

Identification of potential genes for human ischemic cardiomyopathy based on RNA-Seq data

Wan Li^{1,*}, Liansheng Li^{1,*}, Shiyong Zhang^{1,*}, Ce Zhang^{2,*}, Hao Huang¹, Yiran Li¹, Erqiang Hu¹, Gui Deng¹, Shanshan Guo¹, Yahui Wang¹, Weimin Li³ and Lina Chen¹

¹ College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China

² Department of internal medicine, Heilongjiang Commercial Hospital, Harbin, Heilongjiang, China

³ Department of Cardiology, the First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China

* These authors contributed equally to this work

Correspondence to: Lina Chen, **email:** chenlina@ems.hrbmu.edu.cn

Weiming Li, **email:** Liweimin_2009@163.com

Keywords: RNA-Seq; canonical correlation analysis; co-expression network; ischemic cardiomyopathy; Pathology Section

Received: August 02, 2016

Accepted: October 07, 2016

Published: November 12, 2016

ABSTRACT

Ischemic cardiomyopathy (ICM) is an important cause of heart failure, yet no ICM disease genes were stored in any public databases. Mutations of genes provided by RNA-Seq data could set a foundation for a variety of biological processes. This also made it possible to elucidate the mechanism and identify potential genes for ICM. In this paper, an integrated co-expression network was constructed using univariate and bivariate canonical correlation analysis for RNA-Seq data of human ICM samples. Three ICM-related modules were recognized after comparing between Pearson correlation coefficients of ICM samples and normal controls. Furthermore, 32 ICM potential genes were identified from ICM-related modules considering protein-protein interactions. Most of these genes were verified to be involved in ICM and diseases caused it by OMIM and literature. Our study could provide a novel perspective for potential gene identification and the pathogenesis for ICM and other complex diseases.

INTRODUCTION

Ischemic cardiomyopathy (ICM) is an important cause of heart failure. It is a common type of dilated cardiomyopathy (DCM) results from coronary heart disease (CHD), according to National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services. As other complex diseases, ICM is caused by interactions of genetic and environmental factors. Many studies have been conducted on ICM from different aspects, especially from the genome level [1, 2]. However, no ICM disease genes have been stored in public databases.

Nowadays, with the next-generation sequencing technology being widely used [3], studies for dysfunctions caused by gene mutations are still inadequate [4, 5]. Transcriptome data retrieved from next-generation sequencing data, especially RNA-Seq data, were superior to expression profiles from the microarray technology [6]. The high sensitivity of RNA-Seq data could provide

accurate and precise expression data from various levels, such as exon, position and allelic levels [7].

Genes causing complex diseases always participate in common biological processes in various kinds of biological networks [8, 9]. Of all biological networks, co-expression networks could provide information of co-regulation genes that function in regulation processes and relationships of transcriptome components disturbed by environment. Therefore, co-expression networks could be used as effective tools to study gene functions, biological processes and complex disease mechanisms [10]. Traditionally, correlation coefficients between expression values of gene pairs were used to evaluate their relationships and construct co-expression networks. Each expression value was represented by a single variable without considering expression differences caused by exons, positions or alleles [7, 11, 12]. On the contrary, canonical correlation analysis (CCA) could measure correlations between two sets of variables [13]. Thus, CCA could be used to examine complex patterns of

Table 1: Assortativity coefficients for various types of networks.

Network type	Assortativity coefficient
Biological network	-0.2305
Random network	-0.0489
The integrated co-expression network	-0.1233

gene expressions and accurately represent co-expression relationships by considering variations of transcripts. In co-expression networks, modules, or highly interconnected regions, are mutually correlated [14]. Genes in the same modules tend to have similar functions, or involve in common biological processes [15]. Comprehensive understanding of diseases could be provided by analyzing biological data based on network modules.

Hence, to construct co-expression networks to investigate genes in ICM process comprehensively, co-expression gene pairs were identified using univariate and bivariate CCA. Here, CCA was performed for the expression data of the exon, position and allelic levels obtained from RNA-Seq data of human ICM samples. Modules and their functions were further analyzed. Since protein-protein interaction (PPI) could reflect functions and cooperation of proteins/genes [16], sub-modules from PPI networks could help to identify potential genes of ICM (Figure 1).

RESULTS

The integrated co-expression network

Co-expression gene pairs of ICM samples for expression data of exons and positions were identified using the univariate CCA, and those for expression data of ASE were identified using the bivariate CCA (see Materials and Methods). The integrated co-expression network was constructed by integrating 17045 co-expression gene pairs between 1135 genes with non-zero correlations from all three levels.

Topological property of the integrated co-expression network

The assortativity of a biological network, a random network and the integrated co-expression network were measured by their assortativity coefficients (see Materials and Methods). It was showed that the biological network was disassortative (assortativity coefficient <0) as high degree nodes were more likely to connect to low degree nodes. For the random network, the assortativity

coefficient was close to 0. These results were consistent with previous studies [6]. Our integrated co-expression network was disassortative since the assortativity coefficient <0 (Table 1). These results indicated that our integrated co-expression network had similar topology features as biological networks.

ICM-related modules

Modules of the integrated co-expression network could reflect underlying pathological disease mechanisms. Here, network modules were detected using MCODE. After screening by comparing between Pearson correlation coefficients of ICM samples and normal controls (see Materials and Methods), 3 ICM-related modules were recognized and named as Module A, B and C (Figure 2).

BP_Fat of GO and pathways of KEGG enriched for these 3 modules using DAVID were selected as stated in Materials and Methods (P-Value <0.05 , Figure 3). All three modules were enriched in functions of “translation elongation”, “translation” and “generation of precursor metabolites and energy”, and the pathway of “Ribosome”. These enriched functions and pathways have been verified to be associated with ICM and diseases caused it (DCM and CHD).

