

# Analysis of multi-factor regulated network and different clusters in hypertrophic obstructive cardiomyopathy

**Xianyu Qin**

Guangdong Provincial People's Hospital

**Lei Huang**

Guangdong Provincial People's Hospital

**Shaoxian Chen**

Guangdong Provincial People's Hospital

**Pengju Wen**

Guangdong Provincial People's Hospital

**Sichengg Chen**

Guangdong Provincial People's Hospital

**Min Wu**

Guangdong Provincial People's Hospital

**Yueheng Wu** (✉ [edgar\\_wuyh@outlook.com](mailto:edgar_wuyh@outlook.com))

Guangdong Provincial People's Hospital <https://orcid.org/0000-0001-6891-7614>

---

## Research article

**Keywords:** different clusters, hypertrophic obstructive cardiomyopathy, multi-factor regulatory network, weighted gene co-expression network

**Posted Date:** June 8th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-33252/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Medical Genomics on August 6th, 2021. See the published version at <https://doi.org/10.1186/s12920-021-01036-4>.

# Abstract

## Background

Hypertrophic obstructive cardiomyopathy (HOCM) is a kind of hypertrophic cardiomyopathy (HCM) with more severe symptoms. We herein explored the regulatory network between lncRNA, miRNAs, mRNAs, and TFs in the HOCM and divided HOCM samples into different clusters to identify key genes and regulatory factors that are involved in the process of HOCM.

## Methods

Differentially expressed genes and online associated genes were integrated as a candidate gene set. WGCNA method was used to obtain key modules and genes related to HOCM. Subsequently, core genes were used to construct a multi-factor regulatory network and divided HOCM into different clusters. Lastly, the key regulatory factors and significant genes were identified in different clusters.

## Results

The four gene sets from GEO, GENE and OMIM databases were integrated and WGCNA network was used to obtain two modules and 32 core genes. Through online database, miRNA-lncRNA, miRNA-mRNA, and TF-mRNA interaction pairs were obtained to screen the regulatory factors that interact with the co-expressed key genes (N = 32), and finally a multi-factor regulatory network with 175 interaction pairs was obtained. The first 7 regulatory factors were obtained as the core regulatory factors, which included the lncRNAs (XIST, MALAT1, H19), TFs (SPI1 and SP1) and miRNAs (hsa-miR-29b-39 and has-miR-29a-3p). Finally, the unsupervised clustering method was used to divide HOCM samples into 4 clusters. As a result, four genes including COMP, FMOD, AEBP1 and SULF1 showed significant expression in different clusters.

## Conclusion

Bioinformatics method was used to obtain the molecular regulatory lncRNA-miRNA-mRNA and TF network and classified HOCM samples into different clusters. Finally, 32 co-expressed key genes that are of great significance for HOCM typing and 4 genes are considered as important biomarkers for different progressive stages or prognosis of HOCM. These might assist in discovering molecular mechanisms of HOCM in the future.

## 1. Introduction

Hypertrophic cardiomyopathy (HCM) is a common inheritable heart disease, with an estimated incidence rate of 0.2% in the general population (Semsarian et al., 2015; Geske et al., 2018). It is regarded as one of the main reasons for sudden cardiac death (SCD) in young people (Corrado et al., 2001). Of note, about

25–70% of HCM patients had left ventricular outflow tract obstruction LVOTO (Maron et al., 2003; Prinz et al., 2011; Veselka et al., 2017), which is also referred to as hypertrophic obstructive cardiomyopathy (HOCM). HOCM shows stronger symptoms and requires varied treatment strategies when compared with hypertrophic nonobstructive cardiomyopathy (Batzner et al., 2019). Unfortunately, SCD might be the first symptom observed in young individuals. However, the underlying molecular mechanisms of HOCM are still poorly understood. Therefore, it is necessary to study the mechanism of HOCM.

Long non-coding RNAs (lncRNAs) have more than 200 nucleotides and have similar structures with that of mRNAs that are not translated into proteins (Guttman et al., 2009). These participate in a variety of biological processes (Mattick, 2009). Previous studies have revealed that dysregulation of lncRNAs is involved in the pathogenesis of HCM (Yang et al., 2015; Shen et al., 2017). MicroRNAs (miRNAs) are noncoding RNA molecules (about 21 nucleotides) that prevent gene expression through post-transcriptional regulation. Some studies have reported that miRNAs, such as miR-1 (Sayed et al., 2007), miR-451 (Song et al., 2014), miR-22 (Huang et al., 2013), and so on, play an essential role in cardiac hypertrophy.

Further non-coding RNAs (ncRNAs) act as crucial biomarkers and functions as therapeutic targets in cardiovascular diseases. However, the signaling pathways and regulatory networks underlying the pathogenesis of HOCM is largely unknown. So, the discovery of new regulatory targets of HOCM is imperative for future therapeutic development.

In our research, two public datasets (GSE36961 and GSE89714, including HOCM samples and normal samples) and Gene database of NCBI, Online Mendelian Inheritance in Man database (OMIM) were integrated to screen the core genes associated with HOCM. Using the online database, a multi-factor regulatory network was constructed and divided HOCM into 4 clusters according to the core genes. Our research provides useful information in exploring the potential biomarkers for diagnosis and treatment of HOCM.

## **2. Materials And Methods**

### **2.1 Materials**

The datasets of GSE36961 contributed by Hebl VB et al. and GSE89714 contributed by Li Y et al. were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), and the expression profiling was generated using GPL15389 (Illumina HumanHT-12 V3.0 expression beadchip) and GPL11154 (IlluminaHiSeq 2000). The GSE36961 dataset contains a total of 39 control samples and 106 case samples, and the other dataset – GSE89714 included 5 case samples and 4 control samples. The case samples were of HOCM patients who underwent surgical septal myectomy because of LVOTO, and the control samples were donor cardiac tissues. Statistical analysis with the R (version 3.6.0) was performed.

