International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Peripheral Blood HIST1H2AE is a Candidate Epigenetic Biomarker for Coronary Artery Disease: A Multi-Dataset Discovery and Comparative Validation Study

Liugang Xu[1,*], Yajun Wang[1,*], Hongyun Ji[1], Yisi Shan[2], Jinhu Zhang[3], Li Yang[4], Wei Du[1], Jianyu Jiang[1], Jin Chen[1], Ruihong Cao[1]

[1]Department of Cardiology, Zhangjiagang Hospital of Traditional Chinese Medicine Affiliated to Nanjing University of Traditional Chinese Medicine, Zhangjiagang, Jiangsu, 215600, People's Republic of China; [2]Department of Neurology, Zhangjiagang Hospital of Traditional Chinese Medicine Affiliated to Nanjing University of Traditional Chinese Medicine, Zhangjiagang, Jiangsu, 215600, People's Republic of China; [3]Department of Urology, Zhangjiagang Hospital of Traditional Chinese Medicine Affiliated to Nanjing University of Traditional Chinese Medicine, Zhangjiagang, Jiangsu, 215600, People's Republic of China; [4]Department of Gynecology, Zhangjiagang Hospital of Traditional Chinese Medicine Affiliated to Nanjing University of Traditional Chinese Medicine, Zhangjiagang, Jiangsu, 215600, People's Republic of China

*These authors contributed equally to this work

Correspondence: Ruihong Cao; Jin Chen, Department of Cardiology, Zhangjiagang Hospital of Traditional Chinese Medicine Affiliated to Nanjing University of Traditional Chinese Medicine, No. 77 Chang'an South Road, Zhangjiagang, Jiangsu, 215600, People's Republic of China, Tel +86-13584491033; +86-15995988054, Email xuliugang1994@hotmail.com; 89679565@qq.com

Objective: To identify peripheral blood biomarker candidates for coronary artery disease (CAD) and to test whether bioinformatically prioritized genes replicate in an angiography-confirmed cohort, with emphasis on HIST1H2AE as an epigenetically relevant marker.

Methods: Differentially expressed genes (DEGs) were identified from three GEO datasets (GSE42148, GSE98583, GSE12288) after within-dataset normalization; DEGs were then intersected across datasets to prioritize robust candidates while avoiding direct cross-study merging and associated batch effects. Enrichment analysis and protein–protein interaction network prioritization were performed, followed by independent in silico evaluation in GSE20681 using receiver operating characteristic (ROC) analysis. HIST1H2AE and CXCL14 were then tested by qPCR in peripheral blood from 20 participants (10 angiography-confirmed CAD; 10 non-CAD controls). Immune cell proportions in GSE20681 were estimated with CIBERSORTx, and correlations with hub-gene expression were assessed.

Results: Eight DEGs were consistently shared across the three discovery datasets, and HIST1H2AE and CXCL14 were prioritized as hub candidates. In GSE20681, ROC analysis suggested discriminatory ability for both genes (AUC=0.711 for HIST1H2AE; AUC=0.878 for CXCL14). In the angiography-confirmed qPCR cohort, CXCL14 was not consistently different between groups, whereas HIST1H2AE was significantly downregulated in CAD. HIST1H2AE expression showed no significant correlation with estimated immune cell proportions.

Conclusion: This multi-dataset discovery and comparative validation framework prioritizes HIST1H2AE as a peripheral blood biomarker candidate for CAD with an immune-independent expression profile in the analyses performed. Given the small qPCR cohort, these findings are preliminary and require confirmation in larger, independent cohorts and mechanistic studies of chromatin-level regulation.

Keywords: coronary artery disease, HIST1H2AE, CXCL14, peripheral blood, angiography-confirmed, immune-independent, comparative validation framework

https://doi.org/10.2147/IJGM.S593063
Received: 20 January 2026
Accepted: 8 April 2026
Published: 15 April 2026

Introduction

Coronary artery disease (CAD) remains a leading cause of global morbidity and mortality, with patients facing high risks of recurrent cardiovascular events.[1] While coronary angiography serves as the diagnostic gold standard, its invasive nature limits patient acceptance and widespread screening use.[2] Consequently, there is a pressing need to identify less invasive, reliable biomarkers for CAD. CAD pathogenesis reflects a multifactorial process involving genetic susceptibility, immune-inflammatory activation, and epigenetic regulation.[3–5] Because atherosclerosis is driven in part by chronic inflammation,[4] many biomarker studies have emphasized immune-related pathways and mediators within the plaque microenvironment. However, inflammatory signals detected in plaques do not necessarily translate into stable or discriminative markers in peripheral blood, where cell composition and transcript abundance may differ.

Beyond immune activation, epigenetic mechanisms are increasingly recognized as contributors to CAD development and progression.[6] Histone-related processes are particularly relevant because histones and nucleosome organization regulate chromatin accessibility and transcriptional programs in circulating leukocytes—cells that are directly sampled by peripheral blood transcriptomic profiling. Accordingly, altered expression of histone genes (including core histone H2A family members such as HIST1H2AE) in blood may reflect systemic chromatin remodeling or shifts in hematopoietic/leukocyte transcriptional states associated with CAD, providing a potential biomarker axis that is not strictly dependent on immune-cell infiltration patterns.

However, gaps persist in translating mechanistic insights into clinically viable peripheral blood biomarkers. First, the stability and diagnostic performance of immune-associated candidates identified from local lesion studies require rigorous validation in peripheral blood, where expression may differ substantially.[7,8] Second, comparatively fewer studies have evaluated epigenetically relevant, potentially immune-independent biomarkers in blood, despite their theoretical advantages as complementary and possibly more stable circulating signals.

Therefore, the aims of this study were to identify robust CAD-associated DEGs in peripheral blood by integrating multiple independent GEO datasets; to prioritize candidate hub genes using functional enrichment and PPI-based network analyses; and to perform a comparative validation of two biologically distinct candidates: CXCL14 (immune-associated) and HIST1H2AE (histone/epigenetically relevant), using independent in silico evaluation (GSE20681) and qPCR testing in an angiography-confirmed peripheral blood cohort. Additionally, we assessed immune cell infiltration patterns and hub-gene/immune-cell correlations to contextualize whether candidate signals tracked immune composition changes or appeared comparatively immune-independent. The overall workflow is shown in Figure 1.

