BMC Cardiovascular Disorders

RESEARCH

Open Access

Screening of CAD-related secretory genes associated with type II diabetes based on comprehensive bioinformatics analysis and machine learning

Li Xie[1†], Han Xiao[1], Maoyu Zhao[1], Li Xu[1], Si Tang[1] and Youzhu Qiu[1*†]

Abstract

Background Type II diabetes mellitus (T2DM) is strongly linked with a heightened risk of coronary artery disease (CAD). Exploring biological targets common to T2DM and CAD is essential for CAD intervention strategies.

Methods RNA transcriptome data from CAD and T2DM patients and single-cell transcriptional data from myocardial tissue of CAD patients were used for bioinformatics analysis. Differential analysis and Weighted Gene Co-expression Network Analysis (WGCNA) were conducted to identify hub genes associated with the CAD Index (CADi) in these cells. We then intersected these genes with differentially expressed genes in the T2DM dataset to validate the key gene FGF7. Additional analyses included immune analysis, drug sensitivity, competing endogenous RNA (ceRNA) networks, and smooth muscle cell -related functional analysis.

Results An abnormally high proportion of smooth muscle cells was observed in CAD tissues compared to normal cardiomyocytes. The gene FGF7, which encodes the keratinocyte growth factor 7 protein, showed increased expression in both CAD and T2DM and was significantly positively correlated with the CADi (correlation = 0.24, $p < 0.05$). FGF7 expression was inversely correlated with $CD4^+$ and $CD8^+$ T-cell immune infiltration and correlated with the cardiovascular drugs. Overexpression of FGF7 in CAD samples enhanced interactions with mononuclear macrophages and influenced the metabolism of alanine, glutamate, nicotinamide, and retinol. We also identified that hsa-miR-15a-5p, hsa-miR-373-3p, hsa-miR-20a-5p, and hsa-miR-372-3p could regulate FGF7 expression.

Conclusion FGF7 serves as a critical shared biological target for T2DM and CAD, playing a significant role in CAD progression with potential therapeutic implications.

Keywords Coronary artery disease, Type II diabetes mellitus, FGF7, Bioinformatics analysis, Smooth muscle cells

[†]Li Xie and Youzhu Qiu contributed equally to this work.

*Correspondence:
Youzhu Qiu
zeng000xim@163.com
[1]Department of Cardiology, The Second Affiliated Hospital, Army Medical University, Third Military Medical University, Chongqing 400037, China

Background

Cardiovascular disease (CVD) is a leading cause of death and morbidity worldwide, killing approximately 17.9 million people each year, or one person every 40 s [1]. Approximately 80% of deaths are from myocardial infarction (MI) [2]. CAD, also known as coronary heart disease (CHD), is characterized by coronary atherosclerotic lesions that narrow and block the blood lumen, leading to myocardial ischemia and even necrosis [3]. Patients with diabetes, hypertension, obesity, smoking, and a family history are predisposed to CAD. Currently, serological markers such as troponin T/I and creatine kinase-MB are essential for the diagnosis of CAD [4], but these markers depend on the presence of associated symptoms [5]. Recent studies have identified CAD-related transcripts in blood and muscle using chip-based approaches, and some of the chip data obtained has been applied to some clinical practices [6, 7]; however, effective biomarkers for predicting disease progression in CAD patients remain scarce, posing a significant clinical challenge. Therefore, identifying effective biomarkers that predict disease progression in CAD patients is urgent and critical.

Type II diabetes (T2DM) is one of the most prevalent chronic metabolic diseases globally, affecting multiple bodily systems, including significant changes in the musculoskeletal system [8]. It is primarily characterized by insulin resistance, where metabolic tissues such as skeletal muscle and liver fail to utilize insulin effectively, leading to elevated blood sugar levels and hyperglycemia [9]. The global incidence of T2DM is on the rise, partly due to an aging population, making it a leading cause of cardiovascular and neurodegenerative diseases. Diabetes compromises the strength of cardiac and skeletal muscles through impaired energy metabolism, reduced blood flow, and mitochondrial dysfunction [10]. Over recent decades, the incidence of T2DM has surged, becoming the seventh leading cause of death in the United States.

T2DM is strongly linked to an increased risk of myocardial dysfunction and coronary artery disease (CAD), with these conditions often coexisting and influencing each other [11]. Up to 75% of T2DM patients may experience myocardial dysfunction, which frequently goes unrecognized due to its asymptomatic nature, comorbidities such as CAD and obesity, and the absence of standardized diagnostic criteria, leading to delayed treatment [12]. Thus, identifying shared biomarkers between T2DM and CAD is crucial for the diagnosis and treatment of these diseases. In this study, we utilized multiple integrated bioinformatics tools to identify critical genes. We analyzed three CAD datasets and one T2DM dataset to pinpoint key genes linking these diseases. We then developed a binary classification model using machine learning techniques to assess the potential of these key genes in predicting CAD. The Coronary Artery Disease Index (CADi) prediction based on key genes highlighted FGF7, whose potential function was further explored through immune and drug sensitivity analysis.

Materials and methods

Data acquisition

We accessed three CAD-related datasets from the Gene Expression Omnibus Database (GEO, http://www.ncbi.nlm.nih.gov/geo): GSE 12,288 [13], GSE 20,680 [14, 15], and GSE 221,911 [16]. GSE 12,288 comprises a total of 222 samples, including 110 CAD and 112 normal samples, and the sample tissue source is peripheral blood. GSE 20,680 includes 139 samples, with 87 representing CAD and 52 as normal, and the sample tissue source is peripheral blood. GSE 221,911 features 134 samples, consisting of 55 CAD and 79 normal samples, and the sample tissue source is peripheral blood. Additionally, a T2DM dataset, GSE 179,455 [17], was incorporated, containing 19 samples with 5 T2DM, 7 lean and 7 fat individuals, and the sample tissue source is epicardial adipose tissue. All data, including sequencing and chip data, underwent normalization prior to analysis. For single-cell data related to CAD, we utilized GSE131778 [18, 19], which contains data from proximal to mid coronary artery of eight CAD patients, totaling 11,754 cells, and GSE 121,893 [20], comprising two normal samples which cells come from left artery and ventricular cells in heart tissue, totaling 356 cells. We performed quality control, normalization, and batch effect removal on the single-cell data using Seurat V5 software.

CellChat

To infer and analyze intercellular communication in CAD samples, we utilized CellChat, a public library equipped with ligands, receptors, cofactors, and their interactions [21]. CellChat serves as a versatile and user-friendly toolkit, accessible via a web-based resource manager (http://www.cellchat.org/), which facilitates the discovery of novel cell-to-cell communication pathways and the construction of comprehensive cell-to-cell communication maps. In the analysis of cell interactions, expression levels are calculated relative to the total values mapped to the same set of encoded genes across all transcriptomes. Additionally, expression values are averaged for each individual cell population or cell sample, ensuring a standardized approach to data analysis.

Trajectory analysis

Monocle 3 version is used to perform trajectory analysis to predict differentiation paths between different cells in CAD samples without prior knowledge of differentiation time or direction [22].

WGCNA analysis

The WGCNA software package was used to identify key genes significantly associated with CAD [23]. We began by inputting the expression profile of CAD patients from the GSE12288 dataset. Soft thresholds were determined based on the characteristics of the expression matrix, after which adjacency matrices were clustered, and core modules were identified. Pearson correlation coefficients between these modules and CAD cores were calculated to identify modules with the strongest positive correlations for further analysis. We assessed gene significance (GS) and module membership (MM) within these central modules for each genetic trait. Ultimately, we selected the modules that demonstrated the strongest correlations and significance, and screened for genes that might be associated with CAD.

Protein-protein network interaction (PPI)

The STRING database (https://string-db.org/) was used to identify and predict PPIs [24]. After establishing the PPI, Cytoscape was employed to visualize the interactions, and the MCC algorithm of the cytoHubba software calculated the rank value of each protein, determining the top 15 hub genes from the network in descending order of importance [25].

Enrichment analysis

Metascape (metascape.org) is used for functional enrichment analysis, known for its user-friendly and reliable features [26]. The criteria set included a minimum overlap of 3, minimum enrichment of 3, and a P value threshold of 0.05. We then conducted an enrichment analysis of the hub genes using this platform.

Establishment of mechanical learning model

We utilized GSE12288 as the training set and GSE20680 as the test set. Fifteen machine learning algorithms were integrated, including neural network, logistic regression, linear discriminant analysis, quadratic discriminant analysis, KNN nearest neighbors, decision tree classification, random forest classification, XGBoost classification, ridge regression, lasso regression, support vector machine, gradient boosting, stepwise logistic regression, and naive Bayesian model. We implemented 10-fold cross-validation to identify the best lasso features, where each model underwent training validation with all features and with the best features only. Variable importance output was performed for any machine learning algorithm with a feature selection function. We selected each parameter corresponding to the main features, and formed 207 optimal parameter combinations by integrating multiple parameters.

Drug sensitivity analysis

Potentially effective drugs targeting key genes were obtained from Drug-gene interactions database and visualized with Cytoscape (https://dgidb.genome.wustl.edu/) (DGIdb) [27].

