

# Identification of Key Modules and Hub Genes in Hypertrophic Cardiomyopathy Based on Integrative Weighted Gene Co-expression Network Analysis

Jun Jiang (✉ [jiang-jun@zju.edu.cn](mailto:jiang-jun@zju.edu.cn))

Zhejiang University School of Medicine Second Affiliated Hospital <https://orcid.org/0000-0001-6926-9516>

Delong Chen

Zhejiang University School of Medicine Second Affiliated Hospital

Siyuan Xie

Zhejiang University School of Medicine Second Affiliated Hospital

Qichao Dong

Zhejiang University School of Medicine Second Affiliated Hospital

Yi Yu

Zhejiang University School of Medicine Second Affiliated Hospital

Yi Xu

Ningbo City First Hospital

Yuxuan Zhang

Zhejiang University School of Medicine Second Affiliated Hospital

---

## Research

**Keywords:** hypertrophic cardiomyopathy, WGCNA, key module, hub gene, GSEA

**Posted Date:** September 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-915958/v1>

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

[Read Full License](#)

---

# Abstract

## Background

Hypertrophic cardiomyopathy (HCM) is a heterogeneously inherited cardiac disorder with unclear biological pathogenesis. This study aims to identify the key modules and genes involved in the development of HCM.

## Methods

Using weighted gene co-expression network analysis (WGCNA) algorithm, we constructed integrative co-expression networks for the two large sample HCM datasets separately. After selecting clinically significant modules with the same clinical trait, functional enrichment analysis was performed to detect their common pathways. Based on the intramodular connectivity (IC), the shared hub genes were generated, validated, and further explored in gene set enrichment analysis (GSEA).

## Results

The orange and pink modules in GSE141910, the green and brown modules in GSE36961 were mostly related to HCM. Functional enrichment analysis suggested that HCM might exhibit enhanced processes including remodeling of extracellular matrix, activation of abnormal protein signaling, aggregation of calcium ion, and organization of cytoskeleton. SMOC2, COL16A1, RASL11B, TUBA3D, IL18R1 were defined as real hub genes due to their top IC values, significantly different expression levels, and excellent diagnostic performance in both datasets. Moreover, GSEA analysis demonstrated that pathways of the five hub genes were mainly involved in neuroactive ligand-receptor interaction, ECM-receptor interaction, Hedgehog signaling pathway.

## Conclusion

Our study provides more comprehensive insights into the molecular mechanisms of HCM, identifies five hub genes as candidate biomarkers for HCM, which might be theoretically feasible for targeted therapy against HCM.

## Background

Hypertrophic cardiomyopathy (HCM) is one of the most common cardiovascular genetic diseases featured by unexplained left ventricular hypertrophy and heterogeneous clinical phenotype which can be presented as asymptomatic, atrial fibrillation, syncope, heart failure, or even sudden death(1, 2). Despite the effective treatment and a favorable overall prognosis, HCM still remains a growing public health challenge affecting approximately 1 in 500 of the general population(3, 4). Genetic studies demonstrated that HCM is inherited in an autosomal dominant pattern with mutations in genes encoding sarcomere proteins. Genetic testing panels identified more than 1500 mutations in at least 29 genes(5). Some of the mutations have been reported to be pathogenic, pathogenicity is uncertain in others since they are not

easily distinguishable(6). In addition to the novel pharmacotherapies and minimally-invasive procedures, recent years also witnessed the emergence of gene-directed approaches which hold great therapeutic potential, including germline editing, high-fidelity gene repair, allele-specific gene slicing and so on(2). Identification of key genes and related expression networks in the pathogenesis and development of HCM was an important prerequisite for successful targeted therapy but remains a critical challenge.

Weighted gene co-expression network analysis (WGCNA) is a powerful bioinformatics analytical strategy that can be used to investigate gene expression patterns of multiple samples(7). Based on large-scale gene expression profiles such as high-throughput microarray or RNAseq dataset, WGCNA algorithm clusters highly correlated genes into co-expression modules which are considered to possess similar biological functions(8). Then comprehensive information is gained to evaluate the relationship between modules of interest and phenotypic traits. Furthermore, genes with highly connected nodes in a module are usually defined as hub genes, which could play crucial roles in the biological process of the module, or serve as represent representatives of its principal biological functions(9). At present, the application of WGCNA in the field of cardiovascular disease has provided various meaningful perspectives to explore and analyze them at molecular level(10-12).

In this study, by applying WGCNA, we aimed to identify clinically significant modules and hub genes, and further explore their functions in HCM. It is hoped that this study will provide novel insights into the molecular mechanisms of HCM and point out directions for precise treatments.

## Results

### Co-expression Network Construction

After data preprocessing, 17102, 37846 gene expression values were derived from GSE141910, GSE36961, respectively. We performed sample clustering to ensure the high quality of network construction. Finally, 56 samples (28 HCMs and 28 controls) in GSE141910 and 144 samples (105 HCMs and 39 controls) in GSE36961 were left for the following WGCNA analysis while an outlier was removed (Fig. 1a). To build scale-free networks, we set the soft threshold  $\beta = 18$  (scale free  $R^2 = 0.971$ ) for GSE141910,  $\beta = 18$  (scale free  $R^2 = 0.943$ ) for GSE36961 (Fig. 1b). Subsequently, based on the average linkage hierarchical clustering, genes with similar expression profiles were assigned to the same module (Fig. 1c, 1d). The gray module consisted of genes that cannot be clustered into the established modules.

### Identification of Clinically Significant Modules

Here, we identified 20 modules for GSE141910 and 31 modules for GSE36961 (Fig. 2a, 2b). Then, the correlation between the modules and the clinical traits was calculated. As shown in Fig. 2c, the orange ( $r = 0.86$ ,  $P < 0.01$ ) and pink ( $r = -0.89$ ,  $P < 0.01$ ) modules in GSE141910, the green ( $r = 0.81$ ,  $P < 0.01$ ) and brown ( $r = -0.85$ ,  $P < 0.01$ ) modules in GSE36961, were mostly associated with HCM and were identified as clinically significant modules.

## Functional Enrichment Analysis for Genes of Selected Modules

GO biological process (GO-BP) and KEGG enrichment analysis were performed to evaluate the functions of the four selected modules. Terms of GO-BP in the orange and green modules were combined to screen out the shared functional mechanisms of the two datasets. Meanwhile, terms of GO-BP in the pink and brown modules undergone the same processing. Given that relatively few KEGG pathways were enriched, three main ones of each module were extracted. Results were shown in Fig. 3a and 3b.

Specifically, in the orange and green module, the common GO-BP terms were mainly enriched in negative regulation of myoblast differentiation, extracellular matrix (ECM) organization, extracellular structure organization, regulation of presynapse organization, positive regulation of Ras protein signal transduction, regulation of Rho protein signal transduction, regulation of cytosolic calcium ion concentration, sodium ion transport, actin cytoskeleton organization, actin filament-based process; the common KEGG pathway was mainly associated with protein digestion and absorption. Whereas in the pink and brown module, the shared GO-BP terms were significantly related to positive regulation of macrophage derived foam cell differentiation, regulation of peroxisome proliferator activated receptor signaling pathway, macrophage cytokine production, epithelial to mesenchymal transition involved in endocardial cushion formation, ventricular septum morphogenesis, regulation of T-helper 1 type immune response, response to prostaglandin, response to fatty acid, positive regulation of reactive oxygen species metabolic process, ventricular cardiac muscle tissue development; the shared KEGG pathway was mainly involved in complement and coagulation cascades. These findings indicated that HCM patients may experience some unusual process including remodeling of ECM, activation of abnormal protein signaling, aggregation of calcium ion, and organization of cytoskeleton. Yet normal controls tended to exhibit more physiological, metabolic, and immunological functions.

