Development and Validation of an Immune-related Gene Diagnostic Model for Cancerous Change in Oral Leukoplakia

Yang Lin
Department of Oral and Maxillofacial Surgery, the First Affiliated Hospital of Fujian Medical University

Bingju Gao
Department of Oral and Maxillofacial Surgery, the First Affiliated Hospital of Fujian Medical University

Jun Li
Department of Oral and Maxillofacial Surgery, the First Affiliated Hospital of Fujian Medical University

Peng Chen
Department of Oral and Maxillofacial Surgery, the First Affiliated Hospital of Fujian Medical University

Yu Qiu (✉ qy97@163.com)
Department of Oral and Maxillofacial Surgery, the First Affiliated Hospital of Fujian Medical University

Research Article

Keywords: immune-related gene diagnostic model, cancerous change, oral leukoplakia

Posted Date: December 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2317839/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background**: Oral squamous cell carcinoma (OSCC) is usually preceded by oral potentially malignant disorders, such as oral leukoplakia (OLK). The prevention of the transformation of OLK to OSCC may significantly improve clinical outcomes of patients. Recent studies have highlighted the dynamics of immune microenvironment during oral carcinogenesis. Given this context, this study aimed to develop an immune-related gene diagnostic model for predicting malignant transformation of OLK to OSCC.

**Methods**: In this study, we first compared immune cell infiltration between the control and OLK and OSCC from two GEO databases (GSE85195 and GSE23558) using CIBERSORT algorithm. Next, we compared differentially expressed genes between the control and OSCC and OLK and OSCC using the LIMMA package. Finally, developed an immune-related gene diagnostic model using the LASSO regression and used the GSE26549 and TCGA databases to validate the model's predictive ability.

**Results**: The composition of 22 types of immune cells in each sample was presented in a boxplot (figure 1A). The expression of dendritic cells resting, macrophages M0, macrophages M2, and mast cells resting were significantly different between the OLK and OSCC groups (P<0.001). 1432 and 1256 DEGs were identified from GSE85195 and GSE23558, respectively. Overlapping differential DEGs and the 2483 immune-related genes. Sixty-nine candidate genes were integrated into the LASSO regression to identify OLK cancerous immune signatures. The area under ROC curve (AUC) in the training and internal validation cohorts were 1 and 0.994 (figure 4A), respectively. Finally, a model comprising 16 immune signatures was constructed.

**Conclusions**: This study develop and validate an immune-related gene diagnostic model was a promising objective diagnosis scheme to predict cancer risk of OLK to OSCC.

**Background**

Head and neck cancer is the sixth most common cancer occurring worldwide. Squamous cell carcinoma accounts for more than 90% of head and neck cancer cases [1, 2]. Oral squamous cell carcinoma (OSCC) arise from the mucosal lining of the oral cavity, which is often first diagnosed in its late stages, resulting in advanced regional disease and/or metastasis [3]. Furthermore, its delayed diagnosis is often associated with a greater risk of death [4]. OSCC is usually preceded by oral premalignant lesions (OPLs), such as those have been clinically described as oral leukoplakia (OLK) or erythroplakia. A recent systematic review estimated the overall mean proportion of malignant transformation rate for oral leukoplakia to be 9.7% (7.8–11.7%) [5]. Prevention of the transformation of OLK to OSCC may substantially improve clinical outcomes. Nonetheless, one of the greatest challenges in OLK patients is determining the risk of malignant transformation. OLK biopsy and dysplasia grading remain the gold standard for assessing the risk of transformation. However, the diagnosis of oral epithelial dysplasia is subjective and poorly reproducible. Current schemes of OLK biopsy fail to accurately differentiate non-progressive from progressive OPLs. Therefore, many studies have attempted to define cellular/molecular markers to predict cancer risk for oral potentially malignant disorders [6].
The immune system plays an essential role in preventing the cancer development by detecting and eliminating cancer cells. Currently, clinical trials investigating immunotherapies represent an unprecedented advance in head and neck squamous cell carcinomas (HNSCC)\[7, 8\]. The role of immunoprevention strategies in oral carcinogenesis have been emphasized. Recent studies have highlighted the dynamics of immune microenvironment during oral carcinogenesis and have also reported the unexpected association of M2 macrophages gene expression signatures with oral cancer-free survival of patients with OLK [9]. The regulation of the immune microenvironment plays a key role in tumor malignant transformation. Few immune-related gene prognostic models have been developed and validated and are significantly associated with the overall survival of patients with HNSCC [10–12]. Compared with the tertiary prevention, the early diagnostic, therapeutic, and control efforts of progressive OPLs is key in the strategies of cancer secondary prevention. Given this context, to development and validation of an immune-related gene diagnostic model for predicting malignant transformation from OLK to OSCC may help in establishing immune microenvironment regulations to guide clinical management for better prognosis.

