

RESEARCH

Open Access



# The polymorphisms of LCR, E6, and E7 of HPV-58 isolates in Yunnan, Southwest China

Juemin Xi<sup>1,2,3</sup>, Junying Chen<sup>1,2,3</sup>, Miaoling Xu<sup>4</sup>, Hongying Yang<sup>5</sup>, Songjiao Wen<sup>1,2,3</sup>, Yue Pan<sup>1,2,3</sup>, Xiaodan Wang<sup>1,2,3</sup>, Chao Ye<sup>1,2,3,6</sup>, Lijuan Qiu<sup>1,2,3,7</sup> and Qiangming Sun<sup>1,2,3\*</sup>

## Abstract

**Background:** Variations in HPV LCR/E6/E7 have been shown to be associated with the viral persistence and cervical cancer development. So far, there are few reports about the polymorphisms of the HPV-58 LCR/E6/E7 sequences in Southwest China. This study aims to characterize the gene polymorphisms of the HPV-58 LCR/E6/E7 sequences in women of Southwest China, and assess the effects of variations on the immune recognition of viral E6 and E7 antigens.

**Methods:** Twelve LCR/E6/E7 of the HPV-58 isolates were amplified and sequenced. A neighbor-joining phylogenetic tree was constructed by MEGA 7.0, followed by the secondary structure prediction of the related proteins using PSIPRED v3.3. The selection pressure acting on the HPV-58 E6 and E7 coding regions was estimated by Bayes empirical Bayes analysis of PAML 4.8. Meanwhile, the MHC class-I and II binding peptides were predicted by the ProPred-I server and ProPred server. The transcription factor binding sites in the HPV-58 LCR were analyzed using the JASPAR database.

**Results:** Twenty nine SNPs (20 in the LCR, 3 in the E6, 6 in the E7) were identified at 27 nucleotide sites across the HPV-58 LCR/E6/E7. From the most variable to the least variable, the nucleotide variations were LCR > E7 > E6. The combinations of all the SNPs resulted in 11 unique sequences, which were clustered into the A lineage (7 belong to A1, 2 belong to A2, and 2 belong to A3). An insertion (TGTCAGTTTCCT) was found between the nucleotide sites 7280 and 7281 in 2 variants, and a deletion (TTTAT) was found between 7429 and 7433 in 1 variant. The most common non-synonymous substitution V77A in the E7 was observed in the sequences encoding the  $\alpha$ -helix. 63G in the E7 was determined to be the only one positively selected site in the HPV-58 E6/E7 sequences. Six non-synonymous amino acid substitutions (including S71F and K93 N in the E6, and T20I, G41R, G63S/D, and V77A in the E7) were affecting multiple putative epitopes for both CD4<sup>+</sup> and CD8<sup>+</sup> T-cells. In the LCR, C7265G and C7266T were the most variable sites and were the potential binding sites for the transcription factor SOX10.

**Conclusion:** These results provide an insight into the intrinsic geographical relatedness and biological differences of the HPV-58 variants, and contribute to further research on the HPV-58 epidemiology, carcinogenesis, and therapeutic vaccine development.

**Keywords:** HPV-58, LCR, E6, E7, Polymorphism, Phylogenetic analysis, Selection pressure, MHC binding peptides, Transcription factor binding site

\* Correspondence: [sunqiangming1972@yahoo.com](mailto:sunqiangming1972@yahoo.com)

<sup>1</sup>Institute of Medical Biology, Chinese Academy of Medical Sciences, and Peking Union Medical College, Kunming 650118, People's Republic of China

<sup>2</sup>Yunnan Key Laboratory of Vaccine Research and Development on Severe Infectious Diseases, Kunming 650118, People's Republic of China

Full list of author information is available at the end of the article



## Background

Human papillomavirus type 58 (HPV-58) is one of the high-risk members of the papillomavirus family, which are considered to be associated with cervical cancer [1]. Its prevalence is particularly high in East Asia [2]. Based on the viral genomes, HPV-58 variants are classified into 4 variant lineages (A, B, C and D) and 7 sublineages (A1 A2, A3, B1, B2, D1 and D2) [3, 4]. The genome of HPV-58 is a circular double-stranded DNA with 7.9 kb in length, and is comprised of six early open-reading frames (ORFs) (E1, E2, E4, E5, E6, and E7), two late ORFs (L2 and L1), a small non-coding region between E5 and L2, and the long control region (LCR) [5].

