**Research Paper** 

## Comprehensive analysis identifies crucial genes associated with immune cells mediating progression of carotid atherosclerotic plaque

Zhen Li<sup>1,\*</sup>, Junhui Liu<sup>1,\*</sup>, Zhichun Liu<sup>1</sup>, Xiaonan Zhu<sup>1</sup>, Rongxin Geng<sup>1</sup>, Rui Ding<sup>1</sup>, Haitao Xu<sup>1,\*</sup>, Shulan Huang<sup>1,\*</sup>

<sup>1</sup>Department of Neurosurgery III, Renmin Hospital of Wuhan University, Wuhan 430060, Hubei Province, P.R. China \*Equal contribution

Correspondence to: Haitao Xu, Shulan Huang; email: <u>xuhaitaorenmin@163.com</u>, <u>https://orcid.org/0000-0003-1880-3736</u>; <u>Huang\_shulan@outlook.com</u>, <u>https://orcid.org/0009-0001-1187-1130</u> Keywords: carotid atherosclerotic plaque, bioinformatics, RNA sequencing, WGCNA, immunity Received: September 29, 2023 Accepted: December 27, 2023 Published: February 20, 2024

**Copyright:** © 2024 Li et al. This is an open access article distributed under the terms of the <u>Creative Commons Attribution</u> <u>License</u> (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Backgrounds: Carotid atherosclerosis is prone to rupture and cause ischemic stroke in advanced stages of development. Our research aims to provide markers for the progression of atherosclerosis and potential targets for its treatment.

Methods: We performed a thorough analysis using various techniques including DEGs, GO/KEGG, xCell, WGCNA, GSEA, and other methods. The gene expression omnibus datasets GSE28829 and GSE43292 were utilized for this comprehensive analysis. The validation datasets employed in this study consisted of GSE41571 and GSE120521 datasets. Finally, we validated PLEK by immunohistochemistry staining in clinical samples.

Results: Using the WGCNA technique, we discovered 636 differentially expressed genes (DEGs) and obtained 12 co-expression modules. Additionally, we discovered two modules that were specifically associated with atherosclerotic plaque. A total of 330 genes that were both present in DEGs and WGCNA results were used to create a protein-protein network in Cytoscape. We used four different algorithms to get the top 10 genes and finally got 6 overlapped genes (TYROBP, ITGB2, ITGAM, PLEK, LCP2, CD86), which are identified by GSE41571 and GSE120521 datasets. Interestingly, the area under curves (AUC) of PLEK is 0.833. Besides, we found PLEK is strongly positively correlated with most lymphocytes and myeloid cells, especially monocytes and macrophages, and negatively correlated with most stromal cells (e.g., neurons, myocytes, and fibroblasts). The expression of PLEK were consistent with the immunohistochemistry results.

Conclusions: Six genes (TYROBP, ITGB2, ITGAM, PLEK, LCP2, CD86) were found to be connected with carotid atherosclerotic plaques and PLEK may be an important biomarker and a potential therapeutic target.

## **INTRODUCTION**

Atherosclerosis is a chronic progressive inflammationrelated disease that occurs in the arteries and is characterized by the formation of plaques on damaged artery walls. It can occur in a variety of arteries throughout the body, most commonly the coronary, carotid, and peripheral arteries. In comparison to early plaques, advanced carotid atherosclerotic plaque has a higher propensity to rupture, resulting in transient ischemic attack or ischemic stroke, making it a significant contributor to global mortality.

Various applications have utilized data mining, such as sequencing [1], analysis of gene expression using microarrays [2–4], detection of single-nucleotide

polymorphisms [5, 6], and more. WGCNA, a robust technique, can be employed to detect gene coexpression modules, investigate the relationship between the modules and phenotypes, and uncover pivotal genes that govern crucial biological mechanisms [7]. xCell is an innovative approach that utilizes gene signatures to detect 64 different types of immune and stromal cells through comprehensive *in silico* investigations [8]. This method allows accurate calculation of the relative proportions of immune and stromal cell composition in a lesion sample.

Despite the fact that research has shown that persistent inflammation can contribute to the development of atherosclerosis, the precise mechanism behind atherosclerosis remains unknown. Moreover, there have been fewer investigations carried out to examine the association between genes and immune cells in big data related to atherosclerosis.

To examine the immunocyte infiltration microenvironment and determine the important genes linked to carotid atherosclerotic plaque, we employed GSE28829 and GSE43292 as the training datasets for this investigation. The validated hub genes were then further confirmed in GSE41571 and GSE120521, as well as through immunohistochemistry (IHC) analysis of clinical samples.

## **MATERIALS AND METHODS**

## Data source

GSE28829 [9] and GSE43292 [10] were downloaded from the Gene Expression Omnibus (GEO) database [11] (<u>https://www.ncbi.nlm.nih.gov/geo/</u>). The GSE28829 profile includes 16 advanced atherosclerotic plaques and 13 early atherosclerotic plaques on the GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array. The GSE43292 dataset included 32 atheroma plaques and 32 macroscopically intact tissues on the GPL6244 platform, specifically using the HuGene-1\_0-st Affymetrix Human Gene 1.0 ST Array in its transcript (gene) version. The training dataset was identified as GSE28829 and GSE43292.

The Department of Pathology, People's Hospital of Wuhan University provided 23 carotid atherosclerotic plaques and 11 carotid intact tissues. These samples were collected between 2017 and 2022.

## Data preprocessing and DEG screening

Screening and preprocessing of data for differential expression gene (DEG) analysis, the series matrix

files of GES28829 and GSE43292 were annotated using data tables from GPL570 and GPL6244, respectively. This annotation involved replacing the probe name with the official gene symbol to obtain the gene expression matrix. Then, the gene expression matrix of GSE28829 and GSE43292 were combined in RStudio [12-14]. The "sva" R package was used to remove batch effects. Besides, the screening of differentially expressed genes (DEGs) was performed using "limma" R package. DEG screening was performed using thresholds of |Fc (Fold-change)|>1.5 and adjusted P<0.05. Significantly regulated transcripts and genes were visualized in a volcano plot created using the ggplot2 package in the R software. With the "pheatmap" package, the top 50 upregulated and downregulated genes were visualized.

