scientific reports

Check for updates

OPEN

Identification of key proteins and pathways in myocardial infarction using machine learning approaches

Chang Liu[1,4], Xing Zhang[1,4], Qian Xie[1,4], Binbin Fang[1,4], Fen Liu[1,2], Junyi Luo[1], Gulandanmu Aihemaiti[1], Wei Ji[1], Yining Yang[3]✉ & Xiaomei Li[1]✉

Acute myocardial infarction (AMI) is a leading cause of global morbidity and mortality, requiring deeper insights into its molecular mechanisms for improved diagnosis and treatment. This study combines proteomics, transcriptomics and machine learning (ML) to identify key proteins and pathways associated with AMI. Plasma samples from 48 AMI patients and 50 healthy controls (HC) were used for proteomic sequencing. Differentially expressed proteins (DEPs) were identified and analyzed for pathway enrichment. Protein-protein interaction (PPI) networks were constructed, and we conducted a meta-analysis (GSE60993, GSE61144, GSE48060) using an inverse variance model to combine differentially expressed genes (DEGs) identified via LIMMA and FDR adjustment across three studies. Clustering and co-expression analysis were performed using K-Medoids and weighted gene co-expression network analysis (WGCNA). ML feature selection identified hub proteins, which were validated across bulk, single-cell, and spatial datasets for atherosclerosis (ATH) and MI. In this study, we identified 437 DEPs with 291 up-regulated and 146 down-regulated proteins. Functional enrichment analysis revealed key pathways involved in inflammation, immunity, metabolism, and cellular stress responses, among others. Using non-negative matrix factorization (NNMF) and K-Medoids clustering, AMI patients were divided into two clusters (C1 and C2), with distinct protein expression patterns and inflammatory responses. Differential analysis between clusters revealed 200 cluster-specific DEPs, with C1 associated with angiogenesis and vascular remodeling, and C2 linked to cellular stress and apoptosis. A meta-analysis identified 1383 DEGs, and their intersection with DEPs yielded 63 proteins, which were subsequently refined by logistic regression to 36 AMI-associated proteins. Furthermore, a protein co-expression network analysis identified 49 modules, with the turquoise module being strongly associated with AMI highlighting pathways in lipid metabolism, immune response, and tissue repair. From this module, 17 key proteins were selected, and ML further distilled these to nine core features (CAMP, CLTC, CTNNB1, FUBP3, IQGAP1, MANBA, ORM1, PSME1, and SPP1) that are closely linked to immune regulation, apoptosis, and metabolism. These proteins were validated across multiple datasets. Single-cell analysis revealed distinct expression patterns of these proteins across cell types and spatial regions in ATH and MI, emphasizing their roles in inflammation, vascular remodeling, and plaque instability. This study identifies critical proteins and pathways in AMI, offering potential biomarkers and therapeutic targets. The use of ML provides a robust framework for identifying AMI's key molecular.

Keywords Acute myocardial infarction, Proteomics, Functional enrichment analysis, Machine learning, Single-cell sequencing

Acute myocardial infarction (AMI) is a prevalent and severe manifestation of coronary artery disease, representing a leading cause of morbidity and mortality worldwide[1]. According to the 2019 global burden of disease report, cardiovascular diseases caused 18.5 million deaths globally, with about 2.45 million deaths in China in 2021 alone, of which approximately 1 million (40.5%) were due to MI[2]. Despite significant advancements in reperfusion therapy and antithrombotic strategies, AMI remains the principal cause of chronic heart failure[3].

[1]Department of Cardiology, The first Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China. [2]Xinjiang Key Laboratory of Cardiovascular Disease Research, Clinical Medical Research Institute of First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China. [3]Department of Cardiology, People's Hospital of Xinjiang Uyghur Autonomous Region, Urumqi 830001, China. [4]Chang Liu, Xing Zhang, Qian Xie and Binbin Fang contributed equally to this work. ✉email: yangyn5126@163.com; lixm505@163.com

Around 10% of MI survivors live with severely compromised left ventricular (LV) function, often leading to heart failure and adverse LV remodeling[4]. The death of myocardial cells and the resulting excessive inflammation are primary contributors to heart damage in AMI, which can result in heart failure and sudden death[5]. Currently, myocardial enzyme profiles and cardiac troponin are routinely used for AMI diagnosis; however, these markers primarily indicate myocardial damage rather than providing insights into the underlying molecular mechanisms of AMI. In light of this, our study focuses on elucidating the molecular pathways involved in AMI and identifying candidate biomarkers that may contribute to a more nuanced characterization of the disease. Such mechanistic insights could pave the way for the development of targeted therapeutic strategies.

The field of proteomics has been widely applied to identify biomarkers associated with specific phenotypes and genotypes. Plasma proteomics may be the most suitable and rapid approach for MI research, and it is also the most likely to translate proteomics findings into clinical applications. Currently, proteomics research methods have been applied to the study of proteomes in platelets, red blood cells, and monocytes, and have preliminarily elucidated the protein expression profiles in acute coronary syndrome[6,7]. The use of high-dimensional proteomic datasets has become common in a variety of real-world diseases, especially as machine learning (ML) is increasingly applied to the identification and analysis of key molecules[8]. However, the presence of numerous irrelevant and redundant features in these datasets often impairs the accuracy of ML systems[9]. To mitigate this issue, feature selection has emerged as an effective method[10]. Unlike feature extraction, which reduces dimensionality by creating new features from the original ones and may lose some of the original feature relationships[11], feature selection directly chooses relevant features from the initial feature space, thus preserving their original semantic meanings and enhancing the interpretability and reliability of the model[12]. Particle swarm optimization (PSO), an evolutionary algorithm, has proven effective for feature selection due to its robust global search capabilities. Our approach employs an enhanced PSO variant-ISPSO-that integrates sub-feature grouping, probability-based search operators, and adaptive penalty terms, guided by the information gain ratio for feature importance assessment. This strategy ensures a high-quality initial population and improves search efficiency, outperforming traditional feature selection algorithms on high-dimensional datasets[13].

Accordingly, this study aims to characterize the protein expression landscape of AMI using advanced proteomic techniques, thereby uncovering the biological significance and mechanistic underpinnings of protein-level alterations. Furthermore, by integrating machine learning algorithms, we aim to identify diagnostic protein biomarkers that can facilitate precise clinical diagnosis and risk stratification of AMI.

Materials and methods

Sample selection

This study complied with the ethical standards outlined in the Helsinki Declaration II and was approved by the Institutional Review Board of the First Affiliated Hospital of Xinjiang Medical University (approval number: 20220308-105). Written informed consent was obtained from all participants. In this study, the diagnostic criteria for AMI patients were based on a comprehensive assessment of clinical symptoms, cardiac enzyme indicators, ECG changes, and coronary angiography. Patients were required to present with typical symptoms of myocardial ischemia, such as chest pain or discomfort (commonly described as a feeling of pressure, tightness, or squeezing that may radiate to the left arm, neck, or jaw), accompanied by symptoms including shortness of breath, sweating, and nausea; plasma cardiac troponin levels had to exceed the 99th percentile upper reference limit, and creatine kinase (CK) along with its isoenzyme CK-MB were significantly elevated, indicating myocardial injury; the ECG needed to show typical ischemic changes, including ST-segment elevation (≥ 1 mm) in at least two contiguous leads, ST-segment depression or T-wave inversion, and possibly the appearance of pathological Q waves; in addition, all patients underwent coronary angiography, where a stenosis of ≥ 70% in non-left main coronary arteries or a stenosis of ≥ 50% in the left main coronary artery was considered indicative of significant lesions or occlusion. Plasma samples were collected from 48 AMI patients and 50 healthy controls (HC). All plasma samples were collected immediately upon admission to the emergency department, prior to any therapeutic interventions. Clinical details are summarized in Supplementary file1.

Sample preparation and lysis

Plasma samples from AMI patients and HC were processed following established proteomic protocols. Samples were lysed using TCEP buffer (2% sodium deoxycholate, 40 mM 2-chloroacetamide, 100 mM Tris-HCl, 10 mM tris(2-chloroethyl) phosphate, 1 mM PFSM, pH 8.5), supplemented with protease and phosphatase inhibitors. Lysis was performed at 99 °C for 30 min, followed by cooling to room temperature. Samples were then digested with trypsin (Promega, Madison, WI, USA, #V5280) at 37 °C for 18 h. Post-digestion, 10% formic acid was added, samples were vortexed for 3 min, and centrifuged at 12,000 g for 5 min. The supernatant was further extracted with 0.1% formic acid in 50% acetonitrile, vortexed again, centrifuged, transferred to new tubes, and dried using a speed-vac.

First-dimensional reversed-phase separation

Dried peptides were re-dissolved in 10 mM NH4HCO3 (pH 10), vortexed, and centrifuged at 12,000 g for 3 min. Peptides were fractionated using a homemade C18 column, yielding nine fractions at increasing acetonitrile concentrations (6%, 9%, 12%, 15%, 18%, 21%, 25%, 30%, and 35%). These fractions were pooled into three combined fractions (6% + 15% + 25%, 9% + 18% + 30%, 12% + 21% + 35%) and dried by vacuum concentrator for proteomic analysis.

Nano-LC-MS/MS analysis

Peptides were analyzed by nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer coupled with an EASY nLC 1200 system

(Thermo Fisher Scientific). Dried peptides were re-dissolved in Solvent A (0.1% formic acid in water) and loaded onto a 2-cm trap column packed with ReproSil-Pur C18-AQ beads. Peptides were separated on a 15-cm analytical column packed with 1.9 µm ReproSil-Pur C18-AQ beads using a 75-minute gradient (4–100% Solvent B: 0.1% formic acid in 80% acetonitrile) at 600 nL/min. Ionized peptides were analyzed by mass spectrometry in a data-dependent mode, acquiring MS1 spectra at a resolution of 120,000 and MS2 spectra at a rapid scan rate following fragmentation by higher-energy collision dissociation (HCD). Dynamic exclusion was applied for 12 s to prevent repeated scanning of the same ions.

Peptide and protein identification

Raw MS files were processed using the "Firmiana" proteomic cloud platform and matched against the human NCBI RefSeq protein database (32,015 entries, updated 04-07-2013) via Mascot 2.4. Parameters included up to two missed cleavages, 20 ppm precursor tolerance, and 50 mmu product ion tolerance. Carbamidomethylation was set as a fixed modification, while N-acetylation and methionine oxidation were variable modifications. A target-decoy approach ensured a peptide and protein false discovery rate (FDR) below 1%, with a peptide ion score cutoff of 20. Proteins identified by at least two unique peptides were selected for further analysis.

Label-free protein quantification

Protein quantification was performed using "Firmiana," based on extracted-ion chromatograms (XIC) and intensity-based absolute quantification (iBAQ). Peptide and protein abundances were calculated, and "match between runs" was applied for improved comparability across samples. Protein abundance was expressed as fraction of total (FOT), adjusted and scaled for presentation.

Missing value imputation

Missing values for proteins detected in more than 30% of samples were imputed with a small constant (1E-5) and log2 transformed for downstream analysis, as described in previous proteomic studies[14].

Quality control

Mass spectrometry performance was monitored using HEK293T cell lysates analyzed every three days. Spearman's correlation among quality control runs averaged 0.90 (range: 0.82–0.95), confirming data reproducibility and reliability.

Differential expression analysis

Proteins expressed in over 30% of samples were selected for differential expression analysis between AMI and HC. Differentially expressed proteins (DEPs) were identified using the Wilcoxon rank-sum test, with significance criteria set at $p < 0.05$, $|\log2 \text{ fold change}| > 0.5$, and mean expression levels > 0.01 in both AMI and HC groups. DEPs were classified as either up-regulated or down-regulated based on their expression profiles.

Functional enrichment analysis

Functional enrichment analyses were conducted using the ClusterProfiler[15], enrichR[16], Gene Set Enrichment Analysis (GSEA)[17], and Metascape (https://metascape.org/) tools. Statistical significance for enrichment results was defined as an adjusted or raw p-value < 0.05.

