

# Machine Learning Approach for Screening Key Genes of Atherosclerosis and Performing Pan-Cancer Analysis

Tingting Zhao<sup>1,3\*</sup>, Zhenrun Zhan<sup>2</sup>, Xiaoyuan He<sup>3</sup>, Xu Tang<sup>1,3</sup>

<sup>1</sup>Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi 046000, China

<sup>2</sup>Department of Endocrinology, the First Affiliated Hospital, Fujian Medical University, Fuzhou, 350005, China

<sup>3</sup>Changzhi Medical College, Changzhi, Shanxi 046000, China

#Zhenrun Zhan and Tingting Zhao Contributed to the work equally

## Research Article

### ABSTRACT

**Objective:** Atherosclerosis (AS), an immunoinflammatory disease caused by lipids, is a significant factor in coronary heart disease and stroke. Researchers worldwide are working to develop more effective ways of diagnosing and treating it. This article introduces a machine-learning algorithm that screens biomarkers and performs pan-cancer analysis as a reference.

**Patients and methods:** We first downloaded the Gene Expression Omnibus (GEO) dataset containing information on atherosclerotic patients for differential gene screening and analyzed differential mRNAs using Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis. Then, we combined Weighted Gene Co-expression Network Analysis (WGCNA), protein interaction network analysis, and mechanistic learning algorithms to screen for core genes, perform basic experimental validation, and immuno-infiltration analysis. Finally, we used The Cancer Genome Atlas Program (TCGA) and Genotype-Tissue Expression (GTEx) databases for pan-carcinogenic analysis of differentially expressed genes.

**Results:** After screening differentially expressed genes from the GSE28829 dataset, we performed WGCNA and obtained 122 key genes by combining module genes and differential expression genes. We used four machine learning algorithms (XGBoost, RandomForest, Support Vector Machine with Recursive Feature Elimination (SVM-REF), and Generalized Linear Model (GLM)) to calculate the critical genes and obtain the junction of the results, which led us to identify four hub genes (Signaling Lymphocytic Activation Molecule Family Member 8 (*SLAMF8*), Toll-Like Receptor 2 (*TLR2*), Vesicle-Associated Membrane Protein 8 (*VAMP8*), and V-set and Immunoglobulin Domain Containing 4 (*VSIG4*)). Next, we built a diagnostic model and assessed its capabilities. The Receiver Operating Characteristic (ROC) curves of the four genes suggested their critical role in the development of atherosclerosis. We then performed gene correlation analysis, immune infiltration analysis, and Reverse Transcription Polymerase Chain Reaction (RT-PCR) verification of the four core genes and finally screened out *TLR2* for pancreatic carcinoma. Our analysis found that *TLR2*'s expression in patients with various tumors differed from that in healthy individuals, and it was strongly associated with the prognosis of patients with diverse cancers.

**Conclusion:** *TLR2* may be a target for intervention in developing diseases such as atherosclerosis and tumors.

**Keywords:** WGCNA; Atherosclerosis; Pan-cancer; Machine learning; Pan-carcinoma; Database

**Received:** 08-Aug-2024,

Manuscript No. JMB-24-144862;

**Editor assigned:** 12-Aug-2024,

PreQC No. JMB-24-144862 (PQ);

**Published:** 09-Sep-2024,

DOI: 10.4172/2320-3528.13.3.004

**\*For Correspondence:**

Tingting Zhao,

Heping Hospital Affiliated to Changzhi Medical College,

Changzhi, Shanxi 046000, China

**E-mail:** 18835599006@163.com

**Citation:** Zhao T, et al. Machine Learning Approach for Screening Key Genes of Atherosclerosis and Performing Pan-Cancer Analysis. J Microbiol Biotechnol. 2024;13:004

**Copyright:** © 2024 Zhao T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### INTRODUCTION

Cardiovascular heart Disease (CVD) caused by atherosclerosis is the leading cause of death worldwide. The World Health Organization predicts that by 2030, 23.6 million people will die from CVD [1]. Atherosclerosis most commonly occurs in the coronary arteries and carotid arteries, etc. Coronary atherosclerosis often leads to diseases of the cardiovascular system, while carotid atherosclerosis often causes diseases of the cerebrovascular system. Carotid atherosclerotic stenosis is caused by the chronic accumulation of carotid subintimal atherosclerotic plaques, and 20%~30% of ischemic strokes or transient ischemic attacks are related to carotid atherosclerotic stenosis [2]. The atherosclerosis formation mechanism involves vascular endothelial injury, mononuclear macrophage adhesion, lipid deposition, and other factors [3]. Although many factors affect carotid atherosclerosis, the mechanism that affects the atherosclerosis process is not fully understood. Therefore, it is essential to explore further and find the key regulatory genes and potential signaling pathways, particularly to prevent the occurrence of atherosclerosis and delay its progression.

It is of great clinical significance to use gene chip technology to detect and analyze clinical patient specimens, screen valuable genes, and then deeply study the pathogenesis of atherosclerosis. Bioinformatics analysis can screen for differential genes for related diseases at the genomic level. However, independent microarray analysis often results in false positive rates. In this study, we downloaded the GSE28829 dataset from the GEO database, which studies early and

advanced carotid atherosclerosis. We performed a series of preprocessing steps on the original data to screen out the differentially expressed genes associated with carotid atherosclerosis progression. We then conducted functional annotation and signaling pathway enrichment analysis of differentially expressed genes to identify potential signaling pathways. Finally, we used WGCNA combined with machine learning analysis to screen out the hub gene that regulates the progression of carotid atherosclerosis, and pan-cancer analysis was carried out [4,5]. The outcomes of this research will help to improve the comprehension of the molecular mechanism of the onset and progression of carotid atherosclerosis and the correlation of the hub gene with the occurrence of various tumors.

## MATERIALS AND METHODS

### Data set collection and processing

The GSE28829 dataset, containing 13 samples with early and 16 with severe atherosclerosis lesions, was downloaded from the GEO database [6]. The dataset was annotated, and batch effect elimination was performed to acquire an ideal expression matrix file.

### Differentially Expressed Gene (DEG) screening and Gene Set Enrichment Analysis (GSEA)

DEG screening was performed using the "limma" package, with a screening criterion of  $(\log_2 FC) > 0.5$  and  $FDR < 0.01$ . The important genes volcano and expression heat maps were plotted using "ggplot2" and "pheatmap" software packages. Genomic enrichment analysis of candidate genes was also performed.

### Screening WGCNA-based targeting modules and genes and functional enrichment analysis

WGCNA of the gene set was performed to identify atherosclerosis-related modules [7]. The  $\beta$  that meets the standard soft threshold of the scale-free network is selected to construct the gene co-expression network, and then the modules are preliminarily divided.

We calculated correlation coefficients between genes and modules to assess their association, and finally, the hub genes within the module were screened. Finally, these hub genes are functionally annotated and pathway enriched [8].

### Combined with machine learning to screen and validate atherosclerotic biomarkers

XGBoost algorithm, Random forest machine learning algorithm [9], GLM algorithm and SVM-RFE algorithm [10] based on the e1071 packet were used to screen core genes from important DEGs. Venn packets were used to intersect the four to obtain key genes, their expression was compared between groups, the expression of key genes based on the early and advanced atherosclerosis groups was collected.

The ROC curve of key genes was drawn using the pROC package, and their diagnostic efficiency in diagnosing atherosclerosis was evaluated.

### Construction and validation of the signature

A diagnostic nomogram was drawn using R to construct a diagnostic model containing four core genes, and its ROC curve was plotted using the pROC software package to reflect its diagnostic efficiency [11].

### Enrichment analysis of core biomarkers

GSEA of *SLAMF8*, *TLR2*, *VAMP8*, and *VSIG4* was performed to investigate their association with atherosclerosis development differences in signalling pathway enrichment of DEGs and different groups in the GSE28829 dataset were explored using the "GSVA" package.

### Sample collection and quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

Twenty patients who underwent cervical vascular ultrasound were enrolled from May 2022 to November 2022 at Heping Hospital, Affiliated with Changzhi Medical College. Ten patients had carotid atherosclerosis; the remaining 10 were controls without atherosclerosis. Peripheral blood was collected and stored at  $-80^\circ\text{C}$  for later use.

Total RNA was extracted using TRIzol® reagent and cDNA synthesis was performed according to the reverse transcription kit instructions (YEASEN Biotech). qRT-PCR was performed using cDNA as a template, and the results were analyzed with the  $2^{-\Delta\Delta Ct}$  method. Primers used for PCR are provided in Table 1.

