Infection, Genetics and Evolution 73 (2019) 13–20

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research Paper

Human papillomavirus type 16 *E6* and *E7* gene variations associated with cervical cancer in a Han Chinese population

Ziyun Zhou[b,1], Hongying Yang[a,1], Lijuan Yang[a], Yueting Yao[b], Shuying Dai[c], Li Shi[b], Chuanyin Li[b], Longyu Yang[b], Zhiling Yan[a,*], Yufeng Yao[b,**]

[a] Department of Gynaecologic Oncology, The 3rd Affiliated Hospital of Kunming Medical University, Kunming 650118, China
[b] Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Kunming 650118, Yunnan, China
[c] School of Basic Medical Science, Kunming Medical University, Kunming 650500, China

ARTICLE INFO

Keywords:
Human papillomavirus type 16 (HPV16)
E6 gene
E7 gene
Cervical cancer
Variations
Association study

ABSTRACT

Backgroud: Human papillomavirus type 16 (HPV16) is a high-risk HPV subtype and a potent carcinogen. The HPV16 *E6* and *E7* genes are considered oncogenes that play a core role in the development of cervical cancer.
Methods: In the current study, we enrolled 97 HPV16-positive cervical cancer patients (case group) and 136 HPV16-positive asymptomatic individuals (control group) in a study to analyse the association between HPV16 *E6* and *E7* gene variations and cervical cancer.
Results: Our results showed that three HPV16 sub-lineages (A1-A3, A4 and D3) were present; the distribution of these variants between the case and control group was not significantly different ($P = 0.178$). When the distribution of the HPV16 *E6* and *E7* gene variations was compared, the distribution of only A131C (R10R) in the *E6* gene showed a different trend between the case and control groups and C749T (S63F) in the *E7* gene was significantly different between the case and control groups ($P = 0.071$ and $P = 4.861 \times 10^{-10}$, respectively). Regarding the sub-lineages, no variations in the *E6* gene were significantly different between the case and control group for the A4 (As) and A1-A3 (EUR) sub-lineages. However, the distribution of C749T (S63F) in the *E7* gene was significantly different between the case and control groups for the A4 (As) and A1-A3 (EUR) sub-lineages ($P = 1.815 \times 10^{-8}$ and $P = 0.008$). In the current study, we found that the C749T (S63F) variation in the HPV16 *E7* gene was associated with cervical cancer not only in the A4 (As) sub-lineage but also in the A1-A3 (EUR) sub-lineage.
Conclusion: Our study will provide a good reference for further functional studies of the relationship between cervical cancer carcinogenesis and the HPV16 *E6* and *E7* genes.

1. Background

Human papillomavirus (HPV) has been identified as a key aetiological factor in cervical cancer (Walboomers et al., 1999). Persistent high-risk HPV infection is the major factor contributing to the development of cervical cancer (Petry, 2014; Schiffman et al., 2007). Among these high-risk subtypes, HPV16 is present in more than half of cervical cancers (Bosch et al., 1995; Crow, 2012; Hung et al., 2008). Thus, recent studies have focused on the roles of HPV16 and its variants on the molecular basis underlying its oncogenic potential.

In 1993, Ho et al. studied HPV16 variants worldwide and they reported that the organization of HPV 16 LCR variants into a phylogenetic tree identified five major sub-lineages: European (EUR), Asian (As), Asian-American (AA), and two African lineages, African-1 and African-2 (AFR1 and AFR2) (Ho et al., 1993). Subsequently, Yamada et al. described one additional sub-lineage, which is the North American (NA) (Yamada et al., 1997). Cornet et al. further subdivided HPV16 variants into nine sub-lineages (EUR, As, AFR1a, AFR1b, AFR2a, AFR2b, NA, AA1, and AA2) base on *E6* and LCR in 2012 (Cornet et al., 2012). In 2013, Burk et al. reported that HPV16 can be divided into four main variant lineages (A, B, C and D) that including ten sub-lineages base on the whole HPV genome sequencing: A lineage including A1-A3 (previously named EUR) and A4 (As) sub-lineages; B lineage including B1 (AFR1a) and B2 (AFR1b) sub-lineages; C lineage (AFR2); D lineage

* Correspondence to: Z. Yan, Department of Gynaecologic Oncology, The 3rd Affiliated Hospital of Kunming Medical University, Kunming 650118, Yunnan, China.
** Corresponding author.
E-mail addresses: yzlyyf@163.com (Z. Yan), leoyyf@gmail.com (Y. Yao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.meegid.2019.04.008
Received 16 February 2019; Received in revised form 9 April 2019; Accepted 10 April 2019
Available online 11 April 2019

including D1 (NA1), D2 (AA2), and D3 (AA1) sub-lineages (Burk et al., 2013). Besides, studies have reported that the distribution of HPV16 variants is different in different regions, and that HPV16 variants are associated with persistent viral infection and the development of cervical cancer (Berumen et al., 2001; Bontkes et al., 1998; Hildesheim et al., 2001; Sichero et al., 2007; Villa et al., 2000; Xi et al., 1997; 2007).

The HPV16 genome is divided into three major regions: the early region, encoding early viral proteins; the late region, encoding late viral proteins; and the long control region (Burk et al., 2009). Research has shown that the HPV16 E6 and E7 proteins encoded by the viral early *E6* and *E7* genes interact with the p53 and retinoblastoma proteins, respectively (Boyer et al., 1996; Crook et al., 1991; Longworth and Laimins, 2004; Munger and Howley, 2002; zur Hausen, 2002). The E6 protein degrades p53 by forming an E6/E6AP complex to abrogate the growth arrest and apoptosis efficiently induced by p53 (Scheffner et al., 1993), and the E7 protein promotes cell proliferation by degrading the retinoblastoma tumour suppressor protein pRB (Dyson et al., 1989). In addition, the continued expression of the E6 and E7 proteins is necessary for the growth and tumourigenicity of cervical carcinoma cells (von Knebel Doeberitz et al., 1991; 1988).

Variations in the HPV16 *E6* and *E7* genes have been associated with the development of cervical cancer. For example, Andersson et al., Zehbe et al., and Grodzki et al. reported that the oncogenicity of the HPV16 *E6* T350G (L83V) variation was associated with an increased risk of cervical cancer that varied among Danish, Swedish and French populations (Andersson et al., 2000; Grodzki et al., 2006; Zehbe et al., 1998). However, other studies did not find an association between the HPV16 *E6* T350G (L83V) variation and cervical cancer (Cornet et al., 2013a; Freitas et al., 2014). This discrepancy indicates varied roles for HPV16 gene variations in the development of cervical cancer among different populations.

