Integrative analyses of biomarkers and pathways for heart failure

Shaowei Fan
China Academy of Chinese Medical Sciences
Yuanhui Hu (huiyuhui55@sohu.com)
China Academy of Chinese Medical Sciences

Research Article

Keywords: potential key genes, heart failure, bioinformatics, gene expression synthesis, biomarkers

Posted Date: October 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1031511/v1

Version of Record: A version of this preprint was published at BMC Medical Genomics on March 27th, 2022. See the published version at https://doi.org/10.1186/s12920-022-01221-z.

Integrative analyses of biomarkers and pathways for heart failure

Shaowei Fan,[1] Yuanhui Hu[1,*]

[1]Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China
*Correspondence: Yuanhui Hu huiyuhui55@sohu.com

Abstract

Background: Heart failure (HF) is the most common potential cause of death, causing a huge health and economic burden all over the world. So far, some impressive progress has been made in the study of pathogenesis. However, the underlying molecular mechanisms leading to this disease remain to be fully elucidated.

Methods: The microarray data sets of GSE76701, GSE21610 and GSE8331 were retrieved from the gene expression comprehensive database (GEO). After merging all microarray data and adjusting batch effects, differentially expressed genes (DEG) were determined. Functional enrichment analysis was performed based on Gene Ontology (GO) resources, Kyoto Encyclopedia of Genes and Genomes (KEGG) resources, gene set enrichment analysis (GSEA), response pathway database and Disease Ontology (DO). Protein protein interaction (PPI) network was constructed using string database. Combined with the above important bioinformatics information, the potential key genes were selected. The comparative toxicological genomics database (CTD) is used to explore the interaction between potential key genes and HF.

Results: We identified 38 patients with heart failure and 16 normal controls. There were 315 DEGs among HF samples, including 278 up-regulated genes and 37 down-regulated genes. Pathway enrichment analysis showed that most DEGs were significantly enriched in BMP signal pathway, transmembrane receptor protein serine / threonine kinase signal pathway, extracellular matrix, basement membrane, glycosaminoglycan binding, sulfur compound binding and so on. Similarly, GSEA enrichment analysis showed that DEGs were mainly enriched in extracellular matrix and extracellular matrix related proteins. BBS9, CHRD, BMP4, MYH6, NPPA and CCL5 are central genes in PPI networks and modules.

Conclusions: the enrichment pathway of DEGs and go ontology may reveal the molecular mechanism of HF. Among them, target genes EIF1AY, RPS4Y1, USP9Y, KDM5D, DDX3Y, NPPA, HBB, TSIX, LOC28556 and XIST are expected to become new targets for heart failure. Our findings provide potential biomarkers or therapeutic targets for the further study of heart failure and contribute to the development of advanced prediction, diagnosis and treatment strategies.

Keywords: potential key genes, heart failure, bioinformatics, gene expression synthesis, biomarkers

Background

Cardiovascular disease is one of the main causes of human death, including coronary heart disease, hypertension, congenital heart disease, heart failure and other heart related diseases, as well as systemic vascular system related diseases such as atherosclerosis and lower extremity deep venous thrombosis. Among them, heart failure is the most common cardiovascular disease in clinic. It is estimated that about 64.3 million people worldwide suffer from heart failure[1]. As everyone knows, heart failure often follows a variety of other diseases, such as coronary heart disease, hypertension, diabetes, etc. these diseases are characterized by an obvious age-dependent dependency. That is, the higher the risk factor for age, the greater the risk of diseases. However, research shows that the burden of heart failure in young people may be increasing, which means that more young people will enter the ranks of heart failure, and its prevalence tends to be younger[2].

In terms of the mechanism of heart failure, the main reason is the myocardial injury (such as myocardial ischemia and necrosis), the overload of pressure and the overload of the volume, which result in the change of the structure and function of the myocardium, and eventually lead to the dysfunction of the blood pumping or filling of the ventricle, and the main clinical manifestations are fatigue, dyspnea and fluid retention. Among them, myocardial hypertrophy is the key to the occurrence and development of heart failure, but the exact reason for the transformation of myocardial hypertrophy into heart failure is not clear. In this process, it may play a role in the pathological increase of the circulating level of vasoactive substances (such as angiotensin II, catecholamine, endothelin, etc.), and the increase of the content of vasoactive substances will stimulate the myocardium and eventually lead to myocardial hypertrophy by stimulating their respective signal transduction pathways[3]; Oxygen free radicals formed in the early stage of pathological stimulation are involved in the occurrence and development of myocardial hypertrophy; Others include functional hypoxia, metabolic disorder, mitochondrial electron transport uncoupling and inflammation, the oxidation of catecholamine by monoamine oxidase and the activation of NADPH oxidase by angiotensin II and endothelin, which may be involved in the process of heart failure with myocardial hypertrophy as the trigger point. Conversely, cardiac hypertrophy itself can become a new key point to trigger heart failure. When hypertrophic cardiomyocytes are exposed to the environment with increased content of vasoactive substances for a long time, it will lead to subcellular defects, protease activation, metabolic disorders, abnormal calcium regulation, etc., which will increase the dysfunction of heart function, thus aggravating the occurrence and development process of heart failure[4]. Therefore, the cytokines and other molecular mechanisms involved in the occurrence and development of heart failure still need to be further studied, and its specific mechanism needs to be further proved at the molecular level[5]. Heart failure is a major health problem in the world. It is necessary to explore the potential biomarkers and molecular mechanisms related to the occurrence and development of heart failure, so as to provide more targeted and effective treatment strategies for patients and bring benefits. At present, high-throughput technology has developed rapidly and is widely used in various fields. Integrated bioinformatics analysis is expected to become a key technology to clarify the etiology, pathogenesis and treatment of heart failure, which will benefit mankind in terms of human, material and financial resources.

In this study, we analyzed three mRNA expression profiles (GSE76701, GSE21610 and GSE8331) of the same platform (GPL570) downloaded from GEO Database (https://www.ncbi.nlm.nih.gov/geo/) to determine the possible DEGs in the occurrence and development of heart failure, and analyzed their expression, function and interaction, so as to

provide reference for exploring biomarkers or therapeutic targets of heart failure.

Methods

Figure 1 clearly illustrates the flow chart of materials and methods.

Heart failure dataset

Download the heart failure dataset original files of three registered microarray datasets from NCBI GEO Database, including GSE76701, GSE21610 and GSE8331 (Table 1). All these datasets are from the microarray platform of Affymetrix Human Genome U133 Plus 2.0 Array [HG-U133_Plus_2]. In each dataset, human myocardial samples were selected only from HF and normal EF subjects, and finally 38 HF and 16 normal EF group samples were included for subsequent analysis.

Data preprocessing

The GSE76701 dataset contains 4 human HF samples and 4 human health samples, the GSE21610 dataset contains 30 human HF samples and 8 human health samples, and the GSE8331 dataset contains 4 human HF samples and 4 human health samples. A series of matrix text files for the dataset have been obtained. Subsequently, the limma R software package was used for background correction, quartile standardization and probe summary [6-9]. See **Table 1** for details.

Identification of DEGs between HF and healthy samples

In this study, we used the limma R package (version 3.6.3; https://www.r-project.org/)[6]. The DEGs between HF samples and healthy samples were determined by the threshold standard of | log2 (FC) | > 1 and p.adj < 0.05.

