Cellular and Molecular Biology
E-ISSN : 1165-158X / P-ISSN : 0145-5680
www.cellmolbiol.org

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 to select the key genes. Results showed that the useful genes of ICM were successfully screened in the GEO database, 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, CHRDL1, 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.

Doi: http://dx.doi.org/10.14715/cmb/2022.69.1.17

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 lncRNA (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

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 hugene10sttranscriptcluster.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

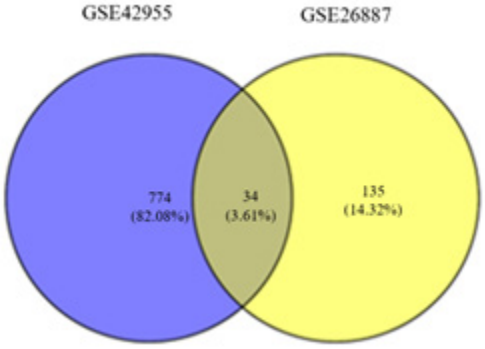

Figure 1. Venn diagram of differentially expressed up-regulated genes in the two datasets.

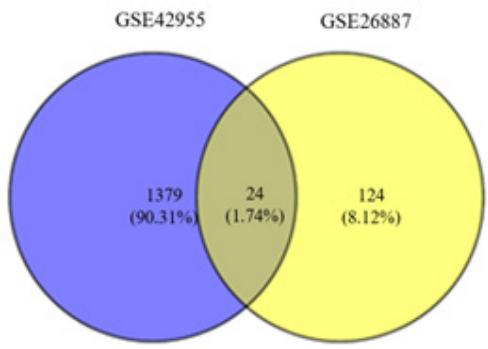

Figure 2. Venn diagram of differentially expressed down-regulated genes in the two datasets.

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 SERPINEI, SERPINA3, TNC, SPPI, S100A8, CYPIB1, ANKRD2, 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 EIFIAY 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

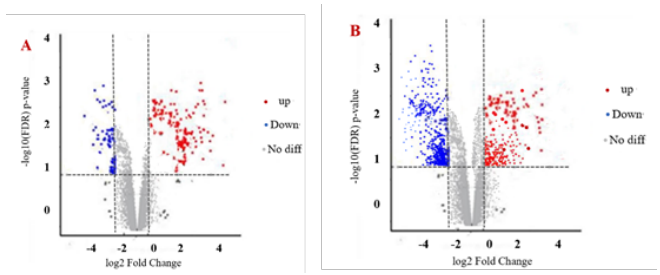

Figure 3. Volcano plot of two datasets. A: GSE26887 volcano plot, B: GSE42955 data plot.

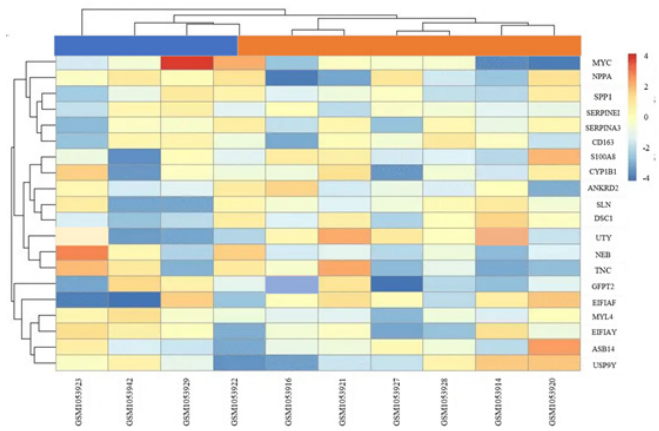

Figure 4. Cluster heatmap of differentially expressed genes. The top blue bars of the figure are normal control samples, orange bars indicate ICM patient samples, orange is up-regulated gene expression level, blue is down-regulated gene expression level, and the shade of the color represents the degree of change in expression level.

Table 1. GO analysis of down-regulated versus up-regulated differential gene expression.

	Category name	Name of article	Number	*P*
Down-regulated gene	Biological processes	Immune response	66	< 0.05
	Biological processes	Inflammatory response	63	0.001
	Biology	Virus defense response	47	< 0.05
	Cell components	T cell receptor signaling pathway	34	< 0.05
	Biological processes	Virus response	31	< 0.05
	Biological processes	Type I interferon signaling pathway	31	< 0.05
	Biological processes	C-type lectin receptor signaling pathway	26	< 0.05
	Biological processes	TAP-dependent antigen processing and presentation	21	< 0.05
	Molecular function	Amino acid biosynthesis	18	< 0.05
		Viral myocarditis	22	< 0.05
Up-regulation process				
	Biological processes	Metabolic pathways	63	< 0.05
	Biological processes	Parkinson's disease	26	< 0.05
	Molecular function	Protein folding	18	< 0.05
	Biological processes	Extracellular matrix tissue	18	< 0.05
	Cell components	P13K-Akt signaling pathway	15	< 0.05
	Biological processes	Huntington's disease	15	< 0.05
	Biological processes	Alzheimer's disease	15	< 0.05
	Molecular function	Oxidative phosphorylation	15	< 0.05
	Biological processes	Mitochondrial translation extension	11	< 0.05
	Biological processes	Protein digestion and absorption	10	< 0.05

