

(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
2017 年 8 月 24 日 (24.08.2017)



(10) 国际公布号
WO 2017/139945 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2016/074027

(22) 国际申请日: 2016 年 2 月 18 日 (18.02.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 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: 章元伟 (ZHANG, Yuanwei); 中国广东省深
圳市盐田区北山工业区综合楼, Guangdong 518083
(CN)。 张涛 (ZHANG, Tao); 中国广东省深圳市盐
田区北山工业区综合楼, Guangdong 518083 (CN)。
曹红志 (CAO, Hongzhi); 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 PROP-
ERTY 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, IR, IS,
JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU,
LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA,
RO, RS, RU, RW, SA, 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, ST, 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, KM, 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: TYPING METHOD AND DEVICE

(54) 发明名称: 分型方法和装置

(57) Abstract: Disclosed is a typing method, comprising the steps of: obtaining sequencing data; aligning a reference genome to a baseline sequence group and obtaining a type set; aligning the sequencing data to a reference sequence group and obtaining an alignment result; converting the alignment result to an alignment result relative to the conversion of the baseline sequence; assembling reads aligned to the same baseline sequence and obtaining a haplotype; determining the types supported by the haplotype; grouping the types and obtaining a first and a second candidate type group; respectively screening the types in the first and the second candidate type group to determine the primary and secondary types; and determining the genotype of the target region based on the difference in the number of reads supporting the primary and secondary types. This typing method is applicable to genotyping of any region, and is particularly applicable to the typing of highly polymorphic regions, such as the typing of HLA-DRB3, 4 and/or 5.

(57) 摘要: 本发明公开一种分型方法, 包括步骤: 获取测序数据; 比对参考基因组至基准序列组, 获得型别集合; 比对测序数据至参考序列组, 获得比对结果; 将比对结果转化为相对于基准序列的转化的比对结果; 对比对上同一基准序列的读段进行组装, 获得单体型; 确定单体型支持的型别; 对型别进行分组, 获得第一和第二候选型别组; 分别对第一和第二候选型别组中的型别进行筛选, 以确定出主要和次要型别; 基于支持主要和次要型别的读段数目的差异, 判定目标区域的基因型。该分型方法适用于任何区域的基因分型, 特别适用于高度多态性区域的分型, 例如对 HLA-DRB3、4 和/或 5 的分型。

WO 2017/139945 A1

分型方法和装置

技术领域

本发明涉及生物信息领域，具体的，本发明涉及分型方法和装置。

背景技术

HLA-DRB3, DRB4 和 DRB5 属于 HLA (人类白细胞抗原) II 类分子的 β 链的同源编码基因。

HLA II 类分子表达在抗原递呈细胞的细胞膜上，能递呈细胞外源的蛋白的肽段，在免疫系统中起着中心作用。HLA-DRB3 被报道与克罗恩病，格雷夫斯病，I 型糖尿病等相关。HLA-DRB4 被报道与儿童急性淋巴细胞白血病，桥本甲状腺炎，变应性肉芽肿性血管炎，白癜风等相关。HLA-DRB5 被报道与瘢痕疙瘩，系统性红斑狼疮，多发性硬化症，嗜睡症等相关。

对 HLA-DRB3, DRB4, DRB5 进行分型有着重要的医学应用和疾病研究价值。

而已有的 HLA-DRB3,4,5 的分型方法主要为外显子 PCR 结合基因测序，或者为长片段 PCR 结合基因测序，涉及必需的 PCR 引物设计、试验步骤较多以及无法应用于高通量全基因组测序或高通量芯片捕获测序等问题。

发明内容

本发明旨在至少解决上述问题至少之一或者提供至少一种可选择的商业手段。

依据本发明的一方面，本发明提供一种分型方法，该方法包括：获取待测样品的测序数据，所述测序数据包括多个来自目标区域的读段；将参考序列组比对至基准序列组，获得型别集合，所述型别集合包含多个型别，所述基准序列组包含一条或多条基准序列，所述基准序列组能够完全覆盖所述目标区域的所有基因序列，不同所述基准序列包含不同的基因全序列，所述参考序列组包括多条参考序列，所述参考序列组能够完全覆盖所述目标区域的编码区序列的所有外显子，不同所述参考序列包含不同的外显子；将所述测序数据比对至所述参考序列组，获得比对结果；将所述比对结果转化为相对于所述基准序列组的比对结果，获得转化的比对结果；基于所述转化的比对结果，分别对对比对上相同基准序列的读段进行组装，获得组装结果，所述组装结果包括多个单体型；比较所述单体型上的变异和所述型别集合中的型别，以确定所述单体型支持的型别；依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组；

基于支持所述型别的单体型和支持所述型别的读段，分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选，以确定出主要型别和次要型别；基于支持所述主要型别和支持所述次要型别的读段的数目的差异，判定所述目标区域的基因型。

依据本发明的另一方面，本发明提供一种计算机可读介质，该计算机可读介质用于存储计算机可执行程序，执行所述程序包括完成上述本发明一方面的分型方法。本领域技术人员可以理解，在执行该计算机可执行程序时，通过指令相关硬件可完成上述分型方法的全部或部分步骤。所称存储介质可以包括：只读存储器、随机存储器、磁盘或光盘等。

依据本发明的再一方面，本发明提供一种分型装置，该装置包括：数据输入单元，用于输入数据；数据输出单元，用于输出数据；存储单元，用于存储数据，其中包括计算机可执行程序；处理器，与所述数据输入单元、所述数据输出单元和所述存储单元连接，用于执行所述计算机可执行程序，执行所述程序包括完成上述分型方法。

依据本发明的又一方面，本发明提供一种分型系统，包括：数据输入模块，用于输入待测样品的测序数据，所述测序数据包括多个来自目标区域的读段；比对模块，包括第一比对模块和第二比对模块，所述第一比对模块用于将参考序列组比对至基准序列组，获得型别集合，所述型别集合包含多个型别，第二比对模块，用于将来自数据输入模块的测序数据比对至所述参考序列组，获得比对结果，所述基准序列组包含一条或多条基准序列，所述基准序列组能够完全覆盖所述目标区域的所有基因序列，不同所述基准序列包含不同的基因全序列，所述参考序列组包括多条参考序列，所述参考序列组能够完全覆盖所述目标区域的编码区序列的所有外显子，不同所述参考序列包含不同的外显子；转化模块，用于将所述比对结果转化为相对于所述基准序列组的比对结果，获得转化的比对结果；组装模块，用于基于所述转化的比对结果，分别对对比上相同基准序列的读段进行组装，获得组装结果，所述组装结果包括多个单体型；单体型支持型别确定模块，用于比较所述单体型上的变异和所述型别集合中的型别，以确定所述单体型支持的型别；聚类模块，用于依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组；主/次要型别确定模块，用于基于支持所述型别的单体型和支持所述型别的读段，分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选，以确定出主要型别和次要型别；基因型确定模块，用于基于支持所述主要型别和支持所述次要型别的读段的数目的差异，判定所述目标区域的基因型。

本发明的方法通过构建目标区域的参考序列和基准序列，构建型别集合，基于读段信息构建单体型，比较型别与单体型以对目标区域进行基因分型。本发明的分型方法适用于

任何区域的基因分型，特别适用于高度多态性区域的分型，例如适用于对 HLA-DRB3,4 和/或 5 的分型，特别是基于包含有任意形式的 HLA-DRB3,4 和/或 5 的序列信息的测序数据、对 HLA-DRB3,4 和/或 5 的分型。本发明的方法不需要针对目标基因的外显子进行 PCR 或者长片段 PCR，降低了实验工作量和实验难度，提高了应用或研究时方案设计的灵活性。

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 是本发明的实施例中的分型方法的流程图。

图 2 是本发明的实施例中的分型方法的流程图。

图 3 是本发明的实施例中的分型方法的流程图。

图 4 是本发明的实施例中的分型装置的结构示意图。

图 5 是本发明的实施例中的分型系统的结构示意图。

图 6 是本发明的实施例中的分型系统的结构示意图。

具体实施方式

如图 1 所示，根据本发明的一个实施方式提供的分型方法，包括以下步骤：

S10 获取测序数据。

获取待测样品的测序数据，所述测序数据包括多个来自目标区域的读段。

所称的测序数据通过对待测样品的核酸序列进行测序文库制备、上机测序获得。根据本发明的实施例，获取所述测序数据，包括：获取待测样本中的核酸，制备所述核酸的测序文库，对所述测序文库进行测序。测序文库的制备方法根据所选择的测序方法的要求进行，测序方法依据所选的测序平台的不同，可选择但不限于 Illumina 公司的 Hisq2000/2500 测序平台、Life Technologies 公司的 Ion Torrent 平台、BGI 的 BGISEQ 平台和单分子测序平台，测序方式可以选择单端测序，也可以选择双末端测序，获得的下机数据是测读出来的片段，称为读段（reads）。

目标区域可以为任何感兴趣的基因区域。根据本发明的实施例，目标区域包括 MHC（主要组织相容性复合体，major histocompatibility complex）基因家族中的成员。哺乳动物的 MHC 基因（MHC gene）呈高度多态性，人类的 MHC 通常被称为 HLA（human leucocyte antigen, HLA），即人类白细胞抗原。根据本发明的一个实施例，待测样品来自人类，目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5 中的至少之一。根据本发明的另一个实

施例，目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5。利用本发明的方法，能够不进行 PCR，只要基于任意形式的包含 HLA-DRB3、HLA-DRB4 和/或 HLA-DRB5 的测序数据，就能对这类高度多态性区域进行准确分型。

