THE ANATOLIAN JOURNAL OF CARDIOLOGY

Official journal of the TURKISH SOCIETY OF CARDIOLOGY

Identification of Vital Genes Common to Dilated and Hypertrophic Cardiomyopathy Using Machine Learning and Bioinformatics

ABSTRACT

Background: Heart failure (HF) is a complex clinical syndrome and a leading cause of hospitalization and mortality worldwide, with significantly increasing mortality rates over time. Dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM) are 2 major myocardial diseases that can lead to progressive HF. Dilated cardiomyopathy is characterized by left ventricular dilation and systolic dysfunction, while HCM primarily presents with left ventricular wall thickening; both are closely associated with mutations in genes encoding cardiac sarcomere contractile proteins and exhibit genetic heterogeneity. Although previous studies have revealed alterations in transcription and protein expression in failing hearts, it may have neglected molecular changes in less frequent cell types.

Methods: The overlapping differentially expressed genes (DEGs) that overlapped between datasets GSE141910 and GSE249925 were identified, and a functional enrichment analysis was carried out. Overlapping genes in LASSO, SVM-RFE, and the key module in Weighted Gene Co-expression Network Analysis were seen as potentially significant. The receiver operating characteristic (ROC) curve was utilized to assess their potential for prognostic stratification. Immune cell infiltration was evaluated with cell-type identification by estimating relative subsets of RNA (CIBERSORT), and the findings were validated using the GSE165303 and GSE36961 datasets. Finally, the JASPAR database was utilized to predict potential transcription factors governing these key genes.

Results: One hundred eighty-three overlapping DEGs were identified in HCM and DCM, and the functional enrichment analysis revealed they were closely related to the inflammatory and immune response. The genes *FRZB*, *FNDC1*, and *SMOC2* were identified as potentially crucial, with ROC curves suggesting their potential utility for risk stratification in both test and validation datasets. The analysis using CIBERSORT revealed a relationship between immune cells that infiltrate and important potential genes. Secreted protein acidic and rich in cysteine-related modular calccium -binding protein 2 (SPARC-SMOC2) was identified as a hub gene with potential utility for patient stratification. Estrogen receptor 1 (ESR1) showed the strongest likelihood to be the regulator.

Conclusion: The research aimed at thoroughly understanding the pathogenesis of HCM and DCM. Discovering SMOC2 could help in understanding the shared mechanisms of HF, potentially serving as therapeutic targets in the future.

Keywords: Bioinformatics, DCM, HCM, machine learning, potential crucial genes

ORIGINAL INVESTIGATION

INTRODUCTION

Heart failure (HF) is a clinical condition characterized by structural and functional myocardial defects, leading to elevated intracardiac pressures and insufficient cardiac output at rest and during physical activity.[1-3] It is estimated that 64.3 million people worldwide are living with HF, with 1, 2, 5, and 10-year mortality rates of approximately 87%, 73%, 57%, and 35%, respectively.[4-6] Heart failure is a diverse clinical syndrome and a primary cause of hospitalizations and fatalities.[7,8]

Dilated cardiomyopathy (DCM) is characterized by left ventricular (LV) dilation and impaired systolic function, typically with preserved LV wall thickness.[9] This condition leads to progressive HF, declining contractility, and an increased risk of ventricular and supraventricular arrhythmias, conduction system abnormalities,

Yang Li
Ran Zhang

Department of Internal Medicine-Cardiovascular, The Second Hospital of Jiaxing, Jiaxing, China

Corresponding author:
Ran Zhang,
zhangran8190@163.com

Received: November 12, 2025
Accepted: July 23, 2026
Available Online Date: August 25, 2026

Cite this article as: Li Y, Zhang R. Identification of vital genes common to dilated and hypertrophic cardiomyopathy using machine learning and bioinformatics. *Anatol J Cardiol.* 2026;XX(X):X-X.

DOI: 10.14744/AnatolJCardiol.2026.5935

thromboembolism, and sudden cardiac or HF-related death.[10,11] As a common and largely irreversible form of myocardial disease, approximately 20%-35% of DCM cases are familial, showing incomplete and age-dependent penetrance and have been linked to over 20 genetic loci.[12,13] Although genetically heterogeneous, DCM is most commonly inherited in an autosomal dominant manner.[14,15] Many of the mutated genes associated with autosomal dominant DCM encode sarcomeric contractile proteins—such as α-cardiac actin, α-tropomyosin, cardiac troponins T, I, and C, β-, and α-myosin heavy chains, and myosin-binding protein C—which are also implicated in hypertrophic cardiomyopathy (HCM).[16]

Conversely, HCM is a heterogeneous disorder defined by LV wall thickening, often caused by mutations in sarcomeric protein genes.[17-19] Hypertrophic cardiomyopathy is 1 of the most prevalent inherited cardiomyopathies and is typically inherited in an autosomal dominant pattern.[20] It arises from diverse mutations in genes encoding cardiac sarcomere contractile proteins.[21] To date, 11 genes have been associated with HCM, most frequently β-myosin heavy chain—the first identified—and myosin-binding protein C.[22] This genetic diversity is accompanied by substantial intragenic heterogeneity, with over 400 distinct mutations identified.[23,24] The phenotypic variability of HCM is attributed not only to the specific disease-causing mutation but also to modifier genes and environmental influences.[19]

Given the complex etiology of HF, integrating molecular phenotyping with detailed clinical phenotypic data holds promise for uncovering key insights into disease mechanisms.[25-27] For instance, tissue-level analyses of RNA and protein expression have revealed disease-specific transcriptional signatures in cardiomyopathies.[28-31] While earlier studies have documented alterations in transcriptional and protein profiles in failing myocardium, many may have overlooked subtle yet significant molecular changes within less abundant cell populations.

In this study, gene expression profiles for DCM and HCM were sourced from the Gene Expression Omnibus (GEO) database. Potential key genes common to both conditions were subsequently identified through the application of Weighted Gene Co-expression Network Analysis (WGCNA) and machine learning techniques. The findings were then validated using independent datasets. The overarching objective was to enhance understanding of the transcriptional diversity of the human heart in both healthy and diseased states, thereby identifying novel therapeutic targets and biomarkers for prognostic stratification in HF.

HIGHLIGHTS

- A total of 183 common differentially expressed genes were identified in both hypertrophic and dilated cardiomyopathy.
- SMOC2 was characterized as a central hub gene with strong diagnostic value.
- ESR1 was proposed as a potential transcriptional regulator of SMOC2.

METHODS

Data Acquisition and Analysis

The GEO database was utilized to search for datasets using "HCM" and "DCM" as keywords, aiming to collect gene expression profiles and clinical data on diseases. The selection criteria are the following: (1) The organism must be *Homo sapiens*; (2) the tissue must originate from myocardial tissue; (3) the datasets should include both a disease group and a healthy control group; and (4) there must be more than 15 samples for WGCNA.

In the end, datasets GSE141910 and GSE249925 were utilized for analysis, while GSE165303 and GSE36961 were employed to verify the findings. Initially, data was downloaded from the GEO database.[32] Subsequently, low-expression probes across all samples were filtered out. The dataset was then subjected to log2 transformation, followed by inter-array normalization using the normalizeBetweenArrays function within the limma package.[33] Thereafter, limma vooma (variance-modelling at the observational level) was applied to estimate the mean−variance relationship and calculate appropriate observation-level weights. For probe annotation, probes mapping to multiple genes were excluded. When multiple probes corresponded to a single gene, only the probe with the maximum signal intensity was retained. Each dataset was analyzed independently. Overlapping genes that showed consistent directional changes across both datasets were retained as robust signatures.

Screening for Genes with Differential Expression

Using the "Limma" R package, differentially expressed genes (DEGs) were identified in the HCM and DCM datasets, with |Log2 fold change|≥1 and P value < .05 as the cut-off criteria. Genes that experienced simultaneous upregulation or downregulation in both datasets were labeled as overlapping. Principal Component Analysis (PCA) and Venn plots were applied to visualize the clustering of sample groups and the overlapping DEGs.

