

Article

# Characterization of Intra-Type Variants of Oncogenic Human Papillomaviruses by Next-Generation Deep Sequencing of the E6/E7 Region

Enrico Lavezzo <sup>1,†</sup>, Giulia Masi <sup>1,†</sup>, Stefano Toppo <sup>1</sup>, Elisa Franchin <sup>1</sup>, Valentina Gazzola <sup>1</sup>, Alessandro Sinigaglia <sup>2</sup>, Serena Masiero <sup>1</sup>, Marta Trevisan <sup>1</sup>, Silvana Pagni <sup>1</sup>, Giorgio Palù <sup>1,\*</sup> and Luisa Barzon <sup>1</sup>

<sup>1</sup> Department of Molecular Medicine, University of Padova, via A. Gabelli 63, 35121 Padova, Italy; enrico.lavezzo@unipd.it (E.L.); giulia.masi@unipd.it (G.M.); stefano.toppo@unipd.it (S.T.); elisa.franchin@unipd.it (E.F.); gazzola.valentina@virgilio.it (V.G.); seremasi@libero.it (S.M.); marta.trevisan@unipd.it (M.T.); silvana.pagni@unipd.it (S.P.); luisa.barzon@unipd.it (L.B.)

<sup>2</sup> Veneto Institute of Oncology IOV IRCCS, via Gattamelata 64, 35128 Padova, Italy; alessandrosinaglia@yahoo.it

\* Correspondence: giorgio.palu@unipd.it; Tel.: +39-049-821-2350; Fax: +39-049-827-2355

† These authors contributed equally to this work.

Academic Editors: Johnson Mak, Peter Walker and Marcus Thomas Gilbert

Received: 24 November 2015; Accepted: 7 March 2016; Published: 14 March 2016

**Abstract:** Different human papillomavirus (HPV) types are characterized by differences in tissue tropism and ability to promote cell proliferation and transformation. In addition, clinical and experimental studies have shown that some genetic variants/lineages of high-risk HPV (HR-HPV) types are characterized by increased oncogenic activity and probability to induce cancer. In this study, we designed and validated a new method based on multiplex PCR-deep sequencing of the E6/E7 region of HR-HPV types to characterize HPV intra-type variants in clinical specimens. Validation experiments demonstrated that this method allowed reliable identification of the different lineages of oncogenic HPV types. Advantages of this method over other published methods were represented by its ability to detect variants of all HR-HPV types in a single reaction, to detect variants of HR-HPV types in clinical specimens with multiple infections, and, being based on sequencing of the full E6/E7 region, to detect amino acid changes in these oncogenes potentially associated with increased transforming activity.

**Keywords:** human papillomavirus; genotyping; next-generation sequencing; E6/E7; subtype; variant; cervical cancer; high-risk HPV; deep sequencing; pyrosequencing

## 1. Introduction

Papillomaviruses are small circular double-stranded DNA viruses that preferentially infect the basal layer of skin and mucosal epithelia. They are highly species-specific and have been identified in a broad variety of animal species, including reptiles, birds, and mammals [1]. To date, nearly 200 different human papillomavirus (HPV) types have been identified [2,3], including twelve classified as oncogenic or “high-risk” HPV types (HR-HPV) [4]. Different papillomavirus types are defined by the presence of 10% or more nucleotide differences between two phylogenetically close genomes that share a common ancestor, while papillomavirus genomes that differ between 10% and 2% are defined as intra-type variants (also defined as “subtypes” or “lineages”) [5,6]. The high genetic diversity of papillomaviruses is the result of a long evolutionary history with the sequential accumulation of genetic changes over millions of years. Actually, papillomaviruses are characterized by a very low

mutation rate because they exploit host replication machinery and proofreading activity to replicate their own genome.

Differences in tissue tropism and ability to promote cell proliferation and transformation have been demonstrated among papillomavirus types. As for HPV types, 12 (*i.e.*, HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59) have been classified as oncogenic types because of the experimental and/or clinical evidence of their etiological role in the development of cervical cancer and of a subset of anogenital and head and neck squamous cell carcinomas [4]. HPV68, defined as a possible oncogenic HPV type because of the mechanistic evidence of a role in cervical tumorigenesis [4], is generally included among HR-HPV types in diagnostic assays [7].

Differences in oncogenic activity have been demonstrated among HR-HPV types (e.g., HPV16 is the most potent HPV type, able to cause cancer at different anatomic sites, at variance with the other oncogenic HPV types that have been associated with cervical cancer but rarely with other types of cancer). In addition, clinical and experimental studies have shown that the presence of nucleotide changes in some HPV intra-type variants may be associated with increased oncogenic activity and probability to induce cancer [6,8–11]. For example, non-European lineages of HPV16 have been associated with an increased risk of persistent viral infection and development of cervical precancer and cancer amongst European populations [11–14]; A1 sublineage of HPV33, B2 sublineage of HPV45, and A1 and A3 sublineages of HPV58 have been found to be overrepresented in cervical cancer [9–11].

While further studies are required to confirm the significance of HPV intra-type genetic variants in conferring an increased risk of cancer development, especially for non-HPV16/18 oncogenic types [6], the available data from the literature indicate that identification of HPV types and variants in HPV-positive subjects may be useful for risk stratification and management strategies [9]. So far, characterization of HPV genetic variants in clinical samples has been performed by polymerase chain reaction (PCR) followed by conventional Sanger sequencing of the complete viral genome or of a specific region of the genome (e.g., the L1 ORF, the E6/E7 ORFs, or the long control region) [9,10,13,15–17]. While being highly accurate, this technique does not allow testing samples with infection by multiple HPV types, which account for approximately 30% of genital HPV infections [18,19]. In addition, reported methods based on PCR-Sanger sequencing generally require prior knowledge of the infecting HPV type since they use HPV type-specific primers for PCR [9,10]. These technical hurdles can be overcome by using massively parallel high-throughput sequencing (or NGS, next-generation sequencing) technologies, which allow sequencing of multiple HPV types and variants in clinical specimens and the identification of minority genetic variants in individual samples [20–29]. Reported NGS-based methods for HPV detection and typing consisted either in deep sequencing of amplicons obtained from the amplification of highly conserved regions in HPV genome or in the unbiased metagenomics sequencing of whole DNA or RNA from biological samples. Methods based on the first approach have been developed and validated as diagnostic tests for HPV genotyping [21–25,28,29], while the metagenomics approach has been applied to the study of the molecular mechanisms of virus-related tumors and has led to the discovery on new tumor viruses, including new HPV types [26,27,30–33].

In this study, we designed and validated a new method for the characterization of intra-type variants of HR-HPVs based on deep sequencing of the whole E6/E7 region. Potential applications of this method include epidemiological studies on the distribution of HPV lineages, investigation of HPV-related cancers to identify HPV variants associated with increased oncogenic potential, and risk-stratification in patients with HR-HPV infection.

## 2. Materials and Methods

### 2.1. Primer Design

A bioinformatics pipeline was implemented with the aim of generating a set of PCR primers able to amplify the whole E6/E7 region of all HPV types classified as oncogenic groups 1 and 2A by the International Agency for Research on Cancer (IARC) [4] in order to identify HR-HPV intra-type variants (or subtypes). The E6/E7 region was chosen as target sequence because it allows a better discrimination of HPV subtypes than other regions of HPV genome [6].

The setup of the pipeline developed for this purpose allowed to two distinct tasks to be accomplished: (1) the possibility to detect and amplify all HR-HPV types in the same reaction and (2) the robustness of the design to the effect of HPV genome variability, at least for all known variants that are deposited in the Nucleotide Database.

To this aim, all nucleotide sequences belonging to HR-HPV types (IARC group 1 and 2A) were downloaded from GenBank (Table 1). For each HPV type, a multiple sequence alignment was generated with USEARCH [34] and visually inspected with Jalview [35], in order to correct alignment artifacts and discard sequences not covering the E6/E7 region (totally or partially). Then, multiple sequence alignments were scanned column by column and all possible single-nucleotide polymorphisms (SNPs) present in at least three different sequences were identified and used to build all haplotypes.

