



# A long non-coding RNAs expression signature to improve prognostic prediction of Wilms tumor in children

Hongyan Zhao<sup>1</sup>, Peng Wang<sup>2</sup>, Gang Wang<sup>3^</sup>, Shuo Zhang<sup>4</sup>, Feng Guo<sup>3^</sup>

<sup>1</sup>Department of Critical Care Medicine, The Second Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China; <sup>2</sup>Department of Critical Care Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China; <sup>3</sup>Department of Pediatric Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China; <sup>4</sup>Department of Hand and Foot Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

*Contributions:* (I) Conception and design: H Zhao, F Guo; (II) Administrative support: P Wang; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: H Zhao, G Wang; (V) Data analysis and interpretation: H Zhao, S Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Feng Guo. Department of Pediatric Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, 324 Jingwu Road, Jinan 250021, China. Email: guofeng85186021@163.com.

**Background:** Wilms tumor (WT) is the most frequent malignancy of the kidney in children, and a subset of patients remains with a poor prognosis. This study aimed to identify key long non-coding RNAs (lncRNAs) related to prognosis and establish a genomic-clinicopathologic nomogram to predict survival in children with WT.

**Methods:** Clinical data of 124 WT patients and the relevant RNA sequencing data including lncRNAs expression signature of primary WT samples were obtained from the Therapeutically Applicable Research to Generate Effective Treatment (TARGET) Data Matrix. Then, lncRNAs associated with overall survival (OS) were identified through univariate Cox, least absolute shrinkage and selection operator (LASSO), and multivariate Cox regression analyses. The risk scores of 124 participants were calculated, and survival analyses were performed between low- and high-risk groups. A genomic-clinicopathologic nomogram was then developed and evaluated by time-dependent receiver operating characteristic (ROC) curves, including the area under the curve (AUC), calibration curve, and decision curve analysis. Subsequently, bioinformatics analyses were performed to explore the potential molecular mechanisms that affect the prognosis of WT. The package “DESeq2” was used to identify differentially expressed protein-coding genes (DEPCGs) between groups. Gene Set Enrichment Analysis (GSEA) was applied to explore the differences in pathways enrichment. The analytical tools CIBERSORTx and ESTIMATE were used to investigate the discrepancies of the immune microenvironment.

**Results:** A total of 10 lncRNAs were selected as independent predictors associated with OS ( $P < 0.05$ ). Participants in the high-risk group had a significantly worse OS and event-free survival (EFS) than those in the low-risk group ( $P < 2 \times 10^{-16}$  and  $P = 2.03 \times 10^{-4}$ , respectively). The risk score and 3 clinicopathological features (gender, cooperative group protocol, and stage) were identified to construct the nomogram (combined model) ( $P = 5.11 \times 10^{-17}$ ). The combined model (1-year AUC: 0.9272, 3-year AUC: 0.9428, 5-year AUC: 0.9259) and risk score model (1-year AUC: 0.9285, 3-year AUC: 0.9399, 5-year AUC: 0.9266) displayed higher predictive accuracy than that of the other models. Subsequently, 105 DEPCGs were identified. The GSEA revealed 4 significant pathways. Analysis with CIBERSORTx demonstrated that monocytes, macrophages M1, activated dendritic cells, and resting mast cells had significant infiltration differences between groups.

**Conclusions:** This study constructed a genomic-clinicopathologic nomogram, which might present a novel and efficient method for treating patients with WT.

<sup>^</sup> ORCID: Gang Wang, 0000-0003-0391-2171; Feng Guo, 0000-0003-0391-2171.

**Keywords:** Wilms tumor (WT); long non-coding RNA (lncRNA); nomograms; prognosis; bioinformatics analysis

Submitted Oct 04, 2020. Accepted for publication Jan 29, 2021.

doi: 10.21037/tp-20-318

View this article at: <http://dx.doi.org/10.21037/tp-20-318>

## Introduction

Wilms tumor (WT), also known as nephroblastoma, is the most frequent malignancy of the kidney in children, constituting nearly 90% of childhood renal tumors (1). Despite the development and advancement of various treatment approaches, 15% of WT patients undergo metastasis or relapse, suggesting poor prognosis after initial treatment (2). Therefore, it is of great clinical benefit to explore prognostic factors in children with different WT stages, which could provide valuable guidance on formulating an accurate therapeutic regime for each patient. Previous studies that were committed to identifying potential prognostic markers for WT patients relied on very few factors, mainly histology analysis and characterization of tumor stage, which has been deemed insufficient as WT in children was shown to be affected by a variety of factors, including genetic and epigenetic changes that underlie WT pathogenesis (1). The prognosis for WT children in clinical practice incorporates a diverse set of measurements, including age, tumor size, loss of heterozygosity (LOH) for chromosomes 16q and 1p, and sensitivity to medications (3-6). Moreover, radiation (7), surgery, microscopic residual disease, diffuse anaplasia (8), lymph node involvement (9), and a combination of lymph node and LOH status (10) have been associated with the prognosis of WT in children. Therefore, it is important to build a prognostic model using multiple variables that underlie a broad spectrum of genetic and clinicopathological factors.

Long non-coding RNAs (lncRNAs) are characterized as non-coding transcripts longer than 200 nucleotides (11). Next-generation and high-throughput sequencing techniques have enabled a significant breakthrough in lncRNA identification and characterization in recent years. Gene expression of various lncRNAs is dysregulated in numerous cancers, and it has been reported to correlate with cancer recurrence, metastasis, and poor prognosis (12-24). Previous studies have suggested that several lncRNAs, such as LINC00473 (25), MIAT (26), SOX21-AS1 (27), and LINP1 (28), play important roles in the pathogenesis of WT. However, it remains largely unknown whether lncRNAs could be used as prognostic markers of WT in children. The present study sought to characterize lncRNA

signature as a robust prediction for WT children's prognosis and construct a prognostic risk index based on comprehensive RNA-sequencing analysis. We established a prognosis model using RNA-sequencing data and subsequently developed and validated a genomic- clinicopathological nomogram that integrated risk scores and traditional clinicopathological factors. Moreover, Gene Set Enrichment Analysis (GSEA) revealed molecular pathways significantly enriched in different risk groups (29,30). Additionally, CIBERSORTx and ESTIMATE algorithms were applied to explore the immune microenvironment discrepancies between the high-risk and low-risk groups (31-33). Altogether, the current study presents a novel and efficient method for the prognosis of patients with WT and provides insight into the molecular mechanisms that affect WT's prognosis. We present the following article in accordance with the TRIPOD reporting checklist (available at <http://dx.doi.org/10.21037/tp-20-318>).

## Methods

### *Study population and RNA-sequencing data processing*

The level 3 RNAseq data of primary WT samples, including raw read counts and reads per kilobase million (RPKM), were obtained from the Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA), and relevant clinical information of the WT patients was downloaded from the Therapeutically Applicable Research to Generate Effective Treatment (TARGET) Data Matrix in April 2020. This study was conducted following the TARGET publication guidelines. The inclusion criteria of this study were as follows: (I) pathological diagnosis of primary WT; (II) clinical information including age at diagnosis, gender, cooperative group protocol, stage, histologic classification, the first event, event-free survival (EFS), vital status, and overall survival (OS) was documented; and (III) complete RNAseq data. The exclusion criteria were as follows: (I) patients with a history of other malignancies; (II) tumor tissues were not used for RNA sequencing analysis; (III) clinical information not available. Thus, 124 WT patients were selected and enrolled in this study (Tables S1,S2).

**Table 1** Clinicopathological characteristics of 124 WT patients

Characteristics	Median (range) or n (%)
Age at diagnosis in days	1,539 (156–5,698)
Gender	
Female	70 (56.45)
Male	54 (43.55)
First event	
None	27 (21.77)
Relapse or progression	97 (78.23)
Vital status	
Alive	74 (59.68)
Dead	50 (40.32)
Protocol	
NWTS-4	2 (1.61)
NWTS-5	122 (98.39)
Stage	
I–II	65 (52.42)
III–IV	59 (47.58)
Histologic classification	
FHWT	82 (66.13)
DAWT	42 (33.87)

WT, Wilms tumor; NWTS-4, the fourth National Wilms' Tumor Study; NWTS-5, the fifth National Wilms' Tumor Study; FHWT, favorable histology Wilms tumor; DAWT, diffuse anaplastic Wilms tumor.

