

Differential gene expression analysis of HNSCC tumors deciphered tobacco dependent and independent molecular signatures

Inayatullah Shaikh¹, Afzal Ansari¹, Garima Ayachit¹, Monika Gandhi¹, Priyanka Sharma¹, Shivarudrappa Bhairappanavar¹, Chaitanya G. Joshi¹ and Jayashankar Das¹

¹Gujarat Biotechnology Research Centre (GBRC), Department of Science and Technology (DST), Government of Gujarat, Gandhinagar 382011, India

Correspondence to: Jayashankar Das, *email:* jayshankardas@gmail.com, jd2-gbrc@gujarat.gov.in

Keywords: tobacco; head and neck squamous cell carcinoma (HNSCC); differentially expressed genes; hub gene; miRNA

Received: June 14, 2019

Accepted: September 16, 2019

Published: October 22, 2019

Copyright: Shaikh et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Head and neck cancer is the sixth most common cancer worldwide, with tobacco as the leading cause. However, it is increasing in non-tobacco users also, hence limiting our understanding of its underlying molecular mechanisms. RNA-seq analysis of cancers has proven as effective tool in understanding disease etiology. In the present study, RNA-Seq of 86 matched Tumor/Normal pairs, of tobacco smoking (TOB) and non-smokers (N-TOB) HNSCC samples analyzed, followed by validation on 375 similar datasets. Total 2194 and 2073 differentially expressed genes were identified in TOB and N-TOB tumors, respectively. GO analysis found muscle contraction as the most enriched biological process in both TOB and N-TOB tumors. Pathway analysis identified muscle contraction and salivary secretion pathways enriched in both categories, whereas calcium signaling and neuroactive ligand-receptor pathway was more enriched in TOB and N-TOB tumors respectively. Network analysis identified muscle development related genes as hub node i. e. ACTN2, MYL2 and TTN in both TOB and N-TOB tumors, whereas EGFR and MYH6, depicts specific role in TOB and N-TOB tumors. Additionally, we found enriched gene networks possibly be regulated by tumor suppressor miRNAs such as hsa-miR-29/a/b/c, hsa-miR-26b-5p etc., suggestive to be key riboswitches in regulatory cascade of HNSCC. Interestingly, three genes PKLR, CST1 and C17orf77 found to show opposite regulation in each category, hence suggested to be key genes in separating TOB from N-TOB tumors. Our investigation identified key genes involved in important pathways implicated in tobacco dependent and independent carcinogenesis hence may help in designing precise HNSCC diagnostics and therapeutics strategies.

INTRODUCTION

Head and Neck Squamous Cell Carcinoma (HNSCC) ranks 6th, amongst the most common cancers in the world, contributing to about 5% of all malignancies globally with a death rate of nearly half of total cases [1–3]. Tobacco usage such as smoking and chewing is the most common cause of head & neck and other cancers [4, 5]. Nonetheless, there are occurrences where individuals who never used tobacco or take liquor developed malignancy [6–9].

In the past decade DNA sequencing technologies such as Next generation sequencing technology (NGS) has identified key genomic signatures involved in cancer development and progression [10–14]. In recent years, large consortium based studies such as The Cancer Genome Atlas (TCGA) has provided detailed molecular map of HNSCC, and improved improved understanding of the role of genes in the pathogenesis of HNSCC [15–17]. Similarly, several studies has demonstrated the role of RNA-sequencing technique in cancers including HNSCC

and uncovered differential gene expression signatures of therapeutics and prognostics potential [11, 12, 18–25]. Identification of altered gene expression signature in cancer will help in identifying key biological pathways, leading to better understanding of the underlying molecular mechanism of the disease and can be use in precise therapy selection in the HNSCC management [26–28].

Role of tobacco in the pathological process has been analysed in recent studies and found genes, which are frequently mutated and abruptly expressed, possibly due to tobacco exposure [29–31]. Despite these studies, our understanding of underlying molecular mechanisms influenced by tobacco in HNSCC patients is limited. Hence, there is need of uncovering key genes and its molecular mechanism in Tobacco induce and non-Tobacco HNSCC tumors. Therefore, in the present study differential gene expression analysis of tobacco smoking (we term as TOB) and non-smokers (we term as N-TOB) in TCGA HNSCC data set was performed for stratifying tobacco dependent and independent biological networks and functional pathways.

RESULTS

Detection of differentially expressed genes

RNA-seq analysis of 86 matched Tumor normal pairs of HNSCC have identified total 5968 upregulated & 5698 downregulated genes in habituated (tobacco smokers) tumors; and in non-Habituated tumors (non-tobacco users), total 6698 & 5896 genes showed up and downregulation respectively with p -value <0.1 . Filtering based on fold change ($padj(FDR) < 0.05$, $\text{Log}_2\text{Foldchange} > 1$) showed total total 6371 and 7003 DEGs in TOB and N-TOB samples respectively. Further filtering was done to consider genes which shows significant ($padj(FDR) < 0.05$, $\text{Log}_2\text{Foldchange} > 2$) expression changes between Tumor and control samples. Total 2194 and 2073 genes were observed to show significantly altered expression in TOB and N-TOB tumors respectively (Figure 1, Supplementary Tables 1, 2). Among DEGs, 954 genes were up regulated and 1240 were down-regulated in TOB tumors. In case of N-TOB tumors, 890 and 1183 genes were found be up and downregulated respectively.

Additionally, Principal Component Analysis was performed using the PCA function from the sklearn Python module. Prior to performing PCA, the raw gene counts were normalized using the logCPM method, filtered by selecting the 2500 genes with most variable expression, and finally transformed using the Z-score method (Supplementary Figure 1).

A list of the top 20 most significantly up- and down-regulated DEGs of TOB and N-TOB tumors was shown in Figure 2. The most significant upregulated genes were CA9, GRIN2D, HOXC-AS2, GPR158, TGFBI

etc. and genes i. e. EDN3, FAM107A, MYZAP, DLG2, SLC38A3, PRH1 etc. were observed to be most significant downregulated genes in TOB tumors. In N-TOB tumors, most significant upregulated genes were i. e. CD276, ADAM12, RPLP0P2 etc. and downregulated were KRT36, CRISP3, MYOC etc.

In order to identify common Transcriptional signature shared between Tobacco smokers and non-tobacco smoker's patients, comparison of DEGs between the two categories was performed. Of the total DEGs, 1344 genes were observed to be altered in both categories whereas, 850 and 729 genes were observed to be unique in TOB and N-TOB tumors respectively (Figure 3).

Gene ontology and pathway enrichment analysis

In order to gain insights into the biological roles of the DEGs in HNSCC, we performed Gene ontology enrichment (p value < 0.05) analysis using Gorilla software. The GO terms for biological process was found to be enriched in muscle contraction (GO:0006936; $P = 2.14E-7$, $P = 1.95E-9$) and retinal homeostasis (GO:0001895; $P = 3.81E-6$, $P = 4.36E-7$) in both TOB and N-TOB tumors (Figure 4).

Further, biological significance for the DEGs was evaluated using KEGG and Reactome pathway enrichment analysis (p -value ≤ 0.05). The most significant pathway based on right sided hypergeometric test and bonferroni (pV) adjustments was muscle contraction ($pV = 2.55E-13$, $pV = 9.09E-16$), Extracellular matrix organization ($pV = 9.28E-09$, $pV = 1.70E-18$) and salivary secretion ($pV = 1.83E-07$, $pV = 5.76E-5$) in both TOB and N-TOB tumors (Figure 5). In addition to this, pathway related to Neuroactive ligand-receptor interaction ($pV = 1.49E-09$, $pV = 1.70E-18$) was observed to be enriched more in N-TOB tumors (Figure 5).

Protein-protein (PPI) interaction analysis of DEGs of Tobacco and non-Tobacco patients

The PPI network of DEGs were constructed for both TOB and N-TOB tumors by mapping genes onto STRING interactome database with higher confidence score cutoff of 0.900. The PPI network consists of 1552 nodes and 2576 edges in Tobacco DEGs, whereas N-TOB DEGs were observed to have 1913 nodes and 2969 edges. Hub genes were identified based on the number of interacting partners they have in the biological network. It was found that, genes such as ACTN2, PLK1, TPM2, MYL2, AURKA, AR, CCNB2, EGFR, TTN, CENPA, were observed to be top 10 hub genes in TOB tumors, whereas N-TOB tumors showed ACTN2, MYH6, MYL2, TPM2, TTN, TNNI2, TNNT3, TNNI1, TNNT1, TNNC1 as hub genes in the network (Figure 6). Of these, hub genes ACTN2, MYL2, and TPM2 were observed to be downregulated in both the cases. However, upregulated

hub genes i. e. PLK1 (*Degree* = 13; *Betweenness* = 1898), AURKA (*Degree* = 11; *Betweenness* = 1216), AR (*Degree* = 10; *Betweenness* = 4872), CCNB2 (*Degree* = 10; *Betweenness* = 1045) and EGFR (*Degree* = 9; *Betweenness* = 2234) were observed in Tobacco tumors.

miRNA enrichment analysis

miRNA Enrichment results were generated by analyzing the up-regulated and down-regulated gene sets using Enrichr. Significant results was obtained using cut-

off of p -value < 0.05 after applying Benjamini-Hochberg correction (FDR < 0.05). The most enriched miRNAs terms were hsa-miR-29 family, which are regulating DEGs in both TOB and N-TOB tumors (Table 1). Interestingly, TOB has more enriched miRNAs terms than N-TOB tumors. microRNA-29a/b/c were enriched miRNAs terms and observed to be upregulating 25, 33 and 24 target genes, in the TOB tumors respectively. In case of N-TOB, only hsa-miR-29b-3p* were found as enriched miRNA, possibly involved in upregulation of 20 target genes. Addition to this, TOB tumors has other enriched miRNAs

