이학박사학위논문

RNA sequencing 에 기반한 분자면역반응의 체계적 연구

A systematic study of molecular immune response based on RNA sequencing

2023 년 2 월

서울대학교 대학원

협동과정 생물정보학

심 규 원

RNA sequencing 에 기반한 분자면역반응의 체계적 연구

지도교수 강봉균

이 논문을 이학박사 학위논문으로 제출함
2022 년 12 월

서울대학교 대학원
협동과정 생물정보학
심 규 원

심규원의 이학박사 학위논문을 인준함
2022년 12 월

위 원 장 이 용 석 (인)

부위원장 강 봉 균 (인)

위 원 김 형 (인)

위 원 이 재 형 (인)

위 원 문 주 영 (인)

ABSTRACT

A systematic study of molecular immune response based on RNA sequencing

Kyuwon Shim

Interdisciplinary Program in Bioinformatics

College of Natural Sciences

Seoul National University

One of the major challenges of immunology is the intricate intertwining of different markers and cell types. Therefore, various attempts are being made to address these issues through RNA sequencing (RNA-seq) technology, which can effectively identify multiple markers and cell types. Though, many recent advances have been made based on single cell RNA seq (scRNA-seq) technology, but bulk RNA seq is still often required due to technical or cost issues. Here, we present a simple, yet robust and cost-efficient immune response analysis pipeline based on bulk RNA-seq that leverages recent advances in scRNA-seq. We have successfully applied the proposed method to aging mouse kidneys and human periodontitis of varying severity, characterized various immune responses at the molecular level in both models and identified biomarker candidate genes to predict the periodontitis. The result in the current study presents a novel insight for the comprehensive understanding of the immune system.

……

Keyword: Chronic inflammation, Immune response, RNA-seq, single cell deconvolution, Inflammaging, Periodontitis

Student number: 2015-20508

CONTENTS

LIST OF FIGURES

Chapter I.

Introduction

Backgrounds

Single cell RNA-sequencing in Immunology

An organism's immune system, which consists of immune organs, immune cells, and immunological chemicals, controls a variety of pathophysiological processes, and preserves physiological equilibrium. Immunology is challenging to study, nevertheless, due to the immune system's diversity and the intricacy of the immunological response. The current exploratory approach, which is based on a single experiment, is ineffective and unstable (Furman & Davis, 2015), making it urgently necessary to simultaneously identify the whole immune system's structure and pathological changes tailored to different diseases.

A single cell RNA sequencing (scRNA-seq) technology assesses the whole transcriptome at the single cell level. By identifying the homogeneity and heterogeneity of individual cells, scRNA-Seq overcomes the limitations of traditional approaches. It is valuable as a new tool for bioinformatics analysis since it can collect cellular data on each individual cell and recognize every type of cell in a sample without bias or prior knowledge, which is crucial for comprehending the variety of the immune system.

Uncovering cellular heterogeneity, cell growth and differentiation, cell-cell interaction, hematopoiesis, and gene regulatory networks to predict immune activities have been the main uses of scRNA-seq in immunology. (Y. Chen et al., 2022; Vegh & Haniffa, 2018).

Single cell deconvolution

Most of the pertinent research is frequently hampered by changes in cell type proportions since only the average expression levels are represented by bulk samples of heterogeneous mixes. By understanding how the mix of cell types changes in diseases like cancer, researchers may be able to find specific cell populations that could be used to treat the disease (Elloumi et al., 2011). For instance, immunotherapy may benefit from the latest findings about lymphocytes and other immune cells and their role in the tumor microenvironment (Hendry et al., 2017; Sharma et al., 2019).

A thorough benchmark study (Avila Cobos et al., 2020) gives a quantitative analysis of the cumulative effects of deconvolution outcomes on data processing, scaling/normalization, marker selection, composition of cell types, and technique selection. They tested 20 deconvolution methods, including five that utilize scRNA-seq data as a baseline, to estimate cell type proportions. Then, a number of more deep learning algorithms were given (Molho et al., 2022; F. Yang et al., 2022). but many deep learning-based models frequently fail to be robust because they only tend to overfit the benchmark dataset. Tree-based classifiers can still be used robustly in those cases, with great interpretability of the results (X. Wang et al., 2019).

Chronic inflammation

Immune cells generate cytokines and enzymes during acute inflammation in order to eradicate the cause of the inflammation and begin

healing (Arango Duque & Descoteaux, 2014; Takeuchi & Akira, 2010). Cytokines are proteins produced by immune cells that govern immunological and inflammatory responses. Interleukin and tumor necrosis factor may stimulate the generation of inflammatory mediators and attract immune cells to the site of inflammation.

Immune cells generate proteases and oxidases, which aid in the breakdown of injured tissue and the elimination of infections (Takeuchi & Akira, 2010). DAMPs are molecules that are produced during inflammation and help in tissue repair and regeneration (Vénéreau et al., 2015). *HMGB1* and ATP have the potential to move immune cells and stem cells to the site of inflammation, induce angiogenesis, and improve stem cell growth.

Uncontrolled inflammation may progress to chronic inflammation, which is characterized by ongoing immune activity and the generation of inflammatory mediators (Gilroy & De Maeyer, 2015). Chronic inflammation has been related to cancer, diabetes, and heart disease (Roh & Sohn, 2018). It might be caused by chronic infections, autoimmune diseases, toxins, or cigarette smoke. Chronic inflammation often resolves in an imperfect manner, resulting in ongoing inflammation and tissue damage.

Inflammaging

Inflammaging is a word used to describe the ongoing, low-grade inflammation generated by the innate immune system's activity. Inflammation of this type has been linked to numerous age-related health issues, such as cardiovascular disease, diabetes, and neurological problems. The response of

the immune system to inflammation is based on macrophages, a kind of white blood cell. Immune system dysregulation, namely macrophage activation, has been associated with DNA damage, cellular senescence, reduced autophagy and mitophagy (cellular health-maintenance processes), and alterations in the microbiota (the collection of microorganisms that live in the body). Chronic inflammation of the immune system, resulting in elevated levels of pro-inflammatory proteins such as IL-6 and TNF-alpha, is referred to as inflammaging (Franceschi et al., 2000). This chronic low-level inflammation has been linked to various age-related disorders including cardiovascular disease diabetes , and neurodegenerative diseases like Alzheimer's disease (Bradt et al., 2014).

In addition, DNA damage has been recognized as a marker of inflammation (Vitale et al., 2013). It is thought that DNA damage contributes to the onset of age-related diseases by disrupting normal cellular processes and causing cellular senescence (a state of permanent cell cycle arrest) (J.-H. Chen et al., 2007). Autophagy and mitophagy are cellular health-maintenance processes that remove faulty or superfluous cellular components (Green et al., 2011).

Chapter II.

Compartment specific immune system rearrangement in aging kidney

Introduction

Kidney function tends to decline with age, which has been connected to an increased prevalence of end-stage renal disease that requires treatment through kidney replacement therapy. Acute kidney injury (AKI) occurs more frequently in older people (Baraldi, 1998) and this demographic has a worse prognosis for recovery from AKI (Paraskevas et al., 2010).

It is beneficial to analyze the transcriptome of kidneys using a comprehensive technique such as RNA-seq to acquire a better knowledge of the shared and unique aspects of aging kidneys and renal disease.

One study reported that inflammation-related pathways were considerably upregulated in aged kidneys, utilizing RNA-seq to assess the transcriptome of entire kidneys (D. Park et al., 2016)However, this method has limitations in that it does not account for differences in gene expression between different cell types within the kidney (Shalek & Benson, 2017), nor does it allow for the detection of changes in low-expression genes such as long non-coding RNA (lncRNA) and alternative splicing (Shalek & Benson, 2017).

The transcriptomes of 3 distinct areas within the kidneys of mice were studied using compartment-specific RNA-seq at 3 different stages of life: early (2 months), mid-life (12 months), and late (24 months). We also compared the transcriptomes of aged kidneys to those of kidneys with unilateral ureteral obstruction (UUO) to detect differences and similarities between the effects of aging and a well-known model of kidney fibrosis (D.

Park et al., 2016). We want to learn more about how the kidney ages and how aging and fibrosis are different by looking at the transcriptomes of the different parts of the kidney and comparing them to each other and to the UUO model.

Materials and Methods

Animals

C57BL/6 mice at the ages of 2, 12, and 24 months and CD11c YFP mice were obtained from the Korea Institute of Basic Science and the Jackson Laboratory, respectively. Each group consisted of 12 male mice. Serum creatinine and urea nitrogen levels were measured using the Vet Test 8008 kit from IDEX. Albuminuria and creatinine in urine were measured using ELISA and analysis kits from ALPCO and R&D Systems, respectively, and reported as the albumin-to-creatinine ratio in mg/gCr. A total of 3 mice underwent ureteral ligation surgery to induce UUO, and after 14 days, the left kidney was collected, and the mouse was euthanized. All animal experiments were conducted in accordance with the guidelines of the Animal Research Ethics Committee at Kyung Hee University.

Isolation of glomerular and nonglomerular fraction

Eighteen kidneys were collected from 9 mice per group to sufficiently extract the mRNAs of glomerular and nonglomerular fractions. The total RNA from the entire kidney was extracted from three mice per group. The extracted kidney was quickly frozen and cut with a cryotome to obtain a 10 m sample for the fraction of the entire kidney in the kidney's center. In the case of the glomerular compartment, mice were anesthetized, and diluted 8×107 Dynabeads were injected into 40 ml of phosphate buffered saline through the heart. The kidneys were removed, minced, and digested

with collagenase (Collagenase A 1 mg/ml in HBSS) while gently stirring at 37℃ for 35 minutes. Collagenase digestion tissue was able to gently pass a 100 μm sized cell filter using a flat pestle and then clean the cell filter with 5 ml of HBSS. The cell suspension was centrifuged at 200 × g for 5 minutes. Finally, glomeruli containing Dynabeads were collected in a magnetic particle concentrator and washed at least three times with HBSS. At this time, the supernatant (tube and epilepsy fraction) was carefully pipetted into a separate tube and stored on ice. Kidney tissue, excluding collagenase digestion at 37℃ during the procedure, was maintained at 4°C.

RNA extraction and RNA sequencing

Samples for RNA-seq were chosen based on their meeting certain quality standards, including RNA mass above 1 μg and RNA integrity number above 6, as determined by the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). These samples were then processed according to the Illumina TruSeq protocol (Illumina Inc., CA, USA), which involves fragmenting, reverse transcribing, and amplifying the RNA with a random oligo-dT primer to create a cDNA library. The cDNA library was sequenced using the Illumina HiSeq platform (Illumina Inc., CA, USA), and the resulting raw data was converted to FASTQ format using the bcl2fastq package (Illumina Inc.).

RNA sequencing data analysis

The FASTQ data was then analyzed using STAR software (Dobin et

al., 2013), with unique, properly mapped reads being used for further analysis. Gene expression was measured using transcripts per million (TPM) (Wagner et al., 2012) and annotated using the GENCODE VM16 annotation (Ensembl release 91). Principal component analysis was performed using the sklearn package (Love et al., 2014), while hierarchical and k-means clustering were done with SciPy and the Morpheus tool (https://software.broadinstitute.org/morpheus). The MuSiC package (R-packaged) was used to deconvolve 26 bulk RNA-seq samples and identify the proportions of various cell subtypes (GEO's open single-cell RNA-seq reference and GSE146912). Differential gene expression was analyzed using DESeq2 (Benjamini-Hochberg method), and genes with at least 2 or 4 times variation between samples at a false detection rate of 5% were considered differentially expressed. These differentially expressed genes were functionally annotated using Gene Ontology and KEGG pathways, as well as the Metascape tool (Y. Zhou et al., 2019), through the DAVID web portal (https://david.ncifcrf.gov/summary.jsp). A gene expression network was also established (Dennis et al., 2003), with Pearson correlation used to identify the top 100 correlated genes for each differentially expressed long non-coding RNA (lncRNA), and functional enrichment analysis performed using the GO term annotation from the Ensembl database. All heat maps were created using Morpheus (https://software.broadinstitute.org/morpheus).

Data availability

RNA sequencing data have been archived in NCBI GenBank in accordance with BioProject ID PRJNA672727 (BioSample SAMN16576419 -

SAMN1657647). Use the Personal Reviewer link to view the data.

https://dataview.ncbi.nlm.nih.gov/object/PRJNA672727?reviewer=v3i8pmd9f99jFPfehs3tdl8ds

Results

In this study, the transcriptome was extracted from 27 mice of varying ages (2 months, 12 months, and 24 months) and 3 distinct kidney compartments: glomerulus (G), nonglomerulus (NG), and whole kidney (WK). After generating and aligning a total of 22.9 million to 159 million sequences to the mm10 mouse genome, the data set was further analyzed by eliminating one of the biological replicates (specifically, the third replicate of 12M glomerular compartment) as determined by a PCA grid analysis (L. Chen et al., 2021). Analysis of the glomerular compartments using Principal Component Analysis (PCA) revealed a distinct transcript composition, which was also supported by the results of hierarchical clustering (Figure 2.1B). Markers for specific cell subtypes were used to verify the separation strategy and cell identity (J. Park et al., 2018) (Figure 2.1C).

The glomerular fractions displayed high levels of transcripts for vascular endothelial proteins (*Nrp1* and *Plat*), glucose proteins (*Podxl* and *Synpo*), and mesenchymal proteins (*Sfrp2* and *Pdgfra*). In contrast to the glomerular compartment, genes associated with different nonglomerular proteins and markers of water quality such as *Slc6a12* and *Ghr* were expressed in the nonglomerular and overall kidney compartments. Deconvolution analysis, using publicly available single-cell RNA-seq data

from the kidney (Chung et al., 2020; The Tabula Muris Consortium et al., 2020), was performed to investigate the heterogeneity of cell subtypes and compare the ratios of different cell types. The results for the glomerular compartment samples indicated that podocytes were present in high proportions, while the proportion of mesangial cells increased with age (Figure 2.1D). However, when the single-cell data from the entire kidney was used in the deconvolution process, podocytes were not detected in the RNA-seq samples (Figure 2.1E). No significant changes in cell subtypes with aging were observed in the nonglomerular or whole kidney compartments (Figure 2.1F and G).

