

Network-based computational approach to identify genetic links between cardiomyopathy and its risk factors

ISSN 1751-8849
 Received on 11th July 2019
 Revised 23rd September 2019
 Accepted on 21st October 2019
 E-First on 26th November 2019
 doi: 10.1049/iet-syb.2019.0074
 www.ietdl.org

Md. Nasim Haidar¹, M. Babul Islam¹, Utpala Nanda Chowdhury², Md. Rezanur Rahman³, Fazlul Huq⁴, Julian M.W. Quinn⁵, Mohammad Ali Moni^{4,5} ✉

¹Department of Electrical and Electronic Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh

²Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh

³Department of Biochemistry and Biotechnology, School of Biomedical Science, Khwaja Yunus Ali University, Sirajgonj 6751, Bangladesh

⁴School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, NSW 2006, Australia

⁵Bone Biology Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia

✉ E-mail: mohammad.moni@sydney.edu.au

Abstract: Cardiomyopathy (CMP) is a group of myocardial diseases that progressively impair cardiac function. The mechanisms underlying CMP development are poorly understood, but lifestyle factors are clearly implicated as risk factors. This study aimed to identify molecular biomarkers involved in inflammatory CMP development and progression using a systems biology approach. The authors analysed microarray gene expression datasets from CMP and tissues affected by risk factors including smoking, ageing factors, high body fat, clinical depression status, insulin resistance, high dietary red meat intake, chronic alcohol consumption, obesity, high-calorie diet and high-fat diet. The authors identified differentially expressed genes (DEGs) from each dataset and compared those from CMP and risk factor datasets to identify common DEGs. Gene set enrichment analyses identified metabolic and signalling pathways, including MAPK, RAS signalling and cardiomyopathy pathways. Protein–protein interaction (PPI) network analysis identified protein subnetworks and ten hub proteins (CDK2, ATM, CDT1, NCOR2, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E and HIST1H4L). Five transcription factors (FOXC1, GATA2, FOXL1, YY1, CREB1) and five miRNAs were also identified in CMP. Thus the authors' approach reveals candidate biomarkers that may enhance understanding of mechanisms underlying CMP and their link to risk factors. Such biomarkers may also be useful to develop new therapeutics for CMP.

1 Introduction

Cardiomyopathy (CMP) is a group of diseases affecting the structure and functioning of the heart, and includes conditions where the heart is affected by ventricular hypertrophy, dilation or fibrotic dysplasia that cause mechanical and electrical dysfunction. CMP may be either cardiac-specific or a part of generalised systemic disorders, but many of these conditions result in cardiovascular damage or progressive heart failure [1]. CMP is the third most prevalent cause of heart failure in the USA [1]. In 2015, about 2.6 million people worldwide were affected by cardiomyopathy and myocarditis [2]. Currently, the most commonly occurring form of CMP is dilated CMP which affects five in 100,000 adults and 0.57 in 100,000 children [3].

The etiology of the cardiomyopathy involves genetic, infectious, metabolic and environmental factors [1]. Lifestyle risk factors include severe obesity, alcohol consumption (AC), long-term high blood pressure, coronary heart disease, and sarcoidosis, but the molecular mechanisms behind the development of CMP and how these risk factors contribute to the progression of the CMP is not well understood. However, we can use our knowledge of CMP risk factors to identify key factors in CMP development by determining the altered gene expression patterns the risk induces that are also seen CMP-affected heart tissues. Using an integrative gene-network-based approach we can then identify candidate causative pathways that can be further examined [4, 5].

Integrative network-based gene or multi-omics analyses are an increasingly common approach used to identify disease-associated biomarkers and therapeutic targets [6]. Such an approach is now commonly used for elucidating molecular mechanisms in different disease such as Alzheimer's disease [7–12], Parkinson's disease [13–15], multiple sclerosis [16], respiratory system diseases [17], colorectal cancer [18] and Thyroid cancers [19–21]. Therefore, in this study, a system biology-based approach was used to identify

molecular biomarker transcripts (i.e. mRNAs), and proteins (hub proteins) and pathways in CMP using CMP-associated risk factors to clarify the genes that may be causative factors for the progression of CMP (Fig. 1). For this purpose, we first identified DEGs, genes whose expression is altered in CMP affected tissues and in risk-factor exposed tissues; these DEGs that were common between CMP and particular CMP-associated risk factors were then identified. These common DEGs, were then studied for their involvement in human biomolecular networks such as protein–protein interaction (PPI) networks to identify central signalling molecules (hub proteins) and molecular pathways. This resulted in the identification of candidate genes that could mediate influences of the CMP risk factors, and these were then cross-validated using gold benchmarking datasets OMIM and dbGaP gene-disease association databases to identify those candidates with known pathological involvement.

2 Materials and methods

2.1 High-throughput microarray gene expression datasets

We analysed gene expression microarray datasets to identify the molecular association of different factors with CMP at the molecular level. All the datasets used in this study were collected from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus [22], and employed Affymetrix Human DNA arrays unless otherwise stated. The utilised gene expression datasets with accession numbers GSE4172, GSE1144, GSE4806, GSE12654, GSE20950, GSE25220, GSE44456, GSE48964, GSE56960 and GSE68231 were analysed in this study. The CMP dataset (GSE4172) was obtained by gene expression profiling of human inflammatory CMP [23]. The ageing (AG) dataset (GSE1144) was obtained by analysing gene expression in skeletal muscle tissue characterised by loss of metabolic and contractile

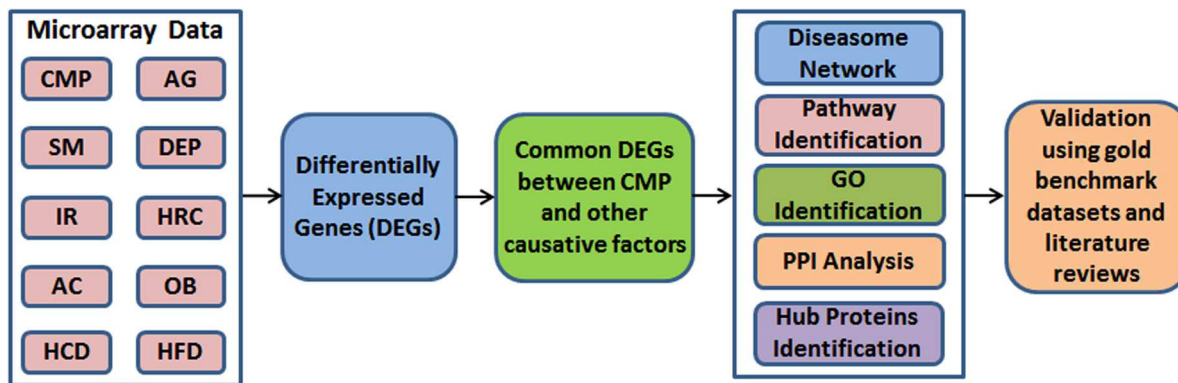


Fig. 1 Overview of the quantitative network-based approach employed in this study. The microarray gene expression datasets of CMP (GSE4172), ageing (GSE1144), smoking (GSE4806), depression (GSE12654), high RM consumption (GSE25220), IR (GSE20950), chronically high AC (GSE44456), morbid obesity (GSE48964), HCD (GSE56960) and HFD datasets (GSE68231) have been obtained from NCBI-GEO

