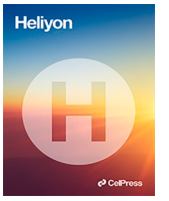
Heliyon 10 (2024) e25572

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Targeting *ERBB2* and *PIK3R1* as a therapeutic strategy for dilated cardiomyopathy: A single-cell sequencing and mendelian randomization analysis

Bin He [a], Liping Quan [a], Chengban Li [a], Wei Yan [a,b], ZhuoHua Zhang [b], LiuFan Zhou [b], Qinjiang Wei [b], Zhile Li [b], Jianjiao Mo [b], Zhen Zhang [b], Xingshou Pan [b], JianJun Huang [c,d,**], Li Liu [b,e,*]

[a] *Graduate School of Youjiang Medical University for Nationalities, Baise, China*
[b] *Department of Cardiology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China*
[c] *College of Clinical Medicine, Youjiang Medical University for Nationalities, Baise, China*
[d] *Department of Neurology, Affiliated Hospital of Youjiang Medical University for Nationalities, Youjiang Medical University for Nationalities, Baise, 533000, Guangxi, China*
[e] *The Key Laboratory for High Incidence Prevention and Treatment in Guangxi Guixi Area, Baise, 533000, Youjiang Medical University for Nationalities, Baise, 533000, Guangxi, China*

ARTICLE INFO

Keywords:
Dilated cardiomyopathy
Single-cell RNA sequencing
Weight gene co-expression network analysis (WGCNA)
Mendelian randomization analysis (MR)
Immune cell infiltration

ABSTRACT

Background: Dilated cardiomyopathy (DCM) is widely recognized as a significant contributor to heart failure. Nevertheless, the absence of pharmaceutical interventions capable of reversing disease progression and improving prognosis underscores the imperative for additional research in this area.

Methods: First, we identified and evaluated three gene sets, namely "SC-DCM", "EP-DCM" and "Drug", using big data and multiple bioinformatics analysis methods. Accordingly, drug-treatable ("Hub") genes in DCM were identified. Following this, four microarray expression profile datasets were employed to authenticate the expression levels and discriminatory efficacy of "Hub" genes. Additionally, mendelian randomization analysis was conducted to ascertain the causal association between the "Hub genes" and heart failure. Finally, the "DGIdb" was applied to identify "Hub" genes-targeted drugs. The "ssGSEA" algorithm assessed the level of immune cell infiltration in DCM.

Results: Enrichment analysis showed that the "SC-DCM" and "EP-DCM" gene sets were closely associated with DCM. *PIK3R1* and *ERBB2* were identified as drug-treatable genes in DCM. Additional analysis using MR supported a causal relationship between *ERBB2* and heart failure, but not *PIK3R1*. Moreover, *PIK3R1* was positively correlated with immune activation, while *ERBB2* was negatively correlated. We found that everolimus was a pharmacological inhibitor for both *PIK3R1* and *ERBB2*. However, no pharmacological agonist was found for *ERBB2*.

Conclusion: PIK3R1 and ERBB2 are drug-treatable genes in DCM. *ERBB2* has a causal effect on heart failure, and its normal expression may play a role in preventing the progression of DCM to heart failure. In addition, there is a cross-expression of *PIK3R1* and *ERBB2* genes in both DCM and

* Corresponding author. Department of Cardiology, Affiliated Hospital of Youjiang Medical University for Nationalities Baise, 533000, China.
** Corresponding author. College of Clinical Medicine, Youjiang Medical University for Nationalities, Baise, 533000, China.
E-mail addresses: jianjun453@163.com (J. Huang), liuli011258@sina.com (L. Liu).

https://doi.org/10.1016/j.heliyon.2024.e25572
Received 19 July 2023; Received in revised form 13 January 2024; Accepted 29 January 2024
Available online 13 February 2024

tumors. The adaptive immune system and *PIK3R1* may be involved in DCM disease progression, while *ERBB2* exerts a protective effect against DCM.

1. Introduction

Dilated cardiomyopathy (DCM) is a genetically heterogeneous disease characterized by enlarged ventricles and decreased myocardial function. Its etiology and pathogenesis include genetic mutations leading to abnormal cell structure and function, abnormal signaling pathways conducive to muscle contractile force production or transmission defects, and inflammatory responses leading to ventricular/myocardial remodeling [1,2]. Given the complex and diverse etiology and pathogenesis of DCM, there is still a lack of specific treatment options that address the etiology and pathogenesis of the disease, and most are mainly symptomatic [3]. Although current therapeutic approaches can improve patients' clinical symptoms and enhance their quality of life and survival, these treatments, including drugs and adjuvant devices, are often accompanied by frequent adverse reactions. These limitations greatly hinder their clinical application. In the face of the heavy medical burden and treatment limitations, developing new and effective therapeutic agents for DCM remains an urgent need.

It is well-established that the development of DCM involves mutations in multiple susceptibility genes. This understanding has been facilitated by the development of high-throughput sequencing technologies such as genome-wide association studies (GWAS), transcriptome (bulk RNA-seq) sequencing, and whole-exome sequencing (WES) [4–6]. The mapping of DCM genetic variants and causative genes at the tissue level has provided novel insights into DCM pathogenesis. Interestingly, recent cellular mapping studies of the normal human heart have shown that cardiac tissue is usually composed of multiple cell types, including cardiomyocytes, smooth muscle cells, endothelial cells, and fibroblasts. In addition, multiple immune cells are activated during the pathogenesis [7]. In recent years, single-cell RNA sequencing (scRNA-seq) technology has been widely used in scientific research to identify specific pathogenic cell types and pathogenic genes [8]. In the present study, we innovatively combined single-cell sequencing data, microarray expression profiling data and drug targets to explore potentially pathogenic and drug-treatable genes in DCM, providing the foothold for developing novel therapeutic agents for DCM at the genetic levels.

This study incorporates the utilization of Curcumin (Cur), an herbal medicine known for its diverse drug targets and potential role in safeguarding against cardiovascular disease. Cur is derived from the rhizome of turmeric and is classified as a polyphenolic compound. Its efficacy in mitigating cardiovascular ailments has been extensively investigated, particularly in relation to its hypolipidemic, antioxidant stress, anti-inflammatory, anti-apoptotic, and anti-myocardial fibrosis properties [9,10]. For example, curcumin can act as an inhibitor of histone acetyltransferase p300 by inhibiting acetylation of a key transcription factor (GATA4) for cardiac development and pathological changes, thereby inhibiting cardiac hypertrophy and heart failure [11,12]. In addition, curcumin inhibits apoptosis-related protein expression by reducing the activation of p53 in diseased myocardium, ultimately inhibiting cardiomyocyte apoptosis [13]. Interestingly, Xiao-Jie Bai et al. showed that activated Na+/Ca^{2+}exchanger (NCX) might be a downstream target of the cardioprotective effect of curcumin, which can inhibit myocardial hypertrophy and improve cardiac function by upregulating NCX expression [14]. These studies suggest that curcumin is involved in the protection against cardiovascular diseases through abundant biological activities and mediating multiple signaling pathways. Herein, we chose curcumin to identify potentially drug-treatable disease-causing genes in DCM.

Mendelian randomization (MR) studies utilize the random assignment of alleles during the transmission from parent to offspring, enabling the establishment of "natural randomized controlled studies." This approach employs genetic variation as an instrumental variable to estimate causality in the association between risk factors and disease outcomes [15]. Consequently, MR studies offer a means to circumvent the prevalent biases arising from confounding factors and reverse causality in observational studies. Notably, the proliferation of large-scale GWAS studies in recent years has further expanded the scope for MR analysis. In this study, we aim to investigate key pathogenic genes in DCM by utilizing single-cell sequencing data and MR analysis. To the best of our knowledge, this study represents a groundbreaking approach in the field. Our methodology involves merging multiple microarray expression profile data of DCM and conducting differential analysis and weighted gene co-expression network analysis (WGCNA) to effectively identify genes that are closely linked to DCM. Furthermore, we employ a comprehensive screening of multiple drug databases to identify potential drug targets. Finally, we combined single-cell sequencing data, microarray expression profile data, drug target data, and large GWAS cohorts and performed various bioinformatics analysis methods such as MR analysis, protein-protein interaction network analysis (PPI), receiver-operating characteristic (ROC) curves, "ssGSEA" immune infiltration analysis, "spearman" correlation analysis, "singscore" and "ssGSEA" scoring, to comprehensively explore pathogenic and drug-treatable genes and immune infiltration pathways in DCM.

2. Materials and methods

2.1. Data sources and presentations

The single-cell dataset GSE184899 and microarray expression datasets (GSE84796 [16], GSE29819 [17], GSE42955 [18], and GSE120895 [19]) related to DCM were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) (**Supplementary file1**). The single-cell dataset of GSE184899 contained day 35 induced pluripotent stem cell (iPSC)-CMs from a congenital dilated cardiomyopathy iPSC line and a control iPSC line, based on the GPL24676 platform. The microarray

expression datasets of GSE84796 and GSE29819 were merged to obtain the training group, and four datasets, GSE84796, GSE29819, GSE42955 and GSE120895, were merged as the validation group (**Supplementary file2**). The training dataset GSE84796 based on the GPL14550 platform contained 10 cases of DCM and 7 controls. In addition, dataset GSE29819, based on the GPL570 platform, contained 12 arrhythmogenic cardiomyopathy (ARVC), 12 DCM, and 12 controls. In the validation group, dataset GSE42955, based on the GPL570 platform, contained 12 DCM, 12 ICM, 5 controls. Another dataset, GSE120895, based on the GPL570 platform, contained 47 DCM and 8 controls. All controls included in the microarray expression datasets were derived from healthy hearts of organ donors or normal hearts. The 12 ARVC samples of GSE29819 and 12 ICM samples of GSE42955 were removed. The merged training and validation datasets were subdivided into Treat (DCM) and Con (non-DCM) groups.

2.2. Single-cell data analysis and annotation

The "Seurat" R software package (version 4.2.0) was used to analyze single-cell data for further dimensionality reduction, cell clustering and annotation [20]. First, single-cell data were filtered by the following criteria: each gene was expressed in at least 3 cells; The number of genes expressed per cell ranged from 200< nFeature_RNA <10000; Mitochondrial and ribosomal content were less than 20%. Next, the "LogNormalize", "ScaleData" and "Harmony" functions were used to normalize the data and remove batch effects. Subsequently, the preprocessed single-cell data were further identified for the top 2000 highly variable genes (HVGs) using the "vst" method. The "PCA" function was used to transform the data into linear dimension and to identify the principal components (PC). Different clusters were defined using the "FindNeighbors" and "FindClusters" functions (with a resolution of 0.8). Subsequently, 20 PC were selected for t-SNE and UMAP analysis [21]. Based on the filtering criteria logFC.filter >0.5 and adjPvalFilter <0.05, the "FindAllMarkers" function was used to filter the significant markers genes in different clusters. Finally, the "SingleR" package was used to annotate the different cell clusters [22]. The set of marker genes in a specified cluster was defined as the "SC-DCM" gene set for subsequent analysis.

2.3Differential analysis and weighted gene co-expression network analysis (WGCNA) of microarray expression data

The expression profiles data of the training group (datasets GSE84796 and GSE29819) were merged and normalized by the R software packages "sva" and "limma" [23]. The "combat" function was applied to remove batch effects between the two datasets [24]. First, the merged and preprocessed data were screened for differentially expressed genes (DEGs) with filtering criteria of log fold change (logFC) > 1 and adjusted P-value <0.05. The top 50 DEGs were visualized by heatmaps and volcano plots using the "pheatmap" and "ggplot2" packages [25]. Next, the merged and preprocessed data were further used to construct an optimal scale-free network (ideal soft threshold $\beta = 6$) by the "goodSampleGenes" and "pickSoftThreshold" functions of the R package "WGCNA" [26]. Subsequently, the network was transformed into a topological overlap matrix for clustering and dimensionality reduction of the data. Subsequently, the different modules and the relationship between modules and clinical traits were identified by the dynamic shearing module and the module eigengenes (MEs). Based on the screening criteria, Module membership (MM) > 0.8 and gene significance (GS) > 0.5, marker genes in modules with significant correlation with the phenotype were identified. Finally, the marker genes in the module and the DEGs were intersected using the "VENN" package. The overlapping genes were defined as "EP-DCM" gene sets for subsequent analysis.

2.4. Drug target acquisition

The Comparative Toxicogenomics Database (CTD, http://ctdbase.org/), SwissTargetPrediction database (SwissTargetPrediction, http://www.swisstargetprediction.ch/), Binding Database (BindingDB, http://bindingdb.org/bind/index.jsp), and Drug-target interactions database (TargetNet, http://targetnet.scbdd.com/home/index/) were searched using the keyword "Curcumin" and the "SMILES" number of curcumin, the curcumin target genes and proteins were identified. The screening conditions of each database were as follows: drug-target interaction "Interactions" > 1 in CTD database, "Probability" > 0 in SwissTargetPrediction database and "Prob" > 0.01 in TargetNet database. It is worth noting that the drug target prediction results for these databases were derived from high-evidence experimental studies and relevant literature. In addition, the predicted drug target proteins in the BindingDB and TargetNet databases were converted to their corresponding gene names by the Uniprot database (https://www.uniprot.org/). Finally, all predicted target genes were combined, and duplicate genes were removed using the R package "venn". The merged genes were defined as the "Drug" gene set for subsequent analysis.