In the current study, we aimed to investigate the distribution of HPV16 *E6* and *E7* gene variations in HPV16-positive asymptomatic individuals and HPV16-positive cervical cancer patients in a Han Chinese population. Moreover, we evaluated the association of HPV16 *E6* and *E7* gene variations with cervical cancer.

2. Methods

2.1. Ethical statement

Each participant was informed of the study aims for obtaining informed consent. This study was approved by the Institutional Review Boards of the 3rd Affiliated Hospital of Kunming Medical University, and was carried out in line with the Helsinki Declaration.

2.2. Subjects

In the present study, the case group included 97 HPV16-positive cervical cancer patients, and 136 HPV16-positive asymptomatic individuals undergoing routine health check-ups were recruited as the control group. All participants were recruited at the 3rd Affiliated Hospital of Kunming Medical University during 2013 to 2017. The diagnosis of cervical cancer was confirmed by pathological examination according to “Diagnosis and Treatment, Obstetrics and Gynaecology” and the FIGO stage (International Federation of Gynaecology and Obstetrics, 2009). Subjects with oncotherapy histories, other malignancies or incomplete clinical data were excluded from the present study. The inclusion criteria for HPV16-positive asymptomatic individuals were HPV16-positivity, female, and the absence of lesions in the cervix. All participants in this study were self-reported as ethnically Han.

2.3. HPV genotyping

For HPV genotyping, the Tellgenplex™ HPV DNA Test, which is a suspension bead array method, was used to identify HPV types. The experimental protocol referred to that in a previous study (Li et al., 2016). Human β-globin was used as an internal control for each reaction.

2.4. Amplification and sequencing of the HPV16 E6 and E7 genes

The HPV16 *E6* and *E7 genes* were amplified using a high-fidelity DNA polymerase with specific primers. The PCR product contained the complete *E6* and *E7* sequences, with sizes of 456 and 297 bp, respectively. The PCRs were performed in a final volume of 25 μl, and each PCR mixture contained 1 × Q5 PCR buffer (containing Mg^{2+}), 200 μM dNTPs, 0.5 μM sense and antisense primers, 50 ng of genomic DNA, and 0.02 U/μl high-fidelity DNA polymerase. The PCR conditions were as follows: an initial denaturation step at 95 °C for 3 min; 35 cycles of denaturation at 95 °C for 10 s, annealing at 51/55 °C for 60/45 s, and extension at 72 °C for 60 s; and a final elongation step at 72 °C for 2 min. The amplified DNA fragments (including the complete *E6* and *E7* gene sequences) were sent to Shanghai Sangon Biotech for sequencing after visualization via 2.0% agarose electrophoresis. The *E6* and *E7* gene PCR and sequencing primers are shown in Table 1.

2.5. HPV16 E6 and E7 gene variant identification and phylogenetic analyses

All of the sequences of the HPV16 *E6* and *E7* genes were assembled by the SeqMan tool included in Lasergene v 7.1 software and were aligned by the ClustalW multiple sequence alignment tool included in Molecular Evolutionary Genetics Analysis (MEGA) v 7.0 software (Kumar et al., 2016). The HPV16 reference sequence HQ644251.1 was used to identify variations in the HPV16 *E6* and *E7* gene sequences. Phylogenetic trees of the HPV16 *E6* and *E7* genes were constructed using MEGA 7.0 software by the neighbour-joining method with 1000-fold bootstrapping. The reference sequences used to construct the phylogenetic tree were obtained from GenBank. The accession numbers are HQ644282.1 (A1), HQ644283.1 (A1), HQ644268.1 (A1), HQ644280.1 (A1), AF536179.1 (A2), HQ644236.1 (A3), HQ644248.1 (A4), HQ644251.1 (A4), AF534061.1 (A4), HQ644235.1 (A4), HQ644290.1 (B1), HQ644238.1 (B1), HQ644240.1 (B1), HQ644298.1 (B2), HQ644249.1 (C), HQ644250.1 (C), AF472509.1 (C), HQ644237.1 (C), HQ644239.1 (C), HQ644257.1 (D1), HQ644279.1 (D2), HQ644281.1 (D2), HQ644263.1 (D2), HQ644277.1 (D2), HQ644247.1 (D3), HQ644253.1 (D3), HQ644255.1 (D3), AF402678.1 (D3).

Table 1
Primers and PCR cycling conditions for HPV16 *E6* and *E7* genes.

HPV16 gene	PCR primers (5′-3′)	Sequencing primers	Annealing conditions
E6	F-CGAAACCGGTTAGTATAA R-GTATCTCCATGCATGATT	F-CGAAACCGGTTAGTATAA R-GTATCTCCATGCATGATT	51 °C/60s
E7	F-ATAATATAAGGGGTCGGTGG R-CATTTTCGTTCTCGTCATCTG	R-CATTTTCGTTCTCGTCATCTG	55 °C/45 s

2.6. Statistical analysis

The ages between case and control groups were compared using the Student's *t*-test. A chi-squared test was performed to analyse the distribution difference of HPV16 variants between case and control group. The frequency of each HPV16 *E6* and *E7* gene variations was determined by direct counting and Fisher's exact test was used to analyse the relationship between variations and cervical cancer. Statistical analyses were performed using SPSS 13 (Chicago, IL), and *P* values less than 0.05 were considered statistically significant.

3. Results

3.1. Subject characteristics

In this study, the case group included 97 HPV16-positive cervical cancer patients with squamous cell carcinoma (SCC), the age distribution was 45.19 ± 9.15 (mean ± SD), and the control group included 136 HPV16-positive asymptomatic individuals, the age distribution was 46.63 ± 9.46 (mean ± SD). There was no significant difference in age between the case and control group ($P = 0.244$).

3.2. Distribution of HPV16 variants in the case and control groups

In the current study, based on the *E6* gene sequences, there were three HPV16 sub-lineages in the case group, namely, the A1-A3 (EUR), A4 (As) and D3 (AA1) sub-lineages, which had distribution frequencies of 28.9%, 70.1% and 1.0%, respectively (Fig. 1). In the control group, only two sub-lineages (A1-A3 and A4) were found, with distribution frequencies of 38.2% and 61.8%, respectively (Fig. 2). Based on the *E7* gene sequences, there were also three HPV16 sub-lineages in the case group, namely, the A1-A3 (EUR), A4 (As) and D3 (AA1) sub-lineages, with distribution frequencies of 28.9%, 70.1% and 1.0%, respectively (Fig. 3). In the control group, only two sub-lineages (A1-A3 and A4) were found, with distribution frequencies of 38.2% and 61.8%, respectively (Fig. 4). There was no difference in the distribution of the HPV16 variants between the case group and control group ($P = 0.178$).