## Hub Genes Identification

Since the selected modules were the most highly positive or negative related to HCM in GSE141910 and GSE36961, we extracted genes of the top 10% intramodular connectivity (IC) as candidate genes. Additionally, module membership (MM) and gene significance (GS) were calculated to obtain further insight into the relationships between module genes and HCM. Venn diagrams were performed and the overlapping genes were defined as hub genes. Subsequently, 3 genes in the positive-related modules (SMOC2, COL16A1, RASL11B), 6 genes in the negative-related modules (TUBA3D, S1PR3, TUBA3E, IL18R1, FKBP5, MYH6) were identified as hub genes (Fig. 3c and 3d, Table 1).

## Validation of Hub Genes

We verified the expression patterns and diagnostic performance of hub genes in both datasets. Firstly, we established Wilcoxon models to detect the differentially expressed hub genes in HCM patients and normal controls. As illustrated in Fig. 4a and 4b, the expression levels of positively-related hub genes were significantly increased in HCM samples, whereas all the negatively-related ones were significantly downregulated compared to normal samples. Moreover, we built logistic regression models and

performed receiver operating curve (ROC) analysis to further assess the diagnostic performance of hub genes. As shown in Fig. 4c, both in GSE141910 and in GSE36961, SMOC2, COL16A1, RASL11B, TUBA3D, IL18R1 manifested excellent diagnostic value as predictive biomarkers for HCM, thus they were identified as real hub genes (AUC: 1 and 0.969, 0.906 and 0.922, 0.927 and 0.894, 1 and 0.985, 0.962 and 0.948, respectively,  $P < 0.01$  or  $P = 1$ . Of note, both AUC = 1 and  $P = 1$  indicated this gene was only expressed in HCM samples).

### Gene Set Enrichment Analysis

To further assess the underlying mechanisms for the five real hub genes involved in HCM, gene set enrichment analysis (GSEA) was implemented. The mainly enriched KEGG pathway of each hub gene was shown in Fig. 5 and Table 2. SMOC2 and RASL11B were mainly involved in the 5 same pathways, including Hedgehog signaling pathway, Th1 and Th2 cell differentiation, ECM-receptor interaction, Protein digestion and absorption, Neuroactive ligand-receptor interaction. Whereas TUBA3D tended to be negative related to these major pathways. In addition, Neuroactive ligand-receptor interaction was also enriched in COL16A1 and IL18R1. These results suggested that the five real hub genes might act synergistically to a certain extent and played essential roles in HCM.

## Discussion

Despite tremendous advances in the management of HCM, there continues to be a global health burden. Current treatment strategies remain symptomatic and aim to reduce complications and prevent sudden cardiac death but are unable to reverse the disease. More targeted therapies are warranted but still insufficiency due to the genetic heterogeneity and unclear pathophysiology.

WGCNA method, whose clustering criterion has biological significance, is capable of identifying the relationships between the modules and disease phenotypes. Of note, the reliability of co-expression network is enhanced as the sample size grows larger. In the present study, we employed integrative WGCNA analysis on two large sample datasets downloaded from GEO. We constructed co-expression networks and detected several clinically significant modules related to the pathogenesis of HCM. Genes in the selected modules were sent to enrichment analysis and the shared module-trait matched genes with top 10% IC values were extracted and verified in expression level and ROC analysis. Finally, we identified 5 real hub genes and investigated their underlying molecular functions.

In the enrichment analysis of the orange module and green module, functional pathways are mainly associated with ECM, homeostasis of cations and movement of cardiomyocytes, revealing some pathological processes of HCM. ECM will be discussed in detail later. It is reported that pathogenic variants in sarcomere protein genes occur in approximately 40%–60% of HCM patients(13), greatly influencing the fundamental contractile units of cardiomyocytes which are composed of thick myosin and thin actin filaments. Furthermore, accumulating evidence has demonstrated the role of  $Ca^{2+}$  and  $Na^{+}$  homeostasis in the occurrence and development of HCM. The abnormal  $Ca^{2+}$  signaling pathways

and alterations in the sensitivity of myofilament to  $\text{Ca}^{2+}$  have been considered as triggers for arrhythmias. A single-cell level study showed that calcium mishandling was a basic modification for the pathogenesis of familial HCM and pharmacological retrieval of  $\text{Ca}^{2+}$  homeostasis blocked progress of hypertrophy(14). In addition, increased late  $\text{Na}^+$  current might be a major contributor to the electrophysiological disorders of HCM cardiomyocytes, implying potential therapeutic targets of its inhibitor(15). Therefore, our findings illustrate some underlying molecular mechanisms of HCM.

After validation and diagnostic evaluation, SMOC2, COL16A1, RASL11B, TUBA3D, IL18R1 are identified as real hub genes.

SMOC2 encodes matricellular proteins that belong to secreted proteins of the SPARC family which can promote matrix assembly thus playing an important role in regulating cell-matrix interactions. Previous studies demonstrated that SMOC2 knockout or targeted by siRNA exhibited decreased inflammatory reaction and attenuated TGF- $\beta$ 1-induced fibroblast-to-myofibroblast transition, suppressing the progression of pulmonary or kidney fibrosis(16, 17). Moreover, Laugier et al. indicated that SMOC2 was upregulated in chronic Chagas disease cardiomyopathy by participating in matrix remodeling and potentially involved in the inflammatory damage in heart(18). Thereby, SMOC2, serves as a vital factor that alters matrix remodeling and inflammation response, might be a potential therapeutic target for HCM.

COL16A1, encoding fibril-associated collagen, is typically considered to regulate fibril-formation and extracellular matrix organization. It has been suggested that TGF- $\beta$  behaved as a modulator of collagen XVI and regulated COL16A1 gene and protein expression through Smad-signaling(19). Particularly, expression of COL16A1 at mRNA level and protein level was found significantly increased in dilated cardiomyopathy patients, with TGF- $\beta$ 1 significantly upregulated and related to COL16A1(20). Additionally, with a similar expression pattern, COL16A1 was associated with left ventricular dysfunction in ischemic cardiomyopathy patients(21). Thus, COL16A1 may facilitate the development of cardiac fibrosis and remodeling, leading to the pathological process of heart failure in HCM.

RASL11B belongs to the small GTPase family and is highly similar to RAS proteins. RASL11B was demonstrated to be involved in pathophysiology processes of inflammation and arteriosclerosis through TGF- $\beta$ 1 mediated pathways(22). Upregulation of RASL11B was found to inhibit cell growth and induce apoptosis in lung adenocarcinoma(23). Furthermore, it has been reported that in Huntington's disease, epigenetic dysregulation might result in the increased susceptibility to transcriptional dysregulation of RASL11B(24). These studies suggest that RASL11B may be associated with the development of HCM and offer novel insight into the mechanism studying of HCM.

TUBA3D belongs to the alpha tubulin family and modulates microtubule cytoskeleton organization. Mutant in TUBA3D was found to be associated with cardiac arrhythmia in sudden unexpected death in epilepsy patients(25). And TUBA3D is also linked with cardiovascular disorders, atrial and ventricular

septal defects commonly reported in Patau Syndrome(26). HCM has been well studied in variants of sarcomere genes, whereas the role of TUBA3D in HCM needs further exploration.

IL18R1, belonging to the interleukin 1 receptor family, encodes a specific cytokine receptor of interleukin 18 (IL18). Generally, IL18 is a proinflammatory cytokine implicated in the injury of different organs. Tirt and colleagues revealed that IL18R1 polymorphisms showed no linkage with IL-18 levels, but were linked with cardiovascular mortality both in single-locus analysis and in haplotype analysis(27). Similarly, it is reported that polymorphic variants of IL18R1 are associated with high risk of myocardial infarction, and minor alleles of IL18R1 are related to high risk of occurrence of multifocal atherosclerosis and arterial hypertension in stable coronary heart disease patients after myocardial infarction(28). Therefore, we hypothesize that in HCM patients, mutation of IL18R1 might lead to a decrease in its detected expression level and whether IL18R1 alters the cardiovascular risk of HCM remains to be verified.

