

RESEARCH

Open Access



Immune-associated biomarkers identification for diagnosing carotid plaque progression with uremia through systematical bioinformatics and machine learning analysis

Chunjiang Liu^{1†}, Liming Tang^{1†}, Yue Zhou¹, Xiaoqi Tang¹, Gang Zhang², Qin Zhu^{3*} and Yufei Zhou^{4*}

Abstract

Background Uremia is one of the most challenging problems in medicine and an increasing public health issue worldwide. Patients with uremia suffer from accelerated atherosclerosis, and atherosclerosis progression may trigger plaque instability and clinical events. As a result, cardiovascular and cerebrovascular complications are more likely to occur. This study aimed to identify diagnostic biomarkers in uremic patients with unstable carotid plaques (USCPs).

Methods Four microarray datasets (GSE37171, GSE41571, GSE163154, and GSE28829) were downloaded from the NCBI Gene Expression Omnibus database. The Limma package was used to identify differentially expressed genes (DEGs) in uremia and USCP. Weighted gene co-expression network analysis (WGCNA) was used to determine the respective significant module genes associated with uremia and USCP. Moreover, a protein–protein interaction (PPI) network and three machine learning algorithms were applied to detect potential diagnostic genes. Subsequently, a nomogram and a receiver operating characteristic curve (ROC) were plotted to diagnose USCP with uremia. Finally, immune cell infiltrations were further analyzed.

Results Using the Limma package and WGCNA, the intersection of 2795 uremia-related DEGs and 1127 USCP-related DEGs yielded 99 uremia-related DEGs in USCP. 20 genes were selected as candidate hub genes via PPI network construction. Based on the intersection of genes from the three machine learning algorithms, three hub genes (FGR, LCP1, and C5AR1) were identified and used to establish a nomogram that displayed a high diagnostic performance (AUC: 0.989, 95% CI 0.971–1.000). Dysregulated immune cell infiltrations were observed in USCP, showing positive correlations with the three hub genes.

Conclusion The current study systematically identified three candidate hub genes (FGR, LCP1, and C5AR1) and established a nomogram to assist in diagnosing USCP with uremia using various bioinformatic analyses and machine learning algorithms. Herein, the findings provide a foothold for future studies on potential diagnostic candidate genes

[†]Chunjiang Liu, Liming Tang contributed equally to this work

*Correspondence:

Qin Zhu

zhuq@njmu.edu.cn

Yufei Zhou

zyf15851852592@163.com; 21111210061@m.fudan.edu.cn

Full list of author information is available at the end of the article



for USCP in uremic patients. Additionally, immune cell infiltration analysis revealed that the dysregulated immune cell proportions were identified, and macrophages could have a critical role in USCP pathogenesis.

Keywords Unstable carotid plaque, Uremia, Diagnostic biomarker, Bioinformatics analysis, Machine learning, Immune cell infiltration

Introduction

Uremia is one of the most challenging medical disorders and a growing public health issue worldwide [1]. It is often complicated by severe arteriosclerosis, consequently, cardiovascular and cerebrovascular complications are more likely to occur compared to the general population [2]. Arteriosclerosis refers to major pathological alterations in cardiovascular and cerebrovascular diseases with uremia, which cannot be fully explained by the classical risk factors, such as age, gender, smoking, obesity, dyslipidemia, hypertension, and diabetes [3]. A previous study described that during the development of atherosclerosis in patients with uremia, atherogenesis is stimulated by lipid disturbances, thrombogenesis, and synthesis of vasoactive substances, growth factors, and mediators of inflammation [4].

Atherosclerosis is a frequently encountered vascular condition that affects the intima of the arteries and can be life-threatening [5]. Its progression may lead to plaque destabilization, resulting in plaque rupture and thrombosis. Moreover, rupture of the carotid artery by an unstable plaque is one of the primary causes of stroke [6]. Even when the rupture of plaques does not result in a stroke, they likely contribute to plaque progression and arterial stenosis [7]. As is well established, uremic patients suffer from accelerated atherosclerosis; i.e., plaques form faster, and the disease is more severe [8]. Earlier studies have signaled that uremia plays a decisive role in stimulating plaque growth, intraplaque hemorrhage, and lesion instability [9]. Nevertheless, even when severe, unstable carotid plaque (USCP) patients are frequently asymptomatic, implying that they are unaware of the illness and do not seek medical help until alarming complications arise [10]. Therefore, there is an urgent need to distinguish stable carotid plaques (SCPs) from USCPs to prevent cerebrovascular diseases; indeed, early diagnosis and prompt treatment of USCPs are crucial. Accordingly, biomarkers are essential for early diagnosis and medical intervention for conditions such as uremia and USCP, which typically lacks overt clinical manifestations.

In recent years, microarray analysis has become a vital method for researching genetic modifications in various diseases owing to advancements in technologies [11]. Biomarkers for uremia have been identified in multiple studies. For instance, FZD10, FOXD4, PPP3R1,

and UCP2 may be valuable uremia diagnostic biomarkers [12]. Another research reported that genes associated with USCP encompassed CD5L, S100A12, CKB, CEMIP, and SH3GLB1 [13]. However, no previous studies have identified the mechanism of uremia-induced USCP at the genetic level. An attempt was made to apply Limma analysis and the weighted gene co-expression network analysis (WGCNA) to screen the gene clusters of connected, shared, and correlating genes in uremia and USCP. Through this method, common risk genes associated with multiple disease phenotypes [14] have been identified.

To the best of our knowledge, this was the first study to simultaneously analyze uremia and USCP. More importantly, the identification of pivotal diagnostic biomarkers for uremia-related USCP was pursued through a series of bioinformatics and machine learning approaches.

Methods

Microarray data and data processing

The study flowchart is illustrated in Fig. 1. Four microarray datasets (GSE37171, GSE41571, GSE163154, and GSE28829) were downloaded from the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database [15]. The search strategy was “Uremia” or “Carotid Plaque” AND “Homo sapiens” AND “Series” AND “Expression profiling by array”. Dataset GSE37171 [16] included data on 75 uremic patients and 40 controls. Datasets GSE41571 [17] and GSE163154 [18] included gene expression data on USCP and SCP patients. Data GSE28829 [19] was utilized as the external validation dataset. Detailed information on the datasets is presented in Table 1.

The data were preprocessed using “affy” in R (R Institute for Statistical Computing, Vienna, Austria) from the Bioconductor project to carry out background calibration and normalization. Multiple probes matching the same gene were calculated as a median expression. Gene symbols were converted from probes and then prepared into matrix files. After merging the GSE41571 and GSE163154 datasets, batch effects and other unwanted variations between the two datasets were eliminated using the R “surrogate variable analysis” (SVA) package from Bioconductor.

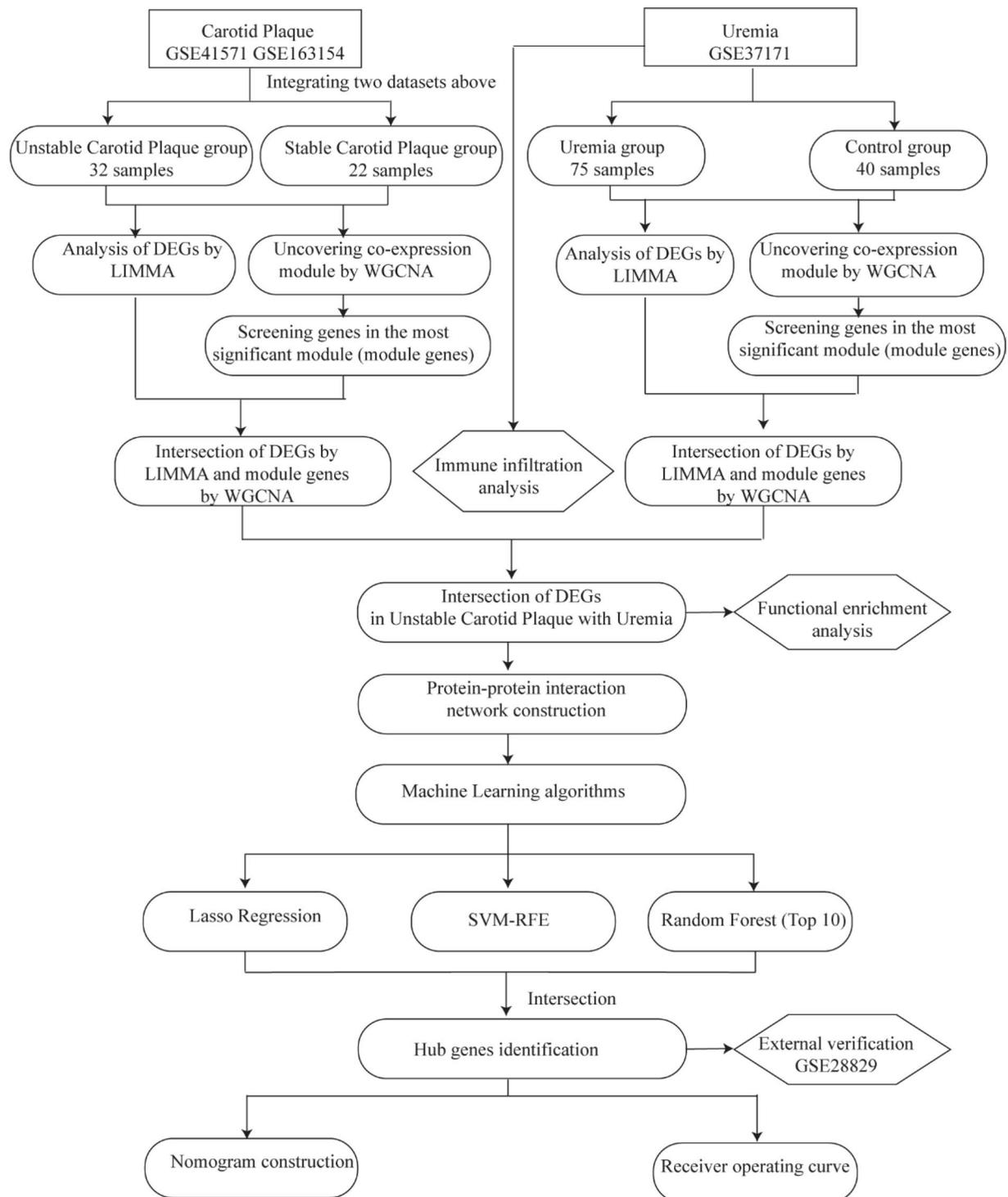


Fig. 1 Study flowchart. *DEGs* differentially expressed genes, *LIMMA* linear models for microarray data, *WGCNA* weighted gene co-expression network analysis, *SVM-RFE* support vector machine-recursive feature elimination, *ROC* receiver operating characteristic curve

