

Identification of prognostic chromatin-remodeling genes in clear cell renal cell carcinoma

Yujing Yang^{1,*}, Chengyuan Wang^{2,*}, Ningde Wei², Ting Hong², Zuyu Sun², Jiawen Xiao³, Jiayi Yao², Zhi Li¹, Tao Liu²

¹Department of Medical Oncology, The First Affiliated Hospital of China Medical University, Shenyang 110001, P.R. China

²Department of Urology, The First Affiliated Hospital of China Medical University, Shenyang 110001, P.R. China

³Department of Medical Oncology, Shenyang Fifth People Hospital, Tiexi District, Shenyang 110001, P.R. China

*Equal contribution

Correspondence to: Zhi Li, Tao Liu; **email:** zli@cmu.edu.cn, cmuliu2016@163.com, <https://orcid.org/0000-0002-3044-8813>

Keywords: clear cell renal cell carcinoma (ccRCC), chromatin-remodeling genes, prognostic biomarkers, BPTF, SIN3A

Received: March 24, 2020

Accepted: September 28, 2020

Published: November 20, 2020

Copyright: © 2020 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The aim of this study was to investigate the effects of chromatin-remodeling genes on the prognosis of patients with clear cell renal cell carcinoma (ccRCC). In TCGA-KIRC patients, two subgroups based on 86 chromatin-remodeling genes were established. The random forest algorithm was used for feature selection to identify BPTF, SIN3A and CNOT1 as characterized chromatin remodelers in ccRCC with good prognostic value. YY1 was indicated to be a transcription factor of genes highly related to BPTF, SIN3A and CNOT1. Functional annotations indicated that BPTF, SIN3A, CNOT1 and YY1 are all involved in the ubiquitin-mediated proteolysis process and that high expression of any of the five associated E3 ubiquitin ligases found in the pathway suggests a good prognosis. Protein network analysis indicated that BPTF has a targeted regulatory effect on YY1. Another independent dataset from International Cancer Genome Consortium (ICGC) showed a strong consistency with results in TCGA. In conclusion, we demonstrate that BPTF, SIN3A and CNOT1 are novel prognostic factors that predict good survival in ccRCC. We predicted that the good prognostic value of chromatin-remodeling genes BPTF and SIN3A is related to the regulation of YY1 and that YY1 regulates E3 ubiquitin ligases for further degradation of oncoproteins in ccRCC.

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC), which accounts for 70-80% of all renal cell carcinoma (RCC) patients, is one of the most lethal malignancies of the urinary system [1]. The genetic changes underlying ccRCC include alterations in genes regulating the hypoxia pathway (for example, VHL) and maintaining chromatin states (for example, PBRM1 and ARID1A) [2]. Prognostic biomarkers play an important role in stratifying patients to avoid both overtreatment and inadequate treatment of ccRCC [3]. Currently, numerous studies on the hypoxia pathway have been reported, but few prognostic factors

are widely accepted in ccRCC. Recent studies have shown that ARID1 and PBRM1 have significant prognostic value and distinguish the sensitivity of patients to therapy in RCC [4–9]. Interestingly, both ARID1 and PBRM1 are chromatin-remodeling genes. This strongly suggests that chromatin-remodeling genes are crucial in determining the prognosis and guiding the treatment of ccRCC. Therefore, the exploration of novel prognostic biomarkers involved in chromatin remodeling in ccRCC is urgently required.

Chromatin remodeling is the process of dynamic modification of the chromatin structure to control gene

expression by allowing regulatory transcription proteins access to condensed DNA. In addition, chromatin remodeling plays epigenetic regulatory roles in many processes related to cancer development, including cell cycle progression, cell death, cell pluripotency, and DNA repair [10–12]. During DNA replication, DNA repair and transcription, the chromatin structure is continuously modified, thereby exposing specific gene regions and allowing DNA-interacting enzymes access to specific regions of DNA. Chromatin remodelers play critical roles in stem and progenitor cell differentiation, lineage commitment and organogenesis during mammalian development [13, 14]. With advances in gene sequencing and an in-depth understanding of the epigenetic regulation of DNA-templated processes, numerous studies have indicated that extensive dysregulation of chromatin remodelers and the resulting inappropriate expression of regulatory genes together lead to oncogenesis [15–18]. Therapeutic targeting of chromatin remodelers has been shown to be effective in controlling tumorigenesis [12]. Furthermore, the prognostic value and therapeutic decision-making significance of chromatin remodelers in ccRCC have been demonstrated [5, 6, 19, 20]. In recent years, chromatin-remodeling genes, including ARID1, PBRM1, BAP1, SETD2, bromodomain PHD-finger transcription factor (BPTF), and SMARCA4, have been the focus of research in various types of tumors [4, 21–24]. However, the prognostic value of chromatin-remodeling genes as an overall indicator has not been evaluated in many types of tumors. Hence, in this study, we explored a prognostic model based on chromatin-remodeling genes and evaluated the prognostic value of several characterized chromatin-remodeling genes in ccRCC.

In our study, we downloaded data from The Cancer Genome Atlas (TCGA) database, International Cancer Genome Consortium (ICGC) and Human Transcription Factor Database (HumanTFDB). Then, a prognostic model based on chromatin-remodeling genes was established, and characterized chromatin-remodeling genes were identified in ccRCC. In this paper, the prognostic value of chromatin-remodeling genes and the associated mechanism in ccRCC were explored. The study of chromatin-remodeling genes provides new insight into the prognostic biomarkers and mechanisms of ccRCC.

RESULTS

Identification of subgroups with distinct prognoses in TCGA-KIRC patients based on consensus clustering

After data processing, a total of 496 TCGA-KIRC patients were included, and 86 chromatin-remodeling

genes were accepted as the definitive input in a first-step consensus clustering analysis. According to the comparison of the cumulative distribution function (CDF) curve from 2 to 10 category numbers (Figure 1B, 1C), two major clusters (cluster A and cluster B) were identified for which model stabilization became the highest with the optimal classification accuracy of 2 categories in the consensus matrix (Figure 1A). Patients belonging to cluster B (n=290) exhibited significantly better prognostic values for overall survival (OS) (P=0.00026) and progression-free survival (PFS) (P=0.00019), while patients in cluster A (n=206) displayed a strong tendency for poor clinical outcomes (Figure 1D, 1E). Univariate and multivariate Cox regression analyses indicated that in addition to age, clinical stage and differentiation grade, cluster B was a significant factor correlated with OS and PFS endpoints in TCGA-KIRC patients (hazard ratio (HR): 0.667; 95% confidence interval (CI): 0.535-0.831; P <0.001; Table 1). As the two clusters show substantial differences in survival outcomes, their associations with clinical parameters and vital mutations were also estimated. Clearly, cluster B was significantly related to a relatively low clinical stage and good histological differentiation (Supplementary Table 1). However, the statuses of some top mutant genes in renal carcinoma patients, including VHL, PBRM1, TP53 and MTOR, which were collected from the Catalogue of Somatic Mutations in Cancer (COSMIC) database, did not show evident differences between the two clusters.

Characterized chromatin-remodeling factor screening by the random forest algorithm

After multiple cross-validations of the random forest (RF) process, the relationship between the number of selected variables and model accuracy based on the consensus cluster A and cluster B grouping was evaluated (Figure 2A). The cut-off value of the model's accuracy was defined as 0.9, and the accuracy reached 90% when at least 3 variables were included in the model (Supplementary Table 2). The model achieved its highest accuracy when 38 variables were included by rank, and the top three variables were CNOT1, SIN3A and BPTF (Supplementary Table 3), which means that these three factors could represent the original chromatin-remodeling model with 90% accuracy. According to the RF algorithm, the importance of the contributions of all 86 chromatin-remodeling factors in the model was also calculated (Figure 2B). The top three variables of the green features, which are features designated as having robust importance, were CNOT1, SIN3A and BPTF. Agreement between the two approaches validated the consistency and enhanced the confidence of the result. The Z-score-processed expression data of the 86 chromatin-remodeling factor

dendrogram clustering heatmap based on prespecified groups were utilized to discover the differences in CNOT1, SIN3A and BPTF expression levels between cluster A and cluster B (Figure 2C). Consensus clustering based on BPTF, SIN3A and CNOT1 was also performed to obtain cluster C and cluster D. Interestingly, we found that BPTF, SIN3A and CNOT1 were all highly expressed in cluster B, a group previously associated with relatively good outcomes, and the patient proportions and distributions of cluster C and cluster D were consistent with those of cluster A and cluster B. Then, comparisons of OS and PFS between cluster C and cluster D for the 496 TCGA-KIRC patients were performed to verify conformity. As shown in Figure 2D, 2E, compared with those in cluster C, the ccRCC patients in cluster D had better OS ($P=0.0011$) and PFS ($P=0.00031$). The risk-predicted accuracy of the combined model including clinical stage, BPTF, SIN3A and CNOT1 was higher than that of the univariate clinical stage model for one-year, three-year and five-year OS (Supplementary Figure 1A,

1B). The prediction of long-term survival varied even more between the two models. In summary, the discovery that a model based on BPTF, SIN3A and CNOT1 could faithfully represent the model resulting from the original input of chromatin-remodeling factors was confirmed in a variety of ways.

Survival analysis and correlation analyses of clinical parameters with BPTF, SIN3A and CNOT1

The OS and PFS estimations of BPTF, SIN3A and CNOT1 were performed for 496 TCGA-KIRC patients. After dividing ccRCC patients into three equal quantile divisions based on selected gene expression, we found that compared with the low expression group, the high BPTF expression group exhibited better OS ($P<0.0001$) and PFS ($P<0.0001$) (Figure 3A, 3B). Likewise, ccRCC patients with high CNOT1 expression had better OS ($P<0.0001$) and PFS ($P=0.00018$) than those with low expression (Figure 3C, 3D), and patients in the high SIN3A expression group also had better OS

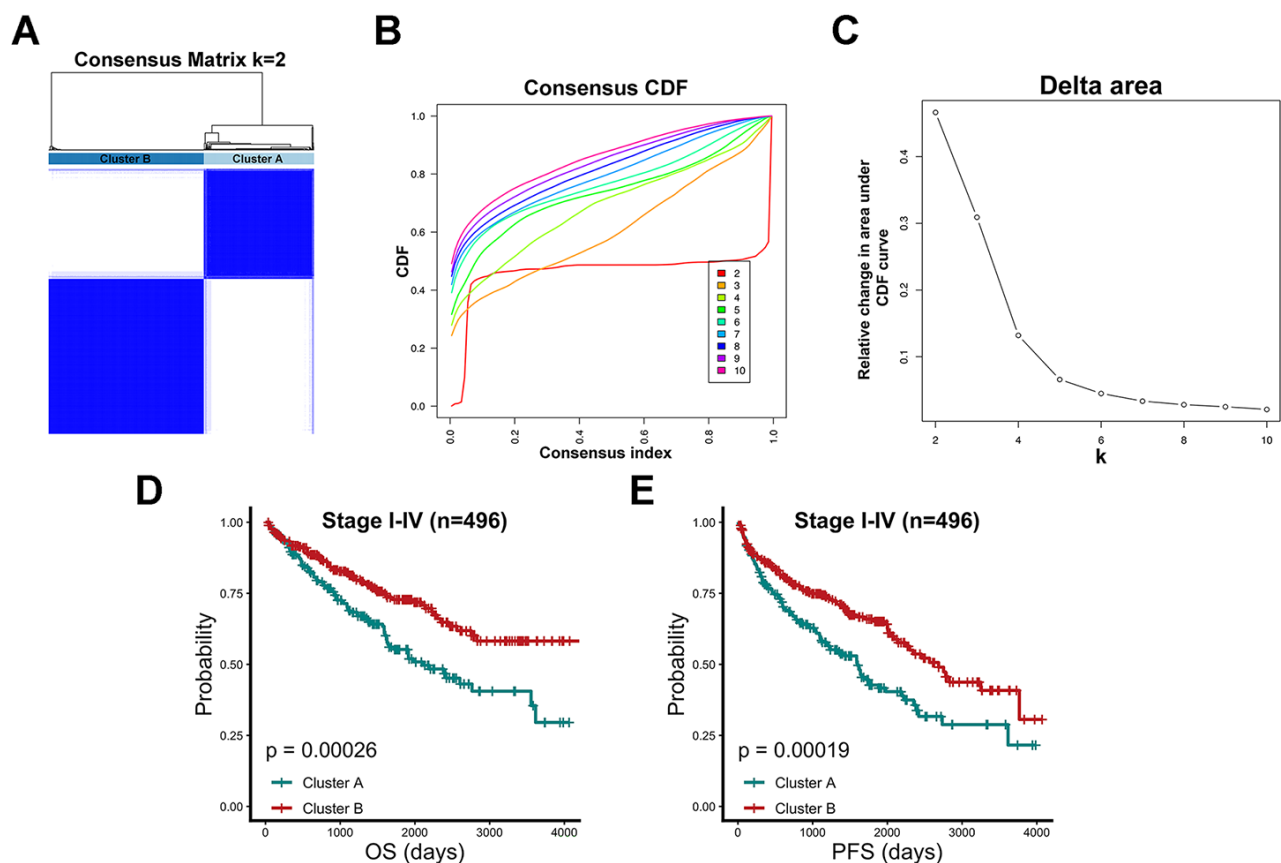


Figure 1. Consensus clustering algorithm identified two clusters with prognostic value. (A) 496 TCGA-KIRC patients were grouped into a consensus matrix with Cluster A (n=206) and Cluster B (n=290) based on 86 chromatin-remodeling factors and consensus clustering algorithm. (B) Consensus matrix among cumulative distribution function (CDF) curves from 2 to 10 clusters. (C) Relative changes in delta area under the CDF curve for each k category from 2 to 10. (D) Comparison of overall prognostic differences between Cluster A and Cluster B upon all stage patients. (E) Comparison of progression-free prognostic differences between Cluster A and Cluster B upon all stage patients.

Table 1. Univariate and multivariate cox regression of Cluster A and Cluster B for overall survival and progression-free survival in TCGA-KIRC (n=496).

Variable	Univariate analysis		Multivariate analysis		
	HR (95% CI)	P value	HR (95% CI)	P value	
TCGA-KIRC OS (n=496)					
Cluster	Cluster B vs Cluster A	0.667(0.535-0.831)	< 0.001	0.767(0.613-0.96)	0.02
Age	Continuous	1.03(1.017-1.044)	< 0.001	1.032(1.018-1.047)	< 0.001
Gender	Male vs Female	0.961(0.766-1.206)	0.731		
Stage	Stage IV vs Stage I-III	3.037(2.419-3.812)	< 0.001	2.672(2.102-3.398)	< 0.001
Grade	Grade III-IV vs Grade I-II	2.06(1.595-2.661)	< 0.001	1.565(1.196-2.048)	0.001
TCGA-KIRC PFS (n=496)					
Cluster	Cluster B vs Cluster A	0.695(0.573-0.843)	< 0.001	0.799(0.656-0.973)	0.025
Age	Continuous	1.021(1.009-1.032)	< 0.001	1.021(1.009-1.033)	< 0.001
Gender	Male vs Female	1.118(0.912-1.371)	0.282		
Stage	Stage IV vs Stage I-III	3.788(3.073-4.668)	< 0.001	3.431(2.761-4.263)	< 0.001
Grade	Grade III-IV vs Grade I-II	2.011(1.618-2.499)	< 0.001	1.652(1.32-2.069)	< 0.001

(P= 0.00081) and PFS (P= 0.00015) outcomes than those in the low expression group (Figure 3E, 3F). Associations between the expression levels of these three characterized chromatin-remodeling genes and clinicopathological parameters were also evaluated. As the clinical stage advanced, the expression levels of BPTF, SIN3A and CNOT1 appeared to decrease gradually (Figure 4A–4C). In regard to differentiation grade, significantly reduced expression was found in poorly differentiated tumor tissue compared with normal renal tissue or neoplastic tumor tissue with a level I grade (Figure 4D–4F).

