CANCER GENOMICS & PROTEOMICS *22*: 188-207 (2025)
doi: 10.21873/cgp.20496

Characterization of the Genomic Landscape in HPV-positive Cervical and Head and Neck Squamous Cell Carcinomas by Whole Genome Next Generation Sequencing

JIANLAN REN[1], NIAN MA[2], TYLER SECKAR[3], SHEYNAZ BASSA[4], NICOLA ZETOLA[5], SURBHI GROVER[6], ZHI WEI[1] and ERLE ROBERTSON[2]

[1]Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, U.S.A.;
[2]Department of Otorhinolaryngology-Head and Neck Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A.;
[3]Department of Cancer Biology, George Washington School of Medicine, George Washington University, Washington, DC, U.S.A.;
[4]Department of Radiation Oncology, Steve Biko Academic Hospital, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa;
[5]Division of Pulmonary Medicine, Department of Medicine, Augusta College of Medicine, Augusta, GA, U.S.A.;
[6]Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A.

Abstract

Background/Aim: In this study, we provide a comprehensive characterization of HPV-positive primary cervical cancers (CC) and HPV-positive head and neck squamous cell carcinomas (HNSCC) through whole genome next-generation sequencing. Human papillomavirus (HPV) infection, recognized as a definitive human carcinogen, is increasingly acknowledged for its role in development of human cancers. HPV-driven cervical cancers are among the leading causes of cancer-related deaths worldwide, while HPV-driven head and neck cancers exhibit distinct biological and clinical characteristics. Recent data has provided convincing evidence that HPV-related cervical cancer, like HPV head and neck cancer also predict better outcomes, with viral integration patterns further predicting disease related outcomes.

Materials and Methods: We designed an experimental study that encompasses four pairs of HPV-positive patient samples with controls, utilizing state-of-the-art Next Generation Sequencing (NGS) technology including whole genome sequencing, transcriptome sequencing and virus integration.

Results: Multiple mutated genes, including TTN, COL6A3, and FLNA, were identified shared between CC and HNSCC. Additionally, we observed a notable proportion of pathways affected by oncogenic alterations, particularly in the RTK-RAS and NOTCH pathways, in both CC and HNSCC. Furthermore, we discovered a shared down-regulation of

Erle Robertson, Department of Otorhinolaryngology-Head and Neck Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A. Tel: +1 2157460114, email: erle@pennmedicine.upenn.edu

Received November 26, 2024 | Revised December 24, 2024 | Accepted January 10, 2025

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
©2025 The Author(s). Anticancer Research is published by the International Institute of Anticancer Research.

the Hedgehog signaling pathway based on transcriptome expression analysis in KEGG. We also identified RUNX2 and TFPI as sites of virus integration, and upstream as well as downstream pathway modulators, and represent potential targets for therapeutic interventions.

Conclusion: Overall, this study showed a thorough comparison between CC and HNSCC from multiple aspects, including gene variations, oncogenic pathways, KEGG enrichment and virus integration sites. However, further studies, which involve larger patient cohorts should be undertaken to further support these findings.

Keywords: HPV-positive cancers, cervical cancer (CC), head and neck squamous cell carcinoma (HNSCC), whole genome sequencing, next-generation sequencing (NGS), viral integration.

Introduction

Human papillomavirus (HPV) is a diverse family of viruses with numerous distinct types capable of infecting humans (1, 2). While many HPV types are relatively benign, several have been identified as significant causative agents in various forms of cancers (3). HPV-related cancers account for 4.5% of all cancers worldwide (4). Among these, HPV is a primary etiological factor in cervical carcinoma (CC), one of the most prevalent and life-threatening malignancies affecting women globally (5). Cervical cancer is the eighth most common cancer worldwide, with more than 600,000 new cases diagnosed annually, predominantly in regions such as Sub-Saharan and East Africa (6). High-risk HPV types, particularly 16 and 18, are established as the leading causes of nearly all cervical cancers (7). Beyond the cervix, HPV has also been implicated in head and neck squamous cell carcinoma (HNSCC), originating from various anatomical sites within the head and neck region (8). Notably, HPV infection has overtaken smoking and alcohol as the primary risk factor for oropharyngeal cancers (9), which are more frequently diagnosed in males (10). However, recent trends indicate a rising incidence of these cancers among females. The true prevalence of HPV-related head and neck cancers remains uncertain, primarily due to inconsistencies in testing practices (4). Despite this, cervical cancer and head and neck cancers are recognized as the most common HPV-related malignancies. Several studies have been conducted to characterize the genomic landscapes of both CC and HNSCC (9, 11). These efforts are driven by the significant differences in clinical outcomes between HPV-related and non-HPV-related head and neck cancers, while nearly all cervical cancers are HPV-related (12, 13).

A study of 228 primary cervical cancer cases revealed notable APOBEC mutagenesis patterns, with *SHKBP1*, *ERBB3*, *CASP8*, *HLA-A*, and *TGFBR2* identified as significantly mutated genes in cervical cancer (7). Amplifications of immune targets, including *CD274* (commonly known as PD-L1) and *PDCD1LG2* (commonly known as *PD-L2*), were also observed, along with the discovery of *BCAR4*, a long non-coding RNA associated with response to lapatinib (7). HPV integration was detected in all HPV18-related samples, while 76% of HPV16-related samples were linked to structural aberrations and increased expression of target genes (7). Additionally, a unique subset of endometrial-like cervical cancers, predominantly HPV-negative, was identified. These tumors exhibited relatively high frequencies of *KRAS*, *ARID1A*, and *PTEN* mutations (11).

The Cancer Genome Atlas (CGA) profiled 279 HNSCCs to provide a comprehensive landscape of somatic genomic alterations. The study identified key distinctions in HPV-associated tumors, highlighting frequent helical domain mutations in the oncogene *PIK3CA*, novel alterations involving loss of *TRAF3*, and amplification of the cell cycle gene *E2F*. In contrast, smoking-related HNSCCs exhibited widespread *TP53* loss-of-function mutations and *CDKN2A* inactivation, alongside common copy number alterations like amplification in 3q26/28 and 11q13/22 (14). Notably,

a subset of oral cavity tumors with favorable prognoses showed limited copy number variations but harbored activating mutations in *HRAS* or *PIK3CA*, combined with inactivating mutations of *CASP8*, *NOTCH1*, and *TP53*. Additional subgroups featured alterations such as *NSD1* inactivation, WNT pathway disruptions involving *AJUBA* and *FAT1*, and *NFE2L2* activation – primarily observed in laryngeal cancers (14).

Besides, alterations in the Hedgehog (*HH*) signaling pathway have been found to correlate with local recurrence of cervical cancer and, head and neck cancer progression (15, 16). The HH pathway appears to be one of the HPV-mediated up-regulated gene pathways, and the expression of the *E7* oncogene is associated with advanced stage of cervical cancer as well as HH pathway overexpression (17, 18).

Notably, the underlying mechanisms behind HPV-positive cancers remain unclear, and the relationship between HPV-positive CC and HPV-positive HNSCC remains to be more carefully explored (19). Therefore, there is a compelling need to unravel the shared and unique aspects of these two malignancies for better treatment strategies. To address this gap in our understanding, we analyzed and compared the oral and cervical genomes in HPV-positive samples to provide insights into the similarities and differences between HPV-associated CC and HNSCC.

Despite advancements in understanding HPV-associated cancers, the mechanisms underlying HPV-positive CC and HNSCC remain inadequately defined, and the interplay between these malignancies is not fully elucidated. This study aims to address these gaps by leveraging state-of-the-art Next Generation Sequencing (NGS) technologies, including whole genome sequencing (WGS) and transcriptome sequencing, to comprehensively compare the genomic landscapes of HPV-positive CC and HNSCC. By identifying shared and distinct molecular features, this research seeks to enhance our understanding of the pathophysiology of these cancers and to uncover potential targets for improved treatment strategies.

Materials and Methods

Sample preparation. Tumor and control samples are from deidentified pathology samples from the Pathology Laboratory at the Ministry of Health Botswana under IRB approval number (#820159). The sequencing data generated and analyzed in the study came from two sources, HiSeq3000 1×50 and MiSeq 2×150, respectively. The HiSeq3000 1×50 data included 12 oral and cervical specimens, 4 for each cancer type totaling 8 samples altogether, and 2 controls each with a total of 4 controls altogether. All 12 samples were pooled into a single lane for HiSeq and each sample was dual indexed. Six oral specimens included 4 tumor samples and 2 control samples, while 6 cervical specimens included 4 tumor samples and 2 control samples. For MiSeq 2×150, the data corresponds to CC and HNSCC specimens used for transcriptomics analyses. A total of 4 samples for each cancer type and 2 controls each were used for the transcriptomics RNA-Sequencing analyses. The tumor samples were run together for sequencing, and the two control samples were run together for sequencing. Therefore, for each cancer type there are six samples, four tumors and two controls, which were sequenced together for whole genome DNA-sequencing and similarly for RNA-Sequencing. The entire program from sample preparation to data analysis is shown in a schematic in Figure 1.