Meanwhile, the genes associated with HCM were obtained from the GENE database of NCBI (<https://www.ncbi.nlm.nih.gov/gene/>) and Online Mendelian Inheritance in Man database (OMIM, <https://omim.org/>).

## 2.2 Data procession

The datasets of R language including the matching probes with genes were processed, and the median value of the expression level of the gene was selected when multiple probes were applied to the same gene. The differentially expressed genes (DEGs) were then obtained by using the R software package “limma”, according to  $\log_2|\text{fold change}|$  ( $|\log_2\text{FC}| > 0.58$ ) and adjusted the p-value ( $\text{adj.P.Val} < 0.05$ ). (Irizarry et al., 2003; Ritchie et al., 2015)

## 2.3 Identification of candidate gene set and function analysis

The DEGs from the two datasets were summed up, and then the HCM-associated genes from GENE and OMIM databases were obtained. Next, the redundancy was removed to obtain candidate gene set, and the expression profile data of candidate gene set in GSE36961 were selected for the weighted co-expression network analysis (WGCNA). The “ClusterProfiler” software package from R language was used to perform enrichment analysis of genes of candidate gene set including KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis and gene ontology (GO) enrichment analysis.

## 2.4 WGCNA

The “WGCNA” software package from R language was used to construct the weighted co-expression network for the candidate gene set. After screening HOCM-related modules, the “ClusterProfiler” software package from R language was used for enrichment analysis of genes from modules. Based on GO analysis and KEGG analysis, the significant GO terms and pathways were identified ( $\text{adj P-Val} < 0.05$ ). Meanwhile, the co-expressed gene set within the module was constructed into an interaction subnetwork, and the core genes of the module were screened according to the node degree of the network using the Cytoscape3.7.2 to visualize the network.

## 2.5 Construction of the multi-factor regulatory network

To screen the TF gene, the lncRNA, and the miRNA that interact with the core genes, the interaction pairs between ncRNAs and their target genes from the starBase v2.0 (<http://starbase.sysu.edu.cn/>) database (including miRNA-lncRNA and miRNA-mRNA) and interaction pairs between TFs and their target genes (TF-mRNA) from the TRRUST v2 database (<https://www.grnpedia.org/trrust/>) were downloaded (Li et al., 2014; Han et al., 2018).

MiRNA-lncRNA pairs should meet the number of supporting experiments that are greater or equal to high stringency ( $\geq 3$ ), and the miRNA-mRNA interaction pairs should also be identified by at least one of the softwares that consisted of targetScan, picTar, RNA22, PITA and miHCMnda/mirSVR. Finally, interaction pairs of core genes containing TF-mRNA, miRNA-lncRNA and miRNA-mRNA were obtained to construct a

multi-factor regulatory network. By calculating the degree of each regulator (miRNA, lncRNA, TF), the regulators that act as core regulators (degree > 5) were screened.

## 2.6 Clusters of HOCM

The key genes that co-expressed to GSE36961 for unsupervised clustering were mapped, and then the HOCM samples (N = 106) were selected to classify the HOCM samples by K-means unsupervised clustering method combined with “tsne” dimension reduction. Firstly, we chose the optimal K value (number of categories) by finding the inflection point of SSE (sum of the squared error, i.e., the sum of the squares of the distances of all points to the center of the cluster to which they belong) to divide HOCM samples into clusters. After that, the expression patterns of these core genes in different subtypes were examined and the genes with significant differences in different clusters were analyzed, which might act as potential marker genes in HOCM subtypes.

## 3. Results

### 3.1 Identification of candidate gene set and function analysis

The steps conducted in our study were shown in Fig. 1A, and the identified DEGs of the datasets GSE110226 and GSE89714 in HOCM were presented in **Supplementary Table 1**. A total of 628 DEGs were identified from GSE36961 ( $|\text{Log}_2\text{FC}| > 0.58$ ,  $\text{adj.P.Val} < 0.05$ ) between HOCM samples and normal tissues, which included 244 upregulated and 384 downregulated genes, and 1483 DEGs were identified from GSE89714 ( $|\text{Log}_2\text{FC}| > 0.58$ ,  $\text{adj.P.Val} < 0.05$ ), which included 903 upregulated and 580 downregulated genes (Fig. 1B). The volcano plots (**Supplementary Fig. 1**) are presented to display the DEGs.

A total of 177 HCM-associated genes were obtained from the GENE database of NCBI and 124 genes from the OMIM database (**Supplementary Table 2**). The DEGs of the two datasets were combined, and the HCM-associated genes were searched in two databases and removed the redundancy to get 2239 candidate genes, and this was considered as a candidate gene set (Fig. 1D). Furthermore, GO analysis and KEGG analysis were performed for candidate genes. The results of KEGG analysis showed that candidate genes showed significant enrichment in HCM and dilated cardiomyopathy pathways, and biological process (BP) of GO enrichment analysis involved in the muscle system process and heart process (Fig. 1E and **Supplementary Table 3**).