Clustering and clinical correlation analysis

The DEPs expression profiles were analyzed and clustered using Non-Negative Matrix Factorization (NNMF) combined with the Dung Beetle Optimization (DBO) algorithm to optimize clustering performance. Data dimensionality was reduced to three dimensions via NNMF to retain key characteristics while simplifying cluster analysis. DBO was applied to automatically optimize the number of clusters (range: 2–20) and distance metrics, maximizing the mean silhouette coefficient to evaluate cluster quality. Algorithm parameters included a population size of 10 and a maximum of 50 iterations.

Clinical differences between two clusters were assessed. DEPs between clusters were identified and further analyzed using enrichR. These proteins were intersected with previously identified DEPs between AMI and HC. Spearman correlation analysis was performed to evaluate relationships among selected proteins and their correlations with blood biochemical parameters, specifically the percentages of monocytes and neutrophils.

GEO data acquisition and processing

Gene expression data from peripheral whole blood samples of AMI patients and HC were obtained from two GEO datasets. GSE60993 includes 17 AMI and 7 HC samples profiled using the Illumina HumanWG-6 v3.0 platform (GPL6884). GSE61144 contains 6 AMI and 10 HC samples analyzed using the Sentrix Human-6 v2 platform (GPL6106). GSE48060 comprises 24 AMI and 21 HC samples derived from blood nucleated cells and measured on the Affymetrix Human Genome U133 Plus 2.0 platform (GPL570).

Raw data were downloaded from GEO, subjected to background correction and normalization, and annotated with gene symbols. For genes with multiple probes, the probe exhibiting the highest expression value was selected for subsequent analysis.

Meta-analysis and AMI-related proteins identification

Outlier samples within each study were identified using Bolstad's relative log expression method[18]. Although GSE48060 comprises blood nucleated cells while the other datasets focus on whole blood, we conducted differential expression analysis for each dataset separately using the LIMMA package[19] with FDR adjustment[20].

For each differentially expressed genes (DEGs), we recorded statistical measures, including t-statistic, beta, p-value, log fold change, Cohen's D, and Hedges' adjusted g.

Meta-analysis was conducted using an inverse variance model, as recommended by Ramasamy et al.[21], to estimate summary effect sizes. Pooled effect sizes were calculated with Hedges' adjusted g using a fixed-effects model via the R package meta[22] to account for between-study heterogeneity. Meta-DEGs were identified based on an FDR-adjusted p-value < 0.05, |meta Hedges| > 0.5, and their presence across all three studies. This strategy minimized the potential impact of tissue source heterogeneity and enabled us to incorporate as many DEGs as possible from both peripheral blood and blood nucleated cells.

Subsequently, the up- and down-regulated proteins identified in proteomic analyses were intersected with up- and down- regulated Meta-DEGs in transcriptomic analysis. The intersected proteins were then subjected to univariate logistic regression analyses to identify proteins associated with AMI.

Weighted gene correlation network analysis

Weighted Gene Co-expression Network Analysis (WGCNA) was performed using the R package WGCNA[23] to construct a co-expression network from quantifiable proteins and identify significant modules. A scale-free topology fit index (R^2) of 0.9 ensured robust network properties. The adjacency matrix was transformed into a topological overlap matrix (TOM) to reduce noise and eliminate weak correlations. Modules were identified using parameters: soft-threshold power (β) = 2, minimum module size = 50, and merge cut height = 0.25 with the "cutreeDynamic" function. The co-expression network was visualized using Cytoscape v3.10.3[24]. Proteins from the key module were intersected with previously identified DEPs from AMI and HC groups, highlighting overlapping candidates for subsequent analyses.

Feature engineering based on ISPSO

In most studies, ML is used to build predictive or diagnostic models. Feature importance is typically evaluated and key features are selected only after the model is built. This post-hoc approach has several drawbacks. It may miss crucial interactions between features that only become apparent after model construction, and it can lead to overfitting, making the model too specific to the training data and less effective on new data.

In this study, we use a wrapper method for feature selection that is integrated into the model construction phase, with the K-Nearest Neighbors (KNN) algorithm as the base learner. Our novel PSO variant, ISPSO, leverages the information gain ratio to assess feature importance and applies a probabilistic approach with penalty terms to efficiently remove redundant and irrelevant features.

We divided 98 samples (48 AMI patients and 50 HC) into training and testing sets in a 7:3 ration and conducted 100 model iterations. In each iteration, the KNN algorithm predicted outcomes using the current feature set. We then analyzed how often and how much each feature contributed to the model's predictions. By combining the model's overall predictive performance with the individual feature contributions, we consistently identified the features that were key to accurate predictions.

We comprehensively assessed the algorithm's performance using a variety of metrics, including minimum classification error (MCE), precision, recall, F1-score, and the area under the receiver operating characteristic (ROC) curve (AUC). This integrated approach not only identifies the most relevant features but also ensures that the model remains robust and generalizable. Ultimately, leveraging both ISPSO and wrapper methods provides a more effective and nuanced means of feature engineering compared to traditional post-hoc approaches.

Multiple data validation

To validate the key proteins, we analyzed several atherosclerosis (ATH) bulk datasets processed from raw data. These include GSE43292 (32 atheroma plaques and 32 intact tissues), GSE100927 (29 atherosclerotic carotid arteries and 12 controls, 26 atherosclerotic femoral arteries and 12 controls, and 14 atherosclerotic infrapopliteal arteries and 11 controls), and GSE120521 (4 stable and 4 unstable carotid plaques). This comprehensive validation ensures consistency and robustness in our analysis.

We utilized the GSE159677 dataset, which comprises calcified atherosclerotic core (AC) plaques and patient-matched proximal adjacent (PA) carotid artery tissues from three patients undergoing carotid endarterectomy. Raw count matrices were downloaded from GEO and processed using Seurat (v5)[25]. Initially, SCTransform normalization was applied to filter out low-quality cells during preliminary clustering[26]. Doublets were detected and removed using the scDblFinder package, which generated artificial doublets based on clustering, with consistently detected doublets over six iterations excluded[27]. To correct for ambient RNA contamination, we applied DecontX from the celda package, and the decontaminated matrices were subsequently integrated into Seurat objects[28]. Quality control criteria retained cells with more than 200 uniquely expressed genes, 800–6000 UMIs, less than 10% mitochondrial reads, and less than 5% hemoglobin reads to minimize erythrocyte contamination. The count matrices were then normalized again using SCTransform (vst.flavor = v2) to adjust for sequencing depth variability and to facilitate variable gene selection and dimensionality reduction. Additionally, cell cycle effects were regressed out to prevent clustering bias. Principal component analysis (PCA) was performed for dimensionality reduction, with the top 30 PCs used for clustering via a Shared-Nearest-Neighbors (SNN) Louvain algorithm. Uniform Manifold Approximation and Projection (UMAP) was subsequently employed for cluster visualization. To address batch effects, we evaluated several integration methods (FastMNNIntegration, CCAIntegration, RPCAIntegration, HarmonyIntegration, and JointPCAIntegration) using Local Inverse Simpson's Index (LISI) scores[29], kBET[30], principal component regression, and Cluster Average Silhouette Widths (cASW) to assess performance. Finally, clusters containing fewer than 400 cells were removed from the analysis, and cell annotation was performed manually.

MI data were obtained from publicly available single-nucleus RNA (snRNA) and spatial transcriptomics RNA (stRNA) datasets published by Kuppe C et al.[31]. For the snRNA dataset, we extracted data specifically from

the following regions: Control (CTRL), ischemic zone (IZ), ischemic-border zone (IZ-BZ), border zone-remote zone (BZ-RZ) and remote zone (RZ). For the stRNA dataset, we selected only the control, IZ, BZ, and RZ regions from samples collected 2 days post-MI, which were processed using Scanpy[32]. To enable statistical analysis of gene expression across different anatomical regions, we integrated the datasets and applied Harmony to remove batch effects. All analyses were subsequently conducted using these pre-processed data.

Protein quantification by enzyme-linked immunosorbent assay (ELISA)

Plasma samples were collected from patients with AMI and HC. The expression levels of CLTC, CAMP, CTNNB1, IQGAP1, MANBA, ORM1, SPP1, FUBP3, and PSME1 were quantified using commercial ELISA kits (Jianglai Bio, Shanghai, China), according to the manufacturer's instructions. Absorbance was measured using a microplate reader, and protein concentrations were calculated based on standard curves generated for each target.

Statistical analysis

R 4.3.0 and Python 3.12.3 were used for bioinformatics analysis, and MATLAB 2022 was employed for feature selection and clustering. For clinical data, comparisons between two groups were performed using t-tests for continuous variables that met the normality assumption, and Wilcoxon rank-sum tests for those that did not. For comparisons of categorical variables across multiple groups, chi-square tests were used when the sample size was sufficient and the expected frequencies met the assumptions of the test. Fisher's exact tests were applied as an alternative when these assumptions were violated due to small sample sizes or low expected cell counts. For transcriptomic and proteomic expression analyses, individual genes were compared between two groups using Wilcoxon rank-sum tests. Likewise, differential expression analyses for clustering data were conducted with Wilcoxon rank-sum tests. A P value < 0.05 was considered statistically significant. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$; $P > 0.05$: NS.

Data sources and software packages used

Data sources

Plasma Samples: Collected from 48 patients with AMI and 50 HC. Gene expression data were sourced from GEO datasets (GSE60993, GSE61144, GSE48060) for AMI and validated with ATH-related datasets (GSE43292, GSE100927, GSE120521, and GSE159677). Single-nucleus RNA and spatial transcriptomics data for MI were obtained from datasets published by Kuppe C et al. Access to the original dataset requires reasonable request and permission from the corresponding author.

Software and packages

Proteomic Analysis: Firmiana Proteomic Cloud Platform (https://phenomics.fudan.edu.cn/firmiana/login/) was used for protein identification and quantification. Mascot (v.2.4, Matrix Science Inc., London, UK, https://www.matrixscience.com/) was employed for database searching against the human NCBI RefSeq protein database. Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, https://www.spectralabsci.com/equipment/thermo-scientific-q-exactive-hf-x-quadrupole-orbitrap-mass-spectrometer-with-easy-nlc-1000/) was used for nano-LC-MS/MS analysis. High-performance liquid chromatography system was used in conjunction with the mass spectrometer. Bioinformatics and Statistical Analysis: R(v.4.3.0, https://www.r-project.org/), EnrichR(v.3.4,https://cran.r-project.org/web/packages/enrichR/index.html), ClusterProfiler(v.4.14.6, https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) package[15] and Metascape (v3.5.20250101, https://metascape.org/) for functional enrichment analysis. STRING (v.11, https://string-db.org/) and Cytoscape(v3.10.3, https://cytoscape.org/)[24] for PPI network construction. Meta package (v.8.0.2, https://cran.r-project.org/web/packages/meta/index.html)[22] was used for meta-analysis of gene expression data. Bolstad's Relative Log Expression Method[18] was used for identifying outlier samples. LIMMA package (v.3.62.2, https://bioconductor.org/packages/release/bioc/html/limma.html)[19] was used for DEGs. WGCNA package (v.1.73, https://cran.r-project.org/web/packages/WGCNA/index.html) was used for constructing and analyzing co-expression networks. Seurat package (v.5.2.1, https://satijalab.org/seurat/)[25] was used for scRNA and snRNA analysis. Scanpy(v.1.11.1, https://scanpy.readthedocs.io/en/stable/)[32] was used for spatial transcriptome data analysis. LISI scores [29], kBET[30], Principal Component Regression, and cASW to assess integration performance. MATLAB (v.R2024b, https://www.mathworks.com/products/matlab.html) was used to implement the ISPSO for feature selection in conjunction with the KNN algorithm. NNMF was applied for dimensionality reduction, while the DBO algorithm was employed to optimize the K-Medoids clustering process.

