Immune Landscape of Periodontitis Unveils Alterations of Infiltrating Immunocytes and Molecular Networks-Aggregating into an Interactive Web-Tool for Periodontitis Related Immune Analysis and Visualization

Xiaoqi Zhang
Department of Orthodontics, West China Hospital of Stomatology, State Key Laboratory of Oral Diseases, National Clinical Research Center of Oral Diseases, Sichuan University

Qingxuan Wang
Department of Orthodontics, West China Hospital of Stomatology, State Key Laboratory of Oral Diseases, National Clinical Research Center of Oral Diseases, Sichuan University

Xinyu Yan
Department of Orthodontics, West China Hospital of Stomatology, State Key Laboratory of Oral Diseases, National Clinical Research Center of Oral Diseases, Sichuan University

Yue Shan
Department of Orthodontics, West China Hospital of Stomatology, State Key Laboratory of Oral Diseases, National Clinical Research Center of Oral Diseases, Sichuan University

Lu Xing
Department of Orthodontics, West China Hospital of Stomatology, State Key Laboratory of Oral Diseases, National Clinical Research Center of Oral Diseases, Sichuan University

Minqi Li
Department of Bone metabolism, School and Hospital of Stomatology, Shandong University & Shandong Key Laboratory of Oral Tissue Regeneration & Shandong Engineering Laboratory for Dental Materials and Oral Tissue Regeneration

Hu Long
Department of Orthodontics, West China Hospital of Stomatology, State Key Laboratory of Oral Diseases, National Clinical Research Center of Oral Diseases, Sichuan University

Wenli Lai (wenlilai@scu.edu.cn)
Sichuan University West China Hospital of Stomatology https://orcid.org/0000-0002-7652-739X

Research

Keywords: periodontitis, immune, immunocyte, gene, pathway, web-tool, bioinformatics

Posted Date: June 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37931/v1

Version of Record: A version of this preprint was published on November 18th, 2020. See the published version at https://doi.org/10.1186/s12967-020-02616-1.

Abstract

Background: Immunity and immunocyte reaction play an essential role in periodontitis progress and we aim to investigate the underlying regulatory network of periodontitis immune alterations.

Methods: CIBERSORT was used to estimated immunocyte fractions in different clinical status. Logistic regression was used to assess the immunocyte weight to periodontitis. Immune-related periodontitis subtypes were identified by the NMF algorithm. GSEA and GSVA were conducted to analyze pathway activity. Immunocytes related gene modules were identified by WGCNA.

Results: Altered immunocytes in healthy-vs-periodontitis, aggressive-vs-chronic, male-vs-female and age were identified. Contributing for periodontitis of immunocytes was calculated and their correlation was also done. Two distinct immune-related periodontitis subtypes were identified and one is characterized by B cell reaction and the other is IL-6 cytokine reactions. 463 statistically significant correlations between 22 immunocytes and pathways were revealed. Immunocytes and clinical phenotypes matched their gene modules, and their functions were annotated. Last, an easy-to-use and user-friendly interactive web-tool were developed for periodontitis related immune analysis and visualization (http://118.24.100.193:3838/tool-PIA/).

Conclusions: This study systematically investigated periodontitis immune atlas and glimpse the underlying mechanism of periodontitis from gene-pathway-immunocyte networks, which can not only inspire researchers but also help them in periodontitis related immune researches.

Background

Periodontitis is one of the three most common oral diseases with high morbidity rate [1]. Epidemiological investigation revealed that periodontitis is the primary cause of tooth loss worldwide [2]. Periodontitis is characterized by the inflammatory destruction of periodontal support tissues, with main clinical manifestations including gingival inflammation, bleeding, periodontal pocket formation, alveolar bone resorption, progressive loss of attachment and tooth loosening and displacement [3]. Bacteria and their products in the plaque biofilm attached to and around the tooth surface can not only directly cause the inflammation and destruction of periodontal supporting tissues, but also trigger the host's immune and inflammatory response[4]. The progression of the disease not only depends on the bacteria but also the host's immune response, because the inappropriate immune response to microorganisms can accelerate the development and progress of periodontitis[5]. Therefore, host immune response, especially the immunocyte reaction, plays a key role in the balance between periodontal tissue repair and destruction. A lot of previous studies have demonstrated both innate and adaptive immunocyte reactions play an essential role in periodontitis progress such as dendritic cell [6], macrophage [7], neutrophil [8] in innate immunity responses, and CD4 T cells [9], CD8 T cell [10], and B cell [11] in adaptive immunity response. Hence, how immunocytes mediate the occurrence and development of periodontal inflammation and

alveolar bone destruction and how the molecular regulation network function in immunocytes reaction and differentiation are the focus of current research on periodontitis.

Nowadays, we have been the post-genome era and the price for making high-throughput data is cheaper and cheaper. This results in tens of thousands of high-throughput data were generated. How to use bioinformatics to analyze and use these data fully is now the biggest challenge. High-throughput data were widely used in various fields including periodontitis, such as Yuko et al. used methylation array identifying periodontal disease-related methylation sites [12], K et al. used miRNA array detecting miR143-3p as a novel salivary biomarker for chronic periodontitis [13], and M et al. used gene array finding distinct periodontitis subtypes [14]. However, in these studies high-throughput data were not fully used, only routine analysis has been performed and only single of the numerous results were used and explained. Besides, analyzing and using these high-throughput data requires a certain basis in bioinformatics and programming. Therefore, it is imperative to better assist researchers by comprehensively analyzing these public data using advanced algorithmic systematically to find more discoveries and to present the results in a simple form to researchers who do not understand bioinformatics.

In this study, we aim to use several advanced bioinformatic methods analyzing the periodontitis related immunocytes from public data in multiple levels and aggregate the results as an interactive web-tool for periodontitis related immune analysis and visualization. As a result, genes, pathways and immunocytes alterations and relationships in periodontitis were systematically revealed and abundant discoveries were presented. Furthermore, a web tool was constructed. These can not only inspire researchers but also help them in periodontitis related immune researches.

Methods

1. Data acquisition and process

The data used in this study was in GSE16134 from the GEO database [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse16134]. It contains expression data of 69 healthy samples and 241 periodontitis samples. The detail information was described in the previous studies[14]. The raw data preprocesses were performed according to the standard procedure for Affymetrix microarray and prepared data for CIBERSORT. The clinical information was obtained from Prof. P.N. Papapanou from Columbia University by personal communication on 7/25/2019.

2. Bioinformatic analysis

CIBERSORT was used to estimate the proportion of infiltrating immunocyte subsets by transforming gene expression matrix to 22 types of immunocyte fractions matrix [15]. The immune-related gene list was obtained from the ImmPort database [http://www.immport.org]. Differentially expressed gene (DEG) analysis was performed by R package “limma”, and was conducted according to its pipeline, which is

described before [16]. Genes with |logFC|>1 and FDR<0.01 were regarded as DEGs in periodontitis. The protein-protein interaction (PPI) network was conducted from the STRING database [https://string-db.org/] and beautified by CytoScape. The Gene Ontology and KEGG function enrichment analysis and GSEA were performed using the R package “clusterProfiler”. GSVA was conducted using the R package “gsva”, which transforms an expression matrix to the gene-set score matrix. R package “NMF” was used for consensus molecular subtyping [17]. The optimal number of subtypes was selected according to cophenetic, dispersion, and silhouette coefficients [18]. WGCNA analysis was conducted by R package “WGCNA”. The top 25% of genes with the largest variation were included and samples far from the main cluster were excluded. The analysis processes of WGCNA were described in detail in a previous study [19, 20].

