

Machine-learning–guided transcriptomic integration identifies GFM1 as a lactylation-related candidate biomarker in aortic dissection

Received: 27 October 2025

Accepted: 10 February 2026

Published online: 14 February 2026

Cite this article as: Chen J., Jiang N., Guo Z. *et al.* Machine-learning–guided transcriptomic integration identifies GFM1 as a lactylation-related candidate biomarker in aortic dissection. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-40139-9>

Junquan Chen, Nan Jiang, Zhigang Guo & Yunpeng Bai

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Machine-Learning-Guided Transcriptomic Integration Identifies
GFM1 as a Lactylation-Related Candidate Biomarker in Aortic
Dissection

Junquan Chen^{1,2}, Nan Jiang¹, Zhigang Guo^{1*}, Yunpeng Bai^{1*}.

1 Department of Cardiovascular Surgery, Tianjin Chest Hospital,
Tianjin University, Tianjin 300222, China

2 Tianjin Medical University, Tianjin 300070, China

*Corresponding authors E-mail(s):

Zhigang Guo: zhigangguovip@163.com;

Yunpeng Bai: 644765651@qq.com.

Contributing authors E-mail(s):

Junquan Chen: chenjunquan@tmu.edu.cn;

Nan Jiang: 358028634@qq.com.

Abstract

Background: Aortic dissection (AD) is a life-threatening aortic disease with limited disease-modifying pharmacologic options.

Lysine lactylation is a metabolism-linked post-translational modification implicated in vascular and inflammatory biology, but its relationship to AD has not been well characterized.

Methods: Public AD transcriptomic datasets were integrated for differential expression analysis and WGCNA. Lactylation-related DEGs were defined by intersecting DEGs with a curated lactylation-related gene set. Candidate genes were prioritized using complementary machine-learning models (LASSO, Random Forest, and XGBoost) as a feature-screening strategy with internal resampling and hold-out validation (cross-validation and a hold-out set). GFM1 expression was assessed by qRT-PCR and western blotting in human aortic tissues. Functional relevance was examined in primary mouse aortic vascular smooth muscle cells (VSMCs) using siRNA knockdown under angiotensin II stimulation (1.0 $\mu\text{mol/L}$, 24 h), with proliferation and migration assessed by CCK-8, EdU, Transwell, and scratch-wound assays.

Results: We identified 217 DEGs and an AD-associated co-expression module. Intersection analysis yielded 11 lactylation-related DEGs, among which GFM1 received consistent support across models. GFM1 showed higher expression in AD tissues, and GFM1 knockdown attenuated angiotensin II-induced VSMCs proliferation and migration.

Conclusions: Integrated transcriptomics and machine-learning-based prioritization nominate GFM1 as a lactylation-related candidate associated with AD, warranting further investigation. These findings are hypothesis-generating: model performance reflects internal evaluation only, and independent external validation and direct lactylation profiling are required to establish generalizability and clarify mechanistic links.

Keywords: Aortic dissection; lactylation; integrative transcriptomics; machine learning; GFM1; vascular smooth muscle cells.

1. Introduction

Aortic dissection (AD) is a catastrophic cardiovascular condition with abrupt onset, rapid progression, and high mortality¹. Pathologically, an intimal tear permits blood to enter the media, creating an intramural hematoma and propagating the separation of the intima from the media. The annual incidence is ~1 per 100,000, with a higher prevalence in men than in women². Without timely intervention, many patients die within hours to days of symptom onset, and mortality is estimated to increase by 1%–2% per hour during the first 24 hours³. Despite advances in perioperative care, no approved disease-modifying pharmacologic therapy exists for AD, surgery/endovascular repair remains the mainstay for many patients⁴. However, substantial perioperative risk and postoperative complications can compromise survival and quality of life, while prolonged intensive care and hospitalization impose considerable economic burdens on patients and society⁵⁻⁸. These challenges

underscore the need to identify reliable biomarkers and actionable targets and to generate testable mechanistic hypotheses for future translational studies.

Lysine lactylation is a recently described post-translational modification (PTM) in which lactate-derived acyl groups can be added to lysine residues on proteins⁹. Unlike many classical PTMs, lactylation has been discussed as a potential link between cellular metabolic state and gene regulation, with reported roles in immune modulation, epigenetic remodeling, and stress adaptation^{10,11}. In cancer, for example, lactylation has been implicated in shaping the tumor microenvironment and influencing proliferation, migration, and survival¹². By contrast, its relevance to vascular pathologies—particularly AD—remains poorly defined.

The pathogenesis of AD involves vascular smooth muscle cells (VSMCs) dysfunction, inflammatory activation, and structural remodeling of the aortic wall^{13,14}. In this study, the term

“lactylation-related genes” refers to genes curated from prior literature/databases that are implicated in lactate metabolism and/or lactylation-associated regulatory programs; it does not necessarily indicate that the corresponding proteins are directly lysine-lactylated in AD. Given lactylation’s regulatory roles in immune and metabolic pathways, it is plausible that lactylation-related genes (LRGs) contribute to AD initiation and progression; however, systematic investigations in AD are scarce. Bioinformatics has become integral to discovering disease biomarkers and dissecting molecular networks in complex conditions ^{15,16}. Integrating bioinformatics with machine learning to delineate disease subtypes and prioritize candidate markers represents a powerful interdisciplinary strategy¹⁷. Importantly, our goal here is candidate nomination and hypothesis generation within a lactylation-related transcriptomic framework, rather than lactylation-level mechanistic proof.

In this study, we integrated public transcriptomic datasets with a machine-learning prioritization pipeline to examine lactylation-related transcriptional signatures in AD. By screening a curated lactylation-related gene set within AD expression profiles, we prioritized GFM1 as a candidate associated with AD and conducted preliminary tissue-level and cellular validation. These results provide hypothesis-generating insights and a rationale for subsequent studies, including validation in independent cohorts and direct lactylation profiling, to establish robustness and mechanism.

2. Methods

2.1 Public data download and processing

The overall workflow is shown in Figure 1. Public AD transcriptomic datasets (GSE52093, GSE98770, GSE153434, GSE147026) were downloaded and preprocessed in R (v4.2.1). For microarray datasets, background correction and between-array

normalization were performed using limma's `normalizeBetweenArrays`¹⁸. Expression values were log₂-transformed when required (i.e., when data were not already on a log₂ scale) to improve comparability while preserving relative differences. After mapping probes to official gene symbols and harmonizing identifiers across platforms, the four datasets were merged, yielding a combined matrix of 27 AD and 24 control aortic samples. Batch effects were adjusted using ComBat (sva package) with study/source as the batch variable¹⁹, with AD status included as a covariate in the model matrix to preserve biological signal. Principal component analysis (PCA) before and after ComBat indicated substantial mitigation of batch effects while preserving group-related variability, providing a basis for downstream analyses.

2.2 Identification of Differentially Expressed Genes

Differential expression was assessed using limma with empirical Bayes moderation. Genes were considered significant at $|\log_2FC| >$

0.26 (≈ 1.2 -fold) and $FDR < 0.05$ (Benjamini-Hochberg). Genes with positive \log_2FC were defined as upregulated in AD, and those with negative \log_2FC as downregulated. Results were visualized using volcano plots (ggplot2) and heatmaps (pheatmap).

2.3 Co-expression Network Construction via WGCNA

Genes were variance-filtered (top 60% by variance) prior to WGCNA²⁰. The soft-thresholding power was selected to approximate scale-free topology (scale-free topology fit index $R^2 \approx 0.8$). A signed network was constructed with $power = 8$, $minModuleSize = 30$, and $mergeCutHeight = 0.25$, followed by topological overlap computation and module detection. Module-trait correlations were calculated, and the module showing the strongest association (highest absolute correlation) with AD status was prioritized for downstream analyses.

