ORIGINAL RESEARCH

Identification of Biomarkers Associated with Oxidative Stress in Aortic Dissection Based on Bulk Transcriptome Analyses

Zhenghao Li *, Changying Li*, Yue Shao , Haoyu Ran, Haoming Shi, Ruiqin Zhou, Xuanyu Liu, Qingchen Wu, Cheng Zhang

Department of Cardiothoracic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China

*These authors contributed equally to this work

Correspondence: Cheng Zhang, Department of Cardiothoracic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China, Email zhangchengcqmu@126.com

Purpose: The aim of this study is to investigate the underlying molecular mechanism of oxidative stress (OS) involved in aortic dissection (AD).

Methods: Datasets of AD and OS-related genes were obtained from the Gene Expression Omnibus (GEO) and the GeneCards database, respectively. Differential expression analysis and weighted gene correlation network analysis (WGCNA) were employed to screen genes. After enrichment analysis, a protein–protein interaction (PPI) network was constructed, and machine learning algorithms were used to determine signature genes. Comprehensive bioinformatics analyses on the signature genes were executed, and a clinical prediction model was established and evaluated. External datasets, in vitro experiment, and Mendelian randomization (MR) analysis were applied to validation.

Results: We identified CCL2, ITGB4, MYC, SOCS3, SPP1 and TEK as OS-related signature genes in AD. The area under the ROC curve of all the signature genes was greater than 0.75. The clinical prediction model based on the signature genes showed satisfactory diagnostic efficacy in both training and validation cohorts. In validation cohort and in vitro experiment, CCL2, MYC, SPP1 and TEK were further validated. However, the MR results showed no causal association between the expression of the signature genes and AD.

Conclusion: This study demonstrated that OS participates in and affects the progression of AD. Six biomarkers associated with OS could be perceived as crucial targets for the diagnosis and treatment of AD.

Keywords: bioinformatics analysis, machine learning, Mendelian randomization, expression quantitative trait locus, aortic dissection

Introduction

Aortic Dissection (AD) is a catastrophic condition caused by the division of the layers comprising the aortic wall. Unrelenting exposure to elevated pulsatile pressure and stress from blood flow can attenuate the structural integrity of the aortic wall, causing a tear in the inner layer and the development of a false lumen.[1] The annual incidence of AD is reported to be about 2.8–6.0 cases per 100,000 people.[2] Nevertheless, only 15%–43% of confirmed acute AD cases are precisely suspected, and a notable portion of AD patients are still overlooked in the emergency department.[3,4] Therefore, early detection and intervention for AD patients is of paramount importance to enhance survival rate.

Oxidative stress (OS) refers to the imbalance of oxidation-antioxidation homeostasis owing to an excess of the free radicals like reactive oxygen species (ROS), which can induce damage to cells and tissues. Recent studies have confirmed that through promoting vascular smooth muscle cell (VSMC) phenotypic switching and damage, increasing expression of matrix metalloproteinases (MMPs), inducing extracellular matrix (ECM) degradation, and mediating other pathophysiological disorders, OS advances the formation and development of AD.[5] However, the molecular mechanism

of OS involved in the onset and progression of AD has been scarcely explored, and the potential regulatory targets are still undiscovered.

With the prosperity of bioinformatics, especially the advances in multi-omics analysis such as genomics and transcriptomics based on high-throughput sequencing and the exponential growth of big data such as genome-wide association studies (GWAS) data in public databases,[6] researchers can have a different perception of diseases from a novel dimension, such as exploring the molecular markers. Mendelian randomization (MR) analysis enhances causal deduction between exposures and outcomes from observational studies by utilizing single nucleotide polymorphisms (SNPs), which are the most common genetic variants, as instrumental variables (IVs). Large-scale molecular quantitative trait locus (QTL) and GWAS data enable us to investigate the latent association between genes and traits.

The objective of this study was to furnish a theoretical underpinning for a deeper comprehension of the impact of OS on the pathophysiological process of AD. In this study, we performed machine learning algorithms to identify the signature genes, depicted clinical prediction nomograms, and substantiated our findings by external datasets, in vitro experiments, and MR analysis.

Materials and Methods

Data Acquisition and Screening of Differentially Expressed Genes (DEGs)

The data from studies of AD in the Gene Expression Omnibus (GEO) database[7] were obtained by R package "GEOquery". The six included expression profiling series were split into training and validation cohorts (Table 1), and all samples of the included expression profiling series were from ascending aorta tissue. The batch effects were adjusted by the "ComBat" function of the R package "sva" when consolidating all the training cohorts. Differential expression analysis was performed using the R package "limma", applying the threshold of adjusted p-value <0.05 and |log2FC| > 1. "Oxidative Stress" was searched in the GeneCards database[8] to attain oxidative stress-related genes (OSRGs), and genes with relevance score greater than the average were included (Supplementary File 1, Table S2).

Weighted Gene Correlation Network Analysis (WGCNA)

R package "WGCNA" was used to construct and analyze correlation networks. To reduce code load and systematic error, we initially selected the top 5000 genes from all genes with the smallest median absolute deviation in the combined gene expression data. Then, based on the scale-free topology criteria, the appropriate soft-threshold for network construction was calculated by "pickSoftThreshold" function. The "blockwiseModules" function was executed to perform automatic network construction, gene clustering, and module detection according to the optimal soft-threshold and dynamic cut tree method, using the following chief parameters: networkType = "unsigned", TOMType = "unsigned", mergeCutHeight = 0.15, minModuleSize = 20. After importing clinical traits into the network, namely the presence or absence of AD, the

Table 1 Information About the Research of AD in the GEO Database

Datasets ID	Sample Size of AD	Sample Size of Normal	Experiment Type	Data Type
GSE98770	6	5	Expression profiling	Training cohort
GSE147026	4	4	Expression profiling	Training cohort
GSE153434	10	10	Expression profiling	Training cohort
GSE190635	4	4	Expression profiling	Training cohort
GSE52093	7	5	Expression profiling	Validation cohort
GSE98770	6	5	miRNA expression profiling	/
GSE107844	3	3	lncRNA expression profiling	/

Abbreviations: AD, aortic dissection; GEO, Gene Expression Omnibus.

Pearson correlation between modules and clinical traits was calculated. Finally, "networkScreening" function was used to identify genes highly related to AD with a screening criterion of weighted p-value < 0.05.

Functional Enrichment Analysis

Gene ontology (GO),[9] which is composed of biological process (BP), cellular component (CC), and molecular function (MF), and Kyoto Encyclopedia of Genes and Genomes (KEGG)[10] pathway enrichment analysis of the overlapped genes of DEGs, OSRGs, and WGCNA genes were performed by R package "clusterProfiler". Enrichment terms with false discovery rate (FDR) < 0.05 were considered statistically significant.

Protein–Protein Interaction (PPI) Network Construction and Hub Gene Computation

The PPI network of the overlapped genes was analyzed using the STRING database[11] with a minimum required interaction score of 0.7 (high confidence). The network was visualized using Cytoscape software (version 3.10.1),[12] and the top 15 hub genes were computed using the maximal clique centrality (MCC) topological algorithm of the "cytoHubba" plugin.[13]

Calculation of Signature Genes by Machine Learning Algorithms

Based on the hub genes, three supervised classification machine learning algorithms, including least absolute shrinkage and selection operator (LASSO), random forest (RF), and support vector machine and recursive feature elimination (SVM-RFE), were implemented to calculate the signature genes. The LASSO algorithm selects the minimum lambda value to determine the optimal variables. The RF algorithm screens the top variables based on the lowest model error rate and the highest variable importance. The SVM-REF algorithm explores the best variables by removing the feature vectors produced by SVM.[14]

Receiver Operating Characteristic (ROC) Curve Analysis and Gene Set Variation Analysis (GSVA)

The ROC curve was used to evaluate the diagnostic value of individual genes using the R package "pROC". The signature genes were initially divided into up- and down-regulated gene sets. According to the single-sample gene set enrichment analysis (ssGSEA) algorithm and the formula "$NES_{total} = NES_{up} - NES_{down}$", the total normalized enrichment score (NES) of each sample in the combined training cohort was calculated. The samples were equally divided into low- and high-NES_{total} groups. The R packages "GSVA" and "limma" were used for performing GSVA. The gene sets involved in the analysis were derived from C5 (ontology gene sets) and C2 (curated gene sets) in the Molecular Signatures Database (MSigDB).[15] The terms with adjusted p-value <0.05 and |log2FC| > 0.58 were considered significantly variant.

Construction of Molecular Networks

Transcription factors (TFs) regulate the expression of corresponding genes by recognizing specific cis-acting elements, such as promoters and enhancers, to form dynamic transcription complexes with ribonuclease activity. The common TFs of the signature genes were retrieved from the TRRUST database.[16] The binding sites of the most significant TF to the promoter region of the signature genes were forecasted in the JASPAR database,[17] with the promoter region being specified from 2.0 kb before to 0.1 kb after of the transcription start site.

To identify the drugs corresponding to the signature genes that serve as potential therapeutic targets, the DGIdb database[18] was utilized to mine and predict the possible drugs for the signature genes.

