



OPEN

Identification of key interactive genes and their potential biological functions in type 2 diabetes and Sjögren's syndrome

Yibing Wang¹, Huilin Zeng¹, Xiangjie Hu², Guangzhao Zhang¹, Liangliang Zhou¹, Xingyuan Yu¹ & Yuchun Zhang³✉

This study aimed to identify crosstalk genes and explore their potential roles in type 2 diabetes (T2D) and Sjögren's syndrome (SS) using bioinformatic analysis. We analyzed multiple publicly available gene expression datasets and screened 16 crosstalk genes. Consequently, genes that may play significant roles in disease processes were identified. Thereafter, we used gene set variation analysis to assess the differences in gene sets among various samples. LASSO regression analysis was performed to determine the optimal diagnostic genes, and predictive models for T2D and SS were constructed. The classification accuracy of these models was evaluated using receiver operating characteristic curves. Among 16 identified crosstalk genes, 11 showed significant differences in expression. These genes were significantly enriched in biological processes. The predictive model generated from *ALDH6A1* and *IL11RA* demonstrated good classification accuracy for T2D and SS samples. Immune infiltration analysis revealed significant differences in specific immune cell types between the disease and control groups, demonstrating a significant correlation with the identified hub genes, highlighting their potential involvement in T2D and SS pathophysiology. This study revealed the crucial role of specific immune cells in T2D and SS pathology, providing new insights into both conditions and the potential targets for future immunotherapy design.

Keywords Type 2 diabetes, Sjögren's syndrome, Crosstalk genes, Immune cells, Bioinformatics analysis

Type 2 Diabetes (T2D) and Sjögren's Syndrome (SS) are two major chronic diseases that pose significant threats to human health worldwide. T2D is a chronic metabolic disorder characterized by high blood glucose levels, with common symptoms including dry mouth, excessive thirst, and frequent urination¹. It is typically accompanied by severe complications such as cardiovascular disease and kidney damage, resulting in substantial economic burdens for patients and society^{2–4}. SS is an autoimmune disease primarily characterized by impaired function of the salivary and lacrimal glands, primarily manifesting as discomforting symptoms such as dry mouth and dry eyes⁵. Epidemiological studies indicate that both diseases have high incidence and prevalence rates, severely affecting the quality of life of affected individuals.

Although there is symptom overlap between T2D and SS in clinical settings, the common pathophysiological mechanisms underlying both conditions have not yet been clearly elucidated. Recent studies suggest that inflammatory responses⁶ and metabolic imbalances⁷ play significant roles in the development of both diseases, indicating potential cross-talk and mutual influence in their pathogenesis. However, existing literature regarding the molecular associations between T2D and SS, as well as their potential signaling pathways, remains relatively limited; moreover, a systematic investigation of the molecular interaction networks and key regulatory genes between both conditions has not been conducted^{8–14}.

In this context, an in-depth exploration of the molecular associations between T2D and SS holds significant scientific importance. First, it will aid in understanding the shared pathological basis of complex diseases; second, it will provide a theoretical foundation for discovering new diagnostic biomarkers and targeted therapeutic strategies. Therefore, in this study, we used various bioinformatics methods (including weighted

¹Department of Endocrinology, Rheumatology and Immunology, Jingzhou Branch of Huaihua Central Hospital (The Second People's Hospital of Huaihua City), Huaihua, China. ²Department of Otorhinolaryngology, Jingzhou Branch of Huaihua Central Hospital (The Second People's Hospital of Huaihua City), Huaihua, China. ³Hospital Infection Control Center, Jingzhou Branch of Huaihua Central Hospital (The Second People's Hospital of Huaihua City), Huaihua, China. ✉email: zhangyuchun2503@163.com

gene co-expression network analysis [WGCNA]¹⁵), based on multiple publicly available gene expression datasets, to (1) systematically identify and characterize the key interactive genes common to T2D and SS and (2) analyze their biological functions and related signaling pathways. By revealing the molecular interaction mechanisms between these two diseases, we aim to provide new theoretical foundations and molecular targets for personalized diagnosis, treatment, and precise interventions.

Results

Technology roadmap

The technology roadmap is present in Fig. S1.

WGCNA

WGCNA was performed on the genes in the GSE18732 dataset to evaluate the co-expression module of the T2D dataset. First, the scale-free fit index (Fig. 1a) under different soft thresholds was calculated and displayed to render the constructed network more consistent with the scale-free topology. The results showed that at a fitting index of 0.80, the minimum soft threshold was 3, which is the optimal soft threshold for constructing a scale-free network. A co-expression network was constructed based on the optimal soft threshold, and

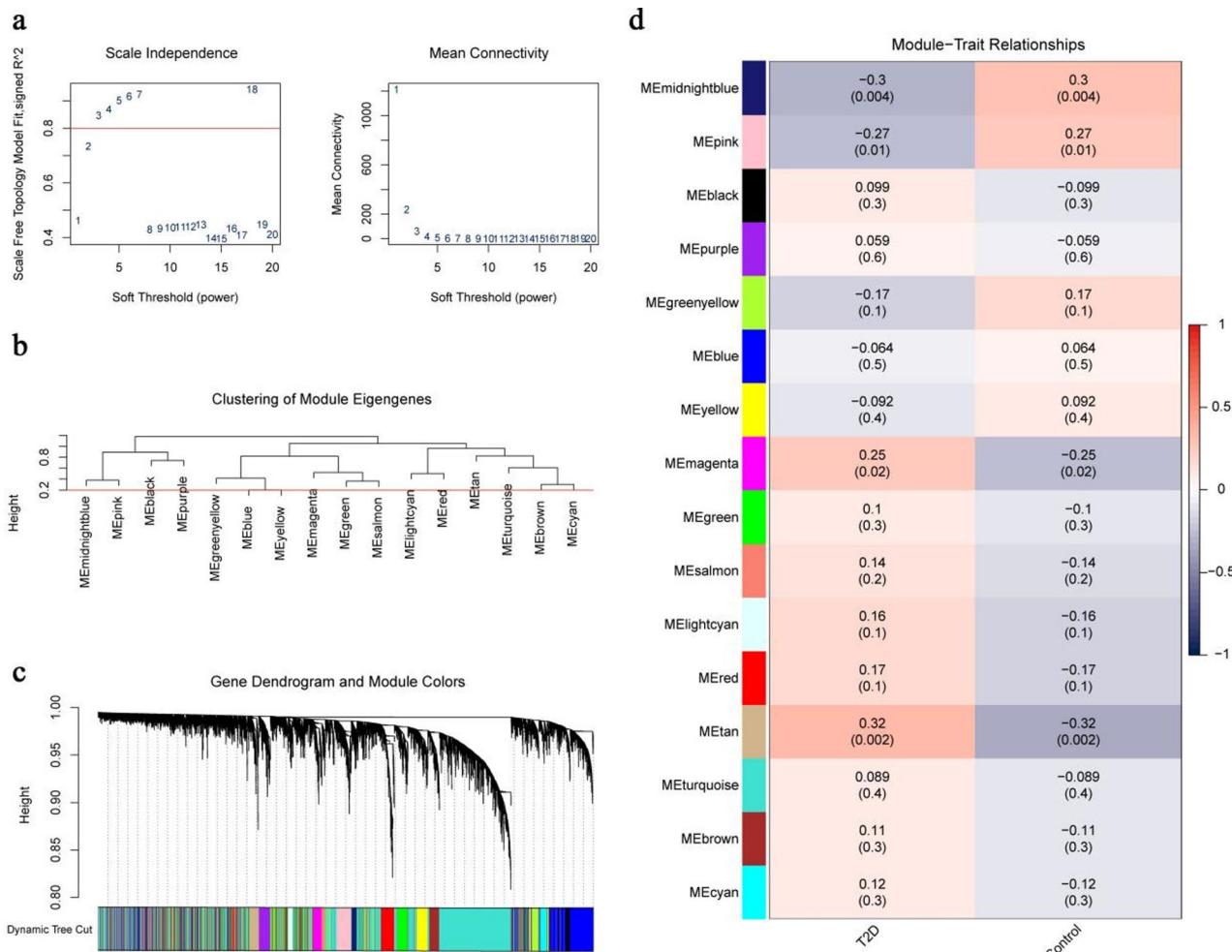


Fig. 1. Weighted gene co-expression network analysis (WGCNA) for type 2 diabetes (T2D) dataset.

(a) Determination of optimal soft-thresholding power. Left panel: Scale-free topology fit index (y-axis, left) and mean connectivity (y-axis, right) plotted against soft-thresholding power (x-axis). Red line indicates the selected soft threshold power ($\beta = 3$). Right panel: Mean network connectivity (y-axis) at different soft-thresholding powers (x-axis). **(b)** Gene cluster dendrogram and module assignment. Genes clustered by topological overlap (y-axis); colored horizontal bars below denote co-expression modules identified by dynamic tree cutting. **(c)** Gene clustering results. Top: Hierarchical clustering dendrogram based on topological overlap dissimilarity (y-axis). Bottom: Color-coded assignment of genes to modules (y-axis). **(d)** Module-trait associations. Correlation analysis between gene module eigengenes (rows) and clinical traits (columns: T2D status). Color scale represents Pearson correlation coefficient (r) values: Red = positive correlation, Purple = negative correlation. Correlation strength: $|r| < 0.3$ (weak/none), $0.3 - 0.5$ (weak), $0.5 - 0.8$ (moderate), > 0.8 (strong).

the genes were clustered and labeled with grouping information using a clustering tree (Fig. 1b). The results showed that using a mergeCutHeight threshold of 0.2, the genes were clustered into the following 16 modules: MEmidnightblue, MEpink, MEblack, MEpurple, MEGreenyellow, MEblue, MEyellow, MEmagenta, MEGreen, MEsalmon, MElightcyan, MERed, METan, MEturquoise, MEBrown, and MECyan. Thereafter, the genes were clustered, and the relationships between the genes and merged modules were visualized (Fig. 1c). Next, the correlation between all genes in the 16 modules and the T2D and control samples was determined according to the expression patterns of module genes (Fig. 1d). Finally, the modules MEmidnightblue and METan with $p < 0.05$, $|r| > 0.30$ were selected.

