


RESEARCH ARTICLE

Open Access



Weighted correlation network bioinformatics uncovers a key molecular biosignature driving the left-sided heart failure

Jiamin Zhou^{1,2}, Wei Zhang³, Chunying Wei^{1,2}, Zhiliang Zhang^{1,2}, Dasong Yi^{1,2}, Xiaoping Peng^{1,2}, Jingtian Peng^{1,2}, Ran Yin^{1,2}, Zeqi Zheng^{1,2}, Hongmei Qi^{1,2}, Yunfeng Wei^{1,2} and Tong Wen^{1,2*} 

Abstract

Background: Left-sided heart failure (HF) is documented as a key prognostic factor in HF. However, the relative molecular mechanisms underlying left-sided HF is unknown. The purpose of this study is to unearth significant modules, pivotal genes and candidate regulatory components governing the progression of left-sided HF by bioinformatical analysis.

Methods: A total of 319 samples in GSE57345 dataset were used for weighted gene correlation network analysis (WGCNA). ClusterProfiler package in R was used to conduct functional enrichment for genes uncovered from the modules of interest. Regulatory networks of genes were built using Cytoscape while Enrichr database was used for identification of transcription factors (TFs). The MCODE plugin was used for identifying hub genes in the modules of interest and their validation was performed based on GSE1869 dataset.

Results: A total of six significant modules were identified. Notably, the blue module was confirmed as the most crucially associated with left-sided HF, ischemic heart disease (ISCH) and dilated cardiomyopathy (CMP). Functional enrichment conveyed that genes belonging to this module were mainly those driving the extracellular matrix-associated processes such as extracellular matrix structural constituent and collagen binding. A total of seven transcriptional factors, including Suppressor of Zeste 12 Protein Homolog (SUZ12) and nuclear factor erythroid 2 like 2 (NFE2L2), adrenergic receptor (AR), were identified as possible regulators of coexpression genes identified in the blue module. A total of three key genes (OGN, HTRA1 and MXRA5) were retained after validation of their prognostic value in left-sided HF. The results of functional enrichment confirmed that these key genes were primarily involved in response to transforming growth factor beta and extracellular matrix.

(Continued on next page)

* Correspondence: went19840201@163.com

¹Department of Cardiology, The First Affiliated Hospital of Nanchang University, No. 17 Yongwaizheng Street, Nanchang 330006, Jiangxi province, China

²Hypertension Research Institute of Jiangxi Province, Nanchang 330006, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusion: We uncovered a candidate gene signature correlated with HF, ISCH and CMP in the left ventricle, which may help provide better prognosis and therapeutic decisions and in HF, ISCH and CMP patients.

Keywords: Weighted gene co-expression network analysis, Ischemic heart disease, Dilated cardiomyopathy, Heart failure

Background

Cardiac arrest, the inability of the heart to perform its pumping function, is a major cause of death and a public health problem [1, 2]. The incidence of cardiac arrest is growing worldwide, especially in the vast majority of developed countries [3]. Left-sided heart failure (HF), also known as left ventricular failure, is the common element associated with heart disorders leading to eventual final heart failure [4]. Greater knowledge of the mechanisms involved in the physio-pathogenesis of the left ventricular failure could allow early identification of patients at risk and timely management, which could reduce the socio-economic damages associated with cardiac arrest.

Advances in molecular biology and especially the advent of latest generation sequencing platforms have allowed the accumulation of a large amount of data on the expression of genes regulating the initiation and development of various diseases [5–8]. Correct analysis of these data could help us to uncover the underlying biological functions of genes in different diseases [9, 10]. Regarding cardiac arrest, a number of biomarkers have been discovered in previous studies [11–13]. For example, myoglobin [12], creatine kinase MB isoenzyme (CK-MB) [14], and troponins [15] have been used as biomarkers to assess myocardial pain and diagnose postoperative myocardial infarction. Through transcriptomic analysis, thousands of genes, screened by differential gene expression analysis, have also been suggested as biological markers for cardiac arrest [16]. However, our knowledge on biomarkers especially associated with left ventricular failure is limited. Previous studies based on differential expression analysis have allowed the discovery of key genes between different presentations leading to HF [17, 18], but their application requires experimental validation. However, before any experimental validation, it is essential to do a preliminary work of accurate selection of all the genes very potentially associated with cardiac arrest. This is possible thanks to bioinformatics approaches which are available today.

The WGCNA approach is a bioinformatics technique that allows the extraction and grouping into modules of a list of genes involved in a given biological process [19]. This technique has been used for the credible discovery of a number of genes associated with various diseases and their sub-characteristics [20–23]. WGCNA allows the correlative identification of genes with a similar expression profile to a given characteristic trait. WGCNA has been used to screen genes for different processes in

cardiovascular disease such as coronary artery disease [24], congestive HF (CHF) and valvular heart disease (VHD). The use of WGCNA for the discovery of biomarkers related to cardiac arrest after acute myocardial infarction (AMI) has also been reported in a previous study [25] which identified six key genes with a great prognostic value for the progression of HF post-AMI. However, more studies are needed to dissect and decipher the genes involved in left ventricular failure.

Thus, in the present study, we conducted a WGCNA data mining in order to uncover key genes potentially involved in the pathogenesis and the progression of left ventricular failure. Our ultimate goal is to make available a list of biomarkers that could guide the identification of patients at risk of left ventricular failure and the design of appropriate management strategies.

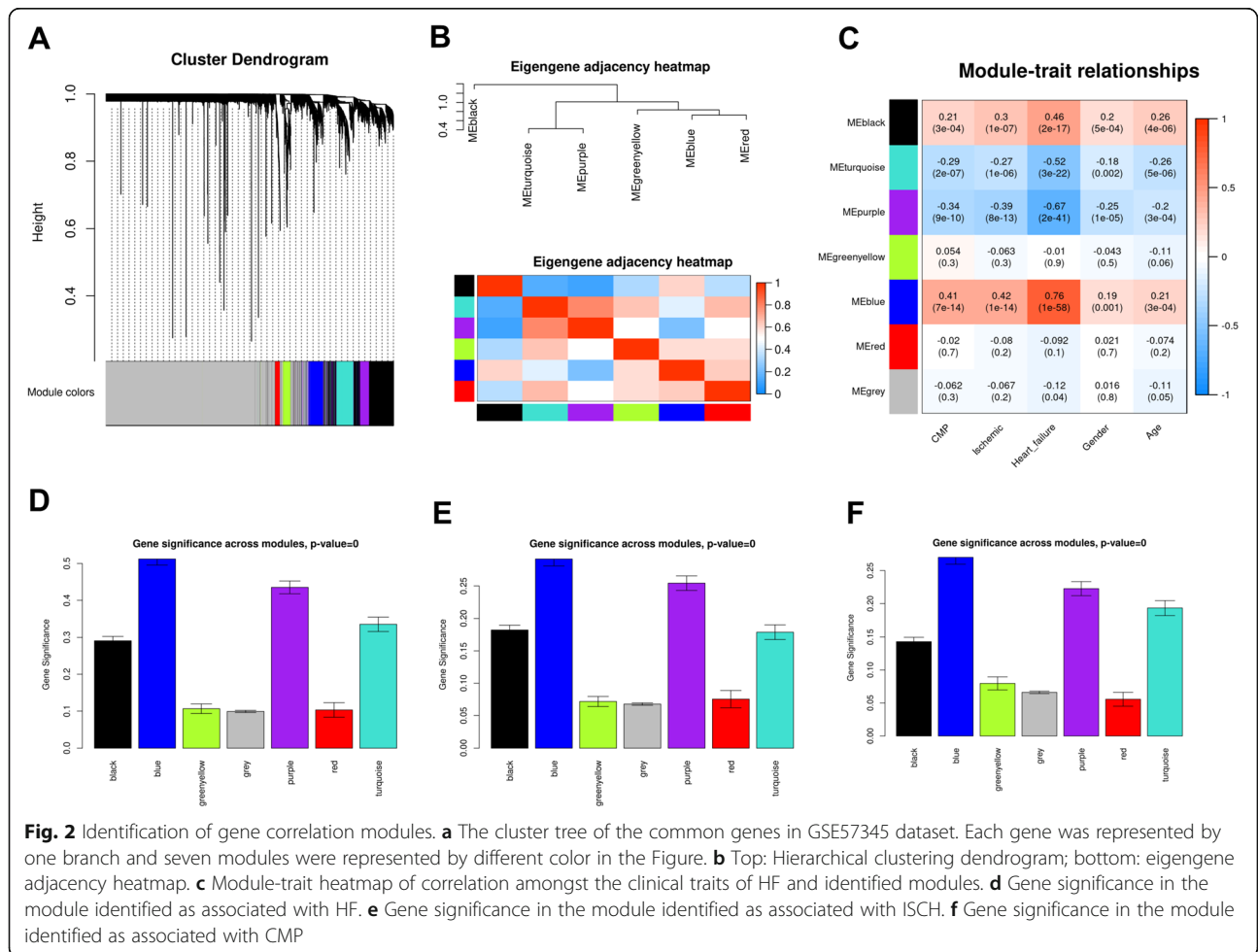
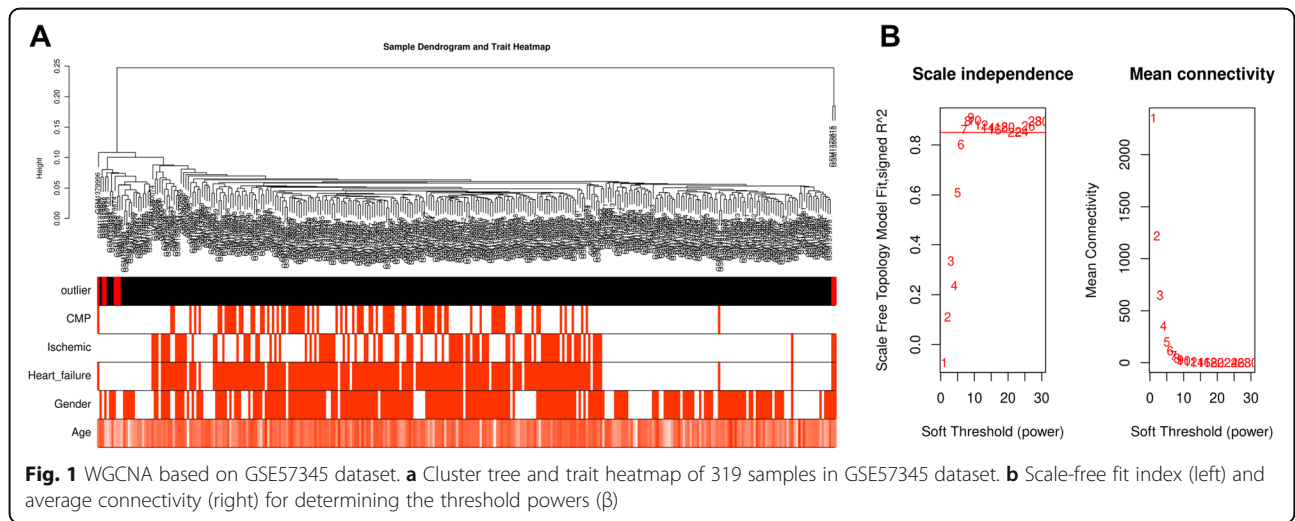
Methods

