RNA-seq profiling of mRNA associated with hypertrophic cardiomyopathy

CHANG-WEI REN, JIA-JI LIU, JIN-HUA LI, JING-WEI LI, JIANG DAI and YONG-QIANG LAI

Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing 100029, P.R. China

Received September 9, 2015; Accepted September 20, 2016

DOI: 10.3892/mmr.2016.5931

Abstract. The mechanisms of hypertrophic cardiomyopathy (HCM) pathogenesis can be investigated by determining the differences between healthy and disease states at the molecular level. In the present study, large-scale transcriptome sequencing was performed to compare mRNA expression in patients with HCM and control groups using an Illumina sequencing platform. Compared with the genome background, 257 differentially expressed genes (DEGs) were identified in which 62 genes were downregulated and 195 genes were upregulated. Reverse transcription-quantitative polymerase chain reaction was performed to validate the expression pattern of certain mRNAs. Gene ontology enrichment and KEGG analysis of mRNAs was conducted to identify the biological modules and pathological pathways associated with the DEGs. To the best of our knowledge, this is the first time study to investigate the differences in mRNA between patients with HCM and normal controls at the transcriptome level. The results of the study will contributed to the understanding of the important molecular mechanisms involved in HCM and aid the selection of key genes to investigate in the future.

Introduction

Myocardial hypertrophy is an initial physiological adaptive response to elevation of blood pressure and pressure overload, it is also a common pathophysiological process involved in various cardiovascular diseases, including hypertension, myocardial infarction, cardiac valvular disease, hypertrophic cardiomyopathy (HCM) (1). Although the intervention of drug can control blood pressure in the normal range, the myocardial hypertrophy will, inevitably, gradually progress to chronic heart failure (2). In fact, myocardial hypertrophy is an independent risk factor for cardiovascular disease, exponentially increased risk of cardiovascular mortality (3,4). Treatment for myocardial hypertrophy is remains confined to dilation of blood vessels, reducing myocardial contraction force and afterload, rarely directly targeting the formation of myocardial hypertrophy. HCM is a common cause of sudden cardiac arrest in young people, including young athletes (5,6). Accordingly, early diagnosis is essential for the prevention of such catastrophic events (7-9).

The transcriptome is the collection of all the produced mRNA transcripts within a species or specific cell types (10). Transcriptome analysis, that investigates gene function and expression from the overall level, revealing the specific molecular mechanisms in biological processes and disease, has been widely used in basic research, clinical diagnosis, drug research and other fields (11,12). Based on the high-throughput sequencing platform, transcriptome sequencing technology enables the detection of the overall transcriptional activity of any mRNA species and transcription at the single nucleotide level. Analysis of transcription and expression levels can also identify unknown and rare transcripts, accurately detecting the variable shear loci and coding sequence nucleoside polymorphisms, providing the most comprehensive transcriptional analysis. Compared with the traditional platform of chip hybridization, transcriptome sequencing without advance design of probes for known sequences, transcriptional activity analysis can be performed for any species. It also provides a more precise digital signal and high-throughput analysis with a broader scope, and is a powerful tool for studying transcriptome complexity (10,13).

The current study performed large-scale transcriptome sequencing comparing HCM patients and control groups using an Illumina sequencing platform. Differentially expressed genes (DEGs) were examined, and gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of mRNAs were performed to identify the biological modules and pathological pathways involved in HCM.

Materials and methods

Tissue collection from patients with HCM and healthy individuals. For the present study, all procedures were approved by the Ethics Committee of the Beijing Anzhen Hospital (Beijing, China). Written informed consent was obtained from all subjects under a general waiver by the institutional review

Correspondence to: Dr Yong-Qiang Lai, Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, 36 Wuluju, Chaoyang, Beijing 100029, P.R. China
E-mail: yongqianglai@yahoo.com

Key words: myocardial hypertrophy, transcriptome, differentially expressed genes, molecular mechanism

board of Beijing Anzhen Hospital. Hypertrophied myocardial tissues used in this study were obtained from 6 patients with HCM that exhibited severe symptoms and underwent septal myectomy at Department of Cardiac Surgery of Beijing Anzhen Hospital, termed the HY group. Additionally, 6 normal myocardial tissues were excised from 6 healthy individuals that were previously declared brain dead during autopsy that had voluntarily donated their body to Renmin Hospital of Wuhan University (Wuhan, China) for research, and were used as the control group, termed the NH group. Patient characteristics were as follows: Age, 35-57 years, 6 males and 6 females.

Transcript profiling using RNA-sequencing (RNA-Seq). Total RNAs were extracted from samples using RNeasy Mini kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. The total RNA samples were initially treated with 1 U/μl DNase I to degrade any potential DNA contamination. Then the mRNA was enriched with 50 μl/sample oligo (dT) magnetic beads (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The samples were mixed with the fragmentation buffer [NEB Next First Strand Synthesis Reaction Buffer (New England BioLabs, Inc., Ipswich, MA, USA)] and fragmented into short fragments of ~200 bps. The first strand of cDNA was synthesized using random hexamer-primers following the manufacturer's protocol (Thermo Fisher Scientific, Inc.). They were incubated for 10 min at 94˚C. The double stranded cDNA was purified with AMPure XP magnetic beads (Beckman Coulter, Inc., Brea, CA, USA). End reparation and 3'-end single nucleotide A addition was performed as previously described (14). Finally, sequencing adaptors (Illumina, Inc., San Diego, CA, USA) were ligated to the fragments and the fragments were amplified by polymerase chain reaction (PCR) according to the Illumina (Illumina, Inc.) instructions. Agilent 2100 Bioanaylzer (Agilent Technologies, Inc., Santa Clara, CA, USA) and ABI StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Inc.) were used to qualify and quantify the sample library. The library products were sequenced via Illumina HiSeq2000 (Illumina, Inc.).

Read mapping and analysis. By base calling, the original image data produced by the sequencing machine is converted into sequences. Prior to starting further analysis, data filtering is preprocessed to obtain clean reads using in-house Perl scripts based on criterion previously reported (15). Short Oligonucleotide Analysis Package (SOAP) software (SOAPaligner/soap2; soap.genomics.org.cn/soapaligner.html) was used to map reads into reference sequences and reference gene set (UCSC genome browser; genome.ucsc.edu/; version hg19), in which no more than 2 mismatches were allowed (16). Annotating the results that from Basic Local Alignment Search Tool (BLAST; version 2.2.23; parameters: -p blastx-e 1e-5 -m7) sequences of the 'non-redundant' NCBI database to the terms of GO using Blast2GO (version 2.2.5; www.blast2go.com; default parameters) (17). Annotation to the KEGG database was also performed with BLAST (parameters: -p blastx -e 1e-5-m 8).

Gene expression analysis. The gene expression level was calculated using the reads per kilobase per million algorithm (18). The NOI-seq method (19) was used to screen differentially expressed genes between two groups which is a non-parametric approach for differential expression analysis. (19). Screening differentially expression genes between the groups requires high correlation among the replicates. The Pearson method was used to obtain coefficients for all genes between every two samples for the reference (data not shown). GO enrichment analysis based on hypergeometric distribution mode provided all GO terms of DEGs that were significantly enriched compared with the genome background, and screened the DEGs with which biological function was significantly correlated. KEGG pathway enrichment identified significantly enriched metabolic pathways or signal transduction pathways associated with the DEGs compared with the whole genome background using hypergeometric distribution. We use WEGO software (wego.genomics.org.cn/cgi-bin/wego/index.pl) to perform the GO function classification of different genes, and understand the functional distribution of different genes at the macro level (20).

Reverse transcription-quantitative PCR (RT-qPCR). Total tissue RNA was isolated from NH and HY cardiac tissue samples using TRIzol reagent (Thermo Fisher Scientific, Inc.) and carried out RT was performed using PrimeScript™ II first strand cDNA Synthesis kit (Takara Bio, Inc., Otsu, Japan) according to the manufacturer's instructions. RT-qPCR was performed using ABI 7900HT cycler (Thermo Fisher Scientific, Inc.) and 5-carboxyfluorescein (Sangon Biotech Co., Ltd., Shanghai, China) was used as the fluorophore. The following PCR primer sequences were used: fibromodulin (*FMOD*) forward (F) 5'-GCTGCTGTATGTGCGGCT-3' and reverse (R) 5'-AAGTTCACGACGTCCACCAC-3'; fibroblast growth factor (*FGF12*) F 5'-TTCTCGGATGGAAAGTCTGG -3' and R 5'-ATCAAGGTGTCCACAGGGCT-3'; potassium voltage-gated channel interacting protein 2 (*KCNIP2*) F 5'-AGCGCGATCCCTCTACCA-3' and R 5'-GGTGACACA CGGTGGACAAT-3'; corin serine peptidase (*CORIN*) F 5'-TGTGCTCTCGTTCTCTTGCT-3' and R 5'-GAATAA CATCGGACCCTTGG-3'; connective tissue growth factor (*CTGF*) F 5'-AATGACAACGCCTCCTGC-3' and R 5'-TGC ACTTTTTGCCCTTCTAA-3'; and hActin F 5'-AGCACA ATGAAGATCAAGATCAT-3' and R 5'-ACTCGTCATACT CCTGCTTGC-3'. PCR conditions were as follows: Pre-incubation at 95˚C for 3 min, following by 40 cycles of 95˚C for 3 sec and 60˚C for 20 sec, with a final cycle of 95˚C for 15 sec, 60˚C for 15 sec and 95˚C for 15 sec. Actin was used as an internal control and relative quantification was calculated by the $2^{-\Delta\Delta Cq}$ method (21).

