Non-coding RNA model improves prognostic prediction in patients with nephroblastoma

Jingzhu Sun[1,a], Zezheng Geng[2,b], Naiwen Liu[3,c], Xiaonan Fang[4,d*]

[1,2,3,4]College of Information Engineering, Shandong Management University, Jinan, China

Abstract. Background. Nephroblastoma (Wilms tumor) is a common abdominal malignancy in children, ranking second among abdominal malignancies in children, but the pathogenesis is still unclear, and further research on their molecular mechanisms is needed.

Method. We obtained lncRNA expression and clinical data from the TARGET database of the GDC data portal of the American Cancer Institute. Effective patient samples were determined based on gene differential expression analysis and clinical data screening. The risk calculation model was established by univariate and multivariate Cox regression analysis, after which the samples were divided into training group and test group to predict the prognosis of patients with nephroblastoma, and then the independent validation on gender was carried out for all samples. Finally, the corresponding target genes of lncRNA were predicted for functional enrichment analysis to explore the enrichment of genes and the interaction between them.

Result. 125 valid samples were identified after screening 136 samples. After experimental analysis, five significant lncRNAs (AC004233.3, AC055764.1, SCAT8, LINC02623, AL118505.1) associated with the prognosis of nephroblastoma were found and validated. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve on the test set was 0.732, which indicated that the model was accurate and the independence verification had good discrimination. In the enrichment analysis, we can intuitively see the significant situation and promotion or inhibition of genes. It can be inferred that the experimental five lncRNAs may regulate the expression of related protein-coding genes and the production process of nephroblastoma, thus affecting the relevant pathways of cancer development.

Conclusions. This study systematically analyzed the lncRNA expression and clinical data of patients with nephroblastoma. The data obtained by establishing Cox proportional hazard regression model can further understand the molecular mechanism related to nephroblastoma and play an auxiliary role in the clinical diagnosis of doctors, thereby improving the diagnosis, treatment and prognosis of patients with nephroblastoma.

1 Introduction

Nephroblastoma is a common pediatric solid tumor, the main population is children, ranking second among abdominal malignancies in children[1]. Its prevalence is about 6% of all tumors in children, 90% of kidney tumors in children, and 1/100,000 in children[2]. Its pathogenesis is still unclear. At present, the main treatment mode is surgery combined with radiotherapy and chemotherapy, but the overall prognosis is not good, which will produce a series of complications and sequelae [3]. Therefore, further studies on its molecular mechanism are needed.

Because of the large number of non-coding RNAs and their role in all aspects of cancer, the focus of research on tumor genes has gradually shifted from coding genes to non-coding genes[4]. Non-coding RNAs are numerous and play a wide role in the occurrence and development of tumors[4]. Long non-coding RNAs are non-coding RNAs that are longer than 200 nucleotides. Although these RNAs do not encode related proteins, they can influence the prognosis of tumors through relatively complex mechanisms[5].

Studies have shown that abnormal expression of microRNAs (miRNAs) can promote cancer cell proliferation and metastasis, and can inhibit cancer cell apoptosis, thereby promoting tumor development[6]. At present, there are relatively few studies on the expression and regulation mechanism of miRNA in nephroblastoma. The main research includes the study on the expression of miR-572 in nephroblastoma tissues by Zhang[7]. By regulating the expression of CDH1 gene, miR-572 can promote the process of epithelial-mesenchymal transition and thus accelerate the metastasis of tumors[7]. The study on the expression of miR-140-5p in nephroblastoma tissues by Liu and Wang et al.indicated miR-140-5pIt can inhibit the expression of TGFBR1 and IGF1R in tumor cells, thereby inhibiting tumor proliferation and promoting apoptosis[8, 9]. In addition, there are studies on miR-539 and miR92a3p, which can inhibit the proliferation, metastasis and invasion of nephroblastoma

[d]14438120180011@sdmu.edu.cn, [a]1295617695@qq.com, [b]972264307@qq.com, [c]sdnwliu@126.com, [d]franknan@126.com

cells by regulating the gene expression of NOTCH1 [10,11].

Up to now, there have been many reports on the mechanism of lncRNA action in nephroblasts. However, they are relatively few in number compared with the rest of the tumors. Through the study of M/Z6455.5Da protein polypeptide, a specific protein marker of nephroblastoma, Sun et al. found that lncRNA UCHL1-AS1 could promote the proliferation, cloning and invasion of nephroblastoma in vitro, and inhibit the apoptosis and cell cycle of nephroblastoma cells. UCHL1-AS1 promotes the growth of nephroblastoma in vitro without significantly affecting the histological structure of the kidney[4].Li et al. Demonstrated that SKI-SC can induce lncRNA RP11-796E2.4 to regulate the expression of pro-apoptotic protein BTG1, thereby causing lethality to nephroblasts [12].

2 Materials and Methods

2.1 Data source

The clinical data and lncRNA expression data of 136 patients with nephroblastoma used in this experiment were obtained from the TARGET database of the GDC data portal of the American Cancer Institute.TARGET (Therapeutic Applicable Research To Generate Effective Treatments, https://ocg.cancer.gov/programs/TARGET) is a database for children's tumors. It uses a comprehensive genomic approach to identify molecular changes during the development of less curable childhood cancers, aiming to use data to guide the development of more effective, less toxic therapies. As the data are from the TAREGET database, further approval by the Ethics Committee is not required.

