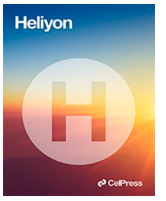
Heliyon 10 (2024) e29652

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

WGCNA and machine learning analysis identifi ed SAMD9 and IFIT3 as primary Sjögren's Syndrome key genes

Shu Liu [a,b,1], Hongzhen Chen [a,1], Lin Tang [b], Mian Liu [a], Jinfeng Chen [b], Dandan Wang [a,b,*]

[a] Department of Rheumatology and Immunology, Nanjing Drum Tower Hospital, Clinical College of Nanjing University of Chinese Medicine, China
[b] Department of Rheumatology and Immunology, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, China

ARTICLE INFO

Keywords:
Primary Sjögren's syndrome
SAMD9
IFIT3
Machine-learning
WGCNA

ABSTRACT

Background: Current treatments for primary Sjögren's Syndrome (pSS) are with limited effect, partially due to the heterogeneity and uncleared mechanism.
Methods: We got GSE40568 (Japan) and GSE40611 (USA), and analyzed them with WGCNA to find key Differentially expressed genes (DEGs) between pSS and healthy salivary glands (SG). Key pSS genes (KPGs) were further selected through 3 machine-learning methods. The expression of KPGs was validated via two other GEO datasets (GSE127952 and GSE154926). Infiltrated immune cells, ceRNA network, and potential compounds were explored.
Results: Our study identified 376 DEGs from the pSS patients, with 186 genes located in the "plum2" module, showing the strongest correlation with clinical characteristics. SAMD9 and IFIT3 emerged as KPGs with excellent diagnostic potential. SAMD9 demonstrated close association with immune cell infiltration. We constructed a lncRNA-miRNA-mRNA network comprising 2 KPGs, 12 miRNAs, 124 lncRNAs, and potential therapeutic targets.
Conclusion: In the investigation of pSS public datasets, our study revealed two potential critical mediators in the pathological process of pSS salivary glands, namely SAMD9 and IFIT3. Furthermore, we put forth a hypothesis regarding the ceRNA network and made predictions regarding potential therapeutic drugs targeting these two genes.

1. Introduction

Primary Sjögren's syndrome (pSS), affects approximately 0.1 % of the general population [1], is a chronic autoimmune disease characterized by infiltration of lymphocytes in the lacrimal and salivary glands (SGs), and inflammation, leading to dryness in the eyes and mouth [2,3]. Current treatment aim to alleviate symptoms, prevent complications, and enhance the quality of life, but with limited effect [4].

Significant attention has been devoted to exploring the mechanism, diagnostic biomarkers, and therapeutic strategies associated with this condition. However, despite these extensive efforts, the precise pathogenesis of the disease remains incompletely understood. At present, more research is needed to address the key pathways and biomarkers [5].

Utilizing an integrative bioinformatic approach, which involves the analysis of publicly available high-throughput omics data and

* Corresponding author.Nanjing Drum Tower Hospital, Nanjing, 210008, China
E-mail address: dandanwangffd@nju.edu.cn (D. Wang).
[1] Shu Liu, and Hong-zhen Chen are co-lead authors.

https://doi.org/10.1016/j.heliyon.2024.e29652
Received 25 August 2023; Received in revised form 10 April 2024; Accepted 11 April 2024
Available online 26 April 2024
2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

drug databases, enables the effective identification of novel molecular mechanisms and potential targets for drug development. We employed a machine learning-based integrative bioinformatic analysis on publicly accessible GEO datasets, with the objective of identifying key genes associated with pSS.

2. Materials and methods

2.1. Data acquisition and integration

The GEO database (https://www.ncbi.nlm.nih.gov/geo/) was used to determine the messenger RNA (mRNA) expression level in the colon tissue of pSS patients [6]. The SGs gene expression data of pSS patients, GSE40568 (Japan) [7] and GSE40611 (USA) [8] were obtained from the GEO database. There were 17 pSS and 18 healthy controls in the GSE40611 dataset, while 5 pSS and 3 healthy control individuals were included in the GSE40568 dataset, summarized in Table 1. The GSE127952 dataset (pSS: n = 8, Control: n = 6; Brazil) and the GSE154926 (pSS: n = 43, Control: n = 7; USA) were used as validation datasets. The "sva" R package was utilized for batch effect correction [9].

2.2. Exploration of the key genes for pSS

Differential gene analysis (DEGs) between the samples pSS patient and control individuals was carried out by using the "limma" package [10], with an adjusted P value < 0.05 and a $|\log2 \text{ fold change}| > 1$ as the criteria for detecting DEGs in the dataset comparing diseased and control samples. Genes with logFC >1 and adjusted P-value <0.05 were considered up-regulated genes, indicating increased expression. Genes with logFC < -1 and adjusted P-value <0.05 were considered down-regulated genes, indicating a decrease in expression. The results of the differential analysis will be used for subsequent analysis.

WGCNA may reveal disease-related gene networks and co-expressed gene modules with important biological implications [11]. Weighted Gene Co-Expression Network Analysis (WGCNA) was adopted to find gene modules correlated to clinical characters. During the process of transforming the weighted adjacency matrix into the topological overlap matrix, the nodes of the matrix were organized into a cluster tree using a hierarchical clustering approach [12]. The dissimilarity threshold for merging modules is 0.3 and the minimum module size was 30.

Finally, we used "VennDiagram" to look for overlap between core module genes and DEGs [13]. DEGs located in the gene modules which has the highest correlation with clinical characters were set as key DEGs (KDEGs), and utilized for function enrichment, including the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and the Gene ontology (GO) analysis. GO [14] enrichment is a commonly used method for large-scale functional enrichment studies which includes biological processes (BP), molecular functions (MF), and cellular components (CC). The KEGG [15] is a widely used database that stores information related to genomes, biological pathways, diseases, drugs, and more. We used the R package "clusterProfiler" [16] to perform GO annotation analysis on differentially expressed genes (DEGs). The criteria for selecting entries were P.adj <0.05 and a false discovery rate (FDR) value (q-value) < 0.05, which were considered statistically significant. The Benjamini-Hochberg (BH) method was used for P-value correction.

