Research Paper

Identification and external validation of the hub genes associated with cardiorenal syndrome through time-series and network analyses

Jingjing Liang[1, 2], Xiaohui Huang[1], Weiwen Li[1, 2], Yunzhao Hu[1, 2]

[1]Department of Cardiology, Shunde Hospital of Southern Medical University, Foshan 528000, China
[2]The Second School of Clinical Medicine, Southern Medical University, Guangzhou 510515, China

Correspondence to: Yunzhao Hu; **email:** huyunzhao4406@163.com, https://orcid.org/0000-0003-2299-3560
Keywords: cardiorenal syndrome, FN1, POSTN, diagnosis, biomarker
Received: December 1, 2021 **Accepted:** January 12, 2022 **Published:** February 8, 2022

Copyright: © 2022 Liang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Cardiorenal syndrome (CRS), defined as acute or chronic damage to the heart or kidney triggering impairment of another organ, has a poor prognosis. However, the molecular mechanisms underlying CRS remain largely unknown. The RNA-sequencing data of the left ventricle tissue isolated from the sham-operated and CRS model rats at different time points were downloaded from the Gene Expression Omnibus (GEO) database. Genomic differences, protein–protein interaction networks, and short time-series analyses, revealed fibronectin 1 (FN1) and periostin (POSTN) as hub genes associated with CRS progression. The transcriptome sequencing data of humans obtained from the GEO revealed that FN1 and POSTN were both significantly associated with many different heart and kidney diseases. Peripheral blood samples from 20 control and 20 CRS patients were collected from the local hospital, and the gene expression levels of FN1 and POSTN were detected by real-time quantitative polymerase chain reaction. FN1 (area under the curve [AUC] = 0.807) and POSTN (AUC = 0.767) could distinguish CRS in the local cohort with high efficacy and were positively correlated with renal and heart damage markers, such as left ventricular ejection fraction. To improve the diagnostic ability, diagnosis models comprising FN1 and POSTN were constructed by logistic regression (F-Score = 0.718), classification tree (F-Score = 0.812), and random forest (F-Score = 1.000). Overall, the transcriptome data of CRS rat models were systematically analyzed, revealing that FN1 and POSTN were hub genes, which were validated in different public datasets and the local cohort.

INTRODUCTION

Chronic heart and kidney disease often occur together and promote each other, leading to progressive deterioration of heart and renal function. The phenomenon in which acute or chronic impairment of the heart or kidney causes dysfunction of another organ was first defined as cardiorenal syndrome (CRS) by Ronco et al. in 2008 [1]. Several evidence-based medicine studies have shown that renal insufficiency is a vital prognostic predictor and risk factor for heart diseases, while cardiovascular deaths account for the largest proportion of renal diseases [2–4]. Microalbuminuria, a hallmark of renal dysfunction, is widely accepted as a predictive factor of cardiovascular events [5–7]. Additionally, patients with chronic heart failure are prone to microalbuminuria and exhibit unfavorable clinical outcomes [8, 9]. In general, a bidirectional relationship between heart and renal diseases has been confirmed.

Several mechanisms have been proposed to explain CRS pathophysiology. From a hemodynamic standpoint, the reduction of cardiac output caused by heart failure directly leads to a decrease in renal blood flow, causing renal ischemia. The decrease in renal

Table 1. The detailed information of the public datasets used in this study.

ID	Platform	Organism	Tissue	Disease type	Control (*n*)	Disease (*n*)	References
GSE98520	GPL14844	Sprague Dawley rat	Myocardium	Cardiorenal syndrome	8	8	[17]
GSE2240	GPL97	Human	Myocardium	Atrial fibrillation	20	10	[53, 54]
GSE161472	GPL11154	Human	Myocardium	Heart failure	37	47	[55]
GSE36961	GPL15389	Human	Myocardium	Hypertrophic cardiomyopathy	39	106	Unknown
GSE66494	GPL6480	Human	Kidney	Chronic kidney disease	8	53	[56]
GSE125779	GPL17586	Human	Kidney	Focal segmental glomerulosclerosis	8	8	[57]
GSE37171	GPL570	Human	Peripheral blood	Uremia	40	75	[58]

blood flow could also activate the renin-angiotensin-aldosterone system (RAAS), thereby impairing the heart and renal functions [10]. From a physiological perspective, inflammation and oxidative stress play essential roles in CRS progression. The overactivation of RAAS promotes the release of inflammatory factors, such as interleukin 6, tumor necrosis factor α, and transforming growth factor (TGF), causing kidney fibrosis and ventricular remodeling [11, 12]. The upregulation of RAAS could further advance the production of reactive oxygen species (ROS), and excessive ROS leads to necrosis of renal and cardiac cells [13]. In addition to these classical hypotheses, other factors such as activation of the sympathetic nervous system, accumulation of uremic toxins, and endoplasmic reticulum stress are known to impact CRS [14, 15]. However, our current understanding of the initiation and development of CRS remains insufficient.

The rapid development of gene sequencing technology and big-data analysis has allowed further clarification of the latent molecular mechanisms underlying CRS. Chen et al. performed transcriptome sequencing of the right ventricle and kidney isolated from CRS mouse models and established lncRNA-miRNA-mRNA competing endogenous RNA networks, thus clarifying the comprehensive regulatory relationships of CRS [16]. Chuppa et al. utilized RNA-sequencing (RNA-seq) to analyze the left ventricle tissue of CRS rat models. They found that miR-21-5p could improve cardiac function by regulating peroxisome proliferator-activated receptor alpha [17]. These efforts broaden our horizons and highlight potential therapeutic targets for CRS. Nevertheless, the number of genome-wide studies on CRS remains limited.

Herein, the transcriptome data of the rat ventricle tissue at weeks 2, 4, 5, and 7 after subtotal nephrectomy were retrieved from the Gene Expression Omnibus (GEO) database. Genomic divergence, protein–protein interaction (PPI) network, and time-series analyses were conducted to identify the hub genes involved in CRS progression. Transcriptome sequencing of different types of heart and renal diseases downloaded from the GEO database was used for preliminary verification. The peripheral blood of the control and CRS subjects was also collected from the Shunde Hospital of Southern Medical University, and real-time quantitative polymerase chain reaction (RT-qPCR) was performed to detect gene expression levels. Finally, multiple machine learning algorithms were used to improve the diagnostic ability based on these hub genes.

MATERIALS AND METHODS

Data collection

The GSE98520 dataset, including the RNA-seq data with fragments per kilobase million format (FPKM) of the left ventricle tissue isolated from the sham-operated and treated rats at weeks 2, 4, 5, and 7 after the 5/6 nephrectomy, was directly obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) as the training dataset. The GEO datasets GSE2240, GSE161472, GSE36961, GSE66494, GSE125779, and GSE36961, containing different heart disease or kidney disease samples, were also downloaded. Detailed information on the public datasets utilized in this study is presented in Table 1. According to the platform annotation files collected from the GEO database, all probe IDs were transformed into gene symbols using R software (version 3.6.3). If multiple probes corresponded to a gene, an average value was adopted. Probes corresponding to multiple genes were excluded from analyses.

Genomic difference analysis and PPI network construction

Genomic differences were analyzed to detect the differentially expressed genes between the sham-operated and subtotal nephrectomy rats at weeks 2, 4, 5, and 7 using the limma package. The filtering threshold to identify associated genes was set at $P < 0.05$. Genes

Table 2. The primer sequence used in this study.

Gene	Primer sequence (5′–3′)
FN1	Forward: CGGTGGCTGTCAGTCAAAG
	Reverse: AAACCTCGGCTTCCTCCATAA
POSTN	Forward: CTCATAGTCGTATCAGGGGTCG
	Reverse: ACACAGTCGTTTTCTGTCCAC
GAPDH	Forward: GGAGCGAGATCCCTCCAAAAT
	Reverse: GGCTGTTGTCATACTTCTCATGG

exhibiting significant expression differences at all four time points were selected and included in the PPI network analysis [18]. The PPI network was based on the STRING database (version 11.5, https://cn.string-db.org/), and the confidence level was set to 0.4 [19]. The CytoHubba plug-in (version 0.1) in the Cytoscape software (version 3.8.0) was used to measure the importance of the genes in the network, and genes with a degree ≥10 were considered as hub genes [20].

Time-series analysis

Time-series analysis was performed using the Short Time-series Expression Miner (STEM, version 1.3.13). The data were normalized to the expression values of the sham-operated samples. The STEM clustering method was utilized to conduct the clustering with the following parameters: the maximum number of model profiles = 50 and maximum unit change in model profiles between time points = 2 [21]. The advance options were all set to the default values.

Functional enrichment analysis

Gene Ontology (GO) functional annotation was conducted using the clusterProfiler package after transforming the gene symbols into Entrez IDs according to the org.Rn.eg.db or org.Hs.eg.db package. Terms with $P < 0.05$ and $Q < 0.05$ were considered statistically significant.

Clinical samples

The study protocol was reviewed and approved by the Ethics Committee of the Shunde Hospital of Southern Medical University (Ethics Approval Number: 20210207). All participants signed an informed consent form. Peripheral blood samples (2 mL) of 20 control and 20 CRS subjects were collected within 24 h of admission and stored in EDTA anticoagulant tubes at 4°C. CRS diagnosis was based on the latest clinical guidelines [22, 23]. The control subjects were defined as those without CRS, severe cardio or renal dysfunction, malignant tumors, severe infection, or other factors which could possibly influence the gene expression level. The clinicopathological features of patients, including age, sex, smoking, diabetes history, left ventricular ejection fraction (LVEF), N-terminal pro-B-type natriuretic peptide (NTproBNP), serum creatinine (Scr), blood urea nitrogen (BUN), and uric acid (UA), were also recorded.

RT-qPCR

Total RNA was extracted using the Trizol-chloroform method (Trizol reagent, Sigma-Aldrich, China) after the red blood cells of the whole blood sample were lysed with erythrocyte lysis buffer (Sigma-Aldrich, Saint Louis, MO, USA) for 15 min at room temperature. The purity and concentration of the RNA were measured using a Nanodrop2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). Complementary DNA was synthesized and amplified using the PrimeScript RT Reagent Kit (Takara, Dalian, China) and SYBR Premix ExTaq kit (Takara, China). The PCR experiments were conducted on an Applied Biosystems 7600 thermocycler (ABI, USA). Gene expression levels were normalized to GAPDH, and the $2^{-\Delta\Delta Ct}$ method was used to calculate the definite RNA expression values. The primer sequences used in this study are listed in Table 2.

Construction of diagnostic models based on machine learning algorithm

To improve the diagnostic ability of CRS, we constructed diagnostic models based on the screened core genes. Logistic regression was performed to establish the diagnostic model, and a nomogram was drawn to visualize the model using the rms package [24]. The classification tree was constructed using the rpart package [25], and the random forest model was developed using the randomForest package with the following parameters: 500 as the ntree and 3 as the mtry [26]. The confusion matrices, accuracy, precision, recall, and F-Score were used to measure the predictive ability of each machine learning diagnosis model.

Identification of the functionally-related genes

The top 20 functionally related genes of the hub genes were identified using the GeneMANIA database

(http://genemania.org/search/homo-sapiens/). The number of the maximum resultant genes was set to 20, and that of maximum resultant attributes was set to 10. The query-dependent weighting method was chosen automatically by the website [27].

