RESEARCH ARTICLE

Exome Sequencing with Validations and Expression of p16/CDKN2A Shows no Association with HPV in Oral Cancers

Vidyarani Shyamsundar[1], Soundara Viveka Thangaraj[2], Arvind Krishnamurthy[3], Sanjana Vimal[2], Pallavi Kesavan[2], Aravindha Babu[1], Masthan KMK[1], Vijayalakshmi Ramshankar[2]*

Abstract

Purpose: Despite many studies attributing HPV infection to oropharyngeal tumorigenesis, its involvement in non-oropharyngeal cancers is ambiguous. We have evaluated the mutation profile of p16 along with protein expression and correlated it with the HPV status in oral cancers. **Methods:** Somatic mutations in p16 were studied by exome sequencing (n=25) and validated by Sequenom Mass spectrometry (n=50). Expression of p16 was studied by immunohistochemistry (IHC) and correlated with HPV16/18 status evaluated by PCR, and IHC (n=221) in oral cancers. **Results:** Out of 25 oral cancer patients' samples sequenced by Exome sequencing, p16 mutations were found in 4 samples (16%). All the p16 mutations were identified in patients with cancers in the site of gingivobuccal complex and not tongue subsite. All the 4 patients with p16 mutations had failed treatment, and showed a significantly poor disease-free survival. Insilico analysis of the types of p16 mutations showed mutated, truncated p16 protein having an increased intrinsic disorder, and all the mutations involved truncation post arginine. Validation of the p16 mutations by mass spectrometry showed 8/50 (16%) of patients harbouring pArg80Ter mutation, of which 7/8 (87.5%) had failed treatment. Overexpression of p16 in >70% of the tumour cells was found in 21.4% (26/121) OSCC patients, 6.75% (5/74) OPML patients and p16 expression was significantly correlated (p=0.001; χ^2 = 25.601) to the grade. All the samples were studied for HPV presence by PCR and IHC. We found that none of the p16 positive tumours showing expression in >70% of the tumour cells harbored HPV both by PCR as well as IHC. **Conclusion:** Our study emphasises the importance of p16 in oral cancers, and shows that oral cancer is not HPV associated and p16 expression is not a surrogate marker for HPV.

Keywords: HPV- oral cancer- p16 surrogate marker- p16 expression- P16 mutation

Introduction

There is a wide geographical variation in the incidence of Oral Squamous Cell Carcinoma (OSCC), with the highest rate found in the Indian subcontinent (India, Pakistan, Sri Lanka, Bangladesh) contributing to 25% of the new incident cases (Warnakulasuriya, 2008). Among these countries, India has the world's highest number of OSCC cases with approximately 1 million cases being diagnosed each year. It is estimated that 1% of the Indian population is living with Oral Pre-Malignant lesions (OPML) (Chaturvedi, 2012). The male to female ratio for OSCC is 2:1, and the most common site is gingivobuccal complex followed by tongue (Krishnamurthy and Ramshankar, 2013).. The scenario in the Western countries is different, with tobacco consumption falling and an increase in tumours arising from oropharynx giving rise to OroPharyngeal Squamous Cell Carcinoma (OPSCC).

This epidemiological change is attributed to high-risk Human Papilloma Virus (HPV), and particularly HPV 16 recognized as a causative agent for this growing subset of OPSCC. Recent evidence shows two- to three-fold increase in the prevalence of HPV-driven OPSCC in Northern America and Europe, the reason being changes in sexual behaviour (D'Souza et al., 2009; Chenevert and Chiosea, 2012).

Despite oral cavity cancer being the most common non-oropharyngeal head and neck cancer subsite where HPV is implicated with a reported prevalence of 0%–90%, the potential role of HPV in OSCC is still unresolved (Castro and Bussoloti Filho, 2006; Elango et al., 2011). This is of paramount importance because, in the coming years, the diagnosis, treatment and follow-up in head and neck cancers may vary according to HPV status as being observed for OPSCCs where HPV presence indicates a more favourable prognosis leading to less intensive

¹Centre for Oral Cancer Prevention, Awareness and Research (COCPAR). Sree Balaji Dental College and Hospital, BIHER University, Pallikaranai, Chennai – 600100, India. ²Department of Preventive Oncology (Research), Cancer Institute (WIA), Adyar, Chennai – 600020, India. ³Department of Surgical Oncology, Cancer Institute (WIA), Chennai – 600 020, India. *For Correspondence: r.vijayalakshmi@cancerinstitutewia.org

treatment regimens, less toxicity and better quality of life (Mirghani et al., 2015).

In the current study, we have evaluated and validated the somatic mutation profile of p16 and studied the expression by immunohistochemistry. We also evaluated the presence of HPV16/18 by PCR and IHC, and have assessed whether p16 expression is a surrogate marker for a transcriptionally active HPV in a large series of oral cancer samples.

Patients and Samples

This study has been approved by the appropriate institutional ethics committee, and has been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards (IRB-SBDCECM 105/13/58, and SBDCH/IEC/03/2016/19) and Cancer Institute WIA; Protocol 1 HNCOG (Cancer Institute, Women's India Association; Protocol 1 Head and Neck Co-operative Oncology Group). This study included prospective OSCC patients (n=25) diagnosed and treated between 2014 and 2015. Tumour tissues were collected from treatment naïve patients at the time of biopsy. All the patients were clinically staged and treated in accordance with the decision of the multi-specialty board. The experiments were undertaken with the understanding and written consent of each subject according to the ICMR (Indian Council of Medical Research) and World Medical Association Declaration of Helsinki guidelines. For validation studies, fresh OSCC samples were collected from prospective patients diagnosed with squamous cell carcinoma of the gingiva–buccal complex (n=50). We additionally collected retrospective archived Formalin Fixed Paraffin wax Embedded (FFPE) tissue blocks of squamous cell carcinoma of the Buccal mucosa (n=121), oral dysplasia (buccal mucosa) (n=74) and normal oral tissue (n=26). Normal buccal mucosal tissues were obtained during surgical extraction of third molar after obtaining written consent and FFPE blocks were made. The clinical and histopathological variables such as age, sex, site, tobacco habits, size of the tumour, pattern of the lesion, clinical stage and histological grade (Broders grading) were recorded.

Materials and Methods

Targeted Exome Deep sequencing

The DNA extraction from the tissues were done using DNA mini kit (QIAamp DNA Mini Kit, QIAGEN, Germany Cat no: 69504), and corresponding blood DNA was isolated using QIAamp DNA Blood Mini Kit (Qiagen, cat no 51104) according to manufacturer's protocol. Exome sequencing of 421 genes of length 1.8Mb was performed at 400× read length following DNA isolation from the prospective OSCC samples (n=25).