2.4 Acquisition of lactylation-related differentially expressed genes

A lactylation-related gene set was curated from public resources by querying lactylation- and lactate metabolism-related programs in the Molecular Signatures Database (MSigDB) gene set collections²¹. Intersecting this gene set with differentially expressed genes (DEGs) yielded lactylation-related DEGs (LR-DEGs), and overlaps were displayed using Venn diagrams. Importantly, this gene-set restriction provides a transcriptomic framework and does not constitute direct quantification of protein or histone lactylation. STRING (confidence > 0.75) was used to derive protein-protein interaction networks, and clusterProfiler²² was used for GO/KEGG^{23,24} enrichment analysis with $p.adjust < 0.05$. For visualization, Figure 2F labels $-\log_{10}(P \text{ value})$ as shown, whereas statistical significance was determined using adjusted P values ($p.adjust$) with a cutoff of 0.05.

2.5 Identification of the key gene

Samples were randomly split into stratified training and validation

cohorts (75%/25%) with a fixed random seed (seed = 2024) for reproducibility. Model tuning was performed within the training set using 5-fold cross-validation. LASSO logistic regression was implemented using glmnet (family = binomial, alpha = 1), and candidate genes were defined as those with non-zero coefficients at the λ_{1se} solution (1-SE rule). Random Forest (ntree = 1000) and XGBoost models were trained on the same training set, and feature importance was extracted from each model. Model performance was summarized as internal AUC on the hold-out validation set; no independent external cohort was used for validation in this study. Cross-model consensus was defined a priori using a vote-based overlap criterion across LASSO, Random Forest, and XGBoost: one vote if a gene was selected by LASSO (non-zero coefficient at λ_{1se}), and one vote each if it ranked among the top three ($K = 3$) features by importance in Random Forest and XGBoost. Genes with a consensus score of 3/3 were prioritized for downstream validation.

2.6 Function enrichment analysis

Complementary pathway interrogation was performed using KOBAS 3.0²⁵ and clusterProfiler. Enrichment results were visualized using dotplots, in which the x-axis (and color scale) represents $-\log_{10}(\text{P value})$. Statistical significance was determined using Benjamini-Hochberg adjusted P values ($p.\text{adjust} < 0.05$). Results were plotted with ggplot2.

2.7 Gene set enrichment analysis

GSEA was conducted in clusterProfiler using the MSigDB C2: CP KEGG collection (c2.cp.kegg.v2022.1.Hs.symbols.gmt), reporting NES, nominal P, and FDR q-values; significance was defined as $\text{FDR} < 0.25$. Plots were generated with GseaVis.

2.8 Protein-protein interaction analysis

STRING-derived networks were complemented with GeneMANIA (default settings) to construct broader interaction maps and highlight putative functional neighborhoods.

2.9 Quantitative reverse transcription PCR (qRT-PCR)

Human ascending aortic tissues (AD vs. control) were obtained from Tianjin Chest Hospital after written informed consent and were approved by the Ethics Committee of Tianjin Chest Hospital (Approval No. 2023LW-001). All methods were performed in accordance with the relevant guidelines and regulations and with the Declaration of Helsinki. Human aortic tissues were then processed for RNA extraction and cDNA synthesis. qRT-PCR was performed in technical triplicates using β -actin as the internal reference. Relative expression was calculated by the $2^{-\Delta\Delta Ct}$ method. The sample size for each group is reported in the corresponding figure legends. Primer sequences were as follows:

Gene	Forward	Reverse
Name		

GFM1	ATGGTGAGGTTGGCATGAGA	TGGGCTGATGAGTCTCTGTT
β -Actin	AGAGCTACGAGCTGCCTGAC	AGCACTGTGTTGGCGTACAG

2.10 Western blotting of human aortic tissues

Protein was extracted from snap-frozen human aortic tissues (AD and control) using RIPA buffer with protease/phosphatase inhibitors. Equal amounts of protein were separated by SDS-PAGE and transferred to PVDF membranes. After blocking (5% BSA), membranes were incubated with primary antibodies against GFM1 and β -ACTIN, followed by HRP-conjugated secondary antibodies. Bands were visualized by enhanced chemiluminescence and quantified by ImageJ; GFM1 intensity was normalized to β -ACTIN. Group sizes and replicate numbers are reported in figure legends.

2.11 Mouse aortic VSMCs culture, siRNA transfection targeting GFM1, and functional assays.

Mouse aortic VSMCs were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C in 5% CO₂. Cells were transfected with siRNA targeting GFM1 (siGFM1) or negative-control siRNA

(siNC) using a lipid-based transfection reagent according to the manufacturer's instructions (final siRNA concentration: 50 nM).

After recovery, cells were stimulated with angiotensin II (AngII, 1.0 μ mol/L for 24 h) or vehicle as indicated. Four experimental conditions were used: Control (untreated), Model (AngII only), NC (siNC + AngII), and siGFM1 (siGFM1 + AngII).

siRNA knockdown efficiency: Knockdown efficiency was quantified 24 h after transfection by qRT-PCR ($2^{-\Delta\Delta C_t}$, normalized to β -actin). Quantified knockdown data are provided in Supplementary Figure S1.

CCK-8: Cell viability/relative proliferation was assessed using the CCK-8 kit. For each independent experiment, each condition was plated in technical triplicates (three wells). Cells were exposed to AngII (or vehicle) as specified, and OD450 was recorded at 24 h after AngII treatment. Technical replicates were averaged to yield one value per biological replicate.

EdU: DNA synthesis was measured using an EdU incorporation assay. Cells were incubated with EdU for 2 h, followed by fixation and nuclear counterstaining. For each condition, EdU-positive nuclei were quantified as a percentage of total nuclei from ≥ 5 random fields per well under identical imaging settings.

Transwell migration: Serum-starved cells were seeded into the upper chamber of 8- μm -pore inserts in serum-free medium, with 10% FBS in the lower chamber as chemoattractant. After 24 h, migrated cells were fixed, stained with crystal violet, and counted in ≥ 5 non-overlapping fields per insert.

Scratch (wound-healing): Confluent monolayers were scratched using a 200- μL pipette tip, rinsed to remove debris, and cultured in 1% FBS medium with AngII or vehicle as indicated. Images were captured at 0 h and 12 h under identical settings; wound area was quantified in ImageJ and expressed as percent closure relative to 0 h.

2.12 Statistical analysis

Analyses were performed in R (v4.2.1) and GraphPad Prism (v9). Data are reported as mean \pm SD or median (IQR) and compared using unpaired t tests or Mann-Whitney U tests (two groups), and one-way ANOVA with Tukey's post hoc or Kruskal-Wallis with Dunn's post hoc (≥ 3 groups); categorical data were compared using χ^2 or Fisher's exact tests. For cell assays, statistics were conducted at the biological-replicate level. Two-sided $P < 0.05$ was considered significant. Transcriptome-wide analyses controlled multiplicity using Benjamini-Hochberg FDR. For cell-based assays, biological replicates refer to independent experiments performed on different days ($n = 3$), whereas technical replicates (e.g., multiple wells per condition and multiple imaged fields per well/insert) were averaged within each experiment before statistical testing.

3 Results

3.1 Characterization of Differentially Expressed Genes and

Functional Associations in AD

PCA showed separation between AD and control samples after batch-effect adjustment, consistent with improved comparability in the integrated dataset (Figure 2A–B). Unsupervised hierarchical clustering broadly recapitulated the group structure (Figure 2C). Comparative transcriptomic analysis identified 217 DEGs (117 upregulated and 100 downregulated) at the prespecified thresholds in AD versus controls (Figure 2D). A radial heatmap depicts expression patterns of the top 50 DEGs (Figure 2E). Enrichment analyses suggested involvement of immune/inflammatory and extracellular-matrix remodeling pathways. KEGG enrichment highlighted cytokine–cytokine receptor interaction, PI3K–Akt signaling, and TNF signaling (Figure 2F), while Gene Ontology terms were predominantly related to myeloid leukocyte differentiation and extracellular matrix organization (Figure 2G). In the protein–protein interaction network, COL1A1, MMP9, and POSTN showed relatively

high connectivity, consistent with central roles within the inferred network (Figure 2H-I).

3.2 Co-expression Network Architecture and Disease-relevant Module Detection via WGCNA

WGCNA identified co-expression modules associated with AD.