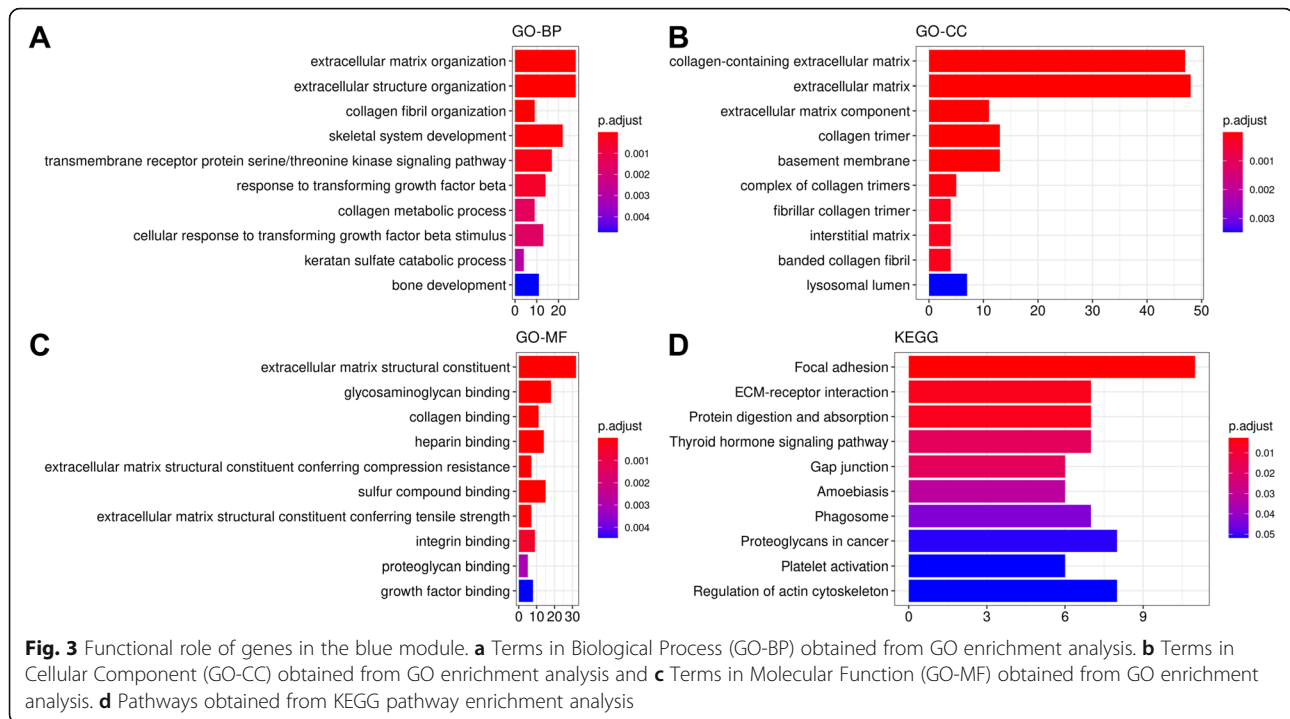
Data sources and data preprocessing

All of the expression datasets used in the present study were obtained from the Gene Expression Omnibus (GEO) datasets (<https://www.ncbi.nlm.nih.gov/gds>). The data for WGCNA construction was the GSE57345 dataset containing 319 samples of 96 patients with ischemic heart disease (ISCH) and 84 patients with dilated cardiomyopathy (CMP) [26]. The platform used for the acquisition of this data was GPL11532 platform. The validation data was the GSE1869 dataset based on the GPL96 platform and containing samples from six non-HF patients and patients with HF after AMI [27]. The R library “affy” was employed for expression data preprocessing with the Robust Multichip Average (RMA) in the R 3.3.1 software. Following the correction of background effect, quantile normalization and log₂-transformation, the datasets were used for subsequent analyses [28].

Construction of coexpression modules

The WGCNA mining of gene modules was achieved with the R library WGCNA [29]. The standardized connectivity (Z.K) approach was used for identifying outliers suggested by WGCNA authors, with the threshold Z. K score < -2 as suggested by the WGCNA authors [30]. The gene co-expression similarity S_{xy} among genes x and y was calculated using the formula $S_{xy} = |\text{cor}(x, y)|$. The correlation of genes was estimated as follows: $axy =$





$|S_{xy}|^{\beta}$. By using the power gradient method, scale independence as well as mean connectivity were subsequently analyzed. An adequate β value was picked out once the degree of independence was higher than 0.85 [29] for generating a scale-free network. Next, the modules were generated using hierarchical clustering based on average linkage. The module eigengene (ME) was then determined and the module-trait correlations were computed by estimating the correlation among MEs and clinical traits for identifying modules relevant to each clinical trait. The module significance (MS) and the average absolute gene significance (GS) were calculated for evaluating the correlation of overall module genes with the clinical trait. GS is the \log_{10} transformation of P values obtained from the linear correlation model based on patient clinical features and gene expression. For the determination of candidate hub genes, the module connectivity (module membership (MM) \geq median, and gene significance (GS) \geq median [31]) was applied for selecting the most significant module genes for network construction and visualization in Cytoscape. Next, the MCODE plugin in Cytoscape was used for subnetwork extraction using the following setting: node score cut-off ≥ 0.2 , degree cut-off ≥ 2 , max depth = 100 and K-core ≥ 2 .

Functional enrichment analysis of genes

The functional enrichment was analyzed by the R library ClusterProfiler [32]. Terms with an enrichment P value of < 0.05 were considered as meaningful ones.

Detection of TFs regulating genes in key module

Module genes were inputted into Enrichr (<http://amp-pharm.mssm.edu/Enrichr/>) to uncover transcription factors (TFs) interacting with these genes. To reduce false-positives, we screened only TFs with targets available in ENCODE and ChEA gene-set libraries and with corrected P value < 0.05 based on the Fisher exact test. The Cytoscape 3.4.0 application (Cytoscape Consortium, San Diego, CA, USA) was used for TF-target gene network to visualization.

Validation of hub genes

The GSE1869 data was employed for verification of the hub genes associated with HF. We used the Wilcoxon test to measure the significance of correlation between hub genes' expression and HF. Additionally, ROC (receiver operating characteristic) analysis was performed on two data sets (GSE57345 and GSE1869) to validate the hub genes, and the area under curve (AUC) of ROC was computed to differentiate HF and non-HF.