Functional annotation of CNOT1, SIN3A and BPTF and property analysis of their highly correlated genes

Since the results of the survival analysis and correlation analysis between clinicopathological parameters and BPTF, SIN3A and CNOT1 in TCGA-KIRC were highly consistent, we investigated the potential mechanisms of these three chromatin-remodeling genes and their highly correlated genes (Pearson $r > 0.5$; $P < 0.001$) in ccRCC. As shown in Figure 5A–5C, gene ontology (GO) analysis indicated that BPTF, SIN3A and CNOT1 function in the biological process of chromatin regulation, which is consistent with current knowledge. In addition, some of their correlated genes were found to be involved in the process of utilizing the autophagic mechanism. Furthermore, the top 5 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis terms indicated that genes associated with ubiquitin-mediated proteolysis pathways were relatively enriched (Figure 5D–5F). Because the genetic annotations and gene sets both largely associated with protein degradation pathways, we focused on the properties of their intersected genes. The Venn diagram

shown in Figure 6A illustrates the number and overlap of the significantly highly correlated genes of BPTF, SIN3A and CNOT1 (Pearson $r > 0.6$; $P < 0.001$) in ccRCC patients. Univariate Cox regression analysis of OS was utilized to evaluate the prognostic value of 165 intersected genes. We found that the majority of these genes (92.7%; 153/165) significantly predicted a good prognosis, while only a small group of them (1.8%; 3/165) were significantly associated with poor survival outcomes or were nonsignificant (5.5%; 9/165) in ccRCC (Figure 6B). Comprehensive GO analysis performed on the 165 intersected genes suggested that processes related to posttranscriptional gene silencing regulation and protein polyubiquitination were the top enriched processes (Figure 6C). An interactive network of the top 15 enriched pathways among the intersected genes determined by KEGG analysis was constructed, and it was very consistently found that ubiquitin-mediated proteolysis pathways ranked as the second highest score in the enrichment analysis (Supplementary Figure 2A). Hence, we hypothesized that BPTF, SIN3A and CNOT1 all play roles in the occurrence and development of ccRCC by affecting autophagy regulation and activating ubiquitin-mediated proteolysis, which might account for their good prognostic value.

Prediction of YY1 as a transcriptional factor of the intersected genes

The iRegulon plugin was used to predict the putative transcription factors of all 165 intersected genes with Cytoscape Java software. According to iRegulon, YY1 was the most likely transcription factor of these genes, and the predictive results of other regulators were not sufficient with strong evidence. After excluding genes that did not match YY1 (44/165; blue module) and

genes with a correlation factor less than 0.3 with YY1 by Pearson's correlation analysis (11/165; gray module), the remaining genes showed at least two evidence-based relationships with YY1, including enriched motif mapping information obtained from the Encyclopedia of DNA Elements (ENCODE), JASPAR and HOMER databases, CHIP-seq track signal

calculations and correlation analysis (Figure 7). Four out of the top 10 enriched pathways identified by KEGG analysis were verified by single-gene YY1 gene set enrichment analysis (GSEA), the TGF-beta signaling pathway, the Wnt signaling pathway, RCC and ubiquitin-mediated proteolysis (Supplementary Figure 2B); these results were consistent with the

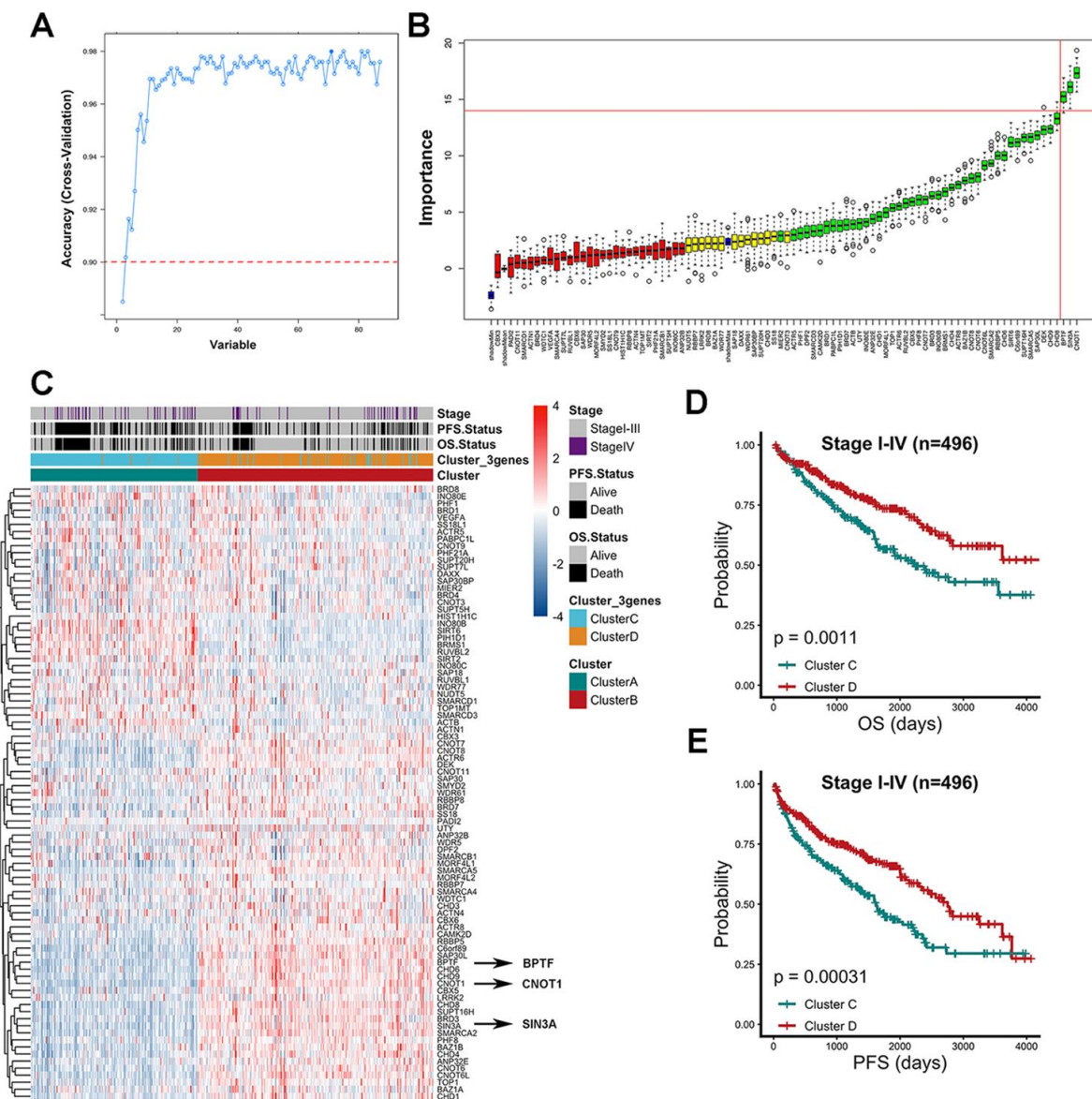


Figure 2. Characterized chromatin-remodeling factors selection based on random forest algorithm. (A) The relationship between the number of variables and model accuracy after multiple cross-validations of the random forest process was shown. The auxiliary line was added at 0.9 accuracy and the first point indicated 2 variables while the following points added one variable compared to the previous one. (B) The importance of each variable in model contribution was shown. The horizontal axis represents the position of 86 chromatin-remodeling factors in the rank of high to low importance from right to left. Red features were confirmed unimportant, yellow features were designated as tentative and green features were confirmed important. Three blue boxplots mapped to minimal, mean and maximum Z scores of shadow attributes respectively. (C) Z score gene expression level of 86 chromatin-remodeling factors among consensus cluster A and cluster B. Clinicopathological stage, outcome status and groups (cluster C and cluster D) of consensus cluster based on BPTF, CNOT1, and SIN3A were annotated in the heatmap. (D) Comparison of overall prognostic differences between Cluster C and Cluster D upon all stage patients. (E) Comparison of progression-free prognostic differences between Cluster C and Cluster D upon all stage patients.

enrichment results for the intersected genes (Supplementary Figure 2A), which could be considered additional evidence supporting YY1 as the potential transcriptional factor due to the similar pathway annotations. Five E3 ubiquitin ligase genes, CBL, UBE4B, TRAF6, HUWE1, and PIAS1, that were enriched in the ubiquitin-mediated proteolysis pathway identified by the previous interactive KEGG analysis (Supplementary Figure 2A), were predicted to be regulated by YY1. Because ubiquitin E3 ligases have an impact on almost every aspect of eukaryotic biological processes by promoting protein ubiquitination and degradation, it was necessary to predict the possible substrates of CBL, UBE4B, TRAF6, HUWE1 and

PIAS1 to elucidate their mechanisms. Based on five relationship predictive methods provided by the UbiBrowser online tool, some potential targets were found and are listed in Supplementary Figure 3, and many of them are best known as oncoproteins in multiple cancer types, such as proteins in the JAK family, AKT1 and BRAF. As shown in Supplementary Figure 4, OS and PFS analyses with patients stratified by the median cut-off values for the five E3 ubiquitin ligases were also performed for the 496 TCGA-KIRC patients. Obviously, patients with high CBL, UBE4B, TRAF6, HUWE1 or PIAS1 expression exhibited better ccRCC outcomes than those with corresponding low expression.

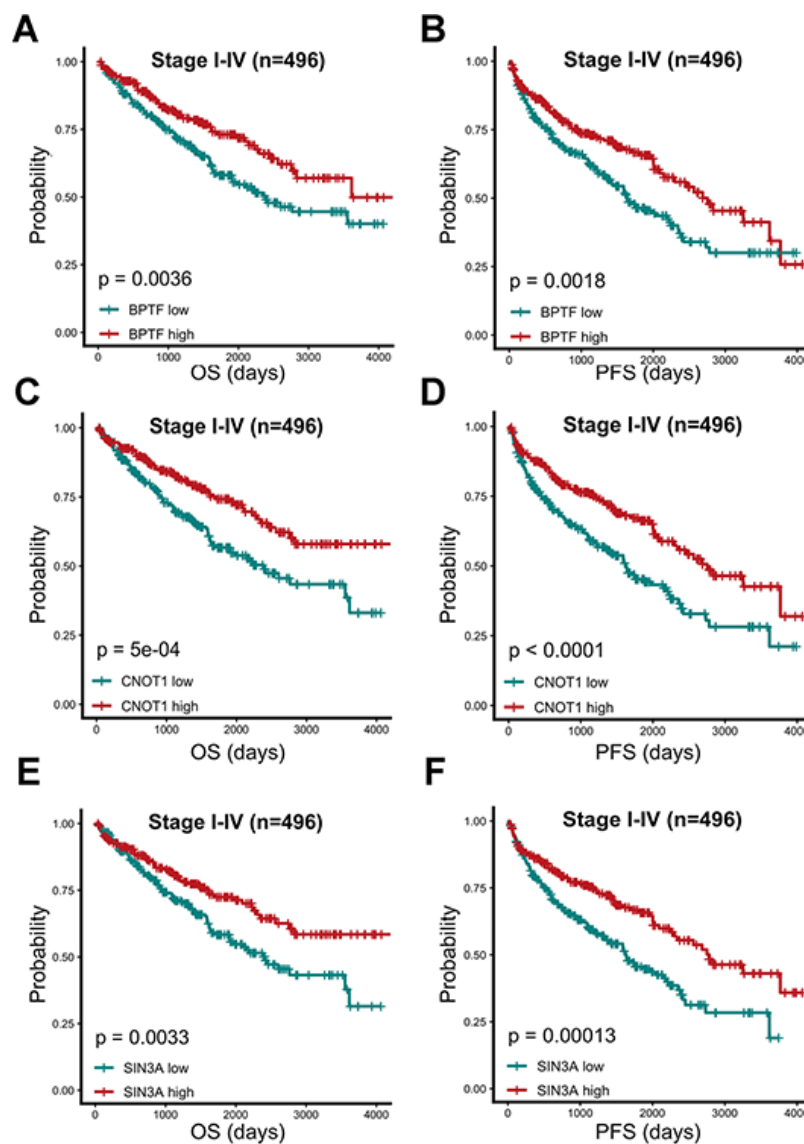


Figure 3. Prognostic analysis of three characterized chromatin-remodeling factors upon 496 TCGA-KIRC patients. Comparison of overall and progression-free prognostic differences between two groups divided by BPTF (A, B), CNOT1 (C, D) and SIN3A (E, F) expression levels based on median expression level.

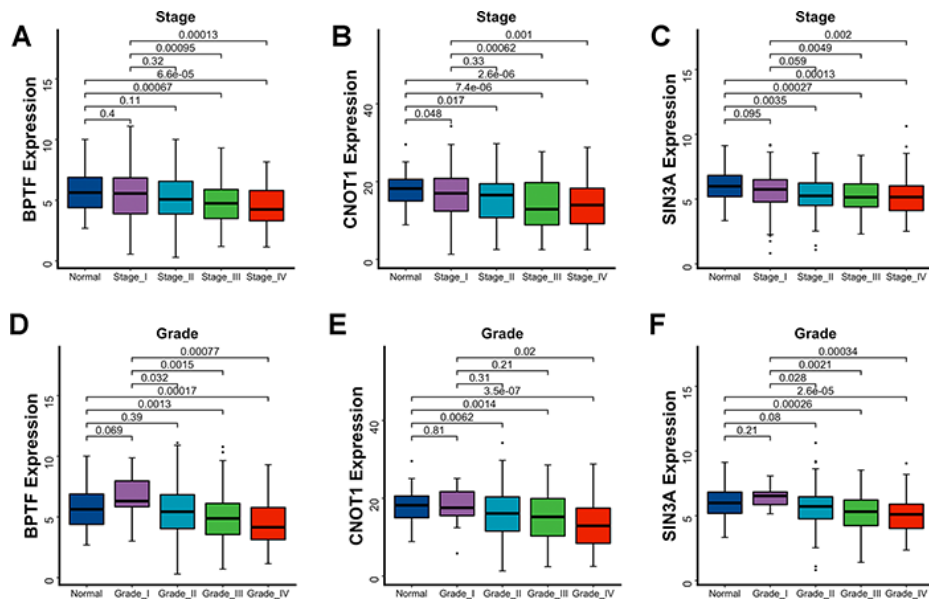


Figure 4. Association between three characterized chromatin-remodeling factors expression levels and clinicopathological parameters. Differentiated expression of BPTF, CNOT1 and SIN3A between normal renal tissue, disparate clinicopathological stages (A–C) and clinicopathological grade (D–F). The Wilcoxon test was used for group comparison.

