

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日

2012 年 1 月 5 日 (05.01.2012)



(10) 国际公布号

W O 2012/000149 A I

- (51) 国际分类号 :
C12N 15/11 (2006.01) C12Q 1/70 (2006.01)
C12Q 1/68 (2006.01)
- (21) 国际申请号 : PCT/CN2010/001812
- (22) 国际申请日 : 2010 年 11 月 12 日 (12.11.2010)
- (25) 申报语言 : 中文
- (26) 公布语言 : 中文
- (30) 优先权 :
201010213736.9 2010 年 6 月 30 日 (30.06.2010) CN
- (71) 申请人 (除美国外的所有指定国): 深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED) [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): 易鑫 (YI, Xin) [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。刘涛 (LIU, Tao) [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。徐佳佳 (XU, Jiajia) [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); 中国北京市西城区阜成门外大街 2 号万通新世界广场 8 层, Beijing 100037 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, ML, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。
- 本国际公布 :
- 包括国际检索报告 (条约第 21 条 (3))。
- 包括说明书序列列表部分 (细则 5.2(a))。

(54) Title: DNA MOLECULAR INDEXES FOR HIGH-THROUGHPUT DETECTING HUMAN PAPILLOMA VIRUS

(54) 发明名称: 用于高通量检测人类乳头瘤病毒的 DNA 分子标签

(57) Abstract: The present invention provides a method for detecting Human Papilloma Virus (HPV), especially a method for detecting HPV based on Solexa sequencing. The present invention also provides indexes, index primers and kits for high-throughput HPV sequencing, detecting or subtyping.

(57) 摘要:

本发明提供了人类乳头瘤病毒 (Human Papilloma virus, HPV) 的检测方法, 特别是基于 Solexa 测序法的 HPV 检测方法。本发明还提供了用于高通量测序、检测或分型的标签、标签引物和试剂盒。



2012/00 14 A1

用于高通量检测人类乳头瘤病毒的 DNA 分子标签

技术领域

本发明涉及人类乳头瘤病毒 (Human Papilloma virus, HPV) 的检测方法,特别是基于 Solexa 测序法的 HPV 检测方法。

背景技术

宫颈癌是世界上女性癌症的第二号杀手,仅次于乳腺癌。每年全世界约有 50 万新发病例,25 万死亡病例,其中发展中国家占 2/3。我国是宫颈癌的高发区,占全球宫颈癌总数的 10%。研究表明,人乳头瘤病毒 (HPV) 与宫颈癌有着密切关系,是重要的致癌因子,同时也是引发宫颈癌的必要条件之一。已经表明,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。基于诱导的良性的,恶化前的或恶性的病变,将 HPV 分别分为低危型 (例如 HPV6, 11, 42, 43 和 44) 和高危型 (例如 HPV16, 18, 31, 33 和 45)。因此,对 HPV 感染的及早发现和正确分型对宫颈癌防治显得至关重要。

目前 HPV 的检测方法主要有以下几种: ① 细胞学检查,其利用宫颈刮片细胞学检查或液基薄层细胞学检查,藉由细胞外观形态的转变来加以诊断。对于 HPV 感染,镜下可见挖空细胞、角化不良以及湿疣外底层细胞。其局限性在于,HPV 感染的诊断的灵敏性和特异性较低。② 免疫组化方法,其通过检测 HPV 的衣壳抗原来进一步明确 HPV 感染,其所得的阳性反应定位明确,判断可靠。但是,衣壳抗原仅在 HPV-DNA 复制成熟后才产生,因此,诊断为阴性的受试者并不能被确定为不被 HPV 感染,该方法灵敏性较低。③ 实时

光定量PCR法 (FQ—PCR) , 其主要是应用荧光检测PCR仪 , 在PCR反应体系中加入荧光基团 , 利用荧光信号积累对PCR过程中每一个循环产生的扩增产物进行实时监测 , 从而对模板的初始浓度进行定量。此方法通量较低。④杂交捕获法 (主要为HC- II系统) , 其是目前唯一获得美国FDA认证的可用于临床的HPV DNA检测方法 , 并且其已获得欧洲CE及中国SDA认证。该方法使用专用的标本采集保存器 ; 已获得专利保护的全长8000bp的RNA探针以及已获得专利保护的特异性第一抗体。其原理是将核酸探针杂交到受测检体的HPV DNA上 , 再通过化学荧光或酶反应借助增幅的信号进行检测。该方法所使用的核酸探针主要分为两种 : 针对低危险型HPV的核酸探针和针对高危险型HPV的核酸探针。该方法可用于HPV的初步筛检 , 但是不能确定HPV的特定类型 , 也不能确定多重感染的情况。

将上述检测方法联合使用 , 可以提高HPV检测的灵敏性并降低检测的假阴性率。然而 , 这些方法的组合成本相对较高 , 一般只适用于经济发达地区的HPV检测和宫颈癌筛查。对于经济较不发达地区 , 特别是在山区和广大农村地区 , 上述检测方法的联合使用存在着较大的局限性。因此 , 开发适合的、低成本的HPV检测方法 , 是亟待解决的问题。

另一方面 , 目前已知的HPV检测方法 , 例如上文描述的那些检测方法 , 通量都较低。当对大规模的样品进行HPV检测时 , 应用上述方法是耗时耗力的 , 并且成本高昂。因此 , 本领域迫切需要新的高通量的、低成本的HPV检测方法。

发明内容

本发明基于Solexa测序法以及PCR标签 (PCR index) , 开发了新型的HPV检测方法以及用于此的试剂盒。根据本发明的方

法和试剂盒不仅可以实现 HPV 的高通量、低成本检测，而且可以实现 HPV 的精确分型。

定义

为了更好地理解本发明，下面提供相关术语的定义和解释。

如本文中所使用的，术语 "PCR" 是指聚合酶链式反应。

如本文中所使用的，术语 "Solexa 测序法" 是指近几年开发的新一代 DNA 测序法，也称为第二代测序法。Solexa 测序法与传统测序法（例如，Sanger 测序法）的不同之处在于，其采用边合成边测序的原理进行 DNA 序列分析。Solexa 测序法具有下述优点：1) 成本低，仅为传统测序成本的 1%；2) 通量高，可以同时多个样品进行测序，并且进行一次的 Solexa 测序法可以产生大约 500 亿 (50G) 个碱基的数据；3) 精确性高 (高于 98.4%)，有效的解决了多聚重复序列的读取问题。另一方面，高测序通量在进行测序的序列的数目确定的情况下，又反过来提高了序列的测序深度（例如，针对每个序列，可以进行多次测序），从而确保了测序结果的可靠性。如本文所使用的，术语 "测序深度" 是指一段 DNA 序列在测序数据中集中出现的次数。测序深度可以通过将测序量除以基因组长度来计算，例如测序深度为 10，表示测了 10 次的整个基因组。

Solexa 测序法的应用十分广泛。其可以用于基因组测序，基因分型，基因多态性研究等等。本发明的方法将 Solexa 测序法用于检测 HPV：通过对待分析的样品进行针对 HPV 的测序，然后使用本领域已知的比对程序，例如 BLAST 和 SOAP，将所得的测序结果与 HPV 数据库中的参考序列进行比对，从而实现对样品所感染的 HPV 的精确分型。本文中使用的 HPV 数据库包含本领域已知



HPV 类型的序列，所述序列可见于例如公共数据库，例如 NCBI 数据库 (<http://www.ncbi.nlm.nih.gov/>)。

如本文中可互换使用的，术语 "PCR 标签 (PCR index)"、"标签 (index)" 或 "引物标签 (primer index)" 是指添加在 PCR 引物 5' 末端的一小段碱基序列，其通过 PCR 扩增可以用于标记 PCR 产物，从而辨别不同模板来源的 PCR 产物的混合物中各个 PCR 产物的模板来源。通过在引物的 5' 末端添加标签，可以对 PCR 产物进行标记，从而可以将多个不同的 PCR 产物混合成一个文库，用于进一步的分析和处理。文库中各个不同的 PCR 产物各自具有独特的标签，从而根据各个 PCR 产物中独特的标签，可以将各个不同的 PCR 产物互相区分开来，并将其与 PCR 模板——对应。