Therefore, in this study, we first compared immune cell infiltration between the control and OLK and OSCC from two Gene Expression Omnibus (GEO) studies to establish cohort. Next, we compared differentially expressed genes (DEGs) between the control and OSCC and OLK and OSCC. Finally, developed and validated an immune-related gene diagnostic model.

**Methods**

**Public data source**

The original RNA-seq transcriptome data and their corresponding clinical information were downloaded from 3 independent GEO (https://www.ncbi.nlm.nih.gov/geo/) sets. The accession number were GSE85195 [13], GSE23558[14], and GSE26549[14]. Including 1 control and 15 OLK and 34 OSCC, 5 control and 27 OSCC, 51 no oral cancer development OLK and 35 oral cancer development OLK, respectively. The GSE85195 and GSE23558 sequencing platform used was the GPL6480. The GSE26549 sequencing platform used was the GPL6244. The between arrays function of linear models was normalized for microarray analysis (LIMMA) package [15], and the robust multi-array average method [16] was used to summarize 1 expression matrix. Data of the TGCA cohort were downloaded from the TCGA database (https://portal.gdc.cancer.gov). Including 32 control and 322 OSCC.

**Immune Cell Infiltration Evaluation**

The GSE85195 and GSE23558 data were merged for immune infiltration analysis and model development. The CIBERSORT [17] algorithm was used to evaluate and compare immune cell leukocyte subsets between the control and OLK and OSCC. The algorithm can transform normalized gene
expression matrix into compositions of infiltrating immune cells, which estimates the abundance of 22 distinct immune cell subsets.

**Differential Expression Analysis**

Differential expression analysis between groups were analyzed for OSCC vs normal and OSCC vs OLK to screen differentially expressed genes. Differential expression analysis was conducted using the limma package. For p-values, false discovery rates (FDRs) were utilized for multiple testing correction. Absolute log2 [fold change (FC)] > 1.0 and FDR < 0.05 were considered the cutoff criteria. Gene Ontology (GO) [18] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [16] enrichment analyses were performed to identify biological processes and pathways, with FDR p-values < 0.05 as cutoff values. Differential expression analysis was performed to explore biological signaling pathways.

**Construction and validation of the prediction model**

Overlapping differential genes from the GEO dataset and immune-related genes from the ImmPort database (https://immport.niaid.nih.gov) were retrieved to obtain candidate immune-related genes. Least absolute shrinkage and selection operator (LASSO) regression was used to construct a prediction model by utilizing the generalized linear model (glmnet) package for identifying OLK to OSCC immune signatures in the GSE85195 and GSE23558 samples. In addition, the GSE26549 and TCGA datasets were used as external validation sets. Receiver operating characteristic (ROC) curves were plotted to assess sensitivity and specificity.

**Statistical analysis**

The R (version 3.6.1, The R Foundation, Vienna, Austria) software package was used to perform all statistical analyses. The Wilcoxon’s test was used to compare two independent non-parametric samples. The predictive accuracy of model was tested via ROC analysis. P-values < 0.05 were considered statistically significant.

**Results**

**Immune cell infiltration analysis**

The expression matrices of GSE85195 and GSE23558 were selected to study immune cell infiltration after removing the batch effect. The proportion of 22 immune cells in each sample was calculated using CIBERSORT. The composition of 22 types of immune cells in each sample was presented in a boxplot (Fig. 1A). The expression pattern of 11 types of immune cells: dendritic cells resting, mast cells resting, mast cells activated, macrophages M0, monocytes, plasma cells, dendritic cells activated macrophages M1, macrophages M2, NK cells activated, and neutrophils were significantly different between the control
and OLK and OSCC groups (P < 0.05). The expression of dendritic cells resting, macrophages M0, macrophages M2, and mast cells resting were also significantly different between the OLK and OSCC groups (P < 0.001).

