scientific reports

Check for updates

OPEN

Identification of biomarkers associated with M1 macrophages in the ST-segment elevation myocardial infarction through bioinformatics and machine learning approaches

Huiying Li[1,2,4], Qiwei Zhu[1,4], Wei Wang[3,4], Yu Bao[1], Yongyi Bai[1]✉, Hongbin Liu[1]✉ & Wenxiu Leng[1]✉

ST-segment elevation myocardial infarction (STEMI) is considered a critical cardiac condition with a poor prognosis. Shortly after STEMI occurs, the increased number of circulating leukocytes including macrophages can lead to the accumulation of more cells in the myocardium, affecting the cardiac immune microenvironment. Identifying serum biomarkers associated with immune infiltration after STEMI is important for diagnosing and treating STEMI. In this work, we aimed to use integrated bioinformatics and machine learning methods to identify new biomarkers. First, candidate genes closely associated with M1 macrophage immune infiltration and STEMI were obtained using the *limma* package, the *CIBERSORTx* package, weighted gene coexpression network analysis (WGCNA), and protein–protein interaction (PPI) networks from the GSE59867 dataset, which comprises peripheral blood mononuclear cell (PBMC) samples. The STEMI patients were subsequently stratified into subtypes using the *ConsensusClusterPlus* package. Furthermore, using machine learning methods, we identified AKT3, GJC2, HMGCL and RBM17 as the genes with the greatest potential to be associated with STEMI subtypes and with M1 macrophage infiltration during the acute phase of STEMI. Finally, the expression profile and diagnostic value of the four feature genes were validated in the GSE59867 and GSE62646 datasets and in 24 patients using real-time PCR. This study revealed logically and comprehensively that AKT3, GJC2, HMGCL and RBM17, which are derived from PBMCs, could enhance the accuracy of STEMI diagnosis and might provide effective treatment options for STEMI patients.

Myocardial infarction (MI) is a severe cardiovascular disorder triggered by inadequate blood perfusion to cardiac tissue due to coronary artery obstruction. According to data from the National Health Commission, there are over one million new cases of MI each year, with a mortality rate where one out of every three heart attack patients tragically passes away. In addition, MI can lead to several serious complications, such as arrhythmias, heart failure, myocardial rupture, and thrombus formation[1,2]. Regrettably, the adverse prognosis of MI is partly attributed to the complexity of its pathogenesis and the current lack of specific therapies for MI subtypes[3]. MI is classified into ST-segment elevation myocardial infarction (STEMI) and non-ST-segment elevation myocardial infarction (NSTEMI) based on the characteristics of the electrocardiogram (ECG). STEMI is the most severe type of MI and is typically associated with complete coronary artery occlusion, requiring urgent intervention, affecting a broader area of myocardial tissue, and carrying a greater risk of arrhythmias[4]. For many years, the elevation of biomarkers such as serum troponin I, troponin T, and myoglobin has been the basis for diagnosing MI. However, these markers cannot differentiate clinical subtypes of MI or guide pathological treatment. The

[1]Department of Cardiology, The Second Medical Center and National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital, 28 Fuxing Road, Haidian, Beijing 100853, China. [2]Medical School of Chinese PLA, 28 Fuxing Road, Haidian, Beijing 100853, China. [3]Department of Cardiology, The Sixth Medical Center of Chinese PLA General Hospital, 6 Fucheng Road, Haidian, Beijing 100037, China. [4]These authors contributed equally: Huiying Li, Qiwei Zhu and Wei Wang. ✉email: baiyongyi301@sina.com; liuhbcad301@163.com; kevin301doc@126.com

current study provides profound insights into the significant challenges that MI poses to patients' long-term health and highlights the urgency of delving into the pathogenesis of STEMI and identifying pathology-related biomarkers for STEMI subtypes to improve prognosis.

Research has shown that immune cells, especially macrophages, play crucial roles in regulating the inflammatory response and immune environment during the onset of STEMI[5]. Shortly after ischaemia occurs, cardiomyocytes and resident macrophages produce inflammatory cytokines and chemokines such as IL-1, IL-6 and TNF, which mediate changes in the cardiac immune microenvironment. These events, in turn, induce the large-scale production and recruitment of neutrophils and monocytes, mainly from haematopoietic stem and progenitor cells in the bone marrow. An increased number of circulating leukocytes, including macrophages, can lead to increased accumulation of cells in the myocardium that actively participate in the inflammatory cascade and is the mechanism of monocytic immune infiltration during STEMI[6]. Macrophages exhibit M1 and M2 phenotypes. The M1 phenotype, in which necrotic cells and matrix debris are phagocytosed, drives the initial cellular response and is an important source of inflammatory mediators in the early stages of injury. M2 macrophages, however, promote cardiac recovery. Though new subtypes of M1 and M2 macrophages such as M2a, M2b, M2c, M2d and M4 have been identified, studying and intervention with polarized M1 and M2 macrophages in the myocardium has shown unique advantages[7,8]. Especially, M1 macrophages have been able to release high levels of proinflammatory cytokines and have exploited high level of plasticity in response to microenvironmental stimuli. Identifying M1 macrophage-associated biomarkers could be important for diagnosis, cytokine regulation, stem cell therapy and immune cell therapy for MI.

In this study, for the first time, we employed differentially expressed genes (DEGs) related to immune infiltration between the normal and STEMI groups to cluster STEMI patients and identify M1 macrophage-associated molecular features. We anticipate that this study will contribute substantively to the theoretical underpinnings of diagnosing and treating patients with acute myocardial infarction, reflecting the rigor of our medical research.

Methods

Dataset collection

To identify microarray datasets pertaining to STEMI, we performed a search in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) using the keyword "myocardial infarction". The dataset selection criteria were as follows: the subjects were limited to "*Homo sapiens*"; the datasets included peripheral blood mononuclear samples from both normal individuals and STEMI patients; and peripheral blood samples from normal individuals were collected from coronary artery disease (CAD) patients who had never experienced MI. Finally, two publicly available independent datasets (GSE59867 and GSE62646, both from platform GPL6244) were utilized to predict and validate the selected molecular features. The GSE59867 dataset encompasses 111 STEMI patients and 46 CAD patients. The GSE62646 dataset includes 28 STEMI patients and 14 CAD patients. Owing to the nature of the GSE59867 dataset as a large cohort study, it is considered the training set, while GSE62646 is designated the validation set.

Preprocessing and differential gene expression analysis

The GSE59867 and GSE62646 datasets were accessed using the *GEOquery* package[9]. When multiple probes corresponded to a single gene, the average gene expression level was used as the final expression profile. The raw counts were subsequently converted into the $\log_2(X+1)$ format, which was implemented for standardization via the normalizeBetweenArrays function in the *Limma* package[10]. We utilized box plots to assess normalization and used a PCA plot for clustering between sample subgroups. Box plots were generated using the *ggplot* package, whereas PCA plots were generated using the *FactoMindR* and *factoextra* packages[11,12]. Using the *hclust* function and hierarchical clustering method, we formed a dendrogram. The static tree cut method defines each contiguous branch below a fixed-height cutoff a cluster and gives detection of outliers[13]. In this way, we removed the outliers above the fixed height. Differential gene expression analysis between the control and MI groups was conducted using the *Limma* package[10], with significance determined based on the criteria of FDR < 0.05 and $|\log_2\text{fold change}| > 0.485$ (fold change > 1.4). Using the *Pheatmap* package, we created a heatmap to visualize the distribution of DEGs between the control and MI groups.

Immune infiltration analysis

CIBERSORTx is a support vector regression algorithm that relies on signature datasets to assess the relative abundance of immune cell populations in individual samples[14]. The LM22 dataset is a human-specific feature dataset encompassing the relative infiltration proportions of 22 distinct immune cell types. Using the *CIBERSORT* package and setting the LM22 dataset as a reference dataset, we assessed the relative proportions of immune cell infiltration in both the STEMI and control samples in the GSE59867 dataset. Next, we compared the infiltration proportions of the 22 distinct immune cell types in the STEMI and control groups. Immune cell types with significant infiltration differences between the STEMI and control groups were identified for subsequent WGCNA.