The function of “generation of precursor metabolites and energy”, “electron transport chain” (enriched by Module A and C), the “Oxidative phosphorylation” pathway (enriched by Modules C) and other functions/pathways all result in the formation of energy. Moreover, mitochondrion is the site where all three modules were mainly localized, which indicated that energy was involved in ICM. It was showed that ICM and other cardiovascular diseases involved the misuse of energy and oxygen with high uptake and oxidation of fatty acids [17]. Moreover, ATP-binding cassette (ABC) B10 expression and heme levels were altered in hearts of patients with ICM. The mitochondrial transporter ABCB10 was reported to export heme out of the mitochondria. Heme plays a critical role in gas exchange, mitochondrial energy production, and antioxidant defense in cardiovascular system [18]. Alterations in energy-metabolism have also been detected in biopsies from patients with ICM [19].

Homeostasis is the process that living organisms use to regulate their internal equilibrium, which keeps the

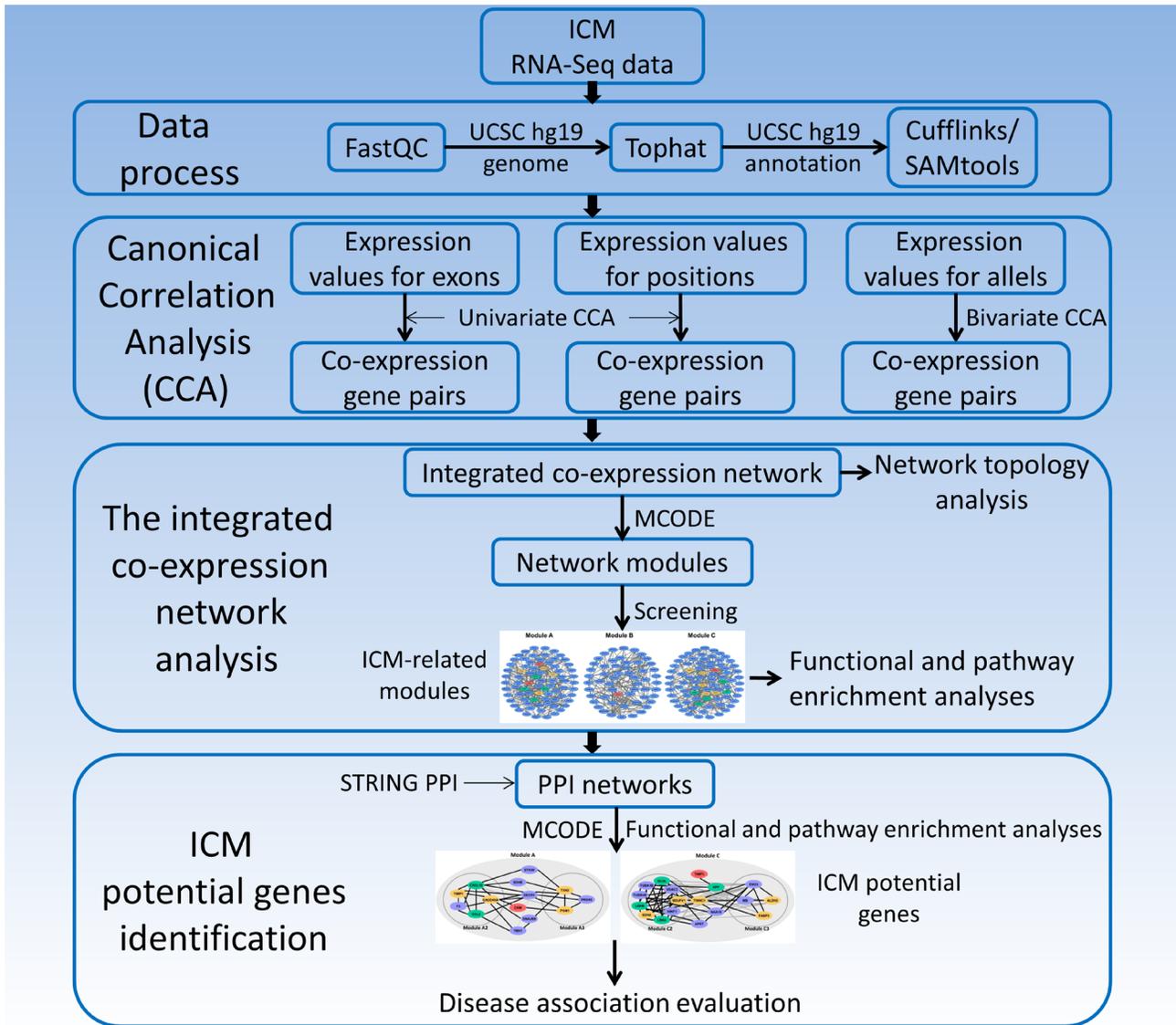


Figure 1: The identification of potential genes for human ischemic cardiomyopathy based on RNA-Seq data.

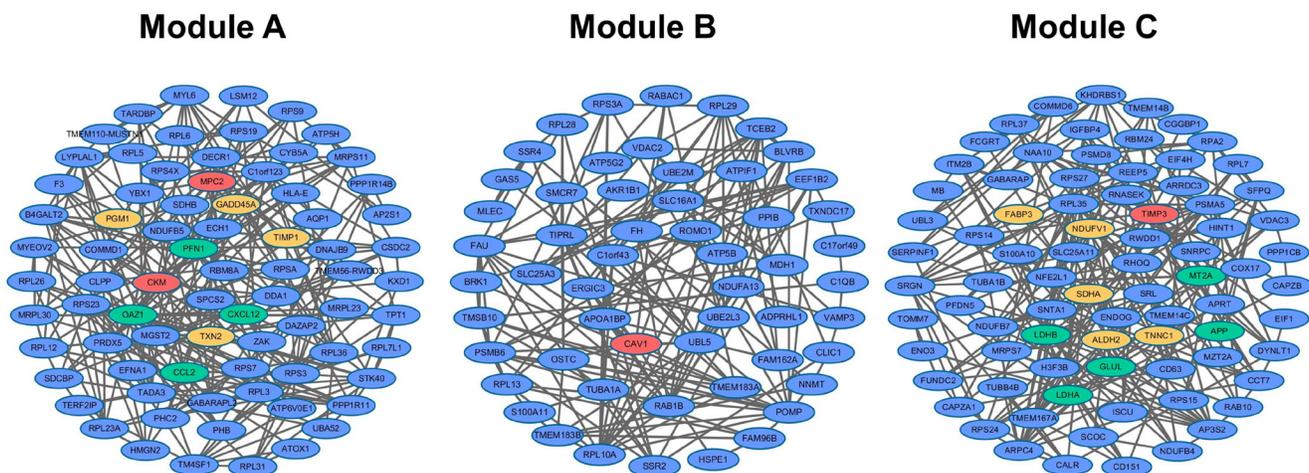


Figure 2: Three ICM-related modules (Module A, B and C). Red, yellow and green genes were verified to be related to ICM, DCM, and CHD, respectively.