### 3.2 WGCNA

Based on 2239 candidate genes, the gene expression profile data from GSE36961 was applied to construct the WGCNA network (**Supplementary Table 4**). The results of the cluster showed no outlier sample, and so further analysis for 145 samples in the GSE36961 dataset was performed. A soft threshold ( $\beta$ ) value of 8 was set (Fig. 2A). Next, the expression matrix was transformed into the adjacent matrix, and then the adjacent matrix was transformed into the topology matrix (TOM). Based on TOM,

the average-linkage hierarchical clustering method was used to cluster the genes, and set a minimum number of genes for each gene network module to be 30. After determining the gene module by dynamic shear method, the eigengenes of each module were calculated at one time, and then clustering analysis was performed on the module. The modules with a relatively close distance were merged into a new module. A height of 0.25 was set, and a total of 5 modules were obtained (Fig. 2B and **Supplementary Table 5**)

The Pearson correlation coefficient between the module eigengenes and HOcm samples was calculated. The bigger value of the module is considered as more important in HOcm. Subsequently, the value of gene significance (GS) of every module was obtained. The bigger value of GS represented a more relative module with HOcm. Finally, the turquoise module (cor = 0.82,  $p < 1e-200$ ) and blue module (cor = 0.72,  $p < 4.3e-25$ ) were considered as the most relative modules with HOcm (Fig. 2C, D and **Supplementary Fig. 2**). Through the BP of GO enrichment analysis of the gene from the two modules, the genes in the turquoise module showed association with the muscle system process and muscle tissue development and the genes from the blue module participated with the muscle system process and the heart process. (Fig. 3 and **Supplementary Table 5**)

### 3.3 Construction of the multi-factor regulatory network

To identify the association of core genes with HOcm, the genes from the two modules were further analyzed. According to the gene expression relationship in the turquoise module, co-expression pairs with a connection threshold value of no less than 0.2 as edges of the co-expression network were chosen to construct a turquoise module network diagram, and the genes were selected with a node degree greater or equal to 3 ( $N = 15$ ) as the core genes of the turquoise module. Similarly, a connection threshold value of no less than 0.07 was set and the node degree of greater or equal to 3 ( $N = 17$ ) was selected as the core gene of the blue module. Also the 32 core genes were used as co-expression key genes for further analysis. (Fig. 4A and **Supplementary Table 6**)

A total of 376 miRNA-lncRNA interactions and 160774 miRNA-mRNA interactions were obtained from the starBase v2.0 database, and a total of human 9396 TF-mRNA interaction pairs from the TRRUST v2 database were obtained (**Supplementary Table 7**). Firstly, the miRNAs and transcription factors that interact with the co-expressed key genes ( $N = 32$ ) were screened, followed by selecting the lncRNAs that interact with these miRNAs, and finally obtained a multi-factor regulatory network with 175 interaction pairs. (Fig. 4B and **Supplementary Table 8**) Using the Mcode plugin of Cytoscape to calculate the degrees of every regulatory factor (miRNA, lncRNA, and TF), the first 7 regulatory factors were obtained as the core regulatory factors, which included the lncRNAs (XIST, MALAT1, H19), TFs (SP11 and SP1) and miRNAs (hsa-miR-29b-39 and has-miR-29a-3p). (Fig. 4C and **Supplementary Table 8**)

To explore the functions and pathways that are regulated by these core regulators, a total of 8397 target genes that interact with the core regulators from miRNA-lncRNA, miRNA-mRNA, and TF-mRNA interaction pairs were selected. Enrichment analysis for these target genes was performed, and the results obtained

the target genes that are related to biological processes such as positive regulation of catabolic process, histone modification, and proteasomal protein catabolic process. (Supplementary Fig. 3 and Supplementary Table 8)

## 3.4 Clusters of HOCM

The 32 co-expressed key genes were mapped to GSE36961 for unsupervised clustering, and an optimal K value of 4 was set. (Fig. 5A and Supplementary Table 9) All HOCM samples were clearly divided into 4 clusters, and the expression of key genes in all HCMs was shown in the heatmap. (Fig. 5B,C and Supplementary Table 9)

Furthermore, the differences of 32 co-expressed key genes in different clusters were analyzed, and the average value of each gene in each class was taken as the gene expression value in that class. As a result, four genes including the COMP, FMOD, AEBP1 and SULF1 showed significant expression in different clusters ( $p < 0.01$ ), (Fig. 5D, E).

## 4. Discussion

HOCM is regarded as a primary risk factor of SCD caused by HCM. Therefore, exploration of the pathogenic mechanism is of utmost importance. In our recent study, two datasets were integrated from the GEO database on HOCM and associated genes of HCM in the GENE and OMIM databases to obtain the candidate gene set. Then, WGCNA method was used to identify the related modules of HOCM. The integration of high-throughput data, online databases and bioinformatic method for scale-free network have widened the disease spectrum and strengthened the evidence. BP analysis indicated that the candidate gene set and genes in most of the relevant modules were targeted to muscle system process, muscle contraction and heart process. Pathway analysis indicated that candidate gene set was mostly enriched in HCM, focal adhesion and dilated cardiomyopathy. These results showed correlation with HOCM, and so, the co-expressed 32 genes with the highest degree were chosen in the two modules as core genes.