例如，当需要对多个样品进行测序时，可以在用于各个样品的引物的 5' 末端添加不同的标签，然后用添加了标签的引物分别对各个样品进行 PCR 反应，从而对各个样品（即，PCR 产物）进行标记。PCR 反应后，可以将来自各个样品的带有不同标签的 PCR 产物混合在一起组成一个文库，然后应用高通量的 Solexa 测序法同时对文库中的各个 PCR 产物进行测序。最终，在所得的测序数据中，通过独特的标签，可以将测序结果与各个 PCR 产物（从而样品模板）——对应。

可以仅在用于 PCR 扩增的引物对的一条引物中引入标签，也可以在引物对的两条引物中都引入标签。当在引物对的两条引物中都引入标签时，每个 PCR 引物对与一对标签组合成一对标签引物，其中正向和反向 PCR 引物的 5' 端分别具有正向标签和反向标签，并且正反标签和正反引物序列是对应的，且正向标签和反向标签可以是相同的，或不同的。

设计标签时需要考虑多种因素，包括：1) 标签序列中应当

超

免 3 个或 3 个以上的单碱基重复序列；2) 所有标签的同一位点中碱基 A 和碱基 C 的总含量应在所有碱基含量的 30%-70%之间，例如，当设计 100 条不同的标签序列时，每一条标签序列的第二个碱基（即，所谓的同一位点）中 A 和 C 占该 100 条序列第二个碱基总量的 30%-70%；3) 标签序列本身的 GC 含量应在 40-60%之间；4) 标签之间的序列差异应大于 4 个碱基；5) 标签序列中应避免出现与用于测序的引物相似度高的序列；6) 当标签序列添加到 PCR 扩增引物上后，应避免 PCR 扩增引物形成发卡结构和二聚体等二级结构。

如本文中所使用的，术语“标签引物（index primer）”是指带有标签的引物，其包含 2 个部分，标签部分和引物部分，其中标签部分用于在 PCR 扩增反应中标记 PCR 产物，而引物部分与模板碱基互补配对，用于扩增模板，并且其中标签部分，任选地通过连接序列，连接至引物部分的 5' 端。

如本文中所使用的，术语“接头（adapter）”或“文库接头（library adapter）”是指经设计的一段碱基序列，其可以连接至文库中的扩增后的 PCR 产物上，从而文库中的所有扩增后的 PCR 产物都可以借助该接头进行测序，例如，使用针对该接头设计的测序引物（而无需使用针对各 PCR 产物设计的特异性测序引物）进行测序。优选地，本发明的接头可以通过“PCR-FREE”方法连接至 PCR 产物上。

如本文中所使用的，术语“PCR-FREE”是指不进行 PCR 反应而直接将接头连接至 PCR 产物上，例如通过使用 DNA 连接酶将接头连接至 PCR 产物上。利用 PCR-Free 方法来构建测序文库是本领域技术人员已知的，参见例如，Nature Methods 6, 291-295 (2009)。“PCR-FREE”方法由于整个过程无需进行 PCR 而具有以

下有利方面：1) 减少了纯化步骤，降低了花费的时间和成本；2) 减少了非特异性扩增；3) 在含有许多序列高度相似的 PCR 产物的文库的构建过程中，避免由 PCR 引入错误，从而提高最后的测序结果的准确性。

如本文中所使用的，本发明的方法和试剂盒可以使用至少 1 种接头。不同的接头可以共有相同的一段序列（在本文中称为“测序序列”），并且可进一步包含不同的特征序列，从而，不同的接头可以共用相同的引物（其针对相同的测序序列进行设计）进行测序，并且利用独特的特征序列可以辨别多个文库的混合物中各个 PCR 产物的文库来源，即，对不同文库来源的 PCR 产物进行进一步的标记。

将标签与具有不同特征序列的接头组合，可以大大提高标记的效率（参见图 1）。例如，使用 100 种标签可以标记 100 个样品，而使用 100 种标签和 200 种具有不同特征序列的接头，可以标记 $100 \times 200 = 20000$ 个样品。

因此，本发明的一个方面提供了一组标签，其包含至少 10 种，优选至少 20 种、至少 30 种、至少 40 种、至少 50 种、至少 60 种、至少 70 种、至少 80 种、至少 90 种或 95 种标签，并且所述标签具有选自 SEQ ID NO: 1-95 的序列。在一个优选实施方案中，该组标签至少包括 SEQ ID NO: 1-10，或 SEQ ID NO: 11-20，或 SEQ ID NO: 21-30，或 SEQ ID NO: 31-40，或 SEQ ID NO: 41-50，或 SEQ ID NO: 51-60，或 SEQ ID NO: 61-70，或 SEQ ID NO: 71-80，或 SEQ ID NO: 81-90，或 SEQ ID NO: 91-95 所示的标签，或者其任何两个或者多个的组合，例如 SEQ ID NO: 1-95 所示的标签。

在本发明的一个优选实施方案中，本发明的标签用于标

SEQ ID NO: 96-1 06 所示的 PCR 引物，从而用于进行高通量 HPV 测序、检测或分型。

本发明的一个方面提供了一种标签引物组，其包含 11 种标签引物，所述标签引物的序列包含标签序列和 PCR 引物序列，并且所述标签序列，任选地通过连接序列，连接至所述 PCR 引物序列的 5' 端，其中

1) 所述标签序列选自 SEQ ID NO: 1-95，且标签引物组中的 11 种标签引物的每一种的标签序列都相同，和

2) 所述 11 种标签引物的 PCR 引物序列分别如 SEQ ID NO: 96-1 06 所示。

本发明的标签引物组可扩增出至少 16 种大约 170bp 的产物，其对应于 HPV 基因组中最为保守的基因区 (L1 区) 中的一段高度保守的 DNA 序列。因此，本发明的标签引物组可用于 HPV 的精确分型。

在一个优选实施方案中，本发明的标签引物组可用于 HPV 测序、检测或分型，从而其可用于医疗用途，例如诊断 HPV 的存在和确定 HPV 的类型等，以及非医疗用途，例如构建 HPV 数据库，鉴定 HPV 的新的类型和亚型，研究 HPV 类型分布的地域性特点，流行病学研究和疫苗研制等。在另一个优选实施方案中，本发明的标签引物组可用于制备试剂盒，所述试剂盒可用于 HPV 测序、检测或分型。

本发明的另一个方面提供了一种标签引物套组，其包含至少 10 个，优选至少 20 个、至少 30 个、至少 40 个、至少 50 个、至少 60 个、至少 70 个、至少 80 个、至少 90 个或 95 个上文描述的标签引物组。优选地，标签引物套组中，各个标签引物组所使用的标签序列互不相同。更优选，标签引物套组中所使用的标签序

列至少包括 SEQ ID NO: 1-10, 或 SEQ ID NO: 11-20, 或 SEQ ID NO: 21-30, 或 SEQ ID NO: 31-40, 或 SEQ ID NO: 41-50, 或 SEQ ID NO: 51-60, 或 SEQ ID NO: 61-70, 或 SEQ ID NO: 71-80, 或 SEQ ID NO: 81-90, 或 SEQ ID NO: 91-95 所示的标签序列, 或者它们任何两个或者多个的组合, 例如 SEQ ID NO: 1-95 所示的标签序列。

