Oral Oncology Reports 14 (2025) 100735

Contents lists available at ScienceDirect

Oral Oncology Reports

journal homepage: www.journals.elsevier.com/oral-oncology-reports

Original Research Article

Deciphering oral cancer subtypes: Integrating differential gene expression and pathway analysis followed by non-negative matrix factorization transcription analysis

Anoop Kumar Tiwari [a], Devansh Jain [a], Jayesh Kumar Tiwari [a], Shyam Kishore [c], Akhilesh Kumar Singh [d], Sushant Kumar Shrivastava [a], Arun Khattri [a,b,*]

[a] Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (BHU), Varanasi, 221005, Uttar Pradesh, India
[b] Uttar Pradesh State Institute of Forensic Science (UPSIFS), Lucknow, 226008, Uttar Pradesh, India
[c] Department of Molecular and Human Genetics, Banaras Hindu University, Varanasi, 221005, Uttar Pradesh, India
[d] Faculty of Dental Sciences, Trauma Centre, IMS-BHU, Varanasi, 221005, Uttar Pradesh, India

ARTICLE INFO

Keywords:
Oral cancer
Molecular and genomic classification
Differential gene expression
Pathway enrichment analysis
Subgroup discovery

ABSTRACT

Oral cancer is a major public health concern around the globe, and its classification relies on factors such as habitual status and tumor stages. However, a significant gap exists in understanding oral cancer patients' molecular and genomic characteristics. This study aims to bridge this gap by analyzing International Cancer Genome Consortium (ICGC's) oral cancer data, which identified 2270 differentially expressed genes related to oral cancer. We employed pathway enrichment analysis, highlighting key pathways including hypoxia, VEGF, PI3K, and TGF-β, and STAT2, E2F4, and SP1 transcription factors enriched in tumor samples compared to normal samples. Moreover, we utilized a non-negative matrix factorization (NMF) technique for unsupervised subtype discovery and identified three distinct tumor subgroups. Each subgroup exhibited unique molecular profiles, with pathways related to TNF-α, NF-κB, and hypoxia enriched across all groups. Notably, transcription factor analysis revealed crucial differences: subgroup A was enriched in EGR1, TP53, and HIF1A; subgroup B showed high levels of CDX2 and HNF4A; while subgroup C was characterized by enrichment in ATF4 and E2F4. These findings suggest the feasibility of classifying oral squamous cell carcinoma (OSCC) patients based on gene expression profiles, laying a foundational framework for future research aimed at personalized treatment strategies.

1. Introduction

Cancer is the leading cause of morbidity and mortality worldwide, characterized by the uncontrolled proliferation of aberrant cells disrupting normal cellular processes [1]. It is predicted that by 2027, the mortalities from cancer will rise to 70 %, with oral, lung, breast, prostate, and colorectal cancers accounting for the majority of cases. Oral cancer, a subtype of head and neck cancer (HNSCC), is the sixth most prevalent cancer worldwide [2,3] and poses a severe threat to global healthcare. In underdeveloped and developing nations, the impact of oral cancer is severe and leads to an increased burden on their healthcare system. According to the WHO report of 2022, the incidence and mortalities in males due to oral cancer cases are higher in the Indian subcontinent than worldwide. It is the most common type of cancer in males and the fourth most prevalent in women [4]. It is linked to various risk factors, including genetic predisposition, alcohol, betel nuts, tobacco, cigarettes, human papillomavirus (HPV), and others [5]. While HPV has been implicated as a potential etiological factor in oropharyngeal cancers, its causative or prognostic role in oral cavity cancers remains less well-established and is subject to ongoing research [6–8]. However, HPV has been detected in some oral cancer cases, suggesting the need for further research to understand its impact. Oral Cancer lowers the quality of life by interfering with vital processes like speech, swallowing, and facial appearance. Early detection and comprehensive management are crucial to improving outcomes and mitigating its profound health impacts.

According to India's Head & Neck Cancers study 2020, there were roughly 135,929 new cases, 300,413 prevalent cases, and 75,290 fatalities from oral cancer [8]. In order to improve patient outcomes, the primary cause of cancer-related death needs to be addressed for possible

* Corresponding author. Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (BHU), Varanasi, 221005, Uttar Pradesh, India.
E-mail address: akhattri@upsifs.ac.in (A. Khattri).

https://doi.org/10.1016/j.oor.2025.100735
Received 20 November 2024; Received in revised form 10 March 2025; Accepted 21 March 2025
Available online 24 March 2025

treatment targets. Current therapeutic approaches for oral cancer encompass antibodies (cetuximab, bevacizumab), radiation, and medication therapy (cisplatin, paclitaxel) [9]. Nonetheless, achieving more personalized and efficacious treatment strategies necessitates delineating molecular and genomic attributes.

The classification of cancer based on tumor site, stage, genetics, and biological pathways involved in cancer formation has been accepted by the World Health Organization (WHO) and the Union for International Cancer Control (UICC) [10]. Several studies have successfully sub-categorized HNSCC into molecular or genomic subtypes. For instance, *Walter* et al. provided a 4-group classification of HNSCC [11], and *Keck* et al. suggested a 3-group classification [12]. However, in the context of oral cancer, particularly among south asian patients, there remains a need for a more comprehensive exploration of distinct genetic features. This need persists as the molecular and genomic characteristics of the disease are still understudied, particularly in India, where oral squamous cell carcinoma (OSCC) is a common type of HNSCC.

In this work, we aim to clarify the biological function of the subgroups identified by unsupervised clustering. We employed the unsupervised clustering method NMF (non-negative matrix factorization) to discover subgroups. This study used differential gene expression (DGE) and pathway enrichment analysis to shed light on the underlying biological pathways linked to cancer. Identifying differentially expressed genes across two groups (normal and tumor) will provide insight into transcriptional changes in cancer cells. We have also employed the transcription factor and pathway enrichment analysis to identify significant functional genes and pathways in tumor samples. To understand the molecular and genetic aspects of oral cancer, it is essential to develop tailored therapeutic strategies that maximize treatment and patient outcomes. Our goal is to provide important insights into the molecular classification of OSCC and expand knowledge of its distinct biological features, which will help develop tailored treatment strategies for patients with oral cancer.