Statistical analysis. The statistical significance of GO enrichment and pathway enrichment were analyzed by using the hypergeometric distribution and false discovery rate was calculated to correct the P-value using Bonferroni correction, with $P \leq 0.05$ considered to indicate a statistically significant difference. For comparison of two groups using the NOI-seq method, a gene was declared as a DEG with probability is ≥ 0.8 and fold change ≥ 2. Data analysis use R version 3.0.1 (www.r-project.org) with the addition of the ggplot2 package (ggplot2.org).

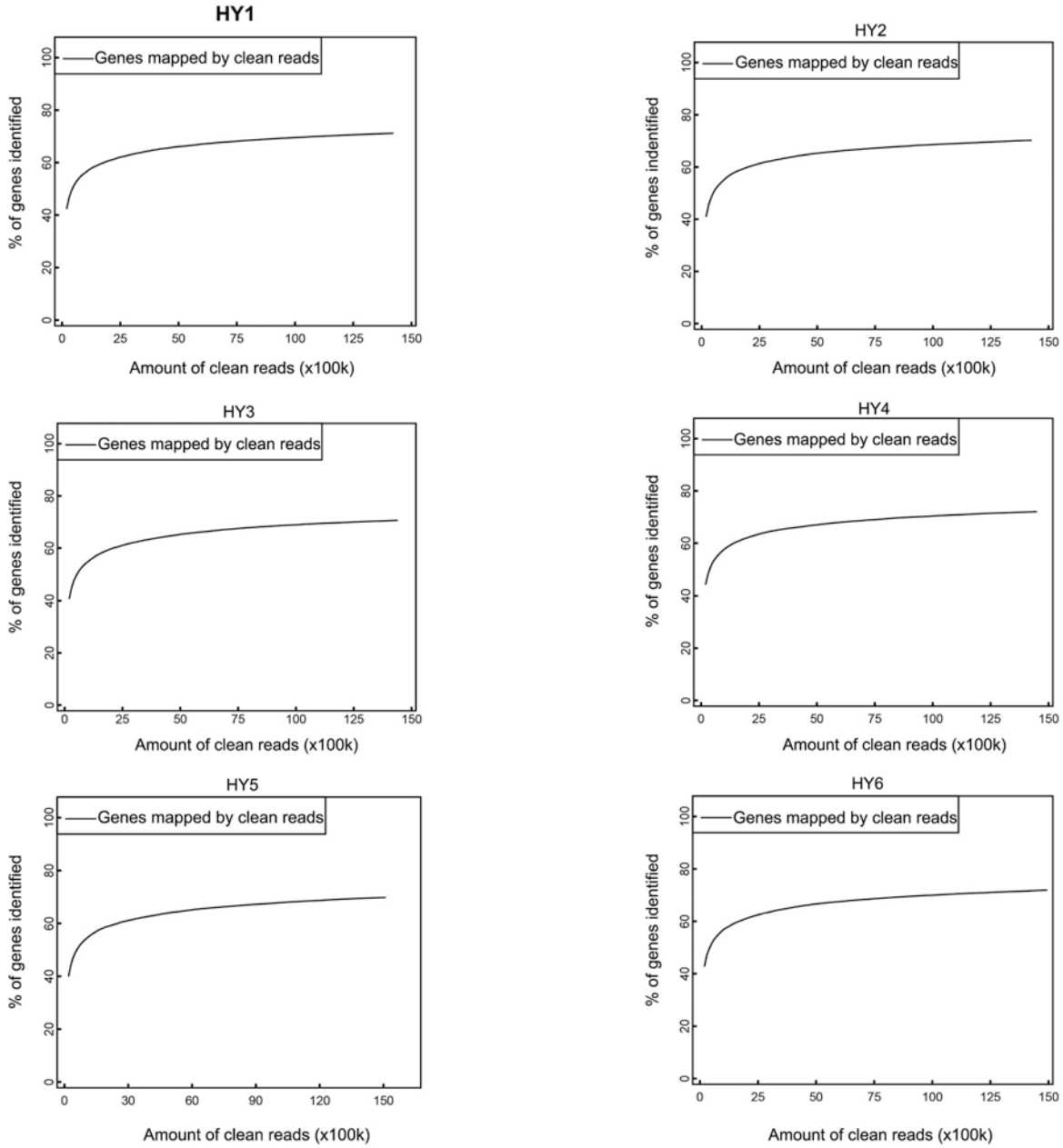

Figure 1. Sequencing saturation analysis of patients with hypertrophic cardiomyopathy that underwent septal myectomy.

Results

Sequencing saturation analysis and gene coverage. Following sequencing using the HiSeq2000 and raw data filtering, reads were mapped to the reference gene set. The average sample obtained 14,459,386 reads used for further analysis. On average, in each sample the percentage of reads mapped into the gene set was 54.38%, ranging from 48.55 to 62.41%; mapped into the genome was 91.33%, ranging from 90.35 to 92.43%. Sequence saturation analysis demonstrated that the higher the number of reads the more genes that were detected. When the number of reads reached a certain level, the number of detected genes was in a constant state, which implied the detected gene number trended toward saturation (Figs. 1 and 2). Gene coverage was analyzed by the ratio of the number of bases in a gene covered by unique mapping reads. Thus, the RNA-Seq data is sufficient for profiling of the transcriptome differences between HCM patients and normal controls.

Sequencing randomness analysis. The randomness of sequencing, immediate impact subsequent bioinformatics analysis, was analyzed by the distribution of reads on reference genes. As reference genes have different lengths, the reads position on gene was standardized to a relative location, which is the ratio between reads position on the gene and gene length. The results of distributions of reads on reference genes of each sample suggested that the randomness of RNA fragmentation is acceptable (Figs. 3 and 4).

DEGs between the groups. The NOI-seq method was performed to screen for DEGs between the HY and NH groups. In total, 19,962 DEGs were detected by comparing the NH group and HY group. Based on filter criteria of probability ≥0.8 and