Statistical analysis

Statistical analysis were performed using R (version 3.6.3) and GraphPad Prism (version 8.4.3). Data are presented as mean ± standard deviation (SD). Student's *t*-test was used to compare gene expression differences in the PCR experiments. The Welch-corrected *t*-test was adopted to compare the differences in gene expression levels between the control and disease groups obtained from the GEO database and clinicopathological parameters of the control and CRS subjects from the Shunde Hospital of Southern Medical University. The association between the hub gene expression level and clinicopathological variables was measured using the Spearman correlation test. The receiver operating curve (ROC) and corresponding area under curve (AUC) were obtained from the pROC package. Unless otherwise specified, $P < 0.05$ was considered to be statistically significant; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

RESULTS

Genomic difference analysis and PPI network construction

The workflow of the study protocol is shown in Figure 1. The R code used in this study is presented in Supplementary Material. The differentially expressed genes between the sham-operated and CRS model rats were analyzed. A total of 286 (Figure 2A), 593 (Figure 2B), 463 (Figure 2C), and 1182 (Figure 2D) genes showing expression differences were screened in 2-, 4-,

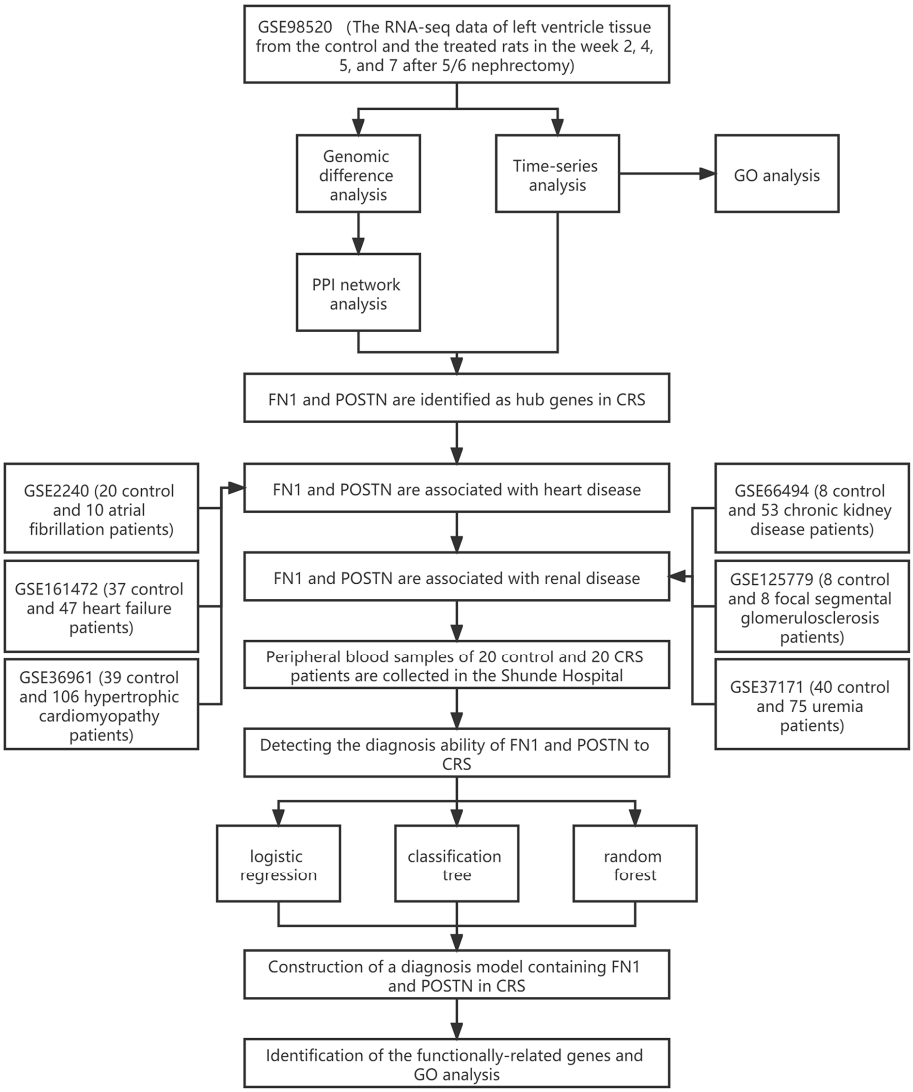


Figure 1. The workflow of the present study.

5-, and 7-week old rats, respectively, and 35 genes were found to overlap (Figure 2E). Subsequently, 35 genes were included in the PPI network (Figure 2F). The CytoHubba app revealed that the degrees of Col1a1, FN1, POSTN, and Col3a1 were greater than or equal to 10, and thus, the four genes were selected for subsequent analysis (Figure 2G, Supplementary Table 1).

Time-series analysis and functional annotation

The time-series analysis of the transcriptome sequencing data of the CRS model rats at 4 time nodes identified 11 different gene clusters ($P < 0.05$) (Figure 3A). A total of 1346 genes were included in these clusters, of which fibronectin 1 (FN1) and periostin (POSTN) were determined by PPI network analysis (Figure 3B). Interestingly, FN1 and POSTN were both members of cluster 41 ($P < 0.001$, Figure 3C). GO enrichment analysis indicated that the genes in cluster 41 were mainly involved in cell cycle- and immune-related pathways, such as negative regulation of metaphase/anaphase transition of the cell cycle, spindle checkpoint, and mast cell granules (Figure 3D). The synthesis of pro-inflammatory mediators always

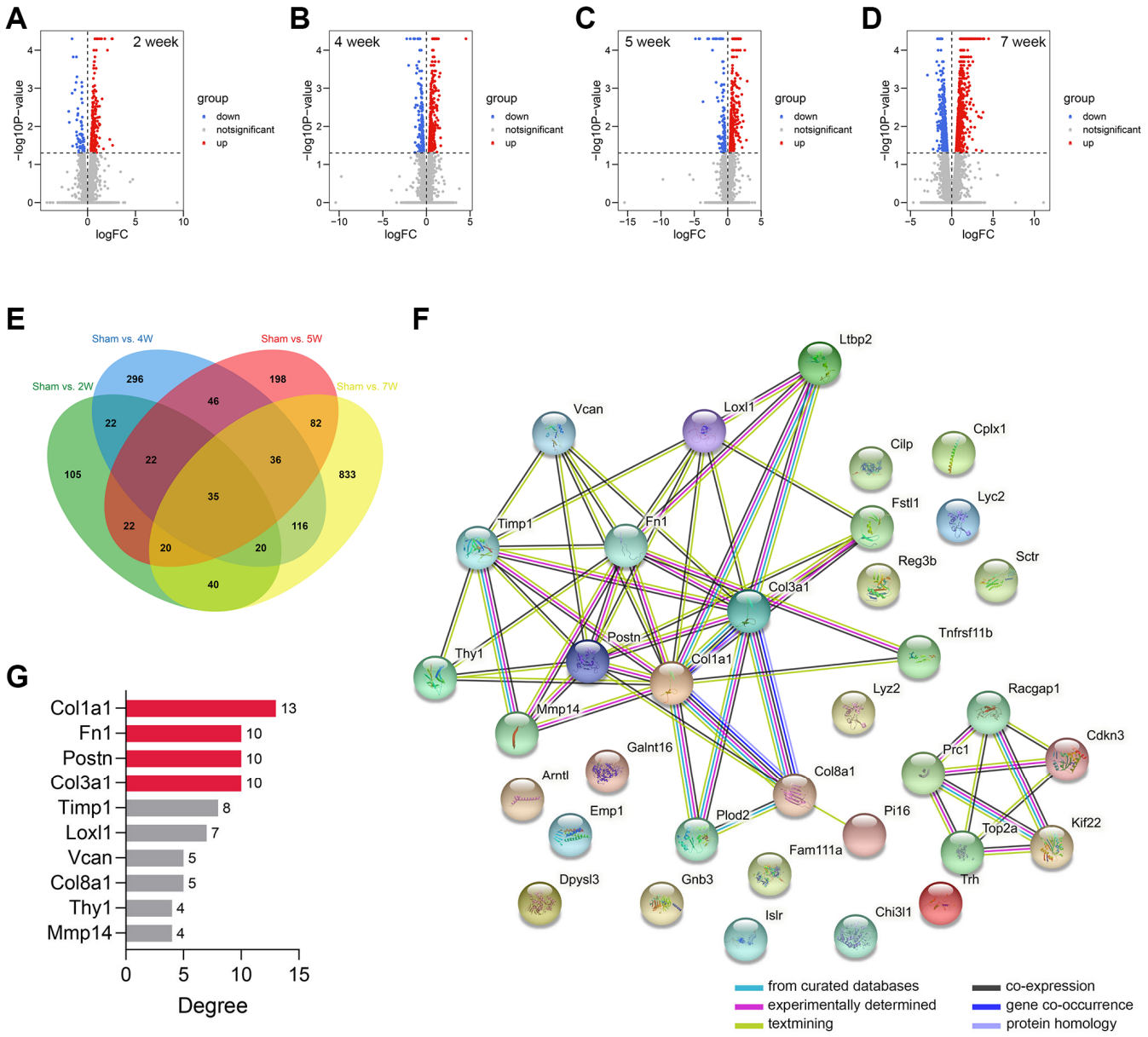


Figure 2. Genomic difference analyses and PPI network construction. (**A**–**D**) The volcano plots showing the differentially-expressed genes between the sham-operated and CRS model rats in week 2 (**A**), week 4 (**B**), week 5 (**C**), and week 7 (**D**). (**E**) A sum of 35 differentially expressed genes were overlapped. (**F**) The PPI network of the 35 genes. (**G**) The Top 10 genes with the highest degree value in the network. Abbreviation: PPI: protein-protein interaction network.

increases in CRS, causing cell death and fibrosis [28]. In addition, cell cycle arrest plays essential roles in pathological processes as the renal and cardiac cells usually undergo cell cycle arrest to prevent possible DNA damage from cell division when the cells undergo cellular stress [29]. Generally, these findings correspond to those of previous studies.

FN1 and POSTN are associated with many different heart and renal diseases

To further verify the role of FN1 and POSTN in CRS, transcriptome data of different heart and kidney diseases were downloaded. We found that FN1 was significantly upregulated in the atrial fibrillation ($P < 0.05$, Figure 4A), heart failure ($P < 0.01$, Figure 4B), hypertrophic cardiomyopathy ($P < 0.001$, Figure 4C), chronic kidney disease ($P < 0.05$, Figure 4D), focal segmental glomerulosclerosis ($P < 0.01$, Figure 4E), and uremia ($P < 0.001$, Figure 4F) samples compared with control samples. Similar trends were also observed in POSTN, as shown in Figure 4A–4F ($P < 0.05$). The indirect evidence partly demonstrated that FN1 and POSTN are strongly associated with many different heart and kidney diseases, thereby influencing the pathogenesis of CRS.

