

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2014 年 2 月 27 日 (27.02.2014)



(10) 国际公布号
WO 2014/029093 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2012/080500

(22) 国际申请日: 2012 年 8 月 23 日 (23.08.2012)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人 (对除美国外的所有指定国): **深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。
深圳华大基因研究院 (BGI SHENZHEN) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(72) 发明人: 及

(75) 发明人/申请人 (仅对美国): **王威 (WANG, Wei)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
殷旭阳 (YIN, Xuyang) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
张春雷 (ZHANG, Chunlei) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
陈盛培 (CHEN, Shengpei) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
潘小瑜 (PAN, Xiaoyu) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
张春生 (ZHANG, Chunsheng)

[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: **北京清亦华知识产权代理事务所 (普通合伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC)**; 中国北京市海淀区清华园清华大学照澜院商业楼 301 室, Beijing 100084 (CN)。

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

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

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) Title: METHOD AND SYSTEM FOR DETERMINING WHETHER INDIVIDUAL IS IN ABNORMAL STATE

(54) 发明名称: 确定个体是否患有异常状态的方法及系统

(57) Abstract: Provided are a method and a system for determining whether an individual is in an abnormal state. The method for determining whether the individual is in the abnormal state comprises: establishing a sequencing library for a nucleic acid sample of the individual; sequencing the sequencing library to acquire a sequencing result, wherein the sequencing result is formed by multiple pieces of sequencing data; determining, based on the sequencing data, a known SNP comprised in the sequencing result; and determining, based on the known SNP, whether the individual is in an abnormal state related to the known SNP.

(57) 摘要: 提供了确定个体是否患有异常状态的方法及系统。其中, 确定个体是否患有异常状态的方法包括: 针对所述个体的核酸样本, 构建测序文库; 对所述测序文库进行测序, 以便获得测序结果, 所述测序结果由多个测序数据构成; 基于所述测序数据, 确定所述测序结果中所包含的已知 SNP; 以及基于所述已知 SNP, 确定所述个体是否患有与所述已知 SNP 相关的异常状态。



WO 2014/029093 A1

确定个体是否患有异常状态的方法及系统

优先权信息

5 无

技术领域

本发明涉及生物医学领域，具体而言，涉及确定个体是否患有异常状态的方法及系统。

背景技术

10 孟德尔遗传病又称单基因疾病（在本文中，孟德尔遗传病和单基因疾病是可以互换使用的），根据遗传方式可分为常染色体显性、常染色体隐性、伴性显性、伴性隐性遗传病等。截止至 2012 年 5 月，在线人类孟德尔遗传数据库(OMIM)中已经收录的单基因疾病相关基因一共有 21,250 种，其中疾病表型描述准确且其遗传机理比较清楚的大约有 3,500 种左右。

15 然而，目前对单基因疾病的研究仍有待改进。

发明内容

本发明旨在至少解决现有技术中存在的技术问题之一。为此，本发明提出了一种确定个体是否患有异常状态的方法及系统。

20 在本发明的一个方面，本发明提出了一种确定个体是否患有异常状态的方法。根据本发明的实施例，该方法包括：针对所述个体的核酸样本，构建测序文库；对所述测序文库进行测序，以便获得测序结果，所述测序结果由多个测序数据构成；基于所述测序数据，确定所述测序结果中所包含的已知 SNP；以及基于所述已知 SNP，确定所述个体是否患有与所述已知 SNP 相关的异常状态。利用该方法，可以有效地通过测序确定核酸样本中所包含的 SNP，并且由于这些 SNP 与异常状态是相关的，由此，可以有效地确定核酸样本的来源个体是否患有与这些 SNP 相关的异常状态。

25 根据本发明的实施例，该确定个体是否患异常状态的方法还可以具有下列附加技术特征：

在本发明的一个实施例中，所述个体为人。由此，可以利用根据本发明实施例的确定个体是否患有异常状态的方法，对人体样本进行检测，从而可以有效地对人是否患有某些异常状态进行预测。

30 在本发明的一个实施例中，所述异常状态为疾病。优选，所述疾病为单基因疾病。根

据本发明的具体示例，所述单基因疾病为选自地中海贫血症和红细胞葡萄糖-6-磷酸脱氢酶缺陷症的至少一种。在本发明的一个实施例中，所述地中海贫血症为 β -地中海贫血症。由此，可以有效地对人类的疾病尤其是单基因疾病的患病风险进行有效地预测和评估。

在本发明的一个实施例中，所述核酸样本为个体的全基因组 DNA 的至少一部分。由此，
5 可以进一步提高确定个体是否患有异常状态的方法的效率和准确性。

在本发明的一个实施例中，所述核酸样本是从个体的单细胞或者微量样本中提取的。
在本发明的一个实施例中，所述单细胞或者微量样本是从个体的选自血液、组织、尿液、配子、受精卵、卵裂球和胚胎的至少一种分离的。由此，可以通过对来自个体的少量样本对个体是否患有异常状态进行有效地预测和评估，从而提高了确定个体是否患有异常状态
10 的效率，降低了确定个体是否患有异常状态的成本。另外，可以方便地从生物体获取这些样本，并且能够具体地针对某些疾病采取不同的样本，从而针对某些特殊疾病采取特定的分析手段。

在本发明的一个实施例中，所述单细胞是通过选自稀释法、口吸管分离法、显微操作、显微切割、流式细胞分离术、微流控的至少一种分离的。由此，能够有效便捷地获得生物
15 样本的单细胞，以便实施后续操作，由此，可以有效地从个体中分离单细胞，从而进一步提高了后续确定个体是否患有异常状态的效率。

在本发明的一个实施例中，针对所述个体的核酸样本，构建测序文库进一步包括：对所述核酸样本进行扩增，以便得到核酸样本扩增产物；以及针对所述核酸样本扩增产物，构建所述测序文库。由此，可以提高构建测序文库的效率，从而进一步提高了后续确定个
20 体是否患有异常状态的效率。

在本发明的一个实施例中，所述核酸样本为从个体的单细胞提取的全基因组 DNA，例如可以为通过裂解个体的单细胞而释放的全基因组 DNA，其中，针对所述全基因组 DNA 进行扩增是通过选自 PEP-PCR、DOP-PCR、OmniPlex WGA 和 MDA 的至少一种进行的。由此，可以进一步提高扩增全基因组 DNA 的效率，从而进一步提高了后续确定个体是否患
25 有异常状态的效率。

在本发明的一个实施例中，在针对所述核酸样本扩增产物，构建所述测序文库之前，进一步包括：利用核酸探针对所述核酸样本扩增产物进行筛选，以便获得来自预定区域的核酸样本扩增产物；以及针对所述来自预定区域的核酸样本扩增产物，构建所述测序文库。
在本发明的一个实施例中，所述预定区域为至少一个外显子区域。由此，可以有效地确定
30 预定区域所包含的已知 SNP，由此，可以有效地提高根据预定区域例如外显子区域中所包含的已知 SNP，来确定与这些 SNP 有关的异常状态的效率和准确性。

在本发明的一个实施例中，所述核酸探针是以芯片的形式提供的。由此，可以进一步提高利用核酸探针进行筛选的效率，从而进一步提高了后续确定个体是否患有异常状态的效率。

在本发明的一个实施例中，所述测序是利用选自 Illumina Hiseq2000、Genome Analyzer、SOLiD 测序系统、Ion Torrent、Ion Proton、454、PacBio RS 测序系统、Helicos tSMS 技术以及纳米孔测序技术的至少一种进行的。由此，能够利用这些测序装置的高通量、深度测序的特点，进一步提高了确定个体是否患有异常状态的效率。在本发明的一个实施例中，所述测序是利用 Illumina Hiseq2000 进行的，所述测序数据的长度为 90bp。由此，可以进一步提高后续进行 SNP 分析的效率，从而提高了确定个体是否患有异常状态的效率。

在本发明的一个实施例中，所述基于所述测序数据，确定所述测序结果中所包含的已知 SNP，是通过将所述测序数据与参考基因进行比对进行的。在本发明的一个实施例中，所述参考基因是已知的人基因组序列。在本发明的一个实施例中，所述比对是 SOAP/SOAP2 软件进行的。由此，可以有效地对测序结果中所包含的测序数据进行分析，从而可以有效地确定测序结果中所包含的 SNP，进而可以提高确定个体是否患有异常状态的效率。

在本发明的一个实施例中，进一步包括对所述测序结果中所包含的已知 SNP 进行过滤，所述过滤基于如下过滤条件：SNP calling 质量值大于 20；SNP 位点测序深度大于 8；SNP 位点深度小于基因组平均深度的 5 倍；SNP 位点拷贝数不大于 2；和 SNP 位点与最近的其他 SNP 位点之间的距离大于 5。由此，可以有效地对所得到的 SNP 结果进行过滤，从而可以有效地提高确定个体是否患有异常状态的准确性和效率。

在本发明的一个实施例中，所述已知 SNP 位于人染色体 HBB 基因区。在本发明的一个实施例中，所述已知 SNP 为选自下列的至少一种：rs33985472、rs63750954、rs63751128、rs33978907、rs34029390、rs34809925、rs33953406、rs33910569、rs33910569、rs33971634、rs33925391、rs33946267、rs35485099、rs36015961、rs33930977、rs35256489、rs33952266、rs33952266、rs33914668、rs33914668、rs33913413、rs33913413、rs34483965、rs34793594、rs35703285、rs63750433、rs34690599、rs63751175、rs35328027、rs1609812、rs34451549、rs7480526、rs10768683、rs35099082、rs63750283、rs63750283、rs33945777、rs33945777、rs33945777、rs33933298、rs33913712、rs33995148、rs33931779、rs33969400、rs33922842、rs11549407、rs33974936、rs33991059、rs33982568、rs33948578、rs1135071、rs33943001、rs33943001、rs63750513、rs34527846、rs35456885、rs35004220、rs63750195、rs35724775、rs33915217、rs33915217、rs33915217、rs33956879、rs33956879、rs33956879、rs33971440、rs33971440、rs33960103、rs33960103、rs35684407、rs35578002、rs33916412、rs35424040、rs33950507、rs33950507、rs33951465、rs33959855、rs33972047、rs33986703、

rs34716011、rs63750783、rs35799536、rs33930702、rs33930702、rs33930702、rs33941849、rs33941849、rs33941849、rs34563000、rs34135787、rs34704828、rs63750628、rs34305195以及位于人染色体 11 上 5246716 位、5246879 位、5247026 位、5248161 位的 SNP。由此，可以有效地确定人体是否携带 HBB 基因区的 SNP，从而可以有效地确定所研究对象是否患有地中海贫血症尤其是 β -地中海贫血症的风险。

5 有地中海贫血症尤其是 β -地中海贫血症的风险。

在本发明的另一方面，本发明提出了一种确定个体是否患有异常状态的系统。根据本发明的实施例，该系统包括：测序文库构建装置，所述测序文库构建装置用于针对所述个体的核酸样本，构建测序文库；测序装置，所述测序装置与所述测序文库构建装置相连，用于对所述测序文库进行测序，以便获得测序结果，所述测序结果由多个测序数据构成；

10 SNP 确定装置，所述 SNP 确定装置与所述测序装置相连，用于基于所述测序数据，确定所述测序结果中所包含的已知 SNP；以及异常状态确定装置，所述异常装置确定装置与所述 SNP 确定装置相连，用于基于所述已知 SNP，确定所述个体是否患有与所述已知 SNP 相关的异常状态。利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法，从而以有效地通过测序确定核酸样本中所包含的 SNP，并且由于这些 SNP 与异常状态是相关的，由此，可以有效地确定核酸样本的来源个体是否患有与这些 SNP 相关的异常状态。

