Identification of Novel Diagnostic Markers for Atherosclerosis Using Machine-Learning Algorithms

Yanshuang Cheng[1] and Yang Shao[2]

[1]*Department of Neurosurgery, The First Hospital of China Medical University, Liaoning, China*
[2]*Department of Cardiology, The First Hospital of China Medical University, Liaoning, China*

ABSTRACT

Objective: To outline immune-cell infiltration and identify diagnostic genes for atherosclerosis (AS) to better understand the potential molecular processes involved in AS development.

Study Design: Descriptive study.

Place and Duration of the Study: Department of Cardiology, The First Hospital of China Medical University, Shenyang, Liaoning, China, from 10th June to 8th October 2024.

Methodology: Relevant datasets were collected from the Gene Expression Omnibus database. Gene set enrichment analysis was conducted on differentially expressed genes (DEGs). Subsequently, three machine-learning algorithms were used to identify the core genes. Receiver operating characteristic (ROC) curves were used to analyse the clinical diagnostic value of the core genes.

Results: A Total of 3,307 DEGs, which were found primarily enriched in inflammation-related pathways. Further analysis of the core genes using three machine-learning algorithms revealed four intersecting genes, *IBSP, PI16*, *MYOC*, and *IGLL5*, which are all inflammation-related genes; they also showed good clinical diagnostic abilities, which were verified using ROC curves (area under the curve: 0.959, 0.946, 0.931, and 0.880, respectively).

Conclusion: *IBSP*, *PI16*, *MYOC*, and *IGLL5* participate in AS pathogenesis by regulating inflammatory reactions. These are novel diagnostic markers and are expected to become potential targets for AS-targeted therapies.

Key Words: *Atherosclerosis, Inflammatory reaction, Machine-learning algorithms, Bioinformatic*

How to cite this article: Cheng Y, Shao Y. Identification of Novel Diagnostic Markers for Atherosclerosis Using Machine-Learning Algorithms. *J Coll Physicians Surg Pak* 2025; **35(05)**:574-579.

INTRODUCTION

The incidence rate of various cardiovascular and cerebrovascular diseases continues to rise, posing a great threat to the safety of patients.[1,2] Atherosclerosis (AS) plays an important role in all these diseases.[3,4] AS often involves the deposition of lipid plaques in the blood on the arterial wall. With the continuous release of inflammatory factors, the arterial wall is in a chronic inflammatory reaction state,[5] ultimately leading to plaque rupture, thrombosis, and luminal stenosis, which in turn trigger a series of major adverse cardiovascular events (MACEs).[6] Currently, the main treatment strategy for AS is the use of statins, which can lower low-density lipoprotein cholesterol levels.[7] However, this therapy has not effectively reduced the occurrence of MACEs.[8] Therefore, deciphering the pathogenic mechanism of AS can help guide and improve the effects of clinical diagnosis, treatment, and outcomes.

Correspondence to: Ms. Yang Shao, Department of Cardiology, The First Hospital of China Medical University, Shenyang, Liaoning, China
E-mail: shaoyang0705_sy@126.com

Received: November 27, 2024; Revised: April 21, 2025; Accepted: April 29, 2025
DOI: https://doi.org/10.29271/jcpsp.2025.05.574

The concept of inflammation, which plays a key role in AS pathogenesis, has received increasing attention over the years, and many researchers have attempted to combine immune and anti-inflammatory treatments to reduce MACEs.[9] For example, the anti-inflammatory medicine canakinumab, which is an interleukin-1β-specific antibody, can significantly reduce the occurrence of MACEs.[10] The aim of this study was to outline immune-cell infiltration and identify diagnostic genes for AS, *via* machine-learning algorithms, to better understand the potential molecular processes involved in AS development.

METHODOLOGY

Using the GPL17077 platform of the gene expression omnibus (GEO) database; available from: (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi), GSE100927, which contains 69 AS tissues and 35 control tissues, was obtained. Inclusion criteria were AS patients with definite diagnoses and signed informed consent. Exclusion criteria were patients with non-AS peripheral artery disease, thrombosis or restenosis.

Differentially expressed genes (DEGs) were identified using the limma R package in R studio (version 4.1.1; R Foundation for Statistical Computing, Vienna, Austria) with a statistical threshold of logFC >1 and a false discovery rate of <0.05.