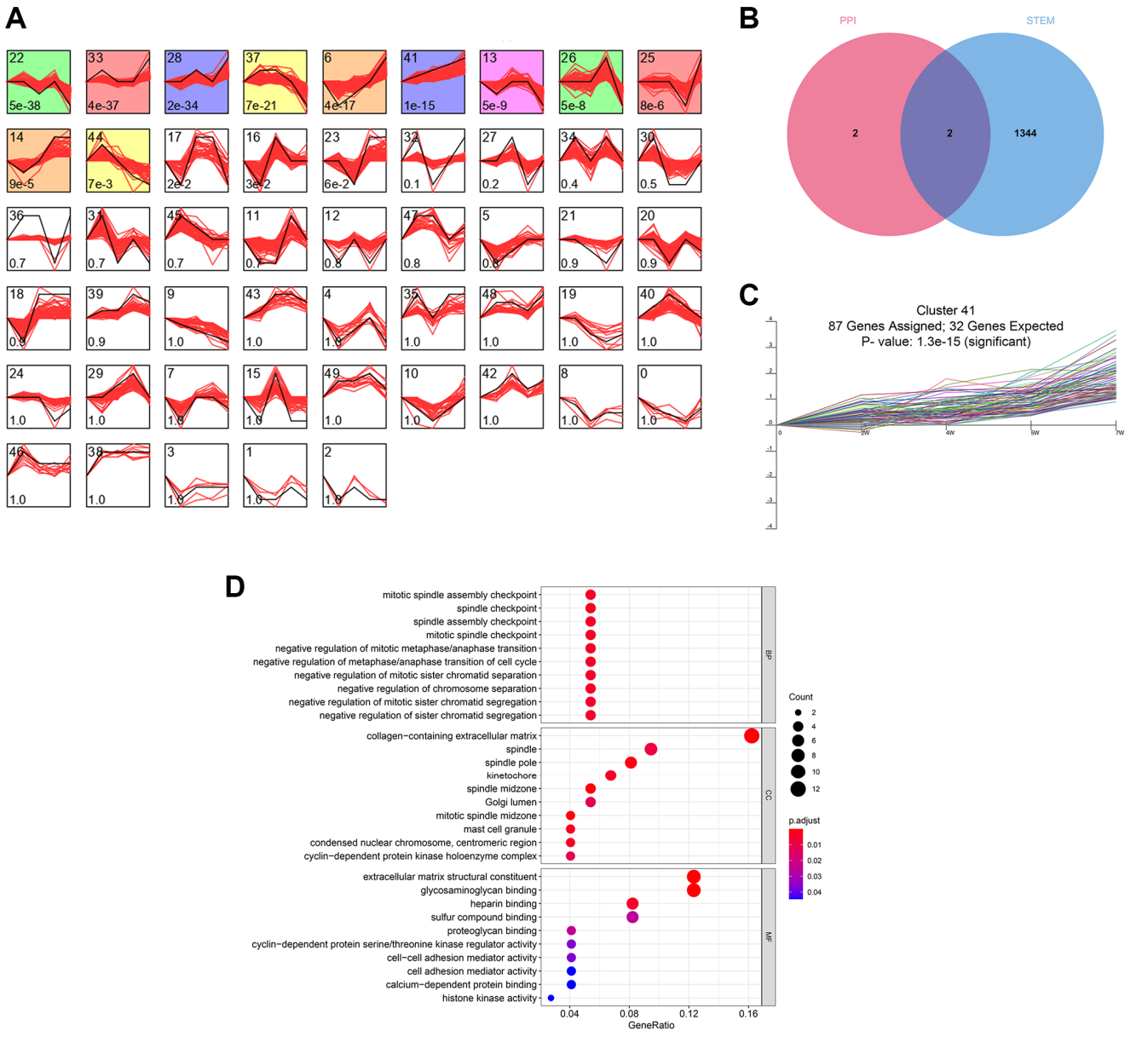


Figure 3. The time-series analyses. (**A**) 11 gene clusters were identified. The number at bottom-left in each cell represent the *P*-value and the number at top-left is the gene cluster ID. (**B**) FN1 and POSTN were con-determined by the network analysis and the time-series analysis. (**C**) FN1 and POSTN were encompassed in cluster 41, of which the *P*-value <0.001. (**D**) The GO functional annotation of the gene cluster 41. Abbreviation: GO: gene ontology.

Table 3. The baseline information of the 20 control and 20 CRS patients.

Parameters	Control (n = 20)	CRS (n = 20)	P-value	Significance
Male (n, %)	13 (65.0%)	15 (75.0%)	0.302	ns
Age (years)	59.1 ± 16.8	55.1 ± 16.8	0.450	ns
Smoking (n, %)	6 (30.0%)	11 (55.0%)	0.015	*
Diabetes history (n, %)	4 (20.0%)	8 (40.0%)	0.068	ns
LVEF (%)	64.2 ± 7.7	58.3 ± 12.2	0.078	ns
NTproBNP (pg/ml)	1474.6 ± 2457.9	1580.7 ± 2317.9	0.889	ns
Scr (μmol/L)	52.1 ± 20.3	168.4 ± 227.7	0.029	*
BUN (mmol/L)	4.8 ± 1.2	9.6 ± 6.9	0.004	**
UA (μmol/L)	309.4 ± 81.6	419.7 ± 143.6	0.005	**

Abbreviations: CRS: cardiorenal syndrome; LVEF: left ventricular ejection fraction; NTproBNP: N-terminal-pro-B-type natriuretic peptide; Scr: serum creatinine; BUN: blood urea nitrogen; UA: uric acid; ns: not significance. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

FN1 and POSTN are promising diagnostic biomarkers of CRS

RT-qPCR was used to measure the expression levels of FN1 and POSTN in the peripheral blood of control and CRS patients. The original CT values of FN1, POSTN, and GAPDH are listed in Supplementary Table 2. The baseline clinicopathological information of the subjects enrolled in this study is displayed in Table 3. Except for smoking history (P < 0.05), Scr (P < 0.05), BUN

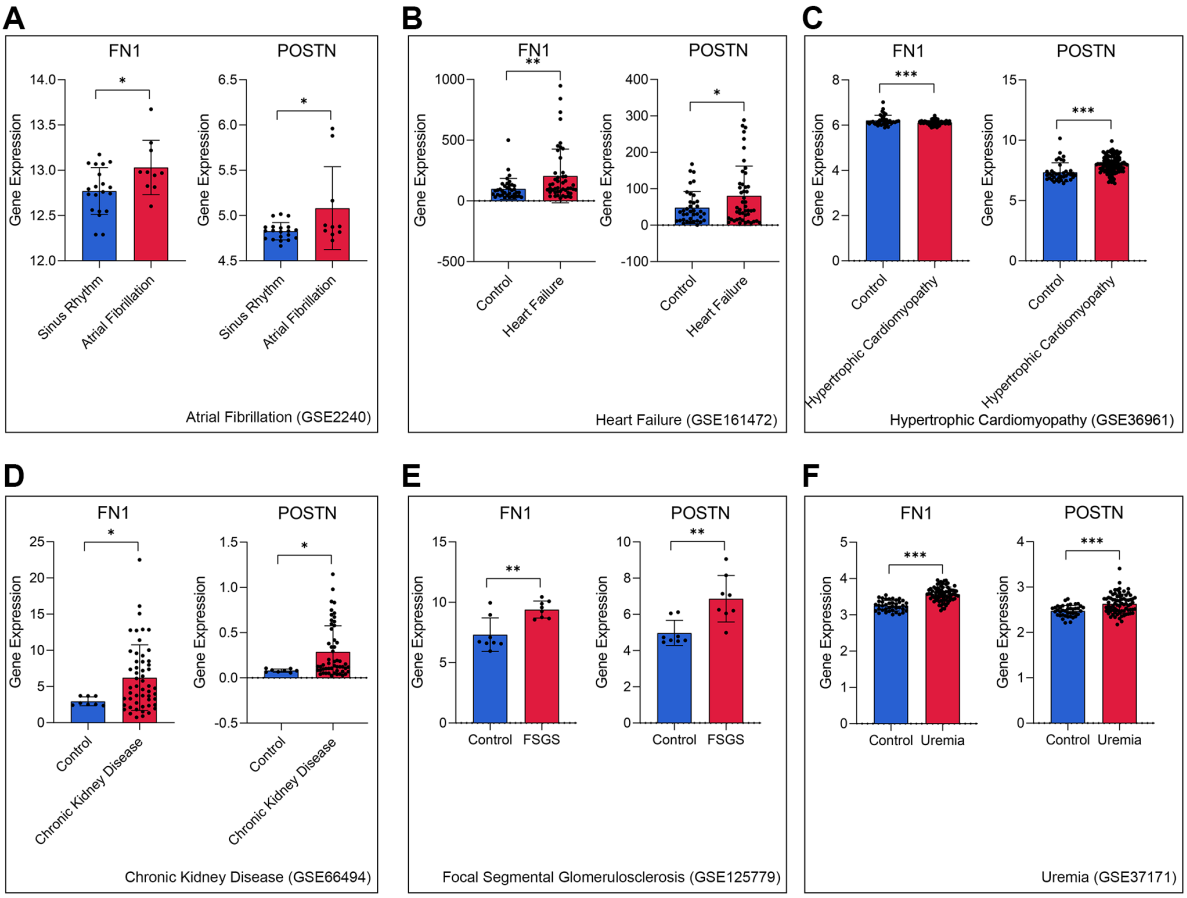


Figure 4. FN1 and POSTN were up-regulated in multiple heart disease and renal diseases samples, including atrial fibrillation (**A**), heart failure (**B**), hypertrophic cardiomyopathy (**C**), chronic kidney disease (**D**), focal segmental glomerulosclerosis (**E**), and uremia (**F**), compared with the control samples. Abbreviations: FSGS: focal segmental glomerulosclerosis. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

($P < 0.01$), and UA ($P < 0.01$), other parameters, including age, sex, diabetes history, LVEF, and NT-proBNP, revealed no significant differences between the control and CRS groups. Compared with the control subjects, the CRS cases exhibited higher expression levels of FN1 ($P < 0.01$, Figure 5A) and POSTN ($P < 0.05$, Figure 5B). ROC analysis indicated that FN1 (AUC = 0.807, Figure 5C) and POSTN (AUC = 0.767, Figure 5C) were both diagnostic biomarkers with high efficacy. The association of the expression values of FN1 and POSTN with routine laboratory tests was also detected. FN1 was significantly associated with LVEF (R = 0.54, $P < 0.05$, Figure 5D), NT-proBNP (R = 0.49, $P < 0.05$, Figure 5E), and Scr (R = 0.57, $P < 0.01$, Figure 5F), but no significant association was observed with BUN (R = 0.43, $P = 0.057$, Figure 5G) or UA (R = 0.38, $P = 0.1$, Figure 5H). POSTN was further significantly correlated with LVEF (R = 0.54, $P < 0.05$, Figure 5I) and BUN (R = 0.41, $P < 0.05$, Figure 5J), while no significant association was observed between POSTN and Scr (R = 0.33, $P = 0.16$, Figure 5K), NT-proBNP (R = 0.43, $P = 0.063$, Figure 5L), and UA (R = 0.35, $P = 0.13$, Figure 5M). The association of CRS with LVEF [30], NTproBNP [31], Scr [32], and BUN [33] has been verified in multiple studies, which strengthens the reliability of FN1 and POSTN.

