

**Original article:**

**COMMON DIFFERENTIALLY EXPRESSED GENES AND PATHWAYS  
CORRELATING BOTH CORONARY ARTERY DISEASE AND  
ATRIAL FIBRILLATION**

Youjing Zheng, Jia-Qiang He\*

Department of Biomedical Sciences and Pathobiology, College of Veterinary Medicine,  
Virginia Tech, Blacksburg, VA 24061, USA

\* **Corresponding author:** Jia-Qiang He, Department of Biomedical Sciences and  
Pathobiology, Virginia Tech, Phase II, Room 252B, Blacksburg, VA 24061, USA.  
Tel: 1-540-231-2032. E-mail: [jiahe@vt.edu](mailto:jiahe@vt.edu)

<https://orcid.org/0000-0002-4825-7046> Youjing Zheng

<https://orcid.org/0000-0002-0640-5960> Jia-Qiang He

<http://dx.doi.org/10.17179/excli2020-3262>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License  
(<http://creativecommons.org/licenses/by/4.0/>).

**ABSTRACT**

Coronary artery disease (CAD) and atrial fibrillation (AF) share common risk factors, such as hypertension and diabetes. The patients with CAD often suffer concomitantly AF, but how two diseases interact with each other at cellular and molecular levels remain largely unknown. The present study aims to dissect the common differentially expressed genes (DEGs) that are concurrently associated with CAD and AF. Two datasets [GSE71226 for CAD] and GSE31821 for AF] were analyzed with GEO2R and Venn Diagram to identify the DEGs. Signaling pathways, gene enrichments, and protein-protein interactions (PPI) of the identified common DEGs were further analyzed with Kyoto Encyclopedia of Gene and Genome (KEGG), Database for Annotation, Visualization and Integrated Discovery (DAVID), and Search Tool for the Retrieval of Interacting Genes (STRING). 565 up- and 1367 down-regulated genes in GSE71226 and 293 up- and 68 down-regulated genes in GSE31821 were identified. Among those, 21 common DEGs were discovered from both datasets, which lead to the findings of 4 CAD and 21 AF pathways, 3 significant gene enrichments (intracellular cytoplasm, protein binding, and vascular labyrinthine layer), and 3 key proteins (membrane metallo-endopeptidase (MME), transferrin receptor 1 (TfR1), and Lyso-some-associated membrane glycoprotein 1 (LAMP1)). Together, these data implied that these three proteins may play a central role in development of both CAD and AF.

**Keywords:** Coronary artery disease, atrial fibrillation, differentially expressed gene, signaling pathway, bioinformatics

**INTRODUCTION**

Cardiovascular disease is the leading cause of death in the developed countries (Virani et al., 2020). Of all cardiovascular diseases (*e.g.*, acute myocardial infarction, heart

failure, valvular heart disease, cerebrovascular accident, transient ischemic attack, peripheral arterial disease, sudden cardiac arrest, ventricular arrhythmia, venous thromboembolism, and pulmonary embolism), coronary artery disease (CAD) is the most common type and contributes the highest rate of death

(Michniewicz et al., 2018; Virani et al., 2020); whereas of all cardiac arrhythmia (*e.g.*, supraventricular tachycardia, ventricular tachycardia, sinus-node dysfunction, and heart block), atrial fibrillation (AF) is the most typical disorder, and it affects about 37.6 million individuals globally in 2017 (Go et al., 2001; Michniewicz et al., 2018; Virani et al., 2020).

Interestingly, studies found that AF is highly associated with the increased risk of many other diseases, such as CAD, stroke, heart failure, diabetes, sudden cardiac death, and mortality, especially within the aging populations (Motloch et al., 2017; Murakami et al., 2017; Virani et al., 2020). In the case of CAD, it was demonstrated that both AF and CAD share the same risk factors and impact on each other (Kristensen et al., 2020; Lieder et al., 2018; Motloch et al., 2017). A systematic review and meta-analysis of 15 cohort studies, for example, demonstrated that AF was associated with a 1.54-fold increased risk of myocardial infarction induced by CAD (Ruddox et al., 2017). Overall, about 17-46.5 % patients with AF suffer concomitantly CAD while the patients with CAD have a low prevalence rate (0.2 % to 5 %) of AF, suggesting the significant effects of AF on promoting morbidity and mortality of concomitant diseases (Michniewicz et al., 2018).

Meanwhile, the outcomes of the patients with CAD is modulated by AF; however, it's still unclear whether the presence of CAD simply increases the risk of AF or changes the impact of other risk factors (Mehta et al., 2003; Pilgrim et al., 2013). The management of AF with concomitant CAD is still a huge

clinical challenge (Gladding et al., 2020). Fully understanding the similarities in the pathogenesis of AF and CAD may reveal the mechanisms underlying both diseases and facilitate discovery of new therapy targets.

Bioinformatic analysis of gene profiles offers a novel approach to explore the underlying mechanisms of disease at the molecular level. This technique has been widely utilized in basic and clinical studies (Kumar et al., 2016), yet only limited data is reported regarding interlinkages of critical genes and signaling molecules between CAD and AF (Kertai et al., 2015). In this paper, we aimed to profile the common differentially expressed genes (DEGs) of CAD and AF by using the sequencing databases of these two diseases and identified potential pathways modulating the development of CAD and AF.

## MATERIALS AND METHODS

### Data sources

The datasets of gene of interest with sequence number GSE71226 and GSE31821 were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). In GSE71226 microarray dataset, 3 samples from the patients with CAD and 3 samples of healthy subjects were included; while in GSE31821 dataset, 4 samples from the patients with AF and 2 samples of healthy subjects were enrolled. Both datasets were collected from GPL570 Platforms ((HG-U133\_Plus\_2) Affymetrix Human Genome U133 Plus 2.0 Array). The detail information is shown in Table 1.

**Table 1: Basic information of datasets used in the study**

Items	GSE71226	GSE31821
Platform	GPL570	GPL570
Diseases	CAD	AF
Research subjects	Asian	Caucasian
Samples	Peripheral blood	Cardiac (atrial) tissue
Number of subjects	6 (3 patients + 3 controls)	6 (4 patients + 2 controls)

### **Identifications of differentially expressed genes**

DEGs between patients and healthy subjects were identified via GEO2R online tools ( $\log_2FC > 1$  or  $\log_2FC < -1$ ,  $p$  value  $< 0.05$ ) (Davis and Meltzer, 2007). The row data were then run in Venn Diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to identify the common DEGs between 2 datasets. The DEGs with  $\log_2FC < -1$  were considered as down-regulated genes, while the DEGs with  $\log_2FC > 1$  were regarded as up-regulated genes. Heatmap of gene expression was made by R package ggplot2 as described previously (Aibar et al., 2015; Walter et al., 2015).

### **Gene ontology enrichment analysis**

The Gene ontology (GO) analysis has become a common way to analyze large scale genomic data (Zheng et al., 2008). Kyoto encyclopedia of genes and genomes (KEGG) (<https://www.genome.jp/kegg/>) is a biological genomic database that focuses on computerization of molecular linkage among genomes, gene functions, and biochemical (metabolic and regulatory) pathways of all organisms under normal and disease conditions (Ogata et al., 1999). Database for Annotation, Visualization and Integration Discovery (DAVID) (<https://david.ncifcrf.gov/>) and Cytoscape software (<https://cytoscape.org/>) were used for the GO enrichment and KEGG pathway analysis of integrated differential genes. DAVID (v6.8) is an online bioinformatic tool that is designed to identify gene and protein functions and visualize different signaling pathways. In these analyses, bar plots were made by R package heatmap to show the ten most significant enriched GO terms (Aibar et al., 2015; Walter et al., 2015)

### **Protein-protein interaction network mapping**

The online Search Tool for the Retrieval of Interacting Genes (STRING) (<https://string-db.org/>) was used to analyze

the protein-protein interaction (PPI) network of the DEGs as described in the previously published study (Szklarczyk et al., 2015). Since PPI is known to modulate a variety of biological process, such as cellular metabolisms, development processes, and cell-to-cell interactions, thus, it could be used to predict key protein(s) that regulate cellular specific functions or to be screened as potential therapeutic target(s) (Rao et al., 2014).