Table 1 Basic information on the GEO datasets used in the study

ID	GSE series	Disease	Samples	Source types	Platform	Group
1	GSE37171	Uremia	75 uremia patients and 40 normal controls	whole blood	GPL570	Discovery cohort
2	GSE41571	CP	5 USCP patients and 6 SCP patients	Plaque	GPL570	Discovery cohort
3	GSE163154	CP	27 USCP patients and 16 SCP patients	Plaque	GPL6104	Discovery cohort
4	GSE28829	CP	16 advanced carotid plaque patients and 13 early carotid plaque patients	Plaque	GPL570	Validation cohort

CP carotid plaque, USCP unstable carotid plaque, SCP stable carotid plaque

Differentially expressed genes (DEGs) identification

The Bioconductor Limma package [20] was used to screen DEGs between uremia and the control in dataset GSE37171 and to identify DEGs between USAP and SAP in the merged dataset (GSE41571, GSE163154). The criteria for identification of DEGs were $|\text{Fold change}| > 1.5$ and $P \text{ value} < 0.05$.

Significant module identification via weighted gene co-expression network analysis (WGCNA)

WGCNA [21] is an efficient method for building co-expression networks extensively applied to analyze large datasets. Using WGCNA, respective significant module genes correlating with uremia and USCP were identified. First, each gene was analyzed by calculating the median absolute deviation (MAD) and eliminating 50% of those with the smallest MAD. The `goodSamplesGenes` function of WGCNA was then used to filter the DEG expression matrix to create a scale-free co-expression network. Next, a weighted adjacency matrix was built based on the soft-thresholding parameter β , originating from co-expression similarity, applying the `pick-Soft-Threshold` function. A topological overlap matrix (TOM) was constructed based on adjacency; TOM has the ability to measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for network generation. Additionally, the corresponding dissimilarity (1-TOM) was determined. Then, hierarchical clustering and dynamic tree-cutting were used to identify the modules. Using the TOM-based dissimilarity measure, with 50 genes (Gene group) as the minimum size, average linkage hierarchical clustering was constructed for the Genes dendrogram; gene modules were formed by classifying genes with similar expression profiles. Finally, the eigengene network was visualized after merging some modules based on the dissimilarity of estimated module eigengenes.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis

GO [22] contains structured and computable information about the functions of genes and their products. During GO analysis, biological process (BP), cellular

component (CC), and molecular function (MF) are identified. As a knowledge base, KEGG [23] provides detailed information about pathways, as well as systematic analyses of gene functions. Functional enrichment analysis was performed with the R package `ClusterProfiler` (version 3.14.3) [24], and the top 10 GO terms were visualized using the R package “`ggplot2`” in each category. The selection criteria were: adjusted $P \text{ value} < 0.05$ and false discovery rate < 0.25 .

Protein–protein interaction (PPI) network construction

PPI was constructed using the Search Tool for Retrieval of Interacting Genes (String) database [25] (version 11.5; www.string-db.org), an online tool for identifying protein interactions; 0.40 was set as the minimum required interaction score. Genes that did not interact with each other were hidden. Next, the file of interaction data was downloaded and visualized using Cytoscape [26] (<http://www.cytoscape.org>). The top 30 genes were identified using three algorithms (betweenness, closeness, and degree) in CytoHubba [27], a plug-in of Cytoscape. Finally, the DEGs were visualized by intersecting the three algorithms using a Venn diagram.

Machine learning

For the identification of candidate hub genes, three different machine learning algorithms were utilized to minimize the risk of bias in potential diagnostic genes. The Support Vector Machine-Recursive Feature Elimination (SVM-RFE) [28] algorithm of the support vector machine was used to identify the optimal variables by deleting SVM-generated eigenvectors. The least absolute shrinkage and selection operator (Lasso) [29] regression analysis was performed using the “`glmnet`” R package; regularization was utilized to reduce prediction errors. With random forest [30], high dimensions of data can be handled, predictive models can be developed, and the importance of each variable can be predicted. The genes from all three algorithms were eventually intersected for further analysis.

Receiver operating characteristic curve (ROC) evaluation and nomogram construction

The Student's *t*-test was used to compare the expression level of each candidate diagnostic gene between the USCP and the SCP groups. In order to evaluate the diagnostic value of each candidate gene, a ROC curve was established; the area under the curve (AUC) and the 95% confidence interval (CI) were subsequently calculated. Using the R package "rms", a nomogram was constructed to provide a clinical perspective on the diagnosis of uremia with USAP. Each gene was scored based on its expression, and after the accumulation of all genes, the total score could be used to predict the incidence of uremia with USAP. Finally, a ROC curve was constructed for the nomogram. Lastly, the validation dataset GSE28829 was applied to analyze the diagnostic value of each candidate gene via the ROC curve and nomogram.

Immune infiltration analysis

Using the "Cibersort" algorithm, the normalized gene expression matrix was converted into the composition of infiltrating immune cells [31]. With the R package "Cibersort", the percentage of different types of immune cells was calculated and displayed in a bar plot between USCP and SCP. Next, the "vioplot" R package was adopted to compare and visualize the proportion of immune cells between USCP and SCP. Then, using the R package "corrplot" [32], a heatmap of the correlation between immune cells and USCP pathogenesis was constructed. Finally, the correlation between immune cell infiltrations and the hub diagnostic genes was analyzed.

Results

DEGs identification via Limma in uremia and USCP

2904 DEGs were screened between the uremia and control groups, of which 495 were up-regulated, and 2409 were down-regulated. A total of 1154 DEGs were screened between the USCP and SCP groups, of which 544 were up-regulated, and 610 were down-regulated. A heatmap and volcano plot were used to display the top 30 up-regulated or down-regulated DEGs and all DEGs, respectively (uremia: Fig. 2A, B, USCP: Fig. 3A, B).

(See figure on next page.)

Fig. 2 The heatmap and volcano plot of DEGs between uremia and the control, module genes recognition in uremia via WGCNA. **A** Heatmap showing the top 30 up-regulated and down-regulated DEGs in uremia compared with the control. Red blocks indicate up-regulation between the two groups, while blue blocks indicate down-regulation. **B** The volcano plot displaying all DEGs between uremia and the control. Significantly up-regulated and down-regulated DEGs with significant upregulations or downregulations are marked in red and green, respectively. **C** The heatmap of uremia-associated modules. Numbers in the top left corner indicate module-uremia correlation, and the P value in the lower right corner indicates the correlation. In uremia, cyan and pink modules show the most significant correlation. **D-E** The correlation between module membership in the cyan/pink module and gene significance in uremia. **F** The Venn plot displays that showing the 2795 DEGs were identified from the intersection between DEGs via Limma and significant module genes via WGCNA in uremia. DEGs differentially expressed genes, WGCNA weighted gene co-expression network analysis

Identification of significant module genes in uremia and USCP via WGCNA

Significant module genes of uremia and USCP were identified using WGCNA. The gray module failed to cluster genes regarded as a junk module. Meanwhile, the cyan module exhibited the most significant positive correlation with uremia ($r=0.54$, $p=4.6 \times 10^{-10}$), while negative correlation was most prominent in the pink module ($r=-0.76$, $p=1.5 \times 10^{-22}$) (Fig. 2C). Figure 2D, E depicts the correlation between module membership in gene significance and the cyan/pink module in uremia. Meanwhile, the dark red module exhibited the most significant positive correlation with USCP ($r=0.77$, $p=1.5 \times 10^{-11}$), whereas negative correlation was most prominent in the brown module ($r=-0.72$, $p=8.3 \times 10^{-10}$) (Fig. 3C). Figure 3D, E delineates the correlation between module membership in gene significance and the cyan/pink module in USCP. As a result, 8942 uremia-related genes and 5978 USCP-related genes were selected.

Finally, the intersection of 2904 DEGs and 8942 module genes regarding uremia yielded 2795 uremia-related DEGs (Fig. 2F), and the intersection of 1154 DEGs and 5978 module genes regarding USCP yielded 1127 USCP-related DEGs (Fig. 3F). The soft threshold selection and gene cluster tree are displayed in Additional file 1: Figures S1 and S2.

Functional enrichment analysis of uremia-related DEGs in USCP

The intersection of 2795 uremia-related DEGs and 1127 USCP-related DEGs yielded 99 uremia-related DEGs in USCP (Fig. 4A). Significantly enriched GO terms of the 99 uremia-related DEGs for BP included "immune system process", "intracellular signal transduction", and "cell activation". Moreover, "cytoplasmic vesicle", "cytoplasmic vesicle part", and "secretory vesicle" were enriched in CC, while the MF of DEGs were highly associated with "cytoskeletal protein binding", "actin binding", and "actin filament binding" (Fig. 4B-D, Additional file 1: Table S1). Figure 4E displays the functional pathway analysis of the 99 DEGs, mainly enriched in "Rheumatoid arthritis", "Lysosome", and "Graft-versus-host disease".