**Table 1.** All primers used for amplification.

| Gene   | Primer Sequence (5'→3') |                         |
|--------|-------------------------|-------------------------|
| GAPDH  | Forward                 | GGAGCGAGATCCCTCCAAAAT   |
|        | Reverse                 | GGCTGTTGTCATACTTCTCATGG |
| SLAMF8 | Forward                 | CTGGAGACTCTGTACCATTCCC  |
|        | Reverse                 | AGCAATGAACACTTGTACCACG  |
| TLR2   | Forward                 | ATCCTCCAATCAGGCTTCTCT   |
|        | Reverse                 | GGACAGGTCAAGGCTTTTACA   |
| VAMP8  | Forward                 | TGTGCGGAACCTGCAAAGT     |
|        | Reverse                 | CTTCTGCGATGTCGTCTTGAA   |
| VSIG4  | Forward                 | GGGGCACCTAACAGTGGAC     |
|        | Reverse                 | GTCTGAGCCACGTTGTACCAG   |

Flow cytometry

Peripheral Blood Mononuclear Cells (PBMCs) were washed twice and re-suspended in RPMI medium (Sigma-Aldrich, USA) supplemented with 5% fetal bovine serum (Gibco, Australia) and stained with FIT-labeled anti-CD206, PE-labeled anti-CD11b, PerCP/Cyanine 5.5-labeled CD27 and APC-labeled anti-CD38 on ice in a dark room for 30 min on surface staining. Fluorescent minus one control tube was also prepared for each antibody. Finally, cell samples were collected and analyzed on a FACSCanto flow cytometer using BD FACSDiva software (BD Biosciences).

Pan-cancer candidate biomarker expression

RNA expression profiles and clinical data for 33 tumors were downloaded from <https://portal.gdc.com>, and pan-cancer analysis was conducted to assess the expression of candidate genes in each tumor. P-values below 0.05 were considered significant, indicating differential expression between tumors and normal tissues.

Analysis of prognosis

RNA expression profiles and clinical information for 33 cancers were obtained from the TCGA database. The 'forest plot' R package was used to perform univariate cox regression analysis to calculate the p-value, HR and 95% CI.

RESULTS

Data processing and screening and analysis of differential genes

Using a boxplot, we started by evaluating the batch effect of the expression matrix from the GSE28829 dataset. Figure 1 displays the differential expression of multiple genes in multiple sample data. The effectiveness of the "sva" software package in eliminating batch effects is shown by comparing Figure 1A to Figure 1B. The expression matrix file analyzed to identify 23 downregulated and 154 upregulated differentially expressed genes, visualized as volcano maps in Figure 2A. We created heat maps of the top 60 upregulated genes with the most significant differences using the "pheatmap" package (Figure 2B).

Figure 1. (A,B): Box plots of the normalized between-sample primary data.

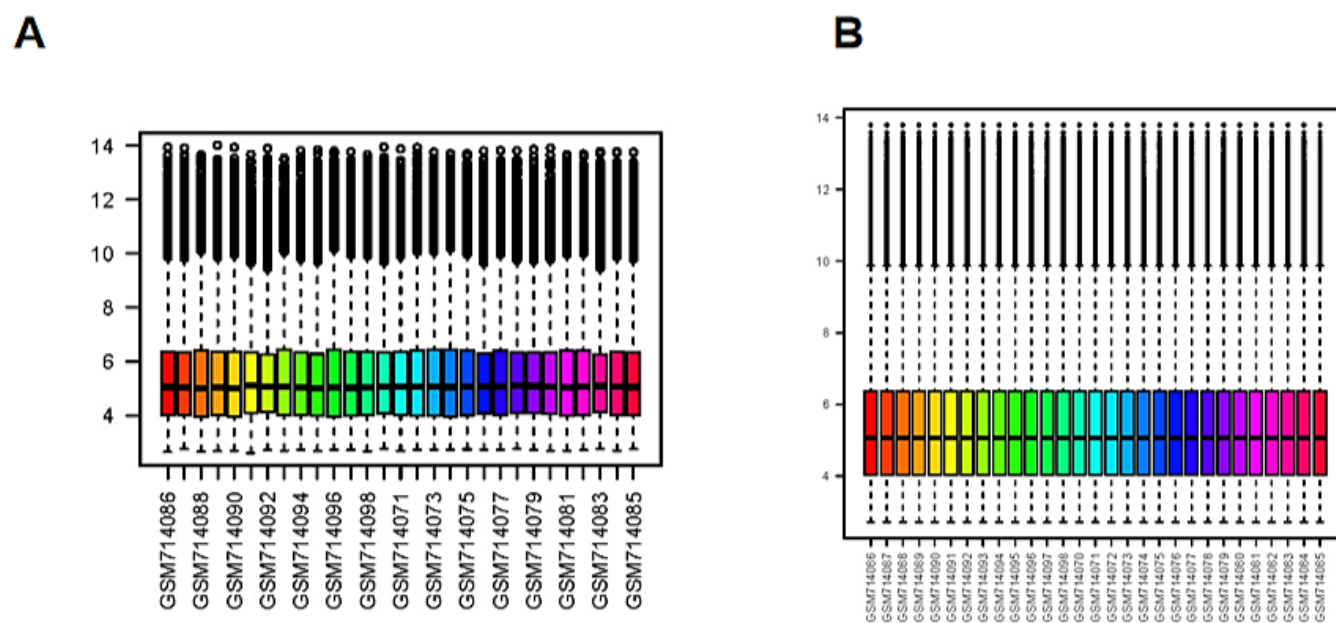
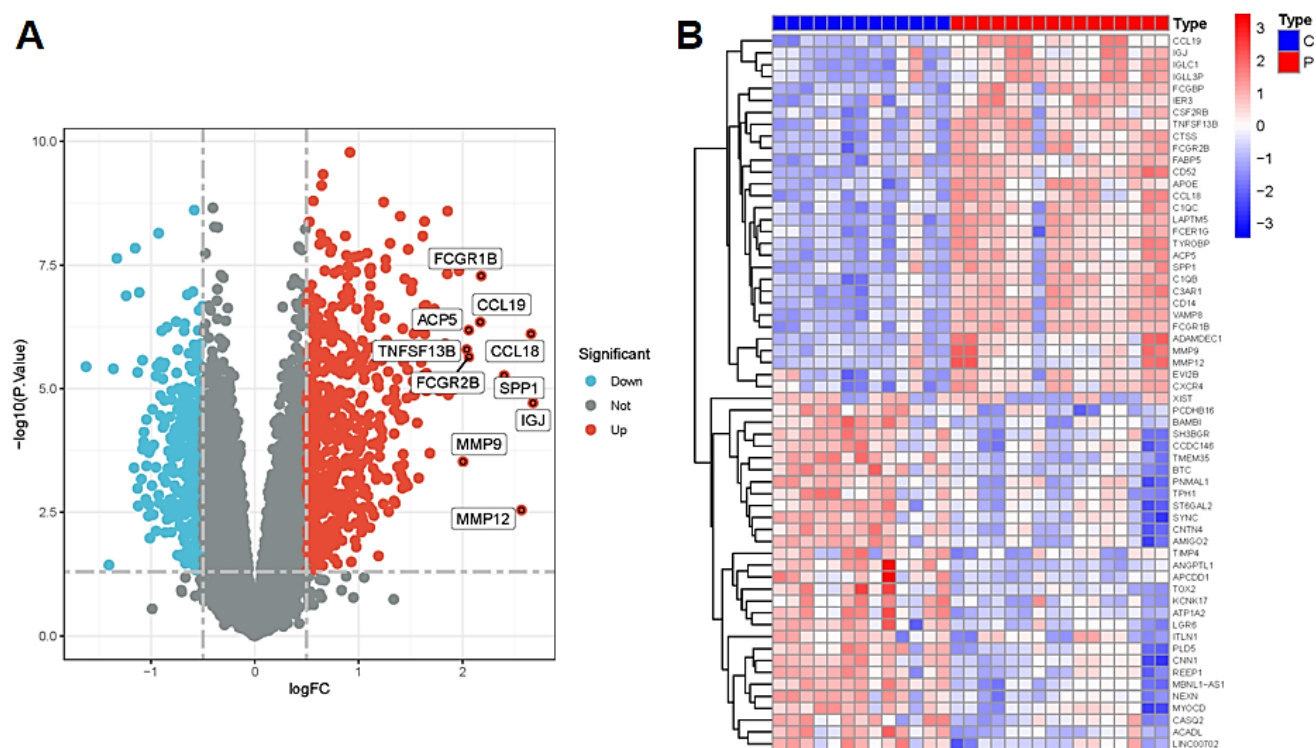


Figure 2. (A) A volcano plot of differentially expressed genes; (B) Heat map of the atherosclerosis and normal groups.

Note: (●) Down; (●) Not; (●) Up; (■) Type C; (■) Type P.

