Am J Transl Res 2020;12(8):4108-4126
www.ajtr.org /ISSN:1943-8141/AJTR0108722

Original Article

Identification and validation of novel metastasis-related signatures of clear cell renal cell carcinoma using gene expression databases

Chun-Guang Ma[1,2*], Wen-Hao Xu[1,2*], Yue Xu[3*], Jun Wang[4*], Wang-Rui Liu[5], Da-Long Cao[1,2], Hong-Kai Wang[1,2], Guo-Hai Shi[1,2], Yi-Ping Zhu[1,2], Yuan-Yuan Qu[1,2], Hai-Liang Zhang[1,2], Ding-Wei Ye[1,2]

[1]Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, P. R. China; [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, P. R. China; [3]Department of Ophthalmology, The First Affiliated Hospital of Soochow University, Suzhou 215000, P. R. China; [4]Department of Urology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, Guangdong, P. R. China; [5]Department of Neurosurgery, Affiliated Hospital of Youjiang Medical College for Nationalities, Guangxi, P. R. China. [*]Equal contributors.

Received February 4, 2020; Accepted July 4, 2020; Epub August 15, 2020; Published August 30, 2020

Abstract: Patients with clear cell renal cell carcinoma (ccRCC) typically face aggressive disease progression when metastasis occurs. Here, we screened and identified differentially expressed genes in three microarray datasets from the Gene Expression Omnibus database. We identified 112 differentially expressed genes with functional enrichment as candidate prognostic biomarkers. Lasso Cox regression suggested 10 significant oncogenic hub genes involved in earlier recurrence and poor prognosis of ccRCC. Receiver operating characteristic curves validated the specificity and sensitivity of the Cox regression penalty used to predict prognosis. The area under the curve indexes of the integrated genes scores were 0.758 and 0.772 for overall and disease-free survival, respectively. The prognostic values of *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* were validated through an analysis of 10 hub genes in 380 patients with ccRCC from a real-world cohort. The expression levels of were of high prognostic value for predicting metastatic potential. These findings will likely significantly contribute to our understanding of the underlying mechanisms of ccRCC, which will enhance efforts to optimize therapy.

Keywords: Clear cell renal cell carcinoma, metastasis, signature, bioinformatics, prognosis

Introduction

Renal cell carcinoma (RCC) is one of the most common malignancies of the urinary system, accounting for approximately 2% of cancer deaths [1]. The worldwide morbidity and mortality rates of RCC are increasing by approximately 2%-3% each decade [2]. Until recently, the rates stabilized or declined in many economically advanced countries [3]. As the most predominant histological subtype, clear cell renal cell carcinoma (ccRCC) represents approximately 70% of RCC cases. Although numerous factors contribute to the pathogenesis of RCC, limited information is available that can be applied to explain the aggressive pathogenesis and progression of ccRCC. The aggressiveness of ccRCC is directly associated with metastatic potential, which is typically not efficiently targeted despite great advances in therapeutic strategies [4]. Metastasis is a significant hallmark of tumor progression and thus the major cause of poor overall survival of patients with ccRCC [5]. Therefore, the identification of new prognostic biomarkers and molecular alterations is urgently required to develop more effective treatments of ccRCC.

Accumulating evidence has demonstrated that gene expression panels and related hallmarks participate in the carcinogenesis and aggressive progression of ccRCC [6-8]. For example, the expression levels of a panel of seven genes associated with the extracellular matrix significantly correlate with metastasis and prognosis of ccRCC [9]. Moreover, multigene panels help predict signatures that detect ccRCC in biopsy specimens [10]. Nomograms employing integrated clinical and gene expression profiles predict pathological nuclear grades and help clinicians to manage personalized regimens [8]. Despite these encouraging advances in strategies to treat metastatic ccRCC, many patients do not achieve longer survival [11]. Therefore, we urgently require methods that identify the underlying biochemical mechanisms associated with metastasis and predict the prognosis of patients with ccRCC. Such methods will facilitate the development of optimal therapeutic strategies.

During the latest decade, high-throughput nucleotide sequencing technologies, combined with bioinformatics analysis, sensitively and specifically detect mRNA expression levels, which identify differentially expressed genes (DEGs) that represent hallmarks associated with the pathogenesis and progression of ccRCC [8, 9, 12]. However, the diversity of genomic alterations and molecular interactions associated with metastasis hinders the identification of patients at high-risk for metastasis who may benefit from available as well as potential therapies. To address these challenges, here we analyzed three transcriptional microarray datasets acquired from the Gene Expression Omnibus (GEO) database to identify DEGs between cancer tissues and adjacent tissues that will serve as biomarkers of ccRCC. We conducted functional pathway enrichment analyses to better understand the associated molecular mechanisms. Moreover, we employed protein-protein interaction (PPI) network analysis to better define the importance of these interactions as they relate to biological processes, molecular functions, and signal transduction [13-15].

Our findings led us to hypothesize that the oncogenicity of significant hub genes correlates with metastasis and that these genes may serve as potential prognostic factors that will facilitate the identification of therapeutic targets for managing ccRCC.

Materials and methods

Raw biological microarray data

Raw transcriptional microarray data from GEO (http://www.ncbi.nlm.nih.gov/geo) [16] were screened for metastatic or primary ccRCC. Corresponding genes were converted into probes and assigned symbols associated with annotation information. We analyzed the datasets GSE22541 (24 primary and 44 metastatic tumors), GSE47352 (4 primary and 4 metastatic tumors), and GSE85258 (14 primary and 14 metastatic tumors) (Affymetrix Human Genome U133 Plus 2.0 Array).

Data normalization and identification of DEGs

Preprocessing and normalization of raw biological data were the first steps in processing the DNA microarrays. This process removes biased microarray data to ensure its uniformity and integrity. Subsequently, background correction, propensity analysis, normalization, and visualization of probe data were performed using the robust multiarray average analysis algorithm 17 in the affy package of R.

DEGs between primary and metastatic ccRCC samples were screened and identified across experimental conditions. Delineating parameters such as adjusted *P* values (adj. *P*), the Benjamini and Hochberg false-discovery rate (FDR), and fold-change values were used for filtering DEGs and balancing between discovery of significant genes and limitations of false-positives. Probe sets without corresponding gene symbols, or genes with more than one overlapping probe-set, were removed or averaged. Log_2FC (fold-change) >1 and adj. *P* value <0.01 were considered to indicate a significant difference.

Protein-protein interaction (PPI) network and functional annotations

We used Search Tool for the Retrieval of Interacting Genes (http://string-db.org) (version 10.0) to predict a PPI network of DEGs and to analyze the degree of interactions between proteins [17]. A significant edge score was considered as an interaction combined score >0.4. Cytoscape (version 3.5) was used to visualize interactive network data [18].

To identify the role of DEGs in ccRCC, we used Gene Ontology (GO) enrichment analysis to extract functional attributes including biological processes (BP), molecular functions (MF), and cellular component (CC) [19]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used for this purpose as well [20]. The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.ncifcrf.gov; Version 6.8) was analyzed to explore the role of development-related signaling pathways in ccRCC [21]. *P*<0.05 was considered to indicate a significant.

ClueGO is a Cytoscape plug-in that visualizes nonredundant biological terms associated with large clusters of genes in a functionally grouped network [22]. The results of KEGG pathway analysis of selected hub genes enriched in the GO term BP were visualized using ClueGO (version 2.5.3) and the Cytoscape plugin CluePedia (version 1.5.3). Potential associations of 24 coordinately expressed hub genes and their potential prognostic values are shown using a heat map. The hub nodes of networks with connectivity degrees >10 were identified. A network comprising 24 genes and their coordinately expressed neighboring genes was constructed using cBioPortal (http://www.cbioportal.org) [23].

Statistical analysis of TCGA data cohort

A lasso Cox regression model was used to identify independent prognostic factors. Functional annotations and enrichment of signaling pathways were predicted and illustrated using DAVID. Phenotype and transcriptomics data of selected hub genes of 515 patients with ccRCC in TCGA data are displayed. Expression profiles were respectively identified as binary variables (high vs low), referring to the median expression level of each hub gene in TCGA cohort data. The primary and secondary endpoints for patients were disease-free survival (DFS) and overall survival (OS), respectively. The duration of follow-up was estimated using the Kaplan-Meier method with 95% confidence intervals (95% CI) and the log-rank test. X-tile software was used to determine the cutoff value [24]. All hypothetical tests were two-sided, and *P* values <0.05 were considered to indicate a significant difference for all tests. Hierarchical partitioning was performed using transcriptional expression profiles of selected oncogenes in a heat map. Color gradients indicate a high (red) or a low (blue) expression level.

The human protein atlas

The Human Protein Atlas project (https://www.proteinatlas.org) contains immunohistochemistry (IHC) data acquired using a tissue microarray analysis of 17 paired, major cancer types [25]. Data include IHC staining intensity, quantity, location, and patients' information. Here we used the Human Protein Atlas (https://www.proteinatlas.org) to identify representative proteins differentially expressed between ccRCC and corresponding normal tissues.

ccRCC patients from a validated cohort

We analyzed samples acquired from 380 ccRCC patients who underwent nephrectomy at the Department of Urology of Fudan University Shanghai Cancer Center (FUSCC; Shanghai, China) from June 2009 to September 2017. Tissue samples were obtained during surgery, and available from the FUSCC tissue bank.

Quantitative real-time PCR (RT-qPCR) analysis

Total cellular RNA, which was isolated using Trizol (Invitrogen, Carlsbad, CA) in accordance with the source's protocols, was reversed-transcribed using a PrimeScript RT reagent kit (Thermo Fisher, USA). Primers were diluted and mixed in RNase-Free ddH_2O, and RT-qPCR was performed using the SYBR Green qPCR method (Takara Biotechnology Co.). The levels of *GAPDH* mRNA were measured to serve as a standard. Specific amplification conditions were performed using the SYBR Green qPCR Master Mix (Applied Biosystems) according to the manufacturer's protocols. Primes sequences were shown in **Table 2**. The relative expression level of the target mRNA was calculated using the $2^{-\Delta\Delta Ct}$ method.

Statistical analysis of FUSCC-cohort data

We evaluated the significance of the associations of DFS and OS with distinct mRNA expression groups of 10 hub genes expressed by patients with ccRCC. The duration of follow-up was estimated using the Kaplan-Meier method (95% CI) and the log-rank test. Integrated scores represent the sums of each significant oncogenic hub-gene weight. Receiver operating characteristic (ROC) curves were generated to validate the specificity and sensitivity of a diagnosis according to high or low integrated scores of significant hub-gene levels. Area under the curve (AUC) analysis was performed to determine diagnostic ability.

Gene set enrichment analysis (GSEA) and related gene networks

TCGA data were subjected to GSEA using Bioconductor category version 2.10.1 package.

Table 1. Functional roles of 10 prognostic hub genes

No.	Gene symbol	Full name	Function
1	ADAMTS9	A Disintegrin And Metalloproteinase With Thrombospondin Motifs 9	Pathway: Metabolism of proteins and O-linked glycosylation. GO: Metalloendopeptidase activity and endopeptidase activity.
2	C1S	Complement C1s	Pathway: Immune response Lectin induced complement pathway and Creation of C4 and C2 activators. GO: Calcium ion binding and serine-type endopeptidase activity.
3	DPYSL3	Dihydropyrimidinase Like 3	Pathway: Semaphorin interactions and Developmental Biology. GO: Hydrolase activity and phosphoprotein binding.
4	H2AFX	H2A Histone Family Member X	Pathway: Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3 and ATM Pathway. GO: Sequence-specific DNA binding and enzyme binding.
5	MINA	MYC-Induced Nuclear Antigen	Pathway: Validated targets of C-MYC transcriptional activation and Chromatin organization.
6	PLOD2	Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 2	Pathway: Collagen chain trimerization and Lysine degradation. GO: Oxidoreductase activity and oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen.
7	RUNX1	Runt Related Transcription Factor 1	Pathway: Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds and Embryonic and Induced Pluripotent Stem Cell Differentiation Pathways and Lineage-specific Markers. GO: DNA-binding transcription factor activity and protein homodimerization activity.
8	SLC19A1	Solute Carrier Family 19 Member 1	Pathway: Metabolism of water-soluble vitamins and cofactors and Antifolate resistance. GO: Oxidoreductase activity and folic acid transmembrane transporter activity.
9	TPX2	Targeting Protein For Xklp2	Pathway: Gene Expression and Cell Cycle, Mitotic. GO: GTP binding and protein kinase binding.
10	TRIB3	Tribbles Pseudokinase 3	Pathway: Class I MHC mediated antigen processing and presentation and RET signaling. GO: Transferase activity, transferring phosphorus-containing groups and protein kinase binding.