The high-risk HPV oncoproteins E6 and E7 can disturb the transcriptional activity of the tumor suppressor protein p53 and the retinoblastoma tumor suppressor gene product pRB, respectively, and induce their proteolytic degradation. E7 can also interact with p53. They are thought to perturb the cell cycle to provide a cellular milieu to favor the replication of the HPV genome DNA and the production of viruses [6]. These effects induce the malignant transformation and immortalization of the infected cells. The HPV LCR is about 850 bp and contains a variety of the regulatory sites for the viral factors and cellular transcriptional factors, like E2 [7, 8], YY1 [9], and sex determining region Y-box 2 (SOX2) [10]. In addition to being useful for the HPV classification, the variations in the LCR have been implicated in the viral persistence and cervical cancer development [11, 12]. The LCR variants have been shown to differentially regulate the replication of HPV throughout the viral life cycle [13] and the transcriptional activity of E6 and E7 [14].

In recent years, there have been many relevant reports about the E6, E7 and LCR of HPV-16. However, studies of the HPV-58 sequences originating in Southwest China are few, especially for the LCR. It is necessary to further research the nucleotide polymorphisms within the HPV-58 LCR, E6, and E7. Besides, the effects of the variations on the immune system's ability of recognition of the viral E6 and E7 antigens were little reported and should be evaluated. In this paper, the single nucleotide polymorphisms (SNPs) in the HPV-58 LCR, E6, and E7 in Southwest China, were identified and the phylogeny was described to determine the circulating lineages. To assess the possible association of polymorphisms in the HPV-58 E6, E7 and LCR with the virus infection, replication and propagation, the amino acid changes in the HPV-58 E6 and E7 which affect multiple putative epitopes for both CD4+ and CD8+ T-cells were described, and the binding sites of the transcription factors in the LCR were analyzed. Our results could provide basic data for further studies on the HPV-58 epidemiology, prevention, and therapeutic vaccine development.

## Methods

### Ethical statement

Each participant was informed of the study aims for obtaining informed consent. This study was approved by the Institutional Ethics Committee of Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College (No. 2009ZC187M), and was carried out in line with the Helsinki Declaration.

### Cervical samples

Two thousand one hundred forty six cervical scrapings of gynecological outpatients aging from 22 to 59 were collected from the First Affiliated Hospital of Kunming Medical University in Southwest China in 2015. The routine cytology and HC2 testing were performed. Then each sample was stored in a 1.5 mL microtube at  $-80^{\circ}\text{C}$  for further study.

### DNA extraction and HPV typing

The DNA of cervical scraped cells was extracted according to the previous publications [15, 16]. The pellet of scraped cells was washed in 1 mL washing buffer containing 50 mM Tris-HCl (pH 8.1), 0.5% Tween 20, and 1 mM EDTA followed by digested in 100  $\mu\text{L}$  buffer containing 50 mM Tris-HCl (pH 8.1), 0.1 mg/mL proteinase K, 0.5% Tween 20 and 1 mM EDTA at  $55^{\circ}\text{C}$  overnight. After the digested pellet was boiled at  $95^{\circ}\text{C}$  for 10 min, each supernatant was collected by centrifugation at 13,000 rpm for 5 min, each supernatant was transferred into a 1.5 mL microtube and stored at  $-80^{\circ}\text{C}$  for amplification.

The nested PCR assay was performed for HPV typing. Firstly, the primer pair MY09 (CGTCCAAAAGGAAA CTGAGC) /MY11 (GCACAGGGACATAACAATGG) was used to amplify the extracted DNA. The 20  $\mu\text{L}$  PCR reaction volume contained 1  $\mu\text{L}$  of the extracted DNA, 10  $\mu\text{L}$  2 $\times$  Power Taq PCR MasterMix (BioTeke), 2  $\mu\text{M}$  MY09 primer and 2  $\mu\text{M}$  MY11 primer. The reaction condition was  $94^{\circ}\text{C}$  for 5 min, followed by 38 cycles of  $94^{\circ}\text{C}$  for 1 min,  $55^{\circ}\text{C}$  for 1 min,  $72^{\circ}\text{C}$  for 1 min, and then  $72^{\circ}\text{C}$  for 10 min. Secondly, the primer pair GP5+ (TTTGTTACTGTGGTAGATACTAC) /GP6+ (GAAAA ATAAACTGTAAATCATATTC) was used as previously described [17, 18]. 5  $\mu\text{L}$  of the first-round products, 25  $\mu\text{L}$  2 $\times$  Power Taq PCR MasterMix (BioTeke), 0.8  $\mu\text{M}$  GP5+ primer and 0.8  $\mu\text{M}$  GP6+ primer were contained in the 50  $\mu\text{L}$  reaction volume. The following thermocycling was used:  $95^{\circ}\text{C}$  for 5 min, followed by 38 cycles of  $95^{\circ}\text{C}$  for 1 min,  $50^{\circ}\text{C}$  for 1 min,  $72^{\circ}\text{C}$  for 50 s, and then  $72^{\circ}\text{C}$  for 10 min.  $\beta$ -globin (forward primer: ACA-CAACTGTGTTCACTAGC, reverse primer: CAACTT-CATCCACGTTTACC) was served to assess the sample quality [19], and the PCR reaction without the template DNA was used as a negative control. After the second-round PCR, the PCR products were examined on the 1%