Functional Enrichment Analysis

To gain a deeper understanding of the roles of the overlapping genes, a functional enrichment analysis was performed. To convert Identity Documents (IDs) in molecular lists, the "Org. Hs. eg. Db" R package was applied. The "clusterProfiler" R package was used to perform enrichment analysis for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.[34] The "ggplot2" R package was used to create visual representations of the enrichment analysis results, focusing on the biological process (BP) in GO enrichment analysis. To explore biological pathways, KEGG pathway enrichment analysis was utilized. The cluster conducted functional enrichment analysis to examine gene biological functions and pathways. P value < .05 was regarded as the cut-off criterion.

An enrichment analysis of Reactome pathways was conducted by comparing cell types and diseases using 2 distinct approaches. Initially, a Gene Set Enrichment Analysis (GSEA)

analysis was excuted utilizing fgsea v1.16.044 on gene sets containing more than 15 but fewer than 500 genes. Utilizing CellBender count data, genes were ranked based on the t-statistic from the differential expression test and subsequently evaluated for enrichment in Reactome pathways from MsigDB. Secondly, ReactomePA was utilized to conduct a hypergeometric enrichment test to highlight significant changes. Genes with a Benjamini–Hochberg adjusted *P*-value less than .01 were selected based on CellRanger and CellBender counts and were split into upregulated or downregulated categories for the specified cell type. Taking into account the number of differentially expressed genes, this second technique permits us to underscore more extreme differential expression results and also offers a control for substantial differences between CellBender and CellRanger tests. Pathways that were enriched in both tests, with a Benjamini–Hochberg corrected 2-sided *P*-value of less than .05 for fgsea and a 1-sided *P*-value of less than .05 for ReactomePA, and showed consistent directionality, were deemed robust results.

Analysis of Gene Co-expression Networks Using Weighted Methods

The WGCNA analysis was conducted using the "WGCNA" R package. First, the expression matrix was obtained, and the 5000 genes with the greatest variability, as determined by the standard deviation of their expression data, were selected for analysis.[35,36] The fitted index, which showed a positive correlation, and the average connectivity, which showed a negative correlation, were plotted using a power scatter plot. To create a scale-free network, $\beta = 9$ was considered the best soft-power value. $R2 > 0.8$ was employed to determine the topological relationship of a scale-free network. Genes were clustered into modules based on their expression similarity using topological overlap matrix analysis. Highly interconnected gene clusters were established with a minimum module size of 30. Pearson correlation analysis was employed to explore the link between module traits and clinical features. In conclusion, genes with significance and module significance values greater than 0.8 were identified as hub genes in the modules.[37]

Employing Machine Learning to Discover Important Genes

LASSO and SVM-RFE, 2 machine learning techniques, were employed alongside WGCNA to identify important genes. The LASSO regression model aims to prevent overfitting by using L1 regularization to shrink the variable regression coefficients through a penalty function. The "glmnet" package in R was utilized for LASSO. Cross-validation plots were then created with Nfolds set to 10. SVM-RFE is a machine learning technique that automates and speeds up variable selection. The R packages "E1071," "kernlab," and "caret" were used for SVM-RFE. The genes were output after identifying the point with the lowest cross-validation error.

Measuring the Reliability of Diagnostic Conclusions

The "pROC" package is used to perform the receiver operating characteristic curve (ROC), also referred to as the sensitivity curve, by computing sensitivity and specificity at various cut-off points in a continuous variable. The area under the curve (AUC) was computed to demonstrate the diagnostic reliability of key genes on both the test and validation datasets. Ranging from 0 to 1, the AUC value reflects diagnostic confidence, with higher values indicating more certainty.

Immune Infiltration Analysis

Using the principle of linear support vector regression, CIBERSORT (https://cibersort.stanford.edu) analyzes and deconvolves the expression matrix of human immune cell subtypes. This technique offers a collection of gene expression values across 22 types of immune cells using a pre-processed reference dataset.[38] The current research set 1000 permutations as the default signature matrix, employing the "ggplot2" R package for visualization. Differences in immune cell infiltration between disease and control were depicted using heatmaps and violin plots. Afterward, the correlation between infiltrating immune cells and potential critical genes was determined through Spearman correlation analysis, where *P*-values under .05 were considered statistically significant. For genes associated with 3 or more infiltrating immune cells, lollipop graphs were used for visualization; otherwise, a correlation scatter plot was preferred.

Analysis of Gene Expression Specific to Cell Types

The most common cell types, including cardiomyocytes, fibroblasts, endothelial cells, pericytes, macrophages, Vascular Smooth Muscle Cells (VSMCs), and lymphocytes, underwent a targeted sub-clustering analysis. Clustering followed the global map approach, adjusting n_neighbors to 10 in sc.pp.neighbors to better reflect local structure. Furthermore, before calculating PCA, the percentage of mitochondrial reads and total Unique Molecular Identifier (UMI) counts per nucleus were removed. Leiden clustering for each cell type was conducted at resolutions ranging from 0.1 to 1.0, increasing by 0.1. Clustering was stopped and the earlier resolution was kept when a sub-cluster appeared without any marker genes having an AUC greater than 0.6. For VSMCs, these criteria were modified to demand at least 2 genes with an AUC over 0.6 to prevent excessive clustering. Marker genes were computed using a method analogous to that of the global cluster analysis. To identify marker genes, more relaxed criteria were applied, necessitating that the gene be expressed in at least 15% of nuclei within the target sub-cluster, have an AUC over 0.55, a positive log fold-change, and an False Discovery Rate (FDR)-adjusted 2-sided *P*-value below .01. As previously described, single-cell Comoositional Date-Analysis (scCODA) was used for composition analysis to identify subpopulations with different abundances in disease conditions. Using sc.tl.score_genes_cell_cycle() with default settings, the nuclei of macrophages were evaluated for their cell-cycle profiles.

Prediction of Transcription Factor

The University of California,Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/) was utilized to predict transcription factors (TFs) for the hub gene. Following this, correlation analyses involving the potential TFs and the targeted hub gene were performed in GSE165303 and GSE36961. Transcription factors that had correlation coefficient values

greater than 0.6, indicating a strong positive correlation, were chosen. To validate the relationship between the chosen TFs and the hub gene, multiple binding sites for the hub gene were forecasted using JASPAR (http://jaspar.genereg.net/).

RESULTS

The overall analytical workflow is summarized in Figure 1. Principal Component Analysis demonstrated clear separation between disease samples (DCM and HCM) and healthy controls, confirming distinct transcriptomic profiles (Figures 2A, B). Differential expression analysis identified 417 DEGs in DCM (GSE141910) and 899 DEGs in HCM (GSE249925) (Figures 2C, D). Intersection of DEGs from both datasets identified 183 overlapping genes, including 94 commonly upregulated and 89 commonly downregulated genes (Figures 2E, F).

Gene Ontology enrichment analysis revealed that the 183 overlapping genes were predominantly involved in inflammatory responses (acute inflammatory response, myeloid cell activation) and immune regulation (Figure 2G). Kyoto Encyclopedia of Genes and Genomes and reactome pathway analyses further identified enrichment in TGF-beta signaling, extracellular matrix (ECM) organization, collagen trimerization, and ECM proteoglycans—processes central to cardiac fibrosis and remodeling (Figures 2H and I). These findings point to shared pathological mechanisms—inflammation and fibrosis—that drive disease progression in both HCM and DCM.

The gene clustering dendrogram revealed that the WGCNA analysis identified 5 distinct modules within the GSE141910 dataset, with each module represented by a unique color (Figure 3A). Notably, the grey module exhibited the most significant positive correlation with DCM ($r = 0.92$, $P < 1e\text{-}134$), leading to the selection of 868 hub genes from this module for further analysis (Figures 3B, C).