## Functional enrichment analysis

The above up- and down-regulated genes were imported into the Database for Annotation, Visualization, and Integrated Discovery (DAVID, v6.8) [15], official gene symbols were selected as identifiers, and Homo sapiens was selected as a species. Finally, the role of genes and gene products in any organism was obtained from Gene Ontology (GO) [16] and metabolic pathways and gene signaling networks were obtained according to Kyoto Encyclopedia of Genes and Genomes (KEGG) [17]. This study shows the top five results sorted by P-value (P<0.05) from high to low.

## Gene set enrichment analysis

GSEA analysis was performed on the above dataset using GSEA.4.2.3 [18] software. Input plaque grouping data and select c2.cp.kegg.v2023.1.Hs.symbols.gmt to assess the associated pathways and molecular mechanisms. The minimum gene set was set to be 5, the maximum gene set to be 5000, and 1000 were resampled. The screening criteria included a normalized enrichment score>1.5 and a false discovery rate (FDR)<0.01.

## Co-expression network construction by WGCNA

We used Weighted Gene Co-Expression Network Analysis (WGCNA) built with the R package "WGCNA" [19] to identify co-expression networks. We used gene expression profiles to calculate the MAD (median absolute deviation) of each gene, removing the top 50% of genes with the smallest MAD. Outliers and samples were removed using the GoodSamplesGenes method in the R package WGCNA, and scale-free coexpression networks were further constructed using WGCNA. The pickSoftThreshold function was used to choose the soft threshold power( $\beta$ ) based on the criterion of scale-free topology. Afterwards, the gene coexpression modules were discovered utilizing the method of constructing networks in a single step. Every module contained at least 30 genes. By utilizing the dynamic cutting tree technique, the threshold was established at 0.25.

#### Selecting important modules

Module genes (ME) of every module were calculated using the R ggcorr software package and correlations within the ME between modules were shown. Based on the moduleTraitCor and moduleTraitPvalue algorithms, the two modules with significant correlation (P<0.05) with plaque features and the strongest positive and negative correlations were assumed to be the key modules for further analysis. Genes with module membership greater than 0.8 (MM, ME correlation with gene expression profile) were screened for further analysis.

#### Functional enrichment analysis on selected modules

After the aforementioned screening, the genes within each module were utilized for conducting BP and KEGG enrichment analyses. After adjusting the Pvalue, the first 10 terms of BP enrichment analysis and the first 9 terms of KEGG enrichment analysis were visualized.

## PPI network construction and screening for key genes

After intersecting the differentially expressed genes (DEGs), we proceeded to identify the genes within the modules. Next, the intersecting genes were submitted to the STRING [20] (version 11.5) platform to determine the protein-protein interactions, with a minimum interaction score of 0.6. Then, we imported the obtained network into Cytoscape software [21] for visualization. In addition, we applied the cytohubba plugin [22] to identify key genes by Degree algorithm. The top 10 genes were filtered from the Degree algorithm. Then, we intersected the top 10 genes from 4 different algorithms and obtained 6 hub genes.

## Validation of hub genes

Six genes were screened from the four algorithms. Then, GSE41571 and GSE120521 were combined to validate genes that may be involved in carotid atherosclerotic plaque progression. The diagnostic and discriminatory value of the six hub genes in the plaque and intact tissue groups was assessed using ROC curve [23].

#### Immune cell infiltration analysis

To understand the immune landscapes of the genetic profile of advanced atherosclerotic plaques, we used a new method based on genetic characterization, xCell, to identify 64 cell types by extensive *in silico* analysis. The 64 different cell types were classified into lymphoid, myeloid, stromal, stem cells, and additional categories. Significant immune cells were then identified between plaque and intact tissue samples using the Wilcoxon threshold test [24] at P < 0.05.

#### Relevance analysis of genes and immune cells

The relationship between the expression of the six pivotal genes and the relative proportion of immune cells in plaque and intact tissue samples was illustrated by Spearman's correlation test analysis, respectively. Correlation coefficient values between gene expression and relative proportion of immune cells can indicate strong, weak or no correlation. P < 0.05 was considered a statistically significant difference by unpaired t-test.

#### Immunohistochemistry and immunofluorescence

IHC [25] staining was performed with the antibody PLEK (Proteintech, 66431-1-lg, 1:500) and CD68(Abcam, ab955, 1:200) following the manufacturer's protocol.

Apply the IHC Profiler in ImageJ 1.54d [26] to calculate the percentage contribution of High Positive, Positive, Low Positive and Negative. The staining value of PLEK and CD86 was calculated as 1-Negative (%). Immunofluorescence [27] staining of human carotid atherosclerotic plaques and intima using CD68(Abcam, ab955, 1:500) and CD163 (Abcam, ab182422, 1:1000) antibodies to investigate macrophage infiltration in plaques.

#### Statistical analysis

Use unpaired t test to compare the Negative percentage contribution in GraphPad Prism 8. Other statistical analysis was performed using R 4.2.3. To compare the variations in expression between the groups, the Wilcox test was employed. The Spearman method was utilized to establish the connections between genes and immune cells. P<0.05 was considered as statistically significant difference.

## RESULTS

#### **Identify DEGs**

Using the gene expression matrix merged with GSE28829 and GSE43292 as the exploration dataset,

PCA analysis showed the samples before and after merging (Figure 1A). Using |Fc(Fold-change)|>1.5 as the dividing line, 385 up-regulated genes and 251 down-regulated genes were shown by volcano plots (Figure 1B). The top 10 up-regulated and downregulated DEGs of |Fc(Fold-change)| are listed in Table 1. The top 100 DEGs are represented by heatmap (Figure 1C).