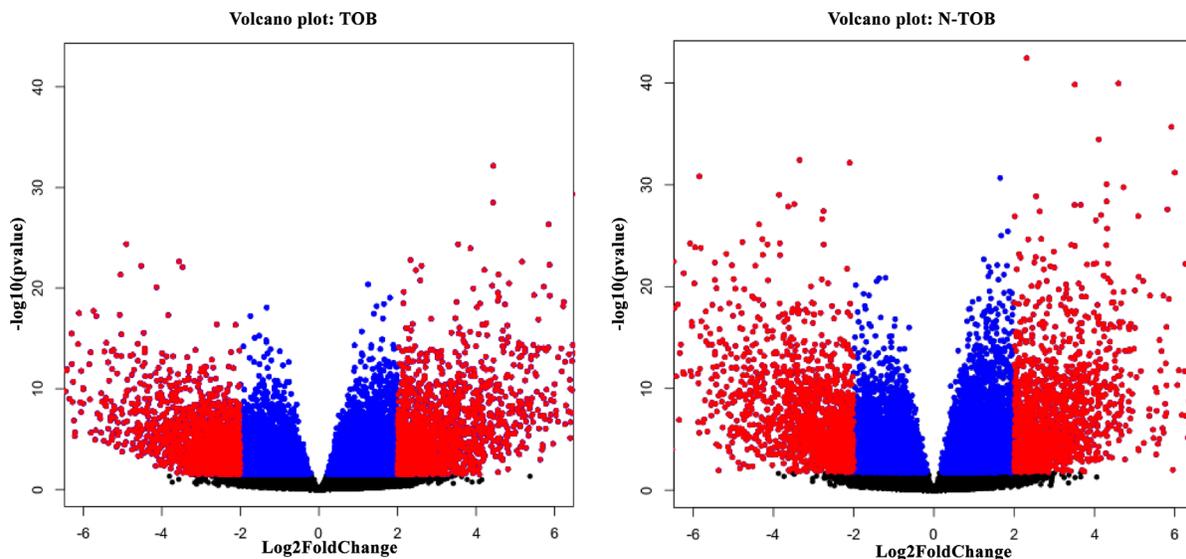


Figure 1: (A) Volcano plot showing gene expression significantly (FDR < 0.05; LogFC > 2) altered (highlighted in red colour) in TOB sample. (B) Volcano plot showing genes significantly (FDR < 0.05; LogFC > 2) altered in expression, highlighted in red colour in N-TOB patients.

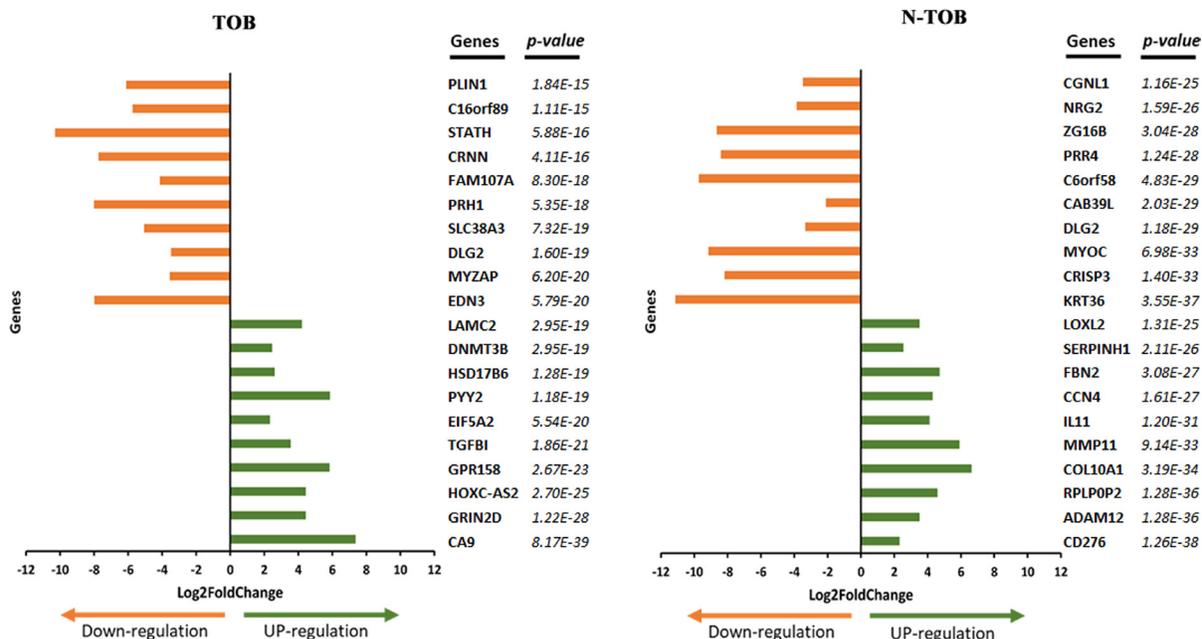


Figure 2: Top 20 DEGs identified in TOB and N-TOB tumors. Genes were ranked based on p -values ≤ 0.05 and adjusted false discovery rate using the Benjamini-Hochberg procedure.

i. e. hsa-miR-215-5p*, hsa-miR-193b-3p*, hsa-miR-192-5p*, hsa-miR-124-3p*, hsa-miR-615-3p*, hsa-miR-26b-5p*, hsa-miR-92a-3p* which regulates 75, 77, 81, 70, 50, 82 and 65 target genes respectively.

Potential biomarker

In order to find candidate genes, which behave differently and can be used to separate TOB tumors from N-TOB tumors, Log2fold change of gene of both tumors were compared. Surprisingly, three genes namely

PKLR1, CST1 and C170rf77 were observed to expressed differently in both TOB and N-TOB tumors, which means the same genes is upregulating in one category, whereas downregulating in another categories of HNSCC tumors (Figure 7).

External validation of the DEGs in non-matched tumors

Further, the consistency of the identified altered gene expression profile was checked in other non-

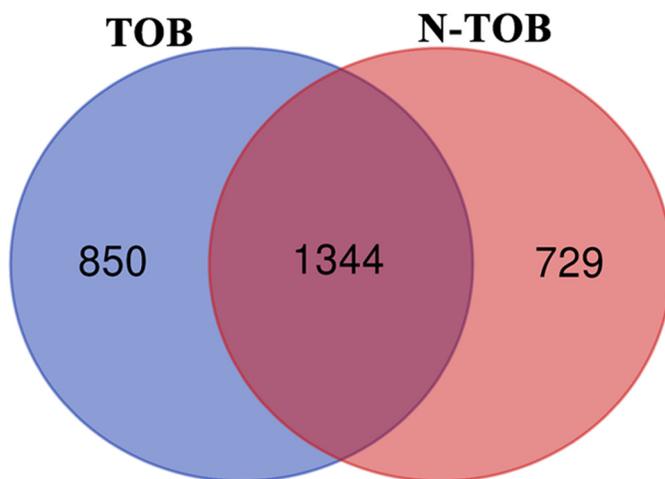


Figure 3: Venn diagram showing DEGs common and unique between Tobacco and non-Tobacco patients. Genes highlighted in blue and pink red are showed altered expression in Tobacco and non-Tobacco patients only; and genes highlighted in dark red are observed to be altered in both Tobacco and non-Tobacco patients.

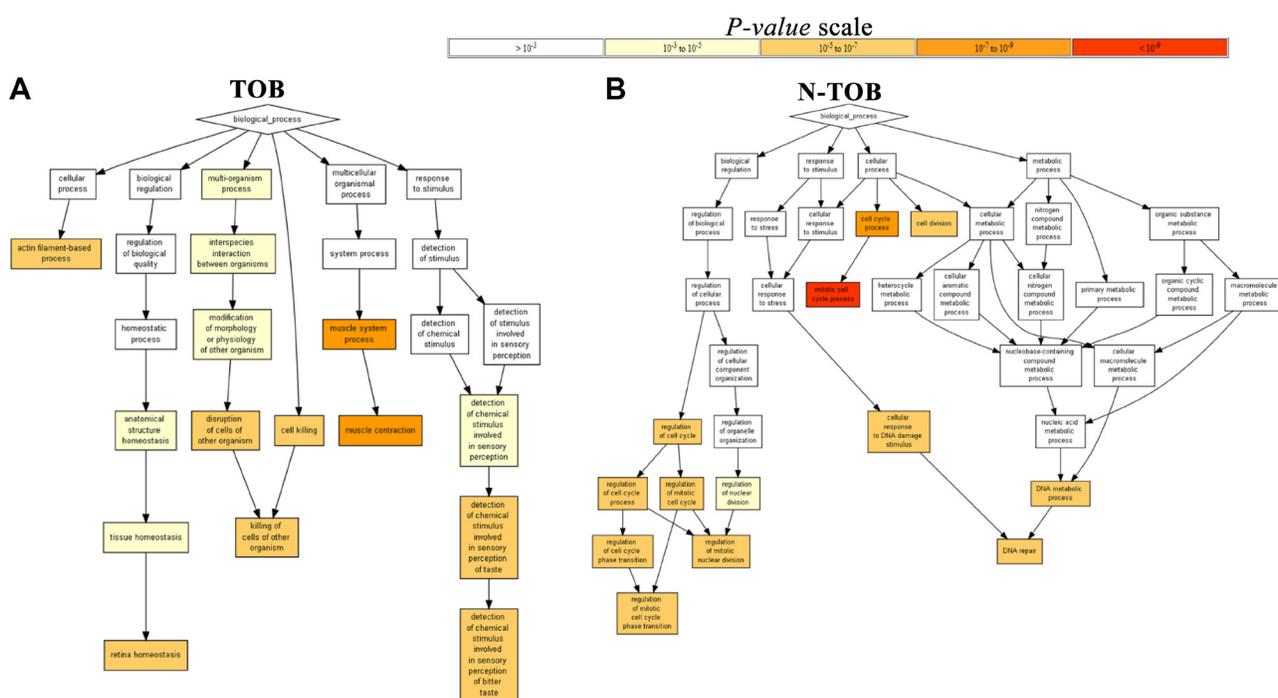


Figure 4: Gene Ontology Enrichment Analysis Results. The figure contains interactive charts displaying the results of the Gene Ontology (GO) enrichment analysis generated using GOrilla. The x-axis indicates the enrichment score for each term. Significant terms are highlighted in bold.

matched tumors samples of tobacco smoking and non-smoking patients. For this, non-matched tumors samples having tobacco smoking history and those who never smoked, were searched in TCGA database. Total 224 and 151 samples of smoking habit and non-smokers patients were selected respectively. For DGE analysis, the normal control from the matched pairs (as used in the above analysis) were taken and process further using same pipeline (see method).

