

(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6月7日 (07.06.2012)



(10) 国际公布号
WO 2012/071685 A1

(51) 国际专利分类号:
C12Q 1/70 (2006.01)

(21) 国际申请号: PCT/CN2010/001943

(22) 国际申请日: 2010年12月2日 (02.12.2010)

(25) 申请语言: 中文

(26) 公布语言: 中文

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

(72) 发明人: 及

(75) 发明人/申请人 (仅对美国): **刘智盛 (LIU, Zhisheng)** [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)。 **田仕林 (TIAN, Shilin)** [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)。 **潘荣 (PAN, Rong)** [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)。

[见续页]

(54) Title: METHOD AND SYSTEM FOR BIOINFORMATICS ANALYSIS OF HPV PRECISE TYPING

(54) 发明名称: HPV 精确分型的生物信息学分析的方法及系统

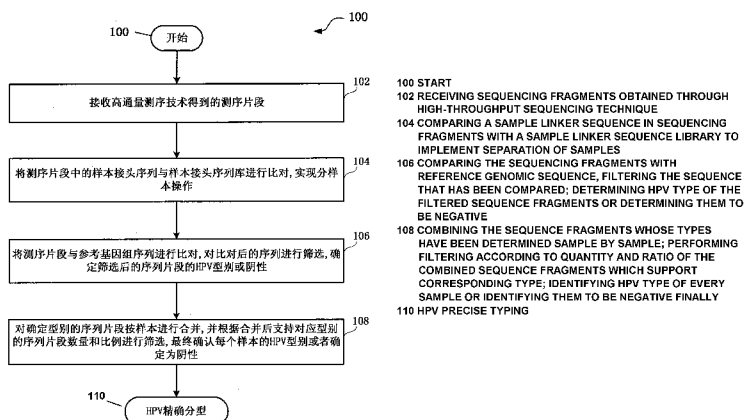


FIG.1

(57) Abstract: A method and system for bioinformatics analysis of HPV precise typing is disclosed in the invention, said method comprises: receiving sequencing fragments obtained through high-throughput sequencing technique; comparing a sample linker sequence in sequencing fragments with a sample linker sequence library to implement separation of samples; comparing the sequencing fragments with reference genomic sequence, filtering the sequence that has been compared; determining HPV type of the filtered sequence fragments or determining them to be negative; combining the sequence fragments whose types have been determined sample by sample; performing filtering according to quantity and ratio of the combined sequence fragments which support corresponding type; identifying HPV type of every sample or identifying them to be negative finally. The method and system for bioinformatics analysis of HPV precise typing provided in the invention utilizes bioinformatics analysis method and technical solution to implement rapid detection of large samples and rapid detection of infected HPV type with relative high sensitivity and specificity.

[见续页]



本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(57) 摘要:

本发明公开一种 HPV 精确分型的生物信息学分析的方法及系统, 该方法包括: 接收高通量测序技术得到的测序片段; 将测序片段中的样本接头序列与样本接头序列库进行比对, 实现分样本操作; 将测序片段与参考基因组序列进行比对, 对比对后的序列进行筛选, 确定筛选后的序列片段的 HPV 型别或阴性; 对确定型别的序列片段按样本进行合并, 并根据合并后支持对应型别的序列片段数量和比例进行筛选; 最终确认每个样本的 HPV 型别或者确定为阴性。本发明提供的 HPV 精确分型的生物信息学分析的方法及系统, 利用生物信息学的分析方法及技术手段, 实现了快速检测大量样本、快速完成对感染 HPV 型别的检测, 具有较高的灵敏度和特异性。

HPV 精确分型的生物信息学分析的方法及系统

技术领域

本发明涉及生物基因工程技术领域，尤其涉及一种 HPV 精确分型的生物信息学分析的方法及系统。

背景技术

人类乳突病毒 (Human Papillomavirus, HPV) 是一种嗜上皮性病毒，根据其致病强弱能力被分为高危型 (例如 HPV16, 18, 31, 33 和 45) 和低危型 (例如 HPV6, 11, 42, 43 和 44) 两大类。在自然人群中，感染率从低于 1% 到高达 50%。超过 100 种的 HPV 能够感染皮肤 (皮肤类型) 或呼吸道和肛门生殖道的粘膜 (粘膜类型)，超过 40 种的 HPV 能够感染子宫颈。HPV 在许多肿瘤的启动、发生、发展甚至恶性的过程中都起到了重要作用，从而被认为是与人类肿瘤关系最为密切的肿瘤病毒。

对 HPV 病毒感染进行准确的检测可提高 HPV 相关肿瘤，特别是妇女宫颈癌的病变筛查敏感性，改善其防治手段。分型与临床的结合研究已经证实不同的 HPV 亚型在致癌性方面存在比较大的差别。HPV 分型检测对于 HPV 治疗方案的制定、感染 HPV 的危险程度、HPV 感染的区域特异性等均具有重要的意义。因此目前的研究认为有必要对各个样本中存在的 HPV 进行分型检测，将有助于更详细的分析各种 HPV 型别的致病性，以达到最佳的临床预防、治疗的效果。

目前，现有技术中用于 HPV 基因分型的检测方法主要包括以下几种：

1、ELISA 法：是采用抗原与抗体的特异反应将待测物与酶连接，通过酶与底物产生颜色反应，用于定量测定。这种方法只

能用于个别亚型的鉴定，现已逐渐被其它检测方法取代。

2、PCR（聚合酶链反应，Polymerase Chain Reaction）检测法：是将提取的 DNA 进行扩增，实现 HPV 感染检测。目前常用通用引物 PCR 和实时荧光定量 PCR。该方法具有假阳性高，手段繁复、费时，不能准确诊断多重感染的缺点。

3、杂交捕获检测法：是利用分子杂交化学发光来放大信号，通过判读光的强弱来确定 HPV 型。该方法具有无法检测 HPV 特定型别和多重感染且费用高等缺点。

4、PCR 结合杂交检测法：是 PCR 与杂交共享的方法。该方法同样具有费时、手段繁复等缺点。

5、基因芯片技术：基因芯片技术，有多种分类，常用的是寡核苷酸原位合成法。该方法具有检测结果不准确，实验条件要求高，费用高等缺点。

综上所述，提供一种具有较高灵敏度和特异度，且准确性高的 HPV 分型检测技术成为本领域亟待解决的技术问题。

发明内容

本发明要解决的一个技术问题是提供一种 HPV 精确分型的生物信息学分析的方法及系统，能够实现高灵敏度和特异性、快速的识别基因序列的 HPV 型。

本发明的一个方面提供了一种 HPV 精确分型的生物信息学分析的方法，该方法包括：接收高通量测序技术得到的测序片段；将测序片段中的样本接头序列与样本接头序列库进行比对，实现分样本操作；将测序片段与参考基因组序列进行比对，对比对后的序列进行筛选，确定筛选后的序列片段的 HPV 型别或阴性；对确定型别的序列片段按样本进行合并，并根据合并后支持对应型别的序列片段数量和比例进行筛选；最终确认每个样本的 HPV 型

别或者确定为阴性。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，该方法还包括：接收到测序序列后，对测序序列进行过滤，去除不合格的序列。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，步骤“对测序序列进行过滤，去除不合格的序列”进一步包括：预先设置不合格碱基的测序质量阈值和比例阈值；当测序序列中碱基的测序质量值低于测序质量阈值，且低于测序质量阈值的碱基个数占整条序列碱基个数的比例超过比例阈值时；则认为测序序列是不合格序列并加以过滤；当测序序列的测序结果中不确定的碱基的个数超过整条序列碱基个数的 10%，则认为测序序列是不合格序列并加以过滤；与测序接头序列库进行比对时，如果测序序列中存在测序接头序列，则测序序列是不合格序列并加以过滤。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，该方法还包括：实现分样本操作后，将样本接头序列从序列片段中去除。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，步骤“将样本接头序列从序列片段中去除”进一步包括：预先设置样本接头序列的测序质量阈值和碱基数阈值；将接头序列中碱基的测序质量值低于测序质量阈值，且碱基的数量超过碱基数阈值的序列去除。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，该方法还包括：步骤 a、将样本接头序列与样本接头序列库中序列进行完全匹配操作；步骤 b、将样本接头序列降解 1-2bp 碱基，与样本接头序列库中序列对应部分进行完全匹配操作；步骤 c、允许样本接头序列仅有一个碱基的插入，即在样本接

头序列起始端进行完全匹配操作，当出现一个碱基无法匹配时将该碱基视为插入碱基，跳过此碱基后继续执行完全匹配操作；步骤 d、允许样本接头序列仅有一个碱基的缺失，即在样本接头序列中模拟缺失任何一个碱基后，进行完全匹配操作；完成步骤 a - d 后，按照优先级顺序：步骤 a > 步骤 b > “步骤 c 或步骤 d” 的顺序确定最终的样本接头序列的比对结果；比对到同一样本接头序列的被认为是来自同一样本的序列，从而区分样本；以及去除样本的序列中的样本接头序列。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，该方法还包括：如果步骤 a - d 四步操作中均无比对结果，或者一个步骤同时比对到两个结果，或者仅有步骤 c 和步骤 d 同时比对出结果；则认为该比对结果是由于无法区分而判定为无效信息，并将相应的整条序列去除。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，步骤“对比对后的序列进行筛选”进一步包括：将高通量测序技术得到的测序片段比对到参考基因组序列上；比对后，筛选并去除比对结果中比对长度低于 70%，或者一致性低于 85% 的序列；保留每条序列比对结果中的最佳结果；保留次优结果；其中，次优结果满足：序列的一致性 * 比对长度、比对得分分别高于或等于最佳结果的 0.9 倍、0.85 倍，且序列与参考序列匹配不相关的概率低于所述最佳结果的 10^3 倍；判断每条序列的最佳结果和次优结果是否比对到同一型别或其亚型，如果是，则保留比对结果仅比对到某一型别的序列作为有效序列，确定有效序列比对的 HPV 型别或阴性。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，该方法还包括：对确定型别的序列片段按样本进行合并后，对样本合并后的序列片段数量进行标准化。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，对样本合并后的序列片段数量进行标准化进一步包括：将各个文库中每个样品所拥有的序列数量，都按照比例缩放到文库的测序量为理想情况下的平均测序量。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，步骤“根据合并后支持对应型别的序列片段数量和比例进行筛选”进一步包括：标准化后，按下述条件的先后顺序进行筛选：如果可用序列数小于阴性对照样本的有效序列片段的平均个数加上其四倍标准差的和，认为实验或者测序操作失败；否则，如果比对结果支持 HPV 型别的序列片段数小于预定阈值，就认为是阴性；如果比对结果支持 HPV 型别的序列片段数占总序列片段数的比例达到预定阈值以上，则认为感染了该型别。