#### **Functional enrichment analysis**

Gene ontology including Biological Function (BP), Cell Component (CC), and Molecular Function (MF) was performed for both upregulated and downregulated genes. The GO terms that were significantly enriched after adjusting the P-value are displayed in Figure 2A–2C, 2E–2G. During the analysis of BP enrichment, the genes that were upregulated primarily participated in the inflammatory and immune responses, whereas the downregulated genes were associated with cell adhesion and homophilic cell adhesion through adhesion molecules on the plasma membrane. The CC term that showed the highest upregulation was 'extracellular region', whereas 'Z disc' exhibited the most significant downregulation. The molecular function terms indicated an increase in the activity of signaling receptors and chemokines. On the other hand, the molecule function showed a decrease in the expression of actin binding and structural constituent of muscle.

Furthermore, we performed KEGG analysis to demonstrate the pathway of differentially expressed genes (Figure 2D). Significant enrichment was observed in the upregulated pathways, including the B cell receptor, chemokine, and NF-kappa B signaling pathway. Not surprisingly, the downregulated genes exhibited enrichment in the contraction of vascular smooth muscle, as well as the cGMP-PKG and cAMP signaling pathway (Figure 2H).

#### Gene set enrichment analysis (GSEA)

GSEA, a computational method, evaluates the statistical significance of a priori customized gene sets, in contrast to traditional enrichment analysis that only considers a single gene. The top five pathways up- and down-regulated by GSEA are presented in Figure 3, respective-ly. The results showed that lysosomal, cytokine-cytokine receptor interaction, and toll-like receptor signaling pathways were upregulated, whereas butanoate metabolism and dilated cardiomyopathy were downregulated.



Figure 1. Principle components analysis (PCA) and analysis of differentially expressed genes (DEGs) between carotid atherosclerotic plaque and intact tissue samples. (A) PCA; (B) volcano plot; (C) heatmap of top 100 DEGs.

Gene symbol	Official full name	Log <sub>2</sub> (fold-change)	Adjusted P-value
Up-regulated			
IGJ	Immunoglobulin J Chain	2.13	7.23E-08
MMP9	Matrix Metallopeptidase 9	1.86	3.60E-07
MMP12	Matrix Metallopeptidase 12	1.84	5.26E-05
FABP4	Fatty Acid Binding Protein 4	1.67	5.41E-06
CHI3L1	Chitinase 3 Like 1	1.61	1.88E-06
FABP5	Fatty Acid Binding Protein 5	1.55	3.48E-08
CD36	CD36 Molecule	1.48	8.17E-06
CCR1	C-C Motif Chemokine Receptor 1	1.43	1.03E-09
MMP7	Matrix Metallopeptidase 7	1.43	3.72E-05
CD52	CD52 Molecule	1.38	4.80E-09
Down-regulated			
CASQ2	Calsequestrin 2	-1.50	1.90E-08
CNTN4	Contactin 4	-1.44	6.55E-09
MYOCD	Myocardin	-1.41	5.97E-08
TPH1	Tryptophan Hydroxylase 1	-1.38	2.43E-06
PLD5	Phospholipase D Family Member 5	-1.33	1.68E-08
HAND2-AS1	HAND2 Antisense RNA 1	-1.30	4.50E-07
ITLN1	Intelectin 1	-1.26	4.42E-04
ACADL	Acyl-CoA Dehydrogenase Long Chain	-1.25	3.59E-09
CNTN3	Contactin 3	-1.24	2.34E-07
FHL5	Four And A Half LIM Domains 5	-1.23	1.90E-08

Table 1. DEGs with top-10 |log<sub>2</sub> (fold-change)| (carotid plaque/intact tissue).

#### **Building a co-expression network**

Construct a co-expression network by enrolling the combined dataset into R. There were no anomalies identified during the process of hierarchical clustering. Figure 4A depicts the construction of a scale-free network through the utilization of soft-thresholds at  $\beta$ =14. Twelve relevant co-expression modules were obtained (Figure 4B).

#### Selection and enrichment analysis of key modules

In order to determine the main modules associated with sample characteristics, we developed a relationship between modules and traits (Figure 4C). We observed two co-expression modules of genes that are strongly associated with the atherosclerotic process. The darkgreen module, consisting of 2364 genes, was positively associated with sample traits in atherosclerotic



**Figure 2.** Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) of upregulated (A–D) and downregulated DEGs (E–H). (A, E) biological process (BP) analysis, (B, F) cellular components (CC) analysis, (C, G) molecular function (MF) analysis, (D, H) KEGG pathway analysis.

plaques. In contrast, the 1886 genes found in the green module were inversely associated with advanced atherosclerotic plaques. To explore the biological functions of the key modules associated with sample traits, we also performed GO and KEGG enrichments. For the darkgreen module, BP enrichment indicated that these genes mainly participated in immune system processes, immune responses, cell activation, and leukocyte activation. KEGG analysis suggested that lysosomes might be involved in these pathways (Figure 5A, 5B). The results from GO and KEGG were highly similar to our previous enrichment analysis. The GO analysis of the green module indicated enrichment in system development, morphogenesis of anatomical structures, and development of the nervous system (Figure 5C). The KEGG findings indicated the involvement of the actin cytoskeleton, tight junctions, and regulation of vascular smooth muscle contraction, which was consistent with the results for downregulated DEGs (Figure 5D).

