

Epithelial-Mesenchymal Transition-Related lncRNAs And SNAI2 Are Potential Biomarkers in Coronary Artery Disease

Xiang Xu

The Second Affiliated Hospital of Kunming Medical University

Renchao Zou

The Second Affiliated Hospital of Kunming Medical University

Xiaoyong Liu

The Second Affiliated Hospital of Kunming Medical University

Qianqian Su (✉ suqianqian@kmmu.edu.cn)

Kunming Medical University

Research Article

Keywords: coronary artery disease, epithelial-mesenchymal transition, support vector machine reverse feature elimination, least absolute shrinkage and selection operator, lncRNA

Posted Date: November 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-940366/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Increasing evidence suggests that epithelial-mesenchymal transformation (EMT) is critical in the development of inflammatory response, atherosclerosis, and coronary artery disease (CAD). However, landscapes of EMT-related lncRNAs and their target genes have not been fully established in CAD.

Methods

lncRNA and mRNA expression profiles obtained from Gene Expression Omnibus (GEO) database were used to identify the differentially expressed mRNAs (DEGs) and lncRNAs (DElncRNAs) between CAD and normal samples. Based on Pearson correlation analysis to identify the EMT-related lncRNAs, the optimal features were identified by receiver operating characteristic (ROC), the least absolute shrinkage and selection operator (LASSO) regression, Support Vector Machine Reverse Feature Elimination (SVM-RFE) algorithms, and logistic regression models were constructed aiming to distinguish CAD from normal samples. The cis and trans-regulatory networks were constructed based on EMT-related lncRNAs. We further estimated the infiltration of the immune cells in CAD patients with the CIBERSORT algorithm, and the correlation between key genes and infiltrating immune cells was analyzed.

Results

In this study, a logistic regression model with powerful diagnostic capability was constructed based on a total of eight EMT-related lncRNAs identified by two machine learning methods. Then, results of the immune analysis revealed three significant immune cell subsets (CD8 T cells, monocytes, and NK cells) in CAD patients and found EMT-related lncRNAs were closely correlated with these immune cell subsets. By Pearson correlation analysis we got 34 “cis” and “trans” genes. Among them, SNAI2, an EMT-TF gene, was found in the trans-regulatory network of EMT-related lncRNAs. Further, through logistic regression and analysis of immune cell infiltration, we found SNAI2 was a potential biomarker for the diagnosis of CAD but also a close correlation between highly expressed SNAI2 and these three immune cell subsets in CAD patients.

Conclusion

In conclusion, these biomarkers have important significance in the diagnosis of CAD patients. Eight EMT-related lncRNAs and SNAI2 can improve our understanding of the molecular mechanism between EMT and CAD.

Introduction

Coronary artery disease (CAD) is a common public health problem, mainly occurring in people over 45 years of age. CAD is now the leading cause of death in the United States, accounting for one in six deaths alone[1]. The American Heart Association has said that cardiovascular disease causes more than 17.3

million deaths a year; By 2030, the number of deaths will exceed 23.6 million[2]. CAD is the most common type of cardiovascular disease. Its pathogenesis is due to coronary artery atherosclerotic lesions resulting in vascular lumen narrowing, or obstruction reduced blood supply to the myocardium, hypoxia or necrosis, and eventually leading to heart failure[3, 4]. However, the pathogenesis of CAD is complex, and there are no apparent symptoms in the early stage. The results of the myocardial enzyme spectrum can be negative, which is only manifested by abnormal ST-T changes in exercise plate electrocardiogram. Although coronary angiography is the gold standard for diagnosing CAD, its high cost and technical requirements, reliance on specific equipment, and potential risk of radionuclide radiation have limited its use. Meanwhile, it is neither practical nor ethical to perform invasive coronary angiography on low-risk patients[5]. The cost of blood biomarker detection is low and easy to promote[6]. Therefore, it is vital to search for more potential biomarkers for the diagnosis and treatment of CAD based on blood sequencing data.

Epithelial-mesenchymal Transition (EMT) is a biological process in which epithelial cells are transformed into cells with a mesenchymal phenotype through a specific procedure. During this process, endothelial cells gradually lose their morphology and function and acquire the phenotypic characteristics of mesenchymal cells such as proliferation, migration, and collagen synthesis[7]. Recent studies have demonstrated that EMT plays a pivotal physiological and pathological role in the development and structural remodeling of the myocardium, blood vessels, and valves, suggesting that EMT may be a worthwhile target for preventing and treating cardiovascular diseases. For example, endocardial EMT generates valvular cells necessary for heart valve formation and complete septal formation[8]. Epicardial EMT also generates cardiac fibroblasts, vascular smooth muscle cells, and surrounding cardiomyocytes necessary for cardiac muscle growth and coronary angiogenesis[9]. In CAD, endothelial cells participate in the formation of fibroblasts through the mesenchymal transformation of epithelial cells and promote cardiac fibrosis[10]. In addition, the EMT of endothelial cells plays a crucial role in the process of atherogenesis[11]. These studies suggest that EMT genes have important significance in the field of cardiovascular disease.

Long-coding RNAs (lncRNAs) are non-coding RNAs with a length of more than 200 nucleotides. It has a wide range of functional activities, including RNA decay, genetic regulation of gene expression, RNA splicing, miRNA regulation, and protein folding[12]. lncRNAs also play an essential role in forming atherosclerosis, CAD, and heart failure[13]. For example, several genome-wide association studies have found that some single nucleotide polymorphisms (SNPs) located in lncRNA-ANRIL are closely related to the susceptibility to atherosclerosis and are also the sites with the most potent genetic susceptibility in CAD[14, 15]. However, among all the identified lncRNAs, only a few have been verified to be involved in the regulation of EMT. For example, metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) is a lncRNA competing with microRNAs (miRNAs), directly interacts with oncogenes and proteins, and is involved in the activation of Wnt/ β -catenin, PI3K/Akt/mTOR, these are typical EMT-related signal pathways[16, 17]. A novel study indicated that MALAT1/miRNA-203/Wnt5a axis was a potential regulate mechanism for CAD[18]. Currently, the number of EMT-related lncRNAs in CAD research is small. How EMT-related lncRNAs regulate the formation and progression of CAD remains unclear.

Therefore, based on previous research, we constructed two machine learning algorithms: LASSO regression algorithm and SVM-RFE algorithm to screen out EMT-related diagnostic lncRNAs in CAD patients. Meanwhile, we construct cis-trans regulatory networks based on EMT-related lncRNAs and explore the potential EMT gene of related molecules in the cis-trans network and the target drugs and structures. We also investigated the correlation between EMT-related diagnostic signatures and immune cell subsets by immune analysis. Through bioinformatics methods, in-depth excavation of the EMT genes in promoting role in coronary atherosclerosis and the potential signal pathways and molecular mechanisms, late for the prevention and treatment of CAD can provide a new train of thought and targets.

Methods And Materials

Data collection

In our study, the microarray gene expression profiling data of CAD was downloaded from the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database with accession number GSE113079[19]. The platform for GSE113079 was GPL20115, Agilent-067406 Human CBC lncRNA + mRNA microarray V4.0 (Probe name version), which contained peripheral blood mononuclear cells (PBMCs) of 93 CAD patients and 48 healthy controls. 200 EMT-related genes were obtained from the Molecular Signatures Database (MsigDB, <http://www.broad.mit.edu/gsea/msigdb/>). Besides, 1639 genes related to TFs were acquired from the database of The Human Transcription Factors (TFBS, <http://tfbsdb.systemsbio.net/>).

Differentially expressed analysis

The limma package in R was used to identify the DElncRNAs and DEGs between CAD and normal samples (**Supplementary Table 1-2**). The lncRNAs/mRNAs met the selection standards of $|\log_2(\text{Fold change})| > 1.5$ and false discovery rate (FDR) < 0.01 were considered as DElncRNAs/DEGs for further study.

Correlation analysis

By mating the listed 200 EMT-related genes in the MsigDB database, differentially expressed EMT genes (DE-EMTs) for CAD were identified. Then, Pearson correlation analysis was operated between the harvested DE-EMTs and DElncRNAs expression data in samples to identify the EMT-related lncRNAs according to the correlation coefficient and P values ($|\text{Cor}| > 0.8$ and $P < 0.05$) (**Supplementary Table 3**).

Diagnostic value of EMT-lncRNAs and SNAI2

To explore the diagnostic ability of EMT-lncRNAs mentioned above, ROC analysis was first performed using the R package pROC, and the EMT-related lncRNAs with AUC > 0.95 were screened for further study. After filtration of EMT-related lncRNAs, candidate diagnostic lncRNAs for CAD were selected via integrated analysis of two algorithms consisting of LASSO and SVM-RFE. Logistic regression was

performed on diagnostic lncRNAs and the SNAI2 gene, respectively, to construct a logistic regression diagnostic model, and the bias residual diagram was drawn (**Supplementary Fig. 1A-B**). 5-fold cross validation was used to evaluate the performance of diagnostic signature. Moreover, the diagnostic value of EMT-related lncRNAs and SNAI2 were assessed by ROC curve analysis using the pROC package in the R language.

EMT-related lncRNAs categorization

Based on modifications of the previous classification[20], we classified lncRNAs according to their gene positions related to the most proximal protein-coding genes. Firstly, the lncRNA genes were regarded as intergenic and genic based on whether they intersect a protein-coding gene. Furthermore, intergenic lncRNAs were categorized as two groups depended on their transcribed from the same or opposite strand: convergent (IC) and divergent (ID). Genic lncRNAs were separated into genic exonic (genic exonic same strand (GES) and genic exonic antisense (GEAS)), genic intronic (genic intronic same strand (GIS) and genic intronic antisense (GIAS)), or overlapping (genic overlapping same strand (GOS) and genic overlapping antisense (GOAS)) based on them overlapped with the exons or introns of a protein-coding gene.

The regulatory mechanisms of diagnostic EMT-lncRNAs

It is reported that lncRNAs regulated transcription of their nearby genes via acting in cis- and trans-manners. For the cis-regulation manner, we first selected the genes located on the same chromosome within a 300 kb region upstream or downstream of the lncRNAs. Subsequently, the Pearson analysis method was performed to analyze the correlation between the harvested lncRNAs and their corresponding genes under the selection criteria of $|Cor| > 0.3$ and $P < 0.05$.