Participant clinicopathological characteristics are listed in *Table 1*. The geneset annotation GENCODE Human (GRCh38.p13) was used to annotate protein-coding genes and lncRNAs were selected for further study (34). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

#### **Identification and validation of the prognostic lncRNA signature**

The “survival” package in the R software (version 3.6.2; <https://www.R-project.org/>) was used to analyze RPKM data [transformed to  $\log_2(\text{RPKM}+1)$ ] through univariate Cox regression analysis, and  $P < 0.05$  was used as the cut-off criterion. Next, through the “glmnet” package (version 3.0-2) in R, the least absolute shrinkage and selection operator

(LASSO) Cox regression was performed to select key lncRNAs among the significant lncRNAs from univariate Cox analysis (35,36). We used 10 times cross-validations to determine the best penalty parameter lambda. The key lncRNAs selected by the LASSO method were considered as candidate variables for inclusion in the multivariate Cox analysis. Then, the independently prognostic lncRNAs were identified in the multivariate Cox analysis to obtain their coefficients ( $\beta$  values), with which the prognostic formula was constructed as follows:  $(\beta_1 \times \log_2(\text{RPKM}+1)$  value of gene1) +  $(\beta_2 \times \log_2(\text{RPKM}+1)$  value of gene2) + ... +  $(\beta_n \times \log_2(\text{RPKM}+1)$  value of genen). Subsequently, the risk score of each WT patient was calculated based on the prognostic formula. Then, the “surv\_cutpoint” function in the “survminer” R package (version 0.4.6) was utilized to achieve the optimal cutoff values for risk scores. According to the cutoff values, we divided the 124 WT participants in the cohort into a low-risk group and a high-risk group. Between the low- and high-risk groups, differences in OS and EFS were examined by log-rank test, and Kaplan-Meier survival curves were constructed by the “survival” package in R software. A  $P$  value  $< 0.05$  was considered statistically significant.

#### **Development and assessment of a predictive nomogram**

Univariate and multivariate Cox regression analyses were conducted to ascertain whether the risk score of lncRNAs and the clinicopathological variables could be prognostic markers related to OS for WT patients, and a  $P < 0.05$  was deemed significant. Then, according to the results of Cox regression analyses, we constructed a genomic-clinicopathologic nomogram to predict 1-, 3-, and 5-year OS through the “rms” package (version 5.1-4) in R software. Using the same R package, we generated a calibration curve to verify the consistency between the nomogram-predicted probability of OS and the actual OS. Furthermore, to assess the accuracy of the predictive nomogram, we performed time-dependent receiver operating characteristic (ROC) curve analyses and calculated the area under the curve (AUC) through the “timeROC” package (version 0.4) in R (37). Additionally, we conducted a decision curve analysis (DCA) to examine the clinical value with the “stdca.R” statistical code (38) in R.

#### **Identification of differentially expressed genes and construction of protein-protein interaction network**

Using the “DESeq2” package (version 1.26.0) in R, the

differentially expressed genes, including protein-coding genes (DEPCGs) and lncRNAs (DELs) were identified by comparing the low-risk group and high-risk group of primary WT, with the screening conditions of  $|\log_2$  fold change  $>1$ , and a false discovery rate (FDR) or adjusted P-value  $<0.05$ . These DEPCGs were then used to construct the protein-protein interaction (PPI) network through the Search Tool for the Retrieval of Interacting Genes (STRING) online database (39). Required interaction score  $>0.4$  was used as the cut-off criterion. Then, the data downloaded from STRING were used to model the PPI network through Cytoscape software (version 3.7.1, <https://cytoscape.org>) (40), and the densely connected regions were identified through the Molecular Complex Detection (MCODE) plug-in Cytoscape with the default criteria (degree cutoff =2, node score cutoff =0.2, Haircut = true, Fluff = false, K-score =2, and Max depth =100) (41).

### GSEA

Before running GSEA, the normalization of the raw read counts of protein-coding genes was performed by “DESeq2”. Then, GSEA (version 4.0.3) was conducted to demonstrate significant differences in protein-coding genes between the low-risk and high-risk groups (29,30). Canonical pathway gene sets (c2.cp.v7.1.symbols.gmt) and Gene Ontology (GO) gene sets including biological process (BP) (c5.bp.v7.1.symbols.gmt), molecular function (MF) (c5.mf.v7.1.symbols.gmt), and cellular component (CC) (c5.cc.v7.1.symbols.gmt) were selected from the Molecular Signatures Database (MSigDB) as the reference gene sets (42–44). The number of permutations was set at 1,000, and the nominal P value  $<0.01$  was considered statistically significant.

### Analysis of immune infiltration level

The analytical tool CIBERSORTx provides an estimation of member cell types’ abundances in a mixed cell population, using gene expression data (31,32). We utilized CIBERSORTx to assess the proportions of 22 immune cell subtypes in the primary WT samples, using the RPKM data of protein-coding genes. The permutations for significance analysis were set at 1000. Then, we performed the 2-sided Wilcoxon test through the “ggpubr” package (version 0.2.5) in R, to compare differences in 22 immune cell subtypes between the low- and high-risk groups. Subsequently, through the “estimate” package (version 1.0.13), the immune and stromal scores were calculated based on the

ESTIMATE algorithm (33), and Welch’s *t*-test performed the comparisons of the stromal and immune scores between the 2 groups with “ggstatsplot” package (version 0.4.0) in R. A P value  $<0.05$  denoted statistical significance.

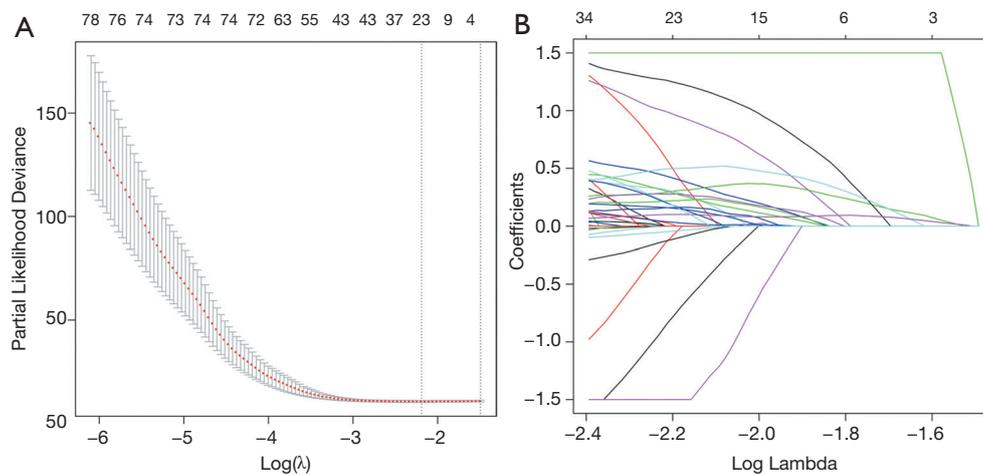
### Statistical analysis

The R software version 3.6.2 and several R packages were used for statistical analyses, and a 2-tailed P value  $<0.05$  denoted statistical significance. The univariate and multivariate Cox regression analyses were performed by the “survival” package. The LASSO Cox regression analyses were calculated by the “glmnet” package (35,36), and 10 times cross-validations were used to determine the best penalty parameter lambda. Kaplan-Meier analyses and survival curves were constructed through the “survival” package. The “rms” package generated the nomogram and the calibration curve. The time-dependent ROC curve analyses were performed by the “timeROC” package (37).