Machine learning, especially cluster and pattern recognition algorithms are now frequently used in science to analyze complex data sets [17], has significantly improved the predictive and accuracy value of key genes identified based on microarrays and next-generation sequencing data [18]. The least absolute shrinkage and selection operator (LASSO) regression and support vector machine-recursive feature elimination (SVM-RFE) algorithm are the most widely used machine learning methods to identify key genes [19]. The Lasso regression builds upon linear regression by adding a penalty term (lambda times the absolute value of the coefficients) to reduce overfitting and improve the model's generalization ability [20]. The SVM-RFE acts as an effective feature selection technique that finds the best variables by deleting the feature vector generated by SVM [21]. The RF modeling has been successfully applied to identify diseases in serum samples and to follow physiological processes in cells since its development by Breiman and became a well-established supervised machine learning technique that is known to perform well for medium to large size data sets while having only minimal requirements on data type and feature correlation [22]. However, few studies have combined LASSO, RF, and SVM-RFE to identify the key genes expressed in salivary glands of pSS. We conducted three machine-learning analyses to select Key pSS genes (KPGs) from the key DEGs.

2.3. Function analysis of the KPGs

The expression of KPGs was examined and graphed in both patients and healthy individuals by boxplots made with the "ggplot2" R

Table 1
A Summary of GEO datasets.

Dataset Type	GEO Number	Platform	pSS Samples	Control Samples	Country	
Testing sets	GSE40568	GPL570	5	3	Japan	2014
	GSE40611	GPL570	17	18	USA	2012
Validation sets	GSE127952	GPL20995	8	6	Brazil	2019
	GSE154926	GPL18573	43	7	USA	2021

pSS: primary Sjögren's syndrome.

package. The precision of each KPG was evaluated using receiver operating characteristic (ROC) curves by the "pROC" software [23], with an area under the curve (AUC) greater than 0.70 indicating potential diagnostic benefits for the disease.

Genes that potentially interacted with the KPGs were predicted using the GeneMANIA and annotated through GO and KEGG enrichment analysis. GeneMANIA is a remarkable database designed to construct protein-protein interaction networks. It can be accessed at http://www.genemania.org. This database provides a visual representation of the functional connections between genes, encouraging further exploration into their roles and functions. The website offers a diverse range of bioinformatics research methods, such as analyzing physical interactions, gene enrichment, gene co-location, gene co-expression, and website prediction [24].

The investigation of the immunological microenvironment, which is made up of many different components, such as mesenchymal tissues, inflammatory cells, immune cells, fibroblasts, and a wide variety of cytokines and chemokines, is essential for understanding immune reactions and disease mechanisms. The CIBERSORT methodology was utilized to determine the percentages of 22 different kinds of immune cells in the salivary glands samples [25]. Additionally, spearman correlation testing was performed on invading immune cells and KPGs using the "corrplot" program.

A ceRNA-based Cytoscape regulatory network [26] for the KPGs was constructed based on Starbase (https://starbase.sysu.edu.cn/starbase2/index.php), miRTarBase (https://mirtarbase.cuhk.edu.cn/), and TargetScan (http://www.targetscan.org/vert_72/) databases [27,28,29]. In the ceRNA network, we used MicroRNAs (miRNAs), the ceRNA control protein translation by binding the target

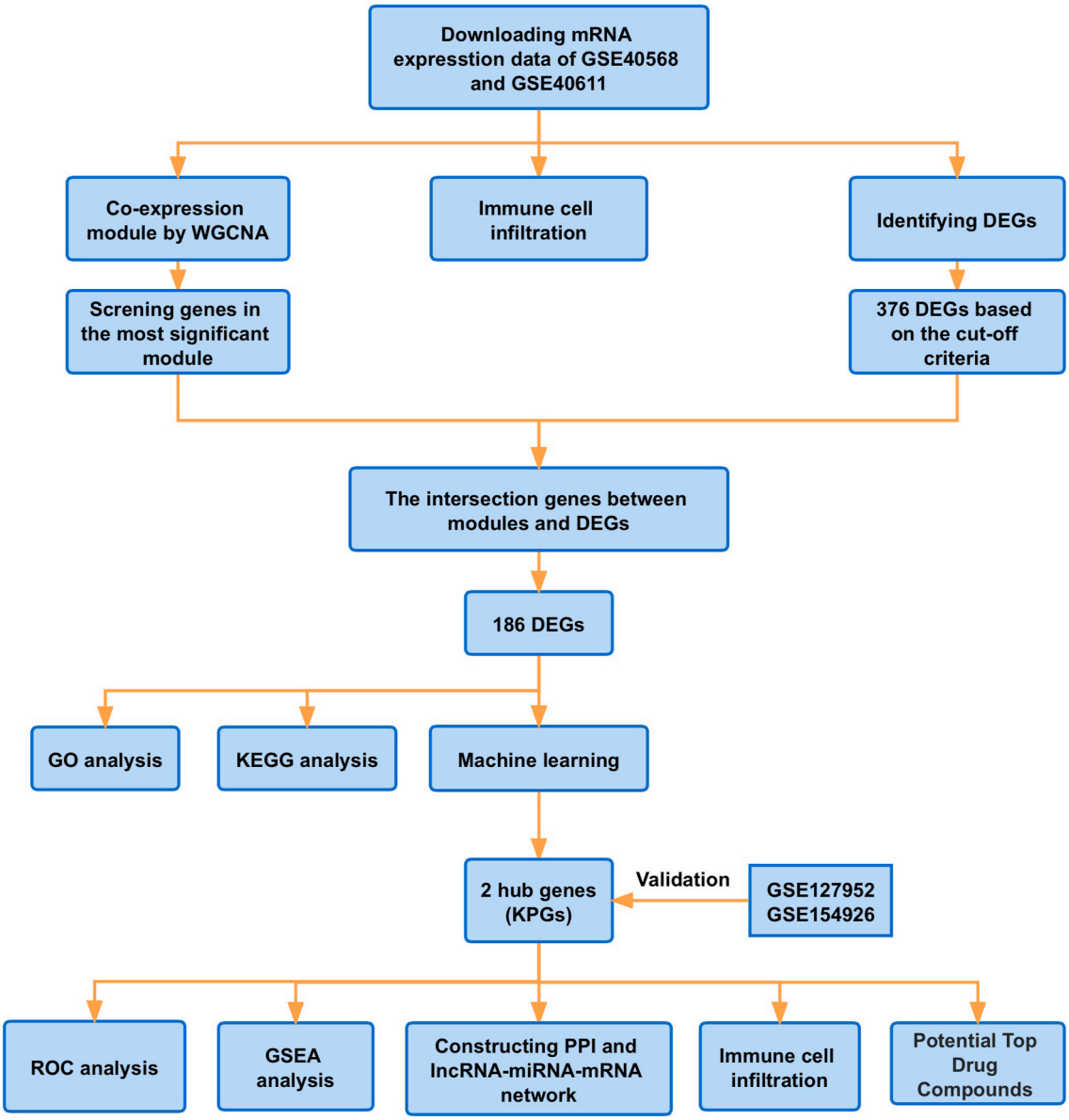


Fig. 1. Study flow chart for this work.

