Original Research Article

Identification of potential biomarkers and candidate small-molecule drugs for heart failure via comprehensive gene microarray analysis

Hailang Liu, Chunyang Yu, Zhongcheng Wei, Qing Zhang*
Department of Cardiology, The Affiliated Huai'an No.1 People's Hospital of Nanjing Medical University, Huai'an City, Jiangsu Province 223300, China

*For correspondence: **Email:** qzhang email@163.com; **Tel:** +86-0517-80872604*

Abstract

Purpose: To identify potential novel biomarkers and to explore new small-molecule drugs for heart failure (HF).
Methods: The Gene Expression Omnibus (GEO) microarray datasets were downloaded for analyzing the differentially expressed genes (DEGs). Venn analysis was performed to calculate the overlapping genes which were then used for Gene Ontology (GO) analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis using cluster Profiler in R package; a protein-protein interaction network (PPI) was constructed using STRING database. The hub genes were selected for small-molecule drug identification, while molecular docking of small-molecule drugs and hub genes was performed using CB-dock2.
Results: Upregulated and downregulated DEGs were obtained from GSE84796, GSE107569 and GSE116250 datasets, respectively. Eleven (11) overlapping genes, which were enriched in collagen fiber tissue, collagen-containing extracellular matrix and collagen fiber-related pathways, were also enriched in AGE-RAGE and relaxin signaling pathways. The PPI network of the DEGs was constructed, and five hub genes, with high connectivity, were significantly upregulated in HF. The five hub genes were ranked as MFAP4, LTBP2, THBS4, COL3A1 and COL1A1. Two targets (COL1A1 and COL3A1) matched potential drugs, and fostamatinib shared by the two targets had the greatest therapeutic value for HF.
Conclusion: Five novel biomarkers and involved signaling pathways have been identified in HF via comprehensive microarray analyses. The results also show that fostamatinib might be a promising drug candidate for HF treatment.

Keywords: Bioinformatics analysis, Biomarker, Candidate small-molecules, Fostamatinib, Heart failure

INTRODUCTION

Heart failure (HF) is a chronic disease with high risk of sudden death; it is caused by increased intracardiac pressure or decreased cardiac output, and the incidence is higher in the elderly persons [1]. Pathological cardiac hypertrophy is considered a major predictor of HF [2].

Cardiac hypertrophy refers to the sustained expansion of the myocardium in response to cardiac injury such as myocardial infarction and hypertension [3]. At the cellular level, cardiac hypertrophy is defined by several key features, including enlarged cardiomyocyte size, heightened protein synthesis, changes in gene expression, and the formation of sarcoma-forming tissue [4].

Angiotensin II (Ang II) is an important active peptide of the renin-angiotensin-aldosterone system (RAAS), which is closely related to the apoptosis, hypertrophy and fibrosis of cardiomyocytes, and may be responsible for the exacerbation of cardiovascular diseases. In addition, Ang II induces cardiac oxidative stress, inflammation and fibrotic responses, ultimately orchestrating events leading to cardiac dysfunction [5].

Comprehensive gene microarray analysis, also known as gene expression profiling or transcriptomics, is a technique used to evaluate the cellular protein levels of thousands of genes simultaneously [6]. It provides valuable information about the activity of genes within a biological sample, such as tissues or cells, and helps to uncover various biological processes, disease mechanisms, and potential therapeutic targets. Comprehensive gene microarray analysis has been widely used in various fields, including genomics, molecular biology, pharmacology, and medical diagnostics [7]. It provides a powerful tool to study gene expression patterns and identify biomarkers, therapeutic targets, and molecular signatures associated with diseases or specific pathogens.

In this study, the GEO microarray datasets were utilized for differentially expressed genes (DEGs), GO and KEGG analysis, as well as PPI network construction. Two targets (COL1A1 and COL3A1) along with their shared target, fostamatinib, were used for potential drug analysis through molecular docking. These findings suggest the candidate biomarkers and key signaling pathways for HF and highlight fostamatinib as a promising HF treatment candidate.

METHODS

DEGs screening

Three gene microarray datasets, GSE84796, GSE107569 and GSE116250, were downloaded from the Gene Expression Omnibus (GEO) database. Normalization of the raw data was performed by the Normalize Between Arrays in R package. The DEGs between HF and normal tissues were identified via Limma Package, based on the criteria, $p < 0.05$ and $| logFC | > 1$. The volcano plot of the DEGs was constructed using gglpot2, while the heatmap of DEGs was constructed using Pheatmap.

