Co-expression network analysis identified novel potential Signature Genes Associated with human left ventricle cardiomyopathies arises from different etiologies

Jiao Tian
  Kunming Institute of Botany Chinese Academy of Sciences

Zhengyuan Wu
  Kunming Institute of Botany Chinese Academy of Sciences

Yingying He
  Yunnan University

SHUBAI LIU (✉ liushubai@mail.kib.ac.cn)
  Kunming Institute of Botany Chinese Academy of Sciences

Research article

Abstract

Background: Heart disease is global pandemic and causes huge healthcare burden to society. However, it is still illusive that the whole transcription disorder pattern of cardiomyopathies arises from different etiologies. The Weighted Gene Co-Expression Network Analysis (WGCNA) was applied to construct and screen functional gene that be significantly related to different cardiomyopathies pathological feature. Through co-expression and protein-protein interaction (PPI) networks enrichment analysis, the hub genes and key pathways were screened, which were correlated to cardiomyopathy traits. To discover the novel disease signature genes, cardiovascular disease bioportal database and were employed to blast and validate, which contained independently investigations of clinical cardiomyopathies cases.

Results: The potential disease signature genes were identified and assorted into three common axes shared among five subtype of cardiomyopathies. Four genes (MDM4, CFLAR, RPS6KB1, PKD1L2) were shared by ischemic and ischemic cardiomyopathy group. The secondary axe contained eight signature genes (MAPK1, MAPK11, MAPK14, LMNA, RAC1, PECAM1, XIAP, CREB1) and was overlapped by Ischemic Cardiomyopathy, Post. Partum Cardiomyopathy, Familiar Cardiomyopathy and Idiopathic Cardiomyopathy. The third axe consisted of two common signature genes (TFAM, RHEB) that shared among the subgroups of Viral Cardiomyopathy, Post. Partum Cardiomyopathy, Familiar Cardiomyopathy and Idiopathic Cardiomyopathy. The majority of disorder functions and pathways were enriched in metabolic processes and pathways of MAPK signaling, protein processing in endoplasmic reticulum, regulation of actin cytoskeleton.

Conclusion: These results strongly suggest that expression disorder of signature genes contribute to the cardiac dysregulation and functional relapse into cardiomyopathies. Taken together, these novel signature genes could be utilized as potential diagnostic biomarkers or therapy targets. It will be benefit the cardiomyopathy precise clinical diagnostics with better outcome. In summary, this study will attract great interest of clinical research scientists as well as medical scientists that work on heart diseases.

Background

Cardiomyopathy is a pathological syndrome that featured with the incapacity of heart to pump and/or fill with blood for body need, and the worst state lapses into heart failure. Heart failure is complex pathophysiological condition with left ventricle myocytes dysfunction. It is caused by many physiological or pathologic processes, such as ischemia, hypertension, pregnancy and diabetes mellitus. As a global pandemic, at least 26 million people were suffered from heart failure and consumed over $30 billion health expenditure in the world [1]. Furthermore, the mortality of heart failure patients is higher than 50% in 5-year [2]. Many historical binary classification systems, like termed with idiopathic, were simplicity and clarity, but they may represent an oversimplification assorting for a complex biologic phenomenon [3]. As a group of heterogeneous complex cardiovascular diseases, the typical features of cardiomyopathies are primary abnormalities in the physiological structure and function of the heart. According to the morphological features and different etiologies, cardiomyopathies are commonly grouped into several

subtypes, including hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), viral cardiomyopathy (VCM), familial cardiomyopathy (FCM), post-partum cardiomyopathy (PCM) and ischemic cardiomyopathy (ISCM) [4]. This new classification scheme is potentially useful to draw relationships between enigmatic etiology and dysfunctional cardiovascular disease, and to promote better understanding of disease.

Early clinical investigations demonstrated that some subtypes of cardiomyopathies were derived from patient's cases with dysregulated gene expression profile, which has initial normal physiological condition and somatic genetics background. It strongly suggests that different etiologies paly critical role in epigenetics transcription change [5]. Recently, to enhance the patient's treatment and healthcare management, it is the trend to discover clinical applicable disease signature genes or biomarker for precise diagnostics through analyzing the genetic disorder and expression profiling of heart failure [6, 7]. Among multiple computerization methodologies, the Weighted Gene Co-Expression Network Analysis (WGCNA) is considered as one of the most useful approaches to discover gene co-expression network based on functional feature through gene expression profiling analysis [8]. Furthermore, WGCNA has been widely applied to screen the novel biomarkers or therapeutic targets for cancer early diagnostics and treatment, like hepatocellular carcinoma [9] and lung cancer [10].

In this study, the WGCNA has been employed to analyze gene expression profiles of several subtypes of cardiomyopathies and discover the highly connected modules that genes significance associated with heart failure. The novel signature genes and key biological pathways were identified and validated through cardiovascular disease bioportal, and were listed as potential novel drug targeted for cardiomyopathies. These results will be helpful to better understand the cardiomyopathies pathogenesis, progression and prognosis, and be benefit for patient's precise diagnostics and treatment.

Results

Identified Gene Modules Correlation with Cardiomyopathies Trait

In this study, WGCNA was applied to investigate the whole transcription profiling of heart failure patient's biopsy arise from eight different etiologies (90 cases, 8 case groups sorting, Table. S1, Fig.1, Fig.S1). It was reasonable to build the co-expression networks with different clinical traits using the Pearson correlation analysis (Fig.2A). The correlation among cardiomyopathies and eight different pathological features were separated by performed hierarchical clustering. To discover the correlated modules to cardiomyopathies phenotype, the genes significance of modules was calculated by the linear mixed effects model to test the association of node to the pathological phenotypes. Next, the associated significance was analyzed, which represents the identified individual gene modules associated with different subtype of cardiomyopathy, respectively (Fig.2B). Calculated with a dynamic tree cutting algorithm, the distinct co-expression modules were identified that significantly related to different pathological features (Fig.2B & Fig.3A). Twenty-three modules were detected through the dataset (Fig.2B,

Table.1). The numbers of significant genes containing in modules were varied from 121 to 14938. The correlation significance of module and pathological features was determined by module significance (MS) correlation and statistics p-value, and significant module varied from different subtype cases. The higher value of module eigengene (ME) correlation, the module is closer correlated to cardiomyopathies (Fig.S3A-G). Through calculation of the linear mixed-effects model, significant modules were identified for specific pathological feature (Fig.3B). In the idiopathic dilated group, six modules were significantly associated with idiopathic dilated status, including turquoise module (t-value = 0.75, p-value = 5e−17), lightcyan module (t-value = 0.54, p-value = 7e−08), tan module (t-value = 0.56, p-value = 2e−08), grey module (t-value = 0.7, p-value = 8e−14), green module (t-value = 0.45, p-value = 1e−05) and the blue module (t-value = 0.72, p-value = 8e−15) (Fig.2B, Fig.3B, Table.1). In ischemic group, six significant modules were identified, including lightyellow module (t-value = 0.43, p-value = 3e−05), greenyellow module (t-value = 0.41, p-value = 8e−05), lightgreen module (t-value = 0.4, p-value = 1e−04), red module (t-value = 0.35, p-value = 8e−04), green module (t-value = 0.24, p-value = 0.03) and the blue module (t-value = 0.44, p-value = 2e−05) (Fig.2B, Fig.3B, Table.1). In idiopathic cardiomyopathy (IdCM) group, three modules were significantly linked to pathological trait, listed as module of Magenta (t-value = 0.43, p-value = 4e−05), Purple (t-value = 0.31, p-value = 0.003) and Brown (t-value = 0.25, p-value = 0.02). In familial cardiomyopathy group, only Magenta module (t-value = 0.22, p-value = 0.04) was significantly correlated to this trait (Fig.2B, Fig.3B). For the Hypertrophic cardiomyopathy (HCM) group, none of module was identified and ignored in next step analysis (Fig.2B). In the Post. Partum cardiomyopathy (PCM) group, Magenta module (t-value = 0.24, p-value = 0.03) was correlated to its trait (Fig.2B, Fig.3B, Table.1). In ischemic cardiomyopathy group, seven modules were significantly correlated to pathological trait, including Black (t-value = 0.56, p-value = 2e−08), Midnightblue (t-value = 0.49, p-value = 2e−06), Darkred (t-value = 0.43, p-value = 3e−05), Cyan (t-value = 0.34, p-value = 0.02), Yellow (t-value = 0.57, p-value = 1e−08), Pink (t-value = 0.29, p-value = 0.007) and Red (t-value = 0.26, p-value = 0.02) (Fig.3B). For Viral cardiomyopathy (VCM), two significant modules, Cyan (t-value = 0.35, p-value = 0.001) and Magenta (t-value = 0.23, p-value = 0.03), were identified (Fig.2B, Fig.3B).

