



Identifying RBM47, HCK, CD53, TYROBP, and HAVCR2 as Hub Genes in Advanced Atherosclerotic Plaques by Network-Based Analysis and Validation

Chiyu Liu^{1,2,3†}, Haifeng Zhang^{1,2†}, Yangxin Chen^{1,2†}, Shaohua Wang^{1,2}, Zhiteng Chen^{1,2}, Zhaoyu Liu^{3*} and Jingfeng Wang^{1,2*}

¹ Department of Cardiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China, ² Laboratory of Cardiac Electrophysiology and Arrhythmia in Guangdong Province, Guangzhou, China, ³ Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China

OPEN ACCESS

Edited by:

Lixin Cheng,
Jinan University, China

Reviewed by:

Vishal Acharya,
Institute of Himalayan Bioresource
Technology (CSIR), India
Hao Zhang,
Jilin University, China

*Correspondence:

Zhaoyu Liu
liuzhy98@mail.sysu.edu.cn
Jingfeng Wang
dr_wjf@vip.163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 04 September 2020

Accepted: 15 December 2020

Published: 15 January 2021

Citation:

Liu C, Zhang H, Chen Y, Wang S,
Chen Z, Liu Z and Wang J (2021)
Identifying RBM47, HCK, CD53,
TYROBP, and HAVCR2 as Hub Genes
in Advanced Atherosclerotic Plaques
by Network-Based Analysis and
Validation. *Front. Genet.* 11:602908.
doi: 10.3389/fgene.2020.602908

Background: Atherosclerotic cardiovascular diseases accounted for a quarter of global deaths. Most of these fatal diseases like coronary atherosclerotic disease (CAD) and stroke occur in the advanced stage of atherosclerosis, during which candidate therapeutic targets have not been fully established. This study aims to identify hub genes and possible regulatory targets involved in treatment of advanced atherosclerotic plaques.

Material/Methods: Microarray dataset GSE43292 and GSE28829, both containing advanced atherosclerotic plaques group and early lesions group, were obtained from the Gene Expression Omnibus database. Weighted gene co-expression network analysis (WGCNA) was conducted to identify advanced plaque-related modules. Module conservation analysis was applied to assess the similarity of advanced plaque-related modules between GSE43292 and GSE28829. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of these modules were performed by Metascape. Differentially expressed genes (DEGs) were mapped into advanced plaque-related modules and module membership values of DEGs in each module were calculated to identify hub genes. Hub genes were further validated for expression in atherosclerotic samples, for distinguishing capacity of CAD and for potential functions in advanced atherosclerosis.

Results: The lightgreen module (MElightgreen) in GSE43292 and the brown module (MEbrown) in GSE28829 were identified as advanced plaque-related modules. Conservation analysis of these two modules showed high similarity. GO and KEGG enrichment analysis revealed that genes in both MElightgreen and MEbrown were enriched in immune cell activation, secretory granules, cytokine activity, and immunoinflammatory signaling. RBM47, HCK, CD53, TYROBP, and HAVCR2 were identified as common hub genes, which were validated to be upregulated in advanced atherosclerotic plaques, to well distinguish CAD patients from non-CAD people and to

regulate immune cell function-related mechanisms in advanced atherosclerosis.

Conclusions: We have identified RBM47, HCK, CD53, TYROBP, and HAVCR2 as immune-responsive hub genes related to advanced plaques, which may provide potential intervention targets to treat advanced atherosclerotic plaques.

Keywords: atherosclerosis, bioinformatics, weighted gene co-expression network analysis, differentially expressed genes, gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes

INTRODUCTION

Atherosclerosis (AS) is the most common underlying cause of cardiovascular diseases, and accounted for a quarter of global deaths (Lee et al., 2019). Clinically the most significant and dangerous process in atherosclerotic cardiovascular disease like coronary atherosclerotic disease (CAD) and stroke is the growth and development of advanced atherosclerotic lesions, which causes either arterial lumen stenosis and blood flow obstruction or plaque rupture to enhance atherosclerotic thromboembolism (Fuster et al., 2005; Kovanen, 2019). However, no effective drugs have yet been available to reverse advanced plaques formation (Gaurav et al., 2015).

High throughput sequencing technology for detecting gene expression is an effective tool to reveal the underlying genes and biological processes during atherosclerotic plaque formation (Tan et al., 2017). Recently, several atherosclerotic gene expression profiling studies have been performed. Chen et al. (2020) identified THRAP3 and RBM43 as potential diagnostic and prognostic targets for ischemic event occurrence in carotid atherosclerosis. Jiao et al. (2020) identified CD40, F11R, TNRC18, and CAMK2G as surrogate diagnostic biomarkers for CAD in non-diabetic patients. Zhang et al. (2019) have identified TNPO1, RAP1B, ZDHHC17, and PPM1B as targets of CAD-related miRNAs. These studies were based on analysis of AS-related expression profile data in human peripheral blood, which were inclined to screen biomarkers for diagnosis and prognosis. However, gene expression profile analysis in diseased tissue is more helpful to reveal the mechanism of disease progression. Presently, there are only a few tissue-based studies of advanced plaque-related expression profile analysis. These studies focused on protein-protein interaction network (PPI) or transcription factor co-regulation network, both of which were based on previous experimental evidence (prior knowledge) (Lin et al., 2014; Wang et al., 2014; Tan et al., 2017; Liu et al., 2018). However, these analysis neglects gene expression information and clinical traits of samples under specific disease condition (Feltrin et al., 2019). Weighted gene co-expression network analysis (WGCNA) is an advanced bioinformatics analysis method that relies exclusively on the expression data values, making it possible to find all possible gene expression correlation under specific disease condition (Zhang and Horvath, 2005; Feltrin et al., 2019). Besides, WGCNA makes full use of the phenotypic traits of samples to identify trait-related gene sets, which may further improve the biological and clinical significance beyond classical methods (Miao et al., 2018; Shi et al., 2019).

In the present study, two human advanced atherosclerotic plaque-related expression profile datasets were obtained. Then advanced plaque-related co-expressed gene modules were identified, whose underlying biological processes were revealed. Finally, the hub genes in advanced plaques were identified and validated for expression in atherosclerotic samples, for distinguishing capacity of coronary atherosclerotic disease and for potential functions in advanced atherosclerosis, which provides new insight for AS progression and promising intervention targets.

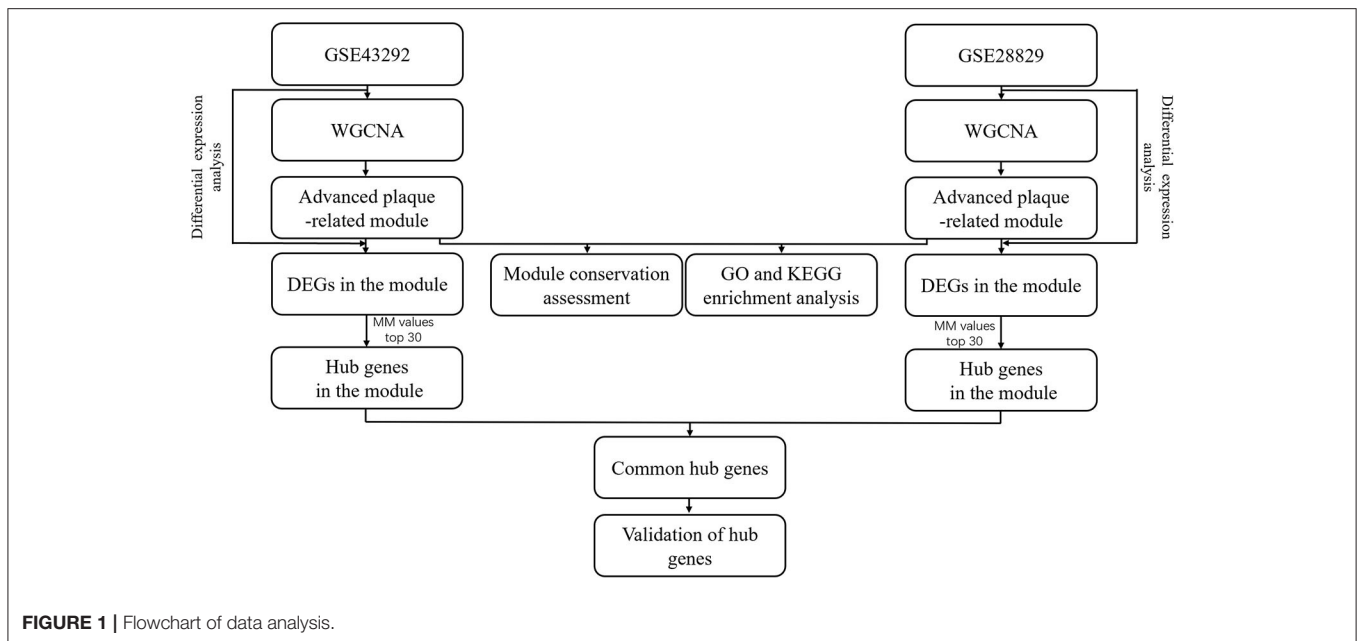
MATERIALS AND METHODS

Data Analysis Workflow

The flowchart of data analysis was presented in **Figure 1**. After advanced atherosclerosis-related datasets, GSE43292 and GSE28829, were obtained, WGCNA was applied to identify their own advanced plaque-related module. Module conservation analysis was applied to assess the similarity of advanced plaque-related modules between GSE43292 and GSE28829. Meanwhile, GO and KEGG enrichment analysis was applied to reveal the underlying biological functions of these modules. After differential expression analysis of each dataset, the identified DEGs were mapped into advanced plaque-related modules and module membership values of DEGs in each module were calculated to identify hub genes. Hub genes were further validated for expression level in advanced atherosclerotic tissues by PCR and differential expression analysis, for distinguishing capacity of coronary atherosclerotic disease by ROC analysis and for potential biological functions by Gene Set Enrichment Analysis.

