REVIEW OPEN

Check for updates

Human papillomavirus 16 genomic variation in head and neck cancer: a systematic review with structural and functional annotation

Ngoc Ha Tran[1,7], Huy Gia Nguyen[2,7], Dayna Sais[1], Nicole Lima[3], Alison Todd[3], Tuan Van Nguyen[1,4], Michael Elliott[5,6] and Nham Tran[1]✉

© The Author(s) 2026

Human papillomavirus type 16 (HPV16)-positive head and neck cancer (HNC) is generally associated with a more favourable prognosis than HPV-negative disease. However, emerging evidence suggests that specific HPV16 genomic variants may be associated with poorer clinical outcomes. We therefore conducted a systematic review of HPV16 single nucleotide polymorphisms (SNPs) reported in HNC and mapped selected variants to viral genomic and protein regions to assess their potential relevance. We searched four major databases (PubMed, Embase, Scopus, and Web of Science) for studies reporting HPV16 SNPs in HNC. Most studies focused on either E6 or E7, while only a few studies examined the entire HPV16 sequence. These SNPs were extracted and mapped to the HPV16 genome, while structural and functional characteristics were assessed using UniProt and the RCSB Protein Data Bank. Fourteen studies were analysed, identifying 30 SNPs in E6 and seven in E7. Structural mapping suggested that core HPV16 oncogenic motifs remain highly conserved, while recurrent variants tend to occur in mutationally permissive regions, such as the E6 interdomain linker. Only one study reported associations between HPV16 variants and poorer patient survival within the upstream regulatory region. However, these findings require independent validation before any clinical relevance can be established. The comprehensive mapping of HPV16 SNPs in HNC provides an important grounding for understanding how these SNPs may be associated with disease progression and patient outcomes. Although many of the identified SNPs are within important HPV16 regions, their associations with disease progression remain unclear. Further studies are required to determine which variants are associated with patient outcomes. A more comprehensive understanding of HPV16 genomic variation and its relevance could provide broader clinical insights across HPV16-associated cancers.

BJC Reports; https://doi.org/10.1038/s44276-026-00251-9

INTRODUCTION

Head and Neck cancer (HNC) represents a group of malignant tumors affecting the upper region of the respiratory and digestive tracts [1] with squamous cell carcinoma being the primary pathology (more than 90%) [2]. According to Global Cancer Statistics 2022 [3], head and neck cancer ranks as the sixth most common cancer worldwide with over 891,453 new cases and 458,107 deaths. Head and neck cancer is associated with well-established risk factors, including tobacco use, alcohol consumption, and infection with human papillomavirus (HPV) [4]. Among these, HPV16 accounts for more than 90% of HPV-driven head and neck cancer cases [5] and is prevalent in oropharyngeal squamous cell carcinoma (OPSCC), where it has been detected in ~72% of tumour specimens in case-control studies [6]. A systematic review reported that HPV is attributable to 47.2% of base of tongue cancers and 66.4% of tonsillar cancers, the two main types of oropharyngeal cancer, respectively [7]. The prevalence of HPV in OPSCC has steadily increased over time, rising from 40.5% before 2000 to 72.2% between 2005 and 2009, while non-oropharyngeal HNCs showed no significant change [8].

Despite the inclusion of HPV16 in the prophylactic HPV vaccine, Gardasil, since 2006, the global burden of HPV16-driven HNC has continued to rise and is anticipated to increase by 30% by 2030 [3, 9]. Since radiotherapy is the first choice of treatment for HNC, patients can develop post-treatment complications and long-term side effects including dermatitis, dry mouth, hearing impairment, osteonecrosis of the jaw, or paresthesia [10]. Accurate detection of HPV16 in HNC is important, as patients with HPV-positive tumors exhibit significantly higher overall survival than HPV-negative patients [11]. However, the pathogenic and carcinogenic mechanisms driven by HPV16 in HNC do require further investigation. Two key viral oncogenes, E6 and E7, target tumor suppressor proteins p53 and retinoblastoma (Rb), respectively, disrupting cell cycle regulation and apoptosis, and promoting chromosomal instability [12, 13]. Other important regions of the HPV genome, such as L1 and L2, contribute to viral capsid formation and infectivity, while

[1]Discipline of Biomedical Engineering, University of Technology Sydney, Sydney, NSW, Australia. [2]Bone Epidemiology and Clinical Research, Garvan Institute of Medical Research, Sydney, NSW, Australia. [3]SpeeDx Pty Ltd, Eveleigh, Sydney, NSW, Australia. [4]School of Population Health, University of New South Wales, Sydney, NSW, Australia. [5]Sydney Medical School, University of Sydney, Sydney, NSW, Australia. [6]Chris O'Brien Lifehouse, Sydney, NSW, Australia. [7]These authors contributed equally: Ngoc Ha Tran, Huy Gia Nguyen. ✉email: nham.tran@uts.edu.au

Received: 9 June 2026 Revised: 7 August 2026 Accepted: 7 September 2026
Published online: 09 November 2026

the upstream regulatory region (URR) modulates viral gene expression and replication [13].

Recent findings suggest that high-risk single nucleotide polymorphisms (SNPs) in HPV16 are associated with poorer prognosis in oropharyngeal cancer [14]. This study used whole-genome sequencing and found patients carrying at least one of eight deleterious high-risk SNPs had a median overall survival of 3.96 years, compared with 18.67 years for patients without these variants. This observation may be explained by the fact that the viral genome is not static but continuously evolves through natural selection [15]. As an oncogenic virus, HPV16 may accumulate or retain specific high-risk variants that enhance its ability to persist, invade host tissues, and contribute to more aggressive disease [16]. These findings highlight the prognostic significance of HPV16 genomic variation and its potential to modulate HNC outcomes.

While HPV16 mutational profiling has been extensively studied in cervical cancer, data on HNC remains limited. Existing studies report heterogeneous distributions of viral variants; however, their functional relevance and clinical impact remain poorly defined. In HPV16, mutation occurrences and their effects on protein function are influenced by several factors, including amino acid substitution similarity [17, 18], residue position within core or non-core regions [19], solvent accessibility (exposed versus buried residues) [20–23], protein structural order or disorder [24], and natural purifying selection [25].

To our knowledge, no study has systematically investigated which genomic regions are permissive to mutation, how variants are constrained by viral oncogenic requirements, or which SNPs structurally may affect host-virus interactions in HNC. Therefore, this review aims to summarise reported HPV16 SNPs in head and neck cancer and map them onto the viral genome for functional characterisation.

METHODS

Search strategy and selection criteria

We conducted this systematic review following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations [26]. Four databases, including PubMed, Embase, Scopus, and Web of Science, were used to search for studies published up to September 14, 2025, and restricted in English language. We targeted studies investigating HPV infections and HPV-related SNPs identified in human samples from individuals with HNC. Studies were eligible if HPV-associated disease was assessed using PCR-based assays, sequencing, in situ hybridisation (ISH), or p16 immunohistochemistry (IHC), and if HPV16-specific variant data were reported. These methods are not equivalent, and p16 immunohistochemistry was considered a surrogate marker of HPV-associated disease rather than a genotype-specific measure of HPV16 infection. The complete search strategy is provided in the Supplementary Table 1. No review protocol was registered prior to conducting this review.

After removing duplicates, we screened titles and abstracts independently. Exclusion criteria included studies that were not original research, used cell lines, lacked HPV SNP data, and did not report the study population. Eligible studies were reviewed in full and further evaluated by the same reviewers. Additional relevant studies were identified through manual screening of the reference lists of articles included and relevant systematic reviews. No formal meta-analysis was performed; rather, findings were summarised narratively and presented in a structured format.

Data extraction

Data were independently extracted using a standardised form. Information collected from each study included the first author, article title, year of publication, country, sample type, total number of participants, number of HPV-positive individuals, number of HPV16-positive individuals, sample site, median age, number of HPV16-positive samples analyzed for SNPs, reference genome used, targeted regions, and SNP positions along with their reported patients.

