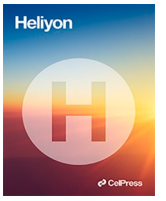
Heliyon 10 (2024) e28685

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Key chromatin regulator-related genes associated with the risk of coronary artery disease regulate the expression of HCFC1, RNF8, TNP1 and SET

Wang Bingyu [a,1], Yang Xi [a,b,1], Lian Jiangfang [a,b], Zhou Jianqing [a,b,*]

[a] Department of Cardiovascular, Ningbo Medical Center Lihuili Hospital, Ningbo, China
[b] Ningbo Institute of Innovation for Combined Medicine and Engineering, Ningbo, China

ARTICLE INFO

Keywords:
Chromatin regulators
Coronary heart disease
HCFC1
RNF8
TNP1
SET

ABSTRACT

Chromatin regulators are indispensable upstream epigenetic regulators.The emergence and progression of atherosclerosis has been demonstrated to be influenced by smooth muscle-related chromatin regulators, such as ZEB2 and MAFF. However, specific chromatin regulators and their possible roles have not been clarified. Information was gathered from 51 patients diagnosed with coronary artery disease (CAD) and 50 individuals in good health from the GEO database. 440 genes were identified as having differential expression across the two datasets, and these genes were linked to cellular reactions. Enrichment of pathways related to histone modification and transcriptional regulatory factors was observed in GO and KEGG analyses. Four machine learning models (RF, SVM, GLM, and XGB) were developed using the expression profiles of 440 chromatin-associated genes in the CAD cohort to pinpoint genes with significant diagnostic potential. After evaluating residuals, root mean square errors, receiver operating characteristic curves, and immune-infiltration, four key genes (HCFC1, RNF8, TNP1, and SET) were identified. Gene expression in different blood vessel levels in atherosclerotic plaques in a mouse model of coronary artery disease showed significant variations. The gene expression levels in macrophages aligned with clinical data from the GEO database as expected. This discovery is crucial for future analysis and the prediction of drug and miRNA targets. In conclusion, we found that the four hub genes are important in the mechanism of CAD. These findings provide new ideas for the study of potential epigenetic predictive markers and therapeutic targets to be used in determining a treatment strategy for CAD.

1. Introduction

Epigenetic regulation involves genomic imprinting, DNA methylation, RNA editing and histone modifications at the transcriptional level [1]. Acting as a connection between genetic elements, it does not change the DNA sequence directly, but regulates gene expression through different methods like DNA methylation, histone modification (including methylation, acetylation, and ubiquitination), chromatin remodeling, noncoding RNA editing, and other processes that control genes [2–4].

At first, it was thought that epigenetics was involved in the process of cell specialization as organisms develop.However, there is

* Corresponding author.Department of Cardiovascular, Lihuili Hospital Facilitated to Ningbo University, Ningbo University, Ningbo, China.
E-mail address: lhlzhoujianqing@nbu.edu.cn (Z. Jianqing).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e28685
Received 28 June 2023; Received in revised form 12 March 2024; Accepted 22 March 2024
Available online 29 March 2024

Abbreviations

ASXL3	Additional sex combs like transcriptional regulator 3
AS	Atherosclerotic disease
β-ET	β-extra terminal protein
BRD	Bromine domain
BRDT	Bromodomain testis-specific protein
CAD	Coronary artery disease
CRs	Chromatin regulators
CRDGs	Chromatin regulators different genes
DCA	Decision curve analysis
DCs	Dendritic Cells
DGIdb	Drug–Gene Interaction Database
FDR	False Discovery Rate
GLM	Generalized Linear Model
GO	Gene Ontology
HCFC1	Host cell factor C1
IMF	Immune function
IMC	Immune cell
KEGG	Kyoto Encyclopedia of Genes and Genomes
SET	SET nuclear proto-oncogene
SVM	Support Vector Machine,
TNP1	Transition protein 1
PPI	Protein–protein interaction
RNF8	Ring finger protein 8
ROC	Receiver operating characteristic
RMSE	Root mean square of residuals error
RF	Random Forest
XGB	eXtreme Gradient Boosting
SMYD1	SET and MYND domain-containing protein 1
SMCs	Smooth muscle cells
TET1	Ten-eleven translocation-1
TAF5	Transcription initiation factor TFIID subunit 5
TGF-β1	Transforming growth factor-β1
TIL	Tumorin filtrating lymphocytes

growing recognition of its causal involvement in cardiovascular diseases and organ fibrosis.To understand this phenomenon better, it is essential to classify DNA methylators and histone modifiers based on their ability to encode and decode diverse cytosine and histone residue modifications [5]. Readers typically include a specific domain that identifies cytosines or their modified colored residues, and determines the state and type of modification [6,7]. In addition to adding and removing certain modifications, writers and erasers can methylate and demethylate specific cytosine or histone residues [6,7]. Chromatin remodelers are regulators of chromatin that alter nucleosomes, either replacing or eliminating them, leading to abnormal epigenetic modifications in cells.Atherosclerosis, also known as, is characterized by the involvement of various cell types such as endothelial cells, smooth muscle cells, and inflammatory macrophages [8]. According to recent research, smooth muscle cell (SMC) phenotypic transformation is controlled by two classical molecules [9]. The protein family containing the bromine domain and beyond the terminal domain includes BRD2, BRD3, and BRD4. Additionally, BRDT is part of this, which is a protein involved in reading epigenetic information and attaches to particular acetylated lysine residues located on the histone tail [10]. RNA polymerase II promotes the assembly of transcriptional complexes including transcription factors [10]. Research in cardiovascular environments has emphasized the importance of BET in regulating gene expression throughout differentiation, cell identity, and transitions between cell states, whether they are normal, abnormal, adaptive, or non-adaptive. The intricate genetic marker of CAD patients in the 2q22.3 region is found in the long-range enhancer of ZEB2, a well-researched transcription factor involved in epithelial mesenchymal transition during cancer progression.ZEB2 influences the change in characteristics of smooth muscle cells by altering chromatin structure, which hinders access and interferes with the Notch and TGF-β signaling pathways, ultimately changing the epigenetic location of smooth muscle cell transformation [11]. Other molecules such as neutrophil-related NETs and macrophage-related TFEB emphasize the importance of epigenetic inheritance in AS [12]. Nevertheless, there has not been a comprehensive study detailing the involvement of CRs in CAD to date. In this study, 870 CRs were analyzed to identify potential therapeutic targets and possible predictive markers for CAD, thereby providing new ideas for the treatment and diagnosis of CAD.

2. Materials and methods

2.1. Animals and experimental design

Six-week-old male ApoE$^{-/-}$ mice weighing between 20 and 22 g were acquired from GemPharmatech Co. in Jiangshu, China. Twenty ApoE$^{-/-}$ mice were given either a high-fat diet (40% of energy from fat, 20% fructose, and 0.8% cholesterol, ten mice) or a standard diet (20% of energy from fat, 20% fructose, and 0.01% cholesterol, ten mice) from Trophic Animal Feed High-Tech Co., Ltd., China, for a duration of 8 weeks. Serum and blood vessels were obtained, while tissues were preserved in 4% paraformaldehyde. Approval for this research was granted by the Institutional Animal Care and Ethics Committee at Ningbo University of Medicine (10489) and conducted following the NIH guidelines.

2.2. Blood lipid test

After incubating at room temperature for 2 h, the supernatant was obtained with mouse plasma, then centrifuged at 1000×*g* for 20 min at 4 °C, and finally stored at −80 °C. The LDL-C, HDL-C, TG, and TC concentrations were determined with kits following the