Function and pathway enrichment analysis of DEGs

Gene Ontology (GO) resources (http://geneontology.org/, Accessed 20 Oct 2021) is a bioinformatics tool that provides a framework and a set of concepts to describe the function of all biological gene products[10]. Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp/) is a database resource integrated the information of genomes, biological pathways, diseases and chemicals[11]. Reactome pathway database (https://reactome.org/) is a path annotation database that collects human biological paths and processes[12]. Enrichr (https://maayanlab.cloud/Enrichr/, Accessed 20 Oct 2021) is an online data processor[13-15]. Disease Ontology (DO) (http://disease-ontology.org, Accessed 20 Oct 2021), represents a comprehensive knowledge base of 8043 genetic, developmental and acquired human diseases[16]. Gene Set Enrichment Analysis (GSEA) (http://software.broadinstitute.org/gsea/index.jsp, Accessed 20 Oct 2021) is a computational method for interpreting gene expression data based on molecular signature database[17, 18]. Before enrichment analysis, the gene symbol code is converted into Entrez ID using the human genome annotation package "org. HS. eg.db". In order to better understand the biological functions and characteristics, the enrichment analysis is carried out with R software, using "Clusterprofiler" Go enrichment analysis package, reactor pathway analysis "Reactomepa" package and do enrichment analysis "DOSE" package. The "GOplot" and "ggplot2" packages of R software are used for visual mapping. The relevant go biological function map is considered to be significantly rich if it meets the p. adjust value < 0.05 and Q value < 0.05. For the important paths related to heart failure, the significance level, nominal p value and false discovery rate The cutoff value of (FDR) Q value is 0.05. For GSEA enrichment analysis, FDR Q value < 0.25 and p.adjust value < 0.05 are used as screening indexes.

Protein-protein interaction (PPI) network and potential key gene analysis

String database (http://string-db.org/, Accessed 20 Oct 2021) is used to construct PPI network to

reveal the general organization principle of functional cell system and predict protein-protein interaction[19]. Through molecular complex detection (MCODE) of Cytoscape, the results of PPI network are modularized analyzed and visualized. Default parameters (degree cutoff) ≥ 2, node score cut-off ≥ 2, K nucleus ≥ 2, maximum depth = 100). In order to select potential key genes, we synthesized the above important bioinformatics information for subsequent analysis. If DEG meets the inclusion criteria (adjusted p value < 0.05 and | log2 FC | ≥ 1), it is considered as potential key genes. In addition, genes with connectivity greater than 5 in PPI network are also included.

Identification of potential key genes associated with heart failure

The Comparative Toxicogenomics Database (CTD, http://ctdbase.org/, Accessed 20 Oct 2021) synthesize information, including chemical gene / protein interactions, chemical diseases and gene disease relationships, to develop hypotheses related to disease mechanisms[20]. Use data in CTD to analyze the association between potential key genes and the risk of heart failure, atrial fibrillation, hypertension and sudden cardiac death.

Results

Dataset evaluation

The gene expression level of the combined GEO series with adjusted batch effect is standardized, and the results before and after standardization are shown in **Figure 2**. Probes corresponding to 21655 genes in GSE76701, GSE21610 and GSE8331 datasets were identified, and DEGs of heart failure were confirmed. The total number of filtered molecules was 21655, of which 85 IDS met the threshold of | log2 (FC) | ≥ 1 & p.adj < 0.05. Under this threshold, 60 were highly expressed in HF group and 25 in normal group; 22 IDS met the threshold of | log2 (FC) | ≥ 1.5 & p.adj < 0.05. Under this threshold, 16 IDS were highly expressed in HF group and 6 IDS were highly expressed in normal group; There are 10 IDs that meet the threshold of | log2 (FC) | ≥ 2 & p.adj < 0.05. Under this threshold, there are 7 highly expressed IDS in HF group (EIF1AY, RPS4Y1, USP9Y, KDM5D, DDX3Y, NPPA and HBB) and 3 highly expressed IDS in normal group (TSIX, LOC28556 and XIST). See **Table 2** for details. **Figure 3** shows the Heatmap plot, Volcano plot, PAC plot and UMAP plot.

Figure 2 Boxplots of gene expressions before and after standardization for 3 selected GEO Database. (A. Before standardization; B. After standardization. NG: Normal Group; HF: Heart Failure Group)

Figure 3 The PCA plot, UMAP plot, Volcano plot and Heatmap plot of gene expressions for 3 selected GEO Database with a screening criteria of |log2 FC| ≥ 1 and adjust P value < 0.05.
(A. PCA plot; B. UMAP plot; C. Volcano plot; D. Heatmap plot. PCA: Principal Component Analysis; UMAP: Uniform Manifold Approximation and Projection; NG: Normal Group; HF: Heart Failure Group)

Functional enrichment analysis of DEGs

GO enrichment analysis

In order to further study the biological functions of 10 DEGs, functional enrichment analysis was carried out, and the results are shown in **Table 3**. GO enrichment analysis showed that the functions of differentially expressed genes were mainly concentrated in the following 11 aspects: GO: 0030509 ~ BMP signaling pathway; GO: 0071772 ~ response to BMP; GO: 0071773 ~ cellular response to BMP stimulus; GO: 0003012 ~ muscle system process; GO: 0007178 ~ transmembrane receptor protein serine / threonine kinase signaling pathway; GO: 0062023 ~ collagen containing

extracellular matrix; GO: 0005604 ~ basement membrane; GO: 0005614 ~ interstitial matrix; GO: 0008201 ~ heparin binding; GO: 0005539 ~ glycosaminoglycan binding; GO: 1901681 ~ sulfur compound binding. For the above GO meeting the requirements, the R language GOplot and ggplot 2 package are used for visual presentation (**Fig. 4**). According to the adjusted screening criteria of P value < 0.05 and Q value < 0.05, there was no enrichment pathway in KEGG.

Figure 4 Enrichment plots by GO.
(A. GO Bar graph, B. Bubble plot, C. Bar graph of GO enrichment pathways, D. chord diagram, E. loop graph. BP: Biological Process; CC: Cellular Component; MF: Molecular Function)

GSEA enrichment analysis

GSEA was used to test the combined GEO dataset to identify functional gene sets associated with heart failure. Finally, five groups of HF related expressions were determined (**Table 4**), of which only three data sets met FDR Q value < 0.25 and p.adjust value < 0.05. These data sets include: (1) involved in encoding core extracellular matrix (including ECM glycoprotein, collagen and proteoglycan), (2) involved in encoding structural ECM glycoprotein, and (3) involved in encoding extracellular matrix and extracellular matrix related proteins (Fig. 3). According to the visualization results of GSEA, the eligible GSEA gene sets are as follows: (1) in NABA_ CORE_ Matrix gene set was significantly enriched (NES = 2.199; p.adjust = 0.037; FDR = 0.035); (2) At NABA_ ECM_ The glycoproteins gene set was significantly enriched (NES = 2.050; p.adjust = 0.037; FDR = 0.035) **(Fig. 5)**.

Figure 5 Enrichment plots by GSEA.
(A. GSEA Visual Analysis, B. GSEA ridgeplot. NES: Normal Enrichment Score; FDR: False Discovery Rate)

Enrichment analysis by Enrichr and Reactome

Using Enrichr through online enrichment analysis, it was identified that 7 highly expressed IDS in HF group (EIF1AY, RPS4Y1, USP9Y, KDM5D, DDX3Y, NPPA and HBB) were related to Erythrocytes take up oxygen and release carbon dioxide, Physiological factors, Erythrocytes take up carbon dioxide and release oxygen, O2/CO2 exchange in erythrocytes, HDMs demethylate histones, YAP1-and WWTR1 (TAZ)-stimulated gene expression; 7 highly expressed IDS in normal group (TSIX, LOC28556 and XIST) were related to Fatty acid metabolism (**Fig. 6**). The ten DEGs identified were enriched and analyzed by Reactome database(https://reactome.org/), and the related biological processes of HF were not enriched.

Figure 6 Enrichment pathways by Enrichr.
(Upper: Related enrichment pathways of up-regulated genes; Lower: Related enrichment pathways of down-regulated genes)

PPI network construction and Hub gene selection

PPI analysis was performed on these DEGs using the string platform, and 42 nodes and 41 interactions were finally determined (**Fig. 7**). In addition, an important module with 3 nodes and 3 edges is selected through MCODE. Two important modules with 6 nodes and 7 edges are selected

through MCODE. BBS2, BBS7 and BBS9 are the hub nodes in module B, FRZB, CHRD, BMP4, MYH6, SLN and NPPA are the hub nodes in module C, and KLRB1, CD3D, CCL5, C3, CFH and FCN3 are the hub nodes in module D. Only BBS9, CHRD, BMP4, MYH6, NPPA and CCL5 were selected as hub genes. In addition, combined with the results of differential expression, enrichment analysis and PPI, BBS9, CHRD, BMP4, MYH6, NPPA and CCL5 were considered as Hub genes for further analysis.