mRNA's 3′ untranslated region [30], to link target mRNAs to their respective long non-coding RNAs (lncRNAs). Additional usage of Starbase for stringent prediction of miRNA-lncRNA interactions [29]. Potential drugs and chemical compounds targeting the KPGs were explored using the DSigDB database (http://tanlab.ucdenver.edu/DSigDB) [31], with a false discovery rate of 0.05 and a composite score of more than 3000 being selected as criteria. Finally, the expression of KPGs was validated in two other GEO datasets.

2.4. Statistics

Data was analyzed with *R* software and Prism (GraphPad Software, La Jolla, CA). Statistical significance was checked by *t*-test and Kruskal–Wallis test. Ns, no significance; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$.

3. Results

Fig. 1 depicts the study workflow. The batch effect between the two test datasets, GSE40568 (Japan) [7] and GSE40611, was corrected by the "sva" package (Fig. S1). A total of 376 DEGs were identified between the patients with pSS and the healthy controls (Fig. 2A and B). The module plum2, consisting of 291 genes (Figure S2 A-E), exhibited the strongest correlation with clinical characteristics both in pSS group and healthy control group (cor = 0.47, $P = 2.1e\text{-}17$, Fig. 2 C, and Figure S2 F-G). Among these 291 genes, 186 were also found to be DEGs between pSS patients and control individuals, leading them to be classified as KDEGs (Fig. 3 A). The function of these KDEGs has primarily associated with immune regulation such as the biological processes of antigen receptor-mediated signaling pathway, leukocyte-mediated immunity, and mononuclear cell differentiation, the immune receptor activity, and the chemokine signaling pathway (Fig. 3B and C). Through the three machine-learning analyses, namely LASSO regression (Fig. 3D and E), SVM-RFE (Fig. 3 F), and RF (Fig. 3G and H), two KPGs, IFIT3 and SAMD9 (Fig. 3D–I), were consistently identified in the top 10 most important genes.

Both IFIT3 and SAMD9 exhibited significant upregulation in pSS patients compared to healthy controls (Fig. 4A and B). Furthermore, ROC analysis demonstrated that they could serve as markers for pSS, as indicated by the AUC values (IFIT3: AUC = 0.959, 95% confidence interval [CI]: 0.874–1.000; SAMD9: AUC = 0.955, 95% CI: 0.879–1.000) (Fig. 4C and D).

The proportion of M1 macrophages, γδ T cells, activated CD4 memory T cells, naïve B cells, and resting dendritic cells was higher in

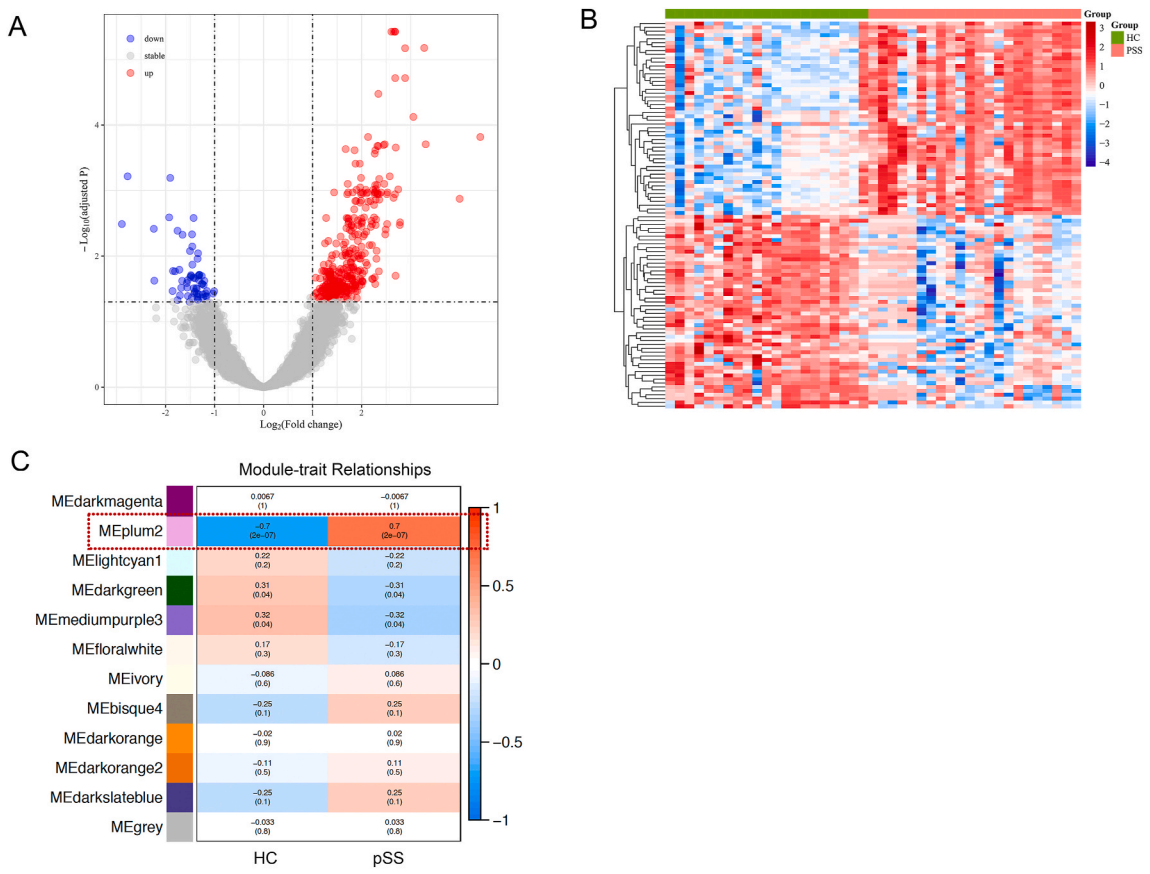


Fig. 2. DEGs between pSS patients and healthy control individuals. Volcano plot (A) and heatmap (B) of DEGs; (C) Module-trait relationships in pSS.

