Original Research Article

Investigating the Effect of Smoking on Transcriptomics of Primary Oral Cavity Tumors using the Gene Expression Omnibus Dataset

Abstract

Smoking has been suspected to have some relation with the occurrence of oral cancers. However, not much research has been documented with advanced analysis. Utilizing Gene Expression Omnibus (GSE) data, a differential expression (DE) analysis was conducted to investigate the impact of smoking on oral carcinogenesis. The dataset underwent quality control, followed by K-nearest neighbors (KNN) clustering of cells. Cell types were assigned by using the sc-type cell marker library and the annotations provided by the data contributors. Subsequently, DE analysis was performed for each identified cell type. The results of the DE analysis revealed significant upregulation of keratin-related genes, extracellular matrix (ECM) related genes, and immunoglobulin-related genes in smoker tissues. Notably, neutrophils and macrophages exhibited elevated expression of the keratin (KRT) gene family. Moreover, normal epithelial cells displayed increased expression of type 1 collagen (COL1A1). Neutrophils showed heightened expression of trefoil factor 3 (TFF3), which is associated with mucosa secretion. Furthermore, macrophages and naive CD4+ T cells exhibited elevated levels of matrix metalloproteinases (MMPs), enzymes involved in ECM degradation. Interestingly, tumor cells and fibroblasts demonstrated elevated expression of S100A7 (S100 calcium-binding protein A7), an antimicrobial peptide known to impact keratinocyte differentiation. These findings shed light on the complex molecular changes that can potentially lead to the remodeling of the local microenvironment.

Keywords: Differential Expression Analysis, Gene Expression Omnibus Data, Primary Oral Cavity Tumor, Transcriptomics

1. Introduction

Smoking is the culprit of major health risks across multiple diseases [1, 2]. Smoking drastically increases the likelihood of developing various respiratory problems, i.e., chronic bronchitis and emphysema, including chronic obstructive pulmonary disease (COPD) [3]. These complications are

recognized by narrowing and blockage of the airways, leading to difficulty breathing and reduced lung function over time [4]. Moreover, smoking is a primary cause of lung cancer, accounting for the vast majority of cases worldwide [5]. The carcinogenic chemicals in tobacco smoke damage cells in the lungs and other parts of the body, triggering the uncontrolled growth of abnormal cells that form tumors [6]. Beyond lung cancer, smoking is linked to cancers of the throat, mouth, esophagus, bladder, pancreas, and more [7].

Secondhand smoke exposure is also profoundly harmful, affecting non-smokers who inhale tobacco smoke exhaled by smokers or released from burning cigarettes [8]. Non-smokers exposed to secondhand smoke face an increased risk of developing respiratory infections, asthma, and even lung cancer [9]. Furthermore, exposure to secondhand smoke raises the chances of heart disease by up to 30%, primarily due to the toxic chemicals and fine particles in cigarette smoke that can damage blood vessels and affect heart function [10]. These combined health impacts underscore the urgent need for comprehensive tobacco control measures to protect both smokers and non-smokers from the devastating consequences of smoking and secondhand smoke exposure [11].

This research aims to elucidate the transcriptomic differences between smokers and non-smokers in relation to oral cancer. While it's well established that smoking leads to cancer, investigating the specific mechanisms underlying smoking-induced carcinogenesis can bring new insight. A causal connection between smoking and cancers in the lung, larynx, oral cavity, pharynx, esophagus, pancreas, bladder, kidney, cervix, and stomach has been thoroughly reported [12]. Notably, a decrease in the smoking population correlates with a reduction in cancer cases, reinforcing this relationship [13]. Carcinogenic compounds, such as tobacco-specific nitrosamines (TSNAs), polycyclic aromatic hydrocarbons (PAHs), aromatic amines, and aldehydes, are known to play pivotal roles. TSNAs encompass 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and N'-nitrosonornicotine (NNN), both inducing lung tumors across species [14]. PAHs, arising from incomplete combustion from sources like tar, automobile engine exhaust, and furnaces, are notable carcinogens [15]. Aromatic amines, including 2naphthylamine and 4-aminobiphenyl (4-ABP), are linked to bladder cancer [16]. Lastly, aldehydes like formaldehyde and acetaldehyde, prevalent in various environments, are recognized carcinogens [17].

Genetic and epigenetic abnormalities triggered by carcinogens can lead to chronic inflammation [18]. This sustained, deregulated inflammation subsequently disrupts expressions in inflammatory pathways.

Molecules like Interleukin(IL)-1β, prostaglandin (PG)E_2, and transforming growth factors (TGF)-β are integral components of these inflammatory pathways [19, 20]. Significantly, epithelial-mesenchymal transition (EMT) plays a pivotal role in inflammation, fibrosis, and cancer development. In cancerous tissue, EMT regulation falters, consequently activating molecules within the inflammatory pathway [21. 22]. For example, NNK has been identified as an EMT inducer by promoting the downregulation of E-cadherin [23]. This interconnected process underscores the intricate relationship between inflammation, EMT, and carcinogenesis.

Previous literature explored oral cancer carcinogenesis by assessing different cancer stages in a mouse model [24]. This investigation revealed that genes associated with stem cells and keratinocytes, specifically MYC targets v1, exhibited significant enrichment. This outcome was further substantiated by the identification of cisplatin-resistant nasopharyngeal carcinoma cell lines displaying the same pattern. Another study investigated epigenetic changes in oral cancer attributed to alcohol and tobacco exposure [25]. The research uncovered hypermethylation in promoter regions of genes with tumor-suppressive roles. Furthermore, extensive global (genome-wide) hypomethylation was observed, accompanied by alterations in methylation patterns across genes, changes in noncoding RNAs, and modifications of histones. These findings provide comprehensive insights into the multifaceted epigenetic alterations associated with oral cancer development.

2. Experimental Methods

2.1 Data Sources

The single-cell RNAseq data (accession number GSE234933) was obtained from Gene Expression Omnibus, and it was disclosed on July 24th, 2023. As of the writing of this article, the contributors of the data have not yet published a paper citing this dataset and anonymously utilized for this study. However, their publication might be included once it becomes available.

2.2 Metadata

The GSE234933 dataset screening was performed to specifically identify primary tumors of the oral cavity. The screening processes facilitated the patients to be categorized into two distinct groups based

on their smoking status: smokers (HN1, HN60, HN67, HN72, and HN75) and non-smokers (HN7, HN30, HN49, and HN74).

2.3 Data Processing Procedures

The downloaded data was processed through the steps illustrated in Figure 1. Briefly, the data was filtered to select high-quality cells and genes for further processing. Next, the filtered data was clustered using KNN clustering. By employing gene marker libraries (sc-type), the cell types of the clusters were identified. Lastly, the differentially expressed genes were identified for each cell type under the smoking condition.