To ensure consistency across studies using different HPV16 reference genomes, all SNP positions were manually aligned to the primary reference genome K02718. Amino acid changes were also determined based on K02718, and synonymous changes were excluded. Study-specific SNP frequency was expressed as n/N (%), where n represents the number of samples carrying the variant and N represents the number successfully assessed for the relevant gene, genomic region or predefined SNP. The denominator was not assumed to represent all enrolled, HPV-positive or HPV16-positive participants unless the original study confirmed that all such samples were assessed at the relevant locus. Where an explicit SNP-specific numerator or denominator was unavailable, the source-reported frequency was retained and n/N was recorded as not reported.

Synonymous coding variants were excluded because this review focused on the structural annotation and protein-level interpretation of amino-acid substitutions. When different nucleotide changes produced the same amino-acid substitution, they were counted as a single protein-level variant for descriptive counting and structural interpretation, while the individual nucleotide changes remained reported in the supplementary data. This exclusion applied only to synonymous variants within coding regions; variants in non-coding regions, including the upstream regulatory region (URR), were retained and analysed separately.

Structural and functional analysis of residues changed by SNPs

For this analysis, we accessed the sequence and structural data for E6 and E7 from the Universal Protein Resource (UniProt, E6: P03126 and E7: P03129) and the RCSB Protein Data Bank (RCSB PDB, E6: 4XR8 and E7: 2B9D, 9UUJ) to map residues affected by SNPs to their positions within functional regions. Structural information for the E6 protein was further used to assess whether the identified residues may influence E6AP and p53 binding within the HPV16 E6/E6AP/p53 ternary complex. We evaluated structural factors including residue similarity, core versus non-core subregions, ordered versus disordered regions, and exposed versus buried residues. At the residue level, we compared the physicochemical similarity of amino acid substitutions caused by SNPs based on the classification framework for the 20 amino acids described in Lehninger Principles of Biochemistry [27].

Quality assessment

The methodological quality of the included studies was assessed using the JBI Critical Appraisal Checklist for Case Series [28]. This checklist consists of ten items evaluating key domains of study quality. Each item was rated as 'Yes', 'No', 'Unclear', or 'Not applicable'. Three reviewers (NHT, HGN and DS) independently assessed the risk of bias in each study. The overall risk of bias was categorised as low, moderate, or high based on the number of items rated as 'Yes'. Discrepancies were resolved through discussion.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

Database search results

In total, we identified 10,844 records from four databases (Pubmed, Embase, Scopus, and Web of Science) and four records

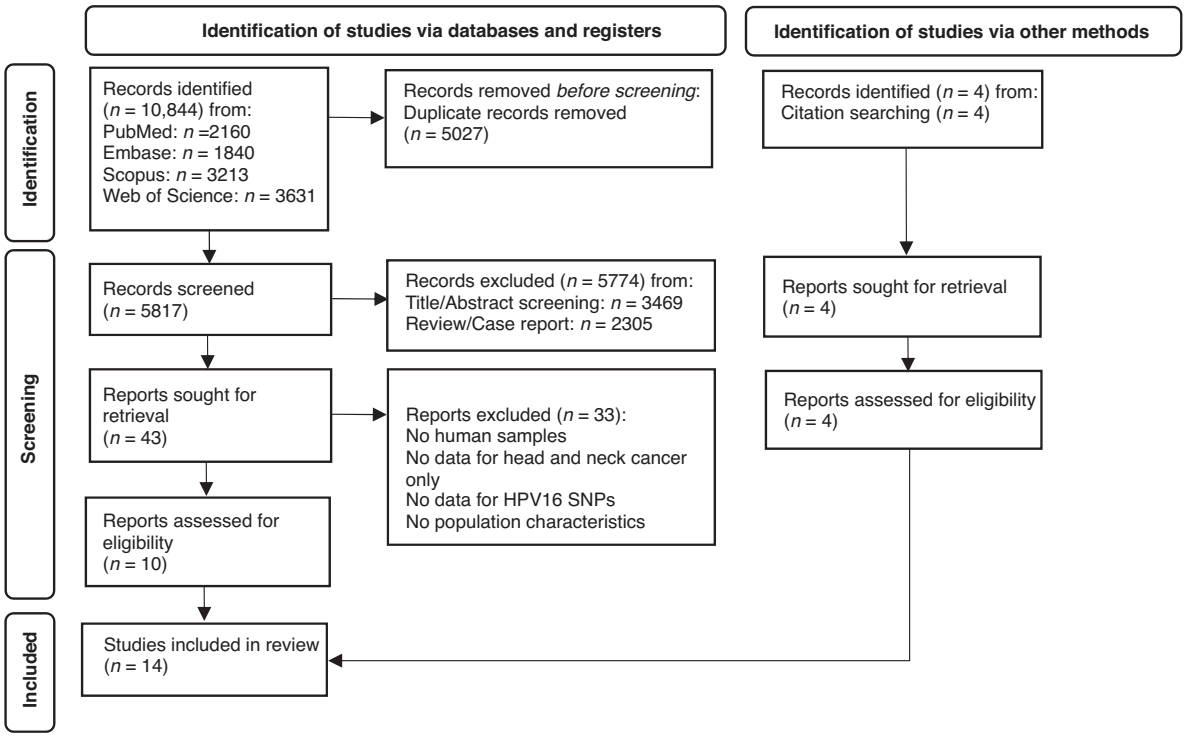


Fig. 1 Study Overview. A flow diagram showing the search strategy and selection criteria.

from manual searching. From this, 5027 duplicates were removed, 5817 records were screened by title and abstract, of which 5774 were excluded at the title/abstract stage. The remaining 43 reports were sought for full-text retrieval, and 33 of these were excluded for specific, stated reasons (no human samples; no data specific to head and neck cancer; no HPV16 SNP data; or no population characteristics reported), leaving 10 eligible studies. These 10 studies, combined with the 4 studies identified through citation searching, yielded the final 14 studies included that met our inclusion criteria (Fig. 1).

Quality assessment

Of the 14 HNC studies, 64.3% (9/14) were rated as having a low risk of bias, 28.6% (4/14) had some concerns, and 7.1% (1/14) was rated as having a high risk of bias. The main sources of bias included incomplete reporting of participant demographics and clinical information, as well as missing clear outcomes or follow-up data. A detailed breakdown of the quality assessment for each study using the JBI checklist is included in Supplementary Table 2.

Study characteristics

Study characteristics are summarised in Table 1 for HNC. The fourteen studies were published between 2000 and 2023 [14, 29–41]. Solid tissue was the primary sample type in 13 studies, while oral rinse was used in one. While two studies assessed survival outcomes, only one comprehensive whole-genome study identified specific high-risk SNPs associated with poorer prognosis. The reported median ages in the 14 studies ranged from 57 to 67.5 years. Four primary tumor sites were reported, including the oral cavity, oropharynx, hypopharynx, and larynx. Among these, the oropharynx was the most common tumor site mentioned in 13 of the 14 studies.

HPV16 positivity and detection methods across included studies

A total of 1578 patients were enrolled (Table 1). HPV16 was the most common type, affecting 1036 individuals. Six studies (Yoshida et al. [41], Kottaridi et al. [39], Joseph et al. [38], Hoffmann et al. [37], Du et al. [33], and LeConte et al. [40] only evaluated patient samples that were HPV positive, with HPV16 comprising 91.2–100% of cases. In the other eight studies, HPV status was not confirmed at enrollment; however, 57% of patients were later identified as HPV-positive, with HPV16 remaining the predominant type at 97.7%. Eleven studies used PCR to detect HPV and HPV16 positivity, except for one study, which used in situ hybridization [38]; another study used sequencing-based method [30], while a single study used PCR/RNA ISH, followed by HPV16 whole-genome sequencing [14] (Supplementary Table 3). These approaches measure different aspects of HPV infection and are not directly interchangeable, particularly when identifying transcriptionally active HPV-driven disease

Number of SNPs across all HPV16 regions. The primary GenBank reference sequences used were K02718 (seven studies), NC_001526 (three studies), and AB889493 (one study) (Supplementary Table 3). Three studies did not report the GenBank reference sequences. The targeted regions within the HPV16 genome varied across studies. With respect to HPV16 E6 and E7 oncogenes, 13 (including one study [14] sequenced whole HPV16 genome) of the 14 studies specifically targeted E6 and/or E7 regions. Among these, nine studies focused exclusively on E6, no studies targeted E7 alone, and three investigated both E6 and E7.

