Translational Oncology 15 (2022) 101256

Contents lists available at ScienceDirect

Translational Oncology

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

Original Research

Predominance of genomically defined A lineage of HPV16 over D lineage in Indian patients from eastern India with squamous cell carcinoma of the cervix in association with distinct oncogenic phenotypes

Paramita Mandal [a,d], Bornali Bhattacharjee [a,e,*], Shrinka Sen [a,f], Amrapali Bhattacharya [a,g], Sweta Sharma Saha [a,h], Rahul Roy Chowdhury [b], Nidhu Ranjan Mondal [b], Biman Chakrabarty [b], Tanmay Chatterjee [b], Sudipta Roy [c], Sharmila Sengupta [a,*]

[a] National Institute of Biomedical Genomics, P.O.: N.S.S., Kalyani, West Bengal 741251, India
[b] Saroj Gupta Cancer Centre and Research Institute, Kolkata, India.
[c] AMRI Group of Hospitals, Kolkata, India
[d] Biomedical Genetics Laboratory, Department of Zoology, The University of Burdwan, Purba Burdwan, West Bengal 713104, India
[e] Amity Institute of Biotechnology, Amity University Kolkata, Kolkata, India
[f] Department of Medical Genetics, University of British Columbia, 4500 Oak Street, Vancouver, British Columbia V6H 3N1, Canada
[g] DBT-NIDAN Kendra on Rare Genetic Diseases, N. R. S. Medical College and Hospital, 138, A. J. C. Bose Road, Kolkata 700014, India
[h] Translational and Clinical Research Institute, Faculty of Medical Sciences, Centre for Cancer, Newcastle University, NE2 4HH, UK

ARTICLE INFO

Keywords:
Cervical cancer
Episomal HPV16 lineages
Phylogenetic networks
E7 expression
Viral copy numbers

ABSTRACT

Human papillomavirus type-16 (HPV16) is classified into lineages, A, B, C and D and 10 sub-lineages portraying variable infectivity, persistence, and cytological outcomes, however, with geographical variations. Our objective was to delineate the distinctive features of lineages among cervical squamous cell carcinoma (SCC) in the eastern region of India. A total of 145 SCC cases and 24 non-malignant specimens, harboring episomal HPV16, were included. The presence of higher proportion of lineage A over D was observed among SCC cases (86.89% A1, 8.97% D1 and 4.14% D2), while only A1 sub-lineage viruses were found among control specimens. Among the A1 viruses, an association of variants in the E5 (Y44L, I65V), E6 (L83V) genes and LCR: C7577T with SCC, with combined Odd's ratio (95% CI) of 20.5(4.61–91.25) was observed. Network analyses revealed the presence of 10 clades of lineage A viruses comprising of 64 HPV16 genomes harboring the risk alleles. High episomal HPV16 DNA copy numbers (adjusted p-value= 0.0271) and E7 mRNA expression (p-value=0.000017) predominated in SCC with lineage A, over D. Our study highlights the distinctive modalities of oncogenicity among different HPV16 lineages.

Introduction

In most countries including India, 80% of cervical cancer cases are of squamous cell origin [1] and human papillomaviruses (HPV) are the established drivers [2]. Of the repertoire of oncogenic HPV infections, HPV16 appears to be uniquely carcinogenic, accounting for over fifty percent of all CaCx cases worldwide [3–5]. It is observed to be correlated with virus persistence and often with increased virus burden within host cells [6]. This leads to enhanced risk of developing prevalent and incident pre-invasive lesions and ultimately CaCx [7]. In India, majority of the CaCx cases are associated with HPV16 infection [8–10]. The predominance of HPV16 among the general population is also evident based on data available from some regions of India [9,11] including an earlier study from our group [12]. A recent study from our group has also shown the preponderance of HPV16 infection among women seeking healthcare at a hospital outpatient department [13].

Worldwide, HPV16 related CaCx prevalence varies across different regions and countries, which is often attributable to several factors related to the virus, besides the host and the environment. These factors include physical status of HPV16 genomes, viral load, methylation in E2 binding sites, HPV16 E7 expression etc as published previously from our laboratory [14,15]. Therefore, studies have often focused on the

* Corresponding authors at: National Institute of Biomedical Genomics, P.O.: N.S.S., Kalyani, West Bengal 741251, India; Amity Institute of Biotechnology, Amity University Kolkata, Kolkata, India
E-mail addresses: bbhattacharjee@kol.amity.edu (B. Bhattacharjee), ssg1@nibmg.ac.in (S. Sengupta).

https://doi.org/10.1016/j.tranon.2021.101256
Received 16 July 2021; Received in revised form 28 September 2021; Accepted 22 October 2021
1936-5233/© 2021 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

potential role of HPV16 genomic variants in modulating the risk for progression to malignancy [16–20]. Based on epidemiological data, it has also been proposed that viral genome variants diverging by about 2% within a given type and differing geographically [21], can contribute to viral pathogenicity through alteration of its carcinogenic potential and immunogenicity [22,23].

HPV16 genome variations have most extensively been studied [21, 24]. Initially, HPV16 genomes were classified into five lineages based on L1 variants. Each lineage was found to be predominant within specific geographical regions and named accordingly as E (European), As (Asian), AA (Asian-American), NA (North American), Af1 (African-1) and Af2 (African-2) [25–27]. Burk et al, in 2013 [28], used whole genome viral variants to further classify HPV16 into ten sublineages (A1, A2, A3, A4, B1, C, D1, D2, D3, D4). Further, in 2016, a comprehensive study by Mirabello et al, identified HPV16 variant lineages and sub-lineages with variable risk towards development of precancers and cancers [29].

However, most of the studies have not characterized the distinguishing attributes beyond nucleotide variants. Previous studies from our laboratory [14,15] have highlighted the biological plausibility of episomal HPV16 driven cervical carcinogenesis as opposed to integration, as well as the role of viral variants in CaCx [30,31]. The current study, an extension of our earlier work, is aimed at identifying the distinctive features of HPV16 lineages among SCC cases from the eastern region of India, with respect to genomic variations, viral DNA load and E7 expression as markers of oncogenicity. Among the two major oncoproteins, E7 over E6 was chosen because of least variability in CaCx [32], role in initiation of the quasi-S phase essential for viral genome replication in differentiated epithelial cells [33], and presence of higher number of HPV16 E7 transcripts in cancers in contrast to HPV18 cancers [34].