2. Material and methods

2.1. Data-collection

This study analysed RNA-seq data of Indian OSCC-GB (gingivabuccal oral squamous cell carcinoma) patients. The patients in this study were participants of the ICGC project in India, which focused on genomic studies of gingivobuccal oral cavity cancer. Detailed clinical data of the patients, including demographic information and environmental exposures, were collected by the Advanced Centre for Research, Treatment, and Education on Cancer, Mumbai. DNA samples were isolated from adjacent normal and tumor tissues from each patient. Sequencing of the collected samples was performed at the National Institute of Biomedical Genomics (NIBMG), Kalyani, West Bengal, India.

2.2. Pre-processing

The transcriptomic data of ICGC OSCC-GB patients was downloaded from the European Genome-Phenome Archive (EGA) portal. The dataset comprises 80 samples, 40 tumors, and 40 normal from the same group of patients. The data for 12 paired samples were available under accession no. EGAD00001003981, while 28 other paired-sample data were available under accession no. EGAD00001004430. The BAM files of all the samples were downloaded and converted to FASTQ format using bamtofastq from the biobambam2 package [13]. Further, the samples were aligned by using STAR aligner with the reference [14] and raw counts were extracted using FeatureCounts [15]. To obtain the HGNC symbol of all genes in the raw counts file, we first matched the ENSEMBL IDs to the HGNC portal. However, we found that only 39,500 symbols were available in the HGNC portal, and some were associated with two or more ENSEMBL IDs. Therefore, we removed duplicate genes and genes with low expression values from the dataset (Fig. 1).

2.3. Differential gene expression (DGE)

To identify differentially expressed genes (DEGs) between 40 pairs of

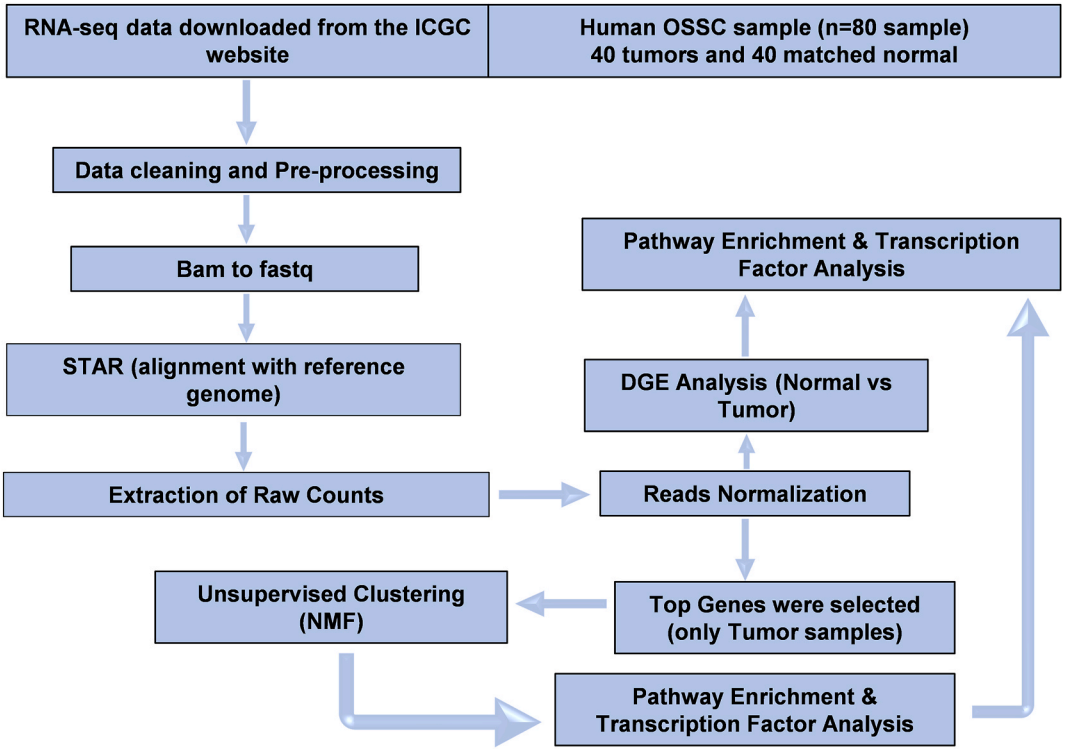

Fig. 1. Data cleaning and processing steps.

tumor and normal samples, we utilized the edgeR package in Rstudio [16], along with FDR <0.05 and logFC >2. The results of the DGE analysis were visualized using the ggplot2 package in R-studio.

2.4. Pathway activity and transcription factor estimation

A transcriptional tutorial user guide was used to gain the pathway enrichment analysis and transcription factor analysis in the RNA-seq data. Footprint-based tools PROGENy & DoRothEA were used to estimate the pathways enrichment and transcription factors, respectively. The obtained result was used to understand the biological processes involved in the cancer progression [17].

2.5. Subtype discovery, pathway enrichment, and transcription factor analysis

In this study, we employed unsupervised clustering through non-negative matrix factorization (NMF) following the method described by Brunet and team [18]. The top 1000 genes from tumor samples were selected in the first step based on their higher interquartile range (IQR) values. Furthermore, different numbers of clusters (k) were evaluated, and it was determined that k = 3 provided the most appropriate clustering solution. Therefore, we selected k = 3 as the optimal choice and performed 100 iterations using the Brunet method, with all other parameters set to their default values in NMF [19]. After that, featured genes from the NMF result were used for the pathway enrichment (PE) and transcription factor (TF) analysis by PROGENy and DoRothEA respectively [17].

3. Results

3.1. Differential gene expression

The DGE analysis revealed that 2270 genes were differentially expressed in the ICGC-OSCC dataset. Of these, most of the genes (1545 genes) were significantly downregulated, while the remaining 725 genes were up-regulated (Fig. 2).