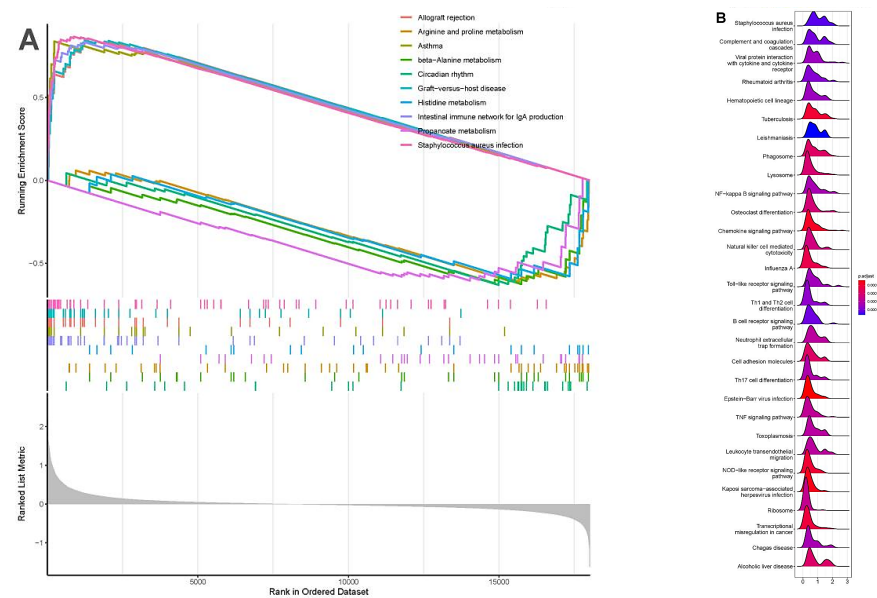


### GSEA analysis

To better understand the role of differentially expressed genes in atherosclerosis, we analyzed atherosclerosis tissues using GSEA. Circadian rhythm, beta-alanine metabolism, arginine and proline metabolism, propanoate metabolism, and histidine metabolism were enriched in the advanced atherosclerosis samples. In contrast, graft-versus-host disease, the intestinal immune network for IgA production, asthma, allograft rejection, and *Staphylococcus aureus* infection were mainly enriched in the early atherosclerosis samples (Figures 3A-3B).

**Figure 3.** (A) GSEA enrichment pathways. (B) The ridge plot of the GSEA.

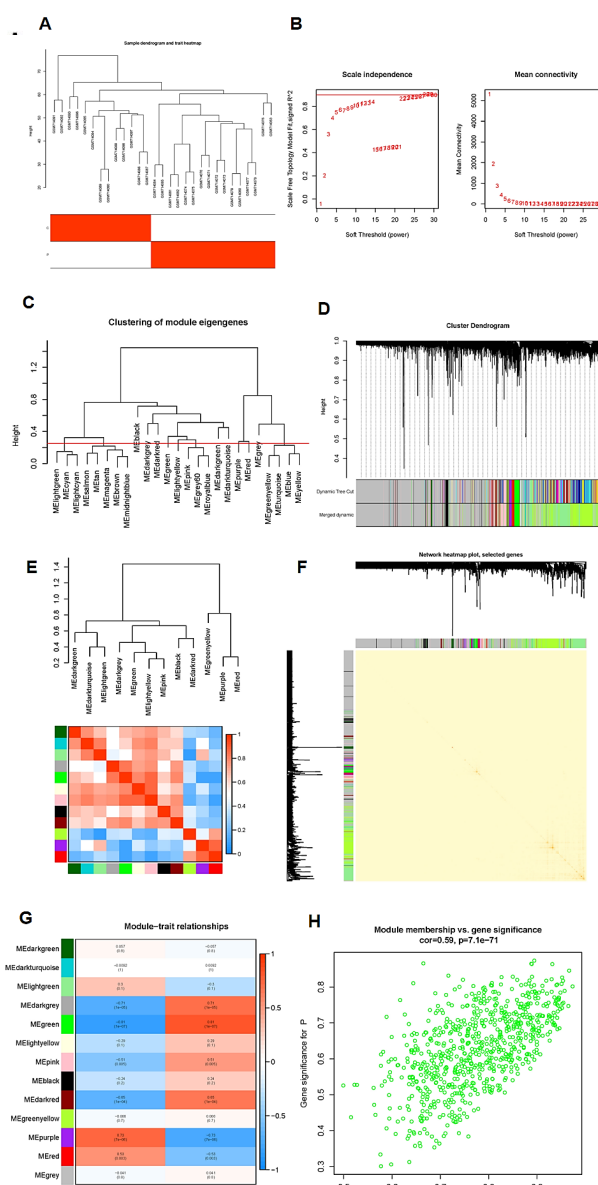
**Note:** (—) Allograft rejection; (—) Arginine and proline metabolism; (—) Asthma; (—) Beta-alanine metabolism; (—) Circadian rhythm; (—) Graft-versus-host disease; (—) Histidine metabolism; (—) Intestinal immune network for IgA production; (—) Propanoate metabolism and (—) *Staphylococcus aureus* infection.



### WGCNA

We clustered 13 early and 16 late samples from the GSE28829 dataset, as shown in Figure 4A. The soft threshold was set to 7 when  $R^2 > 0.9$  and the average connectivity was high (Figure 4B). We set the clustering height to 0.25 to combine the strongly associated modules, as shown in Figures 4C-4D. There was no obvious association between the intermodule correlations (Figure 4E), and the transcriptional correlation analysis within the modules indicated no substantial connection between the modules (Figure 4F). We investigated the association between modules and clinical conditions and found that the green module was positively associated with advanced atherosclerosis ( $r=0.81$ ,  $p=1e-07e$ ) and negatively correlated with normal atherosclerosis ( $r=-0.81$ ,  $p=1e-07e$ ) and negatively correlated with normal atherosclerosis ( $r=-0.81$ ,  $p=1e-07e$ ) (Figure 4G). The findings revealed that the green module was strongly associated with atherosclerosis in the scatter plot of MM vs. GS (Figure 4H). We will continue to analyze the genes in the module.

**Figure 4.** Conducting WGCNA. A) Clustering dendrogram of two sample groups; B) Soft threshold analysis; C) A cutoff value of 0.25 was used for the combination of similar modules; D) Display of the primary and combined modules; E) Module gene correlation; red indicates a positive association and blue indicates a negative association; F) Module gene clustering tree diagram; G) Red indicates a positive correlation between modules and traits and blue indicates a negative correlation; H) MM vs. GS scatter plot of atherosclerosis.

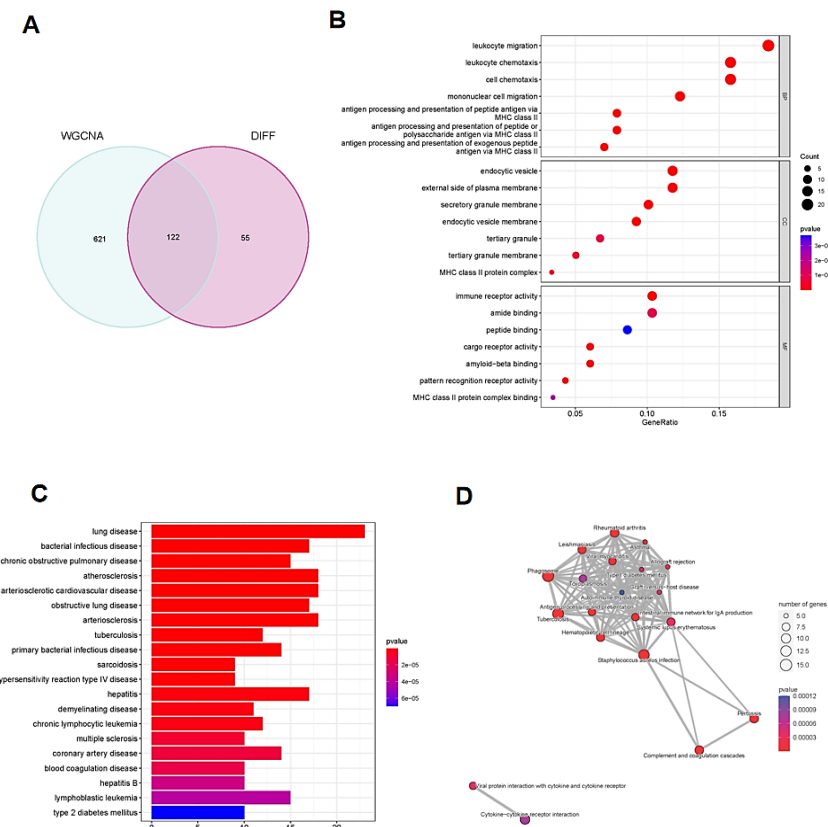


### Functional notes and pathway enrichment

We extracted differing genes within the module using the Venn plot, resulting in 122 genes (Figure 5A). Functional annotation and pathway enrichment were performed to analyze biological functions further.