Analysis of immunoinfiltration

IOBR is an R package focused on decoding the immune microenvironment of tumours [28]. In this study, eight immune algorithms were used to decode the Peripheral blood immune microenvironment of CAD samples from the GSE12288 cohort, including CIBERSORT, ESTIMATE, quanTIseq, TIMER, IPS, MCPCounter, xCell and EPIC [29–33]. Relevant features were provided by the IOBR package and the ssGSEA algorithm was used to quantify the features [34].

MR analysis

Mendelian randomisation (MR) analysis of model genes in CAD and diabetes was performed using the TowSanpleMR package [35]. Diabetes outcome data were obtained from the ebi-a-GCST90013891 dataset of the IEU database; CAD outcome data were obtained from the ebi-a-GCST005195 dataset; and gene expression exposure SNP data were also obtained from the IEU database. After p5e<-8 and linkage disequilibrium elimination, gene SNPs were included in the subsequent MR analysis.

Statistical analysis

All statistical analyses were conducted using the R programming language. Comparison of data between different groups were performed with independent-samples t-test or Kruskal test. A P-value less than 0.05 was considered statistically significant.

Results

Abnormal smooth muscle cell function in CAD patients

To examine the differences in cell subset composition between CAD patients and normal samples, Single R annotation was applied to cells from CAD patients in the GSE131778 dataset and normal myocardial tissue in the GSE121893 dataset. The results, depicted in Fig. 1A, reveal a disproportionately high number of smooth muscle cells in CAD patients, with the cell subset composition ratio displayed in Fig. 1B. Differential enrichment analysis conducted for each cell subset indicated that smooth muscle cells are primarily enriched in pathways related to myelination and tissue regeneration in the central nervous system, with SOX2 and SOX10 identified as marker genes (Fig. 1C). Trajectory analysis suggested that various cells in CAD patients, including chondrocytes and tissue stem cells, have the potential to differentiate into smooth muscle cells (Fig. 1D). Cell communication

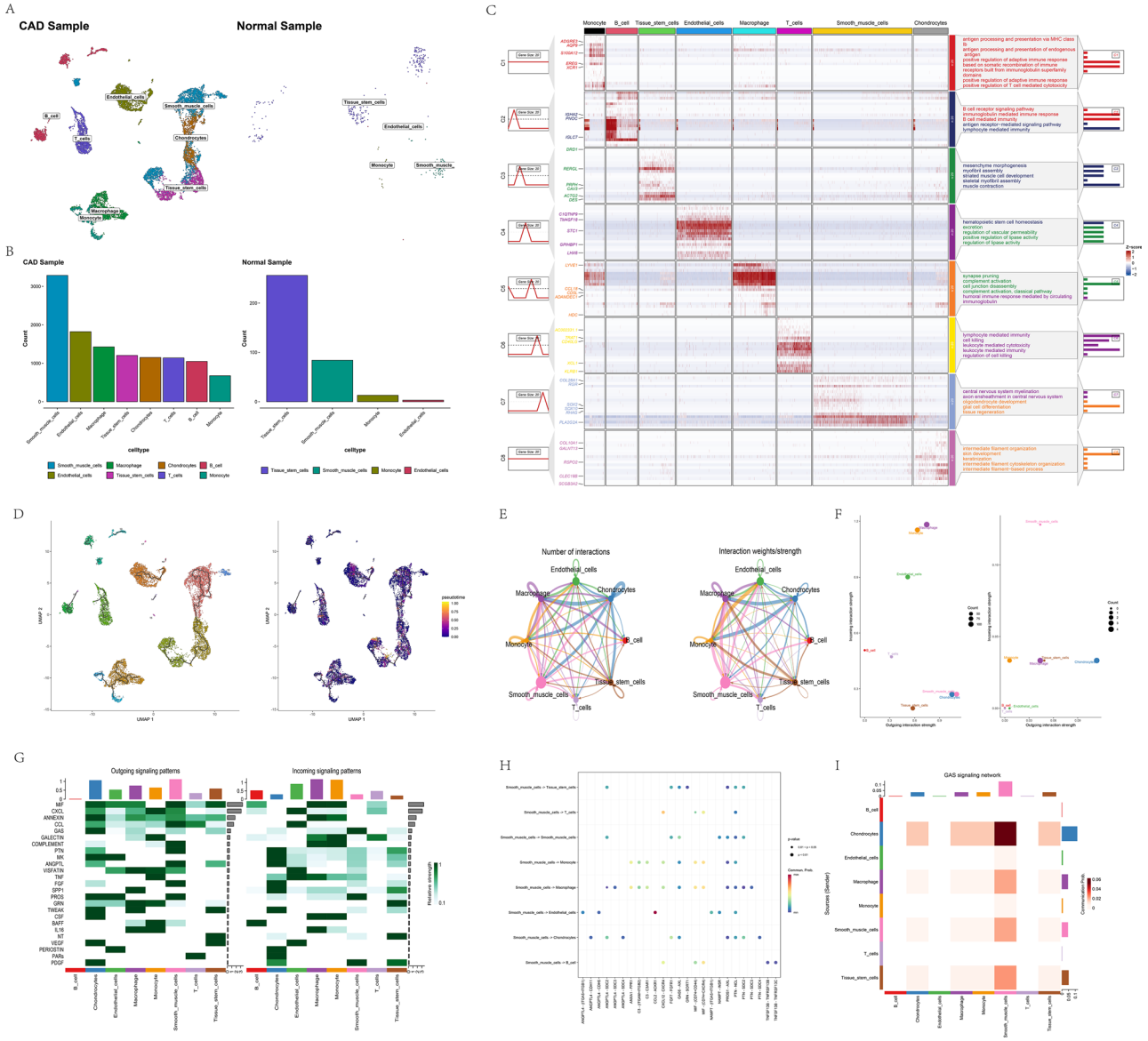


Fig. 1 Abnormal function of smooth muscle cells in CAD patients. (**A**) Cell annotation in CAD patients and normal samples. (**B**) Histogram showing the proportion of cells in CAD patients versus normal samples. (**C**) Differential genes and enrichment heat maps for various cell subsets in CAD patients. (**D**) Cell subset trajectory analysis in CAD. (**E**) Network diagram of cell interaction intensity. (**F**) Bubble plot showing afferent and efferent intensity of signaling pathways. (**G**) Heat map of signaling pathway cellular afferent and efferent intensity. (**H**) Bubble diagram of ligand -receptor interactions between different cells. (**I**) Heat map of GAS signaling pathway interactions among CAD cell subsets

analyses highlighted that smooth muscle cells possess the highest signaling capacity among various pathways, notably in the GAS signaling pathway, demonstrating strong interaction intensities with various cells such as chondrocytes, tissue stem cells, and macrophages (Fig. 1E-I).

Acquisition of hub gene

To identify hub genes in smooth muscle cells from CAD patients, we conducted differential analysis on these cells. As illustrated in Fig. 2A, we identified 2797 differential genes under the condition of a log fold change (logfc) threshold of 0.5, selecting only genes highly expressed in smooth muscle cells. Subsequent WGCNA analysis was then performed on the GSE12288 dataset for these differential genes yielded three gene modules. The turquoise module showed the highest correlation with the CADi with a p-value of 9e-7 and a correlation coefficient of 0.18 (Fig. 2B-E). PCA (Principal Component Analysis) was used for dimensionality reduction in GSE12288 to compare the modular genes between CAD and normal samples (Fig. 2F). The differential analysis results, presented in Fig. 2G, revealed 75 differential genes. The PPI analysis of these 75 genes is depicted in Fig. 2H. The MCC (Maximal Clique Centrality) algorithm of the cytoHubba

