

(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 2012/000153 A I

- (51) 国际分类号 :
C12Q 1/68 (2006.01) C07H 21/00 (2006.01)
C12N 15/11 (2006.01)
- (21) 国际申请号 : PCT/CN2010/001835
- (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 :
201010213719.5 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)。 田埂 (TIAN, Geng) [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)。 蒋愈 (JIANG, Hui) [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, ML, 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))。
- 包括说明书序列列表部分 (细则 5.2(a))。

(54) Title: HIGH RESOLUTION TYPING METHOD OF HLA GENE BASED ON ILLUMINA GA SEQUENCING TECHNOLOGY

(54) 发明名称 基于 ILLUMINAGA 测序技术的 HLA 基因高分辨率分型方法

(57) Abstract: The invention provides a high resolution typing method of HLA gene based on Illumina GA sequencing technology. Also provided are primer indexes used for the method.

(57) 摘要:

本发明提供了基于 Illumina GA 测序技术的 HLA 基因高分辨率分型方法, 还提供了用于上述方法的引物标签。



2012/000153 A1

基于 Illumina GA 测序技术的 HLA 基因高分辨率分型方法

技术领域

5 本发明涉及核酸测序技术领域，特别是 PCR 测序技术领域。另外，本发明还涉及 DNA 分子标签技术和 DNA 不完全打断策略。本发明的方法特别适用于第二代测序技术。本发明的方法涉及 DNA 序列的分型方法，特别是 HLA 基因高分辨率分型方法。

背景技术

10 人类白细胞抗原，即 HLA(human leukocyte antigen, HLA)，是迄今为止发现的多态性最高的基因系统之一，它是调控人体特异性免疫应答和决定疾病易感性个体差异的主要基因系统，与同种异体器官移植的排斥反应密切相关。研究发现，移植时，供受双方的 HLA 相关基因匹配程度越高，分辨率越高，移植物的存活时间越长。

15 HLA-SBT (Sequence Based Typing, 基于 DNA 序列的分型方法)是当前 HLA 高分辨率分型的主要方法。该方法通过 PCR 扩增相应 HLA 的基因区域，对扩增产物测序，测序结果经过专业的分型软件分型，最终得到样本的 HLA 基因型别信息。其具有直观、高分辨且能检测新的等位基因的特点。

20 当前的 HLA-SBT 方法主要是基于 Sanger 测序法，该测序方法不能直接得到样本中单体型（父本或者母本单独的序列信息）的序列信息，而只能得到二倍体型序列信息，这给 HLA 分型结果带来了不确定性（Ambiguity）。为了得到确定的分型结果，需要加测 GSSP（组特异性测序引物，Group Sepecific Sequencing Primer）或者通过克隆测序来解决上述问

基于 Sanger 测序法的测序通量小，以 ABI 公司的 3730 测序仪一天 4000 个测序反应的饱和通量为例，单份样本的 HLA-A/B/DRB1 三个位点的 SBT 分型大概需要 17 个测序反应，一台 3730 测序仪整天不停运转也只能产出约 240 份样本的数据量。

目前 HLA-SBT 主要通过专业分型软件来分型，导入测序峰图质量的好坏对分型软件的峰图识别能力影响很大，当软件识别错误时，需要分型人员能及时发现错误，改正错误。同时，为了避免人为错误，同一批次样本的分型工作往往要由两人以上独立完成，核对无误后，才能确认结果。如果能够实现软件分型的自动化那将大大减少错误的发生率，并且降低人力成本。

基于 Sanger 测序法的 HLA-SBT 实验步骤包括：PCR 扩增、PCR 产物电泳、PCR 产物纯化、测序反应、测序反应产物纯化、测序仪测序、测序结果分型以及后继的加测 GSSP 等，整个实验流程复杂，且实验过程中，不同 PCR 产物不能混合在一起操作，大大地加大了实验工作量。

HLA-SBT 整个实验流程复杂、通量低和成本高等缺点使其很难应用于大规模 HLA 高分辨分型项目。

发明内容

Illumina GA 测序 (Illumina 公司的 Genome Analyzer 测序仪，简称 Illumina GA) 是利用边合成边测序的原理进行 DNA 序列分析，可以检测单体型，其最终产出的数据是一系列的碱基序列，可直接用于与 HLA 数据库中的参考序列直接比对，不存在传统分型软件峰图误判的问题，有利于软件分型的自动化。

Illumina GA 的测序通量大，目前一个实验流程下来可以产生 50G (500 亿) 碱基的数据，平均每天产生 50 亿碱基的数据。高的

据通量可以在测序序列数确定的情况下，使得每条序列获得高的测序深度，确保测序结果的可靠性。

目前还未有将 Illumina GA 应用于 HLA 分型领域的研究，本发明首次将 Illumina GA 测序应用于 HLA 分型领域，结合 DNA 分子标签技术、DNA 不完全打断及 PCR-FREE 建库的 PCR 测序技术，实现 HLA 的低成本，高通量、高准确率、高分辨率的分型。

基于 DNA 分子标签技术，实现了对多样本 PCR 产物的分别标记，使 Illumina 测序文库构建实验环节可把多个样本混合（pooling）成一个文库同时处理，大大简化了实验操作，最终，每个样本的检测结果可以通过其独特的标签（index）序列找回。

DNA 不完全打断技术使 Illumina GA 实际可测通的 PCR 产物长度超过测序仪的测序最大长度，在当前 Illumina GA 测序最大长度 200bp 的情况下，实际可测通的 PCR 产物长度达到 200bp 以上。