Results showed, nearly 83% of matched tumor DEGs ($p_{adj}(FDR) < 0.05$, $\text{Log}_2\text{Foldchange} > 1$), were also observed to be deregulated in non-matched tumors, whereas 88% (1722 of total 2194 in TOB; 1643 of total 2073 in N-TOB) concordance was observed with fold change 2 ($\text{Log}_2\text{Foldchange} > 2$) (Supplementary Tables 3, 4). In addition to similar DEGs pattern, the top differentially expressed genes and three candidate genes of matched pairs were also found to be abruptly expressed in

non-matched tumors. The expression pattern of candidate genes i. e. PKLR1, CST1 and C170rf77 in non-matched samples showed similar pattern with little low fold change as compared to matched tumors (Supplementary Figure 2).

DISCUSSION

Tobacco usage is a major factor in the development of various cancers including Head and neck squamous cell carcinoma (HNSCC). Present study attempted to identify differential expressed genes (DEGs) signature and its underlying possible molecular mechanism in 86 matched tumor normal pairs of 43 HNSCC patients having Tobacco smoking (17 samples) and non-smoking (26 samples) history. Findings was further validated on 375 non-matched HNSCC tumors having tobacco smoking history (tobacco smoking = 224; non-smokers = 151) available in the TCGA database. Total, 2194 (954 up

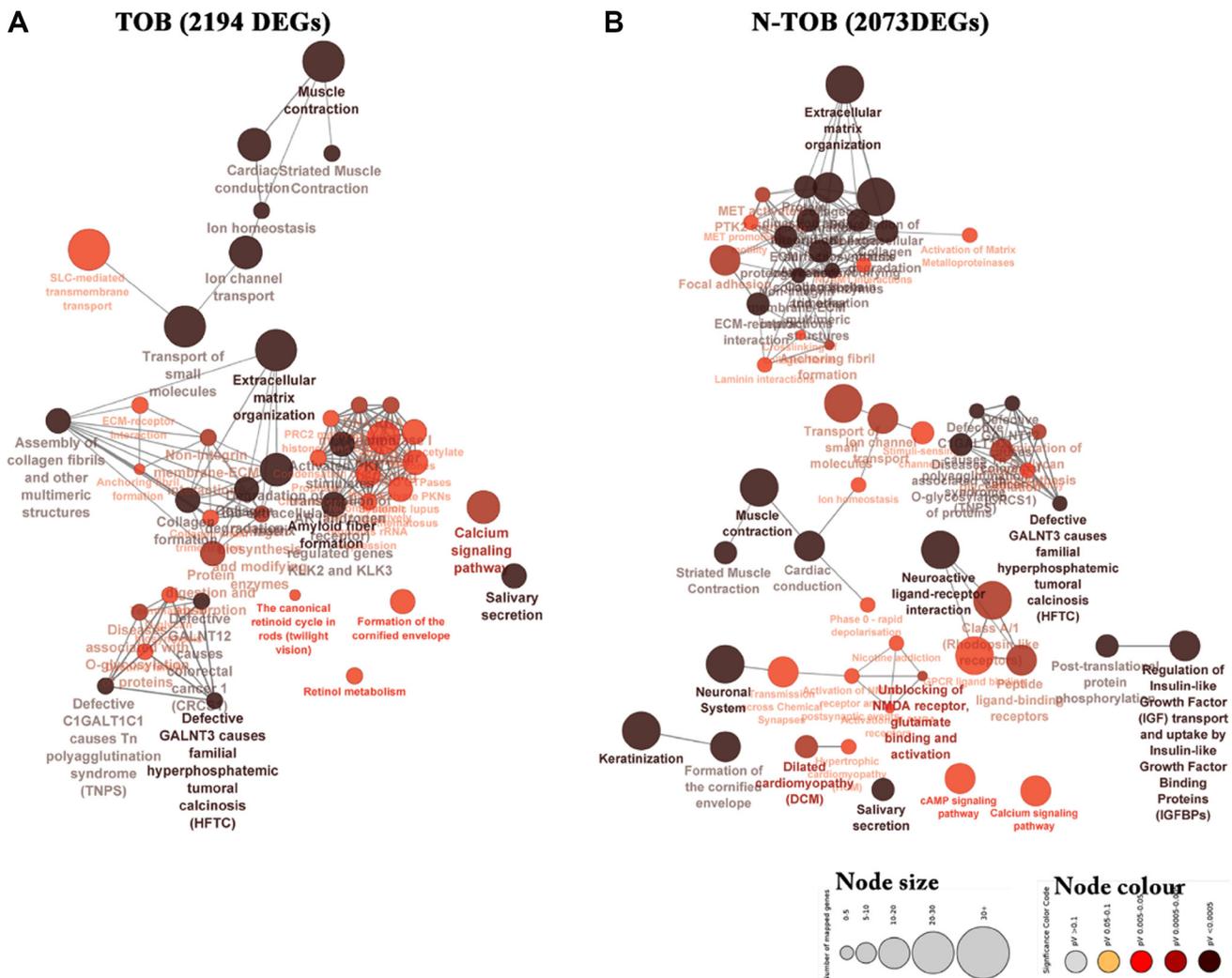


Figure 5: Network representing enriched pathways integrated KEGG and Reactome pathways of both TOB and N-TOB tumors. Highest significance of enriched pathway was obtained using advanced statistical settings such as Hyper-geometric (right-handed) enrichment distribution tests, p -value < 0.05 , and Bonferroni adjustment. The size and colour represents number of DEGs involved and enrichment significance respectively- deeper the colour, the higher the enrichment significance.

regulated and 1240 down regulated) and 2073 (890 up regulated and 1183 down regulated) genes were observed to be differentially expressed in TOB and N-TOB tumors respectively (Figure 1).

In order to uncover the biological roles of the DEGs in each TOB and N-TOB tumors, we performed a gene ontology (GO) enrichment analysis. GO terms related to muscle contraction, retinal homeostasis were enriched in TOB tumors; whereas cell cycle, cell division and DNA repair in N-TOB tumors were observed to be highly enriched biological process. Genes related to muscle contraction has been reported to be altered in Oral squamous cell carcinoma (OSCC) and suggests the presence of myofibroblasts in tumor stroma of patients with lymph node involvement [32].

Pathway enrichment analysis identified pathways related to extracellular matrix organization, muscle contraction, calcium signaling and salivary secretion to be the most significantly enriched pathways in both categories. In case of N-TOB patients, pathways of Neuroactive ligand-receptor interaction were found to

be enriched. Extracellular matrix organization genes regulates cell proliferation, adhesion, differentiation, death and alteration leads to tissue fibrosis and cancer [33], whereas muscle contraction pathways were reported to be deregulated in lymph node metastasis OSCC [32]. Deregulation of the calcium signal is reported to be involved in tumor initiation, angiogenesis, progression and metastasis [34, 35] and a promising pathway for therapeutics strategy designing [26].

In recent years, effects of etiological factors such as lifestyle, diet, environment and exposure on molecular pathogenesis has been well studied due to emergence of the field i. e. molecular pathological epidemiology (MPE) [36, 37]. MPE is an integrative field, which incorporates molecular pathology into epidemiologic research and dissect the link between heterogeneous etiological factors such as environmental, dietary habits, microbiota, lifestyle, exposure and genetic factors with tumor initiation and progression in cancers of breast, prostate, lung and colorectal [38–42]. Moreover, the present study identified calcium-signaling pathway to

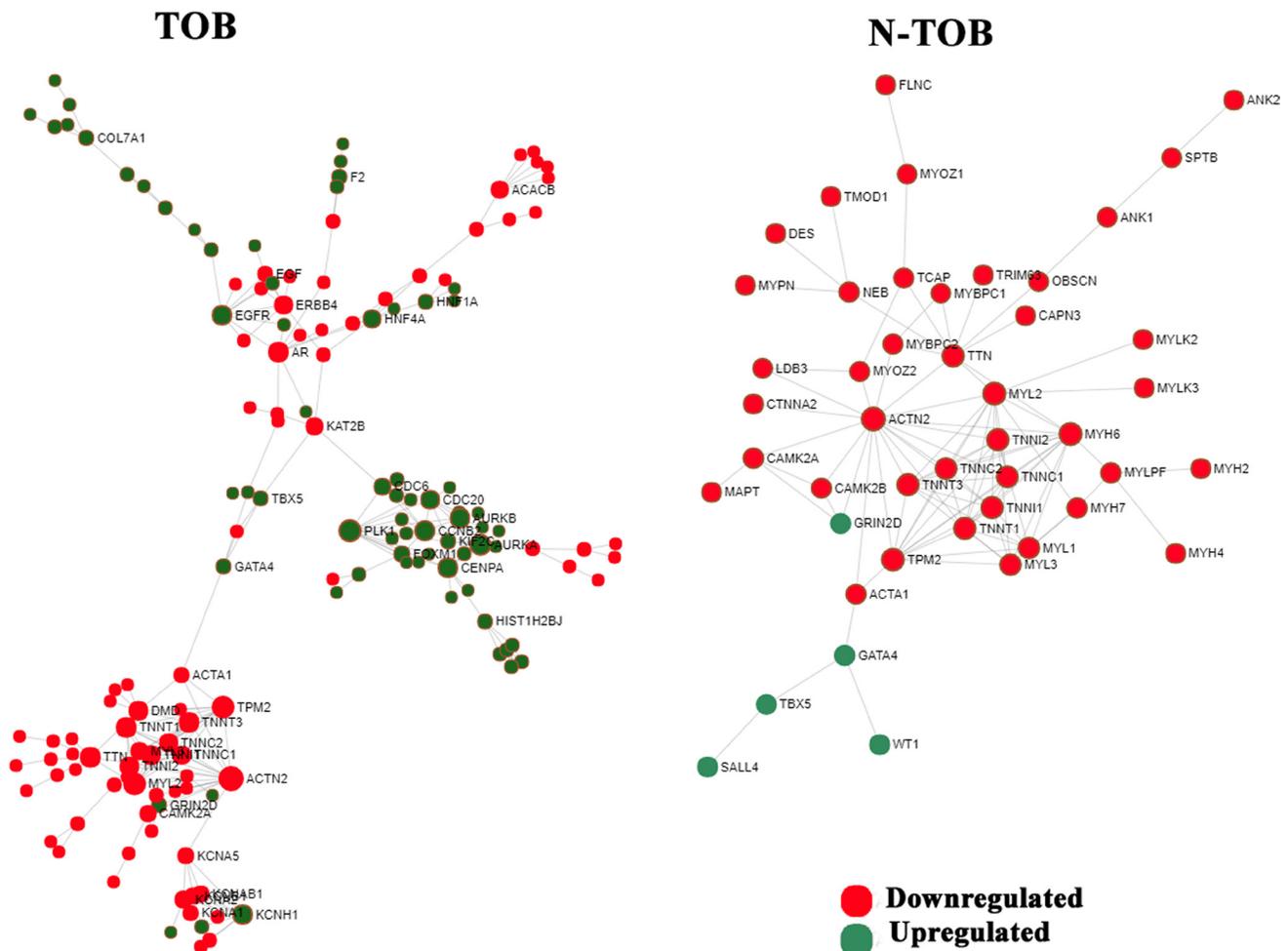


Figure 6: Network based meta-analysis of hub genes. Zero-order interaction network of DEGs obtained from RNA-seq data using force-directed algorithm with Fruchterman-Rengold layout; green nodes represents over-expressed and red nodes represents under-expressed genes.