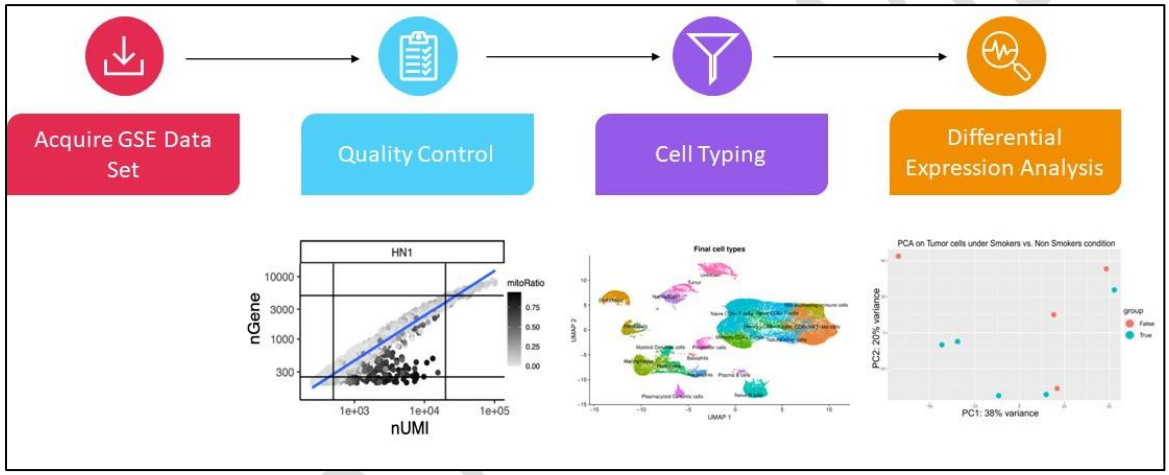

Figure 1: Graphical overview of workflow.

2.4 Data Quality Control

The data contributors provided count matrices in rds file format, which were further processed into Seurat objects using the Seurat R package (v.3.2.2) (Seurat). To enhance data quality and reliability, we applied several filtering steps. Firstly, cells with less than 500 UMIs and 80% biological complexity level (UMIs per gene) were excluded to avoid clustering artifacts caused by cells consisting mainly of themselves. Secondly, cells with a high number of mitochondrial genes were removed as they might indicate cellular damage or stress. Lastly, the cells with fewer than 250 genes and more than 5000 genes were filtered out to ensure consistency in the number of genes per cell. The data quality and the criteria used for filtering are visually depicted in Figure 2. Throughout the quality control process, no potential doublets were removed. Table 1 presents the total number of cells remaining after the quality control steps were applied.

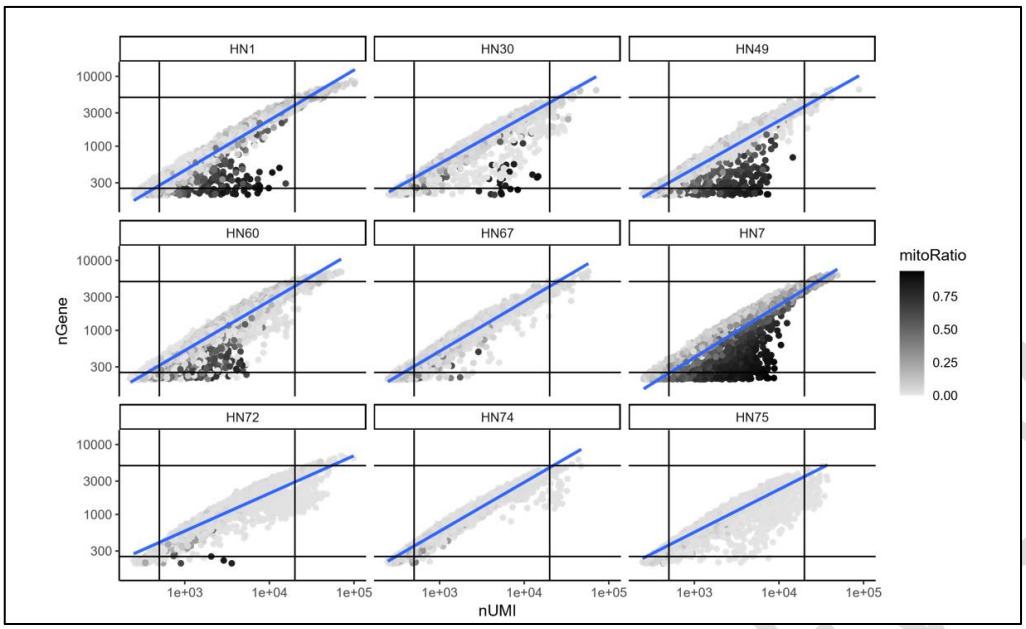

Figure 2: Quality control criteria for each cell visualized. Vertical and horizontal lines are drawn to show the cells that were filtered out. The mitochondrial ratio is depicted with the shade of the data points.

	HN1	HN60	HN67	HN72	HN75	HN7	HN30	HN49	HN74
Smoker	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Num. of cells	2570	7757	4287	3994	9968	1347	5146	3711	11474

Table 1: Number of cells after quality control.

2.5 Normalization Procedures

The count matrices were normalized to values between 0 and 1, following the filtering process. Then, the top 2000 genes with the highest variability were selected, and the ScaleData function was then applied to scale the count matrices using these selected genes. Upon a quick visualization of the data using PCA, it was observed that the mitochondrial ratio did not appear to act as a confounding factor. However, to ensure robustness, the SCTranform function was employed to regress out any potential genetic variability associated with the mitochondrial ratio [26, 27]. The resulting regressed count matrices were then integrated using the FindIntegrationAnchor and Integratedata functions, ensuring data integration across different conditions.

2.6 Cell Type Markers

Forty principal components were computed with the RunPCA function and created an elbow plot to identify the most informative components. Based on this plot, the first 13 principal components were selected for dimension reduction, because the percentage change of variation dropped below 0.1% at the 13th component. Findneighbors and Findcluster functions were employed to determine k-nearest neighbors and clusters. The clustering resolution of 0.6 was used and applied the RunUMAP function to generate a 2-dimensional representation of the data. Next, the markers were identified for each cluster using the FindAllMarkers function. To assign cell types to the clusters, the sc-Type library was utilized, and its auto-detection function suggested that the tissue best matched the immune system library. The data contributors provided cell-type metadata. After inspecting the clustered cells (Figures 3 and 4), it became evident that the sc-Type cell types aligned well with those identified by the contributors. Consequently, we merged both sets of cell type annotations to yield a more detailed cell type classification. Specifically, we retained the cell types originally identified by the contributors, such as fibroblasts, normal epithelial cells, macrophages, and monocytes. However, all other cell types were replaced with sc-type classifications to gain more detailed information. The cancer cells were merged from sc-Type with tumor cells from the contributors to form a larger cluster representing tumor cells. Additionally, we merged non-classical monocytes with macrophages. Thereafter, the erythroid-like and erythroid precursor cells were merged from sc-Types with the "Unknowns" category.

2.7 Differential Expression Analysis

In order to prepare the dataset for differential expression analysis, the count matrices were converted into a single cell experiment object using the SingleCellExperiment package. Next, the counts were aggregated based on cell types and sample IDs. Then, the aggregated counts were used to create a DESeq2 object using DESeqDataSetFromMatrix function from the DESeq2 package [28]. To improve the accuracy of the results, the apeglm algorithm was employed to shrink the logarithmic fold change (LFC) values [29]. Finally, for statistical significance, genes with a p_{adj} value less than 0.05 were selected.

3. Result

3.1 Cell type Identification

Based on the cell type annotation provided by the data contributors, the cell type clusters were identified as shown in Figure 3. Among these clusters, some cells are categorized as "Not Applicable" (NA). The contributors likely removed these cells during their quality control step. However, my quality control process did not remove them.

