frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/fonc.2022.982110

OPEN ACCESS

EDITED BY
Paul J. Galardy,
Mayo Clinic, United States

REVIEWED BY
Wafaa M. Rashed,
Children's Cancer Hospital Egypt, Egypt
Belamy B. Cheung,
Children's Cancer Institute Australia, Australia

*CORRESPONDENCE
Xiaogeng Deng
dengxg@mail.sysu.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Pediatric Oncology, a section of the journal Frontiers in Oncology

RECEIVED 30 June 2022
ACCEPTED 05 October 2022
PUBLISHED 21 October 2022

CITATION
Lou L, Chen L, Wu Y, Zhang G, Qiu R, Su J, Zhao Z, Lu Z, Liao M and Deng X (2022) Identification of hub genes and construction of prognostic nomogram for patients with Wilms tumors. *Front. Oncol.* 12:982110. doi: 10.3389/fonc.2022.982110

Identification of hub genes and construction of prognostic nomogram for patients with Wilms tumors

Lei Lou[1,2†], Luping Chen[1,3†], Yaohao Wu[1†], Gang Zhang[2†], Ronglin Qiu[1], Jianhang Su[1], Zhuangjie Zhao[1], Zijie Lu[1], Minyi Liao[1] and Xiaogeng Deng[1*]

[1]Department of Pediatric Surgery, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China, [2]Department of Pediatric Surgery, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou Medical University, Guangzhou, China, [3]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China

Background: In children, Wilms' tumors are the most common urological cancer with unsatisfactory prognosis, but few molecular prognostic markers have been discovered for it. With the rapid development of high-throughput quantitative proteomic and transcriptomic approaches, the molecular mechanisms of various cancers have been comprehensively explored. This study aimed to uncover the molecular mechanisms underlying Wilms tumor and build predictive models by use of microarray and RNA-seq data.

Methods: Gene expression datasets were downloaded from Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Gene Expression Omnibus (GEO) databases. Bioinformatics methods wereutilized to identified hub genes, and these hub genes were validated by experiment. Nomogram predicting OS was developed using genetic risk score model and clinicopathological variables.

Results: CDC20, BUB1 and CCNB2 were highly expressed in tumor tissues and able to affect cell proliferation and the cell cycle of SK-NEP-1 cells. This may reveal molecular biology features and a new therapeutic target of Wilms tumour.7 genes were selected as prognostic genes after univariate, Lasso, and multivariate Cox regression analyses and had good accuracy, a prognostic nomogram combined gene model with clinical factors was completed with high accuracy.

Conclusions: The current study discovered CDC20,BUB1 and CCNB2 as hub-genes associated with Wilms tumor, providing references to understand the pathogenesis and be considered a novel candidate to target therapy and construct novel nomogram, incorporating both clinical risk factors and gene

model, could be appropriately applied in preoperative individualized prediction of malignancy in patients with Wilms tumor.

KEYWORDS

wilms tumor, hub gene, risk stratification, prognostic model, nomogram

Introduction

Malignant renal tumors comprise 5% of all cancers occurring before the age of 15 years (1). The most common form of urologic cancer in children is Wilms tumor (2).Current treatment strategies *for* Wilms tumor include surgery, chemotherapy and radiotherapy. In high-income countries, 90% of patients with this tumor survive (3). While the OS in high, 20% of patients relapse after first-line therapy and up to 25% of survivors report severe late morbidity of treatment (4, 5).And the treatment of children with bilateral, high-risk tumors remains challenging (6). Novel biomarkers, therefore, are urgently needed in order to predict patients' prognosis and develop novel target-specific therapies. But the introduction of biology-driven approaches to risk stratification and new drug development has been slower in Wilms tumor than in other childhood tumors (7).It is very important to improving prognosis for Wilms tumor patients by improving treatment based on clinical and biological risk factors, and further stratification of current treatment options based on tumor biology (6, 8).

Many researchers are devoted to study about Wilms tumor. In the earlier study, Predisposition loci at 11p13(WT1),17q21 (FWT1), 19q13 (FWT2) and 11p15 were identified by genetic linkage studies of familial Wilms tumor (9–11).With the development of modern genetic technology, the diverse genetic landscape of Wilms tumor is now beginning to be unveiled. Several of the variant genes that have been identified in Wilms tumor which are involved in histone modification during nephrogenesis (BCOR, MAP3K4, BRD7, CREBBP and HDAC4) (12). Recently, Putative predisposition genes include REST, CHEK2, EP300, PALB2 and ARID1A were identified (13, 14). Somatic mutations in SMARCA4, ARID1A, BCORL1, MAP3K4, NONO and MAX has also been discovered in Wilms tumor (15, 16). At present, there are several hot research topics like simultaneous allele loss of 1p and 16q, MYCN gain, loss of p53 function (17–20).These investigations provide novel insights to molecular mechanism of tumorigenesis and development of Wilms tumor. But the biological mechanism and prognostic effect still need further explore.

With the rapid advances in genomic and bioinformatics analysis, a large volume of cancer genomics data is being produced, benefiting cancer research. TARGET (Therapeutically Applicable Research to Generate Effective Treatments) aims to identify driver mutations and therapeutic targets for high-risk pediatric tumors through comprehensive integrative genomics (21). GEO (Gene Expression Omnibus) is another authoritative oncology database. Both of them comprised of massive sequencing data and clinic data about various tumors. Also, bioinformatics analysis of data matrix is widely used nowadays to determine differentially expressed genes (DEGs) and perform various analyses. By combining multiple databases, it is possible to integrate data from independent studies and obtain a greater number of samples for analysis, thus improving the rigor and accuracy of the results. Given the above, we integrated the patient data affected by Wilms tumor from TARGET and GEO database for further analysis by bioinformatics and experiments.

In this study, we tried to identify hub genes to provide further insights into the occurrence and development of Wilms tumor. Also, we wanted establish a simple and effective model to predict the clinical classification and risk stratification of patients with Wilms tumor.

Materials and methods

Patients tissues and cell cultures

The RNA sequencing data and relating clinical information of Wilms tumor patients were downloaded from TARGET data portal (https://ocg.cancer.gov/programs/target). The inclusion criteria were as follows: (i) gene expression data of WT were available in the database; (ii) clinical data including gender, age, stage, and overall survival were complete; Finally, 125 patients were enrolled. For finding the Wilms tumor gene expression datasets,we performed systematic searches within the GEO database (https://www.ncbi.nlm.nih.gov/geo/). Three independent datasets were downloaded from GEO for further comprehensive analysis in this study, including GSE19249, GSE66405 and GSE73209 datasets. The GSE19249 include 6 tumor samples and 6 normal samples; the GSE66405 include 28 tumor samples and 4 normal samples; the GSE73209 include 32 tumor samples and 6 normal samples (Table 1).

This study comprised 13 patients with Wilms tumor who were diagnosed between 2011 and 2015. Tissue specimens were

TABLE 1 Details for 3 GEO dataset.