LASSO regression identified 31 candidate genes, while SVM-RFE selected 48 genes (Figures 3D, E). The intersection of genes from both algorithms and WGCNA hub genes yielded 3 key genes: *FRZB*, *FNDC1*, and *SMOC2* (Figure 3F). The AUC values for *FRZB*, *FNDC1*, and *SMOC2* were 0.981, 0.989, and 0.989 in GSE141910, and 0.972, 0.964, and 0.988 in GSE249925, respectively (Figures 3G, H). Receiver operating characteristic analysis suggested the potential utility of these genes for patient stratification.

The heatmap displays the relative proportions of 22 immune cell types in DCM compared to control samples (Figure 4A). Violin plots depict the distributional differences of immune cells between the DCM and control groups (Figure 4B). In DCM samples, 11 immune cell types exhibited significant differences relative to controls. Specifically, the proportions of naïve B cells, M0 macrophages, M1 macrophages, resting mast cells, monocytes, plasma cells, CD8 T cells, and gamma delta T cells were significantly elevated. Conversely, the proportions of eosinophils, M2 macrophages, and resting memory CD4 T cells were reduced, demonstrating a negative correlation. Additionally, in DCM samples, *FRZB*, *SMOC2*, and *FNDC1* all showed significant positive correlations with M1 macrophages (Figures 4C-E).

The heatmap illustrates the relative proportions of 22 immune cell types in HCM versus control samples (Figure 5A). Violin plots depict the distributional differences of immune cells between the HCM and control groups (Figure 5B). In the HCM samples, 9 immune cell types exhibited significant differences relative to controls: the proportions of naïve B cells, resting mast cells, activated NK cells, resting NK cells, and CD8 T cells were significantly elevated. Conversely, the proportions of activated mast cells, monocytes, neutrophils,

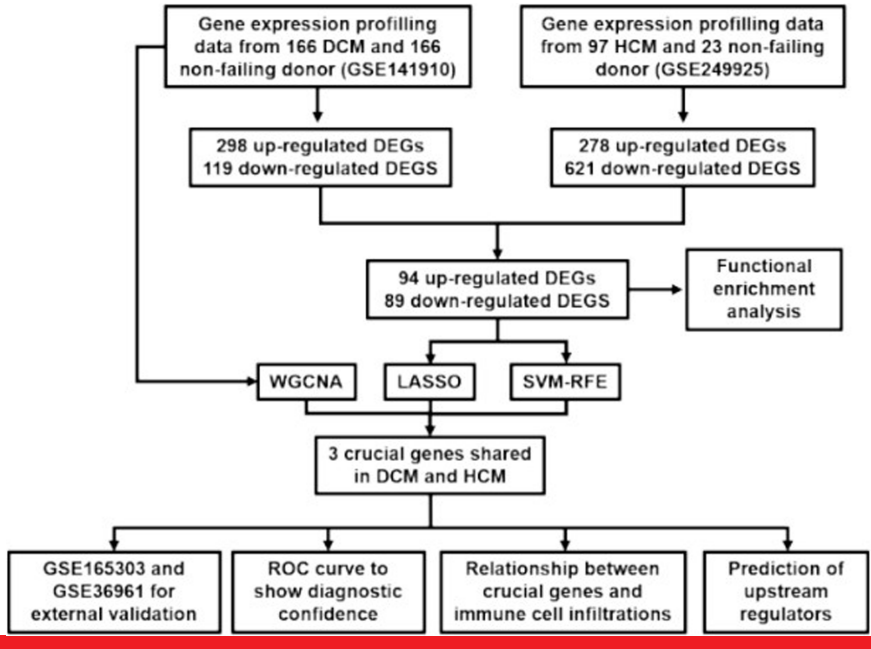

Figure 1. Workflow of process in this study.

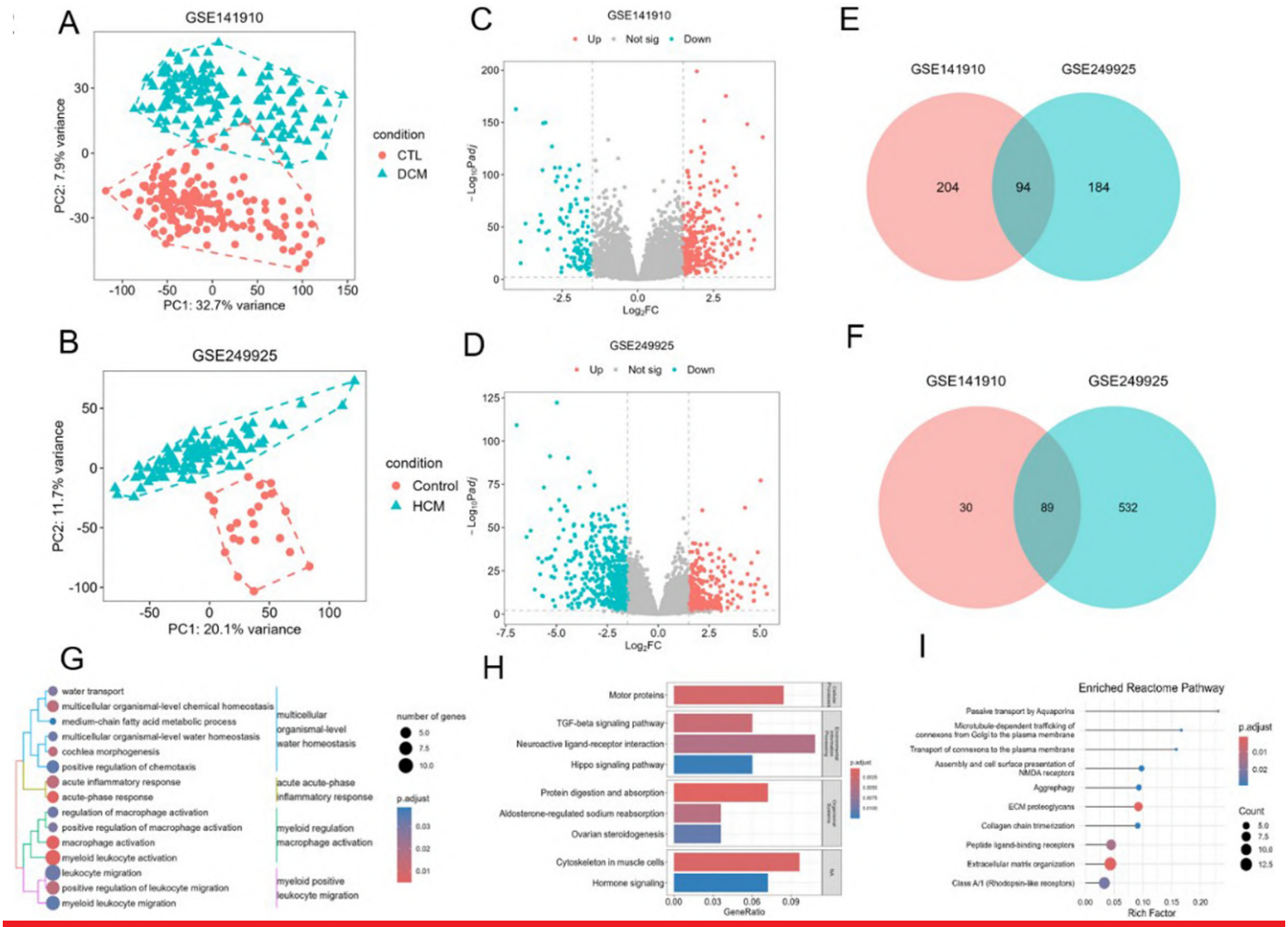

Figure 2. DEGs in GSE141910 and GSE249925. (A) The 2-dimensional PCA cluster diagram in GSE141910. (B) The 2-dimensional PCA cluster diagram in GSE249925. (C) The volcano plot of DEGs in GSE141910. (D) The volcano plot of DEGs in GSE249925. (E) The common up-regulated DEGs in GSE141910 and GSE249925. (F) The common down-regulated DEGs in GSE141910 and GSE249925. (G) The biological process (BP) of GO enrichment analysis of common DEGs shared in GSE141910 and GSE249925. (H) The KEGG pathway enrichment analysis of common DEGs shared in GSE141910 and GSE249925. (I) Enriched Reactome pathway analysis of common DEGs shared in GSE141910 and GSE249925. BP, biological process; DEGs, differentially expressed genes; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PCA, principal component analysis.

and resting CD4 memory T cells were reduced, demonstrating a negative correlation. Within the HCM samples, the genes *FRZB*, *FNDC1*, and *SMOC2* were found to be closely associated with immune cells. Specifically, *FRZB* showed correlations with 5 immune cell types, including resting mast cells and CD8 T cells (Figure 5C). *FNDC1* was correlated with 5 immune cell types, such as resting NK cells and CD8 T cells (Figure 5D). *SMOC2* exhibited positive correlations with resting mast cells and CD8 T cells, while showing negative correlations with activated mast cells and neutrophils (Figure 5E). Of note, these 3 hub genes did not exhibit disease-specific expression patterns capable of distinguishing HCM from DCM, consistent with their role as shared molecular markers of the failing heart.