Table 2. Primer sequences (5'-3') of 10 prognostic hub genes for qRT-PCR

No.	Gene symbol	Forward	Reverse
1	ADAMTS9	CAGAAGGGGCTTGGTTGG	TCGTGTTCCTACCCTATTTTGA
2	C1S	GTTGTCATGGACAGTGAGAG	GCCTAAATTCACCCTGGAAG
3	DPYSL3	GGATCACGAGTGACCGCCTT	TCGTCATTCACGCGCCATGT
4	H2AFX	CGGCAGTGCTGGAGTACCTCA	AGCTCCTCGTCGTTGCGGATG
5	MINA	CCAAAGAACTGCTTTCCTCAGAC	CTACACTGTCCAGCCTCGGTAA
6	PLOD2	CATGGACACAGGATAATGGCTG	AGGGGTTGGTTGCTCAATAAAAA
7	RUNX1	GCCAGGAACCGGCCTTACTC	GCTAGTGTGCCGAGGAAGA
8	SLC19A1	TTGCCCAAGCTATTCTCAGTCGA	CAGAGACACCGCCAGCCACAT
9	TPX2	ACCTTGCCCTACTAAGATT	AATGTGGCACAGGTTGAGC
10	TRIB3	ATGCCCCCTCGGATTTCATC	TTGCCCTGAAAAAGCCCTCC
11	GAPDH	GTCTTC TCCACCATGGAGAAGG	CATGCCAGTGAGCTTCCCGTTCA

For each analysis, the Student *t* test was performed, and the mean levels of differentially expressed genes were calculated. A permutation test (1000 times) was used to identify significantly changed pathways. The default adj. *P* values determined using the BH-FDR method were applied to correct for false-positives [26]. Significantly related genes were defined as those with adj. *P*<0.01 and FDR>0.25. Statistical analyses and graphical presentation were conducted using R software (Version 3.3.2). A detailed PPI network associated with a default set of 10 hub genes was constructed using GeneMANIA (http://genemania.org/).

Results

We first assessed DEGs using three datasets hosted on the GEO platform to identify coordinately-regulated hub genes. Second, we evaluated the significance of the associations between the expression levels of hub genes and prognosis according to TCGA data and then validated their prognostic value using a real-world cohort. Third, we predicted potential functional annotations and hallmark pathways.

Identification of DEGs in ccRCC

After normalization and identification of mRNA microarray data, the DEGs (12,817; 1,817; and 1,766 probe samples in GSE22541, GSE85258, and GSE47352, respectively) were determined as significant according to the analytical and statistical parameters of the data processing steps. The overlap among the three datasets included 112 significant genes (**Figure 1A**).

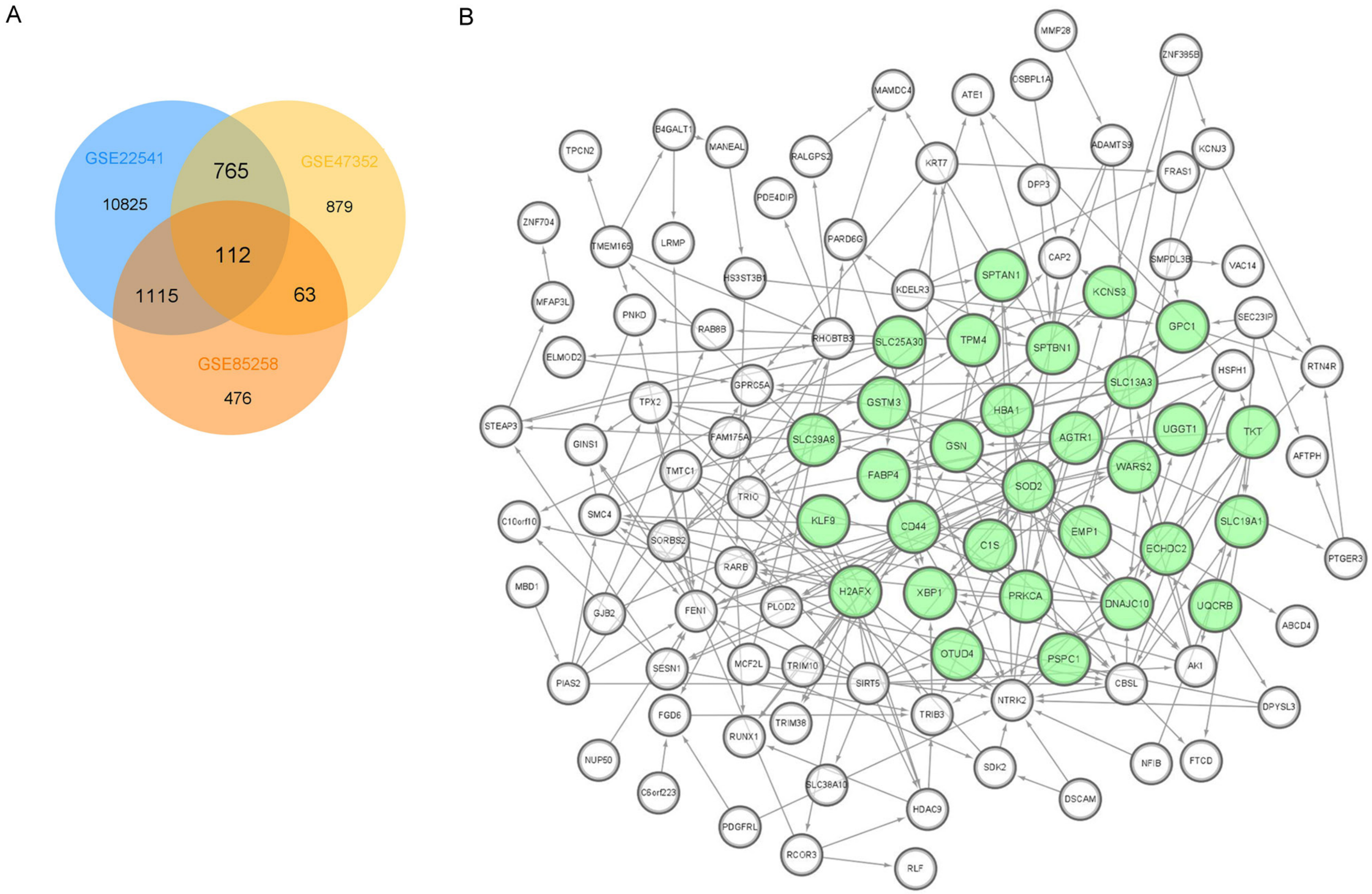
Construction of a PPI network and functional analysis of modules

The PPI network comprising the DEGs is shown in **Figure 1B**, and the network of the DEGs and their coordinately expressed neighboring genes are shown in **Figure 1C**. The enrichment profiles determined using DAVID-GO functional analyses suggested that the 112 DEGs in this module were primarily enriched in GO terms such as double-strand break, positive response to stimulus, sarcomere, external side of plasma membrane, metal ion biding, cation binding, ion binding, and cytoskeletal protein binding (**Figure 1D**).

To further identify underlying signaling pathways, we analyzed KEGG pathways together with GO functional annotations (**Figure 2**). Detailed functional notes and classification pie charts are shown in Supplementary Figure 1. The frequencies of GO terms determined using ClueGO analysis were as follows: 25.0%, negative regulation of protein complex disassembly; 12.5%, negative regulation of response to oxidative stress; 9.38%, cell-cell junction assemble; 6.25%, maintenance of protein location in cell; 6.25%, photoreceptor cell differentiation; 6.25%, inorganic cation import across plasma membrane; and 6.25%, lysosome organization.

Selection of significant hub genes

Lasso Cox regression analysis suggested that C1S, PLOD2, ADAMTS9, AK1, CHD2, DCAF8, DNAJC10, DPP3, DPYSL3, GSTM3, H2AFX, HBA1, RUNX1, MAMDC4, MFAP3L, MINA, OSBPL1A, PIAS2, RBBP6, RCOR3, RHOBTB3,


A
GSE22541
10825
765
GSE47352
879
112
1115
63
GSE85258
476
B
MMP28
ZNF385B
OSBPL1A
MAMDC4
ATE1
B4GALT1
MANEAL
TPCN2
RALGPS2
KRT7
ADAMTS3
KCNJ3
FRAS1
DPP3
PDE4DIP
ZNF704
TMEM165
LRMP
PARD6G
CAP2
SMPDL3B
VAC14
HS3ST3B1
KDELR3
SPTAN1
KCNS3
MFAP3L
PNKD
RAB8B
RHOBTB3
SLC25A30
TPM4
SPTBN1
GPC1
SEC23IP
RTN4R
ELMOD2
GPRC5A
SLC13A3
HSPH1
STEAP3
TPX2
GSTM3
HBA1
FAM175A
SLC39A8
GSN
AGTR1
UGGT1
TKT
AFTPH
GINS1
TMTC1
FABP4
SOD2
WARS2
TRIO
SMC4
KLF9
CD44
C1S
EMP1
ECHDC2
SLC19A1
C10orf10
SORBS2
PTGER3
RARB
H2AFX
XBP1
PRKCA
DNAJC10
UQCRB
MBD1
GJB2
FEN1
PLOD2
ABCD4
OTUD4
PSPC1
PIAS2
SESN1
MCF2L
TRIM10
SIRT5
CBSL
AK1
NTRK2
DPYSL3
FGD6
TRIB3
TRIM38
RUNX1
NUP50
SDK2
NFIB
FTCD
C6orf223
SLC38A10
DSCAM
HDAC9
PDGFRL
RCOR3
RLF

