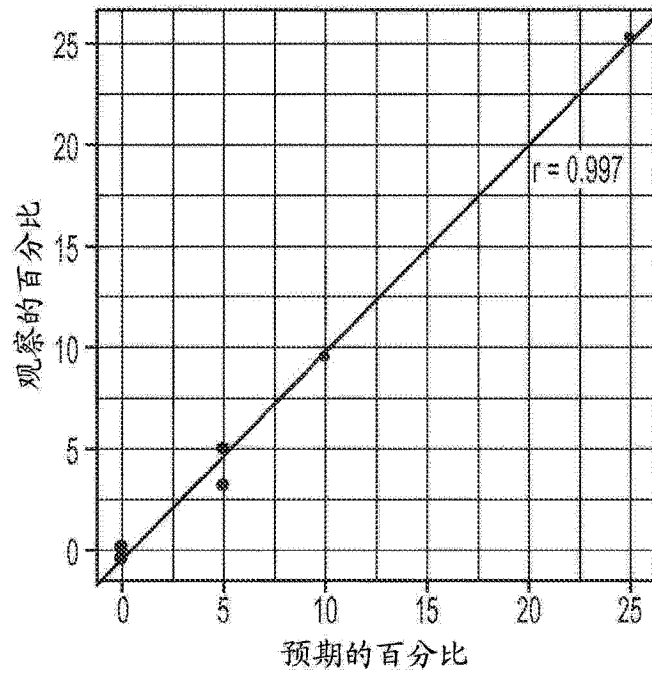


图 4

补充表1:本FAST-SeqS研究中分析的样品

个体	组织	非整倍性	样品名称
n_01	血浆	无	n_01_pls
n_02	血浆	无	n_02_pls
n_03_1	血浆	无	n_03_1_pls
n_03_2	血浆	无	n_03_2_pls
n_04	血浆	无	n_04_pls
n_05	血浆	无	n_05_pls
n_06	血浆	无	n_06_pls
n_07	血浆	无	n_07_pls
n_01	WBC	无	n_01_wbc
n_02	WBC	无	n_02_wbc
n_03_1	WBC	无	n_03_1_wbc
n_03_2	WBC	无	n_03_2_wbc
n_04	WBC	无	n_04_wbc
n_05	WBC	无	n_05_wbc
n_06	WBC	无	n_06_wbc
n_07	WBC	无	n_07_wbc
n_08	脾	无	n_08
n_09	脾	无	n_09
n_10	脾	无	n_10
n_11	脾	无	n_11
n_12	脾	无	n_12
n_13	脾	无	n_13
n_14	脾	无	n_14
n_15	WBC	无	n_15
t21_01	成纤维细胞	21三体 (NA02767)	t21_01
t21_02	成纤维细胞	21三体 (NA04616)	t21_02
t21_03	成纤维细胞	21三体 (NG05120)	t21_03
t21_04	成纤维细胞	21三体 (NG05397)	t21_04
t21_05	成纤维细胞	21三体 (NG07438)	t21_05
t18_01	成纤维细胞	18三体 (NA03623)	t18_01
t18_02	成纤维细胞	18三体 (NG12614)	t18_02
t13_01	成纤维细胞	13三体 (NA03330)	t13_01

