



OPEN Differential expression of glucose metabolism and circadian rhythm-related genes in patients with acute myocardial infarction and their clinical significance

Ya Wang¹, Shuai Liu¹, Chen Liu¹, Jing Li¹, Yueying Lin² & Huijing Zhang^{1,3}✉

Acute myocardial infarction (AMI) is a leading cause of mortality worldwide, with about 7 million deaths annually from cardiovascular disease. Current diagnostic and therapeutic approaches face challenges, highlighting the need for new biomarkers and diagnostic methods. Abnormalities in glucose metabolism and circadian rhythm genes are linked to cardiovascular conditions, and understanding these mechanisms may reveal new therapeutic targets. This study aimed to identify differentially expressed genes related to glucose metabolism and circadian rhythms (GMCRRDEGs) in AMI. Bioinformatics techniques, including differential expression analysis, enrichment analysis, and machine learning models, were used to find prospective biomarkers for early diagnosis and therapeutic intervention. Twelve GMCRRDEGs were found, with six key genes (*JUN*, *EPAS1*, *IL1B*, *ADRB2*, *FOS*, *CD36*) incorporated into the diagnostic model, showing high accuracy in both the training set (AUC = 0.93) and validation set (AUC = 0.91). Enrichment analysis linked GMCRRDEGs to biological processes related to NO synthesis and tumor necrosis factor signaling pathways. Immune infiltration analysis showed significant changes in immune cell abundance, especially in mast cells and neutrophils, among high-risk patients. This study highlights GMCRRDEGs as AMI biomarkers, emphasizing their role in disease mechanisms and immune responses. Further research should validate these findings in larger cohorts to enhance early detection for AMI.

Keywords Acute myocardial infarction, Glucose metabolism abnormalities, Circadian rhythm, Biomarkers, Immune response

Acute myocardial infarction (AMI) is the predominant factor contributing to global mortality rates¹. By 2030, the incidence of AMI is expected to reach 75 million cases². The epidemiology of AMI highlights its high prevalence and mortality rates, necessitating improved strategies for early diagnosis and effective treatment^{3,4}. Current therapeutic approaches include pharmacological interventions and interventional strategies. However, challenges such as diagnostic difficulties and suboptimal treatment outcomes persist, emphasizing the critical need for innovative biomarkers and diagnostic techniques^{3,5}.

The molecular mechanisms underlying AMI involve several pathological processes including myocardial ischemia, apoptosis, and inflammatory responses^{1,3}. Ischemic conditions trigger a series of cellular processes that ultimately culminate in cardiomyocyte death and contribute to the overall damage observed in AMI. The initiation of pro-inflammatory signaling cascades, especially the NF- κ B, MAPK, and PI3K/Akt pathways, is essential for orchestrating these biological responses^{2,6}. These pathways are implicated in the regulation of apoptosis and inflammation, further complicating the disease process, and present potential targets for therapeutic interventions.

Abnormalities in glucose metabolism are intricately linked to several diseases including diabetes, metabolic syndrome, and cardiovascular disease. These conditions often coexist and exacerbate patient morbidity and mortality⁷. Diabetes significantly increases the risk of cardiovascular complications⁸. Metabolic syndrome,

¹Department of Cardiology, Hengshui People's Hospital, Hengshui, Hebei, China. ²Department of Outpatient Department, The Second People's Hospital of Hengshui, Hengshui, Hebei, China. ³Department of Cardiology, Hengshui People's Hospital, no.180 Renmin Road, Taocheng District, Hengshui 053000, Hebei, China. ✉email: huijingzhang@126.com

another closely related condition, encompasses a cluster of metabolic disorders, including obesity, dyslipidemia, and hypertension, which further contribute to cardiovascular diseases^{9,10}. The aforementioned research underscores the need to understand the fundamental mechanisms associated with AMI.

Under normal conditions, the circadian rhythm can coordinate the physiological functions of the cardiovascular system, maintaining the dynamic balance of various indicators such as blood pressure, heart rate, vascular tension, and blood coagulability^{11–13}. However, when the circadian rhythm is disrupted, the tolerance of the cardiovascular system to ischemia and hypoxia is significantly reduced, and the risk of AMI is markedly increased^{14,15}. Consequently, a more comprehensive exploration of the interplay between the blood-brain barrier, circadian rhythm, and AMI not only aids in clarifying the underlying mechanisms of myocardial infarction, but also has the potential to generate novel concepts and therapeutic targets for the formulation of prevention and treatment strategies that are grounded in the understanding of circadian rhythms.

This study employed a comprehensive approach that integrates batch effect removal, differential expression analysis, enrichment analysis, and machine learning model construction to investigate AMI. The advantage of this methodology lies in its ability to consolidate information from multiple datasets, thereby enhancing the reliability of the results, whereas the incorporation of machine learning techniques significantly improves the accuracy of the diagnostic models. The main goal of this study was to identify the essential genes linked to AMI and develop a robust diagnostic framework that can enhance early identification and tailored therapeutic approaches.

Results

Technology roadmap

The technology roadmap is illustrated in Fig. 1.

Merging of AMI data sets

We downloaded two datasets, GSE66360¹⁶ (49 AMI vs. 50 control samples, blood tissue, GPL570 microarray, Homo sapiens, PMID: 21737128) and GSE48060¹⁷ (31 AMI vs. 21 control samples, blood tissue, GPL570 microarray, Homo sapiens, PMID: 25684211), from the GEO database(<https://www.ncbi.nlm.nih.gov/geo/>)¹⁸. Initially, the GSE66360 and GSE48060 datasets were integrated into a single dataset using the R package, sva, to mitigate batch effects. Following this integration, distribution boxplots (Fig. 2A,B) were used to assess and compare the expression values of the combined datasets before and after removing the batch effect. Second,

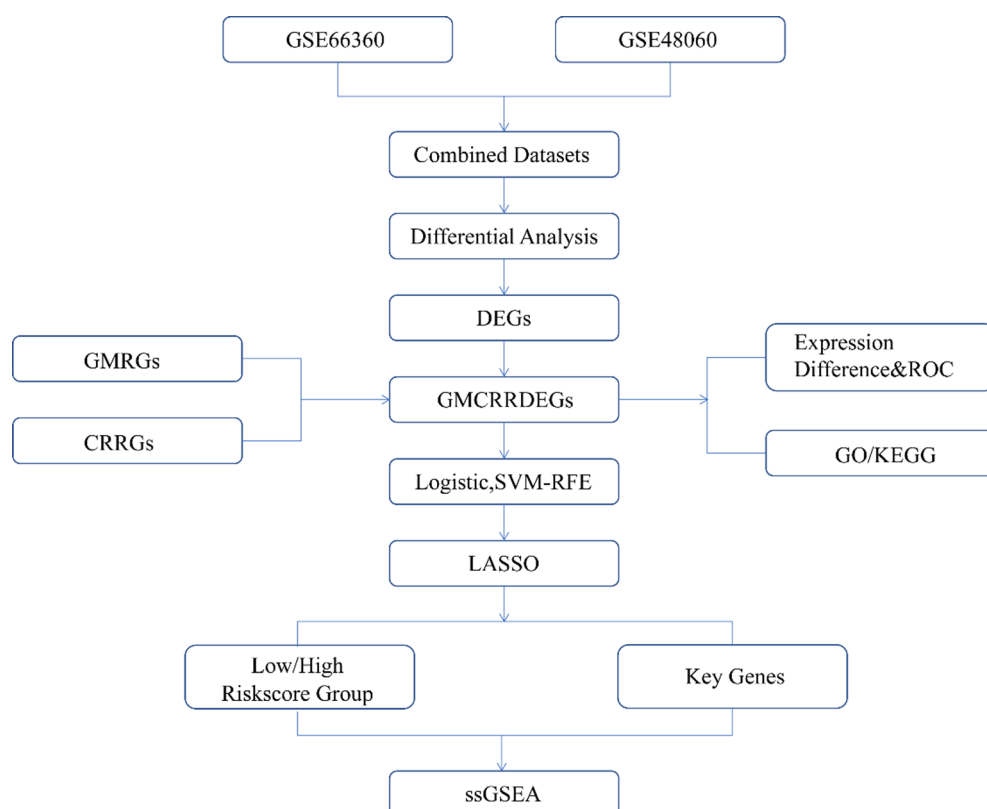


Fig. 1. Technology roadmap. DEGs, Differentially Expressed Genes; GMRGs, Glucose Metabolism-Related Genes; CRRGs, Circadian Rhythm-Related Genes; GMCRRDEGs, Glucose Metabolism and Circadian Rhythm-Related Differentially Expressed Genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, Least Absolute Shrinkage and Selection Operator; ssGSEA, Single-Sample Gene Set Enrichment Analysis.