3. Statistical analysis and visualization

Samples with a CIBERSORT P-value of <0.05 were included in the analysis. R software 3.6.1 was conducted in this study for the statistical analyses and visualization. R package “ggplot2”, “ggpuber” were used to make statistical plots including bar-plots, box-plots, violin-plots volcano-plots, and scatter-plots. R package “pheatmap” was used for heatmaps. The periodontitis related immunocytes and immune genes were identified by univariate logistic analysis. LASSO regression was used for dimension reduction in immunocytes. Multivariate logistic regression was used to identify immunocytes related to periodontitis independently. For differences of immunocyte fractions between different clinical groups were analyzed by a two-sided Wilcoxon rank-sum test. Correlation analysis was performed using the Spearman test. The p-value was adjusted by the FDR method for multiple hypothesis testing. Dichotomous variables were compared using the chi-square test. The web-tool was constructed by R package “shiny”

Results

1. Infiltrating immunocytes landscape in periodontitis

To explore the differences of infiltrating immunocytes between healthy and periodontitis tissues, CIBERSORT algorithm was conducted to reveal the distribution of 22 immunocyte fractions and we found no matter in healthy or periodontitis tissue the plasma cell was the most in content (Figure 1A). We compared the fractions of 22 immunocytes between healthy and periodontitis, and 13 immunocytes content were altered in periodontitis tissue including 5 immunocytes fractions increased (plasma cell, neutrophils, T cells CD4 memory activated, dendritic cells activated and NK cells resting) and 8 immunocytes fractions decreased (Tregs, dendritic cells resting, eosinophils, mast cells resting, macrophages M1, B cells memory, T cells follicular helper and T cells CD4 memory resting) in periodontitis tissue (Figure 1B). Among them, Tregs and neutrophils changed most in decreasing and increasing (Figure 1C). Then, to confirm the variation in immunocytes proportions might be an intrinsic feature that may characterize the individual diversities, we performed PCA analysis and found that the

proportions of immune cells displayed distinct clustering and individual differences (Figure 1D). Among them, plasma cells, B cells memory and mast cells resting were the most contributed in the healthy and periodontitis tissue variation. Correlation analysis was performed in periodontitis samples and all samples. We found in both periodontitis and all samples that B cells memory and plasma cells had a big and significantly negative correlation, so as to plasma cells and macrophages M1 (Figure 1E). Further, the differences of infiltrating immunocytes in different clinical subgroups were also investigated including periodontitis subtype, gender, and age by taking similar analysis strategies (Figure S1).

To explore and identify the robustness and contributing weight of immunocytes that are related to periodontitis independently, we conducted a series of regression algorithms to adjust the immunocyte interaction effect including univariate logistic regression, LASSO regression, and multivariate logistic regression. 12 periodontitis related immunocytes were first identified by univariate analysis (Figure 2A). Then, LASSO regression was performed to make feature selection and dimension reduction so that the unimportant immunocytes could be excluded. Therefore, we screen out 11 important periodontitis related immunocytes through LASSO regression (Figure 2B-C). Last, multivariate logistic regression was conducted on the 11 immunocytes to get the independently periodontitis related immunocytes. Only the dendritic cells resting, dendritic cells activated and T cells follicular helper were independently related to periodontitis (Figure 2D).

2. Identification of immune-related periodontitis subtypes and their characteristics

To identify immune-related periodontitis subtypes, 1214 immune-related genes were compiled, and 702 of them were significantly related to periodontitis (FDR < 0.05). Using the 702-gene panel, 2 distinct immune-related periodontitis subtypes were identified by performing consensus clustering analysis (Figure 3A, Figure S2A-C). Then, 116 genes were selected from all 702 immune-related genes and the sparse structure of the metagenes appears in the localized patterns of the gene contributions, in which the 116 genes were defined as hub genes for each subtype representing the subtype characteristics (Figure 3B, Figure S2D). There were 45 immune-related genes represent subtype 1 (cluster 1) and 71 genes for subtype 2 (cluster 2). To get a further understanding of the regulation effects among these genes, the PPI networks of the 116 hub genes revealed their relationships and importance in the network., in which CXCR4 of cluster-1 has the most links with other genes and PIK3RI of cluster-2 has the most (Figure 3C).

To understand the characteristics and differences between the two immune-related subtypes, GO enrichment analysis and canonical pathway enrichment analysis was performed. Biological processes of GO were conducted and to simplify the results, we grouped categories according to functional theme, which represents the biological characteristics of the cluster (Figure 4A). The characteristics of cluster-1 mainly concerning FC-receptor and immune-associated receptor-mediated immune regulations such as FC/FC-gamma receptor signaling pathway and cell surface receptors. Another characteristic of cluster-1 is about B cells immune processes regulations such as B cells differentiation and B cell activation. As for

cluster-2, the IL-6 cytokine feature is predominant such as regulation of IL-production. Then, canonical pathway analysis revealed the significantly enriched or activated pathways in cluster-1 and cluster-2 (Figure 4B-C). The results were similar as above and we found some famous pathways were shown such as JAK-STAT signaling in cluster-1 and PK3K-AKT in cluster-2.

3. Associations between immune subtypes and immunocytes

To investigated the diversities between the two immune subtypes from cell level, the 22 infiltrating immunocyte fractions were compared and clinical features were also compared. From the results, the two immune subtypes had a large difference in infiltrating immunocytes, in which 13 immunocytes were different (Figure 5A-B). We found cluster-1 had more dendritic cells activated, plasma cells and neutrophils, while B cells memory, T cells CD4 memory resting, T cells follicular helper, NK cells activated, monocytes, macrophages M1, macrophages M2, dendritic cells resting, mast cells activated and mast cells resting were enriched in cluster-2. Besides, we noticed cluster-2 was significantly associated with chronic periodontitis and cluster-1 associated with aggressive periodontitis (Figure 5C), while there is no different distribution between two clusters about gender and age.

4. Differentially expressed genes and gene-sets between healthy and periodontitis

To understand the molecular mechanisms by which genes or gene-sets drive the biological reactions in the periodontitis, DEGs analysis and gene-set enrichment analysis were conducted. Using the criterion of FDR < 0.01 and |logFC| > 1, 171 DEGs were identified and 134 of them were up-regulated in periodontitis while 37 of them were down-regulated. These up-regulated genes may be involved in the progress of periodontitis, and among them, the MZB1, TNFRSF17, and IGLL5 are the top three up-regulated genes. The interaction relationships among these genes were also investigated and presented as a PPI network (Figure S3A). According to outward traversal from a locally dense seed protein and vertex weighting by local neighborhood density, these DEGs were divided into several molecular complex and presented by different shapes. Besides, the top ten hub genes among these genes were also identified with the highest links to nodes, indicating they are in the core position of the network and play an essential role in the regulation of the network.

Not only the individual gene was investigated, but also gene-sets were. We mainly focus on four aspects of gene-sets including immunologic signatures, hallmark gene sets, GO biological processes and canonical pathways. Results showed many enriched gene-sets in periodontitis were associated with immune processes or regulations (Figure S3B-E) such as CD4 T CELL VS B CELL DOWN in immunologic signature, IL6 JAK STAT6 SIGNALING in hallmark, POSITIVE REGULATION OF B CELL DIFFERENTIATION

in GO and T HEPLER PATHWAY in the canonical pathway. All these four aspects implied immune-related response is essential in periodontitis.

5 Identification of infiltrating immunocytes related hallmark pathways

To further understand the molecular mechanisms by which hallmark pathways are involved in immunocyte content, we examined the correlation between the fractions of 22 immunocytes and the activity of 50 hallmark pathways. In total, 463 statistically significant correlations between 22 immunocytes and 50 hallmark pathways were observed, in which the T cells CD4 naive and B cells memory had the most negative correlated pathways as 24 while Plasma cells had the most positive correlated pathways of 21 (Figure 6A). The links with high correlation coefficients were presented (Figure 6B) such as dendritic cells resting is negatively correlated with HALLMARK COAGULATION with -0.56 correlation and plasma cells is positively correlated with HALLMARK UNFOLDED PROTEIN RESPONSE with 0.57 correlation.

6. Infiltrating immunocytes related gene module identified by WGCNA

To understand the molecular mechanisms of alteration in infiltrating immunocytes in detail, the genes or gene modules regulating immunocytes or affected by immunocytes need to be uncovered. Therefore, WGCNA, an advanced bioinformatic algorithm, was conducted to find clusters (modules) of highly correlated genes related to external sample characteristics (immunocytes fractions). The analysis of network topology for thresholding powers was calculated from 1 to 20 and identified 13 was used as network topology for thresholding power considering its relatively mean connectivity power and balanced scale independence (Figure S4A-B). Clustering of module eigengenes was performed in the top 25% variance 4329 genes and 27 modules were identified (Figure 7A-B). Then, the immunocyte fractions and clinical features related gene modules were identified presenting in a heatmap (Figure 7C). For example, plasma cells, we can find the turquoise module was the most positive associated gene module to plasma cells fractions from the heat map and scatter plot (Figure 8A), suggesting the genes in turquoise module is involved in regulation for plasma cell population or affected by increased plasma cell fractions. The genes in the turquoise module with high gene significance and module membership were presented as PPI network so that we could locate the central genes according to links degree (Figure S4C), and we noticed BTK is the hub gene with the most links in the network. GO analysis was also conducted to assess the biological function of the turquoise module and it involved in B cell regulation (Figure 8B). Furthermore, plasma cell-related genes were also investigated (Figure 8C-E, Figure S4D).