3.3. HPV16 E6 gene variations in the case and control groups

Table 2 shows the HPV16 *E6* gene variations in the case and control groups. In the case group, twelve variations were observed; eight were non-synonymous, and the remaining four were synonymous. In the control group, nine variations were detected, eight of which were non-synonymous and one of which was synonymous. The distribution of the A131C (R10R) variation showed a different trend between the case group and control group ($P = 0.071$) (Table 2). For the A4 (As) sub-lineage, three and two variations were found in the case and control groups, respectively. In the case group, two variations were non-synonymous, and only one was synonymous. In the control group, both variations were non-synonymous. The distribution of these variations was not significantly different between the case and control groups in the A4 (As) sub-lineage (Table 2). For the A1-A3 (EUR) sub-lineage, a total of five variations were observed in the case group, all of which were non-synonymous. Eight variations were found in the control group; among these, seven were non-synonymous, and one was synonymous. In the comparison of these variations' distributions between the case group and control group, we found that no distributions were significantly different (Table 2). For the D3 (AA1) sub-lineage, which was found only in the case group, a total of eight variations were identified; and among these, five were non-synonymous (Table 2).

3.4. HPV16 E7 gene variations in the case and control groups

The HPV16 *E7* gene variations in the case and control groups are shown in Table 3. In the case group, twelve variations were observed, five of which were non-synonymous, and the remaining seven of which were synonymous. In the control group, eleven variations were detected; four were non-synonymous and seven were synonymous. We found that the distribution of only the C749T (S63F) variation was significantly different between the case group and the control group ($P = 4.861 \times 10^{-10}$). For the A4 (As) sub-lineage, five variations were found in the case group and six were found in the control group during the analysis of those sequences. In the case group, two variations were nonsynonymous and three were synonymous. In the control group, two variations were non-synonymous, and four were synonymous. The distribution of C749T (S63F) was significantly different between the case and control groups for the A4 (As) sub-lineage ($P = 1.815 \times 10^{-8}$) (Table 3). For the A1-A3 (EUR) sub-lineage, a total of eight variations were observed in the case group; five of these were non-synonymous and three were synonymous. In addition, a total of eight variations were found in the control group, among which four were non-synonymous and four were synonymous. In the comparison of these variations' distributions between the case group and control group, we found that the distribution of C749T (S63F) was significantly different in all the cases and controls, as well as in the A4 (EUR) and A1-A3 (EUR) sub-lineages. ($P = 4.861 \times 10^{-10}$, $P = 1.815 \times 10^{-8}$, $P = 0.008$) (Table 3). For the D3 (AA1) sub-lineage, which was found only in the case group, a total

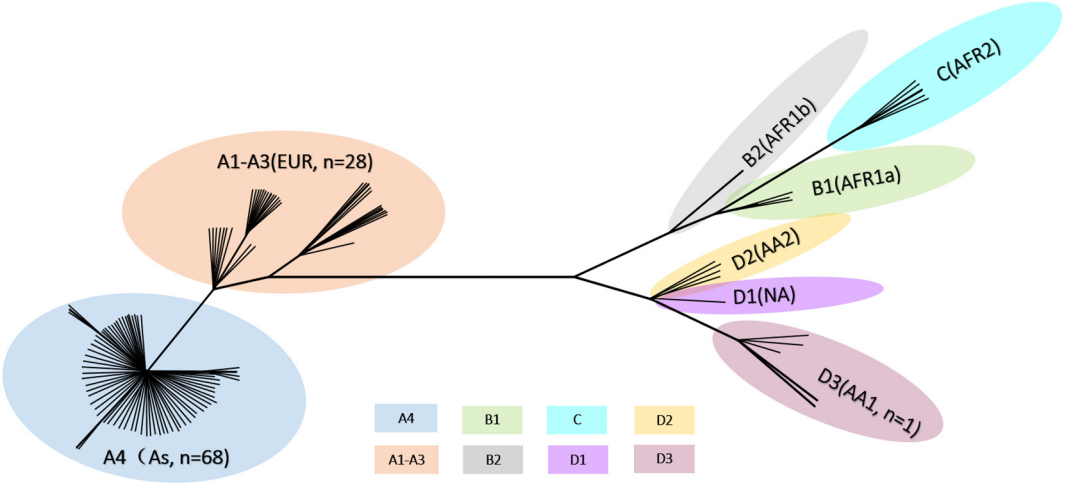

Fig. 1. Phylogenetic tree based on the HPV16 *E6* gene sequences from HPV16-positive cervical cancer patients, corresponding to the HQ644251.1 reference sequence (nucleotide positions 104–559).

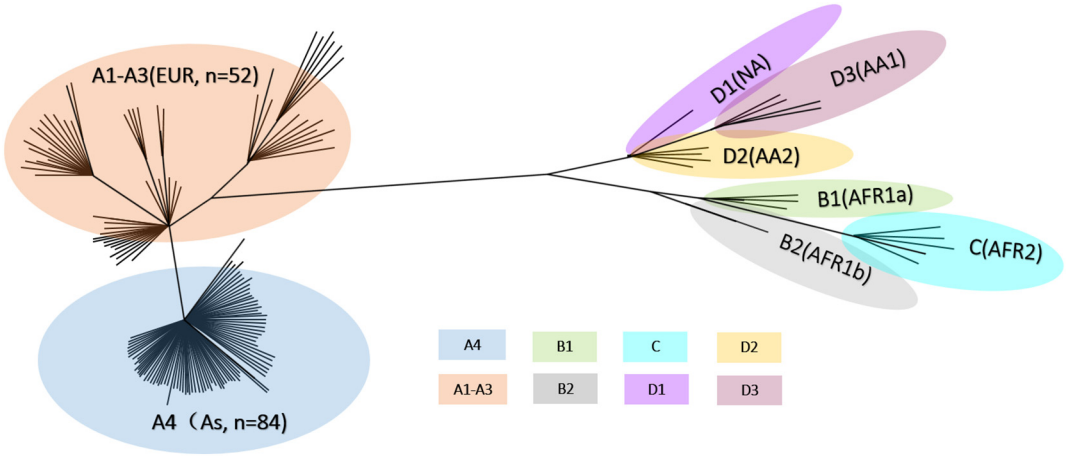

Fig. 2. Phylogenetic tree based on the HPV16 *E6* gene sequences from HPV16-positive asymptomatic individuals, corresponding to the HQ644251.1 reference sequence (nucleotide positions 104–559).