2.2 Data pre-processing

First, 136 total samples of nephroblastoma cancer (up to Jan. 11, 2020), 130 samples of TP (solid tumor), 6 samples of NT (normal tissue) were downloaded to annotate the mRNA and lncRNA by using gene name file (gencode.v22.annotation.gtf). Then, the annotated files were screened and the expression of mRNA and lncRNA was extracted by using R language. The clinical data mainly come from four XLSX files of clinical data downloaded from TARGET, which are then integrated and saved by R language.

After that, we performed differential expression analysis test on lncRNAs. Firstly, we filtered the expression of RNA (filter criteria: at least 25% of the samples, the expression of more than 2), obtained 9157 lncRNAs, and then identified the tumor samples and normal samples by identifying the sample number, and deleted the normal samples. Then, for FDR, fold change performed multiple tests for the significant expression of genes (FDR < 0.01, fold change > 2). Differentially expressed lncRNAs were obtained by further analyzing the data using the edgeR package (3.28.0). Differentially expressed lncRNAs were used to generate volcano maps using the gplots package (3.0.1.2) and ggplot2 package (3.2.1) of the R platform.

2.3 Survival analysis method

Firstly, we calculated the correlation between lncRNA expression level and overall survival time and status in each nephroblastoma patient by univariate Cox regression analysis.Among them, the p value of univariate Cox analysis results was less than 0.05, indicating that the relationship between the expression level of lncRNA and the overall survival of patients was significant. Then, the significant genes in the univariate Cox regression analysis were screened for univariate robustness using the rbsurv package (2.44.0) in R language. The significant genes in the univariate Cox regression analysis were disrupted, and then randomly selected for calculation. The significant lncRNAs were selected. After multiple iterations (100 iterations in this experiment), the stable lncRNAs in the univariate Cox regression analysis were used.Multivariate Cox regression analysis was performed. In multivariate Cox regression analysis, multiple lncRNAs with smaller p value were selected, and the risk assessment coefficients of these significant lncRNAs were used to construct model formulas, and the survival risk values of each nephroblastoma patient were calculated. According to the calculated survival risk coefficient of each patient with nephroblastoma, we divided the overall patients into high-risk group and low-risk group, and presented the survival status of the two groups of patients through Kaplan Meier survival analysis curve. After that, we validated the independence of the prognostic analysis model of the screened lncRNAs by gender, and all the analyses were done under R platform (version 3.6.2).

3 Results

3.1 Differentially expressed lncRNAs in nephroblastoma

Compared with normal tissues, 2282 lncRNAs were identified to have statistically significant differential expression in nephroblastoma samples (FC > 2 and FDR < 0.01), including 1131 upregulated lncRNAs (49.56%) and 1151 downregulated lncRNAs (50.44%). The volcano map in Figure 1 depicts the distribution of all differentially expressed lncRNAs on -log (FDR) and log FC. The proportion of up-regulated lncRNAs to down-regulated lncRNAs is not much different, the FDR difference in up-regulated lncRNAs is more significant, and the logFC difference in down-regulated lncRNAs is more significant.
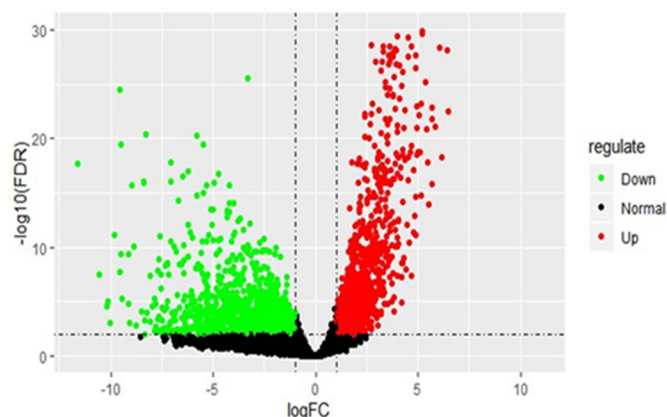
Fig1. Volcano map of lncRNA.

Red dots represent up-regulated genes with fold change > 1, and green dots represent down-regulated genes with fold change < 1.p value < 0.05.

3.2 Cox survival analysis results

Based on the difference analysis data and clinical data, a total of 125 valid samples were randomly divided into training set (n=63) and test set (n=62). After screening, 339 significant univariate lncRNAs were obtained. Robustness screening was performed for these significant univariate lncRNAs, and after multiple iterations, the number of prognostically relevant lncRNAs in

nephroblastoma was finally determined at 10. That is, screening based on sample data analysis resulted in 10 lncRNAs that were significantly associated with overall survival in patients with nephroblastoma (Table 1). Based on the gene expression values of these 10 lncRNAs, we constructed a multivariate Cox regression model for survival prediction of nephroblastoma patients, and extracted the results of the multivariate Cox regression model analysis. Next, the univariate and multivariate analysis results were integrated, and the multivariate model was constructed again to screen the most significant lncRNAs with gene expression. After iteration, five lncRNAs with the most significant gene expression were finally screened out (Table 2).

Table1. Ten lncRNAs were significantly associated with overall survival in patients with nephroblastoma.

lncRNAs	hazard ratio	CI95	P_value
POU6F2_AS1	1.35	1.05-1.75	0.021735
AC124067.3	1.46	1.08-1.97	0.013096
SCAT8	0.54	0.32-0.89	0.015706
AC004233.3	1.61	1.05-2.45	0.027870
AL162713.1	2.28	1.41-3.67	0.000736
AL118505.1	0.49	0.26-0.94	0.031813
AP002370.2	2.51	1.43-4.42	0.001398
LINC02623	0.39	0.17-0.89	0.025790
AC055764.1	0.42	0.25-0.71	0.001003
AL049597.2	2.21	1.11-4.39	0.023125

After robust screening of all highly significant single-factor lncRNAs, the lncRNAs associated with the significant overall survival of Wilms tumor patients were obtained, and the 10 most significant lncRNAs were selected.