Materials and Methods

Bioinformatic Analysis of Public Datasets

This study was designed as a comparative validation framework, integrating computational screening, independent dataset ROC validation, and clinical qPCR testing. Three public mRNA expression datasets (GSE42148, GSE98583, GSE12288) from the Gene Expression Omnibus (GEO) database were analyzed to identify differentially expressed genes (DEGs) between CAD patients and healthy controls (Table S1). As these datasets were generated in different studies and/or on different platforms, we did not merge expression matrices across GEO series. Instead, each dataset was processed and normalized within-study, and DEGs were identified separately using consistent statistical criteria. High-confidence candidates were then defined by taking the intersection of DEGs shared across the three discovery datasets, a strategy intended to reduce sensitivity to between-study batch effects and improve robustness. A bioinformatic workflow, including functional enrichment, protein-protein interaction (PPI) network construction, and in silico validation in an independent cohort (GSE20681), was employed to screen for hub genes. A detailed description of the bioinformatic procedures is provided in the Supplementary Methods. This study was conducted in accordance with the Declaration of Helsinki.

Experimental Validation with Clinical Samples

Study Participants and Ethical Statement

Peripheral blood samples were collected from an angiography-confirmed CAD cohort (n=10, ≥50% stenosis) and controls without CAD (n=10, <50% stenosis) at the Department of Cardiology, Zhangjiagang Hospital of Traditional

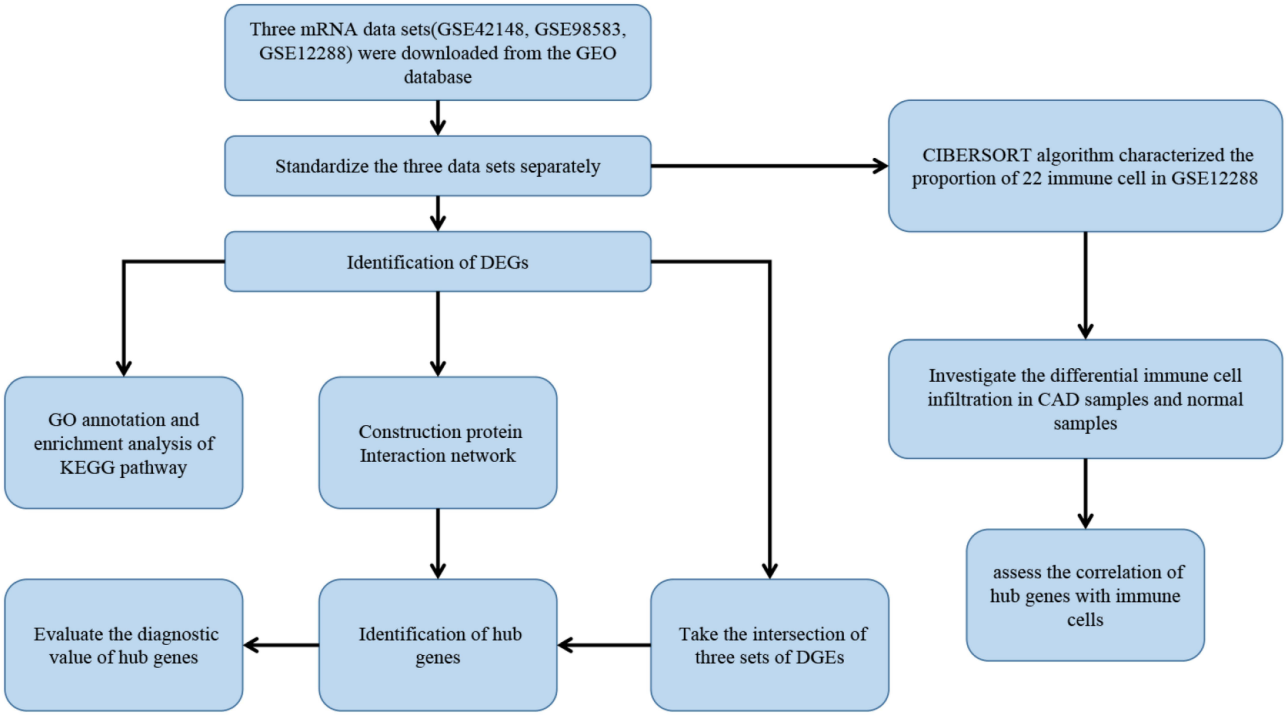

Figure 1 The study workflow.

Chinese Medicine between June 2022 and June 2023. The research was approved by the institutional review board of the Zhangjiagang Hospital of Traditional Chinese Medicine Ethics Committee (Approval No. JS2021-21-1), and all participants provided written informed consent. Exclusion criteria for all participants were as follows: a history of myocardial infarction within the last 3 months, severe valvular heart disease, cardiomyopathy, severe liver or kidney dysfunction, autoimmune diseases, malignant tumors, or any acute or chronic inflammatory diseases. Controls were defined by the absence of angiography-confirmed obstructive CAD (<50% stenosis).

RNA Extraction and Quantitative Real-Time PCR (qPCR)

Total RNA was extracted from peripheral blood samples using TRIzol Reagent. After assessing RNA quality and quantity, cDNA was synthesized from 1 μg of total RNA using the PrimeScript™ RT Reagent Kit. qPCR was performed on a 7500 Fast Real-Time PCR System using SYBR Green. The primer sequences used were:

HIST1H2AE Forward: 5′-TCGGGCAAAAGCTAAACGC-3′
HIST1H2AE Reverse: 5′-TAGTTGCCTTTGCGGAGGAG-3′
CXCL14 Forward: 5′-AAGGGACCCAAGATCCGCTAC-3′
CXCL14 Reverse: 5′-CTTCTCGTTCCAGGCGTTGTA-3′
ACTB (beta-actin) Forward: 5′-CACCCAGCACAATGAAGATCAAGAT-3′
ACTB Reverse: 5′-CCAGTTTTAAATCCTGAGTCAAGC-3′

Relative mRNA expression was calculated using the 2^(-ΔΔCt) method with ACTB as the endogenous control.

Statistical Analysis

For differences in gene expression from the qPCR comparisons between CAD and control groups, data distribution was assessed using the Shapiro–Wilk test. If normality assumptions were satisfied, groups were compared using a two-tailed unpaired Student's *t*-test; otherwise, the Mann–Whitney *U*-test was applied. The Venn diagram was drawn using the jvenn package. GO and KEGG enrichment analyses were analyzed by Fisher's exact test. Immune cell infiltration analysis was performed by Wilcoxon's test. Associations between hub-gene expression and estimated immune cell

proportions were evaluated using Spearman correlation. A two-sided p-value < 0.05 was considered statistically significant. All analyses were performed in R software and GraphPad Prism (for qPCR data analysis and visualization).