## Protein-protein interactions (PPI) and screening of potential genes

Initially, we filtered out these genes from two modules with MM>0.8, indicating that these genes exhibit a greater level of connectivity within the modules. Next, we compared the DEGs with the

highly connected genes. There was an overlap of 330 genes (Figure 6A). We then imported these 330 genes into the STRING database and constructed a PPI network, ensuring a minimum required interaction score of 0.4 (Figure 6B). The results were visualized in Cytoscape and included 333 nodes and 1727 edges. The Degree algorithm utilizing the Cytoscape plugin cytohubba displays all PPI networks (Figure 6C) and the top 10 genes (Figure 6D). The top ten hub genes from the cytohubba plugin were obtained by four algorithms: MCC, MNC, Degree, and EPC. The overlapping hub genes in the four algorithms were verified by a Venn diagram (Figure 6E).

#### Validation hub genes in public databases

TYROBP (Transmembrane Immune Signaling Adaptor), ITGB2 (Integrin Subunit Beta 2), ITGAM (Integrin Subunit Alpha M), PLEK (Pleckstrin), LCP2 (Lymphocyte Cytoplasmic Protein 2), and CD86 (CD86 molecule) were the overlapping hub genes. After extracting the expression matrix profiles, the validation datasets used were GSE41571 and GSE120521. ROC analysis was conducted, revealing the potential diagnostic significance of all these essential genes. Figure 7 shows the ROC profiles and gene expression of TYROBP, ITGB2, ITGAM, PLEK, LCP2 and CD86.



Figure 3. Gene set enrichment analysis (GSEA) of the top 5 upregulated and downregulated GSEA pathways. (A) lysosome; (B) cytokine-cytokine-receptor interaction; (C) NOD-like-receptor signaling pathway; (D) TOLL-like-receptor signaling pathway; (E) B-cellreceptor signaling pathway; (F) butanoate metabolism; (G) dilated cardiomyopathy; (H) propanoate metabolism; (I) hypertrophic cardiomyopathy HCM; (J) tyrosine metabolism.



**Figure 4. Construction of the co-expression network for carotid atherosclerotic plaque.** (**A**) Identification of soft threshold power (β); (**B**) Clustering dendrogram to find co-expression modules; (**C**) The identification of key modules related to sample traits.



Figure 5. The enrichment analysis of key modules. (A, B) BP and KEGG analysis of the darkgreen module; (C, D) BP and KEGG analysis of green module.



**Figure 6. Screening out potential genes.** (A) The overlapping of DEGs and key module genes from WGCNA; (B) Protein-protein interaction (PPI) networks; (C, D) All the genes and the top 10 genes calculated by the Degree algorithm of cytoHubba; (E) The overlapped hub genes from four different algorithms.



Figure 7. ROC curves and statistic of expression for TYROBP, ITGB2, ITGAM, PLEK, LCP2 and CD86. (A) The AUC for TYROBP was 0.933. (B) The AUC for ITGB2 was 0.900. (C) The AUC for ITGAM was 0.722. (D) The AUC for PLEK was 0.833. (E) The AUC for LCP2 was 0.844. (F) The AUC for CD86 was 0.900.

#### Immune cell infiltration analysis

The xCell includes 64 immune cells and stromal cells for quantitative assessment of enrichment of these cells. Figure 8 displays the types of cells and their corresponding enrichment scores. In advanced atherosclerotic plaque, the xCell scores indicated an upregulation of immunity-related cells, particularly the macrophage cells, in both lymphoid and myeloid cells (Figure 8A, 8B). On the other hand, the stromal cells exhibited a decrease in expression levels (Figure 8C). The difference in the total immune score and the microenvironmental score demonstrated immune infiltration in advanced atherosclerotic plaques (Figure 8D).

## Correlation analysis of hub genes and immune cells

To demonstrate and exhibit the connection between hub genes and immune cells in advanced plaque and intact tissue samples (Figure 9), a Spearson test was conducted for correlation analysis. The results showed that CD86, ITGAM, ITGB2, LCP2, PLEK and TYROBP are strongly positively correlated with most lymphocytes and myeloid cells, especially monocytes and macrophages. In contrast, the six hub genes are



negatively correlated with most stromal cells, such as neurons, myocytes, fibroblasts and chondrocytes. As shown, PLEK has a close correlation with macrophages (R=0.85, p<2.2e-16), macrophages M1(R=0.85, p<2.2e-16), macrophages M2(R=0.73, p<2.2e-16).

# IHC validation of PLEK importance in carotid atherosclerotic plaque

We then identified the PLEK and CD68 expression level in 34 samples, including 23 carotid atherosclerotic plaques and 11 intact tissue samples. IHC experiment validated that carotid plaque showed higher expression of PLEK and CD68 compared to intact tissue samples (Figure 10A, 10B). Besides, we validated that PLEK has a close correlation with CD68(R=0.66, p=2.3e-05) (Figure 10C). Compared with intact tissue, Immunofluorescence staining showed increased infiltration of both CD68+ macrophages and CD163+ macrophages in carotid atherosclerotic plaques (Figure 10D).

## DISCUSSION

By combining the datasets GSE28829 and GSE43292, we identified a total of 385 genes that were upregulated and 251 genes that were downregulated in plaque



**Figure 8. xCell analysis.** (A) The enrichment scores of lymphoid cells; (B) Enrichment scores of myeloid cells; (C) Enrichment scores of stromal cells; (D) Total enrichment scores of the immune and stromal microenvironment. Significance level was denoted by \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.001.

samples during our study. DEGs screening in previous research has been conducted using either the microarray dataset GSE28829 or GSE43292. DEG results vary due to the use of various methods and criteria. Liu and his colleagues acquired a total of 758 genes that exhibited differential expression by utilizing GSE28829, meeting the criteria of FDR<0.05 and |log2FC|>0.58. The PPI network shows that ITGAM and ACTN2 have the greatest degree [28]. Guiming Wang et al. screened 513 upregulated genes and 373 downregulated genes. The PPI network constructed with these DEGs included 35 key nodes with degrees≥20, among which SYK, LYN and PIK3CG were the three highest [29]. Julong Guo et al. analyzed GSE41571, GSE120521, E-MTAB-2055 and one non-coding RNA dataset (GSE111794) to discover genetic molecules linked to histologically unstable carotid atherosclerotic plaques and found 10 hub genes. Among them, upregulated genes included HCK, C1QC, CD14, FCER1G, LCP1 and RAC2, while TPM1, MYH10, PLS3 and FMOD were found to be downregulated [30].