Overlapping gene analysis

Venn analysis was performed to calculate the overlapping genes of the screened DEGs above, and the upregulated and downregulated DEGs were individually submitted into the Venn analysis software. Then the results were shown as pie pictures which were drawn with the aid of an online platform (https://www.bioladder.cn/web/#/pro/index).

Enrichment analysis

The gene enrichment analyses included GO analysis and KEGG pathway analyses. These analyses were performed using clusterProfiler in R package and the information of the overlapping genes above was submitted into the software for the analysis. The KEGG and GO analysis were screened for q values < 0.05. Then the GO and KEGG results were shown as histogram and bubble chart respectively, which were drawn using ggplot2 in R package.

PPI network establishment and hub gene analysis

The screened overlapping genes were uploaded into STRING database (https://string-db.org/) to establish the PPI network. The cytoHubba of cytoscape (http://www.cytoscape.org) was mainly used to rank the nodes in the network according to their reported biological functions. The TOP5 core nodes (hub genes) were selected using the density of maximum neighbourhood component (DMNC) method [8]. The histograms involved in the expression of hub genes in GSE84796, GSE107569 and GSE116250 datasets were drawn using GraphPad Prism 8.0 (GraphPad Software, Boston, MA, USA).

Small-molecule drug identification

The potential small-molecule therapeutic candidates targeting the hub genes were identified using the Quartata platform (http://quartata.csb.pitt.edu/index_2.php) according to the provided instructions. The molecular docking of small-molecule drugs and hub genes was performed using CB-dock2 online tools (https://cadd.labshare.cn/cb-dock2/php/index.php)[9].

RESULTS

Differential genes in heart failure

Three gene microarray datasets, including GSE84796, GSE107569 and GSE116250, were downloaded from the Gene Expression Omnibus (GEO) database. Based on these data, the differentially expressed genes (DEGs) were predicted. Specifically, there were 1144 upregulated genes and 817 downregulated genes of HF identified from GSE84796 database via the Limma Package. The volcano plot of DEGs was constructed using gglpot2, and the heatmap of DEGs was constructed using Pheatmap (Figure 1 A). Similarly, a total of 668 upregulated genes and 634 downregulated genes of HF were identified from GSE107569 database, for which the volcano plot and heat maps were also constructed (Figure 1 B). From GSE116250 database, 94 upregulated genes and 19 downregulated genes were screened, and their volcano plot and heat map are shown in Figure 1 C. Further, the Venn diagrams were analyzed to show the overlapping DEGs, and the results illustrated there were 11 overlapping genes which were upregulated, but no overlapping genes were downregulated among the three datasets (Figure 1 D). The details of

these 11 overlapping genes are listed in Figure 1 E.

Overlapping DEGs

To investigate the biological functions and signaling pathways of the DEGs, the clusterProfiler software in R package was used for KEGG and GO analysis, and the histogram and bubble chart were drawn using ggplot2. The histogram of GO analysis revealed that the 11 overlapping DEGs were enriched in collagen fiber tissue, collagen-containing extracellular matrix and collagen fiber-related pathways, which are key factors in promoting fibrosis formation and myocardial remodeling (Figure 2 A). In addition, the bubble chart of the KEGG analysis suggested that these genes were enriched in the AGE-RAGE pathway and the relaxin signaling pathway, which are closely associated with heart-related diseases (Figure 2 B).