For trans prediction, we focused on that lncRNAs might regulate the expression levels of TFs via the trans manner. After selecting the genes correlated with lncRNAs by the Pearson method ($|Cor| > 0.8$ and $P < 0.05$), we further overlapped these genes with identified DEGs and TFs to obtain trans-regulated genes. A lncRNA-mRNA network that included EMT-lncRNAs, cis- and trans-regulated genes was constructed and visualized by the Cytoscape software.

Functional enrichment analysis

GO annotation and KEGG pathway analyses were employed on the DE-EMTs, cis-regulated genes, trans-regulated genes, and SNAI2 regulated genes of CAD, respectively, to explore the latent biological functions and pathways. Besides, the optional pathways related to CAD were predicted by the Comparative Toxicogenomics Database (CTD, <http://ctdbase.org>). Genes related to CAD were predicted by the DisGeNET database(<https://www.disgenet.org/home/>). The KEGG pathways both in the CAD database and KEGG analysis were introduced into the lncRNA-mRNA network to establish a lncRNA-mRNA-pathway network for CAD.

Immunity analysis and its correlation with key genes

We used CIBERSORT[21] for immune infiltration. R script downloaded from CIBERSORT website (<https://cibersort.stanford.edu/>). After obtaining the immune cell expression matrix according to the instructions on the CIBERSORT website, we used the "ggplot2" software package to create a cumulative histogram that visually showed the proportion of 22 immune cell infiltrates in CAD patients. We also used the "vioplot" package to draw violin plots showing differences in expression of 22 infiltrating immune cells. "corrplot" software package in R software was used to calculate Pearson correlation coefficients among immune cells, and the results were displayed by correlation heat map. Pearson correlation coefficients and p-value between identified key genes and infiltrating immune cells were calculated by "cor" and "Hmisc" software packages and then visualized by the "ggcorrplot" software package. In addition, single-sample gene set Enrichment Analysis (ssGSEA)[22] and QuantIseq[23] algorithms were also used compared to assess cellular components between the high SNAI2 gene expression group and low SNAI2 gene expression group. The differences in the immune response under different algorithms were uncovered using a Heatmap.

The drug-gene prediction

The genes cis- and trans-regulated by diagnostic lncRNAs were supposed to be the promising drug targets for searching drugs through the Drug-Gene Interaction database (DGIdb, <https://dgidb.genome.wustl.edu/>) that contained the drug-gene interaction information of several databases[24]. The drug-gene network was visualized by the Cytoscape tool.

Statistical analysis

The subcellular localization of diagnostic EMT-related lncRNAs was predicted by the LncLocator online tool[25]. The clusterProfiler package in R was utilized to perform GO and KEGG analyses. P-value < 0.05 was considered as statistically significant.

Results

Identification of EMT-related genes

We performed the differentially expressed analysis on the GSE113079 dataset. As shown in **Fig. 1A**, 5955 DElncRNAs were identified between CAD and normal samples under $|\log_2(\text{Fold change})|$ and FDR < 0.01 with 3067 were upregulated and 2888 downregulated. Meanwhile, we screened 2868 DEGs between two groups, including 1540 upregulated and 1328 downregulated DEMs (**Fig. 1C**). The expressed levels of DElncRNAs and DEGs were shown in the heatmap plot and displayed in **Fig. 1B, D**, respectively. The 32 DEGs related to EMT were generated by overlapping 200 EMT genes in the MsigDB database and preselected DE-EMTs, in which 21 were upregulated, and 11 were downregulated (**Fig. 1E-F**).

To further reveal the potential mechanism of the role of the above 32 DE-EMT-related genes in CAD, we performed functional enrichment analysis by Metascape software. A total of 262 BP terms, 26 MF terms, and 26 CC terms were enriched in the GO system (**Fig.1G**). We mainly focused on the enrichment results

of GO-BP categories. The 'response to wounding' was the most significantly enriched term, speculating that these genes may be involved in the regulation of the response after coronary artery damage. Next, 'extracellular matrix organization' and 'extracellular structure organization' were significantly enriched, suggesting that these genes may be associated with the acquisition of a mesenchymal phenotype by epithelial cells after EMT. The enrichment of terms related to cell adhesion ('regulation of cell adhesion', 'cell-matrix adhesion', 'positive regulation of cell adhesion', 'cell-substrate adhesion', 'regulation of cell-substrate adhesion', etc.) and cytoskeletal ('actin cytoskeleton organization', 'regulation of actin cytoskeleton organization', 'regulation of cytoskeleton organization', 'positive regulation of cytoskeleton organization', etc.) regulation and extracellular matrix disassembly could also be confirming the possibility of this speculation. Surprisingly, we found that terms related to the dynamic developmental processes of the vasculature ('blood vessel morphogenesis', 'blood vessel development', 'vasculature development', 'angiogenesis', etc.) and cardio ('heart development', 'semi-lunar valve development', 'heart morphogenesis', 'heart valve morphogenesis', 'heart valve development', etc.) were also significantly enriched, implying that, DE-EMT-related genes may be involved in the process of CAD onset and development. Besides, terms related to the regulation of biological processes such as activation and differentiation of B cells, lymphocytes, myeloid leukocytes and terms related to immune response ('negative regulation of immune system process', 'cell activation involved in immune response', 'humoral immune response', etc.) were also closely related to these genes. The enrichment results on the GO system were available in **Supplementary Table. 4**. KEGG pathway enrichment showed that these DE-EMT-related genes were involved in a total of 24 pathways (**Fig. 1H; Supplementary Table.5**). Among them, A and B were the two most significantly enriched pathways. Moreover, these genes were also linked to multiple CAD-related cardiac disease pathways, such as A, B, and C. This evidence further suggested that the 32 CAD-related DE-EMT genes may play a role in the CAD process through a certain EMT mechanism.

Identification of EMT-related lncRNAs

Next, we screened the lncRNAs related to EMT in CAD with correlation analysis. According to the screening criteria, 1141 EMT-related lncRNAs were identified. The Top50 EMT-lncRNAs were illustrated in **Fig. 2A**. Genic lncRNAs stand for the largest category (57.2%) of EMT-related lncRNAs (GEAS = 7.8% and GES = 7%, GIAS = 19% and GIS = 18.4%, GOAS = 2.8% and GOS = 2.3%), following intergenic lncRNAs were 42.7% (IC = 19.9% and ID = 22.8%) (**Fig. 2B**).

Construction of an EMT-related lncRNAs diagnostic signature for CAD

To further detect the diagnostic ability of these EMT-related lncRNAs, the AUC value of each EMT-related lncRNAs was analyzed. 223 EMT-related lncRNAs were screened with an AUC value above 0.95 (**Supplementary Table. 6**). LASSO regression analysis SVM-RFE algorithm was used to identify the optimal diagnostic lncRNAs in the GSE113079 dataset and establish the risk signature for CAD. 16 EMT-related lncRNAs were screened via the LASSO analysis, which intersected with 34 EMT-related lncRNAs obtained from the SVM-RFE algorithm to identify 11 diagnostic lncRNAs for CAD (**Supplementary Table. 7, Fig. 3A-E**). After annotating the diagnostic lncRNAs using the Rsubread package in R, we obtained eight

lncRNAs that were used to construct a diagnostic signature for CAD (**Table 1**), showing accuracy and specificity for the diagnosis of CAD (AUC = 1) (**Fig. 3F**). Besides, the AUC value of each diagnostic lncRNAs was greater than 0.95, which exhibited a better ability to distinguish CAD patients from normal (**Fig. 3G**). Subcellular localization of each lncRNA determines the regulatory models. To investigate the subcellular localization of the diagnostic lncRNAs, we assessed LncLocator online platforms. We uncovered that these diagnostic lncRNAs were mainly located in the cytosol and cytoplasm (**Supplementary Fig. 2**)

Establishment of cis and trans-regulatory network

Previous studies indicated that lncRNAs regulated gene expression via local (cis) and long-distance (trans) mechanisms[26]. In the present study, we identified seven diagnostic lncRNAs regulated their nearby genes via the cis-regulatory manner, except RP11-103H7.3 (**Table 2**). Among them, only CTD-2089N3.3 were significantly correlated with their corresponding gene EMB via Pearson analysis under $|Cor| > 0.3$ and $P\text{-value} < 0.05$ (**Fig. 4A**). Based on the median expression level of EMB, we divided the CAD patients in the GSE113079 dataset into the high-expressed EMB group and low-expressed EMB group. GSEA result suggested that several metabolic- and tumor-related pathways were associated with high-expressed EMB groups, including 'fatty acid metabolism,' 'pyrimidine metabolism,' 'mTOR signaling pathway,' and 'TGF BETA signaling pathway.' In contrast, 'calcium signaling pathway,' 'complement, and coagulation cascades,' 'neuroactive ligand-receptor interaction,' and 'olfactory transduction' were involved in the low-expressed EMB group (**Supplementary Fig. 3**).

By combining 685 genes correlated with lncRNAs with DEGs and identified TFs, 33 genes were identified to be regulated by diagnostic lncRNAs via trans manner, in which SNAI2 was founded to be a differentially expressed EMT gene (**Fig. 4B-C**). Then, a lncRNA-mRNA regulatory network was constructed that contained diagnostic lncRNAs, cis- and trans-regulated genes, which consisted of 42 nodes and 93 edges (**Fig. 4D**). These genes in the regulatory network were mainly involved in nervous development and vitamin metabolism by GO analysis (**Supplementary Fig. 4A**). Combining the pathways related to these genes identified with the KEGG analysis and related to CAD development in the CTD database, 'maturity onset diabetes of the young' and 'transcriptional misregulation in cancer' were discovered (**Fig. 4E-F, Table 3**). Hence, these two KEGG pathways were introduced into the lncRNA-mRNA regulatory network to establish a lncRNA-mRNA-pathway network for CAD that included 46 nodes and 98 edges (**Supplementary Fig. 4B**).