7. An easy-to-use and user-friendly web-tool for periodontitis related immunity analysis

To help periodontitis researchers used the results described and facilitate broad access to results in this work to any aspect of interest about periodontitis, we established an easy-to-use and user-friendly interactive web-tool for researchers working on periodontitis (http://118.24.100.193:3838/tool-PIA/). We named this platform as Periodontitis Immune Atlas (PIA). PIA provides 6 modules for users: “Genes”, “Pathways”, “WGCNA”, “Correlations”, “Immunocytes” and “Functions”, and all the results in these six modules can be downloaded for usage. The “Genes” module (Figure 9A) provides the DEGs between healthy and periodontitis in several forms. Users can browse the DEG list in a table and sort them by the way you want. Besides a gene list of immune-related genes with their category is also presented. The specific gene of interest can be visualized by volcano-plot or box-plot. Two genes’ correlation analysis are also provided. The “Pathway” module (Figure 9B) allows users to inquire about the activity score of specific pathways and presented them by box-plot. Three commonly used pathway sets are involved including hallmarks, KEGG and GO-BP. The correlation between the two pathways can also be investigated. The “WGCNA” module (Figure 9C) provides analysis for phenotype and gene module. Users can find gene modules related to interesting phenotypes and the gene modules function can also be investigated. In the “Correlation” module (Figure 9D), users can fully explore the relationships among genes, pathways, and immunocytes. “Immunocytes” module (Figure 9E) allows users to visualize the status of immunocytes in different clinical status and the correlation between two immunocytes is also included. The “Functions” module (Figure 9F) provides GSEA analysis between different clinical status. Besides, single-gene GSEA can be used as a function prediction for a specific gene in periodontitis. All figures generated in this platform can be downloaded for further analysis and the query results are demonstrated in a comfortable form. The functions provided in the web-tool, which will be continuously updated, may serve as a guide for researchers interested in periodontitis related immunity

Discussion

The immunocytes and immune reaction play a key role in periodontitis progress, uncovering the underlying regulation network and mechanism is urgent. To well address the problem, we conducted several advanced bioinformatic methods to systematically analyze the immunity in periodontitis from three aspects including genes, pathways and immunocytes. First, we used CIBERSORT to convert an expression matrix to the immunocytes fraction matrix. 22 infiltrating immunocyte fractions were analyzed between different clinical statuses. Then a series of statistical algorithms were conducted to estimate the immunocyte weight contributing to periodontitis and identify the independent immunocyte affecting periodontitis [21]. And we found dendritic cells activated, dendritic cells resting and T cells follicular helper may play an essential role in periodontitis [22]. Second, we used immune-related genes that are closely related to periodontitis to perform unsupervised clustering by the NMF algorithm [14, 17]. And two distinct immune subtypes of periodontitis were identified with their molecular features. The subtypes were different from the currently accepted periodontitis classification, but their distinct

molecular signatures representing diversity periodontitis-related immune characteristics. The results suggested that immune-related genes of periodontitis tissues can provide an alternative classification of periodontitis that associated with immune characteristics [23], which helps understand the immune pathobiology of periodontitis. Third, GSVA algorithm was employed to convert expression matrix to a pathway activity score matrix, so that the relationship between pathways and immunocytes can be unveiled [24]. Common pathways positive or negative correlated to some immunocytes were fully investigated, which implies these pathways may be involved in immunocytes regulating or affected by immunocytes. Pathways are the bridge between genes and immunocytes. Last, WGCNA was conducted to investigate the relationship between gene modules and phenotypes. Genes were grouped into modules according to their similarity and correlation between gene module and phenotypes were investigated. Such as the turquoise module is significantly related to plasm cells fractions indicating genes of turquoise module may be involved in the regulation of plasm cells. This analysis revealed the relationship between genes and immunocytes. As far as we know, all the four aspects analysis about periodontitis immune regulation were first time been investigated in our study. They systematically explained the periodontitis immune regulation network from gene-pathway-immunocyte.

At present, there is a lack of such systematic and comprehensive bioinformatics analysis in the periodontal research field, and there are few studies on immune cells using high-throughput data, so many data cannot fully play their role. However, in other fields, particularly oncology, bioinformatics analysis and the use of high-throughput data have become the norm [25, 26], which help researchers identify research targets and predict results. In our study, we made a lot of bioinformatic analysis and generated abundant results that can well inspire other researchers. To facilitate immune-related analysis in periodontitis more easily, we developed the PIA (Periodontitis Immune Atlas) App, an easy-to-use and user-friendly web-tool for comprehensive analysis on gene, pathway and immunocyte data in periodontitis. The PIA App integrates several advance bioinformatics methods and provides customized functions and interactive operation including gene-related analysis, pathway-related analysis, immunocytes related analysis, WGCNA analysis, correlation analysis and functional analysis for users to explore the immune-related researches in a multi-dimensional manner. The PIA app serves as a new approach for users, especially wet-bench scientists with no programming background, to analyze the scientific big data and facilitate data mining [27]. It is an interactive web application, which enables experimental researchers without any computational programming background to perform multiple analysis related to periodontitis immunity. Using the PIA App, one can easily explore the large and systematical data, ask specific scientific questions, and validate their findings.

Analyzing infiltrating immunocyte fractions using a machine learning algorithm from transcriptome data maybe not the best way and single-cell sequencing should be the current best way to precisely analyzing infiltrating immunocytes from a bulk tissue [28]. While, before the emergence of single-cell sequencing data of periodontitis, using CIBERSORT algorithm to analyze infiltrating immunocytes is currently the best approach we can use in periodontitis [29, 30]. Besides, considering the high cost and complex technology of single-cell sequencing, it would be a good thing to use such a simple and effective method to study immune infiltration in periodontitis before the popularization of single-cell sequencing, if the CIBERSORT

can be verified experimentally with high accuracy. We believe this CIBERSORT is accurate enough, after all, it is widely used and verified in other fields [31-33]. Therefore, the results of this study deserve more attention.

Conclusions

In conclusion, we systematically analyzed periodontitis infiltrating immunocytes alterations in different clinical status and identified several important immunocytes related to periodontitis by multiple methods. Two immune-related periodontitis subtypes were identified and the relation to immunocytes, clinical features and biological functions were investigated. Pathways associated with immunocyte fractions were revealed. Immunocytes related genes were also identified. These results can glimpse the underlying mechanism of periodontitis from gene-pathway-immunocyte networks. Further, a web-tool for analyzing and visualizing periodontitis related immune was constructed (http://118.24.100.193:3838/tool-PIA/). These can not only inspire researchers but also help them in periodontitis related immune researches.

Abbreviations

WGCNA: Weighted gene co-expression network analysis; MM: Module membership; GS: Gene significance; GSEA: Gene-set enrichment analysis; GSVA: Gene-set variation analysis; GO Gene ontology; BP: Biological process; KEGG: Kyoto Encyclopedia of Genes and Genomes; PCA: Principal component analysis; LASSO: Least absolute shrinkage and selection operator; DEG: Differentially expressed gene; PPI: Protein-protein interaction;

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and materials:

The datasets generated and analysed during the current study are available in the GEO repository [https://www.ncbi.nlm.nih.gov/geo/].

Competing interests:

The authors declare that they have no competing interests.

Funding:

Fundamental Research Program funded by Department of Science and Technology of Sichuan Province grant number: No. 2018JY0558.

Authors' Contributions:

Conceptualization, Xiaoqi Zhang and Wenli Lai; Data curation, Yue Shan; Formal analysis, Xiaoqi Zhang, Qingxuan Wang and Yue Shan; Funding acquisition, Hu Long and Wenli Lai; Methodology, Xinyu Yan; Resources, Xinyu Yan; Software, Xinyu Yan; Supervision, Minqi Li and Wenli Lai; Validation, Lu Xing, Xinyu Yan, Minqi Li, Hu Long and Wenli Lai; Visualization, Xiaoqi Zhang, Qingxuan Wang and Lu Xing; Writing – original draft, Xiaoqi Zhang and Qingxuan Wang; Writing – review & editing, Hu Long and Wenli Lai.

Acknowledgments:

Special thanks for Dear Prof. P.N. Papapanou from Columbia University to share the clinical information of samples in GSE16134 by personal communication on 7/25/2019.