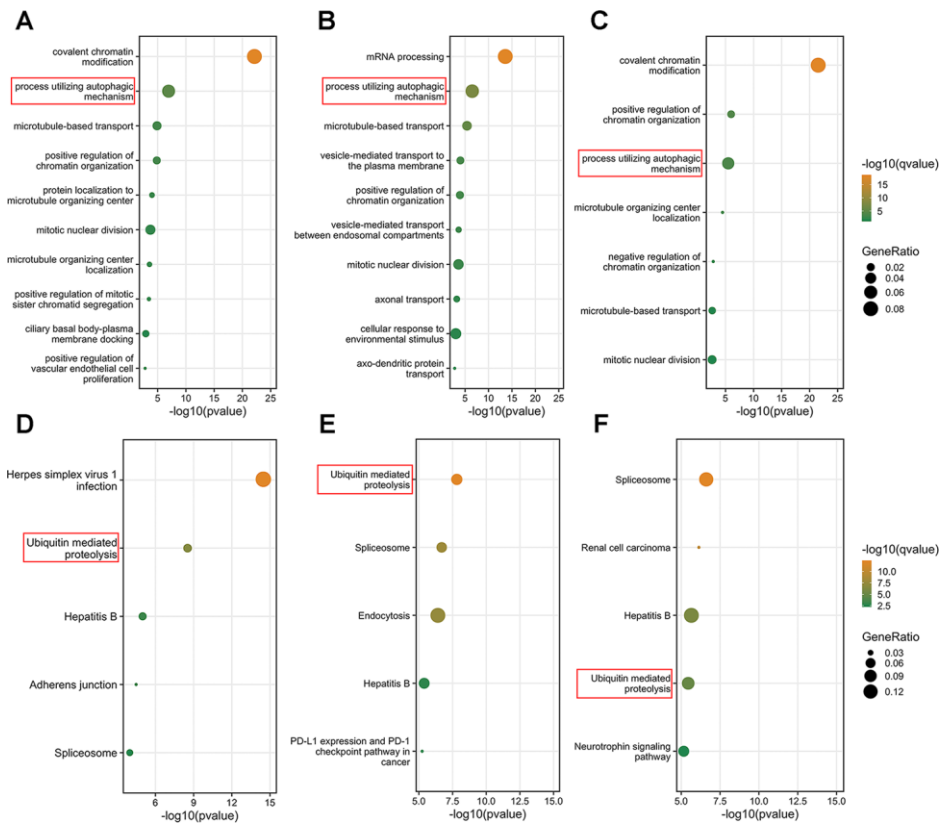


Figure 5. Function prediction of three characterized chromatin-remodeling factors. (A–C) Top 10 biological process enrichment results of genes in which correlation with BPTF, CNOT1 or SIN3A were greater than 0.5. (D–F) Top 5 KEGG pathway enrichment results of genes which correlation with BPTF, CNOT1 and SIN3A were greater than 0.5. The results were ranked by the enrichment scores and displayed corresponding to $-\log_{10}(q\text{-value})$ and gene ratio enriched in the selected pathways.

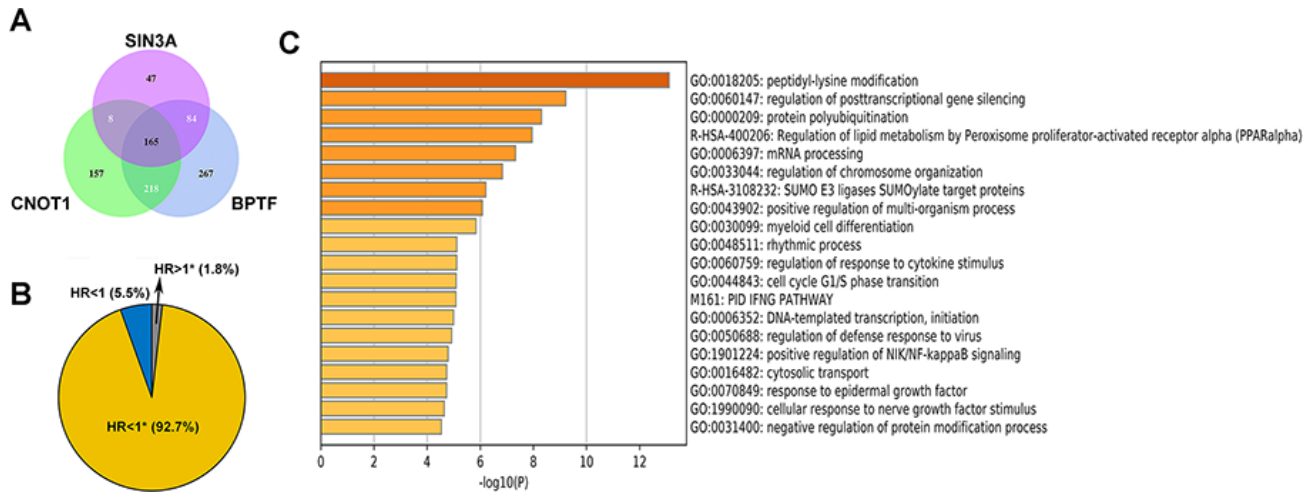


Figure 6. The properties of the intersected genes. (A) Venn diagram shows the number of intersection genes of genes in which correlation with BPTF, CNOT1, and SIN3A was greater than 0.6. (B) Univariate Cox regression results for the overall survival of all 165 intersected genes were displayed by ratio pie plot. "*" means the statistic P-value was significant ($P < 0.05$). (C) Top 20 results of gene ontology analysis upon all 165 intersected genes.

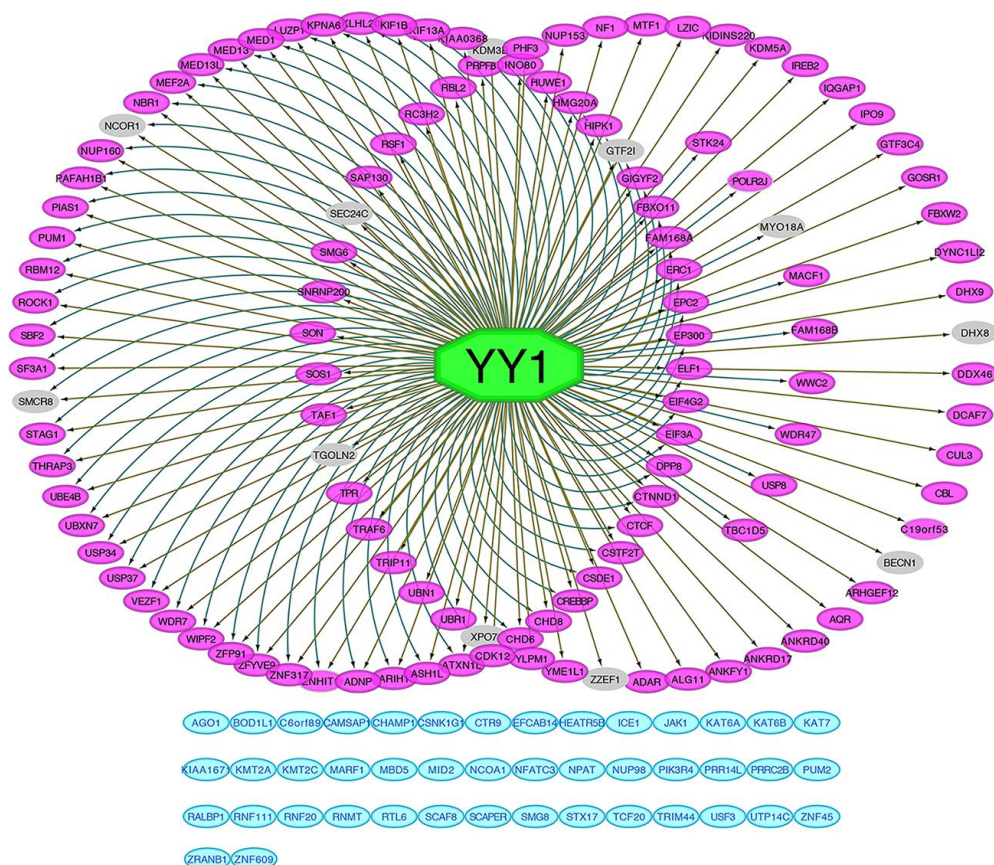


Figure 7. Prediction of YY1 as intersected genes regulator. All 165 intersected genes predictive network of transcription factor YY1 performed by iRegulon. Genes in the left circle were found regulated by YY1 according to both motif's matching and ChIP-seq tracks signals. Genes in the right circle were found regulated by YY1 based on ChIP-seq tracks signals. Genes in the inner right circle were found regulated by YY1 based on motif's matching. Genes in blue modules were not mapped to YY1. Genes in grey color are those that are less than 0.3 correlated with YY1.

Relationship predictions for YY1 and the characterized chromatin remodelers

Another function encoded in iRegulon was exploited to query the targets of BPTF, a transcription-associated chromatin remodeler. According to MSigDB, GeneSigDB and Ganesh Clusters, YY1 and CDKN1A were predicted to be in the top 20 matching results (Figure 8A). Studies have identified CNKN1A as a target gene of YY1 that binds to the BRCA2 protein to have a role in tumor suppression [25]. Through experimental determination and text mining performed with the STRING database, confirmed co-expression relationships were found among the expression of BPTF, SIN3A and YY1 with SIN3A acting as the crosslinking bridge, but CNOT1 did not show any connectivity in the protein network (Figure 8B). One-year, three-year and five-year receiver operating characteristic (ROC) curves of a combined model including clinical stage, BPTF, SIN3A and CNOT1 and YY1 were constructed, and by comparison, this model had better risk-prediction

performance than the previous two models, especially for long-term OS (Supplementary Figure 1A, 1B). The results of Pearson's correlation analyses of BPTF, SIN3A, CNOT1 and YY1 in TCGA-KIRC patients were all positive and significant, and only the coefficient between CNOT1 and YY1 was relatively low (Figure 8C), which was consistent with the protein interaction results. Significantly positive correlations among BPTF, SIN3A, CNOT1 and YY1 could also be acquired with a TCGA pan-cancer analysis performed by Gene Expression Profiling Interactive Analysis (GEPIA) (Supplementary Figure 5); however, their prognostic values, which were consistent in ccRCC, were not as consistent across other cancer types (Supplementary Table 4). Finally, high expression of YY1 was associated with relatively good OS in TCGA-KIRC patients (Figure 8E), which implied that there might be a distinct interaction relationship among BPTF, SIN3A, CNOT1 and YY1 in ccRCC patients that leads to an unclear tumor-suppressive function that promotes improved clinical outcomes.

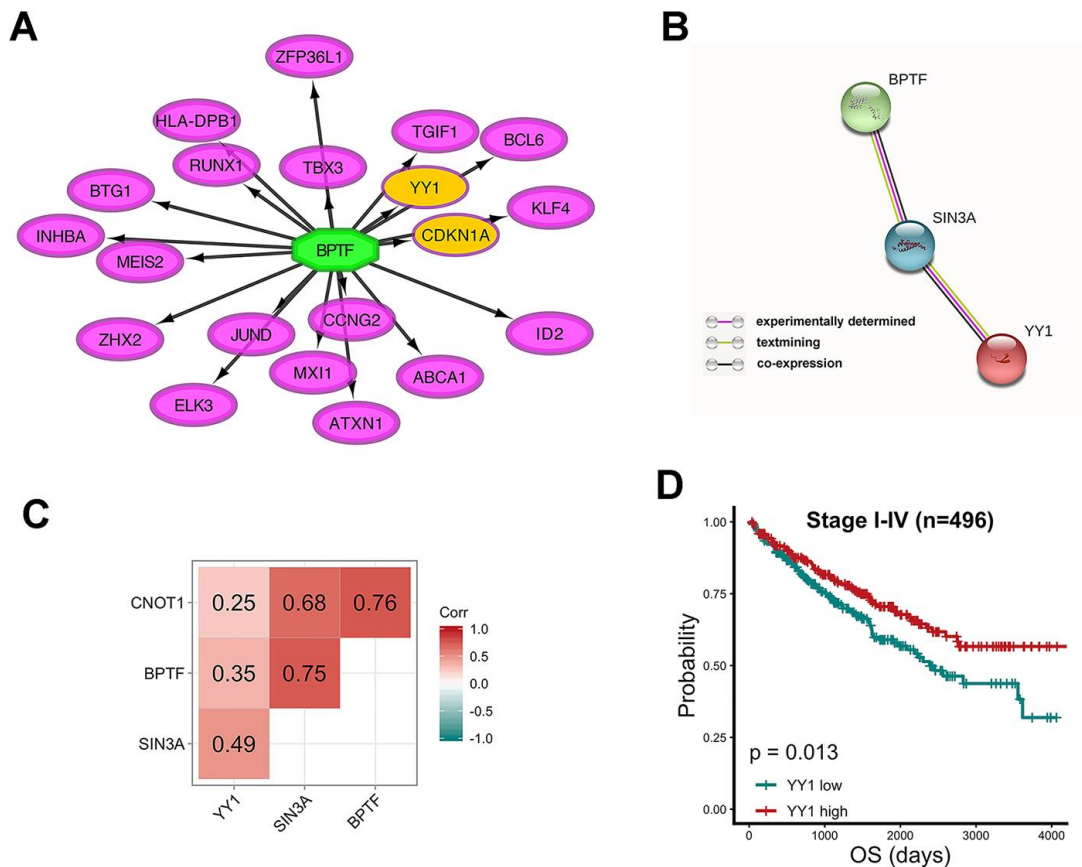


Figure 8. Relationship between characterized chromatin-remodelers and YY1 (A) Top 20 predicted target genes of BPTF. (B) The interactive relationship between BPTF, SIN3A and YY1 produced by STRING. (C) Pearson correlation between the expression levels of BPTF, CNOT1, SIN3A and YY1 upon 496 TCGA-KIRC patients. (D) Comparison of overall prognostic differences between YY1 high and low expression groups by median cut-off value.

Validation in ICGC dataset

We downloaded another RNA-Seq dataset from ICGC which containing 91 patients with renal cancer for validation. Patients in ICGC dataset can be divided into 3 groups with differential expression of BPTF, CNOT1 and SIN3A by consensus clustering. Patients with low BPTF, CNOTA and SIN3A expression have significantly worse OS and no pathways were enriched from KEGG database. Patients with relatively median or high expression of BPTF, CNOT1 and SIN3A suffer better OS and ubiquitin-mediated proteolysis pathway was enriched (NES=1.78, P=0.000) as one of the top 10 enrichment, which was consistent with the result acquired from TCGA (Supplementary Figure 6). Furthermore, patients with high expression level of BPTF and SIN3A have significantly better OS but the result of CNOT1 was insignificant. The expression levels of BPTF, CNOT1 and SIN3A significantly decreased in stage IV patients compared with stage I patients. Since BPTF, CNOT1 and SIN3A were identified as protective factors in the progression of ccRCC in TCGA dataset, that conclusion was validated in ICGC dataset (Supplementary Figure 7). Using iRegulon plugin of cytoscape, YY1 was identified as the most likely predicted transcriptional factor of overlapped correlated genes according to both motifs matching and ChIP-seq tracks signals. Pearson correlation between the expression of BPTF, CNOT1, SIN3A, YY1 and the 5 E3-ligases were significantly high, which suggested closed relationships of them. High expression of CBL and HUWE1 indicated significantly better OS while the similar prognostic tendency could be obtained in PIAS1, TRAF6 and UBE4B but were not significant. The results of transcriptional factor's prediction, correlation analysis and survival analysis in ICGC dataset were consistent with TCGA (Supplementary Figure 8).