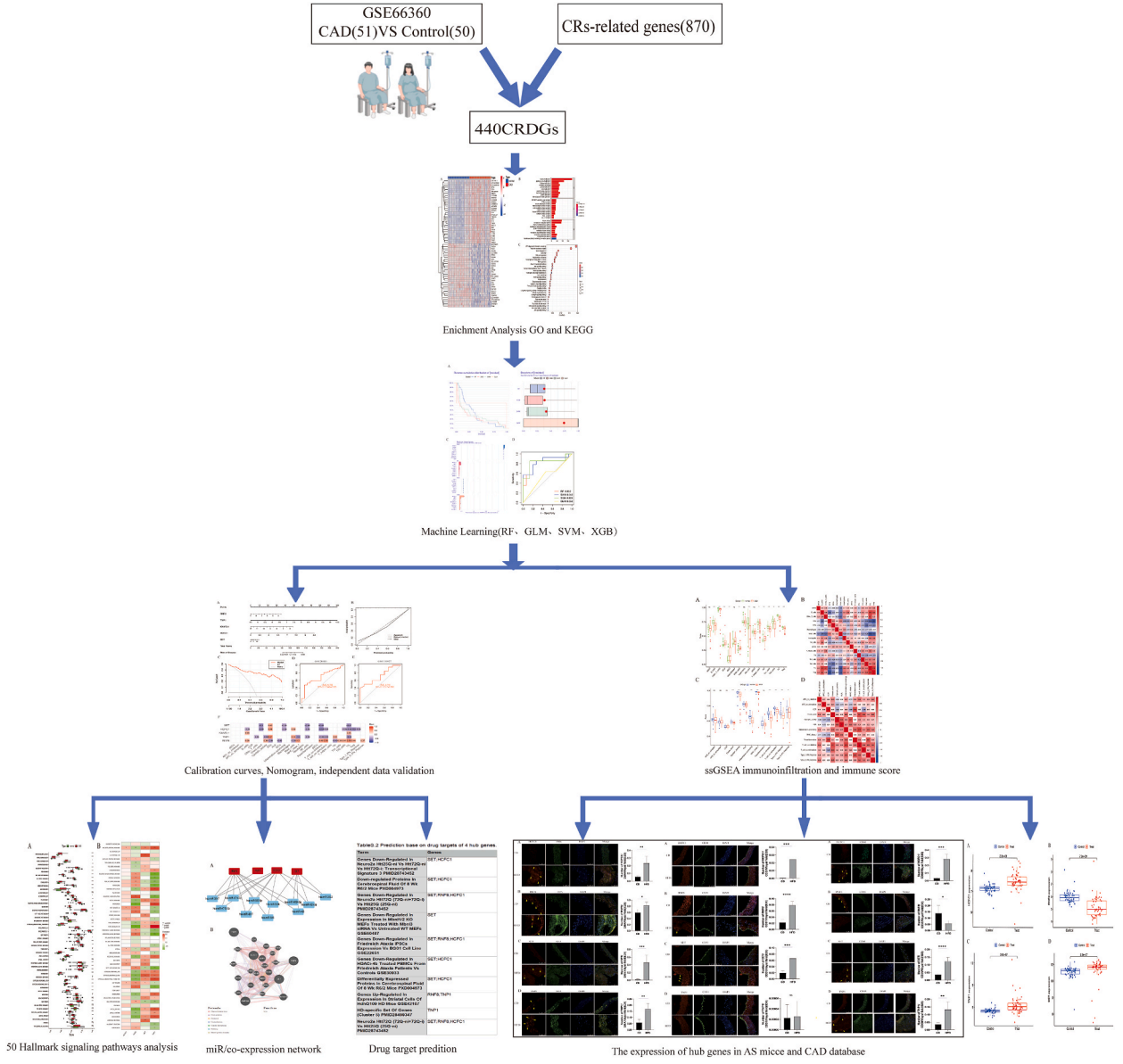


Fig. 1. Study flow chart.

manufacturer's guidelines (Elab. Biotech. Co., Ltd., Wuhan, China).

2.3. Histopathological staining

Extravascular tissue from the abdominal aorta of the mice was separated from the aortic arch. Oil Red O solution (Leigen Co., Ltd., China, DL0010) was added dropwise. A fine needle was used to spread the blood tube horizontally, and a macro lens was used for observation and photography.

2.4. Immunostaining

Aortic vessel samples were preserved in 4% paraformaldehyde for a night, then encased in paraffin before being sliced into 5-μm sections of diseased tissue with a Leica microtome. Hydration was performed using xylene and gradient alcohol dewaxing with water. The samples were incubated with 3% H2O2 for 20 min and washed thrice with phosphate-buffered saline. After high-pressure repair, sheep serum was blocked for 1 h and the corresponding primary antibody was incubated. Following an overnight incubation at 4 °C, the cells were then exposed to fluorescent secondary antibodies for 1 h and DAPI for 5 min the following day. Images were captured using a multi-monitor color digital camera on a laser confocal miniature display (Leica, USA).The fluorescence intensity of CD31/α-SAM/CD68/HCFC1/SET/RNF8/TNP1 was measured in 8–12 tissue images at 63 × magnification. Various antibodies including CD31, CD68, α-SMA, HCFC1, SET, RNF8, and TNP1 were used in the experiment, all at a dilution of 1:100 or 1:400, along with a multicolor fluorescence kit from China.

2.5. Identification of differentially expressed CRDGs

440 genes were found to be expressed differently in 51 patients with CAD compared to 50 healthy individuals, with the top 60 genes identified (Fig. 1A and B). In addition, 870 CRFGs were analyzed as described previously by Liu et al. [13]. A total of 841 CRs intersected with the differential genes in the sample. Following this, 440 CRDGs were chosen for additional examination based on standard of $|\log_2 FC| > 0.5$ and adjusted P-values <0.05 as shown in Table S1. Adjusted P values were calculated using FDR. FDR is calculated as the ratio of false positives to the sum of false positives and true positives, with true positives representing positive test results for positive samples and false positives representing positive test results for negative samples.

2.6. Pretreatment and detection of original expression signal and CRDGs

The original expressed signal was processed using the Bioconductor package of R software [14]. The sample data was standardized and noise was reduced by uploading the CEL file to Affymetrix Quality Assessment and Analysis Tool Bioconductor version 1.30.19. By applying robust multi-array mean methods, the unprocessed expression signal was normalized to the median [15]. Genes that exhibited a 1.5-fold change in expression and passed FDR (Benjamini and Hochberg's false discovery) rate correction with a p value of 0.05 or less were identified as CRDGs [14]. The limma package in R software was used to create maps of volcanoes and averages.

2.7. Functional annotation of CRDGs

Analysis of GO annotation and enrichment of KEGG pathways was performed using the Metascape database [16]. Significantly enriched pathways related to CRF genes were identified based on a p value threshold of <0.05.

2.8. Machine learning-based prediction model construction

The caret R package (version 6.0.93) was utilized to create ML models using two distinct CRG clusters, incorporating the RF, SVM, GLM, and XGB algorithms. The RF [4], acting as a collective machine learning technique, employs numerous separate decision trees to make predictions on classifications or regressions [17]. The Support Vector Machine (SVM) algorithm creates a hyperplane within the feature space to separate positive and negative instances with the largest possible margin [18]. GLM is a multivariate linear regression method that can be used to analyze the relationships between categorical or continuous factors and normally distributed dependent factors [19]. Based on gradient boosting, XGB compares the classification errors and complexity by comparing the boosted trees [20]. DEGs were used as explanatory variables in this study in conjunction with distinct clusters as response variables. Fifty-one samples of CAD were randomly divided into a training set consisting of 75% (numbers = 38) and a validation set consisting of 25% (numbers = 13). The caret package was used to conduct a grid search to optimize the parameters of the models, which were then assessed with default settings through 5-fold cross-validation. We utilized the pROC R package (version 1.18.0) to display AUC. The top five predicted genes were identified as the most important variables for predicting CAD. AUC were analyzed in datasets GSE20680 and GSE100927 to confirm the diagnostic accuracy of the model.

2.9. Construction and validation of a nomogram model

A nomogram model was created to predict CAD using the rms R package (version 6.2.0).A score was given to each predictor, and the total score was determined by adding up all the individual scores.Calibration curve and DCA were used to assess the predictive

accuracy of the nomogram.

2.10. Assessment of immune cell infiltration and its correlation with hub genes

The ssGSEA algorithm was utilized to measure immune cells (IMCs) and immune functions (IMFs) for the purpose of creating an immune map [21]. The Wilcoxon test compared the IMC and IMF fractions in patients with CAD and healthy participants, with a filter set at 0.05. The Spearman's rank correlation coefficient was used to establish the relationship between overlapping genes and immune components, providing essential insights into CRDG-related immune landscapes. The majority of researchers concur that a coefficient below 0.4 suggests a minimal connection, while a coefficient above 0.4 indicates a significant correlation between immunity and CRDGs [22,23].

2.11. Enrichment analysis of the ssGSEA

The signature gene set score was obtained by performing ssGSEA to investigate functionally enriched pathways and hallmark gene sets associated with subgroups using the R GSVA software package [24]. The hallmark (h.all.v7.3) gene sets were obtained from the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/).The correlation between CAD, molecular function, and gene expression was evaluated using Spearman's correlation coefficient. Significantly enriched items were those with a P value less than 0.05.

2.12. Gene-drug interactions

The druggability of genes was assessed by utilizing the drug-gene interaction database known as DGIdb [25]. DGIdb serves as a central database for housing information on interactions between drugs and genes, as well as aggregating information on the potential for drugs to target each gene from various origins. Default filters, such as cytotoxic drugs, immunotherapy drug interactions, approved drugs from a database of nine sources unrelated to the disease, were used in the search. The drug-gene interactions were filtered based on interaction type and gene class, with a critical interaction score above 0.03.

2.13. Target miRNA prediction of hub genes

In this study, the miRNA target prediction tool Enrichr TargetScan 2017 (https://maayanlab.cloud/Enrichr/enrich#) was used to predict target miRNA genes [25]. MicroRNAs (miRNAs) are abundant in eukaryotic organisms and display important biological characteristics including strong conservation, temporal expression patterns, and tissue-specific expression patterns.The significant conservation of miRNAs suggests the presence of homologues in germlines. The dataset GSE66360 was utilized to investigate stage-specific miRNA profiles in atherosclerosis progression, potentially aiding in the identification of crucial miRNAs involved in atherosclerotic pathogenesis. Validation was limited to target genes that overlapped with hub genes. It was considered a key gene to forecast target genes predicted by multiple miRNAs.miRNAs of human origin ($p < 0.05$) were selected, and four key gene interaction networks were constructed and visualized using Cytoscape software 3.7.1.

