MEDS Clinical Medicine (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/medsc.2023.040518
ISSN 2616-1907 Vol. 4 Num. 5

Identification of key gene associated with periodontitis and prediction of therapeutic drugs using machine learning in combination with LIME model explainer

Qilong Huang[1,a,*], Zhaohua Wang[2,b], Jiayu Han[3,c]

[1]Department of Biomedical Engineering, China Medical University, Shenyang, 110122, China
[2]Department of Otorhinolaryngology & Head and Neck Surgery Dalian Friendship Hospital of Dalian Medical University, Dalian, Liaoning, 116000, China.
[3]Yantai Research Institute, Harbin Engineering University, Yantai, Shandong, 264000, China
[a]huangql029@163.com, [b]doctorwang1990@hotmail.com, [c]1005141804@qq.com
**Corresponding author*

Keywords: Periodontitis, LIME, hub gene, molecular docking.

Abstract: Periodontitis is an immune-inflammatory disease characterized by irreversible periodontal attachment loss and bone destruction. In this study, we downloaded two microarray datasets, GSE10334 and GSE16134, from the Gene Expression Omnibus (GEO) database to identify molecular biomarkers and potential mechanisms associated with periodontitis. We performed differential gene expression analysis using the Limma package and co-expression network analysis. Additionally, we used machine learning with L1 regularization and LIME model explainer to identify the most relevant gene, ISL1. Finally, we validated molecular docking experiments using AutoDockTool and PyMOL. GO and KEGG enrichment analyses showed that periodontitis may affect various biological processes, including transcription, gene expression, apoptosis, and proliferation regulation. We found that periodontitis may influence cytokine-cytokine receptor interaction, lipid and atherosclerosis, and IL-17 signaling pathway. Our molecular docking results demonstrated that all of the major targets selected could be stably bound by the active components we chose. In summary, this study provides the hub gene, ISL1. We also identified 9 active components that may play a role in regulating ISL1 in periodontitis.

1. Introduction

Periodontitis is a prevalent chronic immune-inflammatory disease triggered by microbial plaque, which is characterized by gradual loss of soft tissue support and bone resorption[1]. The pathophysiology of periodontitis is marked by an excess of pro-inflammatory factors required for inflammation resolution and insufficient resolution factors[2]. The Fourth National Oral Health Epidemiological Survey in China indicated that 87%-97% of Chinese adults exhibit varying degrees of periodontal disease. If left untreated, periodontitis can result in tooth mobility and eventual tooth loss[3,4]. Furthermore, periodontitis is associated with systemic diseases such as cardiovascular disease[5], Alzheimer's disease[6], diabetes, and insulin resistance. Thus, early diagnosis of

periodontitis is critical for protecting the alveolar bone, maintaining tooth stability, and potentially preventing related diseases[7].

A prerequisite for utilizing machine learning to screen for periodontitis-related genes is the design of strategies that not only perform well on training data but also generalize well to new inputs. Regularization techniques are explicitly designed to reduce testing error and are defined as "modifications to the learning algorithm aimed at reducing the generalization error instead of the training error."[8] In other words, the objective of regularization is to prevent overfitting, reduce generalization error, and enhance generalization ability. Developing more effective regularization strategies has become one of the primary research topics in machine learning. Currently, a variety of regularization strategies exist, with the most basic method entailing adding a penalty term to the original objective function to penalize models with high capacity[9]. The mathematical expression is as follows:

$$\tilde{J}(\theta; X, \boldsymbol{y}) = J(\theta; X, \boldsymbol{y}) + \alpha \cdot \Omega \tag{1}$$

where X and y are the training samples and their corresponding labels, θ is the parameter, J is the objective function, Ω is the penalty term, and α controls the strength of regularization. Different Ω functions have different preferences for the optimal solution of the parameter θ, resulting in varying regularization effects. In deep learning, it is common practice to regularize only the weights and not the biases. The two most commonly used Ω functions are L1 norm and L2 norm. When p = 1, it is the L1 norm, which represents the sum of the absolute values of the nonzero elements in the vector. According to the definition of the LP norm, the mathematical form of the L1 norm is as follows:

$$\| x \|_1 = \sum_{i=1}^{n} | x_i |. \tag{2}$$

The L1 norm is usually used to identify the optimal and sparse feature items[10].

