Bioinformatics analysis of differentially expressed genes in ischemic cardiomyopathy using GEO Database

Rui Yan*, Jiahui Song, Ming Guo, Minghui Hao
Department of Cardiology, Beijing Luhe Hospital, Capital Medical University, Beijing, 101199, China

ARTICLE INFO

Original paper
Article history:
Received: November 19, 2022
Accepted: January 20, 2023
Published: January 31, 2023

Keywords:
bioinformatics, differential expression, ischemic cardiomyopathy (ICM), database, gene expression

ABSTRACT

It was to analyze differentially expressed genes and their expression characteristics in ischemic cardiomyopathy (ICM) by bioinformatics and provide targets for drug therapy of ICM. For this purpose, the gene expression data of ICM in the gene expression omnibus (GEO) database were used, the differentially expressed genes between healthy myocardium and ICM myocardium were screened by R language, and then the differentially expressed genes were analyzed by protein-protein interaction (PPI), gene ontology (GO), and KEGG pathway analysis was performed for the differentially expressed genes in ICM tissues, including the main pathways: viral carcinogenesis, energy metabolism, viral response, oxidative phosphorylation, influenza A, extracellular matrix receptor interaction, Epstein-Barr virus infection, chemokine receptor pathway, phagosome, proteasome, and protein digestion and absorption. PPI network analysis showed that C3, F5, FCGR3A, APOB, PENK, LUM, CHRD1L, FCGR3A, CIQB, and FMOD were critical genes. In conclusion, bioinformatics can screen out the key genes in ICM, which is helpful to understand the treatment of drug targets in ICM patients.

Introduction

Ischemic cardiomyopathy (ICM) is a common and frequent acquired heart disease in middle-aged and elderly people, also known as senior-specific cardiomyopathy, which refers to localized or diffuse fibrosis of the myocardium due to long-term myocardial ischemia, which produces impaired systolic and diastolic function and causes clinical manifestations such as cardiomegaly or hypotension, congestive heart failure, and arrhythmia (1,2). It mainly includes coronary obstruction or stenosis caused by atherosclerotic lesions (3). Most patients with congested ICM occur, mainly in the middle-aged and elderly male population, with a male-to-female ratio of approximately: 5:1 to 7:1 (4-5). Clinical symptoms of ICM include angina pectoris, heart failure, arrhythmia, thrombosis, and embolism (6). The pathogenesis of ICM is mainly due to long-term myocardial ischemia, which is closely related to coronary heart disease, and some patients with ICM present with painless myocardial ischemia or myocardial infarction (7). ICM is associated with environmental and genetic factors, and many studies have investigated gene expression in ICM from different aspects (8-9).

Bioinformatics includes biology, statistics, information technology, and other disciplines. Using data analysis software and online data platform to screen disease genes and explore the pathogenesis of diseases has become a new research method. In recent years, big data and multi-omics technology have been rapidly developed, and bioinformatics has become a hot spot in the study of disease diagnosis and typing, and pathogenesis (10). Long non-coding RNA (lncRNA) is a kind of non-coding RNA, which markedly means all non-coding transcripts participate in many critical biological processes. Gene chip has high throughput characteristics and is more adopted in gene expression analysis. Most of the genome is “non-coding”, which generates many non-coding RNA (ncRNA), including microRNA (miRNA) and IncRNA (11). Emerging evidence suggests that these classes of ncRNA are involved in most aspects of cardiac gene expression, cardiomyocyte proliferation, differentiation, and stress-responsive cardiac remodeling. Recent findings have demonstrated the important function of ncRNA in ischemic and non-ischemic cardiomyopathy. It is expected that ncRNA will become a promising therapeutic target for cardiovascular diseases (12-13). As a complex disease with multiple genes and multiple signaling pathway interactions, ICM-related gene differential expression has still not been fully elucidated. The discovery of new ICM differentially expressed genes and understanding the related signaling pathways have become urgent problems (14-15). The use of public databases to effectively mine and develop gene expression of related diseases also makes it possible to use intelligent systems to study the acquisition of biological life and the occurrence and development of diseases, which has a profound impact on the prevention and treatment of diseases (16). Gene chips can detect a large amount of gene information in a short time, which has become an important means to study the gene expression profile of diseases (17). Relevant studies have been presented in large numbers. Many gene information data are uploaded to public databases. Many studies have not been effectively applied because of

* Corresponding author. Email: yanrui_med@ccmu.edu.cn
Cellular and Molecular Biology, 2023, 69(1): 98-103

Copyright: © 2023 by the C.M.B. Association. All rights reserved.
the differences in research objectives, and many potential molecular markers for clinical diagnosis/prognosis need to be discovered.