health of human bodies. Most human diseases, including cardiovascular diseases, involve the disruption of normal homeostasis [20]. ICM could be caused by dysfunction of such biological processes, e.g. “homeostatic process” and “cellular homeostasis” [1, 21].

“Ribosome” is the site where all three modules were mainly localized. It is the place of “translation”, which is the process that proteins are synthesized. During “translation elongation”, amino acids are brought to the ribosome, and joined to form proteins in the order specified by mRNAs. These biological processes/pathways were greatly associated with DCM and CHD. Mice with mutant mTOR rapidly developed DCM with cardiomyocyte growth defects resulted from impaired protein translation efficiency [22]. The Food and Drug Administration approved drug mipomersen could treat

FH, one disease that could cause CHD, by inhibiting the apolipoprotein translation [23]. Androgen deficiency was shown to play a part in CHD and other cardiovascular diseases. These symptoms resulted from altered or damaged androgen synthesis, regulation or binding. In this process, genomic transcription and translation were also affected [24]. Mencarelli et al. found a different RNA secondary structure could change translation and protein synthesis, which was associated with CHD, type 2 diabetes and hypertension in the carriers [25].

These results exhibited the importance of ICM-related modules detected from the integrated co-expression network. These modules participated in many vital biological processes, which were associated with ICM.



Figure 3: GO functions and KEGG pathways enriched by genes in 4 modules. Node size is proportional to $-\log(P\text{-Value})$ of each enriched function/pathway.

ICM potential genes

Three PPI networks were built for three ICM-related modules, respectively. Since genes in these PPI networks were enriched in biological processes associated with ICM, these PPI networks were also ICM-related. Using MCODE, 3, 2 and 6 sub-modules were recognized from PPI networks of Module A, B and C, respectively. These sub-modules were named as Module A1-A3, B1-B2, and C1-C6. Module A1, B1-B2, C1 and C4-C6 were mainly localized to important subcellular organelles, such as ribosome or mitochondrion. It was worth noting that Module A2, A3, C2 and C3 were significantly enriched in translation, energy or homeostasis-related biological processes. These biological processes were verified to be strongly associated with ICM and its two major causes, DCM and CHD. As a result, 32 genes locating in or mediating Module A2, A3, C2 and C3 could act as ICM potential genes (Figure 4).

Since ICM was caused by DCM and CHD and no ICM disease genes were stored in any public databases, the disease associations of these genes were evaluated by the Online Mendelian Inheritance in Man (OMIM, Updated 9 July 2016) database [26] and recent literature for ICM, DCM or CHD.

For genes in Module A2, the research of Bironaite et al. showed that TIMP1 significantly increased in sera in DCM myocardium [27]. Results from DNA methylation profile between DCM patients and normal individuals showed that selenium deficiency increased the expression of the Gadd45 α , i.e. gene GADD45A [28]. CCL2 was CHD disease gene in OMIM [26]. Huang et al. and Franceschini et al. observed significant association between SNPs of CXCL12, rs1746048-C and rs501120, and an increased risk of CHD in Han Chinese and US cohorts, respectively [29, 30]. Many studies have shown that DCM was one of main clinical manifestations of PGM1 (in Module A3) deficiency [31, 32]. Western blot and histological analysis revealed that TXN2 (Module