Somatic Mutation Calling and Annotation

Base quality score distribution, sequence quality score distribution, average base content per read, GC distribution in the reads, PCR amplification issues, checks for over representation of sequences and adapter trimming was done by different programs and tools namely fast-q-mcf program (Version 1.04.676) for adapter trimming, BWA program (version Ojibwa – 0.7.9) for alignment, Picard tool for duplicates removal, mutation calling using the Strelkasomatic variant caller tool (v2.0.2-29-ged3977a). After the mutations were identified, we passed the same applying the filters of Strelka program and the mutation should be in the target regions. The mutation was identified as a rare variant (<0.01 allele frequency) after comparison with DBMS, 1000 Genome, ExAC, EVS and TWINSUK. Disease annotation was done using Clinical, HGMD, GWAS, Swiss and COSMIC. Mutation consequence prediction was done using the computation prediction tools Polyphen, SIFT, Computational, Mutation Assessor and others. The CScape tool on FATHMM was used to predict the oncogenic status (disease-driver or neutral) of somatic point mutations in the coding and non-coding regions of the cancer genome.

Validation and replication in Sequenom iPLEX MassARRAY platform in prospective gingivo–buccal complex cancers (n=50)

Histopathologically confirmed tumor samples were subjected to DNA extraction. DNA was isolated by DNA mini kit (QIAamp DNA Mini Kit, QIAGEN, and Germany Cat no: 69504). After DNA extraction, the tissue samples were assayed for the analysis of mutation of p16 using MassARRAY Compact mass spectrometer (Sequenom; manufactured by Bruker Instruments, Billerica, USA) and Sequenom real-time detection software. Briefly, 0.5 μM PCR primer mix, containing both forward and reverse primers (as shown in supplementary table S1) were made for studying p16 mutations. PCR reactions in 5μL were performed, followed by the SAP and iPLEX reaction. The samples were then desalted, dispensed to a Spectro CHIP and analyzed with MALDI-TOF MS. The PCR conditions were as follows 95°C for 2 min, followed by 45°C Cycles of 95 for 30 seconds, 56°C for 30 seconds, 72°C for 60 seconds, 72°C for 5 min,4 and 4°C. For the SAP reaction, 2 μl SAP mix was dispensed to each well of the sample plate and incubated on a thermocycler with the following program: 37°C 40 min, 85°C 5 min, extension reaction. The reactions were dispensed to a SpectroCHIP, using a MassARRAY™Nanodispenser after conducting a volume check with real sample from the plate. The resulting spectra were analyzed by SpectroTyper software, which combined the base caller with a clustering algorithm.

Bioinformatics Analysis

The p16 protein mutation detected was subjected to FATHMM and RAPTOR X for In-silico analysis. Functional Analysis Through Hidden Markov Models (FATHMM) (http://fathmm.biocompute.org.uk) is a high-throughput web-server capable of predicting the functional consequences of both coding variants, i.e. non-synonymous Single Nucleotide Variants (nsSNVs), and non-coding variants in the human genome. The CScape tool (http://cscape.biocompute.org.uk) on FATHMM was used to predict the oncogenic status (disease-driver or neutral) of somatic point mutations in the p16 gene (Rogers et al., 2017). The database predicts

the results as oncogenic (high confidence), oncogenic and benign, and also gives a corresponding score (high-confidence predictions are approximately 92% accurate in coding regions and 76% accurate in non-coding regions). In theory, the higher the score, the better the oncogenic character.

RaptorX: Protein Structure and Function Prediction Powered by Deep Learning (http://raptorx.uchicago.edu) predicts protein secondary and tertiary structures, contact and distance map, solvent accessibility, disordered regions, functional annotation and binding sites. The amino acid sequence was retrieved from Ensembl database (http://www.ensembl.org/index.html) and the FASTA format was given as input data into the RaptorX to retrieve the corresponding 3-Dimensional structure (Källberg et al., 2012).

PCR for HPV

DNA was isolated from fresh frozen prospective samples (Exome sequencing samples, n=25, Sequenom iPLEX MassARRAY samples, n=50) and from 10×5µm serial sections of FFPE tissue samples (n=221). The DNA isolation was done using Tissue DNA Isolation Kit (Qiagen, Cat no: 69504) and the high pure FFPET DNA Isolation Kit (Roche, Mannheim, Germany) according to the manufacturer's recommendations. PCR was performed on all the samples after amplification of beta–globin gene as a quality check. The samples were tested for HPV using two sets of primers. The sequences of the primers as well as the conditions have been described previously (Ramshankar et al., 2015). Briefly, two sets of consensus primers used were SPF10 and GP5+/GP6+. The PCR reaction mixture of 25µl included 1XPCR buffer, 2.5mM MgCl2, 0.2mM of each DTP, 40pmoles of each primer, 1U Taq polymerase and 100ng of genomic DNA. Cervical cancer cell line SiHa, CaSki along with cervical cancer paraffin DNA known to harbour HPV DNA were used as positive controls. Normal lymphocyte DNA was used as a negative control.

Immunohistochemistry (IHC) based detection of p16 and HPV 16/18 E6

All the samples used in the study were reviewed by a senior Oral Pathologist (VS) for the confirmation of their pathological status. All the grading was done using ProgRes CapturePro 2.8.8 software (JENOPTIK optical systems) at 40× objective magnification. The IHC detection of p16 and HPV16/18 E6 expression was performed on four-micron sections of FFPE tissues. The sections were deparaffinised in xylene and rehydrated in absolute ethanol. Antigen retrieval was done with 0.05M Tris EDTA Buffer (pH – 9) in pressure cooker for 20 minutes. HPV 16/18 E6 retrieval was done in Citrate buffer (pH-6) by wet autoclaving for 30 minutes. Endogenous peroxidase activity was blocked using 0.03% hydrogen peroxide in distilled water for 10 minutes and then washed with Phosphate Buffered Saline (PBS). Sections were pre-incubated with 2% BSA for 40 minutes for blocking. The sections were incubated with mouse monoclonal antibody against p16 (clone - JC8) (Santacruz, sc-56330) in 1:150 dilution, HPV 16/18 E6 (Santacruz, Clone-C1P5- sc-

460), dilution 1:25, overnight at 4°C. Antigen expression was observed using the SuperSensitive™ Polymer-HRP IHC Detection System (BioGenex Laboratories, San Ramon, CA). Sections were counterstained with Haematoxylin, dehydrated and mounted in DPX. For HPV16/18 E6 staining and p16 expression, tissue sections from HPV16/18 positive Uterine Cervix squamous cell carcinoma FFPE sections were used as positive controls. Primary antibody was replaced with 2% BSA in negative control. Immunostaining of the sections was reviewed along with corresponding Haematoxylin and Eosin stained sections.