3.2. Pathway enrichment analysis

In this study, PROGENy & DoRothEA were used to conduct PE and TF analyses using the top 2270 differentially expressed genes (DEGs). The top 100 responsive genes per pathway were used for pathway enrichment analysis. The results indicated Hypoxia, Phosphatidylinositol 3-kinase (PI3K), vascular endothelial growth factor (VEGF), Transforming growth factor-beta (TGF-β), Janus kinase-signal transducers and activators of transcription (JAK-STAT), Nuclear factor kappa B (NF-κB), and Tumor necrosis factor-alpha (TNF-α) were the most enriched pathways in tumor samples (Fig. 3A). Furthermore, transcription factor analysis revealed that signal transducer and activator of transcription 2 *(STAT2)*, interferon regulatory factor 9 *(IRF9)*, E2F transcription factor 4 *(E2F4)*, interferon regulatory factor 3 *(IRF3)*, specificity protein 1 *(SP1)*, and CCAAT enhancer binding protein alpha *(CEBPA)* were enriched in most tumor samples. In contrast, Aryl-Hydrocarbon-Receptor-Nuclear-Translocator-Like *(ARNTL)*, Myeloid Ecotropic Insertion Site 2 *(MEIS2)*, One Cut Homeobox 1 *((ONECUT1),* also known as hepatocyte nuclear factor 6 *(HNF6))*, and Homeobox B 13 *(HOXB13)* were enriched in normal samples (Fig. 3B). During pathway enrichment analysis, the normalized enrichment scores of the pathways were computed using Progeny. The results showed that JAK-STAT and PI3K had the most normalized enriched scores, consistent with Dorothea's transcription factor analysis (Fig. 3C).

3.3. Subtype discovery, pathway enrichment, and transcription factor analysis

In this study, we used the Non-Negative Matrix Factorization (NMF) technique to categorise the OSCC patient cohort into three groups: A (15 samples), B (7 samples), and C (18 samples) (Fig. 4A). Most featured genes (Fig. 4B) of these subgroups were used for the PE analysis and TF analysis. TNF-α, NF-κB, and Hypoxia were highly enriched pathways in all groups (Fig. 5A). However, it is worth mentioning that the observed transcription factors were different in all three groups. In group A, with 97 genes, Early growth response 1 (*EGR1*), Tumor Protein 53 (*TP53*), Hypoxia-inducible factor 1 subunit alpha (*HIF1A*), Signal Transducer and Activator of Transcription (*STAT3*), Suppressor of Mothers Against Decapentaplegic 3 and 4 (*SMAD3/4*), and Specificity Protein 1 (*SP1*)

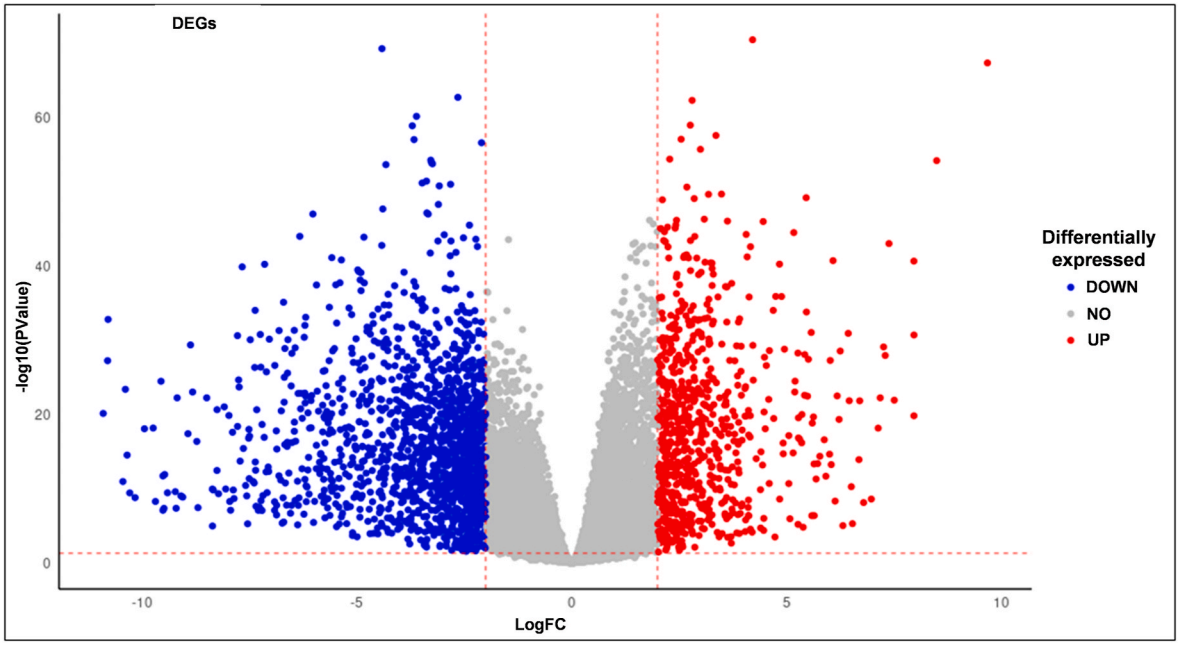

Fig. 2. Differential gene expression analysis. Red dots represent up-regulated genes, while blue dots represent down-regulated genes. All other dots do not meet the criteria for selection of differentially expressed genes.