In this study, we first observed the differential expression of various expression profiles based on GSE10334 and GSE16134 in periodontitis and healthy samples. Functional analysis revealed that the differentially expressed genes mainly involved immune-related biological processes. Additionally, the CIBERSORT algorithm demonstrated significant differences in the abundance of most immune cells between periodontitis and healthy samples. The central genes were identified through L1 regularization and LIME model interpretation, which facilitates an understanding of the pathogenesis of periodontitis and may serve as a therapeutic target.

2. Materials and methods

2.1 Microarray data acquisition

The GEO database (https://www.ncbi.nlm.nih.gov/geo/) provided two datasets: GSE10334, GSE16134. Table 1 gives more information about the gene expression profiles used in this study. GSE10334 [11] and GSE16134[12] based on GPL570 platform included array based gene expression profiles of periodontitis.

Table 1 Characteristics of datasets in this study

GSE series	Platform	Total	Periodontitis	Control
GSE10334	GPL570	247	183	64
GSE16134	GPL570	310	241	69

2.2 Data merging and Differentially Expressed Genes (DEGs) selection

The series matrix files were converted to gene symbol codes using Active Perl 5.30.0 software

(https://www.activestate.com/products/perl/). Then the ‘combat’ function of the ‘SVA’ package of R software was used to adjust batch effects using empirical Bayes models after all microarray data had been merged. Finally, we used the ‘normalize’ function of the ‘Limma’ package in R software to normalize the expressions of the datasets[13]. A gene was defined as a DEG between the periodontitis and normal samples when the adjusted P value was<0.05 and the |log2FC| >1, which were visualized as Volcano plots and heat map plots.

2.3 GO and KEGG Analysis

The Gene Ontology (GO) database comprised categories of Biological Processes (BP), Cellular Composition (CC), and Molecular Function (MF). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were derived from the org.hs.eg.db package, clusterProfiler package (https://github.com/YuLab-SMU/clusterProfiler), and ggplot2 package (version 3.3.6 for visualization) in the R software. Homo sapiens was designated as the species of interest, with a screening threshold of p.adjust<0.05, to obtain the primary enriched functions and pathways.

2.4 Immune infiltration analysis

The CIBERSORT algorithm was utilized to analyze the immune landscape of the microenvironment between the normal and periodontitis groups. The combined dataset served as the gene expression input, with the LM22 gene signature file consisting of 22 immune cell types. The analysis was conducted with 1, 000 permutations, and the resulting CIBERSORT values represented the fraction of immune cell infiltration per sample.

2.5 Screen hub gene

L1 regularization is a common method in linear regression, reducing model complexity and overfitting by adding an L1 norm penalty term to the loss function. This leads to sparse solutions, making it useful in machine learning applications like LASSO regression and sparse coding.

This method is particularly effective in identifying genes related to periodontitis, a prevalent oral disease with genetic links. By screening gene expression profiling data, L1 regularization can pinpoint relevant genes, using the size and sign of model parameters for analysis.

LIME (Local Interpretable Model-Agnostic Explanations) is another valuable tool, capable of explaining predictions from any black-box model. It constructs a locally interpretable model for each instance, making it useful for explaining periodontitis-related features and genes. It trains a black-box model using gene expression profiling data, predicts periodontitis, and then explains the prediction for a particular sample. This process involves generating a similar dataset, calculating feature contributions, selecting important features, and interpreting the model's prediction. This enhances understanding of periodontitis pathogenesis and can provide new diagnostic and treatment insights.