To further validate the findings, external datasets GSE165303 and GSE36961 were analyzed. The analysis revealed that in both validation datasets, the expression levels of *FRZB*, *FNDC1*, and *SMOC2* were significantly elevated in DCM and HCM compared to controls (Figures 6A, C). Receiver operating characteristic curves further suggested the potential utility of these genes for patient stratification in the validation datasets (Figures 6B, D). Single-cell analysis using dataset SCP1303 revealed pronounced expression of *SMOC2* in activated fibroblasts and epicardial cells in HCM and DCM. Fibroblasts and epicardial cells are intricately linked with myocardial fibrosis, a critical pathological hallmark of both HCM and DCM. To elucidate the upstream regulatory mechanisms of *SMOC2*, upstream transcriptional regulators were predicted. Utilizing the UCSC database, it was predicted that the expression of 6 transcription factors—ESR1, ZNF610, GLIS1, DMC1, and E2F1—was significantly positively correlated with *SMOC2*. To enhance the validation of the regulatory interactions between these transcription factors and *SMOC2*, the JASPAR database was utilized to predict their binding motifs. The

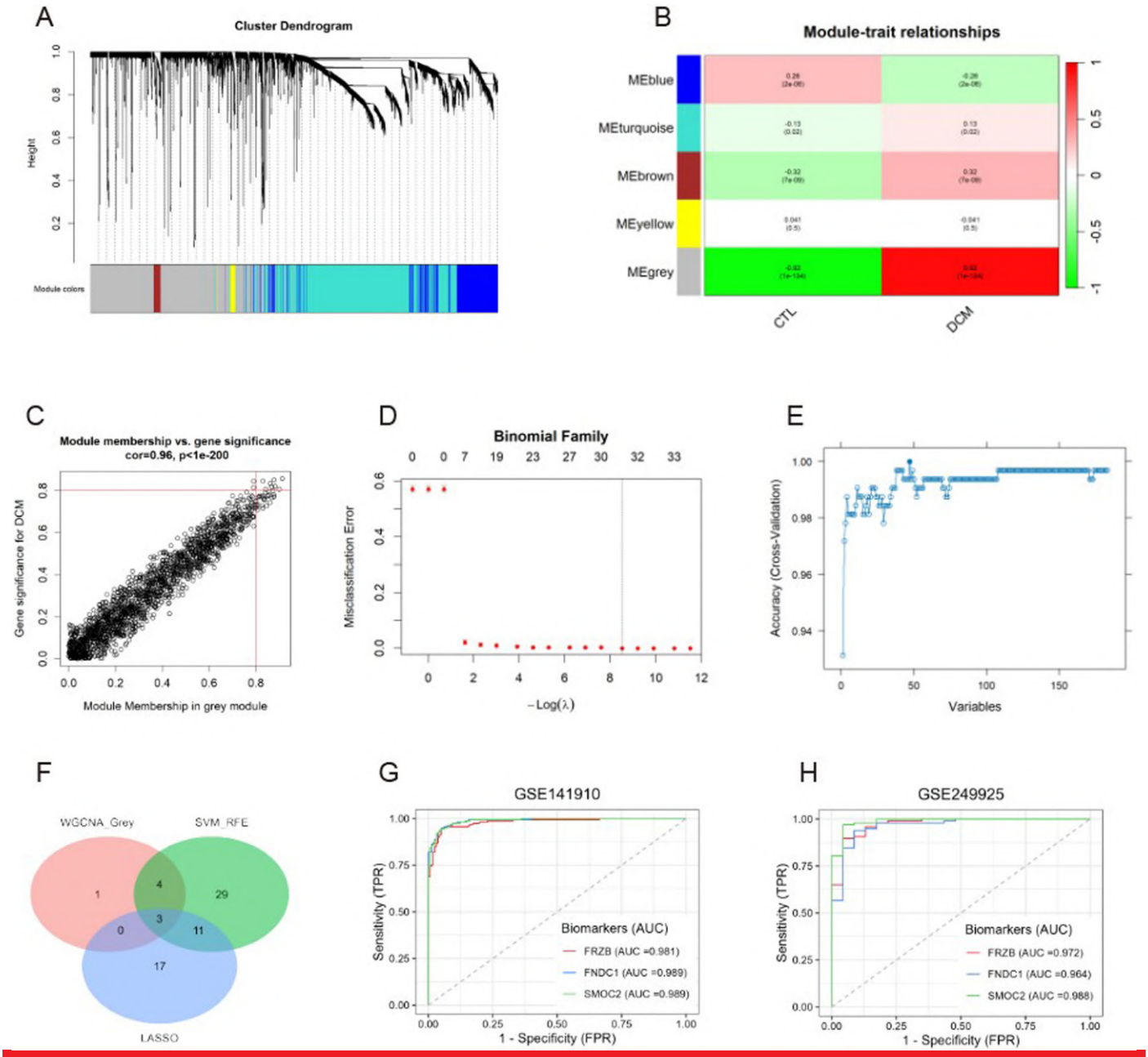

Figure 3. Identification of potential crucial genes by WGCNA, LASSO, and SVM-RFE. (A) The cluster dendrogram of genes in GSE141910. (B) Relationships between modules and clinical traits in GSE141910. (C) The relationship between module membership and gene significance in the grey module. (D) SVM-RFE algorithm for screening key genes in DEGs. (E) LASSO algorithm for determining key genes in DEGs. (F) Venn diagram to show the overlap of genes obtained using WGCNA, LASSO, and SVM-RFE. (G). ROC curve evaluating the potential of 3 hub genes for patient stratification in GSE141910. (H) ROC curve evaluating the potential of 3 hub genes for patient stratification in GSE249925. DEGs, differentially expressed genes; ROC, receiver operating characteristic; WGCNA, weighted gene co-expression network analysis.

motif with the highest score for ZNF610 was identified as MA1713.1 and MA1713.2, characterized by a core binding sequence of TATTACTAATA. For E2F1, the motif with the highest score was MA0024.2, also featuring the binding sequence TATTACTAATA. In the case of ESR1, the highest-scoring motif was MA0112.4, with a binding sequence of AATTTCACAAAAA.