DISCUSSION

In this study, based on 86 chromatin-remodeling genes, two subgroups of TCGA-KIRC patients with distinct prognoses were defined by consensus clustering of ccRCC samples. The clustering of three genes (BPTF, SIN3A and CNOT1) could effectively represent 90% of chromatin-remodeling genes and indicate the prognosis of ccRCC. Furthermore, each of the three genes was also identified to be indicative of a good prognosis, and high expression was related to relatively low clinical stage and good histological differentiation in ccRCC. The results of a functional analysis indicated that BPTF, SIN3A and CNOT1 were all involved in ubiquitin-mediated proteolysis in ccRCC. In addition, YY1 was indicated to be an important transcription factor of genes highly related to BPTF, SIN3A and CNOT1.

Correlation and protein network analyses indicated that BPTF, SIN3A and CNOT1 were highly correlated with YY1 and that YY1 might be a regulatory target of BPTF and co-express with SIN3A in ccRCC. Furthermore, another RNA-Seq dataset from ICGC which containing 91 patients with renal cancer was used to validate the results, indicating the results in TCGA were not database-specific. The highlight of our study is the first demonstration that BPTF, SIN3A, CNOT1 and YY1 are novel biomarkers of good prognosis in ccRCC.

BPTF and SIN3A are new prognostic biomarkers for predicting survival in ccRCC. BPTF is the largest subunit of the nucleosome-remodeling factor (NURF) complex and plays a vital role in chromatin remodeling for gene activation through its association with histone acetylation or methylation [20, 26]. The important roles of BPTF in thymocyte maturation, embryonic development, and T cell homeostasis and function have been confirmed [27–30]. Recent studies have indicated that BPTF promotes tumor cell growth in several types of cancers [26, 31–35]. Interestingly, in this paper, BPTF was first identified as a gene positively associated with a good prognosis in ccRCC. SIN3A is a core component of the histone deacetylation activity-associated transcriptional repressor complex [36]. In mammals, Sin3 proteins can recruit histone deacetylases (HDACs) to chromatin-bound transcription factors to repress the expression of target genes [37]. SIN3A regulates processes important for development and homeostasis including mitochondrial biogenesis, cell death and neuronal fate selection [38, 39]. In breast cancer cells, interference with SIN3A function induces epigenetic reprogramming and differentiation, and the SIN3A/HDAC complex plays an important role in maintaining sensitivity to chemotherapy in breast cancer [40, 41]. In this study, we demonstrated for the first time that high SIN3A expression suggests a good prognosis in ccRCC. Consistent with high BPTF or SIN3A expression suggesting a good prognosis in ccRCC, we confirmed that high expression of BPTF or SIN3A was significantly associated with a relatively low clinical stage and good histological differentiation in ccRCC, respectively. According to the TCGA database and existing studies, BPTF and SIN3A in most types of tumors have no prognostic value. Thus, the clear mechanistic understanding of the unique prognostic significance of BPTF and SIN3A in ccRCC needs to be further developed.

In our study, BPTF, SIN3A and CNOT1 were all involved in processes utilizing the autophagic mechanism and ubiquitin-mediated proteolysis in ccRCC. However, how these three genes participate in the process of ubiquitination remains unknown. Ubiquitination is an important posttranslational modification that controls many biological processes,

such as cell division and differentiation, in all eukaryotes, and it also controls many steps in autophagy [42]. Furthermore, autophagy deficiency induces inhibition of histone ubiquitination [43]. These results are consistent with high expression of BPTF, SIN3A and CNOT1 indicating a good prognosis in ccRCC. We predict that BPTF, SIN3A and CNOT1 enhance genomic stability and inhibit tumor cell formation by positively regulating ubiquitination in ccRCC.

YY1 was identified as the intersected gene regulator in the network based on chromatin-remodeling genes highly related to BPTF, SIN3A and CNOT1. In ccRCC, YY1 was identified as a novel biomarker of a good prognosis.

YY1, a ubiquitously expressed transcription factor, plays an essential role in early embryogenesis, adult tissue formation and the regulation of Th2 cell and B cell differentiation [44–47]. However, the mechanism underlying YY1 involvement in ccRCC progression has not been reported. Interestingly, like BPTF, SIN3A and CNOT1, YY1 is also enriched in ubiquitin-mediated proteolysis. Since YY1 is highly correlated with BPTF, SIN3A and CNOT1 in ccRCC, we conclude that YY1, along with the three chromatin-remodeling genes, is involved in the regulation of ubiquitination.

The possible regulatory effects of BPTF and SIN3A on YY1 were investigated via functional and correlation analyses in our study. Five ubiquitin E3 ligases, CBL, UBE4B, TRAF6, HUWE1 and PIAS1, were predicted to be regulated by YY1. Based on five relationship predictive methods provided by the UbiBrowser online tool, some potential targets were found and many of the substrates of these five E3 ubiquitin ligases, such as proteins in the JAK family, AKT1 and BRAF, are well-known oncoproteins in multiple cancer types. In patients with ccRCC, many of these substrates have been identified as poor prognostic factors, such as: JAK2 [48], JAK3 [49], CSK [50], RET [51], AR [52], AHR [53, 54], IRF4 [55], SHC1 [56], TGFBR2 [57]. Interestingly, consistent with the prognostic value of BPTF, SIN3A, CNOT1, and YY1, high expression of any of the five E3 ubiquitin ligases indicated a good prognosis in ccRCC. In addition, BPTF, SIN3A, CNOT1 and YY1 were highly positively correlated in both pan-cancer and ccRCC datasets. We further mined ccRCC datasets in GEO to verify our conclusions in ccRCC. In GSE36895 containing 29 samples from ccRCC patients, positive correlation was found among BPTF/CNOT1/SIN3A/YY1, in which the correlation between SIN3A and CNOT1/YY1 was significant (data not shown) [58]. This result supports the relationship of BPTF/CNOT1/SIN3A/YY1 found in TCGA. Protein network analysis indicated that BPTF might have a targeted regulatory effect on YY1. Confirmed co-expression relationships were found among BPTF, SIN3A

and YY1, with SIN3A acting as the crosslinking bridge. Based on the above results, we predict that BPTF, SIN3A and YY1 are involved in one pathway that is likely associated with ubiquitination in ccRCC. Therefore, we conclude that the good prognostic value of BPTF and SIN3A is caused by their involvement in the regulation of YY1, which further degrades oncoproteins through five E3 ubiquitin ligases in ccRCC.

The Ccr4-Not complex is a highly conserved regulator of mRNA metabolism, and CNOT1 is the large human subunit of the Ccr4-Not complex [59]. CNOT1, a transcriptional repressor, is crucial for maintaining embryonic stem cells in a pluripotent state, and a specific CNOT1 mutation can lead to holoprosencephaly and the novel syndrome of pancreatic agenesis [60]. CNOT1 can target heterochromatin to regulate gene expression and protect genome integrity [61]. In our study, we found that CNOT1 played an important role in chromatin remodeling, and we demonstrated for the first time that high CNOT1 expression suggested a good prognosis in ccRCC. Interestingly, like BPTF, SIN3A and YY1, CNOT1 is also involved in ubiquitin-mediated proteolysis. However, the mechanism underlying CNOT1 involvement in ccRCC needs further study, and we hope that more research will focus on CNOT1 in ccRCC.

To evaluate the reproducibility of the results in TCGA, another additional independent RNA-Seq dataset from ICGC which containing 91 patients with renal cancer was used to compare the results to our preceding findings in TCGA dataset. Significantly, a strong consistency between TCGA and ICGC database was found, suggesting the results in our study were solid and not database-specific.

In conclusion, we identified three chromatin-remodeling genes, BPTF, SIN3A and CNOT1, that could predict a good prognosis, which was related to a relatively low clinical stage and good histological differentiation in ccRCC. We confirmed that BPTF, SIN3A, CNOT1 and YY1 are all involved in ubiquitin-mediated proteolysis. We predicted that BPTF and SIN3A may be involved in regulating YY1, which regulates E3 ubiquitin ligases to promote further degradation of oncoproteins in ccRCC. The limitation of this study is the lack of experimental validation. Further experiments are necessary to validate the prognostic value of BPTF, SIN3A, CNOT1 and YY1 in ccRCC.

MATERIALS AND METHODS

Datasets preparation

Raw RNA-seq count data, fragment per kilobase per million (FPKM) values and the corresponding

clinicopathological features of 611 ccRCC patients were retrieved from the TCGA using TCGAbiolinks R packages version 2.13.6 [62]. After preprocessing to integrate the OS and PFS outcome information obtained from the TCGA clinical data resource [63], patients who met the following criteria were excluded from this study: (1) prior receipt of neoadjuvant therapy; (2) an unclear histopathological definition; (3) missing survival records; and (4) a follow-up period less than 30 days. Thus, a total of 496 tumor cases and 72 normal renal samples satisfying the inclusion criteria were evaluated in the final research. The mutational annotation format (MAF) of somatic mutation data was also downloaded from the TCGA database. A dataset included (reads per kilobase per million) RPKM expression profile of 91 patients with renal cancer and the corresponding clinical information was downloaded from ICGC database (<https://icgc.org/>) for validation.

Clustering of chromatin-remodeling factors

We required a list of all 136 chromatin-remodeling genes from the Human Transcription Factor Database (HumanTFDB; <http://bioinfo.life.hust.edu.cn/HumanTFDB#!/>) after excluding 6 factors in the chromatin Y family to avoid a sex bias and 4 factors in the histone cluster 1 H1, which have relatively low expression levels in renal tumor samples. The paired Wilcoxon test ($P < 0.01$) was applied to compare differential gene expression between renal cancer tissue and nontumoral tissue samples; subsequently, 105 genes were selected as significant factors related to tumorigenesis. We then took the intersection of these genes and the top 50% variant genes were determined by the median absolute deviation (MAD) of FPKM expression data. Finally, 86 chromatin-remodeling genes were accepted as the definitive input for downstream clustering analysis. Based on the K-Means method and Euclidean distance estimation of the gene expression matrix, a consensus clustering algorithm embedded in the R package ConsensusClusterPlus [64] was performed with the following parameters: the clustering number was no more than 10, the proportion of items and features of clusters were no more than 80%, and the number of subsamples was set as 1000 to define the major differentiated groups between samples. The optimal clustering number was identified according to the CDF curve. The same consensus clustering process was performed in TCGA and ICGC datasets after three characterized chromatin remodeling-associated genes were selected.

Characterized gene selection

Machining learning RF wrapper approaches implemented in the R packages caret [65] and Boruta

[66] were applied as feature selection classifiers. Based on the consensus clustering group, we employed a random resampling iteration procedure performed 10 times with 100-fold cross-validations through the caret package on chromatin-remodeling gene expression data to confirm the accuracy of the prediction and to reduce the variability in model performance. By comparing the importance of the original features with randomly accessible shadow-attributed importance through the Boruta package, nonsignificant or low-contribution features were gradually rejected after 30 rounds of tracing regression, while strong or tentatively confirmed features were sustained to stabilize the result of the selection test. After establishing an integrated risk model of clinical stage and selected chromatin-remodeling genes by multivariate Cox regression, we assessed the one- to five-year predictive accuracy of the computed OS prognostic scores by generating ROC curves and comparing their time-dependent areas under the curve (AUCs). The ROC curves were generated with the “timeROC” R package [67].

Survival analysis

OS and PFS were defined as the clinical endpoints to evaluate survival outcomes. For survival analysis (Kaplan-Meier) performed with the survival (<https://CRAN.R-project.org/package=survival>) R package, patients in TCGA and ICGC cohorts were divided into two groups of equal size based on the median expression value of BPTF, CNOT1, SIN3A, YY1, CBL, UBE4B, TRAF6, HUWE1 and PIAS1. The differences between groups were estimated by the log-rank test. Univariate Cox proportional hazard regression was used to test the HR and significance of the prognostic value of the intersected genes. Pan-cancer survival analysis results of TCGA datasets were retrieved from the GEPIA website (<http://gepia.cancer-pku.cn>).

GSEA

To mine possible biological pathways in which the selected chromatin-remodeling genes might be enriched, GSEA was performed. We estimated the correlations between these genes and all other genes by scoring the Pearson correlation coefficient. The genes that presented a relatively high correlation over 0.5 with BPTF, SIN3A and CNOT1 were collected as GO: BP and KEGG input gene sets for subsequent enrichment analysis with the clusterProfiler R package [68]. Java desktop GSEA software (version 4.0.1) was used to evaluate the c2.cp.kegg.v6.1.symbols enriched annotation on YY1 expression by ranking the Pearson's correlation coefficients to produce the gene list and to evaluate the enriched pathways of patients in ICGC dataset with differential expression of BPTF, CNOT1

and SIN3A. To explore the connection between each enriched module, an interactive web-based analytic platform that could construct a visual relational network, NetworkAnalyst (<https://www.networkanalyst.ca/NetworkAnalyst/home.xhtml>), was utilized to visualize the KEGG pathway results for the intersected genes. GO analysis of the intersected genes was carried out with the Metascape online tool (<http://metascape.org>).

Regulator prediction

With Cytoscape Java version 3.5.0 (<https://cytoscape.org>) software, the iRegulon [69] plugin was used to predict the putative transcription factors of target genes and to generate a regulator interactive network. The signature query parameters were determined by default; the enrichment score threshold was 3, and the maximum false discovery rate (FDR) for motif similarity was strictly defined as 0.001. According to both enriched motifs matching from several databases (including JASPAR, HOMER and Elemento) and ENCODE raw ChIP-seq track signals, some significant predictive transcription factors and their calculated targets were filtered to remain and be ranked on the results panel. Genes in which Pearson's correlation with the predicted regulators was smaller than 0.3 were shadowed in the network analysis of TCGA dataset. Based on the MSigDB, GeneSigDB and Ganesh Clusters gene set sources, the target genes of selected transcriptional regulators were queried and then visualized in a network form with the Query TF-target database function.

Protein network construction

The STRING (<https://string-db.org>) database was used to analyze the coexpression relationships between chromatin remodelers and transcription factors in this study. UbiBrowser (<http://ubibrowser.ncpsb.org/ubibrowser/>) is a website application that provides a confident prediction of human ubiquitin ligase (E3) and corresponding substrate interactions based on a naïve Bayesian computational framework. The biological evidence that supports the predicted result is derived and combined from multiple approaches: published data, protein orthologs, protein domains, protein motifs and network topology.

Statistical analysis

The chi-square test and corrected chi-square test (values lower than 5 but larger than 1 prompted calibration for continuity) were applied to intergroup comparisons and to assess associations with clinical features. Pearson's correlation analysis was used to calculate the correlative

relation between selected genes. Univariate and multivariate Cox regression models were created to detect the significant risk and protective factors for patients. The two-sided Wilcoxon rank sum test was performed to measure the differences between two variables. All statistical analyses were conducted with R version 3.6.2, and P values below 0.05 were considered statistically significant.

Abbreviations

ccRCC: clear cell renal cell carcinoma; TCGA-KIRC: The Cancer Genome Atlas-Kidney Renal Cell Carcinoma; HumanTFDB: Human Transcription Factor Database; FPKM: fragment per kilobase per million; RPKM: reads per kilobase per million; MAF: mutational annotation format; MAD: median absolute deviation; AUC: areas under the curve; GEPIA: Gene Expression Profiling Interactive Analysis; FDR: false discovery rate; CDF: cumulative distribution function; GSEA: gene-set enrichment analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes; OS: overall survival; PFS: progression-free survival; COSMIC: Catalogue of Somatic Mutations in Cancer; GO: Gene Ontology; NURF: nucleosome-remodeling factor; ICGC: International Cancer Genome Consortium.