S20 比对。

比对参考序列组至基准序列组，获得型别集合，所述型别集合包含多个型别。

比对所述测序数据至参考序列组，获得比对结果。

上述两个比对过程是独立进行的，无顺序限制。实际上获得两个比对结果，即为所称的型别集合和比对结果。

所称基准序列组包含一条或多条基准序列，基准序列为预先确定的包含基因全长的序列，可以是预先获得的待测样本所属生物类别的参考模板，例如，若待测样本来源的为人类个体，基准序列可选择 NCBI 数据库提供的 HG19 中包含的目标区域中的基因的基因全序列。根据本发明的一个实施例，一条基准序列为目标区域中的一个基因的基因全序列。若一个基因有多条基因全序列，可以选择其中最长的基因全序列作为该基因的基准序列。所称基准序列组能够完全覆盖目标区域的所有基因序列，不同的基准序列包含不同的基因全序列。

所称参考序列组包括多条参考序列，参考序列为预先确定的包含外显子的序列，可以是预先获得的待测样本所属生物类别的参考模板，例如，若待测样本来源的为人类个体，参考序列可选择 NCBI 数据库提供的 HG19 中包含的目标区域的外显子的序列，进一步地，也可以预先配置包含更多参考序列的资源库，例如依据待测样本来源个体的状态、地域等因素选择或是测定组装出更接近的序列作为参考序列。参考序列组能够完全覆盖目标区域的编码区序列的所有外显子，不同的参考序列包含不同的外显子。

所称的型别为等位基因。某个型别是某个基因的某个等位基因，本质上是某个基因组区域的特定变异的集合。某些国际组织如 WHO 会对某些基因的已知等位基因命名，一般有被命名的等位基因才会被称作型别。根据本发明的一个实施例，目标区域包含 HLA 基因的至少一部分。由于 HLA 基因的等位基因数目众多，且对医学移植有关键作用，因此 WHO 对 HLA 基因的所有已知等位基因都给了命名，例如 DRB3*01:01:01，某个型别就指代某个命名的等位基因。分型是利用各种方法确定目标个体的目标基因的型别的过程。

所称的型别集合中提供的型别是相对于基准序列组的位置信息和变异信息。根据本发明的一个实施例，选取最长的基因全序列作为基准序列，记录所有编码区序列和基因全序

列相对于基准序列的变异，将变异与型别相关联。建立变异信息与型别的关联，利于后续基于检测出的变异建立型别基准、利于基因分型。

所称的比对测序数据至参考序列组以获得比对结果，该比对结果包含比对上所述参考序列组中的任一参考序列的读段的位置信息和变异信息。

在本实施方式中，不将测序数据直接比对至基准序列组，是因为基准序列组只包含每个基因的一种等位基因的序列，利用上述方式，即先将测序数据比对至参考序列组、再基于基准序列转化比对结果，由于参考序列组包含所有等位基因，使得能够增加比对上的读段（reads）的数目，显著提高数据利用率。

根据本发明的一个实施例，目标区域为高度多态性区域，参考序列组通过以下步骤构建：获得包含目标区域的编码区序列和基因全序列；将编码区序列按外显子分割，获得多个外显子序列；从与外显子序列最接近型别的基因全序列提取该外显子序列两侧各 K bp 的序列，加到对应的外显子序列的两侧，获得参考序列组中的参考序列， K 为读段的长度。例如，欲对 HLA-DRB3、HLA-DRB4 和/或 HLA-DRB5 进行分型，可从 IMGT/HLA 数据库下载 HLA 基因型别和序列数据，序列数据包括多条编码区序列和多条基因全序列。可以修改下载的 HLA 基因型别和序列数据的格式，以方便于后续步骤分析。接着，将各个编码区序列按外显子分割，并从最相近型别的基因全长序列提取外显子两侧各 K bp 加在对应的外显子序列的两侧，构成参考序列。 K 取决于读段的长度，延伸外显子序列是为了保证比对时能保留比对至外显子边缘的读段，使得该部分数据能够得以利用，提高数据利用率。需要说明的是，根据读段的长度为均一长度或者为不等长度， K 可变化，再者， K 不须为一个确定值或者为一个精确值，可以为一个数值范围，例如为任一读段的长度的正负 10% 的区间，而且，加到外显子序列两侧的 K 可取不同数值。根据本发明的一个实施例，目标分型区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5，发明人从 IMGT/HLA 数据库获得 94 条已公布的 HLA-DRB3、HLA-DRB4 或 HLA-DRB5 的编码区序列以及 7 条 HLA-DRB3、HLA-DRB4 或 HLA-DRB5 的基因全长序列，各编码区序列或各基因全长序列在长度、包含的外显子个数、特定位置的碱基序列或结构等方面有所差异，构建的参考序列组包含 166 条参考序列。

所称的与外显子序列最接近型别的基因全序列，是指该基因全序列上的相应外显子序列与该外显子序列的匹配度最高，即序列最相似、差异最小。例如，外显子序列上多个特定位置存在不同于参考基因组的碱基，存在某一基因全长序列的对应外显子区域的相应位置也都为与该外显子序列的相同，则确定该基因全长序列与该外显子序列的型别最接近。

所称的“比对上”意同匹配。具体比对时，可以利用已知比对软件进行，例如 SOAP、BWA 和 TeraMap 等，本实施方式对此不作限制。在比对过程中，根据比对参数的设置，一对或一条读段最多允许有 n 个碱基错配 (mismatch)，例如设置 n 为 1 或 2，若读段中有超过 n 个碱基发生错配，则视为该对读段无法比对到参考序列，或者，若错配的 n 个碱基全部位于读段对中的一个读段，则视为该读段对中的该读段无法比对到参考序列。

当匹配为完全匹配，例如当读段与参考序列上存在变异的那段序列无错配，即读段上包含该变异位点或包含该变异发生时应当具有的特征序列，则称这种读段为支持该变异的读段。

S30 获得转化的比对结果。

将比对结果转化为相对于基准序列组的比对结果，获得转化的比对结果。即，将比对结果中的所有位置信息和变异信息转化为相对于基准序列组的位置信息和变异信息，获得转化的比对结果。如此，获得的转化的比对结果与 S20 中的型别集合都是基于基准序列，便于后续比较二者进行基因型判断。

根据本发明的一个实施例，选取每个目标基因的最完整的基因全序列作为该基因的基准序列，将比对结果中的比对位置信息和变异信息转化为相对于基准序列组的位置信息和变异信息，获得转化的比对结果，利于后续基于检出的变异信息对型别进行筛选判断以确定基因型。

S40 组装。

基于所述转化的比对结果，分别对比对上相同基准序列的读段进行组装，获得组装结果，所述组装结果包括多个单体型。

所称单体型(Haplotype)为位于一条染色体特定区域的一组相互关联、并倾向于以整体遗传给后代的单核苷酸多态的组合，又称单倍体型或单元型。

序列的组装可以根据已知的组装方法，本实施方式对此不作限制。根据本发明的一个实施例，该步骤包括：对比对上同一参考序列的读段，利用具有重叠部分并且所述重叠部分完全一致的读段进行所述组装，以获得所述多个单体型。

由于基准序列组中的基准序列均为包含一个基因的全序列，而真核生物的一个基因一般包含多个外显子，所以基于比对上同一基准序列的读段，利用其中具有重叠部分并且重叠部分序列完全一致的读段组装而获得的多个单体型，实质上为包含不同碱基序列或者不同长度的至少一个外显子序列的至少一部分。根据本发明的一个实施例，在获得所述组装

结果之后，过滤掉对所述外显子的覆盖度低于 95% 的单体型。如此，弃去长度相对短的序列，可以提高后续分析步骤所使用的数据的可靠性，减少数据复杂度，利于准确分型。

需要说明的是，S20 中的获得型别集合的步骤，与 S30 和 S40 无进行顺序先后的限制。例如，可以先进行 S20 中的获得型别集合的步骤，再进行 S30 和 S40；也可以先进行 S30 和 S40，再进行 S20 中的获得型别集合的比对步骤。

S50 确定单体型支持的型别。

比较所述单体型上的变异和所述型别集合中的型别，以确定所述单体型支持的型别。

根据本发明的一个实施例，在进行该步骤之前包括进行：分别基于比对上相同基准序列的读段的组装结果，对每个单体型进行打分，基于获得的单体型的分值对单体型进行筛选，获得候选单体型；以候选单体型替代单体型进行后续步骤。如此，利于减少数据干扰，减少需要处理的数据量。

根据本发明的再一个实施例，上述分别基于比对上相同基准序列的读段的组装结果，对每个所述单体型进行打分，基于获得的单体型的分值对所述单体型进行筛选，获得候选

单体型，包括：利用以下公式确定所述单体型的分值 Score，
$$Score = \frac{c^2 \times N}{\log R}$$
，其中，c 为所述单体型对所述外显子的覆盖度，N 为支持所述单体型的读段数目，R 表示所述单体型的

可靠性，
$$R = \frac{\sum [(X_i - X)^2 \div X]}{\log(L-1)}$$
， X_i 为所述单体型的位置 i 的测序深度，i 为所述单体型上的位置的编号，X 为所述单体型的平均深度，L 为所述单体型的长度。利用以上公式，基于单体型自身的平均测序深度、各个位置的测序深度以及长度来评判该单体型的可靠性，进而依据该单体型的可靠性、对其对应的外显子的覆盖度以及支持其的读段的数目，来赋予该单体型分值，该分值能够反映代表该单体型，利于多个单体型的比较评判。

所称的筛选包括：对于比对上同一基准序列的读段的组装结果，取出其中的最高的单体型的打分值所对应的单体型后，去除满足以下条件的所述组装结果中的单体型：与取出的所述单体型的序列不一致并且不一致位点的测序深度低于所述单体型相应位点的测序深度的 20%；重复以上步骤，直至取出不超过 4 个单体型，以获得所述候选单体型。由于目标区域来自二倍体，一个区域至多有两种型别（即为杂合子时），利用该步骤进行单体型筛选，使得每个外显子至多具有相应的 4 个单体型，筛掉不符合真实情况或没有意义的单体型，降低数据的复杂度，利于后续分析的快速进行以及准确分型。