Figure 7 PPI network construction.
(A. The complete PPI network, B-D. The three modules of hub genes)

Identification of potential key genes associated with heart failure

CTD is used to explore the interaction between potential key genes and heart failure. As shown in **Figure 8**, potential key genes for heart failure, atrial fibrillation, hypertension, myocardial infarction, sudden cardiac death and myocarditis. The reasoning score in CTD reflects the association between chemicals, diseases and genes. The results of interaction showed that NPPA, HBB, DDX3Y and XIST had higher scores with heart failure.

Figure 8 The CTD analysis between potential key genes and diseases.

Discussion

In this study, we integrated the gene expression profiles of 38 HF samples and 16 normal samples from 3 geo databases, and analyzed the data using bioinformatics tools. 85 IDS met the threshold of $|\log2(FC)| \geq 1$ & $p.adj < 0.05$. Under this threshold, 60 IDS were highly expressed in HF group and 25 IDS were highly expressed in normal group; 22 IDS met the threshold of $|\log2(FC)| \geq 1.5$ & $p.adj < 0.05$. Under this threshold, 16 IDS were highly expressed in HF group and 6 IDS were highly expressed in normal group; There are 10 IDs that meet the threshold of $|\log2(FC)| \geq 2$ & $p.adj < 0.05$. Under this threshold, there are 7 highly expressed IDS in HF group (EIF1AY, RPS4Y1, USP9Y, KDM5D, DDX3Y, NPPA, HBB) and 3 highly expressed IDS in normal group (TSIX, LOC28556, XIST). These 10 potential key genes (EIF1AY, RPS4Y1, USP9Y, KDM5D, DDX3Y, NPPA, HBB, TSIX, LOC28556, XIST) and some important pathways related to the risk of heart failure have been identified, indicating that these may play an important role in the mechanism of the occurrence and development of heart failure.

Specific genes on the Y chromosome may be a risk factor for heart failure

EIF1AY, RPS4Y1, USP9Y, KDM5D[21] and DDX3Y, which are highly expressed in HF group, are located on Y chromosome. Considering that men account for a high proportion in the sample, such as data set GSE21610. However, gender is indeed a factor that cannot be ignored in cardiovascular diseases, especially in cardiovascular calcification[22]. In male sex hormones, there is a positive correlation between elevated testosterone and cardiovascular calcification, while in female sex hormones, the cardioprotective effect of estrogen is widely recognized. Therefore, when women enter menopause, the risk of cardiovascular disease will increase due to the decrease of estrogen level. When mineral components in the blood are deposited in blood vessels or heart valves, cardiovascular calcification can occur, including vascular calcification and heart valve calcification. According to its location, cardiovascular calcification can be divided into three types: atherosclerotic intimal vascular calcification, medial vascular calcification and aortic valve

calcification[23]. Studies have shown that the occurrence of cardiovascular calcification has become a predictor of cardiovascular disease-related risks[24]. The deposition of maladjusted calcium may lead to coronary atherosclerotic heart disease, aortic stenosis, hypertension and even heart failure, and become the first trigger to push down dominoes. Studies have shown that in vascular smooth muscle cells, estrogen receptors α (ERα). The expression of estrogen receptor was higher than that of estrogen receptor β (ERβ). Estrogen is mainly through ERα Inhibit RANKL signal, so as to reduce osteogenic differentiation and calcification of vascular smooth muscle cells by up regulating BMP and down regulating MGP[25]. In addition, estrogen and aromatase in blood vessels may also play the same role[26]. On the other hand, studies have shown that matrix vesicles (MVS) produced by the outer membrane of cardiovascular cells It also plays a role in cardiovascular calcification. When they are ingested by vascular smooth muscle receptors, they can cause changes in MAPK signal and calcium metabolism. The regulatory response process of calcium binding annexin is activated, while the expression of calcification inhibitors such as MGP and fetuin A is at a low level[27, 28].

In conclusion, some key genes in this study may affect the cardiovascular system with the help of gender related factors, and finally play a promoting role in the occurrence and development of heart failure.

Possible relationship between BMP, extracellular matrix and heart failure

Combining go enrichment analysis and PPI analysis, we pay attention to that BMP and extracellular matrix are relatively high-frequency keywords.

Bone morphogenetic protein BMP is involved in the regulation of embryogenesis and organogenesis, and can participate in the development of cardiovascular structure and function in embryonic stage. When individuals mature, BMP can be used as an important endocrine regulator to participate in cardiovascular, metabolic and hematopoietic activities[29, 30]. Studies have shown that BMP is a transforming growth factor β (TGF-β), a member of the family carries out signal transmission with adjacent cells by means of paracrine or autocrine. Therefore, it plays a role in promoting the early development of organs. On the other hand, some BMP can also play a role in signal transmission in blood circulation, so as to affect distant tissues and organs, and finally complete the role of BMP in cardiovascular, metabolic and hematopoietic functions[31, 32]. Studies have shown that BMP signaling pathway plays an important role in cardiovascular diseases, increasing its activity in vascular inflammation and atherosclerosis, while decreasing its activity in pulmonary hypertension and hereditary hemorrhagic telangiectasia[32]. BMP has TGF-β, the commonness of the family is to bind to serine threonine kinase type I and type II receptors. Among them, the affinity for binding to type I receptors is higher, forming BMPRIA, also known as ALK3, BMPRIB, also known as ALK6, ACVRL1, also known as ALK1, and ACVR1, also known as ALK2 heterotetramer complexes; binding with type II receptors to form BMPR II, ACTRIIA, and ACTRIIB heterotetramer complexes, which can be widely expressed in mesenchymal stem cells In the tissues derived from, especially BMPR II is highly expressed in endothelial and endocardial tissues[33]. Studies have shown that when myocardial pathological hypertrophy is induced by pressure overload, the expression of BMP4 increases, but it does not increase in physiological hypertrophy induced by exercise[34, 35], and this effect can be inhibited by BMP inhibitors[35, 36]. BMP7 has been proved to inhibit cell apoptosis, myocardial fibrosis and anti calcification, and can improve the cardiac function of patients[37]. Studies have shown that BMP is closely related to the development of the heart during embryonic development. In the process of cardiac development, it needs to undergo endocardial to mesenchymal transition (EMT) And migrate as mesenchymal cells to fill the extracellular matrix

separating the endocardial layer and the outer layer of myocardium[38, 39], and finally complete the development of the heart[40-42]. Therefore, the enriched pathways in this study are closely related to these mechanisms, which may provide some new evidence for these mechanisms[43].

Extracellular matrix (ECM) It plays an important role in stabilizing the structure, transmitting signals and stress of cardiomyocytes, vascular cells and stromal cells. Therefore, the regulatory role of ECM is closely related to the occurrence and development of heart failure. When cardiomyocytes are stimulated by external stimuli or their own dysfunction, it can affect the regulatory role of ECM and reduce stress The increase of load can cause the transformation of cardiac fibroblasts into myofibroblasts, that is, "interstitial fibrosis" And promote the synthesis of extracellular matrix, so as to reduce cardiac compliance, myocardial remodeling and accelerate the impairment of diastolic function[44]. Interstitial fibrosis is accompanied by the expansion of collagen area around cardiac microvascular adventitia, secondary to myocardial ischemia and hypoxia, aggravate the imbalance between supply and demand of blood oxygen under stress, and promote the occurrence and development of heart failure[45, 46]. Clinically, angiotensin converting enzyme inhibitors, angiotensin receptor inhibitors β[47, 48]. The application of adrenergic receptor antagonists and diuretics is to regulate the abnormal changes of ECM in patients with heart failure by reducing the load on the heart[49]. Among them, proteoglycan located in ECM plays a role in ECM, which belongs to glycosylated protein[50]. Research shows that fibroblasts' response to mechanical stress may include the following mechanisms: (1) After fibroblasts perceive the increase of mechanical load, they amplify the signal through intracellular cascade reaction and induce myocardial fibrosis by activating transcription factor myocardial related transcription factor (MRTF)[51]; (2) mechanical stress stimulates the key medium in the transformation of myofibroblasts - transforming growth factor TGF-β After treatment, it can promote matrix synthesis[52, 53]; (3) increased pressure load can directly activate renin angiotensin aldosterone system (RAAS), and stimulate fibroblast proliferation and ECM protein synthesis through angiotensin II type 1 receptor (AT1R) signal[54, 55]; (4) The increase of pressure negative charge can induce the expression of miRNA in fibroblasts, further activate MAPK signaling pathway, and finally promote the synthesis of matrix[56].