Table 1: miRNA enrichment analysis results

Rank	miRNA	P-value	FDR	Target	Rank	miRNA	P-value	FDR	Target
		TOB					N-TOB		
1	hsa-miR-215-5p*	6.08E-25	1.62E-21	75 upregulated	1	mmu-miR-29a-3p*	0.000002	0.004102	9 upregulated
2	hsa-miR-193b-3p*	5.10E-23	6.78E-20	77 upregulated	2	mmu-miR-29b-3p*	0.000004	0.00419	9 upregulated
3	hsa-miR-192-5p*	2.30E-21	2.03E-18	81 upregulated	3	hsa-miR-29b-3p*	0.00001	0.006585	20 upregulated
4	hsa-miR-29b-3p*	1.94E-14	1.29E-11	33 upregulated	4	mmu-miR-124-5p	0.002887	1	3 downregulated
5	hsa-miR-29a-3p*	1.34E-08	7.12E-06	25 upregulated	5	mmu-miR-1897-5p	0.017333	1	4 downregulated
6	hsa-miR-29c-3p*	2.73E-08	1.21E-05	24 upregulated	6	mmu-miR-3098-5p	0.019988	1	2 downregulated
7	hsa-miR-124-3p*	6.95E-08	2.64E-05	70 upregulated	7	mmu-miR-467e-5p	0.019988	1	2 downregulated
8	hsa-miR-615-3p*	8.43E-08	2.80E-05	50 upregulated	8	mmu-miR-467h	0.019988	1	2 downregulated
9	hsa-miR-26b-5p*	3.43E-07	1.01E-04	82 upregulated	9	mmu-miR-543-3p	0.023748	1	3 downregulated
10	hsa-miR-92a-3p*	1.06E-06	2.83E-04	65 upregulated	10	hsa-miR-502-5p	0.037752	1	5 downregulated

The table displaying the results of the miRNA enrichment analysis generated using Enrichr. Every row represents a miRNA; significant (p -value < 0.05; FDR < 0.005) miRNAs are highlighted in bold. Displays results generated using the miRTarBase library.

be enriched in TOB tumors, and may have association of smoking exposure with molecular pathogenesis of HNSCC. Calcium signaling is known to be affected by etiologic factors such as cigarette smoke exposure and results into increase calcium levels in cells and plays an important role in proliferation, migration, invasion or tumor growth and metastasis in cancers including drug resistance of pancreatic cancer [43–47]. Hence, in HNSCC it is possible that cigarette smoke increases intracellular

calcium levels and affects normal cellular functions by promoting proliferation, motility, invasion and survival. Therefore, targeting calcium-signaling pathway in future HNSCC research, will provide detailed molecular insights of smoking related cancers and help in developing more effective treatment strategies in HNSCC. Neuroactive ligand-receptor interaction pathway has been reported to be implicated in many cancer types [26, 48–50] including OSCC [51] and represented promising candidates for

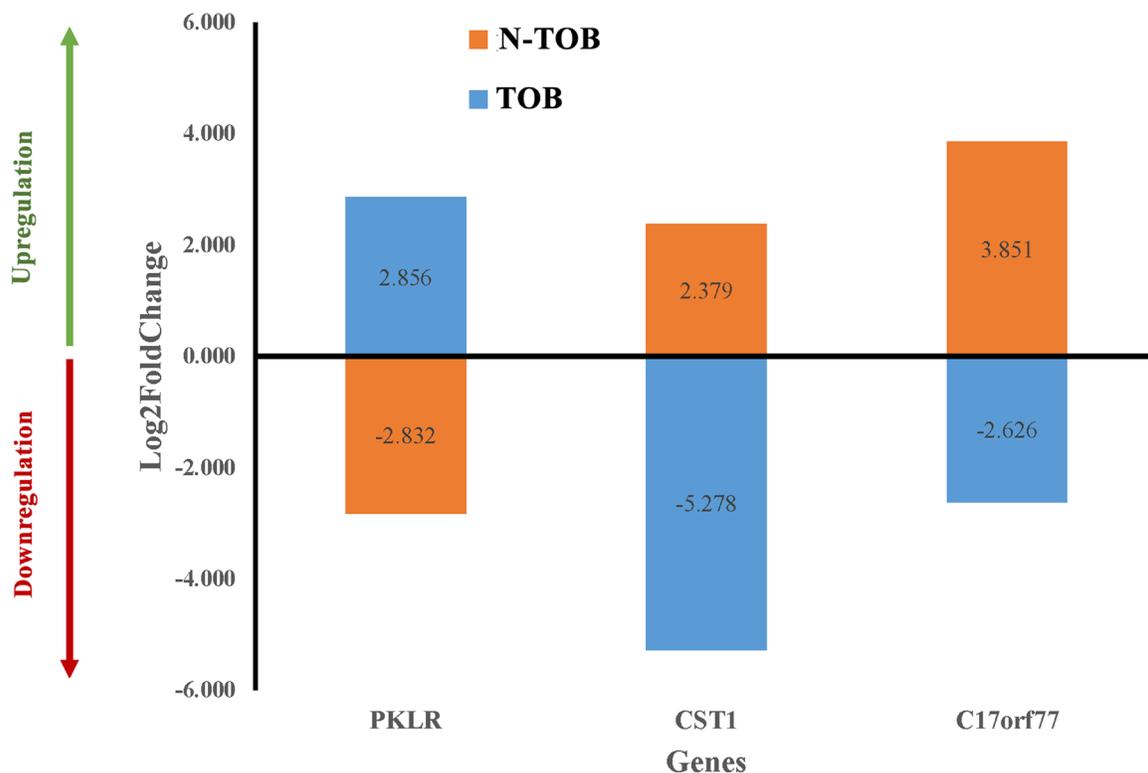


Figure 7: Differential gene expression pattern of key genes in both TOB and N-TOB tumors. Log2fold change values of genes showing opposite regulation of the same gene in between TOB and N-TOB tumors.

therapeutic intervention in OSCC patients [26]. Therefore, targeting Neuroactive ligand-receptor interaction pathway in N-TOB tumors may open new avenues in HNSCC therapeutics intervention.

PPI network analysis elucidated detailed interaction among TOB and N-TOB DEGs and highlighted top 10 centrality hub genes. Interestingly, MYH6 (*Degree* = 13; *Betweenness* = 183) hub gene is observed to be downregulated only in N-TOB tumors, hence suggested to play important role in tobacco independent disease development and progression. Role of myosin heavy chain 6 (MYH6) gene has earlier been associated in familial dilated cardiomyopathy [51, 52]. However, some study using RNAi *in vivo* screen identified MYH6 frequently altered in HNSCC MOC lines [53] and consider as novel putative cancer genes [54]. EGFR gene was observed to be upregulated in TOB tumors only, hence may be possibly involved in tobacco mediated carcinogenesis. Previously it was reported [55] that EGFR was overexpressed in about 30% of human epithelial tumors including HNSCCs [55, 56]. Ansell et al. 2016 found that the amount of EGF had a determinant function in cell proliferation and the response to treatment of cetuximab in tongue cancer, and emerged as a potential predictive biomarker of poor cetuximab response [57]. High level of EGFR enhanced proliferation, promote tumor growth leading to poor prognosis [58–60] and resistant to radiotherapy [61–63].

The prominent nodes with lowest *p*-values signify presence of muscle contraction pathways and associated altered expression of genes such as, TTN, TNNT3, TNN2 and MYL2 in both TOB and N-TOB tumors. This is in concordance with earlier microarray based gene expression profiling study, which revealed a distinction signature pattern belonging to muscle contraction pathway [32]. Furthermore, different studies have already proved the presence of myofibroblasts of tumor stroma by gene expression analysis in lymph nodes establishing their role in metastatic migration, invasion and association with poor survival rate in OSCC patients [64–66]. Troponin (TNN) gene family reported to be altered in TSCC and might serve as future clinical prognostic marker for TSCC [67]. These genes might play a regulatory role in control of cellular locomotion, cytoplasmic streaming, and cytokinesis in non-muscle cells. Further studies are warranted to elucidate the role of muscle related genes in OSCC carcinogenesis.

Polo-like kinase 1 (PLK1) encodes the protein which is important regulators of the cell cycle and cell division. PLK1 has been reported to be upregulated in various cancers [67, 68] and their expression level are associated with a poor prognosis and a lower overall-survival in many cancers [69–71] including HNSCC [72]. Recent study has demonstrated this as a promising target for chemopreventive treatment of preneoplastic cells, and could be applied to prevent HNSCC and local relapses [72, 73]. Aurora kinase A (AURKA) has been

implicated in numerous types of cancer [72–78] and reported to promotes cell migration & invasion [79] tumor progression in patients with laryngeal squamous cell carcinoma (LSCC) [80]. Both PLK1 and AURKA hub genes were observed to be upregulated, which is in concordance with the above earlier reports.