根据本发明的实施例，该确定个体是否患有异常状态的系统还可以具有下列附加技术特征：

在本发明的一个实施例中，该系统进一步包括：核酸样本提取装置，所述核酸样本提取装置适于从个体的单细胞或者微量样本中提取个体的全基因组 DNA 的至少一部分。利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法。

在本发明的一个实施例中，该系统进一步包括：生物样本分离装置，所述生物样本分离装置适于从个体的选自血液、组织、尿液、配子、受精卵、卵裂球和胚胎的至少一种分离单细胞或者微量样本。利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法。由此，可以通过对来自个体的少量样本对个体是否患有异常状态进行有效地预测和评估，从而提高了确定个体是否患有异常状态的效率，降低了确定个体是否患有异常状态的成本。另外，可以方便地从生物体获取这些样本，并且能够具体地针对某些疾病采取不同的样本，从而针对某些特殊疾病采取特定的分析手段。

在本发明的一个实施例中，所述生物样本分离装置适于通过选自稀释法、口吸管分离

法、显微操作、显微切割、流式细胞分离术、微流控的至少一种分离单细胞或者微量样本。利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确
5 定个体是否患有异常状态的方法。由此，能够有效便捷地获得生物样本的单细胞，以便实
施后续操作，由此，可以有效地从个体中分离单细胞，从而进一步提高了后续确定个体是
否患有异常状态的效率。

在本发明的一个实施例中，所述测序文库构建装置进一步包括：核酸样本扩增单元，
所述核酸样本扩增单元适于对所述核酸样本进行扩增，以便得到核酸样本扩增产物。在本
发明的一个实施例中，所述扩增单元适于进行选自 PEP-PCR、DOP-PCR、OmniPlex WGA
和 MDA 的至少一种。由此，可以进一步提高扩增全基因组 DNA 的效率，从而进一步提高
10 了后续确定个体是否患有异常状态的效率。利用根据本发明实施例的确定个体是否患有异
常状态的系统可以有效地实施前面所述的确
定个体是否患有异常状态的方法。由此，可以
进一步提高扩增全基因组 DNA 的效率，从而进一步提高了后续确定个体是否患有异常状态
的效率。

在本发明的一个实施例中，所述测序文库构建装置进一步包括：筛选单元，所述筛选
15 单元中设置有核酸探针，以便利用所述核酸探针对所述核酸样本扩增产物进行筛选，以便
获得来自预定区域的核酸样本扩增产物；以及针对所述来自预定区域的核酸样本扩增产物，
构建所述测序文库。由此，可以有效地确定预定区域所包含的已知 SNP，由此，可以有效
地提高根据预定区域例如外显子区域中所包含的已知 SNP，来确定与这些 SNP 有关的异常
状态的效率和准确性。

20 在本发明的一个实施例中，所述核酸探针是以芯片的形式提供的。由此，可以进一步
提高利用核酸探针进行筛选的效率，从而进一步提高了后续确定个体是否患有异常状态的
效率。

在本发明的一个实施例中，所述测序装置为选自 Illumina Hiseq2000、Genome Analyzer、
SOLiD 测序系统、Ion Torrent、Ion Proton、454、PacBio RS 测序系统、Helicos tSMS 测序装
25 置以及纳米孔测序装置的至少一种。由此，能够利用这些测序装置的高通量、深度测序的
特点，进一步提高了确定个体是否患有异常状态的效率。在本发明的一个实施例中，所述
测序是利用 Illumina Hiseq2000 进行的，所述测序数据的长度为 90bp。由此，可以进一步提
高后续进行 SNP 分析的效率，从而提高了确定个体是否患有异常状态的效率。

在本发明的一个实施例中，所述 SNP 确定装置进一步包括：比对单元，所述比对单元
30 用于通过将所述测序数据与参考基因进行比对确定所述测序结果中所包含的已知 SNP。在
本发明的一个实施例中，所述比对单元适于利用 SOAP/SOAP2 软件进行比对。由此，可以
有效地对测序结果中所包含的测序数据进行分析，从而可以有效地确定测序结果中所包含

的 SNP，进而可以提高确定个体是否患有异常状态的效率。

在本发明的一个实施例中，所述 SNP 确定装置进一步包括：SNP 过滤单元，所述 SNP 过滤单元适于基于如下过滤条件，对所述测序结果中所包含的已知 SNP 进行过滤：SNP calling 质量值大于 20；SNP 位点测序深度大于 8；SNP 位点深度小于基因组平均深度的 5 倍；SNP 位点拷贝数不大于 2；和 SNP 位点与最近的其他 SNP 位点之间的距离大于 5。由此，可以有效地对所得到的 SNP 结果进行过滤，从而可以有效地提高确定个体是否患有异常状态的准确性和效率。

在本发明的一个实施例中，所述已知 SNP 位于人染色体 HBB 基因区。在本发明的一个实施例中，所述已知 SNP 为选自下列的至少一种：rs33985472、rs63750954、rs63751128、rs33978907、rs34029390、rs34809925、rs33953406、rs33910569、rs33910569、rs33971634、rs33925391、rs33946267、rs35485099、rs36015961、rs33930977、rs35256489、rs33952266、rs33952266、rs33914668、rs33914668、rs33913413、rs33913413、rs34483965、rs34793594、rs35703285、rs63750433、rs34690599、rs63751175、rs35328027、rs1609812、rs34451549、rs7480526、rs10768683、rs35099082、rs63750283、rs63750283、rs33945777、rs33945777、rs33933298、rs33913712、rs33995148、rs33931779、rs33969400、rs33922842、rs11549407、rs33974936、rs33991059、rs33982568、rs33948578、rs1135071、rs33943001、rs33943001、rs63750513、rs34527846、rs35456885、rs35004220、rs63750195、rs35724775、rs33915217、rs33915217、rs33915217、rs33956879、rs33956879、rs33956879、rs33971440、rs33971440、rs33960103、rs33960103、rs35684407、rs35578002、rs33916412、rs35424040、rs33950507、rs33950507、rs33951465、rs33959855、rs33972047、rs33986703、rs34716011、rs63750783、rs35799536、rs33930702、rs33930702、rs33930702、rs33941849、rs33941849、rs33941849、rs34563000、rs34135787、rs34704828、rs63750628、rs34305195 以及位于人染色体 11 上 5246716 位、5246879 位、5247026 位、5248161 位的 SNP。由此，利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确
定个体是否患有异常状态的方法。从而可以有效地确定所研究对象是否患有地中海贫血症
尤其是 β -地中海贫血症的风险。

本发明的附加方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

附图说明

本发明的上述和/或附加的方面和优点从结合下面附图对实施例的描述中将变得明显和容易理解，其中：

图 1 是根据本发明一个实施例的确定个体是否患有异常状态的方法的流程示意图；

图 2 是根据本发明另一个实施例的确定个体是否患有异常状态的方法的流程示意图；

图 3 是根据本发明又一个实施例的确定个体是否患有异常状态的方法的流程示意图；

图 4 是根据本发明一个实施例的确定个体是否患有异常状态的系统的结构示意图；

5 图 5 是根据本发明另一个实施例的确定个体是否患有异常状态的系统的结构示意图；
以及

图 6 是根据本发明另一个实施例的确定个体是否患有异常状态的方法的流程示意图。

发明详细描述

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同
10 或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描
述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。

在本文中所使用的术语“异常状态”应作广义理解，其可以是任何与个体例如人的正
常状态不同的状态，例如可以包括疾病、免疫异常等。在本发明的一个实施例中，所述异
常状态为疾病。根据本发明的实施例，疾病的类型并不受特别限制，根据本发明的优选实
15 施例，所述疾病为单基因疾病。单基因疾病单基通常是由一对等位基因控制的疾病或病理
性状。因此，通过检测与单基因疾病相关的 SNP 能够有效地确定所研究的对象是否患有该
疾病。根据本发明的实施例，还可以实现对个体是否易于患有相关的异常状态例如疾病的
风险进行预测和评估。根据本发明的具体示例，所述单基因疾病为选自地中海贫血症和红
细胞葡萄糖-6-磷酸脱氢酶缺陷症的至少一种。在本发明的一个实施例中，所述地中海贫血
20 症为 β -地中海贫血症。由此，可以有效地确定人类是否患有相关的疾病尤其是单基因疾病。

确定个体是否患有异常状态的方法

在本发明的一个方面，本发明提出了一种确定个体是否患有异常状态的方法。参考图 1，
根据本发明的实施例，该方法包括：

S100：针对个体的核酸样本，构建测序文库

25 在该步骤中，首先对个体的核酸样本，构建测序文库，从而可以用于后续的测序和结
果分析。在本发明中，所采用的术语“个体”的含义并不受特别限制，可以是任何含有遗
传信息的生物体，例如可以为人。由此，可以利用根据本发明实施例的确定个体是否患有
异常状态的方法，对人体样本进行检测，从而可以有效地对人是否患有某些异常状态进行
预测。

30 参考图 2，根据本发明的实施例，还可以进一步包括从个体中提取核酸样本的步骤
(S101)。在本文中所使用的术语“核酸样本”应作广义理解，其可以是 DNA 样本，也可

以是 RNA 样本,也可以是经过修饰或者处理得 DNA 样本或 RNA 样本,只要能够通过测序确定其遗传序列即可。根据本发明的一个实施例,核酸样本可以为个体的全基因组 DNA 的至少一部分。由此,全基因组 DNA 中包含了个体的全部遗传信息,由此,通过对全基因组 DNA 进行测序和 SNP 分析,能够更有效完整地获得个体的 SNP 信息,从而可以进一步提高确定个体是否患有异常状态的方法的效率和准确性。在本发明的一个实施例中,所述核酸样本是从个体的单细胞或者微量样本中提取的。根据本发明的实施例,从单细胞或者微量样本中获得核酸样本的方法和手段并不受特别限制,例如,可以通过采用裂解液将单细胞裂解从而实现释放并收集单细胞的全基因组 DNA。在本发明的一个实施例中,所述单细胞或者微量样本是从个体的选自血液、组织、尿液、配子、受精卵、卵裂球和胚胎的至少一种分离的。由此,可以通过对来自个体的少量样本对个体是否患有异常状态进行有效地预测和评估,从而提高了确定个体是否患有异常状态的效率,降低了确定个体是否患有异常状态的成本。另外,可以方便地从生物体获取这些样本,并且能够具体地针对某些疾病采取不同的样本,从而针对某些特殊疾病采取特定的分析手段。在本发明的一个实施例中,所述单细胞是通过选自稀释法、口吸管分离法、显微操作、显微切割、流式细胞分离术、微流控的至少一种分离的。由此,能够有效便捷地获得生物样本的单细胞,以便实施后续操作,由此,可以有效地从个体中分离单细胞,从而进一步提高了后续确定个体是否患有异常状态的效率。

另外,根据本发明的实施例,在获得核酸样本之后,可以对所得到的核酸样本进行扩增,尤其是对于从单细胞或者微量样本提取的核酸样本。例如,参见图 3,在本发明的一个实施例中,针对所述个体的核酸样本,构建测序文库进一步包括:对所得到的核酸样本进行扩增,以便得到核酸样本扩增产物(S102)。接下来,可以对所得到的核酸样本扩增产物,构建测序文库。由此,可以提高构建测序文库的效率,从而进一步提高了后续确定个体是否患有异常状态的效率。