GSE	Authour	Platform	Tumour samples	Normal sample
GSE19249	Abdueva	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus	6	6
GSE66405	Ludwig	Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381	28	4
GSE73209	Karlsson	Illumina HumanHT-12 V4.0 expression beadchip	32	6

obtained from the pathology archives of the Sun Yat-Sen Memorial Hospital. Inclusion criteria were as follows: (1) lesions showing histologic features of Wilms tumor; (2) patients aged ≤14 years at the time of diagnosis; (3) availability of sufficient tissue for genomic assays; The study material included 18 formalin-fixed paraffin-embedded (FFPE) tissue specimens: primary tumors (n = 13), and normal tissues (n = 5).

SK-NEP-1 cell line were cultured in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Bioind), streptomycin (100 μg/mL), and penicillin (100 μg/mL) (Gibco) and incubated at 37°C in a humidified chamber with 5% CO2.

Data pre-processing

Three series matrix files from GEO database were annotated with an official gene symbol using the data table of the microarray platform, and the gene expression matrix files were obtained respectively. The three gene expression matrix files were merged into one by Perl. Gene probes without gene symbols were eliminated and genes with more than one probe were averaged. To ensure the integrity and comparability of the datasets, the batch normalization of merged data was pre-processed by sva package using the R language (22). Patients' mRNA expression data were downloaded from TARGET database and merged as one matrix through Perl software.

Identification of DEGs and enrichment analysis

A package called Limma (3.18.2) from R Bioconductor was used to identify DEGs in GEO data matrix and TARGET data matrix respectively (23).The threshold of DEGs in the data set was set as the adjusted value of $p < 0.05$ and $|\log_2 FC| > 1$ (FC: fold change). The FunRich software was used for analysis of DEGs that overlapped between the two databases and created the Venn diagram (24).Overlapping DEGs visualization was done using a volcano map and heatmap using gglot2 and the pheatmap package (25). We conducted gene ontology analysis (GO) and Kyoto encyclopedia of genes and genomes analysis (KEGG) using the ClusterProfiler package to further understand the overlapping DEGs potential biological processes, cell components, molecular function, and enrichment signaling pathways (26–28). It was deemed significant when p values were less than 0.05.

Construction of protein-protein interaction network

The STRING website (http://string-db.org)was used to identify protein–protein interactions (PPIs) of the overlapping DEGs (29). The interactions with reliability scores more than 0.7 were selected for analysis. After that, Cytoscape software was used for PPI network analysis and visualization (30). we utilized the plugin MCODE to screen hub modules and cytoHubba to calculate the degree of genes in the PPI network and obtain the top 10 hub genes with the highest degree for further studies (31).

qRT-PCR assay

Total RNA from 4 tumor tissues and 4 normal tissues was extracted by using TRIzol reagent. The specific steps were in strict accordance with the protocol. The Nanodrop 2000 UV spectrophotometer was used to detect the extracted RNA concentration and purity. Calculate the volume required for 100ng of RNA which were used for reverse transcription. The cDNA was synthesized by *reverse transcription* for further *experiments*. Primer sequences were designed and produced by Servicebio Company, The GAPDH gene was used as the reference control. *Reverse transcription* was performed in a total of 20ul volume containing 100 ng RNA, 3 *μl* 5x gDNA digester Mix, 5ul 4x HifairIII Supermix plus and RNase-free water supplemented to 20 μl. The reaction conditions were as follows: *25*°C for 5 min, 55°C for 15 min, 85°C for 5 min., 4°C ∞. The synthesized cDNA was stored *in* an ultra-low temperature *freezer* (−20°C). A total of 10 μl of the real-time PCR system was prepared containing: 2 μl of cDNA,1 μl of primer mix,5 μl of SYBR Green Master(Low Rox), and 2 μl of DEPC H_2O supplemented. Cycling mode was set as: 40 cycles of 95 °C 15 s, 60 °C 30 s. Relative expression was analyzed using $2^{-\Delta\Delta Ct}$ method and GAPDH served as internal control. All experiments were conveyed in triplicates. Primer sequences are shown in Table 2.

TABLE 2 Primers used for qRT-PCR.

Name	Sequence (5'-3')
BUB1-F	TCATTCATGGAGACATTAAAC
BUB1-R	CTGAGCATCTCAACACACTG
CDC20-F	GCTTTGAACCTGAACGGTTTTG
CDC20-R	TCTGGCGCATTTTGTGGTTTT
CCNB2-F	CCGACGGTGTCCAGTGATTT
CCNB2-R	TGTTGTTTTGGTGGGTTGAACT
MAD2L1-F	GTTCTTCTCATTCGGCATCAACA
MAD2L1-R	GAGTCCGTATTTCTGCACTCG
KIF11-F	TATTGAATGGGCGCTAGCTT
KIF11-R	TCGTCTGCGAAGAAGAAAGA
PBK-F	AGACCCTAAAGATCGTCCTTCTG
PBK-R	GTGTTTTAAGTCAGCATGAGCAG
TOP2A-F	CTAGTTAATGCTGCGGACAACA
TOP2A-R	CATTTCGACCACCTGTCACTT
ASPM-F	GGCCCTAGACAACCCTAACGA
ASPM-R	AGCTTGGTGTTTCAGAACATCA
KIF2C-F	GTGTCTGTGTTAGGAAACGCCC
KIF2C-R	CACTTGCTAGGAATGGAAATCACAT
CDCA8-F	GAAGGGCAGTAGTCGGGTG
CDCA8-R	TCACGGTCGAAGTCTTTCAGA
H-GAPDH-S	GGAAGCTTGTCATCAATGGAAATC
H-GAPDH-A	TGATGACCCTTTTGGCTCCC

IHC staining

A total of 20 formalin-fixed, paraffin-embedded biopsy tissues were available included 15 wilms tumor tissue and 5 normal tissues. The paraffin sections were deparaffinized by dimethylbenzene and rehydrated by graded alcohol. Then, *EDTA* antigen restore solution was used to repair antigens on slices in a microwave oven at the condition of moderate heat for 8 min, heat preservation for 8 min, and moderate-low heat for 8 min followed by natural cooling. After cooling, the sections were put in 0.3% hydrogen peroxide solution for 20 min to block endogenous peroxidase. After washing with PBS, blocking was performed with 3% BSA for 30 min at room temperature. The blocking solution was discarded, and *the* primary *antibody* was added and incubated overnight at *4*°C in a wet box. *The concentrations* used for *the primary antibodies were* at 1:50 (anti-BUB1, Abcam, ab195268), 1:200 (anti-CDC20, Abcam, ab183479) and 1:100 (anti-CNNB2, Abcam, ab15622) dilutions respectively. Following washes with *PBS*, a corresponding secondary antibody (1: 200, Servicebio, GB23303) was added, incubated at room temperature for 1 h, and washed with PBS.DBA solution was added, color development was checked under the microscope and the solution was washed with distilled water after satisfactory color development. Hematoxylin counterstain, water wash and wash fully with water after differentiation till *return blue*; routine dehydration and transparence; neutral gum mounting. Brownish yellow staining detected by microscopy indicated positive expression.