To evaluate the co-expression modules of the SS dataset, WGCNA was performed on the genes in the SS dataset, GSE40611. First, the scale-free fitting index (Fig. 2a) at different soft thresholds was calculated and displayed to ensure that the constructed network was more consistent with the scale-free topology. The results showed that at a fitting index of 0.80, the minimum soft threshold that conformed to the construction of the scale-free network was 12. Furthermore, a co-expression network was constructed based on the optimal soft threshold, and the genes were clustered and labeled with grouping information using a clustering tree (Fig. 2b). The results showed that at a screening criterion of 0.2, the genes were clustered into the following 16 modules:

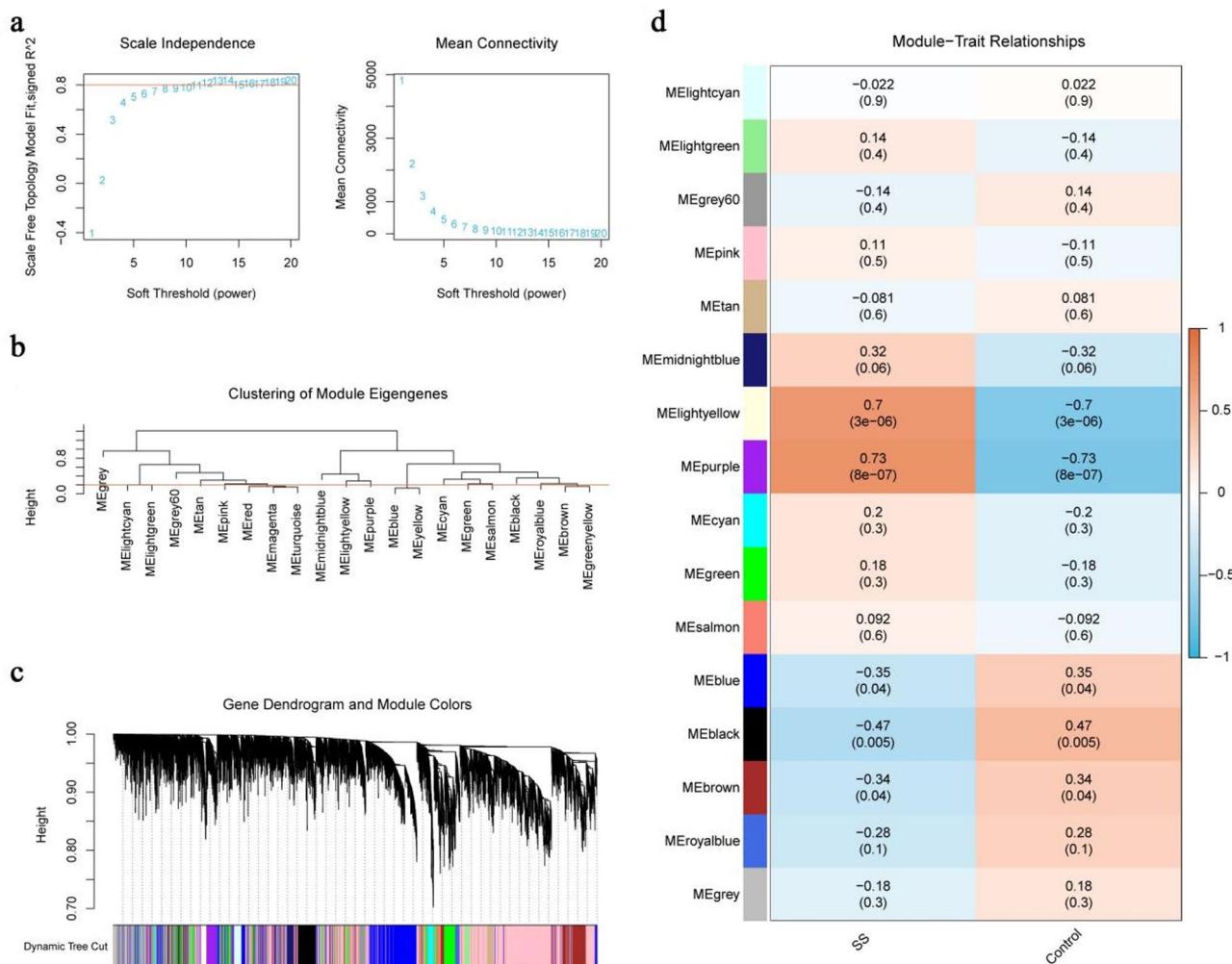
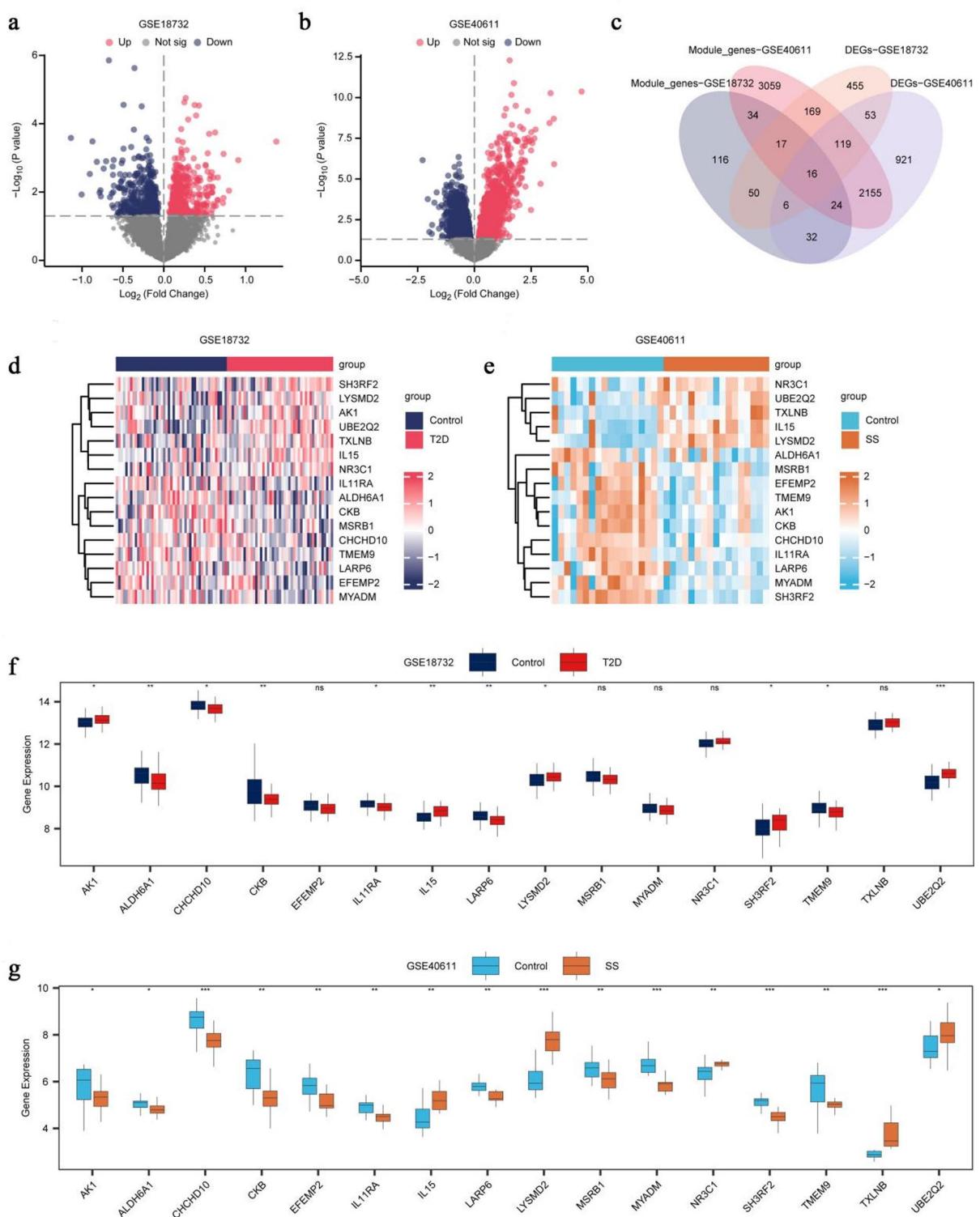


Fig. 2. Weighted gene co-expression network analysis (WGCNA) for Sjogren's syndrome (SS) dataset. (a) Determination of optimal soft-thresholding power. Left panel: Scale-free topology fit index (y-axis, left) and mean connectivity (y-axis, right) plotted against soft-thresholding power (x-axis). Red line indicates the selected soft threshold power ($\beta = 12$). Right panel: Mean network connectivity (y-axis) at different soft-thresholding powers (x-axis). (b) Gene cluster dendrogram and module assignment. Genes clustered by topological overlap (y-axis); colored horizontal bars below denote co-expression modules identified by dynamic tree cutting. (c) Gene clustering results. Top: Hierarchical clustering dendrogram based on topological overlap dissimilarity (y-axis). Bottom: Color-coded assignment of genes to modules (y-axis). (d) Module-trait associations. Correlation analysis between gene module eigengenes (rows) and clinical traits (columns: SS status (Control or SS)). Color scale represents Pearson correlation coefficient (r) values: Red = positive correlation, Brown = negative correlation. Correlation strength: $|r| < 0.3$ (weak/none), $0.3–0.5$ (weak), $0.5–0.8$ (moderate), > 0.8 (strong).



MElightcyan, MElightgreen, MEgrey60, MEpink, MEtan, MEmidnightblue, MElightyellow, MEpurple, MEcyan, MEGreen, MEsalmon, MEblue, MEblack, MEbrown, MEroyalblue, and MEgrey. Thereafter, the genes were clustered and the relationships between the genes and merged modules visualized (Fig. 2c). Next, the correlation between all genes in the 16 modules and the SS and control samples was obtained according to the expression patterns of the module genes (Fig. 2d). Finally, the modules MElightyellow, MEpurple, MEblue, MEblack, and MEbrown with $p < 0.05$ and $|r| > 0.30$ were selected.

