Single-Cell Transcriptome Analysis of the Oral Mucosa

Introduction

Single-cell RNA sequencing (scRNA-seq) is a powerful transcriptomic technique that allows for the profiling of gene expression at the level of individual cells. Unlike bulk RNA sequencing, which measures the collective gene expression of an entire sample, scRNA-seq captures the transcriptomic heterogeneity within a population of cells. This technology provides insights into cellular diversity, enabling the identification of rare cell types, the characterization of dynamic transcriptional changes, and the study of cellular responses to different biological conditions. One of the key distinctions between scRNA-seq and bulk RNA sequencing lies in resolution. While bulk RNA-seq provides an averaged expression profile across all cells in a sample, scRNA-seq dissects this complexity by analyzing each cell independently. This capability is particularly crucial in tissues with diverse cellular compositions, such as the immune system or tumor microenvironment, where distinct cell types coexist with varying functions. By capturing single-cell transcriptomes, scRNA-seq facilitates a deeper understanding of cellular states, transitions, and lineage relationships, which are often masked in bulk analyses. Another major advantage of scRNA-seq is its ability to identify and characterize rare cell populations that would otherwise remain undetected in bulk sequencing. In conditions such as cancer, autoimmune diseases, or neurodegenerative disorders, rare but functionally significant cells may play a critical role in disease progression. Bulk RNA sequencing, which averages gene expression signals from all cells, can obscure these unique signatures. In contrast, scRNA-seq can pinpoint such rare cells and elucidate their contributions to the overall system. Despite these advantages, scRNA-seq comes with technical and analytical challenges. The data generated is inherently sparse due to the low capture efficiency of transcripts from individual cells, leading to dropout events where certain genes appear undetected. Additionally, the high dimensionality of scRNA-seq data necessitates specialized computational methods for normalization, clustering, and downstream interpretation. Unlike bulk RNA-seq, where gene expression levels are more stable due to the pooling of RNA from many cells, single-cell data requires sophisticated correction techniques to account for variations in sequencing depth and technical noise. Overall, scRNA-seq represents a significant advancement in transcriptomics, offering an unprecedented view of cellular heterogeneity and function. Its ability to resolve individual cell states makes it a valuable tool for studying complex biological systems, uncovering novel cell types, and understanding gene regulatory mechanisms. While bulk RNA sequencing remains useful for large-scale expression studies and quantitative comparisons, scRNA-seq has become indispensable for applications requiring single-cell resolution and detailed cellular characterization.

Objective

The primary objective of this project is to analyze single-cell RNA sequencing (scRNA-seq) data to characterize the immune cell landscape within the oral mucosa. By leveraging computational techniques, we aim to identify distinct immune cell subsets, explore their transcriptional profiles, and understand their roles in inflammation and immune resolution. This analysis will provide insights into cellular heterogeneity, gene expression patterns, and potential markers associated with immune responses in the oral environment. Through clustering and differential expression analysis, we seek to uncover key biological processes and cellular interactions that contribute to immune regulation in the oral mucosa.

Data

For this study, we utilized publicly available single-cell RNA sequencing (scRNA-seq) data from the Gene Expression Omnibus (GEO) under accession ID GSE244633. This dataset contains transcriptomic profiles of immune cells isolated from the oral mucosa, providing a detailed view of cellular heterogeneity and gene expression dynamics. The data was originally generated to investigate immune responses during inflammation and resolution in the oral cavity.The dataset consists of preprocessed count matrices, which were downloaded directly from GEO in .h5 format. These files contain gene expression counts for individual cells, serving as the

input for downstream analysis. We processed and analyzed this data using Seurat, a widely used R package for single-cell RNA-seq analysis, to identify cell clusters, detect marker genes, and perform functional annotation of immune cell populations.