## **RESULTS**

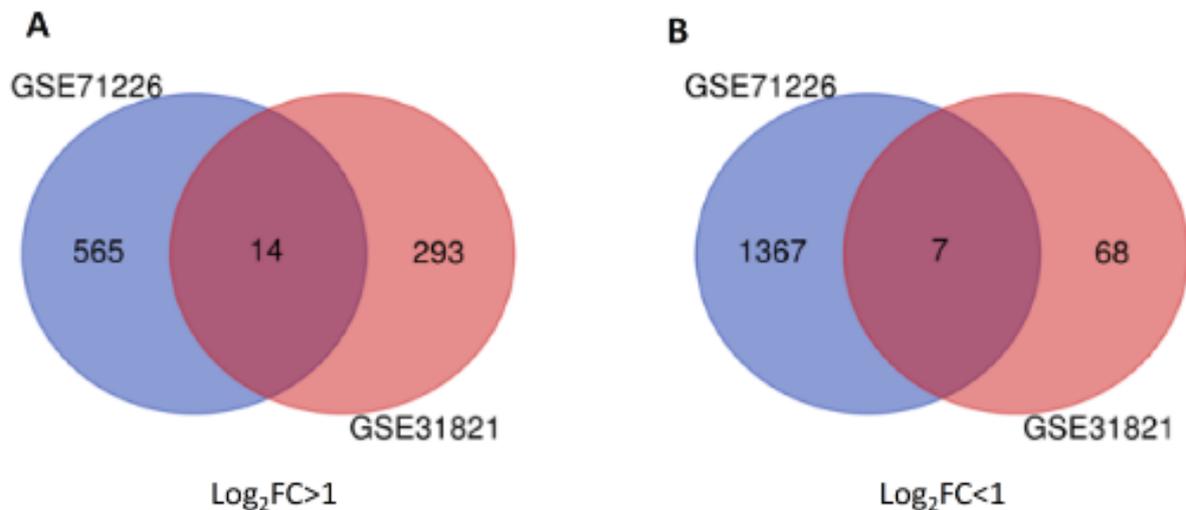
### **Up- and down-regulated genes that were concurrently expressed in the patients with CAD and AF**

To outline the profiles of DEGs, two datasets were analyzed with GEO2R software. From GSE71226 dataset of the patients with CAD patients and healthy subjects, a total of 1932 DEGs was identified, among which 565 genes were up-regulated ( $p < 0.05$ ,  $\log_2FC > 1$ ) and 1367 genes were down-regulated ( $p < 0.05$ ,  $\log_2FC < -1$ ) (Supplementary Table 2 and Figure 1). Similarly, from GSE31821 dataset of the patients with AF and healthy subjects, a total of 361 genes were extracted, among which 293 genes were up-regulated ( $p < 0.05$ ,  $\log_2FC > 1$ ) and 68 genes were down-regulated ( $p < 0.05$ ,  $\log_2FC < -1$ ) (Supplementary Table 3 and Figure 1), suggesting significantly differential expressions of multitudinous genes in the patients with CAD and AF.

To further determine the common DEGs that exist in both datasets, we ran two datasets on Venn Diagram and confirmed 21 common DEGs, which comprises 14 up-regulated genes ( $p < 0.05$ ,  $\log_2FC > 1$ ) and 7 down-regulated genes ( $p < 0.05$ ,  $\log_2FC < -1$ ) (Table 2 and Figure 1). To unveil the expression patterns of the DEGs among all groups, the top 100 DEGs were selected based on the  $p$ -values ( $p < 0.05$ ), and constructed as a cluster heatmap to show the cross-correlation of those genes among each individual.

**Table 2: Twenty-one common differentially expressed genes (DEGs) identified in two datasets**

Up- or down-regulated genes		GSE71226		GSE31831	
		Log FC	p-value	Log FC	p-value
Up-regulation (14)	Lysosomal associated membrane protein 1 (LAMP1)	1.07	0.0387	1.24	0.0392
	Polycomb group ring finger 5 (PCGF5)	1.7	0.0323	1.26	0.0400
	Glutamate-ammonia ligase (GLUL)	3.1	0.0406	1.03	0.0345
	Ankyrin repeat domain 2 (ANKRD2)	1.03	0.0430	2.02	0.0089
	Transducin beta like 1 X-linked receptor 1 (TBL1XR1)	1.81	0.0018	1.41	0.0156
	Membrane metalloendopeptidase (MME)	2.87	0.0041	1.08	0.0387
	Transmembrane p24 trafficking protein 2 (TMED2)	1.16	0.0192	1.34	0.0470
	Hes related family bHLH transcription factor with YRPW motif 1 (HEY1)	1.42	0.0312	1.23	0.0004
	Rho related BTB domain containing 3 (RHOBTB3)	1.51	0.0009	1.26	0.0086
	Cyclin D2 (CCND2)	1.05	0.0409	1.11	0.0018
	Interleukin 13 receptor subunit alpha 1 (IL13RA1)	2.87	0.0116	1.01	0.0428
	Upstream binding transcription factor, RNA polymerase I (UBTF)	1.16	0.0141	1.05	0.0330
	Transferrin receptor (TFRC)	1.38	0.0191	1.24	0.0463
	Mitogen-activated protein kinase kinase kinase kinase 5 (MAP4K5)	1.59	0.0447	1.04	0.0232
Down-regulation (7)	Leucine rich repeat neuronal 3 (LRRN3)	-1.56	0.0448	-2.40	0.0134
	Family with sequence similarity 208-member B (FAM208B)	-1.22	0.0095	-1.25	0.0339
	Brain expressed X-linked 2 (BEX2)	-1.96	0.0342	-2.17	0.0355
	Vacuolar protein sorting 13 homolog D (VPS13D)	-1.56	0.0283	-1.54	0.0246
	Zinc finger protein 493 (ZNF493)	-2.25	0.0006	-1.03	0.0379
	Zinc finger protein 302 (ZNF302)	-1.74	0.0289	-1.27	0.0326
	Phosphatidylinositol transfer protein, cytoplasmic 1 (PITPNC1)	-1.35	0.0100	-1.35	0.0188



**Figure 1: Twenty-one common DEGs from two datasets were identified in Venn Diagram. Panel A** shows 14 up-regulated DEGs (at the center) from GSE71226 dataset (in blue) and GSE31821 (in red). **Panel B** shows 7 down-regulated genes from GSE71226 dataset (in blue) and GSE31821 (in red). DEGs: differentially expressed genes. GSExxxx: gene set enrichment #; Log<sub>2</sub>FC > 1 or < 1: fold changes in logarithms to base 2 between patients and healthy subjects are greater (up-regulated) or lower (down-regulated) than 1.

As shown in Figure 2, there are significant differences of gene expression profiles between healthy subjects and patients. Overall, 5 healthy subjects from two datasets exhibited similar patterns of gene expressions except for group 1 (G1), 3, 5, and 9; while significant differences of expression levels were observed in all groups (G1-9) between 3 CAD patients and 4 AF patients with high expressions in G1, 2, 7 and low expressions in G3, 4, 5, 6, 8, and 9 in the patients with CAD compared to the same groups of the patients with AF. Interestingly, STX17-AS1, BAG1, GYPC, STRADB, S100A9, and HBM are the mostly expressed genes in the CAD patients while ACTA1, FHL2, FABP4, and EGR1 are the mostly expressed genes in the AF patients. The common DEGs between CAD and AF appear in G1 and 2, such as STEAP4, SLC6A8, GYPC, and STRADB. Together, these data demonstrated that majority of the DEGs expressed differently between CAD and AF, but a small group of genes expressed concurrently.