Enriched Genes Significance related to different cardiomyopathies

Compared the module memberships (MM) correlation and Genes Significance (GS) among the all significant modules, the module with most significant value was defined as the best candidate for pathological traits correlation analysis (Table.1, Fig.3B), respectively. These candidates were listed as turquoise module (cor = 0.77, p < 1.0e−200, GS=0.4244) (Idiopathic dilated group, Fig.S3A), lightyellow module (cor = 0.13, p = 3.0e−05, GS=0.2733) (Ischemic group, Fig.S3B), magenta module (cor = 0.65, p = 1.5e−114, GS=0.2783) (Idiopathic cardiomyopathy group, Fig.S3C), magenta module (cor = 0.31, p = 1.7e−22, GS=0.1502) (Familial cardiomyopathy group, Fig.S3D), magenta module (cor = 0.30, p = 4.2e−21, GS=0.1583) (Post. partum cardiomyopathy group, Fig.S3E), yellow module (cor = 0.63, p < 1.0e−200, GS=0.3259) (Ischemic cardiomyopathy group, Fig.S3F) and cyan module (cor = 0.39, p = 8.2e−24, GS=0.2245) (Viral cardiomyopathy group, Fig.S3G), respectively. In addition, the scatter plot of

multiple module memberships (MM) was plotted against the Genes Significance (GS) in each significant module, and the point was represented each gene contained in a module.

Hierarchical Clustering of Eigengene Profiles with cardiomyopathies Traits

Based on the ME's values, the hierarchical clustering was performed between all modules and different cardiomyopathies traits to identify their relationships. Furthermore, the Eigengene dendrogram analysis was performed to build the correlation of candidate module with different subtypes of cardiomyopathy feature, respectively (Fig.S4A-G). In the idiopathic dilated group, turquoise module was tightly clustered with idiopathic dilated (Fig. S4A). In the ischemic group, modules of greenyellow and lightyellow were the closest branch clustered with ischemic (Fig.S4B). In previous step analysis, greenyellow module (t-value = 0.41, p-value = 8e−05, GS = 0.2418) was the secondary higher correlation with ischemic status (Fig.2B, Fig.3B). It suggests that genes containing in greenyellow module would be involve the progress of ischemic. In idiopathic cardiomyopathy group, modules of brown, magenta and purple were clustered with idiopathic cardiomyopathy in a separate branch, and magenta and purple module allocated in same cluster (Fig.S4C). It suggests that module of purple and magenta would be the top two significant associated with cardiomyopathy status (Fig.2B, Table.1). In the familial cardiomyopathy and post. partum cardiomyopathy groups, modules of brown, magenta and purple were tightly clustered with familiar cardiomyopathy, while the magenta module was the most significant associated with disease status (Fig.S4D-4E). In the hypertrophic cardiomyopathy group, no module was associated with its pathological feature. In the ischemic cardiomyopathy group, although module of black and midnightblue were clustered in closer branch, the yellow module was allocated in the adjacent branch (Fig.S4F). Combined with the module-trait relationship correlation and gene significance results, it suggested that module of black, midnightblue and yellow were significant associated with ischemic cardiomyopathy (Fig.2B, Fig. 3B, Table. 1). The module of brown and blue were ignored for next analysis as with higher negative GS values (Table.1, Fig. 2B). The yellow module was the most significant correlation to ischemic cardiomyopathy. In the viral cardiomyopathy group, the modules of magenta, purple and brown were clustered with viral cardiomyopathy in a separated branch, and cyan module had the highest GS value associated with pathological feature (t-value = 0.35, p-value = 1e−03, GS = 0.2245) (Fig.2B, Fig. 3B, Supporting Fig.4G). It suggested that magenta module containing genes involve the progress of viral cardiomyopathy.

Functions and pathways enrichment analysis

   Blast through GenClip2, the enrichments analysis of significance genes contained in the correlated module were summarized as bar chart, including Biological Process, Molecular Function and Cellular Component (Ischemic, Fig.S5A-A; Idiopathic Cardiomyopathy, Fig.S5B-A; Familial Cardiomyopathy, Fig.S5C-A; Post-Partum Cardiomyopathy, Fig.S5D-A; Ischemic Cardiomyopathy, Fig.S5E-A; Viral

Cardiomyopathy, Fig.S5F-A). The Biological Processes were mainly concentrated in subgroups of cellular macromolecular metabolic process, protein metabolic process, organic substance metabolic process and macromolecule modification. The genes significant enriched in molecular functions were summarized and listed (Table.S4). The molecular functions were linked to endoplasmic reticulum functions, cell functions (migration, death, growth, division), DNA binding, protein-protein interaction, kinases activity and signal transduction, iron binding and nucleotide binding, etc. The cellular components were mainly enriched in membrane-bounded organelle, intracellular organelle part, etc. The pathways concentrated on mitogen-activated protein kinase (MAPK) signaling pathway, protein processing in endoplasmic reticulum, regulation of actin cytoskeleton, etc. These results suggested that dysregulation of cardiac functions would be associated with metabolism abnormal and accelerated progress of cardiomyopathies. Furthermore, through gene network literature mining and clustering analysis, these significance genes were clustered and labelled according to cellular functions and pathological feature keywords literature corresponding (Ischemic, Fig.S5A; Idiopathic Cardiomyopathy, Fig.S5B; Familial Cardiomyopathy, Fig.S5C; Post-Partum Cardiomyopathy, Fig.S5D; Ischemic Cardiomyopathy, Fig.S5E; Viral Cardiomyopathy, Fig.S5F). Moreover, the network and connectivity of significance genes were identified as node-connection map, which was correlated to different traits (Ischemic, Fig.S5A-B; Idiopathic Cardiomyopathy, Fig.S5B-B; Familial Cardiomyopathy, Fig.S5C-B; Post-Partum Cardiomyopathy, Fig.S5D-B; Ischemic Cardiomyopathy, Fig.S5E-B; Viral Cardiomyopathy, Fig.S5F-B). The key regulatory genes were labelled with purple border, which was reported in pathological cases, and majority of genes involved the pathway were labelled as un-reported. More interesting, no related key regulatory gene was identified in Post-Partum cardiomyopathy group.

Identification of hub genes for cardiomyopathies

Through co-expression network (MM-GS) filtered, the candidates of hub gene were identified in different groups (5 genes in Ischemic group, Fig.4A; 113 genes in Idiopathic Cardiomyopathy group, Fig.4B; 41 genes in Familiar Cardiomyopathy group, Fig.4C; 65 genes in Post-Partum Cardiomyopathy group, Fig.4D; 83 genes in Ischemic Cardiomyopathy group, Fig.4E; 60 genes in Viral Cardiomyopathy group, Fig.4F). Calculated by the PPI network method, the candidates of hub gene were summarized (12 genes in Ischemic group, Fig.4A; 120 genes in idiopathic cardiomyopathy group, Fig.4B; 277 genes in Familiar Cardiomyopathy group, Fig.4C; 119 genes in Post-Partum Cardiomyopathy group, Fig.4D; 348 genes in Ischemic Cardiomyopathy group, Fig.4E; 49 genes in Viral Cardiomyopathy group, Fig.4F). The real hub genes were determined as described in method section (Table.3), and the numbers of real hub genes were listed (Fig. 4 A-F). The Idiopathic Dilated group was dismissed for further analysis as no identified real hub gene.

Through Venn diagrams analysis, three common axes of hub genes were discovered among these cardiomyopathies groups (Fig.4G). The first axis was **PICALM,** which shared by Ischemic Cardiomyopathy, Idiopathic Cardiomyopathy and Post. Partum Cardiomyopathy groups (Fig.4G), and