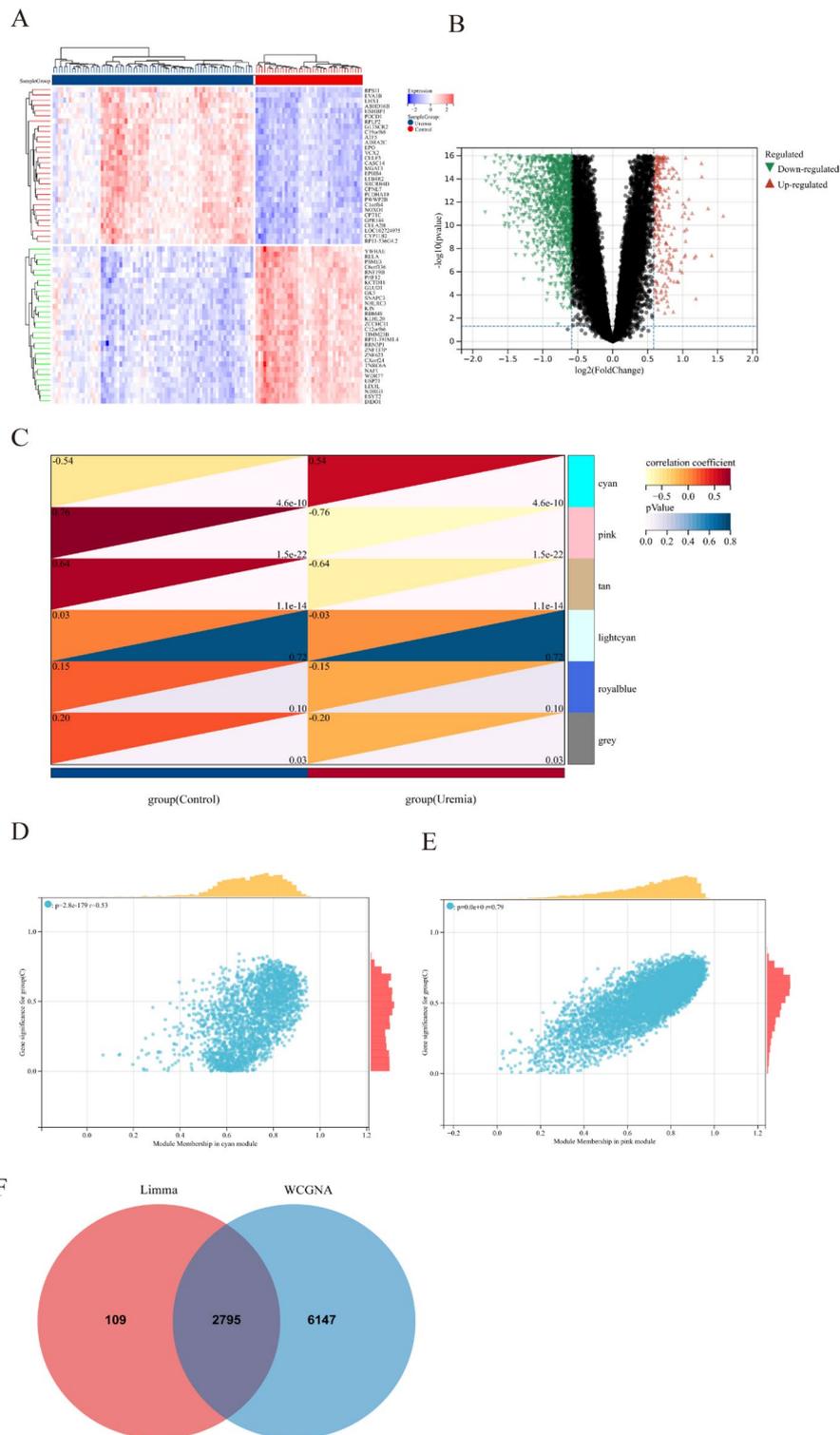


Fig. 2 (See legend on previous page.)

PPI network construction and hub gene selection

Based on the 99 DEGs, a PPI network was preliminarily constructed to identify hub DEGs for uremia with

USCP. 44 DEGs were screened after interacting with the others (Fig. 5A), and 45 DEGs were eliminated due to lack of interaction. Moreover, three different algorithms

(degree, betweenness, closeness) were used in the CytoHubba plug-in to identify intersecting DEGs. The top 30 node genes from the betweenness, closeness and degree algorithms are visualized and presented in Fig. 5B–D. The intersection of 30 genes from the three algorithms was depicted using a Venn diagram, and 20 genes were selected as candidate hub genes (Fig. 5E). Detailed information on the 20 genes is listed in Additional file 1: Table S2.

Candidate hub genes selection via machine learning

Lasso regression, random forest, and SVM-RFE algorithms were adopted to screen candidate hub genes for ROC evaluation. According to the Lasso regression algorithm, 7 potential hub genes corresponding to the lowest point of the curve were identified as the most suitable biomarkers for USCP with uremia diagnosis. Figure 6A illustrates the Lasso regression results. DEG gene importance was calculated using random forest.

The random forest algorithm revealed the error in USCP (Fig. 6B). A list of the ten most important DEGs is shown in Fig. 6C. As a result of SVM-RFE, the top 8 DEGs had the lowest error and highest accuracy in diagnosing USCP with uremia (Fig. 6D, E). Finally, based on the intersection of genes from the three algorithms, three hub genes (FGR, LCP1, and C5AR1) were selected for nomogram construction and diagnostic value evaluation (Fig. 6F).

The diagnostic value evaluation and nomogram construction

Three hub genes were up-regulated in USCP compared to SCP (Fig. 7A). The diagnostic value of each candidate hub gene was assessed using ROC curves. Figure 7B shows the AUC and 95% CI for each gene: FGR (AUC: 0.970, 95% CI 0.934–1.000); LCP1 (AUC: 0.972, 95% CI 0.933–1.000); C5AR1 (AUC: 0.987, 95% CI 0.968–1.000). Each gene presented a satisfactory diagnostic performance based on ROC curve analysis. Finally, the nomogram was built (Fig. 7C), and the AUC under the nomogram was 0.989 (95% CI 0.971–1.000), yielding a high clinical diagnostic value (Fig. 7D). As part of the validation process, the validation dataset of early-stage and advanced-stage atherosclerotic plaques (GSE28829) was adopted

to evaluate the diagnostic value via ROC curve analysis. ROC analysis of the nomogram yielded an AUC of 0.942, demonstrating an equally high clinical diagnostic value (Fig. 8).

Immune cell infiltration analysis

A bar plot (Fig. 9A) was used to display the percentage of 21 types of immune cells in each sample following the application of the Cibersort algorithm; in all samples, CD4 naive T cells were absent because their proportion was zero. The boxplot showed that the proportion of M0 and M1 macrophages were higher in USCP compared with the control, while the proportion of naive B cells and M2 macrophages was lower in SCP (Fig. 9B). Correlation analysis showed that activated mast cells had the highest positive correlation with naive B cells ($r=0.67$), while resting NK cells had the highest negative correlation with activated NK cells ($r=-0.44$) (Fig. 9C). In summary, focusing on macrophage regulation could serve as a potential approach for USCP treatment. Additionally, immune cell infiltration was positively correlated with all three hub DEGs (Fig. 9D).

Discussion

Uremia is frequently complicated by severe arteriosclerosis [33, 34]. Dyslipidemia, inflammation, and oxidative stress have been reported to be the risk factors for arteriosclerosis development in uremic patients [35]. Patients with uremia suffer from accelerated atherosclerosis; atherosclerosis progression may trigger the rupture of plaques, the formation of thrombi in the lumen of vessels, and the occurrence of clinical events, which significantly contribute to the morbidity and mortality of patients [36]. To the best of our knowledge, this is the first study to apply bioinformatics analysis methods and machine learning algorithms to uremia and USCP. Moreover, all the dataset samples used in this study were peripheral blood samples. Accordingly, only the analysis of the expression of the hub genes in peripheral blood from uremic patients was required to estimate the probability of USCP incidence in uremic patients, which is an effective and practical clinical approach. The total score could be determined by adding the scores of each gene after establishing a nomogram. Uremic patients with high total

(See figure on next page.)

Fig. 3 The heatmap and volcano plot of DEGs between USCP and SCP, module genes recognition in uremia via WGCNA. **A** Heatmap showing the top 30 up-regulated and down-regulated DEGs in USCP compared with SCP. Red blocks indicate up-regulation between the two groups, while blue blocks indicate down-regulation. **B** The volcano plot displaying all DEGs between USCP and SCP. Significantly up-regulated and down-regulated DEGs with significant upregulations or downregulations are marked in red and green, respectively. **C** The heatmap of USCP-associated modules. Numbers in the top left corner indicate module-USCP correlation, and the P value in the lower right corner indicates the correlation. In USCP, dark red and brown modules show the most significant correlation. **D–E** The correlation between module membership in dark red/brown module and gene significance in USCP. **F** The Venn plot illustrating the 1127 DEGs were identified from the intersection between DEGs via Limma and significant module genes via WGCNA in USCP. USCP unstable carotid plaque, SCP stable carotid plaque, DEGs differentially expressed genes, WGCNA weighted gene co-expression network analysis