Construction of the diagnostic models based on FN1 and POSTN

Diagnostic models containing FN1 and POSTN were established using multiple machine learning algorithms. The logistic regression model was constructed as follows: logistic score = −2.10 + 1.58*EXP(FN1) +

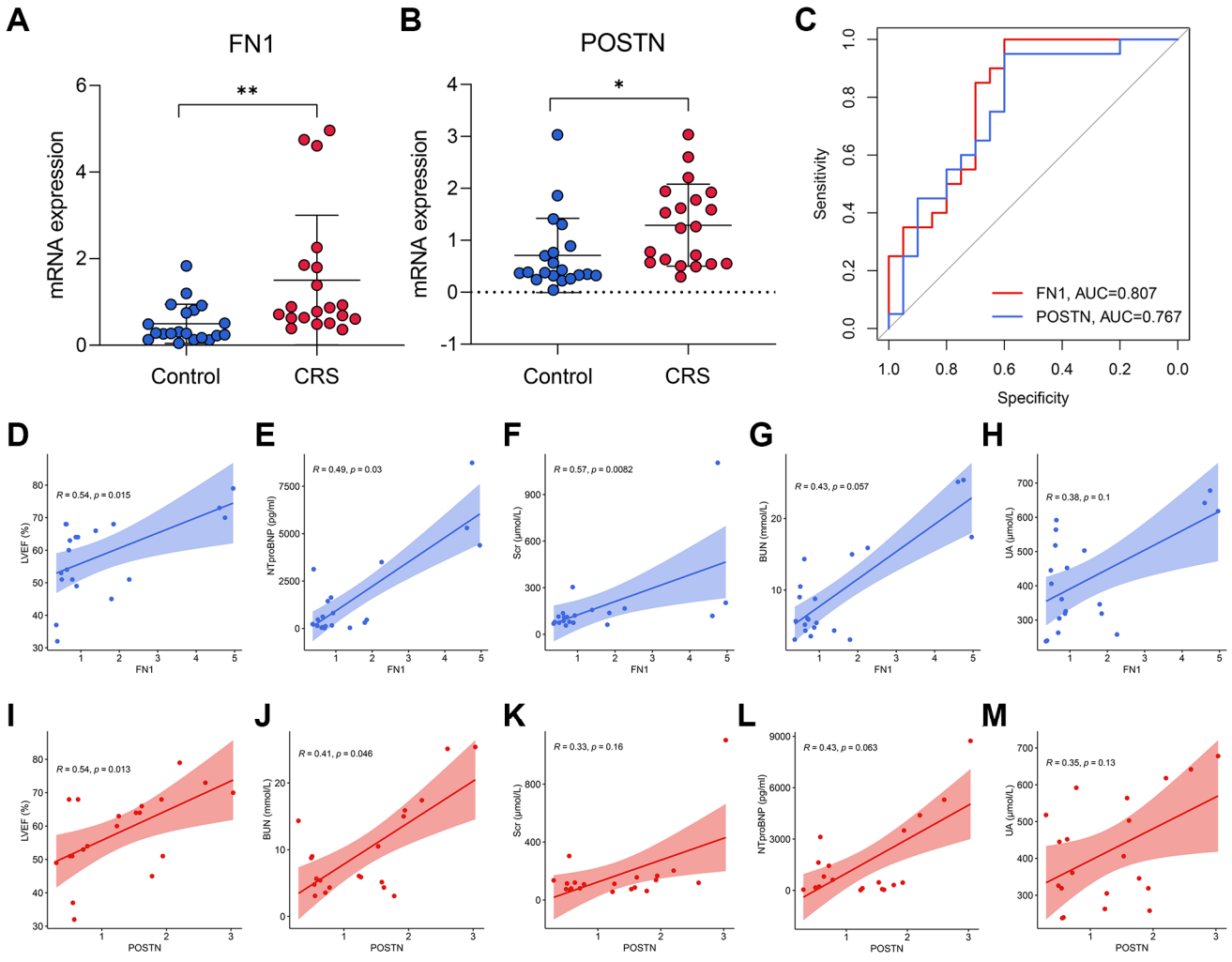


Figure 5. The diagnostic ability of FN1 and POSTN to CRS. (**A**, **B**) FN1 (**A**) and POSTN (**B**) were up-regulated in the peripheral blood sample of CRS patients from the Shunde Hospital of Southern Medical University. (**C**) The ROC analysis showed that FN1 and POSTN could discriminate the CRS samples with high efficacy. (**D**–**H**) The association of FN1 expression with LVEF (**D**), NTproBNP (**E**), Scr (**F**), BUN (**G**), and UA (**H**). (**I**–**M**) The association of POSTN expression with LVEF (**I**), BUN (**J**), Scr (**K**), NTproBNP (**L**), and UA (**M**). Abbreviations: CRS: cardiorenal syndrome; ROC: receiver operating curve; LVEF: left ventricular ejection fraction; NTproBNP: N-terminal-pro-B-type natriuretic peptide; Scr: serum creatinine; BUN: blood urea nitrogen; UA: uric acid. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

0.88*EXP(POSTN), where EXP represented the mRNA expression level of FN1 or POSTN. A nomogram including FN1 and POSTN was constructed to help clinicians better understand the logistic model (Figure 6A). The confusing matrix of indicated that the logistic regression model could predict CRS with high efficacy (Figure 6B). The classification tree is shown in Figure 6C. Compared with the logistic regression model, the classification tree showed a higher predictive ability (Figure 6D). Unexpectedly, it was found that only FN1

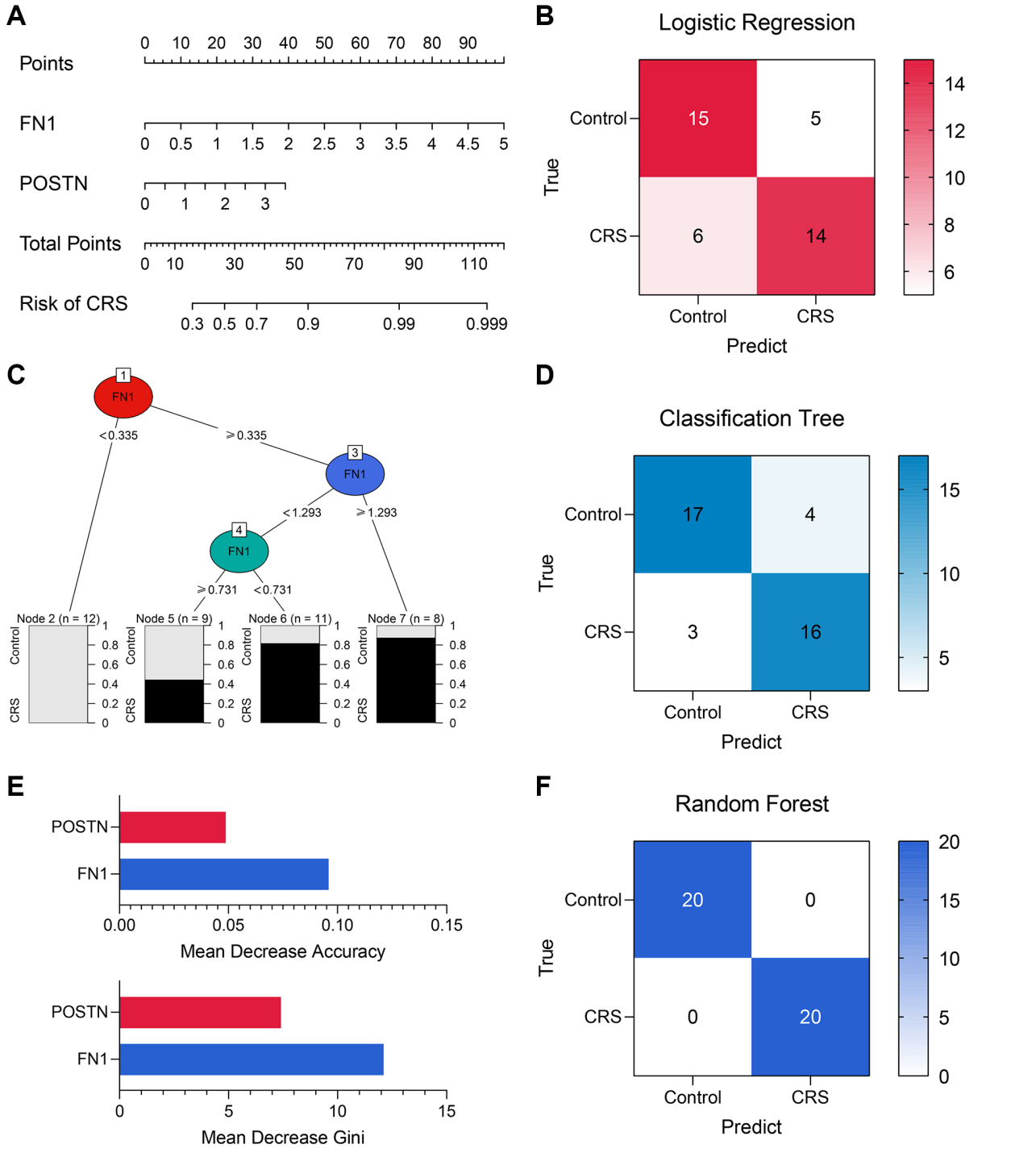


Figure 6. The machine learning models encompassing FN1 and POSTN to diagnose CRS. (**A**) A nomogram was drawn to visualize the logistic regression model. (**B**) The confusion matrix showed the predictive performance of the logistic regression model. (**C**) The classification tree was established to diagnose CRS. (**D**) The confusion matrix of the classification tree model. (**E**) The mean decrease accuracy and mean decrease Gini of the features in the random forest model. (**F**) The confusion matrix exhibited that the random forest model could distinguish the CRS samples with high efficacy. Abbreviation: CRS: cardiorenal syndrome.

Table 4. The performance of the logistic regression model, the classification tree model, and the random forest model.

Model	Accuracy	Precision	Recall	F-Score
Logistic regression model	0.725	0.737	0.700	0.718
Classification tree model	0.825	0.800	0.842	0.812
Random forest model	1.000	1.000	1.000	1.000

was included in the classification tree model, implying that FN1 had better predictive ability than POSTN. In the random forest model, the importance of FN1 and POSTN was quantified as a mean decrease in accuracy and Gini, which were previously described [34]. Compared with POSTN, FN1 exhibited higher mean decreases in accuracy and Gini (Figure 6E), which was in agreement with the classification tree analysis. The random forest model could distinguish the CRS sample with the highest effectiveness (Figure 6F, Supplementary Figure 1A and 1B), which is in alignment with the results of several previous studies that have provided proof of its strong performance [35–37]. The accuracy, precision, recall, and F-score of each model are listed in Table 4. In addition, the predictive ability of CRS for clinicopathological features and established diagnostic models were also compared (Supplementary Figure 1A and 1B). Overall, compared with the traditional biomarkers or single biomarkers like FN1 and POSTN, the combination of FN1 and POSTN through machine learning algorithms, especially random forest, greatly improved the efficiency of CRS diagnosis.

Functionally-related genes of FN1 and POSTN

The top 20 most strongly associated genes of FN1 and their interaction relationships are illustrated in Figure 7A. The size of the gene nodes represents the importance of the genes in the network, and the thickness of the lines is positively correlated with the interaction strength. GO analyses revealed that FN1 and its associated genes mostly participate in extracellular matrix-, immune-, and platelet-related biological processes (Figure 7B). The top 20 POSTN-related genes are shown in Figure 7C, and their functions were mainly associated with the extracellular matrix, amino acids, and TGF (Figure 7D). These findings indicate the possible mechanisms by which FN1 and POSTN are involved in the bidirectional interaction between the heart and kidney.

DISCUSSION

CRS is a complex heart–kidney disease characterized by high morbidity and mortality, resulting in a tremendous social burden worldwide [38]. About 20–40% of patients with acute heart failure suffer from kidney dysfunction, and almost 40–60% of patients with chronic heart failure experience chronic kidney disease [39, 40]. Exploration of the molecular mechanisms in CRS is critical and difficult. The advancement and popularization of gene sequencing technology has provided an opportunity to further elucidate the pathological processes of this disease. Recently, many novel biomarkers associated with CRS have been reported, including cystatin 3, galectin 3, NGAL, and KIM1 [31, 41]. These biomarkers not only provide the clinical guidelines for CRS diagnosis and prognosis, but also suggest the underlying cut-in points for mechanistic studies. However, given the complexity of CRS, the current findings are far from sufficient.

