

(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



(10) 国际公布号
WO 2012/175013 A1

(43) 国际公布日
2012年12月27日 (27.12.2012)

- (51) 国际专利分类号:
C12Q 1/68 (2006.01) C12M 1/34 (2006.01)
C12Q 1/70 (2006.01)
- (21) 国际申请号: PCT/CN2012/077175
- (22) 国际申请日: 2012年6月19日 (19.06.2012)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201110174686.2 2011年6月24日 (24.06.2011) CN
PCT/CN2011/082855 2011年11月24日 (24.11.2011) CN
- (71) 申请人 (对除美国外的所有指定国): **深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。 **深圳华大基因研究院 (BGI SHENZHEN)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
- (72) 发明人; 及
- (75) 发明人/申请人 (仅对美国): **李伟阳 (LI, Weiyang)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **曾玺 (ZENG, Xi)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **徐佳佳 (XU, Jiajia)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **蒋慧 (JIANG, Hui)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **杨玲 (YANG, Ling)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **易赏 (YI, Shang)**

[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **陈盛培 (CHEN, Shengpei)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **管彦芳 (GUAN, Yanfang)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **易鑫 (YI, Xin)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **胡学达 (HU, Xueda)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **李英睿 (LI, Yingrui)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **张秀清 (ZHANG, Xiuqing)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **杨焕明 (YANG, Huanming)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

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

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

[见续页]

(54) Title: SYSTEM AND METHOD FOR DIAGNOSING HUMAN BODY WITH ABNORMAL STATE

(54) 发明名称: 用于确定人体具有异常状态的系统和方法

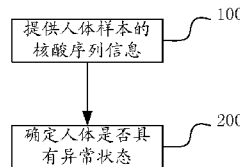


图 1 / Fig. 1

100 PROVIDING THE NUCLEOTIDE SEQUENCE INFORMATION ABOUT A HUMAN SAMPLE
200 DETERMINING WHETHER THE HUMAN BODY HAS AN ABNORMAL STATE

(57) Abstract: Provided are a system and a method for diagnosing a human body with an abnormal state. The method for determining the human body with an abnormal state comprises: providing nucleotide sequence information about a human sample, wherein the nucleotide sequence information about the human sample is obtained based on human sample testing; and determining whether the human body has an abnormal state based on the nucleotide sequence information about the human sample.

(57) 摘要: 提供了用于确定人体具有异常状态的系统和方法。其中, 确定人体具有异常状态的方法包括: 提供人体样本的核酸序列信息, 该人体样本的核酸序列信息是基于对人体样本进行检测而获得的; 以及基于人体样本的核酸序列信息, 确定人体是否具有异常状态。



WO 2012/175013 A1

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

CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。
- 包括说明书序列表部分(细则 5.2(a))。

用于确定人体具有异常状态的系统和方法

优先权信息

本申请请求 2011 年 6 月 24 日向中国国家知识产权局提交的、专利申请号为
5 201110174686.2 的专利申请, 以及 2011 年 11 月 24 日向中国国家知识产权局提交的、
专利申请号为 PCT/CN2011/082855 的专利申请的优先权和权益, 并且通过参照将其全文
并入此处。

技术领域

10 本发明涉及生物学领域。更具体地, 本发明涉及用于确定人体具有异常状态的系
统和方法。

背景技术

HBV 是一种全球性的慢性病毒感染性疾病。我国的乙肝病毒感染率约 60%-70%,
15 而乙肝表面抗原携带率约占总人口的 7.18%, 以此计算, 全国约有 9300 万人携带乙肝
病毒, 其中乙肝患者大约有 3000 万。在全世界, 大约有 45% 的人群生活在慢性 HBV
感染的高发区, 43% 的人群生活在慢性 HBV 感染的中发区, 并且 HBV 为已知的引起肝
硬化以及肝癌的主要原因, 现已知 HBV 基因组与人类基因组整合为肝癌的主要诱因之
一。EBV 现已知为鼻咽癌的主要诱因, 初发症状到死亡的自然病程从 3~13 个月不等,
20 放射治疗后 5 年生存率为 8%~62%。幽门螺旋杆菌为胃癌的主要诱因, 并且慢性胃炎
患者的胃粘膜活检标本中幽门螺杆菌检出率可达 80%~90%, 而消化性溃疡患者更高,
可达 95% 以上, 甚至接近 100%。胃癌由于局部上皮细胞已发生异化, 因此检出率高低
报道不一。以上病原体, 包括 HBV, HCV, HIV, EBV, 幽门螺旋杆菌等, 均能够与
25 宿主基因组进行整合, 并且与引起的相关病症直接相关, 能够对感染者造成持久和高危
性的危害, 是高致病性, 高致癌性的高危型病原微生物。因此, 建立一种无创的诊断及
病程追踪监测手段, 既能够免除对病原体感染患者进行病理组织取样的伤害, 同时可以
定期跟踪检测感染者感染程度以及治疗效果, 提高感染患者的治愈率以及有利于及时地
指导用药, 对于由病原体引起的肿瘤晚期病人也可进行及时检测判断治愈效果以及复发的
可能性, 能够及时有效地对患者进行病情随访并给出最佳建议。

30 然而, 现有技术并不能满足上述要求。

发明内容

本发明旨在至少解决现有技术中存在的技术问题之一。为此, 本发明的一个目的在
于提出一种具有能够有效确定人体异常状态的方法。本发明的另一目的在于提出一种能够
35 有效确定人体异常状态的系统。

本发明是基于发明人的下列发现而完成的: 在人体样本的检测过程中, 对于蛋白质的

检测，通常由于实验条件的限制，无法早期获得人体的状态信息。而利用核酸序列的特性，则可以通过对于离体的人体样本进行核酸分析，尽可能早期地对人体状态进行分析。

在本发明的一个方面，本发明提出了一种确定人体具有异常状态的方法。根据本发明的实施例，该方法包括：提供人体样本的核酸序列信息，所述人体样本的核酸序列信息是基于对所述人体样本进行检测而获得的；以及基于所述人体样本的核酸序列信息，确定所述人体是否具有异常状态。根据该实施例的方法，通过对人体样本的核酸序列信息进行分析，可以根据核酸序列中所包含的信息确定人体是否具有异常状态。由于核酸序列信息与原位状态的核酸信息保持一致，因而可以有效地确定人体是否具有异常状态。

另外，根据本发明上述实施例，上述确定人体具有异常状态的方法，还可以具有如下附加的技术特征：

根据本发明的一个实施例，所述人体样本的核酸序列信息是基于对所述人体样本进行核酸序列检测而获得的。由此，可以通过核酸序列检测方法，容易地获得人体样本的核酸序列信息，从而提高了确定人体具有异常状态的效率。根据进一步的实施例，所述核酸序列检测是借助第二代测序技术或第三代测序技术进行的。由此，可以高效地对人体样本进行核酸序列进行检测，并且能够实现高通量深度测序。本发明的发明人发现基于第二代测序技术和第三代测序技术的高效、高精度的性质，可以实现对人体样本核酸序列信息的高效、高精度检测，能够非常灵敏地对人体样本中痕量的核酸进行检测。

根据本发明的一个实施例，所述样本为所述人体的细胞、组织、血液、体液、尿液、排泄物或其组合。由此，根据本发明的实施例的确定人体具有异常状态的方法能够应用于各种人体样本，并且能够根据不同的人体样本的特点，确定多种异常状态。

根据本发明的一个实施例，所述人体样本为血浆或者血清。根据该实施例，可以利用常规的方法获得人体的血浆和血清样本，并对其进行核酸序列分析，可以直接对人体的异常状态进行确定。

根据本发明的一个实施例，所述核酸序列信息包括所述人体样本中游离核酸的序列信息。由此，可以根据游离核酸的序列信息，对人体的异常状态进行确定。并且基于游离核酸的序列信息与人体正常的核酸序列或者病原体的核酸序列进行比对分析后，可以获得多种人体的异常状态。

根据本发明的一个实施例，所述人体样本中游离核酸的序列信息是通过除去所述人体样本中的细胞后，进行测序检测而获得的。由此，可以提高游离核酸的测序检测的精度和准确度，从而能够进一步有效地确定人体具有异常状态。

根据本发明的一个实施例，所述异常状态选自疾病的发生、疾病的发展阶段、疾病的疗效和预后的至少一种。由此，可以确定与人体密切相关的疾病的发生、发展、疗效以及预后，从而有利于制定有效的治疗方案。根据本发明进一步的实施例，所述疾病是肿瘤性疾病、免疫性疾病、遗传性疾病的至少一种。根据本发明的实施例的确定人体异常状态的方法，能够有效地确定人体是否患有肿瘤性疾病、免疫性疾病、遗传性疾病。根据具体的示例，所述肿瘤性疾病是选自肺癌、肝癌、胃癌、食管癌、结直肠癌、胰腺癌、乳腺癌、

膀胱癌、肾癌、卵巢癌、宫颈癌、甲状腺癌、鼻咽癌、脑胶质瘤的至少一种。根据更进一步的示例，如果所述核酸序列信息包含选自下列至少一种的核酸片段序列：HBV、HPV、EBV、幽门螺旋杆菌，则确定所述人体患有宫颈癌、肝癌、鼻咽癌、胃癌的至少一种。由此，根据本发明的实施例的确定人体具有异常状态的方法，可以基于所检测的核酸序列信息，有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。

5 根据本发明的一个实施例，对核酸进行测序检测之前，可以利用探针去除含有特定序列的核酸，然后对所述去除后剩余的核酸进行测序检测。由此，可以通过探针除去具有特定序列的核酸，从而能够提高对剩余材料进行序列检测的精度和准确性。根据本发明的具体示例，用于去除特定序列的核酸的探针可以结合人体基因组中的共有序列，或者为可以结合人体基因组中甲基化位点的抗体或蛋白。由此，可以进一步提高检测的精度和准确性。

10 根据本发明的实施例，对所述核酸进行测序检测之前，还可以利用探针捕获含有特定序列的核酸，然后对所述含有特定序列的核酸进行测序检测。由此，可以通过探针，预先对进行核酸序列分析的核酸进行筛选，从而能够进一步提高确定人体具有异常状态的方法的效率。根据进一步的实施例，所述探针对于选自下列的至少一种是特异性的：HBV、HPV、EBV、幽门螺旋杆菌，由此，可以有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。

15 因此，在本发明的另一方面，本发明还提供了一种核酸探针集。根据本发明的实施例，该核酸探针集包括多个探针，且具有以下特征：

- (1) 每个探针上具有 1 个或多个生物素标记的 dNTP；和/或
- (2) 生物素标记的 dNTP 在核酸探针集中的丰度为 1: 6-1: 2；和/或
- 20 (3) 核酸探针集的全部核酸序列覆盖对应选自 HBV、HPV、EBV 和幽门螺旋杆菌的至少一种病毒的基因组序列的 70%-100%。

在另一优选例中，本发明的核酸探针集具有 1-20000 个核酸探针；较佳地，核酸探针集具有 1000-5000 个核酸探针；更佳地，核酸探针集具有 2500 个核酸探针。

在另一优选例中，生物素标记的 dNTP 在核酸探针集中的丰度为 1: 4。

25 在另一优选例中，核酸探针集中，探针之间具有部分重叠。

在另一优选例中，核酸探针集(在本文中有时也称为“探针集”)中的探针长度为 100-500 bp；较佳地，探针长度为 200-300 bp；更佳地，探针长度为 250 bp。

在另一优选例中，探针是以病毒基因组作为模板，PCR 法扩增获得的，较佳地，扩增模板为乙型肝炎病毒 (HBV) 基因组、丙型肝炎病毒 (HCV) 基因组、艾滋病病毒 (HIV) 基因组、乳头瘤病毒 (HPV) 基因组，或其组合；更佳地，扩增模板为 B 型 HBV 基因组和/或 C 型 HBV 基因组。

30 进一步，在本发明的又一方面，还提供了一种表面固定有本发明的核酸探针集的核酸芯片。

在本发明的再一方面，还提供了本发明的核酸探针集和核酸芯片的用途，用于检测病毒在待测样本中的整合方式；较佳地，整合方式选自下组：重排、异位、插入、替换，或其组合。

在本发明的另一方面，还提供了一种制备本发明的核酸探针的方法，包括步骤：

a. 获得探针来源样本；

b. 对步骤(a)获得的样本进行PCR扩增，PCR扩增体系的dNTP为生物素标记的dNTP，以便获得带有生物素标记的PCR扩增产物；

5 c. 对步骤(b)获得的生物素标记的PCR扩增产物进行打断，得到片段化的生物素标记的PCR扩增产物，即为探针。

在另一优选例中，步骤(a)所述样本具有以下特征：

样本为含有核酸的病毒样本；和/或

样本为病毒粒子、血清、血液、组织样本、脱落细胞，上皮细胞，或其组合；和/或

10 样本选自下组：乙型肝炎病毒(HBV)、丙型肝炎病毒(HCV)、艾滋病病毒(HIV)、乳头瘤病毒(HPV)，或其组合；和/或

样本为B型HBV和/或C型HBV。

在另一优选例中，步骤(b)具有以下特征：

步骤(b)所述的扩增为对样本中病毒DNA全长进行扩增；和/或

15 步骤(b)所述标记的dNTP为biotin-dNTP，且标记的dNTP能够与链霉素亲和磁珠结合；和/或

步骤(b)所述标记的dNTP与非标记dNTP的比例为1:2-8；优选比例为1:3-6；更优选比例为1:4。

在另一优选例中，步骤(c)所述打断为超声法打断。

20 在另一优选例中，还包括步骤(d)：对步骤(c)获得的探针进行纯化和/或定量。

在另一优选例中，根据本发明的制备核酸探针的方法制备的探针，其长度为100-500bp；较佳地，探针的长度为200-300bp，更佳地，探针的长度为250bp。

在本发明的又一方面，本发明还提供了一种检测病毒在待测样本中基因整合方式的方法，包括步骤：

25 (i) 获得待测样本；

(ii) 对步骤(i)获得的样本进行文库构建；

(iii) 将本发明的探针与步骤(ii)获得的文库进行杂交，捕获与病毒基因整合有关的核酸序列；

(iv) 对步骤(iii)捕获的核酸序列进行扩增，获得与病毒整合有关的扩增产物；

30 (v) 对步骤(iv)获得的扩增产物进行测序，获得与病毒整合方式有关核酸信息。

在另一优选例中，步骤(i)具有以下特征：

所述待测样本为组织、血液、脱落细胞，上皮细胞；和/或

所述待测样本来源于人或非人哺乳动物，较佳地来源于人；和/或

所述待测样本来源于HBV感染者或肝癌患者。

35 在另一优选例中，步骤(iii)具有以下特征：

所述探针为变性的单链DNA；和/或

在杂交液中加入接头封闭分子和标签封闭分子；和/或
所述接头封闭分子的序列如 SEQ ID NO: 8 所示；和/或
所述标签封闭分子的序列如 SEQ ID NO: 9 和 SEQ ID NO: 10 所示。