Results

Workflow

The detailed research process can be found in Fig. 1. This flowchart was created using biorender (https://www.biorender.com/).

Screening for DEPs and functional enrichment analysis

We identified 437 DEPs in AMI cases compared to HC ($|\log_2 FC| > 0.5$, P-value < 0.05), with 291 up-regulated and 146 down-regulated (Supplementary file2). These DEPs were visualized using volcano plots and heatmaps (Fig. 2A-B). Functional enrichment analyses of the DEPs revealed pathways critical to inflammation, immunity, metabolism, and cellular stress responses (Fig. 2C, Supplementary file3). Key pathways include the complement and innate immune systems, neutrophil extracellular trap (NET) formation, and ferroptosis, underscoring immune and inflammatory regulation. Metabolic pathways such as oxidative phosphorylation, protein

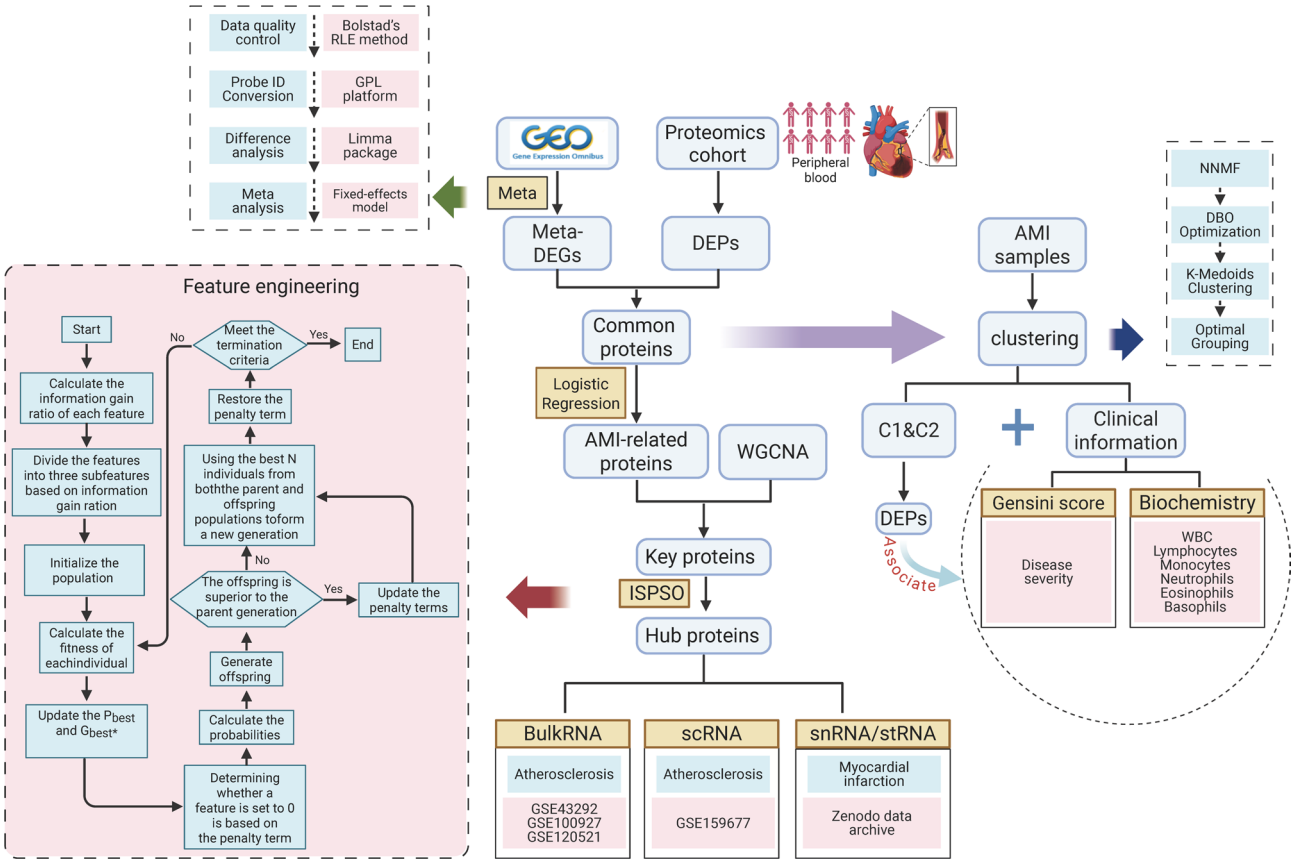


Fig. 1. Flowchart of data quality control, analysis, and feature engineering procedures for proteomics cohort study.

metabolism, and reactive oxygen species signaling indicate oxidative stress and energy dysregulation. Structural and signaling pathways, including supramolecular fiber organization, ECM-receptor interaction, focal adhesion, and VEGFA-VEGFR2 signaling, point to tissue remodeling and vascular changes. Processes such as endocytosis, phagosome formation, and cytoplasmic translation further highlight their roles in processing damage signals and necrotic cells, modulating local inflammation and clearance mechanisms, and ensuring the rapid replenishment of essential proteins during the progression of acute myocardial infarction.

GSEA confirmed these findings, revealing enrichments in vascular remodeling, immune response, oxidative stress, and metabolic dysfunction (Supplementary file4). Notable enrichments include vascular development, extracellular matrix organization, and monocyte chemotaxis, which emphasize inflammation and tissue remodeling; metabolic dysregulation in pyruvate metabolism and fatty acid beta-oxidation; and signaling pathways such as TGFB receptor and Hedgehog signaling, which suggest roles in fibrosis and cellular communication (Fig. 2D). These results underscore the complex interplay of inflammation, metabolism, and tissue repair in AMI pathophysiology.

Using NNMF and the DBO algorithm for K-Medoids clustering, we divided AMI patients into two clusters: C1 ($n=21$) and C2 ($n=27$) (Fig. 3A, Supplementary file5). The majority of clinical characteristics showed no statistically significant differences between the subgroups (Supplementary file6). Specifically, demographic variables (gender, age), risk factors (hypertension, diabetes, smoking, drinking), blood pressure, anthropometric measurements (height, weight), cardiac function (left ventricular ejection fraction, cardiac output), coagulation parameters (thrombin time, prothrombin time, activated partial thromboplastin time, prothrombin time international normalized ratio, D-dimer, fibrinogen), inflammatory markers (interleukin-6), and cardiac enzymes (NT-proBNP, myoglobin, creatine kinase, CK-MB) all had p-values exceeding 0.05. However, fibrin degradation products (FDP) differed significantly between the two subgroups ($p=0.0427$), suggesting a potential biological distinction related to fibrinolysis or coagulation pathways (Supplementary file 6). Heatmaps showed distinct protein expression patterns for each cluster, highlighting unique immune cell profiles (Fig. 3B). C2 had significantly higher Gensini score, WBC count, and neutrophil percentage, while C1 exhibited elevated lymphocytes, monocytes, eosinophils, and basophils ($p<0.05$) (Fig. 3C). These results suggest that while the subclusters represent distinct inflammatory responses at the proteomic and cellular levels, they do not translate into major differences in conventional clinical presentation or risk profiles.

Differential analysis between C1 and C2 revealed 200 cluster-specific DEPs (SupFig.1 A, Supplementary file7). C2 was enriched in pathways related to cellular stress, apoptosis, and programmed cell death-such as apoptosis-induced DNA fragmentation, cellular senescence, and MAPK/ERK signaling targets (Supplementary

file8), suggesting heightened immune regulation, cellular damage, and stress adaptation, likely indicative of a prolonged inflammatory phase. In contrast, C1 was enriched in pathways associated with angiogenesis, vascular remodeling, and extracellular matrix organization, including VEGFA-VEGFR2 signaling, positive regulation of angiogenesis, and extracellular structure organization (Supplementary file 8). Additional enrichment in glutathione metabolism and endothelin signaling points to oxidative stress adaptation and vascular repair mechanisms. These results imply that C1 reflects a repair and remodeling state, while C2 represents an immune-regulatory, stress-adaptive state, offering insights for tailored therapeutic strategies (SupFig.1B). Further analysis identified 15 upregulated and 5 downregulated proteins overlapping with AMI-DEPs (SupFig.1 C). Spearman correlation analysis revealed that MYLK was positively correlated with neutrophil percentage (SupFig.1D). Overall, these findings highlight distinct immune and protein expression profiles between C1 and C2, with potential implications for immune regulation in AMI.

Meta-analysis and ML

We identified 1383 DEGs from three AMI datasets compared to HC (|meta Hedges| > 0.5 and Meta FDR < 0.05) (Fig. 4A, Supplementary file9). By intersecting these DEGs with DEPs, we obtained 43 up-regulated and 20 down-regulated proteins (Fig. 4B). Univariate logistic regression further screened out 36 AMI-related DEPs (Fig. 4C, Supplementary file10).

To explore the biological characteristics of AMI, we built a protein co-expression network using WGCNA. This generated a scale-free network (R^2 = 0.9, soft-threshold power $\beta = 2$) and identified 49 co-expression modules (Fig. 4D-E, Supplementary file11). The turquoise module, which showed the strongest association with AMI, had mean Pearson correlations of 0.203 in our co-expression network and 0.213 in the STRING database (SupFig.2 A). Protein interactions in this module were visualized using STRING (SupFig.2B). Pathway enrichment analysis revealed that turquoise module proteins participate in key processes, including lipid metabolism (e.g., plasma lipoprotein and triglyceride metabolism), immune response (e.g., humoral immunity), blood coagulation and fibrin formation, as well as structural remodeling (e.g., intermediate filament organization and ECM remodeling) (SupFig.2 C). These pathways highlight the interplay of lipid dysregulation, inflammation, vascular processes, and tissue repair in AMI.

We then identified 17 overlapping proteins between the turquoise module and AMI-related DEPs, including ACTR3, CLTCL1, CALML5, CAMP, CLTC, CTNNB1, FGD3, FUBP3, GNS, HPCAL1, IQGAP1, MANBA, ORM1, PSME1, RPS27A, SPP1 and TMPRSS12 (SupFig.3 A). Metascape analysis showed that these proteins are involved in key AMI-related pathways, such as neutrophil degranulation (indicating innate immune responses), the PID RAC1 pathway (linked to actin cytoskeleton organization and cell migration), and RUNX3-mediated transcriptional regulation (associated with immune function and inflammation) (SupFig.3B).

Using the ISPSO algorithm with a KNN model based 17 key proteins expression profiles, we analyzed a cohort of 98 samples-randomly split into a 7:3 ratio-over 100 training iterations (Supplementary file12). The model performed robustly, with mean metrics of training AUC (0.971), valid-gBest (0.096), valid-loss (0.126), valid-precision (0.883), valid-recall (0.874), and validation AUC (0.944) (SupFig.3 C). Based on feature occurrence frequencies greater than 0.5, we identified nine key proteins (CAMP, CLTC, CTNNB1, FUBP3, IQGAP1, MANBA, ORM1, PSME1, and SPP1) as core features.

Protein function exploration and verification

Based on single-protein GSEA (SupFig.4 A, Supplementary file13), SPP1 is involved in activating and positively regulating immune and acute inflammatory responses. PSME1 plays a key role in suppressing intrinsic apoptotic signaling, processing antigens, and modulating inflammatory responses to antigenic stimuli. ORM1 is linked to inflammatory responses, lipid localization, and foam cell differentiation, while MANBA contributes to extrinsic apoptotic signaling and regulates lipid metabolism and plasma lipoprotein levels. IQGAP1 controls cytoskeleton organization, actin filament processes, and blood vessel morphogenesis. FUBP3 is associated with B cell-mediated immunity, inflammasome signaling, and protein secretion. CTNNB1 participates in the negative regulation of canonical Wnt signaling, wound healing, and DNA metabolism. CAMP mediates responses to tumor necrosis factor, innate immune responses, and cholesterol efflux, whereas CLTC regulates endothelial cell migration, lipid storage, and reactive oxygen species metabolism.