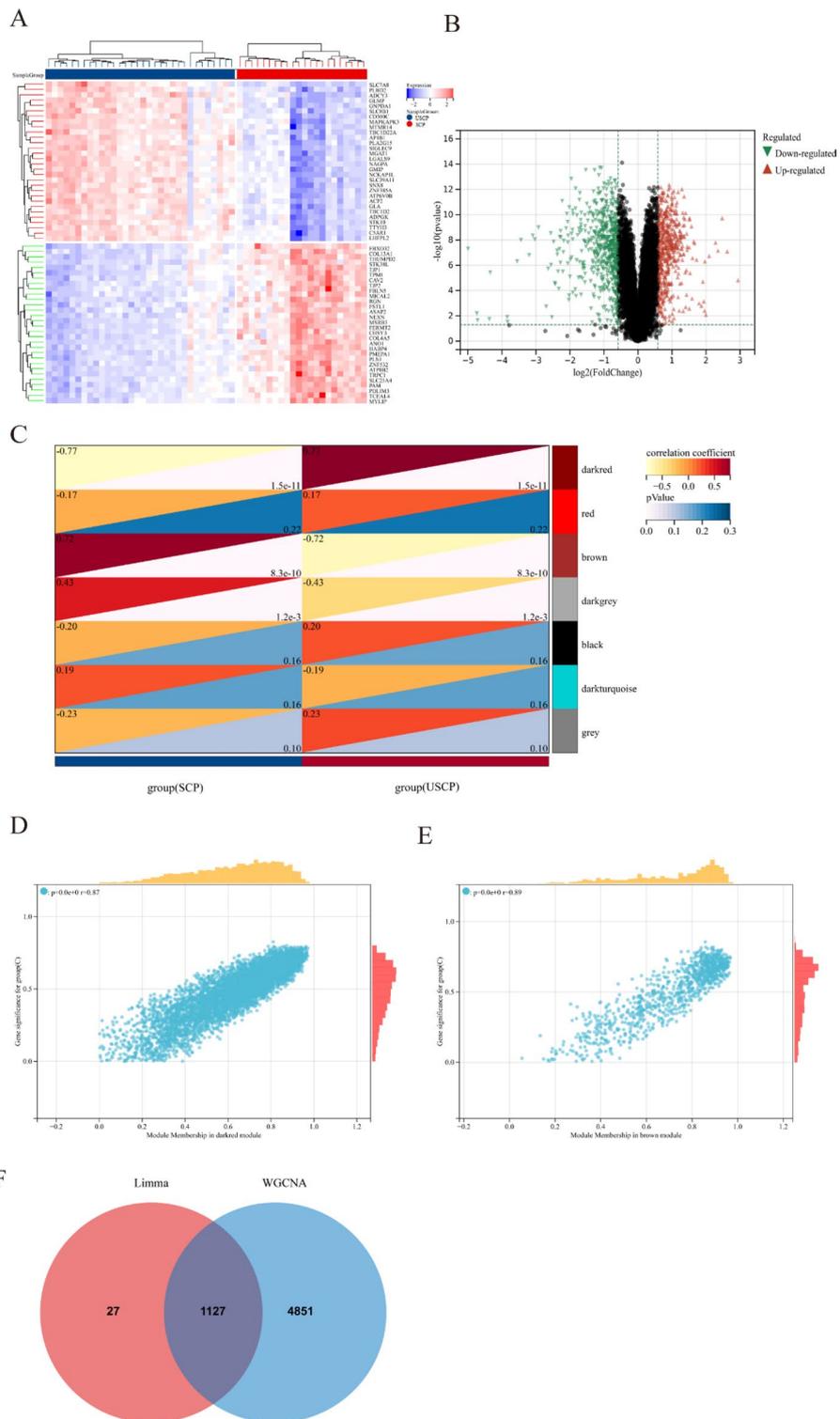


Fig. 3 (See legend on previous page.)

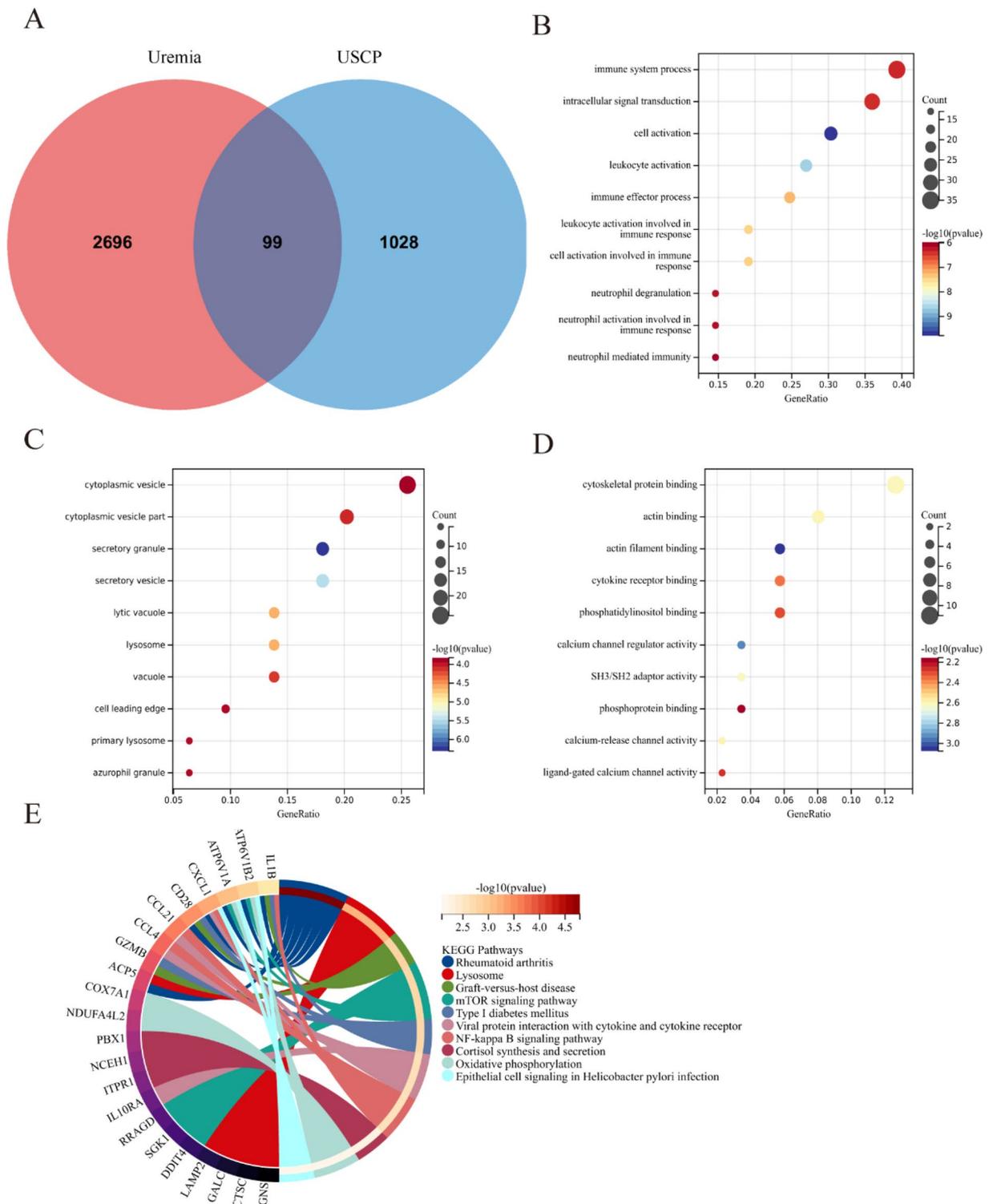


Fig. 4 Functional enrichment analysis of uremia-related DEGs in USCP. **A** The Venn plot displaying the intersection of DEGs between uremia and USCP yielded 99 uremia-related DEGs in USCP. **B–D** GO analysis (BP, CC, and MF) of uremia-related DEGs in USCP. The X-axis represents the gene ratio, Y-axis refers to different ontologies, the circle size represents the gene number, and the color indicates the significance. **E** KEGG pathway analysis of uremia-related DEGs in USCP. *USCP* unstable carotid plaque, *DEGs* differentially expressed genes, *BP* biological function, *CC* cellular component, *MF* molecular function, *GO* Gene Ontology, *KEGG* Kyoto Encyclopedia of Genes and Genomes

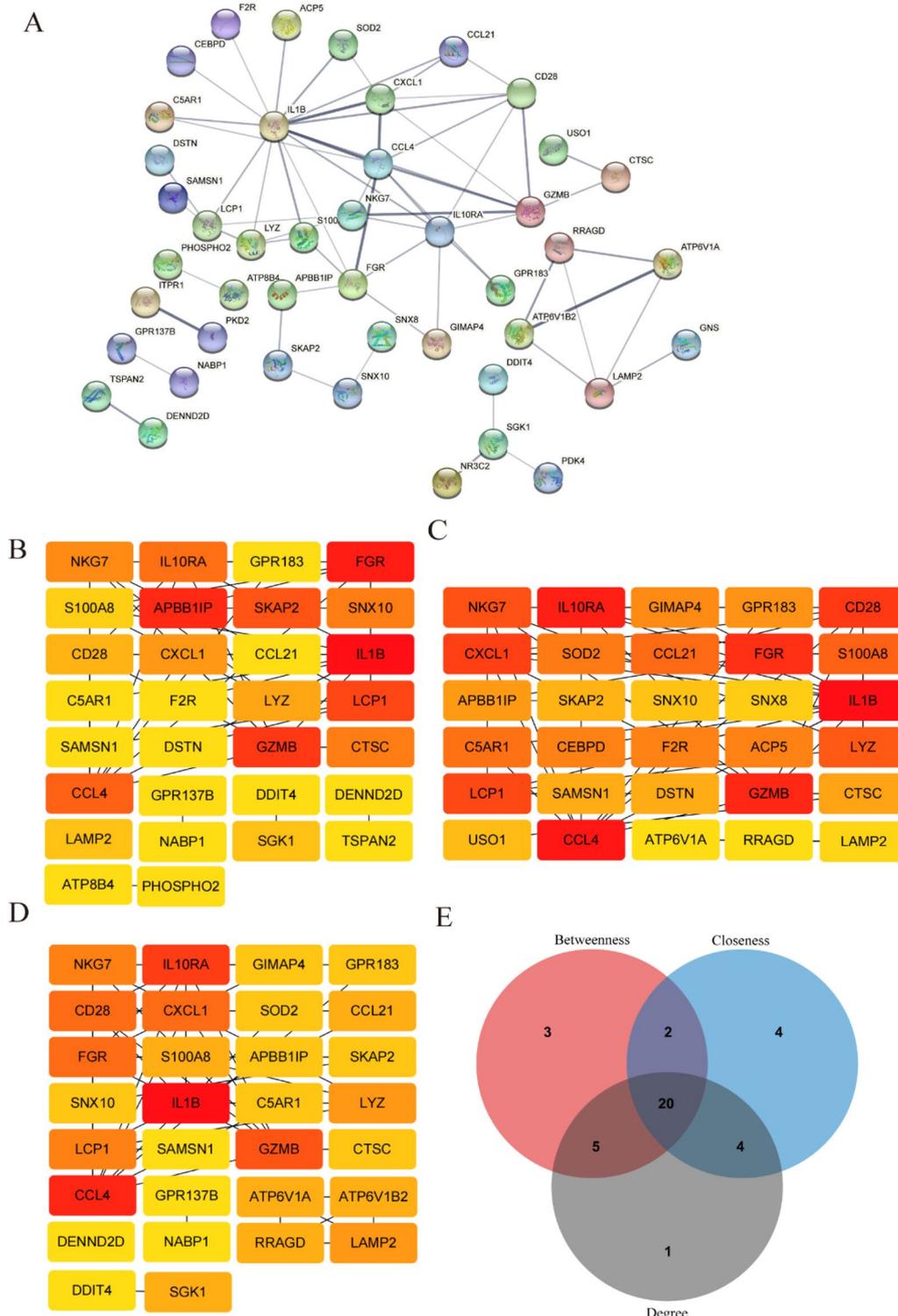


Fig. 5 PPI network construction and hub gene selection. **A** The whole PPI network of 44 uremia-related DEGs in USCP was visualized via String. The remaining 55 DEGs were eliminated due to a lack of interaction. **B–D** With the CytoHubba plug-in of Cytoscape, we selected the top 30 DEGs were selected using three different algorithms. Panels **B–D** illustrate betweenness, closeness, and degree algorithms. In the algorithm, darker colors indicate more significant weights. **E** Based on the intersection of genes from the three algorithms, 20 DEGs were chosen for further analysis. *PPI* protein–protein interaction network, *DEGs* differentially expressed genes, *USCP* unstable carotid plaque