在一个优选实施方案中, 本发明的标签引物套组可用于高通量 HPV 测序、检测或分型, 从而其可用于医疗用途, 例如大规模的 HPV 相关疾病的诊断和 HPV 的精确分型 (为临床诊断和治疗方案选择提供依据) 等, 以及非医疗用途, 例如构建 HPV 数据库, 鉴定 HPV 的新的类型和亚型, 研究 HPV 类型分布的地域性特点, 流行病学研究和疫苗研制等。在另一个优选实施方案中, 本发明的标签引物套组可用于制备试剂盒, 所述试剂盒可用于高通量 HPV 测序、检测或分型。

本发明的另一个方面提供了一种试剂盒, 其包含上文描述的标签引物组或标签引物套组。优选, 本发明的试剂盒还包含至少 1 种, 优选至少 2 种、至少 10 种、至少 20 种、至少 30 种、至少 40 种、至少 50 种、至少 100 种或至少 200 种接头。在一个优选实施方案中, 所述接头适用于 Solexa 测序法, 例如其可用于构建测序文库, 例如所述接头可具有选自 SEQ ID NO: 121-132 的序列。在一个优选实施方案中, 通过 PCR-FREE 方法, 例如 DNA 连接酶法, 使用接头来构建测序文库。

在一个优选实施方案中, 本发明的试剂盒可用于高通量 HPV 测序、检测或分型, 并且如上所述, 其可用于医疗用途和非医疗用途。

本发明的另一个方面提供了用于对一个或多个样品进行 HPV 测序、检测或分型的方法。所述方法包括使用上文描述的标签引物组或标签引物套组或试剂盒对各个样品的 DNA 进行扩增，然后进行测序以获得样品的序列的步骤。

本发明的另一个方面提供了用于对一个或多个样品进行 HPV 测序、检测或分型的方法，其包括下列步骤：

提供 n 个样品， n 为大于等于 1 的整数，所述样品优选地来自哺乳动物，更优选是人，且优选是脱落细胞；可选地，将待分析的 n 个样品分成 m 个小组， m 为整数且 $n \geq m \geq 1$ ；

1) 对于每一个样品，使用一个标签引物组对样品的 DNA 进行扩增，其中

所述标签引物组包含 11 种标签引物，所述标签引物的序列包含标签序列和 PCR 引物序列，并且所述标签序列，任选地通过连接序列，连接至所述 PCR 引物序列的 5' 端，其中

i) 所述标签序列选自 SEQ ID NO: 1-95，并且所述 11 种标签引物的每一种的标签序列都相同，和

ii) 所述 11 种标签引物的 PCR 引物序列分别如 SEQ ID NO: 96-106 所示，

其中，不同的样品使用的标签引物组可以相同或者不同，并且不同的标签引物组使用不同的标签序列；

2) 将步骤 1) 中的使用不同标签引物组进行扩增所获得的扩增产物混合在一起，得到一个或多个 PCR 产物文库；

3) 通过 PCR-FREE 方法，例如 DNA 连接酶法给步骤 2) 中获得的一个或多个 PCR 产物文库分别添加接头，从而构建一个或多个测序文库，其中不同的测序文库所使用的接头可以相同或者不同，且不同的接头共有相同的测序序列，但具有不同的特征序列

4) 任选地, 将步骤 3) 中获得的使用不同接头的测序文库混合在一起, 得到一个或多个文库混合物;

5) 对步骤 3) 中获得的一个或多个测序文库或步骤 4) 中获得的一个或多个文库混合物分别进行测序, 其中利用二代测序技术, 优选 Pair-End 技术 (例如 Solexa、Illumina Hiseq 2000) 进行测序;

6) 根据标签引物组的标签引物序列, 或者根据标签引物组的标签引物序列以及接头的特征序列, 将测序结果与样品——对应; 其中所述样品优选是脱落细胞, 且优选来源于动物, 例如人。

在优选的实施方案中, 本发明的方法使用至少 10 个, 优选至少 20 个、至少 30 个、至少 40 个、至少 50 个、至少 60 个、至少 70 个、至少 80 个、至少 90 个或 95 个上文描述的标签引物组。进一步优选地, 所使用的标签序列至少包括 SEQ ID NO: 1-10, 或 SEQ ID NO: 11-20, 或 SEQ ID NO: 21-30, 或 SEQ ID NO: 31-40, 或 SEQ ID NO: 41-50, 或 SEQ ID NO: 51-60, 或 SEQ ID NO: 61-70, 或 SEQ ID NO: 71-80, 或 SEQ ID NO: 81-90, 或 SEQ ID NO: 91-95 所示的标签序列, 或者它们任何两个或者多个的组合, 例如 SEQ ID NO: 1-95 所示的标签序列。

在本发明的方法的一个优选实施方案中, 本发明的方法使用至少 1 种, 优选至少 2 种、至少 10 种、至少 20 种、至少 30 种、至少 40 种、至少 50 种、至少 100 种或至少 200 种接头, 例如所述接头可具有选自 SEQ ID NO: 121-132 的序列。

在本发明的方法的一个优选实施方案中, 在测序后, 将获得的样品序列与上文描述的 HPV 数据库中的序列进行对比, 从而对样品进行 HPV 精确分型。

在本发明的另一个方面, 本发明基于 Solexa 测序法提供了



于对多个样品进行高通量 HPV 测序、检测或分型的方法，所述样品优选是脱落细胞，且优选来源于动物，例如人，所述方法包括下列步骤：

- 1) 将待分析的样品分成 m 个小组， m 为 ≥ 1 的整数；
- 2) 对每个样品小组进行如下步骤：
 - 2a) 从待分析的样品中提取 DNA；
 - 2b) 根据用于扩增 HPV DNA 的引物组中的所有引物的序列，设计一套标签，标签的个数 n 等于该样品小组的样品数目；
 - 2c) 将 2b) 中设计的每一个标签分别添加至引物组的所有正向引物或所有反向引物或所有引物的序列的 5' 末端，从而提供 n 个标签引物组；
 - 2d) 使用 2c) 中提供的标签引物组对 2a) 中获得的样品 DNA 进行 PCR 扩增，从而提供 PCR 产物，其中针对每个样品 DM，使用不同的标签引物组；和
 - 2e) 将 2d) 中的所有 PCR 产物混合在一起，获得 PCR 产物文库；
- 3) 给 2) 中得到的 PCR 产物文库加上接头，其中各个 PCR 产物文库使用不同的接头，从而构建 m 个测序文库，其中各个接头共有相同的测序序列，而具有互不相同的特征序列；
- 4) 将 m 个测序文库混合在一起，利用二代测序技术，优选 Pair-End 技术（例如 Solexa> Illumina Hiseq 2000）进行测序；得到所有样品的测序结果；
- 5) 根据测序结果中的接头的特征序列、标签的序列以及引物的序列，将获得的测序结果与样品 对应；和任选地，将每个样品的测序结果与 HPV 数据库进行比对，从而实现 HPV 测序、检测或分型。

在一个优选实施方案中，使用本领域技术人员熟知的方法进行 DNA 提取。例如，可以使用自动化的 DNA 提取仪和 DNA 提取试剂盒进行 DNA 提取，例如可商购获得的 KingFisher 自动提取仪，例如美国 Thermo Scientific Kingfisher Flex 全自动磁珠提取纯化系统。

在一个优选的实施方案中，2b) 中的引物组包含 11 条引物，其序列分别如 SEQ ID NO: 96-106 所示。这 11 条引物组成的引物组可扩增出至少 16 种大约 170bp 的产物，其对应于 HPV 基因组中最为保守的基因区 (L1 区) 中的一段高度保守的 DNA 序列。因此，通过对该扩增产物进行精确测序，可以实现 HPV 的精确分型。