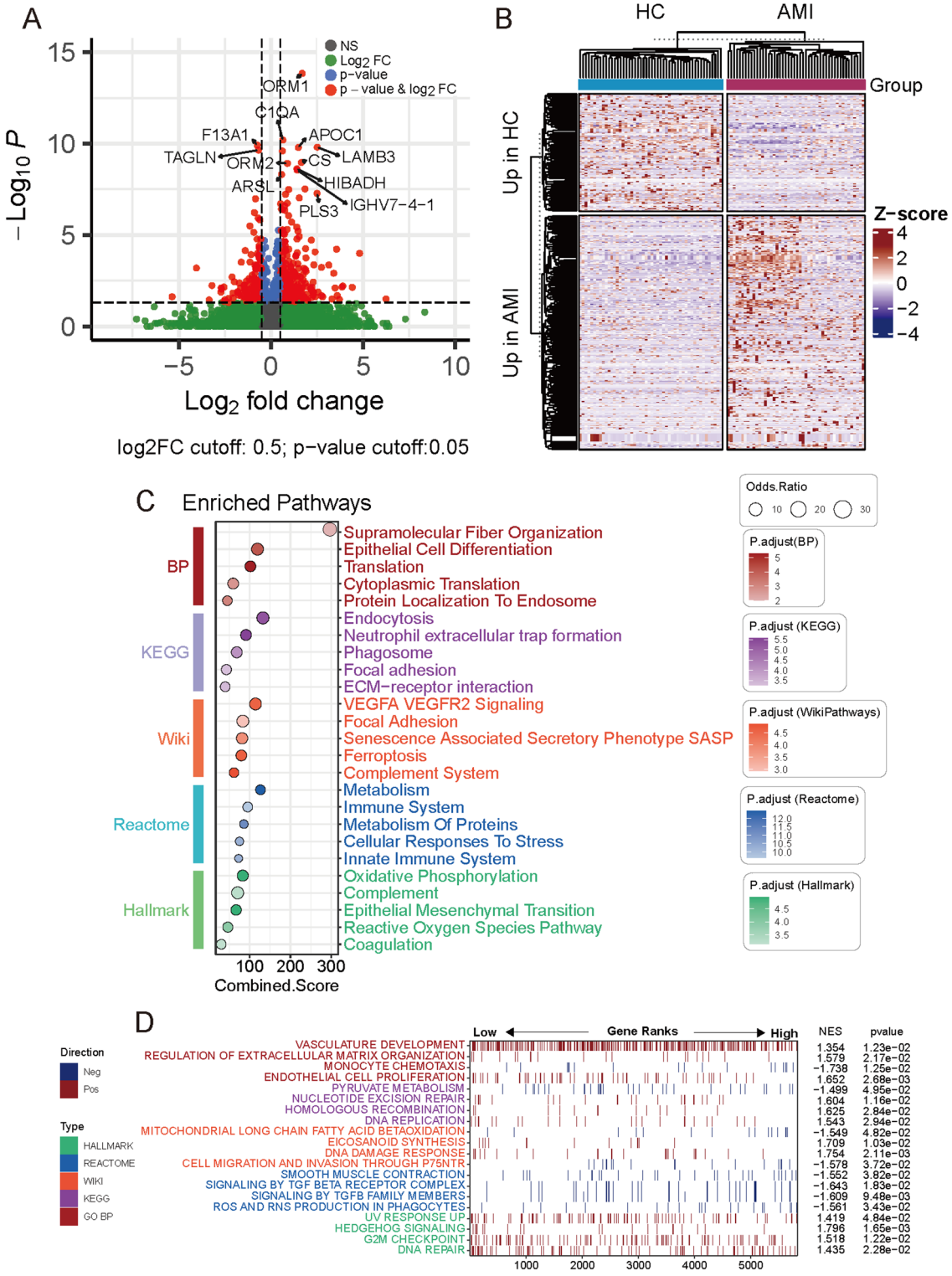
Expression analysis of these nine core proteins showed significant differences across various conditions and vascular sites. In atherosclerotic plaques versus normal tissues, CAMP, CLTC, MANBA, PSME1, and SPP1 were significantly upregulated, while CTNNB1, FUBP3, IQGAP1, and ORM1 showed no significant changes (SupFig.4B). In unstable plaques, ORM1, PSME1, and SPP1 were markedly elevated compared to stable plaques (SupFig.4 C).

Across vascular sites, CAMP and PSME1 showed no differences in expression among the atherosclerotic carotid artery, atherosclerotic femoral artery, and atherosclerotic infra-popliteal artery. In contrast, CLTC, IQGAP1, and ORM1 were significantly upregulated in the atherosclerotic carotid artery, while no significant differences were observed at the other sites. Moreover, CTNNB1 was significantly downregulated in the atherosclerotic carotid artery with no differences elsewhere. FUBP3 was markedly underexpressed in both the atherosclerotic carotid and femoral arteries. Additionally, MANBA and SPP1 were significantly overexpressed across all three vascular sites (SupFig.4D).

Validation at the single cell levels based on ATH

Integration method selection

We developed a standardized pipeline to quality-control and process six raw sequencing libraries, which improved cell cluster separation and cohesion after filtering. We then used a subset of these libraries to benchmark several integration tools (FastMNNIntegration, CCAIntegration, RPCAIntegration, HarmonyIntegration,



and JointPCAIntegration) for batch correction. The performance of these tools was evaluated using UMAP visualization (SupFig.5), cASW (SupFig.6 A), principal component regression (SupFig.6B), LISI scores (SupFig.6 C-D), kBET (SupFig.6E), and metrics for batch removal, preservation of biological variation, and clustering purity. Harmony provided the best balance for cell type separation and clustering performance. Using Harmony, we integrated the libraries to obtain high-quality cells across 20 clusters, with UMAP embeddings confirming effective batch mixing, preservation of biological signals, and well-separated clusters (Fig. 5A).

Cell type annotation

The clusters were annotated into major cell types, including macrophages, synthetic smooth muscle cells (SMCs), fibroblasts, endothelial cells, T cells, and NK cells (Fig. 5B). Cell type identification was based on marker gene expression: B cells were identified by the markers CD79A, CD79B, and MS4A1; endothelial cells by PECAM1,

◀**Fig. 2**. DEPs visualization and functional enrichment analysis. (**A**) The DEPs identified in the differential expression analysis, with log2 fold changes on the x-axis and -log10(p-value) on the y-axis. Significantly up and down-regulated proteins in AMI samples are marked in red, and non-significant genes in grey. (**B**) This heatmap represents the expression levels of DEPs identified in the proteomics analysis of AMI samples. Each row corresponds to a specific protein. Each column represents individual samples, AMI samples are marked in red, healthy samples are marked in blue. The gradient ranges from blue to red, where blue denotes low expression levels and red denotes high expression levels. (**C**) Pathway enrichment for DEPs. The x-axis represents the Combined.Score, the color gradient represents -log10p, and the size gradient represents Odds. Ratio. (**D**) GSEA Enrichment analysis for all proteins. The x-axis represents the protein location. *AMI* acute myocardial infarction, *DEPs* differentially expressed proteins, *GSEA* gene set enrichment analysis, *Hallmark* hallmark gene sets, *Reactome* reactome pathway database; Wiki, wikipathways, *KEGG* Kyoto encyclopedia of genes and genomes, *BP* biological process.

VWF, and CLDN5; smooth muscle cells (SMCs) by ACTA2, MYH11, and TPM2; and fibroblasts by COL6A1, COL6A2, and DCN. Synthetic SMCs were characterized by the co-expression of COL6A2, ACTA2, and DCN, while NK cells expressed NKG7, NCAM1, and KLRD1, and T cells were marked by CD3D, CD3E, and CD3G. Finally, macrophages were identified through the expression of CD68, CD14, and C1QA. (Fig. 5C).

Core protein expression across cell types and spatial regions

The expression levels of the 9 core proteins (CLTC, CAMP, CTNNB1, FUBP3, IQGAP1, MANBA, ORM1, PSME1 and SPP1) varied across cell types (Fig. 5D). SPP1 and MANBA were highly expressed in macrophages, reflecting their roles in immune responses, while IQGAP1, CLTC, CTNNB1 and PSME1 were predominantly expressed in multiple cell types. Specifically, we found that CAMP and FUBP3 are minimally expressed across cell types in ATH, suggesting their elevated levels may be specifically associated with AMI rather than ATH.

In addition, examining all cells collectively revealed that CTNNB1, PSME1, IQGAP1, and CLTC were highly expressed in the proximal adjacent region, whereas SPP1 was highly expressed in the atherosclerotic core (Fig. 5E). Focusing on individual cell types (SupFig.7), CLTC was upregulated in fibroblasts and macrophages but downregulated in endothelial cells within the atherosclerotic core group. CTNNB1 was downregulated in T cells, endothelial cells, SMC, fibroblasts, macrophages, B cells, and synthetic SMC in the atherosclerotic core group. IQGAP1 was downregulated in fibroblasts and macrophages, and MANBA was also downregulated in macrophages in the atherosclerotic core group. PSME1 showed downregulation in T cells, endothelial cells, SMC, macrophages, and synthetic SMC, but was upregulated in fibroblasts in the atherosclerotic core group. Finally, SPP1 was predominantly upregulated in macrophages in the atherosclerotic core group.

Validation at the single cell levels based on MI

Core protein expression and cellular dynamics in MI

We directly utilized the preprocessed data provided by Kuppe C et al.[31], ensuring standardized and reliable analysis. UMAP analysis of snRNA data revealed distinct clustering of cell types across MI regions, including CTRL, IZ, BZ and RZ. Identified cell populations included adipocytes, cardiomyocytes, endothelial cells, fibroblasts, lymphoid cells, mast cells, myeloid cells, neuronal cells, pericytes, and vascular smooth muscle cells (Fig. 6A). FUBP3, MANBA, IQGAP1, CTNNB1, and CLTC are broadly expressed in cardiomyocytes, underscoring their vital roles in maintaining myocardial function and structural integrity. By contrast, SPP1 is predominantly expressed in circulating cells and myeloid cells, reflecting its involvement in immune responses and inflammation. Notably, ORM1 is absent in myocardial tissue, suggesting it exerts its effects systemically rather than acting directly within the myocardium (Fig. 6B). At the tissue-region level, the cellular expression percentages of FUBP3, MANBA, IQGAP1, CTNNB1, and CLTC gradually increase from the infarct zone (IZ) to the remote zone (RZ), with a similar trend observed in both the control and RZ groups (Fig. 6C-D). In cardiomyocytes specifically, these genes also show a progressive upregulation from IZ to RZ (SupFig.8).

In AMI plasma proteomics, FUBP3 and PSME1 were downregulated, whereas ORM1, SPP1, MANBA, IQGAP1, CTNNB1, CLTC, and CAMP were upregulated. Notably, cell-level analyses in both ATH and MI consistently demonstrated that IQGAP1, CTNNB1, and CLTC exhibited an upregulation pattern-being predominantly expressed in the proximal adjacent region in ATH and showing a gradual increase from the IZ to the RZ in MI. The elevated levels of IQGAP1, CTNNB1, and CLTC in the plasma of AMI patients likely reflect a systemic response to myocardial injury, as these proteins are released from affected cells. In tissue analyses, however, the highest expression is observed in the proximal adjacent regions rather than in the infarct core. This discrepancy can be attributed to several factors: the infarct core is characterized by extensive cell death and necrosis, which results in reduced protein synthesis, while the adjacent regions contain viable cells actively engaged in repair, remodeling, and compensatory mechanisms, leading to higher local expression. Furthermore, these proteins may be secreted from the active remodeling regions, accumulating in the circulation and thus appearing upregulated in blood proteomics. Overall, the difference between tissue-level and blood findings underscores the complex interplay between localized tissue repair processes and systemic biomarker release following myocardial infarction. Moreover, the concurrent upregulation of SPP1 and downregulation of PSME1 across AMI plasma proteomics as well as ATH and MI single-cell analyses suggest that distinct molecular pathways are at play in the disease state. Specifically, the increased expression of SPP1, a protein known for its role in promoting inflammation and tissue remodeling, points toward an active inflammatory response that may contribute to disease progression. In contrast, the reduction in PSME1 levels could imply impaired proteasome activity or altered antigen processing, potentially affecting cellular homeostasis and immune regulation.