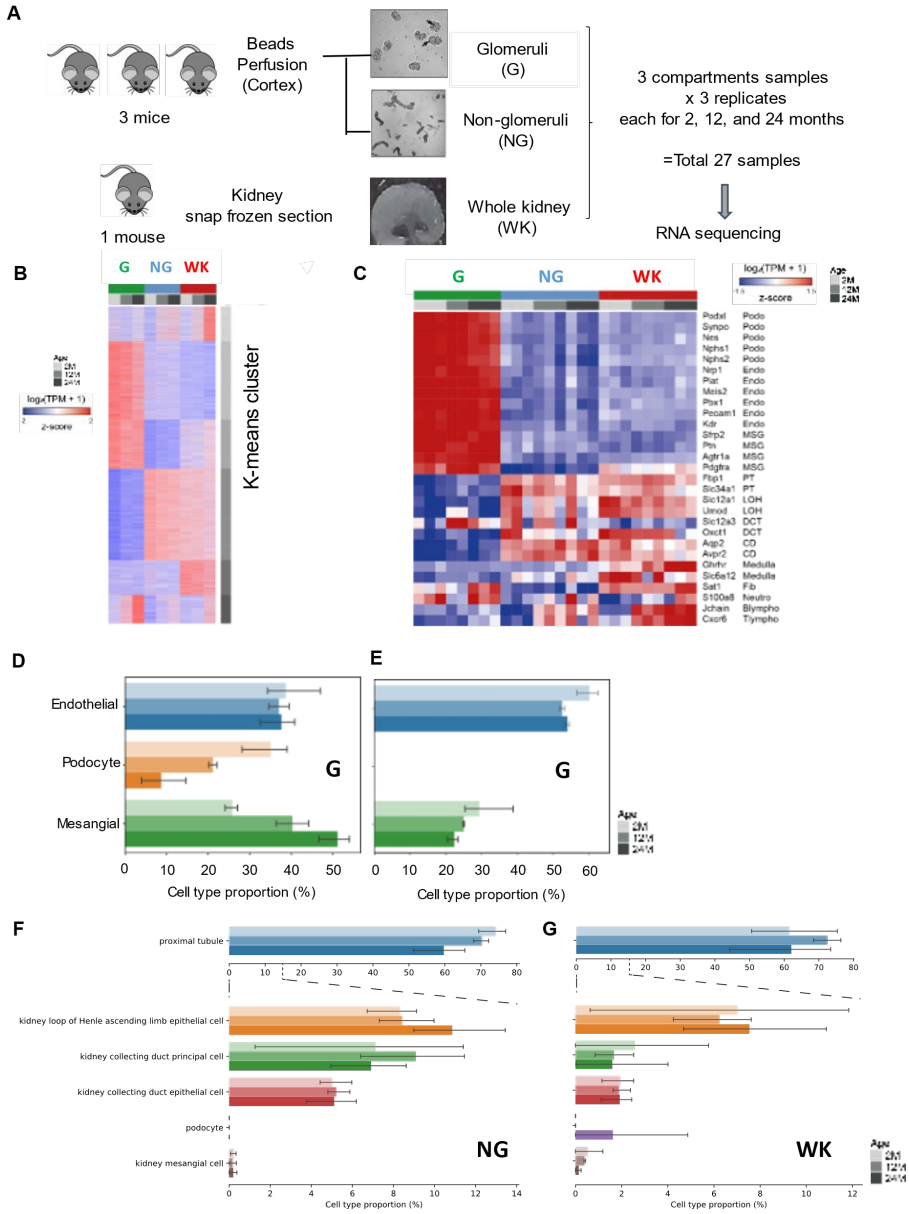

Figure 2.1 Design of experiment and global transcriptome expression patterns.

(A) Schematic of 2 compartments and whole isolated mouse kidneys at various ages. (B) K-means clustering from all detected genes in log2(TPM + 1) values. (C) Heatmap displaying expression of previously identified cell-specific gene markers. Proportions of cell types in the glomerular

compartment based on single-cell references shown in (D), while whole kidney reference in (E). (F) Proportions of cell types in the nonglomerular compartment (left) and whole kidney (right). podocyte; Podo, endothelial cell; Endo, proximal tubular epithelial cell; PT, loop of Henle; LOH, distal convolute tubular cell; DCT, collecting duct, principal cell; CD PC, collecting duct, intercalated cell; CD IC, fibroblast; Fib, natural killer cell; NK, macrophage; Macro, neutrophil; Neutro, lymphocyte; Lympho.

The results of the differential gene expression tests conducted on early (2M to 12M) and late (12M to 24M) phases showed that there were significant changes in gene expression (FDR $\leq$ 0.05) in the glomerular compartments. During the early phase, the upregulated genes were related to ion transfer and immune system processes, while the downregulated genes were involved in responses to lipopolysaccharides and positive regulation of transcription (Figure 2.2A). Some of the specifically upregulated genes in the early phase included *Sost*, *Syp*, *C1q*, and *Gdf15*, which were involved in cell adhesion and complement categories (Figure 2.2B). On the other hand, the downregulated genes during this phase included *Plk2*, *Egr1*, and *Atf3*, which play a role in normal cell division. During the later stages, several genes associated with chemical reactions and immune system processes were activated, while genes involved in ion transfer and transmembrane transport were suppressed (Figure 2.2C). During the later stage of the process, certain genetic factors linked to inflammation, such as *Ifitim3* and *Tnfrsf19*, displayed heightened activity. Additionally, genes that play a role in fibrosis, including *Spon2, Mmp3* and *Angpt4* were seen to be more active during this phase. On the other hand, the activity of *Kcnj1* and *Pvalb*, which are linked to potassium channels and calcium binding, went down at this stage. (Figure 2.2D).

In the nonglomerular compartment, the upregulated genes during the early phase were involved in immune system processes and inflammatory responses, including *Aldh1a1, Igkc, Igha1*, and *Cyp2d12* which are associated with immunoglobulin production and fatty acid metabolism (Figure 2.2F). During the late phase, the main transcriptome features were immune system

processes and phagocytosis (Figure 2.2G), and several genes involved in tubular injury, inflammation, and fibrosis, such as *Lcn2*, *Ltbp2*, *Grem1*, and *Trem2*, were upregulated (Figure 2.2H).

In the whole kidney compartment, a significant number of genes were highly upregulated during the late phase, particularly those involved in immune system processes, inflammatory responses, and chemotaxis (Figure 2.2I, J). Some of the genes that showed the highest level of increased activity were those linked to B cells, which are a type of white blood cell responsible for producing immunoglobulins. These genes included *IgKv*, *Ighv*, and *Fcgr1* (Figure 2.2 K, L).

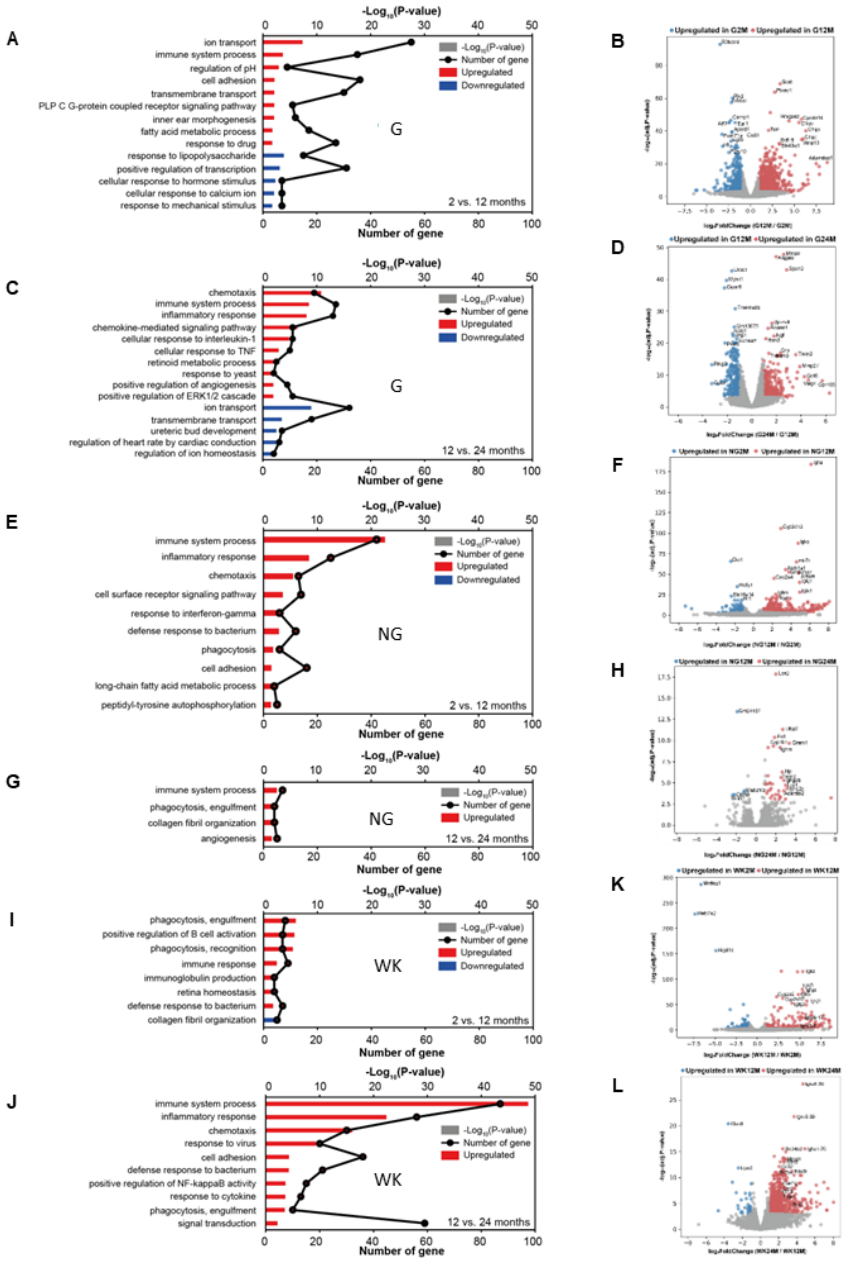

Figure 2.2 Genetic variations in the kidney over time in relation to aging

(A, C, E, G, I, J) Bar charts displaying prevalent GO terms and KEGG pathways related to differentially expressed genes in early and late stages. (B, D, F, H, K, L) Graphs depicting gene expression differences, with red

showing genes that are more active in older mice and blue indicating genes that are more active in younger mice.

Unique inflammatory gene signatures in the glomerular compartments in kidney aging

Our glomerular compartment enables us to study the specific modifications in glomeruli during the aging process. By comparing the expression levels of genes linked to immune system functions, innate immune reactions, inflammatory responses, and chemical reactions, we can observe notable differences between 12 and 24 months of gestation (Figure 2.3A). We concentrated on genes linked to innate immunity, complement pathways, and molecules involved in cell signaling (cytokines and chemokines). We specifically looked at *Gdf15*, *Ccl2*, and *Cxcl13* which play a role in the aging process of the glomeruli. (Figure 2.3B, C). Genes involved in innate immune responses such as *Ly86* and *Spon2*, were significantly upregulated in 24M, on the other hand, *Irf1* was downregulated. One of the primary complement pathways that undergo changes during the aging process is the classical pathway. Specifically, components of this pathway known as *C1q*, including *C1qa*, *C1qb*, and *C1qc*, become more active. (Figure 2.3 F, G). Antigen presentation for immune response via major histocompatibility (MHC) class II and cytokine macrophage migration inhibitors (MIFs), i.e., cell surface receptors for *Cd74*, *H2-Eb1*, and cathepsin S (*Cts*), were also upregulated in glomerular compartments. Through scRNA-seq data, we observed distinct inflammatory gene expression patterns in podocytes, endothelial cells, and mesangial cells across the aging process (Figure 2.3H).

