*Genetics and Molecular Biology*, 46, 2, e20220358 (2023)

DOI: https://doi.org/10.1590/1678-4685-GMB-2022-0358

Short Communication
Genomics and Bioinformatics

# Epigenomic integrative analysis pinpoint master regulator transcription factors associated with tumorigenesis in squamous cell carcinoma of oral tongue

Larissa Miyuki Okano[1]*, Lívia Maria Maciel da Fonseca[2]*, Isabela Dias Erthal[1] and Tathiane Maistro Malta[2]

*[1]Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, SP, Brazil.*
*[2]Universidade de São Paulo, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Ribeirão Preto, SP, Brazil.*

## Abstract

Head and Neck Cancer (HNC) is a heterogeneous group of cancers, which includes cancers arising in the oral cavity, nasopharynx, oropharynx, hypopharynx, and larynx. Epidemiological studies have revealed that several factors such as tobacco and alcohol use, exposure to environmental pollutants, viral infection, and genetic factors are risk factors for developing HNC. The squamous cell carcinoma of oral tongue (SCCOT), which is significantly more aggressive than the other forms of oral squamous cell carcinoma, presents a propensity for rapid local invasion and spread, and a high recurrence rate. Dysregulation in the epigenetic machinery of cancer cells might help uncover the mechanisms of SCOOT tumorigenesis. Here, we used DNA methylation changes to identify cancer-specific enhancers that were enriched for specific transcription factor binding sites (TFBS), and potential master regulator transcription factors (MRTF) associated with SCCOT. We identified the activation of MRTFs associated with increased invasiveness, metastasis, epithelial-to-mesenchymal transition, poor prognosis, and stemness. On the other hand, we found the downregulation of MRTFs associated with tumor suppression. The identified MRTFs should be further investigated to clarify their role in oral cancer tumorigenesis and for their potential use as biological markers.

*Keywords:* Squamous cell carcinoma of oral tongue, head and neck cancer, master regulator, epigenetic.

Received: November 28, 2022; Accepted: May 04, 2023.

Head and Neck Cancer (HNC) is a heterogeneous group of cancers, which includes cancers arising in the oral cavity, nasopharynx, oropharynx, hypopharynx, and larynx. Epidemiological studies have revealed that several factors such as cigarette (tobacco) use, alcohol consumption, exposure to environmental pollutants, and infection through viral agents such as the human papillomavirus (HPV) and the Epstein-Barr virus are risk factors for developing head and neck cancer. Additionally, genetic factors such as Fanconi anemia and genetic diseases, characterized by impaired DNA repair, increase the risk of first developing cancer in the oral cavity by 500-700 times (Johnson *et al.*, 2020). Oral Squamous Cell Carcinoma (OSCC) is one of the most common head and neck neoplasms, is a complex disease that arises in the oral cavity and oropharynx, is associated with tumor relapse and metastasis following traditional treatment, remaining a major clinical challenge in oral cancer management (Fang *et al.*, 2017; Bugshan and Farooq, 2020). One of the most common is Squamous Cell Carcinoma of the Oral Tongue (SCCOT), which is significantly more aggressive than the other forms of OSCC, with a propensity for rapid local invasion and spread, and a high recurrence rate (Franceschi *et al.*, 1993; Wang *et al.*, 2009).

Send correspondence to Tathiane Maistro Malta. Universidade de São Paulo, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Av. do Café, s/n, Vila Monte Alegre, 14040-900 Ribeirão Preto, SP, Brazil.
E-mail: tathimalta@gmail.com
*These authors contributed equally to this study

Some factors have been identified to increase the susceptibility to the development of head and neck cancer, such as genetic abnormalities that impact cell proliferation, characteristics of cell differentiation, changes in cell cycle, angiogenesis, and tumor metabolism. Furthermore, dysregulation in epigenetic machinery such as DNA methylation, histone modification, and non-coding RNAs have a direct influence on genetic activities (Hier *et al.*, 2021).

Epigenetics is a term used to refer to changes that occur in the genome by altering gene expression without changes in the DNA sequence. One of the mechanisms that can lead to these changes is DNA methylation, which is defined by the addition of a methyl group ($CH_3$) to a cytosine nucleotide and is responsible for modulating important cellular processes such as self-renewal and cellular dedifferentiation (Ehrlich and Lacey, 2013). DNA methylation can change the accessibility of transcription factors to regions containing gene promoters and also to distal regulatory regions such as enhancers, it can also remodel and organize the genome into genomic regions of active and inactive transcription.

This epigenetic alteration can lead to dysregulation of pathways linked to cell differentiation, cell cycle control, apoptosis, angiogenesis, and metastasis; therefore DNA methylation and expression analysis have great potential to provide epigenetic markers for diagnosis, prognosis, evaluation risk assessment, and disease monitoring in various types of cancer (Zavridou *et al.*, 2020). In head and neck tumors, several genes were shown to be epigenetically regulated by DNA methylation and can be studied as therapeutic targets (Hier *et al.*, 2021).

Here, we used DNA methylation changes to identify cancer-specific enhancers that were enriched for specific transcription factor binding sites (TFBS) and potential master regulator transcription factors (MRTF) associated with SCCOT.