图 5

补充表2: FAST-SeqS实验的测序特征

ILLUMINA 仪器	测序的 样品	组	实验 ID	每通道 样品数	通过 纯洁 过滤器的 标签	唯一 比对	唯一 比对 (%)	不同 位置
HiSeq 2000	n_01_pls	1	1	4	27,179,424	8,303,313	31%	22,589
HiSeq 2000	n_02_pls	1	2	4	29,081,731	10,081,137	35%	23,632
HiSeq 2000	n_03_1_pls	1	3	4	30,074,519	10,205,766	34%	23,608
HiSeq 2000	n_03_2_pls	1	4	4	34,968,939	12,626,577	36%	24,562
HiSeq 2000	n_04_pls	1	5	4	30,842,018	10,492,951	34%	24,163
HiSeq 2000	n_05_pls	1	6	4	36,048,017	12,997,372	36%	24,314
HiSeq 2000	n_06_pls	1	7	4	32,670,848	11,536,414	35%	23,145
HiSeq 2000	n_07_pls	1	8	4	31,518,407	11,526,042	37%	23,431
HiSeq 2000	n_01_wbc	2	9	4	35,712,331	11,805,456	33%	22,807
HiSeq 2000	n_02_wbc	2	10	4	13,158,351	4,905,067	37%	20,577
HiSeq 2000	n_03_1_wbc	2	11	4	42,674,992	16,015,347	38%	23,454
HiSeq 2000	n_03_2_wbc	2	12	4	19,063,243	7,556,803	40%	21,133
HiSeq 2000	n_04_wbc	2	13	4	39,854,363	14,576,619	37%	23,285
HiSeq 2000	n_05_wbc	2	14	4	25,058,110	9,607,768	38%	21,642
HiSeq 2000	n_06_wbc	2	15	4	31,447,494	11,781,575	37%	22,255
HiSeq 2000	n_07_wbc	2	16	4	27,035,516	10,435,840	39%	21,828
GA Iix	n_08	3	17	2.00	11,233,449	4,002,305	36%	20,873
GA Iix	n_09	3	18	2.00	10,593,346	4,430,374	42%	20,804
GA Iix	n_10	3	19	2.00	9,015,697	3,909,552	43%	20,392
GA Iix	n_11	3	20	2.00	9,976,099	4,381,161	44%	20,581
GA Iix	n_12	3	21	2.00	9,698,517	4,380,884	45%	20,466
GA Iix	n_13	3	22	2.00	10,432,676	4,645,608	45%	20,737
GA Iix	n_14	3	23	2.00	7,871,040	3,514,064	45%	20,135
GA Iix	n_15	3	24	2.00	8,198,416	2,847,661	35%	20,046
GA Iix	t21_01	3	25	2.00	10,381,557	3,515,217	34%	20,595
GA Iix	t21_02	3	26	2.00	7,939,433	2,911,392	37%	20,163
GA Iix	t21_03	3	27	2.00	5,524,580	1,972,803	36%	19,095
GA Iix	t21_04	3	28	2.00	6,476,900	2,326,116	36%	19,653
GA Iix	t18_01	3	29	2.00	3,716,461	1,343,382	36%	18,484
GA Iix	t18_02	3	30	2.00	6,197,319	2,384,437	38%	19,543
GA Iix	t13_01	3	31	2.00	7,373,315	2,598,029	35%	19,809
GA Iix	100% n_15	4	32	2.00	9,905,447	3,964,677	40%	20,612
GA Iix	100% n_15	4	33	2.00	10,540,366	3,412,883	32%	20,674
GA Iix	95% n_15 + 5% t21_05	4	34	2.00	10,618,558	4,243,752	40%	21,136
GA Iix	95% n_15 + 5% t21_05	4	35	2.00	10,713,550	4,261,765	40%	21,095
GA Iix	90% n_15 + 10% t21_05	4	36	2.00	11,170,581	4,453,489	40%	21,030
GA Iix	75% n_15 + 25% t21_05	4	37	2.00	9,444,389	3,761,988	40%	20,950
HiSeq 2000	100% n_15	5	38	4.00	29,844,141	10,239,086	34%	22,194
HiSeq 2000	100% n_15	5	39	4.00	36,518,106	13,771,732	38%	23,612
HiSeq 2000	100% n_15	5	40	4.00	29,971,627	11,108,083	37%	22,689

图 6A

HiSeq 2000	100% n_15	5	41	4.00	42,240,580	15,998,021	38%	23,436
HiSeq 2000	96% n_15 + 4% t21_05	5	42	4.00	22,779,370	7,992,518	35%	21,374
HiSeq 2000	96% n_15 + 4% t21_05	5	43	4.00	20,583,900	8,066,490	39%	21,822
HiSeq 2000	96% n_15 + 4% t21_05	5	44	4.00	13,435,012	5,239,203	39%	20,555
HiSeq 2000	96% n_15 + 4% t21_05	5	45	4.00	16,988,321	6,752,178	40%	21,053
HiSeq 2000	92% n_15 + 8% t21_05	5	46	4.00	31,696,544	10,785,467	34%	22,376
HiSeq 2000	92% n_15 + 8% t21_05	5	47	4.00	33,552,046	12,537,527	37%	23,710
HiSeq 2000	92% n_15 + 8% t21_05	5	48	4.00	31,850,369	11,635,853	37%	22,784
HiSeq 2000	92% n_15 + 8% t21_05	5	49	4.00	37,499,259	14,082,747	38%	23,228