agarose gels. The DNA fragments with 150 bp in length were then sequenced bidirectionally by TSINGKE, China and aligned on NCBI for the HPV typing.

#### HPV-58 LCR, E6, and E7 sequencing

According to the HPV-58 prototype D90400.1, two partially overlapping fragments of the HPV-58 LCR were amplified by using two pairs of primers to obtain the entire LCR. Primers used for the amplification of the complete nucleotide sequence of the HPV-58 E6 and E7 were described previously [16]. All the primers (Additional file 1: Table S1) were synthesized by TSINGKE, China.

The E6, E7 and LCR fragments were amplified in 50  $\mu$ L PCR reaction volumes containing 2  $\mu$ L of extracted DNA, 25  $\mu$ L 2 $\times$  Power Taq PCR MasterMix (BioTeke), 0.4  $\mu$ M forward primer and 0.4  $\mu$ M reverse primer. The PCR condition was: 95  $^{\circ}$ C for 5 min, followed by 40 cycles of 95  $^{\circ}$ C for 30 s, 55  $^{\circ}$ C for 45 s, 72  $^{\circ}$ C for 50 s, and then 72  $^{\circ}$ C for 10 min. The positive fragments were subjected to the bi-directional DNA sequencing (TSINGKE, China) for further analysis. The sequences obtained in this study were submitted to the NCBI GenBank database and got the accession number: MF741321 to MF741331.

#### Variant identification and analysis

To identify the SNPs in the LCR, E6 and E7, the studied HPV-58 LCR/E6/E7 sequences were aligned to the HPV-58 prototype sequence (GenBank: D90400.1) by MEGA (version 7.0) [20] after sequencing.

The secondary structures of the HPV-58 E6 and E7 proteins were predicted by PSIPRED v3.3 (<http://bioinf.cs.ucl.ac.uk/psipred/>) using the default parameters [21].

The neighbor-joining phylogenetic tree based on the HPV-58 LCR/E6/E7 was constructed by MEGA software (version 7.0) using the Kimura 2-parameter model. The number of bootstrap replications was set at 1000. To construct the phylogenetic branches, the following reference sequences were used: D90400.1 (A1), HQ537752 (A2), HQ537758 (A3), HQ537762 (B1), HQ537764 (B2), HQ537774 (C), HQ537768 (D1) and HQ537770 (D2) [3].

To estimate the selection pressure acting on the HPV-58 E6 and E7 coding regions, the non-synonymous and synonymous nucleotide divergence were calculated by CODEML in PAML 4.8. Using the Bayes empirical Bayes (BEB) analysis, the sites with the posterior probability > 95% were identified as the positively selected sites [22–25].

#### MHC binding peptides prediction

The ProPred-I server (<http://www.imtech.res.in/raghava/propred1/>) and ProPred server (<http://www.imtech.res.in/raghava/propred/>) were used to predict the major histocompatibility complex (MHC) class-I binding

peptides [26] and MHC class-II binding peptides [27], respectively.

#### Transcription factor binding sites prediction

To analyze the transcription factor binding sites in the HPV-58 LCR, the JASPAR database (<http://jaspar.genereg.net/>) [28] was used. The sites for YY1, SRY, CEBPA, SOX9, SOX10, POU2F2, JUN, FOS, FOXP3, ETS1, SPIB, ESR2, HSF1, CREB1, PHOX2A, NFIA, ELK4, SMAD3, IRF2, STAT1, HOXC11, NFKB1, RAX, and VAX1 were included. The relative profile score threshold was set at 85%.

## Results

#### HPV-58 prevalence among the collected gynecological outpatient samples

Among the 2146 collected outpatient samples in Southwest China, 248 samples (11.56%, 248/2146) were HPV positive [15]. Following HPV-16 (18.95%, 47/248), HPV-58 was identified as the second most common high-risk type, accounting for 13.71% (34/248) in this population.