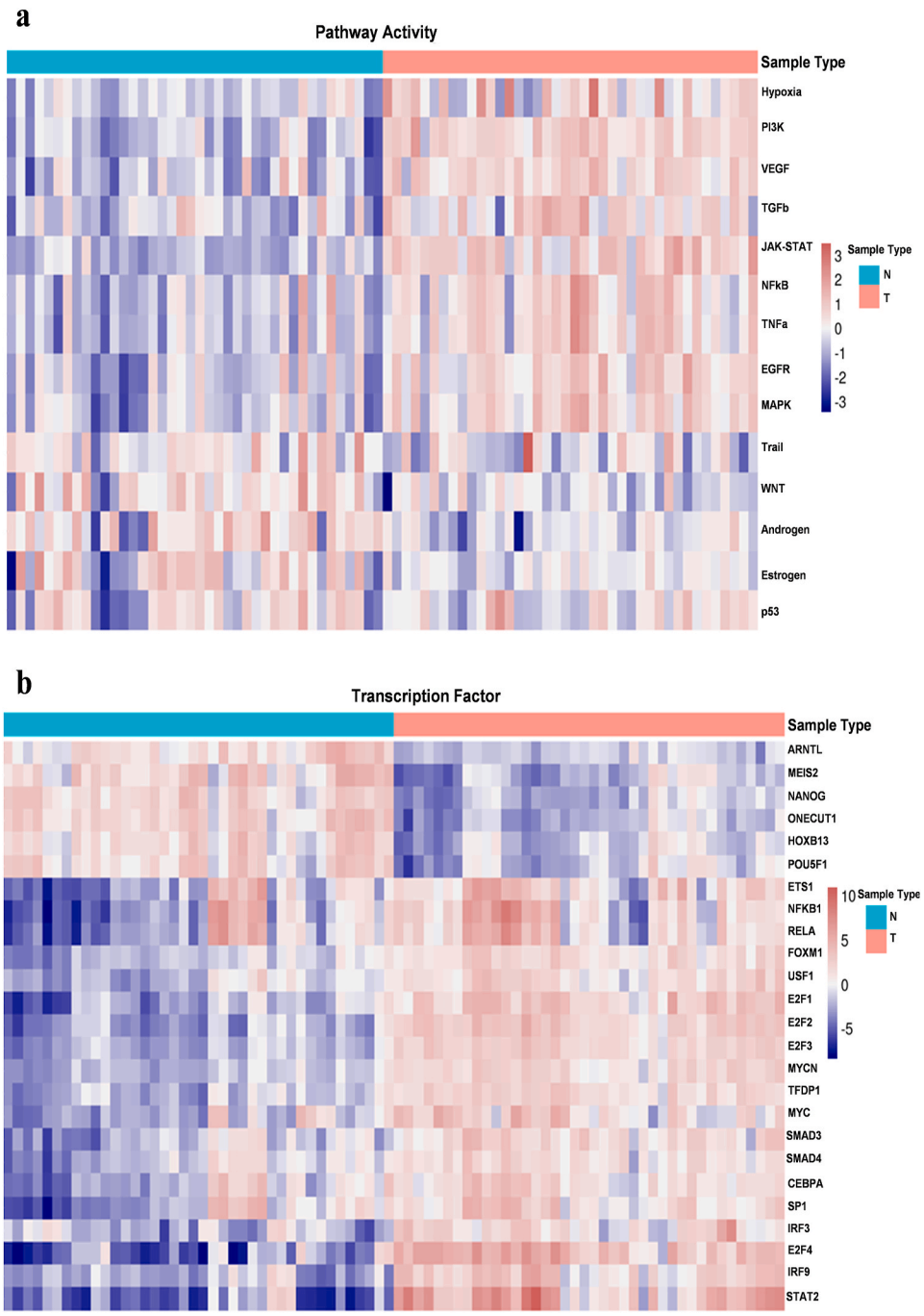

Fig. 3. Comparison of normal and tumor samples (3. a) Pathway enrichment analysis. (3. b) Transcription factors analysis.

were highly upregulated. In group B, with 66 genes, Caudal Type Homeobox 2 (*CDX2*) and Hepatocyte Nuclear Factor 4 alpha (*HNF4A*) were highly enriched. In group C, with 27 genes, Activating Transcription Factor 4 (*ATF4*), Early region 2 binding factor 4 (*E2F4*), and *SPI1* were upregulated (Fig. 5B) (see Supplementary Table 2 for subgroups).

4. Discussion

India has one of the highest incidences of oral cancer in the world, with an incidence rate of one-third of all cases worldwide, causing serious concern for public health [20]. Key risk factors of oral cancer include tobacco, alcohol, betel quid, HPV (mainly HPV16), and genetic predisposition [5,21,22]. Despite these established risk factors, there is a lack of comprehensive studies focusing on the genomic characteristics of oral cancer. This study addresses this gap by analyzing publicly available data comprising 80 samples from 40 patients downloaded from the International Cancer Genome Consortium (ICGC) database. Using DGE analysis, we found 2270 genes differently expressed in tumors than in normal samples (Supplementary Table 1). These differentially expressed genes were further used for PE and TF analysis, which revealed hypoxia, PI3K, VEGF, JAK-STAT, and EGFR pathways are enriched in tumor samples, indicating their potential role in disease progression.

Subsequently, by unsupervised clustering, three subgroups of the tumor samples, A, B, and C, were identified based on their unique gene expression profiles. Further, we elucidated enriched pathways and TFs, revealing that TNF-α, NF-κB, and Hypoxia pathways are enriched in OSCC patients and could be potential targets as these pathways are commonly associated with tumor progression and poor prognosis in