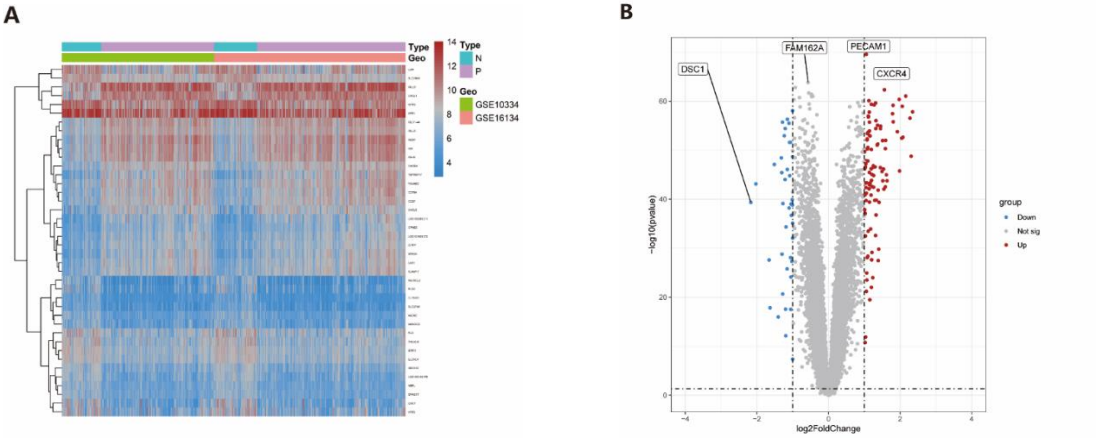
2.6 Molecular Docking Verification

Download 3D structures of 9 potentially active ingredients from the PubChem database(https://pubchem.ncbi.nlm.nih.gov/).The 3D structure of the hub gene is download from the PDB protein database(http://www.rcsb.org/pdb/home/home).Then the protein was dehydrated and ligand extracted with PyMOL software. Then Autodock software was used to conduct molecular simulation docking between 9 potential active ingredients and hub gene, and the binding strength of hub gene and 9 active ingredients was evaluated according to the docking binding energy.

3. Results

3.1 Identification of DEGs

Gene expression of merged GEO series that have been adjusted for batcheffects were standardized. The DEGs were analyzed using the 'Limma'package. After consolidation and normalization, 146 DEGs (|logFC| >1, P < 0.05) between untreated and periodontitis subjects were screened. Among them, 107 genes were upregulated and 39 genes were downregulated. We select 20 upregulated and 20 downregulated show in the heatmap (Figure 1A). A volcano plot was used to show the upregulation and downregulation (as shown in Figure 1B).