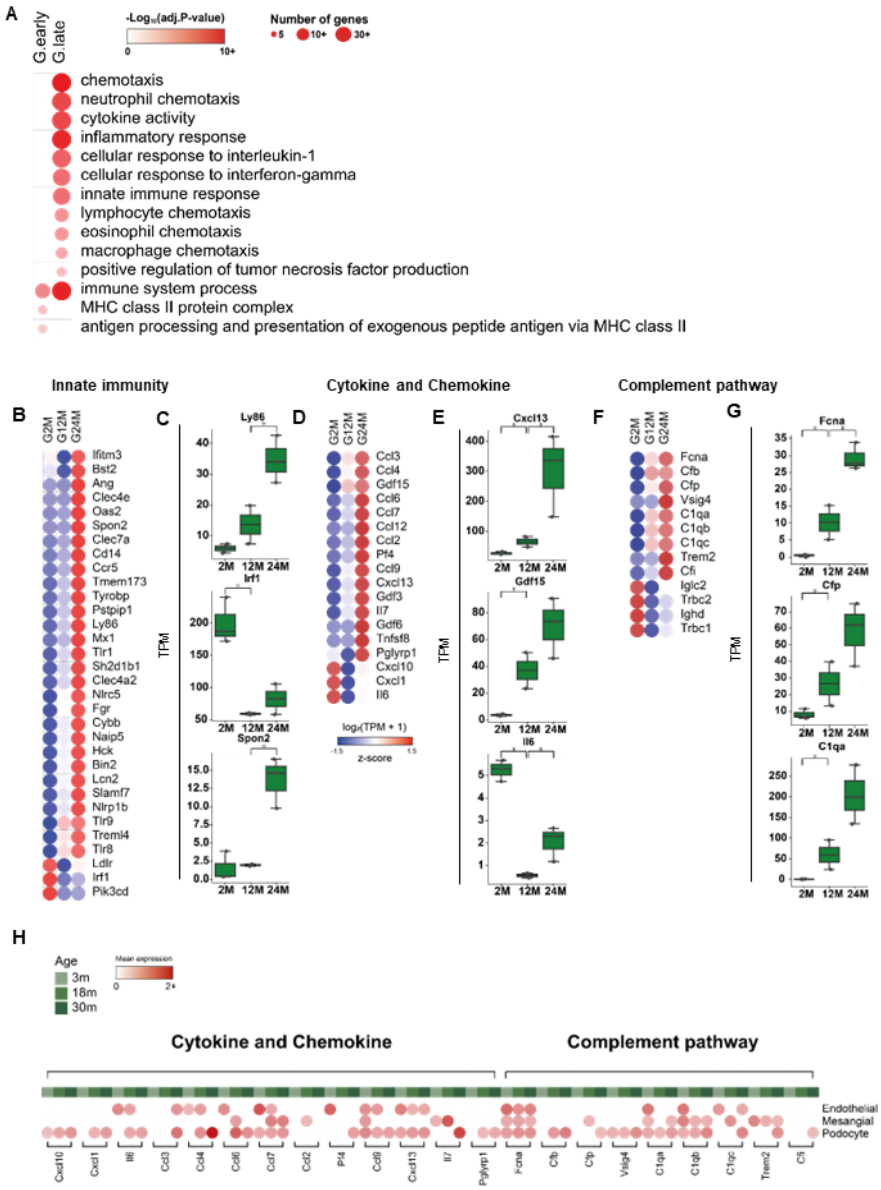

Figure 2.3 Inflammatory gene expression in the glomeruli upon aging

(A) A visual representation of significantly prevalent inflammatory biological processes in the glomeruli, using -log2(adjusted P-value) for transparency and the number of genes for the size of the representation. (B, D, F) Genes with

notable increase in activity during the period between 2-24 months, represented in heatmaps and linked to specific GO terms. (B) Innate immunity. (D) Cytokine and chemokines. (F) Complement pathway. (C, E, G) Graphical representation of the RNA-seq expression levels of specific differentially expressed genes associated with each prevalent inflammatory biological process (H) A heatmap of genes associated with prevalent inflammatory biological processes using single-cell reference data. A one-way ANOVA was applied, followed by Tukey's test.

Investigating the potential functions of lncRNA in the aging of the kidney

Studies conducted in recent times have demonstrated that long non-coding RNA(lncRNA), have a substantial impact on the management of gene expression, including transcription and translation regulation, genome imprinting, and epigenetic regulation (Statello et al., 2021). In the release of GENCODE VM16, over 12,000 long non-coding RNAs were identified and out of those, 9,146 were observed to be active in a group of samples. Among the 250 lncRNAs that displayed variations in expression across all compartments, 74 were found to have a strong association with protein-coding genes that also displayed variations in expression through co-expression network analysis. This analysis revealed that upregulated lncRNAs were functionally linked to the immune system, inflammation, blood coagulation, transport, and lipid metabolism (Figure 2.4A). In opposition, lncRNAs that had a decreased level of expression were linked to the processes of transcribing genetic information and arranging the spindle fibers during cell division (Figure 2.4B). The expression patterns of individual lncRNAs (Figure 2.4C, E, G) were highly similar to those of the correlated neighboring genes, and many differentially expressed genes (DEGs) were major components of the coexpression network for each lncRNA (Figure 2.4 D, F, H) (Shalek & Benson, 2017).

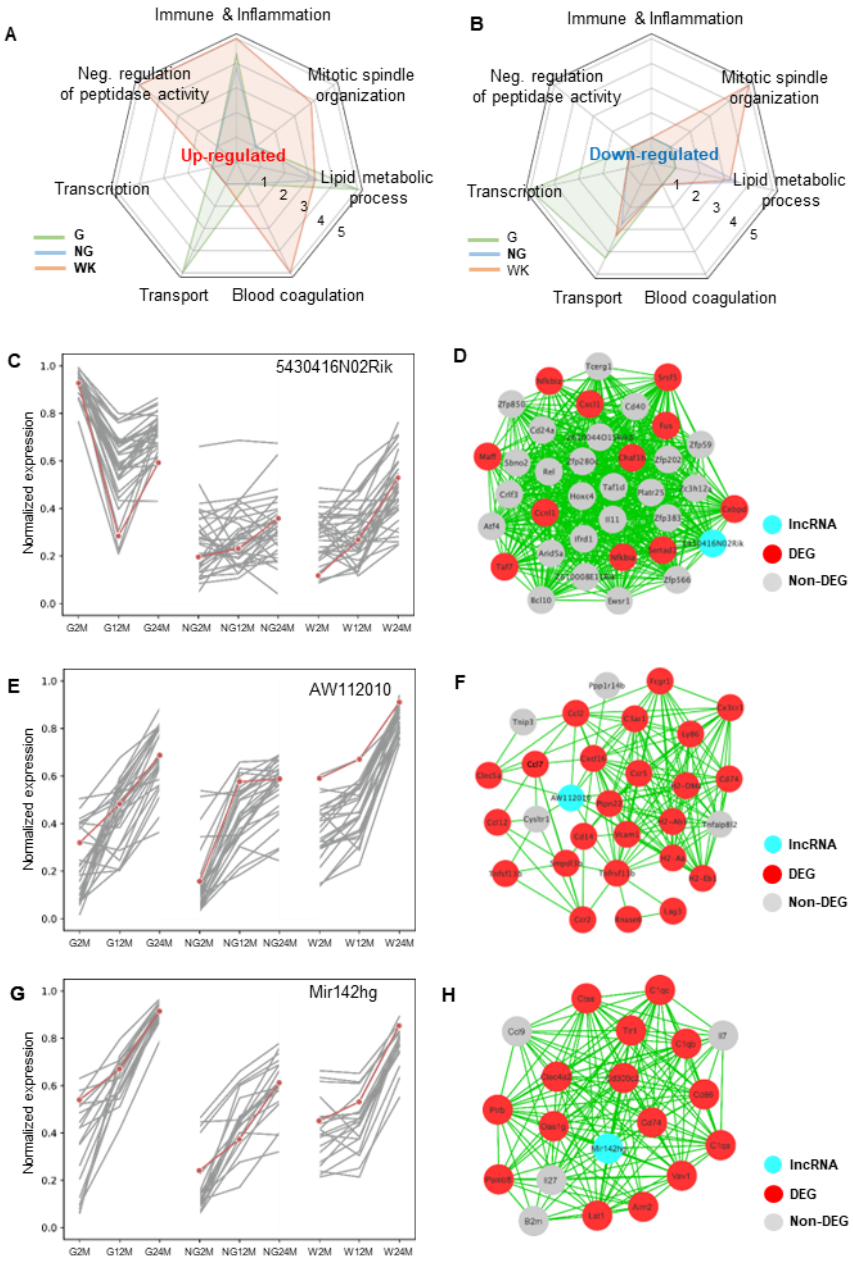

Figure 2.4 Long non-coding RNA in the aging process of the kidney

(A) and (B), Rader plots are displayed, showcasing enriched GO terms linked to DEGs that are significantly correlated with lncRNAs that are either upregulated or downregulated, respectively. (C), (E), and (G)

showcase the coexpression patterns and networks of selected lncRNAs, with the core coexpression network depicted. The networks are shown for three compartments and three timepoints, specifically 2M, 12M and 24M, with red lines indicating the lncRNAs and grey lines representing the neighboring genes. (D), (F), and (H) also show the coexpression pattern of the lncRNAs. Coexpression networks of lncRNAs. Genes with high correlation in the network are represented as green edges, a correlation of 0.9 or greater. lncRNAs are colored cyan in network representations. Red indicates DEGs in the network.

Comparison of gene expression changes in aging and UUO-induced fibrotic kidneys reveal both shared and unique alterations

We have examined the concept of kidney aging as a precursor to chronic kidney disease (CKD). By analyzing the genetic makeup of both aged and injured kidneys, the study aimed to understand the similarities and differences in fibrosis, a condition that contributes to both aging- and injury-related kidney damage. Largely, gene expression profiles in 24M kidneys were distinguished with the ones in the UUO kidneys (Figure 2.5A). The expression of cell subtype-specific markers significantly increased in immune cells, but decreased in tubular cells, and proportionally increased in glomerular cells in UUO kidneys compared to 24M kidneys (Figure 2.5B).

Despite selecting genes that exhibited a fourfold increase in expression, a significantly greater number of genes were found to be either increased or decreased in the UUO group. (Figure 2.5C) Genetic Interactions were further analyzed to identify shared genes in healthy kidney aging and UUO-induced kidney fibrosis (Figure 2.5D). A group of 40 genes that play a significant role in controlling cell growth, differentiating cells, and supporting embryonic development have been identified. Among these genes are *Nr2e3*, *Ak1c18*, *Wnt8*, and *Wntl1*, which have been found to have a suppressing effect on cellular activity. By comparing GO term enrichment between 24M and UUO, we found features that distinguish kidney aging from injury-induced kidney fibrosis (Figure 2.5E).

Many ECM tissues and angiogenesis genes were highly upregulated in UUO while oxidative phosphorylation, immunoglobulin production, and

classical complement pathways were major upregulated GO terms at 24M. Adaptive immune cell regulation revealed the difference between healthy kidney aging and UUO. T cell-related genes were mostly upregulated from UUO (Figure 2.5F), while B cell-related genes were upregulated from 24M and UUO compared to 2M. A number of genes that are linked to B-cell activation were found to be significantly more active in the case of UUOs in response to unfolded proteins. Examples of such genes include *Xbp1* and the *Edem* family of proteins, which are known to enhance degradation within the endoplasmic reticulum. These genes were only upregulated in UUOs, and not in other conditions (Figure 2.5G,H).

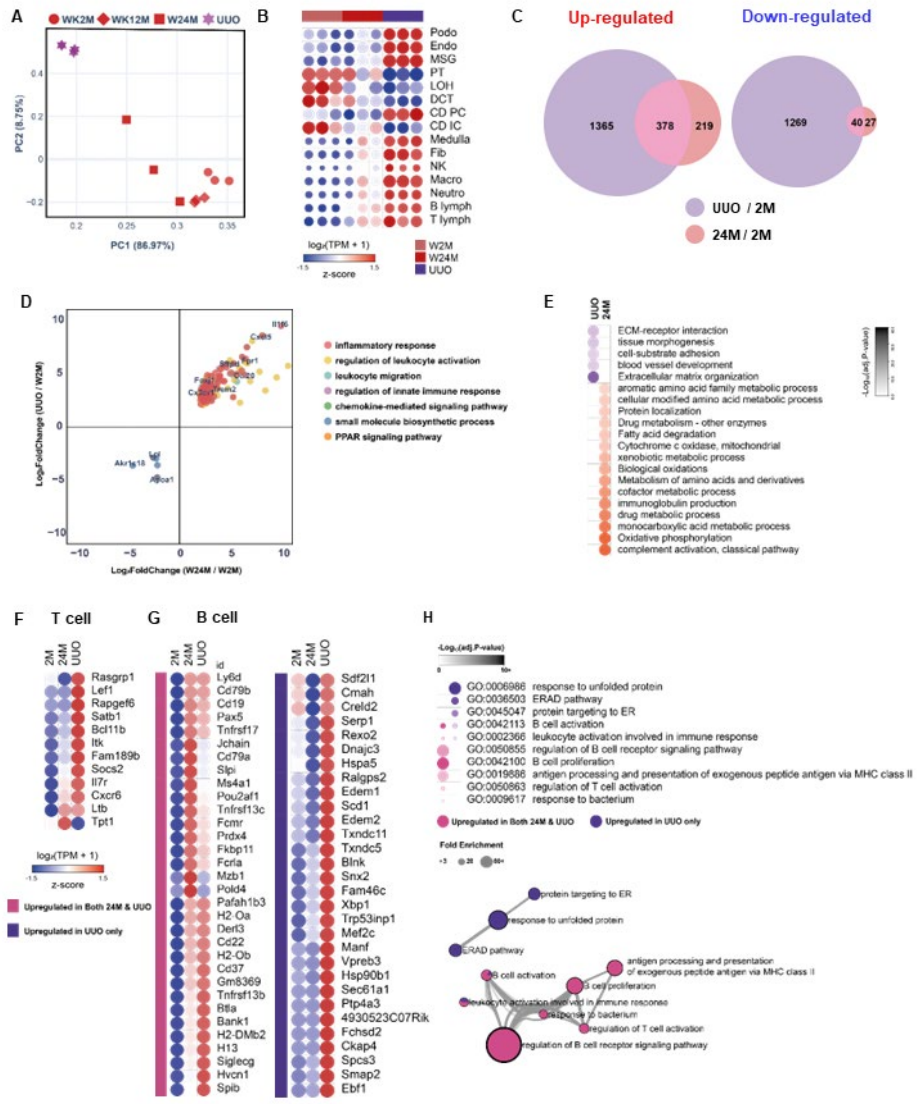

Figure 2.5. A comparison of inflammatory characteristics in aged and fibrotic kidneys resulting from UUO injury

(A), a principal component analysis is conducted on the kidneys from three groups, UUO, 24M, and 2M. (B) illustrates the expression levels of known cell type-specific markers through a heatmap. The differentially expressed genes (DEGs) between UUO and 24M kidneys in comparison to 2M kidneys are illustrated in (C) through Venn diagrams. (D) Scatterplot

displays the log2FC values of all DEGs in the 24M and UUO groups compared to the 2M group, with the top 7 functional terms in relation to size presented in the legend. The circle size in the scatterplot represents the number of enriched functional terms compared to the 2M group. (E) presents a heatmap of significantly overrepresented functional annotations for DEGs between the UUO and 24M groups, while (F) and (G) depict heatmaps of DEGs for each overrepresented Gene Ontology term for T cells and B cells, respectively. (H) A heatmap and network of functional annotations were created for DEGs from UUO and 24M upregulation and DEGs specifically upregulated in UUO. These annotations were found to be highly overrepresented. podocyte; Podo, endothelial cell; Endo, proximal tubular epithelial cell; PT, loop of Henle; LOH, distal convolute tubular cell; DCT, collecting duct, principal cell; CD PC, collecting duct, intercalated cell; CD IC, fibroblast; Fib, natural killer cell; NK, macrophage; Macro, neutrophil; Neutro, lymphocyte; Lympho.