根据本发明的一个实施例，以候选单体型替代单体型进行后续步骤，该步骤包括：比较每个候选单体型上的变异和所述型别集合中的型别，若完全匹配，则确定所述候选单体型支持所述型别。如此，确定候选单体型支持的型别，利于后续基于候选单体型支持情况对型别进一步判定。

根据本发明的一个实施例，确定所述候选单体型支持的型别之后，合并支持相同型别的所述候选单体型。利于后续步骤的快速进行。

S60 确定第一候选型别组和第二候选型别组。

依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组。

目标区域的编码序列包含多个外显子，根据本发明的一个实施例，在进行该步骤之前，过滤掉支持的外显子数目小于外显子总数 30% 的型别。如此，依据该条件去除掉型别集合中的没有意义或者意义小的型别，降低数据的干扰、分型的复杂度，利于后续快速进行以及准确分型。

如图 2 所示，根据本发明的一个实施例，该步骤包括：依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别；基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，以获得所述第一候选型别组和所述第二候选型别组。

根据本发明的一个实施例，所称的依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别，包括：利用以下公式确定所述型别的第一分值 $TScore$ ， $TScore = N \times S$ ，其中， N 为支持所述型别的读段数目， S 为支持所述型别的候选单体型的第一分值的总和；将所有所述型别的第一分值作两两组合，确定所述第一分值之和最高的组合中的型别分别为所述第一候选型别和所述第二候选型别。利用以上公式赋予型别第一分值，该分值能够反映代表该型别，利于后续型别判定。

根据本发明的一个实施例，所述基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，获得所述第一候选型别组和所述第二候选型别组，包括：对于每个所述其它型别，比较第一交集和第二交集的大小，依据比较的结

果将每个所述其它型别归至所述第一候选型别或者所述第二候选型别，以获得所述第一候选型别组和所述第二候选型别组，所述第一交集为支持该其它型别的读段与所述支持第一候选型别的读段的交集，所述第二交集为支持该其它型别的读段与所述支持第二候选型别的读段的交集。若第一交集较第二交集大，即支持该个其它型别的读段落在支持第一候选型别的读段中的数目更多，则将该个其它型别归到与第一候选型别同一组，否则归为第二候选型别所在组。如此，基于第一和第二候选型别，将型别集合中的其它型别分成了两个组。这样处理，利于后续进一步筛选准确分型。

S70 确定主要型别和次要型别。

基于支持所述型别的单体型和支持所述型别的读段，分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选，以确定出主要型别和次要型别。

根据本发明的一个实施例，该步骤包括：基于支持所述型别的读段和所述型别的第一分值，分别对所述第一候选型别组和所述第二候选型别组中的型别进行第二打分，基于获得的型别的第二分值确定出所述主要型别和所述次要型别。

根据本发明的一个实施例，利用以下公式确定所述型别的第二分值 $TScore_New$ ， $TScore_New = N^* \times TScore$ ，其中， N^* 为所述第一候选型别组中的型别的支持读段落在所述支持第二候选型别的读段之外的数目，或者为所述第二候选型别组中的型别的支持读段落在所述支持第一候选型别的读段之外的数目；确定所述第一候选型别组和所述第二候选型别组中的第二分值最高的两个型别为所述主要型别和所述次要型别。所称的主要型别和次要型别，是基于频率的相对大小的相对概念。这里，主要型别和次要型别是基于支持其的读段的相对多和少来区分的。利用以上公式调整前面步骤赋予型别的第一分值，即对第一分值进行修正，得到的第二分值，更能够反映代表该型别，依据第二分值的高低确定出主要型别和次要型别利于后续型别的准确判定。

S80 确定目标区域的基因型。

基于支持所述主要型别和支持所述次要型别的读段的数目的差异，判定所述目标区域的基因型。

根据本发明的一个实施例，该步骤包括：若仅支持所述次要型别的读段数目与仅支持所述主要型别的读段数目的比值大于 0.1，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别，否则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别。发明人基于前述步

骤的中间结果以及大量样本数据分析，发现将仅支持所述次要型别的读段数目与仅支持所述主要型别的读段数目的比例的阈值设置为 0.1，能够简单准确的确定目标区域的基因型。

需要说明的是，本发明中涉及的具体数值大多具有统计意义，因此，如无特殊说明，任意以精确方式表达的数值均可代表一个范围，例如包含该数值正负 10% 的区间。

如图 3 所示，根据本发明的实施例，上述实施方式中的分型方法还可以进一步包括以下步骤：

S90 判定目标区域的拷贝数。

以至少一个拷贝数固定为 2 的区域的平均测序深度为基准深度，计算目标区域的测序深度与所述基准深度的比例，依据计算得到的比例确定所述目标区域的拷贝数。根据本发明的一个实施例，该步骤基于利用测序深度信息，可以判断目标区域 HLA-DRB3、4 和/或 5 的拷贝数。

依据所述目标区域的拷贝数，进一步修正判定的所述目标区域的基因型，以更加准确的确定目标区域的基因型。根据本发明的一个实施例，若所述目标区域的拷贝数为 0，则判定所述目标区域不存在；若所述目标区域的拷贝数为 1，则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别；若所述目标区域的拷贝数为 2，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别。需要说明的是，该处涉及的具体数值具有统计意义，因此，这里以精确方式表达的数值“0”、“1”和“2”均可代表一个范围，例如包含该数值正负 10% 或者正负 20% 的区间。

本领域技术人员可以理解，上述分型方法的全部或部分步骤，可以利用机器可识别语言编写成程序，存储于存储介质中。根据本发明的另一个实施方式提供的一种计算机可读介质，该计算机可读介质用于存储计算机可执行程序，执行所述程序包括完成上述任一实施例中的分型方法的全部或部分步骤。本领域技术人员可以理解，在执行该计算机可执行程序时，通过指令相关硬件可完成上述任一分型方法的全部或部分步骤。所称存储介质可以包括：只读存储器、随机存储器、磁盘或光盘等。

如图 4 所示，根据本发明的再一个实施方式提供的一种分型装置 100，该装置 100 包括：数据输入单元 110，用于输入数据；数据输出单元 120，用于输出数据；存储单元 140，用于存储数据，其中包括计算机可执行程序；处理器 130，与所述数据输入单元、所述数据输出单元和所述存储单元连接，用于执行所述计算机可执行程序，执行所述程序包括完成上述任一实施例中的分型方法的全部或部分步骤。

如图 5 所示, 根据本发明的一个实施方式还提供的一种分型系统 1000, 该系统 1000 能够用于实施上述本发明任一实施例中的分型方法, 该系统 1000 包括: 数据输入模块 1001, 用于输入待测样品的测序数据, 所述测序数据包括多个来自目标区域的读段; 比对模块 1003, 包括第一比对模块 1013 和第二比对模块 1023, 第一比对模块 1013 用于将参考序列组比对至基准序列组, 获得型别集合, 所述型别集合包含多个型别, 第二比对模块 1023 用于将来自数据输入模块的测序数据比对至所述参考序列组, 获得比对结果, 所述基准序列组包含一条或多条基准序列, 所述基准序列组能够完全覆盖所述目标区域的所有基因序列, 不同所述基准序列包含不同的基因全序列, 所述参考序列组包括多条参考序列, 所述参考序列组能够完全覆盖所述目标区域的编码区序列的所有外显子, 不同所述参考序列包含不同的外显子; 转化模块 1005, 用于将所述比对结果转化为相对于所述基准序列组的比对结果, 获得转化的比对结果; 组装模块 1007, 用于基于所述转化的比对结果, 分别对比对上相同基准序列的读段进行组装, 获得组装结果, 所述组装结果包括多个单体型; 单体型支持型别确定模块 1009, 用于比较所述单体型上的变异和所述型别集合中的型别, 以确定所述单体型支持的型别; 聚类模块 1011, 用于依据支持所述型别的单体型和支持所述型别的读段, 将所述型别分成两组, 获得第一候选型别组和第二候选型别组; 主/次要型别确定模块 1013, 用于基于支持所述型别的单体型和支持所述型别的读段, 分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选, 以确定出主要型别和次要型别; 基因型确定模块 1015, 用于基于支持所述主要型别和支持所述次要型别的读段的数目的差异, 判定所述目标区域的基因型。上述对本发明任一实施例中的分型方法的优点和技术特征的描述, 同样适用本发明这一实施方式的分型系统, 在此不再赘述。本领域技术人员能够理解, 上述系统中的功能模块可以包含子模块或者连接其它功能模块以实施分型方法中的可选或优化步骤或处理。

根据本发明的一个实施例, 所述目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5 中的至少之一。

根据本发明的一个实施例, 所述目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5。

根据本发明的一个实施例, 比对模块 1003 中的参考序列组通过以下步骤构建获得: 获得包含所述目标区域的编码区序列和基因全序列; 将所述编码区序列按外显子分割, 获得多个外显子序列; 从与所述外显子序列最接近型别的基因全序列提取该外显子序列两侧各 K bp 的序列, 加到对应的外显子序列的两侧, 获得所述参考序列组中的参考序列, 其中 K 依据读段的长度来确定。

根据本发明的一个实施例，第一比对模块 1013 和/或转化模块 1005 中的一条所述基准序列为所述目标区域中的一个基因的基因全序列。

根据本发明的一个实施例，利用组装模块 1007 进行以下：对比对上相同基准序列的读段，利用具有重叠部分并且所述重叠部分完全一致的读段进行所述组装，以获得所述多个单体型。