The present study systematically analyzed the transcriptomic data of the rat ventricle tissue at weeks 2, 4, 5, and 7 after subtotal nephrectomy through genomic difference detection, PPI network analysis, and time-series analysis; FN1 and POSTN were ultimately identified as hub genes associated with CRS. The validation in different public datasets and local clinical samples indicated that FN1 and POSTN were both significant diagnostic biomarkers for CRS, which verified the findings from the animal experiments. Here, we report that levels of FN1 and POSTN were obviously increased not only in the diseased heart and kidney, but also in the plasma of CRS patients, implying that FN1 and POSTN are underlying cardiorenal connectors. FN1, encoding fibronectin, a protein located in the plasma, cell surface, and extracellular matrix, is mainly involved in cell adhesion and migration [42]. POSTN is also located in the extracellular space and regulates tissue development and regeneration. The vital roles of FN1 and POSTN in heart and kidney disease have previously been reported. Wang et al. found that FN1 was associated with immune infiltration in diabetic nephropathy [43], Su et al. reported that FN1 could regulate the process of renal fibrosis [44], Zhao et al. disclosed that FN1 was involved in rat H9C2 cardiomyocyte growth by regulating cell cycle arrest [45], and Patel et al. discovered that the expression of FN1 increased in patients with sudden cardio death [46]. Cardiofibrosis has a significant influence on the prognosis of heart diseases, which can be regulated by POSTN [47, 48].

POSTN was also found to be associated with renal diseases, including diabetic kidney disease [49], immunoglobulin A nephropathy [50], and polycystic kidney disease [51]. Hence, regardless of the absence of a regulatory relationship of POSTN and FN1 with CRS, FN1 and POSTN may exert important biological functions in the pathogenesis of CRS. Time-series analysis indicated that FN1 and POSTN were both members of cluster 41, and were thus mainly associated with immune-related pathways, which corresponded to the enrichment results of their functionally related genes. Immune-mediated dysregulation is well known for its vital role in CRS development, directly evidenced by the strong upregulation of plasma proinflammatory cytokines in CRS patients [52]. In summary, FN1 and POSTN are latent cardiorenal connectors that may function by regulating the immune response in CRS.

Another highlight of this study is the establishment of CRS diagnostic models comprising FN1 and POSTN. To ameliorate the predictive ability, the diagnostic models were constructed using three different machine learning algorithms, namely logistic regression, classification tree, and random forest. The ROCs revealed that the random forest model could distinguish CRS samples from control samples with high efficacy. Compared with traditional biomarkers, the performance

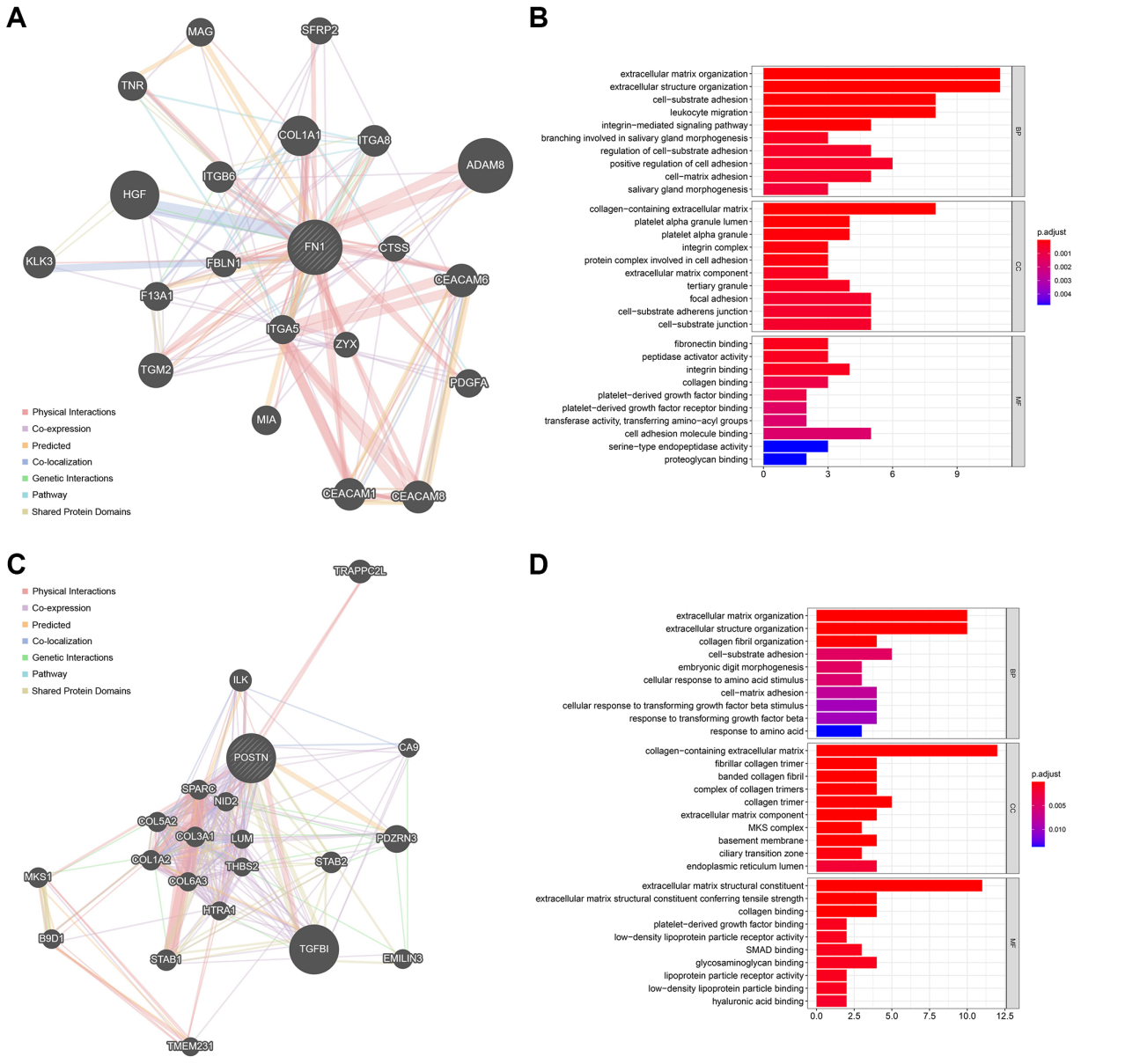


Figure 7. The functionally-related genes of FN1 and POSTN. (**A**) The Top 20 genes showing the highest association with FN1. (**B**) GO functional analysis of the FN1-related genes. (**C**) The Top 20 genes most associated with POSTN. (**D**) GO enrichment analysis of the POSTN-related genes. Abbreviation: GO: gene ontology.

of the random forest model is more satisfactory. In general, novel genetic diagnostic tools based on FN1 and POSTN are presented, providing new choices for clinicians.

The limitations of this study should not be neglected. First, we only performed the direct validation of the diagnostic value of FN1 and POSTN in the local hospital, and external verification in other centers would be beneficial to further clarify the clinical usefulness of these biomarkers. Second, we revealed that FN1 and POSTN serve as novel biomarkers of CRS in animal experiments and different cohorts, but how they affect the progression of CRS remains unclear. Further experiments are needed to explore the processes by which FN1 and POSTN affect the development of CRS.

In summary, the transcriptome sequencing data of the CRS rat models at different time models were systematically analyzed, and FN1 and POSTN were thus identified as novel biomarkers in CRS. These were externally validated in public datasets and local clinical samples, providing novel insights into the molecular mechanisms of CRS.

AUTHOR CONTRIBUTIONS

YZH designed the whole study and provided financial support. JJL developed the algorithm, drew the plots, conducted the experiments, and wrote the original manuscript. XHH and WWL did help to collecting the clinical samples, editing and reviewing.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This project was supported by the Guangdong Basic and Applied Basic Research Fund (Key project of Guangdong-Foshan Joint Fund) (2019B1515120044), Science and Technology Innovation Project from Foshan, Guangdong (FS0AA-KJ218-1301-0006) and Science and Technology Innovation Project from Foshan, Guangdong (FS0AA-KJ218-1301-0010).

REFERENCES

1. Ronco C, Haapio M, House AA, Anavekar N, Bellomo R. Cardiorenal syndrome. J Am Coll Cardiol. 2008; 52:1527–39. https://doi.org/10.1016/j.jacc.2008.07.051 PMID:19007588

2. Collins AJ, Foley RN. A decade after the KDOQI CKD guidelines: impact on the United States and global public policy. Am J Kidney Dis. 2012; 60:697–700. https://doi.org/10.1053/j.ajkd.2012.08.015 PMID:23067633

3. Tyralla K, Amann K. Cardiovascular changes in renal failure. Blood Purif. 2002; 20:462–5. https://doi.org/10.1159/000063551 PMID:12207093

4. Lip GYH, Coca A, Kahan T, Boriani G, Manolis AS, Olsen MH, Oto A, Potpara TS, Steffel J, Marín F, de Oliveira Figueiredo MJ, de Simone G, Tzou WS, et al, and Reviewers. Hypertension and cardiac arrhythmias: a consensus document from the European Heart Rhythm Association (EHRA) and ESC Council on Hypertension, endorsed by the Heart Rhythm Society (HRS), Asia-Pacific Heart Rhythm Society (APHRS) and Sociedad Latinoamericana de Estimulación Cardíaca y Electrofisiología (SOLEACE). Europace. 2017; 19:891–911. https://doi.org/10.1093/europace/eux091 PMID:28881872

5. Figueiredo EL, Leão FV, Oliveira LV, Moreira MC, Figueiredo AF. Microalbuminuria in nondiabetic and nonhypertensive systolic heart failure patients. Congest Heart Fail. 2008; 14:234–8. https://doi.org/10.1111/j.1751-7133.2008.00008.x PMID:18983285

6. Hong Z, Jiang Y, Liu P, Zhang L. Association of microalbuminuria and adverse outcomes in hypertensive patients: a meta-analysis. Int Urol Nephrol. 2021; 53:2311–9. https://doi.org/10.1007/s11255-021-02795-w PMID:33674949

7. Fangel MV, Nielsen PB, Kristensen JK, Larsen TB, Overvad TF, Lip GY, Jensen MB. Albuminuria and Risk of Cardiovascular Events and Mortality in a General Population of Patients with Type 2 Diabetes Without Cardiovascular Disease: A Danish Cohort Study. Am J Med. 2020; 133:e269–79. https://doi.org/10.1016/j.amjmed.2019.10.042 PMID:32205071

8. Odutayo A, Hsiao AJ, Emdin CA. Prevalence of Albuminuria in a General Population Cohort of Patients With Established Chronic Heart Failure. J Card Fail. 2016; 22:33–7. https://doi.org/10.1016/j.cardfail.2015.10.009 PMID:26505813

9. Miura M, Sakata Y, Miyata S, Nochioka K, Takada T, Tadaki S, Ushigome R, Yamauchi T, Takahashi J, Shimokawa H, and CHART-2 Investigators. Prognostic impact of subclinical microalbuminuria in patients with chronic heart failure. Circ J. 2014; 78:2890–8.

