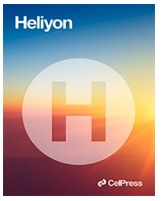
Heliyon 10 (2024) e33433

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Integrative multi-omics analysis reveals cellular and molecular insights into primary Sjögren's syndrome

Yao Tan [a,b], Jiayang Yin [a,b], Zhenkai Wu [c,**], Wei Xiong [a,*]

[a] Department of Ophthalmology, The Third Xiangya Hospital, Central South University, No. 138 Tongzipo Road, Yuelu District, Changsha City, 410013, Hunan Province, China
[b] Postdoctoral Station of Clinical Medicine, The Third Xiangya Hospital, Central South University, Changsha City, 410013, Hunan Province, China
[c] Department of Ophthalmology, Changde Hospital, Xiangya School of Medicine, Central South University, Changde, 415000, Hunan Province, China

ARTICLE INFO

Keywords:
Primary Sjögren's syndrome
Multi-omics analysis
Mendelian randomization
Biomarker
Single-cell RNA sequencing

ABSTRACT

Objective: This study aims to comprehensively analyze genomic, transcriptomic, proteomic, and single-cell sequencing data to unravel the molecular basis of primary Sjögren's syndrome (pSS) and explore potential therapeutic targets.
Methods: Mendelian randomization and single-cell RNA sequencing were employed to analyze pSS data. Differentially expressed genes specific to different blood cell types were identified. Integration of multiomics data facilitated the exploration of genetic regulatory relationships.
Results: The analysis revealed distinct cell clusters representing various immune cell subsets. Several genes, including cathepsin S (CTSS) and glutathione S-transferase omega 1 (GSTO1), were identified as potential biomarkers and therapeutic targets for pSS. Diagnostic utility analysis demonstrated the discriminatory power of CTSS and GSTO1 in distinguishing pSS patients from healthy controls.
Conclusion: The findings highlight the importance of integrating multiomics data for understanding pSS pathogenesis. CTSS and GSTO1 show promise as diagnostic biomarkers and potential therapeutic targets for pSS. Further investigations are warranted to elucidate the underlying mechanisms and develop targeted therapies for this complex autoimmune disease.

1. Introduction

Primary Sjögren's syndrome (pSS) is a globally prevalent autoimmune disease, most commonly affecting the exocrine glands and leading to significant dryness in the eyes and mouth [1]. The disease exhibits a higher predisposition towards women, with an approximately 9:1 female-to-male ratio, and generally manifests in the 4th to 5th decade of life [2]. Despite its significant prevalence, pSS is often underdiagnosed due to its non-specific symptoms and heterogeneous clinical presentation. More severe cases transcend the confines of exocrine gland involvement, leading to systemic manifestations that significantly impact patients' quality of life. The etiology of pSS is multifaceted, involving an interplay between genetic, hormonal, and environmental factors [3]. Though recent advancements in genomics have identified several genetic loci associated with pSS [4], providing valuable insights into the disease's genetic basis, the full extent of its genetic underpinnings is yet to be completely understood.

* Corresponding author.
** Corresponding author.
E-mail addresses: 56631585@qq.com (Z. Wu), weixiong420@csu.edu.cn (W. Xiong).

https://doi.org/10.1016/j.heliyon.2024.e33433
Received 11 August 2023; Received in revised form 16 June 2024; Accepted 21 June 2024
Available online 24 June 2024

The advent of single-cell RNA sequencing (scRNA-seq) has revolutionized the study of complex diseases such as pSS [5]. Leveraging scRNA-seq allows us to dissect the intricate cellular landscape and gene expression profiles with unprecedented single-cell resolution [6]. This methodology enables us to identify distinct gene expression patterns across various blood cell types, unveiling their unique roles in the disease process of pSS [7]. In this study, we employ scRNA-seq to analyze gene expression profiles from both pSS patients and healthy controls, with a special focus on the identification of differentially expressed genes that are specific to different blood cell types. Alongside scRNA-seq, another promising approach in understanding the intricate interplay between genetics and disease is Mendelian randomization (MR) [8]. When combined with a multiomics approach, which simultaneously analyses and integrates multiple layers of biological data (including genomic, transcriptomic, and proteomic information), MR serves as a powerful tool to unravel the complex molecular mechanisms underpinning diseases like pSS.

A focal point of our investigation revolves around the role of cathepsin S (CTSS), a protein-coding gene associated with immune responses [9–11]. Previous studies have implicated CTSS in the pathogenesis of pSS, suggesting it as a potential diagnostic marker and therapeutic target [12,13]. By scrutinizing the expression patterns of CTSS in pSS patients and healthy controls, we aim to evaluate its diagnostic efficacy and explore its potential as a drug target in pSS. Additionally, our study reveals the upregulation of glutathione S-transferase omega 1 (GSTO1) in pSS patients, indicating its involvement in immune dysregulation [14]. GSTO1 is an enzyme involved in detoxification processes and oxidative stress response [15,16]. Importantly, our diagnostic utility analysis demonstrates that GSTO1 expression levels exhibit excellent discriminatory power in distinguishing pSS patients from healthy controls. This highlights the potential of GSTO1 as a promising diagnostic biomarker for pSS. Further investigations into the specific mechanisms by which GSTO1 contributes to pSS pathogenesis and its potential as a therapeutic target are warranted. By comprehensively understanding the roles of CTSS and GSTO1, along with the cellular and molecular basis of pSS, we can develop more effective, targeted therapeutic approaches for this complex disease.

2. Materials and methods

2.1. Data sources

Our analyses utilized scRNA-seq data from the GSE157278 series, comprising samples from 5 primary pSS patients and 5 healthy controls (Fig. 1) [5]. This dataset offered a rich collection of 27,566 cells from the pSS patients and 30,741 cells from control individuals, enabling us to delineate diverse immune cell subsets and identify genes implicated in pSS. Incorporating summary-level data from Genome-wide Association Study (GWAS) of Sjögren's syndrome, we utilized the M13_SJOGREN dataset of the R9 version, sourced from the FinnGen consortium, which included 2495 cases and 365,533 controls [17]. Transcriptomic data were sourced from two Gene Expression Omnibus (GEO) datasets: the GSE84844 series (comprising 30 controls and 30 pSS samples) served as our training set [18], and the GSE48378 series (consisting of 16 controls and 11 pSS samples) served as the validation set [19]. As the data used in this study was publicly available (Supplementary Table 1), anonymized, and de-identified, it was exempt from ethical review approval by the Ethical Review Authority.