One study investigated only the L1 region, while another identified SNPs in L1 alongside E6 and E7; the whole-genome sequencing study also included L1. There was also variation in study objectives (Supplementary Table 3): seven studies aimed to identify all potential SNPs within the target regions, whereas eight focused on specific variants, including E6 SNPs A131G and T350G, E7 SNPs A645C, A647G, and C712A, L1 SNPs A6693C and G6720A.

Regarding SNP identification, of the 1,578 patients initially enrolled in 14 studies, approximately 971 were involved in HPV16 SNP identification (Supplementary Table 3). As noted, most studies focused on the E6 and E7 regions (13 out of 14 studies); we divided the identified SNPs into two groups: those located within E6/E7 and those outside these regions.

Table 1. Study-level characteristics of the studies included in the systematic review.

Study	Year	Country/ region	Specimen type	Study-level cohort	Cohort definition and HPV status	HPV16-positive, n	Samples assessed for HPV16 variation	Tumour site(s)
Gillison et al.	2000	USA	Tumour tissue	253 HNC cases	Unselected cohort: 62 HPV-positive; 56 HPV16-positive	56	52 HPV16-positive tumours with E6 sequence data	Oral cavity, larynx, hypopharynx and oropharynx
Hoffmann et al.	2004	Germany	Tumour tissue	24 patients	HPV16-positive HNSCC cohort	24	21 with complete E6/E7 genotype data	Mixed HNC sites, including oropharyngeal and non-oropharyngeal sites
Agrawal et al.	2008	USA	Oral rinse	135 HNC patients	44 HPV16-positive and 91 HPV16-negative tumour cases; oral-rinse specimens assessed longitudinally	44 tumour cases	22 HPV16-positive oral-rinse specimens; E6 successfully sequenced in 19	Mixed HNC sites; 41 of 44 HPV16-positive cases were oropharyngeal and 3 were of unknown site
Boscolo-Rizzo et al.	2009	Italy	Tumour tissue	77 upper aerodigestive tract SCC cases	Unselected cohort: 8 HPV16-positive cases	8	8 assessed for E6 T350G/ L83V	Oral cavity, larynx, hypopharynx and oropharynx
Castillo et al.	2011	Japan, Pakistan and Colombia	Tumour tissue	71 oral-cavity SCC cases	HNC-only cohort after exclusion of oesophageal cancers: 40 HPV-positive and 39 HPV16-positive	39	23 oral-cancer cases with E6 variant assignments	Oral cavity
Du et al.	2012	Sweden	Tumour tissue	108 tonsillar cancers	HPV16-positive tonsillar cancer cohort	108	108 genotyped for predefined E6 variants A131G and T350G	Tonsil/oropharynx
Joseph et al.	2013	USA	Tumour tissue	4 patients with paired tumours	Four patients with bilateral HPV-associated tonsillar carcinomas	4	4 patients assessed using HPV16 E6 sequencing across paired tumours	Tonsil/oropharynx
Barbieri et al.	2014	Italy	Tumour tissue	81 OPSCC biopsies	60 HPV-positive; 51 HPV16-positive	51	51 HPV16-positive samples assessed for selected L1 variants	Oropharynx
Hassani et al.	2015	Japan	Tumour tissue	24 tonsillar SCC cases	Unselected cohort: 10 HPV16-positive	10	10 HPV16-positive tumours assessed for E6 variation	Tonsil/oropharynx
LeConte et al.	2018	USA	Tumour tissue	137 HPV-positive OPSCC tumours	HPV-positive oropharyngeal cohort; 125 were HPV16-positive	125	92 complete E6 sequences and 89 complete E7 sequences	Oropharynx
Yoshida et al.	2022	Japan	Tumour tissue	70 HPV-related OPSCC cases	HPV16 sequences obtained from 70 HPV-related OPC cases	70	70 assessed across E6, E7 and selected L1 regions	Oropharynx
Kottaridi et al.	2022	Greece	Tumour tissue	40 tumour samples	HPV16-positive oropharyngeal cancer cohort	40	40 assessed for HPV16 E6 variation	Oropharynx
Lang Kuhs et al.	2022	USA	Tumour tissue	460 OPSCC specimens	Specimens underwent HPV16 whole-genome sequencing; 76 failed sequencing quality-control criteria and should not be classified as HPV-negative	NR at pre-QC level	384 QC-passing HPV16 genomes with sufficient sequencing depth and coverage	Oropharynx
Gameiro et al.	2023	Canada and USA	Tumour tissue	161 HPV16-positive HNC samples with variant calls	Combined North American cohort; clinical data were available for 149 cases	94	161 with variant-lineage calls; 98 evaluable for T350G in the combined cohort	Mixed HNC sites; predominantly oropharyngeal

Studies are presented in chronological order. Study-level cohort sizes, HPV-status counts and tumour-site distributions are reported according to the definitions used in the original publications. Some studies enrolled unselected head and neck cancer cohorts, whereas others included only HPV-positive, HPV16-positive or sequencing-evaluable cases. The number of samples assessed for HPV16 variation may therefore differ from the total study cohort or the number of HPV16-positive participants. Anatomical-site counts refer to HPV-positive or HPV16-positive cases, depending on the data reported by each study. NR indicates not reported. *HNC* head and neck cancer, *HNSCC* head and neck squamous cell carcinoma, *HPV* human papillomavirus, *HPV16* human papillomavirus type 16, *OPSCC* oropharyngeal squamous cell carcinoma, *OR* oral rinse, *QC* quality control, *S* solid tumour tissue, *SCC* squamous cell carcinoma.

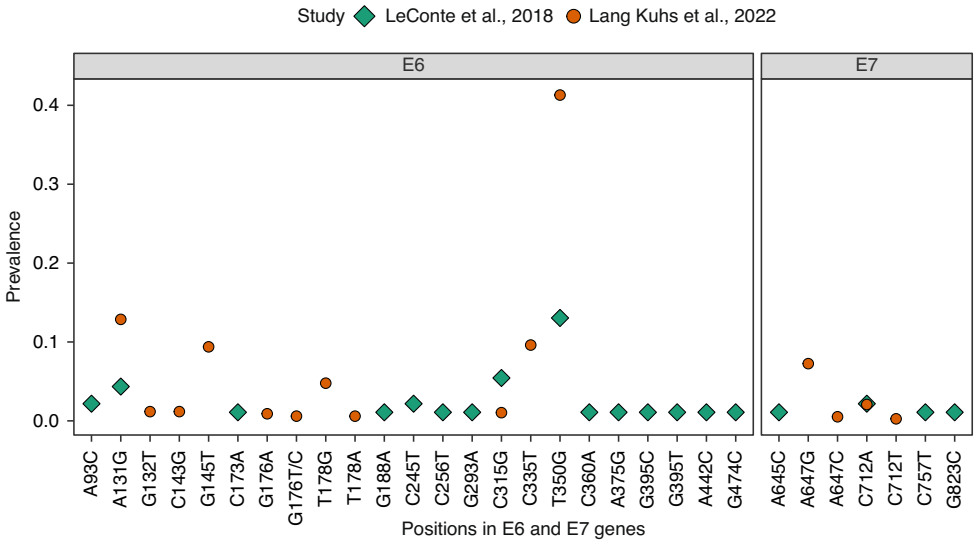


Fig. 2 Comparison of SNP prevalence across HPV16 E6 and E7 regions. Each point represents an individual SNP. For LeConte et al., frequencies are based on gene-specific denominators of 92 complete E6 sequences and 89 complete E7 sequences. For Lang Kuhs et al., values represent source-reported minor allele frequencies among 384 quality-controlled HPV16 genomes. As these studies used different frequency measures and denominators, the values are descriptive.

Across the included studies, 30 non-synonymous SNPs were reported in E6 and seven in E7. However, these counts are subject to ascertainment bias because E6 was investigated more frequently than E7: nine studies focused exclusively on E6, none focused exclusively on E7, and only a limited number assessed both regions. Among the two studies that provided data suitable for direct comparison of both regions (Lang Kuhs et al. [14] and LeConte et al. [40], 23 SNPs were reported in E6 and seven in E7 (Fig. 2 and Supplementary Table 4). This pattern is descriptive and does not establish that E6 is intrinsically more variable than E7.