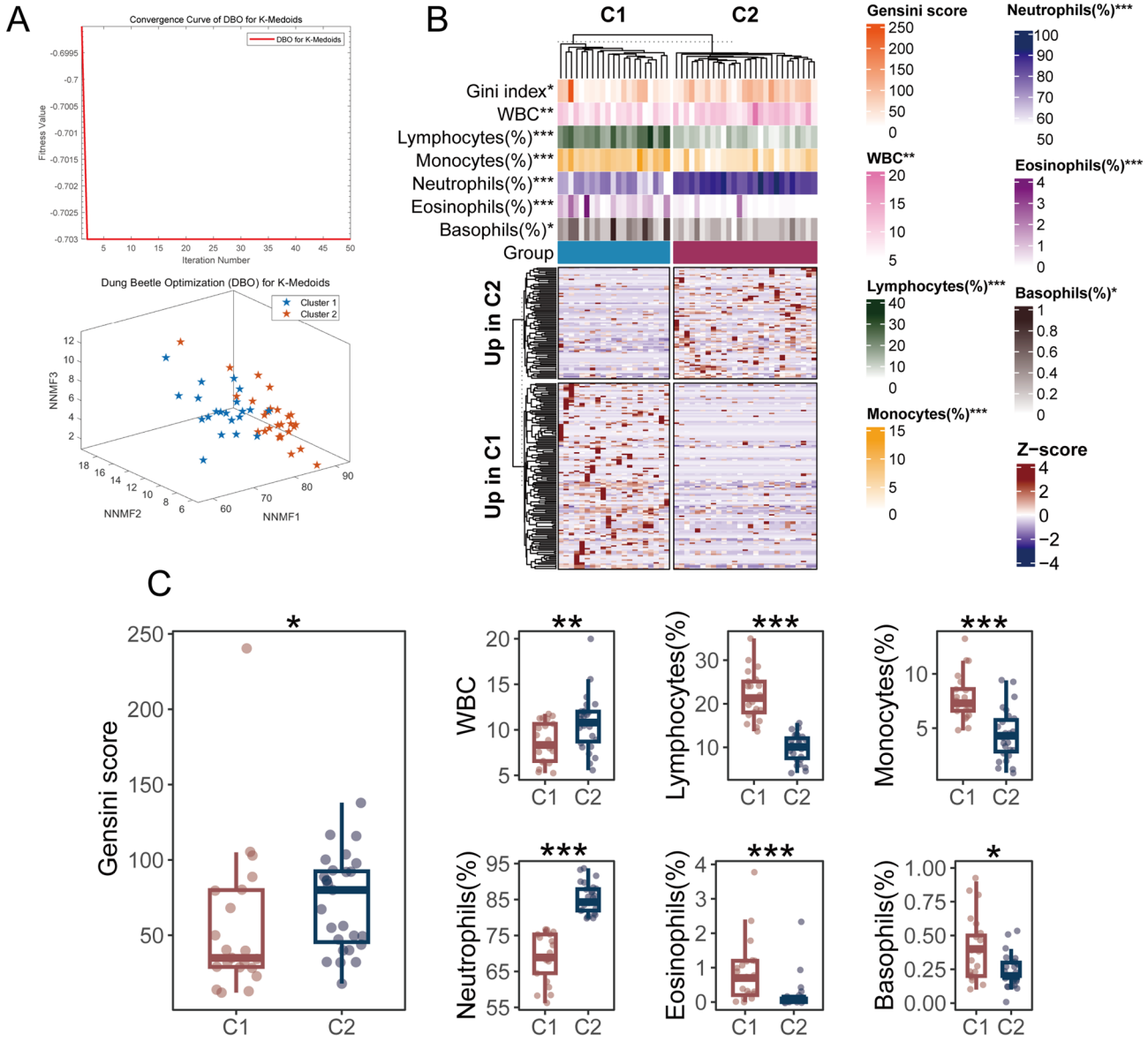


Fig. 3. Multi-dimensional clustering and immune cell distribution analysis. (**A**) The convergence curve of the DBO-based K-Medoids clustering approach demonstrates a steady reduction in objective function value over iterations. A three-dimensional scatterplot (NNMF1, NNMF2, NNMF3) illustrates two distinct clusters (C1: orange; C2: blue). (**B**) Immune cell proportion analysis and gene expression heatmap for C1 and C2 display the Gensini score, WBC counts, and proportions of lymphocytes, monocytes, neutrophils, eosinophils, and basophils. Significant differences between clusters in immune parameters are highlighted. Distinct gene expression patterns are shown for genes upregulated in C1 and C2. (**C**) Boxplots visualize significant differences in the Gensini score, WBC counts, and the proportions of lymphocytes, monocytes, neutrophils, eosinophils, and basophils between C1 and C2. *C1* cluster 1, *C2* cluster 2, *WBC* white blood cells. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Together, these opposing expression patterns not only underscore the complexity of the molecular mechanisms driving these conditions but also highlight the potential of SPP1 and PSME1 as biomarkers or therapeutic targets in the context of cardiovascular and inflammatory diseases.

We utilized spatial transcriptomics data to visualize the spatial distribution of core proteins across different tissue regions two days after MI compared to HC, using the expression levels in HC as a baseline reference to assess core protein expression. The FUBP3, IQGAP1, CTNNB1, CLTC and SPP1 were highly expressed in the IZ, MANBA showed elevated expression in the BZ, and PSME1 was enriched in both the BZ and RZ. In contrast, CAMP and ORM1 exhibited minimal expression in cardiac tissue two days post-MI (Fig. 7A-D, SupFig.9).

Plasma protein levels in AMI and HC

Plasma concentrations of nine proteins-CLTC, CAMP, CTNNB1, FUBP3, IQGAP1, MANBA, ORM1, PSME1, and SPP1-were measured by ELISA in plasma samples from patients with AMI and HC (SupFig.10). Compared to the HC group, levels of CAMP, CTNNB1, IQGAP1, MANBA, ORM1, and SPP1 were significantly elevated in AMI patients ($P < 0.0001$ for each). In contrast, FUBP3 and PSME1 were significantly decreased in the AMI group ($P < 0.001$ and $P < 0.05$, respectively). No significant difference was observed in CLTC levels between the two groups.

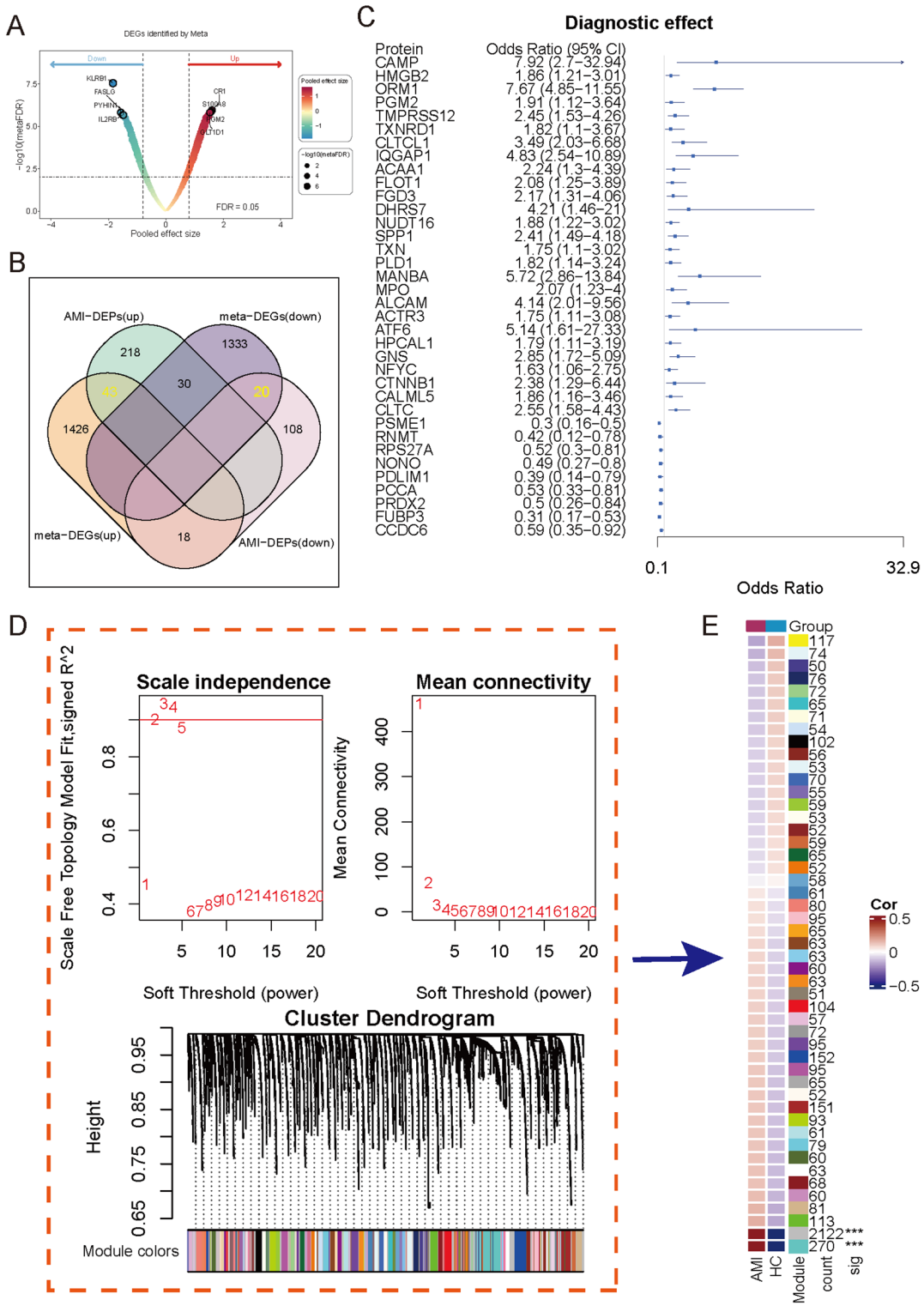
Discussion

In this study, we performed proteomic sequencing on acute AMI patients and HC to identify various DEPs. By integrating meta-analysis of peripheral blood transcriptomes (meta-DEGs) with logistic regression, WGCNA, and advanced feature engineering, we identified key proteins, which were further validated in ATH plaque tissues. Additionally, we explored the cellular and spatial distribution of these key proteins in MI data. Although the enriched pathways of the DEPs were largely similar to those reported in previous studies[33–35], our integrative approach-combining proteomic and transcriptomic data with a feature engineering-oriented screening strategy-led to the discovery of a unique set of key proteins, thereby unveiling potential therapeutic targets.

Functional enrichment analysis underscores the central roles of inflammation, immunity, and metabolic dysfunction in AMI pathophysiology, consistent with prior studies. However, our analysis provides novel insights by revealing specific pathways and molecular interactions that extend beyond traditional inflammation and metabolic disturbances. For instance, while the role of neutrophil extracellular traps (NETs) in thrombotic complications and vascular injury has been previously documented[36,37], our findings highlight specific proteins within this pathway that could serve as potential biomarkers or therapeutic targets in AMI. Similarly, although ferroptosis's contribution to cardiomyocyte death and oxidative stress is recognized[38], the identification of distinct ferroptosis-associated proteins in our dataset suggests new molecular mechanisms and intervention points that could be therapeutically exploited. Moreover, while metabolic dysregulation involving oxidative phosphorylation and reactive oxygen species (ROS) signaling pathways aligns with known mitochondrial dysfunction in AMI[39,40], the detailed characterization of proteins in these pathways provides novel molecular targets for metabolic interventions aimed at reducing ischemic injury. The identification of enriched structural and signaling pathways, such as extracellular matrix (ECM) organization and VEGFA-VEGFR2 signaling, offers deeper insights into mechanisms of vascular remodeling and angiogenesis, highlighting previously underappreciated targets involved in maladaptive remodeling versus reparative angiogenesis[41]. Finally, enrichment in DNA repair mechanisms and cellular stress responses not only underscores the critical adaptive processes during AMI but also identifies specific proteins potentially involved in cellular resilience and repair, offering new avenues for therapeutic modulation to enhance myocardial recovery. Thus, our study advances beyond the existing literature by providing a comprehensive molecular map that highlights novel targets and pathways for personalized treatment strategies in AMI.