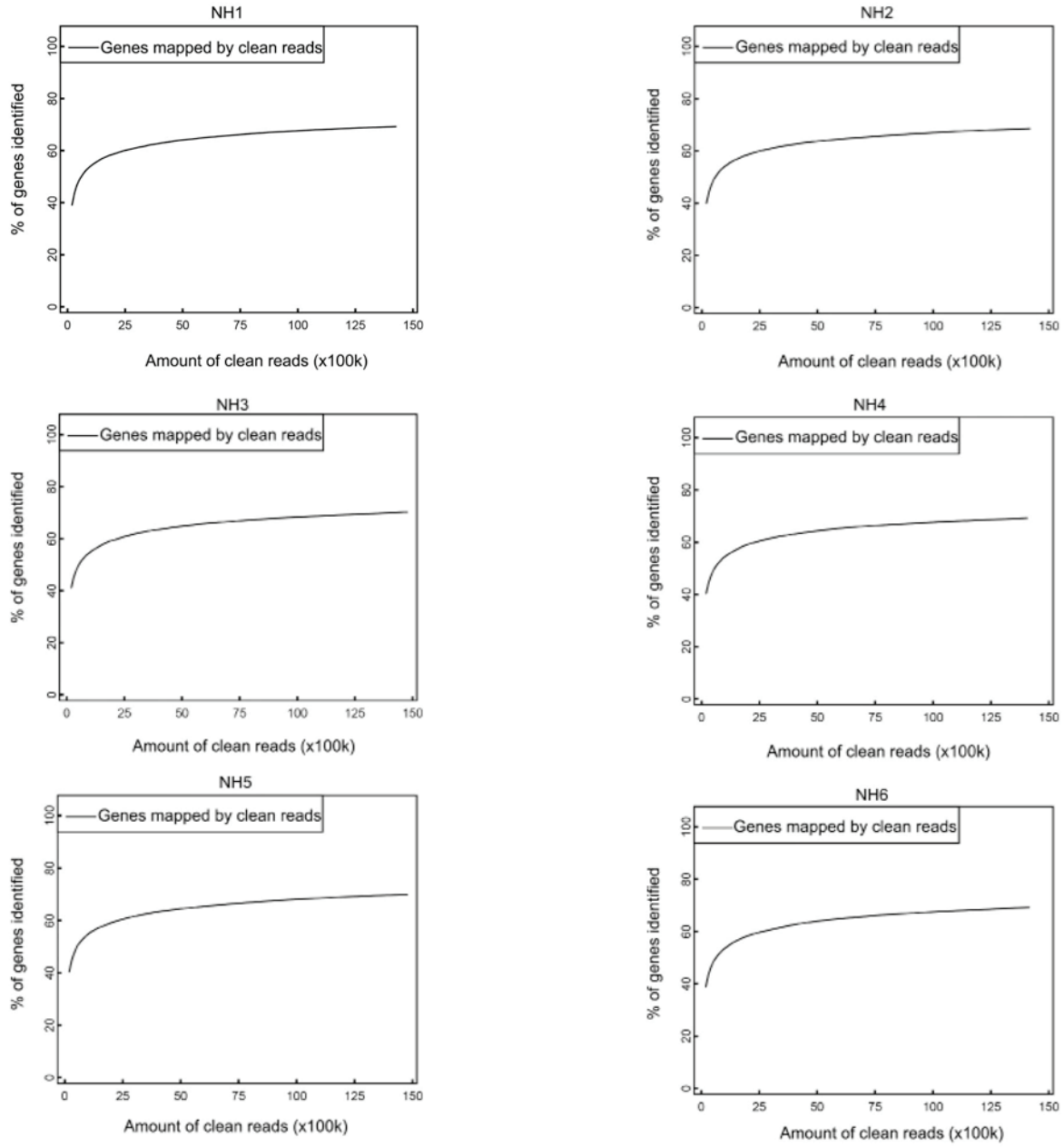

Figure 2. Sequencing saturation analysis of samples from healthy control group patients.

|log2Ratio|≥1, 62 downregulated and 195 upregulated DEGs were identified (Fig. 5). Thus, 1.29% genes were significantly regulated, which indicates that these genes may be HCM-specific genes. To ascertain possible functions of significantly regulated genes, the heart-specific genes were searched in the Pattern Gene Database (22) and are presented in Table I.

RT-qPCR to validate the expression pattern of mRNA. The mRNA expression of certain significantly regulated genes including *CORIN*, *CTGF*, *FGF12*, *FMOD*, *KCNIP2*, was validated using RT-qPCR as they had a significant log2Ratio. All the mRNAs exhibited similar changes in expression pattern as observed by the transcriptome sequencing analysis (Fig. 6).

GO functional classification and enrichment analysis of DEGs. GO annotation is a useful means of assigning functional information to genes employing standardized vocabulary. In the current study, a total of 16,090 genes were annotated into cellular component groups, of which 228 genes were DEGs; 15,165 genes were annotated into molecular function groups, of which 205 genes were DEGs; 14,596 genes were annotated into biological process groups, of which 213 genes were DEGs. Subsequently, GO annotation for DEGs were classified using WEGO software to determine the distribution of gene functions at the macro level, as presented in Fig. 7. Of the biological process categories, the major categories identified were cellular process (12.7%), metabolic process (8.6%) and multicellular organismal process (8.6%). In molecular function, the most common terms were binding (52%), catalytic activity (20.5%) and molecular transducer activity (9.3%). For cellular components, the dominant groups were cell (25%), cell part (25%) and organelle (15%). GO enrichment analysis identified significantly

Table I. Heart-specific genes among the significantly regulated genes.

Gene ID	log2Ratio (HY/NH)	Probability	Symbol	Description
2006	1.671998011	0.856736132	ELN	Elastin
4624	-1.183412466	0.837763417	MYH6	Myosin, heavy chain 6, cardiac muscle, α
8728	1.447600926	0.810635207	ADAM19	ADAM metallopeptidase domain 19
27063	1.131351044	0.838543232	ANKRD1	Ankyrin repeat domain 1 (cardiac muscle)
26287	-1.276392993	0.851026951	ANKRD2	Ankyrin repeat domain 2 (stretch responsive muscle)
10930	-1.372749343	0.85623351	APOBEC2	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
441432	-2.374262995	0.917853922	AQP7P3	Aquaporin 7 pseudogene 3
488	-1.383797756	0.863984905	ATP2A2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
633	1.934662987	0.900285543	BGN	Biglycan
84529	-1.499948403	0.871530909	C15orf41	Chromosome 15 open reading frame 41
92346	-3.153831161	0.835724543	C1orf105	Chromosome 1 open reading frame 105
84808	-1.324498257	0.834518919	C1orf170	Chromosome 1 open reading frame 170
23632	-3.767087943	0.903346358	CA14	Carbonic anhydrase XIV
55799	-1.883350727	0.846184417	CACNA2D3	Calcium channel, voltage-dependent, α 2/δ subunit 3
151887	1.955639687	0.898714224	CCDC80	Coiled-coil domain containing 80
6356	-2.011081118	0.850576095	CCL11	Chemokine (C-C motif) ligand 11
1009	2.112503788	0.871123468	CDH11	Cadherin 11, type 2, OB-cadherin (osteoblast)
25884	-5.120525912	0.927139899	CHRDL2	Chordin-like 2
1158	-1.130350575	0.838391277	CKM	Creatine kinase, muscle
1047	-1.277350137	0.844611428	CLGN	Calmegin
1191	1.039771123	0.813509	CLU	Clusterin
1307	1.583729933	0.827577397	COL16A1	Collagen, type XVI, α 1
1282	1.176982895	0.835489096	COL4A1	Collagen, type IV, α 1
1295	1.960380919	0.878058311	COL8A1	Collagen, type VIII, α 1
1311	3.031852231	0.91569649	COMP	Cartilage oligomeric matrix protein
10699	-3.878815566	0.952008817	CORIN	Corin, serine peptidase
131034	-2.210328025	0.864417393	CPNE4	Copine IV
1490	3.073801351	0.941660822	CTGF	Connective tissue growth factor
1545	1.214028874	0.812515446	CYP1B1	Cytochrome P450, family 1, subfamily B, polypeptide 1
54541	1.416237836	0.845394583	DDIT4	DNA-damage-inducible transcript 4
201140	-4.273538498	0.882605283	DHRS7C	Dehydrogenase/reductase (SDR family) member 7C
1734	2.316538895	0.900131917	DIO2	Deiodinase, iodothyronine, type II
148252	-1.47099063	0.858167184	DIRAS1	DIRAS family, GTP-binding RAS-like 1
10085	1.946972867	0.880167318	EDIL3	EGF-like repeats and discoidin I-like domains 3
2202	1.308741888	0.846304645	EFEMP1	EGF containing fibulin-like extracellular matrix protein 1
2012	1.180723625	0.825653742	EMP1	Epithelial membrane protein 1
2042	2.165840112	0.866184083	EPHA3	EPH receptor A3
2191	3.125633582	0.897688942	FAP	Fibroblast activation protein, α
10160	1.923691924	0.872395886	FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)
2257	-4.360911591	0.970832916	FGF12	Fibroblast growth factor 12
2252	-1.581285406	0.826862706	FGF7	Fibroblast growth factor 7
2274	-1.854040983	0.897775774	FHL2	Four and a half LIM domains 2
387758	2.112065785	0.888169188	FIBIN	Fin bud initiation factor homolog (zebrafish)
161247	-1.179247392	0.834690913	FITM1	Fat storage-inducing transmembrane protein 1
2331	3.364977995	0.948149818	FMOD	Fibromodulin
2335	2.077955587	0.906519053	FN1	Fibronectin 1
2628	1.318314029	0.830960492	GATM	Glycine amidinotransferase (L-arginine:glycine amidinotransferase)
23171	-1.105228276	0.827253448	GPD1L	Glycerol-3-phosphate dehydrogenase 1-like

Table I. Continued.