https://doi.org/10.1253/circj.CJ-14-0787
PMID:25421233

10. Muzi-Filho H, Jannuzzi LB, Bouzan ACS, Alves-Barros S, Alves-Bezerra DS, Pereira-Acácio A, Ferreira BSN, Silva-Pereira D, Costa-Sarmento G, Vieyra A. Histone Deacetylase Activity and the Renin-Angiotensin-Aldosterone System: Key Elements in Cardiorenal Alterations Provoked by Chronic Malnutrition in Male Adult Rats. Cell Physiol Biochem. 2020; 54:1143–62. https://doi.org/10.33594/000000306 PMID:33201644

11. Panico K, Abrahão MV, Trentin-Sonoda M, Muzi-Filho H, Vieyra A, Carneiro-Ramos MS. Cardiac Inflammation after Ischemia-Reperfusion of the Kidney: Role of the Sympathetic Nervous System and the Renin-Angiotensin System. Cell Physiol Biochem. 2019; 53:587–605. https://doi.org/10.33594/000000159 PMID:31535830

12. Linhart C, Ulrich C, Greinert D, Dambeck S, Wienke A, Girndt M, Pliquett RU. Systemic inflammation in acute cardiorenal syndrome: an observational pilot study. ESC Heart Fail. 2018; 5:920–30. https://doi.org/10.1002/ehf2.12327 PMID:30015388

13. Wang J, Zhang W, Wu L, Mei Y, Cui S, Feng Z, Chen X. New insights into the pathophysiological mechanisms underlying cardiorenal syndrome. Aging (Albany NY). 2020; 12:12422–31. https://doi.org/10.18632/aging.103354 PMID:32561688

14. Lee SA, Cozzi M, Bush EL, Rabb H. Distant Organ Dysfunction in Acute Kidney Injury: A Review. Am J Kidney Dis. 2018; 72:846–56. https://doi.org/10.1053/j.ajkd.2018.03.028 PMID:29866457

15. Delgado-Valero B, Cachofeiro V, Martínez-Martínez E. Fibrosis, the Bad Actor in Cardiorenal Syndromes: Mechanisms Involved. Cells. 2021; 10:1824. https://doi.org/10.3390/cells10071824 PMID:34359993

16. Chen K, Huang X, Xie D, Shen M, Lin H, Zhu Y, Ma S, Zheng C, Chen L, Liu Y, Liao W, Bin J, Liao Y. RNA interactions in right ventricular dysfunction induced type II cardiorenal syndrome. Aging (Albany NY). 2021; 13:4215–41. https://doi.org/10.18632/aging.202385 PMID:33494070

17. Chuppa S, Liang M, Liu P, Liu Y, Casati MC, Cowley AW, Patullo L, Kriegel AJ. MicroRNA-21 regulates peroxisome proliferator-activated receptor alpha, a molecular mechanism of cardiac pathology in Cardiorenal Syndrome Type 4. Kidney Int. 2018; 93:375–89. https://doi.org/10.1016/j.kint.2017.05.014 PMID:28760335

18. Li L, Dong L, Xiao Z, He W, Zhao J, Pan H, Chu B, Cheng J, Wang H. Integrated analysis of the proteome and transcriptome in a MCAO mouse model revealed the molecular landscape during stroke progression. J Adv Res. 2020; 24:13–27. https://doi.org/10.1016/j.jare.2020.01.005 PMID:32181013

19. Zhang X, Wang Z, Hu L, Shen X, Liu C. Identification of Potential Genetic Biomarkers and Target Genes of Peri-Implantitis Using Bioinformatics Tools. Biomed Res Int. 2021; 2021:1759214. https://doi.org/10.1155/2021/1759214 PMID:34931168

20. Li S, Zhao J, Lv L, Dong D. Identification and Validation of TYMS as a Potential Biomarker for Risk of Metastasis Development in Hepatocellular Carcinoma. Front Oncol. 2021; 11:762821. https://doi.org/10.3389/fonc.2021.762821 PMID:34858842

21. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics. 2006; 7:191. https://doi.org/10.1186/1471-2105-7-191 PMID:16597342

22. Jentzer JC, Bihorac A, Brusca SB, Del Rio-Pertuz G, Kashani K, Kazory A, Kellum JA, Mao M, Moriyama B, Morrow DA, Patel HN, Rali AS, van Diepen S, Solomon MA, and Critical Care Cardiology Working Group of the Heart Failure and Transplant Section Leadership Council. Contemporary Management of Severe Acute Kidney Injury and Refractory Cardiorenal Syndrome: JACC Council Perspectives. J Am Coll Cardiol. 2020; 76:1084–101. https://doi.org/10.1016/j.jacc.2020.06.070 PMID:32854844

23. Zannad F, Rossignol P. Cardiorenal Syndrome Revisited. Circulation. 2018; 138:929–44. https://doi.org/10.1161/CIRCULATIONAHA.117.028814 PMID:30354446

24. Ye C, Wang H, Li Z, Xia C, Yuan S, Yan R, Yang X, Ma T, Wen X, Yang D. Comprehensive data analysis of genomics, epigenomics, and transcriptomics to identify specific biomolecular markers for prostate adenocarcinoma. Transl Androl Urol. 2021; 10:3030–45. https://doi.org/10.21037/tau-21-576 PMID:34430406

25. Hsieh CH, Rau CS, Wu SC, Liu HT, Huang CY, Hsu SY, Hsieh HY. Risk Factors Contributing to Higher Mortality Rates in Elderly Patients with Acute

Traumatic Subdural Hematoma Sustained in a Fall: A Cross-Sectional Analysis Using Registered Trauma Data. Int J Environ Res Public Health. 2018; 15:2426. https://doi.org/10.3390/ijerph15112426 PMID:30388747

26. Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, Jo Y, Cummins MR. Predicting Pressure Injury in Critical Care Patients: A Machine-Learning Model. Am J Crit Care. 2018; 27:461–8. https://doi.org/10.4037/ajcc2018525 PMID:30385537

27. Yao S, Chen W, Zuo H, Bi Z, Zhang X, Pang L, Jing Y, Yin X, Cheng H. Comprehensive Analysis of Aldehyde Dehydrogenases (ALDHs) and Its Significant Role in Hepatocellular Carcinoma. Biochem Genet. 2021. [Epub ahead of print]. https://doi.org/10.1007/s10528-021-10178-0 PMID:34928471

28. Ruiz-Ortega M, Ruperez M, Lorenzo O, Esteban V, Blanco J, Mezzano S, Egido J. Angiotensin II regulates the synthesis of proinflammatory cytokines and chemokines in the kidney. Kidney Int Suppl. 2002; 62:S12–22. https://doi.org/10.1046/j.1523-1755.62.s82.4.x PMID:12410849

29. Latini R, Aleksova A, Masson S. Novel biomarkers and therapies in cardiorenal syndrome. Curr Opin Pharmacol. 2016; 27:56–61. https://doi.org/10.1016/j.coph.2016.01.010 PMID:26894469

30. Kaddourah A, Goldstein SL, Basu R, Nehus EJ, Terrell TC, Brunner L, Bennett MR, Haffner C, Jefferies JL. Novel urinary tubular injury markers reveal an evidence of underlying kidney injury in children with reduced left ventricular systolic function: a pilot study. Pediatr Nephrol. 2016; 31:1637–45. https://doi.org/10.1007/s00467-016-3360-2 PMID:27139898

31. Gembillo G, Visconti L, Giusti MA, Siligato R, Gallo A, Santoro D, Mattina A. Cardiorenal Syndrome: New Pathways and Novel Biomarkers. Biomolecules. 2021; 11:1581. https://doi.org/10.3390/biom11111581 PMID:34827580

32. Virzì GM, Breglia A, Ankawi G, Bolin C, de Cal M, Cianci V, Vescovo G, Ronco C. Plasma Lipopolysaccharide Concentrations in Cardiorenal Syndrome Type 1. Cardiorenal Med. 2019; 9:308–15. https://doi.org/10.1159/000500480 PMID:31238313

33. Brisco MA, Zile MR, Ter Maaten JM, Hanberg JS, Wilson FP, Parikh C, Testani JM. The risk of death associated with proteinuria in heart failure is restricted to patients with an elevated blood urea nitrogen to creatinine ratio. Int J Cardiol. 2016; 215:521–6. https://doi.org/10.1016/j.ijcard.2016.04.100 PMID:27153048

34. Wang H, Yang F, Luo Z. An experimental study of the intrinsic stability of random forest variable importance measures. BMC Bioinformatics. 2016; 17:60. https://doi.org/10.1186/s12859-016-0900-5 PMID:26842629

35. Yifan C, Jianfeng S, Jun P. Development and Validation of a Random Forest Diagnostic Model of Acute Myocardial Infarction Based on Ferroptosis-Related Genes in Circulating Endothelial Cells. Front Cardiovasc Med. 2021; 8:663509. https://doi.org/10.3389/fcvm.2021.663509 PMID:34262953

36. Roguet A, Eren AM, Newton RJ, McLellan SL. Fecal source identification using random forest. Microbiome. 2018; 6:185. https://doi.org/10.1186/s40168-018-0568-3 PMID:30336775

37. Zheng X, Liu G, Huang R. Identification and Verification of Feature Immune-Related Genes in Patients With Hypertrophic Cardiomyopathy Based on Bioinformatics Analyses. Front Cardiovasc Med. 2021; 8:752559. https://doi.org/10.3389/fcvm.2021.752559 PMID:34765659

38. Rangaswami J, Bhalla V, Blair JEA, Chang TI, Costa S, Lentine KL, Lerma EV, Mezue K, Molitch M, Mullens W, Ronco C, Tang WHW, McCullough PA, and American Heart Association Council on the Kidney in Cardiovascular Disease and Council on Clinical Cardiology. Cardiorenal Syndrome: Classification, Pathophysiology, Diagnosis, and Treatment Strategies: A Scientific Statement From the American Heart Association. Circulation. 2019; 139:e840–78. https://doi.org/10.1161/CIR.0000000000000664 PMID:30852913

39. Lorin J, Guilland JC, Stamboul K, Guenancia C, Cottin Y, Rochette L, Vergely C, Zeller M. Increased Symmetric Dimethylarginine Level Is Associated with Worse Hospital Outcomes through Altered Left Ventricular Ejection Fraction in Patients with Acute Myocardial Infarction. PLoS One. 2017; 12:e0169979. https://doi.org/10.1371/journal.pone.0169979 PMID:28125604

40. Forman DE, Butler J, Wang Y, Abraham WT, O'Connor CM, Gottlieb SS, Loh E, Massie BM, Rich MW, Stevenson LW, Young JB, Krumholz HM. Incidence,

predictors at admission, and impact of worsening renal function among patients hospitalized with heart failure. J Am Coll Cardiol. 2004; 43:61–7. https://doi.org/10.1016/j.jacc.2003.07.031 PMID:14715185

41. Goffredo G, Barone R, Di Terlizzi V, Correale M, Brunetti ND, Iacoviello M. Biomarkers in Cardiorenal Syndrome. J Clin Med. 2021; 10:3433. https://doi.org/10.3390/jcm10153433 PMID:34362216

42. Song G, Liu K, Yang X, Mu B, Yang J, He L, Hu X, Li Q, Zhao Y, Cai X, Feng G. SATB1 plays an oncogenic role in esophageal cancer by up-regulation of FN1 and PDGFRB. Oncotarget. 2017; 8:17771–84. https://doi.org/10.18632/oncotarget.14849 PMID:28147311

43. Wang Y, Zhao M, Zhang Y. Identification of fibronectin 1 (FN1) and complement component 3 (C3) as immune infiltration-related biomarkers for diabetic nephropathy using integrated bioinformatic analysis. Bioengineered. 2021; 12:5386–401. https://doi.org/10.1080/21655979.2021.1960766 PMID:34424825