"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减少引物标签数目的情况下，大大提高可唯一标记的样本数目（图 1）。

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。

本发明，采用基于 DNA 分子标签技术、DNA 不完全打断及 PCR-FREE 建库的 PCR 测序技术，通过对待分析样本分组，再对每组样本通过双向引物标签标记的引物，对 HLA 基因目的片段扩增 (PCR 产物的最大长度取决于测序仪可结合的最大 DNA 长度，当前 Illumina GA 适用的最大 DNA 长度为 700bp，此长度为原始 DNA 长度，没有包括文库接头序列长度)，所得 PCR 产物等量混合，经 DNA 不完全打断处理，构建 PCR-Free 标签测序文库。把各样本组得到的不同标签测序文库等摩尔混合，选择性回收片段长度大于测序仪最大测序长度以上的所有 DNA 片段随后用 Illumina GA 测序仪测序。通过对测序结果中接头标签 (adapter index)、引物标签以及 PCR 引物的序列信息筛选，可获得每个样本的 DNA 序列信息，所得 DNA 序列经过组装与 IMGT HLA 专业数据库中对应数据库的比对，最终可得到样本的 HLA 基因型别。

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, 5 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, 10 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 15 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 20 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。

25 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 HLA 分型的方法，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；

2) 将待分析的 n 个样品分成 m 个小组， m 为整数且 $n > m \geq 1$ ；

5 3) 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
10 引物对中的引物标签彼此不同；

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

5) 打断：将所得的 PCR 产物文库进行不完全打断；

6) 建库：结合文库接头标签技术，将打断后的 PCR 产物文库
15 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；

7) 测序：将回收的 DNA 混合物利用二代测序技术，优选的是
20 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；

8) 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
25 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，在上文所述的方法中，所述结合文库接头标签技术，将打断后的 PCR 产物文库构建 PCR-Free 测序文库是指使用 m 种文库接头给 4) 中得到的 m 个 PCR 产物文库加上接头，其中每一个 PCR 产物文库使用一种不同的文库接头，从而构建 m 个接头标签测序文库；将 m 个接头标签测序文库等摩尔混合在一起构建混合接头标签测序文库。其中连接文库接头的方法是指不通过 PCR 程序直接采用 DNA 连接酶进行连接。

在本发明的一个具体实施方式中，在上文所述的方法中，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 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。

在本发明的一个具体实施方式中，在上文所述的方法中，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 2, 3, 4 号外显子和 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,
5 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, 纯化回收 450-750bp 长度的片段。所述纯化回收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。

在本发明的一个具体实施方式中, 在上文所述的方法中, 所述 DNA 打断后, 在构建 PCR-Free 标签文库的过程中, 对不同组
15 样品的 DNA 用不同的文库接头连接, 从而在其后的分型步骤中, 基于每个样品所用的引物标签和接头标签, 将获得的测序结果与样本一一对应。

利用比对程序把各个样本测序序列定位到其 PCR 产物已知相应的 DNA 参考序列 (Reference Sequence) 上, 通过序列重叠和连锁
20 关系, 从打断后的 DNA 的序列拼接出完整的 PCR 产物序列。

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) 进行测序和拼接, 以及将拼接好的序列与 HLA 数据库 (如 IMGT HLA 专业数据库) 中 HLA 相关序列数据比对, 序列比对结果 100% 匹
25 配的即为对应样本的 HLA-DRB1 基因型别。

发明的有益效果

本发明提供了基于 illumina GA 测序技术的 HLA 基因高分辨率分型方法，从而实现单体型测序、软件分型自动化，提高 HLA 基因分型的通量，降低成本。

附图说明

图 1: 为引物标签和接头标签 (adaptor index) 标记后的 PCR 产物示意图。实验时，通过 PCR 在每个样本的 PCR 产物两端同时引入引物标签，把多个带有不同引物标签的 PCR 产物混合在一起，用于构建测序文库。测序文库构建过程中，当需要构建多个测序文库时，可通过添加带有不同接头标签的文库接头，来标记各个测序文库。文库构建完毕后，带有不同接头标签标记的多个测序文库可以混合在一起同时用 Illumina GA 测序（不同接头标签标记的测序文库之间的引物标签可以相同）。测序结果出来后，通过对测序结果中接头标签和引物标签序列信息的筛选，可获得每个样本的 DNA 序列信息。

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，采用基于引物标签、DNA 不完全打断、文库标签及 PCR-FREE 建库的 PCR 测序方法，对 950 个样本的 HLA-A/B 2, 3, 4 号外显子以及 HLA-DRB1 2 号外显子 (PCR 产物长度大小处于 290bp-500bp 之间) 的基因分型，证明该发明能够实现低成本、高通量、高准确率和高分辨率的 HLA 基因分型。

原理：将待分析的样本均分成 10 组，对每组样本通过 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 扩增产物等体积混合在一起，获得 PCR 产物文库；所得 PCR 产物文库经过超声不完全打断后，构建不同的 PCR-Free 标签测序文库 (其中每一个 PCR 产物文库使用一种不同的接头，从而构建 10 个标签测序文库)；将 10 个标签测序文库等摩尔混合在一起构建混合标签测序文库，混合标签测序文库经 2% 低熔点琼脂糖电泳，割胶纯化回收位于 450-750p 长度范围之间的所有 DNA 条带。回收的 DNA 经 Illumina GA PE-100 测序。通过文库标签和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 (请提供供货商信息) (美国
5 Thermo 公司) 从 950 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 (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作好标记,
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
10 "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, 准备做下一步 PCR 中的模板
用。

