Journal of Inflammation Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Combining Bulk and Single Cell RNA-Sequencing Data to Identify Hub Genes of Fibroblasts in Dilated Cardiomyopathy

Xiaoyan Huang[1,2], Xiangrong Zhao[1,2], Yaping Li[1,2], Yangmeng Feng[1,2], Guoan Zhang[3], Qiyu Wang[4], Cuixiang Xu[1,2]

[1]Shaanxi Provincial Key Laboratory of Infection and Immune Diseases, Shaanxi Provincial People's Hospital, Xi'an, People's Republic of China; [2]Shaanxi Engineering Research Center of Cell Immunology, Shaanxi Provincial People's Hospital, Xi'an, People's Republic of China; [3]Department of Cardiovascular Surgery, Shaanxi Provincial People's Hospital, Xi'an, People's Republic of China; [4]Department of Graduate School, Yan'an University, Yan'an, People's Republic of China

Correspondence: Cuixiang Xu, Email xucuixiang1129@163.com

Background: Dilated cardiomyopathy (DCM) is the second leading cause of heart failure, with intricate pathophysiological underpinnings. In order to shed fresh light on the mechanistic research of DCM, we combined bulk RNA-seq and single-cell RNA-seq (scRNA-seq) data to examine significant cells and genes implicated in the disease.

Methods: This analysis employed publicly accessible bulk RNA-seq and scRNA-seq DCM datasets. The scRNA-seq data underwent normalization, principal component, and t-distribution stochastic neighbor embedding analysis. Cell-to-cell communication networks and activity analysis were conducted using CellChat. Utilizing enrichment analysis, the marker genes' role in the active cells was evaluated. After screening by limma software and weighted gene co-expression network analysis, the differentially expressed genes (DEGs) served as hub genes. Furthermore, these hub genes were subjected to immunological studies, transcription factor expression, and gene set enrichment. Lastly, the expression of the four hub genes and their connection to DCM were verified using the rat models.

Results: Fibroblasts and monocytes were chosen as hub cells from among the eight identified cell clusters; their marker genes intersected with DEGs to yield six hub genes. In addition, the six hub genes and the essential module genes intersected to yield four essential genes (*ASPN, SFRP4, LUM*, and *FRZB)* that were connected to the Wnt signaling pathway and highly expressed in fibroblast. The four hub DEGs had an expression pattern in the DCM rat model experiment results that was in line with the findings of the bioinformatics study. Additionally, there was a strong correlation between decreased cardiac function and the up-regulation of *ASPN, SFRP4, LUM*, and *FRZB*.

Conclusion: Ultimately, bulk RNA-seq and scRNA-seq data identified fibroblasts and monocytes as the main cell types implicated in DCM. The highly expressed genes *ASPN, FRZB*, *LUM*, and *SFRP4* in fibroblasts may aid in the mechanistic investigation of DCM.

Keywords: dilated cardiomyopathy, single cell RNA-seq, bulk RNA-seq, weighted gene co-expression network analysis, wnt signaling pathway, fibroblasts

Introduction

Dilated cardiomyopathy (DCM) is the second most frequent form of coronary artery disease, and heart failure affects over 26 million people worldwide.[1] Left ventricular dilatation, systolic dysfunction without aberrant loading circumstances, and significant coronary artery disease are the clinical characteristics of DCM.[2] About 20% to 35% of DCM cases are known to have a familial illness. These were primarily autosomal dominant cases. Myocarditis, persistent tachyarrhythmia, and endocrine abnormalities are other causes that might lead to DCM.[3] The disease mechanism and treatment methods of DCM are of great interest due to its intricate osteogenesis mechanism and elevated mortality risk. Powerful instruments to assist this interest have recently been made available by more modern research methodologies.

Received: 30 May 2024
Accepted: 2 August 2024
Published: 14 August 2024

Gene insights from transcriptomics can be used to better understand a range of illness characteristics. Applications for RNA-seq exist in various domains, including healthcare.[4] Moreover, a variety of biological problems have propelled bioinformatics advancements and given rise to fresh uses for bulk RNA-seq. The in-depth analysis of gene expression profiles in tissue samples is made easier by the tissue-level transcriptional data that bulk RNA-seq technology offers. However, minute variations amongst individual cells may be obscured by this sequencing method.[5] Conversely, the multilayered condition of individual cells is observed by the application of single-cell RNA-seq (scRNA-seq) technology. A good tool for precisely understanding the transcriptome state of various cell types is scRNA-seq.[6]

Numerous opportunities for illness study are presented by widely used sequencing technology. RNA-seq, however, has been applied extensively to DCM research. Sweet et al used RNA-seq data to show that defective cell-matrix adhesion signatures and cell-cell adhesion signatures are characteristic aspects of DCM.[7] These findings have been validated by experiments. Other researchers have screened out genes that might be important in DCM pathophysiology using DCM RNA-seq data from public databases.[8] DCM research has also made use of scRNA-seq. Specific cardiac cell types in DCM can have distinct disease-associated transcriptome alterations identified by ScRNA-seq.[9] Based on scRNA-seq data, Zhu et al provided a thorough map of the alteration of cardiac phenotype; the features of this map will aid in the investigation of new interventional targets.[10]

Researchers can comprehensively characterize the various cell heterogeneities in tissues thanks to scRNA-seq technique. However, the sample size is constrained due to the high cost of single-cell sequencing.[11] Gene expression profiles from hundreds of people can be created using bulk RNA-seq.[12] Consequently, there are important ramifications for comprehending cell type ratios and the connections among complicated disorders when combining scRNA-seq with bulk RNA-seq.[13] In order to determine the genes and cell types that might be implicated in the course of DCM, we synthesized scRNA-seq and bulk RNA-seq data on the disease from the NCBI gene expression omnibus database.

Methods

Data Acquisition

The scRNA-seq dataset GSE145154 was downloaded from the Gene Expression Omnibus (GEO) database. The scRNA-seq dataset included eight DCM and four standard samples. In addition, two bulk RNA-seq datasets, GSE5406 (86 DCM and 16 standard samples) and GSE57345 (82 DCM and 136 normal samples), were also downloaded from GEO to identify the critical genes in DCM. A system Verilog assertion algorithm was employed to correct the chip data.

Single Cell Analysis

The expression profiles were analyzed using Seurat to identify low-expression genes (nFeature_RNA > 50% and percent. mt < 5%).[14] Data were then standardized, normalized, and PCAed. ElbowPlots were used to determine the ideal number of principal components (PC), while t-distribution stochastic neighbor embedding (TSNE) analyses were utilized to determine cell cluster placements. Finally, we set the logic to extract marker genes for each cell cluster from the single-cell expression profile. The FindAllMarkers threshold is 1. Marker genes with |avg_log2FC| > 1 and p_val_adj < 0.05 were unique to each cell cluster.

Ligand-Receptor Interaction Analysis

To look at the role and activity of each type of cell in the disease, we conducted an analysis of ligand-receptor interactions. Cell subtypes obtained from single-cell analysis were utilized as the cell information, and normalized single-cell expression profiles were imported into CellChat.[15] The degree of proximity between these cells was then measured using the weights and counts of their interactions.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genome (KEGG) Annotation

Unique marker genes were functionally annotated using clusterProfiler to explore their functional relevance comprehensively.[16] GO and KEGG were used to evaluate relevant functional categories. Enriched terms with P < 0.05 and Q < 0.05 were significantly enriched.

Identification of Differentially Expressed Genes (DEGs)

Corrections were made to data from GSE5406 and GSE57345, using the ComBat algorithm. The batch effects amongst the chips were lessened following the adjustment. Additionally, we applied a threshold of P-value < 0.05 and |Log2FC| > 1 to find DEGs between the DCM and standard groups using the limma software. Furthermore, hub genes were discovered to be the genes that intersected DEGs and critical cell marker genes.

Weighted Gene Co-Expression Network Analysis (WGCNA)

We used WGCNA-R to create a weighted gene co-expression network to find co-expressed gene modules.[17] To assess network connectedness, the weighted adjacency matrix was turned into a topological overlap matrix. Hierarchical clustering was utilized to create a topological overlap degree matrix clustering tree. Based on weighted correlation coefficients, genes were classified by expression patterns. Lastly, modules were formed from genes having comparable patterns. The hub genes and intersecting genes between the two modules most strongly related with the trait were key.

Gene Set Enrichment Analysis (GSEA)

GSEA was performed on disease expression profiles to identify differences in signaling pathways between the high and low-key gene expression groups.[18] Gene sets were filtered using 500 and 15 gene sizes. Following 1000 permutations, enriched gene sets were identified with a P-value < 0.05.