Results

Determination of Differentially Expressed Genes (DEGs) in Peripheral Blood CAD Samples

To facilitate biomarker discovery, it is essential first to identify genes whose expression is significantly altered in CAD. This study analyzed publicly available gene expression profile data from the Gene Expression Omnibus (GEO) database. Three independent datasets (GSE42148, GSE98583, and GSE12288) comparing CAD patients and healthy controls' transcriptomes were selected, enhancing our findings' robustness and generalizability.[9] Following the standardization of these datasets, DEGs were screened using established statistical criteria ($p<0.05$, |logFC| thresholds as specified below). A total of 227 DEGs were identified in GSE42148 (161 upregulated, 66 downregulated; Figure 2A, $p<0.05$, $|logFC|>1.0$), 254 DEGs in GSE98583 (141 upregulated, 113 downregulated; Figure 2B, $p<0.05$, $|logFC|>1.0$), and 68 DEGs from GSE12288 (33 upregulated, 35 downregulated; Figure 2C, $p<0.05$, $|logFC|>0.38$). Volcano plots illustrate the distribution of these DEGs (Figure 2A–C). Heatmaps display the expression patterns of the top 50 DEGs with the largest fold changes in each dataset (Supplementary Figure 1).

A widely used strategy for identifying high-confidence candidates involves pinpointing consistently dysregulated genes across multiple patient cohorts. We identified the intersection of the DEG lists derived from the three datasets to implement this approach. This stringent filtering strategy yielded eight common DEGs: MAP7, RIPK4, BAALC, CA6, CXCL14, HIST1H2AE, MS4A3, and GPR15 (Figure 2D). These genes represent promising candidates potentially implicated in the pathogenesis of CAD and merit further functional investigation.

Functional Annotation and Pathway Enrichment Analysis of Candidate Genes

To elucidate the biological functions and pathways potentially influenced by the observed gene expression changes, an essential step toward understanding the mechanistic roles of the identified genes, we performed Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on the identified DEGs using the DAVID platform (https://david.ncifcrf.gov/). The top five enriched GO terms are shown in Table S2 and Supplementary Figure 2A–C. Significant enrichment was observed in fundamental biological processes (BP) including signal transduction, positive regulation of transcription from RNA polymerase II promoter, positive regulation of cell proliferation, negative regulation of transcription from RNA polymerase II promoter, and cell adhesion. Key cellular components (CC) involved were the nucleus, integral component of the plasma membrane, extracellular space, extracellular region, and cell surface. Enriched molecular functions (MF) included transcription factor activity, sequence-specific DNA binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, protein homodimerization activity, and metal ion binding. These findings suggest that the molecular alterations in CAD involve complex disruptions in cellular signaling, transcriptional control, cell-cell interactions, and potentially fundamental metabolic activities.

KEGG pathway analysis (Table S3 and Supplementary Figure 2D) further revealed enrichment in pathways such as serotonergic synapse; neuroactive ligand−receptor interaction; glycine, serine, and threonine metabolism; fat digestion and absorption; cocaine addiction, hinting at diverse systemic processes, including neurological and metabolic pathways, potentially linked to CAD development or its comorbidities.[10]

Protein-Protein Interaction (PPI) Network and Hub Gene Selection

Knowing that proteins exert their functions through complex interactions, we constructed a PPI network using the STRING database and visualized it with Cytoscape (v3.9.0) to investigate the relationships among the identified DEGs and pinpoint potential key regulatory hubs. The resulting network comprised 357 nodes and 774 edges. To further identify tightly interconnected clusters within the network, which often represent functional modules performing specific biological tasks, we utilized the MCODE plug-in with defined criteria (Degree Cutoff = 2, Node Score Cutoff = 0.2,