IHC scoring

The p16 positivity was seen as nuclear brown stain of varying intensity with or without cytoplasmic stain. In each field, 100 cells were counted and the area was marked using measure tool. Similarly, a total of 1,000 cells were counted and percentage positivity was calculated. All the counting was done at the objective magnification of 40×. High nuclear and high cytoplasmic staining in greater than 70% of tumour cells was considered positive as suggested by American Society of Clinical Pathology and College of American Pathologists previously (Chute et al., 2014; Larsen et al., 2014). HPV16/18 E6 staining was scored in OSCC as percentage of tumour cells staining and was compared with the cervical cancer tissue, used as positive control. The positive staining was visualised as intense nuclear stain.

Statistical Analysis

Statistical analysis of all the data was done using SPSS version 16. Non-parametric test like chi-square test was carried out. Survival curves were derived using the Kaplan Meier modules.

Results

The study design is depicted in Figure 1.

Exome Sequencing and Validations

In the prospective sample set (n=25), of OSCC, we had oral tongue squamous cell carcinoma (OTSCC) (n=10) and gingivo–buccal complex (n=15). The median age of patient cohort was 56 years and most were males (68%; n=17). A majority of them were tobacco users (73.1%; n=19) either as chewers or smokers (57.7%; n=15). Twelve patients (46.15%) consumed alcohol regularly. More than 90% of the data obtained in exome sequencing was above Q30 Phred Score. More than 99% of the reads were aligned back to the reference genome and average read duplicates found in the sample was <=5%. The median insert size of the aligned reads was about 150 bp and more than 90%

Table 1. Prediction of Oncogenic Status of P16 Single Nucleotide Polymorphism

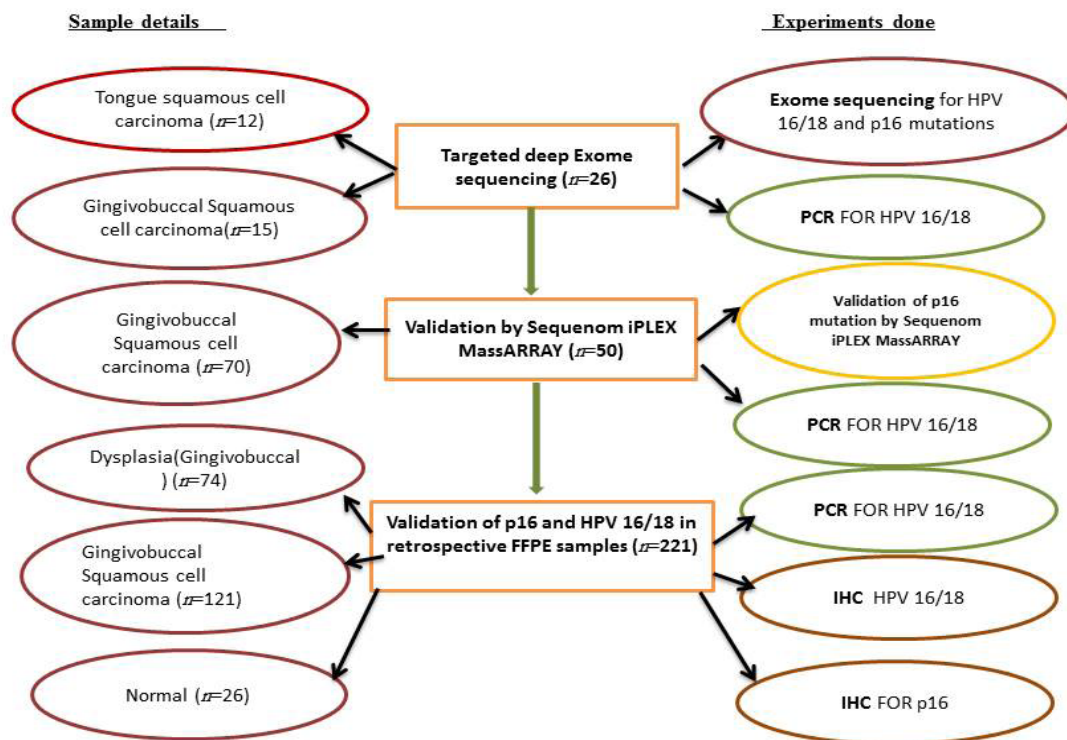| Nucleotide Change	Amino Acid change	CScape Prediction	CScape score
c.172C>T	p.Arg58Ter	Oncogenic	0.788085
c.238C>T	p.Arg80Ter	Oncogenic	0.863989

Figure 1. Schematic Diagram Explaining the Sample Details(Left Hand Side) and Study Design (Right Hand Side)

of the read aligned to the genome with a quality of 60 (Phred Scale). An average of 99.3% of the targeted region in the sample was covered, of which 73% of the targeted region had a depth >100X. Gingivo–buccal complex cancers showed P16 gene abnormalities in 20% (3/15) patients (one sample harboured two mutations). There were 3 types of p16 mutations identified, namely, in-frame deletion (c.83_100del; pVal28_Glu33del) was found in 50% (2/4); 25% (1/4) had nonsense mutation c.238C>T;

p.Arg80Ter and 25% (1/4) had c.172C>T; p.Arg58Ter mutations. All the buccal cancer patients harboring p16 mutations were identified to be tobacco users. Among the non-tobacco users, a mutation (rs121913388) in P16 gene (chromosome 9) was identified, causing termination of protein and displayed a "high impact." The consequence of this point mutation was also involving arginine at 58 position, leading to truncated p16 protein. Interestingly, all the abnormalities in p16 in the form of a mutation
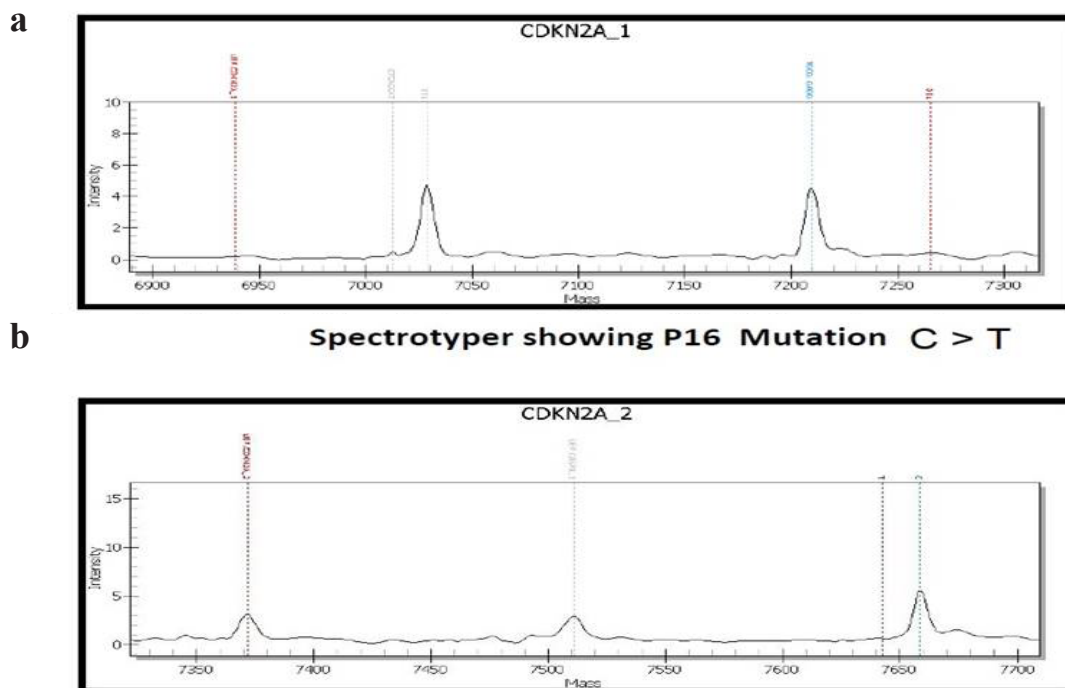