## Results

### Identification and validation of the prognostic lncRNA signature

After filtering of low-expression lncRNAs, 2,927 lncRNAs were generated for further study. Subsequently, univariate Cox analysis was performed to obtain 244 lncRNAs associated with OS of WT patients ( $P<0.05$ ). Among the 244 prognosis-associated lncRNAs, 24 were identified using the LASSO Cox method (Figure 1). Then, through multivariate Cox analysis, ten lncRNAs (TENM3-AS1, AC022098.3, EMX2OS, AC099811.1, AL359710.1, ADAMTSL4-AS1, AC005944.1, MBNL1-AS1, AC002451.1, and AC120498.3) were selected from 24 candidates of lncRNAs as independent predictors associated with OS ( $P<0.05$ ), and the coefficients of the 10 lncRNAs were obtained (Table 2). The formula of the lncRNA-based prognostic index model was imputed as follows:  $[1.3 \times \log_2(\text{RPKM}+1)$  value of TENM3-AS1] +  $[2.627 \times \log_2(\text{RPKM}+1)$  value of AC022098.3] +  $[-0.5553 \times \log_2(\text{RPKM}+1)$  value of EMX2OS] +  $[-7.114 \times \log_2(\text{RPKM}+1)$  value of AC099811.1] +  $[19.99 \times \log_2(\text{RPKM}+1)$  value of AL359710.1] +  $[3.866 \times \log_2(\text{RPKM}+1)$  value of ADAMTSL4-AS1] +  $[-10.4 \times \log_2(\text{RPKM}+1)$  value of AC005944.1] +  $[1.517 \times \log_2(\text{RPKM}+1)$  value of MBNL1-AS1] +  $[-1.431 \times \log_2(\text{RPKM}+1)$  value of AC002451.1] +  $[2.215 \times \log_2(\text{RPKM}+1)$  value of AC120498.3]. Then,



**Figure 1** Candidate lncRNAs selection through LASSO-penalized Cox regression. (A) 10-fold cross-validation with lambda.min. (B) Coefficient profiles of the candidate lncRNAs. lncRNAs, long non-coding RNAs; LASSO, least absolute shrinkage, and selection operator.

the risk score for each participant was calculated, and all participants were divided into high-risk ( $n=54$ ) and low-risk ( $n=70$ ) groups based on the optimal cutoff value (1.5069) for risk scores (Figure 2). Participants in the high-risk group had a significantly worse OS and EFS than those in the low-risk group ( $P<2E-16$  and  $P=2.03E-04$ , respectively) (Figure 3A,B, Table 3). Importantly, subgroup analyses based on the clinicopathological variables (i.e., gender, cooperative group protocol, stage, and histologic classification) indicated that the prognostic value of the risk scores for OS and EFS was independent of all these clinicopathological variables, except for EFS in the subgroup of stage I-II and favorable histology Wilms tumor (FHWT) ( $P=0.348$  and  $P=0.326$ ; Figure 3C,D,E,F,G,H,I,J,K,L,M,N,O,P, Table 3).

#### Development and assessment of a predictive nomogram

Based on the univariate Cox regression analysis results, the risk score of lncRNAs and 3 clinicopathological factors (i.e., gender, cooperative group protocol, and stage) were all significantly related to OS for WT patients (Table 4). The subsequent multivariate Cox analysis (P value for Wald test:  $5.11E-17$ ) further demonstrated that the risk score remained a powerful and independent prognostic factor ( $P=9.22E-16$ ) (Table 4). Through integrating the 4 prognostic factors identified by univariate Cox analysis, a genomic-clinicopathologic nomogram was developed. This combined model is depicted in Figure 4A. The concordance index of the combined OS prediction model was 0.8780 (95% CI: 0.8407–0.9154). The nomograms' predictive

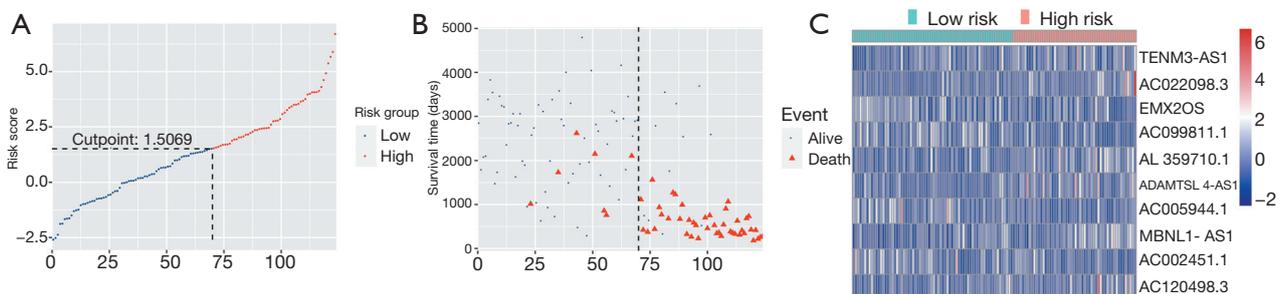
accuracy was further evaluated by time-dependent ROC curves of 1-, 3-, and 5-year OS among the models. The results suggested that the combined model and the risk score model had a significantly greater AUC than that of the gender, cooperative group protocol, and stage models (Figure 4B,C,D; Tables 5,6). Additionally, there was no significant difference in AUC between the combined model and the risk score model (1-year  $P=0.8610$ , 3-year  $P=0.7024$ , 5-year  $P=0.9354$ ) (Table 6). Furthermore, the calibration curve also demonstrated high consistency between the actual proportion of 500-day OS and the nomogram-predicted probability of 500-day OS (Figure 4E). Meanwhile, DCA illustrated that the net benefit of the 500-day decision curve for the combined model was greater than that for the gender, cooperative group protocol, and stage model, further supporting the clinical value of the genomic-clinicopathologic nomogram (Figure 4F).

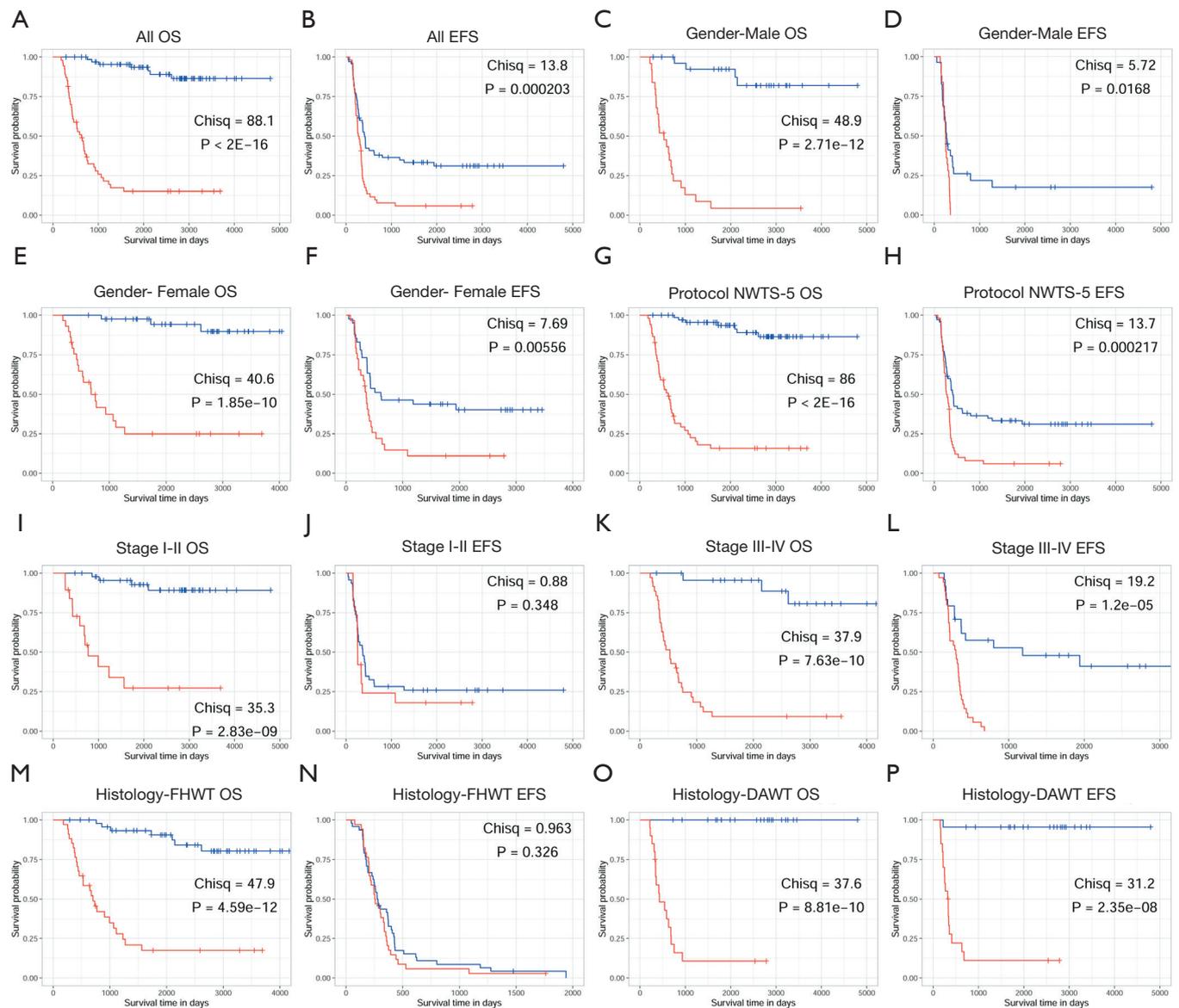
#### Identification of differentially expressed genes and construction of PPI network

