Analysis of Methylation Levels and Prognostic Correlations in Oral Squamous Cell Carcinoma Based on TCGA/GTEx Data

Yuting Chen*, Taotao Sun

Hangzhou Normal University Affiliated Hospital (School of Clinical Medicine, School of Stomatology), Hangzhou, China

Abstract: Objective: To investigate the relationship between DNA methylation and the prognosis of oral squamous cell carcinoma (OSCC), and to introduce normal tissue data from the GTEx database as a control. Methods: DNA methylation data and expression data of OSCC were obtained from the TCGA database, and methylation and expression data of normal tissues were obtained from the GTEx database as controls. Differential methylation site analysis of DNA methylation chips was performed using the CHAMP package in R software. Methylation annotation was conducted using GENCODE 28V probes, and prognostic-related genes were identified through Cox regression analysis. Kaplan-Meier survival analysis, univariate and multivariate analyses, and receiver operating characteristic (ROC) curves were used to evaluate factors affecting prognosis. Results: A total of 38 differentially expressed genes were identified, including 29 upregulated genes (VMP1, SPNS1, EIF4EBP1, NLRC4, KLHL24, HIF1A, TP63, ITGB4, BID, VEGFA, DRAM1, GAA, BNIP3, ITGA3, BIRC5, SPHK1, CXCR4, NRG1, ITGA6, FADD, DDIT3, SERPINA1, IFNG, IL24) and 9 downregulated genes (NRG2, TP53INP2, FOS, HSPB8, PTK6, NRG3, PRKN, CCL2, ULK3). A total of 5213 differentially methylated sites were detected in the transcription start site (TSS) and 14233 in the gene body. After removing duplicate data, 3696 independent sites were obtained. Comparison with normal tissue data from the GTEx database yielded 1522 differentially methylated sites, of which 1521 showed consistent trends in both databases. Univariate Cox regression analysis identified 53 genes with prognostic significance. Multivariate Cox regression analysis based on these 53 genes confirmed that 4 genes (VMP1, HIF1A, VEGFA, IL24) were statistically significant ($P<0.05$). Conclusions: The methylation levels of VMP1, HIF1A, VEGFA, and IL24 are associated with the prognosis of oral squamous cell carcinoma, and the normal tissue data from the GTEx database provide important control support for the study. Hypermethylation of these genes was associated with downregulation of their expression levels, which may impact key functional pathways such as angiogenesis (VEGFA), hypoxia response (HIF1A), and immune regulation (IL24). These findings provide mechanistic insights into the role of DNA methylation in the development and prognosis of OSCC.

1 Introduction

Oral squamous cell carcinoma (OSCC) is one of the major types of cancer worldwide, accounting for 2% of all cancer cases, with a mortality rate close to 50% [1-3]. The development of OSCC is a complex, multi-step process involving the abnormal expression of multiple genes and epigenetic changes [4-6]. In recent years, the rapid accumulation of cancer omics data with the development of medical big data has provided a data foundation for studying the mechanisms of OSCC development, clinical diagnosis, and treatment [7]. DNA methylation, an important epigenetic modification, plays a key role in the development and prognosis of OSCC [8,9]. Studies have shown that abnormal DNA methylation levels are closely related to the occurrence of OSCC and may serve as good indicators for early diagnosis and prognostic assessment of OSCC [10-12]. This study aims to identify differentially methylated genes related to the prognosis of OSCC through the TCGA and GTEx databases, providing new biomarkers for clinical diagnosis and treatment.

2 Materials and Methods

2.1 Differential Gene Screening

Methylation and expression data of oral squamous cell carcinoma (OSCC) were downloaded from the TCGA database (https://portal.gdc.cancer.gov/), including 346 tumor tissues and 50 adjacent normal tissues. Additionally, methylation data of normal tissues were downloaded from the GTEx database (https://gtexportal.org/) as controls, totaling 11 samples. These data were used to analyze DNA methylation levels in OSCC and their relationship with prognosis.