Sample clustering with trait annotation is shown in Figure 3A. Using a scale-free topology framework with a soft-threshold power of 8, the network achieved a scale-free topology fit index of $R^2 = 0.82$ (slope = -1.72 ; Figure 3B-C). Gene dendrogram and dynamic tree cut identified 25 modules (Figure 3D). Modules were further merged based on eigengene similarity ($\text{mergeCutHeight} = 0.25$; Figure 3E). Module-trait analysis showed that the MEviolet module exhibited the strongest association with AD status ($r = 0.40$; $p = 0.006$; Figure 3F) and was prioritized for downstream analyses. Gene significance patterns across modules are summarized in Figure 3G.

3.3 Lactylation-related DEGs and pathway enrichment

Intersecting DEGs with a curated lactylation-related gene set identified eleven LR-DEGs (Figure 4A). Functional annotation suggested enrichment of hypoxia-responsive signaling (including the HIF-1 pathway) as well as immune- and metabolism-related processes (Figure 4B). Gene Ontology mapping highlighted regulation of the MAPK cascade, extracellular localization, and protein domain interactions (Figure 4C). Within the LR-DEG protein-protein interaction network, GFM1, CIDEA, and DDIT4 exhibited relatively higher degree centrality, indicating that these genes may occupy more connected positions within the inferred network (Figure 4D).

3.4 Identification and Validation of Key Genes Using LASSO, Random Forest, and XGBoost Algorithms

Using the eleven LR-DEGs, LASSO regression identified six candidates with non-zero coefficients at the λ_{1se} solution: LIF, TNC, DDIT4, HMOX1, PLAUR, and GFM1 (Figure 5A-D). Random

Forest and XGBoost ranked genes based on feature-importance measures (Figure 5E-H). We quantitatively defined “cross-model consensus” as a vote-based overlap across the three models: one vote for LASSO selection (non-zero coefficient at λ_{1se}) and one vote each for ranking among the top three features by importance in Random Forest and XGBoost, yielding a consensus score of 0-3. The overlap is summarized in the Venn diagram (Figure 5I). Under this pre-defined criterion, GFM1 was the only gene supported by all three models (score = 3/3) and was therefore prioritized for downstream validation. AUC values reflect internal discrimination only and should be interpreted cautiously in the absence of independent external validation.

3.5 Validation of GFM1 expression

In human aortic tissues, GFM1 mRNA and protein levels were higher in AD than in non-diseased aorta, as assessed by qRT-PCR and western blotting with densitometric normalization (Figure 6A-B;

group sizes are indicated in the figure). In cultured murine aortic VSMCs, angiotensin II (AngII) stimulation was associated with increased proliferation/viability and migration, reflected by higher CCK-8 and EdU readouts as well as increased Transwell migration and wound closure (Figure 6C-F). siRNA-mediated GFM1 knockdown attenuated these AngII-associated responses across assays (Figure 6C-F). Collectively, the tissue and cellular data support an association between elevated GFM1 expression and AD, and suggest that GFM1 may contribute to AngII-associated VSMC proliferative and migratory behavior. Statistical comparisons and exact P values are provided in Figure 6.

4 Discussion

AD is a life-threatening cardiovascular emergency for which effective strategies for prevention and treatment remain limited^{1,5}. Despite progress in understanding its pathophysiology, clinical management is still constrained by the lack of widely validated

biomarkers and actionable therapeutic targets²⁶. Addressing these gaps will require molecular insights that enable earlier diagnosis and targeted intervention. Metabolic dysregulation has been reported in AD, but lactate-associated epigenetic regulation remains insufficiently defined and was not directly quantified in the present work. Here, we integrated public transcriptomic data with complementary machine-learning models (LASSO, random forest, XGBoost) as a prioritization strategy to nominate lactylation-related candidates in AD, and consistently prioritized GFM1 for preliminary validation.

Lactate is increasingly recognized as a signaling metabolite that can modulate gene expression through lactate-associated epigenetic programs, including histone lactylation¹⁰. In AD, VSMC dysfunction and extracellular matrix remodeling contribute to wall vulnerability, and metabolic shifts in VSMCs may increase lactate and be accompanied by changes in histone lactylation marks (e.g.,

H4K12la) and inflammatory transcriptional programs^{27,28}. Consistent with this broader vascular literature, enhanced glycolysis supports inflammatory phenotypes in plaque immune cells^{29,30}, and VSMCs can adopt glycolytic states that promote proliferation and migration³¹. TRAP1-related metabolic reprogramming has also been linked to lactate accumulation, H4K12la, and cellular senescence³². Within this context, we identified eleven lactylation-related DEGs enriched for glucose handling, mitochondrial function, and oxidative stress. However, lactylation itself was not directly quantified in AD tissues or in our VSMC experiments, and we did not determine whether GFM1 undergoes lysine lactylation; therefore, our findings should be interpreted cautiously as an association within a lactylation-related transcriptomic framework rather than direct evidence of GFM1 lactylation.

In contrast to studies that emphasize model development for prediction, we used machine-learning methods as supportive tools

to prioritize biologically meaningful targets rather than to build a clinically deployable diagnostic model³³. This integrative, biology-first strategy was designed to enhance the stability of candidate nomination while avoiding overinterpretation of internal performance metrics. In particular, cross-validated/hold-out AUC values were used only to summarize internal discrimination and were not intended to claim external predictive generalizability. Beyond discovery analyses, orthogonal validation by qPCR and western blotting confirmed concordant upregulation of GFM1 in AD. Experimental perturbation of GFM1 in VSMCs yielded reproducible effects on proliferation and migration, providing biological support for the computational prioritization and motivating targeted mechanistic investigation.

GFM1 encodes mitochondrial translation elongation factor G1, which is essential for mitochondrial protein synthesis and respiratory-chain integrity; pathogenic variants have been

associated with combined oxidative phosphorylation deficiency in humans³⁴. Although GFM1 has not been directly characterized in AD, mitochondrial dysfunction and metabolic remodeling are increasingly implicated in VSMCs phenotypic switching and aortic wall remodeling. AngII can promote mitochondrial ROS generation and redox signaling in vascular cells^{35,36}, and mitochondrial homeostasis programs (e.g., PGC-1 α -related pathways and SIRT3) have been linked to AngII-associated oxidative stress and aortopathy-related phenotypes^{37,38}. In addition, pro-glycolytic signaling such as HIF-1 α in VSMCs has been linked to AngII-driven vascular remodeling, and recent work supports a role for VSMCs metabolic reprogramming in AD pathogenesis^{39,40}. Within this framework, GFM1 may represent a mitochondrial node that influences respiratory efficiency and redox tone, which could secondarily favor glycolytic adaptation and lactate accumulation under stress. This provides a rationale for why a mitochondrial

factor emerged as a cross-model consensus candidate in a lactylation-related transcriptomic analysis. However, we did not directly measure mitochondrial function, metabolic flux, or lactylation; thus, these links should be regarded as testable hypotheses. Future work will combine GFM1 perturbation with mitochondrial functional assays and direct lactylation measurements, and will validate robustness in independent cohorts.

Several limitations should be acknowledged. First, the integrated public cohorts differ in sampling protocols, platforms, and clinical characteristics; residual heterogeneity and unmeasured confounding may influence gene prioritization and effect estimates. Second, tissue validation was performed in a single center with a modest sample size, and independent external validation is still required to establish generalizability and clinical value. Third, functional experiments were limited to an AngII-stimulated murine VSMC model and may not fully recapitulate

human AD biology or contributions of other cell types. Fourth, cross-validation/hold-out AUC summarizes internal discrimination only and does not establish external predictive performance. Most importantly, we did not directly quantify protein or histone lactylation in AD tissues or in GFM1-silenced VSMCs; therefore, our findings should be interpreted as association within a lactylation-related transcriptomic framework rather than mechanistic proof at the lactylation level. Future work will prioritize external validation and direct lactylation profiling in tissues and relevant cellular models, together with in vivo and spatial/single-cell approaches to clarify cell-type-specific mechanisms.

