AHEAD OF PRINT

Original papers

Immunology; molecular biology; omics

Integrated single-cell transcriptomics, Mendelian randomization, and machine learning identify *CEBPZ* as an immune-related biomarker in oral lichen planus

Yichen Feng[1,A–F], Shenfeng Wang[2,A–F], Ling Guan[2,A–C,E], Linlin Wang[2,A–E], Anqi Zou[2,A–C], Zhankui Xing[1,C,E], Tingying Xiao[1,A–F]

[1] Department of Stomatology, Second Hospital of Lanzhou University, China
[2] School of Stomatology, Lanzhou University, China

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation; D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online) *Adv Clin Exp Med*. 2026

Address for correspondence
Tingying Xiao
E-mail: mike19952025@163.com

Funding sources
None declared

Conflict of interest
None declared

Received on June 8, 2025
Reviewed on July 22, 2025
Accepted on September 30, 2025

Published online on September 14, 2026

Abstract

Background. Oral lichen planus (OLP) is a chronic, immune-mediated oral mucosal disease with complex pathophysiology and potential for malignant transformation. Understanding its molecular basis is critical for the development of precise diagnostic and therapeutic strategies.

Objectives. We aimed to identify key immune-related biomarkers and characterize cellular dynamics in OLP, with a particular focus on the role of *CEBPZ* in disease pathogenesis.

Materials and methods. We analyzed single-cell RNA sequencing (scRNA-seq) data from OLP lamina propria samples (GSE211630) to identify disease-specific T-cell subpopulations using high-dimensional weighted gene co-expression network analysis (hdWGCNA) for oxidative stress-related gene modules. Summary-data-based Mendelian randomization (SMR) integrated FinnGen genome-wide association study (GWAS; 342,499 Europeans) data with Genotype-Tissue Expression (GTEx) expression quantitative trait loci (eQTL) data to identify causal genes. Machine learning (ML) models (least absolute shrinkage and selection operator (LASSO) and convolutional neural network (CNN)) were developed using bulk RNA-seq datasets (GSE52130 and GSE38616) for diagnostic purposes.

Results. We identified OLP-specific T-cell populations (clusters 0, 3, 5, 7, 13, and 15) with enhanced migration inhibition factor (MIF) pathway signaling toward B cells and monocytes. Two oxidative stress-associated modules contained hub genes, including *CEBPZ*. Summary-data-based Mendelian randomization analysis identified 231 OLP-associated genes, with *CEBPZ* uniquely intersecting LASSO-selected markers (odds ratio (OR) = 1.057, 95% confidence interval (95% CI) = 1.013–1.102, p = 0.010). Machine learning models achieved area under the curve (AUC) values ranging from 0.653 to 0.745, with the CNN model reaching a validation accuracy of 0.735. *CEBPZ* showed elevated expression in OLP T cells and correlated with enhanced MIF-(CD74+CXCR4) signaling.

Conclusions. This integrative approach identifies *CEBPZ* as a pivotal biomarker linking genetic susceptibility, oxidative stress, and immune dysregulation in OLP. Our diagnostic models offer promising tools for OLP management.

Key words: oral lichen planus, machine learning, *CEBPZ*, single-cell transcriptome sequencing, Mendelian randomization

Cite as
Feng Y, Wang S, Guan L, et al. Integrated single-cell transcriptomics, Mendelian randomization, and machine learning identify *CEBPZ* as an immune-related biomarker in oral lichen planus [published online as ahead of print on September 14, 2026]. *Adv Clin Exp Med*. 2026. doi:10.17219/acem/211565

DOI
10.17219/acem/211565

Copyright
Copyright by Author(s)
This is an article distributed under the terms of the Creative Commons Attribution 3.0 Unported (CC BY 3.0) (https://creativecommons.org/licenses/by/3.0/)

Highlights

- Single-cell RNA sequencing identifies disease-specific T-cell subpopulations and immune signaling pathways involved in oral lichen planus (OLP).
- Integrative genomic analyses reveal *CEBPZ* as a key biomarker linking oxidative stress, immune dysregulation, and genetic susceptibility in OLP.
- Enhanced MIF-mediated signaling between T cells, B cells, and monocytes suggests a central mechanism in OLP pathogenesis.
- Machine learning diagnostic models based on transcriptomic data show promising accuracy in identifying OLP-associated molecular signatures.

Introduction

Oral lichen planus (OLP) represents a significant clinical challenge in oral medicine, affecting millions of people worldwide with chronic, painful lesions that impair quality of life and carry a risk of malignant transformation. It is a T-cell-mediated chronic inflammatory disease of the oral mucosa with no recognized cause, presenting as plaques, papules, white striations, erosions, erythema, or blisters, primarily affecting the gingiva, tongue, and buccal mucosa.[1,2] The disease affects 1–2% of the adult population.[3] Oral lichen planus diagnosis relies on histological and clinical characteristics; a biopsy is recommended to confirm the diagnosis and rule out malignancy.[1,2] Annual surveillance is strongly recommended, as research indicates that 1.63% of OLP cases develop into oral squamous cell carcinoma (OSCC) within 7 years.[4,5] However, there is no recognized gold standard treatment, and both systemic and local therapeutic approaches are used. Some clinical manifestations are particularly difficult to treat, significantly reducing patients' quality of life.[2,6] Therefore, there is an urgent need to develop more accurate methods for the early diagnosis and treatment of OLP.

Oral lichen planus may be caused by several factors, including hypersensitivity reactions, psychological stress, viral infections, and autoimmune responses.[7,8] Numerous immune cell types have been implicated in the development of OLP. The disease is characterized by liquefaction degeneration of basal keratinocytes mediated by T lymphocytes. $CD8^+$ T lymphocytes produce granzymes and perforin, which can destroy oral mucosal keratinocytes and accelerate epithelial degradation.[9] Natural killer (NK) cells in OLP lesions damage living cells and produce tumor necrosis factor alpha (TNF-α) and interferon gamma (IFN-γ).[10] B cells are also frequently observed in OLP lesions.[11] The basement membrane of the oral mucosa in OLP may be damaged by particles and inflammatory mediators released by mast cells, as well as by matrix metalloproteinases (MMPs) produced by T cells.[12] Within complex microenvironments, immune cells interact to produce inflammatory mediators, such as chemokines and cytokines, which contribute to the development and progression of OLP. However, the complete single-cell characteristics of the immune response in OLP have not yet been fully characterized.[13]

Recent advances in single-cell RNA sequencing (scRNA-seq) have revolutionized our understanding of cellular heterogeneity in disease states. While previous scRNA-seq studies identified differences in T-cell composition between OLP and normal tissues,[14–16] our study integrates multiple cutting-edge analytical approaches. We combined scRNA-seq with causal inference using summary data-based Mendelian randomization (SMR), network analysis via high-dimensional weighted gene co-expression network analysis (hdWGCNA), and machine learning (ML) to create a comprehensive framework for biomarker discovery and diagnostic tool development. This multimodal approach allows us not only to identify disease-associated cellular changes but also to establish their causal relationships and clinical utility.

Summary data-based Mendelian randomization is a powerful integrative analytical technique that uses genetic variants as instrumental variables to assess causal relationships between gene expression and disease outcomes.[17] By combining summary-level expression quantitative trait loci (eQTL) data with genome-wide association study (GWAS) results, SMR identifies genes whose expression levels are causally linked to disease risk while minimizing the influence of confounding factors.

Using scRNA-seq analysis, we identified a distinct increase in specific T-cell subgroups in patients with OLP, designated as OLP-specific T cells (OLP_T cells). Through hdWGCNA, we identified T-cell gene modules associated with oxidative stress and linked to immune profiles. We developed diagnostic models using ML and constructed a gene–immune convolutional neural network deep learning model incorporating 10 immune cell types and 9 key genes. Summary data-based Mendelian randomization analysis of GWAS and eQTL data further identified risk genes, which were subsequently evaluated through scRNA-seq analysis and cellular communication analysis.

Objectives

This study aimed to identify key immune-related biomarkers and characterize the cellular dynamics involved

in OLP, with a particular focus on understanding the role of *CEBPZ* in disease pathogenesis, by integrating single-cell transcriptomics, genetic causal inference, and ML approaches to uncover novel diagnostic markers and therapeutic targets.

Materials and methods

Ethical approval and consent to participate

This study did not involve direct human or animal experimentation. All analyses were performed using publicly available datasets, including scRNA-seq data from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo) (GSE211630, GSE52130, and GSE38616), GWAS summary data from the FinnGen database (https://www.finngen.fi), and eQTL data from the Genotype-Tissue Expression (GTEx) database (https://gtexportal.org). Therefore, no ethical approval or informed consent was required. All data were handled in accordance with the ethical guidelines and data-use policies of the respective repositories.

Data acquisition

The GEO database provided the scRNA-seq dataset GSE211630, which was processed using the standard Seurat methodology.[18] The Harmony algorithm was used to eliminate batch effects between samples. Bulk RNA-seq datasets GSE52130 and GSE38616 were also obtained from GEO.[19,20] For SMR analysis, GWAS summary statistics were obtained from the FinnGen database (release 8), specifically using endpoint K11_LICHEORAL, comprising 342,499 samples of European ancestry with 20,169,554 single-nucleotide polymorphisms (SNPs). Esophageal mucosa eQTL summary data were obtained from the GTEx database (version 8) and selected as a proxy for oral squamous epithelium because of shared squamous epithelial characteristics and the lack of adequately powered oral tissue eQTL datasets. No sample overlap existed between the FinnGen and GTEx cohorts, and all genomic coordinates were harmonized to the GRCh38/hg38 reference build. Reference genotype data for linkage disequilibrium (LD) calculations were obtained from the Phase 3 European panel of the 1000 Genomes Project (EUR, n = 503). Protein interaction data were obtained from the STRING database.[21]

Single-cell sequencing data processing

Dataset GSE211630 was processed using the standard Seurat pipeline, with quality control parameters including a ribosomal gene percentage threshold of 10%, a mitochondrial gene proportion threshold of 20%, and between 500 and 3,000 detected genes per cell. Highly variable genes were identified using Seurat's FindVariableFeatures function.[22] Cell clustering was performed using the FindNeighbors and FindClusters functions,[23] and cell types were manually annotated using established cell markers.[24]

The pipeline of high dimensional WGCNA

The hdWGCNA package extends traditional WGCNA to single-cell data by accounting for sparsity and high dimensionality.[25] The pipeline involves 5 key steps: preprocessing to eliminate batch effects while preserving biological variation; construction of a gene co-expression network using metacells to overcome sparsity; identification of gene modules with eigengene calculation; module preservation analysis to evaluate robustness; and functional enrichment analysis to determine associated biological pathways.

Pseudotime analysis

Pseudotime analysis was performed using the Monocle2 R package (https://github.com/MaxMeieran/monocle2) to investigate the potential effects of blue and brown module genes on T-cell development in the OLP microenvironment.[26–29] The plot_cell_trajectory function was used to depict the cell differentiation trajectory, and the DDRTree method implemented through the reduceDimension function was used to calculate differentiation states.[30,31] Additionally, the Velocyto package (https://github.com/velocyto-team/velocyto.py) was used to infer cell developmental vectors based on the balance between spliced and unspliced mRNA, with the developmental stage linked to the quantile polarization of cell principal component values.[32]

Cell–cell interaction analysis

Ligand–receptor-mediated interactions and intercellular communication networks were assessed using the R package CellChat (https://github.com/sqjin/CellChat). The volume and intensity of cell-to-cell communication were graphically represented in Fig.1. The migration inhibition factor (MIF) signaling pathway, which mediates inflammatory responses through the receptors CD74, CXCR4, and CD44, was specifically examined because of its role in chronic inflammation. The principal signaling inputs and outputs of each OLP cell subpopulation were assessed using CellChatDB.[33]

Functional enrichment analysis

Potential interaction networks among target module proteins were identified using GeneMANIA.[34,35] Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the Metascape platform,[36] with a p-value threshold of 0.05.