2.2. scRNA-seq data preprocessing and integration

The Seurat package (v4.0.2) was employed for data quality control and preprocessing of the scRNA-Seq data. Genes associated with red blood cells (Hba1, Hba2, Hbb, Hbd, Hbe1, Hbg1, Hbg2, Hbm, Hbq1, and Hbz) were removed. NormalizeData(.), FindVariableFeatures(.), and ScaleData(.) functions were utilized to normalize the scRNA-Seq data. Subsequently, the filtered scRNA-Seq datasets of the pSS and control samples were integrated using the Seurat anchor-based integration method [20]. Integration was based on the expression of the 3000 most variable counts in each sample. While we recognize the potential influence of cell cycle effects on gene expression profiles, our study was specifically focused on identifying gene expression changes in pSS that are independent of cell cycle variations. Consequently, we did not incorporate cell cycle status adjustment in our final analysis, aligning our efforts with the primary objective of elucidating disease-specific transcriptional signatures.

2.3. Dimensionality reduction, clustering, visualization, and cell type recognition

For dimensionality reduction, clustering, and visualization of the scRNA-Seq data, Seurat (v4) in R was employed. Principal component analysis (PCA) was utilized to reduce the dimensionality of the integrated data. The integrated dataset was further visualized using the uniform manifold approximation and projection (UMAP) algorithm. For the clustering of cells, we used the FindNeighbors and FindClusters functions with a specifically chosen resolution of 0.4, determined to provide the most informative and biologically relevant clusters. Cell types were annotated based on the canonical patterns of marker genes. The 58,307 cells were classified into 7 cell types, including myeloid cells, plasmacytoid dendritic cells (pDC), myeloid dendritic cells (mDC), natural killer cells (NK cells), T cells, B cells, and plasma cells. Cell annotation was performed using the CellMarker 2.0 reference database [21], which provides cell marker genes for accurate identification of cell types.

2.4. MR analysis and sensitivity checks

MR employs molecular quantitative trait data (eQTL or pQTL) as genetic instruments to deduce causal effects of gene/protein expression on disease outcomes. By selecting variants associated with expression as instrumental variables (IVs), MR analysis

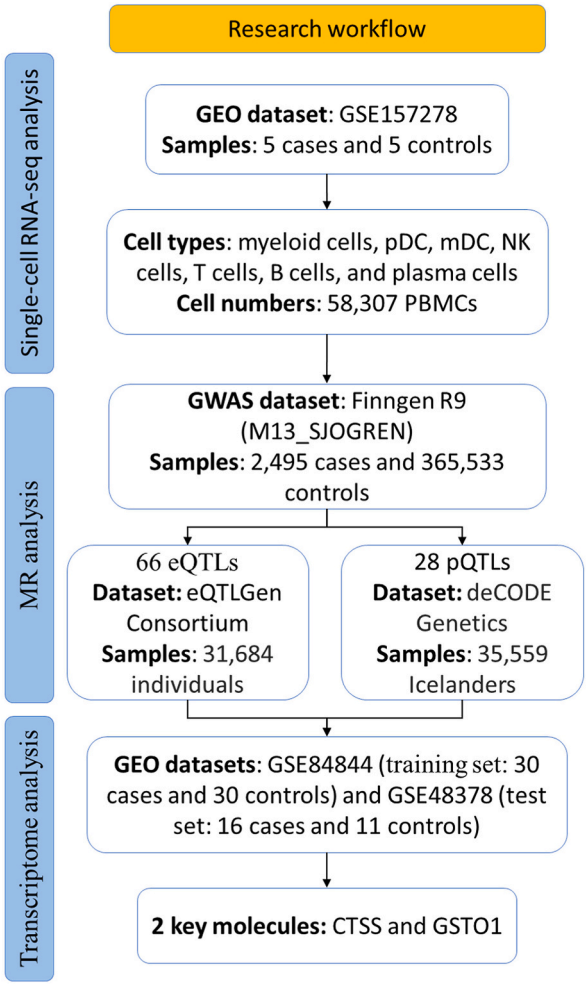
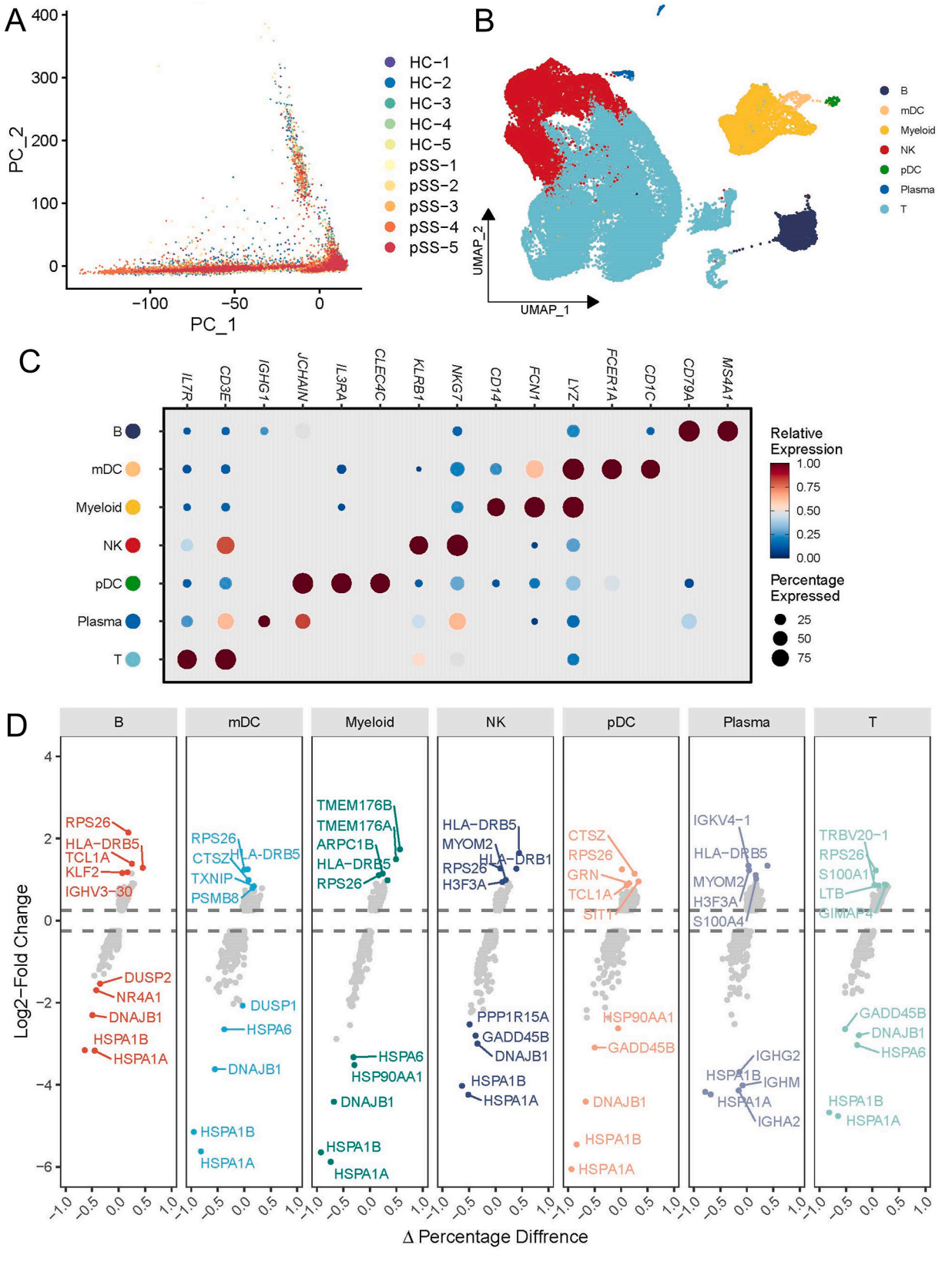


