

The Genetic Variability, Phylogeny and Functional Significance of E6, E7 and LCR in Human Papillomavirus type 52 Isolates in Sichuan, China

Zhilin Song

College of Life Sciences, Sichuan University

Junhang Deng

College of Life Sciences, Sichuan University

Qiufu Li

College of Life Sciences, Sichuan University

Yanru Cui

College of Life Sciences, Sichuan University

Xianping Ding (✉ brainding@scu.edu.cn)

Sichuan University <https://orcid.org/0000-0003-1555-5028>

Jiaoyu He

College of Life Sciences, Sichuan University

Yiran Liu

College of Life Sciences, Sichuan University

Zhuang Ju

College of Life Sciences, Sichuan University

Liyuan Fang

College of Life Sciences, Sichuan University

Research

Keywords: Human papillomavirus, E6, E7, LCR, polymorphism, phylogeny, antigen epitopes, Transcription factor binding site

Posted Date: October 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-67643/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: The genetic variability, phylogeny and functional significance of E6, E7 and LCR in human papillomavirus type 52 isolates in Sichuan, China

Author: Zhilin Song^{1,2}, Junhang Deng^{1,2}, Qiufu Li^{1,2}, Yanru Cui^{1,2}, Xianping Ding^{1,2,*}, Jiaoyu He^{1,2}, Yiran Liu^{1,2}, Zhuang Ju^{1,2}, Liyuan Fang^{1,2}

Correspondence: brainding@scu.edu.cn

1: Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, Sichuan, People's Republic of China.

2: Bio-resource Research and Utilization Joint Key Laboratory of Sichuan and Chongqing, Chongqing, People's Republic of China.

Abstract

Background: Variations in human papillomavirus (HPV) E6 and E7 have been shown to be closely related to the persistence of the virus and the occurrence and development of cervical cancer. Long control region (LCR) of HPV has been shown multiple functions on regulating viral transcription. In recent years, there have been reports on E6/E7/LCR of HPV-16 and HPV-58, but there are few studies on HPV-52, especially for LCR. In this study, we focused on gene polymorphism of the HPV-52 E6/E7/LCR sequences, assessed the effects of variations on the immune recognition of viral E6 and E7 antigens, predicted the effect of LCR variations on transcription factor binding sites and provided more basic data for further study of E6/E7/LCR in Chengdu, China.

Methods: LCR/E6/E7 of the HPV-52 were amplified and sequenced to do polymorphic and phylogenetic analysis. Sequences were aligned with the reference sequence by MEGA 7.0 to identify SNP. A neighbor-joining phylogenetic tree was constructed by MEGA 7.0, followed by the secondary structure prediction of the related proteins using PSIPRED 4.0. The selection pressure of E6 and E7 coding regions were estimated by Bayes empirical Bayes analysis of PAML 4.9. The HLA class-I and II binding peptides were predicted by the Immune Epitope Database (IEDB) server. The B cell epitopes were predicted by ABCpred server. Transcription factor binding sites in LCR were predicted by JASPAR database.

Results: 50 SNP sites (6 in E6, 10 in E7, 34 in LCR) were found. From the most variable to the least variable, the nucleotide variations were LCR>E7> E6. Two deletions were found between the nucleotide sites 7387 to 7391 (TTATG) and 7698 to 7700 (CTT) in all samples. A deletion was found between the nucleotide sites 7287 to 7288 (TG) in 97.56% (40/41) of the samples. The combinations of all the SNP sites and deletions resulted in 12 unique sequences. As shown in the neighbor-joining phylogenetic tree, except for one belonging to sub-lineage C2, others sequences clustered into sub-lineage B2. No positive selection was observed in E6 and E7. 8 non-synonymous amino acid substitutions (including E3Q and K93R in the E6, and T37I, S52D, Y59D, H61Y, D64N and L99R in the E7) were potential affecting multiple putative epitopes for both CD4+ and CD8+ T-cells and B-cells. A7168G was the most variable site (100%) and the binding sites for transcription factor VAX1 in LCR. In addition, the prediction results showed that LCR had the high probability binding sites for transcription factors SOX9, FOS, RAX, HOXA5, VAX1 and SRY.

Conclusion: This study provides basic data for understanding the relation among E6/E7/LCR mutations, lineages and carcinogenesis. Furthermore, it provides an insight into the intrinsic geographical relatedness and biological differences of the HPV-52 variants, and contributes to further research on the HPV-52 therapeutic vaccine development.

Keywords: Human papillomavirus, E6, E7, LCR, polymorphism, phylogeny, antigen epitopes, Transcription factor binding site

Background

According to statistics, about 1.4 million women worldwide suffer from cervical cancer, which is the second most common cancer among women in the world and the main cause of cancer death in some developing countries [1-3]. Cervical cancer not only causes a serious threat to the health of women all over the world, but also causes a very high burden to social economy. Strong epidemiological and molecular evidence accumulated over the past few decades has confirmed the close connection between the persistent infection of high-risk human papillomavirus and the development of cervical cancer, almost all cervical cancer biopsies have found one or more high-risk human papillomavirus [4, 5].

Human papillomavirus (HPV) is a capsid-enclosed small circular double-stranded type of DNA virus of approximately 8kb in size [6]. It can specifically infect human epithelium and mucosa, causing a variety of diseases, and is a large group of virus, which consists of more than 250 different types [7]. According to its pathogenicity, HPV types are classified into high risk and low risk [8]. HPV-16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73 and 82 are the common high-risk types and have been proven to cause more than 96% of cervical cancers; among low-risk HPV types, HPV-6 and HPV-11 are the most common and have been proven to cause 90% of genital warts in men [9-12]. In addition, with the in-depth study of HPV, it is found that high-risk HPV is highly correlated with various malignant tumors such as anal cancer, head and neck cancer, and throat cancer [13-15].