Weighted gene coexpression network analysis (WGCNA) and protein–protein interaction (PPI) network construction

Weighted gene coexpression network analysis (WGCNA) is a bioinformatics method used to construct gene coexpression networks, identify gene modules, and elucidate the coordinated regulation of genes in biological processes, aiding in comprehending complex biological phenomena and disease mechanisms. By utilizing different infiltrated immune cell types and the STEMI/control group as the phenotypes, we constructed a WGCNA network to identify the gene modules closely associated with immune cell infiltration and STEMI. The

WGCNA package was used to construct a gene expression similarity matrix that was subsequently transformed into an adjacency matrix by selecting the optimal soft threshold β to construct a scale-free network[15]. The minimum number of genes per gene module was set to 100. The module genes obtained using WGCNA intersected with the DEGs. Intersecting genes were used to establish a protein–protein interaction (PPI) network. The PPI network was subsequently constructed using the String database (https://www.string-db.org) with an interaction score threshold greater than 0.4. Cytoscape software (version 3.1.1.0) was used to visualize the interaction relationships between genes.

Consensus clustering analysis

The PPI hub genes were subjected to consensus clustering analysis using the *ConsensusClusterPlus* package[16] to delineate STEMI patients into two distinct molecular subtypes. Clustering was performed for 1000 iterations to ensure reliable and stable results. The optimal clustering K values were selected within the range of 2 to 9 for classification. Subsequently, heatmaps were generated to visualize the expression profiles of these genes among the different subtypes. This facilitated the identification of DEGs between the M1 and M2 subtypes. The DEGs between subtypes of STEMI were subsequently intersected with the PPI hub genes, thus identifying the feature genes that are correlated with STEMI subtypes and with M1 macrophage infiltration.

Machine learning and validation in datasets

Random forest, known for its ability to predict stable factors through an ensemble of decision trees, thus enhancing prediction accuracy, was implemented using the 'randomForest' package in R. Least absolute shrinkage and selection operator regression (LASSO), a regression algorithm acclaimed for its variable selection efficiency by minimizing classification error probability, was conducted using the 'glmnet' package in R[17]. The innovation of eXtreme gradient boosting (XGBoost) lies in its incorporation of regularization techniques to prevent overfitting and parallelization strategies to expedite training, achieving state-of-the-art performance in predictive accuracy, scalability, and efficiency[18,19]. XGBoost was performed using the "xgboost" package in R. In this part, feature genes were first obtained by taking the intersection of DEGs between STEMI subtypes and PPI hub genes. Then, the feature genes were subjected to LASSO, random forest and XGBoost models for the identification of the most important feature genes. The overlapping genes filtered by the three machine learning methods were considered the genes with the greatest potential for distinguishing the molecular subtypes. Boxplots of these potential gene expression levels in GSE59867 were subsequently generated using the packages "ggpubr"[20] and "ggplot2"[21]. The gene expression levels were compared among the control group, the STEMI group and the STEMI subtypes. Receiver operating characteristic (ROC) curves were constructed to determine the diagnostic value of the potential genes in predicting the STEMI and STEMI subtypes in both the GSE59867 and GSE62646 datasets.

Validation by real-time PCR

Finally, the potential genes were validated for their expression profiles in peripheral blood mononuclear cells (PBMCs) using real-time PCR. From June 8th to June 29th, 2024, 16 consecutive patients admitted to the PLA general hospital and diagnosed with STEMI were enrolled in our study. The control group included 8 patients with CAD but without STEMI during the same period. We obtained all participants' peripheral plasma samples on admission. For both the STEMI group and the control group, PBMCs were isolated within 1 h of collection using human peripheral blood leukocyte isolation kit (Solarbio Life Sciences, Beijing, China) according to the manufacturer's instructions. The PBMC samples were stored at − 80 ℃ until further analysis. RNA extraction was performed using TRIzol reagent (Invitrogen, CA, USA). The concentration and purity of the RNA samples were checked using an Agilent 2100 Bioanalyzer© and an RNA 6000 Nano Kit (Agilent, Santa Clara, CA, USA). The total RNA was reverse transcribed into cDNA using a PrimeScript™ RT reagent Kit (Takara, Beijing, China) with gDNA Eraser. The primer sequences were synthesized by Sangon Biotech (Shanghai, China), and the primer sequences for each gene are provided in Supplementary Table 2. On a Bio-Rad Real-Time PCR Detection System (T100; Hercules, CA, USA), quantitative real-time PCR was conducted using UltraSYBR Mixture (CWBio, Jiangsu, China) and specific primers (Sangon Biotech, Shanghai, China). The study was approved by the local Ethics Committees of the PLA General Hospital, and was conducted in accordance with the principles of the Declaration of Helsinki (S2021-221-01). All the participants provided written informed consent.

Statistical analysis

Data processing, statistical analyses, and graph plotting were performed using R software, version 4.1.3. The K-S normality tests were conducted to determine whether a certain gene followed a normal distribution. If continuous variables were normally distributed, they were presented as the mean and standard deviation (SD). If continuous variables were not normally distributed, they are presented as medians and interquartile ranges. Student's *t* tests were used to compare two normally distributed variables, whereas the Mann–Whitney *U* test was used to compare two abnormally distributed variables. Correlation analysis was assessed using Spearman's correlation. The parameters for R packages or network platform tools are described in the corresponding sections, with default settings applied unless otherwise specified. For categorical variables, group comparisons were conducted using Fisher's exact test. All tests were conducted bidirectionally with a significance threshold of $P < 0.05$. Significance was defined as a P value < 0.05.

Results

Inhomogeneity of DEGs in STEMI samples

We confirmed the successful standardization of the GSE59867 dataset using the normalizeBetweenArrays function, as evidenced by PCA plots (Fig. 1A) and boxplots (Fig. 1B). We then used the *hclust* package and hierarchical clustering method to form a dendrogram (Supplementary Fig. 1). We set the fixed-height cutoff

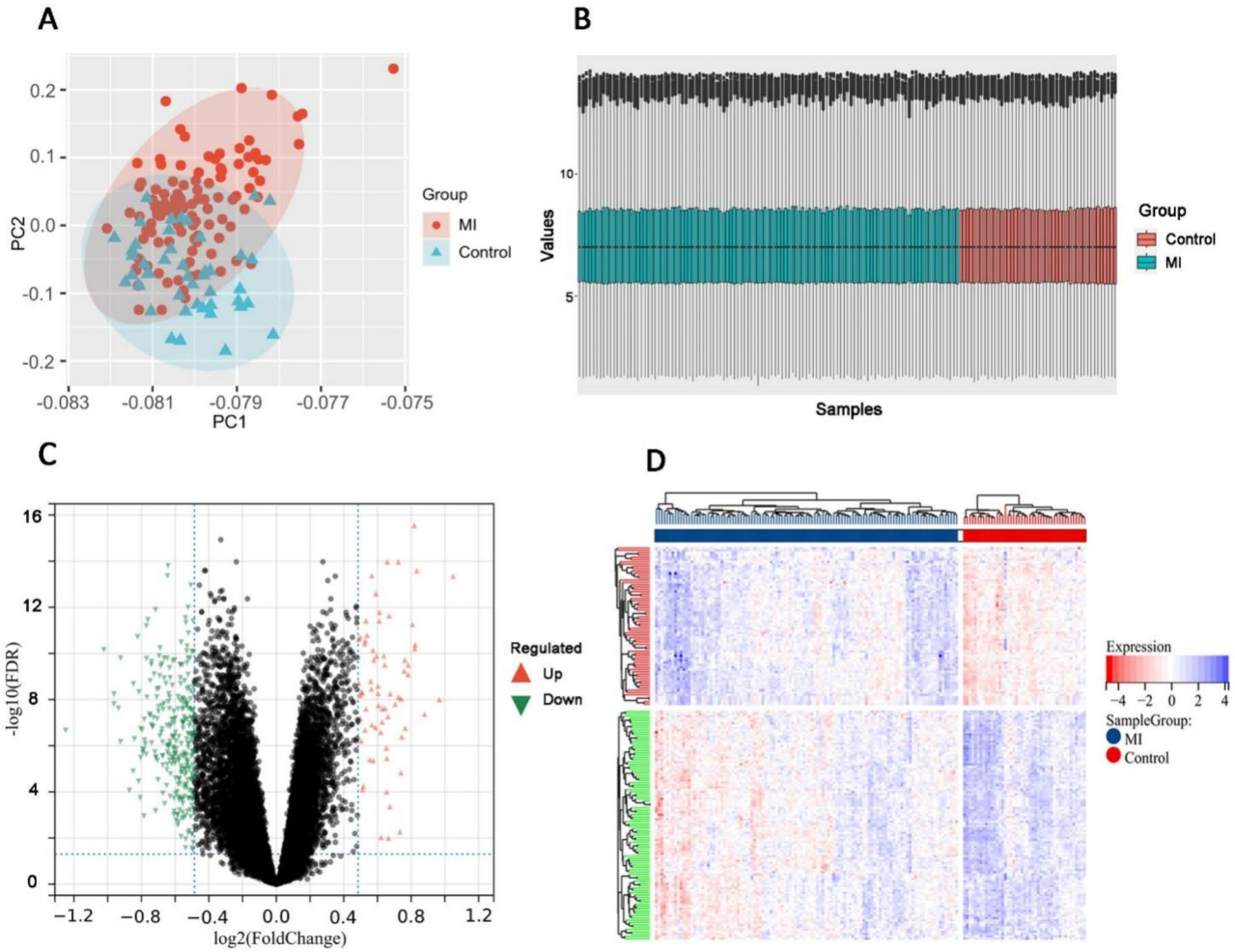

Fig. 1. PCA plot (**A**) and boxplots (**B**) of the GSE59867 dataset after standardization. Volcano plot (**C**) and heatmap (**D**) of DEGs between the control and STEMI samples in the GSE59867 dataset.

