RESEARCH ARTICLE

**Editorial Process:** Submission:12/27/2025 Acceptance:06/10/2026 Published:06/22/2026

# Genomic Landscape of Oral Squamous Cell Carcinoma in Never Smokers and Never Drinkers

**Sinumol George[1,2], Devasena Anantharaman[1,2]***

## Abstract

**Objective:** To identify the mutational landscape of oral cancers of unknown etiology by leveraging publicly available datasets. **Methods:** Oral cancer mutation data in TCGA were accessed to identify never-smoking and never-alcohol-drinking subjects with genomic, epigenomic, and transcriptomic data. Habit-free participants within published individual whole-exome sequencing studies of oral cancer were examined as the replication dataset. **Results:** Somatic mutation analysis of 42 habit-free TCGA oral cancer subjects revealed *MUC16* and *MUC5B* as recurrent mutational targets in 30% of all habit-free oral cancers; the highest after previously reported oral cancer genes such as *TP53, CASP8, CDKN2A, NOTCH1*. Comparison against other habit-free oral cancers not only confirmed these observations but also identified additional mucin genes as frequent mutational targets, supporting their potential causal role. Gene expression and immune fraction analysis of bulk transcriptomic data showed that habit-free oral cancers were enriched for the expression of mesenchymal genes and in T-regulatory cells. **Conclusion:** Together, these data suggest that mucin genes, specifically *MUC16* and *MUC5B*, are important for oral carcinogenesis in the absence of tobacco and alcohol. These results are relevant given the increasing clinical utility of mucins as druggable targets in cancer, particularly *MUC16*.

**Keywords:** Oral cancer- non-smoking/non-drinking- mutation profile- RNA sequencing

*Asian Pac J Cancer Prev,* **27 (6)**, 2323-2334

## Introduction

Oral squamous cell carcinoma represents a significant burden of morbidity and mortality worldwide, with an estimated 389,846 new cases and 188,438 deaths reported in 2022 alone [1]. Tobacco use, in smokeless and smoking forms, betel quid use and alcohol consumption contribute to nearly 70% of all oral cancers (OC) [2-4]. In recent times, around 4% of all OC have been attributed to Human Papillomavirus (HPV) infection, particularly by type 16 [5]. This implies that nearly 25% of OC can be attributed to other causes [6-8]. Among these, environmental factors such as second-hand smoke, heavy metal exposure, lifestyle choices like mate drinking, opium consumption, mouthwash use and oral health factors including history of periodontal disease, poor dentition induced chronic mechanical irritation and oral infections have been proposed. However, the evidence of causation for these potential risk factors is limited [3, 9].

Mutational patterns as signatures of underlying carcinogenic exposures have been characterized, particularly for tobacco smoking, UV exposure, aflatoxin, among others [10-12]. The contribution of recurrent somatic mutations as key driver genes has been established for all major cancers [13-18]. The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), together with other individual efforts have created a comprehensive catalogue of significantly mutated genes in OC that include *TP53*, *CDKN2A, CASP8*, *NOTCH1*, *FAT1* and *PIK3CA* [19-21]. Gene expression profiling of head and neck cancer has identified and validated molecular classes based on enrichment of specific gene sets into basal, mesenchymal and classical subtypes that are characterized by expression of gene signatures related to hypoxia, neuregulin and epithelial to mesenchymal transition (EMT); cell cycle, xenobiotic metabolism; and inflammation, immune response respectively [22]. Taken together, significant advances have been made in our understanding of the molecular pathways towards OC.

Notably, these studies have been predominantly conducted in ever tobacco/ alcohol user patient cohorts (proportion of tobacco/alcohol users in TCGA and ICGC were 70% and 100% respectively) [19-21]. However, oral cancer develops sporadically among individuals with no history of tobacco or alcohol consumption. We hypothesize that oral cancers of such unknown etiology will represent a distinct genomic landscape unique to the molecular pathways underlying them. Here, we present the descriptive genomic analysis using publicly available data in never smokers and never drinkers (also referred to as habit-free) diagnosed with oral cavity cancer.

*[1]Cancer research laboratory-13, BRIC- Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala, India. [2]Regional Centre for Biotechnology, Faridabad, Haryana, India. *For Correspondence: devasena@rgcb.res.in*

## Materials and Methods

*Study design*

To examine the mutational landscape of habit-free OC, we performed primary analysis in eligible subjects within TCGA (https://tcga-data.nci.nih.gov/tcga). External independent whole exome sequencing datasets comprising of habit free subjects were accessed to validate observed mutations (two studies and 20 subjects) [23, 24]. The mutation profile observed in TCGA habit-free OC were contrasted against habit-related OC from TCGA and ICGC (http://www.icgc.org/) for further validation. Transcriptomic datasets of TCGA habit-free subjects was examined to uncover global gene expression patterns, and further integrated with copy number variation (CNV) and methylation datasets to functionally validate mutational findings. Figure S1 shows the overall study design.

*Data accession and mutation analysis*

We obtained genomic, transcriptomic and epigenomic data through the Genomic Data Commons (GDC) portal using the TCGAbiolinks R package v2.31.3 for oral squamous cell carcinoma (last accessed on 7th January, 2025). Patient associated clinical metadata, including demographic and lifestyle information were accessed from CBioportal (https://www.cbioportal.org/). Tumors comprising oral tongue (n = 128), base of tongue, buccal mucosa (n = 23), alveolar ridge (n = 18), floor of mouth (n = 63), hard palate (n = 7) and non-specified sites of the oral cavity were included. The present habit-free analysis included forty-two oral cancer subjects recorded as non-smokers (≤100 cigarettes in lifetime), never alcohol drinkers and HPV-negative based on surrogate P16 positivity ascertained by immunohistochemistry. Subjects with missing data for exposures (n=40) and HPV status were excluded. Somatic variants were processed and summarized using the TCGAbiolinks and maftools packages (v.2.10.5) in R v.4.1.2 [25]. The top 50 most frequently mutated genes were visualized through an oncoplot. Mutational impact was predicted using Sorting Intolerant from Tolerant (SIFT) and variant effector predictor [26, 27]. Tumor mutation burden (TMB) for each sample was calculated by dividing the total number of non-synonymous somatic mutations by the size of the human exome (38 Mb). Patterns of mutual exclusivity and co-occurrence among recurrently mutated genes were assessed using the somaticInteractions function in the R Bioconductor package (v.4.1.2).

Trinucleotide mutational signatures were extracted and compared against reference signatures from the COSMIC (Catalogue of Somatic Mutations in Cancer) database (https://cancer.sanger.ac.uk/cosmic/signatures) using the SomaticSignatures framework. Somatic single-nucleotide variants were classified into 96 trinucleotide contexts [11]. A mutation count matrix (96×n) was generated for all samples and decomposed using non-negative matrix factorization (NMF) to extract underlying mutational signatures. Extracted signatures were normalized to unit sum and compared to COSMIC v3.3 reference signatures using cosine similarity, where values ≥0.85 indicated high concordance. All analyses were performed in R (version 4.1.2).

The replication dataset included 10 OC subjects identified as tobacco and alcohol never users from BioProject PRJNA740146 [24]. Raw reads were aligned to the GRCh38 reference genome using BWA-MEM (v0.7.17) [28]. Duplicate reads were marked, and base quality score recalibration was performed using Genome Analysis Toolkit, Broad Institute (GATK v4.3.0) [29]. Somatic variants were called using GATK Mutect2, employing a panel of normal to suppress sequencing and mapping artefacts. Variants were filtered using FilterMutectCalls and annotated using ANNOVAR [30]. High-confidence non-synonymous mutations (depth ≥20×, allelic depth≥5) that passed filtering were used for downstream analysis. Filtered VCFs were used to examine the mutational profile. Additionally, for replication, summary-level mutation data from 10 never- chewing/ smoking/ alcohol subjects were available from another independent study and were included [23]. Curated mutation summary files included annotated lists of all observed somatic variants with gene names, variant classifications, and functional consequences. Top 50 significantly mutated genes in these analyses were classified into known oral cancer genes, *MUC16*/*MUC5B* related genes based on STRING network analysis [31] and other mucin genes.