Furthermore, GSEA analysis reveals several common pathways of the five hub genes. The neuroactive ligand-receptor interaction pathway has been highly enriched in all real hub genes and shows significant negative correlations in TUBA3D and IL18R1, which were both downregulated in HCM samples, suggesting its important role in the pathological process of HCM via altered cell signaling. Proteins involved in this pathway including calcitonin receptor-like receptor, adrenergic receptor, angiotensin receptor are closely related to cardiac function(29-31), while the deeper correlation between this pathway and HCM needs to be further explored. The ECM-receptor interaction pathway, which behaves as specific interactions between cells and ECM that mainly contains glycoproteins, proteoglycans, glycosaminoglycans and proteases, has been greatly enriched in SMOC2, RASL11B, and TUBA3D. GO enrichment analysis above also showed great importance of ECM involved in HCM. Previous studies have well demonstrated that excessive accumulation of ECM impaired cardiac contractile function and emphasized the role of myocardial fibrosis in the morphogenesis and establishment of HCM(32, 33). Abnormal expression of ECM components like Fibulin-2 and TXNDC5 might augment detrimental interactions in HCM(34, 35) and HCM-related matrix undermined healthy cardiomyocytes with prolonged contractions and poor relaxation(36). Additionally, TGF-beta signaling pathway, which is relatively negative enriched in TUBA3D, participates in pathogenesis and development of HCM. Malgija et al. indicated that Smad-dependent TGFβ1 pathway was depicted as collagen-gene promoter thus causing cardiac fibrosis(33). Hedgehog signaling pathway, which possesses numerous roles since embryonic stage and keeps active in the adult, has been enriched in all real hub genes except COL16A1. Hedgehog signaling pathway was reported to be involved in kidney injury and fibrosis, genetic ablation of specific perivascular mesenchymal stem cells that drive this signaling ameliorated heart and kidney fibrosis and maintained ejection fraction in heart failure model(37). Therefore, our study hypothesizes that abnormally activated Hedgehog signaling pathway is associated with fibrosis of HCM, further investigation into its role in HCM is required. Briefly, GSEA results correspond to the previous functional enrichment findings of selected modules, confirming the key pathological role of the up-expressed genes SMOC2, COL16A1, and RASL11B and the potential risk factor of the down-expressed genes of TUBA3D and IL18R1 in HCM. Our research might provide promising biomarkers for the occurrence and progression of HCM.

Besides, researches have indicated sexual dimorphisms in clinical presentation and prognosis of HCM(38, 39), but our results do not exhibit a significant correlation in this regard, implicating further explorations might gain insights on methylated or proteomics level.

The study has several limitations. First, we only chose modules with the highest correlation in the two co-expression networks for analysis, which might lead to the omission of genes in other highly related modules. Second, the ROC analysis of hub genes requires tests in other datasets, whereas our attempts are failed because of the limited sample size of public database. Third, the identified hub genes need to be validated in experiments with animal models and human samples.

## Conclusion

In summary, our study revealed key modules and underlying biological mechanisms of HCM, clarified the expression levels, diagnostic values, and molecular functions of SMOC2, COL16A1, RASL11B, TUBA3D, IL18R1, thus providing new perspectives on the development and targeted therapy of HCM.

## Methods

### Data Collection and Proccession

The flowchart of this study is shown in Figure 6. We searched the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) for high-throughput sequencing data of HCM. Inclusion criteria were as follows: (1) datasets including HCM patients and healthy controls, (2) samples of human left ventricle tissues, (3) samples containing total RNA sequence, (4) total sample size was greater than 50. Finally, the RNA expression profiles of GSE141910 and GSE36961 were downloaded for analysis.

The RNA-seq dataset GSE141910 contained samples from patients with cardiomyopathies and healthy donors. We extracted the expression profiling data of 28 HCM patients and 28 healthy donors. The microarray dataset GSE36961 included samples from 106 HCM patients and 39 healthy controls. The batch effect was eliminated using the “SVR” package. GSE141910 and GSE36961 were normalized using the “limma” package(40). Clinical features including disease status and gender were extracted for subsequent investigation.

### Co-expression Network Construction

The R package “WGCNA” was applied to construct co-expression networks(7). GSE141910 and GSE36961 were used to build co-expression networks separately. Pearson’s correlation matrices were built after evaluating the correlation coefficients between the sample genes, and outlier genes were discarded(41). We constructed weighted adjacency matrices using the power formula as follows:

$$s_{mn} = |\text{cor}(x_m, x_n)|^\beta$$

Where  $x_m$  and  $x_n$  are expression values of gene m and gene n, cor represents the Pearson's correlation coefficient,  $s_{mn}$  represents the adjacency coefficient between the two genes, and  $\beta$  means the soft threshold calculated by network topology analysis.

Then we generated topological overlap matrix (TOM), which quantitatively evaluated the connectivity of a gene in the network by comparing the weighted correlation with other genes. Based on the TOM matrix, we used average linkage hierarchical clustering and the dynamic tree-cutting algorithm to define and visualize gene modules with similar expression patterns(42). The minimum module size was set as 10 and similar modules were merged with a threshold of 0.25. In addition, modules of highly correlated genes were visualized with the R software and were further explored to detect their potential role positions.

### **Identification of Clinically Significant Modules**

The co-expression module contains genes with high topological overlap similarity. In this study, all modules were summarized by module eigengenes (MEs), which were considered as the major principal component of each module, and were used to describe the overall expression level of the given module(43). MM represented the correlation between individual gene and ME, and was used to describe the reliability of gene in the module(44). GS was defined as the correlation between genes and traits. MM and GS were calculated to further identify clinically significant modules of each dataset.

### **Functional Enrichment Analysis**

To gather further insight into the potential functional mechanisms of the selected modules, we uploaded the data into "https://metascape.org/" to conduct the GO and KEGG enrichment analysis(45-47). The cut-off criterion was set at adjusted  $P < 0.05$ . The findings were illustrated as bubble charts using the "ggplot2" R package.

### **Identification of Hub Genes**

Hub genes are defined as genes with highly IC values in the module calculated by the WGCNA algorithm and have been regarded functionally important. The higher the IC, the more considerable its role in the module, the greater its association with a disease. In this study, we screened out genes with the top 10% IC values in each module as candidate genes, then candidate genes, which possessed the same clinical trait from two datasets, were depicted as Venn diagram(48). The overlapped genes were defined as hub genes and were visualized by Cytoscape3.8.2(49).

### **Validation of Hub Genes**

After extracting the hub genes, the expression level of hub genes in HCM patients and normal controls was verified in the two datasets by building Wilcoxon regression models using "glmnet" R package. Moreover, to evaluate the diagnostic performance of hub genes, ROC analysis was conducted by applying the R package "pROC"(50). Then real hub genes were recognized and sent to GSEA analysis.

## **GSEA**

In the RNA-seq dataset GSE141910, HCM patients were grouped according to high and low real hub gene expression levels. GSEA for each real hub gene was performed using R package “clusterprofiler”(51). The “c2.cp.kegg.v7.4.symbols.gmt” from Molecular Signatures Database was chosen as reference gene sets. Terms with adjusted P value of < 0.05 were considered significant. Major KEGG pathways of each gene were extracted to detect their potential function.

## **Abbreviations**

HCM, Hypertrophic cardiomyopathy; WGCNA, Weighted gene co-expression network analysis; BP, biological process; ECM, extracellular matrix; IC, intramodular connectivity; MM, Module membership; GS, Gene significance; ROC, receiver operating curve; GSEA, Gene set enrichment analysis; TOM, topological overlap matrix; ME, module eigengene.