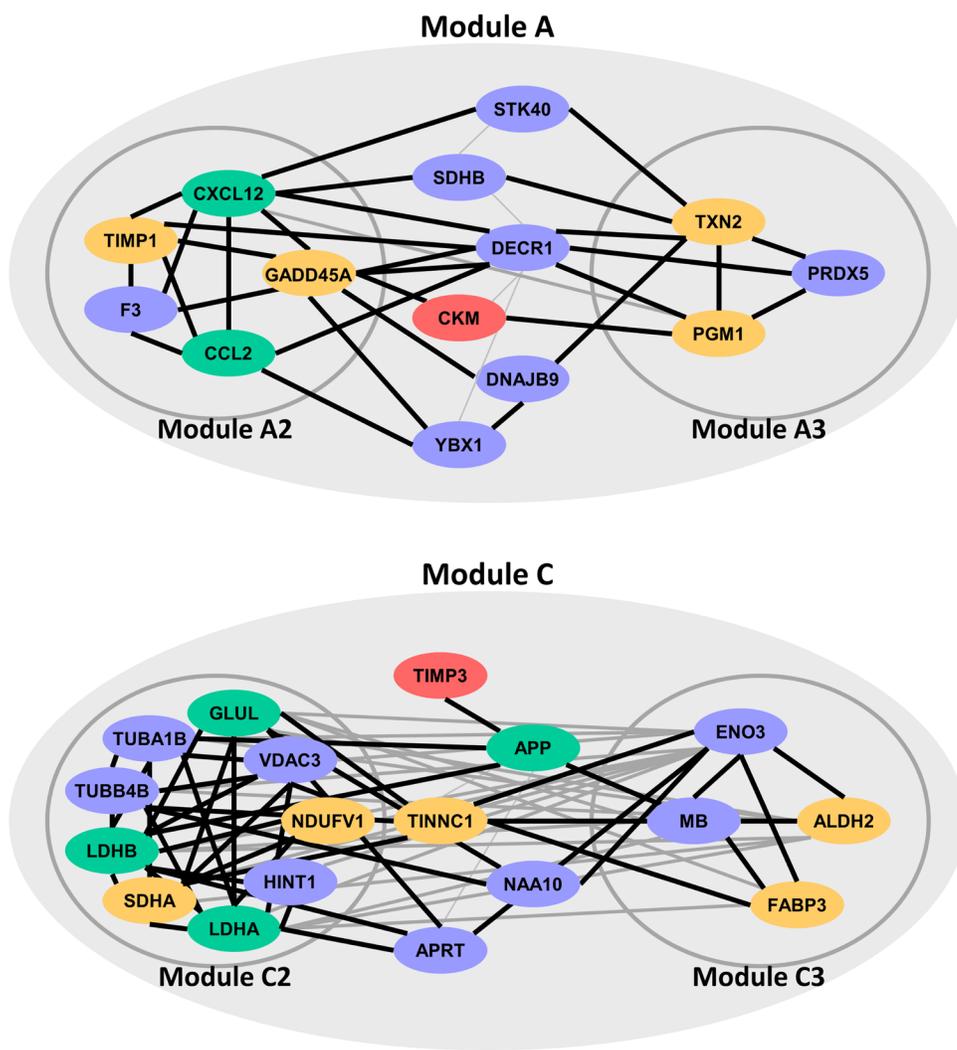


Figure 4: ICM potential genes locating in or mediating four sub-modules (Module A2, A3, C2 and C3). Red, yellow and green genes were verified to be related to ICM, DCM and CHD, respectively.

A3) protein expression levels were reduced in hearts from patients with DCM. Cardiac-specific TXN2 knockout mice developed DCM at 1 month of age with increased heart size, reduced ventricular wall thickness, and a progressive decline in left ventricular contractile function [33]. CKM in Module A that interacted with Module A2 and A3 was reported to be associated with ICM since protein levels of CKM were found to be reduced in ICM patients [34].

Gene SDHA of Module C2 was DCM disease gene stored in OMIM [26]. Ono et al. discovered that NDUFV1 of Module C2 was involved in the pathogenesis of DCM since NDUFV1 production decreased significantly in the myocardium of patients with DCM [35]. The rs10911021 SNP at the locus of gene GLUL of Module C2 has been associated with an increased risk of CHD in individuals with type 2 diabetes [36]. After 24 hours of treatment with lovastatin, which is widely used in prevention and treatment of CHD, the LDHA (in Module C2) mRNA levels went up. When the treatment time was extended to five days, the protein levels of LDHA were up-regulated, while LDHB (in Module C2) mRNA levels and protein levels were both down-regulated [37]. Sun et al. revealed that levels of ALDH2 in Module C3 were down-regulated in hearts from DCM patients [38]. Mean levels of FABP3 (Module C3) in DCM groups were significantly higher than in the control group in many researches. Thus, FABP3 was often used as a plasma biomarker in DCM, and played a significant role in its development [39, 40]. Moreover, TNNC1 of Module C interacting with both Module C2 and C3 were DCM disease genes stored in OMIM [26]. Though TIMP3 of Module C did not interact with Module C2 and C3 directly, it interacted with gene APP that interacted with both Module C2 and C3. APP has been indicated to be involved in CHD by several studies [41, 42]. TIMP3 has also been shown to be significantly down-regulated in patients with ICM [43].

After database and literature evaluation, 17 ICM potential genes were verified to be involved in ICM, DCM and CHD by OMIM and literature. For other ICM potential genes, though no disease association record was found in OMIM or literature, they also participated in ICM associated biological pathways. For example, SDHB in Module A and ENO3 in Module C3 were annotated to the process of “generation of precursor metabolites and energy”. “Cellular homeostasis” was annotated by gene PRDX5 of Module A3.

DISCUSSION

RNA-Seq data from the next-generation sequencing technology could offer high-quality expression data of various levels for transcriptome analysis. Plenty of transcriptome information could be provided by expressions of exons, positions and alleles [44], which should be considered comprehensively. Hence, in this paper, univariate and bivariate CCA was employed to

identify co-expression gene pairs for the exon, position and allelic data from RNA-Seq data, respectively. The integrated co-expression network was constructed by aggregating three sets of co-expression gene pairs. This network was biologically significant and could be used as the context for exploring potential genes and the pathogenesis of diseases.

Additionally, three ICM-related modules were screened from the integrated co-expression network after comparing between Pearson correlation coefficients of ICM samples and normal controls. These modules could reflect alterations from normal to disease states. Therefore, they were significantly enriched in disease associated biological processes, such as translation, energy generation and homeostasis. This indicated that some important genes in these modules might take vital roles in the process of ICM.

To consider real interactions between genes in each ICM-related module, we built one PPI network for each ICM-related module. Genes in these PPI networks were enriched in many vital biological processes, such as “translation elongation”, “translation” and “generation of precursor metabolites and energy”, and the pathway of “Ribosome”, which were associated with ICM. Besides these common biological processes, these PPI networks could be enriched in different biological processes. These different biological processes were also ICM-related processes and could represent different aspects for ICM mechanism. Sub-modules from these PPI networks could provide better understanding of functions and cooperation for genes in ICM. 32 ICM potential genes locating in or mediating sub-modules were identified, which were significantly enriched in translation, energy or homeostasis-related biological processes. Of these ICM potential genes, 2 were verified to be involved in ICM, 9 in DCM, 6 in CHD by OMIM and literature, and others participated in ICM associated biological pathways.

In our method, relationships between genes were detected based on expression values of different levels, i.e. exon, position and allelic levels, from RNA-Seq data of ICM samples. Recently, some sequence analysis tools have been proposed, such as Pse-in-One, repRNA and repDNA [45-47]. The three tools could generate various vectors of important features once sequences of DNAs, RNAs or proteins were given. We speculate that co-regulation/function relationships between genes with sequence variations could be obtained from various vectors of important features generated by the three tools. With relationships from the three tools incorporated into our model, we hope that a much more powerful method considering sequence patterns could be developed in the future.