Therefore, based on previous research, bioinformatics technology was used to select and analyze the gene chip data, explore the associated genes of ICM pathogenesis, and construct a protein-protein interaction (PPI) network with differential expression, screening related genes. It aimed to understand more ICM-related proteomics molecular system biology rules and provide new ideas for the rehabilitation of patients in clinical practice.

Materials and Methods

Public databases
The gene expression omnibus (GEO) database is the largest public gene expression data repository at this stage, and it is a gene expression database (18). It includes a platform (GPL), sample (GSM), and series (GSM). GSM is a dataset that puts all sample information of a certain experiment together; GSM mainly records the conditions for processing a single sample and the basic information of the sample; GPL contains the sequence and a brief description of the array platform. A total of 145,432 series, 4,242,486 samples, and 21,920 platforms were included in the GEO database by February 2021.

KEGG database
KEGG database is a biological database integrating genome, enzymatic pathway, and biochemical, which has the function of understanding biological systems. This database links the information and gene function of the genome, shows the connection between genome information and function in the form of a graph, and describes metabolic pathways more intuitively.

STRING database
The search tool for the retrieval of interacting genes (STRING) database (http://string-db.org/), is a repository that provides PPI data. The protein names corresponding to the selected differential genes can be input into STRING to obtain the interaction network between the encoded proteins in the STRING database, and the connectivity of each differential protein can also be calculated. HUB protein has high connectivity and scores the role between proteins. After the comprehensive score is weighted, the score is obtained. The higher the score, the greater the correlation between genes is shown.

Functional enrichment analysis
Gene ontology (GO) enrichment analysis is adopted, and biological processes, molecular functions, and cellular components can be identified using this information, enabling the extraction of genes with the highest and lowest values from gene expression profiles to produce up and down-regulated genomes. P < 0.05 was taken as the cut-off after the Benjamini-Hochberg correction was used. Biological pathways are also sequences that interact between biological compounds that play critical roles in cytological behavior.

Data sources
In public databases (www.ncbi.nlm.nih.gov), the gene chip datasets GSE26887 and GSE42955 datasets (species: Homo sapiens) were downloaded, and gene expression was detected using the GPL6244 detection platform. Then, Affymetrix Gene Chips Human Gene 1.0 ST chip was selected to detect gene expression. These included RNA expression profiles of genes in 7 ICM patients with diabetes, 5 healthy controls, and 12 ICM patients without diabetes, with 5 healthy samples as the control group, and 12 ICM samples as the ICM group. There was no significant difference in general information, name, age, deformation, smoking, hypertension, BMI, education, and occupation, between the two groups (P > 0.05) GSE26887 dataset and GSE42955 dataset were obtained by the GEO query package in the R language, and then gene chip probes were annotated by hguene10sttranscriptcluster. DB package to obtain gene expression values and gene names, and the data that the probe did not match the gene name and the data with one gene name corresponding to multiple probes were removed.

The top 10 hub genes were selected by the plug-in CytoHubba, and according to the Venny2.1 tool (http://bioinfogp.cnb.csic.es/tools/venny/), a Venn diagram was made.

Results

Venn diagram of differentially expressed genes
There were 19,776 differential genes in GSE42955, and 2,153 genes with significantly altered FC values were selected, including 774 up-regulation genes and 1,379 down-regulation genes, as shown in Figure 1.

As shown in Figure 2, GSE26887 had 18,843 differentially expressed genes, and 259 genes with significantly altered FC values were selected, including 135 up-regulation genes and 124 down-regulation genes. In the two

Figure 1. Venn diagram of differentially expressed up-regulated genes in the two datasets.

Figure 2. Venn diagram of differentially expressed down-regulated genes in the two datasets.
datasets, 58 differential genes were selected, containing 34 genes with increased expression and 24 genes with decreased expression.

**Volcano plot of differentially expressed genes**

As shown in Figure 3, two datasets volcano plots are shown, A is the GSE26887 dataset volcano plot, and B is the GSE42955 dataset volcano plot, with red indicating up-regulated genes and blue indicating down-regulated genes in the plot.