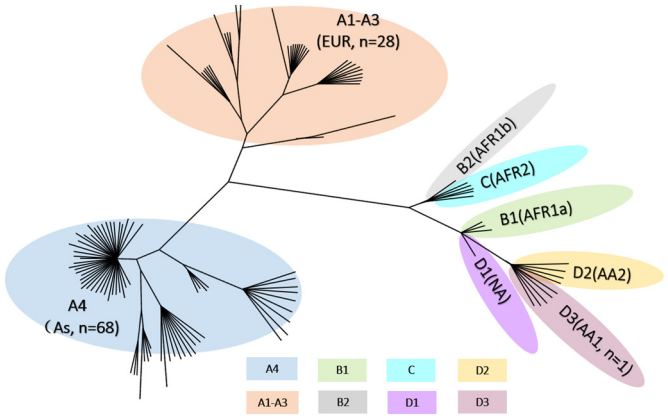

Fig. 3. Phylogenetic tree based on the HPV16 *E7* gene sequences from HPV16-positive cervical cancer patients, corresponding to the HQ644251.1 reference sequence (nucleotide positions 562–858).

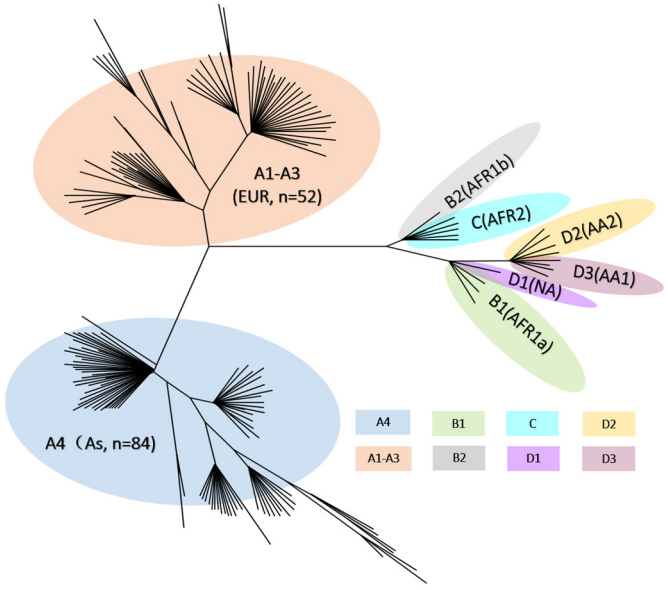

Fig. 4. Phylogenetic tree based on the HPV16 *E7* gene sequences from HPV16-positive asymptomatic individuals, corresponding to the HQ644251.1 reference sequence (nucleotide positions 562–858).

of five variations were identified, and among these variations, only one was non-synonymous (Table 3).

4. Discussion

Among all HPV16 proteins, the HPV16 E6 and E7 proteins encoded by the viral early *E6* and *E7* genes, which are two transforming oncogenes, have been associated with the development of cervical cancer. In the current study, we analysed the distribution of HPV16 variants and *E6* and *E7* gene variations and amino acid substitutions between HPV16-positive cervical cancer patients and HPV16-positive asymptomatic individuals to identify HPV16 *E6* and *E7* gene variations related to cervical cancer.

It is well known that the distribution of HPV variants is associated with the geographical population distribution. For example, in eastern Asia, the A1-A3 (EUR) sub-lineage is predominant in South Korea (54.1%) (Park et al., 2016) and Japan (55.8%) (Shang et al., 2011). In southern Asia, the A4 (As) sub-lineage is common in northeast Thailand (73.9%)(Shang et al., 2011). In China, the A1-A3 (EUR) sub-lineage was predominant in Heilongjiang, Jilin, and the northern Inner Mongolia region, which lies in Northeast China (67.3%) (Shang et al., 2011). However, the A4 (As) sub-lineage was common in Jiangxi and Guangdong provinces, which lie in central and South China (60.3% and 65.5%, respectively) (Shang et al., 2011). However, the A1-A3 (EUR) sub-lineage was also predominant in North India (86.7%)(Shang et al., 2011), which lies in Southern Asia, indicating that the distribution of HPV variants is also related to population or ethnic distributions. Interestingly, the distribution of HPV16 variants was significantly different between Sichuan (EUR: 69.0–76.2% and As: 31.0–23.81%) and Yunnan (EUR: 28.9–38.2% and As: 61.8–70.1%) provinces, although these two provinces are neighbouring regions located in Southwest China. This discrepancy indicated that the distribution of HPV16 variants is not only affected by geography and/or the population genetic background but also by other factors such as the environment.

Recently, studies have shown that different HPV16 variants play key roles in the development of cervical cancer (Cornet et al., 2013b; Smith et al., 2011). Compared with the A1-A3 (EUR) sub-lineage, the non-EUR sub-lineage, such as the A4 (As), D2-D3 (AA), B1-B2 and C (AFR) sub-lineages, are associated with an increased risk of cervical intraepithelial lesions and cervical cancer (Dai et al., 2018; Hang et al., 2016; Hildesheim et al., 2001; Ortiz-Ortiz et al., 2015; Schiffman et al., 2010; Sichero et al., 2007; Sun et al., 2013; Xi et al., 2007). In 2006, Wu et al. reported that the A4 (As) sub-lineage was more frequently found in cervical premalignancies and cancers than the other sub-lineages, which indicated that the A4 (As) sub-lineage might contribute to the high incidence of cervical cancer in China (Wu et al., 2006). In the

Table 2
HPV16 *E6* gene variations and amino acid substitutions in the case and control groups.