在又一个优选的实施方案中，2b) 中设计的标签的个数为至少 10 个，优选至少 20 个，至少 30 个，至少 40 个，至少 50 个，至少 60 个，至少 70 个，至少 80 个，至少 90 个，或至少 100 个。优选，标签可具有选自 SEQ ID NO: 1-95 的序列。在优选的实施方案中，不同样品小组之间的所使用的标签可以相同或者不同。在优选的实施方案中，引入正向引物的标签与引入反向引物的标签可以相同或者不同。在特别优选的实施方案中，2b) 中设计的标签至少包括 SEQ ID NO: 1-10，或 SEQ ID NO: 11-20，或 SEQ ID NO: 21-30，或 SEQ ID NO: 31-40，或 SEQ ID NO: 41-50，或 SEQ ID NO: 51-60，或 SEQ ID NO: 61-70，或 SEQ ID NO: 71-80，或 SEQ ID NO: 81-90，或 SEQ ID NO: 91-95 所示的标签，或者它们任何两个或者多个的组合。在特别优选的实施方案中，2b) 中设计的标签如 SEQ ID NO: 1-95 所示。

在一个优选的实施方案中，使用 PCR-FREE 方法，给 PCR 产物文库加上接头，例如使用 DNA 连接酶。特别地，在本发明的方法中，由于不同 HPV 类型之间的 DNA 序列高度相似，因此，本发

的测序文库的构建必须通过 PCR-FREE 方法来完成。相反,如果通过采用常规的 pool ing PCR 将接头连接至 PCR 产物上来构建测序文库,那么所得到的文库中将含有大量的与原模板不一致的产物,导致不能对原模板进行准确测序。在一个优选的实施方案中,使用的接头的数目为至少 1 种,优选至少 2 种、至少 10 种、至少 20 种、至少 30 种、至少 40 种、至少 50 种、至少 100 种或至少 200 种,例如所述接头可具有选自 SEQ ID NO: 121-1 32 的序列。

在本发明的方法的优选实施方案中,接头是可商购获得的接头,例如可购自 Illumina 公司的 PCR-free Index Adapter 01 igo Mix。在另外的实施方案中,本发明也可使用下列 PCR-free 接头(下划线部分为接头的特征序列)。

PCR-free 接头 1 (SEQ ID NO: 121) : ATCACG

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG

PCR-free 接头 2 (SEQ ID NO:122) : CGATGT

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG

PCR-free 接头 3 (SEQ ID NO:123) : TTAGGC

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG

PCR-free 接头 4 (SEQ ID NO: 124) : TGACCA

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTGACCAATCTCGTATGCCGTCTTCTGCTTG

PCR-free 接头 5 (SEQ ID NO:125) : ACAGTG

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGC
TTG

PCR-free 接头 6 (SEQ ID NO:126) : GCCAAT

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTG

TTG

PCR-free 接头 7 (SEQ ID NO:127) : CAGATC

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTG

TTG

PCR-free 接头 8 (SEQ ID NO:128) : ACTTGA

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGAATCTCGTATGCCGTCTTCTG

TTG

PCR-free 接头 9 (SEQ ID NO:129) : GATCAG

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTG

TTG

PCR-free 接头 10 (SEQ ID NO:130) : TAGCTT

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTTCTG

TTG

PCR-free 接头 11 (SEQ ID NO:131) : GGCTAC

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTTCTG

TTG

PCR-free 接头 12 (SEQ ID NO: 132) : CTTGTA

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGTCTTCTG

TG

在一个优选的实施方案中,本发明的方法使用 Solexa 测序仪 (例如, Illumina Genome Analyzer IIx 测序仪) 进行 Solexa 测序法。在另外的优选的实施方案中, HPV 数据库包含本领域已知的 HPV 类型的序列, 所述序列例如可见于例如公共数据库, 例如 NCBI 数据库 (<http://www.ncbi.nlm.nih.gov/>) 。

在本发明的方法的一个优选实施方案中, 样品可以是脱落细胞。在另一个优选实施方案中, 样品可以来源于动物, 优选哺乳动物, 更优选人。

发明的有益效果

本发明的新型的 HPV 检测方法和用于此的试剂盒相对于现有技术具有以下有利方面。

1) 高通量。通过使用标签与具有不同特征序列的接头, 进行一次本发明的方法, 可检测甚至 10000 份样品, 从而本发明的方法可以广泛用于疾病普查工作, 成为早期诊断疾病的有效手段。

2) 低成本。本发明采用 Solexa 测序法进行测序, 测序成本大大降低 (仅为常规测序法的 1%), 从而 HPV 检测成本大大降低。

3) 可对 HPV 进行精确分型。通过使用多条引物 (例如本发明的 6 条正向引物和 5 条反向引物) 进行扩增, 并将扩增产物的序列信息与 HPV 数据库进行比对, 可以精确地确定 HPV 的类型, 从而为临床诊断和治疗方案选择提供依据。

另外, 使用本发明的方法还有利于发现新的 HPV 类型, 包括已有的类型中的新的亚型和变异体, 为科学研究提供更有效和方便的工具。

下面将结合附图和实施例对本发明的实施方案进行详细描述, 但是本领域技术人员将理解, 下列附图和实施例仅用于说

本发明，而不是对本发明的范围的限定。根据附图和优选实施方案的下列详细描述，本发明的各种目的和有利方面对于本领域技术人员来说将变得显然。

附图说明

图 1 为经标签和具有独特的特征序列的接头标记后的 PCR 产物的示意图。在本发明的示例性方法中，通过 PCR 在每个样品的 PCR 产物两端同时引入标签；把多个带有不同标签的 PCR 产物混合在一起，用于构建测序文库。在测序文库的构建过程中，当需要时，可以构建多个测序文库时，其中通过使用具有不同特征序列的接头，来标记各个测序文库。文库构建完毕后，经具有不同特征序列的接头标记的多个测序文库可以混合在一起，并用 Solexa 测序法进行同时测序（不同的测序文库之间所使用的标签可以相同或不同）。最后，根据测序结果中的接头的特征序列和标签的序列信息，可将测序结果与每个样品 对应。

图 2 为部分 PCR 产物的琼脂糖凝胶电泳图。从电泳图上可观察到，PCR 产物条带的大小为约 170bp。其中，泳道 M 是 50bp DNA 分子梯，泳道 1-14 为随机挑选的 HPV 阳性样本的 PCR 产物。

具体实施方式

采用本发明的方法，对 190 份已知 HC- II 结果的样品进行 HPV 基因分型，结果发现：采用本发明的方法所得的结果不仅与已知的 HC- II 结果一致，而且实现了对 HPV 的精确分型。

实施例 1：样品提取

根据厂商的说明书，使用 KingFisher 自动提取仪（美国

Thermo Scientific Kingfisher Flex 全自动磁珠提取纯化系统)从 190 份已知 HC- II 结果的脱落细胞中提取 DNA。使用程序 "Bioeasy_2 00 μ l Blood DNA.KF. msz", 进行核酸提取。程序结束后, 获得 100 μ l 左右的洗脱产物 (提取的 DNA), 用作下一步 PCR 扩增中的模板。

实施例 2: PCR 扩增

把实施例 1 中获得的 190 份 DNA 依次编号为 1-190, 并平均分为 2 组 (HPV-1 组: 编号 1-95; HPV-2 组: 编号 96-190)。根据用于扩增 HPV DNA 的引物组 (包括 6 条正向引物和 5 条反向引物) 的各条引物的序列 (表 2, SEQ ID NO: 96-107), 设计一套标签, 共 95 个 (表 1, SEQ ID NO: 1-95)。将设计的每一个标签分别添加至引物组的各条引物的序列的 5' 末端, 从而获得 95 个标签引物组, 其中每个标签引物组包括相应的 6 条正向标签引物和 5 条反向标签引物, 并且不同的标签引物组使用不同的标签 (即, 95 个标签引物组与 95 个标签 对应)。