significantly up-regulated in Idiopathic Cardiomyopathy and Ischemic Cardiomyopathy groups (Table.3). PICALM is key regulator in iron homeostasis, clathrin-mediated endocytosis [11, 12]. Overexpression of PICALM impaired endocytosis of Transferrin (Tf) Receptor (TfR) and Epidermal Growth Factor Receptor (EGFR) and disturbed the iron homeostasis [12, 13]. Up to now, it is still illusive that the exactly role and deregulatory mechanism of PICALM in cardiomyopathies. It is strongly suggesting that PICALM work as potential novel biomarker and therapy target for these subcases of cardiomyopathies. The secondary axis, contained genes of PRKACB, MOB1A, CDC40, were shared in Post. Partum Cardiomyopathy and Idiopathic Cardiomyopathy groups. In addition, these genes (PRKACB, MOB1A, CDC40) were significantly overexpressed in Idiopathic cardiomyopathy group, and MOB1A was up-regulated in Post. Partum cardiomyopathy group (Table.3).These genes were linked to the cAMP (cyclic AMP)-dependent protein kinase A (PKA) mediated the exciting-contraction coupling in cardiomyocytes [14], and regulated microtubule stability, cell cycle and cell proliferation & migration, and restrained cardiomyocyte proliferation and size via Hippo pathway [15, 16]. *PRKACB* (protein kinase cAMP-activated catalytic subunit β gene) was linked to congenital heart defect with abnormal over-expression [17]. **MOB1A** (MOB kinase activator 1A) was required for cytokinesis through regulating microtubule stability. It worked as binding partners as well as co-activators of Ndr family protein kinases and mediated phosphor-recognition in core Hippo pathway that restrains cardiomyocyte proliferation during development to control cardiomyocyte size[15, 16]. Overexpression of MOB1A induces centrosomes fail to split and cell size dysregulation[18]. **CDC40** (Cell Division Cycle 40), a splicing factor of cell division cycle 40 homolog, regulates cell cycle and cell proliferation and migration[19]. Overexpression of CDC40 causes abnormally cell proliferation and migration, and linked with carcinogenesis[20]. The third axis consisted of five genes (CREB1, DBT, NCOA2, NUDT21, PIK3C2A) and were overlapped among three groups of Familial / Idiopathic / Post. Partum Cardiomyopathy (Fig.4G). The CREB1 (cAMP-responsive element-binding protein) had been identified as the transcription factor and mediated cAMP stimulation by multiple extracellular signals, such as growth factors and hormones. The CREB1 was the key regulator in heart and linked with heart disease via cAMP-PKA pathway dysregulation [21, 22]. The DBT (dihydrolipoamide branched chain transacylase E2) is an inner-mitochondrial enzyme complex regulated to degrade the branched-chain amino acids isoleucine, leucine, and valine[23]. The DBT was reported as clinical diagnostics biomarker for patients with dilated cardiomyopathy via caused mitochondria dysfunction [24]. NCOA2 (nuclear receptor coactivator 2) is a transcriptional coactivator that functional aid for nuclear hormone receptors, including steroid, thyroid, retinoid, and vitamin D receptors. *NCOA2* promotes muscle cells maintenance and growth, eventually regulates in cardiac cTnT levels[25, 26]. Overexpression of NCOA2 regulated cell proliferation in cardiomyopathy[26, 27]. NUDT21 (nudix hydrolase 21) is a novel of cell fate regulator by alternative polyadenylation chromatin signaling, and suppression of NUDT21 will enhance the cell pluripotent, facilitated trans-differentiation into stem cell[28]. NUDT21 regulates cell proliferation through ERK pathway[29]. Up to now, little knows about the function of NUDT21 in cardiomyocytes. PIK3C2A (phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 alpha) is an enzyme belong to phosphorylate the 3'-OH of inositol ring of phosphatidylinositol (PI) superfamily and regulates multiple signaling pathways. PIK3C2A is mainly expressed in endothelial cells, vascular endothelium, and smooth muscle[30]. Lower expression of PIK3C2A in peripheral blood was used as

significant biomarker for acute myocardial infarction patients [31]. More interesting, these hub genes indicated different expression pattern. The expression level of DBT, NCOA2, NUDT21 and PIK3C2A were significantly upregulated in Idiopathic cardiomyopathy group, and PIK3C2A was up-regulated in Familiar cardiomyopathy group (Table.3). It hints that these hub genes play different regulatory pattern in the progress of these subtype's cardiomyopathies. The fourth axis of hub genes (HNRNPC, UEVLD) were shared by Familiar Cardiomyopathy and Idiopathic Cardiomyopathy groups, and significantly overexpressed in Idiopathic cardiomyopathy group (Table.3). HNRNPC (heterogeneous nuclear ribonucleoprotein C) is RNA binding protein that belong to ubiquitously expressed heterogeneous nuclear ribonucleoproteins subfamily, and mediates pre-mRNAs transport and metabolism between cytoplasm and nucleus [32, 33] and overexpression caused cells multi-nucleation [34]. UEVLD (EV and lactate/malate dehydrogenase domain-containing protein) involves the protein degradation and dysregulated linked with metabolic disease [35]. In this study, the expression level of HNRNPC and UEVLD were significantly up-regulated in Idiopathic cardiomyopathy group (Table.3). Furthermore, through different significant expression analysis, the significant changed hub genes were summarized (Table.3, p<0.05). Combined these results together, it hints that these significantly expressed Hub genes play dominant role and work as common key regulatory nodes in progress of cardiomyopathies.

Disease Signature Genes Identification and expression analysis

The filtered disease signature genes were summarized with the functional annotation of genetic dysregulation correlated to heart diseases phenotypes, including ten signature in the ischemic group and viral cardiomyopathy group, forty signature genes among the groups of familiar cardiomyopathy, Post-partum cardiomyopathy and Idiopathic cardiomyopathy, and 69 signature genes in the ischemic cardiomyopathy group (Table.4). Through Venn Diagram analysis, the common signature genes were determined among different groups (Fig.5A). Four signature genes (MDM4, CFLAR, RPS6KB1, PKD1L2) were shared by Ischemic and Ischemic Cardiomyopathy group (Table.4, Fig.5A) [36-39]. Ischemic cardiomyopathy group did share eight disease signature genes (MAPK1, MAPK11, MAPK14, LMNA, RAC1, PECAM1, XIAP, CREB1) with Post. Partum/Familiar/Idiopathic Cardiomyopathy groups, which dysregulated in cardiomyopathies [22, 40], and genes expression level of MAPK1, MAPK11, LMNA, RAC1 were significantly up-regulated in these cardiomyopathies groups (Table.5, Fig.5A). Two signature genes (TFAM, RHEB) were shared between Viral Cardiomyopathy and Post. Partum / Familiar /Idiopathic Cardiomyopathy groups, which involved in development of cardiac hypertrophy [41] and Mitochondrial Cardiomyopathy [42].

   Furthermore, through different significant expression analysis, the significantly changed signature genes were summarized (Fig.5D-F, Table.5). In ischemic group, MDM4 gene was significantly upregulated (FC=1.0495, p=0.0037) (Table 5, Fig.5B), which genetic deletion associated with cardiomyopathy [39]. In viral cardiomyopathy group, COA5 was overexpressed (FC=1.087485, p<0.0001) (Table 5, Fig.5C), which was upregulated in Ischemia/Reperfusion Injury caused cardiomyopathy [43]. In

idiopathic cardiomyopathy group, seven genes (ADAM10, RAB1A, TFAM, FGF2, ELMOD2, GUF1, ABCC9) were significantly up-regulated, while 8 genes (FHL1, CTNNA3, PDLIM5, LMNA, SIRT4, YME1L1, RHEB, GNB1L) were down-regulated (Fig.5D, Table.5). In idiopathic cardiomyopathy group, four down-regulated genes (NEBL, FHL1, FHL2, SIRT4) and 5 up-regulated genes (GUF1, ELMOD2, ABCC9, FGF2, YME1L1) were significantly changed (Fig.5E, Table. 5). In post-partum cardiomyopathy group, 11 genes (ATL3, ADAM10, ELMOD2, FGF2, GUF1, YMEIL1, up-regulated; FHL1, CTNNA3, PTPN11, GNB1L, SIRT4, down-regulated) were significantly changed (Fig.5F, Table.5). In Ischemic cardiomyopathy group, 31 genes are significantly changed expression, including 12 genes (RAC1, FKTN, EDNRB, ZBTB33, TXN, RALGAPA1, PSEN1, LAMP2, UBR5, SCN4B, SMAD1, MYO6) down-regulated and 19 genes up-regulated expression (POLRMT, AVP, GATA4, CACNB2, MAPK1, NOS3, LAMA3, SOD2, LMNA, MAP1LC3A, MAPK14, TCAP, LRP4, BAD, DES, AKAP8, CASP9, HSPB1, SNTA1) (Fig.5G, Table.5). These results suggest that these common disease signature genes work as novel biomarker and be potential key regulators of the cardiomyopathy progress.

Discussion

In this study, to discover novel signature genes or biomarkers to accelerate the precise clinical diagnostics and interference for cardiomyopathies, the WGCNA pipeline was applied to analyze the gene expression profiling of 90 clinical left ventricle biopsy samples, which represents 8 subtype's cardiomyopathies. The whole transcriptome profile contained 20,283 target genes for promise diagnostic assessment and mainly covered variously biological and cellular processes. It is representative of real pathological satiation and valuable to discover the signature gene of cardiomyopathies. First of all, it was reasonable to build the co-expression networks with different clinical cardiomyopathies traits using the Pearson correlation (Fig. 2A). To discover the related modules to cardiomyopathies phenotype, the genes significance of the modules was calculated by the linear mixed effects model for testing the association of node to the pathological phenotypes. It was identified that the association significance between individual modules of gene expression profile and different cardiomyopathies feature (Fig. 2B). Through the Eigengene dendrogram analysis, the most significantly module was pick out for next analysis (Fig. 3B, Supporting Fig. 4A-G). For the next step, the real hub genes among each significant module were screened by module membership (MM) – Gene Significance and Protein-Protein Interaction Network analysis, and comprised as key interconnected nodes within a functionally network and played important roles in biological functions[44]. There was without any real hub genes identified in the Idiopathic Dilated group and discarded for next analysis. In addition, the Idiopathic Dilated case was treated as unique physiological state without impaired the normal cardiac function and ignored for analysis. Briefly, the next analysis was mainly concentrated on last five subtype's groups, including idiopathic cardiomyopathy (IdCM), familial cardiomyopathy (FCM), post-partum cardiomyopathy (PCM), Ischemic cardiomyopathy (IsCM) and viral cardiomyopathy (VCM). There were four axes of hub genes shared among these cardiomyopathies groups. It was suggesting that these Hub genes work as common key regulator. It was exception that viral cardiomyopathy group did not share hub gene with the others groups. It was possible

