(12) STANDARD PATENT (11) Application No. **AU 2010356776 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Method for detecting Human Papilloma Virus based on Solexa sequencing method

(51) International Patent Classification(s)
C12Q 1/70 (2006.01) ***C12Q 1/68*** (2006.01)

(21) Application No: **2010356776** (22) Date of Filing: **2010.11.15**

(87) WIPO No: **WO12/000151**

(30) Priority Data

(31) Number	(32) Date	(33) Country
201010213722.7	**2010.06.30**	**CN**

(43) Publication Date: **2012.01.05**
(44) Accepted Journal Date: **2014.10.02**

(71) Applicant(s)
BGI Shenzhen Co., Limited

(72) Inventor(s)
Yi, Xin;Xu, Jiajia;Nie, Xifang;Zhao, Meiru

(74) Agent / Attorney
Davies Collison Cave, Level 15 1 Nicholson Street, MELBOURNE, VIC, 3000

(56) Related Art
BINLADEN, J., et al., PLoS ONE, 2007, vol. 2, e197, epages 1-9

(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)

PCT

(10) 国际公布号
WO 2012/000151 A1

(51) 国际专利分类号:
C12Q 1/70 (2006.01) *C12Q 1/68* (2006.01)

(21) 国际申请号: PCT/CN2010/001833

(22) 国际申请日: 2010 年 11 月 15 日 (15.11.2010)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201010213722.7 2010 年 6 月 30 日 (30.06.2010) 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, 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)。 **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)。 **聂喜芳 (NIE, Xifang)** [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)。 **赵美茹 (ZHAO, Meiru)** [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: **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, 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)。

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))。

(54) Title: METHOD FOR DETECTING HUMAN PAPILLOMA VIRUS BASED ON SOLEXA SEQUENCING METHOD

(54) 发明名称 : 基于 SOLEXA 测序法的检测人类乳头瘤病毒的方法

(57) Abstract: The present invention provides a method for detecting Human Papilloma Virus (HPV), especially a method for detecting HPV based on Solexa sequencing method.

(57) 摘要:

本发明提供了人类乳头瘤病毒(HPV)的检测方法，特别是基于 Solexa 测序法的 HPV 检测方法。

WO 2012/000151 A1

Description

A method for detecting Human Papilloma Virus
based on Solexa sequencing method

FIELD OF THE INVENTION

The present invention relates to a method for detecting Human Papilloma Virus (HPV), in particular, to a method for detecting HPV based on Solexa sequencing method.

BACKGROUND OF THE INVENTION

Cervical cancer is the second most common malignant tumor among women in the world, following breast cancer. There are about 500,000 newly reported cases worldwide every year, and nearly 250,000 people die of this disease annually. Cases in developing countries account for approximately 2/3 of the total. Cervical cancer is also prevalent in China. The cases of cervical cancer in China account for 10% of the total. Studies showed that Human Papilloma Virus (HPV) is closely associated with cervical cancer, and is an important carcinogenic factor and one of the prerequisites for causing cervical cancer. It is shown that more than 100 types of HPV can cause infection in skin (skin type) or in mucosa of respiratory tract and anal-genital tract (mucosal type), and more than 40 types of HPV can cause infection in cervices. HPV types are divided into the following two groups depending on the benign, premalignant or malignant pathological changes induced thereby: (1) group of low cancerogenic risk, containing such as types 6, 11, 42, 43, and 44; and (2) group of high cancerogenic risk, containing such as types 16, 18, 31, 33, and 45. Therefore, early detection and correct typing of HPV infections are essential for the prevention and treatment of cervical cancer.

Current HPV detection methods are mainly divided into the following

groups. (1) Cytological examination, which makes diagnosis on the basis of change in cell morphology by using cytological examination of cervical smear or thin-prep cell test (TCT). As for HPV infection, koilocytosis, dyskeratosis and condylomata like basal cell can be seen under microscopy. Its shortcoming lies in low sensitivity and specificity in the diagnosis of HPV infection. (2) Immunohistochemical method, which further confirms HPV infection by detecting the capsid antigen of HPV. The positive reaction thus obtained is definitely located and the result is reliable. However, capsid antigens are not produced until HPV-DNA is replicated and mature. Therefore, subjects diagnosed as negative cannot be taken as unaffected by HPV. The method has low sensitivity. (3) Real-time Fluorescence Quantitative PCR (FQ-PCR), which primarily employs a fluorescence detection PCR Instrument. In the method, a fluorescent group is added to the PCR reaction system, and the amplification product produced in each cycle during PCR is real-time monitored by accumulation of the fluorescent signal, thereby accomplishing the quantitation of the initial concentration of the template. The method has a low throughput. (4) Hybrid capture method (primarily HC-II system), which is the only clinic method for detection of HPV DNA approved by US FDA, and is approved by European CE and Chinese SFDA. The method employs specific specimen collectors and containers, a RNA probe of 8000bp in full length and a specific first antibody, both of which have been granted a patent right. The mechanism for the method is that the nucleic acid probe is hybridized to the HPV DNA of the subject to be tested, and the detection is carried out based on the amplified signals by chemical fluorescence or enzymatic reactions. The nucleic acid probes used in the method are mainly divided into two classes: nucleic acid probes as directed to low-risk HPV and nucleic acid probes as directed to high-risk HPV. The method can be used in primary screening of HPV, but can neither determine the specific type of HPV nor determine multiple infections.

The sensitivity of HPV detection can be increased and the false negative rate can be reduced by using the above detection methods in combination. However, the cost for the combination of these methods was high, and thus the combination of these methods is just feasible in the HPV detection and screening of cervical carcinoma in economy developed areas. For economy less developed areas, especially mountainous areas and most of rural areas, there is a big limitation on using said detection methods in combination. Therefore, there is a need to develop a suitable and low-cost HPV detection method.

In another aspect, the currently known HPV detection methods, such as the detection methods as described above, have low throughput. When conducting HPV detection to samples in large scale, application of said methods are time-consuming and labor-consuming, and the cost thereof is high. Therefore, there is also an urgent need in the art for a new high-throughput and low-cost HPV detection method.

Description of the Invention

The present invention develops a new HPV detection method and a kit for the same, on the basis of Solexa sequencing method and PCR index. The method and kit according to the present invention can not only achieve a high throughput HPV detection with low cost, but also accomplish accurate HPV typing.

Definition

In order to better understand the present invention, the definitions and explanations of the relevant terms are provided herein.

As used in the application, the term "PCR" refers to polymerase chain reaction.

As used in the application, the term "Solexa sequencing method" refers to

a new generation DNA sequencing method, which is developed in current years, also named the second generation sequencing method. The difference between Solexa sequencing method and the traditional sequencing method (such as, Sanger sequencing method) lies in Solexa sequencing method analyses a DNA sequence by conducting sequencing as carrying out synthesis. Solexa sequencing method has the following advantages: 1) low cost, which is 1% of the cost of the t raditional sequencing method; 2) high throughput, capable of conducting sequencing to multiple samples simultaneously, and capable of producing a data of about 50 billion (50G) bases for one Solexa sequencing; 3) high accuracy (greater than 98.4%), which effectively solve the problem relating to readout of poly-repetitive sequences. In another aspect, when the number of the sequences to be sequenced is predetermined, the high sequencing throughput in turn improves the sequencing depth of the sequences (for example, each sequence can be sequenced for more times), thereby ensuring the credibility of the sequencing result. As used in the application, the term "sequencing depth" refers to the times of a fragment of DNA sequence appearing in the sequencing data. Sequencing depth can be calculated by dividing the volume of sequencing data with the length of the genome, for example, a sequencing depth of 10 indicates that the volume of sequencing data is 10 times as the length of the whole genome.

Solexa sequencing method is widely used. It may be applied to genome sequencing, genotype, studies on genetic polymorphism, and the like. In the method according to the present invention, Solexa sequencing method is used to detect HPV: sequencing the samples to be analyzed as directed to HPV, and then aligning the sequencing results with the reference sequences in HPV database by using the alignment programs known in the art, such as BLAST and SOAP, so as to accurately type the HPV infected in the sample. The HPV database used herein comprises the sequences from various HPV types known in the art, which can be found in, for example, public database, such as NCBI

database (http://www.ncbi.nlm.nih.gov/).