2.5. Functional enrichment analysis

In this study, the defined three gene sets ("SC-DCM", "EP-DCM" and "Drug") underwent KEGG and GO enrichment analysis using the R packages "clusterProfiler" and "enrichplot" [27]. The "ggplot2" package was used to visualize the results. In addition, GSEA analysis was conducted on the logFC-ranked DEGs based on the KEGG gene set acquired from the Molecular Signature Database (MsigDB) [28]. The top five significantly enriched pathways were visualized. Adjusted P-values <0.05 were statistically significant.

2.6. Protein-protein interaction (PPI) network construction and analysis

Three different combinations of gene sets ("Drug" and "EP-DCM", "Drug" and "SC-DCM", "Drug" and "EP-DCM" and "SC-DCM")

were imported into the STRING database(https://string-db.org/) to construct the PPI interaction network. The filtering criteria were as follows: species limited to "*Homo sapiens*" and the "minimum required interaction score" was set to "high confidence (0.70)". Next, the "TSV" files containing the PPI network relationships were imported into Cytoscape software for visualization. The size of the nodes in the PPI network was evaluated by computing degree centrality (DC) parameters using the "CytoNCA" plugin. Subsequently, the Cytoscape plugin "MCODE" was applied to discover the interacting dense regions called sub-networks in large PPI networks [29]. The score of each sub-network was obtained as follows: density × number of nodes (V). The density of each sub-network was calculated by the following equation: $(|E|\max = |V|\ (|V| - 1)/2$, where E is the number of subnetwork edges). The key sub-network clusters with high scores and containing disease target genes were extracted from the three large PPI networks. Finally, after merging genes from the three sub-network clusters and removing duplicate genes, the final gene set was defined as the "Drug-DCM" gene set. In addition, the top disease genes with high degree values in the three sub-network clusters were defined as "Hub" genes.

2.7Validation of the expression level and diagnostic accuracy of "Hub" genes and identification of "Hub" genes-targeted drugs

To verify the expression level of the Hub gene and their diagnostic accuracy in DCM (Treat) and non-DCM (Con) patients, the four datasets (GSE84796, GSE29819, GSE42955 and GSE120895) were merged to obtain the validation dataset. The process of data merging and normalization is the same as above. In the validation dataset, the expression levels of the Hub genes were compared between the DCM and non-DCM groups using a boxplot. The diagnostic accuracy of the Hub genes was assessed between the DCM and non-DCM groups using ROC curve analysis. Finally, the Drug Gene Interaction Database (DGIdb) was applied to identify drugs targeting the Hub gene with default parameters [30]. 30 targeted therapy drugs were visualized by Cytoscape software.

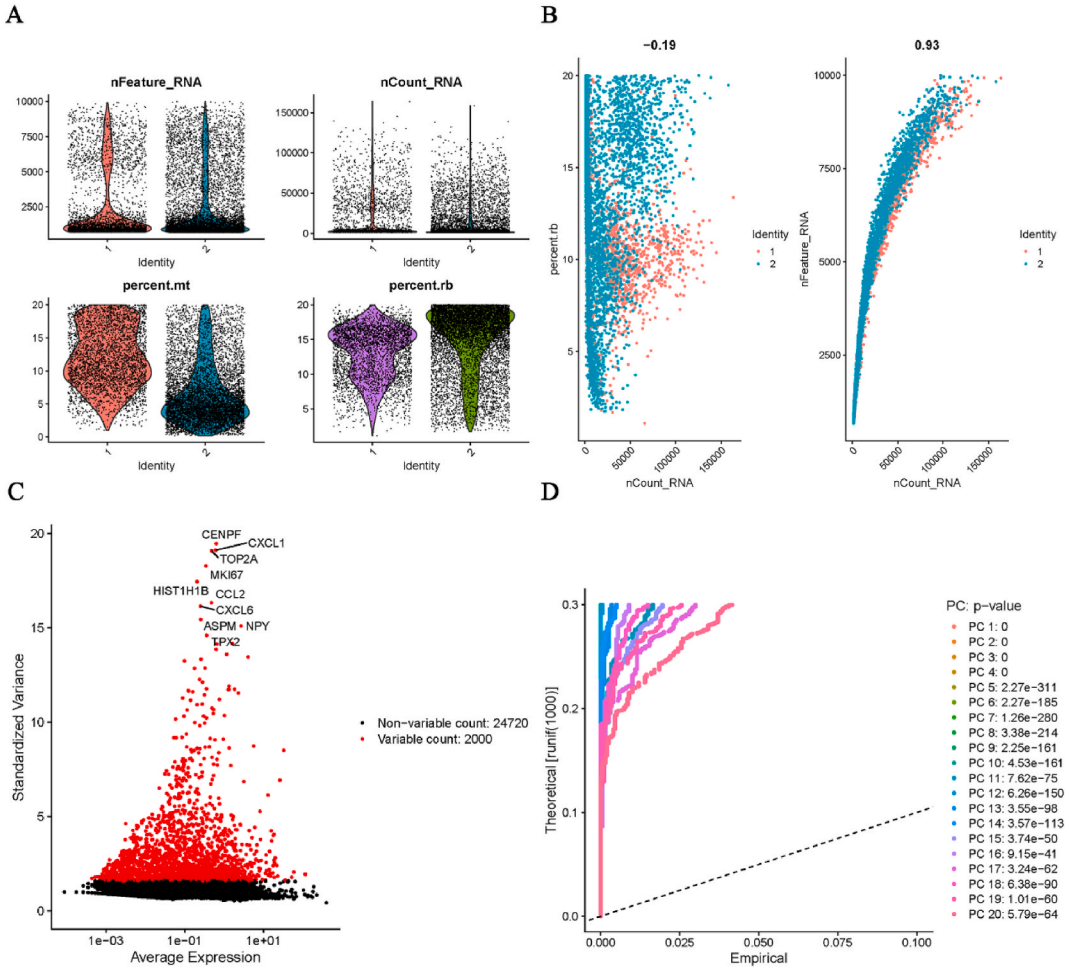
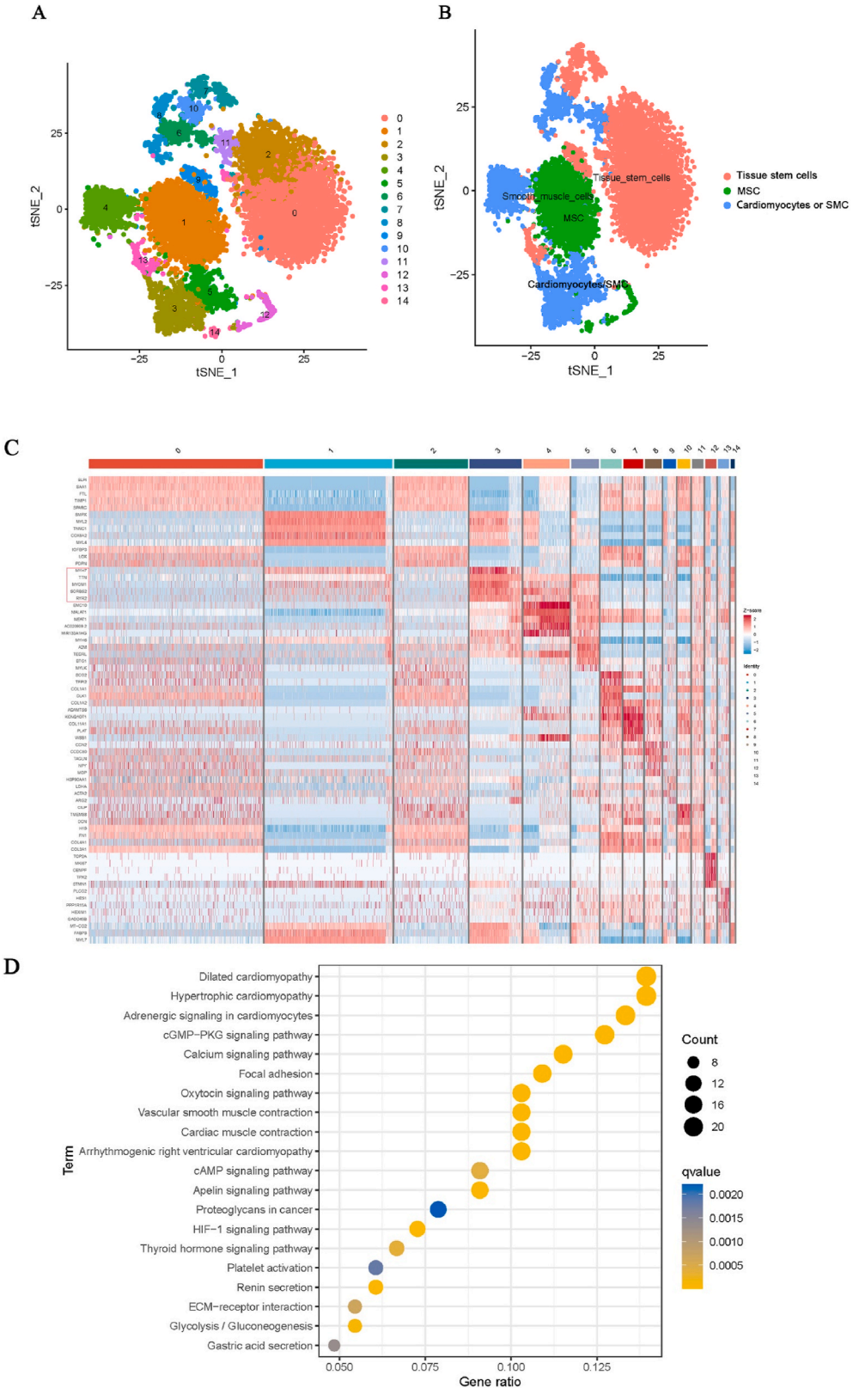


Fig. 1. Quality control and PCA analysis of single-cell datasets. (A) the total number of "nFeature_RNA" and "nCount_RNA", the "percent.mt" and "percent.rb" percentage of each sample. (B) the correlation between "nCount_RNA" and "nFeature_RNA" and "percent.rb". (C) the top 2000 highly variable genes are marked in red. (D) 20 PCs with $P < 0.05$ were identified using the JackStraw function.



(caption on next page)

Fig. 2. Marker genes in cell clustering, cell annotation and enrichment analysis. (A) T-SEN projections of 15 cell clusters in two-dimensional planes. (B) The three cell types were identified by further cell annotation. Different colors represent different cell types. (C) The top 5 significant makers genes in each cell cluster. The redder the color, the higher the Z-score of the gene in each cluster. (D) KEGG enrichment analysis of genes in cluster 3 and cluster 5. The size of bubbles indicates the number of genes, and the more yellow the color, the more significant it is.

2.8. Immune infiltration analysis and Hub gene correlation with infiltrating immune cells

The preprocessed validation dataset was used for "ssGSEA" analysis. The level of infiltration of 28 immune cells ("immune.gmt", acquired from the MSigDB) in the merged gene expression profile was assessed by the "ssGSEA" algorithm [31]. Violin plots were used to visualize the expression levels of 28 immune cells between the DCM and non-DCM groups. Finally, the association between immune cells and the "Hub gene" was assessed by Spearman's correlation analysis. The above results were plotted using the R package "ggplot2".

2.9. Assessment of DCM druggable cell cluster subgroups

The "irGSEA" package was used to evaluate the "Drug-DCM" gene set enrichment scores in different cell clustering subgroups. The enrichment scores were calculated by the "AUCell", "UCell", "singscore" and "ssgsea" algorithms. The results were visualized in heatmaps, density scatter plots and half-violin plots.

2.10. Two-sample mendelian randomization (MR) analysis

Utilizing the Quantitative Trait Loci (eQTL) summary statistics from the Genotype-Tissue Expression (GTEx) database, the identification of eQTLs for *ERBB2* and *PIK3R1* genes was achieved. These eQTLs exhibited a strong correlation with the expression levels of *ERBB2* and *PIK3R1*, with some single nucleotide polymorphisms (SNPs) meeting the essential criteria (correlation, independence, and exclusivity assumptions) for instrumental variables (IVs) in Mendelian randomization (MR) analysis. The F-test statistic ($F = [(N–K-1)/K] \times [R^2/(1\text{-}R^2)]$, N: sample size; K: number of IVs) was used to assess whether these selected SNPs were affected by weak instrumental variables ($F > 10$) [32]. The PhenoScanner database is employed for the purpose of identifying and eliminating single nucleotide polymorphisms (SNPs) that exhibit associations with established confounding factors. In this study, genome-wide association study (GWAS) data pertaining to heart failure were procured from the UK Biobank - Neale lab (IEU GWAS ID: ukb-*d*-HEARTFAIL) for a European cohort comprising 361,194 individuals. Among these, 1405 participants were diagnosed with heart failure, while the remaining 359,789 individuals served as controls. The dataset encompassed a total of 9,858,439 SNPs.