DISCUSSION

Hypertrophic cardiomyopathy and DCM represent the 2 most prevalent and significant primary cardiomyopathies associated with the progression to HF.[39-42] Hypertrophic cardiomyopathy and DCM follow divergent paths toward HF. Dilated cardiomyopathy arises from progressive systolic

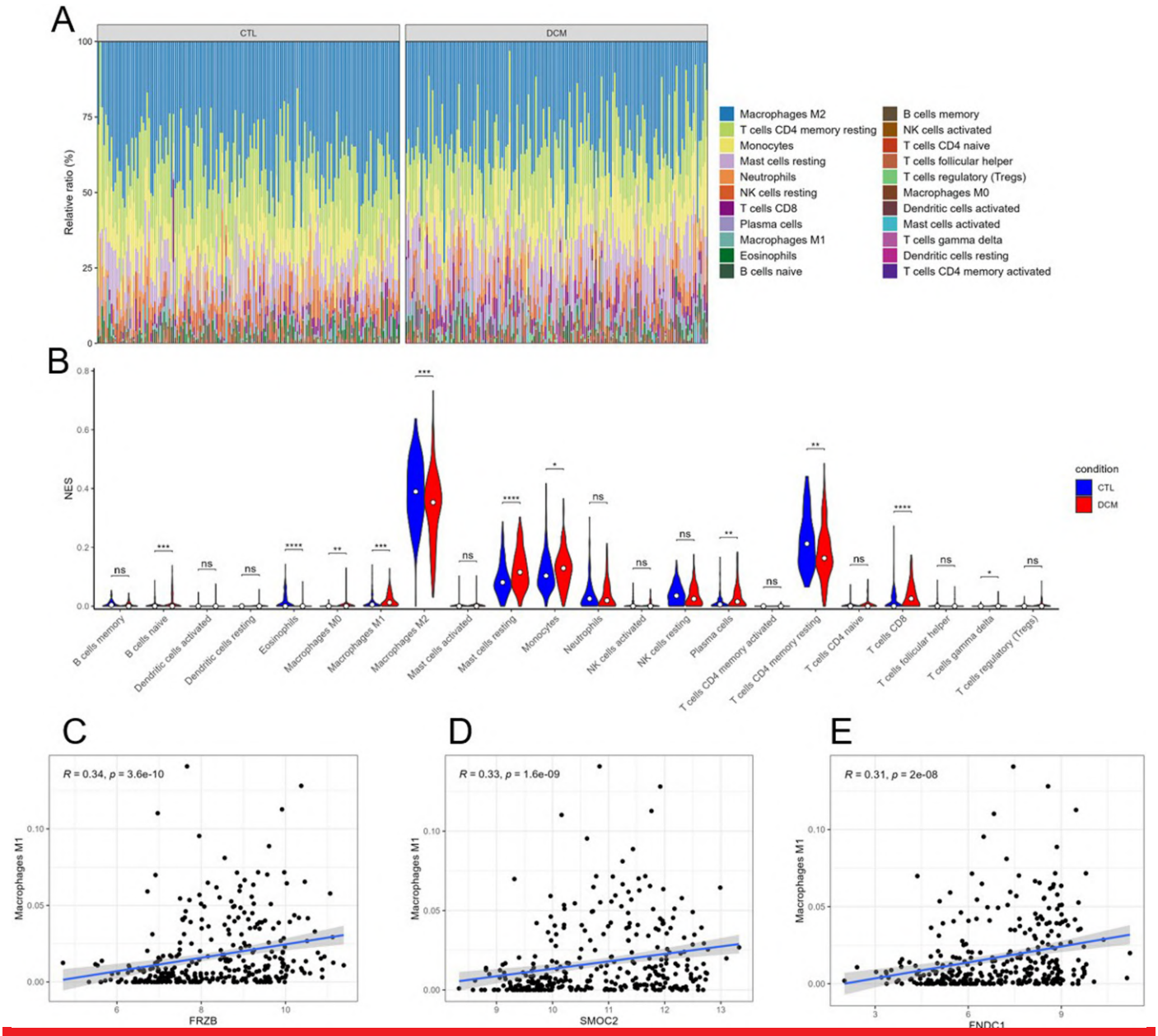

Figure 4. Immune infiltration analysis of crucial genes in DCM. (A) The heatmap of immune cell infiltration between DCM and controls. (B) The violin plot to show the different distribution of each immune cell between DCM and controls in GSE141910. (C) Correlation between FRZB and Macrophages M1. (D) Correlation between SMOC2 and Macrophages M1. (E) Correlation between FNDC1 and Macrophages M1. DCM, dilated cardiomyopathy. * $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$.

pump failure. Conversely, HCM is fundamentally a disease of impaired diastolic filling and compliance; its path to reduced cardiac output is primarily through decreased ventricular preload or outflow obstruction, not intrinsic pump weakness. It is noteworthy that in advanced HCM, a "burned-out" phenotype with systolic dysfunction may emerge, converging hemodynamically with DCM. Thus, despite different primary lesions, a deficit in effective cardiac output represents a common final pathway in both conditions.[43,44] The resultant deficit in cardiac output activates canonical compensatory pathways. These include (1) enhanced sympathetic nervous system activity, which leads to an increased heart rate and a temporary improvement in myocardial contractility[7,45]; and (2) activation of the Renin–Angiotensin–Aldosterone System, which promotes water and sodium retention (thereby increasing blood volume) and vasoconstriction (in an effort to maintain blood pressure).[46,47] While these compensatory mechanisms may offer short-term benefits, they become deleterious over the long term. Activation of the RAAS augments the heart's volume load, resulting in further dilation or increased wall stress. Continuous sympathetic activation elevates heart rate and myocardial oxygen consumption and may precipitate arrhythmias.[18,48] Together, these factors accelerate myocardial injury and functional deterioration, establishing a vicious cycle that culminates in overt HF.