Prediction of regulatory genes of SNAI2

Based on the above results, we uncovered that SNAI2 was found to be a DE-EMT and TFs among all genes regulated by diagnostic lncRNAs. In our study, SNAI2 is obviously higher expressed in CAD groups than normal groups ($p = 5.8e-15$; **Fig. 5A**). Considering the importance of SNAI2, we also detect the diagnostic ability of SNAI2 in CAD patients. The AUC value of SNAI2 was 0.902 (**Fig. 5B**). In addition, we used five-fold cross validation to evaluate the reliability of the SNAI2 gene. Firstly, we randomly divided the samples into five parts, of which four parts were used as training sets to build the logistic regression

model, and the rest were used to verify the model. This process is then repeated five times to reduce errors and improve the sensitivity of the model. The AUC values of the five models were 0.9479, 0.9144, 0.9391, 0.7692, and 0.8766, respectively, indicating that the models had good explanatory power (**Fig. 5C**). Besides, GSEA was performed to investigate the latent biological functions. 'Calcium signaling pathway,' 'linoleic acid metabolism,' 'neuroactive ligand receptor interaction,' and 'olfactory transduction' were mainly associated with the high-expressed SNAI2 group. 'RNA degradation,' 'spliceosome,' 'fatty acid metabolism,' and 'histone metabolism' were involved in the low-expressed SNAI2 group (**Supplementary Fig. 5**). Moreover, we overlapped 234 genes regulated by SNAI2 acquired from the TFBS database and 1576 genes related to CAD acquired from the DisGeNET database to obtain 21 genes regulated by SNAI2 for CAD (**Fig. 5D**). Functional enrichment analysis determined that the harvested 21 genes were concerted on the 'insulin secretion,' 'peptide hormone secretion,' 'long-chain fatty acid biosynthetic process' (**Fig. 5E**). There are no pathways detected by KEGG analysis.

Prediction of the target drugs of genes in the cis and trans-regulatory network

Next, the target drugs of genes regulated by diagnostic lncRNAs were predicted by the DGIdb database. Through DGIdb prediction, a total of 483 drug-gene pairs were identified, and a target-drug network for CAD was constructed, including five genes and 476 drugs. 459 drugs interacted with VDR, which might be promising to treat patients with CAD (**Supplementary Fig. 6A**). The structures of these drugs were illustrated in **Supplementary Fig. 6B-I**.

Immune analysis of EMT-lncRNAs and SNAI2

Enrichment analysis showed that DE-EMT gene was enriched in inflammatory response-related pathways. Therefore, we evaluated the type and fraction of immune cell infiltration between CAD patients and normal samples in the dataset using the CIBERSORT algorithm. The relative proportion of immune cell subtypes was shown in the cumulative histogram (**Fig. 6A**). Our results found an apparent proportion of CD8 T cells, NK cells activated, and monocytes. Moreover, the infiltration of CD8 T cells and NK cell activated were decreased, and the infiltration of monocytes was increased in CAD patients (**Fig. 6B**). By principal component analysis (PCA), immune cell fractions in CAD patients and normal controls showed intergroup bias and individual differences (**Fig. 6C**). In the correlation heatmap (**Fig. 6D**), we found that CD8 T cells were negatively correlated with monocytes and macrophages M0, and positively correlated with NK cells activated. It is consistent with the correlation between seven EMT-related lncRNAs and the immune cells we found, except lncRNA AC109460.4 (**Fig. 7A**). In addition, we also conducted an immune analysis of SNAI2. Then, we divided the samples into high and low groups according to the expression level of SNAI2. It was found that in the high expression level group, the infiltration of monocytes was decreased. In contrast, the infiltration of NK cell activated and CD8 T cells were increased, which was similar to immune cell infiltration in CAD patients (**Fig. 7B**). The heatmap of immune cell compositions based on CIBERSORT, quanTIseq, ssGSEA algorithms is shown in **Fig 8**. It was found that CD8 T cells, monocytes, and NK cells activated had similar immune cell infiltration trends in the CAD and SNAI2 gene high expression groups.

Discussion

EMT plays a critical physiological and pathological role in developing the cardiovascular system, vascular tissue remodeling, and heart valve disease during the embryonic period[27]. However, more research focused on the impact of the EMT in tumor development and treatment. In contrast, few studies have explored the diagnostic value of EMT-related genes or lncRNAs in CAD. Hence, exploring diagnostic biomarkers of EMT genes/lncRNAs in CAD is urgent.

Our analyses uncovered 32 EMT-related DEGs in CAD. KEGG pathway analysis of these DE-EMTs was mainly enriched in the PI3K/Akt signaling pathway. Several reports have shown that PI3K/Akt pathway is a canonical EMT signaling pathway[28, 29]. Meanwhile, we found this signaling pathway plays an essential role in the CAD. A recent study indicated that miRNA-26a-5p activated the PI3K/Akt pathway by targeting Phosphatase and Tensin Homolog (PTEN) and affected the proliferation and apoptosis of endothelial cells isolated from CAD mice[30]. A comparative study also reported that miR-26a-5p could activate the PI3K/Akt signaling pathway through inhibition of PTEN, thereby protecting against myocardial defect/reperfusion injury[31]. These studies have confirmed that activating the PI3K/Akt signaling pathway can prevent myocardial ischemia-reperfusion in animal models. Other studies have also suggested that regulation of the PI3K/Akt signaling pathway plays a vital role in inhibiting myocardial fibrosis, apoptosis, and the inflammatory response[32, 33].

In the present study, we performed a co-expression analysis between EMT genes and DElncRNAs through paired lncRNA and mRNA expression data in CAD patients from GEO. Eight differently expressed EMT-related lncRNAs were found to be diagnosis factors for CAD patients. After a literature review, we found no research had been conducted about the mechanisms of the eight lncRNAs except LINC02747. Previous studies have reported that LINC02747 can upregulate the expression of TFE3 by absorbing miR-608 and ultimately promote the proliferation of renal cell cancer cells (ccRCC)[34]. Gu et al. indicated that miR-608 exerts anti-inflammatory effects by targeting ELANE in monocytes[35]. Our results showed that monocytes were more expressed in the CAD group, so whether the regulation of LINC02747-mir608-ELANE might achieve the reversal of inflammatory response in CAD patients. Other seven EMT-related lncRNAs have not been reported in relevant studies, and reports on how lncRNAs interact with EMT genes have been rarer. However, many 'cis' and 'trans' genes are involved in the formation and development of CAD in the cis-trans regulatory network. For example, EMB, as a 'cis' gene, was enriched in the mTOR signaling pathway in our GSEA analysis. This pathway is closely associated with atherosclerosis, and the pro-inflammatory response of monocytes in CAD requires activation of mTOR[36]. Among 'trans' genes, some studies have reported that VDR gene polymorphisms lead to the development and formation of CAD by affecting changes in serum levels of 25(OH) vitamin D.[37, 38]. Previous study reported VDR in regulating inflammation through inhibiting the NF- κ B pathway and activating autophagy[39]. EBF4 gene promotes the elevation of Cu and leads to the progression of CAD by affecting copper related DNA methylation sites[40]. CTCF gene is essential for cardiogenesis and inhibit cardiomyocytes apoptosis, and can be applied as a therapeutic target for the treatment of heart failure in future.[41, 42]. FLI1 gene is also reported to be closely related to immune dysfunction and platelet disorders[43]. Although the lack of

direct support in the literature, we speculated that these cis-trans genes, under the regulation of lncRNA, affect the formation and development of CAD through immune microenvironment, cell apoptosis, platelet dysfunction and other ways. So far, there has been no study on the role of EMT-related lncRNA in CAD diagnosis. These findings may provide valuable insights into the future diagnosis and treatment of CAD.

The presence of immune cells in the infarct area is vital to initiating the repair process of injured heart tissue. Temporal and spatial regulation of inflammation after infarction is crucial[44, 45]. We evaluate the type and fraction of immune cell infiltration between CAD patients and normal samples in the dataset using the CIBERSORT algorithm. Our results found CD8 T cells and NK cells share a decreased infiltration, and the infiltration of monocytes was increased in CAD patients, which was similar to the previous results[46-48]. In this GEO dataset, CD8 T cells and NK cells are favorable factors for preventing CAD, and monocytes likely promote the occurrence of CAD. Previous studies have suggested that the imbalance of immune regulation is an essential factor in promoting atherosclerosis, heart failure, and chronic kidney disease by monocytes cells[45]. CD8 T cells play a dual role in atherosclerosis. On the one hand, CD8 T cells can secrete many inflammatory cytokines to accelerate the inflammatory response and increase the instability of atherosclerotic plaques. On the other hand, cytotoxic activity against antigen-presenting cells and the presence of regulatory CD8 T cell subsets could suppress immunity and limit atherosclerosis[49]. Ong et al. suggested that NK cells appear to protect the development of cardiac fibrosis by preventing the accumulation of specific inflammatory groups in the heart and directly restricting collagen formation in cardiac fibroblasts[50]. Although the results of our study are similar to these researches, the mechanism of the immune system is still very complex, and some results in the immunotherapy of CAD are not ideal. We need a lot of clinical studies to demonstrate the underlying mechanism. Besides, we also found that except AC109460.4, the other seven lncRNAs related to EMT were significantly negatively correlated with CD8 T cell and NK cell and positively correlated with Treg and monocytes. The results of AC109460.4 were just the opposite. The association between these lncRNAs and the innate immune system is still unclear. More in vivo and in vitro studies are needed to explain the interaction mechanism between these lncRNAs and immune cells in CAD.