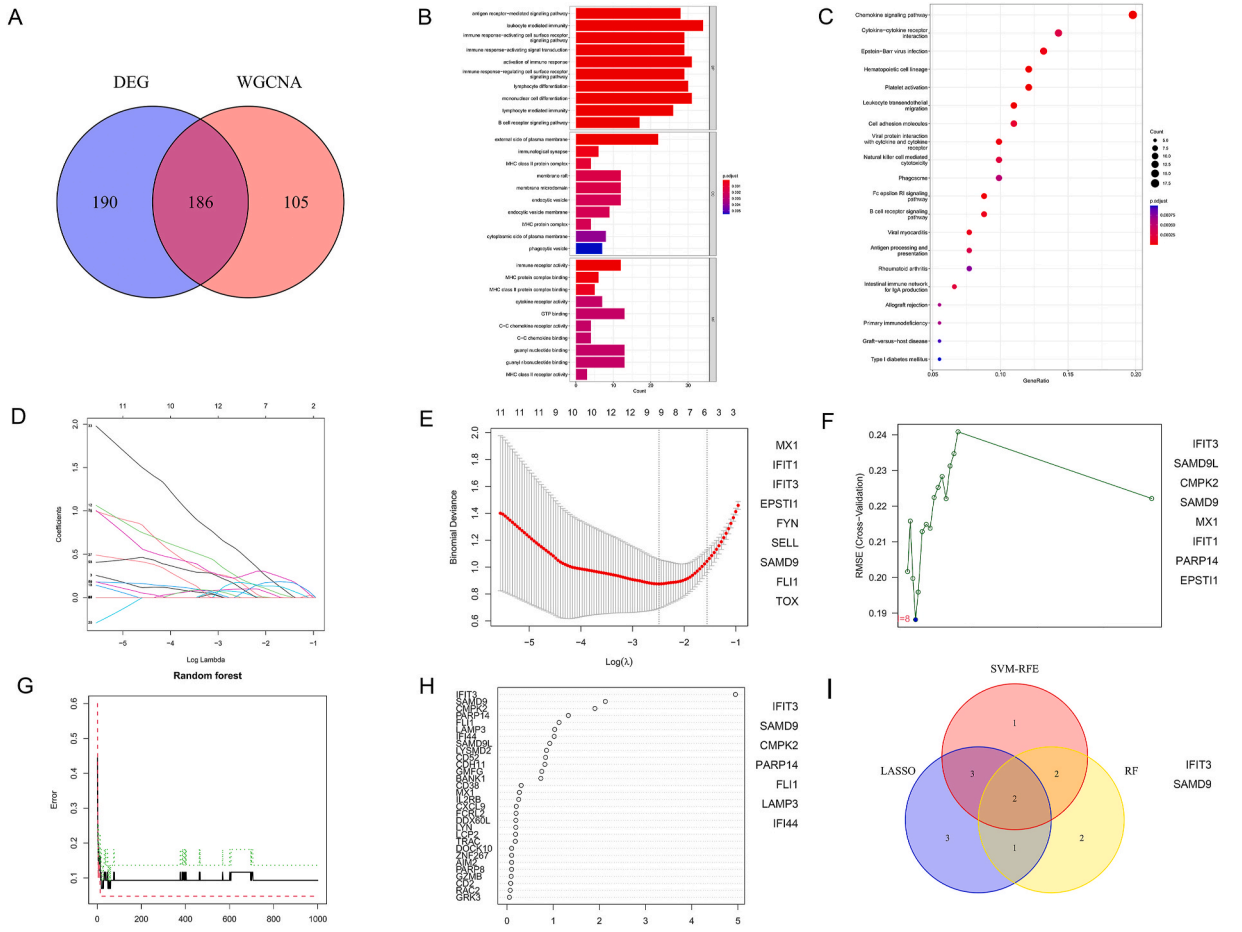


Fig. 3. Exploration of KPGs via machine-learning method. (A) Venn diagram of KDEGs; (B) Enriched GO terms of KDEGs; (C) Enriched KEGG pathways of KDEGs; Lasso regression (D–E), SVM-RF (F), and random forest analysis (G–H) revealed 2 KPGs in venn diagram (I). KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; MF, molecular function; BP, biological process; CC, cellular component; KDEGs, key DEGs; KPGs, Key pSS genes; SVM-RFE, Support Vector Machine Recursive Feature Elimination.

pSS samples, and the proportion of activated NK cells, CD8 T cells, resting NK cells, and activated mast cells were lower in pSS samples, even though there was no statistical significance (Fig. 4E and F). Both the IFIT3 and SAMD9 demonstrated closed associations with immune cell infiltrations (Fig. 4 G). In detail, SAMD9 was positively correlated with M1 macrophages, γδ T cells, and naïve CD4 T cells, negatively correlated with monocytes, resting dendritic cells, activated mast cells, and resting NK cells, while IFIT3 was positively correlated with M1 and M0 macrophages, and negatively correlated with activated mast cells, resting NK cells, monocytes, and resting dendritic cells.

The expression of the two KPGs (IFIT3 and SAMD9) showed an increasing trend, which was validated in the GSE127952 dataset (pSS: n = 8, Control: n = 6; Brazil) and the GSE154926 dataset (pSS: n = 43, Control: n = 7; USA), although only SAMD9 expression in the GSE127952 dataset was upregulated significantly in pSS patients ($P = 0.00056$, Fig. 4H and I).

Among the 20 genes found to interact with these two KPGs (IFIT3 and SAMD9), the top three genes were IFIT1, IFIT2, and IRF9, according to the GeneMANIA database (Fig. 5 A). Functional analysis revealed that these genes play a role in viral responses, adenylyltransferase activity, and virus-related diseases (Fig. 5B and C). The ceRNA network analysis identified 12 miRNAs and 124 lncRNAs (Fig. S3). The top 10 compounds predicted by DSigDB are presented in Table 2, and see the whole compound list in Supplementary files below.