Figure 2. a, Output of Spectrotyper showing p16 wild type in oral cancer DNA genotyped in Sequenom platform; Output of Spectrotyper showing p16 mutant 238C>T in oral cancer DNA genotyped in Sequenom platform

Table 2. Predicted Intrinsic Disorder by PONDR Tool.

Disorders	P16 Wild Type	P16 truncated Arg58	P16 truncated Arg80	P16 – Deleted 28-33 del
Predicted residues	156	58	80	152
No of residues disordered	86	42	66	89
Overall percent disordred	55.13	72.41	82.5	58.55

involved arginine amino acid resulting in a truncated p16 protein. We compared p16 mutations to the 4 databases/ datasets – OncoMD, COSMIC, TCGA and ICGC study. There were no mutations identified in CDKN2A/p16 in tongue cancers.

Patient demography data are shown in supplementary file table S2. Majority (80%) of the patients had locally advanced disease. We found 24% of the patients (6/25) showing no evidence of the disease, and all the 4 patients with p16 mutations belonged to the 76% of the patients whose treatment failed. There was a significant difference in DFS among the patients with wild type and mutant p16 (log rank =40220; p=0.040) as shown in supplementary file S4.

Validation of p16 mutations identified by exome sequencing on Sequenom platform

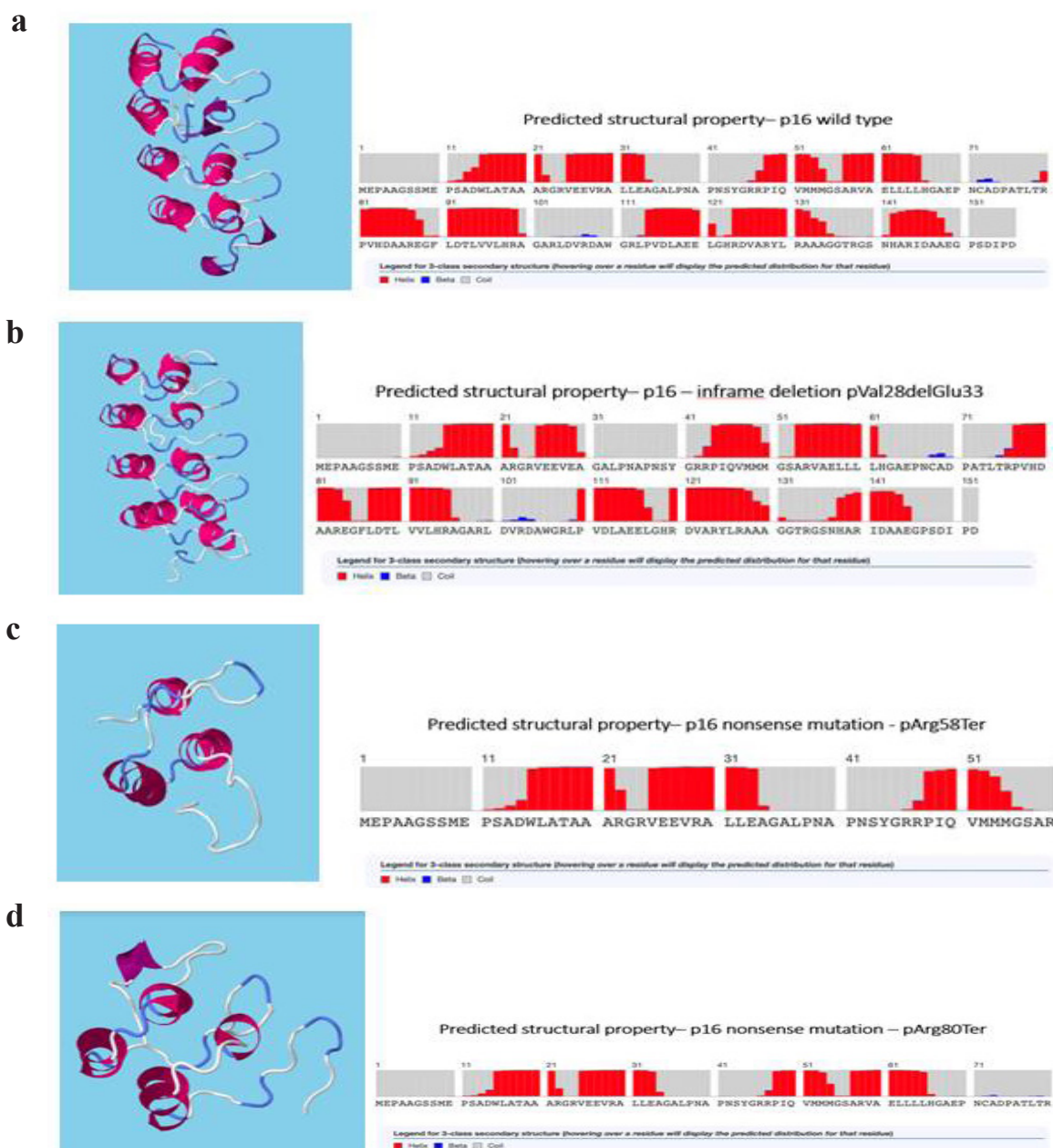Validation of the p16 mutations in prospective oral

Figure 3. a, Predicted 3 D model and structural property of wild type p16 protein; b, Predicted 3 D model and structural property p16 – inframe deletion pVal28delGlu33; c, Predicted 3 D model and structural property of p16 nonsense mutation - pArg58Ter; d, Predicted 3 D model and structural property of p16 nonsense mutation – pArg80Ter