根据本发明的一个实施例，在组装模块 1007 中，在获得所述组装结果之后，过滤掉对所述外显子的覆盖度低于 95% 的单体型。

根据本发明的一个实施例，还包括候选单体型确定模块 1008，用于在利用所述单体型支持型别确定模块 1009 确定单体型支持的型别之前，进行以下：分别基于比对上相同基准序列的读段的组装结果，对每个所述单体型进行打分，基于获得的单体型的分值对所述单体型进行筛选，获得候选单体型；以所述候选单体型替代所述单体型。

根据本发明的一个实施例，在候选单体型确定模块 1008 中，利用以下公式对每个所述

单体型进行打分，以确定所述单体型的分值 Score：
$$Score = \frac{c^2 \times N}{\log R}$$
，其中，c 为所述单体型对所述外显子的覆盖度，N 为支持所述单体型的读段数目，R 表示所述单体型的可靠性，

$$R = \frac{\sum [(X_i - X)^2 \div X]}{\log(L-1)}$$
，Xi 为所述单体型的位置 i 的测序深度，i 为所述单体型上的位置的编号，X 为所述单体型的平均深度，L 为所述单体型的长度。

根据本发明的一个实施例，在候选单体型确定模块 1008 中进行基于获得的单体型的分值对所述单体型进行筛选，包括：对于比对上相同基准序列的读段的组装结果，取出其中的最高的单体型的分值所对应的单体型后，去除满足以下条件的所述组装结果中的单体型：与取出的所述单体型的序列不一致并且不一致位点的测序深度低于所述单体型相应位点的测序深度的 20%；重复以上至最多取出 4 个所述单体型，获得所述候选单体型。

根据本发明的一个实施例，在单体型支持型别确定模块 1009 中进行以下：比较每个候选单体型上的变异和所述型别集合中的型别，若完全匹配，则确定所述候选单体型支持所述型别。

根据本发明的一个实施例，在单体型支持型别确定模块 1009 中，确定所述候选单体型支持的型别之后，合并支持相同型别的所述候选单体型。

根据本发明的一个实施例，还包括型别过滤模块 1010，用于在利用聚类模块 1011 获得第一候选型别组和第二候选型别组之前，过滤掉支持的外显子数目小于外显子总数 30%

的型别。

根据本发明的一个实施例，聚类模块 1011 用于进行以下：依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别；基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，以获得所述第一候选型别组和所述第二候选型别组。

根据本发明的一个实施例，在聚类模块 1011 中进行依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别，包括：利用以下公式确定所述型别的第一分值 $TScore$ ， $TScore = N \times S$ ，其中， N 为支持所述型别的读段数目， S 为支持所述型别的候选单体型的分值的总和；将所有所述型别的第一分值作两两组合，确定所述第一分值之和最高的组合中的型别分别为所述第一候选型别和所述第二候选型别。

根据本发明的一个实施例，在聚类模块 1011 中进行基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，获得所述第一候选型别组和所述第二候选型别组，包括：对于每个所述其它型别，比较第一交集和第二交集的大小，依据比较的结果将每个所述其它型别归至所述第一候选型别或者所述第二候选型别，以获得所述第一候选型别组和所述第二候选型别组，所述第一交集为支持该其它型别的读段与所述支持第一候选型别的读段的交集，所述第二交集为支持该其它型别的读段与所述支持第二候选型别的读段的交集。

根据本发明的一个实施例，主/次要型别确定模块 1013 用于进行以下：基于支持所述型别的读段和所述型别的第一分值，分别对所述第一候选型别组和所述第二候选型别组中的型别进行第二打分，基于获得的型别的第二分值确定出所述主要型别和所述次要型别。

根据本发明的一个实施例，在主/次要型别确定模块 1013 中进行以下：利用以下公式确定所述型别的第二分值 $TScore_New$ ， $TScore_New = N^* \times TScore$ ，其中， N^* 为所述第一候选型别组中的型别的支持读段落在所述支持第二候选型别的读段之外的数目，或者为所述第二候选型别组中的型别的支持读段落在所述支持第一候选型别的读段之外的数目；确定所述第一候选型别组和所述第二候选型别组中的第二分值最高的两个型别为所述主要型别和所述次要型别。

根据本发明的一个实施例，基因型确定模块 1015 用于进行以下：若仅支持所述次要型别的读段数目与仅支持所述主要型别的读段数目的比值大于 0.1，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别，否则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别。

如图 6 所示，根据本发明的一个实施例，还包括拷贝数确定模块 1016，用于进行以下：以至少一个拷贝数固定为 2 的区域的平均测序深度为基准深度，计算所述目标区域的测序深度与所述基准深度的比例，依据计算得的比例确定所述目标区域的拷贝数；依据所述目标区域的拷贝数，判定所述目标区域的基因型。

根据本发明的一个实施例，利用所述拷贝数确定模块 1016 进行依据目标区域的拷贝数，判定所述目标区域的基因型，包括：若所述目标区域的拷贝数为 0，则判定所述目标区域不存在；若所述目标区域的拷贝数为 1，则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别；若所述目标区域的拷贝数为 2，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别。

上述任一实施例中的分型方法、装置和/或系统通过构建目标区域的参考序列和基准序列，构建型别集合，基于读段信息构建单体型，对目标区域进行基因分型。该分型方法适用于任何区域的基因分型，特别适用于高度多态性区域的分型，例如适用于对 HLA-DRB3、4 和/或 5 的分型。该方法不需要针对目标基因进行 PCR，降低了实验工作量和实验难度，提高了应用或研究时方案设计的灵活性。

下面详细描述本发明的实施例，所述实施例中的示例图，其自始至终相同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。

需要说明的是在本文中所使用的术语“第一”、“第二”等仅用于方便描述目的，而不能理解为指示或暗示相对重要性，也不能理解为之间有先后顺序关系。在本发明的描述中，除非另有说明，“多个”的含义是两个或两个以上。

除另有交待，以下实施例中涉及的试剂、仪器或者软件，都是常规市售产品或者开源的，比如从 Illumina 公司购买测序文库制备试剂盒，依照试剂盒说明书进行建库等。

实施例 1

对人类的 HLA-DRB3,4,5 进行分型，包括的一般步骤：

1. 从 IMGT/HLA 数据库下载 HLA 基因型别和序列数据。序列数据包括编码区序列和基因全长序列。

2. 修改下载的 HLA 基因型别和序列数据的格式，以便分析。将编码区序列按外显子分割，并从最相近型别的基因全长序列提取外显子两侧各 100bp（或其他长度，取决于测序读段的长度）序列，加在对应的外显子序列两侧，构成参考序列。延伸外显子序列是为了保证比对时能保留外显子边缘的读段。选取最完整的基因全长序列作为基准序列，记录所有编码区序列和基因全长序列相对于基准序列的变异，将变异与型别相关联。

3. 比对样品测序数据至参考序列，得到比对上的读段的位置信息和变异信息，并转化为相对于基准序列的位置信息和变异信息。

4. 利用比对上的读段的位置信息和变异信息，对每个外显子，利用其中具有重叠部分并且重叠部分序列完全一致的读段进行组装，获得多个单体型。过滤对外显子覆盖度较低的单体型。获得单体型的支持读段数、覆盖度，根据支持读段数和覆盖度打分，按分值从高到底取出单体型，每取出一个单体型就去除与之序列冲突且冲突位点深度较低的其他单体型，得到候选单体型。比较候选单体型变异和每个型别的变异，完全匹配的型别为此单体型支持的型别，一个单体型可能支持多个型别。

5. 提取所有外显子的候选单体型及其对应的型别和支持的读段。过滤支持的外显子数目偏少的型别。根据支持的单体型分值和支持的读段数目，对型别打分。对所有型别的分值做两两组合，选取分值之和最高的组合作为候选型别。根据两个候选型别得到对应的两组候选读段，将所有可能的型别与此两组候选读段比较，依据候选读段的支持情况，将所有型别也分为两组并按候选读段重新打分，两组中分值分别最高的两个型别作为最佳型别。由读段数目更多的那组候选读段得到的最佳型别作为主要型别，另一个最佳型别作为次要型别。若仅支持次要型别的读段数目与仅支持主要型别的读段数目的比值较大，例如大于 0.1，则判断此基因为杂合，保留主要和次要型别，否则判断为纯合，仅保留主要型别。

以下步骤 6 和 7 为可选步骤。

6. 依据测序深度判断 HLA-DRB3,4,5 的拷贝数。与一般的基因不同，HLA-DRB3,4,5 的拷贝数并不固定，对它们中的任何一个，一个人类样本有可能出现 0,1,2 三种基因拷贝数的情况。以拷贝数固定为 2 的区域（其他基因或非编码序列）的测序深度或某个较大区域的平均测序深度作为基准深度，计算 HLA-DRB3,4,5 的深度与基准深度的比例：若比例接近 0，则拷贝数为 0；若比例接近 0.5，则拷贝数为 1；若比例接近 1，则拷贝数为 2。

7. 结合保留的型别信息与拷贝数信息，得到最终的 HLA-DRB3, 4, 5 基因型别。若拷贝数为 0，则该基因不存在，亦无型别；若拷贝数为 1，则选取保留的型别中的主要型别为最终型别；若拷贝数为 2，则选取保留的所有型别为最终型别：仅有主要型别，为纯合，既有主要型别又有次要型别，为杂合。

若只进行步骤 1-5 与进行上述全部步骤发现对目标区域的分型结果不一致，可以以进行步骤 6 和 7 后的分型结果为准。

实施例 2

1. 从 IMGT/HLA 数据库下载 HLA 基因型别和序列数据。序列数据包括编码区序列和基因全长序列。

2. 修改下载的 HLA 基因型别和序列数据的格式，以便分析。将基因编码序列按外显子分割，并从最相近型别的基因全长序列提取外显子两侧各 100bp（或其他长度，取决于测序读段的长度）序列，加在对应的外显子序列两侧，构成参考序列。延伸外显子序列是为了保证比对时能保留外显子边缘的读段。选取最完整的基因全长序列作为基准序列，记录所有编码区序列和基因全长序列相对于基准序列的变异，将变异与型别相关联。