In the analysis of MR, the primary method employed was the inverse variance-weighted (IVW) method [33], while the weighted_median method served to validate the robustness of the IVW results. Heterogeneity was assessed using the Cochran Q statistic, and the presence of horizontal pleiotropy was determined by the intercept term of MR-Egger regression [34]. Additionally, a leave-one-out analysis was conducted to ensure the reliability of the overall effect of the findings. These aforementioned procedures were executed using the "TwoSampleMR" package in R.

3. Results

3.1. Single-cell analysis identified key cell clusters and the "SC-DCM" gene set

The expression characteristics of the single-cell dataset (GSE184899) after quality control and filtering are presented in Fig. 1A. The "nCount_RNA", which represents the total number of genes detected in the cell, was not related to "percent.rb", which represents the ribosome content, but was positively correlated with "nFeature_RNA", which represents the sum of expression of all genes in the cell, with a correlation coefficient of 0.93 (Fig. 1B). The top 2000 highly variable genes (HVGs) are shown in Fig. 1C. According to the cluster tree and elbow plot (Supplementary Fig. 1), 20 PCs with $P < 0.05$ were identified using PCA analysis and are shown in a JackStraw plot (Fig. 1D).

Subsequently, the 20 PCs were clustered into 15 cell subgroups using the t-distributed stochastic neighbor embedding (T-SEN) method (Fig. 2A). Subsequently, the 15 cell subgroups were further annotated into 3 cell types, including cells with contractile function, such as cardiomyocytes or smooth muscle cells (SMC), mesenchymal stem cells (MSC), and tissue stem cells (TSC) (Fig. 2B). The top 5 significant makers genes in each cell cluster are shown in Fig. 2C, Supplementary file 3. We found that the top significant marker genes in clusters 3 and 5, such as *MYH7*, *TTN*, *MYOM1*, *SORBS2* and *RYR2*, were strongly associated with the progression of DCM disease (Fig. 2C). Therefore, genes in clusters 3 and 5 were extracted and underwent KEGG enrichment analysis, which showed significant enrichment in DCM, hypertrophic cardiomyopathy (HCM), and adrenergic signaling in cardiomyocytes (Fig. 2D, Supplementary file 4). These genes were also defined as the "SC-DCM" gene set for subsequent analysis. We found that both clusters 3 and 5 belonged to the cells with contractile function, such as cardiomyocytes or smooth muscle cells type.

3.2. Differential analysis and WGCNA analysis identified the "EP-DCM" gene set

During differential expression analysis, 399 DEGs were obtained based on the filtering conditions. DEGs expression in the samples

was visualized in a heatmap and volcano plot (Fig. 3A and B). GSEA analysis of DEGs showed that the DCM group was mainly enriched in immune-related pathways, such as antigen processing and presentation and natural killer cell-mediated cytotoxicity, while the non-DCM group was mainly enriched in physiological metabolic processes, such as amino sugar and nucleotide sugar metabolism and oxidative phosphorylation (Fig. 3C and D, Supplementary file 5).

During WGCNA analysis, an optimal soft threshold ($\beta = 6$) was used to construct a scale-free network. (Supplementary Fig. 2A). Subsequently, the topological overlap matrix was clustered and displayed in the dynamic tree diagram (Fig. 4A). Six gene co-expression modules, the relationship between modules and clinical traits (DCM(Treat) and non-DCM(Con)) and the gene significance in each module were identified by the dynamic shearing module and the module eigengenes (MEs) (Fig. 4B and C). We found that MEbrown was strongly correlated with the Con group (cor = 0.75, P = 5e-09), while MEred and MEturquoise were strongly correlated with the DCM group (cor = 0.8, P = 1e-10 and cor = 0.69, P = 3e −07). After filtering for GS > 0.5 and MM > 0.8, marker genes in 3 modules significantly correlated with the clinical traits (Fig. 4D, E, and F).

Subsequently, 60 overlapping genes were obtained after taking the intersection of marker genes from the 3 modules with DEGs (Fig. 4G). GO enrichment analysis of overlapping genes showed significant enrichment in biological metabolisms closely related to DCM, such as extracellular matrix organization, collagen-containing extracellular matrix and extracellular matrix structural constituent (Fig. 4H, Supplementary file 6). The overlapping genes were defined as the "EP-DCM" gene set for subsequent analysis.

3.3. Acquisition and analysis of the "drug" gene set

573 curcumin target genes were acquired through CTD, SwissTargetPrediction, BindingDB and TargetNet databases (Fig. 5A, Supplementary file 7). KEGG enrichment analysis of drug target genes showed that lipid and atherosclerosis and PI3K-Akt signaling

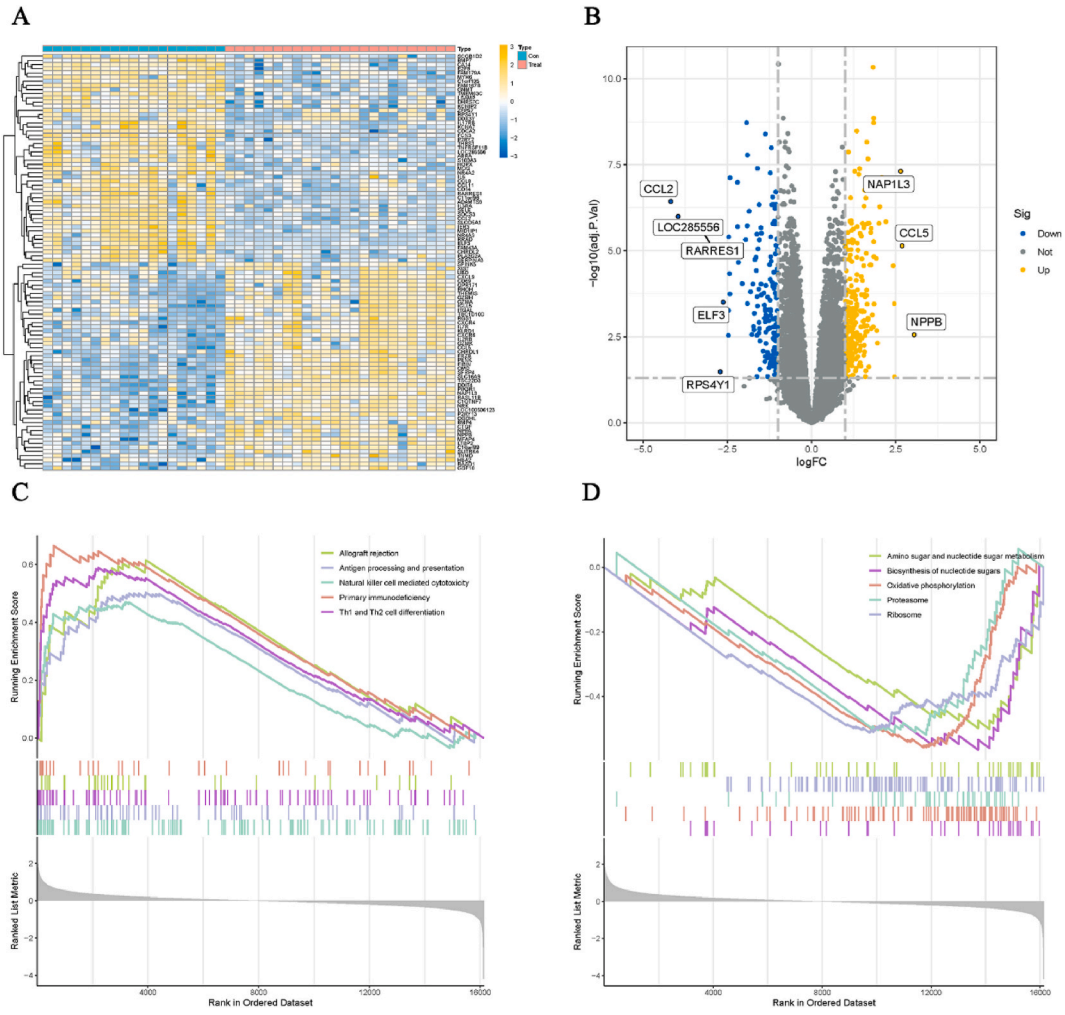


Fig. 3. Screening of DEGs and GSEA analysis of DEGs. (A) Heatmap of DEGs expression in DCM (Treat) and non-DCM (Con) groups. (B) Volcano plots of DEGs expression in Treat and Con groups. (C)The top 5 KEGG gene set scores of GSEA analysis in the Treat group. (D) The top 5 KEGG gene set scores of GSEA analysis in the Con group.

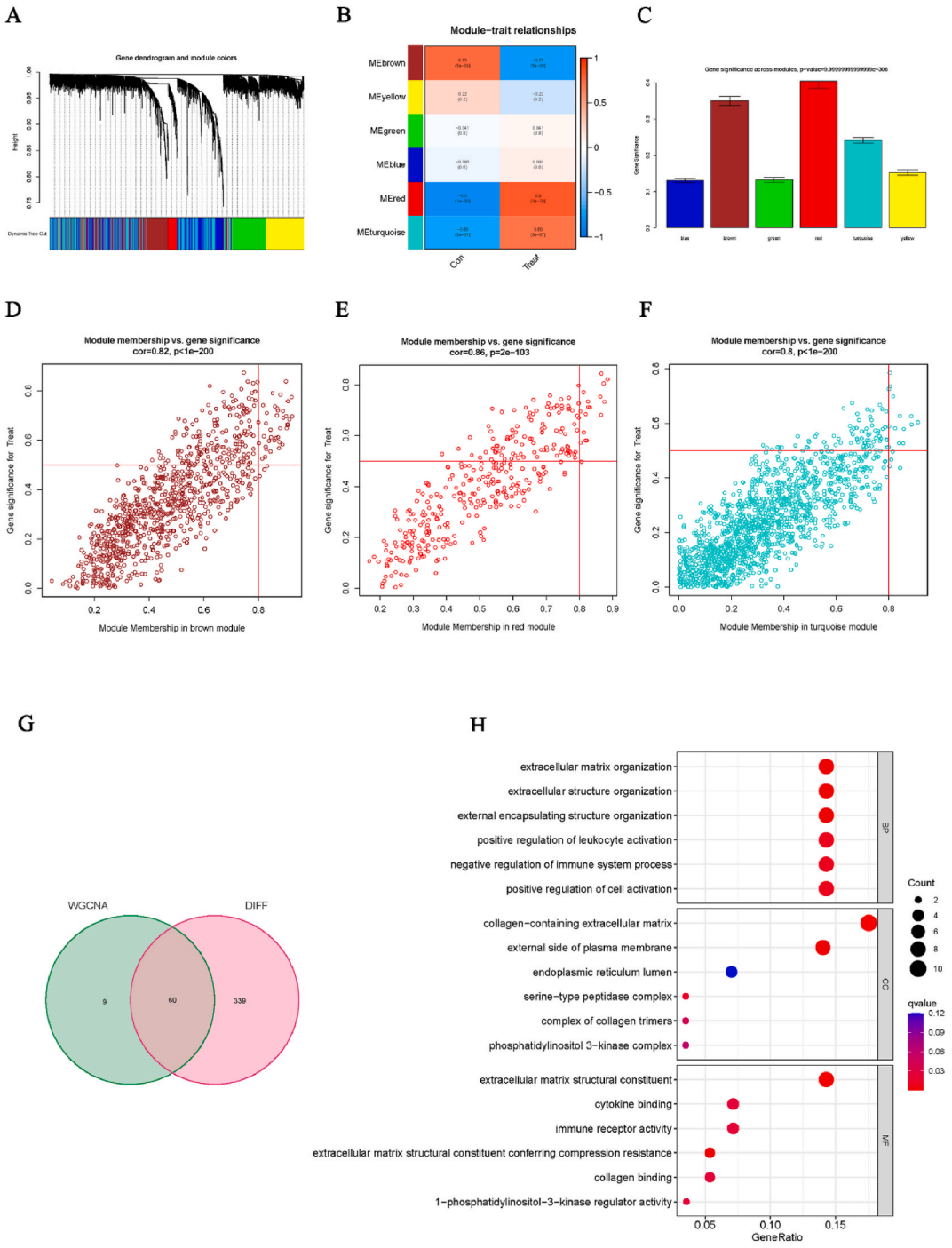


Fig. 4. WGCNA analysis and identification of the "EP-DCM" gene set. (A) Dynamic dendrogram after topological overlap matrix clustering. Each color below represents a co-expression module. (B) Module-trait relationships heatmap, the stronger the color, the stronger the correlation. (C) Distributing gene significance in the six modules. (D, E, and F) In the brown, red and turquoise module, a scatter plot of gene MM vs GS is shown, Genes with GS $>$ 0.5 and MM $>$ 0.8 were screened out. (G) Venn plot of the overlapping genes. (H) Bubble plot representing GO enrichment analysis.