Regulatory Network Analysis of Key Genes

The Cistrome DB database was used to analyze the regulatory link between transcription factors and essential genes. The genome file was set to hg38 and the transcription start site to 10 kb. The regulatory network was displayed using Cytoscape.

Construction of animal Models with DCM

Eight-week-old male Sprague-Dawley (SD) rats weighing 180 ± 20 g were purchased from the Laboratory Animal Center of Xi'an Jiaotong University in China and maintained in a controlled environment at ~23 °C with 40–50% humidity under a 12/12 h dark/light cycle. The rats were randomly divided into two groups: a control group and the DCM group. Four rats per group. Each rat in the DCM group was administered a cumulative dose of 12 mg/kg doxorubicin (DOX, Sigma, St. Louis, USA) four consecutive intraperitoneal cycles for 28 days (injected once every seven days) by the intraperitoneal injection. The control group received injections of the same volume of 0.9% saline at the same time points. Transthoracic echocardiography was performed at fifth week to evaluate heart function while under deep anaesthesia with pentobarbital sodium (50 mg/kg, Sigma, USA). At the fifth week, heart tissue samples were collected, and all of the rats were put to death by breathing in an excessive amount of isoflurane. Prior to sacrifice, an echocardiogram was done, and the body weight was recorded during modeling.

Real-Time Quantitative PCR (RT-qPCR)

Total RNA extraction from cardiac tissue of rats was performed using Total RNA Kit I (Omega Bio-tec, Inc, USA). Total RNA was converted into cDNA on the ice using the PrimeScript™ RT reagent kit (TaKaRa, Dalian, China). Quantitative PCR (qPCR) was performed using the Green PCR Master Mix System (Thermo Fisher Scientific) on an ABI7500 Real-Time PCR System (Thermo Fisher Scientific). GAPDH served as the internal point of reference. The relative fold change was assessed with the $2^{-\Delta\Delta CT}$ method. Similarly, The RT-qPCR experimental operation process of the cardiac tissue in the rat model is also carried out as described above. The primer sequences used are listed in Table 1.

Echocardiography

An intraperitoneal dose of 30 mg/kg pentobarbital was used to anesthetize the rats. The Vevo 3100 small animal ultra-high-resolution ultrasound imaging device was used to do echocardiography. Using long- and short-axis pictures of the left ventricle, the following parameters were determined: left ventricular internal diameters at diastole (LVIDd), systole

Table 1 Primer Sequences of Real-Time PCR

Gene	Primer Sequences
Rat-ASPN	F: 5′- GACGTCTAGAGTTAGCTAGAAGTTT-3′ R: 5′-AGTATCCTCAAGACTCAGGCTCT-3′
Rat-FRZB	F: 5′-AGAGCGCTCCAGGTTACTCT-3′ R: 5′-CCCAGTCCAAGATGACGGAG-3′
Rat-LUM	F: 5′-CACCGGGCTTCAATACTTGC-3′ R: 5′-GCTCTTGACATCAAACTTTTCG-3′
Rat-SFRP4	F: 5′-GTTCCTGCACGATCCCATCA-3′ R: 5′-TGTCTATCCACTTCACATCTTCCG-3′
Rat-GAPDH	F: 5′-ACTCCACGACATACTCAGCA-3′ R: 5′-CATCAACGACCCCTCATT-3′

(LVIDs), and fraction shortening (FS). Three cardiac cycles' worth of data were analyzed to assess cardiovascular function.

Correlation Between Hub Genes and Cardiovascular Parameters in Echocardiography

The parameter (EF%, FS%, LVIDs and LVIDd) was analyzed using the Pearson algorithm, and the results were visualized with the R package "ggplot2".[19]

Immunohistochemistry

Tissue sections fixed in paraffin, measuring 4 μm in thickness, were rehydrated and dewaxed. The antigen was extracted and subsequently subjected to incubation at 4 °C overnight with the following rabbit anti-rat polyclonal antibodies: anti-ASPN (Dilution 1:500, Abcam), anti-FRZB (Dilution 1:200, Abcam), anti-LUM (Dilution 1:4000, Abcam), anti-SFRP4 (Dilution 1:100, Abcam), and anti-β-actin (Dilution 1:1000, Abcam). Following this, the secondary antibody conjugated with HRP was incubated at room temperature for half an hour, and after that, the antigen was stained in a DAB detection kit and restained with hematoxylin for 30 minutes. Ultimately, they were examined using an Olympus Corporation BX41 fluorescent microscope (amplification: × Observe below 400), and quantitatively analyzed via Image-Pro Plus analysis software 6.0 (Media Cybernetics).

Statistical Analysis

GraphPad Prism 9.0 software (GraphPad Software, Inc., San Diego, CA, USA) was used to analyze all the data, which were reported as the mean±standard deviation (SD) of three independent experiments. Dunnett's post-hoc test was used after a one-way analysis of variance or Student's *t*-test to ascertain the statistical differences between the groups. A statistical significance threshold of $P < 0.05$ was applied.

Results

Eight Cell Clusters Were Identified Using scRNA-Seq Data

First, the expression levels in 63,722 cells were detected using GSE145154. Subsequently, nFeature_RNA and nCount_RNA were preliminarily screened according to thresholds (Supplementary Figure 1). Finally, we obtained the expression levels of the 53,767 cells and used them for subsequent analyses. Ten genes with the highest standard deviations were *IGKC, IGHG1*, *JCHAIN, IGHG3*, *SPP1, CCL21*, *IGHG4, TPSAB1*, *IGLC2*, and *IGLC1* (Figure 1A). PCA was used to analyze the batch effect after dimensionality reduction of the samples, and the results showed that the batch effect between the samples was not evident (Figure 1B). The optimal PC was 17, as observed using ElbowPlot (Figure 1C). Finally, using TSNE analysis, we obtained 20 clusters (Figure 1D). Different groups expressed these flag