*Copy number analysis*

Segmented copy number variation (CNV) data for 42 habit-free TCGA OC subjects were obtained from the GDC portal as a masked copy number segment file. CNV segments were matched to the 42 habit-free OC tumors and filtered to retain only those overlapping the genomic coordinates of *MUC16* (chr19) and *MUC5B* (chr11). For each of these genes, overlapping segments were extracted and the dominant segment per sample was defined as the one with the highest absolute $\log_2$ copy-number value. Gene-level CNV calls for *MUC16* and *MUC5B* were summarized using R (v 4.1.2) for downstream integration with methylation and expression analyses.

*Methylation analysis*

DNA methylation data (level 3 β-values) for 42 habit-free TCGA OC subjects were obtained from the GDC portal. CpG probe annotations were retrieved from the Illumina HumanMethylation450 (HM450) manifest file [32], and probes mapping to the promoter and gene body regions of *MUC16* and *MUC5B* were extracted. In the case of *MUC16*, of the total 5 probes, 1 mapped to the gene body and 3 to the promoter (one probe matched the 3'UTR). Similarly, for *MUC5B*, 62 and 8 of total 74 mapped to the gene body and promoter respectively (three probes matched the 3'UTR). For each of these genes, average CpG-level β-values for the gene body and promoter were estimated. Group-wise comparisons between tumor and normal, of the calculated average β-value distribution for promoter and gene body were performed to discern the contribution, if any, of these gene methylation in habit-free oral carcinogenesis. The resulting per-sample methylation estimates for *MUC16* and *MUC5B*, were subsequently integrated with

expression counts of respective genes.

*Differential expression analysis*

The gene expression raw count data for 42 habit-free TCGA OC subjects was obtained from the GDC (project: TCGA-HNSC, harmonized STAR-counts, release 2025-01-07). The dataset included 42 primary tumours and 7 adjacent normals. Only genes expressed at counts ≥10 in at least 25% of samples were retained. Count data were normalized using the DESeq2 package (v1.34.0) [33] with variance stabilizing transformation and principal component analysis. Tumor versus normal were compared to determine differential gene expression. Genes with a log two-fold change ≥ 1 and Benjamini–Hochberg adjusted $p < 0.05$ were considered differentially expressed. Hierarchical clustering of differentially expressed genes (DEGs) was performed and heatmap were generated using R pheatmap package (v1.0.13). Functional annotation was carried out using gene ontology (GO) categories and pathway enrichment was assessed using the KEGG database [34]. Enrichment results were visualized using the ggplot2 package (version 3.0.4). Significance thresholds were set at a false discovery rate (FDR) < 0.25 and $p < 0.05$.

*Molecular subtyping OC subjects*

We classified habit-free oral tumors into established molecular subtypes using a centroid-based approach. RNA-seq expression data were processed by estimating the average of duplicate genes following normalization of the expression values. Subtype reference centroids for the four recognized head and neck cancer subtypes (basal, mesenchymal, classical, and atypical) were obtained from the Keck dataset [35]. Pearson correlation coefficients were calculated between each sample and the subtype centroids, and samples were assigned to the subtype with the highest correlation. PCA of the common genes was performed to visualize sample distribution and subtype separation. All analyses were conducted using R (v4.1.2).

*Immune cell composition analysis*

The relative fractions of immune cell types were estimated using CIBERSORTx, applying the LM22 leukocyte signature (22 immune cell subtypes) and 100 permutations for significance testing [36]. Correlations among immune cell types were assessed, and immune infiltration levels were compared between TCGA habit-free (n = 42) and habit-associated OC using ANOVA, p-value for significance was set at <0.05.

*Statistical analysis*

Two sample tests of proportions were employed to compare patient characteristics of tumors carrying *TP53* and *FBXW7* mutations. Survival associations were tested using univariate and multivariate Cox proportional hazards models adjusted for age, sex and stage as applicable. The association between TMB and focussed gene mutation status of *TP53* and *FBXW7* were examined through logistic regression in univariate and multivariate models adjusted for age, sex and stage at diagnosis (early and late) and odds ratios and 95% confidence intervals were obtained. *MUC16*/*MUC5B* gene expression counts in tumor and normal samples from bulk transcriptomic data were compared to estimate fold change in expression for each participant. Average *MUC16*/*MUC5B* gene expression levels for all forty-two tumors and 95% confidence interval of means was calculated. A single outlier value for *MUC16* was observed with an expression value greater than 650 times the group average. This datapoint was excluded from all analyses pertaining to *MUC16* expression to avoid bias. Impact of *MUC16*/ *MUC5B* gene mutations as deleterious, moderate or unknown estimated as mentioned above were used in stratified analysis. CNV and promoter or gene body methylation status of *MUC16*/ *MUC5B* were determined as described above. Average *MUC16*/*MUC5B* gene expression levels and 95% confidence interval of means for stratified analysis by mutation and copy number status was calculated and P-values were derived using t-test or Analysis of variance (ANOVA), as applicable. Pearson's correlation test was employed to estimate the r-squared value between *MUC16*/ *MUC5B* promoter or gene body methylation and expression. The association between *MUC16*/ *MUC5B* gene mutation status, molecular subtype and TMB was examined through logistic regression models and odds ratios and 95% confidence intervals were obtained. Univariate and multivariate models adjusted for age, sex and stage at diagnosis (early and late) were examined. Survival associations for *MUC16*/ *MUC5B* gene mutations were tested using univariate and multivariate Cox proportional hazards models adjusted for age, sex and stage as applicable. Statistical significance for mutual exclusivity was determined with Fisher's exact test, with a significance threshold of $p < 0.05$. Mutation frequencies between TCGA habit-free and TCGA or ICGC habit-related subjects were compared using two sample tests of proportions, p-value less than 0.05 was considered significant. Tumor infiltration level comparison between habit-free and habit-related subjects were tested using a two-way ANOVA. All statistical tests were performed using STATA statistical software, version18 (StataCorp, College Station, TX), and all reported P-values are two sided. Statistical significance was set at P-value less than 0.05.

## Results

*Mutation Profiling of habit-free oral cancers*

TCGA habit-free oral cancers were frequently seen in oral tongue, tended to be women, diagnosed at later stages and had a median survival of 4.49 years (Table 1). A total of 4146 mutations in 3351 genes were found in 42 samples, with every sample carrying at least one mutation. There were 539 genes mutated in at least two samples or more, with 8 gene mutations present in more than 20% of the samples (8 samples or more). An oncoplot of frequently occurring somatic mutation are depicted in Figure 1. The tumor mutation rate per individual was 2.7 in this group. Somatic mutations in previously known oral cancer genes were replicated in this habit-free group; *TP53* (78%), *TTN* (43%), *CASP8* (26%), *CDKN2A* (26%), *FAT1* (24%), *KMT2D* (19%), *HUWE1* (19%),

Table 1. Demographic and Clinical Characteristics of 42 Habit-Free OC Patients

| Variables | Habit-free oral cancers (n= 42) N (%) |
|---|---|
| Age (years) | |
| Mean ± SE | 64.19± 16.03 |
| Median | 65 |
| Range | 19- 87 |
| Gender | |
| Male | 11 (26.19) |
| Female | 31 (73.80) |
| Cancer subsite [a] | |
| Oral tongue | 24 (57.14) |
| Buccal mucosa | 18 (42.85) |
| Cancer stage at diagnosis [b] | |
| Stage I&II | 16 (38.09) |
| Stage III & IV | 27 (64.28) |
| Overall survival status | |
| Alive | 23 (53.65) |
| Dead | 19 (46.35) |