5 在另一优选例中，在步骤 (v) 中，将所述的扩增产物与固相载体上固定的测序探针进行杂交，进行固相桥式 PCR 扩增，形成测序簇；然后对所述测序簇用“边合成-边测序”法进行测序，从而得到与病毒整合方式有关核酸信息。

在另一优选例中，在步骤 (ii) 中，所述的文库构建为：对打断的基因组 DNA 进行末端修复，加入接头，对具有接头的片段进行扩增，获得的带有接头的扩增混合物即为样本文库。

10 在另一优选例中，所述的接头具有如 SEQ ID NO: 1 和 SEQ ID NO: 2 所示的序列；和/或，所构建的文库具有如 SEQ ID NO: 3 和 SEQ ID NO: 4 所示的标签序列。

在本发明的又一方面，本发明还提供了一种可用于本发明前面所述方法的试剂盒，所述试剂盒包括：

- (1) 第一容器以及位于容器内前面所述的核酸芯片，或本发明前面所述的探针；
- 15 (2) 第二容器以及位于容器内的用于构建样本文库的接头；
- (3) 第三容器以及位于容器内的接头封闭分子；
- (4) 第四容器以及位于容器内的标签封闭分子；
- (5) 检测说明书。

在另一优选例中，所述试剂盒还包括选自下组的试剂：

- 20 用于进行 PCR 扩增所需的试剂、
- 用于进行封闭反应所需的试剂、
- 用于进行杂交反应所需的试剂、
- 用于进行测序反应所需的试剂、
- 或其组合。

25 在本发明的另一方面，本发明还提供了一种可用于本发明前面所述方法的试剂盒，所述试剂盒包括：

前面所述的核酸芯片或前面所述的探针；接头；接头封闭分子；以及标签封闭分子。

30 根据本发明的实施例，在该试剂盒中，所包含的成分被设置在不同的容器中，由此，可以方便使用。根据本发明的实施例，该试剂盒还包括选自下组的试剂：用于进行 PCR 扩增所需的试剂、用于进行封闭反应所需的试剂、用于进行杂交反应所需的试剂、用于进行测序反应所需的试剂、或其组合。

35 在本发明的另一方面，本发明还提供了一种用于确定人体具有异常状态的系统。根据本发明的实施例，该系统包括：核酸序列信息接收器，所述核酸序列信息接收器接收人体样本的核酸序列信息；以及核酸序列信息分析器，所述核酸序列信息分析器与所述核酸序列信息接收器相连，并基于所述人体样本的核酸序列信息，确定所述人体是否具有异常状态。利用该系统，可以有效地实施根据本发明实施例的确定人体具有异常状态的方案，其

具有本发明实施例的确定人体具有异常状态的方法的全部优点，在此不再赘述。

另外，根据本发明上述实施例，上述确定人体具有异常状态的系统，还可以具有如下附加的技术特征：

根据本发明的一个实施例，所述核酸序列信息分析器内预存有选自下列的至少一种：

- 5 人体正常状态的基因组序列、病原体的基因组序列、正常人群的基因组序列。从而可以有效地对核酸序列进行分析，提高了用于确定人体具有异常状态的系统的效率。根据进一步的实施例，所述病原体为选自 HBV、HPV、EBV、幽门螺旋杆菌的至少一种。由此，可以有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。

- 10 根据本发明的一个实施例，进一步包括核酸序列检测装置，所述核酸序列检测装置与
所述核酸序列信息接收器相连，用于对所述人体样本进行核酸序列检测获得所述核酸序列
信息并输送至所述核酸序列信息接收器。由此，可以直接对核酸进行序列检测，并且输送
至核酸序列接收器，进而进行核酸序列分析，从而确定人体是否具有异常状态从而提高了
确定人体具有异常状态的效率。根据进一步的实施例，所述核酸序列检测装置借助第二代
15 测序技术或第三代测序技术。由此，可以高效地对人体样本进行核酸序列进行检测，并且
能够实现高通量深度测序。本发明的发明人发现基于第二代测序技术和第三代测序技术的高
效、高精度的性质，可以实现对人体样本核酸序列信息的高效、高精度检测，能够非常
灵敏地对人体样本中痕量的核酸进行检测。

- 20 根据本发明的一个实施例，进一步包括游离核酸捕获装置，所述游离核酸捕获装置与
所述核酸序列检测装置相连，其中，所述游离核酸捕获装置设置有探针，所述探针适于捕
获含有特定序列的核酸，并且将所述含有特定序列的核酸输送至所述核酸序列检测装置进
行核酸序列检测；或者所述探针适于除去含有特定序列的核酸，并且将经过所述除去的核
酸输送至所述核酸序列检测装置进行核酸序列检测。由此，对所述核酸进行测序检测之前，
利用探针捕获含有特定序列的核酸，然后对所述含有特定序列的核酸进行测序检测。由此，
25 可以通过探针，预先对进行核酸序列分析的核酸进行筛选，从而能够进一步提高确定人体
具有异常状态的方法的效率。根据进一步的实施例，所述探针对于选自下列的至少一种是
特异性的：HBV、HPV、EBV、幽门螺旋杆菌，其来源于前面所述的本发明的核酸探针集，
具有其全部优点，在此不再赘述。由此，可以有效地确定人体是否患有肝癌、宫颈癌、鼻
咽癌或胃癌。或者，可以通过探针除去具有特定序列的核酸，从而能够提高对剩余材料进
行序列检测的精度和准确性。根据本发明的具体示例，用于去除特定序列的核酸的探针可
30 以结合人体基因组中的共有序列，或者为可以结合人体基因组中甲基化位点的抗体或蛋白。
由此，可以进一步提高检测的精度和准确性。

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

35 附图说明

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

容易理解，其中：

- 图 1 是根据本发明一个实施例的确定人体具有异常状态的方法的流程示意图；
图 2 是根据本发明另一个实施例的确定人体具有异常状态的方法的流程示意图；
图 3 是根据本发明又一个实施例的确定人体具有异常状态的方法的流程示意图；
5 图 4 是根据本发明一个实施例的用于确定人体具有异常状态的系统的示意图；
图 5 是根据本发明另一个实施例的用于确定人体具有异常状态的系统的示意图；
图 6 是根据本发明又一个实施例的用于确定人体具有异常状态的系统的示意图；
图 7 是根据本发明又一个实施例的对 HBV 全基因组 PCR 扩增后的电泳检测结果；
图 8 是根据本发明又一个实施例的对 HBV 全长产物打断后的电泳检测结果；
10 图 9 是根据本发明又一个实施例的建库杂交后一种文库的片断大小检测结果；以及
图 10 是根据本发明又一个实施例的建库杂交另一种文库的片断大小检测结果。

发明详细描述

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

在本发明的描述中，需要说明的是，除非另有明确的规定和限定，术语“相连”、“连接”应做广义理解，例如，可以是固定连接，一体地连接，也可以是可拆卸连接；可以是机械连接或电连接，也可以是两个元件内部的连通；可以是直接相连，也可以通过中间媒介间接相连，对于本领域的普通技术人员而言，可以根据具体情况理解上述术语的具体含义。

本发明是基于本发明的下列发现而完成的：在人体样本的检测过程中，对于蛋白质的检测，通常由于实验条件的限制，无法直接对应其原位状态的信息，而核酸序列由于其自身性质比较稳定，因而在对于离体的人体样本进行核酸分析的结果可以直接对应其原位状态的信息，进而可以有效地对人体的状态进行分析。

下面参考附图，首先对根据本发明实施例的确定人体具有异常状态的方法进行详细描述。

参考图 1，根据本发明实施例的确定人体具有异常状态的方法包括以下步骤：

- 步骤 100：提供人体样本的核酸序列信息；以及
30 步骤 200：基于人体样本的核酸序列信息，确定人体是否具有异常状态。

根据该实施例的方法，通过对人体样本的核酸序列信息进行分析，可以根据核酸序列中所包含的信息确定人体是否具有异常状态。由于核酸序列信息与原位状态的核酸信息保持一致，因而可以有效地确定人体是否具有异常状态。

在本发明中，术语“人体”应作广义理解，其并不限于人，其可以是任何能够通过核酸信息预测异常状态的生命体。在本发明中，术语“核酸”可以是任何包含脱氧核糖核苷酸或者核糖核苷酸的聚合物，包括但不限于经过修饰的或者未经修饰的 DNA、RNA，其长

度不受任何特别限制。在本发明中，术语“核酸序列信息”，是指核酸序列所包含的所有信息，包括但不限于核酸的碱基序列、是否被修饰等信息。在本发明中，术语“人体样本”的含义不受特别限制，可以应用于本发明实施例的人体样本的类型，包括但不限于人体的细胞、组织、血液、体液、尿液、排泄物或其组合，更具体的示例包括血浆或血清。更进一步的示例中，采用血浆作为人体样本。本申请的发明人发现选择血浆作为研究样本，背景噪音会较小，检测结果精度高。本领域技术人员可以根据需要选择进行检测和分析的样本类型。由此，根据本发明的实施例的确定人体具有异常状态的方法能够应用于各种人体样本，并且能够根据不同的人体样本的特点，确定多种异常状态。

在本发明中，术语“异常状态”是指人体与正常状态不同的状态，包括但不限于生理状态、心理状态，例如病理状态。根据本发明的一个实施例，异常状态选自疾病的发生、疾病的发展阶段、疾病的疗效和预后的至少一种。由此，可以确定与人体密切相关的疾病的发生、发展、疗效以及预后，从而有利于制定有效的治疗方案。根据本发明进一步的实施例，疾病是肿瘤性疾病、免疫性疾病、遗传性疾病的至少一种。由此，根据本发明的实施例的确定人体异常状态的方法，能够有效地确定人体是否患有肿瘤性疾病、免疫性疾病、遗传性疾病。根据具体的示例，所述肿瘤性疾病是选自肺癌、肝癌、胃癌、食管癌、结直肠癌、胰腺癌、乳腺癌、膀胱癌、肾癌、卵巢癌、宫颈癌、甲状腺癌、鼻咽癌、脑胶质瘤的至少一种。根据更进一步的示例，如果所述核酸序列信息包含选自下列至少一种的核酸片段序列：HBV、HPV、EBV、幽门螺旋杆菌，则确定所述人体患有宫颈癌、肝癌、鼻咽癌、胃癌的至少一种。由此，根据本发明的实施例的确定人体具有异常状态的方法，可以基于所检测的核酸序列信息，有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。

根据本发明的实施例，人体样本的核酸序列信息的来源并不受特别限制。根据本发明的一个实施例，人体样本的核酸序列信息是基于对人体样本进行检测而获得的。根据本发明的实施例，检测人体样本而获得人体样本的核酸序列信息的方法并不受特别限制，可以直接对人体样本进行核酸序列测序分析而获得，也可以通过其他方法例如质谱等。根据本发明的一些示例，可以基于对人体样本直接进行核酸序列检测而获得人体样本的核酸序列信息。即，如图2所示，根据本发明实施例，进一步包括步骤300：对核酸序列进行检测。由此，可以通过核酸序列检测方法，容易地获得人体样本的核酸序列信息，从而提高了确定人体具有异常状态的效率。根据更具体的示例，可以借助第二代测序技术或第三代测序技术进行核酸序列检测。由此，以边合成边测序方法为代表的第二代测序技术，以及以单分子测序为代表的第三代测序方法可以高效地对人体样本进行核酸序列进行检测，并且能够实现高通量深度测序。本申请的发明人发现基于第二代测序技术和第三代测序技术的高效、高精度的性质，可以实现对人体样本核酸序列信息的高效、高精度检测，能够非常灵敏地对人体样本中痕量的核酸进行检测，从而进一步提高确定人体存在异常状态的效率。这里所使用的术语“高通量”是指可以同时大量的核酸进行测序检测，术语“深度”是指可以对核酸进行重复多次检测，例如在实施例1中可以进行100轮测序检测，当然根据样本的不同，重复的次数也可以根据需要进行选择。当然，本领域技术人员能够预见的是，

未来可以采用其他更先进的测序技术。目前可以利用的第二代测序技术包括但不限于 Illumina/HiSeq 2000、Roche/454、ABI/SOLiD。在本发明的一个实施例中，所采用的测序方法为 Illumina/HiSeq 2000。