本发明的另一个方面提供了一种 HPV 精确分型的生物信息学分析的系统，该系统包括：接收模块，用于接收高通量测序技术得到的测序片段；分样本模块，用于将测序片段中的样本接头序列与样本接头序列库进行比对，实现分样本操作；序列型别确定模块，用于将测序片段与参考基因组序列进行比对，对比对后的序列进行筛选，确定筛选后的序列片段的 HPV 型别或阴性；样本型别确定模块，用于对确定型别的序列片段按样本进行合并，并根据合并后支持对应型别的序列片段数量和比例进行筛选；最终确认每个样本的 HPV 型别或者确定为阴性。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，接收模块还用于：接收到测序序列后，对测序序列进行过滤，去除不合格的序列。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，分样本模块还用于：实现分样本操作后，将样本接头序列从序列片段中去除。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，合并筛选模块还用于：对确定型别的序列片段按样本进行合并后，对样本合并后的序列片段数量进行标准化。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，对样本合并后的序列片段数量进行标准化进一步包括：将各个文库中每个样品所拥有的序列数量，都按照比例缩放放到文库的测序量为理想情况下的平均测序量。

本发明提供的 HPV 精确分型的生物信息学分析的方法及系统，利用测序技术与分析手段，实现了高灵敏度和特异度，快速识别并确认 HPV 型别的目的。

附图说明

图 1 示出本发明实施例提供的一种 HPV 精确分型的生物信息学分析的方法的流程图；

图 2 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的另一个实施例的流程图；

图 3 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的另一个实施例的流程图；

图 4 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的另一个实施例的流程图；

图 5 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的一个具体实施方式的流程图；

图 6 示出本发明实施例提供的一种 HPV 精确分型的生物信息学分析的系统结构示意图；

图 7 示出本发明实施例提供的一种 HPV 精确分型的生物信息学分析及系统和分析过程中，各阶段的有效序列占原始序列的比例变化情况示意图；

图 8 示出本发明实施例提供的真实样本与阴性对照样本的有效序列片段数量的分布情况示意图；

图 9 示出本发明实施例提供的每个样本重复 10 次测序及分析后的重复性的结果示意图；

图 10 示出本发明实施例提供的所有真实样本检测出的阴阳性结果与血液阴性样本、临床检测结果的比较示意图；

图 11 示出本发明实施例提供的第二类文库中质粒样本的检测结果的示意图。

具体实施方式

本发明实施例中具体采用的样本包括：328 个患者真实组织样本、血液阴性样本、纯水阴性样本、载入特定 HPV 型别的质粒阳性样本。

在各个实施例中可以采用的上机策略包括：每个测序文库中包含 96 个样本，设计两类文库：第一类为 82 个患者真实组织样本，6 个纯水阴性样本，6 个血液阴性样本，2 个质粒阳性样本；第二类为质粒阳性样本 90 个，6 个纯水阴性样本。每个文库重复测序 10 次以方便验证信息分析时的重复性。因此共上机测序 50 个文库。

下面参照附图对本发明进行更全面的描述，其中说明本发明的示例性实施例。

图 1 示出本发明实施例提供的一种 HPV 精确分型的生物信息学分析的方法的流程图。

如图 1 所示，HPV 精确分型的生物信息学分析的方法 100 包括：步骤 102，接收高通量测序技术得到的测序片段。本发明中采用的高通量测序技术可以为 Illumina GA 测序技术，也可以是现有的其它高通量测序技术。

步骤 104, 将测序片段中的样本接头序列与样本接头序列库进行比对, 实现分样本操作。本发明实施例中采用的样本接头序列库是实验设计的 96 对引物-index (本发明中采用的样本接头序列库可根据实验要求及样品数量要求自行设计, 在设计过程中样本接头序列碱基分布及长度需注意综合考虑检测的样本个数和不同样本接头序列的非同源性。保证不同的样本通过样本接头序列比对实现样本区分)。

步骤 106, 将测序片段与参考基因组序列进行比对, 对比对后的序列进行筛选, 确定筛选后的序列片段的 HPV 型别或阴性。例如, 通过任何一种短序列映射程序 (如 blast 等映射程序), 将高通量测序技术得到的测序片段比对到参考基因组序列上, 其中, 参考基因组序列可取于公共数据库 NCBI, 该公共数据库可以通过如下途径获取 <http://www.ncbi.nlm.nih.gov/gene?term=hvp>。

本发明的一个实施例中, “对比对后的序列进行筛选”进一步包括: 将高通量测序技术得到的测序片段比对到参考基因组序列后, 筛选并去除比对结果中比对长度低于 70%, 或者一致性低于 85% 的序列 (100% 表示两条序列完全一致); 保留每条序列比的对结果中的最佳结果; 保留次优结果; 其中, 次优结果满足: 序列的一致性*比对长度、比对得分分别高于或等于最佳结果的 0.9 倍、0.85 倍, 且序列与参考序列匹配不相关的概率低于所述最佳结果的 10^3 倍; 判断每条序列的最佳结果和次优结果是否比对到同一型别或其亚型, 如果是, 则保留比对结果仅比对到某一型别的序列作为有效序列, 确定有效序列比对的 HPV 型别或阴性。

步骤 108, 对确定型别的序列片段按样本进行合并, 并根据合并后支持对应型别的序列片段数量和比例进行筛选; 最终确认每个样本的 HPV 型别或者确定为阴性。

稍后的其它实施例中还将举例对前述步骤中的具体实现方式

作进一步的详细介绍。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例，利用生物信息学的分析方法及技术手段，实现了快速检测大量样本、快速完成对感染 HPV 型别的检测，具有较高的灵敏度和特异性。

图 2 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的另一个实施例的流程图。

如图 2 所示，HPV 精确分型的生物信息学分析的方法 200 包括：步骤 202、203、204、206 和 208，其中步骤 202、204、206 和 208 可以分别执行与图 1 所示的步骤 102、104、106 和 108 相同或相似的技术内容，为简洁起见，这里不再赘述其技术内容。

如图 2 所示，在步骤 202 之后，执行步骤 203，对测序序列进行过滤，去除不合格的序列。

具体来说，步骤“对测序序列进行过滤，去除不合格的序列”进一步包括：预先设置不合格碱基的测序质量阈值和比例阈值（本发明中低质量阈值由具体测序技术及测序环境而定，例如，测序质量值低于 5 的碱基个数超过整条序列碱基个数的 50% 则认为是不合格序列）。

当测序序列中碱基的测序质量值低于测序质量阈值（如 5），且低于测序质量阈值的碱基个数占整条序列碱基个数的比例超过比例阈值（如 50%）时；则认为测序序列是不合格序列并加以过滤。

当测序序列的测序结果中不确定的碱基（如 Illumina GA 测序结果中的 N）的个数超过整条序列碱基个数的 10%，则认为测序序列是不合格序列并加以过滤。

与测序接头序列库进行比对时，如果测序序列中存在测序接头序列，则测序序列是不合格序列并加以过滤。

本发明提供的 HPV 精确分型的生物信息学分析的方法，通过对测序序列进行过滤，去除不合格的序列，进一步降低了不合格序列的影响，从而提高了检测分析的准确性。

图 3 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的另一个实施例的流程图。

如图 3 所示，HPV 精确分型的生物信息学分析的方法 300 包括：步骤 302、304、305、306 和 308，其中步骤 302、304、306 和 308 可以分别执行与图 1 所示的步骤 102、104、106 和 108 相同或相似的技术内容，为简洁起见，这里不再赘述其技术内容。

如图 3 所示，在步骤 304 之后，执行步骤 305，将样本接头序列从序列片段中去除。

具体来说，步骤“将样本接头序列从序列片段中去除”进一步包括：预先设置样本接头序列的测序质量阈值（如 5）和碱基数阈值（如 3）；将接头序列中碱基的测序质量值低于测序质量阈值，且碱基的数量超过碱基数阈值的序列去除。例如，综合考虑测序条件 and 环境，将本实施例中 10bp（碱基对）的接头序列中测序质量值低于 5 的碱基且个数大于 3 个的序列去除。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例中，进一步地执行：

步骤 a、将样本接头序列与样本接头序列库中序列进行完全匹配操作；

步骤 b、将样本接头序列降解 1-2bp 碱基，与样本接头序列库中序列对应部分进行完全匹配操作；

步骤 c、允许样本接头序列仅有一个碱基的插入，即在样本接头序列起始端进行完全匹配操作，当出现一个碱基无法匹配时将该碱基视为插入碱基，跳过此碱基后继续执行完全匹配操作；

步骤 d、允许样本接头序列仅有一个碱基的缺失，即在样本

接头序列中模拟缺失任何一个碱基后，进行完全匹配操作；

完成步骤 a - d 后，按照优先级顺序：步骤 a > 步骤 b > “步骤 c 或步骤 d” 的顺序确定最终的样本接头序列的比对结果（在处理接头比对的时候，有时候同一个序列会得到不同的比对结果，设置筛选比对结果的优先级可以理解为：步骤 a 的最高，b 次之，c 和 d 的优先级等同）。

比对到同一样本接头序列的被认为是来自同一样本的序列，从而区分样本；以及去除样本的序列中的样本接头序列（可能是 8 - 11bp）。

进一步的，如果步骤 a - d 四步操作中均无比对结果，或者一个步骤同时比对到两个结果，或者仅有步骤 c 和步骤 d 同时比对出结果；则认为该比对结果是由于无法区分而判定为无效信息，并将相应的整条序列去除。