Oligonucleotide transfection

Small interfering RNAs (siRNAs) were purchased from GenePharma (Shanghai, China) based on the following sequences: BUB1-specific siRNA: 5′-GGCCUGCAUUUCUAA GAAUTT-3′ (sense), 5′-AUUCUUAGAAAUGCAGGCCTT-3′ (anti-sense); CDC20-specific siRNA: 5′-GGAAGACCUGCC GUUACAUTT-3′ (sense), 5′-AUGUAACGGCAGGUCUUCC TT-3′(anti-sense);CCNB2-specific siRNA:5′-GCAGCACAU GGCCAAGAAUTT-3′ (sense), 5′-AUUCUUGGCCAUGUG CUGCTT-3′(anti-sense);NC: 5′-UUCUCCGAACGUGUCACG UTT-3′ (sense), 5′-ACGUGACACGUUCGGAGAATT-3′ (anti-sense).Cell transfection was conducted using the Lipofectamine RNAiMax Transfection Reagent (Life Technologies).

Cell proliferation assay

SK-NEP-1 cells were reseeded into 96-well plates (5 × 103 cells per well), and cell proliferation was assessed by the CCK-8 assay (Beyotime). The absorbance of each well was measured at a wavelength of 450 nm using a SPARK 10 M spectrophotometer (Tecan, Austria).

5-Ethynyl -2′ -deoxyuridine assays

EdU assays were performed with the Cell-Light EdU DNA Cell Proliferation Kit (RiboBio, Guangzhou, China). The cells (5 × 103 per well) were seeded into each well of a 96-well plate and incubated for 48 h. Next, the cells were incubated with 100 ul of EdU for 2 h, fixed with 500 μl of 4% paraformaldehyde, and stained with Apollo Dye Solution. DAPI staining was used to identify the nucleic acid. Images of the cells were obtained with an inverted fluorescence microscope (Carl Zeiss, Jena, Germany), and the proportion of EdU-positive cells was calculated.

Cell cycle

For cell cycle assays, cells were fixed overnight at 4°C in 70% ethanol, stained with propidium iodide, and analyzed with a

FACS can flow cytometer (BD Biosciences, San Jose, CA, USA) and ModFit 3.0 software (Verity Software House, Topsham, ME, USA).

Construction of 7-gene prognostic model and nomogram

OS related DEGs were identified by univariate COX regression, *lasso* regression and multivariate *Cox* regression (P<0.05) (32, 33). Only genes with nonzero coefficients in the lasso regression model were chosen to further calculate the risk score. Using the formula below, we calculated the risk score for each patient: risk score = expression level of gene1 × j1 + expression level of gene2 × j2 + … + expression level of gene x × j x, where j represents the coefficient. As a cut-off value, the median risk score was used to subdivide Wilms tumor patients into two risk groups - a high-risk group and a low-risk group.

Thereafter, the R package "survival" was used for Kaplan-Meier (KM) analysis to compare OS between the two groups. The R package "time ROC" was used for ROC curve analysis to evaluate the prognostic model's specificity and sensitivity. In the AUC, 0.5 indicated no discriminating ability, while 1 indicated perfect discrimination. 7-gene prognostic model was constructed.

Building a predictive nomogram

Univariate and multivariate Cox regression analyses were performed to identify the independent prognostic factors among 7-gene prognostic model and clinical characteristics to build the nomogram. "rms" R package was used to assess the prognostic nomogram probability of 0.5-, 1-, and 3-year OS for TARGET-WT patients. The discrimination performance of the nomogram was quantitatively assessed by the C-index and the AUC

Result and discussion

Identification of overlapping DEGs

After data matrix from TARGET and GEO database were merged respectively, DEGs between normal and tumor were identified using R language. In total, 841 DEGs (635 downregulated and 206 upregulated mRNAs),5584 DEGs (4437 downregulated and 5327 upregulated mRNAs) were screened from the GEO data matrix and TARGET data matrix (Figures 1A, B). Among them, 688 DEGs overlapped; 191 were up-regulated, and 497 were down-regulated just like the Venn diagram show. Meanwhile, these overlapping DEGs were visualized in volcano plots (Figure 1C).

Enrichment analysis of the overlapping DEGs

To explore the underlying biological process correlated to WT, the overlapping DEGs analysis were conducted using GO and KEGG enrichment analysis. For GO-biological process (GO-BP) enrichment analysis, the results showed that the genes involved in small molecule catabolic process, urogenital system development, renal system development etc. were significantly enriched. For GO-cellular component (GO-CC) enrichment analysis, the genes involved in apical part of cell, apical plasma membrane, basal plasma membrane, etc. were significantly enriched. For GO-molecular function (GO-MF) enrichment analysis, the genes involved in active transmembrane transporter activity, metal ion transmembrane transporter activity etc. were significantly enriched (Figures 2A, B). The results of the KEGG pathway analysis showed that the genes involved in Carbon metabolism, PPAR signaling pathway, etc. were significantly enriched (Figures 2C, D).

A total of 10 hub genes were selected by the PPI network

The STRING database was applied to construct a PPI network of these 688 genes to explore the core modules and hub genes that played the most importance in modular genes (Figure 3A). PPI network visualization was performed using Cytoscape. The plug-in MCODE in Cytoscape was used to detect the dense connection region of the molecular interaction network, of which had a high possibility of being involved in some important biological processes. Finally, a total of sixteen important modules were obtained, and the module with the highest score was screened out(Figure 3C); top 30 hub genes were screened out according to the node degree (Figure 3B).

Hub genes expression levels in Wilms tumor tissue and normal tissue

The expression of 10 hub genes in tumor and normal tissues was examined using Real-time PCR. Experiment was repeated more than *3* times. In tumor tissues compared with the normal tissues, the mRNA expression of CDC20 (<0.000001) 、CNNB2 (<0.01) 、BUB1(<0.0009) was significantly increased (Figure 4A). Statistical significance among samples was evaluated by paired t-test (*P<0.05).

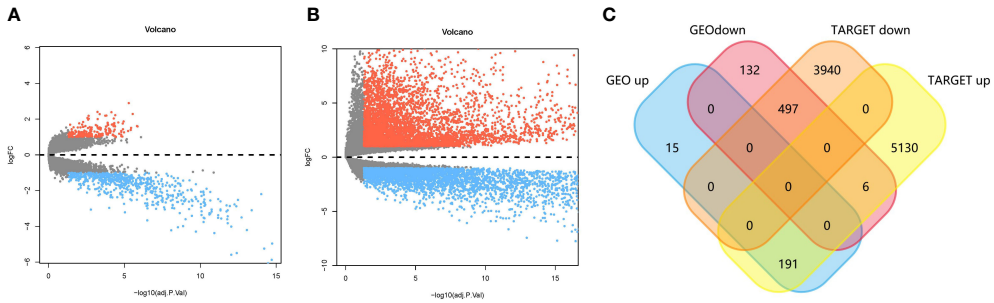

FIGURE 1
(A) and **(B)** Volcano plot of differentially expressed genes of Wilms tumor patients from GEO and TARGET database. Genes with P < 0.05 and logFC >1 or <-1 were considered as DEGs in each series. Blue, down-regulated genes; Gray, non-differential genes; Red, up-regulated genes **(C)** Venn diagram showing the intersection of the DEGs identified in two databases.