References

1. Slots J: **Periodontitis: facts, fallacies and the future.** *Periodontol 2000* 2017, **75:**7-23.
2. Eke PI, Dye BA, Wei L, Slade GD, Thornton-Evans GO, Borgnakke WS, Taylor GW, Page RC, Beck JD, Genco RJ: **Update on Prevalence of Periodontitis in Adults in the United States: NHANES 2009 to 2012.** *J Periodontol* 2015, **86:**611-622.
3. Darveau RP: **Periodontitis: a polymicrobial disruption of host homeostasis.** *Nat Rev Microbiol* 2010, **8:**481-490.
4. Hernandez M, Dutzan N, Garcia-Sesnich J, Abusleme L, Dezerega A, Silva N, Gonzalez FE, Vernal R, Sorsa T, Gamonal J: **Host-pathogen interactions in progressive chronic periodontitis.** *J Dent Res* 2011, **90:**1164-1170.
5. Amano A: **Host-parasite interactions in periodontitis: microbial pathogenicity and innate immunity.** *Periodontol 2000* 2010, **54:**9-14.
6. Xiao W, Dong G, Pacios S, Alnammary M, Barger LA, Wang Y, Wu Y, Graves DT: **FOXO1 deletion reduces dendritic cell function and enhances susceptibility to periodontitis.** *Am J Pathol* 2015, **185:**1085-1093.
7. Garaicoa-Pazmino C, Fretwurst T, Squarize CH, Berglundh T, Giannobile WV, Larsson L, Castilho RM: **Characterization of macrophage polarization in periodontal disease.** *J Clin Periodontol* 2019, **46:**830-

839.

8. White PC, Chicca IJ, Cooper PR, Milward MR, Chapple IL: **Neutrophil Extracellular Traps in Periodontitis: A Web of Intrigue.** *J Dent Res* 2016, **95:**26-34.
9. Hienz SA, Paliwal S, Ivanovski S: **Mechanisms of Bone Resorption in Periodontitis.** *J Immunol Res* 2015, **2015:**615486.
10. Cardoso EM, Arosa FA: **CD8(+) T Cells in Chronic Periodontitis: Roles and Rules.** *Front Immunol* 2017, **8:**145.
11. Zouali M: **The emerging roles of B cells as partners and targets in periodontitis.** *Autoimmunity* 2017, **50:**61-70.
12. Kurushima Y, Tsai PC, Castillo-Fernandez J, Couto Alves A, El-Sayed Moustafa JS, Le Roy C, Spector TD, Ide M, Hughes FJ, Small KS, et al: **Epigenetic findings in periodontitis in UK twins: a cross-sectional study.** *Clin Epigenetics* 2019, **11:**27.
13. Nisha KJ, Janam P, Harshakumar K: **Identification of a novel salivary biomarker miR-143-3p for periodontal diagnosis: A proof of concept study.** *J Periodontol* 2019, **90:**1149-1159.
14. Kebschull M, Demmer RT, Grun B, Guarnieri P, Pavlidis P, Papapanou PN: **Gingival tissue transcriptomes identify distinct periodontitis phenotypes.** *J Dent Res* 2014, **93:**459-468.
15. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nat Methods* 2015, **12:**453-457.
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43:**e47.
17. Gaujoux R, Seoighe C: **A flexible R package for nonnegative matrix factorization.** *BMC Bioinformatics* 2010, **11:**367.
18. Kim H, Park H: **Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis.** *Bioinformatics* 2007, **23:**1495-1502.
19. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4:**Article17.
20. Zhang X, Feng H, Li Z, Li D, Liu S, Huang H, Li M: **Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis.** *Onco Targets Ther* 2018, **11:**6001-6021.
21. Zeng D, Zhou R, Yu Y, Luo Y, Zhang J, Sun H, Bin J, Liao Y, Rao J, Zhang Y, Liao W: **Gene expression profiles for a prognostic immunoscore in gastric cancer.** *Br J Surg* 2018, **105:**1338-1348.
22. Meyle J, Chapple I: **Molecular aspects of the pathogenesis of periodontitis.** *Periodontol 2000* 2015, **69:**7-17.
23. Wang Q, Li M, Yang M, Yang Y, Song F, Zhang W, Li X, Chen K: **Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: implications for immune checkpoint blockade therapy.** *Aging (Albany NY)* 2020, **12:**3312-3339.

24. Li Y, Xiao J, Bai J, Tian Y, Qu Y, Chen X, Wang Q, Li X, Zhang Y, Xu J: **Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types.** *Mol Cancer* 2019, **18:**137.
25. Yang H, Zhang Q, Xu M, Wang L, Chen X, Feng Y, Li Y, Zhang X, Cui W, Jia X: **CCL2-CCR2 axis recruits tumor associated macrophages to induce immune evasion through PD-1 signaling in esophageal carcinogenesis.** *Mol Cancer* 2020, **19:**41.
26. Otani Y, Yoo JY, Chao S, Liu J, Jaime-Ramirez AC, Lee TJ, Hurwitz B, Yan Y, Dai H, Glorioso JC, et al: **Oncolytic HSV infected glioma cells activate NOTCH in adjacent tumor cells sensitizing tumors to gamma secretase inhibition.** *Clin Cancer Res* 2020.
27. Li Y, Ge D, Lu C: **The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization.** *Epigenetics Chromatin* 2019, **12:**71.
28. Papalexi E, Satija R: **Single-cell RNA sequencing to explore immune cell heterogeneity.** *Nat Rev Immunol* 2018, **18:**35-45.
29. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA: **Profiling Tumor Infiltrating Immune Cells with CIBERSORT.** *Methods Mol Biol* 2018, **1711:**243-259.
30. Finotello F, Trajanoski Z: **Quantifying tumor-infiltrating immune cells from transcriptomics data.** *Cancer Immunol Immunother* 2018, **67:**1031-1040.
31. Yazdani S, Callemeyn J, Gazut S, Lerut E, de Loor H, Wevers M, Heylen L, Saison C, Koenig A, Thaunat O, et al: **Natural killer cell infiltration is discriminative for antibody-mediated rejection and predicts outcome after kidney transplantation.** *Kidney Int* 2019, **95:**188-198.
32. Panousis NI, Bertsias GK, Ongen H, Gergianaki I, Tektonidou MG, Trachana M, Romano-Palumbo L, Bielser D, Howald C, Pamfil C, et al: **Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity.** *Ann Rheum Dis* 2019, **78:**1079-1089.
33. Liu R, Hu R, Zeng Y, Zhang W, Zhou HH: **Tumour immune cell infiltration and survival after platinum-based chemotherapy in high-grade serous ovarian cancer subtypes: A gene expression-based computational study.** *EBioMedicine* 2020, **51:**102602.