2.14. Statistical analysis

All data processing, statistical analyses, and plotting were conducted using R version 4.1.0. The importance was evaluated by utilizing either Student's t-test or Wilcoxon rank-sum test when comparing two groups. For the establishment of relationships between continuous variables, Pearson's correlation coefficient was used. The pROC R package was utilized to conduct analysis on the AUC. All tests were conducted with a two-tailed approach, and significance was determined at a $p < 0.05$.

3. Results

3.1. Flow chart of the analysis process

Fig. 1 displays a flowchart outlining the bioinformatics analysis.Using the microarray dataset GSE66360, we used $|\log_2 FC$ calculated differentially expressed mRNA$| > 0.5$ after adjustment for P-values <0.05.Liu et al.reported readings for 870 CRs [13]. Moreover, 870 CRs intersected with the differential genes in the sample.440 genes showed differential expression in both datasets and were associated with CRs. GO and KEGG analyses revealed that histone modifications and transcriptional regulatory factor-related pathways were enriched. In order to pinpoint particular genes that are highly useful for diagnosis, we created four ML models using the expression profiles of 440 CRDGs in the CAD group. Following assessments of residual and RMSE as well as ROC analysis, five crucial genes (HCFC1, RNF8, TNP1, KANSL1, and SET) were discovered. Examination of immune cells, immune function, and their correlation indicated that KANSL1 had a weak association with immunity, leading to its exclusion from further discussion. Following this, a mouse model was created to study AS, revealing notable differences in gene expression at various levels of blood vessels in AS plaques, providing valuable insights for future investigations. Finally, the drug and miRNA targets were predicted. To summarize, we have pinpointed four genes (HCFC1, RNF8, TNP1, and SET) that could have significant implications in the development of CAD.As a result of this study, it has been possible to develop ways to study epigenetics, identify potential predictive markers, and identify therapeutic

targets for the treatment of CAD.

3.2. Chromatin-related differential gene screening, GO and KEGG in patients with CAD and healthy subjects

We identified 440 differentially expressed genes in 51 patients with CAD and 50 healthy subjects (Fig. 2A, top 60). Liu et al. reported readings for 870 CRF genes [13]. Moreover, 870 CRs intersected with the differential genes in the sample. Subsequently, 440 CRDGs underwent additional analysis ($|\log_2 FC| > 0.5$, adjusted P-values <0.05; Table S1). Analysis of the 440 CRDGs showed GO enrichment in histone modifications, binding regions, and transcriptional regulatory activators (Fig. 2B). The majority of enriched KEGG functional categories, such as ATP-dependent chromatin remodeling pathway, polycomb repressive complex, and lysine degradation, were also observed in Fig. 2C.

3.3. Construction and assessment of machine learning models

In order to pinpoint subtype-specific genes that are highly valuable for diagnosis, we created four machine learning (ML) models (RF, SVM, GLM, and XGB) using the expression profiles of the 440 CRDGs in the CAD cohort. The residual distributions of the four ML models were analyzed and visualized using the "DALEX" package. The RF and GLM machine learning algorithms showed relatively low residuals (Fig. 3A–B).Afterwards, the most important 15 feature variables of each model were ordered according to the RMSE (refer to Fig. 3C).Furthermore, the effectiveness of the four ML models in the test set was assessed by generating ROC curves with 5-fold cross-validation. The RF machine learning model described the highest area under the ROC curve (GLM, area under ROC curve = 0.552; SVM, area under ROC curve = 0.843; RF, area under ROC curve = 0.852; XGB, area under ROC curve = 0.829; Fig. 3D). Ultimately, the RF model demonstrated superior performance in distinguishing patients across various clusters. Finally, the five most important variables (HCFC1, RNF8, TNP1, KANSL1, and SET) from the RF model were selected as predictor genes for further analysis.

In order to evaluate the accuracy of the RF model, we initially developed a nomogram to predict the likelihood of important

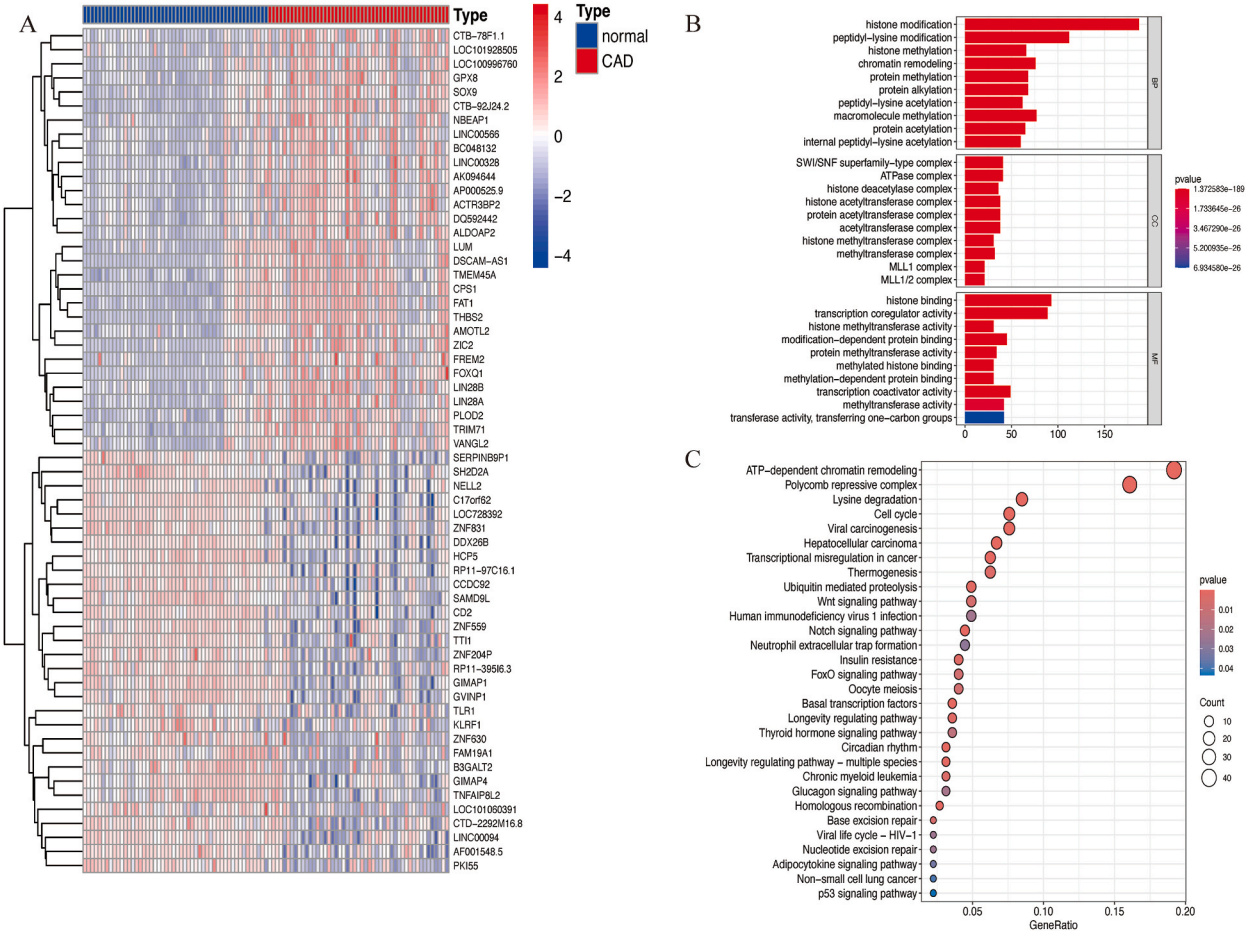


Fig. 2. Functional enrichment analysis of chromatin regulator-related differentially expressed genes (CRDGs) in the CAD GSE66360 dataset. (A) Heatmap showing the top 60 CRDGs in CAD. (B) GO analysis of CRDGs in GSE66360. (C) KEGG analysis of CRDGs in GSE66360.

