73324_Auto_Edited.docx

WORD COUNT	3452	TIME SUBMITTED	15-MAR-2022 03:09PM
		PAPER ID	83660784

Name of Journal: *World Journal of Cardiology*
Manuscript NO: 73324
Manuscript Type: ORIGINAL ARTICLE

Basic Study

Bioinformatics prediction of potential mechanisms and biomarkers underlying dilated cardiomyopathy

Zhou Liu, Ying-Nan Song, Kai-Yuan Chen, Wei-Long Gao, Hong-Jin Chen, Gui-You Liang

Abstract

BACKGROUND

Heart failure is a health burden responsible for high morbidity and mortality worldwide, and dilated cardiomyopathy (DCM) is one of the most common causes of heart failure. DCM is a disease of the heart muscle and is characterized by enlargement and dilation of at least one ventricle alongside impaired contractility with left ventricular ejection fraction < 40%. It is also associated with abnormalities in cytoskeletal proteins, mitochondrial ATP transporter, microvasculature, and fibrosis. However, the pathogenesis and potential biomarkers of DCM remain to be investigated.

AIM

To investigate the candidate genes and pathways involved in DCM patients.

METHODS

Two expression datasets (GSE3585 and GSE5406) were downloaded from the Gene Expression Omnibus database. The differentially expressed genes (DEGs) between the DCM patients and healthy individuals were identified using the R package “linear models for microarray data.” The pathways with common DEGs were analyzed *via*

Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and gene set enrichment analyses. Moreover, a protein-protein interaction network (PPI) was constructed to identify the hub genes and modules. The miRDB database was applied to predict the micro-RNAs (miRNAs) targeting the hub genes. Additionally, immune-cell infiltration in DCM was analyzed using CIBERSORT.

RESULTS

In total, 97 DEGs (47 upregulated and 50 downregulated) were identified. GO analysis showed that the DEGs were mainly enriched in "response to growth factor," "extracellular matrix," and "extracellular matrix structural constituent." KEGG pathway analysis indicated that the DEGs were mainly enriched in "protein digestion and absorption" and "IL-17 signaling pathway." The PPI network suggested that collagen type III alpha 1 chain (COL3A1) and collagen type I alpha 2 chain (COL1A2) contribute to the pathogenesis of DCM. Additionally, visualization of the interactions between miRNAs and the hub genes revealed that hsa-miR-5682 and hsa-miR-4500 interacted with both COL3A1 and COL1A2, and thus these miRNAs might play roles in DCM. Immune-cell infiltration analysis revealed that DCM patients had more infiltrated plasma cells and fewer infiltrated B memory cells, T follicular helper cells, and resting dendritic cells.

CONCLUSION

COL1A2 and COL3A1 and their targeting miRNAs, hsa-miR-5682 and hsa-miR-4500, may play critical roles in the pathogenesis of DCM, which closely relationship to the IL-17 signaling pathway and acute inflammatory response. These results may provide useful clues for the diagnosis and treatment of DCM.

Key Words: Dilated cardiomyopathy; bioinformatics; differentially expressed genes; function enrichment analysis; protein-protein interaction network; immune cell infiltration.

Liu Z, Song YN, Chen KY, Gao WL, Chen HJ, Liang GY. Bioinformatics prediction of potential mechanisms and biomarkers underlying dilated cardiomyopathy. *World J Cardiol* 2022; In press

Core Tip: As commonest causes of heart failure, the diagnosis and therapy for dilated cardiomyopathy (DCM) is still unsatisfactory because of its indistinct pathogenesis and specific biomarkers. Thus, we comprehensively utilized the microarray data from Gene Expression Omnibus database to uncover the biomarker and mechanism underlying DCM development. It's found that collagen type III alpha 1 chain and collagen type I alpha 2 chain, which regulated by hsa-miR-5682 and hsa-miR-4500, may play critical roles in the pathogenesis of DCM through acute inflammatory response and IL-17 signaling pathway. These biomarkers and mechanism need to be further studied.