Materials and methods

Specimens and subjects

The HPV16 positive specimens used for this study are nested to an ongoing natural cohort study [12,14,30,31,35] which has a total of 684 cases at present. Due to lack of awareness, efficient healthcare systems and affordability, majority of the patients in India visit the hospital at an advanced stage of cancer and this cohort is no different. Among the cases, 99.56% (681/684) are squamous cell carcinomas and 69.15% (473/684) have been clinically confirmed to be tumor stage III and above as per FIGO classification (unpublished data). The subset of tumor specimens (n=145) selected for this study were histopathologically confirmed invasive SCC, tumor stage III and above, and most were pathologically classified as moderately differentiated SCC (73.1%; 106/145). These were derived from married subjects (median age: 52 years; range: 23–80 years) attending a cancer referral hospital (Saroj Gupta Cancer center and Research Institute, South 24 Parganas, West Bengal, India and were treatment naive.

The non-malignant specimens (n=24) were normal cervical scrapes confirmed by Pap smear test and derived from married and non-pregnant (or, 6 months post-partum) women (median age: 33 years; range: 24–58 years) with no previous history of cervical dysplasia/malignancy. These were collected through organization of cervical screening camps as well as from the outpatient clinics (Department of Gynecology) of various collaborating hospitals.

All specimens, tumor and non-malignant, were collected from the subjects with written informed consent approved by the institutional ethical committee for human experimentation. Details regarding subjects, specimens, DNA isolation, HPV screening and determination of the physical status of HPV16 genomes (episomal or integrated) have been described earlier from our laboratory [12,14,30,31,35]. For this study, we excluded all specimen samples that harbored purely integrated forms of HPV16 and included only those that portrayed presence of episomal HPV16 genomes, with or without integration into the host genome.

Re-sequencing of HPV16 genome by next-generation sequencing and Sanger method

Whole genome sequence data was generated using next generation sequencing for 8 specimens [36] and for the rest, Sanger sequencing was performed. Fifteen sets of overlapping primers were used for re-sequencing of HPV16 genome, as per the method established earlier in our laboratory [30,31], employing ABI Prism™3100 automated sequencer using dye terminator chemistry.

Variant SNPs and phylogenetic clustering analysis

The HPV16R genome was used for reference guided assembly. Alignment was carried out using Bioedit [37]. The specimens included had amplification at all loci (N= 170; 145 cases and 24 controls). The median-joining (MJ) method [38] of Phylogenetic Network software (http://www.fluxusengineering.com/sharenet.htm) was used to identify distinct evolutionary clustering present within the study population and the case control distribution of each haplotype.

Estimation of HPV16 DNA copy number (total and episomal) and E7 expression among the specimens

The HPV16 DNA copy number (total and episomal) and E7 mRNA expression were estimated in the selected specimens according to the method described earlier by our group [39].

Statistical analyses

The associations of viral non-synonymous variants and LCR, NCR nucleotide changes across lineages with CaCx were determined using Chi square test. To test for normality across datasets the Kolmogorov-Smirnov test was carried out. Significant differences in viral DNA copy number and genome status were evaluated using the Mann Whitney U test and were interpreted based on p-values derived after multiple testing correction using the *Benjamini-Hochberg* method [40]. Differences in E7 mRNA expression were evaluated using the Welch's t test. A p-value less than 0.05 was considered statistically significant. All statistical analyses were done using GraphPad prism version 8.0.0 [www.graphpad.com].

Results

Presence of distinct HPV16 lineages and sublineages among Indian SCC cases from eastern region

A total of 262 single nucleotide variants were identified, out of which 251 were bi-allelic, 10 were tri-allelic and one was tetra-allelic. Among the bi-allelic variants, 101 were synonymous, 99 were non-synonymous of which, 20 were identified to be deleterious and 51 were in the non-coding regions (LCR and NCR-2). All such deleterious variants were rare (frequency < 0.05), except for the E2/E4 variant A3366C. The bi-allelic variant loci frequencies identified across the different regions of the HPV16 genome, is presented in **Table S1.**

Based on earlier reports [17,24], E2 (T3694A), E6 (A532G) and LCR (T7743G, G7834T) variants were used to identify lineages and sub-lineages. A total of two lineages and three sublineages namely, European (A1), North-American (D1) and Asian-American 1 (D2) were identified among the specimens included in this study. The A1 sublineage virus was found both in CaCx (n=126) and non-malignant specimens (n=24), whereas D1(n=6) and D2 (n=13) sublineage viruses were identified only among CaCx cases. Among the lineages and sublineages of HPV16 thus identified, additionally a total of 35 bi-allelic variants were found capable of distinguishing lineages A and D, while 6 could distinguish

between D1 and D2 sublineages (T732C, T2343C, C3161T, A4599C, A6180C, A6316G) (**Table S2**).

Association of non-synonymous and LCR variants among lineage A1 viruses with SCC

Comparative analyses between tumor and non-malignant specimens, of non-synonymous and non-coding region bi-allelic variants among A1 sublineage viruses revealed the presence of 13 variants, with significantly different distributions among cases and controls, after multiple testing correction. Those, proportionately higher among tumor specimens were E5 (A3979C:Y44L, A4042G:I65V), E6 (T350G:L83V), and LCR (C7577T) with a combined odds ratio (95% CI) of 20.5 (4.61–91.25) (Fig. 1**A and B, Table S3**). Five variants were found only among the non-malignant specimens namely, E7 (C790T:R77C), E2 (C3236G:H161D), E2/E4 (A3605G:284D), LCR (A7550G) and NCR-2 (T4222C), portraying significant p-values after multiple testing correction (Fig. 1**C**). Additionally, 4 variants had significantly higher frequencies among non-malignant specimens [E1 (G1363A; G167S), L2 (A5063C; N276T), L2 (A5492C; I418T), and L1 (A6434G; T266A)] (Fig. 1**A–C, Table S3**).