DNA sequencing and mutation calling. Whole genome data pre-processing was conducted in accordance with the GATK Best Practices using GATK 4.0 (20). Briefly, aligned BAM files were processed by read group, with alignments and attributes removed using 'RevertSam' to generate one unaligned BAM (uBAM) file per group (20). Uniform read groups were then assigned using 'AddOrReplace Readgroups'. These uBAM files were converted to interleaved FASTQ format with 'SamToFastq' and underwent quality control *via* 'FastQC.' Adapter sequences were identified and trimmed using 'Trim_galore'. The cleaned FASTQ files were aligned to the b37 genome with 'BWA MEM,' followed by attribute

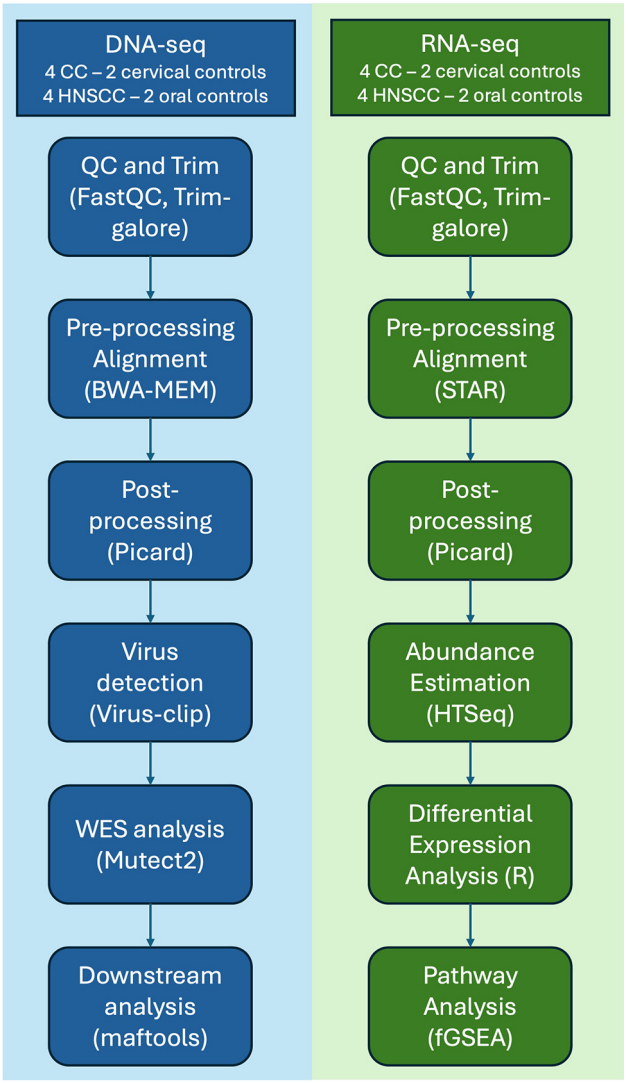

Figure 1. *A schematic showing the work pipeline. The tumors and controls cohorts and next generation sequencing and analysis pipeline. DNA-seq analysis shown of the left and RNA-seq analysis is shown on the right.*

restoration through 'MergeBamAlignment'. Aligned BAM files from multiple read groups were merged, and PCR/optical duplicates marked using 'MarkDuplicates.' Next, base recalibration was performed using 'BaseRecalibrator' and 'ApplyBQSR'. Coverage metrics were generated with 'CollectHsMetrics,' while alignment quality was assessed using 'Validate SamFile' and summarized with 'MultiQC'. Finally, 'Crosscheck Fingerprints' verified sample integrity by confirming all read groups belonged to the same individual, ensuring consistency across samples. Any identified mismatches were flagged and excluded from further analysis (20). The mapping qualities of the samples are summarized in Supplementary Table I. The average reads per sample is approximately 5 million and uniquely mapped reads percentage is 65% on average for Miseq 2×150 sequenced samples as shown in Table I.

Variant detection followed the GATK Best Practices using GATK4. Germline variants were identified from control samples with Mutect2 in artifact-detection mode and compiled into a cohort-wide panel of normal samples. Somatic variants in tumor samples were detected using Mutect2 in matched normal mode, incorporating parameters such as the matched normal sample, reference FASTA, panel of normal variants, and gnomAD germline resources for additional filtering. Cross-sample contamination was assessed with 'GetPileupSummaries' and 'CalculateContamination' for both tumor and matched normal samples. Read orientation artifacts were analyzed using 'CollectF1R2Counts' and modeled through 'LearnReadOrientationModel.' Finally, additional filtering, including artifact-in-normal and contamination thresholds, was applied using 'FilterMutectCalls'. The filtering strategies are default settings with allele frequency in tumor at a minimum frequency of 5% or higher. In addition, gnomAD was used to filter out common germline variants from the somatic calls, with a threshold of 1% allele frequency in population datasets.

RNA-seq pre-processing. RNA data preprocessing followed the mRNA analysis pipeline established by the National Cancer Institute (21). Quality control was first performed on raw FASTQ files using 'FastQC,' and low-quality bases were trimmed with 'Trim_galore'. The cleaned FASTQ files were aligned to the b37 genome using 'STAR' to produce BAM files. Subsequent quality assessment of these BAM files utilized 'Collect RNASeqMetrics,' generating metrics that detailed base distribution across transcripts.

Table I. *Mapping quality shown for the sequencing reads of the CC and HNSCC samples to hg38.*

Metrics	Cervical-Control	Cervical-Tumor	Oral-Control	Oral-Tumor
Number of input reads |	6.773.956	3.189.636	6.799.184	3.830.650
Average input read length |	100	125	104	114
UNIQUE READS:				
Uniquely mapped reads number |	4.911.230	2.862.478	5.907.695	3.007.843
Uniquely mapped reads % |	72.50%	89.74%	86.89%	78.52%
Average mapped length |	95.52	121.67	101.53	111.43
Number of splices: Total |	69.751	18.574	39.496	29.174
Number of splices: Annotated (sjdb) |	15.697	2.955	558	6.557
Number of splices: GT/AG |	40.899	10.709	21.532	16.951
Number of splices: GC/AG |	2.940	675	1.729	984
Number of splices: AT/AC |	100	49	66	38
Number of splices: Non-canonical |	25.812	7.141	16.169	11.201
Mismatch rate per base, % |	1.57%	1.11%	1.26%	1.24%
Deletion rate per base |	0.07%	0.06%	0.06%	0.06%
Deletion average length |	2	1.77	1.79	1.82
Insertion rate per base |	0.02%	0.02%	0.01%	0.01%
Insertion average length |	2.07	1.91	1.86	1.94
MULTI-MAPPING READS:				
Number of reads mapped to multiple loci |	688.113	177.029	406.183	359.142
% of reads mapped to multiple loci |	10.16%	5.55%	5.97%	9.38%
Number of reads mapped to too many loci |	52.436	13.128	57.741	23.077
% of reads mapped to too many loci |	0.77%	0.41%	0.85%	0.60%
UNMAPPED READS:				
Number of reads unmapped: too many mismatches |	0	0	0	0
% of reads unmapped: too many mismatches |	0.00%	0.00%	0.00%	0.00%
Number of reads unmapped: too short |	1.095.410	128.148	388.378	423.852
% of reads unmapped: too short |	16.17%	4.02%	5.71%	11.06%
Number of reads unmapped: other |	26.767	8.853	39.187	16.736
% of reads unmapped: other |	0.40%	0.28%	0.58%	0.44%
CHIMERIC READS:				
Number of chimeric reads |	0	0	0	0
% of chimeric reads |	0.00%	0.00%	0.00%	0.00%

CC: Cervical cancer; HNSCC: head and neck squamous cell carcinoma.

Differential gene expression analysis of RNAseq data. After alignment, BAM files were analyzed using the RNA Expression Workflow to quantify RNA expression levels. Gene-level read counts were obtained with 'HTSeq-Count' (22). These raw counts were normalized using 'DESeq2,' which applies a negative binomial distribution model and its own statistical algorithm. DESeq2 outputs included base means across samples, log2 fold changes, standard errors, test statistics, p-values, and adjusted *p*-values. Significant genes were visualized using a volcano plot to highlight differential expression patterns.

Gene set enrichment analysis. Pathway analysis was performed using 'fgsea', a fast implementation of preranked gene set enrichment analysis (GSEA) (23). Significant genes were ranked and analyzed alongside the KEGG pathway dataset as inputs. 'fgsea' produced outputs including pathway names, enrichment scores, normalized

Table II. *Somatic mutation types identified in CC and HNSCC tumor samples.*

	Cervical			Oral		
ID	Summary	Mean	Median	Summary	Mean	Median
NCBI_Build	hg38	NA	NA	Hg38	NA	NA
Center	NA	NA	NA	NA	NA	NA
Samples	4	NA	NA	4	NA	NA
nGenes	1.617	NA	NA	1.049	NA	NA
Frame_Shift_Del	118	29.5	24	27	6.75	4.5
Frame_Shift_Ins	5	1.25	1.5	4	1	1
In_Frame_Del	22	5.5	5.5	3	0.75	0.5
Missense_Mutation	2.702	675.5	704	1.402	350.5	439
Nonsense_Mutation	185	46.25	47	103	25.75	24.5
Nonstop_Mutation	5	1.25	1.5	3	750.75	1
Splice_Site	92	23	20	55	13.75	11.5
Total	3.129	782.25	815	1.597	399.25	497

CC: Cervical cancer; HNSCC: head and neck squamous cell carcinoma. Sum, mean and median somatic rates are shown in separate columns for both cervical and oral cancers.

enrichment scores (NES), and adjusted *p*-values. Results were prioritized based on the NES values.

Virus integration sites. Virus integration detection was conducted using 'Virus-Clip', a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability (24). First the DNA-seq was aligned to both virus reference genome using BWA-MEM and human reference genome hg38 using BLASTN. Then we extracted soft-clipped reads (potentially containing sequences of virus-integrated human loci) using SAMTools (25). All integration sites were automatically annotated with the affected human genes and their corresponding gene regions using ANNOVAR (26).

Protein-protein interaction. We applied STRING (version 12) for protein-protein interaction analysis, a database of known and predicted protein-protein interactions (27). The interactions encompass both direct (physical) and indirect (functional) associations. These relationships are derived from computational predictions, knowledge transfer across species, and data compiled from primary interaction databases.

Statistical analysis. All statistical analyses were conducted in R version 4.1, and data visualization was performed using Maftools version 2.22 (28). Results were considered statistically significant at a p-value <0.05 in differential expression analysis and pathway enrichment analysis. Statistical significance was determined based on a false discovery rate (FDR) threshold of 0.05 using the Benjamini-Hochberg correction (29).

Results

Frequently of mutated somatic genes of HPV-positive CC and HNSCC. In the study cohort, genomic analysis identified 3129 somatic variations (including SNVs and small indels) in CC, and 1597 somatic variations in HNSCC. Somatic mutation types were summarized in Table II.