The head and neck squamous cell carcinomas (HNSCC) data used were taken from The Cancer Genome Atlas (TCGA), named Genomic Data Commons Data Portal (GDC Data Portal), a reference program in cancer genomics that characterized about 20,000 primary cancers and corresponding normal samples, encompassing 33 different tumor types. We selected 147 HNSC tumor samples from the tongue region and 20 samples from adjacent regions. Of these tumor samples, 74 are HPV-negative, 11 are HPV-positive and 62 patients did not have this clinical data, of samples from adjacent regions 18 are HPV-negative, and 2 are HPV-positive (Table S1). The tongue region was chosen because it is one of the most frequent types of HNSCC, which comprises a heterogeneous group of cancers.

Level 3 gene expression data (RNAseq) and DNA methylation data (Illumina 450K) of TCGA-HNSCC were downloaded using the R/Bioconductor package TCGAbiolinks (Colaprico *et al.*, 2016; Silva *et al.*, 2016). Additionally, probes assessing DNA methylation levels present in distal regions of the genome were identified based on the genomic location provided by GENCODE. A probe is a single-stranded sequence of DNA used to assess the methylation level of a CpG site.

Data visualization and statistical analysis were performed using R software packages (https://www.r-project.org). Samples from the tongue and the adjacent regions were compared using the R/Bioconductor package ELMER (Silva *et al.*, 2019), using the unsupervised mode, to perform both hypomethylation and hypermethylation analyses.

Elmer is an "R" based tool that uses DNA methylation levels to identify enhancer elements in the genome and correlates them with the expression of nearby genes in order to identify transcriptional targets and elucidate epigenetic regulatory mechanisms, it allows the inference of transcription factor (TF) binding motifs and the integration of transcription factor gene expression in order to identify which TFs regulate the biological process explored.

DNA methylation can be used to identify functional changes at transcriptional enhancers and other cis-regulatory modules (CRMs) in tumors and other primary disease tissues. ELMER package reconstructs gene regulatory networks (GRNs) by combining methylation and gene expression data derived from the same set of samples, it uses methylation changes at CRMs as the central hub of these networks, using correlation analysis to associate them with both upstream master regulator (MR) transcription factors and downstream target genes. This analysis process can be briefly described in five steps: Identification of distal enhancer probes; Identification of distal probes that show significantly different DNA methylation levels between the analyzed groups; Identification of potential target genes for differentially methylated probes (probe-gene pairs) - search of nearby genes (10 upstream and 10 downstream) with a negative correlation between DNA methylation and gene expression; Identification of enriched motifs for probes belonging to probe-gene pairs - only the motifs that presented an odds ratio (95% CI) greater than 1.1 and that had a higher incidence in probes were used; and Identification of regulatory TFs whose expression is associated with DNA methylation in the previously identified motifs.

In order not to have a large sampling difference between tumor and non-tumor (adjacent regions), we randomly selected 40% or tumor samples (59 out of 147) from the tongue region and 20 samples of adjacent tissue.

We applied the default probe filter "distal", which uses 158.803 probes that are >2 kbp from any transcription start site as annotated by GENCODE. ELMER was used with the following parameters: get.diff.meth(sig.diff) = 0.3, get.diff.meth(p_value) = 0.001, get.diff.meth (minSubgroupFrac) = 0.4, get.pair(Pe) = 0.001, get.pair(raw.pvalue) = 0.05, get.pair(filter.probes) = FALSE, get.pair(diff.dir)=both, get.pair(permu.size) = 100, get.pair(minSubgroupFrac) = 0.4, get.enriched.motif(lower.OR) = 1.1, get.enriched.motif(min.incidence) = 10, and pathway analysis was performed using Reactome software (Gillespie *et al.*, 2022).

Aberrant DNA methylation is a hallmark of the development of cancer and can affect gene expression, causing dysregulation of important cell mechanisms and leading to tumorigenesis; it could also affect the treatment response and resistance. Most studies focus on the CpG islands regions, once regions of focal hypermethylation in tumors were located primarily at CpG islands (Agrawal *et al.*, 2011; Bergman and Cedar, 2013). However there are studies demonstrating that methylations of distal regulatory sites are closely related to gene expression profiles of transformed cells and bear great relevance to further understanding of the specific regulatory networks, and provide critical information about gene expression control in tumorous and health cells (Aran *et al.*, 2013; Lin *et al.*, 2016; Pan *et al.*, 2020).

To investigate the alterations in the epigenome and its transcriptional implications in SCCOT, we compared DNA methylation data of tumor samples from the tongue to non-tumor samples from adjacent tumors from TCGA and identified 14,659 hypomethylated probes **(**Table S2A). Next, we searched for nearby genes with a negative correlation between DNA methylation and gene expression to identify putative transcriptional targets. We identified 696 probe-gene pairs with significant negative correlation (Figure 1a). A pathway analysis of these genes suggested activation of cell cycle and DNA replication (Figure 1b).

Within the regulatory differentially methylated regions, we identified the enrichment of 69 distal-binding motifs for specific transcription factors. For each enriched motif, the pipeline computes the expression of each gene with the average DNA methylation of all distal probes using probe-gene pairs to identify possible TFs associated with these motifs (Table S2B).

We identified 69 TFs that could potentially bind to these enriched motifs, TFs with TGAGTCA (Fos/Jun family) being the most represented. By computing the anticorrelation of TF gene expression and the DNA methylation levels within their binding sites, we identified HOXC6, SNAI2, SOX12, HMGA2, and POU5F1 candidate MRTFs with increasing expression in SCCOT compared to non-tumor sample (Table 1, Figure 1c).