[a], Oral tongue includes base of tongue and buccal mucosa includes floor of mouth, gum and overlapping sites; [b], 1 subject missing stage information

*NOTCH1* (17%), *PCLO* (14%), *PIK3CA* (14%), *SYNE1* (14%), among others[19]. Further, pathogenic positional mutations in *TP53* (p.R248W, p.R273H, and p.R175H) *CDKN2A* (p.R58*, p.R80*), *CASP8* (p. Q398*, p.R494*), *NOTCH1* (p.R1984*,p.Q803*) and *FAT1* (p.R628*) were also replicated. Notably, in addition to the above known oral cancer genes, this analysis revealed *MUC16* (24%) and *MUC5B* (10%) as frequent mutational targets, together contributing to at least 30% of all mutations in this cohort (12 of 42). Two of the ten mutations in *MUC16* were truncating and categorized as high impact by variant effect predictor tool and one of them (8976759 C>A) was further identified as deleterious based on SIFT prediction. Mutations identified in *MUC16* were distributed in the N- terminal domain and 40% at the C-terminal tandem repeat domain, while MU5B mutations were predominantly localized in the C-terminal domain (Figure S2).

*MUC16, MUC5B mutations in habit-free oral cancers*

In order to ascertain the validity of the observed mutation profile of habit-free OC, we examined additional publicly available data from independent studies of OC that included subjects with no reported tobacco/alcohol history. Results are presented in Table 2. A total of 20 subjects from two independent studies were examined and showed that *MUC16*/*MUC5B* mutations were indeed present, and at higher mutation frequencies, in both the independent replication datasets. The somatic mutation frequency for *MUC16* was 40% in the replication series compared to 24% in TCGA habit-free oral cancers and 20-30% for *MUC5B* in the replication series compared to 10% in TCGA habit-free subjects. Further, genes

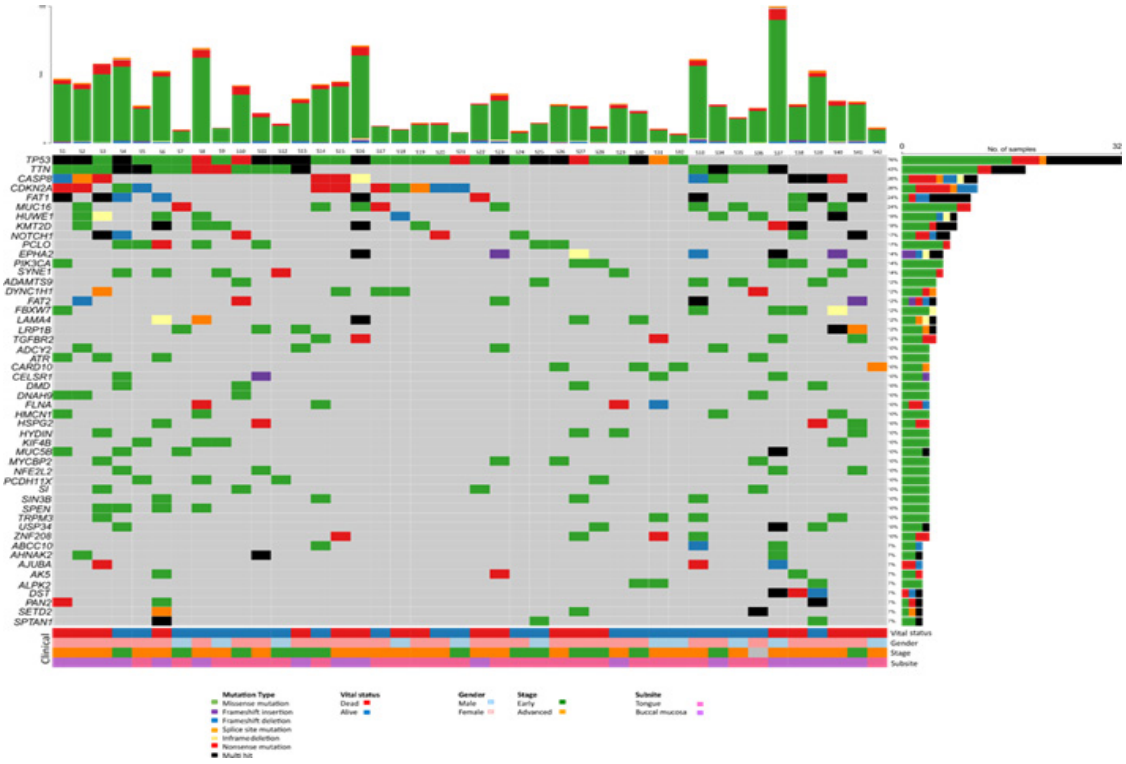


Figure 1. Somatic Mutation Landscape of the TCGA Habit-Free Oral Cancer Cohort. Non-synonymous mutations in 50 most frequently mutated genes across 42 habit-free oral cancers are depicted in this oncoplot. Each column represents an individual tumor sample and each row represents a gene. Colored blocks indicate different mutation types, as shown in the legend. Bar plots at the top indicate the total number of somatic mutations per sample, while bar plots on the right represent the mutation frequency of each gene in the cohort. Clinical annotations, including vital status, tumor stage, anatomical subsite, and gender, are displayed at the bottom of the oncoplot.

Table 2. Replication of Frequently Mutated Genes in Independent Cohorts of Habit-Free Oral Cancer

| Gene Name | Habit-free oral cancer (n=42) | | Replication cohort 1[a] (n=10) | | Replication cohort 2[b] (n=10) | |
|---|---|---|---|---|---|---|
| | Variant class [c] | Mutation Frequency [d] | Mutation Frequency [d] | Variant class | Mutation Frequency [d] | Variant class |
| Known Oral Cancer Genes | | | | | | |
| *TP53* | Missense | 76.19 | 60 | Missense | 50 | Missense |
| *TTN* | Missense | 42.86 | 30 | Missense | 40 | Missense |
| *CASP8* | Truncating | 26.19 | 40 | Missense | 20 | Missense |
| *CDKN2A* | Truncating | 26.19 | 20 | Missense | 10 | Truncating |
| *FAT1* | Truncating | 23.81 | 10 | Truncating | 10 | Truncating |
| *HUWE1* | Missense | 19.05 | 10 | Missense | 20 | Missense |
| *KMT2D* | Missense | 19.05 | 20 | Missense | 10 | Silent |
| *NOTCH1* | Missense | 16.67 | 30 | Missense | 50 | Missense |
| *PCLO* | Missense | 16.67 | 30 | Missense | 0 | Nil |
| *PIK3CA* | Missense | 14.29 | 20 | Missense | 10 | Silent |
| *SYNE1* | Missense | 14.29 | 30 | Missense | 10 | Intronic |
| Mucin and associated genes identified in TCGA | | | | | | |
| *MUC16* | Missense | 23.81 | 40 | Missense | | Missense |
| *MUC5B* | Missense | 9.52 | 40 | Missense | | Silent |
| *HMCN1* | Missense | 9.52 | 20 | Missense | | Truncating |
| *AHNAK2* | Missense | 7.14 | 10 | Missense | | Missense |
| *HYDIN* | Missense | 9.52 | 10 | Missense | | Missense |
| *LRP1B* | Missense | 11.9 | 20 | Missense | | Missense |
| Other Mucin genes observed in the replication cohorts [e] | | | | | | |
| *MUC4* | NA | Nil | 20 | Intronic | | Missense |
| *MUC5AC* | NA | Nil | 10 | Missense | | Missense |
| *MUC3A* | NA | Nil | 10 | Missense | | Missense |
| *MUC15* | NA | Nil | 10 | Truncating | | Missense |
| *MUC17* | NA | Nil | 10 | Missense | | Missense |
| *MUC12* | NA | Nil | 10 | Missense | | NA |
| *MUC2* | NA | Nil | 10 | Missense | | NA |
| *MUC6* | NA | Nil | 20 | Missense | | NA |
| *MUC19* | NA | Nil | 30 | Missense | | NA |
| *MUC20* | NA | Nil | 10 | Missense | | NA |