as 0.025. In this way, GSM1620790 which above the fixed height was removed as an outlier to ensure the robustness of the subsequent filtration results. Differential gene expression analysis was performed to identify genes implicated in STEMI. The findings revealed 362 DEGs between the control and MI groups in GSE59867, comprising 293 downregulated and 69 upregulated genes (Fig. 1C). The heatmap revealed that these DEGs exhibited a clear classification effect between the control and MI groups. However, it also revealed a certain degree of heterogeneity within the STEMI samples (Fig. 1D).

M1 macrophages are positively correlated with STEMI occurrence

To determine the associations between STEMI and immune cells, we evaluated the proportions of infiltrating immune cells in both the control and STEMI samples using the CIBERSORTx algorithm. Compared with the control group, six immune cell types were upregulated, including M1 macrophage, M2 macrophage, CD4 memory T cells, mast cells, CD8 + T cells, and activated memory T cells ($P < 0.05$, Fig. 2A,B). Furthermore, these immune cell types exhibited significant intercorrelations (Fig. 2C). To further investigate the specific gene types associated with STEMI patients, we conducted a WGCNA. Selecting an appropriate soft threshold before constructing a scale-free gene expression network is essential, as it influences the stability of network construction. The first soft threshold of 9, with an R^2 value exceeding 0.85, was selected for conducting the WGCNA (Fig. 3A). As the soft threshold increased, the mean connectivity decreased (Fig. 5B), whereas the frequency of connectivity increased rapidly (Fig. 5C). The logarithm of the mean connectivity was linearly correlated with the logarithm of the connectivity frequency (Fig. 5D). In the module-feature relationship network, both the blue module (with a correlation coefficient of −0.32 and a $P < 0.001$) and the brown module (with a correlation coefficient of 0.4 and a $P < 0.001$) were significantly correlated with M1 macrophages (Fig. 3E). The genes from the brown and blue modules intersected with 362 DEGs between the STEMI patients and the control patients, and we obtained a total of 295 overlapping genes (Fig. 3F). Additionally, the constructed PPI network retained 82 genes for subsequent consensus clustering analysis (Fig. 3G).

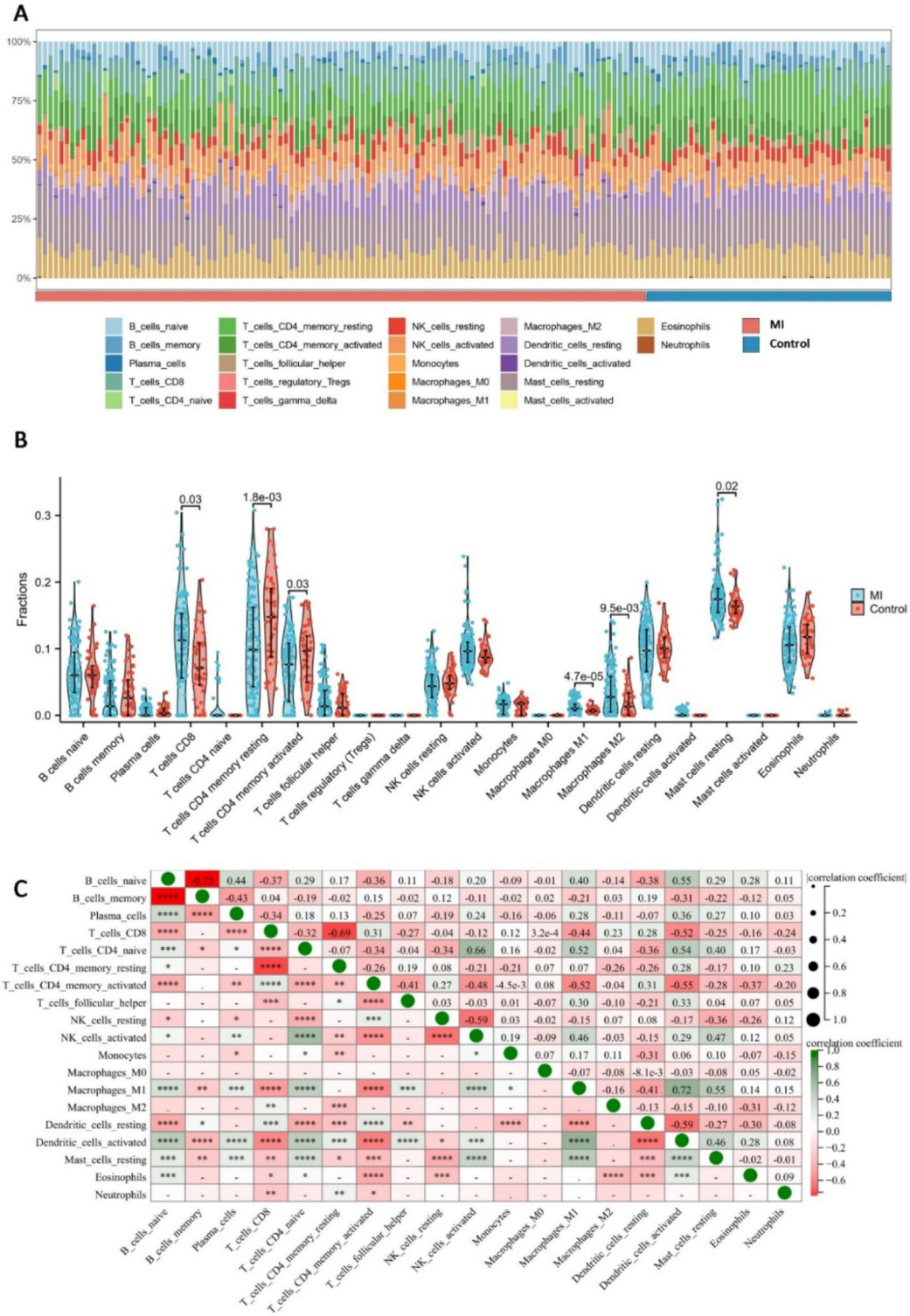

Fig. 2. (**A**) The distribution of immune cell infiltration proportions in both the control and STEMI samples. (**B**) Statistical differences in immune cell profiles between the control and STEMI groups. (**C**) Correlation heatmap of immune cell interactions.

DEGs related to M1 macrophages can effectively distinguish STEMI patients

We conducted consensus clustering analysis to ascertain whether M1 macrophages-related DEGs can classify STEMI patients. The delta area plot demonstrated the highest rate of decrease in K values between 2 and 3 (Fig. 4A). When K was set to 2, the consensus CDF plot displayed a stable variation in CDF values within the range of 0.2 to 0.8 (Fig. 4B); the consensus matrix heatmap revealed a robust clustering effect for two distinct