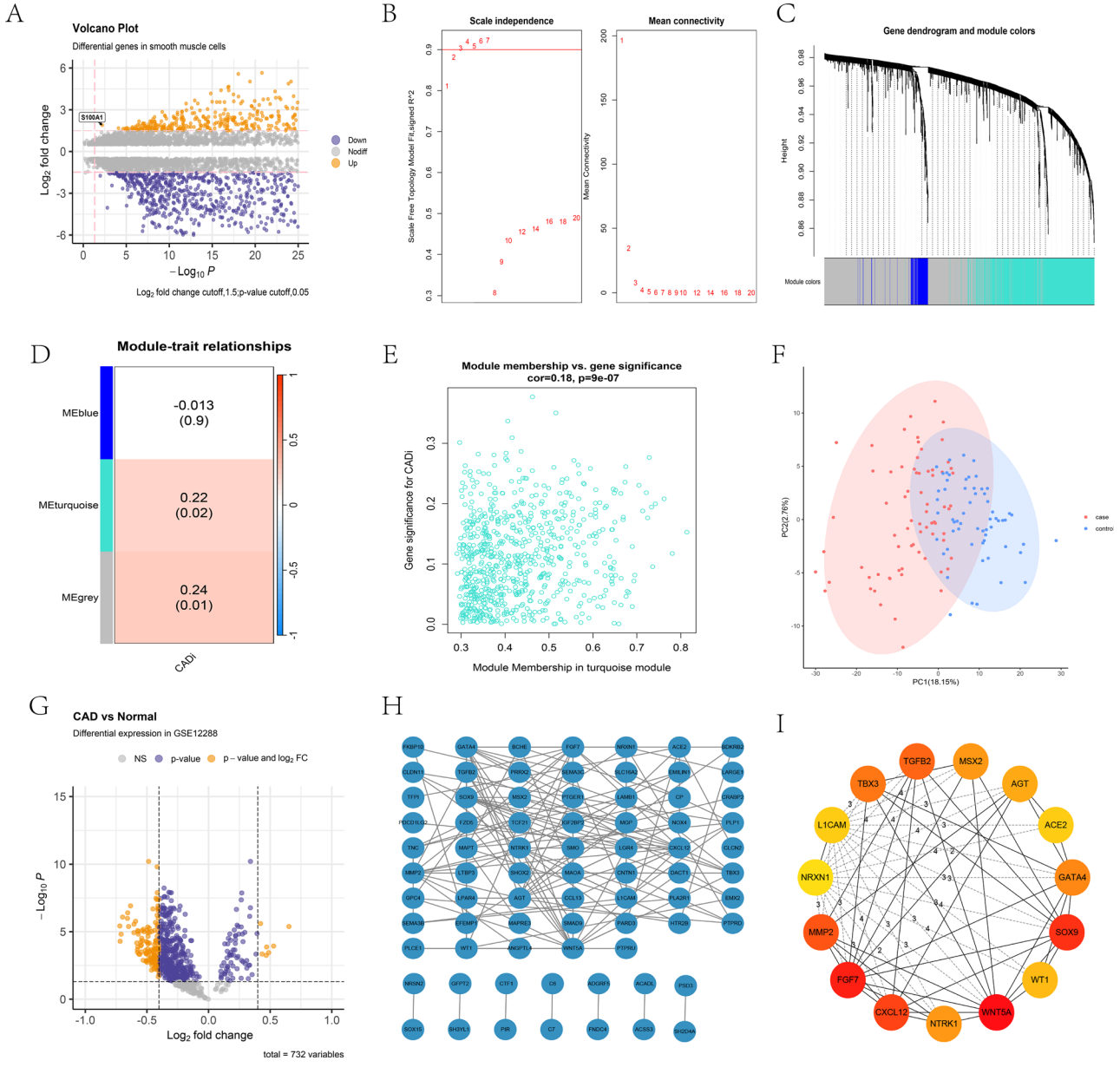


Fig. 2 Identification of the Hub gene. (**A**) Volcano plot of differences between smooth muscle cell subsets and other subsets in CAD patients. (**B**) Selection of the soft threshold. (**C**) Choice of module. (**D**) Association between the module and the CADi. (**E**) Scatter diagram of the association between genes in the turquoise module and CADi. (**F**) PCA results of turquoise module genes between CAD and normal samples in GSE12288 dataset. (**G**) Differential analysis volcano plot. (**H**) PPI network map of differential genes.(**I**) Network maps of the first 15 most related genes obtained using CytoHubba software

software was employed to identify the top 15 key hub genes, with their interaction network displayed in Fig. 2I.

Correlation analysis of hub genes

We conducted correlation analysis on hub genes across three datasets (GSE12288, GSE20680, and GSE221911), with results depicted in Fig. 3A-C. These analyses demonstrated a positive correlation among the hub genes overall (correlation > 0). Additionally, the correlation analysis between hub genes and CADi is shown in Fig. 3D, revealing a significant positive co-expression relationship between FGF7, TBX3, AGT, and CADi ($p < 0.05$, correlation > 0).

Enrichment Analysis of Hub Gene

GSVA analysis was performed on hub genes within the HALLMARK pathway gene set, with the results presented in Fig. 4A. This analysis revealed that the hub gene has a positive expression relationship with the KARS down-regulation pathway, APICAL_JUNCTION, and MYOGENESIS. Conversely, it exhibits a negative expression relationship with DNA_REPAIR, JAK_STAT3, and

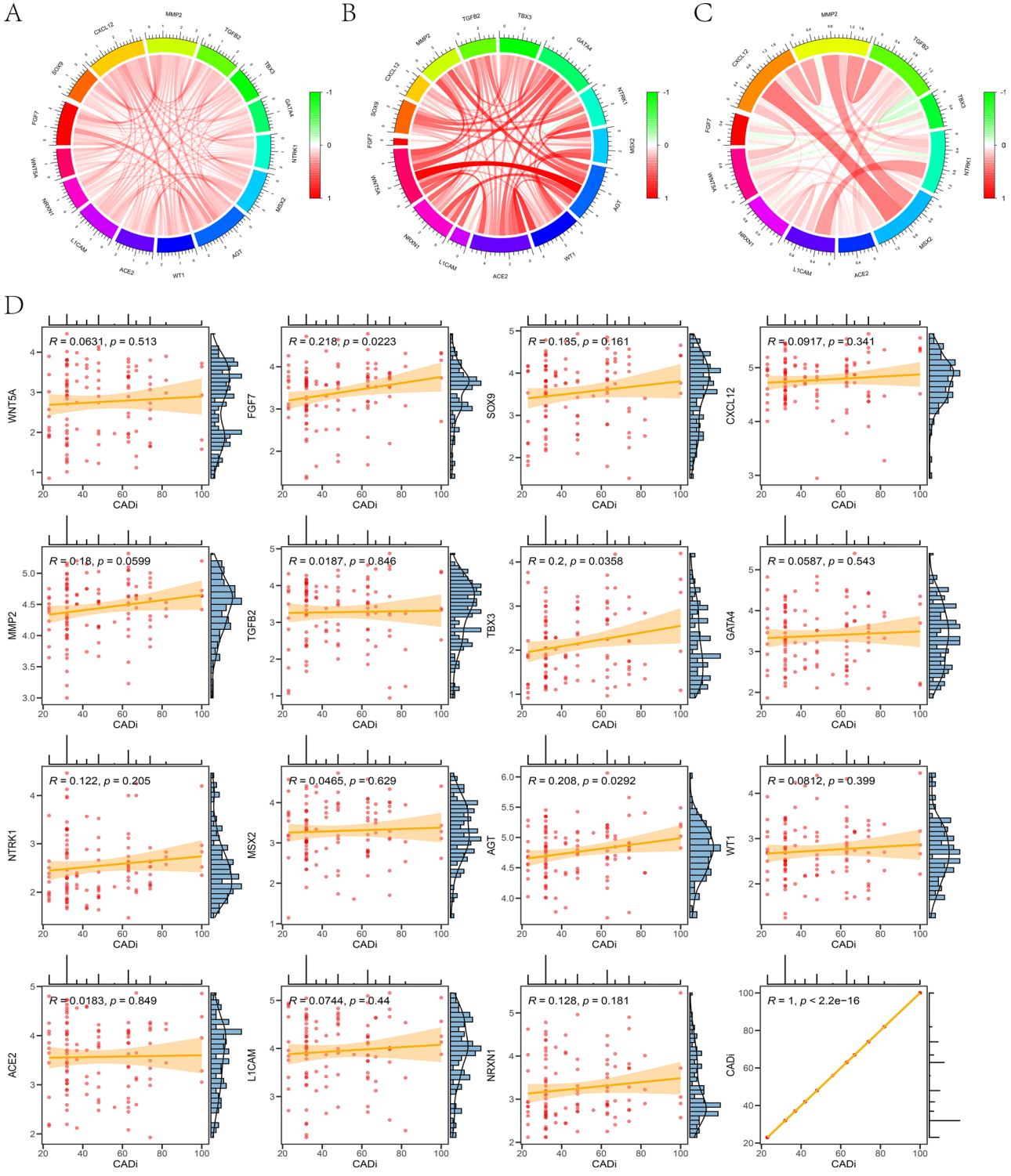


Fig. 3 Correlation analysis. (**A**) Hub gene correlation chord map in the GSE12288 dataset. (**B**) Hub gene correlation chord map in the GSE20680 dataset. (**C**) Hub gene correlation chord map in the GSE221911 dataset. (**D**) Scatter plot of the correlation between the Hub gene and CADi