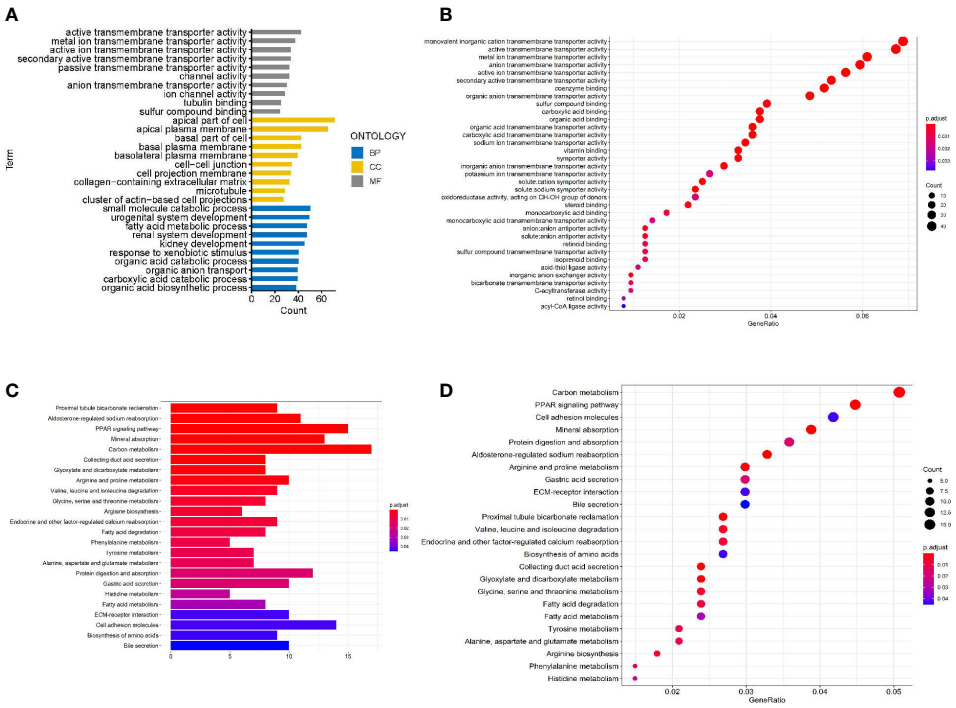

FIGURE 2
Bar charts and bubble map of GO and KEGG pathway analyses. **(A)** GO enrichment bar chart of DEGs: The 10 most significant enriched terms in biological process, cellular component, and molecular function are select. **(B)** GO enrichment bubble map of DEGs. Bubble size represents the number of enriched genes, and bubble color difference represents the significance of target gene enrichment. **(C)** KEGG enrichment analysis bar chart of DEGs. The top 24 pathways from the KEGG enrichment analysis ranked by p-value. **(D)** KEGG enrichment bubble map of DEG.

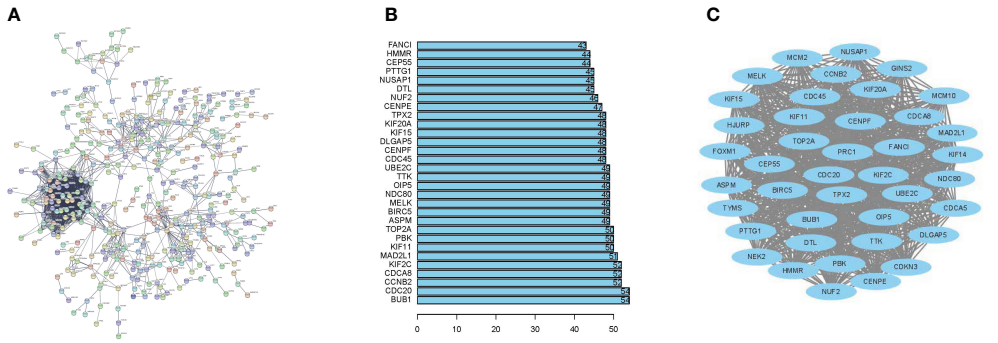

FIGURE 3
Protein-protein interaction (PPI) network, hub gene screening, and hub cluster. **(A)** PPI network (688 DEGs filtered into the PPI network that contained 685 nodes and 1703 edges). **(B)** The predicted association rank (from low to high) of the top 30 genes in the PPI network. **(C)** The key modules identified using the MCODE plugin of Cytoscape, which contains 41 nodes/791 edges.

Moreover, the protein level of CDC20、CNNB2、BUB1 was evaluated by IHC. Brown represents DAB positive immunoreactivity and the blue color is the hematoxylin counterstain. Results from immunohistochemistry showed that CDC20、CNNB2、BUB1 protein was upregulated in tumor tissue (Figure 4B).

Effect of CDC20, CNNB2, and BUB1 knockdown on the proliferation of SK-NEP-1 cells

To determine the effect of CDC20, CNNB2, and BUB1 on SK-NEP-1 cells, CCK-8 and EdU assay was conducted in SK-NEP-1 cells following infection with si-CDC20, si-CNNB2, si-BUB1 and si-NC. The results revealed that CDC20 and CNNB2 knockdown significantly inhibited the proliferation of Wilms tumor cells (*** $p < 0.001$, ** $p < 0.01$, * $P<0.05$; Figures 5A-D).

Effect of CDC20, CNNB2, and BUB1 knockdown on the cell cycle of Wilms tumor cells

To further investigate the role for CDC20, CNNB2, BUB1in tumor growth, a cell cycle analysis was performed using flow cytometry. SK-NEP-1 cells infected with si-CDC20, si-CNNB2, si-BUB1 and si-NC exhibited a notable increase in the percentage of cells in the G0/G1 phase compared with the si-NC groups. the proportion of cells in the S phase showed a variable change depending on the cell type, and the proportion of cells in the G2/M phase also significantly decreased (Figures 5E-H).