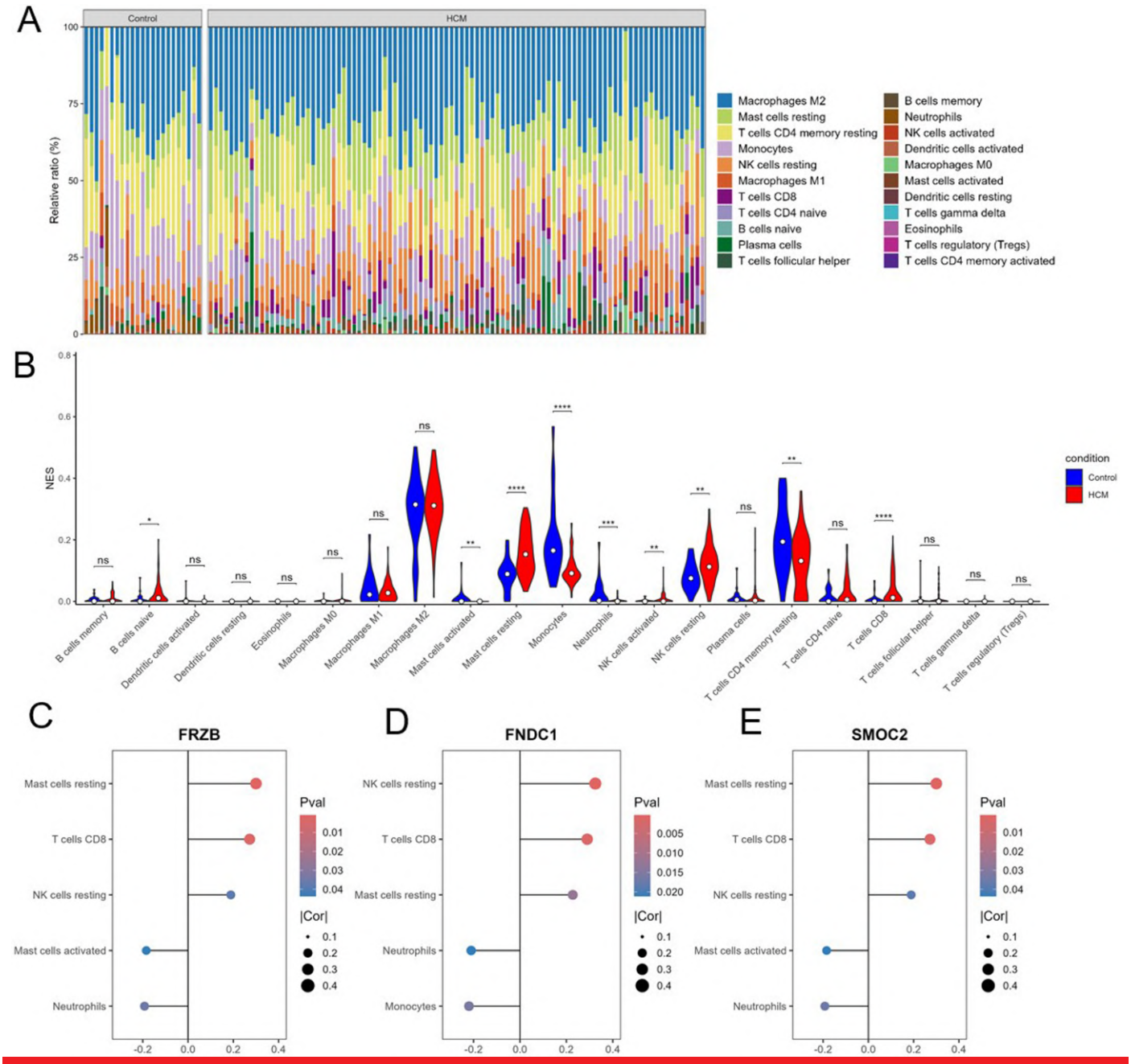

Figure 5. Immune infiltration analysis of crucial genes in HCM. (A) The heatmap of immune cell infiltration between HCM and controls. (B) The violin plot to show the different distribution of each immune cell between HCM and controls in GSE249925. (C) Correlation between FRZB and infiltrating immune cells. (D) Correlation between FNDC1 and infiltrating immune cells. (E) Correlation between SMOC2 and infiltrating immune cells. HCM, hypertrophic cardiomyopathy. * $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$.

Three central genes—*FRZB*, *FNDC1*, and *SMOC2*—were identified as consistently dysregulated in both HCM and DCM. These genes may serve as prognostic biomarkers rather than diagnostic tools to replace established clinical imaging modalities. However, further prospective validation is needed.

Importantly, these 3 hub genes do not differentiate HCM from DCM. Instead, they represent a shared molecular signature of the failing heart, reflecting convergent pathophysiological processes—particularly fibrosis and maladaptive remodeling—that mediate the progression of both diseases to a final common pathway of HF.

SMOC2 is a constituent of the cysteine-rich acidic secretory protein (SPARC) family, known for its role in matricellular functions.[49] Structurally, *SMOC2* comprises an extracellular calcium-binding (EC) domain and a characteristic domain containing 10 cysteines.[50,51] Functionally, *SMOC2* modulates cell-matrix interactions and influences the activity of

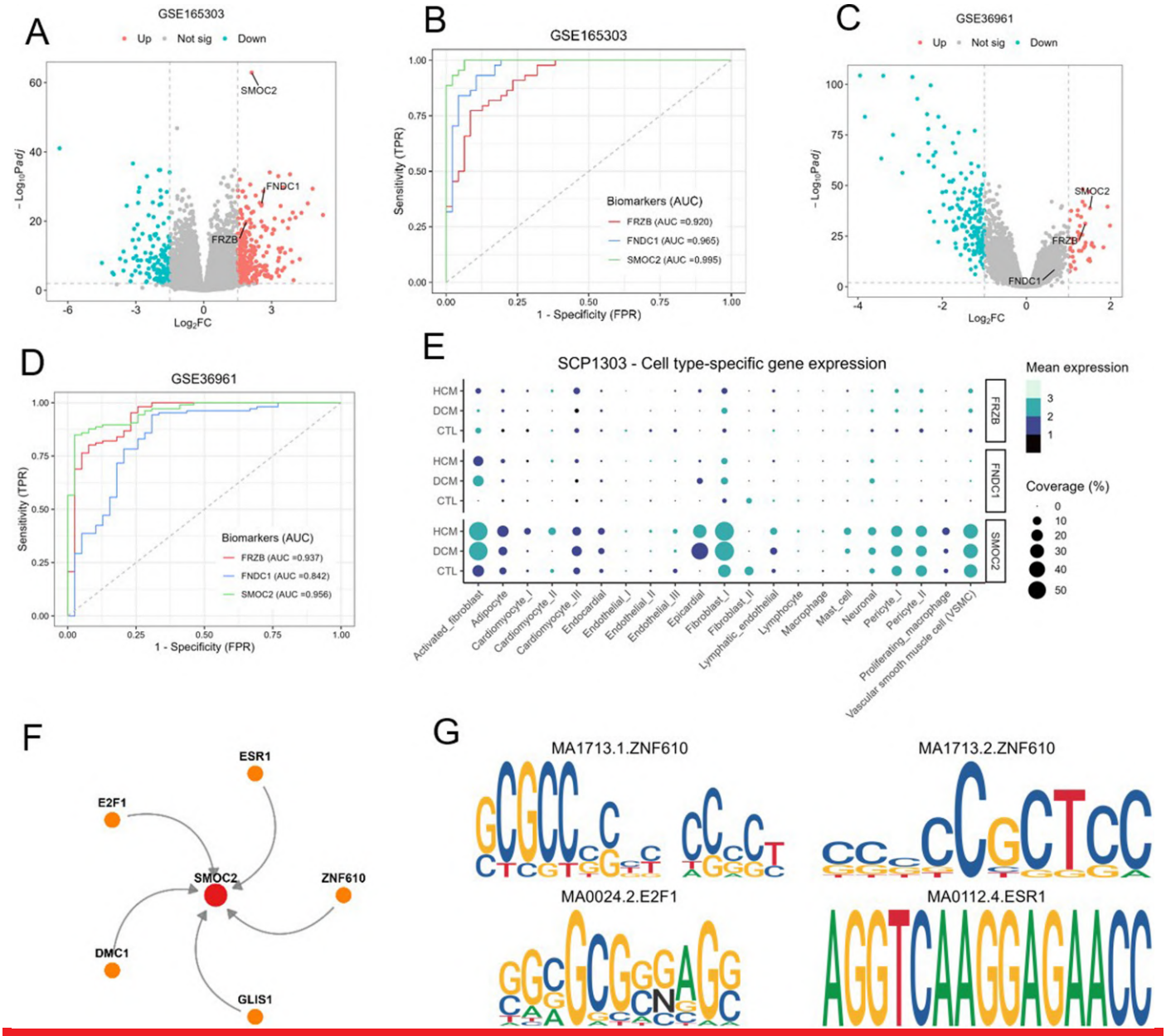

Figure 6. Validation of potential crucial genes and prediction of potential regulators. (A) The volcano plot of DEGs in GSE165303. (B) ROC curve evaluating the potential of 3 hub genes for patient stratification in GSE165303. (C) The volcano plot of DEGs in GSE36961. (D) ROC curve evaluating the potential of 3 hub genes for patient stratification in GSE36961. (E) Cell type-specific expression of FRZB, FNDC1, and SMOC2 in Control, HCM, and DCM samples. (F) The orange nodes in the circle represent 6 transcription factors regulating SMOC2. (G) Binding sites of 3 predicted transcription factors using JASPAR database. DCM, dilated cardiomyopathy; DEGs, differentially expressed genes; HCM, hypertrophic cardiomyopathy; ROC, receiver operating characteristic.

vascular growth factors through its interactions with matrix proteins, cell surface receptors, cytokines, proteases, and other effector molecules.[52,53] This protein has been identified as a potential biomarker for HF, and therapeutic strategies targeting *SMOC2* may offer promising avenues for the treatment of cardiac diseases. For instance, research by Wilk et al[32] demonstrated significant differences in *SMOC2* mRNA expression between non-HF and right HF groups in patients with pulmonary arterial hypertension (PAH) and in left/right ventricular samples from patients post-left ventricular assist device (LVAD) implantation, suggesting *SMOC2* as a viable target for HF treatment. Moreover, elevated plasma levels of *SMOC2* in clinical patients have been associated with an increased risk of rehospitalization in HF patients.[53] Plasma *SMOC2* levels correlate with the severity and progression of HF, serving as a reliable predictor of rehospitalization risk.[53] Mechanistically, experimental studies suggest that METTL3 enhances the m6A methylation of *SMOC2* mRNA, thereby increasing its stability. This post-transcriptional regulation promotes cardiac fibroblast proliferation

and differentiation—a process central to fibrotic remodeling in both HCM and DCM.[50] Emerging evidence further implicates SMOC2 in pathological cardiac remodeling through both fibrotic and inflammatory pathways. *SMOC2* binds to integrin αvβ5 on cardiac fibroblasts, inhibiting the LKB1/AMPKα/FOXO3 pathway, which impairs antioxidant defense, enhances lipid peroxidation, and promotes inflammation and fibrosis. These findings position *SMOC2* as a key mediator of maladaptive myocardial remodeling.[54]

FRZB has been identified as a potential HF marker. In a chronic animal model of repaired Tetralogy of Fallot (TOF), transcriptomic analyses revealed significant dysregulation of *FRZB* in ventricular tissues, pointing to its involvement in ventricular remodeling and cardiac failure.[55] In the context of hypertrophic cardiomyopathy, integrated bioinformatics analyses have identified *FRZB* as a hub gene in the key module showing high positive correlation with HCM, with functional enrichment analyses revealing that extracellular matrix, fibrosis, and neurohormone pathways play important roles.[56] These findings suggest that *FRZB* may contribute to the pathogenesis of both HCM and DCM.