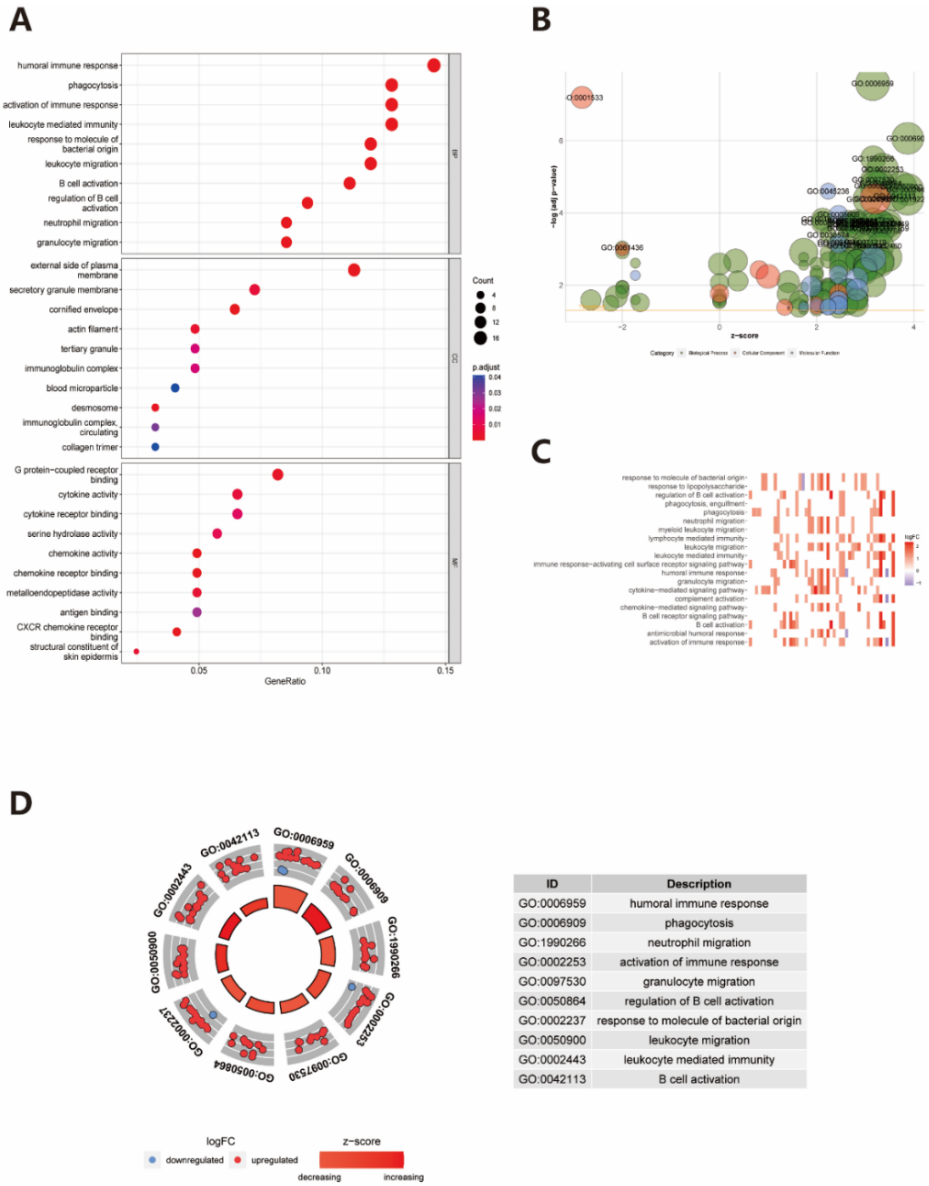


(A) Heat maps of 40 DEGs were selected. (B) volcanal map of all the DEGs.

Figure 1: Identification of DEGs.

3.2 GO term analysis

We analyzed the DEGs using GO analyses to learn more about the biological functions involved in periodontitis samples. As shown in Figure 2A, changes in GO biological processes (BP) mainly included humoral immune response, Phagocytosis and activation of immune response. Genes primarily enriched in CC category were cornified envelope and external side of plasma membrane, Moreover, molecular function (MF) section, changes were significant in chemokine activity, chemokine receptor binding and G protein−coupled receptor binding. As shown in Figure 2.



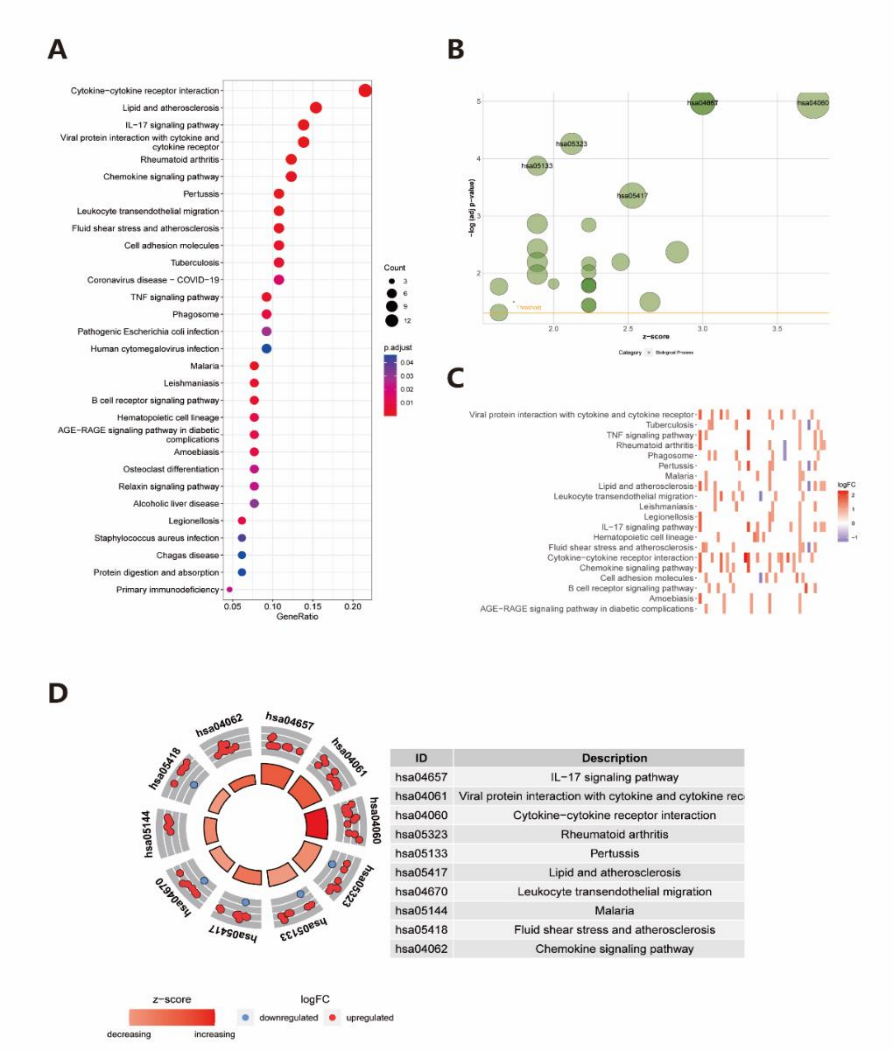
ID	Description
GO:0006959	humoral immune response
GO:0006909	phagocytosis
GO:1990266	neutrophil migration
GO:0002253	activation of immune response
GO:0097530	granulocyte migration
GO:0050864	regulation of B cell activation
GO:0002237	response to molecule of bacterial origin
GO:0050900	leukocyte migration
GO:0002443	leukocyte mediated immunity
GO:0042113	B cell activation