Gene set enrichment analysis (GSEA), available from: https://www.broadinstitute.org/gsea was performed to identify DEGs and their enrichment functions between the two queues to determine the impact of synergistic gene changes on diseases within the gene set.[11]

Three machine-learning algorithms, least absolute shrinkage and selection operator (LASSO),[12] support vector machine recursive feature elimination (SVM-RFE) analysis,[13] and random forest (RF) analysis,[14] were used to investigate the genes that play crucial roles in AS pathogenesis. The core genes were obtained by the intersection of the three algorithms.

Single-sample gene set enrichment analysis (ssGSEA) was based on 29 sets of immune genes to comprehensively evaluate the differences in immune characteristics between the two groups.[15]

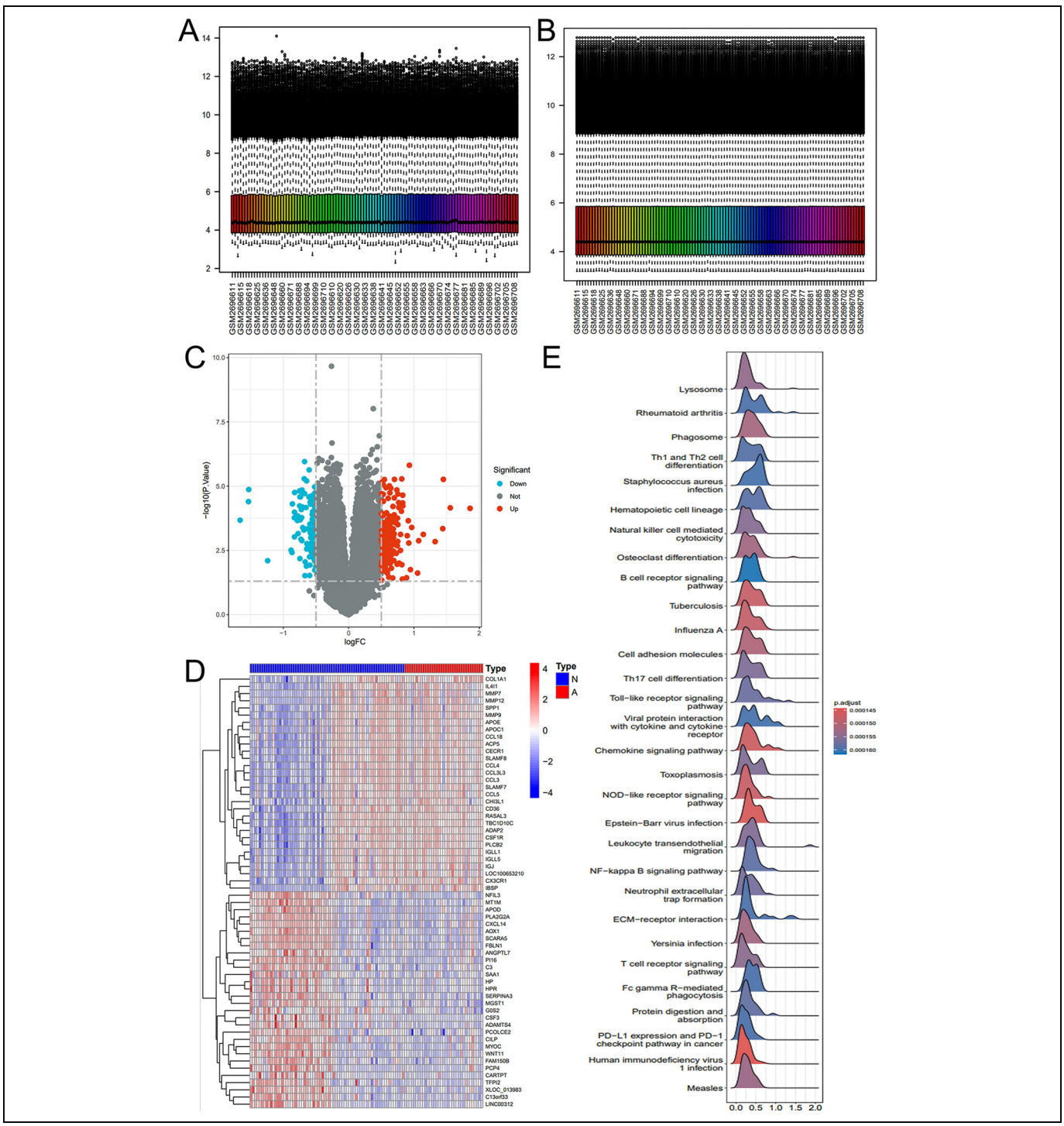

Figure 1: AS pathogenesis may be closely related to inflammation infiltration. (A) GSE100927 dataset, containing 69 AS and 35 control tissues. (B) GSE100927 dataset after calibration processing. (C) Volcano plot showed up- and down-regulated DEGs between AS and control tissues, respectively. (D) Heat map showed the top 30 DEGs. (E) The result of GSEA analysis.

The t-test was used to compare data between different groups. Statistical analysis and data visualisation were performed using the R Studio Software version 4.1.1, and GraphPad Prism version 8.0; GraphPad Software, Boston, MA, USA, respectively. Receiver operating characteristic (ROC) curves were drawn, and the area under the ROC curve (AUC) was calculated using the survival ROC R package. Spearman's rank correlation coefficient was used to evaluate the correlation between the two groups. Statistical significance was set at $p < 0.05$.

RESULTS

GSE100927, which contains 69 AS tissues, was obtained (Figure 1A), and calibration was performed for subsequent analyses (Figure 1B). First, differential expression analysis of AS and control tissue data was conducted, demonstrating that 1,526 and 1,781 genes were up- and down- regulated, respectively (Figure 1C). The top 30 DEGs were displayed on a heat map (Figure 1D). Simultaneously, GSEA was performed to explore the functional differences between AS and control tissues. It was observed that DEGs were mainly enriched in inflammation-related pathways, including Th1 and Th2 cell differentiation, Toll-like receptor signalling pathway, and nucleotide-binding oligomerisation domain (NOD) such as receptor signalling pathway (Figure 1E). These results suggested that AS pathogenesis is closely related to inflammatory infiltration.

Three machine-learning algorithms were used to further investigate the genes that play crucial roles in this process (Figure 2A-C) and successfully intersected four genes: *IBSP*, *PI16*, *MYOC*, and *IGLL5* (Figure 2D, Table I).