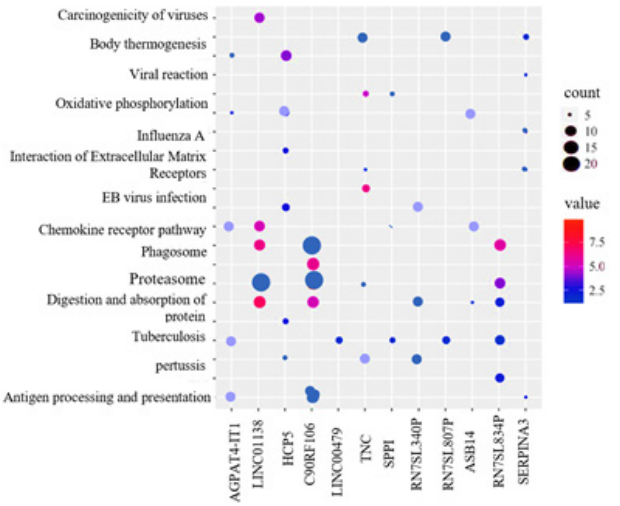

Figure 5. KEGG pathway analysis results. Bubble size represents the amount of enrichment, and color represents the P value level.

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

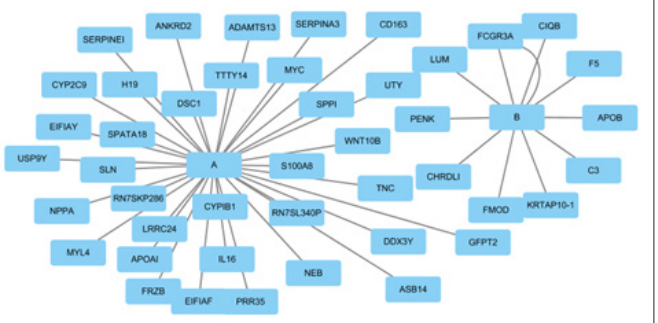

Figure 6. PPI network diagram. A in the figure is the differential gene selected from the gene chip dataset, and B in the figure is the key gene interaction diagram.

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

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 lncRNA 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 IFG1 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 FMOD. 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 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