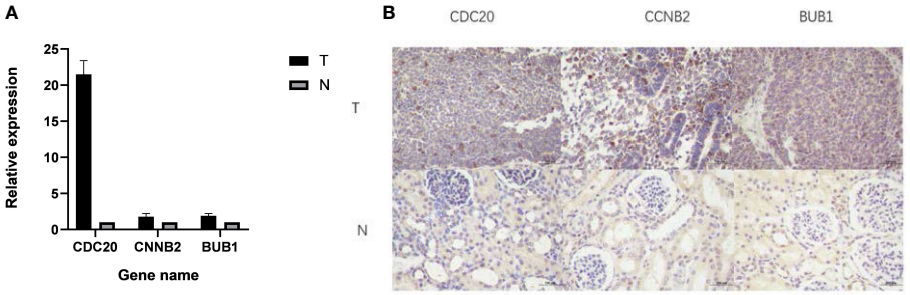

FIGURE 4
CDC20, CCNB2 and BUB1 are highly expressed in tumor tissue. **(A)** Expression levels of CDC20, CNNB2 and BUB1 mRNA determined by real-time PCR in Wilms tumor (T) and normal (N) tissues. **(B)** Expression levels of CDC20,CNNB2 and BUB1 protein determined by IHC in Wilms tumor (T) and normal (N) tissues.

10.3389/fonc.2022.982110

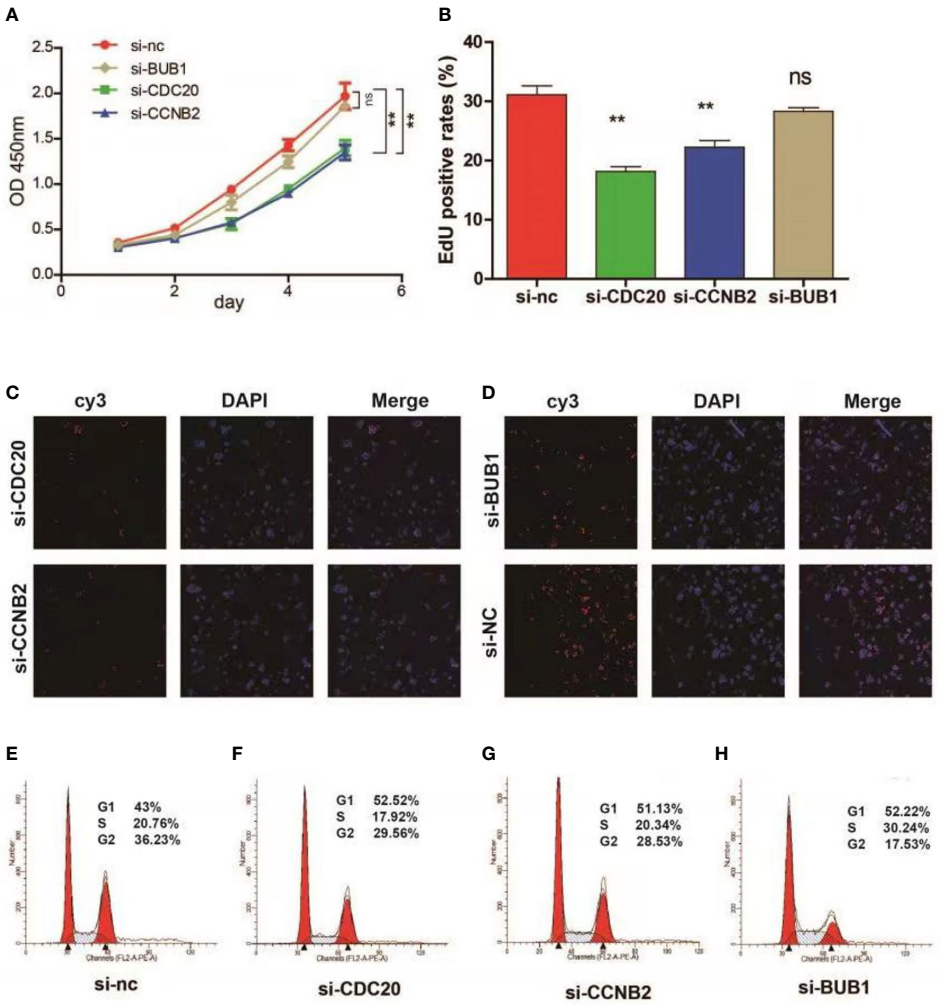

FIGURE 5
CDC20, CNNB2, and BUB1 affects the proliferation of SK-NEP-1 cells. **(A-D)** The cell proliferative ability was assessed by the Cell Counting Kit-8 (CCK-8), and EdU assay after CDC20, CNNB2, and BUB1 was knocked down in SK-NEP-1 cells. **(E-H)** Down-regulation of CDC20, CNNB2, and BUB1 arrests the cell cycle of SK-NEP-1 cells(**P < 0.01, ns, none sense).

Establishment of the 7-gene-based prognostic model in Wilms tumor patients

Univariate *Cox* regression was performed and a total of 43 *OS*-related genes were identified. To construct a more practical model, we employed LASSO regression to screen the most reliable predictive prognostic genes from the above 43 genes, including FCN3, CDC20, and E2F1. The optimal gene model consisting of 18 prognosis DEGs, at the same time, the corresponding coefficients were identified (Figure 6A). Multivariate *Cox* regression analysis further screened 7 prognosis-related genes out of the 18 prognosis-related genes identified from *Lasso* regression analysis. Of the 7 genes identified, HPD and PRC1 were found to be associated with a high risk, showing a HR of >1, whereas CYP27C1,HEY1,EGF,ASPA and TMEM61 were identified as a low-risk gene, with HR <1.

In order to validate the predictive capability of prognostic model, survival curves were generated by K-M survival analysis. Result showed that the predicted survival time of the high-risk group was significantly shorter (P<0.05) (Figure 6B). The 1-,3- and 5-year survival rates of the high-risk group were 83.8%(95% CI=75%-93.5%),42.7%(95%CI=31.7%-57.5%)and 38.8%(95% CI=28%-58.8%), respectively, whereas those of the low-risk group were 93.6%(95%CI=87.8%-99.9%),86.9%(95%CI=78.9%-95.8%)and 81.3%(95%CI=71.9%-92%), respectively. The

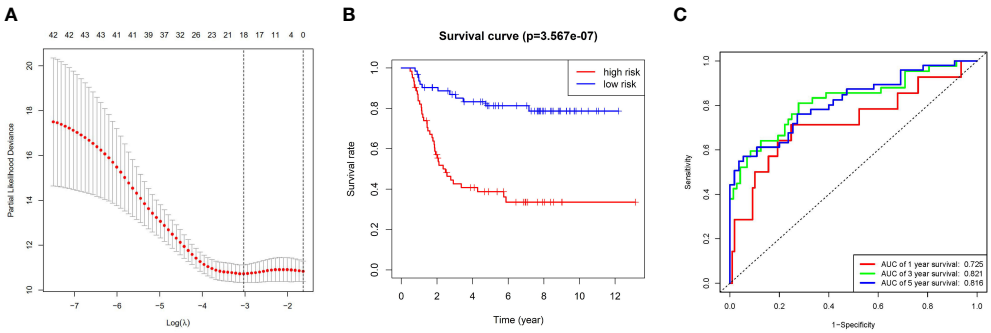

FIGURE 6
Construction of gene prognostic models and their prognostic characteristics. **(A)** Eighteen genes were used to construct the risk models. Coefficients were calculated from univariate Cox regression analysis using LASSO. **(B)** Survival analysis was performed using the Kaplan-Meier method for the overall survival of patients in TARGET. **(C)** The predictive efficiency of the gene prognostic model showed by ROC curves based on 1-, 3- and 5-year survival rates.

prediction capability of the risk score model was evaluated by calculating the area under the *ROC* curve. The 1-, 3-, and 5-years AUC values were 0.727,0.820 and 0.817, respectively (Figure 6C).