Subsequently, gene function analysis of these four core genes was predicted using GSEA, the results showed that *IBSP* was mainly associated with endocrine imbalance and abnormal fat metabolism (Figure 3A, B); *PI16* was mainly associated with cholesterol and amino acid metabolism (Figure 3C, D); *MYOC* was mainly associated with lysosomes and cardiovascular disease (Figure 3E, F); *IGLL5* was mainly associated with glucose and lipid metabolism and inflammatory reactions (Figure 3G, H). In conclusion, these functional abnormalities can contribute to the development of AS.

Nomogram analysis of *IBSP*, *PI16*, *MYOC*, and *IGLL5* was conducted, all of which showed good diagnostic performances. In addition, ROC analysis was conducted on *IBSP*, *PI16*, *MYOC*, and *IGLL5* to determine their potential as diagnostic biomarkers of AS (AUC: 0.959, 0.946, 0.931, and 0.880, respectively).

Based on these findings, the four identified core genes were all found to be related to inflammation. Therefore, ssGSEA was used to explore the specific relationships between the four genes and inflammatory cells and pathways. The result showed that they were significantly correlated with numerous immune-immersion-related pathways. Further- more, Spearman's rank correlation coefficient revealed that each individual gene was significantly associated with immune-inflammatory reaction-related pathways (Table II). These results indicate that inflammatory cells and factors play important roles in AS pathogenesis.

DISCUSSION

This study obtained and calibrated the AS data from a public database, identified the DEGs between the AS and control tissues, performed GSEA on DEGs, and detected that these genes were strongly linked to the inflammatory reaction pathways. The intersecting genes of three machine-learning algorithms were *IBSP*, *PI16*, *MYOC*, and *IGLL5*, suggesting they may be core genes in the occurrence and development of AS. The possible regulatory functions of these four genes revealed that they were all related to inflammatory reactions and metabolic abnormalities, which is consistent with AS pathogenesis.[16,17] Current research indicates a causal relationship between *IBSP*, *MYOC*, and AS, but the specific regulatory mechanisms remain to be elucidated.[18,19] Studies have shown that under high endothelial shear stress, *PI16* inhibits protease activity, and thus plays a protective role during inflammation.[20] Simultaneously, *PI16* also participates in cholesterol transfer and affects AS development.[21] Relatively few reports exist on *IGLL5*; currently, research mainly focuses on its role in haematological and autoimmune diseases.[22,23] It is believed that with further research, the regulatory network between *IGLL5* and AS will gradually be clarified.