	Genome position[a]	Case		Control		*P* value*	Amino acid[a]	Case		Control		*P* value*
		Mutation	Frequency (%)	Mutation	Frequency (%)			Mutation	Frequency (%)	Mutation	Frequency (%)	
All	A131C	3	3.1	0	0.0	0.071	R10R	0	0.0	0	0.0	–
Case (n = 97)	G145 T	1	1.0	0	0.0	0.416	Q14H	1	1.0	0	0.0	0.416
control	G162A	97	100.0	136	100.0	–	R20Q	97	100.0	136	100.0	–
(n = 136)	C168G	6	6.2	6	4.4	0.562	T22S	6	6.2	6	4.4	0.562
	G178 T	29	29.9	52	38.2	0.211	E25D	29	29.9	52	38.2	0.211
	T185G	4	4.1	6	4.4	1.000	L28 V	4	4.1	6	4.4	1.000
	G188C	0	0.0	2	1.5	0.512	E29Q	0	0.0	2	1.5	0.512
	T241G	0	0.0	4	2.9	0.143	A46A	0	0.0	0	0.0	–
	A276G	12	12.4	17	12.5	1.000	N58S	12	12.4	17	12.5	1.000
	T286A	1	1.0	0	0.0	0.416	A61A	0	0.0	0	0.0	–
	A289G	1	1.0	0	0.0	0.416	V62 V	0	0.0	0	0.0	–
	C335T	1	1.0	0	0.0	0.416	H78Y	1	1.0	0	0.0	0.416
	T350G	8	8.2	11	8.1	1.000	L83V	8	8.2	11	8.1	1.000
	A532G	1	1.0	0	0.0	0.416	S143S	0	0.0	0	0.0	–
	G534C	0	0.0	2	1.5	0.512	R144T	0	0.0	2	1.5	0.512
A4 (As)	A131C	3	4.4	0	0.0	0.087	R10R	0	0.0	0	0.0	–
Case (n = 68)	G162A	68	100.0	84	100.0	–	R20Q	68	100.0	84	100.0	–
control	T185G	4	5.9	6	7.1	1.000	L28 V	4	5.9	6	7.1	1.000
(n = 84)												
A1-A3 (EUR)	G162A	28	100.0	52	100.0	–	R20Q	28	100.0	52	100.0	–
Case (n = 28)	C168G	6	21.4	6	11.5	0.326	T22S	6	21.4	6	11.5	0.326
control	G178 T	28	100.0	52	100.0	–	E25D	28	100.0	52	100.0	–
(n = 52)	G188C	0	0.0	2	3.8	0.539	E29Q	0	0.0	2	3.8	0.539
	T241G	0	0.0	4	7.7	0.292	A46A	0	0.0	0	0.0	–
	A276G	12	42.9	17	32.7	0.466	N58S	12	42.9	17	32.7	0.466
	T350G	7	25.0	11	21.2	0.781	L83V	7	25.0	11	21.2	0.781
	G534C	0	0.0	2	3.8	0.539	R144T	0	0.0	2	3.8	0.539
D3 (AA1)	G145 T	1	100.0	–	–	–	Q14H	1	100.0	–	–	–
Case (n = 1)	G162A	1	100.0	–	–	–	R20Q	1	100.0	–	–	–
control	G178 T	1	100.0	–	–	–	E25D	1	100.0	–	–	–
(n = 0)	T286A	1	100.0	–	–	–	A61A	0	0.0	–	–	–
	A289G	1	100.0	–	–	–	V62 V	0	0.0	–	–	–
	C335T	1	100.0	–	–	–	H78Y	1	100.0	–	–	–
	T350G	1	100.0	–	–	–	L83V	1	100.0	–	–	–
	A532G	1	100.0	–	–	–	S143S	0	0.0	–	–	–

[a] The reference HPV16 *E6* gene sequence was HQ644251.1.
* Fisher's exact test *P* value, and *P* values less than 0.05 was considered statistically significant.

current study, we did not observe any association between the A4 (As) sub-lineage and cervical cancer; however, our results showed that the frequency of the A4 (As) sub-lineage was higher in the case group (70.1%) than in the control group (61.8%). One of the reasons that our results showed no difference could be that the relatively modest sample size limited the ability to detect a difference between the frequency of the A4 (As) and A1-A3 (EUR) sub-lineages.

As the HPV16 E6 protein plays an important role in the development of cervical cancer, variations in the HPV16 *E6* gene could also influence the progression of cervical malignancy. Andersson et al. and Grodzki et al. reported that the oncogenicity of HPV16 *E6* T350G (L83V) was associated with an increased risk of cervical cancer that varied among Danish, Swedish and French populations (Andersson et al., 2000; Grodzki et al., 2006; Zehbe et al., 1998). However, other studies did not find an association between the T350G (L83V) variation and cervical cancer (Cornet et al., 2013a; Freitas et al., 2014). In the current study, we did not find that T350G (L83V) was associated with cervical cancer. These discrepant data indicate that the roles of *E6* L83V in the development of cervical cancer vary among populations. However, we found that the distribution of A131C (R10R) showed a different trend between the case group and the control group ($P = 0.071$) and that this variation was synonymous.

In 1994, Fujinaga et al. reported that there were two variation hot spots in the *E7* gene, A647G(N29S) and T846C (S95S), which were detected in nine of 15 and seven of 15 Japanese cervical cancer cases, respectively (Fujinaga et al., 1994). Then, in 2002, Chan et al. observed a high frequency of these variations 58.0% for A647G(N29S) and 52.9% for T846C (S95S) (Chan et al., 2002). In 2006, Wu et al. reported that A647G(N29S), C749T (S63F) and T846C (S95S) were three variation hot spots on the *E7* gene, with variation frequencies of 70.2%, 51.1% and 61.7%, respectively, in Chinese cervical cancers (Wu et al., 2006). In the current study, the frequencies of A647G(N29S), C749T (S63F) and T846C (S95S) were 76.3%, 77.3% and 70.1% in the case group and 69.1%, 36.0%, and 61.8% in the control group, respectively. Previous studies showed that A647G(N29S) was more frequent in the carcinoma group (70.0%) than in the control group (33.3%) or the CIN 3 group (50.0%) in a Korean population (Song et al., 1997). However, in southern China and Japan, the distribution of this variation was not different between case and control group subjects (Chan et al., 2002; Fujinaga et al., 1994). In 2002, Radhakrishna Pillai et al. reported that there were no associations between *E7* variation and tumour stage or patient age in an Indian population (Radhakrishna Pillai et al., 2002). In the current study, we also found that there was no difference in the distribution of A647G(N29S) between the case and control groups. These results suggested that genetic differences between different populations led to viral genomes acquiring different variations. In the current study, we found that the distribution of C749T (S63F) was significantly different between the case and control groups overall, as well as for the A4 (As) and the A1-A3 (EUR) sub-lineages ($P = 4.861 \times 10^{-10}$, $P = 1.815 \times 10^{-8}$, $P = 0.008$). One of the reasons for the association between the C749T (S63F) variation of the HPV16 E7 gene with cervical cancer could be that this variation influences the E7 epitopes and causes viral persistence and cervical cancer. In 2013, we predicted putative cytotoxic lymphocyte (CTL) epitopes on the HPV-16 E6 and E7 proteins using immunoinformatic methods and found that C749T (S63F) was located in the HLA-A*02:01

Table 3
HPV16 *E7* gene variations and amino acid substitutions in the case and control groups.