图 6B

补充表3: 用于FAST-SeqS样品制备和测序的寡核苷酸

引物名称	序列
第1轮 正向-1	<u>CGACGTAAAACGACGGCCAGTNNNNNNNNNNNNNNNNNNNACACAGGGAGGGGAACAT</u>
第1轮 正向-2	<u>CGACGTAAAACGACGGCCAGT<b>HBVDHBVDHBVDHBVDHBVD</b>ACACAGGGAGGGGAACAT</u>
第1轮 反向	<u>CACACAGGAAACAGCTATGACCATGTGCCATGGTGGTTTGCT</u>
第2轮 正向	AATGATACGGCGACACCCGAGATCTACACCGACGTAAAACGACGGCCAGT
第2轮 反向-1	CAAGCAGAAGACGGCATAACGAGAT <b>GATCAG</b> CACACAGGAAACAGCTATGACCATG
第2轮 反向-2	CAAGCAGAAGACGGCATAACGAGAT <b>GCCCAAT</b> CACACAGGAAACAGCTATGACCATG
第2轮 反向-3	CAAGCAGAAGACGGCATAACGAGAT <b>TGACC</b> CACACAGGAAACAGCTATGACCATG
第2轮 反向-4	CAAGCAGAAGACGGCATAACGAGAT <b>TTAGCC</b> CACACAGGAAACAGCTATGACCATG
第2轮 反向-5	CAAGCAGAAGACGGCATAACGAGAT <b>CTTCTA</b> CACACAGGAAACAGCTATGACCATG
通用测序引物	GAGATCTACACCGACGTAAAACGACGGCCAGT
索引测序引物	CATGGTCATAGCTGTTTCCGTGTG

说明:

FAST-1序列

唯一标识符

通用引物序列

索引序列

Illumina接枝序列

图 7

表1-本FAST-SeqS研究中分析的样品

个体	组织	非整倍性	样品名称
n_01	血浆	无	n_01_pls
n_02	血浆	无	n_02_pls
n_03_1	血浆	无	n_03_1_pls
n_03_2	血浆	无	n_03_2_pls
n_04	血浆	无	n_04_pls
n_05	血浆	无	n_05_pls
n_06	血浆	无	n_06_pls
n_07	血浆	无	n_07_pls
n_01	WBC	无	n_01_wbc
n_02	WBC	无	n_02_wbc
n_03_1	WBC	无	n_03_1_wbc
n_03_2	WBC	无	n_03_2_wbc
n_04	WBC	无	n_04_wbc
n_05	WBC	无	n_05_wbc
n_06	WBC	无	n_06_wbc
n_07	WBC	无	n_07_wbc
n_08	脾	无	n_08
n_09	脾	无	n_09
n_10	脾	无	n_10
n_11	脾	无	n_11
n_12	脾	无	n_12
n_13	脾	无	n_13
n_14	脾	无	n_14
n_15	WBC	无	n_15
t21_01	成纤维细胞	21三体 (NA02767)	t21_01
t21_02	成纤维细胞	21三体 (NA04616)	t21_02
t21_03	成纤维细胞	21三体 (NG05120)	t21_03
t21_04	成纤维细胞	21三体 (NG05397)	t21_04
t21_05	成纤维细胞	21三体 (NG07438)	t21_05
t18_01	成纤维细胞	18三体 (NA03623)	t18_01
t18_02	成纤维细胞	18三体 (NG12614)	t18_02
t13_01	成纤维细胞	13三体 (NA03330)	t13_01

图 8

专利名称(译)	快速非整倍性检测		
公开(公告)号	<a href="#">CN104350158A</a>	公开(公告)日	2015-02-11
申请号	CN201380026625.9	申请日	2013-03-22
[标]申请(专利权)人(译)	约翰霍普金斯大学		
申请(专利权)人(译)	约翰霍普金斯大学		
当前申请(专利权)人(译)	约翰霍普金斯大学		
[标]发明人	B沃戈尔斯滕 KW金泽勒 N派帕多普勒斯 伊萨克G金德		
发明人	B·沃戈尔斯滕 K·W·金泽勒 N·派帕多普勒斯 伊萨克·G·金德		
IPC分类号	C12Q1/68 C12N15/11 G01N33/53 G01N33/48		
CPC分类号	C12Q2600/112 C12Q2600/156 C12Q2600/16 C12Q1/6827 C12Q1/6811 C12Q1/6883 C12Q2535/122 C12Q2545/114 C12Q2537/16 C12Q2537/165 C12Q1/6869		
代理人(译)	王思琪 郑霞		
优先权	61/615535 2012-03-26 US 61/659695 2012-06-14 US		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

最近证明，对无细胞的母体血浆DNA的大规模并行测序是胎儿染色体非整倍性的安全有效的筛选方法。在这里，我们报告了一种通过以采用单引物对的PCR替换费力的测序文库制备步骤，实现显著提高通量和降低成本的改进的测序方法。使用这种方法，可容易区分含有少至4%的21三体DNA的样品与整倍体样品。