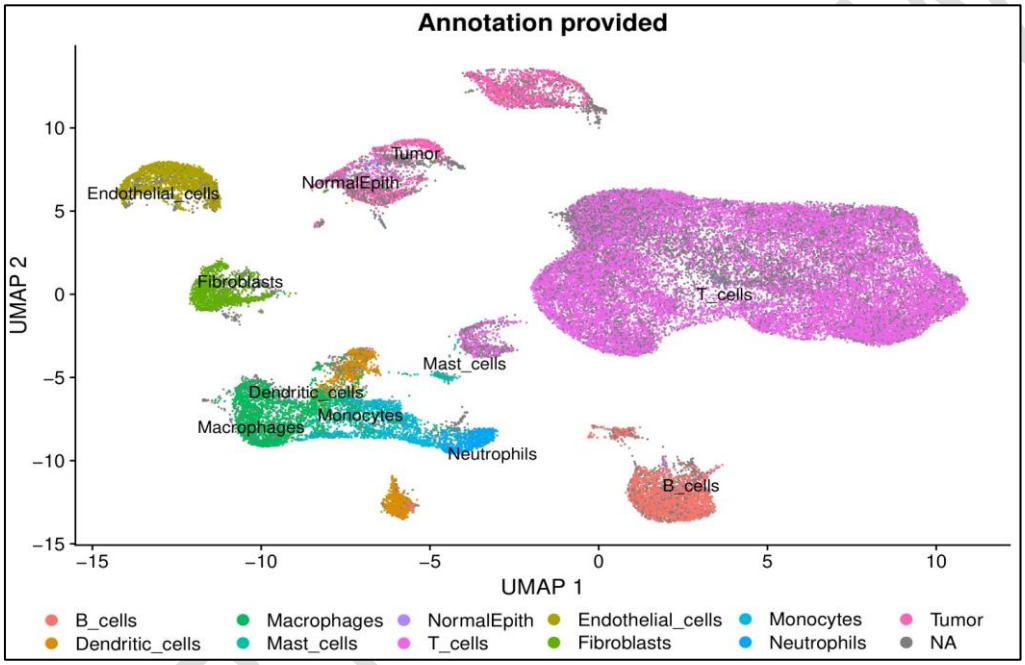

Figure 3: Cell type annotation provided with the raw data mapped to UMAP.

As illustrated in Figure 4, a total of 21 clusters were identified using the sc-Type algorithm. The cell types were similar to those provided by the contributors but with more detailed clustering. Notably, the sc-Type algorithm distinguished between four different T cell types, two B cell types, and two dendritic cell types. It also identified progenitor cells and natural killer cells, which were both merged into T cells by the contributors. However, sc-Type library has higher classification errors for tumor cells. It identified a significant proportion of tumor cells as erythroid cells. It is important to note that the sc-Type library is not specifically designed for tumor cells or support cell (e.g., fibroblast) classification, and this might be the reason for its low reliability in this context.

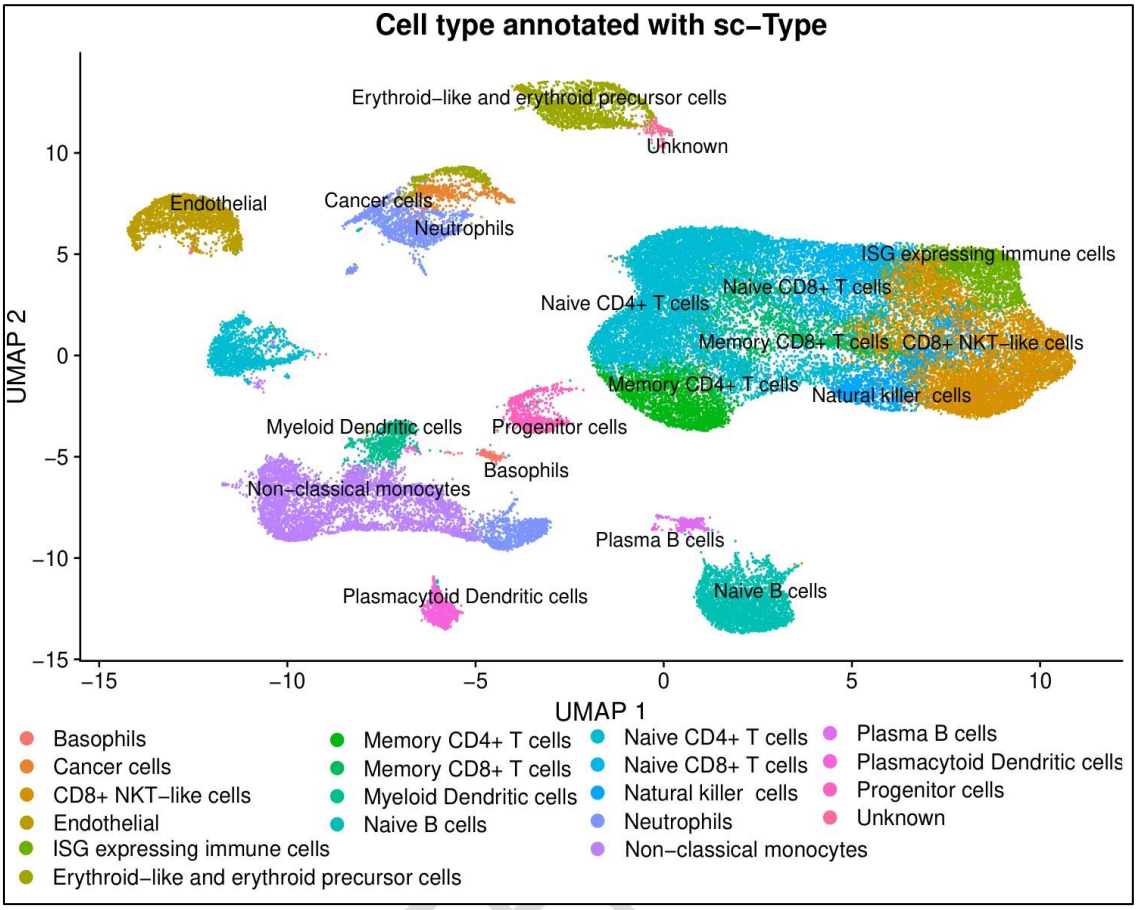

Figure 4: Cell type identified using sc-Type library.

By combining the clustering result from sc-type library and the contributor's annotation, the cell types were finalized as shown in Figure 5.