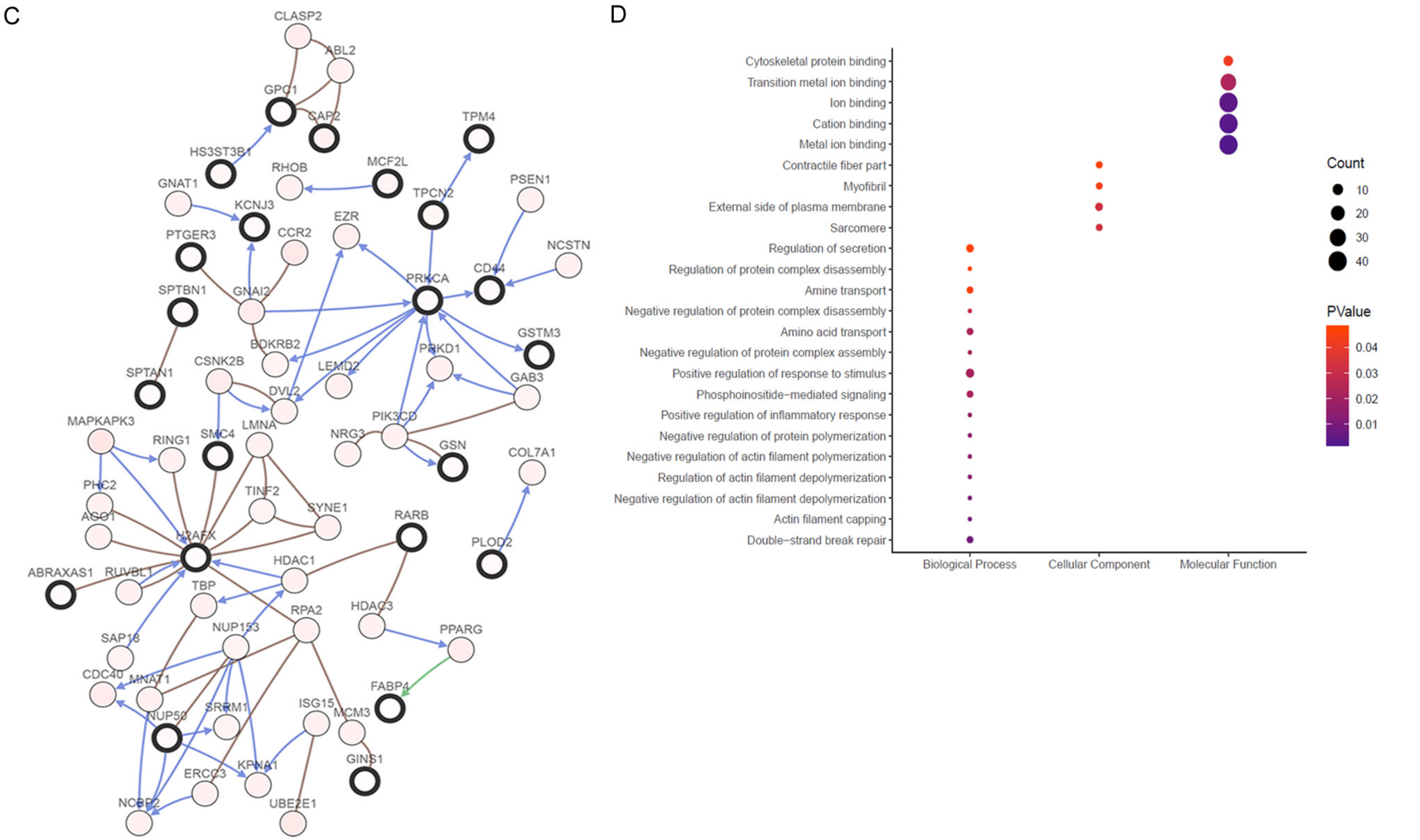


Figure 1. Screening and selection of DEGs using multiple GEO database. A. After normalization and identification of mRNA microarray data, the DEGs (12,817; 1,817; and 1,766 probe samples in GSE22541, GSE85258, and GSE47352, respectively) were determined as significant according to the analytical and statistical parameters of the data processing steps. The overlap among the three datasets included 112 significant genes. B. The PPI network of the DEGs was constructed using Cytoscape. MCODE, a plug-in of Cytoscape, predicts the most significant gene panel, marked in light green. C. DEGs and their co-regulated network were analyzed using cBioPortal. Nodes with bold black outline represent hub genes. Nodes with thin black outline represent the co-expression genes. D. The enrichment profiles from DAVID GO functional analyses of the 112 hub genes suggested that the hub genes in this module were primarily enriched in double-strand break, positive response to stimulus, sarcomere, external side of plasma membrane, metal ion biding, cation binding, ion binding, cytoskeletal protein binding and so one.

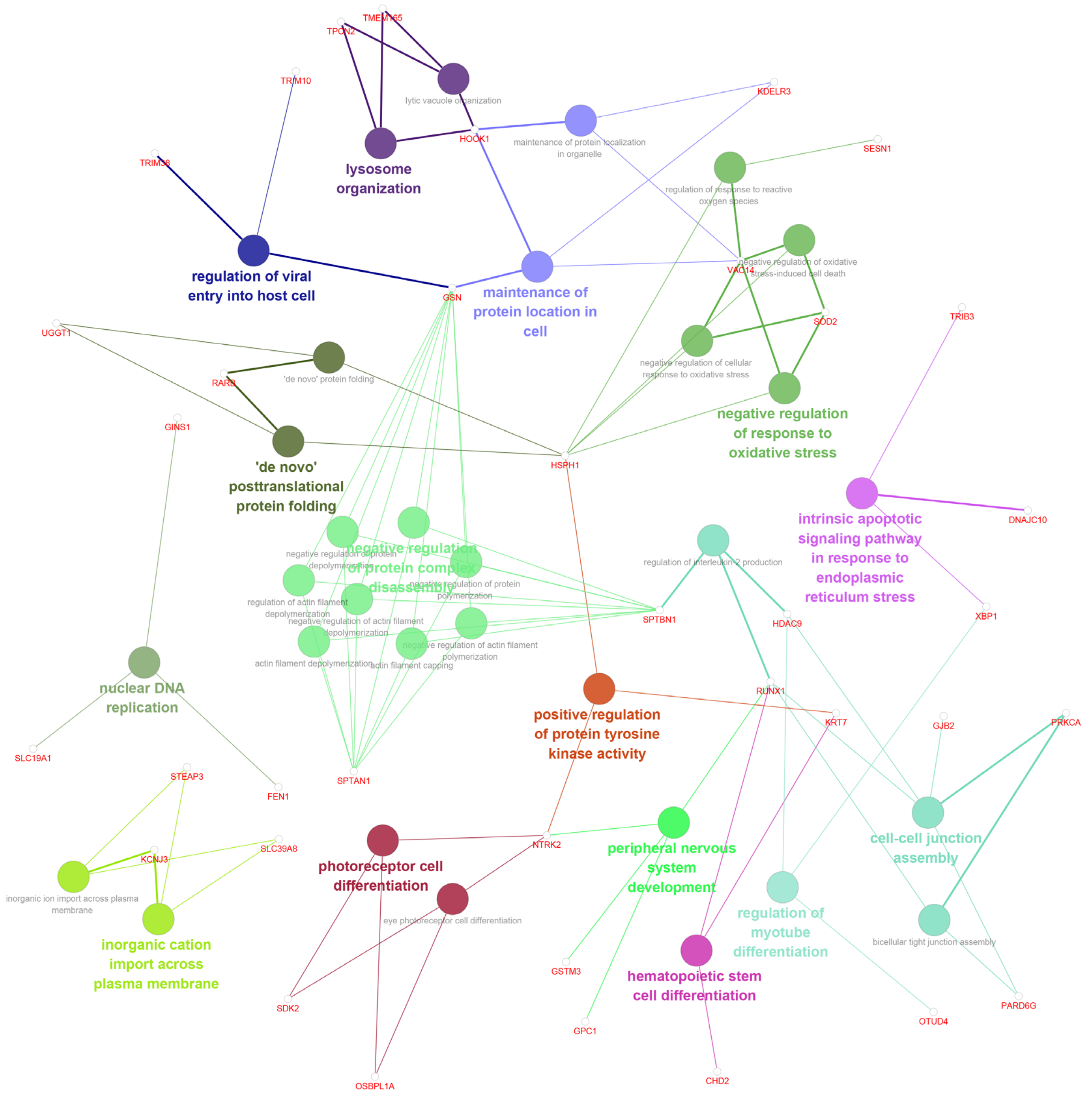

TMEM165
TPCN2
lytic vacuole organization
TRIM10
KDELR3
HOOK1
maintenance of protein localization in organelle
lysosome organization
SESN1
TRIM38
regulation of response to reactive oxygen species
regulation of viral entry into host cell
negative regulation of oxidative stress-induced cell death
VAC14
GSN
maintenance of protein location in cell
SOD2
TRIB3
UGGT1
'de novo' protein folding
RARB
negative regulation of cellular response to oxidative stress
negative regulation of response to oxidative stress
GINS1
HSPH1
'de novo' posttranslational protein folding
intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress
DNAJC10
negative regulation of protein depolymerization
negative regulation of protein complex disassembly
regulation of interleukin-2 production
negative regulation of protein polymerization
regulation of actin filament depolymerization
negative regulation of actin filament depolymerization
SPTBN1
HDAC9
XBP1
negative regulation of actin filament polymerization
actin filament depolymerization
actin filament capping
nuclear DNA replication
RUNX1
positive regulation of protein tyrosine kinase activity
KRT7
GJB2
PRKCA
SLC19A1
STEAP3
FEN1
SPTAN1
SLC39A8
KCNJ3
NTRK2
photoreceptor cell differentiation
peripheral nervous system development
cell-cell junction assembly
inorganic ion import across plasma membrane
inorganic cation import across plasma membrane
eye photoreceptor cell differentiation
regulation of myotube differentiation
bicellular tight junction assembly
GSTM3
hematopoietic stem cell differentiation
SDK2
PARD6G
OTUD4
GPC1
CHD2
OSBPL1A

Figure 2. Functional annotations of significant DEGs. To further identify underlying signaling pathways, we analyzed KEGG pathways together with GO functional annotations using ClueGO and CluePedia, plug-ins of Cytoscape.

SLC19A1, SORBS2, SPTAN1, SPTBN1, SYT13, TPM4, TPX2, TRIB3, UGGT1 were significant weighted prognostic factors. Functional annotation of these genes revealed they were primarily enriched in the terms as follows: regulation of actin filament capping, non-membrane-bounded organelle, sarcomere, myofibril, organelle lumen, nucleolus, actin cytoskeleton, and nuclear lumen (**Figure 3**).

Kaplan-Meier analysis indicated that differentially elevated levels of the hub genes ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2, and TRIB3 significantly correlated with poor prognosis of 513 patients with ccRCC (**Figure 4A-T**). Functional annotations of each prognostic hub gene are listed in **Table 1**. Hierarchical partitioning was performed using the transcriptional expression profiles of selected 10 oncogenes (**Figure 4U**).

External validation of hub DEGs in 380 ccRCC patients from FUSCC cohort

DEGs corresponding to *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* mRNAs were identified using the TCGA database and GSE11151 [27] (**Figure**