unique that the dysregulation expression pattern of viral cardiomyopathy. Furthermore, to deeply dig the correlation of significance genes and different cardiomyopathies, the significance genes were blast through GenCliP2 to mine gene networks and functions connection. The enrichment analysis results contained the biological process, molecular functions, the cellular components and functional pathways. Although the enriched pathways were varied from different subtype's cardiomyopathies group, the key pattern was similar (Supporting Fig. 5A-F). The biological processes were significantly concentrated on cellular metabolic, protein metabolic, organic substance metabolic and macromolecule modification. It suggests that the metabolic process disorder associate with cardiomyopathies. The molecular functions were mainly involving in protein binding, heterocyclic compound binding, purine ribonucleotide binding, iron binding and nucleotide binding, etc. The cellular components were including membrane-bounded organelle, intracellular organelle part, etc. The pathways were concentrated on MAPK signaling pathway, protein kinase C protein processing in endoplasmic reticulum, regulation of actin cytoskeleton, etc (Supporting Fig. 6A-F). These related genes were summarized with functions and pathways (Supporting Table. 4A-F). Furthermore, most of significance genes were labelled as gene-term association not reported. It represents that the regulatory mechanism of these genes are illusive in progress of cardiomyopathies. Through Literature Gene Networks Mining, the genes reported in the regulatory network were identified, which associated with different cardiomyopathies except Post-Partum cardiomyopathy (Supporting Fig. 5D-B). It is possible that less researches and few reports concentrated on Post-Partum cardiomyopathy. For subtype of idiopathic cardiomyopathy (IdCM), familial cardiomyopathy (FCM), and Ischemic cardiomyopathy (IsCM), the highlighted genes were mainly concentrated in MAPK signaling pathway, including MAPK1, MAPK14, CREB1, RAC1. The genes YY1, RAPGEF1, SMAD2, JUND, ATF1, and SRA1 linked with viral cardiomyopathy (Supporting Fig. 5F-B), which are assemble in SMAD signaling pathway [45] and YY1 was overexpressed in heart failure [46]. These results partially matched the key axes of hub genes linked the functions and pathways. These new discovered genes linked to viral cardiomyopathy may give a new view for clinical diagnostics and treatment. It suggests that the dysfunctions of these significance genes associate with metabolism disorder and progress of cardiomyopathies.

In this study, limited by clinical samples accession, blast through the cardiovascular disease BioPortal database was used as validation strategy to explore the significance genes associated with different subtype cardiomyopathies. The number of disease signature genes filtered were varied from different groups (Table.4). Through overlapping analysis, there were three axes of common signature genes identified among five subtype's cardiomyopathies groups (Fig. 5A). The first axis contained 4 disease signature genes (MDM4, CFLAR, RPS6KB1, PKD1L2) shared by ischemic and ischemic cardiomyopathy group. Compared between health and ischemic groups, only MDM4 was significantly overexpressed (FC = 1.0495, p = 0.0037) in ischemic group, while the four signature genes did not significantly change in ischemic cardiomyopathy group. It matches the previously reports that the upregulated MDM4 plays cardio-protective effect in ischemia − refuse injury [47]. Overexpression of MDM4 reflects the self-correction of physiological system in abnormal physiological condition of ischemic. It is a new strategy to develop the interference the ischemic lapse into cardiomyopathies by artificial upregulated MDM4

expression. The secondary axis consisted of eight signature genes (MAPK1, MAPK11, MAPK14, LMNA, RAC1, PECAM1, XIAP, CREB1) and shared by Ischemic Cardiomyopathy with Post. Partum/Familiar/Idiopathic Cardiomyopathy groups. The signature genes (MAPK1, MAPK11 and LMNA down-regulated; RAC1 up-regulated) were significantly changed in ischemic cardiomyopathy group, and only LMNA was significantly down-regulated in Idiopathic Cardiomyopathy group (Table.5, Fig. 5G). These results hint that the disorder expression pattern of these three cardiomyopathies were more complex, although they shared common signature genes. The MAPK pathway plays dominant role in progress of ischemic cardiomyopathy. The signature genes (TFAM, RHEB) were shared by Viral Cardiomyopathy and Post. Partum / Familiar /Idiopathic Cardiomyopathy groups and worked as the third axis. The expression comparison results indicated that only RHEB significantly down-regulated in Idiopathic cardiomyopathy group. It suggests that signature genes (TFAM, RHEB) play less contribution in pathological progress of viral cardiomyopathy and Post. Partum / Familiar /Idiopathic Cardiomyopathy. Combined these results together, it strongly suggests that these genes could be used as promising biomarkers or therapy targets for cardiomyopathies. This progress will be helpful to integrate precise clinical application for different subtype cardiomyopathies.

There are some limitations in this study. Firstly, lacking of large number of clinical patient's samples and more detail of clinical information, it was impossible to track these samples pathological feature with expression profiles and verify these potential biomarkers with original patient's pathological feature. Secondly, due to the nature of bioinformatics analysis, the discovered specific GO pathways and biomarkers were not further investigated. Although validated these genes significant associated with cardiomyopathies feature through cardiovascular disease BioPortal database, and it is necessary to verify these potential biomarkers with more clinic patient's biopsies. It is mandatory and rational to investigate the function and mechanism of these potential biomarkers or targets for cardiomyopathies with animal model and validate with clinical data in the future research.

Conclusions

In summary, this study provides new insight to identify the potential novel key regulatory biomarkers or therapy targets for varied cardiomyopathies induced by different etiologies. The disease signature genes associated with cardiomyopathies were identified and listed as the potential therapy targets for clinical application. In the future research, the detail of regulatory mechanism of these disease signature genes will be deeply investigated and develop novel therapy strategy for cardiomyopathies.

Methods

Clinical Samples information

To establish cardiac transcription profiles of cardiomyopathies, the human left ventricle samples were collected from patient's biopsy undergoing cardiac transplantation whose failure arises from different etiologies (e.g. idiopathic dilated cardiomyopathy, ischemic cardiomyopathy) and from "normal" organ

donors whose hearts cannot be used for transplants. The transcriptional profiles of these samples were measured by Affymetrix Human Genome U133 Plus 2.0 Array. The Changes in transcriptional profiles were correlated with the physiologic profile of heart-failure hearts acquired at the time of transplantation. These samples and data were generated by cardio-genomics lab, Department of Bauer Center for Genomic Research, Harvard University.

Original Gene Expression Data

The gene expression data used in this paper was obtained from the Gene Expression Omnibus (GEO) database in NCBI (Gene Expression Omnibus), and the platform entry number is GPL570. The datasets used for data analysis is available in GEO with tracking number GSE1145 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1145), which was derived from 90 patient's left ventricle biopsies. The datasets contained 20,283 target genes for further analysis (Figure.1, Supporting Table.1). The dataset's probe-sets was annotated with gene symbols, related function, and clinical traits. The biopsy samples were represented 8 subgroups, including normal hearts as control (Health, n=11), and idiopathic dilated (ID, n=15), ischemic (IS, n=11), idiopathic cardiomyopathy (IDCM, n=12), familial cardiomyopathy (FCM, n=5), post-partum cardiomyopathy (PCM, n=4), ischemic cardiomyopathy (ISCM, n=20), viral cardiomyopathy (VCM, n=7), respectively.

Data preprocessing

The GSE1145 dataset was downloaded, and each gene expression values were normalized and performed with log2 transformation. Each probe-set was linked with gene symbol through the Affymetrix annotation file GLP570 (Affymetrix Human Genome U133 Plus 2.0 Array), which contains total of 54,675 probes that correspond to 20,283 genes. The microarray quality was assessed by sample clustering according to the distance between different samples in Pearson's correlation matrices, and a height cut of 170000 was chosen to identify potential microarray outliers. Four normal samples (GSM18444/18445/18447/18448) were detected as outliers and ignored in the subsequent analysis (Supporting Fig.1).