competence [24]. The smoking (SM) dataset (GSE4806) was obtained from gene expression profiles of T-lymphocytes from smokers and non-smokers [25]. The depression (DEP) dataset (GSE12654) was attained by gene expression from the human prefrontal cortex (BA10) [26]. The dataset (GSE20950) for insulin resistance (IR) was obtained by gene expression data from human adipose tissue using an IR patient cohort [27]. The red meat (RM) dietary intervention dataset (GSE25220) was generated using an Agilent-014850 whole human genome microarray data from human colon biopsies before and after participating in a high RM dietary intervention [28]. The AC dataset (GSE44456) was obtained by examining gene expression in post-mortem hippocampus tissues from 20 alcoholics and 19 controls [29]. The obesity (OB) dataset (GSE48964) was obtained by expression data from adipose stem cells (ASCs) from morbidly obese and non-obese individuals [30]. The high-calorie diet (HCD) dataset (GSE56960) was obtained by expression profiling of array of blood cell transcriptome of two different population groups after the ingestion of different caloric doses [31]. The high-fat diet (HFD) dataset (GSE68231) is Affymetrix Human Genome data obtained from human skeleton muscle of five subjects in each group selected before and after three days of an HFD [32].

2.2 Identification of differentially expressed genes

We performed a differential gene expression analysis of the CMP with nine risk factors from transcriptomics datasets. Firstly, we transformed each gene expression data for each disease using the Z-score (or zero mean) normalisation method for both disease and control states. This might resolve the problems regarding mRNA data comparisons using different platforms and experimental setups [33]. Each sample of the gene expression matrix was normalised using mean and standard deviation. The expression value of the gene i in sample j represented by g_{ij} was transformed into Z_{ij} by computing

$$Z_{ij} = \frac{g_{ij} - \text{mean}(g_i)}{\text{SD}(g_i)} \quad (1)$$

where SD is the standard deviation. Comparing values of gene expression for various samples and diseases are made possible by this transformation.

The gene expression datasets were normalised by \log_2 transformation and unpaired student t -test was used. Finally, genes were filtered by setting threshold values with adjusted p -value < 0.05 and absolute log fold change (log FC) > 1.0 to designate statistically significant DEGs.

2.3 Gene set enrichment analysis to identify gene ontology and pathways

To clarify the biological significance of the identified DEGs, gene-set enrichment analysis and pathways analysis were performed to identify the significant gene ontology terms and KEGG pathways

enriched by DEGs via EnrichR [34, 35]. For statistical significance, the adjusted p -value < 0.05 was considered for the significance assessment of enrichment results.

2.4 Identification of transcriptional and/or post-transcription regulators of the DEGs

To identify regulatory transcription factors (TFs) that regulate the DEGs at the transcriptional level, TF-target gene interactions were obtained from the JASPAR database to identify TFs based on topological parameters [36]. The regulatory miRNAs which regulate DEGs at the post-transcriptional level were identified from miRNAs-target gene interactions were obtained from TarBase and miRTarBase based on topological parameters [37–39].

2.5 PPI analysis to identify hub proteins

We reconstructed a PPI network around the proteins encoded by the DEGs using protein interactome database STRING [40]. The PPI network was analysed by Cytoscape (v3.5.1) [41, 42]. An undirected graph representation was used for the PPI network, where the nodes indicate proteins and the edges symbolised the interactions between the proteins. We performed a topological analysis using Cyto-Hubba plugin [43, 44] in Cytoscape to identify highly connected proteins (i.e. hub proteins) in the network and the degree metrics were employed [45, 46].

2.6 Protein–drug interactions analysis

The protein–drug interactions were analysed using the DrugBank database (Version 5.0) to identify potential small molecules that can affect pathways we identified as important in CMP and which may point to therapeutic approaches for CMP [47].

3 Results

3.1 Identification of differentially expressed genes from microarray gene expression datasets

The gene expression datasets of CMP were analysed and a total of 1764 DEGs were identified in CMP patients compared to control samples where 919 genes were up-regulated and 845 genes were down-regulated.

For the investigation of the relationship of the CMP with nine risk factors, we performed several steps of statistical analysis for mRNA microarray data of each risk factor. Thus, we selected the most significant over and under-regulated genes for each risk factor. Our analysis identified a large number of dysregulated genes. The 1122, 400, 72, 356, 482, 25, 157, 255, 739 DEGs were identified in AG, SM, DEP, IR, RM, AC, OB, HCD, HFD datasets. Then, a cross-comparative analysis revealed the common over and under-expressed genes between CMP and above-mentioned risk factors. The CMP shared 48, 32, 3, 28, 29, 2, 6, 7 and 81 significant DEGs with AG, SM, DEP, IR, RM, AC, OB, HCD and HFD, respectively.

The disease association networks centred on the CMP were built to identify statistically significant associations among these risk factors (Figs. 2 and 3 and Table 1). Notably, FAM208B was commonly dysregulated for CMP, AG, HFD and IR. CMP shared 18 DEGs (ESCO2, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4H, HIST1H4I, HIST1H4J, HIST1H4K, HIST1H4L, HIST2H4A, HIST2H4B, HIST4H4, ITGB1, PIEZO2, SCN8A) with HFD. Three DEGs FGFR2, GHRHR and SPTBN1 were common among CMP, AG. CMP shared GABRB1 and HMGA2 dysregulated genes with HFD and RM. Moreover, CMP and HFD shared CCNF, TNPO2 and PPP2R1B dysregulated genes with AG, DEP and SM, respectively.

3.2 Molecular pathway and functional analysis

To clarify the biological roles of the identified common DEGs between CMP and other risk factors, we performed gene ontology analysis to identify the biological process, cellular component and molecular functions enriched by the DEGs (Table 2). There was a total of 741 significant gene ontology groups including leukocyte adhesion to vascular endothelial cell (GO:0061756), cellular response to reactive nitrogen species (GO:1902170), mesoderm formation (GO:0001707), cAMP-mediated signalling (GO:0019933), morphogenesis of an epithelium (GO:0002009), negative regulation of response to biotic stimulus (GO:0002832), leukocyte tethering or rolling (GO:0050901), establishment of epithelial cell apical/basal polarity (GO:0045198), regulation of osteoblast proliferation (GO:0033688), mesodermal cell differentiation (GO:0048333) were identified.

The significantly altered molecular pathways were identified in CMP and other risk factors. A total of 61 pathways were found to be over-represented among several groups out of which some significant pathways are shown in Table 2. The amino acid metabolism pathway such as alanine metabolism pathways, different signalling pathways such as MAPK and RAS signalling pathways, ECM pathways, and alcoholism came into prominence as signalling pathways.

3.3 Identification of regulatory biomolecules

We studied the regulators of the common DEGs utilising DEGs–TFs and DEGs–miRNAs interaction analyses, presented in Table 3. We identified DEGs–TFs interactions (Fig. 4) and DEGs–miRNAs interactions (Fig. 5) and detected central regulatory biomolecules (TFs and miRNAs) using topological parameters. As shown in Table 3, five TFs (FOXC1, GATA2, FOXL1, YY1, and CREB1) and five miRNAs (mir-335-5p, mir-26b-5p, mir-34a-5p, mir-92a-3p, and mir-17-5p) were detected from the DEGs–TFs and DEGs–miRNAs interaction networks, respectively. These biomolecules regulate genes at transcriptional and post-transcriptional levels.