The miRNAs, lncRNAs and TFs that interact with the co-expressed key genes were then screened to obtain a multi-factor regulatory network. To date, several studies have reported the features of ncRNAs in the HOCM (Ntelios et al., 2017; Salman et al., 2018). Nevertheless, the details of RNA crosstalk in HOCM have not been elucidated. In this study, an integrated lncRNA-miRNA-mRNA-TF regulatory network was constructed, expounding the views on gene regulation at the pre-transcriptional and post-transcriptional levels. Moreover, bioinformatics technology was applied to explore the key molecules that are involved in the process of HOCM, which might be considered as optimal candidate markers for future therapy. The first 7 regulatory factors were found as the core regulatory factors, which included the lncRNAs (XIST, MALAT1, H19), TFs (SPI1 and SP1) and miRNAs (hsa-miR-29b-39 and has-miR-29a-3p). XIST, which is named as lncRNA X-inactive specific transcript, has been identified as a necessary regulator of cardiac hypertrophy by regulating miR-101 (Xiao et al., 2019) and miR-330 (Chen et al., 2018). XIST also showed association with heart failure in females (Heidecker et al., 2010). *In vivo* experiments revealed that

knockdown of XIST can inhibit the myocardial cell apoptosis in acute myocardial infarction rat model by regulating miR-449 (Zhang et al., 2019). Besides, H19 has been identified as a regulator that targets PPAR $\alpha$  of cardiac hypertrophy (Liu et al., 2016; Zhou et al., 2019). The results showed that XIST, MALAT1 and H19 possibly regulated other miRNAs involved in cardiac hypertrophy, such as miR-15b(Ooi et al., 2014), miR-19b(Li et al., 2019) and miR-29b (Yang et al., 2020) in our research.

The miRNAs and TFs consistent with other researches were identified. MiR-29 is a regulator of cardiomyocyte hypertrophy through wnt and mTOR signaling pathways (Sassi et al., 2017; Shi et al., 2019). Moreover, SP1 can regulate cardiomyocyte hypertrophy by inducing lncRNA CTBP-AS2 (Luo et al.) and SP1/GATA4 signaling pathways (Sun et al., 2018). However, SPI1 has not been reported in cardiac hypertrophy so far.

So, the core genes were used to divide the HOCM into 4 clusters. Although no difference was observed in the by the expression of a single gene, the HOCM was divided into 4 clusters clearly by combining the 32 co-expressed key genes. Finally, the 4 genes COMP, FMOD, AEBP1 and SULF1 showed significantly different expression in the four clusters. From this point, we speculated that these 32 co-expressed key genes are of great significance for HOCM typing and the 4 genes are considered as important biomarkers due to different progressive stages or prognosis of HOCM. However, there are several limitations to this study. Firstly, our study focused on data mining and analysis without any experimental confirmation; and 2) due to lack of relevant prognostic information, clinical classification of HOCM related to key genes and survival analysis associated with key genes was not conducted.

## 5. Conclusion

In our study, a total of 32 key genes and 7 regulatory factors that might be used as biomarkers in HOCM were identified. Classification of HOCM samples based on the co-expression of key genes demonstrated a better classification effect. Four genes showed significant differences in different subclasses and can be used as marker genes for different hypertrophic cardiomyopathy subclasses. The analysis method used an online database for gene expression, and provided a novel method for the diagnosis and prognosis of HOCM. However, based on the limitations of the present study, additional clinical research and molecular mechanisms are warranted to confirm these findings.

## Declarations

### Acknowledgments

None

### Funding

This study was supported by the Natural Science Foundation of Guangdong Province, China (No. 2016A030313796) and the Key Program of Natural Science Foundation of Guangdong Province, China

(No. 2017B030311010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### **Conflict of interest**

The authors declare that they have no conflict of interest.

### **Ethical statement**

Not applicable.

### **Authors' Contribution**

Min Wu and Yueheng Wu conceived and designed the present study. Xianyu Qin and Lei Huang collected, extracted and analyzed the data. Sicheng Chen, Shaoxian Chen and Pengju Wen performed data analysis and interpretation. All authors wrote the manuscript. All authors read and approved the final manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