GO enrichment analysis showed that leukocyte-mediated immunity, leukocyte migration, endocytic vesicles, the external side of the plasma membrane, immune receptor activity and amide binding were closely related to candidate DEGs (Figure 5B). Differentially Ordered (DO) analysis revealed that candidate DEGs were related to hepatitis, atherosclerosis, arteriosclerosis, lung disease and arteriosclerotic cardiovascular disease (Figure 5C). KEGG analysis showed an association with rheumatoid arthritis, cytokine-cytokine receptor interaction, *Staphylococcus aureus* infection, tuberculosis, and phagosome (Figure 5D).

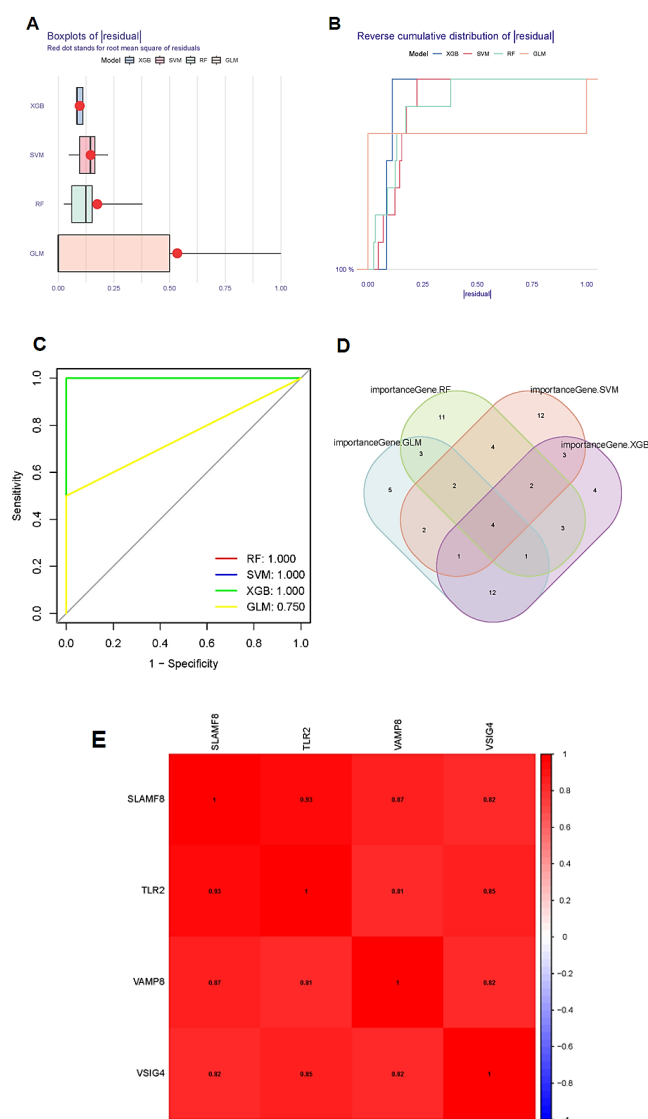
**Figure 5.** (A) Venn diagram of the intersection of WGCNA module genes and DEGs. (B) GO analysis. (C) DO analysis. (D) KEGG analysis.



### Selection of signature genes

We utilized four machine learning algorithms (XGBoost, RandomForest, SVM-REF, and GLM) to identify biomarkers, and each algorithm selected 30 core genes from module difference genes. All four algorithms demonstrated superior sensitivity and specificity (Figure 6A-6C). A Venn plot was used to identify four genes (*SLAMF8*, *TLR2*, *VAMP8* and *VSIG4*) that overlapped in all four methods, indicating significant functional similarities (Figure 6D-6E).

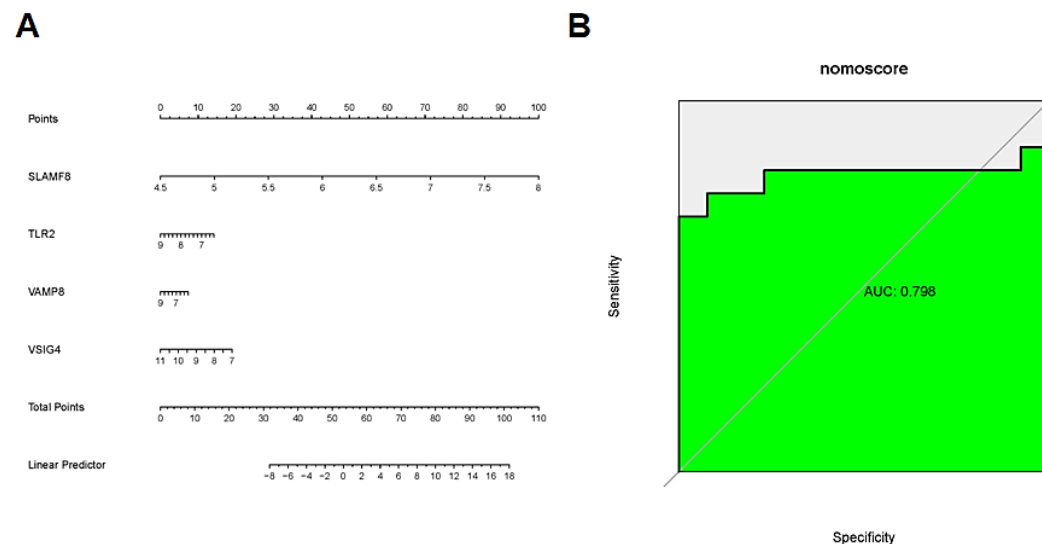
**Figure 6.** (A) Box plots of residual; (B) Reverse cumulative distribution of residual; (C) The efficiency of the four models; (D) The four models screened for overlapping genes; (E) Correlation of overlapping genes.



### Construction and testing of the nomogram for atherosclerosis

We developed an atherosclerosis diagnostic nomogram using rms packages (Figure 7A) based on hub genes (*SLAMF8*, *TLR2*, *VAMP8* and *VSIG4*) and assessed its forecasting ability through ROC curves. The ROC curve demonstrated the high accuracy of the nomogram model (Figure 7B).

Figure 7. (A) The bar graph forecasts the onset of atherosclerosis; (B) Effectiveness of predictive models.



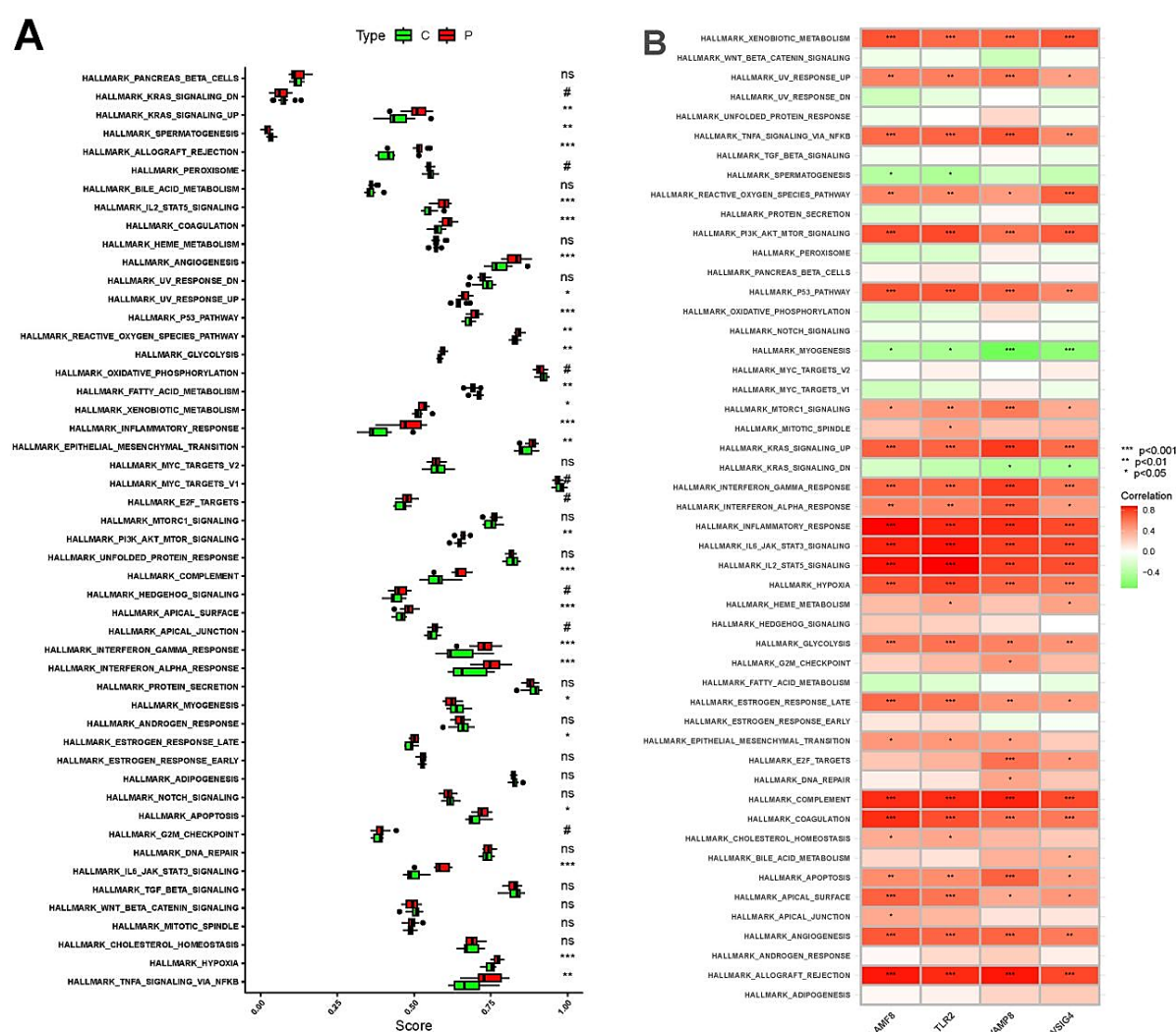
### ssGSEA hallmark analysis in the atherosclerotic group and control group

