

(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

2012 年 1 月 5 日 (05.01.2012)

(10) 国际公布号

W O 2012/000152 A I

- (51) 国际分类号 :
C12Q 1/68 (2006.01)
- (21) 国际申请号 : PCT/CN20 10/00 1834
- (22) 国际申请日 : 2010 年 11 月 15 日 (15.11.2010)
- (25) 申报语言 : 中文
- (26) 公布语言 : 中文
- (30) 优先权 :
201010213717.6 2010 年 6 月 30 日 (30.06.2010) CN
- (71) 申请人 (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, Guangdong 518083 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): 李剑 (LI, Jian) [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, Guangdong 518083 (CN)。刘涛 (LIU, Tao) [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, Guangdong 518083 (CN)。赵美茹 (ZHAO, Meiru) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。张现东 (ZHANG, Xiandong) [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, Guangdong 518083 (CN)。
- (74) 代理人: 中国国际贸易促进委员会专利商标事务所 (CCPIT PATENT AND TRADEMARK LAW OFFICE); 中国北京市阜成门外大街 2 号万通新世界广场 8 层, Beijing 100037 (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, 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, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, 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))。
- 包括说明书序列表部分 (细则 5.2(a))。

(54) Title: PCR-SEQUENCING METHOD BASED ON TECHNOLOGY OF DNA MOLECULAR INDEX AND STRATEGY OF DNA-BREAKING INCOMPLETELY

(54) 发明名称 : 一种基于 DNA 分子标签技术和 DNA 不完全打断策略的 PCR 测序方法

(57) Abstract: The present invention provides a sequencing method for nucleic acid, which method comprises: amplifying the target nucleic acid using indexed primers, breaking the amplicons incompletely and purifying and recovering mixed DNAs, sequencing the mixed DNAs by a next-generation sequencing method to obtain the sequences of the broken DNAs, and assembling the sequences of the broken DNAs into a complete sequence for the target nucleic acid. The present invention further provides primer indexes for the sequencing method.

(57) 摘要 :

本发明提供了核酸的测序方法, 该方法包括: 用标签引物扩增靶核酸; 不完全打断扩增子并纯化回收 DNA 混合物; 通过第二代测序方法测序 DNA 混合物获得打断后的 DNA 序列; 以及将打断后的 DNA 的序列拼接成靶核酸的完整序列。本发明还提供了用于该测序方法的引物标签。



2012 000152 1

一种基于 DNA 分子标签技术和 DNA 不完全打断策略的 PCR 测序方法

技术领域

本发明涉及核酸测序技术领域，特别是 PCR 测序技术领域。另一方面，本发明还涉及 PCR-index/barcode 技术领域。本发明的方法特别适用于第二代测序技术，尤其是第二代测序技术中的 Pair-end 测序技术，还可以用于 HLA 基因分型。

背景技术

PCR 测序 (PCR sequencing) 方法是通过 PCR 的方法得到目的基因 DNA 片段，再通过对所得到的目的基因 DNA 片段进行 DNA 序列检测，从而得到目的基因 DNA 序列信息的一种技术，其具有直观、准确的特点，长久以来广泛应用于基因突变检测以及基因分型等领域。

PCR-index/barcode 技术通过在 PCR 引物的 5' 末端添加引物标签 (primer index) 序列，可在 PCR 过程中对每个样本引入独特的引物标签，使样本在利用第二代 DNA 测序技术 (以 Illumina GA, Roche 454, ABI Solid 等测序仪为代表的可大规模，高通量，单分子测序的测序仪) 检测过程中，除 PCR 环节必须逐个样本处理外，其它实验环节可把多个样本混在一起同时处理，最终每个样本的检测结果可以通过其独特的引物标签序列找回；该方法具有成本低，通量大和可同时检测大量样本的多个不同基因位点的特点。

第二代 DNA 测序技术与第一代 (Sanger 法) DNA 测序技术相比，其可连续测序的长度较短。以 Illumina GA (Illumina 公司的

Genome Analyzer 测序仪,简称 Illumina GA) 为例,当前 Illumina GA 的最大测序长度为 200bp,当 PCR 产物的长度大于 200bp,利用 Illumina GA 测序,直接采取 PCR 测序的方法,不能把整个 PCR 产物测通,得不到被检测 PCR 产物的全部 DNA 序列信息的。短的测序读长限制了第二代测序技术的应用,除了通过逐步改进测序技术来获得更长的实际测序读长外,开发可克服第二代 DNA 测序仪现有测序读长在 PCR 测序应用领域不足的新技术也成为当务之急。

发明内容

当前利用第二代测序技术,对大量样本中的特定基因相关序列同时进行测序分析时,一般会采用 PCR 测序的策略,直接利用引物标签 + 二代测序技术的组合。当所用测序仪的测长可以覆盖整个 PCR 产物的长度时,上述技术就足够满足要求了。但当所用测序仪的测长不够覆盖整个 PCR 产物的长度时,要么更换具有更长测长的第二代测序仪(如用 Roche 454 GS-FLX)替代 Illumina GA,如果测长还不满足要求的话,只能牺牲成本和通量,改用第一代测序仪。

现实情况是, Illumina GA 具有超高测序通量,但其测长才 200bp; Roche 454 GS-FLX 的测长虽然可以达到 500bp,但其测序成本较高,通量较小;第一代测序仪的测长虽然可以达到 1000bp 以上,但其通量和成本没法和第二代测序仪相比。

有没有能够兼顾成本和通量,能提高测序仪可以测通的 PCR 产物长度的技术呢?本专利中的引物标签 + DNA 不完全打断策略 + 二代测序技术的组合能在充分利用第二代测序仪高通量、低成本特点的同时,使测序仪能够测通的 PCR 产物长度达到测序仪本身测长以上,大大扩大了其适用范围。其中,本发明的所采用的

二代测序技术包括第二代测序技术中的 Pair-end 测序技术，以及针对 PCR 模板有 DNA 参考序列 (DNA Reference Sequence) 的 PCR 测序技术。

本发明提供了用于 PCR 测序的方法，通过所述方法减少了自身测序读长较短造成的限制，扩大了第二代 DNA 测序技术在 PCR 测序应用领域的应用。

引物标签的设计根据所应用的实验平台不同而不同，考虑 Illumina GA 测序平台本身的特点，本发明在设计引物标签时主要考虑了以下几点：1：引物标签序列中避免 3 个以上（包括 3 个）单碱基重复序列，2：所有引物标签的同一位点中碱基 A 和碱基 C 的总含量占有所有碱基含量的 30%-70% 之间，3：引物标签序列本身的 GC 含量在 40-60% 之间，4：引物标签之间序列差异度大于 4 个碱基，5：引物标签序列中避免出现与 Illumina GA 测序引物相似度高的序列，6：减少引物标签序列添加到 PCR 引物上后，对 PCR 引物造成的严重发卡 (hairpin)，二聚体 (dimer) 情况的出现。

本发明通过 PCR 反应在 PCR 产物两端各引入一个引物标签（引物标签序列可以相同也可以不同），使 PCR 产物任何一端的引物标签都可以特异的标记 PCR 产物的样本信息。所得 PCR 产物经过不完全打断。所谓不完全打断，指打断终产物中包含完整的未被打断的 PCR 产物和局部打断的 PCR 产物。所述打断方法包括但不限于化学打断方法（例如酶切）和物理打断方法，所述物理打断方法包括超声波打断方法或机械打断方法等。打断的 DNA 经 2% 琼脂糖电泳，割胶纯化回收从测序仪最大读取长度到测序仪可适用的最长 DNA 长度范围之间的所有 DNA 条带（Illumina GA 测序仪可适用的最长 DNA 为 700bp，此长度为原始 DNA 长度，没

有包括文库接头序列长度)。纯化回收方法包括但不限于电泳割胶回收,也可以是磁珠回收等。回收的DNA片段再按照第二代测序仪测序文库构建的流程构建测序文库,然后进行测序,优选为采用PCR-FREE测序文库构建的流程构建测序文库,测序方法优选采用Pair-End方法测序。PCR-Free测序文库构建的流程按照本领域技术人员已知方法构建。在得到测序总数据中,通过引物标签序列可以找到所有所测样本的序列信息,利用BWA把各个测序序列定位到PCR产物的相应DNA参考序列(Reference Sequence)上,再通过测序序列之间的重叠和连锁关系拼接出PCR产物的完整序列(图1)。此处连锁是指由Pair-End测序特点决定的pair-end连锁关系。

"接头(adapter)"或"文库接头(library adapter)"标签技术是指通过对多个测序文库添加不同文库接头(不同文库接头的组成序列不同,序列不同的部分称为接头标签(adapter index)),构建标签测序文库,从而可实现多个不同标签测序文库混合测序,且最终各个标签测序文库的测序结果可相互区分的一种文库标签技术。