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable

### **Consent for publication**

Not applicable

### **Availability of data and materials**

The datasets generated during the current study are available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141910> and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36961>).

### **Competing interests**

The authors declare that they have no competing interests

### **Funding**

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### **Authors' contributions**

DC, SX, and JJ designed the study; SX performed the analysis; DC, QD, YY, YX, and YZ wrote the manuscript and DC was the major contributor. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by the Key Research and Development Program of Zhejiang Province (NO. 2020C03016).

## References

1. Maron BJ. Clinical Course and Management of Hypertrophic Cardiomyopathy. *The New England journal of medicine*. 2018;379(7):655-68.
2. Tuohy CV, Kaul S, Song HK, Nazer B, Heitner SB. Hypertrophic cardiomyopathy: the future of treatment. *European journal of heart failure*. 2020;22(2):228-40.
3. McKenna WJ, Judge DP. Epidemiology of the inherited cardiomyopathies. *Nature reviews Cardiology*. 2021;18(1):22-36.
4. Maron BJ, Ommen SR, Semsarian C, Spirito P, Olivetto I, Maron MS. Hypertrophic cardiomyopathy: present and future, with translation into contemporary cardiovascular medicine. *Journal of the American College of Cardiology*. 2014;64(1):83-99.
5. Wilcox JE, Hershberger RE. Genetic cardiomyopathies. *Current opinion in cardiology*. 2018;33(3):354-62.
6. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al. Genetic Misdiagnoses and the Potential for Health Disparities. *The New England journal of medicine*. 2016;375(7):655-65.
7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008;9:559.
8. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings in bioinformatics*. 2018;19(4):575-92.
9. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature reviews Genetics*. 2004;5(2):101-13.
10. Westerman K, Sebastiani P, Jacques P, Liu S, DeMeo D, Ordovás JM. DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure. *Clinical epigenetics*. 2019;11(1):142.
11. Guarnieri AR, Anthony SR, Gozdiff A, Green LC, Fleifil SM, Slone S, et al. Adipocyte-specific deletion of HuR induces spontaneous cardiac hypertrophy and fibrosis. *American journal of physiology Heart and circulatory physiology*. 2021;321(1):H228-h41.

12. Duan C, Cao Z, Tang F, Jian Z, Liang C, Liu H, et al. miRNA-mRNA crosstalk in myocardial ischemia induced by calcified aortic valve stenosis. *Aging*. 2019;11(2):448-66.
13. Ho CY, Charron P, Richard P, Girolami F, Van Spaendonck-Zwarts KY, Pinto Y. Genetic advances in sarcomeric cardiomyopathies: state of the art. *Cardiovascular research*. 2015;105(4):397-408.
14. Lan F, Lee AS, Liang P, Sanchez-Freire V, Nguyen PK, Wang L, et al. Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells. *Cell stem cell*. 2013;12(1):101-13.
15. Coppini R, Ferrantini C, Yao L, Fan P, Del Lungo M, Stillitano F, et al. Late sodium current inhibition reverses electromechanical dysfunction in human hypertrophic cardiomyopathy. *Circulation*. 2013;127(5):575-84.
16. Luo L, Wang CC, Song XP, Wang HM, Zhou H, Sun Y, et al. Suppression of SMOC2 reduces bleomycin (BLM)-induced pulmonary fibrosis by inhibition of TGF- $\beta$ 1/SMADs pathway. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2018;105:841-7.
17. 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).
18. Laugier L, Frade AF, Ferreira FM, Baron MA, Teixeira PC, Cabantous S, et al. Whole-Genome Cardiac DNA Methylation Fingerprint and Gene Expression Analysis Provide New Insights in the Pathogenesis of Chronic Chagas Disease Cardiomyopathy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2017;65(7):1103-11.
19. Grassel S, Bauer RJ. Collagen XVI in health and disease. *Matrix biology : journal of the International Society for Matrix Biology*. 2013;32(2):64-73.
20. Gil-Cayuela C, Roselló LE, Ortega A, Tarazón E, Triviño JC, Martínez-Dolz L, et al. New Altered Non-Fibrillar Collagens in Human Dilated Cardiomyopathy: Role in the Remodeling Process. *PLoS One*. 2016;11(12):e0168130.
21. Gil-Cayuela C, Rivera M, Ortega A, Tarazón E, Triviño JC, Lago F, et al. RNA sequencing analysis identifies new human collagen genes involved in cardiac remodeling. *Journal of the American College of Cardiology*. 2015;65(12):1265-7.
22. Stolle K, Schnoor M, Fuellen G, Spitzer M, Cullen P, Lorkowski S. Cloning, genomic organization, and tissue-specific expression of the RASL11B gene. *Biochimica et biophysica acta*. 2007;1769(7-8):514-24.
23. Xu B, Qian Y, Hu C, Wang Y, Gao H, Yang J. Dexmedetomidine upregulates the expression of miR-493-5p, inhibiting growth and inducing the apoptosis of lung adenocarcinoma cells by targeting RASL11B. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. 2021;99(4):457-64.

24. Guiretti D, Sempere A, Lopez-Atalaya JP, Ferrer-Montiel A, Barco A, Valor LM. Specific promoter deacetylation of histone H3 is conserved across mouse models of Huntington's disease in the absence of bulk changes. *Neurobiology of disease*. 2016;89:190-201.
25. Friedman D, Kannan K, Faustin A, Shroff S, Thomas C, Heguy A, et al. Cardiac arrhythmia and neuroexcitability gene variants in resected brain tissue from patients with sudden unexpected death in epilepsy (SUDEP). *NPJ genomic medicine*. 2018;3:9.
26. Abuzenadah A, Al-Saedi S, Karim S, Al-Qahtani M. Role of Overexpressed Transcription Factor FOXO1 in Fatal Cardiovascular Septal Defects in Patau Syndrome: Molecular and Therapeutic Strategies. *International journal of molecular sciences*. 2018;19(11).
27. Tiret L, Godefroy T, Lubos E, Nicaud V, Tregouet DA, Barbaux S, et al. Genetic analysis of the interleukin-18 system highlights the role of the interleukin-18 gene in cardiovascular disease. *Circulation*. 2005;112(5):643-50.
28. Ponasenko AV, Tsepokina AV, Khutornaya MV, Sinitsky MY, Barbarash OL. IL18-family Genes Polymorphism Is Associated with the Risk of Myocardial Infarction and IL18 Concentration in Patients with Coronary Artery Disease. *Immunological investigations*. 2021:1-15.
29. Roh J, Chang CL, Bhalla A, Klein C, Hsu SY. Intermedin is a calcitonin/calcitonin gene-related peptide family peptide acting through the calcitonin receptor-like receptor/receptor activity-modifying protein receptor complexes. *The Journal of biological chemistry*. 2004;279(8):7264-74.
30. Wang Y, Yu X, Wang F, Wang Y, Wang Y, Li H, et al. Yohimbine promotes cardiac NE release and prevents LPS-induced cardiac dysfunction via blockade of presynaptic  $\alpha$ 2A-adrenergic receptor. *PLoS One*. 2013;8(5):e63622.
31. Vázquez-Medina JP, Popovich I, Thorwald MA, Viscarra JA, Rodriguez R, Sonanez-Organis JG, et al. Angiotensin receptor-mediated oxidative stress is associated with impaired cardiac redox signaling and mitochondrial function in insulin-resistant rats. *American journal of physiology Heart and circulatory physiology*. 2013;305(4):H599-607.
32. 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. *Journal of molecular and cellular cardiology*. 2013;62:164-78.
33. Malgija B, Kumar NS, Piramanayagam S. Collective transcriptomic deregulation of hypertrophic and dilated cardiomyopathy - Importance of fibrotic mechanism in heart failure. *Computational biology and chemistry*. 2018;73:85-94.
34. Shih YC, Chen CL, Zhang Y, Mellor RL, Kanter EM, Fang Y, et al. Endoplasmic Reticulum Protein TXNDC5 Augments Myocardial Fibrosis by Facilitating Extracellular Matrix Protein Folding and Redox-

Sensitive Cardiac Fibroblast Activation. *Circulation research*. 2018;122(8):1052-68.