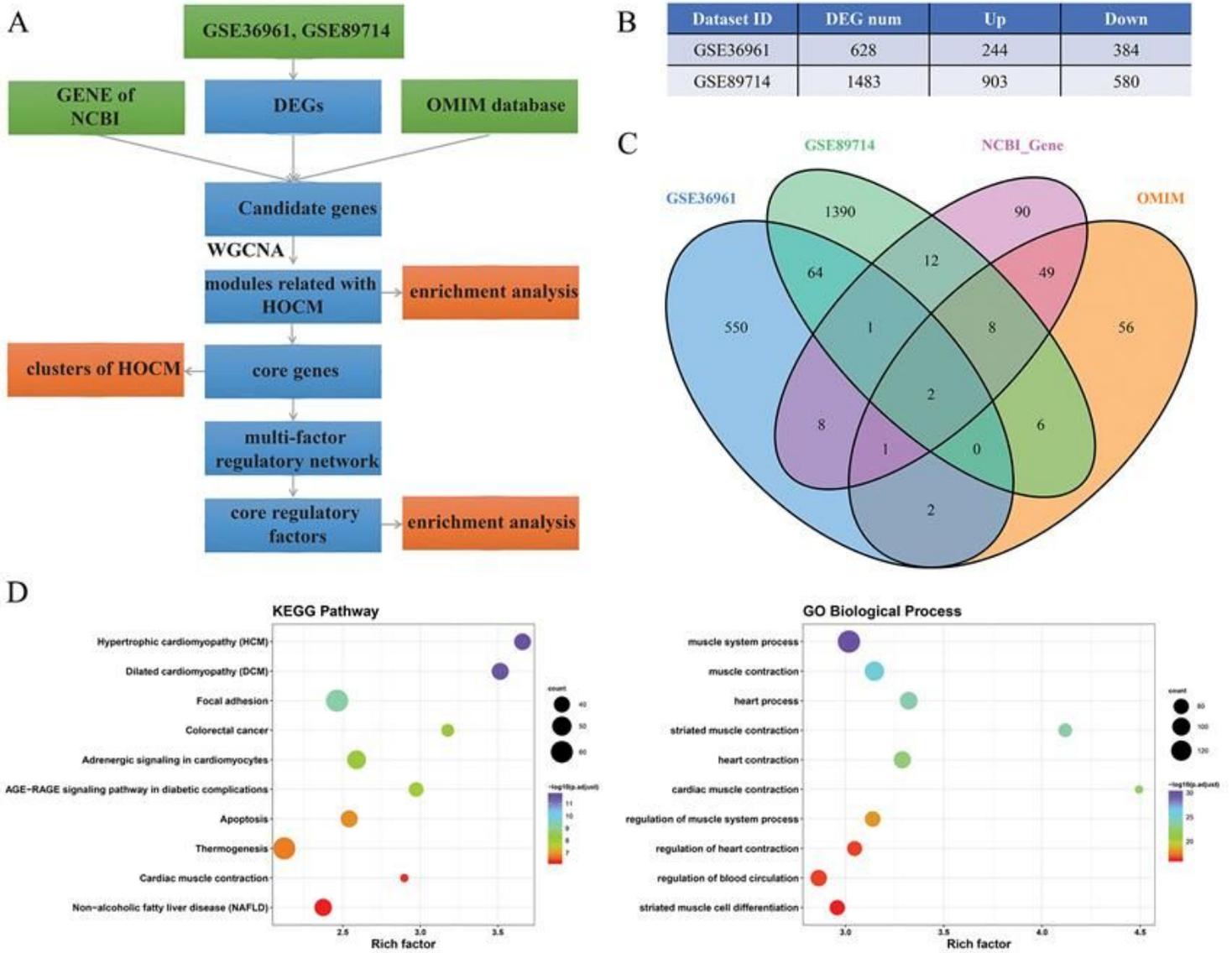
## **References**

1. Batzner A, Schafers HJ, Borisov KV, Seggewiss H. Hypertrophic Obstructive Cardiomyopathy. *Dtsch Arztebl Int.* 2019;116:47–53.
2. Chen Y, Liu X, Chen L, Chen W, Zhang Y, Chen J, Wu X, Zhao Y, Wu X, Sun G. The long noncoding RNA XIST protects cardiomyocyte hypertrophy by targeting miR-330-3p. *Biochem Biophys Res Commun.* 2018;505:807–15.
3. Corrado D, Basso C, Thiene G. Sudden cardiac death in young people with apparently normal heart. *Cardiovasc Res.* 2001;50:399–408.
4. Geske JB, Ommen SR, Gersh BJ. Hypertrophic Cardiomyopathy: Clinical Update. *JACC Heart Fail.* 2018;6:364–75.
5. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009;458:223–7.
6. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon HN, Jung H, Nam S, Chung M, Kim JH, Lee I. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 2018;46:D380-d386.
7. Heidecker B, Lamirault G, Kasper EK, Wittstein IS, Champion HC, Breton E, Russell SD, Hall J, Kittleson MM, Baughman KL, Hare JM. The gene expression profile of patients with new-onset heart failure reveals important gender-specific differences. *Eur Heart J.* 2010;31:1188–96.

8. Huang ZP, Chen J, Seok HY, Zhang Z, Kataoka M, Hu X, Wang DZ. MicroRNA-22 regulates cardiac hypertrophy and remodeling in response to stress. *Circ Res.* 2013;112:1234–43.
9. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003;4:249–64.
10. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42:D92-7.
11. Li X, Teng C, Ma J, Fu N, Wang L, Wen J, Wang TY. miR-19 family: A promising biomarker and therapeutic target in heart, vessels and neurons. *Life Sci.* 2019;232:116651.
12. Liu L, An X, Li Z, Song Y, Li L, Zuo S, Liu N, Yang G, Wang H, Cheng X, Zhang Y, Yang X, Wang J. The H19 long noncoding RNA is a novel negative regulator of cardiomyocyte hypertrophy. *Cardiovasc Res.* 2016;111:56–65.
13. Luo X, He S, Hu Y, Liu J, Chen X. Sp1-induced LncRNA CTBP1-AS2 is a novel regulator in cardiomyocyte hypertrophy by interacting with FUS to stabilize TLR4. *Cardiovasc. Pathol.* 42, 21–29.
14. Maron MS, Olivetto I, Betocchi S, Casey SA, Lesser JR, Losi MA, Cecchi F, Maron BJ. Effect of left ventricular outflow tract obstruction on clinical outcome in hypertrophic cardiomyopathy. *N Engl J Med.* 2003;348:295–303.
15. Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet.* 2009;5:e1000459.
16. Ntelios D, Meditskou S, Efthimiadis G, Pitsis A, Nikolakaki E, Girtovitis F, Parcharidou D, Zegkos T, Koidou S, Karvounis H, Tzimagiorgis G. Elevated plasma levels of miR-29a are associated with hemolysis in patients with hypertrophic cardiomyopathy. *Clin Chim Acta.* 2017;471:321–6.
17. Ooi JY, Bernardo BC, McMullen JR. The therapeutic potential of miRNAs regulated in settings of physiological cardiac hypertrophy. *Future Med Chem.* 2014;6:205–22.
18. Prinz C, Farr M, Hering D, Horstkotte D, Faber L. The diagnosis and treatment of hypertrophic cardiomyopathy. *Dtsch Arztebl Int.* 2011;108:209–15.
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
20. Salman OF, El-Rayess HM, Khalil A, Nemer C, G. and Refaat MM. Inherited Cardiomyopathies and the Role of Mutations in Non-coding Regions of the Genome. *Front Cardiovasc Med.* 2018;5:77.
21. Sassi Y, Avramopoulos P, Ramanujam D, Grüter L, Werfel S, Giosele S, Brunner AD, Esfandyari D, Papadopoulou AS, De Strooper B, Hübner N, Kumarswamy R, Thum T, Yin X, Mayr M, Lagerbauer B, Engelhardt S. Cardiac myocyte miR-29 promotes pathological remodeling of the heart by activating Wnt signaling. *Nat Commun.* 2017;8:1614.
22. Sayed D, Hong C, Chen IY, Lypowy J, Abdellatif M. MicroRNAs play an essential role in the development of cardiac hypertrophy. *Circ Res.* 2007;100:416–24.
23. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol.* 2015;65:1249–54.