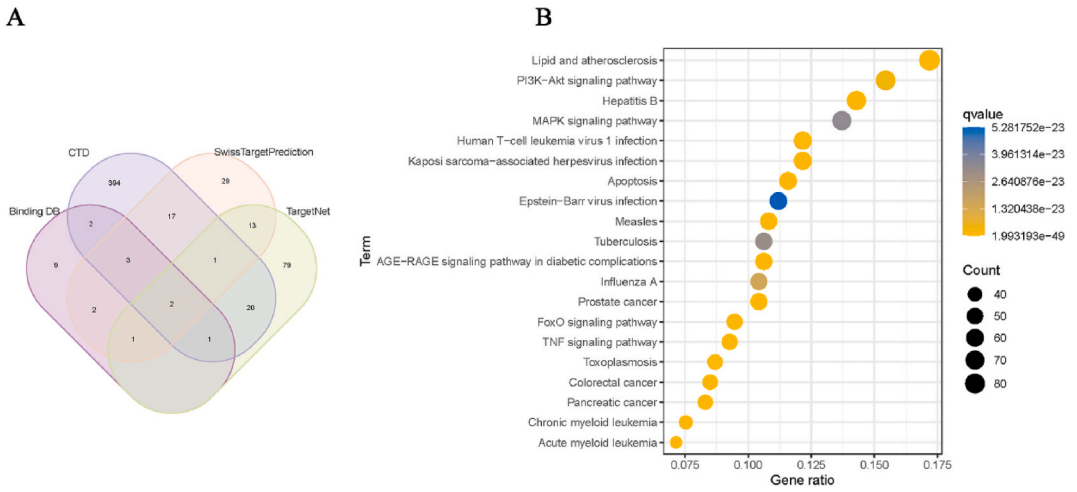


Fig. 5. Acquisition and analysis of the "Drug" gene set. (A) Venn plot of drug-related targets in four databases. (B) KEGG analysis of drug target genes.

pathway were significantly enriched (Fig. 5B. **Supplementary file 8).** These pathways have also been associated with disease progression in DCM. The drug target genes were defined as the "Drug" gene set for subsequent analysis.

3.4. PPI network analysis identified the "Drug-DCM" gene set and Hub genes

According to the study design, three large PPI networks were constructed, namely "Drug" and "EP-DCM" PPI network 1, "Drug" and "SC-DCM" PPI network 2 and "Drug" and "EP-DCM" and "SC-DCM" PPI network 3. PPI network 1 comprised 560 nodes and 12740 edges, with a PPI enrichment p-value <1.0e-16. The key sub-network clusters of PPI network 1 contained 68 nodes and 878 edges, with an "MCODE" score of 13.104 (Fig. 6A). PPI network 2 comprised 780 nodes and 15110 edges, with a PPI enrichment p-value <1.0e-16. The key sub-network clusters of PPI network 2 contained 86 nodes and 1176 edges, with an "MCODE" score = 13.835(Fig. 6B). PPI network 3 included 814 nodes and 15504 edges, with a PPI enrichment p-value: <1.0e-16. The key sub-network clusters of PPI network 3 contained 82 nodes and 1062 edges, with an "MCODE" score = 13.111(Fig. 6C). Finally, after merging genes from the three key sub-network clusters and removing duplicate genes, the final gene set was defined as the "Drug-DCM" gene set. In addition, disease genes with high degree values in the three sub-network clusters, *PIK3R1* and *ERBB2,* were defined as "Hub" genes.

3.5. Validation of "Hub" genes and identification of "Hub" genes-targeted drugs

The expression levels of Hub genes (*PIK3R1* and *ERBB2)* between DCM (Treat) and non-DCM (Con) in the validation datasets (GSE84796, GSE29819, GSE42955 and GSE120895) were visualized in boxplots (Fig. 7A). The expression levels of *PIK3R1* were significantly higher in the Treat group than in the Con group ($P < 0.001$). In contrast, *ERBB2* exhibited a significant correlation with the Con group ($P < 0.05$). ROC curve analysis in the validation datasets assessed the diagnostic accuracy of the *PIK3R1* and *ERBB2* genes between the Treat and Con groups. The AUC values of the *PIK3R1* and *ERBB2* genes were 0.771 and 0.624 (Fig. 7B), which showed that the *PIK3R1* gene had higher diagnostic efficacy as a drug target. The 30 predicted Hub gene-targeted drugs are shown in Fig. 7C. Among them, everolimus was identified as an inhibitory drug for both Hub genes. Given that the *ERBB2* gene was lowly expressed in the DCM group, we sought to identify agonists of the *ERBB2* gene. However, no agonist drugs were predicted in the DGIdb database.

3.6. Immune cell infiltration analysis and its correlation with "Hub" genes

Immune infiltration levels of "immune.gmt" in the validation dataset were analyzed by the "ssGSEA" algorithm. The distribution of the infiltrating immune cells between DCM (Treat) and non-DCM (Con) groups is displayed in Fig. 8A and Supplementary file 9. The results of the infiltration analysis indicated that the infiltration levels of activated CD8 T cell ($p = 0.005$), effector memory CD8 T cell ($p = 0.003$), central memory CD8 T cell ($p < 0.001$), effector memory CD4 T cell ($p < 0.001$), central memory CD4 T cell ($p = 0.002$) were higher in the Treat group than in the Con group. In contrast, eosinophil ($p = 0.03$) infiltration was lower in the Treat group than in the Con group (Fig. 8B). Subsequently, we assessed the correlation between infiltrated immune cells in the Treat group and Hub genes using spearman correlation analysis (Fig. 8C). We found that central memory CD8 T cells were positively correlated with *PIK3R1* ($p < 0.05$). In contrast, MDSC ($p < 0.01$), T follicular helper cells ($p < 0.01$), plasmacytoid dendritic cells ($p < 0.01$), monocytes and macrophages ($p < 0.05$) were negatively correlated with *PIK3R1*. In addition, immature dendritic cells ($p < 0.001$), gamma delta T cells ($p < 0.001$), eosinophils ($p < 0.001$), macrophages ($p < 0.01$), mast cell ($p < 0.01$), neutrophils ($p < 0.01$), type 2 T helper cells (p

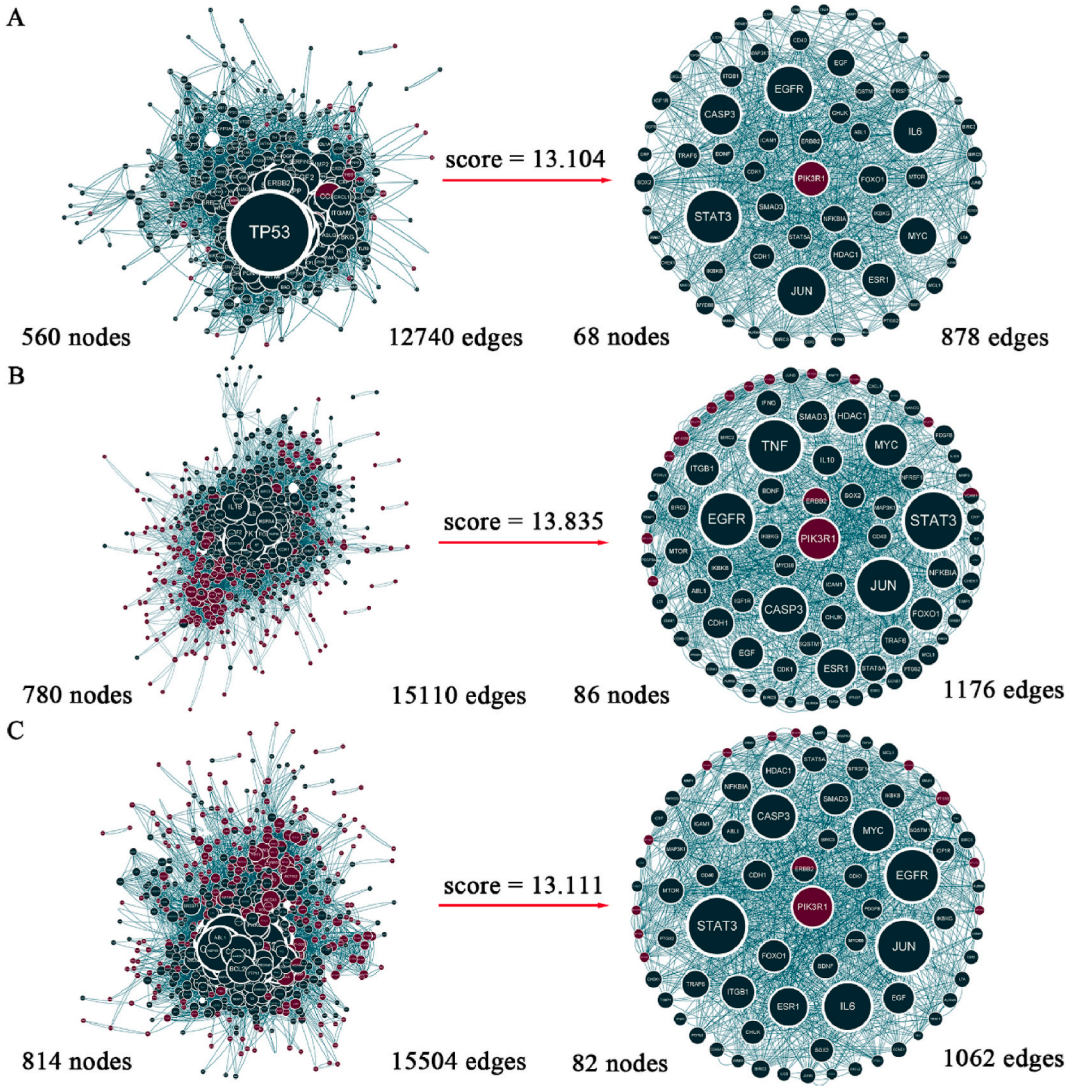


Fig. 6. PPI network analysis. (A) The "Drug" and "EP-DCM" PPI network 1 and the key sub-network clusters of PPI network 1. The red circles represent disease-related genes, and the dark green circles represent drug targets. (B) The "Drug" and "SC-DCM" PPI network 2 and the key sub-network clusters of PPI network 2. (C) The "Drug" and "EP-DCM" and "SC-DCM" PPI network 3 and the key sub-network clusters of PPI network 3.

< 0.01), activated CD4 T cells ($p < 0.05$), and natural killer cells ($p < 0.05$) were negatively correlated with *ERBB2*.

3.7. Assessment of druggable DCM cell subgroups using the "Drug-DCM" gene set

Using the "irGSEA" package, we evaluated druggable cell subgroups. Density scatter plots were used to show the distribution density of the "Drug-DCM" gene set in different cell subgroups based on "singscore" scores and projections of different cell subgroups in low-dimensional space (T-SEN and UMAP) (Fig. 9A and B). Finally, half-violin, ridge plot, and density heatmaps were applied to display the expression and distribution levels of the "Drug-DCM" gene set in different cell subgroups based on "ssGSEA" scores (Fig. 9C, D and E). The results showed that the "Drug-DCM" gene set had a higher enrichment score in cardiomyocytes or smooth muscle cells.

3.8. Two-sample MR analysis results

In order to further investigate the potential causal effect of *PIK3R1* and gene expression on the occurrence of heart failure, a two-sample MR analysis was employed. Following a filtering process, a total of 8 SNPs were identified as instrumental variables (IVs) for *ERBB2*, while 4 SNPs were identified as IVs for *PIK3R1*, each with an F-statistic >10. Subsequently, the genetic variant summary statistics corresponding to the IVs were extracted from GWAS data for heart failure (Table 1, Supplementary file 10 (Table s1)). The