Figures

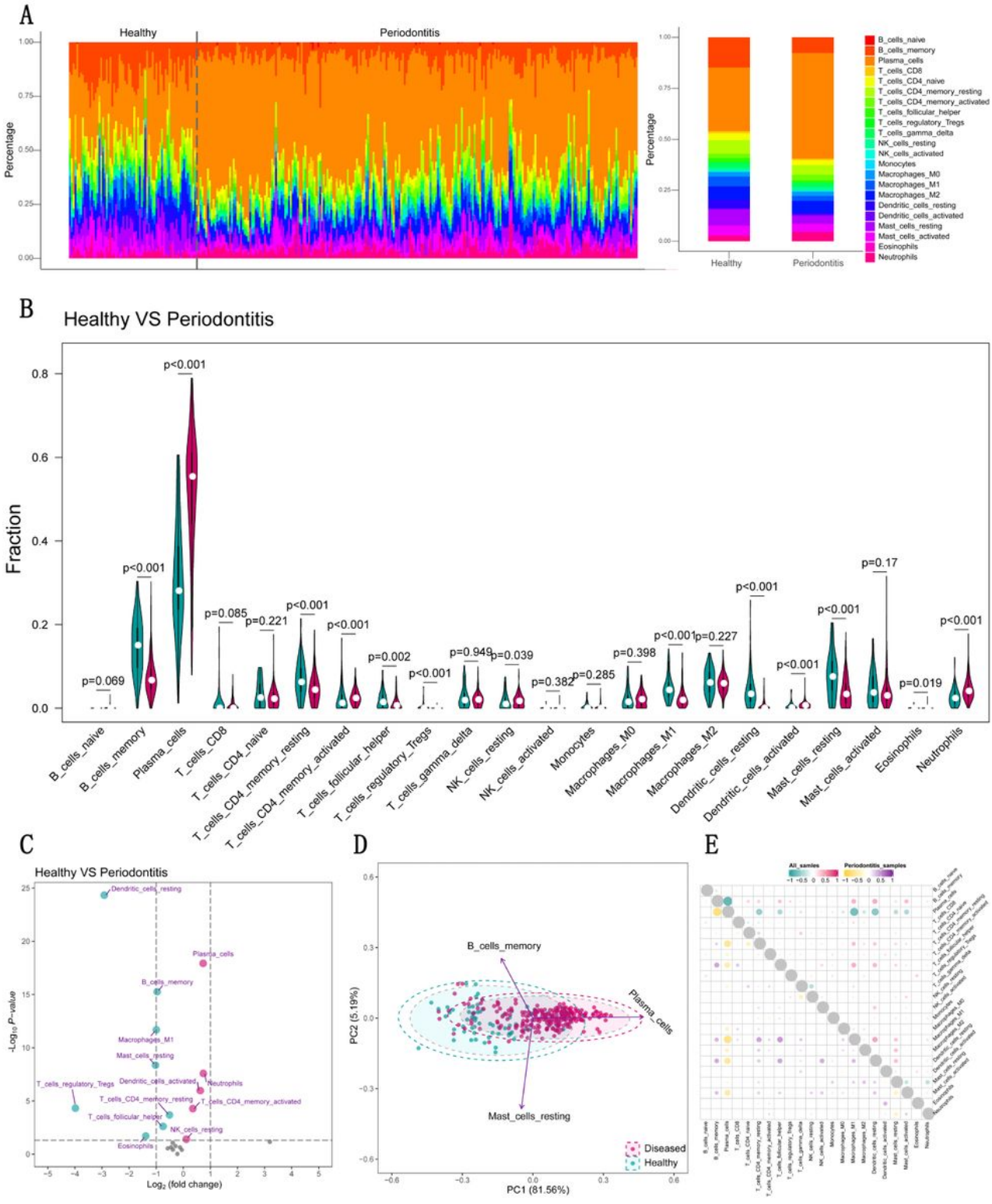

Figure 1

Infiltrating immunocytes differences between healthy and periodontitis samples. (A) Relative fraction of immunocytes identified by CIBERSORT algorithm, which estimates relative subsets of 22 types of immunocyte from known RNA transcripts. The relative distributions of these 22 immunocytes were presented by bar-plots concerning different disease status (all samples for the left and average for the right). (B) Compositional differences of 22 immunocytes between healthy and periodontitis presented by

violin-plot (blue means healthy and red means periodontitis). (C) The volcano-plot demonstrates the fold changes of 22 immunocytes in periodontitis compared with healthy. (D) Principal component analysis (PCA) of 22 infiltrating immunocytes between healthy and periodontitis. The two first principal components (PC1, PC2) which explain the most of the variation are plotted. (E) Correlation matrix of 22 immunocytes proportions. The top right is correlations of 22 immunocytes in all samples and the left bottom is correlations of 22 immunocytes in periodontitis samples.

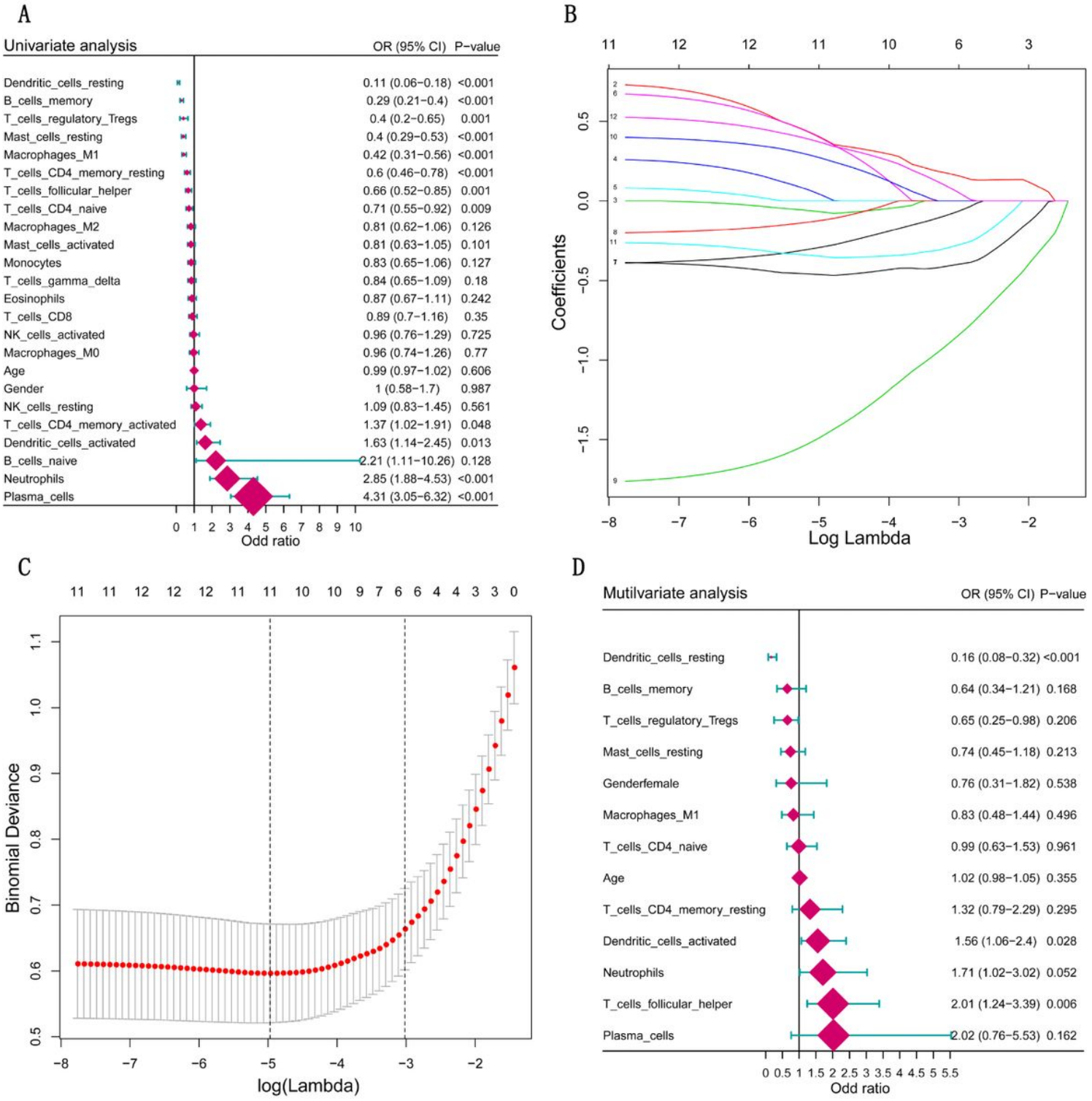

Figure 2

Periodontitis associations of immunocytes (A) Forest-plot demonstrates associations between different immunocyte subsets and periodontitis by univariate logistic regression. (B) Least absolute shrinkage and selection operator (LASSO) coefficient profiles of 22 immunocytes fractions. The dotted line indicates the value chosen by ten-fold cross-validation. (C) Ten-fold cross-validation for tuning parameter selection in the LASSO regression. The partial likelihood deviance is plotted against log (λ), where λ is the tuning parameter. Partial likelihood deviance values are shown, with error bars representing SE. The dotted vertical lines are drawn at the optimal values by minimum criteria and 1-SE criteria. (D) Forest-plot demonstrates the independent associations between periodontitis-related immunocytes and periodontitis by multivariate logistic regression.