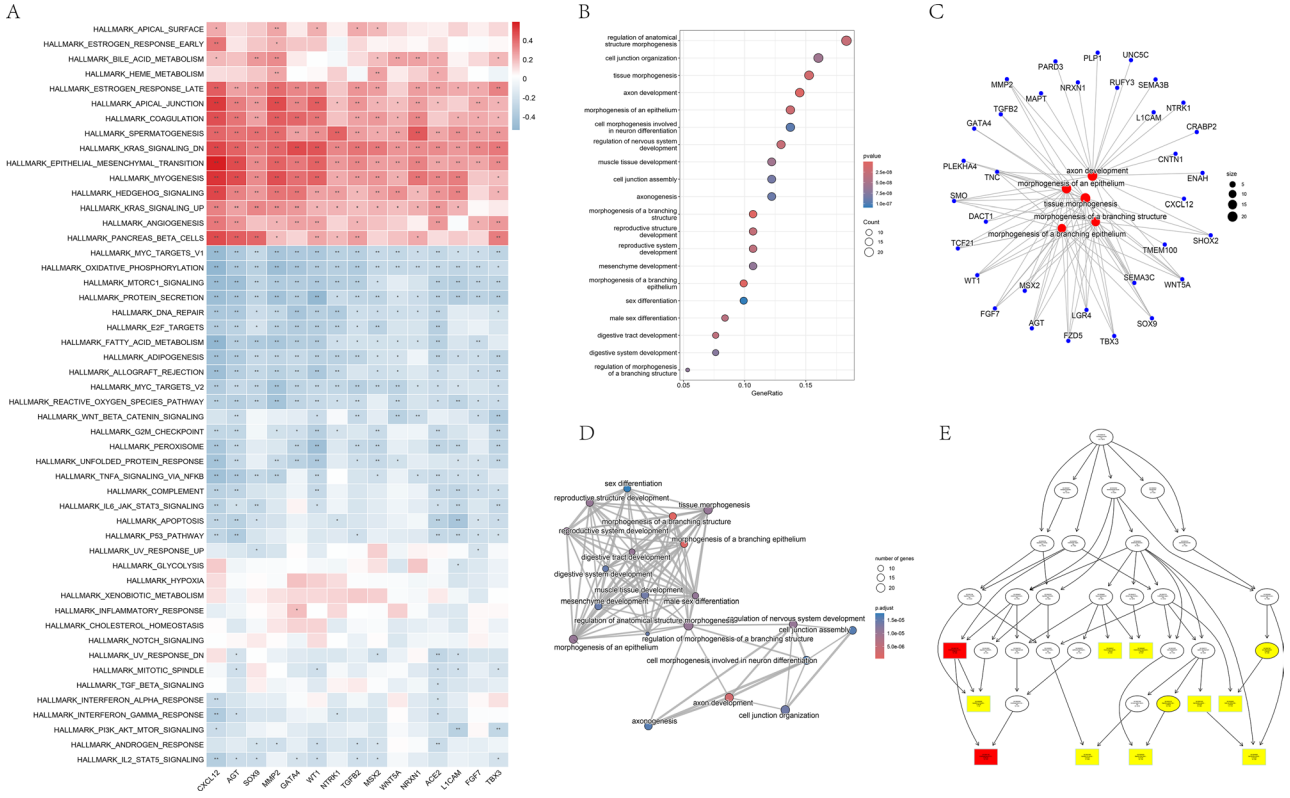


Fig. 4 Enrichment analysis. (**A**) Heat map of the relationship between the Hub gene and HALLMARK pathway. (**B**) Bubble diagram of enrichment analysis. (**C**) Network diagram of enrichment analysis. (**D**) Interaction of pathways. (**E**) Node tree of enrichment analysis

MYC-related pathways. Additionally, GO enrichment analysis indicated that the hub gene is primarily enriched in pathways associated with cell death, regulation of anatomical structure, tissue morphology, and axonal development. Notably, the regulation of structural morphology features prominently as a key regulatory pathway among all pathways studied (Fig. 4B-E).

Machine learning model further verifies hub gene's key role in CAD

To further assess whether the Hub gene can serve as a key target for predicting CAD occurrence, we constructed a model using 15 machine learning algorithms based on the Hub gene. The C-index results, displayed in Fig. 5A, indicate that the combination of GBoost and LASSO yielded the highest C-index (C-index=0.802) in the 10-fold cross-validation model. The C-index in most models hovered around 0.6, suggesting that the Hub gene could be effective biological targets for predicting the occurrence of CAD. In the variable importance results, genes such as CXCL12, GATA4, ACE2, AGT, FGF7, MMP2, NTRK1, and TGFB2 demonstrated strong importance in the model (Fig. 5C).

FGF7 could be a key target for CAD and T2DM

To explore the biological association between T2DM and CAD and identify key targets, we performed PCA dimensionality reduction and differential analysis on diabetes and lean samples in the GSE179455 dataset. The correlation results are presented in Fig. 6A and B. At a significance level of $p<0.05$ and $|\log FC|>0.3$, we identified 758 differential T2DM genes. After intersecting these with Hub genes, three intersection genes were obtained: SOX9, FGF7, and L1CAM (Fig. 6C). The expression differences of these intersection genes among diabetic, lean, and obese samples are shown in Fig. 6D, with FGF7 being highly expressed in diabetic samples and reaching statistical significance ($p<0.05$). The relationship between the expression of intersection genes and CADi is depicted in Fig. 6E and F, revealing a significant positive expression relationship between FGF7 and CADi ($p<0.05$, cor>0), while the other two genes did not show statistical differences. Thus, FGF7 is selected as a key target for both CAD and T2DM.

Immunoassay of FGF7

We conducted immunoassays on CAD samples to investigate the immune mechanisms associated with the FGF7 gene in CAD. Heat maps indicate that high expression of FGF7 results in low infiltration levels of CD8+T

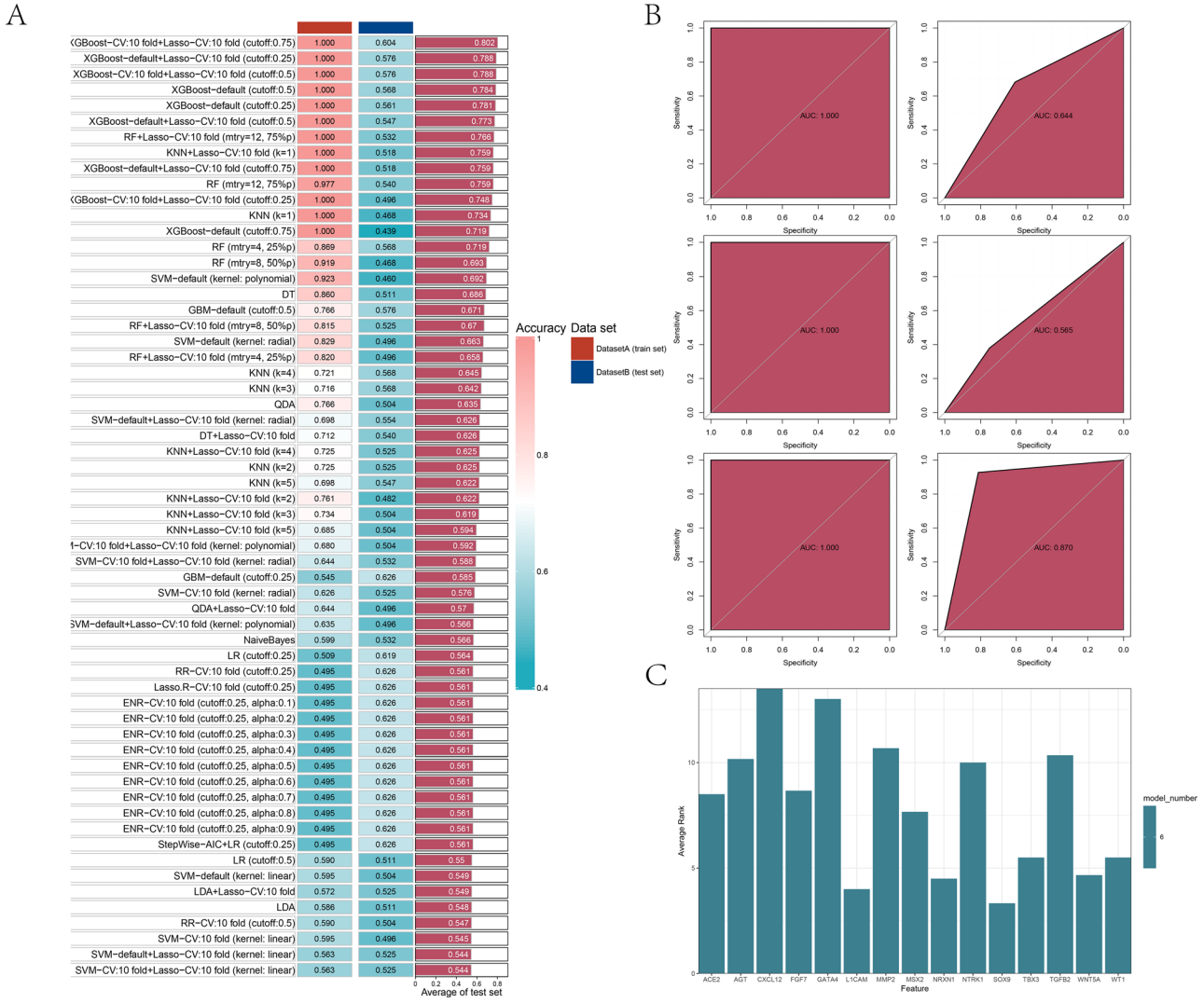


Fig. 5 Machine learning analysis. (**A**) C-index heat map for the Hub gene across different models. (**B**) ROC curve for model performance. (**C**) Histogram of variable importance for the Hub gene in the model

cells, CD4+T cells, monocytes, macrophages, and NK cells (Fig. 7A and C). Additionally, immunomodulator thermograms reveal that high expression of FGF7 is associated with low expression of various types of immunomodulatory molecules, including costimulatory, coinhibitory, and receptor molecules (Fig. 7B). The co-expression butterfly diagram further illustrates that FGF7 co-expresses with a variety of immune cells and immunomodulatory molecules, such as costimulatory molecules CD40 and CD80, and immune cells like NK cells and CD4+T cells (Fig. 7D).