* Corresponding author: c18357757158@126.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

2.2 Data Preprocessing

The DNA methylation chip data from the TCGA database were preprocessed using the ChAMP package in R software. First, the champ.QC function was used for quality control, generating multiple charts including dendrograms, density plots, and multidimensional scaling plots (MDSPlot) to assess data quality. Subsequently, the champ.norm function was used to normalize the data, employing the BMIQ method to correct for Type-II probe bias. The normalized data were used for subsequent differential methylation analysis.

2.3 Differential Methylation Analysis

Differential methylation probe (DMP) analysis was conducted using the champ.DMP function in the ChAMP package to identify CpG sites with significant differences in methylation levels between tumor and normal tissues. Additionally, the champ.DMR function was used to detect differentially methylated regions (DMRs), employing the Bumphunter algorithm to evaluate candidate DMRs. The analysis results were visualized using the DMP, GUI and DMR. The criteria for selecting DMPs included a p-value threshold of <0.05 and an effect size greater than 0.5. These stringent criteria ensured the robustness of the identified DMPs.

2.4 Methylation Annotation and Gene Expression Correlation Analysis

Methylation data obtained from the TCGA database were annotated using GENCODE 28V probes, linking differentially methylated sites with gene expression data. The correlation between methylation levels and gene expression was analyzed to identify genes potentially regulated by methylation.

2.5 Statistical Analysis

All statistical analyses and plotting were completed using R software. Kaplan-Meier methods were used for survival analysis, and Cox regression models were employed for univariate and multivariate analyses. ROC curves were used to evaluate the sensitivity and specificity of the prognostic model. In all statistical tests, a P-value less than 0.05 was considered statistically significant.

3 Results

3.1 Differential Gene Screening

A total of 38 differentially expressed genes were identified, including 29 upregulated genes (VMP1, SPNS1, EIF4EBP1, NLRC4, KLHL24, HIF1A, TP63, ITGB4, BID, VEGFA, DRAM1, GAA, BNIP3, ITGA3, BIRC5, SPHK1, CXCR4, NRG1, ITGA6, FADD, DDIT3, SERPINA1, IFNG, IL24) and 9 downregulated genes (NRG2, TP53INP2, FOS, HSPB8, PTK6, NRG3, PRKN, CCL2, ULK3) (Figure 1).

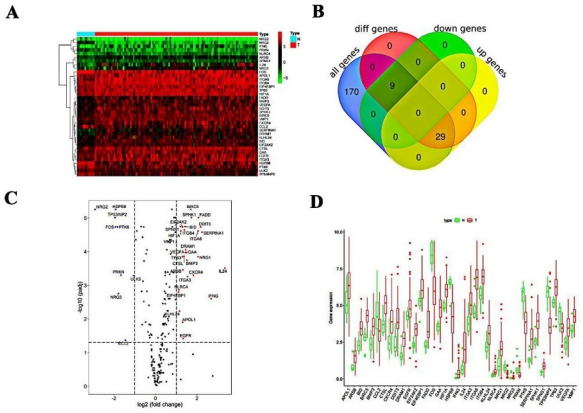

Figure 1. Differential Analysis of OSCC-Related Genes
A. In the differential gene heatmap, red represents high gene expression, and green represents low gene expression, with clustering based on gene expression levels. B. Intersection of differential genes. C. Visualization of differential genes. D. Differential gene expression.

3.2 GO Enrichment and KEGG Pathway Enrichment Analysis of Differential Genes

From the biological process (BP) analysis, these genes were mainly concentrated in autophagy and processes utilizing autophagy mechanisms. From the molecular function (MF) analysis, these genes were mainly concentrated in protein heterodimer activity. Upregulated genes were primarily involved in fluid shear stress and atherosclerosis, ErbB signaling pathways, and rheumatoid arthritis; downregulated genes were mainly involved in human cytomegalovirus infection and human papillomavirus infection (Figure 2).