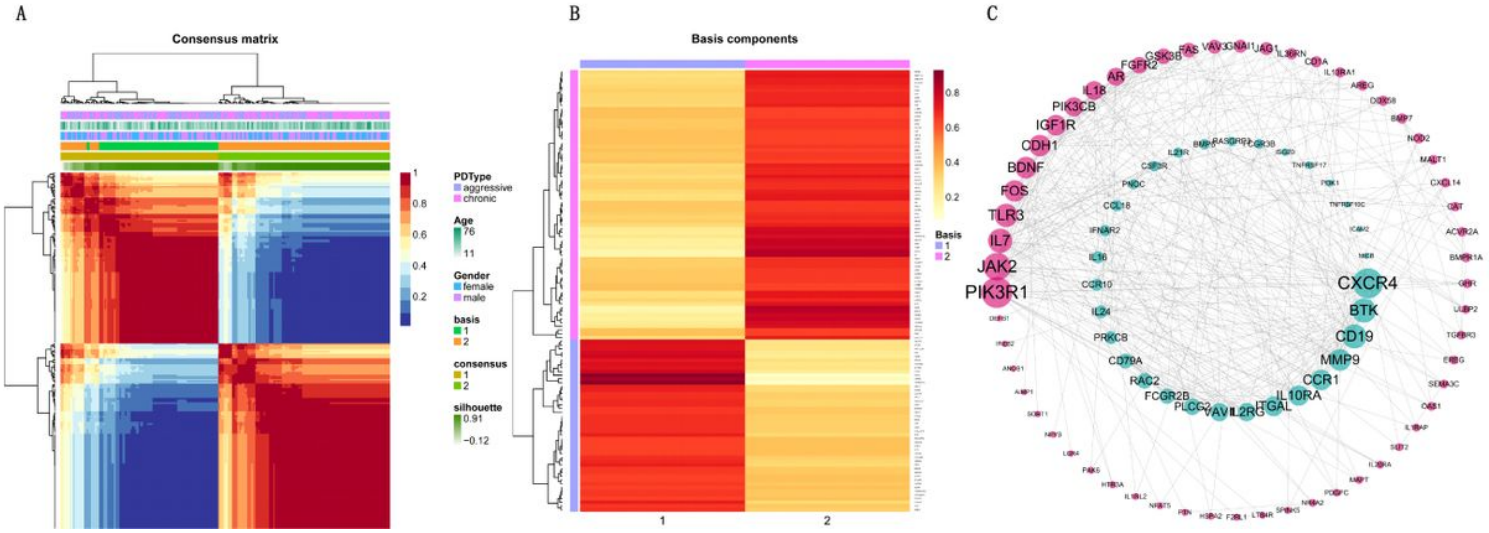

Figure 3

Identification of immune subtypes of periodontitis by consensus clustering. (A) The consensus matrix was obtained from 200 random runs of the Brunet et al.'s algorithm. Values range from 0 to 1. Columns - and rows - were ordered by hierarchical clustering based on the euclidean distance with average linkage. (B) Heatmap of the metagene matrix. Each row corresponds to a gene. The most metagene-specific genes were selected using the Kim and Park's scoring and filtering method. This resulted in the selection of 116 genes. Rows were scaled to sum to one and ordered by hierarchical clustering based on the euclidean distance and average linkage. (C) The protein-protein interaction network of 116 metagenes. The blue is immune subtype cluster-1 and red is subtype cluster-2. Only the genes with interaction with other are presented.

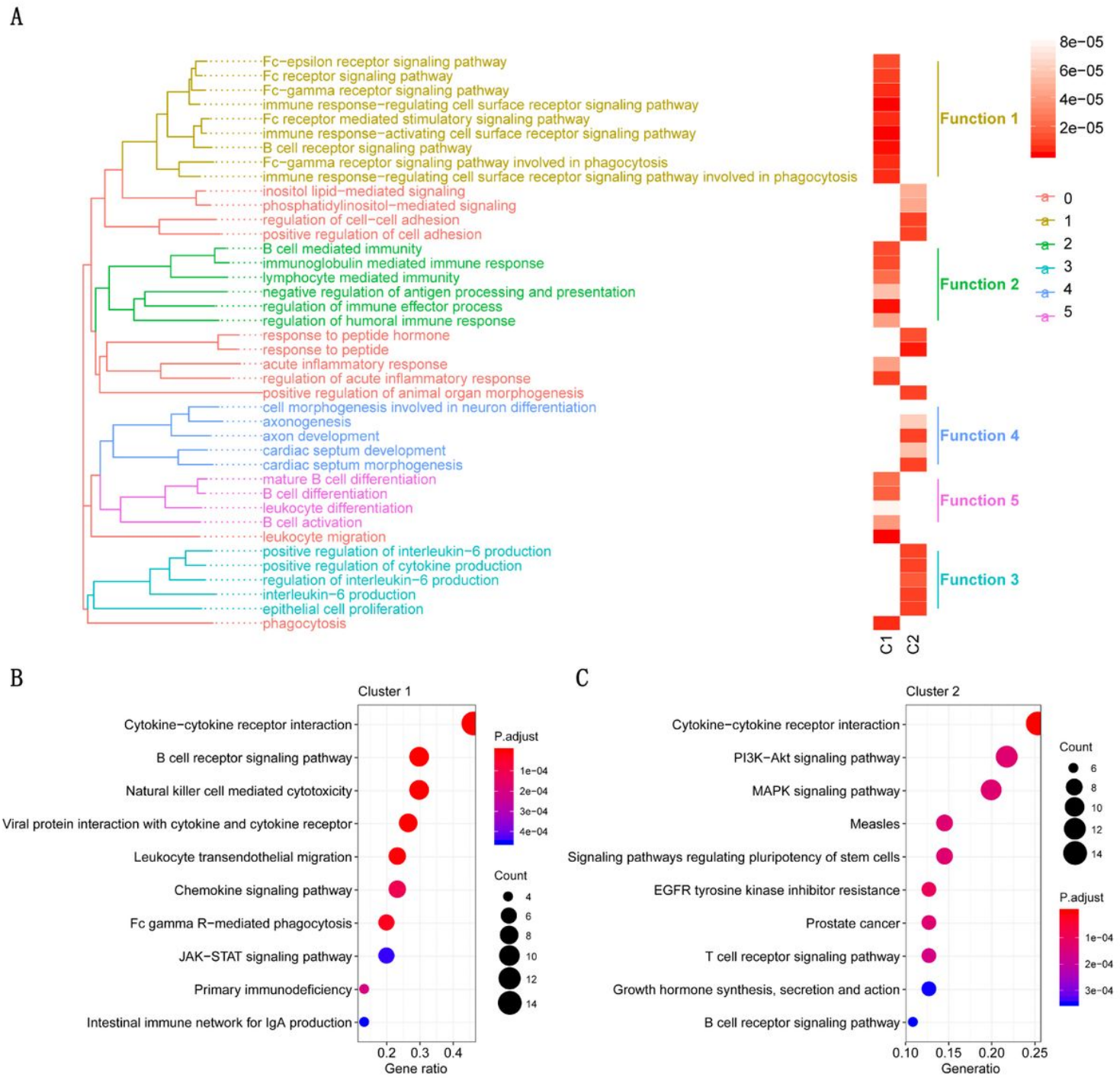

Figure 4

Functions and biological characteristics of immune subtypes. (A) The gene ontology enrichment analysis for the metagenes in two immune subtypes concerning their biological processes. GO categories are grouped according to functional. (B) The KEGG pathway enrichment analysis for the metagenes in two immune subtypes.