#### **Variable and common GO terms between CAD and AF**

To characterize the three critical terms, biological process (BP), molecular function (MF), and cellular component (CC) of the DEGs identified above, the GO (*i.e.*, over-representation or term enrichment) enrichment analysis was performed on two datasets and the results (*i.e.*, terms) are presented as graph (ontology) structure shown in Figure 3.

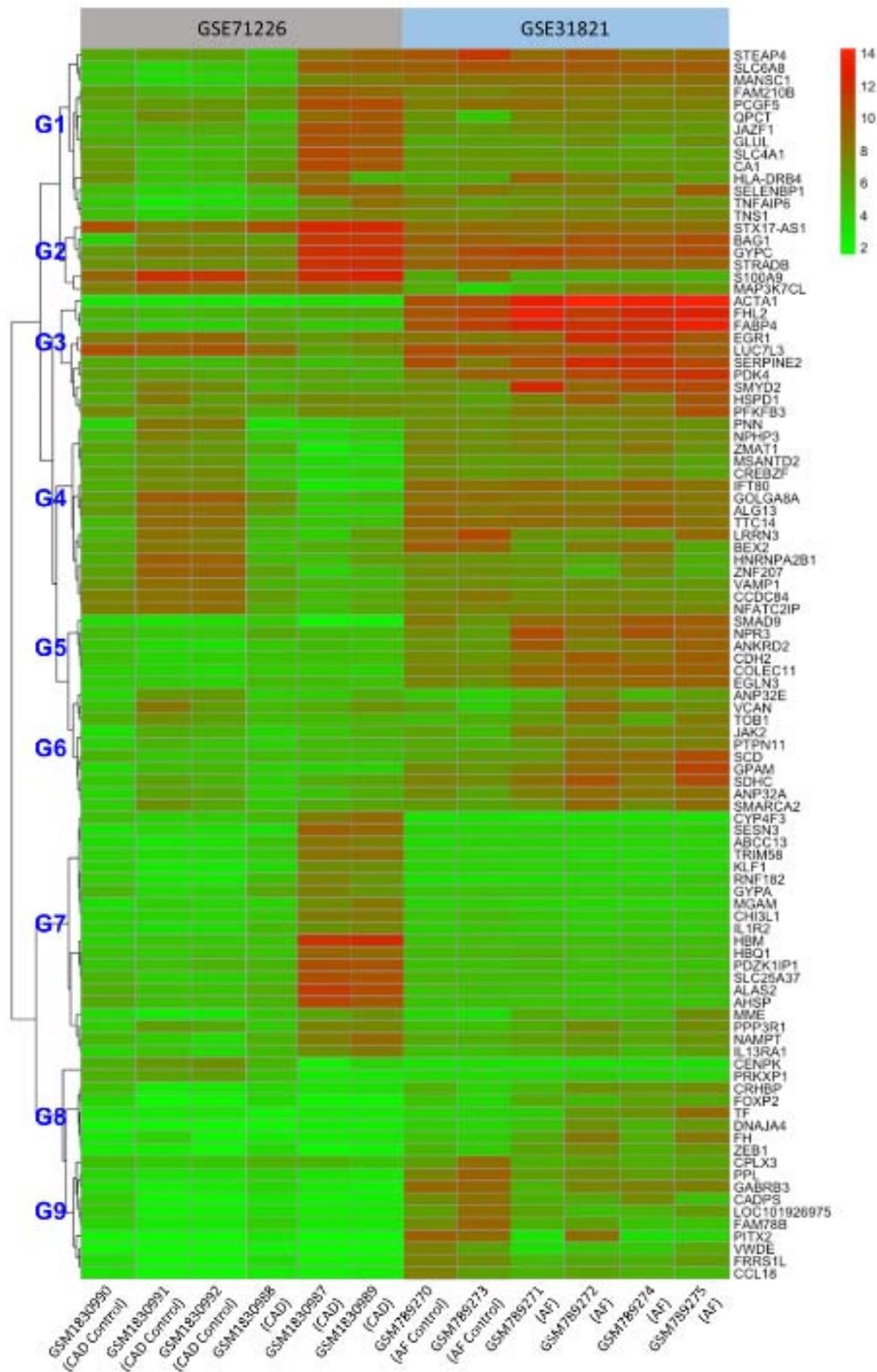
From GSE71226 dataset of the patients with CAD and healthy subjects, it was found that for the BP term, the DEGs were mostly enriched in the regulation of transcription (GO: 0006355,  $p = 1.91E-13$ ); while for MF term, they were mostly enriched in the DNA binding (GO:0003676,  $p = 5.75E-12$ ); and lastly for CC term, they were mostly enriched in the nucleoplasm (GO: 0005654,  $p = 1.24E-21$ ) (Table 3 and Figure 3A). Similarly, from GSE31821 dataset of the patients with AF and healthy subjects, it was found that for the BP term, the DEGs were mostly enriched in the extracellular matrix (GO:0030198,  $p = 2.07E-$

06); while for MF term, they were mostly enriched in the cadherin binding in cell-cell adhesion (GO:0098641,  $p = 1.62E-06$ ), and lastly for CC term, they were mostly enriched in the extracellular exosome (GO: 0070062,  $p = 4.89E-11$ ) (Table 3 and Figure 3B). Together, these data demonstrated that the DEGs identified from the patients with CAD or AF were expressed (enriched) differentially in the aspects of the BP, MF, and CC.

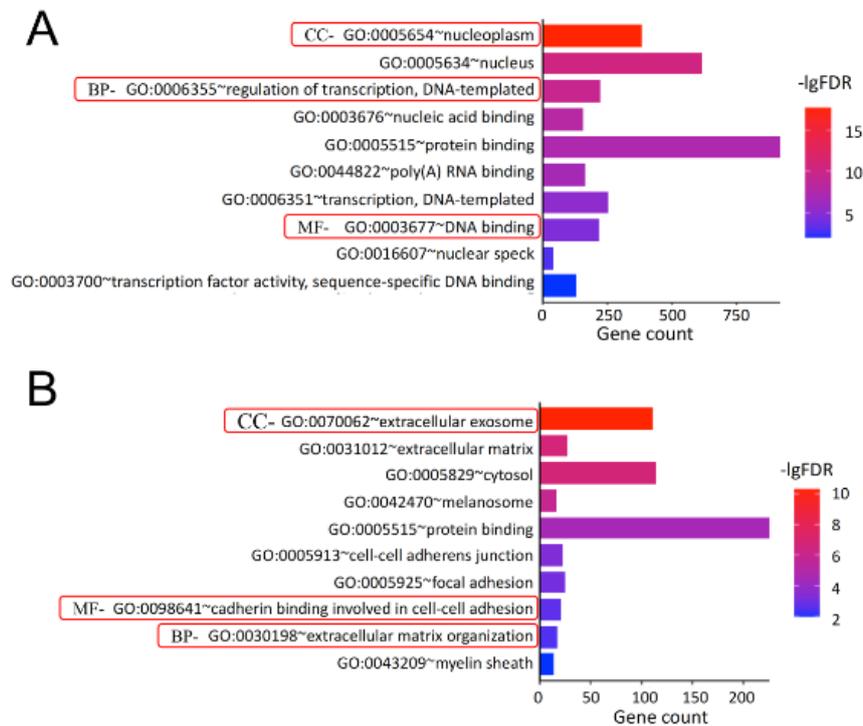
However, when the 21 common DEGs were analyzed, the resulting GO terms are different from the above. Specifically, for the BP, the common DEGs were particularly enriched in the regulation of labyrinthine layer in embryonic blood vessel; while for the MF, they were remarkably enriched in the protein binding, and lastly for the CC, they were substantially enriched in the intracellular cytoplasm (Table 4), implying that those terms (BP, MF, and CC) may represent the common pathogenesis in the development of CAD and AF.