The competitive endogenous RNA (ceRNA) network was constructed to disclose the possible intrinsic mechanisms of interaction between RNAs, including mRNA and non-coding RNA (ncRNA). The differentially expressed microRNAs (DEmiRNAs) in GSE98770 and long non-coding RNAs (DElncRNAs) in GSE107844 were screened by R package "limma". The miRNA–mRNA interactions were predicted by exploring the miRDB,[19] miRWalk,[20] TarBase,[21] and

TargetScan[22] databases. The miRNA–lncRNA interactions were predicted by exploring the miRCode[23] and LncBase[24] databases.

The TF-gene interaction and drug–gene interaction networks were visualized by Cytoscape software, while ceRNA network was visualized by R package "ggalluvial".

Establishment and Assessment of Clinical Prediction Model

The signature genes were incorporated to establish the clinical prediction model. A nomogram was exhibited based on the model to predict the risk of AD. The diagnostic accuracy and predictive capacity of the model were evaluated by concordance index (C-index), calibration curve, Hosmer-Lemeshow Goodness-of-Fit Test, and decision curve analysis (DCA). R packages "rms", "ResourceSelection", "rmda", and "regplot" were used to perform all the procedures above.

Validation in External Dataset

The expression level and the predictive precision of individual signature genes and the capability of the clinical prediction model to distinguish between AD patients and healthy people were verified in validation cohort GSE52093.

Clinical Aortic Specimens Collection

Approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University, this study stringently followed the guidelines stated in the Helsinki Declaration. Patients with recognized connective tissue diseases, bicuspid aortic valve, aortic coarctation, Takayasu's arteritis, or coronary artery disease were excluded. Eighteen patients with acute AD were enrolled. Additionally, 18 aortic specimens for the control group were acquired from healthy donors. Written informed consent was obtained from all participating individuals or their relatives. The incorporated specimens were swiftly cryopreserved after acquisition and used for subsequent in vitro experiments.

Total RNA Extraction and Quantitative Real-Time Reverse Transcription Polymerase Chain Reaction (RT-qPCR)

The customized primers were designed and purchased from Beijing Tsingke Biotech Co., Ltd. Table 2 contains the primer sequences of the signature genes. The aortic tissues were first ground using a tissue grinder. Afterwards, Trizol solution (TaKaRa, Japan) was used to extract total RNA, followed by quantification of RNA concentration in each sample using a nucleic acid concentration detector. The reverse transcription kit Prime Script™RT (TaKaRa, Japan) was used to compound the reverse transcription reaction system and synthesize cDNA, and qPCR was performed using qPCR kit SYBR ® Premix Ex Taq™ (TaKaRa, Japan).

Table 2 The Primer Sequences of the Signature Genes for RT-qPCR

Gene	Forward Primer (5'→3')	Reverse Primer (5'→3')
CCL2	TCGCGAGCTATAGAAGAATCACC	GAATCCTGAACCCACTTCTGCTT
ITGB4	CTCACCAACCTGTACCCGTATTG	TCATCGTTGACCAGGCCATAG
MYC	CAAGAGGCGAACACACAACG	GTCGTTTCCGCAACAAGTCC
SOCS3	TCCAAACAGGGGACACTTCG	GGGGGTGTGACCATTTCCTT
SPP1	CATATGATGGCCGAGGTGATAGT	CTTTCCATGTGTGAGGTGATGTC
TEK	CGGCCAGGTATATAGGAGGAAAC	TTCTCACACGTCCTTCCCATAAA

Abbreviations: RT-qPCR, quantitative real-time reverse transcription polymerase chain reaction; CCL2, C-C motif chemokine ligand 2; ITGB4, integrin subunit beta 4; SOCS3, suppressor of cytokine signaling 3; SPP1, secreted phosphoprotein 1.

Mendelian Randomization (MR)

Expression quantitative trait locus (eQTL) disclosed the associations of SNPs with the gene expression levels. The eQTL data was retrieved from the eQTLGen Consortium[25] and Consortium for the Architecture of Gene Expression (CAGE).[26] The common (effect allele frequency [EAF] > 0.01) cis-eQTL significantly ($p < 5.0 \times 10^{-7}$) associated with the expression level of the signature genes was identified. SNPs were permitted to be in moderate linkage disequilibrium (LD) ($r^2 < 0.01$) to reinforce the strength of IVs. The summary-level data of AD was retrieved from the R10 data of the FinnGen study[27] (Details are presented in Table 3).

The causal effect of the expression levels of the signature genes on AD was estimated using two-sample Mendelian randomization (TSMR) and summary-data-based Mendelian randomization (SMR).[28] During TSMR analysis, the inverse variance weighted (IVW) method was predominantly employed, while the Wald ratio method was employed when only one SNP was used.

Sensitivity tests, including tests for heterogeneity and horizontal pleiotropy, were applied to evaluate the variability and trustworthiness of the MR results, the specific methods included Cochran's Q test, MR-Egger intercept analysis and heterogeneity in dependent instruments (HEIDI) test.

Statistical Analysis

All statistical analyses were conducted using R software (version 4.3.3) and RStudio (version 2023.12.1+402). Student's *t*-Test was used to analyze the comparisons between groups of continuous variables that were normally distributed, and the Mann–Whitney test was used for non-normally distributed variables. A bilateral $p < 0.05$ was considered statistically significant. All available database websites were displayed in Supplementary File 1, Table S1.

Results

Adjusting Batch Effects and Screening of DEGs

The main procedure of this study was illustrated in Figure S1. Boxplots of the GEO expression datasets and principal component analysis (PCA) plots of the samples before and after adjusting batch effects in the combined training cohorts were displayed in Figure S2.

We screened 308 DEGs between AD and normal samples, in which 170 genes are up-regulated and 138 genes are down-regulated (Figure 1A and Supplementary File 1, Table S3). Figure 1B portrays the expression clustering heatmap of the top 15 differentially expressed up- and down-regulated genes.

The gene correlation network was constructed using the optimal soft-threshold of 13 to fulfill the standard of scale-free topology. The top 5000 genes arranged by the median absolute deviation method were clustered into 9 color modules by the dynamic cut tree method (Figure 1C). The correlation heatmap between clinical traits and color modules indicated that the pink module was the most relevant to Normal (Pearson correlation 0.48, p-value 6.0×10^{-4}), while the green module was the most relevant to AD (Pearson correlation 0.59, p-value 1.0×10^{-5}) (Figure 1D). Ultimately, 1421 genes were screened to be highly correlated with AD (Supplementary File 1, Table S4).

Table 3 Information of eQTL and GWAS Data

Phenotype	Resource	Tissue Type	Sample Size	Population
eQTL	eQTLGen Consortium	Blood	31,684	Principally European
	CAGE	Blood	2765	Principally European
AD	FinnGen Consortium (R10)	/	967 cases and 381977 controls	European

Abbreviations: eQTL, expression quantitative trait locus; GWAS, genome-wide association analysis; CAGE, Consortium for the Architecture of Gene Expression.
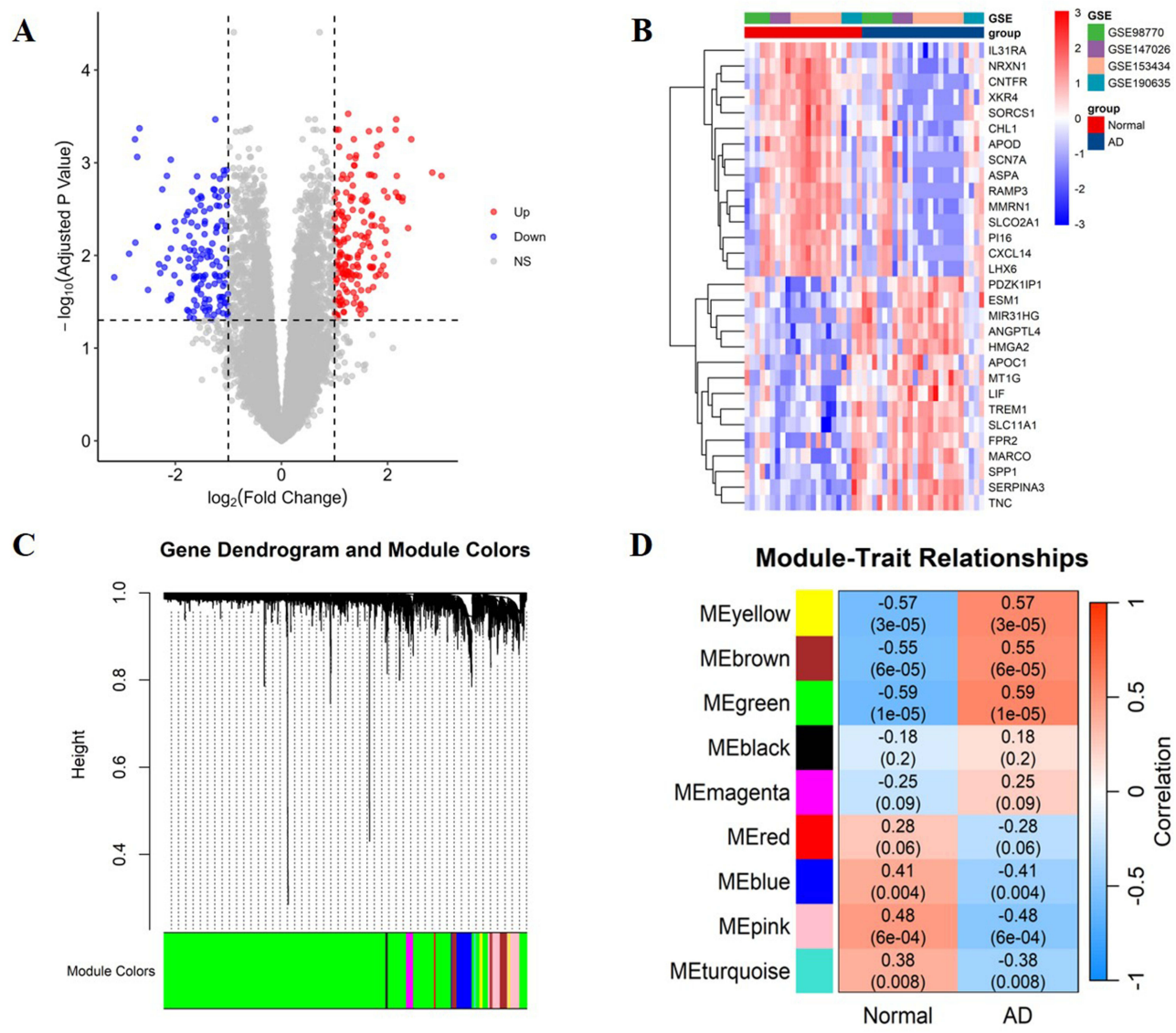
Figure 1 Identification of DEGs in AD and the results of WGCNA. (**A**) Volcano plot of DEGs in AD. (**B**) Clustering heatmap of the top 15 up- and down-regulated genes. (**C**) Dendrogram of genes. Each branch of the tree represents a gene, and each gene is clustered into the corresponding color module. (**D**) The correlation heatmap between clinical traits and color modules.