(A) result of GO enrichment (B) Bubble plot of GO terms (C) Heatmap plot of GO terms (D) circle plot of GO terms.

Figure 2: GO term analysis.

3.3 KEGG pathway enrichment analysis

KEGG pathway analyses were performed using the R software cluster Profiler package. In Figure 3 it showed that DEGs were significantly associated with Cytokine−cytokine receptor interaction, Lipid and atherosclerosis, IL−17 signaling pathway, Viral protein interaction with cytokine and cytokine receptor, Rheumatoid arthritis and Chemokine signaling pathway.



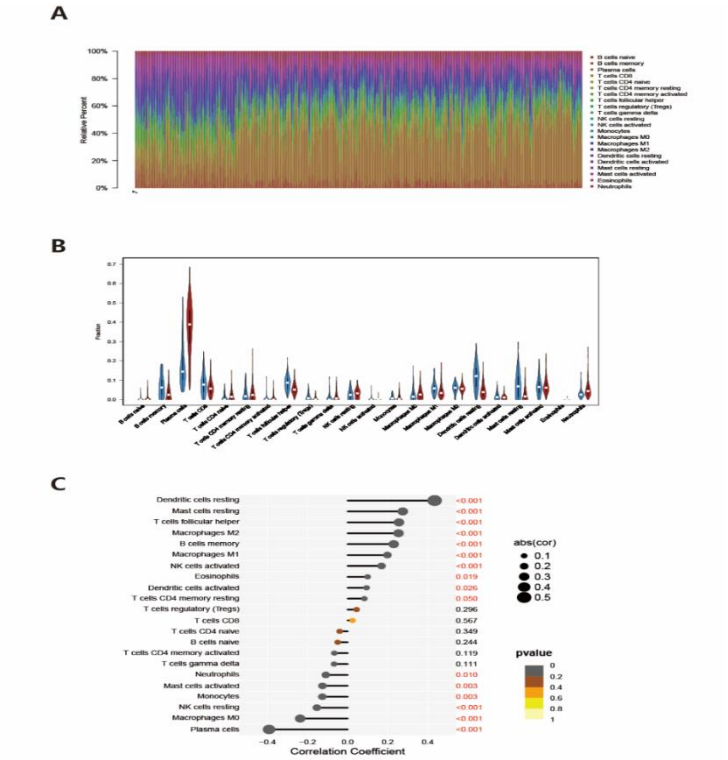
ID	Description
hsa04657	IL−17 signaling pathway
hsa04061	Viral protein interaction with cytokine and cytokine rec
hsa04060	Cytokine−cytokine receptor interaction
hsa05323	Rheumatoid arthritis
hsa05133	Pertussis
hsa05417	Lipid and atherosclerosis
hsa04670	Leukocyte transendothelial migration
hsa05144	Malaria
hsa05418	Fluid shear stress and atherosclerosis
hsa04062	Chemokine signaling pathway

(A) result of KEGG enrichment (B) Bubble plot of KEGG terms (C) Heatmap plot of KEGG terms (D) circle plot of KEGG terms.

Figure 3: KEGG pathway enrichment analysis.

3.4 Immune landscape of periodontitis

Moreover, the CIBERSORT algorithm was used to quantify the proportions of immune cells to evaluate the associations between the dataset and the immune microenvironment (as shown in Figure 4A). After that, the difference in immune infiltration between periodontitis and untreated groups was investigated in 22 immune cell types. The periodontitis group had a significantly higher ratio of Plasma cells (As show in Figure 4B). Next, we assessed the correlation between ISL1 and immune cells (As show in Figure 4C).