Fig. 1. Overview of research workflow. GEO: Gene Expression Omnibus, pDC: plasmacytoid dendritic cells, mDC: myeloid dendritic cells, NK cells: natural killer cells, PBMCs: Human peripheral blood mononuclear cells, GWAS: Genome-wide Association Study, MR: Mendelian randomization, eQTL: expression quantitative trait loci, pQTL: protein quantitative trait loci, CTSS: cathepsin S, GSTO1: glutathione S-transferase omega 1.

investigates the causal relationships between molecular traits and disease. This approach reveals the impact of gene/protein expression on disease development. Genetic variants linked to eQTL or pQTL served as our exposure instruments, while GWAS data on pSS formed the outcome data. For exposure instruments with multiple IVs, we conducted sensitivity checks to ensure our findings were robust. We initially tested for heterogeneity across variant-level MR estimates using the "mr_heterogeneity(.)" function, excluding exposures with significant heterogeneity (p-value <0.05). We also performed a pleiotropy test using MR Egger analysis to detect horizontal pleiotropy amongst IVs (more than 2 IVs), discarding any exposure with significant difference (p-value <0.05). Lastly, we implemented a Leave-one-out sensitivity test for exposures with more than 2 IVs. Each IV was sequentially removed to recalculate MR results, validating the robustness of our findings.

2.5. Data acquisition and selection of instrumental variables for MR analysis

We acquired the summary-level data for *cis*-expression quantitative trait loci (eQTL) in blood gene expression from the esteemed eQTLGen Consortium [22]. The dataset encompasses a substantial cohort of 31,684 blood samples. We conducted MR analysis using genetic instrumental variables from Ferkingstad et al.'s plasma protein quantitative trait loci (pQTL) dataset, involving GWASs on 4179 plasma proteins in 35,559 Icelanders [23]. They reported 18,084 significant pQTL associations, linking genetic variations to plasma protein levels. For robustness, we selected *cis*-single nucleotide polymorphisms (*cis*-SNPs), located within 1 Mb of the relevant gene and showing a genome-wide significant association (eQTL/pQTL p-value$<5 \times 10^{-8}$) with protein levels. Independent instrumental variables were identified via LD clumping in plink, using the 1000 Genomes project's European genotype data, considering SNPs with pairwise LD values (r^2) below 0.001 within a 10 Mb window.



(caption on next page)

Fig. 2. Overview of identified single cells. (A) PCA distribution of control and diseased group samples after batch effect removal. HC: Healthy Control, pSS: Primary Sjögren's syndrome. (B) UMAP of cell profiles, with different color blocks representing related cell clusters. pDC: plasmacytoid dendritic cells, mDC: myeloid dendritic cells, NK cells: natural killer cells. (C) Dot plots depicting the expression levels of selected marker genes across each cell type. (D) Volcano plots highlighting the top 5 marker genes of different blood cell subsets, including genes that are downregulated and upregulated.

2.6. Evaluation of model performance using ROC curve analysis in the pSS dataset

We partitioned the pSS dataset into a training set (n = 60) and a test set (n = 27). Utilizing the R package pROC (v1.18.2), we developed a model based on the training set to assess its performance. The model was then evaluated on the test set, and its efficacy was measured by accuracy, the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. Through this approach, we aimed to identify the most suitable method for our dataset by analyzing the ROC curve and associated metrics.

3. Results

3.1. Analysis on scRNA-seq of pSS

In this study, we obtained scRNA-seq data from 5 pSS patients and 5 healthy controls via microarray GSE157278 from the GEO database. We conducted a detailed analysis of three primary metrics: the gene count per cell (nFeature), the total gene count per cell (nCount), and the percentage of mitochondrial genes (percent.mt) within each cell. Our refined selection criteria, as depicted in Supplementary Fig. S1A, included cells with more than 300 nFeature, ensuring a robust gene expression profile, and with less than 20 % percent.mt to exclude potentially damaged cells. Additionally, our selection favored nCount predominantly below 30,000, balancing transcriptome complexity and data quality. Correlation analysis, as shown in Supplementary Figs. S1B and S1C, indicated a negative correlation between nCount and percent.mt and a positive correlation between nCount and nFeature, validating our selection parameters. Following these criteria, we identified and selected 3000 genes with significant expression variations for in-depth analysis (Supplementary Fig. S1D). The cell profiles, visualized using Uniform Manifold Approximation and Projection (UMAP), are distinctly marked by color blocks to denote sample origins (Supplementary Fig. S1E). In our analysis, we constructed a line graph to compare the cell subset proportions between pSS patients and control subjects (Supplementary Fig. S1F).