The genome of HPV is comprised of an early region (E1, E2, E4, E5, E6 and E7), a late region (L1 and L2), a small non-coding region (NCR) and a long control region (LCR) [16, 17]. Among these genes, the early genes E6 and E7 are regarded as the main oncogenes, and their expression is essential for the transformation and maintenance of the malignant state. The E6 protein binds to the ligase E6AP, forming a complex that can inactivate the p53 tumor suppressor protein, whereas the E7 protein prevents expression of the retinoblastoma (Rb) tumor suppressor protein through ubiquitination [18, 19]. When human papillomavirus integrates into the host cell, E6 and E7 are invariably reserved and uncontrolled expressed so the cell obtains the function of immortalization and transformation [20]. The antigenic epitope of HPV E6 and E7 proteins are ideal targets for therapeutic vaccine design, so it is very important to identify the epitopes of E6 and E7 proteins. The HPV LCR contains the viral origin of replication, the viral early promoter and transcriptional enhancer, the late polyadenylation site and the late regulatory element and which has a variety of the regulatory sites for the viral factors and cellular transcriptional factors, such as E1, E2 and SOX2 [21-24]. The LCR has been shown to be the most variable region of HPV genome, because it does not encode any gene and can able to accumulate and tolerate more mutations. The LCR variants have been shown to differently regulate the replication of HPV throughout the viral life cycle and the transcriptional activity of E6 and E7 [25, 26].

There have been reports about the E6, E7 and LCR of HPV-16 and HPV-58 in recent years. However, studies of the HPV-52 are few, especially for the LCR. In addition, statistics shown that the infection rate of HPV-52 among Sichuan women has increased year by year. Therefore, it is necessary to carry out in-depth research on the nucleotide polymorphisms with the HPV-52 E6, E7 and LCR. This study identified the single nucleotide polymorphism (SNP) in the HPV-52 E6, E7 and LCR in Sichuan,

southwest China, assessed the possible association of polymorphisms in the HPV-52 E6, E7 and LCR with the virus infection, propagation and replication, predicted the high affinity antigen epitope of HPV-52 E6 and E7, and analyzed the binding sites of the transcription factors in the LCR. Our results could provide basic data for further studies on the HPV-52 epidemiology, prevention, and therapeutic vaccine development.

Methods

Ethics statement

All participants were informed of the study aims, and a written informed consent was received from each patient before sample collection, and the patients' privacy have been fully protected. This study was approved by education and research committee and Ethics Committee of Sichuan University, China (approval number SCU20100196494), and was carried out in line with the Helsinki Declaration.

Collection of clinical specimens

3432 cervical scrape cell samples were collected from outpatients who underwent routine cervical screenings at The Affiliate Reproductive Hospital of Sichuan Genitalia Hygiene Research Center, Chengdu SongZiNiao Sterility Hospital and Chengdu Medical College Affiliated Infertility Hospital from December 2018 to December 2019. The sample was stored at $-20\text{ }^{\circ}\text{C}$ in cell preservation solution and all methods were performed in accordance with the relevant guidelines and regulations.

Genomic DNA extraction and HPV typing

HPV DNA was extracted and genotyped using Advanced Fragment Analysis (AFA) based on capillary electrophoresis system with the commercial Human Papillomavirus Genotyping Kit for 25 types (HEALTH Gene technologies, Ningbo, China) according to the manufacturer's guidelines. This kit is able to classify 25 different HPV types (HPV 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, 82, 6, 11, 42, 43, 44 and 83). 46 samples that tested positive for single infection of HPV-52 were picked out for amplification, sequencing and following study.

DNA amplification and sequencing

The primer pairs were designed by Primer premier5.0 according to the HPV-52 reference sequence (GenBank: X74481.1) and the primer sequences were listed in Table 1. All the primers were synthesized by TSINGKE, China. The E6, E7 and LCR fragments were amplified in 50 μl PCR reaction volumes containing 18 μl of extracted DNA, 25 μl Taq 2X PCR Master Mix with Dye (ABclonal), 1 μl forward primer, 1 μl reverse primer and 5 μl Nuclease-free water.

The PCR amplification was performed under the following conditions: an initial 30s denaturation step at $95\text{ }^{\circ}\text{C}$, followed by 30 amplification cycles, with each cycle including a 30 s denaturation step at $95\text{ }^{\circ}\text{C}$, a 30 s annealing step at $52\text{-}55\text{ }^{\circ}\text{C}$ ($52\text{ }^{\circ}\text{C}$ for E6, $55\text{ }^{\circ}\text{C}$ for E7 and LCR), and a 35-65 s (35 s for E7, 45 s for E6 and 65 s for LCR) elongation step at $68\text{ }^{\circ}\text{C}$, and then 5 min final extension at $68\text{ }^{\circ}\text{C}$, ended up and held at $4\text{ }^{\circ}\text{C}$. The PCR products were examined under UV light after electrophoretic separation on a 2% agarose gel. The positive fragments were subjected to the bi-directional DNA sequencing (TSINGKE, China) for the further analysis.

Variant identification and analysis

To identify the SNP in E6, E7 and LCR, the HPV-52 E6/E7/LCR sequences were aligned to the HPV-52 reference sequence (GenBank: X74481.1) by MEGA 7.0 after sequencing. The secondary structures of the HPV-52 E6 and E7 proteins were predicted by PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>) using the default parameters.

The neighbor-joining phylogenetic tree based on the HPV-52 E6/E7/LCR was constructed by MEGA 7.0 using the Kimura 2-parameter model and the number of bootstrap replications was set at 1000. To

construct the phylogenetic branches, the following reference sequences were used: X74481 ((A1), HQ537739 (A2), HQ537740 (B1), HQ537743 (B2), HQ537744 (C1), HQ537746 (C2), HQ537748 (D) [27].