In Table 2, five lncRNAs (AC004233.3, AC055764.1, SCAT8, LINC02623, AL118505.1) all showed negative correlation coefficients, which indicated that the higher the expression of genes, the longer the survival of the corresponding patients.

Table2. The five most significant lncRNA correlation coefficients for gene expression

lncRNA	coef	exp(coef)	z
AC004233.3	-0.1830904	0.8326929	-0.6840959
AC055764.1	-0.8738915	0.4173244	-3.1035840
SCAT8	-0.7212346	0.4861517	-2.3909448
LINC02623	-1.2277135	0.2929617	-2.5253246
AL118505.1	-1.1232171	0.3252318	-2.9299362

3.3 Construction of Survival Analysis Model and Overall Experience Verification

After summarizing the results of the multivariate analysis, we then examined the proportional hazard (PH)

Based on the gene expression values of these five lncRNAs, a risk score formula can be constructed to predict the survival of patients with nephroblastoma:

$$Riskscore = \ln cRNA1 \times coef1 + \ln cRNA2 \times coef2 + \cdots + \ln cRNA5 \times coef5 \tag{1}$$

hypothesis and plotted the PH hypothesis significant gene map (Figure 2).
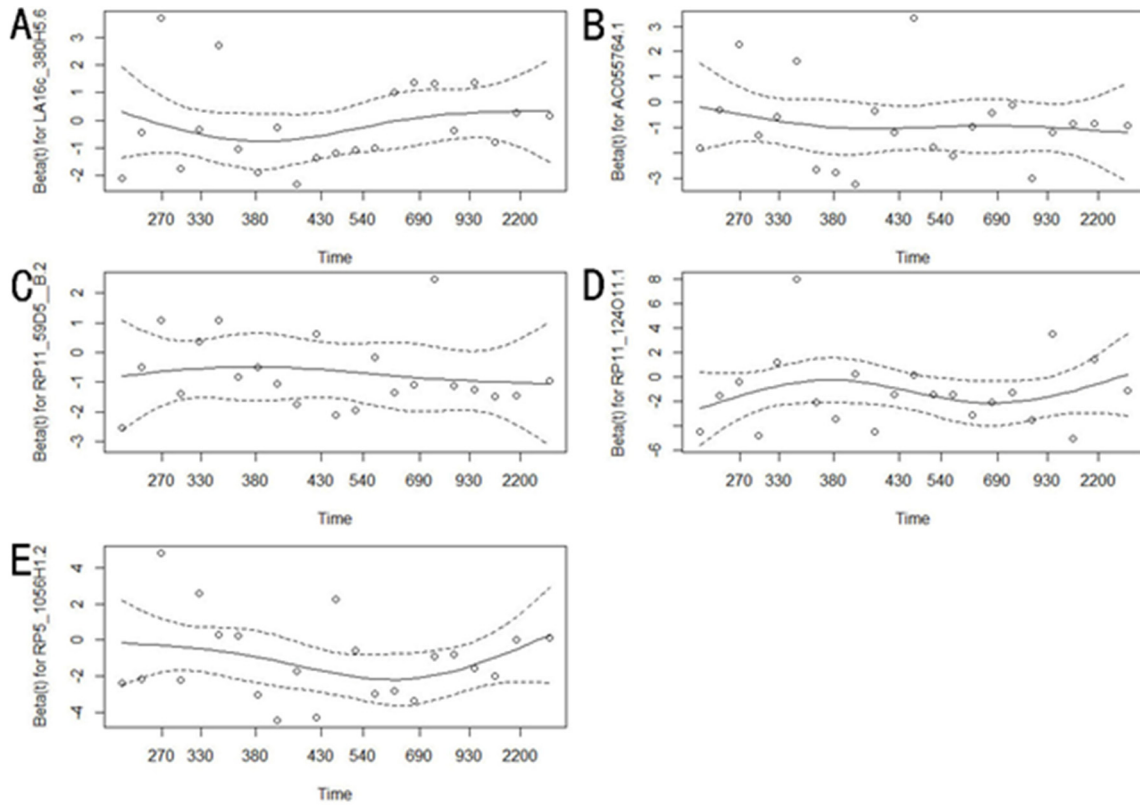
Fig2. Proportional Risk Assumption Diagram for Significant Gene Tests.

Proportional risk (PH) hypotheses were plotted for each of the five significant genes. (A) AC004233.3; (B) AC055764.1; (C) SCAT8 ; (D) LINC02623 ; (E) AL118505.1.