44. Su H, Xie J, Wen L, Wang S, Chen S, Li J, Qi C, Zhang Q, He X, Zheng L, Wang L. LncRNA Gas5 regulates Fn1 deposition via Creb5 in renal fibrosis. Epigenomics. 2021; 13:699–713. https://doi.org/10.2217/epi-2020-0449 PMID:33876672

45. Zhao A, Zhao K, Xia Y, Lyu J, Chen Y, Li S. Melatonin inhibits embryonic rat H9c2 cells growth through induction of apoptosis and cell cycle arrest via PI3K-AKT signaling pathway. Birth Defects Res. 2021; 113:1171–81. https://doi.org/10.1002/bdr2.1938 PMID:34231342

46. Patel M, Rodriguez D, Yousefi K, John-Williams K, Mendez AJ, Goldberg RB, Lymperopoulos A, Tamariz LJ, Goldberger JJ, Myerburg RJ, Junttila J, Shehadeh LA. Osteopontin and LDLR Are Upregulated in Hearts of Sudden Cardiac Death Victims With Heart Failure With Preserved Ejection Fraction and Diabetes Mellitus. Front Cardiovasc Med. 2020; 7:610282. https://doi.org/10.3389/fcvm.2020.610282 PMID:33330671

47. Kalyanasundaram A, Li N, Gardner ML, Artiga EJ, Hansen BJ, Webb A, Freitas MA, Pietrzak M, Whitson BA, Mokadam NA, Janssen PML, Mohler PJ, Fedorov VV. Fibroblast-Specific Proteotranscriptomes Reveal Distinct Fibrotic Signatures of Human Sinoatrial Node in Nonfailing and Failing Hearts. Circulation. 2021; 144:126–43. https://doi.org/10.1161/CIRCULATIONAHA.120.051583 PMID:33874740

48. Zhao J, Lv T, Quan J, Zhao W, Song J, Li Z, Lei H, Huang W, Ran L. Identification of target genes in cardiomyopathy with fibrosis and cardiac remodeling. J Biomed Sci. 2018; 25:63. https://doi.org/10.1186/s12929-018-0459-8 PMID:30115125

49. Gao Z, Sekar A, Li XM, Li XL, Sui LN. Identification of Key Candidate Genes and Chemical Perturbagens in Diabetic Kidney Disease Using Integrated Bioinformatics Analysis. Front Endocrinol (Lausanne). 2021; 12:721202. https://doi.org/10.3389/fendo.2021.721202 PMID:34557161

50. Wu J, Lin Q, Li S, Shao X, Zhu X, Zhang M, Zhou W, Ni Z. Periostin Contributes to Immunoglobulin a Nephropathy by Promoting the Proliferation of Mesangial Cells: A Weighted Gene Correlation Network Analysis. Front Genet. 2021; 11:595757. https://doi.org/10.3389/fgene.2020.595757 PMID:33488671

51. Wallace DP, White C, Savinkova L, Nivens E, Reif GA, Pinto CS, Raman A, Parnell SC, Conway SJ, Fields TA. Periostin promotes renal cyst growth and interstitial fibrosis in polycystic kidney disease. Kidney Int. 2014; 85:845–54. https://doi.org/10.1038/ki.2013.488 PMID:24284511

52. Pastori S, Virzì GM, Brocca A, de Cal M, Clementi A, Vescovo G, Ronco C. Cardiorenal syndrome type 1: a defective regulation of monocyte apoptosis induced by proinflammatory and proapoptotic factors. Cardiorenal Med. 2015; 5:105–15. https://doi.org/10.1159/000371898 PMID:25999959

53. Barth AS, Merk S, Arnoldi E, Zwermann L, Kloos P, Gebauer M, Steinmeyer K, Bleich M, Kääb S, Pfeufer A, Uberfuhr P, Dugas M, Steinbeck G, Nabauer M. Functional profiling of human atrial and ventricular gene expression. Pflugers Arch. 2005; 450:201–8. https://doi.org/10.1007/s00424-005-1404-8 PMID:15877233

54. Barth AS, Merk S, Arnoldi E, Zwermann L, Kloos P, Gebauer M, Steinmeyer K, Bleich M, Kääb S, Hinterseer M, Kartmann H, Kreuzer E, Dugas M, et al. Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: expression of a ventricular-like genomic signature. Circ Res. 2005; 96:1022–9. https://doi.org/10.1161/01.RES.0000165480.82737.33 PMID:15817885

55. Luo X, Yin J, Dwyer D, Yamawaki T, Zhou H, Ge H, Han CY, Shkumatov A, Snyder K, Ason B, Li CM, Homann O, Stolina M. Chamber-enriched gene expression

profiles in failing human hearts with reduced ejection fraction. Sci Rep. 2021; 11:11839. https://doi.org/10.1038/s41598-021-91214-2 PMID:34088950

56. Nakagawa S, Nishihara K, Miyata H, Shinke H, Tomita E, Kajiwara M, Matsubara T, Iehara N, Igarashi Y, Yamada H, Fukatsu A, Yanagita M, Matsubara K, Masuda S. Molecular Markers of Tubulointerstitial Fibrosis and Tubular Cell Damage in Patients with Chronic Kidney Disease. PLoS One. 2015; 10:e0136994. https://doi.org/10.1371/journal.pone.0136994 PMID:26317775

57. Han R, Hu S, Qin W, Shi J, Hou Q, Wang X, Xu X, Zhang M, Zeng C, Liu Z, Bao H. C3a and suPAR drive versican V1 expression in tubular cells of focal segmental glomerulosclerosis. JCI Insight. 2019; 4:e130986. https://doi.org/10.1172/jci.insight.130986 PMID:31292294

58. Scherer A, Günther OP, Balshaw RF, Hollander Z, Wilson-McManus J, Ng R, McMaster WR, McManus BM, Keown PA. Alteration of human blood cell transcriptome in uremia. BMC Med Genomics. 2013; 6:23. https://doi.org/10.1186/1755-8794-6-23 PMID:23809614