本发明提供的 HPV 精确分型的生物信息学分析的方法的一个实施例，将测序片段中的样本接头序列与样本接头序列库进行比对，实现分样本操作后，再将样本接头序列从序列片段中去除，从而确保 HPV 分型分析的真实性和可靠性，为进一步的 HPV 精确分型提供保障。

图 4 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的另一个实施例的流程图。

如图 4 所示，HPV 精确分型的生物信息学分析的方法 400 包括：步骤 402、404、406、408、409 和 410，其中步骤 402、404、406 可以分别执行与图 1 所示的步骤 102、104、106 相同或相似的技术内容，为简洁起见，这里不再赘述其技术内容。

如图 4 所示，在步骤 406 之后，执行步骤 408，对确定型别的序列片段按样本进行合并。具体来说，在步骤 404 中，已经将各个序列是来自哪个样本的关系找到，按照这个关系，将属于同一

个样本的序列归在一起，统计他们与 HPV 参考基因组的比对结果。

步骤 409，对样本合并后的序列片段数量进行标准化。在本发明中，由于是不同文库的样品混在同一个 lane 里面测序，会由于各个文库上机浓度的不均一导致各个文库的样品的测序量有所不同，为了消除这个不同，我们将各个文库每个样品所拥有的序列数量，都按照比例缩放到该文库的测序量为理想情况下的平均测序量。即对每个样本的合并后的序列数量标准化。理想情况下是指各个混合在一个 lane 里面的样品的上样量都一样，且都被平均地测到同等的测序量，即不受实验、测序操作影响下理论的测序量。

标准化公式为：
$$\text{sample_read_num_STD} = \text{sample_read_num_ori} * (150000 / \text{read_num_ori})$$
；其中 sample_read_num_STD 表示标准化后的样本序列数；sample_read_num_ori 表示样本实际序列数；read_num_ori 表示样本对应文库测序下机的序列数。

步骤 410，根据标准化后支持对应型别的序列片段数量和比例进行筛选，最终确认每个样本的 HPV 型别或者确定为阴性。

标准化后，对样本现有信息进行过滤筛选，所采用的筛选条件按顺序如下：可用序列片段数小于一定阈值（如 137），则认为实验或者测序操作失败；比对结果支持 HPV 型别的序列片段数小于一定阈值（如 350），认为检测结果是阴性。比对结果支持 HPV 某型别的序列片段数占总序列片段数的比例达到预定阈值（该阈值的设定在具体实验背景下，需综合考虑检测的真实性及可重复性，如 12%）以上，则认为样本感染了该型别。其中，各部分的具体阈值视具体的实验情况而定，前述给出的具体参数值是通过实际样品，通过统计得出的，采用阴性样本检出数的平均值+4 倍

方差的值作为阴阳性判断的阈值。统计结果分别见图 2、5、6。不同测序平台，所应该用的过滤值会有所不同，实际生产中应该按照大致的方法先予确定；从而达到最终检测出样本感染的所有 HPV 型别或确定检测结果为阴性的目的。

图 5 示出本发明提供的 HPV 精确分型的生物信息学分析的方法的一个具体实施方式的流程图。

如图 5 所示，HPV 精确分型的生物信息学分析的方法 500 包括：步骤 502，接收高通量测序技术得到的测序片段。在本发明实施例中，采用 Illumina GA 高通量测序技术。

步骤 504，接收到测序序列后，对测序序列进行过滤，去除不合格的序列。不合格序列包括：测序质量值低于 5 的碱基个数超过整条序列碱基个数的 50% 则认为是不合格序列；序列中测序结果中 N 的个数超过整条序列碱基个数的 10% 则认为是不合格序列；与测序接头序列库进行比对，若序列中存在测序接头序列则认为是不合格序列。

步骤 506，将每个序列中样本接头序列与样本接头序列库进行比对，实现分样本操作。

步骤 508，将样本接头序列从序列片段中去除。具体来说，将接头序列中有测序质量值低于 5 的碱基个数大于 3 个的序列去除。而后，1) 将样本接头序列与样本接头序列库中序列进行完全匹配操作；2) 假设样本接头序列降解 1-2bp 与样本接头序列库中序列对应部分进行完全匹配操作；3) 允许样本序列仅有一个碱基的插入。在样本接头序列起始端进行完全匹配操作，当出现某碱基无法匹配时认为该碱基为插入碱基，跳过此碱基后继续严格的完全匹配操作；4) 允许样本序列仅有一个碱基的缺失。在样本接头序列中模拟缺失任何一个碱基后，进行完全匹配操作。完成上述四步操作后，按照 1)>2)>3)、4) 的优先级顺序确定最终的样本接

头序列的比对结果，而对于四步操作中四步均无比对结果，一个步骤同时比对到两个结果或仅有且同时 3、4 步骤比对出结果则认为这是由于无法区分，而判定为无效信息，将相应的整条序列去除。比对到同一样本接头序列的被认为是来自该一样本的序列，从而实现样本区分的目的。最后去除序列中样本接头序列部分（8-11bp）。

步骤 510，将测序片段与参考基因组序列进行比对，对比对后的序列进行筛选。确定筛选后的序列片段的 HPV 型别或阴性。本发明实施例采用 blast 映射程序，将高通量测序技术得到的测序片段比对到参考基因组序列上。比对后，筛选掉比对结果中比对长度低于 70%，或者一致性低于 85% 的序列。而后保留每条序列比对结果最好的，即 blast 软件比对输出的第一个比对结果，同时也保留次优结果；其中，次优结果满足：序列的一致性*比对长度、比对得分分别对应的高于或等于所述最佳结果的 0.9 倍、0.85 倍，且序列与参考序列匹配不相关的概率低于所述最佳结果的 10^3 倍。然后判断该序列比对到的型别是否是同一型别（或其亚型），最终仅保留筛选后的比对结果仅比对到某一型别的序列作为有效序列，实现确定每个序列比对的 HPV 型别或确认为阴性。

步骤 512，将确定型别的序列的比对结果按样本合并。具体来说，在步骤 506 中，已经将各个序列是来自哪个样本的关系找到，按照这个关系，将属于同一个样本的序列归在一起，统计他们与 HPV 参考基因组的比对结果。

步骤 514，对每个样本的合并后序列的数量标准化。在本发明中，为了消除这个不同，我们将各个文库每个样品所拥有的序列数量，都按照比例缩放到该文库的测序量为理想情况下的平均测序量。即对每个样本的合并后的序列数量标准化。

标 准 化 公 式 为 ： $\text{sample_read_num_STD} =$

$\text{sample_read_num_ori} * (150000/\text{read_num_ori})$; 其中 $\text{sample_read_num_STD}$ 表示标准化后的样本序列数; $\text{sample_read_num_ori}$ 表示样本实际序列数; read_num_ori 表示样本对应文库测序下机的序列数。

步骤 516, 根据标准化后支持对应型别的序列片段数量和比例进行筛选, 最终确认每个样本的 HPV 型别或者确定为阴性。在本实施例中, 标准化后, 按下述条件的先后顺序进行筛选: 可用序列数小于 137, 认为实验或者测序操作失败; 否则比对结果支持 HPV 型别的序列片段数小于 350, 就认为是阴性。比对结果支持 HPV 某型别的序列片段数占总序列片段数的 12% 以上, 则认为感染了该型别, 最终确定每个样本感染的 HPV 型别或者确定为阴性。

图 6 示出本发明实施例提供的一种 HPV 精确分型的生物信息学分析的系统的结构示意图。

如图 6 所示, 一种 HPV 精确分型的生物信息学分析的系统 600 包括: 接收模块 602、分样本模块 604、序列型别确定模块 606 和样本型别确定模块 608。其中

接收模块 602, 用于接收高通量测序技术得到的测序片段。

分样本模块 604, 用于将测序片段中的样本接头序列与样本接头序列库进行比对, 实现分样本操作。

序列型别确定模块 606, 用于将测序片段与参考基因组序列进行比对, 对比对后的序列进行筛选, 确定筛选后的序列片段的 HPV 型别或阴性。

样本型别确定模块 608, 用于对确定型别的序列片段按样本进行合并, 并根据合并后支持对应型别的序列片段数量和比例进行筛选; 最终确认每个样本的 HPV 型别或者确定为阴性。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个

实施例中，接收模块还用于：接收到测序序列后，对测序序列进行过滤，去除不合格的序列。具体采用的流程细节可参考方法实施例中的描述，这里就不再赘述。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，分样本模块还用于：实现分样本操作后，将样本接头序列从序列片段中去除。具体采用的流程细节可参考方法实施例中的描述，这里就不再赘述。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，合并筛选模块还用于：对确定型别的序列片段按样本进行合并后，对样本合并后的序列片段数量进行标准化。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例中，对样本合并后的序列片段数量进行标准化进一步包括：将各个文库中每个样品所拥有的序列数量，都按照比例缩放文库的测序量为理想情况下的平均测序量。具体采用的流程细节可参考方法实施例中的描述，这里就不再赘述。

本发明提供的 HPV 精确分型的生物信息学分析的系统的一个实施例，利用生物信息学的分析方法及技术手段，实现了快速检测大量样本、快速完成对感染 HPV 型别的检测，具有较高的灵敏度和特异性。

图 7 示出本发明实施例提供的一种 HPV 精确分型的生物信息学分析的方法及系统在分析过程中，各阶段的有效序列占原始序列的比例变化情况示意图。

如图 7 所示，横坐标代表测序文库代码，纵坐标代表有效序列占原始序列的比例。Filter 曲线表示不同测序文库在过滤测序序列后有效序列占原始序列的比例变化情况；Lib_Match 曲线表示不同测序文库在样本区分完成后有效序列占原始序列的比例变化情况；Final 曲线表示不同测序文库在序列 HPV 型别确定后有效序列占