5 根据本发明的实施例，对核酸序列信息进行分析，从而确定异常状态的方法，不受特别限制。可以是在获得核酸序列信息后，与人体的正常基因组信息或者病原体的基因组信息进行比对，得到比对结果后，进行判断人体是否具有异常状态。也可以基于，核酸序列在样本中的含量来判断，人体的异常状态。根据本发明的实施例，可以通过分析人体样本中，游离核酸的含量，来确定人体是否存在异常状态。如实施例 1 所示，本发明的发明人发现，在癌症患者的外周血中，游离核酸的含量远高于正常人外周血中游离核酸的含量（为至少大约 10 倍）。根据本发明另外的实施例，可以分析人体样本中，游离核酸序列中是否存在突变位点或者经修饰的位点例如甲基化位点，从而判断个体是否存在某些特定的异常状态。具体地，可以采用常用的比对软件进行操作，根据本发明的一个示例，采用 SOAP 软件包进行比对分析，由此，能够高效地对核酸序列信息进行分析，并得到准确和精确的结果。

15 根据本发明的实施例，可以用于本发明实施例的确定人体具有异常状态的方法的核酸类型不受特别限制，根据本发明的一个实施例，所采用的核酸序列信息包括人体样本中游离核酸的序列信息。由此，可以根据游离核酸的序列信息，对人体的异常状态进行确定。并且基于游离核酸的序列信息与人体正常的核酸序列或者病原体的核酸序列进行比对分析后，可以获得多种人体的异常状态。为了方便理解，下面对游离核酸进行详细描述。

20 在本发明中，所使用的术语“游离核酸”是指在细胞外的游离状态的核酸，可以是 DNA、RNA 或者其他类型的核酸。本发明的发明人发现，正常状态下，会有少量的核酸由于代谢而进入到外周血中而成为游离核酸，而在异常状态下，例如癌症病人体内，游离核酸的含量大大高于正常状态下的游离核酸含量。发明人发现，在癌症病人体内，游离核酸（如游离 DNA）的含量取决于肿瘤的生物特性，即与肿瘤细胞的恶性程度、侵袭程度、是否发生转移、疾病进程等相关。因而，可以通过分析人体样本中游离核酸的含量，来确定人体是否存在异常状态，例如患有肿瘤性疾病，以及确定肿瘤性疾病的进展阶段，侵袭程度等。从而，为选择有效的治疗方案，提供有利的信息。另外，对于某些由于病原体引起的疾病，在人体样本中，可能会存在这样的游离核酸，即，它属于病原体基因组序列的一部分，这样可以确定人体已经被这些病原体侵染并且处于疾病的发生阶段。甚至有些个体样本中的游离核酸既包含病原体基因组的一部分序列，也包含人体基因组的某些序列，这样可以确定这些病原体基因组已经与人体基因组发生了整合重组。由此，可以判断个体的异常状态的阶段。根据本发明的一个实施例，人体样本中游离核酸的序列信息是通过除去人体样本中的细胞后，进行测序检测而获得的。由此，可以提高游离核酸的测序检测的精度和准确度，从而能够进一步有效地确定人体具有异常状态。

35 另外，根据本发明的实施例，对所述核酸进行测序检测之前，还可以包括步骤 400，如图 3 所示，即利用探针捕获含有特定序列的核酸，然后对含有特定序列的核酸进行测序检

测。由此，可以通过探针，预先对进行核酸序列分析的核酸进行筛选，从而能够进一步提高确定人体具有异常状态的方法的效率。另外，本发明的发明人惊奇地发现，通过此步骤，可以提高检测到与人体基因组发生整合的游离核酸的效率。

本领域技术人员能够理解，所使用的探针类型可以根据检测的目的进行改变，即可以根据所期望的特定序列来选择所使用的探针类型。根据本发明的实施例，特定序列可以是外源核酸序列，也可以是含有突变位点的人基因组部分序列，也可以是含有修饰位点例如甲基化的人基因组部分序列。根据进一步的实施例，所述探针对于选自下列的至少一种是特异性的：HBV、HPV、EBV、幽门螺旋杆菌，其可以以包括多个探针的核酸探针集的形式提供，并且其具有以下特征：（1）每个探针上具有 1 个或多个生物素标记的 dNTP；和/或（2）生物素标记的 dNTP 在核酸探针集中的丰度为 1: 6-1: 2；和/或（3）核酸探针集的全部核酸序列覆盖对应选自 HBV、HPV、EBV 和幽门螺旋杆菌的至少一种病毒的基因组序列的 70%-100%。在另一优选例中，本发明的核酸探针集具有 1-20000 个核酸探针；较佳地，核酸探针集具有 1000-5000 个核酸探针；更佳地，核酸探针集具有 2500 个核酸探针。在另一优选例中，生物素标记的 dNTP 在核酸探针集中的丰度为 1: 4。在另一优选例中，核酸探针集中，探针之间具有部分重叠。在另一优选例中，核酸探针集（在本文中有时也称为“探针集”）中的探针长度为 100-500 bp；较佳地，探针长度为 200-300 bp；更佳地，探针长度为 250 bp。在另一优选例中，探针是以病毒基因组作为模板，PCR 法扩增获得的，较佳地，扩增模板为乙型肝炎病毒（HBV）基因组、丙型肝炎病毒（HCV）基因组、艾滋病病毒（HIV）基因组、乳头瘤病毒（HPV）基因组，或其组合；更佳地，扩增模板为 B 型 HBV 基因组和/或 C 型 HBV 基因组。由此，可以有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。

另外，根据本发明的实施例，可以在对核酸进行测序之前，通过使用特定的探针除去某些含有特定序列的核酸，例如人类基因组中的共有序列，从而提高确定人类具有异常状态的方法的准确性。具体地，可以在对所述核酸进行测序检测之前，利用探针去除含有特定序列的核酸，然后对所述去除后剩余的核酸进行测序检测。这些用于除去含有特定序列的核酸的探针如同用于捕获含有特定序列的核酸的探针一样，它们的类型不受特别限制，可以为核酸、蛋白质以及任何小分子，只要其能够特异性地结合特定的序列即可。另外，为了除去含有特定序列的核酸，所采用的探针能够结合人体基因组中的共有序列，或者也可以是能够结合人体基因组中甲基化位点的抗体或者蛋白。可以根据具体的情况对所采用的探针的类型，以及是需要对样本进行核酸捕获，还是需要特异性去除。

在本发明的另一方面，本发明提供了一种用于确定人体具有异常状态的系统，其可以有效地实施上述根据本发明实施例的确定人体具有异常状态的方法。根据本发明的实施例，参考图 4，该系统包括：核酸序列信息接收器 500、以及核酸序列信息分析器 600。其中，核酸序列信息接收器 500 接收人体样本的核酸序列信息，核酸序列信息分析器 600 与核酸序列信息接收器 500 相连，并基于人体样本的核酸序列信息，确定人体是否具有异常状态。利用该系统，可以有效地实施根据本发明实施例的确定人体具有异常状态的方案，其具有

本发明实施例的确定人体具有异常状态的方法的全部优点，在此不再赘述。

如前所述，根据本发明实施例，对核酸序列信息进行分析的方法不受特别限制，根据具体地示例，可以通过将核酸序列信息与人体正常状态的基因组序列、病原体的基因组序列、正常人群的基因组序列进行比对而确定人体是否具有异常状态。人体正常状态的基因组序列、病原体的基因组序列、正常人群的基因组序列的存放位置不受特别限制，可以存储 5 在远程的数据库中。根据本发明的一个实施例，核酸序列信息分析器 500 内可以预存有选自下列的至少一种：人体正常状态的基因组序列、病原体的基因组序列、正常人群的基因组序列。由此，从而可以有效地对核酸序列进行分析，提高了用于确定人体具有异常状态的系统的效率。根据进一步的实施例，所述病原体为选自 HBV、HPV、EBV、幽门螺旋杆菌的至少一种。由此，可以有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。通过 10 将核酸序列信息与人体正常状态的基因组序列（即将同一身体在不同状态下的基因组信息）进行比对，可以确定人体在一段时间内的状态变化。另外，通过将人体样本的核酸序列信息与正常人群的基因组序列信息进行比对，可以获知与正常人相比的异常状态。

根据本发明的实施例，进行分析的核酸序列信息的来源不受特别限制。根据本发明的一个实施例，参考图 5，本发明的用于确定人体具有异常状态的系统可以进一步包括核酸序 15 列检测装置 700。该核酸序列检测装置 700 与核酸序列信息接收器 500 相连，用于对述人体样本进行核酸序列检测获得核酸序列信息并输送至核酸序列信息接收器 500，进而进行分析和确定人体是否存在异常状态。由此，可以直接对核酸进行序列检测，并且输送至核酸序列接收器 500，进而进行核酸序列分析，从而确定人体是否具有异常状态从而提高了确定人 20 体具有异常状态的效率。根据进一步的实施例，所述核酸序列检测装置借助第二代测序技术或第三代测序技术。由此，可以高效地对人体样本进行核酸序列进行检测，并且能够实现高通量深度测序。本发明的发明人发现基于第二代测序技术和第三代测序技术的高效、高精度的性质，可以实现对人体样本核酸序列信息的高效、高精度检测，能够非常灵敏地对人体样本中痕量的核酸进行检测。

根据本发明的一个实施例，参考图 6，本发明的用于确定人体具有异常状态的系统还可以进一步包括游离核酸捕获装置 800，该游离核酸捕获装置 800 与核酸序列检测装置 700 相 25 连，并且游离核酸捕获装置 800 设置有探针，这些探针适于捕获含有特定序列的核酸，并且将含有特定序列的核酸输送至核酸序列检测装置 700 进行核酸序列检测。由此，对所述核酸进行测序检测之前，利用探针捕获含有特定序列的核酸，然后对所述含有特定序列的核酸进行测序检测。由此，可以通过探针，预先对进行核酸序列分析的核酸进行筛选，从而能够进一步提高确定人体具有异常状态的方法的效率。另外，本发明的发明人惊奇地发现，通过此步骤，可以提高检测到与人体基因组发生整合的游离核酸的效率。 30

本领域技术人员能够理解，所使用的探针类型可以根据检测的目的进行选择，即可以根据所期望的特定序列来选择所使用的探针类型。根据本发明的实施例，特定序列可以是 35 外源核酸序列，也可以是含有突变位点的人基因组部分序列，也可以是含有修饰位点例如甲基化的人基因组部分序列。根据进一步的实施例，所述探针对于选自下列的至少一种是

特异性的：HBV、HPV、EBV、幽门螺旋杆菌，其可以以包括多个探针的核酸探针集的形式提供，并且其具有以下特征：（1）每个探针上具有 1 个或多个生物素标记的 dNTP；和/或（2）生物素标记的 dNTP 在核酸探针集中的丰度为 1: 6-1: 2；和/或（3）核酸探针集的全部核酸序列覆盖对应选自 HBV、HPV、EBV 和幽门螺旋杆菌的至少一种病毒的基因组序列的 70%-100%。在另一优选例中，本发明的核酸探针集具有 1-20000 个核酸探针；更佳地，核酸探针集具有 1000-5000 个核酸探针；更佳地，核酸探针集具有 2500 个核酸探针。在另一优选例中，生物素标记的 dNTP 在核酸探针集中的丰度为 1: 4。在另一优选例中，核酸探针集中，探针之间具有部分重叠。在另一优选例中，核酸探针集（在本文中有时也称为“探针集”）中的探针长度为 100-500 bp；更佳地，探针长度为 200-300 bp；更佳地，探针长度为 250 bp。在另一优选例中，探针是以病毒基因组作为模板，PCR 法扩增获得的，更佳地，扩增模板为乙型肝炎病毒（HBV）基因组、丙型肝炎病毒（HCV）基因组、艾滋病病毒（HIV）基因组、乳头瘤病毒（HPV）基因组，或其组合；更佳地，扩增模板为 B 型 HBV 基因组和/或 C 型 HBV 基因组。由此，可以有效地确定人体是否患有肝癌、宫颈癌、鼻咽癌或胃癌。根据本发明的具体实施例，可以采用 HPV 的 E1 基因区域的特异性序列作为探针。由此，能够高效准确地确定患者体内的 HPV 是否已经与患者体内的基因组发生整合，从而判断个体是否患有宫颈癌。根据本发明的一些具体示例，可以采用 HPV 的全长序列作为探针，由此，能够高效地准确地确定患者体内的 HPV 是否已经与患者体内的基因组发生整合，从而判断个体宫颈病变程度。另外，根据本发明的具体示例，可以采用 HBV 的 X 基因和/或 C 基因中的特异性序列作为探针，由此，能够高效地准确地确定患者体内的 HBV 是否已经与患者体内的基因组发生整合，从而判断个体是否患有肝癌。根据本发明的一些实施例，可以采用 HBV 的全长基因区域的特异性序列作为探针，由此，能够高效地准确地确定患者体内的 HBV 是否已经与患者体内的基因组发生整合，从而判断个体肝炎病变程度。

另外，根据本发明的实施例，还可以采用这样的探针，其能够除去含有特定序列的核酸，从而可以在对核酸进行测序之前，通过使用特定的探针除去某些含有特定序列的核酸，例如人类基因组中的共有序列，从而提高确定人类具有异常状态的方法的准确性。具体地，可以在对所述核酸进行测序检测之前，利用探针去除含有特定序列的核酸，然后对所述去除后剩余的核酸进行测序检测。这些用于除去含有特定序列的核酸的探针如同用于捕获含有特定序列的核酸的探针一样，它们的类型不受特别限制，可以为核酸、蛋白质以及任何小分子，只要其能够特异性地结合特定的序列即可。另外，为了除去含有特定序列的核酸，所采用的探针能够结合人体基因组中的共有序列，或者也可以是能够结合人体基因组中甲基化位点的抗体或者蛋白。可以根据具体的情况对所采用的探针的类型，以及是需要对样本进行核酸捕获，还是需要特异性去除。

为了方便理解，下面提供具体的实施例，对本发明的技术方案进行解释，需要说明的是，这些实施例仅仅是为了说明目的，而不以任何方式限制本发明的范围。除非特别说明，实施例中未注明具体条件的，均为按照常规条件或制造商建议的条件进行。下列实施例中，