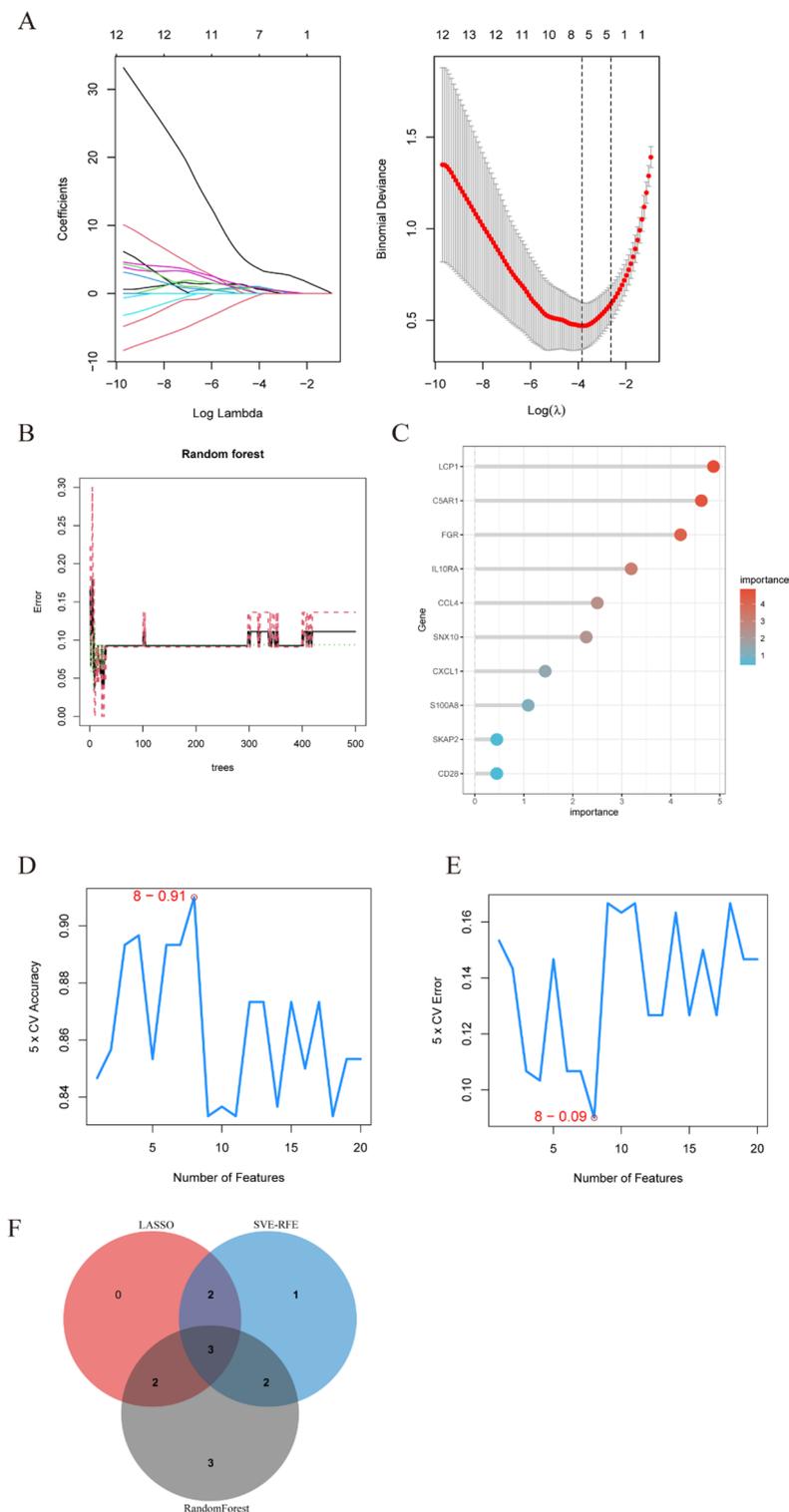


Fig. 6 Candidate diagnostic biomarker identification via machine learning algorithms. **A** Based on the Lasso regression algorithm, 7 genes corresponding to the lowest point of the curve were identified as the most accurate biomarkers for USCP with the diagnosis of uremia with USCP. **B, C** The random forest algorithm shows the error in USCP. Based on the random forest algorithm’s importance score, we selected and ranked the top ten genes. **D, E** 8 genes were selected based on SVM-RFE with the lowest error and highest accuracy. **F** Based on the intersection of genes from three algorithms, three hub genes (*FGR*, *LCP1*, and *C5AR1*) were selected for the next step of nomogram construction and diagnostic value evaluation. *SVM-RFE* support vector machine-recursive feature elimination, *FGR* feline Gardner-Rasheed, *LCP1* lymphocyte cytosolic protein 1, *C5AR1* complement C5a receptor 1

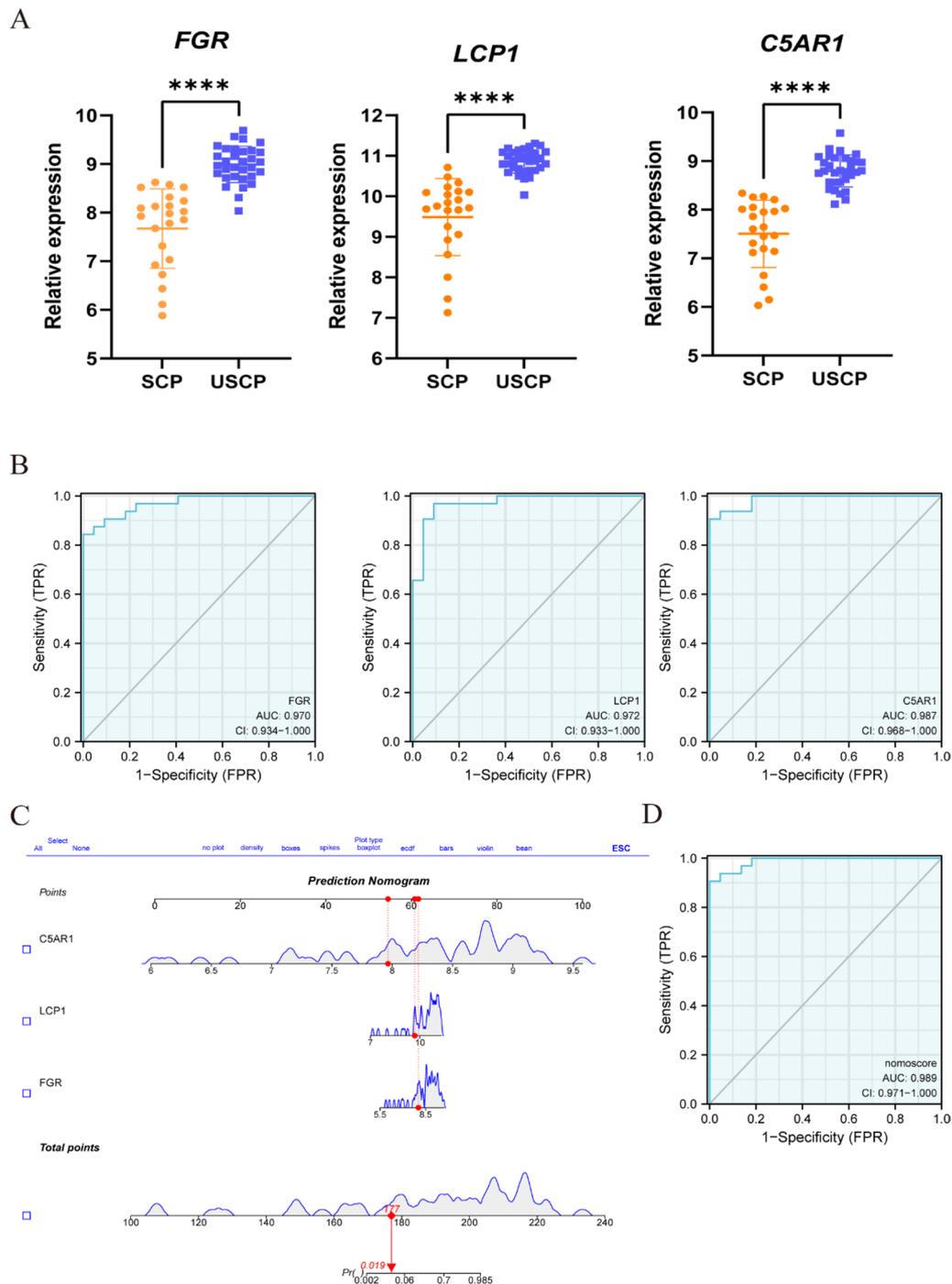


Fig. 7 The diagnostic value of candidate biomarkers and the nomogram construction. **A** Differences in expression between USCP and SCP for the three genes. ****, $P < 0.0001$. **B** The diagnostic value of the three genes in USCP with uremia from the ROC curve. Each panel displays the AUC under the curve and 95% CI. **C** Based on the three genes, a nomogram was constructed for diagnosing USCP with uremia. **D** The diagnostic value of the nomogram in USCP with uremia from the ROC curve. SCP stable carotid plaque, USCP unstable carotid plaque, ROC receiver operating characteristic curve, AUC area under the curve, CI confidence interval