As used herein exchangeably, the terms "PCR index", "index", or "primer index" refers to a small fragment of nucleotide sequence being added to the 5' end of a PCR primer, which can be used to label the PCR product by PCR amplification so as to distinguish the template origin of each of the PCR products in the mixture of PCR products of different template origins. PCR products are labeled by adding an index to the 5' end of the primer and using the index primer to carry out amplification, thereby a library for further analysis and treatment can be obtained b y mixing multiple different PCR products. Each different PCR product in the library has a unique index, and therefore different PCR products can be distinguished from each other based on the unique index in each PCR product, and the PCR products are corresponded to the PCR templates from which they are amplified one by one.

For example, when multiple samples are to be sequenced, a different index may be added to the 5' end of the primer for each sample, and then PCR reaction is conducted to each sample by using the primer having an index added, thereby labeling each sample (i.e., PCR product). After PCR reaction, a library is constructed by mixing the PCR products carrying different indexes from the samples, and then the PCR products in the library are sequenced by the high-throughput Solexa sequencing method. Finally, in the resultant sequencing data, the sequencing results are corresponded to the PCR products (therefore the sample templates) one by one based on the sequence of the unique indexes.

The index may be introduced into one of the pair of primers for PCR amplification or into both primers of the pair of the primers. When introducing indexes into both primers of the pair of the primers, a pair of PCR primers are combined with a pair of indexes to form a pair of index primers, wherein the forward and reverse PCR primers have a forward and reverse indexes, respectively, and the forward index corresponds to the forward PCR

primer and the reverse index corresponds to the reverse PCR primers, and the forward and reverse indexes can be the same or different.

When designing indexes, multiple factors are considered, comprising: 1) single-base repetitive sequences containing 3 or more bases should be avoid in the index sequences; 2) the total content of base A and base C at the same position of all the indexes should be between 30% and 70% of all the bases at the position; for example, when designing 100 different index sequences, the total content of bases A and C at the second position of each of the index sequences (i.e., the so-called same position) should account for 30% to 70% of the bases at the second position of the 100 sequences; 3) GC content of the index sequence per se should be between 40% and 60%; 4) the sequences of the indexes should differ from each other by more than 4 bases; 5) sequences, which are highly similar to the primers for sequencing, should be avoided in the index sequences; 6) after the index sequence is added to the PCR primer, secondary structures such as hairpin structure and dimer should be avoided in PCR index primer.

As used in the application, the term "index primer" refers to a primer carrying an index, which comprises two portions, i.e., the index portion and the primer portion, wherein the index portion is used to label the PCR product in the amplification reaction, while the primer portion base-complementarily matches the template to amplify the template, and the index portion is linked to the 5' end of the primer portion, optionally via a linker sequence.

As used in the application, the term "adapter" or "library adapter" refers to a designed fragment of nucleotide sequence, which may be linked to the amplified PCR product in the library, and therefore all the amplified PCR products in the library can be sequenced by virtue of the adapter, for example, the sequencing is carried out by using the sequencing primer designed as directed to the adapter, without using a specific sequencing primer as directed to the PCR primer. Preferably, the adapter according to the present invention

may be linked to the PCR product by "PCR-FREE" methods.

As used in the application, the term "PCR-FREE" refers to a method wherein an adapter is directly linked to a PCR product without performing PCR reaction, for example, an adapter is linked to a PCR product by using a DNA ligase. It is known for a person skilled in the art to construct a sequencing library by using PCR-Free methods (See, for example, Nature Methods 6, 291-295 (2009)). "PCR-FREE" methods have the following advantages as PCR are not necessary during the whole process: 1) reducing purification steps, reducing the time and cost spent; 2) reducing non-specific amplification; 3) avoiding mistakes introduced by PCR during construction of a library comprising a lot of PCR products of which the sequences are highly homologous, thereby improving the accuracy of the final sequencing results.

As used in the application, the method and kit according to the present invention may employ at least 1 adapter. Different adapters may share a fragment of the same sequence (cited herein as "sequencing sequence"), and can further comprise a different characteristic sequence, and therefore different adapters can be sequenced by the same primer (which is designed as directed to the same sequencing sequence), and the unique characteristic sequence can be used to distinguish the library origin of each of the PCR products in the mixture of multiple libraries, i.e., further labeling the PCR products of different library origins.

The labeling efficiency is greatly improved by the combination of indexes with adapters having different characteristic sequences (see Fig.1). For example, 100 indexes can label 100 samples, while the combinations of 100 indexes and 200 adapters with different characteristic sequences can label 100*200=20000 samples.

Therefore, in one aspect, the present invention provides a group of indexes, comprising at least 10, preferably at least 20, at least 30, at least 40, at least

50, at least 60, at least 70, at least 80, at least 90 or 95 indexes, and said indexes have the sequences selected from the group consisting of SEQ ID NOS: 1-95. In a preferred embodiment, the group of indexes comprises at least the indexes set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS: 51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof, for example, the indexes set forth in SEQ ID NOS: 1-95.

In a preferred embodiment according to the present invention, the indexes according to the present invention are used to label the PCR primers set forth in SEQ ID NOS: 96-106, and therefore are used to carry out high-throughput HPV sequencing, detection or typing.

In one aspect, the present invention provides an index primer group, comprising 11 index primers, the sequence of said index primer comprises an index sequence and a PCR primer sequence, and the index sequence is linked to 5' end of the PCR primer sequence, optionally via a linker sequence, wherein

1) said index sequence is selected from the group consisting of SEQ ID NOS: 1-95, and each of said 11 index primers in the index primer group has the same index sequence, and

2) the PCR primer sequences of said 11 index primers are set forth in SEQ ID NOS: 96-106, respectively.

The index primer group according to the present invention can amplify at least 16 products of about 170 bp, which correspond to a highly conserved DNA sequence in the most conserved gene region (L1 region) of HPV genome. Therefore, the index primer group according to the present invention can be used to accurately type HPV.

In a preferred embodiment, the index primer group according to the present invention is useful in HPV sequencing, detection or typing, and

therefore is useful in medical use, such as diagnosis of the presence of HPV and determination of HPV types, and non-medical use, such as construction of HPV database, identification of a new HPV type and subtype, study on regional characteristics of distribution of HPV types, study on epidemiology and development of vaccines. In another preferred embodiment, the index primer group according to the present invention can be used in manufacture of a kit useful in HPV sequencing, detection or typing.

In another aspect, the present invention provides an index primer set, comprising at least 10, preferably at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90 or 95 index primer groups as described above. Preferably, in the index primer set, the index sequences used in the different index primer groups are different. More preferably, the index sequences used in the index primer set comprise at least the index sequences set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS: 51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof, such as the index sequences set forth in SEQ ID NOS: 1-95.

In a preferred embodiment, the index primer set according to the present invention is useful in high-throughput HPV sequencing, detection or typing, and therefore useful in medical use, such as diagnosis of HPV associated diseases and accurate determination of HPV types (which provides basis for clinic diagnosis and treatment scheme) on large-scale, and non-medical use, such as construction of HPV database, identification of a new HPV type and subtype, study on regional characteristics of distribution of HPV types, study on epidemiology and development of vaccines. In another preferred embodiment, the index primer set according to the present invention can be used in manufacture of a kit useful in HPV sequencing, detection or typing.

In another aspect, the present invention provides a kit comprising the

index primer group or index primer set as described above. Preferably, the kit according to the present invention further comprises at least 1, preferably at least 2, at least 10, at least 20, at least 30, at least 40, at least 50, at least 100 or at least 200 adapters. In a preferred embodiment, the adapter is suitable for Solexa sequencing method; for example, the adapter is useful in construction of a sequencing library, for example, said adapters can have the sequences selected from the group consisting of SEQ ID NOS: 121-132. In a preferred embodiment, adapters are used to construct a sequencing library by PCR-FREE methods, such as DNA ligase assay.

In a preferred embodiment, the kit according to the present invention is useful in high-throughput HPV sequencing, detection or typing, and useful in medical use and non-medical use as described above.

In another aspect, the present invention provides a method for sequencing, detecting, or typing HPV in one or more samples. The method comprises the steps of amplifying DNA in each sample using an index primer group or an index primer set or a kit as described above, and then sequencing to obtain the sequence of the sample.