The clustering analysis identified two distinct AMI subtypes (C1 and C2), characterized by significant differences in immune cell composition and protein expression patterns. These findings underscore the heterogeneity of immune and inflammatory responses observed in AMI, aligning with prior evidence demonstrating divergent inflammatory trajectories among patients[42]. Cluster 2 (C2), distinguished by elevated neutrophil percentages and enhanced expression of proteins associated with apoptosis and cellular stress responses, likely represents patients experiencing a heightened inflammatory state. Clinically, this phenotype may correlate with increased myocardial injury, more extensive tissue damage, or a prolonged inflammatory phase, which can adversely impact patient outcomes, such as increased risk for heart failure or ventricular remodeling. Neutrophils have been consistently linked with adverse clinical outcomes in AMI due to their potent ability to exacerbate myocardial damage through the release of proteolytic enzymes and pro-inflammatory cytokines[43]. Moreover, the enrichment of pathways such as apoptosis-induced DNA fragmentation and MAPK/ERK signaling in this cluster suggests active cellular stress responses, potentially indicating a subgroup with poor prognosis or greater therapeutic needs. In contrast, Cluster 1 (C1) demonstrated a predominance of lymphocytes, monocytes, eosinophils, and basophils, alongside enriched pathways associated with angiogenesis, vascular remodeling, and extracellular matrix organization, including VEGFA-VEGFR2 signaling and endothelin signaling. Clinically, this subtype likely represents patients who are in a reparative or resolving inflammatory phase, characterized by active vascular repair mechanisms and potential for improved myocardial healing and recovery. The significant enrichment in glutathione metabolism further suggests enhanced oxidative stress adaptation, which could confer protective effects against oxidative myocardial injury, potentially indicating a subgroup with a more favorable prognosis or differential responsiveness to anti-inflammatory and reparative therapies. MYLK emerged as a protein significantly upregulated in AMI patients, showing a robust positive correlation with neutrophil percentages. Clinically, MYLK could serve as a valuable biomarker for identifying patients with heightened neutrophil activity and inflammation, thereby facilitating early risk stratification and tailored interventions. The established role of MYLK in neutrophil migration and adhesion via activation of β2 integrins[44] further emphasizes its potential as a therapeutic target. Inhibiting MYLK function could mitigate neutrophil-mediated myocardial damage, offering a novel strategy for therapeutic intervention, especially for patients exhibiting excessive neutrophil-driven inflammation. Distinguishing these two AMI subtypes provides critical insights into patient-specific immune responses and potential clinical trajectories. Identifying such heterogeneity enables clinicians to stratify patients more accurately, customize therapeutic interventions, and predict clinical outcomes more precisely, ultimately improving patient management and prognosis following AMI.

This study integrates meta-analysis, proteomics, and ML approaches to unravel the rare biomarkers, multiple data verification emphasizing the interplay of systemic and localized protein regulation. The findings provide key insights for biomarker discovery and therapeutic strategies. The identification of 43 upregulated and 20



downregulated proteins through the intersection of DEGs and DEPs highlights shared molecular pathways underlying AMI. Pathway enrichment analysis of the turquoise module from WGCNA reveals critical processes such as lipid metabolism, blood coagulation, and ECM remodeling, which align with the established role of dysregulated lipid processing, thrombotic activity, and tissue repair in AMI[45]. Notably, the involvement of neutrophil degranulation and the PID RAC1 pathway underscores the importance of innate immune responses and cytoskeletal reorganization in AMI.

The use of the ISPSO algorithm for feature selection identifies nine key proteins (e.g., CAMP, SPP1, PSME1, MANBA) with high predictive value for AMI. The robustness of this ML approach, demonstrated by high train-AUC (0.971) and valid-AUC (0.944) scores, underscores its potential for biomarker discovery. Findings from plasma provide accessible biomarkers for various cardiovascular diseases due to the minimally invasive nature of blood sampling. Identifying important proteins initially in plasma and subsequently validating these

◀**Fig. 4**. DEGs and network module analysis for diagnostic biomarker identification. (**A**) Volcano plot of DEGs from meta-analysis. DEGs are plotted based on pooled effect size and -log10(FDR-adjusted p-value). Genes meeting statistical significance (FDR < 0.05) and effect size thresholds (> 0.5) are highlighted, with key upregulated (red) and downregulated (green) genes annotated. (**B**) Venn diagram showing overlap of DEGs in different datasets: The overlap of DEGs identified in AMI-DEG (upregulated and downregulated) and meta-DEGs (upregulated and downregulated) is presented. Shared genes across datasets highlight consistent markers potentially related to diagnostic utility in AMI. (**C**) Forest plot showing the diagnostic effect of key proteins: Odds ratios with 95% confidence intervals (CIs) for selected proteins are displayed. Proteins significantly associated with diagnostic utility ($p < 0.05$) are marked. (**D**) WGCNA results. Scale-free topology model fitting (R^2) and mean connectivity analyses identify the optimal power value for constructing gene networks. A cluster dendrogram of gene modules is shown, with module colors assigned based on co-expression patterns. (**E**) Heatmap of module-trait correlations: Module eigengenes are correlated with clinical traits, such as AMI and HC. Correlation coefficients are color-coded, with significant correlations. *DEGs* differentially expressed genes, *WGCNA* weighted gene co-expression network analysis, *AMI* acute myocardial infarction, *HC* healthy control. $^*p < 0.05$.

proteins in cardiac and plaque tissues offers distinct advantages and significant implications for cardiovascular research. Firstly, the initial identification of biomarkers in plasma is minimally invasive and highly practical for clinical translation, allowing rapid screening and repeated sampling. This facilitates the early detection of disease and longitudinal monitoring of disease progression or therapeutic responses[46]. Secondly, validation of these biomarkers directly in cardiac and plaque tissues provides critical biological context, confirming their relevance and specificity to cardiovascular pathology. Tissue-level validation helps elucidate the precise cellular and molecular mechanisms underlying cardiovascular diseases, enhancing the clinical applicability of blood-based biomarkers. Moreover, integrating blood-based discovery with tissue validation addresses the potential discrepancies between systemic circulation and localized pathology. It strengthens the confidence in biomarker relevance and supports the development of targeted therapeutic strategies[47]. Ultimately, this combined approach of plasma identification followed by tissue-level validation maximizes both clinical utility and biological insight, supporting the discovery of robust and translational biomarkers for cardiovascular diseases.

In the context of AMI, CLTC, CAMP, CTNNB1, IQGAP1, MANBA, ORM1, and SPP1 are upregulated, while FUBP3 and PSME1 are downregulated. Notably, MANBA, CAMP, CLTC, FUBP3, PSME1, and ORM1 have not been previously reported in the context of ATH and MI. MANBA has been identified as a therapeutic target for coronary artery calcification[48]. ORM1 is positively correlated with epicardial adipose tissue (EAT) thickness and is associated with acute-phase responses[49,50]. CAMP participates in the cAMP-PKA axis, a pathway known to play a crucial anti-inflammatory role, suggesting its potential in inflammation resolution. CLTC is involved in receptor-mediated intracellular trafficking and macromolecular endocytosis[51], highlighting its potential role in cellular stress responses. IQGAP1, an important scaffold protein regulating mitochondrial function, exacerbates oxidative stress and mtDNA release in ATH by activating the cGAS-STING pathway and promoting NLRP3 inflammasome-mediated pyroptosis, although its role in MI has not been reported[52]. FUBP3 mediates amyloid-beta-induced neuronal NLRP3 expression in Alzheimer's disease, suggesting possible parallels in inflammatory regulation. CTNNB1 (Beta-catenin) promotes macrophage-mediated acute inflammatory responses after MI[53] and connects Wnt and sphingosine-1-phosphate signaling pathways to vascular remodeling and ATH[54]. Extensive studies have confirmed the involvement of Wnt/β-catenin signaling in the mechanisms of MI and ATH[53,55–59]. PSME1, downregulated in AMI, is known to participate in cancer-related pathways and modulate proteasomal activity, improving the production of class I-binding peptides[60]. It has also been introduced as a potential tumor biomarker based on proteomic analyses[60]. Furthermore, studies suggest that targeting PSME1/2 to activate the Wnt/β-catenin signaling pathway can alleviate osteoporosis[61], raising the question of whether its role in ATH and AMI involves protective or promotive effects, potentially via the same pathway. It can be speculated that PSME1 may exert its effects in ATH and AMI through the Wnt/β-catenin signaling pathway, with its mechanism likely depending on the specific pathological context. In the acute phase of AMI, the downregulation of PSME1 may represent an adaptive response to excessive inflammatory activity, serving a protective role by preventing the overactivation of Wnt/β-catenin signaling, which could otherwise amplify inflammation and exacerbate tissue damage. However, in the chronic phase or during the progression of ATH lesions, low PSME1 expression might impair the reparative and anti-inflammatory functions mediated by Wnt/β-catenin signaling, thereby accelerating disease progression. Further studies are required to explore whether PSME1 plays a dual regulatory role by modulating Wnt/β-catenin signaling and to determine its dynamic expression patterns and signaling activity at different stages of these pathologies. Single-cell and spatial distribution analyses reveal that CAMP, ORM1 and FUBP3 are specifically expressed during the acute phase of AMI, suggesting their potential as diagnostic biomarkers for AMI. Further research is needed to elucidate their precise roles and validate their utility in clinical applications.

This study provides innovative insights into the molecular landscape of AMI by integrating proteomics, transcriptomics, ML, and single-cell and spatial transcriptomics. The identification of nine core proteins and the characterization of two distinct AMI subtypes (C1 and C2) with unique immune and protein expression profiles highlight the heterogeneity of AMI and offer potential biomarkers and therapeutic targets. Advanced ML techniques, such as the ISPSO algorithm, achieved high prediction accuracy, underscoring their utility in biomarker discovery. Single-cell and spatial analyses further revealed the cellular and regional dynamics of protein expression. However, limitations include a small sample size and a focus on the acute phase of AMI,