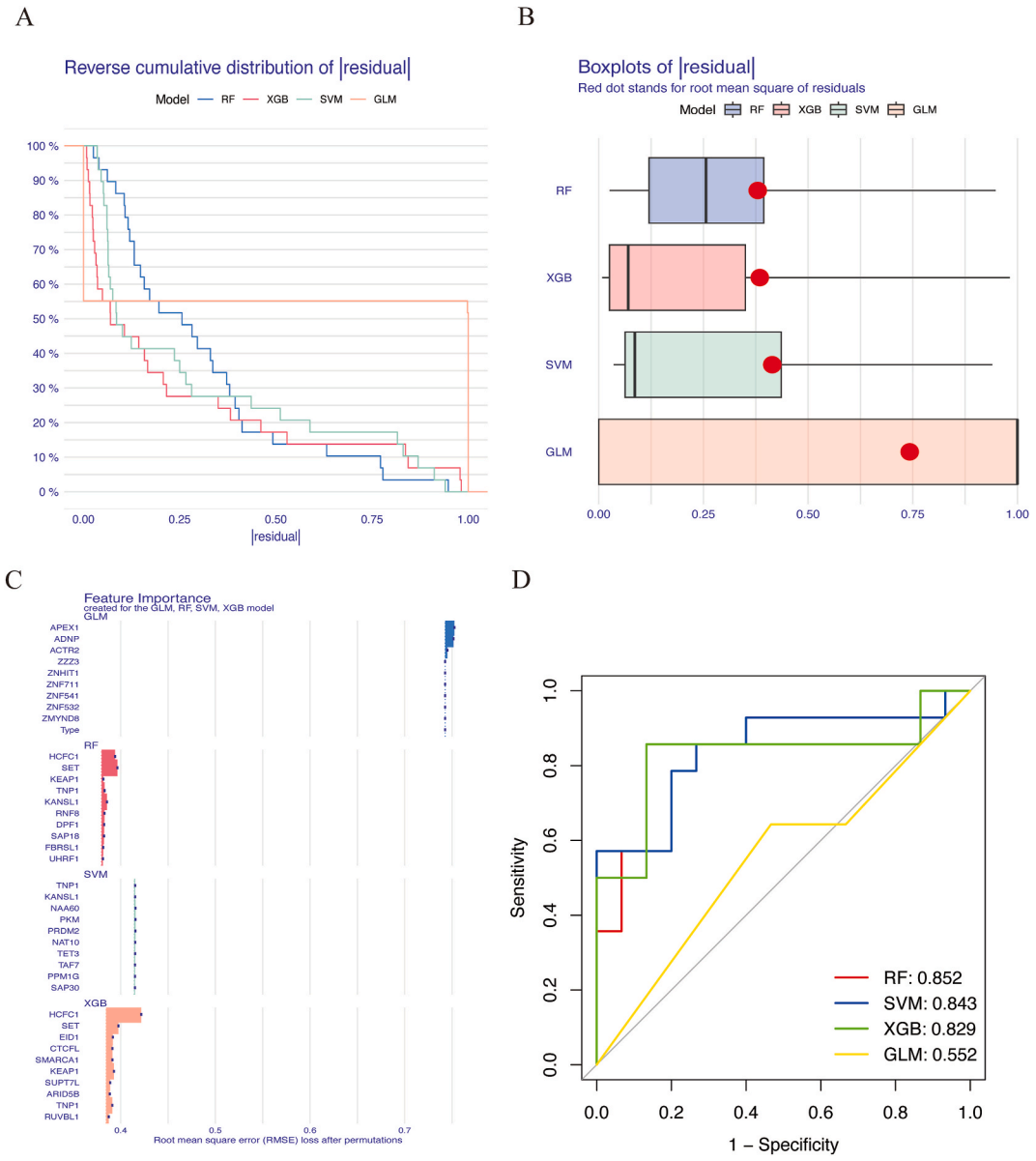


Fig. 3. Construction and evaluation of the RF, SVM, GLM, and XGB machine learning(ML) models. (A) The cumulative residual distribution of each ML model is analyzed. (B) Boxplots are utilized to display the residuals of each ML model, with red dots indicating the RMSE. (C) The significant features in the RF, SVM, GLM, and XGB machine learning models are identified. (D) The ROC analysis is conducted on four ML models using 5-fold-validation in the testing cohort. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

chromatin regulator-associated genes in 51 CAD patients (Fig. 4A). The nomogram's predictive efficiency was evaluated through calibration and DCA. The calibration curves demonstrated minimal error between actual CAD and predicted risk, as depicted in Fig. 4B, while the DCA indicated high accuracy of the nomogram (Fig. 4C).Validation of the prediction model in external datasets showed good performance, with AUC values of 0.703 and 0.718 in the GSE20680 and GSE100927 datasets, as shown in Fig. 4D and E.The findings indicate that the diagnostic model successfully differentiates between CAD patients and healthy subjects. The inclusion of KANSL1 in the ROC analysis impacted the overall results of the database, leading to its exclusion from further analysis.

3.4. Infiltration state of immune cells and their relationship with hub gene expression

Immunity-related studies have been screened as an integral part of the understanding of CAD. Inflammation presents a clear risk of immune cell infiltration, and any dysfunction in the immune system can amplify the inflammatory response. Heatmaps, which

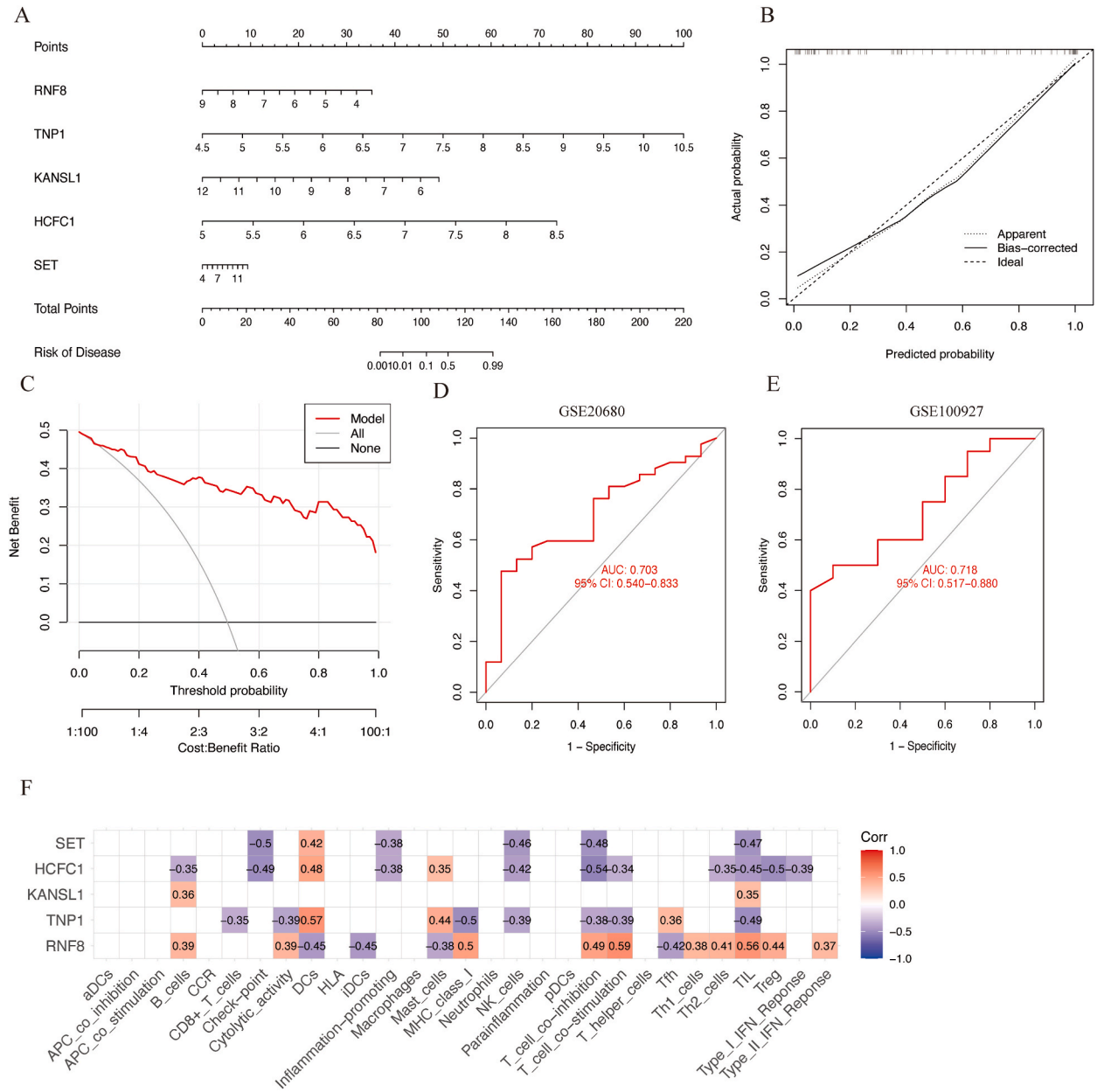


Fig. 4. Validation of the 5-gene-based RF model. (A) Based on a five-gene-based RF model, the nomogram model is utilized to predict CAD risk. (B, C)The calibration curve (B) and DCA(C) were constructed to access the predictive accuracy of the nomogram model. (D, E)The ROC analysis was conducted on the 5-gene-based RF model using 5-fold-validation in the GSE20680 (D) and GSE100927 (E) datasets. (F) Correlation scores of 5 hub genes with immune cells and their functions.

represent the correlations between IMCs or IMFs in batch-corrected validation cohorts (Figures S1A 1C), revealed that the interactions between immune components are associated with the onset and progression of CAD.A Wilcoxon test with a p-value less than 0.05 indicates a significant difference in immune composition between CAD patients and controls on the boxplot (Figure S.1B and 1D).More importantly, Spearman correlation analysis showed that SET and HCFC1 were negatively correlated with NK cells, TIL, checkpoints, inflammation promotion, and T-cell co-inhibition and positively correlated with DCs (Fig. 4F). Additionally, TNP1 and RNF8 levels were inversely correlated with the number of immune cells. In Fig. 4F, there was a difference in RNF8 expression when TNP1 showed a negative correlation with cytolytic activity, MHC class I, T-cell co-inhibition, T-cell co-stimulation, and TILs. From the perspective of immunity, the possible reasons for the inconsistent expression of TNP1 and RNF8 in patients with CAD and healthy individuals are also well explained.