在 96 孔板中对所有样品进行 PCR 反应, 共使用 2 块板 (HPV-1 组和 HPV-2 组各 1 块板)。使用实施例 1 中获得的 DNA 作为模板, 并且在 HPV-1 组和 HPV-2 组 (各 95 个样品) 中, 针对每个样品, 使用不同的标签引物组进行 PCR 扩增 (即, 95 个样品与 95 个标签引物组 对应)。记录下每一个标签引物组 (每一个标签) 对应的样品的编号信息。每块板中还设置一个不添加模板的阴性对照。2 块板中的阴性对照所用的引物分别与样品 1 和样品 96 所用的引物相同。

表 1: 标签与样品的相关信息。

标签编号	标签序列	对应 96 孔板位置	对应样品(组 1)	对应样品(组 2)	SEQ ID NO:
PI-1	GCTGCGACTC	A1	1	96	1
PI-2	GTGTAGATAC	A2	2	97	2
PI-3	CTGATATCTA	A3	3	98	3
PI-4	ACGATGCTAT	A4	4	99	4
PI-5	TAGACTAGAC	A5	5	100	5
PI-6	CTGTCTGTGT	A6	6	101	6
PI-7	GCATACTGAC	A7	7	102	7
PI-8	CTGCTCGCAT	A8	8	103	8
PI-9	CATGAGTAGA	A9	9	104	9
PI-10	TCTCACTATG	A10	10	105	10
PI-11	TGTACTACTA	A11	11	106	11
PI-12	GTAGACTAGT	A12	12	107	12
PI-13	ATATGCTACT	B1	13	108	13
PI-14	CACTCGCTGT	B2	14	109	14
PI-15	CATCACGCAC	B3	15	110	15
PI-16	AGCATGTGAT	B4	16	111	16
PI-17	AGCTAGTAGA	B5	17	112	17
PI-18	GCTATGTAGT	B6	18	113	18
PI-19	TACGATGATG	B7	19	114	19
PI-20	TACGCTGTAC	B8	20	115	20

PI-21	TATGTGTA	B9	21	116	21
PI-22	TGACTCAGAC	B10	22	117	22
PI-23	TGTAAGCTCA	B11	23	118	23
PI-24	GAGACTCGTA	B12	24	119	24
PI-25	CTAGATGTC	C1	25	120	25
PI-26	GATGACTCTC	C2	26	121	26
PI-27	TGAGTGGCAC	C3	27	122	27
PI-28	TGTAGTGA	C4	28	123	28
PI-29	TCATCGTAGA	C5	29	124	29
PI-30	TAGCATCTGT	C6	30	125	30
PI-31	TAGTAGTCGT	C7	31	126	31
PI-32	CTATACGTGC	C8	32	127	32
PI-33	CGACTGTAGA	C9	33	128	33
PI-34	GATGTCATGT	C10	34	129	34
PI-35	GTCGACTG	C11	35	130	35
PI-36	AGCTGACGAT	C12	36	131	36
PI-37	ATGATATAGT	D1	37	132	37
PI-38	ATGTGCTCTA	D2	38	133	38
PI-39	CTGACTCGAT	D3	39	134	39
PI-40	GCTGGAGCTC	D4	40	135	40
PI-41	GAGTCATGTC	D5	41	136	41
PI-42	CATACGCTCA	D6	42	137	42

PI-43	CACTCTCGTC	D7	43	138	43
PI-44	GCACTAGATG	D8	44	139	44
PI-45	AGTACGCATG	D9	45	140	45
PI- 46	TCTGTGACGT	D10	46	141	46
PI- 47	TAGCTCATCT	D11	47	142	47
PI-48	AGCATACACT	D12	48	143	48
PI-49	GCTATAGTCA	E1	49	144	49
PI- 50	CGTCTCATGC	E2	50	145	50
PI-51	GCTACTACGT	E3	51	146	51
PI-52	GAGTGTAATA	E4	52	147	52
PI- 53	GTCATACGTG	E5	53	148	53
PI-54	TATGAGAGAT	E6	54	149	54
PI-55	ATCTGAGTAC	E7	55	150	55
PI-56	CGATAGCATC	E8	56	151	56
PI-57	ACTGATCTCA	E9	57	152	57
PI- 58	CTCGATACTA	E10	58	153	58
PI- 59	CATGTGACTG	E11	59	154	59
PI-60	CGCATCACTA	E12	60	155	60
PI-61	GCAATATATCT	F1	61	156	61
PI - 62	CTGATGCGAC	F2	62	157	62
PI-63	TCTCAGAGTC	F3	63	158	63
PI-64	CAGTGCGAGT	F4	64	159	64

PI-65	ATGTCTGATG	F5	65	160	65
PI-66	GCTAGTAGTC	F6	66	161	66
PI-67	ATGAGTCGTC	F7	67	162	67
PI-68	ATCACTCAGA	F8	68	163	68
PI-69	TCTCTCTGAT	F9	69	164	69
PI-70	CTCTAGTGCT	F10	70	165	70
PI-71	CGTCGTGCTA	F11	71	166	71
PI-72	CGACTACTAT	F12	72	167	72
PI-73	GCACGTCGAT	G1	73	168	73
PI-74	GTAGTGCTCT	G2	74	169	74
PI-75	CTGACGAGCT	G3	75	170	75
PI-76	CTATAGTCTA	G4	76	171	76
PI-77	ACACGCACTA	G5	77	172	77
PI-78	CTCGCACTAC	G6	78	173	78
PI-79	AGATCTCACT	G7	79	174	79
PI-80	ATACTAGTGT	G8	80	175	80
PI-81	ATATCTCGTA	G9	81	176	81
PI-82	TGACTGCGTA	G10	82	177	82
PI-83	TGTAGACGTA	G11	83	178	83
PI-84	AGAGACTATG	G12	84	179	84
PI-85	GTCGAGTCAC	H1	85	180	85
PI-86	TGACAGCTAC	H2	86	181	86

PI-87	CGCTAGACAT	H3	87	182	87
PI-88	CGTAGATATG	H4	88	183	88
PI-89	TGAGTCTGCT	H5	89	184	89
PI-90	TAGTCGTATG	H6	90	185	90
PI-91	CATACACGAC	H7	91	186	91
PI-92	CGCTCAGAGA	H8	92	187	92
PI-93	GTGAGTCTCA	H9	93	188	93
PI-94	GACAGATGAT	H10	94	189	94
PI-95	GCTGTGCGAC	H11	95	190	95

表 2: 未添加标签的用于扩增 HPV DNA 的引物组的各条引物的序列信息。

引物编号	引物序列	SEQ ID NO:
F1	TTTGTTACTGTGGTAGATACTAC	96
F2	TTTGTTACTGTGGGATACTAC	97
F3	TTTGTTACCGTTGTGATACTAC	98
F4	TTTGTTACTAAGGTAGATACCACTC	99
F5	TTTGTTACTGTTGTGGATACAAC	100
F6	TTTGTTACTATGGTAGATACCACAC	101
R1	GAAAAATAAACTGTAAATCATATTCCT	102
R2	GAAAAATAAATTGTAAATCATACTC	103
R3	GAAATATAAATTGTAAATCAAATTC	104
R4	GAAAAATAAACTGTAAATCATATTC	105
R5	GAAAAATAAACTGCAAATCATATTC	106

注：F 表示正向引物，R 表示反向引物。

使用下列 PCR 参数进行扩增：

95℃ 30 秒 → 48℃ 30 秒 → 72℃ 30 秒 (40 个循环)

72℃ 10 分钟 → 12℃ ∞

PCR 反应体系为 25μl, 其组成是 (所有试剂均购自 Enzymatics 公司)：

试剂	体积/反应
水 (HPLC 级)	14.375μl
10x Ex Taq 缓冲液 (Mg ²⁺ plus)	2.5μl
dNTP 混合物 (各 2.5 mM)	2μl
带有标签的 F1/F2/F3/F4/F5/F6 的混合物 (各 7.5pmol)	0.5μl
带有标签的 R1/R2/R3/R4/R5 的混合物 (各 7.5pmol)	0.5μl
Ex Taq HS (5U/μl)	0.125μl
模板 DNA	5μl
总体积	25μl