#### **Numerous but not common pathways were detected in both CAD and AF**

Next, we used DAVID software to map the KEGG pathways of the identified DEGs from both datasets. Briefly, from GSE71226 dataset of the patients with CAD and healthy subjects, four key pathways were determined: 1) mRNA surveillance pathway; 2) eukaryotic ribosome biogenesis pathway; 3) glucagon signaling pathway; and 4) other types of O-glycan biosynthesis (Table 5). However, from GSE31821 dataset of the patients with AF and healthy subjects, twenty-one key pathways were discovered, including 1) Focal adhesion; 2) MAPK; 3) Amoebiasis; 4) Cancer; 5) PI3K-Akt; 6) Wnt signaling; 7) ECM-receptor interaction; 8) Platelet activation; 9) Toxoplasmosis; and 10) Proteoglycans in cancer (see Table 5 for the rest 11 pathways). Interestingly, only the hematopoietic cell lineage signaling pathway was enriched by the common DEGs, and the statistical test is close to significance ( $p=0.085$ ), implying that the hematopoietic cell lineage signaling pathway



**Figure 2: Cluster heatmap of the top 100 DEGs from both GSE71226 and GSE31821 datasets.** The colored codes from green to red indicate expression levels from low (in green) to high (in red). The sample identification # (GSM....) is listed on the bottom (x-axis). The left gene tree was roughly grouped into 9 groups (G1 to 9) for easy reference in text. The right side lists all 100 genes. See Figure 1 and Supplementary Table 1 for other abbreviations.



**Figure 3: GO enrichment analysis of the DEGs. Panel A** shows the results of top 10 GO enrichment in GSE71226 dataset. **Panel B** shows the results of top 10 GO enrichment GSE31821 dataset. In all panels, X-axis represents counts of the DEGs and Y-axis refers the enriched GO terms (BP, MF, and CC). The colored codes from blue to red indicate significance from low to high. GO: gene ontology; BP: biological process; MF: molecular function; CC: cellular components; -lgFDR: Log<sub>10</sub> false discovery rate ( $p$ -value). See Figure 1, 2 and Supplementary Table 1 for other abbreviations.

**Table 3: GO analysis of the top 10 genes in two datasets**

	Category	GO-ID	Description	Count	%	p-value	FDR
GSE71226	GOTERM_CC	0005654	nucleoplasm	386	20.348	1.24E-21	1.89E-18
	GOTERM_CC	0005634	nucleus	617	32.525	2.69E-14	4.10E-11
	GOTERM_BP	0006355	regulation of transcription, DNA-templated	221	11.650	1.91E-13	3.57E-10
	GOTERM_MF	0003676	nucleic acid binding	155	8.1708	5.75E-12	9.44E-09
	GOTERM_MF	0005515	protein binding	919	48.445	3.14E-11	5.16E-08
	GOTERM_MF	0044822	poly(A) RNA binding	167	8.8034	1.21E-10	1.99E-07
	GOTERM_BP	0006351	transcription, DNA-templated	252	13.284	3.32E-09	6.18E-06
	GOTERM_MF	0003677	DNA binding	217	11.439	2.42E-08	3.98E-05
	GOTERM_CC	0016607	nuclear speck	41	2.1613	9.40E-07	0.001433
	GOTERM_MF	0003700	transcription factor activity, sequence-specific DNA binding	128	6.7475	7.01E-06	0.011501
GSE31821	GOTERM_CC	0070062	extracellular exosome	111	29.365	3.47E-14	4.89E-11
	GOTERM_CC	0031012	extracellular matrix	27	7.1429	1.29E-10	1.82E-07
	GOTERM_CC	0005829	cytosol	114	30.158	1.31E-10	1.84E-07
	GOTERM_CC	0042470	melanosome	16	4.2328	9.18E-10	1.29E-06
	GOTERM_MF	0005515	protein binding	226	59.788	2.26E-08	3.34E-05
	GOTERM_CC	0005913	cell-cell adherent junction	23	6.0847	3.35E-07	4.72E-04
	GOTERM_CC	0005925	focal adhesion	25	6.6138	6.46E-07	9.08E-04
	GOTERM_MFT	0098641	cadherin binding involved in cell-cell adhesion	21	5.5556	1.62E-06	0.002393
	GOTERM_BP	0030198	extracellular matrix organization	17	4.4974	2.07E-06	0.003603
	GOTERM_CC	0043209	myelin sheath	14	3.7037	7.93E-06	0.011153

**Table 4: GO analysis of the 21 common DEGs**

Category	GO-ID	Description	Count	%	p-value	FDR
GOTERM_MF_DIRECT	0005515	protein binding	15	78.95	0.0034	3.3369
GOTERM_BP_DIRECT	0060716	labyrinthine layer blood vessel development	2	10.53	0.0191	20.1842
GOTERM_CC_DIRECT	0005622	intracellular	5	26.32	0.0314	28.6846

**Table 5: KEGG pathways identified in two datasets**

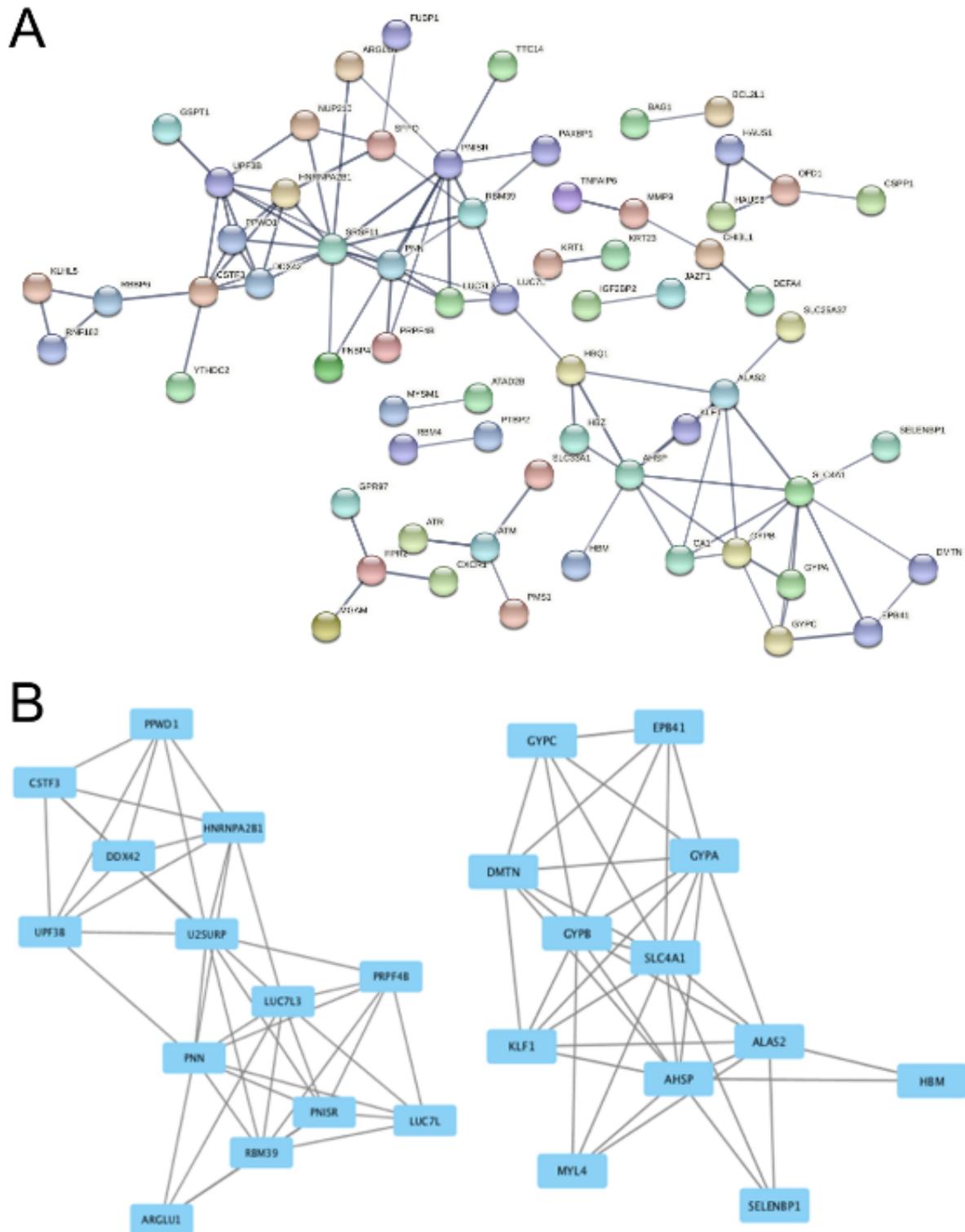
	Terms	Count	%	p-value	FDR
GSE71226	hsa03015: mRNA surveillance pathway	21	1.1071	4.24E-05	0.0558
	hsa03008: Ribosome biogenesis in eukaryotes	16	0.8434	5.01E-03	6.3225
	hsa04922: Glucagon signaling pathway	17	0.8962	7.20E-03	9.0472
	hsa00514: Other types of O-glycan biosynthesis	7	0.3690	7.60E-03	9.5031
GSE31821	hsa04510: Focal adhesion	20	5.2910	7.58E-07	9.6949
	hsa04010: MAPK signaling pathway	19	5.0265	5.55E-05	0.0709
	hsa05146: Amoebiasis	12	3.1746	6.15E-05	0.0787
	hsa05200: Pathways in cancer	25	6.6138	3.82E-05	0.0489
	hsa04151: PI3K-Akt signaling pathway	22	5.8201	1.27E-04	0.1628
	hsa04310: Wnt signaling pathway	13	3.4392	1.61E-04	0.2052
	hsa04512: ECM-receptor interaction	10	2.6455	2.98E-04	0.3803
	hsa04611: Platelet activation	12	3.1746	3.83E-04	0.4889
	hsa05145: Toxoplasmosis	11	2.9100	3.99E-04	0.5086
	hsa05205: Proteoglycans in cancer	15	3.9683	4.53E-04	0.5786
	hsa04921: Oxytocin signaling pathway	12	3.1746	1.27E-03	1.6187
	hsa04974: Protein digestion & absorption	9	2.3810	1.52E-03	1.9285
	hsa05140: Leishmaniasis	8	2.1164	1.87E-03	2.3664
	hsa04810: Regulation of actin cytoskeleton	14	3.7037	2.23E-03	2.8134
	hsa04670: Leukocyte transendothelial migration	10	2.6455	2.25E-03	2.8373
	hsa04062: Chemokine signaling pathway	13	3.4392	2.29E-03	2.8997
	sa04612: Antigen processing & presentation	8	2.1164	2.77E-03	3.4863
	hsa04915: Estrogen signaling pathway	9	2.3810	3.21E-03	4.0328
	hsa04014: Ras signaling pathway	14	3.7037	4.21E-03	5.2515
	hsa04145: Phagosome	11	2.9101	4.22E-03	5.2649
hsa05166: HTLV-I infection	15	3.9683	4.41E-03	5.4975	