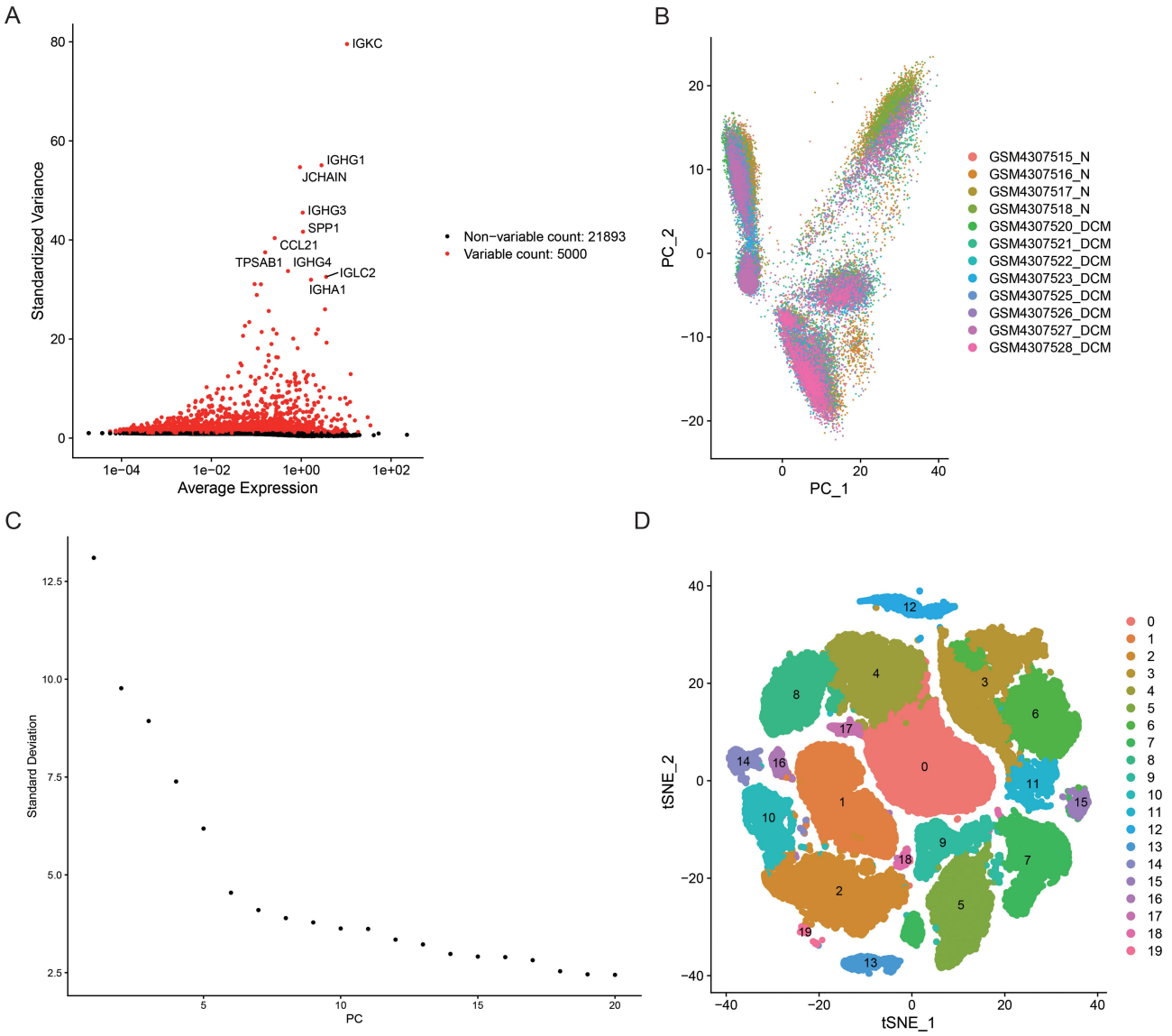

Figure 1 Single cell analysis. (**A**) Genes expression pattern. (**B**) Principal component analysis. (**C**) Screening the optimal number of principal components using ElbowPlot. (**D**) The T-distributed Stochastic Neighbor Embedding (t-SNE) analysis. Each point represents one cell, and the point colors represent the cluster types.

genes (Supplementary Figure 2). Twenty clusters were annotated into eight-cell categories, including T cells, monocytes, endothelial cells, smooth muscle cells, NK cells, fibroblasts, tissue stem cells, and B cells (Figure 2). Subsequently, the FindAllMarkers function was used to extract marker genes specific to each cell cluster from single-cell data.

Fibroblasts and Monocytes Were Two Important Cell Types in DCM

Complex interaction pairings among the cell clusters were revealed by ligand-receptor relationship analysis (Figure 3A and B). We investigated the potential crosstalk between cells by examining ligand and receptor relationships. Sixteen intercellular receptor–ligand relationships were discovered. Among these is a strong receptor–ligand interaction between MIF in endothelial, NK, and tissue stem cells and CD74-CD44 in monocytes. Furthermore, we discovered that fibroblasts, monocytes, and smooth muscle cells were more likely to interact with other cells (Figure 3C). Furthermore, the greatest and second-highest numbers of cells interacted with fibroblasts and monocytes, respectively. As a result, we chose potential gene sets for fibroblast and monocyte marker genes for further investigation (Figure 3D).

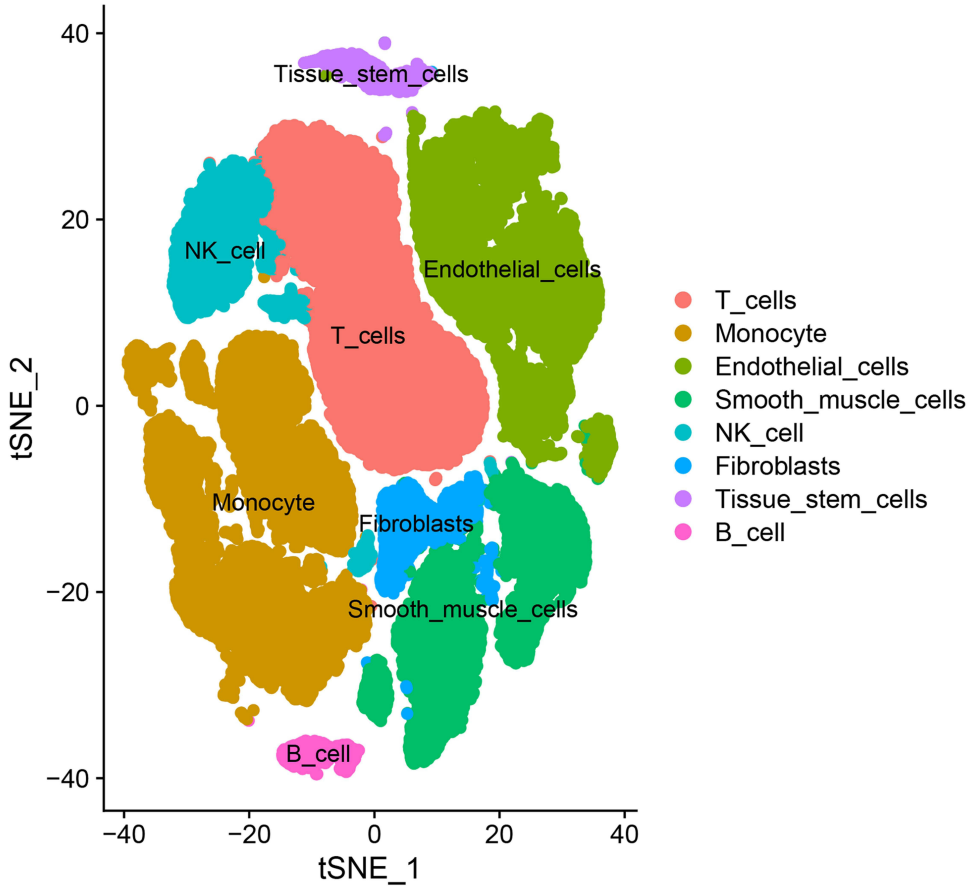

Figure 2 The T-distributed Stochastic Neighbor Embedding (t-SNE) analysis. Each point represents one cell, and the point colors represent the cell types.

Enrichment Analysis Revealed the Potential Function of Candidate Genes

Marker gene functions in fibroblasts and monocytes were examined using GO and KEGG enrichment analysis. These potential genes were primarily involved in extracellular matrix organization, extracellular structure organization, extracellular matrix containing collagen, MHC protein complex, extracellular matrix structural constituents, and collagen binding, according to GO enrichment analysis (Figure 4A). The majority of the significantly enriched signaling pathways were related to phagosomes, viral myocarditis, *Staphylococcus aureus* infection, and graft-versus-host disease (Figure 4B).

Identification of DEGs and Hub Genes

Corrections were made to data from GSE5406 and GSE57345, using the ComBat algorithm. Batch effects between chips were lessened following rectification (Supplementary Figure 3A and B).

Additionally, a total of 16 genes (*SERPINA3, ASPN, LUM, FCN3, FRZB, CD163, SFRP4, LYVE1, IL1RL1, MXRA5, PLA2G2A, MYH6, CYP4B1, SERPINE1, NPPA*, and *HBB*) were obtained when we utilized the limma software to find the DEGs between the DCM group and the standard labor group, (Figure 5A). Among these were seven upregulated DEGs and nine downregulated DEGs. Ultimately, six hub genes for more investigation were obtained by intersecting these 16 genes with the candidate genes (Figure 5B).

Construction of a Co-Expression Network

We performed WGCNA based on the dataset's top 5000 genes in variance to further identify the core genes affecting DCM progression. After the soft threshold was set to 5 (Figure 6A), 14 gene modules were detected: black (246), blue (984), brown (549), cyan (373), green (352), green-yellow (151), gray (27), light cyan (76), magenta (157), midnight blue