35. Ibrahim AM, Roshdy M, Elshorbagy S, Hosny M, Halawa S, Yehia D, et al. An Investigation of Fibulin-2 in Hypertrophic Cardiomyopathy. *International journal of molecular sciences*. 2020;21(19).
36. Sewanan LR, Schwan J, Kluger J, Park J, Jacoby DL, Qyang Y, et al. Extracellular Matrix From Hypertrophic Myocardium Provokes Impaired Twitch Dynamics in Healthy Cardiomyocytes. *JACC Basic to translational science*. 2019;4(4):495-505.
37. Kramann R, Schneider RK, DiRocco DP, Machado F, Fleig S, Bondzie PA, et al. Perivascular Gli1+ progenitors are key contributors to injury-induced organ fibrosis. *Cell stem cell*. 2015;16(1):51-66.
38. Rowin EJ, Maron MS, Wells S, Patel PP, Koethe BC, Maron BJ. Impact of Sex on Clinical Course and Survival in the Contemporary Treatment Era for Hypertrophic Cardiomyopathy. *Journal of the American Heart Association*. 2019;8(21):e012041.
39. Geske JB, Ong KC, Siontis KC, Hebl VB, Ackerman MJ, Hodge DO, et al. Women with hypertrophic cardiomyopathy have worse survival. *European heart journal*. 2017;38(46):3434-40.
40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015;43(7):e47.
41. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*. 2005;4:Article17.
42. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. *Science (New York, NY)*. 2002;297(5586):1551-5.
43. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC systems biology*. 2007;1:54.
44. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS computational biology*. 2008;4(8):e1000117.
45. Zhou YY, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*. 2019;10.
46. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*. 2016;44(D1):D457-62.
47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000;25(1):25-9.
48. Chen T, Zhang H, Liu Y, Liu YX, Huang L. EVenN: Easy to create repeatable and editable Venn diagrams and Venn networks online. *Journal of genetics and genomics = Yi chuan xue bao*. 2021.

49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498-504.
50. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*. 2011;12:77.
51. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012;16(5):284-7.

## Tables

**Table 1** Characteristics of shared hub genes in the selected modules

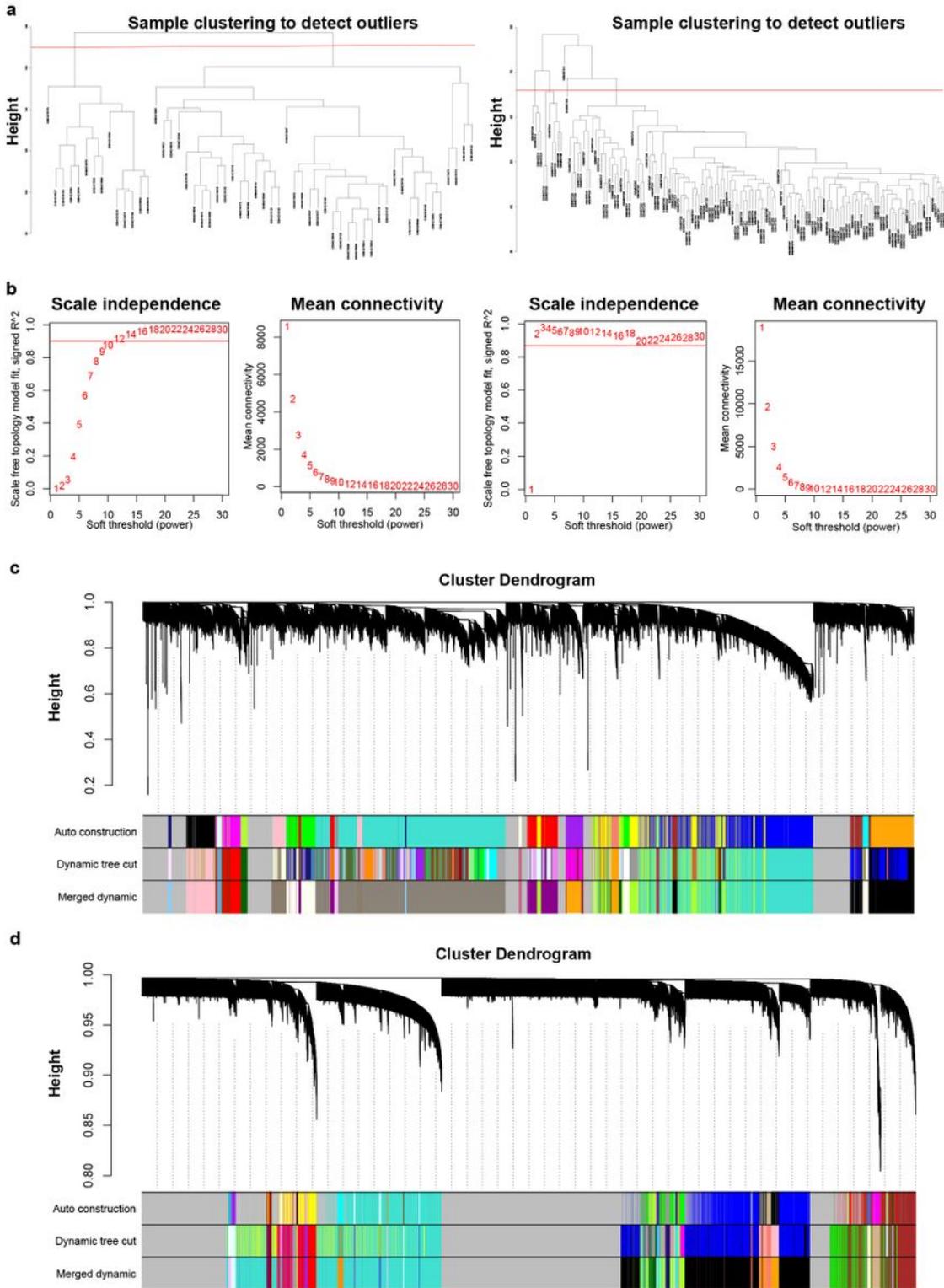
Gene	Intramodular connectivity		Module membership		Gene significance	
	GSE141910	GSE36961	GSE141910	GSE36961	GSE141910	GSE36961
SMOC2	99.386	8.780	0.919	0.750	0.901	0.707
COL16A1	71.615	6.681	0.788	0.687	0.709	0.609
RASL11B	69.249	9.731	0.849	0.771	0.754	0.698
TUBA3D	36.506	27.587	0.872	0.756	-0.902	-0.822
S1PR3	33.624	32.465	0.885	0.808	-0.892	-0.793
TUBA3E	32.658	31.289	0.841	0.772	-0.895	-0.845
IL18R1	31.438	33.720	0.854	0.810	-0.865	-0.716
FKBP5	28.570	25.688	0.814	0.750	-0.857	-0.754
MYH6	27.311	31.908	0.836	0.812	-0.903	-0.835