A total of 105 DEPCGs (24 upregulated and 81 downregulated) and 14 DELs (7 upregulated and 7 downregulated) were identified in the high-risk group (Figure 5A,B). Furthermore, the PPI network of the 105 DEPCGs was generated (Figure 5C). Following MCODE analysis, 2 densely connected regions in the PPI network were identified. The first region consisted of 15 target protein-coding genes, including MYL4, MYBPC2, ACTC1, MYLPE, DES, TRIM63, NEB, TTN, TNNI2,

**Table 2** The results of multivariate Cox analysis to identify the independently predictive lncRNAs from 24 candidates

Gene symbol	Ensemble gene id	Coefficient ( $\beta$ values)	Hazard ratio	P value
TENM3-AS1	ENSG00000177822	1.300197	3.670021	0.000577
AC022098.3	ENSG00000267783	2.62672	13.82834	0.000852
EMX2OS	ENSG00000229847	-0.55528	0.573914	0.001632
AC099811.1	ENSG00000236194	-7.11422	0.000813	0.001753
AL359710.1	ENSG00000237461	19.98674	4.79E+08	0.002196
ADAMTSL4-AS1	ENSG00000203804	3.866465	47.77323	0.014825
AC005944.1	ENSG00000267469	-10.4027	3.03E-05	0.015802
MBNL1-AS1	ENSG00000229619	1.516977	4.558426	0.017516
AC002451.1	ENSG00000231170	-1.43106	0.239055	0.031301
AC120498.3	ENSG00000260403	2.215427	9.165325	0.034148
AL139260.1	ENSG00000228436	1.229792	3.420518	0.053541
RABGAP1L-DT	ENSG00000227373	3.674629	39.434	0.054951
SH3RF3-AS1	ENSG00000259863	0.621723	1.862133	0.075875
LINC00630	ENSG00000223546	1.359918	3.895875	0.083571
AC245140.1	ENSG00000231830	1.778615	5.921647	0.083843
AC233266.2	ENSG00000261600	-0.40106	0.669608	0.094665
TMEM72-AS1	ENSG00000224812	1.521191	4.577673	0.130971
AC009831.1	ENSG00000263823	1.992977	7.337346	0.203017
AC079922.2	ENSG00000237753	-0.89755	0.407569	0.287507
AC091180.6	ENSG00000262039	-2.26552	0.103776	0.303579
LCMT1-AS1	ENSG00000260448	0.651954	1.919288	0.414851
AC226119.1	ENSG00000253917	-0.39665	0.67257	0.468275
POU6F2-AS1	ENSG00000224122	0.142447	1.153092	0.59815
AC018653.3	ENSG00000256967	0.038999	1.03977	0.957213

**Figure 2** Risk score analyses of 124 WT patients in low- and high-risk groups. (A) Risk score distribution against the rank of the risk score. (B) OS status of patients. (C) Heatmap of the expression profiles of the 10 lncRNAs. WT, Wilms tumor; OS, overall survival



**Figure 3** Kaplan-Meier OS and EFS analyses between the high- and low-risk groups represented respectively by the red and blue curves. (A) All patients' OS. (B) All patients' EFS. (C) Male patients' OS. (D) Male patients' EFS. (E) Female patients' OS. (F) Female patients' EFS. (G) NWTS-5 patients' OS. (H) NWTS-5 patients' EFS. (I) Stage I-II patients' OS. (J) Stage I-II patients' EFS. (K) Stage III-IV patients' OS. (L) Stage III-IV patients' EFS. (M) FHWT patients' OS. (N) FHWT patients' EFS. (O) DAWT patients' OS. (P) DAWT patients' EFS. OS, overall survival; EFS, event-free survival; NWTS-5, the fifth National Wilms' Tumor Study; FHWT, favorable histology Wilms tumor; DAWT, diffuse anaplastic Wilms tumor.

TNNC2, MYH3, UNC45B, ACTN2, MYH8, and TNNT2 (Figure 5D). The second region consisted of 4 target protein-coding genes, including ATP1A2, SLC1A2, SLC17A7, and GRM5 (Figure 5E).

**GSEA**

Through canonical pathway gene sets, GSEA identified energy-dependent regulation of mTOR by LKB1-AMPK

**Table 3** The results of OS and EFS analyses for WT patients through log-rank test

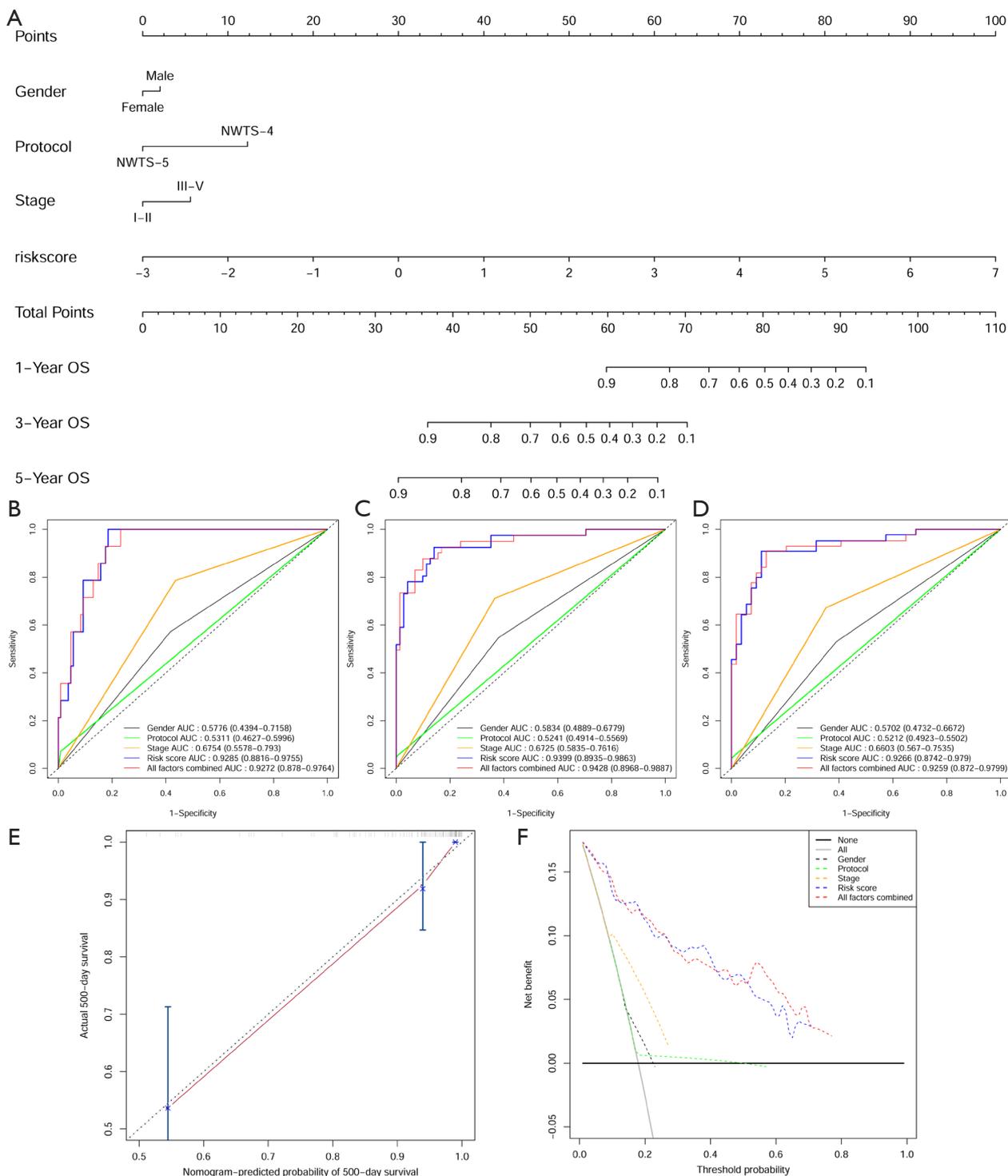
	OS		EFS	
	chi-square value	P value	chi-square value	P value
All	88.15	<2E-16	13.80	2.03E-04
Gender				
Female	40.62	1.85E-10	7.69	5.56E-03
Male	48.89	2.71E-12	5.72	1.68E-02
Protocol				
NWTS-5	85.97	<2E-16	13.68	2.17E-04
Stage				
I-II	35.30	2.83E-09	0.88	0.3482
III-IV	37.85	7.63E-10	19.17	1.20E-05
Histologic classification				
FHWT	47.86	4.59E-12	0.96	0.3264
DAWT	37.57	8.81E-10	31.18	2.35E-08