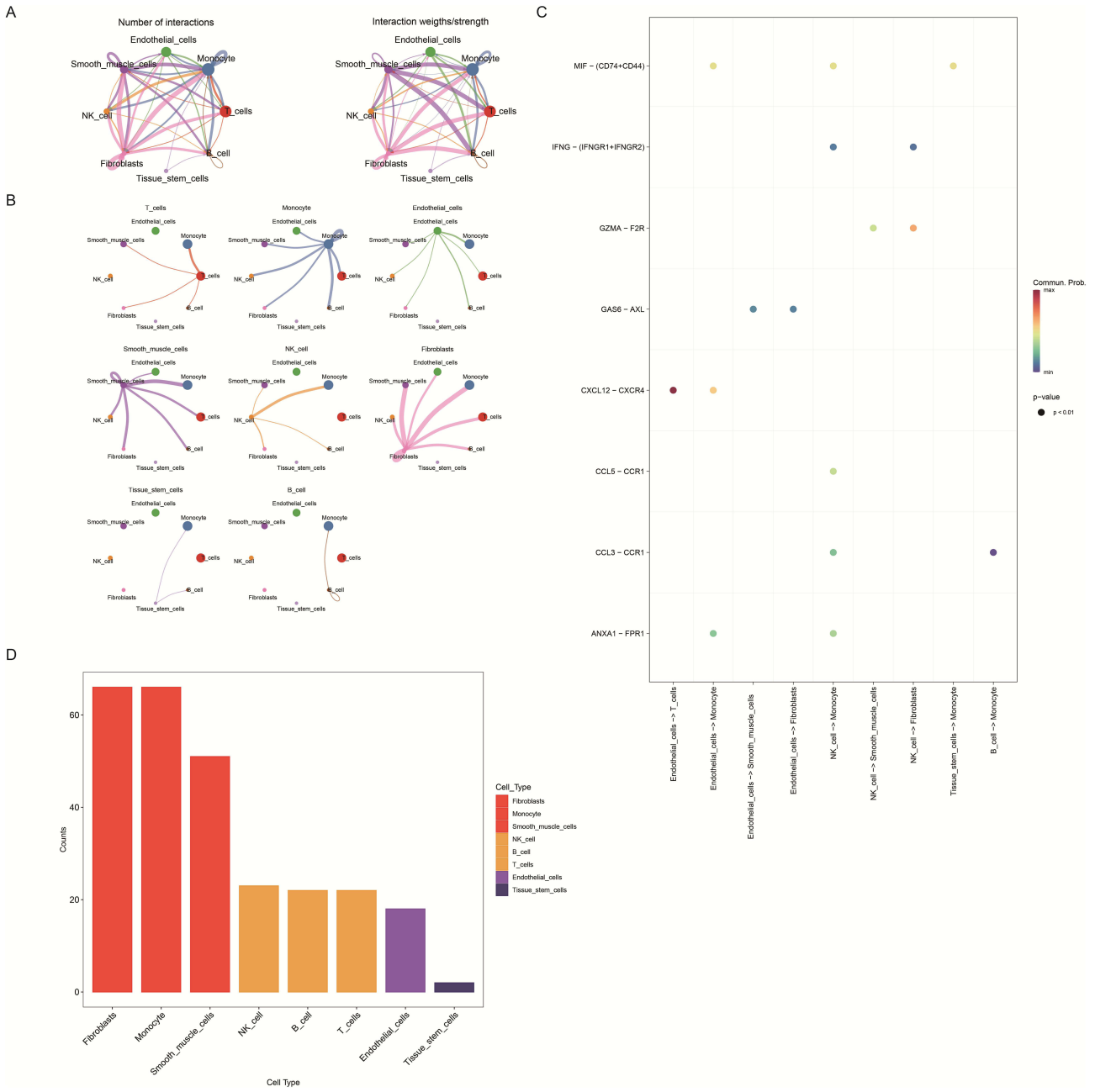

Figure 3 Cell communication analysis. **(A)** and (**B**) Network of eight cell types of intercellular interactions. (**C**) Analysis of cellular cross talk via ligand–receptor interactions. (**D**) Number of marker genes in different cell types.

(87), purple (154), salmon (124), tan (148), and turquoise (1572) (Figure 6B). Subsequently, we selected two black and magenta modules with the strongest positive and negative correlations with disease phenotypes for follow-up analysis (Figure 6C). Ultimately, four essential genes (*ASPN, SFRP4*, *LUM*, and *FRZB)* were obtained by intersecting the genes in the two modules that were chosen with the six hub genes (Figure 6D).

Signaling Pathways and Expression Analysis of Key Genes

According to GSEA results, there was a significant correlation ($P < 0.05$) between low expression of *ASPN* and the signaling pathways for vascular endothelial growth factor (VEGF), glycerophospholipid metabolism, apoptosis, acute myeloid leukemia, insulin, amyotrophic lateral sclerosis, and Erb-B2 receptor tyrosine kinase (ERBB) ($P < 0.05$)

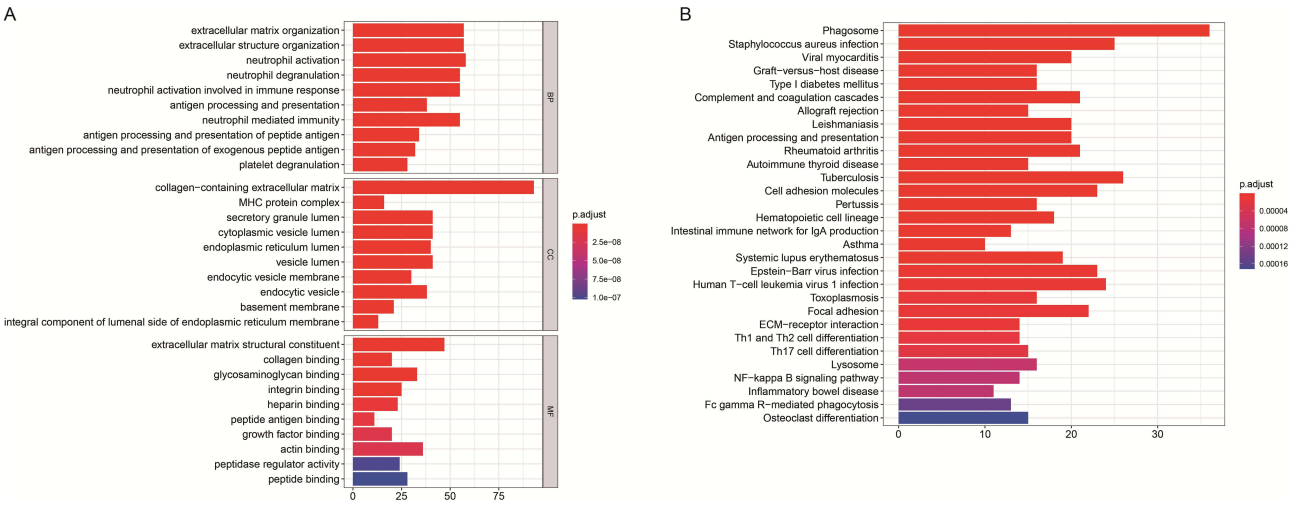

Figure 4 Enrichment analysis of marker genes in fibroblasts and monocyte. (**A**) Gene Ontology analysis. (**B**) Kyoto Encyclopedia of Genes and Genomes analysis. BP, CC, and MF represents biological process, cellular components, and molecular function, respectively.

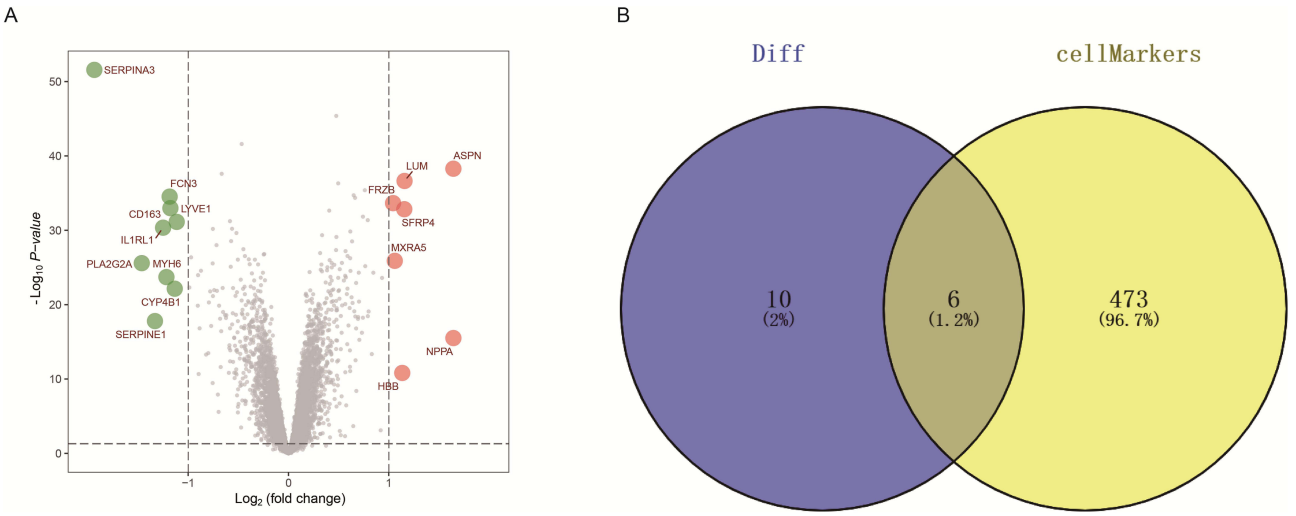

Figure 5 Screening of hub genes. (**A**) Identification of differentially expressed genes between dilated cardiomyopathy and normal samples. (**B**) Intersection of differentially expressed genes and marker genes in fibroblasts and monocyte.

(Supplementary Table 1). *FRZB* was negatively correlated with the VEGF signaling pathway, glycerophospholipid metabolism, galactose metabolism, sphingolipid metabolism, apoptosis, and amino sugar and nucleotide sugar metabolism. *LUM* was negatively correlated with the VEGF signaling pathway, apoptosis, seven amino acid metabolism, galactose metabolism, acute myeloid leukemia, amyotrophic lateral sclerosis, pentose phosphate pathway, endometrial cancer, and ERBB signaling pathway. Furthermore, there was a positive correlation found between elevated *SFRP4* expression and the Wnt signaling pathway, type I diabetes, and the metabolism of beta-alanine.