In another aspect, the present invention provides a method for sequencing, detecting, or typing HPV in one or more samples, comprising the following steps:

providing n samples, wherein n is an integer of ≥ 1, said samples preferably are from mammals, more preferably from human, and preferably are exfoliative cells; optionally, the n samples to be analyzed are divided into m groups, wherein m is an integer and $n \geq m \geq 1$;

1) for each sample, amplifying DNA in the sample with one index primer group, wherein said index primer group comprises 11 index primers, the sequence of said index primer comprises an index sequence and a PCR primer sequence, and the index sequence is linked to 5' end of the PCR primer sequence, optionally via a linker sequence, wherein

2010356776 10 Sep 2014

i) said index sequence is selected from the group consisting of SEQ ID NOS: 1-95, and each of said 11 index primers has the same index sequence, and

ii) the PCR primer sequences of said 11 index primers are set forth in SEQ ID NOS: 96-106, respectively,

wherein the index primer groups used in different samples can be the same or different, and different index primer groups employ different index sequences;

2) mixing the amplification products obtained by conducting amplification with different index primer groups in step 1), so as to obtain one or more PCR product libraries;

3) adding an adapter to one or more PCR product libraries obtained in step 2) by PCR-FREE methods such as DNA ligase assay, thereby constructing one or more sequencing libraries, wherein the adapters used in different sequencing libraries may be the same or different, and different adapters share the same sequencing sequence but have different characteristic sequences,

4) optionally, mixing the sequencing libraries with different adapters as obtained in step 3), so as to obtain one or more library mixtures;

5) conducting sequencing to one or more sequencing libraries obtained in step 3) or one or more library mixtures obtained in step 4) by using the second generation sequencing technology, preferably Pair-End technology;

6) corresponding the sequencing results to the samples one by one, according to the index primer sequences of the index primer group or according to the index primer sequences of the index primer group and the characteristic sequence of the adapter;

wherein said samples are preferably exfoliative cells and are preferably from animals, such as human.

In preferred embodiments, at least 10, preferably at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90 or 95 index primer groups as described above are used in the method according to the present invention. Further preferably, the index sequences used comprise at least the index sequences set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID

NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS: 51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof, for example, the index sequences set forth in SEQ ID NOS: 1-95.

In a preferred embodiment, at least 1, preferably at least 2, at least 10, at least 20, at least 30, at least 40, at least 50, at least 100 or at least 200 adapters are used in the method according to the present invention, and for example, said adapter may have the sequence selected from the group consisting of SEQ ID NOS: 121-132.

In a preferred embodiment of the method according to the present invention, after sequencing, the sequences obtained from the sample is aligned with the sequence in HPV database, so as to accurately type HPV in the sample.

In another aspect of the present invention, the present invention provides a method for high-throughput HPV sequencing, detecting or typing in multiple samples, comprising the following steps:

1) dividing the samples to be analyzed into m groups, m is an integer of ≥ 1;

2) carrying out the following steps to the samples in each group:

2a) extracting DNA from the sample to be analyzed;

2b) designing a set of indexes on the basis of the sequences of all the primers in the primer group for amplifying HPV DNA, wherein the number of the indexes, i.e., n, is equal to the number of the samples in the group;

2c) adding each of the indexes designed in step 2b) to the 5' end of the sequences of all the forward primers or reverse primers or all the primers, respectively, thereby providing n index primer groups;

2d) conducting PCR amplification to the DNA sample obtained in step 2a) by using the index primer group provided in step 2c), thereby providing PCR products, wherein a different index primer group is used for each DNA sample; and

2e) mixing all the PCR products in step 2d), so as to obtain a PCR product library;

2010356776 10 Sep 2014

3) adding an adapter to the PCR product library obtained in step 2), wherein each of the PCR product libraries uses a different adapter so as to construct m sequencing libraries, wherein each adapter shares the same sequencing sequence and has a different characteristic sequence;

4) mixing m sequencing libraries, conducting sequencing by using the second generation sequencing technology, so as to obtain the sequencing results of all the samples;

5) corresponding the sequencing results to the samples one by one, according to the characteristic sequences of the adapters, the sequences of the indexes, and the sequences of the primers in the sequencing results; and optionally, aligning the sequencing result of each sample with the HPV database, so as to achieve HPV sequencing, detection or typing;

wherein said samples are preferably exfoliative cells and are preferably from animals, such as human;

wherein the index designed in step 2b) may have the sequence selected from the group consisting of SEQ ID NOS: 1-95, and preferably comprises at least the index set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS: 51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof.

In a preferred embodiment, DNA is extracted by a method well known to a person skilled in the art. For example, DNA extraction can be carried out by using automatic DNA extractor and DNA extracting kit, for example, commercially available KingFisher automatic extractor, and US Thermo Scientific Kingfisher Flex full automatic bead extraction and purification system.

In a preferred embodiment, the primer group in step 2b) comprises 11

primers, of which the sequences are set forth in SEQ ID NOS: 96-106, respectively. The primer group consisting of 11 primers can be used to amplify to obtain at least 16 products of about 170 bp, which correspond to a highly conserved DNA sequence in the most conserved gene region (L1 region) of HPV genome. Therefore, accurate HPV typing can be accomplished by accurately sequencing the amplification product.

In another preferred embodiment, the number of the indexes designed in step 2b) is at least 10, preferably at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100. Preferably, the index has the sequence selected from the group consisting of SEQ ID NOS: 1-95. In preferred embodiments, the indexes used in different groups of samples may be the same or different. In preferred embodiments, the index introduced into the forward primer may be the same as or different from the index introduced into the reverse primer. In specially preferred embodiments, the indexes designed in step 2b) comprises at least the indexes set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS: 51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof. In specially preferred embodiments, the indexes designed in step 2b) are set forth in SEQ ID NOS: 1-95.

In a preferred embodiment, adapters are added to PCR product libraries by PCR-FREE methods, for example, using DNA ligase. In particular, in the method according to the present invention, since the DNA sequences of different HPV types are highly homologous, the construction of the sequencing library according to the present invention has to be accomplished by PCR-FREE methods. On the contrary, if adapters are added to PCR products to construct a sequencing library by the conventional pooling PCR, the resultant library will comprise a lot of products inconsistent with the original templates, resulting in being unable to accurately sequence the

original templates. In a preferred embodiment, the number of the adapters used is at least 1, at least 2, at least 10, at least 20, at least 30, at least 40, at least 50, at least 100 or at least 200, and for example, said adapter may have the sequence selected from the group consisting of SEQ ID NOS: 121-132.

In preferred embodiments of the method according to the present invention, the adapters are commercially available adapters, such as PCR-free Index Adapter Oligo Mix purchased from Illumina Company. In another embodiments, the present invention can also use the following PCR-free adapters (the portion underlined is the characteristic sequence of the adapter).

PCR-free adapter 1 (SEQ ID NO: 121):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 2 (SEQ ID NO: 122):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 3 (SEQ ID NO: 123):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 4 (SEQ ID NO: 124):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 5 (SEQ ID NO: 125):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 6 (SEQ ID NO: 126):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 7 (SEQ ID NO: 127):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 8 (SEQ ID NO: 128):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGAATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 9 (SEQ ID NO: 129):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 10 (SEQ ID NO: 130):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 11 (SEQ ID NO: 131):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTTCTGCTTG

PCR-free adapter 12 (SEQ ID NO: 132):

5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGTCTTCTGCTTG

In a preferred embodiment, Solexa sequencer (such as Illumina Genome Analyzer II x sequencer) is used in the method according to the present invention to carry out Solexa sequencing. In another preferred embodiment, HPV database comprises the sequences of HPV types known in the art, which for example, can be found in public databases such as NCBI database (http://www.ncbi.nlm.nih.gov/).

In a preferred embodiment of the method according to the present invention, the samples may be exfoliative cells. In another preferred embodiment, the samples may be from animals, preferably mammals, more preferably human.

Beneficial effect of the present invention

The new method for detecting HPV according to the present invention and the kit for the same have the following advantages over the prior art.

1) High throughput. The detection of even 10000 samples can be accomplished by carrying out the method according to the present invention once using indexes and adapters with different characteristic sequences. Therefore, the method according to the present invention can be widely applied to disease survey and become an efficient means for early diagnosis of diseases.

2) Low cost. The present invention conducts sequencing by Solexa sequencing method, the cost for sequencing is greatly reduced (only accounting for 1% of the cost of the conventional sequencing methods), thereby greatly reducing the cost for HPV detection.