“may be” an interactive linkage between CAD and AF that involves membrane metallo-endopeptidase (MME, also known as Neprilysin or Neutral endopeptidase 24.11) (Sankhe et al., 2020), and Tfr1.

#### ***Protein-protein interaction network and molecular analysis***

PPIs, either via strong or weak physical or functional interactions, play fundamental

roles in cellular functions and biological processes of all organisms under normal condition and disease development (Rao et al., 2014). In this respect, the present study used STRING to mine all proteins coded by the DEGs for potential interactions within and between the datasets from the patients with CAD and/or AF.

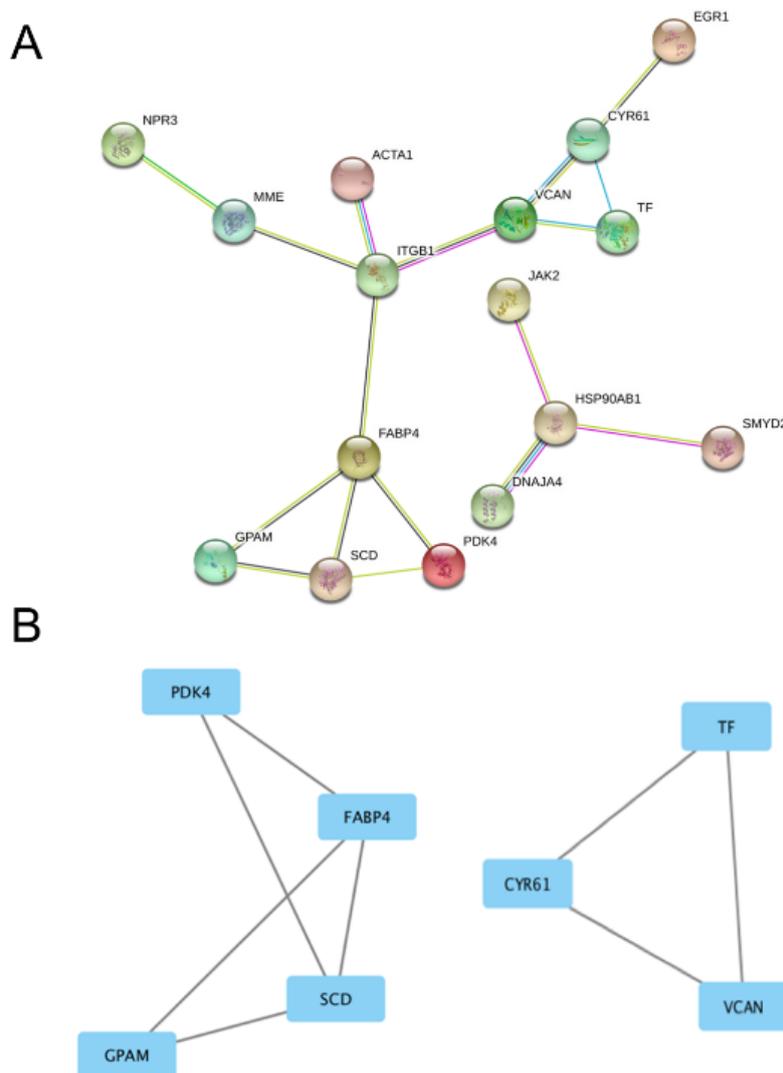


**Figure 4: Protein-protein-interaction analysis of the DEGs in GSE71226 dataset. Panel A** shows the overall network in GSE71226 dataset. **Panel B** shows the derived modules from the network. The rectangles stand for DEGs and the lines stand for their interactions. U2SURP, LUC7L and DDX42 in Module A (on the left side) and GYPC, EPB41 and ALAS2 in Module B (on the right side) are the most important nodes. See Supplementary Table 1 for all abbreviations.

Specifically, from GSE71226 dataset of the patients with CAD and healthy subjects, an intricate PPI network was recognized by STRING analysis. Since the network is so complex as shown in Figure 4A, it is unlikely to decode network(s) of interest; thus, we screened 2 functional modules with the help of Cytoscape software in the network. The results showed that Module A (Figure 4B, the left side) and Module B (Figure 4B, the right side) contain 13 and 12 nodes, respectively. Among them, U2 snRNP-associated SURP motif-containing protein (U2SURP, a RNA binding protein (De Maio et al., 2018)), Luc7-like protein 3 (LUC7L3, a DNA/RNA binding protein (Tufarelli et al., 2001)), and Pinin (PNN, a DNA/RNA binding protein (Hsu et al., 2020)) are the most important nodes in

Module A; while in Module B, it was found that Glycophorin-C (GYPC, an erythrocyte regulatory protein (Jaskiewicz et al., 2018)), Protein 4.1 (EBP41 or Beatty's protein, an erythrocyte structural and regulatory protein (Kiyomitsu and Cheeseman, 2013)), and Alpha-hemoglobin-stabilizing protein (ALAS2, a hemoglobin regulatory protein (Che Yaacob et al., 2020)) are the most important nodes.