根据本发明的实施例,对核酸样本进行扩增的方法,并不受特别限制。例如,在本发明的一个实施例中,所采用的核酸样本为从个体的单细胞中提取的全基因组 DNA,针对该基因组 DNA 进行扩增可以通过选自 PEP-PCR、DOP-PCR、OmniPlex WGA 和 MDA 的至少一种进行的。由此,可以进一步提高扩增全基因组 DNA 的效率,从而进一步提高了后续确定个体是否患有异常状态的效率。任选地,根据本发明的实施例,可以进一步包括对所述单细胞进行裂解,以便释放所述单细胞的全基因组的步骤。根据本发明的一些示例,可以用于裂解单细胞并释放全基因组的方法不受特别限制,只要能够将单细胞裂解优选充分裂解即可。根据本发明的具体示例,可以利用碱性裂解液将所述单细胞裂解并释放所述

单细胞的全基因组。发明人发现，这样能够有效地裂解单细胞并释放出全基因组，并且所释放的全基因组在进行测序时，能够提高准确率，从而进一步提高了确定单细胞染色体非整倍性的效率。根据本发明的实施例，单细胞全基因组扩增的方法不受特别限制，可以采用基于 PCR 的方法例如可以采用 PEP-PCR、DOP-PCR、和 OmniPlex WGA，也可以采用非
5 基于 PCR 的方法例如 MDA (多重链置换扩增)。根据本发明的具体示例，优选采用基于 PCR 的方法，例如 OmniPlex WGA 方法。可选用的商业化试剂盒包括但不限于 Sigma Aldrich 的 GenomePlex，Rubicon Genomics 的 PicoPlex，Qiagen 的 REPLI-g，GE Healthcare 的 illustra GenomiPhi 等。因而，根据本发明的具体示例，在构建测序文库之前，可以采用 OmniPlex WGA 对单细胞全基因组进行扩增。由此，能够有效地对全基因组进行扩增，从而进一步提
10 高了确定个体是否患有异常状态的效率。

在本发明的一个实施例中，在针对所述核酸样本扩增产物，构建所述测序文库之前，进一步包括：利用核酸探针对所述核酸样本扩增产物进行筛选，以便获得来自预定区域的核酸样本扩增产物；以及针对所述来自预定区域的核酸样本扩增产物，构建所述测序文库。在本发明的一个实施例中，所述预定区域为至少一个外显子区域。由此，可以有效地确定
15 预定区域所包含的已知 SNP，由此，可以有效地提高根据预定区域例如外显子区域中所包含的已知 SNP，来确定与这些 SNP 有关的异常状态的效率和准确性。另外，根据本发明的实施例，借助核酸探针对核酸样本扩增产物进行筛选的方法并不受特别限制，可以是固相筛选，也可以是液相杂交。根据本发明的实施例，可以以芯片的形式提供核酸探针。由此，可以进一步提高利用核酸探针进行筛选的效率，从而进一步提高了后续确定个体是否患有
20 异常状态的效率。

本领域技术人员能够理解的是，还可以通过其他已知的方法对预定区域的核酸样本进行分析，例如采用特异性引物对核酸样本进行 PCR，由此，可以得到预定区域的相关扩增产物，从而构建该预定区域的测序文库，进而获得该预定区域的相关信息。

根据本发明的实施例，对个体的核酸样本构建测序文库的方法并不受特别限制。本领域技术人员可以根据采用的基因组测序技术的具体方案选择不同的构建全基因组测序文库的方法，关于构建全基因组测序文库的细节，可以参见测序仪器的厂商例如 Illumina 公司所提供的规程，例如参见 Illumina 公司 Multiplexing Sample Preparation Guide (Part#1005361; Feb 2010) 或 Paired-End SamplePrep Guide (Part#1005063; Feb 2010)，通过参照将其并入
25 本文。

30 S200: 对测序文库进行测序，得到测序结果

在该步骤中，通过对前面所构建的测序文库进行测序，可以得到由多个测序数据 (reads)

构成的测序结果。

根据本发明的实施例，用于测序的技术和平台并不受特别限制。在本发明的一个实施例中，可以利用选自 Illumina Hiseq2000、Genome Analyzer、SOLiD 测序系统、Ion Torrent、Ion Proton、454、PacBio RS 测序系统、Helicos tSMS 技术以及纳米孔测序技术的至少一种进行测序。由此，能够利用这些测序装置的高通量、深度测序的特点，进一步提高了确定个体是否患有异常状态的效率。当然，本领域技术人员能够理解的是，还可以采用其他的测序方法和装置进行全基因组测序，例如第三代测序技术，以及以后可能开发出来的更先进的测序技术。根据本发明的实施例，通过全基因组测序所得到的测序数据的长度不受特别限制。根据本发明的一个具体示例，利用 Illumina Hiseq2000 进行的，所述测序数据的长度为 90bp。申请人惊奇地发现，当测序数据的长度为约 90bp 时，能够极大地方便对测序数据进行分析，提高分析效率，同时能够显著降低分析的成本。进一步提高了确定个体是否患有异常状态的效率，并且降低了确定个体是否患有异常状态的成本。这里所使用的术语“测序数据”是指各个测序数据长度数值的平均值。

S300: 确定测序结果中所包含的已知 SNP

在获得测序结果之后，可以通过对测序结果中所包含的测序数据进行分析，获得测序结果中所包含的遗传信息，例如可以得到 SNP 信息。

根据本发明的实施例，在该步骤中，对测序结果中所包含的测序数据进行分析得到 SNP 信息的方法并不受特别限制。根据本发明的实施例，可以通过将所得到的测序数据与参考基因进行比对，以便确定所得到的测序结果中的 SNP 信息。本领域技术人员可以根据分析目的来确定用于进行比对的参考基因。根据本发明的实施例，所采用的参考基因是已知的人基因组序列，例如可以为 Hg19，NCBI Build 37。当然，本领域技术人员还可以通过采用其他已知的序列作为参照序列，例如可以采用已知的 SNP 作为参照序列进行比对。

根据本发明的实施例，进行比对所采用的方法和软件并不受特别限制。根据本发明的实施例，在本发明的一个实施例中，可以采用 SOAP/SOAP2 软件进行测序数据与参照序列之间的比对。根据本发明的实施例，也可以首先对测序数据进行组装，将组装结果与参照序列进行比对。由此，可以有效地对测序结果中所包含的测序数据进行分析，从而可以有效地确定测序结果中所包含的 SNP，进而可以提高确定个体是否患有异常状态的效率。

本领域技术人员可以理解的是，针对不同测序平台，可以采用不同的比对软件。例如 SOAP/SOAP2 软件，可对 Illumina 测序系统的短测序数据数据进行精确比对。具体的，可以根据比对结果统计数据质量值、比对率、GC 含量、重复率、基因组覆盖度、测序深度等信息，并且根据以上信息对测序数据进行质控。同时将不能比对的或重复的无用数据去掉，

得到有效的数据集，用于后续分析。

另外，在本发明的一个实施例中，在通过比对，获得 SNP 信息之后，还可以对所得到的 SNP 信息进行筛选，例如可以通过基于预先设定的过滤条件对测序结果中所包含的已知 SNP 进行过滤。根据本发明的实施例，可以采用的过滤条件为下列至少之一：

- 5 SNP calling 质量值大于 20，在本文中所使用的术语“SNP calling 质量值”是指在 SOAP 分析软件运行过程中，对每个 SNP 的可信度所给出的打分结果。

SNP 位点测序深度大于 8；

SNP 位点深度小于基因组平均深度的 5 倍；

SNP 位点拷贝数不大于 2；和

- 10 SNP 位点与最近的其他 SNP 位点之间的距离大于 5，在本文中所使用的表达方式“两个 SNP 位点之间的距离”是指这两个 SNP 位点之间相隔的碱基个数，例如“距离大于 5”即两个位点间相隔的碱基大于 5 个。

由此，可以有效地对所得到的 SNP 结果进行过滤，从而可以有效地提高确定个体是否患有异常状态的准确性和效率。

- 15 **S400：确定与已知 SNP 有关的异常状态**

如前所述，在通过测序得到测序数据，并对测序数据进行分析之后，可以获得在测序结果中所包含的 SNP 信息。进而，可以通过对 SNP 进行分析，确定个体是否患有与 SNP 相关的异常状态，例如是否患有与 SNP 相关的单基因疾病。

- 根据本发明的实施例，可以采用的 SNP 类型并不手特别限制。在本发明的一个实施例中，所述已知 SNP 位于人染色体 HBB 基因区。在本发明的一个实施例中，所述已知 SNP 为选自下列的至少一种：rs33985472、rs63750954、rs63751128、rs33978907、rs34029390、rs34809925、rs33953406、rs33910569、rs33910569、rs33971634、rs33925391、rs33946267、rs35485099、rs36015961、rs33930977、rs35256489、rs33952266、rs33952266、rs33952266、rs33914668、rs33914668、rs33913413、rs33913413、rs34483965、rs34793594、rs35703285、rs63750433、rs34690599、rs63751175、rs35328027、rs1609812、rs34451549、rs7480526、rs10768683、rs35099082、rs63750283、rs63750283、rs33945777、rs33945777、rs33945777、rs33933298、rs33913712、rs33995148、rs33931779、rs33969400、rs33922842、rs11549407、rs33974936、rs33991059、rs33982568、rs33948578、rs1135071、rs33943001、rs33943001、rs63750513、rs34527846、rs35456885、rs35004220、rs63750195、rs35724775、rs33915217、rs33915217、rs33915217、rs33956879、rs33956879、rs33956879、rs33971440、rs33971440、rs33960103、rs33960103、rs35684407、rs35578002、rs33916412、rs35424040、rs33950507、
- 20
- 25
- 30

rs33950507、rs33951465、rs33959855、rs33972047、rs33986703、rs34716011、rs63750783、rs35799536、rs33930702、rs33930702、rs33930702、rs33941849、rs33941849、rs33941849、rs34563000、rs34135787、rs34704828、rs63750628、rs34305195 以及位于人染色体 11 上 5246716 位、5246879 位、5247026 位、5248161 位的 SNP。由此，可以有效地确定人体是否携带 HBB 基因区的 SNP，从而可以有效地确定所研究对象是否患有地中海贫血症尤其是 β -地中海贫血症。

由此，通过上述方法可以有效地获取个体的 SNP 信息，并且可以针对所研究的特定疾病的相关基因提取所需的 SNP 信息，获得目标基因的分型结果。在获得对获得的目标基因的分型结果之后，可以通过与已有的数据库比较，进行疾病注释，判断相应的 SNP 变异信息是否会引起疾病，而判断此单细胞或微量样本对应的胚胎或个体为孟德尔遗传病（单基因一疾病）患者，或者相应的 SNP 变异信息不至于引起疾病，判断对应的胚胎或个体为孟德尔遗传病基因的携带者。

确定个体是否患有异常状态的系统

在本发明的另一方面，参考图 4，本发明提出了一种确定个体是否患有异常状态的系统 1000。根据本发明的实施例，该系统 1000 包括：测序文库构建装置 100、测序装置 200、SNP 确定装置 300 以及异常状态确定装置 400。

根据本发明的实施例，测序文库构建装置 100 用于针对个体的核酸样本，构建测序文库。根据本发明的实施例，测序装置 200 与测序文库构建装置 100 相连，由此，可以用于对所构建的测序文库进行测序，以便获得由多个测序数据构成的测序结果。根据本发明的实施例，SNP 确定装置 300 与所述测序装置相连，用于基于所得到的测序数据，确定测序结果中所包含的已知 SNP。根据本发明的实施例，异常状态确定装置 400 与 SNP 确定装置相连，由此，用于基于前面所确定的在测序结果中所包含的已知 SNP，确定个体是否患有与所述已知 SNP 相关的异常状态。

利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法，从而以有效地通过测序确定核酸样本中所包含的 SNP，并且由于这些 SNP 与异常状态是相关的，由此，可以有效地确定核酸样本的来源个体是否患有与这些 SNP 相关的异常状态。