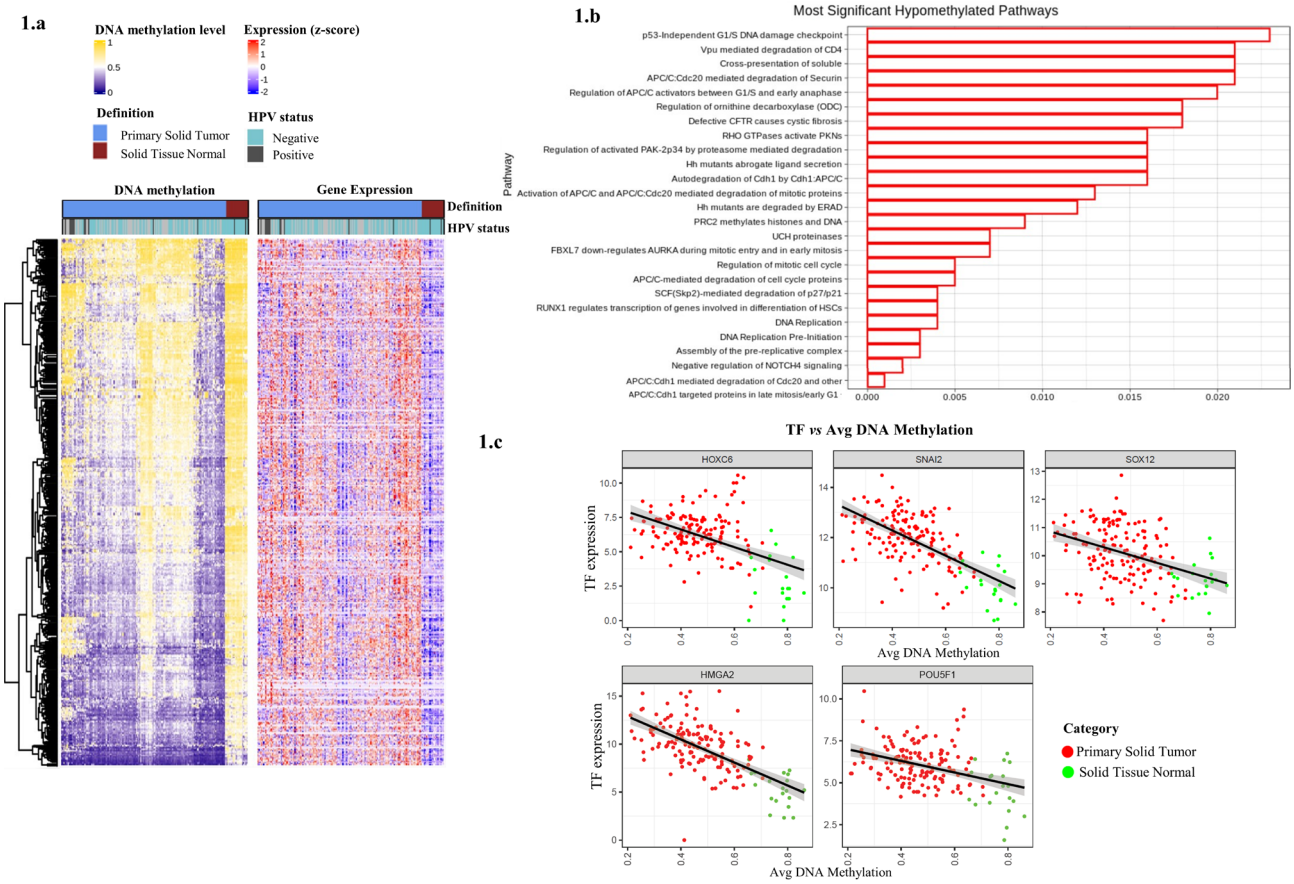


**Figure 1 –** Hypomethylation analysis panel. **a.** Heatmap of hypomethylation analysis of tumor vs. non-tumor samples. **b.** Pathway analysis performed on Reactome showing the most significant hypomethylated pathways. **c.** Scatter plot of expression of the respective TFs (HMGA2, POU5F1, HOXC6, SNAI2, and SOX12) vs DNA methylation.. Source: Author.

**Table 1** – Hypomethylated TF's related to HNSCC: biological function and influence in cancer development.

| TF | Biological Function | Oncogene or tumor suppressive | | Ref |
|---|---|---|---|---|
| | | Cancer | HNSCC | |
| HMGA2 | High cell proliferation, low apoptosis, tumor growth, promote angiogenesis, invasion, drug resistance and metastasis | Oncogene | Oncogene | (Mansoori *et al.*, 2021) |
| HOXC6 | Associated with chemotherapeutic resistance, lymph node metastasis, induce apoptosis and cell growth | Oncogene | Oncogene | (Moon *et al.*, 2012; Zhang *et al.*, 2018) |
| POU5F1 | tumor size, TNM stage, tumor differentiation, tumor invasion depth, lymph node metastasis, distant metastasis, lymphovascular invasion, vascular invasion, tumor number, and tumor recurrence | Oncogene | Oncogene | (He *et al.*, 2020) |
| SNAI2 | lymph node metastasis, reduced overall survival, stem and/or progenitor cell biology, cellular differentiation, vascular remodeling and DNA damage repair | Oncogene | Oncogene | (Wang *et al.*, 2012; Steinbichler *et al.*, 2020) |
| SOX12 | Tumor invasion, tumor differentiation, lymph node metastasis, distant metastasis | Oncogene | Oncogene | (Seok *et al.*, 2021; Su *et al.*, 2022) |

Homeobox C6 (HOXC6) genes belong to the homeoprotein family of transcription factors. The overexpression has been found in several cancer types and OSCC, suggesting that HOX genes are implicated in the development of oral dysplasia and oral SCC and the acquisition of metastatic phenotypes (Zhang *et al.*, 2018).