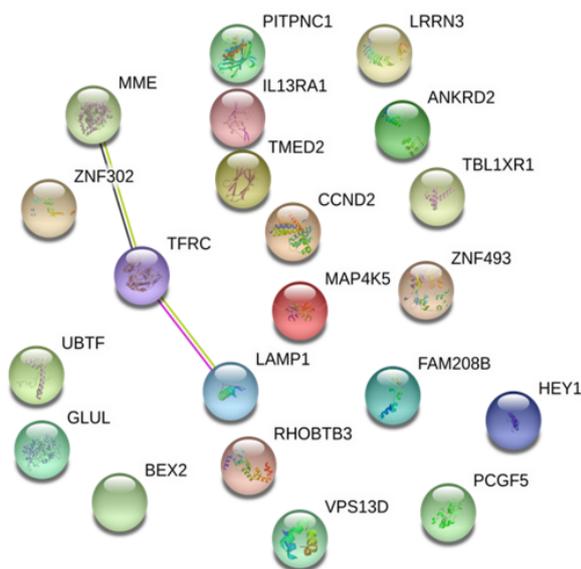
On other hand, when the same approach was used in GSE31821 dataset of the patients with AF and healthy subjects, a relatively simple PPI network (containing 2 clusters with 9 nodes) was identified by STRING analysis (Figure 5). Further, Cytoscape analysis found that Module A (Figure 5B, the left side) and Module B (Figure 5B, the right side)



**Figure 5: Protein-protein-interaction analysis of the DEGs in GSE31821 dataset. Panel A** shows the overall network in GSE31821 dataset. **Panel B** shows the derived modules from the network. The rectangles stand for DEGs and the lines stand for their interactions. PDK4, FABP4, SCD and GPAM in Module A (on the left side) and CYR61, TF and VCAN in Module B (on the right side) are the most important nodes. See Supplementary Table 1 for all abbreviations.

have 4 and 3 nodes, respectively. Among these, Acyl-CoA desaturase (SCD, an enzyme involving biosynthesis of monounsaturated fatty acids (Vanhercke et al., 2011)), Fatty acid-binding protein (FABP4, a fatty acid transport protein (Rezar et al., 2020)), and Glycerol-3-phosphate acyltransferase 1 (GPAM, an enzyme involving glycerolipids biosynthesis (Mitka et al., 2019)) were the mostly critical nodes in Module A; while Serotransferrin (TF, an iron transport protein (Jamnongkan et al., 2019)), Protein CYR61 (CYR61, a cellular growth regulatory protein (Huang et al., 2017)), and Versican core protein (VCAN, an extracellular matrix proteoglycan (Gardela et al., 2020)) are the most critical nodes in Module B.

Interestingly, when the common DEGs from two datasets were analyzed using the same approach, it was found that only three proteins, MME, Transferrin receptor protein 1 (TfR1), and Lysosome-associated membrane glycoprotein 1 (LAMP1, an integral membrane protein with unknown function (Kirschner et al., 2016)), interacted each other while the remaining 18 proteins had no significant influencing characteristics (Figure 6).



**Figure 6: Protein-protein-interaction analysis of the 21 common DEGs in both datasets.** Among 21 common DEGs, MME, TFRC and LAMP1 are the most significant nodes from the PPI network.

Taken together, these data suggested that although the PPI within and between two datasets are complex and most (if not all) functional interactions remain largely unknown, three proteins (MME, TfR1, LAMP1) may be concurrently involved in the development of CAD and AF.

## DISCUSSION

In the present study, we investigated the common DEGs and molecular networks of two datasets consisting of the healthy subjects and the patients with CAD or AF using various bioinformatic tools. Overall, 565 up-regulated and 1367 down-regulated genes were discovered in the dataset from the patients with CAD, while 293 up-regulated and 68 down-regulated genes were revealed in the dataset from the patients with AF. From these genes, 21 common DEGs were highly enriched in the intracellular cytoplasm, protein binding, and labyrinthine layer of vessel in both CAD and AF patients. These common DEGs are involved in 4 pathways in the CAD dataset and 21 pathways in the AF dataset. Further analysis of those pathways identified three important proteins (MME, TfR1, LAMP1) that highly co-expressed in the CAD and AF patients. To the best of our knowledge, this is the first study to investigate the cross-correlation of all DEGs between the CAD and AF datasets. The findings may facilitate a better understanding of the mechanisms underlying the pathogenesis of CAD and AF.

The close relationship between CAD and AF has been well recognized in literature, including the fact such as patients with AF develop a high prevalence of CAD (Michniewicz et al., 2018). It is known that genetic factors contribute importantly to both CAD and AF and studying the genetic basis of cardiovascular disease has made significant contribution to understand disease biology and promote cardiovascular therapy (Yla-Herttuala and Baker, 2017). Numerous studies have identified a number of key genes and critical modules that are associated with CAD

and AF by analyzing the microarrays data using bioinformatic tool and platform (Wang et al., 2016; Zhang et al., 2014). However, the genomic correlations between two diseases have not been fully investigated.

Our studies found that there are significant number of up- and down-regulated genes in each disease, but these genes may not be directly correlated each other within two diseases. However, among those were 21 common DEGs identified from two datasets, including 14 up- and 7 down-regulated genes that could be involved in the pathogenesis of CAD and AF. At the protein levels, three major candidates of MME, TfR1, and LAMP1 that were encoded by the corresponding genes in those 21 common DEGs, were revealed by PPI network analysis, suggesting that these proteins may play a critical role in the development of two diseases.

MME is a 100 kD type II transmembrane glycoprotein and plays an important role by enzymatically modulating the metabolism of glucagon, enkephalins, substance P, neurotensin, oxytocin, bradykinin, and atrial natriuretic peptides (ANP) (Roques, 1998). Among these, ANP is a key peptide synthesized by the heart and contributes critical regulatory roles in normal cardiovascular homeostasis and cardiovascular disease (Munagala et al., 2004). It was reported that MME is up-regulated in the heart of patients with heart failure and in the neutrophils of patients with early phase of acute myocardial infarction (Fielitz et al., 2002; Knecht et al., 2002). MME also controls local antifibrotic peptide bradykinin through the degradation of bradykinin in the extracellular space of heart tissue (Fielitz et al., 2002). Our results indicated that MME is one of the common genes concurrently expressed in both CAD and AF, suggesting that MME could become a therapy target for the AF patients with CAD.

TRFC gene encodes a cell surface receptor, termed transferring receptor 1 (TfR1), necessary for cellular iron uptake via the receptor-mediated endocytosis and it is essential for the function of red blood cell and development of the nervous system (Levy et al.,

1999). Both iron overload and iron deficiency, which are directly controlled by transferring receptor, were found to cause cardiomyopathy and heart failure (Anand and Gupta, 2018; Kremastinos and Farmakis, 2011). The present finding of highly expressed TfR1 in both CAD and AF provides additional evidence regarding the potential role of TfR1 in the pathogenesis of cardiovascular diseases.

LAMP1 is a member of membrane glycoprotein family, and LAMP1/2 are major components of lysosomal membrane (Eskelinen, 2006). Studies demonstrated LAMP1 is involved in autophagy process via mediating fusion between autophagosome and lysosomes; but the detailed mechanism is not fully understood. It was reported that the excessive autophagy by intracellular stress devoted significantly negative impacts on the developments of various cardiovascular diseases, including CAD and heart failure (Martinet et al., 2007). Our study found a remarkable up-regulation of LAMP1, supporting the possible involvement of LAMP1 in both CAD and AF. Surprisingly, mice with LAMP1 deficiency manifest normal lysosomal morphology and function (Andrejewski et al., 1999). The discrepancy could be due to different species or experimental condition that need further investigation.