3. 使用一个 YH cell line 样品(炎黄计划中的细胞样品)的 MHC 区域捕获高通量测序数据。

4. 比对测序数据至参考序列，得到比对上的读段的位置信息和变异信息，并转化为相对于基准序列的位置信息和变异信息。

5. 利用比对上的读段的位置信息和变异信息，对每个外显子，利用其中具有重叠部分并且重叠部分序列完全一致的读段进行组装，获得多个单体型。过滤对外显子覆盖度低于 95% 的单体型。获得单体型的支持读段数、深度和覆盖度，计算单体型的分值 Score。Score 的计算公式：

$$Score = \frac{c^2 \times N}{\log R}$$

c: 单体型对外显子的覆盖度；

N: 支持此单体型的支持读段数；

R: 单体型的可靠性，其计算公式如下：

$$R = \frac{\sum [(X_i - X)^2 \div X]}{\log(L-1)}$$

X_i: 单体型每个位置的深度；

X: 单体型的平均深度；

L: 单体型的序列长度;

按 Score 从高到底取出单体型, 每取出一个单体型就去掉与之序列冲突且冲突位点深度低于取出的单体型的 20% 的其他单体型。最多取出 4 个单体型, 作为候选单体型。比较候选单体型的变异和每个型别的变异, 完全匹配的型别为此单体型支持的型别, 一个单体型可能支持多个型别。

6. 合并支持相同型别的单体型的相关信息。过滤支持的外显子数目小于外显子总数 30% 的型别。计算型别的分值 TScore。TScore 的计算公式:

$$TScore = N \times S$$

N: 支持此型别的读段数;

S: 支持此型别的单体型的分值之和;

对所有型别做两两组合, 选取分值之和最高的组合作为候选型别。根据候选型别中分值较高的型别得到对应的第一组候选读段, 分值较低的型别得到第二组候选读段。将所有型别与此两组候选读段比较, 若支持某型别的读段落在第一组候选读段中的数目更多, 则将此型别归为第一组, 否则归为第二组, 这样将所有型别基于候选也分为了两组。计算型别的分值 TScore_New。TScore_New 的计算公式:

$$TScore_New = N^* \times TScore$$

N^* : 支持此型别的读段落在另一组候选读段之外的数目,

两组中 TScore_New 分别最高的两个型别作为最佳型别。第一组型别中的最佳型别作为主要型别, 第二组中的最佳型别作为次要型别。若仅支持次要型别的读段数目与仅支持主要型别的读段数目的比值大于 0.1, 判断此基因为杂合, 保留主要和次要型别, 否则判断为纯合, 仅保留主要型别。

7. 以拷贝数固定为 2 的区域的平均测序深度作为基准深度, 计算 HLA-DRB3,4,5 的深度与基准深度的比例, 若比例接近 0, 则拷贝数为 0, 若比例接近 0.5, 则拷贝数为 1, 若比例接近 1, 则拷贝数为 2。此样品 HLA-DRB3 拷贝数为 1, HLA-DRB4 拷贝数为 0, HLA-DRB5 拷贝数为 1。

8. 结合保留的型别和拷贝数, 得到 YH cell line 样品 HLA-DRB3,4,5 分型的最终结果如下表 1 所示, 其中“Blank”表示等位基因缺失, 拷贝数为 0, 则表示为 Blank/Blank, 拷贝数为 1, 则表示为 DRB/Blank, 拷贝数为 2, 则表示为 DRB/DRB。

表 1

HLA-DRB3	HLA-DRB4	HLA-DRB5
DRB3*02:02:01/Blank	Blank/Blank	DRB5*01:01:01/Blank

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、“实施方式”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。

尽管已经示出和描述了本发明的实施例，本领域的普通技术人员可以理解：在不脱离本发明的原理和宗旨的情况下可以对这些实施例进行多种变化、修改、替换和变型，本发明的范围由权利要求及其等同物限定。

权利要求书

1. 一种分型方法，其特征在于，包括：

获取待测样品的测序数据，所述测序数据包括多个来自目标区域的读段；

将参考序列组比对至基准序列组，获得型别集合，所述型别集合包含多个型别，

所述基准序列组包含一条或多条基准序列，所述基准序列组能够完全覆盖所述目标区域的所有基因序列，不同所述基准序列包含不同的基因全序列，

所述参考序列组包括多条参考序列，所述参考序列组能够完全覆盖所述目标区域的编码区序列的所有外显子，不同所述参考序列包含不同的外显子；

将所述测序数据比对至所述参考序列组，获得比对结果；

将所述比对结果转化为相对于所述基准序列组的比对结果，获得转化的比对结果；

基于所述转化的比对结果，分别对比对上相同基准序列的读段进行组装，获得组装结果，所述组装结果包括多个单体型；

比较所述单体型上的变异和所述型别集合中的型别，以确定所述单体型支持的型别；

依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组；

基于支持所述型别的单体型和支持所述型别的读段，分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选，以确定主要型别和次要型别；

基于支持所述主要型别和支持所述次要型别的读段的数目的差异，判定所述目标区域的基因型。

2. 权利要求 1 的方法，其特征在于，所述目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5 中的至少之一。

3. 权利要求 1 的方法，其特征在于，所述目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5。

4. 权利要求 1 的方法，其特征在于，所述参考序列组通过以下步骤构建获得：

获得包含所述目标区域的编码区序列和基因全序列；

将所述编码区序列按外显子分割，获得多个外显子序列；

从与所述外显子序列最接近型别的基因全序列提取该外显子序列两侧各 K bp 的序列，加到对应的外显子序列的两侧，获得所述参考序列组中的参考序列，其中 K 依据读段的长度来确定。

5. 权利要求 1 的方法，其特征在于，一条所述基准序列为所述目标区域中的一个基因的基因全序列。

6. 权利要求 1 的方法，其特征在于，所述基于转化的比对结果，分别对比对上相同基准序列的读段进行组装，获得组装结果，所述组装结果包括多个单体型，包括：

对比对上相同基准序列的读段，利用具有重叠部分并且所述重叠部分完全一致的读段进行所述组装，以获得所述多个单体型。

7. 权利要求 1 的方法，其特征在于，在获得所述组装结果之后，过滤掉对所述外显子的覆盖度低于 95% 的单体型。

8. 权利要求 1 的方法，其特征在于，在确定单体型支持的型别之前进行以下：

分别基于比对上相同基准序列的读段的组装结果，对每个所述单体型进行打分，基于获得的单体型的分值对所述单体型进行筛选，获得候选单体型；

以所述候选单体型替代所述单体型。

9. 权利要求 8 的方法，其特征在于，利用以下公式对每个所述单体型进行打分，以确定所述单体型的分值 Score：

$$Score = \frac{c^2 \times N}{\log R}, \text{ 其中,}$$

c 为所述单体型对所述外显子的覆盖度，

N 为支持所述单型型的读段数目，

$$R = \frac{\sum [(X_i - X)^2 \div X]}{\log(L-1)},$$

R 表示所述单型型的可靠性，

X_i 为所述单型型的位置 i 的测序深度，i 为所述单型型上的位置的编号，

X 为所述单型型的平均深度，

L 为所述单型型的长度。

10. 权利要求 8 的方法，其特征在于，基于获得的单体型的分值对所述单体型进行筛选，包括：

对于比对上相同基准序列的读段的组装结果，取出其中的最高的单体型的分值所对应的单体型后，去除满足以下条件的所述组装结果中的单体型：与取出的所述单型型的序列不一致并且不一致位点的测序深度低于所述单型型相应位点的测序深度的 20%；

重复以上至最多取出 4 个所述单型型，获得所述候选单型型。

11. 权利要求 8 的方法，其特征在于，所述比较每个单型型上的变异和所述型别集合

中的型别，以确定所述单体型支持的型别，包括：

比较每个候选单体型上的变异和所述型别集合中的型别，若完全匹配，则确定所述候选单体型支持所述型别。

12. 权利要求 11 的方法，其特征在于，确定所述候选单体型支持的型别之后，合并支持相同型别的所述候选单体型。

13. 权利要求 8 的方法，其特征在于，在进行依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组之前，

过滤掉支持的外显子数目小于外显子总数 30% 的型别。

14. 权利要求 8 的方法，其特征在于，所述依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组，包括：

依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别；

基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，以获得所述第一候选型别组和所述第二候选型别组。

15. 权利要求 14 的方法，其特征在于，所述依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别，包括：

利用以下公式确定所述型别的第一分值 $TScore$ ，

$$TScore = N \times S, \text{ 其中,}$$

N 为支持所述型别的读段数目，

S 为支持所述型别的候选单体型的分值的总和；

将所有所述型别的第一分值作两两组合，确定所述第一分值之和最高的组合中的型别分别为所述第一候选型别和所述第二候选型别。

16. 权利要求 14 的方法，其特征在于，所述基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，获得所述第一候选型别组和所述第二候选型别组，包括：

对于每个所述其它型别，比较第一交集和第二交集的大小，依据比较的结果将每个所述其它型别归至所述第一候选型别或者所述第二候选型别，以获得所述第一候选型别组

和所述第二候选型别组，

所述第一交集为支持该其它型别的读段与所述支持第一候选型别的读段的交集，

所述第二交集为支持该其它型别的读段与所述支持第二候选型别的读段的交集。

17. 权利要求 14 的方法，其特征在于，所述基于支持所述型别的单体型和支持所述型别的读段，分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选，以确定主要型别和次要型别，包括：