It is generally believed that lncRNAs can act in 'trans' to regulate TFs mediated chromatin remodeling and transcription[51]. These lncRNAs recruit protein factors to enhancer and regulate enhancer activity[52]. We constructed cis and trans-regulatory networks based on these eight signatures. In the trans-regulatory network, we obtained 33 differentially expressed TF genes. The most surprising discovery was the screening of SNAI2, an EMT-TF gene (the gene coding product was the transcription factor Snai2). Our results indicated that SNAI2 was not only significantly highly expressed in CAD patients but also strongly positively correlated with LINC01775 and CTD-2089N3.3. The ROC curve showed that the SNAI2 could be a potential biomarker for diagnosing CAD. As a classic EMT-TF gene, SNAI2 has recently been shown to be involved in a broader range of biological processes, including tumor metastasis, heart development, cell differentiation, vascular remodeling, and DNA damage repair[53-55]. Previous studies have reported that the deletion of protein arginine methyltransferase 1 (PRMT1) leads to the accumulation of p53, and enhancing the degradation of SNAI2 can limit the formation of cardiac fibroblasts, coronary smooth muscle cells, and pericytes[56]. Meanwhile, Cooley et al. reported that, by grafting mouse veins to the

femoral artery in mice to simulate human coronary artery bypass grafting (CABG), the results showed that TGF- β /Smad2/3-Snai2 mediated EMT plays a crucial role in venous graft vessel remodeling[57]. These studies have indicated that high expression of SNAI2 can promote the formation of vascular endothelium to EMT and vascular remodeling, which is one of the vital factors in the formation of CAD. At present, the role of SNAI2 in CAD has not been reported, several studies have proven that the vascular endothelial EMT process is involved in atherosclerosis, post-stent stenosis, pulmonary hypertension, and coronary artery remodeling[58-60]. Additionally, the role of EMT can be seen in a range of cell types involved in immunity, such as lymphocytes, NK cells, and myeloid cells, which contribute to inflammatory responses in diverse pathophysiological processes. Ricciardi et al. have reported a decreased viability and proliferation of NK cells and T cells after co-culture with cancer cell lines in which EMT had been induced[61]. In our study, SNAI2 correlated with infiltration of monocytes, CD8 T cells, and NK cells activated. Previous studies have suggested that SNAI2 deletion in mice leads to impaired development of the T-lymphatic system[62]. Subsequent studies also confirmed that Snai2 plays a vital influence in regulating CD8 T cells and targets genes with functions for T cells[63]. Furthermore, our results indicated that the difference in these immune cell infiltrations in the SNAI2 high expression group was similar to the results of CAD patients. These immune cells have been researched to play a role in the formation, erosion, and rupture of coronary plaques[64, 65]. In summary, we inferred that SNAI2 might have significant roles in the occurrence of CAD by regulating innate and adaptive immunity through these immune cells. To confirm our conclusion, more experimental mechanistic research should be carried out in future studies.

Our study should acknowledge some limitations. First, the expression levels of critical lncRNAs in CAD were not verified in clinical samples. Secondly, these EMT-related lncRNAs were investigated in datasets with no access to individual patients' characteristics; thus, we cannot adjust the ROC curve for traditional cardiovascular risk factors. A prospective cohort recruiting CAD patients is needed to confirm the predictive value of EMT-related lncRNAs. Thirdly, the molecular function details of SNAI2 and EMT-related lncRNAs in the progression of CAD have not been further studied. Therefore, molecular biological experiments and flow cytometry analysis are required to validate these findings, and another external validation based on a larger sample is needed.

Conclusions

In conclusion, this comprehensive bioinformatic analysis revealed that SNAI2 and EMT-related lncRNAs could be reliable biomarkers for diagnosing CAD and use decision-making in the treatment of CAD patients. At the same time, based on the eight EMT-related lncRNAs, we constructed the cis and trans-regulatory networks of CAD. Furthermore, the immune analysis suggested that these biomarkers were closely related to immune cells and CAD. These findings provide references for clinicians to understand the molecular mechanism of interaction between CAD and EMT and develop individualized treatment for CAD patients.

Abbreviations

CAD: Coronary artery disease; EMT: Epithelial-mesenchymal Transition; lncRNAs: Long-coding RNAs; MALAT1: Metastasis associated lung adenocarcinoma transcript 1; miRNAs: microRNAs; SNPs: Single nucleotide polymorphisms; SVM-RFE: Support Vector Machine Reverse Feature Elimination; ssGSEA: Single-sample gene set enrichment analysis; GSEA: Gene set enrichment analyses; GO: Gene ontology; BP: Biological processes; MF: Molecular function; CC: Cellular components; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: Differentially expressed genes; DE-EMTs: Differentially expressed EMT genes; DElncRNAs: Differentially expressed lncRNAs; FC: Fold-change; CIBERSORT: Cell Type Identification by Estimating Relative Subsets of RNA Transcripts Algorithm; PCA: Principal component analysis; ccRCC: Clear cell renal cell carcinoma; GEO: Gene Expression Omnibus; LASSO: Least absolute shrinkage and selection operator; FDR: False discovery rate; AUC: The area under the curve; ROC: Receiver operating characteristic; PRMT1: protein arginine methyltransferase 1; CABG: coronary artery bypass grafting.

Declarations

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Xiang Xu designed the experiments, performed the experiments, analyzed the data, prepared figures, reviewed drafts of the paper, approved the final draft.

Qianqian Su designed the experiments, performed the experiments, reviewed drafts of the paper, approved the final draft.

Renchao Zou analyzed the data, contributed analysis tools, prepared figures, approved the final draft.

Xiaoyong Liu analyzed the data, prepared figures, approved the final draft.

Acknowledgments

Not applicable.

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The data used to support the findings of this study are available from the corresponding author upon request. We thank the contributors of the GEO databases for the availability of the data. The gene expression profiles in the GSE113079 dataset were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113079>).

Consent for publication

Not applicable.

Funding

Not applicable.

References

1. Adapala RK, Kanugula AK, Paruchuri S, Chilian WM, Thodeti CK: **TRPV4 deletion protects heart from myocardial infarction-induced adverse remodeling via modulation of cardiac fibroblast differentiation.** *Basic Res Cardiol* 2020, **115**(2):14.
2. Writing Group M, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, de Ferranti S, Despres JP *et al.* **Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association.** *Circulation* 2016, **133**(4):e38-360.
3. Li GM, Zhang CL, Rui RP, Sun B, Guo W: **Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy.** *Eur Rev Med Pharmacol Sci* 2018, **22**(11):3553-3569.
4. Chen CH, Wang SS, Wei EI, Chu TY, Hsieh PC: **Hyaluronan enhances bone marrow cell therapy for myocardial repair after infarction.** *Mol Ther* 2013, **21**(3):670-679.
5. De Marco E, Vacchiano G, Frati P, La Russa R, Santurro A, Scopetti M, Guglielmi G, Fineschi V: **Evolution of post-mortem coronary imaging: from selective coronary arteriography to post-mortem CT-angiography and beyond.** *Radiol Med* 2018, **123**(5):351-358.
6. Rusnak J, Fastner C, Behnes M, Mashayekhi K, Borggrefe M, Akin I: **Biomarkers in Stable Coronary Artery Disease.** *Curr Pharm Biotechnol* 2017, **18**(6):456-471.
7. Georgakopoulos-Soares I, Chartoumpakis DV, Kyriazopoulou V, Zaravinos A: **EMT Factors and Metabolic Pathways in Cancer.** *Front Oncol* 2020, **10**:499.
8. von Gise A, Pu WT: **Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease.** *Circ Res* 2012, **110**(12):1628-1645.

9. Zhou B, von Gise A, Ma Q, Hu YW, Pu WT: **Genetic fate mapping demonstrates contribution of epicardium-derived cells to the annulus fibrosus of the mammalian heart.** *Dev Biol* 2010, **338**(2):251-261.
10. Wilhelmi T, Xu X, Tan X, Hulshoff MS, Maamari S, Sossalla S, Zeisberg M, Zeisberg EM: **Serelaxin alleviates cardiac fibrosis through inhibiting endothelial-to-mesenchymal transition via RXFP1.** *Theranostics* 2020, **10**(9):3905-3924.
11. Souilhol C, Harmsen MC, Evans PC, Krenning G: **Endothelial-mesenchymal transition in atherosclerosis.** *Cardiovasc Res* 2018, **114**(4):565-577.
12. Khan S, Masood M, Gaur H, Ahmad S, Syed MA: **Long non-coding RNA: An immune cells perspective.** *Life Sci* 2021, **271**:119152.
13. Uchida S, Dimmeler S: **Long noncoding RNAs in cardiovascular diseases.** *Circ Res* 2015, **116**(4):737-750.
14. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B: **Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression.** *PLoS Genet* 2010, **6**(4):e1000899.
15. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi MG, Tognoni G *et al.*: **Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p.** *Hum Mol Genet* 2008, **17**(6):806-814.
16. Uthman YA, Ibrahim KG, Abubakar B, Bello MB, Malami I, Imam MU, Qusty N, Cruz-Martins N, Batiha GE, Abubakar MB: **MALAT1: A promising therapeutic target in metastatic colorectal cancer.** *Biochem Pharmacol* 2021, **190**:114657.
17. Xu W, Ding M, Wang B, Cai Y, Guo C, Yuan C: **Molecular mechanism of the canonical oncogenic lncRNA MALAT1 in gastric cancer.** *Curr Med Chem* 2021.
18. Cai Q, Gao ML, Huang LS, Chen HS, Pan LH: **MALAT1/miRNA-203/Wnt5a: A potential mechanism for regulating coronary artery disease.** *Int J Cardiol* 2021, **329**:48.
19. Li L, Wang L, Li H, Han X, Chen S, Yang B, Hu Z, Zhu H, Cai C, Chen J *et al.*: **Characterization of LncRNA expression profile and identification of novel LncRNA biomarkers to diagnose coronary artery disease.** *Atherosclerosis* 2018, **275**:359-367.
20. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG *et al.*: **The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.** *Genome Res* 2012, **22**(9):1775-1789.
21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nat Methods* 2015, **12**(5):453-457.