In the current research, the possible mechanisms of 10 potential key genes involved in the occurrence and development of heart failure have been discussed in the occurrence and development of heart failure. The results show that the above genes may become potential biomarkers and therapeutic targets of heart failure, hoping to provide some ideas for further exploration and research of heart failure Yes, this study still has some limitations: (1) the included samples have limitations: in the included data set, the age, gender, race, nationality, region, living habits and family history of the samples can be called influencing factors. (2) the potential key factors obtained from the analysis need to be experimentally verified in clinical samples, such as RT-qPCR, Western blot, etc.

Conclusions

Our study integrated relatively large sample size data from multiple geographic data sets and identified 10 potential key genes (EIF1AY, RPS4Y1, USP9Y, KDM5D, DDX3Y, NPPA, HBB, TSIX, LOC28556, XIST) by bioinformatics analysis The exploration of potential key genes of heart failure may provide some potential help for further identifying new biomarkers and useful therapeutic targets of heart failure susceptibility.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in the GEO Database repository,

GSE76701 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76701

GSE21610 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21610

GSE8331 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8331

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China
(No.: 81773444).

Authors' contributions

Conception and design: F.S; Data collection and analysis: F.S; Manuscript writing: F.S; Manuscript revising: H.Y.

Acknowledgements

Not applicable.

Reference

[1] Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018.

[2] Groenewegen A, Rutten FH, Mosterd A, et al. Epidemiology of heart failure. European journal of heart failure. 2020;22(8):1342-1356.

[3] Dhalla NS, Saini-Chohan HK, Delfin RL, et al. Subcellular remodelling may induce cardiac dysfunction in congestive heart failure. Cardiovascular Research. 2008;2008(3):3.

[4] Shimizu I, Minamino T. Physiological and pathological cardiac hypertrophy. Journal of Molecular & Cellular Cardiology. 2016;245-262.

[5] Shah AK, Bhullar SK, Elimban V, et al. Oxidative Stress as A Mechanism for Functional Alterations in Cardiac Hypertrophy and Heart Failure. Antioxidants (Basel, Switzerland). 2021;10(6):931.

[6] Smyth GK. limma: Linear Models for Microarray Data. Springer New York. 2005.

[7] Meltzer DPS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23(14):1846-1847.

[8] Gu Z, Roland E, Matthias S. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847.

[9] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics-a Journal of Integrative Biology. 2012;16(5):284-287.

[10] Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000.

[11] Kanehisa M, Sato Y, Furumichi M, et al. New approach for understanding genome variations in KEGG. Nucleic Acids Research. 2019;8;47(D1):D590-D595.

[12] Jupe S, Fabregat A, Hermjakob H. Expression Data Analysis with Reactome. Current Protocols in Bioinformatics. 2015.

[13] Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list

enrichment analysis tool. BMC Bioinformatics,14,1(2013-04-15). 2013;14(1):128-128.
[14] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research. 2016(W1):W90-W97.
[15] Xie Z, Bailey A, Kuleshov MV, et al. Gene Set Knowledge Discovery with Enrichr. Current Protocols. 2021.
[16] Schriml LM, Elvira M, James M, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nuclc Acids Research. 2018(D1):D1.
[17] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):P.15545-15550.
[18] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545-15550. Proceedings of the National Academy of Sciences. 2005;102(43):15545-15550.
[19] Damian S, Gable AL, David L, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2018(D1):D607.
[20] Davis AP, Grondin CJ, Johnson RJ, et al. The Comparative Toxicogenomics Database: update 2019. Nucleic acids research. 2018.
[21] Li N, Dhar SS, Chen TY, et al. JARID1D Is a Suppressor and Prognostic Marker of Prostate Cancer Invasion and Metastasis. Cancer Research. 2016;76(4).
[22] Li Y, Jiang Q, Ding Z, et al. Identification of a Common Different Gene Expression Signature in Ischemic Cardiomyopathy. Genes. 2018;9(1):56.
[23] Woodward HJ, Zhu D, Hadoke PWF, et al. Regulatory Role of Sex Hormones in Cardiovascular Calcification. International journal of molecular sciences. 2021;22(9):4620.
[24] Liu W, Zhang Y, Yu C-M, et al. Current understanding of coronary artery calcification. Journal of geriatric cardiology : JGC. 2015;12(6):668-675.
[25] Osako MK, Nakagami H, Koibuchi N, et al. Estrogen Inhibits Vascular Calcification via Vascular RANKL System. Circulation Research. 2010;20(4):466-475.
[26] Harada N, Sasano H, Murakami H, et al. Localized Expression of Aromatase in Human Vascular Tissues. Circulation Research. 1999;84(11):1285.
[27] Praneet C, Chen NX, Kalisha O, et al. Differential miRNA Expression in Cells and Matrix Vesicles in Vascular Smooth Muscle Cells from Rats with Kidney Disease. Plos One. 2015;10(6):e0131589.
[28] Chen NX, O"Neill KD, Moe SM. Matrix vesicles induce calcification of recipient vascular smooth muscle cells through multiple signaling pathways. Kidney International. 2018;S0085253817305501.
[29] David L, Mallet C, Keramidas M, et al. Bone morphogenetic protein-9 is a circulating vascular quiescence factor. Circulation Research. 2008;102(8):914-922.
[30] Vukicevic S, Grgurevic L. BMP-6 and mesenchymal stem cell differentiation. Cytokine & Growth Factor Reviews. 2009;20(5-6):441-448.
[31] Laux DW, Young S, Donovan JP, et al. Circulating Bmp10 acts through endothelial Alk1 to mediate flow-dependent arterial quiescence. Development. 2013;140(16):3403-3412.
[32] Herrera B, Inman GJ. A rapid and sensitive bioassay for the simultaneous measurement of multiple bone morphogenetic proteins. Identification and quantification of BMP4, BMP6 and