基于DNA分子标签技术和DNA不完全打断策略的PCR测序方法的使用可在不增加引物标签数目的情况下,大大提高可唯一标记的样本数目。

结合文库接头标签技术的PCR-FREE的文库构建方法,是指将文库接头直接连接至测序文库中的DNA片段两端,文库接头的导入过程因为没有PCR的参与,因此称作PCR-Free文库构建。其中接入方法可以采用DNA连接酶进行连接。其整个文库构建过程中无PCR的参与,避免了在高序列相似度的PCR产物混合(pooling)文库的构建过程中,由PCR引入错误而导致最后结

果的不准确性。

在本发明中,通过对正反 PCR 引物添加引物标记,结合 DNA 不完全打断策略,使第二代测序技术应用于 PCR 测序领域时,实际可测得的 PCR 产物长度超过测序仪的最大测序长度。

在本发明的一个方面中,提供了一组引物标签 (primer index),其包括表 1 所示 95 对引物标签中的至少 10 对,或至少 20 对,或至少 30 对,或至少 40 对,或至少 50 对,至少 60 对,或至少 70 对,或至少 80 对,或至少 90 对,或 95 对 (或者所述一组引物标签由表 1 所示 95 对引物标签中的 10 - 95 对 (例如 10 - 95 对,20 - 95 对,30 - 95 对,40 - 95 对,50 - 95 对,60 - 95 对,70 - 95 对,80 - 95 对,90 - 95 对,或 95 对)组成),并且

所述一组引物标签优选地至少包括表 1 所示 95 对引物标签中的 PI-1 至 PI-10, 或 PI-11 至 PI-20, 或 PI-21 至 PI-30, 或 PI-31 至 PI-40, 或 PI-41 至 PI-50, 或 PI-51 至 PI-60, 或 PI-61 至 PI-70, 或 PI-71 至 PI-80, 或 PI-81 至 PI-90, 或 PI-91 至 PI-95, 或者它们任何两个或者多个的组合。

根据本发明另一方面,还提供了所述的引物标签用于 PCR 测序方法的用途,其中特别是,每一对引物标签与用于扩增待测目的序列的 PCR 引物对组合成一对标签引物,正反 PCR 引物的 5 端分别具有 (或者任选通过连接序列连接)正向引物标签和反向引物标签。

在本发明的一个具体实施方式中,所述 PCR 引物是用于扩增 HLA 的特定基因的 PCR 引物,优选是用于扩增 HLA-A/B 2, 3, 4 号外显子和 HLA-DRB1 2 号外显子的 PCR 引物,优选的所述 PCR 引物如表 2 所示。

本发明另一方面中，提供了上文所述一组引物标签与用于扩增待测目的序列的 PCR 引物对组合成的一组标签引物，其中每一对引物标签与 PCR 引物对组合成一对标签引物，正反 PCR 引物的 5' 端分别具有（或者任选通过连接序列连接）正向引物标签和反向引物标签。

在本发明的一个具体实施方式中，上文所述标签引物中的 PCR 引物是用于扩增 HLA 的特定基因的 PCR 引物，优选是用于扩增 HLA-A/B 2,3,4 号外显子和 HLA-DRB1 2 号外显子的 PCR 引物，优选的所述 PCR 引物如表 2 所示。在本发明的另一个具体实施方式中，所述的标签引物用于 PCR 测序方法。

本发明另一方面中，提供了一种测定样品中目的核酸的核苷酸序列的方法，其包括：

1) 提供 n 个样品， n 为大于等于 1 的整数，所述样品优选地来自哺乳动物，更优选是人，特别是人的血样；可选地，将待分析的 n 个样品分成 m 个小组， m 为整数且 $n \geq m \geq 1$ ；

2) 扩增：对于每一个样品，使用一对标签引物，在存在来自该样品的模板时，在适于扩增目的核酸的条件下进行 PCR 扩增，其中，每一对标签引物由包含引物标签的正向标签引物和反向标签引物（均可以是简并引物）构成，其中正向标签引物和反向标签引物所包含的引物标签可以相同或者不同；不同样品所用标签引物对中的引物标签彼此不同；

3) 打断：将所得的 PCR 产物文库进行不完全打断；

4) 测序：将回收的 DNA 混合物利用二代测序技术，优选的是 Pair-End 技术（例如 Illumina GA、Illumina HiSeq 2000）进行测序，获得打断后的 DNA 的序列；

5) 拼接：基于每个样品独特的引物标签将获得的测序结果与

样品一一对应，利用比对程序（例如 Blast, BWA 程序）把各个测序序列定位到 PCR 产物的相应 DNA 参考序列上，通过序列重叠和连锁关系，从打断后的 DNA 的序列拼接出完整的目的核酸。

本发明另一方面，提供了一种测定样品中目的核酸的核苷酸序列的方法，其包括：

1) 提供 n 个样品， n 为大于等于 1 的整数，所述样品优选地来自人，特别是人的血样；可选地，将待分析的 n 个样品分成 m 个小组， m 为整数且 $n \geq m \geq 1$ ；

2) 扩增：对于每一个样品，使用一对标签引物，在存在来自该样品的模板时，在适于扩增目的核酸的条件下进行 PCR 扩增，其中，每一对标签引物由包含引物标签的正向标签引物和反向标签引物（均可以是简并引物）构成，其中正向标签引物和反向标签引物所包含的引物标签可以相同或者不同；不同样品所用标签引物对中的引物标签彼此不同；

3) 混合：将各样品的 PCR 扩增产物混合在一起，获得 PCR 产物文库；

4) 打断：将所得的 PCR 产物文库进行不完全打断；

5) 建库：将打断后的 PCR 产物文库构建 PCR-Free 测序文库，回收位于所用测序仪最大读长长度到所用测序仪适用的最长 DNA 长度范围之间的所有 DNA 条带，可以对文库添加不同的文库接头（adapter）以区分不同的 PCR-Free 测序文库；

6) 测序：将回收的 DNA 混合物利用二代测序技术，优选的是 Pair-End 技术（例如 Illumina GA、Illumina Hiseq 2000）进行测序，获得打断后的 DNA 的序列；

7) 拼接：基于各个文库不同的文库接头序列和每个样品独特的引物标签将获得的测序结果与样品一一对应，利用比对程序（例

如 Blast BWA 程序) 将各个测序序列定位到 PCR 产物的相应 DNA 参考序列上 , 通过序列重叠和连锁关系 , 从打断后的 DNA 的序列拼接出完整的 目的核酸。

在本发明的一个具体实施方式中 , 在上文所述的测序方法中 , 每一对引物标签与 PCR 引物对组合成一对标签引物 , 正反 PCR 引物的 5 端分别具有 (或者任选通过连接序列连接) 正向引物标签和反向引物标签。

在本发明的一个具体实施方式中 , 在上文所述的测序方法中 , 所述 PCR 引物是用于扩增 HLA 的特定基因的 PCR 引物 , 优选是用于扩增 HLA-A/B 的 2 , 3 , 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 引物 , 优选的所述 PCR 引物如表 2 所示。

在本发明的一个具体实施方式中 , 在上文所述的测序方法中 , 所述引物标签针对 PCR 引物进行设计 , 优选针对用于扩增 HLA 的特定基因的 PCR 引物进行设计 , 更优选针对用于扩增 HLA-A/B 的 2 , 3 , 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 引物 , 特别是如表 2 所示的 PCR 引物进行设计 , 所述引物标签特别是包括表 1 所示 95 对引物标签中的至少 10 对 , 或至少 20 对 , 或至少 30 对 , 或至少 40 对 , 或至少 50 对 , 至少 60 对 , 或至少 70 对 , 或至少 80 对 , 或至少 90 对 , 或 95 对 (或者所述一组引物标签由表 1 所示 95 对引物标签中的 10 - 95 对 (例如 10 - 95 对 , 20 - 95 对 , 30 - 95 对 , 40 - 95 对 , 50 - 95 对 , 60 - 95 对 , 70 - 95 对 , 80 - 95 对 , 90 - 95 对 , 或 95 对) 组成) , 并且

所述一组引物标签优选地至少包括表 1 所示 95 对引物标签中的 PI-1 至 PI-10, 或 PI-11 至 PI-20, 或 PI-21 至 PI-30, 或 PI-31 至 PI-40, 或 PI-41 至 PI-50, 或 PI-51 至 PI-60, 或 PI-61 至 PI-70, 或 PI-71 至 PI-80, 或 PI-81 至 PI-90, 或 PI-91 至 PI-95, 或者它

们任何两个或者多个的组合。

在本发明的一个具体实施方式中,在上文所述的测序方法中,所述 DNA 打断包括化学打断方法和物理打断方法,其中所述化学方法包括酶切方法,所述物理打断方法包括超声波打断方法或机械打断方法。所述 DNA 打断后,分离 450-750bp 长度的片段。

本发明另一方面中,提供了一种 HLA 分型方法,包括:使用上文所述的测序方法对来自患者的样品(特别是血样)进行测序,以及将测序结果与 HLA 数据库(如 IMGT HLA 专业数据库)中的标准序列数据比对,序列比对结果 100% 匹配的即为对应样本的 HLA 基因型别。