Evolutionary analysis of whole genome HPV16 haplotypes indicate the presence of four major clusters of A1 sublineage viruses

To further our understanding of the Indian viral genomes, network analyses was executed using the Hamming distances of 97 bi-allelic variants present at ≥ 5% frequency, among the genomes included in this study (N= 169; 145 tumor and 24 non-malignant specimens). A total of 68 distinct haplotypes or nodes were identified. The presence of two distinct clusters of haplotypes representing the two different lineages was reconfirmed. Among the A1 lineage viruses, four major nodes were identified, namely, E-12 (n=26), E-16 (n=13) comprising of both cases and controls and case specific nodes E-41 (n=27) and E-42 (n=17). The co-existence of LCR:C7577T, E5:A3979C, A4042G and E6:T350G variants were found in 10 A1 haplotypes or nodes (E-16, E-18, E-19, E-22, E-26, E-28, E-37, E-38, E-41 and E-42), majority of which were case specific or overrepresented among case samples and across all D haplotypes that represented cases only, except for specimen number T173, which lacked the T350G variant. Several reticulations were observed, which were caused by recurrently occurring mutations (Fig. 2).

High viral episomal DNA copy numbers and high E7 mRNA expression predominated the A lineage viruses among the SCC specimens

Significant difference in viral DNA copy numbers was observed

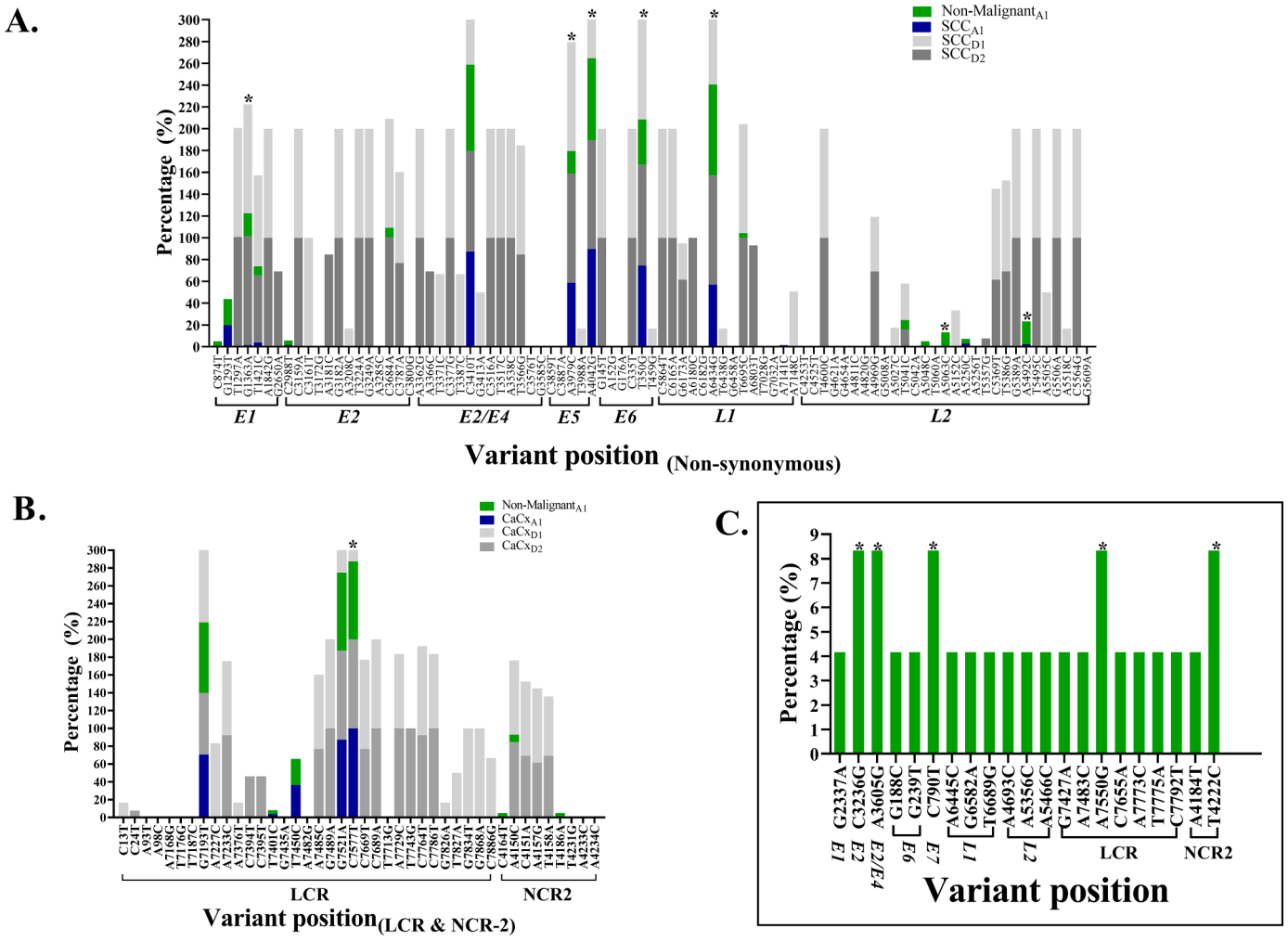

Fig. 1. Differences in distribution of HPV16 non-synonymous and non-coding region bi-allelic variations astericks identified by whole genome sequencing A. The frequencies of non-synonymous variants among different lineages across squamous cell carcinoma (SCC) cases and non-malignant specimens were compared. B. NCR-2 and LCR variations. C. Variants found only among viral DNA isolated from non-malignant specimens (p-values of < 0.05 are denoted by astericks between SCC cases and non-malignant specimens of Al lineage).