3.5. Analysis of ssGSEA marker gene set of hub genes

We investigated the functions of the four central genes in the subsequent signaling cascades. We first used ssGSEA to analyze the signaling pathways enriched by the four genes, particularly the upregulated and downregulated pathways (Figure S2A-D). Subsequently, the marker gene set was utilized to enhance the marker signaling pathways associated with the four hub genes. The findings indicated notable variances in the inflammatory and metabolic pathways, specifically in bile acid, fatty acid, xenobiotic, IL2-STAT5, and heme metabolism (Figure S3A-B). The enhancement of these distinct pathways indicates that CRs are crucial in the onset and progression of CAD. Considering the complex signaling processes and phenotypic modification of genes in vivo, this phenomenon warrants further study.

3.6. Expression of the four genes in CAD tissues

Initially, we examined the genes HCFC1, RNF8, TNP1, and SET utilizing clinical information from the database. The findings indicated that the levels of HCFC1, TNP1, and SET were elevated in individuals diagnosed with CAD (See Fig. 5A–C-D). However, the RNF8 expression was higher in the control group (Fig. 5B).In order to confirm the activity of these four genes in a living organism, ApoE$^{-/-}$ mice were given a high-fat diet for 8 weeks. Oil Red O, HE, and Masson's staining were employed to confirm the extent of AS lesions in the mice (Figure S4A-C,E). Enzyme colorimetry revealed alterations in lipids, with elevated TG, TC, and LDL-C levels observed in the HFD group compared to the control group, while HDL-C levels were decreased (Figure S4D). These results demonstrate that AS modeling was successful. The expression of HCFC1, RNF8, TNP1, and SET in the blood vessels was detected using immunofluorescence staining (Figs. 6–8). CD31 is used to identify the endothelial cell layer, α-SMA to identify vascular smooth muscle cells in the media layer, CD68 to identify macrophages, and DAPI to identify the nucleus.HCFC1, RNF8, SET, and TNP1 showed increased expression in the smooth muscle cells of the HFD group. Our results revealed no notable difference in the expression of HCFC1 and TNP1 in the endothelial cells of normal and HFD-fed mice.In contrast, the endothelial cells of mice fed a high-fat diet exhibited reduced RNF8 expression compared to those fed a normal diet (Fig. 7B). The expression of SET aligned with the anticipated outcomes in our dataset, with HFD-fed mice displaying increased expression in endothelial cells (Fig. 7C). Additionally, our database analysis confirmed that the levels of HCFC1, SET, and TNP1 expression in CD68^{+} macrophages were notably elevated in the group fed a high-fat

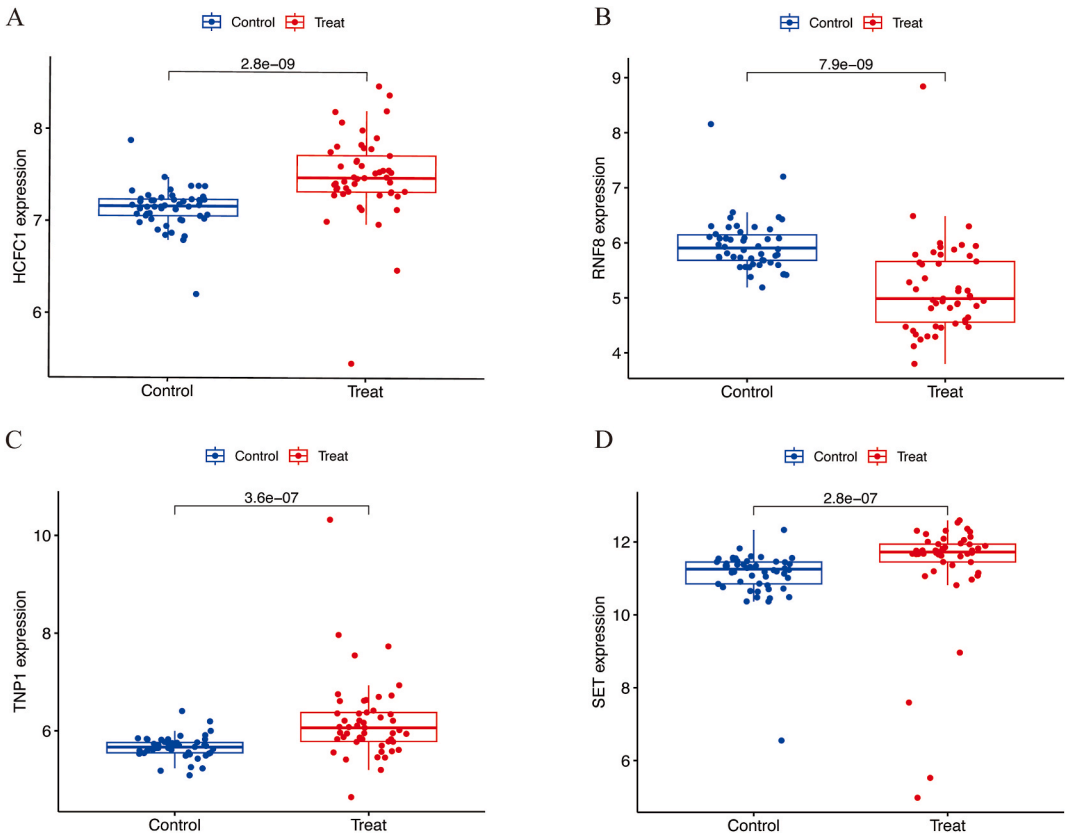


Fig. 5. Analysis of the expression level of four key genes and correlation analysis of CRDGs and clinical factors. (A) mRNA expression of HCFC1. (B) mRNA expression of RNF8. (C) mRNA expression of TNP1. (D) mRNA expression of SET. Control group included healthy subjects whereas Treat group included patients with coronary atherosclerotic disease.