GO and KEGG Enrichment Analysis and PPI Network Construction

A total of 81 overlapped genes of 308 DEGs, 3011 OSRGs and 1421 WGCNA genes were screened, in which 61 genes are up-regulated and 20 genes are down-regulated (Figure 2A). According to GO and KEGG enrichment analysis, 517 GO terms and 14 KEGG pathways were obtained (Supplementary File 1, Tables S5 and S6). The results showed that the overlapped genes were primarily enriched in BPs of regulation of body fluid levels, response to oxygen levels, wound healing, response to hypoxia, and response to decreased oxygen levels. Regarding KEGG pathways, the genes were enriched in rheumatoid arthritis, HIF-1 signaling pathway, ECM-receptor interaction, complement and coagulation cascades, and PI3K-Akt signaling pathway (Figure 2B).

The PPI network of the overlapped genes was constructed by the STRING database to further identify the hub genes (Figure 2C). The "cytoHubba" plugin of Cytoscape software was used to analyze the top 15 hub genes based on the MCC algorithm (Figure 2D).
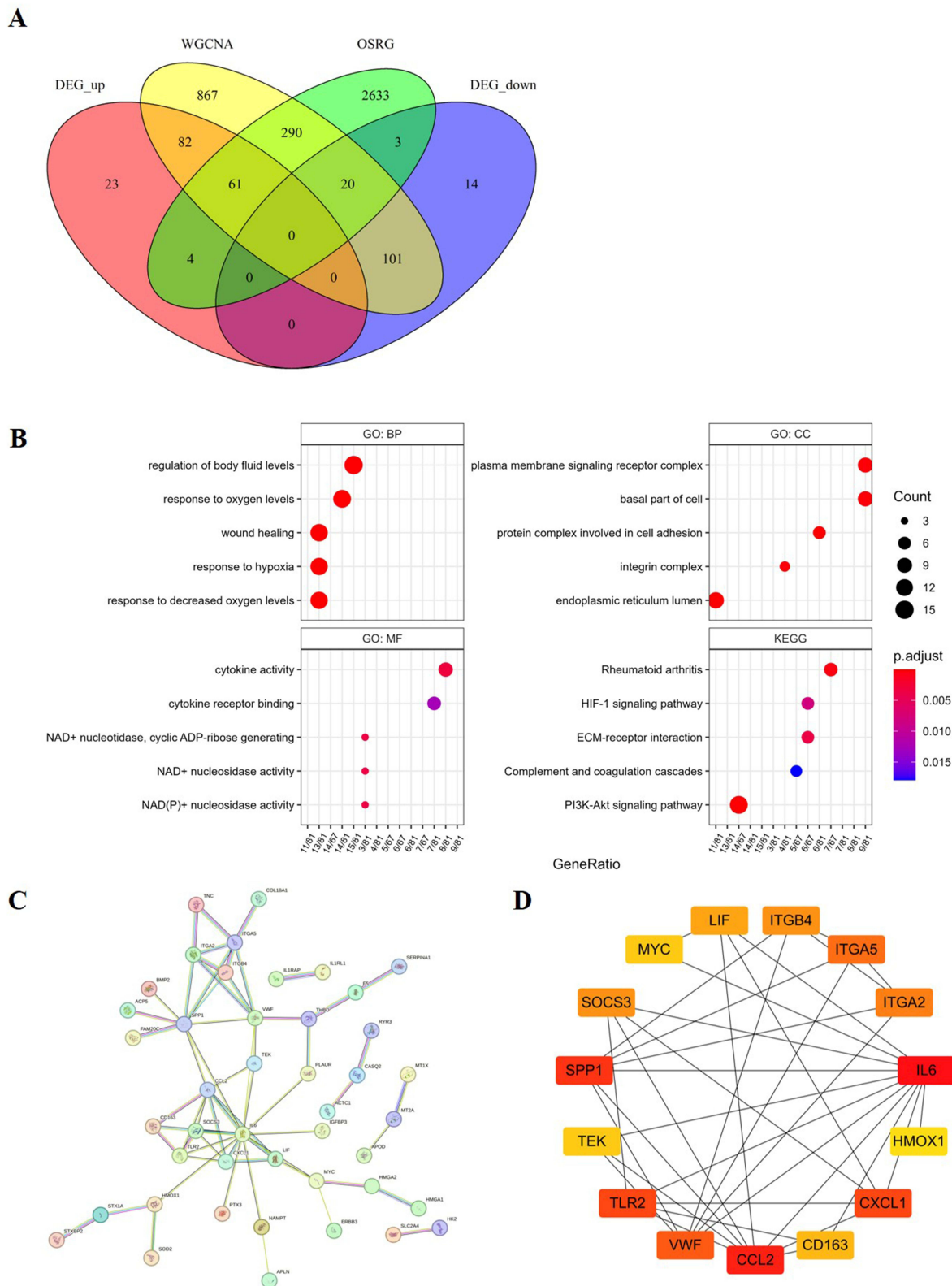
Figure 2 Functional enrichment analysis and PPI network construction. (**A**) Venn diagram of DEGs, OSRGs, and WGCNA genes. (**B**) Bubble plots of the top 5 terms of BP, CC, MF and KEGG pathways in which DEOSGs enriched. (**C**) PPI network of 81 OS-related DEGs constructed by the STRING database. (**D**) The top 15 hub genes screened by the "cytoHubba" plugin in Cytoscape software. A redder node color represents a higher gene ranking.

Identification of Signature Genes

The LASSO algorithm selected 8 genes by choosing the minimum λ as the optimal λ (Figure 3A and B). When the error of the algorithms was minimized, the RF and SVM-RFE screened 8 and 12 genes, respectively (Figure 3C-E). The signature genes jointly identified by the three machine learning algorithms included CCL2, ITGB4, MYC, SOCS3, SPP1, and TEK (Figure 3F).

The area under the curve (AUC) of all ROC curves was above 0.75, indicating satisfactory diagnostic accuracy of individual genes (Figure 4A). All signature genes were confirmed to be highly differentially expressed in the combined training cohort (Figure 4B), and CCL2, MYC, SOCS3, and SPP1 were up-regulated, while ITGB4 and TEK were down-regulated.

Gsva

The results of differential analysis of the enriched MsigDB gene sets showed that 59 and 59 gene ontology gene sets were up- and down-regulated, and 9 and 24 canonical pathways were up- and down-regulated (Supplementary File 1, Tables S7 and S8). Among these, BP of negative regulation of DNA templated DNA replication, CC of basal cortex, MF of aromatic amino acid transmembrane transporter activity, and pathways of SODD/TNFR1 signaling pathway, defective RIPK1-mediated regulated necrosis and extrinsic apoptotic pathways were highly significantly variant (Figure 5).

Construction of TF-Gene Interaction, Drug–Gene Interaction and ceRNA Networks

The TRRUST database detected 9 TFs, including CEBPA, STAT4, STAT3, HDAC2, RELA, NFKB1, STAT1, SP1, and JUN, as the collective regulators of CCL2, MYC, SOCS3, and SPP1 (Figure 6A and Supplementary File 1, Table S9). We searched for the most prominent TF CEBPA in the JASPAR database and predicted transcription factor-binding sites (TFBS) based on gene promoter sequences derived from the NCBI database with a relative profile score > 0.8 (Figure 6B and Supplementary File 1, Table S10).