Next, we validated the model on the training data set. First, the risk value of each patient with nephroblastoma was calculated using this formula, and ranked according to the risk value. The training set was divided into two categories, high-risk and low-risk groups according to the risk value. For more intuitive observation, we presented

the results using Kaplan Meier survival analysis curves (Figure 3). It can be found that the prognosis of the high-risk group is significantly worse than that of the low-risk group. To evaluate the performance of this risk model, we used the receiver operating characteristic curve (ROC curve) to analyze the measurement results. The area under the curve (AUC) of the ROC curve obtained by this model was 0.849 (Fig. 4). This result indicates that the risk calculation model has good prognostic survival prediction performance on the training set.
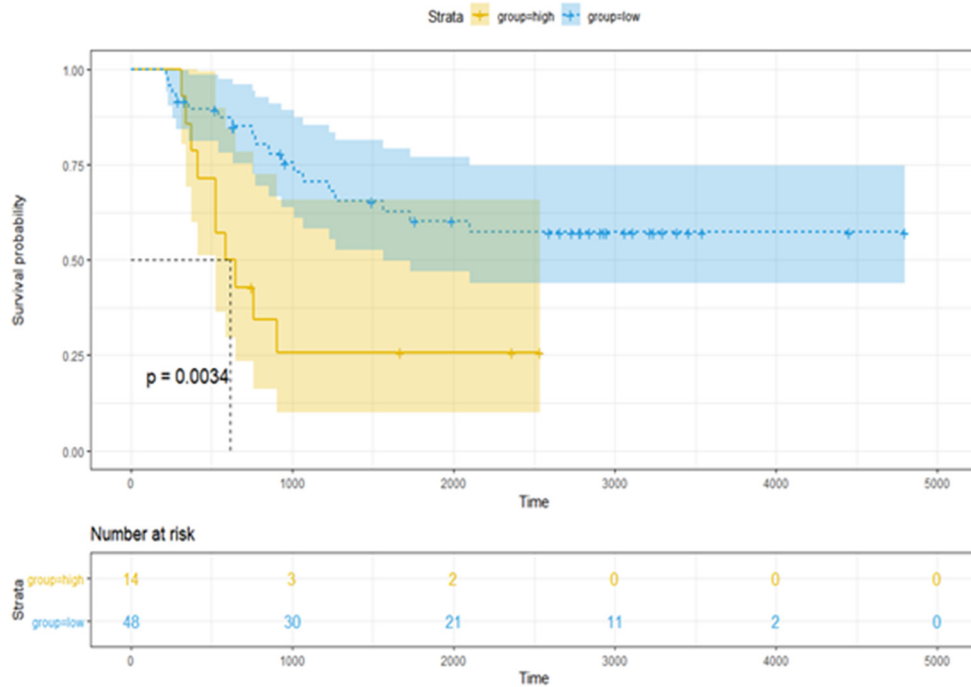
Fig3. Kaplan Meier curves for high-risk and low-risk groups based on the training set.

The blue dotted line is the low-risk group and the yellow solid line is the high-risk group. The bottom layer of the graph is the number of survival samples corresponding to different groups in each time period. p value =0.0034.
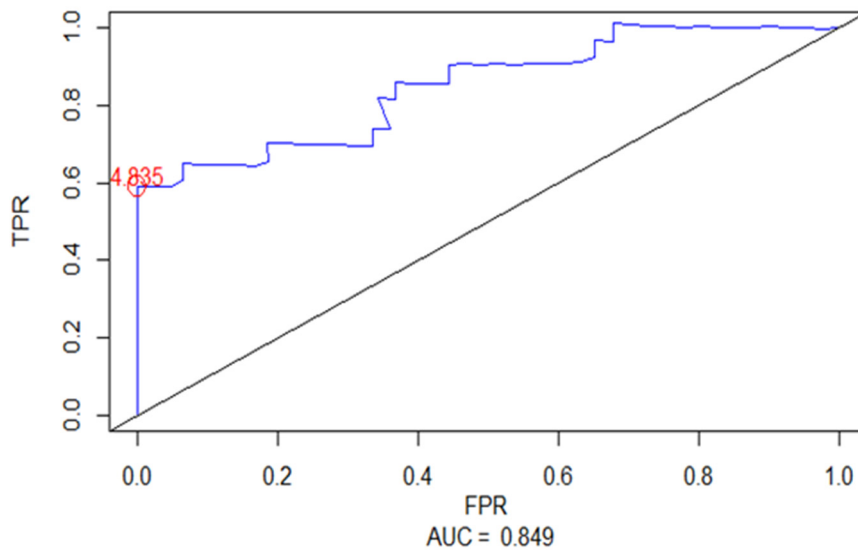

Fig4. Receiver operating characteristic curve based on training set risk calculation model.