Gene ID	log2Ratio (HY/NH)	Probability	Symbol	Description
1404	2.225726918	0.826410179	HAPLN1	Hyaluronan and proteoglycan link protein 1
1839	1.214995299	0.825273019	HBEGF	Heparin-binding EGF-like growth factor
8988	-1.143998603	0.836267241	HSPB3	Heat shock 27kDa protein 3
3400	2.780750996	0.882441639	ID4	Inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
8516	1.980625674	0.829459306	ITGA8	Integrin, α 8
9358	2.750740936	0.894155562	ITGBL1	Integrin, β-like 1 (with EGF-like repeat domains)
30819	-3.66846088	0.953588485	KCNIP2	Kv channel interacting protein 2
63971	1.162360593	0.825398257	KIF13A	Kinesin family member 13A
401265	-1.436986107	0.863475604	KLHL31	Kelch-like family member 31
4016	1.486502481	0.836935177	LOXL1	Lysyl oxidase-like 1
79442	-1.322895489	0.851439402	LRRC2	Leucine rich repeat containing 2
4053	2.796999627	0.929172094	LTBP2	Latent transforming growth factor β binding protein 2
4134	1.066393085	0.825849113	MAP4	Microtubule-associated protein 4
4256	1.412196685	0.86674014	MGP	Matrix Gla protein
4634	-1.160393404	0.843162008	MYL3	Myosin, light chain 3, alkali; ventricular, skeletal, slow
4635	-1.186877706	0.841288448	MYL4	Myosin, light chain 4, alkali; atrial, embryonic
4642	2.207455692	0.873985573	MYO1D	Myosin ID
4856	2.030294086	0.863250175	NOV	Nephroblastoma overexpressed
4878	2.98935631	0.942248606	NPPA	Natriuretic peptide A
4879	2.044931136	0.9038423	NPPB	Natriuretic peptide B
4883	1.626298894	0.868326487	NPR3	Natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)
4958	2.33410036	0.890056107	OMD	Osteomodulin
157310	-1.534651663	0.871235347	PEBP4	Phosphatidylethanolamine-binding protein 4
5224	-1.050384	0.823200748	PGAM2	Phosphoglycerate mutase 2 (muscle)
113791	1.597176924	0.838830445	PIK3IP1	Phosphoinositide-3-kinase interacting protein 1
5320	-2.275020758	0.907515947	PLA2G2A	Phospholipase A2, group IIA (platelets, synovial fluid)
130271	2.334510641	0.803469926	PLEKHH2	Pleckstrin homology domain containing, family H (with MyTH4 domain) member 2
10631	2.750726847	0.933228133	POSTN	Periostin, osteoblast specific factor
5502	-1.144177711	0.814457469	PPP1R1A	Protein phosphatase 1, regulatory (inhibitor) subunit 1A
5627	1.362968106	0.857115186	PROS1	Protein S (α)
9791	-1.817652943	0.887801823	PTDSS1	Phosphatidylserine synthase 1
51715	1.223289264	0.80093344	RAB23	RAB23, member RAS oncogene family
11031	1.555066484	0.852785292	RAB31	RAB31, member RAS oncogene family
153769	-1.040437854	0.815619677	SH3RF2	SH3 domain containing ring finger 2
6508	-1.020533026	0.809541462	SLC4A3	Solute carrier family 4, anion exchanger, member 3
6523	-1.251202925	0.83638413	SLC5A1	Solute carrier family 5 (sodium/glucose cotransporter), member 1
6781	2.021463755	0.828515847	STC1	Stanniocalcin 1
7057	1.536296186	0.856310323	THBS1	Thrombospondin 1
7060	2.593514693	0.927819524	THBS4	Thrombospondin 4
3371	1.559538856	0.814337241	TNC	Tenascin C
7138	1.041963712	0.818049294	TNNT1	Troponin T type 1 (skeletal, slow)
7170	1.82824907	0.888962362	TPM3	Tropomyosin 3
1462	1.888942703	0.875560231	VCAN	Versican

enriched GO terms in DEGs compared with the genome background using hypergeometric analysis. Bonferroni correction was performed on the calculated P-value, taking corrected P≤0.05 as the threshold for significance. GO terms accorded

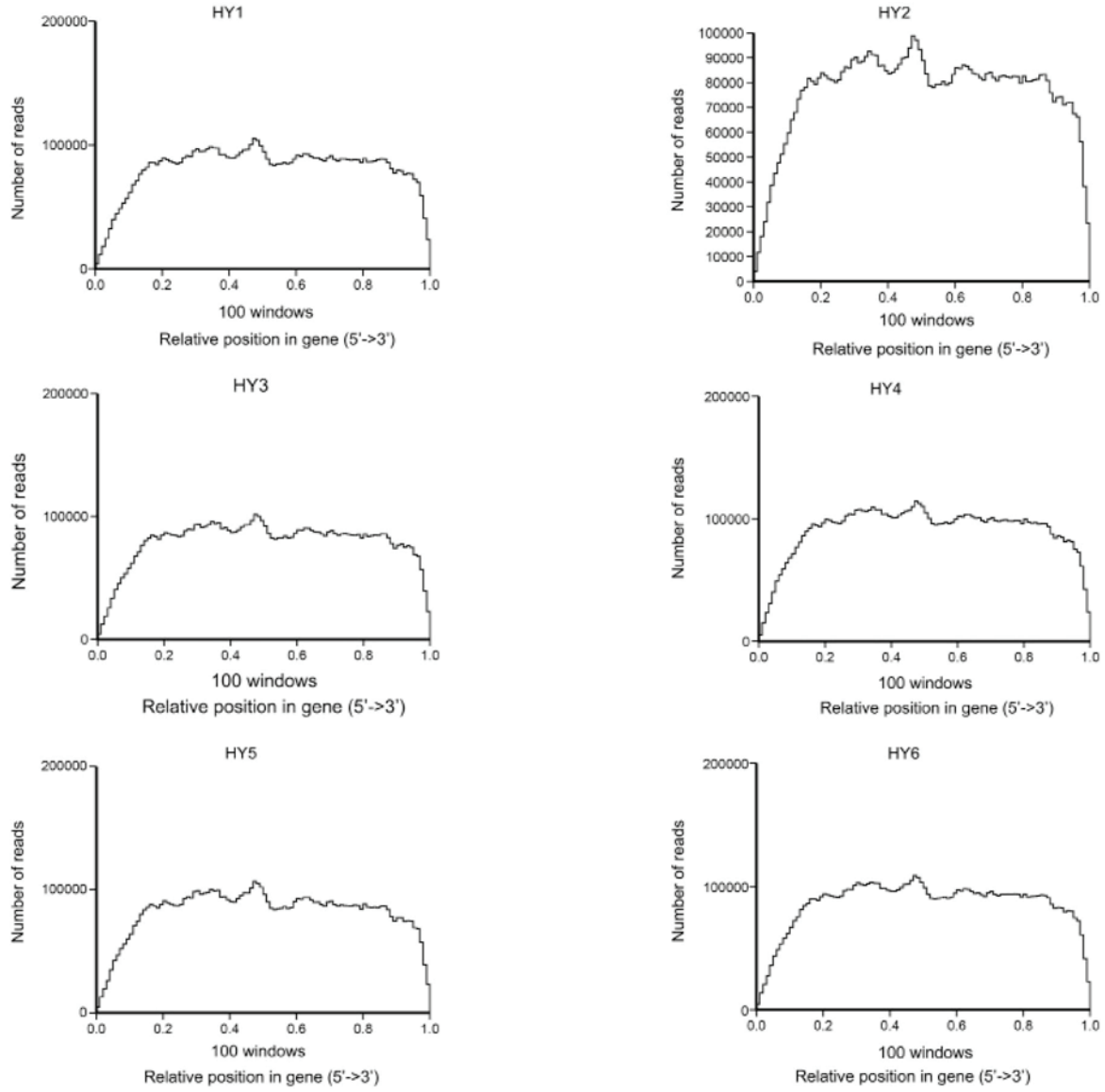

Figure 3. Randomness assessment analysis of with hypertrophic cardiomyopathy that underwent septal myectomy.

with this condition were defined as significantly enriched GO terms in DEGs. This analysis was able to recognize the main biological functions that the DEGs are involved in (Table II).

KEGG pathway enrichment analysis of DEGs. In the organism, genetically distinct genes exert biological functions by coordinating with each other. Pathway enrichment analysis can identify differences in DEGs involved in the major biochemical metabolism and signal transduction pathways. Pathway enrichment analysis based on the KEGG pathway database applied hypergeometric analysis to identify the significantly enriched pathways among DEGs compared with the whole genome. KEGG pathway analysis mapped 225 DEGs to 170 pathways. The top 20 enriched pathway terms are presented in a scatter diagram (Fig. 8). The RichFactor is the ratio between DEG numbers annotated in certain pathway terms and all gene numbers also annotated in same pathway term. A greater RichFactor indicates a greater degree of enrichment. The Q value is the corrected P-value ranging from 0-1, and its lower value indicates greater enrichment.

Discussion

The clinical characteristics of HCM are diverse, including the onset age, degree of disease and sudden death risk. Thus, it is important to further clarify the pathophysiology and investigate the molecular mechanism of the disease, as genetic diagnosis and therapy has become a new research focus and direction (23). With continuously increasing research regarding HCM, the American College of Cardiology Foundation and American Heart Association jointly issued the Guideline for the Diagnosis and Treatment of Hypertrophic Cardiomyopathy. The definition of HCM was updated in these guidelines, and the difference between the new definition and the previous demonstrates the importance HCM-associated genes in the diagnosis of HCM (24). Currently, HCM is considered to be autosomal dominant disease caused by mutations in sarcomere protein genes. At present, eight genes have identified to be associated with the disease, with >1400 mutations (25).