3) Accurate typing of HPV. The types of HPV can be accurately determined by carrying out amplification using multiple primers (such as 6 forward primers and 5 reverse primers according to the present invention) and aligning the sequence information of the amplification products with the HPV database, thereby providing basis for clinic diagnosis and selection of treatment scheme.

In addition, the method according to the present invention also facilitate the discovery of new HPV types, including new subtypes and variants of known types, and provides a more efficient and convenient tool for scientific studies.

The preferable embodiments according to the present invention are described in detail below by combining the drawings with the Examples. However, a person skilled in the art will understand that the following drawings and Examples are intended to illustrate the present invention only, instead of limiting the scope of the present invention. According to the

drawings and the detailed description of the preferred embodiments, the purposes and advantages of the present invention are obvious for a person skilled in the art.

Description of the drawings

Figure 1 is an illustration of the PCR products labeled by an index and an adapter having a unique characteristic sequence. In an exemplary method according to the present invention, indexes are added simultaneously to both ends of the PCR products from each sample by PCR, and multiple PCR products carrying different indexes are mixed together to construct a sequencing library. During the construction of the sequencing library, if necessary, multiple sequencing libraries can be constructed, wherein adapters carrying different characteristic sequences are used to label the sequencing libraries. After construction of libraries is finished, multiple libraries labeled by adapters carrying different characteristic sequences are mixed together and are sequenced simultaneously by Solexa sequencing method (the indexes used in different sequencing libraries may be the same or different). Finally, according to the sequence information of the characteristic sequences of the adapters and indexes in the sequencing results, the sequencing results can be corresponded to the samples one by one.

Figure 2 is an agarose gel electrophorogram of a part of PCR products. It can be seen from the electrophorogram that the band of the PCR products is of about 170 bp, wherein lane M is a molecular ladder of 50bp DNA, lanes 1-14 refer to the PCR products from the randomly selected HPV positive samples.

Examples

190 samples, of which the HC-II results are known, were genotyped for HPV by the method according to the present invention. The results showed

that the results obtained by the method according to the present invention are not only consistent with the known HC-II results, but also accomplish the accurate typing of HPV.

Example 1: Exaction of DNA samples

According to the instruction of the manufacturer, KingFisher automatic extractor (US Thermo Scientific Kingfisher Flex full automatic bead extraction and purification system) was used to extract DNA from 190 samples of exfoliative cells, of which the HC-II results are known. The nucleic acids were extracted by using the program "Bioeasy_200μl Blood DNA_KF.msz". After the program was finished, about 100μl eluting product (extracted DNA) was obtained, which was used as the template for PCR amplification in next step.

Example 2: PCR amplification

The 190 DNA samples obtained in Example 1 were designated as No. 1-190, and were divided averagely into two groups (HPV-1 group: Nos. 1-95 ; HPV-2 group: Nos. 96-190). According to the sequences (Table 2, SEQ ID NOS: 96-106) of the primers in the primer group for amplification of HPV DNA (including 6 forward primers and 5 reverse primers), a set of indexes was designed, including 95 indexes (Table 1, SEQ ID NOS: 1-95). Each of the designed indexes was added to the 5' end of the sequence of each primer of the primer set, respectively, thereby obtaining 95 index primer groups, wherein each index primer group included the corresponding 6 forward index primers and 5 reverse index primers, and different index primer groups employed different indexes (i.e., 95 index primer groups were corresponded to 95 indexes one by one).

PCR reactions were conducted for all the samples in the 96-well plates, and two plates were used (one for HPV-1 group and the other for HPV-2

group). DNA obtained in Example 1 was used as template, and a different index primer group was used to carry out PCR amplification as directed to each of the samples in HPV-1 group and HPV-2 group (each containing 95 samples), that is, 95 samples were corresponded to 95 index primer groups one by one. The number of the sample corresponding to each index primer group (each index) was recorded. In each plate, a negative control was provided without adding template. The primers used in the negative controls in the two plates were the same as the primers used in sample No.1 and No.96, respectively.

Table 1: Information on indexes and samples

Index No.	Index sequences	The corresponding position in the 96-well plate	The corresponding sample (Group 1)	The corresponding sample (Group 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	TATGTGTACT	B9	21	116	21
PI-22	TGACTCAGAC	B10	22	117	22
PI-23	TCGTAGCTCA	B11	23	118	23
PI-24	GAGACTCGTA	B12	24	119	24
PI-25	CTAGATGTCA	C1	25	120	25
PI-26	GATGACTCTC	C2	26	121	26
PI-27	TCAGTCGCAC	C3	27	122	27
PI-28	TGTAGTGAGT	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	GTCTCGACTG	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	CTCACTCGAT	D3	39	134	39
PI-40	GCTGCGACTC	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	GAGTGTACTA	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	GCATATATCT	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	ATCTCTGATG	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

Table 2: The sequence information of the primers of the primer group for amplification of HPV DNA without adding indexes.

Primer No.	Primer sequence	SEQ ID NO:
F1	TTTGTTACTGTGGTAGATACTAC	96
F2	TTTGTTACTGTGGTGGATACTAC	97
F3	TTTGTTACCGTTGTTGATACTAC	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

Note: F represents a forward primer, R represents a reverse primer.

The following PCR parameters were used in amplification:

95℃ 30s → 48℃ 30s → 72℃ 30s (40 cycles)

72℃ 10 min → 12℃ ∞

PCR reaction system was of 25μl, and its composition was as follows (all the agents were purchased from Enzymatics Company):

agents	volume/reaction
H_2O (HPLC grade)	14.375μl
10x Ex Taq Buffer (Mg^{2+} plus)	2.5μl
dNTP mix (each 2.5 mM)	2μl
mixture of F1/F2/F3/F4/F5/F6 carrying an index (each 7.5pmol)	0.5μl
mixture of R1/R2/R3/R4/R5 carrying an index (each 7.5pmol)	0.5μl
Ex Taq HS (5U/μl)	0.125μl
Template DNA	5μl
Total volume	25μl

PCR reaction was carried out in PTC-200 PCR Instrument from Bio-Rad Company. After finishing PCR, 3μl PCR product was taken to perform electrophoresis on 2.5% agarose gel (Fig.2).

Example 3: Mixing and purification of PCR products

The rest PCR products in HPV-1 group and HPV-2 group each were mixed in a 3 ml EP tube (also marked as HPV-1 group and HPV-2 group), and were mixed homogenously under shaking. 500μl DNA was taken from each of the two tubes, and was purified by using the Qiagen DNA Purification kit according to the instruction of the manufacturer, to provide 200μl purified DNA. The DNA concentrations of the purified mixtures were determined as 98 ng/μl (HPV-1 group) and 102 ng/μl (HPV-2 group) by using Nanodrop

8000 (Thermo Fisher Scientific Co.), respectively.

Example 4: Construction of Solexa sequencing library

4.1: End-repair reaction

DNA end-repair reaction was conducted to the purified DNA mixtures in the two tubes obtained in Example 3 by using Thermomixer (Eppendorf Company), respectively. The reaction system for the repair reaction was of 100μl, and its composition was as follow (all the agents were purchased from Enzymatics Company):

Agents	Volume/reaction
DNA obtained in last step	75μl
20x Polynucleotide Kinase Buffer(B904)	10μL
dNTP mix (20mM each)	4μL
T4 DNA Polymerase	5μL
Klenow Fragment	1μL
T4 Polynucleotide Kinase	5μL
Total volume	100μL

Reaction condition: 20℃, 30 mins.

According to the instruction of the manufacturer, the products of the DNA end-repair reaction were purified and recovered by using QIAquick PCR Purification kit. The recovered product was dissolved in 34 μl EB (QIAGEN Elution Buffer) .

4.2: Addition of A at the 3' end

The base A was added to the 3' end of the recovered DNA by using Thermomixer (Eppendorf Company). The reaction system was of 50ul, and its composition was as follow (all the agents were purchased from Enzymatics Company):

Agent	Volume/reaction
DNA obtained in last step	32μl
dATP (1mM, GE Company)	10μl
10x Blue Buffer	5μl
Klenow (3'-5'exo-)	3μl
Total volume	50μl

Reaction conditions: 37℃, 30 mins.

According to the instruction of the manufacturer, the product having the base A added at 3' end was purified and recovered by using MiniElute PCR Purification Kit (QIAGEN Company). The recovered product was dissolved in 20 μl EB.