Construction of a Transcriptional Regulatory Network of Key Genes

To examine the transcriptional regulatory networks involving the four key genes, we projected their transcription factors. According to the findings, transcription factors for *ASPN, FRZB, LUM*, and *SFRP4* were predicted to be 80, 92, 78, and 54. We constructed the transcriptional regulatory network of important genes related to DCM using Cytoscape, based on the aforementioned regulatory connection pairings (Figure 7A). Additionally, we identified six crucial transcription factors (*CEBPB, NR3C1, AR, POLR2A, EP300*, and *CTCF*) by intersecting the transcription factors implicated in these

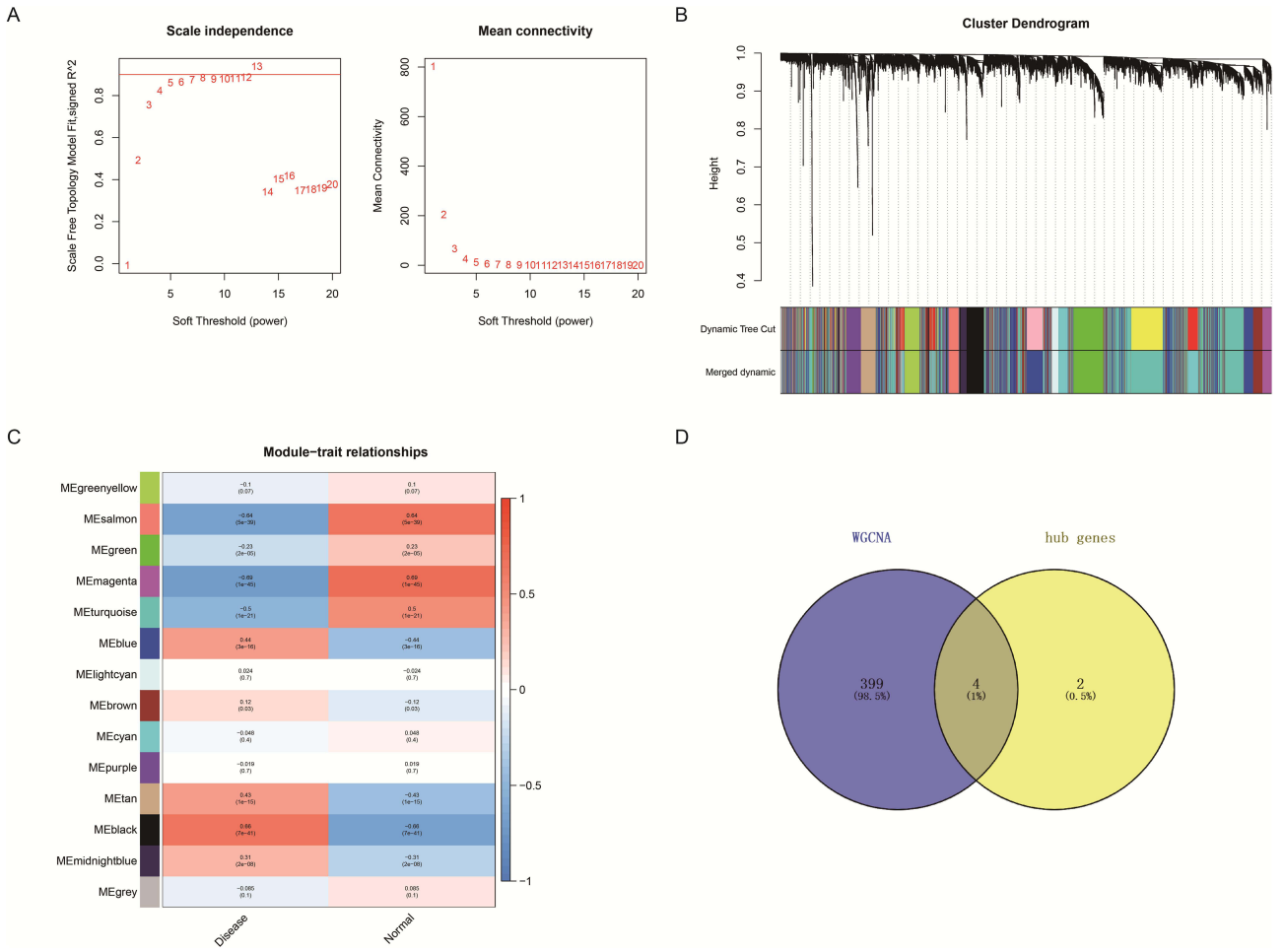

Figure 6 Selection of key genes. (**A**) Selected soft threshold β. (**B**) Cluster dendrogram and module division of genes. (**C**) Module-trait correlation heat map. (**D**) Intersection of hub genes and genes in black and magenta modules.

four genes (Figure 7B). Furthermore, we examined the expression patterns of four key genes in various cell types, and the findings demonstrated that fibroblasts were the primary source of these high levels of expression (Figure 7C).

Relationship Between Key Genes and Immune Factors

According to GSEA results, there was a significant correlation ($P < 0.05$) between low expression of *ASPN* and the signaling pathways for vascular endothelial growth factor (VEGF), glycerophospholipid metabolism, apoptosis, acute myeloid leukemia, insulin, amyotrophic lateral sclerosis, and Erb-B2 receptor tyrosine kinase (ERBB) (Supplementary Figure 4). *FRZB* was negatively correlated with the VEGF signaling pathway, glycerophospholipid metabolism, galactose metabolism, sphingolipid metabolism, apoptosis, and amino sugar and nucleotide sugar metabolism. *LUM* was negatively correlated with the VEGF signaling pathway, apoptosis, seven amino acid metabolism, galactose metabolism, acute myeloid leukemia, amyotrophic lateral sclerosis, pentose phosphate pathway, endometrial cancer, and ERBB signaling pathway. Moreover, high *SFRP4* expression was positively correlated with the Wnt signaling pathway, type I diabetes mellitus, and the metabolism of beta-alanine.

Experimental Validations of four Hub Genes Expression in DCM Rats

During modeling, the DCM group had considerably lower body weight than the Control group ($P<0.05$) (Figure 8A).

The qPCR results of myocardial tissue samples showed elevated mRNA expression levels of *ASPN, FRZB, LUM* and *SFRP4* ($P<0.05$) (Figure 8B). Echocardiography revealed decreased EF% and FS% ($P < 0.05$) in the DCM group