Data on the signature genes and drug interactions were downloaded from the DGIdb database, and entries with interaction scores greater than the average were screened. Only CCL2, MYC, SPP1, and TEK with the interactions were filtered. Among these, CCL2 with Carlumab and Emapticap pegol had the highest interaction scores (Figure 6C).

7 DEmiRNAs and 261 DElncRNAs were identified in GSE98770 and GSE107844. The target mRNAs and lncRNAs of DEmiRNAs were subsequently mined from online databases. The conceivable ceRNA network embraced 2 signature mRNAs, 2 DEmiRNAs, and 4 DElncRNAs (Figure 6D).

Establishment and Assessment of the Clinical Prediction Model

Based on the clinical prediction model established by the signature genes, the nomogram was depicted to predict the risk of AD (Figure 7A). Model efficacy analyses showed that the C-index was 0.90 (95% CI 0.81–1.00), and the p-value of Hosmer-Lemeshow Goodness-of-Fit Test was 0.189. The calibration curve indicated that the model possessed a content accuracy in predicting absolute risk (Figure 7B). Compared with the decision curve of individual genes, the curve of the model had a relatively higher net benefit, indicating that the clinical decisions based on the model could be more profitable for patients (Figure 7C).

Validation by GSE52093 and RT-qPCR

The expression boxplots and the ROC curves of the signature genes in GSE52093 were exhibited in Figure 8A and B. All signature genes except ITGB4 and SOCS3 were significantly differentially expressed, and CCL2, MYC, and SPP1 were up-regulated, while TEK was down-regulated. All signature genes showed satisfactory predictive precision with AUC > 0.75. The ROC curve of the clinical prediction model demonstrated excellent prediction accuracy with an AUC of 1.00 (95% CI 1.00–1.00) (Figure 8C). The results implied that the model was valuable in differentiating AD patients from normal individuals.

RT-qPCR was performed to examine the expression of the signature genes in dissected and normal aortic specimens. Compared with controls, CCL2, MYC, SOCS3 and SPP1 were up-regulated, while ITGB4 and TEK were down-regulated (Figure 8D).
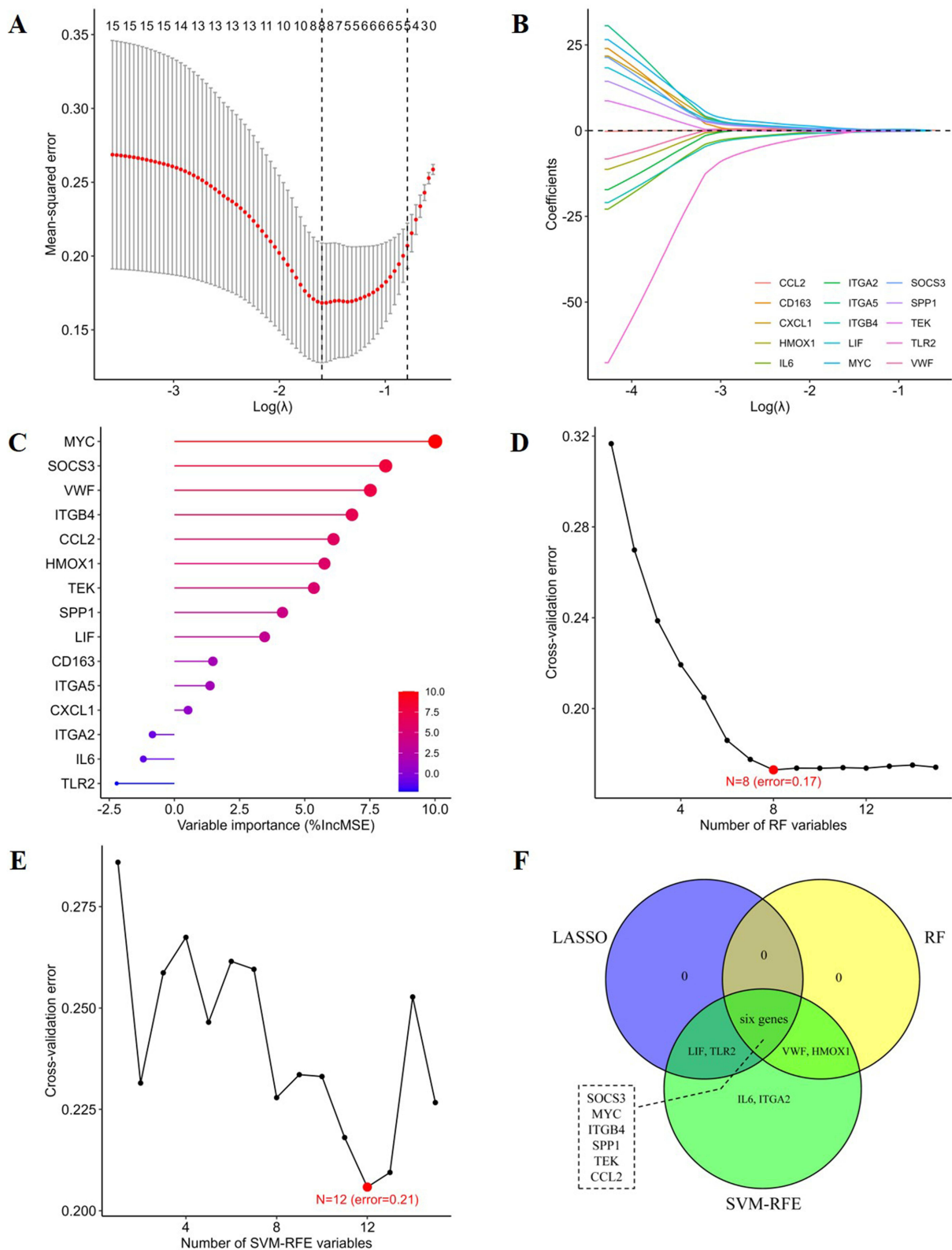
Figure 3 Screening of signature genes by machine learning algorithms. (**A** and **B**) Diagrams of the process of selecting variables by LASSO algorithm. (**C**) The importance ranking of variables obtained by RF algorithm. Graphs to ascertain the optimal variable size by the minimum error of RF (**D**) and SVM-REF (**E**) algorithms. (**F**) Venn diagram of signature genes.
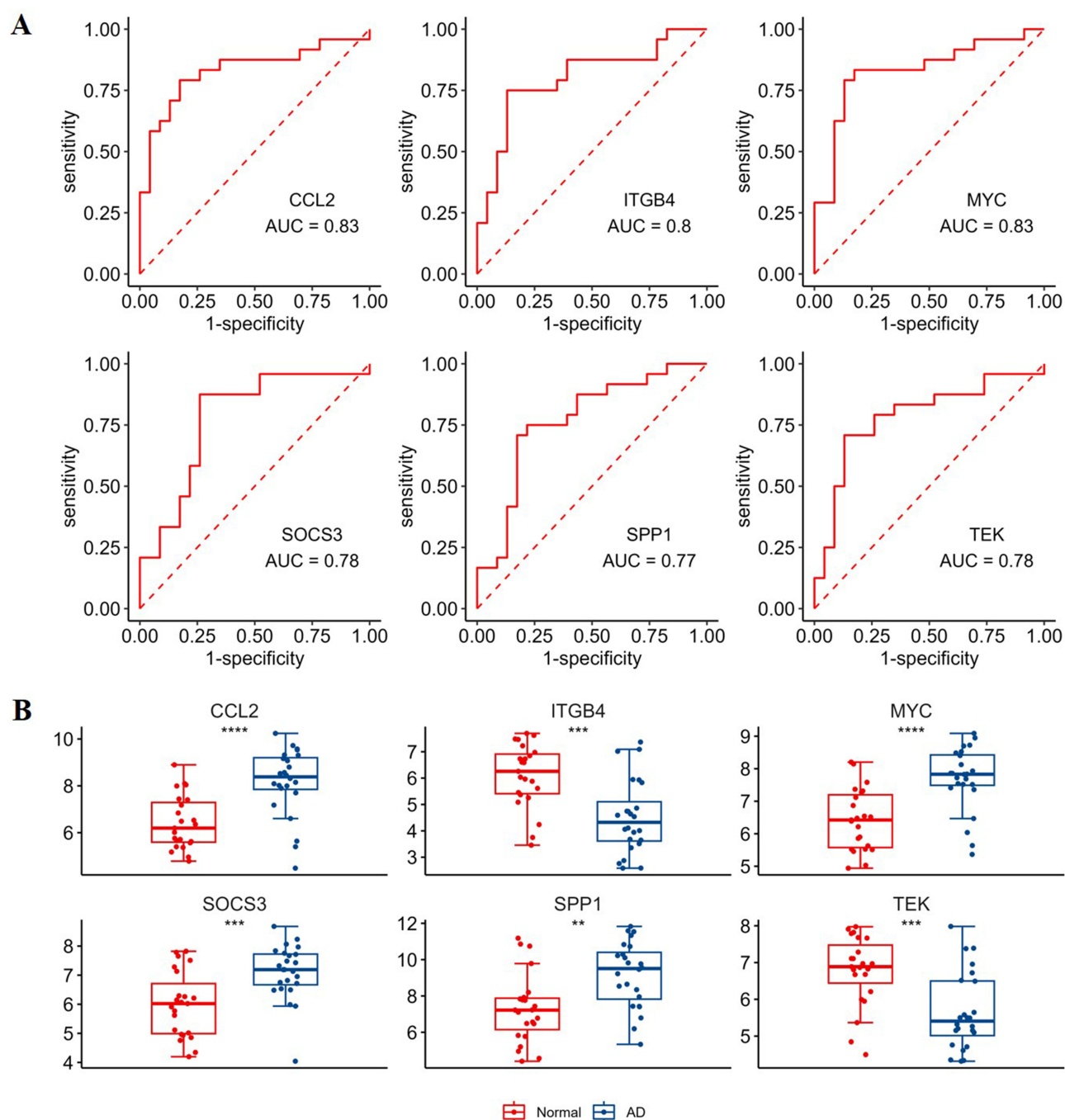
Figure 4 Identification of the signature genes. (**A**) ROC curves of the signature genes in combined training cohort. (**B**) The expression differences of the signature genes between AD and normal samples in combined training cohort (** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