The FOXC1 is a TF that plays a critical role in early cardiomyogenesis [48]. It is also required for the morphogenesis process of cardiac outflow tract [58]. The TF GATA2 expression is high in the thoracic aorta and GATA2 variants are associated with early-onset familial coronary artery disease [49]. FOXL1 is a TF whose elevated expression is associated with good outcomes in human pancreatic ductal adenocarcinoma [50] but does not have a known association with cardiac diseases. The activity of YY1 TF is increased in human heart failure [51]. CREB over-expression is associated with cardiac failure suggesting it plays a significant role in cardiac pathologies [52].

microRNAs (miRNAs) are short single-stranded RNA molecules (~22 nucleotides long) that regulate the expression of genes at post-transcriptional stage. miRNAs are being considered as potential sources of biomarkers for complex disease including neurodegenerative disease and cancers. Therefore, we have identified these miRNAs controlling the DEGs to provide insights into the regulatory biomolecules. Among the miRNAs, mir-335-5p was identified as upregulated in experimental heart failure by experimental animals [53]. Sun *et al.* [59] also predicted mir-335-5p is implicated in hypertrophic CMP pathway by microarray analysis. Jia *et al.* [54] showed mir-26b-5p was associated with suppression of proliferation and enhance the

apoptosis in multiple myeloma cells. It has been proposed the mir-34a-5p could prevent autophagic cell deaths in ischemic hearts and in this way can improve the myocardial injury [55, 60]. The inhibition of mir-92a-3p leads to increase blood vessel growth and recovery of damaged tissues in myocardial infarction mice models, which suggest it may be an important therapeutic target in ischaemic heart disease [56]. The mir-17-5p has been suggested as important prognostic biomarkers in cancer, including hepatocellular carcinoma [57].

3.4 PPI network analysis

The PPI network was constructed using all the distinct 236 differentially expressed genes that were common between the CMP and the risk factors (Figs. 6 and 7). The topological analysis using degree matrices was used to identify highly connected proteins clusters. Each node in the network represents a protein and an edge indicates the interaction between two proteins. We detected ten hub proteins (CDK2, ATM, CDF1, NCOR2, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E and HIST1H4L) in PPI analysis. These hub proteins may be potential drug targets.

3.5 Identification of candidate drugs

A protein–drug interaction network was analysed and we found GABRB1, GRIA3, SLC6A2, GAD2, GAD2, CACNB1, NTRK2, and GRM5 proteins had interaction with 5 drugs/compounds (Ethanol, Amoxapine, L-Glutamic Acid, Amitriptyline, Acamprosate) as shown in Fig. 8.

4 Discussion

In this study, the molecular mechanisms that may link CMP and associated risk factors were investigated. We performed an analysis of gene expression data from CMP tissue analysis and from the risk factors in order to identify the common DEGs shared by CMP and the risk factors. This identified CMP affected tissues shared with 81 genes with tissues and cells affected by HFD exposure; similarly there were shared DEGs seen for AG (48 genes) and SM (32 genes), the conditions that shared most DEGs with CMP. To clarify the biological relevance of the identified DEGs, GO and molecular pathways analysis was performed which revealed pathways with significantly altered activity. Among such pathways, MAPK signalling cascades have been reported to be prominent in the pathogenesis of cardiac and vascular disease [61–63]. Another pathway, RAS signalling, plays a critical role in cardiac hypertrophy, which suggests complexity in developing meaningful therapy for individuals with these RASopathies [64]. Clinical and genetic studies have also revealed close relationships between cell adhesion proteins and the occurrence of various CMPs [65], thus indicating the important role of focal adhesion pathways in CMP. Related to this, extracellular matrix alterations may be a significant factor in the pathogenesis of dilative CMP [66]. Moreover, molecular pathways hypertrophic CMP, arrhythmogenic right ventricular CMP, dilated CMP pathways were notably and consistently seen to be enriched in CMP.

Analysis of PPIs can provide some detailed insights into the central mechanism behind the diseases [9, 11, 12]. Therefore, we reconstructed the PPI networks around the protein encoded by the DEGs. Based on our topological analysis, we detected ten hub proteins (CDK2, ATM, CDT1, NCOR2, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E and HIST1H4L) involved in the CMP. A brief description of hub proteins, their gene ontology and features are presented in Table 4. Among the hubs, CDK2 involved in the regulation of myocardial ischaemia and reperfusion injury [67, 74]. The hub protein apical transverse motion (ATM) is associated with electromechanical dyssynchrony in adult dilated CMP [68]. The ATM protein involved in CMP associated with obesity and IR [75]. The hub protein CDT1 is associated with genotoxic stress, which results in aberrant cell proliferation leading to cancer formation [69], but its association with the CMP is not known. Another hub protein, NCOR2 has been reported to be associated with non-alcoholic fatty liver disease [70], which is one of the prominent risk factors for cardiovascular disease. Yin *et al.*

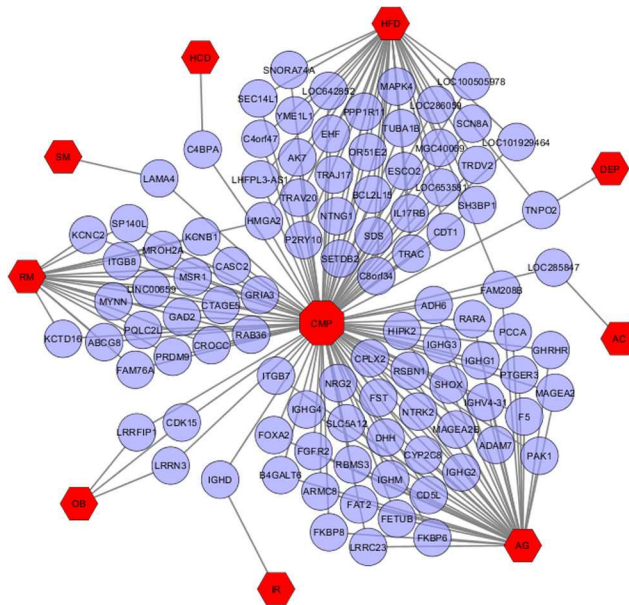


Fig. 2 Network for up-regulated genes of CMP with ageing (AG), smoking (SM), AC, HCD, HFD, IR, obesity (OB), depression (DEP) and high RM intake. The red coloured octagon-shaped node at the centre represents the target CMP and the hexagon-shaped nine nodes represent the risk factors and the other nodes represent the genes that are in common between CMP with the examined risk factors