Results

WGCNA data mining of key modules

Before performing a series of analyses, the GSE57345 dataset was preprocessed. The qualitative assessment of the microarray data was achieved by hierarchical clustering of samples. After detection and elimination of six outliers in the clusters, 313 samples were included in the dendrogram (Fig. 1a). Afterward, $\beta = 7$ was chosen as the power value of soft-threshold (Fig. 1b). Then, a total of

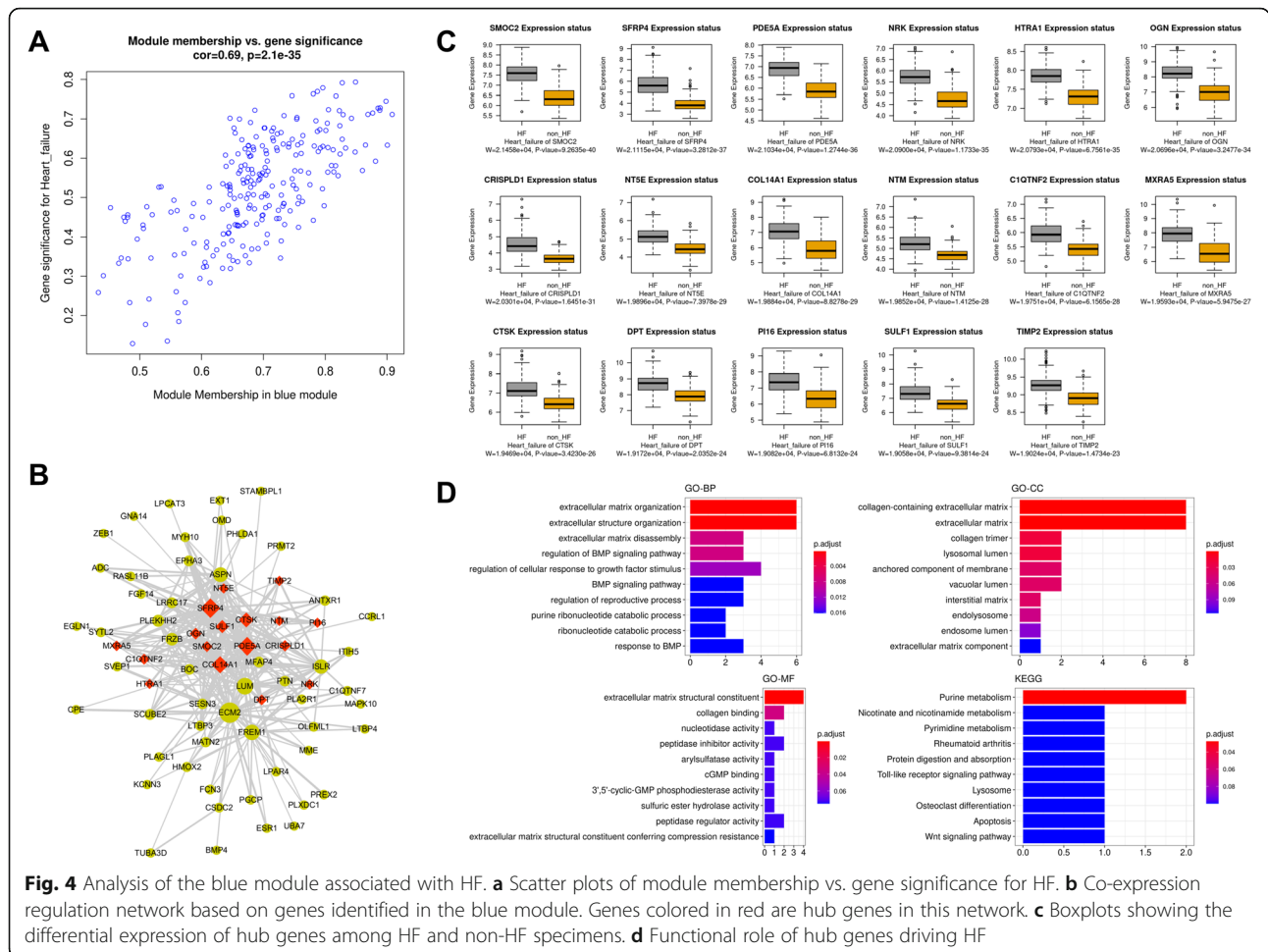


Fig. 4 Analysis of the blue module associated with HF. **a** Scatter plots of module membership vs. gene significance for HF. **b** Co-expression regulation network based on genes identified in the blue module. Genes colored in red are hub genes in this network. **c** Boxplots showing the differential expression of hub genes among HF and non-HF specimens. **d** Functional role of hub genes driving HF

seven modules were generated by average linkage hierarchical clustering based on the 6230 input genes (Fig. 2a). The Eigengene adjacency heatmap showing the correlation and clustering of the modules was reported in Fig. 2b, and hinted that the six significant correlation modules were divisible into two distinct clusters based on their ME correlation with a great level of independence within the modules. After calculating the MS of each module-trait correlation by WGCNA, the correlation between the available clinical features (HF, ISCH, CMP, Gender and Age) in the GSE57345 dataset with each module was as shown in Fig. 2c. The results showed that the blue module was positively and pointedly associated with the HF, ISCH, and CMP traits (Fig. 2c). In addition, the gene significance in relation with HF, ISCH and CMP across modules were as indicated in Figs. 2d-f, respectively. Since the blue module was the most significantly and positively linked with HF, ISCH, and CMP traits, this module was chosen for further analysis of genes associated with HF, ISCH, and CMP.

Functions of genes in the blue module

The clusterProfiler library was run to uncover the biological meaning of the totality of genes identified in the blue module. These genes were majorly enriched in biological processes (GO-BP) associated with extracellular matrix (Fig. 3a). The most enriched cellular components (GO-CC) were collagen-containing extracellular matrix, extracellular matrix component and extracellular matrix (Fig. 3b) while the most representative molecular functions (GO-MF) were extracellular matrix structural constituent, collagen binding and glycosaminoglycan binding (Fig. 3c). In the KEGG pathway analysis, Focal adhesion and Protein digestion were the overrepresented pathways (Fig. 3d).

Identification and functional role of hub genes associated with HF

The module membership vs. gene significance plot was as depicted in Fig. 4a. The MCODE plugin was used for uncovering the hub genes associated with HF in the blue module from the constructed network. We identified 17 hub genes (TIMP2, SMOC2, NRK, NTM, PDE5A,