5. Conclusions

This integrative transcriptomic and machine-learning framework nominates GFM1 as a lactylation-related candidate associated with aortic dissection. The current evidence is associative and

preliminary: we did not directly quantify protein or histone lactylation, and causality has not been established. Therefore, independent external validation in well-characterized cohorts, together with direct lactylation profiling and targeted mechanistic studies, will be necessary to confirm robustness and to clarify biological and potential translational relevance.

Figures

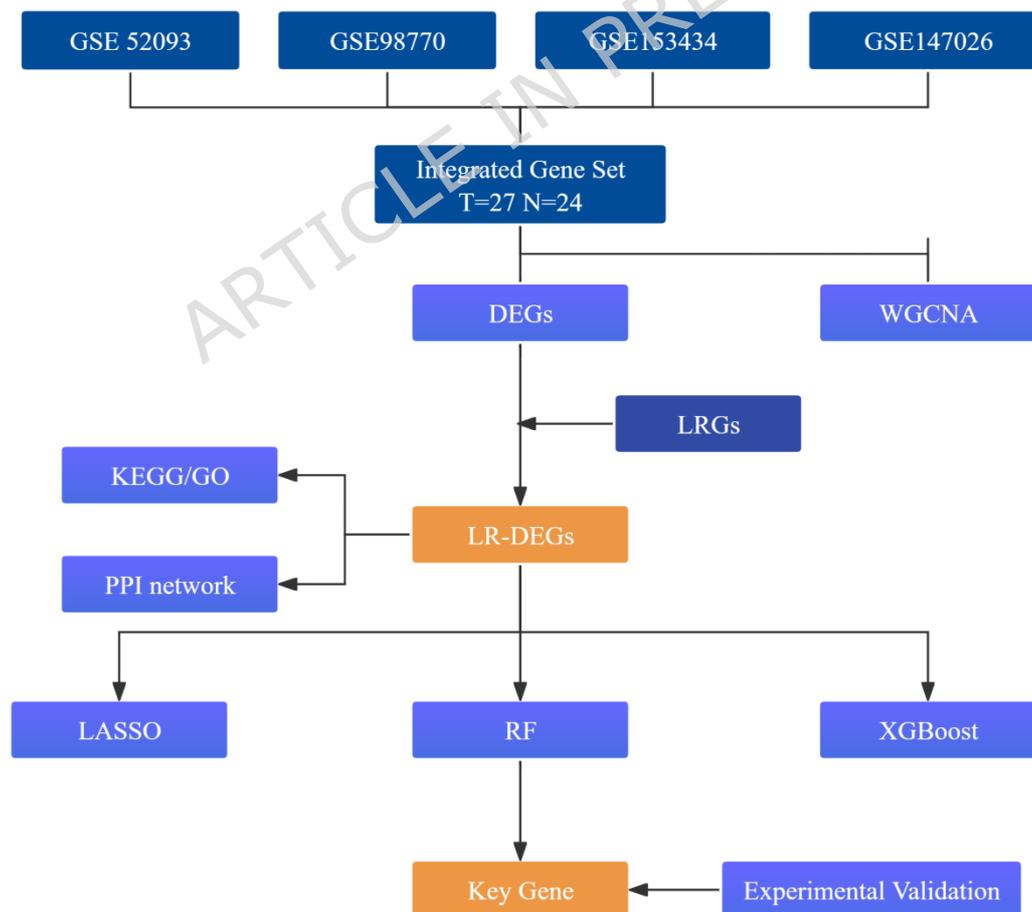


Figure 1 Study design workflow.

Integrated transcriptomic profiles of AD and control aortas were analyzed by differential expression and WGCNA to identify disease-associated genes/modules. Lactylation-related DEGs were defined by intersecting DEGs with a curated lactylation-related gene set, followed by enrichment and PPI analyses. Key candidates were prioritized by machine learning and validated experimentally.

ARTICLE IN PRESS

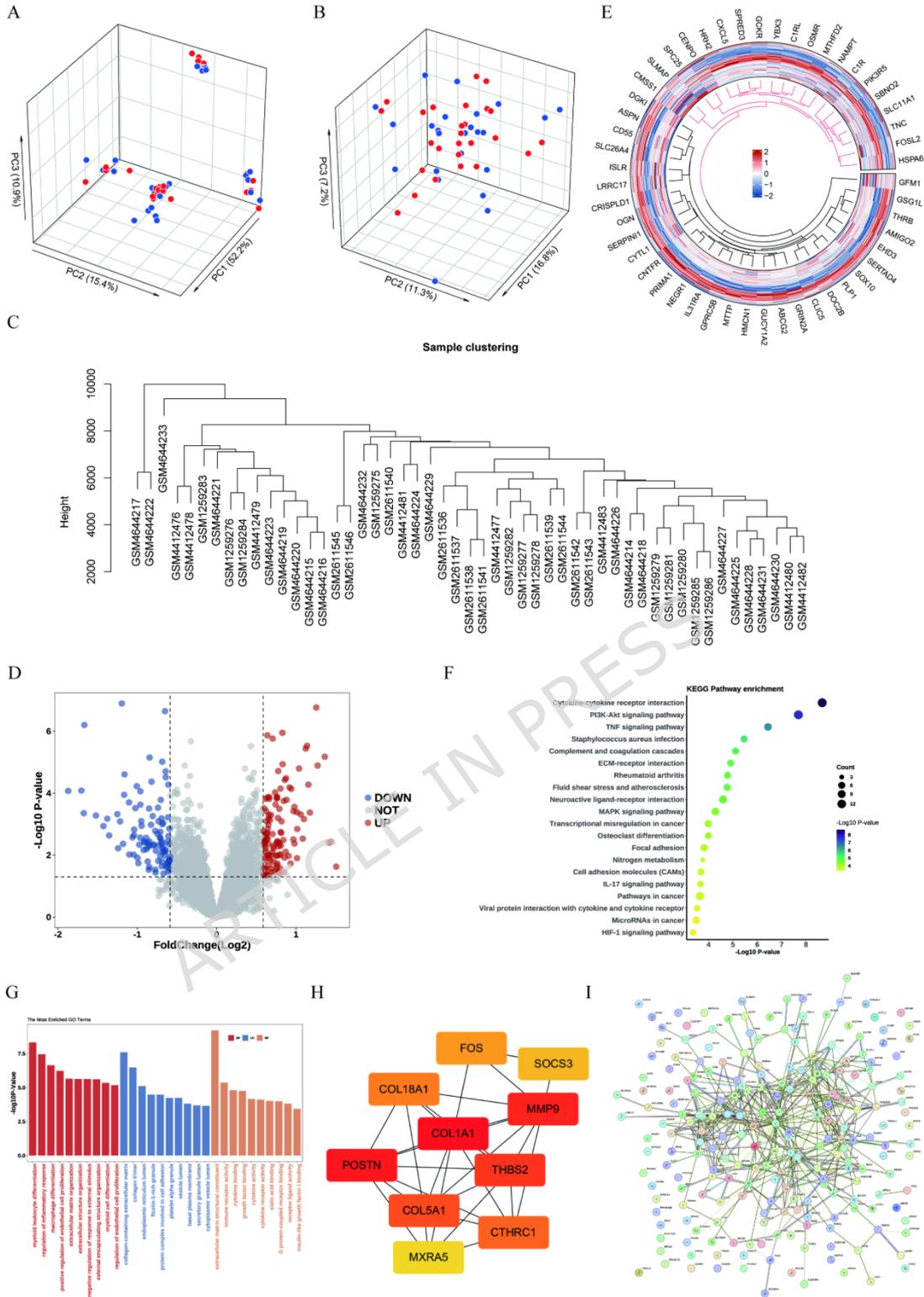


Figure 2 Transcriptomic differences between AD and controls and functional enrichment.