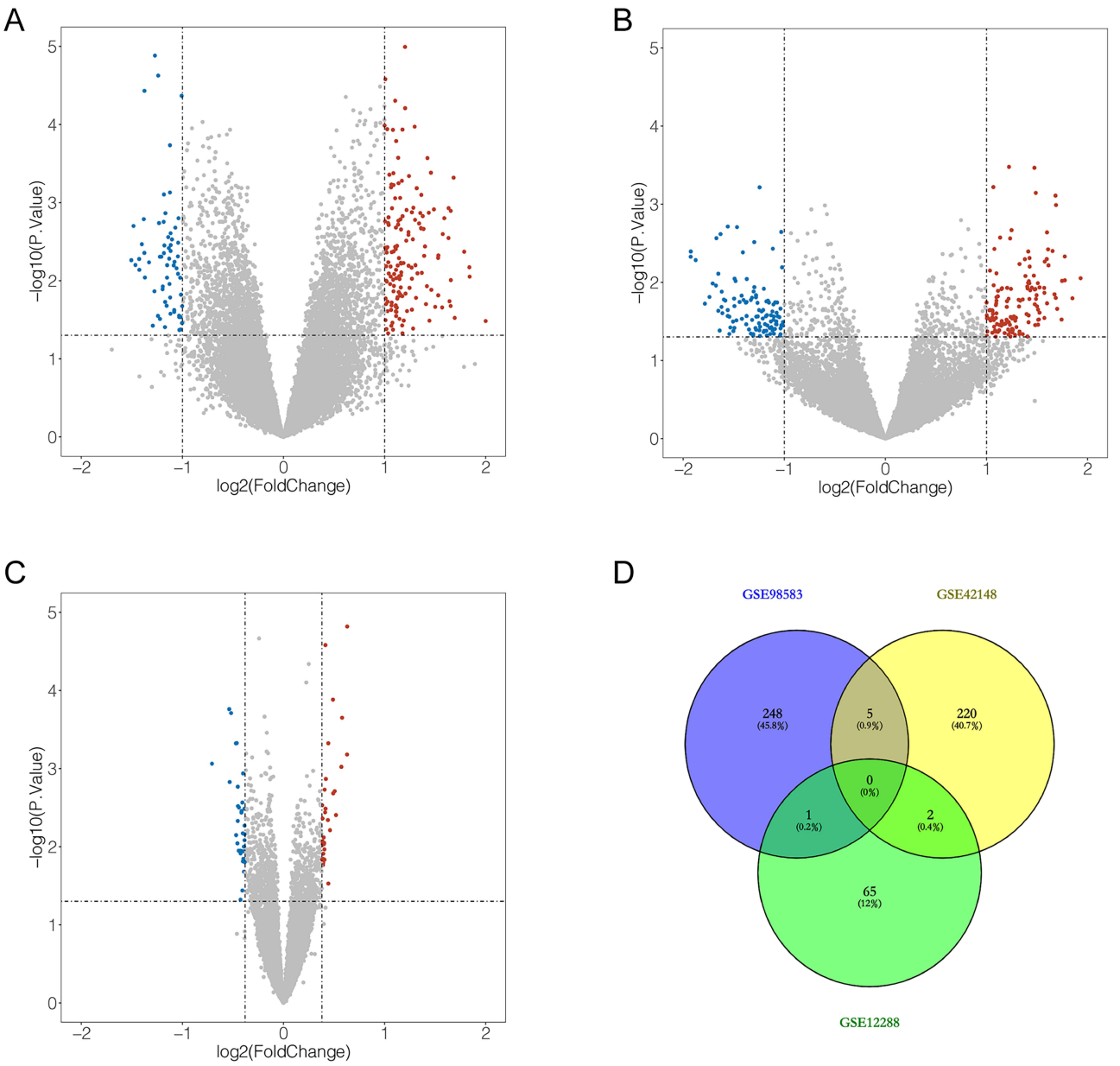

Figure 2 Identification of DEGs from the three human datasets. Volcano plots displaying the distribution of DEGs, where red represents upregulated genes, and blue represents downregulated genes. (**A**) Volcano plot of GSE42148: 227 DEGs (161 upregulated, 66 downregulated; $p < 0.05$, $|\log FC| > 1.0$). (**B**) Volcano plot of GSE98583: 254 DEGs (141 upregulated, 113 downregulated; $p < 0.05$, $|\log FC| > 1.0$). (**C**) Volcano plot of GSE12288: 68 DEGs (33 upregulated, 35 downregulated; $p < 0.05$, $|\log FC| > 0.38$). (**D**) The Venn diagram of the three sets of DEGs.

K-Core = 2, Max Depth = 100). This analysis identified four significant modules with scores ≥ 4.0 (Scores: 9.556, 9.00, 4.993, 4.50; Supplementary Figure 3A–D).

To prioritize central genes likely playing crucial roles within the network, the cytoHubba algorithm, a common method for identifying influential nodes based on network topology, was used next. Critically, by intersecting the high-ranking hub genes derived from the PPI network analysis with the 8 common DEGs identified earlier, HIST1H2AE and CXCL14 were screened. The consistent differential expression observed across multiple datasets, coupled with their central positioning in the interaction network, strongly suggests that HIST1H2AE and CXCL14 are key candidate genes with significant relevance to the pathology of CAD.

Independent ROC Validation in Angiography-Confirmed Cohort

To assess the initial diagnostic potential of the identified hub genes in an independent cohort, we utilized the GSE20681 dataset. Receiver Operating Characteristic (ROC) curves were generated to assess the ability of HIST1H2AE and CXCL14 expression levels to discriminate between CAD patients and healthy controls, and the Area Under the Curve (AUC), a quantitative measure of diagnostic accuracy, was calculated. Furthermore, violin plots were used to visualize the expression differences of these genes between the groups (Figure 3A and B). The validation analysis yielded an AUC of 0.878 for CXCL14 and 0.711 for HIST1H2AE. These in silico results suggest discriminatory potential in this independent dataset and provided the rationale for proceeding to experimental testing; however, ROC performance derived from a single public cohort should be interpreted as exploratory and requires confirmation in prospective clinical cohorts.

Comparative qPCR Validation of HIST1H2AE and CXCL14 in Peripheral Blood

To test the clinical relevance of the hub gene candidates within the comparative validation framework, we performed quantitative real-time PCR (qPCR) on peripheral blood samples from our angiography-confirmed clinical cohort (n=20; 10 CAD vs 10 non-CAD controls). The results revealed a clear distinction between the two candidates. While ROC analysis had indicated diagnostic potential for both, qPCR validation showed that HIST1H2AE was significantly downregulated in CAD patients compared to controls ($p < 0.05$; Figure 3C and D), whereas CXCL14 expression was not consistently altered (Figure 3E and F). Given the limited sample size, these qPCR findings should be interpreted as preliminary evidence and require replication in larger independent cohorts. Within this limitation, the comparative pattern