附图说明

图 1: 为引物标签标记, DNA 打断和 DNA 测序后,序列拼接图示。第 N 号样本的 PCR 产物两端引入了正反引物标签序列 Index-N-F/R (1), PCR 产物经物理方法打断后的产物中包括:一端带有引物标签序列的产物,两端都不带引物标签序列的产物,完全未被打断的产物,割胶纯化回收位于测序仪最大读长长度到测序仪适用的最长 DNA 长度范围之间的所有 DNA 条带用于测序 (2), 利用 Index-N-F/R 在测序结果中找回属于第 N 号样本的 PCR 产物的测序结果,利用 PCR 产物已知的参考序列信息定位各个测序序列相对参考的位置,并根据测序序列之间的重叠和连锁关系组装成完整的 PCR 产物的测序结果 (3, 4)。

图 2: 为 1 号样本 HLA-A/B/DRB1 相应外显子 PCR 产物电泳结果,从电泳图上看,PCR 产物为一系列片段大小 300bp-500bp 的单一条带,其中泳道 M 是分子量标记物 (DL 2000, Takara 公司),泳道 1-7 为 1 号样本的 HLA-A/B/DRB1 各外显子 (A2、A3、A4、

B2、B3、B4、DRBI-2) PCR 扩增产物 , 阴性对照 (N) 无扩增条带。其它样品的结果与此类似。

图 3: 为 HLA-Mix 打断后 DNA 电泳情况(割胶前后), 割胶区域为 450-750bp 区域。其中泳道 M 是分子量标记物 (NEB-50bp DNA Ladder), 泳道 1 是割胶前 HLA-Mix 的电泳情况, 泳道 2 是割胶后 HLA-Mix 的胶图。

图 4: 1 号样本的一致性 (consensus) 序列构建程序截图, 示例说明了根据引物标签和 DNA 片段之间的重叠关系拼接出 PCR 产物的完整序列。

具体实施方式

下面将结合实施例对本发明的实施方案进行详细描述, 但是本领域技术人员将会理解, 下列实施例仅用于说明本发明, 而不应视为限定本发明的范围。

在本发明的实施例中, 通过采取引物标签 + DNA 不完全打断策略 + Illumina GA 测序仪 Pair-End 100 测序技术的组合对 95 个样本的 HLA-A/B 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子的基因分型 (PCR 产物长度大小处于 290bp-500bp 之间), 证明该策略可以在充分发挥第二代测序仪高通量、低成本特点的同时, 可以实现对超过测序仪本身测长以上的基因片段的分型。

原理: 将待分析的样本, 通过 PCR 反应在 HLA-A/B 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 产物两端引入引物标签, 使其特异的标记 PCR 产物的样本信息。将各组内样品的 HLA-A/B/DRB1 三个位点的 PCR 扩增产物混合在一起, 获得 PCR 产物文库; 所得 PCR 产物文库经过超声不完全打断后, 构建 PCR-Free 测序文库, 测序文库经 2% 低熔点琼脂糖电泳, 割胶纯

化回收位于 450bp-750bp 长度范围之间的所有 DNA 条带 (PCR-Free 测序文库的构建过程中在 DNA 片段的两端都添加上了文库接头,使 DNA 片段在电泳图上体现的长度比实际长度大了 250bp 左右,因此此处回收 450bp-700bp 的片段,实际上相当于回收原长度为 200bp-500bp 的 DNA 片段)。回收的 DNA 经 Illumina GA PE-100 测序。通过引物标签序列可以找到所有所测样本的序列信息,再通过已知 DNA 片段的参考序列信息和 DNA 片段序列之间的重叠和连锁关系组装出整个 PCR 产物的序列,再通过与 HLA-A/B/DRB1 相应外显子的标准数据库的比对结果可组装出原 PCR 产物的全序列实现 HLA-A/B/DRB1 的基因分型。

实施例 1

样本提取

使用 KingFisher 自动提取仪 (美国 Thermo 公司)从 95 份已知 HLA-SBT 分型结果的血样 (中国造血干细胞捐献者资料库 (以下称 "中华骨髓库"))中提取 DNA。主要步骤如下:取出 6 个 Kingfisher 自动提取仪配套的深孔板及 1 个浅孔板,根据说明书分别加入一定量配套的试剂并做好标记,将所有已加好试剂的孔板按要求置于相应的位置,选定程序 "Bioeasy -200ul Blood DNA -KF.ms2" 程序,按下 "star" 执行该程序进行核酸提取。程序结束后收集 plate Elution 中的 100ul 左右的洗脱产物即为提取的 DNA。

实施例 2

PCR 扩增

通过合成在 5',末端具有不同引物标签的 PCR 引物制作不同

的 PCR 标签引物，这样不同的 PCR 标签引物可以用于不同的样本，所述 PCR 引物是针对 HLA-A/B 的 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 引物。其后通过 PCR 反应在 PCR 产物两端引入引物标签，从而特异地标记了来自不同样本的 PCR 产物。

以 95 套 PCR 标签引物来分别扩增 95 份 DNA 样本，每套 PCR 标签引物由一对双向引物标签（表 1）和用于扩增 HLA-A/B 的 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 引物（表 2）组成，其中每个正向 PCR 引物的 5' 末端上连接一对引物标签的正向引物标签，而反向 PCR 引物的 5' 末端上连接一对引物标签的反向引物标签。引物标签在引物合成时直接添加在 PCR 引物的 5' 末端。

把实施例 1 的样本提取步骤中所得的 95 份 DNA，依次编号 1-95 PCR 反应在 96 孔板中进行，共 7 板，编号分别为 HLA-P-A2、HLA-P-A3、HLA-P-A4、HLA-P-B2、HLA-P-B3、HLA-P-B4 以及 HLA-P-DRB1-2（A2/3/4、B2/B3/B4、DRB1-2 表示扩增的位点），板内设置一个不添加模板的阴性对照，阴性对照所用引物与模板 1 的对应引物相同。实验的同时，记录下每对引物标签对应的样本编号信息。

表 1，引物标签的相关信息

引物标签 编号	正向引物标签	反向引物标签	对应 96 孔板 位置	对应模板 (组 1)
PI-1	TCGCAGACATCA	TGACACGATGCT	A1	1
PI-2	TACATCGCACTA	TACAGATGCTGA	A2	2
PI-3	CTCGATGAGTAC	ACGTCTAGACAC	A3	3

PI-P	TCTGTACTCA	TGCTGTAG $\pm$ QAC	V4		4
PI-S	TAT $\pm$ GTCTCATV	AGATATTCGAGCT	V5		S
PI-6	1ACVXGC IGAGC	ACGTGTCTATCA	V6		6
PI-7	TCATATCGCGAT	AGV ICG1ATVGC	A7		7
PI-8	ACAGATGCACGC	ATCTCGTGACAG	V8		8
PI-9	TAGATCGTACAT	AC IAGTVACGC	V9		9
PI-10	AC IACVCG1C IC	V1AG1CAGCGX	V10		10
PI-11	AGVC ICGCG1VX	TACTVAGCTGACG	V11		11
PI-Z	V $\pm$ ACTVAG $\pm$ GCTC	TGTATCGTGCTC	V12		12
PI-13	CACGAXGACV IC	TAGTGA(GCGCA	E1		13
AI-u	TGCTGTCTCGAG	CATVAGCAVGTG	B2		14
PI-Si	TGTGCTCGAGTC	TCTDYTCGAGCA	B3		15
PI-16	CACTCG1ACVXC	AGCGA IGTCTAT	B4		16
PI-17	CDYCD1DC1CQC	CDCD IAC IDCAD	B5		17
PI-18	ACGCV1C IXXVC	C1VGXV1CGCAG	B6		18
PI-19	CGAGATGACTCT	TGTATVACCGAT	SZ		19
PI-20	AC1G IC ICGVGC	ACGTACCGCACA	B8		20
PI-21	CACTGCTATVAG	TCTVAGCTCATVAG	89		21
PI-ZZ	ACGCV IC1VGA	C1V IGCVXGVI	EO		22
PI-Z3	TGAGATACAGTA	ATCTGCTVTVAG	ai1		23
PI-Z4	ACTCATCGTGCT	TAGVAGCTGTCAVC	Biz		24
PI-Sr	1ACVXGXC IXX	CAGCVCXVAGV I	D1		25
PI-26	CACAGTACTCGC	CTGC IAG1G1VX	DZ		26
PI-Z7	TGTACTATCATA	TDXG1TVAGVACVC	C3		27
PI-8	C1VGICTGVCG	AGCGAGTCTACT	CD		28