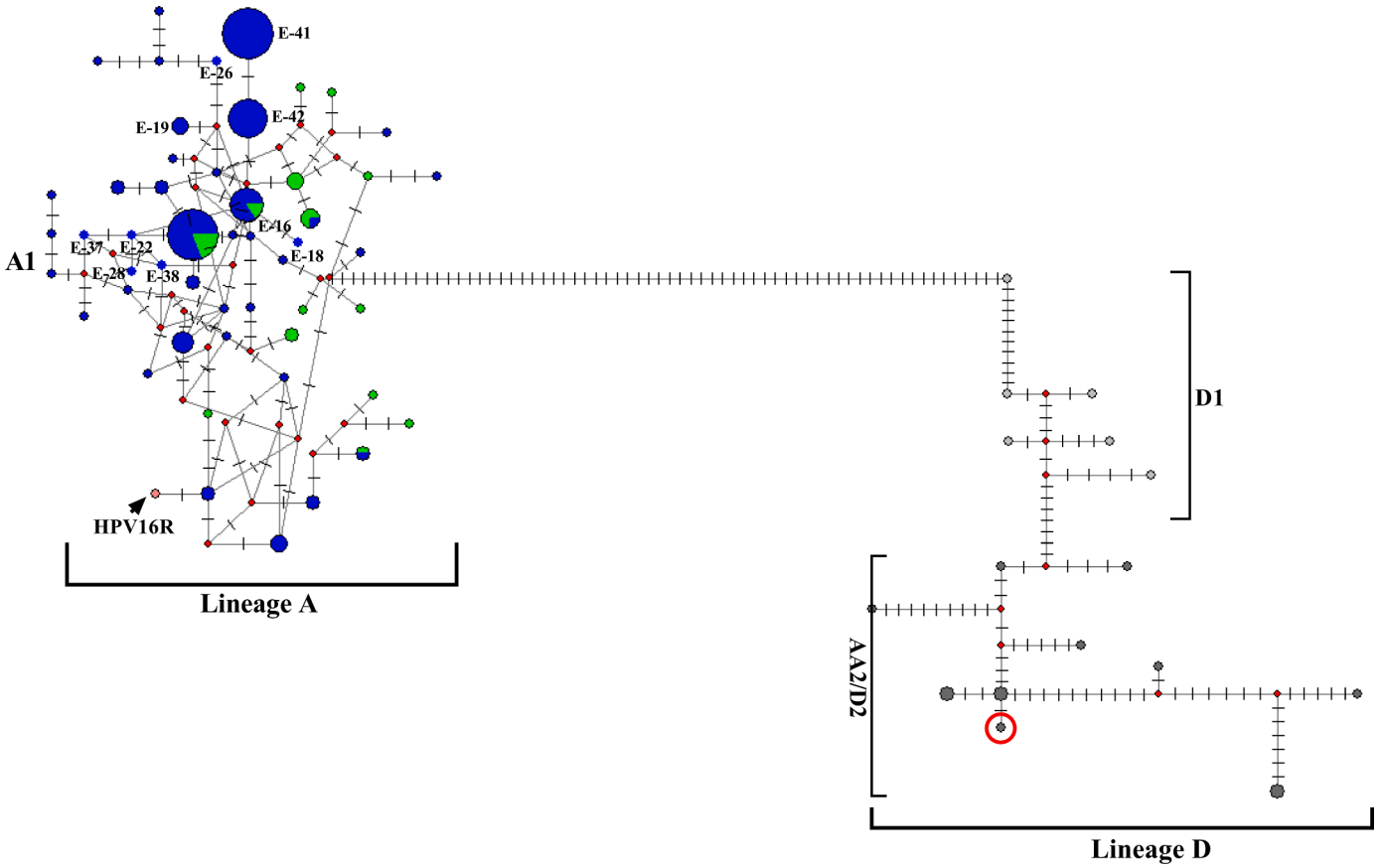

Fig. 2. Phylogenetic network of 169 HPVI6 viral genomes belonging to two distinct lineages and three sub-lineages. Each circle represents a haplotype and the diameter is proportional to the number of genomes belonging to each haplotype. A total of 97 bi-allelic variations were used to construct the haplotypes. Each notch on the horizontal lines indicate a differentiating variant. The colors indicate the sub-lineage Al; SCC (blue) and non malignant ⟨green⟩. LCR:7577T variant and the non-synonymous variants E5:3979C,A4042G, E6:350G are labeled. All the lineage D haplotypes except for one viral isolate (encircled in red) harbored both the non-synonymous variants. The smaller red circles are the median vectors and the arrow indicates the reference viral haplotype NC_001526.4.

among the tumor specimens harboring A lineage viruses (mean viral copy number= 18.39/100 ng genomic DNA) compared to the non-malignant specimens (mean viral copy number= 16.50/100 ng genomic DNA) (adjusted p-value Mann Whitney test = 0.003) as depicted in Fig. 3**A and Table S4**. However, the episomal copy numbers of HPV16 was significantly higher among tumor specimens with A lineage viruses (adjusted p-value Mann Whitney test = 0.0271) and significantly lower among those with D lineage viruses (adjusted p-value Mann Whitney test = 0.0301), compared to A lineage non-malignant specimens. Hence, an enhancement and loss of episomal copy numbers among tumor specimens with A lineage viruses and those with D lineage viruses, respectively, in the process of oncogenic transformation of the cervical epithelium was evident from this observation. Between the tumor specimens harboring the two lineages of viruses, episomal copy numbers of HPV16 was significantly higher among those with A lineage HPV16, in comparison with D lineage tumor specimens (adjusted p-value Mann Whitney test = 0.0000129) as depicted in **Fig. 3B and Table S4**. This was indicative of the relevance of E7 mRNA levels, which was estimated in a subset of CaCx tissues harboring lineage A (n= 14) and D (n=9) lineage viruses. The relative E7 expression, normalized with the housekeeping gene 18S rRNA, was significantly higher (p-value $_{t\text{-test}}$= 0.000017) in lineage A as compared with lineage D (**Fig. 3C and Table S5**) tumor specimens. The findings potentially support the oncogenic role of high copy number episomal A lineage HPV16 viruses with high E7 relative expression in the eastern Indian population. By contrast, the D lineage viruses appear to have a propensity towards integration with low levels of E7 transcripts.

Discussion

In this study, our study design considering tumor and non-malignant samples, analytical protocol based on median-joining network analysis of viral haplotypes (generated employing common sequence variations) of episomal HPV16, provides further insights on the biological relevance of HPV16 lineages and sublineages and their association with CaCx pathogenesis.

To begin with single nucleotide variants, most association studies [41–43,19] have demonstrated the presence of E6:350 G among the HPV16 A lineage viruses as a risk allele for persistence and progression to precancerous lesions among non-A sublineages [19]. By contrast, in this study we demonstrate the presence of non-synonymous variants (E5:3979C, A4042G) and non-coding variation (7577T) in LCR region in conjunction with E6:350G as risk alleles, highlighting the polygenic aspect of viral contribution to disease. Further, irrespective of the HPV16 lineages, our study also reflected the lack of genomic variations within the E7 gene among the CaCx cases, as opposed to a single variation (C790T) among the control samples. This finding is in line with a large-scale study [32] that revealed the absence of variations within E7 gene in precancer and cancer cases as well as those subsequent studies which confirmed the critical need for E7 protein conservation for maintaining oncogenicity [44,45].