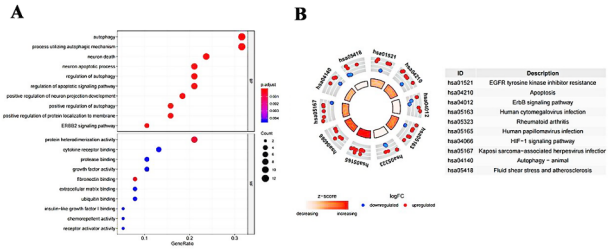

Figure 2. Functional and Pathway Enrichment Analysis of Differential Genes in OSCC
A. GO Enrichment Analysis. B. KEGG Pathway Enrichment Analysis.

3.3 Differential Methylation Sites and Differentially Expressed Genes

Data from 346 tumor tissues and 50 adjacent normal tissues were obtained from the TCGA database, and data from 11 normal tissues were obtained from the GTEx database. Based on database analysis, 93746 differentially methylated sites were identified in OSCC (TCGA). Compared with normal tissues from the GTEx database, a total of 13964 genes exhibited significant differential expression in OSCC (Figure 3).

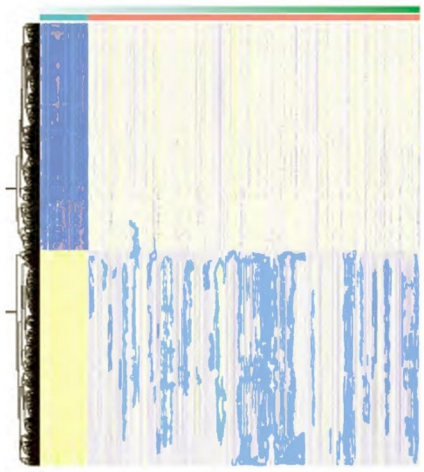

Figure 3. Screening of Differential Methylation Sites and Genes in OSCC

3.4 Differential Methylation Genes: VMP1, HIF1A, VEGFA, and IL24

A total of 5213 differentially methylated sites were detected in the transcription start site (TSS) and 14233 in the gene body. After removing duplicate data, 3696 independent sites were obtained. Comparison with normal tissue data from the GTEx database yielded 1522 differentially methylated sites, of which 1521 showed consistent trends in both databases. Univariate Cox regression analysis identified 53 genes with prognostic significance. Multivariate Cox regression analysis based on these 53 genes confirmed that 4 genes (VMP1, HIF1A, VEGFA, IL24) were statistically significant ($P<0.05$) (Figure 4). Hypermethylation of these genes was associated with downregulation of their expression levels, which may impact key functional pathways such as angiogenesis (VEGFA), hypoxia response (HIF1A), and immune regulation (IL24). These findings provide mechanistic insights into the role of DNA methylation in the development and prognosis of OSCC.

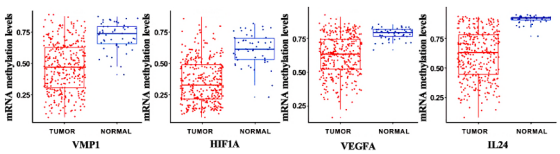

Figure 4. Differential Methylation Sites and Corresponding miRNA Expression Levels of VMP1, HIF1A, VEGFA, and IL24 in OSCC

4 Conclusions

This study systematically analyzed the DNA methylation levels in oral squamous cell carcinoma (OSCC) and their relationship with prognosis by integrating methylation and expression data from the TCGA and GTEx databases. The results showed that the methylation levels of VMP1, HIF1A, VEGFA, and IL24 were significantly elevated in OSCC patients and were significantly correlated with patient survival rates. The methylation levels of these genes may serve as prognostic biomarkers for OSCC, providing new evidence for clinical diagnosis and treatment. Additionally, this study found that HIF1A, TNM stage, and age are independent prognostic indicators significantly correlated with patient survival rates. This suggests that in the prognostic assessment of OSCC, clinical characteristics of patients, in addition to gene methylation levels, should be comprehensively considered to improve the accuracy of prognostic assessment. Future research needs to further validate the expression of these genes' methylation levels in different ethnic groups and clinical samples to determine their feasibility and reliability as prognostic biomarkers for OSCC.