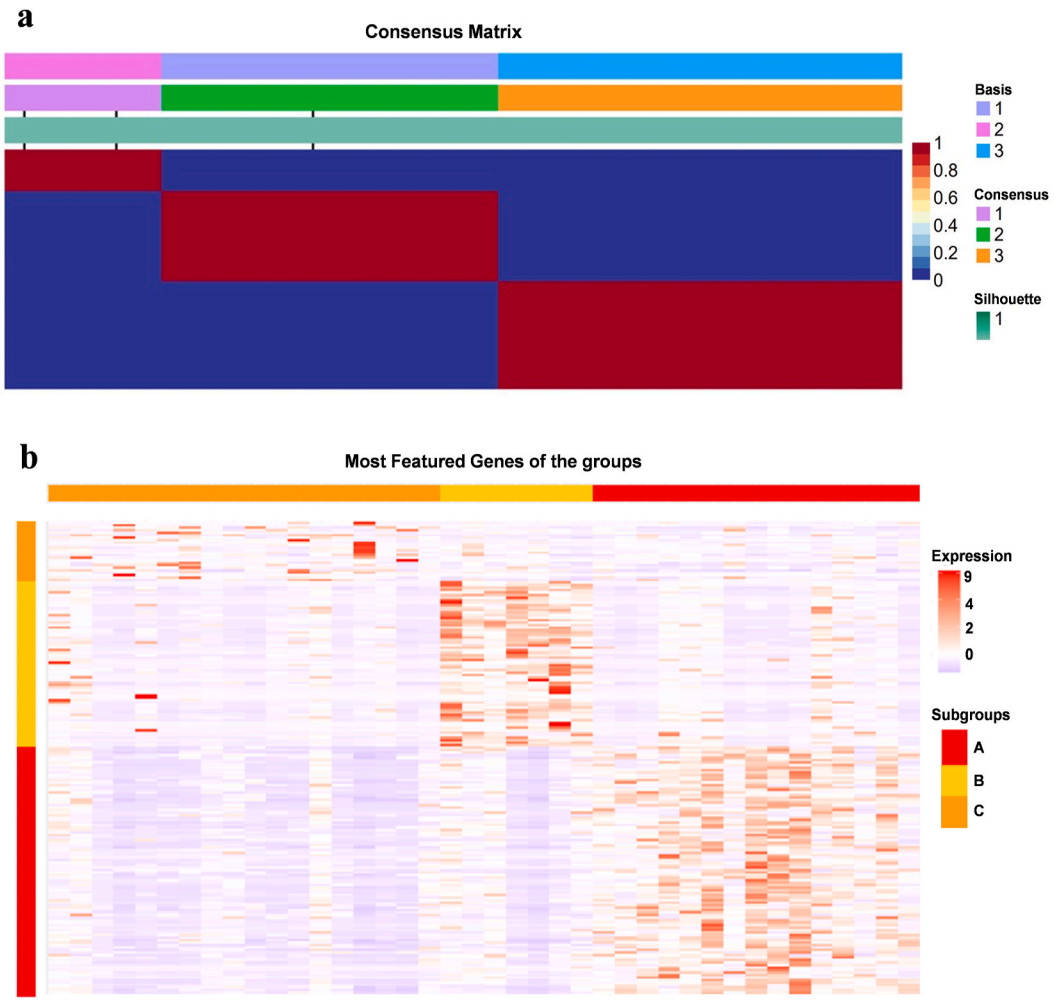

Fig. 4. Subtype discovery (4. a) NMF matrix result. (4. b) Heatmap depicting the most featured genes of each subgroup.

various cancers. For instance, elevated TNF-α and NF-κB pathways, are well known tumor progression and inflammation by enhancing cell survival and tumor growth [22,23]. TNF-α plays a dual role in cancer; initially identified as an anticancer cytokine, it has been observed to promote tumor progression across multiple stages by enhancing cell survival, proliferation, angiogenesis, and metastasis [24]. On the other hand, NF-κB is activated by TNF-α and interleukin-1 (IL-1) and regulates the expression of various inflammatory genes involved in tumorigenesis, linking inflammation to cancer by promoting the expression of genes that support tumor growth, survival, and metastasis. NF-κB activation leads to hypoxia-inducible factor-1 alpha (HIF-1α) upregulation and vascular endothelial growth factor (VEGF), which is crucial for angiogenesis and tumor progression [23]. Moreover, Hypoxia is also considered one of the crucial factors influencing tumor behaviour, treatment responses, and patient prognosis in cancer. It alters signaling pathways that enable cancer cells to survive and thrive in low-oxygen environments and promotes immune tolerance and resistance to therapies [25]. These findings suggest that TNF-α, NF-κB, and hypoxia pathways can potentially be therapeutic targets for oral cancer, as they are also implicated in other cancers.

Further TF analysis identified *STAT2, IRF9, IRF3*, and *E2F4* as enriched transcription factors (TFs) in tumor samples. Among these, E2F4 stands out due to its strong association with poor clinical outcomes. Its overexpression is linked to critical signaling pathways, such as the cell cycle and WNT signaling, both of which play key roles in tumor progression and metastasis [26,27]. Additionally, distinct enriched TFs were identified for each subgroup (A, B, and C), corresponding to unique gene expression profiles. These findings may deepen our understanding of the disease's heterogeneity and uncover potential therapeutic targets.

In subgroup A, several TFs, including *EGR1, TP53, HIFA1*, and *STAT3*, were notably enriched. EGR1 is important in cancer progression, significantly influencing tumor cell proliferation, invasion, and angiogenesis [28]. It is also recognized as a mediator of apoptosis and a promoter of tumor cell growth [29]. Mutations in *TP53*, which are commonly observed in multiple cancers, act as an early driver of tumor development [30]. This disrupts the tumor-suppressing functions of p53 and imparts oncogenic properties, leading to genetic instability and a challenging tumor microenvironment [31,32]. *HIF1A*, another TF enriched in this group, is known to regulate metabolic adaptations, including enhanced glycolysis, by activating glycolytic genes. It promotes tumor cell survival and metastasis by enabling cells to adapt to hypoxic conditions [31,33] Furthermore, *STAT3* was significantly enriched in group A, which is frequently hyperactivating across various malignancies, contributing to the suppression of anti-tumor immunity and the promotion of immunosuppressive factors [34,35]. Moreover, *SMAD3*, *SMAD4* and *SP1* TFs were also enriched in this subgroup, and Previous research has highlighted the pivotal roles of these regulators in cancers, such as breast cancer, colorectal cancer (CRC), and HNSCC [36]. Elevated *SMAD3* expression correlates with significant tumorigenic effects and resistance to certain chemotherapeutics [37]. Mutations in *SMAD3* and *SMAD4* are prevalent, underscoring their involvement in tumorigenesis through the TGF-β signaling pathway [38]. While overexpression of *SP1* is associated with poor prognosis