**Table 1.** High-risk HPV types included in the design of the primer set.

HPV Species	HPV Type	IARC Group	No. of Sequences in Nucleotide Database Covering the HPV E6/E7 Region
$\alpha$ 5	51	1	6
$\alpha$ 6	56	1	126
$\alpha$ 7	18	1	130
	39	1	25
	45	1	104
	59	1	13
	68	2A	92
$\alpha$ 9	16	1	1897
	31	1	276
	33	1	82
	35	1	98
	52	1	363
	58	1	795

Starting from these haplotype sequences, the PCR primers were generated with Hyden [36], a tool for the design of degenerate primers for a given set of DNA sequences. The 454-specific adapters were added to the 5' end of both forward and reverse primers, together with 10-base molecular barcodes (MID: multiplex identifiers) (Roche 454 Life Sciences, Branford, CT, USA).

### 2.2. Samples

The following samples were analyzed in order to evaluate distinct features of the multiplex PCR-NGS-based system for identification of HR-HPV types and subtypes.

#### 2.2.1. HPV-Positive Cell Lines

To assess the sensitivity of designed PCR primer set, HPV16 and HPV18 DNA-positive controls were prepared with DNA purified from CaSki and HeLa cells (which contain about 500 and 50 copies of HPV16 and HPV18 genome per cell, respectively) and added at ten-fold dilutions in the DNA carrier obtained from the HPV-negative HEK293 cells, according to the WHO Human Papillomavirus

Laboratory Manual [37]. These samples were tested in 20 replicates to set the limit of detection (LoD), defined as the lowest HPV load that allowed PCR amplification and Sanger sequencing with a quality score of  $\geq 20$  in 95% of replicate samples.

### 2.2.2. High-Risk HPV-Positive Genital Swabs

To evaluate the ability of the E6/E7 primer set to identify HR-HPV intra-type variants in clinical samples, we analyzed leftover anogenital swab specimens collected in Specimen Transport Medium (STM, Qiagen, Germantown, MD, USA) at Padova University Hospital from patients with a positive result by the INNO-LiPA HPV Genotyping Extra kit (LiPA, Fujirebio Europe, Gent, Belgium). Specimens included in this study, which were stored at  $-80\text{ }^{\circ}\text{C}$  until use, were selected from consecutive samples in order to include a representative series of all HR-HPV types. HPV DNA-negative samples were also included as negative controls.

### 2.3. Primer Validation

The primer set reported in this study was designed for broad applications, including Sanger sequencing and NGS protocols. Thus, the performance of the primer set in the identification of HR-HPV intra-type variants was validated by both Sanger sequencing in 36 samples with a single HPV type infection detected by LiPA and by a NGS-based deep sequencing protocol in eight samples with multiple HPV type infection detected by LiPA.

Total DNA was purified from specimens by using the MagNA Pure 96 Viral NA Small Volume Kit on a MagNA Pure 96<sup>TM</sup> instrument (Roche Diagnostics, Monza, Italy). For PCR amplification, two primer pools composed by forward and reverse primers respectively were prepared, with each primer at 5  $\mu\text{M}$  concentration. A multiplex PCR reaction was performed in a final reaction volume of 50  $\mu\text{L}$  with reaction buffer 10% *v/v*, 3 mM  $\text{MgCl}_2$ , dNTPs 0.8 mM each, oligonucleotide primers 0.03  $\mu\text{M}$  each and two units of AmpliTaq Gold<sup>®</sup> DNA Polymerase (ThermoFisher, Waltham, MA, USA). The amplification was carried out with an initial denaturation step of 10 min at  $95\text{ }^{\circ}\text{C}$ , followed by 40 cycles of amplification ( $95\text{ }^{\circ}\text{C}$ , 30 s;  $60\text{ }^{\circ}\text{C}$ , 1 min 30 s;  $72\text{ }^{\circ}\text{C}$ , 2 min 30 s) and with a final elongation step ( $72\text{ }^{\circ}\text{C}$ , 7 min) in a MyCycler thermal cycler (Bio-Rad, Hercules, CA, USA).

Sanger sequencing was performed by using the single forward or reverse oligonucleotide primers specific for the HR-HPV type identified by LiPA at a 0.3  $\mu\text{M}$  concentration and a BigDye Terminator v3.1 Cycle Sequencing Kit (ThermoFisher) on an ABI PRISM 3130xl Genetic Analyzer System (ThermoFisher). Electropherograms were visualized through the Chromas Lite 2.1.1 software (Technelysium Pty Ltd., Brisbane, Australia) and aligned to reference sequences from PaVE [38] using Jalview [35].

Deep sequencing was performed on a Roche 454 GS FLX+ instrument (Roche 454 Life Sciences) according to the following Roche GS FLX+ protocols: Amplicon Library Preparation Manual (2014); emPCR Amplification Manual Lib-A-SV (2013); Sequencing Method Manual (2013). Due to the insertion of two distinct barcodes in primer sequences, four pools were generated by mixing equimolar amounts of DNA, each one consisting of two samples. The pools were loaded in separate lanes of the sequencing plate.

### 2.4. Data Processing and Analysis

Sequences obtained by 454 deep sequencing were demultiplexed and primers removed from the 5' and (if present) 3' end. The quality control step was performed with USEARCH [34]: sequences that were too short (<250 nt) were removed, while the remaining reads were trimmed starting from the 3' end until all nucleotide positions with a quality score lower than 10 (corresponding to 10% probability of raw read error) were discarded. Such a cut-off is acceptable because potential errors are counterbalanced by the depth of sequencing. After preprocessing, a total of 471,354 reads were obtained (mean per sample  $\pm$  standard deviation:  $57,073 \pm 14,066$ ), with an average length of  $713.9 \pm 122.4$  and an average quality of  $34.3 \pm 6.8$ .

For each sample, HPV typing was performed through BLAST [39] alignments of the processed reads against a custom database of HPV genomes downloaded from PaVE [38], similarly to what was done in [22]. To classify a sequence to a certain genotype, at least 90% of read length was required to be aligned with at least 90% of sequence identity, according to HPV typing guidelines [5]. Finally, a consensus sequence was obtained from all reads assigned to the same type within a sample. No indels (insertion and deletions) were detected in consensus sequences.

All the steps not explicitly performed with third-party tools were carried out with custom-made Perl scripts that are available upon request.

The phylogenetic analysis of E6/E7 consensus sequences obtained from the samples, together with the reference sequences for the different genotypes and variants, was performed with the Maximum Likelihood method based on different models for the different species datasets, as reported in figure legends. The analyses were conducted in MEGA6 [40].

Finally, a score based on SNP data was implemented in order to infer viral variants within HPV types. Briefly, SNPs detected in the E6/E7 region of the HR types and associated to each variant were extracted from previous reports [9,41–48]. Then, all the nucleotide positions reported to harbor an SNP were searched in the sequenced samples and, for each SNP, one point was rewarded to all the variants compatible with it. As an example, let a cytosine (C) in position 31 of E6/E7 region be present in variants A and D of HPV type 16, while a thymine (T) is found in variants B and C. For each sample bearing a C in that position, both variants A and D will be scored 1, while no score will be assigned to variants B and C. By repeating this procedure for all SNPs along the sequence, a cumulative score is obtained for each variant and can be used to infer a classification.

High-risk HPV intra-type variant lineages and sublineages were named according to the alphanumeric nomenclature [6].

### 3. Results

#### 3.1. Multiplex Polymerase Chain Reaction (PCR) Primers for Characterization of High-Risk Human Papillomavirus (HR-HPV) Intra-Type Variants

By using the bioinformatics pipeline described in the Methods section, a set of primer sequences, including 13 forward and 16 reverse primers, were designed in order to target all HR-HPV types and all their nucleotide variants. Nucleotide sequences and identified targets of the primer set are reported in Table 2. Melting temperature of primers ranged from 59 to 61 °C, mean amplicon size was 817 nt and the maximum delta of amplicon lengths was 88 nt. All the primers were pooled together and used in multiplexing.

The sensitivity of the DNA amplification-sequencing method was assessed by testing replicates of serial dilutions of HPV16 and HPV18 DNA-positive cell lines, *i.e.*, CaSki and HeLa cells, respectively. The LoD for HPV16 and HPV18 was estimated as 900 and 700 genome equivalents in 200 µL STM, respectively. These LoD values were lower than the LoD of LiPA, which was estimated as five genome equivalents per 5 µL of purified nucleic acids for both HPV16 and HPV18, as demonstrated by participation to external quality assurance evaluations promoted by WHO Global HPV DNA typing proficiency studies [49,50].