While this study provides valuable insights into the prognostic significance of DNA methylation in OSCC, it has certain limitations. The retrospective design and lack of experimental validation may limit the generalizability of the findings. Additionally, the sample size was limited to certain ethnic groups, and future studies should include diverse ethnic cohorts to validate the results. Despite these limitations, the identified methylation markers hold promise for clinical application in the prognosis of OSCC.

5 Discussion

This study identified differentially methylated genes related to the prognosis of oral squamous cell carcinoma (OSCC) through the TCGA and GTEx databases: VMP1, HIF1A, VEGFA, and IL24. These genes may play important roles in the development and prognosis of OSCC. VMP1 (Vesicle Membrane Protein 1) is a gene associated with apoptosis and autophagy, and its abnormal expression in various cancers has been reported [13-15]. HIF1A (Hypoxia-Inducible Factor 1 Alpha) is a hypoxia-inducible factor closely related to tumor invasion and metastasis. VEGFA (Vascular Endothelial Growth Factor A) is involved in tumor angiogenesis [16,17]. IL24 (Interleukin 24) is a cytokine associated with immune regulation and tumor suppression [18,19]. The methylation levels of these genes are significantly correlated with the survival rates of OSCC patients, indicating that they may serve as prognostic biomarkers for OSCC.

DNA methylation is a common epigenetic modification that regulates gene expression by adding methyl groups to DNA sequences [20-22]. Abnormal DNA methylation levels are closely related to the development of various cancers [23]. In OSCC, changes in DNA methylation levels may affect gene expression, thereby influencing the biological behavior of tumors [24,25]. This study identified differentially methylated genes related to the prognosis of OSCC by analyzing methylation data from the TCGA and GTEx databases. The methylation levels of these genes were significantly elevated in OSCC patients and were significantly correlated with patient survival rates. This suggests that the methylation levels of these genes may serve as prognostic biomarkers for OSCC, providing new evidence

for clinical diagnosis and treatment.

Additionally, this study found that HIF1A, TNM stage, and age are independent prognostic indicators significantly correlated with patient survival rates. This suggests that in the prognostic assessment of OSCC, clinical characteristics of patients, in addition to gene methylation levels, should be comprehensively considered [26,27]. TNM stage is a commonly used cancer staging method in clinical practice, reflecting the size of the tumor, lymph node metastasis, and distant metastasis [28]. Age is also an important factor affecting the prognosis of OSCC, with lower survival rates typically observed in elderly patients [29-32]. Therefore, in the prognostic assessment of OSCC, both gene methylation levels and clinical characteristics of patients should be considered to improve the accuracy of prognostic assessment.

However, this study has certain limitations. First, the data are from the TCGA and GTEx databases, with samples limited to certain ethnic groups, and there may be ethnic differences. Second, the results of this study have not been verified by clinical samples and experimental research, and all the above findings need to be validated through a series of molecular biology experiments. Future research needs to further validate the expression of these genes' methylation levels in different ethnic groups and clinical samples to determine their feasibility and reliability as prognostic biomarkers for OSCC. While this study provides valuable insights into the prognostic significance of DNA methylation in OSCC, it has certain limitations. The retrospective design and lack of experimental validation may limit the generalizability of the findings. Additionally, the sample size was limited to certain ethnic groups, and future studies should include diverse ethnic cohorts to validate the results. Despite these limitations, the identified methylation markers hold promise for clinical application in the prognosis of OSCC.