4.3: Addition of Solexa adapter

Different adapters were added to the two products obtained in last step by using Thermomixer (Eppendorf Company), so as to construct 2 sequencing libraries. The corresponding correlation between the adapters and libraries was recorded.

The reaction system for the addition of the Solexa adapters was of 50ul, and its composition was as follow (all the agents were purchased from Illumina Company):

Agent	Volume/reaction
DNA obtained in last step	11μL
2x Rapid ligation Buffer	15μL
PCR-free Index Adapter oligo mix (25mM)	1μL
T4 DNA Ligase (Rapid, L603-HC-L)	3μL
Total volume	30μL

Reaction condition: 20℃, 15 mins.

According to the instruction of the manufacturer, the reaction products were purified by using Ampure Beads (Beckman Coulter Genomics), and the purified products were dissolved in 17μl deionized water. Agilent Bioanalyzer 2100 (Agilent Company) and Fluorescence quantitative PCR (QPCR) were used to detect the DNA concentration of the products, the results were as follow.

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

Example 5: Solexa sequencing

The concentrations measured by Agilent Bioanalyzer 2100 were used as the standard, and the two products obtained in last step were mixed equimolarly (10pmol DNA for each). According to the instruction of the manufacturer, Solexa sequencer (Illumina Genome Analyzer IIx sequencer) was used in sequencing under Solexa PE-75 program.

Example 6: Result analysis

According to the sequence information of the characteristic sequences of the adapters and the index primers (index portion and primer portion) in the sequencing result, the sequencing results were corresponded to the samples one by one. Then, the sequencing result of each sample was aligned with HPV database by using alignment programs known in the art, such as BLAST and SOAP, thereby accomplishing HPV detection and accurately typing HPV.

The detection results obtained were completely identical to the known results (see Table 3), indicating that the method according to the present

invention can be applied to accurately detect HPV in samples.

Table 3: Detection results of 190 samples.

Sample No.	The known HC-II result (RLU/CO value)	HPV result of the present detection	Sample No	The known HC-II result (RLU/CO value)	HPV result of the present detection
1	14.2	HPV56	96	0.28	negative
2	0.31	negative	97	0.33	negative
3	196.41	HPV16	98	181.29	HPV35, HPV6
4	5.76	HPV18	99	77.32	HPV16
5	0.35	negative	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	negative	106	2.91	HPV68
12	3420.57	HPV18	107	193.13	HPV45
13	0.38	negative	108	2.62	HPV6
14	0.41	negative	109	94.12	HPV16
15	455.06	HPV16	110	792.72	HPV18, HPV31
16	8.93	HPV18	111	31.76	HPV11
17	0.6	negative	112	0.25	negative
18	0.41	negative	113	0.23	negative
19	0.29	negative	114	750.82	HPV56, HPV16
20	27.64	HPV31	115	0.4	negative
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
24	9.55	HPV43	119	6.26	HPV11
25	237.62	HPV39	120	1719.67	HPV16, HPV45
26	1.5	HPV6	121	76.92	HPV51
27	1478.98	HPV68, HPV16	122	1318.02	HPV56, HPV16,

					HPV42
28	115.31	HPV44	123	0.28	negative
29	419.31	HPV16	124	0.33	negative
30	1.81	candHPV89	125	181.29	HPV59
31	2013.61	HPV52, HPV39	126	77.32	HPV68
32	1379.09	HPV54, HPV33	127	110.8	HPV52
33	12.74	HPV42	128	147.25	HPV16
34	1695.31	HPV16, candHPV89	129	0.24	HPV26
35	1410.85	HPV35	130	1.55	HPV11, HPV53
36	1149.25	HPV18	131	2.03	HPV6, HPV66
37	0.24	negative	132	8.45	HPV43
38	1.55	HPV11	133	0.2	negative
39	2.03	HPV11, HPV6	134	0.24	negative
40	8.45	HPV42	135	10.53	HPV11
41	0.2	negative	136	1410.85	HPV16, HPV53, HPV70
42	0.22	negative	137	1149.25	HPV56, HPV81, HPV73
43	0.53	negative	138	0.24	negative
44	10.38	HPV6	139	413.9	HPV45
45	78.21	HPV16	140	17.05	HPV11
46	0.23	negative	141	23.6	HPV52
47	45.42	HPV16, HPV18	142	3379.09	HPV16, HPV35, HPV56
48	0.35	negative	143	0.18	negative
49	148.66	HPV18, candHPV89	144	1.46	HPV18
50	60.27	HPV56	145	1.25	HPV11, HPV26
51	0.28	negative	146	2.13	HPV6, HPV81
52	360.26	HPV56, HPV68	147	872.52	HPV16, HPV45, HPV52
53	50.31	HPV18	148	1.5	HPV18
54	0.18	negative	149	4.33	HPV16
55	0.31	negative	150	0.82	negative
56	196.41	HPV16	151	60.35	HPV59
57	5.76	HPV51	152	0.24	negative
58	0.23	negative	153	0.23	negative

59	0.88	negative	154	0.18	negative
60	0.16	negative	155	1.46	HPV51
61	870.63	HPV52, HPV16	156	11.25	HPV16
62	10.18	HPV42	157	2.13	HPV11
63	0.15	negative	158	0.13	negative
64	1.36	HPV11	159	90.18	HPV58
65	68.2	HPV59	160	0.15	negative
66	0.68	negative	161	602.79	HPV68, HPV16
67	130.41	HPV45	162	132.68	HPV56, HPV11
68	0.26	negative	163	127.08	HPV39, HPV54
69	5.25	HPV6	164	602.79	HPV33
70	0.46	negative	165	276	HPV18
71	8.23	HPV40	166	243.6	HPV45
72	0.28	negative	167	229.44	HPV51
73	100.16	HPV43, HPV44	168	1384.92	HPV16, HPV58, HPV72
74	450.13	HPV41	169	172.64	HPV58
75	127.08	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	negative
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	negative
91	186.06	HPV16	186	0.13	negative
92	0.02	negative	187	60.18	HPV59

2010356776 10 Sep 2014

93	386.06	HPV18, HPV16	188	0.15	negative
94	28.09	HPV6, HPV44	189	1.36	HPV43
95	186.06	HPV68,	190	0.28	negative

In addition, accurate typing of HPV in samples was also accomplished by the method according to the present invention. Table 4 provides the sequences and typing results of the samples corresponding to lane Nos. 1-14 as shown in Fig.2.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

Table 4: The sequences and typing results of the samples corresponding to lane Nos. 1-14 as shown in Fig.2.

Lane No.	HPV type	Sequencing results	Parameter of alignment			SEQ ID NO:
			identity	score	e value	
1	HPV11	GAAAAATAAACTGTAAATCATATTCCTCCACATGGCGCATGTATTCCTTATAATCTGAATTA	96.77	107	6.00E-27	107
2	HPV16	TTTGTTACTGTGGTAGATACTACCCGCAGTACAAATATGTCATTATGTGCTGCCATATCTAC	96.77	107	6.00E-27	108
3	HPV18	TTTGTTACTAAGGTAGATACCACTCGCAGTACCAATTTAACAATATGTGCTTCTACACAGTC	96.77	107	6.00E-27	109
4	HPV31	GAAAAATAAACTGTAAATCATATTCCTCACCATGTCTTAAATACTCTTTAAAATTACTACTT	95.16	99.6	2.00E-24	110
5	HPV33	GAAAAATAAACTGTAAATCATATTCCTCAACATGTCTTATATATTCTTTAAAATTTTCATTT	96.77	107	6.00E-27	111
6	HPV35	GAAAAATAAACTGTAAATCATATTCTTCACCATGCCTTAAATATTCCTTATAATTGTCATTT	98.39	115	3.00E-29	112
7	HPV39	TTTGTTACTGTGGTAGATACTACCCGTAGTACCAACTTTACATTATCTACCTCTATAGAGTC	94.83	91.7	4.00E-22	113
8	HPV43	TTTGTTACTAAGGTAGATACCACTCGTAGTACAAACTTAACGTTATGTGCCTCTACTGACCC	100	101	4.00E-25	114
9	HPV45	GAAAAATAAACTGTAAATCATATTCCTCCACATGTCTACTATAGTGCTTAAACTTAGTAGGA	100	123	1.00E-31	115
10	HPV51	GAAATATAAATTGTAAATCAAATTCTTCCCCATGCCTAATATATTGCTTAAAGTTACTTGGA	91.94	83.8	9.00E-20	116
11	HPV52	TTTGTTACTGTGGTGGATACTACTCGTAGCACTAACATGACTTTATGTGCTGAGGTTAAAAA	98	91.7	4.00E-22	117
12	HPV56	TTTGTTACTGTTGTGGATACAACTAGAAGTACTAACATGACTATTAGTACTGCTACAGAACA	95.16	99.6	2.00E-24	118
13	HPV58	GAAAAATAAACTGTAAATCATATTCCTCAACATGACGTACATATTCCTTAAAATTATCATTT	96.77	107	6.00E-27	119
14	HPV61	TTTGTTACTGTGGTGGATACTACCCGCAGTACTAATTTAGCCATTTGCACTGCTACATCCCC	94	75.8	2.00E-17	120

References

The patents, publications and other materials, which are used to illustrate the present invention or provide further detailed contents about the practice of the present invention, are incorporated by reference herein, and are provided in the following list for the convenience.