基于支持所述型别的读段和所述型别的第一分值，分别对所述第一候选型别组和所述第二候选型别组中的型别进行第二打分，基于获得的型别的第二分值确定出所述主要型别和所述次要型别。

18. 权利要求 17 的方法，其特征在于，包括：

利用以下公式确定所述型别的第二分值 $TScore_New$ ，

$TScore_New = N^* \times TScore$ ，其中，

N^* 为所述第一候选型别组中的型别的支持读段落在所述支持第二候选型别的读段之外的数目，或者为所述第二候选型别组中的型别的支持读段落在所述支持第一候选型别的读段之外的数目；

确定所述第一候选型别组和所述第二候选型别组中的第二分值最高的两个型别为所述主要型别和所述次要型别。

19. 权利要求 1 的方法，其特征在于，所述基于支持所述主要型别和支持所述次要型别的读段的数目的差异，判定所述目标区域的基因型，包括：

若仅支持所述次要型别的读段数目与仅支持所述主要型别的读段数目的比值大于 0.1，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别，

否则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别。

20. 权利要求 1-19 任一方法，其特征在于，还包括：

以至少一个拷贝数固定为 2 的区域的平均测序深度为基准深度，计算所述目标区域的测序深度与所述基准深度的比例，依据计算得的比例确定所述目标区域的拷贝数；

依据所述目标区域的拷贝数，判定所述目标区域的基因型。

21. 权利要求 20 的方法，其特征在于，所述依据目标区域的拷贝数，判定所述目标区域的基因型，包括：

若所述目标区域的拷贝数为 0，则判定所述目标区域不存在；

若所述目标区域的拷贝数为 1，则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别；

若所述目标区域的拷贝数为 2，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别。

22. 一种计算机可读介质，其特征在于，用于存储计算机可执行程序，执行所述程序包括完成权利要求 1-21 任一方法。

23. 一种分型装置，其特征在于，包括：

数据输入单元，用于输入数据；

数据输出单元，用于输出数据；

存储单元，用于存储数据，其中包括计算机可执行程序；

处理器，与所述数据输入单元、所述数据输出单元和所述存储单元连接，用于执行所述计算机可执行程序，执行所述程序包括完成权利要求 1-21 任一方法。

24. 一种分型系统，其特征在于，包括：

数据输入模块，用于输入待测样品的测序数据，所述测序数据包括多个来自目标区域的读段；

比对模块，包括第一比对模块和第二比对模块，

所述第一比对模块用于将参考序列组比对至基准序列组，获得型别集合，所述型别集合包含多个型别，

第二比对模块，用于将来自数据输入模块的测序数据比对至所述参考序列组，获得比对结果，

所述基准序列组包含一条或多条基准序列，所述基准序列组能够完全覆盖所述目标区域的所有基因序列，不同所述基准序列包含不同的基因全序列，

所述参考序列组包括多条参考序列，所述参考序列组能够完全覆盖所述目标区域的编码区序列的所有外显子，不同所述参考序列包含不同的外显子；

转化模块，用于将所述比对结果转化为相对于所述基准序列组的比对结果，获得转化的比对结果；

组装模块，用于基于所述转化的比对结果，分别对比对上相同基准序列的读段进行组装，获得组装结果，所述组装结果包括多个单体型；

单体型支持型别确定模块，用于比较所述单体型上的变异和所述型别集合中的型别，以确定所述单体型支持的型别；

聚类模块，用于依据支持所述型别的单体型和支持所述型别的读段，将所述型别分成两组，获得第一候选型别组和第二候选型别组；

主/次要型别确定模块，用于基于支持所述型别的单体型和支持所述型别的读段，分别对所述第一候选型别组和所述第二候选型别组中的型别进行筛选，以确定出主要型别和次要型别；

基因型确定模块，用于基于支持所述主要型别和支持所述次要型别的读段的数目的差异，判定所述目标区域的基因型。

25. 权利要求 24 的系统，其特征在于，所述目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5 中的至少之一。

26. 权利要求 24 的系统，其特征在于，所述目标区域包括 HLA-DRB3、HLA-DRB4 和 HLA-DRB5。

27. 权利要求 24 的系统，其特征在于，比对模块中的参考序列组通过以下步骤构建获得：

获得包含所述目标区域的编码区序列和基因全序列；

将所述编码区序列按外显子分割，获得多个外显子序列；

从与所述外显子序列最接近型别的基因全序列提取该外显子序列两侧各 K bp 的序列，加到对应的外显子序列的两侧，获得所述参考序列组中的参考序列，其中 K 依据读段的长度来确定。

28. 权利要求 24 的系统，其特征在于，比对模块和/或转化模块中的一条所述基准序列为所述目标区域中的一个基因的基因全序列。

29. 权利要求 24 的系统，其特征在于，利用所述组装模块进行以下：

对比对上相同基准序列的读段，利用具有重叠部分并且所述重叠部分完全一致的读段进行所述组装，以获得所述多个单体型。

30. 权利要求 24 的系统，其特征在于，在所述组装模块中，在获得所述组装结果之后，过滤掉对所述外显子的覆盖度低于 95% 的单体型。

31. 权利要求 24 的系统，其特征在于，还包括候选单体型确定模块，用于在利用所述单体型支持型别确定模块确定单体型支持的型别之前，进行以下：

分别基于比对上相同基准序列的读段的组装结果，对每个所述单体型进行打分，基于获得的单体型分值对所述单体型进行筛选，获得候选单体型；

以所述候选单体型替代所述单体型。

32. 权利要求 31 的系统，其特征在于，在所述候选单体型确定模块中，利用以下公式对每个所述单体型进行打分，以确定所述单体型的分值 Score：

$$Score = \frac{c^2 \times N}{\log R}, \text{ 其中,}$$

c 为所述单体型对所述外显子的覆盖度，

N 为支持所述单体型的读段数目，

$$R = \frac{\sum [(X_i - X)^2 \div X]}{\log(L-1)},$$

R 表示所述单体型的可靠性，

X_i 为所述单体型的位置 i 的测序深度，i 为所述单体型上的位置的编号，

X 为所述单体型的平均深度，

L 为所述单体型的长度。

33. 权利要求 31 的系统，其特征在于，在所述候选单体型确定模块中进行基于获得的单体型分值对所述单体型进行筛选，包括：

对于比对上相同基准序列的读段的组装结果，取出其中的最高的单体型分值所对应的单体型后，去除满足以下条件的所述组装结果中的单体型：与取出的所述单体型的序列不一致并且不一致位点的测序深度低于所述单体型相应位点的测序深度的 20%；

重复以上至最多取出 4 个所述单体型，获得所述候选单体型。

34. 权利要求 31 的系统，其特征在于，在所述单体型支持型别确定模块中进行以下：

比较每个候选单体型上的变异和所述型别集合中的型别，若完全匹配，则确定所述候选单体型支持所述型别。

35. 权利要求 34 的系统，其特征在于，在所述单体型支持型别确定模块中，确定所述候选单体型支持的型别之后，合并支持相同型别的所述候选单体型。

36. 权利要求 31 的系统，其特征在于，还包括型别过滤模块，用于在利用所述聚类模块获得第一候选型别组和第二候选型别组之前，

过滤掉支持的外显子数目小于外显子总数 30% 的型别。

37. 权利要求 31 的系统，其特征在于，所述聚类模块用于进行以下：

依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别；

基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述

第二候选型别，以获得所述第一候选型别组和所述第二候选型别组。

38. 权利要求 37 的系统，其特征在于，在所述聚类模块中进行依据支持所述型别的单体型和支持所述型别的读段，对所述型别进行第一打分，基于获得的型别的第一分值对所述型别进行筛选，获得第一候选型别和第二候选型别，包括：

利用以下公式确定所述型别的第一分值 $TScore$ ，

$$TScore = N \times S, \text{ 其中,}$$

N 为支持所述型别的读段数目，

S 为支持所述型别的候选单体型的分值的总和；

将所有所述型别的第一分值作两两组合，确定所述第一分值之和最高的组合中的型别分别为所述第一候选型别和所述第二候选型别。

39. 权利要求 37 的系统，其特征在于，在所述聚类模块中进行基于支持所述第一候选型别的读段和支持所述第二候选型别的读段对所述型别集合中的其它型别的支持情况，将所述型别集合中的其它型别分别归至所述第一候选型别和所述第二候选型别，获得所述第一候选型别组和所述第二候选型别组，包括：

对于每个所述其它型别，比较第一交集和第二交集的大小，依据比较的结果将每个所述其它型别归至所述第一候选型别或者所述第二候选型别，以获得所述第一候选型别组和所述第二候选型别组，

所述第一交集为支持该其它型别的读段与所述支持第一候选型别的读段的交集，

所述第二交集为支持该其它型别的读段与所述支持第二候选型别的读段的交集。

40. 权利要求 37 的系统，其特征在于，所述主/次要型别确定模块用于进行以下：

基于支持所述型别的读段和所述型别的第一分值，分别对所述第一候选型别组和所述第二候选型别组中的型别进行第二打分，基于获得的型别的第二分值确定出所述主要型别和所述次要型别。

41. 权利要求 40 的系统，其特征在于，在所述主/次要型别确定模块中进行以下：

利用以下公式确定所述型别的第二分值 $TScore_New$ ，

$$TScore_New = N^* \times TScore, \text{ 其中,}$$

N^* 为所述第一候选型别组中的型别的支持读段落在所述支持第二候选型别的读段之外的数目，或者为所述第二候选型别组中的型别的支持读段落在所述支持第一候选型别的读段之外的数目；

确定所述第一候选型别组和所述第二候选型别组中的第二分值最高的两个型别为所述

主要型别和所述次要型别。

42. 权利要求 24 的系统，其特征在于，所述基因型确定模块用于进行以下：

若仅支持所述次要型别的读段数目与仅支持所述主要型别的读段数目的比值大于 0.1，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别，