Among all the reported SNPs, the variants A131G, G145T, T178G, C315G, C335T, T350G, C143G, G176A, C256T and G395C in E6, as well as A645C, A647G, and C712A in E7, were each reported in at least two studies (see Supplementary Table 3). As seen in Fig. 2, the two studies which investigated both E6 and E7 regions identified A131G, C315G, and T350G in E6, along with C712A in E7, highlighting a gap in variant detection among remaining studies.

Since HPV16-driven carcinogenesis is largely mediated by the E6 and E7 oncoproteins through inactivation of p53 and pRB, most studies have focused on these regions, as they are considered the primary functional drivers of transformation. Consequently, potential variants outside the targeted regions, such as E1, E4, E5, L1, L2, and URR, may be overlooked, leading to an incomplete understanding of HPV16 genetic diversity. One study employed whole-genome sequencing of HPV16 in OPSCC identified 284 SNPs, thereby demonstrating the extent of variation that may be missed by region-specific sequencing [14].

Reported study-specific SNP frequencies varied substantially across the included studies (Supplementary Table 3). T350G was reported at frequencies of up to 100%, while G145T and C335T reached 50% and A131G reached 33.3%. Other E6 variants reported at frequencies of 10% or greater included G176A, T178G, T183G and T185C. Among the E7 variants, A647G and C712A reached reported frequencies of 88.5% and 33.3%, respectively. These values are not directly comparable across studies because of differences in tumour site, viral lineage composition, sequencing coverage and whether complete genomic regions or predefined SNPs were assessed. In particular, the high reported frequency of T350G should not be interpreted as a universal population-level frequency in HNC.

Only three studies investigated SNPs outside the E6 and E7 regions, including a whole-genome study [14] and two focusing on the L1 region. The SNP A5316C was reported by Yoshida et al., [41] while A6693C and G6720A were identified in the study led by Barbieri et al., [30] In contrast, the whole-genome sequencing study identified an additional 30 SNPs within the L1 region [14]. A5316C, reported by Yoshida et al. [41], had a prevalence of 40% and was the most common SNP detected in L1.

Structural annotation of E6 and E7 variants. Functional and three-dimensional structural annotations from UniProt and the RCSB Protein Data Bank (RCSB PDB), together with published structural data for E6 and E7, were used to map SNPs to their corresponding residues in the E6 and E7 proteins (Fig. 3). It is important to note that while this structural mapping provides an understanding for exploring biological plausibility, it does not establish functional effect or clinical causality.

30 SNPs in the E6 gene lead to 29 residue changes (two SNPs have the same substitution) in the protein at 24 positions (Fig. 3a). Nine residues were located within the two zinc finger domains (residues 30–66 and 103–139), including C245T(R48W), C256T(C51F), G267A(D56N), G293A(D64N), A442C(E113D), C470A(Q123K), G474C(R124T), G492A(G130D), and C506G(R135G). However, none of these residues were positioned at zinc-binding sites (residues 30, 33, 60, 66, 103, 106, 136, and 139).

The remaining 20 residues were located outside the zinc finger domains, primarily within the N-terminus (12 SNPs) or interdomain linker (8 SNPs). Notably, no SNPs were identified within the PDZ-binding motif (residues 148–151). Eight substitutions at seven positions mapped to broader E6 regions previously implicated in E6AP interaction, including A131G (R10G), G132T (R10I), C256T (C51F), C315G (S71C), C335T (H78Y), T350G (L83V), A375G (Q91R), and G492A (G130D). Eight substitutions at five positions also mapped to regions implicated in p53 interaction, including A131G (R10G), G132T (R10I), C143G (Q14E), G145T (Q14H), C173A (H24N), G395C (D98H), G395T (D98Y) and A442C (E113D). These positional associations will require further mechanistic studies to establish that the affected residues are direct contact sites or that the substitutions alter E6AP or p53 binding.

For recurrent SNPs with relatively high study-specific frequencies, we selected several candidates (A131G (R10G), G145T (Q14H), C335T (H78Y), and T350G (L83V)) to evaluate their structural context and to assess whether they follow established mutational principles. Because reporting methods varied significantly across studies, with

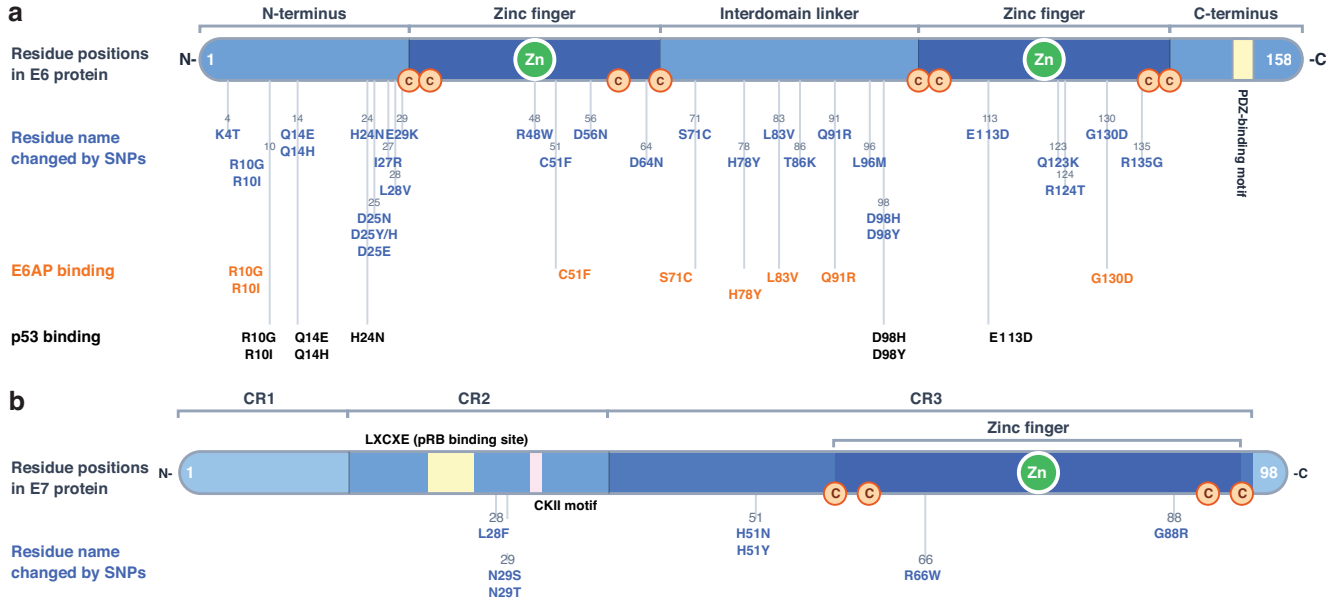


Fig. 3 Linear structural organization of the HPV16 E6 and E7 oncoproteins and mapping of SNP-associated amino acid substitutions. **a** The E6 protein (158 amino acids) is illustrated with its major structural regions, including the N-terminus, two zinc finger domains, the interdomain linker, the C-terminus, and the PDZ-binding motif. Amino acid substitutions resulting from identified SNPs are mapped to their corresponding sequence positions (blue). Specific substitutions previously reported to be involved in E6AP and p53 binding within the HPV16 E6/E6AP/p53 ternary complex are highlighted in orange and black, respectively. **b** The E7 protein (98 amino acids) is depicted with its conserved regions (CR1, CR2, and CR3) and the zinc finger domain. Key functional motifs, including the LXCXE (pRB binding site) and the CKII motif, are highlighted. Amino acid substitutions resulting from SNPs across the E7 sequence are mapped to their corresponding positions (blue). In both panels, cysteine residues essential for zinc coordination are denoted by orange circles ('C'), and coordinated zinc ions are indicated by green circles ('Zn').

some targeting the entire gene and others investigating only specific positions, these prevalence values reflect study-specific frequencies rather than absolute population parameters. The associated amino acid property changes and their corresponding protein positions are summarised in Table 2.

