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

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 used 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, Log2Foldchange>1) showed 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, Log2Foldchange>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 upregulated and 1240 were downregulated in TOB tumors. In N-TOB tumors, 890 and 729 genes were observed to be unique 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:0007382, 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, TNN1, TNN3, TNN1, TNCC1 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 \((padj(FDR)<0.05, \log2\text{Foldchange}>1)\), 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 (\(\log2\text{Foldchange}>2\)) (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 e.g. 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
regulated and 1240 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 stoma 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.](image)
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

Table 1: miRNA enrichment analysis results

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

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.

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 been observed to be downregulated 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, TNTN3, 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]. Cystatin (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.
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

---

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

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

Matched Tumor/Normal pairs of each HNSCC patient was retrieved from TCGA database.
Additional filtering was performed using advanced statistical options such as a two-sided hypergeometric test to calculate the 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 interactome database [97] with a 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 were identified based on two centrality measures such as degree and betweenness centrality.

### miRNA enrichment analysis of DEGs

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