Construction of Weighted Gene Co-Expression Network

The WGCNA package of R (version 1.63) was download and setup by following the protocol described previously [48]. The WGCNA package was used for performing various functions in weighted correlation network analysis, including constructing network, detecting module, calculating topological properties, simulating data, visualization, and interfacing with external software [48]. First of all, data has been checked to exclude the sample with excessive missing values and identification of outlier samples. After data preprocessed, the principal component analysis (PCA) has been applied to double check the data quality. The heart failure and health samples were separated in the PCA plot (Supporting Fig. 1), and the hierarchical clustering on the samples was performed to detect potential outliers. The total 86 samples were used for next step analysis (Figure. 1). The soft threshold $\beta = 7$ was chose to construct the co-expression network as the R^2 reached the peak for the first time when $\beta = 7$. The plot of log10(p(k))

versus log10(k) (Supporting Fig. 2) indicated that the network was close to a scale-free network by using β = 7, where k was the whole network connectivity and p(k) was the corresponding frequency distribution (Supporting Table.1). When β = 7, the R2 is 0.98, ensuring that the network was close to the scale-free network. After the soft thresholding power β was determined, the Topological Overlap Matrix (TOM) and dissTOM = 1−TOM were obtained. After the modules were identified, the T-test was used to calculate the significant p-value of candidate genes, and the gene significance (GS) was defined as mediated p-value of each gene (GS = lgP). Then, the module significance (MS) were defined as the average GS of all the genes involved in the module. The cut-off significant standard was setup as p-value lower than 0.05. In general, the module with the highest MS among all the selected modules will be considered as the one associated with disease. In addition, it was also calculated the relevance between the different etiologies (idiopathic dilated, ischemic, idiopathic cardiomyopathy, familial cardiomyopathy, post-partum cardiomyopathy, ischemic cardiomyopathy, viral cardiomyopathy) of modules and cardiomyopathies phenotypes to identify the most relevant module. In the WGCNA, the module membership (MM): MM(i) = cor (xi, ME) is defined to measure the importance of the gene within the module. The greater absolute value of MM(i), the gene i is more important in the module. The Genes Significance (GS) in the module is highly correlated with MM and the most important element to discover the significant module, indicating that Genes in module is significantly associated with cardiomyopathies feature. The hierarchical clustering analysis was used to identify gene modules and color to indicate modules, which is a cluster of densely interconnected genes in terms of co-expression. For genes that are not assigned to any of the modules, WGCNA places them in a grey module as not co-expressed. The module eigengene (ME) of a module is defined as the first principal component of the module and represents the overall expression level of the module. To identify modules that significantly associated with the traits of different etiologies, it was calculated the correlation of MEs (i.e. the first principle component of a module) [49] with clinical pathological features and identified the most significant associations.

Function & Pathway enrichment analysis of gene significance in module

Genes significance related to different pathological phenotype were blast through the web-based GenCLiP 2.0 and conducted correlation analysis, which can analyze human genes associated with biological functions and molecular networks [50]. The correlation analyses were including Gene Cluster with Literature Profiles, Literature Mining Gene Networks and key functional enrichment & pathway hierarchical clustering. The Gene connectivity was used to measure the connection strength of a gene connects to other genes in global functional pathway & network. The genes labelled with purple border represented the association of pathological feature keywords with published literatures.

Identification of hub genes

The module membership (MM) was defined as the correlation of gene expression profile with module eigengene (ME). The GS measure was defined as (the absolute value of) the correlation between gene and external traits. Hub genes were defined as a gene that in a module play important roles in the biological processes than other genes in the whole network, which were comprised as key interconnected

nodes within a functionally network and played important roles in biological functions [44]. In this study, two methods, co-expression network and PPI network analysis, had been employed to identify the real Hub genes among each significant module. Genes with the highest MM and highest GS in modules were processed as candidates for further functional research [51-53]. In this study, the criterial of screening hub genes were setup as GS > 0.2 and MM > 0.8 with a threshold of P-value <0.05, and hub genes were identified in the most significantly module that correlated to certain clinical trait. In parallel, the protein-protein interaction (PPI) network of the module genes were built in the selected modules through STRING database. The significant module contained genes interaction between genes was defined as positive with a combined with the cutoff of > 0.4 and connectivity degree of ≥ 8 through STRING database [54]. In the PPI network, filtered genes were defined as hub genes. The overlapped hub genes in both co-expression network and PPI network were regarded as "real" hub genes pickup for further analyses.

Exploring Cardiovascular Disease Portal

The identified significant genes blast in the Cardiovascular Disease Portal to filter the genes associated with cardiomyopathies. The Cardiovascular Disease Portal provides easy access to multiple genetic data associated with specific cardiomyopathy types and annotations of 854 genes with verified functions related to human cardiomyopathies [55, 56]. The Cardiovascular Disease Portal integrates data for genes, QTLs and strains associated with the disease(s) highlighted and translational research annotation and associated disease information. The filtered genes will be defined as disease signature genes for cardiomyopathies.

Validations of Hub / Signature Genes Expression

The validation of significant genes expression was performed by comparison between the different cardiomyopathies and health groups. The health group was used as the benchmark. The individual gene expression in each group was presented as means ± standard error of the mean (SEM) that represent cases distribution of group. The expression level comparison used the fold change ratio to quantitatively analyze. The significance of differences genes expression was determined between health and cardiomyopathies cases by using the student t-test with the Prism software (GraphPad Software, Inc. San Diego, CA). A p-value < 0.05 was setup as significant difference standard (*, represents p value <0.05; **, represents $0.001 < p$ value <0.0001; ***, represents p value <0.0001). The standard of significance was setup as up-expression (Fold change > 1.0, p < 0.05) or down-expression (Fold change < 1.0, p < 0.05).

Abbreviations

PCA	principal component analysis
GO	gene ontology
TOM	Topological Overlap Matrix
GS	gene significance
MM	module membership
ME	module eigengene
MS	module significance
WGCNA	weighted gene co-expression network analysis

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during this study are available from online supplementary or contact the corresponding author (SL or YH) to request.

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' Contributions

SL and YH designed the overall project study; YH collected data, performed data analysis, and drafted the manuscript; SL, YH, JT and ZW interpreted and summarized the results; SL and YH wrote and revised the

manuscript; all authors have read and approved the final version of manuscript.

References

1. Braunwald E. The war against heart failure: the Lancet lecture. Lancet. 2015;385(9970):812–24.

2. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. Card Fail Rev. 2017;3(1):7–11.

3. Burke MA, Cook SA, Seidman JG, Seidman CE. Clinical and Mechanistic Insights Into the Genetics of Cardiomyopathy. J Am Coll Cardiol. 2016;68(25):2871–86.

4. Maron BJ, Towbin JA, Thiene G, Antzelevitch C, Corrado D, Arnett D, Moss AJ, Seidman CE, Young JB, American Heart A, et al. Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. Circulation. 2006;113(14):1807–16.

5. Hwang JJ, Allen PD, Tseng GC, Lam CW, Fananapazir L, Dzau VJ, Liew CC. Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. Physiol Genomics. 2002;10(1):31–44.

6. Barth AS, Kuner R, Buness A, Ruschhaupt M, Merk S, Zwermann L, Kaab S, Kreuzer E, Steinbeck G, Mansmann U, et al. Identification of a common gene expression signature in dilated cardiomyopathy across independent microarray studies. J Am Coll Cardiol. 2006;48(8):1610–7.

7. Asakura M, Kitakaze M. Global gene expression profiling in the failing myocardium. Circ J. 2009;73(9):1568–76.

8. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4:Article17.

9. Zhang J, Baddoo M, Han C, Strong MJ, Cvitanovic J, Moroz K, Dash S, Flemington EK, Wu T. Gene network analysis reveals a novel 22-gene signature of carbon metabolism in hepatocellular carcinoma. Oncotarget. 2016;7(31):49232–45.

10. Guo Y, Xing Y. Weighted gene co-expression network analysis of pneumocytes under exposure to a carcinogenic dose of chloroprene. Life Sci. 2016;151:339–47.

11. Ford MG, Pearse BM, Higgins MK, Vallis Y, Owen DJ, Gibson A, Hopkins CR, Evans PR, McMahon HT. Simultaneous binding of PtdIns(4,5)P2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes. Science. 2001;291(5506):1051–5.

12. Scotland PB, Heath JL, Conway AE, Porter NB, Armstrong MB, Walker JA, Klebig ML, Lavau CP, Wechsler DS. The PICALM protein plays a key role in iron homeostasis and cell proliferation. PLoS One. 2012;7(8):e44252.

13. Tebar F, Bohlander SK, Sorkin A. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. Mol Biol Cell. 1999;10(8):2687–702.

14. Susuki-Miyata S, Miyata M, Lee BC, Xu H, Kai H, Yan C, Li JD. Cross-talk between PKA-Cbeta and p65 mediates synergistic induction of PDE4B by roflumilast and NTHi. Proc Natl Acad Sci U S A. 2015;112(14):E1800–9.

15. Wang J, Liu S, Heallen T, Martin JF. The Hippo pathway in the heart: pivotal roles in development, disease, and regeneration. Nat Rev Cardiol. 2018;15(11):672–84.

16. Couzens AL, Xiong S, Knight JDR, Mao DY, Guettler S, Picaud S, Kurinov I, Filippakopoulos P, Sicheri F, Gingras AC. MOB1 Mediated Phospho-recognition in the Core Mammalian Hippo Pathway. Mol Cell Proteomics. 2017;16(6):1098–110.

17. Curti A, Lapucci C, Berto S, Prandstraller D, Perolo A, Rizzo N, Farina A. Maternal plasma mRNA species in fetal heart defects: a potential for molecular screening. Prenat Diagn. 2016;36(8):738–43.