Immune infiltration analysis based on individualized CIBERSORTx signature or MCP_counter

CIBERSORTx was used to deconvolve complex gene expression data into cell type-specific expression patterns.[37] An OLP-specific gene expression signature matrix was generated from GSE211630 for the deconvolution of bulk RNA-seq datasets. The MCP-counter algorithm was used to assess the abundance of 10 cell types.[38]

RNA-seq bulk NMF

Patients with OLP were divided into subgroups using non-negative matrix factorization based on blue and brown module gene expression levels.[39] The cophenetic coefficient was used to determine the optimal number of clusters.

Machine learning and deep learning

Key diagnostic genes were identified using univariate logistic regression. Least absolute shrinkage and selection operator (LASSO) regression was used for variable selection.[40] Ten ML models were evaluated using the mlr3 package (https://cran.r-project.org/web/packages/mlr3/index.html): KNN, LR, LDA, NB, Ranger, ABESS, Glmnet, SVM, XGBoost, and Multinom.[41] GSE52130 served as the training set and GSE38616 served as the validation set, with performance assessed using 10 iterations of five-fold cross-validation. For deep learning, images of 10 immune cell types and 9 genes were generated using the formula N_{j,i} = immune_i/gene_j to investigate the association between immune cells and target genes. Convolutional neural network (CNN) models were developed using the Keras and TensorFlow frameworks.[42,43]

Summary-data-based Mendelian randomization analysis

Comprehensive SMR analysis was conducted using SMR software v. 1.3.1 (https://yanglab.westlake.edu.cn/software/smr) with FinnGen GWAS summary statistics, GTEx esophageal mucosa eQTL data, and the 1000 Genomes European reference panel for LD calculation (workflow shown in Supplementary Fig. 1). All data were harmonized to GRCh38/hg38, alleles were aligned, and palindromic SNPs with a minor allele frequency (MAF) of 0.4–0.6 were removed. Instrument selection used a cis-eQTL window of ±1 Mb with clumping parameters of $r^2 < 0.1$ within 500 kb. An eQTL was considered significant when the p-value of the top associated cis-acting single-nucleotide polymorphism (cis-SNP) was $<5.00 \times 10^{-8}$. The significance threshold for SMR analysis was determined using Bonferroni correction (0.05/N). Oral lichen planus-associated risk genes were defined as those with PSMR (the p-value from SMR analysis indicating a pleiotropic association) <0.05 and PHEIDI (the p-value from the Heterogeneity in Dependent Instruments test) >0.05. The HEIDI test examines whether an association is due to a shared causal variant (PHEIDI >0.05) or linkage (PHEIDI ≤0.05). Default settings were maintained for the minimum number of instruments (≥3 SNPs) and instrument strength (F-statistic >10).

Statistical analyses

Three groups were compared using the Kruskal–Wallis test, and 2 groups were compared using the Wilcoxon test. Pearson's correlation analysis was used to investigate the relationships between 10 immune cell types and 9 key genes. Statistical significance was set at $p < 0.05$. All analyses were performed using R v. 4.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

The change in the composition of T cell subtypes in OLP lamina propria

We acquired the scRNA-seq dataset GSE211630 and used Harmony integration, dimensionality reduction, and clustering to analyze the OLP-specific cellular composition (Supplementary Fig. 2). The Uniform Manifold Approximation and Projection (UMAP) scatter plot showed the distribution of cell subtypes in the OLP lamina propria (LP) (Fig. 1A), while the stacked bar chart displayed the altered proportions of these cell populations (Fig. 1B). T cells showed the most pronounced shift between OLP and normal tissues (Fig. 1A,B). We focused on T cells and performed dimensionality reduction and clustering of T-cell subpopulations. T-cell clusters 0, 3, 5, 7, 13, and 15 showed significant increases in the OLP LP (Fig. 1C,D), which were not observed in normal tissues. Given that these adjacent clusters were distant from other T-cell subpopulations in UMAP space and exhibited specific upregulation in OLP, we designated them as OLP_T cells.

Cell–cell communication analysis revealed that OLP_T cells exhibited more numerous and stronger interactions with other cell types, including B cells, other T cells, and monocytes (Fig. 1E,F). We further investigated the specific pathways through which certain cell types interacted more strongly with one another. OLP_T cells demonstrated enhanced communication through MIF-related pathways compared with other T cells, transmitting more signals to B cells and monocytes (Supplementary Fig. 3A,B). The MIF pathway mediates inflammatory responses through its receptors CD74, CXCR4, and CD44 and plays a crucial role in chronic inflammation. Overall, scRNA-seq

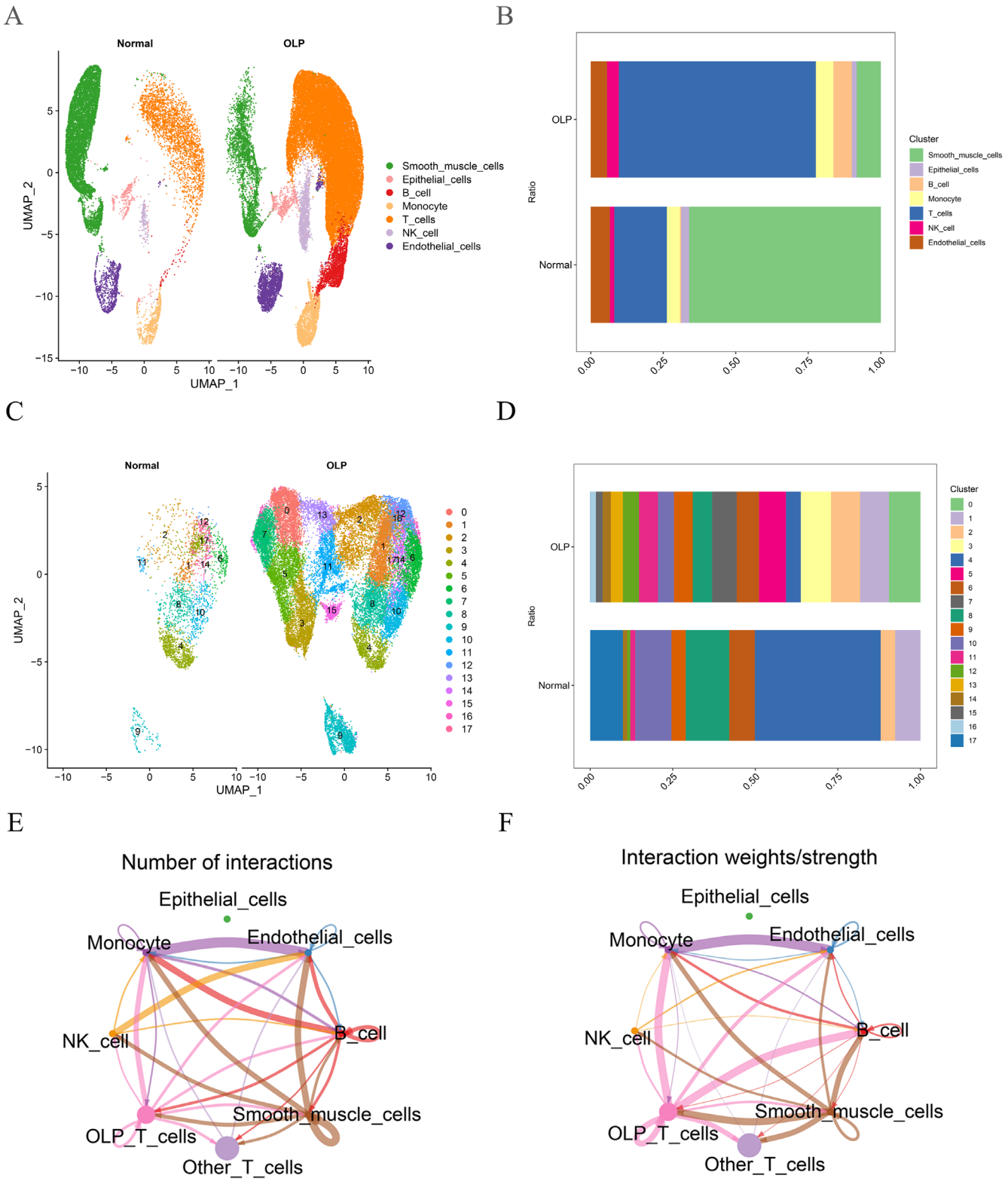

Fig. 1. T-cell composition and interactions in the oral lichen planus (OLP) lamina propria (LP). A. Uniform Manifold Approximation and Projection (UMAP) scatter plot showing the distribution of cell subtypes in the OLP LP; B. Stacked bar chart depicting the proportion of each cell type; C. UMAP scatter plot showing the distribution of T-cell subtypes; D. Stacked bar chart showing T-cell subtype proportions; E,F. Circle plots illustrating the number and strength of cell–cell interactions between T-cell subtypes and other cell types, with colors indicating the source cell populations

NC – normal control; OLP_T – OLP-specific T cells; NK cells – natural killer cells.

analysis revealed an increase in OLP-specific T-cell subtypes in the OLP LP. These cells exhibited close interactions with smooth muscle cells, B cells, and monocytes, suggesting that this population of OLP-specific T-cell subtypes may be involved in the immune-inflammatory process.

hdWGCNA revealed that OLP_T cells were characterized by the blue and brown modules

We examined OLP_T cells using hdWGCNA. A power value of 9 was selected after considering the scale-free topology model fit, median connectivity, and mean connectivity (Fig. 2A). Four modules were identified (Fig. 2B). The blue and brown modules showed specific expression in OLP_T cells (Fig. 2C), with a positive correlation observed between the modules (Fig. 2D). Hub genes, including *CYTIP*, *RPL19*, *SELENOK*, *RPS28*, *FOSL2*, and *CXCR4*, were identified (Fig. 2E). Module correlations are shown in Fig. 2F.

Gene enrichment analysis revealed that the blue and brown module genes were primarily involved in oxidative phosphorylation, adenosine triphosphate (ATP) synthesis coupled with electron transport, and nicotinamide adenine dinucleotide (NADH) dehydrogenase activity (Fig. 3A). KEGG enrichment analysis showed that these genes participated in cellular responses to stress, implying a strong association with oxidative stress (Fig. 3B). This association between *CEBPZ* and oxidative stress is particularly relevant because oxidative stress drives inflammatory responses in chronic diseases such as OLP through the production of reactive oxygen species (ROS), which can damage cellular components and perpetuate the inflammatory cycle. Further analysis of cell–cell communication confirmed strong interactions between OLP_T cells and B cells and monocytes, particularly through the MIF signaling pathway (Fig. 3C–E; Supplementary Fig. 4A). OLP_T cells demonstrated higher signal transmission than other cell types, except in the CXCL signaling pathway (Supplementary Fig. 4B,C), where OLP_T cells were the most potent signal receivers.

Pseudotime analysis reveals early developmental stages of OLP_T cells

Pseudotime analysis revealed that OLP_T cells (clusters 0, 3, 5, 7, 13, and 15) were predominantly in stages 1 or 2 of T-cell development (Fig. 4). This finding suggests that the pathological milieu of the OLP LP may exert an educational effect on OLP_T cells, potentially priming them toward a pro-inflammatory phenotype early in their developmental trajectory. The enrichment of OLP_T cells in early developmental stages contrasts with the more differentiated state of T cells in normal tissues, supporting the hypothesis that the OLP microenvironment actively shapes T-cell differentiation toward pathogenic phenotypes.