15 实施例 2

PCR 扩增

把样本提取步骤中所得的 950 份 DNA 依次编号 1-950, 均分
成 10 组, 每组 95 份 DNA, 分别标记为 HLA-1、HLA-2、HLA-3、
HLA-4、HLA-5、HLA-6、HLA-7、HLA-8、HLA-9、HLA-10。
20 对每组样本分别以 95 套带有双向引物标签 (表 1) 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 (表 2) 来分别扩增 95 份 DNA 样本。PCR 反应在 96 孔板中
进行, 共 7 板, 编号分别为 HLA-X-P-A2、HLA-X-P-A3、
HLA-X-P-A4、HLA-X-P-B2、HLA-X-P-B3、HLA-X-P-B4 以及
25 HLA-X-P-DRB1-2 ("X" 表示样本组号信息 1/2/3/4/5/6/7/8/9/10,
"A2/3/4 B2/3/4 ,DRB1-2" 表示扩增的位点), 每板内设置一个



添加模板的阴性对照，阴性对照所用引物为 PI-1（表 1）标记的引物。实验的同时，记录下每个样本对应的样本组号信息和引物标签信息。

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)	对应模板 (组 n+1, 其中 $1 \leq n < 10$, n 为整数)
PI-1	TCGCAGACATCA	TGACACGATGCT	A1	1	$1+95*n$
PI-2	TACATCGCACTA	TACAGATGCTGA	A2	2	$2+95*n$
PI-3	CTCGATGAGTAC	ACGTCTAGACAC	A3	3	$3+95*n$
PI-4	TCTGTATACTCA	TGCTGTAGTGAC	A4	4	$4+95*n$
PI-5	TATCTGCTCATA	AGATATCGAGCT	A5	5	$5+95*n$
PI-6	TACATGCTGAGC	ACGTGTCTATCA	A6	6	$6+95*n$
PI-7	TCATATCGCGAT	AGATCGTATAGC	A7	7	$7+95*n$
PI-8	ACAGATGCACGC	ATCTCGTGACAG	A8	8	$8+95*n$
PI-9	TAGATCGTACAT	ACTAGTACACGC	A9	9	$9+95*n$
PI-10	ACTACACGTCTC	ATAGTCACGCGT	A10	10	$10+95*n$
PI-11	AGACTCGCGTAT	TACTAGCTGACG	A11	11	$11+95*n$
PI-12	ATACTAGTGCTC	TGTATCGTGCTC	A12	12	$12+95*n$
PI-13	CACGATGACATC	TAGTGAGCGCAC	B1	13	$13+95*n$
PI-14	TGCTGTCTCGAG	CATAGCAGTGTC	B2	14	$14+95*n$
PI-15	TGTGCTCGAGTC	TCTGATCGAGCA	B3	15	$15+95*n$
PI-16	CACTCGTACATC	AGCGATGCTCAT	B4	16	$16+95*n$
PI-17	CGACGTGCTCGC	CGCGTACTGCAG	B5	17	$17+95*n$
PI-18	ACGCATCTATAC	CTAGTATCGCAG	B6	18	$18+95*n$
PI-19	CGAGATGACTCT	TGTATACACGAT	B7	19	$19+95*n$
PI-20	ACTGTCTCGAGC	ACGTAGCGCACA	B8	20	$20+95*n$
PI-21	CATCTGCTATAG	TCTAGCTCATGA	B9	21	$21+95*n$
PI-22	ACGCACTCTAGA	CTATGCACTGAT	B10	22	$22+95*n$
PI-23	TGAGATACAGTA	ATCTGCTATGAC	B11	23	$23+95*n$
PI-24	ACTCATCGTGCT	TAGAGCTGTCAC	B12	24	$24+95*n$
PI-25	TACACTGTCTAT	CAGCACATAGAT	C1	25	$25+95*n$
PI-26	CACAGTACTCGC	CTGCTAGTGTAT	C2	26	$26+95*n$