[a], Previously published whole exome data accessed (PRJNA740146); [b], Previously published whole exome data accessed (PRJNA700466); [c], Variant class represents type of non-synonymous mutation, when mutation of corresponding gene was not noted, represented as NA; [d], denotes number of samples carrying non-synonymous mutations in each gene by the total number of samples. Nil when no variant observed; [e], Genes associated with mucins extracted from STRING network

reported to be associated with *MUC16*/*MUC5B*, such as *HMCN1, AHNAK2, HYDIN* and *LRP1B* were also targeted for mutation at higher rates in the replication series. Interestingly, additional mucin genes, such as *MUC4, MUC5AC, MUC3A, MUC15, MUC17* were also identified as frequent mutational targets in the replication studies examined. Further, 10 to 20% of tumors carried mutations in *MUC12, MUC2* and *MUC6* in any one of the two studies included for replication.

*Co-occurring and exclusive mutations in habit-free oral cancers*

Mutational co-occurrence plot in the 42 habit-fee OC are depicted in Figure 2. These analyses identified *TP53* and *FBXW7* as mutually exclusive mutations in this group. Only 5 tumors harboured *FBXW7* mutations that included the previously known hotspots at Arg479, Arg505 and Val475 deletion [37] (data not shown). Clinical and patient characteristics differed between *TP53* and *FBXW7* mutated tumors (Table 3). While *TP53* mutation carrying tumors showed a TMB of 2.56, *FBXW7* mutated tumors reflected a higher average TMB of 4.81 (p-value: 0.007). This translated to a significant association between *FBXW7* mutation status and high TMB, both in the univariate (OR, 95% CI: 1.88, 1.08- 3.27) and multivariate model (OR, 95% CI: 1.89, 0.98- 3.66). Further, *TP53* mutations carrying tumors were tongue cancers, a subsite of higher prevalence in this cohort (66%); while *FBXW7* mutated tumors were more common in buccal mucosa and related sites (80%) (p-value: 0.07), appeared to be a late event in oral cancer as noted only in advanced stage disease (p-value: 0.06), however did not translate to a

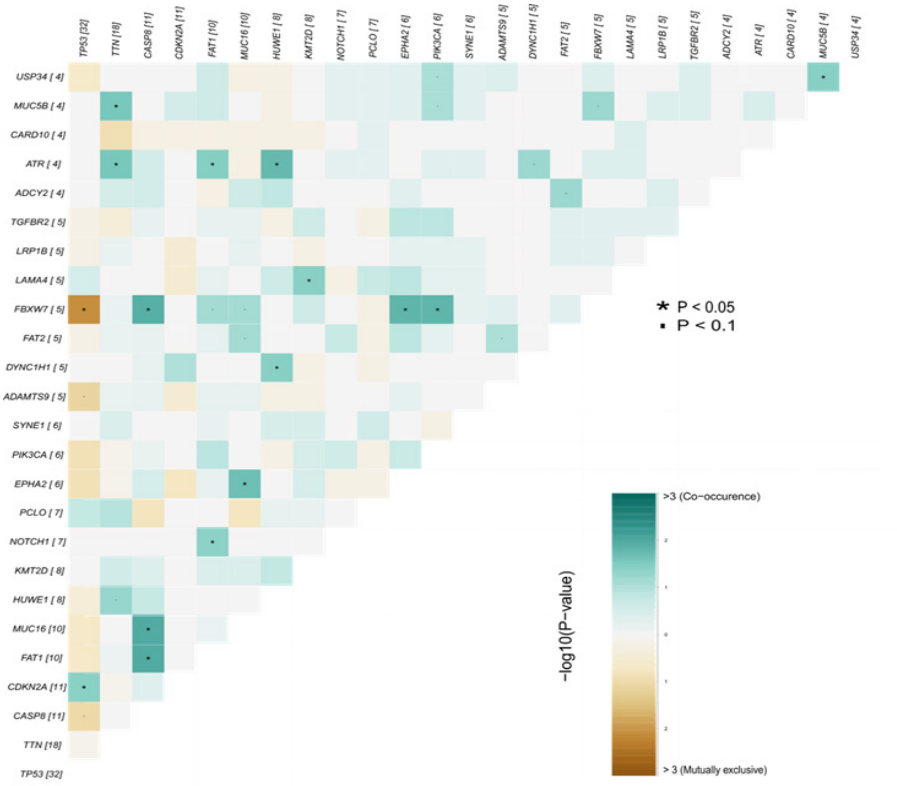


Figure 2. Mutational Co-Occurrence and Exclusivity Analysis in 42 Habit-Free OC Tumors. Pairwise heatmap visualizes the co-occurrence and mutual exclusivity of mutations among the top recurrently mutated genes across 42 habit free oral squamous cell carcinoma samples. Each cell indicates the statistical association between mutations in two genes, color-coded by the $-\log_{10}$ (P-value). Green indicates co-occurrence, and brown indicates mutual exclusivity. Significant interactions are denoted by asterisk.

Table 3. Patient Characteristics between *TP53* and *FBXW7* Mutated Tumors

| Characteristic | Descriptive (N) | *TP53* N=32 N (%) | *FBXW7* N=5 N (%) | P- value [a] |
|---|---|---|---|---|
| Sex | | | | 0.78 |
| | Male (11) | 8 (25.0) | 1 (20.0) | |
| | Female (31) | 24 (75.0) | 4 (80.0) | |
| Subsite | | | | 0.07 |
| | BM (19) | 12 (32.0) | 4 (80.0) | |
| | Tongue (24) | 20 (62.5) | 1 (20.0) | |
| Stage | | | | 0.06 |
| | I-II (16) | 14 (37.5) | 0 | |
| | III-IV (26) | 18 (56.3) | 5 (100.0) | |
| Vital status | | | | |
| | Alive (23) | 18 (56.3) | 1 (20.0) | |
| | Dead (19) | 14 (43.8) | 4 (80.0) | |
| | HR (95% CI) [b] | 0.94 (0.34-2.62) | 1.85 (0.61-5.63) | |
| | HR (95% CI) [c] | 1.04 (0.37-2.97) | 1.00 (0.30-3.35) | |
| TMB [d] (2.8) | | 2.56 | 4.81 | **<0.01** |
| | OR (95% CI)[e] | 0.74 (0.50- 1.10) | **1.88 (1.08- 3.27)** | |
| | OR (95% CI)[f] | 0.75 (0.48- 1.18) | 1.89 (0.98- 3.66) | |

[a], p-value derived from two sample test of proportions; [b], Hazards ratio from univariate Cox proportional models, associated p-values were 0.90 for *TP53* mutated tumors and 0.28 for *FBXW7* mutated tumors; [c], Hazards ratio from multivariate Cox proportional models adjusted for age, sex and early/ late stage at diagnosis, associated p-values were 0.94 for *TP53* mutated tumors and 0.99 for *FBXW7* mutated tumors; [d]TMB, Tumor Mutation Burden; [e], Odds ratio & 95% CI from univariate logistic regression models, associated p-values were 0.14 for *TP53* mutated tumors and 0.03 for *FBXW7* mutated tumors; [f], Odds ratio & 95% CI from multivariate logistic regression models adjusted for age, sex and early/ late stage at diagnosis, associated p-values were 0.75 for *TP53* mutated tumors and 0.06 for *FBXW7* mutated tumors. Bold represents statistically significant values.