所用试剂或仪器未注明生产厂商的，均为可以通过市购获得的常规产品。所使用的测序用的接头和标签序列(Index)来源于Illumina公司的Multiplexing Sample Preparation Oligonutide Kit。

5 实施例 1:

1. 样本文库制备

1.1 样本来源

样品的来源为同一患者的肝癌组织，且此患者肝癌组织有全基因组测序信息。

1.2 样本文库制备

10 文库构建按照 Illumina 公司的标准文库制备流程说明书 (Paired-End Sample Preparation Guide) 进行构建，具体方法如下：

采用 Covaris s2 打断基因组 DNA，末端补平修复，末端加 A，加入接头，建库过程所用的接头序列为：

5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3' (SEQ ID NO: 1)；

15 5'-TACTCTTTCCCTACACGACGCTCTTCCGATCT-3' (SEQ ID NO: 2)

对加入接头的片段进行 PCR，得到样本文库，所构建的文库带有 Index 标签序列，其中 Index 序列如下：

5'-CAAGCAGAAGACGGCATAACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' (SEQ ID NO: 3)；

20 5'-CAAGCAGAAGACGGCATAACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' (SEQ ID NO: 4)。

将样本分别做成两种片段大小文库，对构建文库片段进行片段大小检测，主带为 170 bp 左右和 800 bp 左右。

25 2. 制备 HBV 探针

2.1 引物设计

本实施例中，所设计的引物为：

P1: TTTTTCACCTCTGCCTAATCA (SEQ ID NO: 5)；

P2: AAAAAGTTGCATGGTGCTGG (SEQ ID NO: 6)

30 2.2 PCR 反应体系

PCR 反应体系见表 1。

表 1

LA Taq	0.25 μ l
10 \times LA buffer (Mg+2.00 mM)	2.5 μ l

dNTP (2.5 mM)	4 μ l
P1 (10 μ M)	1.5 μ l
P2 (10 μ M)	1.5 μ l
HBV 病毒基因组 (1 ng/ μ l)	3 μ l
H ₂ O	12.25 μ l
总体积	25 μ l

(注: dNTP 中 biotin-dNTP 与普通 dNTP 的比例是 1: 4, 总浓度为 2.5 mM)

2.3 PCR 反应条件

PCR 反应在 AB-9700PCR 仪上进行, 反应程序见表 2。

表 2

温度	时间	循环数
94 $^{\circ}$ C	3 min	1
94 $^{\circ}$ C	30 s	35
56 $^{\circ}$ C	50 s	
68 $^{\circ}$ C	150 s	
72 $^{\circ}$ C	10 min	
4 $^{\circ}$ C	保持	

5 2.4 PCR 产物纯化及电泳检测

反应结束后, 采用 1%琼脂糖凝胶电泳检测 PCR 产物, 并用 1.2-1.5 倍体积 AMPURE BEADS 纯化, 采用 80 μ l 水溶解。然后采用 250MinElute PCR Purification Kit 纯化, 采用 60 μ l 水溶解。其中, PCR 产物的 1%琼脂糖凝胶电泳检测结果见图 7, 结果表明, 扩增并纯化出了大小约 3.2 K 的 HBV 的片段。

10 其中, HBV 基因组序列 (Hepatitis B virus serotype adr, complete genome) 如下:

CTCGAGGACTGGGGACCCTGCACCGAACATGGAGAGCACAACATCAGGATTCCT
 AGGACCCCTGCTCGTGTTACAGGCGGGGTTTTCTTGTGACAAGAATCCTCACAATA
 CCACAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTCTAGGGGGAGCACCCACGT
 15 GTCCTGGCCAAAATTCGCAGTCCCAACCTCCAATCACTCACCAACCTCTTGTCTCC
 AATTTGTCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATCATATTCCTCTTCATCCTG
 CTGCTATGCCTCATCTTCTTGTGTTCTTCTGGACTACCAAGGTATGTTGCCCGTTTGT
 CCTCTACTTCCAGGAACATCAACTACCAGCACGGGACCATGCAAGACCTGCACGATTC
 CTGCTCAAGGAACCTCTATGTTTCCCTCTTGTGCTGTACAAAACCTTCGGACGGAAA
 CTGCACTTGTATTCCCATCCATCATCCTGGGCTTTCGCAAGATTCCTATGGGAGTGGG
 20 CCTCAGTCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTTCAGTGGTTCGCAGGG
 CTTTCCCCACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCAAGTCTG
 TACAACATCTTGAGTCCCTTTTTACCTCTATTACCAATTTTCTTTTGTCTTTGGGTATACA

TTTGAACCCTAATAAAAACCAAACGTTGGGGCTACTCCCTTAACTTCATGGGATATGTAA
TTGGAAGTTGGGGTACTTTACCACAGGAACATATTGTATTA AAAACTCAAGCAATGTTTT
CGAAA ACTGCCTGTAAATAGACCTATTGATTGGAAAGTATGTC AAAGAATTGTGGGTCT
TTTGGGCTTTGCTGCCCTTTTACACAATGTGGCTATCCTGCCTTGATGCCTTTGTATGC
5 ATGTATAACAATCTAAGCAGGCTTTC ACTTTCTCGCCAACTTATAAGGCCTTTCTGTGTCA
ACAATACCTGCACCTTTACCCCGTTGCCCGGCAACGGTCAGGTCTCTGCCAAGTGTTT
GCTGACGCAACCCCACTGGATGGGGCTTGGCCATAGGCCATCGGCGCATGCGTGGA
CCTTTGTGGTTCCTCTGCCGATCCATACTGCGGAACTCCTAGCAGCTTGTTTTGCTCGC
GACCGGTCTGGAGCAAACTTATCGGGACTGACAACTCGGTGTCTCTCTCGGAAAT
10 ACACCTCCTTCCCATGGCTGCTCGGGTGTGCTGCCAACTGGATCCTGCGCGGGACGTC
CTTTGTCTACGTCCCGTCGGCGCTGAATCCCGCGGACGACCCGTCTCGGGGCCGTTTG
GGCCTCTACCGTCCCTTGCTTTCTCTGCCGTTCCAGCCGACCACGGGGCGCACCTCTCT
TTACGCGGTCTCCCGTCTGTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCAC
CTCTGCACGTCGCATGGAGACCACCGTGAACGGCCACCAGGTCTTGCCCAAGCTCTTA
15 CATAAGAGGACTCTTGGACTCTCAGCAATGTCAACAACCGACCTTGAGGCATACTTCA
AAGACTGTTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAAAGGTCT
TTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGTTCCACCAGCACCATGCAACTTTTTT
ACCTCTGCCTAATCATCTCATGTTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTG
GGTGGCTTTGGGGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCTGTGGAGTT
20 ACTCTCTTTTTTGCTTCTGACTTCTTTCCTTCTATTTCGAGATCTCCTCGACACCGCCTC
TGCTCTGTATCGGGAGGCCTTAGAGTCTCCGGAACATTGTTACCTCACCATACAGCAC
TCAGGCAAGCTATTCTGTGTTGGGGTGAGTTGATGAATTTGGCCACCTGGGTGGGAAG
TAATTTGGAAGACCCAGCATCCAGGGAATTAGTAGTCAGCTATGTCAATGTTAATATGG
GCCTAAAAATCAGACA ACTATTGTGGTTTCATATTCCTGTCTTACTTTTGGAAGAGAA
25 ACTGTTCTTGAGTATTTGGTGTCTTTTGGAGTGTGGATTTCGCACTCCTCCCGCTTACAG
ACCACCAAATGCCCTATCTTATCAACACTTCCGGAACTACTGTTGTTAGACGACGAG
GCAGGTCCCCTAGAAGAAGA ACTCCCTCGCCTCGCAGACGAAGGTCTCAATCGCCGC
GTCGCAGAAGATCTCAATCTCGGGAATCTCAATGTTAGTATCCCTTGGA CTATAAGGT
GGGAACTTTACTGGGCTTTATTCTTCTACTGTACCTGTCCTTAATCCTGAGTGGCAAA
30 CTCCTCCTTTTCTAACATTCATTTACAGGAGGACATTATTAATAGATGTCAACAATATG
TGGGCCCTCTTACAGTTAATGAAAAAAGGAGATTAAAATTAATTATGCCTGCTAGGTTT
TATCCTAACCTTACCAAATATTTGCCCTTGGATAAAGGCATTAAACCTTATTATCCTGAA
CATGCAGTTAATCACTTCAA AACTAGGCATTATTTACATACTCTGTGGAAGGCTGGC
ATTCTATATAAAAGAGAACTACACGCAGCGCTTCATTTTGTGGGTACCATATTCTTG
35 GGAACAAGAGCTACGGCATGGGAGGTTGGTCTTCCAAACCTCGACAAGGCATGGGGA
CGAATCTTTCTGTTCCCAATCCTCTGGGATTCTTTCCCGATCACCAGTTGGACCCTGCG

TTCGGAGCCA ACTCAAACAATCCAGATTGGGACTTCAACCCCAACAAGGATCACTGGC
 CAGAGGCAATCAAGGTAGGAGCGGGAGACTTCGGGCCAGGGTTCACCCCACCACACG
 GCGGTCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGACAACAGTGCCAGCAGC
 GCCTCCTCCTGTTTCCACCAATCGGCAGTCAGGAAGACAGCCTACTCCCATCTCTCCA
 5 CCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGGA ACTCCACAACATTCCACCAAG
 CTCTGCTAGATCCCAGAGTGAGGGGCCTATATTTTCCTGCTGGTGGCTCCAGTTCCGGA
 ACAGTAAACCCTGTTCCGACTACTGTCTCACCCATATCGTCAATCTT (SEQ ID NO: 7)。

2.5 PCR 产物片段化

10 将经过纯化的 PCR 产物全部转移至 Covaris 打断小管并补加 TE 缓冲液至总体积为 80 μl(Nanodrop 检测其总量为 5 μg), Covaris S2 仪器 (基因有限公司) 进行打断, 由此, 获得片段化产物, 其中打断条件见表 3。

表 3

负载比	强度	循环/脉冲	打断时间
10%	5	200	30s × 7 个循环

2.6 片段化产物电泳检测

15 2%琼脂糖凝胶电泳检测片段化产物的大小, 结果见图 8, 结果显示片段化产物的主带在 250-300 bp, 表明所获得的片段化产物即可用作杂交的探针。

2.7 探针保存

使用 MinElute PCR Purification Kit 纯化片段产物, 溶于 40 μl 缓冲液中, 用 Nanodrop 仪检测探针 DNA 的浓度, 使得探针的浓度为 120 ng/μl 左右。得到的探针可以保存在-20℃或-80℃。

20

3. HBV 探针与样本文库进行杂交

3.1 探针变性

探针使用前必须 95℃ 变性 10 分钟, 然后迅速放于冰上冷却形成单链。

25 3.2 选用已确定的整合文库, 文库用量为 1 μg, 探针用量为 600 ng (Nanodrop 定量), 加入接头封闭分子, 接头封闭分子的量与文库量的比值为 1 nmol: 1 μg, 标签封闭分子与文库量的比值为 1 nmol: 1 μg。

接头封闭分子序列为:

5'-AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTTCCG
 ATCT-3' (SEQ ID NO: 8);

30

标签封闭分子序列为:

5'-AAGCAGAAGACGGCATAACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCT
 CTTCCGATCT-3' (SEQ ID NO: 9);

5'-AAGCAGAAGACGGCATAACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCT

CTTCCGATCT-3' (SEQ ID NO: 10)。

5 在一个 1.5 mL 的 EP 管中加入 1 μg 的待杂交文库，1 nmol 接头封闭分子，1 nmol 标签封闭分子，5 μg Cot DNA。盖好管盖，用干净的 50 ml 注射器针在分装的 EP 管盖上戳一个孔，然后置于 60 $^{\circ}\text{C}$ 旋蒸仪中蒸干。使用新的离心管管盖替换戳孔的管盖，并做好标记。EP 管中分别加入 EZ 杂交系统中的两种试剂：2 \times SC Hybridization Buffer 杂交缓冲液 7.5 μl 和 1 \times SC Hybridization Component A 3 μl ，然后 95 $^{\circ}\text{C}$ 变性 10 分钟，在上述杂交混合物加入自制探针 600 ng，探针体积共 5 μl 。震荡混匀后置于离心机全速离心 10 秒，并将样品全部转移到 200 μl PCR 小管中。

杂交混合物中含有的成分见表 4。

10

表 4

文库蒸干混合物	1 μg (文库) + 5 μg (Cot DNA) + 2 nmol 封闭分子
2X SC 杂交缓冲液	7.5 μl
SC 杂交组合物 A (SC Hybridization Component A)	3 μl
探针	600 ng
总体积	15 μl

将 200 μl PCR 小管放置于 PCR 仪上，47 $^{\circ}\text{C}$ 条件下杂交 24 h。

4. 杂交后洗脱

4.1 准备链霉亲和素磁珠(Invitrogen M280)

15 提前从冰箱中拿出链霉亲和素磁珠；漩涡震荡磁珠 1 min，使其充分混匀；在 1.5 mL 的 EP 管中加入 100 μl 磁珠；将 EP 管置于磁力架上至液体澄清，用移液器小心地去除上清；保持 EP 管在磁力架上，加入 200 μl (2 倍体积) 的结合缓冲液 (购于 Agilent 公司)；从磁力架上取下 EP 管，漩涡震荡 10 s，使其混匀；将 EP 管重新放回磁力架至液体澄清，用移液器小心地去除上清；重复清洗两次；用 100 μl 的 Agilent 结合缓冲液 r 悬浮磁珠；将其转入 0.2 ml 的小管中；用磁力架结合磁珠 (将小管靠到磁力架上)，直到液体澄清，用移液器小心去除上清；现在磁珠可以用来结合捕获的 DNA 了。