Results

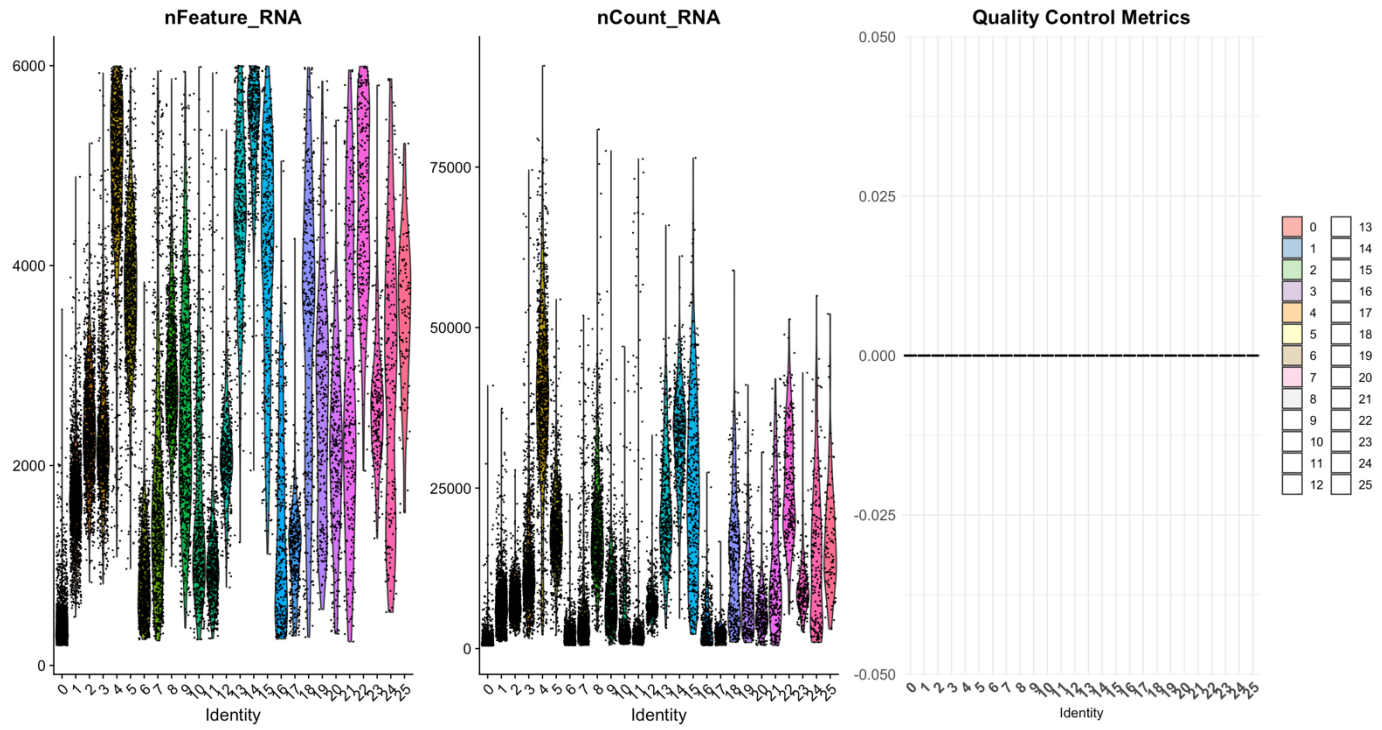

Fig – 1: Violin Plot for Quality Control Metrics

The violin plot presents three key quality control (QC) metrics across different cell clusters: nFeature_RNA (left), nCount_RNA (middle), and percent.mt (right). The nFeature_RNA distribution shows the number of unique genes detected per cell, with variations across clusters indicating differences in transcriptional complexity. The nCount_RNA distribution reflects the total RNA molecules captured per cell, with some clusters displaying a broader range, suggesting high transcriptional activity or potential doublets. The percent.mt plot, which represents the proportion of mitochondrial transcripts, appears empty, possibly due to visualization issues or a lack of mitochondrial gene detection in the dataset. High mitochondrial content is typically associated with cell stress or death, so its accurate representation is crucial. Additionally, the x-axis labels are cluttered, making cluster differentiation difficult. To enhance clarity, the visualization should be refined by adjusting label size and rotation while ensuring mitochondrial content is correctly displayed before proceeding to downstream analysis, including differential expression and cell type annotation.

The UMAP Clustering of Cells plot (Fig-2) provides a visualization of the identified clusters in the dataset, representing different cell populations based on gene expression similarities. Each point in the UMAP space represents a single cell, and cells with similar expression profiles are positioned closer together. The distinct color-coded clusters highlight transcriptionally distinct groups, suggesting different cell types or functional states. The presence of multiple well-separated clusters indicates a high level of heterogeneity in the dataset, which is expected in single-cell transcriptomics. The labelled cluster numbers provide a reference for downstream cell-type annotation and biological interpretation. The clustering structure suggests major immune and non-immune cell populations within the oral mucosa, with potential functional roles in immune regulation and tissue homeostasis.