Differential expression analysis and functional enrichment analysis

Volcano plots were generated to visualize the distribution of expressed mRNAs between the control OLK and OSCC groups (Fig. 2A, E). Red and blue dots in the plots represent significantly upregulated and downregulated mRNAs, respectively. In total, 1432 (688 upregulated and 744 downregulated) and 1256 (497 upregulated and 759 downregulated) DEGs were identified from GSE85195 and GSE23558, respectively. GO and KEGG pathway enrichment analyses of the DEGs were performed (Fig. 2C, D, G, H). We found that many top enriched GO and KEGG pathways overlapped between the two cohorts. For instance, multicellular organism development (GO: 0007275), plasma membrane (GO: 0005886), signal transduction (GO: 0007165), vesicle (GO: 0031982), cytosol (GO: 0005829), animal organ development (GO: 0048513), positive regulation of cellular metabolic process (GO: 0031325), and positive regulation of nitrogen compound metabolic process (GO: 0051173) were shared between the two cohorts for the top 10 GO. ECM-receptor interaction (hsa04512), rheumatoid arthritis (hsa05323), malaria (hsa05144), viral protein interaction with cytokine and cytokine receptor (hsa04061), human papillomavirus infection (hsa05165), focal adhesion (hsa04510), and arachidonic acid metabolism (hsa00590) were shared between the two cohorts for the top 15 KEGG.

Identification of immune-related OLK cancerous genes

Overlapping differential DEGs from the GEO dataset and the 2483 immune-related genes were retrieved from the ImmPort database (https://immport.niaid.nih.gov). Venn diagrams were drawn to obtain the candidate immune-related gene for further analysis (Fig. 3A).

Construction and validation of the immune-related gene OLK cancerous diagnostic model

Sixty-nine candidate genes were integrated into the LASSO regression to identify OLK cancerous immune signatures (Fig. 3B, C). GSE85195 and GSE23558 samples were randomly divided in a 1:1 ratio into training and validation cohorts based on a computer-generated allocation sequence. Then, genes with regression coefficient not equal to 0 in the regression analysis were selected as marker genes for prediction. ROC analysis was used to evaluate the predictive accuracy of the OLK cancerous immune signatures. The area under ROC curve (AUC) in the training and internal validation cohorts were 1 and 0.994 (Fig. 4A), respectively. Finally, a model comprising 16 immune signatures, including ENDOU, CCL14,
SORT1, RORA, RBP7, CD1A, TEK, NPR3, ISG15, TNFRSF12A, ADM, GAST, HLA-A, IL23A, APLN, and SOCS3, was constructed. In addition, the GSE26549 and TCGA cohorts were used to validate the predictive ability of the diagnostic model. The AUC in the training and internal validation cohorts were 0.679 (Fig. 4B) and 0.967(Fig. 4C), respectively.

**Discussion**

OSCC is usually preceded by oral potentially malignant disorders such as OLK. To date, one of the greatest challenges in OLK management is determining the risk of malignant transformation. The regulation of the immune microenvironment plays a key role in the process of malignant transformation of OLK. Few immune-related gene prognostic models have been developed and validated for HNSCC. Given this context, a robust immune-related gene diagnostic model to predict OLK progression to OSCC may help identify immune microenvironment regulations to guide clinical management. In this study, 11 types of immune cells differed between the control OLK and OSCC groups. Dendritic cells resting, macrophages M0, macrophages M2 and mast cells resting were significantly different in the OLK and OSCC groups (P < 0.001) from the two GEO databases (GSE85195 and GSE23558). Our results based on the GEO database samples are in accordance with those of formalin-fixed paraffin-embedded tissue specimen [19] and mouse models of oral carcinogenesis [9] reported in recent years. Our results suggest that increased M2 polarization by macrophage infiltration is associated with the progression of OLK to OSCC. Thus, macrophage infiltration can be used as a predictive marker for malignant transformation. A recent study identified two subtypes of oral premalignant lesion (OPL), namely immune and classical[20], using data generated from 86 OPL cases enrolled in a randomized chemoprevention trial at the University of Texas MD Anderson Cancer Center. Using these data, another study applied the existing signatures derived from HNSCC to OPL and identified only a protective effect of immune-related signatures [21]. It is interesting to note the role of immune-related clusters as being protective against oral malignant transformation.