4.2 将捕获到的 DNA 结合到链霉素磁珠上

25 将杂交混合物吸出来，加到上步准备好的磁珠中；用移液器吹打 10 次混匀；将小管放在 PCR 仪上 47 $^{\circ}\text{C}$ 孵育 45 min (每隔 15 min 拿出来漩涡震荡 3 s 以防止磁珠沉淀)；孵育 45 min 后，将混合物从 0.2 mL 的小管中转入 1.5 mL 的 EP 管中。

4.3 洗涤结合了捕获 DNA 的链霉亲和素磁珠

- 1) 将 EP 管置于磁力架上至液体澄清，用移液器小心地去除上清；
- 2) 加 100 μL 预热到 47 $^{\circ}\text{C}$ 的 1 \times 清洗缓冲液 I；
- 3) 漩涡震荡 10 s，使其混匀；

- 4) 将EP管置于磁力架上至液体澄清,用移液器小心地去除上清;
- 5) 从磁力架上取下EP管,加入200 μl 预热到47°C的1×严谨清洗缓冲液,用移液器吹打混匀10次(该步操作应迅速以尽量使管中液体温度不低于47°C);
- 6) 47°C孵育5 min;
- 5 7) 重复步骤5)-7),总共用1×严谨清洗缓冲液洗两次;
- 8) 将EP管置于磁力架上至液体澄清,用移液器小心地去除上清;
- 9) 加200 μL室温下放置的1×清洗缓冲液I(不用47°C预热的),漩涡震荡2 min,使其混匀;
- 10) 将EP管置于磁力架上至液体澄清,用移液器小心地去除上清;
- 10 11) 加200 μL室温下放置的1×清洗缓冲液II,漩涡震荡1 min,使其混匀;
- 12) 将EP管置于磁力架上至液体澄清,用移液器小心的去除上清;
- 13) 加200 μL室温下放置的1×清洗缓冲液III,漩涡震荡30 s,使其混匀;
- 14) 将EP管置于磁力架上至液体澄清,用移液器小心地去除上清;
- 15) 从磁力架上取下EP管,加入76 μl超纯水(不用将DNA从磁珠上洗脱下来,可以直接进行PCR,取样35 μl进行后面的PCR反应)。

5. PCR 反应

预先从-20°C保存的试剂盒中取出 PFX 聚合酶(购于 Invirtogen 公司), PFX 反应缓冲液(10×), dNTP(10 mM)。引物序列为:

- 20 PCR Flowcell-Primer F (10 pm/μl):
AATGATACGGCGACCACCGAGATC (SEQ ID NO: 11);
- PCR Flowcell-Primer R (10 pm/μl):
CAAGCAGAAGACGGCATA CGA (SEQ ID NO: 12)。

在PCR小管上,每孔按照表格5配置PCR反应体系。

25

表 5

DNA	35 μl
PFX 酶	1 μl
PFX 反应缓冲液(10*)	5 μl
MgSO ₄	2 μl
dNTP(10 mM)	2 μl
Flowcell-primer-F1 (10 pm/μl)	2.5 μl
Flowcell-primer-R1 (10 pm/μl)	2.5 μl
总体积	50 μl

反应条件见表6。

表 6

温度	时间	循环
94℃	2 min	1
94℃	15 s	14
58℃	30 s	
72℃	30 s	
72℃	5 min	1
4℃	保持	

PCR 结束后，每个样品都用 1.5 倍体积的 Amprue Beads 纯化，回收的 PCR 产物溶于 30 μl 超纯水中，Nanodrop1000 测浓度。

5 6. PCR 产物上机测序

上述纯化后的 PCR 产物经 2100 Bioanalyzer (Agilent) 确定大小及插入片段大小见图 9 和图 10，纯化产物大小分别为 271 bp 和 876 bp，QPCR 精确定量后上机测序。在本实施例中，上机测序按照 Illumina/Solexa 官方公布的 c-Bot 和 HISEQ2000Hiseq 2000 说明书进行操作。

10 7. 信息分析

将下机数据除去重复和被接头污染的 reads，统计下机数据的基本信息(文库长度; reads 长度; reads 条数; 碱基数; 重复率等); 分别截取 PE 的两条 reads 前面的 50 bp 碱基，形成一对长为 50 bp 新 reads，即 PE50 reads。将新的 PE50 reads 运用 soap 比对软件(-r 1 -v 2) 分别与人的参考序列 (hg19) 和 HBV 各种参考序列进行比对，从比对结果中挑选出一条 read 比到人的参考序列并且另一条比对到 HBV 参考序列的一对 reads; 这样的 reads 很有可能跨

过 HBV 插入的位点; 统计这部分 reads 比对信息，找到在人类基因组的插入热点区域。杂交结果见表 7。

表 7

样本	L-170	L-800	Genome
测序总读数	6043804	3297107	873813333
断点处总读数	354	112	556
段点位置	Chr19:30315005, 30315366 Chr14:96085140	Chr19:30315005, 303153661 Chr14:96085140	Chr19:30314586-30315653 Chr14:96085100-96085700

表 7 的结果为采用上机数据得到结果，样本 L-170, L-800, Genome 均来自同一肝癌样

本, L-170 为插入片段 170 bp, L-800 为 800bp 文库, genome 为全基因组测序。从表 7 中可以得出自制探针对于捕获基因片段的准确性, 以及片段长度的影响。通过本发明的方法完全可以得到稳定以及可靠的位点, 且所需数据量仅为全基因组测序数据的 1%左右。

5 实施例 2: 对宫颈癌患者外周血样品进行全基因组分析

1、DNA 提取及测序

根据常规方法, 对宫颈癌患者进行静脉取血, 得到宫颈癌患者的外周血样本, 通过离心得到血浆样本。按照 Tiangen Micro Kit (DP316) 微量基因组操作流程从血浆样本提取 DNA, 分别用 Qubit (Invitrogen, the Quant-iT™ dsDNA HS Assay Kit) 定量, 所提取的 DNA 10 总量分别为 5~50 ng。

将所提取到的 DNA, 分别按照制造商提供的标准建库规程 (参见 <http://www.illumina.com/>提供的 Illumina 标准建库说明书) 建立 DNA 文库。简言之, 在 DNA 分子两端加上测序用的接头, 并被加上不同的标签序列, 然后与测序芯片表面互补接头杂交, 使核酸分子成簇生长, 然后在 Illumina HiSeq 2000 上通过 100 轮深度测序循环, 得到长度 15 为 100bp 的 DNA 片段序列。本实施例中, 对于获自肿瘤病人外周血的 DNA 样本分批按照制造商提供的操作说明书 (参见 Illumina 官方公布说明书) 进行上机测序操作。

2、数据分析

根据制造商 Illumina 提供的 Pipeline 操作说明书 (参见 <http://www.illumina.com/>提供的 Pipeline 方法说明书), 将步骤 DNA 测序部分中测得的序列信息经过图形转化获得测序序列 20 信息, 去掉测序质量低的序列之后最终可以获得针对 NCBI 版本 36 的人类基因组参考序列的 ELAND 比对结果。

将获得的数据使用 SOAP 软件包进行比对分析, 使用两个末端测序信息进行比对时去除两个末端均比对至人基因组的序列, 保留其中一条链比对至人基因组的序列, 将另一末端序列比对至 HPV 基因组序列中, 获得 HPV 在人基因组重组信息, 包括人基因组中重组 25 位置以及 HPV 类型。

3、分析结果:

根据数据分析部分中的数据分析流程, 通过使用高通量测序平台对宫颈癌样品进行深度测序以及数据分析, 共检测到 45 个超过 10 条测序序列支持的 HPV 整合基因, 发生整合 HPV 区域均为 E1 区, 发生整合的 HPV 型别为 HPV16 型。

30 表 8、检测出来的支持数超过 10 的 HPV 整合情况

支持数	相关整合基因	染色体	相关样本号	HPV 基因区域	HPV 型别	基因 ID
47	DPP6	chr7	hpv-9,hpv-10,hpv-11	E1	HPV16	1804
40	HIP1	chr7	hpv-9,hpv-10,hpv-11	E1	HPV16	3092
39	BCAS3	chr17	hpv-9,hpv-10,hpv-11	E1	HPV16	54828

38	ARHGEF10L	chr1	hpv-9, hpv-10, hpv-11	E1	HPV16	55160
38	TMCO7	chr16	hpv-9, hpv-10, hpv-11	E1	HPV16	79613
32	PPEF1	chrX	hpv-9, hpv-10, hpv-11	E1	HPV16	5475
31	PRKD2	chr19	hpv-9, hpv-10, hpv-11	E1	HPV16	25865
28	UNC13A	chr19	hpv-9, hpv-10, hpv-11	E1	HPV16	23025
28	DENND1A	chr9	hpv-9, hpv-10, hpv-11	E1	HPV16	57706
28	ZCCHC7	chr9	hpv-9, hpv-10, hpv-11	E1	HPV16	84186
27	GLG1	chr16	hpv-9, hpv-10, hpv-11	E1	HPV16	2734
25	PCCA	chr13	hpv-9, hpv-10, hpv-11	E1	HPV16	5095
25	SMYD3	chr1	hpv-9, hpv-10, hpv-11	E1	HPV16	64754
23	DOCK1	chr10	hpv-9, hpv-10, hpv-11	E1	HPV16	1793
23	INSR	chr19	hpv-9, hpv-10, hpv-11	E1	HPV16	3643
23	C2orf34	chr2	hpv-9, hpv-10, hpv-11	E1	HPV16	79823
22	JUP	chr17	hpv-9, hpv-10, hpv-11	E1	HPV16	3728
22	NSF	chr17	hpv-9, hpv-10, hpv-11	E1	HPV16	4905
21	GPC3	chrX	hpv-9, hpv-10, hpv-11	E1	HPV16	2719
20	CACNB2	chr10	hpv-9, hpv-10, hpv-11	E1	HPV16	783
20	CHD9	chr16	hpv-9, hpv-10, hpv-11	E1	HPV16	80205
19	ADK	chr10	hpv-9, hpv-10, hpv-11	E1	HPV16	132
19	INTS4	chr11	hpv-9, hpv-10, hpv-11	E1	HPV16	92105
18	ACVR1B	chr12	hpv-9, hpv-10, hpv-11	E1	HPV16	91
17	DTNB	chr2	hpv-9, hpv-10, hpv-11	E1	HPV16	1838
16	BMPR2	chr2	hpv-9, hpv-10, hpv-11	E1	HPV16	659
16	UVRAG	chr11	hpv-9, hpv-10, hpv-11	E1	HPV16	7405
15	BPHL	chr6	hpv-9, hpv-10, hpv-11	E1	HPV16	670
14	ADCY2	chr5	hpv-9, hpv-10, hpv-11	E1	HPV16	108
14	NVL	chr1	hpv-9, hpv-10, hpv-11	E1	HPV16	4931
14	SAMHD1	chr20	hpv-9, hpv-10, hpv-11	E1	HPV16	25939
14	VKORC1L1	chr7	hpv-9, hpv-10, hpv-11	E1	HPV16	154807
13	ATP1B3	chr3	hpv-9, hpv-10	E1	HPV16	483
13	DENND2C	chr1	hpv-9, hpv-10	E1	HPV16	163259
12	ACACA	chr17	hpv-9, hpv-10, hpv-11	E1	HPV16	31
12	STYXL1	chr7	hpv-9, hpv-10, hpv-11	E1	HPV16	51657
11	ADD3	chr10	hpv-9, hpv-11	E1	HPV16	120
11	DPP8	chr15	hpv-9, hpv-10, hpv-11	E1	HPV16	54878

11	SLC4A5	chr2	hpv-9, hpv-10, hpv-11	E1	HPV16	57835
10	ABR	chr17	hpv-9, hpv-10, hpv-11	E1	HPV16	29
10	MYOM1	chr18	hpv-9, hpv-10, hpv-11	E1	HPV16	8736
10	DSTN	chr20	hpv-9, hpv-10, hpv-11	E1	HPV16	11034
10	ASTN2	chr9	hpv-9, hpv-10, hpv-11	E1	HPV16	23245
10	FBXO43	chr8	hpv-9, hpv-10, hpv-11	E1	HPV16	286151
10	HCN1	chr5	hpv-9, hpv-10, hpv-11	E1	HPV16	348980

本实施例表明通过由于第二代测序技术能够对肿瘤样品进行深度测序，从而能够快速地对病毒与人体基因的整合进行检测，并且能准确提供被整合的人基因组区域信息以及整合外源序列信息。

5 实施例 3: 对肝癌样品外周血通过目标区域分析

实验方法:

1、DNA 提取及文库构建:

外周血 DNA 提取和文库制备方法与实施例 2 相同，只是样品来源为多名肝癌患者。

2、捕获游离核酸:

10 本实施例中将使用核酸探针芯片 (Nimblegen) 对含有外源序列区域的核酸片段进行捕获。实验流程如下:

a. 样品准备:

组分	重量/体积
Cot-1 DNA	300 μg
DNA 文库 (DNA 提取及文库构建部分中制备获得)	1 μg
PE Block 1.0	1 μL
PE Block 2.0	1 μL

15 其中,核酸探针是以具有 SEQ ID NO: 7 所示的核酸序列的 HBV 基因组 (Hepatitis B virus serotype adr, complete genome) 为模板,以 60 bp 为一个探针长度,每间隔 5 bp 设置一个探针,进行设计得到的。

PE Block 1.0:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCT-3' (SEQ ID NO: 13);

PE Block 2.0:

20 5'-AAGCAGAAGACGGCATAACGAGATATTGGCGTACTGGAGTTCAGACGTGTGCT

CTTCCGATCT-3' (SEQ ID NO: 14)。

- b. 将准备好的样品置于 SpeedVac 中 60℃ 蒸干，然后加入 11.2 μL 超纯水溶解样品。
- c. 全速离心样品 30 秒，分别加入以下两种试剂：18.5 μL 的 2 × SC Hybridization Buffer (Roche NimbleGen 公司) 和 7.3 μL 的 SC Hybridization Component A (Roche NimbleGen 公司)。震荡混匀后置于离心机上全速离心 30 秒，然后于 95℃ 使 DNA 变性 10 分钟。
- d. 根据制造商提供的说明书，将带有相应探针的芯片固定在杂交仪 (Roche NimbleGen 公司) 上，将变性后的样品加入芯片中并封闭芯片，然后设定杂交程序，于 42℃ 杂交 64-72 小时。
- e. 芯片洗涤与样品洗脱：

10

次序	洗涤/洗脱缓冲液 (Roche NimbleGen)	重复洗脱次数	水浴时间	水浴温度
1	1X Wash Buffer II	10 次	/	常温
2	1X Stringent Wash Buffer	10 次	5 min	47.5 °C
3	1X Stringent Wash Buffer	10 次	5 min	47.5 °C
3	1X Wash Buffer I	2 分钟(1 次/秒)	/	常温
5	1X Wash Buffer II	1 分钟(1 次/秒)	/	常温
6	1X Wash Buffer III	10	/	常温
7	NaOH (900 μL)	1	10 min	常温

- f. 将 NaOH 洗脱液回收，并用 32 μL 20%冰醋酸中和，得到中和液。
- g. 将上述中和液用 Qiagen MinElute PCR Purification Kit 纯化，捕获后的样品最后溶解于 138 μL 纯水中。
- h. PCR 扩增捕获的 DNA 文库，分为 6 管 50 μL 反应进行 PCR，PCR 的反应物组成如下：

DNA	138 μL
Phusion DNA 聚合酶	150 μL
PE Post Primer 1.0	6 μL
PE Post Primer 2.0	6 μL
总体积	300 μL

15

- 其中，
- PE Post Primer 1.0: AATGATACGGCGACCACCGAGATC (SEQ ID NO: 15);
 - PE Post Primer 2.0: CAAGCAGAAGACGGCATAACGA (SEQ ID NO: 16)。
- PCR 的反应条件如下：

- (a) . 98°C 30 s
- (b) . 98°C 15 s
- (c) . 62°C 30 s
- (d) . 72°C 30 s
- 5 (e) . 重复 (b) - (d) 步骤 11-19 次 (共扩增 12-20 次)
- (f) . 72°C 5 min
- (g) . 4°C 静置

i. 用 Qiagen QIAquick PCR Purification Kit 纯化 PCR 产物，最后溶于 30 μL 纯水中。

3、高通量测序

10 本实施例中,对于获自肿瘤病人外周血DNA样本分批按照制造商提供的操作说明书(参见 Illumina/Solexa 官方公布的 cBot) 进行上机测序操作。通过 100 轮测序循环, 得到长度为 100 bp 的 DNA 片段序列。

4、数据分析

15 根据制造商 Illumina 提供的 Pipeline 操作说明书 (参见 <http://www.illumina.com/>提供的 Pipeline 方法说明书), 将步骤高通量测序部分中测得的序列信息经过图形转化获得测序序列信息, 去掉测序质量低的序列之后最终可以获得针对 NCBI 版本 36 的人类基因组参考序列的 ELAND 比对结果。

20 获得的数据使用 SOAP 软件包进行比对分析, 使用两个末端测序信息进行比对时去除两个末端均比对至人基因组的序列, 保留其中一条链比对至人基因组的序列, 将另一末端序列比对至 HBV 基因组序列中, 获得 HBV 在人基因组重组信息, 包括人基因组中重组位置以及 HBV 类型。

5、数据结果

染色体	整合染色体区域	插入区域大小	HBV 整合基因	HBV 整合区域
Chr1	27、226、699-27、227、380	681bp	X 基因 C 基因	1383-2116

25 上述结果表明通过表明本发明的用于确定人体具有异常状态方法能够准确有效地检测到肝癌样品中 HBV 病毒整合信息, 即一段 733 bp HBV 序列整合至 1 号染色体区域。

实施例 4: 对肝癌样品外周血及组织通过目标区域捕获分析

实验方法:

1、DNA 提取及测序

30 根据常规方法, 对肝癌患者进行静脉取血, 得到患者的外周血样本, 通过离心得到血浆样本。按照 Tiangen Micro Kit (DP316) 微量基因组操作流程从血浆样本提取 DNA, 并分

别用 Qubit (Invitrogen, the Quant-iT™ dsDNA HS Assay Kit) 定量, 所提取的各样本的 DNA 总量均为 5 ~ 50 ng。

根据常规方法, 取肝癌患者, 癌组织样本对组织样本进行全基因组提取, 取 3 微克进行常规建库, 文库插入片段主带为 170bp。

5 将所提取到的 DNA, 分别按照制造商提供的标准建库规程 (参见 <http://www.illumina.com/> 提供的 Illumina 标准建库说明书) 建立 DNA 文库。简言之, 在 DNA 分子两端加上测序用的接头, 并被加上不同的标签序列, 然后与测序芯片表面互补接头杂交, 使核酸分子成簇生长, 然后在 Illumina HiSeq 2000 上测序 PE101, 得到长度为 100bp 的 DNA 片段序列。

10 2、捕获游离核酸:

本实施例中将使用 HBV (Nimblegen) 核酸探针芯片对含有外源序列区域的核酸片段进行捕获。实验流程如下:

a. 样品准备:

组分	重量/体积
Cot-1 DNA	5 µg
DNA 文库	1 µg
PE Block (1000 µM)	1 µL
PE INDEX Block (1000 µM)	1 µL

15 其中, HBV 核酸探针是由 HBV 的 A, B, C, D, E, F, G, H 八个型别的基因组进行设计得到的, 各基因组序列具体可见公知数据库中的 HBV 基因组序列。具体地, 按照 HBV 基因组长度, 每次在基因组上 10 bp 滑动移动一次, 合成 60-90 bp 长度且带 Bio-tin 标记的探针。具体探针是委托相应的公司合成的。

b. 将准备好的样品置于 SpeedVac 中 60°C 蒸干。

20 c. 全速离心样品 30 秒, 分别加入以下两种试剂: 7.5 µL 的 2 × SC Hybridization Buffer (Roche NimbleGen 公司) 和 3.0 µL 的 SC Hybridization Component A (Roche NimbleGen 公司)。震荡混匀后置于离心机上全速离心 30 秒, 然后于 95°C 使 DNA 变性 10 分钟, 而后加入探针 4.5 µl 于 47°C 杂交 24 小时。

e. 芯片洗涤与样品洗脱: 按照标准的 EZ 洗脱流程进行洗脱, 回收。

25 f. 预先从 -20°C 保存的试剂盒中取出 Pfx 酶 (Invirtogen), Pfx buffer (10*), dNTP (10 mM), PCR Primer F (10 pm/µl), PCR Primer R (10 pm/µl)。

在 PCR 小管上, 每孔按照下面的表格配置 PCR 反应体系:

DNA	35 µl
Pfx 酶	1 µl

Pfx 缓冲液(10*)	5 μ l
MgSO ₄	2 μ l
dNTP(10 mM)	2 μ l
Flowcell-primer-F1 (10 pm/ μ l)	2.5 μ l
Flowcell-primer-R1 (10 pm/ μ l)	2.5 μ l
总体积	50 μ l

其中,

Flowcell-primer-F1: AATGATACGGCGACCACCGAGATC (SEQ ID NO: 17);

Flowcell-primer-R1: CAAGCAGAAGACGGCATAACGA (SEQ ID NO: 18)。

程序如下:

- 5 94°C 2min;
 94°C 15s, 58°C 30s, 72°C 30s, 14 个循环;
 72°C 5min,
 4°C ∞ 。

PCR 结束后, 每个样品都用 1.5 倍体积的 Amprue Beads 纯化, 回收的 PCR 产物溶于
 10 30 μ l 超纯水中, Nanodrop1000 测其浓度为 10 ng/ μ l。

3、高通量测序

本实施例中, 对于获自肿瘤病人外周血 DNA 样本分批按照制造商提供的操作说明书(参见 Illumina/Solexa 官方公布的 cBot) 进行上机测序操作。

4、数据分析

- 15 将下机数据除去重复和被接头污染的 reads, 统计下机数据的基本信息(文库长度; reads 长度; reads 条数; 碱基数; 重复率等); 分别截取 PE 的两条 reads 前面的 50bp 碱基, 形成一对长为 50bp 新 reads, 即 PE50 reads。将新的 PE50 reads 运用 soap 比对软件 (-r 1 -v 2) 分别与人的参考序列 (hg19) 和 HBV 各种参考序列进行比对, 从比对结果中挑选出一条 read 比到人的参考序列并且另一条比对到 HBV 参考序列的一对 reads; 这样的 reads 很有可能跨
 20 过 HBV 插入的位点; 组装这部分 reads, 采用 BWA 比对找到在人类基因组的插入热点区域。

5、数据结果

A 患者组织 DNA, 经过杂交后 1G (碱基) 数据量结果如下:

染色体号	位置	支持数目
chr18	11550976	124
chr18	11550851	119
chr17	18778136	14
chr17	18778051	14
chr17	18778353	11

chr17	18778266	8
-------	----------	---

A 患者血浆 DNA, 经过杂交后 5G (碱基) 数据量结果如下:

染色体号	位置	支持数目
chr18	11550851	18
chr18	11550976	18
chr2	219543496	4
chr2	219543519	2

采用上述方法对同一患者的血浆与肝癌组织中的整合状态进行查找, 发现组织中存在的
的支持数最高的位置, 在血浆中同样存在, 此方法不仅证明组织与血浆游离 DNA 之间的关
系, 并且能够准确有效地找出血浆中的整合位置, 表明本发明的用于确定人体具有异常状
态方法为能够有效地应用于病原体引起染色体整合相关疾病的无创检测。

实施例 5: 对 HPV 患者子宫颈脱落细胞目标区域捕获分析

实验方法:

1、DNA 提取及测序

10 根据常规方法, 对 HPV 患者进行子宫颈脱落细胞取样, 样本按照 Tiangen Micro Kit (DP316) 微量基因组操作流程从血浆样本提取 DNA, 分别用 Qubit (Invitrogen, the Quant-iT™ dsDNA HS Assay Kit) 定量, 所提取的 DNA 总量分别为 100~500 ng。

15 将所提取到的 DNA, 分别按照制造商提供的标准建库规程 (参见 <http://www.illumina.com/>提供的 Illumina 标准建库说明书) 建立 170 bp DNA 文库。简言之, 在 DNA 分子两端加上测序用的接头, 并被加上不同的标签序列, 然后与测序芯片表面互补接头杂交, 使核酸分子成簇生长, 然后在 Illumina HiSeq 2000 上测序 PE101, 得到长度为 100 bp 的 DNA 片段序列。

2、捕获游离核酸:

20 本实施例中将使用(Mygenostics 公司)HPV 核酸探针芯片对含有外源序列区域的核酸片段进行捕获。实验流程如下:

a. 样品准备:

试剂	体积(μL)	体积(μL)
DNA 文库	20/(4 μg)	20/(3 μg)
BL 缓冲液	10	10
P1 block(1000 μM)	4	3
INDEX block(1000 μM)	4	3
Probe	5	5
95 度 7 分钟; 65 度 2 分钟后, 加入 HY 缓冲液 (需要 65 度预热 30 分钟)		
HY 缓冲液	23	23

总体积	72	72
65 度 22 小时		

其中，HPV 核酸探针是由 HPV 的 6, 11, 16, 18, 31, 33, 35, 39, 45, 52, 56, 58, 59, 66, 68, 69, 82 型别的基因组进行设计得到的，各基因组序列具体可见公知数据库中的 HPV 基因组。具体地，按照 HPV 基因组长度，每次在基因组上 10 bp 滑动移动一次，合成 60-90 bp 长度且带 Bio-tin 标记的探针。具体探针是委托相应的公司合成的。

5 b. 芯片洗涤与样品洗脱：按照标准（MyGenostics）洗脱流程进行洗脱，回收，具体步骤如下：

1) 提前将温浴器调至 65℃；

2) 用漩涡混合仪剧烈振荡重悬 Myone beads C1（Invitrogen）至混匀；

3) 每一个杂交反应取 50 μl Myone beads C1 磁珠于一新的 1.5 mL 离心管中，放于磁力架上，然后去除上清；

4) 洗涤磁珠：

a) 加入 50 μL 1*结合缓冲液；

b) 用漩涡混合仪剧烈振荡 5 秒钟重悬磁珠；

c) 将离心管放于磁力架上，待液体变澄清；

15 d) 去除上清液；

e) 重复 2 次“步骤 a 到步骤 d”；

5) 加入 80 μL（根据不同大小捕获区域区域加入量不同，具体按照 KIT 要求加入）1*结合缓冲液重悬磁珠；

6) 加入 64 μL 2*结合缓冲液重悬磁珠（与杂交液体积同等体积），将液体转移到 EPI 后，总体积约为 200 微升；

7) 振荡混匀后在室温下放于 ROATER 上进行 1 小时；

8) 震荡混匀后将样品放于磁力架上，去除上清；

9) 加入 500 μl WB1，振荡混匀后旋转混匀 15 分钟，然后放于磁力架上去上清；

10) 加入 500 μl WB3，振荡混匀后放于 65 度温浴，850 rpm，10 分钟，放于磁力架上去上清；

11) 重复（10）5 次，最后一次完全去除上清；

12) 加入 50 μl 洗脱缓冲液，振荡混匀，室温放置 10 分钟，放于磁力架上取上清转入另一 EP 管（管内包含 70 μl NE 缓冲液）；

13) 采用 QIAquick Minelute 进行纯化，最终溶解 42 μl EB。

30 c. 捕获样品的扩增与纯化

1) 从 -20℃ 冰箱中取出 2*PHUSION MASTER, Flowcell primers(10 μM)，将其置于冰上化冻并充分混匀。

2) 在冰上为每个捕获样品配制一份 Mix，另外加入一个无模板的阴性对照，按以下表

格组分配制反应 Mix 并用移液器混匀:

反应个数	1 个样品
2*PHUSION MASTER	25 μ L
Flowcell primerF(10 μ M)	2 μ L
Flowcell primerR(10 μ M)	2 μ L
样品	21 μ L
总体积	50 μ L

3) 在热循环仪中运行下列程序:

- a. 98 °C 30s
- b. 98 °C 25 s
- 5 c. 65 °C 30 s
- d. 72 °C 30 s
- e. 重复 b-d 步骤(共 15 次循环)
- f. 72 °C 5 min
- g. 4 °C Hold

10 PCR 结束后, 每个样品都用 1.5 倍体积的 Amprue Beads 纯化, 回收的 PCR 产物溶于 30 μ l 超纯水中, Nanodrop1000 测浓度, 并记录, 其浓度为 10 ng/ μ l。

3、高通量测序

本实施例中, 对于获自 HPV 患者的子宫颈脱落细胞 DNA 样本分批按照制造商提供的操作说明书 (参见 Illumina/Solexa 官方公布的 cBot) 进行上机测序操作。通过 100 轮测序
15 循环, 得到长度为 100 bp 的 DNA 片段序列。

4、数据分析

将下机数据除去重复和被接头污染的 reads, 统计下机数据的基本信息(文库长度; reads 长度; reads 条数; 碱基数; 重复率等); 分别截取 PE 的两条 reads 前面的 50bp 碱基, 形成一对长为 50bp 新 reads, 即 PE50 reads。将新的 PE50 reads 运用 Soap 比对软件 (-r 1 -v 2)
20 分别与人的参考序列 (hg19)和 HBV 各种参考序列进行比对, 从比对结果中挑选出一条 read 比到人的参考序列并且另一条比对到 HBV 参考序列的一对 reads; 这样的 reads 很有可能跨过 HBV 插入的位点; 组装这部分 reads, 采用 BWA 比对找到在人类基因组的插入热点区域。

5、数据结果

HPV16 型感染且宫颈病变程度为 CIN 四期病人, 数据量 1G, 结果如下:

染色体	位置	支持数
chr4	124686065	824
chr4	124686035	529

chr2	133034596	16
------	-----------	----

上述方法采用宫颈病变程度为 CIN 四期病人，感染的宫颈脱落细胞进行检测，结果准确发现高频整合位置，证明了表明本发明的用于确定人体具有异常状态方法具有可行性。

工业实用性

- 5 本发明的用于确定人体具有异常状态的系统和方法，能够有效地应用于人体疾病的无创检测，通过对人体样本的核酸序列信息进行分析，可以根据核酸序列中所包含的信息准确地确定人体是否具有异常状态。

10 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。

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

权利要求书

- 1、一种确定人体具有异常状态的方法，其特征在于，包括：
提供人体样本的核酸序列信息，所述人体样本的核酸序列信息是基于对所述人体样本
5 进行检测而获得的；以及
基于所述人体样本的核酸序列信息，确定所述人体是否具有异常状态。
- 2、根据权利要求1所述的确定人体具有异常状态的方法，其特征在于，所述人体样本
的核酸序列信息是基于对所述人体样本进行核酸序列检测而获得的。
- 3、根据权利要求2所述的确定人体具有异常状态的方法，其特征在于，所述核酸序列
10 检测是借助第二代测序技术或第三代测序技术进行的。
- 4、根据权利要求1所述的确定人体具有异常状态的方法，其特征在于，所述样本为所
述人体的细胞、组织、血液、体液、尿液、排泄物或其组合。
- 5、根据权利要求1所述的确定人体具有异常状态的方法，其特征在于，所述人体样本
为血浆或者血清。
- 15 6、根据权利要求1所述的确定人体具有异常状态的方法，其特征在于，所述核酸序列
信息包括所述人体样本中游离核酸的序列信息。
- 7、根据权利要求6所述的确定人体具有异常状态的方法，其特征在于，所述人体样本
中游离核酸的序列信息是通过除去所述人体样本中的细胞后，进行测序检测而获得的。
- 8、根据权利要求1所述的确定人体具有异常状态的方法，其特征在于，所述异常状态
20 选自疾病的发生、疾病的发展阶段、疾病的疗效和预后的至少一种。
- 9、根据权利要求8所述的确定人体具有异常状态的方法，其特征在于，所述疾病是肿
瘤性疾病、免疫性疾病、遗传性疾病的至少一种。
- 10、根据权利要求9所述的确定人体具有异常状态的方法，其特征在于，所述肿瘤性
疾病是选自肺癌、肝癌、胃癌、食管癌、结直肠癌、胰腺癌、乳腺癌、膀胱癌、肾癌、卵
25 巢癌、宫颈癌、甲状腺癌、鼻咽癌、脑胶质瘤的至少一种。
- 11、根据权利要求8所述的确定人体具有异常状态的方法，其特征在于，如果所述核
酸序列信息包含选自下列至少一种的核酸片段序列：HBV、HPV、EBV、幽门螺旋杆菌，
则确定所述人体患有宫颈癌、肝癌、鼻咽癌、胃癌的至少一种。
- 12、根据权利要求1所述的确定人体具有异常状态的方法，其特征在于，
30 对所述核酸进行测序检测之前，利用探针捕获含有特定序列的核酸，然后对所述含有
特定序列的核酸进行测序检测。
- 13、根据权利要求12所述的确定人体具有异常状态的方法，其特征在于，
所述探针对于选自下列的至少一种是特异性的：HBV、HPV、EBV、幽门螺旋杆菌，
任选地，所述探针为核酸探针集，包括多个探针，所述核酸探针集具有以下特征：
35 (1) 每个探针上具有1个或多个生物素标记的dNTP；和/或
(2) 所述生物素标记的dNTP在所述探针集中的丰度为1: 6-1: 2；和/或

(3) 所述探针集的全部核酸序列覆盖对应选自 HBV、HPV、EBV 和幽门螺旋杆菌的至少一种病毒的基因组序列的 70%-100%，

优选地，所述探针集具有 1-20000 个核酸探针；较佳地，所述探针集具有 1000-5000 个核酸探针；更佳地，所述探针集具有 2500 个核酸探针，

5 优选地，所述生物素标记的 dNTP 在所述探针集中的丰度为 1: 4，

优选地，在所述探针集中，所述探针之间具有部分重叠，

优选地，所述探针长度为 100-500 bp；较佳地，所述探针长度为 200-300 bp；更佳地，所述探针长度为 250 bp，

10 优选地，所述探针是以病毒基因组作为模板，PCR 法扩增获得的，较佳地，所述模板为选自 HBV 基因组、HCV 基因组、HIV 基因组和 HPV 基因组的至少一种；更佳地，所述模板为 B 型 HBV 基因组和/或 C 型 HBV 基因组。

14、根据权利要求 1 所述的确定人体具有异常状态的方法，其特征在于，

在对所述核酸进行测序检测之前，利用探针去除含有特定序列的核酸，然后对所述去除后剩余的核酸进行测序检测。

15 15、根据权利要求 14 所述的确定人体具有异常状态的方法，其特征在于，

所述探针可以结合人体基因组中的共有序列，或者为可以结合人体基因组中甲基化位点的抗体或者蛋白。

16、一种用于确定人体具有异常状态的系统，其特征在于，包括：

20 核酸序列信息接收器，所述核酸序列信息接收器接收人体样本的核酸序列信息；以及核酸序列信息分析器，所述核酸序列信息分析器与所述核酸序列信息接收器相连，并基于所述人体样本的核酸序列信息，确定所述人体是否具有异常状态。

17、根据权利要求 16 所述的用于确定人体具有异常状态的系统，其特征在于，

所述核酸序列信息分析器内预存有选自下列的至少一种：人体正常状态的基因组序列、病原体的基因组序列、正常人群的基因组序列。

25 18、根据权利要求 17 所述的用于确定人体具有异常状态的系统，其特征在于，所述病原体为选自 HBV、HPV、EBV、幽门螺旋杆菌的至少一种。

30 19、根据权利要求 18 所述的用于确定人体具有异常状态的系统，其特征在于，进一步包括核酸序列检测装置，所述核酸序列检测装置与所述核酸序列信息接收器相连，用于对所述人体样本进行核酸序列检测获得所述核酸序列信息并输送至所述核酸序列信息接收器。

20、根据权利要求 19 所述的用于确定人体具有异常状态的系统，其特征在于，所述核酸序列检测装置借助第二代测序技术或第三代测序技术。

21、根据权利要求 19 所述的用于确定人体具有异常状态的系统，其特征在于，进一步包括

35 游离核酸捕获装置，所述游离核酸捕获装置与所述核酸序列检测装置相连，其中，

所述游离核酸捕获装置设置有探针，

所述探针适于捕获含有特定序列的核酸，并且将所述含有特定序列的核酸输送至所述核酸序列检测装置进行核酸序列检测；或者

5 所述探针适于去除含有特定序列的核酸，并且将所述去除后的核酸输送至所述核酸序列检测装置进行核酸序列检测。

22、根据权利要求 21 所述的确定人体具有异常状态的方法，其特征在于，

所述适于捕获含有特定序列的核酸的探针对于选自下列的至少一种是特异性的：HBV、HPV、EBV、幽门螺旋杆菌，

任选地，所述探针为核酸探针集，包括多个探针，所述核酸探针集具有以下特征：

- 10 (1) 每个探针上具有 1 个或多个生物素标记的 dNTP；和/或
(2) 所述生物素标记的 dNTP 在所述探针集中的丰度为 1: 6-1: 2；和/或
(3) 所述探针集的全部核酸序列覆盖对应选自 HBV、HPV、EBV 和幽门螺旋杆菌的至少一种病毒的基因组序列的 70%-100%，

15 优选地，所述探针集具有 1-20000 个核酸探针；较佳地，所述探针集具有 1000-5000 个核酸探针；更佳地，所述探针集具有 2500 个核酸探针，

优选地，所述生物素标记的 dNTP 在所述探针集中的丰度为 1: 4，

优选地，在所述探针集中，所述探针之间具有部分重叠，

优选地，所述探针长度为 100-500 bp；较佳地，所述探针长度为 200-300 bp；更佳地，所述探针长度为 250 bp，

20 优选地，所述探针是以病毒基因组作为模板，PCR 法扩增获得的，较佳地，所述模板为选自 HBV 基因组、HCV 基因组、HIV 基因组和 HPV 基因组的至少一种；更佳地，所述模板为 B 型 HBV 基因组和/或 C 型 HBV 基因组。