Hub genes

The PPI network of the DEGs was constructed according to the STRING database. From this network, interactions with a high confidence score which was > 0.4 were selected, and eight genes were retained for refined PPI network construction using Cytoscape (Figure 3 A).
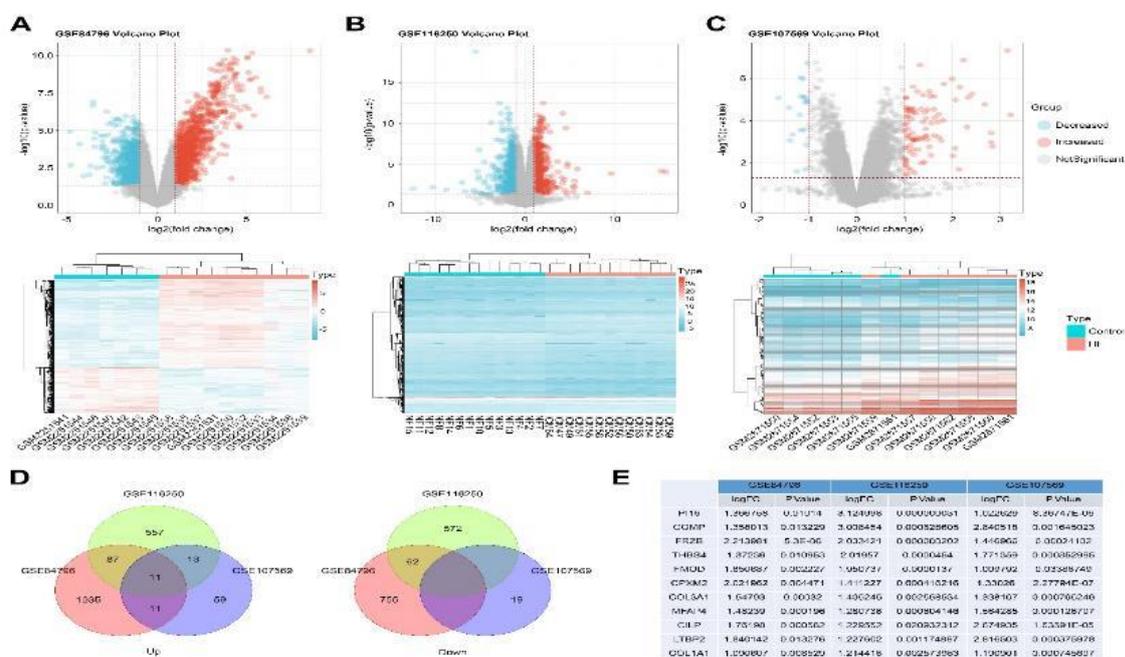