22. Yi M, Nissley DV, McCormick F, Stephens RM: **ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications.** *Sci Rep* 2020, **10**(1):10258.
23. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D *et al.*: **Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data.** *Genome Med* 2019, **11**(1):34.
24. Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL, Wagner AH: **Integration of the Drug-Gene Interaction Database (DGldb 4.0) with open crowdsourcing efforts.** *Nucleic Acids Res* 2021, **49**(D1):D1144-D1151.
25. Cao Z, Pan X, Yang Y, Huang Y, Shen HB: **The IncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier.** *Bioinformatics* 2018, **34**(13):2185-2194.
26. Chen LL: **Linking Long Noncoding RNA Localization and Function.** *Trends Biochem Sci* 2016, **41**(9):761-772.
27. Liu X, Wang Y, Liu F, Zhang M, Song H, Zhou B, Lo CW, Tong S, Hu Z, Zhang Z: **Wdpcp promotes epicardial EMT and epicardium-derived cell migration to facilitate coronary artery remodeling.** *Sci Signal* 2018, **11**(519).
28. Dong WQ, Chao M, Lu QH, Chai WL, Zhang W, Chen XY, Liang ES, Wang LB, Tian HL, Chen YG *et al.*: **Prohibitin overexpression improves myocardial function in diabetic cardiomyopathy.** *Oncotarget* 2016, **7**(1):66-80.
29. Xu Z, Jia K, Wang H, Gao F, Zhao S, Li F, Hao J: **METTL14-regulated PI3K/Akt signaling pathway via PTEN affects HDAC5-mediated epithelial-mesenchymal transition of renal tubular cells in diabetic kidney disease.** *Cell Death Dis* 2021, **12**(1):32.
30. Jing R, Zhong QQ, Long TY, Pan W, Qian ZX: **Downregulated miRNA-26a-5p induces the apoptosis of endothelial cells in coronary heart disease by inhibiting PI3K/AKT pathway.** *Eur Rev Med Pharmacol Sci* 2019, **23**(11):4940-4947.
31. Xing X, Guo S, Zhang G, Liu Y, Bi S, Wang X, Lu Q: **miR-26a-5p protects against myocardial ischemia/reperfusion injury by regulating the PTEN/PI3K/AKT signaling pathway.** *Braz J Med Biol Res* 2020, **53**(2):e9106.
32. Li X, Sun S, Chen D, Yuan T, Chen Y, Wang D, Fang L, Lu Y, Du G: **Puerarin attenuates the endothelial-mesenchymal transition induced by oxidative stress in human coronary artery endothelial cells through PI3K/AKT pathway.** *Eur J Pharmacol* 2020, **886**:173472.
33. Guan BF, Dai XF, Huang QB, Zhao D, Shi JL, Chen C, Zhu Y, Ai F: **Icariside II ameliorates myocardial ischemia and reperfusion injury by attenuating inflammation and apoptosis through the regulation of the**

PI3K/AKT signaling pathway. *Mol Med Rep* 2020, **22**(4):3151-3160.

34. Ju X, Sun Y, Zhang F, Wei X, Wang Z, He X: **Long Non-Coding RNA LINC02747 Promotes the Proliferation of Clear Cell Renal Cell Carcinoma by Inhibiting miR-608 and Activating TFE3.** *Front Oncol* 2020, **10**:573789.
35. Gu W, Wen D, Lu H, Zhang A, Wang H, Du J, Zeng L, Jiang J: **MiR-608 Exerts Anti-inflammatory Effects by Targeting ELANE in Monocytes.** *J Clin Immunol* 2020, **40**(1):147-157.
36. Gao S, Liu W, Zhuo X, Wang L, Wang G, Sun T, Zhao Z, Liu J, Tian Y, Zhou J *et al*: **The activation of mTOR is required for monocyte pro-inflammatory response in patients with coronary artery disease.** *Clin Sci (Lond)* 2015, **128**(8):517-526.
37. Tabaei S, Motallebnezhad M, Tabaei SS: **Vitamin D Receptor (VDR) Gene Polymorphisms and Risk of Coronary Artery Disease (CAD): Systematic Review and Meta-analysis.** *Biochem Genet* 2021, **59**(4):813-836.
38. Moradi N, Fadaei R, Ahmadi R, Mohammad MH, Shahmohamadnejad S, Tavakoli-Yaraki M, Aghajani H, Fallah S: **Role of serum MMP-9 levels and vitamin D receptor polymorphisms in the susceptibility to coronary artery disease: An association study in Iranian population.** *Gene* 2017, **628**:295-300.
39. Bakke D, Sun J: **Ancient Nuclear Receptor VDR With New Functions: Microbiome and Inflammation.** *Inflamm Bowel Dis* 2018, **24**(6):1149-1154.
40. Long P, Wang Q, Zhang Y, Zhu X, Yu K, Jiang H, Liu X, Zhou M, Yuan Y, Liu K *et al*: **Profile of copper-associated DNA methylation and its association with incident acute coronary syndrome.** *Clin Epigenetics* 2021, **13**(1):19.
41. Gomez-Velazquez M, Badia-Careaga C, Lechuga-Vieco AV, Nieto-Arellano R, Tena JJ, Rollan I, Alvarez A, Torroja C, Caceres EF, Roy AR *et al*: **CTCF counter-regulates cardiomyocyte development and maturation programs in the embryonic heart.** *PLoS Genet* 2017, **13**(8):e1006985.
42. Zeng Z, Huang N, Zhang Y, Wang Y, Su Y, Zhang H, An Y: **CTCF inhibits endoplasmic reticulum stress and apoptosis in cardiomyocytes by upregulating RYR2 via inhibiting S100A1.** *Life Sci* 2020, **242**:117158.
43. Antony-Debre I, Bluteau D, Itzykson R, Baccini V, Renneville A, Boehlen F, Morabito M, Droin N, Deswarte C, Chang Y *et al*: **MYH10 protein expression in platelets as a biomarker of RUNX1 and FLI1 alterations.** *Blood* 2012, **120**(13):2719-2722.
44. Jiang YX, Yang SW, Li PA, Luo X, Li ZY, Hao YX, Yu PW: **The promotion of the transformation of quiescent gastric cancer stem cells by IL-17 and the underlying mechanisms.** *Oncogene* 2017, **36**(9):1256-1264.

45. Dounousi E, Duni A, Naka KK, Vartholomatos G, Zoccali C: **The Innate Immune System and Cardiovascular Disease in ESKD: Monocytes and Natural Killer Cells.** *Curr Vasc Pharmacol* 2021, **19**(1):63-76.
46. Yang Y, Xu X: **Identification of key genes in coronary artery disease: an integrative approach based on weighted gene co-expression network analysis and their correlation with immune infiltration.** *Aging (Albany NY)* 2021, **13**.
47. Jabir NR, Firoz CK, Ahmed F, Kamal MA, Hindawi S, Damanhoury GA, Almehdar HA, Tabrez S: **Reduction in CD16/CD56 and CD16/CD3/CD56 Natural Killer Cells in Coronary Artery Disease.** *Immunol Invest* 2017, **46**(5):526-535.
48. Yan W, Zhou L, Wen S, Duan Q, Huang F, Tang Y, Liu X, Chai Y, Wang L: **Differential loss of natural killer cell activity in patients with acute myocardial infarction and stable angina pectoris.** *Int J Clin Exp Pathol* 2015, **8**(11):14667-14675.
49. van Duijn J, Kuiper J, Slutter B: **The many faces of CD8+ T cells in atherosclerosis.** *Curr Opin Lipidol* 2018, **29**(5):411-416.
50. Ong S, Rose NR, Cihakova D: **Natural killer cells in inflammatory heart disease.** *Clin Immunol* 2017, **175**:26-33.
51. Guttman M, Rinn JL: **Modular regulatory principles of large non-coding RNAs.** *Nature* 2012, **482**(7385):339-346.
52. Peng Z, Zhang C, Duan C: **Functions and mechanisms of long noncoding RNAs in lung cancer.** *Onco Targets Ther* 2016, **9**:4411-4424.
53. Hussen BM, Shoorei H, Mohaqiq M, Dinger ME, Hidayat HJ, Taheri M, Ghafouri-Fard S: **The Impact of Non-coding RNAs in the Epithelial to Mesenchymal Transition.** *Front Mol Biosci* 2021, **8**:665199.
54. Heallen T, Zhang M, Wang J, Bonilla-Claudio M, Klysik E, Johnson RL, Martin JF: **Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size.** *Science* 2011, **332**(6028):458-461.
55. Osoegawa K, Iovannisci DM, Lin B, Parodi C, Schultz K, Shaw GM, Lammer EJ: **Identification of novel candidate gene loci and increased sex chromosome aneuploidy among infants with conotruncal heart defects.** *Am J Med Genet A* 2014, **164A**(2):397-406.
56. Jackson-Weaver O, Ungvijanpunya N, Yuan Y, Qian J, Gou Y, Wu J, Shen H, Chen Y, Li M, Richard S *et al.*: **PRMT1-p53 Pathway Controls Epicardial EMT and Invasion.** *Cell Rep* 2020, **31**(10):107739.
57. Cooley BC, Nevado J, Mellad J, Yang D, St Hilaire C, Negro A, Fang F, Chen G, San H, Walts AD *et al.*: **TGF-beta signaling mediates endothelial-to-mesenchymal transition (EndMT) during vein graft**

remodeling. *Sci Transl Med* 2014, **6**(227):227ra234.

58. Cheng SL, Shao JS, Behrmann A, Krchma K, Towler DA: **Dkk1 and MSX2-Wnt7b signaling reciprocally regulate the endothelial-mesenchymal transition in aortic endothelial cells.** *Arterioscler Thromb Vasc Biol* 2013, **33**(7):1679-1689.
59. Kato H, Fu YY, Zhu J, Wang L, Aafaqi S, Rahkonen O, Slorach C, Traister A, Leung CH, Chiasson D *et al.*: **Pulmonary vein stenosis and the pathophysiology of "upstream" pulmonary veins.** *J Thorac Cardiovasc Surg* 2014, **148**(1):245-253.
60. Wu X, Du X, Yang Y, Liu X, Liu X, Zhang N, Li Y, Jiang X, Jiang Y, Yang Z: **Inhibition of miR-122 reduced atherosclerotic lesion formation by regulating NPAS3-mediated endothelial to mesenchymal transition.** *Life Sci* 2021, **265**:118816.
61. Ricciardi M, Zanotto M, Malpeli G, Bassi G, Perbellini O, Chilosi M, Bifari F, Krampera M: **Epithelial-to-mesenchymal transition (EMT) induced by inflammatory priming elicits mesenchymal stromal cell-like immune-modulatory properties in cancer cells.** *Br J Cancer* 2015, **112**(6):1067-1075.
62. Pioli PD, Dahlem TJ, Weis JJ, Weis JH: **Deletion of Snai2 and Snai3 results in impaired physical development compounded by lymphocyte deficiency.** *PLoS One* 2013, **8**(7):e69216.
63. Pioli PD, Whiteside SK, Weis JJ, Weis JH: **Snai2 and Snai3 transcriptionally regulate cellular fitness and functionality of T cell lineages through distinct gene programs.** *Immunobiology* 2016, **221**(5):618-633.
64. Manduteanu I, Simionescu M: **Inflammation in atherosclerosis: a cause or a result of vascular disorders?** *J Cell Mol Med* 2012, **16**(9):1978-1990.
65. Jabir NR, Tabrez S: **Cardiovascular disease management through restrained inflammatory responses.** *Curr Pharm Des* 2016, **22**(7):940-946.