#### Genomic polymorphisms of HPV-58 LCR/E6/E7

12 entire LCR/E6/E7 sequences were obtained and further analyzed. The small copy number of HPV in some women might be a possible explanation for not all sequences successfully amplified. When aligned to the HPV-58 reference sequences (GenBank: D90400.1), a total of 29 single SNPs were identified in 27 nucleotide sites across the LCR, E6 and E7. The combinations of all these SNPs made for 11 unique LCR/E6/E7 sequences (variant IDs: 1 to 11), of which the variant 2 was represented in 2 samples (Table 1). The maximum pairwise difference of the LCR/E6/E7 sequence between any two variants was approximately 0.8%.

In the LCR, 20 SNPs were detected at 18 nucleotide sites (18/794, 2.3% nucleotide variation). C7265G, C7266T and A7793G were the most variable sites and were identified in 27% of the variants (3/11). In addition, an insertion (TGTCAGTTTCCT) was found between the nucleotide sites 7280 and 7281 in 2 variants (variant IDs: 9, 10), and a deletion (TTTAT) was found between the nucleotide sites 7429 and 7433 in 1 variant (variant ID: 2). In the E6, 3 SNPs were detected (3/450, 0.7% nucleotide variation). C307T and A388C were identified in 36% variants (4/11). C321T and A388C resulted in the amino acid changes including S71F and K93 N, respectively. No non-synonymous variations were observed in the E6 sequences encoding the  $\alpha$ -helix or the  $\beta$ -sheet. In the E7, 6 SNPs were detected (6/297, 2.0% nucleotide variation), of which T744G was identified in 64% variants (7/11), and 5 SNPs resulted in amino acid changes including T20I, G41R, G63S, G63D and V77A. The non-synonymous variation V77A was

**Table 1** HPV-58 variants based on the LCR/E6/E7 sequences

GenBank Accession	Variant ID	HPV-58 variable sites																												Lineage
		LCR														E6										E7				
		7172	7194	7265	7266	7304	7440	7525	7540	7556	7714	7749	7755	7784	7793	7807	24	52	54	307	321	388	632	694	744	760	761	803		
D90400.1		T	G	C	C	A	T	A	A	C	A	C	A	T	A	T	T	C	G	C	C	A	C	G	T	G	G	T	A1	
MF741321	1	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	A1	
MF741322	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A1	
MF741323	3	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	C	-	-	G	-	-	C	A1	
MF741324	4	-	-	-	-	-	-	G	-	-	G	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A1	
MF741325	5	-	-	G	T	-	-	-	-	G	-	-	-	-	-	-	-	-	-	T	-	-	-	A	G	-	A	A2		
MF741326	6	-	-	G	T	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	T	C	-	-	G	-	-	C	A1	
MF741327	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	A1	
MF741328	8	C	-	G	T	-	-	-	-	G	-	-	-	-	-	G	-	-	-	T	-	-	-	A	G	-	A	A2		
MF741329	9	-	C	-	-	G	-	C	-	T	C	-	G	-	C	-	-	T	-	T	-	-	T	-	G	A	-	A3		
MF741330	10	-	C	-	-	G	-	-	-	C	-	G	-	-	-	-	-	T	-	T	-	-	T	-	G	A	-	A3		
MF741331	11	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	G	-	-	-	-	C	-	-	G	-	-	C	A1	
AA changes																														
Secondary structure																														

The nucleotides matching the reference sequence (GenBank: D90400.1) are marked with a dash (-). The H and S in the row of secondary structure designate "helix" and "sheet", respectively

detected in the E7 sequences encoding the  $\alpha$ -helix. There was no insertion or deletion detected in either E6 or E7 sequences.

**Phylogenetic analysis**

A neighbor-joining phylogenetic tree based on the HPV-58 LCR/E6/E7 was inferred from 11 obtained unique variants and 8 reference sequences. As shown in Table 1, all the 11 unique variants representing 12 samples, were clustered into the A lineage, including seven A1 (variant IDs: 1, 2, 3, 4, 6, 7, 11, representing 8 samples), two A2 (variant IDs: 5, 8) and two A3 (variant IDs: 9, 10) sublineages with the proportions of 64%, 18%, and 18%, respectively.

**Selective pressure analysis**

The selective pressure analysis results by the Bayes empirical Bayes (BEB) analysis of PAML 4.8 were summarized in Table 2. In the HPV-58 E6 coding sequences, no positive selection was observed. However, in the HPV-58 E7 coding sequences, 63G was determined to be a positively selected site.