Of the somatic variations, multiple genes have 100% somatic mutation enrichment, which is likely due to the size of the samples as shown in Figure 2. The average number of mutations including single nucleotide variants (SNVs) and small insertions and deletions (INDELs) are 657 and 490 for CC and HNSCC, respectively. Among the top enriched genes of CC and HNSCC, many of them have been reported to be

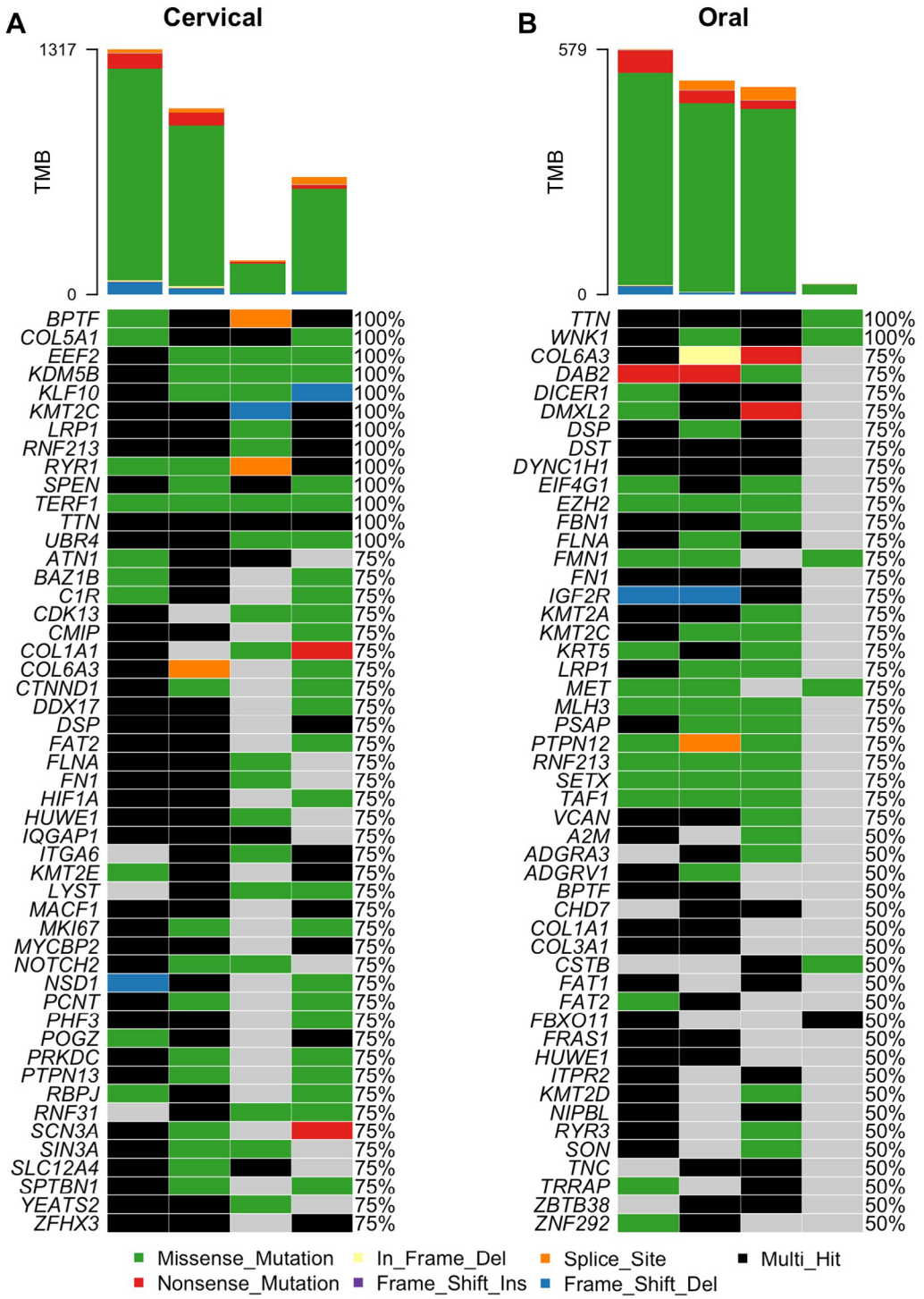

Figure 2. *Mutational oncoplots of most frequently mutated genes with driver mutations in HPV-positive cervical cancer (CC) (A) and head and neck squamous cell carcinomas (HNSCC) (B). The oncoplots show a summary of the most frequent driver mutations of the CC and HNSCC cohorts based on the exome gene list. The oncoplots show the patients in a horizontal orientation and the gene and the corresponding driver mutations per patient in the vertical orientation. The upper area is the tumor mutation burden including all somatic single nucleotide variants (SNVs) and small insertions and deletions (INDELs) per tumor.*

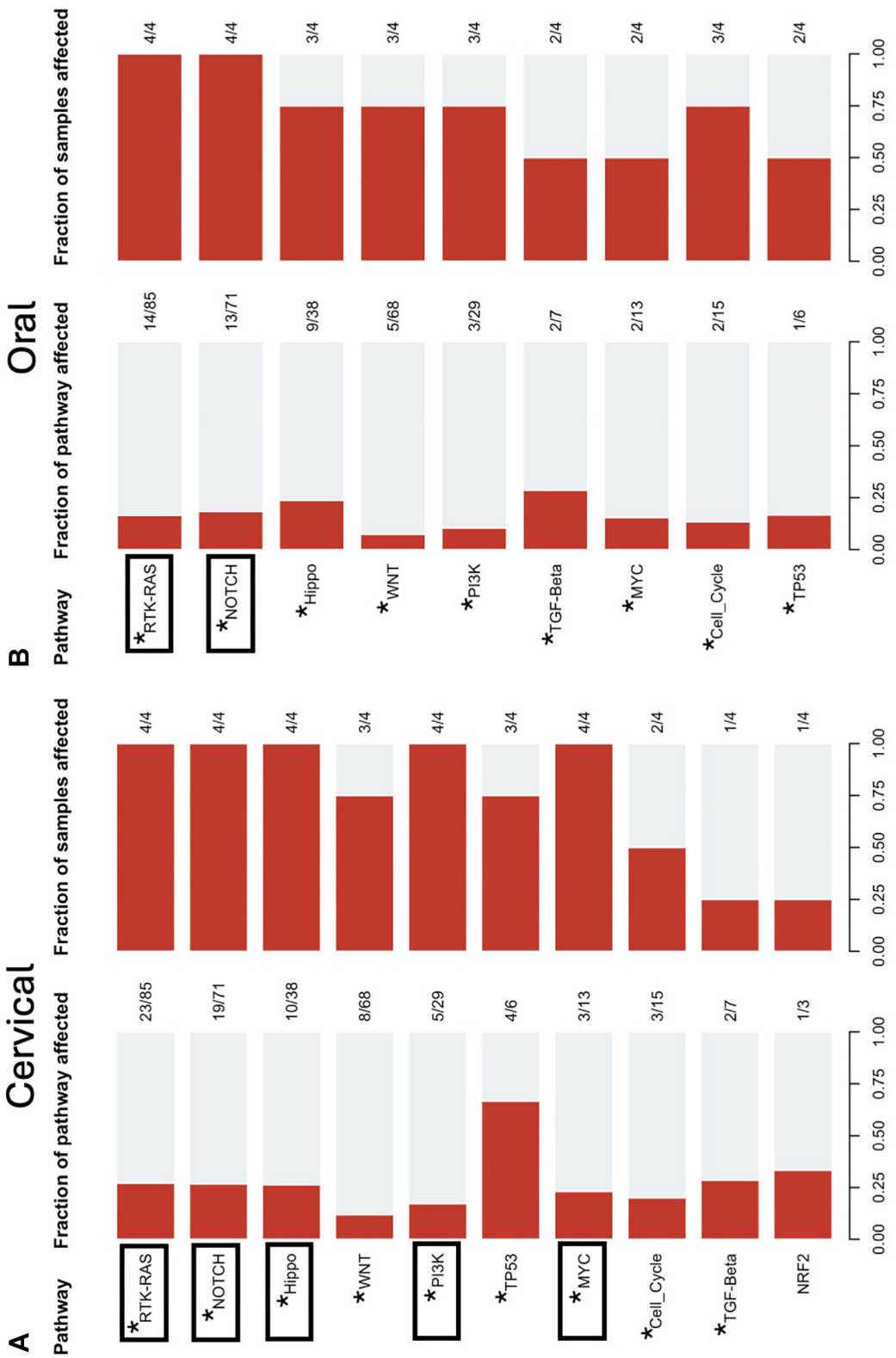

Figure 3. *Continued*

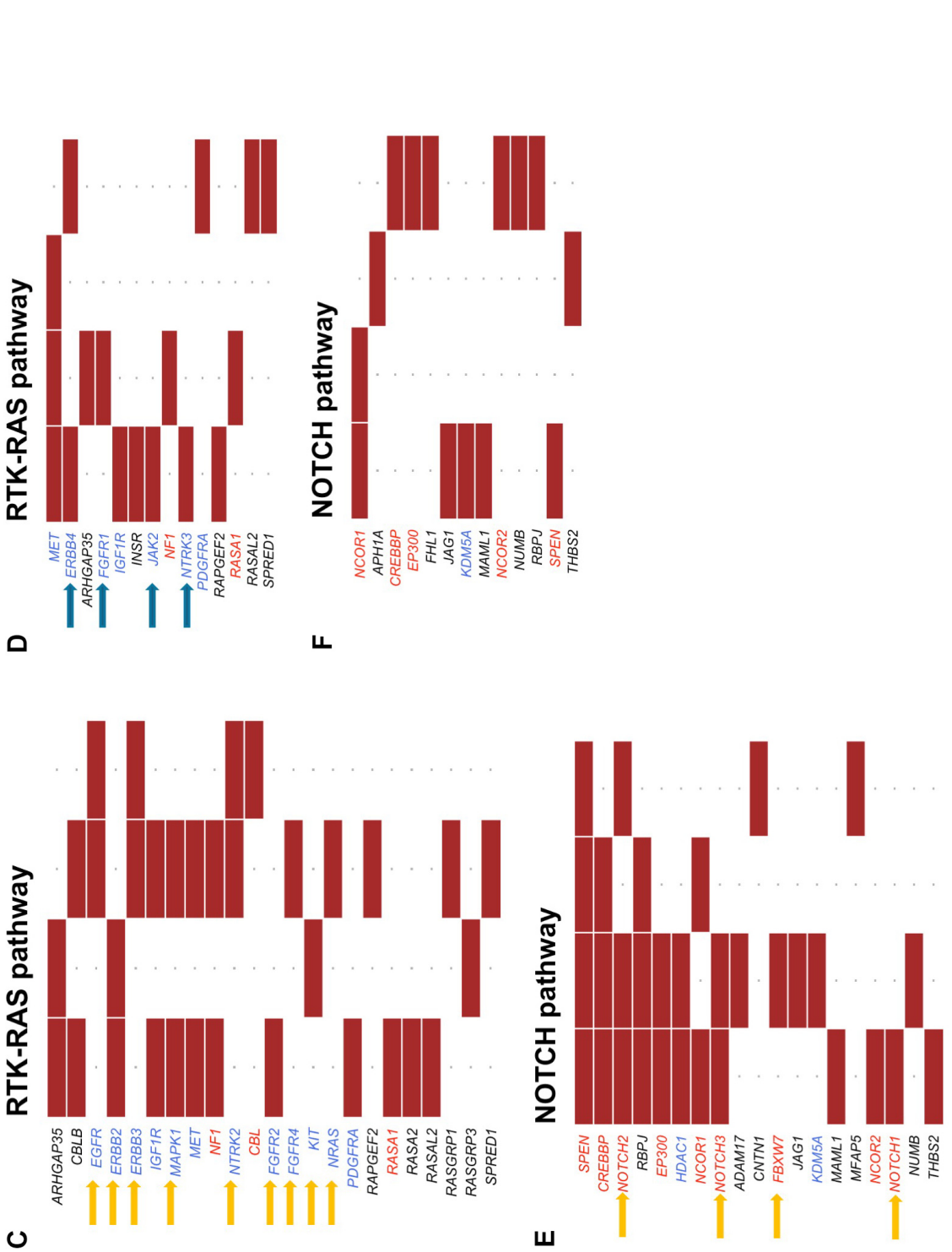

Figure 3. *Overview of oncogenic pathway alterations in the tumor cohort. (A-B) left plots show the oncogenic pathways affected in the HPV-positive cervical cancer (CC) and head and neck squamous cell carcinomas (HNSCC) patients, with the number of affected genes in relation to the total number of genes linked to each pathway. A-B right plots display the fraction of tumors bearing mutations in a given pathway. The squared pathways have a 100% fraction of samples affected; the asterisk marked pathways are found in both CC and HNSCC. (C-F) For each pathway, Maftools allows visualization of the mutated genes for each tumor sample to easily spot the recurrently mutated genes. Oncogenes and tumor suppressor genes are indicated in blue and red, respectively. The mutations exclusively mutated in CC and HNSCC are marked by yellow and blue arrows, respectively.*