Tables

Table 1. Annotating the diagnostic eight EMT-related lncRNAs using the Rsubread package in R.

probe_id	Symbol
p22710	CTD-2089N3.3
p40724_v4	AC113167.2
p2383	LINC02747
p4630	RP11-1152H15.1
p4940	LINC02833
p5961	AC109460.4
p8126	LINC01775
p16363	RP11-103H7.3

Table 2. Seven EMT-related diagnostic lncRNAs regulated their nearby genes via the cis-regulatory manner.

Gene	lncRNA	Symbol	cor	p.value	chromosome	distance
EMB	p22710	CTD-2089N3.3	-0.46194075	8.12E-09	chr5	130172
MEF2C	p40724_v4	AC113167.2	0.156713247	0.063484128	chr5	-294003
MYEOV	p2383	LINC02747	0.23717925	0.004628671	chr11	0
KCNK10	p4630	RP11-1152H15.1	-0.251517089	0.002625124	chr14	0
UNC79	p4940	LINC02833	-0.073789859	0.384523053	chr14	0
LAT	p5961	AC109460.4	-0.244589931	0.003466511	chr16	-9677
LMNB2	p8126	LINC01775	0.05339002	0.529495266	chr19	-1974

Table 3. Combining the pathways related to potential “cis” and “trans” genes identified with the KEGG analysis and related to CAD development in the CTD database.

# Input	DiseaseName	PathwayName	PathwayID	Inference GeneSymbol
coronary heart disease	Coronary Disease	Maturity onset diabetes of the young	KEGG:hsa0495 0	HNF1A
coronary heart disease	Coronary Disease	Transcriptional misregulation in cancer	KEGG:hsa0520 2	MMP3
coronary heart disease	Coronary Disease	Transcriptional misregulation in cancer	KEGG:hsa0520 2	PLAU

Figures

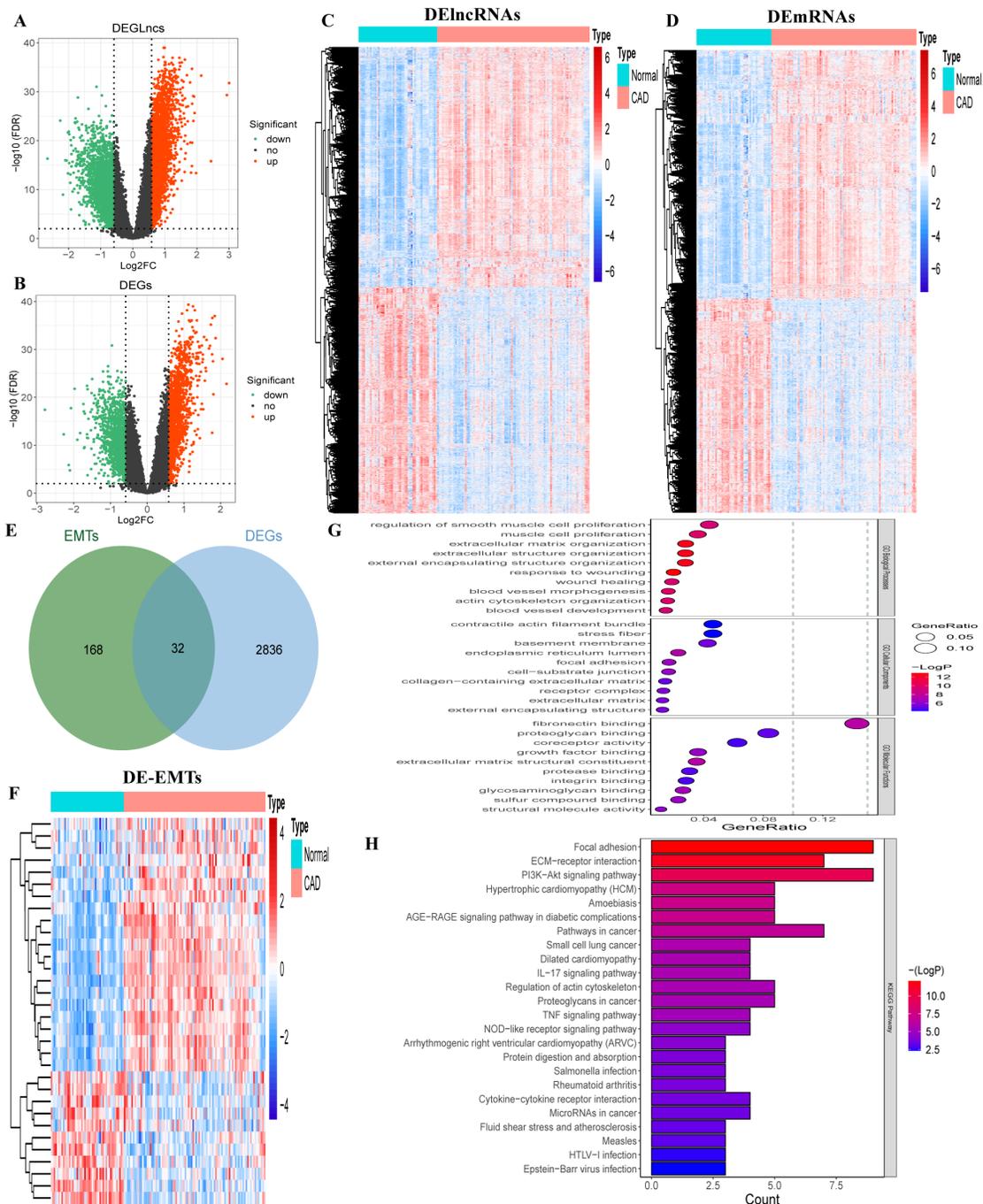


Figure 1

Identification of DE-EMTs in CAD. (A) Volcano plot of lncRNAs expression between CAD and normal groups. (B) Volcano plot of mRNAs expression between CAD and normal groups. (C) Heatmap of DElncRNAs between CAD and normal groups. (D) Heatmap of DE mRNAs between CAD and normal groups. (E) Venn diagram was used for the intersection of DEGs and EMT genes. (F) Heatmap of DE-EMTs between CAD and normal groups. (G) GO enrichment analysis of DE-EMTs: the larger the bubble and longer

columns represent the more genes enriched in this function, the deeper the color of the bubble and bars, the smaller the P-value. (H) KEGG enrichment analysis of DE-EMTs.

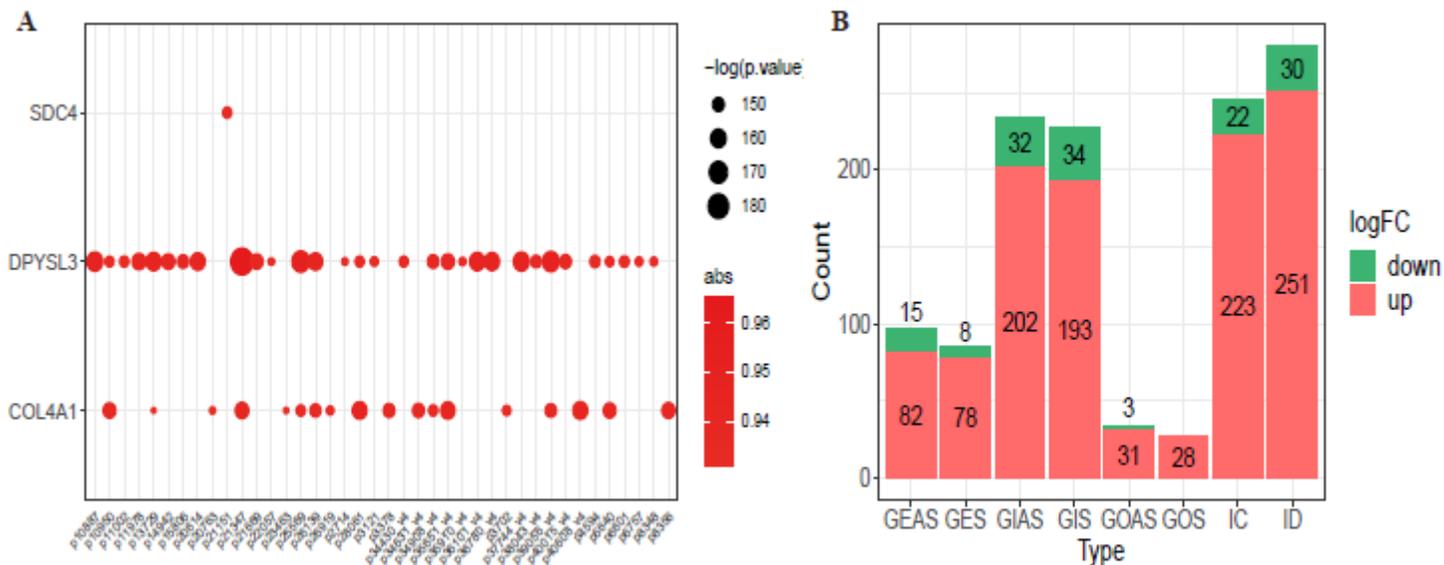


Figure 2

Identification of EMT-related lncRNAs in CAD. (A) Correlation analyses for EMT-related lncRNAs. (B) Classification bar diagram of EMT-related lncRNAs.