Table I: The gene results of three machine-learning algorithms.

LASSO	SVM-REF	RF
IBSP	IGLL5	IBSP
PI16	PI16	PI16
MYOC	IBSP	IGLL5
IGLL5	ACP5	CCL3
-	MYOC	ACP5
-	PLA2G2A	MYOC
-	CCL3	MMP9
-	MMP9	PLA2G2A
-	SPP1	CHI3L1
-	CHI3L1	SPP1

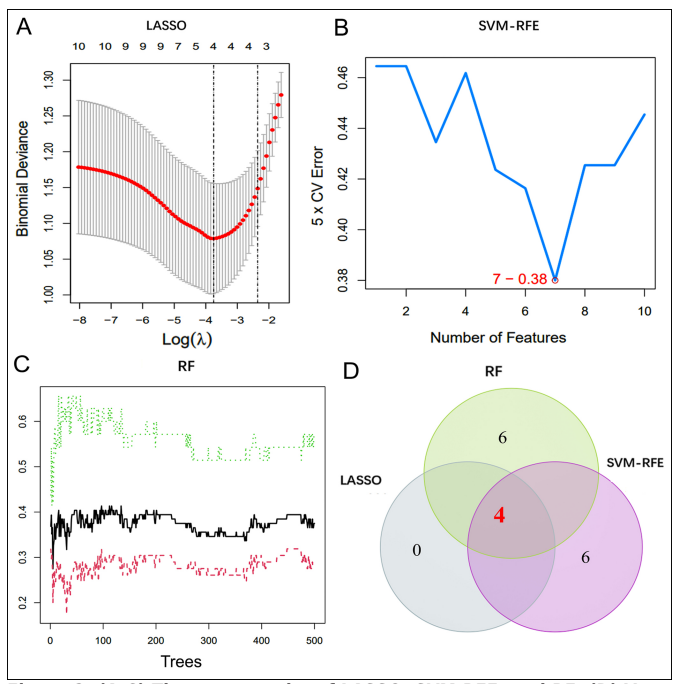

Figure 2: (A-C) The gene results of LASSO, SVM-RFE, and RF. (D) Venn diagram showed four genes, *IBSP*, *PI16*, *MYOC*, and *IGLL5* were intersected.

Table II: The p-value of core genes in different pathways of Spearman's rank correlation.

Id	IBSP	IGLL5	MYOC	PI16
Hallmark_xenobiotic_metabolism			0.032	
Hallmark_wnt_beta_catenin_signalling			p <0.001	
Hallmark_uv_response_up			0.027	0.034
Hallmark_uv_response_dn	0.003	0.017	p <0.001	0.019
Hallmark_unfolded_protein_response	0.032		0.036	
Hallmark_tnfa_signalling_*via*_nfkb				
Hallmark_tgf_beta_signalling		0.006	p <0.001	
Hallmark_spermatogenesis			0.006	
Hallmark_reactive_oxygen_species_pathway			0.004	
Hallmark_protein_secretion			0.008	
Hallmark_pi3k_akt_mtor_signalling			p <0.001	
Hallmark_peroxisome			0.005	
Hallmark_pancreas_beta_cells				
Hallmark_p53_pathway				
Hallmark_oxidative_phosphorylation			0.041	
Hallmark_notch_signalling			0.004	
Hallmark_myogenesis	0.021	0.028	p <0.001	0.003
Hallmark_myc_targets_v2			p <0.001	
Hallmark_myc_targets_v1	0.041		0.003	
Hallmark_mtorc1_signalling			p <0.001	
Hallmark_mitotic_spindle				
Hallmark_kras_signalling_up				
Hallmark_kras_signalling_dn			p <0.001	0.013
Hallmark_interferon_gamma_response		p <0.001	0.017	
Hallmark_interferon_alpha_response		p <0.001	0.020	
Hallmark_inflammatory_response		0.039	0.035	
Hallmark_il6_jak_stat3_signalling	0.011	0.007	0.021	
Hallmark_il2_stat5_signalling		0.041		
Hallmark_hypoxia				
Hallmark_heme_metabolism	0.006			
Hallmark_hedgehog_signalling	p <0.001	0.019	p <0.001	0.002
Hallmark_glycolysis			p <0.001	
Hallmark_g2m_checkpoint			p <0.001	
Hallmark_fatty_acid_metabolism				0.032
Hallmark_estrogen_response_late			p <0.001	p <0.001
Hallmark_estrogen_response_early			p <0.001	p <0.001
Hallmark_epithelial_mesenchymal_transition		0.042	0.031	0.007
Hallmark_e2f_targets			p <0.001	
Hallmark_dna_repair			p <0.001	
Hallmark_complement	0.020	0.002	p <0.001	
Hallmark_coagulation	0.042			
Hallmark_cholesterol_homeostasis			p <0.001	
Hallmark_bile_acid_metabolism				0.008
Hallmark_apoptosis				
Hallmark_apical_surface			p <0.001	0.013
Hallmark_apical_junction		0.013	p <0.001	0.004
Hallmark_angiogenesis				0.043
Hallmark_androgen_response				
Hallmark_allograft_rejection	0.001	p <0.001	p <0.001	
Hallmark_adipogenesis				0.027