3.2. Overview of identified single cells and gene expression patterns

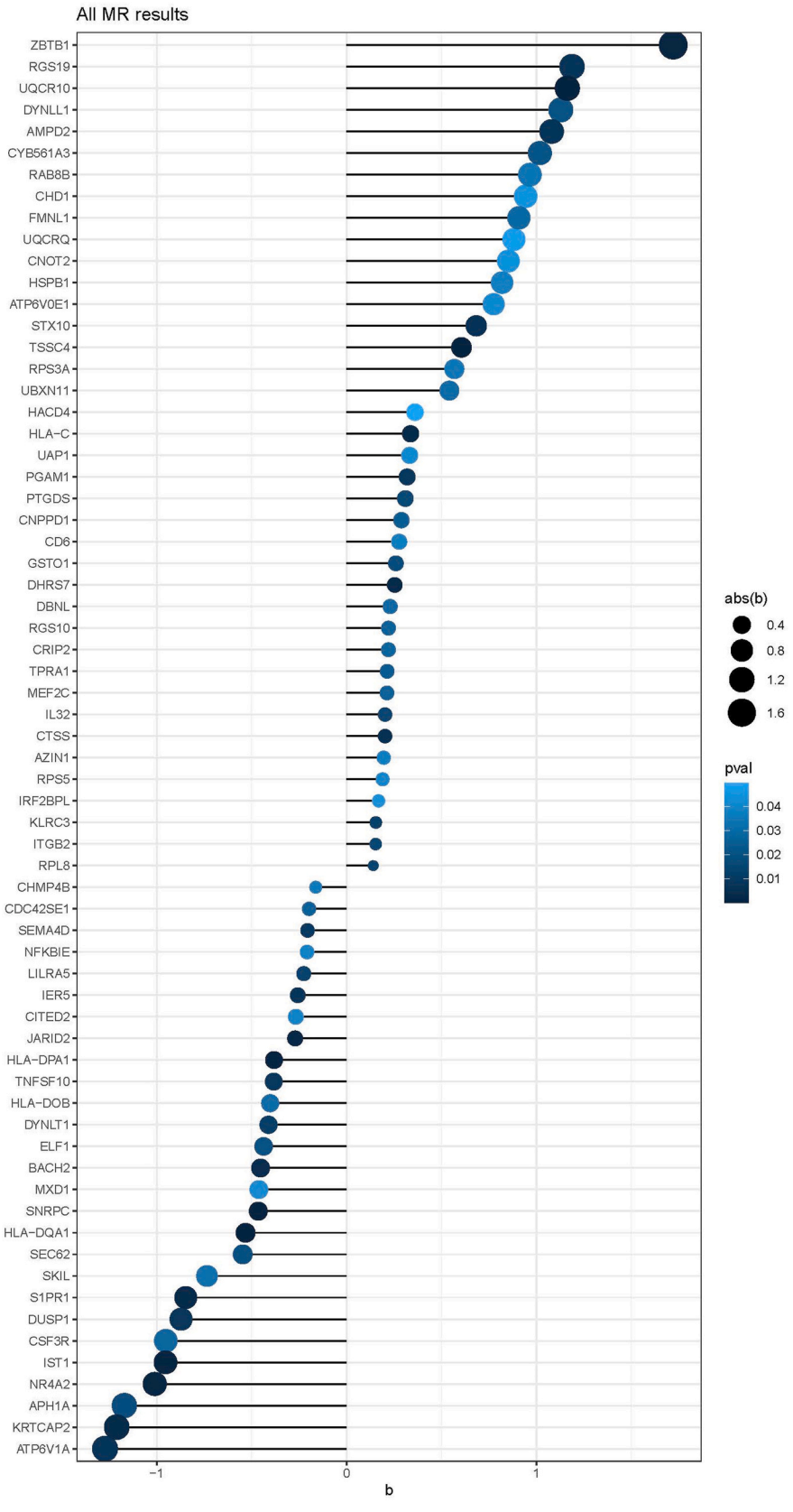
The PCA plot, following SCTransform normalization, effectively displays the intrinsic biological differences between control and diseased samples (Fig. 2A). This representation underscores the biological relevance of our findings, ensuring the observed variance is due to biological diversity rather than batch effects. The analysis of single-cell profiles revealed distinct cell clusters in the UMAP visualization (Fig. 2B). To identify specific gene expression patterns within each cell subset, we performed an analysis of marker genes (Supplementary Table 2). In our study, we identified distinct cell clusters corresponding to specific cell types, including myeloid cells, pDC, mDC, NK cells, T cells, B cells, and plasma cells. The delineation of these clusters was facilitated by the utilization of specific markers such as MS4A1 and CD79A for B cells, CD1C and FCER1A for mDCs, FCN1 and CD14 for myeloid cells, KLRB1 and NKG7 for NK cells, CLEC4C and IL3RA for pDCs, IGHG1 for plasma cells, and IL7R and CD3E for T cells (Fig. 2C). We identified signature genes for each cell subset, highlighting the top five upregulated and downregulated genes, which were visualized via volcano plots (Fig. 2D). This investigation enhanced our understanding of cellular heterogeneity and the associated molecular signatures in pSS.

3.3. Unraveling genetic regulatory relationships in pSS: insights from single-cell screening and MR analysis

In our study, we performed single-cell screening and identified seven distinct cell types, resulting in a total of 3256 differentially expressed genes. After eliminating duplicates, we obtained a final set of 1625 unique genes. To investigate the regulatory relationships between these genes and pSS, we leveraged the IEU open GWAS database to retrieve the corresponding eQTL data. Applying MR analysis, we assessed the causal effects of genetic variations on gene expression levels in the context of pSS. Through this analysis, we identified a total of 39 significant positive eQTLs and 27 significant negative eQTLs associated with the expression of these genes (Fig. 3, Supplementary Table 3). These findings shed light on the potential mechanisms underlying the genetic correlation between pSS and gene expression, offering valuable insights into the molecular basis of the disease.

3.4. Validation and MR analysis of selected pQTLs in relation to pSS

To validate the 66 significant blood-related eQTLs identified previously, we employed plasma protein data for validation. However, out of the 66 eQTLs, 38 lacked corresponding plasma protein associations. Hence, we selected 28 eQTLs with available plasma protein data for further validation. Using a threshold of p-value $<5 \times 10^{-8}$, we screened for SNPs that showed significant associations with the selected plasma proteins. Among the 28 eQTLs, 19 plasma proteins did not have any associated SNPs meeting the significance threshold. Consequently, we proceeded with validation using the remaining 9 plasma proteins that exhibited relevant SNP



(caption on next page)

Fig. 3. The lollipop plot displays the genetic association between primary pSS and 66 expression quantitative trait loci (eQTLs) with a significance level of $p < 0.05$. The beta coefficient (b) in the plot represents the magnitude and direction of the association, providing insights into the strength of the genetic correlation. ($p < 0.05$). pSS: Primary Sjögren's syndrome, MR: Mendelian randomization.

associations. Supplementary Table 4 provides a comprehensive overview of each pQTL's associated SNPs, including their respective F-values. Notably, all F-values surpassed 10, indicating the exclusion of weak instrumental variables.