18. Florindo C, Perdigao J, Fesquet D, Schiebel E, Pines J, Tavares AA. Human Mob1 proteins are required for cytokinesis by controlling microtubule stability. J Cell Sci. 2012;125(Pt 13):3085–90.

19. Ben Yehuda S, Dix I, Russell CS, Levy S, Beggs JD, Kupiec M. Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. RNA. 1998;4(10):1304–12.

20. Kong XX, Lv YR, Shao LP, Nong XY, Zhang GL, Zhang Y, Fan HX, Liu M, Li X, Tang H. HBx-induced MiR-1269b in NF-kappaB dependent manner upregulates cell division cycle 40 homolog (CDC40) to promote proliferation and migration in hepatoma cells. J Transl Med. 2016;14(1):189.

21. Li BX, Gardner R, Xue C, Qian DZ, Xie F, Thomas G, Kazmierczak SC, Habecker BA, Xiao X. Systemic Inhibition of CREB is Well-tolerated in vivo. Sci Rep. 2016;6:34513.

22. Chien PT, Lin CC, Hsiao LD, Yang CM. c-Src/Pyk2/EGFR/PI3K/Akt/CREB-activated pathway contributes to human cardiomyocyte hypertrophy: Role of COX-2 induction. Mol Cell Endocrinol. 2015;409:59–72.

23. Chang CF, Chou HT, Lin YJ, Lee SJ, Chuang JL, Chuang DT, Huang TH. Structure of the subunit binding domain and dynamics of the di-domain region from the core of human branched chain alpha-ketoacid dehydrogenase complex. J Biol Chem. 2006;281(38):28345–53.

24. Ansari AA, Neckelmann N, Villinger F, Leung P, Danner DJ, Brar SS, Zhao S, Gravanis MB, Mayne A, Gershwin ME, et al. Epitope mapping of the branched chain alpha-ketoacid dehydrogenase dihydrolipoyl transacylase (BCKD-E2) protein that reacts with sera from patients with idiopathic dilated cardiomyopathy. J Immunol. 1994;153(10):4754–65.

25. Chen SL, Dowhan DH, Hosking BM, Muscat GE. The steroid receptor coactivator, GRIP-1, is necessary for MEF-2C-dependent gene expression and skeletal muscle differentiation. Genes Dev. 2000;14(10):1209–28.

26. Yu B, Barbalic M, Brautbar A, Nambi V, Hoogeveen RC, Tang W, Mosley TH, Rotter JI, deFilippi CR, O'Donnell CJ, et al. Association of genome-wide variation with highly sensitive cardiac troponin-T

levels in European Americans and Blacks: a meta-analysis from atherosclerosis risk in communities and cardiovascular health studies. Circ Cardiovasc Genet. 2013;6(1):82–8.

27. Gu Q, Mendsaikhan U, Khuchua Z, Jones BC, Lu L, Towbin JA, Xu B, Purevjav E. Dissection of Z-disc myopalladin gene network involved in the development of restrictive cardiomyopathy using system genetics approach. World J Cardiol. 2017;9(4):320–31.

28. Brumbaugh J, Di Stefano B, Wang X, Borkent M, Forouzmand E, Clowers KJ, Ji F, Schwarz BA, Kalocsay M, Elledge SJ, et al. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. Cell. 2018;172(1–2):106–20. e121.

29. Zhang L, Zhang W. Knockdown of NUDT21 inhibits proliferation and promotes apoptosis of human K562 leukemia cells through ERK pathway. Cancer Manag Res. 2018;10:4311–23.

30. Maffei A, Lembo G, Carnevale D. PI3Kinases in Diabetes Mellitus and Its Related Complications. *Int J Mol Sci* 2018, 19(12).

31. Tan B, Liu M, Yang Y, Liu L, Meng F. Low expression of PIK3C2A gene: A potential biomarker to predict the risk of acute myocardial infarction. Medicine. 2019;98(14):e15061.

32. Nakielny S, Dreyfuss G. The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. J Cell Biol. 1996;134(6):1365–73.

33. Williamson DJ, Banik-Maiti S, DeGregori J, Ruley HE. hnRNP C is required for postimplantation mouse development but Is dispensable for cell viability. Mol Cell Biol. 2000;20(11):4094–105.

34. Sun DQ, Wang Y, Liu DG. Overexpression of hnRNPC2 induces multinucleation by repression of Aurora B in hepatocellular carcinoma cells. Oncol Lett. 2013;5(4):1243–9.

35. Markson G, Kiel C, Hyde R, Brown S, Charalabous P, Bremm A, Semple J, Woodsmith J, Duley S, Salehi-Ashtiani K, et al. Analysis of the human E2 ubiquitin conjugating enzyme protein interaction network. Genome Res. 2009;19(10):1905–11.

36. Wang S, Yan R, Wang B, Meng P, Tan W, Guo X. The Functional Analysis of Selenium-Related Genes and Magnesium-Related Genes in the Gene Expression Profile Microarray in the Peripheral Blood Mononuclear Cells of Keshan Disease. *Biol Trace Elem Res* 2019.

37. Haider N, Arbustini E, Gupta S, Liu H, Narula N, Hajjar R, Moorjani N, Westaby S, Semigran MJ, Dec GW, et al. Concurrent upregulation of endogenous proapoptotic and antiapoptotic factors in failing human hearts. Nat Clin Pract Cardiovasc Med. 2009;6(3):250–61.

38. Mackenzie FE, Romero R, Williams D, Gillingwater T, Hilton H, Dick J, Riddoch-Contreras J, Wong F, Ireson L, Powles-Glover N, et al. Upregulation of PKD1L2 provokes a complex neuromuscular disease in the mouse. Hum Mol Genet. 2009;18(19):3553–66.

39. Mandl A, Huong Pham L, Toth K, Zambetti G, Erhardt P. Puma deletion delays cardiac dysfunction in murine heart failure models through attenuation of apoptosis. Circulation. 2011;124(1):31–9.

40. Zhu Y, Yang T, Duan J, Mu N, Zhang T. MALAT1/miR-15b-5p/MAPK1 mediates endothelial progenitor cells autophagy and affects coronary atherosclerotic heart disease via mTOR signaling pathway. Aging. 2019;11(4):1089–109.

41. 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(3):416–24.

42. Zhang D, Li Y, Heims-Waldron D, Bezzerides V, Guatimosim S, Guo Y, Gu F, Zhou P, Lin Z, Ma Q, et al. Mitochondrial Cardiomyopathy Caused by Elevated Reactive Oxygen Species and Impaired Cardiomyocyte Proliferation. Circ Res. 2018;122(1):74–87.

43. Li Q, Cui HH, Yang YJ, Li XD, Chen GH, Tian XQ, Jin C, Dong QT, Huang PS, Xu J. Quantitative Proteomics Analysis of Ischemia/Reperfusion Injury-Modulated Proteins in Cardiac Microvascular Endothelial Cells and the Protective Role of Tongxinluo. Cell Physiol Biochem. 2017;41(4):1503–18.

44. Chen P, Wang F, Feng J, Zhou R, Chang Y, Liu J, Zhao Q. Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. Oncotarget. 2017;8(30):48948–58.

45. Xu J, Kimball TR, Lorenz JN, Brown DA, Bauskin AR, Klevitsky R, Hewett TE, Breit SN, Molkentin JD. GDF15/MIC-1 functions as a protective and antihypertrophic factor released from the myocardium in association with SMAD protein activation. Circ Res. 2006;98(3):342–50.

46. Sucharov CC, Mariner P, Long C, Bristow M, Leinwand L. Yin Yang 1 is increased in human heart failure and represses the activity of the human alpha-myosin heavy chain promoter. J Biol Chem. 2003;278(33):31233–9.

47. Zhao Z, Guan JZ, Wu M, Lai GH, Zhu ZL. Downregulation of microRNA-23b protects against ischemia-reperfusion injury via p53 signaling pathway by upregulating MDM4 in rats. J Cell Biochem. 2019;120(3):4599–612.

48. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

49. Li A, Horvath S. Network neighborhood analysis with the multi-node topological overlap measure. Bioinformatics. 2007;23(2):222–31.

50. Wang JH, Zhao LF, Lin P, Su XR, Chen SJ, Huang LQ, Wang HF, Zhang H, Hu ZF, Yao KT, et al. GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular networks based on free terms. Bioinformatics. 2014;30(17):2534–6.

51. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet. 2006;2(8):e130.

52. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc Natl Acad Sci U S A. 2006;103(46):17402–7.

53. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci U S A. 2006;103(47):17973–8.

54. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447–52.

55. Hayman GT, Laulederkind SJ, Smith JR, Wang SJ, Petri V, Nigam R, Tutaj M, De Pons J, Dwinell MR, Shimoyama M: **The Disease Portals, disease-gene annotation and the RGD disease ontology at the Rat Genome Database**. *Database (Oxford)* 2016, **2016**.

56. Shimoyama M, De Pons J, Hayman GT, Laulederkind SJ, Liu W, Nigam R, Petri V, Smith JR, Tutaj M, Wang SJ, et al. The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. Nucleic Acids Res. 2015;43(Database issue):D743–50.