4. Discussion

Primary Sjögren's syndrome (pSS) is a heterogeneous disease characterized by a wide range of manifestations that vary across different stages of the disease and among different patient subsets [5]. Several hypotheses have been proposed to explain the pathogenesis of pSS, including disruptions in both innate and adaptive immunity, as well as abnormalities at the interface between immune disorders and the neuroendocrine system, leading to dysfunction of lacrimal and secretory glands [32]. Genome-wide association studies (GWAS) have revealed that HLA genes show the strongest association signal with pSS, while epigenetic mechanisms such as

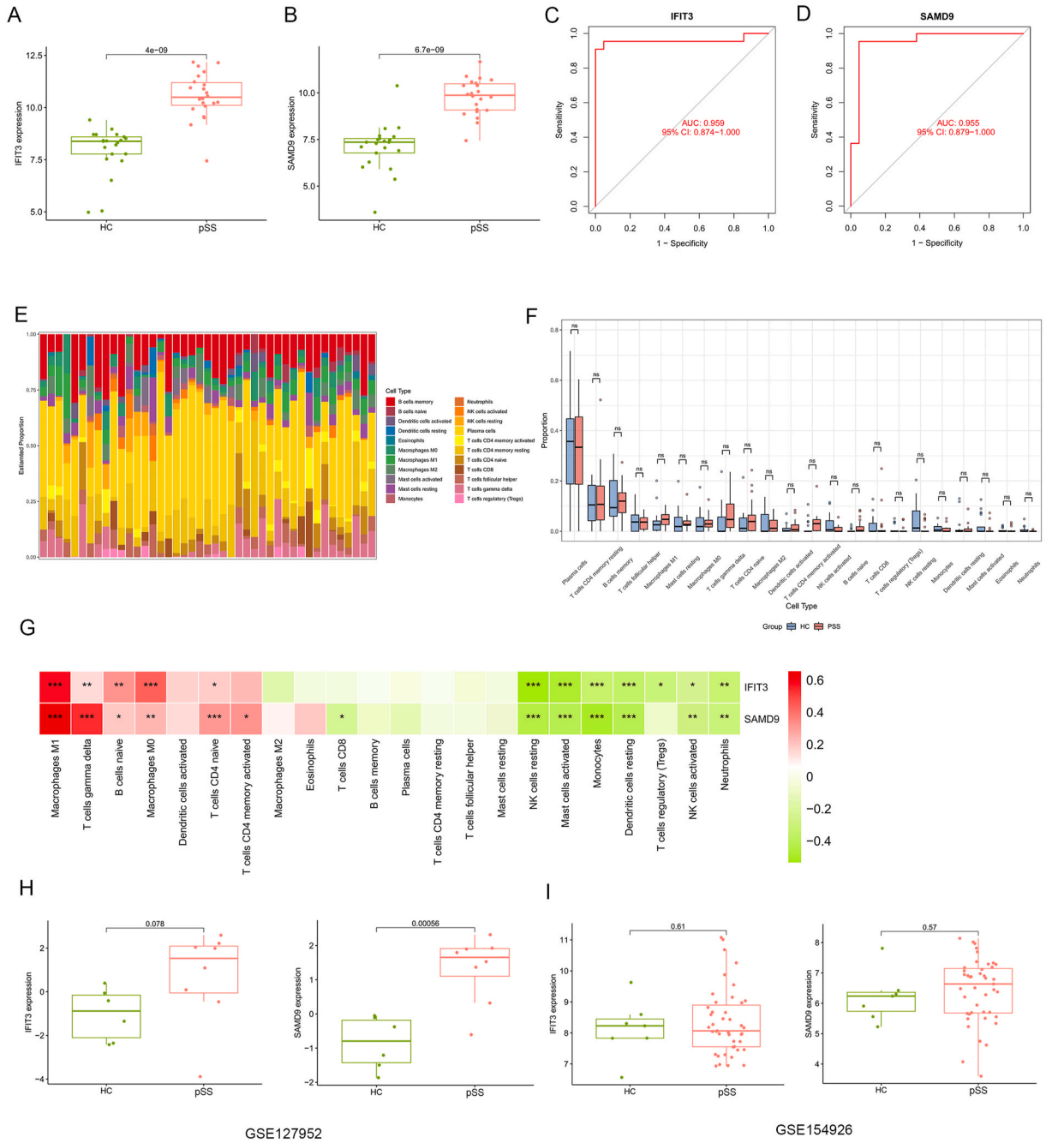


Fig. 4. KPGs expression and prodiction power. IFIT3 (A) and SAMD9 (B) was upregulated in pSS patients; ROC curve of IFIT3 (C) and SAMD9 (D) revealed strong predicting values for pSS; (E) propotion of immune cells in salivary glands (SGs); (F) grouped propotion of immune cells in SGs; (G) KPGs expression in various immune cells; (H) KPGs expression in GSE127952 dataset (pSS: n = 8, HC (health control): n = 6; Brazil); (I) KPGs expression in GSE154926 dataset (pSS: n = 43, HC (health control): n = 7; USA). AUC, area under the curve; ROC, receiver operating characteristic.

histone modifications, DNA methylation, and non-coding RNAs also contribute to the disease [33]. Additionally, ICOS, monocyte-related inflammation, and B cell-related genes have been identified as potentially important factors in pSS [34,35,36,37].

In our study, WGCNA results from DEGs between pSS patients and healthy controls revealed genes associated with immune regulation, which is not surprized. However, through three machine-learning analyses (LASSO regression, RF, and SVM-RFE), IFIT3 and SAMD9 were consistently identified as KPGs with strong power for pSS diagnosis, and further validated by two other GEO datasets from different countries, thus reliability to our findings. IFIT3 has been reported to have important roles in pSS. Type I IFN (IFN–I) activation is a prominent feature of pSS [38]. IFIT3, an IFN-I-stimulated genes, has been found to be common pathogenesis diagnostic