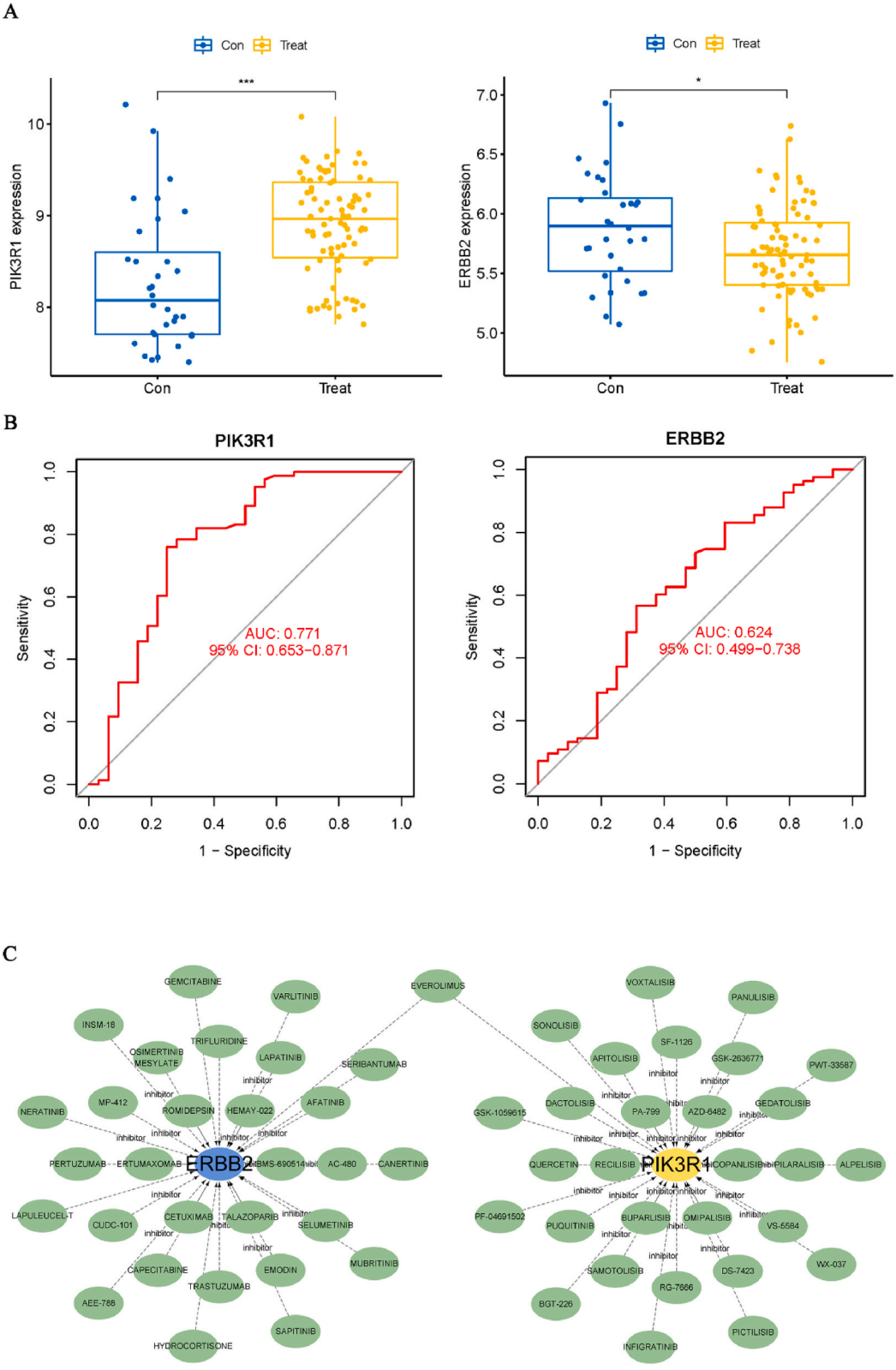
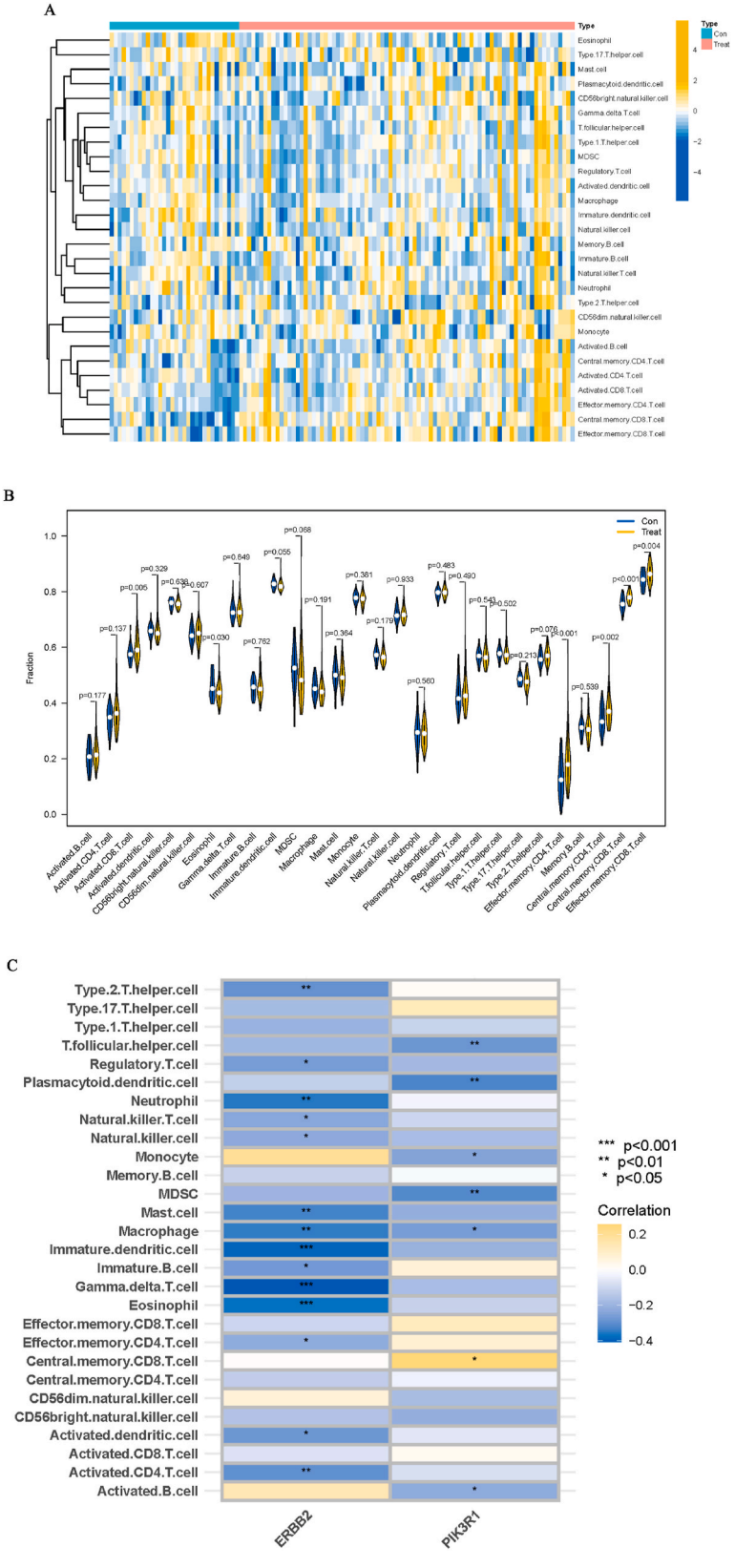


Fig. 7. Validation of "Hub" genes and Identification of "Hub" genes-targeted drugs. (A) Box plots of "Hub" genes (PIK3R1 and ERBB2) expression levels in DCM and non-DCM groups. (B) The diagnostic accuracy of the *PIK3R1* and *ERBB2* genes between the Treat and Con groups. (C) the 30 predicted "Hub" gene-targeted drugs. The green ovals represent different drugs.



(caption on next page)

Fig. 8. Immune cell infiltration analysis, correlation analysis of Hub genes with immune cells. (A) Distribution of 28 infiltrating immune cells in the Treat and Con groups. (B) violin plots showing the differences of 28 immune cells in the Treat and Con groups. (C) Relationship between infiltrated immune cells and two "Hub" genes; the more yellow the color, the more significant. P < (0.001, 0.01, 0.05) indicate "***", "**", "*".

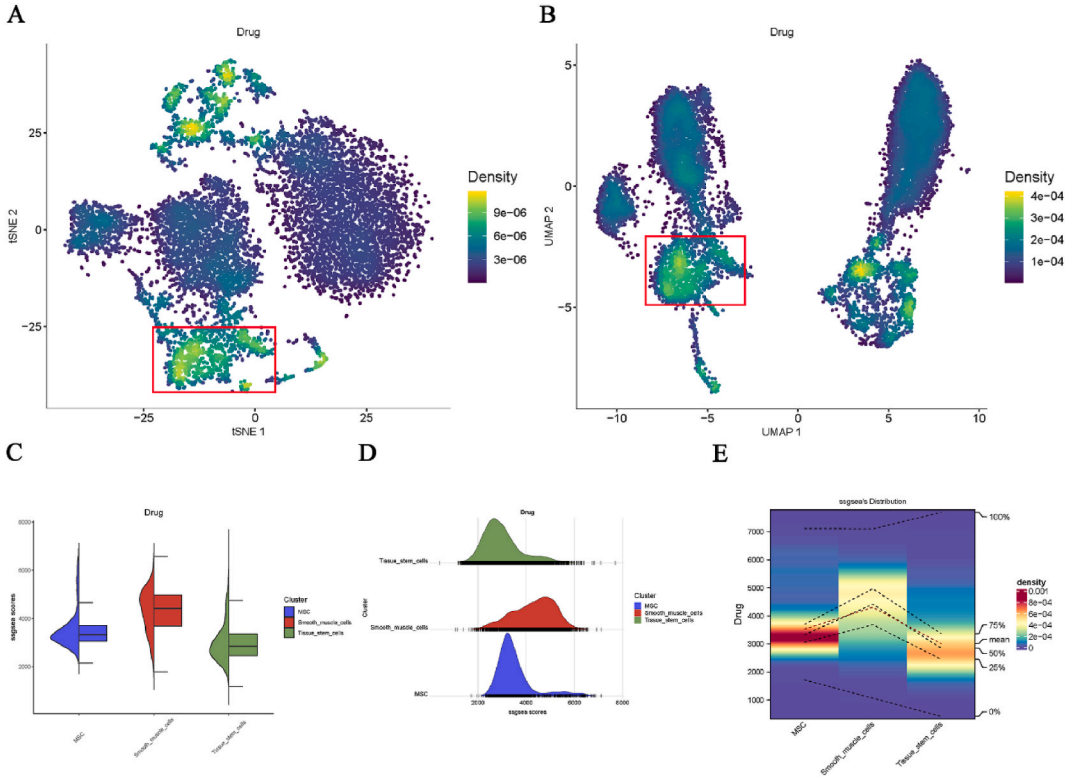


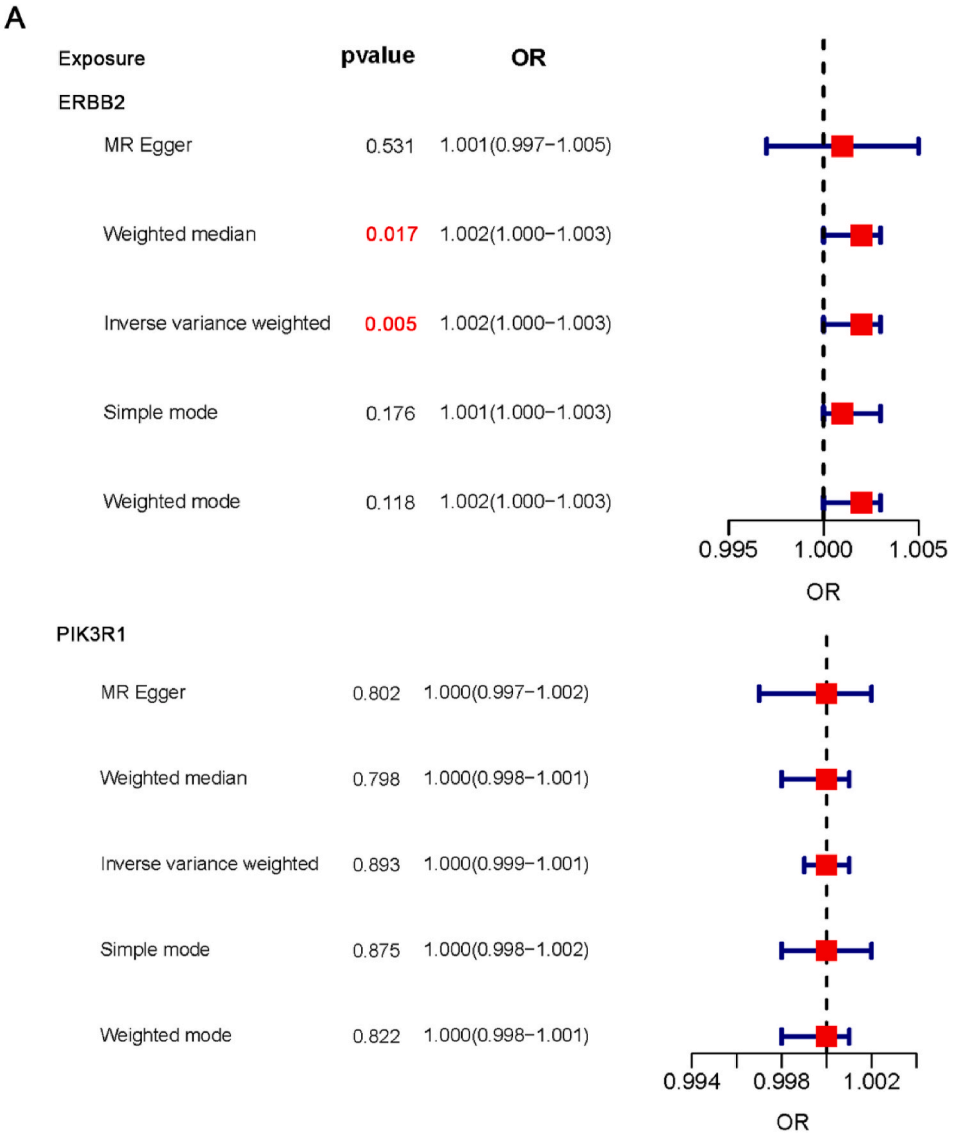
Fig. 9. Assessment of druggable DCM cell subgroups. (A and B) Density scatter plots showing the distribution density of the "Drug-DCM" gene set in different cell cluster subgroups, the more yellow the color, the higher the density. (C) The half-violin plot is presented as a violin plot (left) and a box plot (right). Different colors represent different cell subgroups, the width of the violin represents the distribution level, and the height of the box plot represents the expression level. (D) The horizontal coordinates of the ridgeplot represent different expression levels. (E) Density heatmap showing the expression and distribution levels.