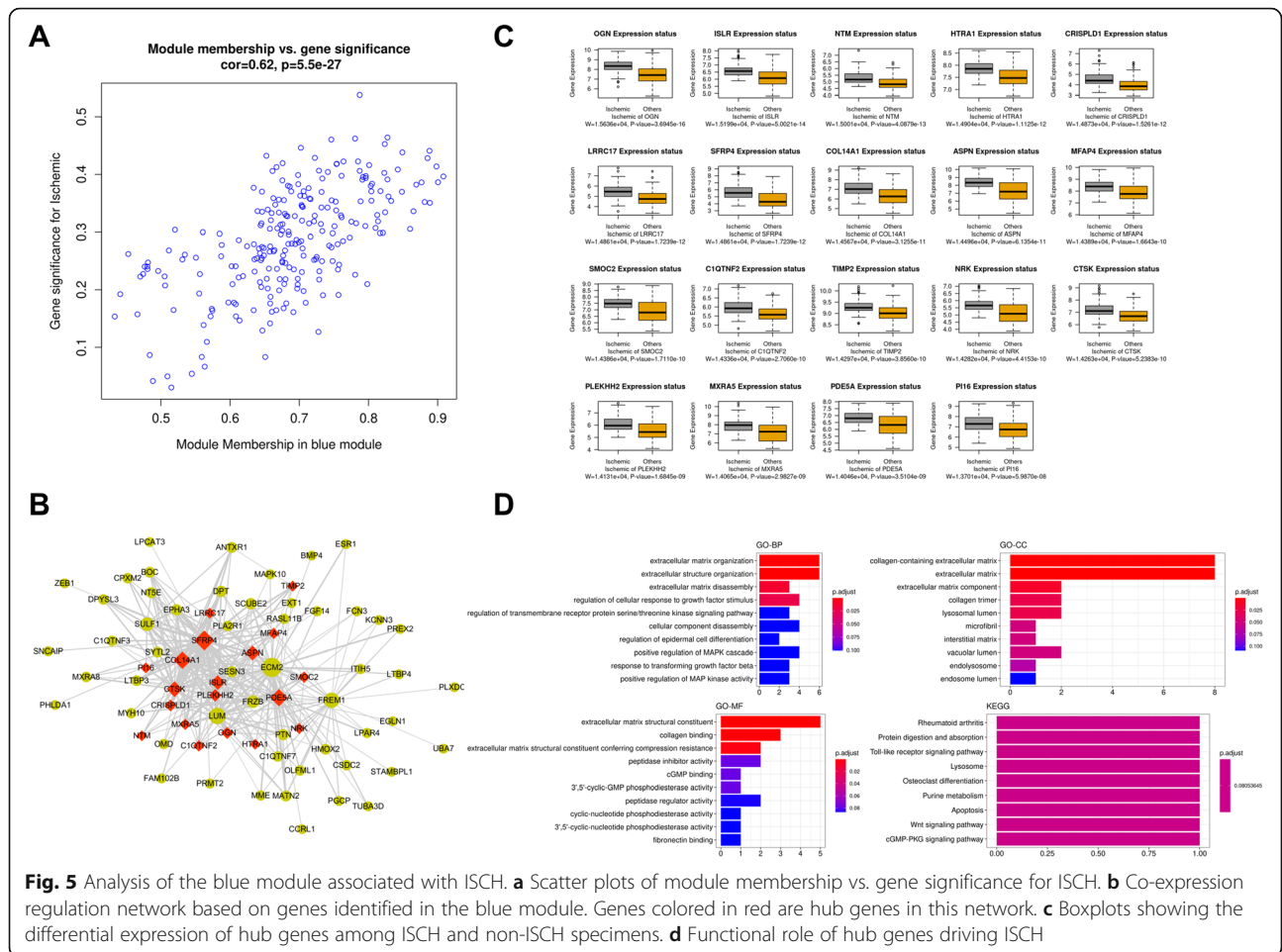


Fig. 5 Analysis of the blue module associated with ISCH. **a** Scatter plots of module membership vs. gene significance for ISCH. **b** Co-expression regulation network based on genes identified in the blue module. Genes colored in red are hub genes in this network. **c** Boxplots showing the differential expression of hub genes among ISCH and non-ISCH specimens. **d** Functional role of hub genes driving ISCH

CTSK, DPT, MXRA5, CRISPLD1, COL14A1, SFRP4, SULF1, OGN, PI16, HTRA1, NT5E, and C1QTNF2) which were colored in red in the network (Fig. 4b). The boxplot showed that these hub genes were significantly upregulated in HF (Fig. 4c). After that, we determined the functional role of hub genes associated with HF in the blue module and found that these genes were those driving the biological processes related to the organization and disassembly of the extracellular matrix (Fig. 4d). In the cellular component ontology, the hub genes were markedly enriched in collagen trimer and extracellular matrix (Fig. 4d). The GO term of collagen binding was the most considerably enriched molecular function (Fig. 4d). Purine metabolism, Nicotinate and nicotinamide metabolism, and Pyrimidine metabolism were the most mainly enriched pathways resulting from the KEGG pathway analysis (Fig. 4d).

Identification and functional role of hub genes associated with ISCH

The module membership vs. gene significance plot was as depicted in Fig. 5a. The MCODE plugin was used for uncovering the hub genes related to ISCH in the blue module from the constructed network. We identified 19

hub genes (NTM, ASPN, LRRIC17, ISLR, TIMP2, SMOC2, PLEKHH2, NRK, CTSK, PDE5A, MXRA5, CRISPLD1, COL14A1, SFRP4, MFAP4, OGN, PI16, HTRA1, and C1QTNF2) which were colored in red in the network (Fig. 5b). The boxplot of the expression of these key genes showed that all these genes were significantly upregulated in ISCH (Fig. 5c). The analysis of the functions of the hub genes in the blue module associated with ISCH indicated that these genes were those mainly controlling the processes related to extracellular matrix (Fig. 5d). In the cellular component ontology, the hub genes were greatly enriched in collagen-containing extracellular matrix (Fig. 5d). The GO terms of extracellular matrix structural constituent as well as the collagen binding were those significantly enriched molecular functions (Fig. 5d). Rheumatoid arthritis and Protein digestion and absorption were the most significantly enriched pathways resulting from the KEGG pathway analysis (Fig. 5d).

Identification and functional role of hub genes associated with CMP

The module membership vs. gene significance plot was as depicted in Fig. 6a. The MCODE plugin was used for

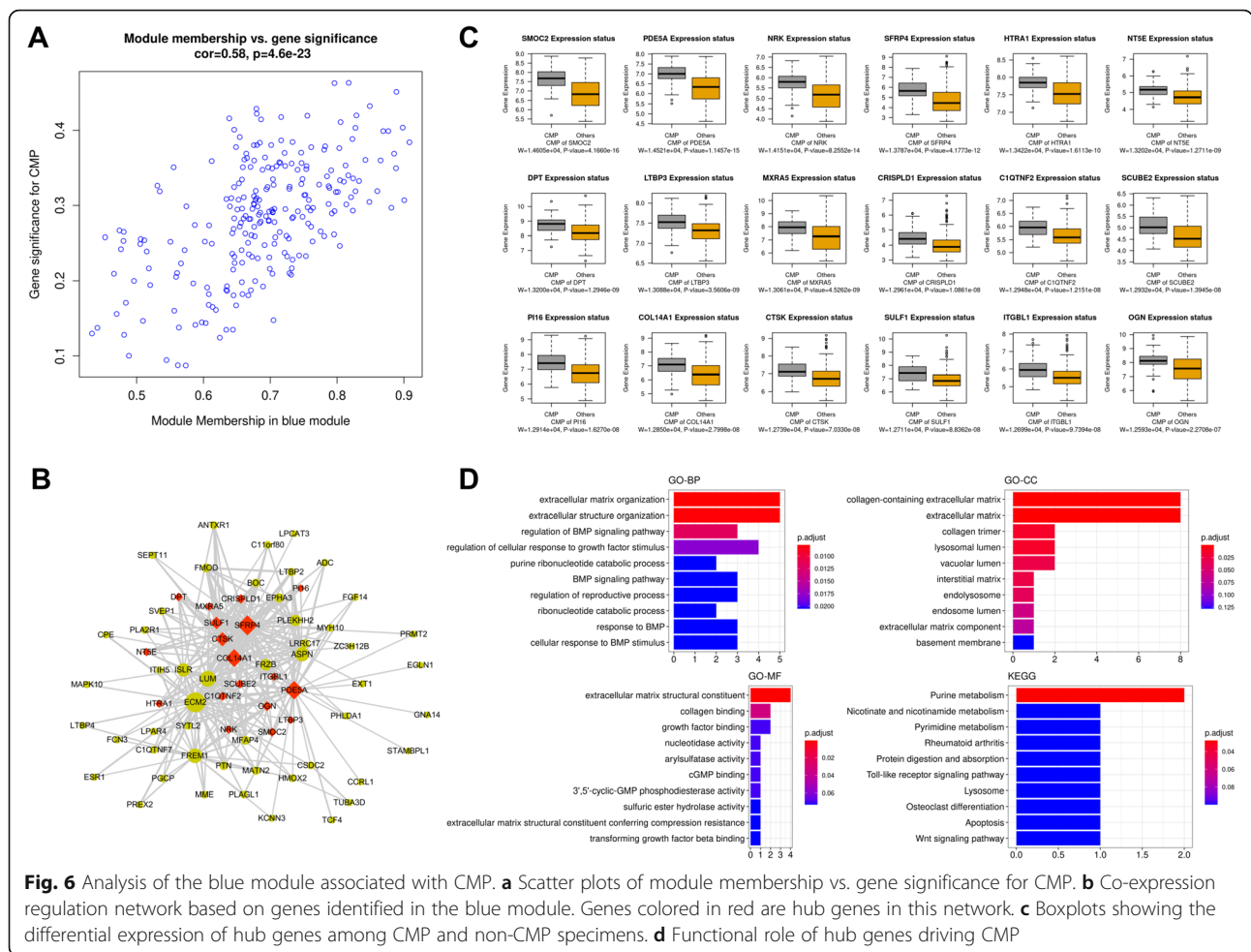


Fig. 6 Analysis of the blue module associated with CMP. **a** Scatter plots of module membership vs. gene significance for CMP. **b** Co-expression regulation network based on genes identified in the blue module. Genes colored in red are hub genes in this network. **c** Boxplots showing the differential expression of hub genes among CMP and non-CMP specimens. **d** Functional role of hub genes driving CMP