Table 3. Expression of p16 in Oral Pre-cancers and Cancers

Tumor	Total samples	Negative for p16 immunoexpression (n=190)	Positive (>70%) for p16 immunoexpression (n=31)
Normal	26	26 (100)	0
Hyperplasia	16	16 (100)	0
Mild Dysplasia	30	29 (96.7)	1 (3.3)
Moderate Dysplasia	16	13 (81.2)	3 (18.75
Severe Dysplasia	16	15 (93.8)	1 (6.25)
Well Differentiated squamous cell carcinoma	60	43 (71.7)	17 (28.3)
Moderately Differentiated squamous cell carcinoma	51	44 (86.3)	7 (13.7)
Poorly Differentiated squamous cell carcinoma	2	2 (100)	0
Verrucous Carcinoma	4	2 (50)	2 (50)

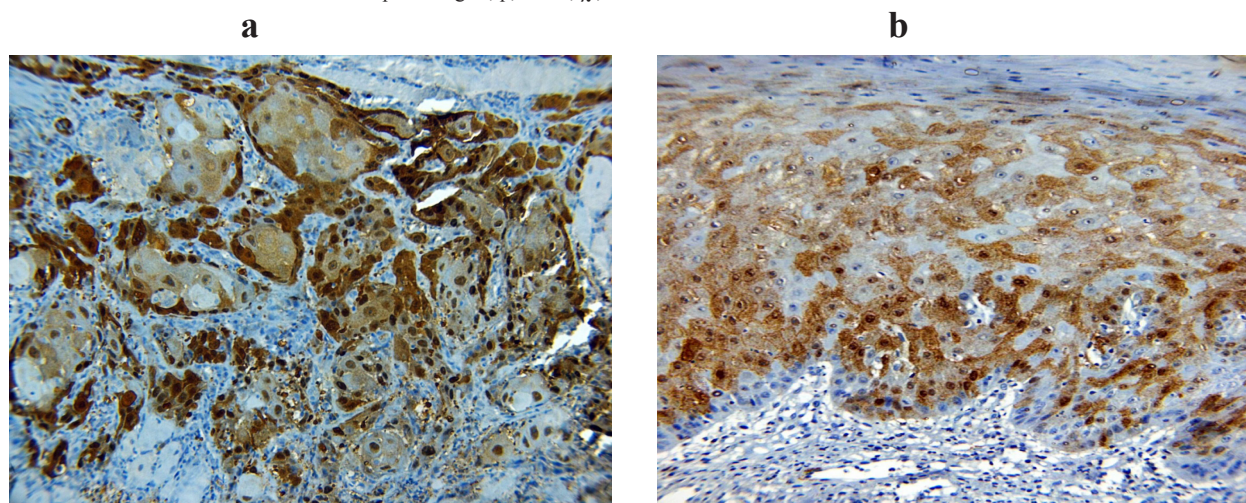Note : The numbers within brackets denote percentages ; p, 0.001; χ^2, 25.601

Figure 4. a, Well differentiated oral squamous cell carcinoma showing p16 positivity in more than 70% of the cells. IHC, 20x magnification; b, Moderate dysplasia of Oral epithelium showing p16 positivity in more than 70% of the cells. IHC, 20x magnification.

cancer samples (n=50) on Sequenom iPLEX MassARRAY (Figure 2a and 2b) showed 16% (8/50) of patients harbouring p.Arg80Ter mutation, out of which 7/8 (87.5%) had failed treatment. All the patients with p16 mutations had tobacco and/or alcohol habits (supplementary table S3). The in-frame deletion (p.Val28_Glu33del) and c.172C>T (p.Arg58Ter) were not found in the validation set.

In-silico analysis of the mutations in p16 protein

CScape analysis for the dataset revealed nonsense mutations in the p16 gene to be oncogenic with low confidence. At positions 58 and 80 amino acid arginine is truncated with the respective oncogenic scores 0.788085 and 0.863989 (Table 1). From the RaptorX data, it was observed that the percentage of coils had increased in the truncated protein structures in comparison to the wild-type protein structure. The alpha helices and beta sheets were found to be reduced in the truncated proteins. Due to the strong structural variations in the mutant proteins, the functionality may vary to contribute to a diseased state. The structure of wild type p16 along with the mutant forms obtained and the percentages of the secondary structures are shown in the figure (Figures 3a, 3b,3c,3d).

We also assessed the property of intrinsic disorder in the mutant p16 proteins using the PONDR tool. The first tool designed specifically for studying the prediction of protein disorder was PONDR (Predictor Of Naturally

Table 4. Size of the Tumour vs p16 Expression

Tumor Size	Total samples	Negative for p16 immunoexpression (n=190)	Positive (>70%) for p16 immunoexpression (n=31)
Normal	26	26 (100)	0
Dysplasia	74	69 (93.2)	5 (6.7)
Tumour< 2 cm	28	24 (85.7)	4 (14.2)
Tumour size 2 - 4 cm	30	20 (66.6)	10 (33.3)
Tumour size < 3 cm	32	27 (84.3)	5 (15.6)
Tumour size < 4 cm	31	24 (77.4)	7 (22.5)

Table 5. Habits vs p16 Expression

Habits	N	Negative for p16 immunoexpression (n=190)	Positive (>70%) for p16 immunoexpression (n=31)
Habits (Alcohol and Tobacco) + Either Alcohol or Tobacco	164	136 (80.4)	28 (17.1)
No Habits	57	54 (94.7)	3 (5.2)

Note : The numbers within brackets denote percentages; p, 0.027; χ^2, 4.892
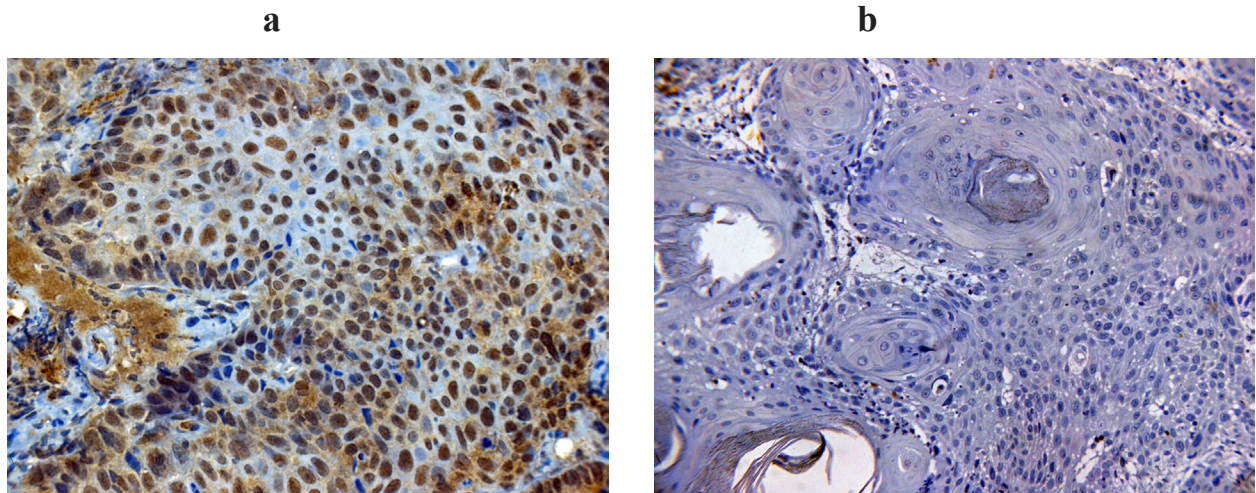
a **b**