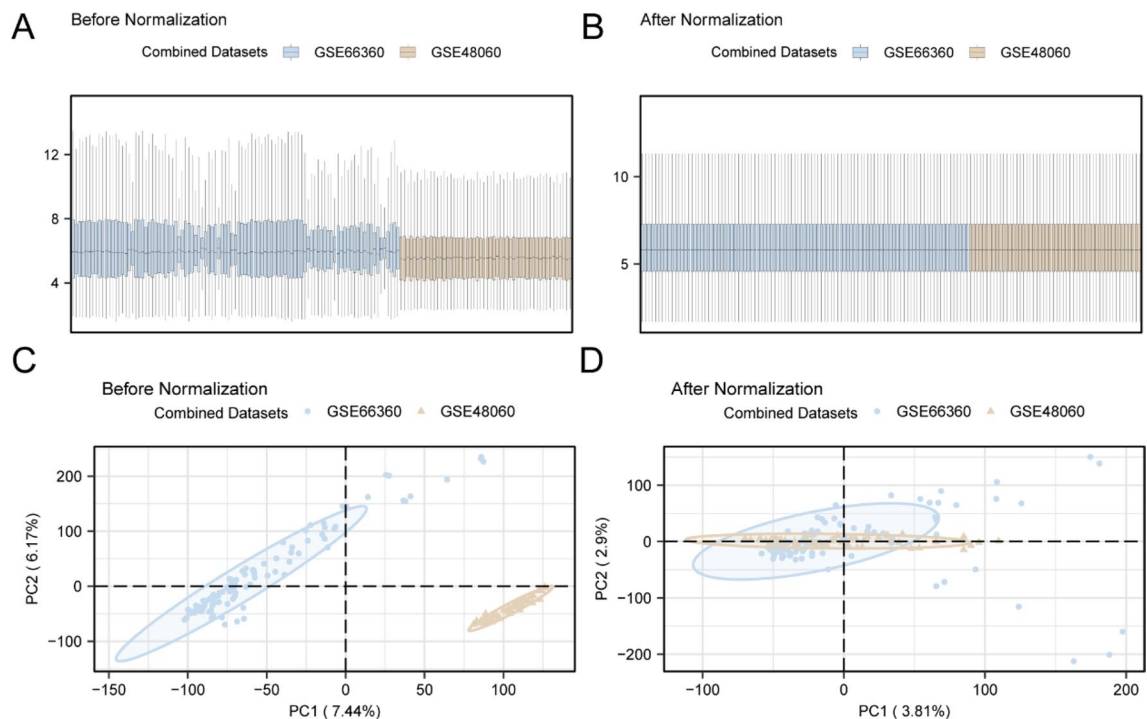


Fig. 2. Batch effects removal. (A) Box plot of the combined GEO datasets distribution before batch removal. (B) Post-batch integrated GEO datasets (combined datasets) distribution boxplots. (C) PCA plot of the datasets before debatching. (D) PCA map of the combined GEO datasets after batch processing. PCA, Principal Component Analysis; AMI, AMI. The AMI (AMI) datasets GSE66360 and GSE48060 are shown in blue and yellow, respectively.

a principal component analysis (PCA) plot (Fig. 2C,D) was employed to evaluate the distribution of low-dimensional features within the datasets, both before and after the removal of batch effects. The outcomes of the distribution boxplot and the PCA plot indicate that the batch effects present in the samples of the AMI dataset were effectively minimized following the batch removal process.

AMI related glucose metabolism and circadian rhythm-related differentially expressed genes (GMCRRDEGs)

This analysis revealed a total of 689 differentially expressed genes (DEGs) that satisfied the criteria of $|\log FC| > 0.5$ and a p -value < 0.05 , within the Combined Datasets. Under this threshold, 256 genes exhibited upregulation ($\log FC > 0.5$ and $p < 0.05$), while 433 genes showed down-regulation ($\log FC < -0.5$ and $p < 0.05$). A volcano plot was constructed using the results obtained from differential analysis of this dataset (Fig. 3A).

To identify Glucose Metabolism and Circadian Rhythm-Related Differentially Expressed Genes (GMCRRDEGs), all DEGs were intermingled with glucose metabolism related-genes (GMRGs) and circadian rhythm-related genes (CRRGs) to create a Venn diagram (Fig. 3B). Twelve GMCRRDEGs were identified, including *NAMPT*, *JUN*, *IL1B*, *PHACTR1*, *TKT*, *ADRB2*, *CD36*, *EPAS1*, *EDN1*, *TNF*, *FOS* and *HP*. Based on the results obtained from the intersection analysis, we assessed the differences in GMCRRDEGs expression across diverse sample groups within the combined datasets. The R package pheatmap was used to create a heatmap that effectively illustrates the analytical outcomes (Fig. 3C).

Through analysis using the R package, RCircos, 12 GMCRRDEGs were found to be located on the human chromosome (Fig. 3D). Chromosome mapping analysis revealed that the majority of candidate chromosomal regions (GMCRRDEGs) were located on chromosome 6, which encompassed genes such as *EDN1*, *PHACTR1*, and *TNF*.

Expression differences and correlation analysis of glucose metabolism and circadian rhythm-related differentially expressed genes (GMCRRDEGs) among the groups of the integrated GEO dataset

A comparative analysis of 12 critical candidate compounds (GMCRRDEGs) across cohorts within the combined datasets revealed distinct expression patterns through intergroup visualization (Fig. 4A). The analysis demonstrated significant differential expression ($p < 0.001$) of multiple biochemical markers, including *HP*, *FOS*, *TNF*, *EDN1*, *EPAS1*, *CD36*, *ADRB2*, *TKT*, *PHACTR1*, *IL1B*, *JUN*, and *NAMPT* between comparative groups. Extreme statistical significance thresholds (indicated by triple asterisks) highlight the strong molecular separation characteristics among these targets.

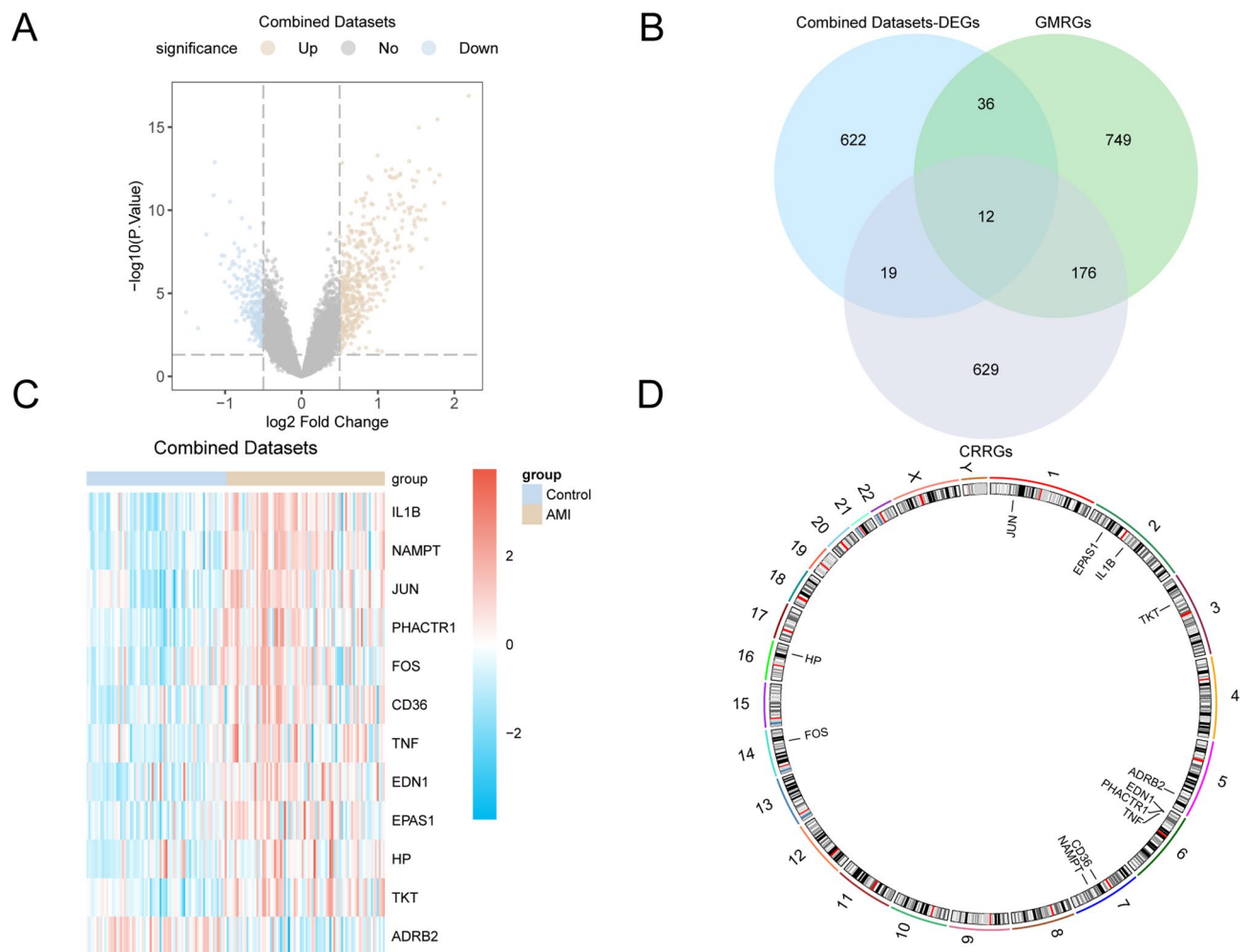


Fig. 3. Differential gene expression analysis. **(A)** Volcano plot of differentially expressed gene analysis between AMI group and Control group in combined GEO datasets. **(B)** Venn diagram of genes associated with GMRGs and CRRGs in all AMI samples from DEGs in the integrated GEO datasets. **(C)** Heat map of GMCRRDEGs in the integrated GEO dataset. **(D)** Chromosomal mapping of GMCRRDEGs. DEGs, differentially expressed genes; GMRGs, glucose metabolism-related genes; CRRGs, circadian rhythm-related genes; GMCRRDEGs, glucose metabolism and circadian rhythm-related differentially expressed genes. Blue is the Control group and yellow is the AMI group. In the heat map, red represents high expression and blue represents low expression.

The ROC curve (Fig. 4B-E) illustrated that the accuracy of FOS expression values for diagnosing the Control and AMI groups was low, as shown by an area under the curve (AUC) between 0.5 and 0.7. The expression values of *HP*, *TNF*, *EDN1*, *EPAS1*, *CD36*, *ADRB2*, *TKT*, *PHACTR1*, *IL1B*, *JUN*, *NAMPT* had a certain accuracy in the diagnosis of the Control and AMI groups ($0.7 < \text{AUC} < 0.9$).

We also calculated the pairwise correlation of the 12 GMCRRDEGs in the combined datasets using a correlation heat map was used for display (Fig. 4F). The results showed that most of the GMCRRDEGs were positively correlated, and the genes *NAMPT* and *IL1B* were significantly positively correlated ($r = 0.774$). Finally, ggplot2 in R was used to produce a scatter plot of the four most strongly correlated gene pairs (Fig. 4G-J).