Establishment of nomogram

As shown in Table 3, univariate and multivariate Cox regression analyses showed that the ender, clinical stage and gene risk score were independent prognostic factors for OS. A nomogram was plotted based on those factors. The C-index was 0.766. The AUC values of the nomogram for predicting 1-,3- and 5-year OS were 0.828,0.837 and 0.827, respectively (Figure 7A, B). To further test the predictive power of the nomogram, ROC curves were drawn to compare the predictive performance between the nomogram and the other 3 independent predictors. The nomogram obtained the highest 1-,3- and 5-year area under the curve (Figures 8A–C).

Discussion

Wilms tumor (*nephroblastoma*) *is* one of *the most common* pediatric solid tumors, and the survival rate of high-risk patients and relapsed patients is low. It is an urgent need to refine risk stratification and find a reliable biomarker for Wilms tumor patients.

In this study, the results of bioinformatics analysis predicted that BUB1, CCNB2 and CDC20 are hub genes of Wilms tumor,

TABLE 3 Univariate and multivariate Cox regression analysis of different prognostic variables for overall survival (OS) prediction in Wilms tumor patients.

Characteristic	Univariate Cox regression		Multivariate Cox regression	
	HR (95%CI)	P value	HR (95%CI)	P value
Age (years)		0		0
0-3	1		——	——
3-6	1.6326 (0.8349, 3.193)	0.152	——	——
>=6	0.7386 (0.3018, 1.808)	0.507	——	——
Gender		0		0
Female	1		1	
Male	1.785 (1.023, 3.114)	0.0414	2.138 (1.205, 3.793)	0.00936
Clinic stage		0		0
I-II	1		1	
III-IV	3.059 (1.686, 5.551)	0.000235	3.513 (1.898, 6.499)	6.28E-05
Gene risk score		0		0
Low risk	1		1	
High risk	4.616 (2.44, 8.731)	2.56E-06	4.438 (2.330, 8.454)	5.81E-06

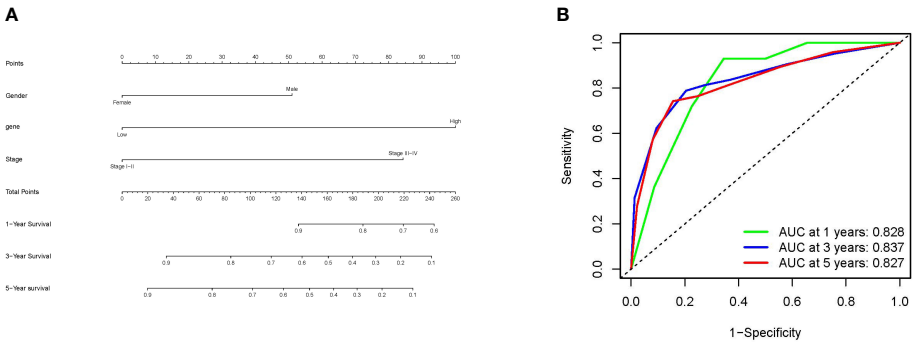

FIGURE 7
Nomogram and its AUC. **(A)** A prognostic nomogram predicting 1-, 2-, and 3-year overall survival of Wilms tumor patient. **(B)** ROC curve for 1-, 3-, and 5-year overall survival predictions for the nomogram.

and this prediction was further proved by the experimental validation. The results proved that CDC20, CNNB2, and BUB1 was upregulated in Wilms tumor and significantly associated with the survival of Wilms tumor patients. Downregulation of CDC20 and CNNB2 markedly suppresses cell growth in cultured SK-NEP-1 cells which was consistent with the results of the functional prediction above. We determined with the cell cycle distribution analysis that the lack of CDC20, CNNB2, and BUB1 led to a G0/G1 phase arrest. We also tried to investigate how CDC20, CNNB2, and BUB1 participates in the development of Wilms tumor.

CDC20 is a cell cycle regulating protein (34). Various malignancies were reported overexpression of CDC20 and high expression of CDC20 was associated with higher tumor grade in bladder, cervical, colon, endometrial, gastric, liver, ovarian, prostatic, and renal carcinomas (35). In cells, BUB1 plays a crucial role in chromosome assembly and kinetochore localization (36).Bub1 also is a specific kinase optimized for CDC20 phosphoryation. CDC20 is the only known physiologically relevant substrate of BuB1 (37). CCNB2 is key protein of the cyclin family and regulates the progression of the G2/M transition during the cell cycle (38). According to Shubbar et al, breast cancer patients' prognoses are affected by CNNB2 overexpression (39). Follow-up analysis has shown that CNNB2 overexpression is associated with poor prognosis. Li et al. found that CCNB2 overexpression in HCC is associated with poor prognosis, and knock-down of CCNB2 inhibits cell proliferation and migration, promotes apoptosis, and causes S phase arrest in HCC cell lines (40). Therefore, more comprehensive studies about molecular mechanism will remain to be lucubrated in the future.

Furthermore, to identify potential gene biomarkers we investigated the differences in gene expression between tumor and normal tissues from Wilms tumor. The differentially expressed genes were screened, and univariate, Lasso and multivariate Cox analyses were used to build a risk model to predict Wilms tumor prognosis. We identified seven genes: PRC1, CYP27C1, HEY1, EGF, HPD, ASPA and TMEM61. Poor prognosis was associated with high levels of PRC1, CYP27C1 and HEY1 expression in Wilms tumor patients. In S. pombe, PCR participates in meiotic recombination, maintenance of heterochromatin structure, and regulation of

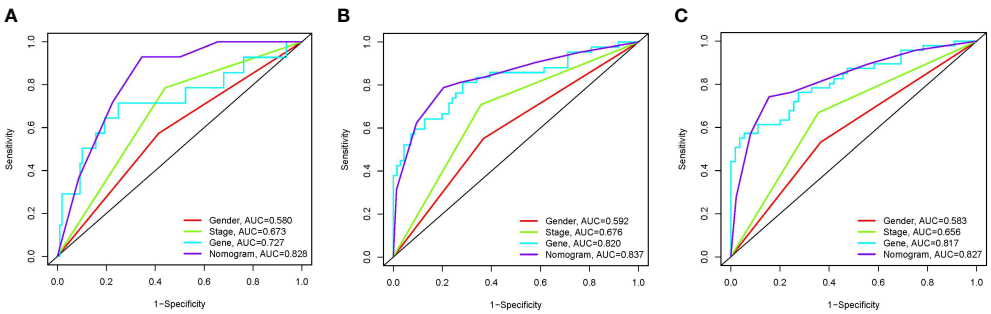

FIGURE 8
The AUC was used to estimate the predictive power of different models. (**A** ROC for 1 year; **B** ROC for 3 years; **C** ROC for 5 years). The predictive ability of the nomogram (purple,1-,3-,5- year AUC= 0.828,0.837,0.827) was better than those of the other risk models.