**Table 2.** Primers targeting the full E6/E7 region of high-risk HPV types.

HPV Species	Primer Type	Primer Sequence *	HPV Type and Variant §	Amplicon Length (HPV Type)
α 5	F	AACCGAAAAGGGTTATGACCGA	51 (all known variants)	815
	R	TCTGCTGTACAACGCGAAGG		
α 6	F	AGGCAGCTTATTCTGTGTGGA	56 (all known variants)	873
	R	CAGAGTGGGCACGTTACTGT		
α 7	F	AGGGAGTRACCRAAAACGGT	18, 39 (all known variants)	814 (HPV18) 807 (HPV39)
	R	GGAATGCTCGAAGGTCGTCT	18 (all known variants)	
	R	CCCGTGAGGCTTCTACTACC	39 (all known variants)	
	F	TGCAACCAAAAACGGTGCAAT	45 (all known variants)	849
	R	TTAGTTGCACACCACGGACA		
	F	GCAATGGCAGCCTTGAGG	59 (all known variants)	806
	R	GTTTGCTGCACACAAAGGACA		
	F	GGTCACGACCGAAAACGG	68 (A2, B, D1, D2, E, F1, F2)	815
	R	AGCAGYTSYAGCTTCCGCA	68 (C1, C2, D1, D2, E, F1, F2)	
	F	GKGACCGRAARCGGTCAT	68 (C1, C2)	830
R	AACAGCTGYTSTAGTGTCGG	68 (A2, B)		
α 9	F	AGGGYGTAACCGAAAACGGT	52, 58 (all known variants)	801 (HPV52) 828 (HPV58)
	R	CCGGGGCACACAACCTTGTA	52 (all known variants)	
	R	ACAGCTAGGGCACACAATGG	58 (A1, A2, A3, B1, B2, D1)	785
	R	GCTGTAGGGTTCGTCTTCA	58 (C, D2)	
	F	AGGGCGTAACCGAAATCGGT	16 (A1, A2, A3, A4, D1, D2, D3)	830
	R	TGAGAACAGATGGGGCACAC	16 (A1, A2, A3, B1, B2, C, D1, D2, D3)	
	F	TTGMACCGAAACCGGTTAGT	16 (B1, B2, C)	807
	R	RCAGATGGGGCACACAATTC	16 (A4)	
	F	GGTGAACCGAAAACGGTTGG	31 (all known variants)	793
	R	GGGGCACACGATTCCAAATG		
F	AAGTAGGGTGTAACCGAAAGCG	33 (all known variants)	787	
R	TGCTGTATGGTTCGTAGGTCAC			
F	ACGGTTGCCATAAAAAGCAGAA	35 (all known variants)	827	
R	TCTCTGTGAACAGCCGGGG			

F: forward; R: reverse; \* The 454-specific adapters were added to the 5' end of both F and R primers, together with a 10-base multiplex identifier; § For each primer, the target genotype is reported; the variants that are covered by the corresponding primer are reported within round brackets.

### 3.2. HPV Typing and Intra-Type Variant Characterization in Single and Multiple Infections

The sequences obtained by Sanger sequencing and the reads obtained by deep sequencing and assigned to each sample were genotyped by means of a BLAST search against a custom database of HPV genome sequences. The multiplex approach of the method, which is designed to target the E6/E7 region of 13 HR-HPV genotypes in the same reaction, allowed detection of the presence of single or multiple infections in all samples.

#### 3.2.1. Sanger Sequencing

A complete agreement was observed between the HPV typing results obtained by LiPA and PCR-Sanger sequencing, as shown in Table 3. However, approximately 10% of cervical swabs positive for high-risk HPVs could not be efficiently amplified by the new PCR primer set, confirming the lower sensitivity of the PCR-sequencing method than LiPA. The lower sensitivity was probably mainly related to the large size of the PCR amplicons generated in our protocol (approximately 800 nt) compared to 65 bp of LiPA amplicons. We did not observe biases in sensitivity associated with specific HPV types, suggesting that the multiplex PCR primer set could amplify the different HR-HPV types with similar efficiency.

Since Sanger sequencing was performed by using only forward and reverse primers specific to the HR-HPV types identified by LiPA as definitely present, HR-HPV types defined as possibly present by LiPA (reported within brackets in Table 3) were not expected by Sanger sequencing. In two cases (samples 890 and 6517), both with HPV58 and possibly HPV52 types as LiPA results, the forward



sequencing primer, which was shared between the two HPV types, confirmed only the presence of HPV58. The same happened for sample 007, typed as HPV18 and possibly HPV39 by LiPA and confirmed as HPV18-positive by Sanger sequencing, even though the forward primer could anneal with both HPV18 and HPV39 sequences.

**Table 3.** Comparison of HPV typing results obtained by Sanger sequencing and LiPA.

Sample ID	LiPA	Sanger Sequencing	
	HPV Type	HPV Type	Scores (Fw, Rw) *
004	HPV16	HPV16	49, 38
088	HPV16	HPV16	48, 42
198	HPV16	HPV16	45, 42
281	HPV16, <i>HPV53</i>	HPV16	46, 45
387	HPV16	HPV16	47, 47
644	HPV16	HPV16	47, 46
007	HPV18 ( <i>HPV39</i> )	HPV18	36, 15
201	HPV31	HPV31	30, 15
235	HPV31 ( <i>HPV52, HPV54</i> )	HPV31	30, 36
995	HPV31 ( <i>HPV52, HPV54</i> )	HPV31	50, 47
351	HPV33, HPV11 ( <i>HPV52, HPV54</i> )	HPV33	31, 46
901	HPV33 ( <i>HPV52, HPV54</i> )	HPV33	29, 45
805	<i>HPV35, HPV44</i>	HPV35	46, 45
843	<i>HPV35, HPV54, HPV69/71, HPVX</i>	HPV35	46, 48
787	<i>HPV39, HPV66</i>	HPV39	46, 48
860	<i>HPV39</i>	HPV39	49, 49
225	<i>HPV45, HPV74</i>	HPV45	45, 31
292	<i>HPV45</i>	HPV45	47, 43
913	<i>HPV45</i>	HPV45	46, 38
096	<i>HPV51</i>	HPV51	37, 43
772	<i>HPV51</i>	HPV51	39, 40
933	<i>HPV51</i>	HPV51	42, 42
014	<i>HPV52</i>	HPV52	37, 41
199	<i>HPV52, HPVX</i>	HPV52	33, 48
082	<i>HPV56</i>	HPV56	15, 30
095	<i>HPV56</i>	HPV56	34, 23
099	<i>HPV58</i>	HPV58	23, 34
108	<i>HPV58</i>	HPV58	16, 26
203	<i>HPV58, HPV54</i>	HPV58	43, 35
206	<i>HPV58</i>	HPV58	47, 35
890	<i>HPV58 (HPV52)</i>	HPV58	28, 43
6517	<i>HPV58, HPV44 (HPV52)</i>	HPV58	26, 14
002	<i>HPV59</i>	HPV59	28, 26
041	<i>HPV59, HPV54</i>	HPV59	49, 45
043	<i>HPV59, HPV43</i>	HPV59	39, 39
541	<i>HPV68 (HPV39)</i>	HPV68	49, 40

Note: The presence of HPV types reported within brackets is defined as possible according to LiPA. HPVX: Presence of HPV type/s not identifiable by LiPA. HPV types marked in italics are non-high-risk HPV types (not included in the design of the NGS primer set). \* Electropherogram quality values calculated by Sequencing Analysis Software 5.3.1 (ThermoFisher): scores were considered high ( $\geq 20$ ), medium ( $\geq 15$  and  $< 20$ ) or low ( $< 15$ ).

### 3.2.2. 454 deep sequencing

A good agreement was also observed between LiPA and the NGS deep-sequencing method (Table 4), with some discrepancies which could be related to the different sensitivity of the two assays for different HPV types. Among the non-HR-HPV types, a high number of reads belonging to HPV34 was found in two samples: this was a non-intended result, due to a good nucleotide pairing with some primers designed for the alpha9 species. This type was not considered in the subsequent analyses.