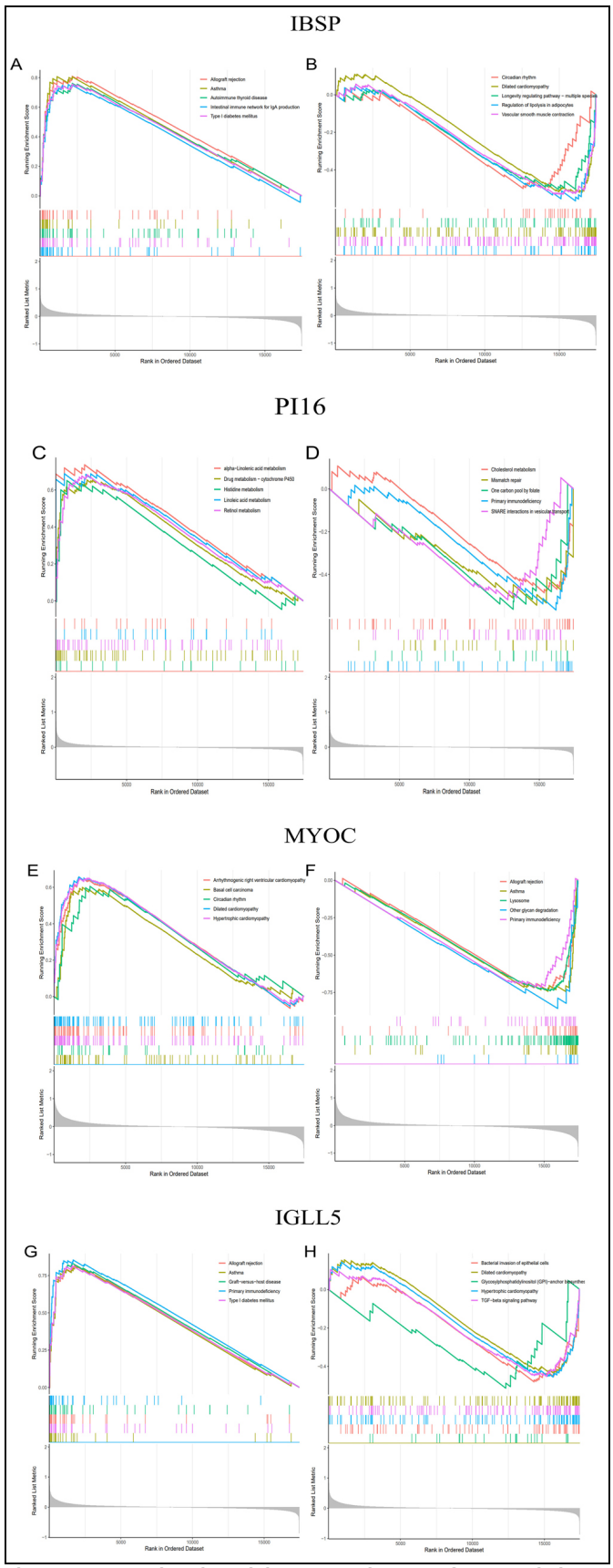

Figure 3: Functional enrichment pathways of up- and down-regulation of *IBSP* (A-B); *PI16* (C-D); *MYOC* E-F; *IGLL5 (*G-H), using GSEA.