To estimate the selection pressure acting on the HPV-52 E6 and E7 protein coding regions, the non-synonymous and synonymous nucleotide divergence were calculated by CODEML in PAML 4.9. Using the Bayes empirical Bayes (BEB) analysis, the sites with the posterior probability > 95% were identified as the positively selected sites [23].

T and B cell antigen epitope prediction

The Immune Epitope Database (IEDB) server (<http://www.iedb.org/>) was used to predict the major HLA-I and HLA-II binding peptides [28, 29]. In our study, we used IEDB recommended methods to predict the epitopes against 34 HLA-I alleles by comprehensive analysis the frequency of Chinese in allele frequency net database(AFND) (<http://www.allelefreqencies.net/hla6006a.asp>) and 27 HLA-II allele [30].(The more details see additional file 1: Table S1 and S2). According to the IEDB recommended method, among HLA-I a low percentile rank (PR) showed a good binder, among HLA-II a low adjusted rank (AR) showed a good binder. In prediction of HLA-I and HLA-II restricted epitopes the $PR \leq 1.0$ and $AR \leq 5.0$ were selected for the further analysis, respectively [29, 31].

The ABCpred server (<https://webs.iitd.edu.in/raghava/abcpred/index.html>) was used to predict B cell antigen epitopes of HPV-52 E6/E7 reference and variant sequences according to the default parameters[32]. The higher predicted score represented a better affinity [29, 31]. The server is able to predict epitopes with 65.93% accuracy using recurrent neural network.

Transcription factor binding sites prediction

To analyze the transcription factor binding sites in the LCR of HPV-52, the JASPAR database (<http://jaspar.genereg.net/>) was used [33]. The sites for CEBPA, CEBPB, CREB1, ELK4, ESR2, ETS1, FOS, FOXC1, FOXL1, FOXP3, HOXA5, HOXC11, HSF1, IRF2, JUN, NFIA, NFKB1, PHOX2A, POU2F2, RAX, SMAD3, SOX9, SOX10, SPIB, SRY, STAT1, VAX1 and YY1 were included. The relative profile score threshold was set at 85% [23].

Results

HPV prevalence in Sichuan China

Of the 3432 collected outpatient samples, 703 samples (20.48%, 703/3432) were HPV positive. By genotyping, 25 HPV types were identified, including high-risk HPV (HPV-16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73 and 82) and low-risk HPV (HPV-6, 11, 42, 43, 44, 81 and 83). Of the types observed, HPV-52 (22.05%, 155/703), 58 (16.07%, 113/703) and 16 (14.37%, 101/703) were the three most common types in the positive samples.

Genomic polymorphisms of HPV-52 E6/E7/LCR

Among the 46 HPV-52 positive samples from the outpatients, only 41 entire E6/E7/LCR sequences were obtained and further analyzed. It is might be the possible explanation for not all sequences successfully amplified that the small copy number of HPV in some samples. Our data showed that the total length of the HPV-52 E6 Open Reading Frame (ORF) was 447 bp and that of E7 was 300 bp, which were consistent with the reference sequence. LCR was only 879 bp, with a 10 bp deletion compared with the reference sequence in 97.56% (40/41) samples. A total of 50 SNP sites and 3 deletions were identified across the E6, E7 and LCR. The combinations of all these SNP sites made for 12 unique E6/E7/LCR sequences (variant No: 1-12, Table 2). The 12 variant had the most variation compared with reference sequence.

For E6, 6 SNP sites were observed compared with the reference sequence and the combinations of

all these SNP made for 5 unique E6 sequences. G350T and A378G were the most variable sites and were observed in 100% and 97.56% (40/41) of the samples, respectively. G108C and A379G, leading to the amino acids substitution of E3Q and K93R, respectively. A378C and A379G together led to the amino acid substitution of K93R. G350T, G356A and A530G were synonymous mutation.

For E7, 10 SNP sites were observed compared with the reference sequence and the combinations of all these SNP made for 2 unique E7 sequences. A801G and C751T were the most variable sites and were observed in 100% and 97.56% (40/41) of the samples, respectively. C662T, AG706/707GA, T727G, C733T, G742A and T848G, leading to the amino acids substitution of T37I, S52D, Y59D, H61Y, D64N and L99R, respectively. T573A, C751T and A801G were synonymous mutation. No deletion or insertion mutation sites were found in either E6 or E7 sequences. The secondary structures also had not changed in either E6 or E7 protein.

For LCR, 34 SNP sites and 3 deletions were observed compared with the reference sequence and the combinations of all these mutations made for 9 unique LCR sequences. G7622A, T7624G, T9659C, G7712C and G7861A were observed in all samples. G7168C, C7207A, G7371, A7657C, A7865G and T13C were observed in 97.56% (40/41) of the samples. Two deletions were found between the nucleotide sites 7387 to 7391 (TTATG) and 7698 to 7700 (CTT) in all samples. A deletion was found between the nucleotide sites 7287 to 7288 (TG) in 97.56% (40/41) of the samples. Also, T7933C and A7938G were the common variable sites and were observed in 36.59% (15/41) and 60.98% (25/41) of the samples, respectively.

Phylogenetic analysis

The neighbor-joining phylogenetic trees were constructed by MEGA 7.0, using the 12 unique HPV-52 E6/E7/LCR variant sequences and 7 sub-lineages reference sequences. The phylogenetic tree (Fig.1) showed that all variants were clustered in sub-lineage B2, except variant NO.12 (sub-lineage C2). It was basically consistent with a study related to HPV-52 in Japan [34].