Figure 1: Differential genes in heart failure. (A) Vvolcano plot and heat map of differentially expressed genes (DEGs) between HF and normal tissues were identified from GSE84796 database via Limma Package; (B) Volcano plot and heat map of DEGs were identified from GSE107569 database; (C) Volcano plot and heat map of DEGs were identified from GSE116250 database; (D) Venn diagrams revealing 11 overlapping DEGs from GSE84796, GSE107569 and GSE116250 datasets; (E) Details of 11 overlapping genes
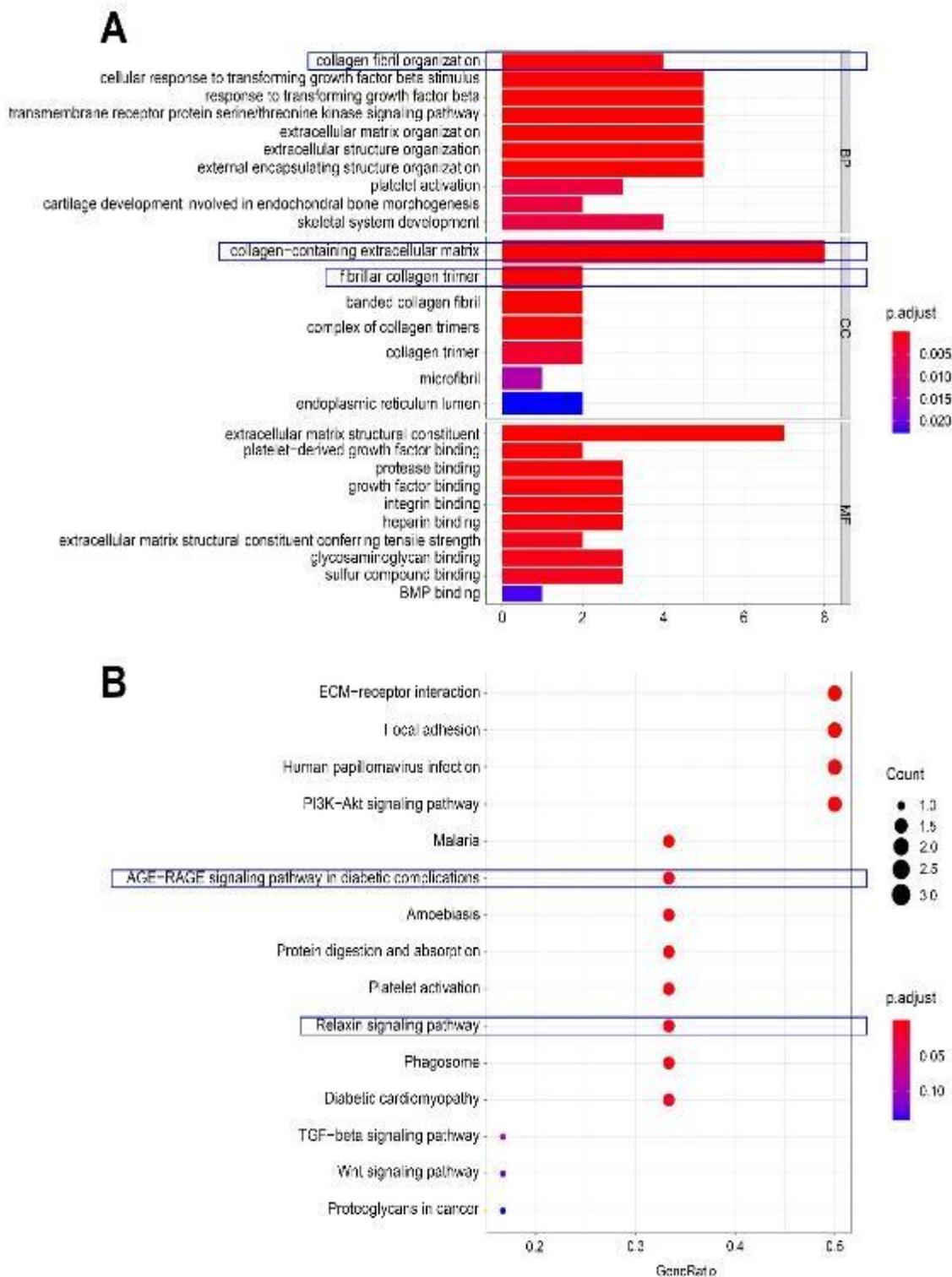
Figure 2: Overlapping DEGs. (A) Histogram of GO analysis of overlapping DEGs; (B) Bubble chart of KEGG analysis