否则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别。

43. 权利要求 24-42 任一系统，其特征在于，还包括拷贝数确定模块，用于进行以下：

以至少一个拷贝数固定为 2 的区域的平均测序深度为基准深度，计算所述目标区域的测序深度与所述基准深度的比例，依据计算得的比例确定所述目标区域的拷贝数；

依据所述目标区域的拷贝数，判定所述目标区域的基因型。

44. 权利要求 43 的系统，其特征在于，利用所述拷贝数确定模块进行依据目标区域的拷贝数，判定所述目标区域的基因型，包括：

若所述目标区域的拷贝数为 0，则判定所述目标区域不存在；

若所述目标区域的拷贝数为 1，则判定所述目标区域为纯合子，组成所述纯合子的两个等位基因均为所述主要型别；

若所述目标区域的拷贝数为 2，则判定所述目标区域为杂合子，组成所述杂合子的主要等位基因和次要等位基因分别为所述主要型别和所述次要型别。

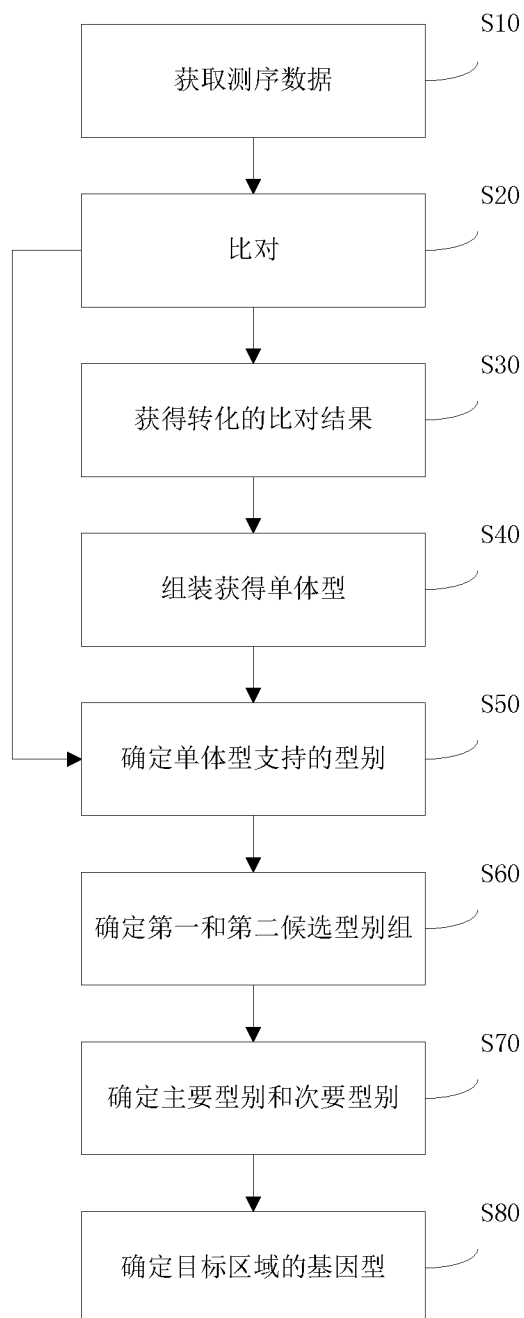


图 1

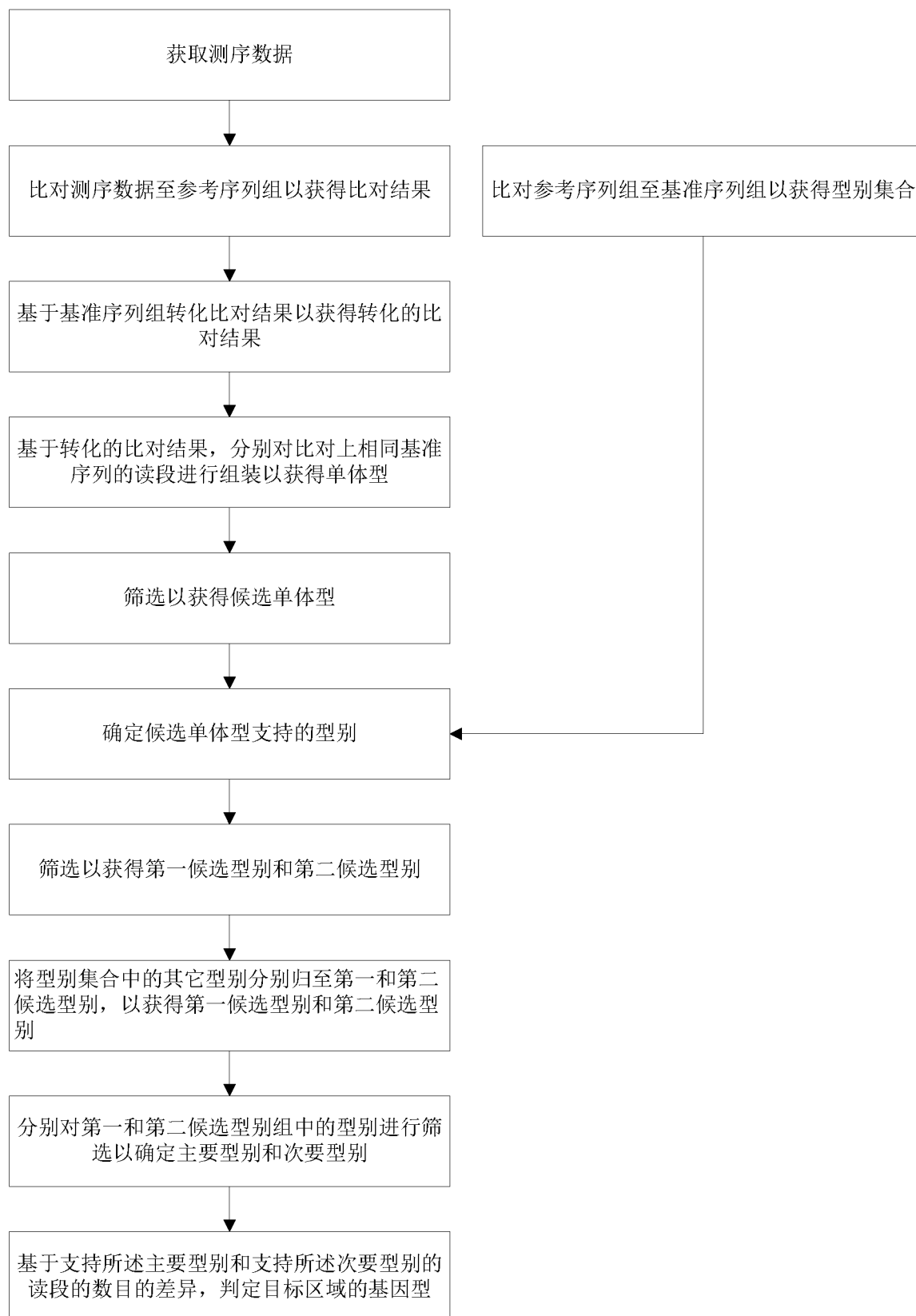


图 2

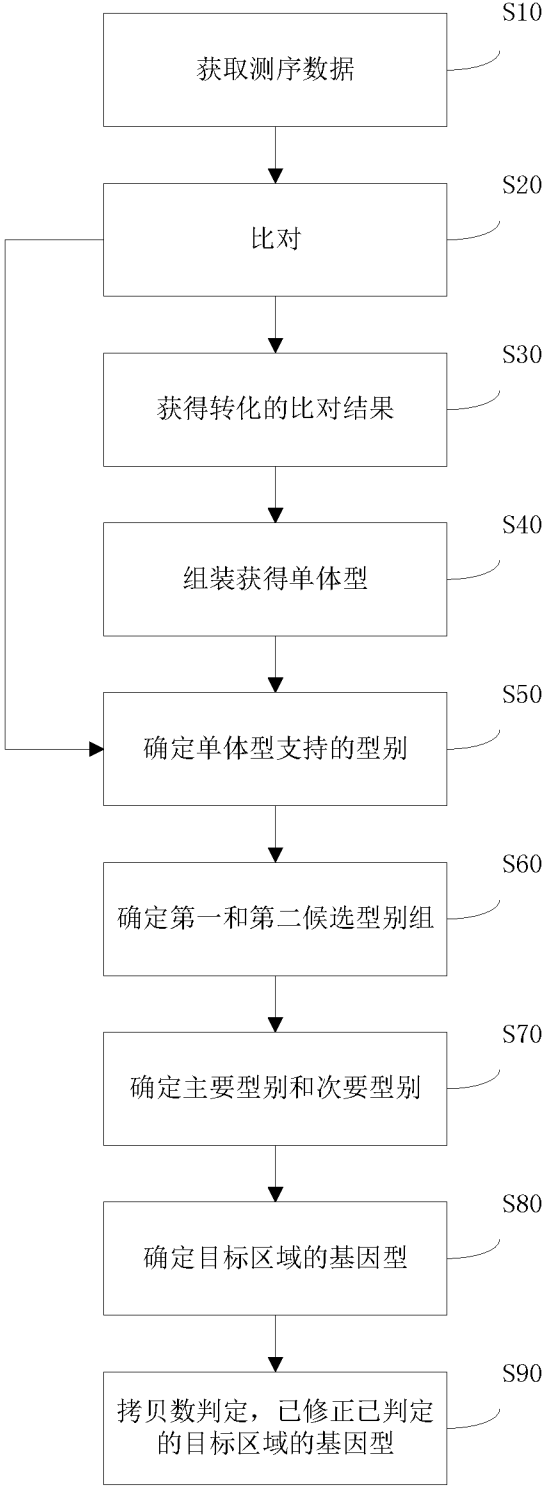


图 3

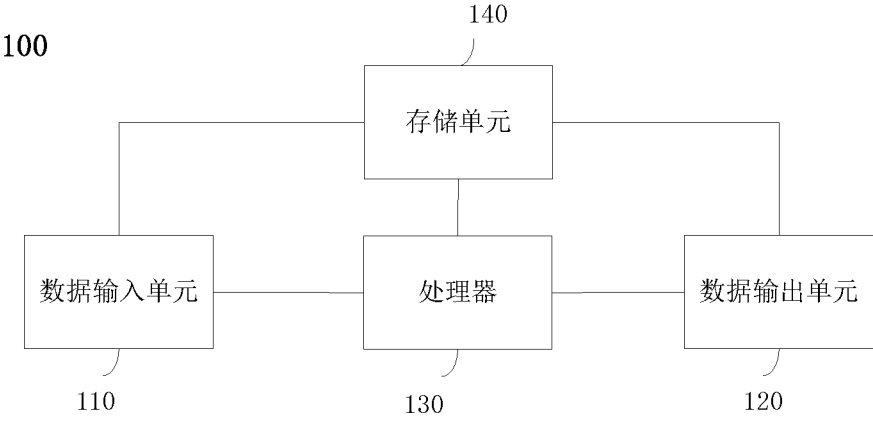


图 4

1000

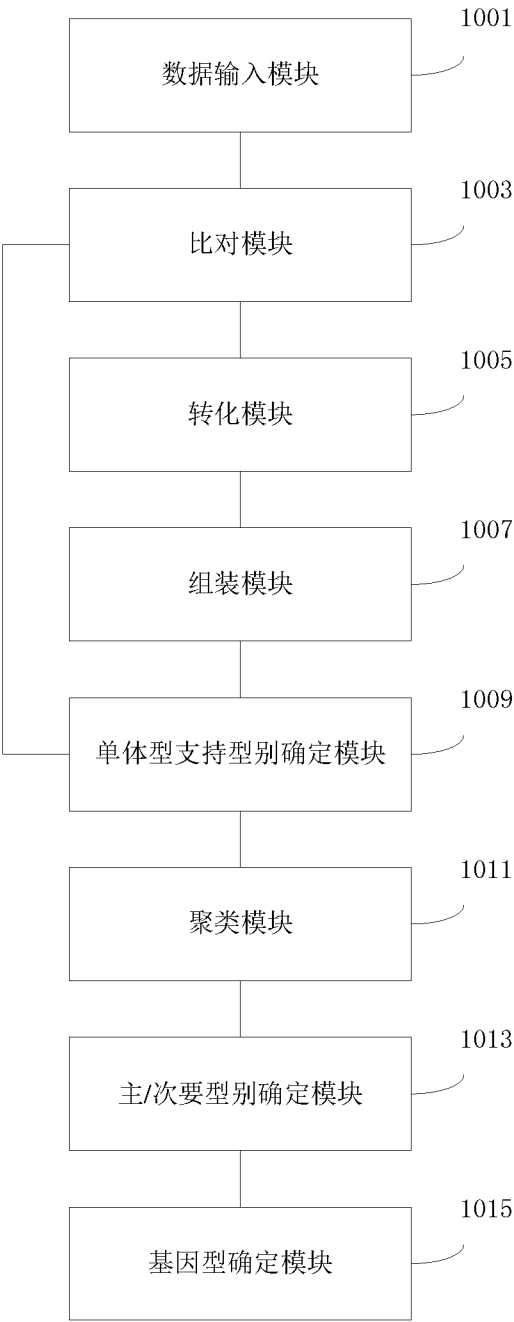


图 5

1000

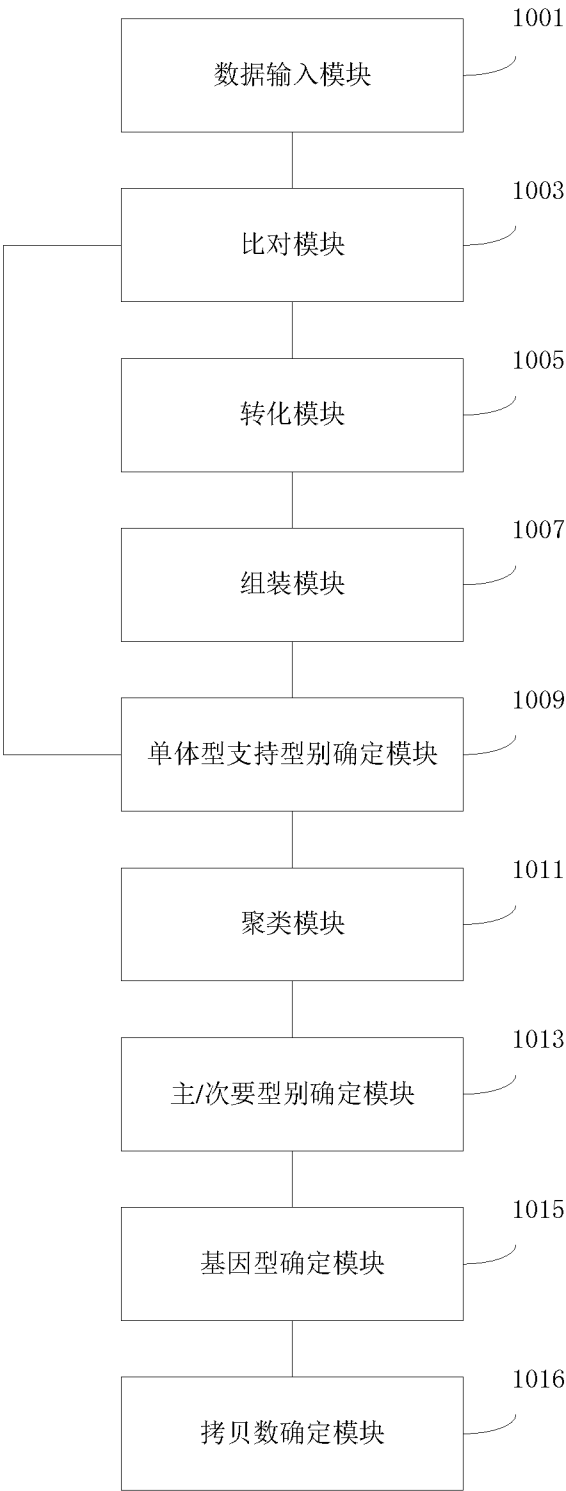


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/074027

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)

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)

CNMED; CPRSABS; CNABS; DWPI; SIPOABS; CNTXT; WOTXT; EPTXT; JPTXT; CNKI; Pubmed; Google scholar: second generation, next generation, sequence, human leucocyte antigen, typing, database, reference sequence, comparison, NGS, next generation sequencing, high-throughput, deep sequencing, HLA, MHC, DRB3, DRB4, DRB5, genotyping, classif+, allele, IMGT/HLA, reference, base, alignment, haplotype

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103074444 A (SUZHOU JINGYIN BIOLOGICAL TECHNOLOGY CO, LTD.), 01 May 2013 (01.05.2013), see the whole document	1-44
A	WANG, C. et al. "High-throughput, high-fidelity HLA genotyping with deep sequencing" PNAS, vol. 109, no. 22, 29 May 2012 (29.05.2012), pages 8676-8681, see the whole document	1-44

☐ 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 17 November 2016 (17.11.2016)	Date of mailing of the international search report 25 November 2016 (25.11.2016)
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 LI, Kangqi Telephone No.: (86-10) 62411034