ROC analysis was conducted to further explore the clinical translational value of *IBSP*, *PI16*, *MYOC*, *IGLL5*, and the AUC values of the four all exceeded (0.85), demonstrating good diagnostic performance. Finally, an immune infiltration analysis of the four genes was conducted, and it was found that both the integration and individual analyses showed a strong correlation with immune-related signalling pathways. This confirmed that *IBSP, PI16*, *MYOC*, and *IGLL5* may participate in AS formation and progression by regulating inflammatory reactions.

This study has some limitations. Although various bio-informatics methods were utilised to identify potential AS-related regulatory genes, further basic research is needed to validate the results. *IBSP*, *PI16*, *MYOC*, and *IGLL5* were identified as potential novel diagnostic markers for AS, however, large-sample clinical patient data is required to validate the findings.

CONCLUSION

IBSP, *PI16*, *MYOC*, and *IGLL5* were found to participate in AS pathogenesis by regulating inflammatory reactions. Thus, they can be considered novel diagnostic markers for AS and are expected to become new targets for AS-targeted therapy.

ETHICAL APPROVAL:

The data for this study were acquired from public databases, did not involve the testing of human and animal samples. Therefore, ethical approval is not applicable.

PATIENTS' CONSENT:

Informed consent was obtained from the participants included in this study.

COMPETING INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

YC: Data acquisition, formal analysis, and original draft.
YS: Supervision, data validation, review, and revision of the work.
Both authors approved the final version of the manuscript to be published.

REFERENCES

1. Benn M, Schwartz M, Nordestgaard BG, Tybjaerg-Hansen A. Mitochondrial haplogroups: Ischemic cardiovascular disease, other diseases, mortality, and longevity in the general population. *Circulation* 2008; **117(19)**:2492-501. doi: 10.1161/CIRCULATIONAHA.107.756809.
2. Hartley A, Marshall DC, Salciccioli JD, Sikkel MB, Maruthappu M, Shalhoub J. Trends in mortality from ischemic heart disease and cerebrovascular disease in Europe: 1980 to 2009. *Circulation* 2016; **133(20)**:1916-26. doi: 10.1161/CIRCULATIONAHA.115.018931.