difference in overall survival (HR, 95% CI for *TP53*: 0.94, 0.34-2.62 compared to HR, 95% CI for *FBXW7*: 1.85, 0.61-5.63). CNV profiling revealed that six *TP53*-mutated samples also harbored deletions encompassing the *FBXW7* locus (chr4:153,249,564–153,269,051) corresponding to the hemizygous loss of *FBXW7* (data not shown). In this group, *FBXW7* mutated tumors also carried mutations in *CASP8*, *EPHA2* and *PIK3CA* (p-value: ≤ 0.02) . Interestingly, *CASP8* (n=11), the third most frequently mutated gene in this group, exhibited high co-occurrence with *FAT1* and *MUC16* (p-value: 0.01) indicating at least 23% of the tumors carried *CASP8-FAT1-MUC16* combined mutations. Additionally, EPHA2-*MUC16* co-occurrence was also seen (p-value: 0.02). In this group, USP34-*MUC5B* mutations were completely collinear (p-value: 0.04) and distinct from all other mutations. These data show substantial co-occurrence of *MUC16*/*MUC5B* mutations with known oral cancer driver mutations. Among previously reported co-occurring mutations, we observed *CDKN2A*/ *TP53* were completely concurrent, *CASP8*/ *FAT1* and *FAT1*/ *NOTCH1* co-occurred in this population (p-value 0.04, 0.01, 0.04 respectively).

*Mutational processes in habit-free oral cancers*

The transition of C:G>T:A and T:A>C:G was contributed by 100% and 97% tumors, whereas transversion of T:A>G:C was contributed by 78% tumors. Overall, the average transition/transversion ratio across the cohort was 1.63 (median= 1.39, SD= 0.75), reflecting a predominance of transition mutations. Deconstructed somatic mutation profiles showed two prominent mutational signatures as shown in Figure 3. One associated with Apolipoprotein B mRNA Editing Catalytic Polypeptide (APOBEC) cytidine deaminase activity (SBS2) and the other with defective DNA mismatch repair (SBS6). APOBEC signature was present in 79% tumors with a cosine similarity of 0.73, while mismatch repair signature was found in 45% tumors with 0.86 cosine similarity contribution. Mutational profiling based on base substitutions confirmed the absence of smoking, chewing and alcohol signatures corresponding to SBS4 (7% representing 3 of 42 tumors showing >15% similarity), SBS29 (none of 42 tumors showing >15% similarity) and SBS16 (one of 42 tumors showing >15% similarity) respectively, thus validating the case inclusion based on clinical history of the participants.

*Gene expression profiles of habit-free oral cancer*

Tumor/ normal differential gene expression from transcriptome data of 42 tumors and 7 adjacent normal tissues were examined to identify the cancer subtype contribution, enriched pathways, immune fraction composition. Global differential gene expression between habit-free tumour and normal samples identified 1780 upregulated and 2319 downregulated genes. Figure S3A and B represent the top 100 differentially expressed genes in a heatmap and significantly altered genes in a volcano plot. These analyses confirmed earlier observations reporting differential expression of known oral cancer genes including *SLC2A1* ($log_2$FC: 1.9), *ITGA6* ($log_2$FC : 1.80), *LAMC2* ($log_2$FC: 3.7), *COL1A2*

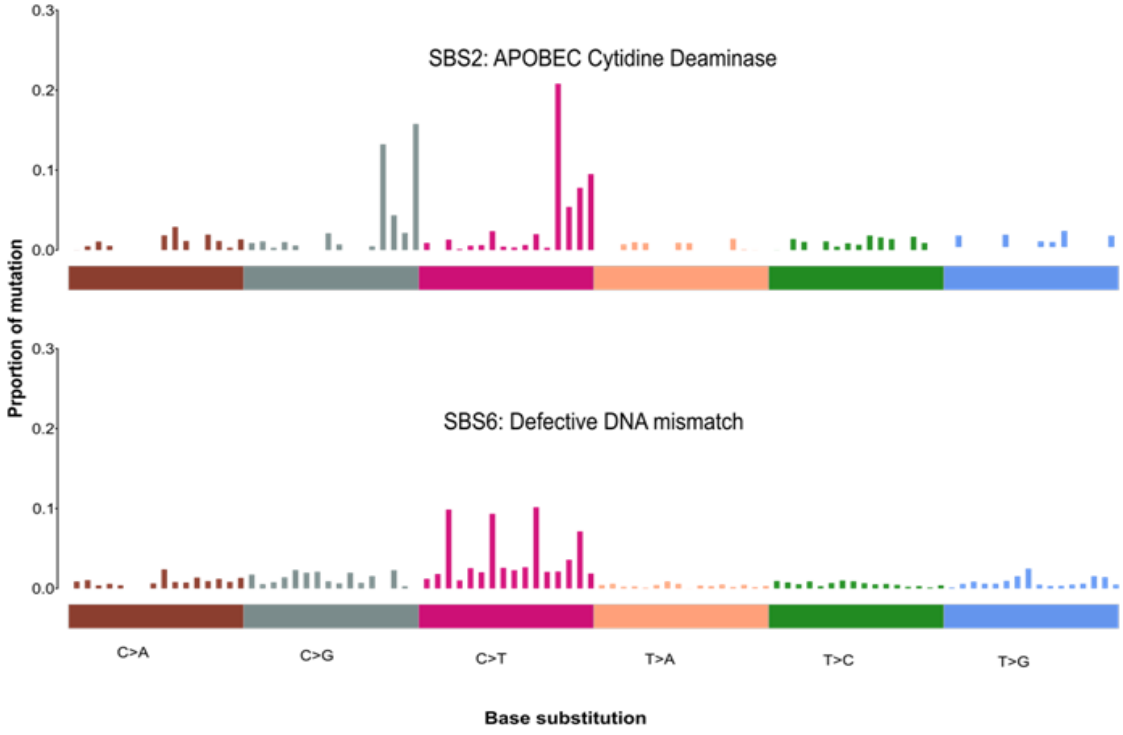


Figure 3. Mutational Processes Underlying Habit-Free Oral Cancers. Mutational signature analysis of habit-free oral tumors showing the extracted trinucleotide substitution patterns and their corresponding best matches to COSMIC single base substitution (SBS) signatures. The X-axis represents the proportion of mutation, and The Y-axis represents the seven base substitutions. Signature 1 reflects APOBEC-mediated mutagenesis (cosine similarity: 073), while signature 2 closely matched defective DNA mismatch repair (cosine similarity: 086).