INTRODUCTION

Dilated cardiomyopathy (DCM) is a progressive myocardial disease. It accounts for 30–40% of heart-failure cases and leads to high mortality worldwide[1]. DCM is characterized by biventricular dilatation, cardiac systolic dysfunction, and ventricular remodeling[2]. Recently, several studies have reported that mutations, myocarditis, hypertension, and ischemia are the induction factors of DCM[3, 4]. Increasing evidence shows that various gene mutations and biomarkers are associated with DCM[5-7]. Mutations in cytoskeletal proteins, including dystrophin[8] and desmin[9], impair muscular force transmission and thereby contribute to the development of DCM. Mutations in lamin A/C, a nuclear membrane protein, usually cause DCM with atrioventricular block and atrial fibrillation [10]. Li *et al* have reported that mutation of aryl hydrocarbon receptor nuclear translocator-like protein-1 (known as BMAL1) plays a critical role in the development of DCM through regulation of mitochondrial fission and mitophagy *via* the mitochondrial protein B cell leukemia/Lymphoma 2 interacting protein 3[11]. Mutations in thin-filament regulatory proteins, including cardiac troponin

T, cardiac troponin I, and α-tropomyosin, can cause DCM with systolic dysfunction by reducing fractional shortening and systolic Ca^{2+} level[12]. Moreover, some biomarkers associated with the development of DCM have been reported. For example, syndecan-1 and 4 may serve as useful biomarkers for predicting adverse cardiovascular events in DCM patients[13, 14]. Carbonic anhydrase 2 and 3 have been shown associated with heart failure and are potential risk biomarkers for DCM[15], and serum fibroblast growth factor 21 Level is linked to the prognosis of DCM[16].

Several studies have sought DCM-related genes and mechanisms *via* bioinformatic methods and found some meaningful results. Huang *et al* have found that Fos proto-oncogene, AP-1 transcription factor subunit, TIMP metallopeptidase inhibitor 1, and serpin family E member 1 may serve as therapeutic targets in DCM[17]; Zhao *et al* have identified 89 differentially expressed genes (DEGs), mainly enriched in the extracellular matrix and biological adhesion signaling pathways, may play significant roles in the development of DCM [18]. However, the main cause(s) and pathogenic mechanism(s) underlying DCM are still elusive, thus DCM is mostly diagnosed late, which causes poor prognosis in turn. More studies are urgently needed to improve the diagnostic and therapeutic efficiency in DCM. The Gene Expression Omnibus (GEO) database includes many DCM-related microarray data, which have not been fully utilized. These data can be used to identify additional candidate biomarkers and pathways to further explore the cause(s) of DCM. In order to investigate the candidate genes and pathways involved in DCM patients, we analyzed the two gene expression data sets GSE3585 and GSE5406. Using the "linear models for microarray data" (limma) method, we identified 97 DEGs between healthy individuals and DCM patients. In addition, we sought the mechanisms commonly regulated by the DEGs *via* Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses and gene set enrichment analyses (GSEA). Moreover, a protein-protein interaction network (PPI) network was applied to identify the hub genes that may contribute to the pathogenesis of DCM and to predict the miRNAs targeting the hub genes. Furthermore, we investigated the pattern of immune-cell infiltration in DCM.

MATERIALS AND METHODS

Microarray data extraction from the GEO database

The mRNA expression profiles GSE3585 and GSE5406 in the GEO database (http://www.ncbi.nlm.nih.gov/geo/), which is a shared platform in which researchers deposit their microarray data related to various diseases, were downloaded. The GSE3585 dataset, generated by Barth *et al*[19], and the GSE5406 dataset, generated by Hannenhalli *et al*[20], consisted of 7 DCM patients and 5 healthy individuals, and 86 DCM patients and 16 healthy individuals, respectively. In total, 114 samples of the left ventricular myocardium, consisting of 93 DCM and 21 healthy samples (control group), were included in this study.