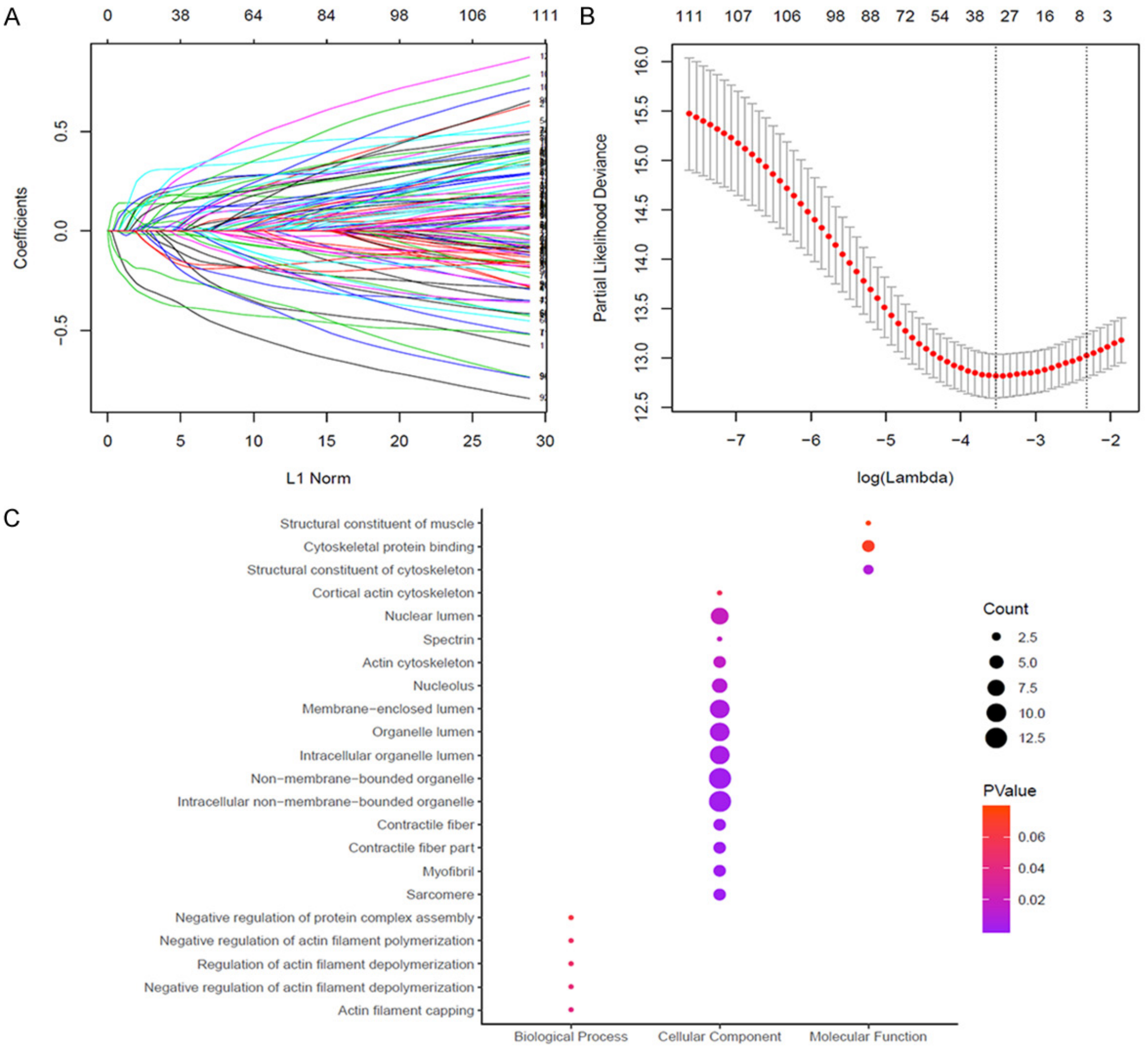
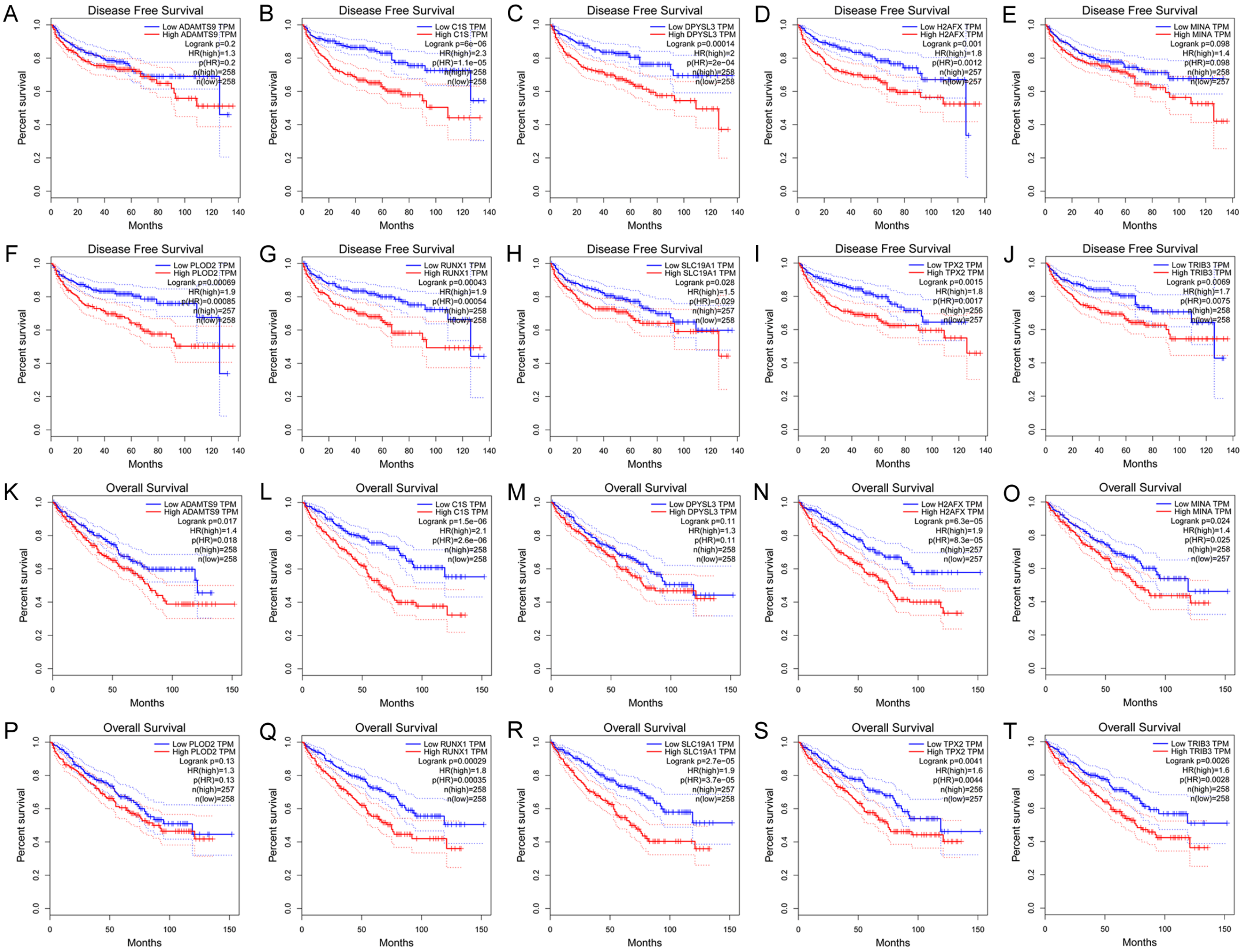


Figure 3. Further identification of hub genes using LASSO regression analysis. A, B. To further select significant prognostic metastasis-related biomarkers, LASSO Cox regression was performed to further search for hub genes associated with metastasis and to narrow the scope of the panel, suggested that a total of 31 genes, including *C1S, PLOD2, ADAMTS9, AK1, CHD2, DCAF8, DNAJC10, DPP3, DPYSL3, GSTM3, H2AFX, HBA1, RUNX1, MAMDC4, MFAP3L, MINA, OSBPL1A, PIAS2, RBBP6, RCOR3, RHOBTB3, SLC19A1, SORBS2, SPTAN1, SPTBN1, SYT13, TPM4, TPX2, TRIB3, UGGT1* are significant weighted prognostic factors. C. GO function annotations of 31 selected genes. Functional annotation of these genes revealed they were primarily enriched in the terms as follows: regulation of actin filament capping, non-membrane-bounded organelle, sarcomere, myofibril, organelle lumen, nucleolus, actin cytoskeleton, and nuclear lumen.

5). Significantly increased levels of these mRNAs were expressed by both cohorts. In contrast, *MINA* mRNA was differentially expressed at lower levels. The levels of the cognate proteins expressed by these hub genes in tumor and adjacent tissues detected using IHC and reported by the Human Protein Atlas are shown in Supplementary Figure 2.

A total of 380 ccRCC patients with available clinical and pathological data were enrolled in this study (**Table 3**). After integration of qRT-PCR and clinical follow-up data for 380 patients with ccRCC, we confirmed that *ADAMTS9, C1S, DPYSL3, H2AFX, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* mRNAs were differentially expressed in the FUSCC cohort. Differentially elevated mRNA levels were significantly associated with shorter DFS and OS (*P*<0.05) (**Figure 6**).

An integrated gene panel was constructed using 10 hub gene (*ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2,*


A
Disease Free Survival
Low ADAMTS9 TPM
High ADAMTS9 TPM
Logrank p=0.2
HR(high)=1.3
p(HR)=0.2
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
B
Disease Free Survival
Low C1S TPM
High C1S TPM
Logrank p=6e−06
HR(high)=2.3
p(HR)=1.1e−05
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
C
Disease Free Survival
Low DPYSL3 TPM
High DPYSL3 TPM
Logrank p=0.00014
HR(high)=2
p(HR)=2e−04
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
D
Disease Free Survival
Low H2AFX TPM
High H2AFX TPM
Logrank p=0.001
HR(high)=1.8
p(HR)=0.0012
n(high)=257
n(low)=257
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
E
Disease Free Survival
Low MINA TPM
High MINA TPM
Logrank p=0.098
HR(high)=1.4
p(HR)=0.098
n(high)=258
n(low)=257
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
F
Disease Free Survival
Low PLOD2 TPM
High PLOD2 TPM
Logrank p=0.00069
HR(high)=1.9
p(HR)=0.00085
n(high)=257
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
G
Disease Free Survival
Low RUNX1 TPM
High RUNX1 TPM
Logrank p=0.00043
HR(high)=1.9
p(HR)=0.00054
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
H
Disease Free Survival
Low SLC19A1 TPM
High SLC19A1 TPM
Logrank p=0.028
HR(high)=1.5
p(HR)=0.029
n(high)=257
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
I
Disease Free Survival
Low TPX2 TPM
High TPX2 TPM
Logrank p=0.0015
HR(high)=1.8
p(HR)=0.0017
n(high)=256
n(low)=257
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
J
Disease Free Survival
Low TRIB3 TPM
High TRIB3 TPM
Logrank p=0.0069
HR(high)=1.7
p(HR)=0.0075
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 20 40 60 80 100 120 140
K
Overall Survival
Low ADAMTS9 TPM
High ADAMTS9 TPM
Logrank p=0.017
HR(high)=1.4
p(HR)=0.018
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
L
Overall Survival
Low C1S TPM
High C1S TPM
Logrank p=1.5e−06
HR(high)=2.1
p(HR)=2.6e−06
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
M
Overall Survival
Low DPYSL3 TPM
High DPYSL3 TPM
Logrank p=0.11
HR(high)=1.3
p(HR)=0.11
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
N
Overall Survival
Low H2AFX TPM
High H2AFX TPM
Logrank p=6.3e−05
HR(high)=1.9
p(HR)=8.3e−05
n(high)=257
n(low)=257
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
O
Overall Survival
Low MINA TPM
High MINA TPM
Logrank p=0.024
HR(high)=1.4
p(HR)=0.025
n(high)=258
n(low)=257
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
P
Overall Survival
Low PLOD2 TPM
High PLOD2 TPM
Logrank p=0.13
HR(high)=1.3
p(HR)=0.13
n(high)=257
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
Q
Overall Survival
Low RUNX1 TPM
High RUNX1 TPM
Logrank p=0.00029
HR(high)=1.8
p(HR)=0.00035
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
R
Overall Survival
Low SLC19A1 TPM
High SLC19A1 TPM
Logrank p=2.7e−05
HR(high)=1.9
p(HR)=3.7e−05
n(high)=257
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
S
Overall Survival
Low TPX2 TPM
High TPX2 TPM
Logrank p=0.0041
HR(high)=1.6
p(HR)=0.0044
n(high)=256
n(low)=257
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150
T
Overall Survival
Low TRIB3 TPM
High TRIB3 TPM
Logrank p=0.0026
HR(high)=1.6
p(HR)=0.0028
n(high)=258
n(low)=258
Percent survival
0.0 0.2 0.4 0.6 0.8 1.0
Months
0 50 100 150

U

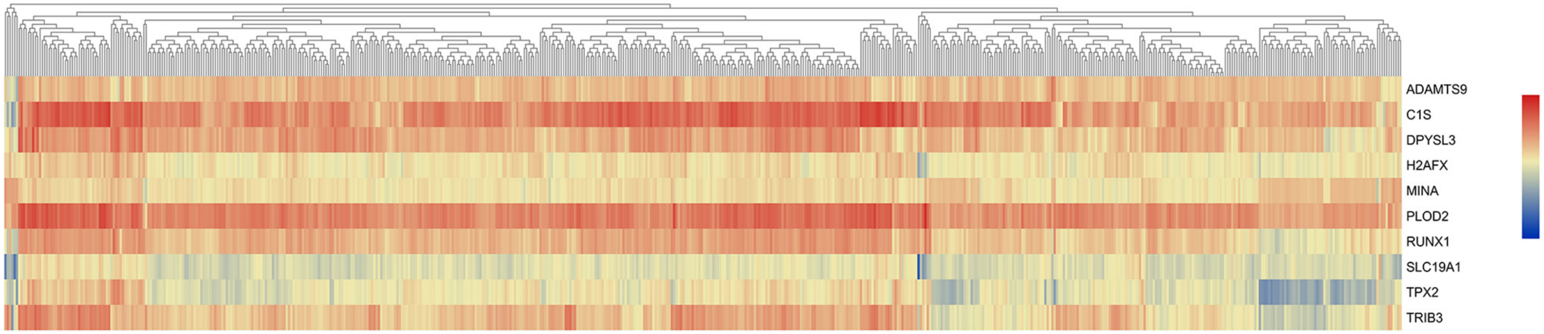


Figure 4. Prognostic value of significant hub genes from TCGA cohort. A-T. Kaplan-Meier analysis indicated that differentially elevated levels of the hub genes *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* significantly correlated with poor prognosis of 513 patients with ccRCC. U. Hierarchical partitioning was performed using transcriptional expression profiles of 10 selected oncogenes.