原始序列的比例变化情况。本实例全部 50 个测序文库的序列有效利用率达到了 80% 以上。

图 8 示出本发明实施例提供的真实样本与阴性对照样本的有效序列片段数量的分布情况示意图。

如图 8 所示，阴性对照样本的有效序列片段的平均数为 19.82。该平均值加四倍的有效序列片段个数的标准差为 136.98。如图 8 所示，将 137 个有效序列片段作为实验或者测序操作成功与否界定值可以有效的将真实样本和阴性对照样本区分开来。

图 9 示出本发明实施例提供的每个样本重复 10 次测序及分析后的重复性的结果示意图。

为了评估将支持 HPV 型别的序列片段数 350 确定为检测结果阴阳性的界定值的有效性，图 9 示出了每个样本重复 10 次测序及分析后的重复性的结果。如图 9 所示，横坐标表示确定检测结果阴阳性的界定值，纵坐标表示所有样本重复率的平均值。本领域技术人员根据图 9 所示可以清楚地知晓，所有样本无论是在香港还是深圳进行测序，当将支持 HPV 型别的序列片段数 350 确定为检测结果阴阳性的界定值时，样本重复分析的重复性高达 99%，充分体现了本发明对于 HPV 检测的稳定性。

图 10 示出本发明实施例提供的所有真实样本检测出的阴阳性结果与血液阴性样本、临床检测结果的比较示意图。

如图 10 所示，血液（Blood DNA）是确定的没有 HPV 感染的阴性样本。在临床上将检测结果大于 1 的患者确认为 HPV 感染阳性。本领域技术人员根据图 10 所示可以清楚地知晓，将支持 HPV 型别的序列片段数 350 确定为检测结果阴阳性的界定值时，本实施例中确认 HPV 感染阴阳性的结果绝大部分与临床检测结果相同。而 350 的阈值又可以把血液阴性样本和阳性样本区分开来，避免了假阳性。由于临床检测结果并不能完全作为阳性对

照，所以本实施例的检测结果已足以证明本发明的精确性。

图 11 示出本发明实施例提供的第二类文库中质粒样本的检测结果的示意图。

如图 11 所示，横坐标表示为质粒中载入 HPV 病毒的型别，纵坐标表示的为实施例分析过程中支持对应 HPV 病毒型别的序列片段所占比例。本领域技术人员根据图 11 所示可以清楚地知晓，将支持 HPV 某型别的序列片段数比例达到 12%以上的样本确定为感染 HPV 的型别，可以有效的特异的检测出样本感染了的具体型别。

样品名称	HPV 检测 型别	样品名称	HPV 检测 型别	样品名称	HPV 检测 型别
sample 1	HBB	sample 33	HBB	sample 65	HBB
sample 2	HBB	sample 34	HBB	sample 66	HBB
sample 3	HPV51	sample 35	HBB	sample 67	HBB
sample 4	HBB	sample 36	HBB	sample 68	HBB
sample 5	HPV35	sample 37	HBB	sample 69	HBB
sample 6	HPV16	sample 38	HBB	sample 70	HBB
sample 7	HBB	sample 39	HBB	sample 71	HBB
sample 8	HBB	sample 40	HBB	sample 72	HBB
sample 9	HBB	sample 41	HBB	sample 73	HBB
sample 10	HBB	sample 42	HBB	sample 74	HBB

sample 11	HBB	sample 43	HPV6	sample 75	HBB
sample 12	-	sample 44	HBB	sample 76	HBB
sample 13	HBB	sample 45	-	sample 77	HBB
sample 14	HPV59	sample 46	HBB	sample 78	HBB
sample 15	HPV16	sample 47	-	sample 79	HBB
sample 16	HBB	sample 48	HBB	sample 80	HBB
sample 17	HBB	sample 49	HBB	sample 81	HBB
sample 18	HBB	sample 50	HBB	sample 82	HBB
sample 19	HBB	sample 51	HBB	质粒(33 型)	HPV33
sample 20	HPV16	sample 52	HBB	质粒(33 型)	HPV33
sample 21	HBB	sample 53	HBB	血液阴性样本	HBB
sample 22	HBB	sample 54	HBB	血液阴性样本	HBB
sample 23	HPV11	sample 55	HBB	血液阴性样本	HBB
sample 24	HBB	sample 56	HBB	血液阴性样本	HBB
sample 25	HBB	sample 57	HBB	血液阴性样本	HBB
sample 26	HBB	sample 58	-	血液阴性样本	HBB
sample 27	HBB	sample 59	HBB	纯水阴性样本	-
sample 28	HBB	sample 60	HBB	纯水阴性样本	-
sample 29	HBB	sample 61	HBB	纯水阴性样本	-
sample 30	HBB	sample 62	HBB	纯水阴性样本	-

sample 31	HBB	sample 63	HBB	纯水阴性样本	-
sample 32	-	sample 64	HBB	纯水阴性样本	-

表 1 样本库的检测结果

表 1 示出本发明实验例提供的一个样本库的检测结果。如表 1 所示，本表为第一类文库的一个样本库检测结果。其中，“HBB”表示检测结果为阴性，“-”表示因样品问题或实验问题导致检测到的序列数低于 137，认为此样品检测失败。

参考前述本发明示例性的描述，本领域技术人员可以清楚的知晓本发明具有以下优点：

1、本发明提供的 HPV 精确分型的生物信息学分析的方法及系统的一个实施例，利用生物信息学的分析方法及技术手段，实现了快速检测大量样本、快速完成对感染 HPV 型别的检测，具有较高的灵敏度和特异性。

2、本发明提供的 HPV 精确分型的生物信息学分析的方法及系统的一个实施例，通过对测序序列进行过滤，去除不合格的序列，进一步降低了不合格序列的影响，从而提高了检测分析的准确性。

3、本发明提供的 HPV 精确分型的生物信息学分析的方法及系统的一个实施例，将测序片段中的样本接头序列与样本接头序列库进行比对，实现分样本操作后，再将样本接头序列从序列片段中去除，从而确保 HPV 分型分析的真实性和可靠性，为进一步的 HPV 精确分型提供保障。

本发明的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本发明限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。本发明中描述的功能模块以及功能模块的划分方式仅为说明本发明的思想，本领域技术人员

根据本发明的教导以及实际应用的需要可以自由改变功能模块的划分方式及其模块构造以实现相同的功能；选择和描述实施例是为了更好说明本发明的原理和实际应用，并且使本领域的普通技术人员能够理解本发明从而设计适于特定用途的带有各种修改的各种实施例。

权 利 要 求 书

1. 一种 HPV 精确分型的生物信息学分析的方法，其特征在于，所述方法包括：

接收高通量测序技术得到的测序片段；

将所述测序片段中的样本接头序列与样本接头序列库进行比对，实现分样本操作；

将所述测序片段与参考基因组序列进行比对，对对比后的序列进行筛选，确定筛选后的序列片段的 HPV 型别或阴性；

对确定型别的序列片段按样本进行合并，并根据合并后支持对应型别的序列片段数量和比例进行筛选；最终确认每个样本的 HPV 型别或者确定为阴性。

2. 根据权利要求 1 所述的方法，其特征在于，所述方法还包括：接收到测序序列后，对测序序列进行过滤，去除不合格的序列。

3. 根据权利要求 2 所述的方法，其特征在于，步骤“对测序序列进行过滤，去除不合格的序列”进一步包括：

预先设置不合格碱基的测序质量阈值和比例阈值；

当测序序列中碱基的测序质量值低于所述测序质量阈值，且低于测序质量阈值的碱基个数占整条序列碱基个数的比例超过所述比例阈值时；则认为所述测序序列是不合格序列并加以过滤；

当所述测序序列的测序结果中不确定的碱基的个数超过整条序列碱基个数的 10%，则认为所述测序序列是不合格序列并加以过滤；

与测序接头序列库进行比对时，如果所述测序序列中存在测序接头序列，则所述测序序列是不合格序列并加以过滤。

4. 根据权利要求 1 所述的方法，其特征在于，所述方法还包括：实现分样本操作后，将所述样本接头序列从序列片段中去除。

5. 根据权利要求 4 所述的方法，其特征在于，步骤“将所述样本接头序列从序列片段中去除”进一步包括：

预先设置所述样本接头序列的测序质量阈值和碱基数阈值；

将接头序列中碱基的测序质量值低于所述测序质量阈值，且所述碱基的数量超过所述碱基数阈值的序列去除。

6. 根据权利要求 5 所述的方法，其特征在于，所述方法还包括：

步骤 a、将样本接头序列与样本接头序列库中序列进行完全匹配操作；

步骤 b、将样本接头序列降解 1-2bp 碱基，与样本接头序列库中序列对应部分进行完全匹配操作；

步骤 c、允许样本接头序列仅有一个碱基的插入，即在所述样本接头序列起始端进行完全匹配操作，当出现一个碱基无法匹配时将该碱基视为插入碱基，跳过此碱基后继续执行完全匹配操作；

步骤 d、允许样本接头序列仅有一个碱基的缺失，即在样本接头序列中模拟缺失任何一个碱基后，进行完全匹配操作；

完成所述步骤 a-d 后，按照优先级顺序：步骤 a>步骤 b>“步骤 c 或步骤 d”的顺序确定最终的样本接头序列的比对结果；

比对到同一样本接头序列的被认为是来自同一样本的序列，从而区分样本；以及
去除所述样本的序列中的样本接头序列。

7. 根据权利要求 6 所述的方法，其特征在于，所述方法还包括：

如果步骤 a - d 四步操作中均无比对结果，或者一个步骤同时比对到两个结果，或者仅有步骤 c 和步骤 d 同时比对出结果；则认为该比对结果是由于无法区分而判定为无效信息，并将相应的整条序列去除。

8. 根据权利要求 1 所述的方法，其特征在于，步骤“对对比后的序列进行筛选”进一步包括：

将高通量测序技术得到的测序片段比对到参考基因组序列上；

比对后，筛选并去除比对结果中比对长度低于 70%，或者一致性低于 85% 的序列；

保留每条序列比对结果中的最佳结果；

保留次优结果；其中，所述次优结果满足：序列的一致性*比对长度、比对得分分别对应的高于或等于所述最佳结果的 0.9 倍、0.85 倍，且序列与参考序列匹配不相关的概率低于所述最佳结果的 10^3 倍；