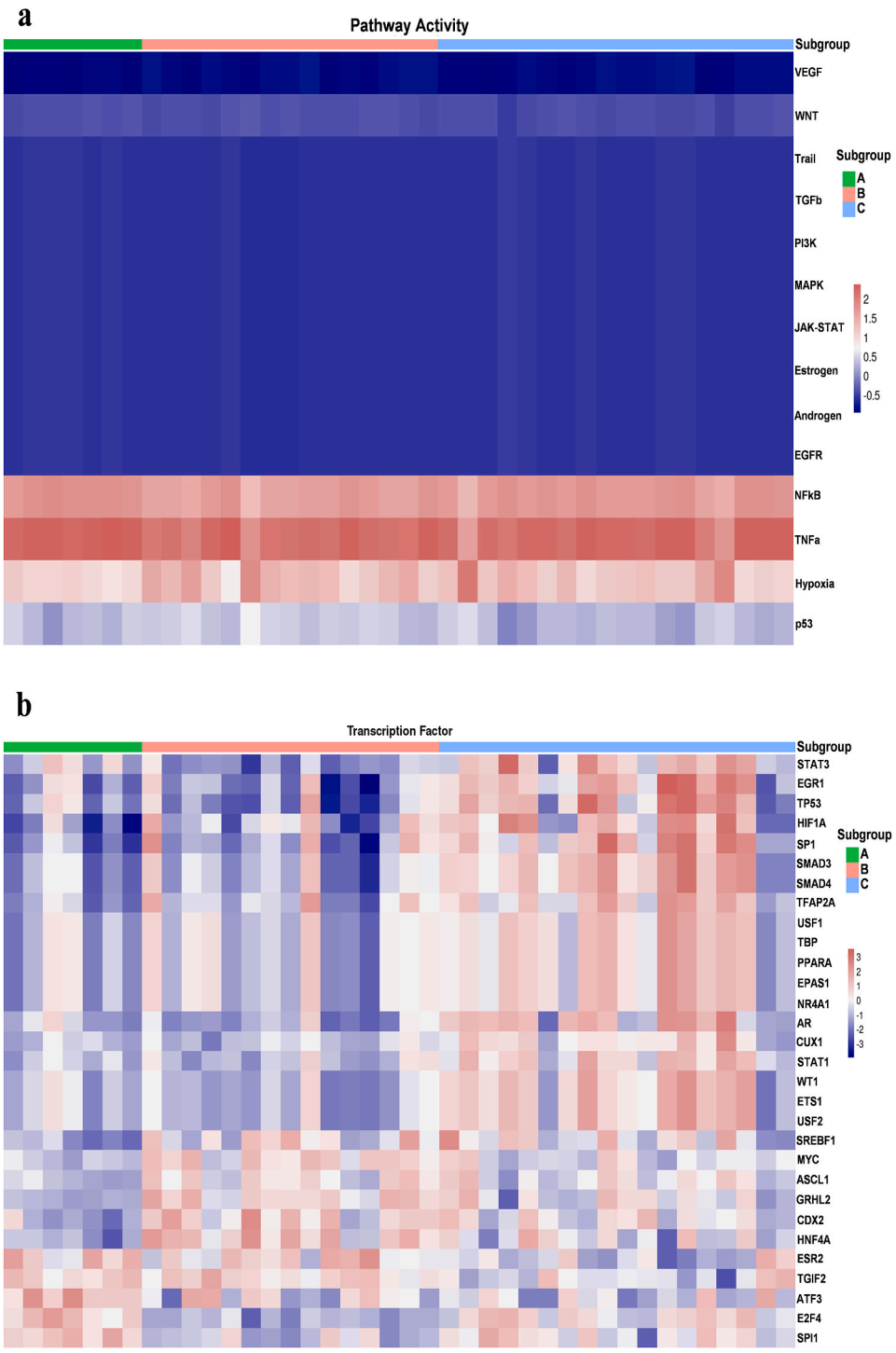

Fig. 5. (5. a) Showcasing the pathway enrichment analysis of each subgroup. (5. b) Identification of transcription factors associated with the subgroups.

across various malignancies, specifically in HNSCC, it has been shown to activate and suppress key oncogenes and tumor suppressors. *SP1* interacts with transcription factors like *HIF-1* and *MYC*, highlighting its collaborative influence on cancer progression [39–42]. The significant enrichment of *EGR1, TP53, HIF1A, STAT3, SMAD3, SMAD4*, and *SP1* in this study suggests that targeting these transcription factors may offer promising therapeutic strategies for treating OSCC in this patient group.

Our study identifies *CDX2* and *HNF4A* as the primary regulators in subgroup B, which have established roles in regulating cancer biology. *CDX2* has been implicated in Wnt/β-catenin signaling, leading to increased cell proliferation, while its overexpression suppresses this pathway by directly activating GSK-3β and Axin2 [43]. Additionally, *CDX2* is a vital biomarker for localised CRC [44]. In the context of HNSCC, *CDX2* enhances the anti-tumor immune responses by promoting CXCL14 expression, stimulating NK-cell migration and cytotoxicity, and suppressing tumor growth [45]. *HNF4α,* another important TF in this subgroup, regulates lipid metabolism, influencing the tumor microenvironment in liver and colorectal cancers [46,47]. Dysregulation of *HNF4α* and *CDX2* plays a crucial role in cancer development, making them potential therapeutic targets in OSCC.

This study also demonstrates significant enrichment of *ATF4*, *E2F*, and *SPI1* in subgroup C, which are involved in crucial processes such as

metabolic adaptation, immune regulation, and cell cycle control. ATF4 has been shown to regulate oxidative stress and metabolic homeostasis in cancer cells, with upregulation linked to increased proliferation and invasion through modulation of the mTORC1 pathway [48–50]. *E2F4* was identified as an enriched factor in our study, consistent with a previous study on HNSCC, which underscores its role in tumor progression and immune regulation, often with poor prognosis. It is associated with key signaling pathways, including the cell cycle and WNT signaling [26,27]. *SPI1* was also identified as a highly enriched factor, similar to its reported roles in gastric cancer (GC) and glioma. Its upregulation is associated with poor prognosis and tumor progression, where it influences the tumor microenvironment through interactions with various immune cell types and is linked to immune activation and cell cycle regulation [51,52]. These findings highlight the roles of *ATF4*, *E2F*, and *SPI1* in cancer biology, warranting further investigation into their potential as therapeutic targets for the treatment of OSCC patients in the given subgroup.

Our study identifies several key pathways and transcription factors, including TNF-α, NF-κB and hypoxia-related pathways, in OSCC, offering a promising therapeutic approach. Given the poor prognosis and resistance to standard therapies in OSCC, targeting the metabolic alterations and signaling pathways could significantly improve patient outcomes.

While this study offers valuable insights into the genomic landscape of OSCC, it is important to consider the relatively small size of 40 patient datasets, which may affect the generalizability of the findings. Further validation with a larger, independent cohort is necessary to confirm these findings' robustness and explore potential ethnic or regional variations in gene expression patterns. Additionally, functional validation of key pathways and transcription factors through in vitro and in vivo models will confirm their precise roles in OSCC progression and therapeutic response.

Summary: Our study identifies key pathways and transcription factors that play significant roles in the progression of OSCC. These include TNF-α, NF-κB, and hypoxia, alongside subgroup-specific transcription factors such as *EGR1, TP53, HIF1A, STAT3, CDX2*, and *HNF4A*. These findings provide a comprehensive framework for understanding genomic heterogeneity in OSCC and highlight potential targets for developing new therapeutic strategies. However, further research, including clinical validation and functional studies, is needed to translate these findings into personalized treatment strategies for OSCC patients.

CRediT authorship contribution statement

Anoop Kumar Tiwari: Writing – original draft, Visualization, Project administration, Investigation, Formal analysis, Data curation. **Devansh Jain:** Formal analysis, Data curation. **Jayesh Kumar Tiwari:** Formal analysis, Data curation. **Shyam Kishore:** Writing – original draft. **Akhilesh Kumar Singh:** Resources, Writing – review & editing. **Sushant Kumar Shrivastava:** Writing – review & editing. **Arun Khattri:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declarations

Approval for data access from ICGC has been obtained under DACO application no. DACO-6289.