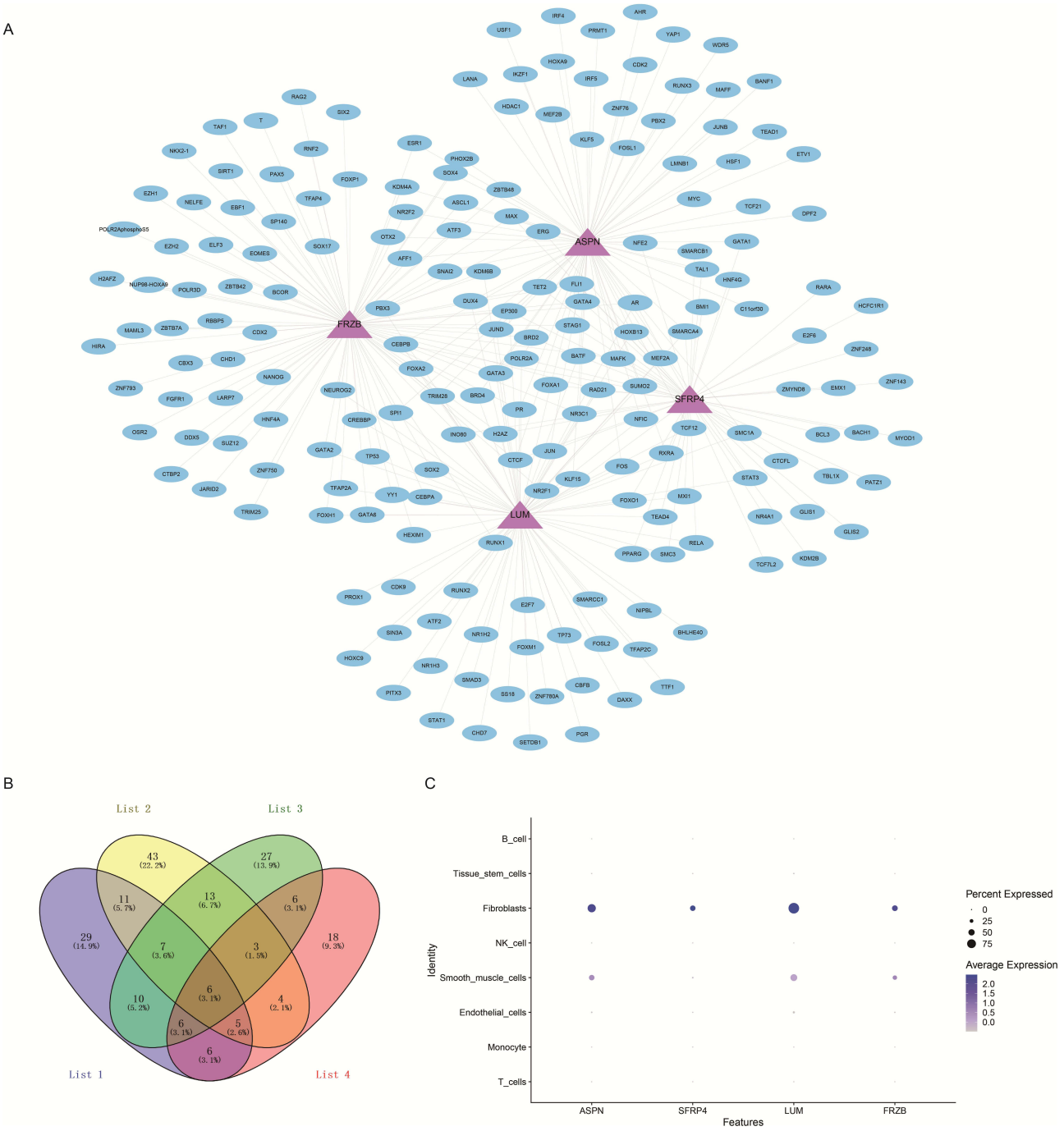

Figure 7 Transcriptional regulation analysis. (**A**) Transcriptional regulatory network. (**B**) Intersection of transcription factors for key genes. (**C**) Expression of key genes in eight cell types.

compared to the Control group, but significantly higher LVIDs ($P < 0.05$) (Figure 8C and D). We examined the relationship between cardiac function and four hub genes (*ASPN, FRZB, LUM*, and *SFRP4*). The number of ASPN PCR cycles correlated positively with EF% ($R = 0.9158$, $P = 0.0014$) and FS% ($R = 0.8745$, $P = 0.0045$) but negatively with LVIDs ($R = -0.7806$, $P = 0.0222$). The number of FRZB PCR cycles correlated positively with EF% ($R = 0.8015$, $P = 0.0168$) and FS% ($R = 0.7628$, $P = 0.0277$) and negatively with LVIDs ($R = -0.7372$, $P = 0.0369$). The number of LUM PCR cycles correlated positively with EF% ($R = 0.8264$, $P = 0.0114$), FS% ($R = 0.7603$, $P = 0.0385$), and negatively with LVIDs ($R = -0.7661$, $P = 0.0266$). The number of SFRP PCR cycles correlated positively with EF% ($R = 0.9476$,

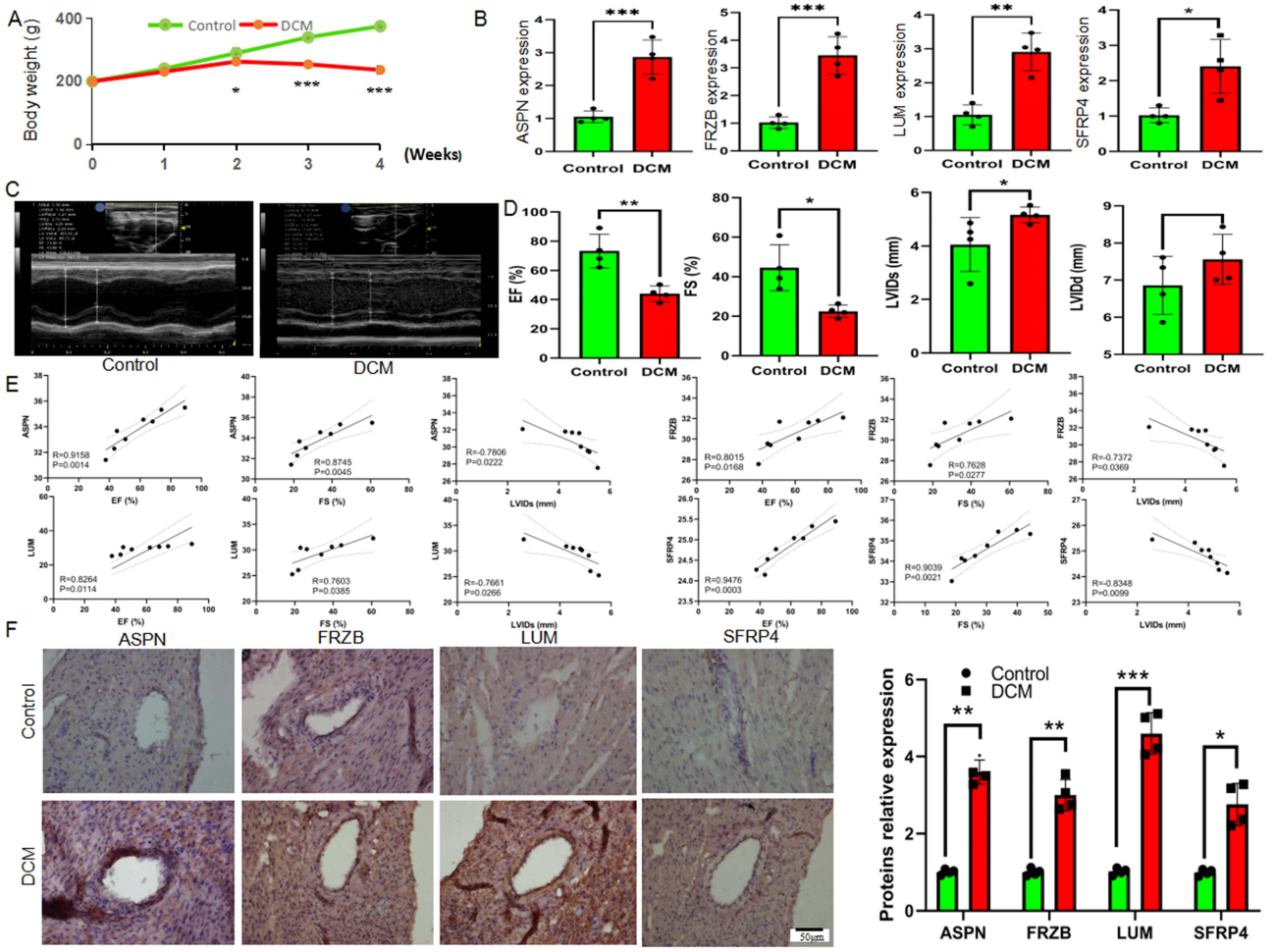

Figure 8 Four hub genes expression in DCM rats. **(A)** the body weight in the Control and DCM group. **(B)** the mRNA level. **(C)** the echocardiography features in the Control and DCM groups. **(D)** the analysis of EF, FS, LVIDs, and LVIDd in the Control and DCM group. **(E)** the correlations between four optimal hub genes mRNA levels, and cardiac functional parameters in the Control and DCM group. **(F)** Protein levels of *ASPN, FRZB, LUM*, and *SFRP4* by immunohistochemistry and quantitative analysis in cardiac tissues in Control and DCM groups. Compared with Control group, **P*<0.05, ***P*<0.01, ****P*<0.001.

P=0.0003), FS% (R = 0.9039, P= 0.0021), and negatively with LVIDs (R = −0.8348, P= 0.0099) (Figure 8E). After that, immunohistochemistry confirmed the protein expression levels of *ASPN, FRZB, LUM*, and *SFRP4* in the Control and DCM groups. Results indicate increased protein expression levels for *ASPN, FRZB, LUM* and *SFRP4* (P<0.05), commensurate with mRNA levels (P<0.05) (Figure 8F).

Discussion

DCM is clinically diverse and varies widely between patients. This difference may be due to genetic-environmental interactions.[20] In recent years, researchers have made great strides in DCM diagnosis and prognosis.[21,22] Due to DCM's clinical variability, future therapeutic and management research is difficult. We chose fibroblast and monocyte marker genes for receptor-ligand interaction analysis in this work. We intersected DEGs from DCM patients and normal individuals to find six hub genes. Following WGCNA, we discovered four crucial genes: *ASPN* (asporin), *SFRP4* (secreted frizzled-related protein 4), *LUM* (lumican), and *FRZB* (frizzled-related protein). The transcription factor regulatory connections and cell type expression of the four essential genes were examined.