Functional and pathway enrichment analysis

Functional annotation of the 12 GMCRRDEG-related genes using gene ontology (GO)/Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment methodologies revealed significant pathway clustering, as shown in Supplementary Table S1. These differentially expressed targets demonstrated pronounced involvement in nitric oxide homeostasis mechanisms, particularly manifesting in three-tiered regulatory hierarchies: (1) activation control of nitric oxide biosynthesis, (2) modulation of nitric oxide metabolic flux, and (3) peptide-responsive cellular signaling dynamics. Additionally, biological specificity has emerged in redox-sensitive macromolecular interaction networks. Several cellular components (CC) were enriched, including endocytic vesicles, specific granules, endocytic vesicle membranes, haptoglobin-hemoglobin complexes, and clathrin-coated vesicles. In the context of MF, the identified enrichment encompassed activities included cytokine activity, R-SMAD binding, SMAD binding, binding of transcription factors specific to RNA polymerase II, and amyloid-beta binding. Furthermore, these genes exhibited significant enrichment within pathways, including the TNF

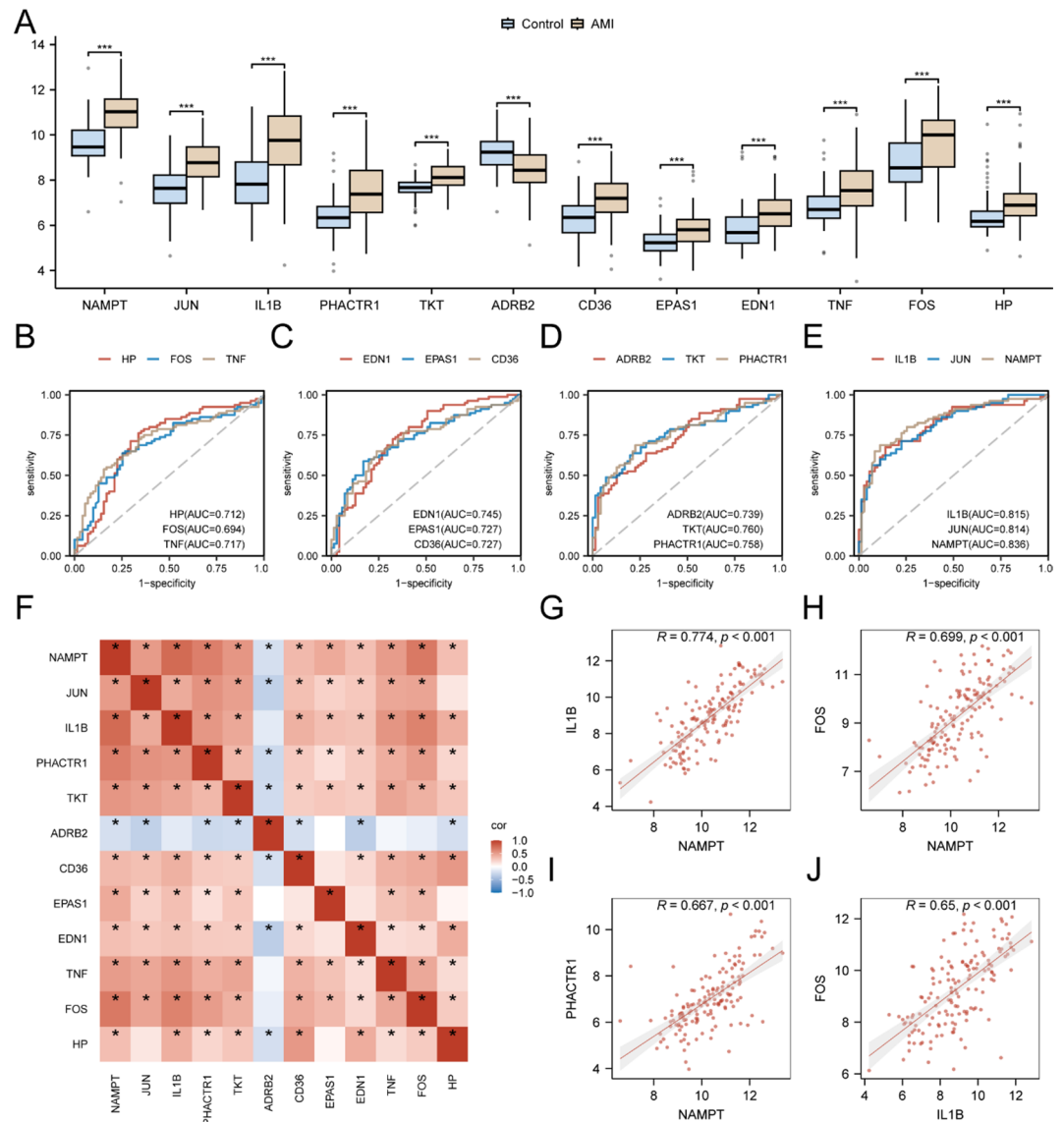


Fig. 4. Differential expression and correlation analysis. **(A)** Group comparison of GMCRRDEGs in the AMI group and the Control group in the combined GEO datasets. **(B–E).** ROC curves of *HP*, *FOS* and *TNF* **(B)**; *EDN1*, *EPAS1* and *CD36* **(C)**; *ADRB2*, *TKT* and *PHACTR1* **(D)**; *IL1B*, *JUN*, and *NAMPT* **(E)** in the combined GEO datasets. **(F)** Heat map of GMCRRDEGs in the AMI group and Control group in the integrated GEO datasets. **(G)** Scatter plot of the association between *NAMPT* and *IL1B* in GMCRRDEGs. **(H)** Scatter plot of the correlation between *NAMPT* and *FOS* in GMCRRDEGs. **(I)** Scatter plot of the correlation between *NAMPT* and *PHACTR1* in GMCRRDEGs. **(J)** Scatter plot of the correlation between *IL1B* and *FOS* in GMCRRDEGs. AMI, AMI; GMCRRDEGs, Glucose Metabolism and Circadian Rhythm-Related Differentially Expressed Genes. In the correlation heat map, red is a positive correlation, blue is a negative correlation, and the depth of color represents the strength of the correlation.

signaling pathway, fluid shear stress, atherosclerosis, pertussis, and leishmaniasis, as well as lipid metabolism and atherosclerosis pathways, as determined by KEGG analysis. Enrichment analysis results are displayed as bubble plots (Fig. 5A).

Network maps of the enrichment analysis were drawn (Fig. 5B–E). The lines illustrate the associated molecules along with the descriptions of the entries. Nodes of larger size signify entries that encompass a higher quantity of molecules.

Establishment of diagnostic model for AMI

To evaluate the diagnostic significance of the 12 GMCRRDEGs in AMI, logistic regression analysis was conducted using the 12 GMCRRDEGs variables. Subsequently, a logistic regression model was developed, and the results were illustrated using a Forest Plot (Fig. 6A). The analysis demonstrated that all 12 GMCRRDEGs exhibited statistical significance within the logistic regression model, as indicated by p-values that were less than 0.05,

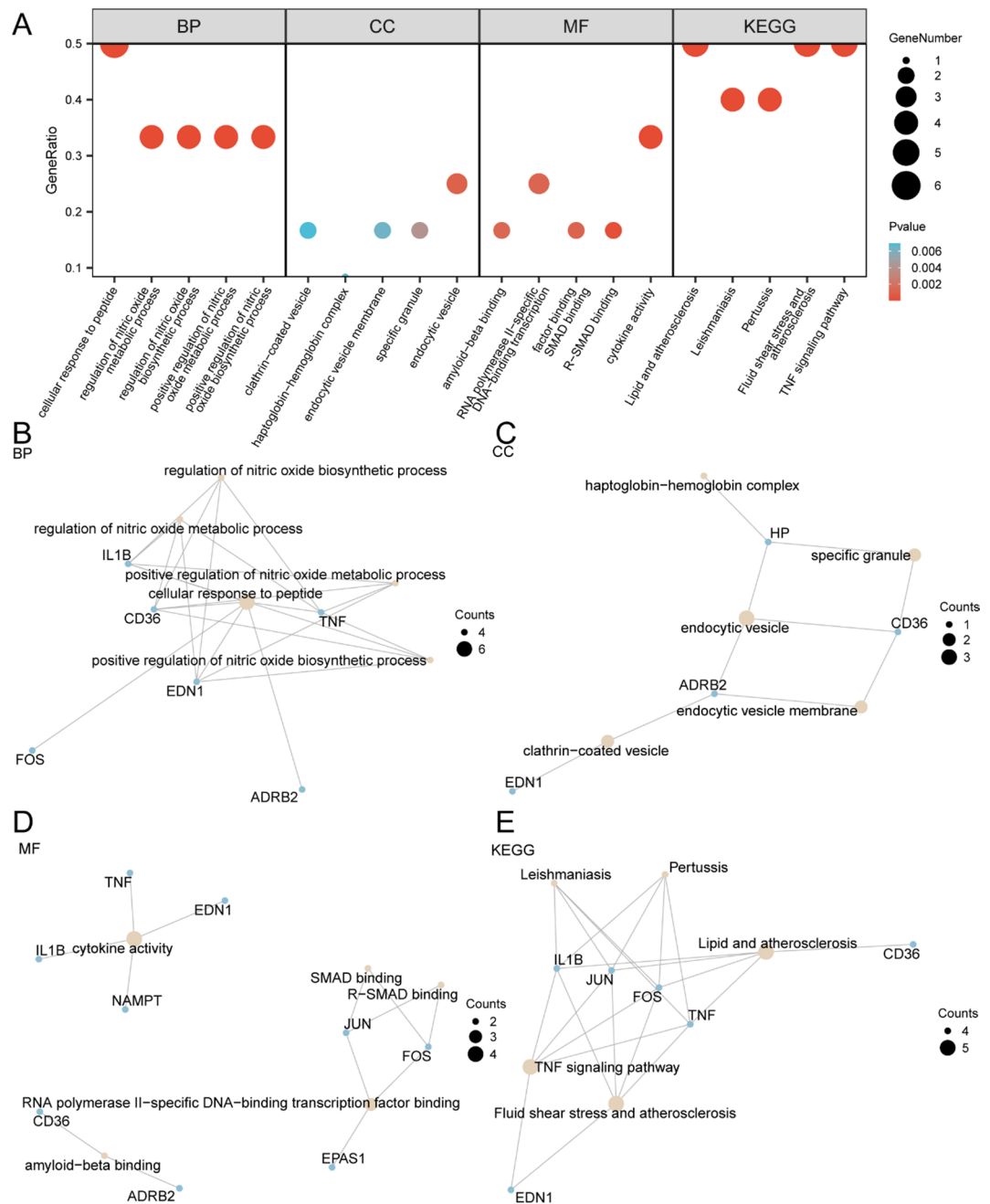


Fig. 5. GO and KEGG enrichment analysis. **(A)** Bubble diagram showing the results of GO and KEGG enrichment analysis of GMCRRDEGs: GO terms and KEGG terms are shown on the abscissa. **(B–E)** The GO and KEGG network diagram showing: BP **(B)**, CC **(C)**, MF **(D)** and KEGG **(E)**. The light-yellow nodes represent items, the light blue nodes represent molecules, and the lines represent the relationship between items and molecules. GMCRRDEGs, glucose metabolism and circadian rhythm-related differentially expressed genes; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological process; CC, cellular component; MF, molecular function. In the bubble plot, bubble size and color represent the number of genes and size of the p-value value, respectively. The redder the color, the smaller the p-value value, and the bluer the larger the p-value value. The screening criteria for GO and KEGG enrichment analysis were p value < 0.05 and FDR value (q value) < 0.25.