Selective pressure analysis

The selective pressure analysis results by the Bayes empirical Bayes (BEB) analysis of PAML 4.9 were indicated that: in the HPV-52 E6 and E7 protein coding sequences, no positive selection sites were observed.

HLA-I and HLA-II binding peptides prediction

To evaluate the impacts of the HPV-52 E6 and E7 sequence polymorphisms on the immune recognition of antigens, the binding peptides for both HLA I and HLA II were predicted by the IEDB servers. Based on the principle of epitope selection described in the methods section, 273 and 275 E6 HLA I predicted epitopes were selected from the E6 reference and variant sequences, respectively (The more details see additional file 2: Table S3 and S4). HLA-A*30:02 was the most frequency, followed by HLA-C*06:02, HLA-C*07:02 and HLA-B*38:01. HLA-A*11:01 (₈₆₋₉₄KTLEERVKK), HLA-A*33:03 (₁₂₇₋₁₃₅NIMGRWTGR) and HLA-A*30:01 (₉₁₋₉₉RVRKPLSEI/₉₁₋₉₉RVKKPLSEI) showed the best binding affinity (PR=0.01). 134 and 114 E7 HLA I predicted epitopes were selected from the E7 reference and variant sequence, respectively (The more details see additional file 3: Table S5 and S6). HLA-A*01:01 was the most frequency, followed by HLA-B*44:02, HLA-B*44:03 and HLA-A*30:02. HLA-B*44:02 (₄₇₋₅₅AEQATSNYY/₄₇₋₅₅AEQATDNYY), HLA-B*44:03 (₄₇₋₅₅AEQATSNYY/₄₇₋₅₅AEQATDNYY) and HLA-B*15:01 (₄₅₋₅₄GQAEQATSNY) showed the best binding affinity (PR=0.01).

556 E6 HLA II predicted epitopes were selected in both of the E6 reference and variant sequence (The more details see additional file 4: Table S7 and S8). The three best epitopes were HLA-DPA1*01:03/DPB1*04:01, HLA-DRB1*03:01 and HLA-DRB4*01:01. Due to K93R, there was

a potentially affected on the epitope HLA-DPA1*02:01/DPB1*01:01. 101 and 120 E7 HLA II predicted epitopes were selected from the E7 reference and variant sequence, respectively (The more details see additional file 5: Table S9 and S10). The three best epitopes were HLA-DQA1*03:01/DQB1*03:02, HLA-DRB3*01:01 and HLA-DQA1*05:01/DQB1*02:01. Due to Y59D, H61Y and D64N, there were a potentially affected on the epitopes HLA-DRB3*02:02 and HLA-DQA1*01:02/DQB1*06:02.

B cell binding peptides prediction

In HPV-52 E6, a total of 13 B cell potential epitopes were predicted in both of the reference and variant sequences. The most potent epitopes were ₁₂₉₋₁₄₄MGRWTGRCSECWRPRP and ₁₀₈₋₁₂₃TPLCPEEKERHVNANK. Due to K93R, the score of epitopes ₈₉₋₁₀₄EERVKKPLSEITIRCI and ₈₁₋₉₆YSLYGKTLSEERVKKPL changed from 0.87 to 0.90 and 0.81 to 0.84, respectively; due to E3Q, the score of epitope ₃₋₁₈EDPATRPRTLHELCEV changed from 0.68 to 0.74. The other prediction epitopes were consistent between reference and variant sequences, and no increased epitopes in the variant sequences (The more details see additional file 6: Table S11).

In HPV-52 E7, a total of 10 B cell potential epitopes were predicted in both of the reference and variant sequences. The most potent epitopes were ₂₃₋₃₈HCYEQLGDSSDEEDTD and ₃₄₋₄₉EEIDGVDRPDGQAEQ. Due to T37I, S52D, Y59D, H61Y and D64N, the prediction epitopes of reference and variant sequences had some differences, only 4 prediction epitopes were completely consistent between reference and variant sequences, an increased high affinity epitope ₄₄₋₅₉DGQAEQATDNYYIVTD was discovered in the variant sequences (The more details see additional file 6: Table S12).

Prediction of the transcription factor binding sites

The online JASPAR database was used to investigate the potential binding sites for the transcription factors in HPV-52 LCR reference and variant sequences. The results showed that LCR region had high-affinity binding sites for transcription factors of SOX9, FOS, RAX, HOXA5, VAX1 and SRY. G7622A and T7624G, G7861A and C7917A/G76C lead to increase the binding sites for FOXC1, RAX/VAX1/HOXA5/PHOX2A and FOXL1, respectively. In addition, the nucleotide sites 21, 7168, 7414, 7580, 7865 and 7983 potentially affected the binding sites for FOXP3, CEBPB, SRY, CEBPB, VAX1 and VAX1/RAX/HOXA5, respectively.

Discussion

According to Marc Arbyn research, approximately 570 000 women developed cervical cancer and 311 000 women died in 2018. Worldwide, cervical cancer was the fourth most common cancer among women, after breast cancer, colorectal cancer, and lung cancer; and it was also the fourth leading cause of cancer death among women, after breast, lung and colorectal cancers. Approximately 84% of all cervical cancers and 88% of all deaths caused by cervical cancer occurred in lower-resource countries, of which 1.8% of women were diagnosed with and 1.3% died from the disease before age 75 years, in the absence of competing causes of death. By contrast, in the highest-resource countries, the cumulative rates of cervical cancer incidence and mortality were two to four times lower than those in lower-resource countries. And China was the country with the highest number of cases (106 000), whereas India was the country with the highest estimated number of cervical cancer deaths (60 000). China and India together contributed 35% to the global burden of cervical cancer cases and deaths[35]. Therefore, it is still particularly important to study cervical cancer in China.