In the present study, a comprehensive transcriptome analysis was performed by comparing HCM tissues and normal

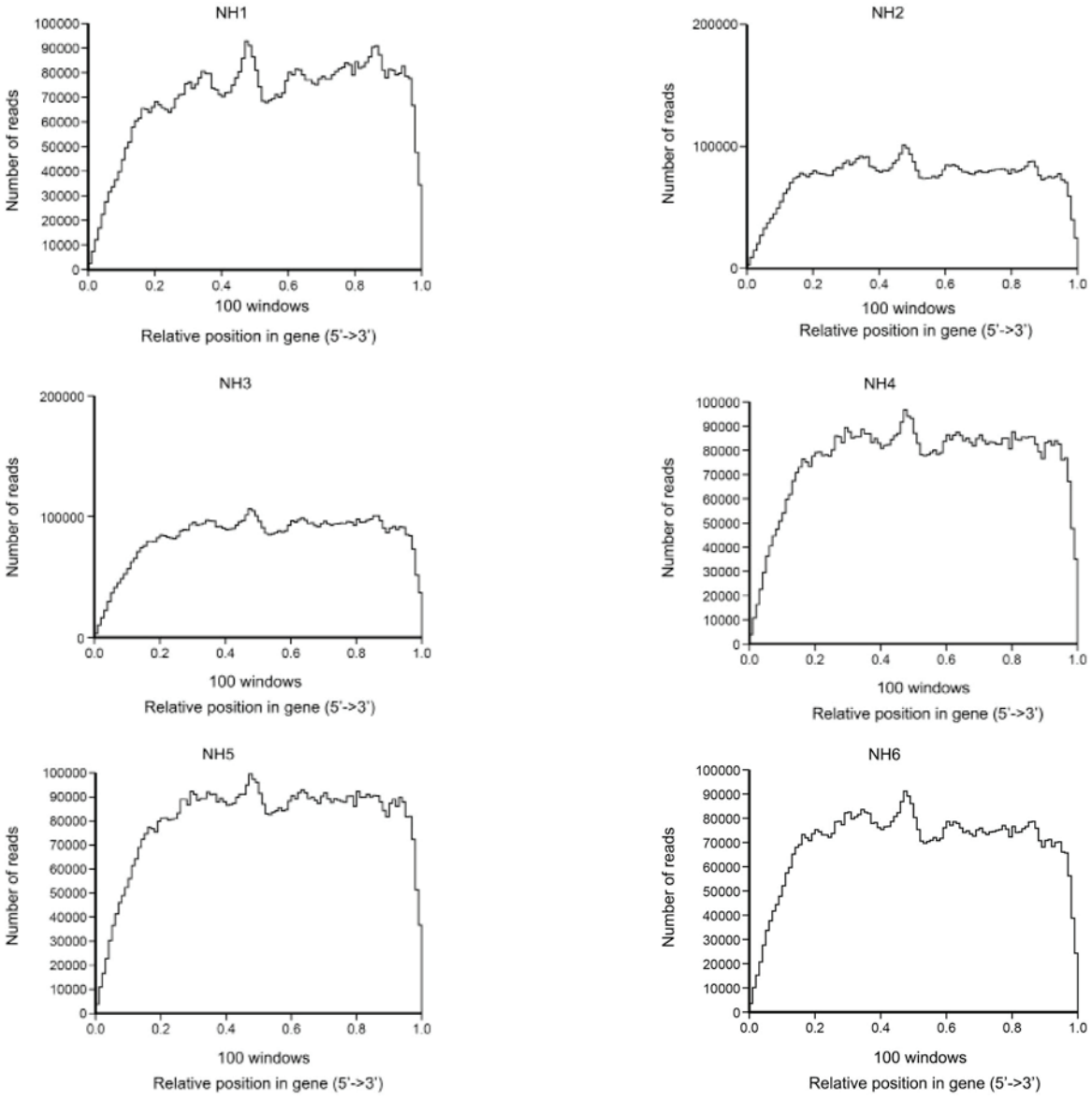

Figure 4. Randomness assessment of healthy control group patients.

myocardial tissue. A total of 257 DEGs were identified in which 195 genes were upregulated and 62 genes were downregulated. Among the regulated genes, 91 genes were heart-tissues specific. The current study focused on differentially expressed genes that were significantly regulated (|log2Ratio|>4). Of these genes, chordin-like 2 (*CHRDL2*), *FGF12*, and dehydrogenase/reductase (SDR family) member 7C (*DHRS7C*) were downregulated and seizure related 6 homolog like (*SEZ6L*), endothelial cell specific molecule 1 (*ESM1*), collagen type X α 1 chain (*COL10A1*), secreted frizzled related protein 4 (*SFRP4*), carbonic anhydrase 3 (*CA3*) were upregulated.

CHRDL2 encodes a member of the secreted chordin family of proteins that have a cysteine-rich pro-collagen repeat domain and associate with members of the transforming growth factor (TGF)-β superfamily (26,27). This gene is expressed in intracellular membrane-bounded organelles and involved in tissue development (GO:0043231). Similarly, *FGF12* encoded a member of the fibroblast growth factor family, which possess broad mitogenic and cell survival activities, and are involved

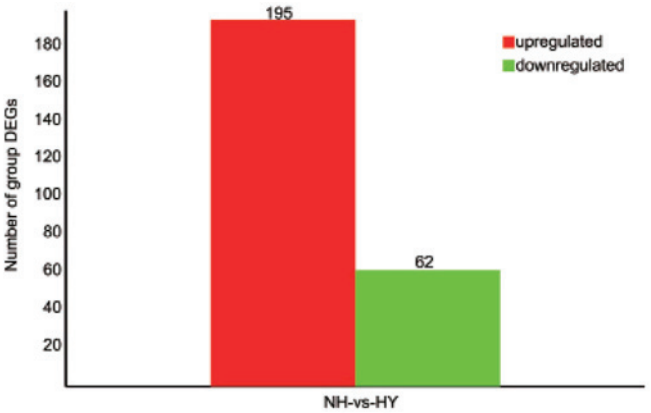

Figure 5. Summary of the number of DEGs. DEG, differentially expressed gene.

in a variety of biological processes, including embryonic development, cell growth, morphogenesis, tissue repair, tumor

Table II. Significantly enriched GO terms in DEGs

A, Cellular component

GO term	Cluster frequency	Genome frequency of use	Corrected P-value
Extracellular region	87 out of 228 genes, 38.2%	1077 out of 16090 genes, 6.7%	1.36E-41
Extracellular region part	86 out of 228 genes, 37.7%	1055 out of 16090 genes, 6.6%	2.42E-41
Extracellular matrix	48 out of 228 genes, 21.1%	312 out of 16090 genes, 1.9%	5.46E-34
Proteinaceous extracellular matrix	25 out of 228 genes, 11.0%	113 out of 16090 genes, 0.7%	4.84E-21
Collagen	9 out of 228 genes, 3.9%	18 out of 16090 genes, 0.1%	9.93E-11
Extracellular matrix part	10 out of 228 genes, 4.4%	50 out of 16090 genes, 0.3%	1.94E-07
Platelet α granule	9 out of 228 genes, 3.9%	56 out of 16090 genes, 0.3%	9.69E-06
Fibrillar collagen	4 out of 228 genes, 1.8%	5 out of 16090 genes, 0.0%	2.25E-05
Contractile fiber	11 out of 228 genes, 4.8%	145 out of 16090 genes, 0.9%	0.0008
Extracellular space	7 out of 228 genes, 3.1%	53 out of 16090 genes, 0.3%	0.00107
Anchoring collagen	3 out of 228 genes, 1.3%	6 out of 16090 genes, 0.0%	0.00631
Myosin II complex	4 out of 228 genes, 1.8%	18 out of 16090 genes, 0.1%	0.01192
Actin cytoskeleton	11 out of 228 genes, 4.8%	204 out of 16090 genes, 1.3%	0.01866
Stored secretory granule	10 out of 228 genes, 4.4%	179 out of 16090 genes, 1.1%	0.02783
Sarcomere	7 out of 228 genes, 3.1%	92 out of 16090 genes, 0.6%	0.03829
Contractile fiber part	7 out of 228 genes, 3.1%	93 out of 16090 genes, 0.6%	0.04092
Myofibril	8 out of 228 genes, 3.5%	126 out of 16090 genes, 0.8%	0.04993

B, Molecular function

GO term	Cluster frequency	Genome frequency of use	Corrected P-value
Pattern binding	26 out of 205 genes, 12.7%	169 out of 15165 genes, 1.1%	3.72E-18
Polysaccharide binding	24 out of 205 genes, 11.7%	159 out of 15165 genes, 1.0%	1.82E-16
Glycosaminoglycan binding	22 out of 205 genes, 10.7%	142 out of 15165 genes, 0.9%	3.01E-15
Protein binding	110 out of 205 genes, 53.7%	4325 out of 15165 genes, 28.5%	3.76E-12
Carbohydrate binding	27 out of 205 genes, 13.2%	363 out of 15165 genes, 2.4%	7.79E-11
Growth factor binding	15 out of 205 genes, 7.3%	87 out of 15165 genes, 0.6%	8.73E-11
Receptor binding	35 out of 205 genes, 17.1%	915 out of 15165 genes, 6.0%	2.90E-06
Structural molecule activity	21 out of 205 genes, 10.2%	423 out of 15165 genes, 2.8%	4.18E-05
Glycoprotein binding	4 out of 205 genes, 2.0%	17 out of 15165 genes, 0.1%	0.00954
Binding	185 out of 205 genes, 90.2%	12212 out of 15165 genes, 80.5%	0.01507
Collagen binding	2 out of 205 genes, 1.0%	3 out of 15165 genes, 0.0%	0.07677
Myosin heavy chain binding	2 out of 205 genes, 1.0%	3 out of 15165 genes, 0.0%	0.07677
Phosphoric diester hydrolase activity	5 out of 205 genes, 2.4%	65 out of 15165 genes, 0.4%	0.26125
Insulin-like growth factor binding	2 out of 205 genes, 1.0%	6 out of 15165 genes, 0.0%	0.37373
Peptidase regulator activity	8 out of 205 genes, 3.9%	182 out of 15165 genes, 1.2%	0.47604

C, Biological process

GO term	Cluster frequency	Genome frequency of use	Corrected P-value
Extracellular structure organization	21 out of 213 genes, 9.9%	124 out of 14596 genes, 0.8%	6.35E-14
Anatomical structure development	90 out of 213 genes, 42.3%	2929 out of 14596 genes, 20.1%	7.24E-11
Developmental process	99 out of 213 genes, 46.5%	3513 out of 14596 genes, 24.1%	4.46E-10
System development	74 out of 213 genes, 34.7%	2405 out of 14596 genes, 16.5%	3.94E-08
Multicellular organismal development	79 out of 213 genes, 37.1%	2679 out of 14596 genes, 18.4%	5.18E-08
Multicellular organismal process	117 out of 213 genes, 54.9%	4897 out of 14596 genes, 33.6%	7.00E-08

Table II. Continued.