certain genes related to sexual differentiation, besides induction of stress-responsive genes under oxidative, heat, reductive, osmotic, and starvation stresses with ATF1 by forming a heterodimer (41, 42) There are few studies on the CYP27C1 gene. CYP27C1 was a member of the cytochrome P450 superfamily of enzymes, which can catalyze a number of reactions associated with drug metabolism (43, 44). Some studies have indicated that CYP27C1 can convert vitamin A1 into A2, The CYP27C1-mediated vitamin A1-to-A2 switch could be a switch for visual sensitivity (44). In Bhavsar-Jog, CHRM1, CYP27C1 and FOXH1 are all linked to pathways related to Alzheimer's disease (45). A deeper understanding of CYP27C1 is needed. HEY1 It was found that the AUCs of the ROC curves for the prognostic model to predict the 1,3 and 5-year survival were0.727,0.820 and 0.817, respectively, indicating that the 7-gene signature performed well for survival prediction.

The nomogram is a widely used prognostic evaluation tool in clinical oncology that can incorporate a variety of prognostic determinants, including molecular and clinicopathological findings (46). In our study, by combining gene signatures and clinic variables, a nomogram is developed to predict OS and to assist clinicians in making personalized treatment decisions. The predictive performance of the nomogram built with the combined model is the best by C-index, AUC and compared with other single risk factor. The best performance of nomogram was confirmed.

The current study has several limitations that should be considered. First, the main information on clinical characteristics and gene expression data for the most part came from white, black or Hispanic populations; so the extrapolation of the findings to other ethnic groups needs to be established. Second, the nomogram was only performed using the TARGET database due to the lack of complete clinical information in other datasets; hence, the nomogram should be further validated through multicenter clinical trials and prospective studies. Third, the molecular mechanisms of the genes identified here and their roles in the pathogenesis and progression of Wilms tumor require further experimental studies. Certainly, future research will have to be done on refining these limitations.

Conclusion

Our study discovers three critical gene signatures of Wilms tumor, they are CDC20, BUB1 and NNB2. They could serve as the basis for subsequent experimental analysis. The prognostic models constructed by nomogram produce more effective predictions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The studies involving human participants were reviewed and approved by Sun Yat-Sen Memorial Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

XD and LL conceived and designed the study. LL performed bioinformatic analysis; LC and LL performed the experiments. DG, LL and LC wrote and critically reviewed the manuscript. All authors discussed the results and commented on the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Guangzhou Production, Education and Research collaborative innovation major project (Projects of people's livelihood) (No.201604020151),the Natural Science Foundation of Guangdong Province (No.2015A030313037, No.2022A1515012222) the Municipal School Joint Project of Guangzhou Science and Technology Bureau (104287750039) Funding by Science and Technology Projects in Guangzhou (202201020395) and the Guangdong Provincial Science and Technology Plan Project (2014A020213021).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001-10: a population-based registry study. *Lancet Oncol* (2017) 18(6):719–31. doi: 10.1016/S1470-2045(17)30186-9

2. Nakata K, Colombet M, Stiller CA, Pritchard-Jones K, Steliarova-Foucher E. Incidence of childhood renal tumours: An international population-based study. *Int J Cancer* (2020) 147(12):3313–27. doi: 10.1002/ijc.33147

3. Groenendijk A, Spreafico F, de Krijger RR, Drost J, Brok J, Perotti D, et al. Prognostic factors for wilms tumor recurrence: A review of the literature. . *Cancers* (2021) 13(13):1–25. doi: 10.3390/cancers13133142

4. Suh E, Stratton KL, Leisenring WM, Nathan PC, Ford JS, Freyer DR, et al. Late mortality and chronic health conditions in long-term survivors of early-adolescent and young adult cancers: a retrospective cohort analysis from the childhood cancer survivor study. *Lancet Oncol* (2020) 21(3):421–35. doi: 10.1016/S1470-2045(19)30800-9

5. Termuhlen AM, Tersak JM, Liu Q, Yasui Y, Stovall M, Weathers R, et al. Twenty-five year follow-up of childhood wilms tumor: a report from the childhood cancer survivor study. *Pediatr Blood Cancer* (2011) 57(7):1210–6. doi: 10.1002/pbc.23090

6. Treger TD, Chowdhury T, Pritchard-Jones K, Behjati S. The genetic changes of wilms tumour. *Nat Rev Nephrol* (2019) 15(4):240–51. doi: 10.1038/s41581-019-0112-0

7. Spreafico F, Fernandez CV, Brok J, Nakata K, Vujanic G, Geller JI, et al. Wilms tumour. *Nat Rev Dis Primers* (2021) 7(1):75. doi: 10.1038/s41572-021-00308-8

8. Zhou N, Yan B, Ma J, Jiang H, Li L, Tang H, et al. Expression of TCF3 in wilms' tumor and its regulatory role in kidney tumor cell viability, migration and apoptosis in vitro. *Mol Med Rep* (2021) 24(3):1–8. doi: 10.3892/mmr.2021.12281

9. Call KM, Glaser T, Ito CY, Buckler AJ, Pelletier J, Haber DA, et al. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 wilms' tumor locus. *Cell* (1990) 60(3):509–20. doi: 10.1016/0092-8674(90)90601-A

10. Karnik P, Chen P, Paris M, Yeger H, Williams BR. Loss of heterozygosity at chromosome 11p15 in wilms tumors: identification of two independent regions. *Oncogene* (1998) 17(2):237–40. doi: 10.1038/sj.onc.1201959

11. Rahman N, Arbour L, Tonin P, Renshaw J, Pelletier J, Baruchel S, et al. Evidence for a familial wilms' tumour gene (FWT1) on chromosome 17q12-q21. *Nat Genet* (1996) 13(4):461–3. doi: 10.1038/ng0896-461

12. Fukuzawa R, Breslow NE, Morison IM, Dwyer P, Kusafuka T, Kobayashi Y, et al. Epigenetic differences between wilms' tumours in white and east-Asian children. *Lancet (London England)* (2004) 363(9407):446–51. doi: 10.1016/S0140-6736(04)15491-3

13. Gadd S, Huff V, Walz AL, Ooms A, Armstrong AE, Gerhard DS, et al. A children's oncology group and TARGET initiative exploring the genetic landscape of wilms tumor. *Nat Genet* (2017) 49(10):1487–94. doi: 10.1038/ng.3940

14. Mahamdallie SS, Hanks S, Karlin KL, Zachariou A, Perdeaux ER, Ruark E, et al. Mutations in the transcriptional repressor REST predispose to wilms tumor. *Nat Genet* (2015) 47(12):1471–4. doi: 10.1038/ng.3440