We used ssGSEA to perform hallmark analysis in patients with advanced-stage and early-stage atherosclerosis. The results showed that most of the hallmark pathways were enriched with more DEGs in the advanced group. The top three pathways with significant enrichment differences were HALLMARK\_WNT\_BETA\_CATENIN\_SIGNALING, HALLMARK\_APICAL\_JUNCTION and HALLMARK\_DNA\_REPAIR (Figure 8A). It suggests that genes in the advanced group are predominantly upregulated by activating WNT signaling through the accumulation of beta-catenin CTNNB1, encoding components of the apical junction complex and involved in DNA repair. Figure 6E shows the high degree of consistency in the regulation of the body by the four signature genes complementing each other.

Therefore, *SLAMF8*, *TLR2*, *VAMP8* and *VSIG4* were upregulated by activating notch signaling, hedgehog signaling, and the mTORC1 complex, downregulated by KRAS activation, upregulated during the formation of blood vessels (angiogenesis) (Figure 8B). And the unfolded protein response, a cellular stress response related to the endoplasmic reticulum, regulated by MYC and NF-κB in response to TNF, is involved in the metabolism of heme and erythroblast differentiation (Figure 8B).

Figure 8. (A) Comparison of hallmark scores between the atherosclerosis group and the control group and (B) Correlation of overlapping genes with hallmark.

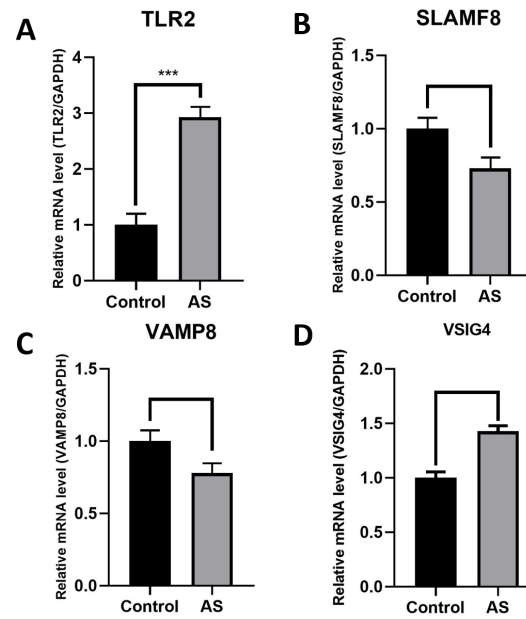
Note: (■) Type C; (■) Type P.



### Quantitative RT-PCR

We used qRT-PCR to detect the transcripts of *SLAMF8*, *TLR2*, *VAMP8* and *VSIG4* in the peripheral blood of atherosclerotic patients and controls. *TLR2* expression was markedly higher in the atherosclerosis subjects compared to control subjects (Figure 9). *TLR2* could serve as a diagnostic marker for atherosclerosis and is important for its prediction and diagnosis, consistent with the results of preliminary bioinformatics analysis.

Figure 9. Expression of overlapping genes in the peripheral blood of patients with atherosclerosis compared with controls.

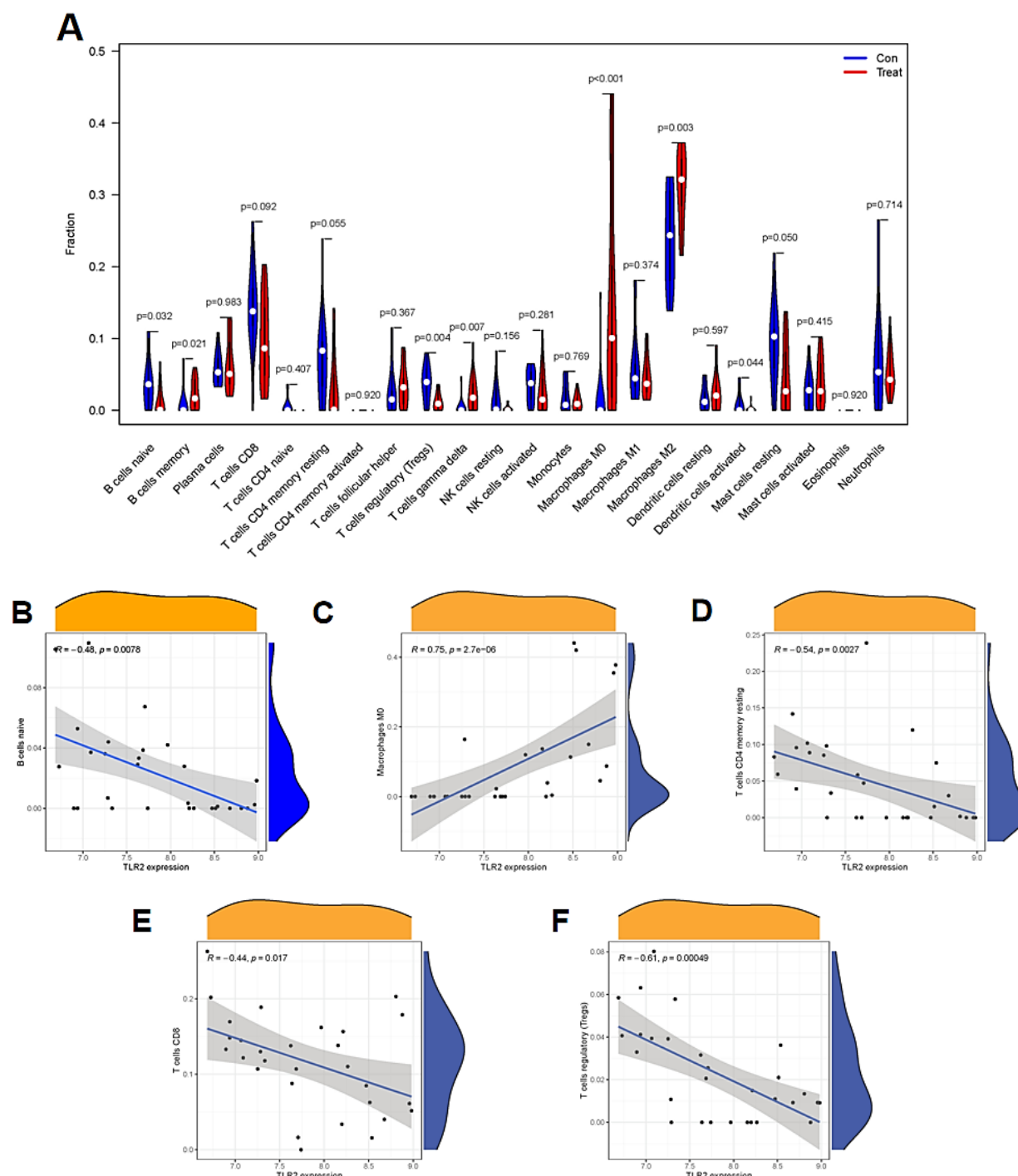


Immuno-infiltrate analysis

Immune infiltration analysis revealed that the infiltration of naive and memory B cells, regulatory T cells (Tregs), gamma delta T cells, M0 and M2 macrophages and activated dendritic cells were significantly associated with the development of atherosclerosis (Figure 10A). We also performed an immunecorrelation analysis on the signature gene *TLR2*. The results showed that *TLR2* expression was positively associated with the infiltration of CD8 T cells, naive B cells, resting memory CD4 T cells, and regulatory T cells (Tregs) and negatively associated with the infiltration of M0 macrophages (Figure 10B-10F).