namely *NAMPT*, *EDN1*, *EPAS1*, *JUN*, *IL1B*, *ADRB2*, *CD36*, *HP*, *TNF*, *TKT*, *FOS*, *PHACTR1*. Second, the SVM-RFE algorithm was used based on the 12 GMCRRDEGs. 5-fold cross-validation was set up, and the average rank of the genes was calculated to obtain the number of genes with the lowest error rate (Fig. 6B) and highest accuracy rate (Fig. 6C) of the model. When the number of genes was 6, the SVM model had the highest accuracy, and the top 6 average-ranked genes were screened for subsequent analyses (Fig. 6D). The 6 GMCRRDEGs identified were *JUN*, *EPAS1*, *IL1B*, *ADRB2*, *FOS*, and *CD36*. Subsequently, they were incorporated into the support

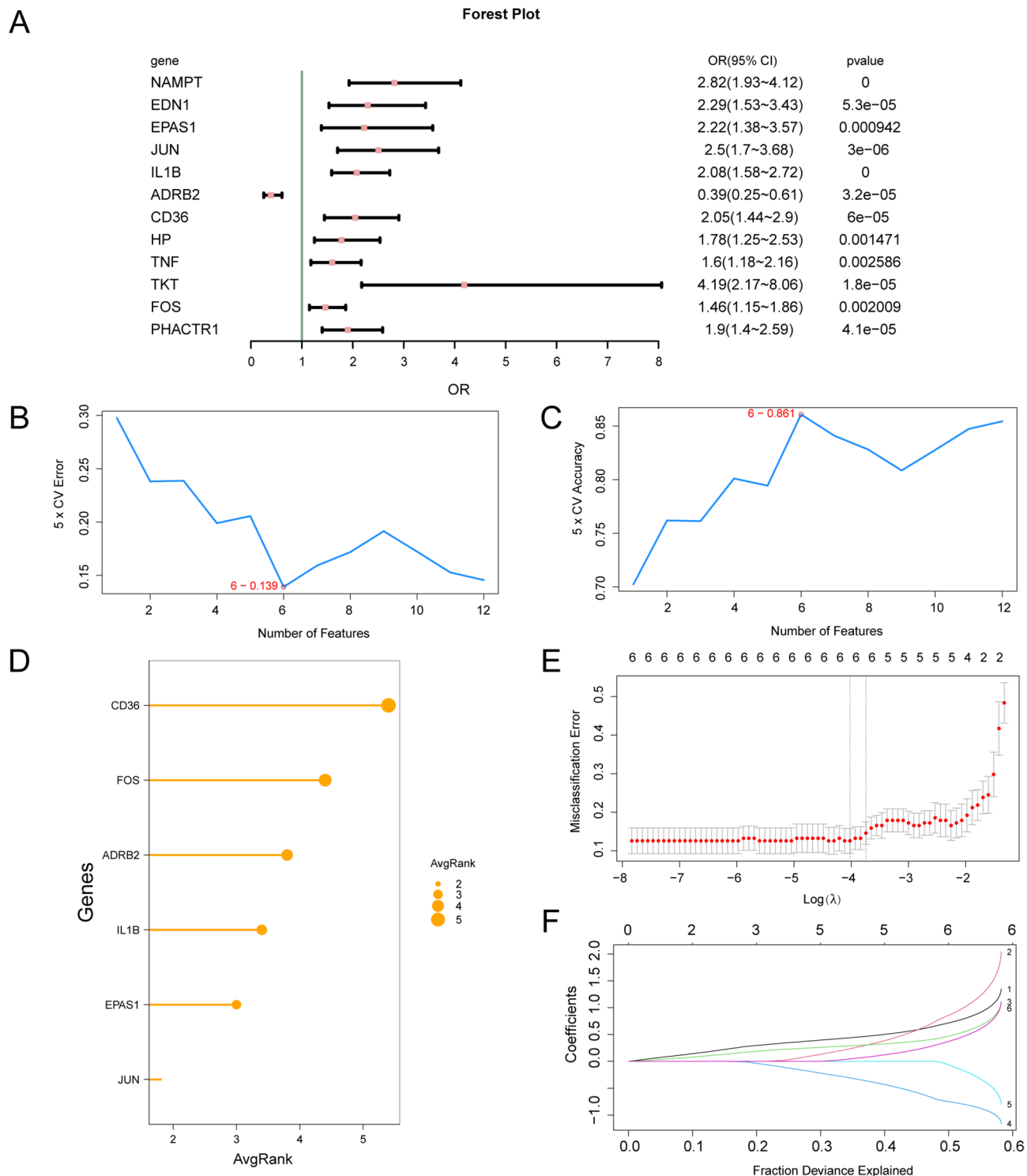


Fig. 6. Diagnostic model of AMI. (A) Forest plot of 12 GMCRRDEGs included in the logistic regression model in the diagnosis model of AMI. (B, C). The number of genes with the lowest error rate (B) and the number of genes with the highest accuracy (C) obtained by SVM-RFE algorithm are visualized. (D) The average importance ranking lollipop plot of the six genes with the lowest error rate obtained by the SVM-RFE algorithm. (E–F). Diagnostic model plot (E) and variable trajectory plot (F) of LASSO regression model. AMI, AMI; SVM, support vector machine; LASSO, least absolute shrinkage and selection operator.

vector machine (SVM) model following their selection through least absolute shrinkage and selection operator (LASSO) regression analysis to develop a diagnostic model for AMI. Visualization was accomplished using the LASSO regression model diagram (Fig. 6E) along with the LASSO variable trajectory diagram (Fig. 6F). The analysis indicated that six GCMRDEGs recognized as key genes were used in the LASSO regression model. The key genes were *JUN*, *EPAS1*, *IL1B*, *ADRB2*, *FOS*, and *CD36*.

Validation of the diagnostic model for AMI

To further verify the effectiveness of the AMI diagnostic models, a nomogram was constructed based on key genes to illustrate their interrelationships (Fig. 7A). The analysis revealed that the expression levels of the pivotal gene *IL1B* exhibited markedly higher significance than other variables included in the diagnostic model of AMI. The expression level of *CD36* was less significant in the AMI diagnostic model than other variables.

The accuracy and resolution of the AMI diagnostic model were assessed by constructing a calibration curve based on our analysis. The predictive performance of the model was evaluated by comparing the actual and predicted probabilities under various conditions, as illustrated in Fig. 7B. The calibration curve pertaining to the diagnostic model for AMI indicated that the calibration line, represented by the dotted line, displayed a slight divergence from the diagonal line that signified the optimal model; however, it maintained proximity to the fitted line. Decision curve analysis (DCA) was used to assess and demonstrate the clinical utility of the AMI diagnostic models developed using key genes in the combined datasets (Fig. 7C). The examination indicated that, within a specific interval, the model's trajectory consistently surpassed the lines representing both positive and negative outcomes. This indicates the greater net benefit and superior performance of the model. Additionally, the ROC curve (Fig. 7D) demonstrated that RiskScore expression levels in the combined datasets achieved high accuracy ($AUC > 0.9$) across different groups.

Patients diagnosed with AMI were classified into two distinct groups, High-Risk and Low-Risk, using the median risk score derived from the diagnostic model as the classification criterion. The risk score was calculated using the following formula:

$$\text{RiskScore} = \text{JUN} * (0.7828) + \text{EPAS1} * (0.9825) + \text{IL1B} * (0.5379) + \text{ADRB2} * (-0.7919) + \text{FOS} * (-0.1603) + \text{CD36} * (0.4531)$$

Risk-stratified expression profiling of six pivotal biomarkers in the AMI cohort (Fig. 7E) revealed distinct transcriptional patterns between the high-risk and low-risk subpopulations. The analysis demonstrated tiered statistical significance across biomarkers: *ADRB2* demonstrated nominal statistical significance ($p < 0.05$), *EPAS1* exhibited robust statistical significance ($p < 0.01$), and *JUN*, *IL1B*, *FOS*, and *CD36* reached stringent significance thresholds ($p < 0.001$). Notably, the overexpression trends of *JUN*, *IL1B*, *FOS*, and *CD36* in high-risk subjects suggested a potential mechanistic convergence in AMI pathogenesis. These transcriptional disparities were quantified through risk-group comparative analysis, with effect sizes proportional to the significance levels.

The results obtained from the functional similarity (Friends) analysis were used to identify genes that have a substantial influence on the biological process of AMI (Fig. 7F). The analysis revealed that *FOS* serves as a vital gene in AMI, being closest to the established critical threshold (cutoff value = 0.7).