SNAI2 is a transcription factor formerly known as Slug. It is involved in several biological processes and has been implicated in epithelial-mesenchymal transition (EMT) during embryonic development and tumor progression (Zhou *et al.*, 2019; Li Z *et al.*, 2022).

The SOX family of transcription factors is characterized by the presence of a DNA-binding high mobility group (HMG) domain. When SOX12 is highly expressed in OSCC, it is positively correlated with OSCC pathological grade, T stage, and N stage. In esophageal cancer (ESCC) overexpression

of SOX12 indicated shorter overall survival (OS) time and disease-free survival (DFS) (Li *et al.*, 2021).

HMGA2 belongs to the family of small high mobility group proteins containing AT-hook DNA-binding domains. Is aberrantly regulated in several human tumors including tumors within the HNSCC type (Meyer *et al.*, 2007; Sterenczak *et al.*, 2014; Loeschke *et al.*, 2016; Zhao *et al.*, 2016; Binabaj *et al.*, 2019). The overexpression is associated with increased invasiveness, metastasis, epithelial-mesenchymal transition and stemness, and poor prognosis (Fang *et al.*, 2017; Li Z *et al.*, 2022).

POU Class 5 Homeobox 1 (POU5F1), also known as OCT-4, is a transcription factor of the POU family. Elevated POU5F1 was associated with tumorigenesis; poor overall survival, disease free survival, and recurrence free survival (RFS); and DSS in several cancers (Reers *et al.*, 2014; Koo *et al.*, 2015; Miyoshi *et al.*, 2018; He *et al.*, 2020; Sun *et al.*, 2021).

To investigate the regulatory factors silenced in SCCOT we analyzed genomic enhancer regions with gain of DNA methylation and associated downregulation of gene expression in tumor samples of the tongue compared to adjacent regions. We identified 6,956 hypermethylated probes and 331 probe-gene pairs with negative correlation (Table S3A, Figure 2a). A pathway analysis of these genes revealed inhibition of genes associated with transcription activity and development in tumor samples (Figure 2b).

Next, we identified the enrichment of 76 TF binding motifs in regions with gain of DNA methylation in tumor samples, ABT4, KLF2 and E2F4 being the motifs most represented (Table S3B). Finally, we identified the downregulation of TF in tumors from the tongue, some of them associated with tumor suppression: HLF, PAX1, RARB, ZNF471, and ZNF582 (Figure 2c, Table 2).

HLF encodes a member of the proline and acidic-rich (PAR) protein family, a subset of the bZIP transcription factors. In Glioma, the overexpression could inhibit proliferation and invasion, in Triple Negative Breast Cancer promotes the ferroptosis resistance in TNBC cells via GGT1 and ultimately facilitates the malignant tumor progression (Liu Q *et al.*, 2021; Li H *et al.*, 2022).

PAX1 is a member of the paired box (PAX) transcription factor family. Inactivation of PAX1 gene by greater promoter methylation and/or somatic mutation may lead to apoptosis resistance, tumor cell proliferation and migration, repression of terminal differentiation, and progression of oral carcinogenesis (Cheng *et al.*, 2016).

Retinoic Acid Receptor Beta (RARB), is a member of the thyroid-steroid hormone receptor superfamily of nuclear transcriptional regulators. Changing in RARB expression is associated with cellular sensitivity to retinoid in numerous cancer cells, including HNSCC cells, betel quid related hypermethylation of RARB will increase the tumorigenesis and poor treatment outcome of oral cancer (Lai *et al.*, 2014).