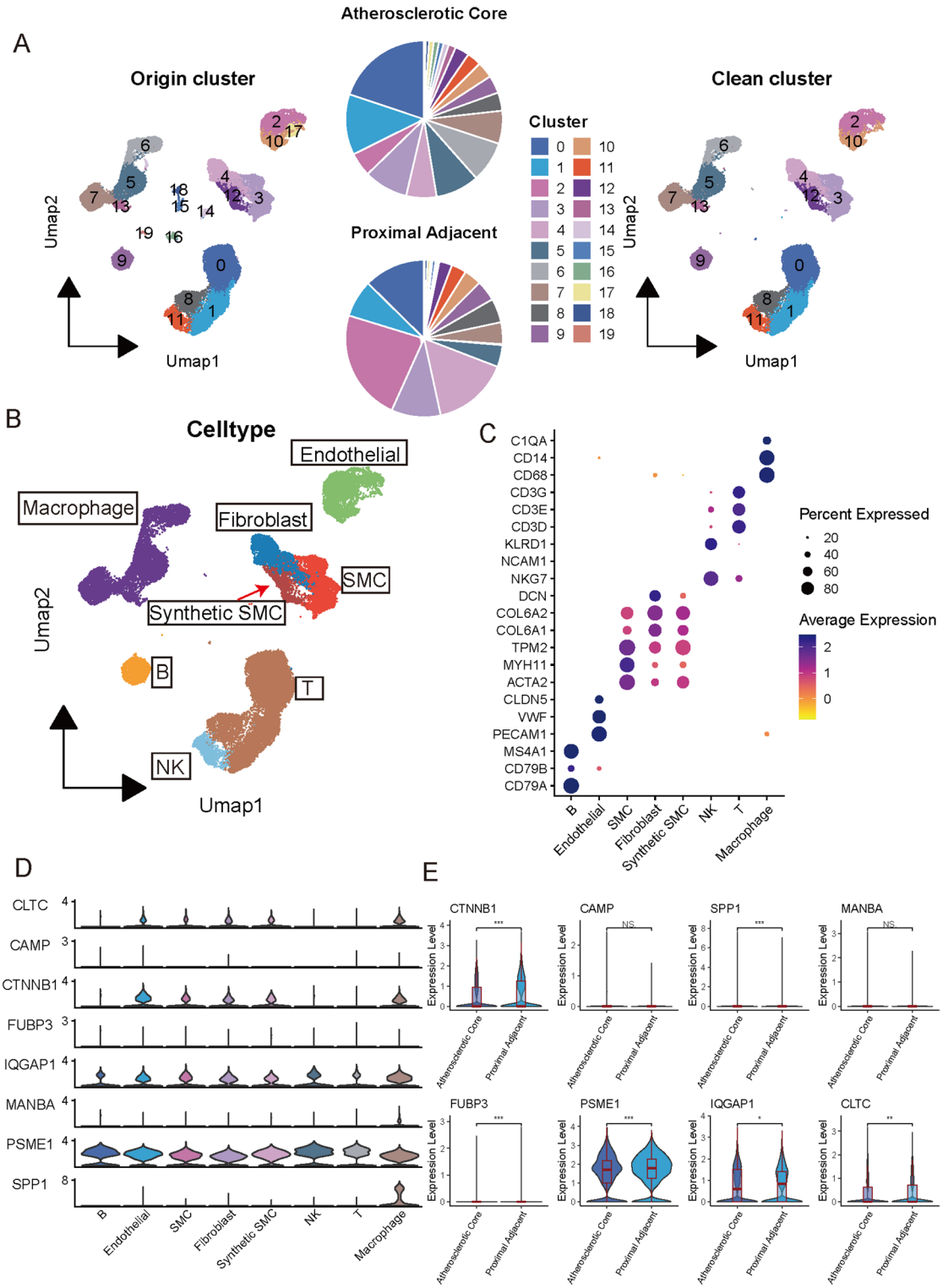


Fig. 5. Single-cell analysis of atherosclerotic plaques. (**A**) UMAP visualization of clusters in the atherosclerotic core and adjacent regions: UMAP plots show distinct cellular clusters in the atherosclerotic core and proximal adjacent regions. Pie charts depict the proportion of clusters in each region, highlighting compositional differences. Clean clustering reveals well-separated groups after data refinement. (**B**) Cell type annotation in the atherosclerotic plaque. UMAP visualization identifies major cell types, including macrophages, fibroblasts, endothelial cells, synthetic smooth muscle cells (SMCs), and natural killer (NK) cells. (**C**) Dot plot of marker gene expression across cell types. The size of dots represents the percentage of cells expressing the marker, while color intensity reflects average expression. (**D**) Violin plots of key gene expression across cell types. (**E**) Violin plots of key gene expression in the atherosclerotic core and proximal adjacent regions. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.001$, ns indicating non-significance.

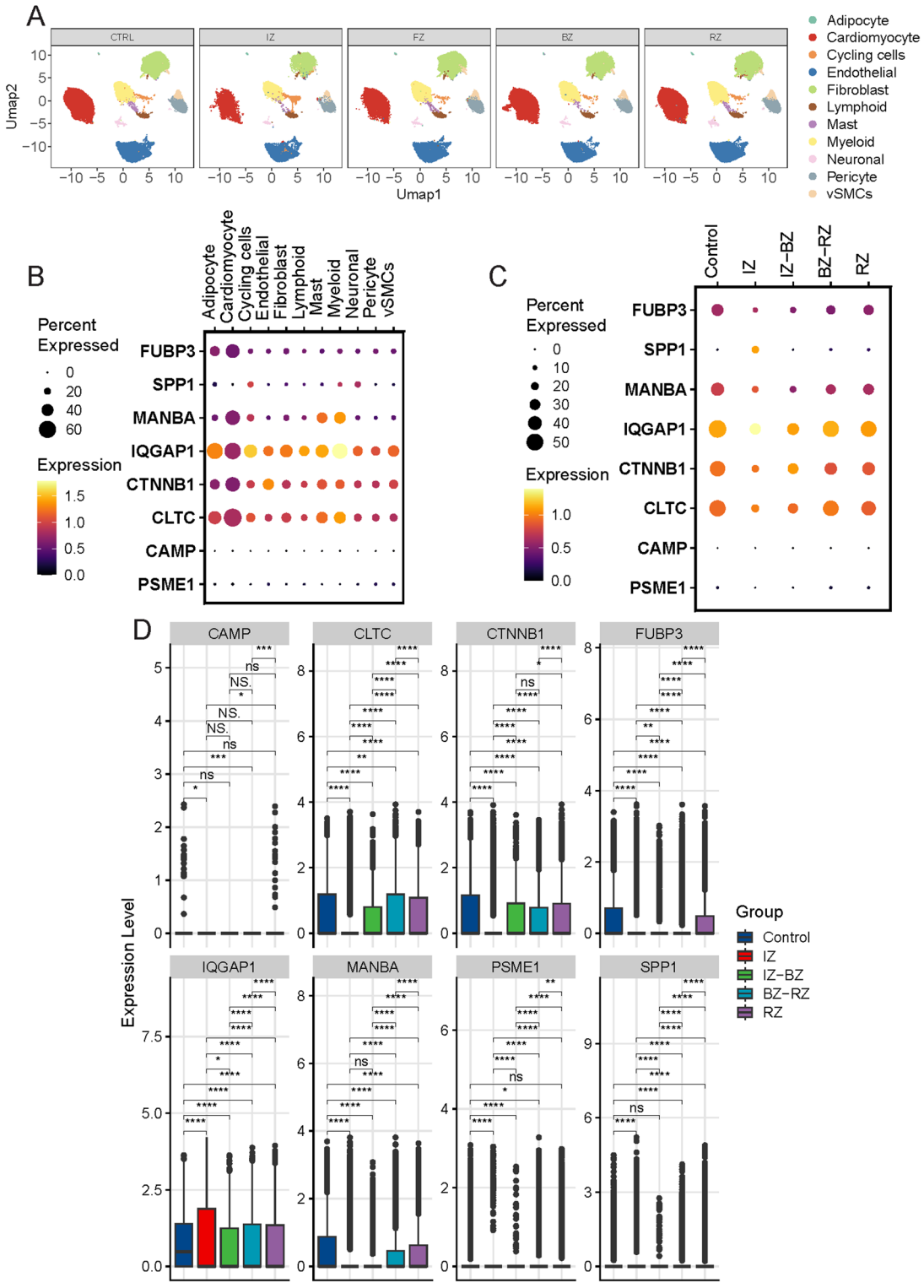


Fig. 6. Single-cell gene expression analysis across zones of MI. (**A**) UMAP visualization of cell type distribution in different zones. UMAP plots depict the distribution of cell types across CTRL, IZ, BZ, and RZ. (**B**) Dot plot of key gene expression across cell types. Dot size represents the percentage of cells expressing the gene, and color indicates average expression levels. (**C**) Dot plot of key gene expression across regions. (**D**) Boxplots of gene expression levels across regions. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns indicating non-significance. *MI* myocardial infarction, *CTRL* control, *IZ* ischemic zone, *BZ* border zone, *RZ* remote zone.

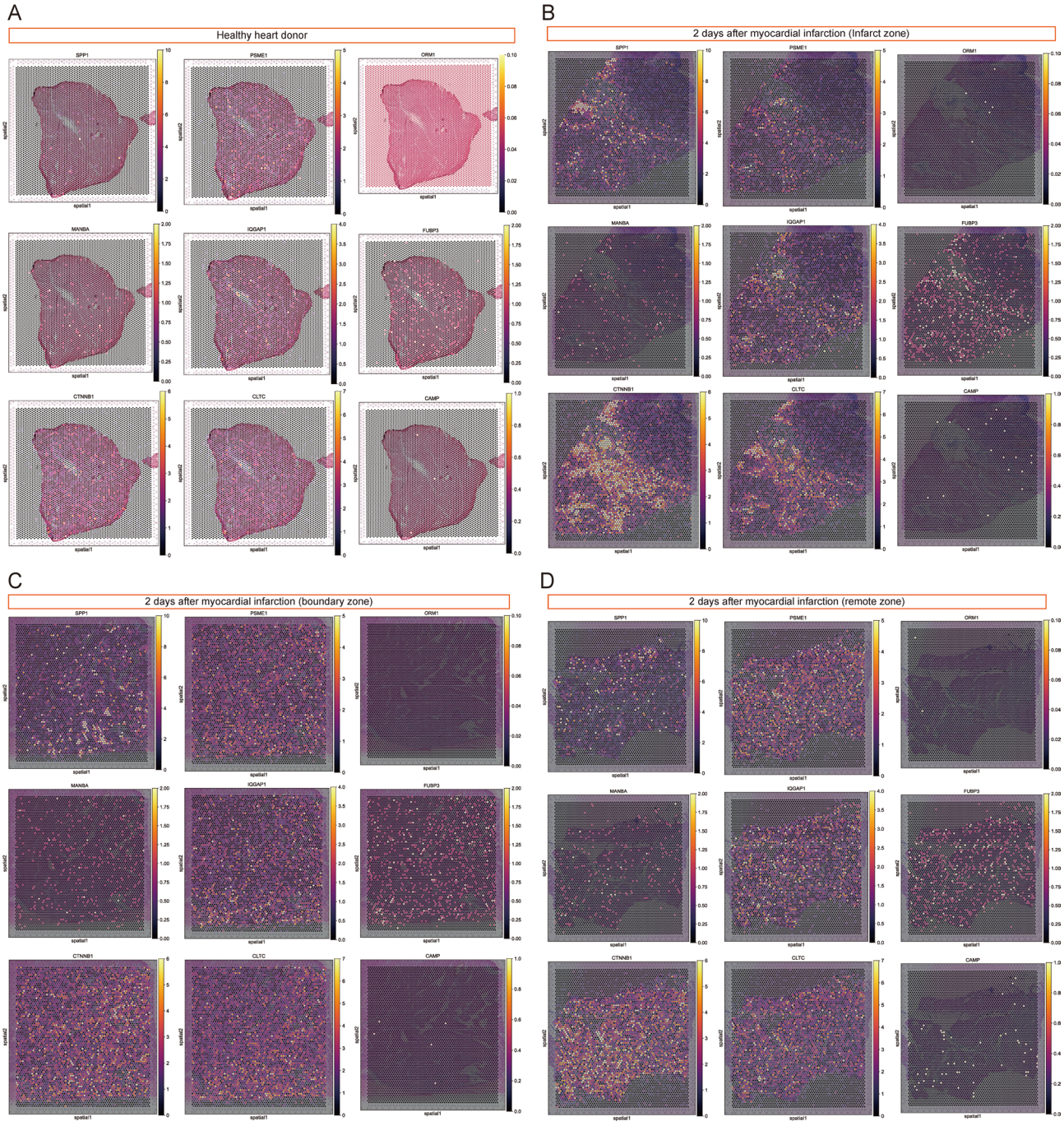


Fig. 7. Spatial expression of key proteins (SPP1, PSME1, ORM1, MANBA, IQGAP1, FUBP3, CTNNB1, CLTC, and CAMP). (**A**) Spatial expression of key genes in a healthy heart donor. (**B**) Spatial expression of key genes in 2 days after MI (IZ). (**C**) Spatial expression of key genes in 2 days after MI (BZ). (**D**) Spatial expression of key genes in 2 days after MI (RZ). *MI* myocardial infarction, *RZ* remote zone, *IZ* infarct zone, *BZ* boundary zone.

necessitating further studies with larger cohorts, longitudinal designs, and functional experiments to validate these findings and explore their broader relevance across cardiovascular conditions.