The cytoHubba in cytoscape was used to rank the nodes in the PPI network, and as a result, five hub genes with high connectivity, whose expression were significantly upregulated in HF, were used for further analyses (Figure 3 B). The five hub genes were ranked as MFAP4, LTBP2, THBS4, COL3A1 and COL1A1 (Figure 3 C). The expression of the five genes in GSE84796, GSE107569 and GSE116250 datasets are shown in histograms (Figure 3 D).
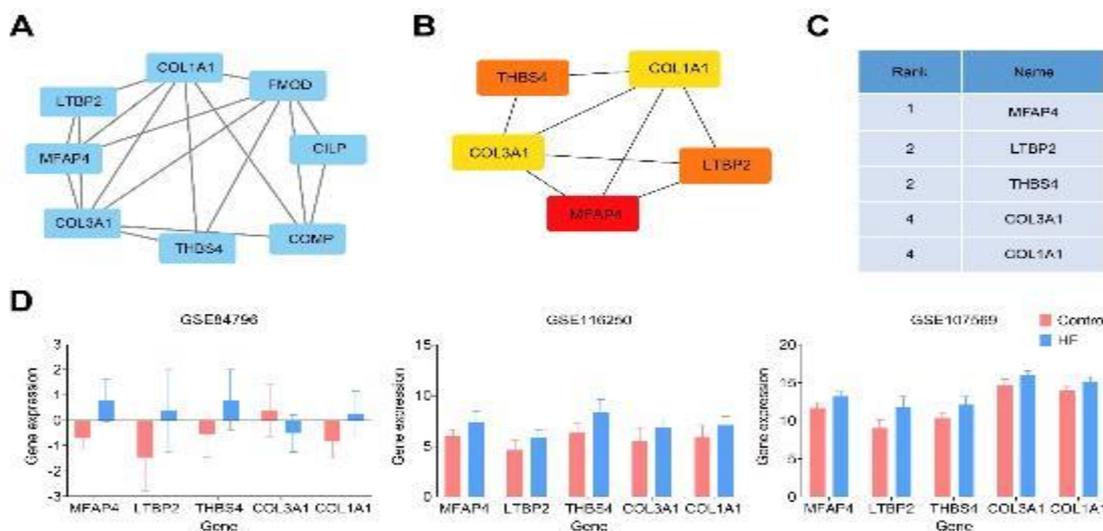
Figure 3: Hub gene analysis. (A) The PPI network of the DEGs with a high confidence score > 0.4; (B) Five hub genes with high connectivity and significant upregulation in HF were selected for deeper analyses; (C) The rank of five hub genes; (D) Histograms of the expression levels of five genes in GSE84796, GSE107569 and GSE116250 datasets
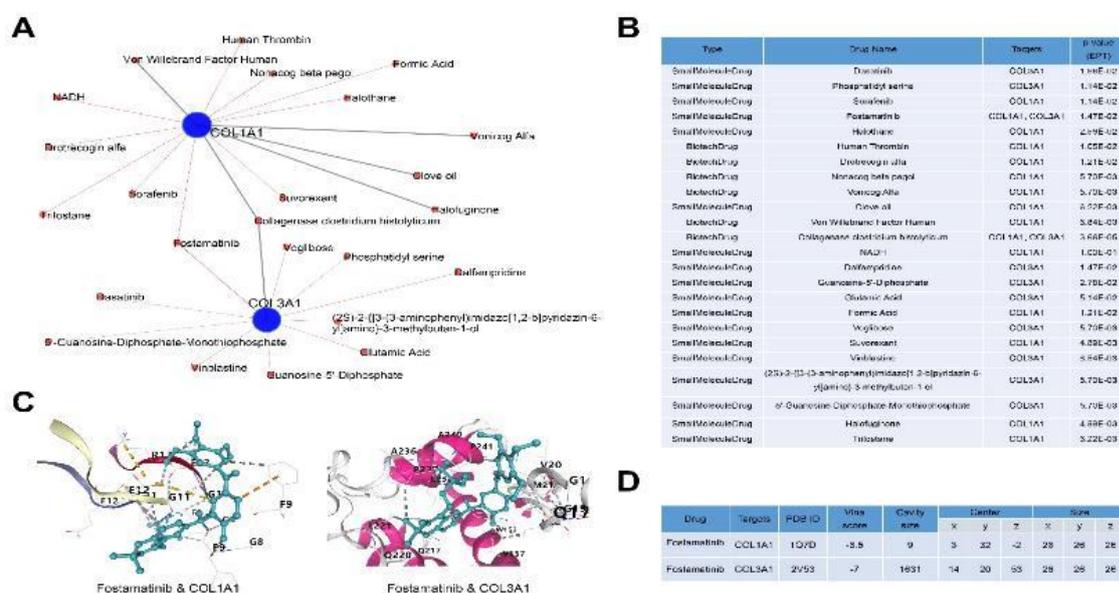

Figure 4: Potential small-molecule drugs. (A) Potential small-molecule drugs identified by bioinformatics analysis of the hub genes. Only COL1A1 and COL3A1 could match potential drugs; (B) Identified drugs shared by COL1A1 and COL3A1, *viz*. fostamatinib and collagenase clostridium histolyticum; (C) CB-Dock analysis to simulate the binding between fostamatinib and its targets COL1A1 and COL3A1; (D) Detailed parameters of molecular docking

Potential small-molecule drugs

To screen the small-molecule therapeutic candidates targeting the five hub genes, bioinformatics analysis of the hub genes using Quartata online tool was performed. All five hub genes were analyzed, but only two targets (COL1A1 and COL3A1) could match potential drug information (Fig. 4A). The drugs shared by the two targets were fostamatinib (small

molecule drug) and collagenase *Clostridium histolyticum* (biological drug; Figure 4 B). Fostamatinib, which is used for treating inflammation, severe infections, kidney and lung damage, was selected for the next docking analysis. CB-Dock analysis was used to simulate the binding between Fostamatinib and its targets COL1A1 and COL3A1 (Fig. 4 C). Detailed parameters of the molecular docking are listed in Figure 4 D. The data shows that fostamatinib

binds with COL1A1 and COL3A1, and thus may be a promising drug candidate for HF treatment.