Crosstalk genes for T2D and SS

We performed a differential analysis of the T2D dataset GSE18732 and SS dataset GSE40611, and a volcano plot was constructed (Fig. 3a, b). A total of 885 differentially expressed genes (DEGs) between T2D and control samples in the GSE18732 dataset met the $|\log FC| > 0$ and $p < 0.05$ threshold; under this threshold, 451 genes were

Fig. 3. Differential gene expression and crosstalk gene analysis. (a) Volcano plot of T2D versus Control differential expression. Type 2 diabetes (T2D) dataset showing differentially expressed genes (DEGs). X-axis: Log \sim 2-fold change ($\log_2 \sim 2 \times FC$); Y-axis: $-\log_{10} \sim$ (adjusted P value). Colored dots: Red = up-regulated DEGs ($\log_2 \sim 2 \times FC > 0$, adj. $P < 0.05$); Blue = down-regulated DEGs ($\log_2 \sim 2 \times FC < 0$, adj. $P < 0.05$); Gray = non-significant genes. (b) Volcano plot of SS versus Control differential expression. Sjögren's syndrome (SS) dataset DEGs. Axes and color scheme identical to panel a. (c) Venn diagram of DEG-module overlaps. Intersections between DEGs and WGCNA module genes. Sets defined: Mistyrose circle = T2D DEGs; Lavender circle = SS DEGs; Silver circle = T2D module genes; Pink circle = SS module genes. Overlapping regions indicate crosstalk genes shared between datasets and analysis methods. (d) and (e) Crosstalk gene expression heatmaps. Normalized expression z-scores across samples. Columns: Samples (GSE18732 in d, GSE40611 in (e)); Rows: Crosstalk genes; Color scale: Red = high expression, Blue = low expression (scale bar shown). Sample groups labeled: Control (c) versus Disease (T2D/SS). (f) and (g) Crosstalk gene expression group comparisons. Violin plots showing expression distribution between groups. Y-axis: Normalized expression value; X-axis: Sample groups (Control vs. Disease). Central line: Median; Box: Interquartile range; Whiskers: $1.5 \times IQR$. Dataset source: GSE18732 (f), GSE40611 (g).

upregulated ($\log_2 FC > 0$ and $p < 0.05$) and 434 genes downregulated ($\log_2 FC < 0$ and $p < 0.05$). In the GSE40611 SS dataset, 3326 DEGs met the $|\log_2 FC| > 0$ and $p < 0.05$ threshold between the SS and control samples. Under this threshold, 1526 genes were upregulated ($\log_2 FC > 0$ and $p < 0.05$) and 1800 genes downregulated ($\log_2 FC < 0$ and $p < 0.05$).

The intersection of the DEGs and module genes of the two datasets was obtained and Venn diagram was constructed to obtain the crosstalk genes of the two datasets (Fig. 3c). In the present study, “crosstalk genes” are defined as the intersection of DEGs and WGCNA-derived module genes that are associated with both T2D and SS; these are genes that are differentially expressed and located within the key co-expression modules in both diseases. The following 16 crosstalk genes were obtained: *AK1, ALDH6A1, CHCHD10, CKB, EFEMP2, IL11RA, IL15, LARP6, LYSMD2, MSRB1, MYADM, NR3C1, SH3RF2, TMEM9, TXLNB*, and *UBE2Q2*. In addition, heat maps were designed to show crosstalk gene expression in the GSE18732 and GSE40611 datasets (Fig. 3d, e).

Furthermore, we used the Mann–Whitney U test to screen crosstalk genes and construct group comparison plots (Fig. 3f, g), which showed that the following 11 genes were significantly different in the GSE18732 and GSE40611 datasets: *AK1, ALDH6A1, CHCHD10, CKB, IL11RA, IL15, LARP6, LYSMD2, SH3RF2, TMEM9*, and *UBE2Q2*.

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis
The biological processes (BPs), cellular components (CCs), molecular functions (MFs), and biological pathways (KEGG) of the 11 crosstalk genes were further explored using gene ontology GO and KEGG pathway enrichment analyses, and the specific results are shown in Table S2. The 11 crosstalk genes were mainly enriched in BPs such as ATP metabolic process, regulation of protein catabolic processes, cytokine-mediated signaling pathway, and other BPs; CCs such as multivesicular body membranes, integral components of mitochondrial inner membranes, and intrinsic components of mitochondrial inner membranes; MFs such as ubiquitin-like protein transferase activity, ubiquitin-like protein transferase activity, sequence-specific mRNA binding, and other MFs. In addition, it was enriched in biological pathways (KEGG) such as the JAK-STAT signaling pathway, cytokine-cytokine receptor interaction, thiamine metabolism, and other biological pathways. The results of the analyses were visualized using a bubble diagram (Fig. 4a).

A loop network diagram of BP, MF, CC, and biological pathway (KEGG) was constructed based on the results of the GO and KEGG pathway enrichment analyses (Fig. 4b–e). The lines indicate the respective molecules and annotations of the respective entries; a greater node size indicates a higher number of molecules encompassed by the entries.

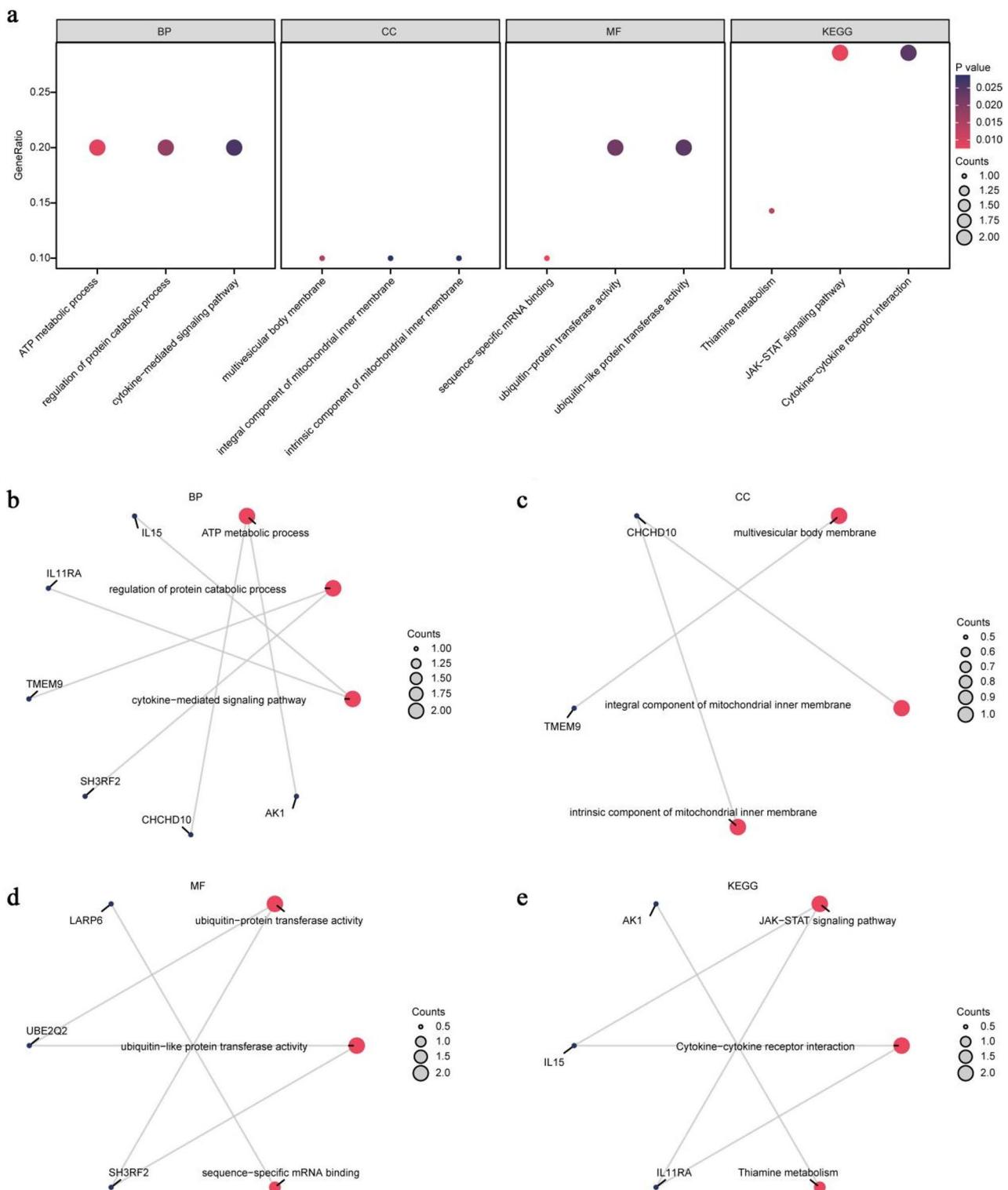
Gene set enrichment analysis (GSEA)

GSEA was used to examine the expression and BPs of all genes in the T2D dataset GSE18732 to determine the effect of their expression levels on T2D. Associations between the affected CCs and their MFs (Fig. S2a) are shown in Table S3. The results showed that all genes in the GSE18732 dataset were significantly enriched in T2D (Fig. S2b), the insulin pathway (Fig. S2c), the MAPK signaling pathway (Fig. S2d), the IL6 pathway (Fig. S2e), the TGF beta signaling pathway (Fig. S2f), T2D (Fig. S2g), and other related biological functions and signaling pathways.

GSEA was used to analyze the expression and BPs of all genes in the GSE40611 dataset to determine the effect of their expression levels on SS. The relationships between the affected CCs and their MFs (Fig. S3a) are shown in Table S4. The results showed that all genes in the GSE40611 dataset were significantly enriched in the negative regulation of the PI3K AKT network (Fig. S3b), the IL18 signaling pathway (Fig. S3c), immune system diseases (Fig. S3d), Notch signaling (Fig. S3e), the inflammatory response pathway (Fig. S3f), the JAK STAT signaling pathway (Fig. S3g), and other related biological functions and signaling pathways.

Gene set variation analysis (GSVA)

To investigate the disparity in the h.all.v7.4. symbols.gmt gene set between the T2D and control samples within the T2D dataset GSE18732, GSVA was performed for all genes in the dataset. Detailed information is provided in Table S5. The five most significant pathways ($p < 0.05$) were screened and visualized using a heat map and bar



graph (Fig. S4a, b). The five pathways included oxidative phosphorylation, KRAS signaling dn, the Hedgehog signaling pathway, Wnt/beta-catenin signaling, and G2M checkpoint.

To investigate the disparity between h.all.v7.4. symbols.gmt gene set in the SS and control samples within the SS dataset GSE40611, GSVA was performed for all genes in the dataset. Detailed information is provided in Table S6. The following five most significant pathways ($p < 0.05$) were screened and visualized using a heat map and bar graph (Fig. S5a, b): interferon alpha response, interferon gamma response, allograft rejection, IL-6/JAK/STAT3 signaling, and KRAS signaling up.