The E7 protein (Fig. 3b), consisting of 98 amino acids, was illustrated with its key subregions and residues, including CR1, CR2, and CR3, zinc finger domain, LXCXE motif responsible for pRB binding, and motif for casein kinase II (CKII) phosphorylation.

Seven SNPs in the E7 gene result in seven amino acid changes at five positions: 28, 29, 51, 66, and 88 (Fig. 3). Two residues (R66W and G88R) were located within the zinc finger domain (residues 58–94). Three residues, including A645C(L28F), A647G(N29S), and A647C(N29T), were mapped to conserved region 2 (CR2; residues 16–38). The remaining variants, C712A(H51N) and C712T(H51Y), were located within conserved region 3 (CR3; residues 38–95) but outside the zinc finger motif. Notably, three residues A645C(L28F), A647G(N29S), and A647C(N29T) are positioned downstream to the LXCXE motif (residues 22–26), which is thought to mediate high-affinity binding to pRB [42], and the recognition site for casein kinase II (CKII) (residues 31, 32), which contributes to E7 structural stability.

Similar to E6, two candidate SNPs in E7 (A647G (N29S) and C712A (H51N)) were selected to evaluate factors contributing to their reported prevalence and to assess whether they follow established mutational principles. The corresponding changes in amino acid properties and protein positions are summarised in Table 2.

DISCUSSION

This review indicates that the distribution of HPV16 genetic variation in head and neck cancer is not random, but it does reflect a complex interplay between the structural constraints of protein evolution [17, 18] and the functional requirement to preserve viral oncogenic fitness [19]. By mapping SNPs to functional protein regions, we have shown that substitutions predominantly occur in mutationally permissive areas such as disordered regions or non-core subregions [24, 43], while critical motifs for viral replication and carcinogenesis remain highly conserved [44]. These findings provide a pathway for identifying high-risk genomic signatures that may hold some prognostic value for HNC patients.

HPV16 infection in HNC

HPV16 was the predominant type among the studies which accounts for 91.2%-100% of all cases [45]. Oropharynx remained the primary site in 13/14 studies. The relationship between HPV16 infection and oropharynx accounts for 86.7% of all HPV16-driven HNC cases [45]. HPV16 favors this specific site due to its deep crypts shielding the virus from immune surveillance [44] and reticulated squamous epithelium lacking tight junctions allowing HPV16 easy access to basal cells [46]. At the molecular level, HPV16 evades immunity through its expression of E5 and E7, which suppresses MHC I expression [44, 47], thereby blocking antigen presentation. Additionally, E6 and E7 are able to regulated and suppress TLR9 and pro-inflammatory cytokines, impairing both innate and adaptive responses [44, 47].

Methodological variation and bias

Methodological approaches varied considerably across the included studies, particularly regarding the choice of reference sequences (predominantly K02718 or NC_001526) and the specific genomic regions analyzed. There is a focus towards the major oncogenes: 13 of the 14 studies utilised targeted sequencing of E6 and E7, with only two examining the L1 region and a single study performing whole-genome sequencing. Consequently, reported SNP frequencies should be interpreted as study-specific observations rather than definitive population-level HPV16 variant frequencies. Furthermore, the studies exhibited significant heterogeneity in their design and objectives, ranging from exhaustive

variant discovery within a specific region to targeted screening for predefined candidate SNPs. As a result, cross-study comparisons of SNP prevalence are inherently challenging and should be interpreted with caution rather than viewed as definitive epidemiological frequencies.

Among the 13 studies that analyzed specifically the E6 and/or E7 oncogenes, nine focused exclusively on E6, three identified SNPs in both E6 and E7; however, none investigated variants in E7 alone. The E6 gene is more frequently targeted for SNP identification than E7 since E6 oncoprotein plays a central role in the degradation of p53, a key tumor suppressor. Variations in E6 can alter its binding affinity to p53 or to E6AP, potentially influencing its oncogenic potential [48]. In addition, E7 is relatively conserved with low genetic variability due to its essential role in binding pRb to maintain carcinogenesis [49].

Targeted SNP sequencing or Whole-genome discovery

The included studies investigated HPV16 variants for different purposes. Some assessed candidate variants or lineage-defining markers, whereas others aimed to identify variants with potential functional or clinical relevance. These categories should not be considered equivalent. Recurrent variants such as E6 T350G/L83V and E7 A647G/N29S may reflect HPV16 lineage or sublineage composition rather than independent adaptation in HNC. Therefore, their reported frequency or recurrence alone does not establish that they are deleterious or clinically high risk.

The SNPs reported in the L1 region illustrate how differing study approaches can lead to substantially different findings. Three studies investigated L1, but with distinct objectives. A5316C, reported by Yoshida et al. [41], had a prevalence of 40% and was the most common SNP identified in that study; however, the analysis was limited to antigenic regions, restricting coverage of other L1 positions. Similarly, the study by Barbieri et al. [30], focused on identifying African lineage variants at positions 6693 and 6720 did not assess SNPs at other loci within L1. In contrast, the whole-genome sequencing study [14] provided a comprehensive assessment, identifying an additional 30 variants within the L1 region. Notably, this study reported SNPs at positions 5962 and 6025 within L1 were associated with poorer survival (HR = 4.39 and 5.71, respectively; 95% CI).

Observed variation in E6 and E7 in HNC patients

More variants were reported in E6 than in E7 across the included studies. It must be noted that E6 was more frequently investigated, no study investigated E7 alone, and only a limited number assessed both regions. The observed difference may therefore reflect selected sequencing coverage and research direction. When extracting data from the two studies that reported SNPs in both E6 and E7 simultaneously, we found 23 SNPs in E6 and 7 in E7 (Fig. 2). Our findings were consistent with previous research in cervical cancer [50], where the sequence variability in the E6 gene was higher than in the E7 gene.

Several SNPs identified in our analysis were also reported at the same positions in that cervical cancer study [50, 51], including positions 131, 145, 176, 178, 185, 188, 267, 335, 350, and 442. Among these, three E6 SNPs, A131G (R10G), G145T (Q14H), and T350G (L83V) had reported study-specific frequencies ranging from 33.3% to 100% and were reported to escape purifying selection during evolution [52].

The E7 oncogene appears to be under stronger evolutionary constraint than E6 [49], consistent with the lower number of variants observed across studies. The difference in sequence variation between E6 and E7 has been linked to APOBEC3 (Apolipoprotein B mRNA Editing Enzyme Catalytic 3) activity [53], and conversely, HPV infection has been associated with increased expression of APOBEC3 enzymes [47]. APOBEC3 proteins are innate antiviral factors that restrict viruses by deaminating cytosine to uracil in viral DNA, which is frequently detected as

Table 2. Candidate non-synonymous SNPs were mapped to structural and functional regions of the HPV16 E6 and E7 proteins.

Protein	SNP	Study	Reported frequency, *n/N* (%)	Residue change	Amino-acid feature comparison	Residue position
E6	A131G (R10G)	Hoffmann et al.	7/21 (33.3%)	Arginine to glycine	Mostly similar	N-terminal disordered region containing exposed residues
E6	G145T (Q14H)	Joseph et al.	2/4 (50.0%)	Glutamine to histidine	Similar	N-terminal disordered region containing exposed residues
E6	C335T (H78Y)	Joseph et al.	2/4 (50.0%)	Histidine to tyrosine	Different	Interdomain linker, exposed residue
E6	T350G (L83V)	Joseph et al.	4/4 (100%)	Leucine to valine	Similar	Interdomain linker, disordered and exposed residue
E7	A647G (N29S)	Yoshida et al.	62/70 (88.6%)	Asparagine to serine	Different	Outside the core functional motif and adjacent to LXCXE (residues 22-26)
E7	C712A (H51N)	Hoffmann et al.	7/21 (33.3%)	Histidine to asparagine	Different	Outside the zinc-finger domain (residues 58-94)

Reported prevalence values are study-specific, and structural location does not establish functional or clinical effects.
SNP single nucleotide polymorphism, *LXCXE* retinoblastoma-protein-binding motif.

C → T substitutions (and G → A in complementary strand) in sequencing data [54].