Validation by MR

The design of the MR study was shown in Figure 9A. A total of 2, 3, 4, 1, 23, and 18 SNPs were identified from the eQTLGen consortium or CAGE for CCL2, ITGB4, MYC, SOCS3, SPP1, and TEK (Supplementary File 1, Table S11). The results of both TSMR and SMR suggested that no significant causal effect between the expression levels of the signature genes mediated by genetic variants and the risk of AD was detected, and the trend of TSMR results was primarily consistent with that of SMR (Figure 9B and C and Supplementary File 1, Tables S12 and 13). The scatter plots,

A

B

Figure 5 The results of GSVA. The top 10 up- and down-regulated gene ontology (**A**) and canonical pathways (**B**) that the low- and high-NES$_{total}$ group variously enriched in.

forest plots, funnel plots, and leave-one-out analysis plots of TSMR and the locus plots and effect plots of SMR were exhibited in Supplementary File 2, Figures S3–8.

Discussion

Oxidative stress has been authenticated to play an important role in cardiovascular diseases, but its possible mechanisms in AD have been rarely reported. This study investigated the underlying molecular mechanisms of OS in AD by thorough bioinformatics approaches on the gene expression profiles of AD and normal samples from the GEO database. Pertaining to GO and KEGG enrichment analysis based on the OS-related DEGs, the BPs were predominantly connected with the response to hypoxia and regulation of body fluid levels, the CCs were principally located in the basal part of cell and endoplasmic reticulum lumen, the MFs were primarily related to the activity of cytokine, NAD+ nucleosidase, and NAD(P)+ nucleosidase, and the pathways contained ECM-receptor interaction, complement and coagulation cascades,
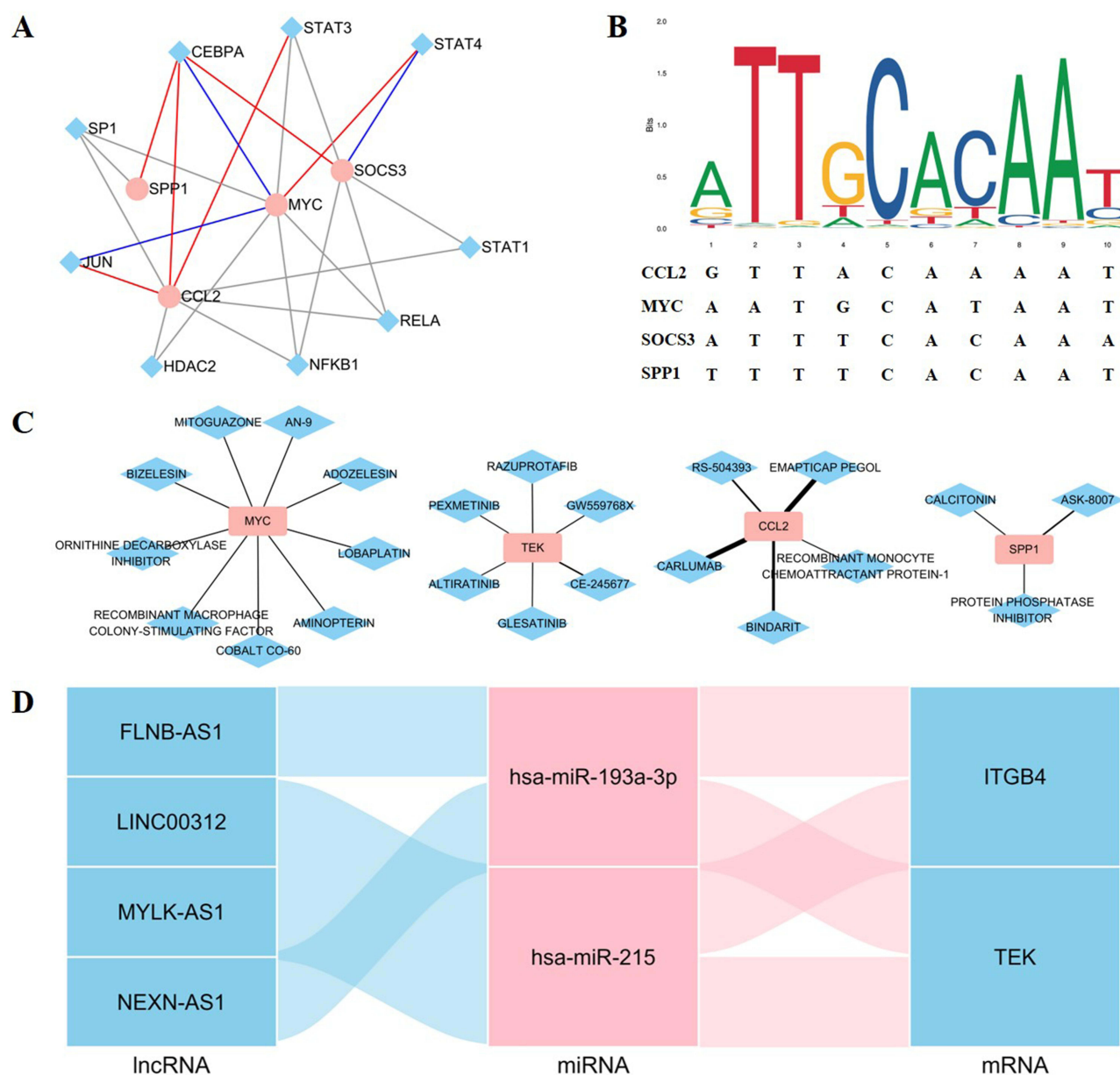
Figure 6 Molecular networks. (**A**) The TF-gene interaction network retrieved from the TRRUST database. The red circle represents gene, the blue diamond represents TF, and the red, blue and grey lines represent activation, repression, and unclear regulation modes. (**B**) The CEBPA motif and the predicted TFBS based on the JASPAR database. (**C**) The drug–gene interaction network mined by the DGIdb database. The red rectangle represents gene, the blue diamond represents drug, and a thicker line represents a higher interaction score. (**D**) Sankey plot of the ceRNA network. The red and blue stratums represent up- and down-regulated RNA.

and PI3K-Akt and HIF-1 signaling pathways. The signature genes were screened by machine learning algorithms thereafter, including 4 up-regulated genes (CCL2, MYC, SOCS3, and SPP1) and 2 down-regulated genes (ITGB4 and TEK). Analyses including GSVA and molecular network construction were subsequently executed to have a deeper cognition of the signature genes.