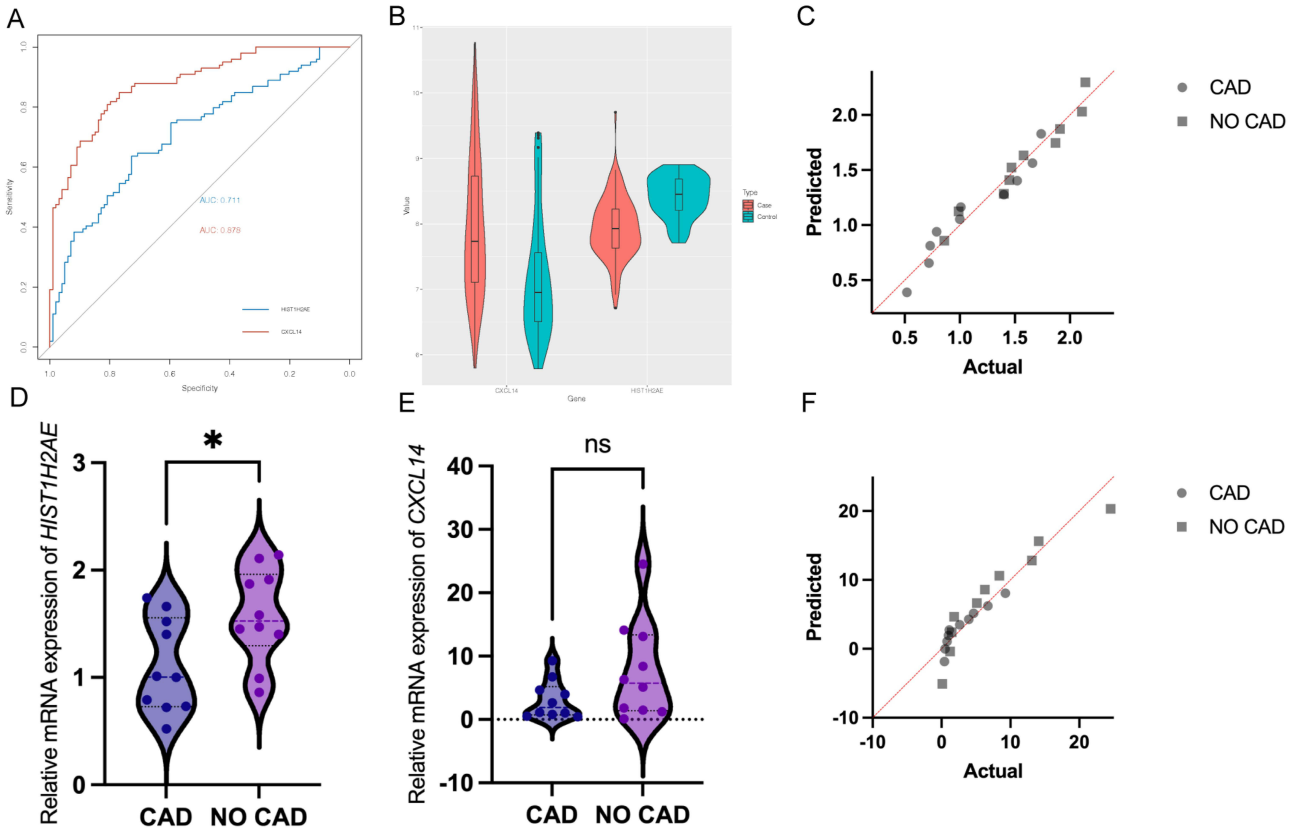

Figure 3 Validation and identification of hub genes related to CAD. (**A**) Receiver operating characteristic (ROC) curve analysis of CXCL14 and HIST1H2AE in GSE20681; the area under the curve (AUC) values for CXCL14 and HIST1H2AE were 0.878 and 0.711, respectively. (**B**) Violin plots showing CXCL14 and HIST1H2AE expression in the validation dataset (GSE20681). Green indicates controls; red indicates CAD. (**C**) Normal Q-Q plot confirming the normal distribution of HIST1H2AE expression data. (**D**) qPCR validation showing that HIST1H2AE mRNA expression was significantly downregulated in CAD patients (n=10) compared with non-CAD controls (n=10). (**E**) qPCR validation showing no significant difference in CXCL14 mRNA expression between CAD patients (n=10) and non-CAD controls (n=10). (**F**) Normal Q-Q plot confirming the normal distribution of CXCL14 expression data. * indicates p<0.05; ns indicates not significant.

https://doi.org/10.2147/IJGM.S593063

—stable downregulation of HIST1H2AE versus non-reproducible CXCL14—supports the utility of the framework for prioritizing candidates for further validation.

Immune Cell Infiltration Analysis in CAD Samples

Given the well-established critical role of inflammation and immune responses in initiating, progressing, and destabilizing atherosclerotic plaques underlying CAD,[11] understanding the immune context associated with the disease and potential biomarkers is vital. Thus, CIBERSORTx algorithm,[12] a robust computational method for deconvolving tissue transcriptomic data, was used to estimate the relative proportions of 22 immune cell types in the transcriptomic data of 112 normal and 110 CAD samples (derived from GSE20681). Comparative analysis revealed significant alterations in the immune landscape of CAD patients (Figure 4). Specifically, the proportions of naive B cells (p=0.023) and M0 macrophages (p=0.004) were significantly elevated in CAD samples compared to normal controls, while the proportion of naive CD4 T cells (p=0.024) was significantly lower. These findings directly highlight immune dysregulation within the CAD environment, emphasize the involvement of specific immune cell subsets (B cells, macrophages, T cells) in the disease's pathological mechanisms, and provide a strong rationale for further investigation into the connections between the identified hub genes and this critical immune context.

Correlation Analysis of Hub Genes with Immune Cell Proportions: Evidence for Immune Independence of HIST1H2AE

To investigate whether our identified hub genes might contribute to or be influenced by the observed immune alterations, potentially providing mechanistic insights into their function within the disease microenvironment,[13] we examined the correlation between the expression levels of HIST1H2AE and CXCL14 and the infiltration levels of various immune cell types. Lollipop plots show the correlations (Figure 5). Notably, CXCL14 expression exhibited significant positive correlations with the proportions of plasma cells (r=0.25, p=0.03) and M0 macrophages (r=0.49, p=0.02) and a significant negative correlation with M1 macrophages (r=–0.15, p=0.03). These specific correlations suggest that CXCL14 may play a role in influencing B-cell differentiation/activity and macrophage polarization (potentially promoting a less pro-inflammatory M0 state while being inversely related to the pro-inflammatory M1 state) within the CAD tissue environment, thereby potentially impacting the local inflammatory balance. In contrast, no significant correlation was observed between HIST1H2AE expression and the analyzed immune cell proportions. Importantly, this lack of correlation supports the potential role of HIST1H2AE as an immune-independent biomarker in CAD.

Discussion

In the present study, we employed an integrated bioinformatic and experimental approach to identify and validate peripheral blood biomarkers for CAD. A key feature is the comparative validation framework in an angiography-characterized peripheral blood cohort, which enabled us to contrast an immune-associated candidate (CXCL14) with an epigenetically relevant candidate (HIST1H2AE). We first analyzed three independent GEO peripheral blood microarray datasets (GSE42148, GSE98583, and GSE12288) to identify DEGs between CAD and controls. To improve robustness across heterogeneous public cohorts, we focused on genes that were consistently differentially expressed across datasets (intersection approach). This yielded eight shared DEGs (MAP7, RIPK4, BAALC, CA6, CXCL14, HIST1H2AE, MS4A3, and GPR15), and subsequent PPI-based prioritization identified HIST1H2AE and CXCL14 as hub genes for downstream validation.

Our principal experimental finding is that HIST1H2AE was significantly downregulated in the peripheral blood of CAD patients in our qPCR cohort. HIST1H2AE is a histone cluster gene encoding a core H2A histone protein required for nucleosome structure and chromatin organization—processes fundamental to transcriptional regulation, DNA repair, DNA replication, and chromosomal stability. Histone regulation and chromatin dynamics have been linked to cardiovascular aging[14] and atherosclerosis.[15] In this context, our observation supports the possibility that altered histone gene expression—specifically HIST1H2AE downregulation—may reflect CAD-associated epigenetic dysregulation in circulating blood cells.[16] Consistent with the plausibility of this signal, Gong et al reported HIST1H2AE as a potential CAD-