significantly associated with cancers. The top genes with 100% significant mutations in CC were *BPTF*, *COL5A1*, *EEF2*, *KDMP5B, KLF10, KMT2C, LRP1, RNF213, RYR1, SPEN, TERF, TTN* and *UBR4.* In addition, the top 100% mutated genes in HNSCC were *TTN* and *WNK1*, while COL6A3 had a 75% mutation rate (Figure 2). We also identified *TTN, COL6A3*, and *FLNA* as significantly mutated shared genes in CC and HNSCC with greater than 50% mutation rates (Figure 2).

Oncogenic pathway analysis in CC and HNSCC. We expanded our analysis of mutated genes to encompass recognized oncogenic signaling pathways. As illustrated in Figure 3A-B, both CC and HNSCC exhibited a notable proportion of pathways that were affected by oncogenic alterations. CC showed that 100% samples were affected in five pathways, including RTK-RAS with 23/85 genes affected in the pathway, *NOTCH* (19/71), *HIPPO* (10/38), *MYC* (3/13) and *PI3K* (5/29) pathways. HNSCC showed that 100% of the samples were affected in two cellular pathways, *RTK-RAS* (14/85) and *NOTCH* (13/71). Notably, CC tended to exhibit a greater proportion of genes in each pathway compared to their HNSCC counterparts. However, we also noticed that many oncogenic pathways had gene mutations in both CC and HNSCC, including *RTK-RAS, NOTCH, HIPPO, WNT, TP53, PI3K, MYC, Cell cycle* and the *TGF-Beta* pathways. It should be noted that the NRF2 oncogenic pathway was exclusively affected in CC (Figure 3).

Key oncogenes and tumor-suppressor genes in CC and HNSCC. In addition to pathways, we also explored genomic alterations, focusing particularly on identifying potential driver mutations. We depicted gene mutations in the *RTK-RAS* and *NOTCH* pathways (Figure 3C-F) in both CC and HNSCC. This analysis unveiled several pivotal genes involved in the pathogenesis of CC and HNSCC, as well as some distinct genes in each group marked by arrows. In the *RTK-RAS* pathway, tumor suppressor genes *IGF1R* and *PDGFRA* were mutated in both groups, while numerous oncogenes, including *EGFR, ERBB2, ERBB3, MAPK1, NTRK2, FGFR2, FGFR4, KIT,* and *NRAS*, were mutated exclusively in CC (Figure 3, yellow arrows). Conversely, tumor suppressor genes, including *ERBB4, FGFR1, JAK2, and NTRK3* (Figure 3, blue arrows), were mutated in HNSCC. Regarding the *NOTCH* pathway comparison, both CC and HNSCC showed mutations in a large number of genes, such as *SPEN*, *CREBBP, EP300, NCOR1, and NCOR2, while NOTCH2, NOTCH3, FBXW7,* and *NOTCH1* mutations were observed only in CC (Figure 3).

To investigate the functional consequences of somatic variants in the most frequently mutated genes in HPV-positive CC and HNSCC, we depicted their relative positions, counts, and types within annotated protein domains (Figure 4). We observed that all HPV-positive CC samples harbored missense mutations in *SPEN*, which disrupts the hormone-inducible transcriptional repressor (30). Particularly noteworthy is a somatic mutation found within the RNA recognition motif region, which is known to bind to a steroid receptor RNA coactivator (31). This binding can modulate the activity of both liganded and nonliganded steroid receptors (Figure 4A). Additionally, we found a 50% somatic mutation rate in *EP300*, encompassing frame shift deletions, nonsense and missense mutations, as well as in-frame deletions. Several of these mutations were identified in *KAT11, KIX, DUF902*, and *CREB* binding domains (Figure 4A). Apart from these genes, we also observed missense mutations in the *EGFR* Recep L domain and *ERBB2* Phospho kinase Tyr domain (Figure 4B-D). In HNSCC samples, we observed a missense mutation in *JAK2* within the *PTK-Jak2-Jak3-rpl1 domain*. Moreover, a nonsense mutation and a missense mutation were harbored in *ERBB4* cytoplasmic fragment (Figure 4E-F).

Transcriptomic sequence analysis reveals key pathways in HPV-positive CC and HNSCC. To expand our investigation, we conducted RNA-seq analysis to discern genes exhibiting alterations between tumor samples and their corresponding normal tissues (Supplementary Material 2). The CC and HNSCC samples showed a closer relationship compared to their control, indicating similar gene expression patterns (Figure 5A). In addition, this analysis identified 212 up-regulated and 1346 down-

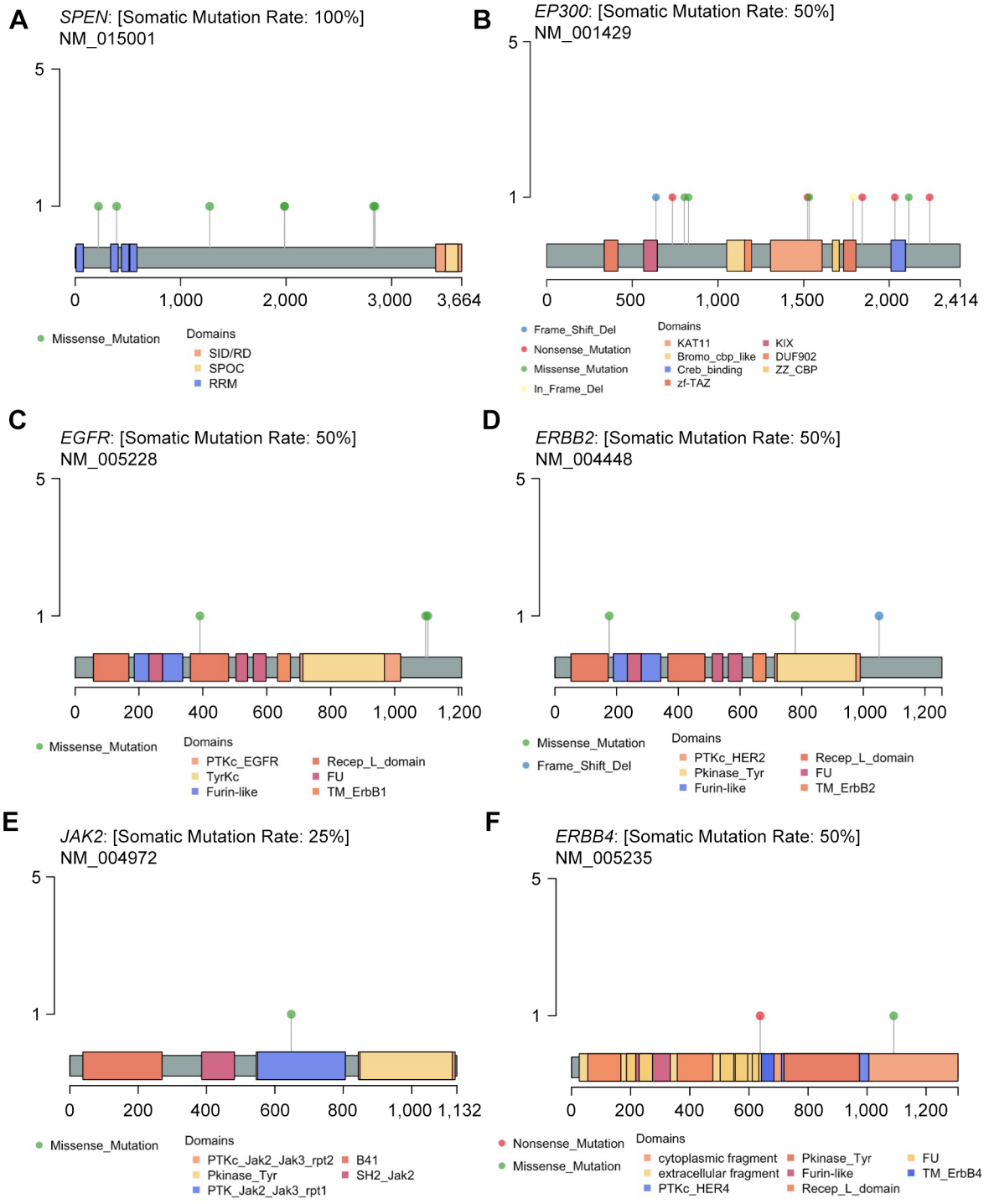

Figure 4. *Somatic mutated variants in genes. (A-D) SPEN, EP300, EGFR, and ERBB2 in HPV-positive cervical cancer (CC). (E-F) JAK2 and ERBB4 in head and neck squamous cell carcinomas (HNSCC). The type of mutation and anatomical localization are annotated below each figure. The variant prevalence and spectrum of these genes are shown in the plots. Green circles indicate missense mutations, blue circles indicate frame shift deletion, red circles indicate nonsense mutations and yellow circles indicate in frame deletion.*