Identification of diagnostic genes

To identify the best diagnostic genes among the crosstalk genes of the two diseases, Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis was performed, and LASSO regression model and variable

Fig. 4. Functional enrichment analysis of crosstalk genes based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). (a) Enrichment bubble plot. Significantly enriched GO terms (Biological Process [BP], Cellular Component [CC], Molecular Function [MF]) and KEGG pathways. X-axis: Gene ratio (Gene counts/Background genes); Y-axis: Enriched term name; Bubble size: Number of associated genes; Bubble color: $-\log_{10}$ (adjusted P value) gradient (red = low P value [high significance], black = high P value [low significance]). Filter criteria: $P < 0.05$, False Discovery Rate (FDR) < 0.25 . (b) BP enrichment network. Circular visualization of Biological Process (BP) terms and associated crosstalk genes. Red nodes: GO-BP terms; Black nodes: Genes; Gray lines: Gene-term associations. (c) CC enrichment network. Cellular Component (CC) term-gene interactions. Node and line schemes identical to panel b. (d) MF enrichment network. Molecular Function (MF) term-gene relationships. Node and line schemes identical to panel b. (e) KEGG pathway network. Kyoto Encyclopedia of Genes and Genomes pathway-gene interactions. Red nodes: KEGG pathways; Black nodes: Genes; Gray lines: Gene-pathway associations.

trajectory maps were constructed for visualization. The results showed four best diagnostic genes (Fig. 5a, b) in the T2D dataset GSE18732 (*ALDH6A1*, *IL11RA*, *IL15* and *LARP6*) and five best diagnostic genes (Fig. 5c, d) in the SS dataset (*AK1*, *ALDH6A1*, *IL11RA*, *LYSMD2*, and *SH3RF2*). The best diagnostic genes of the two datasets were intersected and a Venn diagram (Fig. 5e) was constructed, and the following two diagnostic genes were obtained: *ALDH6A1* and *IL11RA*. The group comparison diagram of the GSE18732 (Fig. 5f) and GSE40611 (Fig. 5g) datasets showed that *ALDH6A1* and *IL11RA* expressions were significantly different between the disease and control groups in both datasets.

Design of the prediction model

Based on two diagnostic genes (*ALDH6A1* and *IL11RA*) and univariate and multivariate binary logistic regression, we constructed predictive models for the T2D and SS datasets. To further validate the predictive model in the T2D dataset, a nomogram demonstrating the relationship between these two genes was constructed (Fig. 6a). The results showed that *ALDH6A1* expression had a significantly higher utility than *IL11RA* expression in the T2D prediction model. The clinical utility of the T2D prediction model was evaluated using decision curve analysis (DCA) (Fig. 6b). The results demonstrated that, within a specific range, the model's line was stable at a higher level than that of all positive and negative values. Moreover, the model had a greater net benefit and superior effect. In addition, the R package pROC was used to construct the receiver operating characteristic (ROC) curve using the RiskScore in the T2D dataset. The ROC curve (Fig. 6c) showed that the RiskScore of the T2D prediction model in the T2D dataset had a certain accuracy (area under the curve [AUC] > 0.7) in the classification of T2D and control samples. The RiskScore was calculated using the following formula:

$$\text{RiskScore} = 23.775 + \text{ALDH6A1} * (-1.066) + \text{IL11RA} * (-1.405)$$

Using the RiskScore formula, we calculated the RiskScore of the T2D dataset GSE29221 and plotted a diagnostic ROC curve (Fig. 6d). Furthermore, the RiskScore of the GSE29221 T2D prediction model had a level of accuracy (AUC > 0.7) in classifying the T2D and control samples.

To further validate the predictive value of GSE40611 in the SS dataset, a nomogram showing the relationship between the two genes was constructed (Fig. 6e). The results showed that *IL11RA* expression level was significantly higher than *ALDH6A1* expression level in the SS prediction model. DCA was used to evaluate and present the clinical utility of the prediction model for SS (Fig. 6f). The outcomes indicated that, within a specific range, the model's line was consistently higher than that of all positive and all negative, with a greater net benefit and superior effect. In addition, the R package pROC was used to construct the ROC curve using RiskScore in the SS dataset. The ROC curve showed (Fig. 6g) that the RiskScore of the SS prediction model had a level of accuracy (AUC > 0.7) in the classification of SS and control samples. The RiskScore was calculated using the following formula:

$$\text{RiskScore} = 25.281 + \text{ALDH6A1} * (-2.673) + \text{IL11RA} * (-2.590)$$

Based on the RiskScore formula for the SS prediction model, we calculated the RiskScore for the SS dataset GSE154926 and plotted a diagnostic ROC curve (Fig. 6h). The results showed that the RiskScore of the GSE154926 SS diagnostic model was accurate in classifying SS and control samples (AUC > 0.7).

Construction of the protein–protein interaction (PPI) network and screening of hub genes

We used the diagnostic genes *ALDH6A1* and *IL11RA* as hub genes and constructed an interaction network between the two hub genes and functionally similar genes using the GeneMANIA website (Fig. S6a). The lines with different colors represent co-expression and share information such as protein domains.

Thereafter, we then obtained miRNAs associated with hub genes using the ENCORI database and Transcription factors (TFs) combined with hub genes through the ChIPBase database. The miRNA-mRNA-TF regulatory network was constructed and visualized using Cytoscape software (Fig. S6b). In the network, there were two hub genes, 12 TFs, and nine miRNAs.

Immune infiltration analysis between disease and control samples

Using the gene expression matrix of T2D samples from the GSE29221 dataset, we quantified the immune infiltration abundance of 28 immune cell subtypes in both T2D and control groups through single-sample GSEA

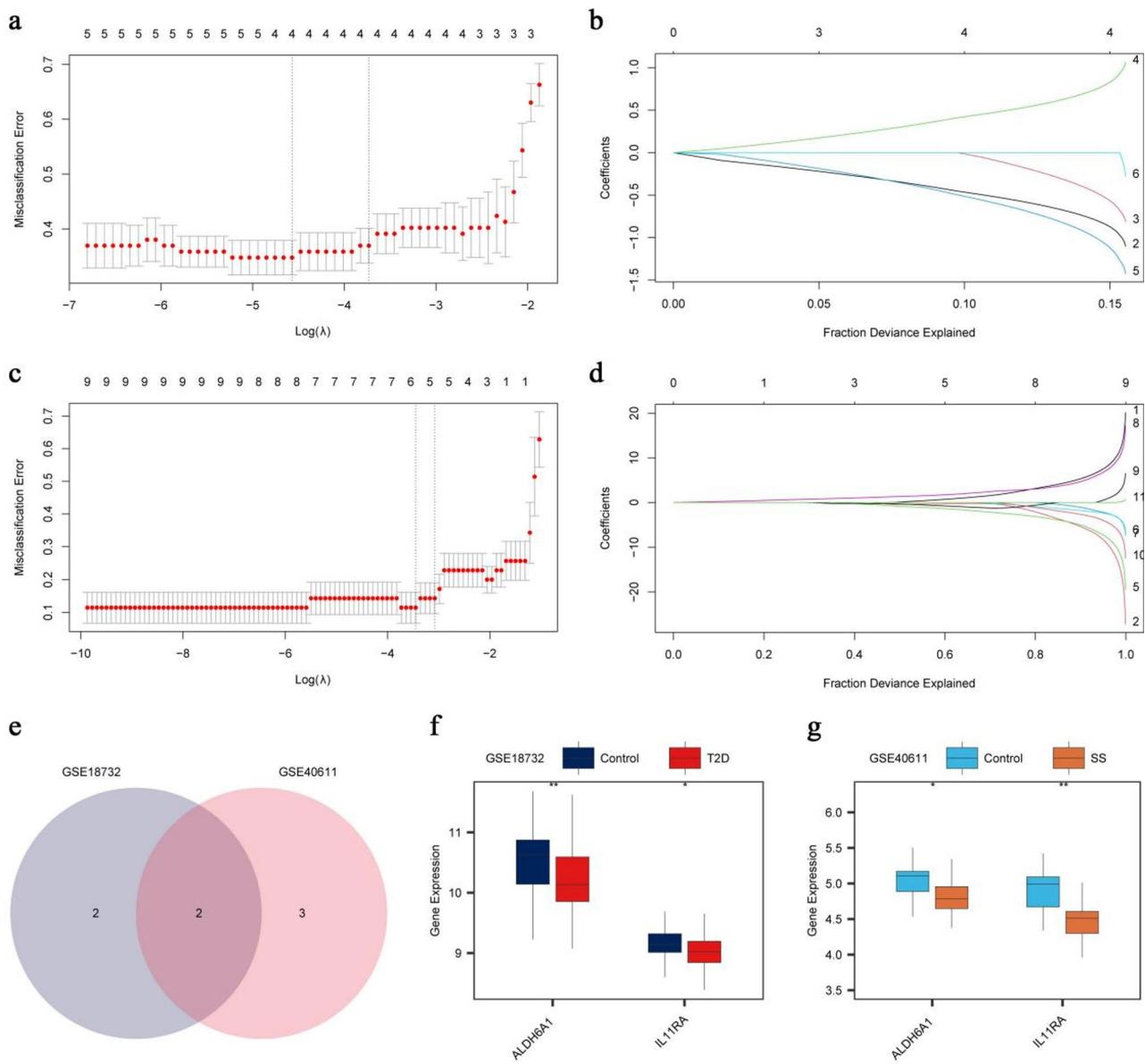


Fig. 5. Diagnostic biomarker screening via LASSO (Least Absolute Shrinkage and Selection Operator) regression. (a) and (b) Diagnostic model (a) and variable trajectory plots (b) of the LASSO regression model for the T2D dataset GSE18732. (c) and (d) Diagnostic model (c) and variable trajectory plots (d) of LASSO regression model for the SS dataset GSE40611. (e) Diagnostic gene intersection. Venn diagram: Shared biomarkers between T2D and SS. Silver circle: T2D diagnostic genes (GSE18732, $n=4$). Pink circle: SS diagnostic genes (GSE40611, $n=5$). Overlap (darkgray): Crosstalk diagnostic genes (*ALDH6A1*, *IL11RA*). (f) and (g) Expression validation. (f) GSE18732 (T2D): Boxplot: *ALDH6A1*/*IL11RA* expression, Y-axis: Normalized expression value, X-axis: Sample groups [Control vs. T2D], Boxplot elements: Center line = median, box = interquartile range (IQR), whiskers = $1.5 \times$ IQR, Black stars: $P < 0.01$ (vs. Control). (g) GSE40611 (SS): Configuration identical to f with SS cohort. Abbreviations: T2D = Type 2 Diabetes; SS = Sjögren's syndrome; IQR = Interquartile range.