(A) 3D PCA before batch correction; (B) 3D PCA after batch correction (blue, controls, $n = 24$; red, AD, $n = 27$). (C) Hierarchical clustering of samples. (D) Volcano plot of DEGs. (E) Circular heatmap of the top 50 DEGs (row Z-scores). (F) KEGG enrichment bubble plot (x-axis and color: $-\log_{10}(P \text{ value})$; size: Count). (G) GO enrichment bar plot (y-axis: $-\log_{10}(P \text{ value})$) across BP/CC/MF. (H) Hub-gene interaction subnetwork. (I) Global PPI network of DEGs. Enrichment significance was assessed using Benjamini-Hochberg adjusted P values ($p.\text{adjust} < 0.05$), whereas plots display $-\log_{10}(P \text{ value})$ for visualization. KEGG: Kyoto Encyclopedia of Genes and Genomes (see References 23 and 24).

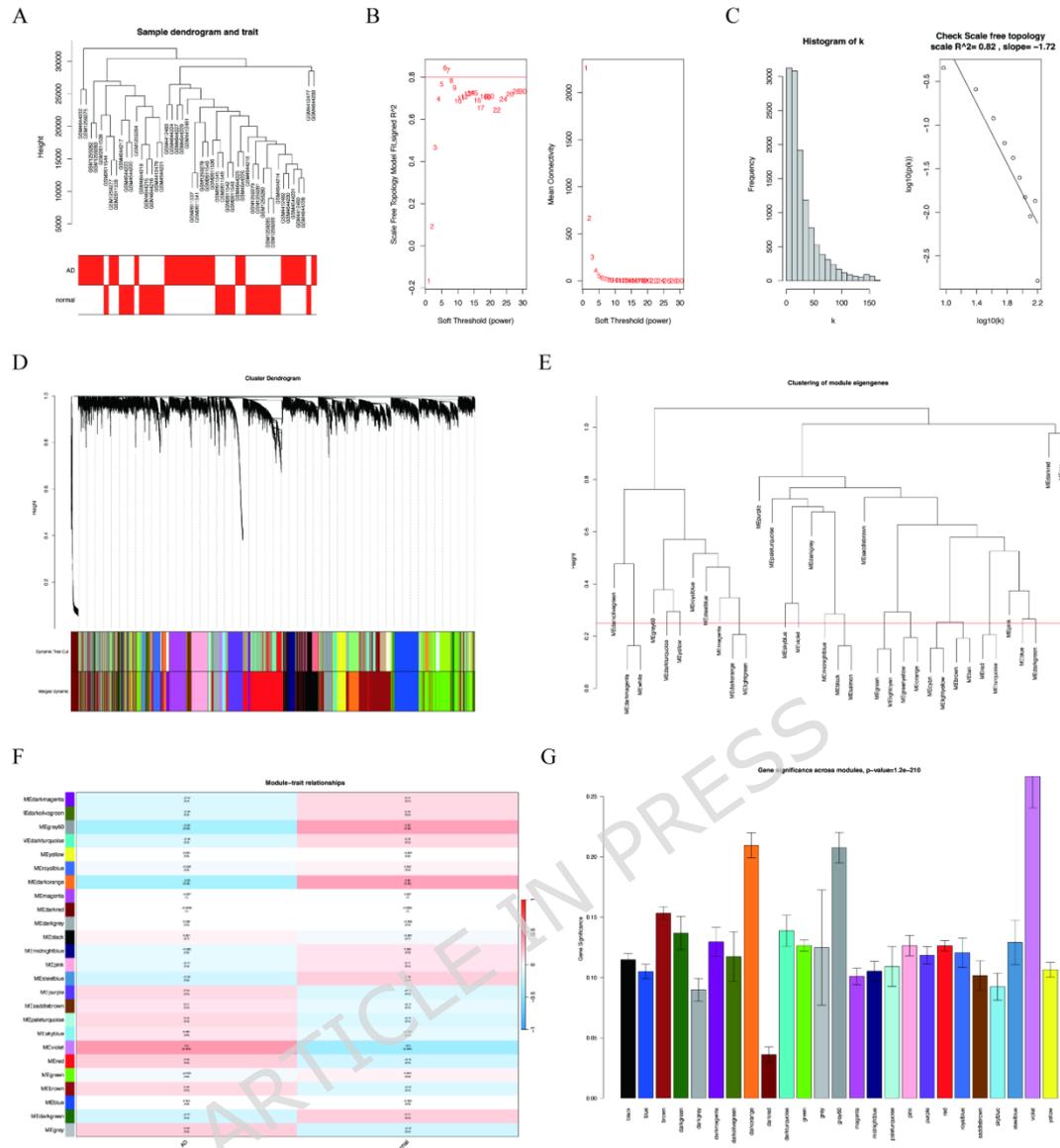


Figure 3 WGCNA network construction and identification of AD-associated modules.

(A) Sample clustering dendrogram of the expression profiles used for WGCNA. The trait heatmap indicates disease status (AD vs control), where red denotes higher trait values and white denotes

lower trait values. (B) Soft-threshold power (β) selection: scale-free topology fit index (left) and mean connectivity (right) across candidate β values. The horizontal line indicates the target scale-free fit; $\beta = 8$ was selected for network construction. (C) Network connectivity distribution (k ; left) and scale-free topology check ($\log_{10}(k)$ versus $\log_{10}(p(k))$; right). (D) Gene dendrogram and module detection. Color bars show modules identified by dynamic tree cut and modules after merging. (E) Clustering of module eigengenes; modules were merged at $\text{mergeCutHeight} = 0.25$ (eigengene correlation > 0.75). (F) Module-trait relationship heatmap showing Pearson correlation coefficients and corresponding P values. (G) Summary of gene significance across modules.

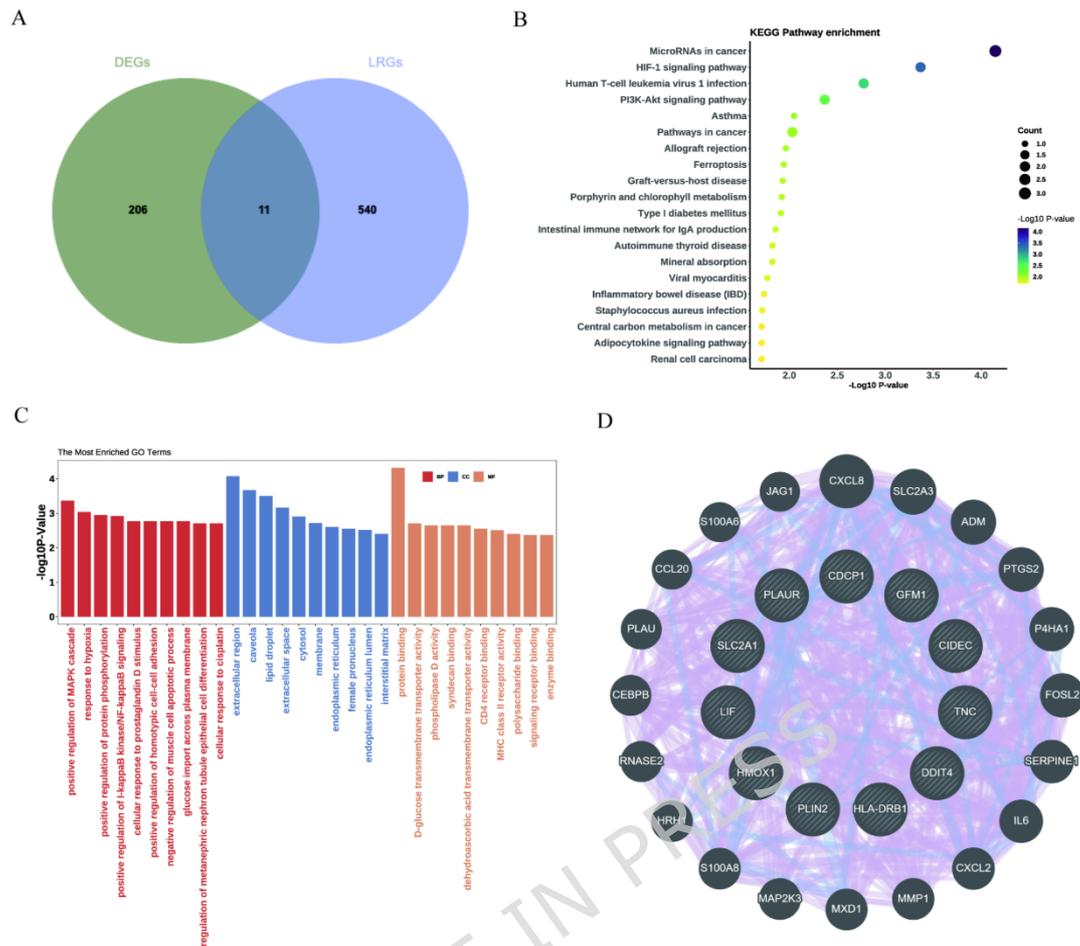


Figure 4 Identification and annotation of lactylation-related DEGs (LR-DEGs).