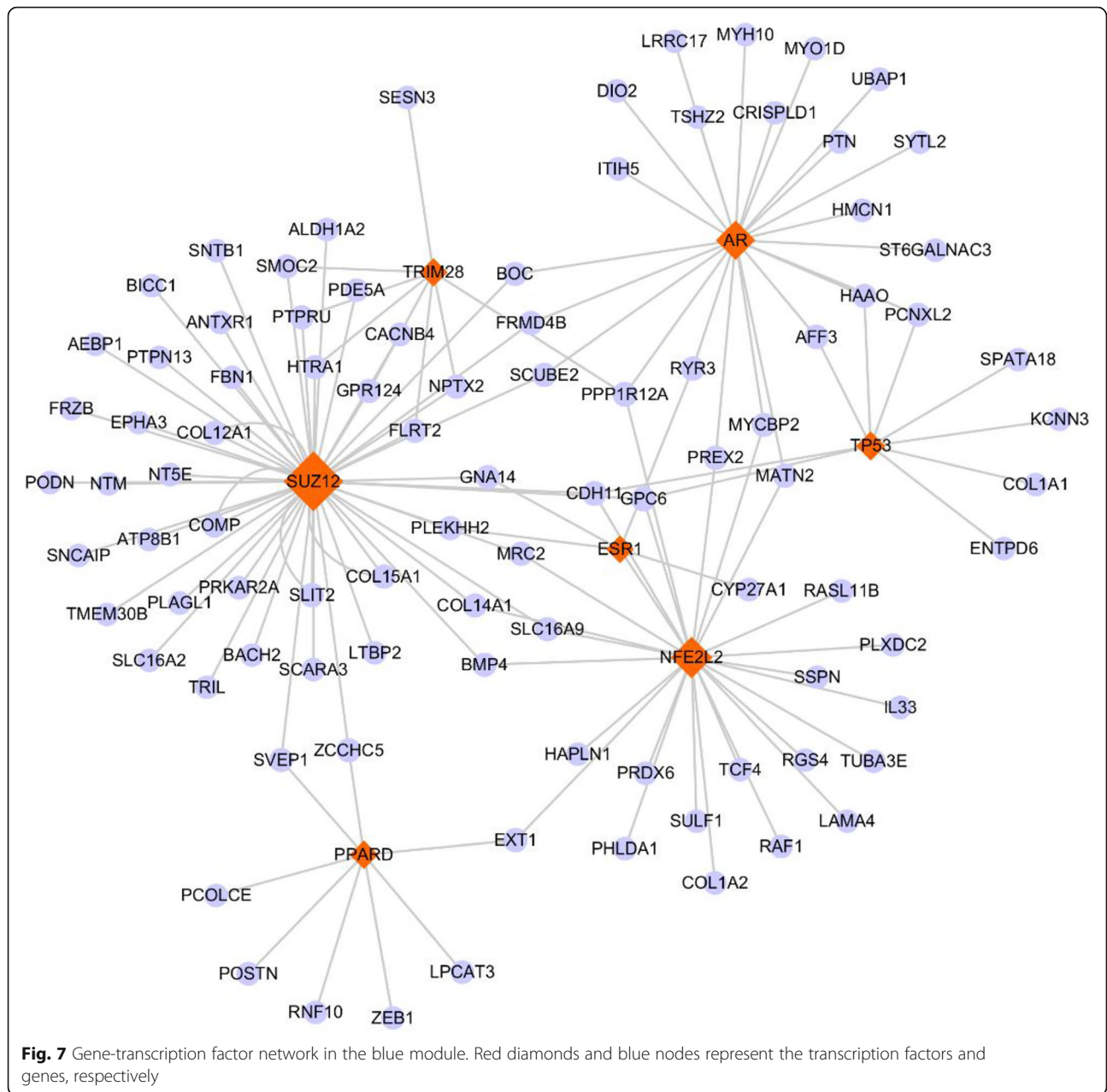
uncovering the hub genes associated with CMP in the blue module from the constructed network. We identified 18 hub genes (SCUBE2, CTSK, ITGBL1, MXRA5, CRISPLD1, SULF1, COL14A1, HTRA1, NT5E, OGN, PI16, C1QTNF2, SMOC2, SFRP4, LTBP3, NRK, PDE5A, and DPT) which were colored in red in the network (Fig. 6b). The boxplot signposted that these hub genes were significantly upregulated in CMP (Fig. 6c). The functional analysis of the hub genes in the blue module associated with CMP indicated their involvement in the biological processes related to extracellular matrix and regulation of BMP signaling pathway (Fig. 6d). In the cellular component ontology, the hub genes were mostly associated with terms related to extracellular matrix and collagen trimer (Fig. 6d). The most enriched GO terms of molecular functions included collagen binding and growth factor binding (Fig. 6d). The result of the KEGG pathway analysis indicated the involvement of hub genes in Purine metabolism, Nicotinate and nicotinamide metabolism, and Pyrimidine metabolism pathways (Fig. 6d).

Identification of transcription factors associated with genes in blue module

In order to explore transcription factors (TFs) controlling gene expression in the blue module, we analyzed ENCODE and ChEA which were the data sources available in Enrich. By setting an adjusted *P*-value cutoff of 0.05, a total of seven TFs were revealed (Additional file 1). The most prevalent TFs were SUZ12 with 47 target genes, NFE2L2 with 26 target genes and AR with 23 target genes. The regulatory network of TF-target gene based on all the genes in blue module and the seven TFs was as displayed in Fig. 7.

Validation of hub genes

All the hub genes of HF were selected for validation by using GSE1869 data set. However, only 11 hub genes were obtained in the GSE1869, which included OGN, HTRA1, MXRA5, TIMP2, PDE5A, DPT, COL14A1, CTSK, SFRP4, SULF1, and NT5E. The differential expression of hub genes between HF and non-HF samples showed that the overexpression of OGN, HTRA1 and MXRA5 were closely related to the occurrence of HF

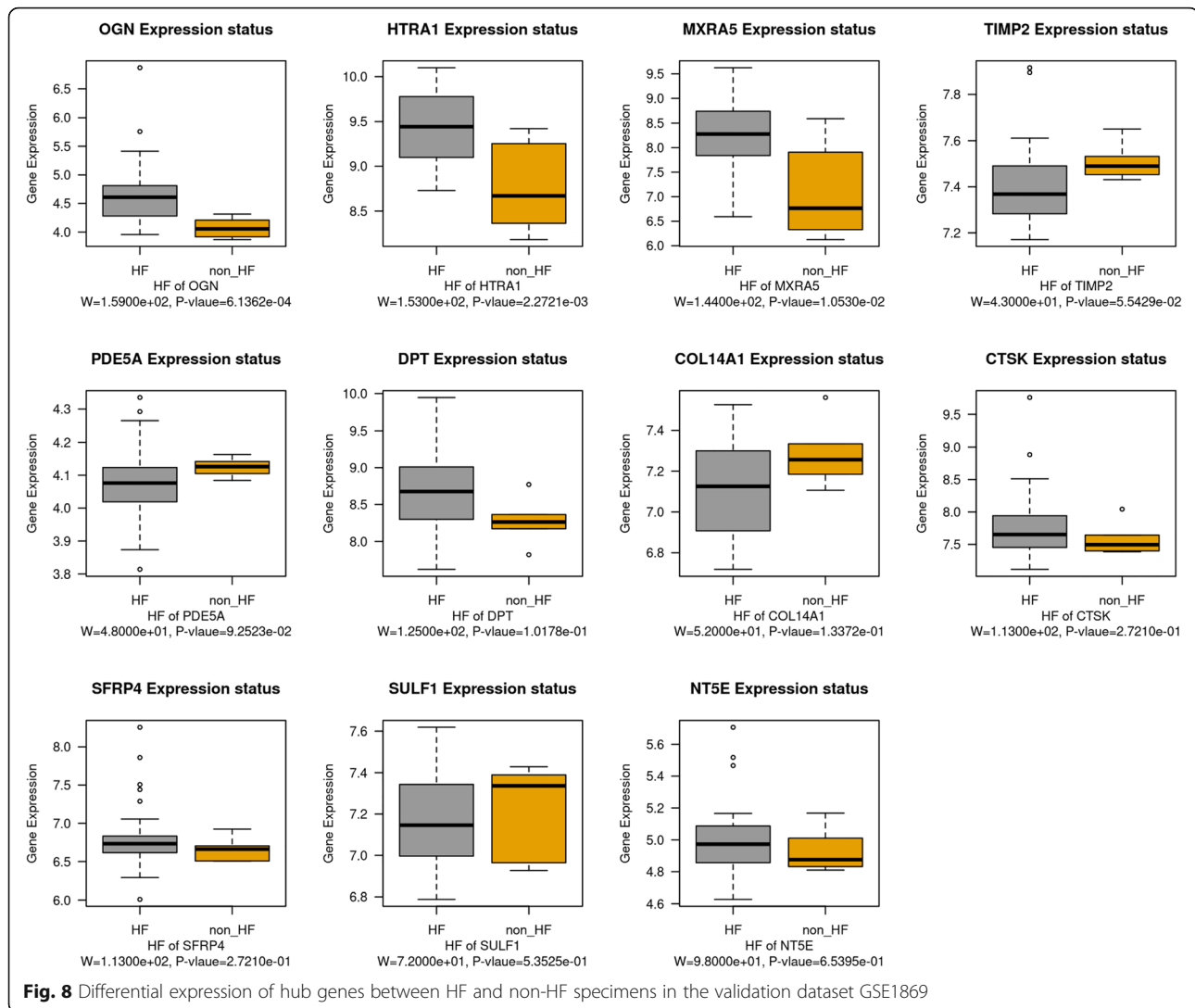


(Fig. 8). Thus, we speculated that such three genes were the key genes of HF progression. To validate the hub genes as predictive biomarkers of HF, we performed and calculated ROC curves [33] and the AUCs [95% confidence intervals (CIs)], respectively. The AUCs of OGN, HTRA1 and MXRA5 in the GSE57345 were respectively 0.912, 0.908 and 0.878, suggesting OGN, HTRA1 and MXRA5 as potential biomarkers of HF (Fig. 9a). The AUCs of OGN, HTRA1 and MXRA5 in the GSE1869 were respectively 0.914, 0.879 and 0.828, further indicating OGN, HTRA1 and MXRA5 as potential biomarkers of HF (Fig. 9b). We also found the AUC of each validated gene was higher than 0.7, indicating that OGN,

HTRA1 and MXRA5 could effectively distinguish HF and non-HF. Therefore, such genes were selected as the true key genes associated with HF. Further functional annotation of these true key genes was performed, and the result showed that these genes were majorly implicated in response to transforming growth factor beta, extracellular matrix and extracellular matrix structural constituent (Fig. 9c).