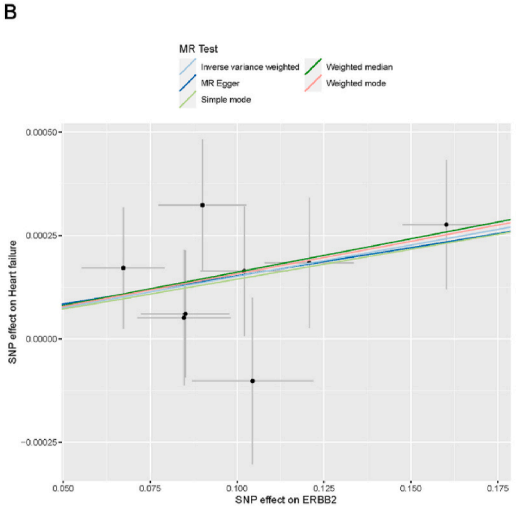
Table 1
Summary statistics of IVs in exposure (*ERBB2*) and outcome.

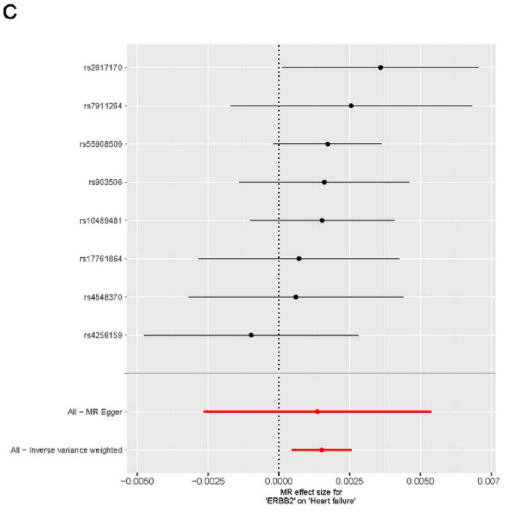
SNPs	R^2	ERBB2				F	Heart failure			
		Beta	Se	Eaf	P-value		Beta	Se	Eaf	P-value
rs10489481(A/G)	0.006289	0.120783	0.012779	0.314428	3.33E-21	198.4427	0.000184	0.000157	0.32078	0.241751
rs17761864(A/C)	0.003173	0.085127	0.0127	0.323707	2.04E-11	99.79548	6.05E-05	0.000154	0.346369	0.694028
rs2617170(C/T)	0.003599	−0.09008	0.012616	0.668027	9.33E-13	113.2417	−0.00032	0.000159	0.694966	0.041605
rs4256159(T/C)	0.002551	0.104451	0.017384	0.135199	1.87E-09	80.19243	−0.0001	0.000202	0.155657	0.613536
rs4848370(T/C)	0.002826	0.084733	0.013397	0.269386	2.53E-10	88.85981	5.12E-05	0.000164	0.274232	0.755009
rs55908509(A/G)	0.011322	−0.16021	0.012601	0.3284	4.96E-37	359.0392	−0.00028	0.000156	0.327465	0.076583
rs7911264(C/T)	0.002261	−0.06728	0.011896	0.515223	1.55E-08	71.05611	−0.00017	0.000147	0.520529	0.241116
rs903506(A/G)	0.004683	−0.10207	0.012525	0.658916	3.67E-16	147.506	−0.00016	0.000157	0.677093	0.294913

Beta: the effect size of SNPs on exposure factors. **Se**: the standard error of Beta. **Eaf**: effect allele frequency.

IVW method yielded significant evidence of a causal relationship between *ERBB2* and heart failure (p-value = 0.005, OR = 1.002, 95% CI: 1.000–1.003), which was further supported by the weighted median regression analysis (p-value = 0.017, OR = 1.002, 95% CI: 1.000–1.003) (Fig. 10 A). Conversely, there was no observed causal association between *PIK3R1* and heart failure according to the IVW analysis (p-value = 0.892, OR = 0.999, 95% CI: 0.998–1.001) (Fig. 10 A). The risk of heart failure is heightened by the presence of these mutant loci that impact the expression of the *ERBB2* gene (Fig. 10 B). The forest plot also illustrates the magnitude of the effect for each mutant locus in *ERBB2* on HF (Fig. 10 C). The absence of potential horizontal pleiotropy and heterogeneity in IVs is indicated by the intercept test and Cochran Q statistic of MR-Egger regression (Supplementary file 10 (Table s2)). Furthermore, the leave-one-out







(caption on next page)

Fig. 10. The relationship between exposure variables (*ERBB2 and PIK3R1*) and the outcome variable (heart failure). (A) The forest plot visually represents the MR analysis results, highlighting statistically significant findings (p-value <0.05) in red. (B) Figure B displays a scatter plot illustrating the trend of *ERBB2*'s effect on heart failure in the presence of SNPs. (C) A forest plot depicting the effect size of each mutant locus within *ERBB2* on the occurrence of heart failure.

sensitivity analysis confirms that the overall causality remains unaffected after removing individual SNP loci (Supplementary file 10 (Supplementary Fig. 3 A and B)), thereby suggesting that the estimated effect of exposure on outcome cannot be attributed to any single IV.

3.9. Discussion

Over the years, the integration of single-cell RNA sequencing (scRNA-seq) technology with various bioinformatics analysis methods has facilitated remarkable advancements in medical research [35]. A notable benefit of single-cell sequencing technology lies in its capacity to effectively and impartially cluster and categorize cells [36]. Interestingly, our findings indicate a strong association between highly-variable genes located at the top of adjacent cell clusters 3 and 5, namely *MYH7, TTN, MYOM1, SORBS2, and RYR2,* and the progression of dilated cardiomyopathy. Among this set of genes, the *MYH7* gene is responsible for encoding the β heavy chain subunit of cardiac myosin, the *TTN* gene encodes a substantial protein present in striated muscle, and the *RYR2* gene encodes a ryanodine receptor located in the cardiac sarcoplasmic reticulum, which is a crucial component of the calcium channel. Any mutations occurring in these genes would directly lead to both structural and functional deficiencies in the heart [37–39]. Consequently, we postulated that clusters 3 and 5 could potentially serve as pivotal pathogenic cell clusters in DCM. To validate this hypothesis, we extracted the genes from these clusters and subjected them to KEGG enrichment analysis, which demonstrated a significant enrichment in DCM, hypertrophic cardiomyopathy, and arrhythmogenic cardiomyopathy, thus corroborating our conjecture. We then performed cellular annotation of the two cell clustering subgroups and found an association with cells with contractile function, such as cardiomyocytes or smooth muscle cells. Smooth muscle cells, which can be classified into synthetic and contractile phenotypes based on the expression of myosin heavy chain isoforms, are present on the inner surface of the endocardium in the contractile phenotype [40]. In normal hearts, the SMC layer thickness is < 30 μm, whereas the endocardial smooth muscle thickness can increase to $\leq$80 μm in patients with DCM [41]. Following this, we designated the genes found in clusters 3 and 5, which exhibit a strong correlation with DCM, as the "SC-DCM" gene set. This selection was made in order to conduct further investigations into the potential suitability of cells with contractile function, implicated in disease causation, as targets for drug therapy. Nevertheless, it is important to acknowledge certain limitations that may impact the reliability of our findings, such as the utilization of single-cell sequencing data obtained from congenital DCM day 35 iPSC-CMs, which do not represent fully mature cardiac tissue cells.

Differential expression analysis is employed to identify genes exhibiting significant upregulated and downregulated expression in microarray expression profile data, whereas WGCNA analysis is utilized to identify co-expression modules that are closely linked to DCM. By integrating the outcomes of both analyses, a distinct set of genes referred to as "EP-DCM" is identified, which displays differential expression between the DCM and non-DCM groups and exhibits a strong association with clinical traits. The GSEA enrichment analysis of differentially expressed genes (DEGs) revealed that the DCM group exhibited significant enrichment in immune-related pathways, specifically antigen processing and presentation as well as natural killer cell-mediated cytotoxicity. Conversely, the non-DCM group demonstrated predominant enrichment in physiological metabolic processes, including amino sugar and nucleotide sugar metabolism, as well as oxidative phosphorylation. This observation aligns with previous studies that have highlighted the crucial involvement of immune-mediated inflammatory injury in the pathogenesis and advancement of DCM [42], In DCM, the activation of antigen-presenting cells, including dendritic cells, as well as innate immune cells, such as natural killer cells, initiates a series of inflammatory and immune responses. These responses subsequently facilitate tissue healing and remodeling by activating compensatory mechanisms. Nevertheless, as time progresses, the persistent nature of remodeling and inflammation leads to chronicity, ultimately resulting in a decline in cardiac function [43].

Following that, we conducted a GO enrichment analysis on the gene set labeled as "EP-DCM". Our analysis revealed a significant enrichment in processes related to extracellular matrix organization, collagen-containing extracellular matrix, and extracellular matrix structural constituent. This finding aligns with the existing body of evidence that supports the notion that alterations in the extracellular matrix are integral to the development and progression of DCM. Moreover, it is well-established that the subsequent myocardial fibrosis resulting from these alterations leads to left ventricular enlargement and thinning, ultimately culminating in structural remodeling and functional deterioration of the heart [44]. Hence, the findings of this study provide compelling evidence that the "EP-DCM" gene set holds significant representativeness and relevance for future investigations.

During the analysis of protein-protein interaction networks, we integrated two gene sets ("SC-DCM" and "EP-DCM") related to DCM with a gene set associated with drug. This integration resulted in the construction of three extensive PPI networks. Subsequently, employing the "MCODE" plugin, we extracted three significant sub-network clusters with high scores, which contained disease target genes. The genes obtained from merging these three sub-network clusters were designated as the "Drug-DCM" gene set. This gene set was then utilized to evaluate the potential drug-treatability of cells with contractile function. Furthermore, the disease genes *PIK3R1* and *ERBB2*, which exhibited high degree values in the three sub-network clusters, were designated as "Hub" genes. The validation datasets were utilized to evaluate the expression levels and diagnostic accuracy of these Hub genes in distinguishing between the DCM and non-DCM groups. Our findings indicate a significant disparity in the expression of *PIK3R1 and ERBB2* between the two groups. Moreover, the drug target *PIK3R1* demonstrated superior diagnostic accuracy.

PIK3R1, also known as p85α, is a regulatory subunit of phosphatidylinositol 3-kinases (PI3K). It is widely acknowledged that the p85α regulatory subunit binds to the catalytic subunit (p110α, β, or δ) to form a heterodimer of PI3K [45]. The catalytic subunit (p110) is called a "tumor suppressor", while the regulatory subunit (p85) is called an "oncogene" [46]. Current evidence suggests that *PIK3R1* is involved in the *PI3K* kinase signaling cascade and plays an important role in immune response, cell growth, proliferation and survival, and insulin resistance. For instance, mutations in *PIK3R1* cause immune deficiency [47–49], SHORT syndrome [50], lipodystrophy and systemic insulin resistance (hypoglycemia) [51,52]. In our investigation, the DCM group exhibited a higher expression level of *PIK3R1* compared to the non-DCM group, suggesting a potential state of immune system over-activation and heightened cellular stress. A separate study has highlighted the potential therapeutic significance of *PIK3R1* as a novel target against hepatocellular carcinoma [53]. Furthermore, in the context of the cardiovascular system, inhibition of the *PIK3R1/Akt/mTOR* signaling pathway demonstrated significant suppression of pathological cardiomyocyte autophagy, along with mitigation of cellular oxidative stress and apoptosis [54]. In our study, the identification of everolimus as a drug target for the *PIK3R1* gene aligns with its established role as a well-recognized inhibitor of the mTOR signaling pathway. Our findings, in line with existing literature, indicate that *PIK3R1* holds promise as a prospective therapeutic target in the context of DCM and exhibits a strong correlation with immune system activation.

ERBB2, also known as *HER-2*, is a ligand-free receptor kinase originally identified as a gene frequently overexpressed in many types of tumors, such as breast cancer. When the *ERBB2* inhibitor trastuzumab (Herceptin) was used to treat tumors, patients experienced significant cardiac toxicities. Accordingly, the mechanism by which *ERBB2* exerts a protective effect on the heart was extensively studied [55]. In DCM, normal expression of the *ERBB2* gene or *ERBB2* kinase activity is essential for cardioprotection. Studies have shown that *ERBB2* can exert multiple protective effects in DCM through the neuromodulin-1 (NRG-1)*/ERBB* signaling pathway, G protein-coupled receptor signaling, and β-adrenergic stress [56,57]. However, mutations in the *ERBB2* gene and reduced activity of *ERBB2* cause severe DCM and heart failure in mice [58]. In response to *ERBB2*'s powerful protective effects on the heart, scientists have designed *ERBB2* agonist drugs and used them to treat heart failure, but the risk of malignancy needs to be considered. Interestingly, one study designed a bivalent NRG-1b agonist, which did not increase the risk of malignancy [59–61]. Despite this cross-talk, *ERBB2* remains a very attractive therapeutic target among DCM.

In our study, we observed a statistically significant decrease in *ERBB2* expression in the DCM group compared to normal controls ($P < 0.05$). This finding suggests that reduced *ERBB2* expression may contribute to the development and progression of DCM, which is consistent with previous literature [55–58]. Based on our analysis of extensive data and pharmacological considerations, our study provides a rationale for considering *ERBB2* as a potential therapeutic target in the treatment of DCM. Interestingly, when we conducted drug target analysis using the DGIdb database, we found that only oncologic drugs were identified, with no cardiovascular system agonist drugs. This observation highlights the potential promising prospects of *ERBB2* agonists in the treatment of DCM.