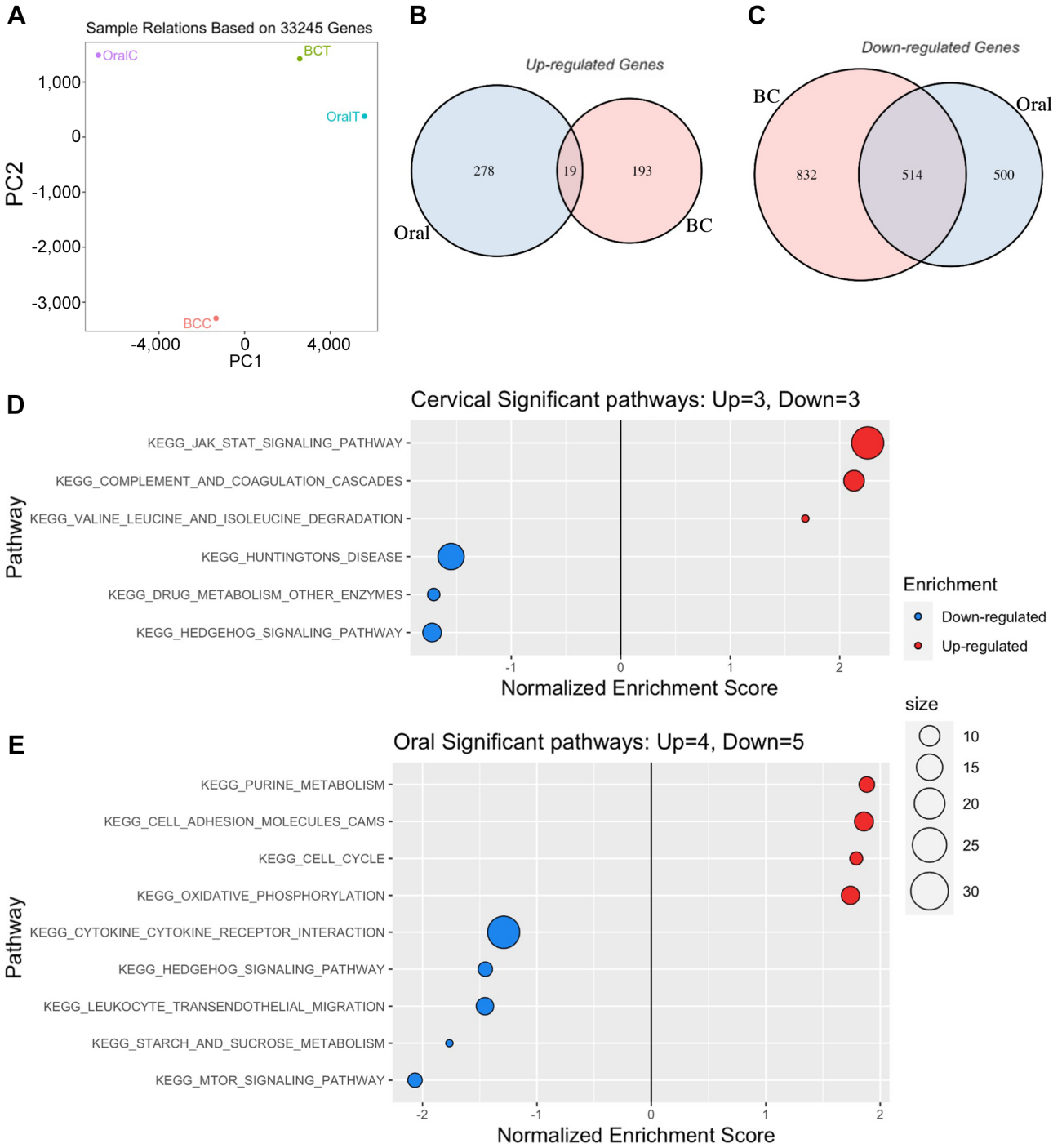

Figure 5. *(A) Sample relation based on 33245 Genes after PCA dimension reduction for oral tumor (OralT), oral control (OralC), bot-cervical tumor (BCT) and bot-cervical control (BCC). X-axis refers to PC1, and y-axis is PC2. (B-C) Venn diagram showing the number of overlap of up-regulated and down-regulated genes between HPV-positive cervical cancer (CC) and head and neck squamous cell carcinomas (HNSCC). (D-E) Enrichment analysis of CC and HNSCC in KEGG pathways. Dotplot with x-axis as normalized enrichment score and y-axis as the name of the pathways. Up-regulated pathways facing right and down-regulated pathways are facing left. The size scale is based on number of affected genes.*

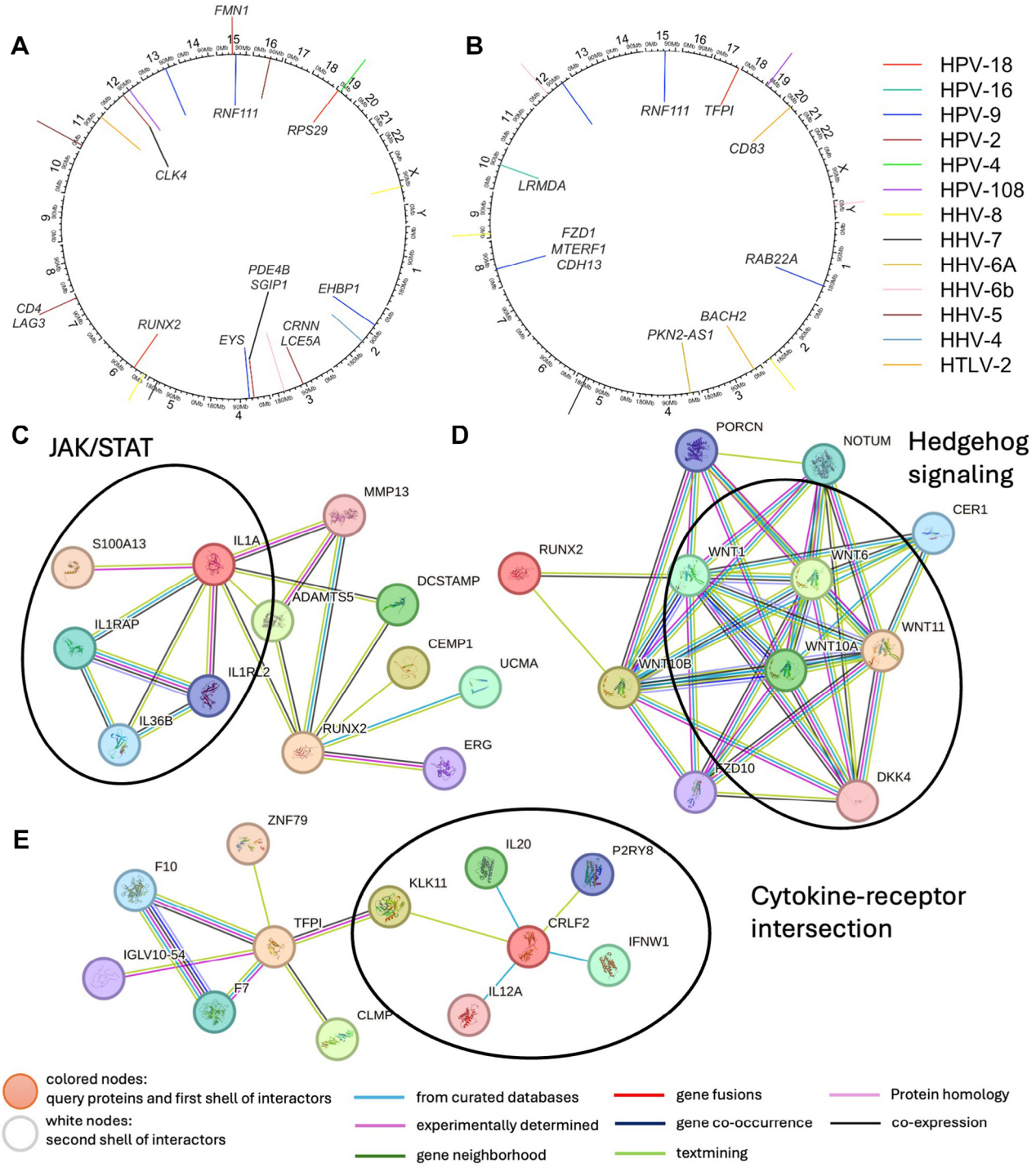

Figure 6. *Virus integration sites cycle plots in HPV-positive cervical cancer (CC) (A) and head and neck squamous cell carcinomas (HNSCC) (B). The circle contains 22 chromosomes as well as the X and Y chromosomes. Integration sites facing inside are from tumors while ones facing outside are from the control group. 13 different virus types are indicated by color by the right side. (C-E) Protein-to-protein interaction maps for RUNX2 and TFPI with JAK/STAT, Hedgehog signaling and cytokine-receptor intersections. Colored nodes are query proteins and first shell of interactors; white nodes are second shell of interactors; edges are assoications based on known or predicted interactions.*

regulated differentially expressed genes (DEGs) out of 6,514 detectable genes in CC, which met the criteria of |logFC| >3. Similarly, 297 up-regulated and 1014 down-regulated DEGs were identified in HNSCC (Supplementary Material 3-4). Nineteen up-regulated genes were shared between CC and HNSCC, including *ABO, ADGRF4, ANGPTL3, CIDEC, DCAF12L2, GPC4, IFNA21, IL37, KCNF1, KERA, KIF7, KRBOX4, MYOZ2, NDUFAF4, NUDT15, OR5K1, PEX7, PIGA,* and *ZNF223*. Among these genes, only PIGA was reported as an oncogene causing paroxysmal nocturnal hemoglobinuria (32). Furthermore, 514 down-regulated genes were shared between CC and HNSCC, where 24 of them are previously reported oncogenes, including *CEBPA, CLP1, EPOR, FEV, FGF3, FH, FOLH1, FOXA1, GATA1, H3-3A, HIRA, HOXC13, HRAS, KMT2B, LMO1, LYL1, MLF1, MN1, NKX2-1, NTHL1, PDCD1, S1PR2, SERPINB4,* and *SH2D1A* (Figure 5B-C).