Model analysis of the survival of patients with nephroblastoma in the training set yielded good results. Therefore, we performed model validation on the test data set and the whole data set (Fig. 5, Fig. 6), and analyzed the measurement results by receiver operating characteristic curve (ROC curve).The area under the curve (AUC) of the ROC curve of the test set and the whole data set obtained by the model were 0.732 and 0.776, respectively (Fig. 7, Fig. 8), which proved that the model had high stability for survival analysis of patients with nephroblastoma and relatively stable ability for prognostic analysis.
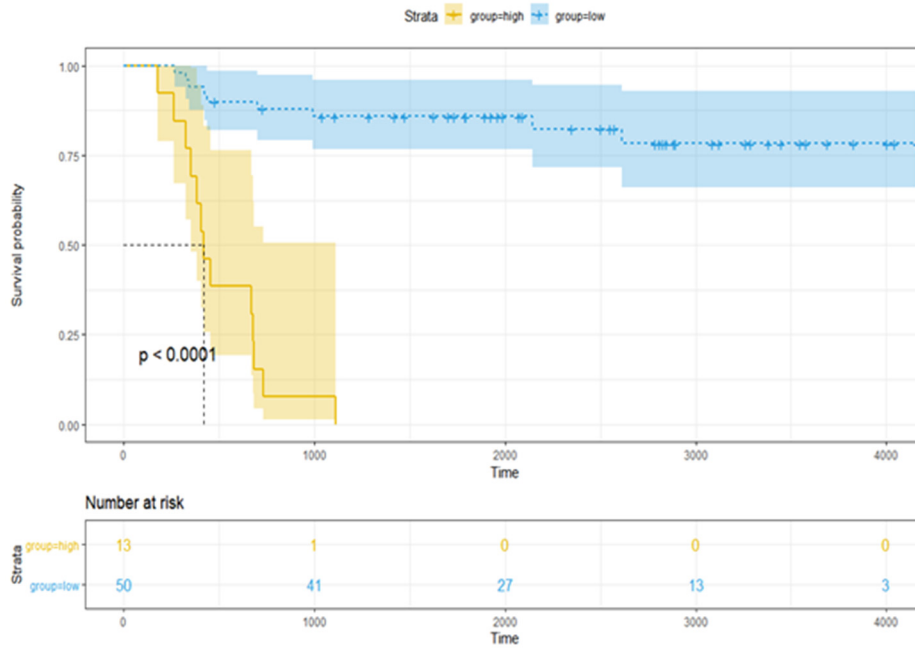
Fig5. Kaplan Meier curves for high-risk and low-risk groups based on the test set.

The blue dotted line is the low-risk group and the yellow solid line is the high-risk group. The bottom layer of the graph is the number of survival samples corresponding to different groups in each time period. p value < 0.0001.
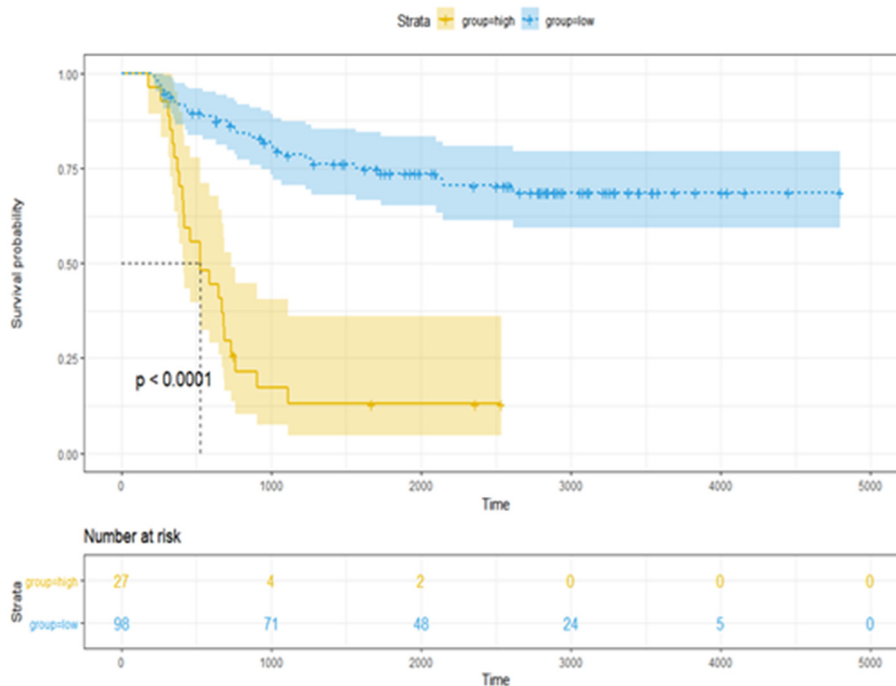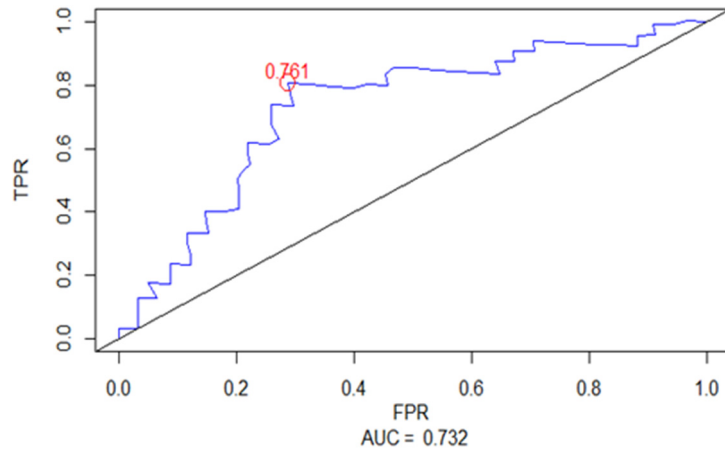

Fig6. Kaplan Meier curves for high-risk and low-risk groups based on the full data set.

The blue dotted line is the low-risk group and the yellow solid line is the high-risk group. The bottom layer of the graph is the number of survival samples corresponding to different groups in each time period. p value < 0.0001.