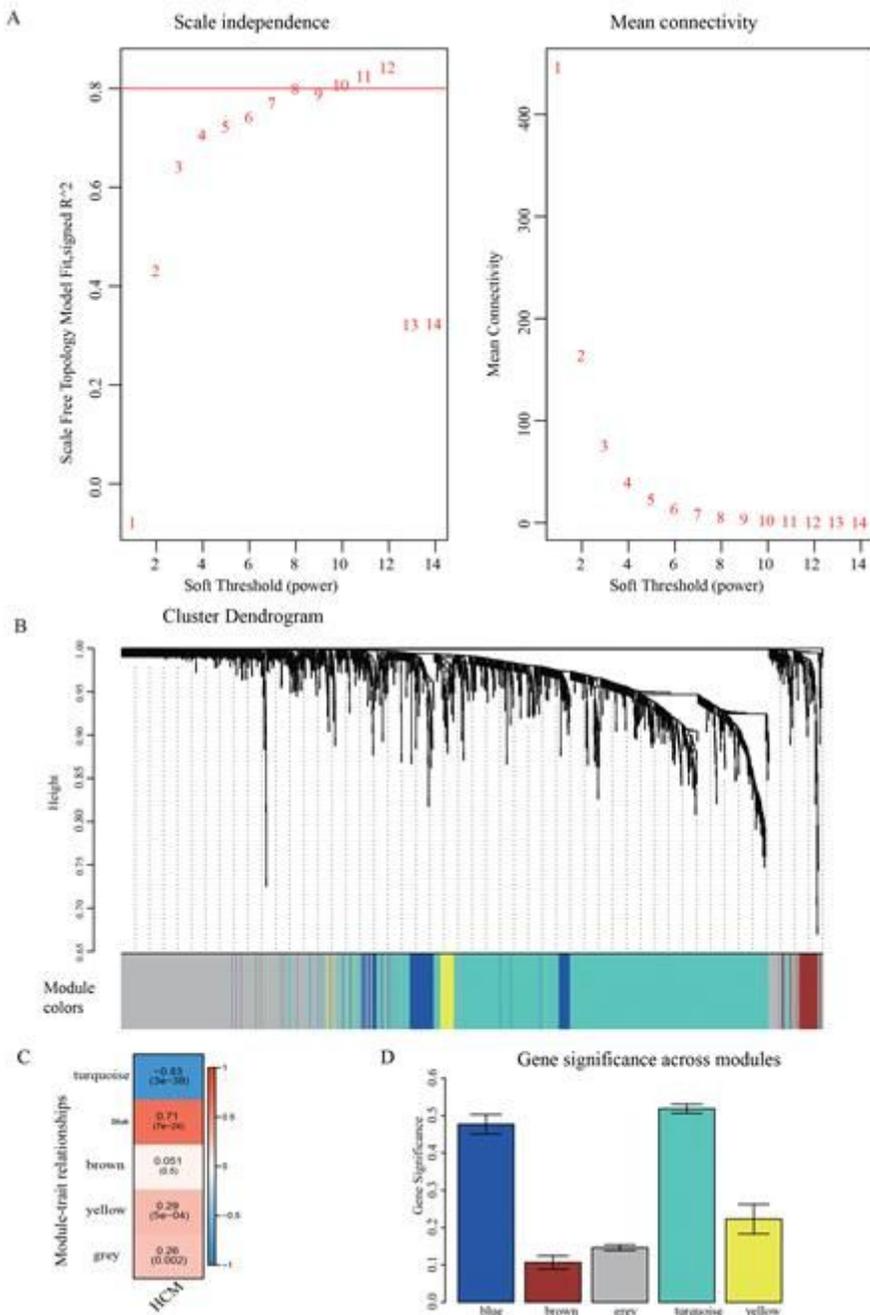
24. Shen S, Jiang H, Bei Y, Xiao J, Li X. Long Non-Coding RNAs in Cardiac Remodeling. *Cell Physiol Biochem.* 2017;41:1830–7.
25. Shi JY, Chen C, Xu X, Lu Q. miR-29a promotes pathological cardiac hypertrophy by targeting the PTEN/AKT/mTOR signalling pathway and suppressing autophagy. *Acta Physiol (Oxf).* 2019;227:e13323.
26. Song L, Su M, Wang S, Zou Y, Wang X, Wang Y, Cui H, Zhao P, Hui R, Wang J. MiR-451 is decreased in hypertrophic cardiomyopathy and regulates autophagy by targeting TSC1. *J Cell Mol Med.* 2014;18:2266–74.
27. Sun S, Li T, Jin L, Piao ZH, Liu B, Ryu Y, Choi SY, Kim GR, Jeong JE, Wi AJ, Lee SJ, Kee HJ, Jeong MH. *Dendropanax morbifera* Prevents Cardiomyocyte Hypertrophy by Inhibiting the Sp1/GATA4 Pathway. *Am J Chin Med.* 2018;46:1021–44.
28. Veselka J, Anavekar NS, Charron P. Hypertrophic obstructive cardiomyopathy. *Lancet.* 2017;389:1253–67.
29. Xiao L, Gu Y, Sun Y, Chen J, Wang X, Zhang Y, Gao L, Li L. The long noncoding RNA XIST regulates cardiac hypertrophy by targeting miR-101. *J Cell Physiol.* 2019;234:13680–92.
30. Yang Q, Wu F, Mi Y, Wang F, Cai K, Yang X, Zhang R, Liu L, Zhang Y, Wang Y, Wang X, Xu M, Gui Y, Li Q. 2020. Aberrant expression of miR-29b-3p influences heart development and cardiomyocyte proliferation by targeting NOTCH2. *Cell Prolif.*, e12764.
31. Yang W, Li Y, He F, Wu H. 2015. Microarray profiling of long non-coding RNA (lncRNA) associated with hypertrophic cardiomyopathy. *BMC Cardiovascular Disorders* 15.
32. Zhang M, Liu HY, Han YL, Wang L, Zhai DD, Ma T, Zhang MJ, Liang CZ, Shen Y. Silence of lncRNA XIST represses myocardial cell apoptosis in rats with acute myocardial infarction through regulating miR-449. *Eur Rev Med Pharmacol Sci.* 2019;23:8566–72.
33. Zhou H, Wang B, Yang YX, Jia QJ, Zhang A, Qi ZW, Zhang JP. 2019. Long Noncoding RNAs in Pathological Cardiac Remodeling: A Review of the Update Literature. *Biomed Res Int* 2019, 7159592.