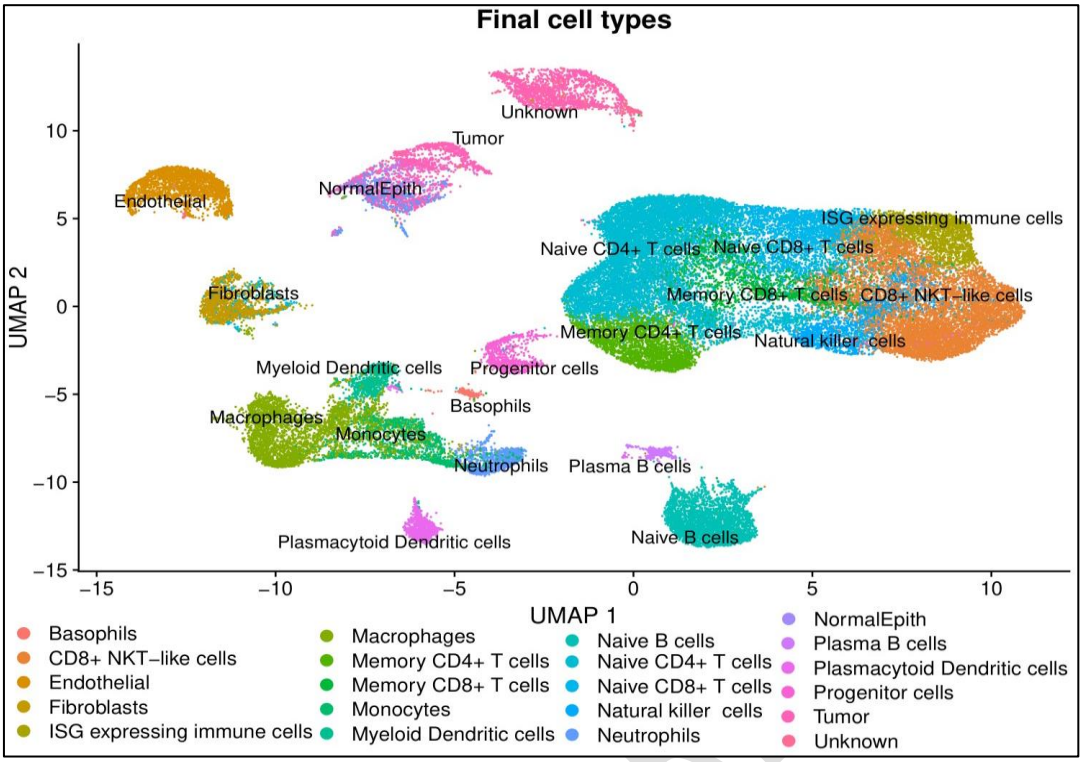

Figure 5: Finalized cell types

3.2 Differential Expression Analysis

3.2.1 Sample-wise PCA

Figure 6 shows how smokers and non-smokers are different from each other. The dimensions of the aggregated count matrices were reduced by principle component analysis (PCA). In addition to tumor cells, all cell types underwent PCA to distinguish smokers from non-smokers. However, no distinctive pattern was discovered in all cell types.

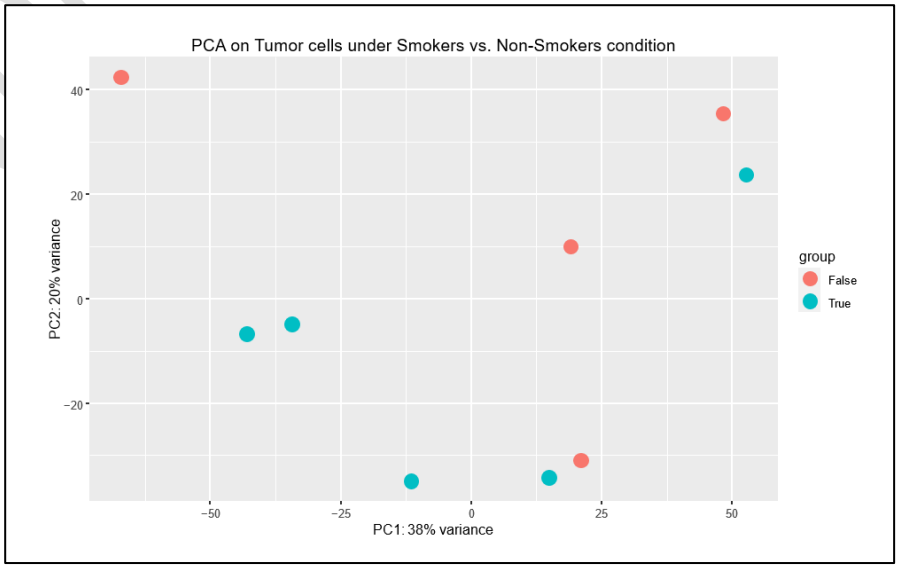

Figure 6: Principal component analysis of data aggregated by sample ID.

3.2.2. Tumor cells

As shown in Figure 7, there are three genes significantly upregulated among smokers: S100A7, SPRR1B and SPRR2A. S100A7 (S100 calcium-binding protein A7) is overexpressed in skin diseases [30]. A study suggests that S100A7 affects keratinocyte differentiation, potentially making the keratinocyte layer tougher. The upregulation of SPRR1B in oral cancer cells is supported by past literature [31].

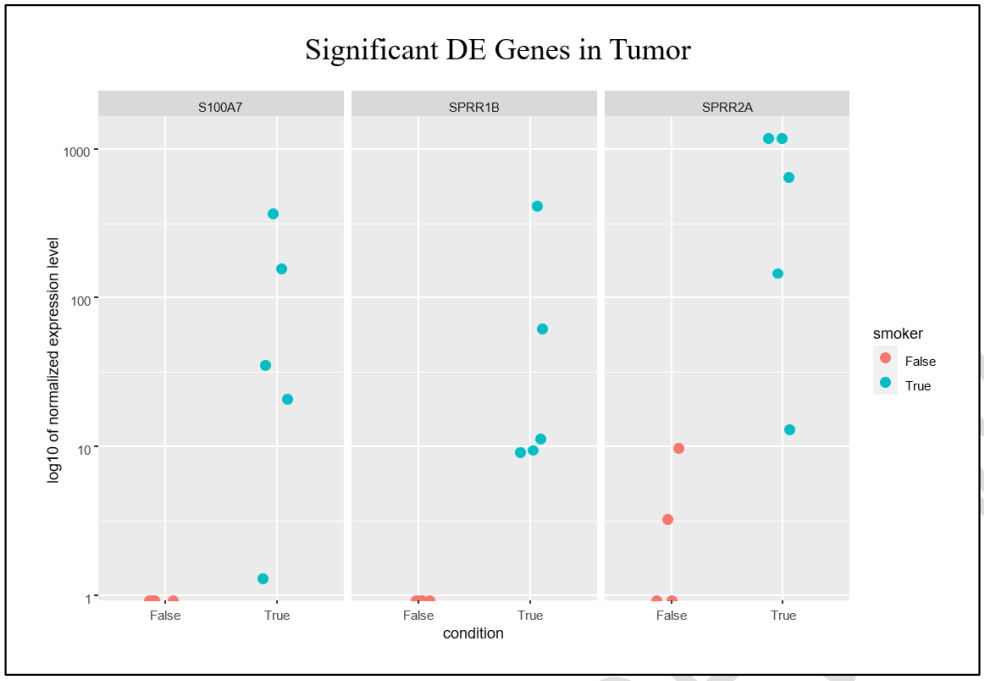

Figure 7: S100A7, SPRR1B and SPRR2A are differentially expressed in tumor.

3.2.3. Progenitor Cells

As presented in Figure 8, there are three genes significantly upregulated among smokers: HPGD, KIT and MS4A2. Downregulation of HPGD is shown to promote cervical cancer proliferation [32]. KIT is a known proto-oncogenic gene [33].