[1].Pectasides D, Kanposioras K, Papaxoinis G et al. Chemotherapy for recurrent cervical cancer. Cancer Treatment Reviews, 2008, 34(7): 603-613.

[2].Brink, A. A., P. J. Snijders, and C. J. Meijer. HPV detection methods. Dis. Markers 2007, 23: 273-281.

[3].IARC. Handbooks of cancer prevention . Cervix cancer screening [R].Lyon: IARC Press, 2005.

[4].Doorbar, J. Molecular biology of human papillomavirus infection and cervical cancer. Clin. Sci.2006, 110: 525-541.

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

[6].Kulmala S, syIjhen. Human papillomavirus testing with the hybrid capture assay and PCR as screening tools.Clin Microbiol, 2004, 42(6): 2470-2475.

[7].Quail, M. et al., A large genome center's improvements to the Illumina sequencing system. Nat. Methods, 2008, 5, 1005-1010.

[8].Brown, C. G. et al., Solexa/Illumina GAPipeline product and product documentation, Illumina Inc, 2006.

[9].Lozano, R. Successfully integrating human papillomavirus testing into your practice . Arch . Patho1. Lab Med, 2003, 127: 991-994.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method for sequencing, detecting, or typing HPV in one or more samples, comprising the following steps:

1) for each sample, amplifying DNA in the sample with one index primer group, wherein

said index primer group comprises 11 index primers, the sequence of said index primer comprises an index sequence and a PCR primer sequence, and the index sequence is linked to 5' end of the PCR primer sequence, optionally via a linker sequence, wherein

i) said index sequence is selected from the group consisting of SEQ ID NOS: 1-95, and each of said 11 index primers has the same index sequence, and

ii) the PCR primer sequences of said 11 index primers are set forth in SEQ ID NOS: 96-106, respectively,

wherein the index primer groups used in different samples can be the same or different, and different index primer groups employ different index sequences;

2) mixing the amplification products obtained by conducting amplification with different index primer groups in step 1), so as to obtain one or more PCR product libraries;

3) adding adapters to one or more PCR product libraries obtained in step 2) by PCR-FREE methods such as DNA ligase assay, thereby constructing one or more sequencing libraries, wherein the adapters used in different sequencing libraries may be the same or different, and different adapters share the same sequencing sequence but have different characteristic sequences,

4) optionally, mixing the sequencing libraries with different adapters obtained in step 3) , so as to obtain one or more library mixtures;

5) conducting sequencing to one or more sequencing libraries obtained in step 3) or one or more library mixtures obtained in step 4) by using second generation sequencing technology;

6) corresponding the sequencing results to the samples one by one, according to the index primer sequences of the index primer group or according to the index primer sequences of the index primer group and the characteristic sequence of the adapter.

2. The method according to claim 1, wherein at least 10 index primer groups are used in step 1).

3. The method of claim 2 wherein the index sequences used comprise at least the index sequences set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS: 51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof.

4. The method according to any one of claims 1 to 3, wherein at least 1adapters are used in step 3).

5. The method of claim 4 wherein said adapter has the sequence selected from the group consisting of SEQ ID NOS: 121-132.

6. The method according to any one of claims 1 to 5, further comprising aligning the sequencing result of said sample with the sequence in HPV database, so as to accurately type HPV in the sample.

7. A method for high-throughput HPV sequencing, detecting or typing in multiple samples, comprising the following steps:

1) dividing the samples to be analyzed into m groups, m is an integer of ≥ 1;

2) carrying out the following steps to the samples in each group:

2a) extracting DNA from the samples to be analyzed;

2b) designing a set of indexes on the basis of the sequences of all the primers in the primer group for amplifying HPV DNA, wherein the number of the indexes, i.e., n, is equal to the number of the samples in the group;

2c) adding each of the indexes designed in step 2b) to the 5' end of the sequences of all the forward primers or reverse primers or all the primers, respectively, thereby providing n index primer groups;

2d) conducting PCR amplification to the DNA sample obtained in step 2a) by using the index primer group provided in step 2c), thereby providing PCR products, wherein for each DNA sample, a different index primer group is used; and

2e) mixing all the PCR products in step 2d), so as to obtain a PCR product library;

3) adding an adapter to the PCR product library obtained in step 2), wherein each of the PCR product libraries uses a different adapter so as to construct m sequencing libraries, wherein each adapter shares the same sequencing sequence and has a different characteristic sequence;

4) mixing m sequencing libraries together, conducting sequencing by using second generation sequencing technology, so as to obtain the sequencing results of all the samples;

5) corresponding the sequencing results to the samples one by one, according to the characteristic sequences of the adapters, the sequences of the indexes, and the sequences of the primers in the sequencing results; and optionally, aligning the sequencing result of each sample with the HPV database, so as to achieve HPV sequencing, detection or typing;

wherein the index designed in step 2b) has the sequence selected from the group consisting of SEQ ID NOS: 1-95.

8. The method of claim 7, wherein the index designed in step 2b) comprises at least the index set forth in SEQ ID NOS: 1-10, or SEQ ID NOS: 11-20, or SEQ ID NOS: 21-30, or SEQ ID NOS: 31-40, or SEQ ID NOS: 41-50, or SEQ ID NOS:

51-60, or SEQ ID NOS: 61-70, or SEQ ID NOS: 71-80, or SEQ ID NOS: 81-90, or SEQ ID NOS: 91-95, or any combination thereof

9. The method according to either claim 7 or claim 8, wherein the primer group in step 2b) comprises 11 primers, of which the sequences are set forth in SEQ ID NOS: 96-106, respectively.

10. The method according to any one of claims 5 to 9, wherein at least 2, adapters are used in step 3).

11. The method of claim 10, wherein said adapters have the sequence selected from the group consisting of SEQ ID NOS: 121-132.

12. The method of any one of claims 1 to 11, wherein said samples are exfoliative cells.

13. The method of claim 12 wherein the cells are from an animal.

14. The method of claim 11 wherein the cells are from a human.

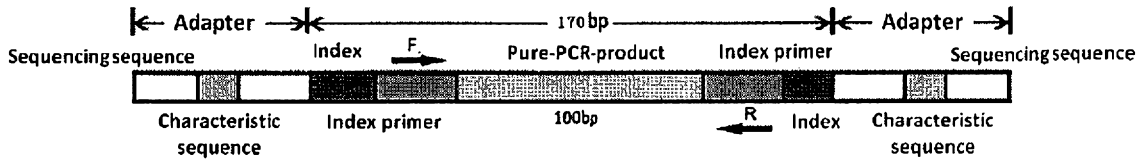

Figure 1: Illustration of the PCR products labeled by an index and an adapter having a unique characteristic sequence. F: forward PCR primer; R: reverse PCR primers.

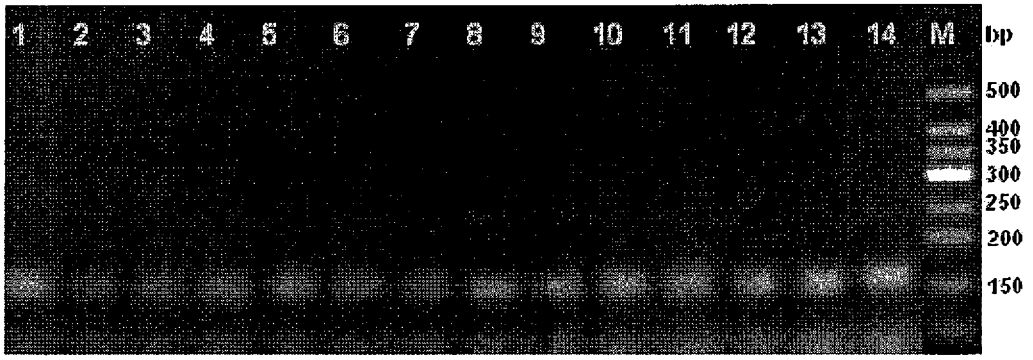

Figure 2: Gel electrophorogram of a part of PCR products.