	Genome position[a]	Case		Control		*P* value*	Amino acid[a]	Case		Control		*P* value*
		Mutation	Frequency (%)	Mutation	Frequency (%)			Mutation	Frequency (%)	Mutation	Frequency (%)	
All	A576G	0	0.0	2	1.5	0.512	T5T	0	0.0	0	0.0	–
Case (n = 97) control (n = 136)	C625T	1	1.0	0	0.0	0.416	L22F	1	1.0	0	0.0	0.416
	C627T	12	12.4	15	11.0	0.836	L22 L	0	0.0	0	0.0	–
	A646C	6	6.2	10	7.4	0.798	S29H	6	6.2	10	7.4	0.798
	G647A											
	A646A	23	23.7	42	30.9	0.211	S29 N	23	23.7	42	30.9	0.211
	G647A											
	G666A	19	19.6	34	25.0	0.347	E35E	0	0.0	0	0.0	–
	T732C	2	2.1	0	0.0	0.172	F57F	0	0.0	0	0.0	–
	C749T	75	77.3	49	36.0	4.861×10^{-10}	S63F	75	77.3	49	36.0	4.861×10^{-10}
	T760C	0	0.0	2	1.5	0.512	L67 L	0	0.0	0	0.0	–
	T789C	1	1.0	0	0.0	0.416	I76I	0	0.0	0	0.0	–
	C790T	3	3.1	3	2.2	0.695	R77C	3	3.1	3	2.2	0.695
	T795G	1	1.0	0	0.0	0.416	T78 T	0	0.0	0	0.0	–
	C840T	0	0.0	1	0.7	1.000	I93I	0	0.0	0	0.0	–
	T843C	23	23.7	27	19.9	0.519	C94C	0	0.0	0	0.0	–
	C846T	29	29.9	52	38.2	0.211	S95S	0	0.0	0	0.0	–
A4 (As)	A576G	0	0.0	2	2.4	0.502	T5T	0	0.0	0	0.0	–
Case (n = 68) control (n = 84)	C627T	8	11.8	9	10.7	1.000	L22 L	0	0.0	0	0.0	–
	T732C	1	1.5	0	0.0	0.447	F57F	0	0.0	0	0.0	–
	C749T	58	85.3	34	40.5	1.815×10^{-8}	S63F	58	85.3	34	40.5	1.815×10^{-8}
	C790T	1	1.5	2	2.4	1.000	R77C	1	1.5	2	2.4	1.000
	C840T	0	0.0	1	1.2	1.000	I93I	0	0.0	0	0.0	–
	T843C	23	33.8	27	32.1	0.863	C94C	0	0.0	0	0.0	–
A1-A3 (EUR)	C625T	1	3.6	0	0.0	0.350	L22F	1	3.6	0	0.0	0.350
Case (n = 28) control (n = 52)	C627T	4	14.3	6	11.5	0.734	L22 L	0	0.0	0	0.0	–
	A646C	6	21.4	10	19.2	1.000	S29H	6	21.4	10	19.2	1.000
	G647A											
	A646A	22	78.6	42	80.8	1.000	S29 N	22	78.6	42	80.8	1.000
	G647A											
	G666A	19	67.9	34	65.4	1.000	E35E	0	0.0	0	0.0	–
	C749T	17	60.7	15	28.8	0.008	S63F	17	60.7	15	28.8	0.008
	T760C	0	0.0	2	3.8	0.539	L67 L	0	0.0	0	0.0	–
	C790T	2	7.1	1	1.9	0.279	R77C	2	7.1	1	1.9	0.279
	C846T	28	100.0	52	100.0	–	S95S	0	0.0	0	0.0	–
D3 (AA1)	A646A	1	100.0	–	–	–	S29 N	1	100.0	–	–	–
Case (n = 1) control (n = 0)	G647A											
	T732C	1	100.0	–	–	–	F57F	0	0.0	–	–	–
	T789C	1	100.0	–	–	–	I76I	0	0.0	–	–	–
	T795G	1	100.0	–	–	–	T78 T	0	0.0	–	–	–
	C846T	1	100.0	–	–	–	S95S	0	0.0	–	–	–

[a] The reference HPV16 *E6* gene sequence was HQ644251.1.
* Fisher's exact test *P* value, and *P* values less than 0.05 was considered statistically significant.

and HLA-A*33:03 epitopes (Yao et al., 2013). Our results indicated that this variation could affect the adaptive immune response to the virus. The other reason could be that the variation is influenced by genetic selection pressure in the host. The apolipoprotein B messenger RNA-editing, enzyme-catalytic, polypeptide-like 3 (APOBEC3) family of cytidine deaminases plays an important role in the innate immune response to viral infections by editing viral genomes(Warren et al., 2017). Wakae et al. reported that C-to-T and G-to-A hypermutations introduced by APOBEC3 accumulated in the HPV16 genome (Wakae et al., 2015). In 2017, Mirabello et al. showed that the HPV16 *E7* gene variation C749T (S63F) could be a specific DNA modification induced by the potentially antiviral activity of human APOBEC3 (Mirabello et al., 2017).

Also in 2017, Mirabello et al. used a novel HPV whole genome sequencing technique to evaluate a large number of HPV16-infected case-control samples and found that the *E7* gene was devoid of variations in precancers/cancers but that the level of variations was increased in the control (Mirabello et al., 2017). However, in the current study, we found that the frequency of *E7* gene variations was almost equal in the case and control groups. One of the reasons for this discrepancy could be the predominant sub-lineages and the sample size. There were smaller numbers of other sub-lineages, especially the A4 (As) sub-lineage, in the Mirabello et al. study. However, the A4 (As) sub-lineage was predominant in our population. Thus, variations in the *E7* gene could be different among sub-lineages. In the future, samples containing a larger proportion of other sub-lineages should be analysed.

In the current study, we found that the HPV16 *E7* gene variation C749T (S63F) played an important role in the development of cervical cancer. Moreover, this variation was associated with cervical cancer not only in the A4 (As) sub-lineage but also in the A1-A3 (EUR) sub-lineage. Our results will provide a good reference for further functional studies of the relationship between cervical cancer carcinogenesis and the HPV16 *E6* and *E7* genes.

Acknowledgements

This work was supported by National Science Foundation of China (81573206 and 31270030), CAMS Innovation Fund for Medical Sciences (CIFMS) (2017-I2M-2-006), Yunnan Applied Basic Research Projects (2016FA034), Fundamental Research Funds for the Central Universities and the PUMC Youth Fund (3332015149), The Association Foundation Program of Yunnan Provincial Science and Technology Department and Kunming Medical University (2017FE467-077 and 2017FE467-012), and Special Funds for high-level health talents of Yunnan Province (D-201669 and L-201615). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Authors' contributions

Zhiling Yan and Yufeng Yao conceived and designed the experiments; Ziyun Zhou, Hongying Yang and Lijuan Yang performed the experiments; Yueting Yao and Li Shi analysed the data; Chuanyin Li, Longyu Yang and Shuying Dai contributed materials; and Yufeng Yao wrote the paper.