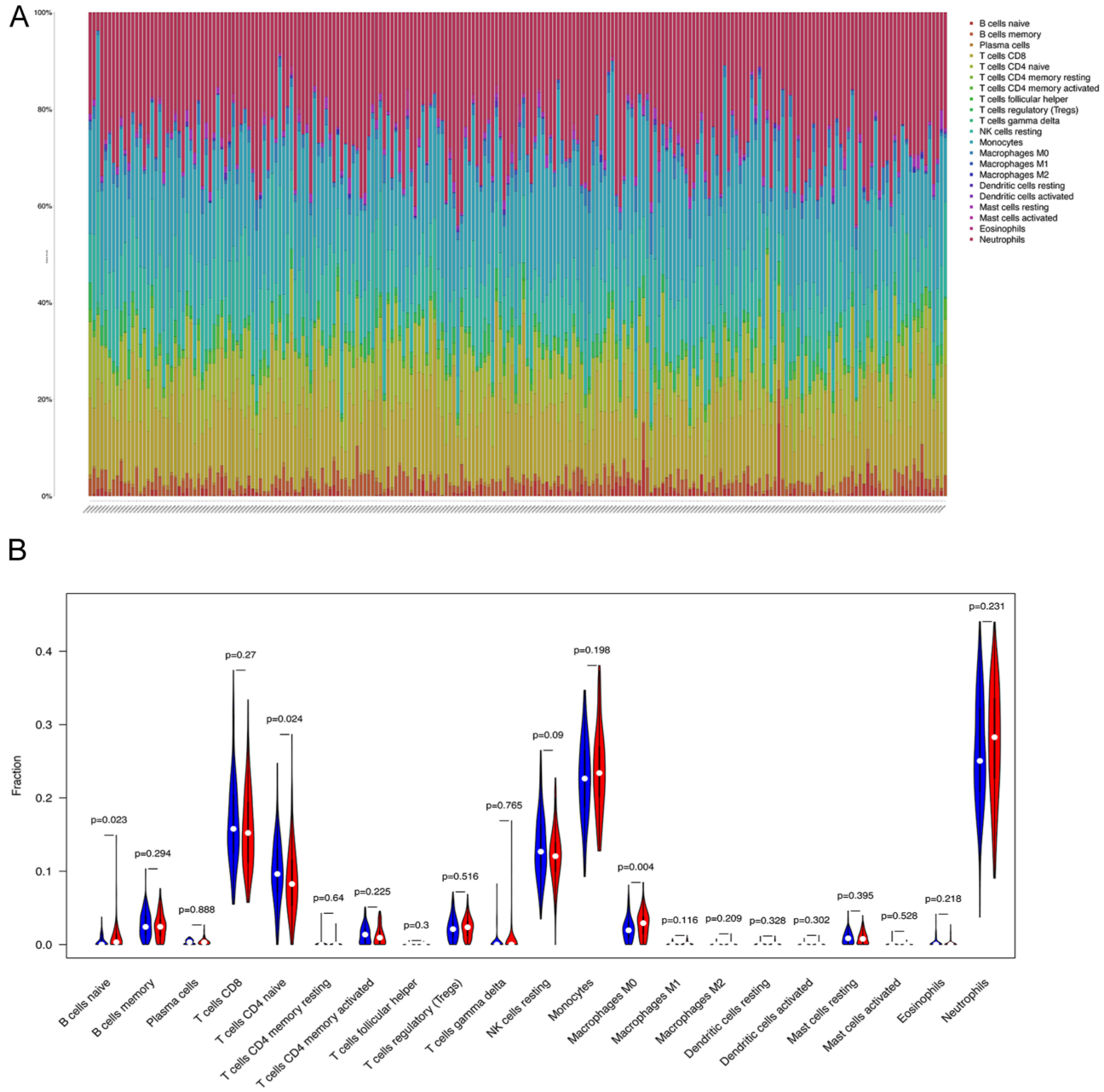

Figure 4 Immune cell composition analysis in GSE20681 using CIBERSORTx. (**A**) Stacked bar plot showing the relative proportions of 22 immune cell types in each sample. (**B**) Group-wise comparison of immune cell proportions between CAD and controls. Blue indicates controls and red indicates CAD.
Abbreviation: CAD, coronary artery disease.

related gene across populations,[6] and Lin et al suggested that healthy dietary patterns may influence gene regulation including HIST1H2AE, with potential cardiovascular relevance.[17] Importantly, HIST1H2AE showed no significant association with estimated immune cell proportions in our correlation analysis. While immune deconvolution from bulk expression has inherent uncertainty, this pattern is compatible with HIST1H2AE capturing a comparatively immune-independent axis of variation (relative to chemokine-type signals). Together, the cross-dataset prioritization plus qPCR confirmation provides a coherent rationale to advance HIST1H2AE into larger-scale validation and mechanistic studies.

In contrast, although bioinformatic analyses supported CXCL14 as a strong candidate—including a high AUC (0.878) in the independent dataset—our qPCR validation did not confirm significant differential expression for CXCL14 in the