Drug sensitivity analysis of FGF7

To identify potential drugs targeting FGF7 in CAD patients, we first grouped samples based on FGF7 expression levels. The PCA plot of the grouped samples (Fig. 8A) shows that the two groups can be distinguished to some extent. Next, we identified differentially expressed genes between the two groups, as shown in Fig. 8B. Using the thresholds of logFC > 0.8 & P.Value < 0.05 for positively correlated genes and logFC < -0.6 & P.Value < 0.05 for negatively correlated genes, we identified four significantly positively correlated genes (PDLIM3, ZNF674, LHCGR, and ZNF442) and four significantly negatively correlated genes (SMAD7, QDPR, SNX6, and PSMD5). The correlation analysis of these differentially expressed genes is shown in Fig. 8C-D, further validating the differential expression levels. To explore potential effective drugs targeting these key genes, we obtained the gene-chemical interaction network from the Drug-Gene Interaction Database (DGIdb) and visualized it using Cytoscape software, as shown in Fig. 8E-F. Notably, COMPOUND 34 and DTRIM24 showed strong interaction scores with FGF7 (interaction scores of

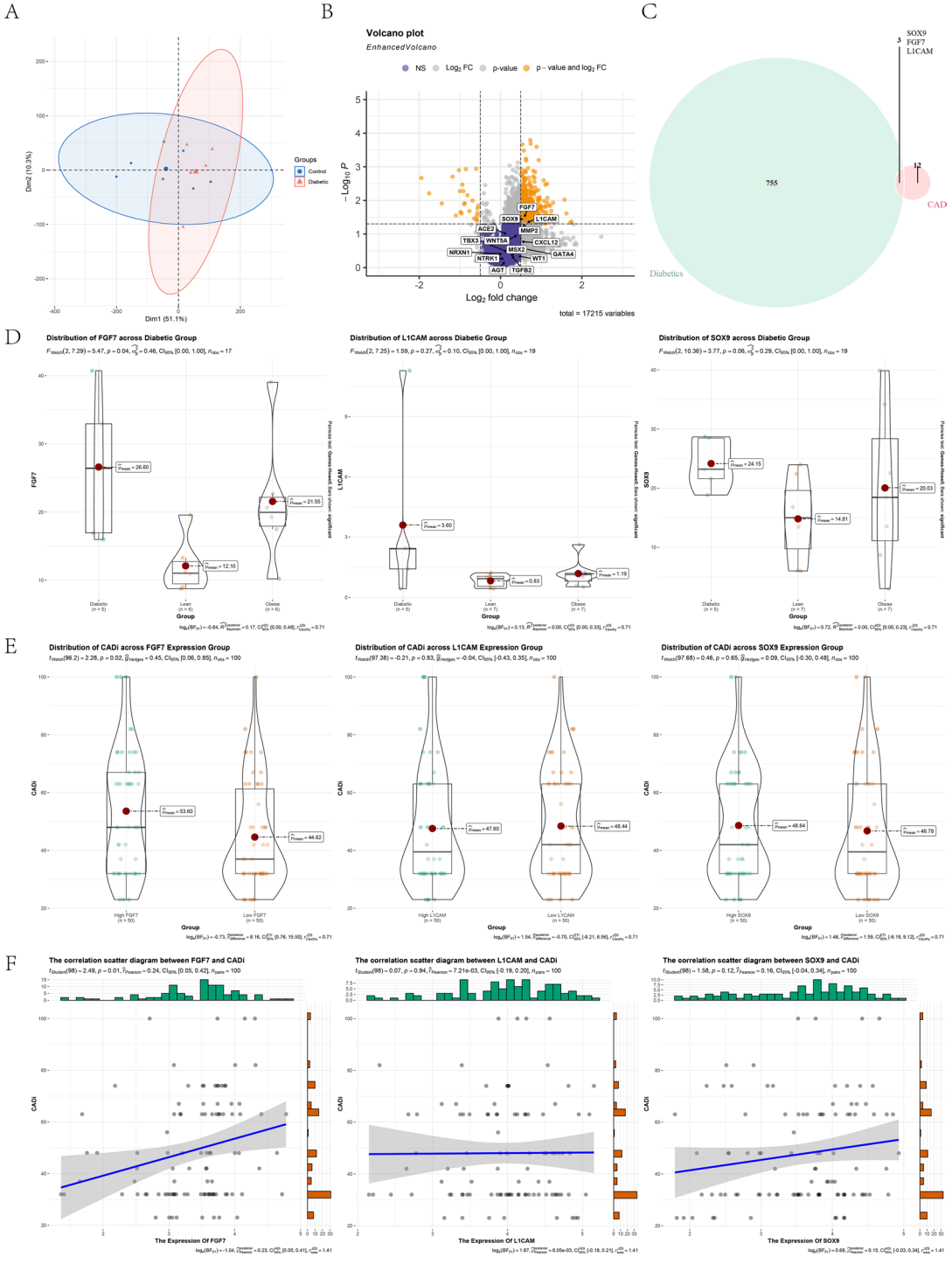


Fig. 6 Key gene screening.(**A**) PCA dimensionality reduction performed on samples from the GSE179455 dataset. (**B**) Volcano plot of differential analysis. (**C**) Intersection of differential genes and Hub genes. (**D**) Expression of intersecting genes in diabetic, lean, and obese individuals. (**E**) Differential expression between CADi and intersecting genes. (**F**) Co-expression relationship between CADi and intersecting genes

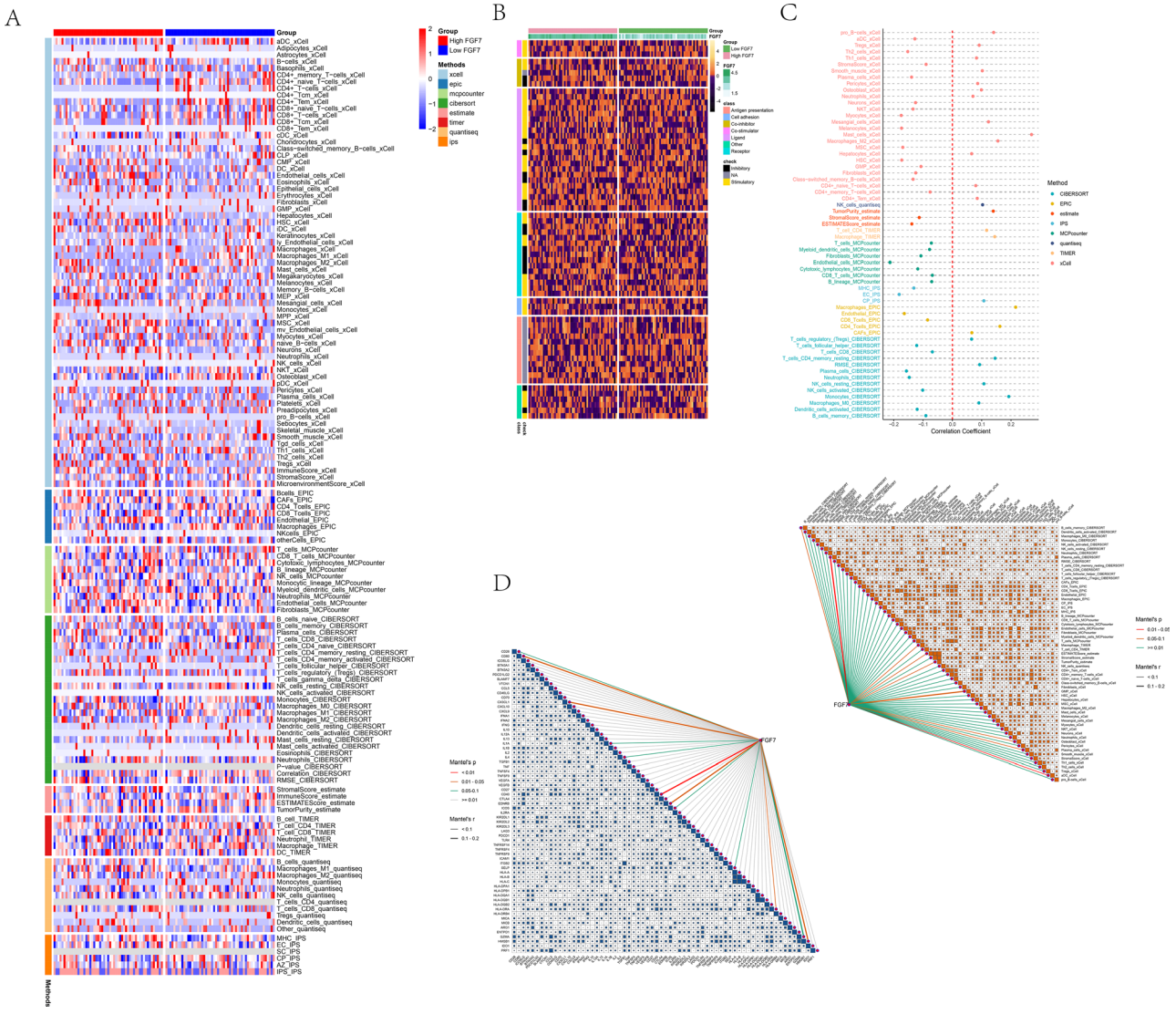


Fig. 7 Immunoassay. (**A**) Heat map showing the relationship between immune cells and FGF7 expression using eight immune microenvironment decoding methods. (**B**) Expression heat map of FGF7 and different immune molecules. (**C**) Bar graph showing correlation between FGF7 and immune cell infiltration score. (**D**) Co-expression butterfly map of FGF7 with immune cells and immune molecules