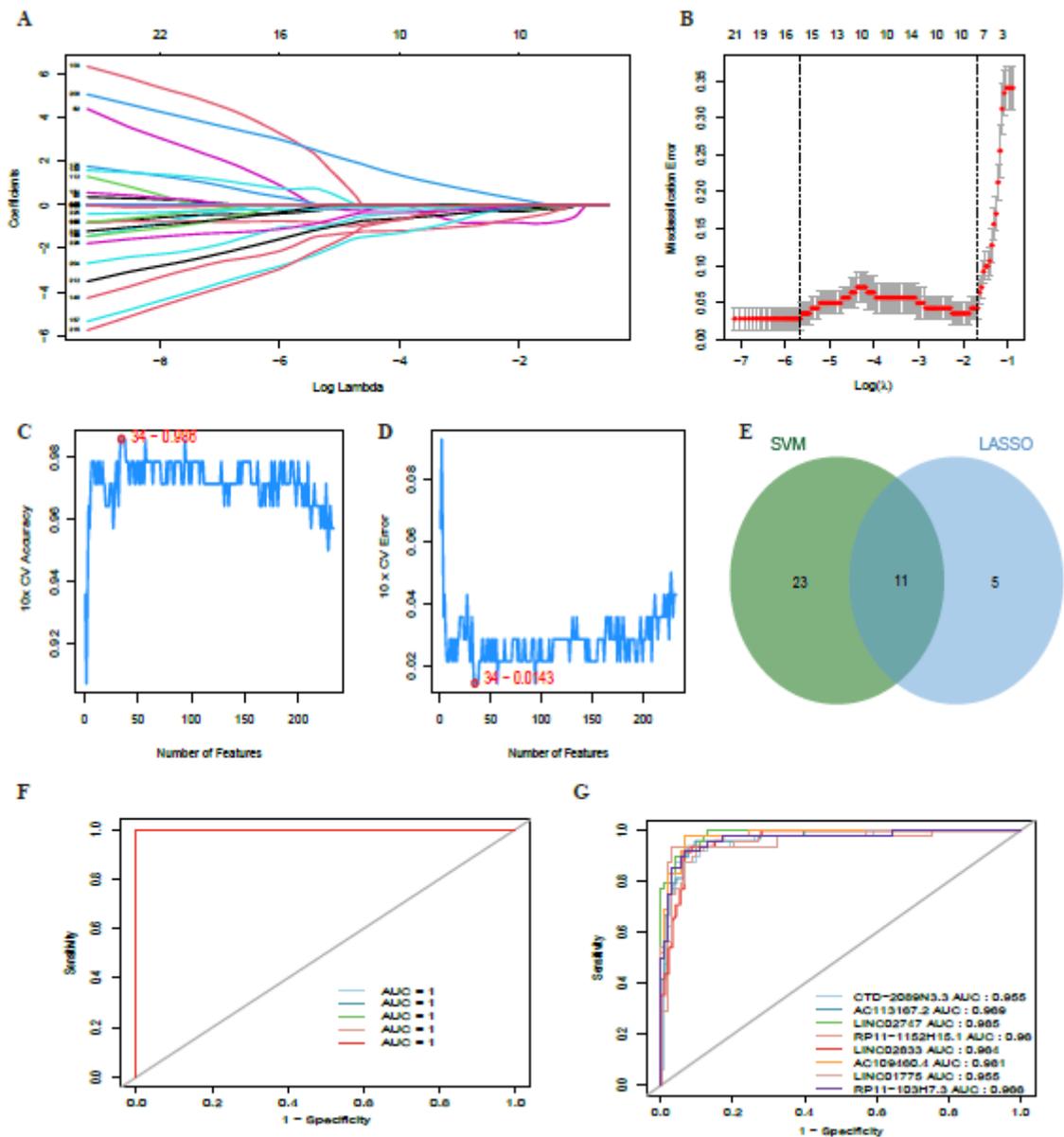


Figure 3

EMT-related lncRNAs diagnostic signature for CAD. (A) LASSO coefficient profiles of the 16 EMT-related lncRNAs selected by the optimal lambda. (B) The diagnostic signature selection of optimal parameter (lambda) in LASSO model. (C-D) Results of SVM-RFE algorithms: the point highlighted indicates the lowest error rate, and the corresponding EMT-related lncRNAs at this point are the best signature selected by SVM-RFE. (E) Venn diagram of overlap EMT-related lncRNAs selected by LASSO and SVM-RFE algorithms. (F) The ROC curve of predicted outcomes of eight EMT-related lncRNAs diagnostic signature by a logistic regression model. (G) ROC analysis results for eight EMT-related lncRNAs in CAD.

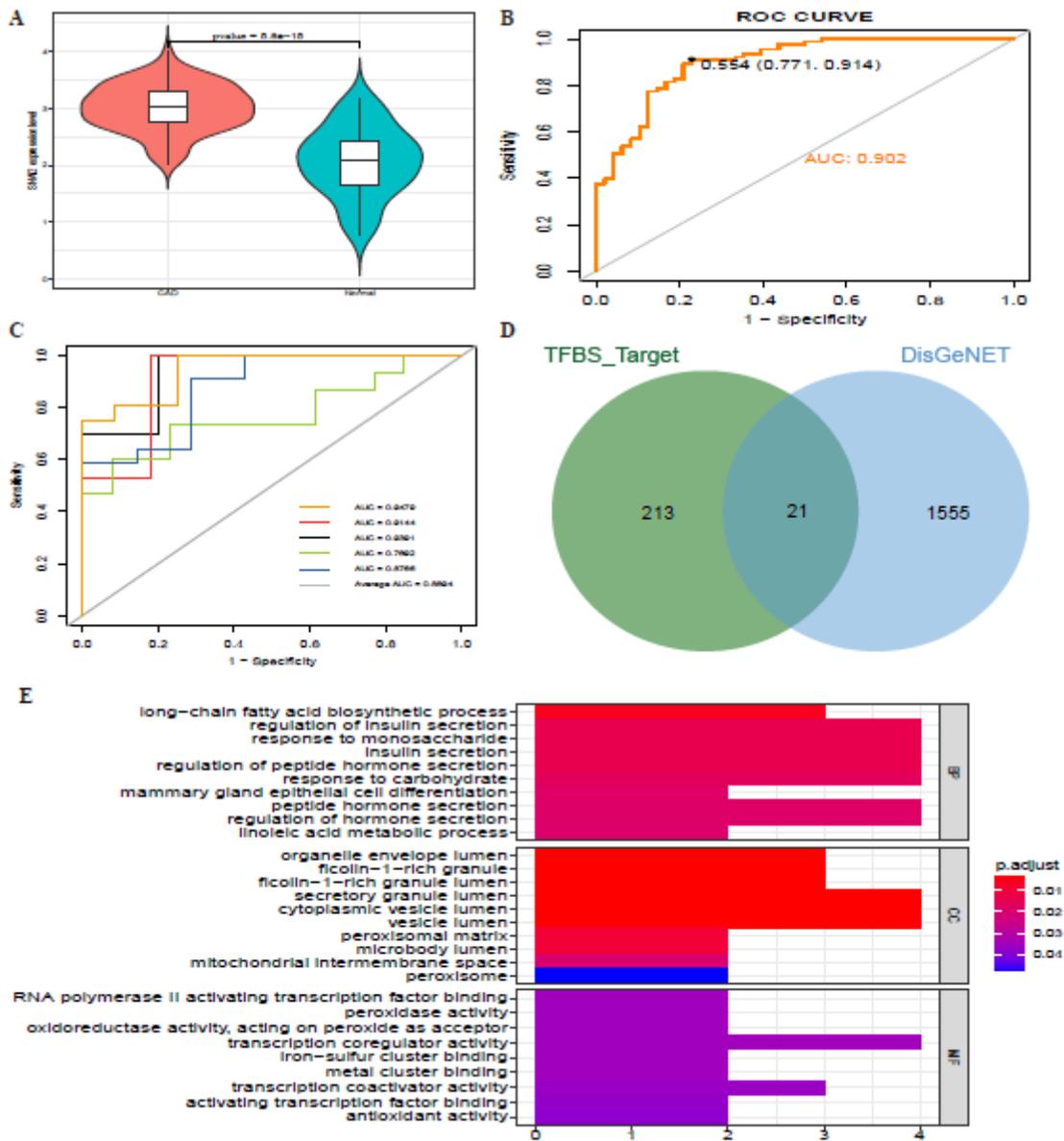


Figure 5

Clinical performance of SNAI2 in CAD. (A) SNAI2 is highly expressed in CAD groups and low in normal groups. (B) ROC analysis results for SNAI2 in CAD. (C) The ROC curve of predicted outcomes of SNAI2 diagnostic signature by a logistic regression model. (D) Venn diagram of 21 CAD-related genes regulated by SNAI2. (E) GO enrichment analysis of 21 CAD-related genes regulated by SNAI2.

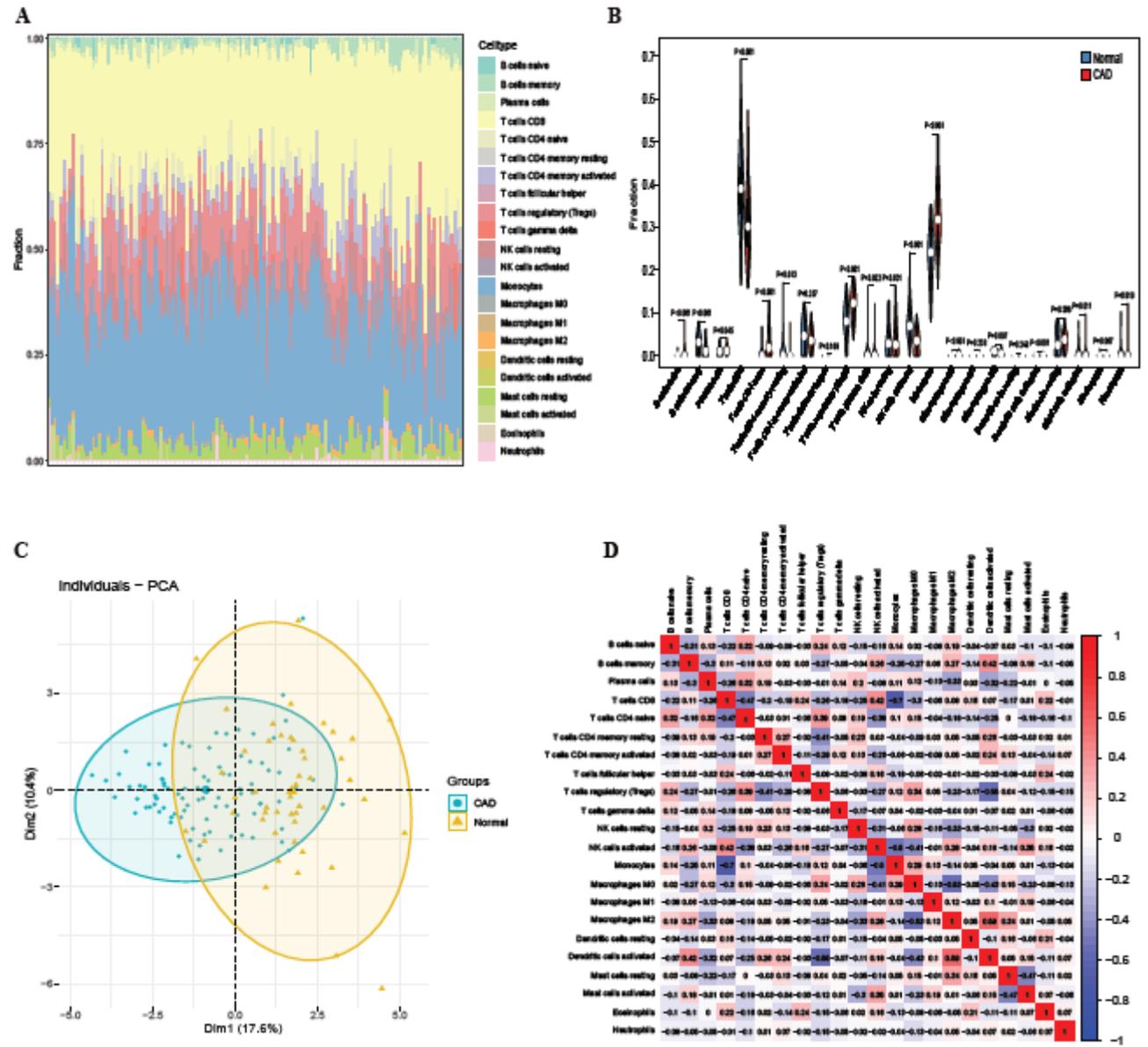


Figure 6

Evaluation and visualization of immune cell infiltration by CIBERSORT algorithm. (A) immune cell types and ratios of CAD groups. (B) Violin plot comparing immune cell compositions between CAD groups and normal groups. (C) PCA for immune cell compositions between CAD groups and normal groups. (D) Pearson correlation heatmap between infiltrating immune cell subpopulations.