**Amino acid changes in the MHC binding peptides of the HPV-58 E6 and E7 proteins**

To evaluate the impacts of the HPV-58 E6 and E7 sequence polymorphisms on the immune recognition of these antigens, the binding peptides specific for both MHC class-I and class-II were predicted by the ProPred-I and ProPred servers, respectively. The results showed that the non-synonymous substitutions within the HPV-58 E6 and E7 were potentially affecting the multiple epitopes. The results were summarized in Table 3. In the HPV-58 E6, a C to T substitution at the nucleotide site 321 led to the change of residue 71 (S to F) in one isolate, where the MHC class-I binding peptide KVCLRLLSK (E6 64–72) and the MHC class-II binding peptide LRLLSKISE (E6 67–75) were located. The A to C substitution at the nucleotide site 388 resulted in the change of residue 93 (K to N), where the MHC class-I binding peptide TLKKCLNEI

**Table 3** The putative MHC binding peptides of HPV-58 E6 and E7 proteins

Protein	Position of amino acid	Peptide sequence	MHC binding
E6	64–72	KVCLRLLSK	Class-I
	91–99	TLKKCLNEI	Class-I
	67–75	LRLLSKISE	Class-II
E7	16–24	HPEPTDLFC	Class-I
	35–43	DEIGLDGPD	Class-I
	59–67	CYTCGTTVR	Class-I
	74–82	TTDVRTLQQ	Class-I
	15–22	LHPEPTDL	Class-II
	37–45	IGLDGPDGQ	Class-II
	77–85	VRTLQQLLM	Class-II

(E6 91–99) was located. In the HPV-58 E7, the non-synonymous substitution T20I was affecting the MHC class-I binding peptide HPEPTDLFC (E7 16–24) and the MHC class-II binding peptide LHPEPTDL (E7 15–22). G41R was affecting the MHC class-I binding peptide DEIGLDGPD (E7 35–43) and the MHC class-II binding peptide IGLDGPDGQ (E7 37–45). G63S/D was affecting the MHC class-I binding peptide CYTCGTTVR (E7 59–67). V77A was affecting the MHC class-I binding peptide TTDVRTLQQ (E7 74–82) and the MHC class-II binding peptide VRTLQQLLM (E7 77–85).

**Analysis of the transcription factor binding sites**

The potential binding sites for the cellular transcription factors in the HPV-58 LCR variants were investigated by the online tool JASPAR. The results indicated the affected transcription factor binding sites due to some of the nucleotide variations spanning nucleotides 7140 to 109. The variations at the nucleotide sites 7265 and 7266 were suggested to be the binding sites for the transcription factor SOX10, and 7304 and 7714 for SOX9. In addition, the nucleotide sites 7172, 7540, 7556, 7749, 7755, 7793, 7807, and 24 potentially affected the binding sites for the transcription factors FOXA1, SRY, ELK4, FOS, FOXP3, ESR2, CEBPA, and HOXC11, respectively.

**Table 2** Tests for the positive selected sites in HPV-58 E6 and E7

Regions	Models	lnL	2ΔlnL	Positively selected sites
E6	M1	-675.08	2.18	NA
	M2	-673.99		NS
	M7	-675.13	2.28	NA
	M8	-673.99		NS
E7	M1	-499.87	18.92	NA
	M2	-490.41		63G**
	M7	-499.84	18.86	NA
	M8	-490.41		63G**

lnL designates the log-likelihood; 2ΔlnL designates 2-fold as the lnL difference between the two models. Two asterisks (\*\*) indicate the posterior probability ≥99% by BEB analysis and to be a positively selected site. NA designates not allowed, and NS designates the sites not reaching the significant level

**Discussion**

A specific geographic and population research on the characterization of each HPV type is important to learn the HPV prevalence, find out the biological evidences of carcinogenesis, and help improving the strategies of vaccine development. So far, the HPV family is comprised of more than 170 different types [29]. They can infect the genital tract, the upper-respiratory tract, or the skin. The high-risk HPV types are considered carcinogenic, such as the types 16, 18, 33 and 58 [1]. In our study, HPV-58, accounting for 13.71% of all the HPV-positive samples, was the second most prevalent high-risk types after HPV-16 in Yunnan,

China. The nucleotide substitution C321T in the HPV-58 E6, leading to the amino acid change (S71F) was the novel variation detected in this study when compared with those from 2010 to 2011 [16].