Immune infiltration analysis of SsgSEA algorithm based on high and low logistic risk scores

The group comparison diagram (Fig. 8A) revealed that all 17 immune cell types showed statistically significant differences ($p < 0.05$), including Activated CD8 T cells, activated dendritic cells, central memory CD8 T cells, eosinophils, gamma delta T cells, immature dendritic cells, macrophages, mast cells, MDSC, memory B cells, monocytes, natural killer cells, natural killer T cells, neutrophils, plasmacytoid dendritic cells, regulatory T cells, and T follicular helper cells. Subsequently, correlation analysis regarding the abundance of 17 immune cell infiltrations within the AMI samples was graphically represented using a correlation heatmap (Fig. 8B-C). The analysis revealed that in the low-risk AMI samples, most immune cells exhibited strong positive correlations, with mast cells and neutrophils showing the strongest significant positive correlations ($r = 0.726$, $P < 0.05$) (Fig. 8B). Within the high-risk cohort, the majority of immune cell types exhibited pronounced positive correlations. Notably, activated dendritic cells and macrophages demonstrated the most significant positive correlation, characterized by an r value of 0.836 and a p -value of less than 0.05 (Fig. 8C). Finally, the association between essential genes and the density of immune cell infiltration was depicted using a correlation bubble plot (Fig. 8D-E). These results indicate that many immune cells were strongly correlated in the low-risk group. Specifically, *CD36* and Gamma delta T cells exhibited the strongest significant positive correlation, with an r value of 0.672 ($p < 0.05$). In the high-risk group, many immune cells showed strong positive correlations, the most notable being between *FOS* and mast cells ($r = 0.77$, $p < 0.05$) (Fig. 8E).

Discussion

During the study design phase, we conducted a systematic search of major public databases such as GEO and ArrayExpress, applying strict selection criteria aligned with the core objectives of our research. We chose to include two GEO datasets, GSE66360 and GSE48060, which both originate from the GPL570 platform. These datasets meet the following criteria: they are derived from human samples, specifically from blood tissues, and they are all based on the GPL570 platform. Combined, they provide a sample size of over 150 cases. Furthermore, we implemented rigorous batch effect removal and normalization steps. As shown in Fig. 2, the batch effects have been nearly completely eliminated, which enhances the overall statistical power and reliability of our results. Finally, the combined datasets were annotated and subjected to standardization and normalization using the R package limma.

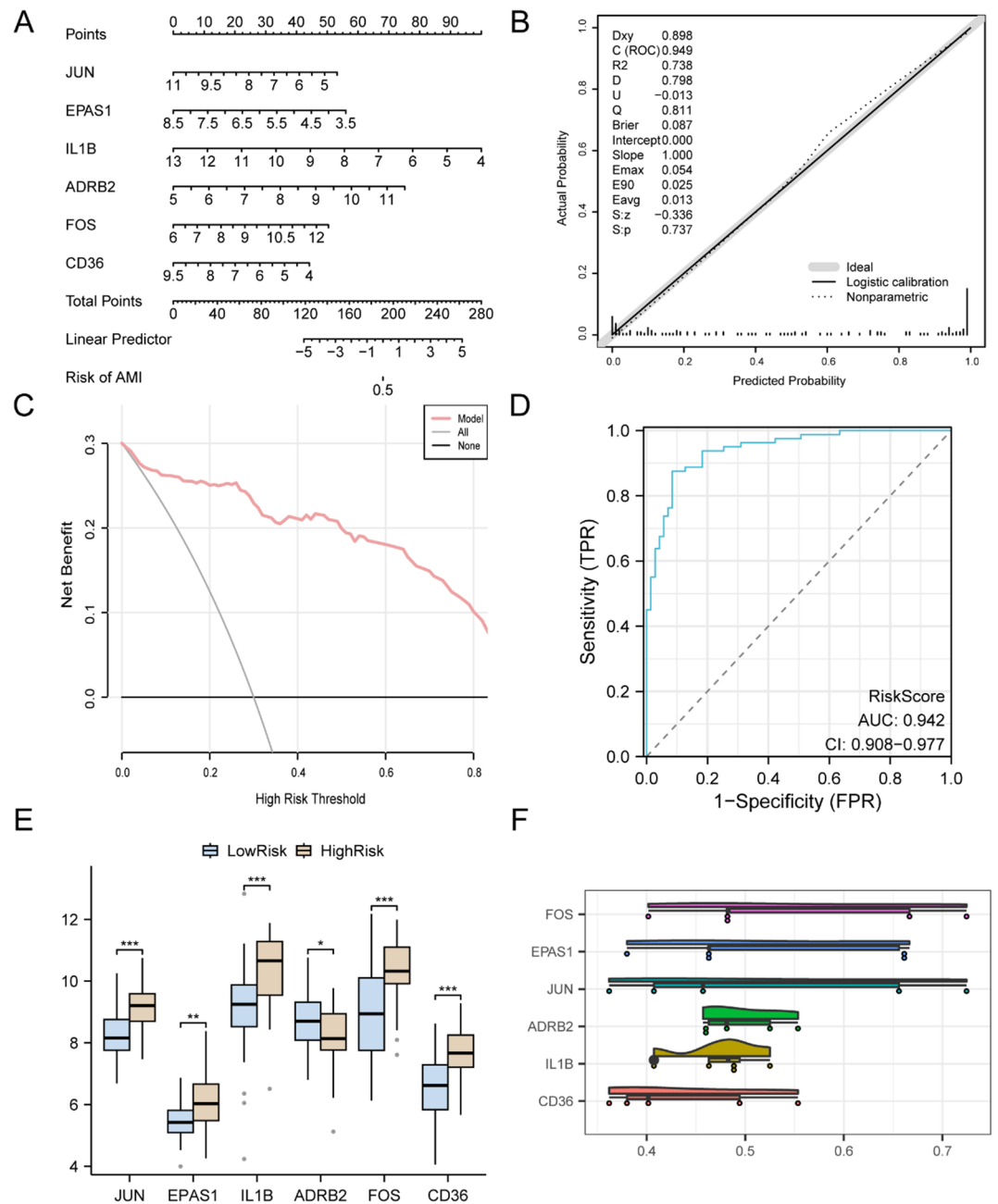


Fig. 7. Diagnostic and validation analysis of AMI. (A) Nomogram of key genes in diagnostic models for diagnosing AMI based on combined GEO datasets. (B,C). Calibration curve plot (B) and decision curve analysis (DCA) plot (C) of the AMI diagnostic model based on the key genes in the integrated GEO datasets. (D) ROC curve of RiskScore in the combined GEO datasets. (E) Group comparison plots of Key Genes in the High-Risk and Low-Risk groups of AMI group. (F) Cloud rain diagram of the results of Friends analysis of key genes. DCA: Decision Curve Analysis; AMI, AMI. * represents p value < 0.05 , statistically significant; *** represents p value < 0.001 , highly statistically significant. In the group comparison graph, blue represents the Low-Risk group and yellow represents the High-Risk group.

Through a comprehensive analysis of the combined datasets, we identified 12 DEGs linked to glucose metabolism and circadian rhythms, including notable genes such as *NAMPT* and *IL1B*. These results underscore the intricate interactions between metabolic mechanisms and circadian regulation in AMI. Moreover, our study employed robust methodologies, including batch-effect removal and differential gene expression analysis, which enhanced the reliability of our results.

During an AMI, cardiac myocytes undergo ischemia and hypoxia, causing a drastic shift in energy metabolism from aerobic oxidation to anaerobic glycolysis. This transition leads to reduced glucose utilization efficiency and the accumulation of lactate, which significantly impairs myocardial contractile function. Numerous clinical and