MATERIALS AND METHODS

Data

RNA-Seq data of human ICM samples, GSE48166, were obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) [48], which contained 15 ICM and 15 normal samples. The reference human genome data were the hg19 file from the UCSC database (<http://genome.ucsc.edu/>) [49]. The human genome annotation data were also from the UCSC database (<http://genome.ucsc.edu/cgi-bin/hgTables?command=start>).

The expression values of exon, position and allelic levels were obtained after processing RNA-Seq data of 15 ICM samples using FastQC [50], Tophat [51], Cufflinks [52] and SAMtools [53]. Expression values of the exon or position level were read counts of exons or positions, and expression values of the allelic level, allele specific expressions (ASE), were read counts of allele pairs of single nucleotide polymorphisms (SNPs). The expression values of 5105 exons, 5280 positions and 5333 allele pairs from 15 ICM samples were obtained after data processing.

Identification of co-expression gene pairs

In this paper, univariate and bivariate CCA were employed to identify co-expression gene pairs, which could detect the maximum correlation between two sets of variables [6].

Univariate canonical correlation analysis

The univariate CCA was performed for the expression data of the exon or position level of two genes, g_1 and g_2 . g_1 had p exons or positions, and g_2 had q exons or positions, $p \leq q$. $X_e^{(1)}$ and $X_e^{(2)}$ were vectors of expression values for the e th exon or the e th position of g_1 and g_2 for ICM samples. Thus, the expression data of the exon or position level of these two genes could be represented as $X^{(1)} = [X_1^{(1)}, \dots, X_p^{(1)}]^T$ and $X^{(2)} = [X_1^{(2)}, \dots, X_q^{(2)}]^T$. Linear combinations of exon or position expression values from two genes, U and V , were represented as $U = a^T X^{(1)}$ and $V = b^T X^{(2)}$, respectively. The maximum correlations between different pairs of U s and V s, $\lambda_1, \lambda_2, \dots, \lambda_p$, $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$ with the significance $p_i = (i = 1, \dots, p)$ were calculated [13].

Since correlations with $p_i \leq 0.05$ was significant, the final correlation, $w_{exon/position}$ between two genes was defined as

$$w_{exon/position} = \frac{\sum_{i=1}^p \lambda_i I(\log P_i)}{\sum_{i=1}^p I(\log P_i)}$$

$$\text{where } I(\log P_i) = \begin{cases} 0 & P_i > 0.05 \\ -\log P_i & P_i \leq 0.05 \end{cases} \cdot w_{exon/position} = 0 \text{ if } \sum_{i=1}^p I(\log P_i) = 0.$$

Gene pairs with non-zero $w_{exon/position}$ were then used to construct the integrated co-expression network [54].

Bivariate canonical correlation analysis

The bivariate CCA was performed for the ASE data of SNPs from two genes, one with s SNPs, the other with t SNPs, $s \leq t$. $Y_i^{(1)}$ and $Y_i^{(2)}$ were vectors of the expression values for two alleles of the i th SNP of one gene, and $Z_1^{(1)}$ and $Z_1^{(2)}$ were vectors of the expression values for two alleles of the i th SNP of one other gene for ICM samples. Thus, the ASE data of these two genes could be represented as $Y = [Y_1^{(1)}, \dots, Y_s^{(1)}, Y_1^{(2)}, \dots, Y_t^{(2)}]^T = [Y^{(1)}, Y^{(2)}]^T$ and $Z = [Z_1^{(1)}, \dots, Z_s^{(1)}, Z_1^{(2)}, \dots, Z_t^{(2)}]^T = [Z^{(1)}, Z^{(2)}]^T$. Linear combinations of ASE from two genes, M and N , were represented as $M = c^T Y [c^{(1)}]^T, Y^{(1)} + [c^{(2)}]^T Y^{(2)}$ and $N = d^T Z [d^{(1)}]^T, Z^{(1)} + [d^{(2)}]^T Z^{(2)}$, respectively. The maximum correlations between different pairs of M s and N s, $\gamma_1, \gamma_2, \dots, \gamma_p$, $\gamma_1^2 \geq \gamma_2^2 \geq \dots \geq \gamma_p^2$ with the significance p_i were calculated [13].

Since correlations with $p_i \leq 0.05$ was significant, the final correlation between two genes w_{allele} was defined as

$$w_{allele} = \frac{\sum_{j=1}^{2s} \gamma_j I(\log P_j)}{\sum_{j=1}^{2s} I(\log P_j)}$$

$$\text{where } I(\log P_j) = \begin{cases} 0 & P_j > 0.05 \\ -\log P_j & P_j \leq 0.05 \end{cases} \cdot w_{allele} = 0 \text{ if } \sum_{j=1}^{2s} I(\log P_j) = 0.$$

Gene pairs with non-zero w_{allele} were then used to construct the integrated co-expression network [54].

Construction and analysis of the integrated co-expression network

Co-expression networks are undirected graphs, where nodes correspond to genes, and edges between genes represent co-expression relationships, i.e. correlations between co-expression gene pairs. The integrated co-expression network was constructed by integrating all co-expression gene pairs with non-zero correlations ($w_{exon/position}$ or w_{allele}) for the expression data of the exon, position and allelic levels obtained from RNA-Seq data of human ICM samples.

Topological analysis of the integrated co-expression network

The topological property of the integrated co-expression network was evaluated by assortativity. Assortativity is an important measure for network topology, which describes the tendency of a node connecting to similar nodes in a network. Assortativity coefficient was used to measure network assortativity by the Pearson correlation coefficient between degrees of connecting node pairs [55]. If the assortativity coefficient >0 , the network was assortative; if the assortativity coefficient <0 , the network was disassortative. To compare assortativity of different types of networks, a biological network and a random network were constructed. The nodes and edges of the biological network were from KEGG pathways, and those for the random network were selected randomly from KEGG pathways.