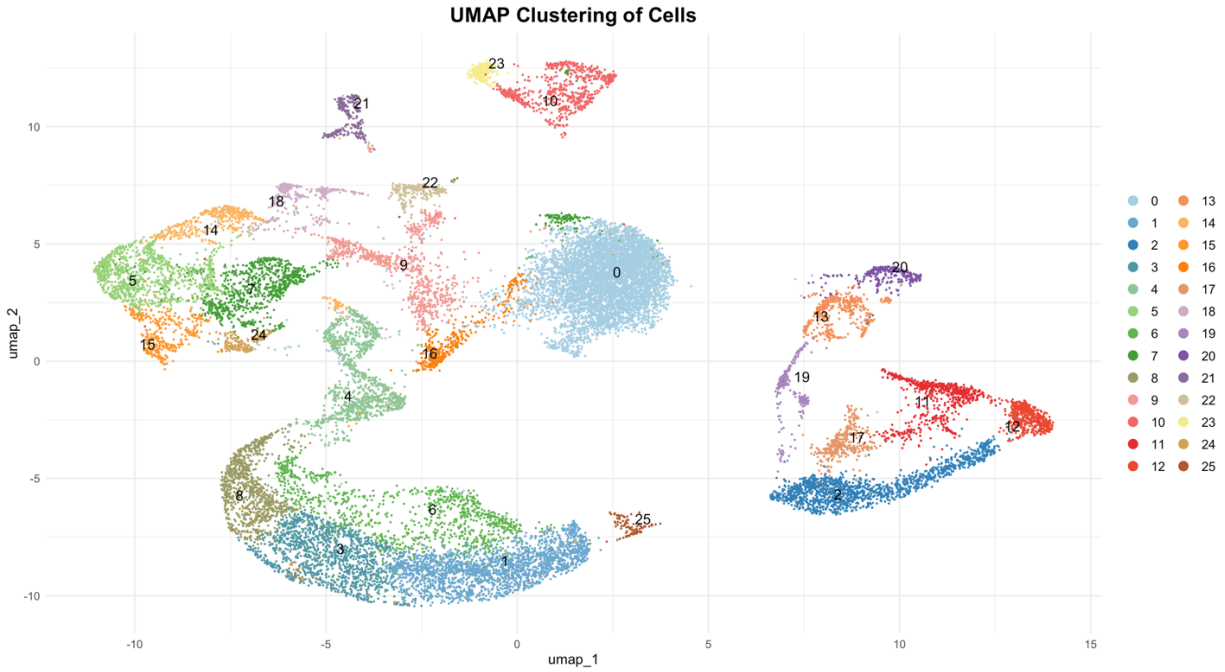

Fig -2: UMAP Clustering Plot

The Cell Type Annotation with SingleR (Fig-3) plot integrates reference-based cell type annotation, assigning specific immune and stromal cell identities to each cluster. The annotation was performed using SingleR, a tool that compares single-cell transcriptomes with curated reference datasets. Different cell types, such as T cells, B cells, macrophages, monocytes, dendritic cells, NK cells, fibroblasts, and endothelial cells, are clearly demarcated in the UMAP space. The presence of diverse immune populations, including monocytes, macrophages, and dendritic cells, suggests an active immune environment in the oral mucosa, potentially involved in resolving inflammation. Additionally, the presence of stromal cells such as fibroblasts and endothelial cells reflects their supportive roles in maintaining tissue integrity. The annotation aligns with the clustering pattern observed in the UMAP, confirming the robustness of the identified clusters and their biological relevance.