Data processing and DEGs identification

The two datasets GSE3585 and GSE5406 were loaded onto the GPL96 platform (Affymetrix Human Genome U133A Array, [HG-U133A]). Additionally, the series matrix and platform annotation for the two databases were downloaded from the GEO database. The probe IDs were transformed into gene symbols. Then, *via* the R package "Surrogate Variable Analysis"(SVA), the two databases were merged, and any batch effect was removed using the "Empirical Bayes" method[21]. The R package "limma" was applied to identify the DEGs between the DCM and healthy myocardium tissues[22]. The screening criteria were set as $P < 0.05$ and $|\log \text{foldchange (FC)}| > 0.589$ ($FC > 1.5$). Volcano and heat maps were generated using the R software.

Gene expression enrichment analysis

The gene expression enrichment analysis in this study included GO analysis (https://www.geneontology.org)[23], KEGG (https://www.genome.jp/kegg)[24] pathway analysis, and GSEA (https://www.gsea-msigdb.org/gsea) analysis[25]. The DEGs were inputted into Metascape (https://metascape.org)[26], (1) the species was selected as *Homo sapiens*, (2) the screening standard was set as $P < 0.05$, (3) the GO terms

of biological process (BP), cellular component (CC) and molecular function (MF) were analyzed and KEGG pathway analysis was performed with the criteria of $P < 0.05$. GSEA interprets the biological function of expression dataset. The expression dataset in the DCM cases *vs* healthy tissues was loaded into GSEA 4.0.3 software, set gene sets database as GO gene set (c5.all.v7.1.symbols.gmt), set number of permutations as 1000, set phenotype labels as control *vs* DCM, set collapse/remap to gene symbols as collapse, set permutation type as phenotype, and the other parameters were set at default parameters. Then the GSEA software was used to obtain the enrichment results. The enriched terms were defined as significant with nominal $P < 0.05$.

PPI network construction and hub-gene identification

PPI network constructed by online website STRING (https://string-db.org/cgi/input.pl)[27] can contribute to the understanding of the interactive relationship among DEGs. The DEGs were inputted into this website, species of *Homo sapiens* was selected, and the identification criterion was set as combined score > 0.4 (medium confidence). Then the profile of interacting node pairs was imported into Cytoscape (https://cytoscape.org)[28] to visualize the PPI network. The top 10 hub genes were identified with the standard of connectivity degree by using the CytoHubba plugin. The plugin Molecular Complex Detection (MCODE) was applied to identify the essential module within the PPI network in Cytoscape with the default parameters (degree cutoff, 2; node score cutoff, 0.2; k-core, 2; and maximum depth, 100).

Construction of the miRNA-mRNA interaction network

miRNAs, a class of small non-coding RNAs, regulate the expression of various genes by binding to their transcripts and play critical roles in DCM progression [29]. By using the miRDB[30] (http://mirdb.org/) database, we predicted miRNAs targeting any of the top 10 hub genes. We then sorted these miRNAs according to their prediction

scores and selected the top 10 miRNAs. Then the mRNA-miRNA pairs were imported into Cytoscape to visualize the miRNA-mRNA network.

Immune cells infiltration analysis

The CIBERSORT (cibersort.stanford.edu) algorithm was applied to analyze the normalized gene expression data, and the proportions of 22 types of immune cells in each sample were analyzed [31]. The gene expression data were normalized *via* "limma" and transformed into the 22 types of immune-cell expression data through the source of CIBERSORT [32] *via* R. Then, the results were filtered out *via* Perl (https://www.perl.org) with $P < 0.05$, and the immune-cell infiltration matrix was obtained. Next, the "vioplot" package was used to draw violin diagrams to visualize the difference in immune-cell infiltration between the DCM and healthy groups in detail. The "ggplot2"[33] package was applied to perform principal component analysis (PCA) and draw a PCA clustering map. The "corrplot" package was used to analyze the correlation among immune-cell infiltration and draw a correlation heatmap.