Detection and function of ICM-related modules

Modules of co-expression network constructed from disease samples could be linked to a particular disease phenotype and help to uncover disease mechanisms [56]. MCODE was employed to detect network modules from the integrated co-expression network [57]. To further screen ICM-related modules, differences between Pearson correlation coefficients of expression values for ICM samples and those for normal controls from GSE48166 were calculated for each module. Then, the difference was compared with those of 1000 random modules, which were constructed by selecting the same number of genes as modules from MCODE. If the real difference was significantly greater than the random ones (permutation test, $p < 0.05$), the module was considered as ICM-related.

Functional and pathway enrichment analyses were performed for these screened ICM-related modules using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>). To understand the significance of genes in the process of ICM, BP_Fat (biological process) of Gene Ontology (GO) [58] and pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG) [59] with P -Value <0.05 were selected.

Identification of ICM potential genes

It was necessary to consider real interactions between genes in ICM-related modules. PPIs of genes in each module were obtained from the STRING database (v10, <http://string-db.org/>) [60]. Three PPI networks were built for these modules, respectively. MCODE was employed to recognize sub-modules from these PPI networks, respectively. Functional and pathway enrichment analyses were also performed

as aforementioned. Genes in sub-modules that were significantly enriched in biological processes associated with ICM (P -Value <0.05) could act as ICM potential genes. In addition, genes mediating this kind of sub-modules, i.e. interacting with sub-modules, could also be ICM potential genes.

CONCLUSIONS

Though no ICM disease genes were stored in public databases, ICM-related modules screened from the integrated co-expression network constructed for ICM RNA-Seq data could provide more genomic and molecular information for biological processes and disease mechanisms. Taking PPIs into consideration, 32 genes locating in or mediating sub-modules were identified as ICM potential genes. 17 genes were verified to be involved in ICM, DCM and CHD by OMIM and literature. Our method will become an effective and powerful tool for identifying potential genes and elucidating the pathogenesis of complex diseases and their subtypes.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

GRANT SUPPORT

This work was supported in part by the National Natural Science Foundation of China (Grant No. 31301040 and 61272388); the Health and Family Planning Commission Scientific Research Subject of Heilongjiang Province (Grant No. 2016-203); the Master Innovation Funds of Heilongjiang Province (Grant No. YJSCX2015-40HYD); the University Student Innovation and Entrepreneurship Training Program in Heilongjiang Province (Grant No. 201610226066 and 201610226012); and the Harbin Applied Technology Research and Development Project (Grant No. 2016RQXXJ105).

REFERENCES

1. Rosello-Lleti E, Carnicer R, Tarazon E, Ortega A, Gil-Cayuela C, Lago F, Gonzalez-Juanatey JR, Portoles M, Rivera M. Human Ischemic Cardiomyopathy Shows Cardiac Nos1 Translocation and its Increased Levels are Related to Left Ventricular Performance. *Sci Rep.* 2016; 6: 24060. doi: 10.1038/srep24060.
2. Herrer I, Rosello-Lleti E, Ortega A, Tarazon E, Molina-Navarro MM, Trivino JC, Martinez-Dolz L, Almenar L, Lago F, Sanchez-Lazaro I, et al. Gene expression network analysis reveals new transcriptional regulators as novel factors in human ischemic cardiomyopathy. *BMC Med Genomics.* 2015; 8: 14. doi: 10.1186/s12920-015-0088-y.
3. Herrer I, Rosello-Lleti E, Rivera M, Molina-Navarro MM,