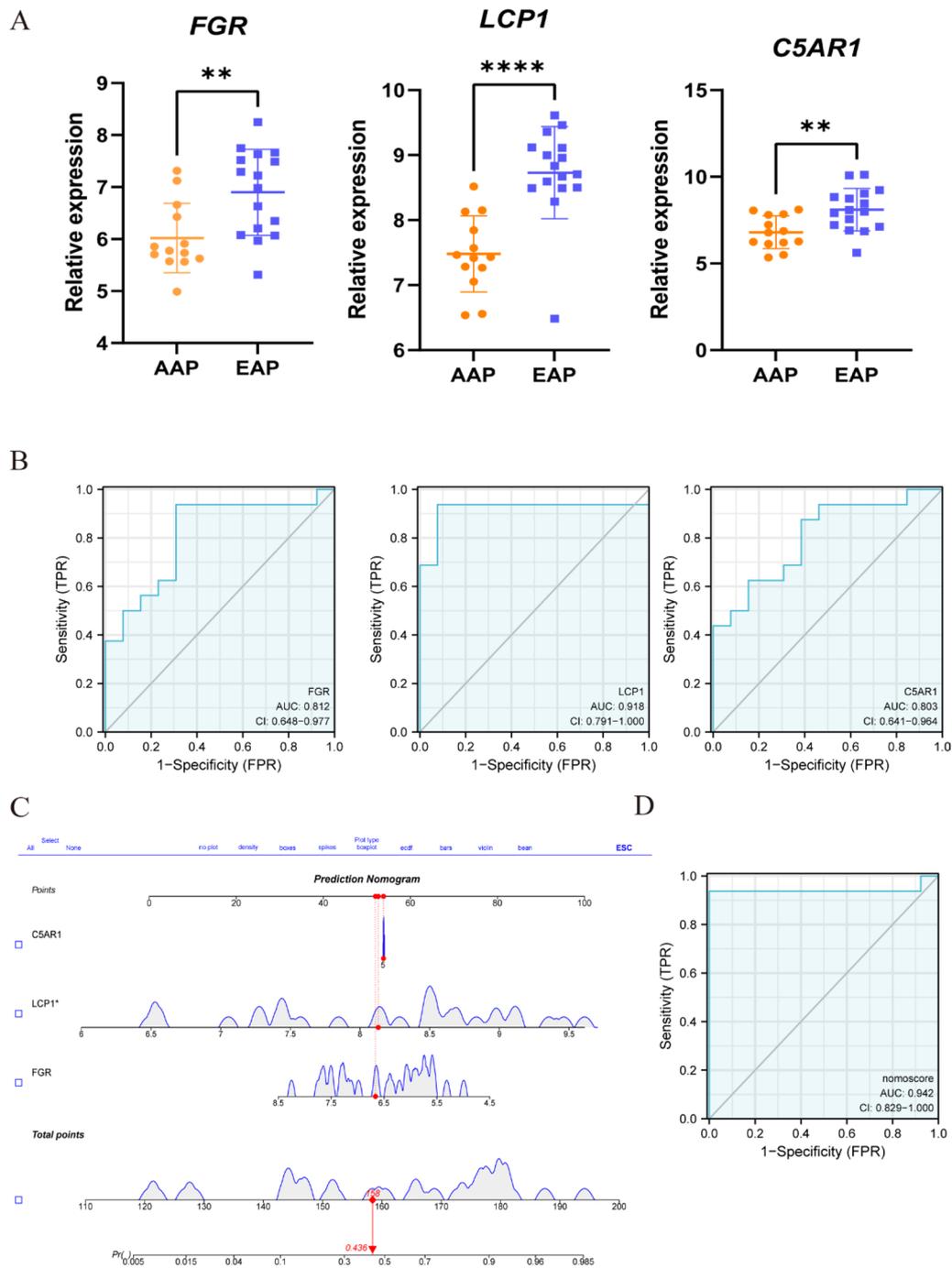


Fig. 8 The diagnostic value of candidate biomarkers and the nomogram construction in the external validation dataset GSE28829. **A** Differences in expression between EAP and AAP for the three genes. **, $P < 0.01$; ****, $P < 0.0001$. **B** The diagnostic value of the three genes in AAP from the ROC curve. Each panel displays the AUC under the curve and 95% CI. **C** Based on the three genes, a nomogram was constructed for AAP. **D** The ROC curve of the nomogram in AAP. AAP advanced-stage atherosclerotic plaque, EAP early-stage atherosclerotic plaques, ROC receiver operating characteristic curve, AUC area under the curve, CI confidence interval

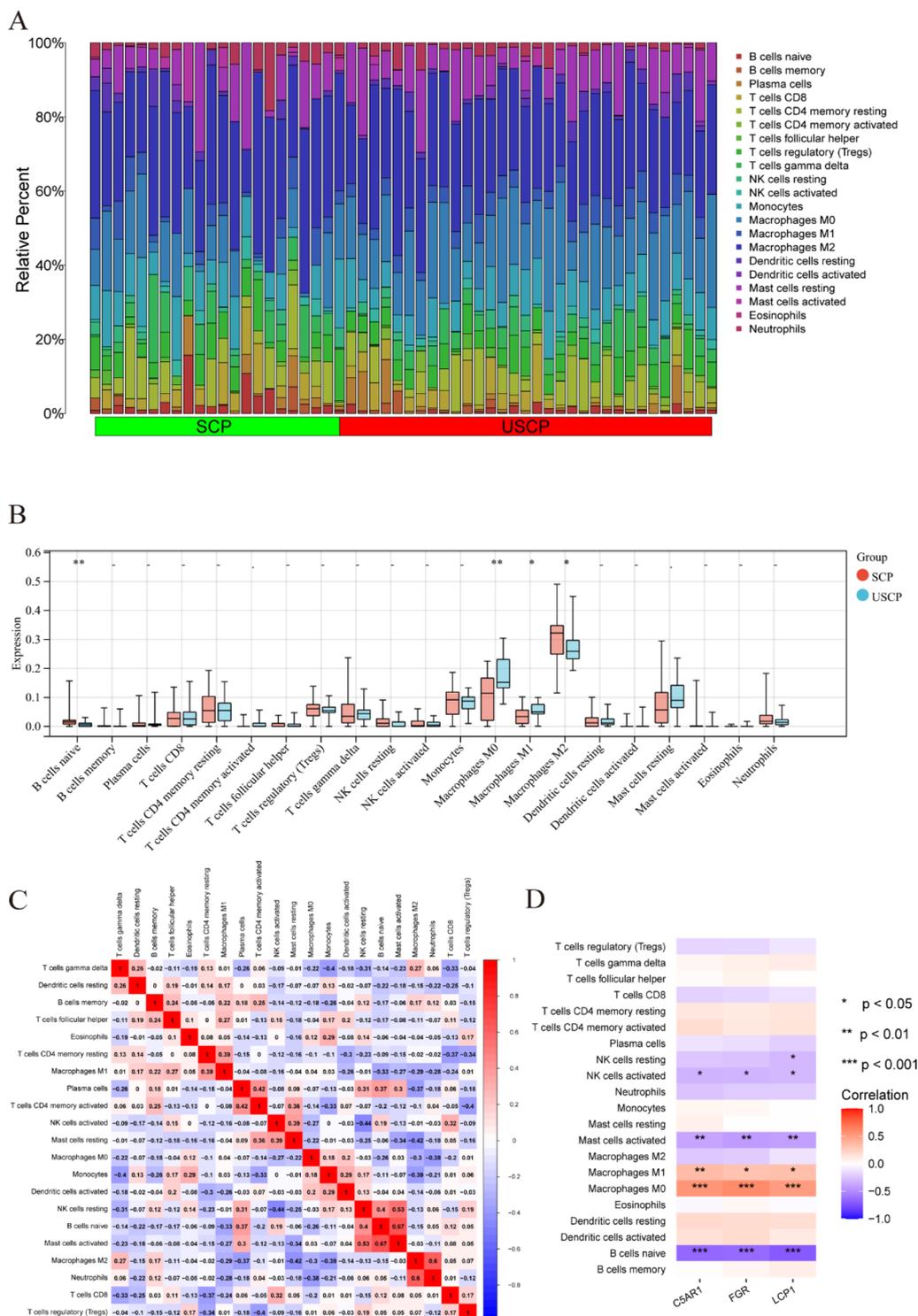


Fig. 9 Alterations in the immunological features from SCP to USCP and correlations between the hub DEGs and immunological features in USCP. **A** The bar plot showing the proportion of immune cells in different samples. **B** The boxplot comparing the expression of immune cells between USCP and SCP. *P < 0.05; **P < 0.01. **C** The correlated heatmap represents the correlation between different immune cells in USCP pathogenesis. **D** Correlation analysis of immune cell infiltrations with the three hub DEGs. *P < 0.05; **P < 0.01; ***P < 0.001. SCP stable carotid plaque, USCP unstable carotid plaque, DEGs differentially expressed genes

scores should be screened, and the prognosis for this patient population can be enhanced with early treatment. Therefore, the nomogram could be extensively employed in clinical practice.

Using multiple integrated bioinformatics and machine learning methods, the current study discovered three biomarkers (FGR, LCP1, and C5AR1). A nomogram was constructed, and its diagnostic value for USCP in uremic patients was evaluated. Additionally, external validation was performed using an additional dataset (GSE28829), which uncovered close relationships between the three hub genes and atherosclerosis progression.

FGR is a member of the protein tyrosine kinases (PTKs) family. FGR gene expression has been shown to be restricted to peripheral blood granulocytes and monocytes, and tissue macrophages [37]. It further contributes to the regulation of immune responses, including neutrophils, monocytes, macrophages, mast cells, phagocytosis, cell adhesion, and migration [38]. Crainciuc et al. [39] revealed that inflammatory injury was prevented in mice through interference with FGR. Meanwhile, Medina et al. [40] found that FGR deficiency in hematopoiesis impeded atherosclerosis plaque formation by limiting endothelial adhesion and transmigration. In the present study, FGR was up-regulated in USCP patients, implying that inflammatory injury and dysregulated immunity resulted in USCP.