PI-29	1 AG AC IG AGCXA	ACV1AC IG AGAC	S3	Z9
PI-30	CAGACDCDIDYG	1ACVXCXCG1AX	C6	30
PI-31	CGGACATCAGG	TAGCGV TG AGAC	D1	31
PI-e1	ACACXCXVAGVX	CTVTCV TG ACAC	C8	32
PI-33	AGCGTATACTAG	CAV1AC1CAGCXA	09	33
PI-34	TGTCCTTCTVTC	ACATGACTCAGG	C10	34
PI-35	CGCTAGACTGTA	XAC1AXVAGXCGA	C11	35
PI-36	ACAG1G1XVGGCG	XGAV1V1GCIACV	C12	36
PI-37	CACTCTATCGAC	XCAGCGGAV1GAG	01	37
PI-38	ACACIG1VAGICV	ACG1AGV1G1VX	az	38
PI-39	CATATGAGATCG	AGCAGAGTGCTC	03	39
PI-P0	CAGCXVXCVA1AC	CACTDCA DACGA	D4	P0
PI-P1	1V1VXCXVGV1	TCA1YDYDCDC	as	P1
PI-ZP	TDTA1GC1CPTC	1CG1DACA DVTC	a6	ZP
PI-P3	TAGTCTVTCCTT	ACGAGCTGATAT	D1	P3
PI-44	AGACXC1GAGXC	C1G1VXVAG1V1CV	D8	P4
PI-45	C1CV1VAGV1V1V1AC	VTCGGGAGXGAC	a9	45
PI-P6	1CGC1CAC1VACV	TGCTCGACVTC	a10	P6
PI-LP	ATAGAGTCTCAT	CGCATAGCGTAT	a11	LP
PI-48	CGAGVACVCTCGC	TGCTVGTCTVACV	ra1	48
PI-49	CAGCV1VAC1VTC	1CG1G1V1VACVAV	E1	P6
PI-SO	CAGC1VTVAGIC1	V1GCVGTVV1C1	a2	SO
PI-S1	TCTATCGATCTCVA	ACAGCGVAGVXC	E3	S1
PI-u-zs	CVTGAGTVXVAGC	CXVAGCXGACGXV	E4	ZS
PI-S3	TAGCATATCGAG	XVACVCG1V1GAG	S3	53

PI-54	ACG AC ICCG IAC	TCATGAGTAGTA	E6	S4
PI-SS	XAGC V XAC VCCG	TGACGGGTATAC	E7	S5
PI-S6	CGTC V AIGC VAG	X V X V AGCG V IGAC	E8	S6
PI-51	TGCAGCGAGTAC	TC DAC DCTA DCG	E9	S7
PI-S8	CG XG ICCG AC VAG V	CAGTCGTG VAGC V	ioa	S8
PI-S9	ACICG VCG IGAC	ACGGGAGTGATA	IIa	S9
PI-60	ATCGTCTG VCG	XGC I V I CAC I G V	E12	60
PI-61	C V I V CXG I V IC I	TACATAGATGTC	VI	61
PI-62	XC I AC ICCG XG AC	CAGGTATAGTGA	VZ	61
PI-63	CTGCAGTAGACA	ACTCATATCGCA	V3	63
PI-64	ACAGGAGCTCAT	CACTCATATCGA	V4	64
PI-S5	XAC VAG VTAG IC I	TCGTG TGTGATA	F5	65
PI-66	XAC VAC ICCG XGC I	TGACGGCTC VTCCT	V6	66
PI-67	IAC V XG IGAC VAG V	ICGTACA I DQ IC	F7	67
PI-68	TGTATGATCTCG	CAC I GTGC X C V I	i8	68
PI-69	CAGC VAC VXC X V	ACXGC V IGAC ICG	V9	69
PI-70	CATACTATCAGC	TGCTGTC VCTAC	PI0	70
PI-71	CAC I X XAC VAG V X	CGAC VCG I V C X V	F11	71
PI-72	AT VTCGT VGC VTC	TCGTG VTC VCT V	VIZ	72
PI-73	XAGC IC I V I VAC V I	ACVCGCTGTGGA	DI	73
PI-Id	TGTCAC VGTG VAC	TCAT VTC VTCGA	GZ	74
PI-75	ATCG VCT VTCCT	CGATCT VTC VAG	G3	75
PI-76	V I VAC I VGC V IC V	TCATGCTGACGA	PD	76
PI-77	CACCTG VCGC IC V	CACTACATCGCT	O5	77
PI-SL	TCGCTCATCTAT	TAGTAC VAG VAGCT	6D	78

PI-79	TGTATCACGAGC	ATGATCGTATAC	G 7	79
PI-80	TACTGCTATCTC	CGCTGCATAGCG	G 8	80
PI-81	CGCGAGCTCGTC	ACTCGATGAGCT	G 9	81
PI-82	TAGAGTCTGTAT	TGTCTATCACAT	G10	82
PI-83	TACTATCGCTCT	TATGTGACATAC	G11	83
PI-84	TAGATGACGCTC	TACTCGTAGCGC	G12	84
PI-85	TCGCGTGACATC	ATCTACTGACGT	H 1	85
PI-86	ACACGCTCTACT	ACAGTAGCGCAC	H 2	86
PI-87	TACATAGTCTCG	CTAGTATCATGA	H 3	87
PI-88	TGAGTAGCACGC	TCGATCATGCAG	H 4	88
PI-89	TAGATGCTATAC	TACATGCACTCA	H 5	89
PI-90	ATCGATGTCACG	CAGCTCGACTAC	H 6	90
PI-91	ATCATATGTAGC	CTCTACAGTCAC	H 7	91
PI-92	TAGCATCGATAT	AGATAGCACATC	H 8	92
PI-93	TGATCGACGCTC	CTAGATATCGTC	H 9	93
PI-94	TGCAGCTCATAG	TACAGACTGCAC	H10	94
PI-95	CGACGTAGAGTC	CAGTAGCACTAC	H11	95

表 2, 未添加引物标签前用于扩增 HLA-A/B/DRB1 相应外显子的 PCR 引物

引物编号	引物序列	引物用途	产物长度
A-F2	CCTCTGYGGGGAGAAGCAA	扩 HLA-A 基因 2 号外显子	480bp
A-R2	ATCTCGGACCCGGAGACTG		
A-F3	CGGGGCCAGGTTCTCACAC	扩 HLA-A 基因 3 号外显子	410bp
A-R3	GGYGATATTCTAGTGTTGGTCCCAA		

A-F4	GTGTCCCATGACAGATGCAAAA	扩 HLA-A 基因 4 号外显子	430bp
A-R4	GGCCCTGACCCTGCTAAAGG		
B-F2	AGGAGCGAGGGGACCGCA	扩 HLA-B 基因 2 号外显子	400bp
B-R2	CGGGCCGGGGTCACTCAC		
B-F3	CGGGGCCAGGGTCTCACA	扩 HLA-B 基因 3 号外显子	370bp
B-R3	GAGGCCATCCCCGGCGAC		
B-F4	GCTGGTCACATGGGTGGTCCTA	扩 HLA-A 基因 4 号外显子	380bp
B-R4	CTCCTTACCCCATCTCAGGGTG		
D2-F1	CACGTTTCTTGGAGTACTCTA	扩 HLA-DRB1 基 因 2 号外显子	300bp
D2-F2	GTTTCTTGTGGCAgCTTAaGTT		
D2-F3	CCTGTGGCAGGGTAAGTATA		
D2-F4	GTTTCTTGAAGCAGGATAAGTT		
D2-F5	GCACGTTTCTTGGAGGAGG		
D2-F6	TTTCCTGTGGCAGCCTAAGA		
D2-F7	GTTTCTTGGAGCAGGTAAAC		
D2-R	CCTCACCTCGCCGCTGCAC		

D2-F1, D2-F2, D2-F3, D2-F4, D2-F5, D2-F6, D2-F7 为扩增 HLA-DRB1 2 号外显子的正向引物, D2-R 为扩增 HLA-DRB1 2 号外显子的反向引物。

HLA-A/B/DRB1 的 PCR 程序如下:

96℃ 2min

95℃ 30s → 60℃ 30s → 72℃ 20s (32cycles)

15℃ ∞

HLA-A/B 的 PCR 反应体系如下所有试剂均购自普洛麦格 (北京) 生物技术有限公司 (Promega)

Promega 5 × buffer I (Mg2+ plus)	5.0ul
dNTP Mixture (各 2.5mM/ ul)	2.0ul
PI _{nr} -A/B-F _{2/3/4} (2pmol/ ul)	1.0ul
PI _{nr} -A/B-R _{2/3/4} (2pmol/ ul)	1.0ul
Promega Taq (5U/ul)	0.2ul
DNA(约 20 ng/ul)	5.0ul
ddH ₂ O	10.8ul
Total	25.0ul

HLA-DRBI 的 PCR 反应体系如下:

Promega 5 × buffer I (Mg2+ plus)	5.0ul
dNTP Mixture (各 2.5mM/ ul)	2.0ul
PI _{nr} -D2-F1 (2pmol/ ul)	1.0ul
PI _{nr} -D2-F2 (2pmol/ ul)	1.0ul
PI _{nr} -D2-F3 (2pmol/ ul)	1.0ul
PI _{nr} -D2-F4 (2pmol/ ul)	1.0ul
PI _{nr} -D2-F5 (2pmol/ ul)	1.0ul
PI _{nr} -D2-F6 (2pmol/ ul)	1.0ul
PI _{nr} -D2-F7 (2pmol/ ul)	1.0ul
PI _{nr} -D2-R (2pmol/ ul)	1.0ul
Promega Taq (5U/ul)	0.2ul
DNA(约 20 ng/ul)	5.0ul

ddH ₂ O	4.8ul
Total	25.0ul

其中 PI_n-A/B/D2-F_{1/2/3/4/5/6/7} 表示引物 5 末端带有第 n 号正向引物标签序列 (表 1) 的 HLA-A/B/DRB1 的 F 引物, PI_n-A/B/D2-R_{2/3/4} 表示引物 5 末端带有第 n 号反向引物标签序列的 HLA-A/B/DRB1 的 R 引物 (此处 n < 95), 其它依次类推。且每个样本对应特定的一套 PCR 引物 (PI_n-A/B/D2-F_{1/2/3/4/5/6/7}, PI_n-A/B/D2-R_{2/3/4})。

PCR 反应在 Bio-Rad 公司的 PTC-200 PCR 仪上运行。PCR 完成后, 取 2ul PCR 产物经 1% 的琼脂糖凝胶电泳检测。图 2 显示了 1 号样本 HLA-A/B/DRB1 相应外显子 PCR 产物电泳结果, DNA 分子标记为 DL 2000 (Takara 公司), 胶图上有一系列片段大小为 300bp-500bp 单一条带, 表明 1 号样本的 HLA-A/B/DRB1 各外显子 (A2、A3、A4、B2、B3、B4、DRB1-2) PCR 扩增成功, 阴性对照 (N) 无扩增条带。其它样品的结果与此类似

实施例 3

PCR 产物混合和纯化

从 96 孔板 HLA-P-A2 剩余的 PCR 产物中 (阴性对照除外) 各取 20ul 混合在一个 3ml 的 EP 管中, 标记为 HLA-A2-Mix, 对其它 6 个 96 孔板进行同样的操作, 分别标记为 HLA-A3-Mix、HLA-A4-Mix、HLA-B2-Mix、HLA-B3-Mix、HLA-B4-Mix 和 HLA-D2-Mix, 震荡混匀, 从 HLA-A2-Mix、HLA-A3-Mix、HLA-A4-Mix、HLA-B2-Mix、HLA-B3-Mix、HLA-B4-Mix 和 HLA-D2-Mix 中各取 200ul 混合在一个 3ml 的 EP 管中, 标记为 HLA-Mix, 从 HLA-Mix 中取 500ul DNA 混合物经 Qiagen DNA Purification kit 试剂盒 (QIAGEN 公司) 过柱纯化 (具体纯化步

骤详见说明书)，纯化所得的 200ul DNA，经 Nanodrop 8000(Thermo Fisher Scientific 公司)测定 HLA-Mix DNA 浓度为 48ng/ul.

实施例 4

PCR 产物的打断，以及 Illumina GA PCR-Free 测序文库的构建

1. DNA 打断

从纯化后的 HLA-Mix 中取总量 5ug 的 DNA 用带 AFA 纤维扣盖的 Covaris 微管在 Covaris S2DNA 打断仪(Covaris 公司)上打断。打断条件如下：

频率扫描 (frequency sweeping)

负载比 (Duty Cycle)	10%
强度 (Intensity)	5
循环/脉冲 (Cycles/Burst)	200
时间 (秒) (Time seconds)	300

2. 打断后纯化

将 HLA-Mix 的所有打断产物用 QIAquick PCR Purification Kit 回收纯化，分别溶于 37.5ul 的 EB (QIAGEN Elution Buffer) 中；

3. 末端修复反应

对打断后纯化的 HLA-Mix 进行 DNA 末端修复反应，体系如下 (试剂均购自 Enzymatics 公司)：

DNA	37.5μL
H ₂ O	37.5μL

多核苷酸激酶缓冲液 (10x Polynucleotide Kinase Buffer(B904))	10 μ L
dNTP 混合物 (每种 10mM) (Solution Set (10mM each))	4 μ L
T4 DNA 聚合酶 (T4 DNA Polymerase)	5 μ L
Klenow 片段 (Klenow Fragment)	1 μ L
T4 多聚核苷酸激酶 (T4 Polynucleotide Kinase)	5 μ L
总体积 (Total volume)	100 μ L

反应条件为：恒温混匀器(Thermomixer, Eppendorf 公司) 20℃温浴 30 min。

反应产物经 QIAquick PCR Purification Kit 回收纯化，溶于 34 μ l 的 EB (QIAGEN Elution Buffer) 中。

4. 3，末端加 A 反应

上一步回收 DNA 的 3，末端加 A 反应，体系如下 (试剂均购自 Enzymatics 公司)：

上一步所得 DNA	32 μ L
10x 蓝色缓冲液 (10x blue buffer)	5 μ L
dATP(1mM, GE 公司)	10 μ L
Klenow (3'-5' exo-)	3 μ L
总体积 (Total volume)	50 μ L

反应条件为：恒温混匀器 (Thermomixer , Eppendorf 公司) 37℃温浴 30 min。

反应产物经 MiniElute PCR Purification Kit(QIAGEN 公司) 回收纯化，溶于 13 μ l 的 EB 溶液 (QIAGEN Elution Buffer) 中。

5. 连接 Illumina GA PCR-Free 文库接头 (adapter)

术语 "PCR-Free 文库接头 (adapter)" 是指经设计的一段碱基,其主要作用是辅助固定 DNA 分子在测序芯片上以及提供通用测序引物的结合位点,PCR-Free 文库接头可以通过 DNA 连接酶将其直接连接至测序文库中的 DNA 片段两端,文库接头的导入过程因为没有 PCR 的参与,因此称作 PCR-Free 文库接头。

加 A 后的产物连接 Illumina GA PCR-Free 文库接头,体系如下 (试剂均购自 Illumina 公司):

上一步所得 DNA	11 μ L
2x 快速连接缓冲液 (2x Rapid ligation buffer)	15 μ L
PCR-free 寡核苷酸接头混合物 (30mM) (PCR-free Adapter oligo mix)	1 μ L
T4 DNA 连接酶 (T4 DNA Ligase) (Rapid, L603-HC-L)	3 μ L
总体积 (Total volume)	30 μ L

反应条件为:恒温混匀器 (Thermomixer, Eppendorf 公司) 20℃ 温浴 15 min。

反应产物经 Ampure Beads(Beckman Coulter Genomics) 纯化后溶于 50ul 去离子水,经荧光定量 PCR (QPCR) 检测到 DNA 浓度结果如下:

	qPCR 检测结果 (nM)
HLA-Mix	78.90

6. 割胶回收

取 30 μ L HLA-Mix 用 2% 低熔点琼脂糖胶进行回收。电泳条件为 100V, 100min。DNA marker 为 NEB 公司的 50bp DNA marker。割胶回收 450-750bp 长度范围的 DNA 片段 (附图 3)。胶回收产

物经 QIAquick PCR Purification Kit (QIAGEN 公司)回收纯化, 纯化后体积为 32ul, 经荧光定量 PCR (QPCR) 检测到 DNA 浓度结果为 10.16nM。

实施例 5

Illumina GA 测序

根据 QPCR 检测结果, 取 10pmol DNA 用 Illumina GA PE-100 程序测序, 具体操作流程详见 Illumina GA 操作说明书 (Illumina GA IIx)。

实施例 6

结果分析

Illumina GA 产出的测序结果是一系列 DNA 序列, 通过查找测序结果中的正反引物标签序列和引物序列, 建立各个引物标签对应样本 HLA-A/B/DRB1 各外显子 PCR 产物测序结果的数据库。通过 BWA(Burrows-Wheeler Aligner) 把各外显子的测序结果定位在相应外显子的参考序列上 (参考序列来源: <http://www.ebi.ac.uk/imgt/hla/>) 同时, 构建各个数据库的一致性 (consensus) 序列, 再对数据库中 DNA 序列进行筛选和测序错误校正。校正后的 DNA 序列通过序列重叠 (overlap) 和连锁 (Pair-End 连锁) 关系可组装成 HLA-A/B/DRB1 各外显子相应的序列。所得 DNA 序列利用与 IMGT HLA 专业数据库中 HLA-A/B/DRB1 相应各外显子的序列数据库比对, 序列比对结果 100% 匹配的即为对应样本的 HLA-A/B/DRB1 基因型别。可参考图 4 示例说明的 1 号样品的 HLA-A 位点的 2 号外显子一致性序列构建程序的截图。