While APOBEC activity may help restrict HPV by introducing deleterious mutations in viral DNA, persistent or dysregulated APOBEC3 expression can increase mutational burden over time. Sublethal editing may therefore expand viral sequence diversity, and sustained APOBEC3 activity can also generate off-target mutations in host DNA during replication and repair.

Structural and functional analysis of amino acid changes for E6 and E7

While several SNPs map to the broader zinc finger domains of E6, none alter the specific cysteine residues (e.g., Cys 30, 33, 60, 66) that are critical for zinc coordination and structural integrity [55]. Also, no SNPs were identified within the PDZ-binding motif at residues 148–151, a feature characteristic of high-risk HPVs that is important for maintaining transforming activity and replicative capacity [56]. These findings may suggest that although the E6 oncogene exhibits more point mutations than E7, these mutations occur outside critical positions, potentially reflecting the HPV capacity to directly avoid disrupting E6 function while preserving oncogenic potential.

Our analysis identified several variant positions that mapped to broader regions implicated in E6AP and p53 interactions within the E6/E6AP/p53 ternary complex. However, positional overlap alone does not establish altered binding or downstream E6 activity, and further mechanistic studies are required to determine whether these variants have functional effects.

We observed two recurrent E7 SNPs, A647G/N29S and C712A/H51N; however, A647G/N29S may also represent a lineage-associated polymorphism, and this frequency should not be interpreted as evidence of independent adaptation.

In our analysis, both N29 and H51 are located outside the core functional regions of E7, including the LXCXE motif in CR2 and the zinc finger in CR3, which may explain their relatively high frequencies. This pattern is also consistent with observations in cervical cancer studies [57, 58]. Overall, fewer SNPs were observed in E7; E7 appears to maintain distinct evolutionary mechanisms that preserve its oncogenic potential [48].

Evolutionary and lineage context of recurrent variants

Several recurrent E6 variants, including A131G, G145T, C335T and T350G, were reported at relatively high frequencies within individual cohorts (summarised in Table 2). These patterns may reflect a combination of viral lineage composition, study design and structural tolerance. In particular, T350G/L83V is a recognised lineage-associated polymorphism. Nevertheless, the distribution of these variants is consistent with established evolutionary principles: substitutions between amino acids with similar physicochemical properties are generally less likely to destabilise protein structure, while mutations are more readily tolerated in non-core regions than in functionally constrained regions subject to strong purifying selection [17–19].

Even within core subregions, substitution rates vary based on structural exposure. Functionally critical "buried" residues such as those mediating zinc coordination are highly conserved and rarely mutate [21–23]. Conversely, intrinsically disordered regions (IDRs) lack rigid tertiary structures and tolerate mutations more readily, as these substitutions have a minimal impact on overall protein stability [24].

More broadly, most protein sites remain under purifying selection, which permits mutations only when they do not compromise protein stability or native conformation [25]. Consistent with these principles, A131G, G145T, C335T and T350G were recurrently reported and reached relatively high frequencies within individual studies. This mutational pattern closely mirrors observations of HPV16 in cervical cancer cohorts [59].

SNPs in the L1 region of HPV16

Two studies restricted their analysis to SNPs within the L1 gene, and one study [14] sequenced the entire HPV16 genome. The most common SNP in L1 was A5316C (N181T), detected in 28 of 70 oropharyngeal cancer patients (40%). This SNP has also been identified as a predominant variant in cervical cancers [60]. Structurally, the L1 protein forms a conserved β-barrel "jellyroll" fold surrounded by flexible, surface-exposed loops (BC, DE, EF, FG, and HI) [61].

These outward-projecting loops determine antigenicity; therefore, mutations within these loops can alter epitope conformation and potentially influence immune recognition [62]. In contrast, mutations in the conserved β-barrel core are typically deleterious due to disruption of capsid assembly or particle stability [61]. In our results, all three SNPs in L1 were surface-exposed residues with the common SNP A5316C (N181T) localising at the hypervariable surface loop (EF loop) [41].

Comparison with cervical cancer findings

Although the impact of HPV16 SNPs has been widely studied in cervical cancer [63] evidence in head and neck cancer remains limited. Within the E6, E7 regions, some of the variants identified in this study have also been observed in cervical cancer, where their functional consequences have been investigated. In particular, both R10G/L83V and Q14H/L83V have been associated with heightened susceptibility to cell-detachment-induced apoptosis [64], suggesting reduced oncogenic activity compared with wild-type E6. Q14H and H78Y modulate E6-dependent immortalisation and transformation, and L83V downregulates E-cadherin compared with other variants, possibly linking population-level oncogenicity to host genetic factors [65]. Clinically, while the L83V variant has been linked to higher rates of persistent infection and progression to high-grade cervical lesions [66] other evidence suggests no significant associations between common variants, including R10G, Q14H, H78Y, and L83V and specific cervical histopathological stages [67]. In our systematic review, no study evaluated the biological impacts from identified SNPs and only one study showed no significant differences between prototype and L83V in clinical T, N, and O staging, as well as disease-free survival in HNC [34].

Mechanistic roles of E6, E7 and L1 variants

Regarding SNPs in E7 oncogene, A647G(N29S) is one of the most common non-synonymous variants in HPV16 E7 reported in cervical cancer [58]. N29S has been also associated with increased risk of CIN2+ and cervical cancer [50] and was explained by its position near the recognition motif for casein kinase II (CKII) (positions 31 and 32) which can enhance CKII-mediated phosphorylation and promote pRb degradation [68]. The impact of C712A (H51N) appears less pronounced with no reported association with cervical carcinoma cases [50]. Therefore, it is necessary to elucidate the molecular and clinical significance of identified SNPs together with their prevalence in cancer, especially for those in critical regions. An example of high-risk SNPs in key functional regions was point mutations at positions 67, 82, 83, and 91 within the CR3 zinc-finger domain abolish E7's interaction with histone deacetylase 1 (HDAC1), thereby preventing E7 from overriding cell-cycle arrest [69]. Notably, our analysis identified 11 variant positions that map to E6AP- or p53-interacting regions in E6 (see Fig. 3). Their clinical significance remains unclear, and it is important to determine whether variants at these positions might exacerbate or mitigate the E6/E6AP/p53 tertiary structure, a key complex in cancer development.

Considering SNPs outside the E6 and E7 regions, the A5316C(N181T) variant in the L1 region was the most frequent in our analysis (40%) and is also commonly observed in cervical cancer [60]. However, its functional and clinical significance remains unvalidated. Among the studies we reviewed, only one

reported position 6025 within the FG loop as a high-risk SNP associated with poorer survival in head and neck cancer, reinforcing the relevance of identifying SNPs outside the E6 and E7 oncogenes [14]. Several positions within L1 have been investigated for biological relevance in cervical cancer, such as the C428W and K430Q variants showed 9-fold and 11-fold reduced susceptibility to neutralisation, respectively [70]. Twenty-seven L1 SNPs are located within immunodominant loops [71] with four reported to induce structural changes [72] highlighting the need for further study on their effects on L1 structure and immune recognition.

High-risk URR variants

The URR contains the transcriptional enhancer and promoter regions as well as the origin of DNA replication [73]. The 5′ segment contains a transcription termination site and a negative regulatory element acting at the level of late mRNA stability [51]. Thus, it is predicted that the URR variations can have various transcriptional activities by changing the activity of specific positions. When mapping, A7173C locates in the 5′ segment of the URR and was reported to locate in the binding site of the transcription factor AP1, FOXA1, and C/EBPβ, which have been shown to regulate HPV16 transcription [14] and the progression of cancer in different cancers, such as breast, liver and lung cancer [51].

SNPs at other positions within URR also affect many biological events in cervical cancer [49, 51], such as C7729A, C7763T, and A7841G. These are located near binding sites for transcription factors YY1, NF1, TEF-1, and Octamer factor-1, which play roles in viral replication and regulation of viral and cellular gene expression, showed significant differences between cervical cancer patients and controls [49, 51]. Similarly, SNPs T24C [49, 51] and C7873G [74] in the 3′ segment of the URR, which includes the p97 promoter, could affect transcription of E6 and E7. Those findings in the URR indicate that, unlike highly polymorphic sites, rare deleterious SNPs may have immediate biological consequences by disrupting or altering key oncogenic signaling pathways [75].