Discussion

As we age, there are a plethora of molecular alterations that occur within the kidneys (Lim et al., 2012). Different parts of the kidney, such as the glomerular and tubulointerstitial sections, are known to work together to perform specific functions. However, the specific impact of aging on each part of the kidney has not yet been studied.

Our study suggests that new methods are necessary to understand the specific effects of aging on the glomerular and tubulointerstitial parts of the kidney. We found a large number of genes that are differently expressed in each compartment, surpassing previous studies. (D. Park et al., 2016). By analyzing transcriptomic data from well-defined compartments at different time points, we were able to uncover various pathways related to renal aging. We also compared the gene expression levels between the glomerular fraction, nonglomerular fraction, and whole kidney to examine their distinct inflammatory characteristics.

The analysis of long non-coding RNAs (lncRNAs) also revealed tissue-specific regulation and previously unreported genetic characteristics during kidney aging (Ignarski et al., 2019). According to Wang et al. long non-coding RNAs (lncRNAs) play various roles in gene regulation and may be connected to the development of diseases (Y.-N. Wang et al., 2021). Also the study at 2021 found that several lncRNAs, such as *H19*, *Pvt1*, and *Sng5*, are associated with kidney disease (Moreno et al., 2021). Xie et al. discovered that the expression of *H19* lncRNA was significantly increased in both in vitro

HK-2 cell fibrosis and in vivo UUO-induced fibrosis models, compared to the 2M model (Xie et al., 2016). By analyzing the correlation between lncRNAs and differentially expressed genes, it is possible to infer the potential function of lncRNAs, even if the role is only tentative. For example, *5430416N02Rik* lncRNA was downregulated in the glomerular compartment at the early stage and was related to transcription functions such as "positive regulation of transcription from RNA polymerase II promoter" and "transcription, DNA-template" through coexpression with other genes (T. Zhao et al., 2020). Additionally, the gene 5430416N02Rik has been discovered to play a role in the growth and reproduction of embryonic stem cells through its involvement in interactions between different chromosomes. On the other hand, the *AW112010* lncRNA was upregulated in the nonglomerular compartment at the early stage and in the whole kidney at the late stage (X. Yang et al., 2020). It was closely related to many immune-related genes and may be linked to immune system functions. Specifically, the expression of *AW112010* was found to be connected to the inflammatory condition of T cells by decreasing *IL-10* expression through histone demethylation.

Upon aging, the transcriptome of glomeruli (Lai et al., 2019) undergoes significant changes compared to that of the nonglomerular compartment and whole kidneys. These changes often involve activation of inflammatory reactions and immune system processes, with cytokines, chemokines, and classical complement pathway activation being particularly prominent in aging glomeruli (Lähnemann et al., 2020). In contrast, in the whole kidney (which is made up of more than 90% nonglomerular tissue), we

see an increase in genes related to immunoglobulin production and plasma cell activation. It is worth noting that the transcriptomes of samples defined as nonglomerular compartments (such as tubules) and the whole kidney tend to differ, possibly due to the influence of glomerular signals or the loss of "gap" signals during nonglomerular sample collection. One possible explanation for these differences is that aging nephrotic cells may enhance anti-inflammatory properties and increase the recruitment of innate and adaptive immune cells. Previous research has also identified inflammation and fibrosis as major active pathological processes in the glomeruli of aged rats (Lai et al., 2019). The heightened activity of the complement system and increased presence of innate immune cells in the aged glomerulus may contribute to the higher incidence of rapid progressive glomerular nephritis seen in older patients (Lähnemann et al., 2020)

In this study, we have shown that the immune response in the kidneys of the elderly undergoes changes over time. These changes can be observed in various functional compartments of the kidney and may contribute to the maintenance of homeostatic equilibrium in the elderly. Our findings suggest that the immune system in the elderly kidney is more complex than previously thought, particularly in the case of chronic kidney disease (CKD). This study makes a valuable contribution to our understanding of the immune system in the elderly kidney and its role in maintaining homeostasis. Future research should focus on further exploring the mechanisms underlying these changes in the immune response and their potential impact on health outcomes in the elderly population.

Chapter III.

RNA sequencing of gingival biopsies reveals molecular signatures reflecting periodontal health status

Introduction

Periodontitis is a serious oral health problem that occurs when specific bacteria in the mouth cause destruction of the periodontal tissue and eventual tooth loss. These bacteria create an imbalanced microbial biofilm, which has traditionally been evaluated by dental professionals through various methods such as measuring plaque, gingival inflammation, bleeding during examination, loss of tooth attachment, and the depth of pockets around the teeth. However, these methods do not effectively identify the root causes of the disease (Socransky et al., 1998). Recent advances in next-generation sequencing have allowed researchers to identify additional periodontal pathogens and the concept of "polymicrobial synergy and dysbiosis" (Jeon et al., 2020), in which the collective activity of the microbiome, including keystone pathogens, disrupts the normal immune response and causes tissue destruction. It is worth noting that individual susceptibility to these pathogens varies and can impact the disease's progression.

To better understand the molecular changes that occur in periodontitis, researchers have utilized transcriptome analysis through techniques such as microarrays and RNA sequencing (RNA-seq). In 2008, a study using microarray technology was conducted to identify genes that may be involved in the progression of periodontitis. However, the findings were inconclusive and varied, with some studies discovering notable differences in gene expression between healthy and diseased gingival tissues (Demmer et al., 2008) and others not observing such differences (Papapanou et al., 2004).

RNA-seq, on the other hand, is an unbiased method with high accuracy in detecting gene expression (Y.-G. Kim et al., 2016) The present study utilized RNA-seq to identify molecular signatures, or biomarkers, that reflect periodontal health status and suggest that RNA-seq may be useful as a diagnostic and predictive tool for monitoring periodontal conditions.

Materials and Methods

Study population and clinical evaluation

The research study, conducted by the Department of Periodontics at Pusan National University Dental Hospital in Yangsan, Korea, included 67 participants. The control group consisted of individuals with clinically healthy periodontal tissues, as indicated by low bleeding scores on probing at 10% of sites and the absence of lesions with probing depths (PD) greater than 3 mm or clinical attachment loss (CAL). The periodontitis group was initially classified based on CAL greater than 3 mm and severe bone loss as seen on radiographic images. A full-mouth clinical examination, including assessments of PD, CAL, gingival index (GI), and plaque index (PI), was conducted by a single practitioner. A periodontal expert then classified the severity of periodontitis into either the moderate or severe category based on the findings of the full-mouth clinical exam.

RNA extraction from gingival tissue and whole transcriptome sequencing (RNA-Seq)

Gingival tissue samples were treated with phosphate-buffered saline and stored at -80°C until RNA isolation using the RNeasy Mini Kit (Qiagen Inc., Valencia, CA). cDNA libraries were constructed using the TruSeq Stranded mRNA LT Sample Prep Kit. The protocol included the extraction of polyA-selected RNA, fragmentation of the RNA, and reverse transcription with random hexamer priming. The libraries were quantified using qPCR and

evaluated with the Agilent Technologies 2100 Bioanalyzer. The prepared sequencing libraries were then sequenced on the Illumina Novaseq platform in paired-end mode with 2x100 base pair reads.

RNA-Seq read data processing and deconvolution analysis

Before conducting the mapping procedure, the sequencer's raw reads were preprocessed to eliminate low-quality and adapter sequence. The processed readings were aligned using HISAT v2.0.5 (D. Kim et al., 2015) to the human genome (hg19). The reference genome sequence of human genome (hg19) and gene annotation data were downloaded from the UCSC table browser (http://genome.uscs.edu). Transcript assembly and abundance estimation using StringTie (Pertea et al., 2015). After alignment, StringTie v1.3.3b and RefSeq gene annotations were used to assemble aligned reads into transcripts and to estimate their abundance. The normalized abundance of the annotated genes was estimated as FPKM (Fragments Per Kilobase of exon per Million fragments mapped) values in each sample and Principal component analysis (PCA) was performed in python via genes with top 1000 variance through sklearn library (Pedregosa et al., 2011). RNA-Seq samples were deconvolved into each cell type proportion by the R package MuSiC with default parameters (X. Wang et al., 2019), via public single cell RNA-seq reference from GEO accession number GSE152042 (Transcriptomic profiling of human gingiva in health and disease) (Caetano et al., 2021). Cells with estimated proportions of zero were excluded from further analysis. The estimated proportion matrices of each RNA-seq samples were grouped by

each labeled disease phenotype for downstream analysis.

Differential gene expression test and enrichment analysis

DESeq2 R package(Love et al., 2014, p. 2) was used to identify differentially expressed genes between groups. At a false discovery rate (FDR) of 5%, genes with changes of at least 2-fold across samples were considered differentially expressed. All enrichment analysis and functional annotations were carried out using Metascape (Y. Zhou et al., 2019) and only the "GO Biological Processes" category has been used to perform functional annotation process to avoid semantic twin terms in other functional pathway databases. A “selective GO cluster” method was applied to choose top clusters as described in the metascape protocol.

GO term cluster network analysis and visualizations

GO term cluster networks were generated based on the metascape functional annotation results. Briefly, every GO term in each cluster was merged into a single node in the network and labeled as a representative GO term. A size of node in the network was based on the number of enriched genes for each GO term cluster. A pie-chart like inner circle in each GO term node was produced based on the proportion of gene count for each group. If a kappa score between GO term nodes was greater than 0.3, an edge connected those two nodes for the purpose of clear visualizations. The subnetwork of GO term cluster networks was defined as a functional “theme” based on a parent GO term, the proportion of shared genes between nodes, and semantic

definitions of GO terms.

PCA and volcano plots were generated by python plotly library (v 4.3.0) and venn diagrams were produced by python library matplotlib-venn and pyvenn. Heatmaps were generated using the online tool Morpheus(https://software.broadinstitute.org/morpheus). The GO term networks were plotted using cytoscape (v3.7.0) (Shannon et al., 2003).

Correlation analysis of gene expression and the clinical parameters

To assess the associations between the expression of gene in the gingival tissue and four clinical parameters, spearman correlation coefficient was calculated. If the p-value for the correlation was less than 0.01 and the correlation coefficient was greater than 0.45 for all clinical parameters, we defined that the gene was highly associated with the clinical parameters.

Development of the clinical severity classifier from transcriptome and biomarker prioritization

Input data consists of normalized counts of all genes as function variable and severity as target variable. Feature variables were scaled through scikit-learn's StandardScaler module and target variable were encoded as ordinal (0: healthy,1: moderate,2: severe). The feature selection process was initially performed with python package 'Minimum-redundancy-maximum-relevance'(mRMR) (Z. Zhao et al., 2019) to extract top 100 feature that has minimum correlation between each gene expression but also have maximum strength(F-statistics) with severity. The input data, filtered by 100 features

from mRMR, was used as input to PyCaret's AutoML pipeline. An additional feature selection (multicollinearity_threshold=0.8) was applied within PyCaret to remove high multicollinearity within features and reduce the number of features from 100 to 44. Train and test sets were split into 0.88 ratio (train:58, test:9) due to the small size of dataset. Stratified cross validation with 10-fold were applied to train set within Pycaret. Different classifier models (Extra Trees, Random Forest, Logistic Regression, CatBoost, Light Gradient Boosting Machine, Gradient Boostin, Decision Tree) were applied and leader board of 7 evalucation metrics (Accuracy, Area under the receiver operating characteristic curve (AUC), Recall, Precision, F1, Kappa, Matthews correlation coefficient (MCC) were generated. SHapley Additive exPlanations(SHAP) values of the best model, Extra trees classifier, were used to sort the priorities of the features. Features with high SHAP values are considered as plausible biomarkers.

Ethics Statement

All donors gave written informed permission, and their rights were safeguarded in accordance with the procedure evaluated and approved by Pusan National University's Institutional Review Board (IRB No. PNUDH-2017-023). This study was conducted in accordance with the Helsinki Declaration as revised in 2013.

Results

Study group samples and RNA sequencing (RNA-Seq) profiles for the study groups

Sixty-seven study subjects; 17 healthy subjects and 50 periodontitis patients were recruited, and the gingival tissue samples were obtained. We divided the periodontitis patients into two different groups; moderate (32) and severe (18) groups based on the parameters from the full-mouth clinical examination (Table 1). The RNA was extracted from the gingival tissue samples and RNA-Seq has been performed to compare the transcriptional expression profiles among the four groups including a healthy group. The sequencing raw reads were preprocessed to evaluate the transcriptional expression. Briefly, the raw read alignment was performed using HISAT2 and hg19 human genome and StringTie and RefSeq gene annotations were used to estimate the gene expression. In total 24,843 genes were expressed in all samples.

We conducted principal compartment analysis (PCA) to characterize the dimension that explains the most variance in the transcriptome expressions (Figure 3.1A). The PCA results shows that 88% of the total variance in gene expression is explained by the first two principal components (PCs). When we divided the PCA plot into two areas based on the second component (PC2), most of the healthy subjects were in the lower region. On the other hand, most patient samples in the severe group were in the upper region. We inferred that the healthy group and severe group could be clearly distinguished based on

the expression profiles. In case of moderate group, the samples were widespread. Therefore, based on the sample distributions and their associated clinical groups, the PC2 could reflect the periodontitis disease progress. Similarly, unsupervised hierarchical clustering analysis for all subjects revealed high correspondence between gene expression profiles and disease status (Figure 3.1B).