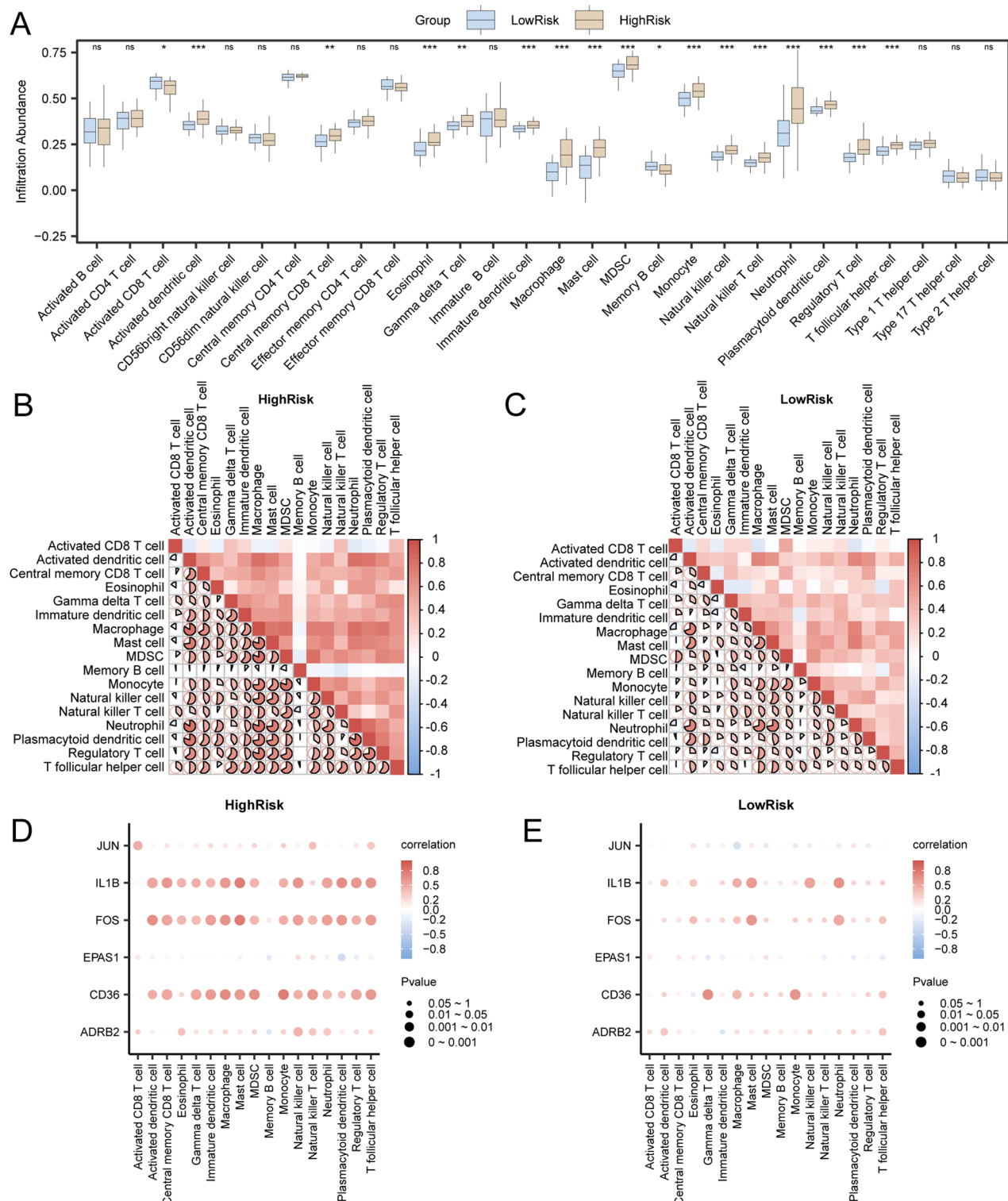


Fig. 8. Immune infiltration analysis by ssGSEA algorithm. (A) Group comparison diagram of immune cells in the Low and High Risk groups in the combined GEO datasets. (B,C) Results of correlation analysis of immune cell infiltration abundance in the Low-Risk (B) and High-Risk (C) groups of AMI samples are presented. (D,E) Bubble plot of correlation between immune cell infiltration abundance and key genes in the Low-Risk (D) and High-Risk (E) groups of AMI samples. ssGSEA, single-sample gene set enrichment analysis; AMI, AMI. ns stands for p value ≥ 0.05 , no statistical significance; * represents p value < 0.05 , statistically significant; ** represents p value < 0.01 , highly statistically significant; *** represents p value < 0.001 extremely statistically significant. The absolute value of the correlation coefficient (r value) ranged from 0.5 to 0.8, indicating a moderate correlation. Low-Risk group (blue) and High-Risk group (yellow). Red is a positive correlation, blue is a negative correlation, and the depth of the color represents the strength of the correlation.

preclinical studies have shown that patients with acute myocardial infarction often experience stress-induced hyperglycemia, and elevated blood glucose levels are closely linked to worse prognoses¹⁹. In the pathological context of AMI, the interplay between glucose metabolism and circadian rhythm genes holds significant biological relevance. For example, *EPAS1*, a hypoxia-inducible factor, enhances the conversion of glucose to lactate in ischemic cardiac tissue by binding to the promoter regions of critical glycolytic enzymes, thereby sustaining energy production. Meanwhile, *IL1B* promotes apoptosis in pancreatic β -cells through activation of the NF- κ B pathway, while concurrently inhibiting the phosphorylation of insulin receptor substrates, which diminishes the ability of myocardial cells to uptake glucose²⁰.

Analysis of the key genes associated with AMI revealed significant findings regarding *JUN*, *IL1B*, and *FOS*, which were notably upregulated in the AMI group. *JUN* is a widely recognized transcription factor that plays an essential role in cellular responses to stress and inflammation²¹, and its elevated expression suggests its potential involvement in the pathological processes of myocardial injury and repair following AMI. Upregulation of *JUN* may indicate an adaptive response to ischemic stress, contributing to the regulation of genes involved in cell survival and inflammation. *IL1B*, which encodes interleukin-1 beta, is a pro-inflammatory cytokine implicated in various cardiovascular diseases²². The significant increase in the AMI cohort underscores its role in mediating inflammatory responses that can exacerbate myocardial damage and influence healing process²³. These findings align with existing literature that highlights *IL1B*'s contribution to myocardial ischemia and subsequent remodeling²⁴, suggesting that targeting *IL1B* could be a therapeutic strategy for managing AMI. *FOS* is another critical component of the AP-1 transcription factor complex²⁵, is involved in cell proliferation and differentiation. The observed upregulation of *FOS* in patients with AMI indicates its potential role in the inflammatory response and cellular signaling pathways activated during myocardial infarction. The significant upregulation of these genes in AMI highlights their potential as therapeutic targets and biomarkers for disease progression and underscores the need for further investigation into their roles in myocardial ischemia and recovery processes.

GO and KEGG enrichment analyses indicated that nitric oxide (NO) synthesis, cellular stress response, and TNF signaling pathway play crucial roles in AMI. NO is vital in various physiological processes, including vasodilation and the immune response, and its dysregulation is linked to the pathogenesis of cardiovascular diseases²⁶. NO significantly contributes to inflammatory responses and vascular function during AMI²⁷; excessive production can lead to oxidative stress and myocardial damage, whereas insufficient levels may affect vasodilation and blood flow restoration²⁸. The involvement of the TNF signaling pathway in inflammation and apoptosis further emphasizes its importance in AMI, as elevated TNF levels can exacerbate myocardial injury and influence the healing process²⁹. Our findings suggest that the interplay between glucose metabolism and circadian rhythms may contribute to the inflammatory response and cellular stress mechanisms observed in AMI. Understanding the molecular mechanisms underlying these pathways can provide insights into developing novel intervention strategies aimed at enhancing cardiac protection during ischemic events, thereby addressing critical issues in cardiovascular research and clinical practice^{30,31}.

The immune infiltration analysis results from this study revealed significant alterations in the abundance of 17 immune cell types in the AMI group, highlighting the strong correlation between mast cells and neutrophils in the high-risk cohort. The observed increase in mast cell and neutrophil infiltration aligns with previous research, indicating that these cells are crucial mediators of inflammation and tissue repair during cardiac events³². These data indicate that the presence of these immune cells is not merely a byproduct of myocardial injury but may actively participate in disease progression, potentially influencing outcomes and recovery. By elucidating the relationship between immune cell infiltration and AMI severity, this study contributes to the growing body of literature emphasizing the importance of the immune system in cardiovascular diseases, paving the way for novel interventions that target immune pathways to mitigate myocardial damage and enhance recovery³³.

In this study, a risk-scoring system based on GMCRRDEGs demonstrated excellent performance in the diagnosis of AMI. The ROC curve showed that the area under the curve (AUC) of the risk score (RiskScore) was greater than 0.9, indicating a high diagnostic accuracy. This provides a rapid and effective method for assessing patient risk levels in clinical practice and helps optimize individualized management strategies. The risk scoring system can be further expanded and optimized in the following ways in practical applications. First, combining the risk scoring system with traditional clinical indicators (such as electrocardiograms and troponin levels) may further improve prediction accuracy. Although traditional clinical indicators are valuable for the diagnosis of AMI, they often reflect only the local characteristics of the disease³⁴. In contrast, gene expression data can reveal the underlying mechanisms of a disease at the molecular level. Combining these approaches may provide a more comprehensive risk assessment. Future studies can conduct multicenter clinical trials to verify the efficacy of the risk-scoring system in combination with traditional clinical indicators to provide a more precise tool for clinical diagnosis. Second, further optimization of the algorithms to enhance model robustness is an important direction for future research. The current risk-scoring system is based on the LASSO regression and SVM-RFE algorithms, which showed good performance in this study. However, there may still be some limitations in dealing with complex and variable clinical data. For example, the introduction of deep learning algorithms, such as neural networks or convolutional neural networks, can better capture nonlinear features and complex relationships in gene expression data. Additionally, increasing the sample size and data diversity to train and validate the model further is a key step in enhancing its robustness.

In this study, six key genes (*JUN*, *EPAS1*, *IL1B*, *ADRB2*, *FOS*, and *CD36*) were identified using the SVM-RFE algorithm, and the LASSO regression model based on these genes performed well. This model provides an efficient method for the precise prediction of AMI risk and lays the foundation for the development of future clinical decision-support systems. However, to further enhance the practicality and reliability of the model, several aspects need to be explored and optimized. First, validating the model's stability in a larger independent cohort is a crucial step for future research. Although the current study achieved good results with an integrated

	GSE66360	GSE48060
Platform	GPL570	GPL570
Species	Homo sapiens	Homo sapiens
Tissue	Blood Tissue	Blood Tissue
Samples in AMI group	49	31
Samples in Control group	50	21
Reference	PMID: 21,737,128	PMID: 25,684,211

Table 1. GEO microarray chip Information.

dataset, the limitations of the sample size and data sources may have affected the generalizability of the model. Therefore, this model must be validated in multiple independent clinical cohorts to assess its stability and accuracy in different populations and settings. Moreover, through multicenter collaboration, collecting more gene expression data from patients with AMI can further optimize model parameters and improve prediction performance. Second, combining other molecular markers or imaging features to enhance the predictive power is also an important direction for future research. In addition to gene expression data, proteomic, metabolomic, and imaging examinations (such as coronary angiography and echocardiography) provide important information for the diagnosis and risk assessment of AMI. Integrating these multi-omics data and imaging features with the current gene-based model may help to construct a more comprehensive and accurate predictive model, providing stronger support for clinical decision-making.

Limitations.

This study has several limitations. The study primarily relied on publicly available databases, which may not entirely encompass the intricacy of MI. The lack of wet lab experiments hinders the verification of our findings, and the comparatively limited sample sizes may have influenced the strength of the results. Additionally, the potential batch effects of integrating multiple datasets could introduce variability, which could confound the analysis. Future studies should incorporate clinical samples to validate these results. We plan to conduct wet lab experiments to further confirm our findings. Ultimately, this research opens new avenues for innovative diagnostic and therapeutic approaches aimed at treating these alterations in patients with AMI.

Conclusion

This study identified 12 differentially expressed genes related to glucose metabolism and circadian rhythms by integrating GEO datasets and applying various bioinformatic methods to successfully construct an efficient and highly accurate (AUC>0.9) diagnostic model for AMI. These findings not only provide new insights into the pathogenesis of AMI but also lay the foundation for the future development of novel diagnostic and therapeutic strategies. Future work will focus on expanding the sample size, multi-center validation, and exploring more potential molecular markers to achieve broader clinical applications.