Figure 10. (A) Differences in the infiltration of 22 immune cells in normal and atherosclerosis samples; (B-F) Scatter diagram of the association of *TLR2* expression with immune cells.

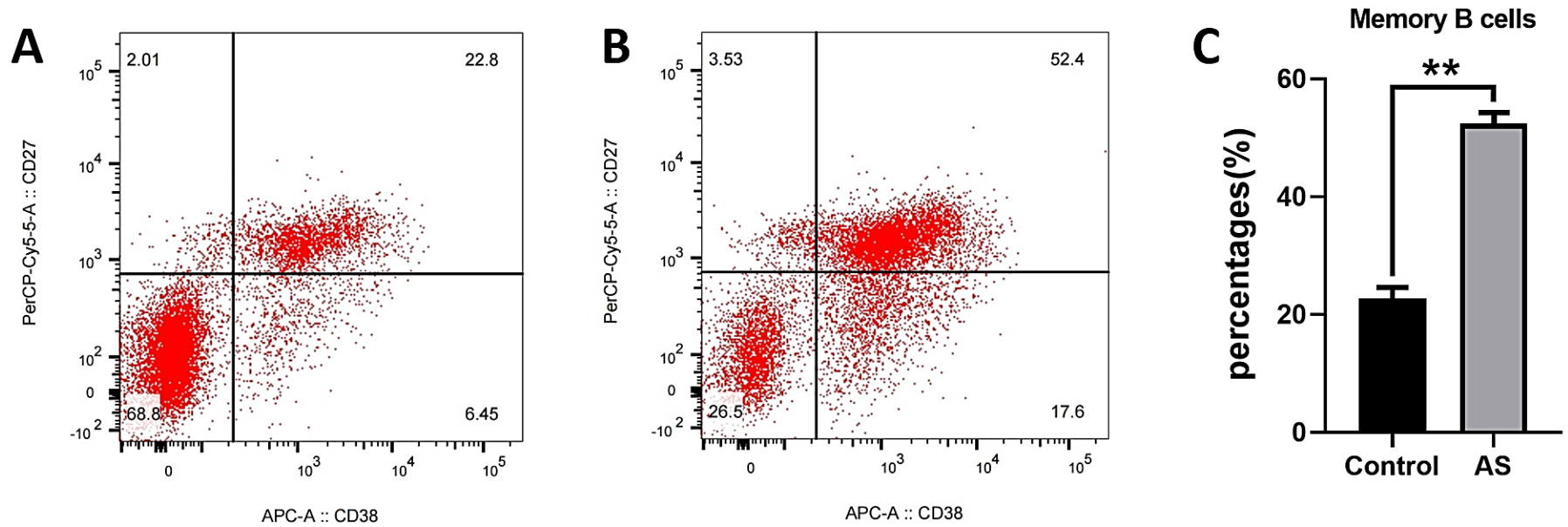
Note: (—) Con; (—) Treat.



Flow cytometry

The number of infiltrating immune cells was detected by flow cytometry analysis. As shown in Figure 11, there was a significant increase in memory B cells in atherosclerosis blood samples. Meanwhile, in the experiment, we found that there was also an increasing trend of M2 macrophages.

**Figure 11.** (A) FACS analysis of memory B cells in the control group; (B) FACS analysis of memory B cells in the atherosclerosis group; (C) Percentage of memory B cells in control and atherosclerosis group.



**Pan-cancer *TLR2* expression**

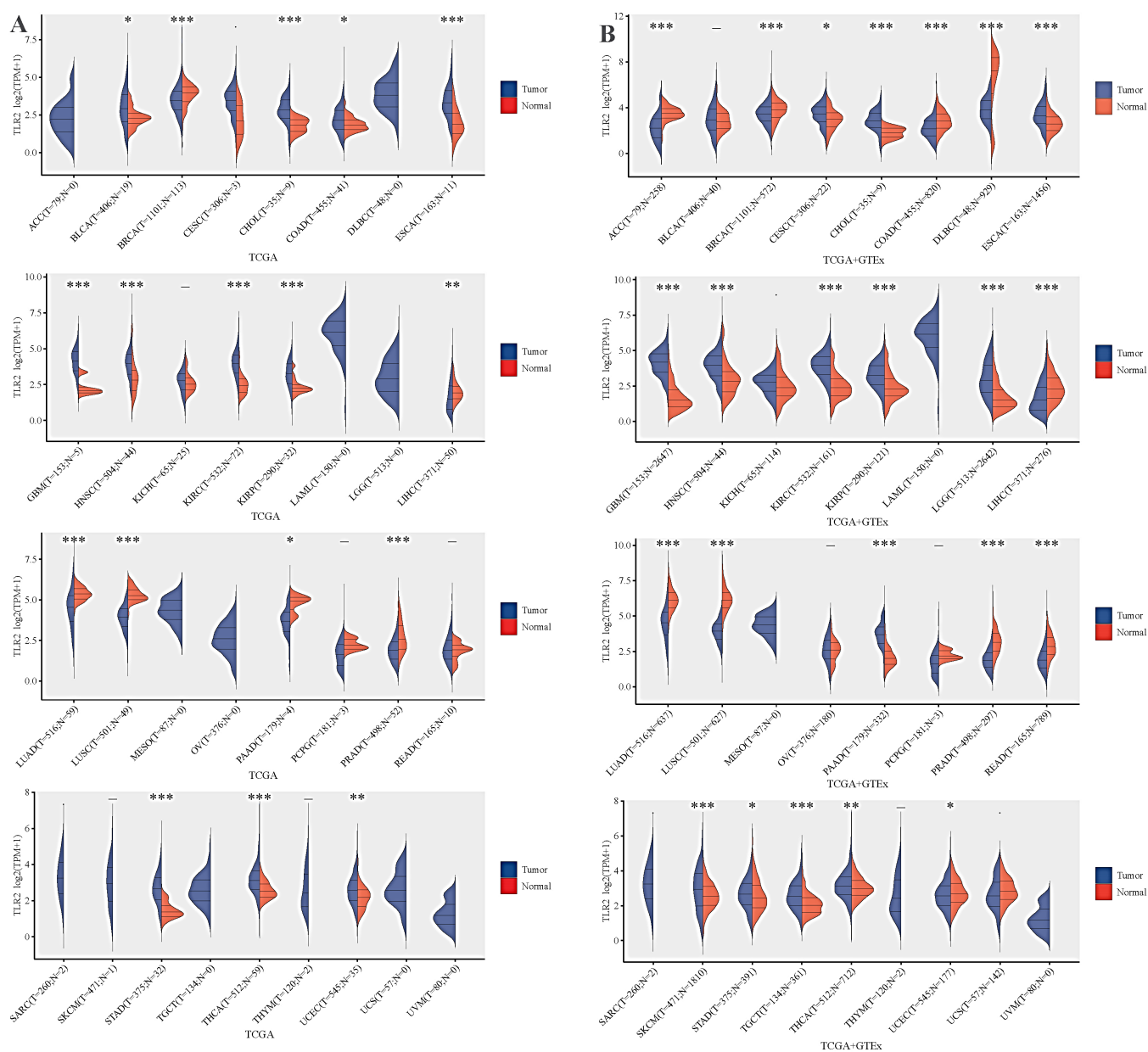
After analyzing *TLR2* expression using PCR and immune infiltration analysis, we delved deeper into the gene's potential impact on immunity and tumor development. Previous research has shown that genetically mediated immune responses can play a role in developing atherosclerosis.

We looked at *TLR2* expression across different tumors in the TCGA and found that it was highly expressed in Bladder Urothelial Carcinoma (BLCA), Cholangiocarcinoma (CHOL) (bile duct cancer), Colon Adenocarcinoma (COAD), Esophageal Carcinoma (ESCA), Glioblastoma Multiforme (GBM), Head and Neck Squamous Cell Carcinoma (HNSC), Kidney Renal Papillary Cell Carcinoma (KIRP), Stomach Adenocarcinoma (STAD), Thyroid Carcinoma (THCA) and Uterine Corpus Endometrial Carcinoma (UCEC), while being expressed at low levels in Breast Cancer (BRCA), Liver Hepatocellular Carcinoma (LIHC), Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), Pancreatic Adenocarcinoma (PAAD), and (PRAD) Prostate Adenocarcinoma (Figure 12A).

After we simultaneously downloaded normal tissue data from the GTex database, *TLR2* was found to be highly expressed in Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC), CHOL, ESCA, GBM, HNSC, Kidney Renal Clear Cell Carcinoma (KIRC), KIRP, Lower Grade Glioma (LGG), PAAD, Skin Cutaneous Melanoma (SKCM), Stomach Adenocarcinoma (STAD), Testicular Germ Cell Tumors (TGCT), and THCA was expressed at low levels in Adrenocortical Carcinoma (ACC), Breast Cancer (BRCA), Colon Adenocarcinoma (COAD), Diffuse Large B-cell Lymphoma (DLBC), Liver Hepatocellular Carcinoma (LIHC), Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), and Rectum Adenocarcinoma (READ) (Figure 12B).