Fig7. Receiver operating characteristic curve based on test set risk calculation model.
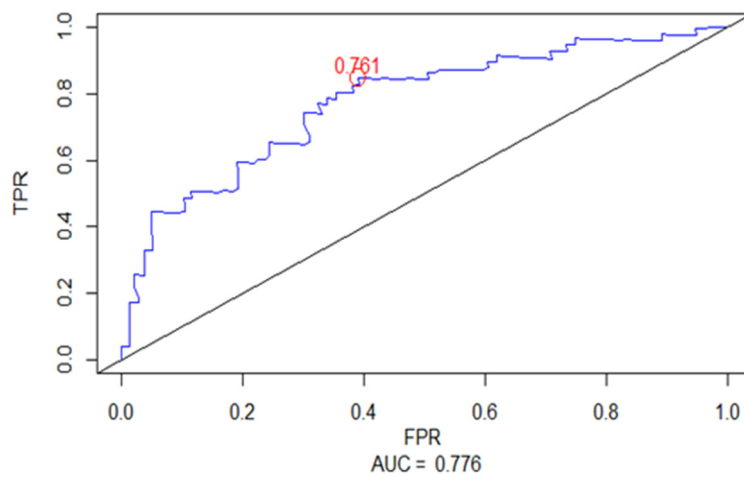

Fig8. Receiver Operational Characteristic Curve of Risk Computing Model Based on All Data Sets.

All the risk values were sorted and the survival time of different individuals was plotted as a sample risk value and survival time distribution map (Figure 9). Black dots represented risk values, red dots represented survival of survival samples, and blue dots represented survival of death samples. It could be found that at low risk values, the survival rate of patients was high and the survival time was long; At high risk values, the survival rate of patients was low and the survival time was short; With the increase of risk values, the survival rate decreased while the survival time was also greatly shortened.
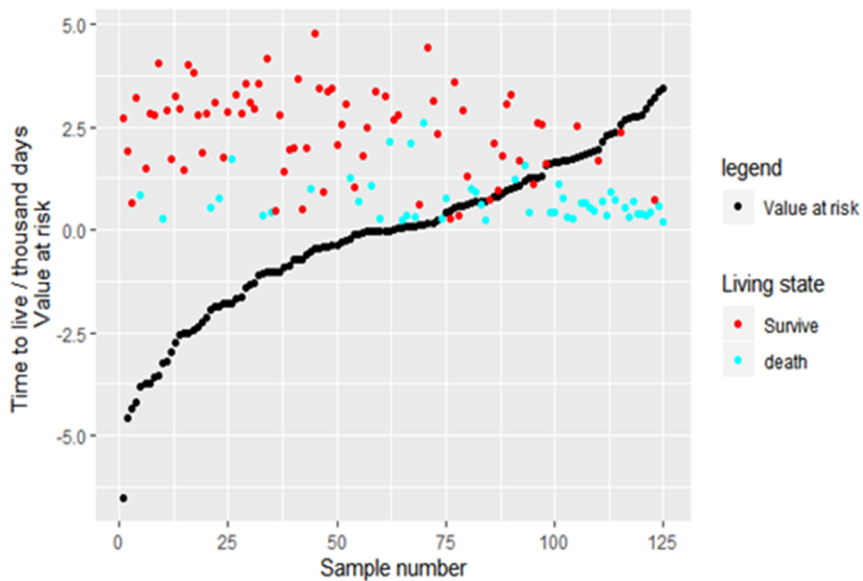

Fig9. Distribution of sample risk value and survival time.

Red dots indicate survival status and blue dots indicated death status, corresponding to Y-axis Time to live/Thousand days.Black dots represented risk values, corresponding to the y-axis Value at risk. The X-axis indicated the sample number.

According to the distribution of these five lncRNAs expression values according to patient risk values, we could draw the expression heat map of lncRNAs (Figure 10). The expression heat map analysis showed that one lncRNA (SCAT8) expression was significantly up-regulated and two lncRNAs (AC004233.3, LINC02623) expression was significantly down-regulated in the low-risk group.In the high-risk group, the expression of five lncRNAs was significantly down-regulated.
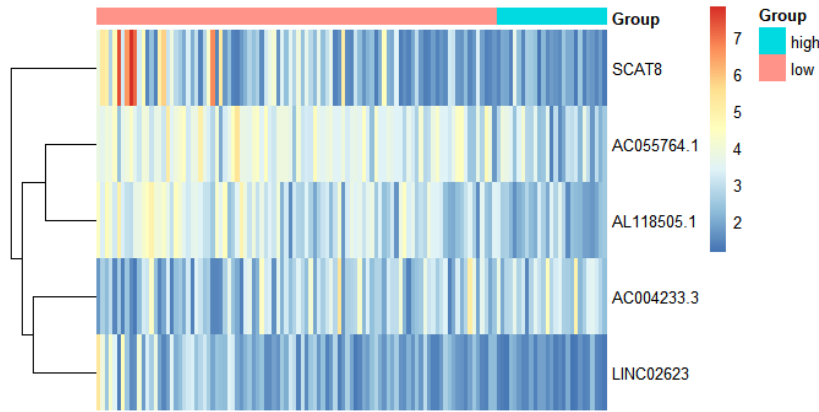

Fig10. Expression heat map of sample lncRNA.

A risk heat map constructed from five lncRNAs.Risk values gradually increased from left to right, with red representing low expression of corresponding genes and blue representing high expression of corresponding genes.