LCP1 encodes the actin-binding protein. Although LCP1 is widely expressed across all hematopoietic lineages, numerous tumor cells are highly up-regulated by LCP1. Interestingly, redox-modified LCP1 inhibits tumor cell actin-based functions [41]. Ge et al. [42] discovered that LCP1 promotes osteosarcoma progression via the activation of the JAK2/STAT3 signaling pathway. Zeng et al. [43] reported that several immune marker sets were highly correlated with LCP1, which was associated with tumor-infiltrating lymphocytes in gastric cancer. In macrophages, LCP1 is required for podosome formation and function [44]. Transcriptome analysis of human aorta atherosclerotic lesions indicated that the candidate gene LCP1 participates in phenotypic changes of smooth muscle cells, immune and inflammatory reactions, as well as oxidative processes [45]. In the present study, LCP1 was up-regulated in USCP patients. However, the relationship between LCP1 and USCP is still unknown.

C5AR1, a G protein-coupled receptor for complement C5a, is a potent immune mediator [46]. A crucial component of innate and adaptive immunity is the complement system, which identifies and eliminates pathogens. C5a expression was significantly elevated in patients with severe atherosclerosis, and C5a was an independent predictor of cardiovascular events [47].

Recruitment of circulating monocytes into the artery wall is a key link in the formation of atherosclerosis [48]. C5a is notably a potent neutrophil chemotactic factor; increased C5a receptor expression facilitates the entry of monocytes into the arterial intima [49], which may generate the conditions for atherosclerosis formation. Interestingly, C5a is positively correlated with atherosclerotic plaque instability [50]. The findings of the current study suggested that C5AR1 expression was higher in USCP than in SCP, indicating that the innate and adaptive immune responses were activated, resulting in USCP.

Chronic systemic inflammation and immune activation are hallmarks of uremia [51]. During the development and progression of atherosclerosis, inflammation and immune responses also play a significant role [52]. As is well documented, immune-related inflammatory state is a key mechanism for the occurrence and development of arteriosclerosis in uremic patients [53]. GO analysis indicated that DEGs were mainly related to immune regulation, such as immune system process, intracellular signal transduction, and cell activation. This study also analyzed the immune infiltration of USCP. The results demonstrated that M0 and M1 macrophages were higher in USCP compared with the control, while the proportion of M2 macrophages was lower in SCP. The proportion of immune cells contributes critically to the initiation and progression of USCP [54]. A previous study has established that the proportion of M0 and M1 macrophages is reportedly higher in USCP, while the proportion of M2 macrophages is lower [55], which is consistent with the findings of the current study. M1 macrophages may play a significant role in the initiation and progression of atherosclerosis, while M2 macrophages are hypothesized to stabilize plaques by virtue of their anti-inflammatory effect [56]. In short, focusing on macrophage regulation could provide potential approaches for the treatment of USCP.

The interrelationship between the three identified hub genes and immunity was examined. The three hub genes (FGR, LCP1, and C5AR1) were positively correlated with immune-related M0 and M1 macrophages. Numerous studies have signaled that plaque stability can be increased by inhibiting M1 macrophage polarization [57, 58]. Therefore, immune-related genes are speculated to be highly involved in the progression of USCP. Most importantly, they can assist in the diagnosis of USCP with uremia and the development of targeted therapies.

However, there were several limitations in this study that need to be taken into account. To begin, the sample size for the unstable and stable carotid plaques was not large; only two datasets were available. Secondly, the basic clinical information of samples in different groups was limited which may cause population bias. Thirdly,

although the validation dataset (GSE28829) was used to assess the diagnostic value, the diagnostic biomarkers and underlying mechanisms should be further validated through experimental studies.

Conclusion

Our study systematically identified three candidate hub genes (FGR, LCP1, and C5AR1) and established a nomogram to assist in the diagnosis of USCP with uremia using various bioinformatic analyses and machine learning algorithms. The findings herein provide a theoretical basis for future studies on potential diagnostic candidate genes for USCP in uremic patients. Additionally, immune cell infiltration analysis revealed the dysregulated immune cell proportions and that macrophages may be implicated in the pathogenesis of USCP.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40001-023-01043-4>.

Additional file 1: Table S1. Functional enrichment analysis of Uremia-related DEGs in USCP. **Table S2.** Complete list of DEGs from three algorithms via CytoHubba plug-in. **Figure S1.** Soft threshold selection and gene cluster tree via WGCNA of Uremia. **Figure S2.** Soft threshold selection and gene cluster tree via WGCNA of USCP.

Acknowledgements

Not applicable.

Author contributions

YZ and QZ conceived and designed the study. CL and LT performed comparative analyses using bioinformatics tools. YZ, XT and GZ participated in data analysis and discussions. CL contributed to image processing and data interpretation. All authors read and approved the final manuscript.

Funding

The study was supported by the Medical Scientific Research Foundation of Zhejiang Province of China (2022KY1298).

Availability of data and materials

Datasets used in the study (GSE37171, GSE41571, GSE163154, and GSE28829) can be downloaded without restriction from the public GEO database. Upon reasonable request, data processing procedures are available from the first and corresponding authors.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no conflict of interests.

Author details

¹Department of General Surgery, Division of Vascular Surgery, Shaoxing People's Hospital (Shaoxing Hospital of Zhejiang University), Shaoxing 312000, China. ²Department of Rehabilitation, The First Affiliated Hospital of Anhui Medical University, Anhui Public Health Clinical Center, Hefei 230000, Anhui, China. ³Hepatobiliary Center/Key Laboratory of Liver Transplantation/NHC Key

Laboratory of Living Donor Liver Transplantation, The First Affiliated Hospital of Nanjing Medical University/Chinese Academy of Medical Sciences/Nanjing Medical University), Nanjing 210000, Jiangsu, China. ⁴Shanghai Medical College, Fudan University, Shanghai 200032, China.

Received: 30 October 2022 Accepted: 4 February 2023

Published online: 23 February 2023

References

- Wong HS, Chang CM, Kao CC, Hsu YW, Liu X, Chang WC, et al. V-J combinations of T-cell receptor predict responses to erythropoietin in end-stage renal disease patients. *J Biomed Sci.* 2017;24(1):43. <https://doi.org/10.1186/s12929-017-0349-5>.
- Raggi P, Bellasi A, Ferramosca E, Islam T, Muntner P, Block GA. Association of pulse wave velocity with vascular and valvular calcification in hemodialysis patients. *Kidney Int.* 2007;71(8):802–7. <https://doi.org/10.1038/sj.ki.5002164>.
- Pencak P, Czerwieńska B, Ficek R, Wyskida K, Kujawa-Szewieczek A, Olszanecka-Glinianowicz M, et al. Calcification of coronary arteries and abdominal aorta in relation to traditional and novel risk factors of atherosclerosis in hemodialysis patients. *BMC Nephrol.* 2013;14:10. <https://doi.org/10.1186/1471-2369-14-10>.
- London GM. Vascular disease and atherosclerosis in uremia. *Nefrologia.* 2005;25(Suppl 2):91–5.
- Bechler SL, Si Y, Yu Y, Ren J, Liu B, Lynn DM. Reduction of intimal hyperplasia in injured rat arteries promoted by catheter balloons coated with polyelectrolyte multilayers that contain plasmid DNA encoding PKC δ . *Biomaterials.* 2013;34(1):226–36. <https://doi.org/10.1016/j.biomaterials.2012.09.010>.
- de Ruijter J, van Sambeek M, van de Vosse F, Lopata R. Automated 3D geometry segmentation of the healthy and diseased carotid artery in free-hand, probe tracked ultrasound images. *Med Phys.* 2020;47(3):1034–47. <https://doi.org/10.1002/mp.13960>.
- Okamoto Y, Ishii S, Croce K, Katsumata H, Fukushima M, Kihara S, et al. Adiponectin inhibits macrophage tissue factor, a key trigger of thrombosis in disrupted atherosclerotic plaques. *Atherosclerosis.* 2013;226(2):373–7. <https://doi.org/10.1016/j.atherosclerosis.2012.12.012>.
- Avramovski P, Janakievska P, Sotiroski K, Sikole A. Accelerated progression of arterial stiffness in dialysis patients compared with the general population. *Korean J Intern Med.* 2013;28(4):464–74. <https://doi.org/10.3904/kjim.2013.28.4.464>.
- Shen Y, Yuan Z, Yin A, Liu Y, Xiao Y, Wu Y, et al. Antiatherogenic effect of pioglitazone on uremic apolipoprotein E knockout mice by modulation of the balance of regulatory and effector T cells. *Atherosclerosis.* 2011;218(2):330–8. <https://doi.org/10.1016/j.atherosclerosis.2011.07.112>.
- Daskalopoulou SS, Daskalopoulos ME, Perrea D, Nicolaidis AN, Liapis CD. Carotid artery atherosclerosis: what is the evidence for drug action? *Curr Pharm Des.* 2007;13(11):1141–59. <https://doi.org/10.2174/138161207780619019>.
- Yoo SM, Choi JH, Lee SY, Yoo NC. Applications of DNA microarray in disease diagnostics. *J Microbiol Biotechnol.* 2009;19(7):635–46.
- Xin G, Zhou G, Zhang W, Zhang X. Construction and validation of predictive model to identify critical genes associated with advanced kidney disease. *Int J Genomics.* 2020;2020:7524057. <https://doi.org/10.1155/2020/7524057>.
- Bao MH, Zhang RQ, Huang XS, Zhou J, Guo Z, Xu BF, et al. Transcriptomic and proteomic profiling of human stable and unstable carotid atherosclerotic plaques. *Front Genet.* 2021;12:755507. <https://doi.org/10.3389/fgene.2021.755507>.
- Yao M, Zhang C, Gao C, Wang Q, Dai M, Yue R, et al. Exploration of the shared gene signatures and molecular mechanisms between systemic lupus erythematosus and pulmonary arterial hypertension: evidence from transcriptome data. *Front Immunol.* 2021;12:658341. <https://doi.org/10.3389/fimmu.2021.658341>.
- Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol.* 2016;1418:93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.