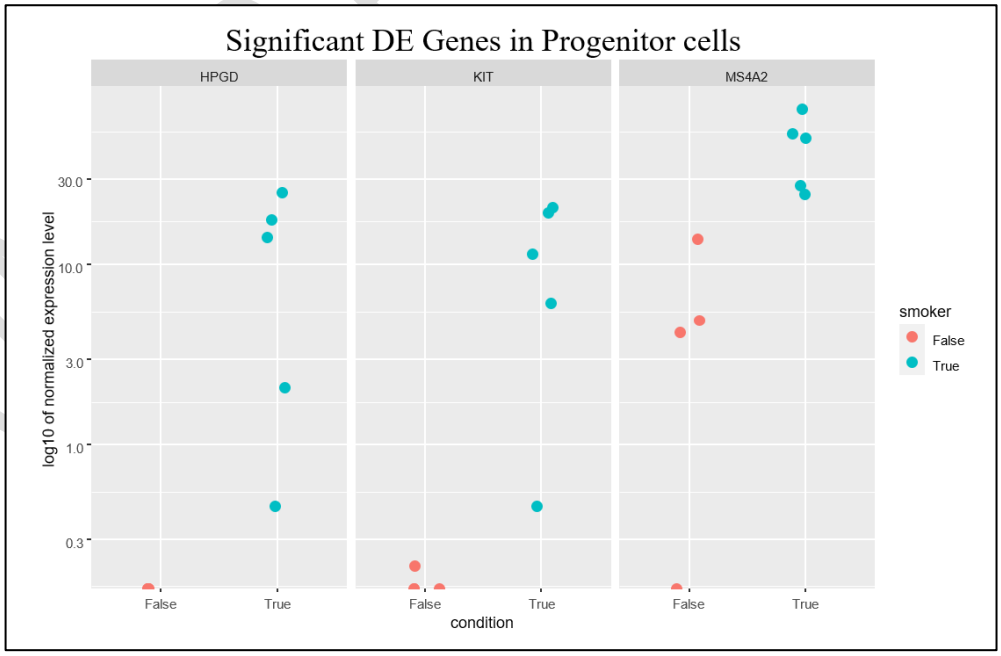

Figure 8: HPGD, KIT and MS4A2 are differentially expressed in progenitor cells.

3.2.4 Plasma B Cells

Fig. 9 presents two genes significantly upregulated among smokers: IGHV3-20 and IGHV7-4-1. Both genes are related to the expression of immunoglobulin heavy chain. The elevated immunoglobulin expression is thought to be elevated in inflammation-related diseases such as multiple sclerosis [34].

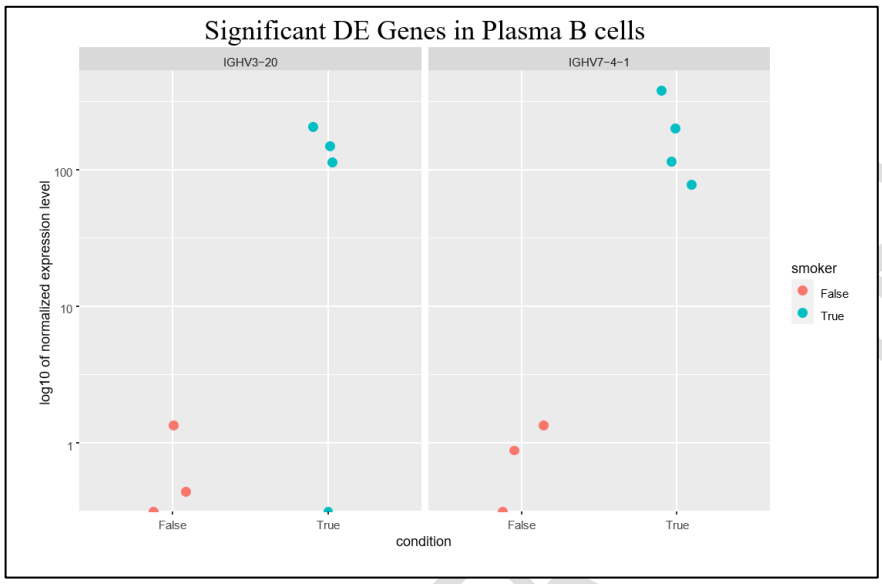

Figure 9: IGHV3-20 and IGHV7-4-1 are differentially expressed in Plasma B cells.

3.2.5. Fibroblasts

As shown in Figure 10, there are three genes significantly upregulated among smokers: IGHV4-39, PI3, and S100A7. It is noteworthy to point out that S100A7 is also over-expressed in tumor cells, and IGHV family genes are over-expressed in plasma B cells.

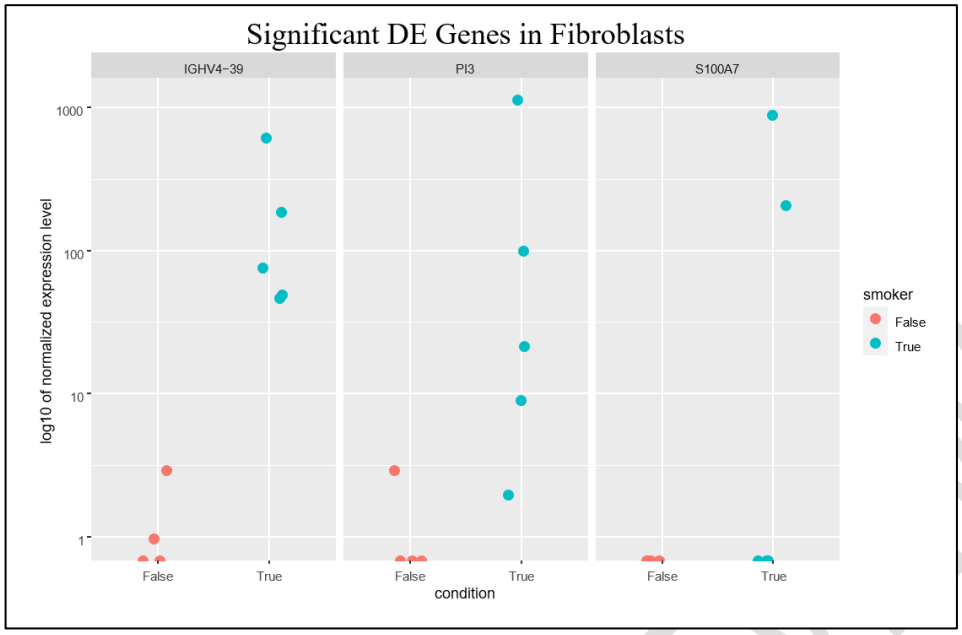

Figure 10: IGHV4-39, PI3, and S100A7 are differentially expressed in fibroblasts.

3.2.6. Normal Epithelial Cells

As plotted in Figure 11, there are two genes significantly upregulated among smokers: COL1A1 and IGKV2-28.

COL1A1 transribes for type 1 collagen while IGKV2-28 is predicted to be part of immunoglobulin complex. A recent paper suggests that elevated COL1A1 expression plays key role in remodeling microenvironment in tumor tissue [35].

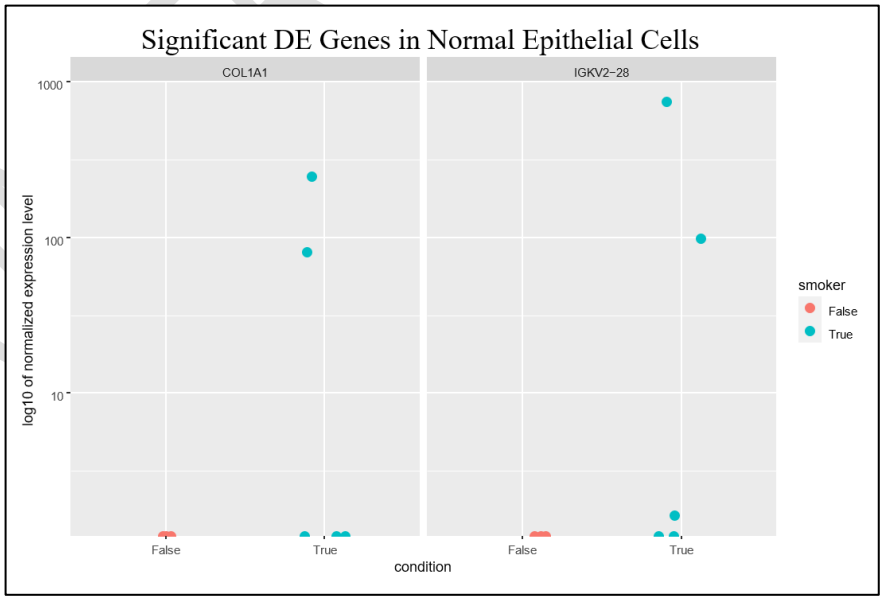

Figure 11: COL1A1 and IGKV2-28 are differentially expressed in normal Epithelial cells.