FNDC1 (also known as *AGS8*) is an ischemia-inducible gene that promotes cardiomyocyte apoptosis under hypoxic stress by interacting with Gβγ subunits and connexin 43 (CX43), leading to increased cell permeability.[57-59] In HCM and DCM, where chronic ischemia and remodeling may occur, *FNDC1* may contribute to disease progression through its pro-apoptotic effects.

While the integrated bioinformatics approach has identified convergent molecular features, several limitations should be acknowledged. First, this study is based on retrospective analysis of tissue transcriptomic data. The immune cell infiltration and transcription factor predictions, while insightful, remain computational inferences derived from gene expression patterns and lack direct experimental corroboration. Future studies employing single-cell RNA sequencing, spatial transcriptomics, or flow cytometry on cardiac tissue would be invaluable to validate and spatially resolve the immune landscape that have been predicted. Second, the predicted regulatory relationships between transcription factors and the core genes require functional validation using techniques such as chromatin immunoprecipitation, luciferase reporter assays, or genetic manipulation in relevant cellular or animal models. Lastly, the prognostic and therapeutic potential of *SMOC2* needs to be prospectively validated in larger, multicenter cohorts and experimental models of cardiomyopathy.

In summary, through bioinformatics approaches, *SMOC2* has been identified and validated as a pivotal gene implicated in both HCM and DCM. *SMOC2* appears to play a significant role in the pathogenesis and progression of these conditions, indicating a shared underlying mechanism. This gene merits further investigation and holds potential as a prognostic biomarker.

CONCLUSION

In conclusion, this study aimed to achieve a comprehensive understanding of the pathogenesis of HCM and DCM, with the potential to elucidate mechanisms underlying HF. Utilizing machine learning and integrated bioinformatics methodologies, *SMOC2* was identified as a potential key gene. *SMOC2* contributes to the understanding of shared mechanisms between HCM and DCM and holds promise as a prognostic biomarker and prospective therapeutic target.

Data Availability: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Ethics Committee Approval: This study is a bioinformatic analysis utilizing public databases and does not require approval from an ethics committee.

Peer-review: Externally peer-reviewed.

Acknowledgments: This work was supported by the Jiaxing Municipal Science and Technology Program Project (2025CGW083).

Author Contributions: Concept – R.Z.; Design – R.Z.; Supervision – Y.L.; Data Collection and/or Processing – Y.L.; Analysis and/or Interpretation – Y.L.; Literature Search – Y.L.; Writing – R.Z.; Critical Review – R.Z.

Declaration of Interests: The authors have no conflict of interest to declare.

AI-Assisted Technologies Statement: No generative artificial intelligence (AI)-assisted technologies were used in the writing, editing, data analysis, or figure creation of this manuscript.

Funding: The authors declared that this study has received no financial support.

REFERENCES

1. Ciarambino T, Menna G, Sansone G, Giordano M. Cardiomyopathies: an overview. *Int J Mol Sci*. 2021;22(14):7722. [CrossRef]
2. Prokopidis K, Isanejad M, Akpan A, et al. Exercise and nutritional interventions on sarcopenia and frailty in heart failure: a narrative review of systematic reviews and meta-analyses. *ESC Heart Fail*. 2022;9(5):2787-2799. [CrossRef]
3. Williams JL, Cavus O, Loccoh EC, et al. Defining the molecular signatures of human right heart failure. *Life Sci*. 2018;196:118-126. [CrossRef]
4. McKenna WJ, Judge DP. Epidemiology of the inherited cardiomyopathies. *Nat Rev Cardiol*. 2021;18(1):22-36. [CrossRef]
5. McKenna WJ, Maron BJ, Thiene G. Classification, epidemiology, and global burden of cardiomyopathies. *Circ Res*. 2017;121(7):722-730. [CrossRef]
6. Wang C, Ba Y, Ni J, Huang R, Du X. Role of telemedicine intervention in the treatment of patients with chronic heart failure: a systematic review and meta-analysis. *Anatol J Cardiol*. 2024;28(4):177-186. [CrossRef]
7. Kemp CD, Conte JV. The pathophysiology of heart failure. *Cardiovasc Pathol*. 2012;21(5):365-371. [CrossRef]
8. Zuo C, Li X, Guan Y, et al. Influence of aging on outcomes of Sacubitril/Valsartan in hypertensive patients with heart failure: a multicenter retrospective study. *Anatol J Cardiol*. 2024;28(4):194-200. [CrossRef]
9. Zhou J, Ahmad F, Parikh S, et al. Loss of adult cardiac myocyte GSK-3 leads to mitotic catastrophe resulting in fatal dilated cardiomyopathy. *Circ Res*. 2016;118(8):1208-1222. [CrossRef]
10. Hamada T, Kubo T, Kitaoka H, et al. Clinical features of the dilated phase of hypertrophic cardiomyopathy in comparison