RESULTS

Identification of DEGs

After merging the two datasets, 97 DEGs, including 47 upregulated and 50 downregulated genes, were obtained in the DCM group compared with the control group. Figure 1 shows the volcano map and heatmap of the 97 DEGs. The details of the top 10 upregulated or downregulated genes are shown in Table 1.

GO and KEGG enrichment analysis

Next, the DEGs were used to perform enrichment analysis for BP, CC, MF, and KEGG pathways. By using the Metascape website, the BP of GO was found significantly enriched in "response to growth factor," "blood vessel development," "regulation of smooth muscle cell proliferation," "muscle tissue development," and "acute inflammatory response" (Figure 2A). The DEGs in CC were mainly enriched in

"extracellular matrix," "cytoplasmic vesicle lumen," "collagen trimer," and "sarcoplasm" (Figure 2B). Regarding MF, the DEGs were mainly enriched in "extracellular matrix structural constituent," "growth factor binding," "alpha-actinin binding," and "calcium ion binding" (Figure 2C). KEGG pathway analysis revealed significant pathway enrichment of DEGs in "protein digestion and absorption," "IL-17 signaling pathway," "AGE-RAGE signaling pathway in diabetic complications," "complement," and "coagulation cascades" (Figure 2D).

GSEA analysis

The GSEA analysis results revealed that, compared with the control group, the DCM group was significantly enriched in GO terms "heart development," "response to ischemia," "vascular smooth muscle cell differentiation," "response to transforming growth factor beta," "stem cell proliferation," and "regulation of mitochondrial fission" (Figure 3).

PPI network and identification of hub genes

To further explore the relationship among the DEGs at the protein level, the PPI network of the 97 DEGs was constructed using STRING with the criterion of combined score > 0.4 and visualized using Cytoscape. The PPI network consisted of 77 nodes and 145 edges (Figure 4A). The top 10 hub genes included collagen type III alpha 1 chain (COL3A1), collagen type I alpha 2 chain (COL1A2), signal transducer and activator of transcription 3 (STAT3), C-C motif chemokine ligand 2 (CCL2), fibromodulin (FMOD), asporin (ASPN), C-X-C motif chemokine ligand 12 (CXCL12), lumican (LUM), heat shock protein 90 alpha family class A member 1 (HSP90AA1), and osteoglycin (OGN) (Figure 4B). The detailed information of these hub genes is provided in Table 2**.** MCODE analysis identified five essential modules, and COL3A1, myosin heavy chain 6, activating transcription factor 3, B cell leukemia/Lymphoma 6 transcription repressor, and pentraxin 3 were the seeds of clusters 1, 2, 3, 4, and 5, respectively (Figure 4C–G).

miRNA–mRNA interaction network

Increasing evidence shows that miRNAs play important roles in the development and progression of DCM. By using the miRDB database, we predicted miRNAs targeting any of the top 10 hub genes. We then sorted these miRNAs according to their prediction scores and selected the top 10 miRNAs. Additionally, the top 100 miRNA–mRNA pairs were visualized using Cytoscape (Figure 5). Consequently, hsa-miR-5682, hsa-miR-4500, hsa-miR-32-3p, and hsa-miR-374a-3p were each found to target ≥ 2 hub genes.