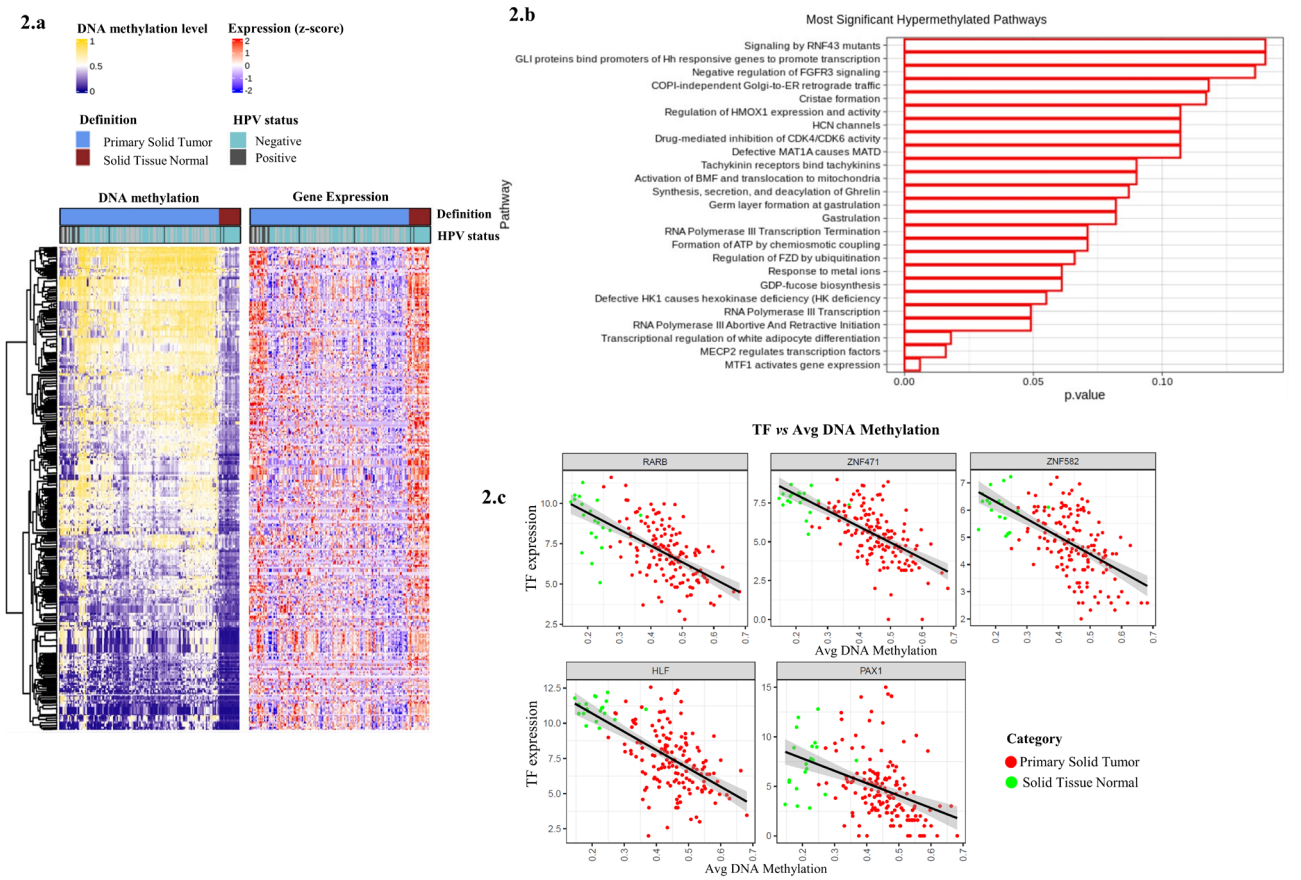


**Figure 2** – Hypermethylation analysis panel. **a.** Heatmap of hypermethylation analysis of tumor vs. non-tumor samples. **b.** Pathway analysis performed on Reactome showing the most significant hypermethylated pathways. **c.** Scatter plot of expression of the respective TFs (HLF, PAX1, RARB, ZNF471, and ZNF582) vs DNA methylation. Source: Author.

**Table 2 –** Hypermethylated TF's related to HNSCC: biological function and influence in cancer development.

| TF | Biological Function | Oncogene or tumor suppressive | | Ref |
|---|---|---|---|---|
| | | Cancer | HNSCC | |
| HLF | Poorer progression-free survivals, cell growth, promotes anaerobic metabolism, distance metastasis and early relapse and progression | oncogene / tumor suppressive | tumor suppressive | (Chen *et al.*, 2020; Liu B *et al.*, 2021; Li *et al.*, 2022) |
| PAX1 | Poor prognosis, apoptosis resistance, tumor cell proliferation and migration, repression of terminal differentiation, and finally progression of oral carcinogenesis | tumor suppressive | tumor suppressive | (Huang *et al.*, 2010; Huang *et al.*, 2014) |
| RARB | involved in suppressing cell growth and tumorigenicity, regulating gene expression and its retinoid-mediated antiproliferative, differentiative, immuno-modulatory, and apoptotic-inducing properties | tumor suppressive | tumor suppressive | (Giannini *et al.*, 1997; Bebek *et al.*, 2012) |
| ZNF471 | poorer survival, suppressed cell proliferation, migration, and invasion | tumor suppressive | tumor suppressive | (Bhat *et al.*, 2016; Wang *et al.*, 2020) |
| ZNF582 | DNA damage response, proliferation, cell cycle control, and neoplastic transformation | tumor suppressive | tumor suppressive | (Liu B *et al.*, 2019; Camuzi *et al.*, 2021) |

ZNF471 e ZNF582 are members of the zinc-finger family. They are involved in all the principal pathways of cancer progression from carcinogenesis to metastasis formation, playing a key role as recruiters of chromatin modifiers or as structural proteins that regulate cancer cell migration and invasion (Huang *et al.*, 2012; Tao *et al.*, 2020).