3.2.7. Neutrophils

As shown in Figure 12, there are two genes significantly upregulated among smokers: KRT4 and TFF3. KRT4 is related to keratin while TFF3 is related to mucosa secretion. A previous literature also found KRT4 as a good prognostic marker for tongue carcinoma [36]. TFF3 is also overexpressed in cervical cancer cells [37].

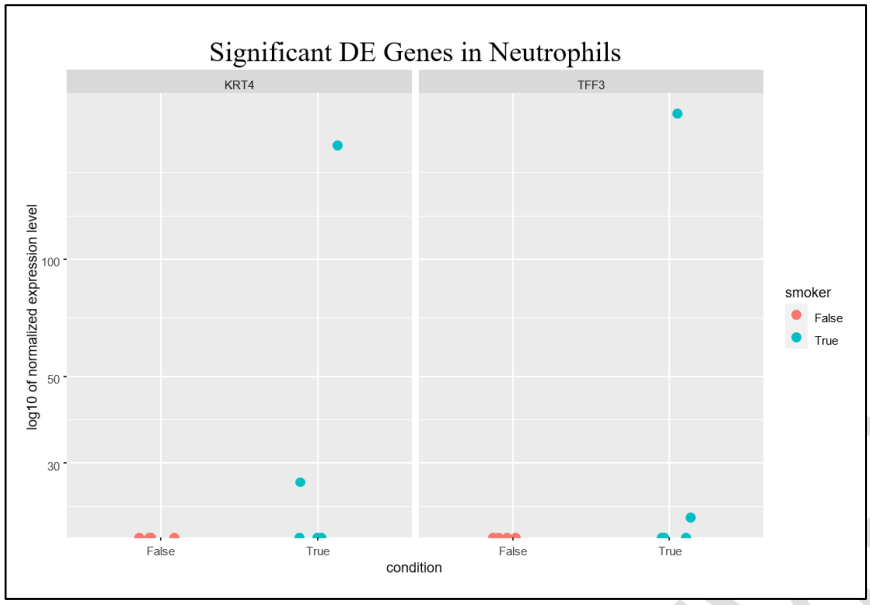

Figure 12: KRT4 and TFF3 are differentially expressed in Neutrophils.

3.2.8. Macrophages

As illustrated in Figure 13, there are two genes significantly upregulated among smokers: IGKV1-33, KRT6A, and MMP10. It is noteworthy to point out that an IGKV family gene is overexpressed in normal epithelial cells, and a keratin-family (KRT) gene is overexpressed in neutrophils. MMP10 is one of matrix metalloproteinases (MMPs).

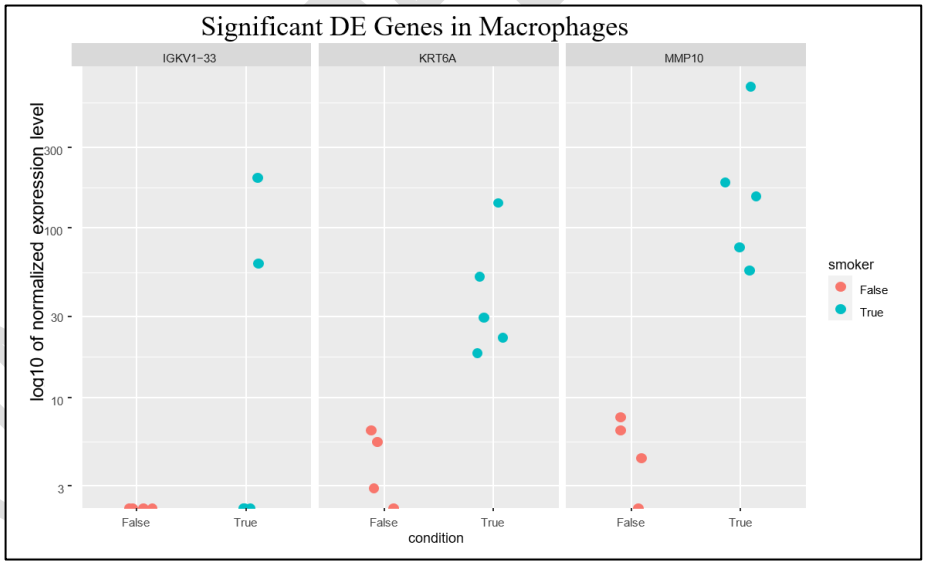

Figure 13: IGKV1-33, KRT6A, and MMP10 are differentially expressed in macrophages.

3.2.9 Monocytes

As shown in Figure 14, there is one gene significantly upregulated among smokers: IGLV3-1. IGLV3-1 is predicted to be part of immunoglobulin that participates in antigen recognition. Elevation of IGLV3-1 expression indicates that monocytes are more activated, considering that their main functions are to recognize antigens and recruit more immune cells.

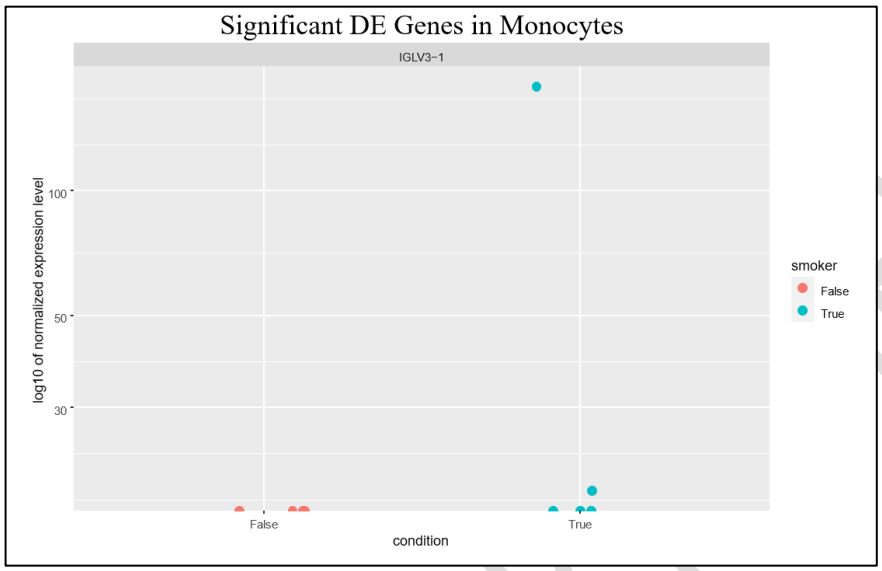

Figure 14: IGLV3-1 is differentially expressed in monocytes.

3.2.10 Naive CD4+ T Cells

Fig. 15 shows the facts that the gene significantly upregulated among smokers: MMP1. MMP1 is one of matrix metalloproteinases (MMPs) similar to MMP10, which is overexpressed in macrophages.