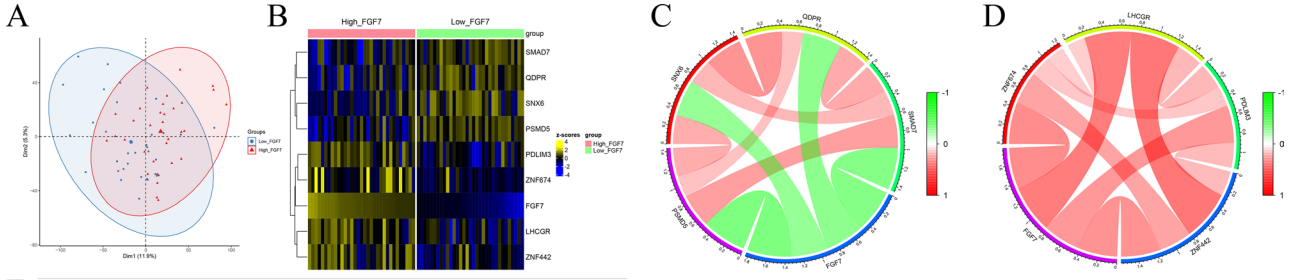
13.126918 and 26.253836, respectively), suggesting that these drugs may play an important role in CAD patients by affecting FGF7 expression. Additionally, drugs associated with the positively correlated gene LHCGR are mainly related to sex hormones, while some drugs associated with the negatively correlated gene SMAD7, such as Alteplase, can be used to treat heart disease.

ceRNA network construction

To further investigate the upstream regulatory targets of FGF7 gene, we obtained first the miRNA targets of FGF7 against two databases miRNet and miRTaRBase, and then the 20 intersection miRNAs (Fig. 9B). Subsequently, for these 20 miRNAs, their related LncRNAs were predicted in the LncACTdb3.0 dataset, and then visualized in Cytoscape. The results are shown in Fig. 9A. The MCC algorithm was used to select key regulatory networks using cytoHubba software. As shown in Fig. 9C, the important miRNAs regulating FGF7 were hsa-miR-15a-5p, hsa-miR-373- 3p, hsa-miR-20a-5p and hsa-miR-372- 3p, and their related LncRNAs were also displayed.

Functional verification of FGF7

To further verify the function of FGF7 in CAD, we first performed MR analysis on the Hub gene, and found that the FGF7-related SNP was deleted under the condition of removing linkage disequilibrium and p5e-8, but the L1CAM gene was positive in type 1 diabetes (p 0.05, OR 1, Fig. 10A and B). At the single-cell level, the



gene	drug	regulatory.approval	indication	interaction.score
FGF7	COMPOUND 34 [PMID: 25974391]	Not Approved		13.12691781
FGF7	DTRIM24	Not Approved		26.25383562
LHCGR	GONADOTROPIN, CHORIONIC	Approved		3.28172945
LHCGR	CHEMBL:CHEMBL429095	Not Approved		0.06563459
LHCGR	LUTROPIN ALFA	Approved	Fertility Agents	16.40864726
LHCGR	LH (MENOTROPINS)	Approved		3.28172945
LHCGR	CHORIOGONADOTROPIN ALFA	Approved	Fertility Agents	6.56345890
LHCGR	GOSERELIN ACETATE	Approved	Hormonal,Antineoplastic Agents	0.72927321
LHCGR	GTX-758	Not Approved	antineoplastic agent	1.64086473
LHCGR	RECOMBINANT HUMAN CHORIONIC GONADOTROPIN	Approved	antineoplastic agent	3.28172945
LHCGR	CTP-37-DT VACCINE	Not Approved	vaccine,antineoplastic agent	3.28172945
LHCGR	CHEMBL:CHEMBL505670	Not Approved		0.09376370
LHCGR	RELUGOLIX	Approved		1.64086473
LHCGR	CHEMBL:CHEMBL448753	Not Approved		6.56345890
LHCGR	[125I]LH	Not Approved		3.28172945
LHCGR	[125I]CHORIONIC GONADOTROPIN (HUMAN)	Not Approved		3.28172945
LHCGR	UROFOLLITROPIN	Approved	Fertility Agents,fertility agent	2.46129709
LHCGR	HCG	Not Approved		3.28172945
SMAD7	URACIL	Approved	antineoplastic agent	2.18781963
SMAD7	ALTEPLASE	Approved	Thrombolytic agent	1.31269178
SMAD7	MONGERSEN	Not Approved		13.12691781
SMAD7	CAPECITABINE	Approved		0.48618214
PSMD5	BORTEZOMIB	Approved	antineoplastic agent	0.30527716
PSMD5	CARFILZOMIB	Approved	antineoplastic agent	0.43756393

Fig. 8 Drug sensitivity analysis. (**A**) PCA plot of samples grouped by the top and bottom thirds of FGF7 expression in CAD patients. (**B**) Heatmap of differentially expressed genes. (**C**) © Co-expression network of FGF7 and negatively correlated genes. (**D**) Co-expression network of FGF7 and positively correlated genes. (**E**) Screening of targeted drugs related to differentially expressed genes of FGF7. (**F**) Network diagram of targeted drugs for differentially expressed genes of FGF7 (red represents drugs for positively correlated genes, blue represents drugs for negatively correlated genes)

expression of FGF7 in smooth muscle cells was divided into FGF7+SMC and FGF7-SMC, and KEGG metabolite analysis was performed. As shown in Fig. 10C, FGF7+SMC cells were more involved in nicotinic acid ester and nicotinamide metabolism, retinol metabolism, alanine, aspartic acid and glutamic acid metabolism, and glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate metabolic pathways than FGF7-SMC cells. FGF7+SMCs showed stronger communication intensity in cellular communication, especially for monocyte/macrophage interactions (Fig. 10D and E). Trajectory analysis showed that FGF7 expression decreased gradually as smooth muscle cells grew and differentiated (Fig. 10F and G).

Discussion

Approximately 2,200 people die from cardiovascular disease (CVD) each day, which translates to about one person every 40 s, and annually, more than one million individuals experience heart attacks, marking it as a global public health issue. Typical symptoms of coronary artery disease (CAD) include chest pain and shortness of breath during exertion; however, these symptoms alone are not definitive for CAD, as most chest pain results from musculoskeletal issues [36]. Studies have identified age, gender, smoking, and increased cholesterol levels as risk factors for the development of CAD [37], yet there remains a significant need for molecular biological markers to predict the occurrence and progression of CAD [16, 38]. Advances in sequencing and chip technology, along with deeper understanding of CAD's molecular biology, are enabling the use of new molecular markers for clinical prognostic assessments and personalized treatment.

The FGF7 gene encodes keratinocyte growth factor 7 (KGF), also known as alveolar epithelial growth factor. This cytokine is crucial for the health and repair of various epithelial tissues [39]. In humans, FGF7 primarily acts through its receptors on epithelial cell surfaces to regulate biological processes including cell proliferation, migration, and differentiation. Additionally, FGF7 is significant in smooth muscle cells, which are vital components of many organs and tissues such as blood vessels, the digestive tract, respiratory system, and reproductive system. The roles of FGF7 in these cells largely pertain to cell proliferation, migration, and extracellular matrix synthesis [40].

In CAD, research indicates that smooth muscle cells are important in the pathogenesis and progression of the condition, primarily through three mechanisms: First,