Unfortunately, the results of the eQTL-based MR were negative, implying that the expression of the signature genes may not be causally related to AD. The expression of genes could be affected by a variety of non-genetic factors, such as regulation of TFs and ncRNA, epigenetic modifications, and environmental influences, which may lead to negative MR results. Based on the above results, we consider that these OS-related signature genes may not be the cause of AD, but they play an important role in the occurrence and development of AD.
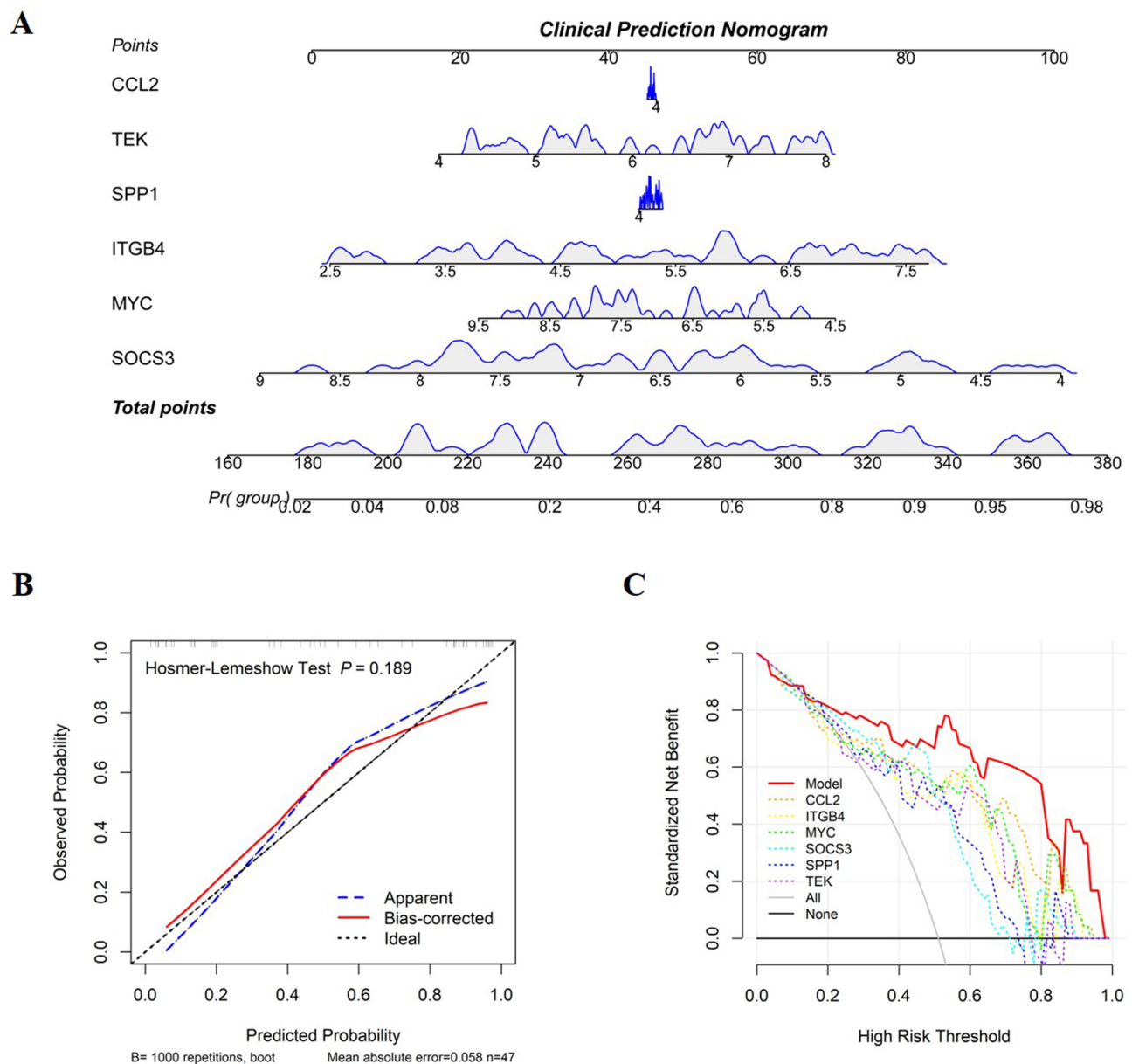
Figure 7 Construction and assessment of the clinical prediction model based on the signature genes. (**A**) Nomogram for predicting the risk of AD. (**B**) Calibration curves and Hosmer-Lemeshow Goodness of Fit Test of the model (p-value > 0.05 indicating a satisfactory calibration). (**C**) Decision curves of the model and individual genes.

C-C motif chemokine ligand 2 (CCL2) exhibits chemotactic activity for monocytes and basophils but not for neutrophils or eosinophils.[29] Its potential role lies in recruiting monocytes into the arterial wall during atherosclerosis progression.[30] The results of bulk sequencing,[31] single-cell sequencing[32] and histological examination[33] confirmed that numerous macrophages infiltrating the aorta of AD patients intensively expressed CCL2. In the suprarenal aorta of apoE (-/-) mice infused with Ang-II, neovascularization and macrophage infiltration were observed, and the expression of matrix metalloproteinase-2 (MMP2), MMP9, CCL2, and VEGF significantly elevated.[34,35]

Integrin subunit beta 4 (ITGB4) is a receptor for laminin, and its heterodimers, namely integrins, facilitate adhesion between cells and the extracellular matrix and serve as signal transducers modulating gene expression and cellular growth. ITGB4 is indispensable for adhesion, proliferation, apoptosis, and senescence, and its down-regulation may participate in the pathological process of AD by reducing vascular endothelial cell adhesion.[36]
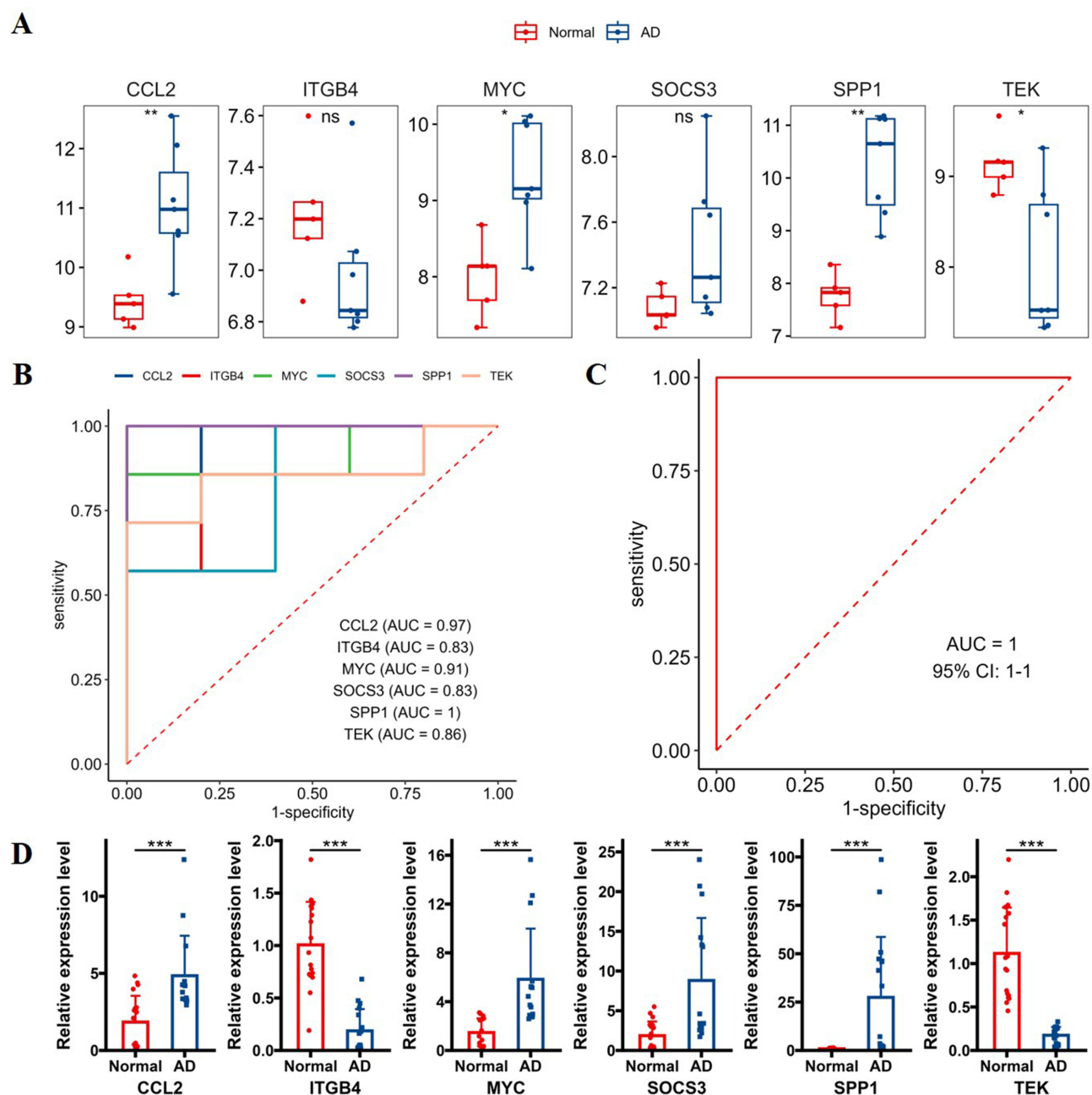
Figure 8 Validation by external dataset and RT-qPCR. (**A**) The expression boxplots of the signature genes in GSE52093. (**B**) The ROC curves of individual signature genes. (**C**) The ROC curve of the clinical prediction model. (**D**) The relative expression level of the signature genes in normal and dissected aortic specimens by RT-qPCR. (* p<0.05, ** p<0.01, *** p<0.001).