Materials and methods

Data download

The two datasets (GSE66360 and GSE48060) were downloaded from the NCBI GEO database using the R package GEOquery³⁵. As detailed in the Introduction (line 84) and summarized in Table 1, both datasets were generated on the GPL570 platform and derived from whole-blood samples of Homo sapiens. GSE66360 comprises 49 AMI cases and 50 controls; GSE48060 comprises 31 AMI cases and 21 controls. All available AMI and control samples were included in the present study.

GEO, gene expression omnibus; AMI, AMI

The GeneCards database³⁶ offers extensive insight into human genes. The search term utilized was “Glucose Metabolism,” and we selected only those genes associated with glucose metabolism that are classified as “Protein Coding” and “Relevance score” > 2. In total, 955 GMRGs were identified. Additionally, a total of 26 GMRGs in the published literature^{37,38} were obtained from PubMed using “Glucose Metabolism” as the key word. A comprehensive total of 973 GMRGs were identified through the processes of consolidation and removal of duplicates from the previously acquired GMRGs. Comprehensive information on these results is provided in supplementary Table S2³⁹.

Similarly, We employed the keyword “circadian rhythm” for our search, subsequently filtering the genes to include only those classified as “Protein Coding” and possessing a relevance score exceeding 2. A total of 701 CRRGs were obtained. Additionally, using “circadian rhythm” as the key word, we obtained a total of 300 CRRGs in the published literature⁴⁰ in PubMed. A total of 836 CRRGs were obtained after combining the CRRGs obtained using the above methods. Comprehensive details can be found in supplementary Table S3.

The R package sva⁴¹ was utilized to remove batch effects from the GSE66360 and GSE48060 datasets, resulting in the formation of integrated datasets. The aggregated datasets included 80 samples originating from the AMI group and 71 samples from the Control group. Subsequently, the combined datasets were normalized using the R package limma⁴², which involved the annotation of probes and various standardization processes. The expression matrices, both before and after eliminating batch effects, underwent PCA⁴³. This analytical approach was employed to convert the data into a reduced-dimensional format, thereby enabling the visualization of these characteristics in either two- or three-dimensional graphical representations. In the subsequent analysis,

the samples from the combined datasets served as the test set and the samples from the GSE48060 dataset functioned as the validation set.

Differentially expressed gene analysis

Based on the sample grouping of the combined datasets, the samples were categorized into AMI and Control groups. Differential analysis was performed on the combined datasets using the limma package in R to examine differences in gene expression levels between the two groups. The R package limma (Version 3.58.1) was used to perform a differential analysis of gene expression between the two groups, and the criteria set for the identification of DEGs comprised a log fold change ($|\log FC|$) threshold exceeding 0.5 and a p -value below 0.05. Genes with $\log FC > 0.5$ and $p < 0.05$, were classified as upregulated DEGs, while those with $\log FC < -0.5$, and $p < 0.05$, were considered downregulated DEGs, and the results of differential analysis were used to draw the volcano plot using the R package ggplot2 (Version 3.4.4).

To obtain GMCRRDEGs associated with AMI, all DEGs were interleaved with GMRGs and CRRGs, and Venn diagrams were drawn to obtain GMCRRDEGs. A heat map was generated for visualization using the R package pheatmap (version 1.0.12). Finally, the chromosomal locations of the GMCRRDEGs were visualized by chromosome mapping using the R package RCircos⁴⁴ (Version 1.2.2).

Differential expression of glucose metabolism and circadian rhythm-related genes (GMCRRDEGs) in the integrated GEO dataset, correlation analysis and ROC curve analysis

To explore variations in GMCRRDEGs expression between the two groups in the combined datasets, a comparative analysis chart was created based on GMCRRDEGs expression levels. The ROC curve for GMCRRDEGs was plotted using the R package pROC (Version 1.18.5), which also facilitated computation of the AUC associated with the ROC analysis. The AUC was employed to evaluate the diagnostic effectiveness of GMCRRDEGs expression for the occurrence of AMI. The AUC values for the ROC curve typically range from 0.5 to 1. An AUC value of 1 indicated a superior diagnostic ability. Particularly, an AUC in the range of 0.5 to 0.7 indicates a low level of accuracy. Conversely, an AUC falling between 0.7 and 0.9 denotes moderate accuracy, while an AUC greater than 0.9 reflects a high degree of accuracy.

To investigate the relationship between GMCRRDEGs, Spearman's algorithm was used to assess the correlation between GMCRRDEGs expression levels within the combined datasets. The results obtained from the correlation analysis were illustrated using the R package pheatmap (version 1.0.12), which facilitated the generation of correlation heat maps. The most relevant Top4 GMCRRDEGs were screened, identified, and a correlation scatter plot was generated using the R package ggplot2 (Version 3.4.4). Correlation coefficients (r values) with absolute values below 0.3 indicated weak or no correlation, those between 0.3 and 0.5 suggested a weak correlation, values ranging from 0.5 to 0.8 denoted a moderate correlation, and coefficients above 0.8 represented a strong correlation.

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

GO analysis⁴⁵ serves as a prevalent approach for conducting large-scale functional enrichment studies, encompassing three main categories: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). KEGG^{46,47} serves as a prominent database that compiles data related to genomes, biological pathways, various diseases, and pharmaceutical agents. We conducted GO and KEGG enrichment analyses of GMCRRDEGs using the R package clusterProfiler⁴⁸. Statistical significance was determined for items that presented a p -value of less than 0.05, alongside a false discovery rate (FDR), commonly known as the q -value, below 0.25.

Construction of diagnostic model for AMI

To develop AMI diagnostic models using the combined datasets, GMCRRDEGs were analyzed using logistic regression. Logistic regression was employed to investigate the relationship between independent and dependent variables when the latter was categorized as a binary variable. The GMCRRDEGs were evaluated using a significance threshold of $p < 0.05$, which led to the development of a logistic regression model. Subsequently, the collective expression of the GMCRRDEGs included in the logistic regression analysis was illustrated using a forest plot.

Subsequently, utilizing the GMCRRDEGs incorporated within the logistic regression model, the support vector machine recursive feature elimination (SVM-RFE)⁴⁹ algorithm was employed, as facilitated by the e1071 package (Version 1.7–14) in R, to identify prospective biomarkers. SVM-RFE relies on the feature selection algorithm of support vector machine (SVM) to select the most important features by recursively eliminating those that contribute the least to the classification.

In conclusion, a LASSO regression analysis was performed using the R package glmnet⁵⁰ (Version 4.1-8). The analysis was configured with parameters including `set.seed(500)` and `family="binomial"`, in accordance with the characteristics identified through the SVM-RFE algorithm. The LASSO regression technique, which is an extension of linear regression, addresses the issue of model overfitting and improves its capacity for generalization by incorporating a penalty term, which is the product of a parameter (λ) and the absolute value of the coefficient. Diagnostic model diagrams and variable trajectory diagrams were used to depict the outcomes of the LASSO regression analysis. These outcomes constituted the diagnostic model for AMI, with GMCRRDEGs incorporated into the model and identified as key genes. Subsequently, the LASSO RiskScore was determined using the risk coefficients obtained from the LASSO regression analysis. The risk scores were calculated using the following formula:

$$RiskScore = \sum_i iCoefficient(gene_i) * mRNAExpression(gene_i)$$

Validation of the diagnostic model for AMI

A Logistic Nomogram⁵¹ is a graphical representation situated within a Cartesian coordinate system that employs a collection of nonintersecting line segments to illustrate the functional associations among several independent variables. To create a nomogram that visually represented the relationships among the critical genes, the R package rms (Version 6.7-1) was employed, leveraging the outcomes derived from the logistic regression analysis. The R package ggDCA⁵² (version 1.1) was used to generate DCA maps based on the key genes identified in the combined datasets.

Moreover, the AMI group was classified into two categories, high- and low-risk, according to the median risk score derived from the AMI diagnostic model. A comparison map was generated to explore the differences in key gene expression between high- and low-risk AMI groups. Finally, the R package pROC⁵³ (Version 1.18.5) was used to plot the ROC curve of the model genes and calculate the AUC value. We evaluated the diagnostic effect of the expression levels of key genes on the occurrence of AMI. The AUC derived from the ROC curve were generally within the range of 0.5–1. An AUC value approaching one reflects superior diagnostic efficacy. Particularly, AUC values ranging from 0.5 to 0.7 are indicative of low diagnostic accuracy, those between 0.7 and 0.9 suggest moderate diagnostic accuracy, and values exceeding 0.9 denote high diagnostic accuracy.

The semantic assessment of GO annotations provides a quantitative framework for evaluating the similarities between genes and genomes, thereby positioning it as an essential element in numerous bioinformatics analytical methodologies. The R package GOSemSim⁵⁴ (Version 2.28.0) was used to assess the functional correlations among key genes. Additionally, functional correlations among these key genes were analyzed by evaluating their functional similarities (friends).

Immune infiltration analysis of the high and low-risk groups of key genes by SsgSEA algorithm

Single-sample gene set enrichment analysis (ssGSEA)⁵⁵ is used to evaluate the presence of specific gene sets in individual biological samples. This method assesses the extent to which various immune cell populations are present in a sample. Different immune cell populations are identified and categorized, including activated CD8 T cells, activated dendritic cells, gamma delta T cells, natural killer cells, and regulatory T cells. Subsequently, ssGSEA generated enrichment scores that reflected the proportional representation of each type of immune cell infiltration across samples. These scores are then compiled into an immune cell infiltration matrix that encompassed data from AMI samples derived from various combined datasets. This approach allows a detailed analysis of immune cell distribution and activity within individual samples, providing insights into the immune landscape of the studied tissues or conditions. We used the R package ggplot2 (Version 3.4.4) map grouping comparison to show the difference in low and high-risk expression by immune cells in the combined datasets of AMI samples between the groups. Subsequently, immune cells that demonstrated significant variance between the two cohorts were selected for in-depth analysis. The Spearman correlation coefficient was used to evaluate the interrelationships among these immune cell types, and the R package pheatmap was used to construct a heatmap that visually depicted the outcomes of the correlation analysis. The AMI cohort was divided into two distinct classifications, namely high- and low-risk, based on the median risk score obtained from the AMI diagnostic framework. To effectively illustrate the correlation analysis results linking key genes with immune cells, a correlation bubble plot was generated using the R package ggplot2 (version 3.4.4).