**Table 2** Results of gene set enrichment analysis for real hub genes

Gene	KEGG pathway	Enrichment score	NES	Adjust P	FDR	Leading edge
SMOC2	Hedgehog signaling pathway	0.653	1.946	0.002	0.002	tags=33%, list=15%, signal=28%
	Th1 and Th2 cell differentiation	0.547	1.746	0.014	0.011	tags=38%, list=16%, signal=32%
	ECM-receptor interaction	0.536	1.715	0.024	0.019	tags=29%, list=9%, signal=26%
	Protein digestion and absorption	0.529	1.681	0.032	0.025	tags=36%, list=13%, signal=31%
	Neuroactive ligand-receptor interaction	0.479	1.683	0.004	0.003	tags=34%, list=12%, signal=31%
COL16A1	Type I diabetes mellitus	0.665	2.150	0.001	0.000	tags=76%, list=27%, signal=55%
	Antigen processing and presentation	0.492	1.775	0.041	0.035	tags=57%, list=31%, signal=39%
	Neuroactive ligand-receptor interaction	0.385	1.592	0.029	0.024	tags=20%, list=8%, signal=19%
RASL11B	ECM-receptor interaction	0.631	1.968	0.000	0.000	tags=49%, list=19%, signal=40%
	Protein digestion and absorption	0.616	1.909	0.002	0.002	tags=40%, list=13%, signal=35%
	Hedgehog signaling pathway	0.585	1.710	0.038	0.033	tags=27%, list=15%, signal=23%
	Th1 and Th2 cell differentiation	0.548	1.712	0.025	0.022	tags=36%, list=14%, signal=31%
	Neuroactive ligand-receptor interaction	0.442	1.509	0.046	0.040	tags=37%, list=15%, signal=32%
TUBA3D	Hedgehog signaling pathway	-0.660	-1.977	0.003	0.002	tags=33%, list=15%, signal=29%

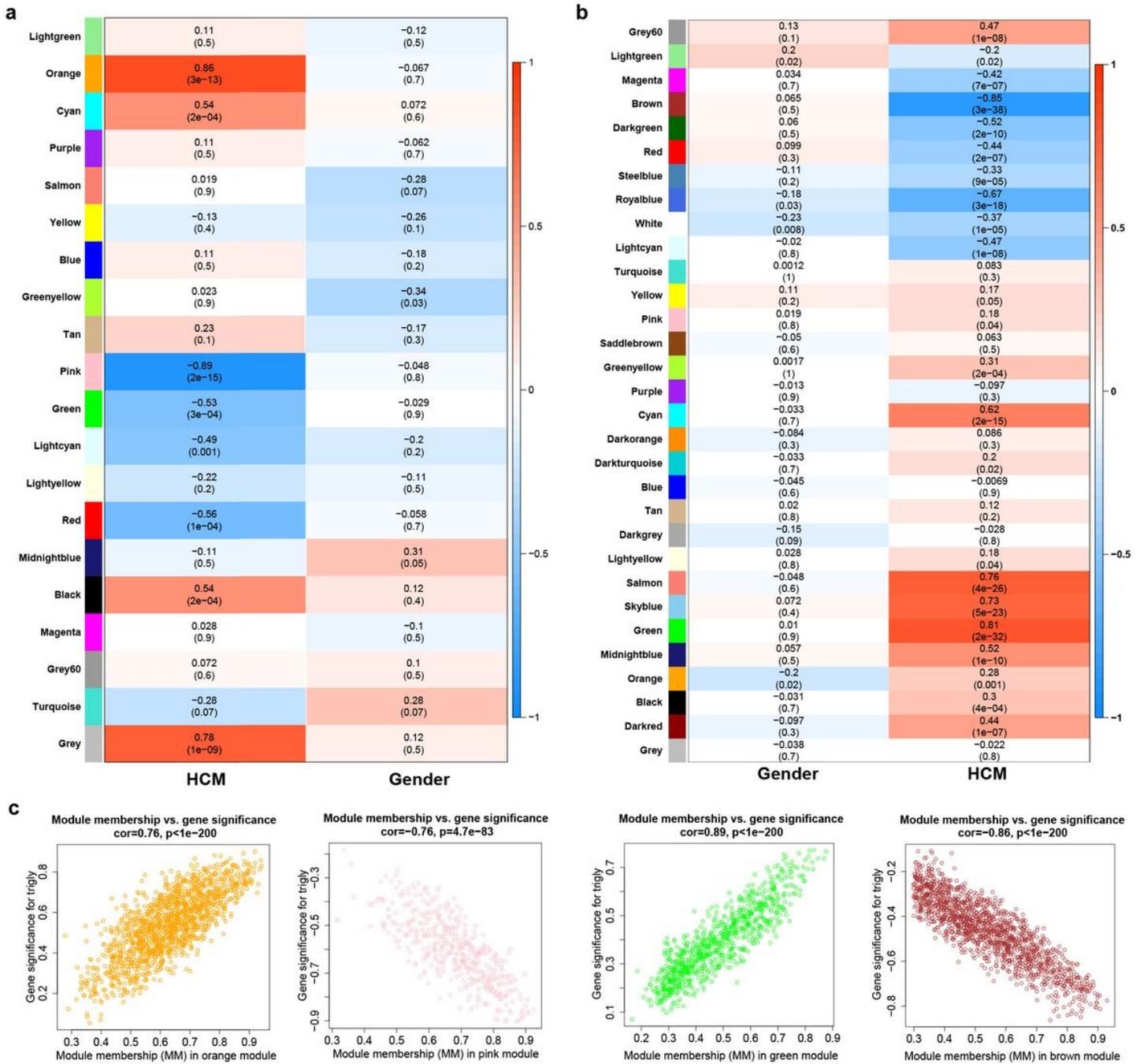
	Protein digestion and absorption	-0.575	-1.799	0.010	0.009	tags=42%, list=14%, signal=37%
	ECM-receptor interaction	-0.567	-1.784	0.010	0.009	tags=30%, list=10%, signal=27%
	TGF-beta signaling pathway	-0.517	-1.646	0.026	0.022	tags=22%, list=13%, signal=19%
	Neuroactive ligand-receptor interaction	-0.494	-1.708	0.003	0.002	tags=35%, list=11%, signal=31%
IL18R1	Type I diabetes mellitus	-0.642	-1.892	0.022	0.018	tags=64%, list=15%, signal=54%
	Hedgehog signaling pathway	-0.582	-1.855	0.017	0.013	tags=31%, list=13%, signal=27%
	Neuroactive ligand-receptor interaction	-0.405	-1.553	0.028	0.023	tags=32%, list=13%, signal=28%