PCR 反应在 Bio-Rad 公司的 PTC-200 PCR 仪上运行。PCR 完成后, 取 3μl PCR 产物在 2.5% 的琼脂糖凝胶上进行电泳检测 (图 2)。

实施例 3: PCR 产物的混合和纯化

把 HPV-1 组与 HPV-2 组中剩余的 PCR 产物各混合在一个 3ml 的 EP 管中 (同样标记为 HPV-1 组和 HPV-2 组), 并震荡混匀。从 2 管混合物中各取出 500μl DNA, 并根据厂商的说明书, 使用 Qiagen DNA Purification 试剂盒进行过柱纯化, 得到 200μl DNA。

使用 Nanodrop 8000 (Thermo Fisher Scientific 公司), 测定纯化后的混合物的 DNA 浓度分别为 98 ng/ μ i (HPV-1 组) 和 102 ng/ μ l (HPV-2 组)。

实施例 4: Solexa 测序文库的构建

4.1: 末端修复反应

使用 Thermomixer (Eppendorf 公司) 对实施例 3 中获得的经纯化的两管 DNA 混合物分别进行 DNA 末端修复反应。修复反应的反应体系为 100 μ l, 其组成是 (所有试剂均购自 Enzymatics 公司) :

试剂	体积/反应
上一步所得的 DNA	75 μ l
20x 多核苷酸激酶缓冲液 (B904)	10 μ L
dNTP 混合物 (各 20mM)	4 μ L
T4 DNA 聚合酶	5 μ L
Klenow 片段	1 μ L
T4 多核苷酸激酶	5 μ L
总体积	100 μ L

反应条件为 : 20℃ , 30 分钟。

根据厂商的说明书, 使用 QIAquick PCR Purification 试剂盒纯化并回收 DNA 末端修复反应的产物。回收的产物溶于 34 μ i EB (QIAGEN Elution Buffer) 中。

4.2: 3' 末端加 A 反应

使用 Thermomixer (Eppendorf 公司), 对回收的 DNA 进行 3' 末端加 A 反应。反应体系为 50 μ l, 其组成是 (所有试剂均购自 Enzymatics 公司) :

试剂	体 积 / 反 应
上一步所得的 DNA	32 μ l
dATP (1mM, GE 公司)	10 μ l
10x Blue 缓冲液	5 μ l
Klen w (3'-5'exo-)	3 μ l
总体积	50 μ l

反应条件为：37°C，30 分钟。

根据厂商的说明书，使用 MiniElute PCR Purification Kit (QIAGEN 公司) 纯化并回收 3' 末端加 A 反应的产物。回收的产物溶于 20 μ l 的 EB 中。

4.3: 添加 Solexa 接头

使用 Thermomixer (Eppendorf 公司) 对上一步获得的 2 种产物分别添加不同的接头以构建 2 个测序文库。记录下接头和文库的对应关系。

添加 Solexa 接头的反应体系为 50 μ l，其组成是（所有试剂均购自 Illumina 公司）：

试剂	体 积 / 反 应
上一步所得的 DNA	11 μ L
2x 快速连接缓冲液	15 μ L
PCR-free Index Adapter Oligo Mix (25mM)	1 μ L
T4 DNA 连接酶(快速, L603-HC-L)	3 μ L
总体积	30 μ L

反应条件为：20°C，15 分钟。

根据厂商的说明书，使用 Ampure Beads (Beckman Coulter Genomic s) 纯化反应产物并将产物溶于 $17\mu\text{l}$ 去离子水。使用 Agilent Bioanalyzer 2100(Agilent 公司 和 荧光定量 PCR(QPCR) 测定产物的 DNA 浓度，结果如下：

	2100 (nM)	qPCR (nM)
HPV-1 组	20.4	24.2
HPV-2 组	21.6	25.8

实施例 5: Solexa 测序

以 Agilent Bioanalyzer 2100 所测浓度为准，将上一步所得的 2 种产物等摩尔混合在一起（各取 1Opmol DNA）。根据厂商的说明书，使用 Solexa 测序仪（Illumina Genome Analyzer II_X 测序仪），用 Solexa PE-75 程序进行测序。

实施例 6: 结果分析

根据测序结果中的接头的特征序列以及标签引物（标签部分和引物部分）的序列的信息，将测序结果与每个样品一一对应。然后，使用本领域已知的比对程序，例如 BLAST 和 SOAP，将每个样品的测序结果与 HPV 数据库进行比对，从而实现 HPV 检测，并对 HPV 进行精确分型。

所得到的检测结果与已知的结果完全一致（参见表 3），表明本发明方法可以用于精确检测样品中的 HPV。

表 3: 190 个样品的检测结果。

样品 编号	原 HC-II 结果 (RLU/CO 值)	本次检测的 HPV 结果	样品 编号	原 HC-II 结果 (RLU/CO 值)	本次检测的 HPV 结果
1	14.2	HPV56	96	0.28	阴性
2	0.31	阴性	97	0.33	阴性
3	196.41	HPV16	98	181.29	HPV35, HPV6
4	5.76	HPV18	99	77.32	HPV16
5	0.35	阴性	100	91.22	HPV39
6	99.86	HPV18, HPV11, HPV16	101	188.92	HPV52
7	128.86	HPV39	102	1352.83	HPV35, HPV11, HPV39
8	35.12	HPV18, HPV6	103	1.39	HPV43
9	498.69	HPV16, HPV56	104	119.5	HPV45, HPV11
10	603.57	HPV18, HPV31, HPV39	105	292.43	HPV56, HPV31
11	0.27	阴性	106	2.91	HPV68
12	3420.57	HPV18	107	193.13	HPV45
13	0.38	阴性	108	2.62	HPV6
14	0.41	阴性	109	94.12	HPV16
15	455.06	HPV16	110	792.72	HPV18, HPV31
16	8.93	HPV18	111	31.76	HPV11
17	0.6	阴性	112	0.25	阴性
18	0.41	阴性	113	0.23	阴性
19	0.29	阴性	114	750.82	HPV56, HPV16
20	27.64	HPV31	115	0.4	阴性
21	1985.41	HPV56, HPV68	116	2.75	HPV31
22	20.71	HPV42	117	396.04	HPV45
23	1795.83	HPV11, HPV16, HPV52	118	354.76	HPV18, HPV16

- u -

24	HPV43	119	6.26	HPV11
25	HPV39	120	1719.67	HPV16, HPV45
26	HPV6	121	76.92	HPV51
27	HPV68, HPV16	122	1318.02	HPV56, HPV16, HPV42
28	HPV44	123	0.28	阴性
29	HPV16	124	0.33	阴性
30	candHPV89	125	181.29	HPV59
31	HPV52, HPV39	126	77.32	HPV68
32	HPV54, HPV33	127	110.8	HPV52
33	HPV42	128	147.25	HPV16
34	HPV16, candHPV89	129	0.24	HPV26
35	HPV35	130	1.55	HPV11, HPV53
36	HPV18	131	2.03	HPV6, HPV66
37	阴性	132	8.45	HPV43
38	HPV11	133	0.2	阴性
39	HPV11, HPV6	134	0.24	阴性
40	HPV42	135	10.53	HPV11
41	阴性	136	1410.85	HPV16, HPV53, HPV70
42	阴性	137	1149.25	HPV56, HPV81, HPV73
43	阴性	138	0.24	阴性
44	HPV6	139	413.9	HPV45
45	HPV16	140	17.05	HPV11
46	阴性	141	23.6	HPV52
47	HPV16, HPV18	142	3379.09	HPV16, HPV35, HPV56
48	阴性	143	0.18	阴性
49	HPV18, candHPV89	144	1.46	HPV18