(A) immune cell distribution (B) the landscape of immune (C) the fraction of immune cells in normal and periodontitis groups.

Figure 4: Immune landscape

3.5 Identification of hub gene

In this study, we applied L1 regularization and LIME model interpretability techniques to identify key genes associated with a particular disease. L1 regularization is a widely used method for feature selection in machine learning, which penalizes model coefficients that are not relevant to the prediction task (As show in Figure 5). LIME is a model-agnostic technique that explains the predictions of any machine learning model by approximating its behavior in the local neighborhood of a given instance (As show in Figure 6).

Using these techniques, we were able to identify ISL1 as a key gene associated with the disease under study. ISL1 is a transcription factor that plays a crucial role in the development of various tissues, including the heart and nervous system. Our analysis suggests that ISL1 may be a potential therapeutic target for this disease.

Overall, our results demonstrate the effectiveness of combining L1 regularization and LIME model interpretability techniques for identifying key genes and potential therapeutic targets in complex diseases. Further studies are needed to validate the role of ISL1 in this particular disease and to explore its potential as a therapeutic target.

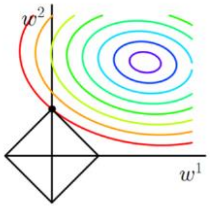


Figure 5: L1 regularization

The isoline in the figure is J0, The black square is the figure of the L function $L = |w^1| + |w^2|$, In the graph, when the J0 isoline intersects the L graph for the first time, it is the optimal solution.

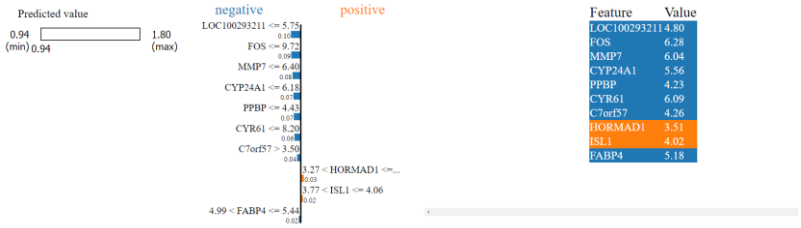


Figure 6: XML visualization generated from LIME's API package

3.7 Molecular Docking Verification

We verified the binding energy of potential chemical components of these 9 compounds on ISL1 using molecular docking technology, and the results are presented in Table 2. It is generally believed that the lower the binding energy of the ligand to the receptor, the more likely the ligand is to interact with the receptor. Our results showed that the binding energy of all 9 predicted active components with ISL1 was less than -6 kcal/mol, which indicates a potential interaction between them. The molecular docking mode is shown in Figure 7.

Table 2: The binding energy of active components to ISL1 by molecular docking.

n	gene	HGNC ID	Compound name	Binding energy (kcal/mol)
1	ISL1	6132	4-(5-benzo(1, 3)dioxol-5-yl-4-pyridin-2-yl-1H-imidazol-2-yl)benzamide	-8.8
2			3-nitrobenzanthrone	-8.2
3			Benzo(a)pyrene	-8.2
4			Dexamethasone	-7.5
5			dorsomorphin	-7.4
6			Fenretinide	-7.2
7			Tretinoin	-7
8			Pioglitazone	-6.9
9			bisphenol A	-6.2

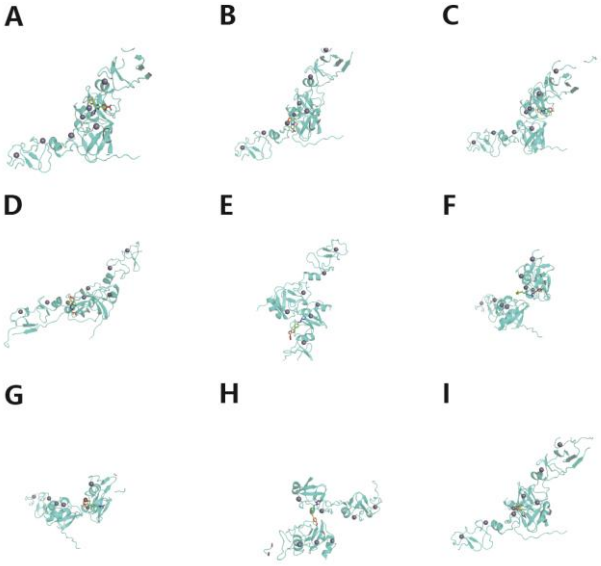


Figure 7: Schematic diagram of molecular docking of 9 potential drugs

4. Discussion

Periodontitis, an immunoinflammatory disease[14,15], can lead to irreversible bone destruction[16]. Traditional treatments, focusing on disrupting dental plaque biofilms, show unsatisfactory prognosis in some populations. Hence, understanding its etiological mechanism is crucial for comprehensive treatment strategies.