**Table 4.** Comparison of HPV typing results obtained by deep sequencing and LiPA.

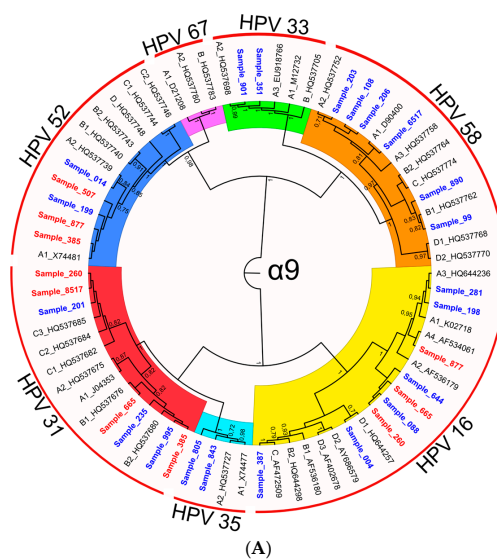
Sample ID	LiPA		454 Deep Sequencing		
	HPV Type	HPV Type	Total No. Reads (% of the Total)	No. Forward Reads	No. Reverse Reads
8517	HPV18, HPV31, (HPV39, HPV52, HPV54)	HPV31 HPV18	50,660 (93.6) 3435 (6.3)	25,876 1738	24,784 1697
507	HPV18, HPV52, HPV66, HPV69/71, HPVX, (HPV39)	HPV18 <i>HPV66</i>	39,860 (72.5) 14,913 (27.1) 151 (0.3)	19,435 7629 81	20,425 7284 70
665	HPV16, HPV31, HPV6, HPV69/71, (HPV52, HPV54)	HPV31 HPV16	69,627 (99.4) 408 (0.6)	36,230 214	33,397 194
731	<b>HPV16</b> , HPV18, (HPV39)	HPV18	42,878 (99.9)	23115	19,763
631	HPV18, HPV56, (HPV39, HPV74)	HPV18 HPV56	62,182 (97.4) 1564 (2.4)	33,380 845	28,802 719
385	HPV18, HPV31, (HPV39, HPV52, HPV54)	HPV31 HPV18 HPV52	45,947 (83.7) 8915 (16.2) 22 (0.04)	26,291 5328 11	19,656 3587 11
877	HPV18, HPV52, HPV66, HPV11, HPVX, (HPV39)	HPV18 <i>HPV34</i> HPV52 <b>HPV16</b>	32,879 (68.6) 13,895 (29) 1012 (2.1) 137 (0.3)	14,681 6048 421 82	18,198 7847 591 55
260	HPV16, HPV39, <b>HPV52</b> , HPV6, HPV69/71, HPVX	HPV39 <i>HPV34</i> HPV16 <b>HPV31</b>	59,838 (72.9) 15,042 (18.3) 6145 (7.5) 942 (1.1)	29,785 7274 2921 423	30,053 7768 3224 519

Notes: The presence of HPV types reported within brackets is defined as possible according to LiPA. HPV types marked in italics are non-high-risk HPV types (not included in the design of the NGS primer set). HPV types marked in bold represent discordant results between LiPA and 454 deep sequencing. HPVX: Presence of HPV type/s not identifiable by LiPA.

### 3.2.3. Characterization of HR-HPV intra-type variants

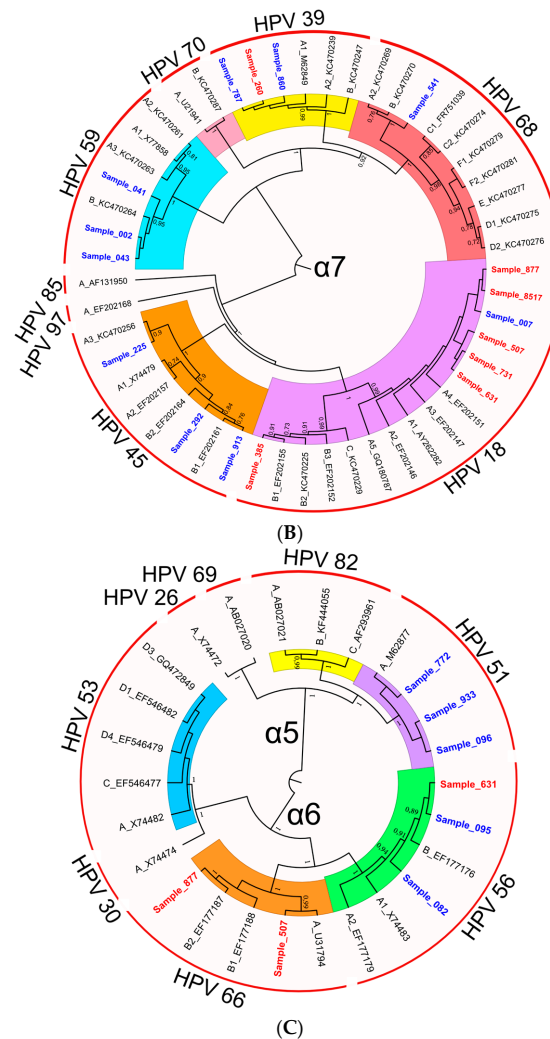
Two distinct methods were implemented to identify intra-type variants: (1) a phylogenetic analysis and (2) a scoring system based on SNPs.

As regards phylogeny, distinct phylogenetic trees were generated for HPV species  $\alpha$ -7,  $\alpha$ -9, and  $\alpha$ -5/ $\alpha$ -6 by using the Maximum Likelihood method. The phylogenetic trees included reference sequences for all known variant lineages/sublineages and the HPV sequences obtained from clinical samples (Figure 1). In the figures, HPV types within each species are highlighted with different colors, while numbers close to the nodes represent the percentage of times in which the sequences subsumed by the node were found in the same cluster in bootstrap iterations.



**Figure 1.** Cont.





**Figure 1.** (A–C) Phylogenetic trees of HPV E6/E7 sequences obtained from clinical samples (red and blue labels for sequences obtained with 454 pyrosequencing and Sanger, respectively) and references (black labels) belonging to the  $\alpha$ -9 (A);  $\alpha$ -7 (B);  $\alpha$ -5 and alpha-6 (C) species inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model [51] (A) and the Hasegawa-Kishino-Yano model [52] (B,C). The percentage of bootstrap replications in which the associated samples clustered together is shown next to the branches (over 1000 iterations). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories: +G, parameter = 0.6006 (A), 1.1395 (B), 1.6034 (C)). Tree branch lengths are proportional to the number of substitutions per site. There were a total of 741, 796, and 695 positions, respectively, in the final dataset of phylogenetic trees (A–C). Evolutionary analyses were conducted in MEGA6 [40] and displayed with FigTree [53]. Sample labels include sample ID; reference sequence labels include the GenBank accession number and the ID of variant lineages/sublineages (variant lineages are designed as A, B, etc.; variant sublineages are designed as A1, A2, etc.).

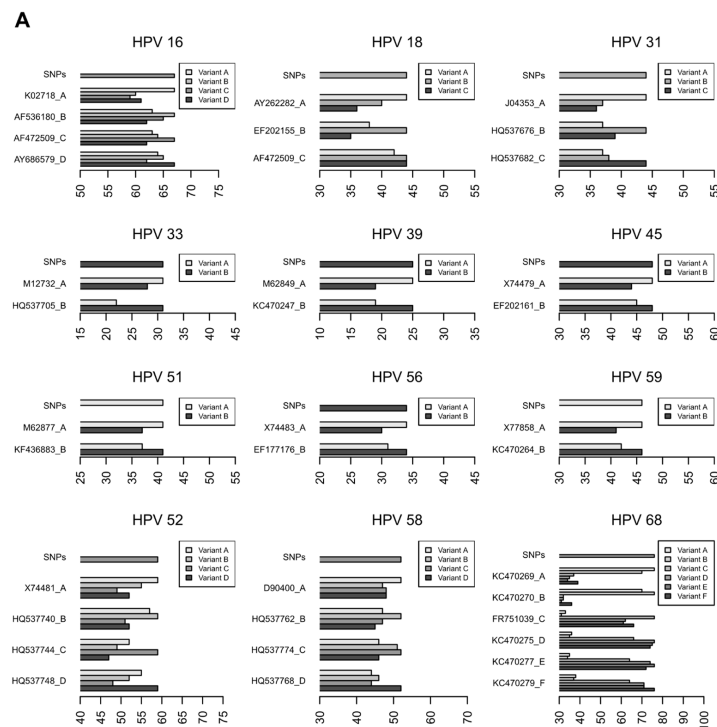
In almost all cases, sample sequences were able to cluster with the reference sequence of a single variant and with a reliable bootstrap value greater than 0.7, as reported in Figure 1 and in Table 5.