Several studies have shown the relevance of studying distal regulatory regions of the gene code, and how epigenetic modulations can affect the overall gene expression (Stadler *et al.*, 2011; Aran *et al.*, 2013; Yao *et al.*, 2015; Lin *et al.*, 2016). Within the hypomethylation analysis, we observed that the candidates master regulators TFs in SCOOT are associated with increased invasiveness; metastasis; epithelial - mesenchymal transition; poor prognosis, overall survival, and recurrence-free survival; and stemness (Zhang *et al.*, 2018; Li *et al.*, 2021; Shahoumi, 2021; Li Z *et al.*, 2022). The majority of these genes, in healthy cells, are associated with cell differentiation during embryonic development (Moon *et al.*, 2012; Zhou *et al.*, 2019; He *et al.*, 2020). On the other hand, in the hypermethylation analysis, we noted that these candidates master regulators TFs, in oral cancers, are associated with tumor suppression and regulation, and their downregulation or inactivation can correlate with tumorigenesis and poor prognosis (Urrutia, 2003; Lai *et al.*, 2014; Cheng *et al.*, 2016). Besides, the ZNF582 and PAX1 are potential biomarkers for differentiating moderate dysplasia or worse (MODY+) oral lesions (Cheng *et al.*, 2018).

These results combined with the literature review indicate that ELMER package is an important and relevant tool to deduce regulatory element landscapes and gene regulatory networks from cancer methylomes (Yao *et al.*, 2015; Silva *et al.*, 2019). Moreover, the understanding of the epigenetic modulations, and how they can provoke global alterations during tumor evolution, is vital in order to characterize clinically heterogeneous malignancies, further understand their physiology, and improve diagnosis and treatment. Since epigenetic alteration of enhancer sites, such as methylation, is related to gene expression profiles of mutated cells, the identified MRTFs can further be used as biological markers for oral cancers, however, to completely understand their mechanisms and relevance in the overall development of the disease, further studies are required.

## Acknowledgments

Support for this work was provided by São Paulo Research Foundation (FAPESP) grants 2018/00583-0, 2019/14928-1, and 2021/08190-0; and by Coordination of Improvement of Higher Education Personnel (CAPES).

## Conflict of Interest

The authors declare that no conflict of interest could be perceived as prejudicial to the impartiality of the reported research.

## Authors Contributions

LO data curation, investigation, methodology, ELMER analysis, visualization, writing-original draft, writing-review and editing, data generation; LMMF investigation, data processing and analysis, writing-review and editing; IDE investigation, methodology, ELMER analysis; TMM project administration, conceptualization, formal analysis, supervision, visualization, writing-review and editing. All authors read and approved the final version.

## References

Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie T-X, Zhang J, Wang J *et al.* (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333:1154-1157.

Aran D, Sabato S and Hellman A (2013) DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol 14:R21.

Bebek G, Bennett KL, Funchain P, Campbell R, Seth R, Scharpf J, Burkey B and Eng C (2012) Microbiomic subprofiles and MDR1 promoter methylation in head and neck squamous cell carcinoma. Hum Mol Genet 21:1557-1565.

Bergman Y and Cedar H (2013) DNA methylation dynamics in health and disease. Nat Struct Mol Biol 20:274-281.

Bhat S, Kabekkodu SP, Noronha A and Satyamoorthy K (2016) Biological implications and therapeutic significance of DNA methylation regulated genes in cervical cancer. Biochimie 121:298-311.

Binabaj MM, Soleimani A, Rahmani F, Avan A, Khazaei M, Fiuji H, Soleimanpour S, Ryzhikov M, Ferns GA, Bahrami A *et al.* (2019) Prognostic value of high mobility group protein A2 (HMGA2) over-expression in cancer progression. Gene 706:131-139.

Bugshan A and Farooq I (2020) Oral squamous cell carcinoma: metastasis, potentially associated malignant disorders, etiology and recent advancements in diagnosis. F1000Res 9:229.

Camuzi D, Simão TA, Dias F, Pinto LFR and Soares-Lima SC (2021) Head and neck cancers are not alike when tarred with the same brush: An epigenetic perspective from the cancerization field to prognosis. Cancers 13:5630.

Chen J, Liu A, Lin Z, Wang B, Chai X, Chen S, Lu W, Zheng M, Cao T, Zhong M *et al.* (2020) Downregulation of the circadian rhythm regulator HLF promotes multiple-organ distant metastases in non-small cell lung cancer through PPAR/NF-κb signaling. Cancer Lett 482:56-71.

Cheng S-J, Chang C-F, Lee J-J, Chen H-M, Wang H-J, Liou Y-L, Yen C and Chiang C-P (2016) Hypermethylated ZNF582 and PAX1 are effective biomarkers for detection of oral dysplasia and oral cancer. Oral Oncol 62:34-43.

Cheng S-J, Chang C-F, Ko H-H, Lee J-J, Chen H-M, Wang H-J, Lin H-S and Chiang C-P (2018) Hypermethylated ZNF582 and PAX1 genes in mouth rinse samples as biomarkers for oral dysplasia and oral cancer detection. Head Neck 40:355-368.

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I *et al.* (2016) TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 44:e71.

Ehrlich M and Lacey M (2013) DNA methylation and differentiation: Silencing, upregulation and modulation of gene expression. Epigenomics 5:553-568.

Fang C-Y, Liew P-L, Chen C-L, Lin Y-H, Fang C-L and Chen W-Y (2017) High HMGA2 expression correlates with reduced recurrence-free survival and poor overall survival in oral squamous cell carcinoma. Anticancer Res 37:1891-1899.

Franceschi D, Gupta R, Spiro RH and Shah JP (1993) Improved survival in the treatment of squamous carcinoma of the oral tongue. Am J Surg 166:360-365.

Giannini F, Maestro R, Vukosavljevic T, Pomponi F and Boiocchi M (1997) All-trans, 13-cis and 9-cis retinoic acids induce a fully reversible growth inhibition in HNSCC cell lines: Implications for in vivo retinoic acid use. Int J Cancer 70:194-200.

Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C *et al.* (2022) The reactome pathway knowledgebase 2022. Nucleic Acids Res 50:D687-D692.

He D, Zhang X and Tu J (2020) Diagnostic significance and carcinogenic mechanism of pan-cancer gene POU5F1 in liver hepatocellular carcinoma. Cancer Med 9:8782-8800.

Hier J, Vachon O, Bernstein A, Ibrahim I, Mlynarek A, Hier M, Alaoui-Jamali MA, Maschietto M and da Silva SD (2021) Portrait of DNA methylated genes predictive of poor prognosis in head and neck cancer and the implication for targeted therapy. Sci Rep 11:10012.

Huang R-L, Chang C-C, Su P-H, Chen Y-C, Liao Y-P, Wang H-C, Yo Y-T, Chao T-K, Huang H-C, Lin C-Y *et al.* (2012) Methylomic analysis identifies frequent DNA methylation of zinc finger protein 582 (ZNF582) in cervical neoplasms. PLoS One 7:e41060.

Huang T-H, Lai H-C, Liu H-W, Lin CJ, Wang K-H, Ding D-C and Chu T-Y (2010) Quantitative analysis of methylation status of the PAX1 gene for detection of cervical cancer. Int J Gynecol Cancer 20:513-519.

Huang Y-K, Peng B-Y, Wu C-Y, Su C-T, Wang H-C and Lai H-C (2014) DNA methylation of PAX1 as a biomarker for oral squamous cell carcinoma. Clin Oral Investig 18:801-808.

Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE and Grandis JR (2020) Head and neck squamous cell carcinoma. Nat Rev Dis Primers 6:92.

Koo BS, Lee SH, Kim JM, Huang S, Kim SH, Rho YS, Bae WJ, Kang HJ, Kim YS, Moon JH *et al.* (2015) Oct4 is a critical regulator of stemness in head and neck squamous carcinoma cells. Oncogene 34:2317-2324.

Lai Z-L, Tsou Y-A, Fan S-R, Tsai M-H, Chen H-L, Chang N-W, Cheng J-C and Chen C-M (2014) Methylation-associated gene silencing of RARB in areca carcinogens induced mouse oral squamous cell carcinoma. Biomed Res Int 2014:378358.

Li C, Zhu M, Zhu J, Lu Q, Shi B, Sun B and Chen H (2021) SOX12 contributes to the activation of the JAK2/STAT3 pathway and malignant transformation of esophageal squamous cell carcinoma. Oncol Rep 45:129-138.

Li H, Yang P, Wang J, Zhang J, Ma Q, Jiang Y, Wu Y, Han T and Xiang D (2022) HLF regulates ferroptosis, development and chemoresistance of triple-negative breast cancer by activating tumor cell-macrophage crosstalk. J Hematol Oncol 15:2.

Li Z, Wu X, Li J, Yu S, Ke X, Yan T, Zhu Y, Cheng J and Yang J (2022) HMGA2-Snai2 axis regulates tumorigenicity and stemness of head and neck squamous cell carcinoma. Exp Cell Res 418:113271.

Lin CY, Erkek S, Tong Y, Yin L, Federation AJ, Zapatka M, Haldipur P, Kawauchi D, Risch T, Warnatz H-J *et al.* (2016) Active medulloblastoma enhancers reveal subgroup-specific cellular origins. Nature 530:57-62.

Liu B, Su Q, Ma J, Chen C, Wang L, Che F and Heng X (2021) Prognostic value of eight-gene signature in head and neck squamous carcinoma. Front Oncol 11:657002.

Liu Q, Ge H, Liu P and Li Y (2021) High Hepatic leukemia factor expression indicates a favorable survival in glioma patients. Medicine (Baltimore) 100:e23980.

Liu X, Zhao X and Gou C (2019) Identification of novel methylated DNA marker ZNF569 for head and neck squamous cell carcinoma. J Cancer 10:2250-2260.

Loeschke S, Ohlmann AK, Bräsen JH, Holst R and Warnke PH (2016) Prognostic value of HMGA2, P16, and HPV in oral squamous cell carcinomas. J Craniomaxillofac Surg 44:1422-1429.

Mansoori B, Mohammadi A, Ditzel HJ, Duijf PHG, Khaze V, Gjerstorff MF and Baradaran B (2021) HMGA2 as a critical regulator in cancer development. Genes (Basel) 12:269.

Meyer B, Loeschke S, Schultze A, Weigel T, Sandkamp M, Goldmann T, Vollmer E and Bullerdiek J (2007) HMGA2 overexpression in non-small cell lung cancer. Mol Carcinog 46:503-511.

Miyoshi N, Fujino S, Ohue M, Yasui M, Takahashi Y, Sugimura K, Tomokuni A, Akita H, Kobayashi S, Takahashi H *et al.* (2018) The POU5F1 gene expression in colorectal cancer: A novel prognostic marker. Surg Today 48:709-715.

Moon S-M, Kim S-A, Yoon J-H and Ahn S-G (2012) HOXC6 is deregulated in human head and neck squamous cell carcinoma and modulates Bcl-2 expression. J Biol Chem 287:35678-35688.