Development of a machine learning diagnostic model based on OLP_T cell-specific genes

The LASSO regression analysis reduced the candidate genes to 9: *SRRM1*, *RPS11*, *CEBPZ*, *RAB9A*, *KLF3*, *CNOT6L*, *RPL18*, *ARID4B*, and *IRF2BP2* (Fig. 5A,B). The UMAP diagram in Supplementary Fig. 5 displays the differential expression of these 9 genes between normal and OLP tissues. These selected genes were incorporated into multiple ML models, including support vector machine (SVM), naïve Bayes, K-nearest neighbors (KNN), linear discriminant analysis (LDA), and XGBoost. GSE52130 served as the training set and GSE38616 served as the validation set. Model performance was assessed using 10 iterations of five-fold cross-validation. Logistic regression (Logreg), Adaptive Best Subset Selection (ABESS), and LDA showed balanced precision and sensitivity in differentiating OLP tissues from normal tissues (Fig. 5C,D). The final ML framework included these 3 techniques, which demonstrated good predictive accuracy in the external validation set (area under the ROC curve (AUC) = 0.745, 0.694, and 0.653, respectively; Supplementary Fig. 6A–C).

Oxidative stress genes in OLP_T cells correlate with immune profiles and enable OLP patient subtyping

Using non-negative matrix factorization (NMF) based on oxidative stress genes, patients with OLP were stratified into 2 subgroups to investigate the clinical significance of these genes in patient classification (Fig. 6A). The consensus matrix heatmap displayed OLP patient subtypes when k = 2 (Fig. 6B). Further analysis at the bulk RNA-seq level revealed strong associations between immune infiltration and patient heterogeneity.

Using CIBERSORTx to generate an OLP-specific signature derived from scRNA-seq data, we determined the abundance of cellular components in the OLP LP and assessed correlations with OLP_T-specific oxidative stress gene expression patterns. Most of these genes showed strong associations with different cell types, suggesting their involvement in modulating the OLP microenvironment. Notably, most oxidative stress genes specific to OLP_T exhibited robust positive correlations with lymphocyte infiltration (Fig. 6C). Subtype 2 showed stronger lymphocyte activation (Fig. 6D) and higher proportions of most immune cell populations (Supplementary Fig. 7A), further supporting stronger immune activation in this subtype.

Establishment of the gene–immune convolutional neural network model

Considering the established relationship between the immunological landscape and oxidative stress genes in OLP_T cells, we developed a gene–immune CNN classifier to create a general diagnostic model while minimizing batch effects. We generated a unique gene–immune heatmap (9 units wide × 10 units long), in which the value of each square for a given OLP patient represents the proportion of a specific immune cell divided by the expression level of the indicated

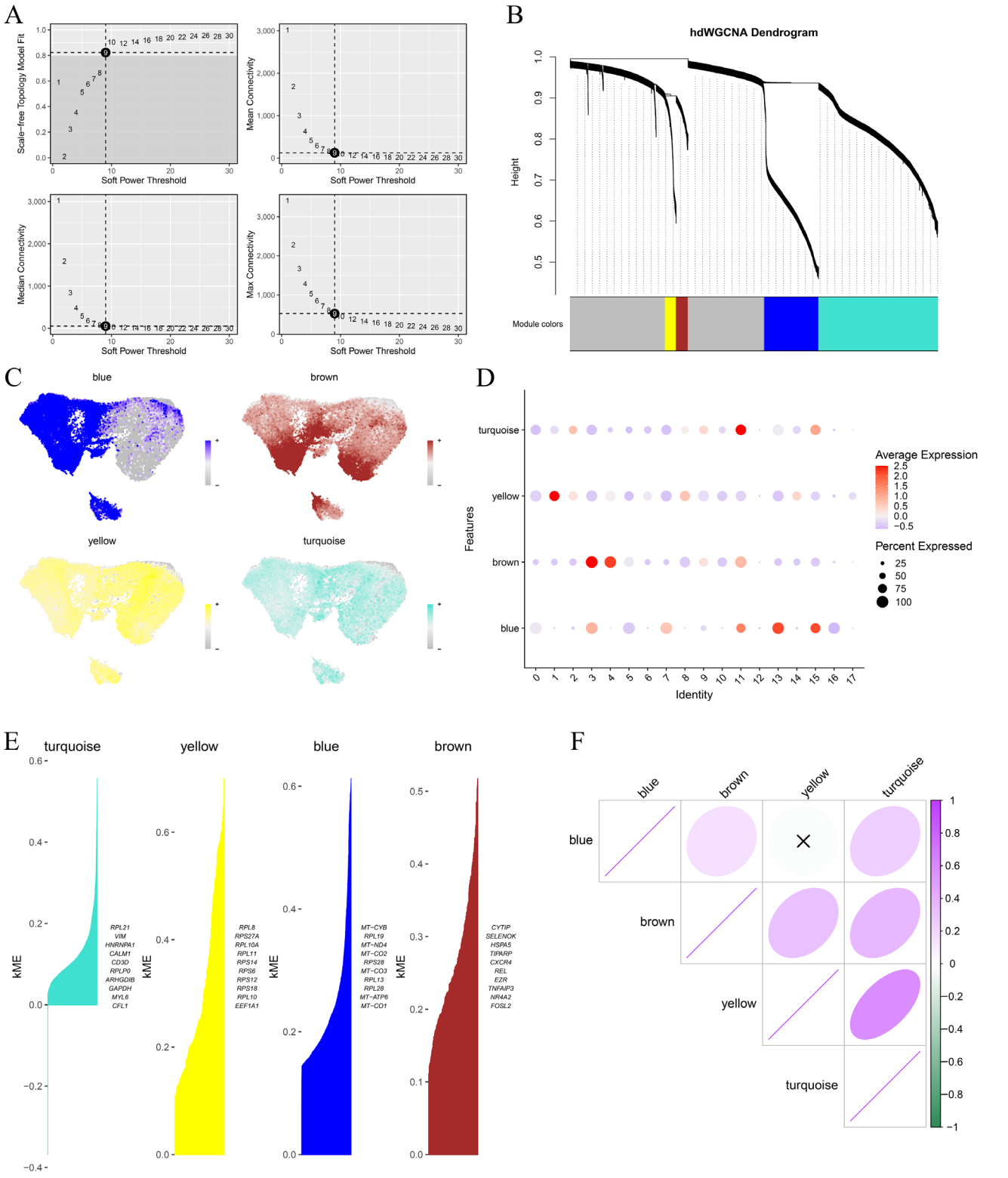

Fig. 2. High-dimensional weighted gene co-expression network analysis (hdWGCNA) of oral lichen planus (OLP)-specific T cells. A. Selection of power value 9 for scale-free network topology; B. Clustering of highly variable genes into 4 modules; C. Feature plots displaying gene expression scores for the blue and brown modules in OLP_T cells; D. Bubble plot showing module scores across T-cell subtypes. E. Hub genes ranked according to eigengene connectivity (kME); F. Correlation matrix showing relationships between modules

OLP_T – OLP-specific T cells.

gene (Fig. 7A). GSE52130 served as the training set, and GSE38616 served as the validation set. After 300 epochs of training (Fig. 7B), the CNN model demonstrated strong predictive accuracy in both datasets (Fig. 7C,D; training AUC = 1.000, testing AUC = 0.735), suggesting potential for broader applications in OLP diagnosis.

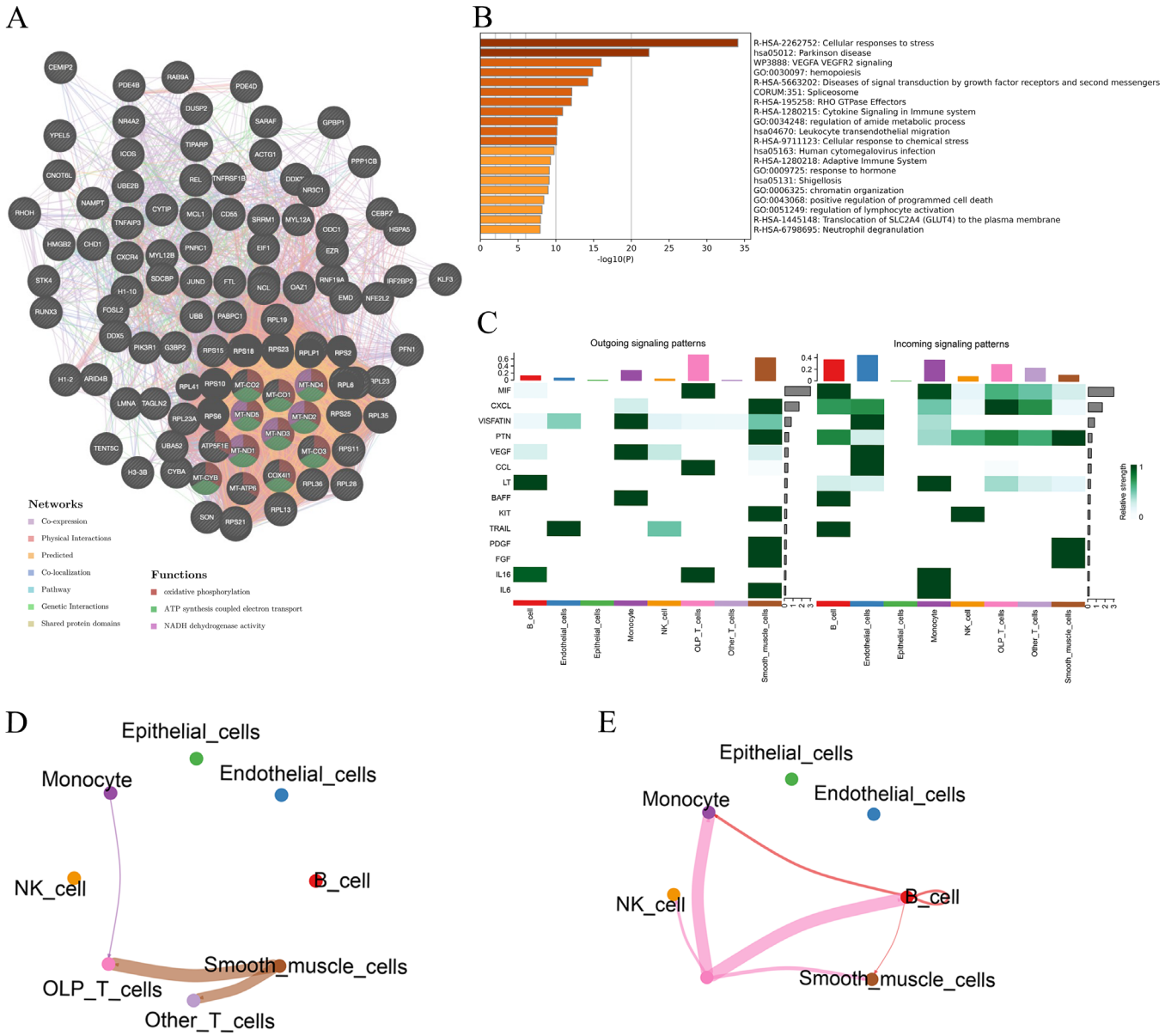

Fig. 3. Functional enrichment and cell interactions of the blue and brown modules. A. Protein–protein interaction network of genes in the blue and brown modules; B. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis linking genes to oxidative stress and cellular responses; C. Heatmap of outgoing and incoming signaling patterns across cell types; D,E. Circle plots showing the number and strength of cell–cell communication between OLP_T cells and other cell types

OLP_T – OLP-specific T cells; NK cells – natural killer cells.