Ethics approval and consent to participate

All procedures performed in this study involving human participants were in accordance with the ethical standards of the Institutional Ethics Committee of Institutional Review Boards of the 3rd Affiliated Hospital of Kunming Medical University and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in the study.

Conflict of interest

The authors have no conflicts of interest to declare.

References

Andersson, S., Alemi, M., Rylander, E., Strand, A., Larsson, B., Sallstrom, J., Wilander, E., 2000. Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions. Br. J. Cancer 83, 307–310.

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., 2001. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. J. Natl. Cancer Inst. 93, 1325–1330.

Bontkes, H.J., van Duin, M., de Gruijl, T.D., Duggan-Keen, M.F., Walboomers, J.M., Stukart, M.J., Verheijen, R.H., Helmerhorst, T.J., Meijer, C.J., Scheper, R.J., Stevens, F.R., Dyer, P.A., Sinnott, P., Stern, P.L., 1998. HPV 16 infection and progression of cervical intra-epithelial neoplasia: analysis of HLA polymorphism and HPV 16 E6 sequence variants. Int. J. Cancer 78, 166–171.

Bosch, F.X., Manos, M.M., Munoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V., 1995. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J. Natl. Cancer Inst. 87, 796–802.

Boyer, S.N., Wazer, D.E., Band, V., 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. Cancer Res. 56, 4620–4624.

Burk, R.D., Chen, Z., Van Doorslaer, K., 2009. Human papillomaviruses: genetic basis of carcinogenicity. Public Health Genomics 12, 281–290.

Burk, R.D., Harari, A., Chen, Z., 2013. Human papillomavirus genome variants. Virology 445, 232–243.

Chan, P.K., Lam, C.W., Cheung, T.H., Li, W.W., Lo, K.W., Chan, M.Y., Cheung, J.L., Xu, L.Y., Cheng, A.F., 2002. Human papillomavirus type 16 intratypic variant infection and risk for cervical neoplasia in southern China. J. Infect. Dis. 186, 696–700.

Cornet, I., Gheit, T., Franceschi, S., Vignat, J., Burk, R.D., Sylla, B.S., Tommasino, M., Clifford, G.M., Group, I.H.V.S., 2012. Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR. J. Virol. 86, 6855–6861.

Cornet, I., Gheit, T., Clifford, G.M., Combes, J.D., Dalstein, V., Franceschi, S., Tommasino, M., Clavel, C., 2013a. Human papillomavirus type 16 E6 variants in France and risk of viral persistence. Infect Agent Cancer 8, 4.

Cornet, I., Gheit, T., Iannacone, M.R., Vignat, J., Sylla, B.S., Del Mistro, A., Franceschi, S., Tommasino, M., Clifford, G.M., 2013b. HPV16 genetic variation and the development of cervical cancer worldwide. Br. J. Cancer 108, 240–244.

Crook, T., Tidy, J.A., Vousden, K.H., 1991. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. Cell 67, 547–556.

Crow, J.M., 2012. HPV: the global burden. Nature 488, S2–S3.

Dai, S., Yao, Y., Yan, Z., Zhou, Z., Shi, L., Wang, X., Sun, L., Zhang, R., Yao, Y., 2018. The association of human papillomavirus type 16 E2 variations with cervical cancer in a Han Chinese population. Infect. Genet. Evol. 64, 241–248.

Dyson, N., Howley, P.M., Munger, K., Harlow, E., 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 243, 934–937.

Freitas, L.B., Chen, Z., Muqui, E.F., Boldrini, N.A., Miranda, A.E., Spano, L.C., Burk, R.D., 2014. Human papillomavirus 16 non-European variants are preferentially associated with high-grade cervical lesions. PLoS ONE 9, e100746.

Fujinaga, Y., Okazawa, K., Nishikawa, A., Yamakawa, Y., Fukushima, M., Kato, I., Fujinaga, K., 1994. Sequence variation of human papillomavirus type 16 E7 in pre-invasive and invasive cervical neoplasias. Virus Genes 9, 85–92.

Grodzki, M., Besson, G., Clavel, C., Arslan, A., Franceschi, S., Birembaut, P., Tommasino, M., Zehbe, I., 2006. Increased risk for cervical disease progression of French women infected with the human papillomavirus type 16 E6-350G variant. Cancer Epidemiol. Biomark. Prev. 15, 820–822.

Hang, D., Yin, Y., Han, J., Jiang, J., Ma, H., Xie, S., Feng, X., Zhang, K., Hu, Z., Shen, H., Clifford, G.M., Dai, M., Li, N., 2016. Analysis of human papillomavirus 16 variants and risk for cervical cancer in Chinese population. Virology 488, 156–161.

Hildesheim, A., Schiffman, M., Bromley, C., Wacholder, S., Herrero, R., Rodriguez, A., Bratti, M.C., Sherman, M.E., Scarpidis, U., Lin, Q.Q., Terai, M., Bromley, R.L., Buetow, K., Apple, R.J., Burk, R.D., 2001. Human papillomavirus type 16 variants and risk of cervical cancer. J. Natl. Cancer Inst. 93, 315–318.

Ho, L., Chan, S.Y., Burk, R.D., Das, B.C., Fujinaga, K., Icenogle, J.P., Kahn, T., Kiviat, N., Lancaster, W., Mavromara-Nazos, P., et al., 1993. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J. Virol. 67, 6413–6423.

Hung, C.F., Ma, B., Monie, A., Tsen, S.W., Wu, T.C., 2008. Therapeutic human papillomavirus vaccines: current clinical trials and future directions. Expert. Opin. Biol. Ther. 8, 421–439.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Li, Z., Liu, F., Cheng, S., Shi, L., Yan, Z., Yang, J., Shi, L., Yao, Y., Ma, Y., 2016. Prevalence of HPV infection among 28,457 Chinese women in Yunnan Province, Southwest China. Sci. Rep. 6, 21039.

Longworth, M.S., Laimins, L.A., 2004. Pathogenesis of human papillomaviruses in differentiating epithelia. Microbiol. Mol. Biol. Rev. 68, 362–372.

Mirabello, L., Yeager, M., Yu, K., Clifford, G.M., Xiao, Y., Zhu, B., Cullen, M., Boland, J.F., Wentzensen, N., Nelson, C.W., Raine-Bennett, T., Chen, Z., Bass, S., Song, L., Yang,

Q., Steinberg, M., Burdett, L., Dean, M., Roberson, D., Mitchell, J., Lorey, T., Franceschi, S., Castle, P.E., Walker, J., Zuna, R., Kreimer, A.R., Beachler, D.C., Hildesheim, A., Gonzalez, P., Porras, C., Burk, R.D., Schiffman, M., 2017. HPV16 E7 genetic conservation is critical to carcinogenesis. Cell 170, 1164–1174 (e1166).