The persistent infection with high-risk HPV types is the main cause in triggering the development of cervical cancer, such as the types 16, 18, 52 and 58 [36]. HPV 52 is one of the most relevant HPV types

especially in Southeast Asia, where it causes up to 20% of all cervical cancer [37, 38]. Through preliminary exploration in the early stage, we found that HPV-52 accounting for 22.05% of all the HPV-positive samples was the most common high-risk types, followed by HPV-58 and HPV-16 in Sichuan, China. In addition, other studies have shown that the distribution of HPV-52 has a certain regional distribution, mainly related to cervical cancer in Asian countries like China and South Korea [39, 40]. Compared with the reference sequence, the HPV-52 variants were clustered in sub-lineage B2 and C2, no variants belonged to the lineage A and D in our study. In addition, the HPV-52 LCR variants from Sichuan, China have not been reported yet. Our study showed that the variation of the HPV-52 LCR was showed in a higher ratio than those of the E6 and E7, the nucleotide variations were LCR > E7 > E6 found in 3.82%, 3.33% and 1.34%, respectively. The most common non-synonymous substitution in the HPV-58 E6 was A278G (K93R), and E7 was only one sample with non-synonymous. The viral proteins E6 and E7 function as the main regulators of HPV-induced tumorigenesis, and changes in amino acids may influence the transforming activity of the E6 and E7 oncoproteins [41]. Identifying new variants in HPV-52 E6/E7 may inform the rational design of new vaccines specifically for women in southwest China.

The amino acid changes in E6 and E7 oncoproteins can influence the HLA binding peptides and have a significant immunological effect on the immune system's ability of recognition of these viral antigens. Unlike prophylactic HPV vaccines, which are used to generate neutralizing antibodies against viral particles, therapeutic HPV vaccines are used to stimulate cell-mediated immune responses to specifically target and kill infected cells. In the cell mediated immune responses, cytotoxic T lymphocytes (CTLs) were considered as the major eradicators of both HPV-infected cells and cervical cancer [42]. HPV oncoproteins E6 and E7 are responsible for the malignant progression of HPV-associated diseases and are consistently expressed in HPV-associated diseases and cancer lesions. E6 and E7 act as the promising specific tumor antigens and are available as the therapeutic targets [43]. Furthermore, therapeutic HPV vaccines targeting E6 and E7 can circumvent the problem of immune tolerance against selfantigens because these virus encoded oncogenic proteins are foreign proteins to human bodies [44]. In our study, we had identified the high affinity epitopes of E6/E7 oncoproteins for T cells and B cells, providing certain basic data for the development of therapeutic vaccines.

Mutations on LCR may influence the binding sites and the function of it. The HPV LCR which contains the binding sites for both viral and cellular factors, has shown regulatory functions on replication of HPV, transcriptional activity of the E6/E7 and the other interaction through the virus life cycle [25, 38]. In our study, VAX1, CEBPB, FOXL1, PHOX2A and HOXA5 were the transcription factors that may be affected in HPV-52. One study had shown that VAX1 was closely related to bladder cancer recurrence [45]. CEBPB is a leucine-zipper transcription factor that regulates growth and differentiation of hematopoietic and epithelial cells. One study based on breast cancer found that CEBPB was a novel transcriptional regulator of CLDN4. The upregulation of CEBPB-CLDN4 signaling caused the migration and invasion of cancer cell [46]. FOXL1 is a member of the Forkhead box (FOX) superfamily and was reported to be dysregulated in various types of cancers [47]. PHOX2A was a transcription factor involving in cell proliferation and migration in lung cancer [48]. HOXA5 is a member of the homeobox (HOX) family and is upregulated in many types of tumors [48, 49].

In addition, SOX10, a transcription factor of the sex determining region Y (SRY)-related high motility group (HMG)-box gene family, playing an important role in cancer progression, including tumorigenesis, changes in the tumor microenvironment, and metastasis [50, 51]. Studies had shown that FOS was closely related to the pathogenesis of bone tumors in mice [52]. The study by Yang M et al. showed that in addition to stimulating PKR activity, RAX can positively regulate both SV40 large T

antigen-dependent DNA replication and transcription in a mechanism that may alter the interaction of the cellular factor(s) with the SV40 enhancer via the dsRNA-binding domains of RAX [53]. This function of RAX may have implications for regulation of HPV replication and transcription because of the many similarities between the viral and cellular processes. SOX9, FOS, RAX and SRY were the high probability binding sites in HPV-52 LCR.

Conclusion

In conclusion, this study investigated the gene polymorphisms and phylogeny of high-risk HPV-52 E6/E7/LCR, the possible influence of non-synonymous substitutions in E6/E7 on the T-cell and B-cell response and the impacts of the LCR variations on the bindings of the cellular transcription factors from Southwest China. Knowledge of genetic variation in HPV may be useful as an epidemiologic correlate of cervical cancer risk, or may even provide critical information for developing diagnostic probes. Although our study showed some limitations on sample capacity and source, it provided more basic data for the further immunotherapeutic approaches and vaccine development strategies. It also helps performing further study to demonstrate the biological function of HPV-52 E6/E7/LCR variants and the effect of multiple infection of high-risk HPV on tumor progression. The TFBS we found are still need deeper exploration for the potential of them to be marker in diagnosis and therapy.