International application No.
PCT/CN2016/074027

Form PCT/ISA/210 (patent family annex) (July 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) 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)) CNMED;CPRSABS;CNABS;DWPI;SIPOABS;CNTXT;WOTXT;EPTXT;JPTXT;CNKI;Pubmed;Google scholar;第二代, 高通量, 下一代, 测序, 人白细胞抗原, 组织相容性复合体, 分型, 基因型, 数据库, 参考序列, 基准序列, 比对, 等位基因, 单体型, NGS, next generation sequencing, high-throughput, deep sequencing, HLA, MHC, DRB3, DRB4, DRB5, genotyping, classif+, allele, IMGT/HLA, reference, base, alignment, haplotype											
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 103074444 A (苏州晶因生物科技有限公司) 2013年 5月 1日 (2013 - 05 - 01) 参见全文</td> <td>1-44</td> </tr> <tr> <td>A</td> <td>WANG, C. 等. "High-throughput, high-fidelity HLA genotyping with deep sequencing" PNAS, 第109卷, 第22期, 2012年 5月 29日 (2012 - 05 - 29), 第8676-8681页 参见全文</td> <td>1-44</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 103074444 A (苏州晶因生物科技有限公司) 2013年 5月 1日 (2013 - 05 - 01) 参见全文	1-44	A	WANG, C. 等. "High-throughput, high-fidelity HLA genotyping with deep sequencing" PNAS, 第109卷, 第22期, 2012年 5月 29日 (2012 - 05 - 29), 第8676-8681页 参见全文	1-44
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求									
A	CN 103074444 A (苏州晶因生物科技有限公司) 2013年 5月 1日 (2013 - 05 - 01) 参见全文	1-44									
A	WANG, C. 等. "High-throughput, high-fidelity HLA genotyping with deep sequencing" PNAS, 第109卷, 第22期, 2012年 5月 29日 (2012 - 05 - 29), 第8676-8681页 参见全文	1-44									
☐ 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 2016年 11月 17日		国际检索报告邮寄日期 2016年 11月 25日									
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 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 李康琦 电话号码 (86-10)62411034									

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/074027

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 103074444 A	2013年 5月 1日	无	