Using the filtered set of 9 pQTLs as exposure variables and pSS as the outcome variable, we conducted MR analysis. The inverse variance weighted (IVW) method was employed to calculate the odds ratios (ORs) with 95 % confidence intervals (CIs) for each pQTL. According to the IVW method, the dominant effect OR for CTSS was 1.41 (95 % CI: 1.23–1.63, $p < 0.001$). For GSTO1, the dominant effect OR was 0.94 (95 % CI: 0.88–1.00, $p = 0.04$). Additionally, the dominant effect OR for HSPB1 was 1.14 (95 % CI: 1.02–1.27, $p = 0.02$). As for TSSC4, which had only one SNP identified, MR analysis was performed using the Wald ratio. The dominant effect OR for TSSC4 was 0.22 (95 % CI: 0.07–0.74, $p = 0.01$). The remaining 5 pQTLs yielded negative results (Fig. 4, Supplementary Table 5). Sensitivity analysis demonstrated no evidence of heterogeneity or horizontal pleiotropy between the four pQTLs with positive results and pSS. Cochran's Q statistic and the intercept from MR Egger regression indicated no statistically significant evidence of horizontal pleiotropy between CTSS and pSS ($p > 0.05$) (Supplementary Tables 6 and 7).

Plasma protein	nSNP	OR (95%CI)	p value	Heterogeneity
CRIP2				
Wald ratio	1	0.57(0.01-45.40)	0.80	-
CTSS				
MR Egger	13	1.47(1.14-1.91)	0.01	0.18
Weighted median	13	1.43(1.22-1.69)	<0.001	-
Inverse variance weighted	13	1.41(1.23-1.63)	<0.001	0.23
Simple mode	13	1.52(1.19-1.95)	0.01	-
Weighted mode	13	1.41(1.21-1.65)	<0.001	-
DBNL				
Wald ratio	1	1.10(0.47-2.58)	0.82	-
GSTO1				
MR Egger	45	0.87(0.77-0.97)	0.02	0.18
Weighted median	45	0.93(0.86-1.02)	0.11	-
Inverse variance weighted	45	0.94(0.88-1.00)	0.04	0.14
Simple mode	45	0.88(0.74-1.03)	0.13	-
Weighted mode	45	0.91(0.85-0.98)	0.02	-
HSPB1				
MR Egger	24	1.14(0.90-1.43)	0.28	0.14
Weighted median	24	1.13(0.96-1.33)	0.14	-
Inverse variance weighted	24	1.14(1.02-1.27)	0.02	0.18
Simple mode	24	1.28(0.96-1.70)	0.11	-
Weighted mode	24	1.08(0.89-1.31)	0.47	-
LILRA5				
MR Egger	37	0.87(0.73-1.04)	0.13	0.69
Weighted median	37	0.91(0.78-1.06)	0.23	-
Inverse variance weighted	37	0.96(0.86-1.06)	0.39	0.66
Simple mode	37	0.99(0.71-1.39)	0.96	-
Weighted mode	37	0.87(0.73-1.04)	0.13	-
SEMA4D				
MR Egger	37	0.96(0.87-1.06)	0.45	0.30
Weighted median	37	0.97(0.88-1.07)	0.58	-
Inverse variance weighted	37	0.99(0.92-1.06)	0.68	0.32
Simple mode	37	1.00(0.87-1.14)	0.97	-
Weighted mode	37	0.95(0.87-1.04)	0.27	-
TSSC4				
Wald ratio	1	0.22(0.07-0.74)	0.01	-
UAP1				
Wald ratio	1	0.60(0.21-1.69)	0.33	-

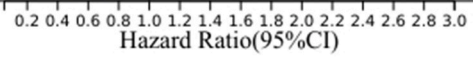


Fig. 4. MR Analysis of plasma protein quantitative trait loci (pQTLs) in Relation to pSS. The MR analysis investigates the potential causal relationships between selected pQTLs and pSS. SNP: single nucleotide polymorphism, OR: odds ratio, CI: confidence interval, pSS: Primary Sjögren's syndrome, MR: Mendelian randomization.