BMP9 in bovine and human serum. BMC Cell Biology. 2009;10(1):20.
[33] Wu X, Sagave J, Rutkovskiy A, et al. Expression of bone morphogenetic protein 4 and its receptors in the remodeling heart. Life Sciences. 2014;97(2):145-154.
[34] Sun B, Sheng Y, Huo R, et al. Bone morphogenetic protein-4 contributes to the down-regulation of Kv4.3 K+ channels in pathological cardiac hypertrophy. Biochemical & Biophysical Research Communications. 2013;436(4):591-594.
[35] Sun B, Rong H, Yue S, et al. Bone morphogenetic protein-4 mediates cardiac hypertrophy, apoptosis, and fibrosis in experimentally pathological cardiac hypertrophy. Hypertension. 2013;61(2):352.
[36] Pachori AS, Custer L, Hansen D, et al. Bone morphogenetic protein 4 mediates myocardial ischemic injury through JNK-dependent signaling pathway. Journal of Molecular & Cellular Cardiology. 2010;48(6):1255-1265.
[37] Narasimhulu A, Singla. The Role of Bone Morphogenetic Protein 7 (BMP-7) in Inflammation in Heart Diseases. Cells. 2020;9(2):280.
[38] Selleri L, Zappavigna V, Ferretti E. 'Building a perfect body': control of vertebrate organogenesis by PBX-dependent regulatory networks. Genes & development. 2019;33(5-6):258-275.
[39] Tam PP, Parameswaran M, Kinder SJ, et al. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. Development. 1997;124(9):1631-1642.
[40] Waller BR, Wessels A. Cardiac Morphogenesis and Dysmorphogenesis. Developmental Biology Protocols. 2000;135:151-161.
[41] Srivastava D, Olson EN. A genetic blueprint for cardiac development. Nature. 2000;407(6801):221-226.
[42] Kruithof B, Duim SN, Moerkamp AT, et al. TGFβ and BMP signaling in cardiac cushion formation: lessons from mice and chicken. Differentiation. 2012;84(1).
[43] Morrell NW, Bloch DB, Dijke PT, et al. Targeting BMP signalling in cardiovascular disease and anaemia. Nature Reviews Cardiology. 2016;13(2).
[44] Ytrehus K, Hulot J-S, Perrino C, et al. Perivascular fibrosis and the microvasculature of the heart. Still hidden secrets of pathophysiology? Vascular Pharmacology. 2018:S153718911730469X.
[45] Dai Z, Aoki T, Fukumoto Y, et al. Coronary perivascular fibrosis is associated with impairment of coronary blood flow in patients with non-ischemic heart failure. Journal of Cardiology. 2012;60(5-6):416-421.
[46] Varagic J, Frohlich ED, Diez J, et al. Myocardial fibrosis, impaired coronary hemodynamics, and biventricular dysfunction in salt-loaded SHR. American Journal of Physiology Heart & Circulatory Physiology. 2006;290(4):H1503.
[47] Kate H, Ida L, Andrew MC, et al. The Soft- and Hard-Heartedness of Cardiac Fibroblasts: Mechanotransduction Signaling Pathways in Fibrosis of the Heart. J Clin Med. 2017;6(5):53.
[48] Rienks M, Papageorgiou AP, Frangogiannis NG, et al. Myocardial Extracellular Matrix An Ever-Changing and Diverse Entity. Circulation Research. 2014;114(5):872-888.
[49] Christensen G, Herum KM, Lunde IG. Sweet, yet underappreciated: Proteoglycans and extracellular matrix remodeling in heart disease. Matrix Biology. 2018;S0945053X17302718.
[50] Karamanos NK. Matrix pathobiology—central roles for proteoglycans and heparanase in health and disease. Febs Journal. 2017;284(1).

[51] Kennedy L, Xu SW, Ca Rter DE, et al. Fibroblast adhesion results in the induction of a matrix remodeling gene expression program. Matrix Biology. 2008;27(4):274-281.
[52] Lighthouse JK, Small EM. Transcriptional control of cardiac fibroblast plasticity. Journal of Molecular & Cellular Cardiology. 2016;91:52-60.
[53] Zhao XH, Laschinger C, Arora P, et al. Force activates smooth muscle alpha-actin promoter activity through the Rho signaling pathway. Journal of Cell Science. 2007;120(Pt 10):1801-1809.
[54] Vincent S, Anne K, Chow ML, et al. Integrins αvβ5 and αvβ3 promote latent TGF-β1 activation by human cardiac fibroblast contraction. Cardiovascular Research. 2014(3):407-417.
[55] Ping K, Shinde AV, Su Y, et al. Opposing Actions of Fibroblast and Cardiomyocyte Smad3 Signaling in the Infarcted Myocardium. Circulation. 2017;137(7):CIRCULATIONAHA.117.029622.
[56] Frangogiannis NG. The Extracellular Matrix in Ischemic and Nonischemic Heart Failure. Circulation Research. 2019;125(1):117-146.

Figures

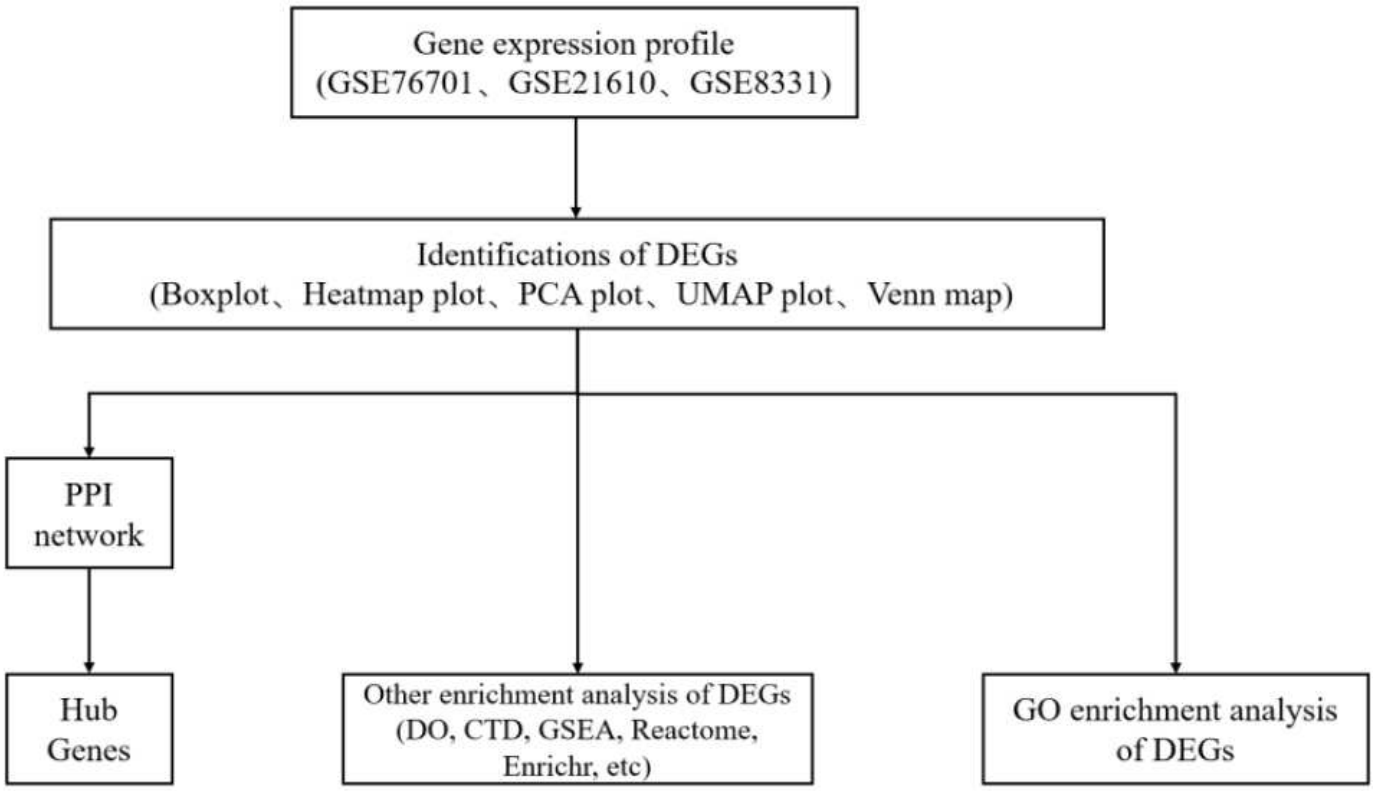

Figure 1

Study design (flow diagram of study). PPI Protein-protein interaction; DO Disease Oncology; CTD The Comparative Toxicogenomics Database; GSEA Gene Set Enrichment Analysis; GO Gene Oncology.