Function analysis involved genes that were upregulated and downregulated, respectively. In the biological

process of gene ontology, "inflammatory response" and "immune response" were found to be enriched with upregulated genes, while the "cell adhesion", "homophilic cell adhesion via plasma membrane adhesion molecules" and "cell-matrix adhesion" were involved in down-regulation of genes. KEGG pathway analysis further indicated up-regulation of genes of immunityrelated pathways whereas "the vascular smooth muscle contraction" was downregulated. Extensive experimenttal and clinical data now confirm that atherosclerosis is a persistent inflammatory condition that typically remains stable until a disruption in the arterial surface's integrity takes place, like endothelial erosion and plaque rupture [31]. In addition, experience from genome-wide association studies, advanced in vivo imaging, transgenic lineage tracing mice, and clinical interventional studies suggests that both innate and adaptive immune mechanisms can accelerate or inhibit atherosclerosis [32]. The role of cell adhesion is crucial in the regulation of cell migration and proliferation, particularly for VSMC in atherosclerosis. Certain initial research has indicated that the attachment of cells played a role in maintaining the health of blood vessels and safeguarding them from intraplaque hemorrhaging [33]. Hence, the compromised attachment of



**Figure 9. Correlation between gene expressions and the relative percentages of 64 cell types.** (A) The heatmap of correlation between six hub genes and lymphoid cells, myeloid cells and stomal cells. (B–E) Scatterplots illustrate the exact relationship between the PLEK expression and the relative proportion of macrophages M0(R=0.85, p<2.2e-16), macrophages M1(R=0.85, p<2.2e-16), macrophages M2(R=0.73, p<2.2e-16) and immuneScore (R=0.88, p<2.2e-16).

vascular smooth muscle cells could potentially contribute to the advancement of plaques.

The results of GSEA analysis were the same as those of KEGG analysis. In our GSEA results, lysosomal processes were upregulated. It is indicated that advanced atherosclerotic plaques exhibit autophagy mediated by lysosomes [34]. Additionally, lysosomal enzymes played a role in the destabilization of atherosclerotic plaques, including cathepsin S [35]. Thus, upregulation of the lysosome pathway may be involved in plaque progression and lead to its eventual rupture. Interestingly, butanoate, propanoate, and tyrosine metabolism were enriched in the downregulated pathway. Short-chain fatty acids (propionate and butyrate) have been reported to be associated with energy metabolism. For example, propionate and butyrate, the major metabolites of dietary fiber, are major products of bacterial metabolism and important sources of energy [36]. It has been previously reported that acetate, butyrate, and propionate play an important role in atherosclerosis by modulating Treg cell production and inhibiting histone deacetylases (HDACs) [37, 38]. Adhesion molecules are known to promote adhesion between leukocytes and endothelial cells. Li et al. found that butyrate and propionate reduced the expression of vascular cell adhesion molecule-1(VCAM-1) [39]. Therefore, modulation of short-chain fatty acid metabolism could serve as an innovative approach to impede the formation of atherosclerotic plaques.

We then constructed a WGCNA co-expression network to mine the set of co-expression genes associated with the sample trait "Plaque". The results revealed that two modules (darkgreen and green) were strongly associated with advanced plaques. The biological roles and KEGG pathways of the modules aligned with the DEGs. The darkgreen module was also enriched for several other immune-related biological functions, such as "leukocyte activation", "leukocyte mediated immunity" and KEGG pathways like "leukocyte transendothelial migration". The process of leukocyte transendothelial migration (LTEM) plays a crucial role in initiating an inflammatory immune response and sustaining chronic inflammation. Atherosclerosis leads to unregulated movement of leukocyte and leakage of blood vessels due to the compromised protective function of the endothelium [40]. Various leukocytes, especially monocytes, transit through leaky vessels to the subendothelium, polarize into macrophages, and contribute to the development of atherosclerosis.

We used the overlapped genes between DEGs and WGCNA to establish a protein-protein interaction



Figure 10. Validation of PLEK importance in carotid atherosclerotic plaque. (A) Staining images of PLEK protein expression in carotid atherosclerotic plaque and intact tissue samples. (B) Statistic of PLEK protein expression in carotid atherosclerotic plaque and intact tissue samples. (C) The scatterplot of correlation between PLEK and CD68. (D) Macrophage infiltration of plaque versus intact tissue by immunofluorescence.