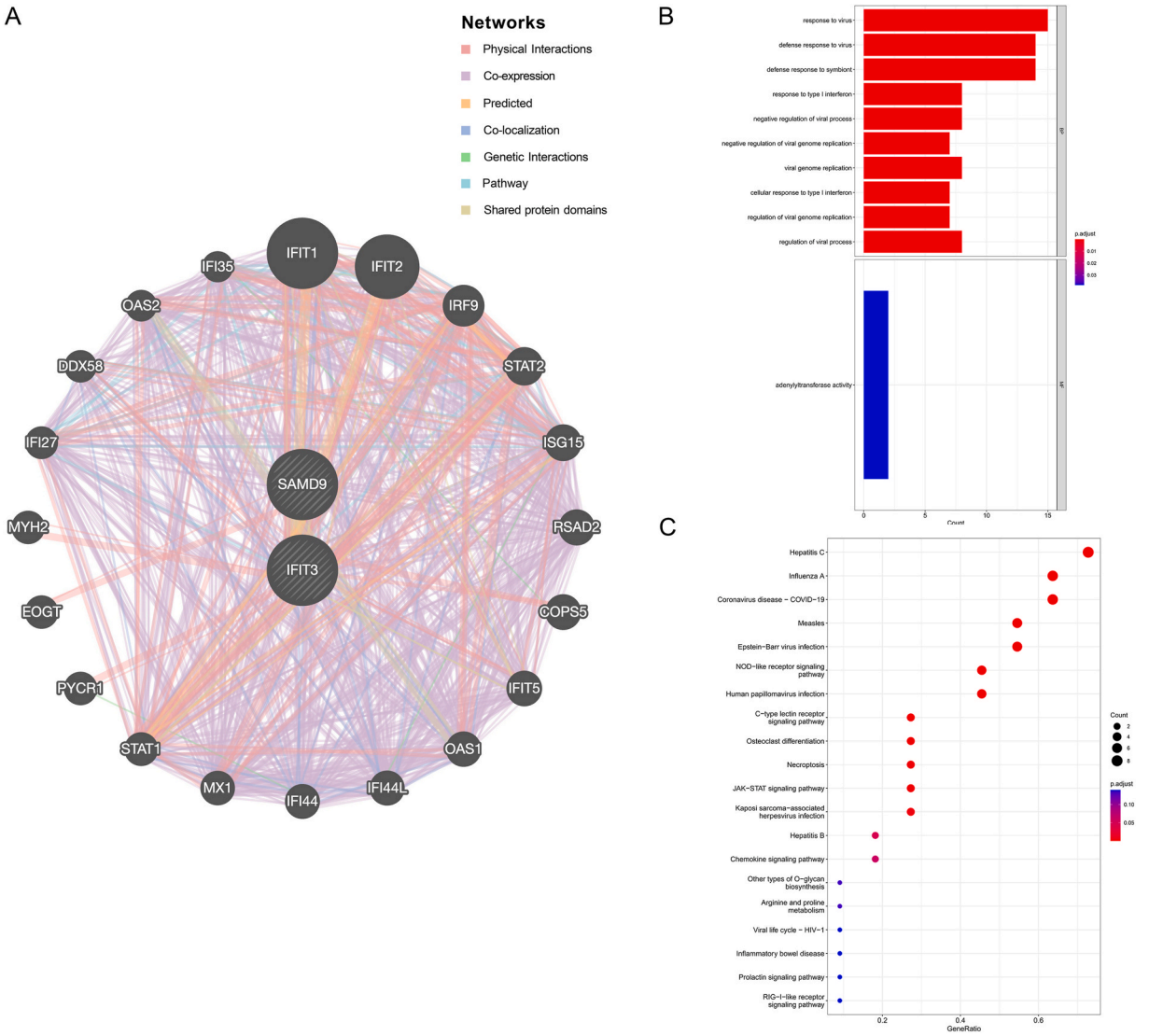


Fig. 5. Gens interacted with KPGs. Genes that potentially interacted with the KPGs (SAMD9 & IFIT3) were predicted using the GeneMANIA (A) and annotated through GO (B) and KEGG (C) enrichment analysis. KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; MF, molecular function; BP, biological process; CC, cellular component.

Table 2
The Top 10 Compounds predicted by DSigDB.

Compounds	P value	Adjusted P value	Combined Score
prenylamine HL60 UP	3.51E-06	2.21E-04	501385.4037
suloctidil HL60 UP	4.94E-05	0.00155453	393866.3146
naftopidil HL60 DOWN	0.002198806	0.024652725	5821.7173
hmba CTD 00000732	0.002598331	0.024652725	4755.879686
atovaquone MCF7 UP	0.002997776	0.024652725	4000.606286
chlorophyllin CTD 00000324	0.002997776	0.024652725	4000.606286
Gadodiamide hydrate CTD 00002623	0.004295424	0.024652725	2589.626019
DICHLOROMETHANE CTD 00006313	0.004894053	0.024652725	2211.014619
clioquinol PC3 UP	0.004894053	0.024652725	2211.014619
ETHYLBENZENE CTD 00000178	0.004993807	0.024652725	2157.56848

markers for COVID-19 and pSS [39]. The pSS serum had an enhanced capability of inducing IFIT3 in monocytic cell line [40], while IFIT3 in PBMCs of pSS patients can be inhibited by the BX795, an inhibitor of the IFN-I prmote gene TBK1 [41]. Khuder et al. also proclaimed that IFIT3 in SGs could represent a promising diagnostic biomarker for pSS [42].

A series of miRNAs and lncRNAs were found to interact with the KPGs (IFIT3 and SAMD9) in our study. Among these non-coding RNAs, miR-137 was reported to downregulated in tear sample of pSS patients [43], miR-300 was reported to upregulated in pSS patients' monocytes [44], miR-21-5p was found to mediate immunomodulation effect of extracellular vesicles derived from mesenchymal stromal cell for Sjögren's Syndrome model of mice [45]. These published articles further validated results of our study.

The pSS is an autoimmune disease characterized by lymphocytic infiltration and destruction of SGs [42]. Hongxiao Gong et al. found that in the SG tissues of pSS patients and mice model, the proportion of macrophages was increased, while the proportion of monocytes was decreased [46]. The distribution of immune cells in SGs presented with same trends in our results but with no statistical significance maybe due to the small sample. On the other hand, the upregulated KPGs in pSS tissue thus presented a positive correlation with macrophages and a negative correlation with monocytes, again validating the trend of immune cell proportion change.

We found the relationship between IFIT3 and SAMD9, which may involved in cell proliferation and apoptosis, and, also a series of miRNAs and lncRNAs, further explored potential drugs and compounds targeting the KPGs (IFIT3 and SAMD9). Nevertheless, our study has some limitations. First, the diagnostic efficacy of the KPGs should be validated using large-scale datasets, prospective trials, and controlled in vitro/vivo experiments, which would be undertaken in the future. Second, the immune infiltration analysis results contrasted with the existing theory and previous studies, suggesting the heterogeneous nature of pSS across various disease stages and patient subsets. So next generation RNA sequence method based on large scale patients population were neeeded for explicit mechansims of pSS among different disease stages and patient subsets.