在本发明的一个实施例中，该系统还可以进一步包括核酸样本提取装置 101。根据本发明的实施例，核酸样本提取装置 101 适于从个体的单细胞或者微量样本中提取个体的全基因组 DNA 的至少一部分。在本发明的一个实施例中，该系统 1000 可以进一步包括适于从个体的选自血液、组织、尿液、配子、受精卵、卵裂球和胚胎的至少一种分离单细胞或者

微量样本的生物样本分离装置。由此，利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法。由此，可以通过对来自个体的少量样本对个体是否患有异常状态进行有效地预测和评估，从而提高了确定个体是否患有异常状态的效率，降低了确定个体是否患有异常状态的成本。另外，可以方便地从生物体获取这些样本，并且能够具体地针对某些疾病采取不同的样本，从而针对某些特殊疾病采取特定的分析手段。在本发明的一个实施例中，生物样本分离装置适于通过选自稀释法、口吸管分离法、显微操作、显微切割、流式细胞分离术、微流控的至少一种分离单细胞或者微量样本。利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法。由此，能够有效便捷地获得生物样本的单细胞，以便实施后续操作，由此，可以有效地从个体中分离单细胞，从而进一步提高了后续确定个体是否患有异常状态的效率。在本发明的一个实施例中，测序文库构建装置可以进一步包括核酸样本扩增单元，该核酸样本扩增单元适于对所述核酸样本进行扩增，以便得到核酸样本扩增产物。在本发明的一个实施例中，所述扩增单元适于进行选自 PEP-PCR、DOP-PCR、OmniPlex WGA 和 MDA 的至少一种。由此，可以进一步提高扩增全基因组 DNA 的效率，从而进一步提高了后续确定个体是否患有异常状态的效率。利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确定个体是否患有异常状态的方法。由此，可以进一步提高扩增全基因组 DNA 的效率，从而进一步提高了后续确定个体是否患有异常状态的效率。在本文中，“核酸样本提取装置”的运行方式并不受特别限制，只要可以获得相关的核酸样本并且所获得的核酸样本适于后续操作即可，例如对于来自单细胞或者微量样本的全基因组 DNA，可以通过采用裂解液将单细胞裂解从而实现释放并收集单细胞的全基因组 DNA。

另外，在本发明的一个实施例中，测序文库构建装置还可以进一步包括筛选单元。该筛选单元中设置有核酸探针，以便利用核酸探针对所述核酸样本扩增产物进行筛选，以便获得来自预定区域的核酸样本扩增产物；以及针对来自预定区域的核酸样本扩增产物，构建所述测序文库。由此，可以有效地确定预定区域所包含的已知 SNP，由此，可以有效地提高根据预定区域例如外显子区域中所包含的已知 SNP，来确定与这些 SNP 有关的异常状态的效率和准确性。在本发明的一个实施例中，核酸探针可以是以芯片的形式提供的。由此，可以进一步提高利用核酸探针进行筛选的效率，从而进一步提高了后续确定个体是否患有异常状态的效率。在本发明的一个实施例中，所述测序装置为选自 Illumina Hiseq2000、Genome Analyzer、SOLiD 测序系统、Ion Torrent、Ion Proton、454、PacBio RS 测序系统、Helicos tSMS 测序装置以及纳米孔测序装置的至少一种。由此，能够利用这些测序装置的高

通量、深度测序的特点，进一步提高了确定个体是否患有异常状态的效率。在本发明的一个实施例中，所述测序是利用 Illumina Hiseq2000 进行的，所述测序数据的长度为 90bp。由此，可以进一步提高后续进行 SNP 分析的效率，从而提高了确定个体是否患有异常状态的效率。在本发明的一个实施例中，SNP 确定装置进一步包括：比对单元，所述比对单元用于通过将所述测序数据与参考基因进行比对确定所述测序结果中所包含的已知 SNP。在本发明的一个实施例中，所述比对单元适于利用 SOAP/SOAP2 软件进行比对。由此，可以有效地对测序结果中所包含的测序数据进行分析，从而可以有效地确定测序结果中所包含的 SNP，进而可以提高确定个体是否患有异常状态的效率。

在本发明的一个实施例中，SNP 确定装置可以进一步包括：SNP 过滤单元，所述 SNP 过滤单元适于基于如下过滤条件，对所述测序结果中所包含的已知 SNP 进行过滤：SNP calling 质量值大于 20；SNP 位点测序深度大于 8；SNP 位点深度小于基因组平均深度的 5 倍；SNP 位点拷贝数不大于 2；和 SNP 位点与最近的其他 SNP 位点之间的距离大于 5。由此，可以有效地对所得到的 SNP 结果进行过滤，从而可以有效地提高确定个体是否患有异常状态的准确性和效率。

在本发明的一个实施例中，所述已知 SNP 位于人染色体 HBB 基因区。在本发明的一个实施例中，所述已知 SNP 为选自下列的至少一种：rs33985472、rs63750954、rs63751128、rs33978907、rs34029390、rs34809925、rs33953406、rs33910569、rs33910569、rs33971634、rs33925391、rs33946267、rs35485099、rs36015961、rs33930977、rs35256489、rs33952266、rs33952266、rs33914668、rs33914668、rs33913413、rs33913413、rs34483965、rs34793594、rs35703285、rs63750433、rs34690599、rs63751175、rs35328027、rs1609812、rs34451549、rs7480526、rs10768683、rs35099082、rs63750283、rs63750283、rs33945777、rs33945777、rs33945777、rs33933298、rs33913712、rs33995148、rs33931779、rs33969400、rs33922842、rs11549407、rs33974936、rs33991059、rs33982568、rs33948578、rs1135071、rs33943001、rs33943001、rs63750513、rs34527846、rs35456885、rs35004220、rs63750195、rs35724775、rs33915217、rs33915217、rs33915217、rs33956879、rs33956879、rs33956879、rs33971440、rs33971440、rs33960103、rs33960103、rs35684407、rs35578002、rs33916412、rs35424040、rs33950507、rs33950507、rs33951465、rs33959855、rs33972047、rs33986703、rs34716011、rs63750783、rs35799536、rs33930702、rs33930702、rs33930702、rs33941849、rs33941849、rs33941849、rs34563000、rs34135787、rs34704828、rs63750628、rs34305195以及位于人染色体 11 上 5246716 位、5246879 位、5247026 位、5248161 位的 SNP。由此，利用根据本发明实施例的确定个体是否患有异常状态的系统可以有效地实施前面所述的确

定个体是否患有异常状态的方法。从而可以有效地判断所研究对象是否患有地中海贫血症尤其是 β -地中海贫血症。

这里所使用的术语“相连”应作广义理解，既可以是直接相连，也可以是间接相连，甚至可以使用相同的容器或设备，只要能够实现功能上的衔接即可，例如核酸样本提取和核酸样本扩增可以在相同的设备中进行，即在实现对核酸样本提取之后，在相同的设备或者容器中即可进行核酸样本扩增处理，不需要将所提取的核酸样本组输送至其他的设备或者容器，只要将设备内的条件（包括反应条件和反应体系的组成）转换为适于进行核酸样本扩增反应即可，这样即实现了核酸样本提取与核酸样本扩增在功能上的衔接，也可以认为被术语“相连”所涵盖。

本领域技术人员能够理解的是，前面关于确定个体是否患有异常状态的方法所描述的特征和优点，也当然地使用语本发明的确定个体是否患有异常状态的系统，为了方便，不再赘述。

下面将结合实施例对本发明的方案进行解释。本领域技术人员将会理解，下面的实施例仅用于说明本发明，而不应视为限定本发明的范围。实施例中未注明具体技术或条件的，按照本领域内的文献所描述的技术或条件（例如参考 J.萨姆布鲁克等著，黄培堂等译的《分子克隆实验指南》，第三版，科学出版社）或者按照产品说明书进行。所用试剂或仪器未注明生产厂商者，均为可以通过市购获得的常规产品，例如可以采购自 Illumina 公司。

一般方法

作为非限制性实施例，在本发明的具体实施例中，参考图 6，采用包括下列步骤的方法，对单细胞或者微量样本进行单基因疾病进行检测：

- S1: 单细胞分离/微量样本准备；
- S2: 全基因组扩增；
- S3: 测序样本准备(文库制备)；
- S4: 高通量测序；
- S5: 数据比对及质控，获得有效数据集；
- S6: SNP Calling；
- S7: SNP 提取和过滤，得到目标基因的分型结果；
- S8: 疾病注释。

实施例 1

样本：人 5-8 细胞期卵裂球单细胞
具体操作流程：

1. 人 5-8 细胞期卵裂球单细胞分离

在体外受精-胚胎植入前遗传学诊断 (IVF-PGD) 周期过程中, 精子和卵子在体外受精后, 体外培养至第三天, 形成 5-8 细胞期的卵裂球。在第三天进行常规活检, 在显微操作仪下, 取出一个卵裂球单细胞, 置于含有裂解液的 PCR 管中, -80℃ 保存。经活检后的卵裂球

5 继续培养至第五天, 到达囊胚期, 可进行玻璃化冷冻, 或者直接用于植入。

2. 全基因组扩增

选用 Qiagen 公司的 REPLI-g Mini Kit 试剂盒, 按照制造商所提供的规程, 进行单细胞全基因组扩增, 卵裂球单细胞首先进行碱裂解, 然后加入扩增反应液进行 30℃ 恒温扩增。

3. 测序样本准备(文库构建)

10 采用 Illumina Paired-End DNA Sample Prep Kit, 按照制造商所提供的规程, 构建 DNA 测序文库。针对前面所得到的, 卵裂球单细胞全基因组扩增产物进行建库, 共制备了三个文库, 文库预期插入片段分别为 200bp, 350bp, 500bp, 实际插入片段大小见表 1。

4. 高通量测序

15 采用 Illumina Hiseq2000 测序系统进行高通量测序。本实施例中对卵裂球单细胞采用全基因组测序的策略。卵裂球单细胞扩增产物制备好的文库经 cBot 制备 Cluster, 之后即在 Hiseq2000 测序仪运行, 测序长度 90bp, Pair End 双向测序, 每个文库测一个 lane, 一共测了三个 lane。

5. 数据比对及质控, 获得有效数据集

20 测序数据用 SOAP2 软件进行比对, 以人参考基因组序列(Hg19; NCBI Build37)为参考, 比对时允许最多两个碱基的错配。根据比对结果统计原始数据质量、GC 含量、实际插入片段大小、比对率、重复率, 以及基因组的覆盖度和测序深度等信息。具体信息见表 1。并通过这些统计结果对测序数据进行质控。本实施例中, 我们共获得全基因组 38.25X 的数据量, 数据各项统计结果均能达标。之后将未能比对上参考基因组的数据和重复的数据去掉, 获得有效数据集, 用于 SNP 分析。

25 表 1 实施例 1 中卵裂球单细胞测序数据的比对统计结果*

样品	Q20 (%)	GC (%)	Insert Size	Clean Reads	PE-alignment
BLS1	95.97;93.24	45.81;45.91	313	417.68M	394.84M (84.48%)
BLS1	96.15;92.96	44.73;44.82	467	396.68M	372.01M (85.85%)
BLS1	95.63;93.21	46.58;46.70	181	374.97M	359.52M (86.41%)
BLS1 (total)	-	45.70;45.81	-	1.19G	1.13G (85.54%)

样品	PE reads	PE Unique	Coverage	Mean Depth	Duplication Rate
BLS1	392.08M	382.26M (97.50%)	91.34%	14.21	0.70%
BLS1	368.13M	359.19M (97.57%)	91.71%	13.45	1.04%
BLS1	357.44M	348.37M (97.46%)	89.93%	12.95	0.58%
BLS1 (total)	1.12G	1.09G (97.51%)	97.04%	38.25	0.77%