1. Panza JA, Chrzanowski L, Bonow RO. Myocardial Viability Assessment before Surgical Revascularization in Ischemic Cardiomyopathy: JACC Review Topic of the Week. J Am Coll Cardiol 2021; 78(10): 1068-1077.
2. Beesley SJ, Weber G, Sarge T, Nikravan S, Grissom CK, Lanspa MJ, Shahul S, Brown SM. Septic Cardiomyopathy. Crit Care Med 2018; 46(4): 625-634.
3. Bansal SS, Ismahil MA, Goel M, Zhou G, Rokosh G, Hamid T, Prabhu SD. Dysfunctional and Proinflammatory Regulatory T-Lymphocytes Are Essential for Adverse Cardiac Remodeling in Ischemic Cardiomyopathy. Circulation 2019; 139(2): 206-221.
4. Divoky L, Maran A, Ramu B. Gender Differences in Ischemic Cardiomyopathy. CurrAtheroscler Rep 2018; 20(10): 50.
5. Del Buono MG, Moroni F, Montone RA, Azzalini L, Sanna T, Abbate A. Ischemic Cardiomyopathy and Heart Failure After Acute Myocardial Infarction. CurrCardiol Rep 2022; 24(10): 1505-1515.
6. Ródenas-Alesina E, Jordán P, Herrador L, Espinet-Coll C, Pizzi MN, Romero-Farina G, Aguadé-Bruix S, Ferreira-González I. Q waves in ischemic cardiomyopathy. Int J Cardiovasc Imaging 2021; 37(6): 2085-2092.
7. Richardson TD, Kanagasundram AN, Stevenson WG. Epicardial Ablation of Ventricular Tachycardia in Ischemic Cardiomyopathy. Card Electrophysiol Clin 2020; 12(3): 313-319.
8. Reith S, Kaestner W, Marx N, Burgmaier M. Parachute-Implantation bei schwerer ischämischer Herzinsuffizienz [Parachute Implantation in Severe Ischemic Cardiomyopathy]. Dtsch Med Wochenschr. 2017 Apr;142(8):586-594. German. doi: 10.1055/s-0042-109889. Epub 2017 Apr 21. PMID: 28431444.
9. Razeghian-Jahromi I, Matta AG, Canitrot R, Zibaeenezhad MJ, Razmkhah M, Safari A, Nader V, Roncalli J. Surfing the clinical trials of mesenchymal stem cell therapy in ischemic cardiomyopathy. Stem Cell Res Ther 2021; 12(1): 361.
10. Zhai Z, Qin T, Liu F, Han L, Zhou H, Li Q, Xia Z, Li J. Identification of atrial fibrillation-related circular RNAs and constructing the integrative regulatory network of circular RNAs, microRNAs and mRNAs by bioinformatics analysis. Cell Mol Biol (Noisy-le-grand) 2020; 66(7): 161-168.
11. Bührke A, Bär C, Thum T. Nichtkodierende RNA: Innovative Regulatoren mit therapeutischer Perspektive [Non-coding RNA : Innovative regulators with therapeutic perspective]. Herz. 2018 Mar;43(2):115-122. German. doi: 10.1007/s00059-017-4660-4. PMID: 29236145.
12. Liu DD, Wang HZ, Han X, Han CP, Ren FB.Identification of potential key genes associated with cardiac fibrosis by RNA sequencing data analysis. Acta Media Mediterranea 2019; 35: 2315.
13. Gao J, Xu W, Wang J, Wang K, Li P. The Role and Molecular Mechanism of Non-Coding RNAs in Pathological Cardiac Remodeling. Int J Mol Sci 2017; 18(3): 608.
14. Chen C, Huang H, Wu CH. Protein Bioinformatics Databases and Resources. Methods Mol Biol 2017; 1558: 3-39.
15. Huang X, Liu S, Wu L, Jiang M, Hou Y. High Throughput Single Cell RNA Sequencing, Bioinformatics Analysis and Applications. Adv Exp Med Biol 2018; 1068: 33-43.
16. Alkhnbashi OS, Meier T, Mitrofanov A, Backofen R, Voß B. CRISPR-Cas bioinformatics. Methods 2020; 172: 3-11.
17. Backofen R, Engelhardt J, Erxleben A, Fallmann J, Grüning B, Ohler U, Rajewsky N, Stadler PF. RNA-bioinformatics: Tools, services and databases for the analysis of RNA-based regulation. J Biotechnol 2017; 261: 76-84.
18. Barrett, T., etal. NCBI GEO:mining tens of millions of expression profiles-database and tools update. Nucleic Acids Res 2007; 35(Database issue): p.D760-5.
19. Ibrahim B, McMahon DP, Hufsky F, Beer M, Deng L, Mercier PL, Palmarini M, Thiel V, Marz M. A new era of virus bioinformatics. Virus Res 2018; 251: 86-90.
20. Chen JH, Wang LL, Tao L, Qi B, Wang Y, Guo YJ, Miao L. Identification of MYH6 as the potential gene for human ischaemic cardiomyopathy. J Cell Mol Med 2021; 25(22): 10736-10746.
21. Alimadadi A, Aryal S, Manandhar I, Joe B, Cheng X. Identification of Upstream Transcriptional Regulators of Ischemic Cardiomyopathy Using Cardiac RNA-Seq Meta-Analysis. Int J Mol Sci 2020; 21(10): 3472.
22. Li Y, Jiang Q, Ding Z, Liu G, Yu P, Jiang G, Yu Z, Yang C, Qian J, Jiang H, Zou Y. Identification of a Common Different Gene Expression Signature in Ischemic Cardiomyopathy. Genes (Basel) 2018; 9(1): 56.
23. Li GM, Zhang CL, Rui RP, Sun B, Guo W. Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy. Eur Rev Med Pharmacol Sci 2018; 22(11): 3553-3569.
24. Wang J, Xie S, Cheng Y, Li X, Chen J, Zhu M. Identification of potential biomarkers of inflammation-related genes for ischemic cardiomyopathy. Front Cardiovasc Med 2022; 9: 972274.

25. Qiu LL, Ding XJ, Zhu HT, Gao LW, Tang JF, Liu XQ. Comparative RNA profile analysis of idiopathic dilated cardiomyopathy and ischemic cardiomyopathy. Genet Mol Res. 2016; 15(1).
26. Zhang N, Yang C, Liu YJ, Zeng P, Gong T, Tao L, Li XA. Analysis of susceptibility genes and myocardial infarction risk correlation of ischemic cardiomyopathy based on bioinformatics. J Thorac Dis. 2022; 14(9): 3445-3453
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