DISCUSSION

Heart failure (HF) is a chronic disease with high risk of sudden death and its incidence is higher in the elderly. Thus, it is necessary to develop novel biomarkers and potential drugs for its diagnosis and treatment. High-throughput sequencing has acquired increasing use in analyzing genetic or epigenetic regulations in diseases, including HF [10]. Bioinformatics analysis is a common tool for investigating biomarkers and identifying novel drug. In this study, comprehensive microarray analysis was used to study the specific biomarkers for HF and to develop small-molecule drugs based on HF gene expression profiles.

The purpose of this work is to analyze the multi-chip data, screen out the genes significantly expressed in HF disease as potential HF disease targets, and predict drugs through these targets, in order to provide potential solutions for the treatment of HF. By analyzing GSE84796, GSE107569, and GSE116250 chips, this subject obtained 11 differentially expressed common genes. The 11 genes were mainly enriched in the pathways of extracellular matrix and collagen fiber formation. The extracellular matrix and collagen are important components to promote myocardial fibrosis and cardiac hypertrophy [11,12]. KEGG enrichment found the enrichment of these genes in the AGE-RAGE pathway and the relaxin signaling pathway. On the one hand, increased levels of AGE pathway indicators promote the risk of HF [13,14], and on the other hand, the relaxin signaling pathway is closely related to myocardial remodeling, which is also associated with HF [15]. Therefore, the enrichment analysis of overlapping DEGs facilitates the discovery of biological mechanisms and potential functional implications. The overlapping DEGs exhibit significant associations with the development and advancement of HF. These shared DEGs have the potential to serve as novel biomarkers for diagnosis, prognosis, or personalized therapeutic interventions.

Hub gene analysis refers to the identification and analysis of highly connected genes, known as hub genes, in biological networks. Hub genes are crucial nodes within a network that have a high degree of connectivity, meaning that they are strongly linked to many other genes or molecules in the network. Analyzing hub genes can provide insights into their central roles in biological processes, disease mechanisms, and regulatory networks. In the present study, one-step analysis obtained 5 hub genes, namely MFAP4, LTBP2,

THBS4, COL3A1, COL1A1. Among them, the top-ranked gene has also been reported in heart diseases, with the expression of MFAP4 shown to facilitate the development of angiotensin II-induced atrial fibrosis and atrial fibrillation [16].

Identifying potential small-molecule drugs involves a process known as drug discovery or drug development. This process aims to identify and develop compounds that have the potential to be effective in treating specific diseases or conditions. In this work, small molecule drugs identification for HF treatment was carried out. It was discovered that fostamatinib has anti-inflammatory effects, may treat severe infection, improve kidney and lung injury [17], and is a potential drug for HF treatment. *In vitro* and *in vivo* tests to evaluate the efficacy, toxicity, and pharmacological properties of fostamatinib are, however, necessary to determine the actual potential of the drug clinical practice.

CONCLUSION

DEGs obtained through integrated microarray analysis, and several novel biomarkers and biological pathways have been found to participate in the pathogenesis of HF. Furthermore, five hub genes have been selected and ranked as MFAP4, LTBP2, THBS4, COL3A1 and COL1A1. Two targets (COL1A1 and COL3A1) matched potential drugs, of which fostamatinib, shared by the two targets, has the greatest therapeutic potential for the treatment of HF. Taken together, these findings may contribute to the development of novel biomarkers for HF diagnosis, as well as present new candidate drugs for HF management.

DECLARATIONS

Ethical approval

None provided.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest

No conflict of interest associated with this work.

Contribution of Authors

We declare that this work was done by the authors named in this article and all liabilities pertaining to claims relating to the content of this article will be borne by the authors. Hailang Liu and Qing Zhang designed the study and carried them out; Hailang Liu, Chunyang Yu, Zhongcheng Wei supervised the data collection, analyzed and interpreted the data, prepared the manuscript for publication and reviewed the draft of the manuscript. All authors read and approved the manuscript for publication.

Open Access

REFERENCES

1. Duflos C, Troude P, Strainchamps D, Ségouin C, Logeart D, Mercier G. Hospitalization for acute heart failure: the in-hospital care pathway predicts one-year readmission. Sci Rep 2020; 10(1): 1-8.