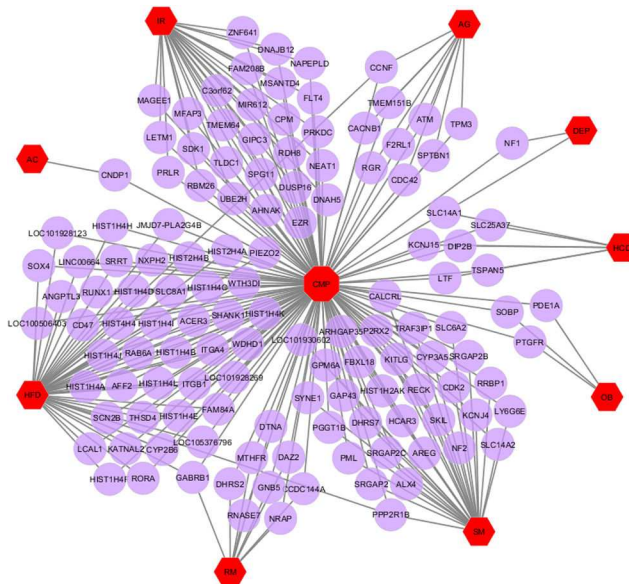


Fig. 3 Network for down-regulated genes of CMP with ageing (AG), smoking (SM), AC, HCD, HFD, IR, obesity (OB), depression (DEP) and high RM intake. The red coloured octagon-shaped node at the centre represents the target CMP and the hexagon-shaped nine nodes represent the risk factors and the other nodes represent the genes that are common to CMP and the risk factors

[72] have reported the dysregulation of HIST1H4B in rat cardiomyocytes. The other hub proteins, HIST1H4A, HIST1H4D, HIST1H4E and HIST1H4L were not reported to CMP yet. These identified hubs proteins might be considered as candidate biomarkers or, if their biological role is confirmed, as potential drug targets.

Based on the network-based approach, our analyses revealed novel relationships between CMP and other susceptibility/causative factors. This study identified potential biomarkers, which may be candidates for the development of prognostic strategies and treatments. Since the common pathways may indicate ways by the risk factors influence CMP, such pathways and their hub genes identified in this study may have important pathogenic roles in CMP. To examine this and so to validate the results of this systems biology approach, we also analysed the DEGs associated with CMP and each of the risk factors with OMIM databases and dbGAP databases using the valid gold benchmark the disease-gene associations (Table 5). The DEGs of nine risk factors were identified as showing suggestive links that may promote CMP

development and progression. This analysis furnishes new hypotheses that may point the way to establishing mechanistic links between the CMP and the various risk factors that we examined.

5 Conclusion

In this study, the genetic association of CMP with various disease was identified from comprehensive transcriptomics analyses incorporated with human biomolecular networks to reveal candidate biomarkers at RNA level (transcripts and miRNAs) and protein levels (hub proteins); identified as potential key signalling and regulatory biomolecules in CMP; we also identified possible molecular pathways with CMP involvement. Protein-drug interaction studies revealed eight gene products that had detectable in silico interaction with four compounds including, Amoxapine, L-Glutamic Acid, Amitriptyline and Acamprosate, which are all compounds already available for therapeutic use apart from glutamate, which is a nutrient and neurotransmitter. Thus, new gene-based recommendations for disease diagnosis and possible

Table 1 Significant Gene Ontology terms related to CMP and risk factors such as smoking, IR, AC, high caloric diet, high RM intake, depression, HFD, obesity and ageing

Category	GO ID	Term/pathway	Genes	Risk factors	P-value
gene ontology biological process	GO:2000146	negative regulation of cell motility	NF1, SRGAP2C, SRGAP2, NF2, RECK, SRGAP2B	DEP, SM	4.62×10 ⁻⁷
	GO:0070828	heterochromatin organisation	MTHFR, HMGA2	HFD, RM	9.07×10 ⁻⁵
	GO:2000257	regulation of protein activation cascade	IGHG3, IGHG4, IGHG1, IGHG2, C4BPA	AG, HCD	1.35×10 ⁻⁴
	GO:0030334	regulation of cell migration	NF1, SRGAP2C, SRGAP2, NF2, RECK, SRGAP2B	DEP, SM	1.37×10 ⁻⁴
	GO:0030449	regulation of complement activation	IGHG3, IGHG4, IGHG1, IGHG2, C4BPA	AG, HCD	1.40×10 ⁻⁴
	GO:0002920	regulation of humoral immune response	IGHG3, IGHG4, IGHG1, IGHG2, C4BPA	AG, HCD	1.61×10 ⁻⁴
	GO:0002697	regulation of immune effector process	IGHG3, IGHG4, IGHG1, IGHG2, C4BPA	AG, HCD	1.66×10 ⁻⁴
	GO:0002673	regulation of acute inflammatory response	IGHG3, IGHG4, IGHG1, IGHG2, C4BPA	AG, HCD	2.08×10 ⁻⁴
	GO:0070613	regulation of protein processing	IGHG3, IGHG4, IGHG1, IGHG2, C4BPA	AG, HCD	2.58×10 ⁻⁴
GO:1902531	regulation of intracellular signal transduction	CDC42, PAK1, F2RL1, ATM, HIPK2, CDK2, NF2, SRGAP2, PML, ARHGAP35	AG, SM	5.20×10 ⁻⁴	
gene ontology	GO:0005887	integral components of plasma membrane	SLC14A1, KCNJ15, TSPAN5, PTGFR	HCD, OB	1.10×10 ⁻²
	GO:0030424	axon	NTRK2, PAK1, NF1, KCNB1, KCNC2	AG, DEP, RM	1.79×10 ⁻²
cellular component	GO:0071437	invadopodium	PAK1, EZR	AG, IR	1.81×10 ⁻²
	GO:0030425	dendrite	KCNB1, KCNC2, NF1	DEP, RM	3.21×10 ⁻²
	GO:0005856	cytoskeleton	CDC42, TPM3, RARA, SPTBN1, LRRFIP1	AG, OB	3.60×10 ⁻²
	GO:0005096	GTPase activator activity	SRGAP2C, SRGAP2, NF1, ARHGAP35, SRGAP2B	SM, DEP	6.51×10 ⁻⁴
gene ontology molecular function	GO:0030695	GTPase regulator activity	SRGAP2C, SRGAP2, NF1, ARHGAP35, SRGAP2B,	SM, DEP	9.41×10 ⁻⁴
	GO:0015204	urea transmembrane transporter activity	SLC14A1, SLC14A2	DEP, HCD, SM	1.05×10 ⁻³
	GO:0042887	amide transmembrane transporter activity	SLC14A1, SLC14A2	DEP, HCD, SM	1.35×10 ⁻³
	GO:0004955	prostaglandin receptor activity	PTGFR, PTGER3	OB, AG	3.30×10 ⁻³
	GO:0015467	G-protein activated inward rectifier potassium channel activity	KCNJ15, KCNJ4	HCD, SM	3.84×10 ⁻³
	GO:0022838	substrate-specific channel activity	SLC14A1	DEP, HCD	4.19×10 ⁻³
	GO:0005249	voltage-gated potassium channel activity	KCNB1, KCNC2, KCNJ15	RM, HCD	6.88×10 ⁻³
	GO:0005242	inward rectifier potassium channel activity	KCNJ15, KCNJ4	HCD, SM	7.68×10 ⁻³
	GO:0003680	AT DNA binding	HMGA2	RM, HFD	1.15×10 ⁻²

treatment were demonstrated in this study. The molecular signatures and repurposing of the drugs presented in this study may thus deserve attention for use in further experimental studies of CMP.