PI-27	TGTACTATCATA	TGTGATAGACAC	C3	27	27+95*n
PI-28	CTAGTACTGACG	AGCGAGTCTACT	C4	28	28+95*n
PI-29	TAGACTGAGCTA	ACATACTGAGAC	C5	29	29+95*n
PI-30	CAGACGCCGTGAG	TACATCTCGTAT	C6	30	30+95*n
PI-31	CGCGACATCAGC	TAGCGATGAGAC	C7	31	31+95*n
PI-32	ACACTCATAGAT	CTATCATGACAC	C8	32	32+95*n
PI-33	AGCGTATACTAG	CATACCTCAGCTA	C9	33	33+95*n
PI-34	TGTCGTGCTATC	ACATGACTCAGC	C10	34	34+95*n
PI-35	CGCTAGACTGTA	TACTATAGTCGA	C11	35	35+95*n
PI-36	ACAGTGTAGCGC	TGATATGCTACA	C12	36	36+95*n
PI-37	CACTCTATCGAC	TCACGGCGATGAG	D1	37	37+95*n
PI-38	ACACTCTAGTCA	ACGTAGATCTAT	D2	38	38+95*n
PI-39	CATATGAGATCG	AGCAGAGTGCTC	D3	39	39+95*n
PI-40	CAGCTATCATAC	CACTGCAGACGA	D4	40	40+95*n
PI-41	TATACTCTAGAT	TGCATAGAGCGC	D5	41	41+95*n
PI-42	TGTATGCTCGTC	TCGTGACAGATC	D6	42	42+95*n
PI-43	TAGTGATGCTCT	ACGAGCTGATAT	D7	43	43+95*n
PI-44	AGACTCTGAGTC	CTGATAGTATCA	D8	44	44+95*n
PI-45	CTCATAGACTAC	ATCGCGAGTGAC	D9	45	45+95*n
PI-46	TCGCTCACTACA	TGTCTCGACATC	D10	46	46+95*n
PI-47	ATAGAGTCTCAT	CGCATAGCGTAT	D11	47	47+95*n
PI-48	CGAGACACTCGC	TCGTAGTCTACA	D12	48	48+95*n
PI-49	CAGCATACTATC	TCGTGATACAGA	E1	49	49+95*n
PI-50	CAGCTATAGTCT	ATGCAGATATCT	E2	50	50+95*n
PI-51	TCTATCGATGCA	ACACGCCAGATCG	E3	51	51+95*n
PI-52	CATGAGTATAGC	CTAGCTGACGTA	E4	52	52+95*n
PI-53	TAGCATATCGAG	TACACGTATGAG	E5	53	53+95*n
PI-54	ACGACTCGCTAC	TCATGACTAGTA	E6	54	54+95*n
PI-55	TAGCATACACGC	TGACCGCGTATAC	E7	55	55+95*n
PI-56	CGTCATATGCAG	TATAGCGATGAC	E8	56	56+95*n
PI-57	TGCAGCGAGTAC	TCGACCGCTAGCG	E9	57	57+95*n
PI-58	CGTGTGCAGACA	CAGTCTGTAGCA	E10	58	58+95*n
PI-59	ACTCGACGTGAG	ACGCGAGTGATA	E11	59	59+95*n
PI-60	ACTCGTCTGACG	TGCTATCACTGA	E12	60	60+95*n

PI-61	CATACTGTATCT	TACATAGATGTC	F1	61	61+95*n
PI-62	TCTACTCGTGAC	CACGTATAGTGA	F2	62	62+95*n
PI-63	CTGCACTAGACA	ACTCATATCGCA	F3	63	63+95*n
PI-64	ACACGAGCTCAT	CACTCATATCGA	F4	64	64+95*n
PI-65	TACAGATAGTCT	TCGTCTGTGATA	F5	65	65+95*n
PI-66	TACTACTCGTGCT	TGACGCTCATCT	F6	66	66+95*n
PI-67	TACATGTGACGA	TCGTACATGCTC	F7	67	67+95*n
PI-68	TGTATGATCTCG	CACTGTGCTCAT	F8	68	68+95*n
PI-69	CAGTACACTCTA	ACTGCATGATCG	F9	69	69+95*n
PI-70	CATACTATCACG	TCGTGTCACTAC	F10	70	70+95*n
PI-71	CACTATACAGAT	CGACACGTACTA	F11	71	71+95*n
PI-72	ATATCGTAGCAT	TCGTGATCACTA	F12	72	72+95*n
PI-73	TAGTCTATACAT	AGACGCTGTCGA	G1	73	73+95*n
PI-74	TGTCACAGTGAC	TCATATGATCGA	G2	74	74+95*n
PI-75	ATCGACTATGCT	CGATCATATGAG	G3	75	75+95*n
PI-76	ATACTAGCATCA	TCATGCTGACGA	G4	76	76+95*n
PI-77	CACTGACGCTCA	CACTACATCGCT	G5	77	77+95*n
PI-78	TCGCTCATCTAT	TAGTACAGAGCT	G6	78	78+95*n
PI-79	TGTATCACGAGC	ATGATCGTATAC	G7	79	79+95*n
PI-80	TACTGCTATCTC	CGCTGCATAGCG	G8	80	80+95*n
PI-81	CGCGAGCTCGTC	ACTCGATGAGCT	G9	81	81+95*n
PI-82	TAGAGTCTGTAT	TGTCTATCACAT	G10	82	82+95*n
PI-83	TACTATCGCTCT	TATGTGACATAC	G11	83	83+95*n
PI-84	TAGATGACGCTC	TACTCGTAGCGC	G12	84	84+95*n
PI-85	TCGCGTGACATC	ATCTACTGACGT	H1	85	85+95*n
PI-86	ACACGCTCTACT	ACAGTAGCGCAC	H2	86	86+95*n
PI-87	TACATAGTCTCG	CTAGTATCATGA	H3	87	87+95*n
PI-88	TGAGTAGCACGC	TCGATCATGCAG	H4	88	88+95*n
PI-89	TAGATGCTATAC	TACATGCACTCA	H5	89	89+95*n
PI-90	ATCGATGTCACG	CAGCTCGACTAC	H6	90	90+95*n
PI-91	ATCATATGTAGC	CTCTACAGTCAC	H7	91	91+95*n
PI-92	TAGCATCGATAT	AGATAGCACATC	H8	92	92+95*n
PI-93	TGATCGACGCTC	CTAGATATCGTC	H9	93	93+95*n
PI-94	TGCAGCTCATAG	TACAGACTGCAC	H10	94	94+95*n

PI-95	CGACGTAGAGTC	CAGTAGCACTAC	HII	95	95+95*n
-------	--------------	--------------	-----	----	---------

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	GGYGATATTCTAGTGTGGTCCCAA		
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 :