判断每条序列的最佳结果和次优结果是否比对到同一型别或其亚型，如果是，则保留比对结果仅比对到某一型别的序列作为有效序列，确定所述有效序列比对的 HPV 型别或阴性。

9. 根据权利要求 1 所述的方法，其特征在于，所述方法还包

括:

对确定型别的序列片段按样本进行合并后, 对所述样本合并后的序列片段数量进行标准化。

10. 根据权利要求 9 所述的方法, 其特征在于, 对所述样本合并后的序列片段数量进行标准化进一步包括:

将各个文库中每个样品所拥有的序列数量, 都按照比例缩放到所述文库的测序量为理想情况下的平均测序量。

11. 根据权利要求 9 所述的方法, 其特征在于, 步骤“根据合并后支持对应型别的序列片段数量和比例进行筛选”进一步包括:

标准化后, 按下述条件的先后顺序进行筛选:

如果可用序列数小于阴性对照样本的有效序列片段的平均个数加上其四倍标准差的和, 认为实验或者测序操作失败;

否则, 如果比对结果支持 HPV 型别的序列片段数小于预定阈值, 就认为是阴性;

如果比对结果支持 HPV 型别的序列片段数占总序列片段数的比例达到预定阈值以上, 则认为感染了该型别。

12. 一种 HPV 精确分型的生物信息学分析的系统, 其特征在于, 所述系统包括:

接收模块, 用于接收高通量测序技术得到的测序片段;

分样本模块, 用于将所述测序片段中的样本接头序列与样本接头序列库进行比对, 实现分样本操作;

序列型别确定模块, 用于将所述测序片段与参考基因组序列进行比对, 对比对后的序列进行筛选, 确定筛选后的序列片段的 HPV 型别或阴性;