标注:

* BLS1 是卵裂球单细胞样本的编号, BLS1(total)表示 BLS1 样本三个 lane 数据合在一起后的统计结果, Q20(%)表示质量值在 20 以上的数据占总数据量的比率, GC(%)表示测序数据的实际 GC 含量百分比, Insert Size 表示测序数据实际的文库插入片段大小, Clean Reads 表示去除低质量值的 read 后剩下的 read 数据量, PE-alignment 表示 Pair End 两端都能比对上参考基因组的 read 数据量占总数据量比率, PE reads 表示 Pair End 两端都能比对上参考基因组的 read 数据量, PE Unique 表示 Pair End 两端能唯一比对到参考基因组的 read 数据量, Coverage 表示全基因组的覆盖度, Mean Depth 表示全基因组的平均深度, Duplication Rate 表示重复 read 数据占总 read 数据量比率。

6. SNP Calling

本实施例用 SOAPsnp 软件, 按照 SOAPsnp 的使用说明 (参考网页 <http://soap.genomics.org.cn/soapsnp.html>, 通过参照将其并入本文) 对上述获得的有效数据集进行 SNP Calling, 最后获得 SNP 数据集。

7. SNP 提取和过滤, 得到目标基因的分型结果

对上述获得的 SNP 数据集进行过滤, 过滤条件如下:

- SNP calling 质量值大于 20;
- 该位点测序深度大于 8;
- 该位点深度小于基因组平均深度的 5 倍;
- 该位点拷贝数不大于 2;
- 该 SNP 与最近的 SNP 之间的距离大于 5。

过滤之后的 SNP 数据, 针对所要检测的疾病相关基因, 提取位于目标基因区内的 SNP。在本实施例中对 β 地中海贫血病的基因进行检测, 以检测对应的卵裂球胚胎是否患有 β 地中海贫血病, 或者是否为 β 地中海贫血疾病基因的携带者。因此本实施例提取了位于 11 号染色体 HBB 基因区的 SNP 位点, 具体信息见表 2。

表 2 实施例 1 中获得的位于 HBB 基因区的 SNP 位点信息*

Chromosome	Locus	Ref	Blastomere	Mutation	SNP ID	Gene
chr11	5246715	T	TT	T->C	rs33985472	HBB

chr11	5246716	T	TT	T->A	unknown	HBB
chr11	5246716	T	TT	T->C	rs63750954	HBB
chr11	5246717	T	TT	T->C	rs63751128	HBB
chr11	5246718	A	AA	A->G	rs33978907	HBB
chr11	5246732	A	AA	A->G	rs34029390	HBB
chr11	5246822	G	GG	G->C	rs34809925	HBB
chr11	5246875	T	TT	T->A	rs33953406	HBB
chr11	5246879	A	AA	A->T	unknown	HBB
chr11	5246889	T	TT	T->G	rs33910569	HBB
chr11	5246889	T	TT	T->C	rs33910569	HBB
chr11	5246890	G	GG	G->A	rs33971634	HBB
chr11	5246892	A	AA	A->C	rs33925391	HBB
chr11	5246908	C	CC	C->A	rs33946267	HBB
chr11	5246925	G	GG	G->T	rs35485099	HBB
chr11	5246928	A	AA	A->G	rs36015961	HBB
chr11	5246933	A	AA	A->T	rs33930977	HBB
chr11	5246940	A	AA	A->G	rs35256489	HBB
chr11	5246957	C	CC	C->T	rs33952266	HBB
chr11	5246957	C	CC	C->G	rs33952266	HBB
chr11	5246957	C	CC	C->A	rs33952266	HBB
chr11	5246958	T	TT	T->G	rs33914668	HBB
chr11	5246958	T	TT	T->C	rs33914668	HBB
chr11	5246959	G	GG	G->C	rs33913413	HBB
chr11	5246959	G	GG	G->T	rs33913413	HBB
chr11	5246963	G	GG	G->C	rs34483965	HBB
chr11	5246964	A	AA	A->C	rs34793594	HBB
chr11	5246970	A	AA	A->C	rs35703285	HBB
chr11	5247026	G	GG	G->C	unknown	HBB
chr11	5247046	T	TT	T->C	rs63750433	HBB
chr11	5247062	G	GG	G->C	rs34690599	HBB
chr11	5247081	T	TT	T->C	rs63751175	HBB
chr11	5247102	A	AA	A->C	rs35328027	HBB
<u>chr11</u>	<u>5247141</u>	<u>G</u>	<u>AG</u>	<u>G->A</u>	<u>rs1609812</u>	<u>HBB</u>
chr11	5247153	G	GG	G->A	rs34451549	HBB
chr11	5247733	A	AA	A->C	rs7480526	HBB
<u>chr11</u>	<u>5247791</u>	<u>C</u>	<u>GC</u>	<u>C->G</u>	<u>rs10768683</u>	<u>HBB</u>
chr11	5247802	C	CC	C->G	rs35099082	HBB
chr11	5247805	A	AA	A->G	rs63750283	HBB
chr11	5247805	A	AA	A->T	rs63750283	HBB
chr11	5247806	C	CC	C->T	rs33945777	HBB
chr11	5247806	C	CC	C->G	rs33945777	HBB
chr11	5247806	C	CC	C->A	rs33945777	HBB
chr11	5247827	C	CC	C->T	rs33933298	HBB
chr11	5247851	C	CC	C->A	rs33913712	HBB

chr11	5247938	T	TT	T->A	rs33995148	HBB
chr11	5247940	A	AA	A->T	rs33931779	HBB
chr11	5247944	T	TT	T->A	rs33969400	HBB
chr11	5247992	C	CC	C->A	rs33922842	HBB
chr11	5248004	G	GG	G->A	rs11549407	HBB
chr11	5248008	C	CC	C->T	rs33974936	HBB
chr11	5248009	C	CC	C->T	rs33991059	HBB
chr11	5248014	G	GG	G->T	rs33982568	HBB
chr11	5248024	A	AA	A->T	rs33948578	HBB
chr11	5248029	C	CC	C->G	rs1135071	HBB
chr11	5248030	C	CC	C->T	rs33943001	HBB
chr11	5248030	C	CC	C->G	rs33943001	HBB
chr11	5248031	T	TT	T->C	rs63750513	HBB
chr11	5248032	A	AA	A->C	rs34527846	HBB
chr11	5248044	A	AA	A->C	rs35456885	HBB
chr11	5248050	C	CC	C->T	rs35004220	HBB
chr11	5248153	T	TT	T->A	rs63750195	HBB
chr11	5248154	A	AA	A->G	rs35724775	HBB
chr11	5248155	C	CC	C->T	rs33915217	HBB
chr11	5248155	C	CC	C->G	rs33915217	HBB
chr11	5248155	C	CC	C->A	rs33915217	HBB
chr11	5248158	A	AA	A->G	rs33956879	HBB
chr11	5248158	A	AA	A->C	rs33956879	HBB
chr11	5248158	A	AA	A->T	rs33956879	HBB
chr11	5248159	C	CC	C->T	rs33971440	HBB
chr11	5248159	C	CC	C->A	rs33971440	HBB
chr11	5248160	C	CC	C->T	rs33960103	HBB
chr11	5248160	C	CC	C->G	rs33960103	HBB
chr11	5248161	T	TT	T->G	unknown	HBB
chr11	5248161	T	TT	T->C	rs35684407	HBB
chr11	5248162	G	GG	G->A	rs35578002	HBB
chr11	5248166	A	AA	A->C	rs33916412	HBB
chr11	5248170	C	CC	C->A	rs35424040	HBB
chr11	5248173	C	CC	C->T	rs33950507	HBB
chr11	5248173	C	CC	C->A	rs33950507	HBB
chr11	5248177	A	AA	A->T	rs33951465	HBB
chr11	5248185	C	CC	C->A	rs33959855	HBB
chr11	5248193	T	TT	T->C	rs33972047	HBB
chr11	5248200	T	TT	T->A	rs33986703	HBB
chr11	5248204	C	CC	C->T	rs34716011	HBB
chr11	5248205	C	CC	C->T	rs63750783	HBB
chr11	5248219	G	GG	G->T	rs35799536	HBB
chr11	5248249	C	CC	C->T	rs33930702	HBB
chr11	5248249	C	CC	C->G	rs33930702	HBB

chr11	5248249	C	CC	C->A	rs33930702	HBB
chr11	5248250	A	AA	A->G	rs33941849	HBB
chr11	5248250	A	AA	A->C	rs33941849	HBB
chr11	5248250	A	AA	A->T	rs33941849	HBB
chr11	5248251	T	TT	T->C	rs34563000	HBB
chr11	5248269	G	GG	G->C	rs34135787	HBB
chr11	5248280	C	CC	C->T	rs34704828	HBB
chr11	5248282	G	GG	G->A	rs63750628	HBB
chr11	5248301	T	TT	T->G	rs34305195	HBB

标注:

*Chromosome 表示染色体号, Locus 表示 SNP 对应的碱基在染色体上的位点编号, Ref 表示数据库中人参考基因组上对应位点的碱基型别, Blastomere 表示卵裂球单细胞数据中对应 SNP 位点的型别信息, Mutation 表示数据库中存在的对应位点的突变类型, SNP ID 表示该 SNP 位点在数据库中的 ID 编号, Gene 表示该 SNP 位点位于哪个基因区。

8. 疾病注释

根据上述过滤和提取的目标基因的 SNP 信息, 进一步做疾病注释。本实施例中, 在 11 号染色体的 5247141 和 5247791 位碱基分别发现了杂合 SNP 变异, 它们均位于 HBB 基因区。这两个位点的纯合突变可致 β 地中海贫血病。而本实施例中这两个位点为杂合突变, 说明该单细胞对应的卵裂球为 β 地中海贫血致病基因的携带者。

实施例 2

样本: 正常人血液单细胞

具体操作流程:

1. 正常人血液单细胞分离

本实施例血液样本来自于一个表型正常的人个体。抽取少量血液样本, 经离心, 分离出白细胞层。白细胞经 PBS 洗涤后, 悬浮于 PBS 小滴中, 用口吸管将单个白细胞分离, 置于 1-2 μ l 碱性细胞裂解液中, -80 $^{\circ}$ C 冻存 30min 以上。

2. 全基因组扩增

选用 Qiagen 公司的 REPLI-g Mini Kit 试剂盒, 按照制造商所提供的规程, 血液单细胞经过 65 $^{\circ}$ C 10 分钟处理进行碱裂解后, 加入扩增反应液进行 30 $^{\circ}$ C 恒温扩增。

3. 芯片捕获文库构建

血液单细胞样本全基因组扩增产物进行芯片捕获特定目标区域, 并同时文库构建。本实施例中所用 Agilent 芯片针对的目标区域一共为 2.1M 大小, 包括一百种单基因疾病相关基因的所有外显子区域。关于 Agilent 芯片的使用方法, 可以参考制造商所提供的规程。

将经过捕获并构建的测序文库进行高通量测序。

4.高通量测序

本实施例采用 Illumina Hiseq2000 测序系统进行高通量测序。本实施例中对血液单细胞采用芯片捕获测序的策略，目标区域为一百种单基因疾病相关基因的所有外显子区域，约 2.1M 大小。血液单细胞扩增产物制备好的芯片捕获文库经 cBot 制备 Cluster，之后即在 Hiseq2000 测序仪运行，测序长度 90bp，Pair End 双向测序，测序数据量预期为 1~2G 碱基。