Additional files

Additional file 1: Table S1. HLA-I alleles and lengths selected for prediction. **Table S2.** HLA-II alleles and lengths selected for prediction

Additional file 2: Table S3. Prediction results for HPV52 E6 reference sequence HLA-I epitope peptides. **Table S4.** Prediction results for HPV52 E6 variant sequence HLA-I epitope peptides

Additional file 3: Table S5. Prediction results for HPV-52 E7 reference sequence HLA-I epitope peptides. **Table S6.** Prediction results for HPV-52 E7 variant sequence HLA-I epitope peptides

Additional file 4: Table S7. Prediction results for HPV-52 E6 reference sequence HLA-II epitope peptides. **Table S8.** Prediction results for HPV-52 E6 variant sequence HLA-II epitope peptides

Additional file 5: Table S9. Prediction results for HPV-52 E7 reference sequence HLA-II epitope peptides. **Table S10.** Prediction results for HPV-52 E7 variant sequence HLA-II epitope peptides.

Additional file 6: Table S11. Predicted linear B cell epitopes of HPV-52 E6. **Table S12.** Predicted linear B cell epitopes of HPV-52 E7.

Abbreviations

DNA: Deoxyribonucleic Acid; HPV: Human Papillomavirus; IEDB: The Immune Epitope Database; LCR: long control region; MEGA: Molecular Evolutionary Genetics Analysis; NCBI: National Center for Biotechnology Information; PCR: Polymerase Chain Reaction; SNP: single nucleotide polymorphism; PAML: Phylogenetic Analysis by Maximum Likelihood; TFBS: transcription factor binding site. the amino acid shorthand A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W and Y: alanine, cysteine, aspartic acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparagine, proline, glutamine, arginine, serine, threonine, valine, tryptophan and tyrosine.

Declarations

Ethical approval and consent to participants

All participants were informed of the study aims, and a written informed consent was received from

each patient before sample collection, and the patients' privacy have been fully protected. This study was approved by education and research committee and Ethics Committee of Sichuan University, China (approval number SCU20100196494), and was carried out in line with the Helsinki Declaration.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and GenBank.

Competing interests

The authors declare that they have no competing interests.

Funding

The present study was funded by Key Scientific Research Foundation Projects of Sichuan Province (No.2018JY0601).

Authors' contributions

Conception of the study and supervising the progress of study: XD and ZS. Conduct experiments: ZS. Sample collection: ZS, YC and QL. HPV typing: ZS, JD and YC. Other parts of experiments: LF, JD, YL, ZJ and JH. Data analysis: ZL and LF. Draft of manuscript: ZS. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was funded by Key Scientific Research Foundation Projects of Sichuan Province and supported by following hospitals: The Affiliate Reproductive Hospital of Sichuan Genitalia Hygiene Research Center, Chengdu SongZiNiao Sterility Hospital and Chengdu Medical College Affiliated Infertility Hospital and others that participated in this study.

Authors' information

1: Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, Sichuan, People's Republic of China.

2: Bio-resource Research and Utilization Joint Key Laboratory of Sichuan and Chongqing, Chongqing, People's Republic of China.

References

1. Schiffman, M. and D. Solomon, *Cervical-Cancer Screening with Human Papillomavirus and Cytologic Cotesting*. New England Journal of Medicine, 2013. **369**(24): p. 2324-2331.
2. Tewari, K.S., et al., *Improved survival with bevacizumab in advanced cervical cancer*. N Engl J Med, 2014. **370**(8): p. 734-43.
3. Gaffney, D.K., et al., *Too many women are dying from cervix cancer: Problems and solutions*. Gynecol Oncol, 2018. **151**(3): p. 547-554.
4. zur Hausen, H., *Papillomaviruses in the causation of human cancers - a brief historical account*. Virology, 2009. **384**(2): p. 260-5.
5. Torre, L.A., et al., *Global cancer statistics, 2012*. CA Cancer J Clin, 2015. **65**(2): p. 87-108.

6. Harari, A., Z. Chen, and R.D. Burk, *Human papillomavirus genomics: past, present and future*. *Curr Probl Dermatol*, 2014. **45**: p. 1-18.
7. Bzhalava, D., et al., *Deep sequencing extends the diversity of human papillomaviruses in human skin*. *Scientific Reports*, 2014. **4**(1).
8. Chen, Z., et al., *Characteristics of HPV prevalence in Sichuan Province, China*. *Int J Gynaecol Obstet*, 2015. **131**(3): p. 277-80.
9. Geng, Y. and L. Liu, *Human papillomavirus genotypes and infection among women in Changzhou, China*. *Hum Vaccin Immunother*, 2019. **15**(7-8): p. 1884-1888.
10. Flanagan, M.B., *Primary High-Risk Human Papillomavirus Testing for Cervical Cancer Screening in the United States: Is It Time?* *Arch Pathol Lab Med*, 2018. **142**(6): p. 688-692.
11. Li, N., et al., *Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication*. *Int J Cancer*, 2011. **128**(4): p. 927-35.
12. Ferenczi, A., et al., *Sequence variation of human papillomavirus type 31 long control region: phylogenetic and functional implications*. *J Med Virol*, 2013. **85**(5): p. 852-9.
13. Mehanna, H., et al., *Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer--systematic review and meta-analysis of trends by time and region*. *Head Neck*, 2013. **35**(5): p. 747-55.
14. Berman, T.A. and J.T. Schiller, *Human papillomavirus in cervical cancer and oropharyngeal cancer: One cause, two diseases*. *Cancer*, 2017. **123**(12): p. 2219-2229.
15. Aimagambetova, G. and A. Azizan, *Epidemiology of HPV Infection and HPV-Related Cancers in Kazakhstan: a Review*. *Asian Pac J Cancer Prev*, 2018. **19**(5): p. 1175-1180.
16. Tommasino, M., *The human papillomavirus family and its role in carcinogenesis*. *Semin Cancer Biol*, 2014. **26**: p. 13-21.
17. Xi, J., et al., *Genetic variability and functional implication of the long control region in HPV-16 variants in Southwest China*. *PLoS One*, 2017. **12**(8): p. e0182388.
18. Narisawa-Saito, M. and T. Kiyono, *Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins*. *Cancer Sci*, 2007. **98**(10): p. 1505-11.
19. Strati, K. and P.F. Lambert, *Role of Rb-dependent and Rb-independent functions of papillomavirus E7 oncogene in head and neck cancer*. *Cancer Res*, 2007. **67**(24): p. 11585-93.
20. Chenzhang, Y., et al., *Identification of the impact on T- and B- cell epitopes of human papillomavirus type-16 E6 and E7 variant in Southwest China*. *Immunol Lett*, 2017. **181**: p. 26-30.
21. Amaro-Filho, S.M., et al., *HPV DNA methylation at the early promoter and E1/E2 integrity: A comparison between HPV16, HPV18 and HPV45 in cervical cancer*. *Papillomavirus Res*, 2018. **5**: p. 172-179.
22. Martinez-Ramirez, I., et al., *SOX2 as a New Regulator of HPV16 Transcription*. *Viruses*, 2017. **9**(7).
23. Xi, J., et al., *The polymorphisms of LCR, E6, and E7 of HPV-58 isolates in Yunnan, Southwest China*. *Virol J*, 2018. **15**(1): p. 76.
24. Graham, S.V., *Keratinocyte Differentiation-Dependent Human Papillomavirus Gene Regulation*. *Viruses*, 2017. **9**(9).
25. Guan, P., et al., *Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer*. *Int J Cancer*, 2012. **131**(10): p. 2349-59.
26. Burk, R.D., Z. Chen, and K. Van Doorslaer, *Human papillomaviruses: genetic basis of carcinogenicity*. *Public Health Genomics*, 2009. **12**(5-6): p. 281-90.
27. Zhang, Y., et al., *Genetic variability in E6, E7, and L1 genes of human papillomavirus genotype 52*