- Tarazon E, Ortega A, Martinez-Dolz L, Trivino JC, Lago F, Gonzalez-Juanatey JR, Bertomeu V, Montero JA, Portoles M. RNA-sequencing analysis reveals new alterations in cardiomyocyte cytoskeletal genes in patients with heart failure. *Lab Invest*. 2014; 94: 645-53. doi: 10.1038/labinvest.2014.54.
4. Lee JH, Gao C, Peng G, Greer C, Ren S, Wang Y, Xiao X. Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts. *Circ Res*. 2011; 109: 1332-41. doi: 10.1161/CIRCRESAHA.111.249433.
 5. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res*. 2011; 21: 1728-37. doi: 10.1101/gr.119784.110.
 6. Hong S, Chen X, Jin L, Xiong M. Canonical correlation analysis for RNA-seq co-expression networks. *Nucleic Acids Res*. 2013; 41: e95. doi: 10.1093/nar/gkt145.
 7. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009; 10: 57-63. doi: 10.1038/nrg2484.
 8. Kugler KG, Mueller LA, Graber A, Dehmer M. Integrative network biology: graph prototyping for co-expression cancer networks. *PLoS One*. 2011; 6: e22843. doi: 10.1371/journal.pone.0022843.
 9. Emmert-Streib F, Dehmer M. Networks for systems biology: conceptual connection of data and function. *IET Syst Biol*. 2011; 5: 185-207. doi: 10.1049/iet-syb.2010.0025.
 10. Auffray C, Hood L. Editorial: Systems biology and personalized medicine - the future is now. *Biotechnol J*. 2012; 7: 938-9. doi: 10.1002/biot.201200242.
 11. Uyar B, Chu JS, Vergara IA, Chua SY, Jones MR, Wong T, Baillie DL, Chen N. RNA-seq analysis of the *C. briggsae* transcriptome. *Genome Res*. 2012; 22: 1567-80. doi: 10.1101/gr.134601.111.
 12. Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol*. 2012; 10: 618-30. doi: 10.1038/nrmicro2852.
 13. Tang CS, Ferreira MA. A gene-based test of association using canonical correlation analysis. *Bioinformatics*. 2012; 28: 845-50. doi: 10.1093/bioinformatics/bts051.
 14. Roy S, Bhattacharyya DK, Kalita JK. Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinformatics*. 2014; 15 Suppl 7: S10. doi: 10.1186/1471-2105-15-S7-S10.
 15. Pena-Castillo L, Mercer RG, Gurinovich A, Callister SJ, Wright AT, Westbye AB, Beatty JT, Lang AS. Gene co-expression network analysis in *Rhodobacter capsulatus* and application to comparative expression analysis of *Rhodobacter sphaeroides*. *BMC Genomics*. 2014; 15: 730. doi: 10.1186/1471-2164-15-730.
 16. Zamanian-Azodi M, Rezaei-Tavirani M, Rahmati-Rad S, Hasanzadeh H, Rezaei Tavirani M, Seyyedi SS. Protein-Protein Interaction Network could reveal the relationship between the breast and colon cancer. *Gastroenterol Hepatol Bed Bench*. 2015; 8: 215-24. doi:
 17. Iozzo P. Metabolic toxicity of the heart: insights from molecular imaging. *Nutr Metab Cardiovasc Dis*. 2010; 20: 147-56. doi: 10.1016/j.numecd.2009.08.011.
 18. Bayeva M, Khechaduri A, Wu R, Burke MA, Wasserstrom JA, Singh N, Liesa M, Shirihai OS, Langer NB, Paw BH, Ardehali H. ATP-binding cassette B10 regulates early steps of heme synthesis. *Circ Res*. 2013; 113: 279-87. doi: 10.1161/CIRCRESAHA.113.301552.
 19. Haas S, Jahnke HG, Moerbt N, von Bergen M, Aharinejad S, Andrukhova O, Robitzki AA. DIGE proteome analysis reveals suitability of ischemic cardiac in vitro model for studying cellular response to acute ischemia and regeneration. *PLoS One*. 2012; 7: e31669. doi: 10.1371/journal.pone.0031669.
 20. Kotas ME, Medzhitov R. Homeostasis, inflammation, and disease susceptibility. *Cell*. 2015; 160: 816-27. doi: 10.1016/j.cell.2015.02.010.
 21. Heidrich F, Sossalla S, Schotola H, Vorkamp T, Ortmann P, Popov AF, Coskun KO, Rajab TK, Friedrich M, Sohns C, Hinz J, Bauer M, Quintel M, et al. The role of phospho-adenosine monophosphate-activated protein kinase and vascular endothelial growth factor in a model of chronic heart failure. *Artif Organs*. 2010; 34: 969-79. doi: 10.1111/j.1525-1594.2010.01121.x.
 22. Mazelin L, Panthu B, Nicot AS, Belotti E, Tintignac L, Teixeira G, Zhang Q, Risson V, Baas D, Delaune E, Derumeaux G, Taillandier D, Ohlmann T, et al. mTOR inactivation in myocardium from infant mice rapidly leads to dilated cardiomyopathy due to translation defects and p53/JNK-mediated apoptosis. *J Mol Cell Cardiol*. 2016; 97: 213-25. doi: 10.1016/j.yjmcc.2016.04.011.
 23. Marbach JA, McKeon JL, Ross JL, Duffy D. Novel treatments for familial hypercholesterolemia: pharmacogenetics at work. *Pharmacotherapy*. 2014; 34: 961-72. doi: 10.1002/phar.1441.
 24. Carruthers M. Testosterone deficiency syndrome: cellular and molecular mechanism of action. *Curr Aging Sci*. 2013; 6: 115-24. doi:
 25. Mencarelli M, Zulian A, Canello R, Alberti L, Gilardini L, Di Blasio AM, Invitti C. A novel missense mutation in the signal peptide of the human POMC gene: a possible additional link between early-onset type 2 diabetes and obesity. *Eur J Hum Genet*. 2012; 20: 1290-4. doi: 10.1038/ejhg.2012.103.
 26. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 2015; 43: D789-98. doi: 10.1093/nar/gku1205.
 27. Bironaite D, Daunoravicius D, Bogomolovas J, Cibiras S,