OS, overall survival; EFS, event-free survival; WT, Wilms tumor; NWTS-5, the fifth National Wilms' Tumor Study; FHWT, favorable histology Wilms tumor; DAWT, diffuse anaplastic Wilms tumor.

**Table 4** The results of univariate and multivariate Cox analyses of OS for WT patients

Risk factors	N	Univariate Cox			Multivariate Cox (P= 5.11 E-17)		
		HR	95% CI	P value	HR	95% CI	P value
Age	124	1	0.9997-1	0.7487			
Gender							
Female	70	1		0.04857	1		0.5577
Male	54	1.751	1.004-3.055		1.2060	0.6447-2.2561	
Protocol							
NWTS-4	2	1		0.0369	1		0.1381
NWTS-5	122	0.2194	0.05277-0.9119		0.3222	0.0721-1.4393	
Stage							
I-II	65	1		1.60E-04	1		0.1072
III-IV	59	3.151	1.736-5.718		1.6741	0.8943-3.1336	
Histologic classification							
FHWT	82	1		0.7081			
DAWT	42	1.118	0.6228-2.008				
Risk score	124	2.62	2.124-3.231	2.17E-19	2.5147	2.0083-3.1489	9.22E-16

OS, overall survival; WT, Wilms tumor; HR, hazard ratio; CI, confidence interval; NWTS-4, the fourth National Wilms' Tumor Study; NWTS-5, the fifth National Wilms' Tumor Study; FHWT, favorable histology Wilms tumor; DAWT, diffuse anaplastic Wilms tumor.



**Figure 4** Construction and validation of the nomogram in WT patients. (A) The nomogram integrating the risk score of lncRNAs and 3 clinicopathological factors (gender, cooperative group protocol, and stage) to predict 1-, 3-, and 5-year OS. The time-dependent ROC curves for predicting probabilities of patients with 1-year (B), 3-year (C), and 5-year (D) OS. The calibration curve (E) and the DCA curve (F) of the nomogram for predicting probabilities of patients with 500-day OS. WT, Wilms tumor; lncRNAs, long non-coding RNAs; OS, overall survival; ROC, receiver operating characteristic; DCA, decision curve analysis.

**Table 5** 1-, 3-, and 5-year AUCs (95% CI) among models

	1-year		3-year		5-year	
	AUC	95% CI	AUC	95% CI	AUC	95% CI
Gender	0.5776	0.4394–0.7158	0.5834	0.4889–0.6779	0.5702	0.4732–0.6672
Protocol	0.5311	0.4627–0.5996	0.5241	0.4914–0.5569	0.5212	0.4923–0.5502
Stage	0.6754	0.5578–0.793	0.6725	0.5835–0.7616	0.6603	0.567–0.7535
Risk score	0.9285	0.8816–0.9755	0.9399	0.8935–0.9863	0.9266	0.8742–0.979
Combined	0.9272	0.878–0.9764	0.9428	0.8968–0.9887	0.9259	0.872–0.9799

AUC, the area under the time-dependent receiver operating characteristic curve; CI, confidence interval.

**Table 6** The results (P values) of comparing AUCs between two models

	1-year	3-year	5-year
Combined vs. Gender	1.19E-06	1.63E-11	9.02E-11
Combined vs. Protocol	1.68E-24	1.59E-51	5.89E-41
Combined vs. Stage	9.74E-06	1.50E-10	1.91E-09
Combined vs. Risk score	0.8610	0.7024	0.9354
Risk score vs. Gender	4.27E-07	3.75E-11	1.97E-10
Risk score vs. Protocol	2.34E-22	1.04E-46	3.09E-41
Risk score vs. Stage	3.24E-05	6.66E-09	2.72E-08

AUC, the area under the time-dependent receiver operating characteristic curve.