SMR analysis identifies OLP-related gene targets

SMR analysis of GWAS and eQTL data identified regulatory genes causally associated with OLP using significance thresholds of PSMR < 0.05 and PHEIDI > 0.05. A total of 231 genes were found to be pleiotropically linked to OLP (Supplementary Table 1). Among these, the top 10 genes are highlighted because of their significant associations (Table 1). Notably, *CEBPZ* was the only gene that overlapped with the LASSO regression results (Fig. 8A). CCAAT/enhancer-binding proteins (CEBPs), including *CEBPZ*, have been reported to regulate cytokine production and macrophage function, thereby enhancing inflammatory responses.[44,45] Our findings showed a correlation between OLP risk and increased *CEBPZ* expression (odds ratio (OR) = 1.057, 95% confidence interval (95% CI): 1.013–1.102, p = 0.010; Fig. 8B).

We acknowledge that esophageal mucosa eQTLs served as proxies for oral epithelium because of the lack of adequately powered oral tissue eQTL data. Although both tissues share squamous epithelial characteristics, tissue-specific regulatory mechanisms may differ. Future studies should validate these findings using oral tissue-specific eQTLs. Sensitivity analyses using whole-blood eQTLs or other immune-relevant tissues would strengthen causal inferences.

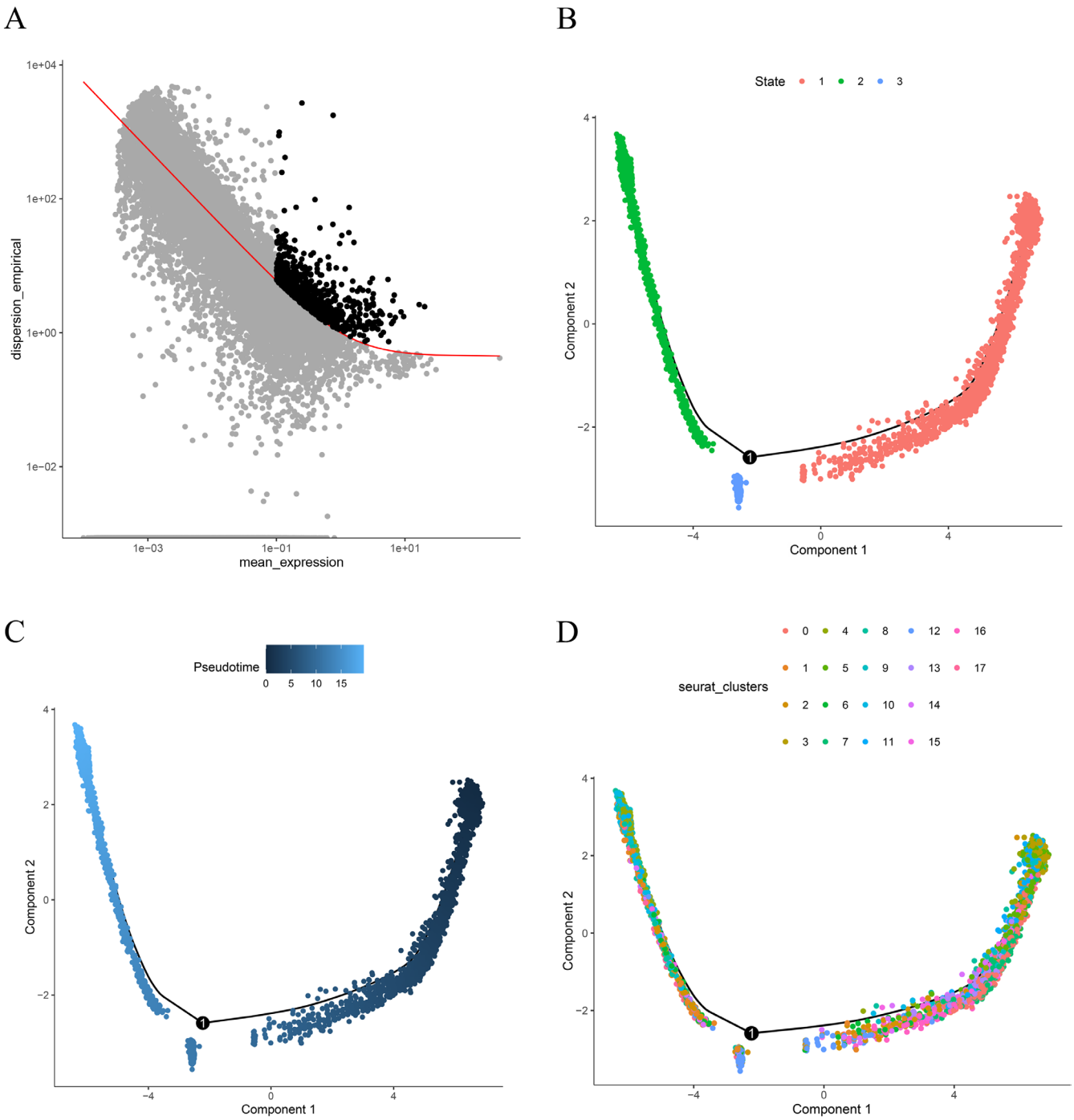

Fig. 4. Pseudotime analysis of T-cell subtypes in oral lichen planus (OLP). A,B. Trajectory plots showing developmental states of OLP_T cell clusters characterized by the blue and brown modules; C. Monocle time-series plot of T-cell development; D. Time-series plot of T-cell subtype differentiation stages

OLP_T – OLP-specific T cells.

Expression levels of SMR outcome gene in oral lichen planus

CEBPZ expression was significantly higher in OLP patients than in healthy individuals across all cells and T cells (Fig. 9A,B). The UMAP plot and enrichment analysis demonstrated *CEBPZ* expression across all OLP cell populations (Fig. 9C,D). Expression was increased in T cells, epithelial cells, and smooth muscle cells (Supplementary Fig. 8A). Unsupervised developmental inference analysis indicated that *CEBPZ* expression increased with cell differentiation (Supplementary Fig. 8B–D).

Given that OLP is a chronic T-cell-mediated inflammatory disease, and consistent with our previous analyses, we focused on *CEBPZ* expression in T cells. We divided T cells into *CEBPZ*+ (high-expression) and *CEBPZ*– (low-expression) groups according to their expression levels. Using CellChat to calculate receptor–ligand

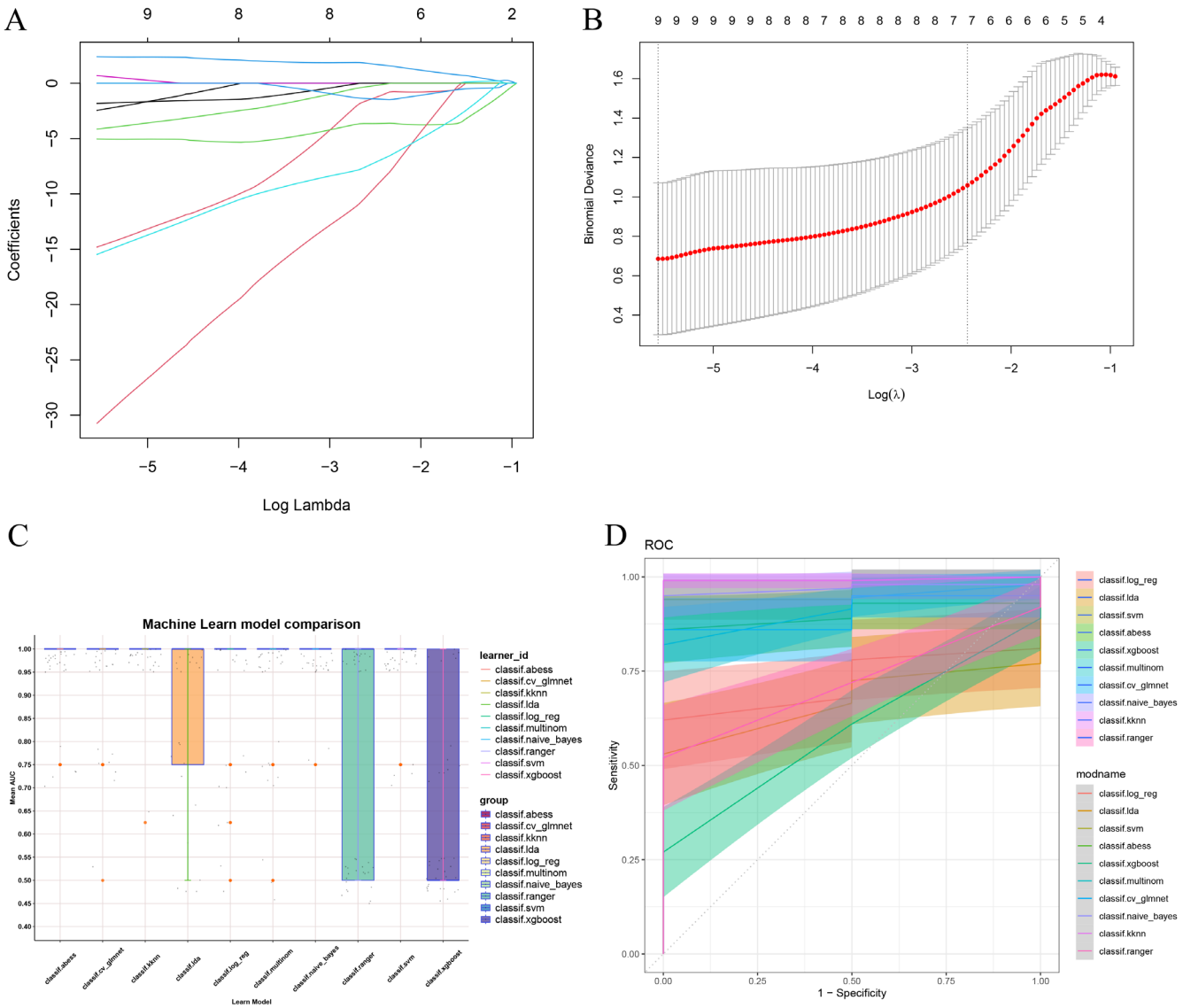

Fig. 5. Machine learning (ML) diagnostic model for oral lichen planus (OLP). A. Least absolute shrinkage and selection operator (LASSO) regression coefficient plot; B. Penalty parameter selection plot; C. Average area under the curve (AUC) from cross-validation for ML models; D. Receiver operating characteristic (ROC) curves with confidence intervals

Table 1. Top genes identified using summary data-based Mendelian randomization (SMR) analysis for oral lichen planus

No.	ProbeID	Gene	P_SMR	P_HEIDI
61	ENSG00000138792	*ENPEP*	0.0007440228	0.5696891
81	ENSG00000135587	*SMPD2*	0.000723646	0.4871803
71	ENSG00000183323	*CCDC125*	0.0009409793	0.5892023
127	ENSG00000121691	*CAT*	0.0008846088	0.402772
86	ENSG00000010030	*ETV7*	0.001591347	0.5681294
123	ENSG00000164978	*NUDT2*	0.001405524	0.4440702
85	ENSG00000204420	*C6orf25*	0.002674887	0.5592639
217	ENSG00000104852	*SNRNP70*	0.002695897	0.732081
224	ENSG00000099949	*LZTR1*	0.003013516	0.9810746
31	ENSG00000115816	*CEBPZ**	0.02483984	0.7211582

CEBPZ was the only gene overlapped with least absolute shrinkage and selection operator (LASSO) regression results (see Fig. 9A). P_SMR – p-value from SMR analysis indicating pleiotropic association with OLP; P_HEIDI – p-value from the Heterogeneity in Dependent Instruments (HEIDI) test assessing heterogeneity.