Figure 5. a, Squamous cell carcinoma of Uterine cervix showing positive nuclear staining for HPV 16/18 E6 protein. IHC (Positive Control), 20x magnification; b, Well differentiated oral squamous cell carcinoma showing negative staining for HPV 16/18 E6 protein. IHC, 20x magnification

Disordered Regions, http://www.pondr.com) (Romero et al., 1997, Garner et al., 1998 and Garner et al., 1999). It is based on artificial neural networks. Interestingly the truncated p16 protein showed a dramatic increase in the intrinsic disorder compared to the wild type p16 (Table 2). Intrinsic disorder of the proteins is involved in critical functions of the transcription machinery and it would be interesting to explore the oncogenic property of the truncated p16 protein.

Expression of p16

In the validation set (n=221), comprising squamous cell carcinoma of the gingivo–buccal complex (n=121), OPML (buccal mucosa) (n=74) and absolute normals (n=26) the median age of the patient cohort was 49 years. Despite tobacco use being the most common etiologic agent, we had 12% (14/117) of the patients among OSCC and 26.9% (21/78) among OPMLs without any association with tobacco and alcohol habits. We found betel quid chewing habits significantly associated (p=0.001; χ^2 =26.4) among females whereas males opting for more commercial tobacco preparation. Expression of p16 was done in a series of normal (n=26), OPML (n=74) and OSCC patients (n=121). Overexpression of p16 in >70% of the tumour cells (Figure 4a and 4b) was seen in 14% (31/221), of which 21.4% (26/121) were OSCC samples and 6.4% (5/78) were OPML. None of the normal samples (n=26) expressed p16. Expression of p16 was significantly (p=0.001; χ^2 = 25.601) correlated to the differentiation status (Table 3) and T size of the tumours (p=0.002; χ^2 = 18.708) of the oral pre-cancers and cancers (Table 4). Expression of p16 was also significantly correlated (p=0.027; χ^2 = 4.892) to patients with alcohol and/or tobacco habits, compared to patients with no habits of tobacco and alcohol (Table 5).

Correlation with HPV assessed by PCR and immunohistochemistry

The DNA isolated from the prospective samples (n=75) (exome sequence and Sequenom iPLEX MassARRAY samples) were studied for HPV16/18 by PCR using GP5+/GP6+ primers and found to be negative for HPV. None of the samples displayed the presence of HPV. We further performed PCR to validate these findings in FFPE sections (n=221). All the samples were amplified for housekeeping gene beta globin to confirm the adequacy of the extracted DNA followed by HPV DNA, using two sets of primers, SPF10 and GP5+/GP6+. None of the samples tested were positive for HPV DNA. Presence of HPV 16/18 E6 protein is the proof of both presence and oncogenic functional expression. Antibody detecting both HPV16/18 E6 oncoprotein was stained for protein expression in the validation set (n=221). There was a complete correlation (Kappa =1) with the PCR data, as no cases expressed HPV 16/18 E6 (Figure 5a), whereas positive control showed strong positivity for HPV16/18 E6 (Figure 5b).

Discussion

We report exome sequencing based finding pertaining to p16 mutations along with validations in a large series of samples (n=296) by mass spectrometry and IHC in the current study.

Exome sequencing data in the current study have shown that frequency of p16 genetic abnormalities were seen in 16% OSCC of gingivo–buccal complex samples;

p53 was the only gene with a mutation frequency higher than p16 (data not shown in this publication). Truncation of the p16 protein is the most common consequence seen in our series of cases. We found only p.Arg80Ter mutation in the validation series of patients among the other mutations identified in the exome sequencing data. Therefore a higher sample size may be needed for validation. We speculate that probably p.Arg80Ter is more common among the other p16 mutations reported in this study. Interestingly, all the patients harbouring p16 mutations were tobacco users. The positive p16 immunoexpression was found to be correlating with patients having tobacco and/or alcohol habits in our IHC studies as well. Among the non-tobacco users, we found SNP rs121913388 in P16 (chromosome 9) involving arginine, which is interesting and needs to be explored.

We found patients with p16 mutations having an increased risk of failure that needs to be explored further. This is similar to the previous study, which also indicated that p16 expression may be useful in risk stratification of patients more likely to fail (Ramshankar et al., 2014).

In-silico analysis of p16 mutated proteins led to several interesting insights. We found both the nonsense mutations leading to truncated p16 proteins having excess of coils in the secondary structure, and there was a dramatic increase in the intrinsic disorder of these truncated p16 proteins. Disordered proteins are often found to interact with many different cellular targets and perform promiscuous and moonlighting functions (Oldfield et al., 2008, Tompa et al., 2005, Hsu et al., 2013).

More importantly, our study shows that p16 expression is not a surrogate marker for HPV presence in OSCC. We did not find the presence of HPV DNA by PCR as well as protein expression by IHC in all our samples. Several previous studies have supported this assumption that HPV is not biologically active in a majority of OSCC and enhanced p16 expression can be seen in HPV negative tumours (Reuschenbach et al., 2013, Lingen et al., 2013, Kouvousi et al., 2013). In contrast to oropharyngeal squamous cell carcinoma (Ang et al., 2010, Gillison et al., 2000, Marur et al., 2010), where the prognostic relevance of biologically active HPV infection has been shown, there is no clear association demonstrated between HPV and OSCC. There could be HPV independent mechanisms that cause p16 expression, which suggests that it need not be concordant with the presence of HPV as found in cervical cancers and some oropharyngeal cancers.

Immunoexpression of p16 significantly correlates with the grade, and is associated with differentiation status of the disease. Expression of p16 seen in OSCC is probably due to HPV independent mechanisms, which needs to be explored further. Oral cancers are caused by excessive tobacco and alcohol use, characterized by numerous genetic and epigenetic alterations. In the current study we have followed a clear scoring system set at >70% of diffused cytoplasmic and nuclear staining. Our findings have been reflected previously by other studies (Lingen et al., 2013; Poling et al., 2014). A recent study conducted on Indian population has also reported a similar observation (Gheit et al., 2017), implying that there may

be geographical differences pertaining to p16 expression and HPV correlation, and OSCC cannot have p16 as a surrogate marker for HPV. Another important aspect pertaining to p16 is the choice of antibody, and several studies have used E6H4 clone of CINtech p16 histology kit (Roche). We have used another commercially available clone JC8 that had been previously used in several studies, showing that it is equally comparable to E6H4 as the findings of the study are similar to the reports using E6H4 (Haller et al., 2010; Lassen et al 2010).