AUTHOR CONTRIBUTIONS

ZL and TL conceived the study. CW and YY collected data. YY, NW, TH, ZS JX and JY participated in statistical analysis. CW and YY wrote the manuscript. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

I would like to extend my sincere thanks to my supervisor, Dr. Tao Liu, a responsible and resourceful scholar, who has helped me finish the manuscript with valuable guidance.

CONFLICTS OF INTEREST

We declare no conflicts of interest for this manuscript.

FUNDING

The study was received funding from National Natural Science Foundation of China (81101938).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; 69:7–34. <https://doi.org/10.3322/caac.21551> PMID:30620402

2. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013; 499:43–49.
<https://doi.org/10.1038/nature12222> PMID:[23792563](https://pubmed.ncbi.nlm.nih.gov/23792563/)
3. Zisman A, Pantuck AJ, Dorey F, Said JW, Shvarts O, Quintana D, Gitlitz BJ, deKernion JB, Figlin RA, Belldegrun AS. Improved prognostication of renal cell carcinoma using an integrated staging system. *J Clin Oncol*. 2001; 19:1649–57.
<https://doi.org/10.1200/JCO.2001.19.6.1649> PMID:[11250993](https://pubmed.ncbi.nlm.nih.gov/11250993/)
4. Li J, Wang W, Zhang Y, Cieřlik M, Guo J, Tan M, Green MD, Wang W, Lin H, Li W, Wei S, Zhou J, Li G, et al. Epigenetic driver mutations in ARID1A shape cancer immune phenotype and immunotherapy. *J Clin Invest*. 2020; 130:2712–26.
<https://doi.org/10.1172/JCI134402> PMID:[32027624](https://pubmed.ncbi.nlm.nih.gov/32027624/)
5. Lichner Z, Scorigas A, White NM, Girgis AH, Rotstein L, Wiegand KC, Latif A, Chow C, Huntsman D, Yousef GM. The chromatin remodeling gene ARID1A is a new prognostic marker in clear cell renal cell carcinoma. *Am J Pathol*. 2013; 182:1163–70.
<https://doi.org/10.1016/j.ajpath.2013.01.007> PMID:[23416164](https://pubmed.ncbi.nlm.nih.gov/23416164/)
6. Hakimi AA, Chen YB, Wren J, Gonen M, Abdel-Wahab O, Heguy A, Liu H, Takeda S, Tickoo SK, Reuter VE, Voss MH, Motzer RJ, Coleman JA, et al. Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *Eur Urol*. 2013; 63:848–54.
<https://doi.org/10.1016/j.eururo.2012.09.005> PMID:[23036577](https://pubmed.ncbi.nlm.nih.gov/23036577/)
7. Laguna MP. Re: clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *J Urol*. 2013; 190:493–94.
<https://doi.org/10.1016/j.juro.2013.04.068> PMID:[23845327](https://pubmed.ncbi.nlm.nih.gov/23845327/)
8. Zheng N, Shabek N. Ubiquitin ligases: structure, function, and regulation. *Annu Rev Biochem*. 2017; 86:129–57.
<https://doi.org/10.1146/annurev-biochem-060815-014922> PMID:[28375744](https://pubmed.ncbi.nlm.nih.gov/28375744/)
9. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, Norton C, Bossé D, Wankowicz SM, Cullen D, Horak C, Wind-Rotolo M, Tracy A, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science*. 2018; 359:801–06.
<https://doi.org/10.1126/science.aan5951> PMID:[29301960](https://pubmed.ncbi.nlm.nih.gov/29301960/)
10. Hargreaves DC, Crabtree GR. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res*. 2011; 21:396–420.
<https://doi.org/10.1038/cr.2011.32> PMID:[21358755](https://pubmed.ncbi.nlm.nih.gov/21358755/)
11. Lai AY, Wade PA. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer*. 2011; 11:588–96.
<https://doi.org/10.1038/nrc3091> PMID:[21734722](https://pubmed.ncbi.nlm.nih.gov/21734722/)
12. Marazzi I, Greenbaum BD, Low DH, Guccione E. Chromatin dependencies in cancer and inflammation. *Nat Rev Mol Cell Biol*. 2018; 19:245–61.
<https://doi.org/10.1038/nrm.2017.113> PMID:[29184195](https://pubmed.ncbi.nlm.nih.gov/29184195/)
13. Janout V, Janoutová G. Epidemiology and risk factors of kidney cancer. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2004; 148:95–101.
PMID:[15523556](https://pubmed.ncbi.nlm.nih.gov/15523556/)
14. Basta J, Rauchman M. The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res*. 2015; 165:36–47.
<https://doi.org/10.1016/j.trsl.2014.05.003> PMID:[24880148](https://pubmed.ncbi.nlm.nih.gov/24880148/)
15. Helming KC, Wang X, Roberts CW. Vulnerabilities of mutant SWI/SNF complexes in cancer. *Cancer Cell*. 2014; 26:309–17.
<https://doi.org/10.1016/j.ccr.2014.07.018> PMID:[25203320](https://pubmed.ncbi.nlm.nih.gov/25203320/)
16. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. *Annu Rev Biochem*. 2009; 78:273–304.
<https://doi.org/10.1146/annurev.biochem.77.062706.153223> PMID:[19355820](https://pubmed.ncbi.nlm.nih.gov/19355820/)
17. Chabanon RM, Morel D, Postel-Vinay S. Exploiting epigenetic vulnerabilities in solid tumors: novel therapeutic opportunities in the treatment of SWI/SNF-defective cancers. *Semin Cancer Biol*. 2020; 61:180–98.
<https://doi.org/10.1016/j.semancer.2019.09.018> PMID:[31568814](https://pubmed.ncbi.nlm.nih.gov/31568814/)
18. Hasan N, Ahuja N. The emerging roles of ATP-dependent chromatin remodeling complexes in pancreatic cancer. *Cancers (Basel)*. 2019; 11:1859.
<https://doi.org/10.3390/cancers11121859> PMID:[31769422](https://pubmed.ncbi.nlm.nih.gov/31769422/)
19. Chromatin-remodeling genes promote immunotherapy resistance. *Cancer Discov*. 2018; 8:OF1.
<https://doi.org/10.1158/2159-8290.CD-NB2018-006> PMID:[29367196](https://pubmed.ncbi.nlm.nih.gov/29367196/)
20. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD. A PHD finger of NURF couples histone H3

- lysine 4 trimethylation with chromatin remodelling. *Nature*. 2006; 442:86–90.
<https://doi.org/10.1038/nature04815> PMID:[16728976](https://pubmed.ncbi.nlm.nih.gov/16728976/)
21. Leung JY, Kim WY. Bap1 and Pbrm1: determinants of tumor grade and mTOR activation in VHL-deficient mouse models of renal cell carcinoma. *Cancer Discov*. 2017; 7:802–04.
<https://doi.org/10.1158/2159-8290.CD-17-0610>
PMID:[28765116](https://pubmed.ncbi.nlm.nih.gov/28765116/)
 22. Niu N, Lu P, Yang Y, He R, Zhang L, Shi J, Wu J, Yang M, Zhang ZG, Wang LW, Gao WQ, Habtezion A, Xiao GG, et al. Loss of Setd2 promotes kras-induced acinar-to-ductal metaplasia and epithelia-mesenchymal transition during pancreatic carcinogenesis. *Gut*. 2020; 69:715–26.
<https://doi.org/10.1136/gutjnl-2019-318362>
PMID:[31300513](https://pubmed.ncbi.nlm.nih.gov/31300513/)
 23. Richart L, Carrillo-de Santa Pau E, Río-Machín A, de Andrés MP, Cigudosa JC, Lobo VJ, Real FX. BPTF is required for c-MYC transcriptional activity and in vivo tumorigenesis. *Nat Commun*. 2016; 7:10153.
<https://doi.org/10.1038/ncomms10153>
PMID:[26729287](https://pubmed.ncbi.nlm.nih.gov/26729287/)
 24. Xue Y, Meehan B, Fu Z, Wang XQ, Fiset PO, Rieker R, Levins C, Kong T, Zhu X, Morin G, Skerritt L, Herpel E, Venneti S, et al. SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. *Nat Commun*. 2019; 10:557.
<https://doi.org/10.1038/s41467-019-08380-1>
PMID:[30718506](https://pubmed.ncbi.nlm.nih.gov/30718506/)
 25. Su J, Sui Y, Ding J, Li F, Shen S, Yang Y, Lu Z, Wang F, Cao L, Liu X, Jin J, Cai Y. Human INO80/YY1 chromatin remodeling complex transcriptionally regulates the BRCA2- and CDKN1A-interacting protein (BCCIP) in cells. *Protein Cell*. 2016; 7:749–60.
<https://doi.org/10.1007/s13238-016-0306-1>
PMID:[27535137](https://pubmed.ncbi.nlm.nih.gov/27535137/)
 26. Zhao X, Zheng F, Li Y, Hao J, Tang Z, Tian C, Yang Q, Zhu T, Diao C, Zhang C, Chen M, Hu S, Guo P, et al. BPTF promotes hepatocellular carcinoma growth by modulating hTERT signaling and cancer stem cell traits. *Redox Biol*. 2019; 20:427–41.
<https://doi.org/10.1016/j.redox.2018.10.018>
PMID:[30419422](https://pubmed.ncbi.nlm.nih.gov/30419422/)
 27. Landry JW, Banerjee S, Taylor B, Aplan PD, Singer A, Wu C. Chromatin remodeling complex NURF regulates thymocyte maturation. *Genes Dev*. 2011; 25:275–86.
<https://doi.org/10.1101/gad.2007311>
PMID:[21289071](https://pubmed.ncbi.nlm.nih.gov/21289071/)
 28. Ma Y, Liu X, Liu Z, Wei S, Shang H, Xue Y, Cao Y, Meng A, Wang Q. The chromatin remodeling protein bptf promotes posterior neuroectodermal fate by enhancing Smad2-activated wnt8a expression. *J Neurosci*. 2015; 35:8493–506.
<https://doi.org/10.1523/JNEUROSCI.0377-15.2015>
PMID:[26041917](https://pubmed.ncbi.nlm.nih.gov/26041917/)
 29. Stankiewicz P, Khan TN, Szafranski P, Slattery L, Streff H, Vetrini F, Bernstein JA, Brown CW, Rosenfeld JA, Rednam S, Scollon S, Bergstrom KL, Parsons DW, et al, and Deciphering Developmental Disorders Study. Haploinsufficiency of the chromatin remodeler BPTF causes syndromic developmental and speech delay, postnatal microcephaly, and dysmorphic features. *Am J Hum Genet*. 2017; 101:503–15.
<https://doi.org/10.1016/j.ajhg.2017.08.014>
PMID:[28942966](https://pubmed.ncbi.nlm.nih.gov/28942966/)
 30. Wu B, Wang Y, Wang C, Wang GG, Wu J, Wan YY. BPTF is essential for T cell homeostasis and function. *J Immunol*. 2016; 197:4325–33.
<https://doi.org/10.4049/jimmunol.1600642>
PMID:[27799308](https://pubmed.ncbi.nlm.nih.gov/27799308/)
 31. Pan Y, Yuan F, Li Y, Wang G, Lin Z, Chen L. Bromodomain PHD-finger transcription factor promotes glioma progression and indicates poor prognosis. *Oncol Rep*. 2019; 41:246–56.
<https://doi.org/10.3892/or.2018.6832> PMID:[30542695](https://pubmed.ncbi.nlm.nih.gov/30542695/)
 32. Dar AA, Majid S, Bezrookove V, Phan B, Ursu S, Nosrati M, De Semir D, Sagebiel RW, Miller JR 3rd, Debs R, Cleaver JE, Kashani-Sabet M. BPTF transduces MITF-driven pro-survival signals in melanoma cells. *Proc Natl Acad Sci USA*. 2016; 113:6254–58.
<https://doi.org/10.1073/pnas.1606027113>
PMID:[27185926](https://pubmed.ncbi.nlm.nih.gov/27185926/)
 33. Li Y, Li J, Luo M, Zhou C, Shi X, Yang W, Lu Z, Chen Z, Sun N, He J. Novel long noncoding RNA NMR promotes tumor progression via NSUN2 and BPTF in esophageal squamous cell carcinoma. *Cancer Lett*. 2018; 430:57–66.
<https://doi.org/10.1016/j.canlet.2018.05.013>
PMID:[29763634](https://pubmed.ncbi.nlm.nih.gov/29763634/)
 34. Gong YC, Liu DC, Li XP, Dai SP. BPTF biomarker correlates with poor survival in human NSCLC. *Eur Rev Med Pharmacol Sci*. 2017; 21:102–07.
PMID:[28121349](https://pubmed.ncbi.nlm.nih.gov/28121349/)
 35. Xiao S, Liu L, Lu X, Long J, Zhou X, Fang M. The prognostic significance of bromodomain PHD-finger transcription factor in colorectal carcinoma and association with vimentin and e-cadherin. *J Cancer Res Clin Oncol*. 2015; 141:1465–74.
<https://doi.org/10.1007/s00432-015-1937-y>
PMID:[25716692](https://pubmed.ncbi.nlm.nih.gov/25716692/)
 36. Zhu F, Zhu Q, Ye D, Zhang Q, Yang Y, Guo X, Liu Z, Jiapaer Z, Wan X, Wang G, Chen W, Zhu S, Jiang C, et al. Sin3a-Tet1 interaction activates gene transcription and