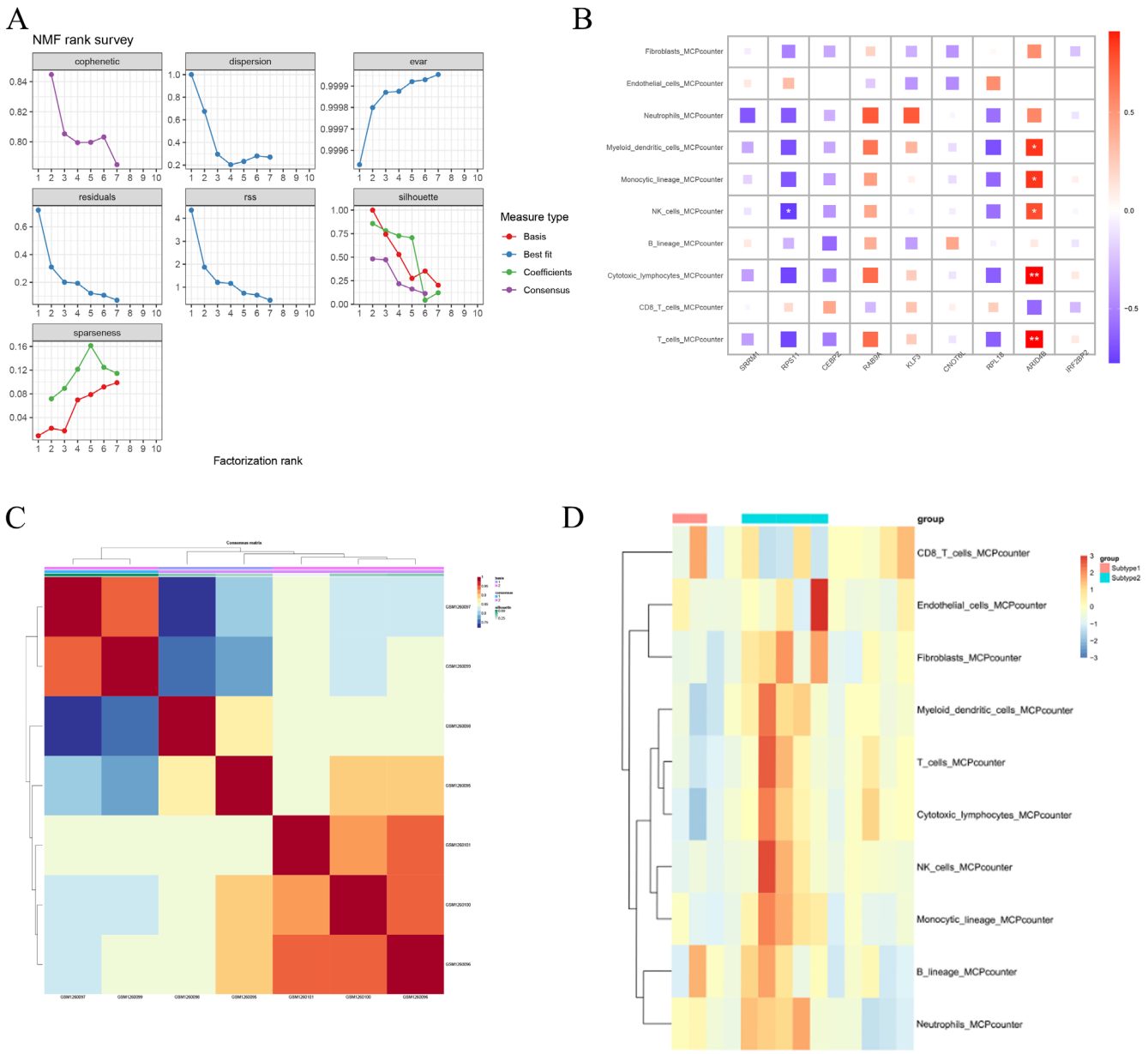

Fig. 6. Oral lichen planus (OLP) patient subtyping and immune infiltration. A. Non-negative matrix factorization (NMF)-based clustering of OLP patients; B. Consensus matrix heatmap (k = 2); C. Correlation heatmap between 9 genes and immune cell types; D. Microenvironment Cell Populations (MCP)-counter comparison of immune cell infiltration between subgroups

interaction strength, we found that signaling pathways between *CEBPZ*+ T cells and monocytes were significantly enhanced compared with those between *CEBPZ*– T cells (Fig. 10A). The enhanced interactions involving "MIF-(CD74+CXCR4)" and "MIF-(CD74+CD44)" were consistent with the enhanced communication observed in OLP_T cells through *MIF*-related pathways (Fig. 10B). According to recent studies, *MIF*, *CD74*, *CXCR4*, and *CD44* are associated with inflammation. *MIF* mediates the chronicity and potential malignant transformation of OLP. A protein–protein interaction (PPI) network generated using the STRING database identified interactions among *CEBPZ*, *MIF*, *CD74*, *CD44*, and *CXCR4* (Fig. 10C).

Discussion

Oral lichen planus is an immune-mediated inflammatory disease that primarily affects the tongue, gingiva, and buccal mucosa. Research indicates that 1.63% of OLP cases develop into OSCC within 7 years, leading the World Health Organization (WHO) to classify OLP as a potentially malignant condition. Early diagnosis and ongoing monitoring are crucial for the management of OLP. However, current diagnostic methods lack precision and are noninvasive.[46] The development of high-throughput sequencing technology has enabled rapid and precise analysis of gene expression differences between healthy and diseased tissues.[47] However, using bulk sequencing techniques to determine

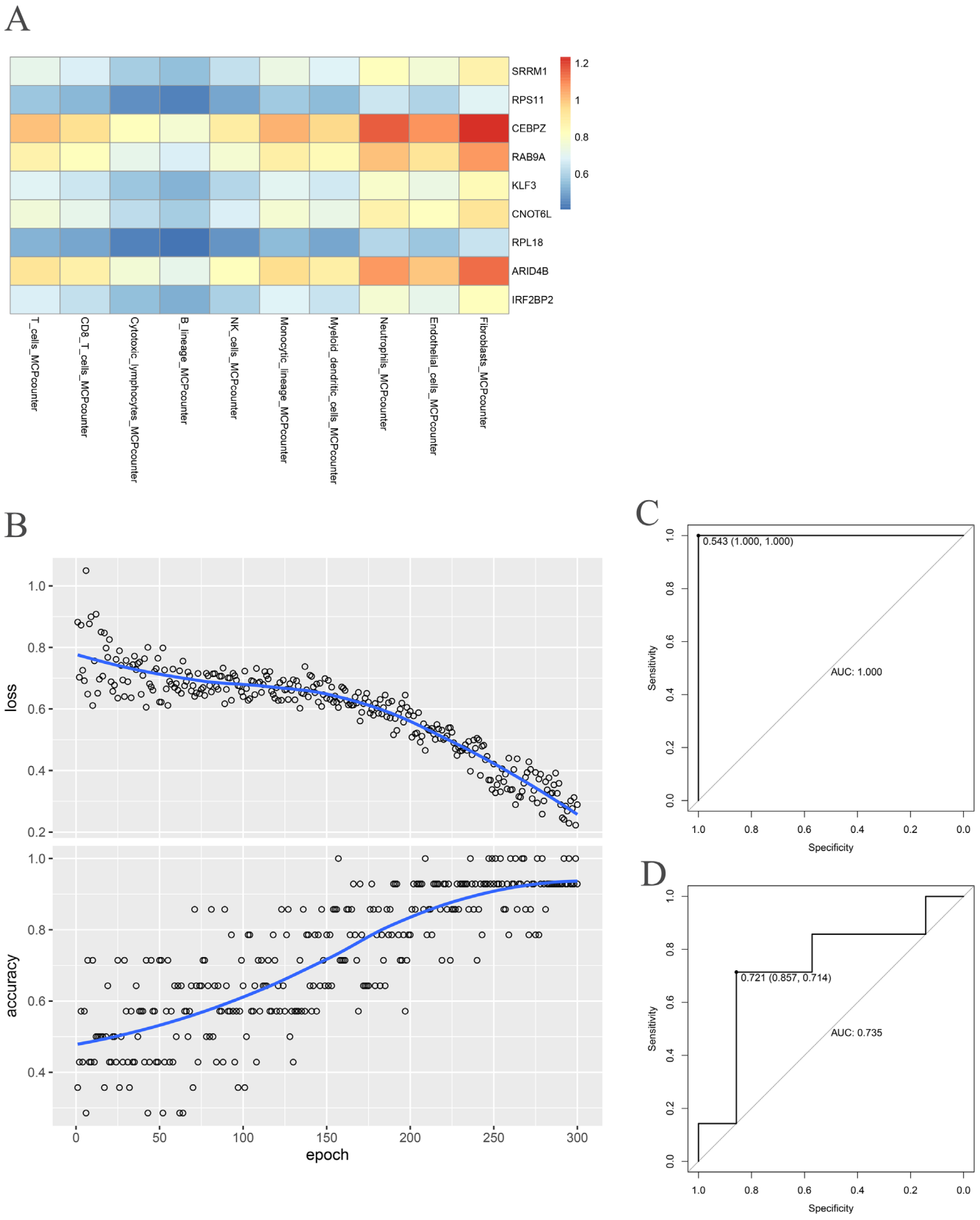

Fig. 7. Gene–immune convolutional neural network (CNN) model. A. Schematic representation of the CNN architecture integrating 9 genes and 10 immune cell types; B. Training curve over 300 epochs; C,D. Model performance in the training (area under the curve (AUC) = 1.000) and validation (AUC = 0.735) datasets

cellular composition and the involvement of specific subpopulations in diseased tissues remains challenging.[48] Single-cell sequencing technology addresses these limitations by enabling detailed analysis of cellular heterogeneity and communication within diseased tissues.[49] SMR analysis helps identify disease-associated genes while minimizing

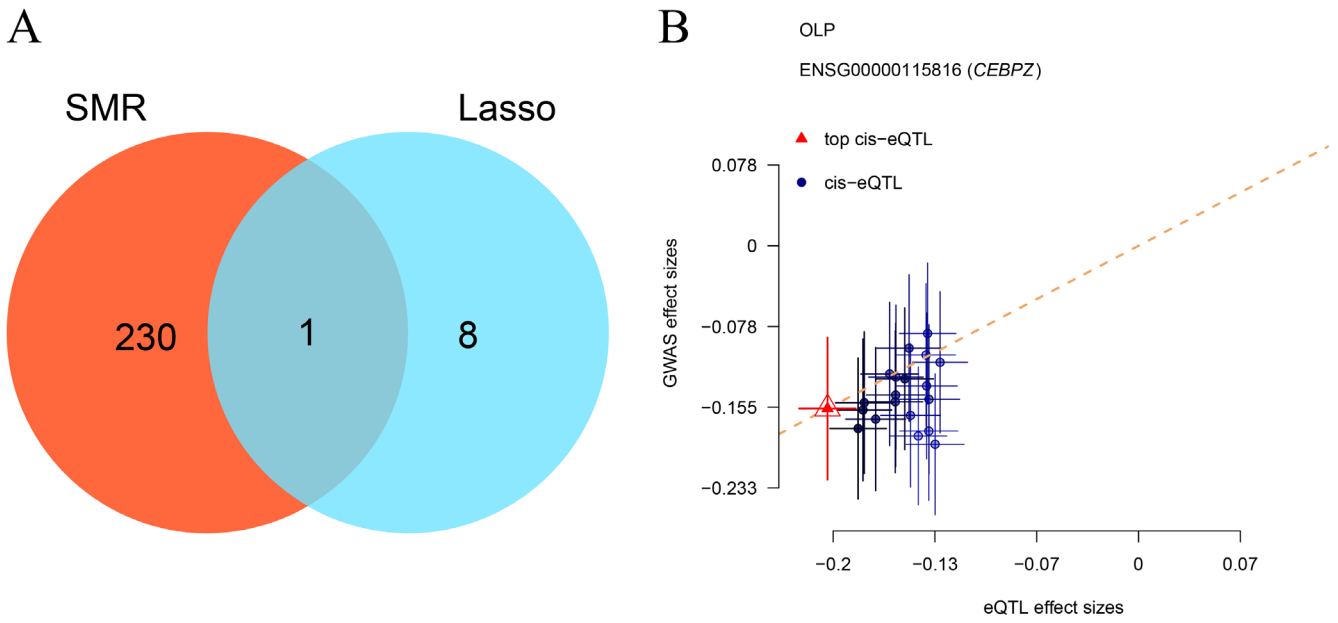

Fig. 8. Summary data-based Mendelian randomization (SMR) analysis of oral lichen planus (OLP)-associated genes. A. Venn diagram showing the overlap between SMR and lasso least absolute shrinkage and selection operator (LASSO) results; B. Mendelian randomization locus plot showing the association between *CEBPZ* expression and OLP risk (odds ratio (OR) = 1.057, 95% confidence interval (95% CI) = 1.013–1.102, p = 0.010)

confounding factors and providing important information regarding causality, thereby offering a dynamic view of disease progression.[50]