5 96℃ 2min
95℃ 30s → 60℃ 30s ^ 72℃ : 20s (32cycles)
15℃ ∞

HLA-A/B 的 PCR 反应体系如下

Promega 5 × buffer I (Mg²⁺ plus)	5.0ul
dNTP Mixture (各 2.5mM)	2.0ul
PI_{nf}-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
ddH₂O	10.8ul
Total	25.0ul

10 HLA-DRB1 的 PCR 反应体系如下:

Promega 5 × buffer I (Mg²⁺ plus)	5.0ul
dNTP Mixture (各 2.5mM)	2.0ul
PI_{nf}-D2-F1 (2pmol/ ul)	1.0ul
PI_{nf}-D2-F2 (2pmol/ ul)	1.0ul
PI_{nf}-D2-F3 (2pmol/ ul)	1.0ul
PI_{nf}-D2-F4 (2pmol/ ul)	1.0ul
PI_{nf}-D2-F5 (2pmol/ ul)	1.0ul
PI_{nf}-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_{nr} 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_{nr} 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

15 PCR 产物混合和纯化

对第 "X" 组 ("X" 为 1/2/3/4/5/6/7/8/9/10) 样本, 从 96 孔板 HLA-X-P-A2 剩余的 PCR 产物中 (阴性对照除外) 各取 20 ul 混合在一个 3ml 的 EP 管中, 标记为 HLA-X-A2-Mix, 对第 "X" 组样本的其它 6 个 96 孔板进行同样的操作, 分别标记为 HLA-X-A3-Mix、HLA-X-A4-Mix、HLA-X-B2-Mix、

HLA-X-B3-Mix、HLA-X-B4-Mix 和 HLA-X-D2-Mix，震荡混匀，
 从 HLA-X-A2-Mix、HLA-X-A3-Mix、HLA-X-A4-Mix、
 HLA-X-B2-Mix、HLA-X-B3-Mix、HLA-X-B4-Mix 和
 HLA-X-D2-Mix 中各取 200ul 混合在一个 3ml 的 EP 管中，标记
 5 为 HLA-X-Mix。从中各取 500ul DNA 混合物经 Qiagen DNA
 Purification kit 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
 司测定的 DNA 浓度分别为：

	HLA-1 -Mix	HLA-2 -Mix	HLA-3 -Mix	HLA-4 -Mix	HLA-5 -Mix	HLA-6 -Mix	HLA-7 -Mix	HLA-8 -Mix	HLA-9 -Mix	HLA-10 -Mix
浓度值 (ng/ul)	48.0	52.1	49.3	50.2	47.6	48.5	49.1	48.6	51.3	50.8

10 实施例 4

Illumina GA 测序文库构建

I.DNA 打断

从纯化后的 HLA-X-Mix 中各取总量 5ug 的 DNA 用带 AFA
 纤维扣盖的 Covaris 微管在 Covaris S2(Covaris 公司)上打断。打
 15 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-X-Mix 的所有打断产物用 QIAquick PCR
 20 Purification Kit(QIAGEN 公司)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-X-Mix 进行 DNA 末端修复反应 , 体系如下 (试剂均购自 Enzymatics 公司) :

DNA	37.5 μ L
H ₂ O	37.5 μ L
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

5 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

10 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°C 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 后的产物分别连接不同的 Illumina GA PCR-Free index 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 寡核苷酸标签接头混合物(30mM) (PCR-free index 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°C 温浴 15 min。

15 样本组与文库接头的对应关系如下

样本组 编号	HLA-1	HLA-2	HLA-3	HLA-4	HLA-5	HLA-6	HLA-7	HLA-8	HLA-9	HLA-10
文库接 头编号	1	2	3	4	5	6	7	8	9	10

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 摩尔浓度结果如下：

	HLA-1	HLA-2	HLA-3	HLA-4	HLA-5	HLA-6	HLA-7	HLA-8	HLA-9	HLA-10
	-Mix	-Mix	-Mix	-Mix	-Mix	-Mix	-Mix	-Mix	-Mix	-Mix
摩尔浓度 (nM)	78.90	72.13	79.33	80.21	77.68	78.50	89.12	78.60	81.32	80.82

6. 割胶回收

将 HLA-1-Mix 、 HLA-2-Mix 、 HLA-3-Mix, HLA-4-Mix 、
 5 HLA-5-Mix, HLA-6-Mix 、 HLA-7-Mix 、 HLA-8-Mix, HLA-9-Mix
 和 HLA-10-Mix 等摩尔混合 (终浓度 72.13nM/ul) ，标记为
 HLA-Mix-10, 取 30 μ L HLA-Mix-10 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
 10 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 9.96nM 。

实施例 5

15 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) 。

20

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
 测序结果中的接头标签序列、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/DRBI 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/DRBI 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/DRBI 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/DRBI 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 950 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/DRBI		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' 40:01	DRB1 *14:54
7	A*11:01	A*30:01	B*43:02	B *45:18	DRB1' ^04:04	DRB1 *07:01
8	A*01:01	A*02:01	B *5:03	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	ORBV ^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 *15:01	B *45:05	DRB1*04:06	DRB1 *15:01
13	A*01:01	A*11:02	B*07:02	B *45:02	DRB1' "09:01	DRB1' *15:01
14	A*01:01	A*02:01	B*52:01	B *67:01	ORBV "15:02	DRB1' *16:02
15	A*01:01	A*02:05	B*15:17	B*50:01	ORBV *07:01	DRB1: *15:01
16	A*01:01	A*11:01	B*37:01	B *40:02	ORBV *10:01	DRB1*12:02
17	A*24:07	A*32:01	B *35:05	B*40:01	ORBV ^03:01	DRB1 *04:05
18	A*11:01	A*24:02	B *13:01	B *35:01	ORBV *16:02	DRB1 *16:02
19	A*11:01	A*11:01	B *40:02	B *55:12	ORBV *04:05	DRB1' "15:01
20	A*02:11	A*24:02	B*40:01	B*40:06	DRB1' 41:01	DRB1*15:01