The evidence – describing HPV prevalence in oral cavity – emerging from the general population in India is highly discordant.

Among the non-malignant subjects, two studies have reported a prevalence of 2.75% for HPV 16 and 22% for HPV 18 in salivary rinses (Kulkarni et al., 2017) and 65% positivity for HPV16/18 tested in oral smears (Pattanshetty et al., 2014). However, Western literature has uniformly reported a low prevalence (<10%) in normal oral mucosa (Kristoffersen et al., 2012). In India, prevalence of HPV in OSCC is also reported to be different across the different regions with 33.6% in the eastern region (Nagpal et al., 2002), 28% in north (Kumar et al., 2002), 6% in the western region (Koppikar et al., 2005) and studies with data based on developed nations report a variable prevalence ranging from 10% to 80% (Kumaraswamy and Vidhya, 2011).

Many recent Indian studies have evaluated the contribution and role of HPV in head and neck cancers (Bhosale et al., 2016). But most of the studies had oropharyngeal cases also included in their study, and reported a positivity of HPV between 15% and 70%. A systematic review, summarizing 60 different studies comprising 4195 patients with OSCC by Isayeva et al., (2012) shows that 16.8% (705 patients) had tumours with HPV 16 DNA. The rates of prevalence have been reported as 0% to 94.7% with a weighted prevalence of 20.2%. All these variations could be attributed to misclassification due to widespread use of successive versions of World Health Organisation's (WHO) International Classification of Diseases (ICD). Some of these studies have grouped base of tongue cancers as OSCC even though base of the tongue belongs to oropharynx. These studies have also attempted to study the role of HPV in oral cancer using different methods like PCR to detect HPV DNA, E6/E7 mRNA expression, serological assays, IHC and in-situ hybridisation based assays. This is applicable in OPSCCs as well, where up to 50% of HPV DNA positive tumours have been negative for HPV E6/HPVE7 mRNA expression (Smeets et al., 2006).

A large international study comprising 3,680 patients from 29 different countries with 1,260 oral cancer samples shows 0% prevalence of HPV 16 DNA and 0% prevalence of E6 mRNA. According to this study, central South Asia has been reported to have 0% HPV attributable fractions in oral cancer, implying that the role of HPV in oral cancer is probably anecdotal. This study has confirmed the role of HPV in oropharyngeal cancers but has drastically downplayed the role of HPV in non-oropharyngeal cancers including oral cancer (Castellsagué et al., 2016). Our

results are similar to the findings of this study.

The current study suggests that probably HPV has no role in oral cancers comprising both buccal and tongue subsites from the Indian population. Our series of OSCC was negative for HPV DNA by PCR, HPV RNA by transcriptome sequencing, and HPV E6 onco-protein expression as well. Most of our patients were tobacco users, but even among non-tobacco users, HPV presence was negative. None of our normal oral FFPE tissue samples were positive for HPV, proving that there may not be latent infection in oral cavity of normal samples.

In oropharyngeal cancers, p16 has been reported as a surrogate marker for HPV infection as reported for cervical cancers. Studies show a moderate agreement between p16 and HPV 16 RNA positivity in oropharyngeal cancers (Bussu et al., 2013). Some studies have shown an assay algorithm with double positivity of p16 expression, and HPV DNA could be used to predict HPV RNA positivity (Smeets et al., 2007). However specificity of p16 has remained low with ranges between 79% and 82% in oropharyngeal cancers compared to gold standard E6/E7 mRNA.

In conclusion, the current study implies that p16 is not a surrogate marker for HPV in oral cancers, as the p16 expression was found in the absence of HPV. The current study also shows that HPV is absent in oral cancer, raising an important question: whether it is worthwhile to study HPV prevalence in oral cancers and, more so, as no clinical parameters have been identified to select patients who would benefit from HPV testing in oral cancers. It is therefore important to classify the head and neck cancers into the oropharyngeal and non-oropharyngeal context before the role of HPV is focused upon. HPV independent p16 expression mechanisms probably play a role in oral cancers that can be explored further. Factors other than HPV including the tobacco habits and mutations associated with oral cancers have to be focused upon in detail for an effective oral cancer management.

P16 mutations have to be explored further in OSCC; they may have more important implications given the association with patients whose treatment failed.

Author Contribution Statement

Study concept- Vidyarani Shyamsundar, Ramshankar Vijayalakshmi. Study design - Vidyarani Shyamsundar, Ramshankar Vijayalakshmi. Data acquisition- Vidyarani Shyamsundar, Thangaraj Soundara Viveka, Sanjana Vimal, Pallavi Kesavan, Krishnamurthy Arvind. Quality control of algorithms: Vidyarani Shyamsundar, Thangaraj Soundara Viveka, Sanjana Vimal, Pallavi Kesavan, Krishnamurthy Arvind. Data analysis and interpretation: Vidyarani Shyamsundar, Thangaraj Soundara Viveka, Ramshankar Vijayalakshmi, Aravindha Babu, Masthan KMK. Statistical analysis: Vidyarani Shyamsundar, Thangaraj Soundara Viveka, Ramshankar Vijayalakshmi. Manuscript preparation and editing: Vidyarani Shyamsundar, Krishnamurthy Arvind, Ramshankar Vijayalakshmi. Manuscript review: Vidyarani Shyamsundar, Thangaraj Soundara Viveka, Ramshankar Vijayalakshmi, Krishnamurthy Arvind, Sanjana Vimal, Pallavi Kesavan, Aravindha Babu, Masthan KMK.

References

Ang KK, Harris J, Wheeler R, et al (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*, **363**, 24-5.

Bhosale PG, Pandey M, Desai RS, et al (2016). Low prevalence of transcriptionally active human papilloma virus in Indian patients with HNSCC and leukoplakia. *Oral Surg Oral Med Oral Pathol Oral Radiol*, **122**, 609-8.

Bussu F, Sali M, Gallus R, et al (2013). HPV infection in squamous cell carcinomas arising from different mucosal sites of the head and neck region. Is p16 immunohistochemistry a reliable surrogate marker?. *Br J Cancer*, **108**, 1157-2.

Castellsagué X, Alemany L, Quer M, et al (2016). ICO international HPV in head and neck cancer study group. HPV Involvement in Head and Neck Cancers: Coprehensive Assessment of Biomarkers in 3680 Patients. *J Natl Cancer Inst*, **108**, djv403.

Castro TP, Bussoloti Filho I (2006). Prevalence of human papillomavirus (HPV) in oral cavity and oropharynx. *Braz J Otorhinolaryngol*, **72**, 272-2.

Chaturvedi P (2012). Effective strategies for oral cancer control in India. *J Cancer Res Ther*, **1**, S55-6.