Studies have indicated that OLP has a major immunological component.[16] The current study found TNF-α expression throughout the subepithelial T-cell infiltrate and TNF receptor 1 (TNFR1) expression in basal and suprabasal epithelial cells in OLP. Oral lichen planus is characterized by a balance between pro-inflammatory (IFN-γ, TNF-α) and immunosuppressive (transforming growth factor beta 1 (TGF-β1)) cytokines. Insufficient immunosuppression associated with decreased TGF-β1 activity may promote hyperactive immune responses in OLP. Conversely, high TGF-β1 activity may impair antitumor immunity and thereby promote OLP carcinogenesis.[51,52]

Corticosteroids, used as first-line treatment for OLP, primarily suppress the $CD4^+$ Th-cell subgroup and downregulate NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) signaling, thereby reducing IFN-γ levels and exerting immunomodulatory effects.[53,54] Tacrolimus, a second-line treatment, inhibits cytokine production and T-cell activation, reduces Treg proliferation, inhibits mast cell activation, and affects Langerhans cell activity.[55–57] Studies based on the immune landscape of OLP have identified potential therapeutic targets, including β-integrins and ICAM-1 inhibitors for controlling adhesion molecule expression, RANTES inhibitors for stabilizing mast cells, Fas cell surface death receptor (FAS) or Fas ligand (FASL) inhibitors, and TNF-α receptor inhibitors for reducing keratinocyte apoptosis.[58,59] However, further research is required before these approaches can be translated into clinical practice based on the OLP immune microenvironment.

Based on oral scRNA-seq transcriptomic features, we developed a 9-gene classifier that may assist in OLP diagnosis and highlights the potentially important role of specific T-cell subsets characterized by pronounced oxidative stress in OLP pathophysiology. The expression patterns of these 9 genes correlate with oxidative stress, inflammation-related pathways, and treatment-related responses in T cells.[2,60,61] The application of scRNA-seq has revealed the heterogeneity of immune cells in OLP and their interactions with epithelial cells, which is crucial for understanding disease progression and developing targeted therapies.

CCAAT/enhancer-binding proteins including CEBPA, CEBPB, CEBPD, CEBPE, CEBPG, and CEBPZ, play important roles in physiological and pathological processes.[62–64] *CEBPZ* functions as either an activator or a suppressor depending on cellular context and is associated with programmed cell death, cell-cycle arrest, and cellular stress.[65] Strong correlations have been reported between *CEBPZ* expression and methylation in acute myeloid leukemia,[66] and overexpression has also been observed in esophageal squamous cell carcinoma.[62] Nearly all 9 LASSO-selected genes were strongly associated with oxidative stress, inflammation, and OLP pathogenesis. The intrinsic OLP score was markedly elevated according to single-sample Gene Set Enrichment Analysis (ssGSEA) based on genes from the blue and brown modules, suggesting worsening immunological dysregulation and inflammation within the oral microenvironment. Strong correlations between the expression patterns of blue- and brown-module genes and immune infiltration suggest that these genes may participate in immune regulation.

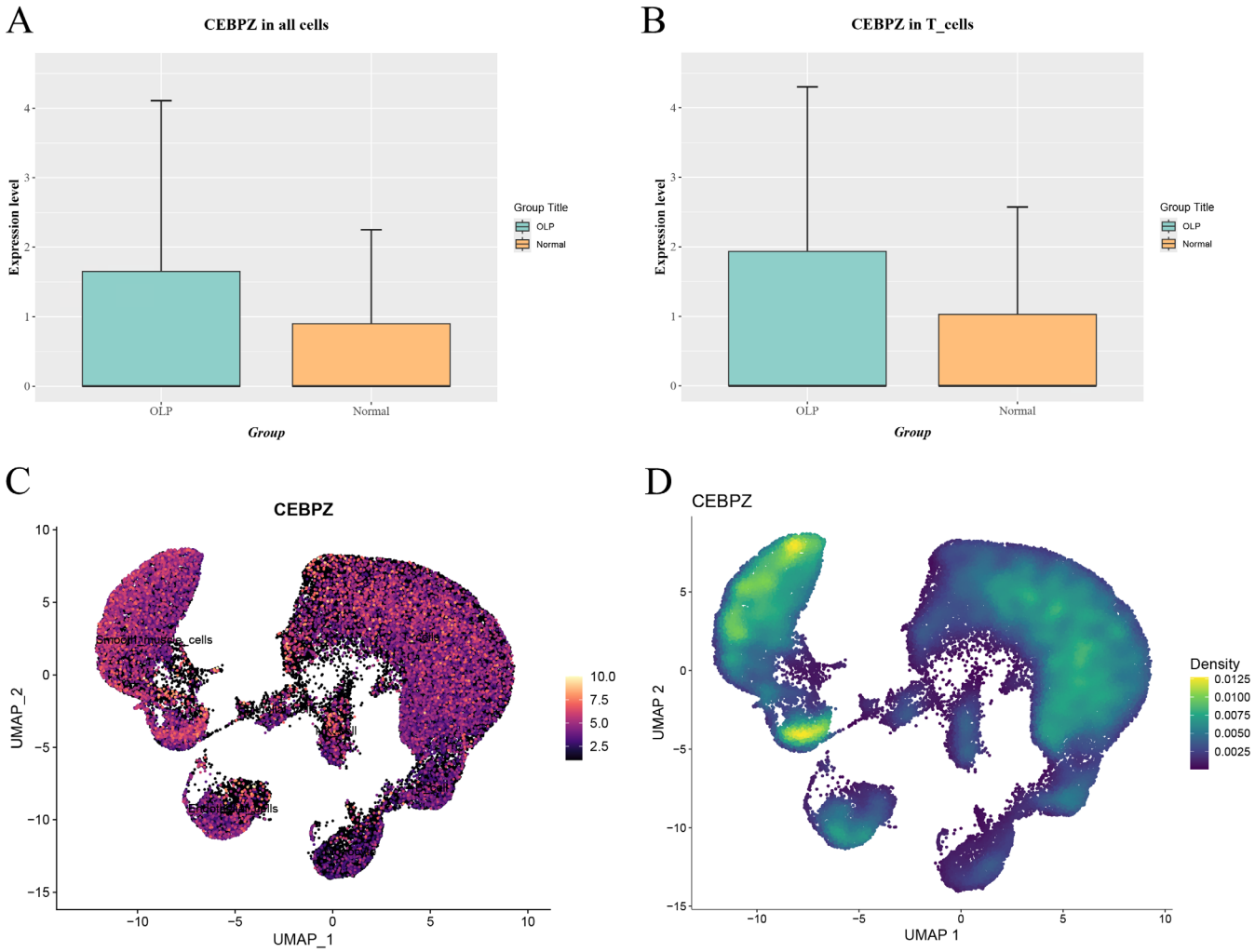

Fig. 9. *CEBPZ* expression in oral lichen planus (OLP). A,B. Box plots of *CEBPZ* expression in all cells and T cells comparing OLP and normal groups; C. Uniform Manifold Approximation and Projection (UMAP) plot showing the distribution of *CEBPZ* expression across all OLP cells; D. Density plot showing CEBPZ expression enrichment

By combining SMR and scRNA-seq analyses, we found that high *CEBPZ* expression is associated with activated MIF pathway signaling between T cells and monocytes. *CEBPZ* may influence the expression of *MIF*, *CXCR4*, *CD74*, and *CD44* through the MIF pathway, creating a feed-forward loop that perpetuates inflammation and oxidative stress in the OLP microenvironment. Studies have demonstrated that T-cell-mediated inflammation in OLP is associated with increased *TRIM21* expression.[67] *TRIM21* links ubiquitinated NF-κB via K63, thereby activating the NF-κB signaling pathway and promoting inflammation.[68] Corticosteroids are first-line therapies that target NF-κB and have considerable therapeutic significance in OLP treatment. T cells may trigger keratinocyte apoptosis through TNF-α binding to TNF receptor 1 (TNFR1), CD95L binding to CD95, or granzyme B entering through perforin-created membrane pores.[69–71] This finding highlights the potential significance of *CEBPZ* in immune modulation and provides a novel direction for investigating its potential as a therapeutic target or biomarker. While clinical examination and histological assessment remain the preferred approaches for evaluating OLP, molecular diagnostic methods such as oral tissue RNA sequencing may significantly enhance diagnostic accuracy.

The clinical implications of *CEBPZ* extend beyond diagnosis. *CEBPZ* expression could potentially guide treatment selection, with high expression indicating patients who may benefit from MIF pathway-targeted therapies or oxidative stress interventions. Integration of *CEBPZ* into ML models could enable risk stratification and identify patients at higher risk of malignant transformation who require more intensive surveillance.