Interestingly, three genes i. e. PKLR, CST1 and C17orf77 were observed to expressed differently in the TOB and N-TOB tumors, means there is upregulation in TOB, whereas downregulation in N-TOB and vice versa. Surprisingly, role of PKLR gene in HNSCC cancer not well studied. However, recent study, have demonstrated that PKLR promotes colon cancer cell metastatic colonization of the liver by increasing glutathione synthesis, which is the primary endogenous antioxidant [80, 81]. CST1 gene encodes Cysteine proteases (CST1) enzymes, generally involve in protein degradation, which is associated with a diversity of diseases and facilitates the development and progression of cancer cells [82–84]. Cystatins (CST1) play important roles in tumor invasion and metastasis in colorectal cancer [84–86]. Recent study reported that elevated expression of CST1 may promotes breast cancer progression and predicts a poor prognosis [87]. Role of Chromosome 17 Open Reading Frame 77 (C17ORF77) gene in cancer has not been well studied yet. Therefore, these three genes will be gene of interest in future studies for better understanding of molecular insight of tobacco habituated and non-habituated associated HNSCC.

Additionally, the present study also identified enriched riboswitches i. e. miRNAs regulating their target genes for understanding regulatory cascade of tobacco and non-tobacco induced carcinogenesis. microRNA-29a/b/c were observed to be the most enriched miRNAs (*P*-value < 0.05) across both TOB and N-TOB categories, as it regulate genes, which are detected to be upregulated in the present study. In earlier studies, miR-29 family members (i. e., miR-29a, miR-29b, and miR-29c) are reported to be frequently downregulated in many cancers [87–91] including their antitumor functions in HNSCC [91–93]. Therefore, downregulation of *miR-29* family leading to loss of tumor suppressor activity, which subsequently results in the upregulation of several oncogenic genes in HNSCC. TOB tumors possesses other enriched miRNAs i. e. hsa-miR-215-5p*, hsa-miR-92a-3p*, hsa-miR-26b-5p* etc. Among these, the largest number of upregulated i. e. 82 genes were found associated with miR-26b, which is reported to be downregulated in various cancers, including OSCC, indicating its tumor suppressive nature in oral cancer progression [93, 94]. hsa-miR-26b-5p was also reported as tumor suppressors in cancers of oral cavity [2, 89] and suggested that, loss of tumor-suppressive miR-26a/b enhanced cancer cell migration and invasion in OSCC [2]. Therefore, identification of gene networks regulated by these tumor suppressor miRNAs may provide novel insights into designing therapeutic strategies in HNSCC.

Table 2: Patients details including age, stage, Tobacco habit, alcohol consumption status, demography and gender shown in the table

Study Characteristic	No.
Age at diagnosis, years	
Mean \pm SD	64.9 \pm 14.5
Median	42
Clinical stage	
Stage I	2
Stage II	15
Stage III	8
Stage IVa	17
Not reported	1
Tobacco habit (Smoking)	
Yes	17
No	26
Alcohol consumption status	
Yes	22
No	19
Not reported	2
Gender	
Male	29
Female	14
Demographic details	
White	39
Asian	1
Black or african american	2
not reported	1

Matched Tumor/Normal pairs of each HNSCC patient was retrieved from TCGA database.

METHODS

Identification of relevant datasets and differential expression analysis

RNA-seq data was retrieved from The Cancer Genome Atlas, TCGA-HNSC project for gene expression profiling of HNSCC patients for our study. Total 86 Tumor/Normal pairs samples of 43 HNSCC patients were found having tobacco habit details (Table 2). Of these, 34 sample pairs from 17 Tobacco habituated patients and 52 sample pairs from 26 non-habituated patients were considered for further analysis. Raw read counts (HTSeq read counts) of total 86 Tumor/Normal pairs were downloaded using TCGA GDC Data Transfer Tool (Figure 8).

Raw read counts were normalized to log₁₀-Counts Per Million (logCPM), followed by the application of a log₁₀-transform using DESeq R package and differentially expressed genes (DEGs) were identified for each TOB and N-TOB tumor.

Gene ontology (GO) and pathway enrichment analysis

In order to understand the biological significance of the DEGs, we performed Gene Ontology (GO) enrichment analysis using GOrilla tool [95] with *p*-value threshold of 0.05. Enriched pathways of DEGs were identified using KEGG and Reactome pathway databases integrated in ClueGO [95, 96] v2.3.5 plugin of Cytoscape 3.4.0 with

p-value cut-off 0.05. Additional filtering was performed using advance statistical option such as Two-sided hypergeometric test for calculating importance of each term and Bonferroni step-down for *p*-value correction.

PPI network construction

Protein-protein interaction (PPI) of analysis was performed for DEGs of both categories to identify key proteins and their significance in the complex biological systems. We used STRING interectome database [97] with high confidence score cut-off (0.900) and constructed PPI network of DEGs using NetworkAnalyst [98]. Further, filtering of zero-order network was performed to retain

only seed proteins that directly interact with each other and finally highly interconnected hub nodes was identified based on two centrality measure such as degree and betweenness centrality.

miRNA enrichment analysis of DEGs

microRNAs enrichment analysis was performed to identify regulatory cascade of DEGs of HNSCC. Two databases *viz.* TargetScan [98, 99] and MiRTarBase [100] integrated in Enrichr [100, 101] tool was used to retrieved enriched (*p*-value<0.05) riboswitches involve in transcriptional regulation of DEGs.

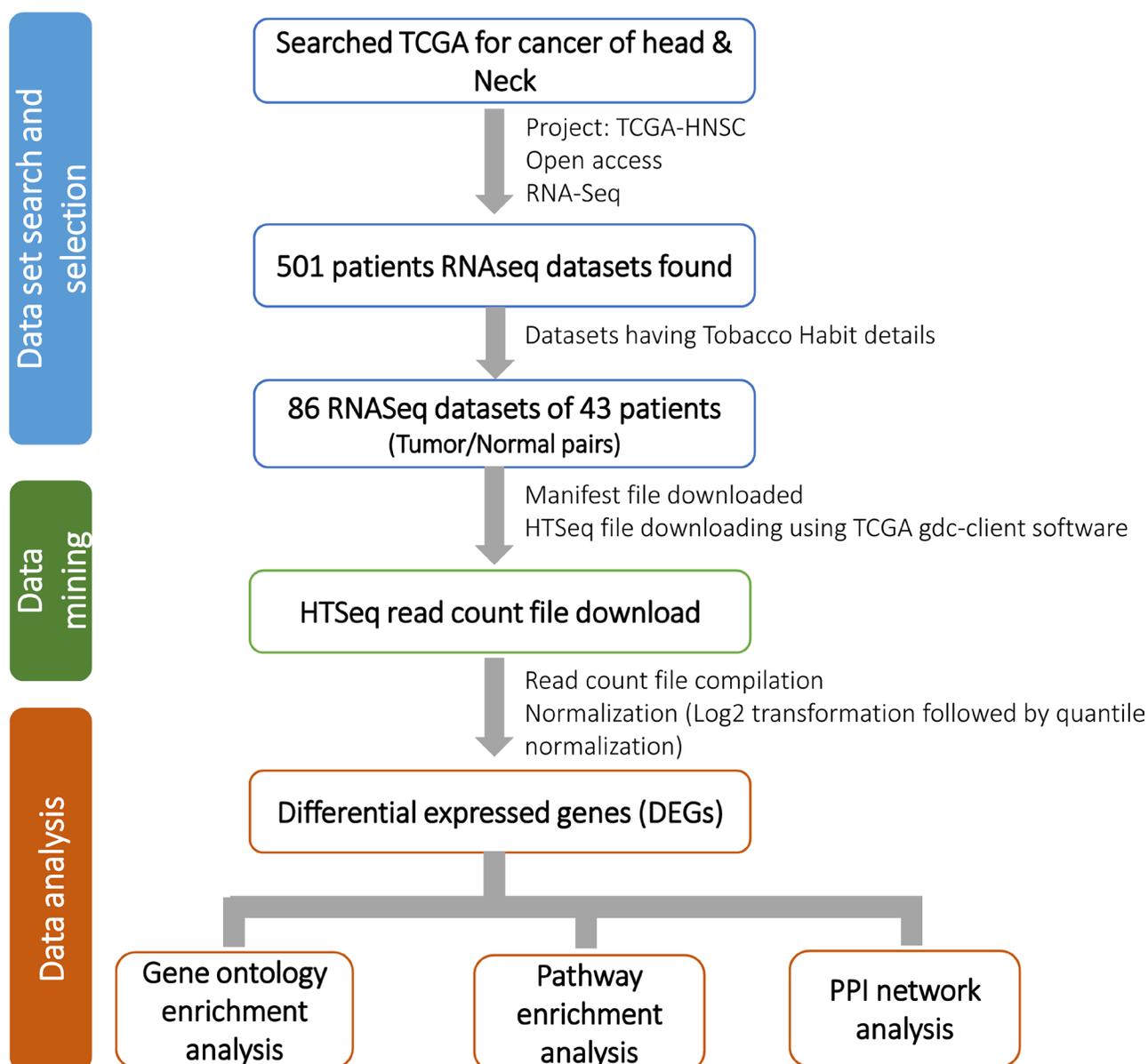


Figure 8: Detail workflow of the study. Study workflow consists of three components i. e. 1) data set search and selection of relevant data; 2) mining of the selected data (RNA-seq read count) from TCGA; and 3) data analysis which includes differential gene expression (DEGs) analysis and annotations.

CONCLUSIONS

In conclusion, we performed analysis of HNSCC RNA-seq data and identified key deregulated genes associated with functional pathways and biological networks, which may be contributing in Tobacco dependent and independent carcinogenesis of the disease. Pathway analysis identified key DEGs involves in calcium signaling, and suggests playing key role in tobacco dependent molecular pathogenesis in HNSCC. Our findings suggest that, three genes i. e. PKLR, CST1 and C17orf77 may hold putative role in both pathogenesis of smoking and non-smoking related HNSCC tumors and can be consider as potential biomarker for separating these tumors from each other. The identified differentially expressed genes can be integrated in multiple biological pathways and provide improved understanding of key molecular mechanisms in smoking and non-smoking associated HNSCC tumors, and be useful in precision therapy selection. However, further research is needed to understand influence of smoking and other etiological factors such as environment, lifestyle, diet, alcohol consumption and HPV infection on the molecular pathogenesis of the disease.

Author contributions

JD, PS, SBB and IS conceived and designed the study. IS, SBB and AA performed the gene expression and network analysis whereas GA and MG performed the pathway and miRNA analysis. CJ and JD supervised the overall study and provided necessary resources. IS, GA, PS and JD wrote the manuscript.