In the HPV-58 genome, each region differs in the sequence diversity. The HPV-58 LCR variants from Yunnan, Southwest China have not been reported yet. Here, the variation of the HPV-58 LCR was showed in a higher ratio than those of the E6 and E7 in Yunnan, Southwest China. From the most variable to the least variable, the nucleotide variations were LCR > E7 > E6 found in 2.3%, 2.0% and 0.7%, respectively, of the studied variants. Additionally, the most common substitutions in the HPV-58 LCR were C7265G, C7266T and A7793G. And the most common non-synonymous substitutions in the HPV-58 E6 and E7 were A388C (K35 N) and T803C (V77A), respectively. V77A in the E7 was observed in the sequences encoding the  $\alpha$ -helix. Whether this substitution changed the secondary structure of the E7 protein and then affected the potential ability of the E7 to cause cervical cancer needs to be further studied. Besides, 63G in the HPV-58 E7 coding sequences was determined to be a positively selected site, indicating an adaptive molecular sequence evolution.

The HPV variants co-evolve with the specific populations and their prevalence varies across the geographic regions. In this study, phylogenetic analysis based on the HPV-58 LCR/E6/E7 revealed a high prevalence of the A lineage, among which the A1 sublineage with a proportion of 64% accounted for the majority of HPV-58 variants obtained. No variant belonged to the lineages B, C, or D. Nevertheless, the B lineage was found in Sichuan [30] and Henan, China [4]; D lineage was found in Hong Kong [2] and Liaoning [31], China, suggesting that the HPV-58 variant lineages are differently distributed around China.

Though phylogenetically related, the HPV-58 variants can differ in pathogenicity. The occurrence of C632T and G760A in the HPV-58 E7, which were identified in our study and resulted in the amino acid changes including T20I and G63S, respectively, has been implied to be linked to an increased risk for cervical cancer [32, 33]. The amino acid changes in the HPV-58 E6 and E7 oncoproteins may influence the MHC binding peptides and then have the significant immunological effects on the immune system's ability of recognition of these viral antigens. The non-synonymous substitution T20I potentially affected the MHC class-I and class-II binding peptides, probably leading to the loss of the CD8+ cytotoxic T cell or CD4+ helper T cell responses, respectively. In addition, G63S/D (a positively selected site) affected the MHC class-I binding peptide, and V77A, the most common non-synonymous substitution in the HPV-58 E7, made the predicted MHC class-II binding peptide VRTLQQLM (E7, 77–85) disappear. The HPV E6 and E7 act as the promising specific tumor antigens and are available as the therapeutic

targets. Adoptive transfer of dendritic cells with the HPV E6 and E7 antigens is able to induce the CD8+ T-cell mediated immune responses [34, 35]. Therefore, it is essential to perform further functional studies concerning the sequence polymorphisms of the HPV-58 E6 and E7 to find out the biological proof for tumorigenicity and to optimize the strategies for the HPV prevention and therapy.

The HPV LCR has been shown to contain kinds of the binding sites for both viral and cellular factors and regulate the replication of HPV and the transcriptional activity of the E6 and E7 through the interactions between the regulatory protein complex and the viral DNA [10, 13, 36]. Certain nucleotide changes in the LCR may influence the tumorigenic process. For example, the variations within the binding sites for the cellular factor YY1 in the LCR were found to increase the viral oncogene expression [9]. In the present study, the variations at the nucleotide sites 7265 and 7266, which were detected as the most variable sites in HPV-58 LCR, were the potential binding sites for SOX10, a transcription factor of the sex determining region Y (SRY)-related high motility group (HMG)-box gene family [37]. Sox10 has been suggested to function as a tumor suppressor in digestive cancers [38] and in prostatic carcinoma [39] since it was silenced or down-regulated in these tumors. A previous study on the HPV-16 LCR showed that SOX2 from the same SRY-HMG-box gene family acted as a transcriptional repressor, down-regulating the expression of the HPV E6 and E7 oncoproteins in a context of squamous cell carcinomas [10].

## Conclusions

This study provides a useful data for understanding the phylogenetic relatedness of the HPV-58 variants, the possible influence of non-synonymous substitutions in the HPV-58 E6/E7 on the T-cell responses, and the impacts of the HPV-58 LCR variations on the bindings of the cellular transcription factors. Further studies of the HPV-58 variant genomes from a large population of the asymptomatic carriers and cervical cancer patients are required to establish the carcinogenesis of HPV-58 and to make improvements to the immunotherapeutic approaches and vaccine development strategies.

## Additional file

**Additional file 1: Table S1.** The PCR primers for validating the variations of HPV-58 LCR, E6 and E7. (DOC 33 kb)

## Abbreviations

HPV-58: Human papillomavirus type 58; LCR: long control region; SNPs: single nucleotide polymorphisms; SOX: sex determining region Y-box; MHC: major histocompatibility complex; the amino acid acronyms A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V and Y: alanine, cysteine, aspartic acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparagine, proline, glutamine, arginine, serine, threonine, valine and tyrosine

### Acknowledgements

We thanked Dr. Penghua Wang in the Department of Microbiology & Immunology, School of Medicine, New York Medical College for the reviewing and proofreading the manuscript.