## Figures



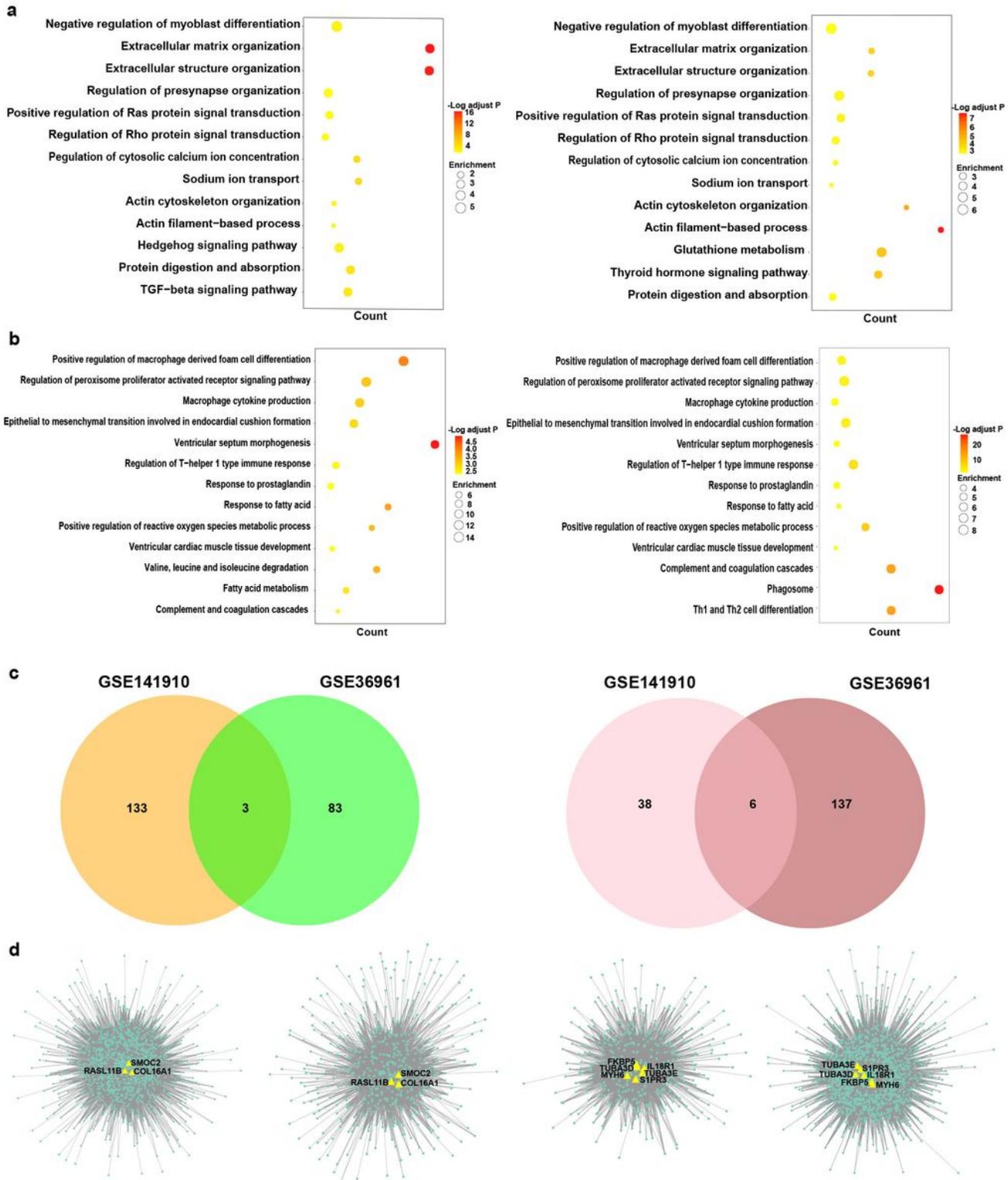
**Figure 1**

Construction of weighted co-expression networks in the two datasets. (a) Sample clustering to detect outliers and for GSE141910 (left), GSE36961 (right), and outlier GSM907212 was removed. (b) The soft threshold power and the scale-free fit index for GSE141910 (left), GSE36961 (right). (c, d) Dendrograms of GSE141910, GSE36961, respectively.



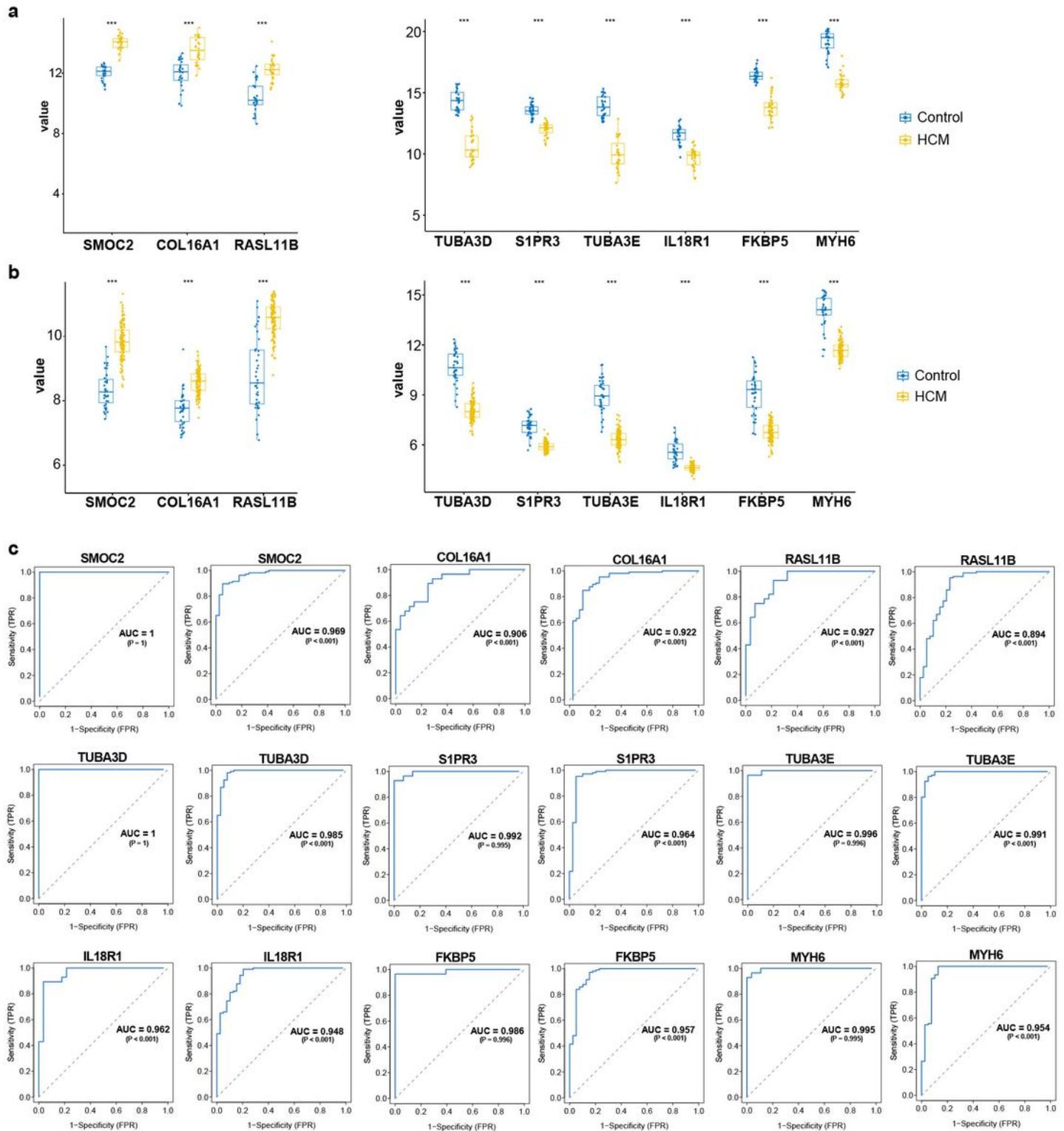
**Figure 2**

Identification of clinically significant modules. (a, b) Correlation heatmaps between modules and clinical traits for GSE141910, GSE36961, respectively. Each unit contained correlation coefficient and corresponding P value). (c) Module membership versus gene significance plot of orange and pink module, green and brown module, respectively.