Limitations of the study

This study has several limitations that should be considered when interpreting the results. First, the analyses relied on publicly available datasets from predominantly European populations, which may lack comprehensive clinical metadata, potentially limiting the generalizability of the findings to other ethnic groups. Second, the cross-sectional nature of the data restricts the ability

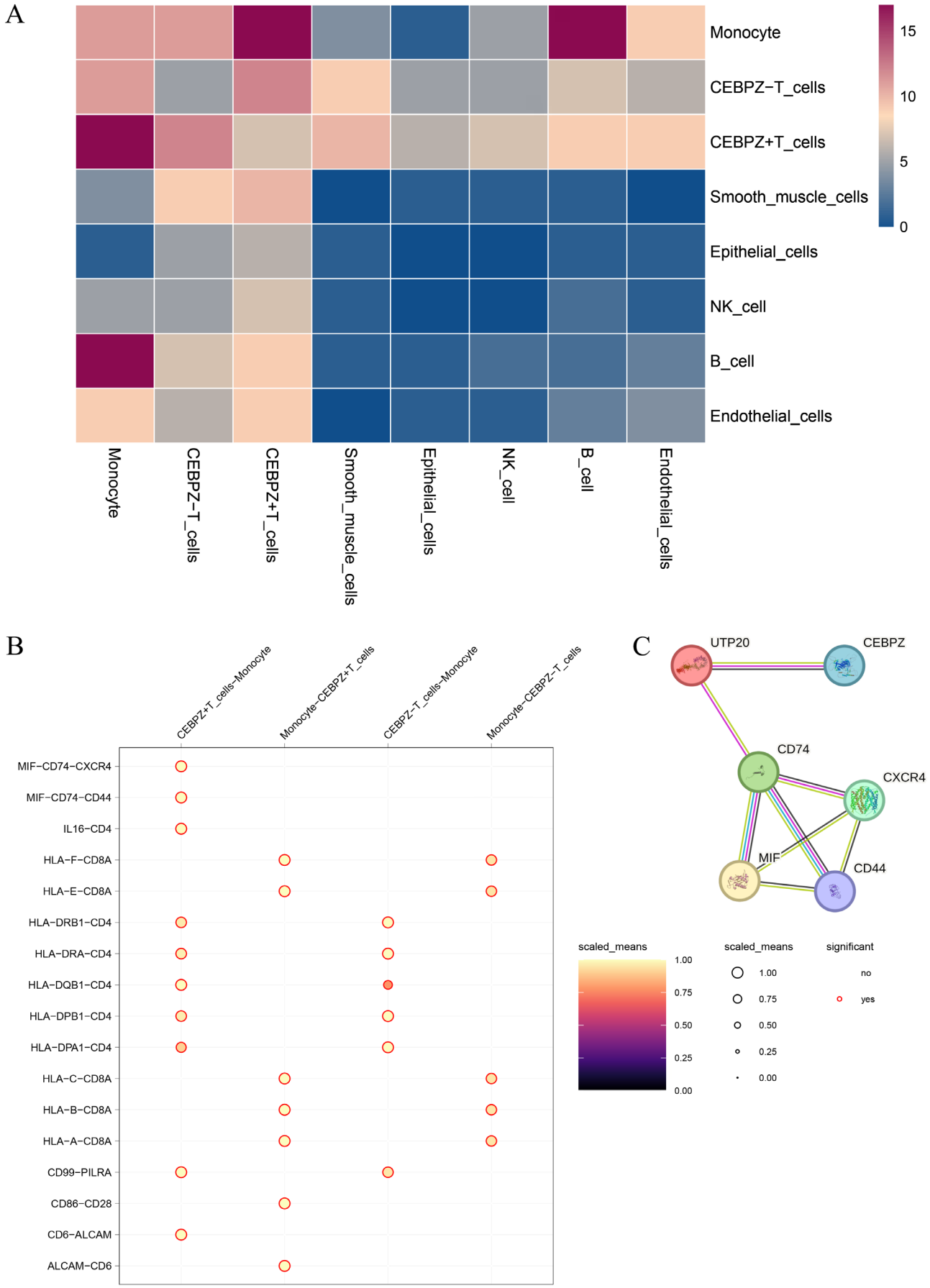

Fig. 10. Role of *CEBPZ* in oral lichen planus (OLP) immune interactions. A. Heatmap of cell-type interaction frequencies; B. Dot plot of receptor–ligand interactions between *CEBPZ*+ and *CEBPZ*– T cells and monocytes; C. Protein–protein interaction (PPI) network generated using the STRING database

to infer temporal changes beyond the genetic associations established through SMR. Third, esophageal mucosa eQTLs served as proxies for oral epithelium; although both tissues share squamous epithelial characteristics, tissue-specific regulatory mechanisms may differ, and future validation using oral tissue-specific eQTLs will be essential. Fourth, experimental validation of identified biomarkers, such as *CEBPZ*, is needed to confirm their functional roles through in vitro and in vivo studies. Fifth, potential batch effects between different datasets, despite computational correction, may have influenced the results. Finally, ML models require prospective validation in independent, clinically characterized cohorts to assess their real-world applicability. Future studies incorporating longitudinal samples, functional assays, and clinical trials will be essential to strengthen the translational impact.

Conclusions

This study pioneers an integrative genomic approach to unravel the immune-mediated mechanisms of OLP by combining scRNA-seq, SMR, hdWGCNA, and ML. We identified a novel T-cell subpopulation in the OLP LP characterized by heightened oxidative stress and enhanced MIF pathway interactions with monocytes and B cells. *CEBPZ* emerged as a pivotal biomarker at the intersection of SMR and LASSO analyses, with elevated expression correlating with immune infiltration and OLP pathogenesis. The developed diagnostic models (AUC: 0.653–0.745) offer promising tools for clinical application. These findings deepen our understanding of the molecular landscape of OLP and highlight *CEBPZ* as a potential therapeutic target, paving the way for precision medicine strategies in OLP.

Supplementary data

The supplementary materials are available at https://doi.org/10.5281/zenodo.18939928. The package contains the following files:

Supplementary Fig. 1. SMR analysis workflow schematic. Step-by-step diagram illustrating the SMR analysis pipeline, including data acquisition, quality control, instrument selection, SMR/HEIDI analysis, and multi-omic integration.

Supplementary Fig. 2. Quality control of the scRNA-seq dataset. UMAP plots showing the GSE211630 dataset before and after quality-control preprocessing.

Supplementary Fig. 3. Signaling pathways in cell–cell communication. Bubble plots highlighting significant signaling pathways involved in cell–cell communication between T-cell subtypes and other cell types, identified using CellChat.

Supplementary Fig. 4. MIF and CXCL signaling pathway analysis. A. Role of T-cell subtypes in the MIF signaling pathway network; B. Role of T-cell subtypes in the CXCL signaling pathway network; C. Scatter plot of outgoing vs incoming interaction strengths across cell types.

Supplementary Fig. 5. Expression of LASSO-selected genes. UMAP plots displaying expression patterns of 9 LASSO-selected genes in OLP and healthy tissues.

Supplementary Fig. 6. External validation of ML models. ROC curves for the Logreg, ABESS, and LDA models in the external validation set GSE38616 (AUC = 0.745, 0.694, and 0.653, respectively).

Supplementary Fig. 7. Immune cell proportions between OLP subtypes. Box plot comparing immune cell composition between Subtype 1 and Subtype 2 across 10 MCP-counter cell types.

Supplementary Fig. 8. *CEBPZ* expression across cell types and developmental trajectory. A. Dot plot of *CEBPZ* expression across cell types; B. RNA velocity analysis showing developmental vectors; C. UMAP of T-cell clusters in OLP and normal tissues; D. *CEBPZ* expression distribution across T-cell clusters.

Supplementary Table 1. Complete SMR analysis results. Excel file containing all 231 genes identified through SMR analysis, together with their corresponding statistical values.

Data Availability Statement

The datasets analyzed in this study are publicly available. Single-cell RNA sequencing data (GSE211630, GSE52130, and GSE38616) were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo). GWAS summary data for oral lichen planus were obtained from the FinnGen database (release 8; https://www.finngen.fi). Esophageal mucosa eQTL data were retrieved from the GTEx database (version 8; https://gtexportal.org). Protein interaction data were obtained from the STRING database (https://string-db.org). The complete results of the SMR analysis, including 231 genes, are provided in Supplementary Table 1. All data were processed in accordance with the guidelines and policies of the respective repositories. No new data were generated in this study.

Consent for publication of personal information

Not applicable.

Use of AI and AI-assisted technologies

Not applicable.

ORCID iDs

Yichen Feng https://orcid.org/0009-0002-2256-7799
Shengfeng Wang https://orcid.org/0009-0008-4721-8193
Ling Guan https://orcid.org/0000-0001-9912-6305
Linlin Wang https://orcid.org/0009-0005-6437-126X
Anqi Zou https://orcid.org/0009-0009-9555-6514
Zhankui Xing https://orcid.org/0009-0000-7105-096X
Tingying Xiao https://orcid.org/0000-0003-2037-7254