**Table 5.** HPV type and variant classification through phylogenetic analysis and SNP score.

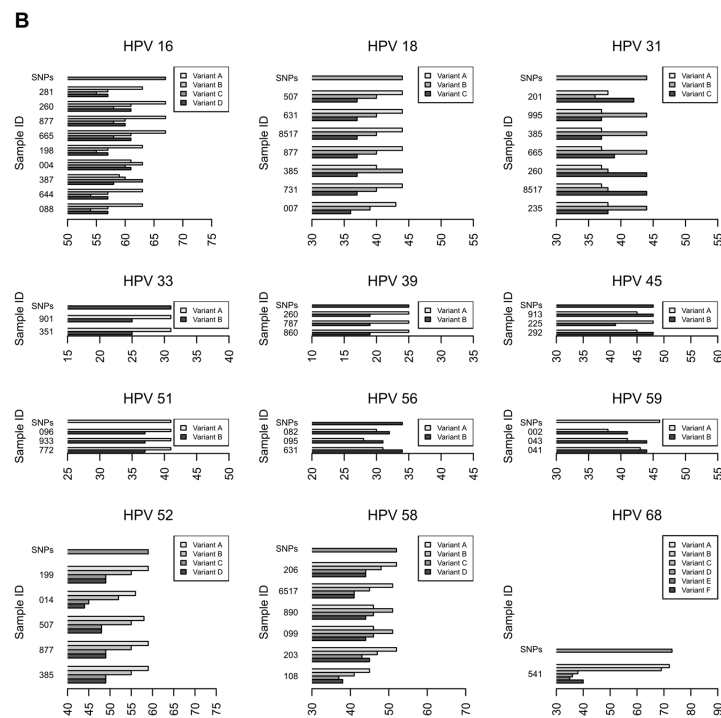
HPV Type	Bootstrap Value	HPV Intra-Type Variant		Sample IDs
		Phylogenetic Analysis	SNP Score	
16	1.00	A	A	281, 260, 877, 665, 198, 644, 088
16	1.00	C	C	387
16	0.70	D	A, B, C, D *	004
18	0.99	A	A	507, 631, 665, 8517, 877, 731, 007
18	0.91	B	B	385
31	1.00	B	B	995, 385, 665, 235
31	1.00	C	C	201, 631, 260, 8517
33	1.00	A	A	901, 351
35	1.00	A	_*	805, 843
39	0.99	A	A	260, 787, 860
45	0.90	A	A	225
45	0.84	B	B	913, 292
51	1.00	A	A	096, 933, 772
52	0.75	A	A	199, 014, 507, 877, 385
56	0.91	B	B	082, 095, 631
58	0.78	A	A	206, 6517, 203, 108
58	0.96	B	B	890, 099
59	0.95	B	B	002, 043, 041
68	1.00	A, B *	A, B *	541

\* Unable to discriminate among indicated lineages. \*\* The SNP score for HPV35 was not computed since only one variant is known.

In order to strengthen and validate the results of HPV subtype classification, we developed an alternative method based on a scoring system of SNPs associated with the different HPV variants. In this method, each nucleotide position in the E6/E7 region presenting variability within an HPV type was evaluated to build a ranking of variants, as detailed in the Methods section. The results are presented in Figure 2A (for variant reference genomes) and Figure 2B (for clinical samples), where the scores obtained by the different variants are reported in the x axis: each sample was assigned to the HPV variant that obtained the highest score.



**Figure 2.** Cont.



**Figure 2.** HPV variant assignment based on the SNP score. SNP score is reported in the x axis for the reference variant genomes (A) and for the samples sequenced in this study (B). HPV sequences detected in samples were assigned to the variant that obtained the highest score. The maximum score achievable for each genotype is reported in each panel and is labeled as “SNPs.”

As shown in Figure 2, some ambiguities occur: e.g., variant C of HPV18 obtained the same score for both lineages B and C (Figure 2A). This happened because the known intra-lineages variability is heterogeneous and there are no specific SNPs of lineage C, while there are SNP positions of lineage B where multiple nucleotides are allowed. From these premises, B *vs.* C and C *vs.* B comparisons are not symmetrical, because B is a superset of C in this particular region of the genome and, inevitably, C reaches the same score when compared to itself and B. Similarly, in some clinical samples, HPV sequences did not obtain the maximum score for any variants, e.g., HPV16 in sample 004. In these cases, possible sources of bias might be considered, such as the incompleteness of available information about lineage variability or the misclassification of the sequences used to build the scoring matrix.

The comparison of the results obtained by the two methods for the classification of HPV variants (*i.e.*, phylogenetic analysis and SNP scoring system) showed an agreement in 52 out of 54 samples (96.3%, Table 5) positive for HPV types for which more than one intra-type variant is known (*i.e.*, only one variant is known for HPV35). The two exceptions included sample 004, which was assigned to HPV16 variant D by phylogeny while the SNP score was unable to discriminate among lineages, and sample 541 which was equally assigned to HPV68 variants A and B by both methods. In the first case, both classifications are quite uncertain, since the bootstrap value for the clade containing sample 004 and HPV16 variant D1 is 0.7 while the SNP score is similar among all HPV16 variants, with no clear separation between the highest score and the others; in the second case, variants A and B of HPV68 are clustered together by phylogeny and obtain a comparable SNP score, suggesting that the E6/E7 region does not provide enough information to discriminate between them.

#### 4. Discussion

In this study, we designed and validated a new PCR-NGS-based method for the characterization of HR-HPV intra-type variants in clinical specimens. The set of PCR primers that we developed allows amplification of all HR-HPV types, which are subsequently characterized into viral variants

through a double classification based on phylogeny and on a score which considers the SNPs within the E6/E7 region. While many studies have been published on the characterization of the genetic variability of intra-type HPV variants [6,9–11,13–17], this is the first attempt to exploit such information to implement a method that combines an NGS-based approach with a bioinformatics pipeline in a new test for HPV variant typing. Proof-of-principle validation experiments were performed on a Roche 454 GS FLX + platform, but this method could be easily adapted to other NGS platforms capable to generate long reads, such as PacBio (Pacific Biosciences, Menlo Park, CA, USA).

Advantages of this method for HPV variant detection as compared to methods reported in the literature are represented by its ability (i) to detect variants of all HR-HPV types in a single reaction; (ii) to detect variants of HR-HPV types in clinical specimens with multiple infections; (iii) being based on sequencing of the full E6/E7 region, to detect amino acid changes in these oncogenes potentially associated with increased transforming activity.

The novel HPV subtyping algorithm based on the SNP score implemented in this study could be used as an alternative to phylogeny for variant classification, since it showed comparable performance. This algorithm can be provided as an online platform, as it only requires *a priori* knowledge of variable sites within the E6/E7 region and their association with HPV intra-type variants. In our evaluation, in some cases, the SNP score method was not able to unambiguously discriminate among HR-HPV intra-type variants. Nonetheless, with the growing availability of sequence data in public databases, an improvement in the discriminating power of the method is expected, helping to overcome the limitations observed for few specific variants. In alternative, an approach based on the detection of haplotypes, instead of independent variable positions, could be adopted.

In the literature, methods for HPV variant detection have been generally designed for testing single HPV types [17,54–58]. By using a multiplexed PCR reaction associated with deep sequencing of amplicons, the NGS-based method reported here allowed to detect and genotype all HR-HPV types directly from clinical specimens without prior knowledge on the HPV types present in the sample. The sensitivity of the method, tested in control HPV16 and HPV18-positive cell lines, was good (up to 900 and 700 copies/reaction respectively), allowing HPV subtyping in clinically relevant infections.