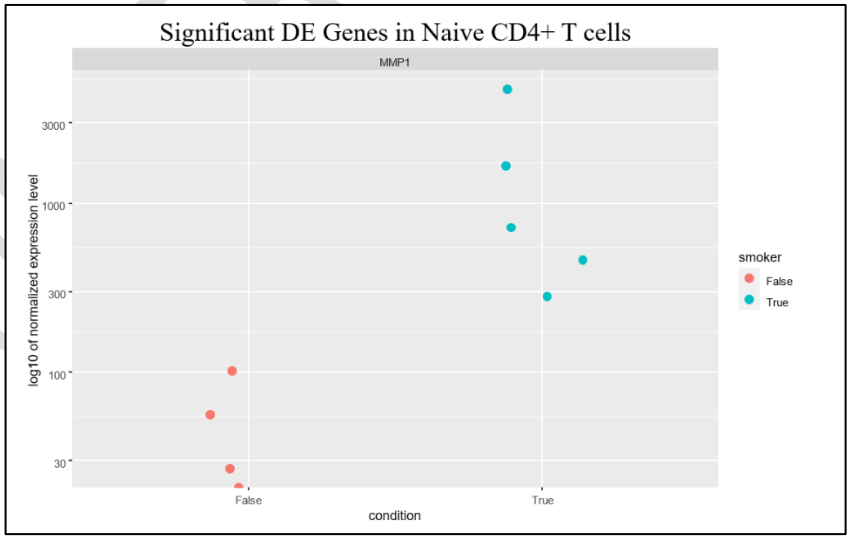

Figure 15: MMP1 is differentially expressed in naive CD4+ T cells.

4. Discussion

4.1 Quality Control on the Data

In Figure 2, samples HN7, HN1, HN49, and HN60 exhibited cells with high mitochondrial gene expression. The high mitochondrial gene expression is often associated with cell death and autolysis. It is possible that these cells were subjected to harsh conditions before the scRNAseq experiment. Although I regressed out mitochondrial gene expression as a confounding factor, compensating for cell loss remains a difficult task. In future studies, I plan to merge data from different experiments to enhance the reliability of the results, creating a larger dataset with a substantial sample size and higher cell counts. This approach aims to counteract the effects of cell loss and ensure more robust conclusions.

4.2 Keratin and ECM-Related Genes

Three keratinocyte-related genes (S100A7, KRT4, KRT6A) were found to be overexpressed in three different cell types: tumor cells, neutrophils, fibroblasts, and macrophages. Additionally, COL1A1, a major component of keratin, showed overexpression in normal epithelial cells. The pathway for destruction of extracellular matrix also seems to be overexpressed, as shown by MMP1 and MMP10 overexpression in macrophages and naive CD4+ T cells. These findings provide evidence that smoking most likely reduces the ECM and replaces it with collagen and keratin. This will have an effect of stiffening the oral tissue. S100A7 also increases the tight junction between cells, making the tissue in general much difficult to penetrate. This can be thought to be the tissue's response to protect itself from outside substance, namely carcinogens from tobacco. The microenvironment of the oral tumor tissue would likely be remodeled to reflect the harsher condition. Further systematic analysis, such as pathway analysis, is essential to gain a deeper understanding of the mechanism of smoking-induced molecular changes. Such analysis will provide valuable insights into the underlying molecular pathways and mechanisms involved in the effects of smoking on carcinogenesis.

4.3 Immunoglobulin-related Genes

In tissue from smokers, there is notable overexpression of six antibody-related genes, including S100A7, which is a gene closely associated with immune responses. The immunoglobulin heavy chain (IGHV)

constitutes a primary component of antibodies. The principal function of plasma B cells revolves around secreting antibodies that target and neutralize antigens. Therefore, the substantial expression of IGHV family genes in plasma B cells is a plausible finding. Interestingly, the presence of immunoglobulin kappa variable cluster (IGKV) expression in fibroblasts and macrophages was unexpected, as these cells are not known for being significant antibody expression. Such expression could potentially stem from noise originating from sequence alignment errors, warranting further investigation.

5. Conclusion

In this paper, the impact of smoking on oral carcinogenesis was assessed through differential expression analysis. The DE analysis showed elevation of keratin-related genes, extracellular matrix (ECM))-related genes, and immunoglobulin-related genes in smoker tissues. The upregulated keratin (KRT, COL1A1, S100A7) and ECM (MMPs) related genes have the capability to remodel the microenvironment of the tissue. The harsher environment of smoker tissue could have triggered the cell's protection pathway to morph the surrounding microenvironment in favor of slowing down penetration of tobacco's carcinogens.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

References

[1] Coa K. I., Augustson E., and Kaufman A., "The impact of weight and weight-related perceptions on smoking status among young adults in a text-messaging cessation program", *Nicotine & Tobacco Research*, 20(5), pp. 614-619, 2018.

[2] Alghamdi A. S., and Zeried F., "The effect of smoking on the ocular surface, tear film and central corneal thickness", *International Journal of Natural Sciences: Current and Future Research Trends*, 15(1), pp. 1-14, 2022.

[3] Wheaton A. G., Croft J. B. VanFrank B., Croxton T. L., Punturieri A., Postow L, and Greenlund K. J., "Chronic obstructive pulmonary disease and smoking status – United States, 2017", *Morbidity and Mortality Weekly Report*, Centers for Disease Control (CDC) USA, 68(24), pp. 533-538, 2019.

[4] Khan K . S., Jawaid S., Memon U. A., Perera T., Khan U., Farwa U. E., Jindal U., Afzal M. S., Razzaq W., Abdin Z. U., Khawaja U. A.. "Management of Chronic Obstructive Pulmonary Disease (COPD) Exacerbations in Hospitalized Patients From Admission to Discharge: A Comprehensive Review of

Therapeutic Interventions." *Cureus*. 15(8):e43694, 2023. doi: 10.7759/cureus.43694. PMID: 37724212; PMCID: PMC10505355.

[5] Hoeng J., Maeder S., Vanscheeuwick P., and Peitsch M. C., "Assessing the lung cancer risk reduction potential of candidate modified risk tobacco products", *Intem Emergency Medicine*, 14(6), pp. 821-834, 2019.

[6] Centers for Disease Control and Prevention (US). "How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta (GA)" *Centers for Disease Control and Prevention (US)*; 2010. 5, Cancer. Available from: https://www.ncbi.nlm.nih.gov/books/NBK53010/

[7] Pichandi, S., Pasupathi, P., Rao, Y.Y., Farook J., Ambika, A., Ponnusha B.S., Sathiyamoorthy, Virumandy, Rajaram and Boopathi, B. "The Effect of Smoking on Cancer-A review." *Int J Biol Med Res.* 2. 593 – 602, 2011.

[8] Napierala M, Merritt TA, Miechowicz I, Mielnik K, Mazela J, Florek E., "The effect of maternal tobacco smoking and second-hand tobacco smoke exposure on human milk oxidant-antioxidant status", Environ Res. 170, pp. 110-121, 2019.

[9] Štěpánek L., Ševčíková J., Horáková D., Patel MS, Durďáková R.. "Public Health Burden of Secondhand Smoking: Case Reports of Lung Cancer and a Literature Review." *Int J Environ Res Public Health*. 19(20):13152, 2022. doi: 10.3390/ijerph192013152. PMID: 36293731; PMCID: PMC9603183.

[10] Gallucci G., Tartarone A., Lerose R., Lalinga A.V., Capobianco A.M. "Cardiovascular risk of smoking and benefits of smoking cessation." *J Thorac Dis.* 12(7):3866-3876, 2020. doi: 10.21037/jtd.2020.02.47. PMID: 32802468; PMCID: PMC7399440.