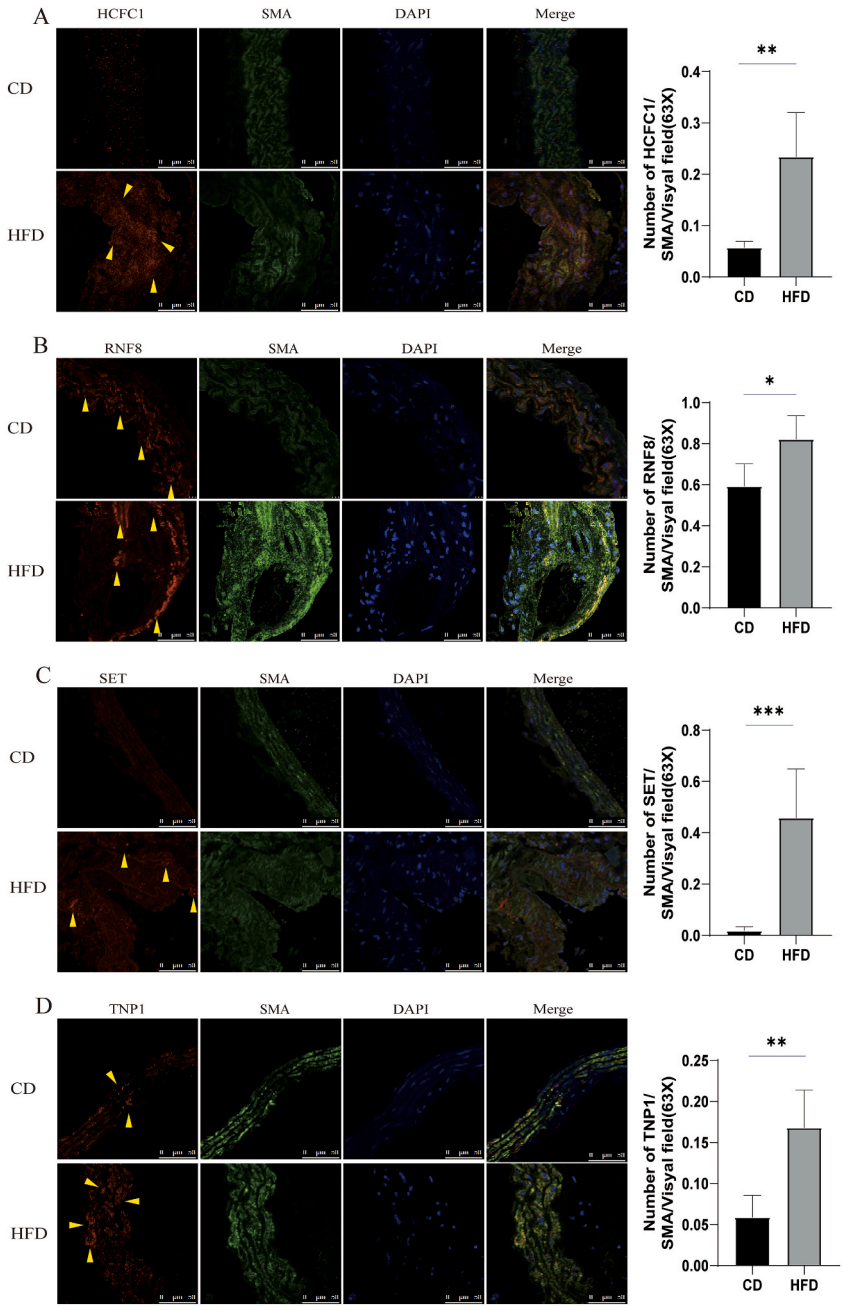


Fig. 6. Expression of four hub genes in the α-SMA-labeled smooth muscle cells. (A–D) The expression of (A) HCFC1, (B) RNF8, (C) SET, and (D) TNP1 in mouse arterial vessels was detected by immunofluorescence. Green represents smooth muscle cells, DAPI represents the nucleus, and red represents target proteins. Positive cells are marked with triangles. The magnification is 630. *$p < 0.05$ compared to the CD group. Data are shown as the mean ± SEM (n = 5). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

diet (HFD) (Fig. 8A,C-D). RNF8 was significantly more abundant in the macrophage layer of the CD group compared to the HFD-fed group.This finding aligns with the results obtained from our database analysis, indicating that RNF8 expression was observed in both the HFD and CD groups but was more pronounced in the CD group (Fig. 8B). Consequently, RNF8 may exert a substantial influence on immune metabolism and contribute to the progression of AS.

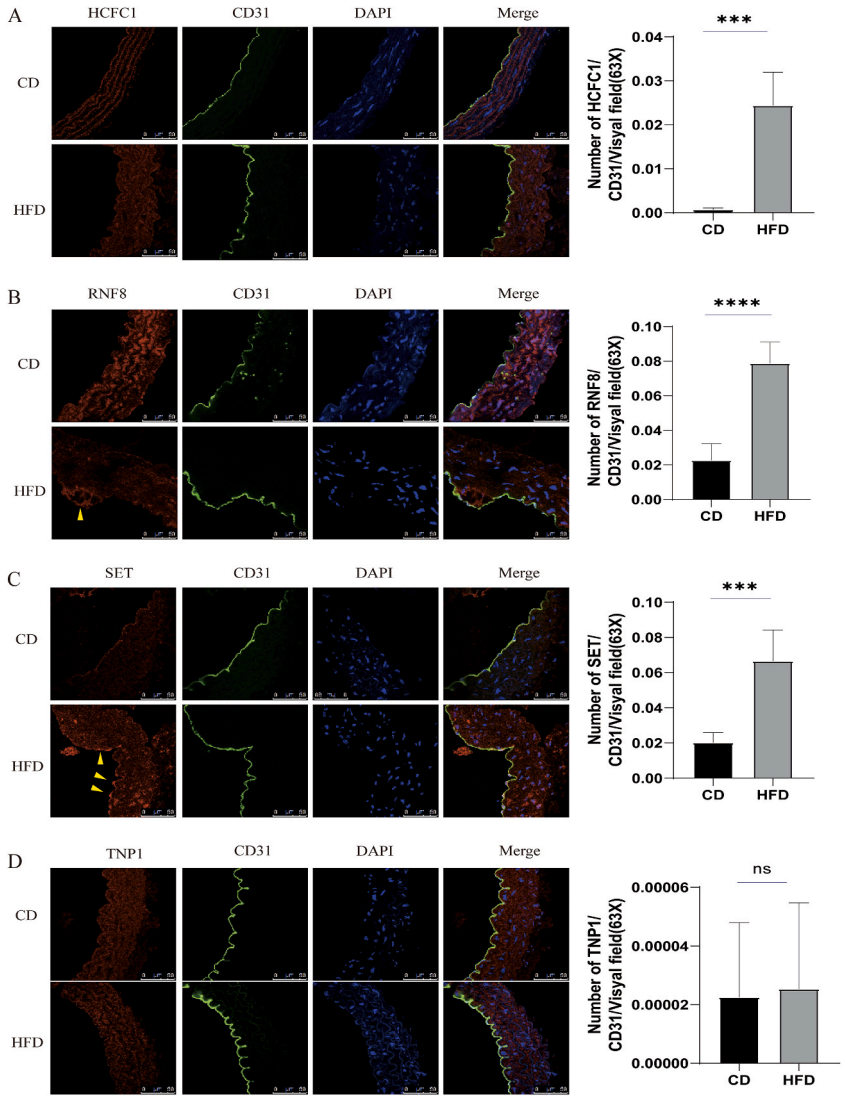


Fig. 7. Expression of four hub genes in the CD31-labeled endothelial cells. (A–D) The expression of (A) HCFC1, (B) RNF8, (C) SET, and (D) TNP1 in mouse arterial vessels was detected by immunofluorescence. Green represents endothelial cells, DAPI represents the nucleus, and red represents target proteins. Positive cells are marked with triangles. The magnification is 630. $^{*}p < 0.05$ compared to the CD group. Data are shown as the mean ± SEM (n = 5). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.7. Forecasts derived from the drug and miRNA targets associated with the four central genes

We used Maayanlab (https://maayanlab.cloud/Enrichr/) to predict drug targets based on the four hub signaling pathways and functional analysis of genes. We selected the top ten target drugs for presentation (Table S2). With the development of modern technology, miRNAs have become important targets for the treatment of gene-related diseases. Maayanlab was also used to predict the miRNA targets. The findings indicated that hsa-miR-423-3p and hsa-miR-3132 controlled HCFC1 and SET, while hsa-miR-4715-5p, hsa-miR-645, hsa-miR-556-5p, hsa-miR-4317, and hsa-miR-4714-3p regulated RNF8 and HCFC1 (Figure S5A). Moreover, hsa-miR-496, hsa-miR-4694-5p, and hsa-miR-550b controlled the expression of RNF8 and SET, while hsa-miR-3147 regulated RNF8, TNP1, and HCFC1 (Figure S5A).Supplementary Table 3 shows the full names and related functions. By using the GeneMANIA database, we explored these genes' co-expression networks and associated functions.These genes showed a complex PPI network with 70.9% physical interactions, 16.01% co-expression, 4.96% predicted interactions 3.22% colocalization interactions, 2.63% genetic interactions, 0.55% shared protein domains, and 1.74% pathway interactions (Figure S.5B). The findings offer an in-depth examination of the potential diagnostic significance of CRFs in individuals with CAD, the impact of following medications, and predictive miRNA targets, thus presenting fresh perspectives for the diagnosis and management of CAD.