The region targeted in HPV genome for variant detection included the full sequence of the E6 and E7 oncogenes. The most accurate method for the identification of HPV variants/lineages and sublineages is obviously based on full genome sequences [6]. Among the different regions in the HPV genome used for HPV subtyping, such as the NCR, URR, E6/E7, E2, and L1, the URR and E6/E7 regions were shown to provide more accurate subtyping results than other genome regions [44,59], although they generally provide reliable discrimination at lineage but not at sublineage resolution [6,9]. Since the E6 and E7 oncogenes are the key viral factors necessary for tumor formation, nucleotide variants leading to changes in amino acid sequence may be associated with differential risk of transformation and disease progression. Hence, HPV typing and variant detection methods based on the full E6/E7 region offer the advantage of providing genetic information linked to disease pathogenesis. This information might be exploited for the identification of viral genetic traits associated with increased oncogenicity and for the development of antiviral drugs and therapeutic vaccines targeting E6 and E7 oncogenes.

Sanger sequencing, which still represents the gold standard sequencing method, was used in the present study to benchmark the performance of the E6/E7 multiplex primer set in amplifying different HR-HPV types from clinical specimens and identifying their lineages/sublineages. Classification of HR-HPV lineages and sublineages was performed following the nomenclature and the reference HPV genomes proposed by Burk *et al.* [6]. Analysis of a set of clinical samples carrying HR-HPV types as single infection showed that sequencing the target E6/E7 region allowed identification of HPV lineages in most cases. The overall analytical sensitivity of the PCR-sequencing method was lower than LiPA, due to the larger size of the amplicons required for HPV subtyping. However, this method is expected to provide valid results for clinically relevant HR-HPV infections, which are

characterized by a relatively high viral load [60]. At variance, analysis of samples characterized by high fragmentation of DNA, e.g., archival formalin-fixed, paraffin-embedded samples, would not be suitable for this method.

After validation of the multiplex PCR protocol by Sanger sequencing, the method was implemented in a 454 deep sequencing protocol. A proof-of-principle sequencing run, performed with eight clinical specimens containing multiple HR-HPV types, showed that the NGS method could correctly identify HPV types and characterize intra-type variants in multiple infections. The inclusion of MID sequences in PCR primers allowed pooling of samples in the sequencing run, thus leading to a reduction of sequencing costs per sample. In our proof-of-principle run, we used a high depth of coverage (*i.e.*, approximately 50,000 reads per sample) that allowed identification of HPV types that were not identified by LiPA probably because of their relatively low concentration in comparison with co-infecting HPV types. A good sensitivity of deep sequencing in the detection of HPV types representing a small proportion (e.g., less than 5% of HPV genome equivalents) in multiple infections was also observed in our previous studies with different primer sets targeting the L1 region [22,23]. Application of a lower depth of coverage would be needed if more samples were pooled, such as in the diagnostic setting, where low costs and high throughput are required. Although obtained from a limited number of samples, our data suggest that a lower depth (e.g., 1000–5000 reads per sample) could be used without loss in sensitivity for clinically relevant HR-HPV infections.

Theoretically, the NGS-based method allows discrimination of different variants of the same HPV type even when they are present in the same sample. This is possible thanks to deep sequencing, since two (or more) different consensus sequences can be obtained from the reads belonging to a certain genotype within a sample. We did not observe such event in the present series of samples, while we detected some cases with multiple infection of variants of the same HPV type in our previous study in a larger series of patients [22].

## 5. Conclusions

In conclusion, this study reports the design and validation of a NGS-based method for characterization of HR-HPV intra-type variants in clinical samples. Notably, the primer set can be adapted to Sanger sequencing and different NGS platforms. This method could be used both for epidemiological studies on the distribution of HPV lineages and for diagnostic purposes. Diagnostic applications may include second line testing in patients with HR-HPV infection, to identify infections associated with increased risk of persistence and disease progression [9,13]. This test could be used together with other biomarkers, e.g., presence of HPV16/18/45, HR-HPV E6/E7 mRNA expression, p16INK4a expression and HPV genome methylation, which have been also associated with an increased risk of progression to malignancy or have been used to define the causative role of HPV in cancer [61,62]. Characterization of HR-HPV variants, especially for HPV types other than HPV16 and HPV18, will be a relevant biomarker in programs for the secondary prevention of HPV-related cancer in the era of prophylactic HPV vaccination.

**Acknowledgments:** This study was funded by the Italian Ministry for Education, University, and Research (MIUR) (grant No. 2010Z9FLM8) and by the University of Padova (grant numbers CPDA138081/13 and GRIC13AAI9).

**Author Contributions:** Enrico Lavezzo, Giulia Masi and Luisa Barzon conceived and designed the experiments; Giulia Masi, Elisa Franchin, Alessandro Sinigaglia, Valentina Gazzola, Marta Trevisan, and Silvana Pagni performed the experiments; Enrico Lavezzo and Stefano Toppo analyzed the data; Luisa Barzon, Enrico Lavezzo, and Giulia Masi drafted the manuscript; Giorgio Palù coordinated the study and critically revised the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bernard, H.U. Taxonomy and phylogeny of papillomaviruses: An overview and recent developments. *Infect. Genet. Evol.* **2013**, *18*, 357–361. [[CrossRef](#)] [[PubMed](#)]

2. Bzhalava, D.; Eklund, C.; Dillner, J. International standardization and classification of human papillomavirus types. *Virology* **2015**, *476*, 341–344. [[CrossRef](#)] [[PubMed](#)]
3. Kocjan, B.J.; Bzhalava, D.; Forslund, O.; Dillner, J.; Poljak, M. Molecular methods for identification and characterization of novel papillomaviruses. *Clin. Microbiol. Infect.* **2015**, *21*, 808–816. [[CrossRef](#)] [[PubMed](#)]
4. Bouvard, V.; Baan, R.; Straif, K.; Grosse, Y.; Secretan, B.; El Ghissassi, F.; Benbrahim-Tallaa, L.; Guha, N.; Freeman, C.; Galichet, L.; *et al.* A review of human carcinogens-part B: Biological agents. *Lancet Oncol.* **2009**, *10*, 321–322. [[CrossRef](#)]
5. Bernard, H.U.; Burk, R.D.; Chen, Z.; van Doorslaer, K.; zur Hausen, H.; de Villiers, E.M. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **2010**, *401*, 70–79. [[CrossRef](#)] [[PubMed](#)]
6. Burk, R.D.; Harari, A.; Chen, Z. Human papillomavirus genome variants. *Virology* **2013**, *445*, 232–243. [[CrossRef](#)] [[PubMed](#)]
7. Poljak, M.; Cuzick, J.; Kocjan, B.J.; Iftner, T.; Dillner, J.; Arbyn, M. Nucleic acid tests for the detection of alpha human papillomaviruses. *Vaccine* **2012**, *30*, F100–F106. [[CrossRef](#)] [[PubMed](#)]
8. Chakrabarti, O.; Veeraraghavalu, K.; Tergaonkar, V.; Liu, Y.; Androphy, E.J.; Stanley, M.A.; Krishna, S. Human papillomavirus type 16 E6 amino acid 83 variants enhance E6-mediated MAPK signaling and differentially regulate tumorigenesis by notch signaling and oncogenic Ras. *J. Virol.* **2004**, *78*, 5934–5945. [[CrossRef](#)] [[PubMed](#)]
9. Xi, L.F.; Schiffman, M.; Koutsky, L.A.; Hughes, J.P.; Winer, R.L.; Mao, C.; Hulbert, A.; Lee, S.K.; Shen, Z.P.; Kiviat, N.B. Lineages of oncogenic human papillomavirus types other than type 16 and 18 and risk for cervical intraepithelial neoplasia. *JNCI J. Natl. Cancer Inst.* **2014**, *106*. [[CrossRef](#)] [[PubMed](#)]
10. Chen, A.A.; Heideman, D.A.; Boon, D.; Gheit, T.; Snijders, P.J.; Tommasino, M.; Franceschi, S.; Clifford, G.M. Human papillomavirus 45 genetic variation and cervical cancer risk worldwide. *J. Virol.* **2014**, *88*, 4514–4521. [[CrossRef](#)] [[PubMed](#)]
11. Schiffman, M.; Rodriguez, A.C.; Chen, Z.; Wacholder, S.; Herrero, R.; Hildesheim, A.; Desalle, R.; Befano, B.; Yu, K.; Safaean, M.; *et al.* A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia. *Cancer Res.* **2010**, *70*, 3159–3169. [[CrossRef](#)] [[PubMed](#)]
12. Villa, L.L.; Sichero, L.; Rahal, P.; Caballero, O.; Ferenczy, A.; Rohan, T.; Franco, E.L. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. *J. Gen. Virol.* **2000**, *81*, 2959–2968. [[CrossRef](#)] [[PubMed](#)]
13. Berumen, J.; Ordonez, R.M.; Lazcano, E.; Salmeron, J.; Galvan, S.C.; Estrada, R.A.; Yunes, E.; Garcia-Carranca, A.; Gonzalez-Lira, G.; Madrigal-de la Campa, A. Asian-american variants of human papillomavirus 16 and risk for cervical cancer: A case-control study. *JNCI J. Natl. Cancer Inst.* **2001**, *93*, 1325–1330. [[CrossRef](#)] [[PubMed](#)]
14. Cornet, I.; Gheit, T.; Iannacone, M.R.; Vignat, J.; Sylla, B.S.; Del Mistro, A.; Franceschi, S.; Tommasino, M.; Clifford, G.M. HPV16 genetic variation and the development of cervical cancer worldwide. *Br. J. Cancer* **2013**, *108*, 240–244. [[CrossRef](#)] [[PubMed](#)]
15. Chan, P.K.; Zhang, C.; Park, J.S.; Smith-McCune, K.K.; Palefsky, J.M.; Giovannelli, L.; Coutlee, F.; Hibbitts, S.; Konno, R.; Settheetham-Ishida, W.; *et al.* Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequence variations. *Int. J. Cancer* **2013**, *132*, 2528–2536. [[CrossRef](#)] [[PubMed](#)]
16. De Boer, M.A.; Peters, L.A.W.; Aziz, M.F.; Siregar, B.; Cornain, S.; Vrede, M.A.; Jordanova, E.S.; Fleuren, G.J. Human papillomavirus type 18 variants: Histopathology and E6/E7 polymorphisms in three countries. *Int. J. Cancer* **2005**, *114*, 422–425. [[CrossRef](#)] [[PubMed](#)]
17. Chan, P.K.; Luk, A.C.; Park, J.S.; Smith-McCune, K.K.; Palefsky, J.M.; Konno, R.; Giovannelli, L.; Coutlee, F.; Hibbitts, S.; Chu, T.Y.; *et al.* Identification of human papillomavirus type 58 lineages and the distribution worldwide. *J. Infect. Dis.* **2011**, *203*, 1565–1573. [[CrossRef](#)] [[PubMed](#)]
18. Barzon, L.; Militello, V.; Pagni, S.; Franchin, E.; Dal Bello, F.; Mengoli, C.; Palu, G. Distribution of human papillomavirus types in the anogenital tract of females and males. *J. Med. Virol.* **2010**, *82*, 1424–1430. [[CrossRef](#)] [[PubMed](#)]