Our study conducted an in-depth analysis of biological pathways using Kyoto Encyclopedia of Genes and Genomes (KEGG) (33), with the aim of identifying the specific pathways enriched in CC and HNSCC. This investigation provides insight into the molecular mechanisms underlying these malignancies (see Supplementary Material 5-8, Figure 5D-E), with the pathways selected based on an adjusted P-value of less than 0.1 with false discovery rate (FDR) adjustment. We found three up-regulated pathways in CC, including *JAK/STAT*, complement and coagulation cascades, and valine, leucine and isoleucine degradation; three down-regulated pathways, including Huntington's disease, drug metabolism and other enzymes, and the hedgehog signaling pathway (Figure 5D-E). On the other hand, we found four up-regulated pathways in HNSCC, including purine metabolism, cell adhesion molecules, cell cycle and oxidative phosphorylation; as well as five down-regulated pathways, including cytokine-receptor interaction, hedgehog signaling, leukocyte trans-endothelial migration, starch and sucrose metabolism and *MTOR signaling*. We showed that hedgehog-signaling pathway was down-regulated in both CC and HNSCC, while many other pathways are either immune related or tumor proliferation related (Figure 5D-E).

Viral integration analysis. In our study, an investigation into the virome composition of CC and HNSCC samples revealed a complex and diverse viral landscape. The analysis of virus composition uncovered shared viral entities between both CC and HNSCC samples, while also highlighting notable differences. Some viral entities that were detected in both CC and HNSCC samples were *HHV-4, HHV-5, HHV-6A, HHV-6B, HHV-7, HHV-8*, and a number of HPV types including *HPV-108, HPV-16, HPV-18, HPV-2, HPV-9,* and *HPV-4*, highlighting the presence of HPV infections at these anatomical sites (Figure 6A-B). Notably, HPV-18 and HPV-9 were found to have higher counts in CC samples compared to HNSCC samples, indicating a potentially more significant role in CC malignancies. In CC, *FMN1, RUNX2, EYS, EHBP1,* and *NUP93* were identified as viral integration sites. In HNSCC, viral integration sites were detected in *TFPI, ADAMTS13, GLRA1, NEDD4, CD83, PLA2G2A, RIN2, LRMDA,* and *CDH13* (Figure 6A-B).

Integrated analysis reveals RUNX2 and TFPI as key factors in HPV-positive CC and HNSCC. Our study delved into the intricate relationships among the genome, transcriptome, and viral integrations in both CC and HNSCC samples. To explore the regulatory mechanisms underlying the expression of significant genes, we utilized the QIAGEN Ingenuity Pathway Analysis (5) program. This analytical approach allowed us to identify key upstream regulators of transcriptionally significant genes, shedding light on the intricate interplay between viral integrations and gene expression. In CC samples, a notable finding was the insertion of the HPV-18 virus on chromosome 6, coinciding with the presence of the *RUNX2* gene (Figure 5A). This gene was identified both as an upstream regulator in the IPA analysis and as a site of viral detection within the CC cohorts. Several genes regulated by *RUNX2* support various aspects of tumor progression. Therefore, we plotted the interaction map between *RUNX2* and *JAK/STAT,* and the hedgehog signaling pathway (Figure 5C-D). *RUNX2* interacts with *IL1A* and *ADAMTS5* directly, which affects the expression of *IL1RAP, IL1RL2* and *IL36B*; it also has direct regulation of *WNT1* and *WNT10B*, which showed down regulation in their

transcription expression. Similarly, although TFPI is not upstream, its recognized inhibitor behavior showed strong interaction with the cytokine-receptor intersection pathway through interaction with *KLK11* (Figure 5E). *KLK11* was initially isolated from the human hippocampus and is considered a potential biomarker for prostate, ovarian, and breast cancers (34). Its expression in peripheral tissues might be stimulated by tumor necrosis factor α (*TNFα*) secretion from tumor cells. TNFα, a key inflammatory cytokine, is notably overexpressed in advanced breast cancer (35). Notably, its involvement in a range of different biological responses were previously described. For example, increased expression of metalloproteases, epithelial–mesenchymal transition, angiogenesis, inflammation, and drug resistance (34-36). The role of TFPI in HPV-positive oral cancer prognosis, and its underlying mechanism will be more clearly understood with additional investigations, which utilizes a larger study sample size.

Discussion

Cervical cancer (CC) and head and neck squamous cell carcinoma (HNSCC) are prevalent and deadly malignancies. Despite the availability of effective prophylactic vaccines targeting key carcinogenic HPV types, vaccine uptake remains low (11). Furthermore, the impact of vaccines will only be appreciated 3 decades after initiation (37). Nonetheless, the countries where the diseases are most prevalent are the ones with the lowest levels of vaccine uptake (38). The interaction between the host genome and virus drives the clinical presentation and therapeutic response that is distinct from the non-virally associated cancers. Therefore, understanding the molecular mechanisms and genetic predisposition of these cancers are of paramount importance for clinical intervention. Over the last few decades, oral HPV infection has surpassed smoking and alcohol as the primary risk factor for oropharyngeal cancers and are more common in males. However, there appears to be a recent increase in incidence among females. Our study aims to provide a comparison between oral and cervical cancers in women with positive HPV tests to a negative HPV control group, providing insights into the similarities and differences between HPV-associated CC and HNSCC.

In this study, we did a thorough analysis based on somatic mutations, oncogenic pathways, enrichment pathways, and virus integrations between HPV-positive CC and HNSCC. We found both similarities and unique characteristics between these two types of cancers. Broadly, the genes with high mutation rates are shown in Figure 2, among which many are reported as oncogenes or tumor suppressors (14, 39-46). We also found several important oncogenic pathways which were previously reported to be associated with HPV-positive cancer progression (47-49). The mutated genes in the *RTK-RAS* and *NOTCH* pathways were different between CC and HNSCC, which were reported to be the leading cause for cancer progression in these cancers (Figure 2 and Figure 3). Moreover, we found that the CC and HNSCC relationship is closer compared their control groups based on gene expression. Nineteen up-regulated shared genes and 514 down-regulated genes were found between the CC and HNSCC tumors. Enrichment analysis using the KEGG pathway analyses showed a greater number of up-regulated pathways and down-regulated pathways in CC and HNSCC (Figure 4). In our virus integration analyses, we found several key factors that were linked to cancer progression and their interactions with the regulation of the pathways identified (Figure 5).

Integration across multiple platforms revealed a diverse array of mutations in CC. Among the top 10 mutated genes, with mutation rates varying from 12% to 33% were *TTN*, *PIK3CA*, *MUC4*, *KMT2C*, *MUC16*, *KMT2D*, *SYNE1*, *FLG*, *DST*, and *EP300* (50). *EP300* regulates *cAMP*-gene expression by binding specifically to phosphorylated *CREB* protein and functions as a histone acetyltransferase, controlling transcription through chromatin remodeling, which is crucial in the processes of cell proliferation and differentiation. Additionally, we observed mutations in *KLF10*, which acts as a tumor suppressor through the *TGFβ* signaling pathway (46); *KMT2C*, a frequently mutated epigenetic modifier gene essential in DNA

replication fork restart (51); and *RNF213*, whose role in CC necessitates further investigation and validation as a prognostic factor (51). *SPEN* mutation showed a significant correlation with tumor mutation burden (TMB) and microsatellite instability (MSI) in pan-cancer and exhibited promise as a biomarker, especially in patients undergoing Immune Checkpoint Inhibitors (ICI) therapy, regardless of cancer type (52). In HNSCC, the top mutated genes include *P53*, *CDKN2A*, *CASP8*, *FAT1*, *NOTCH1*, *HRAS*, *PIK3CA*, *MLL2*, and *FBXW7* (53). Additionally, we detected a mutation in *WNK1*, suggesting a potential involvement of WNK family of genes in cancers (54). Most of these top mutated genes aligned with the previously published results including *NOTCH1*, *HRAS* and *WNK1* as seen in the Cancer Genome Atlas Project (9).

Furthermore, the oncogenic pathways enriched due to the mutated genes may offer some insights into signaling pathways associated with CC and HNSCC that can be therapeutically targeted such as *PI3K/AKT* and *NOTCH* pathways. Mutations affecting *RTK* signaling often result in cell transformation, observed in several different malignancies. These mutations impact *RTKs* or downstream pathways like *MAP* kinase and *PI3K/AKT*, leading to increased cell proliferation, survival, invasion, and metastasis (47). *NOTCH* activity alteration, either gain or loss, promotes HPV-associated cancer development. Additionally, HPV oncogenes likely disrupt *NOTCH* tumor-suppressive activities, enabling the oncogenic pathways activated by *NOTCH* to support tumor growth (55). *NOTCH* signaling can both promote and inhibit tumor development in different cancers. In CC progression, partially differentiated progenitors, which proliferate faster than basal cells, are likely the origin of tumor cells (48). In addition, the NOTCH pathway plays a crucial role in maintaining oral tissue homeostasis and that of other organ systems (49). Therefore, HNSCC-specific *NOTCH* pathway therapy would need to target specific *NOTCH* isoforms to avoid systemic and gastrointestinal toxicities. Apart from *NOTCH*, deregulation of *RTK-RAS/PI3K* was reported in HPV-associated CC and HNSCC, leading to further deregulation in *MYC* and *HIPPO* pathways (14). Additionally, the *NRF2* pathway is highly mutated in both CC and HNSCC. *NRF2* has been linked to treatment resistance, with significant down-regulation of *NRF2*-signaling observed in the subtype of HPV-positive HNSCC (56).

KEGG enrichment pathway analyses showed that CC and HNSCC have significantly different enriched pathways. In CC, we observed an up-regulation of *JAK-STAT* signaling. Previous studies have shown that inhibiting *JAK2* reduces *STAT3* activation, leading to decreased cell proliferation and increased sensitivity to Cisplatin (57). Additionally, we noted a down-regulation of Hedgehog signaling, which is known to be activated in various cancers, including human hepatocellular carcinoma (58). In HNSCC, we observed an up-regulation of cell cycle pathway, consistent with prior studies (59). For instance, Erufosine has been shown to down-regulate cell cycle processes in OSCC cells, highlighting its potential as a cell cycle inhibitor in treatment of HNSCC, either alone or in combination with other therapies (59). Furthermore, we also observed activation of oxidative phosphorylation (OXPHOS). Resistant cancer cells often exhibit reprogrammed metabolism, leading to an energy metabolism shift towards OXPHOS, mediated by certain oncogenes (18). Similarly, we found down-regulation of Hedgehog signaling in HNSCC. Moreover, we also found down-regulation of cytokine-receptor interaction in both CC and HNSCC. These cytokine-mediated processes were shown to be significantly associated with positive regulation of the immune response in various immune cells (60). Importantly, engineered cytokines such as *IFN*, *IL-2*, and *IL-4* are well-known to have therapeutic potential in regulating the immune response (61).