3.4 Independence test of survival analysis model

The overall validation of the model preliminarily proved the stability of the model. In order to further verify the prognostic analysis ability of the risk assessment model for the survival of nephroblastoma, we tested the risk assessment of the model by gender.Firstly, the original data were divided into two groups according to gender, and the risk assessment of the two groups of data was conducted using the model.The results of the analysis showed that different gender patient models could be clearly divided into high-risk and low-risk groups (Figures 11 and 12), and p value < 0.0001 indicated a significant relationship between the two.The analysis results showed that the model had stable prognostic analysis ability for nephroblastoma survival.
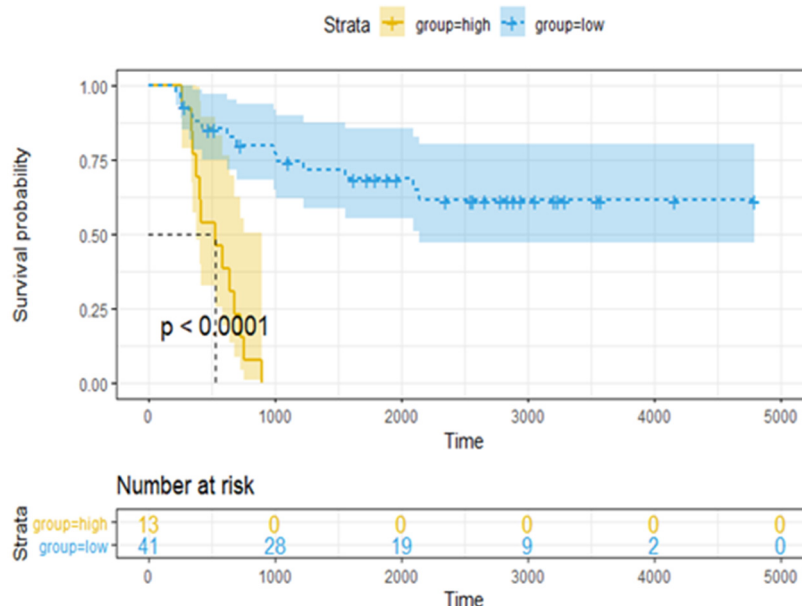

Fig11. Kaplan Meier curves for high-risk and low-risk groups for data from the male group of the independence test.

The blue dotted line is the low-risk group and the yellow solid line is the high-risk group. The bottom layer of the graph is the number of survival samples corresponding to different groups in each time period. p value < 0.0001.

Fig12.   Kaplan Meier curves for high-risk and low-risk groups for data from the female group of the independence test.

The blue dotted line is the low-risk group and the yellow solid line is the high-risk group. The bottom layer of the graph is the number of survival samples corresponding to different groups in each time period. p value < 0.0001.
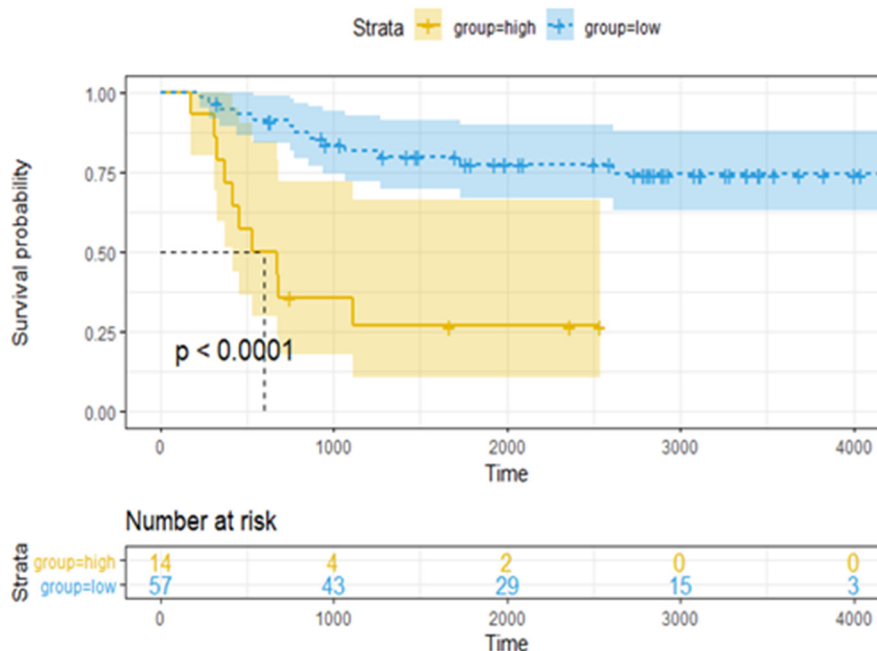
3.5 Functional enrichment analysis