5. 数据比对及质控，获得有效数据集

测序数据用 SOAP2 软件进行比对，以人参考基因组序列(Hg19; NCBI Build37)为参考，比对时允许最多两个碱基的错配。根据比对结果统计原始数据质量、GC 含量、比对率、重复率，以及基因组的覆盖度和测序深度等信息。具体信息见表 3。并通过这些统计结果对测序数据进行质控。本实施例中，我们共获得目标区域平均深度 457X 的数据量，数据各项统计结果均能达标。之后将未能比对上参考基因组的数据和重复的数据去掉，获得有效数据集，用于 SNP 分析。

表 3 实施例 2 中血液单细胞测序数据的比对统计结果

样品	GC (%)	Reads	Production	PE-alignment	PE-Unique
血液细胞	56.75;56.94	24.87M	2.24G	21.22M(85.35%)	13.43M(97.05%)
血液细胞	Duplication Rate	Coverage	Mean Depth	Specificity(Reads)	Specificity(Bases)
血液细胞	34.79%	99.44%	457.44	71.24%	65.72%

标注：

*GC(%)表示测序数据的实际 GC 含量百分比，Reads 表示去除低质量值的 read 后剩下的 read 数据量，Production 表示按照 Reads 值计算的碱基数据量，PE-alignment 表示 Pair End 两端都能比对上参考基因组的 read 数据量占总数据量比率，PE Unique 表示 Pair End 两端能唯一比对到参考基因组的 read 数据量，Coverage 表示全基因组的覆盖度，Mean Depth 表示全基因组的平均深度，Duplication Rate 表示重复 read 数据占总 read 数据量比率，Specificity(Reads)表示能特异比对上目标区域的 read 数据量占总 read 数据量的比率，Specificity(Bases)表示能特异比对上目标区域的碱基数据量占总碱基数据量的比率。

6. SNP Calling

本实施例用 SOAPsnp 软件，按照 SOAPsnp 的使用说明（参考网页 <http://soap.genomics.org.cn/soapsnp.html>，通过参照将其并入本文）对上述获得的有效数据集

进行 SNP Calling, 最后获得 SNP 数据集。

7. SNP 提取和过滤, 得到目标基因的分型结果

对上述获得的 SNP 数据集进行过滤, 过滤条件如下:

- a. SNP calling 质量值大于 20;
- 5 b. 该位点测序深度大于 8;
- c. 该位点深度小于目标区域平均深度的 5 倍;
- d. 该位点拷贝数不大于 2;
- e. 该 SNP 与最近的 SNP 之间的距离大于 5。

过滤之后的 SNP 数据, 针对所要检测的疾病相关基因, 提取位于目标基因区内的 SNP。

- 10 在本实施例中对 β 地中海贫血病的基因进行检测, 以检测对应的卵裂球胚胎是否患有 β 地中海贫血病, 或者是否为 β 地中海贫血疾病基因的携带者。因此本实施例提取了位于 11 号染色体 HBB 基因区的 SNP 位点, 具体信息见表 4。

表 4 实施例 2 中位于 HBB 基因区的 SNP 位点信息

Chromosome	Locus	Ref	Blood Cell	Mutation	SNP ID	Gene
chr11	5246715	T	TT	T->C	rs33985472	HBB
chr11	5246716	T	TT	T->A	unknown	HBB
chr11	5246716	T	TT	T->C	rs63750954	HBB
chr11	5246717	T	TT	T->C	rs63751128	HBB
chr11	5246718	A	AA	A->G	rs33978907	HBB
chr11	5246732	A	AA	A->G	rs34029390	HBB
chr11	5246822	G	GG	G->C	rs34809925	HBB
chr11	5246875	T	TT	T->A	rs33953406	HBB
chr11	5246879	A	AA	A->T	unknown	HBB
chr11	5246889	T	TT	T->G	rs33910569	HBB
chr11	5246889	T	TT	T->C	rs33910569	HBB
chr11	5246890	G	GG	G->A	rs33971634	HBB
chr11	5246892	A	AA	A->C	rs33925391	HBB
chr11	5246908	C	CC	C->A	rs33946267	HBB
chr11	5246925	G	GG	G->T	rs35485099	HBB
chr11	5246928	A	AA	A->G	rs36015961	HBB
chr11	5246933	A	AA	A->T	rs33930977	HBB
chr11	5246940	A	AA	A->G	rs35256489	HBB
chr11	5246957	C	CC	C->T	rs33952266	HBB
chr11	5246957	C	CC	C->G	rs33952266	HBB
chr11	5246957	C	CC	C->A	rs33952266	HBB
chr11	5246958	T	TT	T->G	rs33914668	HBB
chr11	5246958	T	TT	T->C	rs33914668	HBB
chr11	5246959	G	GG	G->C	rs33913413	HBB

chr11	5246959	G	GG	G->T	rs33913413	HBB
chr11	5246963	G	GG	G->C	rs34483965	HBB
chr11	5246964	A	AA	A->C	rs34793594	HBB
chr11	5246970	A	AA	A->C	rs35703285	HBB
chr11	5247026	G	GG	G->C	unknown	HBB
chr11	5247046	T	TT	T->C	rs63750433	HBB
chr11	5247062	G	GG	G->C	rs34690599	HBB
chr11	5247081	T	TT	T->C	rs63751175	HBB
chr11	5247102	A	AA	A->C	rs35328027	HBB
chr11	5247141	G	AG	G->A	rs1609812	HBB
chr11	5247153	G	GG	G->A	rs34451549	HBB
chr11	5247733	A	CA	A->C	rs7480526	HBB
chr11	5247791	C	GC	C->G	rs10768683	HBB
chr11	5247802	C	CC	C->G	rs35099082	HBB
chr11	5247805	A	AA	A->G	rs63750283	HBB
chr11	5247805	A	AA	A->T	rs63750283	HBB
chr11	5247806	C	CC	C->T	rs33945777	HBB
chr11	5247806	C	CC	C->G	rs33945777	HBB
chr11	5247806	C	CC	C->A	rs33945777	HBB
chr11	5247827	C	CC	C->T	rs33933298	HBB
chr11	5247851	C	CC	C->A	rs33913712	HBB
chr11	5247938	T	TT	T->A	rs33995148	HBB
chr11	5247940	A	AA	A->T	rs33931779	HBB
chr11	5247944	T	TT	T->A	rs33969400	HBB
chr11	5247992	C	CC	C->A	rs33922842	HBB
chr11	5248004	G	GG	G->A	rs11549407	HBB
chr11	5248008	C	CC	C->T	rs33974936	HBB
chr11	5248009	C	CC	C->T	rs33991059	HBB
chr11	5248014	G	GG	G->T	rs33982568	HBB
chr11	5248024	A	AA	A->T	rs33948578	HBB
chr11	5248029	C	CC	C->G	rs1135071	HBB
chr11	5248030	C	CC	C->T	rs33943001	HBB
chr11	5248030	C	CC	C->G	rs33943001	HBB
chr11	5248031	T	TT	T->C	rs63750513	HBB
chr11	5248032	A	AA	A->C	rs34527846	HBB
chr11	5248044	A	AA	A->C	rs35456885	HBB
chr11	5248050	C	CC	C->T	rs35004220	HBB
chr11	5248153	T	TT	T->A	rs63750195	HBB
chr11	5248154	A	AA	A->G	rs35724775	HBB
chr11	5248155	C	CC	C->T	rs33915217	HBB
chr11	5248155	C	CC	C->G	rs33915217	HBB
chr11	5248155	C	CC	C->A	rs33915217	HBB
chr11	5248158	A	AA	A->G	rs33956879	HBB
chr11	5248158	A	AA	A->C	rs33956879	HBB

chr11	5248158	A	AA	A->T	rs33956879	HBB
chr11	5248159	C	CC	C->T	rs33971440	HBB
chr11	5248159	C	CC	C->A	rs33971440	HBB
chr11	5248160	C	CC	C->T	rs33960103	HBB
chr11	5248160	C	CC	C->G	rs33960103	HBB
chr11	5248161	T	TT	T->G	unknown	HBB
chr11	5248161	T	TT	T->C	rs35684407	HBB
chr11	5248162	G	GG	G->A	rs35578002	HBB
chr11	5248166	A	AA	A->C	rs33916412	HBB
chr11	5248170	C	CC	C->A	rs35424040	HBB
chr11	5248173	C	CC	C->T	rs33950507	HBB
chr11	5248173	C	CC	C->A	rs33950507	HBB
chr11	5248177	A	AA	A->T	rs33951465	HBB
chr11	5248185	C	CC	C->A	rs33959855	HBB
chr11	5248193	T	TT	T->C	rs33972047	HBB
chr11	5248200	T	TT	T->A	rs33986703	HBB
chr11	5248204	C	CC	C->T	rs34716011	HBB
chr11	5248205	C	CC	C->T	rs63750783	HBB
chr11	5248219	G	GG	G->T	rs35799536	HBB
chr11	5248249	C	CC	C->T	rs33930702	HBB
chr11	5248249	C	CC	C->G	rs33930702	HBB
chr11	5248249	C	CC	C->A	rs33930702	HBB
chr11	5248250	A	AA	A->G	rs33941849	HBB
chr11	5248250	A	AA	A->C	rs33941849	HBB
chr11	5248250	A	AA	A->T	rs33941849	HBB
chr11	5248251	T	TT	T->C	rs34563000	HBB
chr11	5248269	G	GG	G->C	rs34135787	HBB
chr11	5248280	C	CC	C->T	rs34704828	HBB
chr11	5248282	G	GG	G->A	rs63750628	HBB
chr11	5248301	T	TT	T->G	rs34305195	HBB

标注:

* Chromosome 表示染色体号, Locus 表示 SNP 对应的碱基在染色体上的位点编号, Ref 表示数据库中人参考基因组上对应位点的碱基型别, Blood Cell 表示血液单细胞数据中对应 SNP 位点的型别信息, Mutation 表示数据库中存在的对应位点的突变类型, SNP ID 表示该 SNP 位点在数据库中的 ID 编号, Gene 表示该 SNP 位点位于哪个基因区。

8. 疾病注释

上述过滤和提取的目标基因的 SNP 信息, 进一步做疾病注释。本实施例中, 未在 HBB 基因区检出任何杂合或纯合的变异位点, 说明本实施例中的血液单细胞对应的个体不是 β 地中海贫血疾病患者, 也不是 β 地中海贫血疾病基因携带者, 其 β 地中海贫血疾病基因区为正常基因型。

至此，本发明的实施例实现了一种基于高通量测序对单细胞或微量样本检测孟德尔遗传病（单基因疾病）的方法。

工业实用性

本发明的，能够有效地应用于样品 DNA 的 DNA 测序文库的构建以及测序，并且获得
5 的文库质量好，测序结果准确。

尽管本发明的具体实施方式已经得到详细的描述，本领域技术人员将会理解。根据已经公开的所有教导，可以对那些细节进行各种修改和替换，这些改变均在本发明的保护范围之内。本发明的全部范围由所附权利要求及其任何等同物给出。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示意性实施例”、“示
10 例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