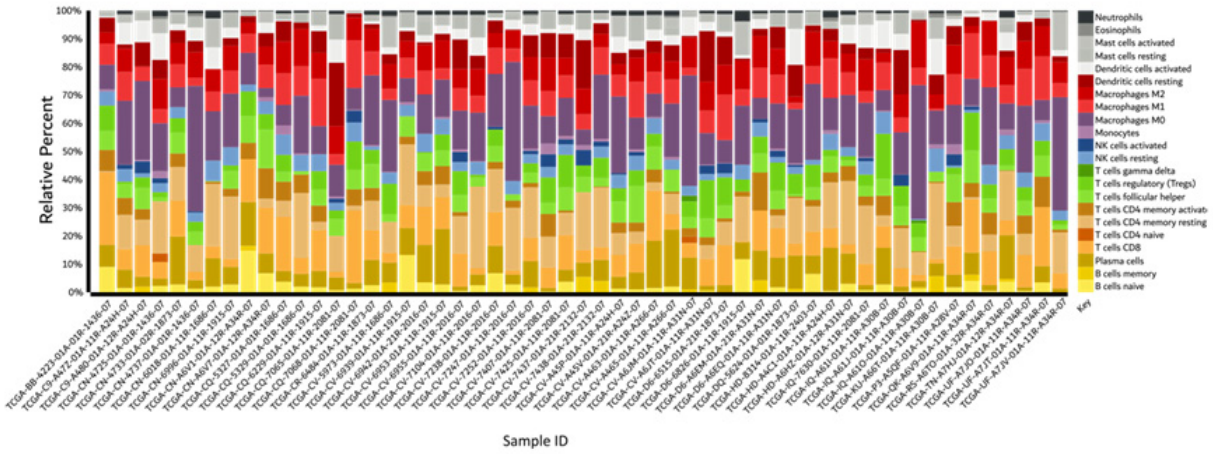


Figure 4. Immune Cell Composition of Habit-Free Oral Cancers. Stacked barplot illustrates the relative abundance of 22 immune cell types across 42 habit-free oral tumors. Each vertical bar corresponds to a single tumor sample, with colored segments representing the estimated proportions of distinct immune cell populations. Immune cell proportions were estimated from bulk RNA-seq expression profiles using the CIBERSORT deconvolution algorithm. The cumulative height of each bar represents 100% immune cell content per sample.

($log_2FC$ : 2.1), *COL1A1* ($log_2FC$ : 3.1), *TNC* ($log_2FC$: 2.28) and *CD276* $log_2FC$ : 1.7)[35]. Four of the five top differentially expressed genes (DEGs) in the habit free subset belonged to the MAGE family, MAGEB2 ($log_2FC$ = 8.6, FDR = 1.79511e-4), MAGEA12 ($log_2FC$ = 8.3, FDR = 2.6945e-4), and MAGEA3 ($log_2FC$ = 8.0, FDR = 2.78e-05) showed the highest fold changes. Conversely, KRTAP13-1 ($log_2FC$ = -12, FDR = 0.008), BPIFA1 ($log_2FC$ = -11, FDR = 2.01e-09), and KRT36 ($log_2FC$ = -10, FDR = 2.44e-32) were among the most downregulated genes. Gene Ontology based enrichment of biological process, cellular component, and molecular function analysis showed extracellular matrix reorganization, muscle structure, and transmembrane transporter activities (FDR < 0.001) were key biological processes relevant to habit-free OC (Figure S3C). As expected, human papillomavirus infection pathway, nicotine addiction and cytochrome p450 related xenobiotic metabolism pathways reflective of smoking/alcohol were not observed in this habit-free group.

Centroid based prediction of subtype contribution classified 69% of habit-free oral cancers as mesenchymal subtype (n=23) characterized by the expression of CD8A, ICOS, LAG3, HLA-DRA, VIM, ZEB1, ZEB2, MMP2, MMP9. A significant proportion of subjects also classified as basal subtype (n=20, 47%) based on HIF1A, CA9, VEGFA, EGFR, AREG, NRG1, CDH3, KRT1 and KRT9 expression (Figure S4).

Immune fraction deconvolution analysis showed presence of macrophages, T-regulatory cells, resting CD4+ and CD8+ T memory cells as the most common immune cell subtypes in habit-free OC (Figure 4).

*Functional relevance of MUC16/MUC5B genes in habit- free oral cancer*

Given the emergence of mucins as important mutational targets in this analysis, we examined their potential functional relevance by integrating available transcriptomic, copy number and methylation data in TCGA. These data are shown in Table S1. In general, both *MUC16* ($log_2FC$= -1.34, FDR= 0.19) and *MUC5B* ($log_2FC$= -4.01, FDR= 0.03) were downregulated in habit-free OC compared to normal tissues. To further elucidate the relevance of *MUC16*/*MUC5B* mutations, we examined the association of non-synonymous mutations in these genes with the corresponding expression. *MUC16* expression counts were lower in tumors carrying *MUC16* mutations compared to wildtype tumors (mean expression in *MUC16* mutated vs wildtype tumors was 0.18 vs 2.55), albeit not statistically significant. Further, the two tumors with predicted potential high impact mutations reflected the lowest *MUC16* expression levels (mean expression in tumors with *MUC16* high impact mutations= 0.05 compared to *MUC16* mutations of unknown relevance= 2.55). No association was noted between *MUC5B* mutations and its expression. We noted that chromosomal gains or losses at the *MUC16* and *MUC5B* locus were uncommon. Only four of the 42 habit-free tumors (<10%) showed copy number alterations at these loci and these were equally distributed as gains or losses. In the case of *MUC16*, three tumors showed copy number loss, and 1 tumor reflected amplification, while for *MUC5B* locus, 2 tumors each showed losses and gains. Similarly, neither gene body nor promoter methylation in *MUC16*/*MUC5B* was associated with corresponding gene expression counts.

In order to understand the cellular phenotype as the functional consequence of *MUC16*/*MUC5B* mutations, we examined their association with gene expression-based tumor subtype classification, TMB and overall survival (Table S2). While *MUC5B* mutations were equally distributed in tumors classified as basal or mesenchymal, based on their gene expression program, *MUC16* mutations were enriched in basal tumors. Tumors carrying *MUC16* mutations were less likely to express mesenchymal genes than basal subtype related genes (OR, 95% CI: 0.12, 0.02-0.73). This observation remained significant following adjustment for age, sex and stage at

diagnosis (OR, 95% CI: 0.12, 0.19-0.72). In this analysis, no association was noted between *MUC16* mutation and TMB compared (OR, 95% CI: 1.39, 0.90- 1.96). *MUC5B* mutations were associated with a significant increase in TMB (OR, 95% CI: 1.79, 1.02-3.15). These results were slightly attenuated upon further adjustment for age, sex and stage at diagnosis. We examined the clinical utility of *MUC16*/*MUC5B* gene mutations in predicting overall patient survival. Neither *MUC16*, nor *MUC5B* mutations were associated with all-cause mortality following OC diagnosis (OR, 95% CI for *MUC16*: 1.28, 0.46-3.56 and OR, 95% CI for *MUC5B:* 0.88, 0.20- 3.83), either in the univariate or multivariate models.

## Discussion

Tobacco smoking, chewing and alcohol consumption are strong risk factors for OC and are known to induce DNA damage and alter gene expression patterns [38, 39]. Little is known on the causation of oral cancers in the absence of these risk factors. We undertook a comprehensive analysis to uncover the mutational and gene expression changes associated with habit-free OC. Our results show that habit-free oral cancers are characterized by a lower tumor mutational burden, recurrent mutations in multiple mucin genes (particularly *MUC16* and *MUC5B*), reflect APOBEC and DNA mismatch repair as underlying mutational processes, show a preponderance towards expression of mesenchymal genes, enrichment of extra cellular matrix components genes and significant tumor infiltration by CD4+, CD8+ T cells and macrophages.

Majority of the frequently occurring mutations observed in this habit-free group have been reported in oral cancers previously, albeit with differing mutation frequencies [19-21] . These results confirm that a core set of mutational targets, potentially including *TP53*, *NOTCH1*, *CASP8*, *PIK3CA*, *FAT1* may act as drivers in oral cancer progression. This analysis also showed *MUC16* and *MUC5B* as frequent mutational targets in habit-free oral cancers. Importantly, these results are strengthened by replication in other independent habit-free OC datasets. Further, the identification of additional mucin gene members in the replication series could be indicative of their potential targeting in oral carcinogenesis. Mutations in the mucin gene family have been reported in many cancers, including head and neck cancer [40, 41]. Mucin gene mutations have been proposed as a marker for overall survival in cancer due to their association with higher TMB and improved immune cell infiltration in tumors [42, 43]. Although no direct association of *MUC16*/*MUC5B* was noted with overall survival, *MUC5B* mutated tumors had a higher overall mutation burden. Further, we observed that the habit-free tumors were enriched for tumor infiltrating CD4+, CD8+ T-cell subpopulations. Therefore, it is possible that mucin mutations in habit-free OC may result in a proinflammatory tumor microenvironment. If validated, this finding could have significant clinical implications for targeted therapy of habit-free oral cancers.