A recent comprehensive analysis of various types of cancer demonstrated that *PIK3R1* exhibits low expression levels in the majority of tumors [62]. However, our study reveals a contrasting finding, indicating that *PIK3R1* is highly expressed in cases of DCM. Moreover, this phenomenon of cross-expression is also observed in the case of *ERBB2*, as previously discussed. Consequently, our extensive investigation provides a more comprehensive characterization of these two genes, highlighting their cross-expression effects in both DCM and cancer. The elevated expression of *PIK3R1* in DCM is intricately linked to the heightened immune and inflammatory responses, which on the one hand safeguards the heart's functionality, and on the other hand triggers irreversible cardiac remodeling. Conversely, the diminished expression of *PIK3R1* in the majority of cancers can be attributed to the tumors inducing mutations in the *PIK3R1* gene. Consequently, this weakens the regulatory subunit (p85α) of PI3K's ability to inhibit the catalytic subunit (p110α), ultimately suppressing the human immune response against cancer. Furthermore, *ERBB2* exhibits significant upregulation in the majority of malignancies as a proto-oncogene, thereby exerting a profound impact on the prognosis of various tumors, notably breast cancer. This effect is achieved through the activation of diverse growth factors, leading to enhanced cellular proliferation, differentiation, and survival. Conversely, the expression of the *ERBB2* gene is notably diminished in DCM, potentially attributable to the fact that mutations in the *ERBB2* gene are implicated in the pathogenesis of this condition. Given the crucial role of *ERBB2* in cardiac development and the maintenance of normal myocardial function, its reduced expression becomes critical for the preservation of differentiated myocardium. This valuable insight can aid future researchers in their pursuit of breakthroughs in the development of therapeutic agents for DCM.

Over time, DCM consistently progresses to heart failure. In order to investigate the potential causal relationship between the cross-expression effects of *PIK3R1* and *ERBB2* in tumor and heart diseases, we employed expression quantitative trait loci (eQTLs) with highly correlated levels of *PIK3R1* and *ERBB2* gene expression as instrumental variables (IVs). Additionally, we utilized a large genome-wide association study (GWAS) cohort consisting of individuals with impaired cardiac structure and function as the outcome measure. Through the application of Mendelian randomization, we conducted an unbiased estimation of their causality. The findings from our MR analysis revealed a causal association between the *ERBB2* gene and impaired cardiac function as well as heart failure. However, no such causal relationship was observed for the *PIK3R1* gene. Hence, it is posited that the preservation of regular cardiac function necessitates the unimpeded expression of *ERBB2.* Conversely, diminished levels of *ERBB2* gene expression and mutations in associated loci may intensify the decline in cardiac function. Consequently, the potential efficacy of *ERBB2*-targeted medications in managing myocardial disease appears promising, as they have the capacity to not only decelerate disease progression and enhance myocardial prognosis in its initial stages, but also potentially restore impaired cardiac function. However, it is important to acknowledge that the clinical targeting of drugs for cardiac disease must consider the potential cross-expression effects that could be oncogenic. Furthermore, our study did not establish a causal association between *PIK3R1* and heart failure, possibly due to the elevated expression of *PIK3R1* in non-oncological diseases involving immune system involvement. Conversely, targeted agents against *PIK3R1* may play a significant role in regulating immune system dysregulation and imbalance.

During immune infiltration analysis, we found that CD4 and CD8 T cells exhibited significantly higher levels of immune infiltration in DCM than in the non-DCM group. This result indicates that activation of the adaptive immune system is closely related to the progression of DCM disease. An increasing body of evidence suggests the important role of activation of the adaptive immune system in developing DCM [42]. For instance, in the event of myocardial injury, the activation of "pattern recognition receptors" (part of the innate immune system) on cardiomyocytes triggers the transmission of inflammatory signals to antigen-presenting cells (APCs), which subsequently relay these signals to adaptive immune cells, including T cells. These resident adaptive immune cells serve the dual purpose of safeguarding the myocardium against pathogens and overseeing the processes of cardiac healing and remodeling. This regulatory function is crucial for the maintenance of cardiac homeostasis and the coordination of normal tissue function within the heart [63]. During analysis of the relationship between *PIK3R1* and *ERBB2* and immune cell infiltration, we found that *PIK3R1* was positively correlated with adaptive immune CD8 T cells and negatively correlated with the innate immune system, such as plasmacytoid dendritic cells ($p < 0.01$), monocyte and macrophage ($p < 0.05$). In addition, in a previous study, we showed that *PIK3R1* was highly expressed in the DCM group, suggesting that inhibiting the expression of *PIK3R1* may play a therapeutic role in DCM by suppressing the adaptive immune response. In contrast, the expression of *ERBB2* was negatively correlated with the activation of the immune system, which argues that *ERBB2* expression plays a protective role against DCM from the perspective of immune cell infiltration, consistent with the literature. Overall, from an immune system perspective, we validated the potential immune mechanisms of *PIK3R1* and *ERBB2* to provide a theoretical basis for future studies.

Finally, we evaluated druggable DCM cell subgroups using the "Drug-DCM" gene set. The results showed that the "Drug-DCM" gene set had a higher enrichment score in cells with contractile function, consistent with our findings that cells with contractile function are potential targets for drug therapy. However, despite the inclusion of studies on the therapeutic effects of curcumin in cardiovascular disease in the introduction, our assessment of druggable genes in DCM based on curcumin's drug targets is still limited. This is due to the fact that curcumin, being a single traditional Chinese drug, has not been fully validated by high-evidence studies for its therapeutic role in DCM. This phenomenon can be attributed, in part, to the classification of curcumin as a BCS class II drug, characterized by inadequate intestinal absorption, swift metabolism, and substantial systemic elimination, thereby constraining its bioavailability. Nevertheless, it is important to consider that different drugs may target common disease targets. Therefore, we constructed three extensive PPI networks and analyzed the disease genes, rather than the drug genes, within the sub-networks. Furthermore, it should be noted that our Mendelian randomization study exclusively incorporated the GWAS cohort from a substantial European population. Nevertheless, it is crucial to acknowledge that variations in ethnicity could potentially influence the outcomes. Consequently, it is imperative to conduct prospective studies encompassing extensive sample sizes or additional molecular and human trials to comprehensively assess the implications of our findings.

4. Conclusion

In conclusion, our study examined the pathogenic and drug-treatable genes associated with DCM through the utilization of big data analysis and pharmacological approaches. The findings of this research offer a foundation for the development of novel DCM drugs aimed at enhancing the prognosis of individuals affected by this condition.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability statement

The data were obtained from the GEO and UK Biobank online repositories. The data and login links can be found in the article and supplementary materials. Data included in article/supp. material/referenced in article.

Funding

This study was sponsored by funds from the Chinese National Natural Science Foundation of China (no.82360080), the Natural Science Foundation of Guangxi Province (no. 2023GXNSFAA026010and 2018GXNSFBA050059), Middle-Aged and Young Teachers in Colleges and Universities in Guangxi Basic Ability Promotion Project (no.2021KY0534), the First Batch of High-level Talent Scientific Research Projects of the Affiliated Hospital of Youjiang Medical University for Nationalities in 2019((no.R20196316).

CRediT authorship contribution statement

Bin He: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Liping Quan:** Software, Data curation. **Chengban Li:** Software. **Wei Yan:** Investigation, Formal analysis. **ZhuoHua Zhang:** Validation, Software. **LiuFan Zhou:** Formal analysis, Data curation. **Qinjiang Wei:** Software, Formal analysis. **Zhile Li:** Software, Formal analysis.

Jianjiao Mo: Visualization, Investigation. **Zhen Zhang:** Software, Conceptualization. **Xingshou Pan:** Supervision, Conceptualization. **JianJun Huang:** Validation, Supervision. **Li Liu:** Writing – review & editing, Conceptualization.

Declaration of competing interest

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

Acknowledgments

It was the GEO network and UK Biobank online repositories that helped make this study possible. We have asked the researcher's house for help with the language. Thanks to FreeScience (https://www.home-for-researchers.com) for helping with the English language.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e25572.

References

[1] A.G. Japp, A. Gulati, S.A. Cook, M.R. Cowie, S.K. Prasad, The diagnosis and evaluation of dilated cardiomyopathy, J. Am. Coll. Cardiol. 67 (25) (2016) 2996–3010, https://doi.org/10.1016/j.jacc.2016.03.590.
[2] D. Reichart, C. Magnussen, T. Zeller, S. Blankenberg, Dilated cardiomyopathy: from epidemiologic to genetic phenotypes: a translational review of current literature, J. Intern. Med. 286 (4) (2019) 362–372, https://doi.org/10.1111/joim.12944.
[3] R.G. Weintraub, C. Semsarian, P. Macdonald, Dilated cardiomyopathy, Lancet (London, England) 390 (10092) (2017) 400–414, https://doi.org/10.1016/S0140-6736(16)31713-5.
[4] V. Regitz-Zagrosek, C. Hengstenberg, T.P. Cappola, R. Isnard, E. Arbustini, S.A. Cook, et al., Genome-wide association analysis in dilated cardiomyopathy reveals two new players in systolic heart failure on chromosomes 3p25.1 and 22q11.23, Eur. Heart J. 42 (20) (2021) 2000–2011, https://doi.org/10.1093/eurheartj/ehab030.
[5] U. Tayal, S. Prasad, S.A. Cook, Genetics and genomics of dilated cardiomyopathy and systolic heart failure, Genome Med. 9 (1) (2017) 20, https://doi.org/10.1186/s13073-017-0410-8.
[6] J. Ramchand, M. Wallis, I. Macciocca, E. Lynch, O. Farouque, M. Martyn, et al., Prospective evaluation of the utility of whole exome sequencing in dilated cardiomyopathy, J. Am. Heart Assoc. 9 (2) (2020) e013346, https://doi.org/10.1161/JAHA.119.013346.
[7] E. Martini, P. Kunderfranco, C. Peano, P. Carullo, M. Cremonesi, T. Schorn, et al., Kallikourdis, single-cell sequencing of Mouse heart immune infiltrate in Pressure overload-driven heart failure reveals extent of immune activation, Circulation 140 (25) (2019) 2089, https://doi.org/10.1161/CIRCULATIONAHA.119.041694. –2107.
[8] E. Shapiro, T. Biezuner, S. Linnarsson, Single-cell sequencing-based technologies will revolutionize whole-organism science, Nat. Rev. Genet. 14 (9) (2013) 618–630, https://doi.org/10.1038/nrg3542.
[9] B. Kocaadam, N. Şanlier, Curcumin, an active component of turmeric (Curcuma longa), and its effects on health, Crit. Rev. Food Sci. Nutr. 57 (13) (2017) 2889–2895, https://doi.org/10.1080/10408398.2015.1077195.
[10] H. Li, A. Sureda, H.P. Devkota, V. Pittalà, D. Barreca, A.S. Silva, et al., Curcumin, the golden spice in treating cardiovascular diseases, Biotechnol. Adv. 38 (2020) 107343, https://doi.org/10.1016/j.biotechadv.2019.01.010.
[11] Y. Sunagawa, M. Funamoto, S. Sono, K. Shimizu, S. Shimizu, M. Genpei, et al., Curcumin and its demethoxy derivatives possess p300 HAT inhibitory activity and suppress hypertrophic responses in cardiomyocytes, J. Pharmacol. Sci. 136 (4) (2018) 212–217, https://doi.org/10.1016/j.jphs.2017.12.013.
[12] J.W. Thompson, J. Wei, K. Appau, H. Wang, H. Yu, M.G. Spiga, et al., Bnip3 binds and activates p300: possible role in cardiac transcription and myocyte morphology, PLoS One 10 (8) (2015) e0136847, https://doi.org/10.1371/journal.pone.0136847.
[13] A. Ray, S. Rana, D. Banerjee, A. Mitra, R. Datta, S. Naskar, et al., Improved bioavailability of targeted Curcumin delivery efficiently regressed cardiac hypertrophy by modulating apoptotic load within cardiac microenvironment, Toxicol. Appl. Pharmacol. 290 (2016) 54–65, https://doi.org/10.1016/j.taap.2015.11.011.
[14] X.J. Bai, J.T. Hao, J. Wang, Wang, W.F. Zhang, C.P. Yan, J.H. Zhao, et al., Curcumin inhibits cardiac hypertrophy and improves cardiovascular function via enhanced Na^{+}/Ca^{2+} exchanger expression after transverse abdominal aortic constriction in rats, Pharmacol. Rep. : PR. 70 (1) (2018) 60–68, https://doi.org/10.1016/j.pharep.2017.07.014.
[15] N.M. Davies, M.V. Holmes, G. Davey Smith, Reading Mendelian Randomisation Studies: a Guide, Glossary, and Checklist for Clinicians, *BMJ*, vol. 362, Clinical research ed., 2018, p. k601, https://doi.org/10.1136/bmj.k601.
[16] L. Laugier, A.F. Frade, F.M. Ferreira, M.A. Baron, P.C. Teixeira, S. Cabantous, et al., Whole-genome cardiac DNA methylation fingerprint and gene expression analysis provide new insights in the pathogenesis of chronic chagas disease cardiomyopathy, Clin. Infect. Dis. : Official Pub. Infect. Dis. Soc. America 65 (7) (2017) 1103–1111, https://doi.org/10.1093/cid/cix506.
[17] A. Gaertner, P. Schwientek, P. Ellinghaus, H. Summer, S. Golz, A. Kassner, Kassner, et al., Myocardial transcriptome analysis of human arrhythmogenic right ventricular cardiomyopathy, Physiol. Genom. 44 (1) (2012) 99–109, https://doi.org/10.1152/physiolgenomics.00094.2011.
[18] M.M. Molina-Navarro, E. Roselló-Lletí, A. Ortega, E. Tarazón, M. Otero, L. Martínez-Dolz, et al., Differential gene expression of cardiac ion channels in human dilated cardiomyopathy, PLoS One 8 (12) (2013) e79792, https://doi.org/10.1371/journal.pone.0079792.
[19] E. Witt, E. Hammer, M. Dörr, K. Weitmann, D. Beug, K. Lehnert, et al., Correlation of gene expression and clinical parameters identifies a set of genes reflecting LV systolic dysfunction and morphological alterations, Physiol. Genom. 51 (8) (2019) 356–367, https://doi.org/10.1152/physiolgenomics.00111.2018.
[20] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, Nat. Biotechnol. 36 (5) (2018) 411–420, https://doi.org/10.1038/nbt.4096.
[21] Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, et al., Integrated analysis of multimodal single-cell data, Cell 184 (13) (2021) 3573–3587, https://doi.org/10.1016/j.cell.2021.04.048. .e29.
[22] D. Aran, A.P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, et al., Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage, Nat. Immunol. 20 (2) (2019) 163–172, https://doi.org/10.1038/s41590-018-0276-y.