Discussion

The most frequent subtypes of HF include ischemic heart disease (ISCH) and dilated cardiomyopathy (CMP). ISCH is due to the shrinkage of blood supply to the



myocardium, while the heart of CMP becomes weakened and enlarged [34]. ISCH and CMP are conducive to symptoms similar to HF, but accumulating findings suggested that these two subtypes might respond differently to therapy [35, 36]. The new era of high-throughput technologies has experienced tremendous progress in the development of computational algorithms for bioinformatics purposes [37, 38]. In this current study, we employed the WGCNA data mining approach to identify genes that are significantly altered upon HF. Then we discovered the crucial modules markedly related to HF development, ISCH and CMP. Finally, the practicality of the key genes as prognostic biomarkers for HF was also evaluated.

In the WGCNA, we screened six significant gene modules from the GSE57345 dataset. The blue module was uncovered as the most crucially correlated with the status of HF, ISCH and CMP in patients, thus we chose the blue module as the main module for the subsequent

analysis. Our study hinted that genes clustered in the correlation gene expression module were chiefly implicated in regulation of ECM (extracellular matrix), including ECM organization and ECM receptor interaction. The ECM network plays an important role in cardiac homeostasis, not only by providing structural support, but also by transducing key signals to cardiomyocytes, vascular cells, and interstitial cells [39]. It is known that the alterations of ECM homeostasis may lead to diastolic or systolic dysfunction in heart and consequent development of HF [40]. A study conducted by Tsoutsman revealed that modulation of CCN2 on early ECM changes might provide a new therapeutic target in the treatment of HF [41].

After identifying the hub genes in the blue module underlying the studied traits (HF, ISCH and CMP), we identified 12 hub genes as those common to the three studied traits including SMOC2, NRK, PDE5A, CTSK, MXRA5, CRISPLD1, COL14A1, SFRP4, OGN, PI16,

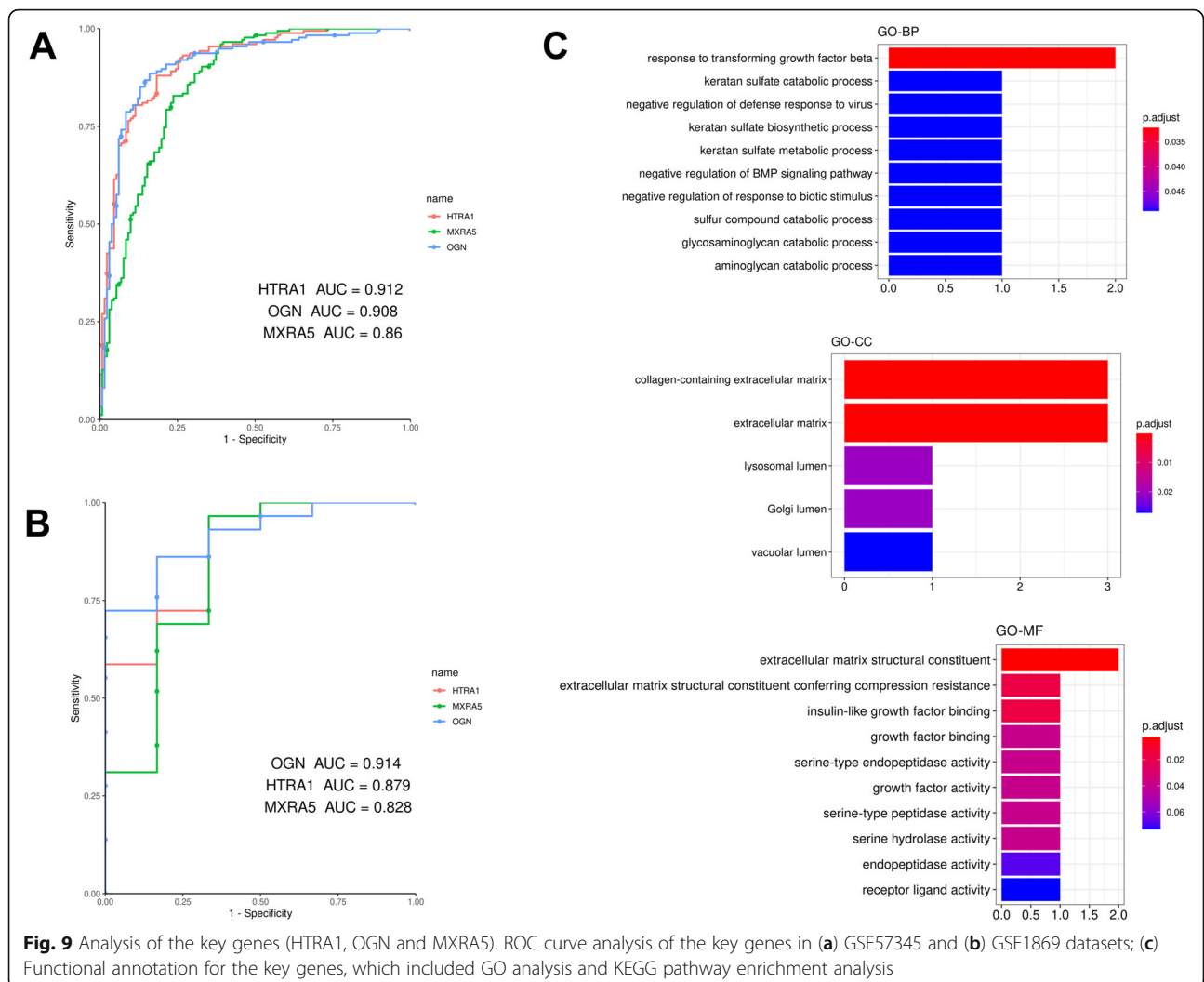


Fig. 9 Analysis of the key genes (HTRA1, OGN and MXRA5). ROC curve analysis of the key genes in (a) GSE57345 and (b) GSE1869 datasets; (c) Functional annotation for the key genes, which included GO analysis and KEGG pathway enrichment analysis

HTRA1 and C1QTNF2. Next, we found the functional enrichment of the hub genes associated with HF or ISCH or CMP were similar with those for the blue module. Williams and his colleagues found that SMOC2 was differentially expressed in failing right ventricular, and was potential targets for further study on HF [42]. A study revealed that SMOC2 could modulate fibroblast proliferation and extracellular matrix deposition [43]. Similarly, a study revealed that expression of C1QTNF2 related to anti-fibrotic function [44]. Thus, SMOC2 and C1QTNF2 might play a similar function to protect from cardiac fibrosis. Multiple studies have reported the pathophysiological role of cyclic guanosine monophosphate (cGMP) signaling in HF. Increased levels of cGMP have been demonstrated to exhibit cardioprotective effects in many cardiovascular diseases. PDE5A is a leading factor contributing to cGMP signaling and cardiac hypertrophy. Multiple studies suggested that PDE5A inhibitor could effectively limit myocardial injury caused by stresses [45–47]. As a lysosomal cysteine

protease, CTSK has been intensively investigated in the osteoporosis [48–50]. In recent years, reports hinted that activation of lysosomal cysteine protease might exert a deleterious role in the progression of cardiometabolic diseases [51]. Researchers have conveyed that CTSK may become an alternative therapeutic target for cardiac disease [52]. In a latest study, CRISPLD1 has been suggested to used be as a novel conserved target in the management of HF [53]. MXRA5 is a cancer related gene and several studies indicated the potential value of this gene as a novel therapeutic target for various cancers including colorectal cancer [54], non-small cell lung cancer [55] and glioblastoma multiform [56]. There are few reports about function of MXRA5 in HF. Studies have shown that SFRP4 is expressed in cardiomyocytes, and elevated during HF. Similarly to our results, studies demonstrated that OGN is significantly up-regulated in CMP and ISCH by reducing cardiac inflammation and injury, and, thus, could become a promising biomarker for HF [57]. The roles of NRK, COL14A1, PI16 and

HTRA1 in HF have not been investigated in the past. It is worth noting that most of these common genes have been proven as genes associated with the regulation of ECM, which corroborated with the functional annotation of the hub genes. Therefore, these genes (SMOC2, NRK, PDE5A, CTSK, MXRA5, CRISPLD1, COL14A1, SFRP4, OGN, PI16, HTRA1 and C1QTNF2) might be the biomarkers of HF, which could regulate organization of ECM to affect HF progression. Our present findings hinted that these genes could be crucial for understanding the pathogenesis of HF and even constitute relevant therapeutic targets.