Furthermore, to check the heterogeneity of the RNA-Seq samples and compare the cell-type proportions among groups, we conducted computational deconvolution analysis for all samples from individual 67 subjects using recent single-cell expression datasets, 12,422 isolated live cells in human gingival tissues (Caetano et al., 2021). Interestingly, the results shows that the epithelial cell proportions were gradually decreased along with the disease progress from mild to severe (Figure 3.1C). Instead, the populations of endothelial, transit-amplifying, stromal and immune cells such as B and T cells, were increasing during the progression and persistence of disease as similarly shown in the previous study (Luo et al., 2021; Rangel-Huerta & Maldonado, 2017; Williams et al., 2021).

Table 1. Participants' demographics and clinical parameters (mean ± standard deviation)

Characteristic	Healthy (n=17)	Periodontitis (n=50)	
		Moderate (n=32)	Severe (n=18)
Age (years)	32.82 ± 13.21	51.13 ± 11.28	51.22 ± 8.48
Gender	[1]M: 9, [2]F: 8	M: 13, F: 19	M: 7, F: 11
Full mouth examination			
[3]PD (mm)	2.33 ± 0.28	3.57 ± 0.78	4.13 ± 1.23
[4]CAL (mm)	2.34 ± 0.28	3.98 ± 0.95	4.65 ± 1.35
[5]GI	0.12 ± 0.08	0.95 ± 0.68	0.95 ± 0.50
[6]PI	17.07 ± 19.27	58.91 ± 24.71	71.51 ± 20.18

[1]M: Male, [2]F: Female, [3]PD: probing depth, [4]CAL: clinical attachment level, [5]GI: gingval index, [6]PI: plaque index

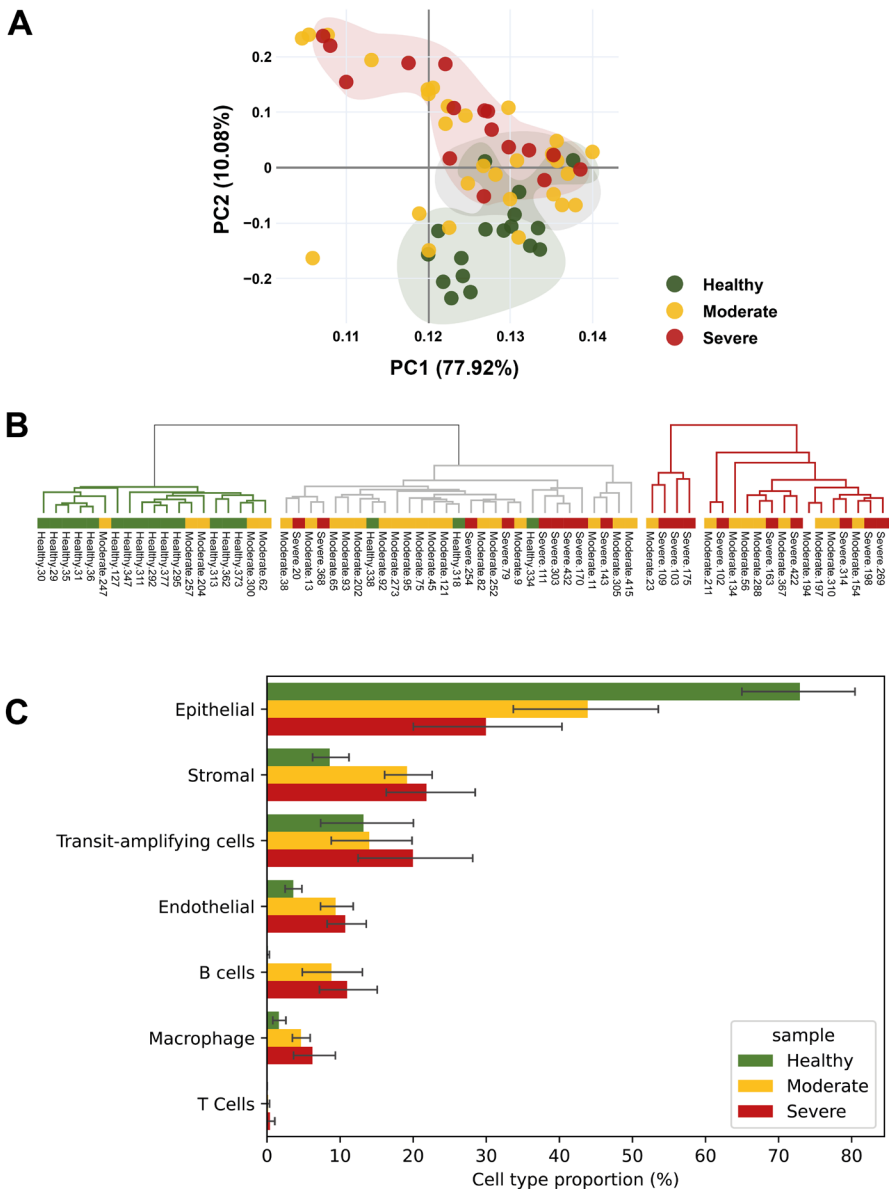

Figure 3.1. Expression profiles of global transcriptome

(A) Principal component analysis (PCA) based on $\log_2$(TPM +1) that describes global transcriptomic expression profile. The two shades (red and green color) represent the clusters of subject status and the grey shade region is the intermixed proportion of all subjects. (B) The hierarchical clustering of all samples. (C) The predicted cell type proportions based on

deconvolution analysis.

Comparing gene expression between healthy individuals and those with periodontal disease

To identify gene signatures and associated biological implications in the periodontal patient groups, we performed differential gene expression tests between each periodontal patient group and healthy subject group. To identify genes that are differentially expressed between two groups, we established the following criteria: a fold change of at least 1 in the log2 scale and a false discovery rate of no more than 0.05. Genes that meet these criteria are referred to as differentially expressed genes (DEGs). In both comparisons, healthy vs moderate and healthy vs severe, the number of upregulated genes in the periodontitis groups were greater than the number of downregulated genes (Table 2 and Figure 3.2A and 3.2B). We identified that the total number of DEGs in the comparison between the severe patient group and healthy subject group was greater than the total number of DEG in the comparison of healthy and moderate group, which agrees with the observation in the PCA analysis. For both pairwise comparisons, the number of upregulated genes for the patient group are about 2~3 times larger than the downregulated genes.

Top 50 upregulated genes included immune related genes (*CHIA, IFNK, ANKRD1, CCL7, CXCL5, IGLL1, MMP3, ORM1, ORM2, TNR, CCL25, INHBE, GH1, SST*) and binding molecule genes (*ACTN3, ACOXL, UGT2B17, UCP1, CRYBA4, OIT3, PCDHA2, PVALB*). The 50 most downregulated genes included nucleotide metabolic process genes (*EGF, NOS1, PLP1, FLG2, ELOVL4, H19, ST6GAL2, GAPDHS, MLXIPL, UGT1A8, RORC, HMGCS2,*

PRKAA), signaling and transport related genes (*CACNG4, COBL, MAP2, EXPH5, CTTNBP2, CDH19, CRLF1, IL12RB2, NPR3, PCSK2, ARC, SH3GL3, PKDREJ, TRPM1, PMEL, SLC24A5, EGF, NOS1, PLP1*) and cornification related genes (*COBL, MAP2, EXPH5, DCT, DSC1, EPB41L4B, CLDN20, SLITRK5, CDSN, KRTAP3-2, KRT40, KRT31, KRT3, ASPRV1, UPK1B, TRIM63, NEFM, PLP1, FLG2*). Next, we assess the genetic changes in the functional aspects by Gene Ontology (GO) term enrichment test for each pairwise comparison. As shown in Figure 3.2C and 3.2D, there were common enriched functional pathways related to immune system processes such as inflammatory response, leucocyte migration, myeloid leucocyte activation and positive regulation of immune response. Meanwhile, the enriched pathways for the downregulated genes were peptide cross-linking and water homeostasis are signs of functional and structural decline in the tight-fitting gingival epithelium. Additionally, downregulated xenobiotic metabolic pathways suggest a lack of immunological response.

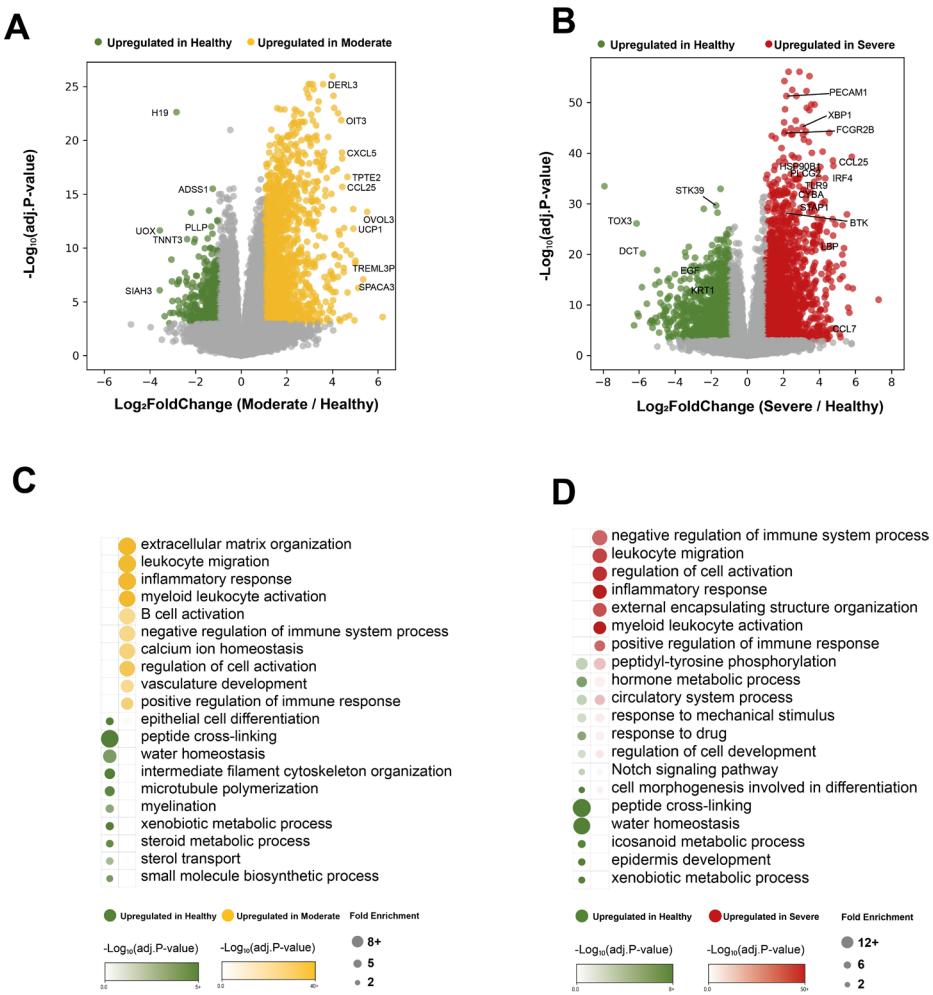

Figure 3.2. Differential gene expression analysis

(A) A volcano plot showing the expression levels of various genes in healthy and moderate individuals, with significantly differentially expressed genes highlighted in green for healthy and yellow for moderate. (B) A visualization of gene expression differences between healthy and severe individuals, with significantly differentially expressed genes marked in green for healthy and red for severe. (C) A list of Gene Ontology terms enriched in the comparison of healthy and moderate individuals, with the circle size representing the number of genes in the term and the adjusted p-value displayed through a

color gradient. (D) A presentation of Gene Ontology terms enriched in the comparison between healthy and severe groups, with the size of the circle indicating the number of genes in the term and the adjusted p-value represented through color.

To delineate the detail gene regulations in the disease progression, we compared the up- and downregulated genes between both pairwise comparisons. In case of upregulated genes, 1,520 genes were common DEGs in both moderate and severe groups and 159 and 726 genes were group-specific DEGs in moderate and severe groups respectively (Figure 3.3A). In case of the downregulate genes, a significant number of genes (905) specifically belong to severe group (Figure 3.3B). Functional annotation analysis for the group-specific upregulated genes showed that many GO terms associated with immune responses were enriched severe group specific genes, on the other hand, GO terms associated with regeneration processes such as “muscle structure development”, “extracellular matrix organization” and “tissue morphogenesis” were enriched in the moderate group specific genes (Figure 3.3C). In case of the group specific downregulated genes, there are many GO terms enriched in severe group specific genes because of high number of group specific genes and most of the GO terms are related in cellular homeostasis (Figure 3.3D).