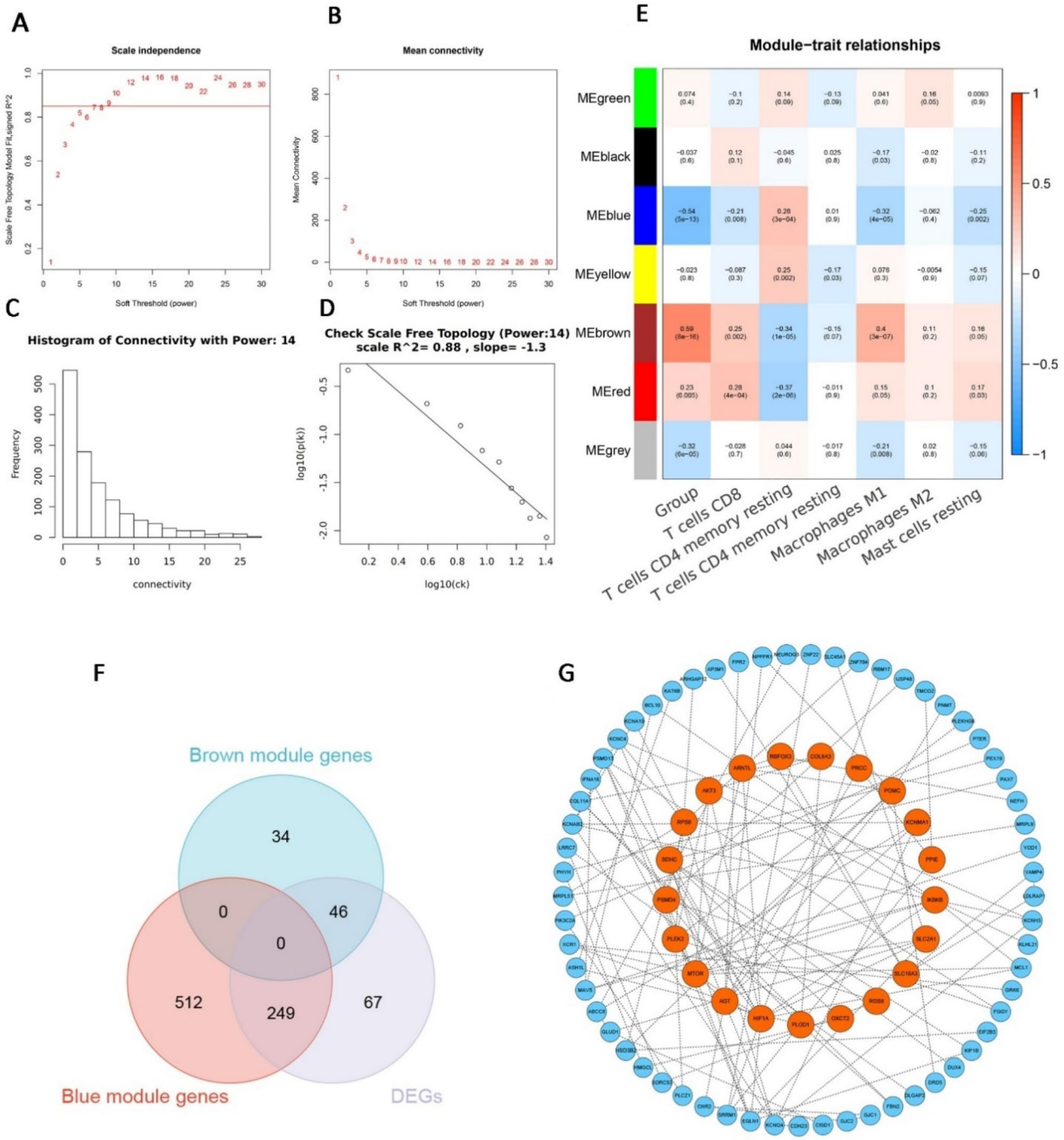

Fig. 3. (**A**) Relationship between the soft threshold and R square value. (**B**) Relationships between the soft threshold and mean connectivity. (**C**) Frequency distribution diagram of the connectivity of gene nodes. (**D**) Linear correlation between the logarithm of the mean connectivity and the logarithm of the connectivity frequency. Selection of a soft threshold for constructing the WGCNA network. (**E**) Correlation heatmap between immune cell infiltration proportions and module genes. (**F**) Venn diagram of genes associated with M1 macrophages and DEGs. (**G**) PPI network of overlapping genes between M1 macrophage-related genes and DEGs.

groups of patients (Fig. 4C). Considering all factors, K = 2 was considered the optimal number for classifying STEMI patients. Furthermore, M1 macrophage-related genes were downregulated in M1-type patients and upregulated in M2-type patients (Fig. 4D). The findings revealed that the classification of STEMI patients by M1 macrophages is highly relevance. To further delineate differences between immune subtypes, we identified differentially expressed genes (DEGs) between the two subtypes. The results revealed 200 DEGs between the two subtypes, as shown in the volcano plot and heatmap (Fig. 4D,E; FC = 1.5).

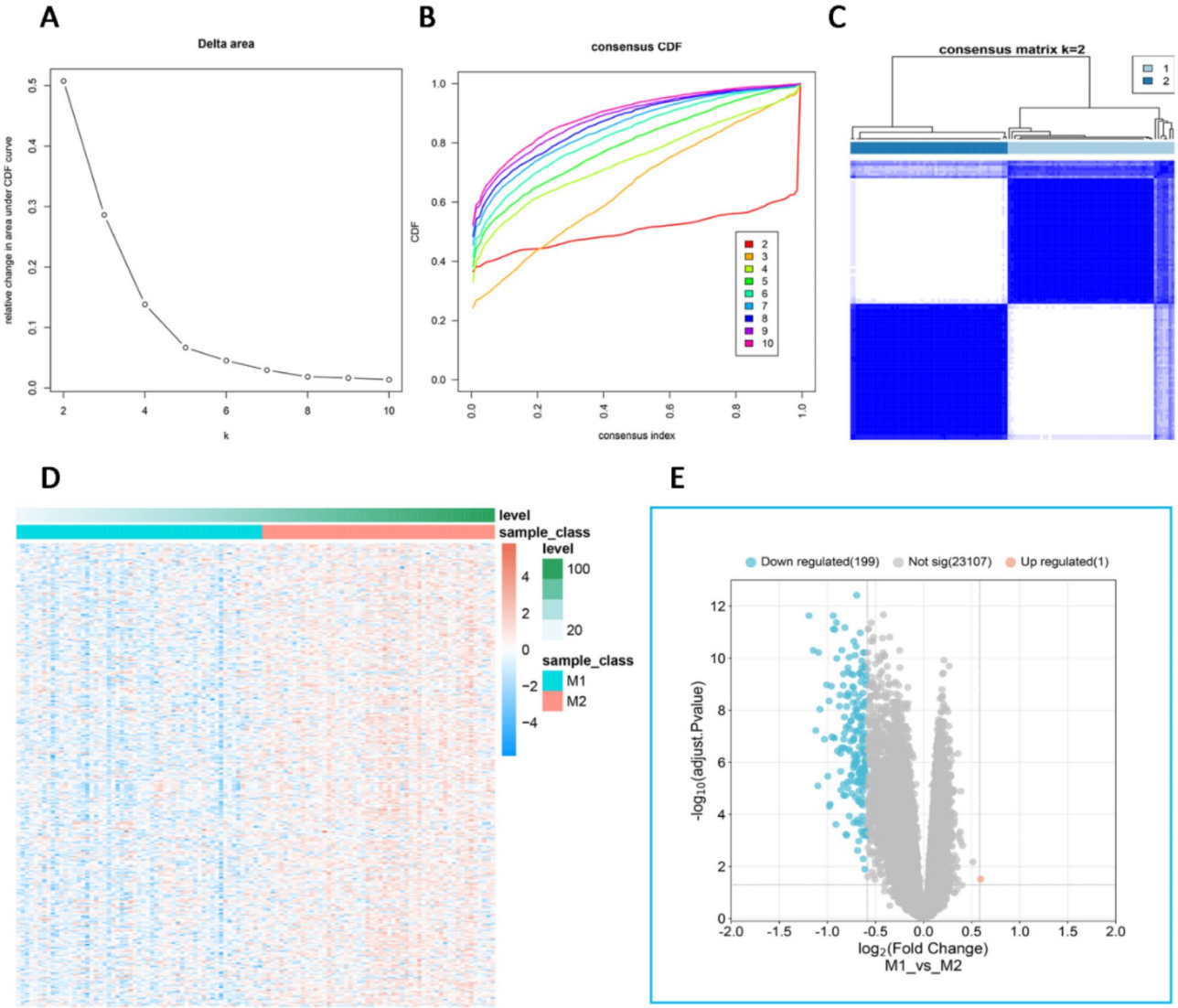

Fig. 4. Delta area (**A**), CDF (**B**), and consensus matrix (**C**) plots of clustered subtypes. (**D**) Expression distribution of M1 macrophages in the two subtypes. (**E**) Volcano plot of DEGs in the two subtypes.

Cultivating feature gene selection between STEMI subtypes using machine learning algorithms

The 31 feature genes were obtained by taking the intersection of 200 DEGs between STEMI subtypes and 82 PPI hub genes (Fig. 5A). Then, machine learning algorithms were used to select the most important feature genes. First, the XGBoost algorithm was employed for analysis. Figure 5B shows the trend of the Cox negative log-likelihood value during the training process. The top 10 important features are shown in Fig. 5C,G. Second, LASSO regression analysis was performed on 31 feature genes, and the cross-validation method was used for iterative analysis. The results revealed that the model's root mean square error was lowest when $\lambda = 0.09$ and when there were 8 variables (Fig. 5D,E,G). Third, we created a random forest model by constructing decision trees and partitioning genes using the log-rank rule. The Gini coefficient method was used to reduce the precision and mean square error. The top 15 feature genes of variable importance were output (Fig. 5F), and the 10 feature genes whose "Mean Decrease Gini" coefficient was greater than 1.6 were selected for further screening (Fig. 5G). By integrating the candidate genes selected by the above three machine learning algorithms, four genes (AKT3, GJC2, HMGCL, and RBM17) were considered feature genes for the immune subtypes (Fig. 5G,H).