MYC, a proto-oncogene, contributes to cell cycle regulation, cellular transformation, and apoptosis. By binding to the promoter of vascular endothelial growth factor A (VEGFA), it enhances VEGFA synthesis and thus promotes sprouting angiogenesis.[37] In the polycystin-1 (PC1)-mediated MEK/ERK/myc signaling pathway, the down-regulation of PC1 in VSMCs upregulates phospho-MEK, phospho-ERK, and myc, inducing the phenotypic switch of VSMCs, which may be a pivotal pathophysiological cause of AD.[38]

Induced by various cytokines, suppressor of cytokine signaling 3 (SOCS3) blocks cytokine signal transmission by attaching to tyrosine kinase receptors and engaging in the JAK/STAT pathway. SOCS3 in macrophage protects stressed aorta from tissue destruction and the development of AD by impeding excessive inflammatory responses, promoting tissue repair responses, and maintaining the function of VSMCs.[39] Targeted removal of SOCS3 from smooth muscle cells
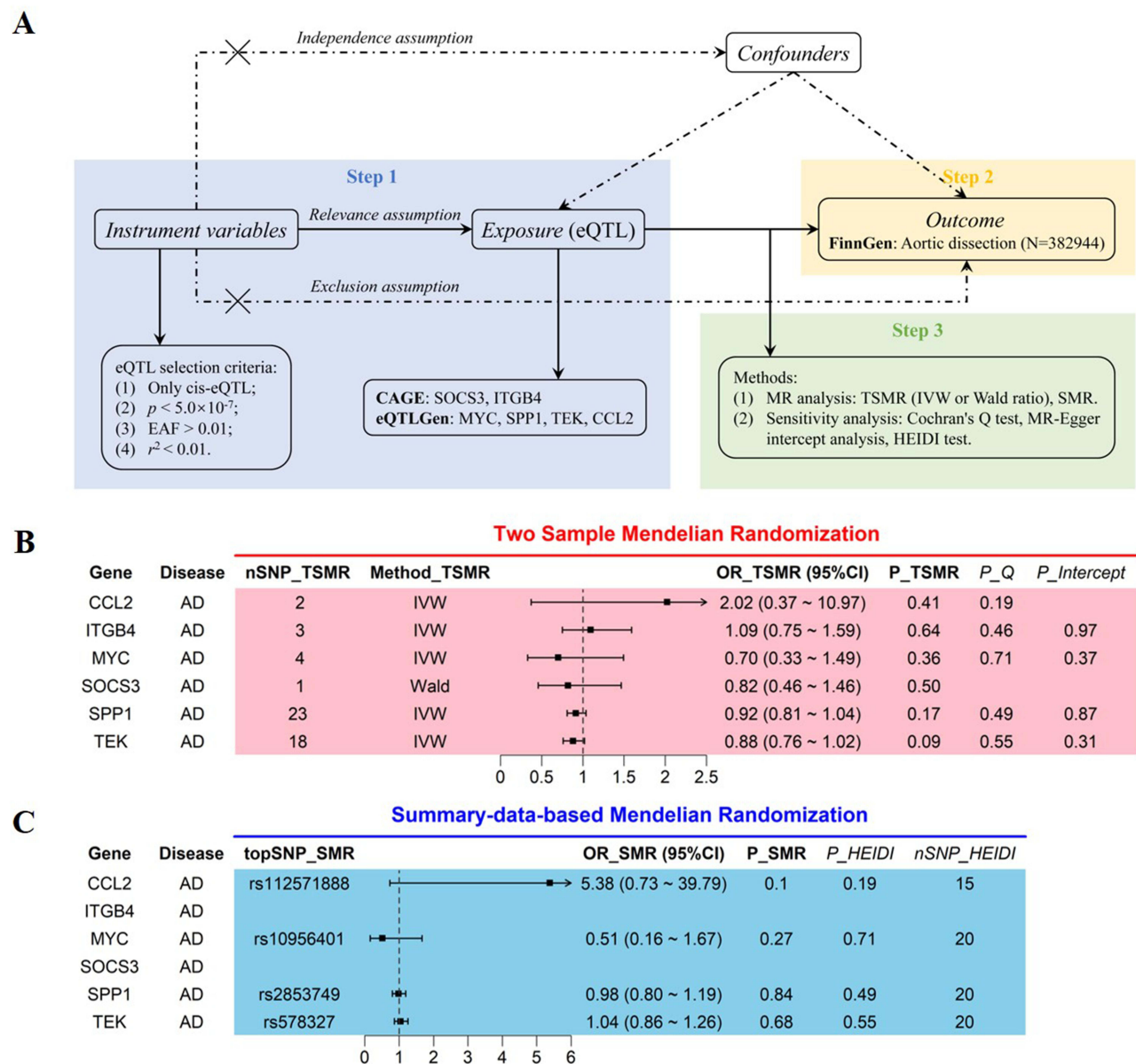
Figure 9 The study design and results of MR analysis. (**A**) The diagram of the MR study design. Relevance assumption, genetic variants are associated with exposures; Independence assumption, genetic variants are not associated with outcomes through confounders; exclusion assumption, genetic variants affect outcomes simply via exposures, rather than directly affect the outcomes. (**B**) The forest plots of the results of TSMR. (**C**) The forest plots of the results of SMR.

(SMC) led to chronic inflammation of the aortic wall, resulting in increased fibroblasts and decreased tissue destruction, which safeguarded the aorta against AD.[40] This accentuates that, depending on the cell type, disease severity may be promoted or mitigated by gene expression and the activation of consequential signaling pathways.

Secreted phosphoprotein 1 (SPP1) acts as a cytokine that is essential in type I immunity by upregulating of INF-γ and IL-12 while downregulating IL-10. SPP1 can be regulated to participate in VSMC apoptosis[41] and inhibit angiotensin II (Ang II)-induced phenotypic switch of aortic SMC.[42] While under normal physiological conditions, SPP1 has been corroborated to contribute to the regulation of vascular tone in resistance arteries and tissue perfusion.[43]

TEK encodes a receptor that is part of the protein tyrosine kinase Tie2 family and functions as a constituent of the angiopoietin (Ang)-Tie system, which is the second endothelial cell-specific ligand-receptor signaling system, in addition to the VEGF receptor pathway. The Ang-Tie system controls the formation of new blood vessels, changes in blood vessel remodeling and permeability, and immune response to maintain the stability of blood vessels.[44,45] Furthermore, TEK

regulates multiple functions of endothelial cell, such as cell adhesion, migration, proliferation and spreading, and promotes the restructuring of the actin cytoskeleton. Nonetheless, there was limited evidence on the mechanisms of ITGB4 and TEK in AD, hence more investigations are required to elucidate their association with AD.

Notwithstanding, this study had several limitations likewise. Firstly, restricted by the low prevalence of AD, the small sample size obtained from the GEO database may lead to biased results, so more samples are needed to ameliorate the analysis. Secondly, the lack of clinical information in the GEO database highlights the necessity to collect the corresponding clinical indicators to explore the relationship between OS and the prognosis of AD patients. Thirdly, the negative results of MR indicated that it is not the genetic variation of the expression of the signature genes that led to the occurrence of AD, but the chronological order of the changes of the gene expression level and the occurrence of AD is still unknown. Finally, this study has only conducted the identification and simple validation of the OS-related DEGs, hence more profound studies are required to further investigate the molecular mechanisms.

Conclusion

This study provided novel insights into the role of OS in AD, and four biomarkers associated with OS (CCL2, MYC, SPP1, and TEK) were identified and validated. Moreover, further studies based on these signature genes may contribute to explore the potential pathogenesis and therapeutic targets.

Abbreviations

AA, aortic aneurysm; AD, aortic dissection; AUC, area under curve; BP, biological process; CAGE, Consortium for the Architecture of Gene Expression; CC, cellular component; CCL2, C-C motif chemokine ligand 2; ceRNA, competitive endogenous RNA; DCA, decision curve analysis; DEG, differentially expressed gene; DElncRNA, differentially expressed long non-coding RNA; DEmiRNA, differentially expressed microRNA; ECM, extracellular matrix; eQTL, expression quantitative trait locus; FDR, false discovery rate; GEO, Gene Expression Omnibus; GM-CSF, granulocyte macrophage colony stimulating factor; GO, gene ontology; GSVA, gene set variation analysis; GWAS, genome-wide association analysis; HEIDI, heterogeneity in dependent instruments; ITGB4, integrin subunit beta 4; IV, instrumental variable; IVW, inverse variance weighted; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; LD, linkage disequilibrium; MCC, maximal clique centrality; MF, molecular function; MMP, matrix metalloproteinase; MR, Mendelian randomization; MSigDB, Molecular Signatures Database; ncRNA, non-coding RNA; NES, normalized enrichment score; OS, oxidative stress; OSRG, oxidative stress-related gene; PCA, principal component analysis; PPI, protein–protein interaction; RF, random forest; ROC, receiver operating characteristic; RT-qPCR, quantitative real-time reverse transcription polymerase chain reaction; SMC, smooth muscle cell; SMR, summary-data-based Mendelian randomization; SNP, single nucleotide polymorphism; SOCS3, suppressor of cytokine signaling 3; SPP1, secreted phosphoprotein 1; ssGSEA, single-sample gene set enrichment analysis; SVM-RFE, support vector machine and recursive feature elimination; TF, transcription factor; TFBS, transcription factor-binding site; TSMR, two sample Mendelian randomization; VEGFA, vascular endothelial growth factor A; VSMC, vascular smooth muscle cell; WGCNA, weighted gene correlation network analysis.