STRENGTH, LIMITATIONS AND CONCLUSION

This review has several limitations that must be acknowledged. First, the literature search was restricted to English-language studies, which introduces potential bias. Second, the included studies are highly heterogeneous in their experimental designs, targeted genomic regions, and objectives. Some studies performed comprehensive sequencing of specific genes or the entire genome, while others only genotyped a few predefined candidate SNPs. This data heterogeneity makes direct cross-study comparisons of SNP prevalence difficult, and the reported frequencies should be viewed as exploratory.

Also, synonymous coding variants were excluded because the review focused on the structural annotation and protein-level interpretation of amino-acid substitutions. However, synonymous variants are not necessarily biologically neutral and may influence splicing, codon usage, RNA secondary structure, transcript stability, regulatory motifs, or viral lineage assignment. These potential effects were outside the scope of this review and were not assessed. This exclusion applied only to synonymous variants within coding regions; variants in non-coding regions, including the upstream regulatory region, were retained and analysed separately.

Furthermore, the reported association between HPV16 SNPs and survival rate was based on a single whole-genome sequencing study. That study reported eight high-risk SNPs adjusted for age, cancer stage, smoking history, and treatment but not for HPV transcriptional activity. However, the underlying cohort (384 patients from two US centers, 97.66% Caucasian) has not been externally validated, so this finding should be interpreted with caution pending independent validation. All the included studies also used different methods to detect HPV and HPV16 positivity, and these approaches are not directly comparable. Furthermore, the single oral-rinse study sampled a different biological compartment from the 13 tumour-tissue studies, limiting direct comparison.

Overall, this review summarises the HPV16 genomic variants reported in head and neck cancer and maps selected variants to relevant viral proteins and regulatory regions. The available evidence provides preliminary insight into the potential biological and clinical relevance of these SNPs, although substantial methodological differences limit any firm conclusions. Future studies should adopt standardised approaches to specimen collection, processing, HPV16 detection and SNP analysis. Combined with mechanistic validation and prospectively collected clinical data, these approaches may improve understanding of the biological and prognostic significance of HPV16 alteration in head and neck cancer patients.

DATA AVAILABILITY

Data used in this study are available in a public, open-access repository. All data relevant to the study are included in the article or in the Supplementary files.

REFERENCES

1. Owens D, Paleri V, Jones AV. Head and neck cancer explained: an overview of management pathways. Br Dent J. 2022;233:721–5.
2. Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. Annu Rev Pathol: Mechan Dis. 2009;4:49–70.
3. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J Clin. 2024;74:229–63.
4. Galati L, Chiocca S, Duca D, et al. HPV and head and neck cancers: Towards early diagnosis and prevention. Tumour Virus Res. 2022;14:200245.
5. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer. 2017;141:664–70.
6. D'Souza G, Kreimer AR, Viscidi R, et al. Case–control study of human papillomavirus and oropharyngeal cancer. N Engl J Med. 2007;356:1944–56.
7. Abogunrin S, Di Tanna GL, Keeping S, Carroll S, Iheanacho I. Prevalence of human papillomavirus in head and neck cancers in European populations: a meta-analysis. BMC Cancer. 2014;14:968.
8. Mehanna H, Beech T, Nicholson T, 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;35:747–55.
9. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J Clin. 2018;68:394–424.
10. Brook I. Late side effects of radiation treatment for head and neck cancer. Radiat Oncol J. 2020;38:84–92.
11. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010;363:24–35.
12. Pinatti LM, Walline HM, Carey TE. Human papillomavirus genome integration and head and neck cancer. J Dent Res. 2018;97:691–700.
13. Tran NH, Sais D, Tran N. Advances in human papillomavirus detection and molecular understanding in head and neck cancers: Implications for clinical management. J Med Virol. 2024;96:e29746.
14. Lang Kuhs KA, Faden DL, Chen L, et al. Genetic variation within the human papillomavirus type 16 genome is associated with oropharyngeal cancer prognosis. Ann Oncol. 2022;33:638–48.
15. Wilke CO, Bloom JD. Protein Biophysics and Evolution. In: Encyclopedia of Evolutionary Biology. Elsevier; 2016:339–43.
16. von Knebel Doeberitz M. HPV 16 and its variants: minor genome variations make a big difference. J Natl Cancer Inst. 2016;108.
17. Epstein CJ. Non-randomness of amino-acid changes in the evolution of homologous proteins. Nature. 1967;215:355–9.
18. James JE, Lascoux M. Amino acid properties, substitution rates, and the nearly neutral theory. Genome Biol Evol. 2025;17:evaf025.
19. Franzosa EA, Xia Y. Structural determinants of protein evolution are context-sensitive at the residue level. Mol Biol Evol. 2009;26:2387–95.
20. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44:W344–350.