所有 95 个样本，得到的分型结果与原已知分型结果完全相符，其中 1-32 号样本的具体结果如下：

样本编号	原 HLA-A/B/DRB1 型别					
1	A*02:03	A*11:01	B*38:02	B*48:01	DRB1*14:54	DRB1*15:01
2	A*01:01	A*30:01	B*08:01	B*13:02	DRB1*03:01	DRB1*07:01
3	A*01:01	A*02:01	B*45:11	B*47:01	DRB1*13:02	DRB1*15:01
4	A*24:08	A*26:01	B*40:01	B*51:01	DRB1*04:04	DRB1*09:01
5	A*01:01	A*24:02	B*54:01	B*55:02	DRB1*04:05	DRB1*09:01
6	A*01:01	A*03:02	B*15:11	B*37:01	DRB1*10:01	DRB1*14:54
7	A*11:01	A*30:01	B*43:02	B*15:18	DRB1*04:04	DRB1*07:01
8	A*01:01	A*02:01	B*35:03	B*81:01	DRB1*11:01	DRB1*15:01
9	A*02:06	A*31:01	B*27:07	B*40:02	DRB1*03:01	DRB1*13:02
10	A*01:01	A*66:01	B*37:01	B*49:01	DRB1*10:01	DRB1*13:02
11	A*01:01	A*03:01	B*35:01	B*52:01	DRB1*01:01	DRB1*15:02
12	A*11:01	A*11:01	B*45:01	B*15:05	DRB1*04:06	DRB1*15:01
13	A*01:01	A*11:02	B*07:02	B*15:02	DRB1*09:01	DRB1*15:01
14	A*01:01	A*02:01	B*52:01	B*67:01	DRB1*15:02	DRB1*16:02
15	A*01:01	A*02:05	B*15:17	B*50:01	DRB1*07:01	DRB1*15:01
16	A*01:01	A*11:01	B*37:01	B*40:02	DRB1*10:01	DRB1*12:02
17	A*24:07	A*32:01	B*35:05	B*40:01	DRB1*03:01	DRB1*04:05
18	A*11:01	A*24:02	B*13:01	B*35:01	DRB1*16:02	DRB1*16:02
19	A*11:01	A*11:01	B*40:02	B*55:12	DRB1*04:05	DRB1*15:01
20	A*02:11	A*24:02	B*40:01	B*40:06	DRB1*11:01	DRB1*15:01
21	A*01:01	A*02:06	B*51:01	B*57:01	DRB1*07:01	DRB1*12:01
22	A*01:01	A*29:01	B*07:05	B*15:01	DRB1*04:05	DRB1*07:01
23	A*01:01	A*02:07	B*37:01	B*46:01	DRB1*04:03	DRB1*10:01
24	A*24:85	A*30:01	B*43:02	B*55:02	DRB1*07:01	DRB1*15:01
25	A*11:01	A*31:01	B*07:06	B*51:01	DRB1*12:02	DRB1*14:05
26	A*01:01	A*11:01	B*46:01	B*57:01	DRB1*07:01	DRB1*08:03
27	A*01:01	A*02:01	B*45:18	B*37:01	DRB1*04:01	DRB1*15:01
28	A*01:01	A*24:02	B*37:01	B*46:01	DRB1*09:01	DRB1*10:01
29	A*26:01	A*66:01	B*40:40	B*41:02	DRB1*2:01	DRB1*15:01
30	A*02:01	A*29:02	B*43:02	B*45:01	DRB1*03:01	DRB1*12:02
31	A*01:01	A*11:03	B*45:01	B*57:01	DRB1*07:01	DRB1*15:01
32	A*11:01	A*26:01	B*35:03	B*38:01	DRB1*11:03	DRB1*14:04

样本编号	测得的 HLA-A/B/DRB1 型别					
1	A*02:03	A*11:01	B*38:02	B*48:01	DRB1*14:54	DRB1*15:01

2	A*01 :01	A*30:01	B*08:01	B*13:02	DRB1*03:01	DRB1*07:01
3	A*01 :01	A*02:01	B*15:01	B*47:01	DRB1*13:02	DRB1*15:01
4	A*24 :08	A*26:01	B*40:01	B*51:01	DRB1*04:04	DRB1*09:01
5	A*01 :01	A*24:02	B*54:01	B*55:02	DRB1*04:05	DRB1*09:01
6	A*01 :01	A*03:02	B*15:01	B*37:01	DRB1*10:01	DRB1*14:54
7	A*01 :01	A*30:01	B*13:02	B*15:18	DRB1*04:04	DRB1*07:01
8	A*01 :01	A*02:01	B*35:03	B*81:01	DRB1*11:01	DRB1*15:01
9	A*02 :06	A*31:01	B*27:07	B*40:02	DRB1*03:01	DRB1*13:02
10	A*01 :01	A*66:01	B*37:01	B*49:01	DRB1*10:01	DRB1*13:02
11	A*01 :01	A*03:01	B*35:01	B*52:01	DRB1*01:01	DRB1*15:02
12	A*01 :01	A*01:01	B*15:01	B*15:05	DRB1*04:06	DRB1*15:01
13	A*01 :01	A*01:02	B*07:02	B*15:02	DRB1*09:01	DRB1*15:01
14	A*01 :01	A*02:01	B*52:01	B*67:01	DRB1*15:02	DRB1*16:02
15	A*01 :01	A*02:05	B*15:17	B*50:01	DRB1*07:01	DRB1*15:01
16	A*01 :01	A*01:01	B*37:01	B*40:02	DRB1*10:01	DRB1*12:02
17	A*24 :07	A*32:01	B*35:05	B*40:01	DRB1*03:01	DRB1*04:05
18	A*01 :01	A*24:02	B*13:01	B*35:01	DRB1*16:02	DRB1*16:02
19	A*01 :01	A*11:01	B*40:02	B*55:12	DRB1*04:05	DRB1*15:01
20	A*02 :11	A*24:02	B*40:01	B*40:06	DRB1*11:01	DRB1*15:01
21	A*01 :01	A*02:06	B*51:01	B*57:01	DRB1*07:01	DRB1*12:01
22	A*01 :01	A*29:01	B*07:05	B*15:01	DRB1*04:05	DRB1*07:01
23	A*01 :01	A*02:07	B*37:01	B*46:01	DRB1*04:03	DRB1*10:01
24	A*24 :85	A*30:01	B*13:02	B*55:02	DRB1*07:01	DRB1*15:01
25	A*01 :01	A*31:01	B*07:06	B*51:01	DRB1*12:02	DRB1*14:05
26	A*01 :01	A*01:01	B*46:01	B*57:01	DRB1*07:01	DRB1*08:03
27	A*01 :01	A*02:01	B*15:18	B*37:01	DRB1*04:01	DRB1*15:01
28	A*01 :01	A*24:02	B*37:01	B*46:01	DRB1*09:01	DRB1*10:01
29	A*26 :01	A*66:01	B*40:40	B*41:02	DRB1*12:01	DRB1*15:01
30	A*02 :01	A*29:02	B*13:02	B*45:01	DRB1*03:01	DRB1*12:02
31	A*01 :01	A*01:03	B*15:01	B*57:01	DRB1*07:01	DRB1*15:01
32	A*01 :01	A*26:01	B*35:03	B*38:01	DRB1*11:03	DRB1*14:04

注：HLA-DRBI 型别中的 DRB1*1201 不排除 DRB1*1206/1210/1217 的可能性，DRB1*1454 不排除 DRB1*1401 的可能性，因为上述等位基因在 HLA-DRBI 2 号外显子的序列完全相同。

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。

参考文献

- [1]. <http://www.ebi.ac.uk/imgt/hla/stats.html>
- [2]. Tiercy J M. Molecular basis of HLA polymorphism: implications in clinical transplantation. [J]. Transpl Immunol, 2002, 9: 173-180.
- [3]. C.Antoine, S.Muller, A.Cant, et al. Long-term survival and transplantation of haemopoietic stem cells for immunodeficiencies: report of the European experience. 1968-99. [J]. The Lancet, 2003,9357:553-560.
- [4]. H. A. Erlich, G. Opelz, J. Hansen, et al. HLA DNA Typing and Transplantation. [J].Immunity, 2001,14:347-356.
- [5]. Lillo R, Balas A, Vicario JL, et al. Two new HLA class allele, DPBI*02014, by sequence-based typing. [J]. Tissue Antigens, 2002, 59: 47-48.
- [6]. A. Dormoy, N. Froelich, Leisenbach, et al. Mono-allelic amplification of exons 2-4 using allele group-specific primers for sequence-based typing (SBT) of the HLA-A, -B and -C genes: Preparation and validation of ready-to-use pre-SBT mini-kits. [J]. Tissue Antigens, 2003, 62: 201-216.
- [7]. Elaine R. Mardis. The impact of next-generation sequencing technology on genetics. [J]. Trends in

Genetics.2008,24:133-141.

[8]. Christian Hoffmannl, Nana Minkahl, Jeremy Leipzig. DNA barcoding and pyrosequencing to identify rare HIV drug resistance mutations. [J]. Nucleic Acids Research,2007,1-8.

[9]. Shannon J.Odelberg, Robert B.Weiss , Akira Hata. Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. [J]. Nucleic Acids Research.1995, 23:2049-2057.

[10]. Sayer D , Whidborne R ,Brestovac B . HLA -DRB1 DNA sequencing based typing: an approach suitable for high throughput typing including unrelated bone marrow registry donors. [J]. Tissue Antigens. 2001 ,57(1):46-54.