As regards lineages, in conformity with earlier reported classification schemes [28–30], we identified the presence of two lineages and three sublineages - European (A1), North-American (D1) and Asian-American 1 (D2) of viruses. The A1 sublineage was prominent among both cases and controls, while the D lineage HPV16 genomes were found only

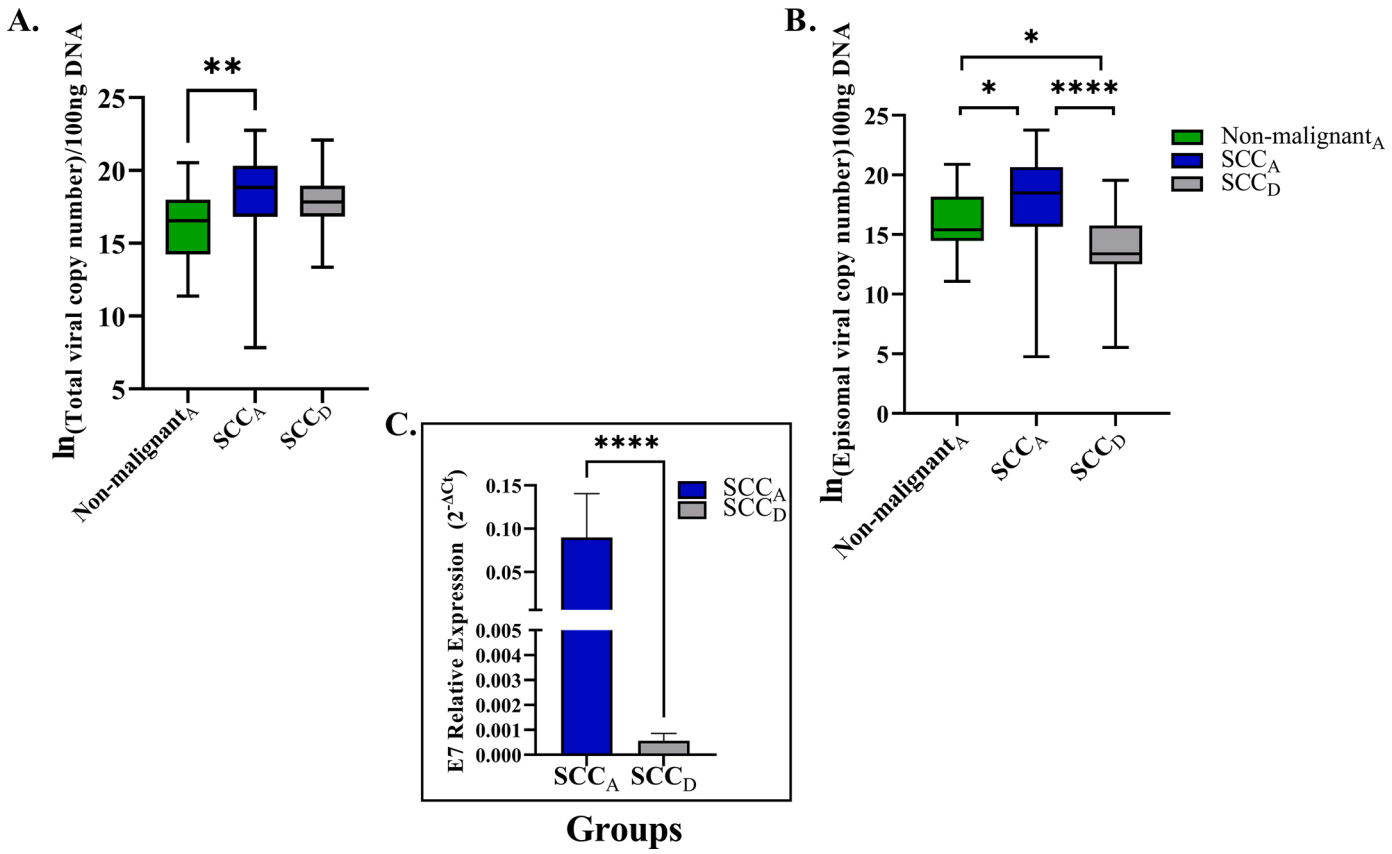

Fig. 3. Distributions of viral copy numbers and percentages of episomal genomes across non-malignant specimens and SCC lineages A. Differences in natural log(ln) viral copy numbers depicted by box-plots. The boxes indicate the interquartile ranges and the median values are indicated by horizontal black lines. B. Differences in proportions of episomal genomes. C. Relative expression of E7 gene transcripts. (corrected p-values of <0.05 are denoted by astericks).

among the case samples. Besides re-confirming our earlier observations [30] through these results, we also render support to observations recorded by other groups that relied on sequence data of a few HPV16 genes [16,46,47] portraying the predominance of A sublineage of HPV16 in the Indian population. Concurrently, the A sublineage also had high haplotype diversity. Since viruses are known to co-evolve with the host [48], the higher prevalence and haplotype diversity corresponding to the A sublineages (A1) could be attributable to a longer history of evolution with the host [48]. Likewise, the lower prevalence in the population, of the D sublineages (D1 and D2) of HPV16 could thus be associated with a shorter history, that is, they appear to be fairly new in the Indian population, with respect to the A1 sublineage. However, the presence of haplotype diversity among the A1 viruses also implies that intra-lineage variability needs to be considered while defining causality. Additionally, this study also identifies the available pool of haplotypes in the Indian population corresponding to the three sublineages of HPV16 prevailing in India, harnessing the strength of network analysis.

Some epidemiological studies have identified that HPV16 sublineages confer variable risks towards CaCx development, along with early or late onset of invasive tumors [49,50]. There are reports [51,52] that support D lineages as more oncogenic and associated with adenocarcinoma, as opposed to A lineages [50]. It has also been shown by some studies [53,54] that D lineages of HPV16 exists in the episomal form, with a higher propensity of replication. A study also revealed the aggressiveness of non-A HPV16 lineages, associated with tumor progression and radioresistance [55]. Another study identified D3 and A4 sublineages of HPV16 lineage to be associated with invasive CaCx and showed that patients with lineage B viruses exhibited worse recurrence free survival [56]. In contrast, a study also demonstrated that women with European variants were at higher risk of developing lymph node metastasis than those with non-European variants in CaCx [57].