**Figure 12.** (A) Expression of *TLR2* across cancers in the TCGA database. (B) Expression of *TLR2* in pan-cancer in the TCGA+GTEx datasets.

**Note:** (■) Tumor; (■) Normal.

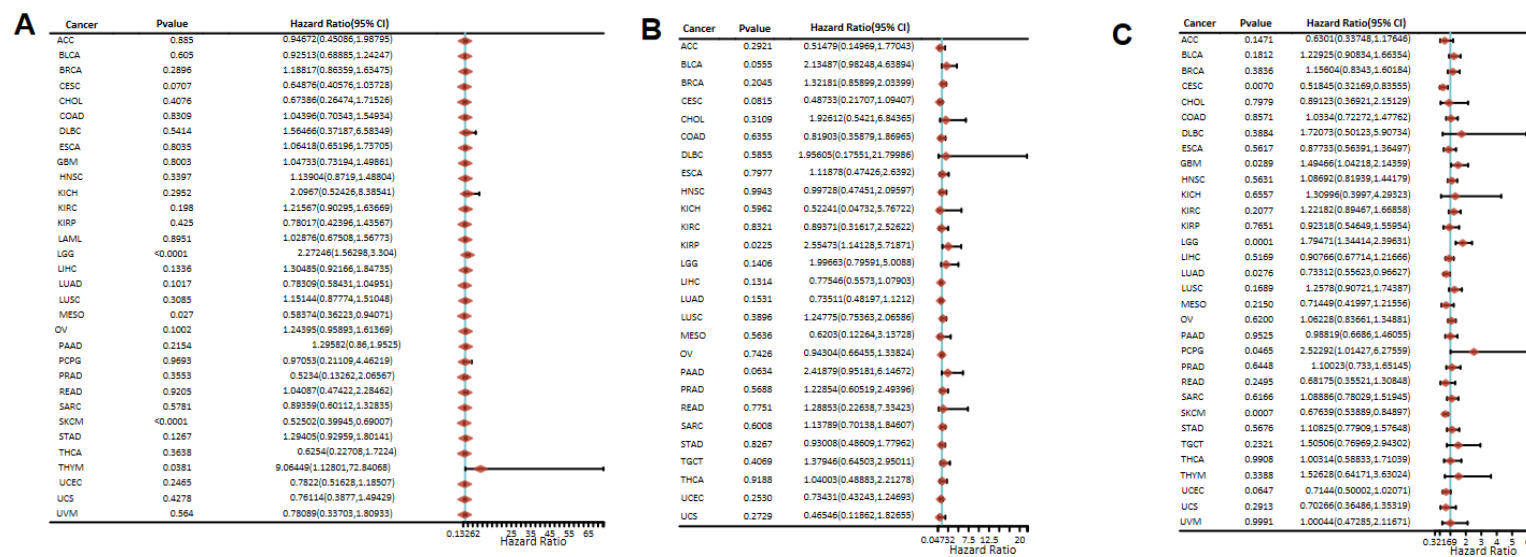




Prognostic value of *TLR2* in pancreatic carcinoma

We conducted a pan-cancer analysis to understand the association between *TLR2* expression and patient prognosis, focusing on Progression-Free Survival (PFS), Disease-Free Survival (DFS), and Overall Survival (OS). Our cox regression analysis of 33 tumors revealed that *TLR2* expression was significantly associated with OS in SKCM (<0.001), LGG (<0.001), Mesothelioma (MESO) (<0.05), and Thymoma (THYM) (<0.05), with *TLR2* being a conservative element of SKCM and MESO, and a risk element of THYM and LGG (Figure 13A). Furthermore, *TLR2* expression was significantly associated with DFS for KIRP (<0.05), where *TLR2* was identified as a risk factor (Figure 13B). We also found that *TLR2* expression was significantly associated with PFS in LGG (<0.001), SKCM (<0.001), CESC (<0.01), LUAD (<0.05), GBM (<0.05), and PCPG (<0.05), with *TLR2* being a conservative element of SKCM, CESC, and LUAD, and a risk element of LGG, GBM, and PCPG (Figure 13C).

**Figure 13.** The correlation between *TLR2* and the prognosis of cancer patients. (A) The cox regression model was used to perform a pan-cancer analysis of the association between *TLR2* expression and OS. (B) The cox regression model was used to perform a pan-cancer analysis of the relationship between *TLR2* expression and DFS. (C) The cox regression model was used to perform a pan-cancer analysis of the relationship between *TLR2* expression and PFS.



DISCUSSION

Atherosclerosis is a widespread inflammatory disease that affects large- sized and medium-sized arteries in the body and is responsible for various cardiovascular diseases, including stroke and ischemic heart disease. Stroke is the second most common cause of death globally and one of the leading causes of disability in adults [12]. A significant proportion of strokes and transient ischemic attacks are associated with atherosclerotic stenosis of the carotid arteries, with approximately 20%-30% of cases resulting from this condition. Another 15% of cases are associated with the progressive growth and rupture of atherosclerotic plaques [13,14]. As a result, understanding the pathogenesis and progression of carotid atherosclerosis is critical in preventing and treating cerebrovascular accidents. Next, we explore new gene targets to provide new atherosclerosis treatment and prevention strategies. Recent years have seen significant developments in bioinformatics analysis and molecular mechanisms studies. Through enrichment analysis of cellular components, molecular functions, and biological processes, we can better understand genes biological functions, regulatory mechanisms, and how gene variants affect disease progression and protein expression. WGCNA has also become widely used to screen for key gene modules highly correlated with phenotypes [15]. In the context of carotid atherosclerosis, WGCNA has started to study genes and mechanisms, although further studies in immune infiltration and functional analysis are still required [16,17]. Therefore, the goal is to establish a reliable evaluation system by combining WGCNA and machine learning techniques to validate the core genes and molecular pathways involved in patients with carotid atherosclerosis. It may lead to identifying potential therapeutic targets for carotid atherosclerotic plaques. Among the 177 DEGs that were filtered, 154 genes were upregulated, and 23 genes were downregulated. Gene set enrichment analysis of target DEGs showed they were primarily associated with propanoate metabolism, histidine metabolism, beta-alanine metabolism, allograft rejection, and the intestinal immune network for IgA production.

WGCNA analysis was performed on 29 samples, and 13 modules were clustered. GO and DO functional annotation and KEGG signaling pathway enrichment analysis of differential module genes showed that they were mainly associated with complement and coagulation cascades, phagosome, cytokine-cytokine receptor interaction, positive regulation of cytokine production, leukocyte-mediated immunity, leukocyte migration, arteriosclerosis and hepatitis. The development of atherosclerosis is due to the increase in cytokines, inflammatory cytokines, and chemokines, immune-related inflammatory mechanisms, and the infiltration of immune cells in the endothelium. This process contributes to plaques' formation, rupture, and thrombosis [18-20]. Four machine learning methods were used to filter out four core genes, *SLAMF8*, *TLR2*, *VAMP8* and *VSIG4*, from the target modules. The expression levels of these four core genes differed significantly between patients with early and advanced carotid atherosclerosis, and the diagnostic models constructed were diagnostically effective. Furthermore, these four genes had a high degree of bio-functional identity.

Several studies have demonstrated a link between these four core genes and the pathogenesis of carotid atherosclerosis. The Signaling Lymphocyte Activation Molecule (SLAM) family receptors are expressed in various immune cell types and contain a transmembrane fragment and a cytoplasmic domain with tyrosine motifs. *SLAMF8* is a non-classical receptor of the SLAM family, a cell surface receptor expressed on various immune cells [21]. The deletion of *SLAMF8* can inhibit the activation of the Mitogen-activated protein kinase (MAPK) signaling pathway by downregulating the expression of Toll-like receptor-4 (*TLR4*) and can also inhibit the NF-κB signaling pathway to reduce the secretion of inflammatory cytokines. It has been proven a therapeutic target for hepatitis and rheumatoid arthritis [22,23]. Toll-Like Receptors (TLRs) are pattern recognition receptors that participate in innate immunity and play a critical role in the first line of defense against microbial invasion [24]. Recent studies have demonstrated that blocking Toll-like receptor 2 (*TLR2*) can slow down or reverse the induction of

atherosclerosis by *Chlamydia pneumoniae* infection, and *TLR2* can be used as a new therapeutic target for the prevention and treatment of atherosclerosis [25]. Vesicle-associated membrane protein 8 (*VAMP8*) is a novel oncoprotein that promotes cell proliferation and drug resistance in gliomas, but its role in carotid atherosclerosis is unclear due to insufficient research [26]. Lastly, v-set and immunoglobulin domain containing 4 (*VSIG4*) participate in the pathogenesis of atherosclerosis by promoting the migration of inflammatory cells to the lesion area and inducing chemokines that activate macrophages [27,28].