SUPPLEMENTARY MATERIALS

The R code used in this study

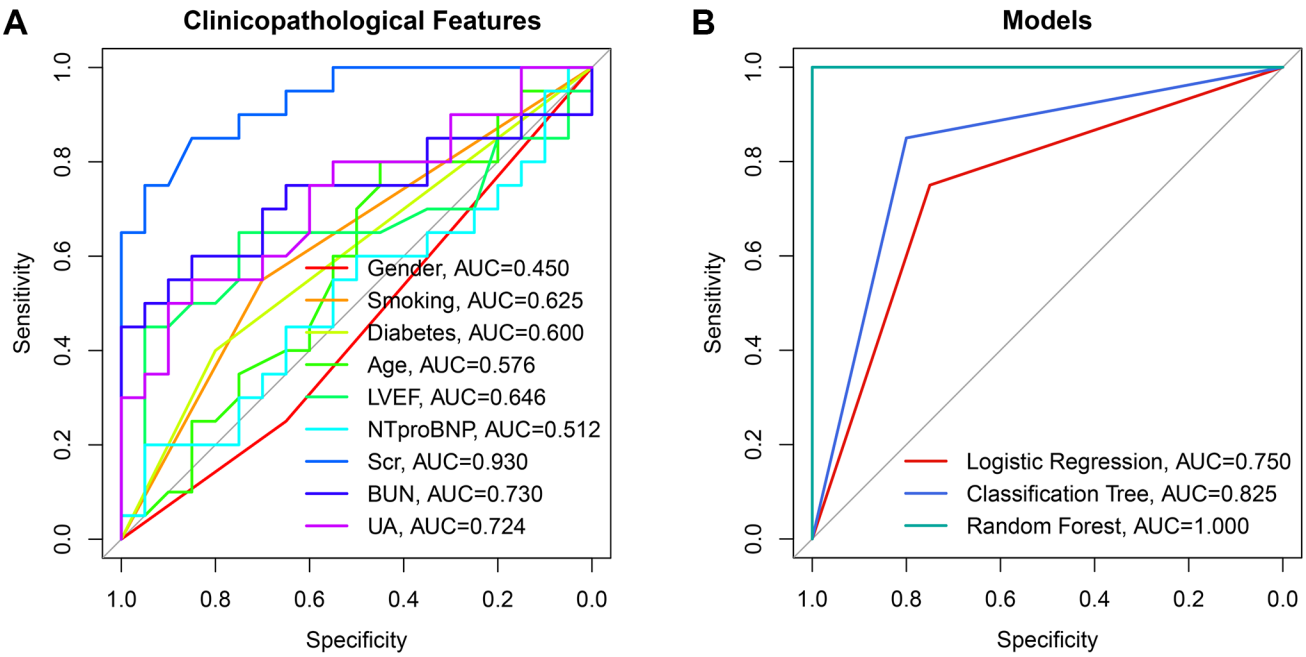
```
#Genomic difference analysis
library(limma)
logFoldChange=0
adjustP=0.05
rt=read.table("GSE98520_Supplemental_Table_1_5-6Nx_vs._Sham_Wk2.txt",sep="\
t",header=T,check.names=F)
rt=as.matrix(rt)
rownames(rt)=rt[,1]
exp=rt[,2:ncol(rt)]
dimnames=list(rownames(exp),colnames(exp))
rt=matrix(as.numeric(as.matrix(exp)),nrow=nrow(exp),dimnames=dimnames)
rt=avereps(rt)
modType=c(rep("con",conNum),rep("treat",treatNum))
design <- model.matrix(~0+factor(modType))
colnames(design) <- c("con","treat")
fit <- lmFit(rt,design)
cont.matrix<-makeContrasts(treat-con,levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)
allDiff=topTable(fit2,adjust='fdr',number=200000)
write.table(allDiff,file="All_limma.xls",sep="\t",quote=F)
#The volcano plot
library(ggpubr)
library(ggthemes)
deg.data <- read.table("score.cor.txt", header = T, sep = "\t")
head(deg.data)
deg.data$logp <- -log10(deg.data$FDR)
deg.data$group = "notsignificant"
deg.data$group[which((deg.data$FDR < 0.05) & (deg.data$logFC > 0))] = "up"
deg.data$group[which((deg.data$FDR < 0.05) & (deg.data$logFC < 0))] = "down"
table(deg.data$group)
deg.data$label = ""
deg.data <- deg.data[order(deg.data$FDR),]
up.genes <- head(deg.data$gene[which(deg.data$group == "up")], 0)
down.genes <- head(deg.data$gene[which(deg.data$group == "down")], 0)
deg.top10.genes <- c(as.character(up.genes),as.character(down.genes))
deg.data$label[match(deg.top10.genes, deg.data$gene)] <- deg.top10.genes
pdf(file="volcano.pdf",
                width = 5,
                height = 4,
)
ggscatter(deg.data, x = "logFC", y = "logp", color = "group", palette =  c("#4169E1","#BBBBBB","#E3170D"), size
= 1, label = deg.data$label, font.label =  8, repel = T, xlab = "Spearman Correlation", ylab = "-log10P",) +
theme_base() +

geom_hline(yintercept = 1.30, linetype = "dashed") + geom_vline(xintercept = c(0,0), linetype = "dashed")
dev.off()
#the differentially expressed genes between the sham-operated and subtotal nephrectomy rats in week 4, 5, and 7 are
obtained using the same algorithms.

#The protein-protein interaction network was constructed in STRING database.
#The time-series analyses were conducted by the Short Time-series Expression Miner  software.
```

```
#GO functional enrichment
library("org.Rn.eg.db")
rt=read.table("DEGs.txt",sep="\t",check.names=F,header=T)
genes=as.vector(rt[,1])
entrezIDs <- mget(genes, org.Rn.egSYMBOL2EG, ifnotfound=NA)
entrezIDs <- as.character(entrezIDs)
out=cbind(rt,entrezID=entrezIDs)
write.table(out,file="id.txt",sep="\t",quote=F,row.names=F)
library("clusterProfiler")
library("enrichplot")
library("ggplot2")
rt=read.table("id.txt",sep="\t",header=T,check.names=F)

rt=rt[is.na(rt[,"entrezID"])==F,]

gene=rt$entrezID
kk <- enrichGO(gene = gene,
                    OrgDb = org.Rn.eg.db,
                    pvalueCutoff =0.05,
                    qvalueCutoff = 0.05,
                    ont="all",
                    readable =T)
write.table(kk,file="GO.txt", sep="\t",quote=F, row.names = F)

pdf(file="bubble_go.pdf",width = 10,height = 8)
dotplot(kk,showCategory = 10,split="ONTOLOGY") + facet_grid(ONTOLOGY~.,  scale='free')
dev.off()

#The calculation of the areas under the receiver operating curve.
library(pROC)
inputFile="input.txt"
outFile="ROC.pdf"

rt=read.table(inputFile,header=T,sep="\t",check.names=F,row.names=1)
y=colnames(rt)[1]
bioCol=c("#E3170D","#4169E1","#03A89E","#03A89E")
if(ncol(rt)>4){
      bioCol=rainbow(ncol(rt))}

pdf(file=outFile,width=4.5,height=4.5)
roc1=roc(rt[,y], as.vector(rt[,2]))
aucText=c( paste0(colnames(rt)[2],", AUC=",sprintf("%0.3f",auc(roc1))) )
plot(roc1, col=bioCol[1])
for(i in 3:ncol(rt)){
       roc1=roc(rt[,y], as.vector(rt[,i]))
       lines(roc1, col=bioCol[i-1])
       aucText=c(aucText, paste0(colnames(rt)[i],", AUC=",sprintf("%0.3f",auc(roc1))) )
}
legend("bottomright", aucText, lwd=2,bty="n", col=bioCol[1:(ncol(rt)-1)])  dev.off()

#The Spearman correlation analyses
library(ggplot2)
library(ggpubr)
rt <- read.table("pre_cor_log.txt", sep = "\t", header = T, check.names = F, row.names = 1)
rt <- t(rt)
```

```
rt <- as.data.frame(rt)
class(rt)
str(rt)
gene1name <- colnames(rt)[1]
gene2name <- colnames(rt)[2]
pdf(file="spearman_cor.pdf",
        width = 4,
        height = 4,

)
ggscatter(rt,
                x = gene1name,
                y = gene2name,
                add = "reg.line",
                conf.int = TRUE,
                cor.coef = TRUE,
                cor.method = "spearman",
                xlab = gene1name,
                ylab = gene2name,color = "blue")
dev.off()

#The logistic regression and nomogram
rt=read.table("input.txt",header=T, sep="\t",check.names=F, row.names=1)
mymodel <- glm(status~FN1 + POSTN,family=binomial(link = "logit"),data = rt)
ddist <- datadist(rt)
options(datadist="ddist")
mymodelSummary = summary(mymodel)
outTab=data.frame()
outTab=rbind(outTab,
        cbind(coef=coefficients(mymodel),
                OR=matrix(exp(coefficients(mymodel)))[,1],
                OR.95L=exp(confint(mymodel))[,"2.5 %"],
                OR.95H=exp(confint(mymodel))[,"97.5 %"],
                pvalue=mymodelSummary$coefficients[,"Pr(>|z|)"])
)
outTab=cbind(id=row.names(outTab), outTab)
write.table(outTab, file="multilogit.xls", sep="\t",row.names=F, quote=F)
rt<-read.table("input.txt",header=T,sep="\t",row.names=1)
ddist <- datadist(rt)
options(datadist="ddist")
mymodel<-lrm(status~., data=rt, x=T, y=T)
mynom<-  nomogram(mymodel,  fun=plogis,fun.at=c(0.3,0.5,0.7,0.9,0.99,0.999),lp=F,
funlabel="Risk of CRS")
pdf("Nom.pdf",8,7)
plot(mynom)
dev.off()
library(pROC)
train=read.table("pre.txt",header=T,sep="\t",check.names=F,row.names=1)
ddist <- datadist(train)
options(datadist="ddist")
test=read.table("test.txt", header=T,sep="\t",check.names=F,row.names=1)
test=test[,colnames(train)]
rt=rbind(train,test)
mymodel <- glm(status~FN1 + POSTN, family=binomial(link = "logit"), data = train)
predict <- predict.glm(mymodel,type = "response",newdata = train)
predict =ifelse(predict>0.5, 1, 0)
```

```
train$predict = predict
write.csv(train, "train_predict.csv")
true_value = train[,1]
modelroc <- roc(true_value,predict)
pdf("ROC_train_LR.pdf",6,6)
plot(modelroc, print.auc=TRUE, auc.polygon=TRUE,legacy.axes=TRUE,
     max.auc.polygon=TRUE, print.thres=TRUE,auc.polygon.col="#4169E1")
dev.off()

#Classification tree
library(rpart)
library(partykit)
library(caret)
rt <- read.table("input.txt",header=T,sep="\t",check.names = F,row.names = 1)
rt$status <-as.factor(rt$status)
tree.biop <- rpart(status ~ ., data = rt)
tree.biop$cptable
cp <- min(tree.biop$cptable[3])
prune.tree.biop = prune(tree.biop, cp <- cp)
plot(as.party(prune.tree.biop))
rparty.test <- predict(prune.tree.biop, newdata = rt,
                       type = "class")
table(rparty.test, rt$status)

#Random forest
library(ROCit)
library(randomForest)
library(rms)
library(pROC)
train_data = read.table("train.sva.txt",header=T,sep="\t",check.names = F,row.names = 1)
train_data$Group = as.factor(train_data$Group)
train_data$Group = as.factor(train_data$Group)
train_randomforest <- randomForest(Group ~.,
                                   data = train_data,
                                   ntree =500,
                                   mtry=3,
                                   importance=TRUE ,
                                   proximity=TRUE)
train_randomforest$importance
library(pROC)
pre_ran <- predict(train_randomforest,newdata=train_data)
train_data$pre_ran <- pre_ran
train_data$pre_ran <- as.numeric(train_data$pre_ran)
obs_p_ran = data.frame(prob=pre_ran,obs=train_data$Group)
table(train_data$Group,pre_ran,dnn=c("True value","Predicted value"))
```

Supplementary Figure



Supplementary Figure 1. The diagnosis ability of the clinicopathological parameters (**A**) and the established models (**B**).

Supplementary Tables

Supplementary Table 1. The PPI network analysis via CytoHubba.

Node_name	MCC	DMNC	MNC	Degree	EPC	Bottle Neck	EcCentricity	Closeness	Radiality	Betweenness	Stress	Clustering Coefficient
Kif22	6	0.46346	3	3	2.349	1	0.125	3.5	0.625	0	0	1
Pi16	1	0	1	1	2.559	1	0.25	6	1.28571	0	0	0
FN1	314	0.47886	10	10	8.878	7	0.25	11.83333	2.19643	15.56667	42	0.53333
POSTN	318	0.49882	10	10	8.793	3	0.375	12	2.25	18.23333	54	0.55556
Plod2	6	0.46346	3	3	5.435	1	0.375	8.5	1.875	0	0	1
Timp1	288	0.58315	8	8	8.423	1	0.25	10.83333	2.08929	3.9	16	0.71429
Loxl1	168	0.5854	7	7	8.365	1	0.25	10.33333	2.03571	2.66667	10	0.7619
Tnfrsf11b	2	0.30779	2	2	4.927	1	0.25	7.83333	1.76786	0	0	1
Vcan	120	0.64826	5	5	7.536	1	0.25	9.33333	1.92857	0	0	1
Thy1	24	0.56839	4	4	6.887	1	0.25	8.83333	1.875	0	0	1
Col8a1	13	0.47366	4	5	6.563	2	0.375	9.5	1.98214	26.66667	54	0.5
Mmp14	24	0.56839	4	4	6.898	1	0.25	8.83333	1.875	0	0	1
Ltbp2	24	0.56839	4	4	6.919	1	0.25	8.83333	1.875	0	0	1
Col3a1	300	0.47886	10	10	8.711	3	0.375	12	2.25	19.73333	54	0.53333
Fstl1	24	0.56839	4	4	6.665	1	0.25	8.83333	1.875	0	0	1
Col1a1	350	0.39597	13	13	9.283	2	0.375	13.5	2.41071	51.23333	112	0.39744
Prc1	12	0.47366	4	4	2.521	1	0.25	4	0.6875	0.66667	2	0.83333
Top2a	12	0.47366	4	4	2.523	1	0.25	4	0.6875	0.66667	2	0.83333
Racgap1	12	0.47366	4	4	2.59	1	0.25	4	0.6875	0.66667	2	0.83333
Cdkn3	6	0.46346	3	3	2.321	1	0.125	3.5	0.625	0	0	1

Supplementary Table 2. The raw data of the RF-qPCR experiments.

ID	Group	CT(FN1)	CT(POSTN)	CT(GAPDH)
2021016	CRS	22.714	21.62269	17.47264
2021035	CRS	21.62767	23.80968	17.20525
2021044	CRS	21.07704	22.63808	17.18994
2021045	CRS	21.36649	23.64455	16.87597
2021061	CRS	21.78654	22.12539	17.33394
2021072	CRS	19.53786	23.48876	16.93788
2021075	CRS	21.59798	22.45887	17.65224
2021079	CRS	20.83566	23.42787	16.70008
2021084	CRS	20.00147	23.39505	18.47383
2021087	CRS	21.25389	22.97611	18.36617
2021088	CRS	21.53118	23.40699	16.77462
2021096	CRS	18.06519	24.10553	16.60071
2021097	CRS	22.15883	23.26074	17.02379
2021112	CRS	21.6212	24.72875	18.31842
2021119	CRS	21.28442	21.61209	16.47575
2021131	CRS	20.43167	23.94418	18.85962

2021145	CRS	21.78432	23.92037	17.80561
2021157	CRS	20.9538	22.39475	18.02395
2021159	CRS	21.75587	22.55037	17.49426
2021165	CRS	21.72398	22.82194	17.40736
2021003	Control	22.04717	24.19026	13.92721
2021008	Control	22.29514	23.86475	16.58527
2021013	Control	22.21988	22.81359	15.47096
2021015	Control	22.8975	23.48162	16.10068
2021017	Control	22.17563	23.79683	15.873
2021021	Control	22.32354	23.76648	17.50554
2021022	Control	21.82626	23.32713	16.35305
2021026	Control	22.69801	22.79801	17.94128
2021029	Control	21.21703	23.83898	17.70134
2021031	Control	22.15845	24.01224	16.21806
2021032	Control	20.26998	21.77754	16.41206
2021034	Control	22.57254	23.01443	15.82979
2021036	Control	22.10823	25.4372	18.03494
2021038	Control	19.83442	24.09309	16.93175
2021039	Control	22.8405	23.61109	17.03301
2021040	Control	20.72439	24.23108	16.52985
2021041	Control	22.49104	22.74978	16.83147
2021042	Control	22.12857	23.5926	16.46831
2021046	Control	23.01731	22.66537	17.41212
2021047	Control	22.55935	22.80843	18.6565