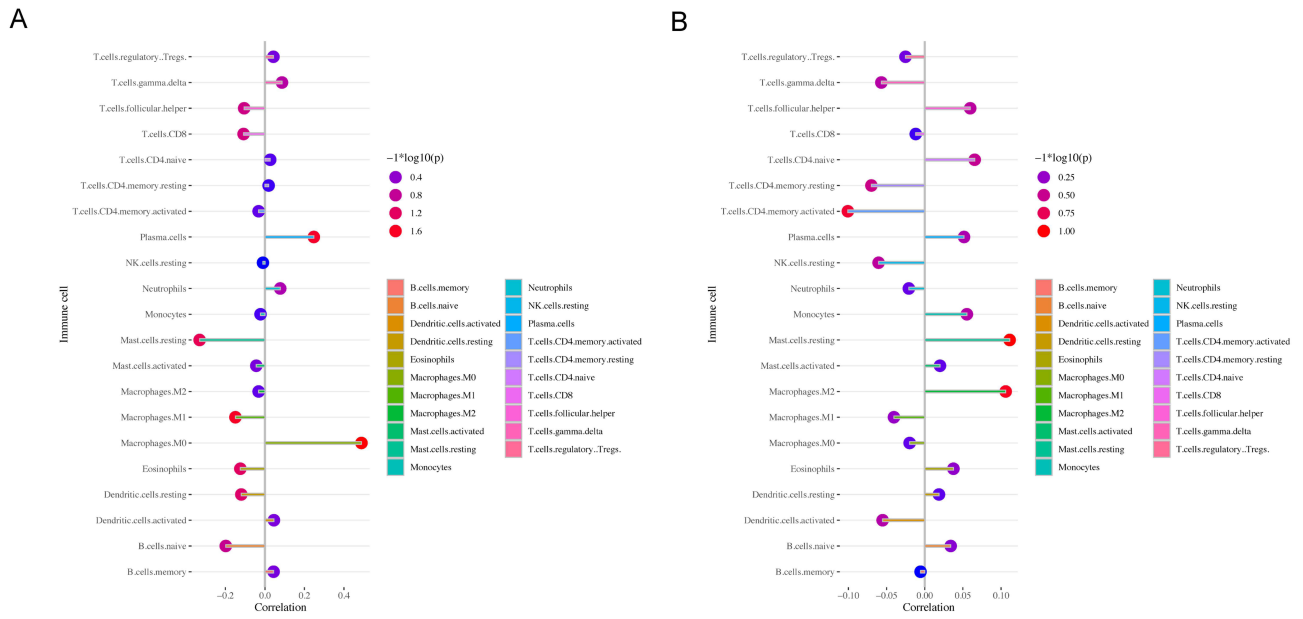

Figure 5 Correlation between hub-gene expression and estimated immune cell proportions in GSE20681. The length of the bars represents the magnitude of the correlation. The color of the dot represents the p-value. The bluer the color, the larger the p-value, and the redder the color, the smaller the p-value. (**A** and **B**) Correlation of CXCL14 and HIST1H2AE with the proportion of immune cell infiltration, respectively.

angiography-characterized cohort. CXCL14 is a chemokine involved in immune cell migration[18] and has pleiotropic roles across antimicrobial defense, neuronal circuits, and metabolism.[19] Mechanistically, CXCL14 can antagonize the CXCL12/CXCR4 axis,[20] which has been implicated in atherosclerosis,[21] and CXCL14 may also participate in thrombosis-related processes through direct interaction with CXCR4,[22] potentially relevant to coronary events. The discrepancy between strong in silico performance and lack of qPCR confirmation underscores a key translational challenge: immune-associated transcript signals may be more context-dependent (eg, influenced by cohort composition, treatment status, timing relative to disease activity, and technical factors), and therefore require experimental confirmation before clinical interpretation.[23] This divergence is also consistent with our immune correlation results: CXCL14 showed relationships with specific immune subsets, including plasma cells and macrophage subsets, whereas HIST1H2AE did not. Taken together, these findings suggest that CXCL14 may reflect immune composition/state variation in the analyzed public cohort rather than a stable CAD-associated expression shift detectable by qPCR in our clinical sampling context.

Consistent with the well-established inflammatory contribution to CAD,[11] immune cell composition analysis showed increased naïve B cells and M0 macrophages and decreased naïve CD4 T cells in CAD samples. Macrophage accumulation and polarization dynamics shape plaque inflammation, with macrophages contributing to lipid uptake, foam cell formation, and lesion progression.[24] Adaptive immune remodeling is also important in atherosclerosis; differentiation of naïve CD4 T cells into effector/memory populations has been described, and reduced naïve CD4 T cells have been reported in atherosclerosis alongside expansion of activated subsets.[25] In our correlation analysis, CXCL14 tracked with particular immune subsets, whereas HIST1H2AE did not, further supporting the concept that these two genes may capture different biological axes in CAD: one more immune-linked (CXCL14) and one more consistent with chromatin/epigenetic regulation (HIST1H2AE).

From a clinical translation perspective, HIST1H2AE holds promise as a non-invasive peripheral blood biomarker candidate and may have utility for screening or risk stratification. However, CAD is heterogeneous, and single markers rarely provide sufficient accuracy alone. Future work should evaluate HIST1H2AE in combination with established clinical variables and/or additional molecular features to build multimodal diagnostic or prognostic models, and to test whether HIST1H2AE improves performance beyond conventional risk factors.

This study has a few limitations. First, the experimental qPCR validation cohort was small, limiting statistical power and generalizability; this may partly explain the lack of significant difference for CXCL14. Therefore, the diagnostic

performance and robustness of HIST1H2AE should be confirmed in larger, independent, ideally multi-center cohorts. Second, because of limited sample size, distributional assumptions are difficult to verify with high confidence; future studies could pre-specify normality checks and consider non-parametric or resampling-based approaches to improve robustness. Third, this work relied on public microarray datasets; differences in platforms, preprocessing pipelines, and batch effects across GEO studies may influence DEG detection and downstream network/ROC results, even though intersecting DEGs across multiple datasets increased stringency.[9] More explicit harmonization and fully reproducible reporting of thresholds and code would further strengthen reliability. Fourth, this study did not include in vitro or in vivo experiments to define causality. Future investigations should elucidate underlying epigenetic mechanisms, such as effects on chromatin accessibility and downstream gene programs, by which HIST1H2AE dysregulation may contribute to atherosclerosis.[14–16] Overall, the current findings should be interpreted as an initial clinical validation signal rather than a definitive diagnostic test.

Conclusions

This study integrates multi-dataset bioinformatic screening with independent in silico evaluation and qPCR testing in an angiography-confirmed peripheral blood cohort to propose a comparative validation framework for CAD biomarker prioritization. Within this framework, CXCL14 showed apparent diagnostic potential in computational analyses but did not demonstrate consistent differential expression in our qPCR cohort, indicating that immune-associated candidates may be context-dependent in peripheral blood. In contrast, HIST1H2AE was significantly downregulated in the qPCR cohort and showed no detectable correlation with estimated immune cell proportions, supporting it as an epigenetically relevant biomarker candidate with a comparatively immune-independent profile in this analytical setting. Given the small experimental sample size, these results should be considered preliminary and hypothesis-generating. Larger, independent (ideally prospective) studies and mechanistic investigations of chromatin-level regulation are needed to confirm reproducibility, quantify diagnostic performance, and determine whether HIST1H2AE provides added value beyond established clinical risk factors.

Abbreviations

CAD, Coronary Artery Disease; GEO, Gene Expression Database; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, Protein-Protein Interaction; BP, Biological Processes; CC, Cell Components; MF, Molecular Functions; DEGs, Differentially Expressed Genes.

Data Sharing Statement

The gene expression profiles of GSE42148 were downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42148).

The gene expression profiles of GSE98583 were downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98583). The gene expression profiles of GSE98583 were downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12288).

The gene expression profiles of GSE20681 were downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2068).

Ethics Approval and Informed Consent

The research was approved by the institutional review board of the Zhangjiagang Hospital of Traditional Chinese Medicine Ethics Committee (Approval No. JS2021-21-1). This study was conducted in accordance with the Declaration of Helsinki.