(ssGSEA). First, the differences in the expression of infiltrating immune cells among the different groups were presented through a group comparison plot. The group comparison diagram (Fig. 7a) showed three immune cells with statistical significance ($p < 0.05$): macrophages, natural killer T cells, and neutrophils.

A correlation heat map was used to show the correlation results of the infiltration levels of the three immune cells in the T2D dataset (Fig. 7b). The results showed a positive correlation between the infiltration levels of the three immune cells.

Finally, the correlation between hub genes and immune cell infiltration levels was demonstrated using a correlation bubble plot (Fig. 7c). *IL11RA* showed the strongest positive correlation with macrophages ($r = 0.390$).

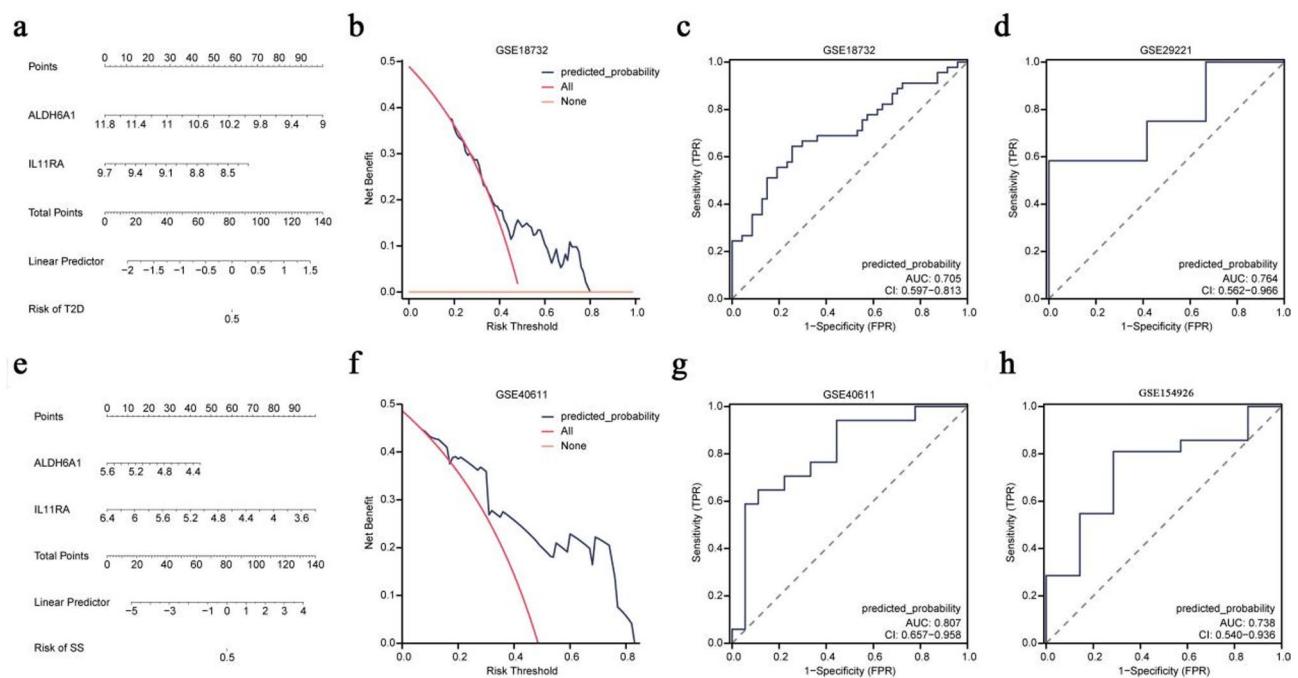


Fig. 6. Construction and validation of diagnostic models for Type 2 Diabetes (T2D) and Sjögren's syndrome (SS). **(a)** T2D diagnostic nomogram (GSE18732). Predicts T2D risk based on two biomarkers. Points scale (top axis): Score for each gene expression level; Total points (middle axis): Summed scores; Risk probability (bottom axis): Predicted T2D probability. Gray vertical lines: Projection path for sample calculation. **(b)** Decision curve analysis (DCA) for T2D model. X-axis: Threshold probability; Y-axis: Net benefit; Darkslateblue curve: Diagnostic model; Mistyrose line: "Treat all" strategy; Darksalmon line: "Treat none" strategy. **(c)** and **(d)** ROC curves for T2D validation: **(c)** T2D dataset (GSE18732) evaluation: X-axis: 1—Specificity; Y-axis: Sensitivity; Blue curve: Diagnostic model (AUC = 0.705, 95% CI 0.597–0.813), **(d)** Independent T2D cohort (GSE29221) validation: Components identical to **(c)** (AUC = 0.764, 95% CI 0.562–0.966), **(e)** SS diagnostic nomogram (GSE40611). Configuration identical to **a** with SS risk prediction. Pink vertical lines: Sample calculation path. **(f)** DCA for SS model. Component scheme identical to **(b)**. **(g)** and **(h)** ROC curves for SS validation: **(g)** SS dataset (GSE40611) evaluation: Red curve: Diagnostic model (AUC = 0.807, 95% CI 0.657–0.958), **(h)** Independent SS cohort (GSE154926) validation: Components identical to **g** (AUC = 0.738, 95% CI 0.540–0.936).

The expression matrix of SS samples from the SS dataset GSE154926 was used to compute the immune infiltration levels of 28 immune cells in SS and control samples using the ssGSEA algorithm. Initially, the expression differences of the infiltration levels of immune cells in distinct groups were presented using a group comparison plot. The group comparison diagram (Fig. 8a) indicated that the following immune cell types showed statistical significance ($p < 0.05$): central memory CD8+ T cells, effector memory CD4+ T cells, and immature dendritic cells.

The correlation results of the three immune cell infiltration levels in the immune infiltration analysis of the SS dataset GSE154926 were presented in a correlation heat map (Fig. 8b). The results showed a positive correlation between the three immune cell infiltration levels.

Finally, the correlation between hub genes and immune cell infiltration levels was determined using a correlation bubble plot (Fig. 8c). The results of the correlation bubble plot showed that *ALDH6A1* had the strongest positive correlation with immature dendritic cells ($r = 0.438$).

Discussion

In recent years, T2D and SS have attracted considerable attention owing to their clinical significance. T2D, a metabolic disorder, presently affects approximately 463 million people globally and the figure is expected to reach 700 million by 2045¹⁶. Its complications, such as cardiovascular disease and kidney failure^{2,3}, place a significant economic burden on global healthcare. Meanwhile, SS primarily impacts women, affecting 0.54.8% of the population. SS causes lacrimal and salivary gland dysfunctions, resulting in chronic dryness and fatigue¹⁷. Recent research has increasingly focused on the interaction between T2D and SS; however, substantial gaps persist in our understanding of the molecular mechanisms between these two chronic diseases.

Therefore, the present study used advanced multivariate statistical techniques, including GSVA and LASSO regression analyses, to conduct a comprehensive bioinformatics analysis of samples from the GSE18732 and GSE40611 datasets, which are related to T2D and SS, respectively. This large-scale systematic exploration enhanced the reliability and generalizability of the study findings. By performing an intersection analysis of the

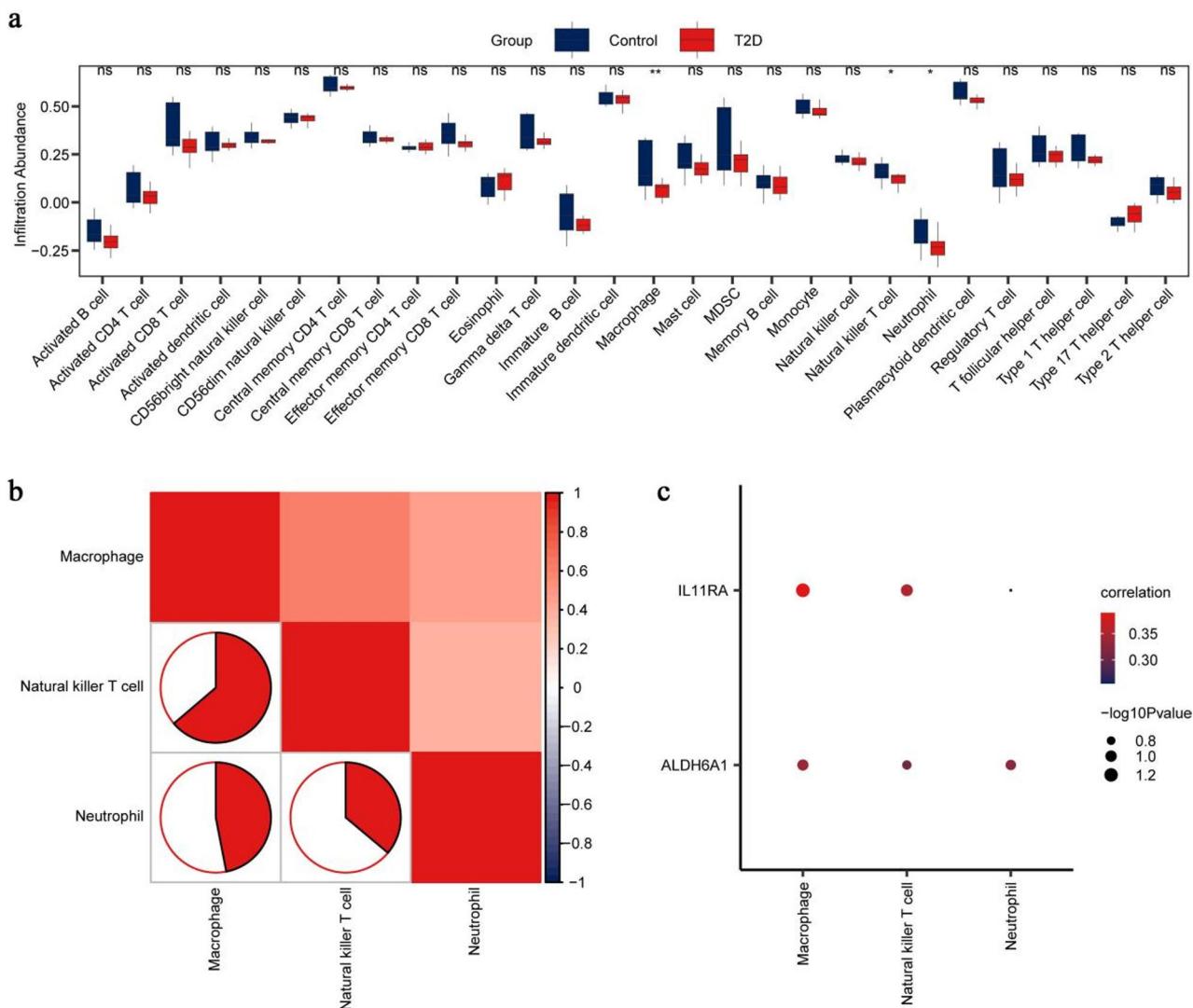


Fig. 7. Immune microenvironment analysis via single-sample Gene Set Enrichment Analysis (ssGSEA) in Type 2 Diabetes (T2D). **(a)** Immune cell infiltration comparison. Differential abundance of 28 immune cell types between T2D ($n = 12$) and Control ($n = 12$) samples in GSE29221. Y-axis: ssGSEA enrichment score; X-axis: Immune cell types; Boxplot colors: Crimson = T2D group, Darkslategray = Control group; Statistical indicators: *** = $P < 0.001$, ** = $P < 0.01$, * = $P < 0.05$, ns = $P \geq 0.05$. **(b)** Immune cell correlation network. Pairwise Spearman correlations between immune cells. Rows/Columns: Immune cell types; Color scale: Crimson = Positive correlation ($r > 0$), Darkslategray = Negative correlation ($r < 0$); Color intensity: Correlation strength ($|r|$). **(c)** Hub gene-immune cell correlations. Bubble plot showing relationships between immune cells (X-axis) and hub genes (Y-axis). Bubble size: $|Correlation coefficient (r)|$; Bubble color: Crimson = Positive correlation, Darkslategray = Negative correlation.