The BP, MF, and CC are three terms commonly used in the GO enrichment analysis to reveal the involvement of genes of interest at different biological levels (Walter et al., 2015). The present study found that the common DEGs for CAD and AF that were enriched mostly in the intracellular cytosol appear to be involved in the development of the labyrinthine layer during embryonic vessel development. Rinckenberger and Werb (2000) demonstrated that the CC is involved in the labyrinthine layer of the placenta blood vessel progression and connected with cardiovascular system development; but future investigation is needed to address the biological role of the vascular labyrinthine layer in cardiovascular abnormalities, such as CAD and AF.

Signaling pathway (or signaling cascade or biochemical cascade) is a series of cellular and molecular reactions that always take place in cells under normal and diseases conditions, including the development of CAD and AF. By running KEGG pathway analysis, Poci (2019) found that mRNA surveillance pathway, ribosome biogenesis, and glucagon signaling pathway were the major pathways that affect CAD. In the case of glucagon pathway, Ali et al. (2015) demonstrated that glucagon administration impairs survival following ischemia in non-diabetic mouse and promote cardiomyocytes apoptosis. The present study identified 4 signaling pathways that were likely associated with CAD, supporting the above findings. By contrast, 21 pathways, such as Focal adhesion and MAPK pathways (see Table 5), were generated out of the AF dataset using the same approach. Among these, MAPK pathway is probably the most important signaling pathway related to the pathogenesis of cardiovascular disease, including AF (Zhang et al., 2003). A study found that the MAPK pathway is involved in occurrence of AF in patients with rheumatic heart disease after cardiac surgery through promoting atrial fibrosis (Zhang et al., 2017), which is consistent with our present finding on MAPK pathway.

In conclusion, the present study identified 21 common DEGs out of thousands of genes in the two datasets collected from the patients with CAD or AF. These common DEGs were highly enriched in the intracellular cytoplasm, protein binding, and vascular labyrinthine layer in patients. Three important protein candidates (MME, TfR1, and LAMP1) may play crucial roles in the disease development of both CAD and AF. We realized that the study comes with limitations. The subjects between the CAD and the AF have different ethnic background and medical history. The sample source and size should also be improved. Future studies using animal models with CAD and AF should be conducted to validate the hypothesis.

### **Acknowledgments**

None.

### **Conflict of interest**

The authors declare no conflict of interest.

### **Funding**

This work was supported by the NIH grant (1R15HL140528-01 for JQH), One-Health seed grant (PJ6SPVHJ for JQH) by the College of Veterinary Medicine at Virginia Tech and the Edward Via College of Osteopathic Medicine, Interdisciplinary Graduate Education Programs of Regenerative Medicine (IGEP-RM, for YJZ), and IRC Seed Grant (#178391 for JQH) by the College of Veterinary Medicine at Virginia Tech. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### **Supplementary material (online only)**

The source datasets are available at <https://www.ncbi.nlm.nih.gov/geo/>.

Supplementary Table 1 - All abbreviations used in the this paper;

Supplementary Table 2 - All DEGs in GSE71226;

Supplementary Table 3 - All DEGs in GSE31821.

## **REFERENCES**

- Aibar S, Fontanillo C, Droste C, De Las Rivas J. Functional gene networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering. *Bioinformatics*. 2015;31:1686-88. doi: 10.1093/bioinformatics/btu864.
- Ali S, Ussher JR, Baggio LL, Kabir MG, Charron MJ, Ilkayeva O, et al. Cardiomyocyte glucagon receptor signaling modulates outcomes in mice with experimental myocardial infarction. *Mol Metab*. 2015;4:132-43. doi: 10.1016/j.molmet.2014.11.005.
- Anand IS, Gupta P. Anemia and iron deficiency in heart failure: Current concepts and emerging therapies. *Circulation*. 2018;138:80-98. doi: 10.1161/CIRCULATIONAHA.118.030099.

- Andrejewski N, Punnonen EL, Guhde G, Tanaka Y, Lüllmann-Rauch R, Hartmann D, et al. Normal lysosomal morphology and function in LAMP-1-deficient mice. *J Biol Chem.* 1999;274:12692-701. doi: 10.1074/jbc.274.18.12692.
- Che Yaacob NS, Islam MA, Alsaleh H, Ibrahim IK, Hassan R. Alpha-hemoglobin-stabilizing protein (AHSP): A modulatory factor in beta-thalassemia. *Int J Hematol.* 2020;111:352-9. doi: 10.1007/s12185-019-02806-8.
- Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23:1846-7. doi: 10.1093/bioinformatics/btm254.
- De Maio A, Yalamanchili HK, Adamski CJ, Gennarino VA, Liu Z, Qin J, et al. RBM17 Interacts with U2SURP and CHERP to regulate expression and splicing of RNA-processing proteins. *Cell Rep.* 2018;25:726-36 e7. doi: 10.1016/j.celrep.2018.09.041.
- Eskelinen EL. Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. *Mol Aspects Med.* 2006;27:495-502. doi: 10.1016/j.mam.2006.08.005.
- Fielitz J, Dendorfer A, Pregla R, Ehler E, Zurbrügg H, Bartunek J, et al. Neutral endopeptidase is activated in cardiomyocytes in human aortic valve stenosis and heart failure. *Circulation.* 2002;105:286-9. doi: 10.1161/hc0302.103593.
- Gardela J, Jauregi-Miguel A, Martinez CA, Rodriguez-Martinez H, Lopez-Bejar M, Alvarez-Rodriguez M. Semen modulates the expression of NGF, ABHD2, VCAN, and CTEN in the reproductive tract of female rabbits. *Genes (Basel).* 2020;11:758-74. doi: 10.3390/genes11070758.
- Gladding PA, Legget M, Fatkin D, Larsen P, Doughty R. Polygenic risk scores in coronary artery disease and atrial fibrillation. *Heart Lung Circ.* 2020;29:634-40. doi: 10.1016/j.hlc.2019.12.004.
- Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: The AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA.* 2001;285:2370-5. doi: 10.1001/jama.285.18.2370.
- Hsu SY, Mukda S, Leu S. Expression and distribution pattern of pnn in ischemic cerebral cortex and cultured neural cells exposed to oxygen-glucose deprivation. *Brain Sci.* 2020;10:708-18. doi: 10.3390/brainsci10100708.
- Huang YT, Lan Q, Lorusso G, Duffey N, Ruegg C. The matricellular protein CYR61 promotes breast cancer lung metastasis by facilitating tumor cell extravasation and suppressing anoikis. *Oncotarget.* 2017;8:9200-15. doi: 10.18632/oncotarget.13677.
- Jamnonngkan W, Lebrilla CB, Barboza M, Techasen A, Loilome W, Sithithaworn P, et al. Discovery of sero-transferrin glycoforms: Novel markers for diagnosis of liver periductal fibrosis and prediction of cholangiocarcinoma. *Biomolecules.* 2019;9:538-50. doi: 10.3390/biom9100538.
- Jaskiewicz E, Peyrard T, Kaczmarek R, Zerka A, Jodłowska M, Czerwinski M. The Gerbich blood group system: Old knowledge, new importance. *Transfus Med Rev.* 2018;32:111-6. doi: 10.1016/j.tmr.2018.02.004.
- Kertai MD, Li YJ, Ji Y, Qi W, Lombard FW, Shah SH, et al. Genome-wide association study of new-onset atrial fibrillation after coronary artery bypass grafting surgery. *Am Heart J.* 2015;170:580-90.e28. doi: 10.1016/j.ahj.2015.06.009.
- Kirschner A, Thiede M, Blaschke F, Richter GH, Gerke JS, Baldauf MC, et al. Lysosome-associated membrane glycoprotein 1 predicts fratricide amongst T cell receptor transgenic CD8<sup>+</sup> T cells directed against tumor-associated antigens. *Oncotarget.* 2016;7:56584-97. doi: 10.18632/oncotarget.10647.
- Kiyomitsu T, Cheeseman IM. Cortical dynein and asymmetric membrane elongation coordinately position the spindle in anaphase. *Cell.* 2013;154:391-402. doi: 10.1016/j.cell.2013.06.010.
- Knecht M, Pagel I, Langenickel T, Philipp S, Scheuermann-Freestone M, Willnow T, et al. Increased expression of renal neutral endopeptidase in severe heart failure. *Life Sci.* 2002;71:2701-12. doi: 10.1016/s0024-3205(02)01990-2.
- Kremastinos DT, Farmakis D. Iron overload cardiomyopathy in clinical practice. *Circulation.* 2011;124:2253-63. doi: 10.1161/CIRCULATIONAHA.111.050773.
- Kristensen KE, Knage CC, Nyhegn LH, Mulder BA, Rienstra M, Van Gelder IC, et al. Subclinical atherosclerosis is associated with incident atrial fibrillation: a systematic review and meta-analysis. *Europace.* 2020;22:991-1000. doi: 10.1093/europace/eaab030.
- Kumar D, Dash D. Proteogenomic tools and approaches to explore protein coding landscapes of eukaryotic genomes. *Adv Exp Med Biol.* 2016;926:1-10. doi: 10.1007/978-3-319-42316-6\_1.