with those of dilated cardiomyopathy. *Clin Cardiol*. 2010;33(7):E24-E28. [CrossRef]
11. Chen Z, Li Y, Wang Y, et al. Cardiomyocyte-restricted low density lipoprotein receptor-related Protein 6 (LRP6) deletion leads to lethal dilated cardiomyopathy partly through Drp1 signaling. *Theranostics*. 2018;8(3):627-643. [CrossRef]
12. Knollmann BC, Roden DM. A genetic framework for improving arrhythmia therapy. *Nature*. 2008;451(7181):929-936. [CrossRef]
13. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat Rev Cardiol*. 2013;10(9):531-547. [CrossRef]
14. Vite A, Gandjbakhch E, Prost C, et al. Desmosomal cadherins are decreased in explanted arrhythmogenic right ventricular dysplasia/cardiomyopathy patient hearts. *PLOS One*. 2013;8(9):e75082. [CrossRef]
15. Adadi N, Radi FZ, Lahrouchi N, et al. Inherited dilated cardiomyopathy in a large Moroccan family caused by LMNA mutation. *Anatol J Cardiol*. 2018;20(1):65-68. [CrossRef]
16. Johnston JR, Landim-Vieira M, Marques MA, et al. The intrinsically disordered C terminus of troponin T binds to troponin C to modulate myocardial force generation. *J Biol Chem*. 2019;294(52):20054-20069. [CrossRef]
17. Okamoto R, Hirashiki A, Cheng XW, et al. Usefulness of serum cardiac troponins T and I to predict cardiac molecular changes and cardiac damage in patients with hypertrophic cardiomyopathy. *Int Heart J*. 2013;54(4):202-206. [CrossRef]
18. Katz AM, Rolett EL. Heart failure: when form fails to follow function. *Eur Heart J*. 2016;37(5):449-454. [CrossRef]
19. Marian AJ, Braunwald E. Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ Res*. 2017;121(7):749-770. [CrossRef]
20. Despond EA, Dawson JF. Classifying cardiac actin mutations associated with hypertrophic cardiomyopathy. *Front Physiol*. 2018;9:405. [CrossRef]
21. Raimoglou D, İzgi C, Enar R, et al. Structural and functional impact of Adrenoceptor Beta-1 gene polymorphism in patients with hypertrophic cardiomyopathy and response to beta-blocker therapy. *Anatol J Cardiol*. 2024;28(3):150-157. [CrossRef]
22. van Velzen HG, Schinkel AFL, Oldenburg RA, et al. Clinical characteristics and long-term outcome of hypertrophic cardiomyopathy in individuals with a MYBPC3 (myosin-binding protein C) founder mutation. *Circ Cardiovasc Genet*. 2017;10(4):e001660. [CrossRef]
23. Walsh R, Offerhaus JA, Tadros R, Bezzina CR. Minor hypertrophic cardiomyopathy genes, major insights into the genetics of cardiomyopathies. *Nat Rev Cardiol*. 2022;19(3):151-167. [CrossRef]
24. Spudich JA. Hypertrophic and dilated cardiomyopathy: four decades of basic research on muscle lead to potential therapeutic approaches to these devastating genetic diseases. *Biophys J*. 2014;106(6):1236-1249. [CrossRef]
25. Liu Y, Morley M, Brandimarto J, et al. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*. 2015;105(2):83-89. [CrossRef]
26. Verdonk K, Danser AHJ, van Esch JHM. Angiotensin II type 2 receptor agonists: where should they be applied? *Expert Opin Investig Drugs*. 2012;21(4):501-513. [CrossRef]
27. Pang J, Bao Y, Mitchell-Silbaugh K, Veevers J, Fang X. Barth syndrome cardiomyopathy: an update. *Genes (Basel)*. 2022;13(4):656. [CrossRef]
28. Maatz H, Jens M, Liss M, et al. RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J Clin Invest*. 2014;124(8):3419-3430. [CrossRef]
29. Jiang X, Chen Y, Liu X, et al. Uncovering inherited cardiomyopathy with human induced pluripotent stem cells. *Front Cell Dev Biol*. 2021;9:672039. [CrossRef]
30. Law ML, Cohen H, Martin AA, Angulski ABB, Metzger JM. Dysregulation of calcium handling in Duchenne muscular dystrophy-associated dilated cardiomyopathy: mechanisms and experimental therapeutic strategies. *J Clin Med*. 2020;9(2):520. [CrossRef]
31. Carreira IM. Molecular characterization of dilated cardiomyopathy. *Rev Port Cardiol (Engl Ed)*. 2019;38(2):141-142. [CrossRef]
32. Wilk JB, Herbert A, Shoemaker CM, Gottlieb DJ, Karamohamed S. Secreted modular calcium-binding protein 2 haplotypes are associated with pulmonary function. *Am J Respir Crit Care Med*. 2007;175(6):554-560. [CrossRef]
33. Smyth GK. limma: linear Models for microarray Data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, eds. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer New York; 2005:397-420. [CrossRef]
34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16(5):284-287. [CrossRef]
35. Zhang T, Wong G. Gene expression data analysis using Hellinger correlation in weighted gene co-expression networks (WGCNA). *Comput Struct Biotechnol J*. 2022;20:3851-3863. [CrossRef]
36. Zhang X, Feng H, Li Z, et al. Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis. *Onco Targets Ther*. 2018;11:6001-6021. [CrossRef]
37. Zhang D, Hao W, Niu Q, Xu D, Duan X. Identification of the co-differentially expressed hub genes involved in the endogenous protective mechanism against ventilator-induced diaphragm dysfunction. *Skelet Muscle*. 2022;12(1):21. [CrossRef]
38. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457. [CrossRef]
39. Malgija B, Kumar NS, Piramanayagam S. Collective transcriptomic deregulation of hypertrophic and dilated cardiomyopathy - Importance of fibrotic mechanism in heart failure. *Comput Biol Chem*. 2018;73:85-94. [CrossRef]
40. Marian AJ, Roberts R. Molecular basis of hypertrophic and dilated cardiomyopathy. *Tex Heart Inst J*. 1994;21(1):6-15.
41. Alpert NR, Warshaw DM. Human heart failure: dilated versus familial hypertrophic cardiomyopathy. *Adv Exp Med Biol*. 2003;538:77-87; discussion 87-88. [CrossRef]
42. Chaffin M, Papangeli I, Simonson B, et al. Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature*. 2022;608(7921):174-180. [CrossRef]
43. El Hadi H, Freund A, Desch S, Thiele H, Majunke N. Hypertrophic, dilated, and arrhythmogenic cardiomyopathy: where are we? *Biomedicines*. 2023;11(2):524. [CrossRef]
44. Friedrich FW, Carrier L. Genetics of hypertrophic and dilated cardiomyopathy. *Curr Pharm Biotechnol*. 2012;13(13):2467-2476. [CrossRef]
45. Polovina M, Tschöpe C, Rosano G, et al. Incidence, risk assessment and prevention of sudden cardiac death in cardiomyopathies. *Eur J Heart Fail*. 2023;25(12):2144-2163. [CrossRef]
46. Schwinger RHG. Pathophysiology of heart failure. *Cardiovasc Diagn Ther*. 2021;11(1):263-276. [CrossRef]
47. Tanai E, Frantz S. Pathophysiology of heart failure. *Compr Physiol*. 2015;6(1):187-214. [CrossRef]
48. Zelis R, Mason DT. Compensatory mechanisms in congestive heart failure—the role of the peripheral resistance vessels. *N Engl J Med*. 1970;282(17):962-964. [CrossRef]

49. Ren Y, Wu Y, He W, Tian Y, Zhao X. SMOC2 plays a role in heart failure via regulating TGF-beta1/Smad3 pathway-mediated autophagy. *Open Med (Wars)*. 2023;18(1):20230752. [CrossRef]
50. Rui H, Zhao F, Yuhua L, Hong J. Suppression of SMOC2 alleviates myocardial fibrosis via the ILK/p38 pathway. *Front Cardiovasc Med*. 2022;9:951704. [CrossRef]
51. He Y, Pan X, Liu Z, et al. METTL3 silencing suppresses cardiac fibrosis post myocardial infarction via m6A modification of SMOC2. *J Cell Mol Med*. 2025;29(17):e70829. [CrossRef]
52. Wang X, Wang M, Zhou Z, et al. SMOC2 promoted vascular smooth muscle cell proliferation, migration, and extracellular matrix degradation by activating BMP/TGF-β1 signaling pathway. *J Clin Biochem Nutr*. 2023;73(2):116-123. [CrossRef]
53. Chen X, Zhong X, Luo D, Lei Y, Huang R. Plasma SMOC2 predicts prognosis in patients with heart failure: a prospective cohort. *Int J Gen Med*. 2024;17:1651-1664. [CrossRef]
54. Yun X, Ling Y, Gaoyuan L, et al. SMOC2 accelerates myocardial fibrosis following myocardial infarction by promoting lipid peroxidation through inhibition of the LKB1/AMPKα/FOXO3 pathway. *J Adv Res*. 2026:S2090-1232(26)00241-9. [CrossRef]
55. Charron S, Roubertie F, Benoist D, et al. Identification of region-specific myocardial gene expression patterns in a chronic swine model of repaired tetralogy of Fallot. *PLOS One*. 2015;10(8):e0134146. [CrossRef]
56. Ma Z, Wang X, Lv Q, et al. Identification of underlying hub genes associated with hypertrophic cardiomyopathy by integrated bioinformatics analysis. *Pharmgenomics Pers Med*. 2021;14:823-837. [CrossRef]
57. Sato M, Jiao Q, Honda T, et al. Activator of G protein signaling 8 (AGS8) is required for hypoxia-induced apoptosis of cardiomyocytes: role of G betagamma and connexin 43 (Cx43). *J Biol Chem*. 2009;284(45):31431-31440. [CrossRef]
58. Sato M, Cismowski MJ, Toyota E, et al. Identification of a receptor-independent activator of G protein signaling (AGS8) in ischemic heart and its interaction with Gbetagamma. *Proc Natl Acad Sci U S A*. 2006;103(3):797-802. [CrossRef]
59. Davis S, Meltzer PS. GE. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846-1847. [CrossRef]