diagnostic genes for both diseases, we identified common candidate biomarkers that offer a novel perspective for cross-disease research.

The primary results of this study indicate that *ALDH6A1* and *IL11RA* are common and important diagnostic genes for T2D and SS. This finding uncovers potential molecular connections between the two disorders and provides a basis for selecting cross-feature biomarkers. *IL11RA*, as the primary receptor for *IL11*, forms a signaling complex with gp130 that activates the JAK-STAT pathway, thereby affecting endocrine and immune responses. In T2D, *IL11* signaling is associated with inflammation in metabolic syndrome; in SS, it affects immune cell functionality and tissue inflammation. Furthermore, *IL11* modulates energy metabolism in adipocytes and myocytes, thereby affecting triglyceride metabolism, blood glucose levels, and insulin sensitivity. *ALDH6A1*, an aldehyde dehydrogenase, plays a critical role in amino acid metabolism and oxidative stress regulation. In addition, it helps remove harmful aldehydes, reducing oxidative stress-related cellular damage^{18,19}. In both T2D and SS, oxidative stress correlates with cellular apoptosis and tissue damage. *ALDH6A1* regulates immune responses in SS, and its dysfunction may exacerbate autoimmune inflammation. Notably, when constructing predictive models, the expression level of *ALDH6A1* was significantly higher than that of *IL11RA*, indicating that

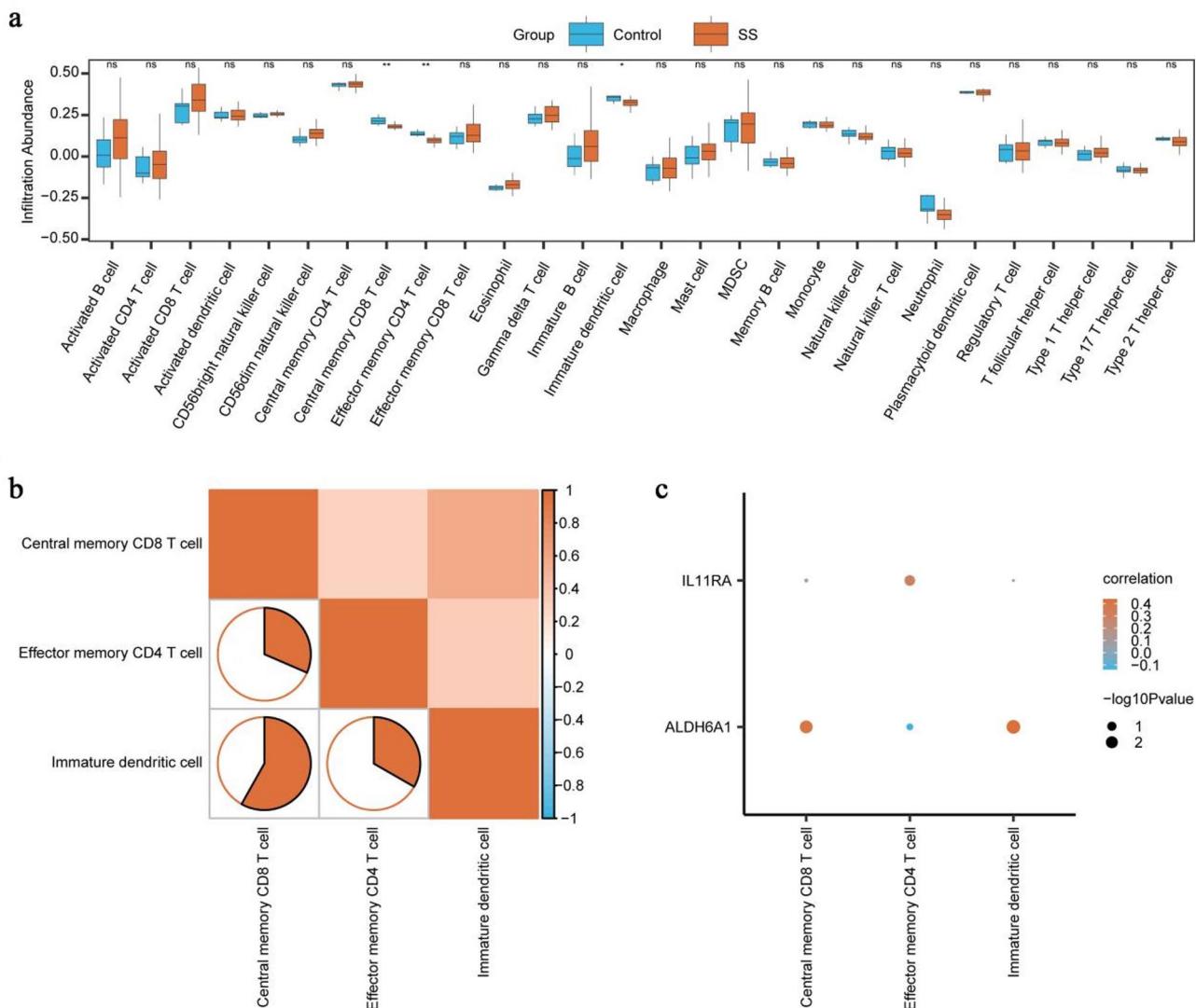


Fig. 8. Immune microenvironment analysis via single-sample Gene Set Enrichment Analysis (ssGSEA) in Sjögren's syndrome (SS). **(a)** Immune cell abundance comparison. Differential infiltration of 24 immune cell types between SS (n=42) and Control (n=7) samples in GSE154926. Y-axis: ssGSEA enrichment score; X-axis: Immune cell types; Boxplot colors: Peru = SS group, Mediumturquoise = Control group; Statistical indicators: *** = $P < 0.001$, ** = $P < 0.01$, * = $P < 0.05$, ns = $P \geq 0.05$. **(b)** Immune cell correlation network. Pairwise Spearman correlations between infiltrating immune cells. Rows/Columns: Immune cell types; Color gradient: Peru = Positive correlation ($r > 0$), Mediumturquoise = Negative correlation ($r < 0$); Color intensity: Absolute correlation strength ($|r|$). **(c)** Hub gene-immune cell relationships. Bubble plot showing interactions between immune cells (X-axis) and hub genes (Y-axis). Bubble size: |Spearman correlation coefficient (r)|; Bubble color: Peru = Positive correlation, Mediumturquoise = Negative correlation.

ALDH6A1 plays a more vital role in T2D pathogenesis. In addition, *IL11RA* expression was strongly correlated with macrophage infiltration, providing new insights into the pathology of SS. The effectiveness of these two genes in classifying T2D and SS was validated using a constructed predictive model and risk score assessment. The AUC value of the ROC curve exceeded 0.7, suggesting that the model had good classification performance. The innovation of this study lies in its comprehensive analysis and data mining, which uncovered the molecular mechanisms and clinical relevance of T2D and SS in multiple dimensions. These results enrich the biological context of these two conditions and provide guidance for future studies. Future studies should validate the functions of *ALDH6A1* and *IL11RA* using animal experiments or clinical samples, thereby laying the foundation for the clinical application of these biomarkers.

In the present study, we conducted a comprehensive analysis of DEGs associated with T2D and SS. We identified 885 DEGs in the T2D dataset, including 451 upregulated and 434 downregulated genes, whereas the SS dataset revealed 3326 DEGs, including 1526 upregulated and 1800 downregulated genes. These significant differences in gene expression highlight the complex molecular underpinnings of both conditions and lay a crucial foundation for exploring their pathological mechanisms. Identifying DEGs is of the utmost importance

for understanding the changes in T2D- and SS-related biological pathways. Notably, many genes upregulated in T2D are associated with insulin signaling and metabolic pathways, such as the PI3K-AKT signaling pathway, which plays an integral role in glucose metabolism and insulin response. Conversely, downregulated genes affect cellular stress response mechanisms, thereby worsening insulin resistance. In SS, the upregulated genes likely reflect enhanced immune responses and inflammation. This study corroborates previous findings that genes upregulated in insulin resistance and inflammation are often related²⁰, supporting the reliability of our results.

Furthermore, our pathway enrichment analysis of the DEGs revealed critical signaling pathways such as the NF-κB pathway, which is essential for immune regulation. The presence of batch effects can introduce biases in differential expression characteristics, which complicate biological interpretation. Therefore, we used normalization techniques and principal component analysis to ensure that the observed gene expression differences accurately reflected biological signals.

In our study, we identified 16 crosstalk genes, 11 of which showed significant differences, including *AK1*, *ALDH6A1*, and *CHCHD10*. These genes share common pathogenic pathways under T2D and SS, indicating their potential in dual therapeutic strategies. The JAK-STAT pathway has previously been associated with inflammation and immune responses under autoimmune conditions²¹; this study provides novel insights into the interplay between these pathways.

Notably, our findings represent the first recognition of the exact roles of crosstalk genes within these processes. Validation using multicenter clinical samples is recommended to assess their diagnostic efficacy across diverse patient populations, thereby underscoring the significance of these results in enhancing our understanding of the interactions between T2D and SS and presenting new opportunities for therapeutic interventions.