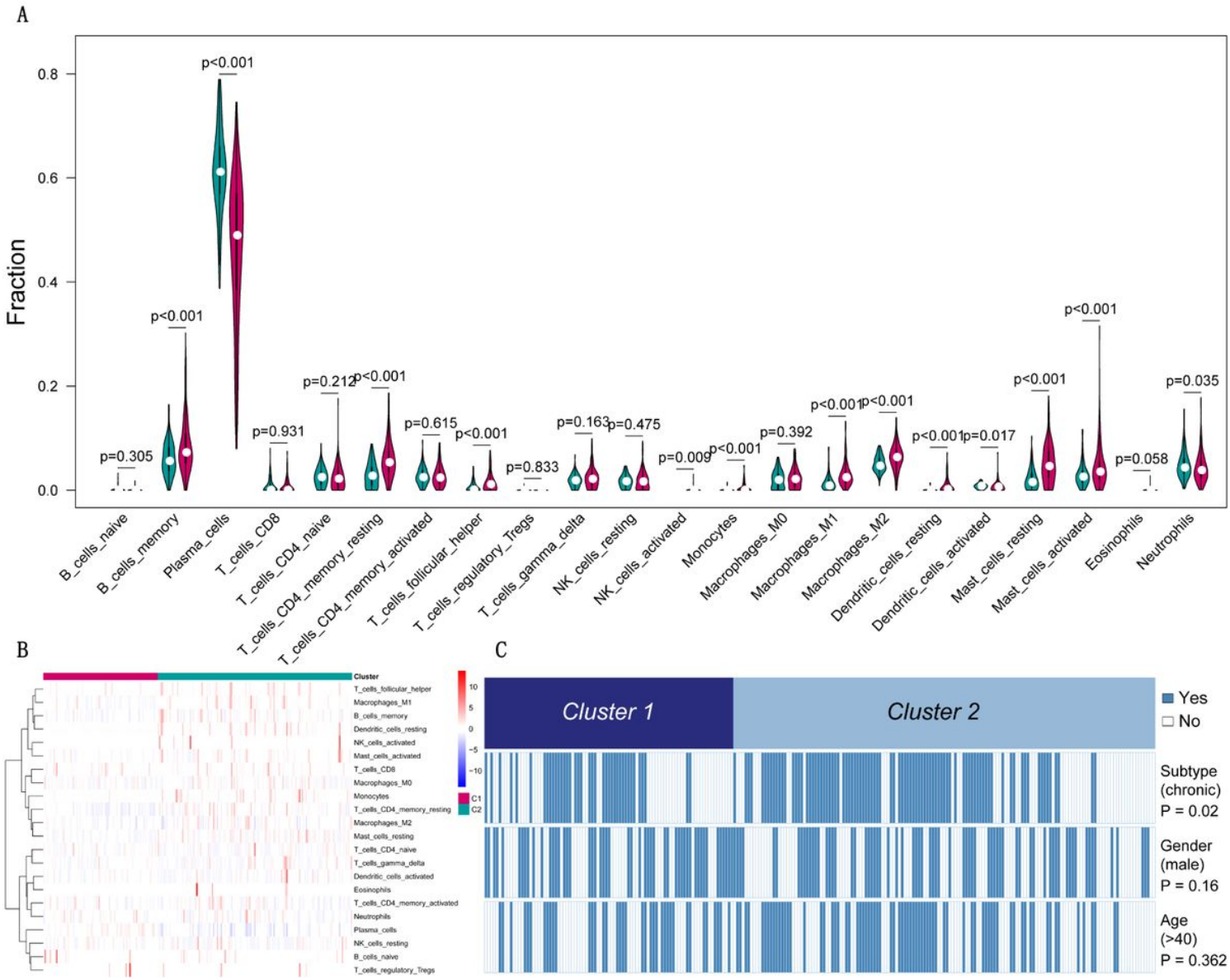

Figure 5

Correlations between immune subtypes and immunocytes. (A) Compositional differences of 22 immunocytes between immune subtype cluster 1 and cluster 2 which was presented by violin-plot (blue means healthy and red means periodontitis). (B) Heatmap demonstrated relationship between 22 immunocytes fractions and the two immune subtypes. (C) Comparing periodontitis subtypes, genders, and ages between immune subtype cluster 1 and cluster 2. The heatmap illustrates the association of different clinical characters with cluster 1 and cluster 2 patients. Statistical significance was performed by Chi-square test.

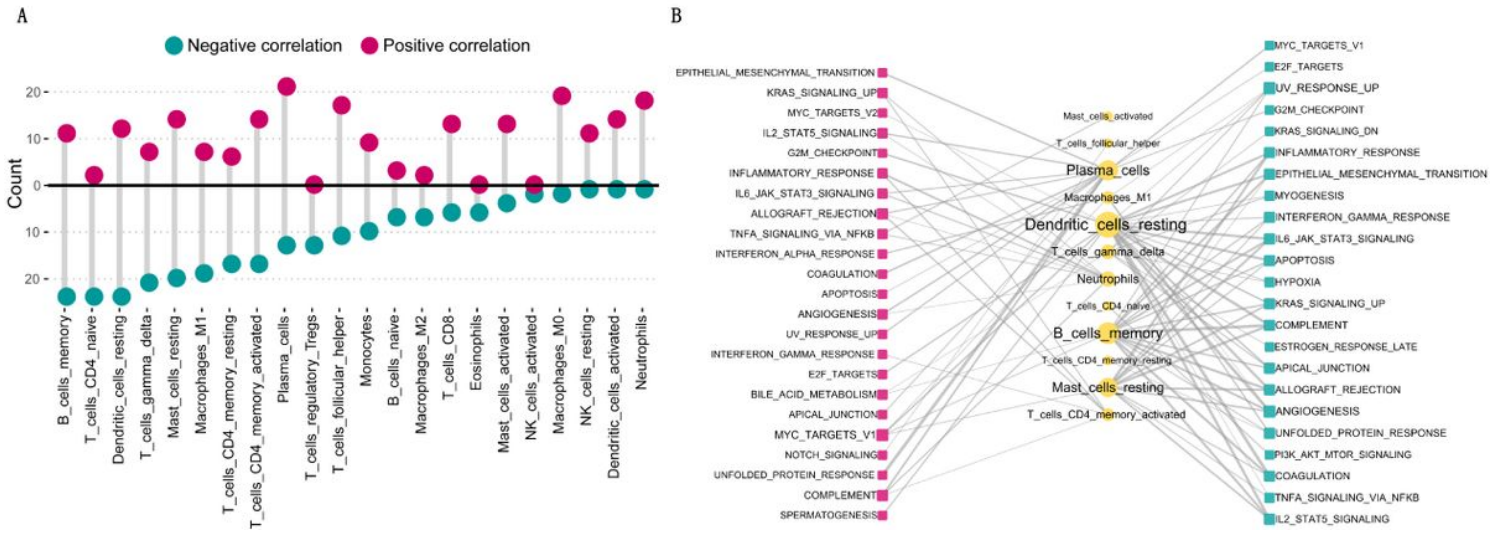

Figure 6

Correlations between 22 immunocytes fractions and activity of 50 important biological hallmark-related pathways. (A) The number of significant pathways is correlated with individual immunocyte. The upper panel is for positively correlated pathways, and the bottom panel is for negatively correlated pathways. (B) Network diagram demonstrating the correlation between immunocytes and pathways. Red represents a positive correlation, and blue represents a negative correlation. The size of the nodes corresponds to the number of links, and thickness of the line represents the correlation coefficient.

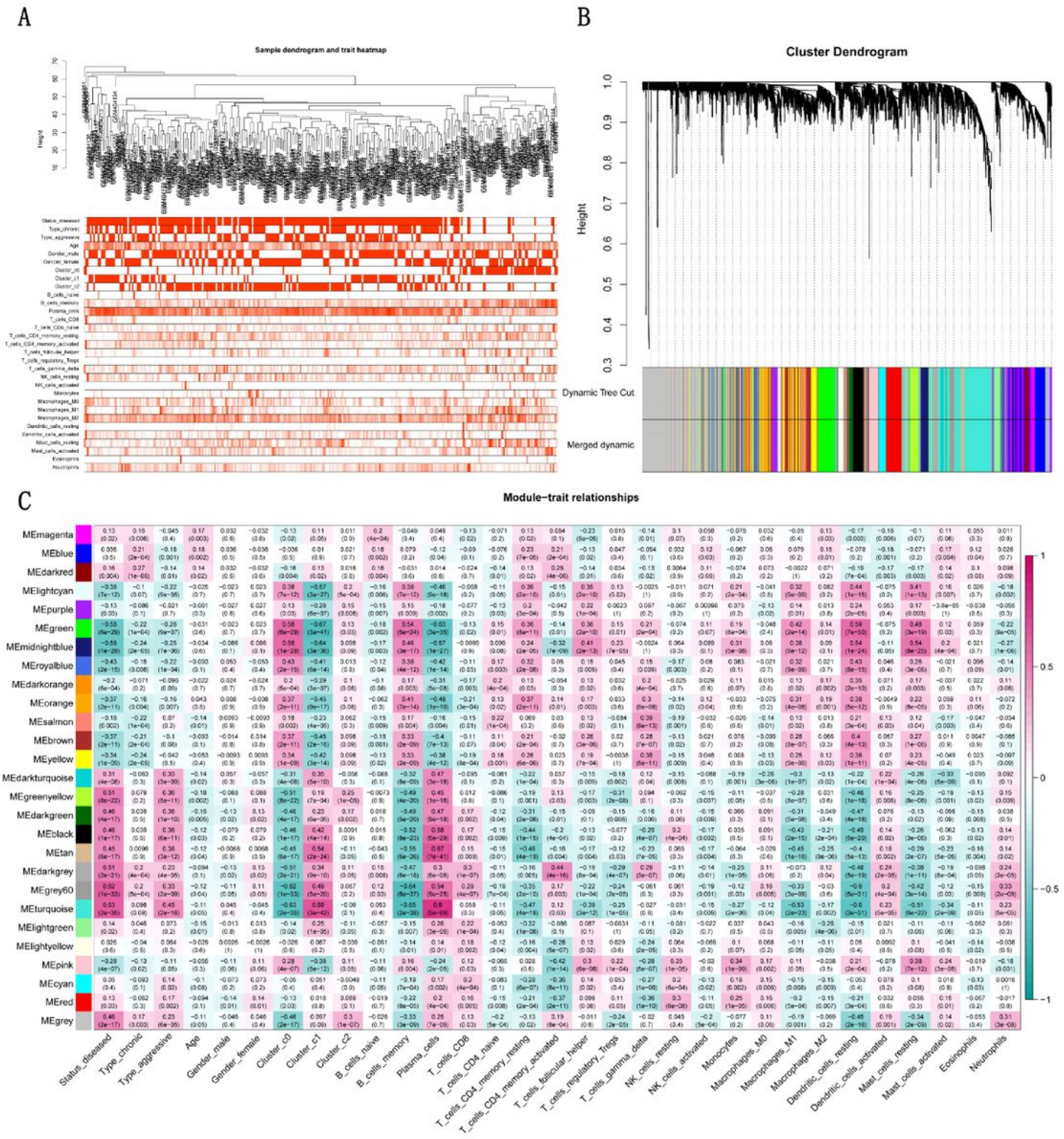

Figure 7

Identification of immunocytes and clinical characteristics related gene modules. (A) The sample clustering was based on the expression data of all samples. The top 25% variation genes were used for the analysis by WGCNA and outlier samples were excluded. (B) Gene dendrogram obtained by average linkage hierarchical clustering. The color row underneath the dendrogram shows the module assignment

determined by the Dynamic Tree Cut, which 27 modules were identified. (C) Heatmap of the correlation between module eigengenes and the immunocytes fractions.