Immune-cell infiltration in DCM

Immune cells infiltrate into the myocardium upon myocardial injury[28]. Thus, a violin plot was constructed to investigate the difference in immune-cell infiltration between the DCM and control groups (Figure 6A). Compared with the control group, the DCM group had more infiltrated plasma cells and fewer infiltrated B memory cells, T follicular helper cells, and resting dendritic cells, whereas there was no significant difference in the remaining 18 types of immune cells. However, principal component analysis (PCA) results showed that the control and DCM groups could not be well distinguished according to the infiltration patterns of the 22 types of immune cells (Figure 6B). We generated a correlation heatmap to assess the correlation among the 19 immune cells that were found infiltrated in the DCM or control group. As shown in Figure 6C, the number of infiltrated B memory cells were positively correlated with that of the infiltrated resting dendritic cells, activated natural killer (NK) cells and T follicular helper cells; the number of infiltrated activated NK cells was negatively correlated with that of infiltrated resting NK cells; and the number of infiltrated resting memory CD4 T cells was negatively associated with that of infiltrated B memory cells and regulatory T cells.

DISCUSSION

DCM is one of the main reasons of sudden cardiac death and heart failure. It is a heterogeneous disease caused by various types of pathogenic factors, including genetic,

infectious, hormonal, and environmental factors[34]. The causes of DCM should be explored in depth to improve the diagnosis, treatment, and prognosis of DCM patients. Therefore, it is of great significance to elucidate the genetic mechanisms involved in DCM.

In this study, 97 DEGs, consisting of 47 upregulated and 50 downregulated genes, were identified between the DCM and control groups. GO of BP and GSEA analysis revealed that the DEGs were not only enriched in the development of the cardiovascular system, such as in the development of the muscle tissue, heart, and blood vessels, but also in the etiology of DCM, such as in acute inflammatory response and mitochondrial fission (Figures 2 and 3). Growing evidence shows that infiltration of inflammatory cells is associated with pathogenesis of DCM[35-37], and prelamin A accumulation [38] and Myosin Binding Protein C3 mutation [39] can promote DCM pathogenesis *via* regulation of inflammation. Xia *et al* have reported that dynamin-related protein 1 (Drp1, myocardial fission protein) is significantly upregulated in DCM patients. Moreover, Sacubitril (known as LCZ696), a novel inhibitor of the angiotensin receptor neprilysin, can alleviate the cardiac dysfunction in doxorubicin-induced DCM and reduce apoptosis by inhibiting mitochondrial fission *via* the Drp1-mediated pathway[40].

Regarding CC, the DEGs were enriched in sarcoplasm. Previous studies have reported that mutations in phospholamban (related to abnormal contractility)[41] and BCL2-associated athanogene 3 (alter the cardiac response)[42] are closely associated with DCM in the sarcoplasmic reticulum. GO analysis of MF indicated that the DEGs were enriched in alpha-actinin binding and calcium ion binding. Alpha-actinin and calcium ion are critical for myocardial contraction[43, 44]. Other studies have demonstrated that most DCM patients exhibit abnormalities related to calcium ion and α-actinin, which cause decreased heart contractility [45-47]. Cardiac troponin contributes to myocardial contraction [48]. Mutations in cardiac troponin T (TNNT2), troponin C (TNNC1) and troponin I (TNNI3) are mainly related to DCM pathogenesis [49, 50]

KEGG pathway analysis showed that the DEGs were significantly enriched in the IL-17 signaling pathway. It has been reported that DCM induced by viral myocarditis, accompanying autoimmune dysfunction, affects secretion of the IL-17 cytokine by Th17 cells, and IL-17 itself promotes myocardial cell injury[51]. Wang *et al* have reported that elevated IL-17 Levels are significantly associated with DCM incidence and progression[52]. Additionally, the serum levels of other inflammatory factors, such as interleukin-6, tumor necrosis factor-α, and IL-21 are significantly increased in DCM patients[53]. Thus, the IL-17 signaling pathway may be one of the major signaling pathways involved in the development of DCM.