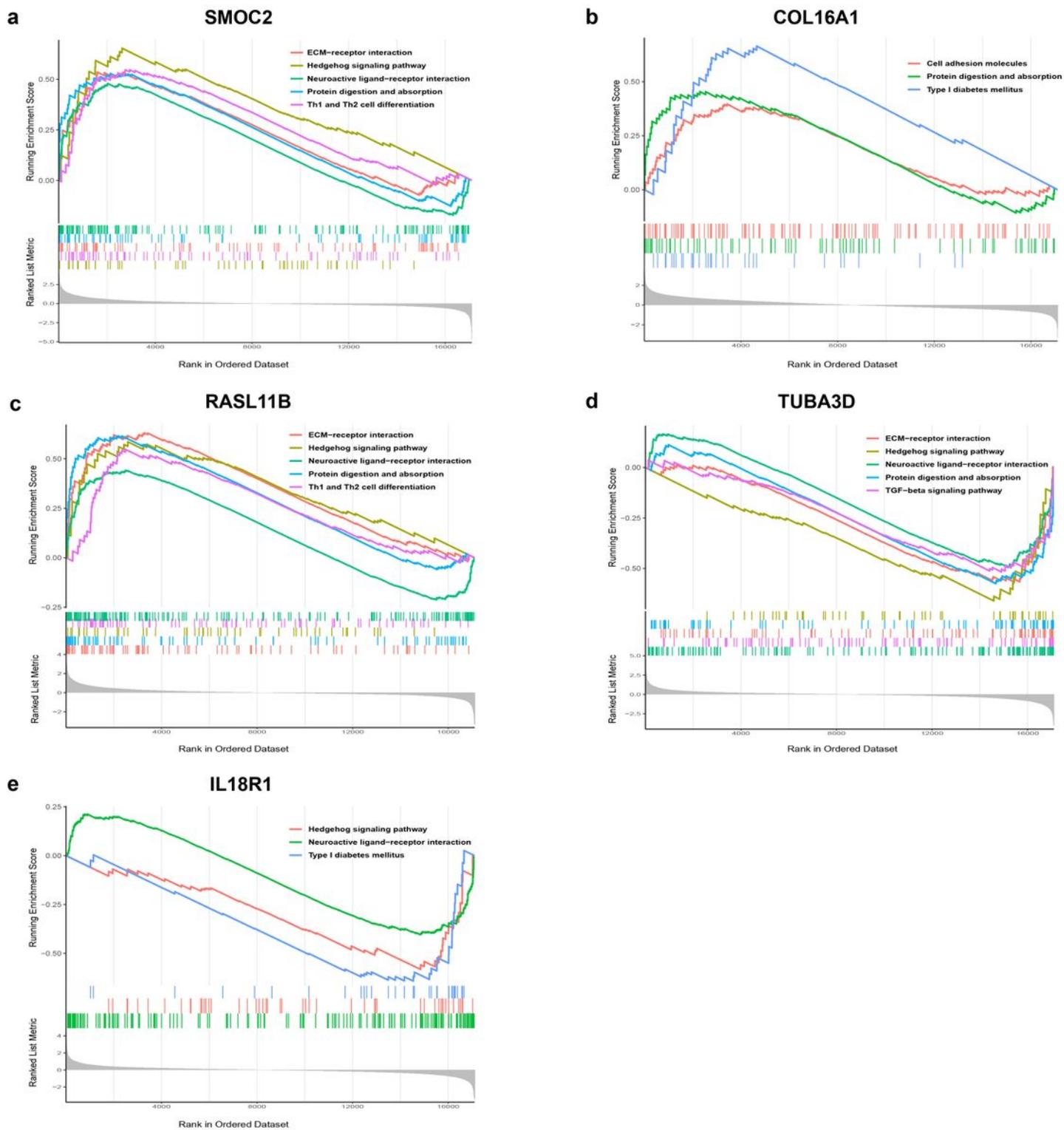
**Figure 3**

Functional enrichment analysis and detection of hub genes. (a) Shared GO biology processes (the top ten) and mainly KEGG pathways (the bottom three) for module orange, green. (b) Shared GO biology processes (the top ten) and mainly KEGG pathways (the bottom three) for pink, and brown. (c) Venn diagram to detect shared hub genes of orange and green module (left), pink and brown module (right). (d) Interaction networks in orange, green, pink, and brown module.



**Figure 4**

Validation of Hub Genes. (a, b) Differential expression of hub genes in GSE141910, GSE36961, respectively, \*\*\* means P value < 0.001. (c) ROC analysis of corresponding hub genes in GSE141910, GSE36961, respectively. Abbreviations: AUC, area under the curve; ROC, receiver operating characteristic.



**Figure 5**

Gene set enrichment analysis of real hub genes. (a) SMOC2, (b) COL16A1, (c) RASL11B, (d) TUBA3D, (e) IL18R1.

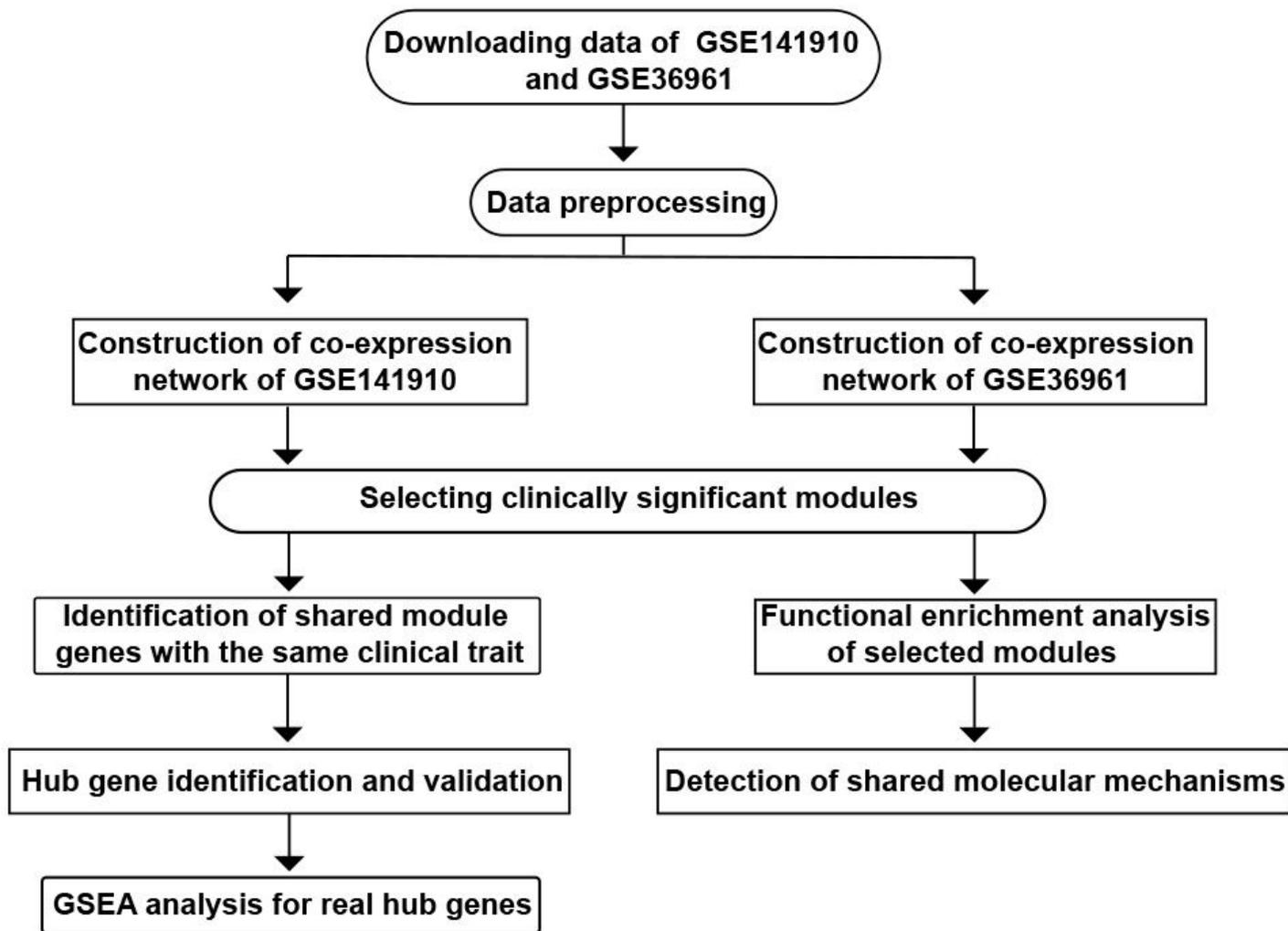


Figure 6

Flow diagram of the whole study.