5. Conclusion

In conclusion, our study identified two genes, SAMD9 and IFIT3, as the poential critical mediators in the pathological process of pSS salivary glands. We also predicted a ceRNA network that could be the upstream and downstream nodes for the mechanisms of SAMD9 and IFIT3 mediated pSS progress. Given the possible drug target role of these two genes, potential drugs and chemical compounds were also explored.

Data availability statement

All data used in this study are publicly available downloaded from GSE40568, GSE40611, GSE127952 and GSE154926 in GEO database without any access restrictions, and the analyzed scripts by *R* software are available upon request. Data included in article/supp. Material/referenced in article.

Funding statement

This study was supported by Jiangsu Provincial Key Research and Development Program (BE2020621).

CRediT authorship contribution statement

Shu Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Hong-zhen Chen:** Writing – original draft, Validation, Methodology, Investigation. **Lin Tang:** Validation, Data curation. **Mian Liu:** Validation, Software. **Jinfeng Chen:** Visualization, Validation, Data curation. **Dan-dan Wang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the GEO, miRTarBase, TargetScanand and StarBase for allowing them to upload the useful datasets.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e29652.

References

[1] G. Maciel, C.S. Crowson, E.L. Matteson, D. Cornec, Prevalence of primary Sjögren's syndrome in a US population-based cohort, Arthritis Care Res. 69 (10) (2017) 1612–1616, https://doi.org/10.1002/acr.23173.

[2] R.I. Fox, Sjögren's syndrome, Lancet 366 (9482) (2005) 321–331, https://doi.org/10.1016/s0140-6736(05)66990-5.

[3] T.D. Steinmetz, G.M. Verstappen, S.R. Schulz, L. de Wolff, R. Wilbrink, A. Visser, J. Terpstra, et al., Association of circulating antibody-secreting cell maturity with disease features in primary Sjögren's syndrome, Arthritis Rheumatol. 75 (6) (2023) 973–983, https://doi.org/10.1002/art.42422.

[4] L.L. Marshall, G.A. Stevens, Management of primary Sjögren's syndrome, Consult. Pharm. 33 (12) (2018) 691–701, https://doi.org/10.4140/TCP.n.2018.691.

[5] S. Zandonella Callegher, I. Giovannini, S. Zenz, V. Manfrè, M.H. Stradner, A. Hocevar, M. Gutierrez, et al., Sjögren syndrome: looking forward to the future, Ther Adv Musculoskelet Dis 14 (2022) 1759720x221100295, https://doi.org/10.1177/1759720x221100295.

[6] B. Tanya, T.D. B, W.S. E, L. Pierre, R. Dmitry, E. Carlos, K.I. F, et al., NCBI GEO: mining tens of millions of expression profiles–database and tools update, Nucleic Acids Res. 35 (Database issue) (2007), https://doi.org/10.1093/nar/gkl887.

[7] H. Tsuboi, Y. Nakai, M. Iizuka, H. Asashima, C. Hagiya, S. Tsuzuki, T. Hirota, et al., DNA microarray analysis of labial salivary glands in IgG4-related disease: comparison with Sjögren's syndrome, Arthritis Rheumatol. 66 (10) (2014) 2892–2899, https://doi.org/10.1002/art.38748.

[8] S. Horvath, A.N. Nazmul-Hossain, R.P. Pollard, F.G. Kroese, A. Vissink, C.G. Kallenberg, F.K. Spijkervet, et al., Systems analysis of primary Sjögren's syndrome pathogenesis in salivary glands identifies shared pathways in human and a mouse model, Arthritis Res. Ther. 14 (6) (2012) R238, https://doi.org/10.1186/ar4081.

[9] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (6) (2012) 882–883, https://doi.org/10.1093/bioinformatics/bts034.

[10] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47, https://doi.org/10.1093/nar/gkv007.

[11] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559, https://doi.org/10.1186/1471-2105-9-559.

[12] L. Song, P. Langfelder, S. Horvath, Comparison of co-expression measures: mutual information, correlation, and model based indices, BMC Bioinf. 13 (2012) 328, https://doi.org/10.1186/1471-2105-13-328.

[13] H. Chen, P.C. Boutros, VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R, BMC Bioinf. 12 (2011) 35, https://doi.org/10.1186/1471-2105-12-35.

[14] Y. Guangchuang, Gene ontology Semantic similarity analysis using GOSemSim, Methods Mol. Biol. (2020) 2117, https://doi.org/10.1007/978-1-0716-0301-7_11.

[15] K. M, G. S, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000), https://doi.org/10.1093/nar/28.1.27.

[16] W.L. Yu G, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012 May) 284–287.

[17] S. Moon, S. Chatterjee, P.H. Seeberger, K. Gilmore, Predicting glycosylation stereoselectivity using machine learning, Chem. Sci. 12 (8) (2020) 2931–2939, https://doi.org/10.1039/d0sc06222g.

[18] R. Chen, X. Liu, S. Jin, J. Lin, J. Liu, Machine learning for drug-target interaction prediction, Molecules 23 (9) (2018), https://doi.org/10.3390/molecules23092208.

[19] S. Wei, J. Lu, J. Lou, C. Shi, S. Mo, Y. Shao, J. Ni, et al., Gastric cancer tumor microenvironment characterization reveals stromal-related gene signatures associated with macrophage infiltration, Front. Genet. 11 (2020) 663, https://doi.org/10.3389/fgene.2020.00663.

[20] J. Gao, P.W. Kwan, D. Shi, Sparse kernel learning with LASSO and Bayesian inference algorithm, Neural Network. 23 (2) (2010) 257–264, https://doi.org/10.1016/j.neunet.2009.07.001.

[21] Q. Wang, X. Liu, Screening of feature genes in distinguishing different types of breast cancer using support vector machine, OncoTargets Ther. 8 (2015) 2311–2317, https://doi.org/10.2147/ott.S85271.

[22] J. Riedel, M. Lettow, M. Grabarics, M. Götze, R.L. Miller, G.J. Boons, G. Meijer, et al., Predicting structural motifs of glycosaminoglycans using cryogenic infrared spectroscopy and random forest, J. Am. Chem. Soc. 145 (14) (2023) 7859–7868, https://doi.org/10.1021/jacs.2c12762.

[23] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77.

[24] D. Warde-Farley, S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, et al., The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function, Nucleic Acids Res. 38 (Web Server issue) (2010) W214–W220, https://doi.org/10.1093/nar/gkq537.

[25] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (5) (2015) 453–457, https://doi.org/10.1038/nmeth.3337.