Acknowledgments

This paper has been uploaded to Research Square as a preprint: https://www.researchsquare.com/article/rs-1298849/v1. It is an early version and has not been peer-reviewed.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by Zhangjiagang Science and Technology Project (ZKYL2343; ZKYL2363), Jiangsu Pharmaceutical Vocational College School-Land Collaborative Innovation Research Project (20239620; 202490604), Suzhou Municipal Health Commission Science and Technology Project (QNXM2024066), and Youth Science and Technology Project of Zhangjiagang Health Commission (ZJGQNKJ 202413).

Disclosure

The authors have no relevant financial or non-financial interests to disclose.

References

1. Mozaffarian D, Benjamin EJ, Go AS, et al. Executive summary: heart disease and stroke statistics--2016 update: a report from the American heart association. *Circulation*. 2016;133(4):447–454. doi:10.1161/CIR.0000000000000366
2. Mehilli J, Presbitero P. Coronary artery disease and acute coronary syndrome in women. *Heart*. 2020;106(7):487–492. doi:10.1136/heartjnl-2019-315555
3. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. 2017;18(6):331–344. doi:10.1038/nrg.2016.160
4. Tedgui A, Mallat Z. Cytokines in atherosclerosis: pathogenic and regulatory pathways. *Physiol Rev*. 2006;86(2):515–581. doi:10.1152/physrev.00024.2005
5. Yang YL, Wu CH, Hsu PF, et al. Systemic immune-inflammation index (SII) predicted clinical outcome in patients with coronary artery disease. *Eur J Clin Invest*. 2020;50(5):e13230. doi:10.1111/eci.13230
6. Xu S, Pelisek J, Jin ZG. Atherosclerosis Is an Epigenetic Disease. *Trends Endocrinol Metab*. 2018;29(11):739–742. doi:10.1016/j.tem.2018.04.007
7. Li GM, Zhang CL, Rui RP, Sun B, Guo W. Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy. *Eur Rev Med Pharmacol Sci*. 2018;22(11):3553–3569. doi:10.26355/eurrev_201806_15182
8. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457–462. doi:10.1093/nar/gkv1070
9. Hussain I, Qureshi M, Ismail M, Iftikhar H, Zywiolek J, Lopez-Gonzales JL. Optimal features selection in the high dimensional data based on robust technique: application to different health database. *Heliyon*. 2024;10(17):e37241. doi:10.1016/j.heliyon.2024.e37241
10. Wang W, Zhang X, Lyu J, et al. Neuroimaging findings from cerebral structure and function in coronary artery disease. *J Magn Reson Imaging*. 2025;61(5):2066–2082. doi:10.1002/jmri.29624
11. Ramoni D, Carbone F, Kraler S, Di Vece D, Montecucco F, Liberale L. Inflammation-driven plaque erosion in atherosclerosis: a focus on complement system pathways. *Curr Atheroscler Rep*. 2025;27(1):42. doi:10.1007/s11883-025-01279-x
12. Conning-Rowland M, Cheng CW, Brown O, et al. Application of CIBERSORTx and BayesPrism to deconvolution of bulk RNA-seq data from human myocardium and skeletal muscle. *Heliyon*. 2025;11(4):e42499. doi:10.1016/j.heliyon.2025.e42499
13. Zhu R, Zhao Y, Yin H, Shu L, Ma Y, Tao Y. Identification of immune-related hub genes and potential molecular mechanisms involved in COVID-19 via integrated bioinformatics analysis. *Sci Rep*. 2024;14(1):29964. doi:10.1038/s41598-024-81803-2
14. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362–d368. doi:10.1093/nar/gkw937
15. Zhang W, Song M, Qu J, Liu GH. Epigenetic modifications in cardiovascular aging and diseases. *Circ Res*. 2018;123(7):773–786. doi:10.1161/CIRCRESAHA.118.312497
16. Beik-Khormizi M, Zare-Khormizi MR, Firoozabadi AD, Vakili M, Hekmatimoghaddam S, Pourrajab F. Alteration in epigenetic profile in subclinical atherosclerosis and in high uric acid. *Sci Rep*. 2025;15(1):21079. doi:10.1038/s41598-025-06842-9
17. Gong R, Chen MH, Peng LS, Wei SL. Common genes in coronary artery disease from Europe, Asia and North America regardless of race and lifestyle. *Eur Rev Med Pharmacol Sci*. 2015;19(6):1092–1100.
18. Lin H, Rogers GT, Lunetta KL, et al. Healthy diet is associated with gene expression in blood: the Framingham Heart Study. *Am J Clin Nutr*. 2019;110(3):742–749. doi:10.1093/ajcn/nqz060
19. Dai C, Basilico P, Cremona TP, et al. CXCL14 displays antimicrobial activity against respiratory tract bacteria and contributes to clearance of *Streptococcus pneumoniae* pulmonary infection. *J Immunol*. 2015;194(12):5980–5989. doi:10.4049/jimmunol.1402634
20. Takahashi M, Takahashi Y, Takahashi K, et al. CXCL14 enhances insulin-dependent glucose uptake in adipocytes and is related to high-fat diet-induced obesity. *Biochem Biophys Res Commun*. 2007;364(4):1037–1042. doi:10.1016/j.bbrc.2007.10.120
21. Tanegashima K, Suzuki K, Nakayama Y, et al. CXCL14 is a natural inhibitor of the CXCL12-CXCR4 signaling axis. *FEBS Lett*. 2013;587(12):1731–1735. doi:10.1016/j.febslet.2013.04.046

22. Puca AA, Carrizzo A, Spinelli C, et al. Single systemic transfer of a human gene associated with exceptional longevity halts the progression of atherosclerosis and inflammation in ApoE knockout mice through a CXCR4-mediated mechanism. *Eur Heart J*. 2020;41(26):2487–2497. doi:10.1093/eurheartj/ehz459
23. Tong W, Duan Y, Yang R, et al. Foam cell-derived CXCL14 muti-functionally promotes atherogenesis and is a potent therapeutic target in atherosclerosis. *J Cardiovasc Transl Res*. 2020;13(2):215–224. doi:10.1007/s12265-019-09915-z
24. Barrett TJ. Macrophages in atherosclerosis regression. *Arterioscler Thromb Vasc Biol*. 2020;40(1):20–33. doi:10.1161/ATVBAHA.119.312802
25. Olson NC, Doyle MF, Jenny NS, et al. Decreased naive and increased memory CD4(+) T cells are associated with subclinical atherosclerosis: the multi-ethnic study of atherosclerosis. *PLoS One*. 2013;8(8):e71498. doi:10.1371/journal.pone.0071498