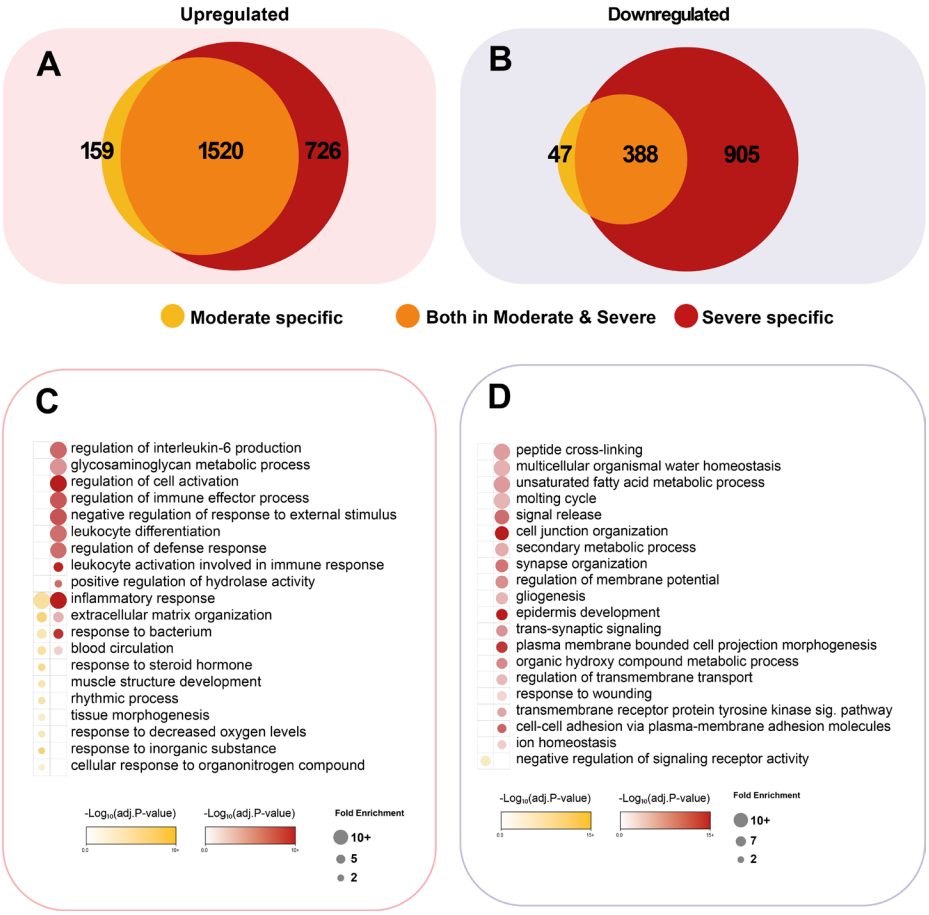

Figure 3.3. Periodontist group specific DEGs

(A) A Venn diagram of upregulated genes in moderate (yellow) and severe (red) periodontitis status. (B) A Venn diagram of downregulated genes in moderate (yellow) and severe periodontitis status. (C) Enriched GO terms for the periodontitis group specific upregulated genes. Yellow: Moderate group specific GO terms, Red: Severe group specific GO terms. (D) Enriched GO terms for the periodontitis group specific downregulated genes. Yellow: Moderate group specific GO terms, Red: Severe group specific GO terms.

Identification of genes significantly correlated with clinical parameters

To identify the expressed genes in RNA-Seq data linked with four clinical measures (PD, CAL, GI, and PI), correlations were established between the normalized gene expression abundance and the clinical measurements using the spearman correlation technique. When we established the criteria for the strong relationship between gene and parameter according to the method, 2,035 genes were positively correlated and 534 were negatively associated with all four parameters. As the Venn diagram depicts (Figure 3.4), 459 genes were positively correlated with all 4 clinical parameters. Functional enrichment analysis shows typical immune system process such as lymphocyte activation, mature B cell differentiation, leukocyte activation. And cell redox homeostasis (Morris et al., 2022). It's interesting to note that endoplasmic reticulum (ER) stress-related terms were also prevalent, which are known to be the cause and the effect of the chronic inflammation (Chipurupalli et al., 2021; Hasnain et al., 2012). N-linked glycosylation has also been investigated for its potential autoimmune associations. or causing ER stress with misfolded protein (Pandey et al., 2022; X. Zhou et al., 2021).

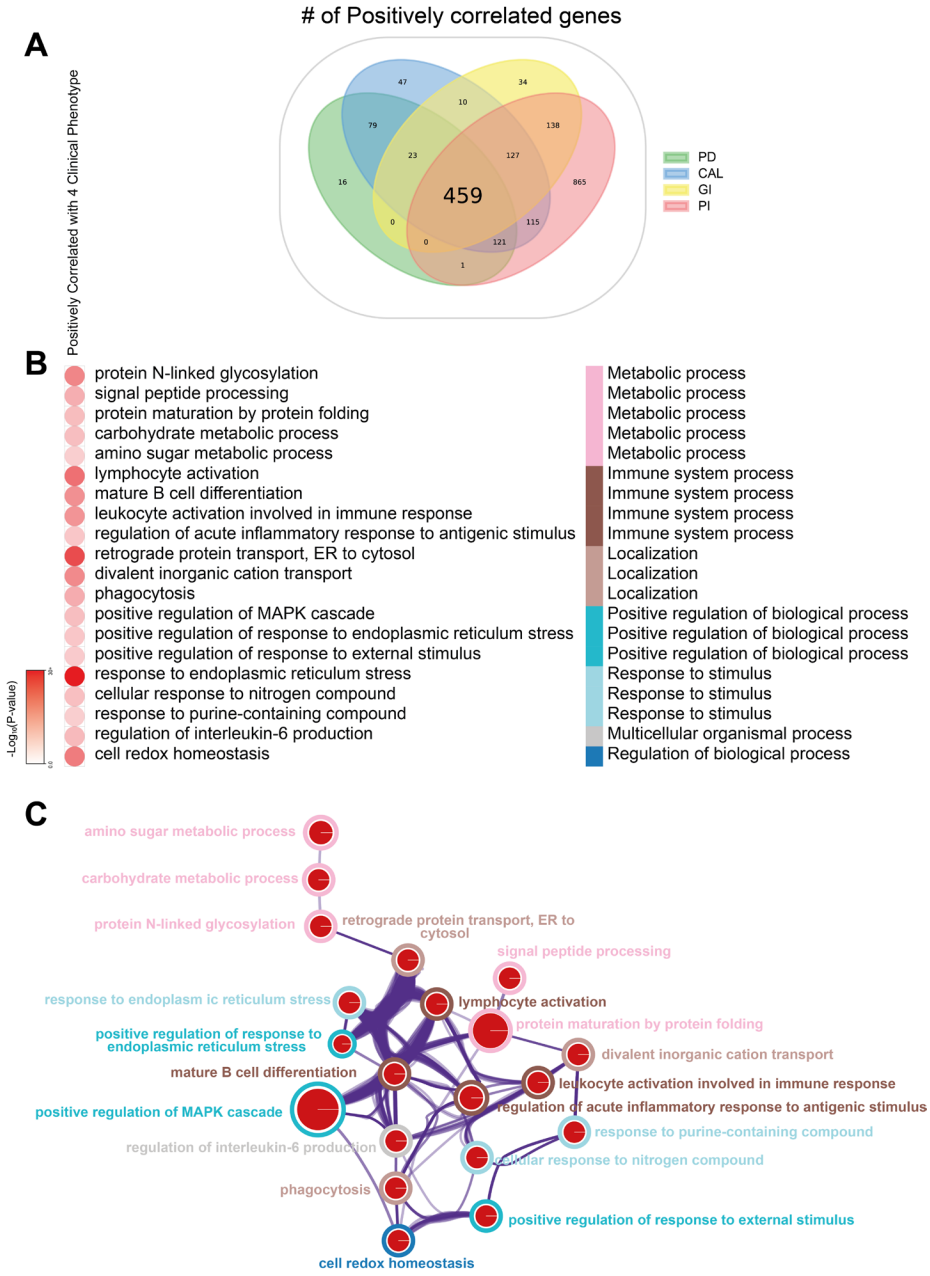

Figure 3.4. Highly correlated genes with clinical parameters

A Spearman correlation analysis has been conducted to evaluate the connection between gene expression levels and certain variables (PD, CAL, GI and PI). (A) Venn diagrams of genes highly correlated (correlation

coefficient > 0.45 & adjusted P-value < 0.01) with the clinical parameters. (B) A heatmap of representative enriched GO terms for positively correlated genes. Right color bars show the representative parent GO terms for the individual enriched GO terms. (C) A graph of Gene Ontology (GO) terms was created based on the results of a cluster analysis. The size of the nodes in the graph represents the number of genes that were found to be enriched in each GO cluster. Edges were drawn between two GO terms if their kappa score was greater than 0.3 (Details in the methods section).

Identification of biomarkers from transcriptome derived clinical severity classifiers

To identify biomarkers that can be used to diagnose and predict clinical severity, we applied different machine learning algorithms to classify the clinical severity based on the transcriptome data. After applying standard scaling on normalized count data, total of 100 genes were selected as initial features, retrieved based on mRMR algorithm (Z. Zhao et al., 2019). Figure 3.5B illustrates that the majority of Gene Ontology terms associated with the 100 potential biomarker genes pertain to B cell function, including the regulation of lymphocyte activation, the adaptive immune response, the B cell receptor signaling pathway, and the regulation of pattern recognition receptor signaling pathways. After additional feature selection by filtering genes with high multicollinearity, 44 genes were finally selected as feature variables. The research project utilized gene expression data from 67 individuals, which had been normalized and included 44 unique genes. This data was randomly rearranged and separated into two subsets: a training set with 58 subjects and a test set with 9 subjects. 7 different classifiers were applied with 10-fold stratified cross validation. The extra trees classifier was chosen as the clinical severity classifier since it scored the highest in all seven metrics (Table 2). A PCA plot based on 44 biomarker candidate gene expression profiles showed the samples were distinctively distributed in PC1 and PC2 among groups. SHAP values (Lundberg & Lee, 2017) were calculated on the best extra tree classifier to determine which genes has the greatest impact on the model output (Table 3, Figure 3.5 C).

Table 2. Evaluation metrics of different machine learning classifiers

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Extra Trees Classifier	**0.8967**	**0.9824**	**0.8333**	**0.8575**	**0.8653**	**0.8066**	**0.838**
Random Forest Classifier	0.8467	0.9672	0.75	0.7931	0.8017	0.717	0.7627
Logistic Regression	0.8767	0.9444	**0.8333**	0.8203	0.8359	0.7795	0.8154
CatBoost Classifier	0.83	0.9367	0.75	0.7783	0.7835	0.6852	0.7203
Light Gradient Boosting Machine	0.7433	0.8669	0.6833	0.6947	0.6985	0.5563	0.6067
Gradient Boosting Classifier	0.7267	0.8353	0.6556	0.6897	0.683	0.5354	0.5751
Decision Tree Classifier	0.6433	0.6892	0.5556	0.6264	0.6115	0.3847	0.4052

Table 3. SHAP value of each feature from the best clinical severity classifier

Gene	SHAP	Gene	SHAP
CRYBA4	0.061921	HMHB1	0.020089
CRACDL	0.049924	KRTAP4-12	0.020001
BRD7	0.043643	C9orf50	0.019941
COL4A3	0.038649	MIR6754	0.019914
SNX3	0.037277	MIR4730	0.01894
PNMT	0.036835	SNORA71E	0.018718
LINC01694	0.036339	SNORA27	0.01824
UBE2F-SCLY	0.032915	MIR6516	0.017799
GTF2H2B	0.032909	HNRNPCL3	0.017103
LOC101927727	0.031248	GTSCR1	0.01617
CTXND1	0.030545	NUP210L	0.015702
UOX	0.028447	UBE2MP1	0.014374
CES2	0.027039	MIR6853	0.01382
ADORA2A	0.026413	SNORA47	0.013767
CLHC1	0.025285	SPATA31E1	0.012818
SPART-AS1	0.025168	MIR3164	0.01025
IGFL3	0.024453	ANKRD66	0.009412
USP17L2	0.023053	SH2D7	0.009253
FSTL5	0.022962	LOC100507387	0.006664
GAPDHS	0.021512	LOC100506274	0.004594
LOC105375545	0.021034	MIR4519	0.002755
LINC01777	0.020775	ROS1	0.001331

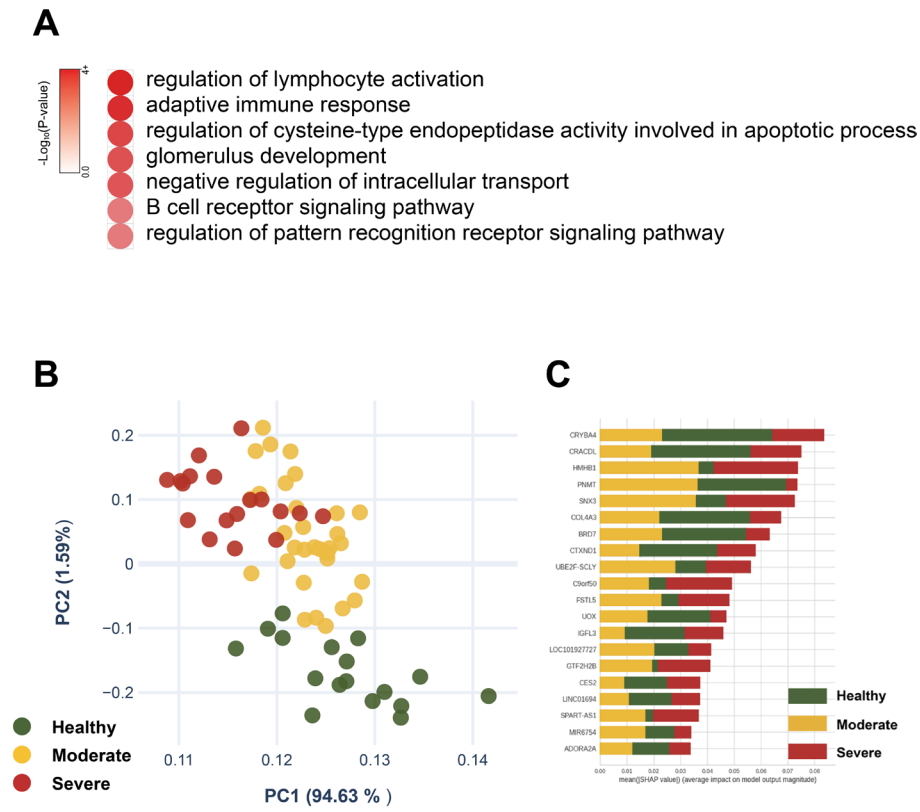

Figure 3.5. Expression profiles and priorities of biomarker candidate genes and enriched GO terms

(A) A heatmap of representative enriched GO terms for initially selected top 100 biomarker candidate genes. (B) A PCA based on the 44 biomarker candidate gene expression profiles.

(C) Genes with higher SHAP values have a stronger effect on the classifier output performance.