[11] Bafunno D., Catino A.., Lamorgese V., Del Bene G., Longo V., Montrone M., Pesola F., Pizzutilo P., Cassiano S., Mastrandrea A., Ricci D., Petrillo P., Varesano N., Zacheo A., Galetta D. "Impact of tobacco control interventions on smoking initiation, cessation, and prevalence: a systematic review." *J Thorac Dis.* 12(7):3844-3856, 2020. doi: 10.21037/jtd.2020.02.23. PMID: 32802466; PMCID: PMC7399441.

[12] Centers for Disease Control and Prevention (US), National Center for Chronic Disease Prevention and Health Promotion (US), , and Office on Smoking Health (US). Cancer. In *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*. Centers for Disease Control and Prevention (US), 2010.

[13] Smoking Cessation: A Report of the Surgeon General. 2023

[14] Xia B., Blount B. C., Guilot T., et. Al., "Tobacco-specific nitrosamines (NNAL, NNN, NAT, and NAB) exposures in the US population assessment of tobacco and health (PATH) study wave 1 (2013-2014), *Nicotine Tob Res,* 23(3), pp. 573-583, 2021.

[15] McCreery M. Q., and Balmain A., "Chemical carcinogenesis models of cancer: Back to the future", Annu. Rev. Cancer Biol, 1, pp. 295-312, 2017.

[16] Kanipakam Y., Santhanam V., Rajaram S., Muthanandam S., and Arumugam S.D., "Historical perspectives of chemical carcinogenesis: A Review", SBV *Journal of Basic, Clinical and Applied Health Science*, 4(2), pp. 46-48, 2021.

[17] Becker L. C., Cherian P. A., Bergfeld W. F., et al., "Safety Assessment of Hydrogen Peroxide as Used in Cosmetics.", *International Journal of Toxicology*, 43(3_suppl), pp. 5S-63S, 2024. doi:10.1177/10915818241237790

[18] O'Keeffe L. M., Taylor G., Huxley R. R., Mitchell P., Woodward M., Peters, S. A. E., "Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis", Epidemiology, *BMJ Open*, 8:e021611, 2018. doi:10.1136/bmjopen-2018-021611

[19] Goepp M., Crittenden S., Zhou Y., Rossi A., Narumiya S., Yao C., "Prostaglandin E2 directly inhibits the conversion of inducible regulatory T cells through EP2 and EP4 receptors via antagonizing TGF-b signaling", *Immunology*, 2021, pp. 164:777–791, 2021.

[20] Bydoun, M., Sterea, A., Weaver, I.C.G. *et al.*, "A novel mechanism of plasminogen activation in epithelial and mesenchymal cells.", *Sci Rep* 8, 14091, 2018. https://doi.org/10.1038/s41598-018-32433-y

[21] Agraval, Hina & Chu, Hong, "Lung Organoids in Smoking Research: Current Advances and Future Promises.", *Biomolecules*, 12. 1463, 2022. 10.3390/biom12101463.

[22] Bhat G. R., Sethi I., Sadida H. Q., Rah B., et al., "Cancer cell plasticity: from cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance", *Cancer and Metastasis Reviews*, 43, pp. 197–228, 2024. https://doi.org/10.1007/s10555-024-10172-z

[23] Hu-Nan S., Chen-Xi R., Yi-Xi G., Dan-Ping X., and Taeho K. "Regulatory function of peroxiredoxin I on 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced lung cancer development. *Oncology Letters*, 21(6):465, 2021.

[24] Ling-Yu H., Yi-Ping H., Yen-Yun W., Daw-Yang H., Shih Sheng J., Wen-Tsung H., Wei-Fan C., Ko-Jiunn L., and Tze-Ta H. "Single-Cell Analysis of Different Stages of Oral Cancer Carcinogenesis in a Mouse Model." *International Journal of Molecular Sciences*, 21(21):8171, 2020.

[25] Yasmine G., Juliana L. S., and Mariana B. "Tobacco and alcohol-induced epigenetic changes in oral carcinoma." *Current Opinion in Oncology*, 30(3):152, 2018.

[26] Andrew B., Paul H., Peter S., Efthymia P., and Rahul S. "Integrating single-cell transcriptomic data across different conditions, technologies, and species." *Nature Biotechnology*, 36(5):411– 420, 2018.

[27] Christoph H. and Rahul S. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression." *Genome Biology*, 20(1):296, 2019.

[28] Michael I. L., Wolfgang H., and Simon A. "Moderate estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15(12):550, 2014.

[29] Anqi Z., Joseph G I., and Michael I L. "Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences." *Bioinformatics*, 35(12):2084–2092, 2019.

[30] F. Hattori, C. Kiatsurayanon, K. Okumura, H. Ogawa, S. Ikeda, K. Okamoto, and F. Niyonsaba. "The antimicrobial protein S100A7/psoriasin enhances the expression of keratinocyte differentiation markers and strengthens the skin's tight junction barrier." *British Journal of Dermatology*, 171(4):742–753, 2014.

[31] Yoshitaka M., Yoshihiko H., Toshihiko T., Akihiro M., Jyunki F., Yasuaki T., Tomohide T., Takayuki K., Junichi K., Takanori S., Akari T., Kenji N., Akira Y., Hiroyoshi H., and Noriyuki S. "Small proline-rich protein-1B is overexpressed in human oral squamous cell cancer stem-like cells and is related to their growth through activation of MAP kinase signal." *Biochemical and Biophysical Research Communications*, 439(1):96–102, September 2013.

[32] Shuihong Y., Jingyun X., Kaixuan Z., Pengxia S., Qin Yan, Weifei F., Wan L., and Chun L. "Down-regulation of HPGD by miR-146b-3p promotes cervical cancer cell proliferation, migration, and anchorage-independent growth through activation of STAT3 and AKT pathways." *Cell Death & Disease*, 9(11):1–10, 2018.

[33] Emre K., Niek A. P., Inge U., Veere A. M. van K., Roxanna D., André V., Jamila L., and Onno K. "KIT promotes tumor stroma formation and counteracts tumor-suppressive TGFβ signaling in colorectal cancer." *Cell Death & Disease*, 13(7):1–10, 2022.

[34] Øivind T., Christine S., Solveig M. A., Evert-Jan K., Jeroen J.G., Paul Van Der V., Kjell-Morten M., Vidar M. S., and Lars B. "Upregulation of Immunoglobulin-related Genes in Cortical Sections from Multiple Sclerosis Patients." *Brain Pathology*, 20(4):720–729, 2010.

[35] Mingyue L., Jiaying W., Conghui W., Lili X., Junfen X., Xing X., and Weiguo L. "Microenvironment remodeled by tumor and stromal cells elevates fibroblast-derived COL1A1 and facilitates ovarian cancer metastasis." *Experimental Cell Research*, 394(1):112153, 2020.

[36] Tomohiro E., Satoshi F., Mari T., Youichi H., Shogo N., Tetsuro W., Tomoko Y., Ryuichi H., Atsushi O., and Makoto T. "Gene expression profiling to predict recurrence of advanced squamous cell carcinoma of the tongue: Discovery and external validation." *Oncotarget*, 8(37):61786–61799, 2017.

[37] Zhaohu Y., Dandan C., Xiaojie C., Huikuan Y., and Yaming W. "Overexpression of trefoil factor 3 (TFF3) contributes to the malignant progression in cervical cancer cells." *Cancer Cell International*, 17(1):7, 2017.