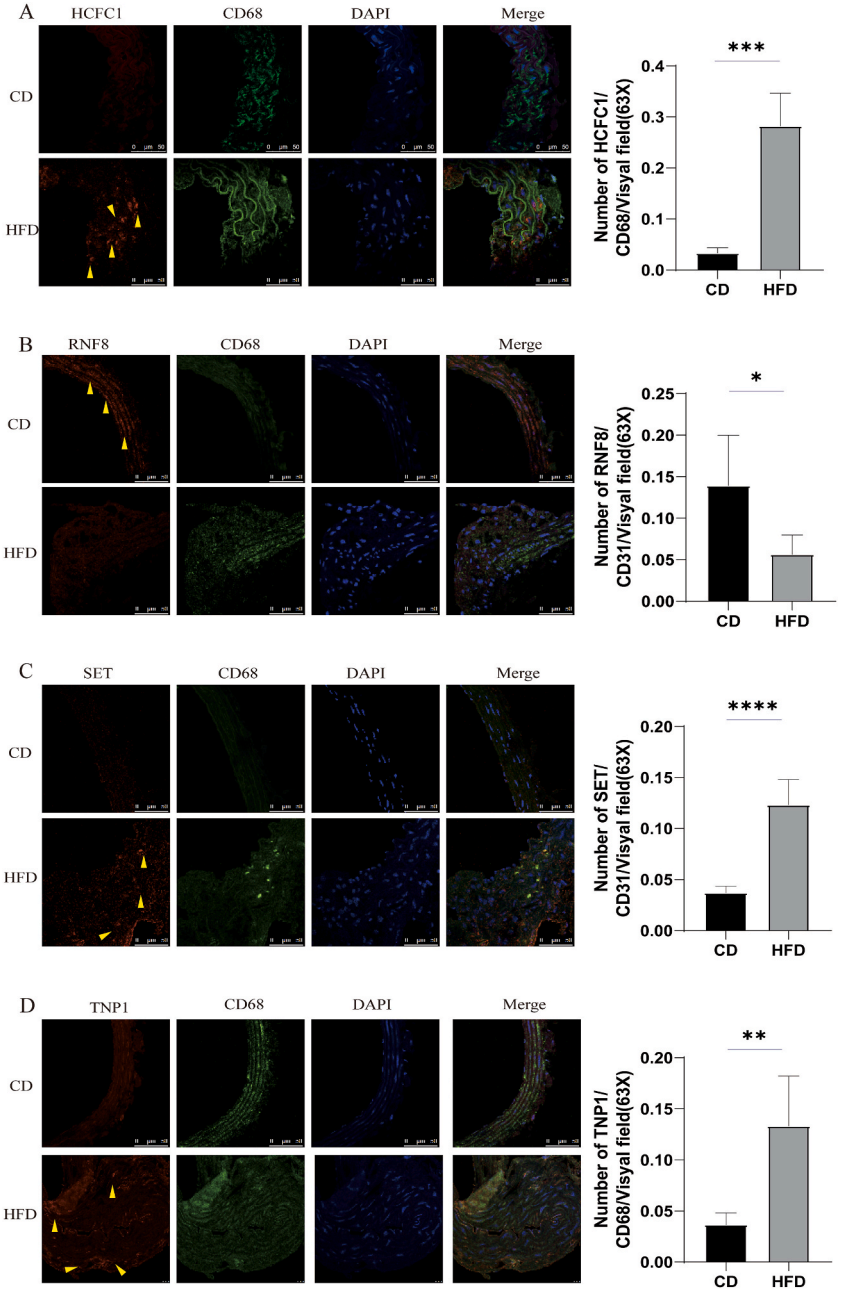


Fig. 8. Expression of four hub genes in the CD68-labeled macrophages. (A–D) The expression of (A) HCFC1, (B) RNF8, (C) SET, and (D) TNP1 in mouse arterial vessels was detected by immunofluorescence. Green represents macrophages, DAPI represents the nucleus, and red represents target proteins. Positive cells are marked with triangles. The magnification is 630. *$p < 0.05$ compared to the CD group. Data are shown as the mean ± SEM (n = 5). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

In this research, ssGSEA was utilized to examine the impact of CRFs on immune cell infiltration, revealing that the primary influence of these four genes was on the adaptive immune response of T cells and macrophages. Moreover, the effects of CRFs on atherosclerotic diseases are immune related. Furthermore, machine learning was performed to screen for molecules related to AS diagnosis, namely HCFC1, RNF8, TNP1, and SET. The marker gene set was utilized for enriching the marker pathways of the four central genes, revealing notable variances in metabolic pathways, particularly in bile acid, fatty acid, xenobiotic, IL2-STAT5, and heme metabolism. This further suggests that these molecules affect AS progression.

HCFC1, a member of the host cell factor family, contains a protein with five Kelch repeats, a fibronectin-like pattern, and six HCF

repeats, each containing a unique cleavage signal. During the development of the heart in mammals, HCFC1 is expressed and linked to congenital heart disease. The research conducted by Zhe et al. demonstrated that the suppression of HCF, the Drosophila counterpart of the mammalian HCFC1, in cardiac tissue results in aberrant cardiac morphology, tissue fibrosis, impaired cardiac function, and decreased H3K4 monomethylation. Intracellular cobalamin metabolism disorders are associated with a hemizygous variant of HCFC1. These conditions may impact the advancement of illness at different points, resulting in unusual heart structure, tissue scarring, and functional heart abnormalities in the prenatal period. Moreover, HCFC1 was discovered to have a receptor for SREBP-1, an important controller of genes related to cholesterol and fatty acid processing, aligning with our research showing a rise in HCFC1 levels in macrophages labeled with CD68 and smooth muscle cells labeled with α-SMA. This suggests that HCFC1 influences the development of CAD by influencing lipid metabolism.

TNP1 is essential in the substitution of histones with protamines during the elongation of mammalian spermatids [26]. Sperm cells in the condensing stage are placed onto nucleosomes, aiding in the attraction and modification of the protamines that cause histones to be removed. Current studies on this gene have focused primarily on its role in sperm production [27]. However, it is important to highlight the findings of Hasegawa et al., who demonstrated that anti-TNP1 autoantibodies are recognized and deposited in the glomerulus as immune complexes (IC) and that 14.7% of patients with systemic lupus erythematosus (SLE) exhibit a response to TNP1 [28]. These results suggest that TNP1 contributes to the development of SLE, which may have implications for immune prediction in patients with CAD. Although limited direct research on the association between TNP1 and CAD is available, this study indicates that TNP1 may affect immune pathways involving dendritic cells (DCs) or tumor-infiltrating lymphocytes (TILs).

The versatile protein SET is involved in multiple cellular functions, such as apoptosis, transcription, nucleosome assembly, and histone chaperoning [29]. Isoform 2 of SET functions to inhibit apoptosis by blocking the activity of GZMA-activated DNase NME1. During cytotoxic T lymphocyte-induced cell death, GZMA cuts SET, breaking its connection with NME1 and releasing the suppression of NME1 [29]. Both isoforms 1 and 2 function as strong blockers of phosphatase 2A.Isoforms 1 and 2 also hinder EP300/CREBBP- and PCAF-mediated histone acetyltransferases and nucleosomes, possibly by blocking the access of histone lysine residues to acetylases [30].

The primary focus of inhibition is on histone H4. Blocking histone acetyltransferase (HAT) activity leads to the inhibition of HAT-dependent gene expression and impedes the process of active DNA demethylation. Both isoforms promote DNA replication within the adenovirus genome when bound to viral core proteins, with isoform 2 exhibiting greater activity. Phosphorylation of the SET protein facilitates apoptosis through the hyperactivation of P53 and the nuclear import of NM23-H1 [31]. An in vivo experiment indicated that increased levels of SET might function through Akt/PTEN to either promote cell survival or serve as a sensor for oxidative stress-induced cell death [32]. Frequent SET overexpression is linked to negative results and leads to the inhibition of protein phosphatase 2A in acute myeloid leukemia [33]. These results are consistent with our experimental results showing increased SET expression in the three cell types in HFD-fed mice. Another research demonstrated that increased levels of SET lead to the development of human promonocytic cell line U937 into dendritic cells, showing characteristic morphological changes and expressing dendritic cell surface markers CD11b and CD86.U937 cells undergo dendritic cell-like differentiation through SET-induced calcium signaling and activation of the MAPK/ERK pathway [34]. The p300 and SET, an oncomodulator, both positively and negatively control the cardiovascular transcription factor KLF5 by interacting with its DNA-binding domain and acetylating it. Additionally, SET negatively impacts KLF5's abilities in DNA binding, transactivation, and cell proliferation. KLF5-mediated gene activation leads to the down-regulation of the negative regulator SET. On the other hand, the coactivator/acetylase p300 binds to and adds acetyl groups to the KLF5 DBD, leading to the activation of its transcription. Interestingly, SET inhibits the acetylation of KLF5, and non-acetylated KLF5 mutants show reduced transcriptional activation, with cell growth complementing the effects of SET [30].

RNF8 functions as a ubiquitin E3 ligase with two conserved regions: the N-terminal FHA domain, which has a strong attraction to phosphorylated peptide units in target proteins, and the C-terminal RING domain. This enzyme is primarily responsible for the catalytic activity of E3 ligases [35]. The original curiosity for research was sparked by its crucial role in DNA repair and the creation of sperm [35,36]. Additionally, RNF8 is involved in important functions during cell division, sperm production, safeguarding telomere ends, and programmed cell death, as well as participating in the cellular reaction to DNA damage [37]. EndMT cells are involved in atherosclerotic plaque formation [38]. The specific transcription factor Snail mediates this transformation. Snail interacts with the ubiquitin E3 ligase Ring1B, which recruits Ring1B and its parallel Ring1A to inhibit the target promoter via its carboxyl zinc finger [39]. Deletion of Ring1A and Ring1B leads to decreased Snail attachment to specific chromatin sites and mono-ubiquitination of histone 2A at K119.This hinders the transcription and movement of cells mediated by Snail [39]. Moreover, the initiation of EndMT through TGF-β is dependent on the activation of target genes and the interaction among different signaling pathways. Inhibition of Smads6/7 may lead to the activation of SMAD-specific E3 ubiquitin protein ligase 1 (SMURF1) to aid in the breakdown of activated TβR1 through ubiquitination, ultimately hindering the advancement of EndMT [40]. SMURF1 is able to suppress the activation of TGF-β1/Smad3/4-induced vascular endothelial growth factor, leading to a reduction in the angiogenesis linked to EndMT [41]. Recent research has shown that RNF8 is capable of triggering the activation of the GSK3β/β-catenin and PI3K/Akt signaling pathways, as well as interacting with slug to enhance K63 ligase ubiquitination and activate downstream twist in various tumor disease models, ultimately advancing the epithelial mesenchymal transition process [42,43]. Our research showed an increase in RNF8 levels specifically in the endothelial cell layer of the high-fat group, with no notable changes in the stromal cell layer. This discrepancy can be attributed to the intricate immune regulation and ubiquitination mechanisms associated with RNF8, thereby warranting further exploration of its role in coronary artery disease.