- from Southwest China. *Gene*, 2016. **585**(1): p. 110-118.
28. Kumar, A., et al., *Identification of human papillomavirus-16 E6 variation in cervical cancer and their impact on T and B cell epitopes*. *J Virol Methods*, 2015. **218**: p. 51-8.
 29. He, J., et al., *Identification of variants and therapeutic epitopes in HPV-33/HPV-58 E6 and E7 in Southwest China*. *Virology*, 2019. **16**(1): p. 72.
 30. Martini, S., *HLA allele frequencies and reference sets with maximal population coverage*. 2019.
 31. Yao, Y., et al., *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*, 2013. **31**(18): p. 2289-94.
 32. Saha, S. and G.P.S. Raghava, *Prediction of continuous B-cell epitopes in an antigen using recurrent neural network*. *Proteins: Structure, Function, and Bioinformatics*, 2006. **65**(1): p. 40-48.
 33. Fornes, O., et al., *JASPAR 2020: update of the open-access database of transcription factor binding profiles*. *Nucleic Acids Res*, 2020. **48**(D1): p. D87-D92.
 34. Tenjimbayashi, Y., et al., *Whole-genome analysis of human papillomavirus genotypes 52 and 58 isolated from Japanese women with cervical intraepithelial neoplasia and invasive cervical cancer*. *Infect Agent Cancer*, 2017. **12**: p. 44.
 35. Arbyn, M., et al., *Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis*. *The Lancet Global Health*, 2020. **8**(2): p. e191-e203.
 36. Burk, R.D., A. Harari, and Z. Chen, *Human papillomavirus genome variants*. *Virology*, 2013. **445**(1-2): p. 232-43.
 37. Chan, P.K., et al., *Meta-analysis on prevalence and attribution of human papillomavirus types 52 and 58 in cervical neoplasia worldwide*. *PLoS One*, 2014. **9**(9): p. e107573.
 38. Marušič, M. and J. Plavec, *Towards Understanding of Polymorphism of the G-rich Region of Human Papillomavirus Type 52*. *Molecules*, 2019. **24**(7).
 39. Huang, S.L., et al., *Comparison between the Hybrid Capture II Test and an SPF1/GP6+ PCR-based assay for detection of human papillomavirus DNA in cervical swab samples*. *J Clin Microbiol*, 2006. **44**(5): p. 1733-9.
 40. Soeda, E., et al., *Repression of HPV16 early region transcription by the E2 protein*. *Virology*, 2006. **351**(1): p. 29-41.
 41. Boulet, G., et al., *Human papillomavirus: E6 and E7 oncogenes*. *Int J Biochem Cell Biol*, 2007. **39**(11): p. 2006-11.
 42. Nilges, K., et al., *Human papillomavirus type 16 E7 peptide-directed CD8+ T cells from patients with cervical cancer are cross-reactive with the coronavirus NS2 protein*. *J Virol*, 2003. **77**(9): p. 5464-74.
 43. Yang, A., et al., *Current state in the development of candidate therapeutic HPV vaccines*. *Expert Review of Vaccines*, 2016. **15**(8): p. 989-1007.
 44. Yang, A., et al., *Perspectives for therapeutic HPV vaccine development*. *J Biomed Sci*, 2016. **23**(1): p. 75.
 45. Zhao, Y., et al., *Methylcap-seq reveals novel DNA methylation markers for the diagnosis and recurrence prediction of bladder cancer in a Chinese population*. *PLoS One*, 2012. **7**(4): p. e35175.
 46. Wang, F., et al., *A novel PAK4-CEBPB-CLDN4 axis involving in breast cancer cell migration and invasion*. *Biochem Biophys Res Commun*, 2019. **511**(2): p. 404-408.
 47. Qin, Y., et al., *Forkhead box L1 is frequently downregulated in gallbladder cancer and inhibits cell growth through apoptosis induction by mitochondrial dysfunction*. *PLoS One*, 2014. **9**(7): p. e102084.
 48. Fang, L., et al., *Genetic variability, phylogeny and functional implication of the long control region in human papillomavirus type 16, 18 and 58 in Chengdu, China*. *Virology*, 2020. **17**(1): p. 106.