The expression and predictive performance of the 4 feature genes in GSE59867 and GSE62646

To further validate the reliability of the selected immune subtype feature genes, we assessed the expression and predictive performance of these feature genes in the GSE59867 and GSE62646 datasets. The validation dataset GSE62646 was also standardized, as shown in the boxplot (Supplementary Fig. 2). After cluster analysis, no samples were excluded from the GSE62646 dataset (Supplementary Fig. 3). As anticipated, in both the

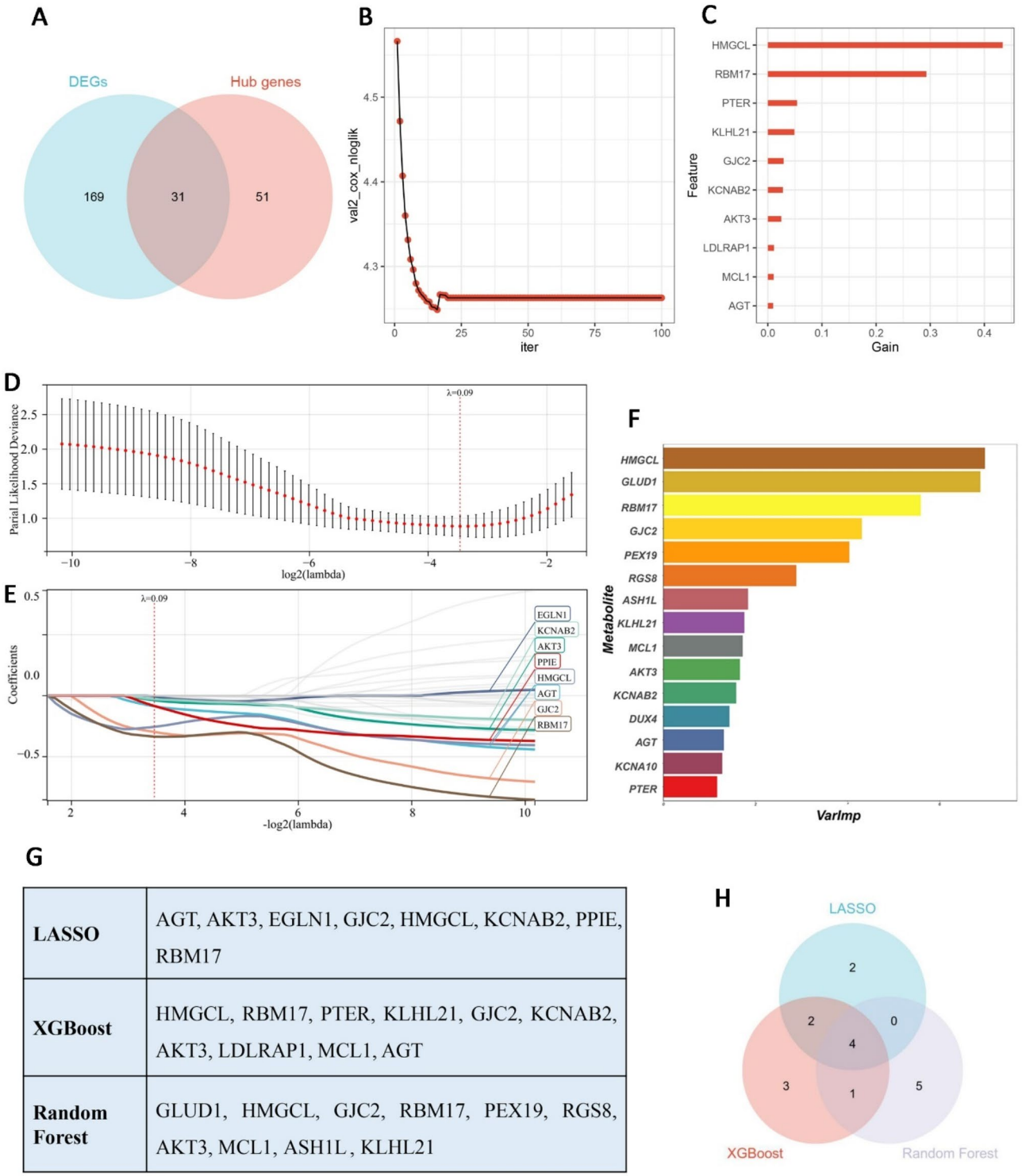

Fig. 5. (**A**) Overlapping genes between subtype-related DEGs and immune-related DEGs. (**B**) Plot of the number of iterations of the training process of the XGBoost algorithm versus the Cox negative log-likelihood value. (**C**) Bar chart of the top 10 genes and their corresponding importance scores screened by the XGBoost algorithm. (**D**) Cross-validation curve for LASSO regression analysis. (**E**) Path diagram of the LASSO coefficients and the 8 feature genes. (**F**) Bar chart of the top 15 genes and their variable importance scores screened using the random forest model. (**G**) Key genes obtained using different types of machine learning. (**H**) Venn diagram of the potential genes selected through three machine learning algorithms.

GSE59867 and GSE62646 datasets, the expression levels of AKT3, GJC2, HMGCL, and RBM17 were lower than those in the control group ($P < 0.05$, Fig. 6A). Furthermore, in the M1-type STEMI patient group, the expression levels of AKT3, GJC2, HMGCL, and RBM17 were lower than those in the M2-type STEMI patient group ($P < 0.05$, Fig. 6B). Notably, the area under the curve (AUC) values for AKT3, GJC2, HMGCL, and RBM17 were all greater than 0.5, regardless of sex, between control and STEMI patients or between M1-type and M2-type STEMI patients, indicating that these genes have strong diagnostic value (Fig. 6C,D). Finally, the diagnostic performance of AKT3, GJC2, HMGCL, and RBM17 in STEMI patients was confirmed through AUCs in the validation set GSE62646. The AUCs of AKT3, GJC2, HMGCL, and RBM17 in predicting STEMI in GSE62646 were 0.758, 0.732, 0.543 and 0.811, respectively (Fig. 6E).

Patient characteristics and real-time PCR verification of the key genes

The STEMI group comprised 14 Chinese men and 2 Chinese women with an average age of 62.7 ± 14.1 years. The control group comprised 6 Chinese men and 2 Chinese women with an average age of 65.5 ± 11.7 years.