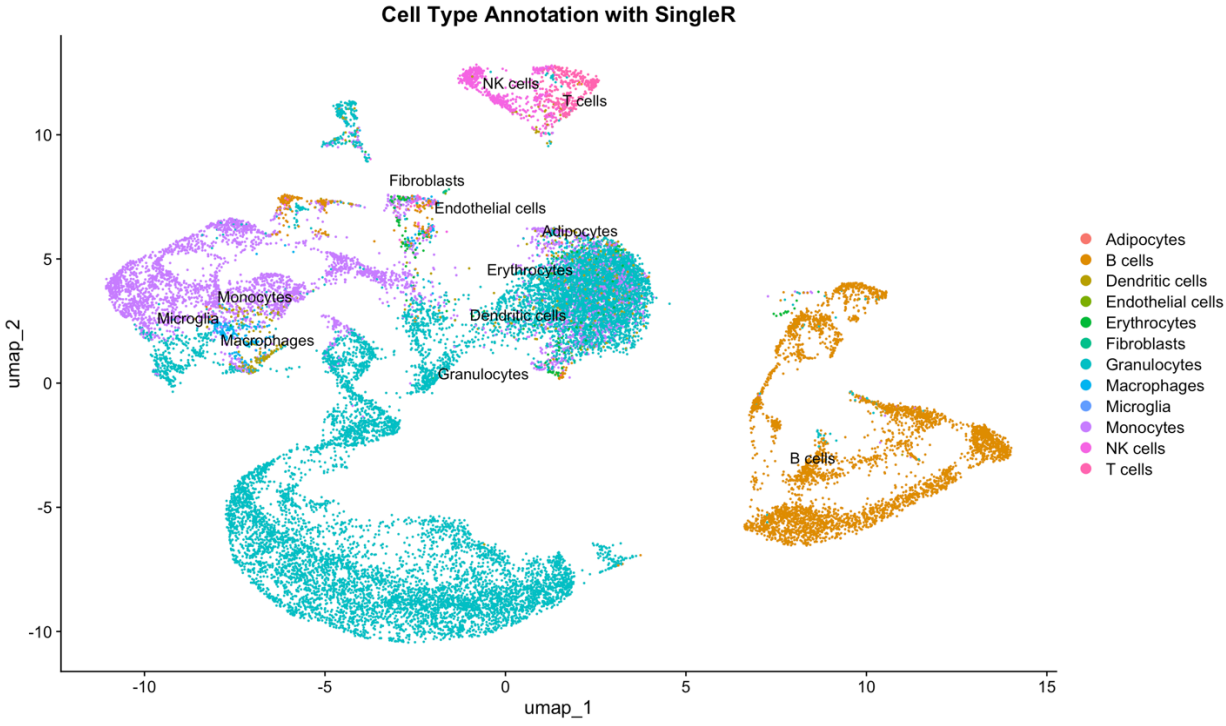

Fig-3 Cell Type Annotation Plot

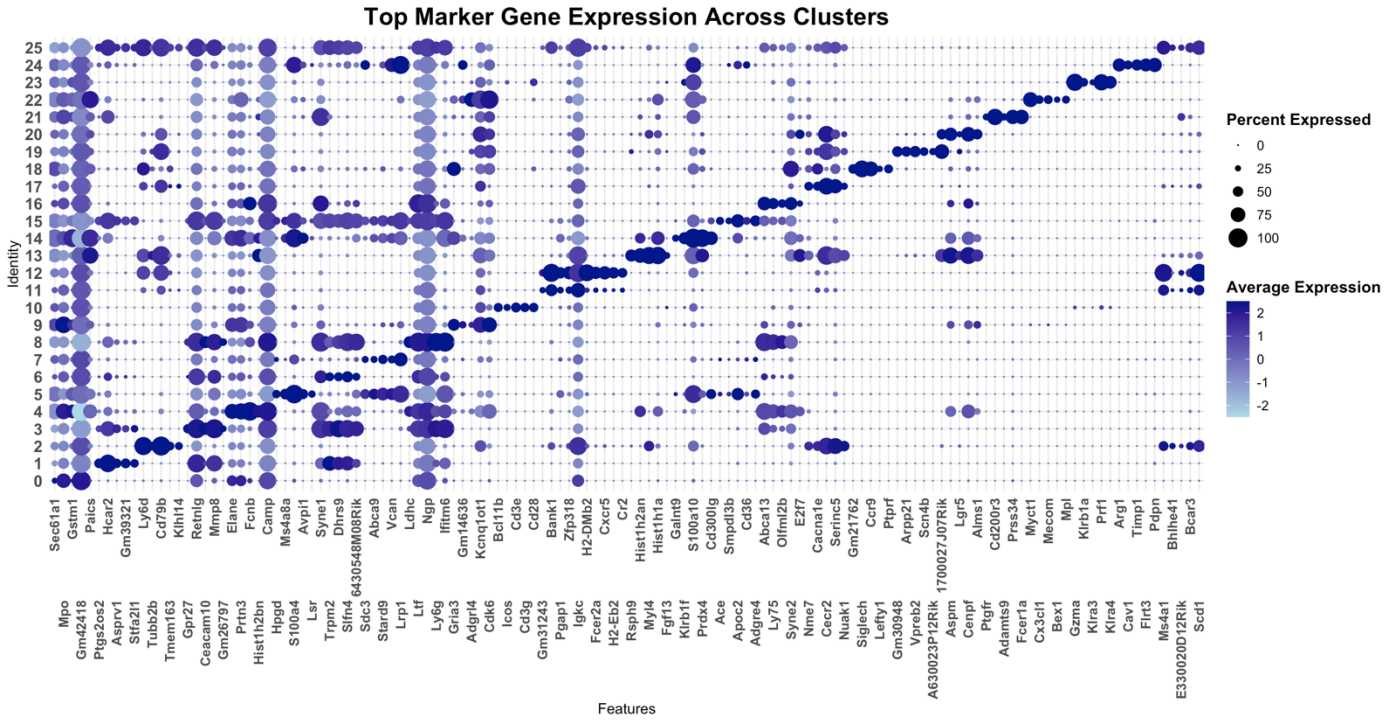

Fig-5: Gene Expression Clusters

The Dot Plot provides an overview of the expression patterns of key marker genes across different cell clusters. Each row represents a distinct cluster, while each column corresponds to a marker gene. The size of the dots reflects the proportion of cells within a given cluster that express the respective gene, while the colour intensity indicates the average expression level. Darker shades represent higher expression levels, while lighter shades or absence of dots suggest low or no expression. From the plot, we observe that certain marker genes exhibit strong expression in specific clusters, indicating their potential roles in defining distinct cell types. For example, genes such as Mpo, Cd79b, and Cx3cr1 show high expression in select clusters, which could correspond to immune cell subsets such as neutrophils, B cells, and monocytes/macrophages, respectively. The differential expression pattern across clusters further supports the heterogeneity of the cellular composition within the dataset, highlighting the presence of multiple immune and stromal populations. The structured distribution of marker expression reinforces the robustness of the clustering process, suggesting that each cluster represents a transcriptionally distinct population. This visualization aids in validating the cell type annotations and identifying genes that play key roles in defining specific cellular states within the oral mucosa.