Pan J, Silva TC, Gull N, Yang Q, Plummer JT, Chen S, Daigo K, Hamakubo T, Gery S, Ding L-W *et al.* (2020) Lineage-specific epigenomic and genomic activation of oncogene HNF4A promotes gastrointestinal adenocarcinomas. Cancer Res 80:2722-2736.

Reers S, Pfannerstill A-C, Maushagen R, Pries R and Wollenberg B (2014) Stem cell profiling in head and neck cancer reveals an Oct-4 expressing subpopulation with properties of chemoresistance. Oral Oncol 50:155-162.

Seok J, Gil M, Dayem AA, Saha SK and Cho S-G (2021) Multi-omics analysis of SOX4, SOX11, and SOX12 expression and the associated pathways in human cancers. J Pers Med 11:823.

Shahoumi LA (2021) Oral cancer stem cells: Therapeutic implications and challenges. Front Oral Health 2:685236.

Silva TC, Colaprico A, Olsen C, D'Angelo F, Bontempi G, Ceccarelli M and Noushmehr H (2016) TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. F1000Res 5:1542.

Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, Lin D-C and Berman BP (2019) ELMER v.2: An R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. Bioinformatics 35:1974-1977.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480:490-495.

Steinbichler TB, Dudas J, Ingruber J, Glueckert R, Sprung S, Fleischer F, Cidlinsky N, Dejaco D, Kofler B, Giotakis AI *et al.* (2020) Slug is A surrogate marker of epithelial to mesenchymal transition (EMT) in head and neck cancer. J Clin Med 9:2061.

Sterenczak KA, Eckardt A, Kampmann A, Willenbrock S, Eberle N, Länger F, Kleinschmidt S, Hewicker-Trautwein M, Kreipe H, Nolte I *et al.* (2014) HMGA1 and HMGA2 expression and comparative analyses of HMGA2, Lin28 and let-7 miRNAs in oral squamous cell carcinoma. BMC Cancer 14:694.

Su Z, Bao W, Yang G, Liu J and Zhao B (2022) SOX12 promotes thyroid cancer cell proliferation and invasion by regulating the expression of POU2F1 and POU3F1. Yonsei Med J 63:591-600.

Sun S, Yang H, Wang F and Zhao S (2021) Oct4 downregulation-induced inflammation increases the migration and invasion rate of oral squamous cell carcinoma. Acta Biochim Biophys Sin (Shanghai) 53:1440-1449.

Tao C, Luo J, Tang J, Zhou D, Feng S, Qiu Z, Putti TC, Xiang T, Tao Q, Li L *et al.* (2020) The tumor suppressor Zinc finger protein 471 suppresses breast cancer growth and metastasis through inhibiting AKT and Wnt/β-catenin signaling. Clin Epigenetics 12:173.

Urrutia R (2003) KRAB-containing zinc-finger repressor proteins. Genome Biol 4:231.

Wang A, Liu X, Sheng S, Ye H, Peng T, Shi F, Crowe DL and Zhou X (2009) Dysregulation of heat shock protein 27 expression in oral tongue squamous cell carcinoma. BMC Cancer 9:167.

Wang C, Liu X, Huang H, Ma H, Cai W, Hou J, Huang L, Dai Y, Yu T and Zhou X (2012) Deregulation of Snai2 is associated with metastasis and poor prognosis in tongue squamous cell carcinoma. Int J Cancer 130:2249-2258.

Wang Q, Wu J, Huang H, Jiang Y, Huang Y, Fang H, Zheng G, Zhou X, Wu Y, Lei C *et al.* (2020) lncRNA LIFR-AS1 suppresses invasion and metastasis of non-small cell lung cancer via the miR-942-5p/ZNF471 axis. Cancer Cell Int 20:180.

Yao L, Shen H, Laird PW, Farnham PJ and Berman BP (2015) Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. Genome Biol 16:105.

Zavridou M, Mastoraki S, Strati A, Koutsodontis G, Klinakis A, Psyrri A and Lianidou E (2020) Direct comparison of size-dependent versus EpCAM-dependent CTC enrichment at the gene expression and DNA methylation level in head and neck squamous cell carcinoma. Sci Rep 10:6551.

Zhang F, Ren C-C, Liu L, Chen Y-N, Yang L and Zhang X-A (2018) HOXC6 gene silencing inhibits epithelial-mesenchymal transition and cell viability through the TGF-β/smad signaling pathway in cervical carcinoma cells. Cancer Cell Int 18:204.

Zhao X-P, Zhang H, Jiao J-Y, Tang D-X, Wu Y-L and Pan C-B (2016) Overexpression of HMGA2 promotes tongue cancer metastasis through EMT pathway. J Transl Med 14:26.

Zhou W, Gross KM and Kuperwasser C (2019) Molecular regulation of Snai2 in development and disease. J Cell Sci 132:jcs235127.

## Supplementary material

The following online material is available for this article:

Table S1 – Identification of tumor samples from the tongue region and non-tumor samples from GDC Data Portal.

Table S2: S2A – Hypomethylated probes identified from the correlation between tumor samples from the tongue region and non-tumor samples. S2B: Transcription factors binding motifs and TF lists associated with DNA hypomethylation in SCCOT.

Table S3: S3A – Hypermethylated probes identified from the correlation between tumor samples from the tongue region and non-tumor samples. S3B: Transcription factors binding motifs and TF lists associated with DNA hypermethylation in SCCOT.

*Associate Editor: Emmanuel Dias Neto*