[11]. Iwanka Kozarewa, Zemin Ning, Michael A Quail. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. [J]. Nature Methods. 2009, 6:291 - 295.

权 利 要 求

1. 一种测定样品中目的核酸的核苷酸序列的方法，其包括：

1) 提供 n 个样品， n 为大于等于 1 的整数，所述样品优选地来自哺乳动物，更优选是人，特别是人的血样；可选地，将待分析的 n 个样品分成 m 个小组， m 为整数且 $n \geq m \geq 1$ ；

2) 扩增：对于每一个样品，使用一对或多对标签引物，在存在来自该样品的模板时，在适于扩增目的核酸的条件下进行 PCR 扩增，其中，每一对标签引物由包含引物标签的正向标签引物和反向标签引物（均可以是简并引物）构成，其中正向标签引物和反向标签引物所包含的引物标签可以相同或者不同；不同样品所用标签引物对中的引物标签彼此不同；

3) 混合：当 $n > 1$ 时，将各样品的 PCR 扩增产物混合在一起；

4) 打断：将所得的扩增产物进行不完全打断，并进行纯化回收；

5) 测序：将回收的 DNA 混合物利用二代测序技术，优选的是 Pair-End 技术（例如 Illumina GA、Illumina HiSeq 2000）进行测序，获得打断后的 DNA 的序列；和

6) 拼接：基于每个样品独特的引物标签将获得的测序结果与样品一一对应，利用比对程序（例如 Blast、BWA 程序）把各个测序序列定位到 PCR 产物的相应 DNA 参考序列上，通过序列重叠和连锁关系，从打断后的 DNA 的序列拼接出完整的目的核酸。

2. 权利要求 1 所述的方法，其中每一对引物标签与 PCR 引物对组合成一对标签引物，正反 PCR 引物的 5' 端分别具有（或者任选通过连接序列连接）正向引物标签和反向引物标签。

3. 权利要求 1 所述的方法，其中所述 PCR 引物是用于扩增

HLA-A/B 的 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 引物, 优选的所述 PCR 引物如表 2 所示。

4. 权利要求 1 所述的方法, 其中所述引物标签针对 PCR 引物进行设计, 优选针对用于扩增 HLA 的特定基因的 PCR 引物进行设计, 更优选是用于扩增 HLA-A/B 的 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子的 PCR 引物, 特别是如表 2 所示的 PCR 引物进行设计, 所述引物标签特别是包括表 1 所示 95 对引物标签中的至少 10 对, 或至少 20 对, 或至少 30 对, 或至少 40 对, 或至少 50 对, 至少 60 对, 或至少 70 对, 或至少 80 对, 或至少 90 对, 或 95 对 (或者所述一组引物标签由表 1 所示 95 对引物标签中的 10 - 95 对 (例如 10 - 95 对, 20 - 95 对, 30 - 95 对, 40 - 95 对, 50 - 95 对, 60 - 95 对, 70 - 95 对, 80 - 95 对, 90 - 95 对, 或 95 对) 组成), 并且

所述一组引物标签优选地至少包括表 1 所示 95 对引物标签中的 PI-1 至 PI-10, 或 PI-11 至 PI-20, 或 PI-21 至 PI-30, 或 PI-31 至 PI-40, 或 PI-41 至 PI-50, 或 PI-51 至 PI-60, 或 PI-61 至 PI-70, 或 PI-71 至 PI-80, 或 PI-81 至 PI-90, 或 PI-91 至 PI-95, 或者它们任何两个或者多个的组合。

5. 权利要求 1 所述的方法, 其中所述 DNA 打断包括化学打断方法和物理打断方法, 其中所述化学方法包括酶切方法, 所述物理打断方法包括超声波打断方法或机械打断方法。

6. 权利要求 1 所述的方法, 其中所述 DNA 打断后, 纯化回收从测序仪最大读取长度到测序仪可适用的最长 DNA 长度范围之间的所有 DNA 条带, 其中所述纯化回收方法包括但不限于电泳割胶回收, 也可以是磁珠回收。

7. 权利要求 1 所述的方法, 其中所述方法还可以包括权利要求 1 所述的步骤 1) - 4), 以及如下步骤:

5) 建库 : 将打断后的 PCR 产物文库构建 PCR-Free 测序文库 , 可以对文库添加不同的文库接头 (adapter) 以区分不同的 PCR-Free 测序文库 , 纯化回收位于所用测序仪最大读长长度到所用测序仪适用的最长 DNA 长度范围之间的所有 DNA 条带 , 优选地是 450-750bp 长度范围的 DNA 片段 ;

6) 测序 : 将回收的 DNA 混合物利用二代测序技术 , 优选的是 Pair-End 技术 (例如 Illumina GA、Illumina Hiseq 2000) 进行测序 , 获得打断后的 DNA 的序列 ;

7) 拼接 : 基于各个文库不同的文库接头序列和每个样品独特的引物标签将获得的测序结果与样品——对应 , 利用比对程序 (例如 Blast , BWA 程序 把各个测序序列定位到 PCR 产物的相应 DNA 参考序列上 , 通过序列重叠和连锁关系 , 从打断后的 DNA 的序列拼接出完整的目的核酸。

8. 权利要求 1 - 7 中任一项所述的方法用于 HLA 分型的用途 , 其特征在于包括 : 使用权利要求 1 - 7 中任一项的方法对来自患者的样品 (特别是血样) 进行测序 , 以及将测序结果与 HLA 数据库 (如 IMGT HLA 专业数据库) 中 HLA-DRB1 2 号外显子的序列数据比对 , 序列比对结果 100 % 匹配的即为对应样本的 HLA-DRB1 基因型别。

9. 一组引物标签 , 其包括表 1 所示 95 对引物标签中的至少 10 对 , 或至少 20 对 , 或至少 30 对 , 或至少 40 对 , 或至少 50 对 , 至少 60 对 , 或至少 70 对 , 或至少 80 对 , 或至少 90 对 , 或 95 对 (或者所述一组引物标签由表 1 所示 95 对引物标签中的 10 - 95 对 (例如 10 - 95 对 , 20 - 95 对 , 30 - 95 对 , 40 - 95 对 , 50 - 95 对 , 60 - 95 对 , 70 - 95 对 , 80 - 95 对 , 90 - 95 对 , 或 95 对) 组成) , 并且

所述一组引物标签优选地至少包括表 1 所示 95 对引物标签中

的 PI-1 至 PI-10 , 或 PI-11 至 PI-20 , 或 PI-21 至 PI-30 , 或 PI-31 至 PI-40, 或 PI-41 至 PI-50, 或 PI-51 至 PI-60, 或 PI-61 至 PI-70, 或 PI-71 至 PI-80, 或 PI-81 至 PI-90, 或 PI-91 至 PI-95, 或者它们任何两个或者多个的组合。

10. 权利要求9所述的一组引物标签用于PCR测序方法的用途, 其中特别是, 每一对引物标签与用于扩增待测目的序列的PCR引物对组合成一对标签引物, 正反PCR引物的5端分别具有(或者任选通过连接序列连接)正向引物标签和反向引物标签。

11. 权利要求10所述的用途, 其中PCR引物是用于扩增HLA的特定基因的PCR引物, 优选是用于扩增HLA-A/B的2, 3, 4号外显子以及HLA-DRB1 2号外显子的PCR引物, 优选的所述PCR引物如表2所示。

12. 权利要求9的一组引物标签与用于扩增待测目的序列的PCR引物对组合成的一组标签引物, 其中每一对引物标签与PCR引物对组合成一对标签引物, 正反PCR引物的5端各具有(或者任选通过连接序列连接)一个引物标签。

13. 权利要求12所述的标签引物, 其中所述PCR引物是用于扩增HLA的特定基因的PCR引物, 优选是用于扩增HLA-A/B的2, 3, 4号外显子以及HLA-DRB1 2号外显子的PCR引物, 优选的所述PCR引物如表2所示。