3. Lavallee PC, Charles H, Albers GW, Caplan LR, Donnan GA, Ferro JM, *et al*. Effect of atherosclerosis on 5-year risk of major vascular events in patients with transient ischemic attack or minor ischemic stroke: An international prospective cohort study. *Lancet Neurol* 2023; **22(4)**:320-9. doi: 10.1016/S1474-4422(23)00067-4.
4. Fani L, van der Willik KD, Bos D, Leening MJ, Koudstaal PJ, Rizopoulos D, *et al*. The association of innate and adaptive immunity, subclinical atherosclerosis, and cardiovascular disease in the Rotterdam study: A prospective cohort study. *PLoS Med* 2020; **17(5)**:e1003115. doi: 10.1371/journal.pmed.1003115.
5. Kong P, Cui ZY, Huang XF, Zhang DD, Guo RJ, Han M. Inflammation and atherosclerosis: signalling pathways and therapeutic intervention. *Signal Transduct Target Ther* 2022; **7(1)**:131. doi: 10.1038/s41392-022-00955-7.
6. Badimon L, Vilahur G. Thrombosis formation on atherosclerotic lesions and plaque rupture. *J Intern Med* 2014; **276(6)**:618-32. doi: 10.1111/joim.12296.
7. Pinkosky SL, Newton RS, Day EA, Ford RJ, Lhotak S, Austin RC, *et al*. Liver-specific ATP-citrate lyase inhibition by bem-pedoic acid decreases LDL-C and attenuates atherosclerosis. *Nat Commun* 2016; **7**:13457. doi: 10.1038/ncomms13457.
8. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, *et al*. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med* 2017; **376(18)**:1713-22. doi: 10.1056/NEJMoa1615664.
9. Gistera A, Hansson GK. The immunology of atherosclerosis. *Nat Rev Nephrol* 2017; **13(6)**:368-80. doi: 10.1038/nrneph.2017.51.
10. Duivenvoorden R, Senders ML, van Leent MMT, Perez-Medina C, Nahrendorf M, Fayad ZA, *et al*. Nano immuno- therapy to treat ischemic heart disease. *Nat Rev Cardiol* 2019; **16(1)**:21-32. doi: 10.1038/s41569-018-00 73-1.
11. Hu Y, Comjean A, Attrill H, Antonazzo G, Thurmond J, Chen W, *et al*. PANGEA: A new gene set enrichment tool for Drosophila and common research organisms. *Nucleic Acids Res* 2023; **51(W1)**:W419-26. doi: 10.1093/nar/gkad331.
12. Zhang G, Su L, Lv X, Yang QK. A novel tumor doubling time-related immune gene signature for prognosis prediction in hepatocellular carcinoma. *Cancer Cell Int* 2021; **21(1)**:522. doi: 10.1186/s12935-021-02227-w.
13. Zhang Z, Wang S, Zhu Z, Nie B. Identification of potential feature genes in non-alcoholic fatty liver disease using bioinformatics analysis and machine learning strategies. *Comput Biol Med* 2023; **157**:106724. doi: 10.1016/j.compbiomed. 2023.106724.
14. Ohanyan H, van de Wiel M, Portengen L, Wagtendonk A, den Braver NR, de Jong TR, *et al*. Exposome-wide association study of body mass index using a novel meta-analytical approach for random forest models. *Environ Health Perspect* 2024; **132(6)**:67007. doi: 10.1289/EHP13393.
15. Chen Y, Feng Y, Yan F, Zhao Y, Zhao H, Guo Y, *et al*. A novel immune-related gene signature to identify the tumor microenvironment and prognose disease among patients with oral squamous cell carcinoma patients using ssGSEA: A bioinformatics and biological validation study. *Front Immunol* 2022; **13**:922195. doi: 10.3389/fimmu.2022.922195.
16. Peng Z, Ye M, Ding H, Feng Z, Hu K. Spatial transcriptomics atlas reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment components in colorectal cancer. *J Transl Med* 2022; **20(1)**:302. doi: 10.1186/s12967-022-03510-8.
17. Ali L, Schnitzler JG, Kroon J. Metabolism: The road to inflammation and atherosclerosis. *Curr Opin Lipidol* 2018; **29(6)**:474-80. doi: 10.1097/MOL.0000000000000550.
18. Ye Z, Wang XK, Lv YH, Wang X, Cui YC. The integrated analysis identifies three critical genes as novel diagnostic biomarkers involved in immune infiltration in atherosclerosis. *Front Immunol* 2022; **13**:905921. doi: 10.3389/ fimmu.2022.905921.
19. Arslan S, Sahin NO, Bayyurt B, Berkan O, Yilmaz MB, Asam M, *et al*. Role of lncRNAs in remodeling of the coronary artery plaques in patients with atherosclerosis. *Mol Diagn Ther* 2023; **27(5)**:601-10. doi: 10.1007/s40291-023-00659-w.
20. Suarez-Rivero JM, Pastor-Maldonado CJ, de la Mata M, Villanueva-Paz M, Povea-Cabello S, Alvarez-Cordoba M, *et al*. Atherosclerosis and Coenzyme Q_{10}. *Int J Mol Sci* 2019; **20(20)**:5195. doi: 10.3390/ijms20205195.
21. Aguilar-Ballester M, Herrero-Cervera A, Vinue A, Martinez-Hervas S, Gonzalez-Navarro H. Impact of cholesterol metabolism in immune cell function and atherosclerosis. *Nutrients* 2020; **12(7)**:2021. doi: 10.3390/nu12072021.
22. Bernstein N, Chapman MS, Nyamondo K, Chen Z, Williams N, Mitchell E, *et al*. Analysis of somatic mutations in whole blood from 200,618 individuals identifies pervasive positive selection and novel drivers of clonal hematopoiesis. *Nat Genet* 2024; **56(6)**:1147-55. doi: 10.1038/s41588-024-01755-1.
23. Wu B, He Y, Yang D, Liu RX. Identification of hub genes and therapeutic drugs in rheumatoid arthritis patients. *Clin Rheumatol* 2021; **40(8)**:3299-309. doi: 10.1007/s10067-021-05650-6.