样本型别确定模块，用于对确定型别的序列片段按样本进行合并，并根据合并后支持对应型别的序列片段数量和比例进行筛选；最终确认每个样本的 HPV 型别或者确定为阴性。

13. 根据权利要求 12 所述的系统，其特征在于，所述接收模块还用于：接收到测序序列后，对测序序列进行过滤，去除不合格的序列。

14. 根据权利要求 12 所述的系统，其特征在于，所述分样本模块还用于：实现分样本操作后，将所述样本接头序列从序列片段中去除。

15. 根据权利要求 12 所述的系统，其特征在于，所述样本型别确定模块还用于：对确定型别的序列片段按样本进行合并后，对所述样本合并后的序列片段数量进行标准化。

16. 根据权利要求 15 所述的系统，其特征在于，对所述样本合并后的序列片段数量进行标准化进一步包括：

将各个文库中每个样品所拥有的序列数量，都按照比例缩放到所述文库的测序量为理想情况下的平均测序量。

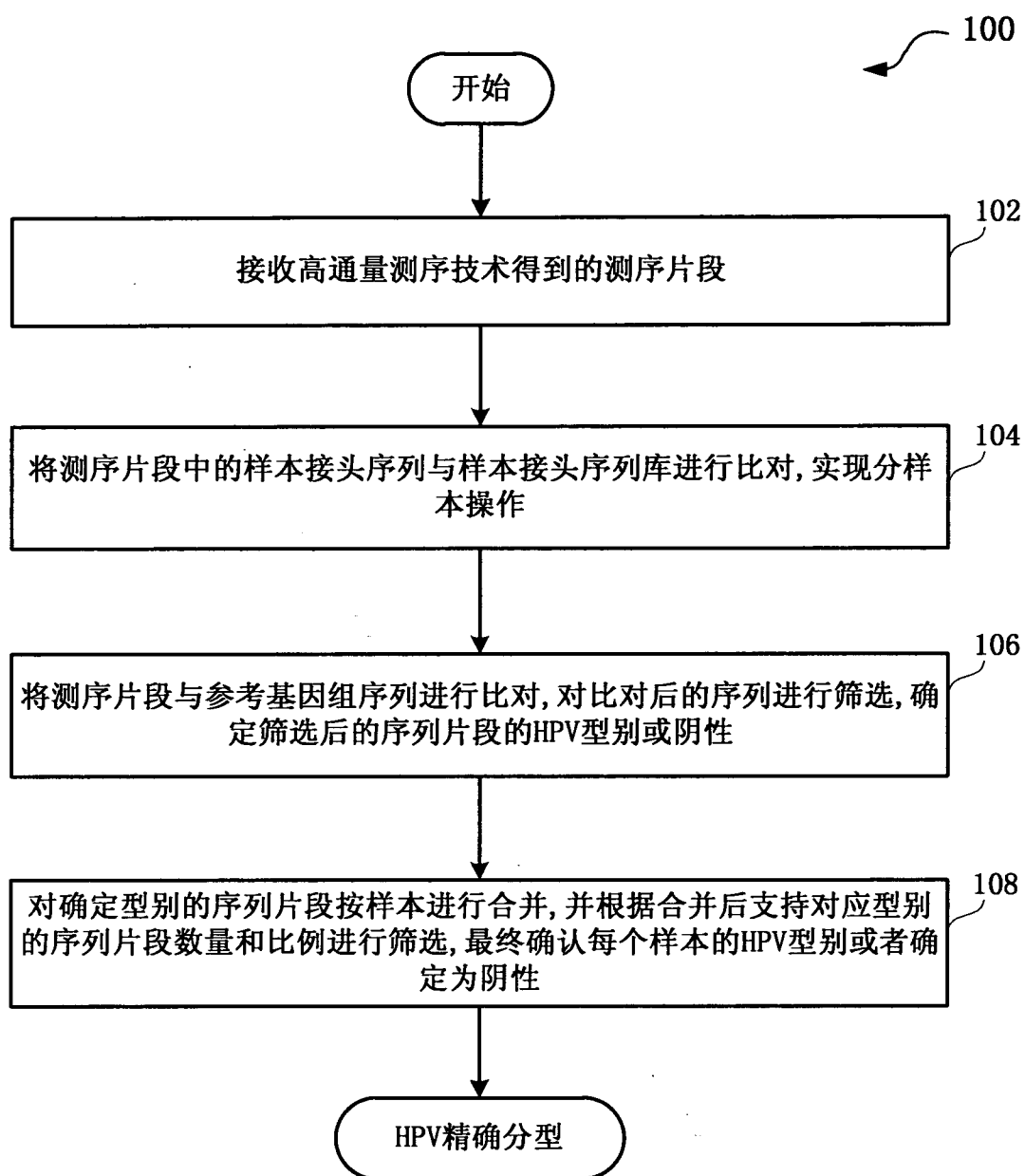


FIG.1

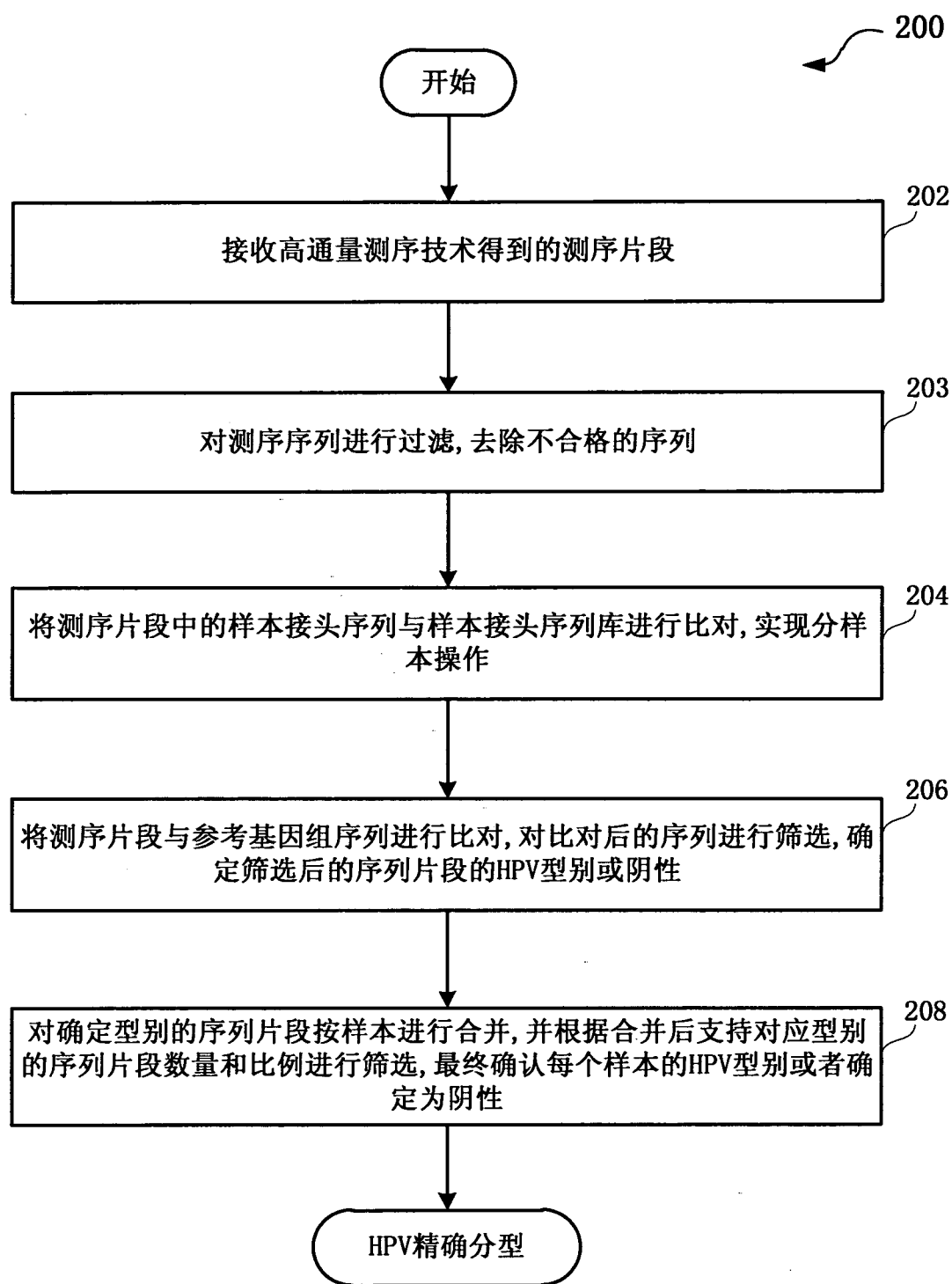


FIG.2

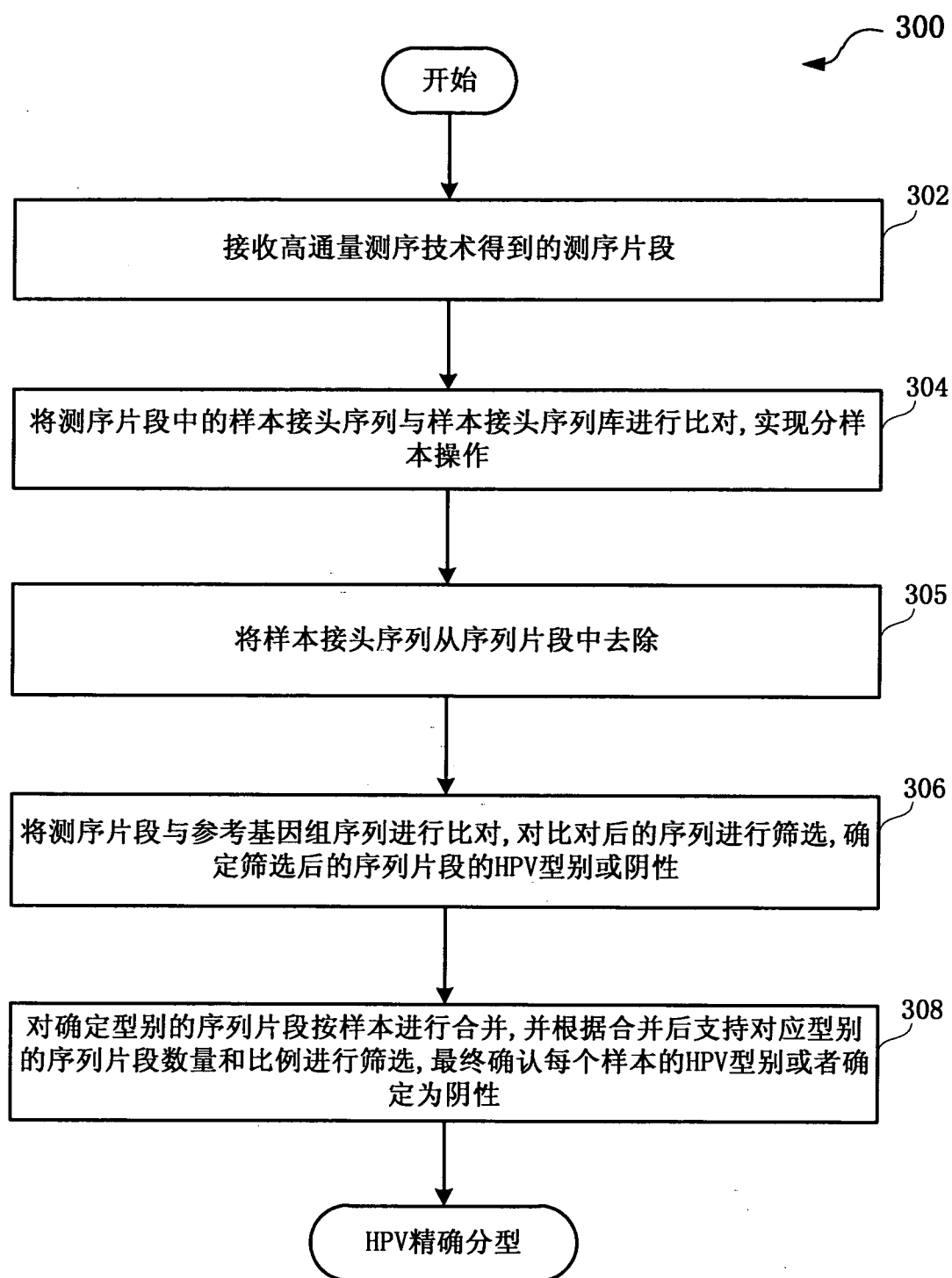


FIG.3

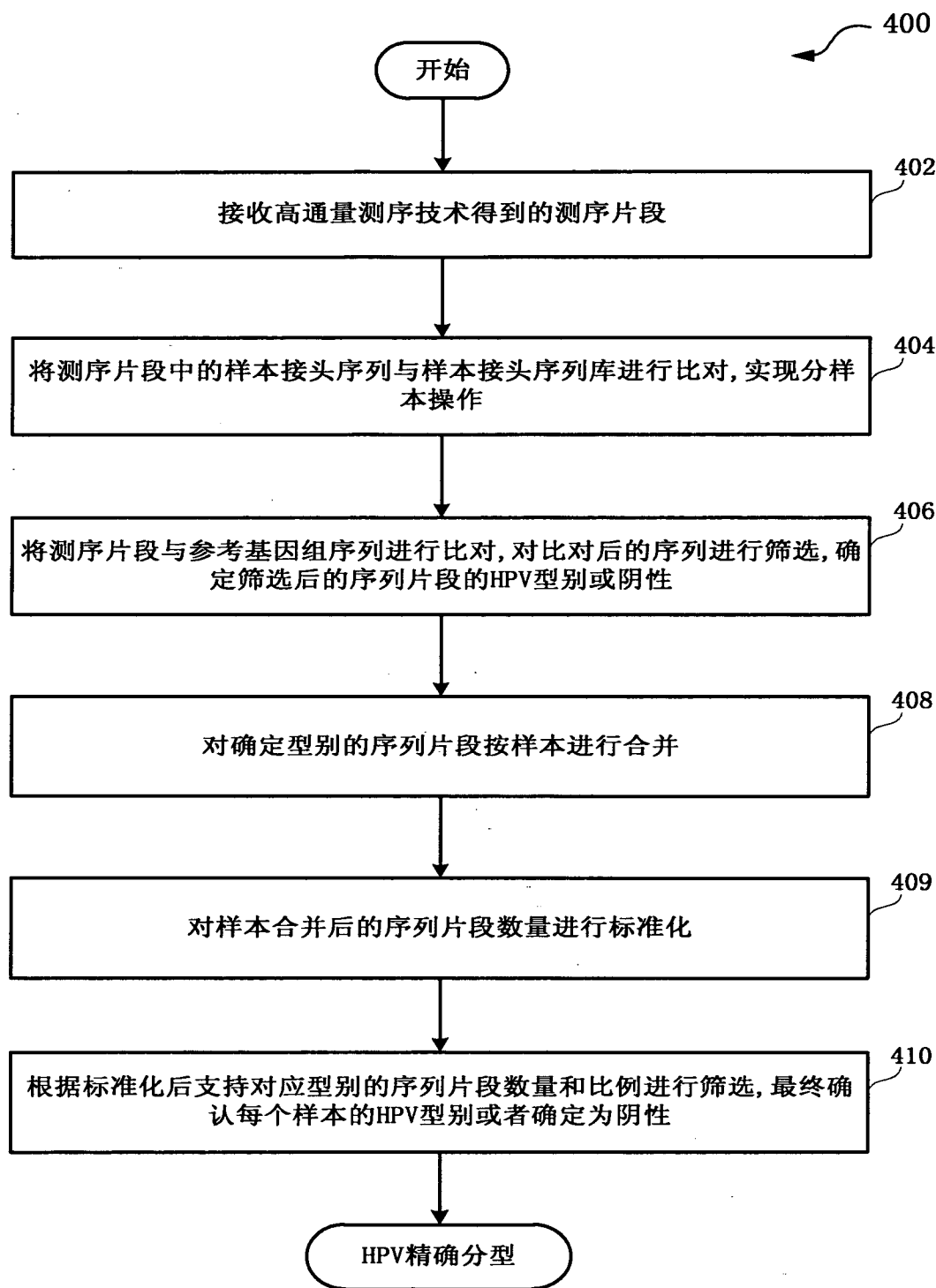


FIG.4

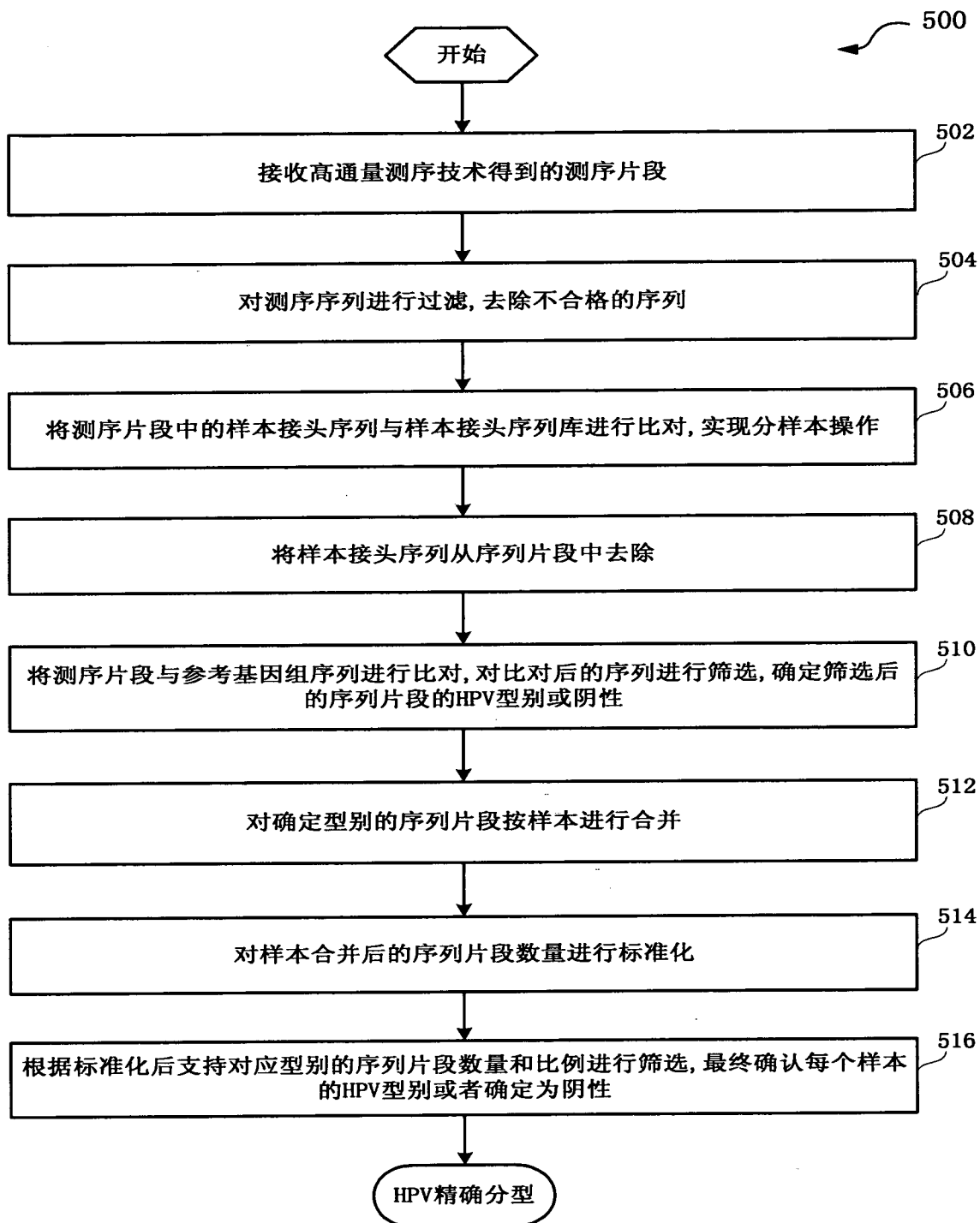
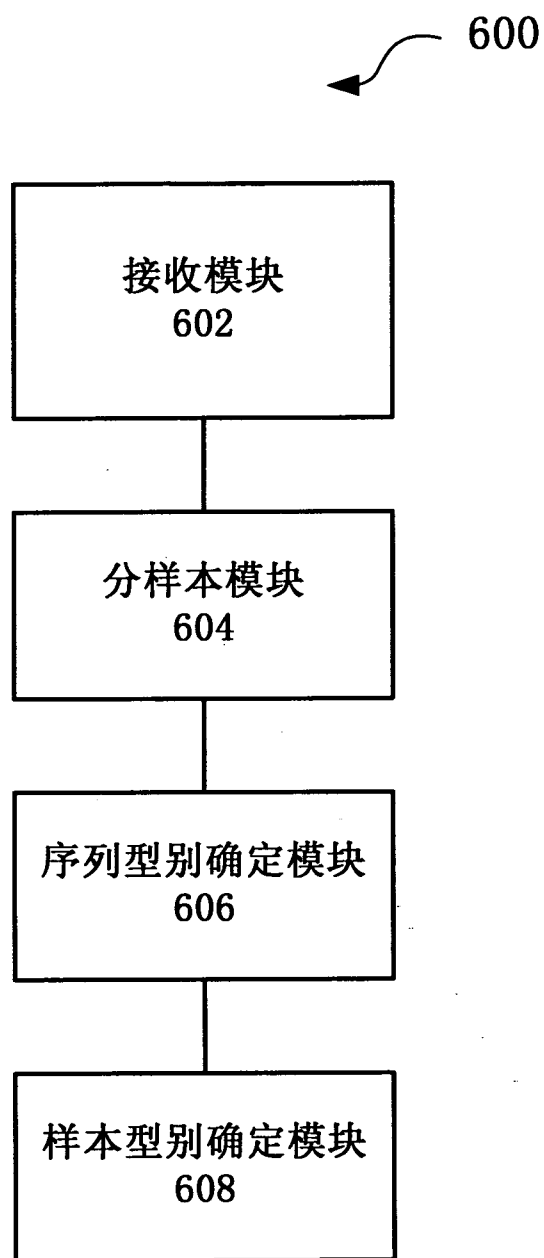
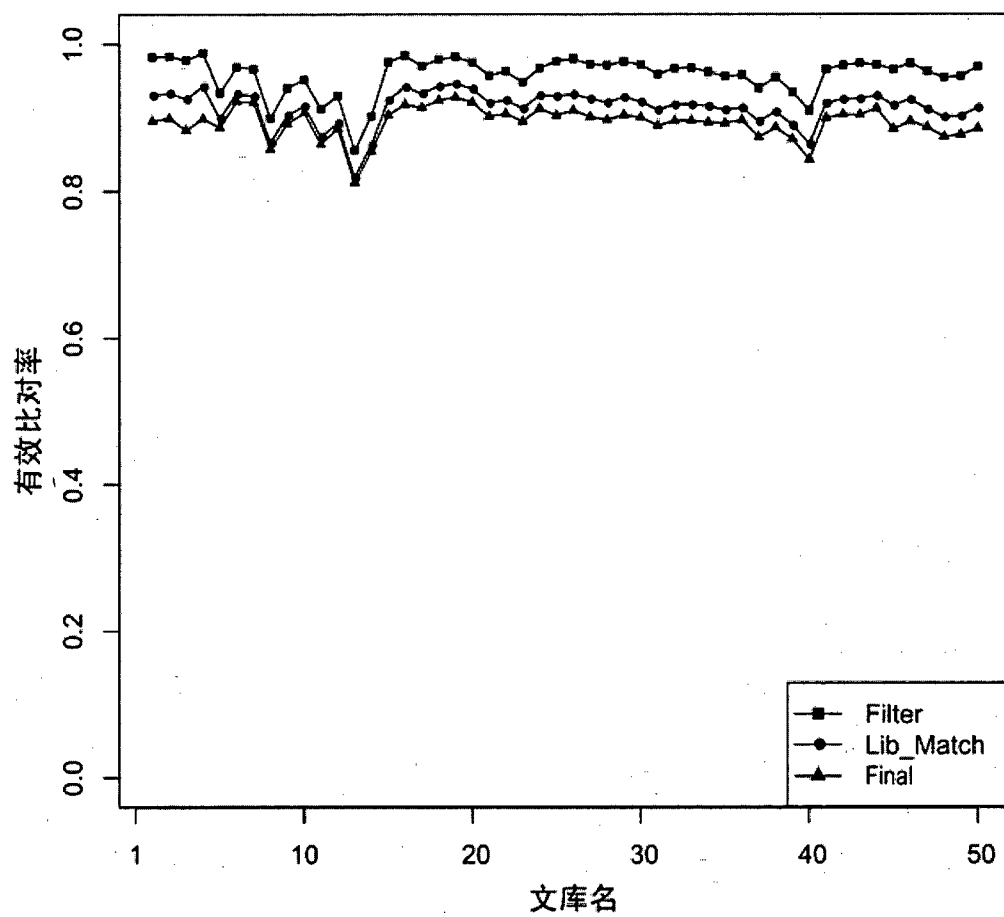
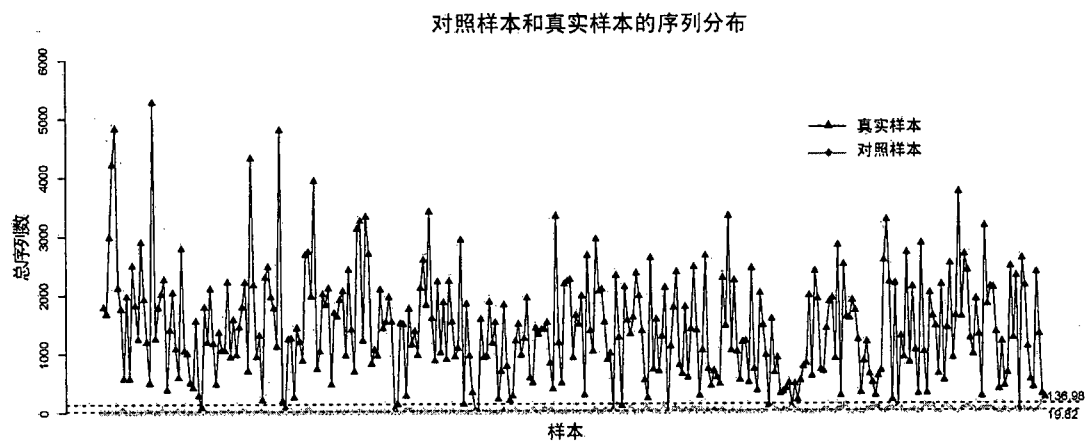