49. Zhang, H., J.H. Zhao, and Z.M. Suo, *Knockdown of HOXA5 inhibits the tumorigenesis in esophageal squamous cell cancer*. *Biomed Pharmacother*, 2017. **86**: p. 149-154.
50. Castillo, S.D. and M. Sanchez-Cespedes, *The SOX family of genes in cancer development: biological relevance and opportunities for therapy*. *Expert Opinion on Therapeutic Targets*, 2012. **16**(9): p. 903-919.
51. Grimm, D., et al., *The role of SOX family members in solid tumours and metastasis*. *Semin Cancer Biol*, 2019.
52. Fittall, M.W., et al., *Recurrent rearrangements of FOS and FOSB define osteoblastoma*. *Nature Communications*, 2018. **9**(1).
53. Yang, M., T. Ito, and W.S. May, *A novel role for RAX, the cellular activator of PKR, in synergistically stimulating SV40 large T antigen-dependent gene expression*. *J Biol Chem*, 2003. **278**(40): p. 38325-32.

Figures

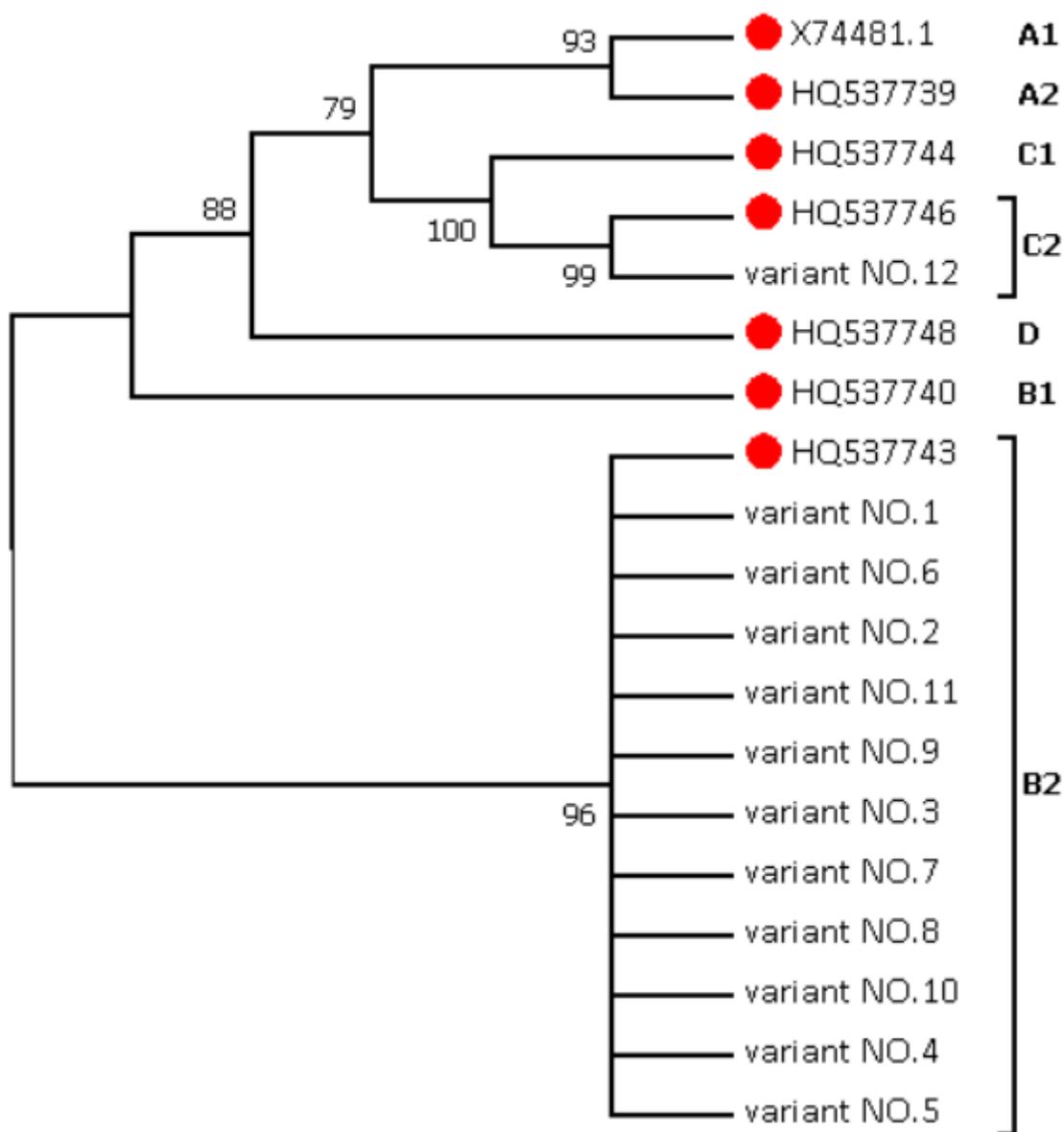


Figure 1

Neighbor-joining phylogenetic tree based on HPV-52 E6/E7/LCR sequences, and reference sequences of sub-lineages were marked with red dot.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1HPV52E6E7andLCRprimersequences.pdf](#)
- [Table2HPV52E6E7andLCRnucleotidemutationsites.pdf](#)

- [Additionalfile6.pdf](#)
- [Additionalfile5.pdf](#)
- [Additionalfile4.pdf](#)
- [Additionalfile3.pdf](#)
- [Additionalfile2.pdf](#)
- [Additionalfile1.pdf](#)