This work was supported by CAMS Innovation Fund for Medical Sciences (CIFMS) (grant number 2016-I2M-3-026), the National "Twelfth Five Year" "major new drug discovery" technology major projects (grant number 2012ZX09104-302), the National Natural Science Foundation of China (grant number 81171946), the Natural Science Foundation of Yunnan Province (grant number 2016FA029), and the Jointly Supported Foundation of the National Project in Yunnan Province (grant number 2013GA018).

### Funding

This study was funded by CAMS Innovation Fund for Medical Sciences (CIFMS) (grant number 2016-I2M-3-026), the National "Twelfth Five Year" "major new drug discovery" technology major projects (grant number 2012ZX09104-302), the National Natural Science Foundation of China (grant number 81171946), the Natural Science Foundation of Yunnan Province (grant number 2016FA029), and the Jointly Supported Foundation of the National Project in Yunnan Province (2013GA018). The funding bodies have no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

The DNA sequences obtained in this study were available in NCBI GenBank database with the accession number: MF741321 to MF741331.

### Authors' contributions

Conception of the study and supervising the progress of study: QS. Draft of manuscript: JX. Performing the majority parts of experiment: JX and JC. Sample collection: MX and HY. Performing the parts experiments: SW, YP, and XW. Data analysis: JX, SW, CY and LQ. Review of the manuscript: All the authors. All authors read and approved the final manuscript.

### 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 Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in the study.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Institute of Medical Biology, Chinese Academy of Medical Sciences, and Peking Union Medical College, Kunming 650118, People's Republic of China. <sup>2</sup>Yunnan Key Laboratory of Vaccine Research and Development on Severe Infectious Diseases, Kunming 650118, People's Republic of China. <sup>3</sup>Yunnan Provincial Key Laboratory of Vector-borne Diseases Control and Research, Pu'er 665000, People's Republic of China. <sup>4</sup>The First Affiliated Hospital of Kunming Medical University, Kunming 650032, People's Republic of China. <sup>5</sup>The Third Affiliated Hospital of Kunming Medical University, Yunnan Provincial Tumor Hospital, Kunming 650118, People's Republic of China. <sup>6</sup>Kunming Medical University, Kunming 650500, People's Republic of China. <sup>7</sup>Institute of Pediatric Disease Research in Yunnan, the Affiliated Children's Hospital of Kunming Medical University, Kunming 650228, People's Republic of China.

Received: 17 January 2018 Accepted: 16 April 2018

Published online: 25 April 2018

### References

- Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003;348:518–27.
- Chan PK, Zhang C, Park JS, Smith-McCune KK, Palefsky JM, Giovannelli L, et al. Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequence variations. *Int J Cancer*. 2013;132:2528–36.
- Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. *Virology*. 2013;445:232–43.
- Liu Y, Pan Y, Gao W, Ke Y, Lu Z. Whole-genome analysis of human papillomavirus types 16, 18, and 58 isolated from cervical Precancer and Cancer samples in Chinese women. *Sci Rep*. 2017;7:263.
- Ho L, Chan SY, Burk RD, Das BC, Fujinaga K, Icenogle JP, et al. 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*. 1993;67:6413–23.
- Münger K, Howley PM. Human papillomavirus immortalization and transformation functions. *Virus Res*. 2002;89:213–28.
- Romanczuk H, Thierry F, Howley PM. Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J Virol*. 1990;64:2849–59.
- Soeda E, Ferran MC, Baker CC, McBride AA. Repression of HPV16 early region transcription by the E2 protein. *Virology*. 2006;351:29–41.
- Dong XP, Stubenrauch F, Beyer-Finkler E, Pfister H. Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *Int J Cancer*. 1994;58:803–8.
- Martínez-Ramírez I, Del-Castillo-Falconi V, Mitre-Aguilar IB, Amador-Molina A, Carrillo-García A, Langley E, et al. SOX2 as a new regulator of HPV16 transcription. *Viruses*. 2017;9. pii: E175. <https://doi.org/10.3390/v9070175>.
- Pientong C, Wongwarissara P, Ekaklaksananan T, Swangphon P, Kleeboak P, Kongyingyoes B, et al. Association of human papillomavirus type 16 long control region mutation and cervical cancer. *Virol J*. 2013;10:30.
- Mosmann JP, Monetti MS, Frutos MC, Kiguen AX, Venezuela RF, Cuffini CG. Mutation detection of E6 and LCR genes from HPV 16 associated with carcinogenesis. *Asian Pac J Cancer Prev*. 2015;16(3):1151–7.
- Hubert WG. Variant upstream regulatory region sequences differentially regulate human papillomavirus type 16 DNA replication throughout the viral life cycle. *J Virol*. 2005;79:5914–22.
- Pande S, Jain N, Prusty BK, Bhambhani S, Gupta S, Sharma R, et al. Human papillomavirus type 16 variant analysis of E6, E7, and L1 genes and long control region in biopsy samples from cervical Cancer patients in North India. *J Clin Microbiol*. 2008;46:1060–6.
- Xi J, Chen J, Xu M, Yang H, Luo J, Pan Y, et al. Genetic variability and functional implication of the long control region in HPV-16 variants in Southwest China. *PLoS One*. 2017;12:e0182388.
- Yang L, Yang H, Chen J, Huang X, Pan Y, Li D, et al. Genetic variability of HPV-58 E6 and E7 genes in Southwest China. *Infect Genet Evol*. 2014;21:395–400.
- de Roda Husman AM, Walboomers JM, van den Brule AJ, Meijer CJ, Snijders PJ. The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. *J Gen Virol*. 1995;76:1057–62.
- Qu W, Jiang G, Cruz Y, Chang CJ, Ho GY, Klein RS, et al. PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J Clin Microbiol*. 1997;35:1304–10.
- Ayatollahi M, Zakerinia M, Haghshenas M. Molecular analysis of Iranian families with sickle cell disease. *J Trop Pediatr*. 2005;51:136–40.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
- Buchan DW, Minnici F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res*. 2013;41:W349–57.
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
- Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 1997;13:555–6.

24. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005;22:1107–18.
25. Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol.* 2013;30:2723–4.
26. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC class-I binding sites. *Bioinformatics.* 2003;19:1009–14.
27. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics.* 2001;17:1236–7.
28. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2016;44:D110–5.
29. Ghittoni R, Accardi R, Chiocca S, Tommasino M. Role of human papillomaviruses in carcinogenesis. *Ecancermedicalscience.* 2015;9:526.
30. Chen Z, Jing Y, Wen Q, Ding X, Wang T, Mu X, et al. E6 and E7 gene polymorphisms in human papillomavirus Types-58 and 33 identified in Southwest China. *PLoS One.* 2017;12:e0171140.
31. Liu JH, Lu ZT, Wang GL, Zhou WQ, Liu C, Yang LX, et al. Variations of human papillomavirus type 58 E6, E7, L1 genes and long control region in strains from women with cervical lesions in Liaoning province, China. *Infect Genet Evol.* 2012;12:1466–72.
32. Chan PK, Lam CW, Cheung TH, Li WW, Lo KW, Chan MY, et al. Association of human papillomavirus type 58 variant with the risk of cervical cancer. *J Natl Cancer Inst.* 2002;94:1249–53.
33. Ding T, Wang X, Ye F, Cheng X, Ma D, Lu W, et al. Distribution of human papillomavirus 58 and 52 E6/E7 variants in cervical neoplasia in Chinese women. *Gynecol Oncol.* 2010;119:436–43.
34. Murakami M, Gurski KJ, Marincola FM, Ackland J, Steller MA. Induction of specific CD8+ T-lymphocyte responses using a human papillomavirus-16 E6/E7 fusion protein and autologous dendritic cells. *Cancer Res.* 1999;59:1184–7.
35. Yang A, Jeang J, Cheng K, Cheng T, Yang B, Wu TC. Current state in the development of candidate therapeutic HPV vaccines. *Expert Rev Vaccines.* 2016;15:989–1007.
36. Cornet I, Gheit T, Franceschi S, Vignat J, Burk RD, Sylla BS, et al. Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR. *J Virol.* 2012;86:6855–61.
37. Miettinen M, McCue PA, Sarlomo-Rikala M, Biernat W, Czapiewski P, Kopczynski J, et al. Sox10—a marker for not only schwannian and melanocytic neoplasms but also myoepithelial cell tumors of soft tissue: a systematic analysis of 5134 tumors. *Am J Surg Pathol.* 2015;39:826–35.
38. Tong X, Li L, Li X, Heng L, Zhong L, Su X, et al. SOX10, a novel HMG-box-containing tumor suppressor, inhibits growth and metastasis of digestive cancers by suppressing the Wnt/ $\beta$ -catenin pathway. *Oncotarget.* 2014;5:10571–83.
39. Zhong WD, Qin GQ, Dai QS, Han ZD, Chen SM, Ling XH, et al. SOXs in human prostate cancer: implication as progression and prognosis factors. *BMC Cancer.* 2012;12:248.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