权利要求书

- 1、一种确定个体是否患有异常状态的方法，其特征在于，包括：
针对所述个体的核酸样本，构建测序文库；
对所述测序文库进行测序，以便获得测序结果，所述测序结果由多个测序数据构成；
5 基于所述测序数据，确定所述测序结果中所包含的已知 SNP；以及
基于所述已知 SNP，确定所述个体是否患有与所述已知 SNP 相关的异常状态。
- 2、根据权利要求 1 所述的方法，其特征在于，所述个体为人。
- 3、根据权利要求 1 所述的方法，其特征在于，所述异常状态为疾病。
- 4、根据权利要求 3 所述的方法，其特征在于，所述疾病为单基因疾病。
- 10 5、根据权利要求 4 所述的方法，其特征在于，所述单基因疾病为选自地中海贫血症和红细胞葡萄糖-6-磷酸脱氢酶缺陷症的至少一种。
- 6、根据权利要求 5 所述的方法，其特征在于，所述地中海贫血症为 β -地中海贫血症。
- 7、根据权利要求 1 所述的方法，其特征在于，所述核酸样本为个体的全基因组 DNA 的至少一部分。
- 15 8、根据权利要求 1 所述的方法，其特征在于，所述核酸样本是从个体的单细胞或者微量样本中提取的。
- 9、根据权利要求 8 所述的方法，其特征在于，所述单细胞或者微量样本是从个体的选自血液、组织、尿液、配子、受精卵、卵裂球和胚胎的至少一种分离的。
- 10、根据权利要求 9 所述的方法，其特征在于，所述单细胞是通过选自稀释法、口吸
20 管分离法、显微操作、显微切割、流式细胞分离术、微流控的至少一种分离的。
- 11、根据权利要求 1 所述的方法，其特征在于，针对所述个体的核酸样本，构建测序文库进一步包括：
对所述核酸样本进行扩增，以便得到核酸样本扩增产物；以及
针对所述核酸样本扩增产物，构建所述测序文库。
- 25 12、根据权利要求 11 所述的方法，其特征在于，所述核酸样本为从个体的单细胞中提取的全基因组 DNA，
其中，
针对所述全基因组 DNA 进行扩增是通过选自 PEP-PCR、DOP-PCR、OmniPlex WGA 和 MDA 的至少一种进行的。
- 30 13、根据权利要求 11 所述的方法，其特征在于，在针对所述核酸样本扩增产物，构建所述测序文库之前，进一步包括：

利用核酸探针对所述核酸样本扩增产物进行筛选，以便获得来自预定区域的核酸样本扩增产物；以及

针对所述来自预定区域的核酸样本扩增产物，构建所述测序文库。

14、根据权利要求 13 所述的方法，其特征在于，所述核酸探针是以芯片的形式提供的。

5 15、根据权利要求 13 所述的方法，其特征在于，所述预定区域为至少一个外显子区域。

16、根据权利要求 1 所述的方法，其特征在于，所述测序是利用选自 Illumina Hiseq2000、Genome Analyzer、SOLiD 测序系统、Ion Torrent、Ion Proton、454、PacBio RS 测序系统、Helicos tSMS 技术以及纳米孔测序技术的至少一种进行的。

17、根据权利要求 16 所述的方法，其特征在于，所述测序是利用 Illumina Hiseq2000
10 进行的，所述测序数据的长度为 90bp。

18、根据权利要求 1 所述的方法，其特征在于，所述基于所述测序数据，确定所述测序结果中所包含的已知 SNP，是通过将所述测序数据与参考基因进行比对进行的。

19、根据权利要求 18 所述的方法，其特征在于，所述参考基因是已知的人基因组序列。

20、根据权利要求 18 所述的方法，其特征在于，所述比对是 SOAP/SOAP2 软件进行的。

15 21、根据权利要求 20 所述的方法，其特征在于，进一步包括对所述测序结果中所包含的已知 SNP 进行过滤，所述过滤基于如下过滤条件：

SNP calling 质量值大于 20；

SNP 位点测序深度大于 8；

SNP 位点深度小于基因组平均深度的 5 倍；

20 SNP 位点拷贝数不大于 2；和

SNP 位点与最近的其他 SNP 位点之间的距离大于 5。

22、根据权利要求 1 所述的方法，其特征在于，所述已知 SNP 位于人染色体 HBB 基因区。

23、根据权利要求 22 所述的方法，其特征在于，所述已知 SNP 为选自下列的至少一种：

25 rs33985472、rs63750954、rs63751128、rs33978907、rs34029390、rs34809925、rs33953406、
rs33910569、rs33910569、rs33971634、rs33925391、rs33946267、rs35485099、rs36015961、
rs33930977、rs35256489、rs33952266、rs33952266、rs33952266、rs33914668、rs33914668、
rs33913413、rs33913413、rs34483965、rs34793594、rs35703285、rs63750433、rs34690599、
rs63751175、rs35328027、rs1609812、rs34451549、rs7480526、rs10768683、rs35099082、
30 rs63750283、rs63750283、rs33945777、rs33945777、rs33945777、rs33933298、rs33913712、
rs33995148、rs33931779、rs33969400、rs33922842、rs11549407、rs33974936、rs33991059、

rs33982568、rs33948578、rs1135071、rs33943001、rs33943001、rs63750513、rs34527846、rs35456885、rs35004220、rs63750195、rs35724775、rs33915217、rs33915217、rs33915217、rs33956879、rs33956879、rs33956879、rs33971440、rs33971440、rs33960103、rs33960103、rs35684407、rs35578002、rs33916412、rs35424040、rs33950507、rs33950507、rs33951465、rs33959855、rs33972047、rs33986703、rs34716011、rs63750783、rs35799536、rs33930702、rs33930702、rs33930702、rs33941849、rs33941849、rs33941849、rs34563000、rs34135787、rs34704828、rs63750628、rs34305195 以及位于人染色体 11 上 5246716 位、5246879 位、5247026 位、5248161 位的 SNP。

24、一种确定个体是否患有异常状态的系统，其特征在于，包括：

10 测序文库构建装置，所述测序文库构建装置用于针对所述个体的核酸样本，构建测序文库；

测序装置，所述测序装置与所述测序文库构建装置相连，用于对所述测序文库进行测序，以便获得测序结果，所述测序结果由多个测序数据构成；

SNP 确定装置，所述 SNP 确定装置与所述测序装置相连，用于基于所述测序数据，确定所述测序结果中所包含的已知 SNP；以及

异常状态确定装置，所述异常装置确定装置与所述 SNP 确定装置相连，用于基于所述已知 SNP，确定所述个体是否患有与所述已知 SNP 相关的异常状态。

25、根据权利要求 24 所述的系统，其特征进一步包括：

20 核酸样本提取装置，所述核酸样本提取装置适于从个体的单细胞或者微量样本中提取个体的全基因组 DNA 的至少一部分。

26、根据权利要求 25 所述的系统，其特征进一步包括：

生物样本分离装置，所述生物样本分离装置适于从个体的选自血液、组织、尿液、配子、受精卵、卵裂球和胚胎的至少一种分离单细胞或者微量样本。

27、根据权利要求 26 所述的系统，其特征在于，所述生物样本分离装置适于通过选自 25 稀释法、口吸管分离法、显微操作、显微切割、流式细胞分离术、微流控的至少一种分离单细胞或者微量样本。

28、根据权利要求 24 所述的系统，其特征在于，所述测序文库构建装置进一步包括：

核酸样本扩增单元，所述核酸样本扩增单元适于对所述核酸样本进行扩增，以便得到核酸样本扩增产物。

30 29、根据权利要求 24 所述的系统，其特征在于，所述扩增单元适于进行选自 PEP-PCR、DOP-PCR、OmniPlex WGA 和 MDA 的至少一种。

30、根据权利要求 29 所述的系统，其特征在于，所述测序文库构建装置进一步包括：

筛选单元，所述筛选单元中设置有核酸探针，以便利用所述核酸探针对所述核酸样本扩增产物进行筛选，以便获得来自预定区域的核酸样本扩增产物；以及

针对所述来自预定区域的核酸样本扩增产物，构建所述测序文库。

5 31、根据权利要求 30 所述的系统，其特征在于，所述核酸探针是以芯片的形式提供的。

32、根据权利要求 24 所述的系统，其特征在于，所述测序装置为选自 Illumina HiSeq2000、Genome Analyzer、SOLiD 测序系统、Ion Torrent、Ion Proton、454、PacBio RS 测序系统、Helicos tSMS 测序装置以及纳米孔测序装置的至少一种。

33、根据权利要求 24 所述的系统，其特征在于，所述 SNP 确定装置进一步包括：

10 比对单元，所述比对单元用于通过将所述测序数据与参考基因进行比对确定所述测序结果中所包含的已知 SNP。

34、根据权利要求 33 所述的系统，其特征在于，所述比对单元适于利用 SOAP/SOAP2 软件进行比对。

35、根据权利要求 33 所述的系统，其特征在于，所述 SNP 确定装置进一步包括：

15 SNP 过滤单元，所述 SNP 过滤单元适于基于如下过滤条件，对所述测序结果中所包含的已知 SNP 进行过滤：

SNP calling 质量值大于 20；

SNP 位点测序深度大于 8；

SNP 位点深度小于基因组平均深度的 5 倍；

20 SNP 位点拷贝数不大于 2；和

SNP 位点与最近的其他 SNP 位点之间的距离大于 5。

36、根据权利要求 24 所述的系统，其特征在于，所述已知 SNP 位于人染色体 HBB 基因区。

37、根据权利要求 36 所述的系统，其特征在于，所述已知 SNP 为选自下列的至少一种：

25 rs33985472、rs63750954、rs63751128、rs33978907、rs34029390、rs34809925、rs33953406、rs33910569、rs33910569、rs33971634、rs33925391、rs33946267、rs35485099、rs36015961、rs33930977、rs35256489、rs33952266、rs33952266、rs33952266、rs33914668、rs33914668、rs33913413、rs33913413、rs34483965、rs34793594、rs35703285、rs63750433、rs34690599、rs63751175、rs35328027、rs1609812、rs34451549、rs7480526、rs10768683、rs35099082、
30 rs63750283、rs63750283、rs33945777、rs33945777、rs33945777、rs33933298、rs33913712、rs33995148、rs33931779、rs33969400、rs33922842、rs11549407、rs33974936、rs33991059、

rs33982568、rs33948578、rs1135071、rs33943001、rs33943001、rs63750513、rs34527846、
rs35456885、rs35004220、rs63750195、rs35724775、rs33915217、rs33915217、rs33915217、
rs33956879、rs33956879、rs33956879、rs33971440、rs33971440、rs33960103、rs33960103、
rs35684407、rs35578002、rs33916412、rs35424040、rs33950507、rs33950507、rs33951465、
5 rs33959855、rs33972047、rs33986703、rs34716011、rs63750783、rs35799536、rs33930702、
rs33930702、rs33930702、rs33941849、rs33941849、rs33941849、rs34563000、rs34135787、
rs34704828、rs63750628、rs34305195 以及位于人染色体 11 上 5246716 位、5246879 位、5247026
位、5248161 位的 SNP。