GO term	Cluster frequency	Genome frequency of use	Corrected P-value
Extracellular matrix organization	10 out of 213 genes, 4.7%	43 out of 14596 genes, 0.3%	3.43E-07
Enzyme linked receptor protein signaling pathway	32 out of 213 genes, 15.0%	716 out of 14596 genes, 4.9%	1.12E-05
Locomotion	35 out of 213 genes, 16.4%	838 out of 14596 genes, 5.7%	0.0000121
Response to wounding	17 out of 213 genes, 8.0%	225 out of 14596 genes, 1.5%	0.0000254
Muscle contraction	13 out of 213 genes, 6.1%	127 out of 14596 genes, 0.9%	0.0000319
Anatomical structure morphogenesis	43 out of 213 genes, 20.2%	1230 out of 14596 genes, 8.4%	0.0000414
Cell motility	26 out of 213 genes, 12.2%	538 out of 14596 genes, 3.7%	0.0000634
Localization of cell	26 out of 213 genes, 12.2%	538 out of 14596 genes, 3.7%	0.0000634
Cell migration	23 out of 213 genes, 10.8%	431 out of 14596 genes, 3.0%	0.0000643
Cellular component movement	26 out of 213 genes, 12.2%	542 out of 14596 genes, 3.7%	0.0000734
Transmembrane receptor protein serine/ threonine kinase signaling pathway	15 out of 213 genes, 7.0%	188 out of 14596 genes, 1.3%	8.23E-05
Muscle system process	16 out of 213 genes, 7.5%	230 out of 14596 genes, 1.6%	2.00E-04
Cell development	25 out of 213 genes, 11.7%	568 out of 14596 genes, 3.9%	6.50E-04
Cellular component organization	73 out of 213 genes, 34.3%	2968 out of 14596 genes, 20.3%	9.60E-04
Cell growth	15 out of 213 genes, 7.0%	237 out of 14596 genes, 1.6%	1.59E-03
Regulation of cell size	16 out of 213 genes, 7.5%	280 out of 14596 genes, 1.9%	2.74E-03
Transforming growth factor β receptor signaling pathway	8 out of 213 genes, 3.8%	65 out of 14596 genes, 0.4%	3.42E-03
Tissue development	28 out of 213 genes, 13.1%	749 out of 14596 genes, 5.1%	3.43E-03
Cellular component organization or biogenesis	73 out of 213 genes, 34.3%	3099 out of 14596 genes, 21.2%	0.00501
Regulation of cell adhesion	10 out of 213 genes, 4.7%	121 out of 14596 genes, 0.8%	0.00842
Regulation of cell-substrate adhesion	7 out of 213 genes, 3.3%	55 out of 14596 genes, 0.4%	0.01099
Complement activation	5 out of 213 genes, 2.3%	23 out of 14596 genes, 0.2%	0.01318
Organ development	44 out of 213 genes, 20.7%	1580 out of 14596 genes, 10.8%	0.01345
Regulation of anatomical structure size	18 out of 213 genes, 8.5%	394 out of 14596 genes, 2.7%	0.01493
Striated muscle cell development	7 out of 213 genes, 3.3%	58 out of 14596 genes, 0.4%	0.0157
Response to stimulus	76 out of 213 genes, 35.7%	3380 out of 14596 genes, 23.2%	0.01659
Growth	19 out of 213 genes, 8.9%	438 out of 14596 genes, 3.0%	1.79E-02
Regulation of multicellular organismal process	31 out of 213 genes, 14.6%	976 out of 14596 genes, 6.7%	2.62E-02
Muscle cell development	7 out of 213 genes, 3.3%	65 out of 14596 genes, 0.4%	3.34E-02
Regulation of cellular component size	16 out of 213 genes, 7.5%	345 out of 14596 genes, 2.4%	3.66E-02
Response to chemical stimulus	47 out of 213 genes, 22.1%	1810 out of 14596 genes, 12.4%	3.92E-02
Muscle fiber development	6 out of 213 genes, 2.8%	47 out of 14596 genes, 0.3%	4.51E-02

GO, gene ontology; DEGs, differentially expressed genes.

growth and invasion (28). However, *SFRP4*, that is involved in positive regulation of cell differentiation and tissue remodeling was downregulated in HCM patients, with that |log2Ratio| of 4.14. It is suggested that cardiac hypertrophy is predominantly caused by changes to myocardial cell structure, rather than the number of myocardial cells. The result may be explained that, under normal circumstances, the cardiac muscle cells are assembled into a parallel arrangement of straight lines. Whereas in HCM patients, myocardial cells have a disorderly arrangement, and abnormal cell connections, with diversity in the diameter and the length of individual cells (29,30). Myocardial cell nuclear size may also vary. The HCM phenotype is associated with diastolic dysfunction, leading to increased susceptibility to arrhythmia, which is an important factor involved in HCM morbidity and mortality (31).

DHRS7C is strongly conserved in vertebrates, and mRNA and protein expression levels are highest in heart and skeletal muscle followed by skin, but not detectable in other organs (32,33). Lu *et al* (34) demonstrated the downregulation of *DHRS7C* in multiple rat heart failure models and in human heart failure patient biopsies. It provides evidence for HCM is an important cause of heart failure-associated disability over