19. Barzon, L.; Militello, V.; Pagni, S.; Palu, G. Comparison of INNO-LiPA genotyping extra and hybrid capture 2 assays for detection of carcinogenic human papillomavirus genotypes. *J. Clin. Virol.* **2012**, *55*, 256–261. [[CrossRef](#)] [[PubMed](#)]
20. Wood, H.M.; Bolt, R.; Hunter, K.D. Is next-generation sequencing an important tool in HPV subtype diagnosis? *Expert Rev. Mol. Diagn.* **2012**, *12*, 663–665. [[CrossRef](#)] [[PubMed](#)]
21. Ekstrom, J.; Bzhalava, D.; Svenback, D.; Forslund, O.; Dillner, J. High throughput sequencing reveals diversity of human papillomaviruses in cutaneous lesions. *Int. J. Cancer* **2011**, *129*, 2643–2650. [[CrossRef](#)] [[PubMed](#)]
22. Barzon, L.; Militello, V.; Lavezzo, E.; Franchin, E.; Peta, E.; Squarzon, L.; Trevisan, M.; Pagni, S.; Dal Bello, F.; Toppo, S.; *et al.* Human papillomavirus genotyping by 454 next generation sequencing technology. *J. Clin. Virol.* **2011**, *52*, 93–97. [[CrossRef](#)] [[PubMed](#)]
23. Militello, V.; Lavezzo, E.; Costanzi, G.; Franchin, E.; Di Camillo, B.; Toppo, S.; Palu, G.; Barzon, L. Accurate human papillomavirus genotyping by 454 pyrosequencing. *Clin. Microbiol. Infect.* **2013**, *19*, E428–E434. [[CrossRef](#)] [[PubMed](#)]
24. Arroyo, L.S.; Smelov, V.; Bzhalava, D.; Eklund, C.; Hultin, E.; Dillner, J. Next generation sequencing for human papillomavirus genotyping. *J. Clin. Virol.* **2013**, *58*, 437–442. [[CrossRef](#)] [[PubMed](#)]
25. Kukimoto, I.; Maehama, T.; Sekizuka, T.; Ogasawara, Y.; Kondo, K.; Kusumoto-Matsuo, R.; Mori, S.; Ishii, Y.; Takeuchi, T.; Yamaji, T.; *et al.* Genetic variation of human papillomavirus type 16 in individual clinical specimens revealed by deep sequencing. *PLoS ONE* **2013**, *8*, e80583. [[CrossRef](#)] [[PubMed](#)]
26. Conway, C.; Chalkley, R.; High, A.; MacLennan, K.; Berri, S.; Chengot, P.; Alsop, M.; Egan, P.; Morgan, J.; Taylor, G.R.; *et al.* Next-generation sequencing for simultaneous determination of human papillomavirus load, subtype, and associated genomic copy number changes in tumors. *J. Mol. Diagn.* **2012**, *14*, 104–111. [[CrossRef](#)] [[PubMed](#)]
27. Meiring, T.L.; Salimo, A.T.; Coetzee, B.; Maree, H.J.; Moodley, J.; Hitzeroth, I.I.; Freeborough, M.J.; Rybicki, E.P.; Williamson, A.L. Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits. *Virol. J.* **2012**, *9*. [[CrossRef](#)] [[PubMed](#)]
28. Garnaes, E.; Kiss, K.; Andersen, L.; Therkildsen, M.H.; Franzmann, M.B.; Filtenborg-Barnkob, B.; Hoegdall, E.; Krenk, L.; Josiassen, M.; Lajer, C.B.; *et al.* A high and increasing HPV prevalence in tonsillar cancers in Eastern Denmark, 2000–2010: The largest registry-based study to date. *Int. J. Cancer* **2015**, *136*, 2196–2203. [[CrossRef](#)] [[PubMed](#)]
29. Yi, X.; Zou, J.; Xu, J.; Liu, T.; Hua, S.; Xi, F.; Nie, X.; Ye, L.; Luo, Y.; Xu, L.; *et al.* Development and validation of a new HPV genotyping assay based on next-generation sequencing. *Am. J. Clin. Pathol.* **2014**, *141*, 796–804. [[CrossRef](#)] [[PubMed](#)]
30. Bzhalava, D.; Muhr, L.S.; Lagheden, C.; Ekstrom, J.; Forslund, O.; Dillner, J.; Hultin, E. Deep sequencing extends the diversity of human papillomaviruses in human skin. *Sci. Rep.* **2014**, *4*. [[CrossRef](#)] [[PubMed](#)]
31. Johansson, H.; Bzhalava, D.; Ekstrom, J.; Hultin, E.; Dillner, J.; Forslund, O. Metagenomic sequencing of “HPV-negative” condylomas detects novel putative HPV types. *Virology* **2013**, *440*, 1–7. [[CrossRef](#)] [[PubMed](#)]
32. Tang, K.W.; Alaei-Mahabadi, B.; Samuelsson, T.; Lindh, M.; Larsson, E. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat. Commun.* **2013**, *4*. [[CrossRef](#)] [[PubMed](#)]
33. Chandrani, P.; Kulkarni, V.; Iyer, P.; Upadhyay, P.; Chaubal, R.; Das, P.; Mulherkar, R.; Singh, R.; Dutt, A. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. *Br. J. Cancer* **2015**, *112*, 1958–1965. [[CrossRef](#)] [[PubMed](#)]
34. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)] [[PubMed](#)]
35. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)] [[PubMed](#)]
36. Linhart, C.; Shamir, R. The degenerate primer design problem. *Bioinformatics* **2002**, *18* (Suppl. 1), S172–S181. [[CrossRef](#)] [[PubMed](#)]
37. WHO. *Human Papillomavirus Laboratory Manual*, 1st ed.; Available online: [http://www.who.int/immunization/hpv/learn/hpv\\_laboratory\\_manual\\_who\\_ivb\\_2009\\_2010.pdf](http://www.who.int/immunization/hpv/learn/hpv_laboratory_manual_who_ivb_2009_2010.pdf) (accessed on 8 March 2016).
38. Papillomavirus Episteme. Available online: <http://pave.niaid.nih.gov/> (accessed on 31 October 2015).
39. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]