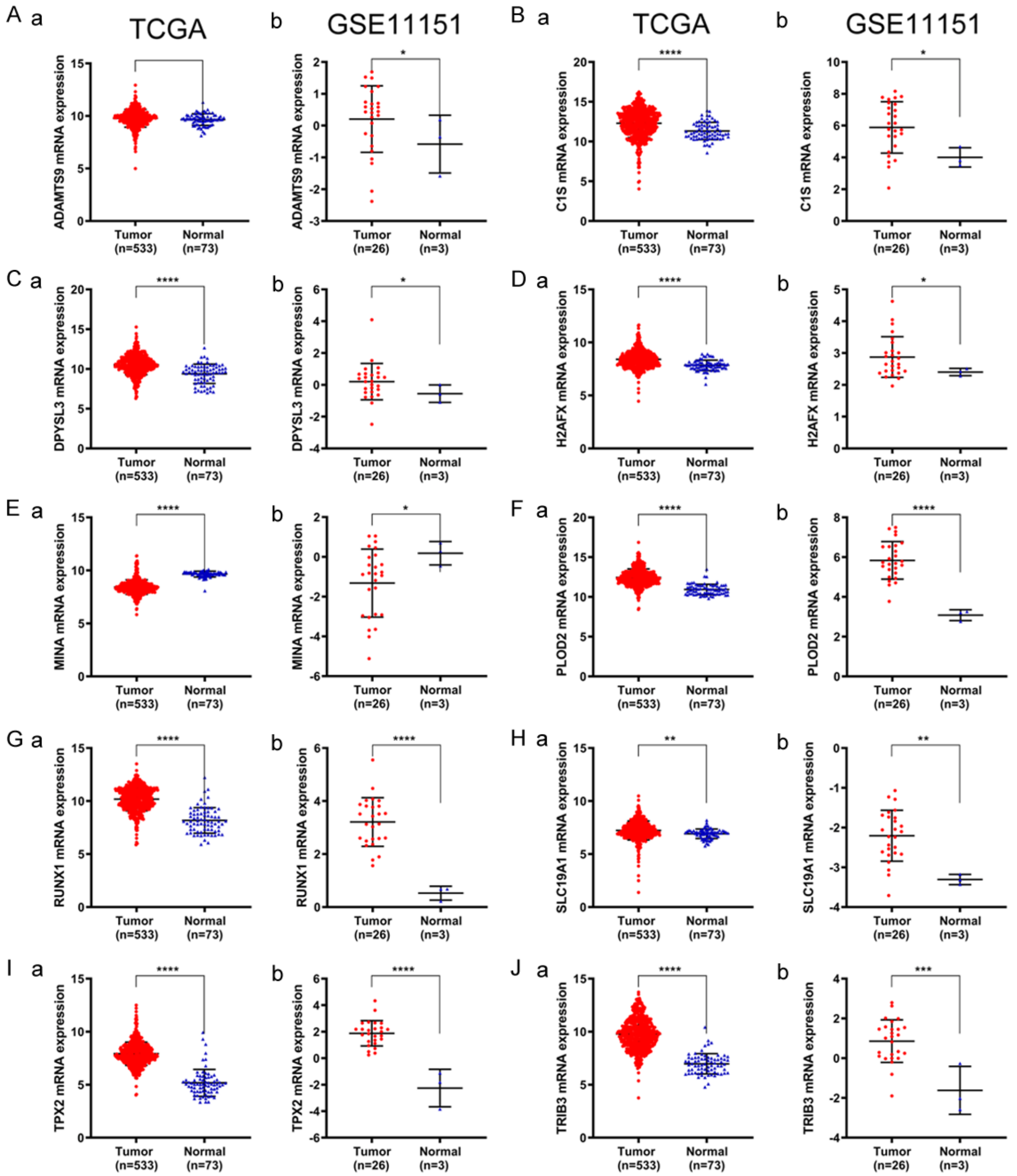


Figure 5. Differential expression validation of 10 hub genes in TCGA and GSE11151. Transcriptional expression of *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2, TRIB3* between were tumor and normal samples were generated using TCGA database and GSE11151.

TRIB3), which may serve as an independent panel to predict the progression of ccRCC. Kaplan-Meier analysis revealed significant associations of mRNA levels with DFS (P<0.0001) and OS (P<0.0001) (**Figure 7A**, **7B**). ROC curve analysis indicated the ability of the gene model to predict metastasis (AUCs of the integrated model were 0.758 for OS and 0.772 for DFS) (**Figure 7C**).

Significantly associated genes and their signaling pathways

GSEA identified 100 DEGs with positive or negative correlations. Among them, we used the hallmarks *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* to perform GSEA. **Figure 8** shows the three most relevant hallmark pathways of each

Table 3. Clinicopathological characteristics baseline in 380 ccRCC patients from FUSCC cohort (FUSCC: Fudan university shanghai cancer center)

Characteristics	FUSCC cohort (N=380)
N (%)	
Age	
<60 years	253 (66.6)
≥60 years	127 (33.4)
Gender	
Male	258 (67.9)
Female	122 (32.1)
BMI	
<25 kg/m^2	231 (60.8)
≥25 kg/m^2	149 (39.2)
pT stage	
T1-T2	307 (80.8)
T3-T4	73 (19.2)
pN stage	
N0	334 (87.9)
N1	46 (12.1)
pM stage	
M0	310 (81.6)
M1	70 (18.4)
AJCC stage	
I-II	292 (76.8)
III-IV	88 (23.2)
ISUP grade	
G1-G2	182 (47.9)
G3-G4	198 (52.1)

hub gene. The most significant genes associated with 10 hub genes, and their relationships are shown in **Figure 9A**, and a detailed PPI network related to this gene set shows additional interrelated nodes at the protein level (**Figure 9B**).

Discussion

Genetic and epigenetic alterations contribute to the development and progression of RCC [28]. ccRCC, which is the most aggressive histological subtype of ccRCC, is associated with elevated mortality owing to its high metastatic potential [29]. Unfortunately, the underlying mechanisms of oncogenesis and metastasis of ccRCC are unknown. We reasoned that microarray technology represents a powerful tool to fill this gap in our knowledge, because it enables comprehensive mRNA expression profiling of ccRCC through its ability to identify and characterize new biomarkers involved in tumorigenesis and progression [30, 31]. Here we conducted an integrated systematic analysis microarray data of well-characterized primary and metastatic ccRCCs to identify unique gene expression profiles characteristic of tumor aggressiveness. We identified 112 DEGs and 10 prognostic hub genes associated with functional annotations in different expression levels. GSEA was used to visualize significantly enriched gene-set hallmarks of 10 selected hub genes. These findings provide the basis for further screening and identification of promising biomarkers of tumor aggressiveness.

Several scoring systems combine clinical and pathological features to determine prognosis and to accurately predict survival outcomes of patients with primary and metastatic ccRCC [32-34]. These predictive models improve the specificity and accuracy of predicting survival outcomes of patients ccRCC, and the incorporated clinicopathological parameters are surrogate measures of major fundamental mechanisms that determine the aggressiveness of malignant tumors. Moreover, proteogenomic characterization of cancers that aims to develop therapeutic strategies will benefit from the acquisition of knowledge of the details of the biological mechanisms that contribute to the oncogenesis and progression of ccRCC.

At present, there have been multiple clinical research models based on multi-center or retrospective studies to predict prognosis, which helps to predict clinical prognosis from the perspective of multiple prognostic factors. These models have the potential to change our clinical practice and guide clinicians to implement individualized research strategies, which has strong clinical guidance value. In 2018, RCClnc4 classifier has been demonstrated to have precise prognostic significance in early ccRCC using four LncRNAs [35]. A retrospective analysis and multicentre validation study also constructed a six-SNP-based classifier for predicting recurrence in localised renal cell carcinoma [36]. Interestingly, there are few researches to analyze the genomic difference between of metastatic and primary ccRCC. The present study represents the first attempt to establish a gene regulatory network incorporating metas-