Ethics Approval and Informed Consent

The study complies with the Declaration of Helsinki. The study has been approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University, and informed consents were obtained from all participating individuals or their relatives.

Acknowledgments

Author Contributions

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

Funding

Disclosure

The author(s) report no conflicts of interest in this work.

References

1. Kim JB, Kim K, Lindsay ME, et al. Risk of rupture or dissection in descending thoracic aortic aneurysm. *Circulation*. 2015;132(17):1620–1629. doi:10.1161/CIRCULATIONAHA.114.015177
2. Howard DP, Banerjee A, Fairhead JF, Perkins J, Silver LE, Rothwell PM. Population-based study of incidence and outcome of acute aortic dissection and premorbid risk factor control: 10-year results from the oxford vascular study. *Circulation*. 2013;127(20):2031–2037. doi:10.1161/CIRCULATIONAHA.112.000483
3. Spinelli D, Benedetto F, Donato R, et al. Current evidence in predictors of aortic growth and events in acute type B aortic dissection. *J Vasc Surg*. 2018;68(6):1925–1935.e8. doi:10.1016/j.jvs.2018.05.232
4. Lau C, Leonard JR, Iannacone E, Gaudino M, Girardi LN. Surgery for acute presentation of thoracoabdominal aortic disease. *Semin Thorac Cardiovasc Surg*. 2019;31(1):11–16. doi:10.1053/j.semtcvs.2018.07.018
5. Xu S, Han X, Wang X, et al. The role of oxidative stress in aortic dissection: a potential therapeutic target. *Front Cardiovasc Med*. 2024;11:1410477. doi:10.3389/fcvm.2024.1410477
6. Gauthier J, Vincent AT, Charette SJ, Derome N. A brief history of bioinformatics. *Brief Bioinform*. 2019;20(6):1981–1996. doi:10.1093/bib/bby063
7. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013;41(Database issue): D991–5. doi:10.1093/nar/gks1193
8. Stelzer G, Rosen N, Plaschkes I, et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinf*. 2016;54:1.30.1–1.30.33. doi:10.1002/cpbi.5
9. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet*. 2000;25 (1):25–29. doi:10.1038/75556
10. Jin Z, Sato Y, Kawashima M, Kanehisa M. KEGG tools for classification and analysis of viral proteins. *Protein Sci*. 2023;32(12):e4820. doi:10.1002/pro.4820
11. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–d613. doi:10.1093/nar/gky1131
12. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504. doi:10.1101/gr.1239303
13. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(Suppl 4(Suppl 4)):S11. doi:10.1186/1752-0509-8-S4-S11
14. Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinf*. 2018;19(1):432. doi:10.1186/s12859-018-2451-4
15. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–425. doi:10.1016/j.cels.2015.12.004
16. Han H, Cho JW, Lee S, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res*. 2018;46(D1):D380–d386. doi:10.1093/nar/gkx1013
17. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2024;52(D1):D174–d182. doi:10.1093/nar/gkad1059
18. Freshour SL, Kiwala S, Cotto KC, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res*. 2021;49(D1):D1144–d1151. doi:10.1093/nar/gkaa1084
19. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48(D1):D127–d131. doi:10.1093/nar/gkz757
20. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*. 2018;13(10): e0206239. doi:10.1371/journal.pone.0206239
21. Skoufos G, Kakoulidis P, Tastsoglou S, et al. TarBase-v9.0 extends experimentally supported miRNA-gene interactions to cell-types and virally encoded miRNAs. *Nucleic Acids Res*. 2024;52(D1):D304–d310. doi:10.1093/nar/gkad1071
22. McGeary SE, Lin KS, Shi CY, et al. The biochemical basis of microRNA targeting efficacy. *Science*. 2019;366(6472). doi:10.1126/science.aav1741
23. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28 (15):2062–2063. doi:10.1093/bioinformatics/bts344

24. Karagkouni D, Paraskevopoulou MD, Tastsoglou S, et al. DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res*. 2020;48(D1):D101–d110. doi:10.1093/nar/gkz1036

25. Võsa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300–1310. doi:10.1038/s41588-021-00913-z

26. Lloyd-Jones LR, Holloway A, McRae A, et al. The genetic architecture of gene expression in peripheral blood. *Am J Hum Genet*. 2017;100 (2):228–237. doi:10.1016/j.ajhg.2016.12.008

27. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613 (7944):508–518. doi:10.1038/s41586-022-05473-8

28. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48 (5):481–487. doi:10.1038/ng.3538

29. Weber M, Uguccioni M, Baggiolini M, Clark-Lewis I, Dahinden CA. Deletion of the NH2-terminal residue converts monocyte chemotactic protein 1 from an activator of basophil mediator release to an eosinophil chemoattractant. *J Exp Med*. 1996;183(2):681–685. doi:10.1084/jem.183.2.681

30. Li YS, Shyy YJ, Wright JG, Valente AJ, Cornhill JF, Kolattukudy PE. The expression of monocyte chemotactic protein (MCP-1) in human vascular endothelium in vitro and in vivo. *Mol Cell Biochem*. 1993;126(1):61–68. doi:10.1007/BF01772208

31. Chen F, Han J, Tang B. Patterns of immune infiltration and the key immune-related genes in acute type a aortic dissection in bioinformatics analyses. *Int J Gen Med*. 2021;14:2857–2869. doi:10.2147/IJGM.S317405

32. Zhang B, Zeng K, Guan RC, et al. Single-cell RNA-seq analysis reveals macrophages are involved in the pathogenesis of human sporadic acute type A aortic dissection. *Biomolecules*. 2023;13(2):399.

33. Kimura S, Sato H, Shimajiri S, et al. Association of troponin I and macrophages in cardiac tamponade with Stanford type A aortic dissection. *Heliyon*. 2023;9(10):e20791. doi:10.1016/j.heliyon.2023.e20791

34. Ortega R, Collado A, Selles F, et al. SGLT-2 (Sodium-Glucose Cotransporter 2) inhibition reduces Ang II (Angiotensin II)-induced dissecting abdominal aortic aneurysm in ApoE (Apolipoprotein E) knockout mice. *Arterioscler Thromb Vasc Biol*. 2019;39(8):1614–1628. doi:10.1161/ ATVBAHA.119.312659

35. Martorell S, Hueso L, Gonzalez-Navarro H, Collado A, Sanz MJ, Piqueras L. Vitamin D receptor activation reduces angiotensin-II-induced dissecting abdominal aortic aneurysm in apolipoprotein E-knockout mice. *Arterioscler Thromb Vasc Biol*. 2016;36(8):1587–1597. doi:10.1161/ ATVBAHA.116.307530

36. Wang L, Dong Z, Zhang Y, Miao J. The roles of integrin β4 in vascular endothelial cells. *J Cell Physiol*. 2012;227(2):474–478. doi:10.1002/ jcp.22769

37. Shi Y, Xu X, Zhang Q, et al. tRNA synthetase counteracts c-Myc to develop functional vasculature. *Elife*. 2014;3:e02349. doi:10.7554/eLife.02349

38. Feng J, Ge S, Zhang L, Che H, Liang C. Aortic dissection is associated with reduced polycystin-1 expression, an abnormality that leads to increased ERK phosphorylation in vascular smooth muscle cells. *Eur J Histochem*. 2016;60(4):2711. doi:10.4081/ejh.2016.2711

39. Ohno-Urabe S, Aoki H, Nishihara M, et al. Role of macrophage Socs3 in the pathogenesis of aortic dissection. *J Am Heart Assoc*. 2018;7(2). doi:10.1161/JAHA.117.007389

40. Hirakata S, Aoki H, Ohno-Urabe S, et al. Genetic deletion of Socs3 in smooth muscle cells ameliorates aortic dissection in mice. *JACC Basic Transl Sci*. 2020;5(2):126–144. doi:10.1016/j.jacbts.2019.10.010

41. Zhang X, Che Y, Mao L, et al. H3.3B controls aortic dissection progression by regulating vascular smooth muscle cells phenotypic transition and vascular inflammation. *Genomics*. 2023;115(5):110685. doi:10.1016/j.ygeno.2023.110685

42. Chen S, Chen H, Zhong Y, et al. Insulin-like growth factor-binding protein 3 inhibits angiotensin II-induced aortic smooth muscle cell phenotypic switch and matrix metalloproteinase expression. *Exp Physiol*. 2020;105(11):1827–1839. doi:10.1113/EP088927

43. Peter BF, Lidington D, Harada A, et al. Role of sphingosine-1-phosphate phosphohydrolase 1 in the regulation of resistance artery tone. *Circ Res*. 2008;103(3):315–324. doi:10.1161/CIRCRESAHA.108.173575

44. Dumont DJ, Gradwohl GJ, Fong GH, Auerbach R, Breitman ML. The endothelial-specific receptor tyrosine kinase, tek, is a member of a new subfamily of receptors. *Oncogene*. 1993;8(5):1293–1301.

45. Eklund L, Kangas J, Saharinen P. Angiopoietin-Tie signalling in the cardiovascular and lymphatic systems. *Clin Sci*. 2017;131(1):87–103. doi:10.1042/CS20160129