16. Scherer A, Günther OP, Balshaw RF, Hollander Z, Wilson-McManus J, Ng R, et al. Alteration of human blood cell transcriptome in uremia. *BMC Med Genomics*. 2013;6:23. <https://doi.org/10.1186/1755-8794-6-23>.
17. Lee K, Santibanez-Koref M, Polvikoski T, Birchall D, Mendelow AD, Keavney B. Increased expression of fatty acid binding protein 4 and leptin in resident macrophages characterises atherosclerotic plaque rupture. *Atherosclerosis*. 2013;226(1):74–81. <https://doi.org/10.1016/j.atherosclerosis.2012.09.037>.
18. Jin H, Goossens P, Juhasz P, Eijgelaar W, Manca M, Karel JMH, et al. Integrative multiomics analysis of human atherosclerosis reveals a serum response factor-driven network associated with intraplaque hemorrhage. *Clin Transl Med*. 2021;11(6):e458. <https://doi.org/10.1002/ctm2.458>.
19. Döring Y, Manthey HD, Drechsler M, Lievens D, Megens RT, Soehlein O, et al. Auto-antigenic protein-DNA complexes stimulate plasmacytoid dendritic cells to promote atherosclerosis. *Circulation*. 2012;125(13):1673–83. <https://doi.org/10.1161/circulationaha.111.046755>.
20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. <https://doi.org/10.1093/nar/gkv007>.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. <https://doi.org/10.1186/1471-2105-9-559>.
22. Resource TGO. 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–d338. <https://doi.org/10.1093/nar/gky1055>.
23. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353–d361. <https://doi.org/10.1093/nar/gkw1092>.
24. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7. <https://doi.org/10.1089/omi.2011.0118>.
25. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49(D1):D605–d612. <https://doi.org/10.1093/nar/gkaa1074>.
26. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol*. 2019;20(1):185. <https://doi.org/10.1186/s13059-019-1758-4>.
27. Su W, Zhao Y, Wei Y, Zhang X, Ji J, Yang S. Exploring the Pathogenesis of Psoriasis Complicated With Atherosclerosis via Microarray Data Analysis. *Front Immunol*. 2021;12:667690. <https://doi.org/10.3389/fimmu.2021.667690>.
28. Huang ML, Hung YH, Lee WM, Li RK, Jiang BR. SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *ScientificWorldJournal*. 2014;2014:795624. <https://doi.org/10.1155/2014/795624>.
29. Yang C, Delcher C, Shenkman E, Ranka S. Machine learning approaches for predicting high cost high need patient expenditures in health care. *Biomed Eng Online*. 2018;17(Suppl 1):131. <https://doi.org/10.1186/s12938-018-0568-3>.
30. Blanchet L, Vitale R, van Vorstenbosch R, Stavropoulos G, Pender J, Jonkers D, et al. Constructing bi-plots for random forest: Tutorial. *Anal Chim Acta*. 2020;1131:146–55. <https://doi.org/10.1016/j.jaca.2020.06.043>.
31. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7. <https://doi.org/10.1038/nmeth.3337>.
32. Hu K. Become competent within one day in generating boxplots and violin plots for a novice without prior R experience. *Methods Protoc*. 2020. <https://doi.org/10.3390/mps3040064>.
33. Chang CT, Shen MY, Hsieh JY, Chang CM, Liao HY, Chen FY, et al. Increased electronegativity of high-density lipoprotein in uremia patients impairs its functional properties and is associated with the risk of coronary artery disease. *Atherosclerosis*. 2018;278:147–55. <https://doi.org/10.1016/j.atherosclerosis.2018.09.009>.
34. Sarnak MJ, Amann K, Bangalore S, Cavalante JL, Charytan DM, Craig JC, et al. Chronic kidney disease and coronary artery disease: JACC state-of-the-art review. *J Am Coll Cardiol*. 2019;74(14):1823–38. <https://doi.org/10.1016/j.jacc.2019.08.1017>.
35. François, Dialysis MJSI. THE CLINICAL EPIDEMIOLOGY OF CARDIOVASCULAR DISEASES IN CHRONIC KIDNEY DISEASE: Uremia-Related Metabolic Cardiac Risk Factors in Chronic Kidney Disease. 2003.
36. Tsao CW, Aday AW, Almarazooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the american heart association. *Circulation*. 2022;145(8):e153–639. <https://doi.org/10.1161/cir.0000000000001052>.
37. Patel M, Faulkner L, Katz DR, Brickell PM. The c-fgr proto-oncogene: expression in Epstein-Barr-virus-infected B lymphocytes and in cells of the myelomonocytic and granulocytic lineages. *Pathobiology*. 1991;59(4):289–92. <https://doi.org/10.1159/000163665>.
38. Gutkind JS, Robbins KC. Translocation of the FGR protein-tyrosine kinase as a consequence of neutrophil activation. *Proc Natl Acad Sci USA*. 1989;86(22):8783–7. <https://doi.org/10.1073/pnas.86.22.8783>.
39. Crainiciuc G, Palomino-Segura M, Molina-Moreno M, Sicilia J, Aragones DG, Li JLY, et al. Behavioural immune landscapes of inflammation. *Nature*. 2022;601(7893):415–21. <https://doi.org/10.1038/s41586-021-04263-y>.
40. Medina I, Cougoule C, Drechsler M, Bermudez B, Koenen RR, Sluimer J, et al. Hck/Fgr kinase deficiency reduces plaque growth and stability by blunting monocyte recruitment and intraplaque motility. *Circulation*. 2015;132(6):490–501. <https://doi.org/10.1161/circulationaha.114.012316>.
41. Balta E, Hardt R, Liang J, Kirchgessner H, Orlik C, Jahraus B, et al. Spatial oxidation of L-plastin downmodulates actin-based functions of tumor cells. *Nat Commun*. 2019;10(1):4073. <https://doi.org/10.1038/s41467-019-11909-z>.
42. Ge X, Liu W, Zhao W, Feng S, Duan A, Ji C, et al. Exosomal transfer of LCP1 promotes osteosarcoma cell tumorigenesis and metastasis by activating the JAK2/STAT3 signaling pathway. *Mol Ther Nucleic Acids*. 2020;21:900–15. <https://doi.org/10.1016/j.omtn.2020.07.025>.
43. Zeng Q, Li L, Feng Z, Luo L, Xiong J, Jie Z, et al. LCP1 is a prognostic biomarker correlated with immune infiltrates in gastric cancer. *Cancer Biomark*. 2021;30(1):105–25. <https://doi.org/10.3233/cbm-200006>.
44. De Clercq S, Boucherie C, Vandekerckhove J, Gettemans J, Guillabert A. L-plastin nanobodies perturb matrix degradation, podosome formation, stability and lifetime in THP-1 macrophages. *PLoS ONE*. 2013;8(11):e78108. <https://doi.org/10.1371/journal.pone.0078108>.
45. Timofeeva AV, Goriunova LE, Khaspekov GL, Il'inskaia OP, Sirotkin VN, Andreeva ER, et al. Comparative transcriptome analysis of human aorta atherosclerotic lesions and peripheral blood leukocytes from essential hypertension patients. *Kardiologiia*. 2009;49(9):27–38.
46. Li K, Wu KY, Wu W, Wang N, Zhang T, Choudhry N, et al. C5aR1 promotes acute pyelonephritis induced by uropathogenic *E. coli*. *JCI Insight*. 2017. <https://doi.org/10.1172/jci.insight.97626>.
47. Speidl WS, Katsaros KM, Kastl SP, Zorn G, Huber K, Maurer G, et al. Coronary late lumen loss of drug eluting stents is associated with increased serum levels of the complement components C3a and C5a. *Atherosclerosis*. 2010;208(1):285–9. <https://doi.org/10.1016/j.atherosclerosis.2009.07.030>.
48. van Tuijl J, Joosten LAB, Netea MG, Bekkering S, Rixsen NP. Immunometabolism orchestrates training of innate immunity in atherosclerosis. *Cardiovasc Res*. 2019;115(9):1416–24. <https://doi.org/10.1093/cvr/cvz107>.
49. Masias C, Vasu S, Cataland SR. None of the above: thrombotic microangiopathy beyond TTP and HUS. *Blood*. 2017;129(21):2857–63. <https://doi.org/10.1182/blood-2016-11-743104>.
50. Wezel A, de Vries MR, Lagrauw HM, Foks AC, Kuiper J, Quax PH, et al. Complement factor C5a induces atherosclerotic plaque disruptions. *J Cell Mol Med*. 2014;18(10):2020–30. <https://doi.org/10.1111/jcmm.12357>.
51. Vaziri ND, Pahl MV, Crum A, Norris K. Effect of uremia on structure and function of immune system. *J Ren Nutr*. 2012;22(1):149–56. <https://doi.org/10.1053/j.jrn.2011.10.020>.
52. He B, Lu Z, Jiang H. Interferon regulatory factors: New targets for intervention of cardiovascular diseases. *Int J Cardiol*. 2015;181:355–6. <https://doi.org/10.1016/j.ijcard.2014.12.084>.
53. Kalil RS, Flanagan M, Stanford W, Haynes WG. Dissociation between progression of coronary artery calcification and endothelial function in hemodialysis patients: a prospective pilot study. *Clin Nephrol*. 2012;78(1):1–9. <https://doi.org/10.5414/cn.106830>.
54. Gao J, Shi L, Gu J, Zhang D, Wang W, Zhu X, et al. Difference of immune cell infiltration between stable and unstable carotid artery atherosclerosis. *J Cell Mol Med*. 2021;25(23):10973–9. <https://doi.org/10.1111/jcmm.17018>.
55. Chinetti-Gbaguidi G, Colin S, Staels B. Macrophage subsets in atherosclerosis. *Nat Rev Cardiol*. 2015;12(1):10–7. <https://doi.org/10.1038/nrcardio.2014.173>.

56. Bories GFP, Leitinger N. Macrophage metabolism in atherosclerosis. *FEBS Lett.* 2017;591(19):3042–60. <https://doi.org/10.1002/1873-3468.12786>.
57. Cai Y, Wen J, Ma S, Mai Z, Zhan Q, Wang Y, et al. Huang-Lian-Jie-Du decoction attenuates atherosclerosis and increases plaque stability in high-fat diet-induced ApoE(-/-) mice by inhibiting M1 macrophage polarization and promoting M2 macrophage polarization. *Front Physiol.* 2021;12:666449. <https://doi.org/10.3389/fphys.2021.666449>.
58. Zhang X, Liu MH, Qiao L, Zhang XY, Liu XL, Dong M, et al. Ginsenoside Rb1 enhances atherosclerotic plaque stability by skewing macrophages to the M2 phenotype. *J Cell Mol Med.* 2018;22(1):409–16. <https://doi.org/10.1111/jcmm.13329>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