Herein, we considered a homogeneous group of CaCx samples, all histopathologically confirmed as SCC and all harbored episomal HPV16 (pure or concomitant with viral integration). Our analysis highlighted the molecular differences between the A and D lineage viruses, with respect to viral load in association with physical status and E7 expression, the key measures of oncogenicity. This study reflected significantly higher and lower episomal copy numbers among the A and D lineage harboring tumor specimens, respectively, over the A lineage virus associated non-malignant specimens. The viral load being at equivalent levels between the A and D lineages among the tumor specimens, the episomal copy numbers were significantly higher among the A lineage viruses. Taken together, this was suggestive of loss of episomal viral DNA concomittant with the higher integration rate of the D lineage of HPV16 and enhancement of episomal copy numbers among the A lineage tumor specimens in eastern region of India. Moreover, higher E7 expression among cases with A lineages of HPV16 does preclude this as less oncogenic, compared to the D lineages recorded in our data set. The differential role of epigenomic regulation of E7 expression among the two lineages of HPV16 cannot be ruled out at this point. In support of this, in an ongoing study, our preliminary findings have revealed an overrepresentation of methylation (5/11; 45.45%) in the CpG within the NF1 binding site (as defined in the TRANSFAC database) at position 7553 within the upstream regulatory region LCR, in contrast to its absence in any of the tumor specimens harboring episomal A1 lineage (0/28; 0%) (data not shown). Overall, such observations further corroborate the oncogenic potential of HPV16 A lineages in the sampled population, while highlighting the distinctive modalities of cervical carcinogenesis among lineages and justifying the oncogenic role of episomal HPV16, as well.

In recent times, therapeutic vaccines are being employed to generate

cell-mediated immunity rather than neutralizing antibodies against transformed cells and E6 and E7 oncoproteins are the major targets. Several strategies have been investigated for enhancing CD4+ and CD8+ T-cell responses, [58]. However, the immunological impact on different HPV16 variant lineages is essential for optimizing the adoptive cell therapy approaches and vaccine development strategies [59,60].

In conclusion, our study lends support to the body of literature correlating the oncogenicity of HPV16 lineages albeit with the limitation of a small number of HPV16 positive non-malignant specimens. However, these findings lend support to our hypothesis that the episomal HPV16 employs multiple modalities in a lineage specific manner for the manifestation of the oncogenic status. Besides, variable levels of E7 gene expression among the two prevalent lineages of HPV16, suggests that the molecular milieu of the CaCx cases harboring the two lineages are likely to differ not only with respect to their capacity of pRb inactivation, but also based on the interactions of the E7 protein with other host cell molecular factors. Therefore, the CaCx cases harboring the two lineages of HPV16 might also reflect variable outcomes with respect to disease prognosis and therapy response. However, more work needs to be done to understand the role of distinct A1 haplotypes as prognosis determinants. Our study highlighting the immense heterogeneity of HPV16 lineages and sublineages in CaCx cases from eastern India, with respect to disease risk and possible prognostic and therapeutic implications, calls for the implementation of a routine HPV vaccination based primary prevention program in India.

CRediT authorship contribution statement

Paramita Mandal: Conceptualization, Data curation, Software, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Bornali Bhattacharjee:** Conceptualization, Supervision, Software, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Shrinka Sen:** Conceptualization, Data curation, Software, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Amrapali Bhattacharya:** Investigation, Methodology, Validation, Visualization, Writing – review & editing. **Sweta Sharma Saha:** Investigation, Methodology, Validation, Visualization, Writing – review & editing. **Rahul Roy Chowdhury:** Resources, Validation, Visualization, Writing – review & editing. **Nidhu Ranjan Mondal:** Resources, Validation, Visualization, Writing – review & editing. **Biman Chakrabarty:** Resources, Validation, Visualization, Writing – review & editing. **Tanmay Chatterjee:** Resources, Validation, Visualization, Writing – review & editing. **Sudipta Roy:** Resources, Validation, Visualization, Writing – review & editing. **Sharmila Sengupta:** Conceptualization, Supervision, Funding acquisition, Project administration, Methodology, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

We thank Saroj Gupta Cancer center and Research Institute (SGCC & RI, Thakurpukur, South 24 Parganas, West Bengal, India), Child in Need Institute (Pailan, South 24 Parganas, West Bengal, India), Leirik Memorial Hospital (Imphal, Manipur, India), Sir Thutob Namgyal Memorial Hospital (Gangtok, Sikkim, India), Vivekananda Tribal Hospital (Jagdalpur and at Katekalyan, Chattisgarh, India), Medical College and Hospital (Kolkata, West Bengal, India) and College of Medicine and JNM Hospital (Kalyani, Nadia, West Bengal, India) for their support in sample collection; Dr. Neeta Sarkar Roy (National Institute of Biomedical Genomics, Kalyani), Dr. Damayanti Das for her support through this work; all members of National Institute of Biomedical Genomics, Kalyani, India for their technical support during the work; National Institute of Biomedical Genomics, Kalyani, India for funding support; and special thanks also to Council of Scientific and Industrial Research, India and University Grant Commission, India for providing Fellowships to Dr. Paramita Mandal (JRF and SRF) and Dr. Shrinka Sen (JRF and SRF) to work on this project.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101256.