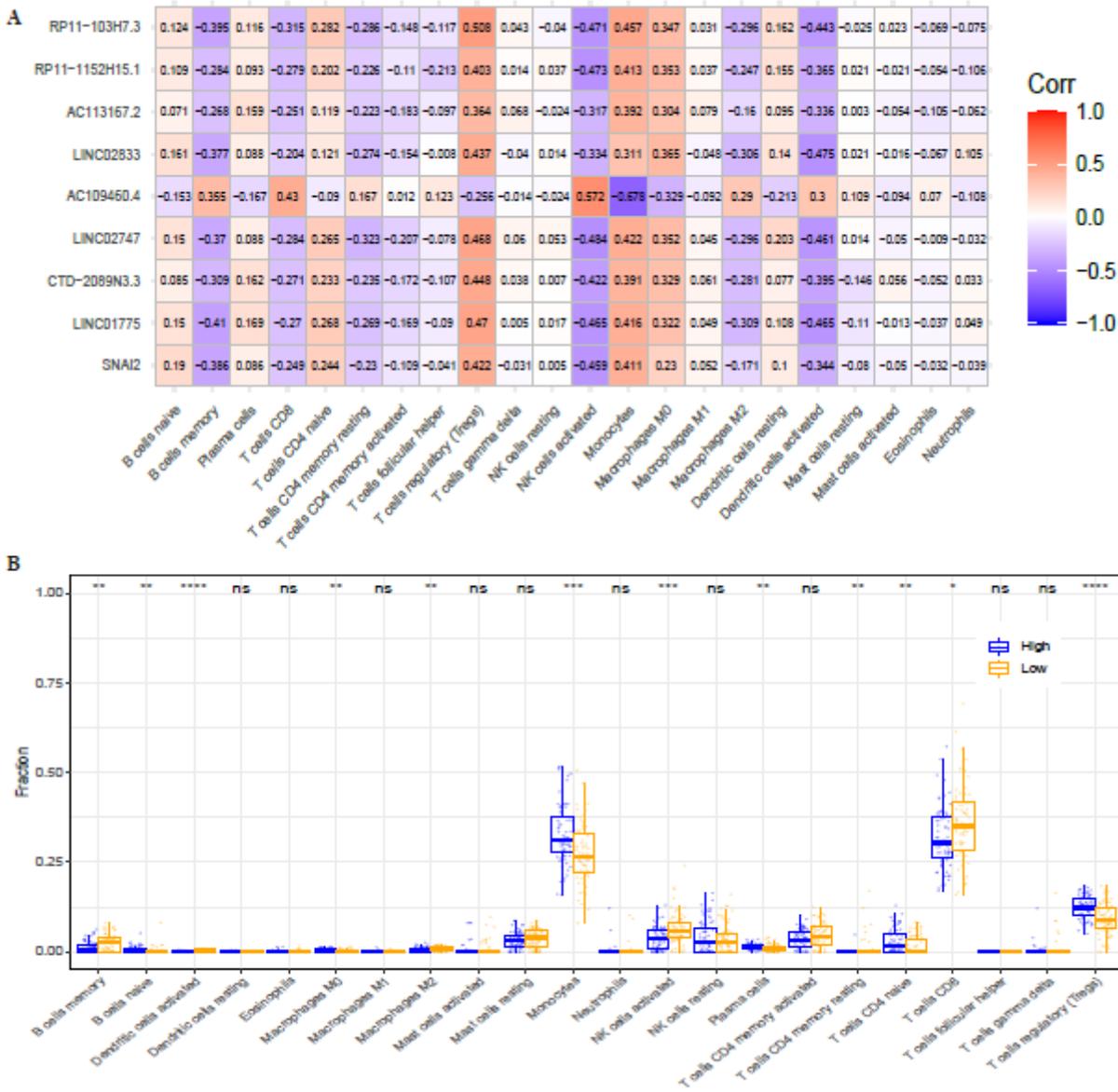


Figure 7

Immune analysis of SNAI2. (A) Pearson correlation heatmap between eight EMT-related diagnostic signatures and infiltrating immune cells. (B) Boxplot comparing immune cell compositions between high SNAI2 expression level groups and low SNAI2 expression level groups.

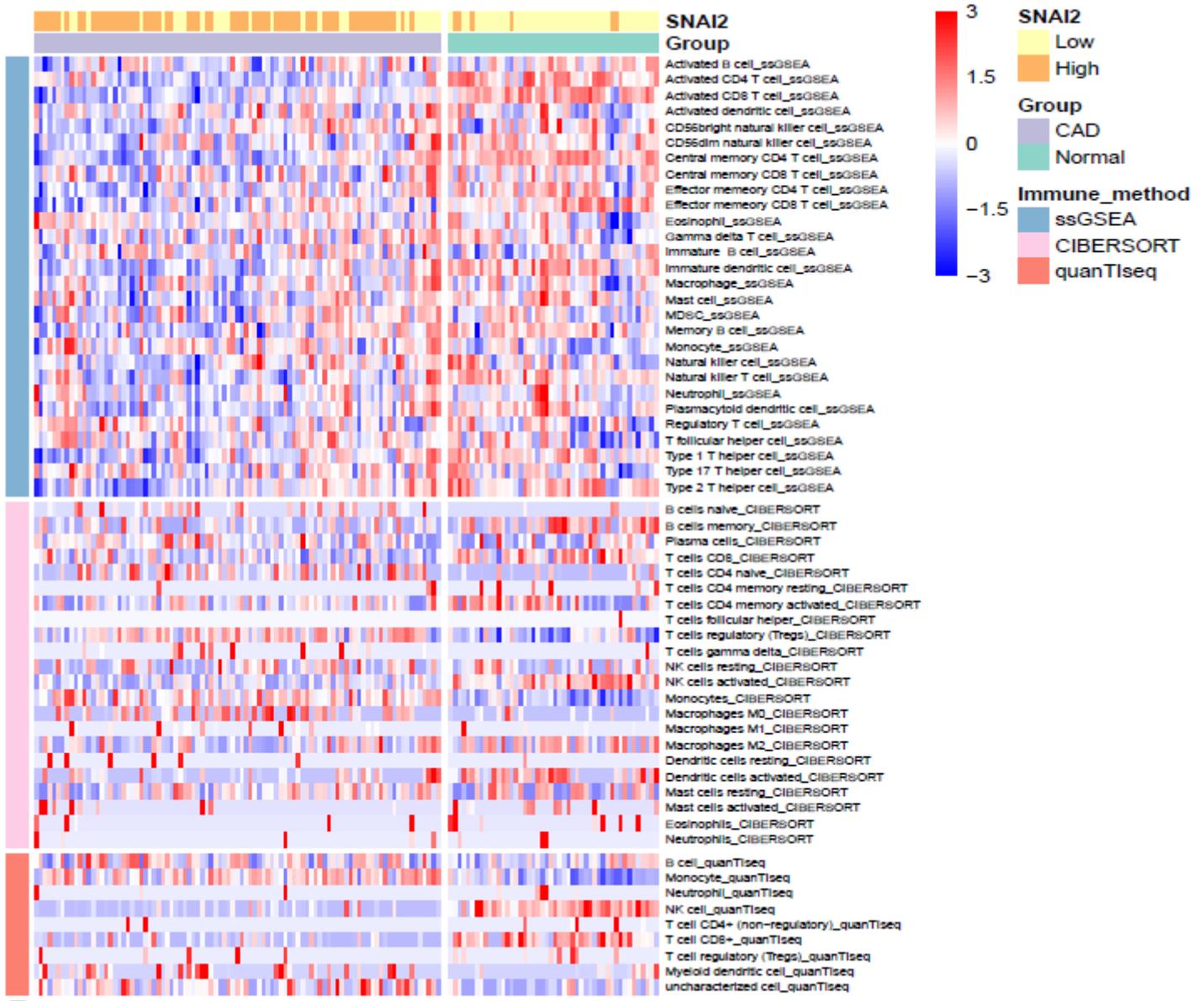


Figure 8

Heatmap for immune cell compositions based on ssGSEA, CIBERSORT, and quanTiseq algorithm among different groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig.1.pdf](#)
- [SupplementaryFig.2.pdf](#)
- [SupplementaryFig.3.pdf](#)
- [SupplementaryFig.4A.pdf](#)
- [SupplementaryFig.4B.pdf](#)

- [SupplementaryFig.5.pdf](#)
- [SupplementaryFig.6A.pdf](#)
- [SupplementaryFig.6B.pdf](#)
- [SupplementaryFig.6C.pdf](#)
- [SupplementaryFig.6D.pdf](#)
- [SupplementaryFig.6E.pdf](#)
- [SupplementaryFig.6F.pdf](#)
- [SupplementaryFig.6G.pdf](#)
- [SupplementaryFig.6H.pdf](#)
- [SupplementaryFig.6I.pdf](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTable7.xlsx](#)