ACKNOWLEDGMENTS

The authors wish to acknowledge all the patients and investigators who contributed to The Cancer Genome Atlas (TCGA) studies analyzed here. We are grateful to acknowledge Gujarat State Biotechnology Mission (GSBTM) and Gujarat Biotechnology Research Centre (GBRC), Department of Science & Technology (DST), Government of Gujarat, India for providing necessary support throughout the research.

CONFLICTS OF INTEREST

The authors declare no competing interests.

FUNDING

This study was a part of intramural research project (GBRC/GoG-DST/JD2/HLT-13/2017-2018) of Gujarat Biotechnology Research Centre (GBRC) and supported by Department of Science & Technology (DST), Government of Gujarat, India.

REFERENCES

1. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW, and Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? *Lancet*. 2012; 380:258–71. [https://doi.org/10.1016/S0140-6736\(12\)60735-1](https://doi.org/10.1016/S0140-6736(12)60735-1). [PubMed]
2. Fukumoto I, Hanazawa T, Kinoshita T, Kikkawa N, Koshizuka K, Goto Y, Nishikawa R, Chiyomaru T, Enokida H, Nakagawa M, Okamoto Y, Seki N. MicroRNA expression signature of oral squamous cell carcinoma: functional role of microRNA-26a/b in the modulation of novel cancer pathways. *Br J Cancer*. 2015; 112:891–900. <https://doi.org/10.1038/bjc.2015.19>. [PubMed]
3. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010; 127:2893–917. <https://doi.org/10.1002/ijc.25516>. [PubMed]
4. Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, Boyle P. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer*. 2008; 122:155–64. <https://doi.org/10.1002/ijc.23033>. [PubMed]
5. Momi N, Kaur S, Rachagani S, Ganti AK, Batra SK. Smoking and microRNA dysregulation: a cancerous combination. *Trends Mol Med*. 2014; 20:36–47. <https://doi.org/10.1016/j.molmed.2013.10.005>. [PubMed]
6. Farshadpour F, Hordijk GJ, Koole R, Slootweg PJ. Non-smoking and non-drinking patients with head and neck squamous cell carcinoma: a distinct population. *Oral Dis*. 2007; 13:239–43. <https://doi.org/10.1111/j.1601-0825.2006.01274.x>. [PubMed]
7. Subramanian J, Govindan R. Lung cancer in never smokers: a review. *J Clin Oncol*. 2007; 25:561–70. <https://doi.org/10.1200/JCO.2006.06.8015>. [PubMed]
8. Kruse AL, Bredell M, Grätz KW. Oral squamous cell carcinoma in non-smoking and non-drinking patients. *Head Neck Oncol*. 2010; 2:24. <https://doi.org/10.1186/1758-3284-2-24>. [PubMed]
9. Scheidt JH, Yurgel LS, Cherubini K, de Figueiredo MA, Salum FG. Characteristics of oral squamous cell carcinoma in users or non users of tobacco and alcohol. *Rev Odonto Ciênc*. 2012; 27:69–73. <https://doi.org/10.1590/S1980-65232012000100013>.
10. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012; 481:506–10. <https://doi.org/10.1038/nature10738>. [PubMed]
11. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattri J, et al. The clonal and mutational evolution spectrum of primary triple-negative breast

- cancers. *Nature*. 2012; 486:395–99. <https://doi.org/10.1038/nature10933>. [PubMed]
12. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, Guillory J, Ha C, Dijkgraaf GJ, et al. Recurrent R-spondin fusions in colon cancer. *Nature*. 2012; 488:660–64. <https://doi.org/10.1038/nature11282>. [PubMed]
 13. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, et al, and Oslo Breast Cancer Consortium (OSBREAC). The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012; 486:400–04. <https://doi.org/10.1038/nature11017>. [PubMed]
 14. Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, Miller CA, Welch JS, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013; 502:333–39. <https://doi.org/10.1038/nature12634>. [PubMed]
 15. Agrawal N, Frederick MJ, Pickering CR, Bettgowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. 2011; 333:1154–57. <https://doi.org/10.1126/science.1206923>. [PubMed]
 16. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011; 333:1157–60. <https://doi.org/10.1126/science.1208130>. [PubMed]
 17. Gaykalova DA, Mambo E, Choudhary A, Houghton J, Buddavarapu K, Sanford T, Darden W, Adai A, Hadd A, Latham G, Danilova LV, Bishop J, Li RJ, et al. Novel insight into mutational landscape of head and neck squamous cell carcinoma. *PLoS One*. 2014; 9:e93102. <https://doi.org/10.1371/journal.pone.0093102>. [PubMed]
 18. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, Virtanen C, Bradner JE, Bader GD, Mills GB, Pe'er D, Moffat J, Neel BG. Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell*. 2016; 164:293–309. <https://doi.org/10.1016/j.cell.2015.11.062>. [PubMed]
 19. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489:519–25. <https://doi.org/10.1038/nature11404>. [PubMed]
 20. Gonzalez HE, Gujrati M, Frederick M, Henderson Y, Arumugam J, Spring PW, Mitsudo K, Kim HW, Clayman GL. Identification of 9 genes differentially expressed in head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg*. 2003; 129:754–59. <https://doi.org/10.1001/archotol.129.7.754>. [PubMed]
 21. Chen J, Fu G, Chen Y, Zhu G, Wang Z. Gene-expression signature predicts survival benefit from postoperative chemoradiotherapy in head and neck squamous cell carcinoma. *Oncol Lett*. 2018; 16:2565–78. <https://doi.org/10.3892/ol.2018.8964>. [PubMed]
 22. Kim E, Palackdharry S, Yaniv B, Mark J, Tang A, Wilson K, Patil Y, Janssen E, Conforti L, Takiar V, Hinrichs B, Wise-Draper TM. Gene expression signature after one dose of neoadjuvant pembrolizumab associated with tumor response in head and neck squamous cell carcinoma (HNSCC). *J Clin Oncol*. 2018; 36:6059. https://doi.org/10.1200/JCO.2018.36.15_suppl.6059.
 23. Sok JC, Kuriakose MA, Mahajan VB, Pearlman AN, DeLacure MD, Chen FA. Tissue-specific gene expression of head and neck squamous cell carcinoma *in vivo* by complementary DNA microarray analysis. *Arch Otolaryngol Head Neck Surg*. 2003; 129:760–70. <https://doi.org/10.1001/archotol.129.7.760>. [PubMed]
 24. Johnson ME, Cantalupo PG, Pipas JM. Identification of Head and Neck Cancer Subtypes Based on Human Papillomavirus Presence and E2F-Regulated Gene Expression. *MSphere*. 2018; 3:e00580–17. <https://doi.org/10.1128/mSphere.00580-17>. [PubMed]
 25. Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, Taylor AM, Wang C, Walter V, Akbani R, Byers LA, Creighton CJ, Coarfa C, et al, and Cancer Genome Atlas Research Network. Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. *Cell Rep*. 2018; 23:194–212.e6. <https://doi.org/10.1016/j.celrep.2018.03.063>. [PubMed]
 26. Zhang H, Liu J, Fu X, Yang A. Identification of Key Genes and Pathways in Tongue Squamous Cell Carcinoma Using Bioinformatics Analysis. *Med Sci Monit*. 2017; 23:5924–32. <https://doi.org/10.12659/MSM.905035>. [PubMed]
 27. Zhang X, Cha IH, Kim KY. Highly preserved consensus gene modules in human papilloma virus 16 positive cervical cancer and head and neck cancers. *Oncotarget*. 2017; 8:114031–40. <https://doi.org/10.18632/oncotarget.23116>. [PubMed]
 28. Braicu C, Catana C, Calin GA, Berindan-Neagoe I. NCRNA combined therapy as future treatment option for cancer. *Curr Pharm Des*. 2014; 20:6565–74. <https://doi.org/10.2174/1381612820666140826153529>. [PubMed]
 29. Rawal RM, Joshi MN, Bhargava P, Shaikh I, Pandit AS, Patel RP, Patel S, Kothari K, Shah M, Saxena A, Bagatharia SB. Tobacco habituated and non-habituated subjects exhibit different mutational spectrums in head and neck squamous cell carcinoma. *3 Biotech*. 2015; 5:685–96. <https://dx.doi.org/10.1007/s13205-014-0267-0>. [PubMed]
 30. Irimie AI, Braicu C, Cojocneanu R, Magdo L, Onaciu A, Ciocan C, Mehterov N, Dudea D, Buduru S, Berindan-Neagoe I. Differential Effect of Smoking on Gene Expression in Head and Neck Cancer Patients. *Int J Environ Res Public Health*. 2018; 15:1558. <https://doi.org/10.3390/ijerph15071558>. [PubMed]