Table 2 Some significant KEGG pathways those are common among inflammatory CMP and other risk factors such as SM (smoking), IR, AC, HCD (high caloric diet), RM (high red meat intake), depression, HFD (high fat diet), obesity and ageing

KEGG ID	Pathway	Genes in pathway	Risk factors
hsa00410	beta-Alanine metabolism	CNDP1, GAD2	AC, RM
hsa04010	MAPK signalling pathway	CACNB1, CDC42, NTRK2, PAK1, FGFR2, NF1	AG, DEP
hsa04014	RAS signalling pathway	CDC42, PAK1, FGFR2, NF1	AG, DEP
hsa04510	focal adhesion	CDC42, PAK1, ITGB7, LAMA4, ARHGAP35	AG, SM
hsa05032	morphine addiction	PDE1A, GABRB1, GNB5	OB, RM
hsa05200	pathways in cancer	CDC42, TPM3, PTGER3, RARA, FGFR2, KITLG, LAMA4, CDK2, PML	AG, SM
hsa05202	transcriptional misregulation in cancer	RARA, ATM, ITGB7, HMGA2, GRIA3	AG, RM
hsa05410	hypertrophic cardiomyopathy (HCM)	CACNB1, TPM3, ITGB7, ITGB1, ITGA4, SLC8A1	AG, HFD
hsa05412	arrhythmogenic right ventricular cardiomyopathy (ARVC)	CACNB1, ITGB7, ITGB1, ITGA4, SLC8A1	AG, HFD
hsa05414	dilated cardiomyopathy	CACNB1, TPM3, ITGB7, ITGB1, ITGA4, SLC8A1	AG, HFD

Table 3 Summary of transcriptional and/or post-transcriptional regulators (TFs and microRNAs) of the differentially expressed genes

	Symbol	Description	Feature
TFs	FOXC1	Forkhead Box C1	play critical role in early cardiomyogenesis [48]
	GATA2	GATA Binding Protein 2	afflicted with early onset familial coronary artery disease [49]
	FOXL1	Forkhead Box L1	associated with good outcome in human pancreatic ductal adenocarcinoma [50]
	YY1	YY1 TF	increased in human heart failure [51]
	CREB1	CAMP Responsive Element Binding Protein 1	cardiac failure is afflicted with CREB [52]
microRNAs	mir-335-5p	MicroRNA 335	upregulated in heart failure; involved in hypertrophic cardiomyopathy pathway [53]
	mir-26b-5p	MicroRNA 26	associated with suppression of proliferation and enhance the apoptosis in multiple myeloma cells [54]
	mir-34a-5p	MicroRNA 34	could prevent autophagic cell deaths in ischaemic hearts and in this way can improve the myocardial injury [55]
	mir-92a-3p	MicroRNA 92	increase blood vessel growth and recovery of damaged tissues in myocardial infarction [56]
	mir-17-5p	MicroRNA 17	prognostic markers of hepatocellular carcinoma [57]

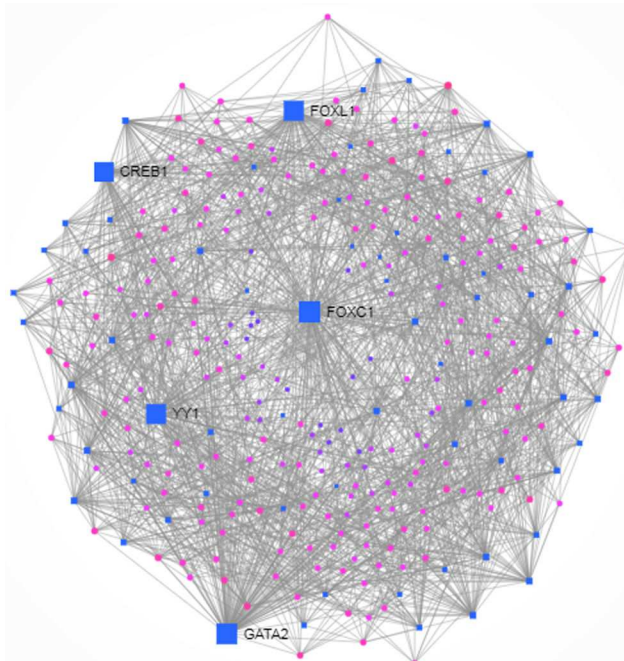


Fig. 4 Differentially expressed genes and TF interactions were analysed to identify the TFs that regulate the differentially expressed genes in CMP

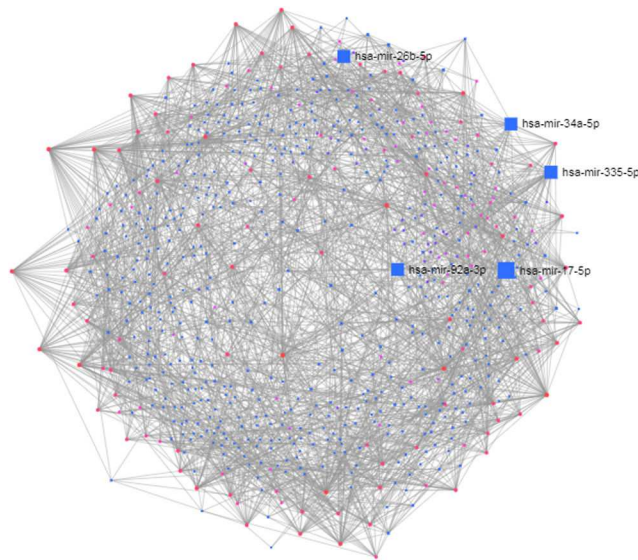


Fig. 5 Differentially expressed genes and microRNAs interactions were analysed to identify the microRNAs that regulate the differentially expressed genes in CMP

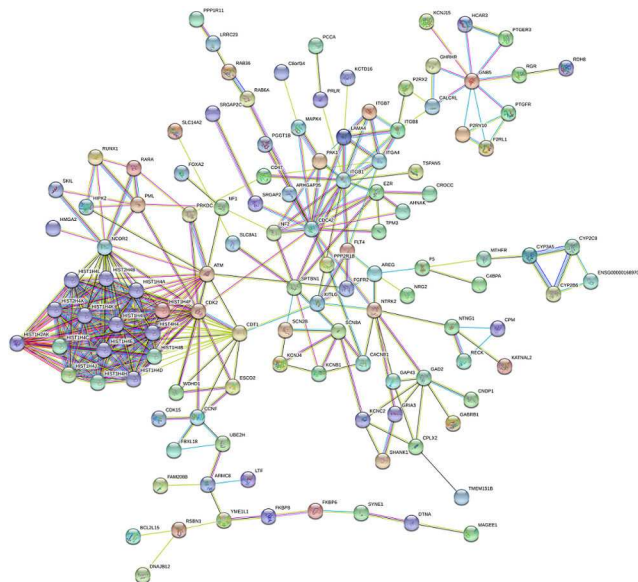


Fig. 6 PPI network of differentially expressed genes that were common between CMP and other risk factors