Sequence Listing

<110> BGI Shenzhen Co., Limited

<120> A method for detecting Human Papilloma Virus based on Solexa sequencing method

<130> IDC100038

<160> 132

<170> PatentIn version 3.2

<210> 1
<211> 10
<212> DNA
<213> artificial

<400> 1
gctgcgactc 10

<210> 2
<211> 10
<212> DNA
<213> artificial

<400> 2
gtgtagatac 10

<210> 3
<211> 10
<212> DNA
<213> artificial

<400> 3
ctgatatcta 10

<210> 4
<211> 10
<212> DNA
<213> artificial

<400> 4
acgatgctat 10

<210> 5
<211> 10
<212> DNA
<213> artificial

<400> 5
tagactagac 10

<210> 6
<211> 10
<212> DNA
<213> artificial

<400> 6
ctgtctgtgt 10

<210> 7
<211> 10
<212> DNA
<213> artificial

<400> 7
gcatactgac 10

<210> 8
<211> 10
<212> DNA
<213> artificial

<400> 8
ctgctcgcat 10

<210> 9
<211> 10
<212> DNA
<213> artificial

<400> 9
catgagtaga 10

<210> 10
<211> 10
<212> DNA
<213> artificial

<400> 10
tctcactatg 10

<210> 11
<211> 10
<212> DNA
<213> artificial

<400> 11
tgtactacta 10

<210> 12
<211> 10
<212> DNA
<213> artificial

<400> 12
gtagactagt 10

<210> 13
<211> 10
<212> DNA
<213> artificial

<400> 13
atatgctact 10

<210> 14
<211> 10
<212> DNA
<213> artificial

<400> 14
cactcgctgt 10

<210> 15
<211> 10
<212> DNA
<213> artificial

<400> 15
catcacgcac 10

<210> 16
<211> 10
<212> DNA
<213> artificial

<400> 16
agcatgtgat 10

<210> 17
<211> 10
<212> DNA
<213> artificial

<400> 17
agctagtaga 10

<210> 18
<211> 10
<212> DNA
<213> artificial

<400> 18
gctatgtagt 10

<210> 19
<211> 10
<212> DNA
<213> artificial

<400> 19
tacgatgatg 10

<210> 20
<211> 10
<212> DNA
<213> artificial

<400> 20
tacgctgtac 10

<210> 21
<211> 10
<212> DNA
<213> artificial

<400> 21
tatgtgtact 10

<210> 22
<211> 10
<212> DNA
<213> artificial

<400> 22
tgactcagac 10

<210> 23
<211> 10
<212> DNA
<213> artificial

<400> 23
tcgtagctca 10

<210> 24
<211> 10
<212> DNA
<213> artificial

<400> 24
gagactcgta 10

<210> 25
<211> 10
<212> DNA

<213> artificial

<400> 25
ctagatgtca 10

<210> 26
<211> 10
<212> DNA
<213> artificial

<400> 26
gatgactctc 10

<210> 27
<211> 10
<212> DNA
<213> artificial

<400> 27
tcagtcgcac 10

<210> 28
<211> 10
<212> DNA
<213> artificial

<400> 28
tgtagtgagt 10

<210> 29
<211> 10
<212> DNA
<213> artificial

<400> 29
tcatcgtaga 10

<210> 30
<211> 10
<212> DNA
<213> artificial

<400> 30
tagcatctgt 10

<210> 31
<211> 10
<212> DNA
<213> artificial

<400> 31
tagtagtcgt 10

<210> 32
<211> 10
<212> DNA
<213> artificial

<400> 32
ctatacgtgc 10

<210> 33
<211> 10
<212> DNA
<213> artificial

<400> 33
cgactgtaga 10

<210> 34
<211> 10
<212> DNA
<213> artificial

<400> 34
gatgtcatgt 10

<210> 35
<211> 10
<212> DNA
<213> artificial

<400> 35
gtctcgactg 10

<210> 36
<211> 10
<212> DNA
<213> artificial

<400> 36
agctgacgat 10

<210> 37
<211> 10
<212> DNA
<213> artificial

<400> 37
atgatatagt 10

<210> 38
<211> 10

<212> DNA
<213> artificial

<400> 38
atgtgctcta 10

<210> 39
<211> 10
<212> DNA
<213> artificial

<400> 39
ctcactcgat 10

<210> 40
<211> 10
<212> DNA
<213> artificial

<400> 40
gctgcgactc 10

<210> 41
<211> 10
<212> DNA
<213> artificial

<400> 41
gagtcatgtc 10

<210> 42
<211> 10
<212> DNA
<213> artificial

<400> 42
catacgctca 10

<210> 43
<211> 10
<212> DNA
<213> artificial

<400> 43
cactctcgtc 10

<210> 44
<211> 10
<212> DNA
<213> artificial

<400> 44

gcactagatg 10

<210> 45
<211> 10
<212> DNA
<213> artificial

<400> 45
agtacgcatg 10

<210> 46
<211> 10
<212> DNA
<213> artificial

<400> 46
tctgtgacgt 10

<210> 47
<211> 10
<212> DNA
<213> artificial

<400> 47
tagctcatct 10

<210> 48
<211> 10
<212> DNA
<213> artificial

<400> 48
agcatacact 10

<210> 49
<211> 10
<212> DNA
<213> artificial

<400> 49
gctatagtca 10

<210> 50
<211> 10
<212> DNA
<213> artificial

<400> 50
cgtctcatgc 10

<210> 51

<211> 10
<212> DNA
<213> artificial

<400> 51
gctactacgt 10

<210> 52
<211> 10
<212> DNA
<213> artificial

<400> 52
gagtgtacta 10

<210> 53
<211> 10
<212> DNA
<213> artificial

<400> 53
gtcatacgtg 10

<210> 54
<211> 10
<212> DNA
<213> artificial

<400> 54
tatgagagat 10

<210> 55
<211> 10
<212> DNA
<213> artificial

<400> 55
atctgagtac 10

<210> 56
<211> 10
<212> DNA
<213> artificial

<400> 56
cgatagcatc 10

<210> 57
<211> 10
<212> DNA
<213> artificial

<400> 57
actgatctca 10

<210> 58
<211> 10
<212> DNA
<213> artificial

<400> 58
ctcgatacta 10

<210> 59
<211> 10
<212> DNA
<213> artificial

<400> 59
catgtgactg 10

<210> 60
<211> 10
<212> DNA
<213> artificial

<400> 60
cgcatcacta 10

<210> 61
<211> 10
<212> DNA
<213> artificial

<400> 61
gcatatatct 10

<210> 62
<211> 10
<212> DNA
<213> artificial

<400> 62
ctgatgcgac 10

<210> 63
<211> 10
<212> DNA
<213> artificial

<400> 63
tctcagagtc 10

<210> 64
<211> 10
<212> DNA
<213> artificial

<400> 64
cagtgcgagt 10

<210> 65
<211> 10
<212> DNA
<213> artificial

<400> 65
atctctgatg 10

<210> 66
<211> 10
<212> DNA
<213> artificial

<400> 66
gctagtagtc 10

<210> 67
<211> 10
<212> DNA
<213> artificial

<400> 67
atgagtcgtc 10

<210> 68
<211> 10
<212> DNA
<213> artificial

<400> 68
atcactcaga 10

<210> 69
<211> 10
<212> DNA
<213> artificial

<400> 69
tctctctgat 10

<210> 70
<211> 10
<212> DNA
<213> artificial

<400> 70
ctctagtgct 10

<210> 71
<211> 10
<212> DNA
<213> artificial

<400> 71
cgtcgtgcta 10

<210> 72
<211> 10
<212> DNA
<213> artificial

<400> 72
cgactactat 10

<210> 73
<211> 10
<212> DNA
<213> artificial

<400> 73
gcacgtcgat 10

<210> 74
<211> 10
<212> DNA
<213> artificial

<400> 74
gtagtgctct 10

<210> 75
<211> 10
<212> DNA
<213> artificial

<400> 75
ctgacgagct 10

<210> 76
<211> 10
<212> DNA
<213> artificial

<400> 76
ctatagtcta 10

<210> 77
<211> 10
<212> DNA
<213> artificial

<400> 77
acacgcacta 10

<210> 78
<211> 10
<212> DNA
<213> artificial

<400> 78
ctcgcactac 10

<210> 79
<211> 10
<212> DNA
<213> artificial

<400> 79
agatctcact 10

<210> 80
<211> 10