23、根据权利要求 21 所述的确定人体具有异常状态的方法，其特征在于，

25 所述适于除去含有特定序列的核酸的探针为可以结合人体基因组中的共有序列，或者为可以结合人体基因组中甲基化位点的抗体或者蛋白。

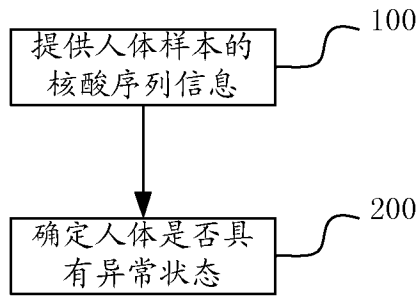


图 1

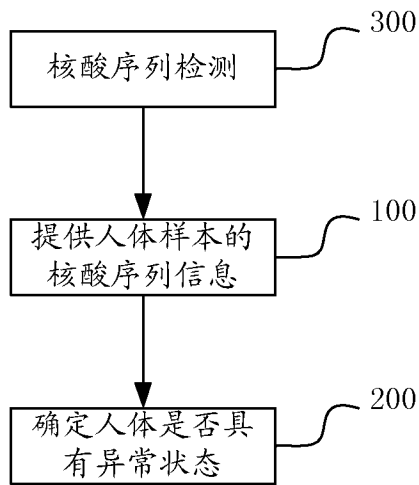


图 2

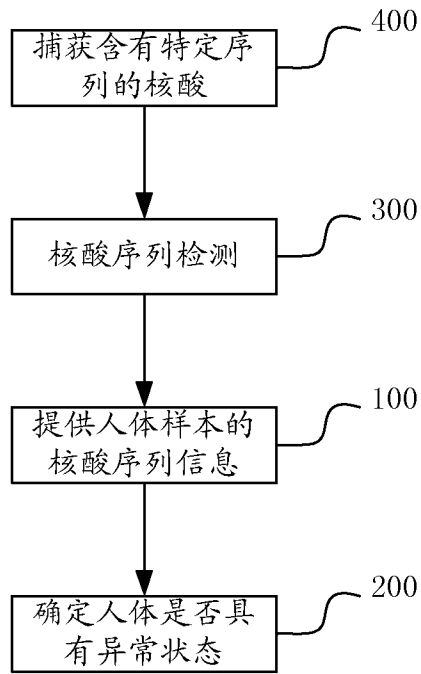


图 3

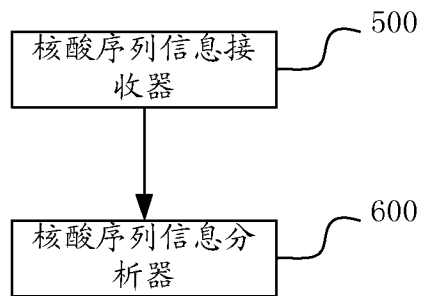


图 4

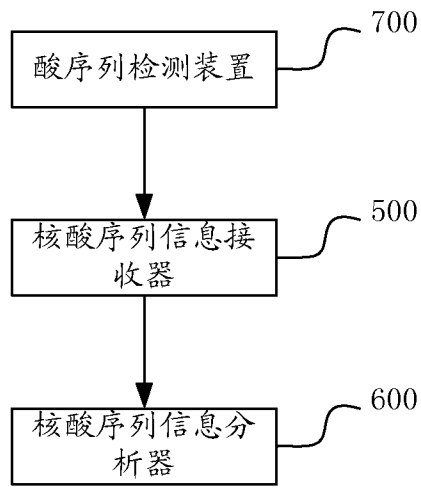


图 5

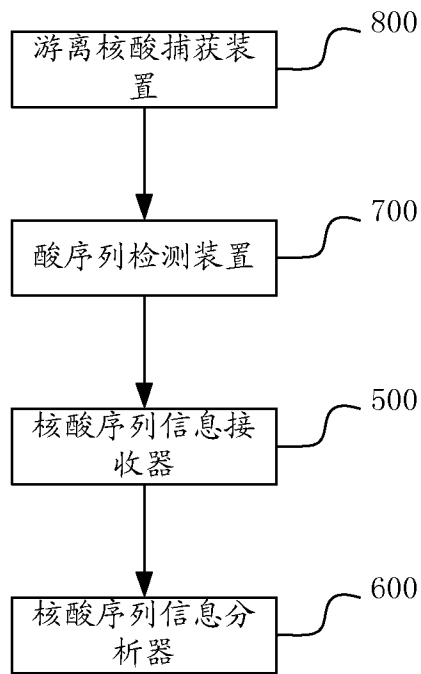


图 6

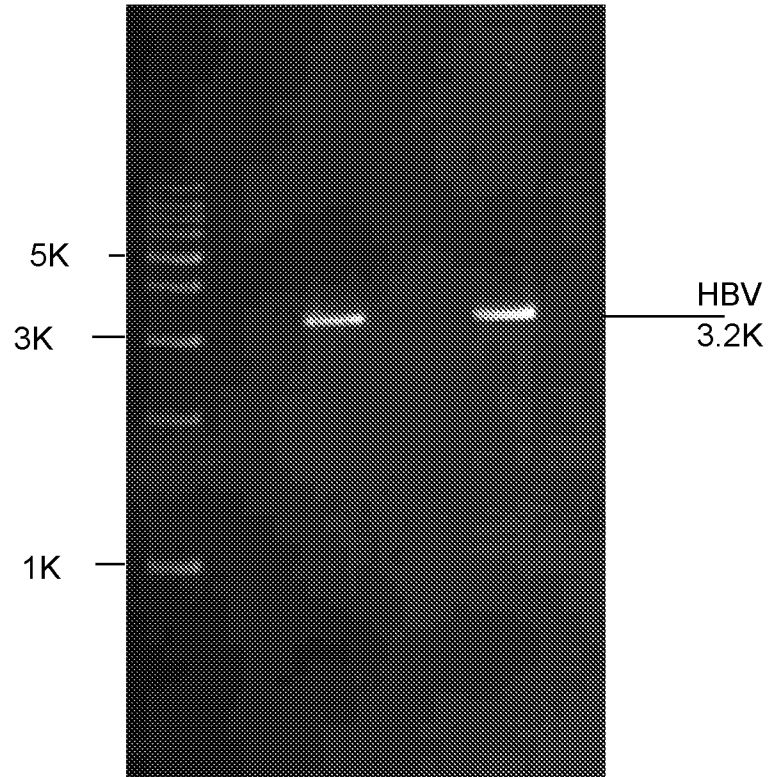


图 7

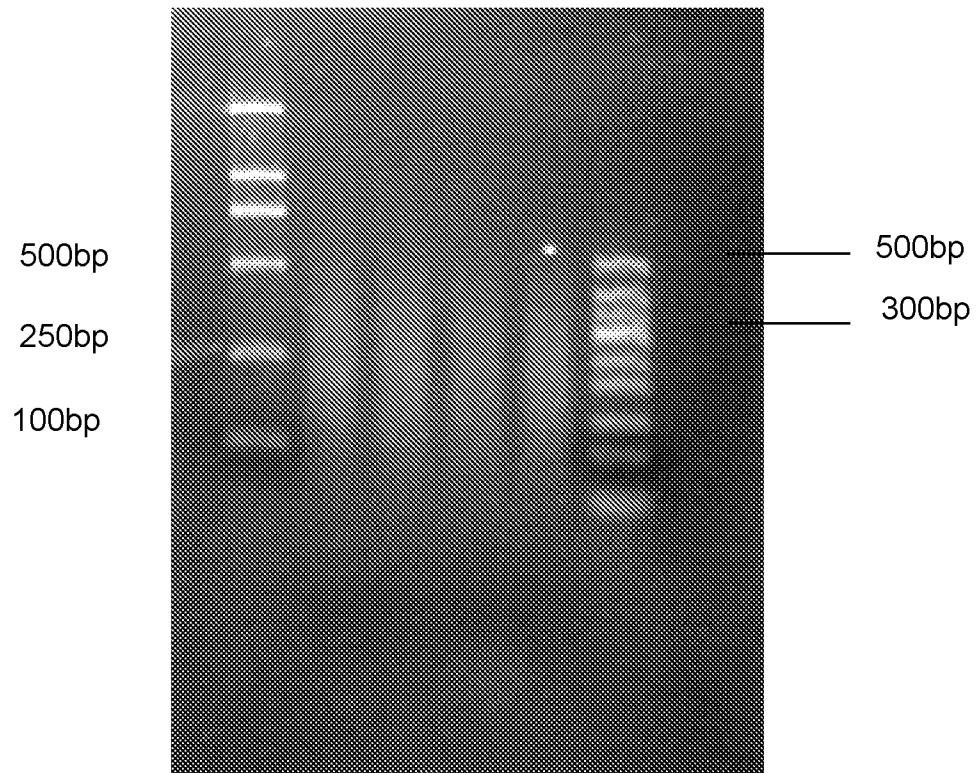


图 8

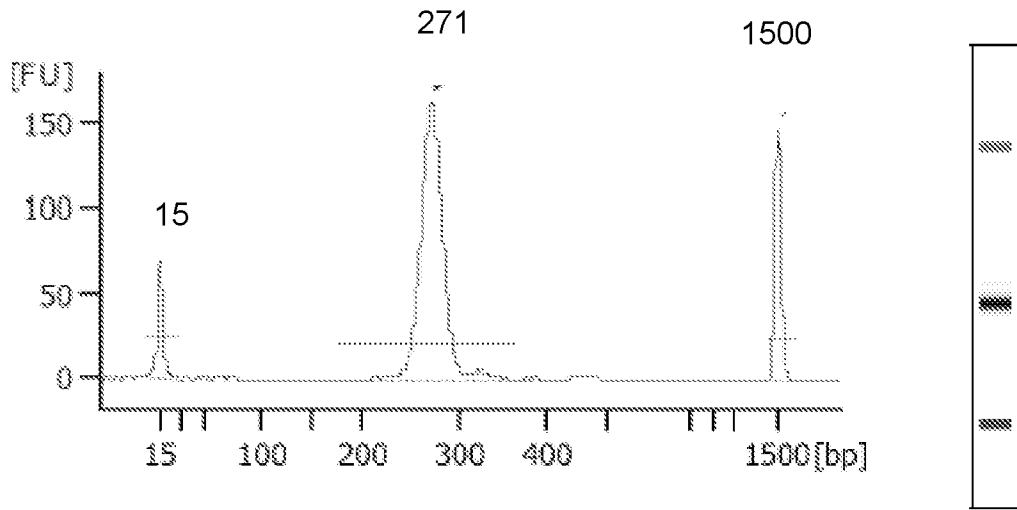


图 9

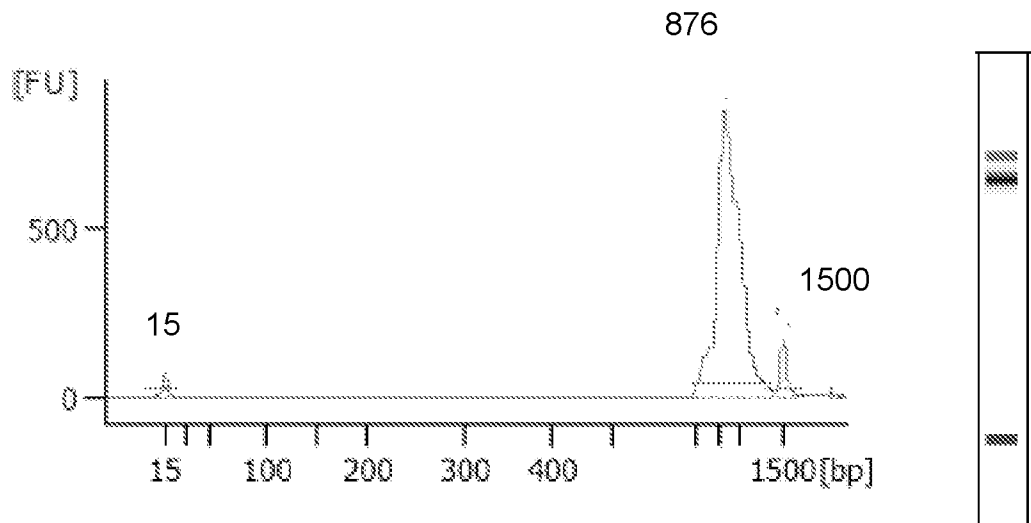


图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/077175

A. CLASSIFICATION OF SUBJECT MATTER

See the 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, C12M

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)

Data bases: CNPAT, CNKI, ISI Web of knowledge, NCBI

Key Words: HBV, HPV, EBV, sequence, sequencing, solexa, probe, primer, biotin, cancer, tumor, helicobacter pylori

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN101921874A (BGI SHENZHEN CO LTD) , 22 Dec. 2010 (22.12.2010), See claim 9, background of the description, exmaple 5, example 6	1-23

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

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p>Date of the actual completion of the international search</p> <p style="text-align: center;">22 Aug. 2012 (22.08.2012)</p>	<p>Date of mailing of the international search report</p> <p style="text-align: center;">20 Sep. 2012(20.09.2012)</p>
-------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------

<p>Name and mailing address of the ISA</p> <p>State Intellectual Property Office of the P. R. China</p> <p>No. 6, Xitucheng Road, Jimenqiao</p> <p>Haidian District, Beijing 100088, China</p> <p>Facsimile No. (86-10)62019451</p>	<p>Authorized officer</p> <p style="text-align: center;">ZHANG, Qi</p> <p>Telephone No. (86-10) 62411036</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/077175

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN101921874A	22.12.2010	NONE	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/077175

CLASSIFICATION OF SUBJECT MATTER

C12Q1/68 (2006.01) i

C12Q1/70 (2006.01) i

C12M1/34 (2006.01) i

<p>A. 主题的分类</p> <p style="text-align: center;">见附加页</p> <p>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</p>												
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>IPC: C12Q, C12M</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>数据库: CNPAT, CNKI, ISI Web of knowledge, NCBI:</p> <p>关键词: 序列、测序、探针、引物, 生物素、癌、肿瘤, HBV, HPV, EBV, 幽门螺旋杆菌, sequence, sequencing, solexa, probe, primer, biotin, cancer, tumor, helicobacter pylori</p>												
<p>C. 相关文件</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">类 型*</th> <th style="width: 60%;">引用文件, 必要时, 指明相关段落</th> <th style="width: 30%;">相关的权利要求</th> </tr> </thead> <tbody> <tr> <td style="vertical-align: top;">X</td> <td style="vertical-align: top;">CN101921874A (深圳华大基因科技有限公司) 22.12 月 2010 (22.12.2010) 见权利要求 9, 说明书背景技术, 实施例 5, 实施例 6</td> <td style="vertical-align: top;">1-23</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	X	CN101921874A (深圳华大基因科技有限公司) 22.12 月 2010 (22.12.2010) 见权利要求 9, 说明书背景技术, 实施例 5, 实施例 6	1-23				
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求										
X	CN101921874A (深圳华大基因科技有限公司) 22.12 月 2010 (22.12.2010) 见权利要求 9, 说明书背景技术, 实施例 5, 实施例 6	1-23										
<p><input type="checkbox"/> 其余文件在 C 栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p>												
<p>* 引用文件的具体类型:</p> <table style="width: 100%;"> <tr> <td style="width: 50%;">“A” 认为不特别相关的表示了现有技术一般状态的文件</td> <td style="width: 50%;">“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</td> </tr> <tr> <td>“E” 在国际申请日的当天或之后公布的在先申请或专利</td> <td>“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</td> </tr> <tr> <td>“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)</td> <td>“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</td> </tr> <tr> <td>“O” 涉及口头公开、使用、展览或其他方式公开的文件</td> <td>“&” 同族专利的文件</td> </tr> <tr> <td>“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</td> <td></td> </tr> </table>			“A” 认为不特别相关的表示了现有技术一般状态的文件	“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	“E” 在国际申请日的当天或之后公布的在先申请或专利	“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	“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)	“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	“O” 涉及口头公开、使用、展览或其他方式公开的文件	“&” 同族专利的文件	“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	
“A” 认为不特别相关的表示了现有技术一般状态的文件	“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											
“E” 在国际申请日的当天或之后公布的在先申请或专利	“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											
“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)	“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											
“O” 涉及口头公开、使用、展览或其他方式公开的文件	“&” 同族专利的文件											
“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												
<p>国际检索实际完成的日期 22.08 月 2012 (22.08.2012)</p>		<p>国际检索报告邮寄日期 20.9 月 2012 (20.09.2012)</p>										
<p>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</p>		<p>受权官员 张起 电话号码: (86-10) 62411036</p>										

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2012/077175

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN101921874A	22.12.2010	无	

主题的分类:

C12Q1/68 (2006.01) i

C12Q1/70 (2006.01) i

C12M1/34 (2006.01) i