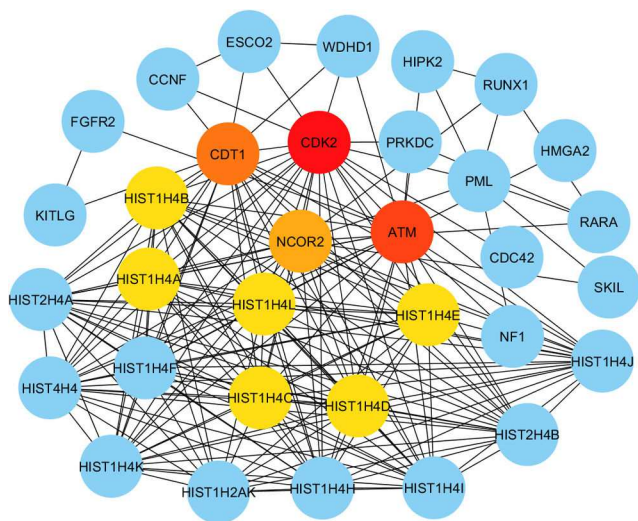


Fig. 7 Simplified PPI network of the common differentially expressed genes between CMP and the risk factors. Ten significant hub proteins are marked as red, orange and yellow, respectively

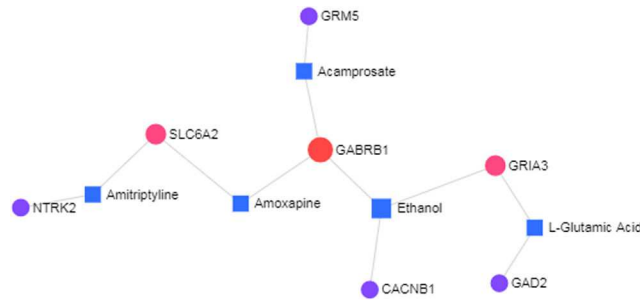


Fig. 8 Protein–drug interactions network. The interactions between drugs and the hub node (*TUBB*) were represented. The area of the node represents the degree of interaction in the network

Table 4 Summary of hub proteins identified from PPI network

Symbol	Description	Gene ontology	Feature
CDK2	Cyclin Dependent Kinase 2	transferase activity	involved in the regulation of myocardial ischaemia and reperfusion injury [67]
ATM	ATM Serine/Threonine Kinase	transferase activity	associated with electromechanical dyssynchrony in adult dilated CMP [68]
CDT1	Chromatin Licensing and DNA Replication Factor 1	—	genotoxic stress which result in aberrant cell proliferation [69]
NCOR2	Nuclear Receptor Corepressor 2	sequence-specific DNA binding	associated with non-alcoholic fatty liver disease [70]
HIST1H4A	Histone Cluster 1 H4 Family Member A	histone binding	implicated a strong involvement of inflammatory-immune pathways [71]
HIST1H4B	Histone Cluster 1 H4 Family Member B	histone binding	dysregulation of HIST1H4B in rat cardiomyocytes [72]
HIST1H4C	Histone Cluster 1 H4 Family Member C	histone binding	hyperinsulinemic hypoglycemia familial 1 [73]
HIST1H4D	Histone Cluster 1 H4 Family Member D	histone binding	involved in meiosis and signalling pathways by Rho GTPases [73]
HIST1H4E	Histone Cluster 1 H4 Family Member E	histone binding	involved in meiosis and signalling pathways by Rho GTPases [73]
HIST1H4L	Histone Cluster 1 H4 Family Member L	histone binding	involved in meiosis and signalling pathways by Rho GTPases [73]

Table 5 Gene-disease association analysis of differentially expressed genes of nine risk factors using OMIM and dbGAP databases for CMP

Risk fac./Causes	Gene	Adjusted P-value
AG	SGCD, TNNT2, TNNC1, LDB3, TTN	1.85×10^{-3}
SM	OPRD1, SMAD1, BUB1, PKP2	2.18×10^{-1}
DEP	HNF4A, TSHR	4.18×10^{-2}
IR	WVVOX, MCTP2, BACH2, ZNF783, PPARA	4.62×10^{-1}
RM	HMGA2, ADCY2, MYH6	3.34×10^{-1}
AC	VWF	2.42×10^{-1}
OB	ABCC9, TFAP2A, DMD, FOS	1.45×10^{-1}
HCD	SMAD1, DSC2	1.02×10^{-1}
HFD	AKAP13, CREBBP, TGFB3, HNF4A, HMGA2, FMN2, ADAMTS7, TGFB3	2.29×10^{-1}

6 References

- [1] Maron, B.J., Towbin, J.A., Thiene, G., *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), pp. 1807–1816
- [2] Disease, G., Incidence, I., Prevalence, C.: ‘Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015’, *Lancet*, 2016, **388**, (10053), pp. 1545–1602
- [3] Towbin, J.A., Lowe, A.M., Colan, S.D., *et al.*: ‘Incidence, causes, and outcomes of dilated cardiomyopathy in children’, *J. Am. Med. Assoc.*, 2006, **296**, (15), pp. 1867–1876
- [4] Chowdhury, U.N., Ahmad, S., Islam, M.B., *et al.*: ‘Network-based identification of genetic factors in ageing, lifestyle and type 2 diabetes that influence in the progression of alzheimer’s disease’, 2018
- [5] Sakib, N., Chowdhury, U.N., Islam, M.B., *et al.*: ‘A system biology approach to identify the genetic markers to the progression of Parkinson’s disease for aging, lifestyle and type 2 diabetes’, *bioRxiv*, 2018, p. 482760
- [6] Barabási, A.L., Gulbahce, N.: ‘Network medicine: a network-based approach to human disease’, *Nat. Rev. Genetics*, 2011, **12**, (1), pp. 56–68
- [7] Rahman, M.R., Islam, T., Zaman, T., *et al.*: ‘Identification of molecular signatures and pathways to identify novel therapeutic targets in Alzheimer’s disease: insights from a systems biomedicine perspective’, *Genomics*, 2019, in press, Available at <https://www.sciencedirect.com/science/article/pii/S0888754319301843?via%3Dihub>
- [8] Satu, M.S., Howlader, K.C., Akhund, T.M.N.U., *et al.*: ‘Bioinformatics approach to identify diseaseome and comorbidities effect of mitochondrial dysfunctions on the progression of neurological disorders’, *bioRxiv*, 2018, p. 483065
- [9] Rahman, M.R., Islam, T., Turanji, B., *et al.*: ‘Network-based approach to identify molecular signatures and therapeutic agents in Alzheimer’s disease’, *Comput. Biol. Chem.*, 2019, **78**, pp. 431–439
- [10] Rahman, M., Islam, T., Shahjaman, M., *et al.*: ‘Discovering biomarkers and pathways shared by Alzheimer’s disease and ischemic stroke to identify novel therapeutic targets’, *Medicina*, 2019, **55**, (5), p. 191
- [11] Rahman, M.R., Islam, T., Shahjaman, M., *et al.*: ‘Identification of common molecular biomarker signatures in blood and brain of Alzheimer’s disease’, *BioRxiv*, 2019, p. 482828