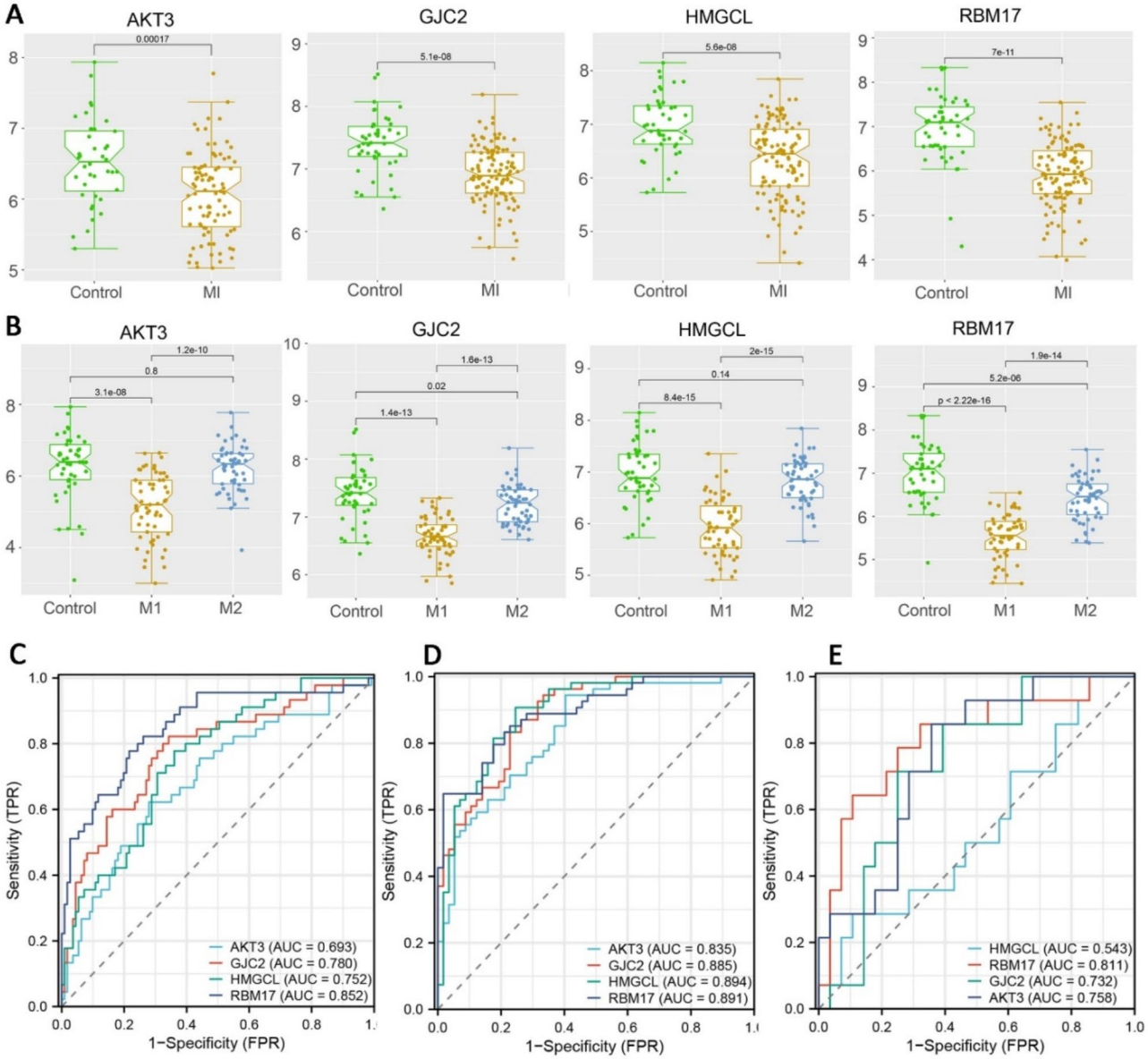

Fig. 6. (**A**) The expression levels of AKT3, GJC2, HMGCL, and REM17 in the STEMI and CAD groups in the GSE59867 dataset. (**B**) The expression levels of AKT3, GJC2, HMGCL, and REM17 in the M1 and M2 subgroups and in the CAD group in the GSE59867 dataset. (**C**) AUC curves of AKT3, GJC2, HMGCL, and REM17 between control and STEMI patients in the GSE59867 dataset. (**D**) AUC curves of AKT3, GJC2, HMGCL, and REM17 between M1-type and M2-type STEMI patients in the GSE62646 dataset. (**E**) AUC curves of AKT3, GJC2, HMGCL, and REM17 between control and STEMI patients in the validation dataset GSE62646.

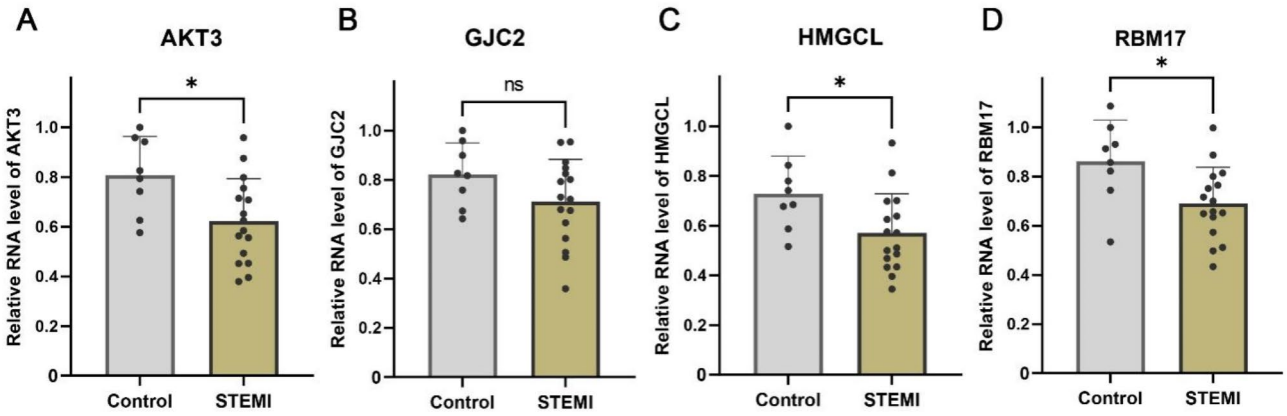

Fig. 7. Box plots showing the expression levels of AKT3 (**A**), GJC2 (**B**), HMGCL (**C**), and RBM17 (**D**) in the STEMI group compared with those in the control group (*$P < 0.05$).

The baseline characteristics of the patients, including blood pressure, heart rate, history of hyperlipidaemia, hypertension, diabetes, and medication history (ACEI/ARB, β-blockers, antiplatelets, calcium antagonists, and statins), are presented in Supplementary Table 1. The CRP, cTnT, NT-proBNP, creatinine and IL-6 levels in the STEMI group were greater than those in the control group ($P < 0.05$, Supplementary Table 1). Real-time PCR analysis revealed differential expression of the AKT3, GJC2, HMGCL and RBM17 genes between the STEMI and control groups, with AKT3, HMGCL and RBM 17 being downregulated significantly in the STEMI group (Fig. 7, Supplementary Table 2).

Discussion.

In the present study, we first determined the research subject and related datasets. Using the CIBERSORTx algorithm and WGCNA, we revealed a comprehensive interaction between M1 macrophages and STEMI, consistent with many other studies. Using the *ConsensusClusterPlus* package, we delineated STEMI patients into two distinct molecular subtypes. To screen for important genes, a protein–protein interaction (PPI) network and Venn diagrams were utilized. The intersection of the PPI hub genes and DEGs between the STEMI subtypes constituted the main feature gene of interest. Finally, we obtained four key genes through machine learning and validation in datasets or in patients: AKT3, GJC2, HMGCL, and RBM17. Real-time PCR revealed that all four genes mentioned above were downregulated in the STEMI patients, among which AKT3, HMGCL and RBM17 were downregulated significantly ($P < 0.05$).

AKT, a serine–threonine protein kinase, comprises three isoforms, namely, Akt1, Akt2, and Akt3. Among these, Akt3 is pivotal downstream of the PI3K signalling pathway, a highly conserved pathway found across eukaryotes. This pathway critically influences cardiac metabolism by fostering myocardial cell growth and survival, promoting coronary neovascularization, maintaining cardiac contractile function, and facilitating autophagy. Akt3 has been shown to specifically inhibit cholesterol ester accumulation and the formation of foam cells[22], a pivotal initial event in the development of atherosclerosis. In rats, there is a notable reduction in AKT3 expression in the spinal cord following myocardial ischaemia–reperfusion injury[23]. M1 macrophages are activated mainly by bacterial lipopolysaccharide (LPS) and IFNγ. Previous studies have shown that the PI3K–Akt pathway negatively regulates LPS signalling and gene expression in monocytes/macrophages. The activation or overexpression of PI3K or Akt kinases results in reduced macrophage activation by LPS, thus restricting proinflammatory and promoting anti-inflammatory responses[24]. These findings are consistent with those of the present study, which revealed a decrease in AKT3 expression and an increase in M1 macrophage activity during the acute phase of MI. The activation of AKT3-related pathways may be benefit myocardial tissue recovery during acute myocardial infarction. In addition, Akt signalling serves as an integrative hub for various extracellular and intracellular signals that orchestrate macrophage biology. This includes regulating pro- and anti-inflammatory cytokine production, phagocytosis, autophagy, apoptosis, and metabolic processes[24]. The outcomes of this study provide compelling evidence of increased M1 macrophage infiltration and reduced AKT3 expression during the acute phase of STEMI, potentially revealing a novel avenue for the treatment of myocardial infarction.

RBM17, alternatively known as SPF45, is a protein-coding gene with significant involvement in RNA spliceosomes, intricate molecular complexes crucial for maintaining cell survival and overall cellular integrity. RBM17 encodes an RNA-binding protein that is a vital component within the splice complex, actively participating in the second catalytic step of alternative splicing[25]. Notably, RBM17 is frequently overexpressed in many tumours and plays a crucial role in cancer progression, while the downregulated expression of RBM17 mRNA is accompanied by the induction of cell cycle arrest and apoptosis[26]. A recent original article identified RBM17 as a novel response biomarker for immunotherapy in bladder cancer, as it was associated with increased activity in the cell cycle and therapeutic responses[27]. The targeted silencing of RBM17 impedes cell proliferation[25,28]. Many studies have investigated the relationship between RBM17 and cellular immunotherapy, but the interaction between RBM17 and macrophages has not yet been reported. In our study, for the first time, we found that the expression of RBM17 was downregulated during the acute phase of STEMI, which might indicate an impaired regeneration of cardiac cells and increased cardiac cell death. Therefore, it was speculated

that activating RBM17 might be one of the methods used to respond to immune activation and regulate the cell cycle of myocardial cells in the acute stage of STEMI.