As regulators of gene expression, TFs are closely linked with the pathogenesis of various diseases. Herein, we explored the possible regulation of genes in the blue module by TFs, and identified seven TFs, which included SUZ12, NFE2L2, TRIM28, AR, TP53, PPAR and ESR1. A new report revealed that SUZ12 (Suppressor of Zeste 12 Protein Homolog) could mediate the downregulation of myocyte enhancer factor 2A, thereby preventing cardiac hypertrophy [58]. It is well known that NFE2L2 (nuclear factor erythroid 2 like 2) is a critical TF that can induce adaptive responses against oxidative stress (OS) for maintaining cellular redox balance [59]. Accumulating researches demonstrated that the upregulation of NFE2L2 could protect the myocardium from ischemic injury and may be of therapeutic benefit in the treatment of ISCH [60]. AR (adrenergic receptor) overactivation is reported as a factor involved in the pathogenesis of HF [61]. Taken together, our findings indicated that these TFs formed a compact regulatory network with genes uncovered from the blue module, and the changes of the activities of these TFs may play crucial functions in the initiation and progression of HF, ISCH and CMP.

In addition, we verified the hub genes associated with HF based on the GSE1869 data set. However, among these 17 hub genes, only OGN, HTRA1, MXRA5, TIMP2, PDE5A, DPT, COL14A1, CTSK, SFRP4, SULF1, and NT5E were obtained in this data set. Then we used Wilcoxon test to determine whether these genes were significantly linked with HF status. A total of three genes were significantly correlated with HF status ($P < 0.05$), which included OGN, HTRA1 and MXRA5. We further validated their value for HF progression and prognosis by performing ROC curve analysis. In many researches, the ROC curve was applied to assess the performance of diagnosis, and the area under the ROC curve (AUC) is a commonly used summary index for comparison among multiple ROC curves [62–64]. In this study, the ROC analysis proved that the selected three key genes could be used as potential biomarkers with great specificity and sensitivity for prognosis in HF patients.

Conclusions

The current study applied the WGCNA data mining and other bioinformatics approaches to identify and validate the major module and corresponding hub genes involved in the progression of HF, ISCH and CMP. The blue module was identified as the common key module among three studied traits. Eleven hub genes, namely SMOC2, NRK, PDE5A, CTSK, MXRA5, CRISPLD1, COL14A1, SFRP4, OGN, PI16, HTRA1 and C1QTNF2 are likely to be prognostic biomarkers for development of HF. Afterward, TFs (SUZ12, NFE2L2, TRIM28, AR, TP53, PPAR and ESR1) were predicted as key regulators that contribute to the pathophysiological outcomes of HF. Finally, we screened three key genes (OGN, HTRA1 and MXRA5) which showed great specificity and sensitivity for HF prognosis in further validation. Though this study is a preliminary investigation, our findings proposed new potential therapeutic targets of HF, and provide novel insights in the pathogenesis of HF.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12920-020-00750-9>.

Additional file 1. ENCODE analysis of transcription factors.

Abbreviations

CK-MB: Creatine kinase MB isoenzyme; CHF: Congestive heart failure; VHD: Valvular heart disease; CMP: Dilated cardiomyopathy; GO: Gene ontology; TF: Transcription factor; ROC: Receiver operating characteristic; ISCH: Ischemic heart disease; ECM: Extracellular matrix; cGMP: Cyclic guanosine monophosphate; CTSK: Cathepsin K; OS: Oxidative stress; OGN: Osteoglycin

Acknowledgements

Not applicable.

Authors' contributions

TW conceptualized, designed the study and supervised the research group. JZ drafted the manuscript. JZ, WZ and CW interpreted the data. ZLZ, DY and XP collected data from publicly available datasets and contributed to data analysis, interpretation of the results and manuscript drafting. JP, RY, ZQZ, HQ and YW analyzed the data. JZ and TW have substantively revised the manuscript and contributed to the acquisition of funding. All the authors have read and approved the final submitted version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (81960054), the Natural Science Foundation of Jiangxi Province (20192BAB205010) and the Technological Project of Health Commission of Jiangxi Province (20195128). The funders had no role in the design or execution of this study.

Availability of data and materials

The datasets, GSE57345 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57345>) and GSE1869 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1869>), analyzed during the current study are available in the Gene Expression Omnibus (GEO) datasets at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/gds>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Cardiology, The First Affiliated Hospital of Nanchang University, No. 17 Yongwaizheng Street, Nanchang 330006, Jiangxi province, China. ²Hypertension Research Institute of Jiangxi Province, Nanchang 330006, China. ³Department of Respiratory Medicine, The First Affiliated Hospital of Nanchang University, Nanchang 330006, China.

Received: 12 April 2020 Accepted: 25 June 2020

Published online: 03 July 2020

References

- Benjamin E, Muntner P, Alonso A, Bittencourt M, Callaway C, Carson A, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation*. 2019;139:e56–e528.
- Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: systematic review of 67 prospective studies. *Resuscitation*. 2010;81(11):1479–87.
- Atwood C, Eisenberg MS, Herlitz J, Rea TD. Incidence of EMS-treated out-of-hospital cardiac arrest in Europe. *Resuscitation*. 2005;67(1):75–80.
- Little WC. Heart failure with a Normal left ventricular ejection fraction: diastolic heart failure. *Trans Am Clin Climatol Assoc*. 2008;119:93–102.
- Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of Cancer driver genes and mutations. *Cell*. 2018;174(4):1034–5.
- Thompson R, Johnston L, Taruscio D, Monaco L, Bérout C, Gut IG, et al. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *J Gen Intern Med*. 2014;29(Suppl 3):S780–7.
- Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, et al. Influenza research database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res*. 2017;45(D1):D466–d74.
- Cheng L, Leung K-S. Quantification of non-coding RNA target localization diversity and its application in cancers. *J Mol Cell Biol*. 2018;10(2):130–8.
- Cheng L, Liu P, Wang D, Leung KS. Exploiting locational and topological overlap model to identify modules in protein interaction networks. *BMC Bioinformatics*. 2019;20(1):23.
- Cheng L, Fan K, Huang Y, Wang D, Leung KS. Full characterization of localization diversity in the human protein Interactome. *J Proteome Res*. 2017;16(8):3019–29.
- Torella D, Rota M, Nurzynska D, Musso E, Monsen A, Shiraishi I, et al. Cardiac stem cell and myocyte aging, heart failure, and insulin-like growth factor-1 overexpression. *Circ Res*. 2004;94(4):514–24.
- Sabatasso S, Mangin P, Fracaso T, Moretti M, Docquier M, Djonov V. Early markers for myocardial ischemia and sudden cardiac death. *Int J Legal Med*. 2016;130(5):1265–80.
- Mueller C, Twerenbold R, Reichlin T. Early diagnosis of myocardial infarction with sensitive cardiac troponin assays. *Clin Chem*. 2019;65(3):490–1.
- Apple FS, Quist HE, Doyle PJ, Otto AP, Murakami MM. Plasma 99th percentile reference limits for cardiac troponin and creatine kinase MB mass for use with European Society of Cardiology/American College of Cardiology consensus recommendations. *Clin Chem*. 2003;49(8):1331–6.
- Shroff GR, Akkina SK, Miedema MD, Madlon-Kay R, Herzog CA, Kasiske BL. Troponin I levels and postoperative myocardial infarction following renal transplantation. *Am J Nephrol*. 2012;35(2):175–80.
- Tissier R, Hocini H, Tchitchek N, Deye N, Legriel S, Pichon N, et al. Early blood transcriptomic signature predicts patients' outcome after out-of-hospital cardiac arrest. *Resuscitation*. 2019;138:222–32.
- Argenziano MA, Doss MX, Tabler M, Sachinidis A, Antzelevitch C. Transcriptional changes associated with advancing stages of heart failure underlie atrial and ventricular arrhythmogenesis. *PLoS One*. 2019;14(5):e0216928.
- Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS One*. 2010;5(1):e8564.
- Liu W, Li L, Ye H, Tu W. Weighted gene co-expression network analysis in biomedicine research. *Sheng Wu Gong Cheng Xue Bao*. 2017;33(11):1791–801.
- Zuo Z, Shen JX, Pan Y, Pu J, Li YG, Shao XH, et al. Weighted gene correlation network analysis (WGCNA) detected loss of MAGI2 promotes chronic kidney disease (CKD) by Podocyte damage. *Cell physiol biochem*. 2018;51(1):244–61.
- Qin D, Wei R, Liu S, Zhu S, Zhang S, Min L. A Circulating miRNA-Based Scoring System Established by WGCNA to Predict Colon Cancer. *Anal Cell Pathol (Amst)*. 2019;2019:1571045.
- Liu H, Liu M, You H, Li X, Li X. Oncogenic network and hub genes for natural killer/T-cell lymphoma utilizing WGCNA. *Front Oncol*. 2020;10:223.
- Cheng L, Nan C, Kang L, Zhang N, Liu S, Chen H, et al. Whole blood transcriptomic investigation identifies long non-coding RNAs as regulators in sepsis. *J Transl Med*. 2020;18(1):217.
- Peng XY, Wang Y, Hu H, Zhang XJ, Li Q. Identification of the molecular subgroups in coronary artery disease by gene expression profiles. *J Cell Physiol*. 2019. <https://doi.org/10.1002/jcp.28324>.
- Niu X, Zhang J, Zhang L, Hou Y, Pu S, Chu A, et al. Weighted gene co-expression network analysis identifies critical genes in the development of heart failure after acute myocardial infarction. *Front Genet*. 2019;10:1214.
- Liu Y, Morley M, Brandimarto J, Hannehalli S, Hu Y, Ashley EA, et al. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*. 2015;105(2):83–9.
- Kittleson MM, Minhas KM, Irizarry RA, Ye SQ, Edness G, Breton E, et al. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. *Physiol Genomics*. 2005;21(3):299–307.
- Liu X, Li N, Liu S, Wang J, Zhang N, Zheng X, et al. Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review. *Front Bioeng Biotechnol*. 2019;7:358.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Oldham MC, Langfelder P, Horvath S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst Biol*. 2012;6:63.
- Lunnon K, Ibrahim Z, Proitsi P, Lourdasamy A, Newhouse S, Sattlecker M, et al. Mitochondrial dysfunction and immune activation are detectable in early Alzheimer's disease blood. *J Alzheimer's dis*. 2012;30(3):685–710.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
- Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J*. 2017;34(6):357–9.
- Liew CC, Dzau VJ. Molecular genetics and genomics of heart failure. *Nat Rev Genet*. 2004;5(11):811–25.
- Hannehalli S, Putt ME, Gilmore JM, Wang J, Parmacek MS, Epstein JA, et al. Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation*. 2006;114(12):1269–76.
- Kittleson MM, Ye SQ, Irizarry RA, Minhas KM, Edness G, Conte JV, et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation*. 2004;110(22):3444–51.
- Liu X, Xu Y, Wang R, Liu S, Wang J, Luo Y, et al. A network-based algorithm for the identification of moonlighting noncoding RNAs and its application in sepsis. *Brief Bioinform*. 2020. <https://doi.org/10.1093/bib/bbz154>.
- Cheng L, Leung K-S. Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome. *Bioinformatics*. 2018;34(20):3519–28.
- Frangogiannis NG. The extracellular matrix in ischemic and nonischemic heart failure. *Circ Res*. 2019;125(1):117–46.
- Kirk JA, Cingolani OH. Thrombospondins in the transition from myocardial infarction to heart failure. *J Mol Cell Cardiol*. 2016;90:102–10.
- Tsoutsman T, Wang X, Garchow K, Riser B, Twigg S, Semsarian C. CCN2 plays a key role in extracellular matrix gene expression in severe hypertrophic cardiomyopathy and heart failure. *J Mol Cell Cardiol*. 2013;62:164–78.
- Williams JL, Cavus O, Loccoch EC, Adelman S, Daugherty JC, Smith SA, et al. Defining the molecular signatures of human right heart failure. *Life Sci*. 2018;196:118–26.
- Gerarduzzi C, Kumar RK, Trivedi P, Ajay AK, Iyer A, Boswell S, et al. Silencing SMOC2 ameliorates kidney fibrosis by inhibiting fibroblast to myofibroblast transformation. *JCI Insight*. 2017;2(8):e90299.