[23] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47, https://doi.org/10.1093/nar/gkv007.
[24] C. Chen, K. Grennan, J. Badner, D. Zhang, E. Gershon, L. Jin, et al., Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods, PLoS One 6 (2) (2011) e17238, https://doi.org/10.1371/journal.pone.0017238.
[25] K. Ito, D. Murphy, Application of ggplot2 to pharmacometric graphics, CPT Pharmacometrics Syst. Pharmacol. 2 (10) (2013) e79, https://doi.org/10.1038/psp.2013.56.
[26] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559, https://doi.org/10.1186/1471-2105-9-559.
[27] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, et al., clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innovation 2 (3) (2021) 100141, https://doi.org/10.1016/j.xinn.2021.100141.
[28] A. Liberzon, A description of the molecular signatures database (MSigDB) web site, Methods Mol. Biol. 1150 (2014) 153–160, https://doi.org/10.1007/978-1-4939-0512-6_9.
[29] G.D. Bader, C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinf. 4 (2003) 2, https://doi.org/10.1186/1471-2105-4-2.
[30] A.H. Wagner, A.C. Coffman, B.J. Ainscough, N.C. Spies, Z.L. Skidmore, K.M. Campbell, et al., DGIdb 2.0: mining clinically relevant drug-gene interactions, Nucleic Acids Res. 44 (D1) (2016) D1036–D1044, https://doi.org/10.1093/nar/gkv1165.
[31] G. Bindea, B. Mlecnik, M. Tosolini, A. Kirilovsky, M. Waldner, A.C. Obenauf, et al., Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer, Immunity 39 (4) (2013) 782–795, https://doi.org/10.1016/j.immuni.2013.10.003.
[32] T.M. Palmer, D.A. Lawlor, R.M. Harbord, N.A. Sheehan, J.H. Tobias, N.J. Timpson, et al., Using multiple genetic variants as instrumental variables for modifiable risk factors, Stat. Methods Med. Res. 21 (3) (2012) 223–242, https://doi.org/10.1177/0962280210394459.
[33] J. Yang, T. Ferreira, A.P. Morris, S.E. Medland, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, P. A. Madden, et al., Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits, Nat. Genet. 44 (4) (2012) 369, https://doi.org/10.1038/ng.2213. S3.
[34] J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression, Int. J. Epidemiol. 44 (2) (2015) 512–525, https://doi.org/10.1093/ije/dyv080.
[35] E. Papalexi, R. Satija, Single-cell RNA sequencing to explore immune cell heterogeneity, Nat. Rev. Immunol. 18 (1) (2018) 35–45, https://doi.org/10.1038/nri.2017.76.
[36] D.T. Paik, S. Cho, L. Tian, H.Y. Chang, J.C. Wu, Single-cell RNA sequencing in cardiovascular development, disease and medicine, Nat. Rev. Cardiol. 17 (8) (2020) 457–473, https://doi.org/10.1038/s41569-020-0359-y.
[37] E.M. McNally, J.R. Golbus, M.J. Puckelwartz, Genetic mutations and mechanisms in dilated cardiomyopathy, J. Clin. Invest. 123 (1) (2013) 19–26, https://doi.org/10.1172/JCI62862.
[38] J.S. Ware, S.A. Cook, Role of titin in cardiomyopathy: from DNA variants to patient stratification, Nat. Rev. Cardiol. 15 (4) (2018) 241–252, https://doi.org/10.1038/nrcardio.2017.190.
[39] E.M. McNally, L. Mestroni, Dilated cardiomyopathy: genetic determinants and mechanisms, Circ. Res. 121 (7) (2017) 731–748, https://doi.org/10.1161/CIRCRESAHA.116.309396.
[40] T. Suzuki, M. Aikawa, M. Kuro-o, M. Watanabe, K. Kimura, Y. Yazaki, R. Nagai, Presence of contractile-type smooth muscle cells in the endocardium, Cardiology 87 (1) (1996) 23–27, https://doi.org/10.1159/000177055.
[41] H. Okada, G. Takemura, H. Kanamori, A. Tsujimoto, K. Goto, I. Kawamura, et al., Phenotype and physiological significance of the endocardial smooth muscle cells in human failing hearts, Circulation. Heart failure 8 (1) (2015) 149–155, https://doi.org/10.1161/CIRCHEARTFAILURE.114.001746.
[42] D. Harding, M.H.A. Chong, N. Lahoti, C.M. Bigogno, R. Prema, S.A. Mohiddin, Dilated cardiomyopathy and chronic cardiac inflammation: pathogenesis, diagnosis and therapy, J. Intern. Med. 293 (1) (2023) 23–47, https://doi.org/10.1111/joim.13556.
[43] B. He, L.P. Quan, C.Y. Cai, D.Y. Yu, W. Yan, Q.J. Wei, et al., Dysregulation and imbalance of innate and adaptive immunity are involved in the cardiomyopathy progression, Frontiers in cardiovascular medicine 9 (2022) 973279, https://doi.org/10.3389/fcvm.2022.973279.
[44] L. Louzao-Martinez, A. Vink, M. Harakalova, F.W. Asselbergs, M.C. Verhaar, C. Cheng, Characteristic adaptations of the extracellular matrix in dilated cardiomyopathy, Int. J. Cardiol. 220 (2016) 634–646, https://doi.org/10.1016/j.ijcard.2016.06.253.
[45] C.L. Neal, J. Xu, P. Li, S. Mori, J. Yang, N.N. Neal, et al., Overexpression of 14-3-3ζ in cancer cells activates PI3K via binding the p85 regulatory subunit, Oncogene 31 (7) (2012) 897–906, https://doi.org/10.1038/onc.2011.284.
[46] T.A. Donlon, R. Chen, K.H. Masaki, B.J. Willcox, B.J. Morris, Association with longevity of phosphatidylinositol 3-kinase regulatory subunit 1 gene variants stems from protection against mortality risk in men with cardiovascular disease, Gerontology 68 (2) (2022) 162–170, https://doi.org/10.1159/000515390.
[47] C.J. Nunes-Santos, G. Uzel, S.D. Rosenzweig, PI3K pathway defects leading to immunodeficiency and immune dysregulation, J. Allergy Clin. Immunol. 143 (5) (2019) 1676–1687, https://doi.org/10.1016/j.jaci.2019.03.017.
[48] M.C. Deau, L. Heurtier, P. Frange, F. Suarez, C. Bole-Feysot, P. Nitschke, et al., A human immunodeficiency caused by mutations in the PIK3R1 gene, J. Clin. Invest. 125 (4) (2015) 1764–1765, https://doi.org/10.1172/JCI81746.
[49] C.L. Lucas, A. Chandra, S. Nejentsev, A.M. Condliffe, K. Okkenhaug, PI3Kδ and primary immunodeficiencies, Nat. Rev. Immunol. 16 (11) (2016) 702–714, https://doi.org/10.1038/nri.2016.93.
[50] D.A. Dyment, A.C. Smith, D. Alcantara, J.A. Schwartzentruber, L. Basel-Vanagaite, C.J. Curry, et al., Mutations in PIK3R1 cause SHORT syndrome, Am. J. Hum. Genet. 93 (1) (2013) 158–166, https://doi.org/10.1016/j.ajhg.2013.06.005.
[51] J.N. Winnay, M.H. Solheim, E. Dirice, M. Sakaguchi, H.L. Noh, H.J. Kang, et al., PI3-kinase mutation linked to insulin and growth factor resistance in vivo, J. Clin. Invest. 126 (4) (2016) 1401–1412, https://doi.org/10.1172/JCI84005.
[52] Y. Terauchi, Y. Tsuji, S. Satoh, H. Minoura, K. Murakami, A. Okuno, et al., Increased insulin sensitivity and hypoglycaemia in mice lacking the p85 alpha subunit of phosphoinositide 3-kinase, Nat. Genet. 21 (2) (1999) 230–235, https://doi.org/10.1038/6023.
[53] X. Ai, L. Xiang, Z. Huang, S. Zhou, S. Zhang, T. Zhang, et al., Overexpression of PIK3R1 promotes hepatocellular carcinoma progression, Biol. Res. 51 (1) (2018) 52, https://doi.org/10.1186/s40659-018-0202-7.
[54] H. Zhan, F. Huang, Q. Niu, M. Jiao, X. Han, K. Zhang, et al., Downregulation of miR-128 ameliorates ang II-induced cardiac remodeling via SIRT1/PIK3R1 multiple targets, Oxid. Med. Cell. Longev. 2021 (2021) 8889195, https://doi.org/10.1155/2021/8889195.
[55] S.A. Crone, Y.Y. Zhao, L. Fan, Y. Gu, S. Minamisawa, Y. Liu, et al., ErbB2 is essential in the prevention of dilated cardiomyopathy, Nat. Med. 8 (5) (2002) 459–465, https://doi.org/10.1038/nm0502-459.
[56] A. Negro, B.K. Brar, Y. Gu, K.L. Peterson, W. Vale, K.F. Lee, erbB2 is required for G protein-coupled receptor signaling in the heart, Proc. Natl. Acad. Sci. U.S.A. 103 (43) (2006) 15889–15893, https://doi.org/10.1073/pnas.0607499103.
[57] P. Sysa-Shah, C.G. Tocchetti, M. Gupta, P.P. Rainer, X. Shen, B.H. Kang, et al., Bidirectional cross-regulation between ErbB2 and β-adrenergic signalling pathways, Cardiovasc. Res. 109 (3) (2016) 358–373, https://doi.org/10.1093/cvr/cvv274.
[58] C. Ozcelik, B. Erdmann, B. Pilz, N. Wettschureck, S. Britsch, N. Hübner, et al., Conditional mutation of the ErbB2 (HER2) receptor in cardiomyocytes leads to dilated cardiomyopathy, Proc. Natl. Acad. Sci. U.S.A. 99 (13) (2002) 8880–8885, https://doi.org/10.1073/pnas.122249299.
[59] Z. Vermeulen, V.F. Segers, G.W. De Keulenaer, ErbB2 signaling at the crossing between heart failure and cancer, Basic Res. Cardiol. 111 (6) (2016) 60, https://doi.org/10.1007/s00395-016-0576-z.
[60] S.M. Jay, E. Kurtagic, L.M. Alvarez, S. de Picciotto, E. Sanchez, J.F. Hawkins, et al., Engineered bivalent ligands to bias ErbB receptor-mediated signaling and phenotypes, J. Biol. Chem. 286 (31) (2011) 27729–27740, https://doi.org/10.1074/jbc.M111.221093.

[61] S.M. Jay, A.C. Murthy, J.F. Hawkins, J.R. Wortzel, M.L. Steinhauser, L.M. Alvarez, et al., An engineered bivalent neuregulin protects against doxorubicin-induced cardiotoxicity with reduced proneoplastic potential, Circulation 128 (2) (2013) 152–161, https://doi.org/10.1161/CIRCULATIONAHA.113.002203.
[62] Y. Liu, D. Wang, Z. Li, X. Li, M. Jin, N. Jia, et al., Pan-cancer analysis on the role of PIK3R1 and PIK3R2 in human tumors, Sci. Rep. 12 (1) (2022) 5924, https://doi.org/10.1038/s41598-022-09889-0.
[63] E. Wang, R. Zhou, T. Li, Y. Hua, K. Zhou, Y. Li, et al., The molecular role of immune cells in dilated cardiomyopathy, Medicina (Kaunas, Lithuania) 59 (7) (2023) 1246, https://doi.org/10.3390/medicina59071246.