References

1. Nukaly HY, Halawani IR, Alghamdi SMS, et al. Oral lichen planus: A narrative review navigating etiologies, clinical manifestations, diagnostics, and therapeutic approaches. *J Clin Med*. 2024;13(17):5280. doi:10.3390/jcm13175280
2. Didona D, Caposiena Caro RD, Sequeira Santos AM, Solimani F, Hertl M. Therapeutic strategies for oral lichen planus: State of the art and new insights. *Front Med (Lausanne)*. 2022;9:997190. doi:10.3389/fmed.2022.997190
3. Li C, Tang X, Zheng X, et al. Global prevalence and incidence estimates of oral lichen planus: A systematic review and meta-analysis. *JAMA Dermatol*. 2020;156(2):172. doi:10.1001/jamadermatol.2019.3797
4. Tsushima F, Sakurai J, Uesugi A, et al. Malignant transformation of oral lichen planus: A retrospective study of 565 Japanese patients. *BMC Oral Health*. 2021;21(1):298. doi:10.1186/s12903-021-01652-7
5. Zotti F, Nocini R, Capocasale G, Fior A, Peretti M, Albanese M. Malignant transformation evidences of oral lichen planus: When the time is of the essence. *Oral Oncol*. 2020;104:104594. doi:10.1016/j.oraloncology.2020.104594
6. Shavit E, Hagen K, Shear N. Oral lichen planus: A novel staging and algorithmic approach and all that is essential to know. *F1000Res*. 2020;9:206. doi:10.12688/f1000research.18713.1
7. Lucchese A, Di Stasio D, Romano A, et al. Correlation between oral lichen planus and viral infections other than HCV: A systematic review. *J Clin Med*. 2022;11(18):5487. doi:10.3390/jcm11185487
8. Scotto G, Fazio V, Massa S, Lo Muzio L, Spirito F. COVID-19 and oral lichen planus: Between an "intriguing plot" and the "fata morgana effect." *J Clin Med*. 2023;12(14):4829. doi:10.3390/jcm12144829
9. Vičić M, Sotošek V, Brajac I, Kaštelan M, Prpić-Massari L. The possible involvement of granulysin mediated cytotoxicity in keratinocytes disruption in lichen planus. *Med Hypotheses*. 2021;149:110546. doi:10.1016/j.mehy.2021.110546
10. Weber B, Schlapbach C, Stuck M, et al. Distinct interferon-gamma and interleukin-9 expression in cutaneous and oral lichen planus. *Acad Dermatol Venereol*. 2017;31(5):880–886. doi:10.1111/jdv.13989
11. Mahdavi N, Aminishakib P, Soltani N. Presence of B cells and plasma cells in oral lichen planus. *J Dent (Shiraz)*. 2020;21(3):209–214. doi:10.30476/DENTJODS.2020.81804.0
12. Xia DN, Tan YQ, Yang JY, Zhou G. Omega-3 polyunsaturated fatty acids: A promising approach for the management of oral lichen planus. *Inflamm Res*. 2020;69(10):989–999. doi:10.1007/s00011-020-01388-0
13. El-Howati A, Thornhill MH, Colley HE, Murdoch C. Immune mechanisms in oral lichen planus. *Oral Dis*. 2023;29(4):1400–1415. doi:10.1111/odi.14142
14. Huang J, Hu Y, Wang S, et al. Single-cell RNA sequencing in autoimmune diseases: New insights and challenges. *Pharmacol Ther*. 2025;267:108807. doi:10.1016/j.pharmthera.2025.108807
15. Awuah WA, Ahluwalia A, Ghosh S, et al. The molecular landscape of neurological disorders: Insights from single-cell RNA sequencing in neurology and neurosurgery. *Eur J Med Res*. 2023;28(1):529. doi:10.1186/s40001-023-01504-w
16. Li Q, Wang F, Shi Y, et al. Single-cell immune profiling reveals immune responses in oral lichen planus. *Front Immunol*. 2023;14:1182732. doi:10.3389/fimmu.2023.1182732
17. Wu Y, Zhang CY, Wang L, Li Y, Xiao X. Genetic insights of schizophrenia via single cell RNA-sequencing analyses. *Schizophr Bull*. 2023;49(4):914–922. doi:10.1093/schbul/sbad002
18. Qing M, Yang D, Shang Q, et al. $CD8^+$ tissue-resident memory T cells induce oral lichen planus erosion via cytokine network. *eLife*. 2023;12:e83981. doi:10.7554/eLife.83981
19. Danielsson K, Coates P, Ebrahimi M, Nylander E, Wahlin Y, Nylander K. Genes involved in epithelial differentiation and development are differentially expressed in oral and genital lichen planus epithelium compared to normal epithelium. *Acta Derm Venerol*. 2014;94(5):526–530. doi:10.2340/00015555-1803
20. Gassling V, Hampe J, Açil Y, Braesen JH, Wiltfang J, Häsler R. Disease-associated miRNA-mRNA networks in oral lichen planus. *PLoS One*. 2013;8(5):e63015. doi:10.1371/journal.pone.0063015
21. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucl Acids Res*. 2023;51(D1):D638–D646. doi:10.1093/nar/gkac1000
22. Li Y, Wang J, Wang F, Gao C, Cao Y, Wang J. Identification of specific cell subpopulations and marker genes in ovarian cancer using single-cell RNA sequencing. *Biomed Res Int*. 2021;2021(1):1005793. doi:10.1155/2021/1005793
23. Gu Z, Hübschmann D. SimplifyEnrichment: A Bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinformatics*. 2023;21(1):190–202. doi:10.1016/j.gpb.2022.04.008
24. Fu Y, Guo Z, Wang Y, et al. Single-nucleus RNA sequencing reveals the shared mechanisms inducing cognitive impairment between COVID-19 and Alzheimer's disease. *Front Immunol*. 2022;13:967356. doi:10.3389/fimmu.2022.967356
25. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods*. 2023;3(6):100498. doi:10.1016/j.crmeth.2023.100498
26. Fang Z, Li J, Cao F, Li F. Integration of scRNA-Seq and bulk RNA-Seq reveals molecular characterization of the immune microenvironment in acute pancreatitis. *Biomolecules*. 2022;13(1):78. doi:10.3390/biom13010078
27. Liu Q, Zhang P, Yuan X, et al. Investigate the stemness of adult adipose-derived stromal cells based on single-cell RNA-sequencing. *Cell Biol Int*. 2022;46(12):2118–2131. doi:10.1002/cbin.11898
28. Ionkina AA, Balderrama-Gutierrez G, Ibanez KJ, et al. Transcriptome analysis of heterogeneity in mouse model of metastatic breast cancer. *Breast Cancer Res*. 2021;23(1):93. doi:10.1186/s13058-021-01468-x
29. Qiang M, Dai Z. Biomarkers of UVB radiation-related senescent fibroblasts. *Sci Rep*. 2024;14(1):933. doi:10.1038/s41598-023-51058-4
30. Joglekar MV, Dong CX, Wong WKM, Dalgaard LT, Hardikar AA. A bird's eye view of the dynamics of pancreatic β-cell heterogeneity. *Acta Physiol*. 2021;233(2):e13664. doi:10.1111/apha.13664
31. Das S, Rai A, Rai SN. Differential expression analysis of single-cell RNA-Seq data: Current statistical approaches and outstanding challenges. *Entropy (Basel)*. 2022;24(7):995. doi:10.3390/e24070995
32. Zhang F, Li X, Tian W. Unsupervised inference of developmental directions for single cells using VECTOR. *Cell Rep*. 2020;32(8):108069. doi:10.1016/j.celrep.2020.108069
33. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell–cell communication using CellChat. *Nat Commun*. 2021;12(1):1088. doi:10.1038/s41467-021-21246-9
34. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. *Nucl Acids Res*. 2018;46(W1):W60–W64. doi:10.1093/nar/gky311
35. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*. 2008;9(Suppl 1):S4. doi:10.1186/gb-2008-9-s1-s4
36. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. doi:10.1038/s41467-019-09234-6
37. Rusk N. Expanded CIBERSORTx. *Nat Methods*. 2019;16(7):577–577. doi:10.1038/s41592-019-0486-8
38. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218. doi:10.1186/s13059-016-1070-5
39. Shao C, Höfer T. Robust classification of single-cell transcriptome data by nonnegative matrix factorization. *Bioinformatics*. 2017;33(2):235–242. doi:10.1093/bioinformatics/btw607
40. Liang J, Wang C, Zhang D, et al. VSOLassoBag: A variable-selection oriented LASSO bagging algorithm for biomarker discovery in omic-based translational research. *J Genet Genomics*. 2023;50(3):151–162. doi:10.1016/j.jgg.2022.12.005
41. Liu Y, Liang Z, Yang J, et al. Diagnostic and comparative performance for the prediction of tuberculous pleural effusion using machine learning algorithms. *Int J Med Inform*. 2024;182:105320. doi:10.1016/j.ijmedinf.2023.105320
42. Yu L, Li B, Jiao B. Research and implementation of CNN based on TensorFlow. *IOP Conf Ser Mater Sci Eng*. 2019;490:042022. doi:10.1088/1757-899X/490/4/042022
43. Lee H, Song J. Introduction to convolutional neural network using Keras: An understanding from a statistician. *Commun Stat Appl Methods*. 2019;26(6):591–610. doi:10.29220/CSAM.2019.26.6.591

44. Chang SS, Cheng CC, Chen YR, Chen FW, Cheng YM, Wang JM. Epithelial *CEBPD* activates fibronectin and enhances macrophage adhesion in renal ischemia–reperfusion injury. *Cell Death Discov.* 2024;10(1):328. doi:10.1038/s41420-024-02082-4
45. Spek CA, Aberson HL, Butler JM, De Vos AF, Duitman J. *CEBPD* potentiates the macrophage inflammatory response but *CEBPD* knockout macrophages fail to identify *CEBPD*-dependent pro-inflammatory transcriptional programs. *Cells*. 2021;10(9):2233. doi:10.3390/cells10092233
46. Kashyap B, Hyvärinen E, Laitinen I, Kullaa AM. Salivary metabolomics in patients with oral lichen planus: A preliminary study based on NMR spectroscopy. *Clin Oral Invest*. 2024;28(1):103. doi:10.1007/s00784-023-05389-1
47. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol*. 2021;82(11):801–811. doi:10.1016/j.humimm.2021.02.012
48. Li X, Wang CY. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci*. 2021;13(1):36. doi:10.1038/s41368-021-00146-0
49. Bai X, Li Y, Zeng X, Zhao Q, Zhang Z. Single-cell sequencing technology in tumor research. *Clin Chim Acta*. 2021;518:101–109. doi:10.1016/j.cca.2021.03.013
50. Yu X, Liang J, Xu J, et al. Identification and validation of circulating microRNA signatures for breast cancer early detection based on large scale tissue-derived data. *J Breast Cancer*. 2018;21(4):363. doi:10.4048/jbc.2018.21.e56
51. Kamath VV, Krishnamurthy S, Satelur KP, Rajkumar K. Transforming growth factor-β1 and TGF-β2 act synergistically in the fibrotic pathway in oral submucous fibrosis: An immunohistochemical observation. *Indian J Med Paediatr Oncol*. 2015;36(2):111–116. doi:10.4103/0971-5851.158842
52. Liu ZW, Zhang YM, Zhang LY, et al. Duality of interactions between TGF-β and TNF-α during tumor formation. *Front Immunol*. 2022;12:810286. doi:10.3389/fimmu.2021.810286
53. Lodi G, Manfredi M, Mercadante V, Murphy R, Carrozzo M. Interventions for treating oral lichen planus: Corticosteroid therapies. *Cochrane Database Syst Rev*. 2020;2020(2):CD001168. doi:10.1002/14651858.CD001168.pub3
54. Andabak-Rogulj A, Vindiš E, Aleksijević L, et al. Different treatment modalities of oral lichen planus: A narrative review. *Dent J*. 2023;11(1):26. doi:10.3390/dj11010026
55. Wang Y, Lian X, Zhong J, et al. Serum exosomal microRNA let-7i-3p as candidate diagnostic biomarker for Kawasaki disease patients with coronary artery aneurysm. *IUBMB Life*. 2019;71(7):891–900. doi:10.1002/iub.2015
56. Huang D, Li Y, Liu Y. Tacrolimus and the treatment of pulmonary fibrosis. *Am J Respir Crit Care Med*. 2023;208(11):1241–1242. doi:10.1164/rccm.202308-1445LE
57. Wallin EF, Hill DL, Linterman MA, Wood KJ. The calcineurin inhibitor tacrolimus specifically suppresses human T follicular helper cells. *Front Immunol*. 2018;9:1184. doi:10.3389/fimmu.2018.01184
58. Deng X, Wang Y, Jiang L, Li J, Chen Q. Updates on immunological mechanistic insights and targeting of the oral lichen planus microenvironment. *Front Immunol*. 2023;13:1023213. doi:10.3389/fimmu.2022.1023213
59. Zhang Q, Zhang S, Chen J, Xie Z. The interplay between integrins and immune cells as a regulator in cancer immunology. *Int J Mol Sci*. 2023;24(7):6170. doi:10.3390/ijms24076170
60. Yang J, Zhang J, Lu R, Tan Y, Du G, Zhou G. T cell-derived exosomes induced macrophage inflammatory protein-1α/β drive the trafficking of $CD8^+$ T cells in oral lichen planus. *J Cell Mol Med*. 2020;24(23):14086–14098. doi:10.1111/jcmm.16020
61. Zhong EF, Chang A, Stucky A, et al. Genomic analysis of oral lichen planus and related oral microbiome pathogens. *Pathogens*. 2020;9(11):952. doi:10.3390/pathogens9110952
62. Huang Y, Lin L, Shen Z, et al. *CEBPG* promotes esophageal squamous cell carcinoma progression by enhancing PI3K-AKT signaling. *Am J Cancer Res*. 2020;10(10):3328–3344. PMID:33163273. PMCID:PMC7642652.
63. Ren Q, Liu Z, Wu L, et al. C/EBPβ: The structure, regulation, and its roles in inflammation-related diseases. *Biomed Pharmacother*. 2023;169:115938. doi:10.1016/j.biopha.2023.115938
64. Hartl L, Duitman J, Bijlsma MF, Spek CA. The dual role of C/EBPδ in cancer. *Crit Rev Oncol Hematol*. 2023;185:103983. doi:10.1016/j.critrevonc.2023.103983
65. Ramji DP, Foka P. CCAAT/enhancer-binding proteins: Structure, function and regulation. *Biochem J*. 2002;365(3):561–575. doi:10.1042/bj20020508
66. Zhang F, Shen Z, Xie J, et al. *CEBPA* bZIP in-frame mutations in acute myeloid leukemia: Prognostic and therapeutic implications. *Blood Cancer J*. 2024;14(1):59. doi:10.1038/s41408-024-01042-6
67. Wei W, Wang Y, Sun Q, et al. Enhanced T-cell proliferation and IL-6 secretion mediated by overexpression of *TRIM21* in oral lesions of patients with oral lichen planus. *J Oral Pathol Med*. 2020;49(4):350–356. doi:10.1111/jop.12938
68. Gong X, He S, Cai P. Roles of *TRIM21*/Ro52 in connective tissue disease-associated interstitial lung diseases. *Front Immunol*. 2024;15:1435525. doi:10.3389/fimmu.2024.1435525
69. Pocino K, Carnazzo V, Stefanile A, et al. Tumor necrosis factor-alpha: Ally and enemy in protean cutaneous sceneries. *Int J Mol Sci*. 2024;25(14):7762. doi:10.3390/ijms25147762
70. Sica M, Roussel M, Legembre P. CD95/Fas stoichiometry in future precision medicine. *Cell Death Differ*. 2025;32(9):1570–1577. doi:10.1038/s41418-025-01493-9
71. Paulsen M, Janssen O. Pro- and anti-apoptotic CD95 signaling in T cells. *Cell Commun Signal*. 2011;9(1):7. doi:10.1186/1478-811X-9-7