HMGCL, 3-hydroxy-3-methylglutaryl-CoA lyase, is a vital mitochondrial matrix protein and a key enzyme in both fatty acid and leucine metabolism. The final step in ketogenesis is catalysed by HMGCL to form the ketone body acetoacetate, which can then be converted to the other ketone bodies β-hydroxybutyrate (BHB) and acetone. A deficiency or absence of HMGCL may result in the aberrant accumulation of precursor substrates, such as 3-hydroxy-3-methylglutaric acid and 3-methylglutaric acid, and a lack of ketone bodies. This metabolic disturbance may contribute to organ damage, notably affecting critical organs such as the brain, heart, and liver, as ketone bodies impact a wide range of immune functions[29–31]. Innate immune cell-intrinsic ketogenesis is also dispensable for organismal metabolism and inflammation[32]. Studies have shown that HMGCL might be involved in immune escape and macrophage polarization[33]. Goldberg et al. reported that conditional ablation of HMGCL in neutrophils and macrophages or in bone marrow cells can affect glucose homeostasis in mice[32]. Accordingly, we infer that low expression of HMGCL in the myocardial infarction region is likely to affect cardiomyocyte and macrophage metabolism and death.

GJC2, spanning approximately 9.9 kb in length, comprises two exons. This gene encodes connexin 47 (Cx47), a gap junction protein comprising 439 amino acids. Connexins constitute a family of integral membrane proteins that are crucial in establishing gap junctional intercellular communication[34]. This communication facilitates the transfer of intercellular nutrients, ions, second messengers, and small molecules, thereby influencing various cellular processes, including cell death[35]. Cx47 regulates calcium signalling, the phosphorylation of ERK1/2, and the promotion of oligodendrocyte precursor cell proliferation. Interference with Cx47 through RNA silencing results in reduced calcium ion influx, diminished phosphorylation of ERK1/2, and decreased oligodendrocyte precursor cell proliferation[36]. Cx47 ablation induces severe inflammation; thus, a prominent immune response occurs in mice lacking Cx47[37,38]. In immune cells such as macrophages, the expression of GJC2/Cx47 and their intercellular communication may affect the activation state and function of macrophages. It has long been recognized that an intense oedematous reactions and inflammatory responses confined to the postischemic region appear early after STEMI. Reduced connexin expression and an aggravated immune response might be the mechanisms that cause oedematous and inflammatory reactions immediately after myocardial infarction. To date, there is no literature indicating a clear association between GJC2 and macrophage polarization. We found for the first time an association between the GJC2 gene and STEMI or M1 macrophages. In the real-time PCR analysis, the GJC 2 levels did not significantly differ between the STEMI and the control group, but they tended to decrease, which was consistent with the results of the bioinformatics analysis. Increasing the sample size may result in significant differences between the two groups.

STEMI is a life-threatening condition with a significant incidence and mortality rate. Diagnosis traditionally relies on classic symptoms of myocardial ischaemia, electrocardiographic findings, and biomarkers such as creatine kinase, troponin I, cardiac troponin T, and myoglobin. However, these traditional biomarkers lack significance in guiding STEMI treatment. Current evidence suggests that various immune cells participate in immune regulation following acute myocardial infarction, working together to clear necrotic tissue and rebuild damaged myocardium[39,40]. Failure to resolve immune responses adequately may contribute to extracellular matrix remodelling and interstitial myocardial fibrosis after STEMI[41]. Macrophages are crucial immune cells that serve as the primary responsive cells after myocardial infarction, and they regulate multiple stages following STEMI[5,42]. Identifying potential biomarkers associated with immune infiltration during the acute phase of STEMI, especially those related to macrophage function, holds considerable practical importance. Macrophages can differentiate from monocytes in the blood after passing through blood vessels. They can exhibit functional plasticity and polarize into two activated forms, M1 and M2, each with distinct immune functions. M1 macrophage polarization promotes myocardial cell damage after STEMI, whereas M2 macrophage polarization can inhibit myocardial cell damage, playing a critical role in the occurrence and prognosis of STEMI[43]. Though this view of proinflammatory M1 macrophages and M2 macrophages suppressing inflammation seems to be an oversimplification because these cells exploit very high level of plasticity in response to microenvironmental stimuli and thus represent a large scale of different immunophenotypes with overlapping functional properties[44], our construction of a WGCNA network revealed that DEGs were strongly correlated with M1-type macrophages and weakly correlated with M2-type macrophages consistent with many other studies. We ultimately identified four feature genes related to M1 macrophage infiltration and helps with subtype classification when STEMI occurs. The AUCs of AKT3, GJC2, and RBM17 in predicting STEMI in GSE62646 were 0.758, 0.732, and 0.811, respectively. HMGCL did not show a diagnostic advantage in the GSE62646 dataset. However, it was also significantly downregulated in STEMI patients. The advantage of the current analysis is that the GSE59867 dataset is a human dataset, the analysis steps are detailed and logical, and it was validated using a different dataset from the GEO database and 24 Chinese human blood samples. This research provides a more comprehensive understanding of the molecular basis of STEMI pathogenesis and may facilitate the development of novel therapeutic strategies.

Limitations

This work has several limitations. First, this study was conducted with data from online databases, which has inherent biases such as selection bias, publication bias and data quality bias. Second, following our research direction, we only found two suitable datasets, one as the experimental set and one as the validation set. The sample size of the validation set is relatively small. Future analyses may integrate multiple datasets to provide a more robust analysis. Third, the validation method on patients was not comprehensive enough. Future studies should be conducted among larger number of patients to validate the robustness and generalizability of our findings. Finally, no specific molecular pathways involving these four genes have been confirmed. Molecular biology studies of these four genes should be conducted to confirm their relationship with immune infiltration

after MI, and precision therapeutic strategies targeting these genes or their associated pathways could be explored in the future, thus promoting further development of treatments for patients with STEMI.

Conclusions

In this investigation, we identified AKT3, GJC2, HMGCL, and RBM17 as pivotal genes intricately tied to the immune microenvironment orchestrated by M1 macrophages in the early stages of STEMI through rigorous bioinformatics scrutiny. These genes exhibit exceptional diagnostic precision in delineating STEMI and could emerge as robust diagnostic biomarkers. This discovery could also open novel avenues for therapeutic intervention in STEMI management.

Data availability

The datasets supporting the conclusions of this article are available in the GEO database: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59867 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62646. Further inquiries can be directed to the corresponding author.

Received: 4 August 2024; Accepted: 3 February 2025
Published online: 01 April 2025