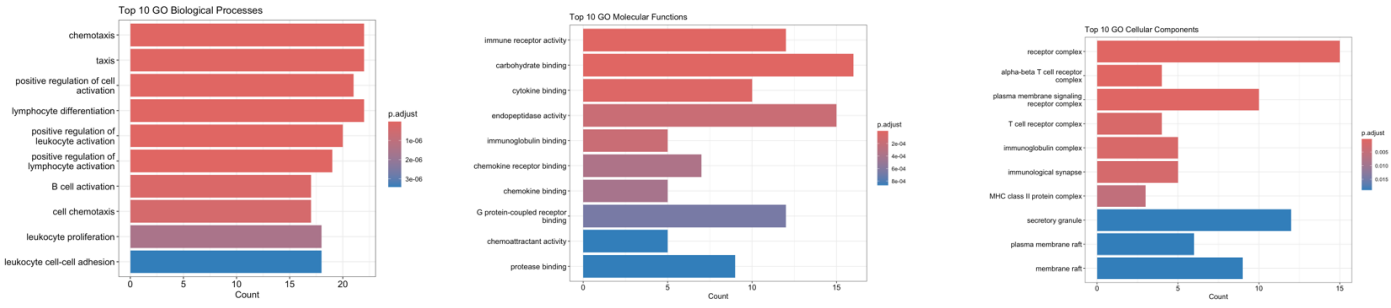

Fig-5 GO Plots

The Gene Ontology (GO) analysis provides insight into the biological processes, molecular functions, and cellular components associated with differentially expressed genes in the dataset. The first plot illustrates the top 10 GO biological processes, highlighting key immune-related functions such as chemotaxis, lymphocyte differentiation, leukocyte proliferation, and cell-cell adhesion. These processes are crucial for immune cell migration, activation, and response to inflammation, reinforcing the study's focus on immune regulation within the oral mucosa. The second plot presents the top GO molecular functions, emphasizing activities such as immune

receptor binding, cytokine binding, chemokine receptor interactions, and protease activity. These functions are integral to immune cell communication and response to external stimuli. The third plot depicts the GO cellular components, showing enriched terms related to receptor complexes, immunoglobulin complexes, and membrane-associated structures. The presence of terms such as “T cell receptor complex” and “MHC class II protein complex” suggests a significant involvement of antigen-presenting cells and adaptive immune responses. Collectively, these GO analyses highlight the key immunological roles of the identified cell clusters, providing a deeper understanding of the immune landscape within the oral mucosa.