Using a multi-omics approach, we identified *RUNX2* as a crucial gene associated with HPV integration sites. Previous studies have demonstrated an indirect modulation of the *PI3K/AKT* signaling pathway by *RUNX2* in both prostate and breast cancer cells (16). Our research demonstrates interactions between *RUNX2*, the *JAK/STAT* and Hedgehog signaling pathways through various mediators, including *IL1A*, *WNT1*, and *WNT10B*. Notably,

WNT1 was found to be down-regulated in CC, supporting our conclusions that *WNT1* is involved in driving the oncogenic phenotype in this cancer. Additionally, we discovered that *TFPI*, a tissue factor pathway inhibitor (62), was adjacent to inserted HPV-18 on chromosome 17 in HNSCC. *TFPI* exhibits anti-thrombotic actions and can associate with lipoproteins in plasma (63). We now show the involvement of *TFPI,* and the cytokine-receptor interactions are likely through *KLK11*, which is important for maintaining tumor environment stability, and inhibition of invasiveness, neoplasm growth, and metastasis formation. *TFPI* has also been shown to induce apoptosis and inhibit angiogenesis, significantly contributing to tumor growth inhibition (64).

In this study, we explored the potential genetic and gene-reprogramming connections that occur in CC and HNSCC, emphasizing their interactions to the somatic mutations, enriched pathways, and HPV oncogenes present in these cancers. Our findings shed light on oncogenic pathways including *RTK/RAS* and *NOTCH*, which were present in both cancer types, and corroborated in part by other previous studies (47-49, 55, 65). Moreover, we underscore the importance of the *RUNX2* and *TFPI* genes as insertion sites for oncogenic HPV genomes, influencing downstream transcriptional expression, and pathways such as *JAK/STAT*, Hedgehog-signaling, and cytokine-receptor interactions. These discoveries underscore the intricate nature of gene reprogramming activities in these cancers and revealing numerous areas that were previously unknown. This opens new avenues for future research. Notably, we did not explore the link between *PI3K/AKT* and *NOTCH* mutations with patient survival outcomes in our cohort due to lack of sufficient clinical data. However, our plan is to expand our study in the future, where we can do follow-up studies on patients to assess the impact of these gene regulatory changes on outcomes. Consequently, further experimentation is imperative to refine treatment strategies and to chart future research avenues important for successful interventions.

Supplementary Material

Available at: https://figshare.com/articles/dataset/Supplementary/27911886

Conflicts of Interest

The Authors declare that they have no conflicts of interest related to this study.

Authors' Contributions

E.R. conceived and designed the study. J.R. performed the experiments and data analysis. J.R., S.G., T.S., S.B., N.Z., and N.M. contributed to data interpretation and manuscript drafting. E.R. and Z.W. supervised the study and critically revised the manuscript. All Authors read and approved the final version of the manuscript.

References

1 Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV, Sidransky D: Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 92(9): 709-720, 2000. DOI: 10.1093/jnci/92.9.709

2 Park NJ, Choi Y, Lee D, Park JY, Kim JM, Lee YH, Hong DG, Chong GO, Han HS: Transcriptomic network analysis using exfoliative cervical cells could discriminate a potential risk of progression to cancer in HPV-related cervical lesions: a pilot study. Cancer Genomics Proteomics 20(1): 75-87, 2023. DOI: 10.21873/cgp.20366

3 Dong H, Shu X, Xu Q, Zhu C, Kaufmann AM, Zheng ZM, Albers AE, Qian X: Current status of human papillomavirus-related head and neck cancer: from viral genome to patient care. Virol Sin 36(6): 1284-1302, 2021. DOI: 10.1007/s12250-021-00413-8

4 Roman BR, Aragones A: Epidemiology and incidence of HPV-related cancers of the head and neck. J Surg Oncol 124(6): 920-922, 2021. DOI: 10.1002/jso.26687

5 Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, Lee S, Hadjipanayis AG, Ivanova EV, Wilkerson MD, Protopopov A, Yang L, Seth S, Song X, Tang J, Ren X, Zhang J, Pantazi A, Santoso N, Xu AW, Mahadeshwar H, Wheeler DA, Haddad RI, Jung J, Ojesina AI, Issaeva N, Yarbrough WG, Hayes DN, Grandis JR, El-Naggar AK, Meyerson

M, Park PJ, Chin L, Seidman JG, Hammerman PS, Kucherlapati R, Cancer Genome Atlas Network: Characterization of HPV and host genome interactions in primary head and neck cancers. Proc Natl Acad Sci USA 111(43): 15544-15549, 2014. DOI: 10.1073/pnas.1416074111

6 Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3): 229-263, 2024. DOI: 10.3322/caac.21834

7 Cancer Genome Atlas Research Network, Albert Einstein College of Medicine, Analytical Biological Services, Barretos Cancer Hospital, Baylor College of Medicine, Beckman Research Institute of City of Hope, Buck Institute for Research on Aging, Canada's Michael Smith Genome Sciences Centre, Harvard Medical School, Helen F. Graham Cancer Center &Research Institute at Christiana Care Health Services, HudsonAlpha Institute for Biotechnology, ILSbio, LLC, Indiana University School of Medicine, Institute of Human Virology, Institute for Systems Biology, International Genomics Consortium, Leidos Biomedical, Massachusetts General Hospital, McDonnell Genome Institute at Washington University, Medical College of Wisconsin, Medical University of South Carolina, Memorial Sloan Kettering Cancer Center, Montefiore Medical Center, NantOmics, National Cancer Institute, National Hospital, Abuja, Nigeria, National Human Genome Research Institute, National Institute of Environmental Health Sciences, National Institute on Deafness &Other Communication Disorders, Ontario Tumour Bank, London Health Sciences Centre, Ontario Tumour Bank, Ontario Institute for Cancer Research, Ontario Tumour Bank, The Ottawa Hospital, Oregon Health &Science University, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, SRA International, St Joseph's Candler Health System, Eli &Edythe L. Broad Institute of Massachusetts Institute of Technology &Harvard University, Research Institute at Nationwide Children's Hospital, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, University of Bergen, University of Texas MD Anderson Cancer Center, University of Abuja Teaching Hospital, University of Alabama at Birmingham, University of California, Irvine, University of California Santa Cruz, University of Kansas Medical Center, University of Lausanne, University of New Mexico Health Sciences Center, University of North Carolina at Chapel Hill, University of Oklahoma Health Sciences Center, University of Pittsburgh, University of São Paulo, Ribeirão Preto Medical School, University of Southern California, University of Washington, University of Wisconsin School of Medicine & Public Health, Van Andel Research Institute, Washington University in St Louis: Integrated genomic and molecular characterization of cervical cancer. Nature 543(7645): 378-384, 2017. DOI: 10.1038/nature21386

8 Herrero R, Castellsagué X, Pawlita M, Lissowska J, Kee F, Balaram P, Rajkumar T, Sridhar H, Rose B, Pintos J, Fernández L, Idris A, Sánchez MJ, Nieto A, Talamini R, Tavani A, Bosch FX, Reidel U, Snijders PJ, Meijer CJ, Viscidi R, Muñoz N, Franceschi S, IARC Multicenter Oral Cancer Study Group: Human papillomavirus and oral cancer: The International Agency for Research on Cancer multicenter study. J Natl Cancer Inst 95(23): 1772-1783, 2003. DOI: 10.1093/jnci/djg107

9 Gravitt PE: Unraveling the epidemiology of oral human papillomavirus infection. Ann Intern Med 167(10): 748, 2017. DOI: 10.7326/M17-2628

10 Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, Wentzensen N, Jordan RC, Altekruse S, Anderson WF, Rosenberg PS, Gillison ML: Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 29(32): 4294-4301, 2011. DOI: 10.1200/JCO.2011.36.4596

11 Qiu L, Feng H, Yu H, Li M, You Y, Zhu S, Yang W, Jiang H, Wu X: Characterization of the genomic landscape in cervical cancer by next generation sequencing. Genes (Basel) 13(2): 287, 2022. DOI: 10.3390/genes13020287

12 Rischin D, Young RJ, Fisher R, Fox SB, Le QT, Peters LJ, Solomon B, Choi J, O'Sullivan B, Kenny LM, McArthur GA: Prognostic significance of p16INK4A and human papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. J Clin Oncol 28(27): 4142-4148, 2010. DOI: 10.1200/JCO.2010.29.2904

13 LaFramboise WA, Pai RK, Petrosko P, Belsky MA, Dhir A, Howard PG, Becich MJ, Holtzman MP, Ahrendt SA, Pingpank JF, Zeh HJ, Dhir R, Bartlett DL, Choudry HA: Discrimination of low- and high-grade appendiceal mucinous neoplasms by targeted sequencing of cancer-related variants. Mod Pathol 32(8): 1197-1209, 2019. DOI: 10.1038/s41379-019-0256-2

14 Cancer Genome Atlas Network: Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517(7536): 576-582, 2015. DOI: 10.1038/nature14129

15 Chaudary N, Pintilie M, Hedley D, Fyles AW, Milosevic M, Clarke B, Hill RP, Mackay H: Hedgehog pathway signaling in cervical carcinoma and outcome after chemoradiation. Cancer 118(12): 3105-3115, 2012. DOI: 10.1002/cncr.26635

16 Yang Y, Yang C, Yang Q, Lu S, Liu B, Li D, Li D, Zhang P, Xu P, Lang J, Zhou J: Elucidating Hedgehog pathway's role in HNSCC progression: insights from a 6-gene signature. Sci Rep 14(1): 4686, 2024. DOI: 10.1038/s41598-024-54937-6

17 Mir BA, Ahmad A, Farooq N, Priya MV, Siddiqui AH, Asif M, Manzoor R, Ishqi HM, Alomar SY, Rahaman PF: Increased expression of HPV-E7 oncoprotein correlates with a reduced level of pRb proteins *via* high viral load in cervical cancer. Sci Rep 13(1): 15075, 2023. DOI: 10.1038/s41598-023-42022-3