Differential gene data analysis was performed in both datasets. Figure 4 is a clustered heatmap. It was found that the normal tissue and differential expression data within the ICM group were good, showing a good expression consistency, which also showed that the extracted data in the dataset could be used for the next step of the analysis.

**GO functional enrichment analysis**

Differentially expressed genes in ICM tissues were enriched, and GO analysis (with P < 0.01 as the inclusion criterion) and KEGG analysis (with P < 0.01 as the inclusion criterion) were performed. Genes SERPINE1, SERPINA3, TNC, SPPI, S100A8, CYP1B1, ANKR2D, CD163, MYC, and GFPT2 were up-regulated in the gene chip dataset GSE26887 dataset. Genes ASB14, USP9Y, SLN, FRZB, UTY, MYL4, DSC1, NEB, NPPA, and EIF1AY were down-regulated (Table 1).

**KEGG pathway analysis of differential genes**

KEGG pathway analysis of differentially expressed genes in ICM tissues included 14 pathways, which were: viral carcinogenesis, body heat production, viral response, oxidative phosphorylation, influenza A, extracellular

<table>
<thead>
<tr>
<th>Category name</th>
<th>Name of article</th>
<th>Number</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down-regulated gene</td>
<td>Biological processes</td>
<td>66</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Immune response</td>
<td>63</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Inflammatory response</td>
<td>47</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Virus defense response</td>
<td>34</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>T cell receptor signaling pathway</td>
<td>31</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Biological processes</td>
<td>31</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Type I interferon signaling pathway</td>
<td>26</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>C-type lectin receptor signaling pathway</td>
<td>21</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>TAP-dependent antigen processing and presentation</td>
<td>18</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Amino acid biosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Viral myocarditis</td>
<td>22</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Up-regulation process</td>
<td>Biological processes</td>
<td>63</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Metabolic pathways</td>
<td>26</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Parkinson’s disease</td>
<td>18</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Extracellular matrix tissue</td>
<td>18</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>P13K-Akt signaling pathway</td>
<td>15</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Huntington’s disease</td>
<td>15</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Alzheimer’s disease</td>
<td>15</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Oxidative phosphorylation</td>
<td>15</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Mitochondrial translation extension</td>
<td>11</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>Protein digestion and absorption</td>
<td>10</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>
matrix receptor interaction, Epstein-Barr virus infection, chemokine receptor pathway, phagosomes, proteasomes, protein digestion and absorption, tuberculosis, pertussis, antigen processing and presentation (Figure 5).

PPI network diagram
STRING was selected for the network construction of PPI (https://string-db.org/), genes were input, and differential genes in ICM were selected for input to construct PPI network maps (Figure 6).

Discussion
Bioinformatics is also database mining, which has the advantages of large sample size, low cost, simplicity, and high efficiency. It uses different databases to obtain a large amount of data, develops relevant algorithms and software, integrates and compares the mechanism of biological development, and identifies relevant biomarkers. It has become a way to investigate disease gene mutations. Transcriptome sequencing technology and expression microarray technology are widely used in gene differential expression, and these technologies can determine transcriptional differences in different pathological conditions and find brand-new biomarkers to provide more information for the occurrence and development of diseases (19). Enrichment analysis is the main method to identify the biological process, molecular function, and cellular components involved in differential genes, and KEGG enrichment analysis is the main method to identify the cellular signaling pathways involved in differential genes. GO and pathway analysis was performed on genes differentially expressed in ICM by bioinformatics methods, and the results showed that significant enrichment in fibrotic function and pathways was associated with endoplasmic reticulum stress. ICM belongs to a type of end-stage coronary atherosclerosis. If the body is in an aerobic and oxygen-supply imbalance environment for a long time, it will cause myocardial apoptosis and necrosis, and the heart will also fibrosis.