Data Sources

The datasets of GSE43292, GSE28829, GSE18443, GSE31947, and GSE12288 were obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). In dataset GSE43292, samples collected from carotid endarterectomy in hypertensive patients were divided into early lesions group (stages I and II of the Stary classification) and advanced plaques group (stage IV and over of the Stary classification). This dataset was performed using the platform GPL6244 [Affymetrix Human Gene 1.0 ST Array, transcript (gene) version]. In dataset GSE28829, atherosclerotic carotid artery segments obtained during autopsy were also divided into early lesions group (stages I and II) and advanced plaques group (stage IV or above). This dataset was performed using the platform GPL570 [Affymetrix Human Genome U133 Plus 2.0 Array]. Both dataset GSE18443



and GSE31947 contained aortic samples from ApoE^{-/-} mice fed with high fat diet for 16 weeks (advanced plaques group) and wildtype mice fed with high fat diet for 16 weeks (control group). These samples were sequenced using the platform GPL3677 (Rosetta/Merck Mouse 44k 1.0 microarray). Dataset GSE12288 contained peripheral blood samples from patients with coronary atherosclerotic disease (CAD group, coronary artery stenosis >50% in at least one major coronary artery) and controls (Non-CAD group, no angiographically detectable coronary artery stenosis). This dataset was performed using the platform GPL96 [Affymetrix Human Genome U133A Array]. GSE18443, GSE31947, and GSE12288 were used as validation datasets.

Data Preprocessing

Data files stored in a raw format (.CEL files) were preprocessed using Robust Multichip Average algorithm (RMA) of oligo package within Bioconductor (<http://www.bioconductor.org/>) in R3.5.2 software (Carvalho and Irizarry, 2010; Wang et al., 2014). After background adjustment, normalization, and log transformation by RMA, the raw data was converted into a gene expression matrix. The “nsFilter” function contained in genefilter package was applied to filter out 50% probes based on IQR. Then the probe IDs were replaced by gene symbols. If more than one probe corresponded to one gene, the median expression value of all probes was analyzed.

Constructing Co-expressed Gene Modules and Identifying Advanced Plaque-Related Modules

The analysis processes of WGCNA included: (1) Samples were clustered by hierarchical clustering analysis and samples not conforming to the experimental grouping (outliers) were detected. (2) Appropriate soft threshold value (β) was chosen

to make the co-expression network approximate biologically significant scale-free topology (scale independence >0.8). (3) The co-expression similarity between gene i and j was defined as adjacency value between genes: $a_{ij} = |S_{ij}|^\beta$ ($S_{ij} = |\text{cor}(i, j)|$). (4) The adjacency matrix was transformed into a topological overlap matrix (TOM), and modules were detected by hierarchical clustering analysis of the gene dendrogram. (5) The module eigengene (ME) value of each module was summarized by the first principal component of the module expression levels. The advanced plaque-related modules were identified by correlating ME values with the clinical traits of samples using hierarchical clustering analysis and spearman correlation analysis. Advanced plaque-related module was defined as the module showing the highest correlation coefficient with advanced plaque ($P < 0.05$) and the closest location with advanced plaque on the hierarchical cluster dendrogram.

Assessing the Similarity of Advanced Plaque-Related Modules Between Dataset GSE43292 and GSE28829

The module conservation analysis processes included: (1) Common soft threshold value (β) was chosen to make the co-expression network of both datasets approximate biologically significant scale-free topology. (2) Common soft threshold value was used to calculate adjacency values in the individual datasets. (3) The adjacency matrix in the individual datasets was transformed into topological overlap matrix (TOM). (4) The TOM in GSE28829 was scaled such that the 95th percentile equals the 95th percentile of the TOM in GSE43292. (5) The consensus TOM was calculated by taking the component-wise (“parallel”) minimum of the TOMs in individual datasets. (6) Based on consensus TOM, consensus modules were detected by hierarchical clustering analysis of the gene dendrogram. (7)

The overlapping gene counts of each pair of GSE43292 (or GSE28829)—consensus modules were calculated to relate the modules in GSE43292 (or GSE28829) to the consensus modules, and the Fisher's exact test was used to assign a p -value to each of the pairwise overlaps. (8) If a module in dataset GSE43292 had the same consensus counterpart with a module in dataset GSE28829, these two modules were considered to be similar.

Functional Enrichment Analysis of Genes in Advanced Plaque-Related Modules

To reveal the underlying biological functions of the advanced plaque-related modules, genes in them were introduced into the DAVID database (<https://david.ncicrf.gov/>) for Gene Ontology (GO) enrichment analysis and Metascape database (<https://metascape.org>) for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (Zhou et al., 2019). GO describes the biological processes (BP), subcellular localization (CC), and molecular function (MF) enriched in gene sets (Ashburner et al., 2000; Mi et al., 2017; Consortium, 2019). And KEGG describes the pathways enriched in gene sets (Kanehisa and Sato, 2020). Terms with a $P < 0.01$, a minimum overlap of 3 and a minimum enrichment of 1.5 were identified.

Identifying Hub Genes in Advanced Plaques

Differentially expressed genes (DEGs) of GSE43292 and GSE28829 were identified based on $|\log_2\text{fold change (FC)}| > 1$ and $P < 0.05$ by using Limma package (Smyth, 2005). DEGs were then intersected with genes of advanced plaque-related module in each dataset to obtain DEGs in advanced plaque-related modules. Module membership (MM) value of each gene was calculated by WGCNA and the genes with top 30 MM values were identified as hub genes in each module.

Aortic Tissue Collection of Animal Models

After being fed with high fat diet or normal chow for 16 weeks, ApoE^{-/-} mice were fasted for 14 h, and then anesthetized and euthanized. The aortic tissues were removed from the ascending aorta to the ileal bifurcation and snap frozen in liquid nitrogen for RNA analysis.

RNA Extraction and Quantitative Real-Time PCR

Total RNA from aortic tissues was extracted using TRIzol reagent (Invitrogen) and purified using isopropanol, 75% ethanol and RNase-free water following the introductions of manufacturer. cDNA was synthesized using the PrimeScriptTM RT Master Mix kit (Takara). Quantitative real-time PCR was performed using the TB Green[®] Premix Ex TaqTM kit (Takara) by Applied Biosystems Quanstudio DX (Thermo Fisher Scientific) after setting up the appropriate protocol. β -actin was used as an internal reference. The $2^{-\Delta\Delta Ct}$ method was applied to analyze the results of PCR. All the primers were designed online (<https://pga.mgh.harvard.edu/primerbank/>) and were listed in **Supplementary Table 1**.

Differential Expression Validation of Hub Genes in Dataset GSE18443, GSE31947, and GSE12288

Differential expression analysis of hub genes was performed by limma package in R3.5.2 software. $P < 0.05$ was used to screen differentially expressed hub genes.

Receiver Operating Characteristic (ROC) Curve Validation of Hub Genes in Dataset GSE12288

ROC was plotted and area under the curve (AUC) was calculated by *GraphPad Prism 8* software to evaluate the capability of hub genes to distinguish samples of CAD group from those of non-CAD group. Efficacy evaluation: AUC = 0.5, non-efficiency; AUC > 0.5 but < 0.7, modest-efficiency; AUC > 0.7, high-efficiency.

Gene Set Enrichment Analysis (GSEA) of Hub Genes

GSEA run in the GSEA software using dataset GSE28829. Based on the median expression of each hub gene, samples were divided into high expression group and low expression group. In run program settings, gene set "c2.cp.kegg.v6.2.symbols.gmt" was chosen as the reference gene set and the parameter "number of permutations" was set as 1,000. The top three gene sets with the lowest NOM ($p < 0.01$) were considered significantly enriched.