50	HPV56	145	1.25	HPV11, HPV26
51	阴性	146	2.13	HPV6, HPV81
52	HPV56, HPV68	147	872.52	HPV16, HPV45, HPV52
53	HPV18	148	1.5	HPV18
54	阴性	149	4.33	HPV16
55	阴性	150	0.82	阴性
56	HPV16	151	60.35	HPV59
57	HPV51	152	0.24	阴性
58	阴性	153	0.23	阴性
59	阴性	154	0.18	阴性
60	阴性	155	1.46	HPV51
61	HPV52, HPV16	156	11.25	HPV16
62	HPV42	157	2.13	HPV11
63	阴性	158	0.13	阴性
64	HPV11	159	90.18	HPV58
65	HPV59	160	0.15	阴性
66	阴性	161	602.79	HPV68, HPV16
67	HPV45	162	132.68	HPV56, HPV11
68	阴性	163	127.08	HPV39, HPV54
69	HPV6	164	602.79	HPV33
70	阴性	165	276	HPV18
71	HPV40	166	243.6	HPV45
72	阴性	167	229.44	HPV51
73	HPV43, HPV44	168	1384.92	HPV16, HPV58, HPV72
74	HPV41	169	172.64	HPV58
75	HPV39, HPV6	170	855.24	HPV16, candHPV89

76	602. 79	HPV45	171	126. 47	HPV51
77	276	HPV16	172	86. 62	HPV44, HPV11
78	243. 6	HPV6, HPV70, HPV39	173	879. 37	HPV18, HPV58
79	229. 44	HPV35	174	119. 39	HPV56
80	1384. 92	HPV52, HPV56, HPV11	175	0. 61	阴 性
81	172. 64	HPV26, HPV42	176	18. 02	HPV16
82	855. 24	HPV35, HPV6	177	16. 06	HPV18
83	620. 69	HPV52	178	60. 69	HPV56, HPV11
84	128. 02	HPV11	179	2. 45	HPV11
85	514. 84	HPV33	180	94. 93	HPV39
86	68. 3	HPV58	181	1635. 3	HPV16, HPV35, HPV51
87	402. 15	HPV59, HPV16	182	754. 64	HPV33, candHPV89
88	51. 72	HPV33	183	0. 23	HPV11
89	1. 78	HPV6	184	20. 28	HPV18
90	56. 7	HPV11, HPV31	185	0. 16	阴 性
91	186. 06	HPV16	186	0. 13	阴 性
92	0. 02	阴 性	187	60. 18	HPV59
93	386. 06	HPV18, HPV16	188	0. 15	阴 性
94	28. 09	HPV6, HPV44	189	1. 36	HPV43
95	186. 06	HPV68,	190	0. 28	阴 性

此外，本发明的方法还可以对样品中的 HPV 进行精确分型。

表 4 中提供了图 2 所示的泳道 1- 14 所对应的样品的测序序列和分型结果。

表 4: 图 2 所示的泳道 1-14 所对应的样品的测序序列和分型结果。

泳道 编号	HPV 型别	测序序列	比对参数			SEQ ID NO:
			同一性	评分	e 值	
1	HPV11	GAAAAATAAACTGTAAATCATATTCTCCACATGGCGCATGTATTCCTTATAATCTGAATTA	96.77	107	6.00E-27	107
2	HPV16	TTTGTTACTGTGGTAGATACTACCCGCAGTACAAAATATGTCATTTATGTGCTGCCATATCTAC	96.77	107	6.00E-27	108
3	HPV18	TTTGTTACTAAGGTAGATACCACCTCGCAGTACCAATTTAACAAATATGTGCTTCTACACAGTC	96.77	107	6.00E-27	109
4	HPV31	GAAAAATAAACTGTAAATCATATTCTCCACCATGTCTTAAATACTCTTTTAAAAATTACTACTTT	95.16	99.6	2.00E-24	110
5	HPV33	GAAAAATAAACTGTAAATCATATTCTCTCAACATGTCTTATATATTCTTTTAAAAATTTTCATTT	96.77	107	6.00E-27	111
6	HPV35	GAAAAATAAACTGTAAATCATATTCTTCCACCATGGCCTTAAATAATTTCCTTATAATTGTCAATTT	98.39	115	3.00E-29	112
7	HPV39	TTTGTTACTGTGGTAGATACTACCGGTAGTACCACCTTTACATTTATCTACCTCTATAGAGTC	94.83	91.7	4.00E-22	113
8	HPV43	TTTGTTACTAAGGTAGATACCACCTCGTAGTACAAAACCTTAACGTTATGTGCTTCTACTGACCC	100	101	4.00E-25	114
9	HPV45	GAAAAATAAACTGTAAATCATATTCTCTCCACATGTCTACTATAGTGCTTAAACTTAGTAGGA	100	123	1.00E-31	115
10	HPV51	GAAATATAAAATTGTAAATCAAAATTCTTCCCATGGCCTAATATATTGGCTTAAAGTTACTTGGA	91.94	83.8	9.00E-20	116
11	HPV52	TTTGTTACTGTGGGATACTACTCGTAGCAGCACTAACATGACTTTATGTGCTGAGGTTAAAAA	98	91.7	4.00E-22	117
12	HPV56	TTTGTTACTGTGGGATACAACTAGAAAGTACTAACATGACTATTAGTACTGCTACAGAGAACA	95.16	99.6	2.00E-24	118
13	HPV58	GAAAAATAAACTGTAAATCATATTCTCTCAACATGACGTACATATTCTCTTAAAAATTTATCATTT	96.77	107	6.00E-27	119
14	HPV61	TTTGTTACTGTGGGATACTACCCCGCAGTACTAATTTAGCCATTTGGCACTGCTACATCCCC	94	75.8	2.00E-17	120

参 考 文 献

本文中用于举例说明本发明或提供关于本发明的实施的另外的详细内容的专利、出版物和其他材料通过引用合并入本文，并且为方便起见按下列文献目录提供。

[[11]] . PPeecttaass iiddeess DD., KKaannppooss iioorraass K . PPaappaaxxooliinniiss e eett a ll..
CChheemmooththeerraappy ffoorr rreeccuurrrreenntt cceerrvviiccaall ccaannnceerr.. CCaannnceerr
TTreeaattmeenntt RReevviiwwss,, 22000088,, 3344 ((77)) : 660033--6611 33..

[[22]] . BBriinnkk,, AA.. AA.. PP.. JJ.. SSnnii jjdeerrss,, aanndd CC.. JJ.. MMeeii jjeerr.. HHPPVV
ddeetteecctt iioon mmeethhoodss.. DDiiss.. MMAarrkkeerrss 22000077 2233:: 227733--228811..

[[33]] . I AARRCC.. HHaannddbboookss ooff ccaannnceerr ppreevveenntt iioon.. CCeerrvviixx ccaannnceerr
ssccreeennniing [[RR]] . LLyyoonn:: IIAARRCC PPreessss,, 22000055..

[4]. Doorbar, J. .MMoolleecuullaar bbiioollooggy ooff hhuumaan
papi llomavirus infect ion and cervical cancer. Cl in.
Sci. 2006, 110: 525-541.

[5] . Cox T, Cuzick J. HPV DNA tes ting in cervical cancer
screening: From evidence to policies. Gyneeol Oncol, 2006,
103: 8-11.

[6] . Kulmala S, syl jhen. Human papi llomavirus tes ting
with the hybr id capture assay and PCR as screening
tool s. Cl in Microbiol, 2004, 42 (6) : 2470-2475.

[7] . Quail, M. et al. , A large genome center's
improvements to the I llumina sequencing sys tem. Nat.
Methods, 2008, 5 , 1005-101 0.