Munger, K., Howley, P.M., 2002. Human papillomavirus immortalization and transformation functions. Virus Res. 89, 213–228.

Ortiz-Ortiz, J., Alarcon-Romero Ldel, C., Jimenez-Lopez, M.A., Garzon-Barrientos, V.H., Calleja-Macias, I., Barrera-Saldana, H.A., Leyva-Vazquez, M.A., Illades-Aguiar, B., 2015. Association of human papillomavirus 16 E6 variants with cervical carcinoma and precursor lesions in women from southern Mexico. Virol. J. 12, 29.

Park, J.S., Shin, S., Kim, E.C., Kim, J.E., Kim, Y.B., Oh, S., Roh, E.Y., Yoon, J.H., 2016. Association of human papillomavirus type 16 and its genetic variants with cervical lesion in Korea. APMIS 124, 950–957.

Petry, K.U., 2014. HPV and cervical cancer. Scand. J. Clin. Lab. Invest. 244, 59–62 Supplementum. (discussion 62).

Radhakrishna Pillai, M., Sreevidya, S., Pollock, B.H., Jayaprakash, P.G., Herman, B., 2002. Human papillomavirus type 16 E6 and E7 gene variations in Indian cervical cancer. Gynecol. Oncol. 87, 268–273.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., Howley, P.M., 1993. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 75, 495–505.

Schiffman, M., Castle, P.E., Jeronimo, J., Rodriguez, A.C., Wacholder, S., 2007. Human papillomavirus and cervical cancer. Lancet 370, 890–907.

Schiffman, M., Rodriguez, A.C., Chen, Z., Wacholder, S., Herrero, R., Hildesheim, A., Desalle, R., Befano, B., Yu, K., Safaeian, M., Sherman, M.E., Morales, J., Guillen, D., Alfaro, M., Hutchinson, M., Solomon, D., Castle, P.E., Burk, R.D., 2010. A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia. Cancer Res. 70, 3159–3169.

Shang, Q., Wang, Y., Fang, Y., Wei, L., Chen, S., Sun, Y., Li, B., Zhang, F., Gu, H., 2011. Human papillomavirus type 16 variant analysis of E6, E7, and L1 [corrected] genes and long control region in [corrected] cervical carcinomas in patients in Northeast China. J. Clin. Microbiol. 49, 2656–2663.

Sichero, L., Ferreira, S., Trottier, H., Duarte-Franco, E., Ferenczy, A., Franco, E.L., Villa, L.L., 2007. High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. Int. J. Cancer 120, 1763–1768.

Smith, B., Chen, Z., Reimers, L., van Doorslaer, K., Schiffman, M., Desalle, R., Herrero, R., Yu, K., Wacholder, S., Wang, T., Burk, R.D., 2011. Sequence imputation of HPV16 genomes for genetic association studies. PLoS ONE 6, e21375.

Song, Y.S., Kee, S.H., Kim, J.W., Park, N.H., Kang, S.B., Chang, W.H., Lee, H.P., 1997. Major sequence variants in E7 gene of human papillomavirus type 16 from cervical cancerous and noncancerous lesions of Korean women. Gynecol. Oncol. 66, 275–281.

Sun, Z., Lu, Z., Liu, J., Wang, G., Zhou, W., Yang, L., Liu, C., Wang, B., Ruan, Q., 2013. Genetic variations of E6 and long control region of human papillomavirus type 16 from patients with cervical lesion in Liaoning, China. BMC Cancer 13, 459.

Villa, L.L., Sichero, L., Rahal, P., Caballero, O., Ferenczy, A., Rohan, T., Franco, E.L., 2000. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J. Gen. Virol. 81, 2959–2968.

von Knebel Doeberitz, M., Oltersdorf, T., Schwarz, E., Gissmann, L., 1988. Correlation of modified human papilloma virus early gene expression with altered growth properties in C4-1 cervical carcinoma cells. Cancer Res. 48, 3780–3786.

von Knebel Doeberitz, M., Bauknecht, T., Bartsch, D., zur Hausen, H., 1991. Influence of chromosomal integration on glucocorticoid-regulated transcription of growth-stimulating papillomavirus genes E6 and E7 in cervical carcinoma cells. Proc. Natl. Acad. Sci. U. S. A. 88, 1411–1415.

Wakae, K., Aoyama, S., Wang, Z., Kitamura, K., Liu, G., Monjurul, A.M., Koura, M., Imayasu, M., Sakamoto, N., Nakamura, M., Kyo, S., Kondo, S., Fujiwara, H., Yoshizaki, T., Kukimoto, I., Yamaguchi, K., Shigenobu, S., Nishiyama, T., Muramatsu, M., 2015. Detection of hypermutated human papillomavirus type 16 genome by next-generation sequencing. Virology 485, 460–466.

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J., Munoz, N., 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J. Pathol. 189, 12–19.

Warren, C.J., Westrich, J.A., Doorslaer, K.V., Pyeon, D., 2017. Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. Viruses 9.

Wu, Y., Chen, Y., Li, L., Yu, G., He, Y., Zhang, Y., 2006. Analysis of mutations in the E6/E7 oncogenes and L1 gene of human papillomavirus 16 cervical cancer isolates from China. J. Gen. Virol. 87, 1181–1188.

Xi, L.F., Koutsky, L.A., Galloway, D.A., Kuypers, J., Hughes, J.P., Wheeler, C.M., Holmes, K.K., Kiviat, N.B., 1997. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. J. Natl. Cancer Inst. 89, 796–802.

Xi, L.F., Koutsky, L.A., Hildesheim, A., Galloway, D.A., Wheeler, C.M., Winer, R.L., Ho, J., Kiviat, N.B., 2007. Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18. Cancer Epidemiol. Biomark. Prev. 16, 4–10.

Yamada, T., Manos, M.M., Peto, J., Greer, C.E., Munoz, N., Bosch, F.X., Wheeler, C.M., 1997. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J. Virol. 71, 2463–2472.

Yao, Y., Huang, W., Yang, X., Sun, W., Liu, X., Cun, W., Ma, Y., 2013. HPV-16 E6 and E7 protein T cell epitopes prediction analysis based on distributions of HLA-A loci across populations: an in silico approach. Vaccine 31, 2289–2294.

Zehbe, I., Voglino, G., Delius, H., Wilander, E., Tommasino, M., 1998. Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms. Lancet 352, 1441–1442.

zur Hausen, H., 2002. Papillomaviruses and cancer: from basic studies to clinical application. Nat. Rev. Cancer 2, 342–350.