- [12] Islam, T., Islam, M.R.R., Shahjaman, M., *et al.*: 'Blood-based molecular biomarker signatures in Alzheimer's disease: insights from systems biomarker perspective', *bioRxiv*, 2018, p. 481879
- [13] Moni, M.A., Rana, H.K., Islam, M.B., *et al.*: 'A computational approach to identify blood cell-expressed Parkinson's disease biomarkers that are coordinately expressed in brain tissue', *Comput. Biol. Med.*, 2019, **113**, p. 103385
- [14] Rana, H.K., Akhtar, M.R., Islam, M.B., *et al.*: 'Genetic effects of welding fumes on the development of respiratory system diseases', *bioRxiv*, 2018, p. 480855
- [15] Moni, M.A., Rana, H.K., Islam, M.B., *et al.*: 'Detection of Parkinson's disease using blood and brain cells transcript profiles', *bioRxiv*, 2019, p. 483016
- [16] Islam, T., Rahman, M.R., Karim, M.R., *et al.*: 'Detection of multiple sclerosis using blood and brain cells transcript profiles: insights from comprehensive bioinformatics approach', *Inf. Med. Unlocked*, 2019, **16**, p. 100201
- [17] Rana, H.K., Akhtar, M.R., Ahmed, M.B., *et al.*: 'Genetic effects of welding fumes on the progression of neurodegenerative diseases', *Neurotoxicology*, 2019, **71**, pp. 93–101
- [18] Rahman, M., Islam, T., Gov, E., *et al.*: 'Identification of prognostic biomarker signatures and candidate drugs in colorectal cancer: insights from systems biology analysis', *Medicina*, 2019, **55**, (1), p. 20
- [19] Hossain, M.A., Asa, T.A., Huq, F., *et al.*: 'A network-based approach to identify molecular signatures and comorbidities of thyroid cancer'. Proc. of Int. Joint Conf. on Computational Intelligence, Springer, Singapore, 2020, pp. 235–246
- [20] Hossain, M.A., Asa, T.A., Rahman, M.R., *et al.*: 'Network-based approach to identify key candidate genes and pathways shared by thyroid cancer and chronic kidney disease', *Inf. Med. Unlocked*, 2019, **16**, p. 100240
- [21] Hossain, M.A., Asa, T.A., Quinn, J.M., *et al.*: 'Network-based genetic profiling, and therapeutic target identification of thyroid cancer', *bioRxiv*, 2018, p. 480632
- [22] Barrett, T., Wilhite, S.E., Ledoux, P., *et al.*: 'Ncbi gene: archive for functional genomics data sets-update', *Nucleic Acids Res.*, 2012, **41**, (D1), pp. D991–D995
- [23] Wittchen, F., Suckau, L., Witt, H., *et al.*: 'Genomic expression profiling of human inflammatory cardiomyopathy (dcmi) suggests novel therapeutic targets', *J. Mol. Med.*, 2007, **85**, (3), pp. 257–271
- [24] Ferrer-Martinez, A., Montell, E., Montori-Grau, M., *et al.*: 'Long-term cultured human myotubes decrease contractile gene expression and regulate apoptosis-related genes', *Gene*, 2006, **384**, pp. 145–153
- [25] Büttner, P., Mosig, S., Funke, H.: 'Gene expression profiles of t lymphocytes are sensitive to the influence of heavy smoking: a pilot study', *Immunogenetics*, 2007, **59**, (1), pp. 37–43
- [26] Iwamoto, K., Kakiuchi, C., Bundo, M., *et al.*: 'Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders', *Mol. Psychiatry*, 2004, **9**, (4), p. 406
- [27] Hardy, O.T., Perugini, R.A., Nicoloso, S.M., *et al.*: 'Body mass index-independent inflammation in omental adipose tissue associated with insulin resistance in morbid obesity', *Surg. Obes. Relat. Dis.*, 2011, **7**, (1), pp. 60–67
- [28] Hebels, D.G., Svejce, K.M., de Kok, M.C., *et al.*: 'N-nitroso compound exposure-associated transcriptomic profiles are indicative of an increased risk for colorectal cancer', *Cancer Lett.*, 2011, **309**, (1), pp. 1–10
- [29] McClintick, J.N., Xuei, X., Tischfield, J.A., *et al.*: 'Stress-response pathways are altered in the hippocampus of chronic alcoholics', *Alcohol*, 2013, **47**, (7), pp. 505–515
- [30] Oñate, B., Vilahur, G., Camino-López, S., *et al.*: 'Stem cells isolated from adipose tissue of obese patients show changes in their transcriptomic profile that indicate loss in stemcellness and increased commitment to an adipocyte-like phenotype', *BMC Genomics*, 2013, **14**, (1), p. 625
- [31] Gille, D., Zangger, N., Soneson, C., *et al.*: 'Caloric dose-responsive genes in blood cells differentiate the metabolic status of obese men', *J. Nutr. Biochem.*, 2017, **43**, pp. 156–165
- [32] Kakehi, S., Tamura, Y., Takeno, K., *et al.*: 'Increased intramyocellular lipid/ impaired insulin sensitivity is associated with altered lipid metabolic genes in muscle of high responders to a high-fat diet', *Am. J. Phys.-Endocrinol. Metabolism*, 2015, **310**, (1), pp. E32–E40
- [33] Xu, H., Moni, M.A., Liò, P.: 'Network regularised Cox regression and multiplex network models to predict disease comorbidities and survival of cancer', *Comput. Biol. Chem.*, 2015, **59**, pp. 15–31
- [34] Kuleshov, M.V., Jones, M.R., Rouillard, A.D., *et al.*: 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic Acids Res.*, 2016, **44**, (W1), pp. W90–W97
- [35] Moni, M.A., Liò, P.: 'Genetic profiling and comorbidities of zika infection', *J. Infect. Dis.*, 2017, **216**, (6), pp. 703–712
- [36] Khan, A., Fornes, O., Stigliani, A., *et al.*: 'Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework', *Nucleic Acids Res.*, 2017, **46**, (D1), pp. D260–D266
- [37] Sethupathy, P., Corda, B., Hatzigeorgiou, A.G.: 'Tarbase: A comprehensive database of experimentally supported animal microRNA targets', *Rna*, 2006, **12**, (2), pp. 192–197
- [38] Moni, M.A., Liò, P.: 'Network-based analysis of comorbidities risk during an infection: sars and hiv case studies', *BMC Bioinform.*, 2014, **15**, (1), p. 1
- [39] Hsu, S.D., Lin, F.M., Wu, W.Y., *et al.*: 'Mirtarbase: a database curates experimentally validated microRNA-target interactions', *Nucleic Acids Res.*, 2010, **39**, (suppl_1), pp. D163–D169
- [40] Szklarczyk, D., Morris, J.H., Cook, H., *et al.*: 'The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible', *Nucleic Acids Res.*, 2017, **45**, (D1), pp. D362–D368, <https://doi.org/10.1093/nar/gkw937>
- [41] Smoot, M.E., Ono, K., Ruscheinski, J., *et al.*: 'Cytoscape 2.8: new features for data integration and network visualization', *Bioinformatics*, 2010, **27**, (3), pp. 431–432
- [42] Moni, M.A., Xu, H., Liò, P.: 'Cytocom: a cytoscape app to visualize, query and analyse disease comorbidity networks', *Bioinformatics*, 2014, **31**, (6), pp. 969–971