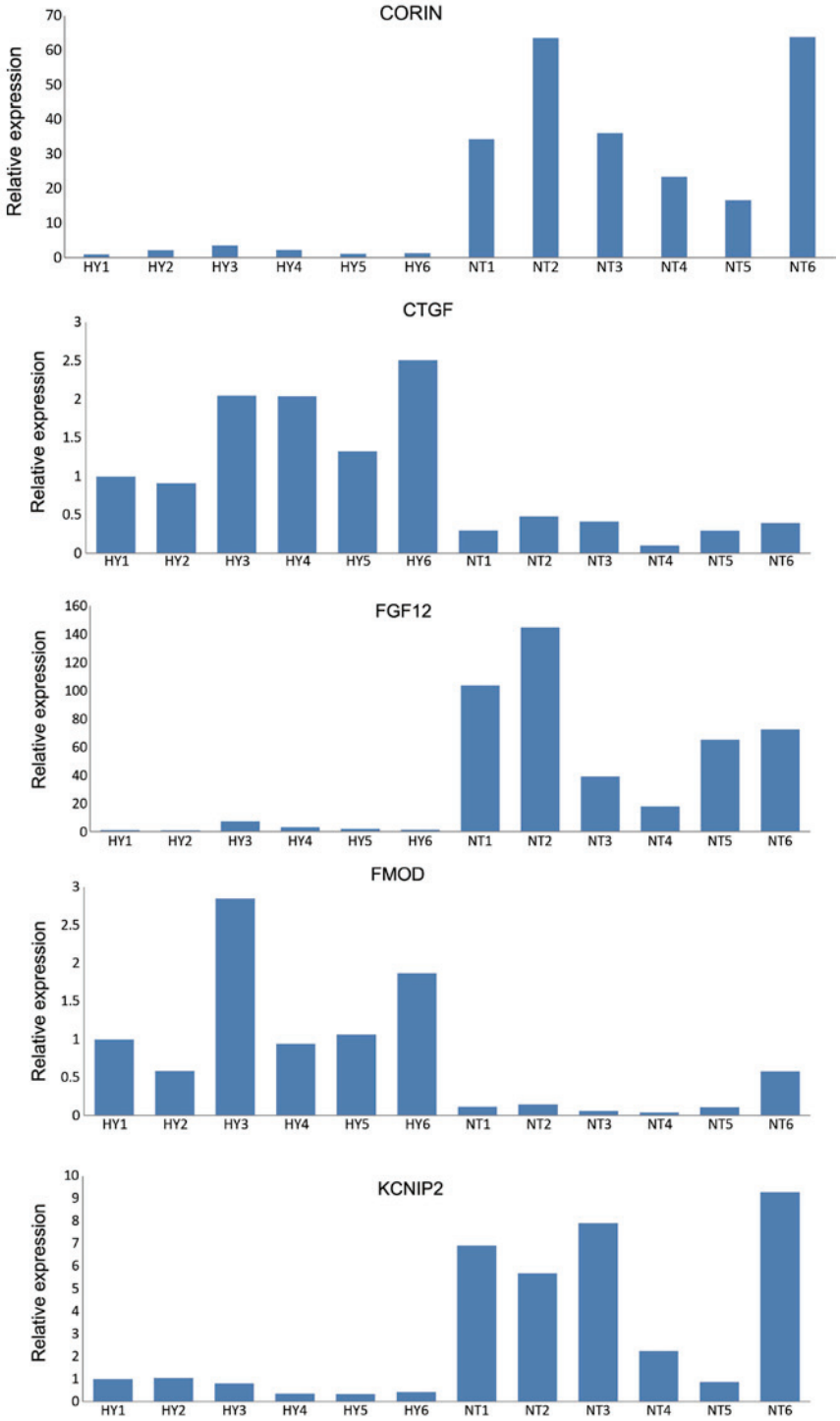

Figure 6. Reverse transcription-quantitative polymerase chain reaction analysis of genes in HCM. NH, healthy control group; HY, patients with HCM. HCM, hypertrophic cardiomyopathy; CORIN, corin, serine peptidase; CTGF, connective tissue growth factor; FGF12, fibroblast growth factor 12; FMOD, fibromodulin; KCNIP2, potassium voltage-gated channel interacting protein 2.

a wide range of ages on the basis of molecular genetics. *CA3* was also differentially expressed between the normal group and the HCM group in the present study, and was upregulated in patients with HCM. *CA3* has previously been demonstrated to be increased in the muscle tissue of hypoxia-trained athletes (35). It is suggested that the observed increase in *CA3* may be a mechanism to act as an anti-oxidizing agent against damaging molecules. Furthermore, the functions of *SEZ6L*, *ESM1* and *COL10A1* in association with HCM were unknown, however, the results of the present study may provide useful information for understanding the human physiology and the genetic etiology of HCM in further research.

CORIN is a membrane-bound serine proteinase distributed in cardiac tissue, involved in modification prior to atrial

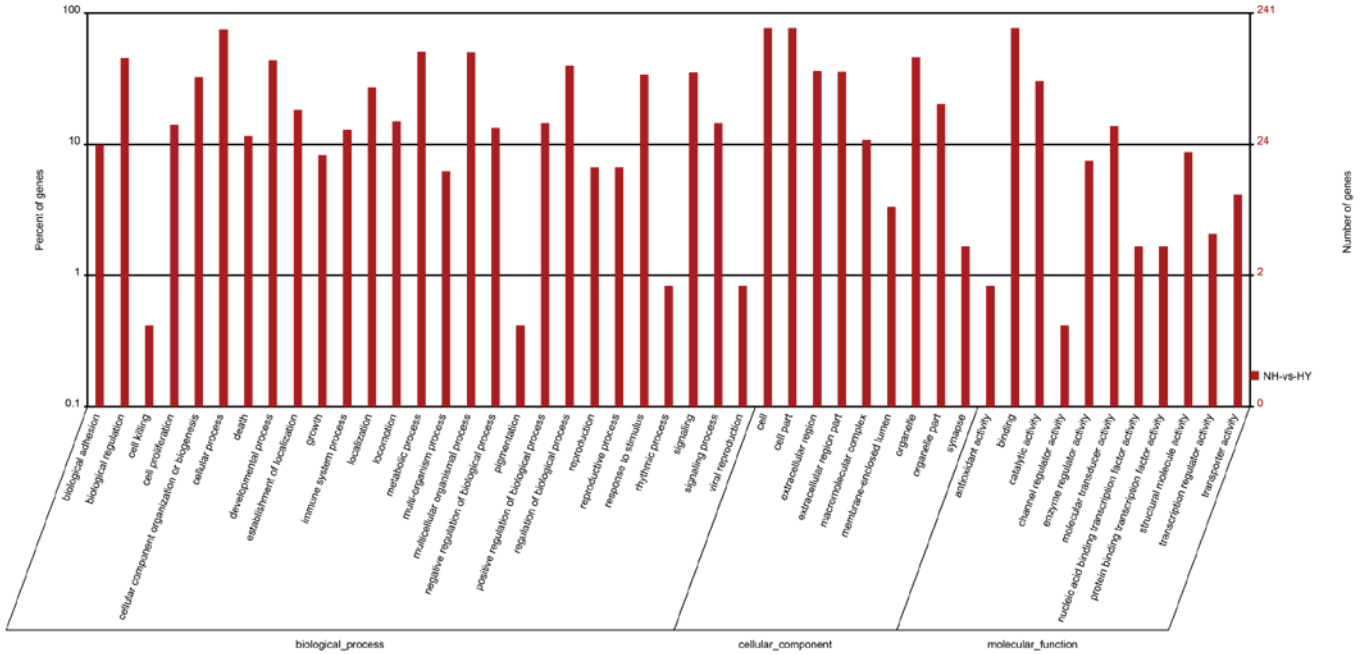

Figure 7. Gene ontology functional classification of differentially expressed genes.

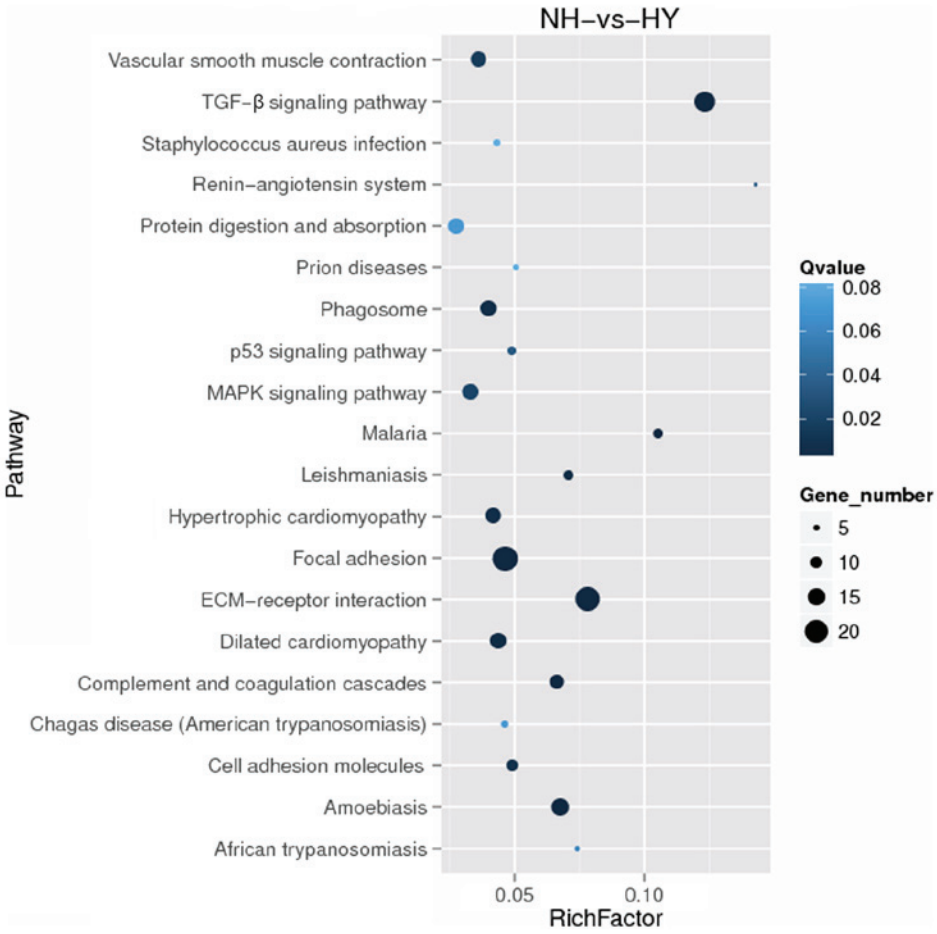

Figure 8. Scatter plot of top 20 KEGG pathway enrichment statistics. KEGG, Kyoto Encyclopedia of Genes and Genomes; NH, healthy control group; HY, patients with HCM; TGF-β, transforming growth factor-β; MAPK, mitogen-activated protein kinase; ECM, extracellular matrix.

peptide precursor release from cells, and is a natriuretic peptide-converting enzyme. The enzyme activates atrial natriuretic peptide, which has important role in regulating blood pressure (36). Thus, RT-qPCR was performed to validate

the RNA-Seq results of *CORIN, CTFG, FGF12, FMOD* and *KCNIP2*. Notably, relative to β-actin expression level, the expression values of *CORIN* ranged from 16-63 in the normal group, and 1-3.59 in patients with HCM. Thus, *CORIN* may be useful as a novel biomarker to detect HCM and patient stratification.