Constructing Transcriptional Regulatory Network and Identifying Key Transcription Factors in Advanced Plaque-Related Modules

TRRUST v2 (www.grnpedia.org/trrust) is a database of reference TF–target regulatory interactions in humans and mice based on literature curation, which applies a network-based algorithm to prioritize key TFs for the given transcriptionally responsive genes (Han et al., 2018). The DEGs in advanced plaque-related modules were introduced into TRRUST v2 database and $P < 0.001$ was used to screen significantly enriched transcriptional regulatory networks.

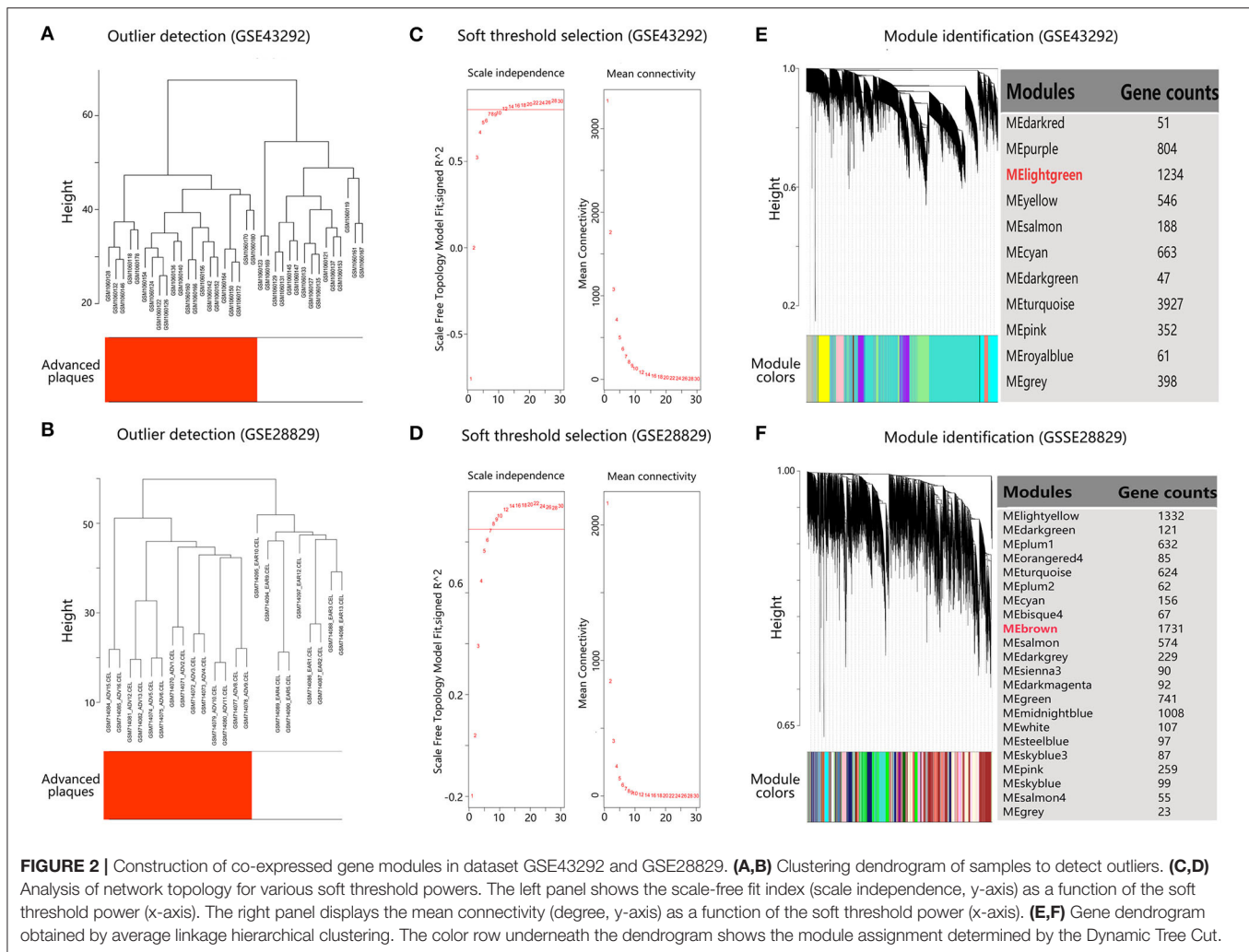
Constructing Protein-Protein Interaction (PPI) Network in Advanced Plaque-Related Modules

The DEGs in advanced plaque-related modules were introduced into STRING database (<https://string-db.org/>) to construct PPI network. The *Cytohubba* plugin of *Cytoscape* software was used to perform network analysis and identify top 10 genes with the highest degree.

RESULTS

Construction of Co-expressed Gene Modules

To identify co-expressed gene sets, WGCNA was applied. The microarray quality was first evaluated by sample clustering. As shown in **Figures 2A,B**, samples were clustered into two clusters



in each dataset, corresponding to the early lesions group and advanced plaques group, respectively. No outliers were detected in the clusters. Next, 12 and 8 were chosen as the soft threshold for GSE43292 and GSE28829, respectively, which ensured a scale-free network (**Figures 2C,D**). As a result, 11 co-expressed modules in GSE43292 were identified and 22 co-expressed modules in GSE28829 were identified (**Figures 2E,F**).

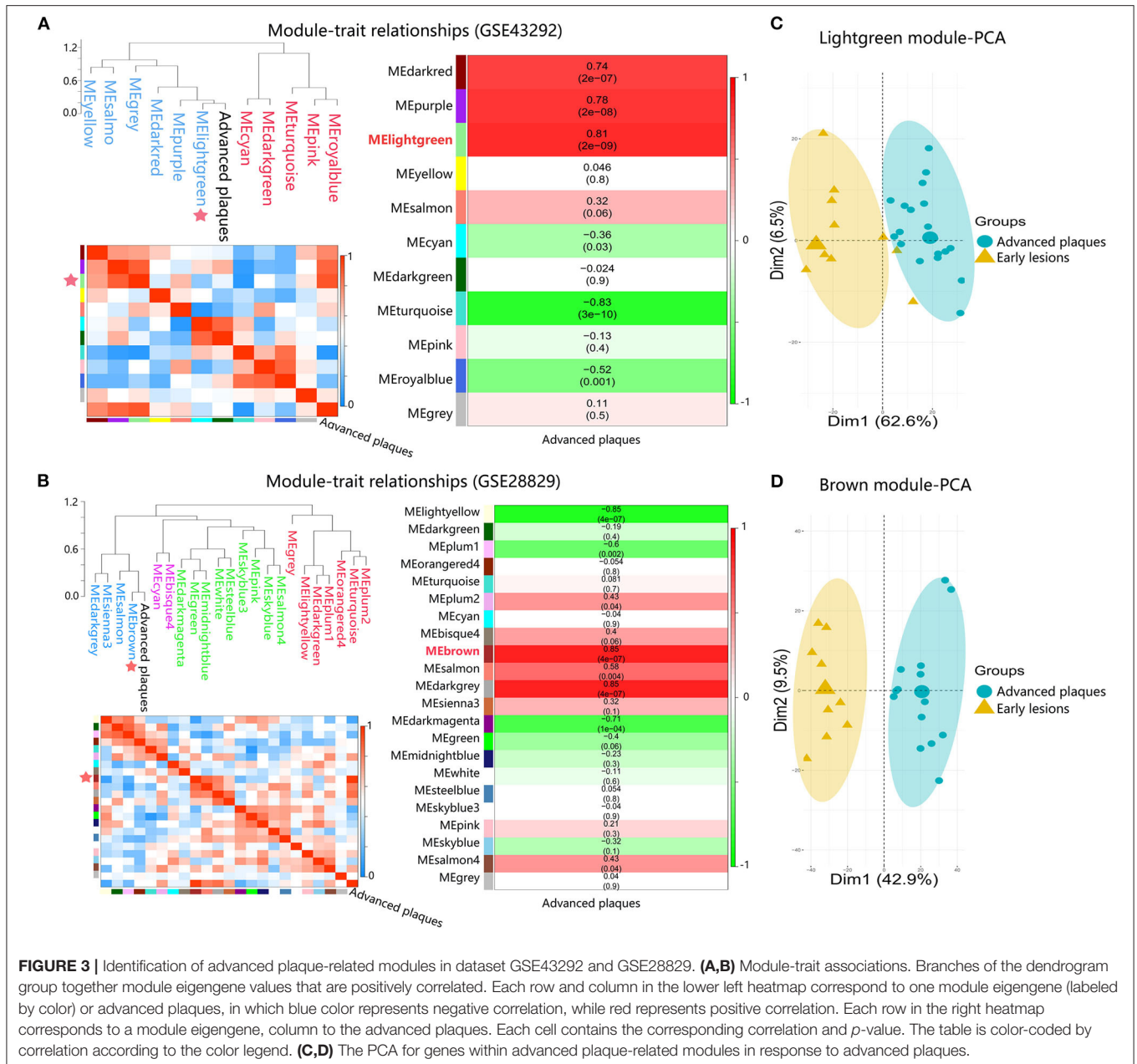
Identification of Advanced Plaque-Related Modules

To identify advanced plaque-related modules, module eigengene (ME) values of modules were correlated with the clinical traits of samples by applying average linkage hierarchical clustering algorithm and spearman correlation coefficient. As shown in **Figures 3A,B**, the lightgreen module (MElightgreen) in GSE43292 and the brown module (MEbrown) in GSE28829 showed the highest association with advanced plaques. Principal component analysis results showed that genes in MElightgreen and MEbrown could well distinguish samples of advanced plaques group from that of early lesions group, indicating that genes in these two modules

contributed to plaque progression from early to advanced stages (**Figures 3C,D**). Hence, MElightgreen and MEbrown were identified as advanced plaque-related modules for GSE43292 and GSE28829, respectively.