Discussion

As the periodontitis has an inherent complication that involves the close interaction of a complex microbiome and systemic immune responses, there have been numerous attempts to unravel its molecular mechanisms and identify diagnostic biomarkers for the state of periodontal disease. Early studies focused on the combination of clinical parameters, but it provided shallow insight into the disease's pathophysiology (Brägger, 2005; Demmer et al., 2008; Hefti, 1997; Meseli et al., 2017). Other branches of the research were dealing with the microorganisms to find key factors of periodontal diseases (Kwack et al., 2022, p. 20; Socransky et al., 1998). Recent developments in the sequencing technology and the idea of "polymicrobial synergy and dysbiosis" have successfully discovered the keystone pathogens associated with the host immune responses, causing the loss of periodontal tissues (Kwack et al., 2022; Lamont & Hajishengallis, 2015). But the each individual has a varied innate susceptibility to microbial species, therefore, the prognosis of periodontal diseases is different (Jeon et al., 2020).

Transcriptome-based systemic immune response analysis was successfully used in this field to address these issues. Many attempts have been made to pinpoint the precise processes and elements involved in the onset of a disease, from microarray to single cell RNA-sequencing (Beikler et al., 2008; Y. Chen et al., 2022). In this study, in addition to identify the transcriptome-level systemic immune response signatures that differs between the different periodontitis severity groups, we also employed machine

learning approaches to construct classifiers and generate a prioritized list of possible biomarker gene candidates.

PCA and hierarchical clustering analysis demonstrated that healthy subjects and periodontitis patients was distinguishable based on the gene expression profiles, although moderate and severe subjects are not clearly discernible. Interesting, the deconvolution analysis based on the scRNA-Seq data successfully assessed the proportional changes of the cell type upon the periodontitis progression. The loss of gingival epithelium could be inferred from a decrease of the epithelial cell proportion, and an increase in the fraction of transit amplifying cells. Therefore, we assumed that the regeneration processes was started from the fast tract differentiation accompanying with the inflammation upon the destruction of gingival tissues (Rangel-Huerta & Maldonado, 2017). Evidently, the increased proportion of immune cells (B cell, T cell, Macro) indicated that the immune cell infiltration occurs and the increase of endothelial cells (blood vessels) and stromal cells also was observed in which has already been shown in the previous studies (Luo et al., 2021; Williams et al., 2021).

Differential gene expression analysis showed that the upregulated genes were dominant DEGs in both groups. The functional annotations of DEGs based on GO terms for the upregulated genes in both groups were similar and most GO terms were primarily related to immune response. When we looked closer into the group-specific DEGs, the majority of upregulated genes were common in both groups, on the other hand, the many downregulated genes were belong to the severe group. GO term enrichment

analysis showed that the many group-specific upregulated DEGs in the severe group were associated with the immune system response functions such as . While the moderate group-specific upregulated DEGs were related to the regeneration processes. Interestingly, Morris et al. reported that the oxygen levels is decreased upon the inflammation progress (Morris et al., 2022) and we observed the several moderate group-specific upregulated DEGs related to the response to decreased oxygen levels.

To investigate the associations between molecular signatures and phenotype characteristics of periodontitis, we measured the correlation of gene expression profiles of the entire subjects and the clinical parameters. Of the 450+ genes analyzed, a positive correlation was identified with certain clinical measures and the functional enrichment analysis demonstrated that immune system processes including leukocyte activation, B cell differentiation, and lymphocyte activation were enriched functions in the correlated genes as expected. Previously, it has been showed that endoplasmic reticulum (ER) stress cause and effect to the chronic inflammation (Chipurupalli et al., 2021; Hasnain et al., 2012) and the autoimmune implications of N-linked glycosylation by misfolded protein stressing in the ER have been reported (Pandey et al., 2022; X. Zhou et al., 2021). Intriguingly, the functional annotation analysis detected several GO terms related to endoplasmic reticulum (ER) stress. Therefore, we inferred that the positively correlated genes closely linked to the phenotype of periodontitis.

To validate the concept of the positive relationship between the molecular signatures and the phenotypic traits, we used the machine learning

techniques to develop the classifiers to predict the periodontal status based on the gene expression profiles. Although there were some trials on predicting periodontal diseases with machine learning or deep learning classifier using microarray datasets and medical images (E.-H. Kim et al., 2020; Rhee, 2019; Ning et al., 2021), they did not consider the severity of the periodontal diseases. In this study, we used the feature selection methods, mRMR technique, recursive feature extraction and addition methods. We did not use clinical information in our classifier because it was based only on transcriptome data. While the classifier's performance might improve if clinical information were included, it is unlikely to have a significant impact as the severity of periodontitis was already determined using clinical parameters. Initially, a selection of 100 potential biomarker genes was made that were closely related to immune processes, including the regulation of lymphocyte activation, the adaptive immune response, the B cell receptor signaling pathway, and the regulation of pattern recognition receptor signaling pathway which can be interpreted as the most important factors that segregates the severity of periodontitis could be contributed from B cell's malfunction. Furthermore, we prioritized the initial biomarkers, set the final 44 candidates and evaluated the candidates using machine learning algorithms. The evaluation results showed that the constructed classifiers using the candidates performed well to predict the periodontal disease status. Though it is not clear exactly what causes periodontitis to become severe, there are indications that it may be related to issues with the way the immune system's B cells recognize or respond to the infection.

In summary, we have profiled the transcriptome of gingival tissues in the periodontitis patients and corresponding healthy subjects and showed the distinct molecular signatures in different three groups. Further, the possible biomarker candidate genes were identified and evaluated. Therefore, the present study provides an important foundation for developing methods to dragonize and screen the periodontal diseases.

Chapter IV.

Conclusion

Conclusion

Inflammaging, or the chronic low-grade inflammation that occurs during aging, has been linked to a number of age-related diseases. In this study, we used RNA sequencing to investigate the molecular basis of inflammaging in both mouse kidneys and human gingival tissue.

In Chapter 2, we used compartment-specific RNA sequencing on aging mouse kidneys to test whether bulk RNA-seq could still capture similar inflammaging-related transcriptome profiles. The result showed compartment specific RNA sequencing can still capture the cell type proportional changes and compartment specific immune patterns. Upon aging, the immune system of their glomeruli, nonglomeruli, and whole kidneys exhibits unique molecular patterns. Additionally, this study analyzed the various expression levels of long non-coding RNA in each of these areas. Lastly, the study showed that the chronic inflammatory condition (UUO) is largely different form inflammaging.

In Chapter 3, we scrutinized the chronic inflammation using the samples of human gingival tissue from individuals with periodontitis and corresponding healthy controls were subjected to RNA sequencing. We managed to tackle this chronic inflammatory problem by combining differential expression profiles between two groups, scRNA-seq deconvolution and the association of clinical parameters and molecular signatures. Finally, we successfully applied machine learning approaches to build plausible periodontitis severity classifier and prioritize gene candidates

for the biomarkers of periodontitis.

In summary, this study used RNA sequencing to investigate the molecular basis of inflammaging in both mouse kidneys and human gingival tissue. Through analysis of compartment-specific RNA in mice and application of machine learning approaches to human tissue samples, the study identified unique molecular patterns and potential biomarkers for age-related inflammation and chronic inflammation in the form of periodontitis. The study also found that chronic inflammatory conditions such as periodontitis have distinct molecular features compared to inflammaging. The contribution of this study is the identification of unique molecular patterns and potential biomarkers for age-related inflammation and chronic inflammation in periodontitis, as well as the development of a periodontitis severity classifier and identification of gene candidates for biomarkers of periodontitis. As future work, it would be interesting to further investigate the potential therapeutic applications of these findings, such as targeting specific molecular pathways to treat or prevent age-related inflammatory diseases and periodontitis. It would also be useful to replicate these findings in larger, more diverse populations to confirm their generalizability.

References

Arango Duque, G., & Descoteaux, A. (2014). Macrophage Cytokines: Involvement in Immunity and Infectious Diseases. *Frontiers in Immunology*, *5*. https://doi.org/10.3389/fimmu.2014.00491

Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P., & De Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nature Communications*, *11*(1), 5650. https://doi.org/10.1038/s41467-020-19015-1

Baraldi, A. (1998). Acute renal failure of medical type in an elderly population. *Nephrology Dialysis Transplantation*, *13*(90007), 25–29. https://doi.org/10.1093/ndt/13.suppl_7.25

Beikler, T., Peters, U., Prior, K., Eisenacher, M., & Flemmig, T. F. (2008). Gene expression in periodontal tissues following treatment. *BMC Medical Genomics*, *1*, 30. https://doi.org/10.1186/1755-8794-1-30

Bradt, B., Bauer, J., Cole, G. M., Cooper, N. R., Eikelenboom, P., Fiebich, B. L., Finch, C. E., Frautschy, S., Griffin, W. S. T., Hull, M., Landreth, G., Lue, F., Mrak, R., Mackenzie, I. R., McGeer, L., O'Banion, M. K., Pachter, J., Pasinetti, G., Plata, C., … Wyss, T. (2014). *Inflammation and Alzheimer's disease*.

Brägger, U. (2005). Radiographic parameters: Biological significance and clinical use. *Periodontology 2000*, *39*(1), 73–90. https://doi.org/10.1111/j.1600-0757.2005.00128.x

Caetano, A. J., Yianni, V., Volponi, A., Booth, V., D'Agostino, E. M., & Sharpe, P. (2021). Defining human mesenchymal and epithelial heterogeneity in

response to oral inflammatory disease. *ELife*, *10*, e62810. https://doi.org/10.7554/eLife.62810

Chen, J.-H., Hales, C. N., & Ozanne, S. E. (2007). DNA damage, cellular senescence and organismal ageing: Causal or correlative? *Nucleic Acids Research*, *35*(22), 7417–7428. https://doi.org/10.1093/nar/gkm681

Chen, L., Chou, C.-L., & Knepper, M. A. (2021). A Comprehensive Map of mRNAs and Their Isoforms across All 14 Renal Tubule Segments of Mouse. *Journal of the American Society of Nephrology*, *32*(4), 897–912. https://doi.org/10.1681/ASN.2020101406

Chen, Y., Wang, H., Yang, Q., Zhao, W., Chen, Y., Ni, Q., Li, W., Shi, J., Zhang, W., Li, L., Xu, Y., Zhang, H., Miao, D., Xing, L., & Sun, W. (2022). Single-cell RNA landscape of the osteoimmunology microenvironment in periodontitis. *Theranostics*, *12*(3), 1074–1096. https://doi.org/10.7150/thno.65694

Chipurupalli, S., Samavedam, U., & Robinson, N. (2021). Crosstalk Between ER Stress, Autophagy and Inflammation. *Frontiers in Medicine*, *8*, 758311. https://doi.org/10.3389/fmed.2021.758311

Chung, J.-J., Goldstein, L., Chen, Y.-J. J., Lee, J., Webster, J. D., Roose-Girma, M., Paudyal, S. C., Modrusan, Z., Dey, A., & Shaw, A. S. (2020). Single-Cell Transcriptome Profiling of the Kidney Glomerulus Identifies Key Cell Types and Reactions to Injury. *Journal of the American Society of Nephrology*, *31*(10), 2341–2354. https://doi.org/10.1681/ASN.2020020220

Demmer, R., Behle, J. H., Wolf, D. L., Handfield, M., Kebschull, M., Celenti, R., Pavlidis, P., & Papapanou, P. N. (2008). Transcriptomes in Healthy and Diseased Gingival Tissues. *Journal of Periodontology*, *79*(11), 2112–2124. https://doi.org/10.1902/jop.2008.080139

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., &

Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, *4*(9), R60. https://doi.org/10.1186/gb-2003-4-9-r60

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Elloumi, F., Hu, Z., Li, Y., Parker, J. S., Gulley, M. L., Amos, K. D., & Troester, M. A. (2011). Systematic bias in genomic classification due to contaminating non-neoplastic tissue in breast tumor samples. *BMC Medical Genomics*, *4*, 54. https://doi.org/10.1186/1755-8794-4-54

Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., & De Benedictis, G. (2000). Inflamm-aging: An Evolutionary Perspective on Immunosenescence. *Annals of the New York Academy of Sciences*, *908*(1), 244–254. https://doi.org/10.1111/j.1749-6632.2000.tb06651.x

Furman, D., & Davis, M. M. (2015). New approaches to understanding the immune response to vaccination and infection. *Vaccine*, *33*(40), 5271–5281. https://doi.org/10.1016/j.vaccine.2015.06.117

Gilroy, D., & De Maeyer, R. (2015). New insights into the resolution of inflammation. *Seminars in Immunology*, *27*(3), 161–168. https://doi.org/10.1016/j.smim.2015.05.003

Green, D. R., Galluzzi, L., & Kroemer, G. (2011). Mitochondria and the Autophagy–Inflammation–Cell Death Axis in Organismal Aging. *Science*, *333*(6046), 1109–1112. https://doi.org/10.1126/science.1201940

Hasnain, S. Z., Lourie, R., Das, I., Chen, A. C., & McGuckin, M. A. (2012). The interplay between endoplasmic reticulum stress and inflammation.