Fibroblasts are among the non-immune cells that are crucial to the immunological processes. Acting as intermediaries between immunological and adaptive cells, cardiac fibroblasts actively promote immune responses. When cardiac fibroblasts are stimulated by IL-17A to produce granulocyte and monocyte chemokines, which in turn stimulate adaptive T cells, they can draw in innate T cells and worsen immunological responses.[23] Once cardiac fibroblasts are injured, they become myofibroblasts and cause cardiac fibrosis. This transition involves pathological processes, such as ventricular dilation, cardiomyocyte

hypertrophy, and, ultimately, heart failure.[24] Serum fibroblast growth factor 21 levels are associated with DCM severity and prognosis and can be used as a prognostic marker for DCM.[25] Furthermore, fibroblast molecular imaging can offer additional data to aid in the diagnosis of heart failure or DCM.[26]

The study discovered four important genes (*ASPN, FRZB*, *LUM*, and *SFRP4*) that play a role in numerous diseases, including gastric, non-small cell, lung, and bladder cancer.[27–29] ASPN is an extracellular matrix (ECM) protein that binds to type I collagen and induces collagen mineralization.[30] ASPN was upregulated in DCM, which is consistent with our findings. ASPN may regulate DCM via the ECM signaling pathway as a miR-129-5p target protein.[31] Animal models of ischemic cardiomyopathy showed increased expression of ASPN after transverse coarctation of the aorta or coronary artery closure.[32] In addition, ASPN has been reported to increase apoptosis and fibrosis in H9C2 cardiomyocytes induced by glycated low-density lipoproteins.[33] FRZB acts as a regulator of the Wnt signaling pathway through its interaction with Wnt and plays a regulatory role in cell growth and differentiation in specific cell types.[34] FRZB is highly expressed in DCM, which makes it a good biomarker and diagnostic tool.[35] LUM is a class II SLRP significantly upregulated in various lung, colon, and gastric tumors.[36] LUM is also upregulated in patients with heart failure. Gene knockout investigations in mice revealed that LUM deficiency reduced collagen cross-linking in pressure-loaded hearts, resulting in ventricular dilatation, systolic dysfunction, and mortality.[37] As a member of the secreted glycoprotein family, *SFRP4* is associated with various cardiovascular diseases.[38] Administration of recombinant SFRP4 reduces fibrotic scar size and improves cardiac function after ischemic injury.[39] Recombinant SFRP4 also alleviates atherosclerosis in ApoE-/- mice by reducing inflammation and oxidative stress, which may be related to the Wnt signaling pathway.[40]

Using GSEA, *ASPN*, *FRZB*, and *LUM* were found to negatively correlate with the VEGF signaling pathway. These three genes were substantially expressed in DCM. VEGF is downregulated in DCM,[41] supporting our findings. Angiogenesis in the ischemic heart is stimulated by VEGF, a fibroblast growth factor family member.[42] VEGF-B preserves angiogenesis by activating VEGF receptors, making it a potential DCM therapy.[43] Cardiac fibrosis involves Wnt signaling. Wnt signaling is quiescent in normal adult cardiomyocytes.[44] It is, however, activated in disease states.[45] In this study, we found the Wnt signaling pathway was favorably linked with increased SFRP4 expression. Heart valve development and embryonic heart specification require Wnt signaling. Wnt/β-catenin plays a role in the endothelial-to-mesenchymal transition in DCM formation.

Targeting transcription factors that are active in diseases is an alternative strategy for disease treatment.[46] This study examined six transcription factors that regulate important genes: (*CEBPB, NR3C1*, *AR, POLR2A*, *EP300*, and *CTCF*). Further in vitro and in vivo research are needed to determine the regulatory roles of transcription factors in DCM.

There were certain limitations on this investigation. First, because DCM is such a profoundly varied condition, further research on it will require a larger sample size and clinical data. But as this work is mostly focused on bioinformatics, more in vivo and in vitro tests will be needed to confirm the findings. Second, in the future research, we will further explore the role of identified genes in DCM and extend the research to other cell types or signaling pathways. It will provide strategic basis for the development of new drugs and precise treatment of DCM.

Conclusion

In conclusion, we integrated scRNA-seq and bulk RNA-seq data and discovered that fibroblasts and monocytes are two significant cell types implicated in DCM. Four vital genes (*ASPN, FRZB*, *LUM*, and *SFRP4*) highly expressed in fibroblasts may be involved in DCM and targeted for DCM treatment.

Abbreviations

DCM, Dilated cardiomyopathy; scRNA-seq, single-cell RNA-seq; DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; PCA, principal component analysis; PC, principal components; TSNE, t-distribution stochastic neighbor embedding; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genome; WGCNA, weighted gene co-expression network analysis; GSEA, Gene set enrichment analysis; RT-qPCR, Real-time quantitative PCR; LVEF, left ventricular ejection fraction; FS, fraction shortening; LVIDs, left ventricular internal diameters at systole; LVIDd, left ventricular internal diameters at diastole.

Data Sharing Statement

The original data presented in the study are included in the article/Supplementary materials, which will be made available by the authors.

Ethics Approval

Human data obtained approval from the Ethics Committee of Shaanxi Provincial People's Hospital (No: SPPH-LLBG -17-3.2). All animal use procedures and ethics were reviewed and approved by the Biomedical Ethics Committee of Health Science Center of Xi'an Jiaotong University (No: XJTUAE2023-110). All animal experiments were in accordance with the guide for the care and use of laboratory animals established by United States National Institutes of Health (Bethesda, MD, USA).

Acknowledgments

This work benefited from previous cohort studies. The author sincerely thanks all relevant researchers for the data shared and published.

Funding

This study was supported by Shaanxi Province Innovation Capability Support Plan (2024RS-CXTD-84), the Science and Technology Program of Xi'an (23YXYJ0186), and the Basic Natural Science Foundation of Shaanxi Province (2024JC-YBMS-654), Talent funding program of Shaanxi Provincial People's Hospital, International Science and Technology Cooperation Program Project of Key Research and Development Plan of Shaanxi Province (2020KWZ-20), the Key Projects of Shaanxi Provincial Department of Education (22JS035).

Disclosure

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Eldemire R, Mestroni L, Taylor M. Genetics of dilated cardiomyopathy. *Annu Rev Med*. 2024;75(1):417–426. doi:10.1146/annurev-med-052422-020535
2. Mahmaljy H, Yelamanchili VS, Singhal M Dilated cardiomyopathy. StatPearls Publishing Copyright© 2022, StatPearls Publishing LLC; 2024.
3. Jefferies JL, Towbin JA. Dilated cardiomyopathy. *Lancet*. 2010;375(9716):752–762. doi:10.1016/S0140-6736(09)62023-7
4. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. *Nat Biotechnol*. 2023;41(6):773–782. doi:10.1038/s41587-022-01448-2
5. Fan Y, Andrusivova Z, Wu Y, et al. Expansion spatial transcriptomics. *Nat Methods*. 2023;20(8):1179–1182. doi:10.1038/s41592-023-01911-1
6. Su M, Pan T, Chen QZ, et al. Data analysis guidelines for single-cell RNA-seq in biomedical studies and clinical applications. *Mil Med Res*. 2022;9(1):68. doi:10.1186/s40779-022-00434-8
7. Sweet ME, Cocciolo A, Slavov D, et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. *BMC Genomics*. 2018;19(1):812. doi:10.1186/s12864-018-5213-9
8. Alimadadi A, Munroe PB, Joe B, Cheng X. Meta-analysis of dilated cardiomyopathy using cardiac RNA-seq transcriptomic datasets. *Genes*. 2020;11(1):60. doi:10.3390/genes11010060
9. Cui YH, Wu CR, Xu D, Tang JG. Exploration of neuron heterogeneity in human heart failure with dilated cardiomyopathy through single-cell RNA sequencing analysis. *BMC Cardiovasc Disord*. 2024;24(1):86. doi:10.1186/s12872-024-03739-9
10. Zhu Y, Yang X, Zu Y. Integrated analysis of wgcna and machine learning identified diagnostic biomarkers in dilated cardiomyopathy with heart failure. *Front Cell Dev Biol*. 2022;10:1089915. doi:10.3389/fcell.2022.1089915
11. Wen L, Tang F. Recent advances in single-cell sequencing technologies. *Precis Clin Med*. 2022;5(1):pbac2. doi:10.1093/pcmedi/pbac002
12. Tan Z, Chen X, Zuo J, Fu S, Wang H, Wang J. Comprehensive analysis of scrna-seq and bulk RNA-seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model. *J Transl Med*. 2023;21(1):223. doi:10.1186/s12967-023-04056-z
13. Fang Z, Li J, Cao F, Li F. Integration of scrna-seq and bulk RNA-seq reveals molecular characterization of the immune microenvironment in acute pancreatitis. *Biomolecules*. 2022;13(1):78. doi:10.3390/biom13010078
14. Santra TS, Tseng FG. Single-cell analysis 2.0. *Cells*. 2022;12(1):154. doi:10.3390/cells12010154
15. Peng L, Yuan R, Han C, et al. Cellenboost: a boosting-based ligand-receptor interaction identification model for cell-to-cell communication inference. *IEEE Trans NanoBiosci*. 2023;22(4):705–715. doi:10.1109/TNB.2023.3278685
16. Wu T, Hu E, Xu S, et al. Clusterprofiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
17. Langfelder P, Horvath S. Wgcna: an r package for weighted correlation network analysis. *BMC Bioinf*. 2008;9(1):559. doi:10.1186/1471-2105-9-559
18. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. Gsea-p: a desktop application for gene set enrichment analysis. *Bioinformatics*. 2007;23(23):3251–3253. doi:10.1093/bioinformatics/btm369
19. Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. Ggtranscript: an r package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics*. 2022;38(15):3844–3846. doi:10.1093/bioinformatics/btac409
20. Haas GJ, Zareba KM, Ni H, Bello-Pardo E, Huggins GS, Hershberger RE. Validating an idiopathic dilated cardiomyopathy diagnosis using cardiovascular magnetic resonance: the dilated cardiomyopathy precision medicine study. *Circ Heart Fail*. 2022;15(5):e8877. doi:10.1161/CIRCHEARTFAILURE.121.008877
21. Mcgurk KA, Halliday BP. Dilated cardiomyopathy - details make the difference. *Eur J Heart Fail*. 2022;24(7):1197–1199. doi:10.1002/ejhf.2586