References

1. Jenča, D. et al. Heart failure after myocardial infarction: incidence and predictors. *ESC Heart Fail.* **8**, 222–237. https://doi.org/10.1002/ehf2.13144 (2021).
2. Damluji, A. A. et al. Mechanical complications of acute myocardial infarction: a scientific statement from the American Heart Association. *Circulation* **144**, e16–e35. https://doi.org/10.1161/cir.0000000000000985 (2021).
3. Roth, G. A. et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010 (2020).
4. Bhatt, D. L., Lopes, R. D. & Harrington, R. A. Diagnosis and treatment of acute coronary syndromes: a review. *JAMA* **327**, 662–675. https://doi.org/10.1001/jama.2022.0358 (2022).
5. Peet, C., Ivetic, A., Bromage, D. I. & Shah, A. M. Cardiac monocytes and macrophages after myocardial infarction. *Cardiovasc. Res.* **116**, 1101–1112. https://doi.org/10.1093/cvr/cvz336 (2020).
6. Jin, K. et al. Single-cell RNA sequencing reveals the temporal diversity and dynamics of cardiac immunity after myocardial infarction. *Small Methods* **6**, e2100752. https://doi.org/10.1002/smtd.202100752 (2022).
7. Yang, Y. et al. Macrophages after myocardial infarction: mechanisms for repairing and potential as therapeutic approaches. *Int. Immunopharmacol.* **143**, 113562. https://doi.org/10.1016/j.intimp.2024.113562 (2024).
8. Wang, C. et al. Macrophage polarization and its role in liver disease. *Front. Immunol.* **12**, 803037. https://doi.org/10.3389/fimmu.2021.803037 (2021).
9. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847. https://doi.org/10.1093/bioinformatics/btm254 (2007).
10. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. https://doi.org/10.1093/nar/gkv007 (2015).
11. Njiru, L. G. et al. Restoring soil nutrient stocks using local inputs, tillage and sorghum-green gram intercropping strategies for drylands in Eastern Kenya. *Heliyon*. https://doi.org/10.1016/j.heliyon.2023.e20926 (2023).
12. Le, S. E., Josse, J. & Husson, F. C. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
13. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720. https://doi.org/10.1093/bioinformatics/btm563 (2008).
14. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*. **12**, 453–457. https://doi.org/10.1038/nmeth.3337 (2015).
15. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. **9**, 559. https://doi.org/10.1186/1471-2105-9-559 (2008).
16. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170 (2010).
17. Long, Q., Zhang, X., Ren, F., Wu, X. & Wang, Z. M. Identification of novel biomarkers, shared molecular signatures and immune cell infiltration in heart and kidney failure by transcriptomics. *Front. Immunol.* **15**. https://doi.org/10.3389/fimmu.2024.1456083 (2024).
18. Yaqoob, A., Verma, N. K., Aziz, R. M. & Shah, M. A. Optimizing cancer classification: a hybrid RDO-XGBoost approach for feature selection and predictive insights. *Cancer Immunol. Immunother.* **73**, 261. https://doi.org/10.1007/s00262-024-03843-x (2024).
19. Yagin, B. et al. Cancer metastasis prediction and genomic biomarker identification through machine learning and eXplainable artificial intelligence in breast cancer research. *Diagnostics (Basel)*. https://doi.org/10.3390/diagnostics13213314 (2023).
20. Kassambara, A. & Kassambara, M. A. Package 'ggpubr'. *R package version 0.1* 6 (2020).
21. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer, 2016).
22. Ding, L. et al. Akt3 deficiency in macrophages promotes foam cell formation and atherosclerosis in mice. *Cell Metab.* **15**, 861–872. https://doi.org/10.1016/j.cmet.2012.04.020 (2012).
23. Li, S. Y. et al. Quantitative proteomics reveal the alterations in the spinal cord after myocardial ischemia–reperfusion injury in rats. *Int. J. Mol. Med.* **44**, 1877–1887. https://doi.org/10.3892/ijmm.2019.4341 (2019).
24. Vergadi, E., Ieronymaki, E., Lyroni, K., Vaporidi, K. & Tsatsanis, C. Akt signaling pathway in macrophage activation and M1/M2 polarization. *J. Immunol.* **198**, 1006–1014. https://doi.org/10.4049/jimmunol.1601515 (2017).
25. Han, Y. et al. Downregulation of RNA binding motif protein 17 expression inhibits proliferation of hypopharyngeal carcinoma FaDu cells. *Oncol. Lett.* **15**, 5680–5684. https://doi.org/10.3892/ol.2018.8012 (2018).
26. Li, C. et al. Exploration of the effects of the CYCLOPS gene RBM17 in hepatocellular carcinoma. *PLoS One* **15**, e0234062. https://doi.org/10.1371/journal.pone.0234062 (2020).
27. Song, B., Wu, P., Wan, C., Sun, Q. & Kong, G. Integrating single cell and bulk RNA sequencing data identifies RBM17 as a novel response biomarker for immunotherapy in bladder cancer. *Virchows Arch.* https://doi.org/10.1007/s00428-024-03952-z (2024).
28. Fukumura, K. et al. SPF45/RBM17-dependent, but not U2AF-dependent, splicing in a distinct subset of human short introns. *Nat. Commun.* **12**, 4910. https://doi.org/10.1038/s41467-021-24879-y (2021).
29. Karagiannis, F. et al. Impaired ketogenesis ties metabolism to T cell dysfunction in COVID-19. *Nature* **609**, 801–807. https://doi.org/10.1038/s41586-022-05128-8 (2022).

30. Liu, S. et al. β-Hydroxybutyrate impairs the release of bovine neutrophil extracellular traps through inhibiting phosphoinositide 3-kinase-mediated nicotinamide adenine dinucleotide phosphate oxidase reactive oxygen species production. *J. Dairy Sci.* **105**, 3405–3415. https://doi.org/10.3168/jds.2021-21174 (2022).
31. Ang, Q. Y. et al. Ketogenic diets alter the gut microbiome resulting in decreased intestinal Th17 cells. *Cell* **181**, 1263–1275. .e1216 (2020).
32. Goldberg, E. L., Letian, A., Dlugos, T., Leveau, C. & Dixit, V. D. Innate immune cell-intrinsic ketogenesis is dispensable for organismal metabolism and age-related inflammation. *J. Biol. Chem.* **299**, 103005. https://doi.org/10.1016/j.jbc.2023.103005 (2023).
33. Jiang, M. et al. The role of amino acid metabolism of tumor associated macrophages in the development of colorectal cancer. *Cells* https://doi.org/10.3390/cells11244106 (2022).
34. K Abrams, C. Mechanisms of diseases associated with mutation in GJC2/Connexin 47. *Biomolecules* https://doi.org/10.3390/biom13040712 (2023).
35. dong-xiao, L. et al. Clinical and molecular genetic analysis of three pedigrees with Pelizaeus-Merzbacher-like disease. *Chin. J. Appl. Clin. Pediatr.* **29**, 1240–1245 (2014).
36. Ana, H., Shuangshuang, Q. & Jianhua, F. Expression of connexin 47 and its effects on the myelination of oligodendrocytes in hypoxia-ischemia-induced periventricular leukomalacia in newborn rats. *J. China Med. Univ.* **49**, 306–312. https://doi.org/10.12007/j.issn.0258-4646.2020.04.004 (2020).
37. Zhao, Y. et al. Oligodendroglial connexin 47 regulates neuroinflammation upon autoimmune demyelination in a novel mouse model of multiple sclerosis. *Proc. Natl. Acad. Sci. USA.* **117**, 2160–2169. https://doi.org/10.1073/pnas.1901294117 (2020).
38. Papaneophytou, C. P. et al. Regulatory role of oligodendrocyte gap junctions in inflammatory demyelination. *Glia* **66**, 2589–2603. https://doi.org/10.1002/glia.23513 (2018).
39. Xie, Y., Wang, Y., Zhao, L., Wang, F. & Fang, J. Identification of potential biomarkers and immune cell infiltration in acute myocardial infarction (AMI) using bioinformatics strategy. *Bioengineered* **12**, 2890–2905. https://doi.org/10.1080/21655979.2021.1937906 (2021).
40. Kologrivova, I., Shtatolkina, M., Suslova, T. & Ryabov, V. Cells of the immune system in cardiac remodeling: main players in resolution of inflammation and repair after myocardial infarction. *Front. Immunol.* **12**, 664457. https://doi.org/10.3389/fimmu.2021.664457 (2021).
41. Brakenhielm, E., González, A. & Díez, J. Role of cardiac lymphatics in myocardial edema and fibrosis: JACC review topic of the week. *J. Am. Coll. Cardiol.* **76**, 735–744. https://doi.org/10.1016/j.jacc.2020.05.076 (2020).
42. Jian, Y. et al. Crosstalk between macrophages and cardiac cells after myocardial infarction. *Cell Commun. Signal.* **21**, 109. https://doi.org/10.1186/s12964-023-01105-4 (2023).
43. Liu, S. et al. M1-like macrophage-derived exosomes suppress angiogenesis and exacerbate cardiac dysfunction in a myocardial infarction microenvironment. *Basic Res. Cardiol.* **115**, 22. https://doi.org/10.1007/s00395-020-0781-7 (2020).
44. Strizova, Z. et al. M1/M2 macrophages and their overlaps—myth or reality? *Clin. Sci. (Lond)* **137**, 1067–1093. https://doi.org/10.1042/cs20220531 (2023).

Author contributions
H.L., Q.Z. and W. W. wrote the main manuscript text. Y. B, H. L and W.L. provided research ideas and supervised manuscript writing. Y. B. and H. L collected clinical data and made analysis. All authors reviewed the manuscript and agreed to submit.

Funding
This work was supported by the Key Project of Military Health Care [22BJZ26], the Special Project of Military Health Care [22BJZ32], the National Natural Science Foundation of China [82270270], the Opening Foundation of the National Clinical Medical Research Center for Gerontology [NCRCG-PLAGH 2022018], and the Innovation and Cultivation Fund of the Sixth Medical Center of the PLA General Hospital [CXPY202202].

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-89125-7.

Correspondence and requests for materials should be addressed to Y.B., H.L. or W.L.