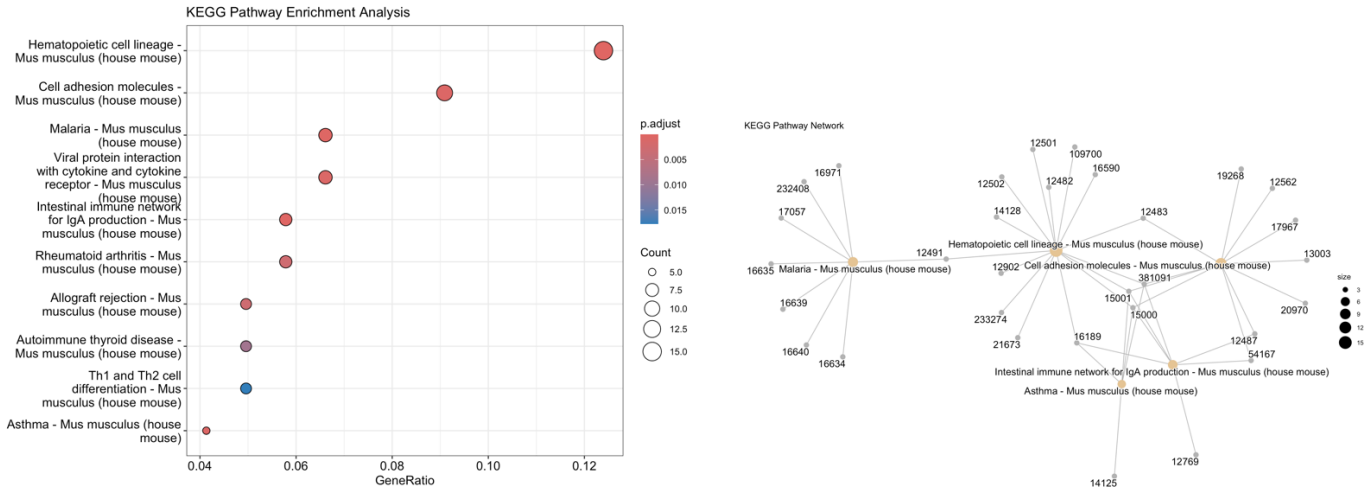

Fig-6 KEGG Pathway Analysis

The KEGG pathway enrichment analysis provides insights into the biological pathways that are significantly associated with the differentially expressed genes (DEGs) identified in this study. The first plot represents the KEGG pathway network, illustrating the interconnections between various pathways enriched in the dataset. Larger nodes indicate pathways with more significant involvement, such as Hematopoietic cell lineage, Cell adhesion molecules, and Intestinal immune network for IgA production, which are central to immune response regulation. The second plot presents a KEGG pathway enrichment bubble chart, where pathways are ranked by their enrichment scores (GeneRatio), with the color gradient representing the adjusted p-values. The most enriched pathways include Hematopoietic cell lineage, Cell adhesion molecules, and Immune-related signaling pathways such as Viral protein interaction with cytokines and cytokine receptors and Autoimmune thyroid disease. These findings suggest that the differentially expressed genes are predominantly involved in immune function, inflammation, and cellular communication, which aligns with the study’s objective of characterizing immune cell heterogeneity and response in the oral mucosa.

Conclusion and Discussion

This study successfully reproduced the single-cell RNA sequencing (scRNA-seq) analysis of immune cell populations in the oral mucosa, providing insights into the heterogeneity of immune responses in inflammation and its resolution. Through clustering and annotation, we identified distinct immune cell subtypes, including macrophages, monocytes, dendritic cells, T cells, and B cells, highlighting their unique transcriptional signatures. The GO enrichment analysis further revealed key biological processes such as leukocyte activation, cell chemotaxis, and immune receptor activity, reinforcing the functional relevance of these immune populations. KEGG pathway analysis identified critical pathways involved in immune regulation, such as hematopoietic cell lineage and cytokine signaling, indicating the activation of immune responses during inflammation. The findings align with the original study, demonstrating that single-cell resolution allows for a more granular understanding of immune cell diversity compared to bulk RNA-seq approaches, which often mask cell-type-specific variations. The annotation of immune cell types provides valuable insights into how these populations contribute to oral immune homeostasis and inflammatory processes, such as periodontitis. However, potential limitations include technical biases in scRNA-seq, dropout events, and variability in gene expression, which

may affect cell-type identification. Future studies could integrate additional datasets or perform spatial transcriptomics to further validate these findings and explore cell-cell interactions in the immune microenvironment.