18 Zhou Y, Huang J, Jin B, He S, Dang Y, Zhao T, Jin Z: The emerging role of Hedgehog signaling in viral infections.

Front Microbiol 13: 870316, 2022. DOI: 10.3389/fmicb.2022.870316
19 Sathish N, Wang X, Yuan Y: Human papillomavirus (HPV)-associated oral cancers and treatment strategies. J Dent Res 93(7 Suppl): 29S-36S, 2014. DOI: 10.1177/0022034514527969
20 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9): 1297-1303, 2010. DOI: 10.1101/gr.107524.110
21 Withanage MHH, Liang H, Zeng E: RNA-Seq experiment and data analysis. Methods Mol Biol 2418: 405-424, 2022. DOI: 10.1007/978-1-0716-1920-9_22
22 Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12): 550, 2014. DOI: 10.1186/s13059-014-0550-8
23 Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A: Fast gene set enrichment analysis. bioRxiv, 2016. DOI: 10.1101/060012
24 Ho DW, Sze KM, Ng IO: Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. Oncotarget 6(25): 20959-20963, 2015. DOI: 10.18632/oncotarget.4187
25 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H: Twelve years of SAMtools and BCFtools. Gigascience 10(2): giab008, 2021. DOI: 10.1093/gigascience/giab008
26 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16): e164, 2010. DOI: 10.1093/nar/gkq603
27 Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C: The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51(D1): D638-D646, 2023. DOI: 10.1093/nar/gkac1000
28 Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP: Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28(11): 1747-1756, 2018. DOI: 10.1101/gr.239244.118
29 Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 57(1): 289-300, 1995. DOI: 10.1111/j.2517-6161.1995.tb02031.x
30 Légaré S, Cavallone L, Mamo A, Chabot C, Sirois I, Magliocco A, Klimowicz A, Tonin PN, Buchanan M, Keilty D, Hassan S, Laperrière D, Mader S, Aleynikova O, Basik M: The estrogen receptor cofactor SPEN functions as a tumor suppressor and candidate biomarker of drug responsiveness in hormone-dependent breast cancers. Cancer Res 75(20): 4351-4363, 2015. DOI: 10.1158/0008-5472.CAN-14-3475
31 Tao Y, Zhang Q, Wang H, Yang X, Mu H: Alternative splicing and related RNA binding proteins in human health and disease. Signal Transduct Target Ther 9(1): 26, 2024. DOI: 10.1038/s41392-024-01734-2
32 Yuan X, Li Z, Baines AC, Gavriilaki E, Ye Z, Wen Z, Braunstein EM, Biesecker LG, Cheng L, Dong X, Brodsky RA: A hypomorphic PIGA gene mutation causes severe defects in neuron development and susceptibility to complement-mediated toxicity in a human iPSC model. PLoS One 12(4): e0174074, 2017. DOI: 10.1371/journal.pone.0174074
33 Kanehisa M, Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1): 27-30, 2000. DOI: 10.1093/nar/28.1.27
34 Madhusudan S, Foster M, Muthuramalingam SR, Braybrooke JP, Wilner S, Kaur K, Han C, Hoare S, Balkwill F, Talbot DC, Ganesan TS, Harris AL: A phase II study of Etanercept (Enbrel), a tumor necrosis factor α inhibitor in patients with metastatic breast cancer. Clin Cancer Res 10(19): 6528-6534, 2004. DOI: 10.1158/1078-0432.CCR-04-0730
35 Morizane S, Yamasaki K, Kabigting FD, Gallo RL: Kallikrein expression and cathelicidin processing are independently controlled in keratinocytes by calcium, vitamin D(3), and retinoic acid. J Invest Dermatol 130(5): 1297-1306, 2010. DOI: 10.1038/jid.2009.435
36 Esteva FJ, Hortobagyi GN: Prognostic molecular markers in early breast cancer. Breast Cancer Res 6(3): 109-118, 2004. DOI: 10.1186/bcr777
37 Combrisson E, Jerbi K: Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods 250: 126-136, 2015. DOI: 10.1016/j.jneumeth.2015.01.010
38 Kutz JM, Rausche P, Gheit T, Puradiredja DI, Fusco D: Barriers and facilitators of HPV vaccination in sub-saharan Africa: a systematic review. BMC Public Health 23(1): 974, 2023. DOI: 10.1186/s12889-023-15842-1
39 Oh JH, Jang SJ, Kim J, Sohn I, Lee JY, Cho EJ, Chun SM, Sung CO: Spontaneous mutations in the single TTN gene represent high tumor mutation burden. NPJ Genom Med 5: 33, 2020. DOI: 10.1038/s41525-019-0107-6
40 Szymonowicz KA, Chen J: Biological and clinical aspects of HPV-related cancers. Cancer Biol Med 17(4): 864-878, 2020. DOI: 10.20892/j.issn.2095-3941.2020.0370
41 Wang J, Pan W: The biological role of the collagen alpha-3 (VI) chain and its cleaved C5 domain fragment endotrophin in cancer. Onco Targets Ther 13: 5779-5793, 2020. DOI: 10.2147/OTT.S256654
42 Savoy RM, Ghosh PM: The dual role of filamin A in cancer: can't live with (too much of) it, can't live without it. Endocr

Relat Cancer 20(6): R341-R356, 2013. DOI: 10.1530/ERC-13-0364
43 Acs G, Zhang PJ, McGrath CM, Acs P, McBroom J, Mohyeldin A, Liu S, Lu H, Verma A: Hypoxia-inducible erythropoietin signaling in squamous dysplasia and squamous cell carcinoma of the uterine cervix and its potential role in cervical carcinogenesis and tumor progression. Am J Pathol 162(6): 1789-1806, 2003. DOI: 10.1016/S0002-9440(10)64314-3
44 Chua CW, Chiu YT, Yuen HF, Chan KW, Man K, Wang X, Ling MT, Wong YC: Suppression of androgen-independent prostate cancer cell aggressiveness by FTY720: Validating Runx2 as a potential antimetastatic drug screening platform. Clin Cancer Res 15(13): 4322-4335, 2009. DOI: 10.1158/1078-0432.CCR-08-3157
45 Tandon M, Chen Z, Pratap J: Runx2 activates PI3K/Akt signaling *via* mTORC2 regulation in invasive breast cancer cells. Breast Cancer Res 16(1): R16, 2014. DOI: 10.1186/bcr3611
46 Memon A, Lee WK: KLF10 as a tumor suppressor gene and its TGF-β signaling. Cancers (Basel) 10(6): 161, 2018. DOI: 10.3390/cancers10060161
47 Regad T: Targeting RTK signaling pathways in cancer. Cancers (Basel) 7(3): 1758-1784, 2015. DOI: 10.3390/cancers7030860
48 Rodrigues C, Joy LR, Sachithanandan SP, Krishna S: Notch signalling in cervical cancer. Exp Cell Res 385(2): 111682, 2019. DOI: 10.1016/j.yexcr.2019.111682
49 Harada H, Kettunen P, Jung HS, Mustonen T, Wang YA, Thesleff I: Localization of putative stem cells in dental epithelium and their association with Notch and FGF signaling. J Cell Biol 147(1): 105-120, 1999. DOI: 10.1083/jcb.147.1.105
50 Xu Y, Luo H, Hu Q, Zhu H: Identification of potential driver genes based on multi-genomic data in cervical cancer. Front Genet 12: 598304, 2021. DOI: 10.3389/fgene.2021.598304
51 Lee SY, Chae DK, Lee SH, Lim Y, An J, Chae CH, Kim BC, Bhak J, Bolser D, Cho DH: Efficient mutation screening for cervical cancers from circulating tumor DNA in blood. BMC Cancer 20(1): 694, 2020. DOI: 10.1186/s12885-020-07161-0
52 Li YD, Huang H, Ren ZJ, Yuan Y, Wu H, Liu C: Pan-cancer analysis identifies SPEN mutation as a predictive biomarker with the efficacy of immunotherapy. BMC Cancer 23(1): 793, 2023. DOI: 10.1186/s12885-023-11235-0
53 Riaz N, Morris LG, Lee W, Chan TA: Unraveling the molecular genetics of head and neck cancer through genome-wide approaches. Genes Dis 1(1): 75-86, 2014. DOI: 10.1016/j.gendis.2014.07.002
54 Jung JU, Jaykumar AB, Cobb MH: WNK1 in malignant behaviors: a potential target for cancer? Front Cell Dev Biol 10: 935318, 2022. DOI: 10.3389/fcell.2022.935318
55 Das T, Zhong R, Spiotto MT: Notch signaling and human papillomavirus-associated oral tumorigenesis. Adv Exp Med Biol 1287: 105-122, 2021. DOI: 10.1007/978-3-030-55031-8_8
56 Nakano-Kobayashi A, Canela A, Yoshihara T, Hagiwara M: Astrocyte-targeting therapy rescues cognitive impairment caused by neuroinflammation *via* the Nrf2 pathway. Proc Natl Acad Sci USA 120(33): e2303809120, 2023. DOI: 10.1073/pnas.2303809120
57 Morgan EL, Macdonald A: JAK2 inhibition impairs proliferation and sensitises cervical cancer cells to cisplatin-induced cell death. Cancers (Basel) 11(12): 1934, 2019. DOI: 10.3390/cancers11121934
58 Tada M, Kanai F, Tanaka Y, Tateishi K, Ohta M, Asaoka Y, Seto M, Muroyama R, Fukai K, Imazeki F, Kawabe T, Yokosuka O, Omata M: Down-regulation of hedgehog-interacting protein through genetic and epigenetic alterations in human hepatocellular carcinoma. Clin Cancer Res 14(12): 3768-3776, 2008. DOI: 10.1158/1078-0432.CCR-07-1181
59 Ansari SS, Sharma AK, Zepp M, Ivanova E, Bergmann F, König R, Berger MR: Upregulation of cell cycle genes in head and neck cancer patients may be antagonized by erufosine's down regulation of cell cycle processes in OSCC cells. Oncotarget 9(5): 5797-5810, 2017. DOI: 10.18632/oncotarget.23537
60 Liu Y, Zhao M, Qu H: Identification of cytokine-induced cell communications by pan-cancer meta-analysis. PeerJ 11: e16221, 2023. DOI: 10.7717/peerj.16221
61 Spangler JB, Moraga I, Mendoza JL, Garcia KC: Insights into cytokine-receptor interactions from cytokine engineering. Annu Rev Immunol 33: 139-167, 2015. DOI: 10.1146/annurev-immunol-032713-120211
62 Nguyen AB, Iqbal O, Block RC, Mousa SA: Prevention and treatment of atherothrombosis: Potential impact of nanotechnology. Vascul Pharmacol 148: 107127, 2023. DOI: 10.1016/j.vph.2022.107127
63 Broze GJ Jr, Girard TJ: Tissue factor pathway inhibitor: structure-function. Front Biosci (Landmark Ed) 17(1): 262-280, 2012. DOI: 10.2741/3926
64 Sierko E, Wojtukiewicz MZ, Kisiel W: The role of tissue factor pathway inhibitor-2 in cancer biology. Semin Thromb Hemost 33(7): 653-659, 2007. DOI: 10.1055/s-2007-991532
65 Bray SJ: Notch signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol 7(9): 678-689, 2006. DOI: 10.1038/nrm2009