Using L1 regularization and LIME, we identified ISL1 as a key gene in periodontitis. ISL1, a protein, regulates Bmp4 transcription[17], and research suggests that dental epithelial stem cells could potentially generate new teeth. However, understanding of their regulation isn't sufficient for successful implementation. Animal studies show Fgf10 as a major regulator of dental epithelial stem cell niche[18], with Shh signaling activity maintaining the stem cell niche. The FAK-YAP-mTOR pathway regulates the balance between stem cell proliferation and differentiation into enamel-forming cells[19,20], with ISL1 expression and Shh signaling pathway activity crucial for proper enamel pattern.

Molecular docking validated the active ingredients that may have regulatory effects on the HUB gene ISL1. Among them, Benzo[a]pyrene hydrocarbon receptor signaling inhibits osteoblastic differentiation and collagen synthesis of human periodontal ligament cells[21]. Dorsomorphin attenuates Jagged1-induced mineralization in human dental pulp cells[22]. Fenretinide has been shown to have an anti-inflammatory effect. Fenretinide inhibited chemokine [23] and chemokine receptor expression [24] in vitro. In animal studies, fenretinide suppressed chronic arthritis induced by administration of streptococcal cell wall [25] and decreased the mRNA levels of proinflammatory mediators in the spinal cord after a spinal cord injury[26]. The application of topical tretinoin acid gel resulted in a 50 percent reduction in the incidence of oral leukoplakia. [27].

However, it is imperative to note that our study is subject to certain limitations. Firstly, the samples utilized in this investigation lacked essential clinicopathological information. As such, our identification of diagnostic markers was limited solely to the transcriptomic level. Secondly, the current transcriptomic datasets available for periodontitis in GEO were restricted, rendering validation of our findings challenging due to inadequate data. Thirdly, the outcomes of our bioinformatics analysis alone may not suffice to establish conclusive evidence, and as such, experimental validation is necessary to confirm our findings.

5. Conclusion

In this study, we analyzed the immunoregulatory effects, affected biological processes, and signaling pathways of periodontitis. We identified the most relevant gene in periodontitis, ISL1, by analyzing the dataset using L1 regularization and the LIME interpretability model. This finding provides new insights into the prevention and treatment of periodontitis. Additionally, we identified nine active ingredients that may play a role in regulating the ISL1 gene, which could contribute to further research on the pathogenesis of periodontitis.

References

[1] Hajishengallis G, Korostoff J M. Revisiting the Page & Schroeder model: the good, the bad and the unknowns in the periodontal host response 40 years later [J]. Periodontology 2000, 2017, 75(1): 116-151.

[2] Van Dyke T E. The management of inflammation in periodontal disease[J]. Journal of periodontology, 2008, 79: 1601-1608.

[3] Liu Z, Jiang M, Li Y. Extracellular vesicles in chronic periodontitis[J]. Chinese Journal of Tissue Engineering Research, 2023, 27(1): 99.

[4] Wang P, Wang B, Zhang Z, et al. Identification of inflammation-related DNA methylation biomarkers in periodontitis patients based on weighted co-expression analysis[J]. Aging (Albany NY), 2021, 13(15): 19678.