- Vitkus D, Zurauskas E, Zasytyte I, Rucinskas K, Labeit S, Venalis A, Grabauskiene V. Molecular mechanisms behind progressing chronic inflammatory dilated cardiomyopathy. *BMC Cardiovasc Disord.* 2015; 15: 26. doi: 10.1186/s12872-015-0017-1.
28. Yang G, Zhu Y, Dong X, Duan Z, Niu X, Wei J. TLR2-ICAM1-Gadd45alpha axis mediates the epigenetic effect of selenium on DNA methylation and gene expression in Keshan disease. *Biol Trace Elem Res.* 2014; 159: 69-80. doi: 10.1007/s12011-014-9985-8.
 29. Huang Y, Zhou J, Ye H, Xu L, Le Y, Yang X, Xu W, Huang X, Lian J, Duan S. Relationship between chemokine (C-X-C motif) ligand 12 gene variant (rs1746048) and coronary heart disease: case-control study and meta-analysis. *Gene.* 2013; 521: 38-44. doi: 10.1016/j.gene.2013.02.047.
 30. Franceschini N, Carty C, Buzkova P, Reiner AP, Garrett T, Lin Y, Vockler JS, Hindorff LA, Cole SA, Boerwinkle E, Lin DY, Bookman E, Best LG, et al. Association of genetic variants and incident coronary heart disease in multiethnic cohorts: the PAGE study. *Circ Cardiovasc Genet.* 2011; 4: 661-72. doi: 10.1161/CIRCGENETICS.111.960096.
 31. Loewenthal N, Haim A, Parvari R, Hershkovitz E. Phosphoglucomutase-1 deficiency: Intrafamilial clinical variability and common secondary adrenal insufficiency. *Am J Med Genet A.* 2015; 167A: 3139-43. doi: 10.1002/ajmg.a.37294.
 32. Lee Y, Stiers KM, Kain BN, Beamer LJ. Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency. *J Biol Chem.* 2014; 289: 32010-9. doi: 10.1074/jbc.M114.597914.
 33. Huang Q, Zhou HJ, Zhang H, Huang Y, Hinojosa-Kirschenbaum F, Fan P, Yao L, Belardinelli L, Tellides G, Giordano FJ, Budas GR, Min W. Thioredoxin-2 inhibits mitochondrial reactive oxygen species generation and apoptosis stress kinase-1 activity to maintain cardiac function. *Circulation.* 2015; 131: 1082-97. doi: 10.1161/CIRCULATIONAHA.114.012725.
 34. Teixeira PC, Santos RH, Fiorelli AI, Bilate AM, Benvenuti LA, Stolf NA, Kalil J, Cunha-Neto E. Selective decrease of components of the creatine kinase system and ATP synthase complex in chronic Chagas disease cardiomyopathy. *PLoS Negl Trop Dis.* 2011; 5: e1205. doi: 10.1371/journal.pntd.0001205.
 35. Ono H, Nakamura H, Matsuzaki M. A NADH dehydrogenase ubiquinone flavoprotein is decreased in patients with dilated cardiomyopathy. *Intern Med.* 2010; 49: 2039-42. doi:
 36. Prudente S, Shah H, Bailetti D, Pezzolesi M, Buranasupkajorn P, Mercuri L, Mendonca C, De Cosmo S, Niewczas M, Trischitta V, Doria A. Genetic Variant at the GLUL Locus Predicts All-Cause Mortality in Patients With Type 2 Diabetes. *Diabetes.* 2015; 64: 2658-63. doi: 10.2337/db14-1653.
 37. Guo WZ, Ji H, Yan ZH, Li L, Li D, Lu CL. Lovastatin changes activities of lactate dehydrogenase A and B genes in rat myocardial cells. *Chin Med J (Engl).* 2011; 124: 423-8. doi:
 38. Sun A, Cheng Y, Zhang Y, Zhang Q, Wang S, Tian S, Zou Y, Hu K, Ren J, Ge J. Aldehyde dehydrogenase 2 ameliorates doxorubicin-induced myocardial dysfunction through detoxification of 4-HNE and suppression of autophagy. *J Mol Cell Cardiol.* 2014; 71: 92-104. doi: 10.1016/j.yjmcc.2014.01.002.
 39. Wang X, Zhou L, Jin J, Yang Y, Song G, Shen Y, Liu H, Liu M, Shi C, Qian L. Knockdown of FABP3 impairs cardiac development in Zebra fish through the retinoic acid signaling pathway. *Int J Mol Sci.* 2013; 14: 13826-41. doi: 10.3390/ijms140713826.
 40. Ozbek M, Erdogan M, Dogan M, Akbal E, Ozturk MA, Ureten K. Serum heart-type fatty acid binding protein levels in acromegaly patients. *J Endocrinol Invest.* 2011; 34: 576-9. doi: 10.3275/7259.
 41. Cojocaru M, Cojocaru IM, Silosi I. Lipoprotein-associated phospholipase A2 as a predictive biomarker of sub-clinical inflammation in cardiovascular diseases. *Maedica (Buchar).* 2010; 5: 51-5. doi:
 42. Li F, Gong Q, Dong H, Shi J. Resveratrol, a neuroprotective supplement for Alzheimer's disease. *Curr Pharm Des.* 2012; 18: 27-33. doi:
 43. Kandalam V, Basu R, Abraham T, Wang X, Awad A, Wang W, Lopaschuk GD, Maeda N, Oudit GY, Kassiri Z. Early activation of matrix metalloproteinases underlies the exacerbated systolic and diastolic dysfunction in mice lacking TIMP3 following myocardial infarction. *Am J Physiol Heart Circ Physiol.* 2010; 299: H1012-23. doi: 10.1152/ajpheart.00246.2010.
 44. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, Harmanci A, Leng J, Bjornson R, Kong Y, Kitabayashi N, Bhardwaj N, Rubin M, Snyder M, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol.* 2011; 7: 522. doi: 10.1038/msb.2011.54.
 45. Liu B, Liu F, Wang X, Chen J, Fang L, Chou KC. Pse-in-One: a web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Res.* 2015; 43: W65-71. doi: 10.1093/nar/gkv458.
 46. Liu B, Liu F, Fang L, Wang X, Chou KC. repRNA: a web server for generating various feature vectors of RNA sequences. *Mol Genet Genomics.* 2016; 291: 473-81. doi: 10.1007/s00438-015-1078-7.
 47. Liu B, Liu F, Fang L, Wang X, Chou KC. repDNA: a Python package to generate various modes of feature vectors for DNA sequences by incorporating user-defined physicochemical properties and sequence-order effects. *Bioinformatics.* 2015; 31: 1307-9. doi: 10.1093/bioinformatics/btu820.
 48. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman

- PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013; 41: D991-5. doi: 10.1093/nar/gks1193.
49. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 2015; 43: D670-81. doi: 10.1093/nar/gku1177.
 50. Ramirez-Gonzalez RH, Leggett RM, Waite D, Thanki A, Drou N, Caccamo M, Davey R. StatsDB: platform-agnostic storage and understanding of next generation sequencing run metrics. *F1000Res.* 2013; 2: 248. doi: 10.12688/f1000research.2-248.v2.
 51. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25: 1105-11. doi: 10.1093/bioinformatics/btp120.
 52. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28: 511-5. doi: 10.1038/nbt.1621.
 53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25: 2078-9. doi: 10.1093/bioinformatics/btp352.
 54. Xulvi-Brunet R, Li H. Co-expression networks: graph properties and topological comparisons. *Bioinformatics.* 2010; 26: 205-14. doi: 10.1093/bioinformatics/btp632.
 55. Newman ME. Assortative mixing in networks. *Phys Rev Lett.* 2002; 89: 208701. doi: 10.1103/PhysRevLett.89.208701.
 56. Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, Sahni N, Thibault D, Voung L, Guo F, Ghiassian SD, Gulbahce N, Baribaud F, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum Mol Genet.* 2015; 24: 3005-20. doi: 10.1093/hmg/ddv001.
 57. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003; 4: 2. doi: 10.1186/1471-2108-4-2.
 58. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25: 25-9. doi: 10.1038/75556.
 59. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28: 27-30. doi: 10.1093/nar/28.1/27.
 60. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015; 43: D447-52. doi: 10.1093/nar/gku1003.