21	A*01:01	A*02:06	B*51:01	B*57:01	DRB1*07:01	DRB1*12:01
23	V*01:01	A*02:07	B*37:01	B*46:01	DRB1*04:03	DRB1*10:01
24	A*24:85	V*30:01	B*13:02	B*55:02	DRB1*07:01	DRB1*15:01
25	V*11:01	V*31:01	B*07:06	B*51:01	DRB1*12:02	DRB1*14:05
26	V*01:01	A*11:01	B*16:01	B*57:01	DRB1*07:01	DRB1*14:54
27	V*01:01	A*02:01	B*15:18	B*37:01	DRB1*10:01	DRB1*15:01
28	A*01:01	A*24:02	B*37:01	B*46:01	DRB1*09:01	DRB1*10:01
29	V*26:01	V*66:01	B*40:40	B*17:02	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*11:03	B*15:01	B*57:01	DRB1*07:01	DRB1*15:01
32	V*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-V/B/DR1 型

1	A*02:03	A*11:01	B*38:02	B*48:01	DRB1*14:54	DRB1*15:01
2	V*01:01	A*30:01	B*08:01	B*17:02	DRB1*03:01	DRB1*07:01
3	V*01:01	A*02:01	B*15:11	B*47:01	DRB1*13:02	DRB1*15:01
4	V*24:08	V*26:01	B*40:01	B*51:01	DRB1*0	DRB1*09:01
5	A*01:01	A*24:02	B*51:01	B*55:02	DRB1*04:03	DRB1*15:01
6	A*01:01	A*03:02	B*15:11	B*37:01	DRB1*10:01	DRB1*14:54
7	A*11:01	V*30:01	B*13:02	B*15:18	DRB1*04:04	DRB1*07:01
8	V*01:01	A*02:01	B*35:03	B*18:01	DRB1*11:01	DRB1*15:01
9	V*02:06	A*31:01	B*27:07	B*40:02	DRB1*03:01	DRB1*13:02
10	A*01:01	V*66:01	B*37:01	B*9:01	DRB1*10:01	DRB1*13:02
11	V*01:01	V*03:01	B*35:01	B*52:01	DRB1*10:10	DRB1*15:01
12	V*11:01	A*11:01	B*15:01	B*15:05	DRB1*04:06	DRB1*15:01
13	V*01:01	V*12:02	B*07:02	B*15:02	DRB1*09:01	DRB1*15:01
14	V*01:01	A*02:01	B*52:01	B*67:01	DRB1*15:02	DRB1*15:01
15	A*01:01	V*02:05	B*15:17	B*15:01	DRB1*07:01	DRB1*15:01
16	A*01:01	A*11:01	B*13:01	B*35:01	DRB1*16:02	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	V*ZD:02	B*13:01	B*35:01	DRB1*16:02	DRB1*16:02
19	V*11:01	A*11:01	B*40:02	B*52:12	DRB1*04:05	DRB1*15:01
20	V*02:11	A*24:02	B*40:01	B*40:06	DRB1*11:01	DRB1*15:01
21	V*01:01	V*02:06	B*13:01	B*57:01	DRB1*07:01	DRB1*12:02
22	V*01:01	V*29:01	B*07:05	B*15:01	DRB1*04:03	DRB1*15:01
23	V*01:01	A*02:07	B*37:01	B*46:01	DRB1*04:03	DRB1*15:01
24	A*24:85	V*30:01	B*13:02	B*55:02	DRB1*07:01	DRB1*15:01
25	V*11:01	V*31:02	B*15:01	B*57:01	DRB1*12:02	DRB1*14:05
26	V*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*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*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

注：HLA-DRB1 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 2 号外显子的序列完全相同。同理对于 HLA-A/B 位点中 2、3、4 号外显子序列完全相同的结果取常见型。

采用本发明的技术路线，对 950 份已知 HLA-SBT 分型结果的样本进行 HLA-A/B/DRB1 位点的基因分型，结果发现：采用本发明的技术路线所得的分型结果与原结果完全一致。

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.

5 [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.

10 [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.

15 [9]. Shannon J.Odelberg, Robert B.Weiss , Akira Hata. Template-switching during DNA synthesis by Therm us aquatic us DNA polymerase I. [J]. Nucleic Acids Research. 1995, 23:2049-2057.

20 [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(l):46-54.