network. We used four different algorithms to get the top 10 genes and finally got 6 overlapped genes. To validate the differential expression of these six genes, we collected two external datasets (GSE41571 and GSE120521) that included both stable and ruptured human atherosclerotic plaques. ROC curve analysis of the six genes. TYROBP. ITGB2. ITGAM. PLEK. LCP2, CD86 was performed. The AUC was 0.933 for TYROBP, 0.900 for ITGB2, 0.722 for ITGAM, 0.833 for PLEK, 0.844 for LCP2, 0.900 for CD86. Analysis of tissue expression data on porcine atherosclerosis induced by high lipid factors indicates that TYROBP, ITGB2, and ITGAM are key genes influencing the progression of ankylosing spondylitis [41]. It has been shown that upregulation of ITGB2 synergistically affects leukocyte adhesion and migration to the vascular wall, thereby influencing the progression of ankylosing spondylitis [41]. Transcriptome sequencing analysis of ox-LDL-treated endothelial cells showed that ITGAM is a key gene that influences endothelial cell apoptosis and thus promotes atherosclerosis [42]. The PLEK gene encodes the pleckstrin protein, which serves as the primary substrate for platelets and leucocytes' protein kinase C [43]. It plays a crucial role in various processes, such as G protein-coupled receptor signaling pathway, actin cytoskeleton organization, and promoting supramolecular fiber organization. The relationship between PLEK and atherosclerosis is not yet fully understood. However, there is evidence suggesting that PLEK might have a significant impact on chronic inflammatory conditions like CVD, rheumatoid arthritis (RA), and ulcerative colitis (UC) [44]. LCP2 (lymphocyte cytoplasmic protein 2) encodes an adaptor protein that is a substrate for the protein tyrosine kinase pathway activated by the Tcell antigen receptor (TCR) and is thought to play a role in TCR-mediated intracellular signaling. CD86, also known as cluster of differentiation 86, is a protein produced by the CD86 gene. It is found on antigen-presenting cells (APCs) and plays a role in delivering costimulatory signals to T cells [45]. CD86, a biomarker of M1-type macrophages, is markedly expressed in vulnerable arterial plaques.M1type macrophages release ROS and pro-inflammatory cytokines, such as TNF- $\alpha$ , IL-1 $\beta$ , IL-6, and IL-12, which damage endothelial cells and blood vessels and promote atherosclerosis [46]. Additionally, CD86 was observed to be present on fully developed dendritic cells and certain T cells [47], but the exact mechanism remains uncertain.

Given the significant importance of inflammation in carotid atherosclerotic plaques, we proceeded to perform an ssGSEA analysis utilizing 64 different immune cell types. Unsupervised hierarchical clustering

revealed the immune difference between the carotid plaque and intact tissue groups. Therefore, we concluded that the presence of immune cells contributed to the advancement of carotid atherosclerotic lesions. According to the xCell results, there was a significant alteration in the ratio of various immune cells. The majority of immune cells, including B-cells, CD4+ memory T cells, DC, macrophages, and monocytes, exhibited upregulation. On the other hand, the stromal cells like chondrocytes, fibroblasts, osteoblasts, and smooth muscle cells exhibited a decrease in expression. During our analysis of immune infiltration, we observed a substantial prevalence of macrophages, which exhibited a notable rise during the progression of atherosclerosis. The classical model of macrophage polarization describes two opposing phenotypic states: the "classical" pro-inflammatory M1 macrophage and the "alternative" anti-inflammatory M2 macrophage. Nevertheless. advancements in the functional characterization demonstrate that macrophage phenotypes are not confined to the M1 and M2 extremes. Instead, they encompass a continuous range of phenotypes linked to varying cytokine production and functional traits [48, 49]. In our research, the levels of both M1 macrophages and M2 macrophages were notably increased in the advanced plaques, and all six hub genes exhibit a strong correlation with these two types of macrophages. The true function of macrophages in carotid atherosclerosis remains unclear, and it is imperative to determine the future studies' examination of the correlation between these genes and macrophages.

We conducted a comprehensive analysis and identified six hub genes involved in carotid atherosclerotic plaques. These six hub genes, TYROBP, ITGAM, ITGB2, CD86, PLEK, LCP2 may be important biomarkers for the progression of atherosclerosis and potential treatment targets. However, there still exist some inevitable difficulties in this study. For example, thanks to a lack of atherosclerosis-related datasets, we use ruptured plaque and stable plaque datasets to validate the hub genes we get from the advanced carotid plaque and early plaque samples. Besides, for financial reasons, we only verified PLEK expression in immunehistochemistry.

## CONCLUSIONS

Through bioinformatics analysis of the microarray datasets, we identified six hub genes (TYROBP, ITGAM, ITGB2, PLEK, LCP2, CD86) that may be involved in the progression of atherosclerosis correlated with immune cells, which provides clues for us to explore the pathogenesis and therapeutic approaches of cardiovascular diseases.

## **AUTHOR CONTRIBUTIONS**

LZ, LJH, XHT and HSL conceived the project and participated in the study design and interpretation of the results. LZ wrote the manuscript. LZC and ZXN participated in the study design and helped draft the manuscript. GRX and DR participated in the interpretation of the data and critically reviewed the manuscript. LZ and LJH conducted subsequent immunohistochemical and immunofluorescence experiments. All authors have read and approved the final manuscript.

## **CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest.

## ETHICAL STATEMENT AND CONSENT

The experimental part of this study was approved by the Ethics Committee of Wuhan University People's Hospital (Approval No. WDRY2023-K076). The informed consent was waived by the Ethics Committee.

## FUNDING

This study was supported by the Fundamental Research Funds for the Central Universities (2042021kf0090).

## **REFERENCES**

- Brown AA, Viñuela A, Delaneau O, Spector TD, Small KS, Dermitzakis ET. Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. Nat Genet. 2017; 49:1747–51. <u>https://doi.org/10.1038/ng.3979</u> PMID:29058714
- Martorell-Marugan J, Toro-Dominguez D, Alarcon-Riquelme ME, Carmona-Saez P. MetaGenyo: a web tool for meta-analysis of genetic association studies. BMC Bioinformatics. 2017; 18:563. <u>https://doi.org/10.1186/s12859-017-1990-4</u> PMID:<u>29246109</u>
- Stears RL, Martinsky T, Schena M. Trends in microarray analysis. Nat Med. 2003; 9:140–5. <u>https://doi.org/10.1038/nm0103-140</u> PMID:<u>12514728</u>
- Toro-Domínguez D, Martorell-Marugán J, López-Domínguez R, García-Moreno A, González-Rumayor V, Alarcón-Riquelme ME, Carmona-Sáez P. ImaGEO: integrative gene expression meta-analysis from GEO database. Bioinformatics. 2019; 35:880–2. https://doi.org/10.1093/bioinformatics/bty721

#### PMID:30137226

 Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res. 2000; 10:853–60. https://doi.org/10.1101/gr.10.6.853