References

[1] International Collaboration of Epidemiological Studies of Cervical Cancer, Comparison of risk factors for invasive squamous cell carcinoma and adenocarcinoma of the cervix: collaborative reanalysis of individual data on 8097 women with squamous cell carcinoma and 1,374 women with adenocarcinoma from 12 epidemiological studies, Int. J. Cancer 120 (4) (2007) 885–891.
[2] E.M. Burd, Human papillomavirus and cervical cancer, Clin. Microbiol. Rev. 16 (1) (2003) 1–17.
[3] P. Guan, R. Howell-Jones, N. Li, L. Bruni, S. de Sanjosé, S. Franceschi, G. M. Clifford, Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer, Int. J. Cancer 131 (2012) 2349–2359.
[4] S. De Sanjose, W.G. Quint, L. Alemany, D.T. Geraets, J.E. Klaustermeier, B. Lloveras, S. Tous, A. Felix, L.E. Bravo, H.R. Shin, et al., Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, Lancet Oncol. 11 (2010) 1048–1056.
[5] S. Taylor, E. Bunge, M. Bakker, X. Castellsagué, The incidence, clearance and persistence of non-cervical human papillomavirus infections: A systematic review of the literature, BMC Infect. Dis. 16 (2016) 293.
[6] N. Munoz, F.X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, et al., Epidemiologic classification of human papillomavirus types associated with cervical cancer, N. Engl. J. Med. 348 (2003) 518–527.
[7] L.L Villa, L. Sichero, P. Rahal, O. Caballero, A. Ferenczy, et al., Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia, J. Gen. Virol. 81 (2000) 2959–2968.
[8] B.C. Das, V. Gopalkrishna, S. Hedau, S. Katiyar, Cancer of the uterine cervix and human papillomavirus infection, Curr. Sci. 78 (2000) 52–63.
[9] A Sowjanya, M. Jain, U.R. Poli, S. Padma, M. Das, et al., Prevalence and distribution of high-risk humanpapillomavirus (HPV) types in invasive squamous cell carcinoma of the cervix andin normal women in Andhra Pradesh, India, BMC Infect. Dis. 5 (2005) 116.
[10] R.K. Singh, S. Maulik, S. Mitra, R.K. Mondal, P.S. Basu, et al., Human papillomavirus prevalence in postradiotherapy uterine cervical carcinoma patients: correlation with recurrence of the disease, Int. J. Gynecol. Cancer 6 (2006) 1048–1054.
[11] S. Franceschi, R. Rajkumar, P.J.F. Snijders, A. Arslan, C. Mahé, et al., Papillomavirus infection in rural women in southern India, Br. J. Cancer 92 (2005) 601–606.
[12] P. Laikangbam, S. Sengupta, P. Bhattacharya, C. Duttagupta, D. Singh, et al., A comparative profile of the prevalence and age distribution of human papillomavirus type 16/18 infections among three states of India with focus on northeast India, Int. J. Gynecol. Cancer 17 (2007) 107–117.
[13] A. Bhattacharya, S. Sen, P. Mandal, S. Sharma Saha, S. Sarkar, O.P. Pathak, L. Biswas, J. Roy, R. Banerjee, R. Roy Chowdhury, M. Pal, A. Mukherjee, S. Sengupta, Prevalence and age-wise distribution of Human Papillomavirus type 16/18 infections among hospital screened women of a peri-urban area in West Bengal: impact of socio-demographic factors, Cancer Epidemiol. 54 (2018) 31–37.
[14] B. Bhattacharjee, S. Sengupta, HPV16 E2 gene disruption and polymorphisms of E2 and LCR: some significant associations with cervical cancer in Indian women, Gynecol. Oncol. 100 (2006) 372–378.
[15] D. Das Ghosh, B. Bhattacharjee, S. Sen, L. Premi, I. Mukhopadhyay, et al., Some novel insights on HPV16 related cervical cancer pathogenesis based on analyses of LCR methylation, viral load, E7 and E2/E4 expressions, PLoS One 7 (9) (2012) e44678.
[16] N. Sathish, P. Abraham, A. Peedicayil, G. Sridharan, G. Chandy, HPV 16 E6 sequence variations in Indian patients with cervical neoplasia, .Cancer Lett. 229 (2005) 93–99.
[17] M. Sun, L. Gao, Y. Liu, Y. Zhao, X. Wang, et al., Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China, PLoS One 7 (5) (2012) e36577.
[18] I. Cornet, T. Gheit, M.R. Iannacone, J. Vignat, B.S. Sylla, et al., HPV16 genetic variation and the development of cervical cancer worldwide, Br. J. Cancer 108 (2013) 240–244.
[19] T. Gheit, I. Cornet, G.M. Clifford, T. Iftner, C. Munk, et al., Risks for persistence and progression by human papillomavirus type 16 variant lineages among a population-based sample of Danish women, Cancer Epidemiol. Biomark. 20 (2011) 1315–1321.
[20] I. Cornet, T. Gheit, S. Franceschi, J. Vignat, R.D. Burk, et al., Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR, J. Virol. 86 (12) (2012) 6855–6861.