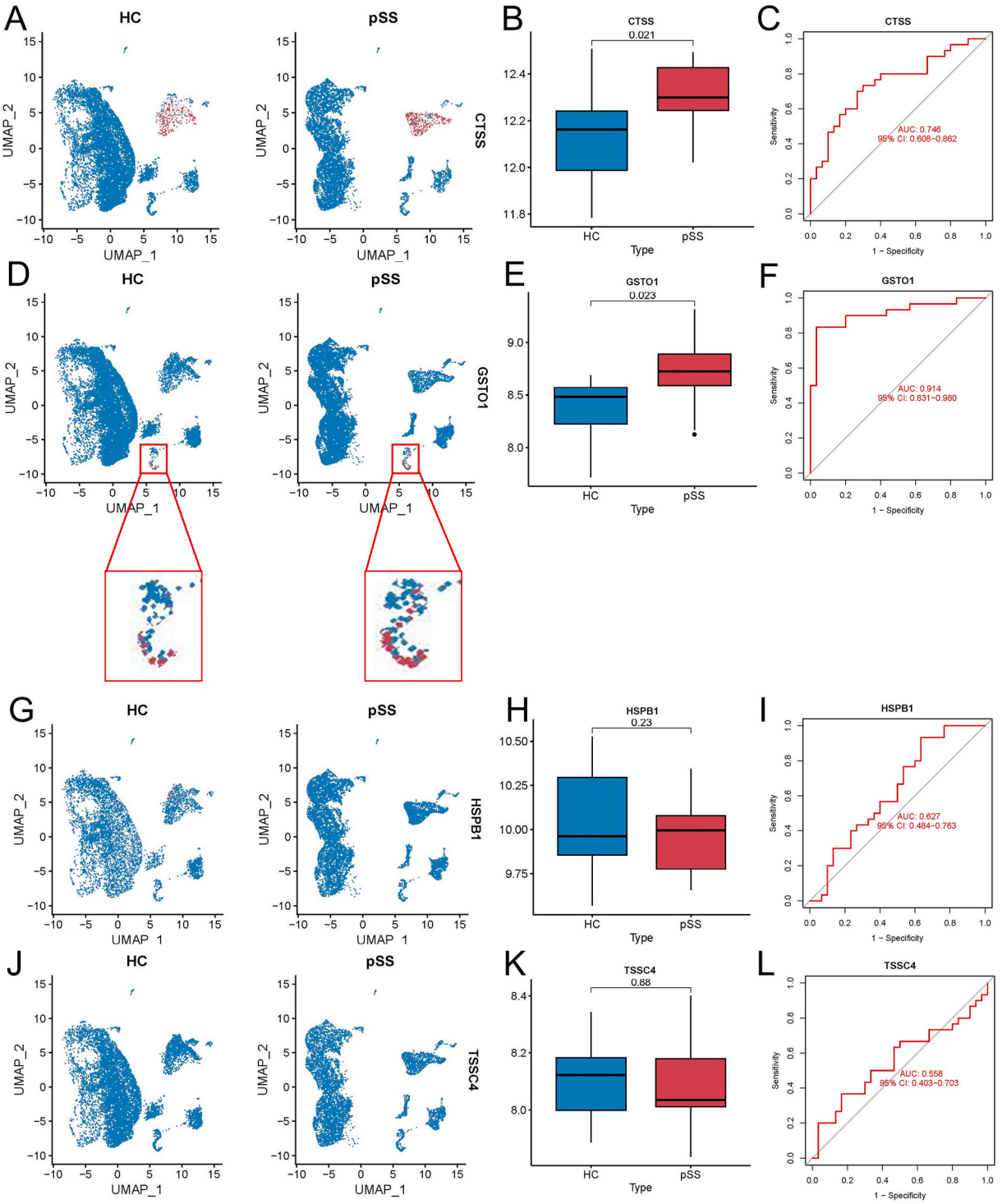


Fig. 5. Validation of CTSS and GSTO1 Expression in pSS. (A) CTSS Expression in the pSS of scRNA-seq. (B)Transcriptomic Validation of CTSS Expression in pSS. (C) Diagnostic Utility of CTSS in pSS. (D) GSTO1 Expression in the pSS of scRNA-seq. (E) Transcriptomic Validation of GSTO1 Expression in pSS. (F) Diagnostic Utility of GSTO1 in pSS. (G–I) HSPB1 Expression in pSS (No Significant Differences). (J–L) TSSC4 Expression in pSS (No Significant Differences). HC: Healthy Control, pSS: Primary Sjögren's syndrome. UMAP: uniform manifold approximation and projection.

The MR analysis findings provide insights into the potential causal relationships between the selected pQTLs and pSS. The positive associations observed for CTSS, GSTO1, HSPB1, and the inverse association for TSSC4 suggest their potential involvement in the pathogenesis of pSS. However, further investigations are required to validate and elucidate the underlying mechanisms of these associations.

3.5. Validation of CTSS and GSTO1 overexpression in pSS and diagnostic utility

To further validate the results obtained from MR analysis, we examined the expression levels of CTSS and GSTO1 using scRNA-seq and transcriptomic data (GSE84844 and GSE48378). Consistent with the previous findings, scRNA-seq analysis showed a significant upregulation of CTSS expression in the pSS group compared to the healthy control group, particularly in bone marrow cells (Fig. 5A). This result was further confirmed in the transcriptomic validation analysis, where CTSS expression was significantly enhanced in the pSS group compared to controls (Fig. 5B). Furthermore, we investigated the diagnostic utility of CTSS in distinguishing between healthy controls and pSS patients. The analysis of the ROC curve revealed that CTSS had a moderate ability to differentiate between the two groups, with an AUC value of 0.746 (Fig. 5C). These findings support the potential diagnostic value of CTSS in identifying pSS patients.

Similarly, scRNA-seq analysis revealed a significant upregulation of GSTO1 expression in the pSS group, with a notable enrichment in a specific subset of T cells, as depicted in the magnified section of Fig. 5D. The transcriptomic validation using the GSE84844 dataset confirmed the increased expression of GSTO1 in the pSS group (Fig. 5E). Moreover, we assessed the diagnostic utility of GSTO1 in distinguishing between NC and pSS patients. The ROC curve analysis revealed excellent diagnostic performance for GSTO1, with an AUC value of 0.914 (Fig. 5F). These findings suggest that GSTO1 may serve as a promising diagnostic biomarker for pSS.

On the other hand, the results for HSPB1 (Fig. 5G–I) and TSSC4 (Fig. 5J-L) showed no significant differences in expression between the pSS and control groups. These findings suggest that HSPB1 and TSSC4 may not be strongly associated with pSS pathology.

4. Discussion

4.1. Integration of single-cell transcriptomics, genomics, and proteomics in Sjögren's syndrome

This study represents the first comprehensive multi-omics investigation combining single-cell transcriptomics, genomics, and proteomics in Sjögren's syndrome, specifically focusing on primary Sjögren's syndrome (pSS). By integrating scRNA-seq data, GWAS data, and plasma protein data, we aimed to gain a deeper understanding of the molecular mechanisms underlying pSS. This integrative approach allowed us to explore cellular heterogeneity, identify key genes associated with pSS, investigate genetic regulatory relationships, and validate potential biomarkers for diagnosis.

4.2. CTSS and its implications in pSS pathogenesis

One of the key findings of our study is the potential involvement of CTSS in the pathogenesis of pSS. CTSS is a lysosomal cysteine protease that plays a role in antigen processing and presentation, as well as tissue remodeling and inflammation [24]. Our MR analysis identified a significant positive association between CTSS expression and pSS, suggesting that CTSS may contribute to the development and progression of the disease. This finding is consistent with previous studies that have implicated CTSS in autoimmune and inflammatory conditions [9,10,13]. We have identified CTSS expression in bone marrow cells as a potential contributor to pSS pathogenesis. Given the bone marrow's role in immune cell production, the overexpression of CTSS could reflect heightened antigen processing activities that promote autoimmune responses [25]. Our findings indicate that the dysregulation of CTSS in bone marrow cells may be a critical step in the cascade of events leading to the manifestation of pSS, offering a novel perspective on the disease mechanism. Our diagnostic utility analysis revealed that CTSS expression levels have moderate discriminatory power in distinguishing pSS patients from healthy controls. These findings highlight CTSS as a potential therapeutic target and diagnostic biomarker for pSS. Moreover, the role of CTSS in autoimmune pathogenesis extends beyond pSS. Recent studies have highlighted a significant relationship between CTSS overexpression and the exacerbation of lupus pathogenesis, notably influencing key immune pathways such as TLR7 expression and IFN-α production [26].