We show that *TP53* and *FBXW7* are mutually exclusive mutational targets in this patient cohort. In addition, there were 6 tumors carrying *TP53* mutations where FBXW7 was deleted (data not shown), suggestive potential co-inactivation with *TP53* in these tumors. Both *TP53* and *FBXW7* are established tumor suppressor genes and known to be mutational targets in multiple cancers [37, 44-47]. The oncogenic activity of mutant *P53* has been well characterized [46, 48-50]. Contrarily, *FBXW7* is the substrate recognition component of the ubiquitin ligase assembly known to degrade several cell fate regulating proteins, including p53 [37, 44, 45]. Mutations at Arg465 and Arg505in the *FBXW7* gene, also seen in this series of habit-free tumors, are known to abrogate its function thereby leading to *TP53* accumulation [37, 51]. Therefore, depending on the cellular context, both *TP53* and *FBXW7* can function as a tumor suppressor or an oncogene [45]. It is also important to note that these mutually exclusive mutations were uncommon in this cohort and require further validation in larger cohorts.

This analysis identified APOBECs, an enzyme family that catalyse cytosine deamination, as the predominant mutation signature in habit-free OC. APOBEC signatures are thought to be reflective of endogenous mutational processes, and have been previously reported in oral cancers [52] .These enzymes are known to contribute to mutation selection during the process of carcinogenesis [53] resulting in higher mutation rates of specific driver genes. APOBEC activity is also known to induce tumor immune signalling [54]. The identification of specific immune cell types is suggestive of tumor immune activation, however, the mutation rate of these habit-free OCs were lower than that known for OC in general, indicating that the contribution of APOBEC signatures in these cancers are likely to be complex.

In summary, our observations confirm several previously reported features in oral cancer, including the mutational targeting of *TP53*, *CAP8, CDKN2A, EPAH2, PIK3CA*, etc [19-21]. We report a significant proportion of habit-free OC tumors exhibit a mesenchymal gene expression program. Recent single cell RNA sequencing studies, in addition to demonstrating tumor heterogeneity, have confirmed the existence of a mesenchymal subset in oral squamous cell carcinoma [55-57]. Similar to our observations, tumor infiltration with monocytes, CD4+, CD8+ and T-regs, have also been reported in oral squamous cell carcinoma [58]. Interestingly, mutational processes involving APOBEC hyperactivity and defective DNA repair have been confirmed in oral cancer recently[58]. Taken together, our results suggest *MUC16*/*MUC5B* mutations could be potential drivers of oral cancer among habit-free subjects. We postulate that intricate interplay of mucin deregulation, mesenchymal transition, tumor immune microenvironment modification underlies tumorigenesis in habit-free individuals.

## Author Contribution Statement

SG: Data analysis, Visualization, Methodology & Writing. DA: Conceptualization, Methodology, Writing and Project supervision. All authors reviewed the final manuscript.

## Acknowledgements

We acknowledge TCGA, ICGC, and NCBI databases for the resources used in this analysis.

*Supplemental information*

Supplemental Information includes 4 Figures and 2 Tables.

This study did not involve the recruitment of human participants or animal subjects. Only publicly available data were used in this study. This manuscript contains no personal data or information.

*Declaration of interests*

The authors declare no competing interests.

*Research Funding*

This work was supported by funding from the Department of Biotechnology, Government of India (grant number BT/PR23140/MED/29/1260/2018). Ms. George was supported by a fellowship from the University Grants Commission (UGC), Government of India.

*Data/code availability*

All the data set used in this study are publicly available under TCGA HNSCC project (https://portal.gdc.cancer.gov/projects/TCGA-HNSC), ICGC ORCA INDIA project (http://www.icgc.org/) and NCBI SRA project PRJNA740146 and PRJNA700466. The analysis pipeline used in this study is available at https://github.com/sinumol/Habit-free-OC_Analysis_APJCP.

*Web resources*

https://tcga-data.nci.nih.gov/tcga
http://www.icgc.org/
https://www.cbioportal.org/
https://cancer.sanger.ac.uk/cosmic/signatures

## References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63. https://doi.org/10.3322/caac.21834.
2. Personal habits and indoor combustions. Volume 100 e. A review of human carcinogens. IARC Monogr Eval Carcinog Risks Hum. 2012;100(Pt E):1-538.
3. Bouvard V, Nethan ST, Singh D, Warnakulasuriya S, Mehrotra R, Chaturvedi AK, et al. Iarc perspective on oral cancer prevention. N Engl J Med. 2022;387(21):1999-2005. https://doi.org/10.1056/NEJMsr2210097.
4. Rumgay H, Nethan ST, Shah R, Vignat J, Ayo-Yusuf O, Chaturvedi P, et al. Global burden of oral cancer in 2022 attributable to smokeless tobacco and areca nut consumption: A population attributable fraction analysis. Lancet Oncol. 2024;25(11):1413-23. https://doi.org/10.1016/S1470-2045(24)00458-3.
5. IWGotEoCRt H. Biological agents. Volume 100 B. A review of human carcinogens. IARC Monogr Eval Carcinog Risks Hum. 2012;100(Pt B):1-441.
6. Anantharaman D, Marron M, Lagiou P, Samoli E, Ahrens W, Pohlabeln H, et al. Population attributable risk of tobacco and alcohol for upper aerodigestive tract cancer. Oral Oncol. 2011;47(8):725-31. https://doi.org/10.1016/j.oraloncology.2011.05.004.
7. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: Pooled analysis in the international head and neck cancer epidemiology consortium. J Natl Cancer Inst. 2007;99(10):777-89. https://doi.org/10.1093/jnci/djk179.
8. Hashibe M, Brennan P, Chuang SC, Boccia S, Castellsague X, Chen C, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: Pooled analysis in the international head and neck cancer epidemiology consortium. Cancer Epidemiol Biomarkers Prev. 2009;18(2):541-50. https://doi.org/10.1158/1055-9965.epi-08-0347.
9. World Health Organization, International Agency for Research on Cancer. Oral Cancer Prevention. IARC Handbooks of Cancer Prevention. Volume 19 [Internet]. France: IARC; 2023 [citado el 9 de diciembre 2025] [Internet].
10. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3(1):246-59. https://doi.org/10.1016/j.celrep.2012.12.008.
11. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21. https://doi.org/10.1038/nature12477.
12. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101. https://doi.org/10.1038/s41586-020-1943-3.
13. Tomczak K, Czerwinska P, Wiznerowicz M. The cancer genome atlas (tcga): An immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68-77. https://doi.org/10.5114/wo.2014.47136.
14. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333-9. https://doi.org/10.1038/nature12634.
15. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45(10):1113-20. https://doi.org/10.1038/ng.2764.
16. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67-73. https://doi.org/10.1038/nature12113.
17. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489(7417):519-25. https://doi.org/10.1038/nature11404.
18. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609-15. https://doi.org/10.1038/nature10166.
19. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in *NOTCH1*. Science. 2011;333(6046):1154-7. https://doi.org/10.1126/science.1206923.
20. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011;333(6046):1157-60. https://doi.org/10.1126/science.1208130.
21. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and