- is required for embryonic stem cell pluripotency. *Nucleic Acids Res.* 2018; 46:6026–40.
<https://doi.org/10.1093/nar/gky347> PMID:[29733394](https://pubmed.ncbi.nlm.nih.gov/29733394/)
37. Hassig CA, Fleischer TC, Billin AN, Schreiber SL, Ayer DE. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell.* 1997; 89:341–47.
[https://doi.org/10.1016/s0092-8674\(00\)80214-7](https://doi.org/10.1016/s0092-8674(00)80214-7) PMID:[9150133](https://pubmed.ncbi.nlm.nih.gov/9150133/)
 38. Dannenberg JH, David G, Zhong S, van der Torre J, Wong WH, Depinho RA. mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes Dev.* 2005; 19:1581–95.
<https://doi.org/10.1101/gad.1286905> PMID:[15998811](https://pubmed.ncbi.nlm.nih.gov/15998811/)
 39. Pile LA, Spellman PT, Katzenberger RJ, Wassarman DA. The SIN3 deacetylase complex represses genes encoding mitochondrial proteins: implications for the regulation of energy metabolism. *J Biol Chem.* 2003; 278:37840–48.
<https://doi.org/10.1074/jbc.M305996200> PMID:[12865422](https://pubmed.ncbi.nlm.nih.gov/12865422/)
 40. Yang Y, Huang W, Qiu R, Liu R, Zeng Y, Gao J, Zheng Y, Hou Y, Wang S, Yu W, Leng S, Feng D, Wang Y. LSD1 coordinates with the SIN3A/HDAC complex and maintains sensitivity to chemotherapy in breast cancer. *J Mol Cell Biol.* 2018; 10:285–301.
<https://doi.org/10.1093/jmcb/miy021> PMID:[29741645](https://pubmed.ncbi.nlm.nih.gov/29741645/)
 41. Farias EF, Petrie K, Leibovitch B, Murtagh J, Chornet MB, Schenk T, Zelent A, Waxman S. Interference with Sin3 function induces epigenetic reprogramming and differentiation in breast cancer cells. *Proc Natl Acad Sci USA.* 2010; 107:11811–16.
<https://doi.org/10.1073/pnas.1006737107> PMID:[20547842](https://pubmed.ncbi.nlm.nih.gov/20547842/)
 42. Grumati P, Dikic I. Ubiquitin signaling and autophagy. *J Biol Chem.* 2018; 293:5404–13.
<https://doi.org/10.1074/jbc.TM117.000117> PMID:[29187595](https://pubmed.ncbi.nlm.nih.gov/29187595/)
 43. Wang Y, Zhu WG, Zhao Y. Autophagy substrate SQSTM1/p62 regulates chromatin ubiquitination during the DNA damage response. *Autophagy.* 2017; 13:212–13.
<https://doi.org/10.1080/15548627.2016.1245262> PMID:[27791533](https://pubmed.ncbi.nlm.nih.gov/27791533/)
 44. Vella P, Barozzi I, Cuomo A, Bonaldi T, Pasini D. Yin yang 1 extends the myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res.* 2012; 40:3403–18.
<https://doi.org/10.1093/nar/gkr1290> PMID:[22210892](https://pubmed.ncbi.nlm.nih.gov/22210892/)
 45. Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, Lachanski CV, Gillis DR, Phillips-Cremins JE. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* 2017; 27:1139–52.
<https://doi.org/10.1101/gr.215160.116> PMID:[28536180](https://pubmed.ncbi.nlm.nih.gov/28536180/)
 46. Hwang SS, Kim YU, Lee S, Jang SW, Kim MK, Koh BH, Lee W, Kim J, Souabni A, Busslinger M, Lee GR. Transcription factor YY1 is essential for regulation of the Th2 cytokine locus and for Th2 cell differentiation. *Proc Natl Acad Sci USA.* 2013; 110:276–81.
<https://doi.org/10.1073/pnas.1214682110> PMID:[23248301](https://pubmed.ncbi.nlm.nih.gov/23248301/)
 47. Kleiman E, Jia H, Loguercio S, Su AI, Feeney AJ. YY1 plays an essential role at all stages of b-cell differentiation. *Proc Natl Acad Sci USA.* 2016; 113:E3911–20.
<https://doi.org/10.1073/pnas.1606297113> PMID:[27335461](https://pubmed.ncbi.nlm.nih.gov/27335461/)
 48. Kang MA, Lee J, Ha SH, Lee CM, Kim KM, Jang KY, Park SH. Interleukin4Rα (IL4Rα) and IL13Rα1 are associated with the progress of renal cell carcinoma through janus kinase 2 (JAK2)/forkhead box O3 (FOXO3) pathways. *Cancers (Basel).* 2019; 11:1394.
<https://doi.org/10.3390/cancers11091394> PMID:[31540495](https://pubmed.ncbi.nlm.nih.gov/31540495/)
 49. de Martino M, Gigante M, Cormio L, Praticchizzo C, Cavalcanti E, Gigante M, Ariano V, Netti GS, Montemurro E, Mancini V, Battaglia M, Gesualdo L, Carrieri G, Ranieri E. JAK3 in clear cell renal cell carcinoma: mutational screening and clinical implications. *Urol Oncol.* 2013; 31:930–37.
<https://doi.org/10.1016/j.urolonc.2011.07.001> PMID:[21868263](https://pubmed.ncbi.nlm.nih.gov/21868263/)
 50. Ghatalia P, Yang ES, Lasseigne BN, Ramaker RC, Cooper SJ, Chen D, Sudarshan S, Wei S, Guru AS, Zhao A, Cooper T, Della Manna DL, Naik G, et al. Kinase gene expression profiling of metastatic clear cell renal cell carcinoma tissue identifies potential new therapeutic targets. *PLoS One.* 2016; 11:e0160924.
<https://doi.org/10.1371/journal.pone.0160924> PMID:[27574806](https://pubmed.ncbi.nlm.nih.gov/27574806/)
 51. Wang L, Zhang Y, Gao Y, Fan Y, Chen L, Liu K, Meng Q, Zhao C, Ma X. Prognostic and predictive values of subcellular localisation of RET in renal clear-cell carcinoma. *Dis Markers.* 2016; 2016:6870470.
<https://doi.org/10.1155/2016/6870470> PMID:[27092013](https://pubmed.ncbi.nlm.nih.gov/27092013/)
 52. Wang K, Sun Y, Tao W, Fei X, Chang C. Androgen receptor (AR) promotes clear cell renal cell carcinoma (ccRCC) migration and invasion via altering the circHIAT1/miR-195-5p/29a-3p/29c-3p/CDC42 signals. *Cancer Lett.* 2017; 394:1–12.
<https://doi.org/10.1016/j.canlet.2016.12.036> PMID:[28089832](https://pubmed.ncbi.nlm.nih.gov/28089832/)

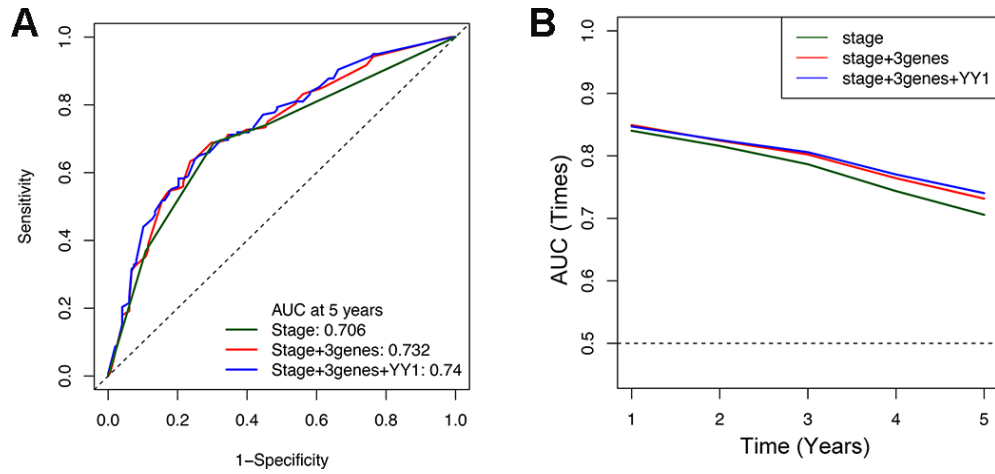
53. Ishida M, Mikami S, Shinojima T, Kosaka T, Mizuno R, Kikuchi E, Miyajima A, Okada Y, Oya M. Activation of aryl hydrocarbon receptor promotes invasion of clear cell renal cell carcinoma and is associated with poor prognosis and cigarette smoke. *Int J Cancer*. 2015; 137:299–310.
<https://doi.org/10.1002/ijc.29398> PMID:25523818
54. Butz H, Szabó PM, Nofech-Mozes R, Rotondo F, Kovacs K, Mirham L, Girgis H, Boles D, Patocs A, Yousef GM. Integrative bioinformatics analysis reveals new prognostic biomarkers of clear cell renal cell carcinoma. *Clin Chem*. 2014; 60:1314–26.
<https://doi.org/10.1373/clinchem.2014.225854> PMID:25139457
55. Dannenmann SR, Thielicke J, Stöckli M, Matter C, von Boehmer L, Cecconi V, Hermanns T, Hefermehl L, Schraml P, Moch H, Knuth A, van den Broek M. Tumor-associated macrophages subvert t-cell function and correlate with reduced survival in clear cell renal cell carcinoma. *Oncoimmunology*. 2013; 2:e23562.
<https://doi.org/10.4161/onci.23562> PMID:23687622
56. Zhao Y, Wang Y, Zhao E, Tan Y, Geng B, Kang C, Li X. PTRF/CAVIN1, regulated by SHC1 through the EGFR pathway, is found in urine exosomes as a potential biomarker of ccRCC. *Carcinogenesis*. 2020; 41:274–83.
<https://doi.org/10.1093/carcin/bgz147> PMID:31605605
57. Parker AS, Lohse CM, Wu K, Kreinest P, Copland JA, Hilton T, Wehle M, Cheville JC, Blute M. Lower expression levels of the transforming growth factor beta receptor type II protein are associated with a less aggressive tumor phenotype and improved survival among patients with clear cell renal cell carcinoma. *Hum Pathol*. 2007; 38:453–61.
<https://doi.org/10.1016/j.humpath.2006.08.029> PMID:17188329
58. Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, Leng N, Pavia-Jiménez A, Wang S, Yamasaki T, Zhrebker L, Sivanand S, Spence P, Kinch L, Hambuch T, Jain S, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet*. 2012; 44:751–59.
<https://doi.org/10.1038/ng.2323> PMID:22683710
59. Winkler GS, Mulder KW, Bardwell VJ, Kalkhoven E, Timmers HT. Human Ccr4-not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. *EMBO J*. 2006; 25:3089–99.
<https://doi.org/10.1038/sj.emboj.7601194> PMID:16778766
60. De Franco E, Watson RA, Weninger WJ, Wong CC, Flanagan SE, Caswell R, Green A, Tudor C, Lelliott CJ, Geyer SH, Maurer-Gesek B, Reissig LF, Lango Allen H, et al. A specific CNOT1 mutation results in a novel syndrome of pancreatic agenesis and holoprosencephaly through impaired pancreatic and neurological development. *Am J Hum Genet*. 2019; 104:985–89.
<https://doi.org/10.1016/j.ajhg.2019.03.018> PMID:31006513
61. Sugiyama T, Thillainadesan G, Chalamcharla VR, Meng Z, Balachandran V, Dhakshnamoorthy J, Zhou M, Grewal SI. Enhancer of rudimentary cooperates with conserved RNA-processing factors to promote meiotic mRNA decay and facultative heterochromatin assembly. *Mol Cell*. 2016; 61:747–59.
<https://doi.org/10.1016/j.molcel.2016.01.029> PMID:26942678
62. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44:e71.
<https://doi.org/10.1093/nar/gkv1507> PMID:26704973
63. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, et al, and Cancer Genome Atlas Research Network. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018; 173:400–16.e11.
<https://doi.org/10.1016/j.cell.2018.02.052> PMID:29625055
64. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010; 26:1572–73.
<https://doi.org/10.1093/bioinformatics/btq170> PMID:20427518
65. Kuhn M. (2020). caret: Classification and Regression Training.
66. Kursa MB, Rudnicki WR. Feature Selection with the {Boruta} Package. *Journal of Statistical Software*. 2010; 36:1–13.
<https://doi.org/10.18637/jss.v036.i11>
67. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013; 32:5381–97.
<https://doi.org/10.1002/sim.5958> PMID:24027076
68. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–87.
<https://doi.org/10.1089/omi.2011.0118> PMID:22455463

69. Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, et al. iRegulon: from a gene list to a

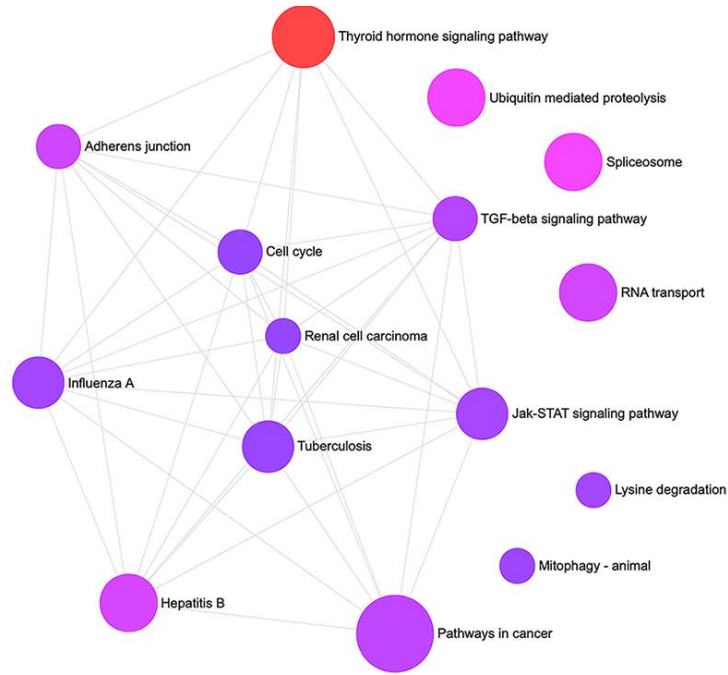
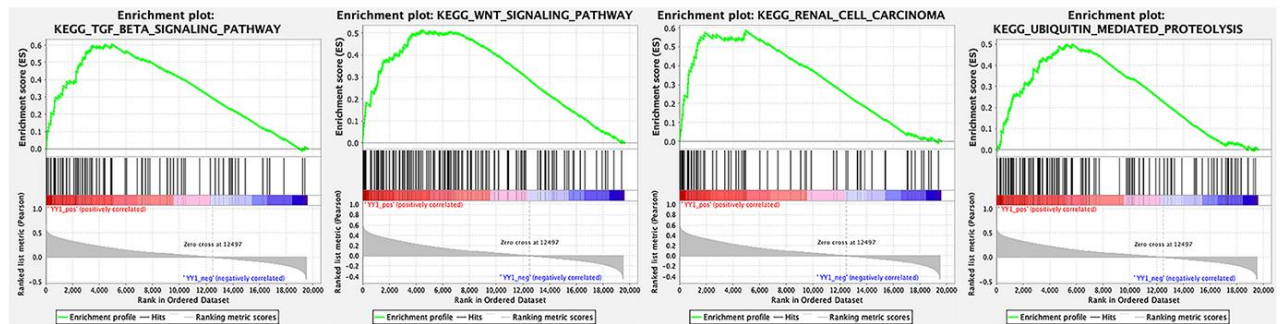
gene regulatory network using large motif and track collections. PLoS Comput Biol. 2014; 10:e1003731.
<https://doi.org/10.1371/journal.pcbi.1003731>
PMID:[25058159](https://pubmed.ncbi.nlm.nih.gov/25058159/)