(A) Venn diagram showing the overlap between DEGs and the lactylation-related gene set, yielding 11 LR-DEGs. (B) KEGG pathway enrichment of LR-DEGs displayed as a bubble plot; the x-axis indicates $-\log_{10}(\text{P value})$, bubble size denotes Count, and bubble color represents $-\log_{10}(\text{P value})$. (C) GO enrichment of LR-DEGs;

bars are grouped by BP/CC/MF as indicated by colors, and the y-axis shows $-\log_{10}(\text{P value})$. (D) Protein-protein interaction (PPI) network of LR-DEGs; nodes represent genes and edges represent predicted interactions. Enrichment significance was assessed using Benjamini-Hochberg adjusted P values ($p.\text{adjust} < 0.05$), whereas plots display $-\log_{10}(\text{P value})$ for visualization. KEGG: Kyoto Encyclopedia of Genes and Genomes (see References 23 and 24).

ARTICLE IN PRESS

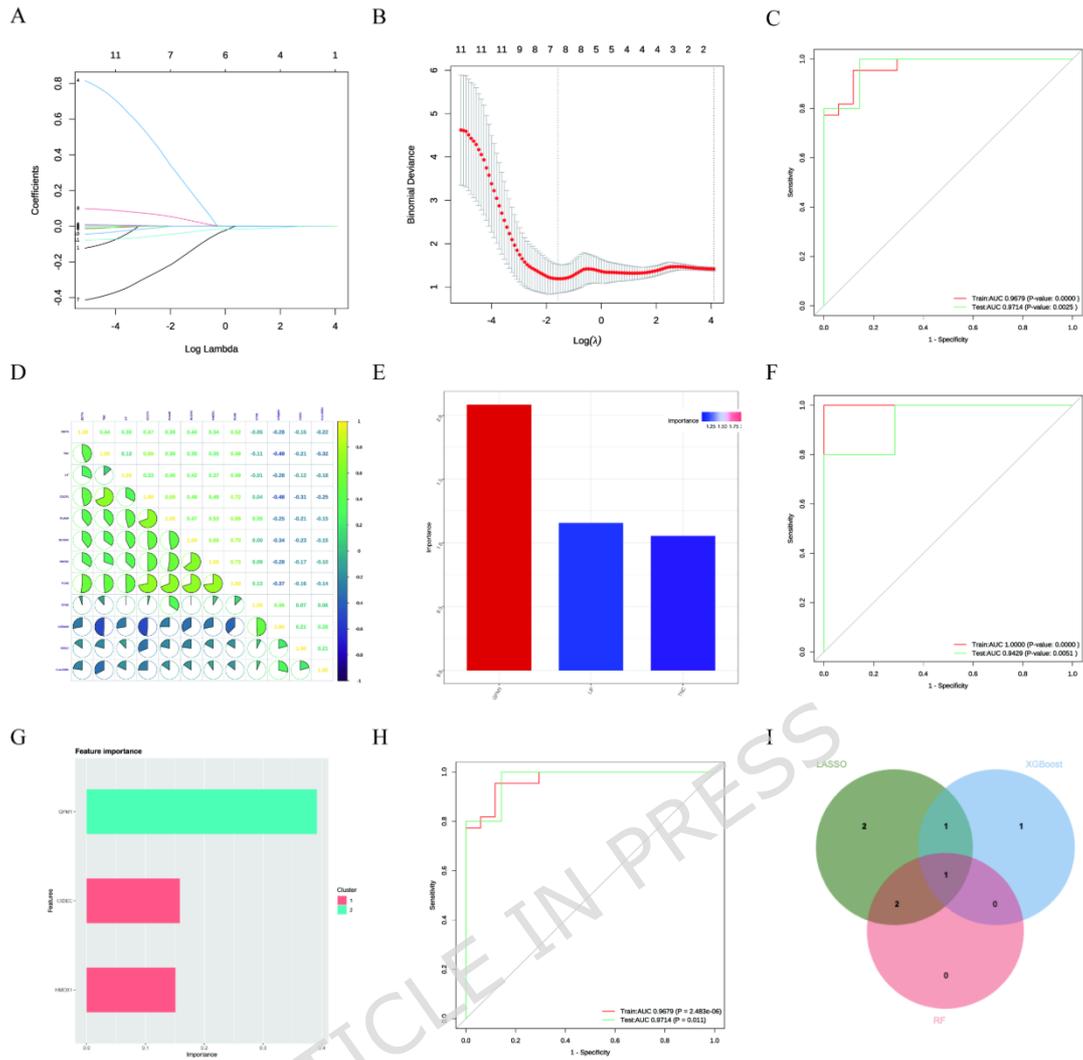


Figure 5 Machine-learning prioritization and model performance.

(A) LASSO coefficient profiles across $\log(\lambda)$. (B) Cross-validation curve for LASSO; the two vertical dotted lines indicate λ_{\min} and λ_{1se} , and λ was selected using the 1-SE rule. (C) ROC curves for the LASSO model (train/test curves with AUC values shown in the panel). (D) Correlation matrix of candidate genes; color indicates

correlation direction and magnitude ($|r|$). (E) Random Forest feature-importance ranking (top features shown). (F) ROC curves for the Random Forest model (train/test metrics shown in the panel). (G) XGBoost feature-importance ranking; features are grouped into clusters (e.g., Cluster 1/2) as labeled. (H) ROC curves for the XGBoost model (train/test metrics shown in the panel). (I) Venn diagram showing the overlap of model-specific candidate sets; “cross-model consensus” was defined as genes supported by all three models (LASSO-selected genes at λ_{1se} and the top three features by importance in Random Forest and XGBoost), and the intersection identifies the final key candidate (GFM1).