15. Ferrucci F, Ciaccio R, Monticelli S, Pigini P, di Giacomo S, Purgato S, et al. MAX to MYCN intracellular ratio drives the aggressive phenotype and clinical outcome of high risk neuroblastoma. *Biochim Biophys Acta Gene Regul Mech* (2018) 1861(3):235–45. doi: 10.1016/j.bbagrm.2018.01.007

16. Torrezan GT, Ferreira EN, Nakahata AM, Barros BD, Castro MT, Correa BR, et al. Recurrent somatic mutation in DROSHA induces microRNA profile changes in wilms tumour. *Nat Commun* (2014) 5:4039. doi: 10.1038/ncomms5039

17. Grundy PE, Breslow NE, Li S, Perlman E, Beckwith JB, Ritchey ML, et al. Loss of heterozygosity for chromosomes 1p and 16q is an adverse prognostic factor in favorable-histology wilms tumor: a report from the national wilms tumor study group. *J Clin Oncol* (2005) 23(29):7312–21. doi: 10.1200/JCO.2005.01.2799

18. Williams RD, Chagtai T, Alcaide-German M, Apps J, Wegert J, Popov S, et al. Multiple mechanisms of MYCN dysregulation in wilms tumour. *Oncotarget* (2015) 6(9):7232–43. doi: 10.18632/oncotarget.3377

19. Maschietto M, Williams RD, Chagtai T, Popov SD, Sebire NJ, Vujanic G, et al. TP53 mutational status is a potential marker for risk stratification in wilms tumour with diffuse anaplasia. *PloS One* (2014) 9(10):e109924. doi: 10.1371/journal.pone.0109924

20. Wegert J, Vokuhl C, Ziegler B, Ernestus K, Leuschner I, Furtwängler R, et al. TP53 alterations in wilms tumour represent progression events with strong intratumour heterogeneity that are closely linked but not limited to anaplasia. *J Pathol Clin Res* (2017) 3(4):234–48. doi: 10.1002/cjp2.77

21. Walz AL, Ooms A, Gadd S, Gerhard DS, Smith MA, Guidry Auvil JM, et al. Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology wilms tumors. *Cancer Cell* (2015) 27(2):286–97. doi: 10.1016/j.ccell.2015.01.003

22. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinf (Oxford England)* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034

23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

24. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* (2015) 15(15):2597–601. doi: 10.1002/pmic.201400515

25. Ginestet C. ggplot2: Elegant graphics for data analysis. *Journal of the Royal Statistical Society* (2011) 174(1):245–6. doi: 10.1111/j.1467-985X.2010.00676_9.x

26. Chen L, Zhang YH, Wang S, Zhang Y, Huang T, Cai YD. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PloS One* (2017) 12(9):e0184129. doi: 10.1371/journal.pone.0184129

27. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27

28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *Omics J Integr Biol* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

29. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* (2017) 45(D1):D362–d8. doi: 10.1093/nar/gkw937

30. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with cytoscape 3. *Curr Protoc Bioinf* (2014) 47:1–24. doi: 10.1002/0471250953.bi0813s47

31. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* (2014) 8 Suppl 4(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11

32. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol* (2014) 21(4):686–94. doi: 10.1007/s12350-014-9908-2

33. Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

34. Liu L, Lin J, He H. Identification of potential crucial genes associated with the pathogenesis and prognosis of endometrial cancer. *Front Genet* (2019) 10:373. doi: 10.3389/fgene.2019.00373

35. Dong S, Huang F, Zhang H, Chen Q. Overexpression of BUB1B, CCNA2, CDC20, and CDK1 in tumor tissues predicts poor survival in pancreatic ductal adenocarcinoma. *Biosci Rep* (2019) 39(2):1–10. doi: 10.1042/BSR20182306

36. Raaijmakers JA, van Heesbeen R, Blomen VA, Janssen LME, van Diemen F, Brummelkamp TR, et al. BUB1 is essential for the viability of human cells in which the spindle assembly checkpoint is compromised. *Cell Rep* (2018) 22(6):1424–38. doi: 10.1016/j.celrep.2018.01.034

37. Kang J, Yang M, Li B, Qi W, Zhang C, Shokat KM, et al. Structure and substrate recruitment of the human spindle checkpoint kinase Bub1. *Mol Cell* (2008) 32(3):394–405. doi: 10.1016/j.molcel.2008.09.017

38. Shi Q, Wang W, Jia Z, Chen P, Ma K, Zhou C. ISL1, a novel regulator of CCNB1, CCNB2 and c-MYC genes, promotes gastric cancer cell proliferation and tumor growth. *Oncotarget* (2016) 7(24):36489–500. doi: 10.18632/oncotarget.9269

39. Shubbar E, Kovács A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdóttir K, et al. Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. *BMC Cancer* (2013) 13:1. doi: 10.1186/1471-2407-13-1

40. Li R, Jiang X, Zhang Y, Wang S, Chen X, Yu X, et al. Cyclin B2 overexpression in human hepatocellular carcinoma is associated with poor prognosis. *Arch Med Res* (2019) 50(1):10–7. doi: 10.1016/j.arcmed.2019.03.003

41. Sansó M, Vargas-Pérez I, García P, Ayté J, Hidalgo E. Nuclear roles and regulation of chromatin structure by the stress-dependent MAP kinase Sty1 of schizosaccharomyces pombe. *Mol Microbiol* (2011) 82(3):542–54. doi: 10.1111/j.1365-2958.2011.07851.x

42. Orosz E, Antal K, Gazdag Z, Szabó Z, Han KH, Yu JH, et al. Transcriptome-based modeling reveals that oxidative stress induces modulation of the AtfA-dependent signaling networks in aspergillus nidulans. *Int J Genomics* (2017) 2017:6923849. doi: 10.1155/2017/6923849

43. Johnson KM, Phan TTN, Albertolle ME, Guengerich FP. Human mitochondrial cytochrome P450 27C1 is localized in skin and preferentially desaturates trans-retinol to 3,4-dehydroretinol. *J Biol Chem* (2017) 292 (33):13672–87. doi: 10.1074/jbc.M116.773937

44. Morshedian A, Toomey MB, Pollock GE, Frederiksen R, Enright JM, McCormick SD, et al. Cambrian Origin of the CYP27C1-mediated vitamin A(1)-to-A(2) switch, a key mechanism of vertebrate sensory plasticity. *R Soc Open Sci* (2017) 4(7):170362. doi: 10.1098/rsos.170362

45. Bhavsar-Jog YP, Van Dornshuld E, Brooks TA, Tschumper GS, Wadkins RM. Co-Localization of DNA i-Motif-Forming sequences and 5-hydroxymethyl-cytosines in human embryonic stem cells. *Mol (Basel Switzerland)* (2019) 24(19):1–12. doi: 10.3390/molecules24193619

46. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol* (2015) 16(4):e173–80. doi: 10.1016/S1470-2045(14)71116-7