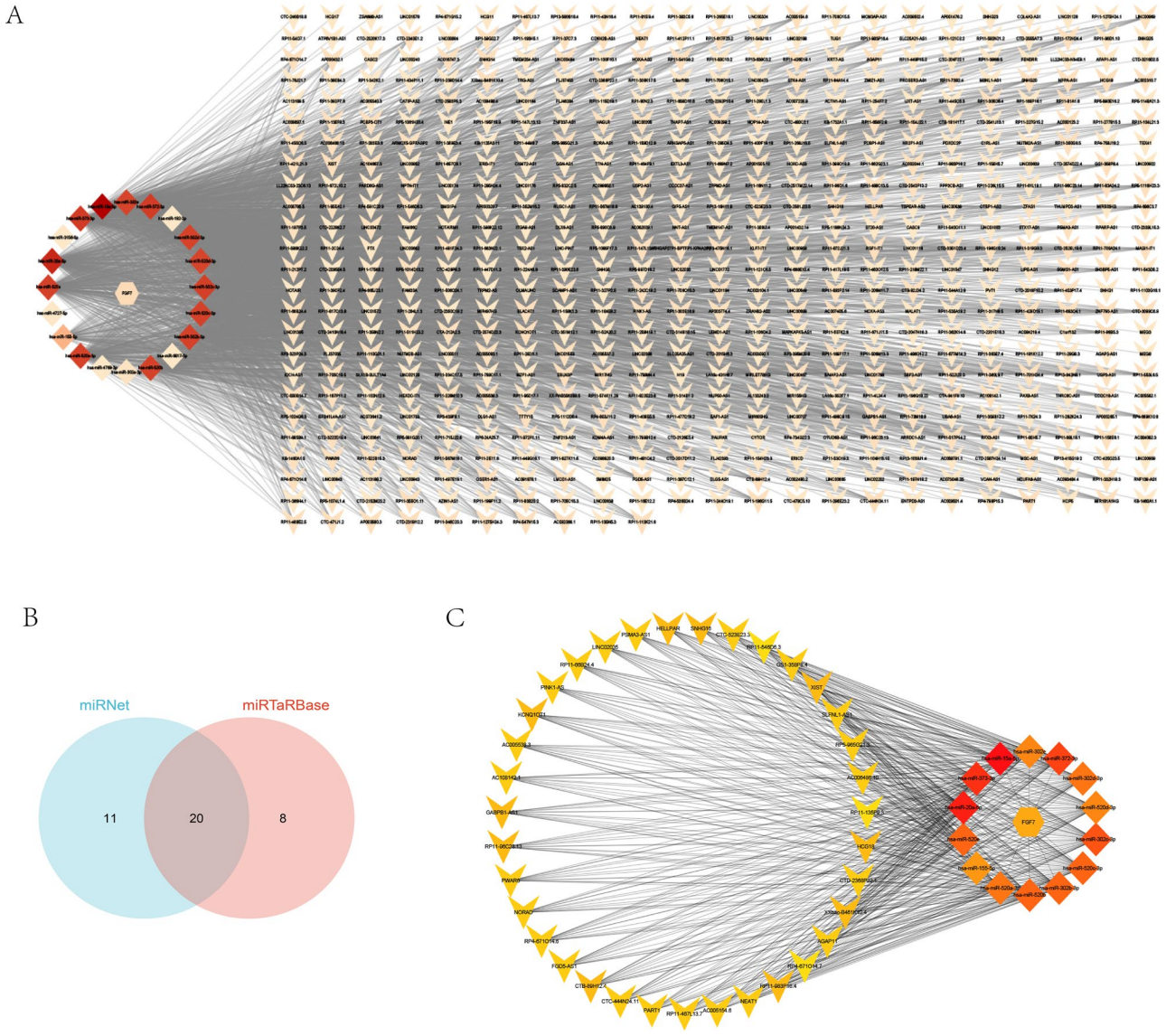


Fig. 9 CeRNA network of FGF7. (**A**) Network diagram showing interactions between FGF7, miRNAs, and LncRNAs. (**B**) miRNA targets of FGF7 intersecting in miRNet and miRTaRBase databases. (**C**) Key ceRNA network obtained using cytoHubba software

within the structure and function of the coronary artery wall, such as in atherosclerosis development, smooth muscle cells are damaged and activated, contributing to the formation and evolution of pathological plaques. Second, smooth muscle cells regulate the vasoconstriction and relaxation of the coronary artery. In CAD, abnormal activation of these cells can lead to improper vasoconstriction, impacting coronary artery hemodynamics. Lastly, the proliferation and migration of smooth muscle cells are crucial for the stability and progression of atherosclerotic plaques. Following arterial wall injury, smooth muscle cells can proliferate and migrate to the plaque areas, aiding in plaque formation and repair [41–43].

From the single-cell transcription profile of CAD myocardium, we initially identified abnormal subsets of smooth muscle cells and analyzed their functions. These cells were predominantly enriched in tissue regeneration and myelination pathways. Notably, tissue stem cells and chondrocytes were found to have the potential to differentiate into smooth muscle cells. Differential analysis and WGCNA were conducted to identify CAD-index-related hub genes in smooth muscle cells, including WNT5A, FGF7, SOX9, CXCL12, MMP2, TGFB2, TBX3, GATA4, NTRK1, MSX2, AGT, WT1, ACE2, L1CAM, and NRXN1. The average C -index in machine learning was around 0.6, confirming the relatively significant role these hub genes play in predicting CAD onset. Subsequently, from the biological link between T2DM and CAD, we

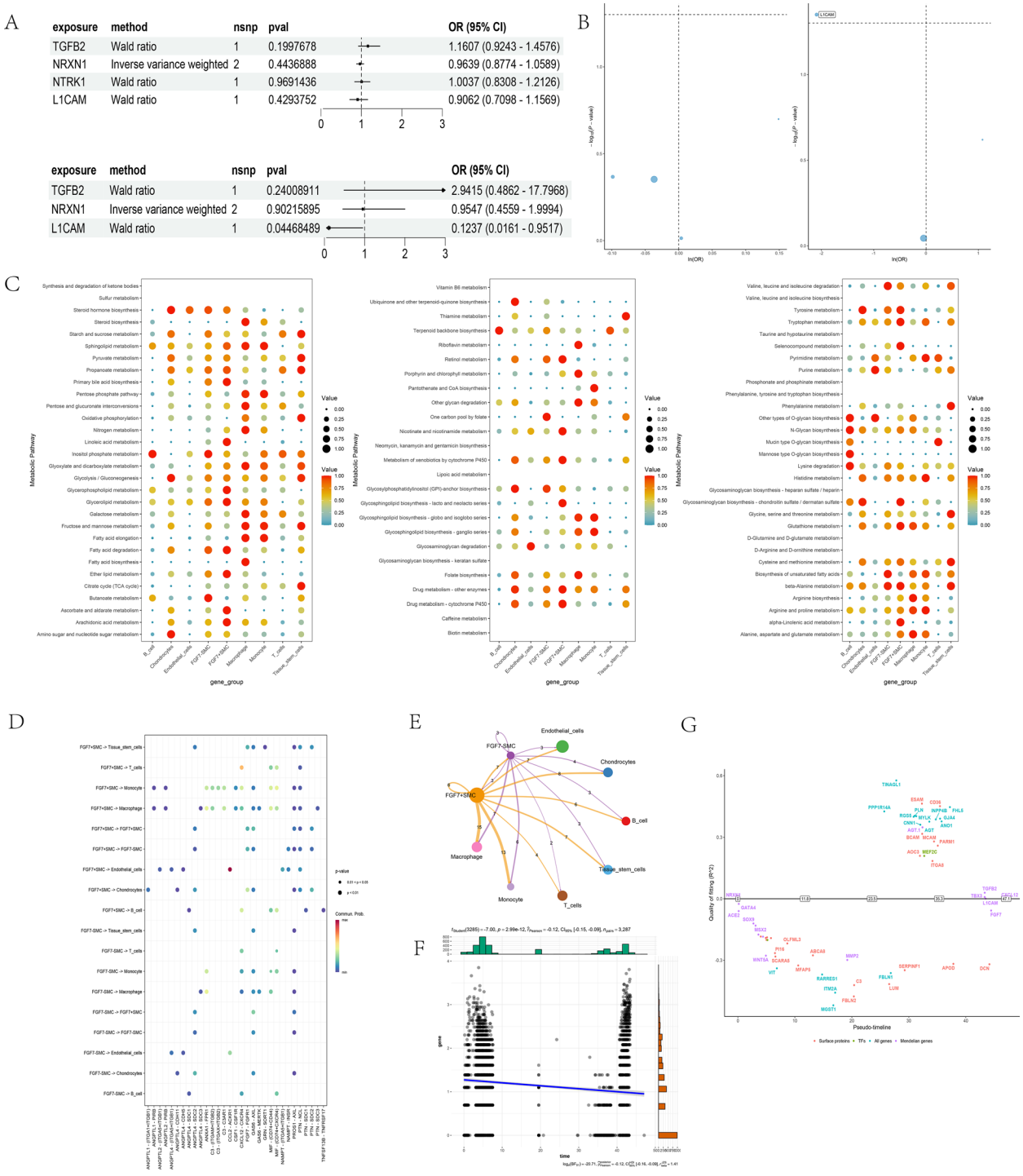


Fig. 10 functional verification FGF7. (**A**-**B**) MR analysis of Hub gene in CAD and Diabetic diseases. (**C**) Differential bubble diagram of FGF7 expression and metabolites of smooth muscle cells in CAD. (**D**)FGF7 expression and signal molecule interaction bubble diagram of smooth muscle cells in CAD. (**E**) The network map of interaction intensity between FGF7 expression of smooth muscle cells and other cells in CAD. (**F**) The expression trend of FGF7 along with the growth and differentiation of smooth muscle cells. (**G**) The expression of model genes with the growth and differentiation of smooth muscle cells

pinpointed a key gene, FGF7, which is crucial for both conditions.

Our study identified FGF7 as a risk factor for the development of coronary artery disease (CAD), showing a correlation coefficient of 0.24 ($p=0.01$) and elevated expression in Type II diabetes mellitus (T2DM) compared to normal subjects (logfc>0.3, $p<0.05$). Immunoassay results in CAD samples indicated that FGF7

expression negatively correlated with the infiltration of CD4+T cells, CD8+T cells, NK cells, and other immune killer cells, suggesting that FGF7 may promote CAD development by inhibiting immune cell recruitment. Drug sensitivity analysis identified key drugs that can affect FGF7-related targets, including the thrombolytic drug Alteplase, and the proteasome inhibitors Bortezomib and Carfilzomib.