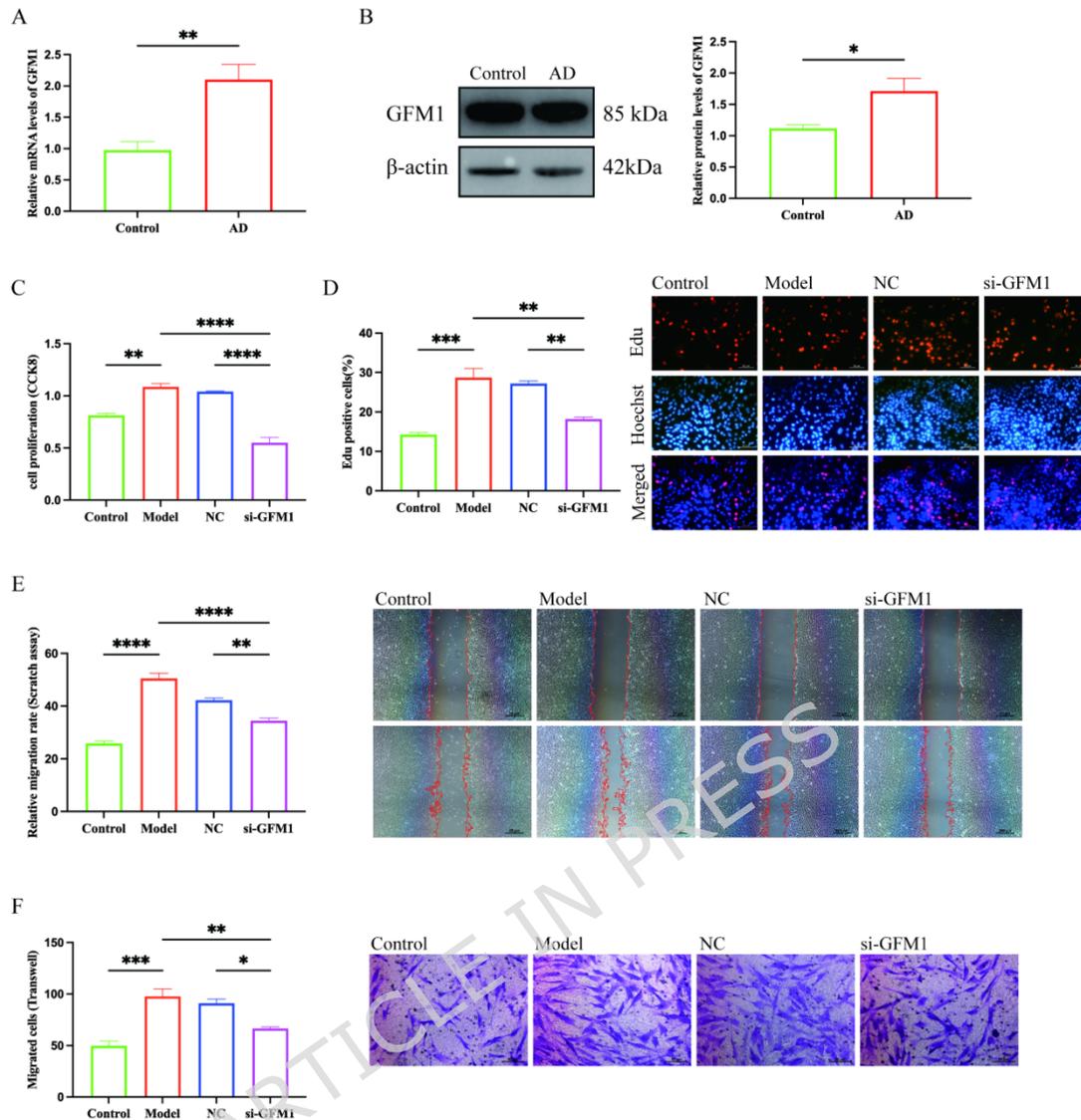
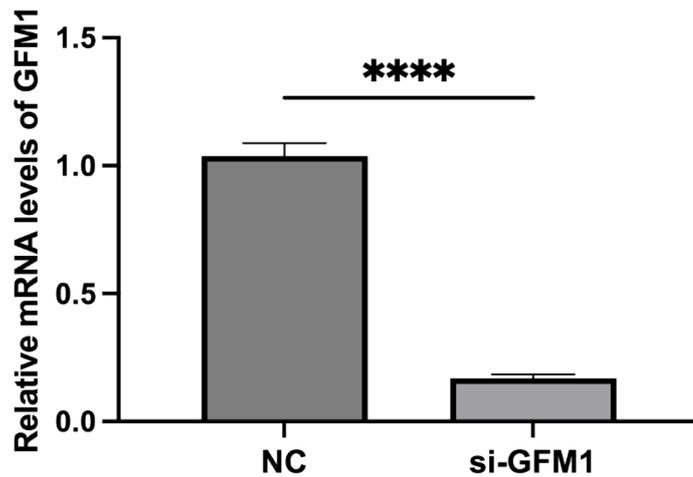


Figure 6. Validation of GFM1 expression and functional assays in mouse aortic VSMCs.

(A) Relative GFM1 mRNA expression in human aortic tissues determined by qRT-PCR (β -actin as the reference gene). (B) Western blot and densitometric quantification of GFM1 protein in human aortic tissues (β -actin as the loading control). (C) CCK-8 assay (24 h)

under four conditions: Control (untreated), Model (AngII 1.0 $\mu\text{mol/L}$ for 24 h), NC (siNC + AngII), and siGFM1 (siGFM1 + AngII). (D) EdU incorporation assay (2 h incubation) with nuclear counterstain (blue) and quantification of EdU-positive cells. (E) Wound-healing assay at 0 h and 12 h with quantified migration/closure. (F) Transwell migration assay (24 h) with crystal violet staining and quantification of migrated cells. Data are presented as mean \pm SD. Two-group comparisons were performed using an unpaired two-tailed Student's t-test or Mann-Whitney U test, as appropriate. Comparisons among ≥ 3 groups were performed using one-way ANOVA with Tukey's post hoc test or Kruskal-Wallis test with Dunn's multiple-comparison test, as appropriate. $P < 0.05$ was considered statistically significant.

Scale bars: 50 μm (EdU, Transwell); 200 μm (wound-healing).



Supplementary Figure S1. qRT-PCR analysis of GFM1 knockdown efficiency in VSMCs 24 h after siRNA transfection (50 nM).

Expression levels were normalized to β -actin and are shown as mean \pm SD (n = 3 independent experiments). Statistical significance was assessed using an unpaired two-tailed Student's t-test. $P < 0.05$ was considered statistically significant.

Declarations

Ethics statement

Clinical trial number: not applicable. The studies involving human participants were reviewed and approved by the Ethics Committee of Tianjin Chest Hospital (approval No. 2023LW-001). The patients/participants provided their written informed consent

to participate in this study. All methods were performed in accordance with the relevant guidelines and regulations and with the Declaration of Helsinki. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article. No live animals were used in this study. Primary mouse aortic VSMCs were commercially obtained. According to the vendor's documentation, these primary cells were procured under protocols approved by the vendor's Institutional Animal Care and Use Committee (IACUC) and in accordance with relevant guidelines. Because the study involved only commercially sourced primary cells and no procedures on live animals at our institution, institutional animal ethics approval was not required.

Data Availability

The integrated transcriptomic data are available from GEO (GSE52093, GSE98770, GSE153434, GSE147026). Full uncropped

gels and blots have been provided as Supplementary File 1. De-identified raw data underlying qPCR/WB and cell-based assays are available from the corresponding author upon reasonable request under an institutional agreement.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Tianjin Key Medical Discipline (Specialty) Construction Project (No. TJYXZDXK-3-030C) and the Tianjin Education Commission Research Program Project (No. 2022YGYB09), and the Tianjin Municipal Science and Technology Program (No. 20JCZDJC00810).

Author Contributions

J.C. conceived the study, designed the analyses, performed data integration and WGCNA/ML modeling, prepared the figures, and drafted the manuscript. N.J. processed samples and performed qRT-

PCR validation, curated clinical/experimental data, revised figures, and contributed to manuscript editing. Z.G. and Y.B. are co-corresponding authors. Z.G. provided clinical resources and supervision, oversaw project administration, and critically revised the manuscript. Y.B. provided senior supervision, secured funding, and critically reviewed and improved the manuscript; Y.B. serves as guarantor of the work. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

We appreciate the surgeons and nurses from the Department of Cardiovascular Surgery for their advice and helpful information.

Thanks to the faculty of the Institute of Cardiovascular Disease.

References

- 1 Bosson, E. & Eagle, K. Epidemiology and management of aortic disease: aortic aneurysms and acute aortic syndromes. *Nature reviews. Cardiology* **18**, 331-348, doi:10.1038/s41569-020-00472-6 (2021).
- 2 Carrel, T., Sundt, T., von Kodolitsch, Y. & Czerny, M. Acute aortic dissection. *Lancet (London, England)* **401**, 773-788, doi:10.1016/s0140-6736(22)01970-5 (2023).