## Figures



**Figure 1**

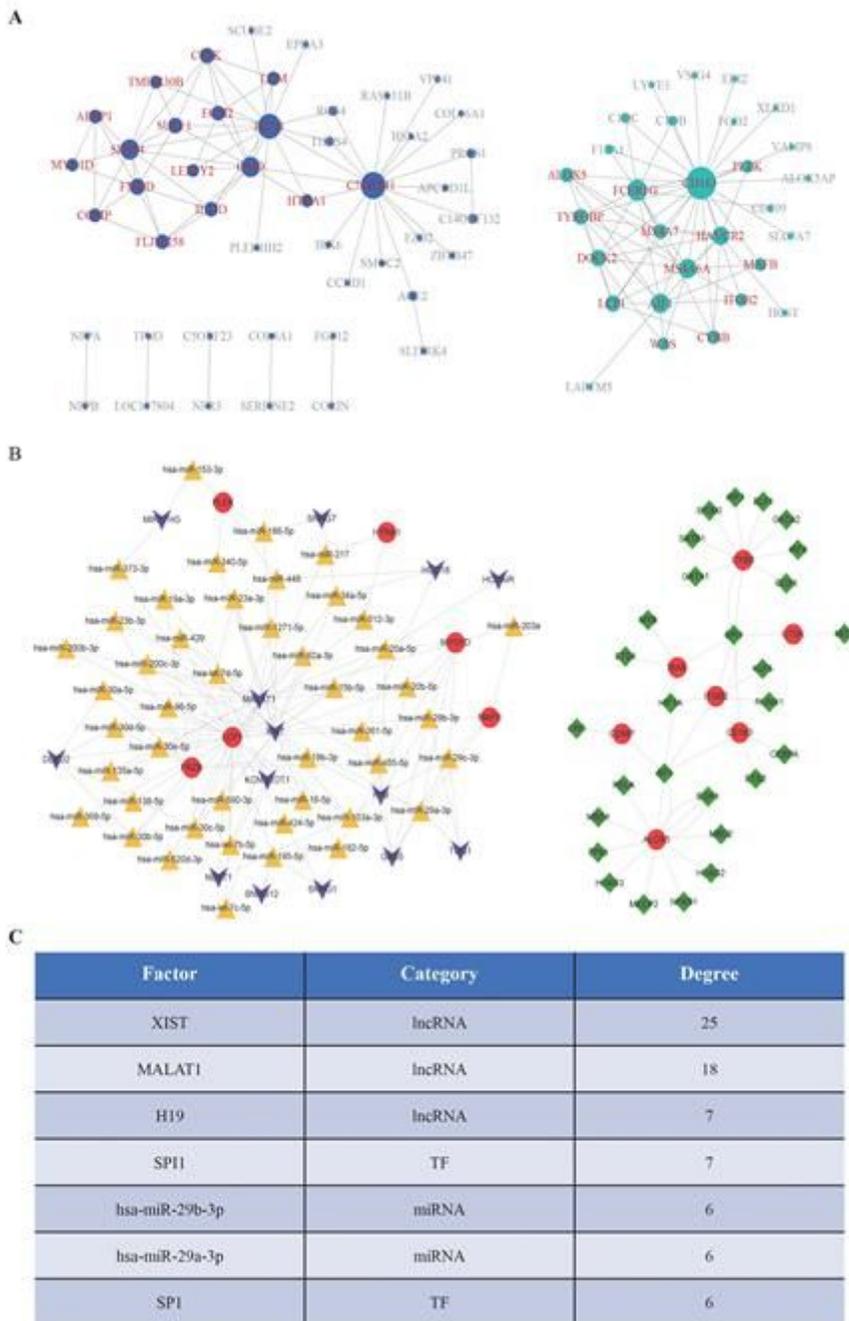
Identification of candidate gene set and function analysis A) Workflow used for bioinformatics analyses; B) DEGs in two datasets; C) Identification of candidate gene set; D) BP and KEGG pathway analyses of candidate gene set. BP: biological process; KEGG: Kyoto Encyclopedia of Genes and Genomes.