PMID:10854416

- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, et al. Large-scale identification, mapping, and genotyping of singlenucleotide polymorphisms in the human genome. Science. 1998; 280:1077–82. <u>https://doi.org/10.1126/science.280.5366.1077</u> PMID:9582121
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005; 4:17. <u>https://doi.org/10.2202/1544-6115.1128</u> PMID:<u>16646834</u>
- Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017; 18:220. <u>https://doi.org/10.1186/s13059-017-1349-1</u> PMID:29141660
- Döring Y, Manthey HD, Drechsler M, Lievens D, Megens RT, Soehnlein O, Busch M, Manca M, Koenen RR, Pelisek J, Daemen MJ, Lutgens E, Zenke M, et al. Auto-antigenic protein-DNA complexes stimulate plasmacytoid dendritic cells to promote atherosclerosis. Circulation. 2012; 125:1673–83. <u>https://doi.org/10.1z161/CIRCULATIONAHA.111.046755</u> PMID:22388324
- Ayari H, Bricca G. Identification of two genes potentially associated in iron-heme homeostasis in human carotid plaque using microarray analysis. J Biosci. 2013; 38:311–5. <u>https://doi.org/10.1007/s12038-013-9310-2</u> PMID:23660665
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets-update. Nucleic Acids Res. 2013; 41:D991–5. <u>https://doi.org/10.1093/nar/gks1193</u> PMID:23193258
- Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hyper variable genes and improves power to detect differential expression. Ann Appl Stat. 2016; 10:946–63.

https://doi.org/10.1214/16-AOAS920 PMID:<u>28367255</u>

 Ruegg KC, Anderson EC, Harrigan RJ, Paxton KL, Kelly JF, Moore F, Smith TB. Genetic assignment with isotopes and habitat suitability (gaiah), a migratory bird case study. Methods in Ecology and Evolution. 2017; 8:1241–52. https://doi.org/10.1111/2041-210X.12800

14. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3:Article3.

https://doi.org/10.2202/1544-6115.1027 PMID:<u>16646809</u>

- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003; 4:P3. <u>https://doi.org/10.1186/gb-2003-4-5-p3</u> PMID:12734009
- Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43:D1049–56. https://doi.org/10.1093/nar/gku1179 PMID:25428369
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017; 45:D353–61.

https://doi.org/10.1093/nar/gkw1092 PMID:27899662

- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–50. <u>https://doi.org/10.1073/pnas.0506580102</u> PMID:16199517
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. <u>https://doi.org/10.1186/1471-2105-9-559</u> PMID:<u>19114008</u>
- 20. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41:D808–15. <u>https://doi.org/10.1093/nar/gks1094</u> PMID:<u>23203871</u>
- Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol. 2011; 696:291–303.

https://doi.org/10.1007/978-1-60761-987-1 18 PMID:21063955

22. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014 (Suppl 4); 8:S11.

https://doi.org/10.1186/1752-0509-8-S4-S11 PMID:<u>25521941</u>

- Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010; 5:1315–6. <u>https://doi.org/10.1097/JTO.0b013e3181ec173d</u> PMID:20736804
- 24. Dexter F. Wilcoxon-Mann-Whitney test used for data that are not normally distributed. Anesth Analg. 2013; 117:537–8. <u>https://doi.org/10.1213/ANE.0b013e31829ed28f</u> PMID:23966647
- 25. Taylor CR, Levenson RM. Quantification of immunohistochemistry--issues concerning methods, utility and semiquantitative assessment II. Histopathology. 2006; 49:411–24. https://doi.org/10.1111/j.1365-2559.2006.02513.x PMID:<u>16978205</u>
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012; 9:671–5. <u>https://doi.org/10.1038/nmeth.2089</u> PMID:<u>22930834</u>
- 27. Odell ID, Cook D. Immunofluorescence techniques. J Invest Dermatol. 2013; 133:e4. <u>https://doi.org/10.1038/jid.2012.455</u> PMID:<u>23299451</u>
- Liu W, Zhao Y, Wu J. Gene expression profile analysis of the progression of carotid atherosclerotic plaques. Mol Med Rep. 2018; 17:5789–95. <u>https://doi.org/10.3892/mmr.2018.8575</u> PMID:<u>29436628</u>
- 29. Wang G, Kuai D, Yang Y, Yang G, Wei Z, Zhao W. Screening of potential gene markers for predicting carotid atheroma plaque formation using bioinformatics approaches. Mol Med Rep. 2017; 15:2039–48. https://doi.org/10.3892/mmr.2017.6273

PMID:28260035

 Guo J, Ning Y, Su Z, Guo L, Gu Y. Identification of hub genes and regulatory networks in histologically unstable carotid atherosclerotic plaque by bioinformatics analysis. BMC Med Genomics. 2022; 15:145.