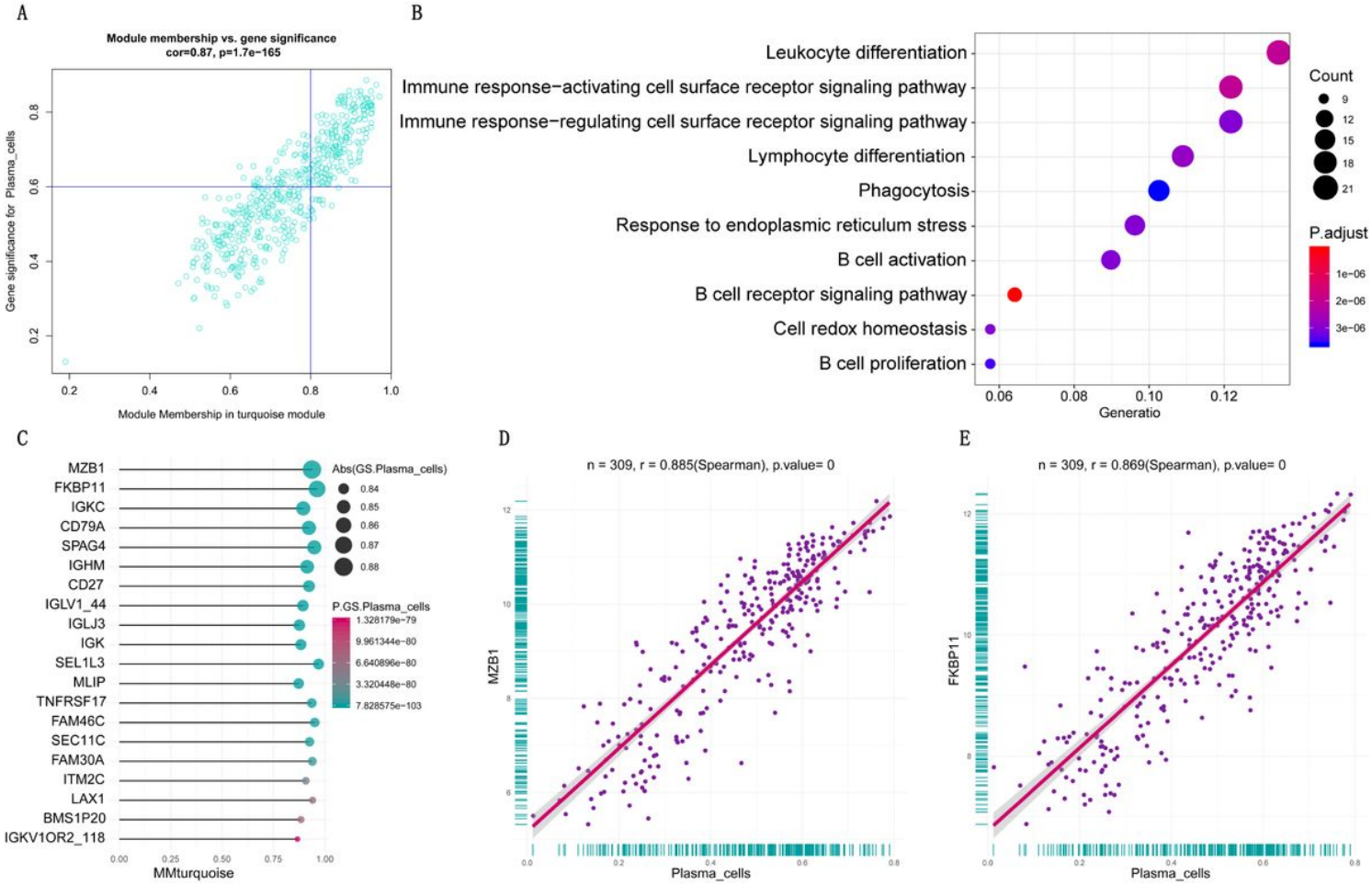

Figure 8

Plasma cells related genes. (A) A scatterplot of gene significance (GS) for plasma cell fraction versus module membership (MM) in the turquoise module. GS and MM exhibit a very significant correlation, implying that hub genes of the turquoise module also tend to be highly correlated with plasma cell fraction. (B) GO enrichment analysis concerning of biological processes revealed the functions of hub genes in turquoise module. (C) The top ten hub genes in turquoise module and their GS and MM were presented by bar-pot. (D-F) The most plasma cell fraction correlated two genes were presented by scatter plot.

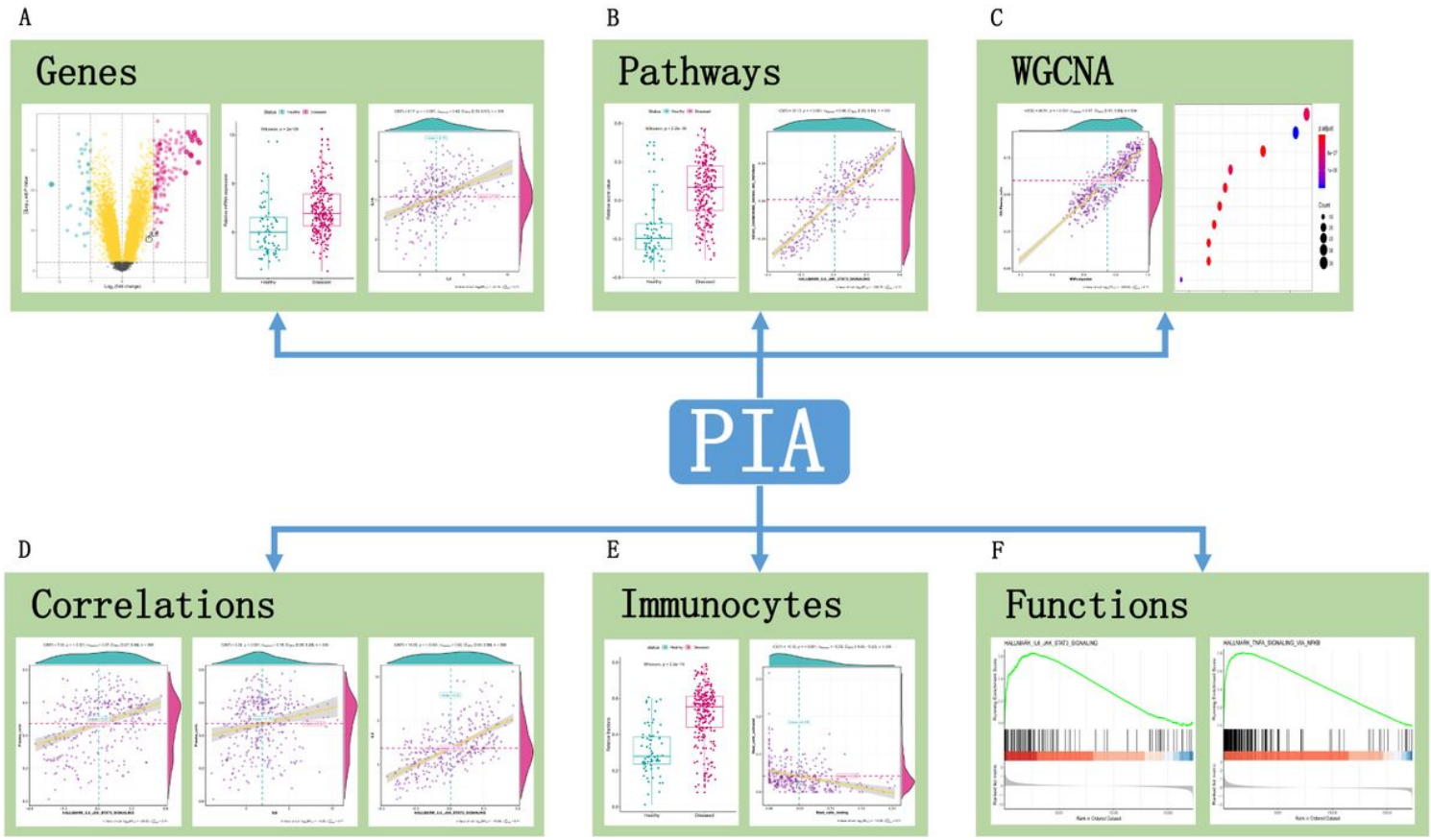

Figure 9

Overview of Periodontitis Immune Atlas app.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfigures.docx