Myocyte necrosis is a pathological process that is irreversible, and cardiac depression is an effective treatment that can replace the damaged heart, but there are still many limitations between donors and immunity that are worth exploring in depth. Many studies have focused on elucidating the coding genes of proteins and their molecular mechanisms, and it is very important to find related differential genes. Chen et al. (2021) (20) investigated the underlying central genes and pathways of ICM and explored the possible related mechanisms, MYH6 was found to be a potential gene for human ICM, and MYH6 expression was closely related to ICM and heart failure (HF). However, whether this marker can be used as a predictor in blood samples requires further experimental validation. ICM is a severe myocardial dysfunction that results from coronary heart disease, which belongs to the late stage of coronary heart disease and is the main factor of HF. Alimadadi et al. (2020) (21) stated that several downstream genes affected by TBX5, etc. are concerned with diseases that are related to ICM. Li et al. (2018) (22) investigated genes that changed evidently when ICM-induced HF is developed and found a total of 255 common DEG. GO, pathway enrichment, and PPI analysis suggested that nucleic acid binding proteins, enzymes, and transcription factors account for a not small proportion of DEG, with the changes of immune system signaling and cytokine signaling most obvious, 7 central genes and 9 transcription factors were definite. Li et al. (2018) (23) pointed out normal different genes of coronary atherosclerotic heart disease and ICM were screened, and pathway analysis and PPI network analysis of genes different were performed, and 575 genes with differences were included from GSE71226, having 350 increasingly expressed genes and 225 decreasingly expressed genes, and the differences had statistical meaning (P < 0.05, fold change > 1). 75 differential genes were selected from GSE9128. A total of eight common differential genes were selected and functional annotation and pathway analysis were performed to facilitate further investigation of interactions between differentially expressed genes.

Wang et al. (2022) (24) found potential biomarkers of SERPINA3, FCN3, PTN, CD163, and SCUBE2 associated with inflammatory responses in ICM. The proposed nomogram may provide a useful tool for clinicians. Qiu et al. (2016) (25) performed a comparative RNA mapping analysis of idiopathic dilated cardiomyopathy and ICM, and it revealed that the genes MYC and FN1 are hub genes, which master each module of DCM-specific and IC-specific DEG, respectively. They found commonalities and differences in gene expression profiles and molecular pathways between different cardiomyopathies. Zhang et al. (2022) (26) performed GO functional and KEGG

---

**Figure 5.** KEGG pathway analysis results. Bubble size represents the amount of enrichment, and color represents the P value level.

**Figure 6.** PPI network diagram. A in the figure is the differential gene selected from the gene chip dataset, and B in the figure is the key gene interaction diagram.
pathway analysis, and it revealed that DEG was enriched in metabolic pathways, oxidative phosphorylation, extracellular matrix receptor interaction, and so on, which were strongly associated with fibrosis, collagen catabolic process, and inflammatory response function, and a Hub gene regulatory network correlated with ICM IncRNA was built. Bioinformatics analysis of DEG of ICM was carried out; it was successful to build the Hub gene regulatory network of ICM. The analysis revealed that DEG was enriched in metabolic pathways, oxidative phosphorylation, extracellular matrix receptor interaction, Epstein-Barr virus infection, chemokine receptor pathway, phagosomes, proteasomes, etc. Li and Chen (2022) (27) carried out an enrichment analysis, and it suggested that normal genes were correlated with inflammatory response, immune response, PI3K/AKT, NF-κB, and TNF pathways. In addition, mmu-miR-92a-3p and mmu-miR-27b-3p were identified as central miRNA, and TNF, IL1B, and IGF1 were screened as critical nodes.

DEG is based on the AGE-RAGE signaling pathway. Advanced glycation end products are a group of heterogeneous proteins. AGE-RAGE signaling is activated to promote inflammatory mediators. Critical genes identified were C3, F5, FCGR3A, APOB, PENK, LUM, CHRDL1, FCGR3A, CIQB, and FMO2. Fibrosis is an important pathophysiological basis for the development of ICM. Retinoic acid binds to intracellular micro-formic acid-binding protein to play a role in inhibiting phosphorylation, alleviating isoproterenol-induced myocardial remodeling, and improving cardiac function.

Bioinformatics analysis of the gene chip of ICM revealed the pathogenesis of ICM was associated with endoplasmic reticulum stress and fibrosis, which could provide some reference for treating ICM. Limitations are that the data are derived from a database and not experimentally validated, and there may be some shortcomings in the analytical interpretation of the data due to the complexity of the body as well as the limitations of the database. There are many other core genes associated with ICM mechanisms, and they need to be further explored to provide a basis for IVM drug target therapy from different angles.

References


27. Li N, Chen Q. Analysis of the miRNA-mRNA Regulatory Network Reveals the Biomarker Genes in the Progression of Myocardial Ischemic Reperfusion. J Healthc Eng. 2022; 2022: 2045619.