**Figure 2**

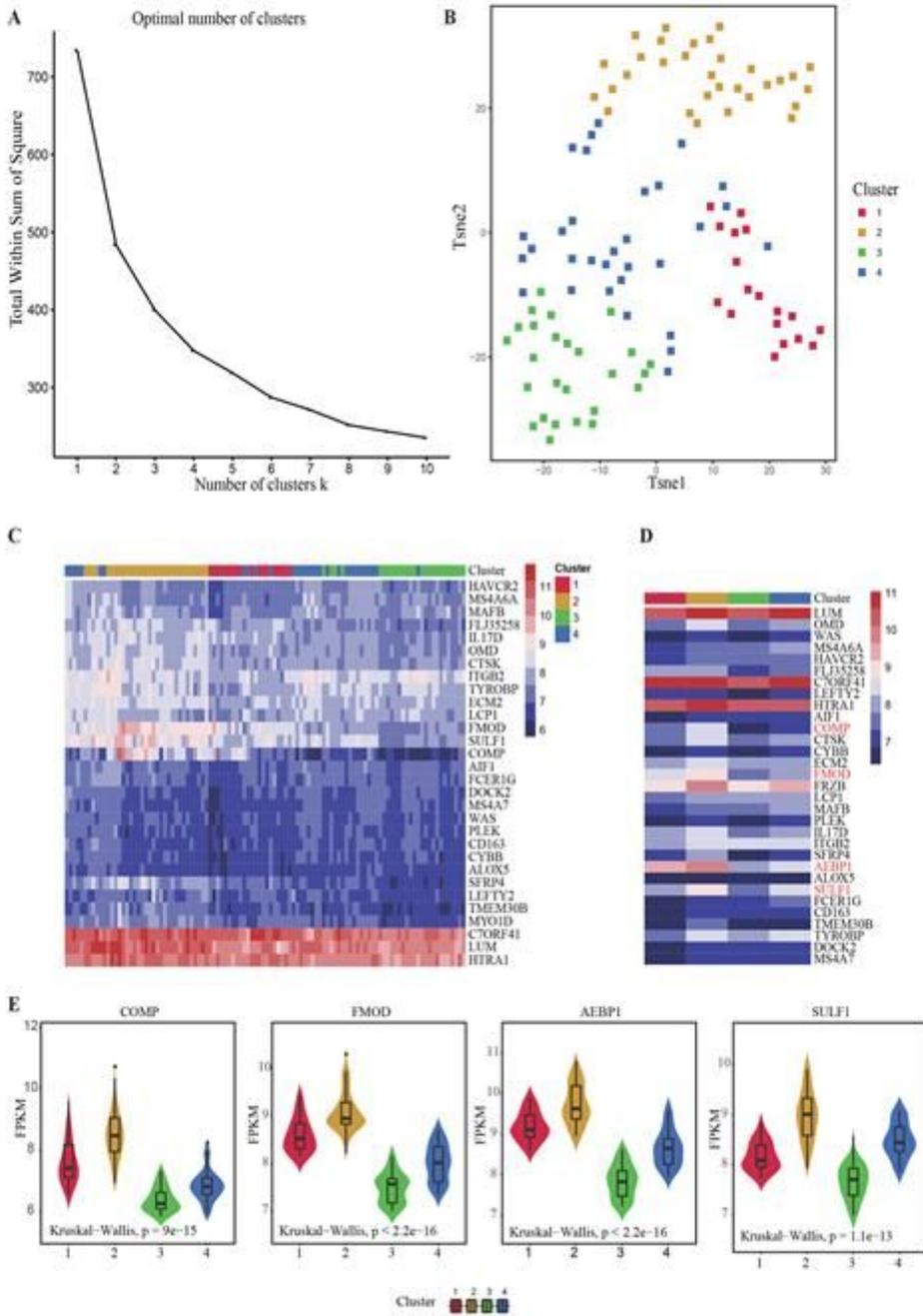
Overview of WGCNA network construction of the candidate gene set A) Identification of soft threshold; B) The candidate gene set was divided into 5 modules; C) The module-trait relationships of HOCM in 5 modules; D) Gene significance (GS) in the modules, in which the larger the GS score is, the larger the difference is.





**Figure 4**

Construction of multi-factor regulatory network A) Core genes in the blue module were 17 (left), and core genes in the turquoise module were 15 (right), core genes were labeled in the red color; and B) the construction of multi-factor regulatory network was done by Cytoscape software. The green rhombus represent transcription factor. The red circles represent genes (mRNAs). The orange triangles represent miRNAs. The purple arrows represent lncRNAs. C) Seven key regulatory factors were found in the network. The higher the degree is, the more important the regulatory function in the network is.



**Figure 5**

Clusters of HOCM. A)  $K = 4$  was selected as the optimal number of clusters since the  $K$  value is decreased by a negligible amount. B) tSNE algorithm provided each sample with a unique x- and y-coordinate (tSNE1 and tSNE2) according to each sample's gene expression of 32 core genes. All HOCM samples were clearly divided into 4 clusters; C) The expression of core genes in all HOCMs was shown in the heatmap; D) The expression of core genes in 4 clusters was shown in the heatmap; E) COMP, FMOD, AEBP1 and SULF1 showed significant expression in different clusters.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement10.xlsx](#)
- [supplement11.xlsx](#)
- [supplement12.xlsx](#)
- [supplement13.xlsx](#)
- [supplement14.xlsx](#)
- [supplement15.xlsx](#)
- [supplement16.xlsx](#)
- [supplement17.xlsx](#)
- [supplement18.xlsx](#)
- [supplement19.tif](#)
- [supplement20.tif](#)