Further, we predicted upstream targets of FGF7, constructed ceRNA networks, and identified significant miRNA targets including hsa-miR-15a-5p, hsa-miR-373-3p, hsa-miR-20a-5p, and hsa-miR-372- 3p that could affect FGF7 expression. Analysis also showed that cells expressing FGF7+in smooth muscle (FGF7+SMC) had stronger interaction intensities with other cells and decreased expression as smooth muscle cells grew and differentiated. Metabolic analysis indicated that FGF7+SMC cells were more enriched in pathways such as nicotinic acid ester and nicotinamide metabolism, retinol metabolism, and alanine, aspartate, and glutamate metabolism than FGF7-SMC cells.

Despite its significance and impact, this study has several limitations. ①This study mainly relied on bioinformatics analysis and transcriptome data in public databases, which may limit the comprehensiveness of the data and the depth of experimental verification. ②Although we have screened out FGF7 gene through Various bioinformatics methods, the specific mechanism of action of the gene in CAD and type II diabetes has not been fully verified in experiments. ③We have identified drugs that may affect the expression of the FGF7, but the clinical efficacy and safety of these drugs have not yet been verified in large-scale clinical trials. ④This study only used common bioinformatics analysis methods to mine common targets for CAD and type 2 diabetes. The single research method may lead to single analysis results. Future studies should aim to overcome these limitations by expanding sample sources, improving data analysis methods, and strengthening experimental validation, thus providing more robust support and guidance for CAD research and clinical practice.

Conclusions

In conclusion, our study demonstrates that FGF7 expression is reduced in CAD and exhibits a positive correlation with CADi. Additionally, FGF7 expression level emerges as a significant feature in our machine learning models. We found that FGF7 expression is linked to the activity of immune killer cells and sensitivity to statins. Moreover, the regulation of FGF7 expression may be influenced by miRNAs and LncRNAs. Thus, FGF7 plays a role in the pathophysiology of CAD and holds potential therapeutic value.

Acknowledgements
Not applicable.

Author contributions
LX and HX conceived and designed the study. MYZ and LX performed the analysis procedures. ST and YZQ analyzed the results; LX and YZQ contributed analysis tools. LX and YZQ contributed to the writing of the manuscript. All authors reviewed and edited the manuscript.

Funding
This study was Sponsored by Natural Science Foundation of Chongqing, China (Project No.: cstb2024nscq-msx1017).

Data availability
The datasets generated and/or analysed during the current study are available in the GEO Database (http://www.ncbi.nlm.nih.gov/geo) repository (accession number：GSE12288, GSE20680, GSE221911).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 13 May 2024 / Accepted: 15 October 2024
Published online: 05 November 2024

References

1. Benjamin EJ, Chiuve SE, Das SR. Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, de Ferranti SD, Floyd J. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation. 2017;135(10):e146-e603.
2. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, Chiuve SE, Cushman M. Heart Disease and Stroke Statistics-2018 update: a Report from the American Heart Association. Circulation. 2018;137(12):e67–492.
3. Yang DZ. BCL3-PVRL2-TOMM40 SNPs, gene-gene and gene-environment interactions on dyslipidemia. Sci Rep. 2018;8(1):6189. Lin WX.
4. Jeck WR, Wang K, Burd CE, Marzluff WF. Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA. 2013;19(2):141 – 57.
5. Zhou HX. Jia EZ. Comprehensive Analysis of mRNA expression profiling and identification of potential diagnostic biomarkers in coronary artery disease. ACS Omega. 2021;6(37):24016–26. Zhang S.
6. Rosenberg S, Beineke P, Wingrove JA, Sager PT. Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, Sager PT Sehnert AJ. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. Ann Intern Med. 153(7):425 – 34.
7. Wingrove JA, Sehnert AJ, Elashoff MR, Buellesfeld L. Daniels SE, Sehnert AJ, Tingley W, Elashoff MR, Rosenberg S, Buellesfeld L. Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis. Circ Cardiovasc Genet. 2008;1(1):31 – 8.
8. Wahli W. PPARs and microbiota in Skeletal Muscle Health and Wasting. Int J Mol Sci. 2020;21(21):8056. Duszka K.
9. Kolnes AJ. The role of skeletal muscle glycogen breakdown for regulation of insulin sensitivity by exercise. Front Physiol. 2011;2:112. Lai YC.
10. Hegarty BD, Ye J, Kraegen EW. Furler SM, Ye J, Cooney GJ, Kraegen EW. The role of intramuscular lipid in insulin resistance. Acta Physiol Scand. 178(4):373 – 83.
11. Sarwar ERF, Gao N, Seshasai P, Gobin SR, Kaptoge R. S. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375 (9733): 2215-22.

12. Ofstad AP. Myocardial dysfunction and cardiovascular disease in type II diabetes. Scand J Clin Lab Invest. 2016;76(4):271–81.
13. Kraus WE. Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. PLoS ONE. 2009;4(9):e7037. Chibout SD.
14. Voros S. Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients. BMC Med Genomics. 2011;4:26. Kraus WE, et al.
15. Rosenberg S. Wingrove JA. A whole blood gene expression-based signature for smoking status. BMC Med Genomics. 2012;5:58. Kraus WE.
16. Shah P. Wargowsky R. RNAseq profiling of blood from patients with coronary artery disease: signature of a T cell imbalance. J Mol Cell Cardiol Plus. 2023;4:100033. Falk Z.
17. Vyas V, Blythe H, Wood EG, Sandhar B, Sarker SJ, Balmforth D, Ambekar SG, Yap J, Edmondson SJ et al. Obesity and diabetes are major risk factors for epicardial adipose tissue inflammation. JCI Insight. 202;6(16):e145495.
18. Chowdhury RR, Huang X, Sasagawa K, Cheng P., D'Addabbo J, Huang X, Veizades S, Sasagawa K, Louis DM, Cheng P Sokol J. Human Coronary Plaque T Cells Are Clonal and Cross-React to Virus and Self. Circ Res. 2022; 130(10):1510–1530.
19. Wirka RC, Paik DT, Nguyen T, Kundu R, Coller J. Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J Koyano TK. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. Nat Med. 2019; 25(8):1280–1289.
20. Wang L, Yu P, Zhou B, Song J, Li Z, Zhang M, Guo G, Wang Y, Chen X, Han L, Hu S. Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. Nat Cell Biol. 2020;22(1):108–19.
21. Myung P. Nie Q Inference Anal cell-cell Communication Using CellChat Nat Commun. 2021;12(1):1088. Plikus MV.
22. Mikkelsen TS. Dynamics Regulators cell fate Decisions are Revealed Pseudotemporal Ordering Single Cells Nat Biotechnol. 2014;32(4):381–6. Rinn JL.
23. Langfelder P. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. Horvath S.
24. Szklarczyk D, Lyon D, Wyder S, Simonovic M. Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M Doncheva NT. 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.
25. Shannon P, Ozier O, Wang JT, Amin N. Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498 – 504.
26. Benner C. Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. Tanaseichuk O.
27. Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021;22(6):bbab260. Gruener RF.
28. Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, Zhou R, Qiu W, Huang N, Sun L, Li X, Bin J, Liao Y, Shi M, et al. IOBR: Multi-omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and signatures. Front Immunol. 2021;12:687975.
29. Selves J. Laurent-Puig P, Sautès-Fridman C, Fridman WH, De Reyniès A. estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17(1):218. Petitprez F.
30. Hoang CD. Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7. Xu Y.
31. Signoretti S. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016;17(1):174. Rodig S.
32. Hackl H. Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18(1):248–62. Rieder D.
33. Aran D. Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18(1):220. Hu Z.
34. Hänzelmann S. Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
35. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
36. Kohn MA, Gupta M. Kwan E, Gupta M Tabas JA. Prevalence of acute myocardial infarction and other serious diagnoses in patients presenting to an urban emergency department with chest pain. J Emerg Med. 2005; 29(4):383 – 90.
37. Pencina MJ., D'Agostino RB Sr, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. Circulation. 2009;119(24):3078-84.
38. Patel MR, Dai D, Redberg RF, Brindis RG. Peterson ED, Dai D, Brennan JM, Redberg RF, Anderson HV, Brindis RG, Douglas PS. Low diagnostic yield of elective coronary angiography. N Engl J Med. 2010;362(10):886 – 95.
39. Zinkle A. Mohammadi, Structural Biology of the FGF7 subfamily. Front Genet. 2019;10:102. Mohammadi M.
40. Wang L. Integr Bioinformatic Anal Reveals Underlying Mol Mechanism Potential Drugs Pulmonary Arterial Hypertens Aging (Albany NY). 2021;13(10):14234–57. Huang X.
41. Iyer D. Smad3 regulates smooth muscle cell fate and mediates adverse remodeling and calcification of the atherosclerotic plaque. Nat Cardiovasc Res. 2022;1(4):322–33. Nagao M.
42. Sol-Church K. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. Nat Genet. 2022;54(6):804–16. Farber E.
43. Yang Q. Genetic regulation of SMC Gene expression and splicing predict causal CAD genes. Circ Res. 2023;132(3):323–38. Gilani H.

Publisher's note