After analyzing the PCR results, we identified *TLR2* as the gene for further investigation into immune infiltration and pan-cancer analysis. In the advanced stage of carotid atherosclerosis, the immune infiltration analysis showed higher levels of memory B cells, gamma delta T cells, M0 macrophages and M2 macrophages compared to the early stage. Dyslipidemia and immune inflammation are related to atherosclerosis formation and although B cells play an important role in maintaining immune system stability, they can also be harmful in autoimmune diseases and promote atherosclerosis [29-31]. In contrast, an increase in M2 macrophages may lead to the resolution of atherosclerotic inflammation and plaque resolution, potentially a protective and restorative factor in advanced atherosclerosis [32]. Previous research has found that T cells and M0 and M2 macrophages account for the largest proportion of the leukocyte population in advanced unstable plaque, consistent with our study's findings [33,34]. Our study also confirmed previous research showing that *TLR2* is mainly associated with naive B cells, M0 macrophages, memory CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and Tregs. Additionally, *TLR2* is a Toll-like receptor that regulates various immune cell subsets, and its expression has been linked to the progression of atherosclerosis [35]. Interestingly, recent studies have established a link between carotid atherosclerosis and the development of several cancers, including colorectal adenoma [36,37]. Similarly, radiotherapy to the neck in nasopharyngeal carcinoma patients has increased the risk of carotid atherosclerosis [38]. Our next step is to explore the role of *TLR2*, a core gene in carotid atherosclerosis, in different types of cancer. Our pan-cancer analysis found that *TLR2* was a prognostic protective factor for SKCM and MESO but a risk factor for THYM and Lower Grade Glioma (LGG). Although previous research has shown that *TLR2*-mediated innate immunity affects the development of various tumors, its expression and relationship to the prognosis of specific tumors require further investigation. This research could help identify therapeutic targets and prognostic markers.

## CONCLUSION

In this study, we conducted a comprehensive analysis of the immune profile and pan-cancer relevance of the core genes and after applying a series of algorithms, enrichment and pathway analyses, we identified *TLR2* as a hub gene. We investigated the immune activities and pathways involved and their impact on the prognosis of different tumors. Based on our findings, we concluded that *TLR2* is a potential intervention target for both carotid atherosclerosis and multiple tumors and may become a promising area of research for developing therapeutic strategies for cardiovascular and cerebrovascular diseases and cancer.

## ACKNOWLEDGEMENTS

This study was supported by the Shanxi Province Graduate Education Innovation Project (2022Y737), the Basic Research Program of Shanxi Province (202203021212010), and the Youth Start-up Fund of Heping Hospital affiliated to Changzhi Medical College (HPYJ202225).

## DECLARATIONS

The public data that support the findings of this study are available from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), and the experimental data are available from the corresponding author upon reasonable request.

## REFERENCES

1. Napoli C, et al. Primary prevention of atherosclerosis: A clinical challenge for reversing epigenetic mechanisms? *Circulation*. 2012;125:2363-2373.
2. Gokaldas R, et al. Carotid stenosis: From diagnosis to management, where do we stand? *Curr Atheroscler Rep*. 2015;17:480.
3. Wolf D, et al. Immunity and inflammation in atherosclerosis. *Circ Res*. 2019;124:315-327.
4. Zhan Z, et al. Identification and evaluation of key biomarkers of acute myocardial infarction by machine learning. *Intel Comput Theo Appl*. 2022;13394:103-115.
5. Kolluri S, et al. Machine learning and artificial intelligence in pharmaceutical research and development: A Review. *AAPS J*. 2022;24:19.
6. Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
7. Langfelder P, et al. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
8. Yu G, et al. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284-287.
9. Ishwaran H, et al. Consistency of random survival forests. *Stat Probab Lett*. 2010;80:1056-1064.
10. Lin X, et al. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2012;910:149-155.
11. Robin X, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
12. Migdalski A, et al. New insight into biology, molecular diagnostics and treatment options of unstable carotid atherosclerotic plaque: A narrative review. *Ann Transl Med*. 2021;9:1207.

13. Carballo-Perich L, et al. Clinical parameters and epigenetic biomarkers of plaque vulnerability in patients with carotid stenosis. *Int J Mol Sci.* 2022;23:5149.
14. Touzé E. Natural history of asymptomatic carotid artery stenosis. *Rev Neurol.* 2008;164:793-800.
15. Zhang T, et al. Gene expression data analysis using Hellinger correlation in Weighted Gene Co-expression Networks (WGCNA). *Comput Struct Biotechnol J.* 2022;20:3851-3863.
16. Chen M, et al. Weighted gene co-expression network analysis identifies crucial genes mediating progression of carotid plaque. *Front Physiol.* 2021;12:601952.
17. Zheng L, et al. Contribution of *FBLN5* to unstable plaques in carotid atherosclerosis via mir128 and mir532-3p based on bioinformatics prediction and validation. *Front Genet.* 2022;13:821650.
18. Loppnow H, et al. Vascular cells contribute to atherosclerosis by cytokine- and innate-immunity-related inflammatory mechanisms. *Innate Immun.* 2008;14:63-87.
19. Sterpetti AV. Inflammatory cytokines and atherosclerotic plaque progression. Therapeutic implications. *Curr Atheroscler Rep.* 2020;22:75.
20. Tedgui A, et al. Cytokines in atherosclerosis: Pathogenic and regulatory pathways. *Physiol Rev.* 2006;86:515-581.
21. Veillette A. Immune regulation by SLAM family receptors and SAP-related adaptors. *Nat Rev Immunol.* 2006;6:56-66.
22. Zeng X, et al. Combined deficiency of *SLAMF8* and *SLAMF9* prevents endotoxin-induced liver inflammation by downregulating *TLR4* expression on macrophages. *Cell Mol Immunol.* 2020;17:153-162.
23. Qin W, et al. Knockout of *SLAMF8* attenuates collagen-induced rheumatoid arthritis in mice through inhibiting *TLR4*/NF- $\kappa$ B signaling pathway. *Int Immunopharmacol.* 2022;107:108644.
24. Aderem A, et al. Toll-like receptors in the induction of the innate immune response. *Nature.* 2000;406:782-787.
25. Miao G, et al. *TLR2*/*CXCR4* coassociation facilitates *Chlamydia pneumoniae* infection-induced atherosclerosis. *Am J Physiol Heart Circ Physiol.* 2020;318:1420-1435.
26. Chen Y, et al. *VAMP8* facilitates cellular proliferation and temozolomide resistance in human glioma cells. *Neuro Oncol.* 2015;17:407-418.
27. Hertle E, et al. The complement system in human cardiometabolic disease. *Mol Immunol.* 2014;61:135-148.
28. Zhao C, et al. Identification of an alveolar macrophage-related core gene set in acute respiratory distress syndrome. *J Inflamm Res.* 2021;14:2353-2361.
29. Libby P, et al. Inflammation in atherosclerosis. *Nature.* 2002;420:868-874.
30. Vodjgani M, et al. Analysis of class-switched memory B cells in patients with common variable immunodeficiency and its clinical implications. *J Investig Allergol Clin Immunol.* 2007;17:321-328.
31. Rodríguez-Bayona B, et al. Decreased frequency and activated phenotype of blood CD27 IgD IgM B lymphocytes is a permanent abnormality in systemic lupus erythematosus patients. *Arthritis Res Ther.* 2010;12:R108.
32. Rahman K, et al. Inflammatory Ly6C<sup>hi</sup> monocytes and their conversion to M2 macrophages drive atherosclerosis regression. *J Clin Invest.* 2017;127:2904-2915.
33. Businaro R. Neuroimmunology of the atherosclerotic plaque: A morphological approach. *J Neuroimmune Pharmacol.* 2013;8:15-27.
34. Adamson S, et al. Phenotypic modulation of macrophages in response to plaque lipids. *Curr Opin Lipidol.* 2011;22:335-342.
35. Li B, et al. Infection and atherosclerosis: TLR-dependent pathways. *Cell Mol Life Sci.* 2020;77:2751-2769.
36. Kim H, et al. Association between colorectal adenoma and carotid atherosclerosis in Korean adults. *Int J Environ Res Public Health.* 2018;15:2762.
37. Tapia-Vieyra JV, et al. Atherosclerosis and cancer; a resemblance with far-reaching implications. *Arch Med Res.* 2017;48:12-26.
38. Yuan C, et al. Association between genetic polymorphisms and carotid atherosclerosis in patients treated with radiotherapy for nasopharyngeal carcinoma. *Radiat Oncol.* 2015;10:39.