In a study on a CTSS inhibitor for primary Sjögren's syndrome, RO5459072 showed limited therapeutic efficacy [12]. Our research reveals that CTSS is mainly expressed in myeloid cells, suggesting that the suboptimal treatment response may be due to selective inhibition in myeloid cells, while other cell types remain unaffected. Inadequate dosage or duration might have hindered optimal CTSS inhibition in relevant cells. Considering the complex nature of pSS and involvement of multiple immune pathways, combination therapies targeting diverse pathways may enhance treatment outcomes. Further investigation is needed to understand CTSS's role in different cell types and develop more effective strategies for pSS.

4.3. GSTO1 as a promising diagnostic biomarker for pSS

Another noteworthy finding from our study is the upregulation of glutathione S-transferase omega 1 (GSTO1) expression in pSS patients. This observation raises intriguing questions about the role of GSTO1 in the pathogenesis of pSS. However, our MR analysis revealed inconsistent results between pQTL and eQTL analyses, suggesting a discrepancy between the protective effect observed in pQTL and the potential disease-promoting effect indicated by eQTL. The discordance may be attributed to the differences in sample

types used for eQTL (primarily whole blood) and pQTL (mainly plasma proteins) analyses. Tissue-specific gene regulation and post-translational modifications can influence protein levels differently than mRNA expression, leading to distinct findings. The specific reasons for this discrepancy require further investigation. Nevertheless, the significant upregulation of GSTO1 expression in T cells of pSS patients highlights its involvement in the immune dysregulation observed in pSS. Notably, our diagnostic utility analysis demonstrated the excellent discriminatory power of GSTO1 expression levels in distinguishing pSS patients from healthy controls, suggesting its potential as a diagnostic biomarker for pSS. Future investigations are warranted to elucidate the underlying mechanisms by which GSTO1 contributes to pSS pathogenesis and to evaluate its therapeutic potential. In addition, the significance of GSTO1 extends to other autoimmune conditions. It has been identified as one of the common marker genes associated with SLE, demonstrating its involvement in the complex interplay of immune responses across various sample types [27].

4.4. Other identified genes and pathways

In addition to CTSS and GSTO1, our multi-omics analysis identified other genes with pSS. The analysis of single-cell profiles revealed distinct cell clusters representing various immune cell subsets, including myeloid cells, pDC, mDC, NK cells, T cells, B cells, and plasma cells. The identification of signature genes for each cell subset enhances our understanding of the cellular heterogeneity and molecular signatures underlying pSS. Furthermore, our MR analysis identified multiple significant eQTLs associated with the expression of genes implicated in pSS, providing insights into the genetic regulatory relationships involved in the disease. These findings contribute to the growing body of knowledge on the genetic basis of pSS and offer potential targets for future investigations.

4.5. Limitations and future directions

While our study provides significant insights into the molecular landscape of pSS, it is important to acknowledge certain limitations. One such limitation is the resolution of our scRNA-seq, which did not allow for the reliable discrimination of immune cell subtypes with immunosuppressive qualities, such as Tregs. This is partly due to the lack of highly specific markers that can confidently identify such subtypes within our dataset, as well as technical limitations in sequencing depth and cell capture rate. Furthermore, while our study offers a broad overview of gene expression patterns across cell types, the constraints of the available genetic instrumental variables limited our ability to explore the full spectrum of plasma proteins implicated in pSS. The existing literature suggests intricate dysregulation of B cells in pSS [28,29], which our dataset may not have captured in its entirety. Additionally, our study would benefit from functional validation of the identified genes and pathways to elucidate their roles in pSS pathogenesis more concretely. Future research should aim to incorporate broader datasets, including those currently under access restrictions, and undertake experimental studies to validate the functional implications of our findings.

In conclusion, our study represents a significant advancement in understanding the molecular mechanisms underlying primary Sjögren's syndrome (pSS) through the integration of single-cell transcriptomics, genomics, and proteomics. Our findings shed light on the cellular heterogeneity, genetic regulatory relationships, and potential biomarkers associated with pSS. However, further validation and functional studies are necessary, and the integration of other omics data is warranted to unravel the complexities of pSS pathogenesis fully. Ultimately, these insights may lead to the development of targeted therapies and improved management strategies for individuals living with pSS.

Funding

This study was supported by the National Natural Science Foundation of China (82071006 and 82371104), the Natural Science Foundation of Hunan Province (2023JJ30851 and 2024JJ6609) and the Postdoctoral Fellowship Program of CPSF (GZC20233180).

Data availability statement

The data that support the findings of this study are available from the GEO repository, with accession numbers GSE157278, GSE84844, and GSE48378. Additional summary-level data from the GWAS of Sjögren's syndrome used in this study can be accessed from the FinnGen consortium with the dataset ID M13_SJOGREN. The eQTL data from the eQTLGen Consortium and pQTL data from Ferkingstad et al. are also publicly available.

CRediT authorship contribution statement

Yao Tan: Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Jiayang Yin:** Resources, Data curation. **Zhenkai Wu:** Supervision, Investigation. **Wei Xiong:** Writing – review & editing, Resources, Project administration.

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.

Acknowledgments

The authors would like to acknowledge the participants and investigators of the FinnGen study, as well as the valuable data provided by the eQTLGen consortium, GEO database, deCODE genetics project, and MRC IEU OpenGWAS database.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33433.