权利要求

1. 一种 HLA 分型方法，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

5 自哺乳动物，更优选是人，特别是人的血样；

2) 将待分析的 n 个样品分成 m 个小组， m 为整数且 $n \geq m \geq 1$ ；

3) 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
10 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；

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

15 5) 打断：将所得的 PCR 产物文库进行不完全打断；

6) 建库：结合文库接头标签技术，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，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
20 的所有 DNA 条带，具体而言是 450-750bp 长度范围的 DNA 片段；

7) 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；

8) 拼接：基于各个文库不同的文库接头序列和每个样品独特
25 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；和

9) 分型：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
5 100%匹配的即为对应样本的HLA-DRB1基因型别。

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
10 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，更优选的所述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用于扩增HLA的特定基因的PCR引物进行设计，优选针对用于扩增
15 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，特别是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
20 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
25 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



或者多个的组合。

5. 权利要求1所述的测序方法,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。

5 6. 权利要求1所述的测序方法,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所述的测序方法,所述结合文库接头标签技术,将打断后的PCR产物文库构建PCR-Free测序文库是指使用m种文库接头给2)中得到的m个PCR产物文库加上接头,其中
10 每一个PCR产物文库使用一种不同的文库接头,从而构建m个接头标签测序文库;将m个接头标签测序文库等摩尔混合在一起构建混合接头标签测序文库。其中连接文库接头的方法是指不通过PCR程序直接采用DNA连接酶进行连接。

8. 组引物标签,其包括表1所示95对引物标签中的至少10
15 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

20 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。

25 9. 权利要求8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

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。

10. 权利要求9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-DRB12号外显子的PCR引物，优选的所述PCR引物如表2所示。

11. 权利要求8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10 任选通过连接序列连接）一个引物标签。

12. 权利要求11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-DRB12号外显子的PCR引物，优选的所述PCR引物如表2所示。

15 13. 权利要求11所述的标签引物用于PCR测序方法的用途。



Fig1

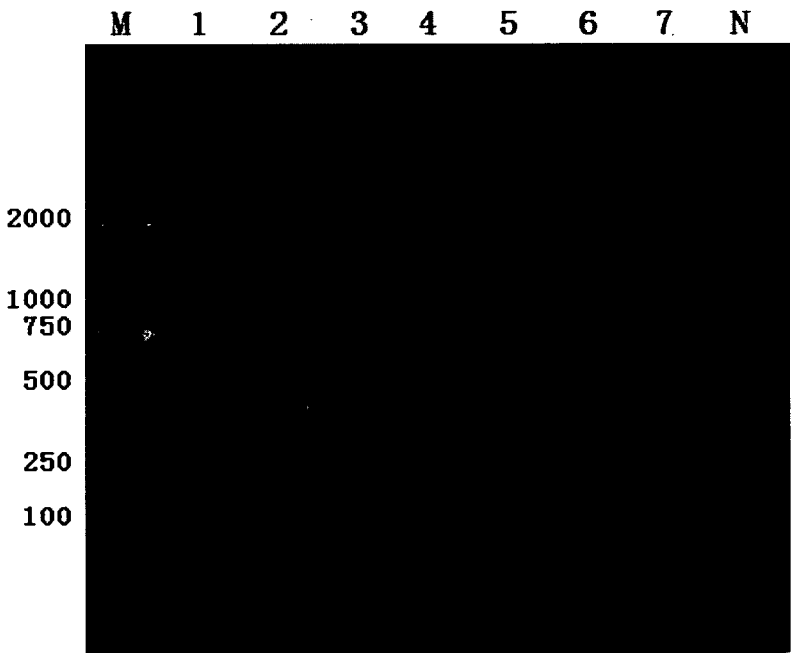


Fig 2

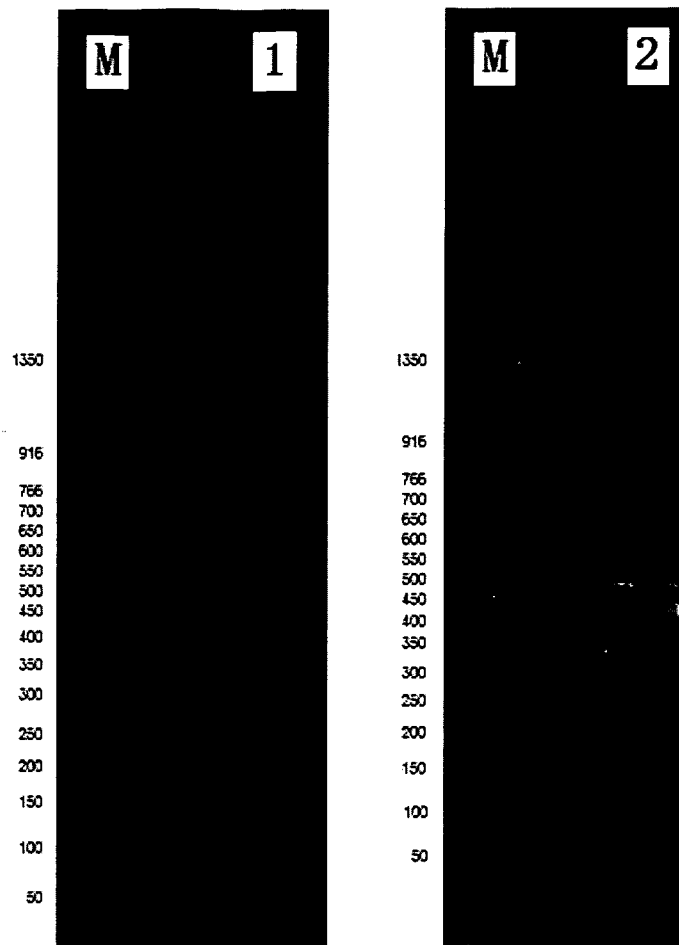


Fig 3



Fig 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1835

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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; C12N; C07H