[21] T. Yamada, M.M. Manos, J. Peto, C.E. Greer, N. Munoz, et al., Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective, J. Virol. 71 (3) (1997) 2463–2472.
[22] H.U. Bernard, The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses, J. Clin. Virol. 32 (1) (2005) S1–S6. Suppl.
[23] H.U. Bernard, I.E. Calleja-Macias, S.T. Dunn, Genome variation of human papillomavirus types: phylogenetic and medical implications, Int. J. Cancer 118 (5) (2006) 1071–1076.
[24] S.Y. Chan, L. Ho, C.K. Ong, V. Chow, B. Drescher, et al., Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind, J. Virol. 66 (1992) 2057–2066.
[25] L. Ho, S.Y. Chan, V. Chow, T. Chong, S.K. Tay, et al., Sequence variants of human papillomavirus type 16 in clinical samples permit veri®cation and extension of epidemiological studies and construction of a phylogenetic tree, J. Clin. Microbiol. 29 (1991) 1765–1772.
[26] L. Ho, S.Y. Chan, R.D. Burk, B.C. Das, K. Fujinaga, 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. 67 (1993) 6413–6423.
[27] T. Yamada, C.M. Wheeler, A.L. Halpern, A.C.M. Stewart, A. Hildesheim, et al., Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments, J. Virol. 69 (1995) 7743–7753.
[28] R.D. Burk, A. Harari, Z. Chen, Human papillomavirus genome variants, Virology 445 (1-2) (2013) 232–243.
[29] L. Mirabello, M. Yeager, M. Cullen, J.F. Boland, Z. Chen, N. Wentzensen, X. Zhang, K. Yu, Q. Yang, J. Mitchell, D. Roberson, S. Bass, Y. Xiao, L. Burdett, T. Raine-Bennett, T. Lorey, P.E. Castle, R.D. Burk, M. Schiffman, HPV16 sublineage associations with histology-specific cancer risk using HPV whole-genome sequences in 3200 women, J. Natl. Cancer Inst. 108 (9) (2016) djw100.
[30] B. Bhattacharjee, N.R. Mandal, S. Roy, S. Sengupta, Characterisation of sequence variations within HPV16 isolates among Indian women: prediction of causal role of rare non-synonymous variations within intact isolates in cervical cancer pathogenesis, Virology 377 (2008) 143–150.
[31] P. Mandal, B. Bhattacharjee, D. Das Ghosh, N.R. Mondal, R. Roy Chowdhury, et al., Differential expression of HPV16 L2 gene in cervical cancers harboring episomal HPV16 genomes: influence of synonymous and non-coding region variations, PLoS One 8 (6) (2013) e65647.
[32] L. Mirabello, M. Yeager, K Yu, HPV16 E7 genetic conservation is critical to carcinogenesis, Cell 170 (6) (2017) 1164–1174. .e6.
[33] A. Vats, O. Trejo-Cerro, M. Thomas, L. Banks, Human papillomavirus E6 and E7: what remains? Tumour Virus Res. 11 (2021), 200213.
[34] The Cancer Genome Atlas Research. Integrated genomic and molecular characterization of cervical cancer, Nature 543 (2017) 378–384.
[35] P. Bhattacharya, S. Sengupta, Predisposition to HPV16/18-related cervical cancer because of proline homozygosity at codon 72 of p53 among Indian women is influenced by HLA-B*07 and homozygosity of HLA-DQB1*03, Tissue Antigens 70 (2007) 283–293.
[36] P. Mandal, B. Bhattacharjee, S. Sen, A. Bhattacharya, R. Roy Chowdhury, N. R. Mondal, S. Sengupta, Complete genome sequences of eight human papillomavirus type 16 Asian American and European variant isolates from cervical biopsies and lesions in Indian women, Genome Announc. 4 (3) (2016), e0024316.
[37] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.
[38] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol 16 (1999) 37–48.
[39] D. Das, B. Bhattacharjee, S. Sen, I. Mukhopadhyay, S. Sengupta, Association of viral load with HPV16 positive cervical cancer pathogenesis: Causal relevance in isolates harboring intact viral E2 gene, Virology 402 (2010) 197–202.
[40] Y Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B (Methodol.) 57 (1995) 289–300.
[41] I. Zehbe, G. Voglino, E. Wilander, H. Delius, A. Marongiu, et al., p53 codon 72 polymorphism and various human papillomavirus 16 E6 genotypes are risk factor for cervical cancer development, Cancer Res. 61 (2001) 608–611.
[42] M. Grodzki, G. Besson, C. Clavel, S. Franceschi, A. Arslan, et al., Increased risk for cervical disease progression of French women infected with the human papillomavirus 16 E6-350 G variant, Cancer Epidemiol. Biomark. Prev. 15 (2006) 820–822.
[43] S. Nicolás-Párraga, C. Gandini, V.N. Pimenoff, et al., HPV16 variants distribution in invasive cancers of the cervix, vulva, vagina, penis, and anus, Cancer Med. 5 (10) (2016) 2909–2919.
[44] E.A. White, K. Munger, Crowd control: E7 conservation is the key to cancer, Cell 170 (6) (2017) 1057–1059.
[45] X. Zhe, H. Xin, Z. Pan, et al., Genetic variations in E6, E7 and the long control region of human papillomavirus type 16 among patients with cervical lesions in Xinjiang, China, Cancer Cell Int. 19 (2019) 65.
[46] M.R. Pillai, R. Hariharan, J.M. Babu, S. Lakshmi, S.V. Chiplunkar, et al., Molecular variants of HPV-16 associated with cervical cancer in Indian population, Int. J. Cancer 125 (2009) 91–103.
[47] S. Pande, N. Jain, B.K. Prusty, S. Bhambhani, S. Gupta, et al., HPV-16 variant analysis of E6, E7, L1 genes and long control region in cervical cancer patients from North India, J. Clin. Microbiol. 46 (2008) 1060–1066.
[48] P.M. Sharp, P. Simmonds, Evaluation of the evidence for virus host coevolution, Curr. Opin. Virol. 1 (2011) 436–441.
[49] L.F. Xi, et al., Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia, J. Natl. Cancer Inst. 89 (1997) 796–802.
[50] J. Berumen, R.M. Ordonez, E. Lazcano, J. Salmeron, S.C. Galvan, R.A. Estrada, E. Yunes, A. Garcia-Carranca, G. Gonzalez-Lira, A. Madrigal-de la Campa, Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study, J. Natl. Cancer Inst. 93 (2001) 1325–1330.
[51] R.M. Ordóñez, A.M. Espinosa, D.J. Sánchez-González, J. Armendáriz-Borunda, J. Berumen, Enhanced oncogenicity of Asian-American human papillomavirus 16 is associated with impaired E2 repression of E6/E7 oncogene transcription, J. Gen. Virol. 85 (2004) 1433–1444.
[52] L. Casas, S.C. Galvan, R.M. Ordonez, N. Lopez, M. Guido, et al., Asian-American variants of human papillomavirus type 16 have extensive mutations in the E2 gene and are highly amplified in cervical carcinomas, Int. J. Cancer 83 (1999) 449–455.
[53] L. Sichero, S. Ferreira, H. Trottier, E. Duarte-Franco, A. Ferenczy, et al., High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18, Int. J. Cancer 120 (2007) 1763–1768.
[54] L.F. Xi, L.A. Koutsky, A. Hildesheim, D.A. Galloway, C.M. Wheeler, et al., Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18, Cancer Epidemiol. Biomark. Prev. 16 (2007) 4–10.
[55] P. Moreno-Acosta, A. Vallard, M. Molano, et al., HPV-16 variants' impact on uterine cervical cancer response to radiotherapy: a descriptive pilot study, Cancer Radiother. 21 (2) (2017) 104–108.
[56] J.S. Rader, S.W. Tsaih, D. Fullin, et al., Genetic variations in human papillomavirus and cervical cancer outcomes, Int. J. Cancer 144 (9) (2019) 2206–2214.
[57] E.A. Kidd, B.A. Siegel, F. Dehdashti, et al., Lymph node staging by positron emission tomography in cervical cancer: relationship to prognosis, J. Clin. Oncol. 28 (2010) 2108–2113.
[58] A.R. Garbuglia, D. Lapa, C. Sias, M.R. Capobianchi, P. Del Porto, The use of both therapeutic and prophylactic vaccines in the therapy of papillomavirus disease, Front. Immunol. 11 (2020) 188.
[59] A. Kumar, S. Hussain, G. Sharma, R. Mehrotra, L. Gissmann, B.C. Das, M. Bharadwaj, Identification and validation of immunogenic potential of India specific HPV-16 variant constructs: *in-silico & in-vivo* insight to vaccine development, Sci. Rep. 5 (2015) 15751.
[60] Y. Liu, Y. Pan, W. Gao, Y. Ke, Z. Lu, Whole-genome analysis of human papillomavirus types 16, 18, and 58 isolated from cervical precancer and cancer samples in chinese women, Sci. Rep. 7 (1) (2017) 263.