References

1. Ma Tianying, Liu Honggang. Research Progress on Tumor Budding in Head and Neck Squamous Cell Carcinoma [J/OL]. Journal of Clinical and Experimental Pathology, 2025, (01): 93-98 [2025-01-23].
2. Niu Jiaxin, Zeng Ju, Zhang Cong, et al. Research Progress on the Immune Role of Dendritic Cells in Head and Neck Squamous Cell Carcinoma [J/OL]. Journal of Clinical and Experimental Pathology, 2025, (01): 81-85+92 [2025-01-23].
3. Ye Jiaxue, Wang Shiyuan. Overview of the Antitumor Effects of Honokiol [J]. Journal of Shandong University of Traditional Chinese Medicine, 2025, 49(01): 126-134.
4. Muhetaibayer Hairoula, Zhang Li. Research Progress on Hypoxia-Induced Heterogeneity of Tumor-Derived Exosomes [J]. Chinese Journal of Biologicals, 2025, 38(01): 122-128.
5. Fan Die, Yao Xinyi, Sun Ruijuan, et al. Research Progress on the Role of Lysosomes in Tumorigenesis and Metastasis [J/OL]. Science China Life Sciences, 1-19 [2025-01-23].
6. Wang Bin, Feng Yi. Research Progress on Volatile Organic Compounds in Exhaled Breath [J/OL]. Journal of Sun Yat-sen University (Medical Science Edition), 1-11 [2025-01-23].
7. Jiang Zongying, Zhang Cong, Jiang Xinwei. Research Progress on Heterogeneous Nuclear Ribonucleoprotein F in Malignant Tumors [J]. Journal of Nanjing Medical University (Natural Science Edition), 2025, 45(01): 111-118.
8. Wang Ziting, Yang Guantong, Fan Jianchun, et al. The Role of Cancer-Associated Fibroblasts in the Formation of Premetastatic Niches [J/OL]. Chinese Journal of Comparative Medicine, 1-10 [2025-01-23].
9. Wang Shan, Sun Xiangyu, Bai Yang. Clinical Study on CT and MRI in Evaluating Mandibular Involvement in Oral Cancer [J]. Chinese Journal of CT and MRI, 2025, 23(01): 33-34.
10. Hu Shanshan, Liu Yayun, Sheng Deqiao, et al. Research Progress on the Antitumor Effects of Usnic Acid [J/OL]. Pharmaceutical Care & Research, 1-19 [2025-01-23].
11. Tian Yuwei, Wang Jiajia, Liu Hongshan, et al. Effect and Mechanism of Long Non-coding RNA SNHG1 on the Biological Behavior of Human Gallbladder Cancer Cells [J]. Chinese Journal of Practical Diagnosis and Therapy, 2025, 39(01): 21-26.
12. Dong Wenbo, Yu Haoyang, Zhu Yifan, et al. Role of Respiratory Tract Microbiota in the Tumor Immune Microenvironment and Immunotherapy of Non-Small Cell Lung Cancer [J/OL]. Journal of China Medical University, 1-5 [2025-01-23].
13. Jing Xiaoli, Zhang Jun, Wang Rong, et al. Mechanism of Salidroside Upregulating the miR-99a/IGF-1R Axis on the Biological Effects of Gastric Cancer SNU-216 Cells [J]. Chinese Journal of Gerontology, 2025, 45(01): 138-141.
14. Zhao Yuqian, Zhou Huansi, Jin Mingjing, et al. Exploration of the Effect of Gelsemine on the Proliferation and Apoptosis of Tongue Squamous Cell Carcinoma Based on Network Pharmacology and Molecular Docking [J]. Drug Evaluation Research, 2025, 48(01): 110-120.
15. Wan Lu, Tong Qiang, Liu Xiaobo. Research Progress on the Relationship Between Gut Microbiota and Gastrointestinal Tumors [J]. Medical Innovation of China, 2025, 22(01): 179-183.
16. Huang Weibo, Guo Shuangyan, Lyu Jieli, et al. Research Progress on the Chemical Constituents and Pharmacological Effects of Cornus officinalis Sieb. et Zucc. [J]. Journal of Xinxiang Medical University, 2025, 42(01): 79-84.
17. Zhang Xuan, Li Tongtong, Guo Yan, et al. Expression and Clinical Significance of GSDMB in Oral Squamous Cell Carcinoma Tissues [J]. Journal of Qingdao University (Medical Edition), 2024, 60(06): 801-806.
18. Wu Mei, Liang Yanjing, Peng Xuepei, et al. Analysis of Factors Influencing Speech Function After Oral Cancer Surgery [J]. International Journal of Stomatology, 2025, 52(01): 42-49.
19. Sun Ruizhe, Ni Qianwei, Gao Zhan. Application