- [43] Chen, S.H., Chin, C.H., Wu, H.H., *et al.*: 'Cyto-hubba: a cytoscape plug-in for hub object analysis in network biology', 2009
- [44] Moni, M.A., Liò, P.: 'How to build personalized multi-omics comorbidity profiles', *Front. Cell. Dev. Biol.*, 2015, **3**, p. 28
- [45] Calimlioglu, B., Karagoz, K., Sevimgolu, T., *et al.*: 'Tissue-specific molecular biomarker signatures of type 2 diabetes: an integrative analysis of transcriptomics and protein-protein interaction data', *Omics: a J. Integrative Biol.*, 2015, **19**, (9), pp. 563–573
- [46] Moni, M.A., Liò, P.: 'Comor: a software for disease comorbidity risk assessment', *J. Clin. Bioinform.*, 2014, **4**, (1), p. 1
- [47] Wishart, D.S., Feunang, Y.D., Guo, A.C., *et al.*: 'Drugbank 5.0: a major update to the drugbank database for 2018', *Nucleic Acids Res.*, 2017, **46**, (D1), pp. D1074–D1082
- [48] Lambers, E., Arnone, B., Fatima, A., *et al.*: 'Foxc1 regulates early cardiomyogenesis and functional properties of embryonic stem cell derived cardiomyocytes', *Stem Cells*, 2016, **34**, (6), pp. 1487–1500
- [49] Connelly, J.J., Wang, T., Cox, J.E., *et al.*: 'Gata2 is associated with familial early-onset coronary artery disease', *PLoS Genet.*, 2006, **2**, (8), p. e139
- [50] Zhang, G., He, P., Gaedcke, J., *et al.*: 'Foxl1, a novel candidate tumor suppressor, inhibits tumor aggressiveness and predicts outcome in human pancreatic cancer', *Cancer Res.*, 2013, **73**, (17), pp. 5416–5425
- [51] Sucharov, C.C., Mariner, P., Long, C., *et al.*: 'Yin yang 1 is increased in human heart failure and represses the activity of the human α -myosin heavy chain promoter', *J. Biol. Chem.*, 2003, **278**, (33), pp. 31233–31239
- [52] Watson, P.A., Birdsey, N., Huggins, G.S., *et al.*: 'Cardiac-specific overexpression of dominant-negative creb leads to increased mortality and mitochondrial dysfunction in female mice', *Am. J. Physiol.-Heart Circulatory Phys.*, 2010, **299**, (6), pp. H2056–H2068
- [53] Van-Rooij, E., Sutherland, L.B., Thatcher, J.E., *et al.*: 'Dysregulation of microRNAs after myocardial infarction reveals a role for mir-29 in cardiac fibrosis', *Proc. Natl. Acad. Sci.*, 2008, **105**, (35), pp. 13027–13032
- [54] Jia, C.M., Tian, Y.Y., Quan, L.N., *et al.*: 'Mir-26b-5p suppresses proliferation and promotes apoptosis in multiple myeloma cells by targeting jag1', *Pathol.-Res. Pract.*, 2018, **214**, (9), pp. 1388–1394
- [55] Shao, H., Yang, L., Wang, L., *et al.*: 'MicroRNA-34a protects myocardial cells against ischemia-reperfusion injury through inhibiting autophagy via regulating $\text{tnf } \alpha$ expression', *Biochem. Cell Biol.*, 2017, **96**, (3), pp. 349–354
- [56] Bonauer, A., Carmona, G., Iwasaki, M., *et al.*: 'MicroRNA-92a controls angiogenesis and functional recovery of ischemic tissues in mice', *Science*, 2009, **324**, (5935), pp. 1710–1713
- [57] Chen, L., Jiang, M., Yuan, W., *et al.*: 'Mir-17-5p as a novel prognostic marker for hepatocellular carcinoma', *J. Invest. Surg.*, 2012, **25**, (3), pp. 156–161
- [58] Seo, S., Kume, T.: 'Forkhead transcription factors, foxc1 and foxc2, are required for the morphogenesis of the cardiac outflow tract', *Dev. Biol.*, 2006, **296**, (2), pp. 421–436
- [59] Sun, C., Tong, L., Zhao, W., *et al.*: 'Microarray analysis reveals altered circulating microRNA expression in mice infected with coxsackievirus b3', *Exp. Ther. Med.*, 2016, **12**, (4), pp. 2220–2226
- [60] Sun, T., Li, M.Y., Li, P.F., *et al.*: 'MicroRNAs in cardiac autophagy: small molecules and big role', *Cells*, 2018, **7**, (8), p. 104
- [61] Muslin, A.J.: 'Mapk signaling in cardiovascular health and disease: molecular mechanisms and therapeutic targets', *Clin. Sci. (Lond)*, 2008, **115**, (7), pp. 203–218
- [62] Wang, Y.: 'Mitogen-activated protein kinases in heart development and diseases', *Circulation*, 2007, **116**, (12), pp. 1413–1423
- [63] Rose, B.A., Force, T., Wang, Y.: 'Mitogen-activated protein kinase signaling in the heart: angels versus demons in a heart-breaking tale', *Physiol. Rev.*, 2010, **90**, (4), pp. 1507–1546
- [64] Gelb, B.D., Tartaglia, M.: 'Ras signaling pathway mutations and hypertrophic cardiomyopathy: getting into and out of the thick of it', *J. Clin. Invest.*, 2011, **121**, (3), pp. 844–847
- [65] Li, J.: 'Alterations in cell adhesion proteins and cardiomyopathy', *World. J. Cardiol.*, 2014, **6**, (5), p. 304
- [66] Kapelko, V.I.: 'Extracellular matrix alterations in cardiomyopathy: the possible crucial role in the dilative form', *Exp. Clinical Cardiol.*, 2001, **6**, (1), p. 41
- [67] Liem, D.A., Zhao, P., Angelis, E., *et al.*: 'Cyclindependent kinase 2 signaling regulates myocardial ischemia/reperfusion injury', *J. Mol. Cell. Cardiol.*, 2008, **45**, (5), pp. 610–616
- [68] Hui, W., Slorach, C., Friedberg, M.K.: 'Apical transverse motion is associated with interventricular mechanical delay and decreased left ventricular function in children with dilated cardiomyopathy', *J. Am. Soc. Echocardiogr.*, 2018, **31**, (8), pp. 943–950
- [69] Pozo, P., Cook, J.: 'Regulation and function of cdt1; a key factor in cell proliferation and genome stability', *Genes*, 2016, **8**, (1), p. 2
- [70] Sookoian, S., Pirola, C.J.: 'Systems biology elucidates common pathogenic mechanisms between nonalcoholic and alcoholic-fatty liver disease', *PLOS One*, 2013, **8**, (3), p. e58895
- [71] Meng, H., Liang, Y., Hao, J., *et al.*: 'Comparison of rejection-specific genes in peripheral blood and allograft biopsy from kidney transplant', *Transplant. Proc.*, 2018, **50**, pp. 115–123
- [72] Yin, H.K., Li, X.Y., Jiang, Z.G., *et al.*: 'Progress in neuregulin/erbB signaling and chronic heart failure', *World. J. Hypertens.*, 2015, **5**, (2), pp. 63–73
- [73] Safran, M., Dalah, I., Alexander, J., *et al.*: 'Genecards version 3: the human gene integrator', *Database*, 2010, **2010**, baq020
- [74] Liao, H.S., Kang, P.M., Nagashima, H., *et al.*: 'Cardiac-specific overexpression of cyclin-dependent kinase 2 increases smaller mononuclear cardiomyocytes', *Circ. Res.*, 2001, **88**, (4), pp. 443–450

- [75] Huisamen, B., Strijdom, H., Collop, N., *et al.*: 'A possible role for the atm protein in the myocardial pathology associated with obesity and insulin resistance', *Cardiovasc. Res.*, 2014, **103**, p. S118