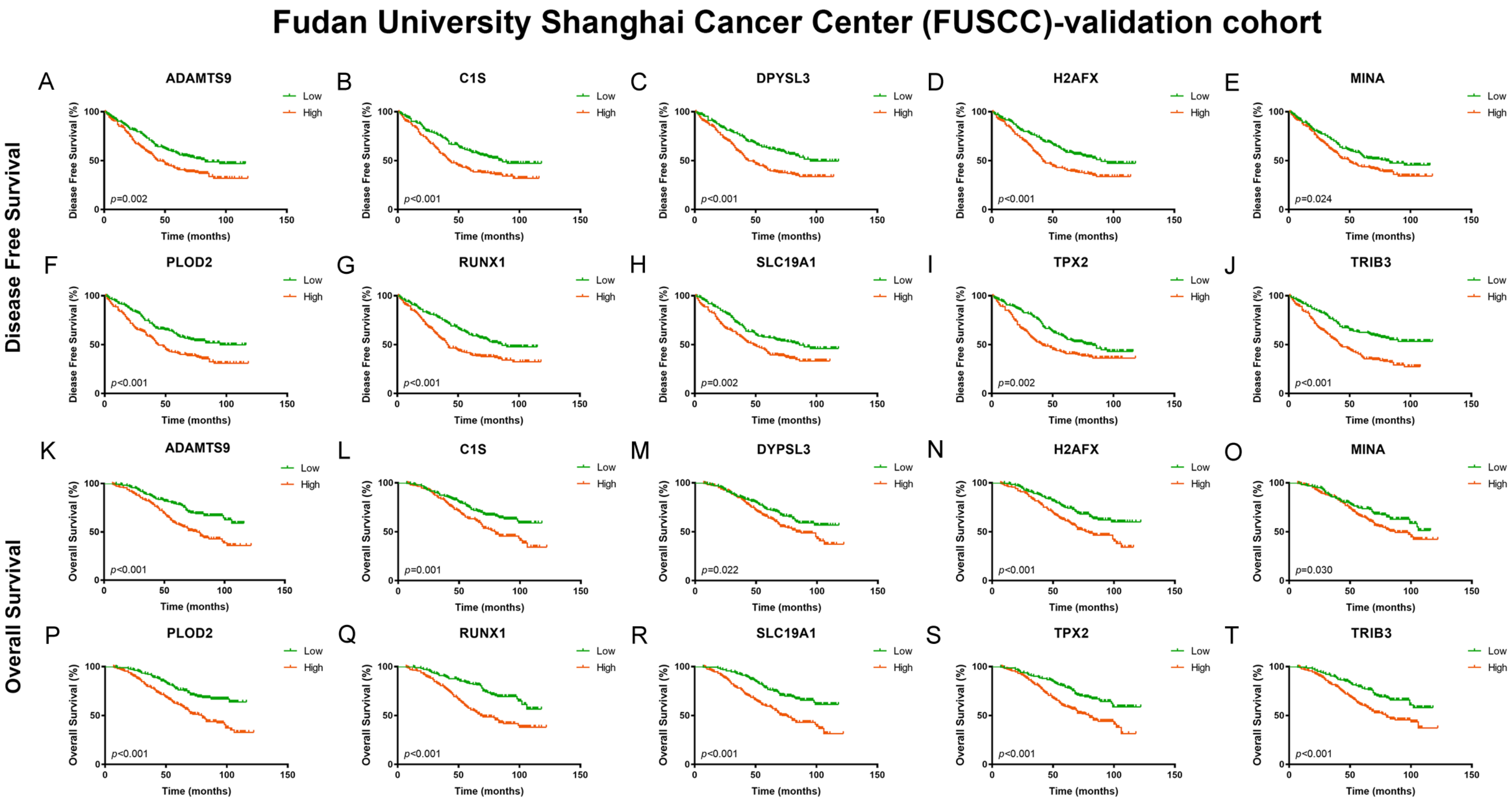


Figure 6. External validation for prognostic value of 10 hub genes and prediction model in 380 ccRCC patients from FUSCC cohort. After integration of qRT-PCR and clinical follow-up data for 380 patients with ccRCC, we confirmed that *ADAMTS9, C1S, DPYSL3, H2AFX, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* mRNAs were differentially expressed in the FUSCC cohort. Differentially elevated mRNA levels were significantly associated with shorter DFS and OS (*P*<0.05).

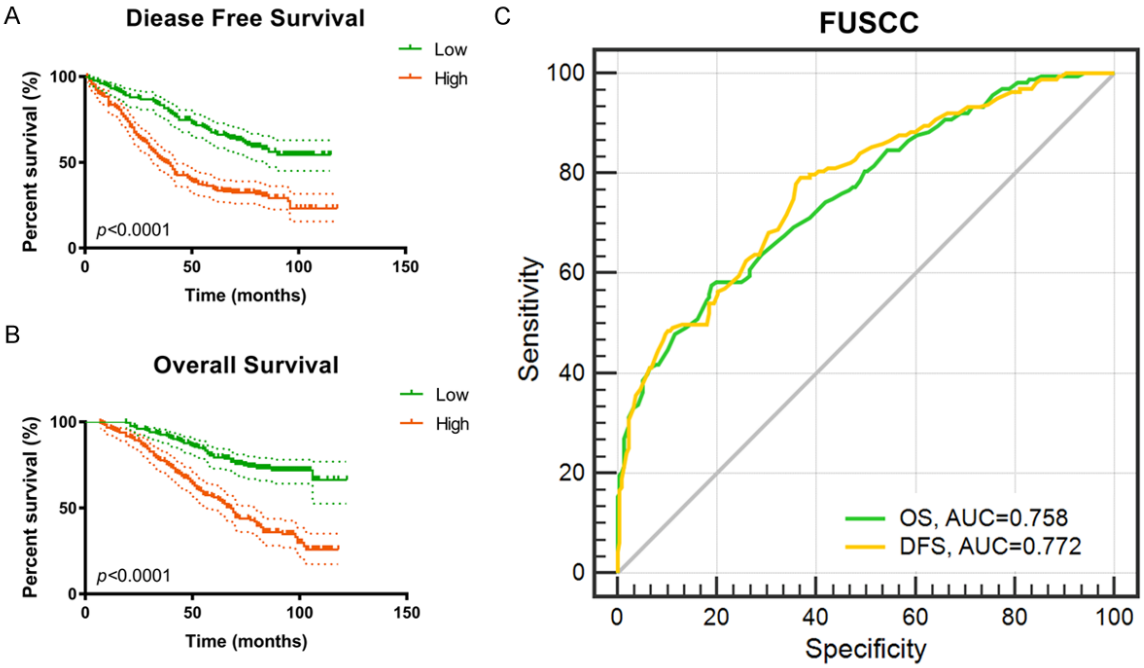


Figure 7. External validation for prognostic value of integrated prediction model in 380 ccRCC patients from FUSCC cohort. A, B. An integrated gene panel was constructed using 10 hub genes (*ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, *TRIB3*), which may serve as an independent panel to predict the progression of ccRCC. Kaplan-Meier analysis revealed significant associations of mRNA levels with DFS (*P*<0.0001) and OS (*P*<0.0001). C. ROC curve analysis indicated the ability of the gene model to predict metastasis (AUCs of the integrated model were 0.758 for OS and 0.772 for DFS).

tasis-related and identifies the hub genes associated with the metastatic potential of ccRCC. Moreover, we show here that the alterations of the expression levels of *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* were significantly associated with shorter OS and DFS, indicating that these significant DEGs may play important roles in the aggressive malignant phenotype of ccRCC.

The limitations of this study are as follows: First, the microarray data were unbalanced with respect to the numbers of ccRCC and control tissues, which were restricted in quantity and acquired from the GEO databases. Second, these microarray data comprised relatively few ccRCC samples. Furthermore, only 513 patients were enrolled from the TCGA cohort with corresponding transcriptome data and relatively complete phenotypic data. Third, a prospective cohort was not analyzed, and identification of the underlying mechanisms was not addressed. Fourth, only the mRNA levels of hub genes are shown. Thus, further functional analyses of validated cohorts are required to verify these findings.

In conclusion, the present study identifies DEGs and hub genes that may be involved in earlier recurrence and poor prognosis of ccRCC. The expression levels of *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2,* and *TRIB3* have high prognostic value and may help us better understand the underlying mechanisms of oncogenesis and progression of ccRCC.

Acknowledgements

We would like to present our sincere gratitude to ccRCC patients from the FUSCC cohort. We also would like to thank Cui-Zhu Zhang, Wei-Yi Yang, and Ying Shen for helping with follow-up of patients. We thank Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. This work is supported by Grants from the National Key Research and Development Project (No. 2019YFC1316000) and the National

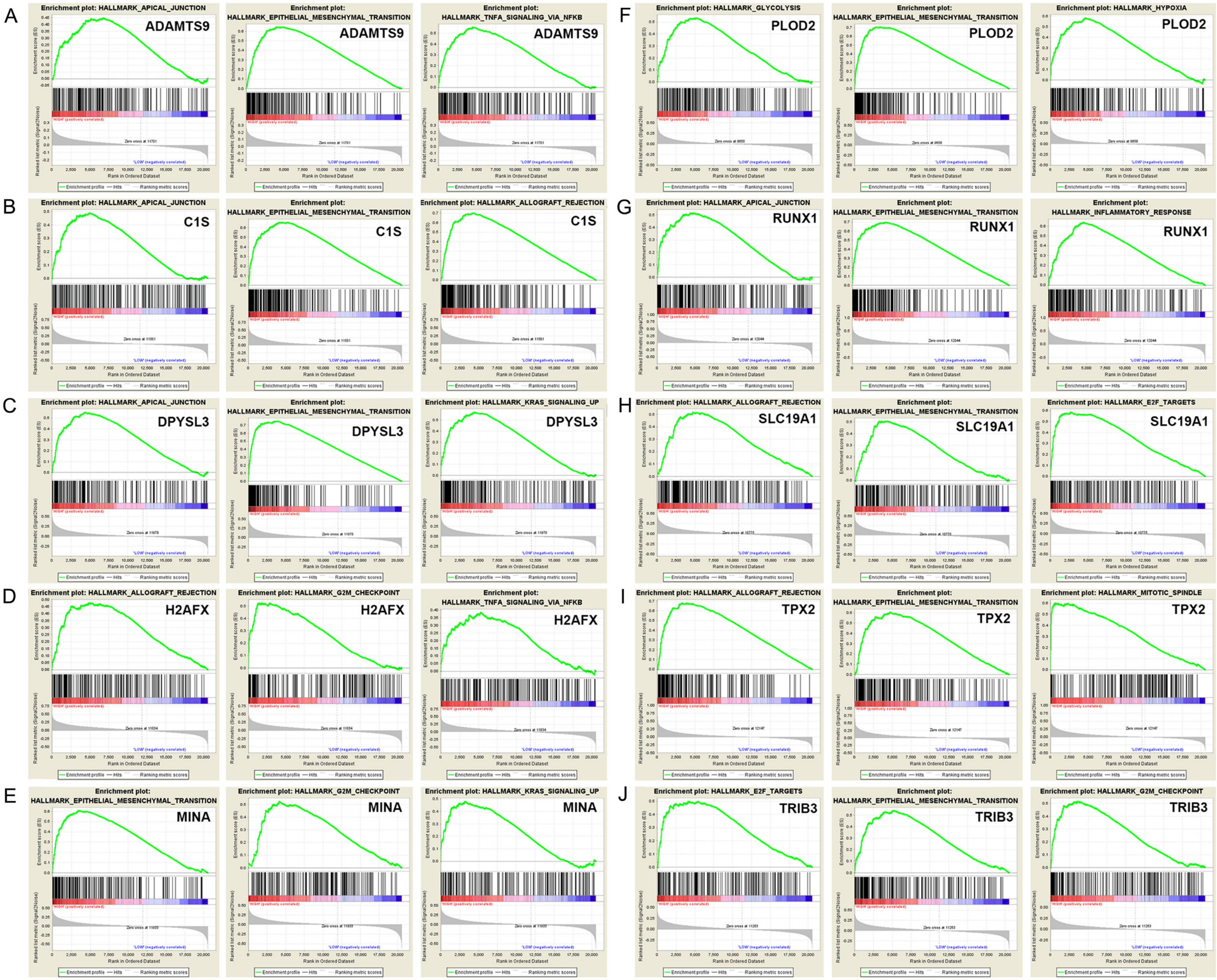

A
Enrichment plot: HALLMARK_APICAL_JUNCTION
ADAMTS9
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
ADAMTS9
Enrichment plot: HALLMARK_TNFA_SIGNALING_VIA_NFKB
ADAMTS9
F
Enrichment plot: HALLMARK_GLYCOLYSIS
PLOD2
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
PLOD2
Enrichment plot: HALLMARK_HYPOXIA
PLOD2
B
Enrichment plot: HALLMARK_APICAL_JUNCTION
C1S
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
C1S
Enrichment plot: HALLMARK_ALLOGRAFT_REJECTION
C1S
G
Enrichment plot: HALLMARK_APICAL_JUNCTION
RUNX1
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
RUNX1
Enrichment plot: HALLMARK_INFLAMMATORY_RESPONSE
RUNX1
C
Enrichment plot: HALLMARK_APICAL_JUNCTION
DPYSL3
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
DPYSL3
Enrichment plot: HALLMARK_KRAS_SIGNALING_UP
DPYSL3
H
Enrichment plot: HALLMARK_ALLOGRAFT_REJECTION
SLC19A1
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
SLC19A1
Enrichment plot: HALLMARK_E2F_TARGETS
SLC19A1
D
Enrichment plot: HALLMARK_ALLOGRAFT_REJECTION
H2AFX
Enrichment plot: HALLMARK_G2M_CHECKPOINT
H2AFX
Enrichment plot: HALLMARK_TNFA_SIGNALING_VIA_NFKB
H2AFX
I
Enrichment plot: HALLMARK_ALLOGRAFT_REJECTION
TPX2
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
TPX2
Enrichment plot: HALLMARK_MITOTIC_SPINDLE
TPX2
E
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
MINA
Enrichment plot: HALLMARK_G2M_CHECKPOINT
MINA
Enrichment plot: HALLMARK_KRAS_SIGNALING_UP
MINA
J
Enrichment plot: HALLMARK_E2F_TARGETS
TRIB3
Enrichment plot: HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
TRIB3
Enrichment plot: HALLMARK_G2M_CHECKPOINT
TRIB3
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'HIGH' (positively correlated)
'LOW' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

Figure 8. GSEA was used to perform hallmark analysis for *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2* and *TRIB3*. GSEA identified 100 DEGs with positive or negative correlations. Among them, we used the hallmarks *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, and *TRIB3* to perform GSEA. Results suggested that the most involved three significant pathways of each oncogene included apical junction, epithelial mesenchymal transition, allograft rejection, G2M checkpoint, mitotic spindle, Kras signaling pathway up, inflammatory response, TNF-alpha signaling via NF-*κB*, etc.