Pathway enrichment analysis of DEGs can identify the most important pathways involved in a disease. The top 20 pathways demonstrated to be associated with the DEGs in the present study are as follows: Vascular smooth muscle contraction; TGF-β signaling pathway; *Staphylococcus aureus* infection; renin-angiotensin system; protein digestion and absorption; prion diseases; phagosome; p53 signaling pathway; mitogen-activated protein kinase signaling pathway; malaria; leishmaniasis; HCM; focal adhesion; extracellular matrix-receptor interaction; dilated cardiomyopathy; complement and coagulation cascades; chagas disease (American trypanosomiasis); cell adhesion molecules; amoebiasis; and African trypanosomiasis. Fig. 8 presents the other pathways in addition to cardiovascular diseases and the circulatory system, such as infectious diseases, signal transduction, endocrine system, digestive system, neurodegenerative diseases, transport and catabolism, cell growth and death, cell communication, signaling molecules and interaction, immune system. Notably, six pathways were associated with infectious diseases; this phenomenon requires further investigation. Recently, a number of genetic studies have indicated that HCM is a genetic disorder associated with numerous mutations. Furthermore, HCM is considered to result from the interaction between pathogenic genes and the environment. The occurrence of HCM may involve multiple independent signaling pathways (37). The specific treatment strategy of HCM requires targeting of specific mechanisms involved in the disease process.

Large-scale transcriptome sequencing comparing the cardiac tissue of patients with HCM and control groups was performed in the present study using an Illumina sequencing platform. A large number of DEGs were identified, which are interesting starting points for investigating the molecular mechanisms of HCM. To the best of our knowledge, this is the first study to demonstrated the changes between patients with HCM and healthy patients at the transcriptome level, which will be important to understand the major molecular mechanisms of HCM and select the key genes to investigate in the future.

Acknowledgements

The authors thank Dr Li Hongliang (Cardiovascular Research Institue; Wuhan University, Wuhan, China) for providing the non-HCM heart tissue samples. This work was supported by The National Natural Science Foundation of China (grant no. 81370328) and the Beijing Natural Science Foundation (grant no. 7132705).

References

1. Frey N, Katus HA, Olson EN and Hill JA: Hypertrophy of the heart: A new therapeutic target? Circulation 109: 1580-1589, 2004.
2. Rosendorff C: Hypertension and coronary artery disease: A summary of the American Heart Association scientific statement. J Clin Hypertens (Greenwich) 9: 790-795, 2007.
3. Frenneaux MP: Assessing the risk of sudden cardiac death in a patient with hypertrophic cardiomyopathy. Heart 90: 570-575, 2004.
4. Artham SM, Lavie CJ, Milani RV and Ventura HO: Obesity and hypertension, heart failure, and coronary heart disease-risk factor, paradox, and recommendations for weight loss. Ochsner J 9: 124-132, 2009.
5. Halabchi F, Seif-Barghi T and Mazaheri R: Sudden cardiac death in young athletes; a literature review and special considerations in Asia. Asian J Sports Med 2: 1-15, 2011.
6. Wang P, Mao B, Luo W, Wei B, Jiang W, Liu D, Song L, Ji G, Yang Z, Lai YQ and Yuan Z: The alteration of Hippo/YAP signaling in the development of hypertrophic cardiomyopathy. Basic Res Cardiol: 109: 435, 2014.
7. Liu X, Jiang T, Piao C, Li X, Guo J, Zheng S, Zhang X, Cai T and Du J: Screening mutations of MYBPC3 in 114 unrelated patients with hypertrophic cardiomyopathy by targeted capture and next-generation sequencing. Sci Rep 5: 11411, 2015.
8. Marian AJ: Contemporary treatment of hypertrophic cardiomyopathy. Tex Heart Inst J 36: 194-204, 2009.
9. Ehlermann P, Weichenhan D, Zehelein J, Steen H, Pribe R, Zeller R, Lehrke S, Zugck C, Ivandic BT and Katus HA: Adverse events in families with hypertrophic or dilated cardiomyopathy and mutations in the MYBPC3 gene. BMC Med Genet 9: 95, 2008.
10. Wang Z, Gerstein M and Snyder M: RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63, 2009.
11. Saliba AE, Westermann AJ, Gorski SA and Vogel J: Single-cell RNA-seq: Advances and future challenges. Nucleic Acids Res 42: 8845-8860, 2014.
12. Cui Y and Paules RS: Use of transcriptomics in understanding mechanisms of drug-induced toxicity. Pharmacogenomics 11: 573-585, 2010.
13. Ozsolak F and Milos PM: RNA sequencing: Advances, challenges and opportunities. Nat Rev Genet 12: 87-98, 2011.
14. 14. Podnar J, Deiderick H, Huerta G and Hunicke-Smith S: Next-generation sequencing RNA-seq library construction. Curr Protoc Mol Biol 106: 1-19, 2014.
15. Wei X, Ju X, Yi X, Zhu Q, Qu N, Liu T, Chen Y, Jiang H, Yang G, Zhen R, *et al*: Identification of sequence variants in genetic disease-causing genes using targeted next-generation sequencing. PloS One 6: e29500, 2011.
16. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K and Wang J: SOAP2: An improved ultrafast tool for short read alignment. Bioinformatics 25: 1966-1967, 2009.
17. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M: Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676, 2005.
18. Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621-628, 2008.
19. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A and Conesa A: Differential expression in RNA-seq: A matter of depth. Genome Res 21: 2213-2223, 2011.
20. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L and Wang J: WEGO: A web tool for plotting GO annotations. Nucleic Acids Res 34 (Web Server issue): W293-W297, 2006.
21. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408, 2001.
22. Pan JB, Hu SC, Shi D, Cai MC, Li YB, Zou Q and Ji ZL: PaGenBase: A pattern gene database for the global and dynamic understanding of gene function. PloS One 8: e80747, 2013.
23. Ho CY: Hypertrophic cardiomyopathy. Heart Fail Clin 6: 141-159, 2010.
24. Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, *et al*: 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: Executive summary: A report of the American college of cardiology foundation/american heart association task force on practice guidelines. Circulation 124: 2761-2796, 2011.
25. Roma-Rodrigues C and Fernandes AR: Genetics of hypertrophic cardiomyopathy: Advances and pitfalls in molecular diagnosis and therapy. Appl Clin Genet 7: 195-208, 2014.
26. Oren A, Toporik A, Biton S, Almogy N, Eshel D, Bernstein J, Savitsky K and Rotman G: hCHL2, a novel chordin-related gene, displays differential expression and complex alternative splicing in human tissues and during myoblast and osteoblast maturation. Gene 331: 17-31, 2004.

27. Wu I and Moses MA BNF-1, a novel gene encoding a putative extracellular matrix protein, is overexpressed in tumor tissues. Gene 311: 105-110, 2003.
28. Kok LD, Tsui SK, Waye M, Liew CC, Lee CY and Fung KP: Cloning and characterization of a cDNA encoding a novel fibroblast growth factor preferentially expressed in human heart. Biochem Biophys Res Commun 255: 717-721, 1999.
29. Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, Dubourg O, Kühl U, Maisch B, McKenna WJ, *et al*: Classification of the cardiomyopathies: A position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. Euro Heart J 29: 270-276, 2008.
30. Tam SK, Gu W, Mahdavi V and Nadal-Ginard B: Cardiac myocyte terminal differentiation. Potential for cardiac regeneration. Ann N Y Acad Sci 752: 72-79, 1995.
31. Varnava AM, Elliott PM, Baboonian C, Davison F, Davies MJ and McKenna WJ: Hypertrophic cardiomyopathy: Histopathological features of sudden death in cardiac troponin T disease. Circulation 104: 1380-1384, 2001.
32. Frey N and Olson EN: Cardiac hypertrophy: The good, the bad, and the ugly. Annu Rev Physiol 65: 45-79, 2003.
33. Heineke J and Molkentin JD: Regulation of cardiac hypertrophy by intracellular signalling pathways. Nat Rev Mol Cell Biol 7: 589-600, 2006.
34. Lu B, Tigchelaar W, Ruifrok WP, van Gilst WH, de Boer RA and Silljé HH: DHRS7c, a novel cardiomyocyte-expressed gene that is down-regulated by adrenergic stimulation and in heart failure. Eur J Heart Fail 14: 5-13, 2012.
35. Zoll J, Ponsot E, Dufour S, Doutreleau S, Ventura-Clapier R, Vogt M, Hoppeler H, Richard R and Flück M: Exercise training in normobaric hypoxia in endurance runners. III. Muscular adjustments of selected gene transcripts. J Appl Physiol (1985) 100: 1258-1266, 2006.
36. Kjaer A and Hesse B: Heart failure and neuroendocrine activation: Diagnostic, prognostic and therapeutic perspectives. Clin Physiol 21: 661-672, 2001.
37. Jiang X, Deng KQ, Luo Y, Jiang DS, Gao L, Zhang XF, Zhang P, Zhao GN, Zhu X and Li H: Tumor necrosis factor receptor-associated factor 3 is a positive regulator of pathological cardiac hypertrophy. Hypertension 66: 356-367, 2015.