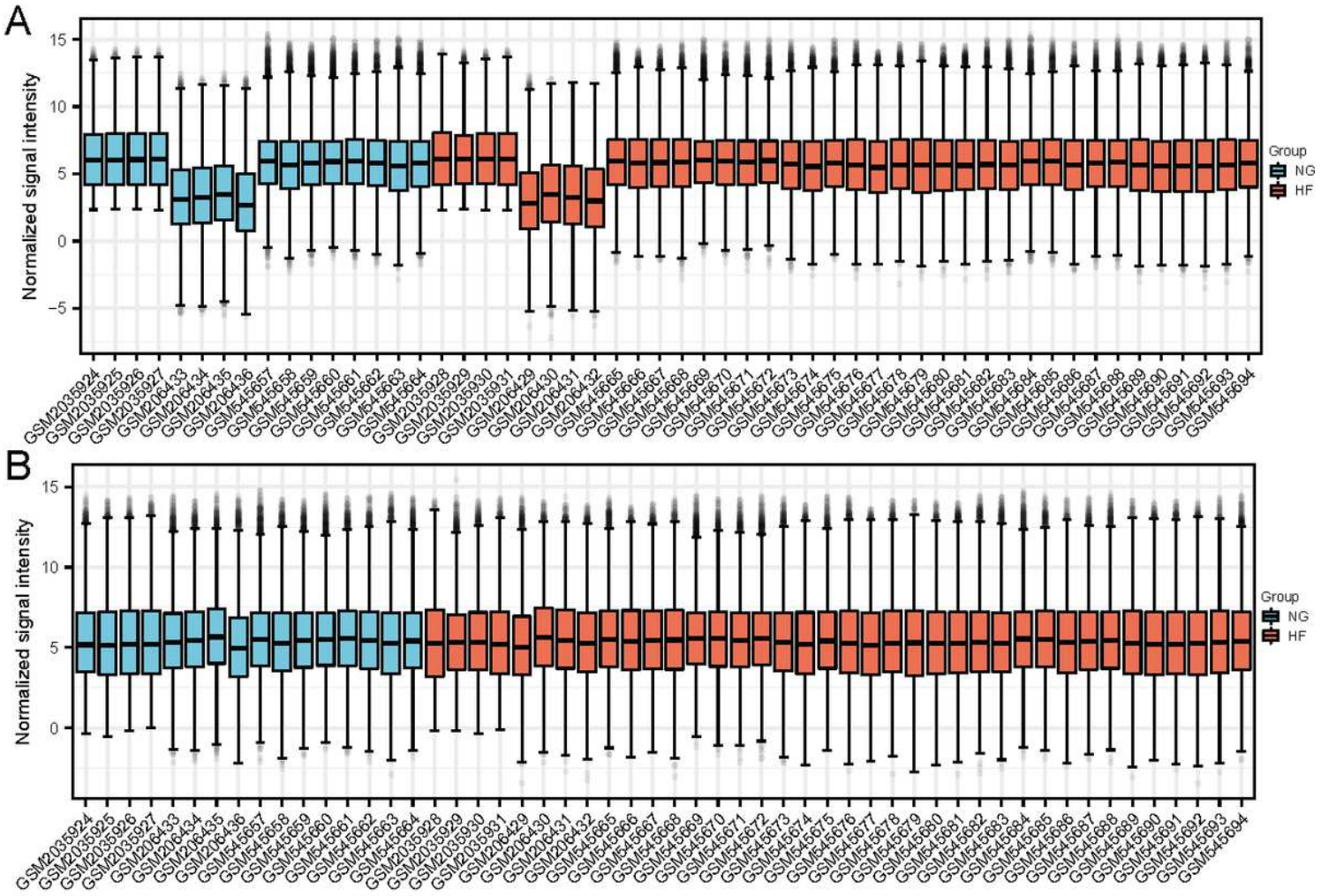

Figure 2

Boxplots of gene expressions before and after standardization for 3 selected GEO Database. (A. Before standardization; B. After standardization. NG: Normal Group; HF: Heart Failure Group)

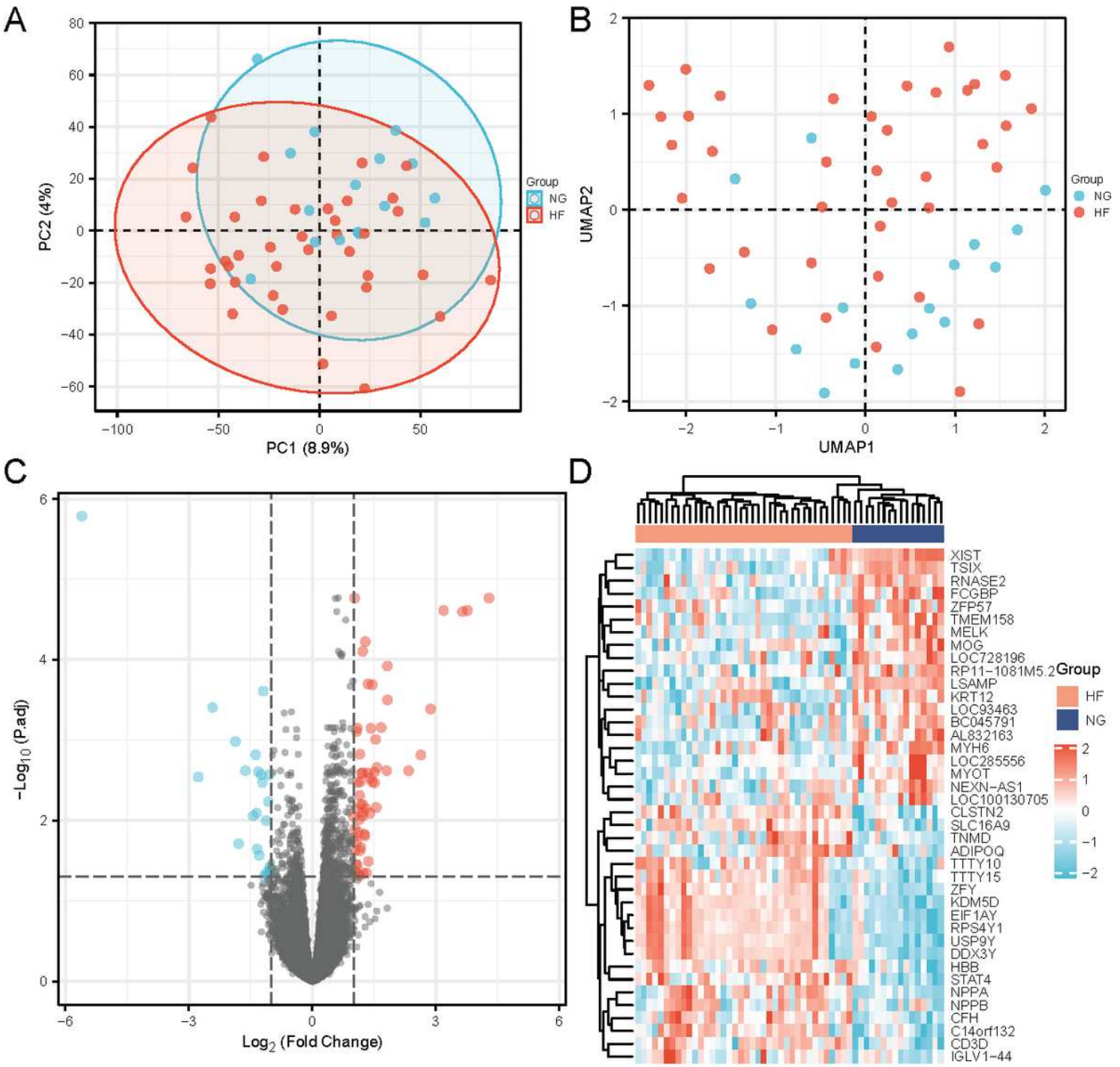

Figure 3

The PCA plot, UMAP plot, Volcano plot and Heatmap plot of gene expressions for 3 selected GEO Database with a screening criteria of |log2 FC| ≥ 1 and adjust P value < 0.05. (A. PCA plot; B. UMAP plot; C. Volcano plot; D. Heatmap plot. PCA: Principal Component Analysis; UMAP: Uniform Manifold Approximation and Projection; NG: Normal Group; HF: Heart Failure Group)