SUPPLEMENTARY MATERIALS

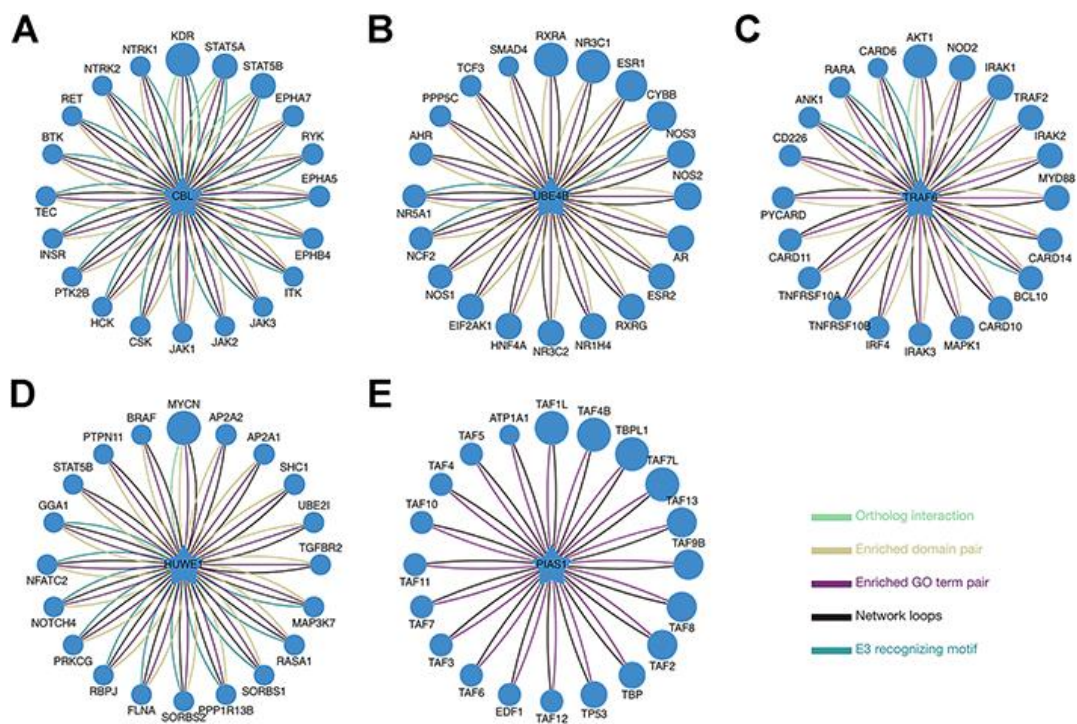
Supplementary Figures



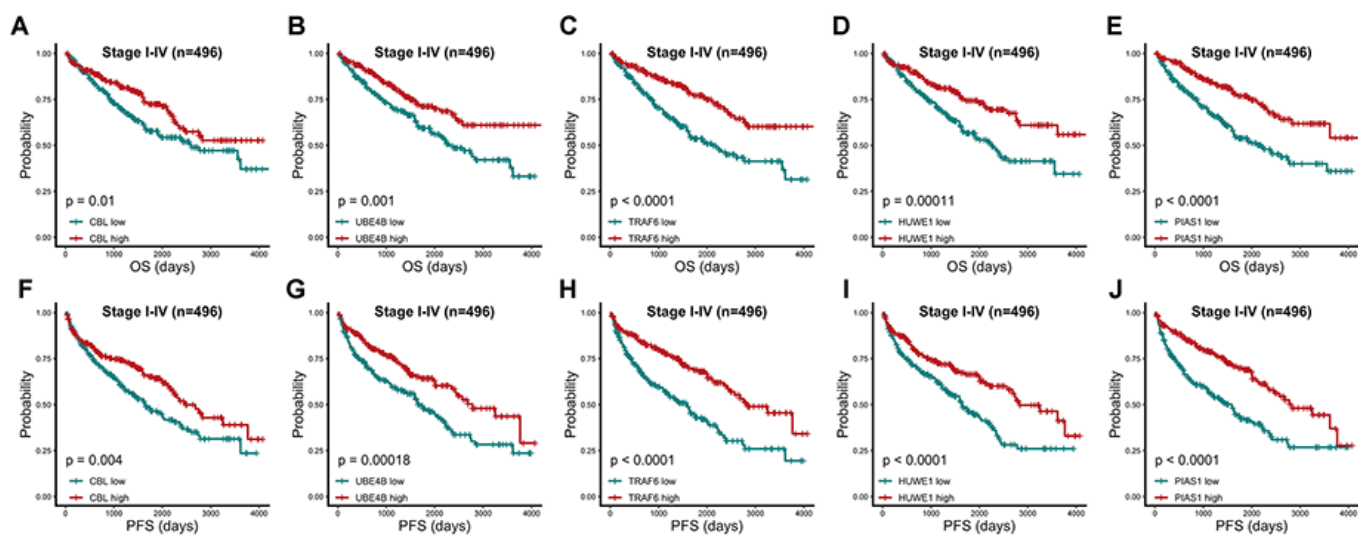
Supplementary Figure 1. Predicted overall survival from one to five years of different models. (A) AUC performance of five years survival among clinical stage model, clinical stage+BPTF+CNOT1+SIN3A model and clinical stage+ BPTF+CNOT1+SIN3A+YY1 model. **(B)** Comparison of time-dependent ROC curves of three different models.

A**B**

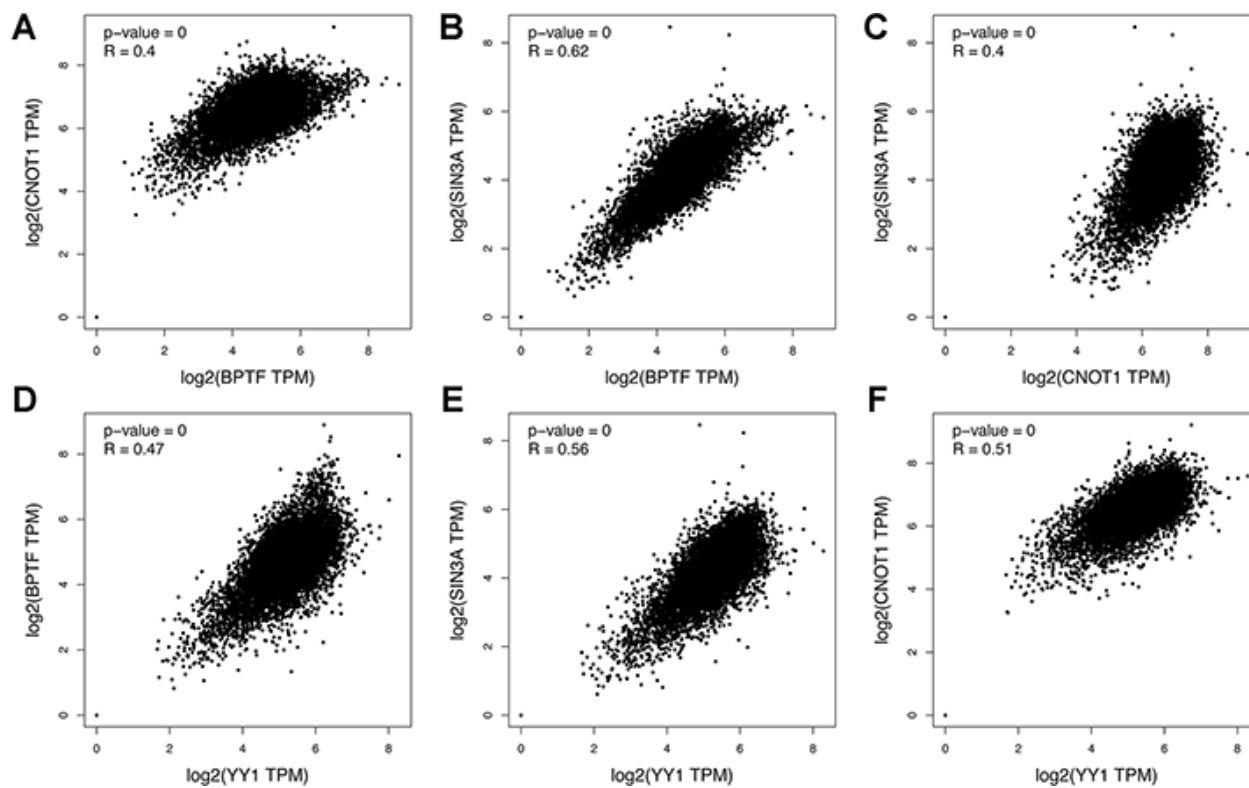
Supplementary Figure 2. GSEA of intersected genes and YY1. (A) An interactive network of the top 15 KEGG enriched pathways upon 165 intersected genes. **(B)** Four out of the top 10 KEGG pathway enrichment results of YY1 single gene GSEA.



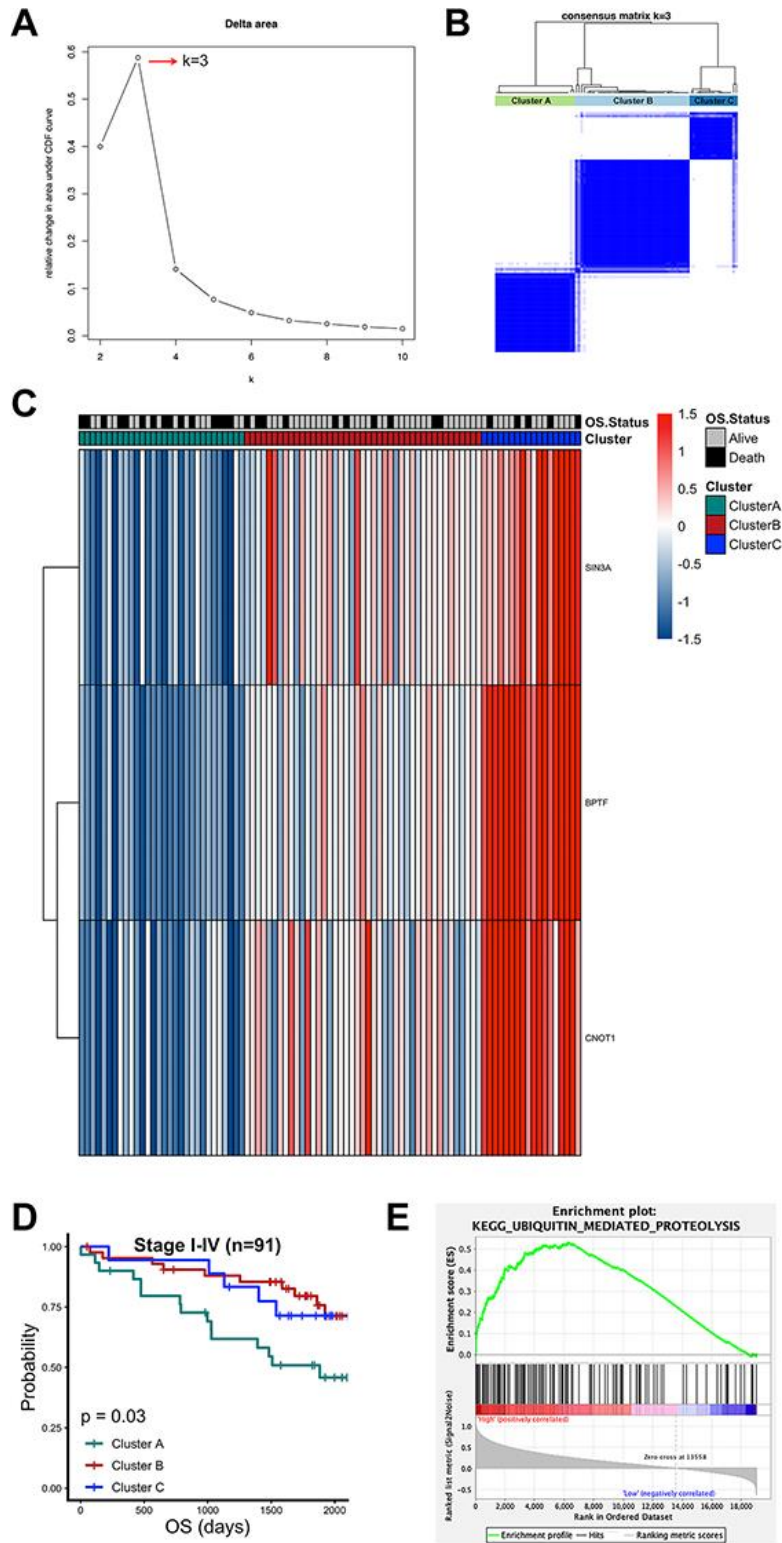
Supplementary Figure 3. Target prediction of E3 ubiquitin ligases. (A–E) Possible substrates of the five identified E3 ubiquitin ligases predicted by multiple approaches. The size of the target stands for the confidence of their interaction with E3.



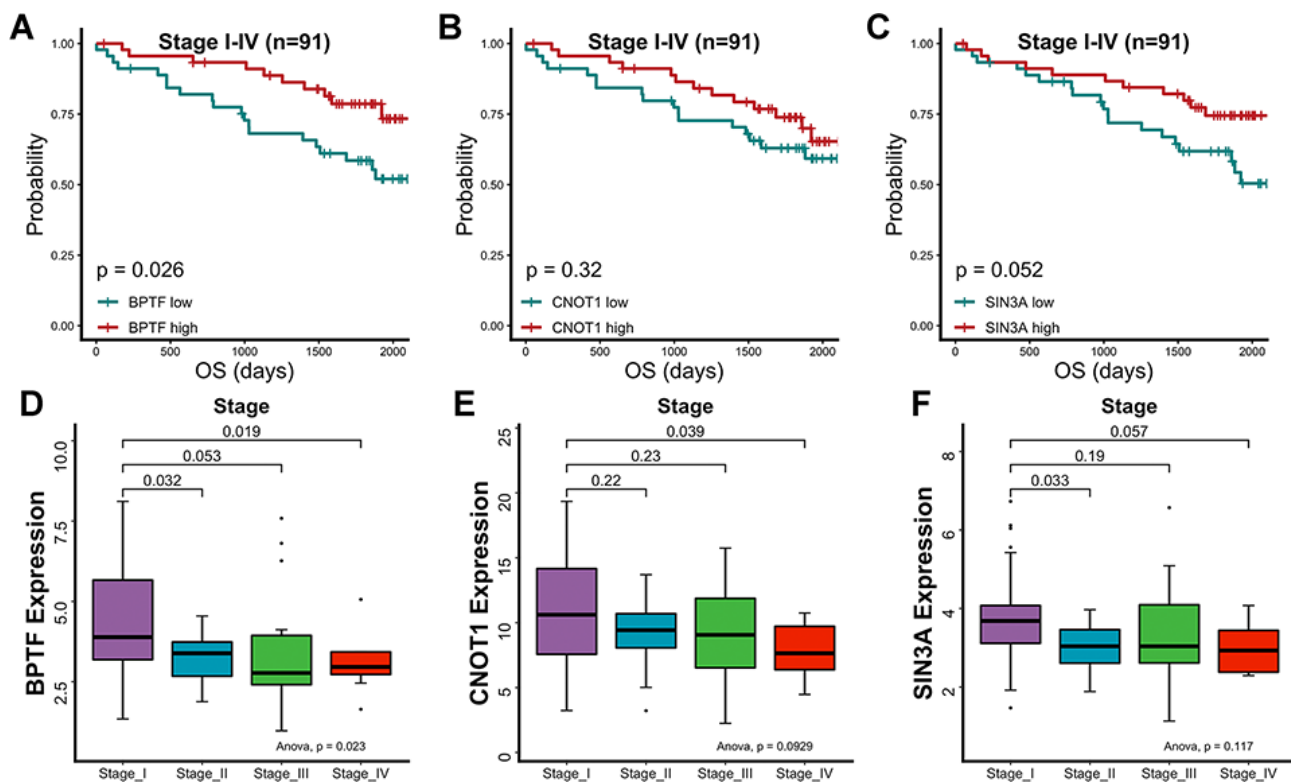
Supplementary Figure 4. Survival analysis of E3 ubiquitin ligases. (A–J) Overall survival outcomes of 5 E3 ubiquitin ligases upon median cut-off values.