31. Ghasemi F, Prokopec SD, MacNeil D, Mundi N, Gameiro SF, Howlett C, Stecho W, Plantinga P, Pinto N, Ruicci KM, Khan MI, Yoo J, Fung K, et al. Mutational analysis of head and neck squamous cell carcinoma stratified by smoking status. *JCI Insight*. 2019; 4:e123443. <https://doi.org/10.1172/jci.insight.123443>. [PubMed]
32. Mazzoccoli G, Castellana S, Carella M, Palumbo O, Tiberio C, Fusilli C, Capocéfalo D, Biagini T, Mazza T, Lo Muzio L. A primary tumor gene expression signature identifies a crucial role played by tumor stroma myofibroblasts in lymph node involvement in oral squamous cell carcinoma. *Oncotarget*. 2017; 8:104913–27. <https://doi.org/10.18632/oncotarget.20645>. [PubMed]
33. Hynes RO. The extracellular matrix: not just pretty fibrils. *Science*. 2009; 326:1216–19. <https://doi.org/10.1126/science.1176009>. [PubMed]
34. Cui C, Merritt R, Fu L, Pan Z. Targeting calcium signaling in cancer therapy. *Acta Pharm Sin B*. 2017; 7:3–17. <https://doi.org/10.1016/j.apsb.2016.11.001>. [PubMed]
35. Pinton P, Giorgi C, Siviero R, Zecchini E, Rizzuto R. Calcium and apoptosis: ER-mitochondria Ca²⁺ transfer in the control of apoptosis. *Oncogene*. 2008; 27:6407–18. <https://doi.org/10.1038/onc.2008.308>. [PubMed]
36. Ogino S, Stampfer M. Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. *J Natl Cancer Inst*. 2010; 102:365–67. <https://doi.org/10.1093/jnci/djq031>. [PubMed]
37. Ogino S, Chan AT, Fuchs CS, Giovannucci E. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut*. 2011; 60:397–411. <https://doi.org/10.1136/gut.2010.217182>. [PubMed]
38. Sisti JS, Collins LC, Beck AH, Tamimi RM, Rosner BA, Eliassen AH. Reproductive risk factors in relation to molecular subtypes of breast cancer: results from the nurses' health studies. *Int J Cancer*. 2016; 138:2346–56. <https://doi.org/10.1002/ijc.29968>. [PubMed]
39. Hirko KA, Willett WC, Hankinson SE, Rosner BA, Beck AH, Tamimi RM, Eliassen AH. Healthy dietary patterns and risk of breast cancer by molecular subtype. *Breast Cancer Res Treat*. 2016; 155:579–88. <https://doi.org/10.1007/s10549-016-3706-2>. [PubMed]
40. Amirian ES, Petrosino JF, Ajami NJ, Liu Y, Mims MP, Scheurer ME. Potential role of gastrointestinal microbiota composition in prostate cancer risk. *Infect Agent Cancer*. 2013; 8:42. <https://doi.org/10.1186/1750-9378-8-42>. [PubMed]
41. Li W, Qiu T, Ling Y, Guo L, Li L, Ying J. Molecular pathological epidemiology of colorectal cancer in Chinese patients with KRAS and BRAF mutations. *Oncotarget*. 2015; 6:39607–13. <https://doi.org/10.18632/oncotarget.5551>. [PubMed]
42. Shinjo K, Okamoto Y, An B, Yokoyama T, Takeuchi I, Fujii M, Osada H, Usami N, Hasegawa Y, Ito H, Hida T, Fujimoto N, Kishimoto T, et al. Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characters of lung adenocarcinoma. *Carcinogenesis*. 2012; 33:1277–85. <https://doi.org/10.1093/carcin/bgs154>. [PubMed]
43. Wang J, Li MD. Common and unique biological pathways associated with smoking initiation/progression, nicotine dependence, and smoking cessation. *Neuropsychopharmacology*. 2010; 35:702–19. <https://doi.org/10.1038/npp.2009.178>. [PubMed]
44. Sheppard BJ, Williams M, Plummer HK, Schuller HM. Activation of voltage-operated Ca²⁺-channels in human small cell lung carcinoma by the tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Int J Oncol*. 2000; 16:513–18. <https://doi.org/10.3892/ijo.16.3.513>. [PubMed]
45. Kimura R, Ushiyama N, Fujii T, Kawashima K. Nicotine-induced Ca²⁺ signaling and down-regulation of nicotinic acetylcholine receptor subunit expression in the CEM human leukemic T-cell line. *Life Sci*. 2003; 72:2155–58. [https://doi.org/10.1016/S0024-3205\(03\)00077-8](https://doi.org/10.1016/S0024-3205(03)00077-8). [PubMed]
46. Schüller HM. Nitrosamine-induced lung carcinogenesis and Ca²⁺/calmodulin antagonists. *Cancer Res*. 1992; 52:2723s–26s. [PubMed]
47. Schaal C, Padmanabhan J, Chellappan S. The Role of nAChR and Calcium Signaling in Pancreatic Cancer Initiation and Progression. *Cancers (Basel)*. 2015; 7:1447–71. [PubMed]
48. Liu X, Wang J, Sun G. Identification of key genes and pathways in renal cell carcinoma through expression profiling data. *Kidney Blood Press Res*. 2015; 40:288–97. <https://doi.org/10.1159/000368504>. [PubMed]
49. Fang ZQ, Zang WD, Chen R, Ye BW, Wang XW, Yi SH, Chen W, He F, Ye G. Gene expression profile and enrichment pathways in different stages of bladder cancer. *Genet Mol Res*. 2013; 12:1479–89. <https://doi.org/10.4238/2013.May.6.1>. [PubMed]
50. Wu X, Zang W, Cui S, Wang M. Bioinformatics analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *Eur Rev Med Pharmacol Sci*. 2012; 16:1582–87. [PubMed]
51. Zhang G, Bi M, Li S, Wang Q, Teng D. Determination of core pathways for oral squamous cell carcinoma via the method of attract. *J Cancer Res Ther*. 2018; 14:S1029–34. <https://doi.org/10.4103/0973-1482.206868>. [PubMed]
52. Klos M, Mundada L, Banerjee I, Morgenstern S, Myers S, Leone M, Kleid M, Herron T, Devaney E. Altered myocyte contractility and calcium homeostasis in alpha-myosin heavy chain point mutations linked to familial dilated cardiomyopathy. *Arch Biochem Biophys*. 2017; 615:53–60. <https://doi.org/10.1016/j.abb.2016.12.007>. [PubMed]
53. Schramek D, Sendoel A, Segal JP, Beronja S, Heller E, Oristian D, Reva B, Fuchs E. Direct *in vivo* RNAi screen unveils myosin IIa as a tumor suppressor of squamous

- cell carcinomas. *Science*. 2014; 343:309–13. <https://doi.org/10.1126/science.1248627>. [PubMed]
54. Onken MD, Winkler AE, Kanchi KL, Chalivendra V, Law JH, Rickert CG, Kallogjeri D, Judd NP, Dunn GP, Piccirillo JF, Lewis JS Jr, Mardis ER, Uppaluri R. A surprising cross-species conservation in the genomic landscape of mouse and human oral cancer identifies a transcriptional signature predicting metastatic disease. *Clin Cancer Res*. 2014; 20:2873–84. <https://doi.org/10.1158/1078-0432.CCR-14-0205>. [PubMed]
 55. Kuan CT, Wikstrand CJ, Bigner DD. EGF mutant receptor vIII as a molecular target in cancer therapy. *Endocr Relat Cancer*. 2001; 8:83–96. <https://doi.org/10.1677/erc.0.0080083>. [PubMed]
 56. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Peadarallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, Imielinski M, Hu X, Ling S, et al, and Cancer Genome Atlas Research Network. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet*. 2016; 48:607–16. <https://doi.org/10.1038/ng.3564>. [PubMed]
 57. Ansell A, Jedlinski A, Johansson AC, Roberg K. Epidermal growth factor is a potential biomarker for poor cetuximab response in tongue cancer cells. *J Oral Pathol Med*. 2016; 45:9–16. <https://doi.org/10.1111/jop.12310>. [PubMed]
 58. Klijn JG, Look MP, Portengen H, Alexieva-Figusch J, van Putten WL, Foekens JA. The prognostic value of epidermal growth factor receptor (EGF-R) in primary breast cancer: results of a 10 year follow-up study. *Breast Cancer Res Treat*. 1994; 29:73–83. <https://doi.org/10.1007/BF00666183>. [PubMed]
 59. Fischer-Colbrie J, Witt A, Heinzl H, Speiser P, Czerwenka K, Sevelde P, Zeillinger R. EGFR and steroid receptors in ovarian carcinoma: comparison with prognostic parameters and outcome of patients. *Anticancer Res*. 1997; 17:613–19. [PubMed]
 60. Magné N, Pivot X, Bensadoun RJ, Guardiola E, Poissonnet G, Dassonville O, Francoual M, Formento JL, Demard F, Schneider M, Milano G. The relationship of epidermal growth factor receptor levels to the prognosis of unresectable pharyngeal cancer patients treated by chemoradiotherapy. *Eur J Cancer*. 2001; 37:2169–77. [https://doi.org/10.1016/S0959-8049\(01\)00280-5](https://doi.org/10.1016/S0959-8049(01)00280-5). [PubMed]
 61. Sartor CI. Biological modifiers as potential radiosensitizers: targeting the epidermal growth factor receptor family. *Semin Oncol*. 2000; 27:15–20. [PubMed]
 62. Newby JC, Johnston SR, Smith IE, Dowsett M. Expression of epidermal growth factor receptor and c-erbB2 during the development of tamoxifen resistance in human breast cancer. *Clin Cancer Res*. 1997; 3:1643–51. [PubMed]
 63. Chen X, Yeung TK, Wang Z. Enhanced drug resistance in cells coexpressing ErbB2 with EGF receptor or ErbB3. *Biochem Biophys Res Commun*. 2000; 277:757–63. <https://doi.org/10.1006/bbrc.2000.3731>. [PubMed]
 64. Marsh D, Suchak K, Moutasim KA, Vallath S, Hopper C, Jerjes W, Upile T, Kalavrezos N, Violette SM, Weinreb PH, Chester KA, Chana JS, Marshall JF, et al. Stromal features are predictive of disease mortality in oral cancer patients. *J Pathol*. 2011; 223:470–81. <https://doi.org/10.1002/path.2830>. [PubMed]
 65. Sobral LM, Bufalino A, Lopes MA, Graner E, Salo T, Coletta RD. Myofibroblasts in the stroma of oral cancer promote tumorigenesis via secretion of activin A. *Oral Oncol*. 2011; 47:840–46. <https://doi.org/10.1016/j.oraloncology.2011.06.011>. [PubMed]
 66. Luksic I, Suton P, Manojlovic S, Virag M, Petroveckii M, Macan D. Significance of myofibroblast appearance in squamous cell carcinoma of the oral cavity on the occurrence of occult regional metastases, distant metastases, and survival. *Int J Oral Maxillofac Surg*. 2015; 44:1075–80. <https://doi.org/10.1016/j.ijom.2015.05.009>. [PubMed]
 67. Yang X, Wu K, Li S, Hu L, Han J, Zhu D, Tian X, Liu W, Tian Z, Zhong L, Yan M, Zhang C, Zhang Z. MFAP5 and TNNC1: potential markers for predicting occult cervical lymphatic metastasis and prognosis in early stage tongue cancer. *Oncotarget*. 2017; 8:2525–35. <https://doi.org/10.18632/oncotarget.12446>. [PubMed]
 68. Zitouni S, Nabais C, Jana SC, Guerrero A, Bettencourt-Dias M. Polo-like kinases: structural variations lead to multiple functions. *Nat Rev Mol Cell Biol*. 2014; 15:433–52. <https://doi.org/10.1038/nrm3819>. [PubMed]
 69. Nogawa M, Yuasa T, Kimura S, Tanaka M, Kuroda J, Sato K, Yokota A, Segawa H, Toda Y, Kageyama S, Yoshiki T, Okada Y, Maekawa T. Intravesical administration of small interfering RNA targeting PLK-1 successfully prevents the growth of bladder cancer. *J Clin Invest*. 2005; 115:978–85. <https://doi.org/10.1172/JCI23043>. [PubMed]
 70. Renner AG, Dos Santos C, Recher C, Bailly C, Créancier L, Kruczynski A, Payrastra B, Manenti S. Polo-like kinase 1 is overexpressed in acute myeloid leukemia and its inhibition preferentially targets the proliferation of leukemic cells. *Blood*. 2009; 114:659–62. <https://doi.org/10.1182/blood-2008-12-195867>. [PubMed]
 71. Strebhardt K, Ullrich A. Targeting polo-like kinase 1 for cancer therapy. *Nat Rev Cancer*. 2006; 6:321–30. <https://doi.org/10.1038/nrc1841>. [PubMed]
 72. Knecht R, Elez R, Oechler M, Solbach C, von Ilberg C, Strebhardt K. Prognostic significance of polo-like kinase (PLK) expression in squamous cell carcinomas of the head and neck. *Cancer Res*. 1999; 59:2794–97. [PubMed]
 73. de Boer DV, Martens-de Kemp SR, Buijze M, Stigter-van Walsum M, Bloemena E, Dietrich R, Leemans CR, van Beusechem VW, Braakhuis BJ, Brakenhoff RH. Targeting PLK1 as a novel chemopreventive approach to eradicate preneoplastic mucosal changes in the head and neck. *Oncotarget*. 2017; 8:97928–40. <https://doi.org/10.18632/oncotarget.17880>. [PubMed]