FIG.5

**FIG.6**

**FIG.7**

**FIG.8**

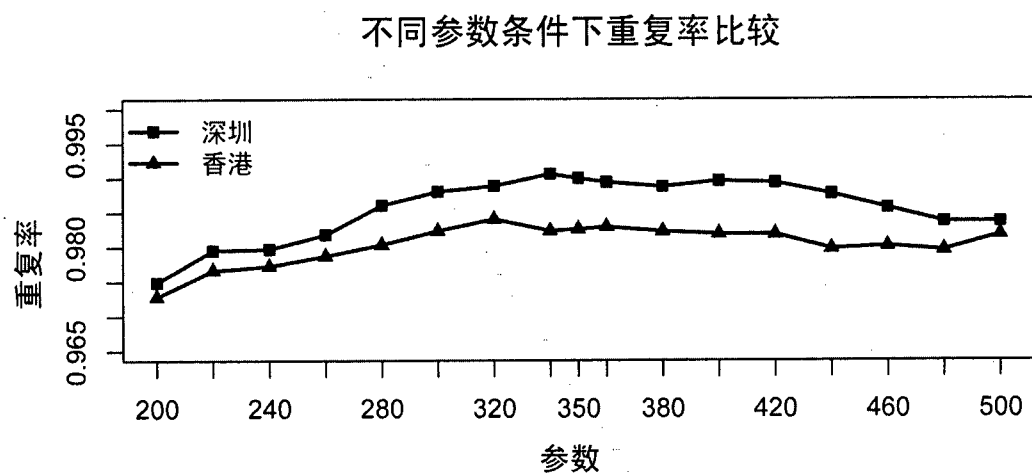
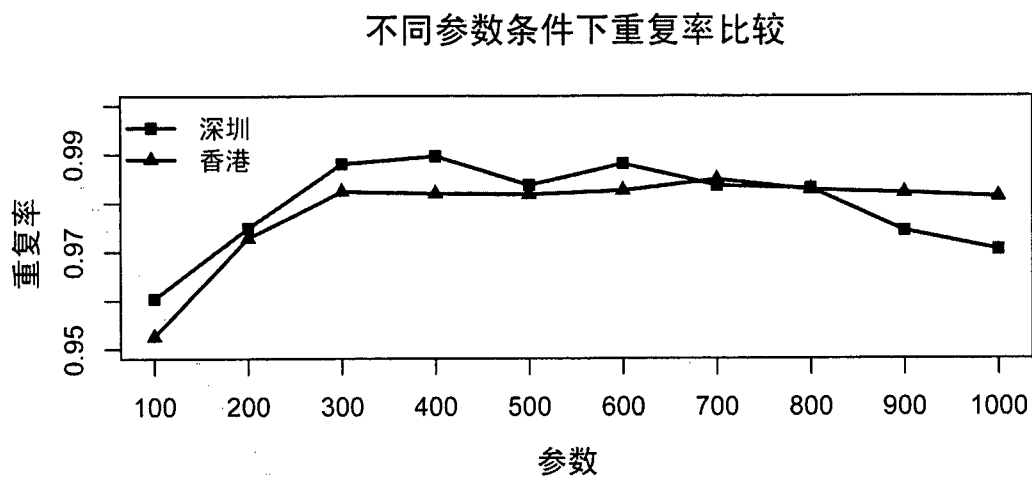


FIG.9

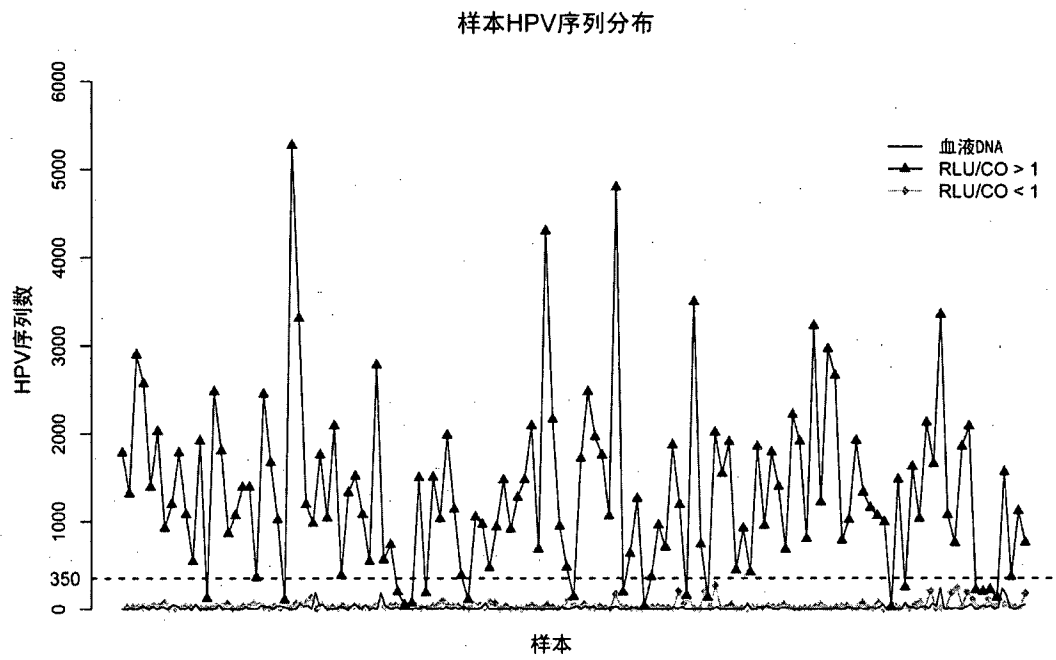


FIG.10

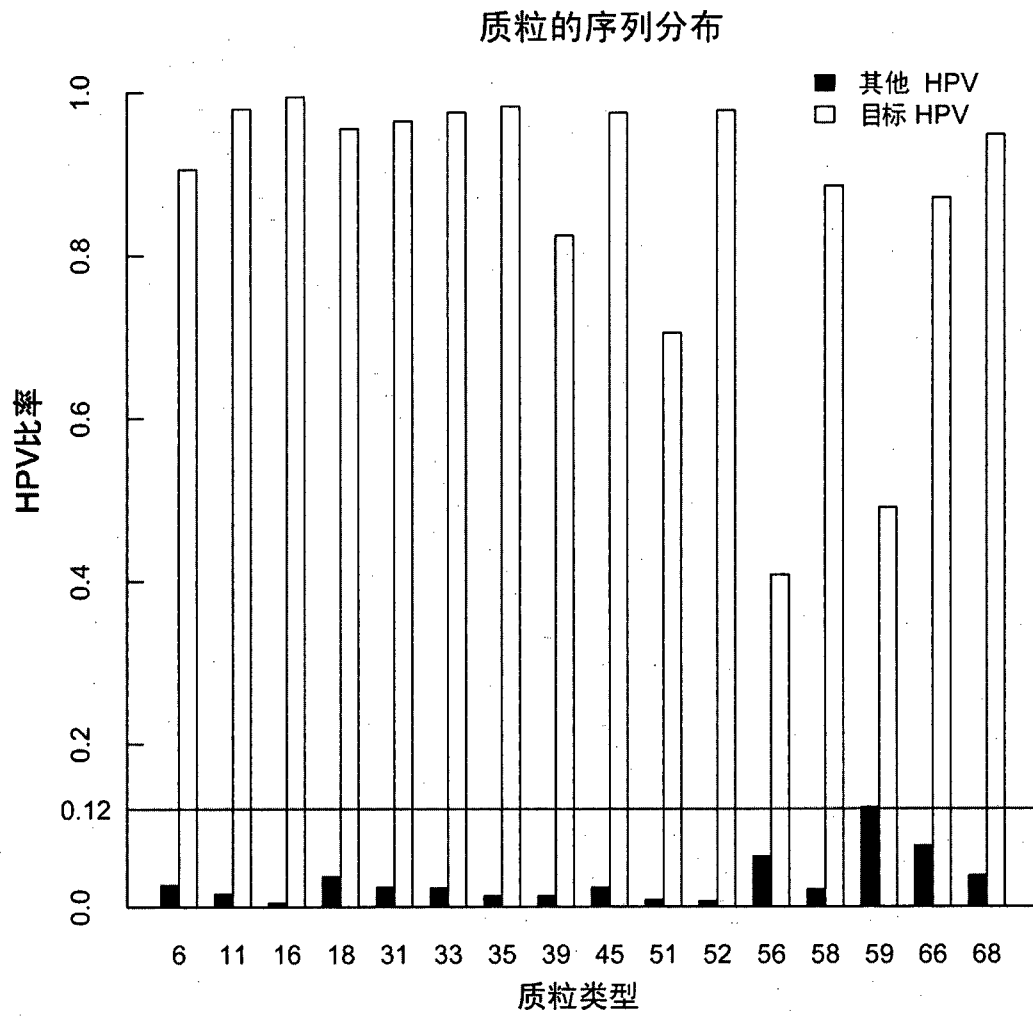


FIG.11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/001943

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/70 (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, G01N

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)
IEEE, WPI, EPODOC: bioinformatics, HPV, typing, precise, genomic, sequence, sample, match, negative, filter
CNKI, CNPAT: bioinformatics, HPV, typing, precise, genomic, sequence, sample, match, negative, filter (in Chinese words)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 101397590 A (HANGZHOU DIAN MEDICAL TEST CENTER CO., LTD.) 01 Apr. 2009 (01.04.2009) page 5 in the description	1-16
A	CN 101435002 A (BGI SHENZHEN CO., LTD.) 20 May 2009 (20.05.2009) the whole document	1-16
A	CN 101838709 A (ZHONGSHAN UNIVERSITY) 22 Sep. 2010 (22.09.2010) the whole document	1-16

☐ 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 26 May 2011 (26.05.2011)	Date of mailing of the international search report 28 Jul. 2011 (28.07.2011)
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 LIU, Chengen Telephone No. (86-10) 82245140

International application No.
PCT/CN2010/001943

Form PCT/ISA /210 (patent family annex) (July 2009)

A. 主题的分类

C12Q 1/70 (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, G01N

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))

IEEE, WPI, EPODOC: bioinformatics, HPV, typing, precise, genomic, sequence, sample, match, negative, filter

CNKI, CNPAT: 生物信息学, HPV, 分型, 精确, 基因, 序列, 样本, 匹配, 阴性, 筛选

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
A	CN 101397590 A(杭州迪安医学检验中心有限公司)01.4 月 2009(01.04.2009) 说明书第 5 页	1-16
A	CN 101435002 A(深圳华大基因科技有限公司)20.5 月 2009(20.05.2009) 全文	1-16
A	CN 101838709 A(中山大学)22.9 月 2010(22.09.2010) 全文	1-16

☐ 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
26.5 月 2011(26.05.2011)国际检索报告邮寄日期
28.7 月 2011 (28.07.2011)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授权官员

刘承恩
电话号码: (86-10) 82245140

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2010/001943

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 101397590 A	01.04.2009	无	
CN 101435002 A	20.05.2009	无	
CN 101838709 A	22.09.2010	无	