Availability of data and materials

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

(https://egaarchive.org/studies/EGAS00001002851).

Twenty-four samples are submitted under the accession number EGAD00001003981, containing 12 samples only, and 56 files contain 28 samples available in the study/accession no. EGAD00001004430.

Ethical approval

The data utilized in this study are sourced from The ICGC oral cancer cohort and were downloaded from the EGA website. Permission for further use and analysis is implied given their availability in the public domain. The ownership of the original data has been acknowledged. This study does not require any animal or ethical approval certificate. Furthermore, it adheres to all relevant guidelines for publication.

Funding

A.K. acknowledges the financial support from the Department of Biotechnology, Government of India, in the form of Ramalingaswami Fellowship. The funding agency played no role in the study design, analysis, and interpretation of results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I sincerely acknowledge the ICGC team and author for providing data access under application no. DACO-6289.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oor.2025.100735.

References

[1] Cooper GM. The cell: a molecular approach. second ed. Sunderland (MA): Sinauer Associates; 2000.
[2] Borse V, Konwar AN, Buragohain P. Oral cancer diagnosis and perspectives in India. Sens Netw 2020;1:100046. https://doi.org/10.1016/j.sintl.2020.100046.
[3] Usman S, Jamal A, Teh M-T, Waseem A. Major molecular signaling pathways in oral cancer associated with therapeutic resistance. Front Oral Health 2021;1. https://doi.org/10.3389/froh.2020.603160.
[4] WHO. International agency for research on cancer. 2022.
[5] Kulkarni MR. Head and neck cancer burden in India. Int J Head Neck Surg 2013;4:29–35. https://doi.org/10.5005/jp-journals-10001-1132.
[6] Vani NV, Rama R, Madhanagopal R, Vijayalakshmi R, Swaminathan R. Human papillomavirus–attributable head and neck cancers in India—a systematic review and meta-analysis. JCO Glob Oncol 2024. https://doi.org/10.1200/GO.23.00464.
[7] Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. Nat Rev Clin Oncol 2022;19:306–27. https://doi.org/10.1038/s41571-022-00603-7.
[8] Foundation for Head and Neck Oncology. Tobacco usage & oral cancer. 2020. https://fhno.org/blog/Tobacco-Usage-And-Oral-Cancer. [Accessed 10 August 2024].
[9] Goel B, Tiwari AK, Pandey RK, Singh AP, Kumar S, Sinha A, et al. Therapeutic approaches for the treatment of head and neck squamous cell carcinoma–An update on clinical trials. Transl Oncol 2022;21:101426. https://doi.org/10.1016/j.tranon.2022.101426.
[10] Carbone A. Cancer classification at the crossroads. Cancers (Basel) 2020;12:980. https://doi.org/10.3390/cancers12040980.
[11] Walter V, Yin X, Wilkerson MD, Cabanski CR, Zhao N, Du Y, et al. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. PLoS One 2013;8:e56823. https://doi.org/10.1371/journal.pone.0056823.
[12] Keck MK, Zuo Z, Khattri A, Stricker TP, Brown CD, Imanguli M, et al. Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes. Clin Cancer Res 2015;21:870–81. https://doi.org/10.1158/1078-0432.CCR-14-2481.
[13] German Tischler. biobambam2. https://gitlab.com/german.tischler/biobambam2. [Accessed 8 June 2022].
[14] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