Conclusion

This study identifies nine core proteins involved in inflammation, immune regulation, and tissue remodeling in AMI, highlighting heterogeneity through two distinct subtypes. Proteins like SPP1 drive macrophage-mediated inflammation, while PSME1 may offer protective effects. Notably, CAMP, ORM1, and FUBP3 emerge as potential acute-phase-specific proteins for AMI. These findings enhance our understanding of AMI pathophysiology and suggest potential biomarkers and therapeutic targets.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (https://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD062794.

Received: 12 December 2024; Accepted: 27 May 2025
Published online: 04 June 2025

References

1. Tsao, C. W. et al. Heart disease and stroke Statistics-2023 update: A report from the American heart association. *Circulation* **147** (8), e93–e621 (2023).
2. Roth, G. A., Mensah, G. A. & Fuster, V. The global burden of cardiovascular diseases and risks: A compass for global action. *J. Am. Coll. Cardiol.* **76** (25), 2980–2981 (2020).
3. Shi, H. et al. Astragaloside IV prevents acute myocardial infarction by inhibiting the TLR4/MyD88/NF-κB signaling pathway. *J. Food Biochem.* **45** (7), e13757 (2021).
4. Del Buono, M. G. et al. Ischemic cardiomyopathy and heart failure after acute myocardial infarction. *Curr. Cardiol. Rep.* **24** (10), 1505–1515 (2022).
5. Heusch, G. Myocardial ischemia/reperfusion: translational pathophysiology of ischemic heart disease. *Med* **5** (1), 10–31 (2024).
6. Zhang, K. et al. Clinical characteristics and biomarkers feature analysis using a proteomics platform in young patients with acute coronary syndrome. *Front. Cardiovasc. Med.* **11**, 1384546 (2024).
7. Shin, M. et al. *Biomarker discovery of acute coronary syndrome using proteomic approach*. *Molecules* **26**(4). (2021).
8. Reel, P. S. et al. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* **49**, 107739 (2021).
9. Pudjihartono, N. et al. A review of feature selection methods for machine Learning-Based disease risk prediction. *Front. Bioinform.* **2**, 927312 (2022).
10. Wang, Y. et al. A hybrid feature selection algorithm and its application in bioinformatics. *PeerJ Comput. Sci.* **8**, e933 (2022).
11. N, A. S. et al. Enhancing machine learning-based sentiment analysis through feature extraction techniques. *PLoS One.* **19** (2), e0294968 (2024).
12. Wang, Z. et al. Information-Theory-based nondominated sorting ant colony optimization for multiobjective feature selection in classification. *IEEE Trans. Cybern.* **53** (8), 5276–5289 (2023).
13. Gao, J. et al. Information gain ratio-based subfeature grouping empowers particle swarm optimization for feature selection. **286**, 111380 (2024).
14. Ge, S. et al. A proteomic landscape of diffuse-type gastric cancer. *Nat. Commun.* **9** (1), 1012 (2018).
15. Yu, G. et al. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16** (5), 284–287 (2012).
16. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44** (W1), W90–W97 (2016).
17. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102** (43), 15545–15550 (2005).
18. Heber, S. & Sick, B. Quality assessment of affymetrix genechip data. *Omics* **10** (3), 358–368 (2006).
19. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** (7), e47 (2015).
20. Korthauer, K. et al. A practical guide to methods controlling false discoveries in computational biology. *Genome Biol.* **20** (1), 118 (2019).
21. Ramasamy, A. et al. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* **5** (9), e184 (2008).
22. Mathew, J. L. Systematic reviews and Meta-Analysis: A guide for beginners. *Indian Pediatr.* **59** (4), 320–330 (2022).
23. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
24. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13** (11), 2498–2504 (2003).
25. Butler, A. et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36** (5), 411–420 (2018).
26. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20** (1), 296 (2019).
27. Germain, P. L. et al. Doublet identification in single-cell sequencing data using ScDblFinder. *F1000Res* **10**, p979 (2021).
28. Yang, S. et al. Decontamination of ambient RNA in single-cell RNA-seq with decontx. *Genome Biol.* **21** (1), 57 (2020).
29. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods.* **16** (12), 1289–1296 (2019).
30. Büttner, M. et al. A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods.* **16** (1), 43–49 (2019).
31. Kuppe, C. et al. Spatial multi-omic map of human myocardial infarction. *Nature* **608** (7924), 766–777 (2022).
32. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19** (1), 15 (2018).
33. Valdes-Marquez, E. et al. Proteomic profiling identifies novel independent relationships between inflammatory proteins and myocardial infarction. *Eur. J. Prev. Cardiol.* **30** (7), 583–591 (2023).
34. Wei, Q. et al. Plasma proteomics implicate glutamic oxaloacetic transaminases as potential markers for acute myocardial infarction. *J. Proteom.* **308**, 105286 (2024).
35. Xie, Y., Zhang, H. & Huang, T. Quantitative proteomics reveal three potential biomarkers for risk assessment of acute myocardial infarction. *Bioengineered* **13** (3), 4939–4950 (2022).
36. Dong, Y. et al. *Recent insights into neutrophil extracellular traps in cardiovascular diseases. J. Clin. Med.* **11**(22). (2022).
37. Ibrahim, N. et al. Neutrophil extracellular traps in cardiovascular and aortic disease: A narrative review on molecular mechanisms and therapeutic targeting. *Int. J. Mol. Sci.* **25**(7). (2024).
38. Wu, X. et al. Ferroptosis and lipid metabolism in acute myocardial infarction. *Rev. Cardiovasc. Med.* **25** (5), 149 (2024).
39. Shen, T. et al. Dioscin alleviates cardiac dysfunction in acute myocardial infarction via rescuing mitochondrial malfunction. *Front. Cardiovasc. Med.* **9**, 783426 (2022).
40. Xu, C. et al. Mitophagy-regulated mitochondrial health strongly protects the heart against cardiac dysfunction after acute myocardial infarction. *J. Cell. Mol. Med.* **26** (4), 1315–1326 (2022).
41. Hilgendorf, I., Frantz, S. & Frangogiannis, N. G. Repair of the infarcted heart: cellular effectors, molecular mechanisms and therapeutic opportunities. *Circ. Res.* **134** (12), 1718–1751 (2024).
42. Li, T. et al. Cardiac repair after myocardial infarction: A two-sided role of inflammation-mediated. *Front. Cardiovasc. Med.* **9**, 1077290 (2022).
43. Ma, Y. *Role of neutrophils in cardiac injury and repair following myocardial infarction*. *Cells*, **10**(7). (2021).
44. Xu, J. et al. Nonmuscle myosin light-chain kinase mediates neutrophil transmigration in sepsis-induced lung inflammation by activating beta2 integrins. *Nat. Immunol.* **9** (8), 880–886 (2008).
45. Shah, R. V. et al. Dietary metabolic signatures and cardiometabolic risk. *Eur. Heart J.* **44** (7), 557–569 (2023).
46. Welsh, P. et al. Cardiac troponin T and troponin I in the general population. *Circulation* **139** (24), 2754–2764 (2019).

47. Ridker, P. M. From C-Reactive protein to Interleukin-6 to Interleukin-1: moving upstream to identify novel targets for atheroprotection. *Circ. Res.* **118** (1), 145–156 (2016).
48. Chen, L. et al. Systematic identification of therapeutic targets for coronary artery calcification: an integrated transcriptomic and proteomic Mendelian randomization. *Front. Cardiovasc. Med.* **11**, 1419440 (2024).
49. Choy, M. et al. Association between epicardial adipose tissue and incident heart failure mediating by alteration of natriuretic peptide and myocardial strain. *BMC Med.* **21** (1), 117 (2023).
50. Wan, J. J. et al. Role of acute-phase protein ORM in a mice model of ischemic stroke. *J. Cell. Physiol.* **234** (11), 20533–20545 (2019).
51. Royle, S. J., Bright, N. A. & Lagnado, L. Clathrin is required for the function of the mitotic spindle. *Nature* **434** (7037), 1152–1157 (2005).
52. An, C. et al. IQGAP1 promotes mitochondrial damage and activation of the MtDNA sensor cGAS-STING pathway to induce endothelial cell pyroptosis leading to atherosclerosis. *Int. Immunopharmacol.* **123**, 110795 (2023).
53. Huang, L. et al. Beta-catenin promotes macrophage-mediated acute inflammatory response after myocardial infarction. *Immunol. Cell. Biol.* **96** (1), 100–113 (2018).
54. Oliveira-Paula, G. H. et al. The β-catenin C terminus links Wnt and sphingosine-1-phosphate signaling pathways to promote vascular remodeling and atherosclerosis. *Sci. Adv.* **10** (11), eadg9278 (2024).
55. Qi, Z. et al. Novel treatment from a botanical formulation Si-Miao-Yong-an Decoction inhibits Vasa vasorum angiogenesis and stabilizes atherosclerosis plaques via the Wnt1/β-catenin signalling pathway. *Pharm. Biol.* **61** (1), 1364–1373 (2023).
56. Li, Y. et al. Autophagy-Sirtuin1(SIRT1) alleviated the coronary atherosclerosis (AS)in mice through regulating the proliferation and migration of endothelial progenitor cells (EPCs) via wnt/β-catenin/GSK3β signaling pathway. *J. Nutr. Health Aging.* **26** (3), 297–306 (2022).
57. Gao, J. H. et al. CXCL12 promotes atherosclerosis by downregulating ABCA1 expression via the CXCR4/GSK3β/β-catenin(T120)/TCF21 pathway. *J. Lipid Res.* **60** (12), 2020–2033 (2019).
58. Liu, J. et al. Lcz696 alleviates myocardial fibrosis after myocardial infarction through the sFRP-1/Wnt/β-Catenin signaling pathway. *Front. Pharmacol.* **12**, 724147 (2021).
59. Zhang, Q. et al. Signaling pathways and targeted therapy for myocardial infarction. *Signal. Transduct. Target. Ther.* **7** (1), 78 (2022).
60. Wang, Q. et al. The prognostic value of the proteasome activator subunit gene family in skin cutaneous melanoma. *J. Cancer.* **10** (10), 2205–2219 (2019).
61. Zhang, F. et al. Carfilzomib alleviated osteoporosis by targeting PSME1/2 to activate Wnt/β-catenin signaling. *Mol. Cell. Endocrinol.* **540**, 111520 (2022).

Author contributions

The study was conceived and designed by X.L., and Y.Y.; C.L., X.Z., Q.X. and B.F. carried out the bioinformatic analyses; Data interpretation was performed by X.L., and Y.Y.; C.L., X.Z., Q.X., B.F. and X.L. were responsible for drafting the manuscript, while F.L., J.L., G.A., and W.J. conducted the literature search; Y.Y. and X.L. contributed to the manuscript revision. All authors have read and approved the final version of the manuscript.

Funding

This research was funded by the Key Research and Development Task Special in Xinjiang Uygur Autonomous Region (2022B03022-2), the Special Fund Project for Central Guidance of Local Science and Technology Development (ZYYD2022C21), "Tianshan Talents" High-Level Talent Training Program in Medicine and Healthcare (2023TSYCLJ0035, 2022TSYCCX0033, 2022TSYCLJ0028), National Natural Science Foundation of China (8216020109).

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

All procedures involving human blood samples were conducted in accordance with the ethical standards of the Institutional Review Board of Xingjiang Medical College, which granted approval under permit number 20220308-105.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-04401-w.

Correspondence and requests for materials should be addressed to Y.Y. or X.L.