14. 权利要求12所述的标签引物用于PCR测序方法的用途。

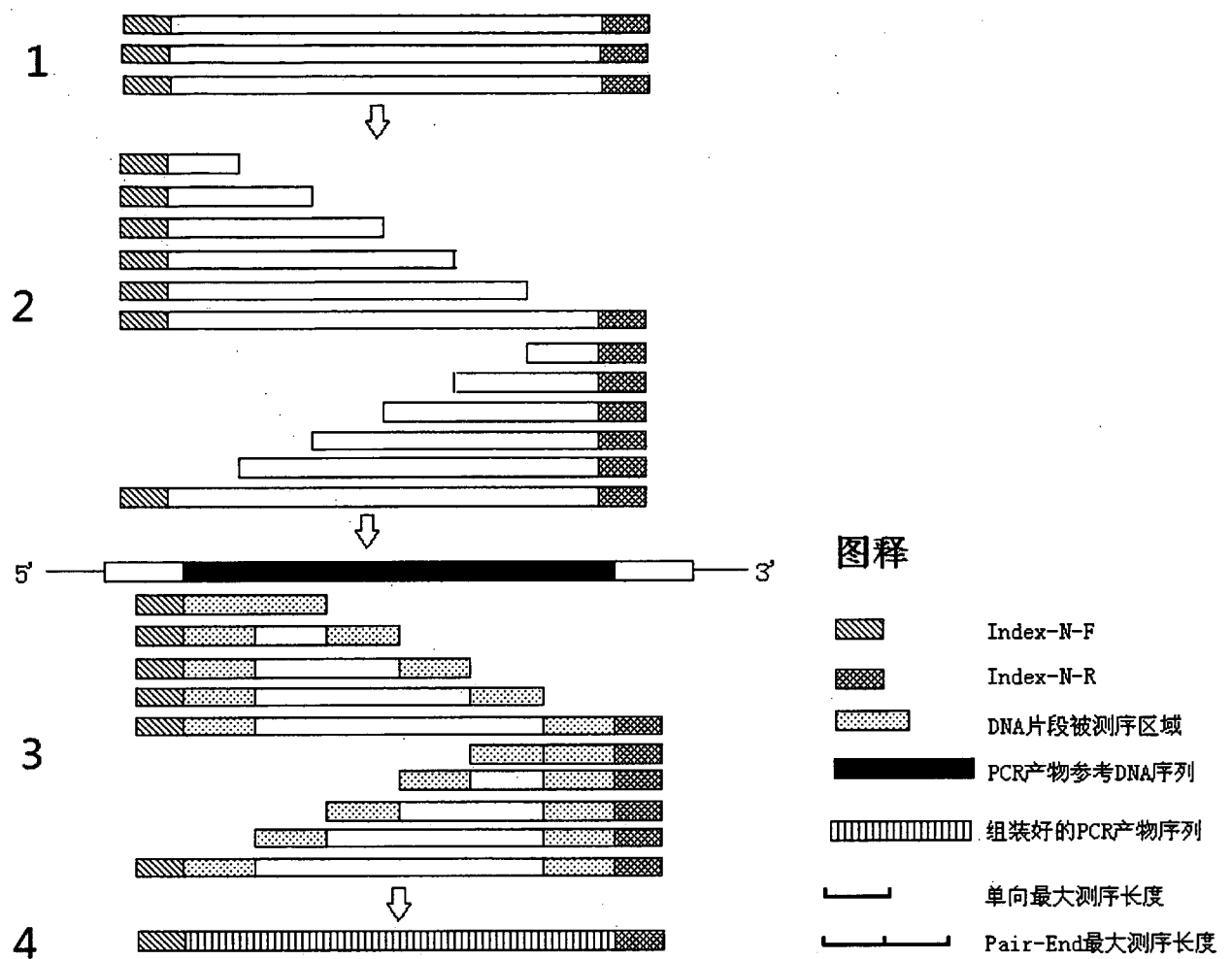


Fig. 1

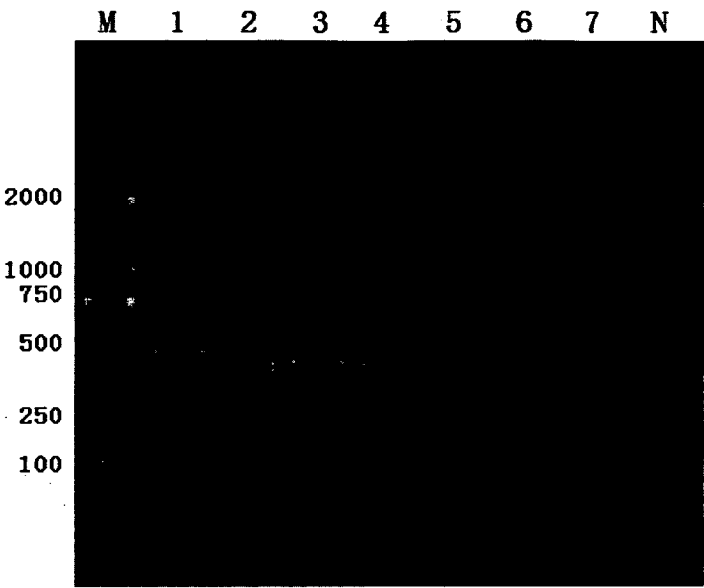


Fig. 2

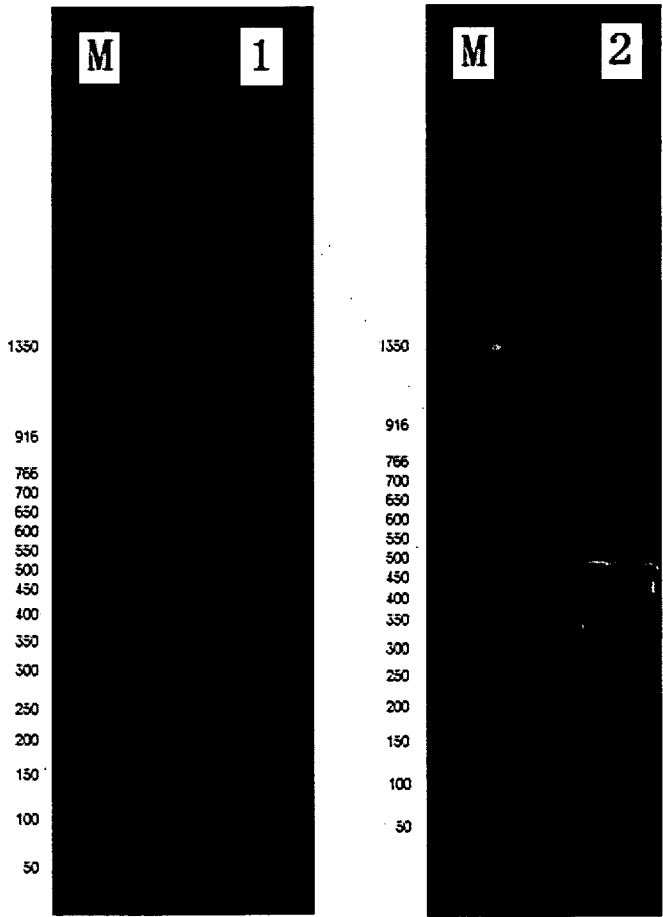


Fig. 3



Fig. 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1834

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)

CNPAT, CNKI, WPI, EPODOC, ISI-WEB OF KNOWLEDGE, NCBI :sequence+, PCR, pair end, paired end, solexa, Illumina, multi+,
 PCR-index, PCR- barcode, primer index, next- generation sequencing, sequence search on SEQ ID NO: 1-210

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Richard Cronn et al., Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology, Nucleic Acids Research, 2008, Vol. 36, No. 19, Published online 27 August 2008, e122	1-3,5-8
A	WO2008045575 A2 (CRAIG VENTER INST INC J [US] et al.) 17 April 2008 (17.04.2008), see the whole document	1-14

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search
 3 March 2011 (03.03.2011)

Date of mailing of the international search report
 07 Apr. 2011 (07.04.2011)

Name and mailing address of the ISA/CN
 The State Intellectual Property Office, the P.R.China
 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
 100088
 Facsimile No. 86-10-62019451

Authorized officer
 QIU,Fuen
 Telephone No. (86-10) 62411034

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1834

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN101654691A (SHENZHEN BGI TECHNOLOGY CO LTD) 24 February 2010 (24.02.2010), see the whole document	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN20 10/00 1834

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
M O 2008045575 A 2	17.04.2008	W O 2008045575 A 3	16.10.2008
		U S 2010311602 A I	09.12.2010
CN101654691A	24.02.2010	NONE	

国际检索报告

国际申请号
PCT/CN2010/001834

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))

CNPAT, CNKI, WPI, EPODOC, ISI-WEB OF KNOWLEDGE, NCBI: sequence+, PCR, pair end, paired end, solexa, Illumina, imulti+, PCR-index, PCR- barcode, primer index, next-generation sequencing, 测序, 引物标签, 引物标记, 引物索引, 第二代测序, 多重, 基于 SEQ ID NO: 1-210 的序列检索

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	Richard Cronn 等, Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology, Nucleic Acids Research, 2008, 第 36 卷第 19 期, 2008 年 8 月 27 日在线公开, el22	1-3,5-8
A	WO2008045575 A 2 (CRAIG VENTER INST INC J [US] 等) 17.4 月 2008 (17.04.2008), 参见全文	1-14

因 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 % 或 % 后公布的在先申请或 %

"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
3.3 月 2011 (03.03.2011)

国际检索报告邮寄日期
07.4 月 2011 (07.04.2011)

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 10008S
传真号: (86-10)62019451

受权官员
邱福恩
电话号码: (86-10) 62411034

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
A	CN101654691A (深圳华大基因科技有限公司) 24.2 月 2010 (24.02.2010), 参见全文	1-14

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2010/001834

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
W O 2008045575 A 2	17.04.2008	W O 2008045575 A 3	16.10.2008
		U S 2010311602 A I	09.12.2010
CN101654691A	24.02.2010	无	