Based in the above premise, we further screened for differential genes in the control OLK and OSCC samples. In total, 1432 and 1256 DEGs were identified from GSE85195 and GSE23558, respectively. Many top enriched GO and KEGG pathways overlapped between the two cohorts. The overlap of the enrichment of epithelial malignant transformation-related pathways is reasonable. Interestingly, the high enrichment of immune-related pathways, for instance, IL-17 signaling pathway (hsa04657) and TNF signaling pathway (hsa04668), also suggests that immune microenvironment changes may be a hub process in OLK's transformation to OSCC. Overlapping differential DEGs from the GEO dataset and immune-related genes were retrieved from the ImmPort database to obtain candidate immune-related genes for further analysis. Next, 69 immune-related genes were used to construct an immune-related gene OLK cancerous diagnosis model. Finally, we constructed a model comprising 16 genes with the LASSO algorithm. Moreover, we used the GSE26549 and TCGA cohorts to validate the model's predictive ability. The results showed that our immune-related gene diagnostic model was a promising objective diagnosis model to predict cancer risk of OLK to OSCC.
This study had some limitations. First, in contrast to overt HNSCC, very few studies provide data with OLK in public open databases. The model OLK data was only constructed and validated based on data from 3 GEO studies lacked external data verification with a large sample size. Large prospective cohorts are needed to validate our results. Second, using bioinformatics methods for data analysis to obtain diagnostic models required follow-up and in vivo and in vitro experiments to study regulatory mechanisms and pathways.

In conclusion, we constructed a 16-immune-related gene diagnostic model to predict OLK’s transformation to OSCC, which could be applicable in the clinical setting for refining prevention strategies. We believe that this study will contribute to the identification of adequate prognosticators for OLK’s malignant transformation in the future.

Conclusions

The results showed that our immune-related gene diagnostic model was a promising objective diagnosis scheme to predict cancer risk of OLK to OSCC, which could be applicable in the clinical setting for refining prevention strategies.

Abbreviations

DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; GO, Gene Ontology; HNSCC, head and neck squamous cell carcinomas; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; OLK, oral leukoplakia; OSCC, oral squamous cell carcinoma; TCGA, The Cancer Genome Atlas

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here:


Competing interests

The authors declare that they have no competing interests.
Funding

This work was supported by the Natural Science Fundation of Fujian Province (No. 2021J01719).

Authors' contributions

YL and YQ wrote the manuscript. BG, JL and PC revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References


Figures
Immune cell leukocyte subsets between the control, OLK and OSCC groups. (A) Boxplot of the immune cell leukocyte subsets between the control and 297 oral leukoplakia (OLK) and oral squamous cell carcinoma (OSCC) groups. (B, C) Correlation heatmap and heat map of immune cell leukocyte between the control and OLK and OSCC groups.
Figure 2

Differential expression analysis between OSCC vs normal and OSCC vs OLK.
(A, B, C, D) Volcano plot, top20 heatmap, top10 Gene Ontology (GO) circles plot, and top15 Kyoto Encyclopedia of Genes and Genomes (KEGG) dotplot of the significantly differential genes between OSCC vs normal groups. (E, F, G, H) Volcano plot, top20 heatmap, top10 GO circles plot, and top15 KEGG dotplot of the significantly differential genes between OSCC vs OLK groups.

Figure 3

Identification of Immune-related OLK Cancerous Genes.
Venn diagrams of differentially expressed genes (DEGs) from aDEGs (OSCC vs normal), bDEGs (OOSCC
vs OLK) and the 2483 immune-related genes. (B, C) Regression and cross-validation of least absolute shrinkage and selection operator (LASSO) regression analysis.

Figure 4

The predictive accuracy of the prognostic signature in the discovery cohort. (A) The receiver operating characteristic (ROC) curve in the training and internal validation cohorts. (B) The ROC curve in the GSE26549 external validation cohorts. (C) The ROC curve in the The Cancer Genome Atlas (TCGA) external validation cohorts.