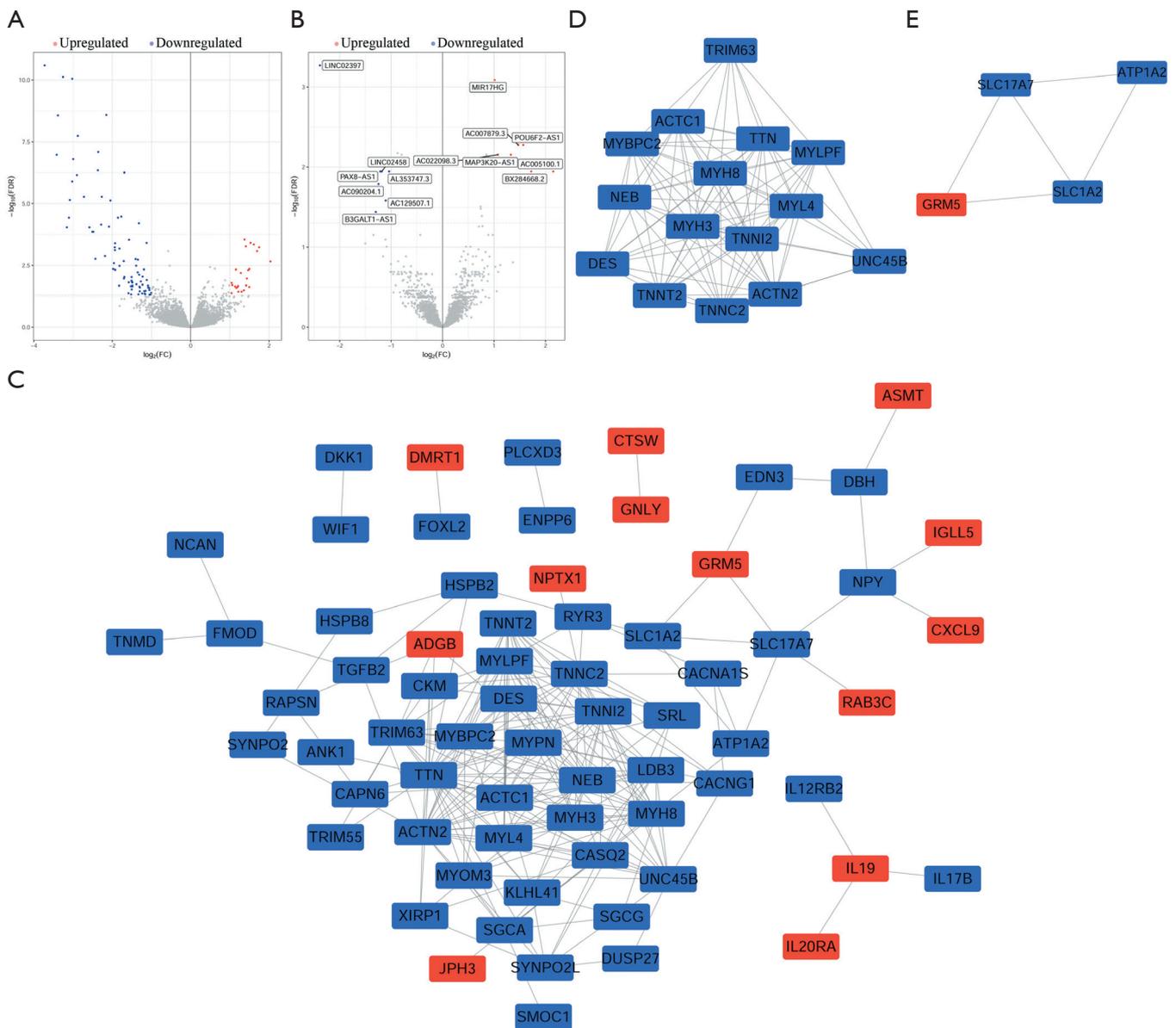
from the Reactome subset, which was significantly enriched in the high-risk group [(nominal  $P=0.007859$ , normalized enrichment score (NES) = 1.728] (Figure 6A). By contrast, 3 other pathways, including the HNF3B pathway (nominal  $P=0.009434$ , NES = -1.631) from the PID subset (Figure 6B), olfactory transduction (nominal  $P=0.009634$ , NES = -1.599) from the Kyoto Encyclopedia of Genes and Genomes (KEGG) subset (Figure 6C), and crosslinking of collagen fibrils (nominal  $P=0.009940$ , NES = -1.730) from the Reactome subset (Figure 6D), were significantly enriched in the low-risk group. Through GO gene sets, GSEA identified 11 GO terms consisted of 5 BP, 1 CC, and 5 MF, which were significantly enriched in the high-risk group (nominal  $P<0.01$ ) (Figure 6E), and 19 GO terms consisted of 16 BP, 2 CC, and 1 MF, which were significantly enriched in the low-risk group (nominal  $P<0.01$ ; Figure 6E).

### Analysis of immune infiltration level

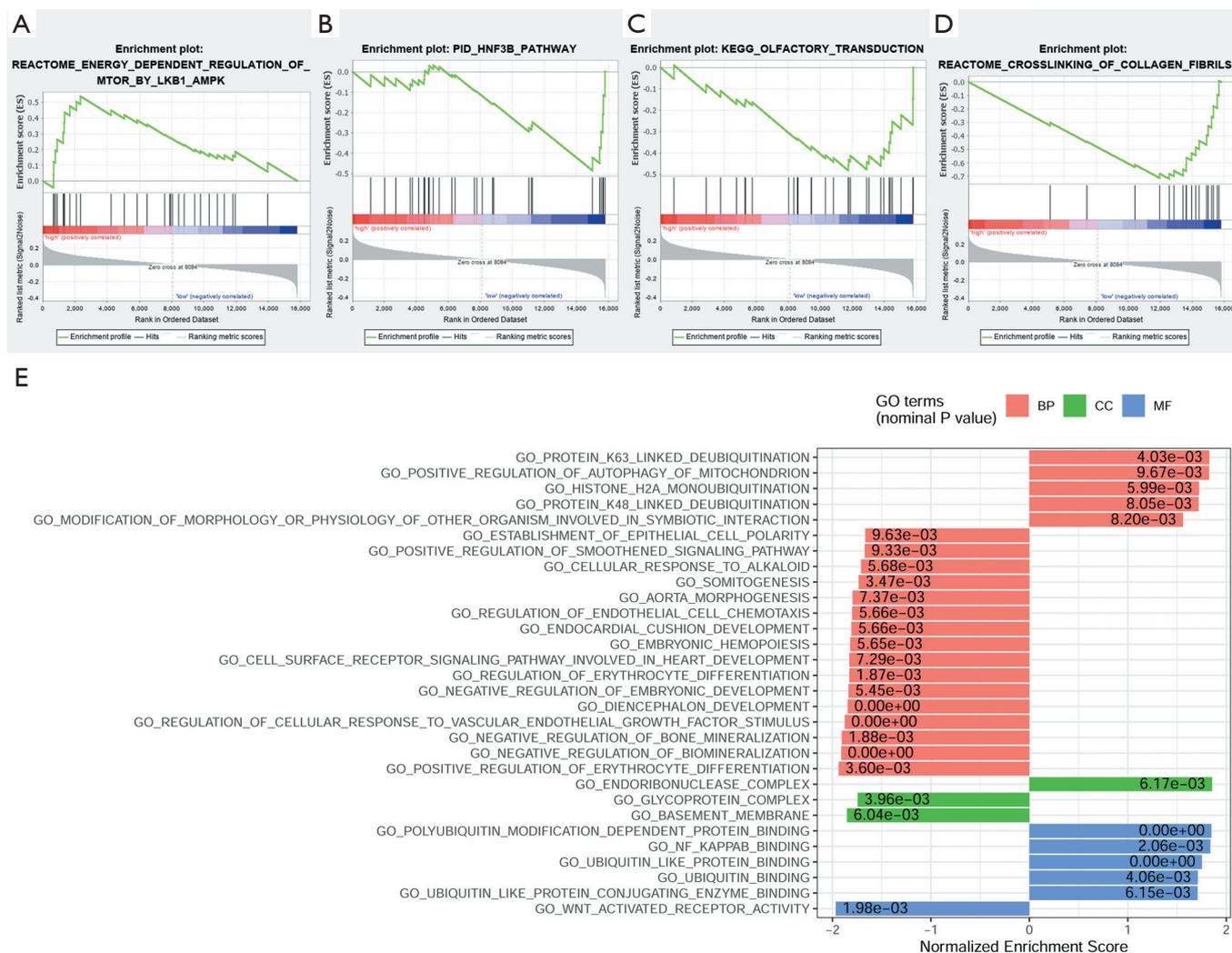
The CIBERSORTx analysis demonstrated that the proportion of monocytes activated dendritic cells and resting mast cells were significantly reduced in the high-risk group compared to that in the low-risk group ( $P=0.0464$ ,  $P=0.0241$ , and  $P=0.00266$ ), and the proportion of M1 macrophages (pro-inflammatory) in the low-risk group was substantially lower than that in the high-risk group ( $P=0.00944$ ; Figure 7A). However, the stromal and immune scores showed no significance between groups ( $P=0.1339$  and  $P=0.8536$ ; Figure 7B,C).

### Discussion

In clinical practice, it is apparent that tumor progression and prognosis of patients are affected by multiple factors, yet many studies have only focused on single or few clinicopathologic factors for WT patients' prognosis (7-9). For instance, Tang *et al.* constructed nomograms to predict OS of children with WT based on 5 clinicopathologic factors, but AUCs of 3- and 5-year OS were not more than 0.74 (45). Nonetheless, prognostic tumor biomarkers also play a vital role in the diagnosis and treatment of tumors. For example, LOH for chromosomes 16q and 1p was demonstrated as a specific marker for increased relapse risk in favorable-histology WT (46). Several retrospective studies of heterogeneously-treated patients suggested an association between a gain of chromosomes 1q and tumor recurrence (47,48). Our previous study also indicated that astrocyte elevated gene-1 overexpression in histologically favorable WT was associated with poor patient prognosis (49).