74. Baba Y, Noshio K, Shima K, Irahara N, Kure S, Toyoda S, Kirkner GJ, Goel A, Fuchs CS, Ogino S. Aurora-A expression is independently associated with chromosomal instability in colorectal cancer. *Neoplasia*. 2009; 11:418–25. <https://doi.org/10.1593/neo.09154>. [PubMed]
75. Kitzen JJ, de Jonge MJ, Verweij J. Aurora kinase inhibitors. *Crit Rev Oncol Hematol*. 2010; 73:99–110. <https://doi.org/10.1016/j.critrevonc.2009.03.009>. [PubMed]
76. Milam MR, Gu J, Yang H, Celestino J, Wu W, Horwitz IB, Lacour RA, Westin SN, Gershenson DM, Wu X, Lu KH. STK15 F31I polymorphism is associated with increased uterine cancer risk: a pilot study. *Gynecol Oncol*. 2007; 107:71–74. <https://doi.org/10.1016/j.ygyno.2007.05.025>. [PubMed]
77. Ogawa E, Takenaka K, Katakura H, Adachi M, Otake Y, Toda Y, Kotani H, Manabe T, Wada H, Tanaka F. Perimembrane Aurora-A expression is a significant prognostic factor in correlation with proliferative activity in non-small-cell lung cancer (NSCLC). *Ann Surg Oncol*. 2008; 15:547–54. <https://doi.org/10.1245/s10434-007-9653-8>. [PubMed]
78. Yang SB, Zhou XB, Zhu HX, Quan LP, Bai JF, He J, Gao YN, Cheng SJ, Xu NZ. Amplification and overexpression of Aurora-A in esophageal squamous cell carcinoma. *Oncol Rep*. 2007; 17:1083–88. <https://doi.org/10.3892/or.17.5.1083>. [PubMed]
79. Wu M, Ye X, Deng X, Wu Y, Li X, Zhang L. Upregulation of metastasis-associated gene 2 promotes cell proliferation and invasion in nasopharyngeal carcinoma. *OncoTargets Ther*. 2016; 9:1647–56. <https://doi.org/10.2147/OTT.S96518>. [PubMed]
80. Zhang H, Chen X, Jin Y, Liu B, Zhou L. Overexpression of Aurora-A promotes laryngeal cancer progression by enhancing invasive ability and chromosomal instability. *Eur Arch Otorhinolaryngol*. 2012; 269:607–14. <https://doi.org/10.1007/s00405-011-1629-4>. [PubMed]
81. Nguyen A, Loo JM, Mital R, Weinberg EM, Man FY, Zeng Z, Paty PB, Saltz L, Janjigian YY, de Stanchina E, Tavazoie SF. PKLR promotes colorectal cancer liver colonization through induction of glutathione synthesis. *J Clin Invest*. 2016; 126:681–94. <https://doi.org/10.1172/JCI83587>. [PubMed]
82. Löser R, Pietzsch J. Cysteine cathepsins: their role in tumor progression and recent trends in the development of imaging probes. *Front Chem*. 2015; 3:37. <https://doi.org/10.3389/fchem.2015.00037>. [PubMed]
83. Berry DA, Cronin KA, Plevritis S, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer E. Effect of Screening and Adjuvant Therapy on Mortality From Breast Cancer. *Obstet Gynecol Surv*. 2006; 61:179–80. <https://doi.org/10.1097/01.ogx.0000201966.23445.91>.
84. Cruz MR, Prestes JC, Gimenes DL, Fanelli MF. Fertility preservation in women with breast cancer undergoing adjuvant chemotherapy: a systematic review. *Fertil Steril*. 2010; 94:138–43. <https://doi.org/10.1016/j.fertnstert.2009.02.055>. [PubMed]
85. Hirai K, Yokoyama M, Asano G, Tanaka S. Expression of cathepsin B and cystatin C in human colorectal cancer. *Hum Pathol*. 1999; 30:680–86. [https://doi.org/10.1016/S0046-8177\(99\)90094-1](https://doi.org/10.1016/S0046-8177(99)90094-1). [PubMed]
86. Saleh Y, Sebzda T, Warwas M, Kopec W, Ziolkowska J, Siewinski M. Expression of cystatin C in clinical human colorectal cancer tissues. *J Exp Ther Oncol*. 2005; 5:49–53. [PubMed]
87. Dai DN, Li Y, Chen B, Du Y, Li SB, Lu SX, Zhao ZP, Zhou AJ, Xue N, Xia TL, Zeng MS, Zhong Q, Wei WD. Elevated expression of CST1 promotes breast cancer progression and predicts a poor prognosis. *J Mol Med (Berl)*. 2017; 95:873–86. <https://doi.org/10.1007/s00109-017-1537-1>. [PubMed]
88. Goto Y, Kurozumi A, Nohata N, Kojima S, Matsushita R, Yoshino H, Yamazaki K, Ishida Y, Ichikawa T, Naya Y, Seki N. The microRNA signature of patients with sunitinib failure: regulation of UHRF1 pathways by microRNA-101 in renal cell carcinoma. *Oncotarget*. 2016; 7:59070–86. <https://doi.org/10.18632/oncotarget.10887>. [PubMed]
89. Koshizuka K, Nohata N, Hanazawa T, Kikkawa N, Arai T, Okato A, Fukumoto I, Katada K, Okamoto Y, Seki N. Deep sequencing-based microRNA expression signatures in head and neck squamous cell carcinoma: dual strands of pre-miR-150 as antitumor miRNAs. *Oncotarget*. 2017; 8:30288–304. <https://doi.org/10.18632/oncotarget.16327>. [PubMed]
90. Itesako T, Seki N, Yoshino H, Chiyomaru T, Yamasaki T, Hidaka H, Yonezawa T, Nohata N, Kinoshita T, Nakagawa M, Enokida H. The microRNA expression signature of bladder cancer by deep sequencing: the functional significance of the miR-195/497 cluster. *PLoS One*. 2014; 9:e84311. <https://doi.org/10.1371/journal.pone.0084311>. [PubMed]
91. Goto Y, Kojima S, Nishikawa R, Kurozumi A, Kato M, Enokida H, Matsushita R, Yamazaki K, Ishida Y, Nakagawa M, Naya Y, Ichikawa T, Seki N. MicroRNA expression signature of castration-resistant prostate cancer: the microRNA-221/222 cluster functions as a tumour suppressor and disease progression marker. *Br J Cancer*. 2015; 113:1055–65. <https://doi.org/10.1038/bjc.2015.300>. [PubMed]
92. Kinoshita T, Nohata N, Hanazawa T, Kikkawa N, Yamamoto N, Yoshino H, Itesako T, Enokida H, Nakagawa M, Okamoto Y, Seki N. Tumour-suppressive microRNA-29s inhibit cancer cell migration and invasion by targeting laminin-integrin signalling in head and neck squamous cell carcinoma. *Br J Cancer*. 2013; 109:2636–45. <https://doi.org/10.1038/bjc.2013.607>. [PubMed]
93. Fukumoto I, Kikkawa N, Matsushita R, Kato M, Kurozumi A, Nishikawa R, Goto Y, Koshizuka K, Hanazawa T, Enokida H, Nakagawa M, Okamoto Y, Seki N. Tumour-suppressive microRNAs (miR-26a/b, miR-29a/b/c and

- miR-218) concertedly suppressed metastasis-promoting LOXL2 in head and neck squamous cell carcinoma. *J Hum Genet.* 2016; 61:109–18. <https://doi.org/10.1038/jhg.2015.120>. [PubMed]
94. Gao J, Liu QG. The role of miR-26 in tumors and normal tissues (Review). *Oncol Lett.* 2011; 2:1019–23. <https://doi.org/10.3892/ol.2011.413>. [PubMed]
 95. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009; 10:48. <https://doi.org/10.1186/1471-2105-10-48>. [PubMed]
 96. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009; 25:1091–93. <https://doi.org/10.1093/bioinformatics/btp101>. [PubMed]
 97. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017; 45:D362–68. <https://doi.org/10.1093/nar/gkw937>. [PubMed]
 98. Xia J, Benner MJ, Hancock RE. NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res.* 2014; 42:W167–74. <https://doi.org/10.1093/nar/gku443>. [PubMed]
 99. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* 2015; 4:e05005. <https://doi.org/10.7554/eLife.05005>. [PubMed]
 100. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2018; 46:D296–302. <https://doi.org/10.1093/nar/gkx1067>. [PubMed]
 101. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016; 44:W90–7. <https://doi.org/10.1093/nar/gkw377>. [PubMed]