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, EPODOC, WPI, ISI web of Knowledge, EMBASE, NCBI: human leukocyte antigen,, human leucocyte antigen, HLA, typing, sequenc+ ,next generation sequencing, second generation| sequencing, PCR, Illumina, pair-end, genome analyzer, primer, adapter, adaptor, index, label, tab, tag, mark, barcode, library, PCR-FREE, 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.
E	CN 101921840 A (SHENZHEN BGI CO LTD) 22 December 2010 (22.12.2010) see paragraphs [0015]-[0040] in description, Table 1, Table 2, Examples, claims 7-14	1-13
E	CN 101921842 A (SHENZHEN BGI CO LTD) 22 December 2010 (22.12.2010) see paragraphs [0016]-[0033] in description, Table 2, Table 3, Examples	1-13

☒ 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
24 March 2011 (24.03.2011)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

XING, Ying
Telephone No. (86-10) 62411044

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1835

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010/0086914 A I (GORDON BENTLEY [US]; HENRY ERLICH [US]; RUSSELL HIGUCHI [US]) 8 April 2010 (08.04.2010) see paragraphs [0005]-[0011], [0028], [0034]-[0035], [0039]-[0045], [0067], Fig.I, claim 1	1-7
A		8-13
Y	Michael L. Metzker. Sequencing technologies – the next generation. NATURE REVIEWS GENETICS. January 2010, Vol. 11 pages 31-46	1-7
A		8-13
A	CN 101654691 A (SHENZHEN BGI CO LTD) 24 February 2010 (24.02.2010) see the whole document	1-13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/001835

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item item1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of:

a. a sequence listing filed or furnished

☐ on paper

☒ in electronic form

b. time of filing or furnishing

☐ contained in the application as filed

☒ filed together with the application in electronic form

☐ furnished subsequently to this Authority for the purposes of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2010/001835

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101921840 A	22.12.2010	NONE	
CN 101921842 A	22.12.2010	NONE	
U S 2010086914 A I	08.04.2010	W O 2009049889 A I	23.04.2009
		EP 2203567 A I	07.07.2010
		CA 2701411 A I	23.04.2009
CN 101654691 A	24.02.2010	NONE	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1835

CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i

C12N 15/11 (2006.01) i

C07H 21/00 (2006.01) i

A. 主题的分类

见附加页

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; C12N; C07H

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, EPODOC, WPI, ISI web of Knowledge, EMBASE, NCBI: 人(类)白细胞抗原, 分型, 测序, 第二代测序, 引物, 接头, 标签, 标记, 索引, 条码, 文库, 不完全打断, human leukocyte antigen, human leucocyte antigen, HLA, typing, sequenc+, next generation sequencing, second generation sequencing, PCR, IUumina, pair-end, genome analyzer, primer, adapter, adaptor, index, label, tab, tag, mark, barcode, library, PCR-FREE, 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
E	CN 101921840 A (深圳华大基因科技有限公司) 22.12 月 2010 (22.12.2010) 参见说明书第 [0015]-[0040] 段, 表 1, 表 2, 实施例, 权利要求 7-14	1-13
E	CN 101921842 A (深圳华大基因科技有限公司) 22.12 月 2010 (22.12.2010) 参见说明书第 [0016]-[0033] 段, 表 2, 表 3, 实施例	1-13



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 J% % %

"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

24.3 月 201 1 (24.03.201 1)

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 100088

传真号: (86-10)62019451

受权官员

幸颖

电话号码: (86-10) 62411044

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
Y	US 2010/0086914 A I (GORDON BENTLEY [US]; HENRY ERLICH [US]; RUSSELL HIGUCHI [US]) 8.4 月 2010 (08.04.2010) 参见第 [0005]-[0011] 、 [0028] 、 [0034]-[0035] 、 [0039]-[0045] 、 [0067] 段，图 1，权利要求 1	1-7
A		8-13
Y	Michael L. Metzker. Sequencing technologies – the next generation. NATURE REVIEWS GENETICS. 1 月 2010, 第 11 卷，第 31-46 页	1-7
A		8-13
A	CN 101654691 A (深圳华大基因科技有限公司) 24.2 月 2010 (24.02.2010) 参见全文	1-13

第 I 栏 **K** 核苷酸和/或氨基酸序列(续第 1 页第 I.c 项)

1、关于国际申请中所公开的任何对要求保护的发明必要的核苷酸和/或氨基酸序列,国际检索是在下列基础上进行的:

a. 序列表的提交或提供

☐ 纸件形式

☒ 电子形式

b. 提交或提供时间

包含在申请提交时的国际申请中

☒ 以电子形式与国际申请一起提交

☐ 为检索之用随后提交本单位

2、☐ 另外,在提交/提供了多个版本或副本的序列表的情况下,提供了关于随后提交的或附加的副本中的信息与申请时提交的申请中的信息相同或未超出申请时提交的申请中的信息范围(如适用)的所需声明。

3. 补充意见

国际检索报告

关于同族专利的信息

国际申请号 PCT/CN2010/001835

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN 101921840 A	22.12.2010	无	
CN 101921842 A	22.12.2010	无	
US 2010086914 A I	08.04.2010	WO 2009049889 A 1 EP 2203567 A 1 CA 2701411 A 1	23.04.2009 07.07.2010 23.04.2009
CN 101654691 A	24.02.2010	无	

主题的分类

072100 (2006-01) 1
1511 (2006-01) 1
1681 (2006-01) 1