Figures
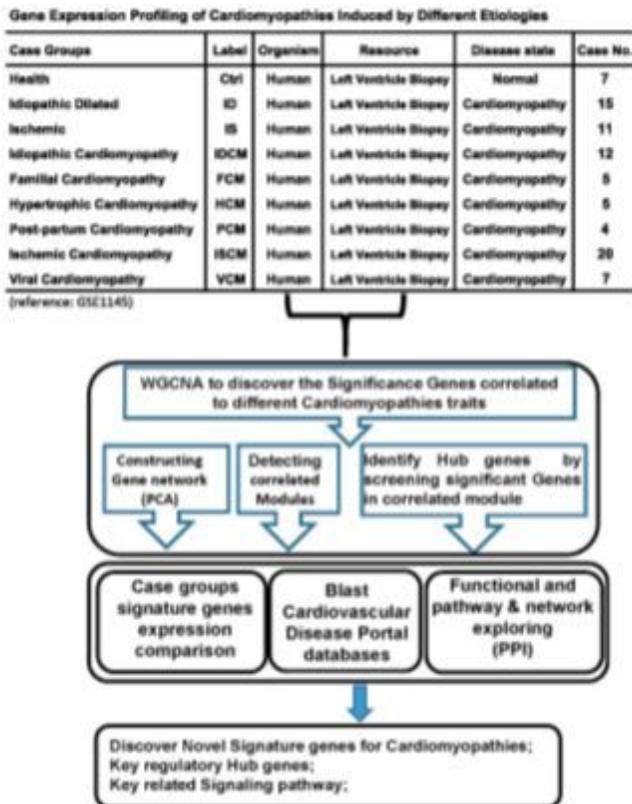
Figure 1

Work flow of the different etiologies caused cardiomyopathies biopsy samples for WGCNA.
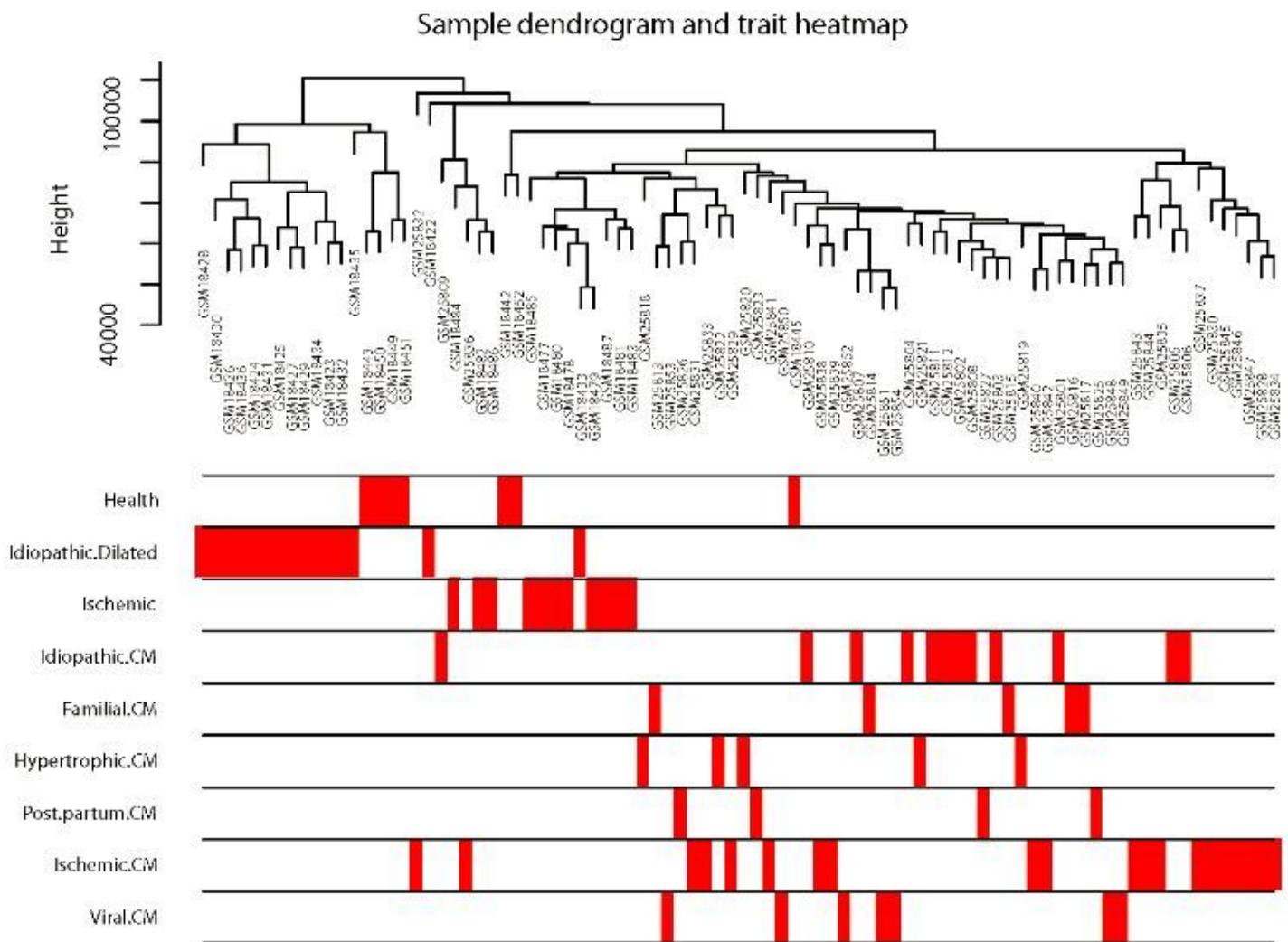
Figure 2

Sample dendrogram and Clinic Feature traits heatmap. (A). Clustering dendrogram of samples based on their Euclidean distance. The different etiologies caused cardiomyopathies (CM) were idiopathic dilated (ID, n=15), ischemic (IS, n=11), idiopathic cardiomyopathy (IdCM, n=12), familial cardiomyopathy (FCM, n=5), post-partum cardiomyopathy (PCM, n=4), ischemic cardiomyopathy (IsCM, n=20) and viral cardiomyopathy (VCM, n=7). The white color means a low value, red means a high value.

Figure 3

Sample dendrogram and Clinic Feature traits heatmap. (B) The correlation heatmap of Module-Cardiomyopathy etiologies traits. Each row corresponds to a module; each column corresponds to a different etiologies feature. Identification of modules significant associated with clinical features and Clustering dendrogram of genes, with dissimilarity based on topological overlap, is merged with assigned module colors. Grey bars represent Genes that do not belong to any other modules and are not co-expressed. Each cell contains the test statistic value and its corresponding p value from the linear mixed-effects model. CM is the abbreviation of Cardiomyopathy.
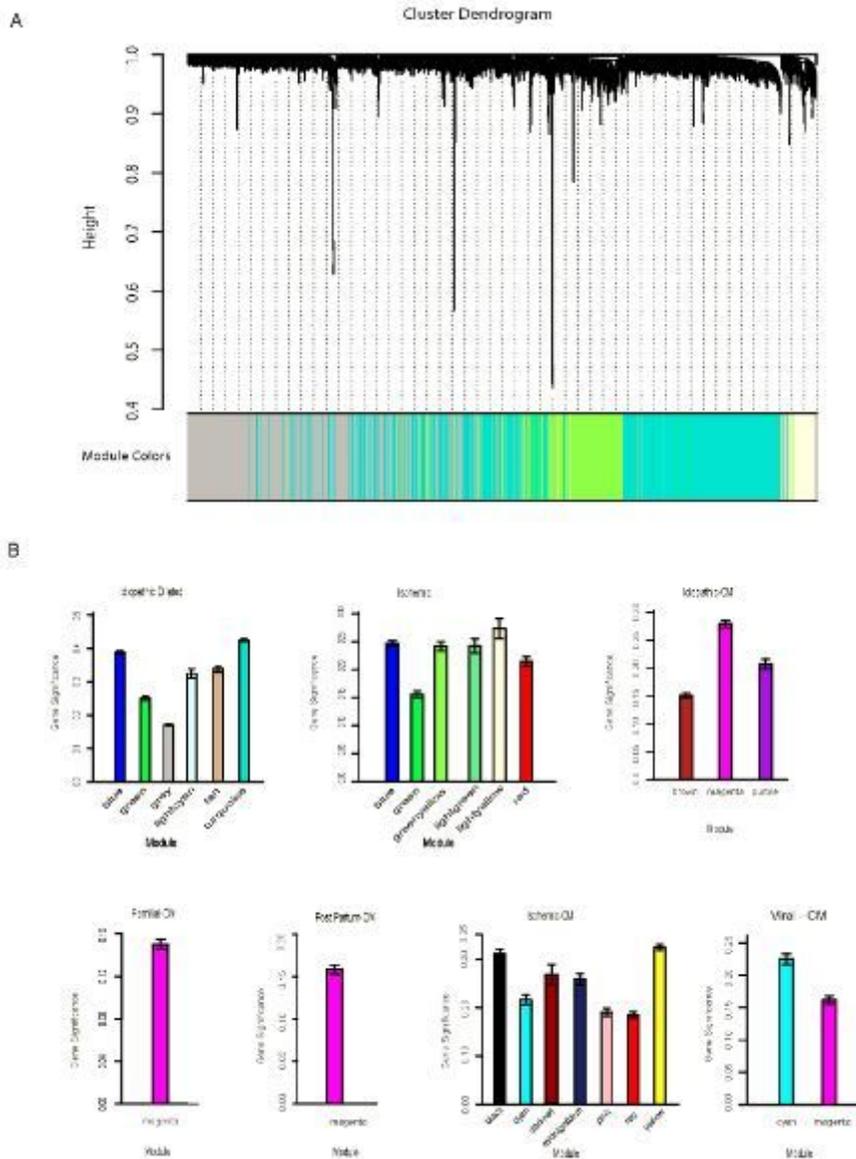
Figure 4