Statistical analysis

Statistical analyses were performed using R programming language (v4.2.2). Continuous variables exhibiting normal distribution were evaluated for intergroup differences using the independent Student's t-test, whereas non-normally distributed data were subjected to nonparametric analysis using the Mann-Whitney U test (Wilcoxon rank-sum test). Multigroup comparisons were conducted using the Kruskal-Wallis rank-sum method. The correlation patterns between molecular markers were quantified using Spearman's rank-order correlation coefficients. All statistical tests followed two-tailed verification principles, with significance thresholds set at $p < 0.05$, unless stated otherwise.

Data availability

The dataset GSE66360 and GSE48060 for this study can be found in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). All data generated or analysed during this study are included in this published article.

Received: 21 April 2025; Accepted: 1 September 2025

Published online: 26 September 2025

References

- Wang, Z., Zhang, M. & Fu, Y. Downregulated Circulating long non-coding RNA GAS6-AS1 screens and predicts AMI. *Anatol. J. Cardiol.* **27**, 167–172 (2023).
- Stevens, W. et al. Estimating the future burden of cardiovascular disease and the value of lipid and blood pressure control therapies in China. *BMC Health Serv. Res.* **16**, 175 (2016).
- Salari, N. et al. The global prevalence of myocardial infarction: a systematic review and meta-analysis. *BMC Cardiovasc. Disord.* **23**, 206 (2023).
- Młynarska, E. et al. From atherosclerotic plaque to myocardial infarction—the leading cause of coronary artery occlusion. *Int. J. Mol. Sci.* **25**, 7295 (2024).
- Elbadawi, A. et al. Outcomes of AMI in patients with rheumatoid arthritis. *Am. J. Med.* **133**, 1168–1179e4 (2020). EPub.

6. Merecz-Sadowska, A., Sitarek, P., Śliwiński, T. & Zajdel, R. Anti-inflammatory activity of extracts and pure compounds derived from plants via modulation of signaling pathways, especially PI3K/AKT in macrophages. *Int. J. Mol. Sci.* **21**, 9605 (2020).
7. Fu, R. et al. Fasting stress hyperglycemia ratio and in-hospital mortality after AMI in patients with different glucose metabolism status: results from China AMI registry. *Diabetes Res. Clin. Pract.* **196**, 110241 (2023).
8. Zhang, Y. et al. Association between the stress hyperglycemia ratio and severity of coronary artery disease under different glucose metabolic States. *Cardiovasc. Diabetol.* **22**, 29 (2023).
9. Fahed, G. et al. Metabolic syndrome: updates on pathophysiology and management in 2021. *Int. J. Mol. Sci.* **23**, 786 (2022).
10. Hayden, M. R. Overview and new insights into the metabolic syndrome: risk factors and emerging variables in the development of type 2 diabetes and cerebrocardiovascular disease. *Med. (Kaunas)*. **59**, 561 (2023).
11. Rabinovich-Nikitin, I., Kirshenbaum, E. & Kirshenbaum, L. A. Autophagy, clock genes, and cardiovascular disease. *Can. J. Cardiol.* **39**, 1772–1780 (2023).
12. Delisle, B. P. et al. Circadian regulation of cardiac arrhythmias and electrophysiology. *Circ. Res.* **134**, 659–674 (2024).
13. Sartor, F. et al. Circadian clock and hypoxia. *Circ. Res.* **134**, 618–634 (2024).
14. Yu, X., Zhang, X., Bilal, H., Shi, C. & Sun, L. Exploring potential biomarkers for AMI by combining circadian rhythm gene expression and immune cell infiltration. *Sci Rep* **15**, 4012 (2025).
15. Zhao, Y. et al. Disruption of circadian rhythms by shift work exacerbates reperfusion injury in myocardial infarction. *J. Am. Coll. Cardiol.* **79**, 2097–2115 (2022).
16. Muse, E. D. et al. A whole blood molecular signature for AMI. *Sci. Rep.* **7**, 12268 (2017).
17. Suresh, R. et al. Transcriptome from Circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction. *J. Mol. Cell. Cardiol.* **74**, 13–21 (2014).
18. Taguchi, K. et al. Genome-wide gene expression profiling of randall's plaques in calcium oxalate stone formers. *J. Am. Soc. Nephrol.* **28**, 333–347 (2017).
19. Ruan, W. et al. BMAL1-HIF2A heterodimer modulates circadian variations of myocardial injury. *Nature* **641** (8064), 1017–1028 (2025).
20. Martchenko, A. & Brubaker, P. L. Inflaming the clock: effects of Interleukin 1 β on circadian rhythmicity of pancreatic β cells. *Endocrinology* **162** (1), bqaa115 (2021).
21. Janus, P. et al. HSF1 can prevent inflammation following heat shock by inhibiting the excessive activation of the ATF3 and JUN&FOS genes. *Cells* **11** (16), 2510 (2022).
22. Oikonomou, E. et al. Interleukin-1 in coronary artery disease. *Curr. Top. Med. Chem.* **22**, 2368–2389 (2022).
23. De Luca, G. et al. Interleukin-1 and systemic sclerosis: getting to the heart of cardiac involvement. *Front. Immunol.* **12**, 653950 (2021).
24. Huang, J., Kuang, W. & Zhou, Z. IL-1 signaling pathway, an important target for inflammation surrounding in myocardial infarction. *Inflammopharmacology* **32**, 2235–2252 (2024).
25. Karakaslar, E. O. et al. Transcriptional activation of Jun and Fos members of the AP-1 complex is a conserved signature of immune aging that contributes to inflammaging. *Aging Cell.* **22**, e13792 (2023).
26. Andrabi, S. M. et al. Nitric oxide: physiological functions, delivery, and biomedical applications. *Adv. Sci. (Weinh)*. **10**, e2303259 (2023).
27. Chen, P. et al. Efficacy and safety of Panax Notoginseng saponin injection in the treatment of AMI: a systematic review and meta-analysis of randomized controlled trials. *Front. Pharmacol.* **15**, 1353662 (2024).
28. Russo, I., Barale, C., Melchionda, E., Penna, C. & Pagliaro, P. Platelets and cardioprotection: The role of nitric oxide and carbon oxide. *Int. J. Mol. Sci.* **24**, 6107 (2023).
29. Galeone, A., Grano, M. & Brunetti, G. Tumor necrosis factor family members and myocardial ischemia-reperfusion injury: state of the Art and therapeutic implications. *Int. J. Mol. Sci.* **24**, 4606 (2023).
30. Lal, H., Verma, S. K., Wang, Y., Xie, M. & Young, M. E. Circadian rhythms in cardiovascular metabolism. *Circ. Res.* **134**, 635–658 (2024).
31. Tian, H., Zhao, X., Zhang, Y. & Xia, Z. Abnormalities of glucose and lipid metabolism in myocardial ischemia-reperfusion injury. *Biomed. Pharmacother.* **163**, 114827 (2023).
32. Rurik, J. G., Aghajanian, H. & Epstein, J. A. Immune cells and immunotherapy for cardiac injury and repair. *Circ. Res.* **128**, 1766–1779 (2021).
33. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020). <https://pubmed.ncbi.nlm.nih.gov/32442275/>
34. Bock, C. et al. Enhancing the diagnosis of functionally relevant coronary artery disease with machine learning. *Nat. Commun.* **15**, 5034 (2024).
35. Davis, S. & Meltzer, P. S. GE. GEOquery: a Bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* **23**, 1846–1847 (2007).
36. Stelzer, G. et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* **54**, 1.30.1–1.30.33 (2013).
37. Wang, S., Zhang, L., Yu, Z., Chai, K. & Chen, J. Identification of a glucose Metabolism-related signature for prediction of clinical prognosis in clear cell renal cell carcinoma. *J. Cancer.* **11**, 4996–5006 (2020).
38. Zhang, C. et al. A novel glucose metabolism-related gene signature for overall survival prediction in patients with glioblastoma. *BioMed Res. Int.* **2021**, 8872977 (2021).
39. Li, D., Liang, J., Zhang, W., Wu, X. & Fan, J. A distinct glucose metabolism signature of lung adenocarcinoma with prognostic value. *Front. Genet.* **13**, 860677 (2022).
40. He, Y., Chen, Y., Dai, X. & Huang, S. Dysregulation of circadian clock genes associated with tumor immunity and prognosis in patients with colon cancer. *Comput. Math. Methods Med.* **2022**, 4957996 (2022).
41. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
42. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
43. Ben Salem, K. Ben Abdelaziz, A. Principal component analysis (PCA). *Tunis Med.* **99**, 383–389 (2021).
44. Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for circos 2D track plots. *BMC Bioinform.* **14**, 244 (2013).
45. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
46. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
47. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
48. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
49. Sanz, H., Valim, C., Vegas, E., Oller, J. M. & Reverter, F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinform.* **19**, 432 (2018).
50. Engbrechtsen, S. & Bohlin, J. Statistical predictions with Glmnet. *Clin. Epigenet.* **11**, 123 (2019).

51. Wu, J. et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun. (Lond)*. **40**, 301–312 (2020).
52. Van Calster, B. et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur. Urol.* **74**, 796–804 (2018).
53. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).
54. Yu, G. et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976–978 (2010).
55. Xiao, B. et al. Identification and verification of immune-related gene prognostic signature based on SsGSEA for osteosarcoma. *Front. Oncol.* **10**, 607622 (2020).

Author contributions

YW analyzed and interpreted the data and was a major contributor in writing the manuscript. JL and YYL conducted the feasibility research of the data. YW, SL, and CL performed the examination of the study. SL and HJZ made significant contributions during the revision. All authors read and approved the final manuscript.

Funding

Our work was supported by the Medical Science Research Project of Hebei Province [Grants nos. 20251599].

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/s41598-025-18337-8>.

Correspondence and requests for materials should be addressed to H.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025