Module Conservation Analysis of MElightgreen and MEbrown

Next module conservation analysis was performed to validate the similarity between MElightgreen and MEbrown, which were derived from different datasets. As shown in **Supplementary Figure 1**, a power of 12 was chosen as common soft threshold value to ensure scale-free topology for each dataset. And after scaling topological overlap matrix (TOM) of the two datasets for mitigating the effect of different statistical properties, they were comparable across datasets (**Supplementary Figure 2**). Subsequently, consensus TOM was calculated by combining the TOMs of two datasets and 45 consensus modules were identified (**Figure 4A**). By further calculating the overlapping gene counts between individual module in each dataset and consensus modules, MElightgreen in GSE43292 and MEbrown in GSE28829 had the most overlapping gene counts with the

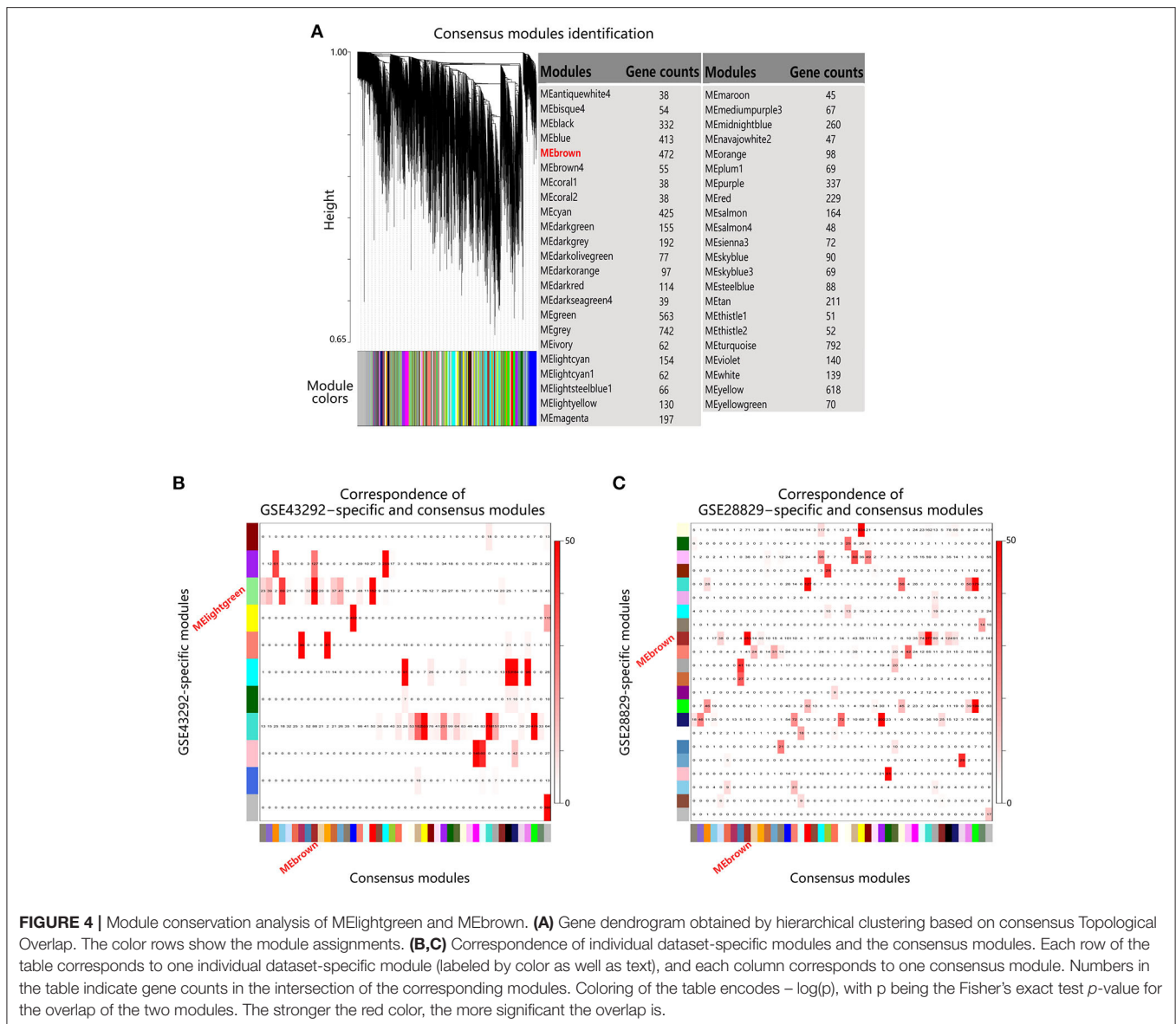


same module in consensus modules (Figures 4B,C), suggesting that MElightgreen in dataset GSE43292 and MEbrown in GSE28829 had high similarity.

Functional Enrichment Analysis of Genes in Advanced Plaque-Related Modules

Then GO and KEGG analysis was executed to assess the functions of genes in MElightgreen and MEbrown (Figure 5). In MElightgreen, the GO terms of BP were enriched in immune cell activation, adhesion, proliferation, and phagocytosis. The GO terms of CC were enriched in plasma membrane, secretory

granule, and phagocytic cup. The GO terms of MF were enriched in virus receptor, hijacked molecular function, Toll-like receptor, pattern recognition receptor, cargo receptor activity, and cytokine binding. The pathway terms were mainly enriched in immune response against infection, phagosome, rheumatoid arthritis, osteoblast differentiation, cytokine-cytokine receptor interaction, primary immunodeficiency, transcriptional misregulation in cancer, ferroptosis, apoptosis, endocytosis, and many signaling pathways like chemokine signaling pathway. In MEbrown, the GO terms of BP were enriched in immune cell activation, adhesion, proliferation. The GO terms of CC were enriched in plasma membrane, secretory granule, immunological



synapse, extracellular matrix, and endocytic vesicle. The GO terms of MF were enriched in chemokine activity, proteoglycan binding, lipopolysaccharide, cytokine activity, pattern recognition receptor activity, and phosphotyrosine residue binding. The pathway terms were mainly enriched in immune response against infection, rheumatoid arthritis, osteoblast differentiation, leukocyte transendothelial migration, platelet activity, lysosome, transcriptional misregulation in cancer, primary immunodeficiency, proteoglycans in cancer, arachidonic acid metabolism, regulation of actin cytoskeleton, axon guidance, and many signaling pathways like chemokine signaling pathway. Collectively, these results suggested that genes within advanced plaque-related modules were involved in immune-related functions.

Identification of Hub Genes in Advanced Plaque Progression

To further identify hub genes in advanced plaque progression, we first mapped differentially expressed genes (DEGs) into advanced plaque-related modules. As shown in **Figure 6A**, with the threshold of $|\log_2(\text{fold change (FC)})| > 1$ and $P < 0.05$, 796 DEGs were identified in GSE43292 with 483 upregulated and 313 downregulated. 390 DEGs were identified in GSE28829 with 304 upregulated and 86 downregulated. After intersection of these DEGs with advanced plaque-related modules, 325 DEGs were identified in MElightgreen and 250 DEGs were identified in MEbrown (**Figure 6B**). Further, the top 30 genes in MM values of MElightgreen and MEbrown were identified as their respective hub genes (**Figure 6C** and **Supplementary Table 2**). Notably,