Progress of Digital Technology in Brachytherapy for Maxillofacial Malignant Tumors [J]. International Journal of Stomatology, 2025, 52(01): 18-24.
20. Li Jingzhe, Zhang Suxin. Research Progress on Phosphatidylinositol 3-Kinase/Protein Kinase B Pathway Inhibitors in Oral Squamous Cell Carcinoma [J]. International Journal of Stomatology, 2025, 52(01): 34-41.
21. Zhu Weijing, Zhang Jiaxing, Wei Yucai, et al. Pan-cancer Analysis of CLDN1 and Its Functional Study in Gastric Cancer [J]. Biomedical Conversion, 2024, 5(04): 70-81.
22. Li Jian, Liang Ye, Zhao Baodong. Research Progress on Metal-Organic Frameworks in Oral Medicine [J]. Chinese Journal of Oral Implantology, 2024, 29(06): 602-609.
23. Ge Bin, Cui Huixian, Liang Kun, et al. Relationship Between the NLRP3/IL-1β, IL-18 Pathway in Peripheral Blood and Prognosis of Children with Severe Mycoplasma Pneumoniae Pneumonia and Its Application Value [J]. Shaanxi Medical Journal, 2024, 53(12): 1658-1662.
24. Yang Wencong, Pan Zhen, Wang Xueting. Incidence and Risk Factors of Myocardial Injury Secondary to Severe Pneumonia in Children [J]. Maternity and Child Health Care of China, 2024, 39(23): 4701-4704.
25. Tang Yufan, Liu Jing, Xue Meng, et al. Analysis of Related Factors Influencing In-hospital Death in Children with Severe Pneumonia and Construction of a Risk Prediction Model [J]. Translational Medicine, 2024, 13(03): 313-317.
26. Huang Lijun, Lin Dan. Application Research of Active Risk Nursing Based on the Children's Medical Coaching Model in Children with Severe Mycoplasma Pneumoniae Pneumonia [J]. Primary Medical Forum, 2024, 28(32): 112-115.
27. Wang Mengnan, Liu Ying, Huang Dongdong. Analysis of the application effect of ambroxol hydrochloride combined with acetylcysteine in the treatment of severe pneumonia in children [J]. Chinese Prescription Drug, 2024, 22(11): 120-123.
28. Qi Yingxiang, Yuan Chao. Predictive value of novel biomarkers for severe Mycoplasma pneumoniae pneumonia in children [J]. Journal of Shaoyang University (Natural Science Edition), 2024, 21(05): 37-44.
29. Song Rui, Li Qun, Liang Huiru. Efficacy of phentolamine combined with antibiotics in the treatment of severe pneumonia in children and its impact on inflammatory factors, lung function, blood gas indicators, and immune function [J]. Aerospace Medicine, 2024, 35(10): 1204-1207.
30. LIU Z, WU C, XIE N, et al. Long non-coding RNA MEG3 inhibits the proliferation and metastasis of oral squamous cell carcinoma by regulating the WNT/β-catenin signaling pathway [J]. Oncology Letters, 2017, 14(4): 4053-4058.
31. WANG Z, YANG B, ZHANG M, et al. lncRNA epigenetic landscape analysis identifies EPIC1 as an oncogenic lncRNA that interacts with MYC and promotes cell-cycle progression in cancer [J]. Cancer Cell, 2018, 33(4): 706-720.
32. TANG H, WU Z, ZHANG Y, et al. Identification and functional analysis of a five-long noncoding RNA prognostic signature for endometrial cancer patients [J]. DNA and Cell Biology, 2019, 38(12): 1480-1498.