Immunology & Cell Biology, *90*(3), 260–270. https://doi.org/10.1038/icb.2011.112

Hefti, A. F. (1997). Periodontal probing. *Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists*, *8*(3), 336–356. https://doi.org/10.1177/10454411970080030601

Hendry, S., Salgado, R., Gevaert, T., Russell, P. A., John, T., Thapa, B., Christie, M., van de Vijver, K., Estrada, M. V., Gonzalez-Ericsson, P. I., Sanders, M., Solomon, B., Solinas, C., Van den Eynden, G. G. G. M., Allory, Y., Preusser, M., Hainfellner, J., Pruneri, G., Vingiani, A., … Fox, S. B. (2017). Assessing Tumor-infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group: Part 1: Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma In Situ, Metastatic Tumor Deposits and Areas for Further Research. *Advances in Anatomic Pathology*, *24*(5), 235–251. https://doi.org/10.1097/PAP.0000000000000162

Ignarski, M., Islam, R., & Müller, R.-U. (2019). Long Non-Coding RNAs in Kidney Disease. *International Journal of Molecular Sciences*, *20*(13), 3276. https://doi.org/10.3390/ijms20133276

Jeon, Y. S., Cha, J. K., Choi, S. H., Lee, J. H., & Lee, J. S. (2020). Transcriptomic profiles and their correlations in saliva and gingival tissue biopsy samples from periodontitis and healthy patients. *Journal of Periodontal & Implant Science*, *50*(5), 313–326. https://doi.org/10.5051/jpis.1905460273

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), Article 4. https://doi.org/10.1038/nmeth.3317

Kim, E.-H., Kim, S., Kim, H.-J., Jeong, H., Lee, J., Jang, J., Joo, J.-Y., Shin, Y., Kang, J., Park, A. K., Lee, J.-Y., & Lee, S. (2020). Prediction of Chronic Periodontitis Severity Using Machine Learning Models Based On Salivary Bacterial Copy Number. *Frontiers in Cellular and Infection Microbiology*, *10*, 571515. https://doi.org/10.3389/fcimb.2020.571515

Kim, Y.-G., Kim, M., Kang, J. H., Kim, H. J., Park, J.-W., Lee, J.-M., Suh, J.-Y., Kim, J.-Y., Lee, J.-H., & Lee, Y. (2016). Transcriptome sequencing of gingival biopsies from chronic periodontitis patients reveals novel gene expression and splicing patterns. *Human Genomics*, *10*, 28. https://doi.org/10.1186/s40246-016-0084-0

Kwack, K. H., Jang, E.-Y., Yang, S. B., Lee, J.-H., & Moon, J.-H. (2022). Genomic and phenotypic comparison of Prevotella intermedia strains possessing different virulence in vivo. *Virulence*, *13*(1), 1133–1145. https://doi.org/10.1080/21505594.2022.2095718

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., Vallejos, C. A., Campbell, K. R., Beerenwinkel, N., Mahfouz, A., Pinello, L., Skums, P., Stamatakis, A., Attolini, C. S.-O., Aparicio, S., Baaijens, J., Balvert, M., Barbanson, B. de, Cappuccio, A., … Schönhuth, A. (2020). Eleven grand challenges in single-cell data science. *Genome Biology*, *21*(1), 31. https://doi.org/10.1186/s13059-020-1926-6

Lai, R. W., Lu, R., Danthi, P. S., Bravo, J. I., Goumba, A., Sampathkumar, N. K., & Benayoun, B. A. (2019). Multi-level remodeling of transcriptional landscapes in aging and longevity. *BMB Reports*, *52*(1), 86–108. https://doi.org/10.5483/BMBRep.2019.52.1.296

Lamont, R. J., & Hajishengallis, G. (2015). Polymicrobial synergy and dysbiosis in inflammatory disease. *Trends in Molecular Medicine*, *21*(3), 172–183.

https://doi.org/10.1016/j.molmed.2014.11.004

Lim, J. H., Kim, E. N., Kim, M. Y., Chung, S., Shin, S. J., Kim, H. W., Yang, C. W., Kim, Y.-S., Chang, Y. S., Park, C. W., & Choi, B. S. (2012). Age-Associated Molecular Changes in the Kidney in Aged Mice. *Oxidative Medicine and Cellular Longevity*, *2012*, 1–10. https://doi.org/10.1155/2012/171383

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions* (arXiv:1705.07874). arXiv. http://arxiv.org/abs/1705.07874

Luo, Q., He, F., & Cao, J. (2021). A stromal and immune cell infiltration-based score model predicts prognosis and chemotherapy effect in colorectal cancer. *International Immunopharmacology*, *99*, 107940. https://doi.org/10.1016/j.intimp.2021.107940

Meseli, S. E., Kuru, B., & Kuru, L. (2017). Relationships between initial probing depth and changes in the clinical parameters following non-surgical periodontal treatment in chronic periodontitis. *Journal of Istanbul University Faculty of Dentistry*, *51*(3), 11–17. https://doi.org/10.17096/jiufd.40993

Molho, D., Ding, J., Li, Z., Wen, H., Tang, W., Wang, Y., Venegas, J., Jin, W., Liu, R., Su, R., Danaher, P., Yang, R., Lei, Y. L., Xie, Y., & Tang, J. (2022). *Deep Learning in Single-Cell Analysis* (arXiv:2210.12385). arXiv. https://doi.org/10.48550/arXiv.2210.12385

Moreno, J. A., Hamza, E., Guerrero-Hue, M., Rayego-Mateos, S., García-Caballero, C., Vallejo-Mudarra, M., Metzinger, L., & Metzinger-Le Meuth, V. (2021). Non-Coding RNAs in Kidney Diseases: The Long and Short of Them. *International Journal of Molecular Sciences*, *22*(11), 6077.

https://doi.org/10.3390/ijms22116077

Morris, G., Gevezova, M., Sarafian, V., & Maes, M. (2022). Redox regulation of the immune response. *Cellular & Molecular Immunology*, *19*(10), 1079–1101. https://doi.org/10.1038/s41423-022-00902-0

Pandey, V. K., Sharma, R., Prajapati, G. K., Mohanta, T. K., & Mishra, A. K. (2022). N-glycosylation, a leading role in viral infection and immunity development. *Molecular Biology Reports*, *49*(8), 8109–8120. https://doi.org/10.1007/s11033-022-07359-4

Papapanou, P. N., Abron, A., Verbitsky, M., Picolos, D., Yang, J., Qin, J., Fine, J. B., & Pavlidis, P. (2004). Gene expression signatures in chronic and aggressive periodontitis: A pilot study. *European Journal of Oral Sciences*, *112*(3), 216–223. https://doi.org/10.1111/j.1600-0722.2004.00124.x

Paraskevas, K. I., Bessias, N., Koupidis, S. A., Tziviskou, E., Mikhailidis, D. P., & Oreopoulos, D. G. (2010). Incidence of end-stage renal disease in the elderly: A steadily rising global socioeconomic epidemic. *International Urology and Nephrology*, *42*(2), 523–525. https://doi.org/10.1007/s11255-009-9691-1

Park, D., Kim, B.-C., Kim, C.-H., Choi, Y. J., Jeong, H. O., Kim, M. E., Lee, J. S., Park, M. H., Chung, K. W., Kim, D. H., Lee, J., Im, D.-S., Yoon, S., Lee, S., Yu, B. P., Bhak, J., & Chung, H. Y. (2016). RNA-Seq analysis reveals new evidence for inflammation-related changes in aged kidney. *Oncotarget*, *7*(21), 30037–30048. https://doi.org/10.18632/oncotarget.9152

Park, J., Shrestha, R., Qiu, C., Kondo, A., Huang, S., Werth, M., Li, M., Barasch, J., & Suszták, K. (2018). Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science*, *360*(6390), 758–763. https://doi.org/10.1126/science.aar2131

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*(null), 2825–2830.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), Article 3. https://doi.org/10.1038/nbt.3122

Rangel-Huerta, E., & Maldonado, E. (2017). Transit-Amplifying Cells in the Fast Lane from Stem Cells towards Differentiation. *Stem Cells International*, *2017*, 1–10. https://doi.org/10.1155/2017/7602951

Rhee, J.-K. (2019). Prediction for Periodontal Disease using Gene Expression Profile Data based on Machine Learning. *Journal of the Korea Institute of Information and Communication Engineering*, *23*(8), 903–909. https://doi.org/10.6109/jkiice.2019.23.8.903

Roh, J. S., & Sohn, D. H. (2018). Damage-Associated Molecular Patterns in Inflammatory Diseases. *Immune Network*, *18*(4), e27. https://doi.org/10.4110/in.2018.18.e27

Shalek, A. K., & Benson, M. (2017). Single-cell analyses to tailor treatments. *Science Translational Medicine*, *9*(408), eaan4730. https://doi.org/10.1126/scitranslmed.aan4730

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sharma, A., Merritt, E., Hu, X., Cruz, A., Jiang, C., Sarkodie, H., Zhou, Z., Malhotra, J., Riedlinger, G. M., & De, S. (2019). Non-Genetic Intra-Tumor Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary Dynamics in Lung Tumors. *Cell Reports*, *29*(8), 2164-2174.e5. https://doi.org/10.1016/j.celrep.2019.10.045

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., & Kent, R. L. (1998). Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*, *25*(2), 134–144. https://doi.org/10.1111/j.1600-051x.1998.tb02419.x

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Takeuchi, O., & Akira, S. (2010). Pattern Recognition Receptors and Inflammation. *Cell*, *140*(6), 805–820. https://doi.org/10.1016/j.cell.2010.01.022

The Tabula Muris Consortium, Almanzar, N., Antony, J., Baghel, A. S., Bakerman, I., Bansal, I., Barres, B. A., Beachy, P. A., Berdnik, D., Bilen, B., Brownfield, D., Cain, C., Chan, C. K. F., Chen, M. B., Clarke, M. F., Conley, S. D., Darmanis, S., Demers, A., Demir, K., … Zou, J. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, *583*(7817), 590–595. https://doi.org/10.1038/s41586-020-2496-1

Vegh, P., & Haniffa, M. (2018). The impact of single-cell RNA sequencing on understanding the functional organization of the immune system. *Briefings in Functional Genomics*, *17*(4), 265–272. https://doi.org/10.1093/bfgp/ely003

Vénéreau, E., Ceriotti, C., & Bianchi, M. E. (2015). DAMPs from Cell Death to New Life. *Frontiers in Immunology*, *6*. https://doi.org/10.3389/fimmu.2015.00422

Vitale, G., Salvioli, S., & Franceschi, C. (2013). Oxidative stress and the ageing

endocrine system. *Nature Reviews Endocrinology*, *9*(4), Article 4. https://doi.org/10.1038/nrendo.2013.29

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, *131*(4), 281–285. https://doi.org/10.1007/s12064-012-0162-3

Wang, X., Park, J., Susztak, K., Zhang, N. R., & Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-018-08023-x

Wang, Y.-N., Yang, C.-E., Zhang, D.-D., Chen, Y.-Y., Yu, X.-Y., Zhao, Y.-Y., & Miao, H. (2021). Long non-coding RNAs: A double-edged sword in aging kidney and renal disease. *Chemico-Biological Interactions*, *337*, 109396. https://doi.org/10.1016/j.cbi.2021.109396

Williams, D. W., Greenwell-Wild, T., Brenchley, L., Dutzan, N., Overmiller, A., Sawaya, A. P., Webb, S., Martin, D., Hajishengallis, G., Divaris, K., Morasso, M., Haniffa, M., & Moutsopoulos, N. M. (2021). Human oral mucosa cell atlas reveals a stromal-neutrophil axis regulating tissue immunity. *Cell*, *184*(15), 4090-4104.e15. https://doi.org/10.1016/j.cell.2021.05.013

Xie, H., Xue, J.-D., Chao, F., Jin, Y.-F., & Fu, Q. (2016). Long non-coding RNA-H19 antagonism protects against renal fibrosis. *Oncotarget*, *7*(32), 51473–51481. https://doi.org/10.18632/oncotarget.10444

Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., Lu, H., & Yao, J. (2022). ScBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nature Machine Intelligence*, *4*(10), Article 10. https://doi.org/10.1038/s42256-022-00534-z

Yang, X., Bam, M., Becker, W., Nagarkatti, P. S., & Nagarkatti, M. (2020). Long

Noncoding RNA AW112010 Promotes the Differentiation of Inflammatory T Cells by Suppressing IL-10 Expression through Histone Demethylation. *The Journal of Immunology*, *205*(4), 987–993. https://doi.org/10.4049/jimmunol.2000330

Zhao, T., Cai, M., Liu, M., Su, G., An, D., Moon, B., Lyu, G., Si, Y., Chen, L., & Lu, W. (2020). LncRNA 5430416N02Rik Promotes the Proliferation of Mouse Embryonic Stem Cells by Activating Mid1 Expression through 3D Chromatin Architecture. *Stem Cell Reports*, *14*(3), 493–505. https://doi.org/10.1016/j.stemcr.2020.02.002

Zhao, Z., Anand, R., & Wang, M. (2019). Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 442–452. https://doi.org/10.1109/DSAA.2019.00059

Zhou, X., Motta, F., Selmi, C., Ridgway, W. M., Gershwin, M. E., & Zhang, W. (2021). Antibody glycosylation in autoimmune diseases. *Autoimmunity Reviews*, *20*(5), 102804. https://doi.org/10.1016/j.autrev.2021.102804

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

국문 초록

면역학의 주요 어려움 중 하나는 서로 다른 마커와 세포 유형이 복잡하게 얽혀 있다는 점이다. 따라서 이러한 문제점을 다양한 마커와 세포 유형을 효과적으로 식별할 수 있는 RNA 시퀀싱 기술을 통해 해결하기 위한 다양한 시도가 이루어지고 있다. 그러나 단일세포 RNA 시퀀싱 기술을 통해 이뤄낸 많은 발전에도 불구하고 기술적 또는 비용적 문제로 인해 여전히 bulk RNA 시퀀싱이 선호되기도 한다. 본 연구에서는 bulk RNA 시퀀싱을 기반으로 하면서도, 단일세포 RNA 시퀀싱의 최근 발전을 활용하여 간단하면서도 강력하고 비용 효율적인 면역 반응 분석 파이프라인을 제시하였다. 신장 노화 마우스 모델과 다양한 중증도를 가진 인간 치주염 환자군을 본 연구에서 제안하는 방법을 이용하여 분자 레벨에서 분석하였으며, 다양한 면역 반응을 확인하고 치주염 질환을 진단할 수 있는 유전자를 성공적으로 동정하였다. 본 연구의 결과는 면역 체계를 이해하는데 포괄적이고 새로운 시야를 제공할 수 있을 것이다.

……..……..……..……..……..……..……..……..

주요어: 만성염증, 면역반응, RNA 시퀀싱, 단일 세포 디컨볼루션, (염증노화, 치주염

학번: 2015-20508