[26] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.

[27] V. Agarwal, G.W. Bell, J.W. Nam, D.P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs, Elife 4 (2015), https://doi.org/10.7554/eLife.05005.

[28] H.Y. Huang, Y.C. Lin, J. Li, K.Y. Huang, S. Shrestha, H.C. Hong, Y. Tang, et al., miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database, Nucleic Acids Res. 48 (D1) (2020) D148–d154, https://doi.org/10.1093/nar/gkz896.

[29] J.H. Li, S. Liu, H. Zhou, L.H. Qu, J.H. Yang, starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Res. 42 (Database issue) (2014) D92–D97, https://doi.org/10.1093/nar/gkt1248.

[30] L. He, G.J. Hannon, MicroRNAs: small RNAs with a big role in gene regulation, Nat. Rev. Genet. 5 (7) (2004) 522–531, https://doi.org/10.1038/nrg1379.

[31] M. Yoo, J. Shin, J. Kim, K.A. Ryall, K. Lee, S. Lee, M. Jeon, et al., DSigDB: drug signatures database for gene set analysis, Bioinformatics 31 (18) (2015) 3069–3071, https://doi.org/10.1093/bioinformatics/btv313.

[32] R.I. Fox, C.M. Fox, J.E. Gottenberg, T. Dörner, Treatment of Sjögren's syndrome: current therapy and future directions, Rheumatology 60 (5) (2021) 2066–2074, https://doi.org/10.1093/rheumatology/kez142.

[33] J. Imgenberg-Kreuz, A. Rasmussen, K. Sivils, G. Nordmark, Genetics and epigenetics in primary Sjögren's syndrome, Rheumatology 60 (5) (2021) 2085–2098, https://doi.org/10.1093/rheumatology/key330.

[34] J. Inamo, K. Suzuki, M. Takeshita, Y. Kassai, M. Takiguchi, R. Kurisu, Y. Okuzono, et al., Identification of novel genes associated with dysregulation of B cells in patients with primary Sjögren's syndrome, Arthritis Res. Ther. 22 (1) (2020) 153, https://doi.org/10.1186/s13075-020-02248-2.

[35] S. Lei, Y. Zhang, Identification of the key genes and pathways involved in B cells in primary Sjögren' s syndrome, Bioengineered 12 (1) (2021) 2055–2073, https://doi.org/10.1080/21655979.2021.1930753.

[36] A.P. Lopes, C.P.J. Bekker, M.R. Hillen, S.L.M. Blokland, A.C. Hinrichs, A. Pandit, A.A. Kruize, et al., The transcriptomic profile of monocytes from patients with Sjögren's syndrome is associated with inflammatory parameters and is mimicked by circulating mediators, Front. Immunol. 12 (2021) 701656, https://doi.org/10.3389/fimmu.2021.701656.

[37] J. Luo, X. Liao, L. Zhang, X. Xu, S. Ying, M. Yu, L. Zhu, et al., Transcriptome sequencing reveals potential roles of ICOS in primary Sjögren's syndrome, Front. Cell Dev. Biol. 8 (2020) 592490, https://doi.org/10.3389/fcell.2020.592490.

[38] E. Huijser, J. Göpfert, Z. Brkic, C.G. van Helden-Meeuwsen, S. Jansen, T. Mandl, P. Olsson, et al., Serum interferon-α2 measured by single-molecule array associates with systemic disease manifestations in Sjögren's syndrome, Rheumatology 61 (5) (2022) 2156–2166, https://doi.org/10.1093/rheumatology/keab688.

[39] H. Luo, X. Zhou, Bioinformatics analysis of potential common pathogenic mechanisms for COVID-19 infection and primary Sjogren's syndrome, Front. Immunol. 13 (2022) 938837, https://doi.org/10.3389/fimmu.2022.938837.

[40] M.E. Wildenberg, C.G. van Helden-Meeuwsen, J.P. van de Merwe, H.A. Drexhage, M.A. Versnel, Systemic increase in type I interferon activity in Sjögren's syndrome: a putative role for plasmacytoid dendritic cells, Eur. J. Immunol. 38 (7) (2008) 2024–2033, https://doi.org/10.1002/eji.200738008.
[41] I.L.A. Bodewes, E. Huijser, C.G. van Helden-Meeuwsen, L. Tas, R. Huizinga, V. Dalm, P.M. van Hagen, et al., TBK1: a key regulator and potential treatment target for interferon positive Sjögren's syndrome, systemic lupus erythematosus and systemic sclerosis, J. Autoimmun. 91 (2018) 97–102, https://doi.org/10.1016/j.jaut.2018.02.001.
[42] S.A. Khuder, I. Al-Hashimi, A.B. Mutgi, N. Altorok, Identification of potential genomic biomarkers for Sjögren's syndrome using data pooling of gene expression microarrays, Rheumatol. Int. 35 (5) (2015) 829–836, https://doi.org/10.1007/s00296-014-3152-6.
[43] Y.J. Kim, Y. Yeon, W.J. Lee, Y.U. Shin, H. Cho, Y.K. Sung, D.R. Kim, et al., Comparison of MicroRNA expression in tears of normal subjects and sjögren syndrome patients, Invest. Ophthalmol. Vis. Sci. 60 (14) (2019) 4889–4895, https://doi.org/10.1167/iovs.19-27062.
[44] A.E. Williams, K. Choi, A.L. Chan, Y.J. Lee, W.H. Reeves, M.R. Bubb, C.M. Stewart, et al., Sjögren's syndrome-associated microRNAs in CD14(+) monocytes unveils targeted TGFβ signaling, Arthritis Res. Ther. 18 (1) (2016) 95, https://doi.org/10.1186/s13075-016-0987-0.
[45] H. Kim, M.J. Lee, E.H. Bae, J.S. Ryu, G. Kaur, H.J. Kim, J.Y. Kim, et al., Comprehensive molecular profiles of functionally effective MSC-derived extracellular vesicles in immunomodulation, Mol. Ther. 28 (7) (2020) 1628–1644, https://doi.org/10.1016/j.ymthe.2020.04.020.
[46] H. Gong, X. Qiu, P. Li, R. Zhao, B. Wang, L. Zhu, X. Huo, Immune infiltration analysis reveals immune cell signatures in salivary gland tissue of primary Sjögren's syndrome, Front. Med. 10 (2023) 1033232, https://doi.org/10.3389/fmed.2023.1033232.