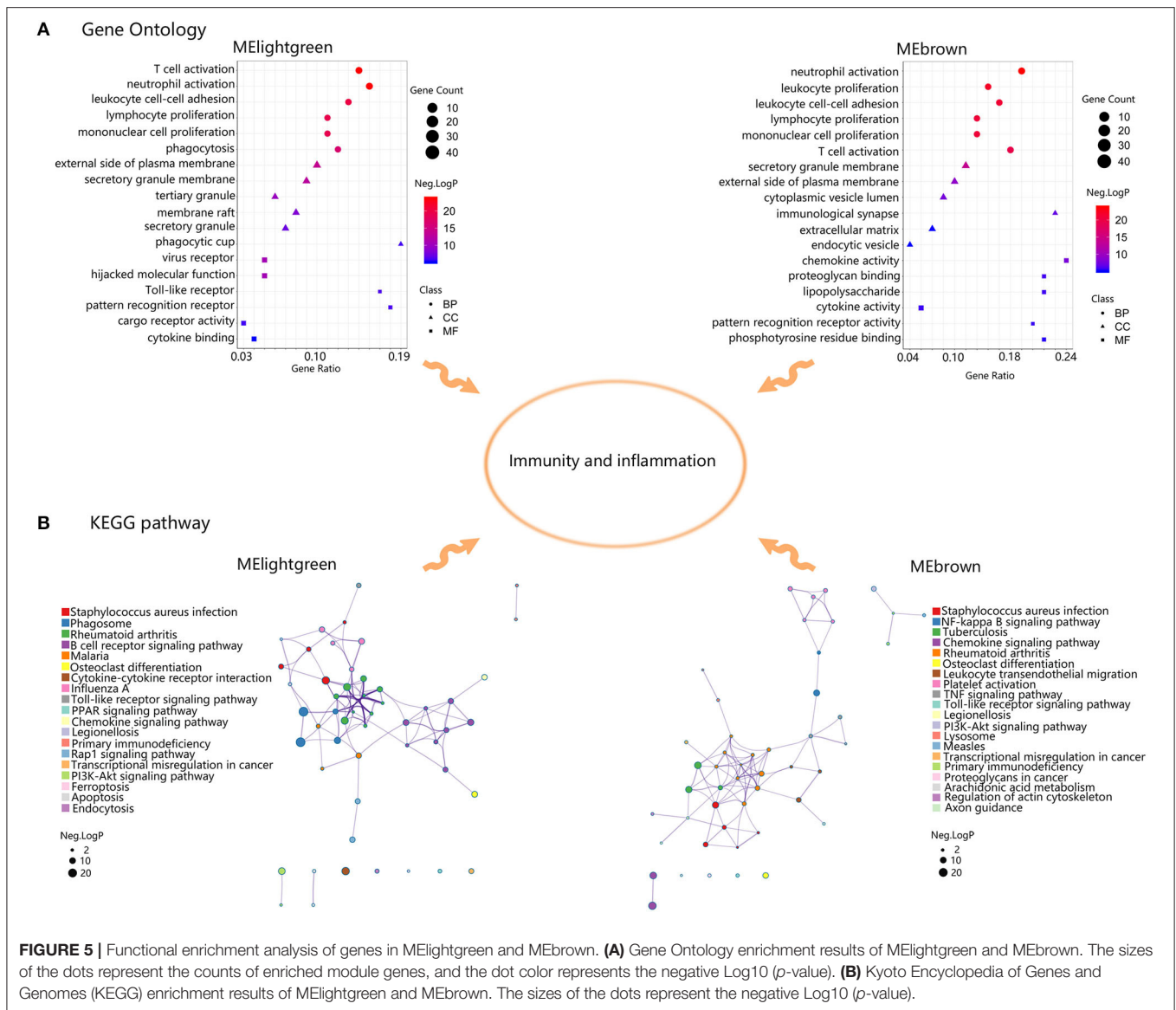


FIGURE 5 | Functional enrichment analysis of genes in MElightgreen and MEbrown. **(A)** Gene Ontology enrichment results of MElightgreen and MEbrown. The sizes of the dots represent the counts of enriched module genes, and the dot color represents the negative Log₁₀ (*p*-value). **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment results of MElightgreen and MEbrown. The sizes of the dots represent the negative Log₁₀ (*p*-value).

TYROBP, HCK, CD53, RBM47, and HAVCR2 were identified as common hub genes in the two datasets, indicating their key roles in advanced plaque formation.

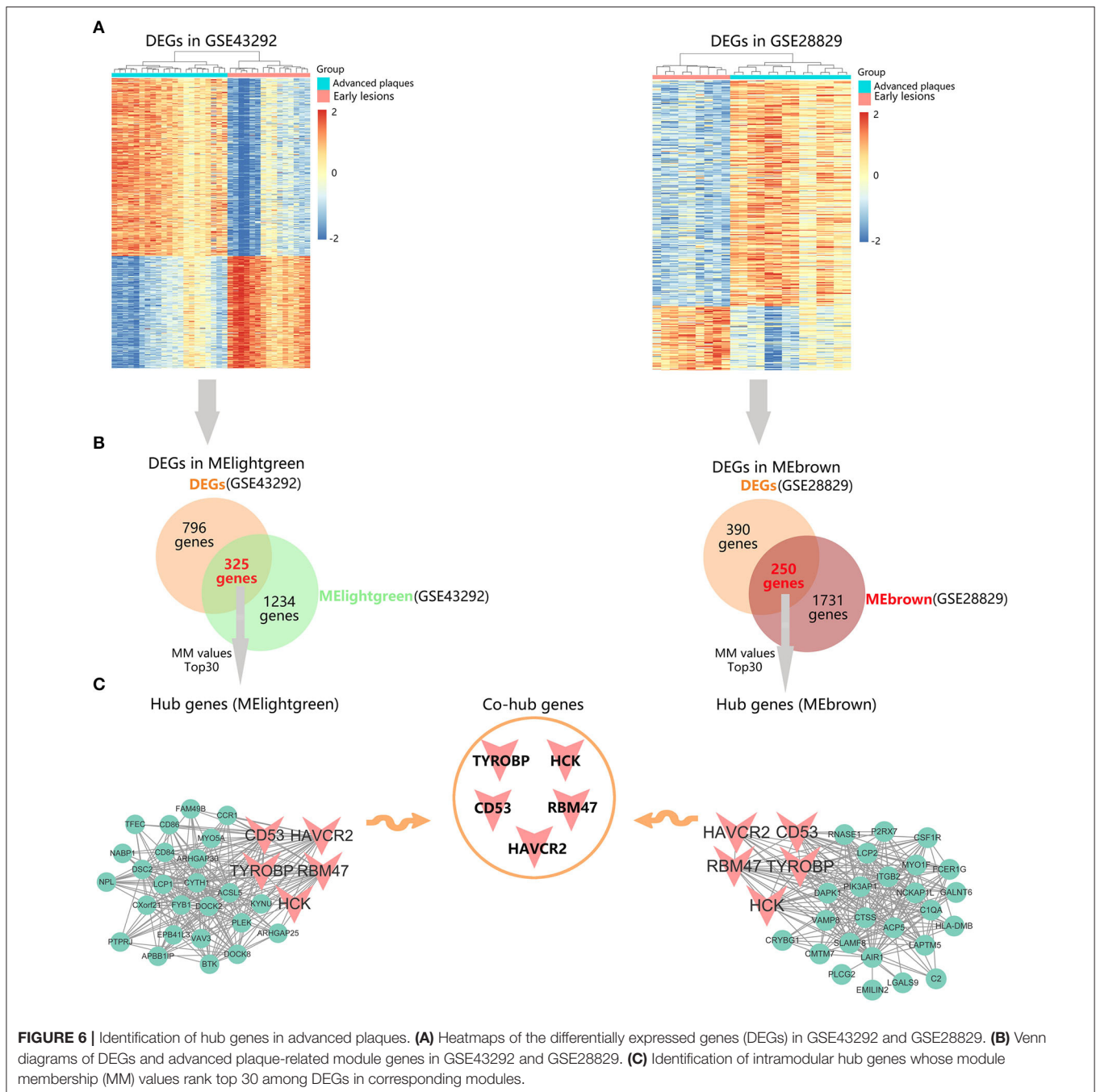
Validation of Hub Gene Expression in Advanced Atherosclerosis

To validate the expression of these five hub genes in advanced plaque formation, we constructed a mouse model by feeding ApoE^{-/-} mice with high fat-diet for 16 weeks. As shown in **Figure 7A**, mRNA levels of Rbm47, Havcr2, Hck, Cd53, and Tyrobp were significantly upregulated in advanced plaques group, compared to control group ($P < 0.05$). We further validated the expression of these genes in three other datasets. As shown in **Figure 7B**, with the threshold of $P < 0.05$, Havcr2, Hck, Cd53, and Tyrobp were significantly upregulated in advanced plaques group in both dataset GSE18443 and GSE31947

(the platform GPL3677 did not detect Rbm47 expression). Besides, CD53, HCK, TYROBP, and RBM47 were significantly upregulated in patients with coronary atherosclerotic disease in dataset GSE12288 (**Supplementary Figure 3**, the platform GPL96 did not detect HAVCR2 expression).