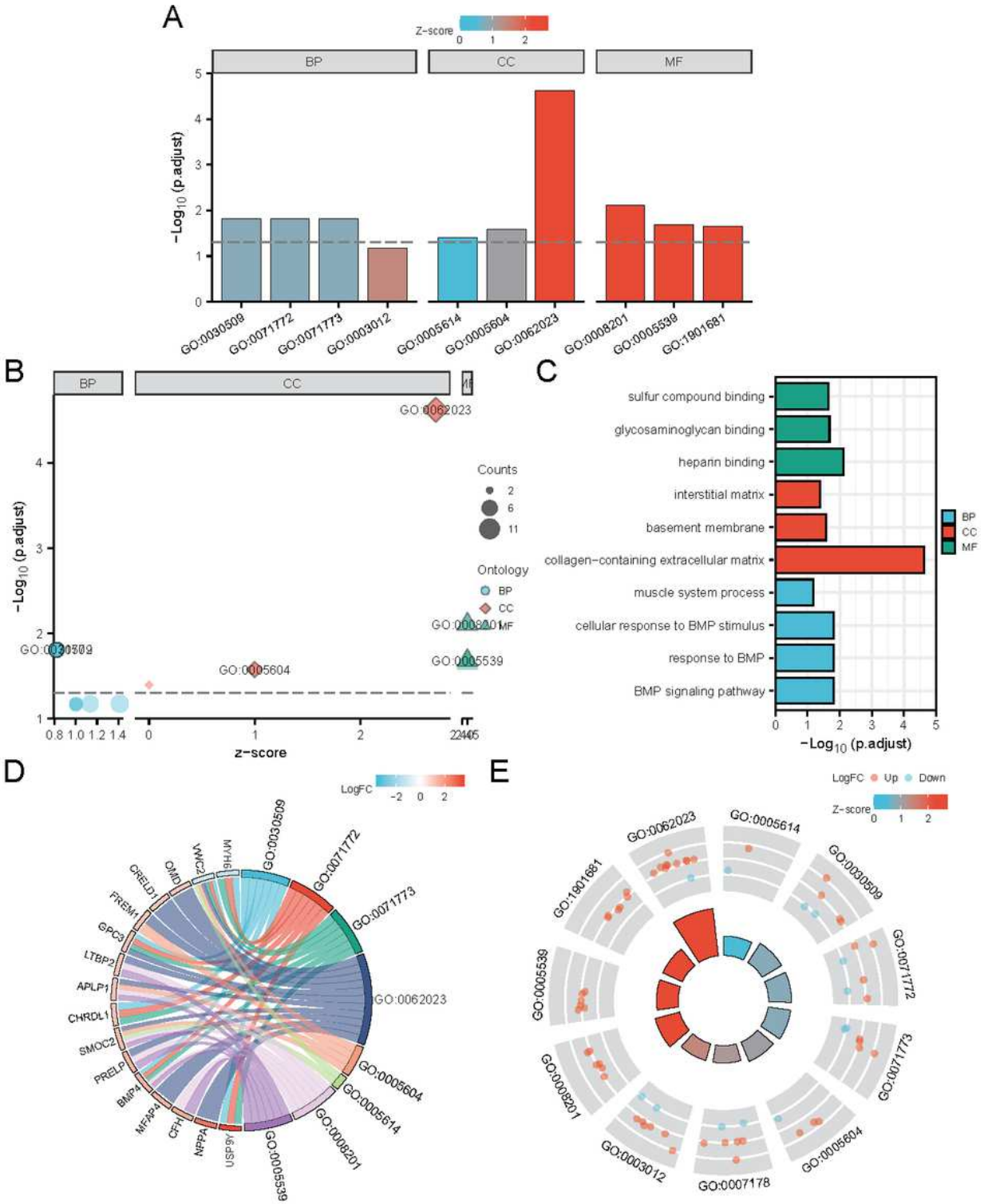

Figure 4

Enrichment plots by GO. (A. GO Bar graph, B. Bubble plot, C. Bar graph of GO enrichment pathways, D. chord diagram, E. loop graph. BP: Biological Process; CC: Cellular Component; MF: Molecular Function)

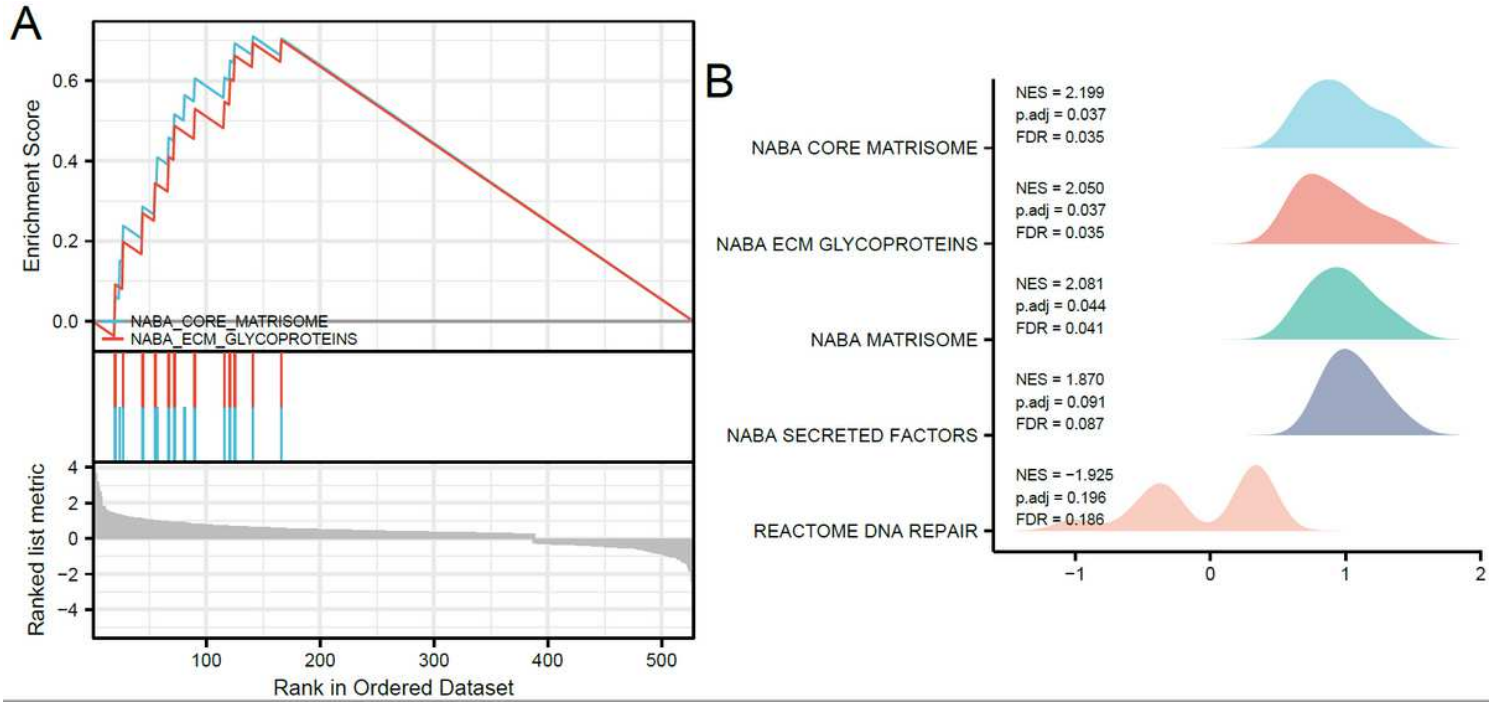

Figure 5

Enrichment plots by GSEA. (A. GSEA Visual Analysis, B. GSEA ridgeplot. NES: Normal Enrichment Score; FDR: False Discovery Rate)

Figure 6

Enrichment pathways by Enrichr. (Upper: Related enrichment pathways of up-regulated genes; Lower: Related enrichment pathways of down-regulated genes)