Through construction of PPI and miRNA–mRNA interaction networks, we identified the hub genes and the miRNAs targeting them. The hub genes, COL1A2 and COL3A1, encode the pro-alpha2/1 chains of type I and III collagens, respectively. Collagens I and III, the main collagens of cardiac extracellular matrix, are classical biomarkers of cardiac fibrosis in DCM[54]. Mihailovici *et al* have reported that collagens I and III are upregulated in DCM patients compared with matched healthy controls[55]. Additionally, Zhao *et al* have reported that COL1A2 may participate in DCM pathogenesis by regulating the cardiac remodeling characterized by collagen deposition in the extracellular matrix[18, 56]. Consistent with our results, Zhang *et al* have identified signal transducer and activator of transcription 3 (STAT3) as a hub gene in DCM *via* bioinformatic analysis. Other studies also indicate a role of STAT3 in DCM. Podewski *et al* have shown that STAT3 protein level is significantly decreased in the cardiomyocytes of patients with end-stage DCM[57]. Moreover, inhibition of the IL-6–mediated STAT3 signaling pathway can improve myocardial remodeling through reducing myocardial apoptosis in a mouse model of DCM[58]. Thus, the roles of COL1A2, COL3A1, and STAT3 in DCM should be further investigated. Moreover, the identified miRNAs hsa-miR-5682, hsa-miR-4500, hsa-miR-32-3p, and hsa-miR-374a-3p may participate in DCM pathogenesis through their interaction with ≥ two hub genes. Previous studies have also suggested that miRNAs play significant roles in DCM. It has been found that miR-21, miR-29a, and miR-133b are differentially expressed in DCM patients[59]. miR-133a

expression is associated with fibrosis, myocyte necrosis, LV function, and clinical outcome in patients with inflammatory DCM[60]. Moreover, Satoh *et al* have shown that a low let-7i level can serve as an independent predictor of cardiac death and HF (relative risk = 3.76)[61].

Immune cells commonly infiltrate into the myocardium upon various types of cardiac damage[49, 62]. Over-activating of immune cell could be investigated in pathological examination about cardiac biopsy specimens in DCM patients. Noutsias *et al* have reported that upregulation of genes associated with T cells exacerbates DCM progression[63]. Therefore, our study also assessed for the correlation between the DEGs and immune-cell infiltration. The results indicated infiltration of 19 types of immune cells in DCM pathogenesis. Notably, compared with the control group, the DCM group had more infiltrated plasma cells and fewer infiltrated B memory cells, T follicular helper cells, and resting dendritic cells. However, Liu *et al* have demonstrated that, compared with healthy controls, DCM patients have more infiltrated T follicular helper cells and fewer T follicular regulatory cells, and infiltration of T follicular regulatory cells is positively correlated with left ventricular ejection fraction[53].

Our study has some limitations. First, the gene expression data were acquired from a public database. Moreover, we did not experimentally verify the relevance of the identified DEGs with DCM and their enriched functions or hub genes. Likewise, we did not verify the predicted miRNA–mRNA interactions and their relevance.

CONCLUSION

In summary, we firstly identified that COL1A2 and COL3A1 may be both presumably regulated by hsa-miR-5682 and hsa-miR-4500, and play significant roles in the pathogenesis of DCM through acute inflammatory response and IL-17 signaling pathway. These results may provide useful biomarkers for the diagnosis and treatment of DCM, but further research is needed to clarify the roles of the predicted genes and pathways.

73324_Auto_Edited.docx

ORIGINALITY REPORT

2%
SIMILARITY INDEX

PRIMARY SOURCES

1	Maria Calderon-Dominguez, Thalía Belmonte, Maribel Quezada-Feijoo, Monica Ramos-Sánchez et al. "Emerging role of microRNAs in dilated cardiomyopathy: evidence regarding etiology", Translational Research, 2020 Crossref	24 words — 1%
2	assets.researchsquare.com Internet	24 words — 1%
3	toolshed.g2.bx.psu.edu Internet	21 words — 1%
4	www.hindawi.com Internet	21 words — 1%

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	< 1%
EXCLUDE BIBLIOGRAPHY	ON		