- Levy JE, Jin O, Fujiwara Y, Kuo F, Andrews NC. Transferrin receptor is necessary for development of erythrocytes and the nervous system. *Nat Genet.* 1999;21:396-9. doi: 10.1161/ATVBAHA.108.179481.
- Lieder H, Breithardt G, Heusch G. Fatal attraction - A brief pathophysiology of the interaction between atrial fibrillation and myocardial ischemia. *Int J Cardiol.* 2018;254:132-5. doi: 10.1016/j.ijcard.2017.11.119.
- Martinet W, Knaapen MW, Kockx MM, De Meyer GR. Autophagy in cardiovascular disease. *Trends Mol Med.* 2007;13:482-91. doi: 10.1016/j.molmed.2007.08.004.
- Mehta RH, Dabbous OH, Granger CB, Kuznetsova P, Kline-Rogers EM, Anderson FA Jr, et al. Comparison of outcomes of patients with acute coronary syndromes with and without atrial fibrillation. *Am J Cardiol.* 2003;92:1031-6. doi: 10.1016/j.amjcard.2003.06.001.
- Michniewicz E, Mlodawska E, Lopatowska P, Tomaszuk-Kazberuk A, Malyszko J. Patients with atrial fibrillation and coronary artery disease - Double trouble. *Adv Med Sci.* 2018;63:30-5. doi: 10.1016/j.advms.2017.06.005.
- Mitka I, Ropka-Molik K, Tyra M. Functional analysis of genes involved in glycerolipids biosynthesis (GPAT1 and GPAT2) in pigs. *Animals (Basel).* 2019;9:308-21. doi: 10.3390/ani9060308.
- Motloch LJ, Reda S, Larbig R, Wolff A, Motloch KA, Wernly B, et al. Characteristics of coronary artery disease among patients with atrial fibrillation compared to patients with sinus rhythm. *Hellenic J Cardiol.* 2017;58:204-12. doi: 10.1016/j.hjc.2017.03.001.
- Munagala VK, Burnett JC Jr, Redfield MM. The natriuretic peptides in cardiovascular medicine. *Curr Probl Cardiol.* 2004;29:707-69. doi: 10.1016/j.cpcardiol.2004.07.002.
- Murakami N, Tanno M, Kokubu N, Nishida J, Nagano N, Ohnishi H, et al. Distinct risk factors of atrial fibrillation in patients with and without coronary artery disease: a cross-sectional analysis of the BOREAS-CAG Registry data. *Open Heart.* 2017;4:e000573. doi: 10.1136/openhrt-2016-000573.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 1999;27:29-34. doi: 10.1093/nar/27.1.29.
- Pilgrim T, Kalesan B, Zanchin T, Pulver C, Jung S, Mattle H, et al. Impact of atrial fibrillation on clinical outcomes among patients with coronary artery disease undergoing revascularisation with drug-eluting stents. *EuroIntervention.* 2013;8:1061-71. doi: 10.4244/EIJV8I9A163.
- Pocai A. Modulation of glucagon signaling: A metabolic approach for heart failure? *JACC Basic Transl Sci.* 2019;4:173-5. doi: 10.1016/j.jacbts.2019.03.006.
- Rao VS, Srinivas K, Sujini GN, Kumar GN. Protein-protein interaction detection: Methods and analysis. *Int J Proteomics.* 2014;2014:147648-60. doi: 10.1155/2014/147648.
- Rezar R, Jirak P, Gschwandtner M, Derler R, Felder TK, Haslinger M, et al. Heart-type fatty acid-binding protein (H-FABP) and its role as a biomarker in heart failure: What do we know so far? *J Clin Med.* 2020;9:164-78. doi: 10.3390/jcm9010164.
- Rinkenberger J, Werb Z. The labyrinthine placenta. *Nat Genet.* 2000;25:248-50. doi: 10.1038/76985.
- Roques BP. Cell surface metalloproteinases involved in blood pressure regulation: Structure, inhibition and clinical perspectives. *Pathol Biol (Paris).* 1998;46:191-200.
- Ruddox V, Sandven I, Munkhaugen J, Skattebu J, Edvardsen T, Otterstad JE. Atrial fibrillation and the risk for myocardial infarction, all-cause mortality and heart failure: A systematic review and meta-analysis. *Eur J Prev Cardiol.* 2017;24:1555-66. doi: 10.1177/2047487317715769.
- Sankhe R, Pai SRK, Kishore A. Tumour suppression through modulation of neprilysin signaling: A comprehensive review. *Eur J Pharmacol.* 2020;891:173727-42. doi: 10.1016/j.ejphar.2020.173727.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447-52. doi: 10.1093/nar/gku1003.
- Tufarelli C, Frischauf AM, Hardison R, Flint J, Higgs DR. Characterization of a widely expressed gene (LUC7-LIKE; LUC7L) defining the centromeric boundary of the human alpha-globin domain. *Genomics.* 2001;71:307-14. doi: 10.1006/geno.2000.6394.
- Vanhercke T, Shrestha P, Green AG, Singh SP. Mechanistic and structural insights into the regioselectivity of an acyl-CoA fatty acid desaturase via directed molecular evolution. *J Biol Chem.* 2011;286:12860-9. doi: 10.1074/jbc.M110.191098.
- Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics-2020 update: A report from the American Heart Association. *Circulation.* 2020;141:e139-96. doi: 10.1161/CIR.0000000000000757.

Walter W, Sanchez-Cabo F, Ricote M. GOpilot: An R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015;31:2912-4. doi: 10.1093/bioinformatics/btv300.

Wang T, Wang B. Identification of microRNA-mRNA interactions in atrial fibrillation using microarray expression profiles and bioinformatics analysis. *Mol Med Rep*. 2016;13:4535-40. doi: 10.3892/mmr.2016.5106.

Yla-Herttuala S, Baker AH. Cardiovascular gene therapy: Past, present, and future. *Mol Ther*. 2017;25:1095-1106. doi: 10.1016/j.ymthe.2017.03.027.

Zhang D, Chen X, Wang Q, Wu S, Zheng Y, Liu X. Role of the MAPKs/TGF-beta1/TRAF6 signaling pathway in postoperative atrial fibrillation. *PLoS One*. 2017;12:e0173759. doi: 10.1371/journal.pone.0173759.

Zhang W, Elimban V, Nijjar MS, Gupta SK, Dhalla NS. Role of mitogen-activated protein kinase in cardiac hypertrophy and heart failure. *Exp Clin Cardiol*. 2003;8:173-83.

Zhang X, Cheng X, Liu H, Zheng C, Rao K, Fang Y, et al. Identification of key genes and crucial modules associated with coronary artery disease by bioinformatics analysis. *Int J Mol Med*. 2014;34:863-9. doi: 10.3892/ijmm.2014.1817.

Zheng Q, Wang XJ. GOEAST: A web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res*. 2008;36:W358-63. doi: 10.1093/nar/gkn276.