22. Sanbe A. Dilated cardiomyopathy: a disease of the myocardium. *Biol Pharm Bull*. 2013;36(1):18–22. doi:10.1248/bpb.b212023
23. Yue E, Yu Y, Wang X, Liu B, Bai Y, Yang B. Anthocyanin protects cardiac function and cardiac fibroblasts from high-glucose induced inflammation and myocardial fibrosis by inhibiting il-17. *Front Pharmacol*. 2020;11:593633. doi:10.3389/fphar.2020.593633
24. Ko T, Nomura S, Yamada S, et al. Cardiac fibroblasts regulate the development of heart failure via htra3-tgf-beta-igfbp7 axis. *Nat Commun*. 2022;13(1):3275. doi:10.1038/s41467-022-30630-y
25. Gu L, Jiang W, Zheng R, Yao Y, Ma G. Fibroblast growth factor 21 correlates with the prognosis of dilated cardiomyopathy. *Cardiology*. 2021;146(1):27–33. doi:10.1159/000509239
26. Shi X, Lin X, Huo L, Li X. Cardiac fibroblast activation in dilated cardiomyopathy detected by positron emission tomography. *J Nucl Cardiol*. 2022;29(2):881–884. doi:10.1007/s12350-020-02315-w
27. Son HJ, Choi EJ, Yoo NJ, Lee SH. Cancer-related gene mutations of aspn in colon cancers. *Pathol Res Pract*. 2020;216(11):153154. doi:10.1016/j.prp.2020.153154
28. Liu H, Mei Y, Ma X, Zhang X, Nie W. Frzb is regulated by the transcription factor egr1 and inhibits the growth and invasion of triple-negative breast cancer cells by regulating the jak/stat3 pathway. *Clin Breast Cancer*. 2022;22(7):690–698. doi:10.1016/j.clbc.2022.05.010
29. Yu P, He W, Zhang Y, et al. Sfrp4 is a potential biomarker for the prognosis and immunotherapy for gastric cancer. *J Oncol*. 2022;2022:8829649. doi:10.1155/2022/8829649
30. Kalamajski S, Aspberg A, Lindblom K, Heinegard D, Oldberg A. Asporin competes with decorin for collagen binding, binds calcium and promotes osteoblast collagen mineralization. *Biochem J*. 2009;423(1):53–59. doi:10.1042/BJ20090542
31. Yang Y, Liu P, Teng R, et al. Integrative bioinformatics analysis of potential therapeutic targets and immune infiltration characteristics in dilated cardiomyopathy. *Ann Transl Med*. 2022;10(6):348. doi:10.21037/atm-22-732
32. Song H, Chen S, Zhang T, et al. Integrated strategies of diverse feature selection methods identify aging-based reliable gene signatures for ischemic cardiomyopathy. *Front Mol Biosci*. 2022;9:805235. doi:10.3389/fmolb.2022.805235
33. Li XL, Yu F, Li BY, et al. The protective effects of grape seed procyanidin b2 against asporin mediates glycated low-density lipoprotein induced-cardiomyocyte apoptosis and fibrosis. *Cell Biol Int*. 2020;44(1):268–277. doi:10.1002/cbin.11229
34. Leyns L, Bouwmeester T, Kim SH, Piccolo S, De Robertis EM. Frzb-1 is a secreted antagonist of wnt signaling expressed in the spemann organizer. *Cell*. 1997;88(6):747–756. doi:10.1016/s0092-8674(00)81921-2
35. Fang C, Lv Z, Yu Z, et al. Exploration of dilated cardiomyopathy for biomarkers and immune microenvironment: evidence from RNA-seq. *BMC Cardiovasc Disord*. 2022;22(1):320. doi:10.1186/s12872-022-02759-7
36. Kumar V, Ramnarayanan K, Sundar R, et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. *Cancer Discov*. 2022;12(3):670–691. doi:10.1158/2159-8290.CD-21-0683
37. Guo Y, Ning B, Zhang Q, et al. Identification of hub diagnostic biomarkers and candidate therapeutic drugs in heart failure. *Int J Gen Med*. 2022;15:623–635. doi:10.2147/IJGM.S349235
38. Senyigit A, Uzun H, Gultepe I, Konukoglu D. The relationship between carotid intima-media thickness and serum secreted frizzled-related protein-4 and dipeptidyl peptidase-4 in diabetic patients with cardiovascular diseases. *Bratisl Lek Listy*. 2019;120(3):188–194. doi:10.4149/BLL_2019_032
39. Matsushima K, Suyama T, Takenaka C, et al. Secreted frizzled related protein 4 reduces fibrosis scar size and ameliorates cardiac function after ischemic injury. *Tissue Eng Part A*. 2010;16(11):3329–3341. doi:10.1089/ten.TEA.2009.0739
40. Guan H, Liu T, Liu M, Wang X, Shi T, Guo F. Sfrp4 reduces atherosclerosis plaque formation in apoe deficient mice. *Cardiol Res Pract*. 2023;2023:8302289. doi:10.1155/2023/8302289
41. Hung JH, Yang CM, Yang CH, Ho TC, Lai TT, Hsieh YT. Dilated choroidal vascular channels in pachychoroid neovasculopathy. *Clin Exp Ophthalmol*. 2023;51(1):44–57. doi:10.1111/ceo.14178
42. Meng XM, Liu SB, Deng T, et al. Loss of histone methyltransferase kmt2d attenuates angiogenesis in the ischemic heart by inhibiting the transcriptional activation of vegf-A. *J Cardiovasc Transl Res*. 2023;16(5):1032–1049. doi:10.1007/s12265-023-10373-x
43. Paradies P, Carlucci L, Woitek F, et al. Intracoronary gene delivery of the cytoprotective factor vascular endothelial growth factor-b(167) in canine patients with dilated cardiomyopathy: a short-term feasibility study. *Vet Sci*. 2019;6(1). doi:10.3390/vetsci6010023
44. Cai D, Wang X, Sun Y, et al. Patient-specific ipsc-derived cardiomyocytes reveal aberrant activation of wnt/beta-catenin signaling in scn5a-related Brugada syndrome. *Stem Cell Res Ther*. 2023;14(1):241. doi:10.1186/s13287-023-03477-3
45. Fan Y, Ho BX, Pang J, et al. Wnt/beta-catenin-mediated signaling re-activates proliferation of matured cardiomyocytes. *Stem Cell Res Ther*. 2018;9(1):338. doi:10.1186/s13287-018-1086-8
46. Lambert M, Jambon S, Depauw S, David-Cordonnier MH. Targeting transcription factors for cancer treatment. *Molecules*. 2018;23(6). doi:10.3390/molecules23061479