[5] Sanz M, Herrera D, Kebschull M, et al. Treatment of stage I–III periodontitis—The EFP S3 level clinical practice guideline [J]. Journal of clinical periodontology, 2020, 47: 4-60.
[6] Liccardo D, Marzano F, Carraturo F, et al. Potential bidirectional relationship between periodontitis and Alzheimer's disease [J]. Frontiers in Physiology, 2020, 11: 683.
[7] Bui F Q, Almeida-da-Silva C L C, Huynh B, et al. Association between periodontal pathogens and systemic disease[J]. Biomedical journal, 2019, 42(1): 27-35.
[8] Nusrat I, Jang S B. A comparison of regularization techniques in deep neural networks[J]. Symmetry, 2018, 10(11): 648..
[9] Pereyra G, Tucker G, Chorowski J, et al. Regularizing neural networks by penalizing confident output distributions[J]. arXiv preprint arXiv:1701.06548, 2017..
[10] Vidaurre D, Bielza C, Larranaga P. A survey of L1 regression[J]. International Statistical Review, 2013, 81(3): 361-387.
[11] R.T. Demmer, J.H. Behle, D. L. Wolf, M. Handfield, M. Kebschull, R. Celenti, P. Pavlidis, P.N. Papapanou, Transcriptomes in healthy and diseased gingival tissues, Journal of periodontology 79(11) (2008) 2112-24.
[12] P.N. Papapanou, J.H. Behle, M. Kebschull, R. Celenti, D.L. Wolf, M. Handfield, P. Pavlidis, R.T. Demmer, Subgingival bacterial colonization profiles correlate with gingival tissue gene expression, BMC microbiology 9 (2009) 221.
[13] Diboun I, Wernisch L, Orengo C A, et al. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma[J]. BMC genomics, 2006, 7(1): 1-14.
[14] Kinane D F, Stathopoulou P G, Papapanou P N. Periodontal diseases[J]. Nature reviews Disease primers, 2017, 3(1): 1-14.
[15] P.N. Papapanou, M. Sanz, N. Buduneli, T. Dietrich, M. Feres, D.H. Fine, T.F. Flemmig, R. Garcia, W.V. Giannobile, F.J.J.o.p. Graziani, Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions, 89 (2018) S173-S182.
[16] Branco-de-Almeida L S, Cruz-Almeida Y, Gonzalez-Marrero Y, et al. Treatment of localized aggressive periodontitis alters local host immunoinflammatory profiles: A long-term evaluation [J]. Journal of clinical periodontology, 2021, 48(2): 237-248.
[17] Jing Y, Ren Y, Witzel H R, et al. A BMP4-p38 MAPK signaling axis controls ISL1 protein stability and activity during cardiogenesis[J]. Stem Cell Reports, 2021, 16(8): 1894-1905.
[18] Harada H, Toyono T, Toyoshima K, et al. FGF10 maintains stem cell compartment in developing mouse incisors[J]. 2002..
[19] Hu J K H, Du W, Shelton S J, et al. An FAK-YAP-mTOR signaling axis regulates stem cell-based tissue renewal in mice[J]. Cell stem cell, 2017, 21(1): 91-106. e6.
[20] Naveau A, Zhang B, Meng B, et al. Isl1 controls patterning and mineralization of enamel in the continuously renewing mouse incisor[J]. Journal of Bone and Mineral Research, 2017, 32(11): 2219-2231.
[21] Monnouchi S, Maeda H, Yuda A, et al. Benzo [a] pyrene/aryl hydrocarbon receptor signaling inhibits osteoblastic differentiation and collagen synthesis of human periodontal ligament cells[J]. Journal of periodontal research, 2016, 51(6): 779-788.
[22] Manokawinchoke J, Watcharawipas T, Ekmetipunth K, et al. Dorsomorphin attenuates Jagged1-induced mineralization in human dental pulp cells[J]. International Endodontic Journal, 2021, 54(12): 2229-2242.
[23] Vilela R M, Lands L C, Meehan B, et al. Inhibition of IL-8 release from CFTR-deficient lung epithelial cells following pre-treatment with fenretinide[J]. International immunopharmacology, 2006, 6(11): 1651-1664.
[24] Villablanca E J, Zhou D, Valentinis B, et al. Selected natural and synthetic retinoids impair CCR7-and CXCR4-dependent cell migration in vitro and in vivo[J]. Journal of Leucocyte Biology, 2008, 84(3): 871-879.
[25] Haraqui B, Wilder R L, Allen J B, et al. Dose-dependent suppression by the synthetic retinoid, 4-hydroxyphenyl retinamide, of streptococcal cell wall-induced arthritis in rats[J]. International journal of immunopharmacology, 1985, 7(6): 903-916.
[26] López-Vales R, Redensek A, Skinner T A A, et al. Fenretinide promotes functional recovery and tissue protection after spinal cord contusion injury in mice[J]. Journal of Neuroscience, 2010, 30(9): 3220-3226.
[27] Kulkarni V, Bhatavadekar N B, Uttamani J R. Effect of nutrition on periodontal disease: a systematic review[J]. Journal of the California Dental Association, 2014, 42(5): 303-311.