[8] . Brown, C. G. et al. , Solexa/1 llumina GAPI pipeline
product and product documentat ion, I llumina Inc, 2006.

[9] . Lozano , R . Successful ly integrat ing human
papi llomavirus tes t ing into your pract ice. Arch. Pathol.
Lab Med, 2003, 127: 991-994.

权 利 要 求

1. 一组标签，其包含至少10种，优选至少20种、至少30种、至少40种、至少50种、至少60种、至少70种、至少80种、至少90种或95种标签，所述标签具有选自SEQ ID NO: 1-95的序列，并且优选所述标签用于标记SEQ ID NO: 96-106所示的PCR引物；优选该组标签至少包括SEQ ID NO: 1-10，或SEQ ID NO: 11-20，或SEQ ID NO: 21-20，或SEQ ID NO: 31-40，或SEQ ID NO: 41-50，或SEQ ID NO: 51-60，或SEQ ID NO: 61-70，或SEQ ID NO: 71-80，或SEQ ID NO: 81-90，或SEQ ID NO: 91-95所示的标签，或者其任何两个或者多个的组合。
2. 权利要求1的一组标签用于进行高通量HPV测序、检测或分型的用途，优选其用于标记SEQ ID NO: 96-106所示的PCR引物。
3. 一种标签引物组，其包含11种标签引物，所述标签引物的序列包含标签序列和PCR引物序列，并且所述标签序列，任选地通过连接序列，连接至所述PCR引物序列的5'端，其中
 - 1) 所述标签序列选自SEQ ID NO: 1-95，并且所述11种标签引物的每一种的标签序列都相同，和
 - 2) 所述11种标签引物的PCR引物序列分别如SEQ ID NO: 96-106所示。
4. 权利要求3的标签引物组的用途，其用于HPV测序、检测或分型，或用于制备试剂盒，所述试剂盒用于HPV测序、检测或分型

5. 一种标签引物套组，其包含至少10个，优选至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个或95个权利要求3的标签引物组，其中，每个标签引物组所使用的标签序列互不相同，优选所使用的标签序列至少包括SEQ ID NO: 1-10，或SEQ ID NO: 11-20，或SEQ ID NO: 21-20，或SEQ ID NO: 31-40，或SEQ ID NO: 41-50，或SEQ ID NO: 51-60，或SEQ ID NO: 61-70，或SEQ ID NO: 71-80，或SEQ ID NO: 81-90，或SEQ ID NO: 91-95所示的标签序列，或者它们任何两个或者多个的组合。

6. 权利要求5的标签引物套组的用途，其用于高通量HPV测序、检测或分型，或用于制备诊断试剂盒，所述试剂盒用于高通量HPV测序、检测或分型。

7. 一种试剂盒，其包含权利要求3的标签引物组或权利要求5的标签引物套组，并且优选还包含至少1种，优选至少2种、至少10种、至少20种、至少30种、至少40种、至少50种、至少100种或至少200种接头，优选所述接头适用于Solexa测序法，例如适用于构建测序文库，例如所述接头可具有选自SEQ ID NO: 121-132的序列。

8. 权利要求7的试剂盒用于高通量HPV测序、检测或分型的用途。

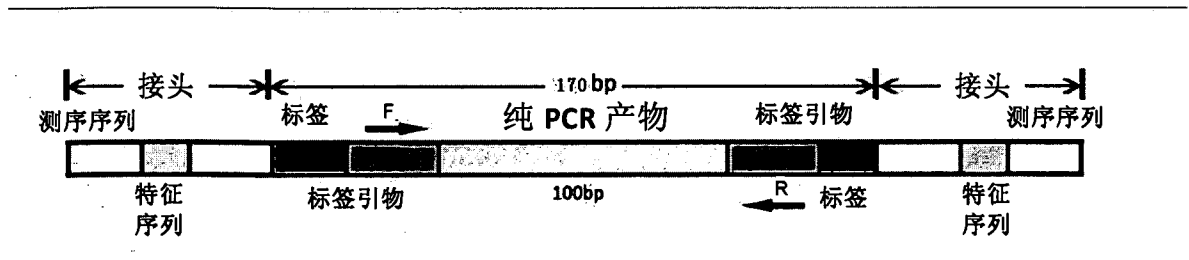


Fig. 1. 经标签和具有独特的特征序列的接头标记后的 PCR 产物的示意图。F: 正向 PCR 引物; R: 反向 PCR 引物。



Fig. 2. 部分 PCR 产物的凝胶电泳图。

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/001812

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12N 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)

WPI, EPODOC, CPRS, CNKI, BIOSIS, GenBank

human papilloma virus, HPV, primer, index, barcode, tag, Solexa, sequence search on SEQ ID NO: 1-106

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	CN101921874A (SHENZHEN BGI TECHNOLOGY CO LTD) 22 Dec. 2010 (22.12.2010) see whole document	1-8
A	CN101435002A (SHENZHEN BGI TECHNOLOGY CO LTD) 20 May 2009 (20.05.2009) see whole document	1-8
A	CN101250593A (GUANGZHOU YISHAN BIOLOGIC TECHNOLOGY CO LTD) 27 Aug. 2008 (27.08.2008) see whole document	1-8
A	Jonas Binladen, et al. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. PLoS ONE, February 2007, Issue 2, e197	1-8



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

11 Mar. 2011 (11.03.2011)

Date of mailing of the international search report

07 Apr. 2011 (07.04.2011)

Name and mailing address of the ISA/CN

The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451

Authorized officer

SU Lin

Telephone No. (86-10) 62411030

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/001812

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

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

a. a sequence listing filed or furnished

☐ on paper

☐ in electronic form

b. time of filing or furnishing

☐ contained in the application as filed

☐ filed together with the application in electronic form

☐ furnished subsequently to this Authority for the purposes of search

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

3. Additional comments:

International application No.
PCT/CN2010/001812

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 18 12

CLASSIFICATION OF SUBJECT MATTER

C12N15/11 (2006.01) i
C12Q1/68 (2006.01) i
C12Q1/70 (2006.01) i

A. 主题的分类

参见附加页

按照国际专利分类(IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: C12N 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))

WPI, EPODOC, CPRS, CNKI, BIOSIS, GenBank

人, 人类, 乳头瘤病毒, 疣病毒, 引物, 标签, 标记, 索引, human papilloma virus, HPV, primer, index, barcode, tag.,
Solexa, SEQ ID NO: 1-106 的序列检索

C. 相关文献

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
E	CN101921874A (深圳华大基因科技有限公司)22. 12 月 2010 (22.12.2010) 参见全文	1-8
A	CN101435002A (深圳华大基因科技有限公司)20.5 月 2009 (20.05.2009) 参见全文	1-8
A	CN101250593A (广州益善生物技术有限公司)27.8 月 2008 (27.08.2008) 参见全文	1-8
A	Jonas Binladen, et al. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. PLoS ONE, 2007 年 2 月, 第 2 期, e197	1-8



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% % 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

11.3 月 2011 (11.03.2011)

国际检索报告邮寄日期

07.4 月 2011 (07.04.2011)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 10008S

传真号: (86-10)62019451

受权官员

苏林

电话号码: (86-10) 62411030

第I栏 ~~核苷酸和/或氨基酸序列~~(续第1页第I.c项)

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

a. 序列表的提交或提供

☐ 纸件形式

IE1 电子形式

b. 提交或提供时间

包含在申请提交时的国际申请中

以电子形式与国际申请一起提交

为检索之用随后提交本单位

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

3. 补充意见

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2010/001812

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN101921874 A	22. 12.2010	无	
CN 101435002 A	20.05.2009	无	
CN 101250593 A	27.08.2008	无	

主题的分类

C12N15/11 (2006.01) i

C12Q1/68 (2006.01) i

C12Q1/70 (2006.01) i