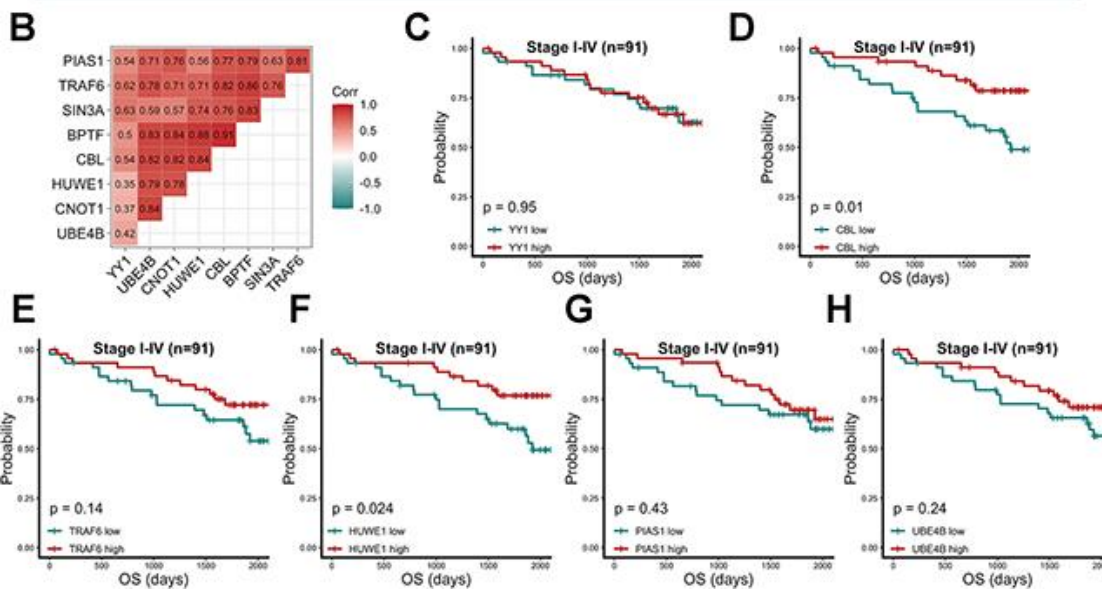
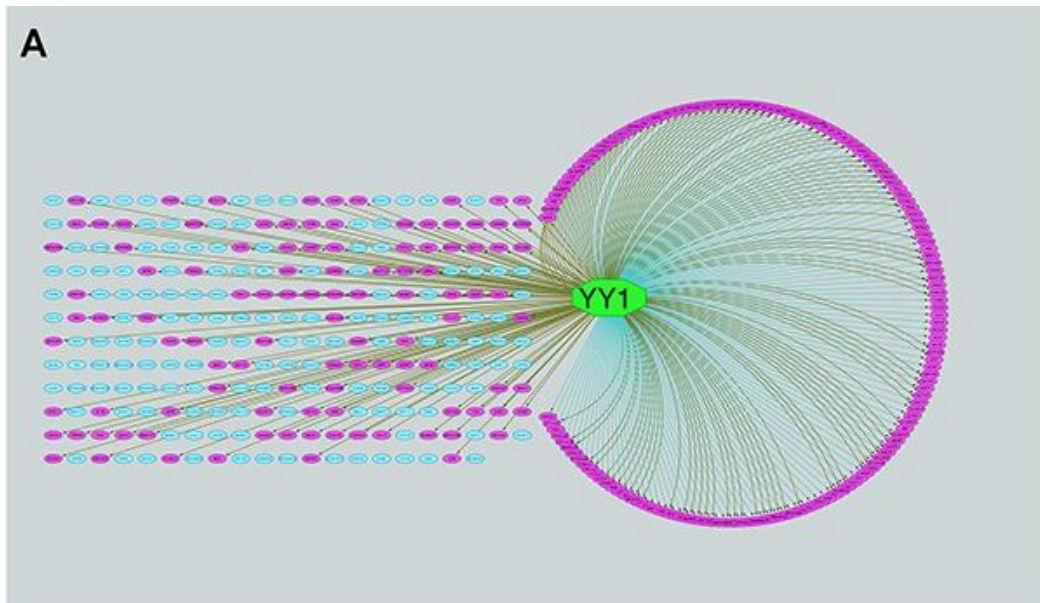
Supplementary Figure 5. TCGA pan-cancer correlation exploration. (A–F) Pan-cancer Pearson correlation between $\log_2(\text{TPM})$ expression of BPTF, CNOT1, SIN3A and YY1 performed by GEPIA.



Supplementary Figure 6. ICGC patients grouping based on the expression of BPTF, CNOT1 and SIN3A. (A, B) Consensus matrix among CDF curves increase from 2 to 10 clusters and groups acquisition. **(C)** The comparison of the expression level of BPTF, CNOT1 and SIN3A among groups. **(D)** Comparison of overall survival outcome among groups. **(E)** KEGG pathways enrichment of patients in cluster B and cluster C.



Supplementary Figure 7. Validation of the prognostic values and clinicopathological parameters comparison of BPTF, CNOT1 and SIN3A in ICGC dataset. (A–C) Validation of the overall survival. (D–F) Expression level among different stage.



Supplementary Figure 8. Validation of transcriptional factor prediction of correlated genes and prognostic value of E3 ubiquitin ligases in ICGC dataset. (A) Transcriptional factor prediction of 398 overlapped correlated genes (Pearson $r > 0.6$, $p < 0.001$) of BPTF, CNOT1 and SIN3A. **(B)** Pearson correlation analysis between BPTF, CNOT1, SIN3A, YY1 and 5 E3 ubiquitin ligases. **(C–H)** Overall survival analysis of YY1, CBL, TRAF6, HUWE1, PIAS1 and UBE4B.

Supplementary Tables

Supplementary Table 1. Association between Cluster A, Cluster B, clinicopathological parameters and common renal carcinoma genomic mutation characteristics in TCGA-KIRC (n=496).

	Cluster A (n=206)	Cluster B (n=290)	Total (n=496)	P value
Age				
Mean±SD	61.4 (11.7)	59.8 (12.3)	60.5 (12.1)	0.175
Gender				
Female	67 (32.5%)	106 (36.6%)	173 (34.9%)	0.406
Male	139 (67.5%)	184 (63.4%)	323 (65.1%)	
Stage				
I	84 (40.8%)	166 (57.2%)	250 (50.4%)	0.002*
II	22 (10.7%)	30 (10.3%)	52 (10.5%)	
III	60 (29.1%)	58 (20.0%)	118 (23.8%)	
IV	40 (19.4%)	36 (12.4%)	76 (15.3%)	
Grade				
I-II	72 (35.0%)	149 (51.4%)	221 (44.6%)	3e-04*
III-IV	131 (63.6%)	136 (46.9%)	267 (53.8%)	
VHL status				
Wild-type	62 (35.0%)	95 (51.4%)	157 (44.6%)	0.652
Mutant	56 (63.6%)	98 (46.9%)	154 (53.8%)	
PBRM1 status				
Wild-type	69 (33.5%)	108 (37.2%)	177 (35.7%)	0.751
Mutant	49 (23.8%)	85 (29.3%)	134 (27.0%)	
TP53 status				
Wild-type	113 (54.9%)	191 (65.9%)	304 (61.3%)	0.146
Mutant	5 (2.4%)	2 (0.7%)	7 (1.4%)	
MTOR status				
Wild-type	107 (51.9%)	184 (63.4%)	291 (58.7%)	0.165
Mutant	11 (5.3%)	9 (3.1%)	20 (4.0%)	

Supplementary Table 2. Referenced model accuracy corresponding to the number of variables included in features selection.

Variables	Accuracy	Kappa	AccuracySD	KappaSD
2	0.885	0.76685814	0.15621254	0.31530365
3	0.90183333	0.79766067	0.14147878	0.28676197
4	0.91633333	0.8291625	0.11966986	0.23714195
5	0.91233333	0.81825341	0.12669989	0.26149403
6	0.927	0.85192641	0.11011625	0.22121793
7	0.95016667	0.89943057	0.09585972	0.19280668
8	0.956	0.91039794	0.09246348	0.1873606
9	0.94566667	0.88912754	0.09456677	0.19169125
10	0.95366667	0.9039627	0.08562768	0.17805629
11	0.9695	0.93625874	0.07313161	0.15326379
12	0.9695	0.93695804	0.07313161	0.15161329
13	0.9655	0.92932401	0.08186432	0.16845566
14	0.967	0.93201632	0.08140806	0.16819983
15	0.969	0.93516317	0.07970273	0.16676898
16	0.9695	0.93631702	0.07846218	0.16410564
17	0.9715	0.94086247	0.07662548	0.15939759
18	0.9735	0.94465035	0.06906848	0.14463067
19	0.9675	0.93317016	0.08020649	0.16558167
20	0.9735	0.94395105	0.06906848	0.14639705
21	0.9715	0.93946387	0.07662548	0.1625547
22	0.9695	0.93631702	0.07846218	0.16410564
23	0.9695	0.93631702	0.07846218	0.16410564
24	0.9695	0.93631702	0.07846218	0.16410564
25	0.96833333	0.93337995	0.07299509	0.15425558
26	0.9735	0.94395105	0.06906848	0.14639705
27	0.9735	0.94395105	0.06906848	0.14639705
28	0.978	0.9527972	0.06289321	0.1354076
29	0.9775	0.95164336	0.0645008	0.13876189
30	0.9755	0.9477972	0.0668539	0.14268294
31	0.978	0.9527972	0.06289321	0.1354076

32	0.9755	0.9477972	0.0668539	0.14268294
33	0.9735	0.94400932	0.07468959	0.15770879
34	0.974	0.94370629	0.06759953	0.14682918
35	0.978	0.9527972	0.06289321	0.1354076
36	0.96783333	0.93228438	0.07957008	0.16766191
37	0.9715	0.93940559	0.07115746	0.15160319
38	0.97183333	0.94061772	0.07042022	0.14888174
39	0.9755	0.9477972	0.0668539	0.14268294
40	0.974	0.94446387	0.07333333	0.15647684
41	0.978	0.9527972	0.06289321	0.1354076
42	0.9755	0.94785548	0.07264658	0.1542658
43	0.974	0.94440559	0.06759953	0.14506925
44	0.9755	0.9477972	0.0668539	0.14268294
45	0.976	0.94895105	0.06531973	0.13945516
46	0.978	0.9534965	0.06289321	0.13344907
47	0.976	0.94895105	0.06531973	0.13945516
48	0.974	0.94370629	0.06759953	0.14682918
49	0.976	0.94900932	0.07123726	0.1512849
50	0.976	0.94825175	0.06531973	0.14130776
51	0.972	0.93991841	0.07531751	0.16137269
52	0.9715	0.93870629	0.07115746	0.15326508
53	0.9735	0.94325175	0.06906848	0.14813903
54	0.9715	0.93940559	0.07115746	0.15160319
55	0.9675	0.93037296	0.08020649	0.1714723
56	0.9735	0.94325175	0.06906848	0.14813903
57	0.976	0.94825175	0.06531973	0.14130776
58	0.972	0.93916084	0.06974702	0.15201314
59	0.978	0.9527972	0.06289321	0.1354076
60	0.9715	0.93800699	0.07115746	0.15490595
61	0.9695	0.93416084	0.07313161	0.15809319
62	0.9735	0.94400932	0.07468959	0.15770879
63	0.976	0.94900932	0.07123726	0.1512849
64	0.978	0.9527972	0.06289321	0.1354076
65	0.9775	0.95234266	0.0645008	0.13685734
66	0.974	0.94370629	0.06759953	0.14682918
67	0.976	0.94895105	0.06531973	0.13945516
68	0.976	0.94825175	0.06531973	0.14130776
69	0.9675	0.92961538	0.075	0.16264989
70	0.976	0.94755245	0.06531973	0.14313293
71	0.98	0.95664336	0.06030227	0.13112135
72	0.9715	0.93870629	0.07115746	0.15326508
73	0.976	0.94825175	0.06531973	0.14130776
74	0.978	0.95285548	0.0690191	0.14756065
75	0.98	0.95664336	0.06030227	0.13112135
76	0.976	0.94755245	0.06531973	0.14313293
77	0.974	0.94376457	0.07333333	0.1581101
78	0.976	0.94825175	0.06531973	0.14130776
79	0.974	0.94446387	0.07333333	0.15647684
80	0.9715	0.93870629	0.07115746	0.15326508
81	0.98	0.95734266	0.06030227	0.12907677
82	0.978	0.9520979	0.06289321	0.1373346
83	0.98	0.95664336	0.06030227	0.13112135
84	0.9755	0.9470979	0.0668539	0.14448852
85	0.9755	0.9477972	0.0668539	0.14268294
86	0.9675	0.93037296	0.08020649	0.1714723
87	0.976	0.94825175	0.06531973	0.14130776

Supplementary Table 3. Order of the 38 variables to be included according to the highest accuracy of the model.

Variable order	Selected genes
1	CNOT1
2	SIN3A
3	BPTF
4	CHD8
5	CHD9
6	DEK
7	SIRT6
8	SAP30L
9	C6orf89
10	SUPT16H
11	SMARCA5
12	CHD6
13	RBBP5
14	SMARCA2
15	CNOT6L
16	CNOT8
17	CNOT6
18	BAZ1B
19	CHD4
20	BRMS1
21	ACTR8
22	INO80B
23	BRD3
24	CNOT7
25	PHF8
26	CBX5
27	RUVBL2
28	TOP1
29	ACTR6
30	MORF4L1
31	CHD1
32	ANP32E
33	INO80E
34	ACTB
35	BRD7
36	PABPC1L
37	SMARCD3
38	UTY
39	PIH1D1
40	DPF2
41	BRD1
42	CAMK2D
43	MIER2

44	ACTR5
45	PHF1
46	SS18
47	CNOT3
48	DAXX
49	BRD8
50	SAP30BP
51	CHD3
52	WDR77
53	SUPT20H
54	BAZ1A
55	WDR61
56	NUDT5
57	RBBP7
58	SAP18
59	LRRK2
60	INO80C
61	ACTN4
62	ANP32B
63	SIRT2
64	CNOT9
65	SMARCB1
66	HIST1H1C
67	PHF21A
68	RBBP8
69	SMYD2
70	SS18L1
71	SUPT5H

Supplementary Table 4. TCGA pan-cancer survival analysis of BPTF, CNOT1 and SIN3A.

		BLCA	BRCA	CESC	CHOL	COAD	ESCA	GBM
BPTF	HR (high)	0.86	1.1	1.2	0.38	0.96	1.2	0.78
	Logrank p	0.32	0.51	0.42	0.049	0.86	0.54	0.17
SIN3A	HR (high)	1	1	0.85	0.58	1.2	0.86	0.99
	Logrank p	0.9	0.98	0.48	0.27	0.55	0.51	0.95
CNOT1	HR (high)	1	1.1	0.86	0.8	0.67	1.7	0.92
	Logrank p	0.86	0.42	0.53	0.64	0.11	0.03*	0.65
		HNSC	KIHC	KIRP	LAML	LIHC	LUAD	LUSC
BPTF	HR (high)	0.93	3.8	0.92	1.1	1.4	1.4	1.1
	Logrank p	0.62	0.075	0.78	0.84	0.065	0.025*	0.36
SIN3A	HR (high)	1	3.6	1.4	1.2	1.1	1.1	1.1
	Logrank p	0.93	0.087	0.32	0.5	0.6	0.67	0.71
CNOT1	HR (high)	1.1	0.45	0.85	1.4	1.3	1.3	1.1
	Logrank p	0.7	0.24	0.6	0.19	0.17	0.12	0.47
		OV	PAAD	PRAD	READ	STAD	THCA	UCEC
BPTF	HR (high)	1.1	0.76	1.3	0.51	1.1	2.3	1.1
	Logrank p	0.59	0.19	0.72	0.17	0.7	0.12	0.84
SIN3A	HR (high)	0.94	1.3	0.79	0.6	1.2	1.7	1.2
	Logrank p	0.59	0.26	0.72	0.28	0.33	0.3	0.59
CNOT1	HR (high)	0.95	1.1	1.8	0.53	1.3	1.8	0.98
	Logrank p	0.67	0.53	0.35	0.19	0.083	0.26	0.96