- 3 Rylski, B., Schilling, O. & Czerny, M. Acute aortic dissection: evidence, uncertainties, and future therapies. *European heart journal* **44**, 813-821, doi:10.1093/eurheartj/ehac757 (2023).
- 4 Eric M, I. *et al.* 2022 ACC/AHA Guideline for the Diagnosis and Management of Aortic Disease: A Report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. *Circulation* **146**, doi:10.1161/cir.0000000000001106 (2022).
- 5 Yuanjia, Z. *et al.* Type A Aortic Dissection-Experience Over 5 Decades: JACC Historical Breakthroughs in Perspective. *J Am Coll Cardiol* **76**, doi:10.1016/j.jacc.2020.07.061 (2020).
- 6 Liu, H. *et al.* Inflammatory risk stratification individualizes anti-inflammatory pharmacotherapy for acute type A aortic dissection. *Innovation (Cambridge (Mass.))* **4**, 100448, doi:10.1016/j.xinn.2023.100448 (2023).
- 7 Liu, H. *et al.* Prognostic implication of residual inflammatory trajectories in acute type I aortic dissection: dual-center prospective cohort study. *International journal of surgery (London, England)* **110**, 3346-3356, doi:10.1097/jjs9.0000000000001245 (2024).
- 8 Jaffar-Karballai, M., Tran, T., Oremakinde, O., Zafar, S. & Harky, A. Malperfusion in Acute Type A Aortic Dissection: Management Strategies. *Vascular and endovascular surgery* **55**, 721-729, doi:10.1177/15385744211017116 (2021).
- 9 Chen, A. *et al.* Lactylation, a Novel Metabolic Reprogramming Code: Current Status and Prospects. *Frontiers in immunology* **12**, 688910, doi:10.3389/fimmu.2021.688910 (2021).
- 10 Xiaolu, L. *et al.* Lactate metabolism in human health and disease. *Signal Transduct Target Ther* **7**, doi:10.1038/s41392-022-01151-3 (2022).
- 11 Di, Z. *et al.* Metabolic regulation of gene expression by histone lactylation. *Nature* **574**, doi:10.1038/s41586-019-1678-1 (2019).

- 12 Weihao, L. *et al.* Tumor-derived lactate promotes resistance to bevacizumab treatment by facilitating autophagy enhancer protein RUBCNL expression through histone H3 lysine 18 lactylation (H3K18la) in colorectal cancer. *Autophagy* **20**, doi:10.1080/15548627.2023.2249762 (2023).
- 13 Georgia, S. *et al.* Aortic Wall Inflammation in the Pathogenesis, Diagnosis and Treatment of Aortic Aneurysms. *Inflammation* **45**, doi:10.1007/s10753-022-01626-z (2022).
- 14 Zhen, Y. *et al.* Abdominal Aortic Aneurysm: Roles of Inflammatory Cells. *Front Immunol* **11**, doi:10.3389/fimmu.2020.609161 (2021).
- 15 Ling, C. *et al.* CRISPR-Cas9 screening identified lethal genes enriched in necroptosis pathway and of prognosis significance in osteosarcoma. *J Gene Med* **25**, doi:10.1002/jgm.3563 (2023).
- 16 Sanchita, B. *et al.* ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data* **5**, doi:10.1038/sdata.2018.15 (2018).
- 17 Orlov, Y., Anashkina, A., Klimontov, V. & Baranova, A. Medical Genetics, Genomics and Bioinformatics Aid in Understanding Molecular Mechanisms of Human Diseases. *International journal of molecular sciences* **22**, doi:10.3390/ijms22189962 (2021).
- 18 Gordon K, S. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, doi:10.2202/1544-6115.1027 (2004).
- 19 W Evan, J., Cheng, L. & Ariel, R. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, doi:10.1093/biostatistics/kxj037 (2007).
- 20 Peter, L. & Steve, H. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, doi:10.1186/1471-2105-9-559 (2008).
- 21 Arthur, L. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**,

- doi:10.1016/j.cels.2015.12.004 (2015).
- 22 Guangchuang, Y., Li-Gen, W., Yanyan, H. & Qing-Yu, H. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, doi:10.1089/omi.2011.0118 (2012).
- 23 M, K. & S, G. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, doi:10.1093/nar/28.1.27 (1999).
- 24 Minoru, K., Miho, F., Yoko, S., Mari, I.-W. & Mao, T. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* **49**, doi:10.1093/nar/gkaa970 (2020).
- 25 Dechao, B. *et al.* KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res* **49**, doi:10.1093/nar/gkab447 (2021).
- 26 Hong, L., Si-Chong, Q., Yong-Feng, S., Hai-Yang, L. & Hong-Jia, Z. Prognostic Impact of Systemic Coagulation-Inflammation Index in Acute Type A Aortic Dissection Surgery. *JACC Asia* **2**, doi:10.1016/j.jacasi.2022.06.007 (2022).
- 27 Jorge, O. *et al.* Extracellular Tuning of Mitochondrial Respiration Leads to Aortic Aneurysm. *Circulation* **143**, doi:10.1161/circulationaha.120.051171 (2021).
- 28 Ahmad A, C. *et al.* The pyruvate-lactate axis modulates cardiac hypertrophy and heart failure. *Cell Metab* **33**, doi:10.1016/j.cmet.2020.12.003 (2020).
- 29 Bo, Y. *et al.* Macrophage polarisation and inflammatory mechanisms in atherosclerosis: Implications for prevention and treatment. *Heliyon* **10**, doi:10.1016/j.heliyon.2024.e32073 (2024).
- 30 Guo, Z., Guan, Z. & Zhang, N. Role of vascular endothelial dysfunction and metabolic eprogramming of immune cells in the development and ogression of atherosclerosis. *Front Physiol* **16**, doi:10.3389/fphys.2025.1624225 (2025).
- 31 Ankan, S., Sandip V, P., Kanwaljit, C. & Manish, J. Gamut of

- glycolytic enzymes in vascular smooth muscle cell proliferation: Implications for vascular proliferative diseases. *Biochim Biophys Acta Mol Basis Dis* **1870**, doi:10.1016/j.bbadis.2024.167021 (2024).
- 32 Xuesong, L. *et al.* TRAP1 drives smooth muscle cell senescence and promotes atherosclerosis via HDAC3-primed histone H4 lysine 12 lactylation. *Eur Heart J* **45**, doi:10.1093/eurheartj/ehae379 (2024).
- 33 Li, B. *et al.* Identification of immune microenvironment subtypes and clinical risk biomarkers for osteoarthritis based on a machine learning model. *Frontiers in molecular biosciences* **11**, 1376793, doi:10.3389/fmolb.2024.1376793 (2024).
- 34 Shen, Y., Yan, K., Dong, M., Yang, R. & Huang, X. GFM1[Analysis of gene mutations in a family with combined oxidative phosphorylation deficiency 1]. *Zhejiang da xue xue bao. Yi xue ban = Journal of Zhejiang University. Medical sciences* **49**, 574-580, doi:10.3785/j.issn.1008-9292.2020.10.04 (2020).
- 35 Molina-Berenguer, M. *et al.* Dysfunctional mitochondrial translation and combined oxidative phosphorylation deficiency in a mouse model of hepatoencephalopathy due to Gfm1 mutations. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **36**, e22091, doi:10.1096/fj.202100819RRR (2022).
- 36 Sergey I, D. & Rafal R, N. Angiotensin II-induced production of mitochondrial reactive oxygen species: potential mechanisms and relevance for cardiovascular disease. *Antioxid Redox Signal* **19**, doi:10.1089/ars.2012.4604 (2012).
- 37 Qingbin, Z., Junfang, Z. & Huifang, W. PGC-1 α limits angiotensin II-induced rat vascular smooth muscle cells proliferation via attenuating NOX1-mediated generation of reactive oxygen species. *Biosci Rep* **35**, doi:10.1042/bsr20150112 (2015).
- 38 Lin, Q., Shaolei, Y., Tingting, Y. & Yan, H. Sirt3 Protects Against

- Thoracic Aortic Dissection Formation by Reducing Reactive Oxygen Species, Vascular Inflammation, and Apoptosis of Smooth Muscle Cells. *Front Cardiovasc Med* **8**, doi:10.3389/fcvm.2021.675647 (2021).
- 39 Dan, Q. *et al.* Hypoxia inducible factor 1 α in vascular smooth muscle cells promotes angiotensin II-induced vascular remodeling via activation of CCL7-mediated macrophage recruitment. *Cell Death Dis* **10**, doi:10.1038/s41419-019-1757-0 (2019).
- 40 Jizhong, W. *et al.* GMRSP encoded by lncRNA H19 regulates metabolic reprogramming and alleviates aortic dissection. *Nat Commun* **16**, doi:10.1038/s41467-025-57011-5 (2025).