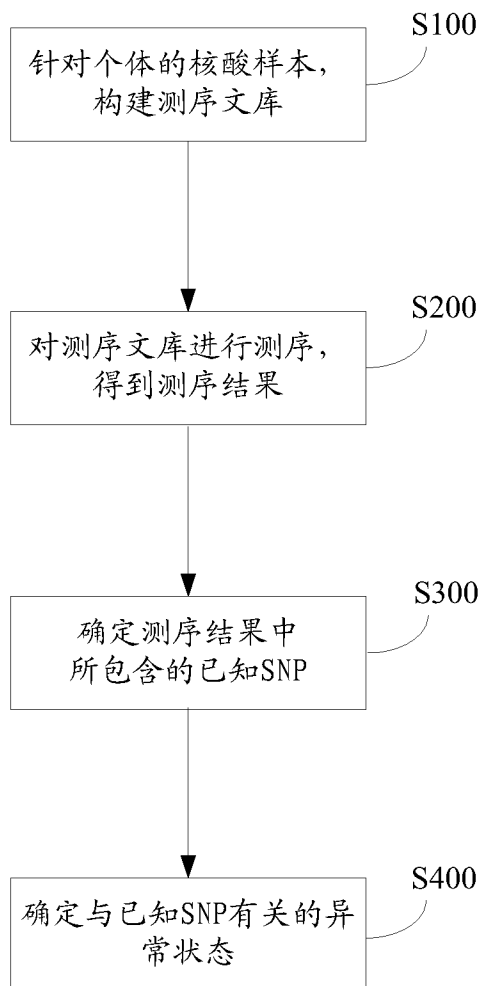


图 1

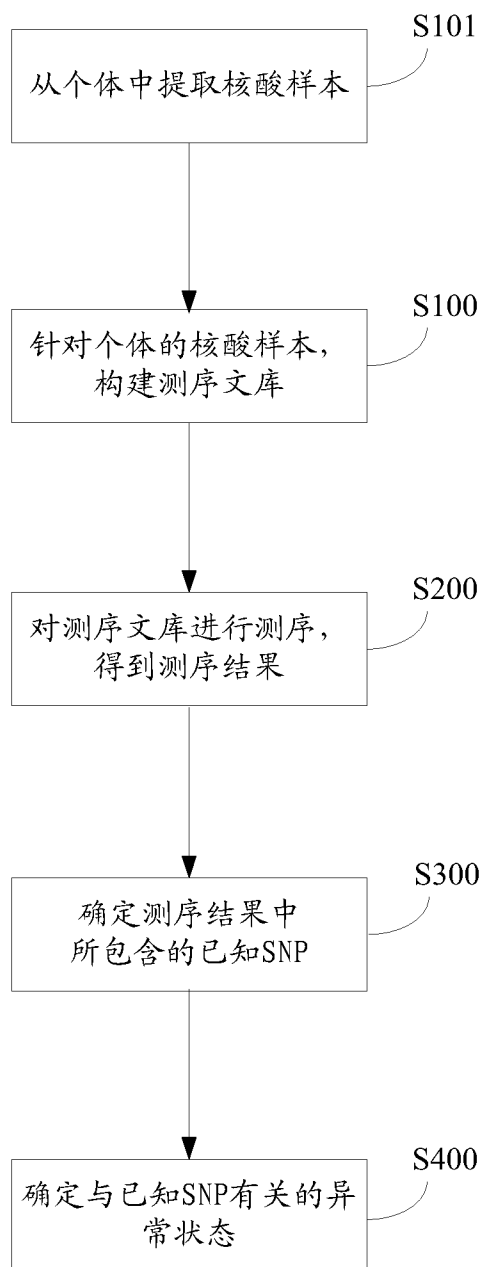


图 2

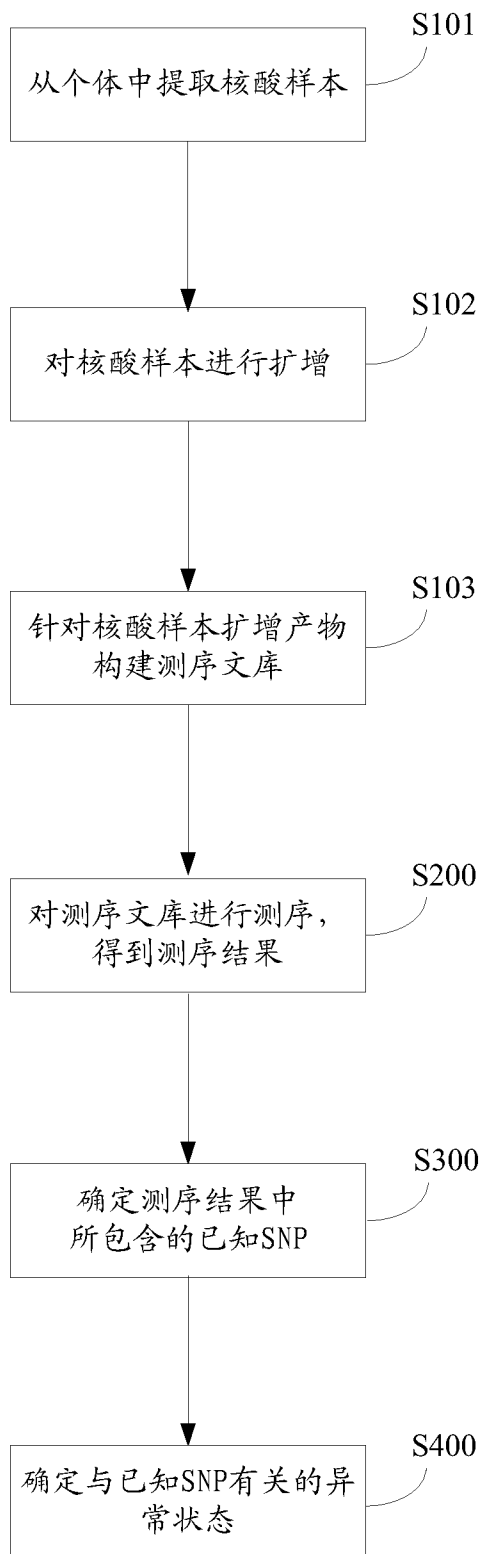


图 3

1000



图 4

1000



图 5

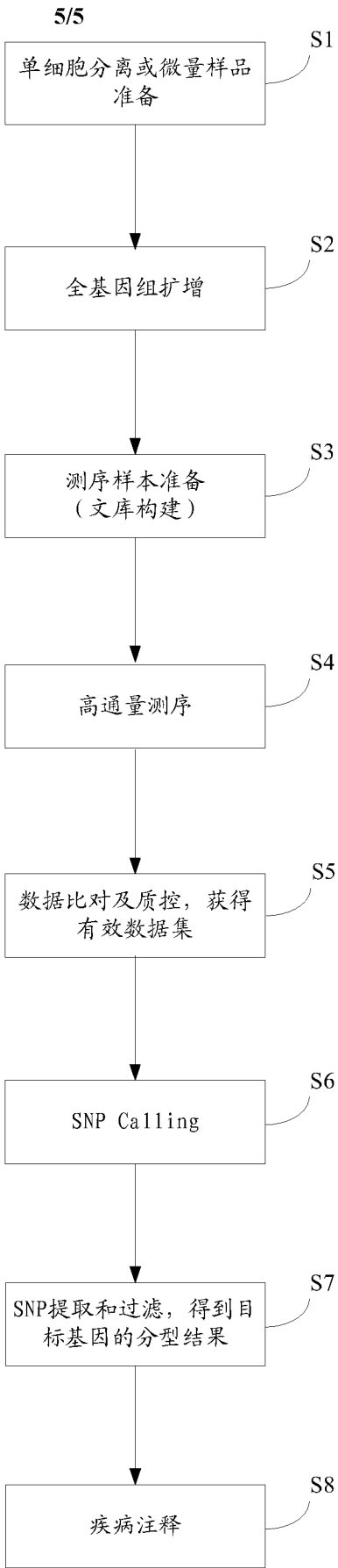


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/080500**A. CLASSIFICATION OF SUBJECT MATTER**

C12Q 1/68 (2006.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12Q

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

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

CNKI, CNABS, NCBI, VEN, ISI-WEB OF KNOWLEDGE, GOOGLE SCHOLAR: SNP, glucose-6-phosphate dehydrogenase deficiency, database establishment, PCR, comparison, filter, monogenic diseases, variant, single nucleotide polymorphism, rhalassemia, cooley anemia, mediterranean anemia, thalassaemia, G6PD, glucose, phosphate, dehydrogenase, library, sequencing, probe, chip, align, discard, exclude

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WEI, Xiaoming et al. Identification of Sequence Variants in Genetic Disease-Causing Genes Using Targeted Next-Generation Sequencing. PLOS ONE. 21 December 2011, vol. 6, no. 12, pages 1-10, see the whole document, particularly abstract, and page 2, right column, page 3, left column, page 5, and page 6, right column	1-37

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

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

Date of the actual completion of the international search 22 May 2013 (22.05.2013)	Date of mailing of the international search report 06 June 2013 (06.06.2013)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer TIAN, Yuan Telephone No.: (86-10) 62411047

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/080500

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LAM, K.W.G. et al. Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted Massively Parallel Sequencing of Maternal Plasma: Application to β Thalassemia. CLINICAL CHEMISTRY. 15 August 2012, vol. 58, no. 10, pages 1-9, see the whole document, particularly abstract, and pages 2-5	1-37
A	CN 101374963 A (KEYGENE N.V.), 25 February 2009 (25.02.2009), see the whole document	1-37
A	SABATH, D.E. et al. A Multiplex Approach to the Molecular Diagnosis of B-Thalassemia. THE JOURNAL OF MOLECULAR DIAGNOSTICS. 2011, vol. 13, no. 4, pages 369-370, see the whole document	1-37

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2012/080500

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101374963 A	25.02.2009	EP 1966393 A1	10.09.2008
		US 2009269749 A1	29.10.2009
		EP 1966393 B1	25.07.2012
		ES 2391837 TT3	30.11.2012
		US 2013059739 A1	07.03.2013
		IN 200802660 P2	30.01.2009
		EP 2363504 A1	07.09.2011
		WO 2007073165 A1	28.06.2007
		JP 2009520497 A	28.05.2009

A. 主题的分类 C12Q 1/68 (2006.01) i 按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) IPC: C12Q 包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNKI, CNABS, NCBI, VEN, ISI-WEB OF KNOWLEDGE, GOOGLE SCHOLAR: 单核苷酸多态性, SNP, 单基因病, 地中海贫血, 磷酸脱氢酶缺乏, 文库, 建库, 测序, PCR, 探针, 芯片, 比对, 过滤, monogenic diseases, variant, single nucleotide polymorphism, rhalassemia, cooley anemia, mediterranean anemia, thalassaemia, G6PD, glucose, phosphate, dehydrogenase, library, sequencing, probe, chip, align, discard, exclude		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	Xiaoming Wei 等. Identification of Sequence Variants in Genetic Disease-Causing Genes Using Targeted Next-Generation Sequencing. PLoS ONE.21.12 月 2011, 第 6 卷, 第 12 期, 第 1-10 页 参见全文, 特别是摘要、第 2 页右栏、第 3 页左栏、第 5 页和第 6 页右栏	1-37
☒ 其余文件在 C 栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 22.5 月 2013 (22.05.2013)		国际检索报告邮寄日期 06.6 月 2013 (06.06.2013)
ISA/CN 的名称和邮寄地址: 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号: (86-10)62019451		受权官员 田园 电话号码: (86-10) 62411047

C(续). 相关文件		
类 型	引用文件, 必要时, 指明相关段落	相关的权利要求
X	Kwan-Wood G. Lam等. Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted Massively Parallel Sequencing of Maternal Plasma: Application to β Thalassemia. Clinical Chemistry.15.8月2012, 第58卷, 第10期, 第1-9页 参见全文, 特别是摘要、第 2-5 页	1-37
A	CN101374963 A (凯津公司) 25.2 月 2009 (25.02.2009) 参见全文	1-37
A	Daniel E. Sabath 等. A Multiplex Approach to the Molecular Diagnosis of β -Thalassemia. The Journal of Molecular Diagnostics. 2011, 第 13 卷, 第 4 期, 第 369-370 页 参见全文	1-37

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2012/080500

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN101374963 A	25.02.2009	EP1966393 A1	10.09.2008
		US2009269749 A1	29.10.2009
		EP1966393B1	25.07.2012
		ES2391837T T3	30.11.2012
		US2013059739 A1	07.03.2013
		IN200802660P2	30.01.2009
		EP2363504 A1	07.09.2011
		WO2007073165 A1	28.06.2007
		JP2009520497 A	28.05.2009