Gene dendrogram and module membership for Clinic Feature traits. (A) Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure. (B) The total Module memberships vs. gene significance cluster for etiologies trait. Distribution of average gene significance and errors in the significant modules associated with the different etiologies.
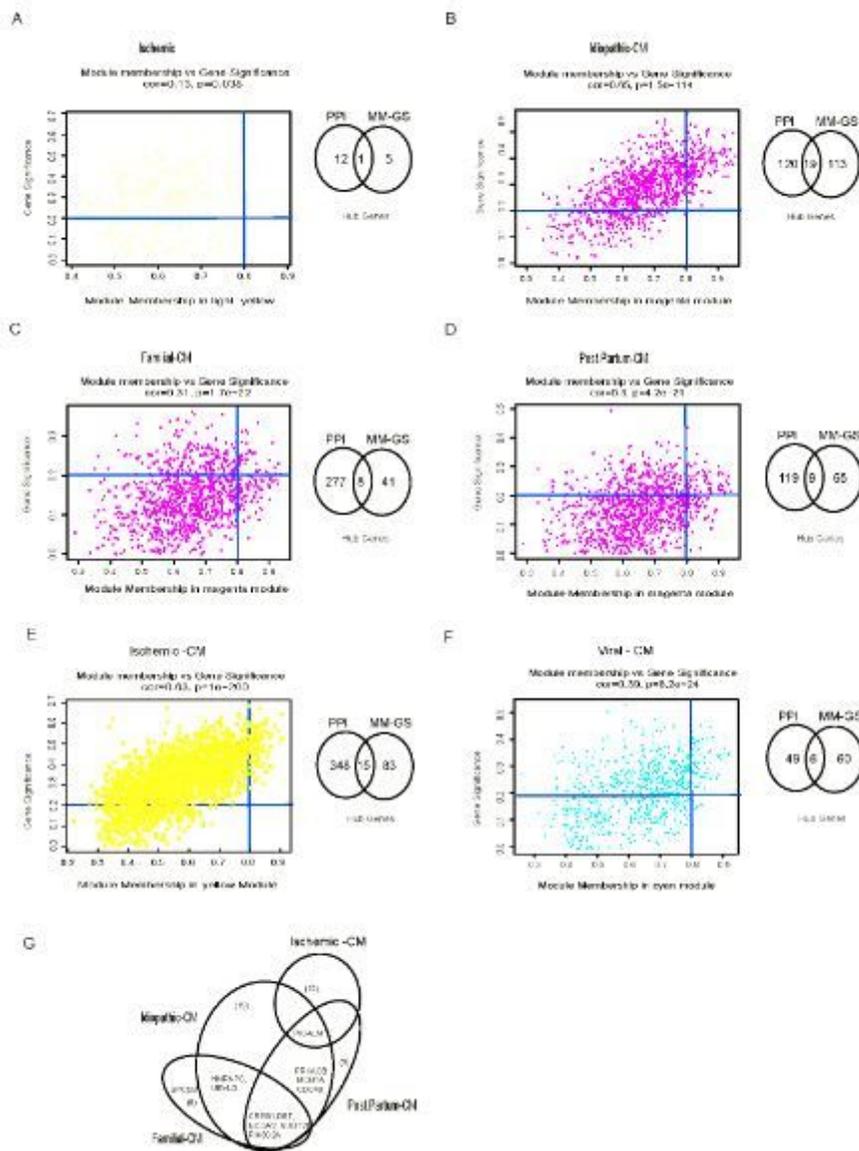
Figure 5

The scatterplots of Gene Significance (GS) for etiologies vs. Module Membership (MM) in the all cases. There is a highly significant correlation between GS and MM in this module that the most important (central) elements of module tend to be highly correlated with cardiomyopathies. Module membership vs gene significance is correlating to different etiologies for cardiomyopathies. Hub genes detected by GS-MM and protein-protein network (PPI). (A) – (F) Scatter plot of module eigengenes in module for all cases, and the real hub genes indicated by Venn diagram for co-expression and PPI network. (G) The Venn diagram showed the overlapped real hub for different cardiomyopathies.
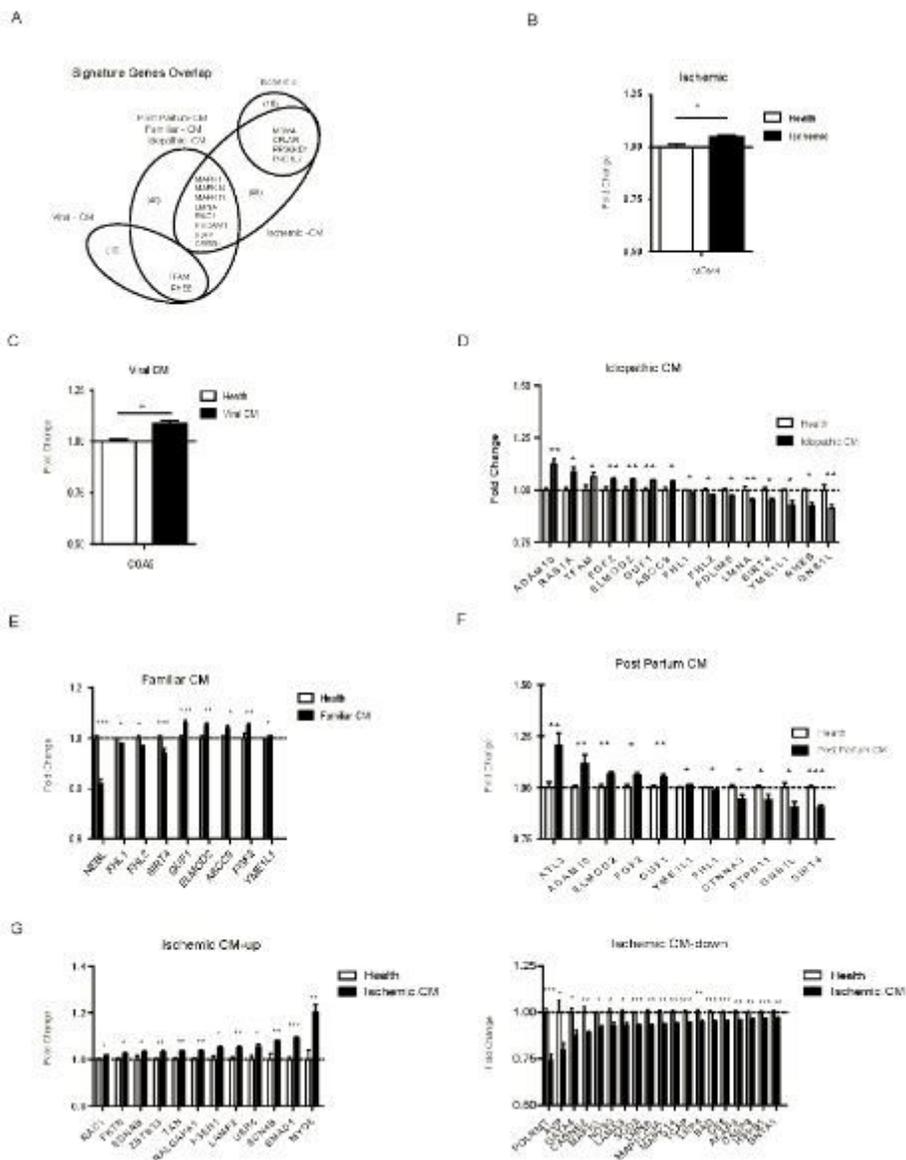
Figure 6

Validation of Disease Signature genes expression. (A) The Venn diagram showed the overlapped disease signature genes for different cardiomyopathies. (B) ~ (G). Expression levels of signature genes between subtype of cardiomyopathies and healthy control. (***, represent p value <0.0001; **, represent 0.001<p value <0.0001; *, represents p value <0.05).

Supplementary Files

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

- SupTable4SGeneSignificanceEnrichmentSummary.xlsx
- SupplementarymaterialFigureTable.pdf
- DrLiuSupplementaryfiles.pdf