ROC Validation of Hub Genes in Coronary Atherosclerotic Disease

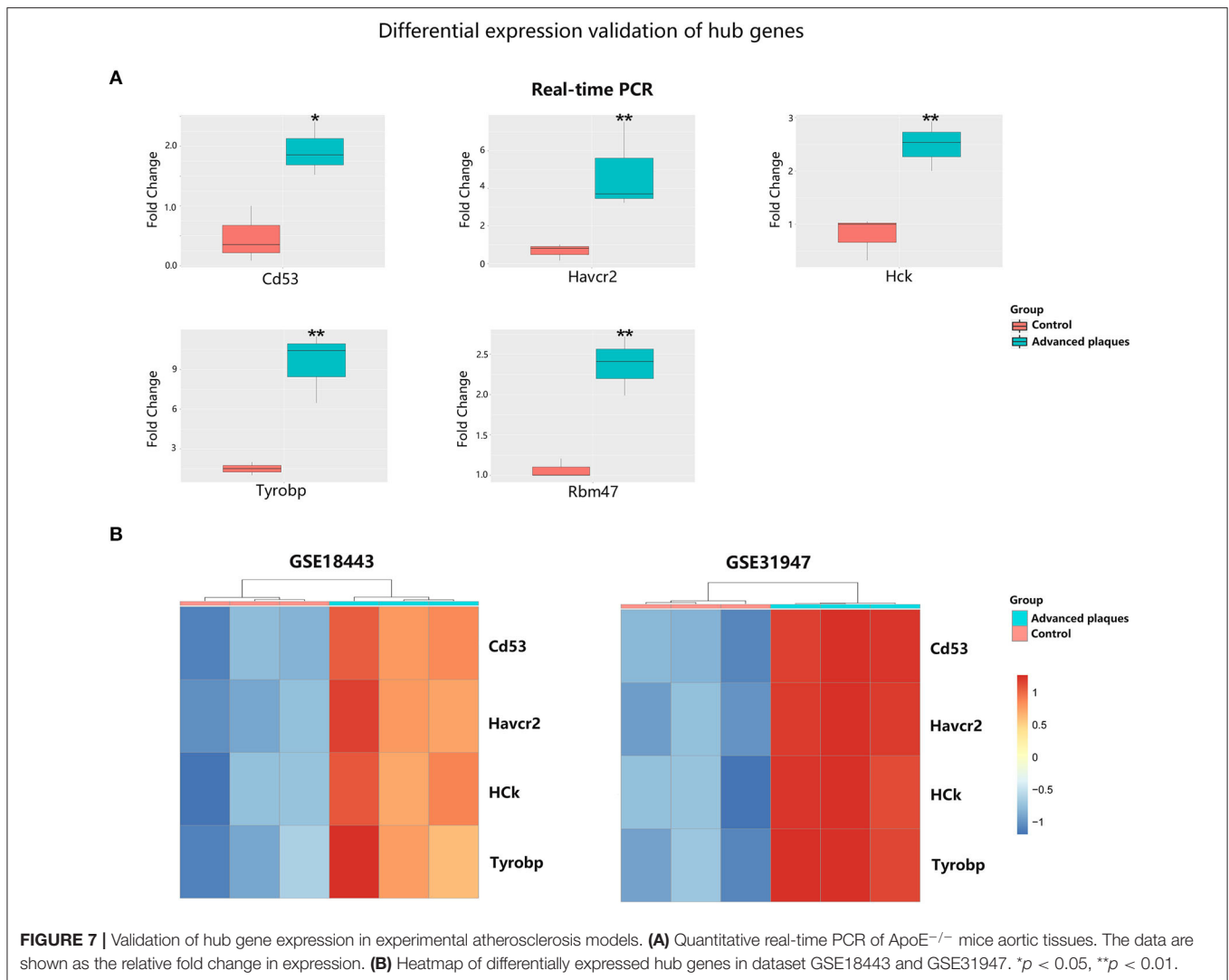
ROC analysis was performed to evaluate the capacity of these hub genes to distinguish samples of CAD group from those of non-CAD group. As shown in **Figures 8A–D**, the expression of CD53, HCK, RBM47 represented high efficiency in distinguishing capacity of CAD (AUC > 0.7). The expression of TYROBP showed modest efficiency in distinguishing capacity of CAD (AUC = 0.6969). The platform GPL96 did not detect HAVCR2 expression.



GSEA of Hub Gene Function in Advanced Atherosclerosis

To further explore the potential functions of CD53, HCK, RBM47, TYROBP, and HAVCR2 in advanced atherosclerosis, GSEA was performed on these hub genes, respectively. As shown in **Supplementary Figure 4**, genes in high expression groups of CD53, HCK, RBM47, TYROBP, and HAVCR2 were all enriched in lysosome; genes in high expression groups of HCK and RBM47 were both enriched in complement and coagulation

casades and sphingolipid metabolism; genes in high expression group of CD53 were enriched in Toll like receptor signaling and vibrio cholerae infection; genes in high expression group of HAVCR2 were enriched in T cell receptor signaling and B cell receptor signaling genes in high expression group of TYROBP were enriched in NOD like receptor signaling and natural killer cell mediated cytotoxicity. Collectively, these results suggested that these hub genes regulated immune cell function-related mechanisms in advanced atherosclerosis.



Identifying Key Transcriptional Regulatory Networks in Advanced Plaque-Related Modules

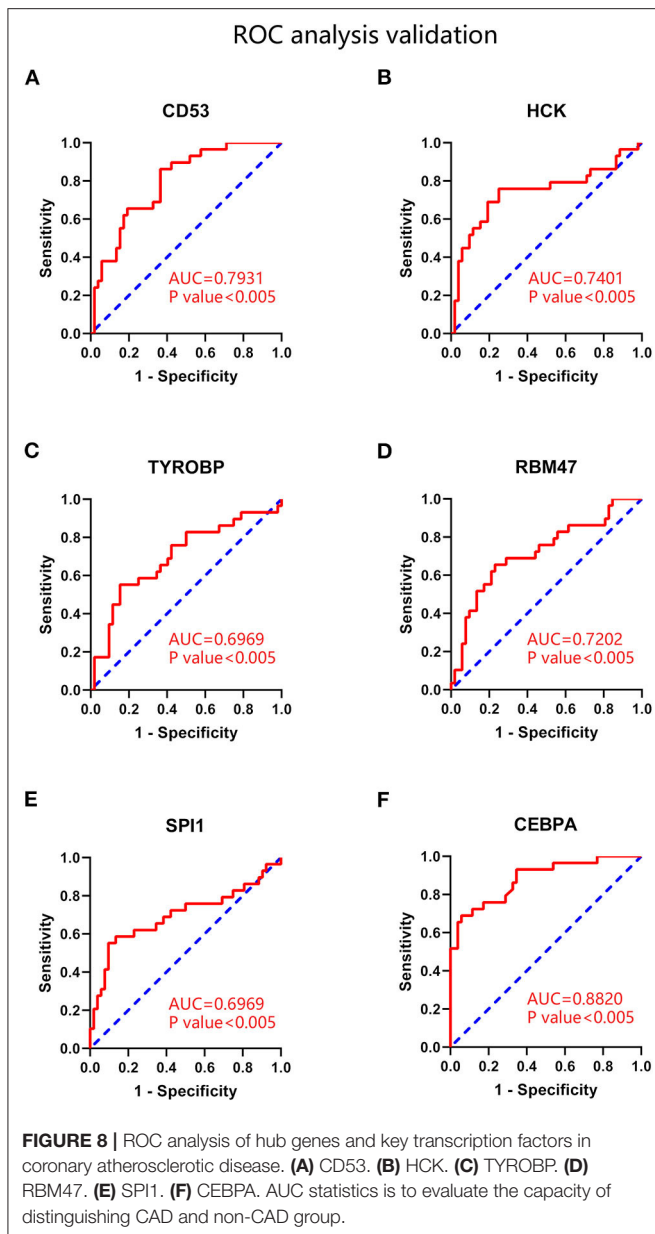
To identify key transcriptional regulatory networks that control these hub genes, TRRUST database was used. As shown in **Supplementary Table 3**, there were 12 transcription factor-mediated regulatory networks were significantly enriched in MELightgreen, and 24 transcription factor-mediated regulatory networks were significantly enriched in MEBrown. SPI1, RELA, NFKB1, ETS1, SP1, ERG, STAT1, STAT3, TRERF1, ETS2, and CEBPA mediated regulatory networks were enriched in both MELightgreen and MEBrown.

Then differential expression analysis and ROC analysis were applied to validate these transcription factors. As shown in **Supplementary Table 4**, only SPI1 and CEBPA were significantly upregulated in advanced plaques (*P* < 0.05). Subsequent ROC analysis demonstrated that the expression of CEBPA showed high efficiency in distinguishing capacity of CAD (AUC = 0.8820). The expression of SPI1 showed modest

efficiency in distinguishing capacity of CAD (AUC = 0.6969), as shown in **Figures 8E,F**. These results indicated that SPI1 and CEBPA played key transcriptional regulatory roles in advanced atherosclerosis.