molecular subgroups. Nat Commun. 2013;4:2873. https://doi.org/10.1038/ncomms3873.
22. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell. 2004;5(5):489-500. https://doi.org/10.1016/s1535-6108(04)00112-6
23. Desai SS, K RR, Jain A, Bawa PS, Dutta P, Atre G, et al. Multidimensional mutational profiling of the indian hnscc sub-population provides irak1, a novel driver gene and potential druggable target. Front Oncol. 2021;11:723162. https://doi.org/10.3389/fonc.2021.723162.
24. Patel K, Bhat FA, Patil S, Routray S, Mohanty N, Nair B, et al. Whole-exome sequencing analysis of oral squamous cell carcinoma delineated by tobacco usage habits. Front Oncol. 2021;11:660696. https://doi.org/10.3389/fonc.2021.660696.
25. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747-56. https://doi.org/10.1101/gr.239244.118.
26. Ng PC, Henikoff S. Sift: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-4. https://doi.org/10.1093/nar/gkg509.
27. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17(1):122. https://doi.org/10.1186/s13059-016-0974-4.
28. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754-60. https://doi.org/10.1093/bioinformatics/btp324.
29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. https://doi.org/10.1101/gr.107524.110.
30. Wang K, Li M, Hakonarson H. Annovar: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. https://doi.org/10.1093/nar/gkq603.
31. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. String v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D13. https://doi.org/10.1093/nar/gky1131.
32. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the illumina infinium humanmethylation450 beadchip array. Epigenetics Chromatin. 2013;6(1):4. https://doi.org/10.1186/1756-8935-6-4.
33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome Biol. 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8.
34. Kanehisa M, Goto S. Kegg: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30. https://doi.org/10.1093/nar/28.1.27.
35. Keck MK, Zuo Z, Khattri A, Stricker TP, Brown CD, Imanguli M, et al. Integrative analysis of head and neck cancer identifies two biologically distinct hpv and three non-hpv subtypes. Clin Cancer Res. 2015;21(4):870-81. https://doi.org/10.1158/1078-0432.CCR-14-2481.
36. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with cibersort. Methods Mol Biol. 2018;1711:243-59. https://doi.org/10.1007/978-1-4939-7493-1_12.
37. Akhoondi S, Sun D, von der Lehr N, Apostolidou S, Klotz K, Maljukova A, et al. Fbxw7/hcdc4 is a general tumor suppressor in human cancer. Cancer Res. 2007;67(19):9006-12. https://doi.org/10.1158/0008-5472.CAN-07-1320.
38. Hecht SS. Cigarette smoking and lung cancer: Chemical mechanisms and approaches to prevention. Lancet Oncol. 2002;3(8):461-9. https://doi.org/10.1016/s1470-2045(02)00815-x.
39. Hecht SS, Hatsukami DK. Smokeless tobacco and cigarette smoking: Chemical mechanisms and cancer prevention. Nat Rev Cancer. 2022;22(3):143-55. https://doi.org/10.1038/s41568-021-00423-4.
40. Hollingsworth MA, Swanson BJ. Mucins in cancer: Protection and control of the cell surface. Nat Rev Cancer. 2004;4(1):45-60. https://doi.org/10.1038/nrc1251.
41. Liu B, Hu FF, Zhang Q, Hu H, Ye Z, Tang Q, et al. Genomic landscape and mutational impacts of recurrently mutated genes in cancers. Mol Genet Genomic Med. 2018;6(6):910-23. https://doi.org/10.1002/mgg3.458.
42. Yang Y, Zhang J, Chen Y, Xu R, Zhao Q, Guo W. Muc4, *MUC16*, and ttn genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pan-cancer. Clin Transl Med. 2020;10(4):e155. https://doi.org/10.1002/ctm2.155.
43. Wang Z, Hou H, Zhang H, Duan X, Li L, Meng L. Effect of *MUC16* mutations on tumor mutation burden and its potential prognostic significance for cutaneous melanoma. Am J Transl Res. 2022;14(2):849-62.
44. Tan Y, Sangfelt O, Spruck C. The fbxw7/hcdc4 tumor suppressor in human cancer. Cancer Lett. 2008;271(1):1-12. https://doi.org/10.1016/j.canlet.2008.04.036.
45. Galindo-Moreno M, Giraldez S, Limon-Mortes MC, Belmonte-Fernandez A, Saez C, Japon MA, et al. P53 and fbxw7: Sometimes two guardians are worse than one. Cancers (Basel). 2020;12(4). https://doi.org/10.3390/cancers12040985.
46. Olivier M, Hollstein M, Hainaut P. *TP53* mutations in human cancers: Origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010;2(1):a001008. https://doi.org/10.1101/cshperspect.a001008.
47. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of *TP53* mutations in human cancer. Nat Genet. 2018;50(10):1381-7. https://doi.org/10.1038/s41588-018-0204-y.
48. Mantovani F, Collavin L, Del Sal G. Mutant p53 as a guardian of the cancer cell. Cell Death Differ. 2019;26(2):199-212. https://doi.org/10.1038/s41418-018-0246-9.
49. Chiang YT, Chien YC, Lin YH, Wu HH, Lee DF, Yu YL. The function of the mutant p53-r175h in cancer. Cancers (Basel). 2021;13(16). https://doi.org/10.3390/cancers13164088.
50. Baslan T, Morris JPt, Zhao Z, Reyes J, Ho YJ, Tsanov KM, et al. Ordered and deterministic cancer genome evolution after p53 loss. Nature. 2022;608(7924):795-802. https://doi.org/10.1038/s41586-022-05082-5.
51. Fan J, Bellon M, Ju M, Zhao L, Wei M, Fu L, et al. Clinical significance of fbxw7 loss of function in human cancers. Mol Cancer. 2022;21(1):87. https://doi.org/10.1186/s12943-022-01548-2.
52. Ghosh A, Das C, Ghose S, Maitra A, Roy B, Majumder PP, et al. Integrative analysis of genomic and transcriptomic data of normal, tumour, and co-occurring leukoplakia tissue triads drawn from patients with gingivobuccal oral cancer identifies signatures of tumour initiation and progression. J Pathol. 2022;257(5):593-606. https://doi.org/10.1002/path.5900.
53. Faden DL, Thomas S, Cantalupo PG, Agrawal N, Myers

J, DeRisi J. Multi-modality analysis supports apobec as a major source of mutations in head and neck squamous cell carcinoma. Oral Oncol. 2017;74:8-14. https://doi.org/10.1016/j.oraloncology.2017.09.002.
54. Cannataro VL, Gaffney SG, Sasaki T, Issaeva N, Grewal NKS, Grandis JR, et al. Apobec-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. Oncogene. 2019;38(18):3475-87. https://doi.org/10.1038/s41388-018-0657-6.
55. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell. 2017;171(7):1611-24 e24. https://doi.org/10.1016/j.cell.2017.10.044.
56. Kim S, Kee HJ, Kim D, Jang J, Jeong HO, Sim NS, et al. Multiregional single-cell transcriptomics reveals an association between partial emt and immunosuppressive states in oral squamous cell carcinoma. iScience. 2025;28(9):112988. https://doi.org/10.1016/j.isci.2025.112988.
57. Chen J, Yang J, Li H, Yang Z, Zhang X, Li X, et al. Single-cell transcriptomics reveal the intratumoral landscape of infiltrated t-cell subpopulations in oral squamous cell carcinoma. Mol Oncol. 2021;15(4):866-86. https://doi.org/10.1002/1878-0261.12910.
58. Ghosh A, Singh S, Mallick TR, Chakravarty S, Varsha Bhagat S, Das C, et al. Comprehensive genomic and digital pathology profiling of tobacco-chewer female oral cancer patients simultaneously with integration of single-cell datasets identifies clinically actionable patient subgroups. Clin Transl Med. 2025;15(7):e70386. https://doi.org/10.1002/ctm2.70386.