40. Tamura, K.; Stecher, G.; Peterson, D.; Filipowski, A.; Kumar, S. Mega6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)] [[PubMed](#)]
41. Schlecht, N.F.; Burk, R.D.; Palefsky, J.M.; Minkoff, H.; Xue, X.A.; Massad, L.S.; Bacon, M.; Levine, A.M.; Anastos, K.; Gange, S.J.; *et al.* Variants of human papillomaviruses 16 and 18 and their natural history in human immunodeficiency virus-positive women. *J. Gen. Virol.* **2005**, *86*, 2709–2720. [[CrossRef](#)] [[PubMed](#)]
42. Yang, L.; Yang, H.; Wu, K.; Shi, X.; Ma, S.; Sun, Q. Prevalence of HPV and variation of HPV 16/HPV 18 E6/E7 genes in cervical cancer in women in South West China. *J. Med. Virol.* **2014**, *86*, 1926–1936. [[CrossRef](#)] [[PubMed](#)]
43. Perez, S.; Cid, A.; Inarrea, A.; Pato, M.; Lamas, M.J.; Couso, B.; Gil, M.; Alvarez, M.J.; Rey, S.; Lopez-Miragaya, I.; *et al.* Prevalence of HPV 16 and HPV 18 lineages in Galicia, Spain. *PLoS ONE* **2014**, *9*, e104678. [[CrossRef](#)] [[PubMed](#)]
44. Chen, Z.; Schiffman, M.; Herrero, R.; DeSalle, R.; Anastos, K.; Segondy, M.; Sahasrabudde, V.V.; Gravitt, P.E.; Hsing, A.W.; Burk, R.D. Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70. *PLoS ONE* **2013**, *8*, e72565. [[CrossRef](#)] [[PubMed](#)]
45. Zhang, L.; Liao, H.; Yang, B.; Geffre, C.P.; Zhang, A.; Zhou, A.; Cao, H.; Wang, J.; Zhang, Z.; Zheng, W. Variants of human papillomavirus type 16 predispose toward persistent infection. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 8453–8459. [[PubMed](#)]
46. Arroyo, S.L.; Basaras, M.; Arrese, E.; Hernaez, S.; Andia, D.; Esteban, V.; Garcia-Etxebarria, K.; Jugo, B.M.; Cisterna, R. Human papillomavirus (HPV) genotype 18 variants in patients with clinical manifestations of HPV related infections in Bilbao, Spain. *Virol. J.* **2012**, *9*. [[CrossRef](#)] [[PubMed](#)]
47. Mosmann, J.P.; Monetti, M.S.; Frutos, M.C.; Kiguen, A.X.; Venezuela, R.F.; Cuffini, C.G. Mutation detection of E6 and LCR genes from HPV 16 associated with carcinogenesis. *Asian Pac. J. Cancer Prev.* **2015**, *16*, 1151–1157. [[CrossRef](#)] [[PubMed](#)]
48. Chen, A.A.; Gheit, T.; Franceschi, S.; Tommasino, M.; Clifford, G.M. Human papillomavirus 18 genetic variation and cervical cancer risk worldwide. *J. Virol.* **2015**, *89*, 10680–10687. [[CrossRef](#)] [[PubMed](#)]
49. Eklund, C.; Forslund, O.; Wallin, K.L.; Dillner, J. Global improvement in genotyping of human papillomavirus DNA: The 2011 HPV LabNet International Proficiency Study. *J. Clin. Microbiol.* **2014**, *52*, 449–459. [[CrossRef](#)] [[PubMed](#)]
50. Eklund, C.; Forslund, O.; Wallin, K.L.; Zhou, T.; Dillner, J. The 2010 global proficiency study of human papillomavirus genotyping in vaccinology. *J. Clin. Microbiol.* **2012**, *50*, 2289–2298. [[CrossRef](#)] [[PubMed](#)]
51. Tamura, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* **1992**, *9*, 678–687. [[PubMed](#)]
52. Hasegawa, M.; Kishino, H.; Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **1985**, *22*, 160–174. [[CrossRef](#)] [[PubMed](#)]
53. Molecular evolution, phylogenetics and epidemiology. Available online at: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 10 November 2015).
54. Geraets, D.T.; van Doorn, L.J.; Kleter, B.; Colau, B.; Harper, D.M.; Quint, W.G. Long-term follow-up of HPV16-positive women: Persistence of the same genetic variant and low prevalence of variant co-infections. *PLoS ONE* **2013**, *8*, e80382. [[CrossRef](#)] [[PubMed](#)]
55. Meza-Menchaca, T.; Williams, J.; Rodriguez-Estrada, R.B.; Garcia-Bravo, A.; Ramos-Ligonio, A.; Lopez-Monteon, A.; Zepeda, R.C. A low density microarray method for the identification of human papillomavirus type 18 variants. *Sensors* **2013**, *13*, 12975–12993. [[CrossRef](#)] [[PubMed](#)]
56. Larsson, G.L.; Helenius, G.; Andersson, S.; Elgh, F.; Sorbe, B.; Karlsson, M.G. Human papillomavirus (HPV) and HPV 16-variant distribution in vulvar squamous cell carcinoma in Sweden. *Int. J. Gynecol. Cancer* **2012**, *22*, 1413–1419. [[CrossRef](#)] [[PubMed](#)]
57. Xi, L.F.; Schiffman, M.; Koutsky, L.A.; Hulbert, A.; Lee, S.K.; Defilippis, V.; Shen, Z.; Kiviat, N.B. Association of human papillomavirus type 31 variants with risk of cervical intraepithelial neoplasia grades 2–3. *Int. J. Cancer* **2012**, *131*, 2300–2307. [[CrossRef](#)] [[PubMed](#)]
58. Sanchez, G.I.; Kleter, B.; Gheit, T.; van Doorn, L.J.; de Koning, M.N.; de Sanjose, S.; Alemany, L.; Bosch, X.F.; Tommasino, M.; Munoz, N.; *et al.* Clinical evaluation of polymerase chain reaction reverse hybridization assay for detection and identification of human papillomavirus type 16 variants. *J. Clin. Virol.* **2011**, *51*, 165–169. [[CrossRef](#)] [[PubMed](#)]

59. Chen, Z.; Schiffman, M.; Herrero, R.; Desalle, R.; Anastos, K.; Segondy, M.; Sahasrabudde, V.V.; Gravitt, P.E.; Hsing, A.W.; Burk, R.D. Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. *PLoS ONE* **2011**, *6*, e20183. [[CrossRef](#)] [[PubMed](#)]
60. Poljak, M.; Kocjan, B.J.; Ostrbenk, A.; Seme, K. Commercially available molecular tests for human papillomaviruses (HPV): 2015 update. *J. Clin. Virol.* **2015**. [[CrossRef](#)] [[PubMed](#)]
61. Ndiaye, C.; Mena, M.; Alemany, L.; Arbyn, M.; Castellsague, X.; Laporte, L.; Bosch, F.X.; de Sanjose, S.; Trottier, H. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: A systematic review and meta-analysis. *Lancet Oncol.* **2014**, *15*, 1319–1331. [[CrossRef](#)]
62. Barzon, L.; Cappellesso, R.; Peta, E.; Militello, V.; Sinigaglia, A.; Fassan, M.; Simonato, F.; Guzzardo, V.; Ventura, L.; Blandamura, S.; *et al.* Profiling of expression of human papillomavirus-related cancer miRNAs in penile squamous cell carcinomas. *Am. J. Pathol.* **2014**, *184*, 3376–3383. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).