2. Luo M, Chen P-p, Yang L, Wang P, Lu Y-l, Shi F-g, Gao Y, Xu S-f, Gong Q-h, Xu R-x. Sodium ferulate inhibits myocardial hypertrophy induced by abdominal coarctation in rats: involvement of cardiac PKC and MAPK signaling pathways. Biomed Pharmacother 2019; 112: 108735.

3. Nakamura M, Sadoshima J. Mechanisms of physiological and pathological cardiac hypertrophy. Nat Rev Cardiol 2018; 15(7): 387-407.

4. Agbo E, Li M, Wang Y, Saahene R, Massaro J, Tian G. Hexarelin protects cardiac H9C2 cells from angiotensin II-induced hypertrophy via the regulation of autophagy. Pharmazie 2019; 74(8): 485-491.

5. Swiderski J, Gadanec LK, Apostolopoulos V, Moore GJ, Kelaidonis K, Matsoukas JM, Zulli A. Role of Angiotensin II in Cardiovascular Diseases: Introducing Bisartans as a Novel Therapy for Coronavirus 2019. Biomolecules 2023; 13(5): 787.

6. Ma J, Cai X, Kang L, Chen S, Liu H. Identification of novel biomarkers and candidate small-molecule drugs in cutaneous melanoma by comprehensive gene microarrays analysis. J Cancer 2021; 12(5): 1307.

7. Wang W, Wang J, Wang J, Liu J, Pei J, Li W, Wang Y, Banerjee S, Xu R, Meng Z. Whole-exome sequencing identified a homozygous novel RAG1 mutation in a child with omenn syndrome. Allergol Immunopath 2022; 50(6): 32-46.

8. Vaghasia H, Sakaria S, Prajapati J, Saraf M, Rawal RM. Interactive bioinformatics analysis for the screening of hub genes and molecular docking of phytochemicals present in kitchen spices to inhibit CDK1 in cervical cancer. Comput bioL Med 2022; 149: 105994.

9. Li N, Liu K, Yu M, Li S, Liu M, Cai W, Tian A. Systematic elucidation of the traditional Chinese medicine prescription Danxiong particles via network pharmacology and molecular docking. Trop J Pharm Res 2022; 21(5): 981-987.

10. Laimoud M, Abdel-Moaty M, Elshobary M, Mohamed T, Ahmed W. Predictors of non-invasive ventilation failure in adult patients with cardiac dysfunction presenting with community-acquired pneumonia: an Egyptian multicenter prospective study. Signa Vitae 2022; 1: 11.

11. Frangogiannis NG. The extracellular matrix in ischemic and nonischemic heart failure. Circ Res 2019; 125(1): 117-146.

12. Wu G, Wang Z, Shan P, Huang S, Lin S, Huang W, Huang Z. Suppression of Netrin-1 attenuates angiotension II-induced cardiac remodeling through the PKC/MAPK signaling pathway. Biomed Pharmacother 2020; 130(110495.

13. Koyama Y, Takeishi Y, Arimoto T, Niizeki T, Shishido T, Takahashi H, Nozaki N, Hirono O, Tsunoda Y, Nitobe J. High serum level of pentosidine, an advanced glycation end product (AGE), is a risk factor of patients with heart failure. J Card Fail 2007; 13(3): 199-206.

14. Chen X, Zhang J, Li T, Ding Z, Zou C. COL3A1 induces ischemic heart failure by activating AGE/RAGE pathway. Signa Vitae 2022; 18(6).

15. Wong NR, Mohan J, Kopecky BJ, Guo S, Du L, Leid J, Feng G, Lokshina I, Dmytrenko O, Luehmann H. Resident cardiac macrophages mediate adaptive myocardial remodeling. Immunity 2021; 54(9): 2072-2088. e2077.

16. Wang H, Liu M, Wang X, Shuai W, Fu H. MFAP4 deletion attenuates the progression of angiotensin II-induced atrial fibrosis and atrial fibrillation. Europace 2022; 24(2): 340-347.

17. Prendecki M, Gulati K, Turner-Stokes T, Bhangal G, Chiappo D, Woollard K, Cook HT, Tam FW, Roufosse C, Pusey CD. Characterisation of an enhanced preclinical model of experimental MPO-ANCA autoimmune vasculitis. J Pathol 2021; 255(2): 107-119.