This study introduced a predictive model based on *ALDH6A1* and *IL11RA*. This model demonstrated strong classification accuracy in distinguishing T2D from SS, with an AUC > 0.7. However, ethnicity, sex, and age variations must be considered to ensure the generalizability of the model. This requires external validation and adjustments for subgroup-specific differences in gene expression.

In addition, the present study found significant changes in immune cells in T2D, including macrophages, natural killer T cells, and neutrophils. In addition, notable shifts in central memory CD8+ T cells were observed in the SS group. These findings indicate the critical role of these cells in disease development and provide insights into immunotherapy. Macrophages play a significant role in the onset and development of metabolic diseases, particularly in T2D. T2D is characterized by insulin resistance and chronic low-grade inflammation, which are closely related to the functional and phenotypic changes in macrophages. Studies have revealed that macrophages polarize to the M1 phenotype in the T2D environment, promoting inflammatory responses that lead to tissue damage and insulin resistance.

IL-11 is a pro-inflammatory cytokine that may play a critical role in T2D progression by binding to the *IL-11RA*. Its overexpression can lead to sustained activation of macrophages, exacerbating inflammation and affecting pancreatic β-cell function, resulting in inadequate insulin secretion. Studies have found that *IL-11RA* expression is significantly upregulated in macrophages from patients with T2D and animal models, indicating its important regulatory role in the inflammatory process^{22,23}. Therefore, adjusting macrophage polarization may offer a novel treatment for T2D. Moreover, the activation of natural killer T cells and modulation of neutrophils can target insulin resistance and provide therapeutic opportunities.

This study provides support for the use of immune infiltration data in precision medicine. The study suggests the need for large-scale sequencing and single-cell transcriptomic analyses to create specific immune cell profiles for T2D and SS. In addition, it proposes the design of targeted immunotherapy strategies and establishment of dynamic monitoring systems for immune function assessment to aid personalized clinical decisions.

Overall, although the predictive model based on *ALDH6A1* and *IL11RA* provides promising avenues for early screening of T2D and SS, further research is essential for broader applicability and optimization, along with careful consideration of ethical issues, to protect patient rights and promote sustainable model use. With ongoing refinement, this model holds substantial potential for enhancing clinical diagnostic efficiency and patient management, thereby advancing immunotherapy and therapeutic research on these diseases through improved understanding of immune cell changes.

Although this study reveals important molecular mechanisms between T2D and SS through systematic bioinformatics analysis, it is essential to acknowledge several limitations. First, the limited sample size may have affected the statistical power and reliability of the results, thereby restricting the generalizability of the study's conclusions. Second, this research is based on publicly available databases and multi-cohort data, primarily focusing on molecular-level associations and mechanism exploration, without including experimental validation. Owing to practical constraints such as research timeline and available resources, relevant functional validation through animal experiments or clinical samples could not be completed in this study. This limitation partially affects the biological significance of the findings and the direct confirmation of their clinical application value. We aim to conduct in-depth functional and mechanistic validations targeting key genes and molecular pathways by integrating animal experiments and clinical samples in subsequent research to further enhance the scientific rigor and clinical translational potential of our study. Finally, because the data were sourced from specific public databases, there may be inherent biases; subsequent research should broaden the sample scope and integrate multiple datasets to improve the reliability and generalizability of the analysis results.

Through multilevel bioinformatics analysis, this study identified 16 genes involved in crosstalk (a phenomenon in which genes interact and affect their functions), including *AK1*, *ALDH6A1*, and *CHCHD10*. Among these 16 genes, 11 showed significant differences in expression. Crosstalk genes are abundant in various biological processes and pathways, such as ATP metabolic processes and the JAK-STAT signaling pathway. A predictive model based on *ALDH6A1* and *IL11RA* showed good classification accuracy for differentiating T2D from SS, with an AUC value > 0.7. Our research emphasizes the role of specific immune cells in disease progression and offers a new understanding of the pathophysiology of these two conditions. These findings not only improve the

comprehension of T2D and SS but also provide potential targets for the design of future immunotherapy. *IL-11* interacts with macrophages through *IL-11RA*, thereby affecting inflammatory response. The elevated expression of *IL-11RA* in T2D and SS may exacerbate inflammation. Thus, targeting *IL-11RA* represents a promising therapeutic strategy, potentially achieved through monoclonal antibodies that block *IL-11* from binding to *IL-11RA*, thereby reducing inflammation and improving insulin sensitivity. The effects of *IL-11* involve the JAK-STAT signaling pathway. JAK inhibitors, such as tofacitinib, have shown efficacy in various autoimmune diseases, suggesting their potential utility in T2D and SS. The PI3K-AKT pathway closely relates to insulin signaling, and *IL-11* may affect insulin resistance through this pathway. Therefore, targeting this pathway with PI3K inhibitors may improve the pathophysiology of T2D. Dysfunction of *ALDH6A1* in immature dendritic cells is related to the onset of SS. Enhancing *ALDH6A1* expression or activity can promote the maturation and immune regulatory functions of these dendritic cells, thereby improving autoimmune responses. Gene editing technologies, such as CRISPR/Cas9, can be used to upregulate *ALDH6A1* expression, providing a novel therapeutic target for SS.

Future studies should validate these results by using larger and more diverse samples to increase the external validity of the research. In addition, exploring the clinical application of these findings, such as leveraging immune infiltration information to promote personalized treatment, is an important direction for further research. Through such explorations, we anticipate achieving significant advancements in the management of these diseases and improving the quality of life of patients.

Methods

Data download

We used the R package GEOquery²⁴ (Version 2.70.0) to download the T2D (GSE18732 and GSE29221) and SS datasets (GSE40611 and GSE154926) from the gene expression omnibus (GEO) database²⁵ (<https://www.ncbi.nlm.nih.gov/geo/>). The GSE18732, GSE29221, GSE40611, and GSE154926 datasets contained data from *Homo sapiens*. Specifically, the GSE18732 dataset was on skeletal muscle, microarray platform GPL9486; the GSE29221 dataset was on skeletal muscle, microarray platform GPL6947. The tissue in the GSE40611 dataset was the parotid gland and the chip platform was GPL570. The tissue in the GSE154926 dataset was a minor salivary gland and the chip platform was GPL18573. The GSE18732 dataset contained 45 T2D and 47 control samples; the GSE29221 dataset contained 12 T2D and 12 control samples. The GSE40611 dataset contained 17 SS and 18 control samples; the GSE154926 dataset contained 42 SS and seven control samples. These samples were all included in the analysis, and the specific details of the datasets are presented in Table S1.

Differential analysis

Based on the sample grouping of the T2D dataset, the samples were categorized into T2D and control samples. The R package limma²⁶ (Version 3.54.2) was used to analyze the differences in genes between the T2D and control samples. We set the thresholds for DEGs to $|\log FC| > 0$ and $p < 0.05$. Genes with $\log FC > 0$ and $p < 0.05$ were considered upregulated DEGs; meanwhile, those with $\log FC < 0$ and $p < 0.05$ were considered downregulated DEGs. The results of the differential analysis were displayed by generating a volcano plot using the R package ggplot2 (Version 3.4.4).

Furthermore, limma was used for the differential analysis of genes in the SS and control specimens. The threshold of DEGs was consistent with that of the T2D dataset.

WGCNA

WGCNA¹⁵ is a systems biology method for describing gene association patterns between different samples and identifying gene sets with highly synergistic changes. In addition, the method identifies candidate biomarker genes or therapeutic targets by evaluating intramodular connectivity and module-trait correlations. WGCNA was performed using the WGCNA package of R software²⁷ (Version 1.72–5). First, the correlation coefficient between any two genes was determined, and the weighted value of the correlation coefficient was used to ensure adherence of the connections between the genes in the network to the scale-free network. Subsequently, hierarchical clustering trees were established based on the correlation coefficients between genes. Distinct branches of the cluster tree signified different gene modules, with diverse colors denoting different modules; thereafter, the module significance was computed. The scale-free fitting index was 0.80; the soft threshold values were the best soft threshold values; the combined shear height of the modules was set to 0.2; and the minimum distance was set to 0.2. The minimum module gene number for the WGCNA in GSE18732 was set at 60; the minimum number of module genes for WGCNA in the GSE40611 dataset was set to 100. Afterward, the correlation between the two groups of disease and control samples and the different modules were measured, and the genes in each module were recorded. Finally, all modules with $p < 0.05$ and $|r| > 0.30$ were retained, and the genes in the modules were regarded as module genes.

GO and KEGG pathway enrichment analyses

GO analysis²⁸ is commonly used for large-scale functional enrichment studies, including BPs, CCs, and MFs. Meanwhile, the KEGG²⁹ is a widely used database that stores information on genomes, biological pathways, diseases, and drugs. We used the R package clusterProfiler³⁰ (Version 4.4.4) for GO and genetic pathways (KEGG) enrichment analyses; item selection criteria for the value $p < 0.05$.

GSEA

GSEA³¹ is used to assess predefined gene sets and analyze trends in gene distribution related to phenotypic correlations. This approach facilitates the identification of the contribution of specific gene sets to the observed phenotypes. In the present study, the genes in the GSE18732 dataset were first ranked according to their logFC values; thereafter, the genes were evaluated using GSEA via the R package clusterProfiler³⁰ (Version 4.4.4). The

parameters employed in the GSEA were as follows: seed number, 2025; number of calculations, 1000, minimum number of genes in each gene set, 10; and maximum number of genes in each gene set, 500. Through Molecular Signatures Database (MSigDB)³². For c2 gene set cp. All. V2022.1. Hs. Symbols. The GMT [all Canonical Pathways] (3050) was used for GSEA and the screening criteria was $p < 0.05$.

GSVA

GSVA³³ is a nonparametric and unsupervised analysis approach. It is primarily used to assess the enrichment outcomes of gene sets in a microarray nuclear transcriptome by transforming the expression matrix of gene sets of different samples into an expression matrix of gene sets between samples. This allows the evaluation of the enrichment of distinct pathways in various samples. The h.all.v7.4.symbols.gmt gene set was obtained from the MSigDB³², and GSVA was performed on all genes in the dataset to compute the functional enrichment disparities between the two sample groups. The selection criterion for GSVA was $p < 0.05$. Five pathways with the most significant p-values were selected for presentation.

LASSO regression analysis

LASSO regression analysis is based on linear regression analysis with the incorporation of a penalty term (lambda \times absolute value of the slope). This curtails overfitting of the model and enhances its generalization capability. In the present study, the outcomes of the LASSO regression analysis were presented visually using diagnostic model graphs and variable trajectory diagrams.