Chenevert J, Chiosea S (2012). Incidence of human papillomavirus in oropharyngeal squamous cell carcinomas: now and 50 years ago. *Hum Pathol*, **43**, 17-2.

Chute DJ, Aramouni GT, Brainard JA, et al (2014). Hybrid Capture 2 human papilloma virus testing for head and neck cytology specimens. *J Am Soc Cytopathol*, **3**, 173-2.

D'Souza G, Agrawal Y, Halpern J, Bodison S, Gillison ML (2009). Oral sexual behaviors associated with prevalent oral human papillomavirus infection. *J Infect Dis*, **199**, 1263-9.

Elango KJ, Suresh A, Erode EM, et al (2011). Role of human papilloma virus in oral tongue squamous cell carcinoma. *Asian Pac J Cancer Prev*, **12**, 889-6.

Garner E, Cannon P, Romero P et al (1998). Predicting disordered regions from amino acid sequence: Common themes despite differing structural characterization. Genome Inform. Ser. Workshop Genome Inform, **9**, pp 201-3.

Garner E, Romero P, Dunker AK, et al (1999). Predicting binding regions within disordered proteins. Genome Inform. Ser. Workshop Genome Inform, **10**, pp 41.

Gheit T, Anantharaman D, Holzinger D, et al (2017). HPV-AHEAD study group. Role of mucosal high-risk human papillomavirus types in head and neck cancers in central India. *Int J Cancer*, **141**, 143-1.

Gillison ML, Koch WM, Capone RB, et al (2000). Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst*, **92**, 709.

Haller F, Agaimy A, Cameron S, et al (2010). Expression of p16INK4A in gastrointestinal stromal tumours (GISTs): two different forms exist that independently correlate with poor prognosis. *Histopathology*, **56**, 305-8.

Hsu WL, Oldfield CJ, Xue B, (2013). Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci*, **22**, 258-3.

Isayeva T, Li Y, Maswahu D, Brandwein-Gensler M (2012). Human papillomavirus in non-oropharyngeal head and neck cancers: a systematic literature review. *Head Neck Pathol*, **6**, 104-0.

Källberg M, Wang H, Wang S, et al (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*, **7**, 1511-2.

Koppikar P, deVilliers EM, Mulherkar R (2005). Identification of human papillomaviruses in tumors of the oral cavity in an Indian community. *Int J Cancer*, **113**, 946-0.

Kouvousi M, Xesfyngi D, Tsimplaki E, et al (2013). Prevalence of human papillomavirus in 45 Greek patients with oral cancer. *J Oncol*, **2013**, 756510.

Krishnamurthy A, Ramshankar V (2013). Early stage oral tongue cancer among non-tobacco users--an increasing trend observed in a South Indian patient population presenting at a single centre. *Asian Pac J Cancer Prev*, **14**, 5061-5.

Kristoffersen AK, Enersen M, Kverndokk E, et al (2012). Human papillomavirus subtypes in oral lesions compared to healthy oral mucosa. *J Clin Virol*, **53**, 364-6.

Kulkarni SS, Kulkarni SS, Vastrad PP, et al (2011). Prevalence and distribution of high risk human papillomavirus (HPV) Types 16 and 18 in Carcinoma of cervix, saliva of patients with oral squamous cell carcinoma and in the general population in Karnataka, India. *Asian Pac J Cancer Prev*, **12**, 645-6.

Kumaraswamy KL, Vidhya M(2011). Human papilloma virus and oral infections: an update. *J Cancer Res Ther*, **7**, 120-7.

Kumar R, Rai AK, Das D, et al (2002). Alcohol and tobacco increases risk of high risk HPV infection in head and neck cancer patients: study from North-East region of India. *PLoS One*, **10**, e0140700.

Larsen CG, Gyldenløve M, Jensen DH, et al (2014). Correlation between human papillomavirus and p16 overexpression in oropharyngeal tumours: a systematic review. *Br J Cancer*, **110**, 1587-4.

Lassen P, Eriksen JG, Hamilton-Dutoit S, et al (2010). Danish Head and Neck Cancer Group (DAHANCA). HPV-associated p16-expression and response to hypoxic modification of radiotherapy in head and neck cancer. *Radiother Oncol*, **94**, 30-5.

Lingen MW, Xiao W, Schmitt A, et al (2013). Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. *Oral Oncol*, **49**, 1-8.

Marur S, D'Souza G, Westra WH, Forastiere AA (2010). HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol*, **11**, 781-9.

Mirghani H, Amen F, Blanchard P, et al (2015). Treatment de-escalation in HPV-positive oropharyngeal carcinoma: ongoing trials, critical issues and perspectives. *Int J Cancer*, **136**, 1494-3.

Nagpal JK, Patnaik S, Das BR (2002). Prevalence of high risk human papilloma virus types and its association with P53 codon 72 polymorphism in tobacco addicted oral squamous cell carcinoma (OSCC) patients of Eastern India. *Int J Cancer*, **97**, 649-3.

Oldfield CJ, Meng J, Yang JY, et al (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*, **9**, 1, S1.

Pattanshetty S, Kotrashetti VS, Nayak R, et al(2014). PCR based detection of HPV 16 and 18 genotypes in normal oral mucosa of tobacco users and non-users. *Biotechnic Histochem*, **89**, 433-9.

Poling JS, Ma XJ, Bui S, et al (2014). Human papillomavirus (HPV) status of non-tobacco related squamous cell carcinomas of the lateral tongue. *Oral Oncol*, **50**, 306.

Ramshankar V, Soundara VT, Shyamsundar V, Ramani P, Krishnamurthy A (2014). Risk stratification of early stage oral tongue cancers based on HPV status and p16 immunoexpression. *Asian Pac J Cancer Prev*, **15**, 8351-9.

Reuschenbach M, Kansy K, Garbe K, et al (2013). Lack of evidence of human papillomavirus-induced squamous cell carcinomas of the oral cavity in southern Germany. *Oral Oncol*, **49**, 937-2.

Rogers MF, Shihab HA, Gaunt TR, Campbell C (2017). CScape: a tool for predicting oncogenic single-point mutations in the cancer genome. *Sci Rep*, **7**, 11597.

Romero P, Obradovic Z, Kissinger CR, et al (1997). Identifying disordered proteins from amino acid sequences. Proc. *IEEE Int Conf Neural Networks*, **1**, 90-5.

Smeets SJ, Braakhuis BJ, Abbas S, et al (2006). Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. *Oncogene*, **25**, 2558-4.

Smeets SJ, Hesselink AT, Speel EJ, et al (2007). A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. *Int J Cancer*, **121**, 2465-2.

Tompa P, Szász C, Buday L (2005). Structural disorder throws new light on moonlighting. *Trends Biochem Sci*, **30**, 484-9.

Warnakulasuriya S (2008). Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*, **45**, 309-6.