https://doi.org/10.1186/s12920-022-01257-1 PMID:<u>35773742</u>

- Tabas I, Lichtman AH. Monocyte-Macrophages and T Cells in Atherosclerosis. Immunity. 2017; 47:621–34. <u>https://doi.org/10.1016/j.immuni.2017.09.008</u> PMID:<u>29045897</u>
- 32. Wolf D, Ley K. Immunity and Inflammation in Atherosclerosis. Circ Res. 2019; 124:315–27. <u>https://doi.org/10.1161/CIRCRESAHA.118.313591</u> PMID:<u>30653442</u>
- 33. Haasdijk RA, Den Dekker WK, Cheng C, Tempel D, Szulcek R, Bos FL, Hermkens DMA, Chrifi I, Brandt MM, Van Dijk C, Xu YJ, Van De Kamp EHM, Blonden LAJ, et al. THSD1 preserves vascular integrity and protects against intraplaque haemorrhaging in ApoE-/- mice. Cardiovasc Res. 2016; 110:129–39. https://doi.org/10.1093/cvr/cvw015 PMID:26822228
- 34. Martinet W, De Meyer GRY. Autophagy in atherosclerosis. Curr Atheroscler Rep. 2008; 10:216–23. <u>https://doi.org/10.1007/s11883-008-0034-y</u> PMID:<u>18489849</u>
- Rodgers KJ, Watkins DJ, Miller AL, Chan PY, Karanam S, Brissette WH, Long CJ, Jackson CL. Destabilizing role of cathepsin S in murine atherosclerotic plaques. Arterioscler Thromb Vasc Biol. 2006; 26:851–6. <u>https://doi.org/10.1161/01.ATV.0000203526.75772.4b</u> PMID:<u>16410454</u>
- Demigné C, Yacoub C, Morand C, Rémésy C. Interactions between propionate and amino acid metabolism in isolated sheep hepatocytes. Br J Nutr. 1991; 65:301–17. https://doi.org/10.1079/bjn19910088 PMID:2043606
- Arpaia N, Campbell C, Fan X, Dikiy S, van der Veeken J, deRoos P, Liu H, Cross JR, Pfeffer K, Coffer PJ, Rudensky AY. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature. 2013; 504:451–5. <u>https://doi.org/10.1038/nature12726</u> PMID:24226773
- 38. Yoo JY, Sniffen S, McGill Percy KC, Pallaval VB, Chidipi BJM. Gut Dysbiosis and Immune System in Atherosclerotic Cardiovascular Disease (ACVD). Microorganisms. 2022; 10:108. <u>https://doi.org/10.3390/microorganisms10010108</u> PMID:35056557
- 39. Li M, van Esch BC, Henricks PAJ, Garssen J, Folkerts G. Time and Concentration Dependent Effects of Short Chain Fatty Acids on Lipopolysaccharide- or Tumor Necrosis Factor α-Induced Endothelial Activation. Front Pharmacol. 2018; 9:233. <u>https://doi.org/10.3389/fphar.2018.00233</u> PMID:29615908
- 40. Sluiter TJ, van Buul JD, Huveneers S, Quax PHA, de Vries MR. Endothelial Barrier Function and Leukocyte

Transmigration in Atherosclerosis. Biomedicines. 2021; 9:328.

https://doi.org/10.3390/biomedicines9040328 PMID:<u>33804952</u>

- 41. Pan Y, Yu C, Huang J, Rong Y, Chen J, Chen M. Bioinformatics analysis of vascular RNA-seq data revealed hub genes and pathways in a novel Tibetan minipig atherosclerosis model induced by a high fat/cholesterol diet. Lipids Health Dis. 2020; 19:54. <u>https://doi.org/10.1186/s12944-020-01222-w</u> PMID:<u>32213192</u>
- 42. Oksala N, Pärssinen J, Seppälä I, Klopp N, Illig T, Laaksonen R, Levula M, Raitoharju E, Kholova I, Sioris T, Kähönen M, Lehtimäki T, Hytönen VP. Kindlin 3 (FERMT3) is associated with unstable atherosclerotic plaques, anti-inflammatory type II macrophages and upregulation of beta-2 integrins in all major arterial beds. Atherosclerosis. 2015; 242:145–54.

https://doi.org/10.1016/j.atherosclerosis.2015.06.058 PMID:26188538

- 43. Gailani D, Fisher TC, Mills DC, Macfarlane DE. P47 phosphoprotein of blood platelets (pleckstrin) is a major target for phorbol ester-induced protein phosphorylation in intact platelets, granulocytes, lymphocytes, monocytes and cultured leukaemic cells: absence of P47 in non-haematopoietic cells. Br J Haematol. 1990; 74:192–202. https://doi.org/10.1111/j.1365-2141.1990.tb02565.x
  - PMID:2317454
- Lundmark A, Davanian H, Båge T, Johannsen G, Koro C, Lundeberg J, Yucel-Lindberg T. Transcriptome analysis reveals mucin 4 to be highly associated with periodontitis and identifies pleckstrin as a link to systemic diseases. Sci Rep. 2015; 5:18475. <u>https://doi.org/10.1038/srep18475</u>
  PMID:26686060
- 45. Meletta R, Müller Herde A, Dennler P, Fischer E, Schibli R, Krämer SD. Preclinical imaging of the costimulatory molecules CD80 and CD86 with indium-111-labeled belatacept in atherosclerosis. EJNMMI Res. 2016; 6:1.

https://doi.org/10.1186/s13550-015-0157-4 PMID:26728358

- 46. Shapouri-Moghaddam A, Mohammadian S, Vazini H, Taghadosi M, Esmaeili SA, Mardani F, Seifi B, Mohammadi A, Afshari JT, Sahebkar A. Macrophage plasticity, polarization, and function in health and disease. J Cell Physiol. 2018; 233:6425–40. https://doi.org/10.1002/jcp.26429 PMID:29319160
- Ewing MM, Karper JC, Abdul S, de Jong RC, Peters HA, de Vries MR, Redeker A, Kuiper J, Toes RE, Arens R, Jukema JW, Quax PH. T-cell co-stimulation by CD28-CD80/86 and its negative regulator CTLA-4 strongly

influence accelerated atherosclerosis development. Int J Cardiol. 2013; 168:1965–74. <u>https://doi.org/10.1016/j.ijcard.2012.12.085</u> PMID:<u>23351788</u>

 Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, De Nardo D, Gohel TD, Emde M, Schmidleithner L, Ganesan H, Nino-Castro A, Mallmann MR, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity. 2014; 40:274–88. https://doi.org/10.1016/j.immuni.2014.01.006 PMID:24530056

49. Colin S, Chinetti-Gbaguidi G, Staels B. Macrophage phenotypes in atherosclerosis. Immunol Rev. 2014; 262:153–66. https://doi.org/10.1111/imr.12218 PMID:25319333