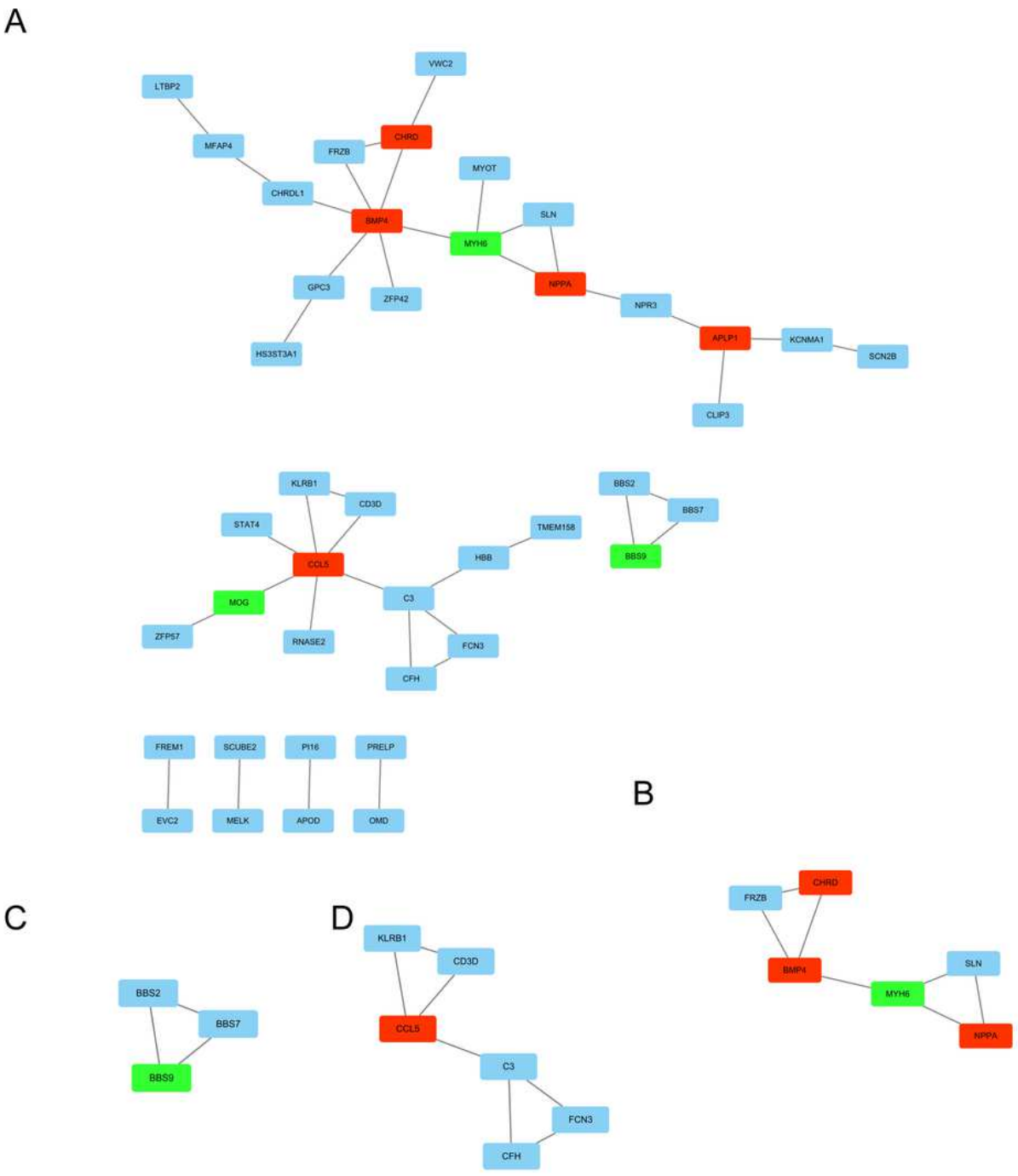

Figure 7

PPI network construction. (A. The complete PPI network, B-D. The three modules of hub genes)

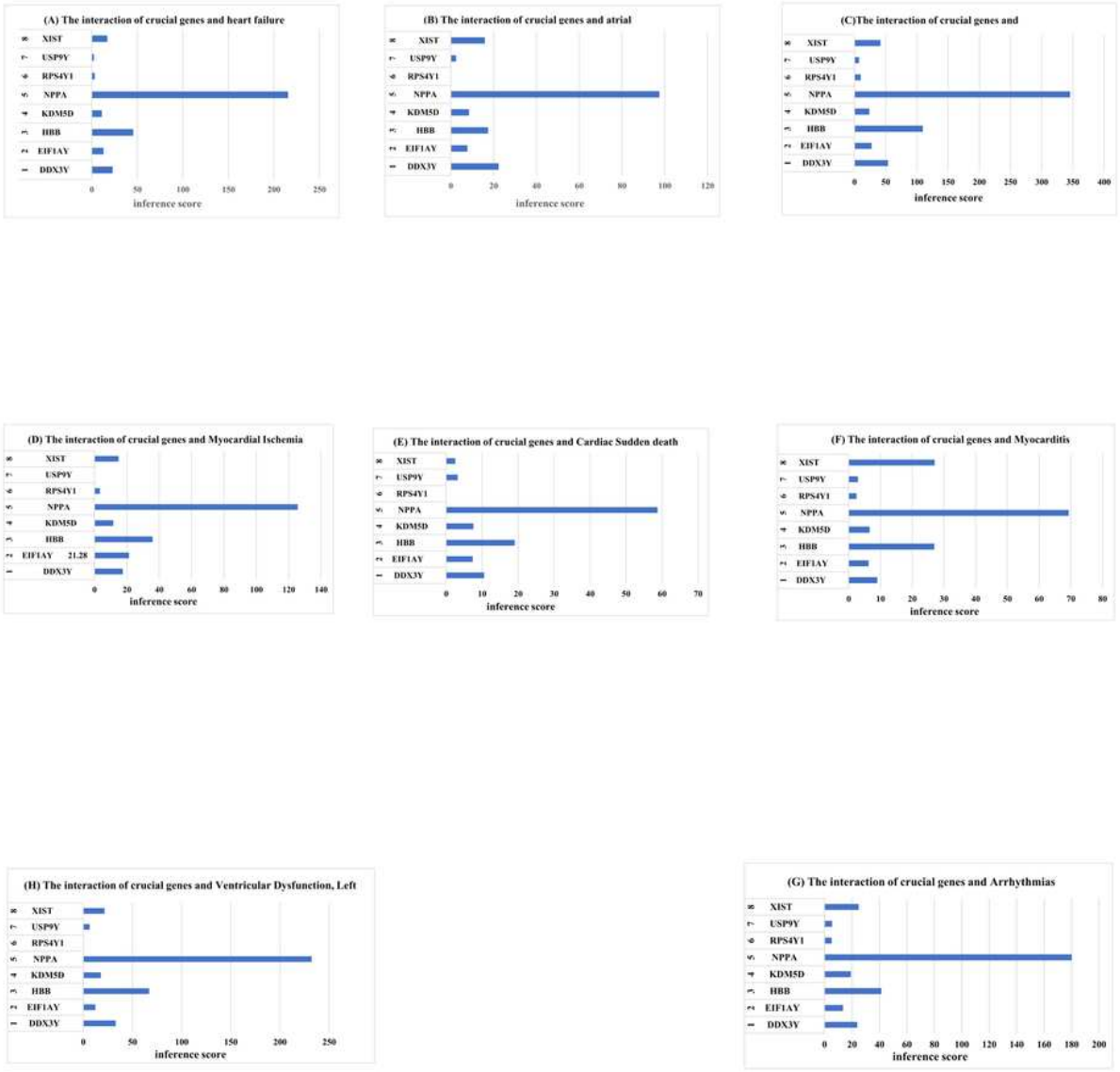

Figure 8

The CTD analysis between potential key genes and diseases.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.pdf
- Table2.pdf
- Table3.pdf
- Table4.pdf