PPI Network Construction and Analysis

To better understand the protein-protein interactions of hub genes in advanced plaques based on prior knowledge, PPI network was constructed. As shown in **Supplementary Tables 5, 6**, there were 2,972 interactions identified in the PPI network of MELightgreen, and 2,240 interactions identified in the PPI network of MEBrown based on previous evidence. By calculating node degree, the 10 genes with the highest degree in MELightgreen and MEBrown were identified separately (**Supplementary Table 7**). Importantly, PTPRC, TYROBP, ITGB2, CD86, PLEK, LCP2, TLR2, TLR8, and CSF1R were the common high-degree genes.



DISCUSSION

In this study, two advanced atherosclerotic plaque-related microarray datasets were analyzed by WGCNA to identify gene sets (modules) and their hub genes that were most related to advanced plaques. Functional enrichment analysis showed that advanced plaques-related modules were enriched in immune-related mechanism. The common hub genes of advanced plaque-related modules between two datasets were further validated by quantitative real-time PCR in advanced atherosclerotic tissue and by other datasets. Besides, these hub genes were also validated for distinguishing capacity of CAD through ROC analysis and for potential functions in advanced atherosclerosis through GSEA.

Weighted gene co-expression network analysis (WGCNA) was the main analytical method in this study. It can be used to identify gene sets (modules) of highly co-expressed genes, to summarize such gene sets using module eigengene (ME) values or intramodular hub genes, to correlate modules with one another and with clinical traits of samples through eigengene network methodology, and to calculate module membership (MM) values of genes (Zhang and Horvath, 2005). In addition, WGCNA is based on the idea of correlation networks rather than individual genes which can effectively identify candidate biomarkers or therapeutic targets. This analytical method has been successfully applied in diverse biomedical fields, like cancer, model organism genetics and analysis of neuroimaging data (Carlson et al., 2006; Ghazalpour et al., 2006; Horvath et al., 2006; Weston et al., 2008). In the present study, we identified advanced plaque-related gene co-expression networks through WGCNA and found they were associated with immune-related pathways, indicating the central role of immune mechanisms in advanced plaque formation. In recent decades, immune and inflammatory mechanisms have become increasingly important in experimental atherosclerosis studies (Minelli et al., 2020). Many initial factors such as hemodynamic disorder, endothelial dysfunction and subendothelial buildup of cholesterol-carrying LDL trigger innate and adaptive immune responses, which affect plaque progression through a complex interaction network, balancing pro-atherogenic inflammatory and atheroprotective anti-inflammatory responses (Libby and Hansson, 2015). Additionally, interleukin 1 β -targeted monoclonal antibody canakinumab significantly lowered the rate of recurrent cardiovascular events in previous myocardial infarction patients, indicating the potential of immunotherapy in atherosclerosis progression (Ridker et al., 2011). However, rather than broad-spectrum anti-inflammatory therapy, investigation of specific inhibition of key targets is warranted (Minelli et al., 2020).

In this study, RBM47, HCK, CD53, TYROBP, and HAVCR2 were identified as hub genes in plaque progression. RNA-Binding-Motif-protein-47 (RBM47), a component of the editosome, has been implicated in the editing of Apob mRNA *in vivo* and *in vitro*, which affects the production of Apob functional proteins APOB100 and APOB48 (Fossat et al., 2014). ApoB is a cholesterol-carrying component of LDL with a well-known atherogenic effect on artery subendothelial lipid retention and accumulation (Skalen et al., 2002). But it remains to be clarified that RBM47 functions as an RNA editing factor for ApoB in immune cells and atherosclerotic plaques tissue. HCK is a member of the Src family of tyrosine kinases, which transmits membrane receptor signals and plays an important role in survival, proliferation migration and phagocytosis of immune cells (Wang et al., 2020). HCK has also been reported to participate in leukocyte adhesion and metastasis, which may promote atherosclerotic plaque formation (Medina et al., 2015). However, the specific mechanism of HCK in atherosclerosis development still need further elaboration. CD53 is a member of the tetraspanin family and mediates signal transduction functions (Yeung et al., 2020). It has been suggested that CD53 regulated the growth of T cells and natural killer cells and has been extensively studied in infectious diseases, autoimmune diseases,

and immunodeficiency diseases (Kim et al., 2004; Tohami et al., 2004; Pedersen-Lane et al., 2007). Interestingly, CD53 has been shown to curb inflammatory cytokines secretion and pathway activation of THP-1 cells, which is the human monocyte and widely used to construct atherosclerotic models *in vitro* (Lee et al., 2013). Hepatitis A virus cellular receptor 2 (HAVCR2, also known as Tim-3), a cell surface receptor, was widely believed to play an inhibitory role by associating with some molecules like ZAP70, LCP2, LCK, and FYN and blocking phosphorylation of key components of signaling pathways (Monney et al., 2002). It was reported that HAVCR2 could regulate macrophage activation and inhibit T-helper type 1 lymphocyte-mediated auto- and alloimmune responses and promote immunological tolerance (Monney et al., 2002; Sánchez-Fueyo et al., 2003). Amanda C Foks, et al. showed that anti-Tim-3 antibody treatment promoted plaque formation in atherosclerosis model, increased lesional percentages of macrophages and CD4(+) T cells and enhanced their activation (Foks et al., 2013). However, the immunosuppressive mechanisms of HAVCR2 against atherosclerosis remains to be elucidated. TYROBP, also known as DNAX-activating protein of 12 kDa (DAP12), encodes immune-related signal transduction adaptin on the membrane and works through an immunoreceptor tyrosine-based activation motif (Kobayashi et al., 2015). DAP12 was widely involved in the proliferation, survival, differentiation, and polarization of immune cells, especially monocyte-macrophage system (Otero et al., 2009; Kobayashi et al., 2015). Interestingly, TYROBP was also identified as the high-degree gene in the PPI network of advanced plaques, indicating its core role in advanced atherosclerosis.

CONCLUSIONS

In summary, our present study has identified RBM47, HCK, CD53, TYROBP, and HAVCR2 as hub genes in advanced plaque progression by weighted gene co-expression network analysis. They were validated to be upregulated during advanced plaque progression, to well distinguish CAD patients from non-CAD people and to regulate immune cell function-related mechanisms in advanced atherosclerosis. Our results provide novel molecular

targets for future research. These hub genes may also serve as promising therapeutic intervention targets to inhibit advanced plaque progression.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee, SYSU.

AUTHOR CONTRIBUTIONS

ZL and JW directed the project. HZ and SW designed experiments. CL performed experiments and drafted the manuscript. YC and ZC performed data analysis. All authors reviewed the manuscript, read, and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (81900387, 81870334, 81870170), Guangdong Basic and Applied Basic Research Found (2019A1515011806, 2020A1515011237), Fundamental Research Funds for the Central Universities (19ykpy97), the Science and Technology Program of Guangzhou City of China (201803040010), Guangdong Science and Technology Department (2017B030314026).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.602908/full#supplementary-material>

REFERENCES

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Carlson, M. R., Zhang, B., Fang, Z., Mischel, P. S., Horvath, S., and Nelson, S. F. (2006). Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* 7:40. doi: 10.1186/1471-2164-7-40
- Carvalho, B. S., and Irizarry, R. A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26, 2363–2367. doi: 10.1093/bioinformatics/btq431
- Chen, S., Yang, D., Liu, Z., Li, F., Liu, B., Chen, Y., et al. (2020). Crucial gene identification in carotid atherosclerosis based on peripheral blood mononuclear cell (PBMC) data by weighted (gene) correlation network analysis (WGCNA). *Med. Sci. Monit.* 26:e921692. doi: 10.12659/MSM.921692
- Consortium, T. G. O. (2019). The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res.* 47, d330–d338. doi: 10.1093/nar/gky1055
- Feltrin, A. S., Tahira, A. C., Simões, S. N., Brentani, H., and Martins, D. C. Jr. (2019). Assessment of complementarity of WGCNA and NER1 results for identification of modules associated to schizophrenia spectrum disorders. *PLoS ONE* 14:e0210431. doi: 10.1371/journal.pone.0210431
- Foks, A. C., Ran, I. A., Wasserman, L., Frodermann, V., Ter Borg, M. N., de Jager, S. C., et al. (2013). T-cell immunoglobulin and mucin domain 3 acts as a negative regulator of atherosclerosis. *Arterioscl. Thromb. Vasc. Biol.* 33, 2558–2565. doi: 10.1161/ATVBAHA.113.301879
- Fossat, N., Tourle, K., Radziewicz, T., Barratt, K., Liebhold, D., Studdert, J. B., et al. (2014). C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep.* 15, 903–910. doi: 10.15252/embr.201438450
- Fuster, V., Moreno, P. R., Fayad, Z. A., Corti, R., and Badimon, J. J. (2005). Atherothrombosis and high-risk plaque: part I: evolving concepts. *J. Am. Coll. Cardiol.* 46, 937–954. doi: 10.1016/j.jacc.2005.03.074