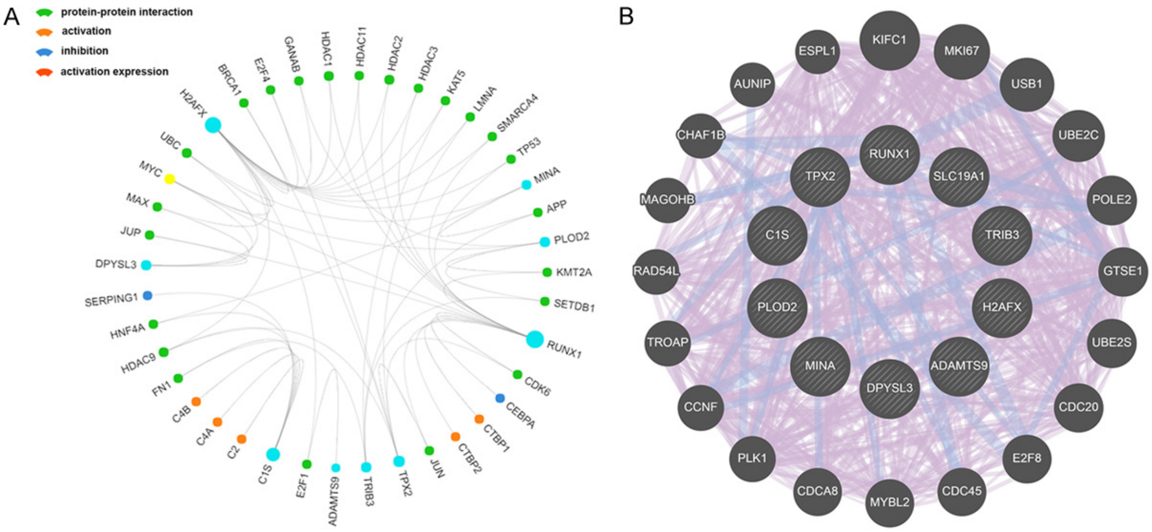


Figure 9. PPI network of nodes from integrated prediction model. A. The most significant genes associated with ten hub genes and their relationship, including protein-protein interaction, activation, inhibition and activation expression, were visualized in a circos plot. B. A detailed PPI network related to this gene sets were constructed to show more interrelated nodes at protein level.

Natural Science Foundation of China (No. 81802525 and 81772706).

Disclosure of conflict of interest

None.

Abbreviations

ccRCC, clear cell renal cell carcinoma; DEGs, differentially expressed genes; PPI, Protein-protein interaction; TCGA, the Cancer Genome Atlas; DFS, disease-free survival; OS, overall survival; HR, hazard ratio; CI, confidence interval; ROC, receiver operating characteristic curve; AUC, Area under curve; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene set enrichment analysis.

Address correspondence to: Drs. Ding-Wei Ye, Hai-Liang Zhang, Yuan-Yuan Qu and Yi-Ping Zhu, Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, P. R. China. Tel: +86-21-64175590-2805; Fax: +86-21-64434556; E-mail: dwyelie@163.com (DWY); zhanghl918@163.com (HLZ); quyy1987@163.com (YYQ); qdzhuyiping@aliyun.com (YPZ)

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018; 68: 7-30.

[2] Gupta K, Miller J, Li J, Russell M and Charbonneau C. Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. Cancer Treat Rev 2008; 34: 193-205.

[3] Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S and Kiemeney LA. The epidemiology of renal cell carcinoma. Eur Urol 2011; 60: 615-21.

[4] Kroeger N, Seligson D, Signoretti S, Yu H, Magyar C, Huang J, Belldegrun A and Pantuck A. Poor prognosis and advanced clinicopathological features of clear cell renal cell carcinoma (ccRCC) are associated with cytoplasmic subcellular localisation of Hypoxia inducible factor-2alpha. Eur J Cancer 2014; 50: 1531-40.

[5] Oliveira R, Ivanovic R, Ramos Moreira Leite K, Viana N, Pimenta R, Junior J, Guimarães V, Morais D, Abe D, Nesrallah A, Srougi M, Nahas

W and Reis S. Expression of micro-RNAs and genes related to angiogenesis in ccRCC and associations with tumor characteristics. BMC Urol 2017; 17: 113.

[6] Grasmann G, Smolle E, Olschewski H and Leithner K. Gluconeogenesis in cancer cells - repurposing of a starvation-induced metabolic pathway? Biochim Biophys Acta Rev Cancer 2019; 1872: 24-36.

[7] Xu W, Xu Y, Wang J, Wan F, Wang H, Cao D, Shi G, Qu Y, Zhang H and Ye D. Prognostic value and immune infiltration of novel signatures in clear cell renal cell carcinoma microenvironment. Aging (Albany NY) 2019; 11: 6999-7020.

[8] Wu J, Xu W, Wei Y, Qu Y, Zhang H and Ye D. An integrated score and nomogram combining clinical and immunohistochemistry factors to predict high ISUP grade clear cell renal cell carcinoma. Front Oncol 2018; 8: 634.

[9] Ho T, Serie D, Parasramka M, Cheville J, Bot B, Tan W, Wang L, Joseph R, Hilton T, Leibovich B, Parker A and Eckel-Passow J. Differential gene expression profiling of matched primary renal cell carcinoma and metastases reveals upregulation of extracellular matrix genes. Ann Oncol 2017; 28: 604-610.

[10] Scelo G, Purdue M, Brown K, Johansson M, Wang Z, Eckel-Passow J, Ye Y, Hofmann J, Choi J, Foll M, Gaborieau V, Machiela M, Colli L, Li P, Sampson J, Abedi-Ardekani B, Besse C, Blanche H, Boland A, Burdette L, Chabrier A, Durand G, Calvez-Kelm F, Prokhortchouk E, Robinot N, Skryabin K, Wozniak M, Yeager M, Basta-Jovanovic G, Dzamic Z, Foretova L, Holcatova I, Janout V, Mates D, Mukeriya A, Rascu S, Zaridze D, Bencko V, Cybulski C, Fabianova E, Jinga V, Lissowska J, Lubinski J, Navratilova M, Rudnai P, Szeszenia-Dabrowska N, Benhamou S, Cancel-Tassin G, Cussenot O, Baglietto L, Boeing H, Khaw K, Weiderpass E, Ljungberg B, Sitaram R, Bruinsma F, Jordan S, Severi G, Winship I, Hveem K, Vatten L, Fletcher T, Koppova K, Larsson S, Wolk A, Banks R, Selby P, Easton D, Pharoah P, Andreotti G, Freeman L, Koutros S, Albanes D, Männistö S, Weinstein S, Clark P, Edwards T, Lipworth L, Gapstur S, Stevens V, Carol H, Freedman M, Pomerantz M, Cho E, Kraft P, Preston M, Wilson K, Gaziano J, Sesso H, Black A, Freedman N, Huang W, Anema J, Kahnoski R, Lane B, Noyes S, Petillo D, Teh B, Peters U, White E, Anderson G, Johnson L, Luo J, Buring J, Lee I, Chow W, Moore L, Wood C, Eisen T, Henrion M, Larkin J, Barman P, Leibovich B, Choueiri T, Lathrop G, Rothman N, Deleuze J, McKay J, Parker A, Wu X, Houlston R, Brennan P and Chanock S. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nat Commun 2017; 8: 15724.

[11] Ni D, Ma X, Li H, Gao Y, Li X, Zhang Y, Ai Q, Zhang P, Song EL, Huang QB, Fan Y and Zhang X. Downregulation of FOXO3a promotes tumor metastasis and is associated with metastasis-free survival of patients with clear cell renal cell carcinoma. Clin Cancer Res 2014; 20: 1779-90.

[12] Clark D, Dhanasekaran S, Petralia F, Pan J, Song X, Hu Y, Leprevost F, Reva B, Lih T, Chang H, Ma W, Huang C, Ricketts C, Chen L, Krek A, Li Y, Rykunov D, Li Q, Chen L, Ozbek U, Vasaikar S, Wu Y, Yoo S, Chowdhury S, Wyczalkowski M, Ji J, Schnaubelt M, Kong A, Sethuraman S, Avtonomov D, Ao M, Colaprico A, Cao S, Cho K, Kalayci S, Ma S, Liu W, Ruggles K, Calinawan A, Gümüş Z, Geiszler D, Kawaler E, Teo G, Wen B, Zhang Y, Keegan S, Li K, Chen F, Edwards N, Pierorazio P, Chen X, Pavlovich C, Hakimi A, Brominski G, Hsieh J, Antczak A, Omelchenko T, Lubinski J, Wiznerowicz M, Linehan W, Kinsinger C, Thiagarajan M, Boja E, Mesri M, Hiltke T, Robles A, Rodriguez H, Qian J, Fenyö D, Zhang B, Ding L, Schadt E, Chinnaiyan A, Zhang Z, Omenn G, Cieslik M, Chan D, Nesvizhskii A, Wang P and Zhang H; Clinical Proteomic Tumor Analysis Consortium. Integrated proteogenomic characterization of clear cell renal cell carcinoma. Cell 2019; 179: 964-983, e31.

[13] Wu J, Vallenius T, Ovaska K, Westermarck J, Mäkelä T and Hautaniemi S. Integrated network analysis platform for protein-protein interactions. Nat Methods 2009; 6: 75-7.

[14] Bapat S, Krishnan A, Ghanate A, Kusumbe A and Kalra R. Gene expression: protein interaction systems network modeling identifies transformation-associated molecules and pathways in ovarian cancer. Cancer Res 2010; 70: 4809-19.

[15] Sharan R, Ulitsky I and Shamir R. Network-based prediction of protein function. Mol Syst Biol 2007; 3: 88.

[16] Edgar R, Domrachev M and Lash A. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002; 30: 207-10.

[17] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C and Jensen L. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41: D808-15.

[18] Smoot M, Ono K, Ruscheinski J, Wang P and Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 2011; 27: 431-2.

[19] Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M,

Rubin G and Sherlock G. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 2000; 25: 25-9.

[20] Kanehisa M. The KEGG database. Novartis Found Symp 2002; 247: 91-101; discussion 101-3, 119-28, 244-52.

[21] Huang D, Sherman B, Tan Q, Collins J, Alvord W, Roayaei J, Stephens R, Baseler M, Lane H and Lempicki R. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 2007; 8: R183.

[22] Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W, Pagès F, Trajanoski Z and Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009; 25: 1091-3.

[23] Cerami E, Gao J, Dogrusoz U, Gross B, Sumer S, Aksoy B, Jacobsen A, Byrne C, Heuer M, Larsson E, Antipin Y, Reva B, Goldberg A, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-4.

[24] Camp R, Dolled-Filhart M and Rimm D. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004; 10: 7252-9.

[25] Asplund A, Edqvist P, Schwenk J and Pontén F. Antibodies for profiling the human proteome-the human protein atlas as a resource for cancer research. Proteomics 2012; 12: 2067-77.

[26] Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, Paulovich A, Pomeroy S, Golub T, Lander E and Mesirov J. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-50.