21. Bowie JU, Lüthy R, Eisenberg D. A Method to identify protein sequences that fold into a known three-dimensional structure. Science. 1991;253:164–70.
22. Echave J. Beyond stability constraints: a biophysical model of enzyme evolution with selection on stability and activity. Mol Biol Evol. 2019;36:613–20.
23. Ristriani T, Masson M, Nominé Y, et al. HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way Junctions. J Mol Biol. 2000;296:1189–203.
24. Huang H, Sarai A. Analysis of the relationships between evolvability, thermodynamics, and the functions of intrinsically disordered proteins/regions. Comput Biol Chem. 2012;41:51–57.
25. Cvijović I, Good BH, Desai MM. The effect of strong purifying selection on genetic diversity. Genetics. 2018;209:1235–78.
26. Tugwell P, Tovey D. PRISMA 2020. J Clin Epidemiol. 2021;134:A5–A6.
27. Boyle J. Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. Biochem Mol Biol Educ. 2005;33:74–75.
28. Munn Z, Barker TH, Moola S, et al. Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. *JBI Database of Systematic Reviews and Implementation Reports*. 2019; Publish Ahead of Print.
29. Agrawal Y, Koch WM, Xiao W, et al. Oral human papillomavirus infection before and after treatment for human papillomavirus 16–positive and human papillomavirus 16–negative head and neck squamous cell carcinoma. Clin Cancer Res. 2008;14:7143–50.
30. Barbieri D, Nebiaj A, Strammiello R, et al. Detection of HPV16 African variants and quantitative analysis of viral DNA methylation in oropharyngeal squamous cell carcinomas. J Clin Virol. 2014;60:243–9.
31. Boscolo-Rizzo P, Da Mosto MC, Fuson R, Frayle-Salamanca H, Trevisan R, Del Mistro A. HPV-16 E6 L83V variant in squamous cell carcinomas of the upper aerodigestive tract. J Cancer Res Clin Oncol. 2009;135:559–66.
32. Castillo A, Koriyama C, Higashi M, et al. Human papillomavirus in upper digestive tract tumors from three countries. World J Gastroenterol. 2011;17:5295–304.
33. Du J, Nordfors C, Näsman A, et al. Human Papillomavirus (HPV) 16 E6 variants in tonsillar cancer in comparison to those in cervical cancer in Stockholm, Sweden. PLoS One. 2012;7:e36239.
34. Gameiro SF, Salnikov MY, Zeng PYF, Barrett JW, Nichols AC, Mymryk JS. HPV16 intratypic variants in head and neck cancers: a North American perspective. Viruses. 2023;15:2411.
35. Gillison ML, Koch WM, Capone RB, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst. 2000;92:709–20.
36. Hassani S, Castillo A, Ohori J, et al. Molecular pathogenesis of human papillomavirus type 16 in tonsillar squamous cell carcinoma. Anticancer Res. 2015;35:6633–8.
37. Hoffmann M, Lohrey C, Hunziker A, Kahn T, Schwarz E. Human papillomavirus type 16 E6 and E7 genotypes in head-and-neck carcinomas. Oral Oncol. 2004;40:520–4.
38. Joseph AW, Ogawa T, Bishop JA, et al. Molecular etiology of second primary tumors in contralateral tonsils of human papillomavirus-associated index tonsillar carcinomas. Oral Oncol. 2013;49:244–8.
39. Kottaridi C, Resta P, Leventakou D, et al. The T350G variation of human papillomavirus 16 E6 gene prevails in oropharyngeal cancer from a small cohort of Greek patients. Viruses. 2022;14:1724.
40. LeConte BA, Szaniszlo P, Fennewald SM, et al. Differences in the viral genome between HPV-positive cervical and oropharyngeal cancer. PLoS One. 2018;13:e0203403.
41. Yoshida T, Ogawa T, Nakanome A, et al. Investigation of the diversity of human papillomavirus 16 variants and L1 antigenic regions relevant for the prevention of human papillomavirus-related oropharyngeal cancer in Japan. Auris Nasus Larynx. 2022;49:1033–41.
42. Todorovic B, Hung K, Massimi P, et al. Conserved region 3 of human papillomavirus 16 E7 contributes to deregulation of the retinoblastoma tumor suppressor. J Virol. 2012;86:13313–23.
43. Tompa P, Fersht A. Structure and function of intrinsically disordered proteins. 0 ed: Chapman and Hall/CRC; (2009).
44. Best SR, Niparko KJ, Pai SI. Biology of human papillomavirus infection and immune therapy for HPV-related head and neck cancers. Otolaryngologic Clin North Am. 2012;45:807–22.
45. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. Cancer Epidemiol, Biomark Prev. 2005;14:467–75.
46. Pan C, Issaeva N, Yarbrough WG. HPV-driven oropharyngeal cancer: current knowledge of molecular biology and mechanisms of carcinogenesis. Cancers Head Neck. 2018;3:12.
47. Ferris RL, Westra W. Oropharyngeal carcinoma with a special focus on HPV-related squamous cell carcinoma. Annu Rev Pathol: Mechanisms Dis. 2023;18:515–35.
48. Martinez-Zapien D, Ruiz FX, Poirson J, et al. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature. 2016;529:541–5.
49. Mirabello L, Yeager M, Yu K, et al. HPV16 E7 genetic conservation is critical to carcinogenesis. Cell. 2017;170:1164–e1166.
50. Dai M-Z, Qiu Y, Di X-H, Shi W-W, Xu H-H. Association of cervical carcinogenesis risk with HPV16 E6 and E7 variants in the Taizhou area, China. BMC Cancer. 2021;21:769.
51. Dai S, Li C, Yan Z, et al. Association of human papillomavirus type 16 long control region variations with cervical cancer in a Han Chinese population. Int J Med Sci. 2020;17:931–8.
52. Chen Z, Terai M, Fu L, Herrero R, DeSalle R, Burk RD. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. J Virol. 2005;79:7014–23.
53. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 2003;424:99–103.
54. Warren CJ, Xu T, Guo K, et al. APOBEC3A functions as a restriction factor of human papillomavirus. J Virol. 2015;89:688–702.
55. Barbosa MS, Lowy DR, Schiller JT. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. J Virol. 1989;63:1404–7.
56. Pim D, Bergant M, Boon SS, et al. Human papillomaviruses and the specificity of PDZ domain targeting. FEBS J. 2012;279:3530–7.
57. Lee CW, Bae JH, Lee SJ, et al. Distribution of human papillomavirus type 16 E6 and E7 gene variants in the progression of cervical dysplasia in Korean women. J Obstet Gynaecol Res. 2011;37:1320–6.
58. Yang L, Yang H, Wu K, Shi X, Ma S, Sun Q. Prevalence of HPV and variation of HPV 16/HPV 18 E6/E7 genes in cervical cancer in women in South West China. J Med Virol. 2014;86:1926–36.
59. Qmichou Z, Khyatti M, Berraho M, et al. Analysis of mutations in the E6 oncogene of human papillomavirus 16 in cervical cancer isolates from Moroccan women. BMC Infect Dis. 2013;13:378.
60. Wang X, Han S, Li X, Wang X, Wang S, Ma L. Prevalence and distribution of human papillomavirus (HPV) in Luoyang city of Henan province during 2015–21 and the genetic variability of HPV16 and 52. Virol J. 2022;19:37.
61. Bishop B, Dasgupta J, Klein M, et al. Crystal structures of four types of human papillomavirus l1 capsid proteins. J Biol Chem. 2007;282:31803–11.
62. Carpentier GS, Fleury MJJ, Touzé A, et al. Mutations on the FG surface loop of human papillomavirus type 16 major capsid protein affect recognition by both type-specific neutralizing antibodies and cross-reactive antibodies. J Med Virol. 2005;77:558–65.
63. Ortiz-Ortiz J, Alarcón-Romero LDC, Jiménez-López MA, et al. Association of human papillomavirus 16 E6 variants with cervical carcinoma and precursor lesions in women from Southern Mexico. Virol J. 2015;12:29.
64. Zehbe I, Richard C, DeCarlo CA, et al. Human papillomavirus 16 E6 variants differ in their dysregulation of human keratinocyte differentiation and apoptosis. Virology. 2009;383:69–77.
65. Togtema M, Jackson R, Richard C, Niccoli S, Zehbe I. The human papillomavirus 16 European-T350G E6 variant can immortalize but not transform keratinocytes in the absence of E7. Virology. 2015;485:274–82.
66. Grodzki M, Besson G, Clavel C, et al. Increased Risk for Cervical Disease Progression of French Women Infected with the Human Papillomavirus Type 16 E6-350G Variant. Cancer Epidemiol, Biomark Prev. 2006;15:820–2.
67. Khezeli Z, Shoja Z, Kaffashian M, Soleimani-Jelodar R, Jalilvand S. The Lineage and Sublineage Investigation of Human Papillomavirus Type 16 in Tehran, Iran, During 2022–3: A Cross-Sectional Study. Health Sci Rep. 2025;8:e70680.
68. Zine El Abidine A, Tomaić V, Bel Haj Rhouma R, et al. A naturally occurring variant of HPV-16 E7 exerts increased transforming activity through acquisition of an additional phospho-acceptor site. Virology. 2017;500:218–25.
69. Brehm A, Nielsen SJ, Miska EA, et al. The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. EMBO J. 1999;18:2449–58.
70. Ning T, Wolfe A, Nie J, Huang W, Chen XS, Wang Y. Naturally occurring single amino acid substitution in the L1 major capsid protein of human papillomavirus type 16: alteration of susceptibility to antibody-mediated neutralization. J Infect Dis. 2017;216:867–76.
71. King AJ, Sonsma JA, Vriend HJ, et al. Genetic diversity in the major capsid L1 protein of HPV-16 and HPV-18 in the Netherlands. PLOS One. 2016;11:e0152782.
72. Minhas S, Kashif M, Nisar H, Idrees M, Ansari F. Whole-genome analysis and evolutionary characterization of cervical and oral human papillomavirus 16. Exp Biol Med. 2023;248:2332–40.
73. Xi J, Chen J, Xu M, et al. Genetic variability and functional implication of the long control region in HPV-16 variants in Southwest China. PLOS One. 2017;12:e0182388.
74. Pientong C, Wongwarissara P, Ekalaksananan T, et al. Association of human papillomavirus type 16 long control region mutation and cervical cancer. Virol J. 2013;10:30.
75. Brown T Mutation, Repair and Recombination. In: Genomes. 2nd edition. Oxford: Wiley-Liss; (2002).

ACKNOWLEDGEMENTS

We sincerely thank Dr. Duy Khuong Hoang for his support in classifying data.

AUTHOR CONTRIBUTIONS

NHT, HGN, DS, NT were responsible for study concept and design. NHT, HGN performed the literature search, title screening, abstract screening. NHT and HGN were responsible for full-text screening and data extraction. NHT, HGN, DS, NT wrote the first draft of the report. All authors were responsible for data interpretation and had access to all the data. ME, NL, AT, TVN, and NT reviewed and edited the article. NT supervised the study. All authors approved the final version to be published.

FUNDING

This work was supported by SPHERE Cancer CAG PhD Scholarship Top-up Award (2024–2026) and the Garnett Passe and Rodney Williams Memorial Foundation (2025–2027).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44276-026-00251-9.

Correspondence and requests for materials should be addressed to Nham Tran.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026