[15] Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 2019;47. https://doi.org/10.1093/nar/gkz114. e47–e47.
[16] Robinson MD, McCarthy DJ, Smyth GK. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40. https://doi.org/10.1093/bioinformatics/btp616.
[17] Hernansaiz-Ballesteros R, Holland CH, Dugourd A, Saez-Rodriguez J. FUNKI: interactive functional footprint-based analysis of omics data. Bioinformatics 2022; 38:2075–6. https://doi.org/10.1093/bioinformatics/btac055.
[18] Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USA 2004;101:4164–9. https://doi.org/10.1073/pnas.0308531101.
[19] Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinf 2010;11:367. https://doi.org/10.1186/1471-2105-11-367.
[20] Borse V, Konwar AN, Buragohain P. Oral cancer diagnosis and perspectives in India. Sens Netw 2020;1:100046. https://doi.org/10.1016/j.sintl.2020.100046.
[21] Ramshankar V, Soundara VT, Shyamsundar V, Ramani P, Krishnamurthy A. Risk stratification of early stage oral tongue cancers based on HPV status and p16 immunoexpression. Asian Pac J Cancer Prev APJCP 2014;15:8351–9. https://doi.org/10.7314/APJCP.2014.15.19.8351.
[22] Xie B-W, Guan B, Chen W, Zhou M, Gu Q, Liu Y, et al. Tumor-derived extracellular vesicles delivering TNF-α promotes colorectal cancer metastasis via the NF-kB/LAMB3/AKT axis by targeting SNAP23. Arch Biochem Biophys 2023;741:109605. https://doi.org/10.1016/j.abb.2023.109605.
[23] Chrysanthakopoulos Nikolaos Andreas, V E. The role of cytokines, chemokines and NFkB in inflammation and cancer. J Case Rep Med Hist 2023. https://doi.org/10.54289/JCRMH2300114.
[24] Khan A, Zhang Y, Ma N, Shi J, Hou Y. NF-κB role on tumor proliferation, migration, invasion and immune escape. Cancer Gene Ther 2024. https://doi.org/10.1038/s41417-024-00811-6.
[25] Chen Z, Han F, Du Y, Shi H, Zhou W. Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions. Signal Transduct Targeted Ther 2023;8:70. https://doi.org/10.1038/s41392-023-01332-8.
[26] Qi L, Ren Z, Li W. E2F4 transcription factor is a prognostic biomarker related to immune infiltration of head and neck squamous cell carcinoma. Sci Rep 2022;12: 12132. https://doi.org/10.1038/s41598-022-16541-4.
[27] Li Y, Huang Y, Li B, Yang K. Roles of E2F family members in the diagnosis and prognosis of head and neck squamous cell carcinoma. BMC Med Genom 2023;16: 38. https://doi.org/10.1186/s12920-023-01470-6.
[28] Wang B, Guo H, Yu H, Chen Y, Xu H, Zhao G. The role of the transcription factor EGR1 in cancer. Front Oncol 2021;11. https://doi.org/10.3389/fonc.2021.642547.
[29] Gitenay D, Baron VT. Is EGR1 a potential target for prostate cancer therapy? Future Oncol 2009;5:993–1003. https://doi.org/10.2217/fon.09.67.
[30] Silwal-Pandit L, Langerød A, Børresen-Dale A-L. TP53: mutations in breast and ovarian cancer. Cold Spring Harb Perspect Med 2017;7:a026252. https://doi.org/10.1101/cshperspect.a026252.
[31] Chen X, Zhang T, Su W, Dou Z, Zhao D, Jin X, et al. Mutant p53 in cancer: from molecular mechanism to therapeutic modulation. Cell Death Dis 2022;13:974. https://doi.org/10.1038/s41419-022-05408-1.
[32] Wang H, Guo M, Wei H, Chen Y. Targeting p53 pathways: mechanisms, structures and advances in therapy. Signal Transduct Targeted Ther 2023;8:92. https://doi.org/10.1038/s41392-023-01347-1.
[33] Peng K, Zhuo M, Li M, Chen Q, Mo P, Yu C. Histone demethylase JMJD2D activates HIF1 signaling pathway via multiple mechanisms to promote colorectal cancer glycolysis and progression. Oncogene 2020;39:7076–91. https://doi.org/10.1038/s41388-020-01483-w.
[34] Zou S, Tong Q, Liu B, Huang W, Tian Y, Fu X. Targeting STAT3 in cancer immunotherapy. Mol Cancer 2020;19:145. https://doi.org/10.1186/s12943-020-01258-7.
[35] Hu Y, Dong Z, Liu K. Unraveling the complexity of STAT3 in cancer: molecular understanding and drug discovery. J Exp Clin Cancer Res 2024;43:23. https://doi.org/10.1186/s13046-024-02949-5.
[36] Millet C, Zhang YE. Roles of Smad3 in TGF-β signaling during carcinogenesis. Crit Rev Eukaryot Gene Expr 2007;17:281–93. https://doi.org/10.1615/CritRevEukarGeneExpr.v17.i4.30.
[37] Chen Z, Wang Y, Lu X, Chen H, Kong Y, Rong L, et al. The immune regulation and therapeutic potential of the SMAD gene family in breast cancer. Sci Rep 2024;14: 6769. https://doi.org/10.1038/s41598-024-57189-6.
[38] Fleming NI, Jorissen RN, Mouradov D, Christie M, Sakthianandeswaren A, Palmieri M, et al. *SMAD2* , *SMAD3* and *SMAD4* mutations in colorectal cancer. Cancer Res 2013;73:725–35. https://doi.org/10.1158/0008-5472.CAN-12-2706.
[39] Vizcaíno C, Mansilla S, Portugal J. Sp1 transcription factor: a long-standing target in cancer chemotherapy. Pharmacol Ther 2015;152:111–24. https://doi.org/10.1016/j.pharmthera.2015.05.008.
[40] Beishline K, Azizkhan-Clifford J. Sp1 and the 'hallmarks of cancer. FEBS J 2015; 282:224–58. https://doi.org/10.1111/febs.13148.
[41] Kimura K, Jackson TLB, Huang RCC. Interaction and collaboration of SP1, HIF-1, and MYC in regulating the expression of cancer-related genes to further enhance anticancer drug development. Curr Issues Mol Biol 2023;45:9262–83. https://doi.org/10.3390/cimb45110580.
[42] Jumaniyazova E, Aghajanyan A, Kurevlev S, Tskhovrebova L, Makarov A, Gordon K, et al. SP1 gene methylation in head and neck squamous cell cancer in HPV-negative patients. Genes (Basel) 2024;15:281. https://doi.org/10.3390/genes15030281.
[43] Yu J, Liu D, Sun X, Yang K, Yao J, Cheng C, et al. CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/β-catenin signaling via transactivation of GSK-3β and Axin2 expression. Cell Death Dis 2019;10:26. https://doi.org/10.1038/s41419-018-1263-9.
[44] Badia-Ramentol J, Gimeno-Valiente F, Duréndez E, Martínez-Ciarpaglini C, Linares J, Iglesias M, et al. The prognostic potential of CDX2 in colorectal cancer: harmonizing biology and clinical practice. Cancer Treat Rev 2023;121:102643. https://doi.org/10.1016/j.ctrv.2023.102643.
[45] Wang H, Nan S, Wang Y, Xu C. CDX2 enhances natural killer cell–mediated immunotherapy against head and neck squamous cell carcinoma through up-regulating CXCL14. J Cell Mol Med 2021;25:4596–607. https://doi.org/10.1111/jcmm.16253.
[46] Zhao Y, Tang H, Xu J, Sun F, Zhao Y, Li Y. HNF4A-Bridging the gap between intestinal metaplasia and gastric cancer. Evol Bioinf Online 2024;20. https://doi.org/10.1177/11769343241249017.
[47] Qu N, Luan T, Liu N, Kong C, Xu L, Yu H, et al. Hepatocyte nuclear factor 4 a (HNF4α): a perspective in cancer. Biomed Pharmacother 2023;169:115923. https://doi.org/10.1016/j.biopha.2023.115923.
[48] Wang M, Lu Y, Wang H, Wu Y, Xu X, Li Y. High ATF4 expression is associated with poor prognosis, amino acid metabolism, and autophagy in gastric cancer. Front Oncol 2021;11. https://doi.org/10.3389/fonc.2021.740120.
[49] Wu D, Liang J. Activating transcription factor 4: a regulator of stress response in human cancers. Front Cell Dev Biol 2024;12. https://doi.org/10.3389/fcell.2024.1370012.
[50] Wang Y, Ali M, Zhang Q, Sun Q, Ren J, Wang W, et al. ATF4 transcriptionally activates SHH to promote proliferation, invasion, and migration of gastric cancer cells. Cancers (Basel) 2023;15:1429. https://doi.org/10.3390/cancers15051429.
[51] Huang J, Chen W, Jie Z, Jiang M. Comprehensive analysis of immune implications and prognostic value of SPI1 in gastric cancer. Front Oncol 2022;12. https://doi.org/10.3389/fonc.2022.820568.
[52] Du B, Gao W, Qin Y, Zhong J, Zhang Z. Study on the role of transcription factor SPI1 in the development of glioma. Chin Neurosurg J 2022;8:7. https://doi.org/10.1186/s41016-022-00276-2.