[27] Yusenko M, Ruppert T and Kovacs G. Analysis of differentially expressed mitochondrial proteins in chromophobe renal cell carcinomas and renal oncocytomas by 2-D gel electrophoresis. Int J Biol Sci 2010; 6: 213-24.

[28] Liao L, Testa J and Yang H. The roles of chromatin-remodelers and epigenetic modifiers in kidney cancer. Cancer Genet 2015; 208: 206-14.

[29] Gumz M, Zou H, Kreinest P, Childs A, Belmonte L, LeGrand S, Wu K, Luxon B, Sinha M, Parker A, Sun L, Ahlquist D, Wood C and Copland J. Secreted frizzled-related protein 1 loss contributes to tumor phenotype of clear cell renal cell carcinoma. Clin Cancer Res 2007; 13: 4740-9.

[30] Yang J, Shi S, Xu W, Qiu Y, Zheng J, Yu K, Song X, Li F, Wang Y, Wang R, Qu Y, Zhang H and Zhou X. Screening, identification and validation of CCND1 and PECAM1/CD31 for predicting prognosis in renal cell carcinoma patients. Aging (Albany NY) 2019; 11: 12057-12079.

[31] Dondeti V, Wubbenhorst B, Lal P, Gordan J, D'Andrea K, Attiyeh E, Simon M and Nathanson K. Integrative genomic analyses of sporadic clear cell renal cell carcinoma define disease subtypes and potential new therapeutic targets. Cancer Res 2012; 72: 112-21.

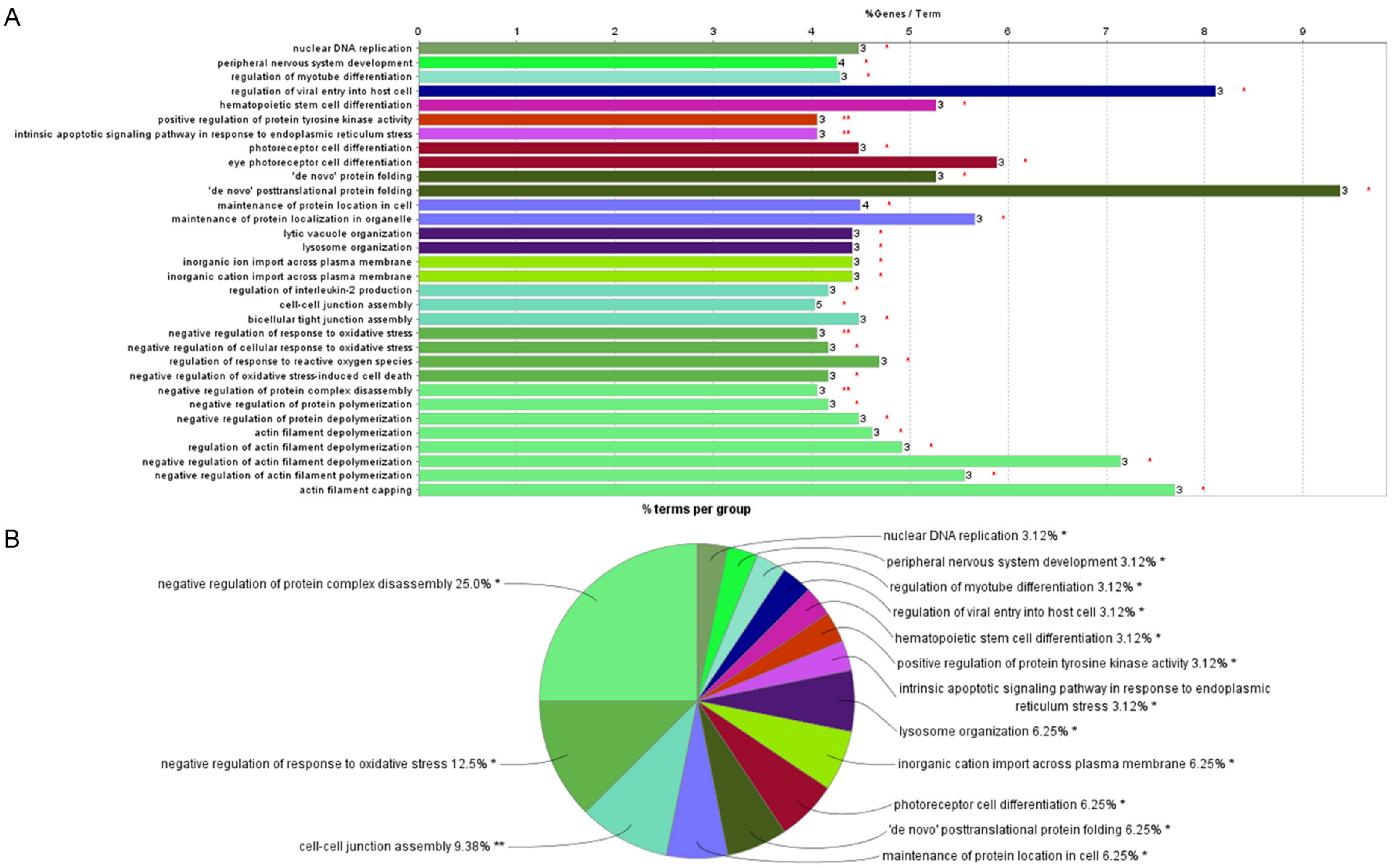
[32] Sorbellini M, Kattan M, Snyder M, Reuter V, Motzer R, Goetzl M, McKiernan J and Russo P. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma. J Urol 2005; 173: 48-51.

[33] Zisman A, Pantuck A, Dorey F, Said J, Shvarts O, Quintana D, Gitlitz B, deKernion J, Figlin R and Belldegrun A. Improved prognostication of renal cell carcinoma using an integrated staging system. J Clin Oncol 2001; 19: 1649-57.

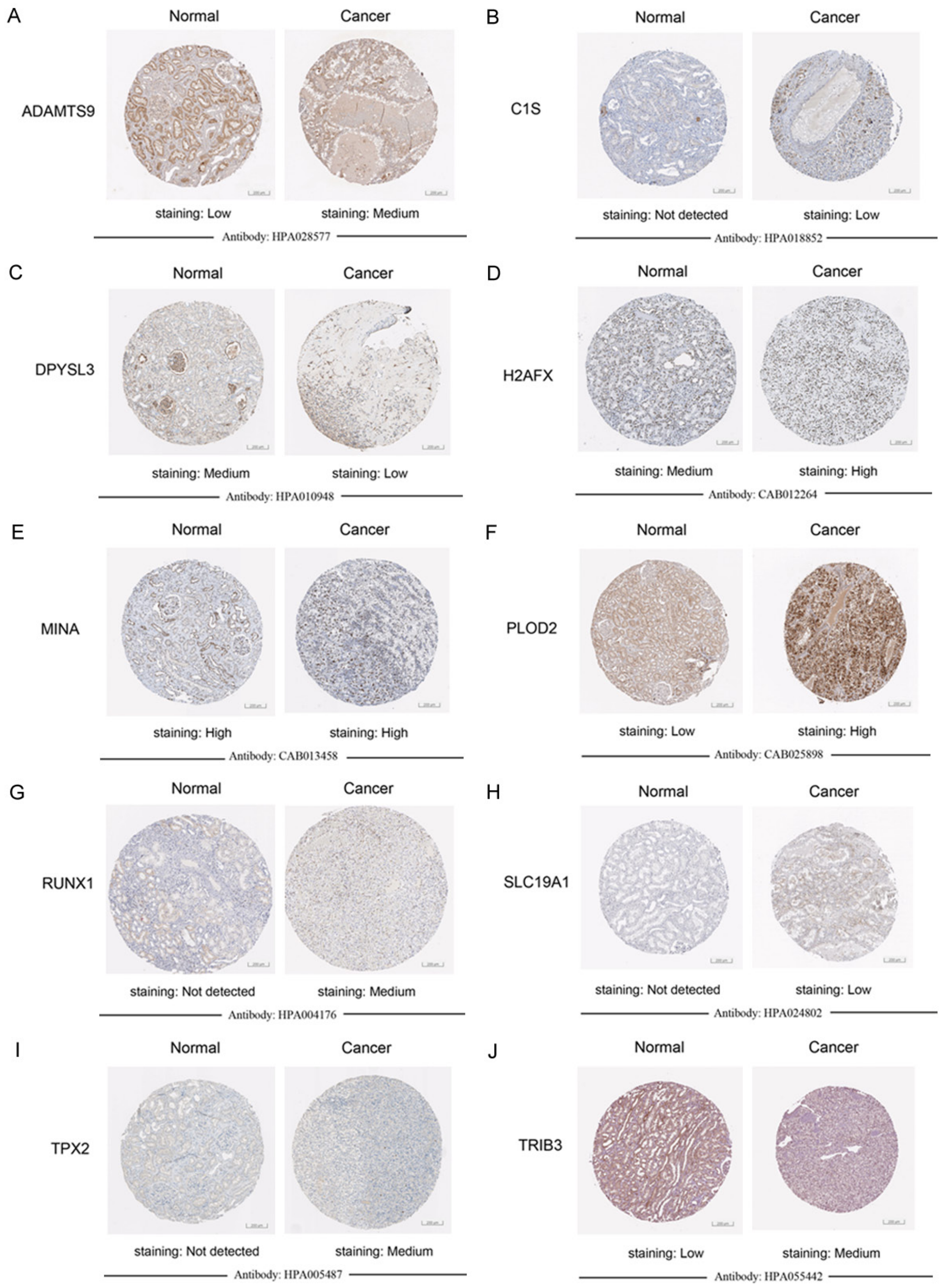
[34] Leibovich B, Han K, Bui M, Pantuck A, Dorey F, Figlin R and Belldegrun A. Scoring algorithm to predict survival after nephrectomy and immunotherapy in patients with metastatic renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer 2003; 98: 2566-75.

[35] Qu L, Wang Z, Chen Q, Li Y, He H, Hsieh J, Xue S, Wu Z, Liu B, Tang H, Xu X, Xu F, Wang J, Bao Y, Wang A, Wang D, Yi X, Zhou Z, Shi C, Zhong K, Sheng Z, Zhou Y, Jiang J, Chu X, He J, Ge J, Zhang Z, Zhou W, Chen C, Yang J, Sun Y and Wang L. Prognostic value of a long non-coding RNA signature in localized clear cell renal cell carcinoma. Eur Urol 2018; 74: 756-763.

[36] Wei J, Feng Z, Cao Y, Zhao H, Chen Z, Liao B, Wang Q, Han H, Zhang J, Xu Y, Li B, Wu J, Qu G, Wang G, Liu C, Xue W, Liu Q, Lu J, Li C, Li P, Zhang Z, Yao H, Pan Y, Chen W, Xie D, Shi L, Gao Z, Huang Y, Zhou F, Wang S, Liu Z, Chen W and Luo J. Predictive value of single-nucleotide polymorphism signature for recurrence in localised renal cell carcinoma: a retrospective analysis and multicentre validation study. Lancet Oncol 2019; 20: 591-600.



Supplementary Figure 1. The detailed functional annotations and classification pie charts using CluePedia from Cytoscape are provided. Of the ClueGO analysis, 25.0% of terms belonged to the negative regulation of protein complex disassembly, 12.5% to negative regulation of response to oxidative stress, 9.38% to cell-cell junction assemble, 6.25% to maintenance of protein location in cell, 6.25% to photoreceptor cell differentiation, 6.25% inorganic cation import across plasma membrane, and 6.25% to lysosome organization.



Supplementary Figure 2. IHC staining was used to describe differential proteomic expression of *ADAMTS9, C1S, DPYSL3, H2AFX, MINA, PLOD2, RUNX1, SLC19A1, TPX2*, *TRIB3*. Antibody staining density described staining status between normal and tumor samples.