**Figure 5** Identification of differentially expressed genes and construction of PPI network. Volcano plots of differentially expressed protein-coding genes (A) and lncRNAs (B) between the low-risk group and high-risk group. (C) PPI network of differentially expressed protein-coding genes in the high-risk group. (D,E) Two densely connected regions recognized by MCODE. The upregulated and downregulated genes in the high-risk group were represented by red and blue, respectively. PPI, protein-protein interaction; lncRNAs, long non-coding RNAs; MCODE, Molecular Complex Detection.

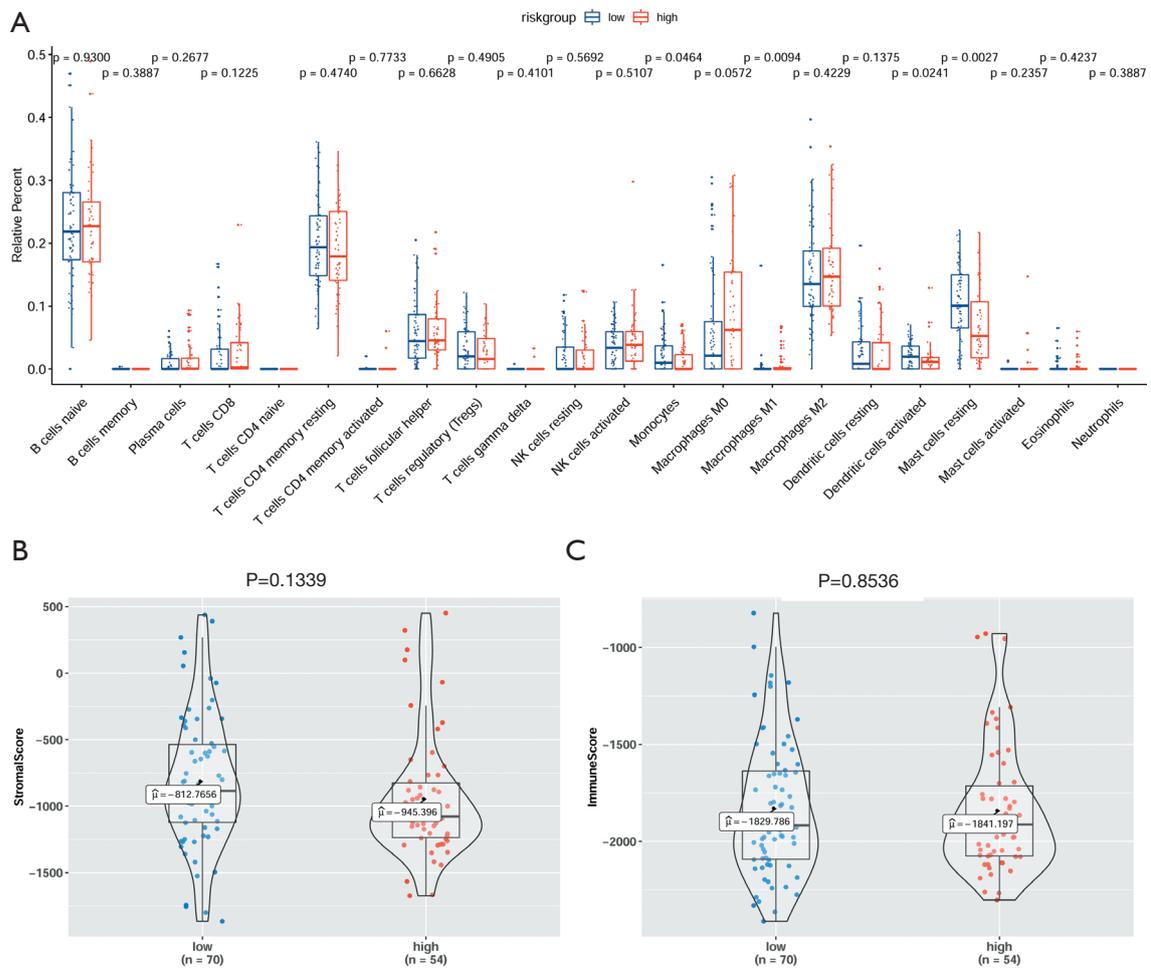


**Figure 6** The results of GSEA between the low- risk group and high-risk group. (A) Energy-dependent regulation of mTOR by LKB1-AMPK from the Reactome subset, which was significantly enriched in the high-risk group. (B,C,D) Three pathways were significantly enriched in the low-risk group. (E) Histogram including nominal P-values of GO enrichment analysis. GSEA, Gene Set Enrichment Analysis; GO, Gene Ontology.

Also, Gong *et al.* have reported a 5-microRNA signature model to predict WT patients' prognosis based on the TARGET database, but AUC was only 0.767 (50). Very few studies currently focus on the development of prognostic models that integrate the clinicopathologic features and biomarkers based on comprehensive sequencing analysis for WT patient prognosis.

In this study, we performed multiple analyses to identify the prognostic lncRNA signature, including filtering of low-expression lncRNAs, identification of lncRNAs associated

with OS through univariate Cox analysis and LASSO method, and identification of independently predictive lncRNAs through multivariate Cox analysis. As a result, the potential minimum number of robust lncRNAs to predict WT patients' prognosis was obtained. Accordingly, a risk-score formula was constructed, and the subsequent log-rank tests and Kaplan-Meier analyses further confirmed the predictive value of the risk score for OS and EFS in the low- and high-risk groups of WT patients. Among the 10 prognostic lncRNAs in the present study, the low expression of EMX2OS (also known



**Figure 7** Analysis of immune cell infiltration levels. Comparisons of the percentage of 22 immune cells (A), the stromal scores (B), and the immune scores (C) between the low- and high-risk groups of WT patients. WT, Wilms tumor.

as EMX2-AS1 or NCRNA00045) was correlated with a poor prognosis in participants with WT [ $\beta = -0.55528$ , hazard ratio (HR) = 0.573914; Table 2]. Interestingly, Gu *et al.* also reported that downregulation of EMX2OS was an independent prognosis factor for shorter recurrence-free survival of classical papillary thyroid cancer (51).

Nomograms also called nomographs or alignment charts, which are widely used as prognostic devices in the field of medicine and oncology, can generate the individual probability of clinical events by integrating multiple prognostic variables (52). In this study, a genomic-clinicopathologic predictive nomogram was developed through incorporating multiple prognostic factors including gender, cooperative group protocol, stage, and risk score to predict 1-, 3-, and 5-year OS of children with WT,

which has shown stronger prediction with better AUCs (1-, 3-, and 5-year OS in the combined model of 0.9272, 0.9428, and 0.9259, respectively) (45,50). The combined model and the risk score model demonstrated significantly higher predictive accuracy, which was further verified by the time-dependent ROC curve and calibration curve analyses. Importantly, there was no significant difference in AUC between the combined and the risk score models, which indicated that the risk score played a critical role in the genomic-clinicopathologic nomogram. Hence, as a powerful and independent prognostic factor, the risk score achieved extremely high accuracy without combining clinicopathologic factors.

Nonetheless, several limitations need to be addressed in the present study. First, WT's sample size was relatively small,

which might lead to bias that could reduce the accuracy of the prognostic model. Second, an external validation cohort for the prognostic model was not available, and all the data analyzed were collected from the TARGET Data Matrix, which might also result in biased discovery. Moreover, the results of bioinformatics were not verified through reverse transcription polymerase chain reaction (RT-qPCR). Finally, besides mRNA sequencing data, other potential biomarkers such as LOH at 1p/16q, gain of 1q, microRNA, and DNA methylation were not integrated into the study.

## Conclusions

In summary, through constructing a genomic-clinicopathologic nomogram, we presented a novel and efficient method for the prognosis of patients with WT and provided insight into the molecular mechanisms that affect WT's prognosis.

## Acknowledgments

The results published here are based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (<https://ocg.cancer.gov/programs/target>) initiative, phs000471.