- Gaurav, C., Saurav, B., Goutam, R., and Goyal, A. K. (2015). Nano-systems for advanced therapeutics and diagnosis of atherosclerosis. *Curr. Pharma. Design* 21, 4498–4508. doi: 10.2174/1381612821666150917094215
- Ghazalpour, A., Doss, S., Zhang, B., Wang, S., Plaisier, C., Castellanos, R., et al. (2006). Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* 2:e130. doi: 10.1371/journal.pgen.0020130
- Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46, d380–d386. doi: 10.1093/nar/gkx1013
- Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., et al. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17402–17407. doi: 10.1073/pnas.0608396103
- Jiao, M., Li, J., Zhang, Q., Xu, X., Li, R., Dong, P., et al. (2020). Identification of four potential biomarkers associated with coronary artery disease in non-diabetic patients by gene co-expression network analysis. *Front. Genet.* 11:542. doi: 10.3389/fgene.2020.00542
- Kanehisa, M., and Sato, Y. (2020). KEGG mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35. doi: 10.1002/pro.3711
- Kim, T. R., Yoon, J. H., Kim, Y. C., Yook, Y. H., Kim, I. G., Kim, Y. S., et al. (2004). LPS-induced CD53 expression: a protection mechanism against oxidative and radiation stress. *Mol. Cells* 17, 125–131.
- Kobayashi, M., Konishi, H., Takai, T., and Kiyama, H. (2015). A DAP12-dependent signal promotes pro-inflammatory polarization in microglia following nerve injury and exacerbates degeneration of injured neurons. *Glia* 63, 1073–1082. doi: 10.1002/glia.22802
- Kovanen, P. T. (2019). Mast cells as potential accelerators of human atherosclerosis—from early to late lesions. *Int. J. Mol. Sci.* 20:4479. doi: 10.3390/ijms20184479
- Lee, H., Bae, S., Jang, J., Choi, B. W., Park, C. S., Park, J. S., et al. (2013). CD53, a suppressor of inflammatory cytokine production, is associated with population asthma risk via the functional promoter polymorphism–1560 C>T. *Biochim. Biophys. Acta* 1830, 3011–3018. doi: 10.1016/j.bbagen.2012.12.030
- Lee, K. K., Stelzle, D., Bing, R., Anwar, M., Strachan, F., Bashir, S., et al. (2019). Global burden of atherosclerotic cardiovascular disease in people with hepatitis C virus infection: a systematic review, meta-analysis, and modelling study. *Lancet Gastroenterol. Hepatol.* 4, 794–804. doi: 10.1016/S2468-1253(19)30227-4
- Libby, P., and Hansson, G. K. (2015). Inflammation and immunity in diseases of the arterial tree: players and layers. *Circul. Res.* 116, 307–311. doi: 10.1161/CIRCRESAHA.116.301313
- Lin, M., Zhao, L., Zhao, W., and Weng, J. (2014). Dissecting the mechanism of carotid atherosclerosis from the perspective of regulation. *Int. J. Mol. Med.* 34, 1458–1466. doi: 10.3892/ijmm.2014.1960
- Liu, W., Zhao, Y., and Wu, J. (2018). Gene expression profile analysis of the progression of carotid atherosclerotic plaques. *Mol. Med. Rep.* 17, 5789–5795. doi: 10.3892/mmr.2018.8575
- Medina, I., Cougoule, C., Drechsler, M., Bermudez, B., Koenen, R. R., Sluimer, J., et al. (2015). Hck/Fgr kinase deficiency reduces plaque growth and stability by blunting monocyte recruitment and intraplaque motility. *Circulation* 132, 490–501. doi: 10.1161/CIRCULATIONAHA.114.012316
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45, d183–d189. doi: 10.1093/nar/gkw1138
- Miao, L., Yin, R. X., Pan, S. L., Yang, S., Yang, D. Z., and Lin, W. X. (2018). Weighted gene co-expression network analysis identifies specific modules and hub genes related to hyperlipidemia. *Cell. Physiol. Biochem.* 48, 1151–1163. doi: 10.1159/000491982
- Minelli, S., Minelli, P., and Montinari, M. R. (2020). Reflections on atherosclerosis: lesson from the past and future research directions. *J. Multidiscip. Healthc.* 13, 621–633. doi: 10.2147/JMDH.S254016
- Monney, L., Sabatos, C. A., Gaglia, J. L., Ryu, A., Waldner, H., Chernova, T., et al. (2002). Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* 415, 536–541. doi: 10.1038/415536a
- Otero, K., Turnbull, I. R., Poliani, P. L., Vermi, W., Cerutti, E., Aoshi, T., et al. (2009). Macrophage colony-stimulating factor induces the proliferation and survival of macrophages via a pathway involving DAP12 and beta-catenin. *Nat. Immunol.* 10, 734–743. doi: 10.1038/ni.1744
- Pedersen-Lane, J. H., Zurier, R. B., and Lawrence, D. A. (2007). Analysis of the thiol status of peripheral blood leukocytes in rheumatoid arthritis patients. *J. Leukocyte Biol.* 81, 934–941. doi: 10.1189/jlb.0806533
- Ridker, P. M., Thuren, T., Zalewski, A., and Libby, P. (2011). Interleukin-1 β inhibition and the prevention of recurrent cardiovascular events: rationale and design of the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS). *Am. Heart J.* 162, 597–605. doi: 10.1016/j.ahj.2011.06.012
- Sánchez-Fueyo, A., Tian, J., Picarella, D., Domenig, C., Zheng, X. X., Sabatos, C. A., et al. (2003). Tim-3 inhibits T helper type 1-mediated auto- and alloimmune responses and promotes immunological tolerance. *Nat. Immunol.* 4, 1093–1101. doi: 10.1038/ni987
- Shi, W., Zou, R., Yang, M., Mai, L., Ren, J., Wen, J., et al. (2019). Analysis of genes involved in ulcerative colitis activity and tumorigenesis through systematic mining of gene co-expression networks. *Front. Physiol.* 10:662. doi: 10.3389/fphys.2019.00662
- Skalen, K., Gustafsson, M., Rydberg, E. K., Hulten, L. M., Wiklund, O., Innerarity, T. L., et al. (2002). Subendothelial retention of atherogenic lipoproteins in early atherosclerosis. *Nature* 417, 750–754. doi: 10.1038/nature00804
- Smyth, G. K. (2005). “Limma: linear models for microarray data,” in *Bioinformatics and Computational Biology Solution Using R and Bioconductor*, eds R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, and S. Dudoit (New York, NY: Springer), 397–420.
- Tan, X., Zhang, X., Pan, L., Tian, X., and Dong, P. (2017). Identification of key pathways and genes in advanced coronary atherosclerosis using bioinformatics analysis. *BioMed. Res. Int.* 2017:4323496. doi: 10.1155/2017/4323496
- Tohami, T., Drucker, L., Radnay, J., Shapira, H., and Lishner, M. (2004). Expression of tetraspanins in peripheral blood leukocytes: a comparison between normal and infectious conditions. *Tissue Antigens* 64, 235–242. doi: 10.1111/j.1399-0039.2004.00271.x
- Wang, J., Wei, B., Cao, S., Xu, F., Chen, W., Lin, H., et al. (2014). Identification by microarray technology of key genes involved in the progression of carotid atherosclerotic plaque. *Genes Genetic Syst.* 89, 253–258. doi: 10.1266/ggs.89.253
- Wang, Z., Ying, C., Zhang, A., Xu, H., Jiang, Y., and Lou, M. (2020). HCK promotes glioblastoma progression by TGF β signaling. *Biosci. Rep.* 40:BSR20200975. doi: 10.1042/BSR20200975
- Weston, D. J., Gunter, L. E., Rogers, A., and Wulschleger, S. D. (2008). Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Syst. Biol.* 2:16. doi: 10.1186/1752-0509-2-16
- Yeung, L., Anderson, J. M. L., Wee, J. L., Demaria, M. C., Finsterbusch, M., Liu, Y. S., et al. (2020). Leukocyte tetraspanin CD53 restrains α (3) integrin mobilization and facilitates cytoskeletal remodeling and transmigration in mice. *J. Immunol.* 205, 521–532. doi: 10.4049/jimmunol.1901054
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:Article17. doi: 10.2202/1544-6115.1128
- Zhang, X., Sun, R., and Liu, L. (2019). Potentially critical roles of TNPO1, RAP1B, ZDHHC17, and PPM1B in the progression of coronary atherosclerosis through microarray data analysis. *J. Cell. Biochem.* 120, 4301–4311. doi: 10.1002/jcb.27715
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Liu, Zhang, Chen, Wang, Chen, Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.