Construction of the prediction model

A generalized linear model was used for univariate and multivariate binary logistic regression analyses. The screening criterion of univariate binary logistic regression was $p < 0.05$, and genes that met the screening criterion were included in multivariate binary logistic regression for model construction. Thereafter, the R package rms (version 6.4-0) was used to construct a nomogram based on the results of multivariate logistic regression analysis, and a DCA plot was used to evaluate the diagnostic model.

The PPI network

The GeneMANIA database³⁴ (<https://genemania.org/>) is used to generate hypotheses regarding gene function, analyze gene lists, and prioritize genes for functional analyses. Given a list of query genes, GeneMANIA identifies functionally similar genes by using a large set of genomic and proteomic data. In this mode, each functional genomic dataset is weighted according to the predicted query value. Besides, GeneMANIA has been used to predict gene function. Given a query gene, GeneMANIA identifies genes that are likely to share functions based on how the gene interacts with it. We predicted functionally similar hub genes using the GeneMANIA online website to construct a PPI network.

Construction of a regulatory network

miRNAs play a significant role in the modulation of biological development and evolution. Moreover, they regulate a variety of target genes, and the same target gene is regulated by multiple miRNAs. To analyze the relationship between hub genes and miRNAs, hub genes-related miRNAs were obtained from the ENCORI database³⁵.

TFs control gene expression by interacting with hub genes at the posttranscriptional stage. We retrieved TFs from the ChIPBase database³⁶ (<http://rna.sysu.edu.cn/chipbase/>) to analyze their regulatory effect on hub genes.

Immune infiltration analysis

ssGSEA³⁷ measures the relative abundance of each immune cell infiltrate. Initially, each infiltrating immune cell type was marked, including activated CD8+ T cells, activated dendritic cells, gamma-delta T cells, natural killer cells, and various human immune cell subtypes such as regulatory T cells. Subsequently, the enrichment scores computed using ssGSEA were used to determine the relative abundance of immune cell infiltration in each sample, thereby obtaining the immune cell infiltration matrix. Afterward, R package ggplot2 (Version 3.4.4) was used to construct group comparison maps to illustrate the expression disparities in immune cells between the two groups. Subsequently, immune cells with notable differences between the two groups were selected for subsequent analysis, and the correlation between immune cells was calculated using the Spearman algorithm. The correlation between hub genes and immune cells was calculated using Spearman's algorithm. The R package ggplot2 (version 3.4.4) was used to construct a correlation bubble plot to illustrate the correlation analysis results for the hub genes and immune cells.

Statistical analysis

All data processing and analyses in this study were conducted using R software (Version 4.2.2). In the comparison of continuous variables between the two groups, the statistical significance of normally distributed variables was evaluated using the independent Student's *t* test, except otherwise stated. The Mann-Whitney U test (Wilcoxon rank-sum test) was used to assess the disparities among variables that did not follow a normal distribution. The Kruskal-Wallis test was used to compare three or more groups. Spearman's correlation analysis was used to calculate the correlation coefficients between the different molecules. All statistical *p* values were two-sided if not specified, and $p < 0.05$ was considered statistical significance.

Data availability

The datasets analyzed during the current study are available at the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) repository, including GSE18732, GSE29221, GSE40611, and GSE154926.

Received: 23 March 2025; Accepted: 29 July 2025

Published online: 06 August 2025

References

- Wang, L. et al. Prevalence and treatment of diabetes in China, 2013–2018. *JAMA* **326**, 2498–2506. <https://doi.org/10.1001/jama.2021.22208> (2021).
- Penno, G. et al. Insulin resistance, diabetic kidney disease, and all-cause mortality in individuals with type 2 diabetes: A prospective cohort study. *BMC Med.* **19**, 66. <https://doi.org/10.1186/s12916-021-01936-3> (2021).
- Georgianos, P. I., Vaios, V., Koufakis, T. & Liakopoulos, V. Slowing the progression of chronic kidney disease in patients with type 2 diabetes using four pillars of therapy: The time to act is now. *Drugs* **84**, 1337–1346. <https://doi.org/10.1007/s40265-024-02091-8> (2024).
- Mullins, C. D. et al. CKD Progression and economic burden in individuals with CKD associated with type 2 diabetes. *Kidney Med.* **4**, 100532. <https://doi.org/10.1016/j.kxme.2022.100532> (2022).
- Fox, R. I. Sjögren's syndrome. *Lancet* **366**, 321–331. [https://doi.org/10.1016/s0140-6736\(05\)66990-5](https://doi.org/10.1016/s0140-6736(05)66990-5) (2005).
- Alhalwani, A. Y., Abudawood, K., Qadizadah, A., Jambi, S. & Sannan, N. S. Immunoglobulin A levels and its correlation with neutrophil-to-lymphocyte ratio as inflammatory biomarkers for dry eye disease in type 2 diabetes: A retrospective study. *Front. Immunol.* **14**, 118462. <https://doi.org/10.3389/fimmu.2023.118462> (2023).
- Gutiérrez-Sámerón, M. et al. Metabolic and hormonal remodeling of colorectal cancer cell signalling by diabetes. *Endocr. Relat. Cancer* **28**, R191-r206. <https://doi.org/10.1530/erc-21-0092> (2021).
- Jo, S. & Fang, S. Therapeutic strategies for diabetes: immune modulation in pancreatic β cells. *Front. Endocrinol. (Lausanne)* **12**, 716692. <https://doi.org/10.3389/fendo.2021.716692> (2021).
- Shimizu, T., Nakamura, H. & Kawakami, A. Role of the innate immunity signaling pathway in the pathogenesis of Sjögren's syndrome. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms22063090> (2021).
- Šrámek, J., Němcová-Fürstová, V. & Kovář, J. Molecular mechanisms of apoptosis induction and its regulation by fatty acids in pancreatic β -cells. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms22084285> (2021).
- Kaneto, H. et al. Molecular mechanism of pancreatic β -cell failure in type 2 diabetes mellitus. *Biomedicines* <https://doi.org/10.3390/biomedicines10040818> (2022).
- Shirakawa, J. Signaling pathways that regulate adaptive β -cell proliferation for the treatment of diabetes. *J. Diabetes Investig.* **14**, 735–740. <https://doi.org/10.1111/jdi.14002> (2023).
- Baldini, C. et al. Pathogenesis of Sjögren's disease: One year in review 2024. *Clin. Exp. Rheumatol.* **42**, 2336–2343. <https://doi.org/10.55563/clinexprheumatol/i8iszc> (2024).
- Song, Y. & Zhou, W. Role of TLR7 in the pathogenesis of primary Sjögren's syndrome. *Clin. Exp. Rheumatol.* **42**, 2513–2519. <https://doi.org/10.55563/clinexprheumatol/cmmkod> (2024).
- Liu, W., Li, L., Ye, H. & Tu, W. Weighted gene co-expression network analysis in biomedicine research. *Sheng Wu Gong Cheng Xue Bao* **33**, 1791–1801. <https://doi.org/10.13345/j.cjb.170006> (2017).
- Williams, R. et al. Global and regional estimates and projections of diabetes-related health expenditure: Results from the international diabetes federation diabetes atlas. *Diabetes Res. Clin. Pract.* **162**, 108072. <https://doi.org/10.1016/j.diabres.2020.108072> (2020).
- Tarn, J. et al. Primary Sjögren's syndrome: Longitudinal real-world, observational data on health-related quality of life. *J. Intern. Med.* **291**, 849–855. <https://doi.org/10.1111/joim.13451> (2022).
- Zhang, Y. et al. Atlas of lysine propionylation of human right atrium and ALDH6A1-NADH pathway in new-onset atrial fibrillation after coronary surgery. *Commun. Biol.* **8**, 858. <https://doi.org/10.1038/s42003-025-08264-9> (2025).
- Sweeney, M. et al. Interleukin 11 therapy causes acute left ventricular dysfunction. *Cardiovasc. Res.* **120**, 2220–2235. <https://doi.org/10.1093/cvr/cvaa224> (2024).
- Deng, X., Liu, D., Li, M., He, J. & Fu, Y. Association between systemic immune-inflammation index and insulin resistance and mortality. *Sci. Rep.* **14**, 2013. <https://doi.org/10.1038/s41598-024-51878-y> (2024).
- Li, X. et al. JAK-STAT1 signaling pathway is an early response to helicobacter pylori infection and contributes to immune escape and gastric carcinogenesis. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms23084147> (2022).
- Witcoski Junior, L. et al. Metabolic reprogramming of macrophages in the context of type 2 diabetes. *Eur. J. Med. Res.* **29**, 497. <https://doi.org/10.1186/s40001-024-02069-y> (2024).
- Wong, A., Sun, Q., Latif, I. I. & Karwi, Q. G. Metabolic flux in macrophages in obesity and type-2 diabetes. *J. Pharm. Pharm. Sci.* **27**, 13210. <https://doi.org/10.3389/jpps.2024.13210> (2024).
- Davis, S. & Meltzer, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847. <https://doi.org/10.1093/bioinformatics/btm254> (2007).
- Barrett, T. et al. NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–995. <https://doi.org/10.1093/nar/gks1193> (2013).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
- Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–d426. <https://doi.org/10.1093/nar/gky1038> (2019).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
- Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).
- Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740. <https://doi.org/10.1093/bioinformatics/btr260> (2011).
- Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. <https://doi.org/10.1186/1471-2105-14-7> (2013).
- Franz, M. et al. GeneMANIA update 2018. *Nucleic Acids Res.* **46**, W60-w64. <https://doi.org/10.1093/nar/gky311> (2018).
- Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–97. <https://doi.org/10.1093/nar/gkt1248> (2014).
- Zhou, K. R. et al. ChIPBase v2.0: Decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* **45**, D43–D50. <https://doi.org/10.1093/nar/gkw965> (2017).
- Xiao, B. et al. Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma. *Front. Oncol.* **10**, 607622. <https://doi.org/10.3389/fonc.2020.607622> (2020).

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We acknowledge GEO database for providing their platforms and contributors for uploading their meaningful datasets.

Author contributions

Yuchun Zhang designed and supervised the study. Yibing Wang, Huilin Zeng, and Guangzhao Zhang analyzed the data and wrote the original draft. Xiangjie Hu, Liangliang Zhou, and Xingyuan Yu edited the draft. All authors contributed to the article and approved the submitted version.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-14044-6>.

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025