44. Hicks DF, Goossens N, Blas-García A, Tsuchida T, Wooden B, Wallace MC, et al. Transcriptome-based repurposing of apigenin as a potential anti-fibrotic agent targeting hepatic stellate cells. *Sci Rep*. 2017;7:42563.
45. Pryszyzhna O, Burgoyne JR, Scotcher J, Grover S, Kass D, Eaton P. Phosphodiesterase 5 inhibition limits doxorubicin-induced heart failure by attenuating protein kinase G Ia oxidation. *J Biol Chem*. 2016;291(33):17427–36.
46. Westermann D, Becher PM, Lindner D, Savvatis K, Xia Y, Fröhlich M, et al. Selective PDE5A inhibition with sildenafil rescues left ventricular dysfunction, inflammatory immune response and cardiac remodeling in angiotensin II-induced heart failure in vivo. *Basic Res Cardiol*. 2012;107(6):308.
47. Mátyás C, Németh BT, Oláh A, Török M, Ruppert M, Kellermayer D, et al. Prevention of the development of heart failure with preserved ejection fraction by the phosphodiesterase-5A inhibitor vardenafil in rats with type 2 diabetes. *Eur J Heart Fail*. 2017;19(3):326–36.
48. Brömme D, Panwar P, Turan S. Cathepsin K osteoporosis trials, pycnodysostosis and mouse deficiency models: commonalities and differences. *Expert Opin Drug Discov*. 2016;11(5):457–72.
49. Mukherjee K, Chattopadhyay N. Pharmacological inhibition of cathepsin K: a promising novel approach for postmenopausal osteoporosis therapy. *Biochem Pharmacol*. 2016;117:10–9.
50. Helali AM, Iti FM, Mohamed IN. Cathepsin K inhibitors: a novel target but promising approach in the treatment of osteoporosis. *Curr Drug Targets*. 2013;14(13):1591–600.
51. Wu H, Du Q, Dai Q, Ge J, Cheng X. Cysteine protease Cathepsins in atherosclerotic cardiovascular diseases. *J Atheroscler Thromb*. 2018;25(2):111–23.
52. Guo R, Hua Y, Ren J, Bornfeldt KE, Nair S. Cardiomyocyte-specific disruption of Cathepsin K protects against doxorubicin-induced cardiotoxicity. *Cell Death Dis*. 2018;9(6):692.
53. Khadjeh S, Hindmarsh V, Weber F, Cyganek L, Vidal RO, Torkieh S, et al. CRISPLD1: a novel conserved target in the transition to human heart failure. *Basic Res Cardiol*. 2020;115(3):27.
54. Wang GH, Yao L, Xu HW, Tang WT, Fu JH, Hu XF, et al. Identification of MXRA5 as a novel biomarker in colorectal cancer. *Oncol Lett*. 2013;5(2):544–8.
55. He Y, Chen X, Liu H, Xiao H, Kwapong WR, Mei J. Matrix-remodeling associated 5 as a novel tissue biomarker predicts poor prognosis in non-small cell lung cancers. *Cancer Biomark*. 2015;15(5):645–51.
56. Rahane CS, Kutzner A, Heese K. A cancer tissue-specific FAM72 expression profile defines a novel glioblastoma multiform (GBM) gene-mutation signature. *J Neuro-Oncol*. 2019;141(1):57–70.
57. Van Aelst LN, Voss S, Carai P, Van Leeuwen R, Vanhoutte D, Sanders-van Wijk S, et al. Osteoglycin prevents cardiac dilatation and dysfunction after myocardial infarction through infarct collagen strengthening. *Circ Res*. 2015;116(3):425–36.
58. Yu J, Yang Y, Xu Z, Lan C, Chen C, Li C, et al. Long noncoding RNA Ahit protects against cardiac hypertrophy through SUZ12 (suppressor of Zeste 12 protein homolog)-mediated Downregulation of MEF2A (Myocyte enhancer factor 2A). *Circ Heart Fail*. 2020;13(1):e006525.
59. Ma YF, Zhao L, Coleman DN, Gao M, Looor JJ. Tea polyphenols protect bovine mammary epithelial cells from hydrogen peroxide-induced oxidative damage in vitro by activating NFE2L2/HMOX1 pathways. *J Dairy Sci*. 2019;102(2):1658–70.
60. Calvert JW, Elston M, Nicholson CK, Gundewar S, Jha S, Elrod JW, et al. Genetic and pharmacologic hydrogen sulfide therapy attenuates ischemia-induced heart failure in mice. *Circulation*. 2010;122(1):11–9.
61. Wang X, Zhao M, Wang X, Li S, Cao N, Liu H. The application of dynamic models to the exploration of $\beta(1)$ -AR Overactivation as a cause of heart failure. *Comput Math Methods Med*. 2018;2018:1613290.
62. Saha-Chaudhuri P, Heagerty PJ. Dynamic thresholds and a summary ROC curve: assessing prognostic accuracy of longitudinal markers. *Stat Med*. 2018;37(18):2700–14.
63. Guo J, Cui Z, Zheng Y, Li X, Chen Y. Comparison of Epstein-Barr virus serological tools for the screening and risk assessment of nasopharyngeal carcinoma: a large population-based study. *Pathol Oncol Res*. 2020. <https://doi.org/10.1007/s12253-020-00808-0>.
64. Wang Y, Chen L, Wang G, Cheng S, Qian K, Liu X, et al. Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis. *J Cell Physiol*. 2019;234(7):10225–37.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