<212> DNA
<213> artificial

<400> 80
atactagtgt 10

<210> 81
<211> 10
<212> DNA
<213> artificial

<400> 81
atatctcgta 10

<210> 82
<211> 10
<212> DNA
<213> artificial

<400> 82
tgactgcgta 10

<210> 83
<211> 10
<212> DNA

<213> artificial

<400> 83
tgtagacgta 10

<210> 84
<211> 10
<212> DNA
<213> artificial

<400> 84
agagactatg 10

<210> 85
<211> 10
<212> DNA
<213> artificial

<400> 85
gtcgagtcac 10

<210> 86
<211> 10
<212> DNA
<213> artificial

<400> 86
tgacagctac 10

<210> 87
<211> 10
<212> DNA
<213> artificial

<400> 87
cgctagacat 10

<210> 88
<211> 10
<212> DNA
<213> artificial

<400> 88
cgtagatatg 10

<210> 89
<211> 10
<212> DNA
<213> artificial

<400> 89
tgagtctgct 10

<210> 90
<211> 10
<212> DNA
<213> artificial

<400> 90
tagtcgtatg 10

<210> 91
<211> 10
<212> DNA
<213> artificial

<400> 91
catacacgac 10

<210> 92
<211> 10
<212> DNA
<213> artificial

<400> 92
cgctcagaga 10

<210> 93
<211> 10
<212> DNA
<213> artificial

<400> 93
gtgagtctca 10

<210> 94
<211> 10
<212> DNA
<213> artificial

<400> 94
gacagatgat 10

<210> 95
<211> 10
<212> DNA
<213> artificial

<400> 95
gctgtgcgac 10

<210> 96
<211> 23

<212> DNA
<213> artificial

<400> 96
tttgttactg tggtagatac tac 23

<210> 97
<211> 23
<212> DNA
<213> artificial

<400> 97
tttgttactg tggtggatac tac 23

<210> 98
<211> 23
<212> DNA
<213> artificial

<400> 98
tttgttaccg ttgttgatac tac 23

<210> 99
<211> 25
<212> DNA
<213> artificial

<400> 99
tttgttacta aggtagatac cactc 25

<210> 100
<211> 23
<212> DNA
<213> artificial

<400> 100
tttgttactg ttgtggatac aac 23

<210> 101
<211> 25
<212> DNA
<213> artificial

<400> 101
tttgttacta tggtagatac cacac 25

<210> 102
<211> 27
<212> DNA
<213> artificial

<400> 102

gaaaaataaa ctgtaaatca tattcct 27

<210> 103
<211> 25
<212> DNA
<213> artificial

<400> 103
gaaaaataaa ttgtaaatca tactc 25

<210> 104
<211> 25
<212> DNA
<213> artificial

<400> 104
gaaatataaa ttgtaaatca aattc 25

<210> 105
<211> 25
<212> DNA
<213> artificial

<400> 105
gaaaaataaa ctgtaaatca tattc 25

<210> 106
<211> 25
<212> DNA
<213> artificial

<400> 106
gaaaaataaa ctgcaaatca tattc 25

<210> 107
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 107
gaaaaataaa ctgtaaatca tattcctcca catggcgcat gtattcctta taatctgaat 60

ta 62

<210> 108
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 108
tttgttactg tggtagatac tacccgcagt acaaatatgt cattatgtgc tgccatatct 60

ac 62

<210> 109
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 109
tttgttacta aggtagatac cactcgcagt accaatttaa caatatgtgc ttctacacag 60

tc 62

<210> 110
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 110
gaaaaataaa ctgtaaatca tattcctcac catgtcttaa atactcttta aaattactac 60

tt 62

<210> 111
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 111
gaaaaataaa ctgtaaatca tattcctcaa catgtcttat atattcttta aaattttcat 60

tt 62

<210> 112
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 112
gaaaaataaa ctgtaaatca tattcttcac catgccttaa atattcctta taattgtcat 60

tt 62

<210> 113
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 113
tttgttactg tggtagatac tacccgtagt accaacttta cattatctac ctctatagag 60

tc 62

2010356776 10 Sep 2014

<210> 114
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 114
tttgttacta aggtagatac cactcgtagt acaaacttaa cgttatgtgc ctctactgac 60

cc 62

<210> 115
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 115
gaaaaataaa ctgtaaatca tattcctcca catgtctact atagtgctta aacttagtag 60

ga 62

<210> 116
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 116
gaaatataaa ttgtaaatca aattcttccc catgcctaat atattgctta aagttacttg 60

ga 62

<210> 117
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 117
tttgttactg tggtggatac tactcgtagc actaacatga ctttatgtgc tgaggttaaa 60

aa 62

<210> 118
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 118
tttgttactg ttgtggatac aactagaagt actaacatga ctattagtac tgctacagaa 60

ca 62

<210> 119
<211> 62
<212> DNA

<213> HPV DNA sequence

<400> 119
gaaaaataaa ctgtaaatca tattcctcaa catgacgtac atattcctta aaattatcat 60

tt 62

<210> 120
<211> 62
<212> DNA
<213> HPV DNA sequence

<400> 120
tttgttactg tggtggatac tacccgcagt actaatttag ccatttgcac tgctacatcc 60

cc 62

<210> 121
<211> 63
<212> DNA
<213> artificial

<400> 121
gatcggaaga gcacacgtct gaactccagt cacatcacga tctcgtatgc cgtcttctgc 60

ttg 63

<210> 122
<211> 63
<212> DNA
<213> artificial

<400> 122
gatcggaaga gcacacgtct gaactccagt caccgatgta tctcgtatgc cgtcttctgc 60

ttg 63

<210> 123
<211> 63
<212> DNA
<213> artificial

<400> 123
gatcggaaga gcacacgtct gaactccagt cacttaggca tctcgtatgc cgtcttctgc 60

ttg 63

<210> 124
<211> 63
<212> DNA
<213> artificial

<400> 124

gatcggaaga gcacacgtct gaactccagt cactgaccaa tctcgtatgc cgtcttctgc 60

ttg 63

<210> 125
<211> 63
<212> DNA
<213> artificial

<400> 125
gatcggaaga gcacacgtct gaactccagt cacacagtga tctcgtatgc cgtcttctgc 60

ttg 63

<210> 126
<211> 63
<212> DNA
<213> artificial

<400> 126
gatcggaaga gcacacgtct gaactccagt cacgccaata tctcgtatgc cgtcttctgc 60

ttg 63

<210> 127
<211> 63
<212> DNA
<213> artificial

<400> 127
gatcggaaga gcacacgtct gaactccagt caccagatca tctcgtatgc cgtcttctgc 60

ttg 63

<210> 128
<211> 63
<212> DNA
<213> artificial

<400> 128
gatcggaaga gcacacgtct gaactccagt cacacttgaa tctcgtatgc cgtcttctgc 60

ttg 63

<210> 129
<211> 63
<212> DNA
<213> artificial

<400> 129
gatcggaaga gcacacgtct gaactccagt cacgatcaga tctcgtatgc cgtcttctgc 60

ttg 63

<210> 130
<211> 63
<212> DNA
<213> artificial

<400> 130
gatcggaaga gcacacgtct gaactccagt cactagctta tctcgtatgc cgtcttctgc 60

ttg 63

<210> 131
<211> 63
<212> DNA
<213> artificial

<400> 131
gatcggaaga gcacacgtct gaactccagt cacggctaca tctcgtatgc cgtcttctgc 60

ttg 63

<210> 132
<211> 63
<212> DNA
<213> artificial

<400> 132
gatcggaaga gcacacgtct gaactccagt caccttgtaa tctcgtatgc cgtcttctgc 60

ttg 63