5. Conclusion

The advancement of contemporary science and technology in genomics has led to CRs playing a crucial part in diagnosing and treating CAD. The GEO database and basic experiments demonstrated that HCFC1, RNF8, TNP1, and SET may play roles in the formation of CAD. Nevertheless, the precise functions of these four genes must be confirmed through experiments involving a substantial amount of clinical samples.

Ethics approval

The research involving animals was authorized by the Institutional Animal Care and Ethics Committee at Ningbo University School of Medicine (10489) and conducted following the guidelines outlined in the NIH Guideline.In this study, summary-level data were de-identified and made publicly available.The ethical review committee approved the participation of each NCBI in their respective institutions.

Funding

Funding for this project was provided by a grant from the Zhejiang Medical and Health Science and Technology Project (2023J217) and the Ningbo Health and Technology Project (2023Y02).

Data availability statement

The published article contains all data produced or examined during this study. The data produced and examined in this research can be searched on the GEO database (GEO Accession viewer (nih.gov)/GSE20680/GSE66360/GSE100927). Inquiries about accessing the software should be sent to the authors responsible.

CRediT authorship contribution statement

Wang Bingyu: Writing – review & editing, Writing – original draft, Software, Data curation. **Yang Xi:** Writing – review & editing, Methodology, Funding acquisition. **Lian Jiangfang:** Supervision, Methodology, Investigation. **Zhou Jianqing:** Supervision, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28685.

References

[1] M. Ikeuchi, K. Sugimoto, A. Iwase, Plant callus: mechanisms of induction and repression, Plant Cell 25 (9) (2013) 3159–3173.
[2] C. Plass, et al., Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer, Nat. Rev. Genet. 14 (11) (2013) 765–780.
[3] A. Gonzalez-Perez, A. Jene-Sanz, N. Lopez-Bigas, The mutational landscape of chromatin regulatory factors across 4,623 tumor samples, Genome Biol. 14 (9) (2013) r106.
[4] Y.A. Medvedeva, et al., EpiFactors: a comprehensive database of human epigenetic factors and complexes, Database 2015 (2015) bav067.
[5] O. Ram, et al., Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells, Cell 147 (7) (2011) 1628–1639.
[6] Q. Wang, et al., CR Cistrome: a ChIP-Seq database for chromatin regulators and histone modification linkages in human and mouse, Nucleic Acids Res. 42 (Database issue) (2014) D450–D458.
[7] R. Zheng, et al., Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis, Nucleic Acids Res. 47 (D1) (2019) D729–d735.
[8] Y. Higashi, et al., Insulin-like growth factor-1 receptor deficiency in macrophages accelerates atherosclerosis and induces an unstable plaque phenotype in apolipoprotein E-deficient mice, Circulation 133 (23) (2016) 2263–2278.
[9] M.R. Alexander, G.K. Owens, Epigenetic control of smooth muscle cell differentiation and phenotypic switching in vascular development and disease, Annu. Rev. Physiol. 74 (2012) 13–40.
[10] M. Tsume-Kajioka, et al., BET proteins are essential for the specification and maintenance of the epiblast lineage in mouse preimplantation embryos, BMC Biol. 20 (1) (2022) 64.
[11] G. Qian, et al., Abdominal aortic aneurysm formation with a focus on vascular smooth muscle cells, Life 12 (2) (2022).

[12] M. von Scheidt, et al., Transcription factor MAFF (MAF basic leucine zipper transcription factor F) regulates an atherosclerosis relevant network connecting inflammation and cholesterol metabolism, Circulation 143 (18) (2021) 1809–1823.
[13] J. Lu, et al., FACER: comprehensive molecular and functional characterization of epigenetic chromatin regulators, Nucleic Acids Res. 46 (19) (2018) 10019–10033.
[14] R.C. Gentleman, et al., Bioconductor: open software development for computational biology and bioinformatics, Genome Biol. 5 (10) (2004) R80.
[15] A. Coppe, et al., Motif discovery in promoters of genes co-localized and co-expressed during myeloid cells differentiation, Nucleic Acids Res. 37 (2) (2009) 533–549.
[16] J. Yang, et al., Analysis of tumor suppressor genes based on gene ontology and the KEGG pathway, PLoS One 9 (9) (2014) e107202.
[17] S.J. Rigatti, Random forest, J. Insur. Med. 47 (1) (2017) 31–39.
[18] L. Tang, Y. Tian, C. Yang, Nonparallel support vector regression model and its SMO-type solver, Neural Netw 105 (2018) 431–446.
[19] C. Gold, P. Sollich, Model selection for support vector machine classification, Neurocomputing 55 (1–2) (2003) 221–249.
[20] R.P. Sheridan, et al., Correction to extreme gradient boosting as a method for quantitative structure-activity relationships, J. Chem. Inf. Model. 60 (3) (2020) 1910.
[21] D.A. Barbie, et al., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, Nature 462 (7269) (2009) 108–112.
[22] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesth. Analg. 126 (5) (2018) 1763–1768.
[23] Y. Zhu, et al., Evaluation of the immune factors in the tumor environment before and after the treatment of cetuximab combined with chemotherapy, World J. Surg. Oncol. 11 (2013) 226.
[24] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[25] K.C. Cotto, et al., DGIdb 3.0: a redesign and expansion of the drug-gene interaction database, Nucleic Acids Res. 46 (D1) (2018) D1068–d1073.
[26] H. Luerssen, et al., Nucleotide sequence of the gene for human transition protein 1 and its chromosomal localization on chromosome 2, Genomics 8 (2) (1990) 324–330.
[27] Y. Okada, et al., Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis, Nature 450 (7166) (2007) 119–123.
[28] S. Onishi, et al., Novel autoantigens associated with lupus nephritis, PLoS One 10 (6) (2015) e0126564.
[29] Y. Wu, et al., CBX8 together with SET facilitates ovarian carcinoma growth and metastasis by suppressing the transcription of SUSD2, Mol. Cancer Res. 20 (11) (2022) 1611–1622.
[30] S. Miyamoto, et al., Positive and negative regulation of the cardiovascular transcription factor KLF5 by p300 and the oncogenic regulator SET through interaction and acetylation on the DNA-binding domain, Mol. Cell Biol. 23 (23) (2003) 8528–8541.
[31] M. Wu, et al., Phosphorylation of SET mediates apoptosis via P53 hyperactivation and NM23-H1 nuclear import, Neurobiol. Aging 69 (2018) 38–47.
[32] A.M. Leopoldino, et al., Accumulation of the SET protein in HEK293T cells and mild oxidative stress: cell survival or death signaling, Mol. Cell. Biochem. 363 (1–2) (2012) 65–74.
[33] I. Cristóbal, et al., Overexpression of SET is a recurrent event associated with poor outcome and contributes to protein phosphatase 2A inhibition in acute myeloid leukemia, Haematologica 97 (4) (2012) 543–550.
[34] A. Kandilci, G.C. Grosveld, SET-induced calcium signaling and MAPK/ERK pathway activation mediate dendritic cell-like differentiation of U937 cells, Leukemia 19 (8) (2005) 1439–1445.
[35] F. Li, et al., Silencing of E3 ubiquitin ligase RNF8 enhances ionizing radiation sensitivity of medulloblastoma cells by promoting the deubiquitination of PCNA, Oncol. Res. 26 (9) (2018) 1365–1373.
[36] R. Rai, et al., The E3 ubiquitin ligase Rnf8 stabilizes Tpp1 to promote telomere end protection, Nat. Struct. Mol. Biol. 18 (12) (2011) 1400–1407.
[37] M.J. Halaby, et al., Synergistic interaction of Rnf8 and p53 in the protection against genomic instability and tumorigenesis, PLoS Genet. 9 (1) (2013) e1003259.
[38] Y. Takafuji, et al., Humoral factors secreted from adipose tissue-derived mesenchymal stem cells ameliorate atherosclerosis in Ldlr-/- mice, Cardiovasc. Res. 115 (6) (2019) 1041–1051.
[39] J. Chen, et al., Snail recruits Ring1B to mediate transcriptional repression and cell migration in pancreatic cancer cells, Cancer Res. 74 (16) (2014) 4353–4363.
[40] J. Liu, et al., To Ub or not to Ub: a regulatory question in TGF-β signaling, Trends Biochem. Sci. 47 (12) (2022) 1059–1072.
[41] L. Fu, et al., The functions and regulation of Smurfs in cancers, Semin. Cancer Biol. 67 (Pt 2) (2020) 102–116.
[42] H.J. Lee, et al., The DNA damage transducer RNF8 facilitates cancer chemoresistance and progression through twist activation, Mol Cell 63 (6) (2016) 1021–1033.
[43] B.L. Lee, et al., Molecular basis for K63-linked ubiquitination processes in double-strand DNA break repair: a focus on kinetics and dynamics, J. Mol. Biol. 429 (22) (2017) 3409–3429.