*Funding:* This study was supported by Shandong Provincial Natural Science Foundation of China (Grant No. ZR2020MH048) and the Science and Technology Development Plan Project of Shandong Province, China (Grant No. 2019GSF108061).

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <http://dx.doi.org/10.21037/tp-20-318>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tp-20-318>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Treger TD, Chowdhury T, Pritchard-Jones K, et al. The genetic changes of Wilms tumour. *Nat Rev Nephrol* 2019;15:240-51.
2. Brok J, Lopez-Yurda M, Tinteren H V, et al. Relapse of Wilms' tumour and detection methods: a retrospective analysis of the 2001 Renal Tumour Study Group-International Society of Paediatric Oncology Wilms' tumour protocol database. *Lancet Oncol* 2018;19:1072-81.
3. Dome JS, Graf N, Geller JL, et al. Advances in wilms tumor treatment and biology: Progress through international collaboration. *J Clin Oncol* 2015;33:2999-3007.
4. Dome JS, Fernandez C V, Mullen EA, et al. Children's Oncology Group's 2013 blueprint for research: Renal tumors. *Pediatr Blood Cancer* 2013;60:994-1000.
5. Dome JS, Perlman EJ, Graf N. Risk stratification for wilms tumor: current approach and future directions. *Am Soc Clin Oncol Educ Book* 2014;215-23.
6. Pritchard-Jones K, Kelsey A, Vujanic G, et al. Older age is an adverse prognostic factor in stage I, favorable histology Wilms' tumor treated with vincristine monochemotherapy: A study by the United Kingdom Children's Cancer Study Group, Wilm's Tumor Working Group. *J Clin Oncol* 2003;21:3269-75.
7. Kaste SC, Dome JS, Babyn PS, et al. Wilms tumour: prognostic factors, staging, therapy and late effects. *Pediatr Radiol* 2008;38:2-17.
8. D'Angelo P, Di Cataldo A, Terenziani M, et al. Factors possibly affecting prognosis in children with Wilms' tumor diagnosed before 24 months of age: A report from the Associazione Italiana Ematologia Oncologia Pediatrica (AIEOP) Wilms Tumor Working Group. *Pediatr Blood Cancer* 2017;64:e26644.
9. Ehrlich PF, Anderson JR, Ritchey ML, et al. Clinicopathologic findings predictive of relapse in children with stage III favorable-histology wilms tumor. *J Clin Oncol* 2013;31:1196-201.

10. Fernandez CV, Mullen EA, Chi YY, et al. Outcome and prognostic factors in stage III favorable-Histology wilms tumor: A report from the children's oncology group study AREN0532. *J Clin Oncol* 2018;36:254-61.
11. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: Insights into functions. *Nat Rev Genet* 2009;10:155-9.
12. Ulitsky I, Bartel DP. XlincRNAs: Genomics, evolution, and mechanisms. *Cell* 2013;154:26.
13. Bolha L, Ravnik-Glavač M, Glavač D. Long Noncoding RNAs as Biomarkers in Cancer. *Dis Markers* 2017;2017:7243968.
14. Clark MB, Mercer TR, Bussotti G, et al. Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat Methods* 2015;12:339-42.
15. Yan X, Hu Z, Feng Y, et al. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell* 2015;28:529-40.
16. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015;47:199-208.
17. Ji Q, Liu X, Fu X, et al. Resveratrol inhibits invasion and metastasis of colorectal cancer cells via MALAT1 mediated Wnt/ $\beta$ -catenin signal pathway. *PLoS One* 2013;8:e78700.
18. Qiu MT, Hu JW, Yin R, et al. Long noncoding RNA: An emerging paradigm of cancer research. *Tumour Biol* 2013;34:613-20.
19. Zhang H, Chen Z, Wang X, et al. Long non-coding RNA: a new player in cancer. *J Hematol Oncol* 2013;6:37.
20. Shi T, Gao G, Cao Y. Long Noncoding RNAs as Novel Biomarkers Have a Promising Future in Cancer Diagnostics. *Dis Markers* 2016;2016:9085195.
21. de Kok JB, Verhaegh GW, Roelofs RW, et al. DD3PCA3, a very sensitive and specific marker to detect prostate tumors. *Cancer Res* 2002;62:2695-8.
22. Luo JH, Ren B, Keryanov S, et al. Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas. *Hepatology* 2006;44:1012-24.
23. Zhang Y, Shields T, Crenshaw T, et al. Imprinting of human H19: Allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. *Am J Hum Genet* 1993;53:113-24.
24. Bartonicek N, Maag JLV, Dinger ME. Long noncoding RNAs in cancer: Mechanisms of action and technological advancements. *Mol Cancer* 2016;15:43.
25. Zhu S, Fu W, Zhang L, et al. LINC00473 antagonizes the tumour suppressor miR-195 to mediate the pathogenesis of Wilms tumour via IKK $\alpha$ . *Cell Prolif* 2018;51:e12416.
26. Zhao XS, Tao N, Zhang C, et al. Long noncoding RNA MIAT acts as an oncogene in Wilms' tumor through regulation of DGCR8. *Eur Rev Med Pharmacol Sci* 2019;23:10257-63.
27. Zhang J, Hou T, Qi X, et al. SOX21-AS1 is associated with clinical stage and regulates cell proliferation in nephroblastoma. *Biosci Rep* 2019;39:BSR20190602.
28. Zhu KR, Sun QF, Zhang YQ. Long non-coding RNA LINP1 induces tumorigenesis of Wilms' tumor by affecting Wnt/ $\beta$ -catenin signaling pathway. *Eur Rev Med Pharmacol Sci* 2019;23:5691-8.
29. Subramanian A, Tamayo P, Mootha VK, et al. 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.
30. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34:267-73.
31. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453-7.
32. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773-82.
33. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.
34. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;47:D766-73.
35. Simon N, Friedman J, Hastie T, et al. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw* 2011;39:1-13.
36. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1-22.
37. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;32:5381-97.
38. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 2008;8:53.
39. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with

- increased coverage and integration. *Nucleic Acids Res* 2013;41:D808-15.
40. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-504.
  41. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2.
  42. Kanehisa M, Sato Y, Furumichi M, et al. New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 2019;47:D590-5.
  43. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes - Release 72.1, December 1, 2014 . *Nucleic Acids Res* 2000;28:27-30.
  44. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019;28:1947-51.
  45. Tang F, Zhang H, Lu Z, et al. Prognostic Factors and Nomograms to Predict Overall and Cancer-Specific Survival for Children with Wilms' Tumor. *Dis Markers* 2019;2019:1092769.
  46. Grundy PE, Breslow NE, Li S, 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;23:7312-21.
  47. Lu YJ, Hing S, Williams R, et al. Chromosome 1q expression profiling and relapse in Wilms' tumour. *Lancet* 2002;360:385-6.
  48. Hing S, Lu YJ, Summersgill B, et al. Gain of 1q is associated with adverse outcome in favorable histology Wilms' tumors. *Am J Pathol* 2001;158:393-8.
  49. Guo F, Zhang LJ, Liu W, et al. Astrocyte elevated gene-1 overexpression in histologically favorable Wilms tumor is related to poor prognosis. *J Pediatr Urol* 2014;10:317-23.
  50. Gong Y, Zou B, Chen J, et al. Potential five-microRNA signature model for the prediction of prognosis in patients with wilms tumor. *Med Sci Monit* 2019;25:5435-44.
  51. Gu Y, Feng C, Liu T, et al. The downregulation of lncRNA EMX2OS might independently predict shorter recurrence-free survival of classical papillary thyroid cancer. *PLoS One* 2018;13:e0209338.
  52. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: More than meets the eye. *Lancet Oncol* 2015;16:e173-80.

(English Language Editors: J. Jones and J. Chapnick)

**Cite this article as:** Zhao H, Wang P, Wang G, Zhang S, Guo F. A long non-coding RNAs expression signature to improve prognostic prediction of Wilms tumor in children. *Transl Pediatr* 2021;10(3):525-540. doi: 10.21037/tp-20-318