References

[1] X. Mariette, L.A. Criswell, Primary Sjögren's syndrome, N. Engl. J. Med. 378 (10) (2018) 931–939.
[2] R. Patel, A. Shahane, The epidemiology of Sjögren's syndrome, Clin. Epidemiol. 6 (2014) 247–255.
[3] C.P. Mavragani, H.M. Moutsopoulos, The geoepidemiology of Sjögren's syndrome, Autoimmun. Rev. 9 (5) (2010) A305–A310.
[4] G.E. Thorlacius, A. Björk, M. Wahren-Herlenius, Genetics and epigenetics of primary Sjögren syndrome: implications for future therapies, Nat. Rev. Rheumatol. 19 (5) (2023) 288–306.
[5] X. Hong, S. Meng, D. Tang, T. Wang, L. Ding, H. Yu, H. Li, D. Liu, Y. Dai, M. Yang, Single-cell RNA sequencing reveals the expansion of cytotoxic CD4+ T lymphocytes and a landscape of immune cells in primary Sjögren's syndrome, Front. Immunol. (2021) 3688.
[6] T. Stuart, R. Satija, Integrative single-cell analysis, Nat. Rev. Genet. 20 (5) (2019) 257–272.
[7] J. Liu, H. Gao, C. Li, F. Zhu, M. Wang, Y. Xu, B. Wu, Expression and regulatory characteristics of peripheral blood immune cells in primary Sjögren's syndrome patients using single-cell transcriptomic, iScience 25 (12) (2022) 105509.
[8] C.A. Emdin, A.V. Khera, S. Kathiresan, Mendelian randomization, JAMA 318 (19) (2017) 1925–1926.
[9] L. Xie, S. Zhang, L. Huang, Z. Peng, H. Lu, Q. He, R. Chen, L. Hu, B. Wang, B. Sun, et al., Single-cell RNA sequencing of peripheral blood reveals that monocytes with high cathepsin S expression aggravate cerebral ischemia-reperfusion injury, Brain Behav. Immun. 107 (2023) 330–344.
[10] T. Zuo, Q. Xie, J. Liu, J. Yang, J. Shi, D. Kong, Y. Wang, Z. Zhang, H. Gao, D.B. Zeng, et al., Macrophage-derived cathepsin S remodels the extracellular matrix to promote liver fibrogenesis, Gastroenterology (2023).
[11] S.J. Kim, S. Schätzle, S.S. Ahmed, W. Haap, S.H. Jang, P.K. Gregersen, G. Georgiou, B. Diamond, Increased cathepsin S in Prdm1(-/-) dendritic cells alters the T (FH) cell repertoire and contributes to lupus, Nat. Immunol. 18 (9) (2017) 1016–1024.
[12] D. Bentley, B.A. Fisher, F. Barone, F.A. Kolb, G. Attley, A randomized, double-blind, placebo-controlled, parallel group study on the effects of a cathepsin S inhibitor in primary Sjögren's syndrome, Rheumatology (2023).
[13] S.F. Hamm-Alvarez, S.R. Janga, M.C. Edman, S. Madrigal, M. Shah, S.E. Frousiakis, K. Renduchintala, J. Zhu, S. Bricel, K. Silka, et al., Tear cathepsin S as a candidate biomarker for Sjögren's syndrome, Arthritis Rheumatol. 66 (7) (2014) 1872–1881.
[14] S. Li, L. Wang, Z. Xu, Y. Huang, R. Xue, T. Yue, L. Xu, F. Gong, S. Bai, Q. Wu, et al., ASC deglutathionylation is a checkpoint for NLRP3 inflammasome activation, J. Exp. Med. 218 (9) (2021).
[15] K. Ramkumar, S. Samanta, A. Kyani, S. Yang, S. Tamura, E. Ziemke, J.A. Stuckey, S. Li, K. Chinnaswamy, H. Otake, et al., Mechanistic evaluation and transcriptional signature of a glutathione S-transferase omega 1 inhibitor, Nat. Commun. 7 (2016) 13084.
[16] D. Menon, P.G. Board, A role for glutathione transferase Omega 1 (GSTO1-1) in the glutathionylation cycle, J. Biol. Chem. 288 (36) (2013) 25769–25779.
[17] M.I. Kurki, J. Karjalainen, P. Palta, T.P. Sipila, K. Kristiansson, K.M. Donner, M.P. Reeve, H. Laivuori, M. Aavikko, M.A. Kaunisto, et al., FinnGen provides genetic insights from a well-phenotyped isolated population, Nature 613 (7944) (2023) 508–518.
[18] S. Tasaki, K. Suzuki, A. Nishikawa, Y. Kassai, M. Takiguchi, R. Kurisu, Y. Okuzono, T. Miyazaki, M. Takeshita, K. Yoshimoto, et al., Multiomic disease signatures converge to cytotoxic CD8 T cells in primary Sjögren's syndrome, Ann. Rheum. Dis. 76 (8) (2017) 1458–1466.
[19] M. Sjöstrand, A. Ambrosi, S. Brauner, J. Sullivan, S. Malin, V.K. Kuchroo, A. Espinosa, M. Wahren-Herlenius, Expression of the immune regulator tripartite-motif 21 is controlled by IFN regulatory factors, J. Immunol. 191 (7) (2013) 3753–3763.
[20] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data, Cell 177 (7) (2019) 1888–1902. e1821.
[21] C. Hu, T. Li, Y. Xu, X. Zhang, F. Li, J. Bai, J. Chen, W. Jiang, K. Yang, Q. Ou, CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data, Nucleic Acids Res. 51 (D1) (2023) D870–D876.
[22] U. Vosa, A. Claringbould, H.J. Westra, M.J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Yazar, et al., Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression, Nat. Genet. 53 (9) (2021) 1300–1310.
[23] E. Ferkingstad, P. Sulem, B.A. Atlason, G. Sveinbjornsson, M.I. Magnusson, E.L. Styrmisdottir, K. Gunnarsdottir, A. Helgason, A. Oddsson, B.V. Halldorsson, Large-scale integration of the plasma proteome with genetics and disease, Nat. Genet. 53 (12) (2021) 1712–1721.
[24] P. Smyth, J. Sasiwachirangkul, R. Williams, C.J. Scott, Cathepsin S (CTSS) activity in health and disease-A treasure trove of untapped clinical potential, Mol. Aspect. Med. 88 (2022) 101106.
[25] X.W. Cheng, M. Narisawa, H. Wang, L. Piao, Overview of multifunctional cysteinyl cathepsins in atherosclerosis-based cardiovascular disease: from insights into molecular functions to clinical implications, Cell Biosci. 13 (1) (2023) 1–19.
[26] J. Lee, S. Jang, M. Choi, M. Kang, S.G. Lim, S.Y. Kim, S. Jang, J. Ko, E. Kim, J. Yi, et al., Overexpression of cathepsin S exacerbates lupus pathogenesis through upregulation TLR7 and IFN-α in transgenic mice, Sci. Rep. 11 (1) (2021) 16348.
[27] P.F. Bing, W. Xia, L. Wang, Y.H. Zhang, S.F. Lei, F.Y. Deng, Common marker genes identified from various sample types for systemic lupus erythematosus, PLoS One 11 (6) (2016) e0156234.
[28] H.M. Ibrahem, B cell dysregulation in primary Sjögren's syndrome: a review, Japanese Dental Science Review 55 (1) (2019) 139–144.
[29] P. Soret, C. Le Dantec, E. Desvaux, N. Foulquier, B. Chassagnol, S. Hubert, C. Jamin, G. Barturen, G. Desachy, V. Devauchelle-Pensec, et al., A new molecular classification to drive precision treatment strategies in primary Sjögren's syndrome, Nat. Commun. 12 (1) (2021) 3523.