To further clarify the roles of the survival-related lncRNAs we found in nephroblastoma and predict the corresponding target genes of lncRNAs for functional enrichment analysis, we first screened the target genes corresponding to the prognostic lncRNAs through the ENCORI website (http://starbase.sysu.edu.cn/), and then used all the genes of the human genome as the target genes. Background Functional enrichment analysis was performed, mainly using GO enrichment analysis and KEGG pathway enrichment visualization analysis[13]. GO enrichment analysis of biological process (BP), cell composition (CC) and molecular function (MF) of selected target genes was performed based on R language cluster Profiler package (3.14.3) (Fig. 13-15). In BP enrichment analysis, genes were mainly enriched in histone modification, Golgi vesicle transport, regulation of autophagy, Macroautophagy, regulation of intracellular transport, peptidyl-lysine modification, negative regulation of phosphorylation, endomembrane system organization, Axonogenesis, positive regulation of cellular protein localization. In CC enrichment analysis, genes were mainly enriched in ubiquitin ligase complex, cell leading edge, focal adhesion, cell-substrate junction, cell-substrate adherens junction, chromosome, centromeric region, cell cortex, mitochondrial outer membrane, chromosomal region, outer membrane. In MF enrichment analysis, genes were mainly enriched in small GTPase bindingz, transcription coactivator activity, cadherin binding, Ras GTPase binding, protein serine/threonine kinase activity, ubiquitin-like protein transferase activity, ubiquitin-protein transferase activity, ubiquitin-like protein ligase binding, ubiquitin protein ligase activity.Binding, double-stranded RNA binding, can clearly see the enrichment results from the graph, intuitively observe the enrichment and significant situation of genes.In KEGG pathway enrichment analysis, genes were mainly enriched in Proteoglycans in cancer, Cellular senescence, Longevity regulation pathway, Endocytosis, Pancreatic cancer, Axon guidance, Shigellosis, ErbB signaling pathway, Autophagy - animal, Focal adhesion.Significant channels were screened for KEGG channel enrichment analysis (Figure 16). According to the ID column of Table 3, relevant pathways can be downloaded from KEGG official website (https://www.kegg.jp/) and enriched pathway maps could be drawn using the pathview package of R language (1.26.0), which could more intuitively see the significant situation and promotion or inhibition of genes.Based on the above results, we conducted a model to construct a significant lncRNA that may regulate the expression of related protein-coding genes and the production process of nephroblastoma, thus affecting the relevant pathways of cancer development.
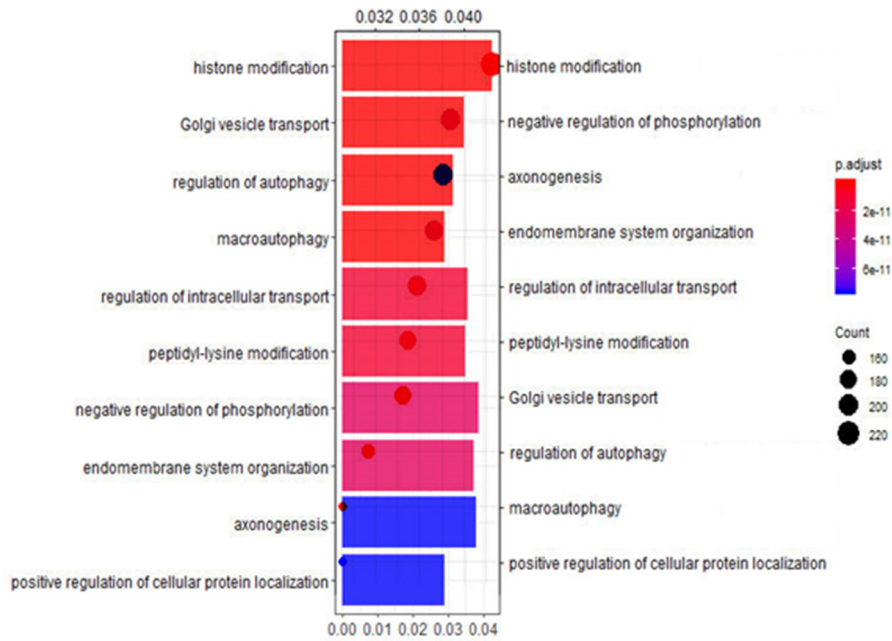
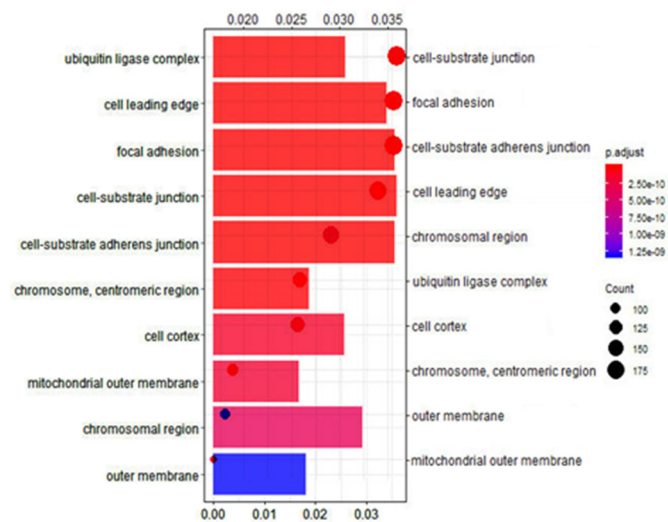**Fig13.**   Column and dot plot of biological process for GO enrichment analysis.

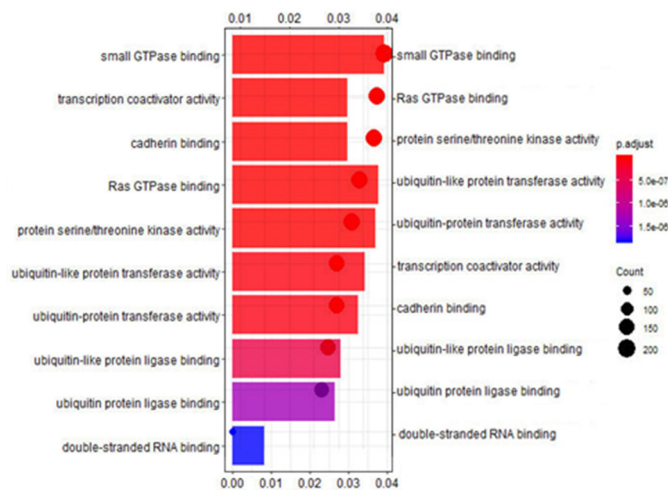**Fig14.**   Column and dot plot of cell composition by GO enrichment analysis.

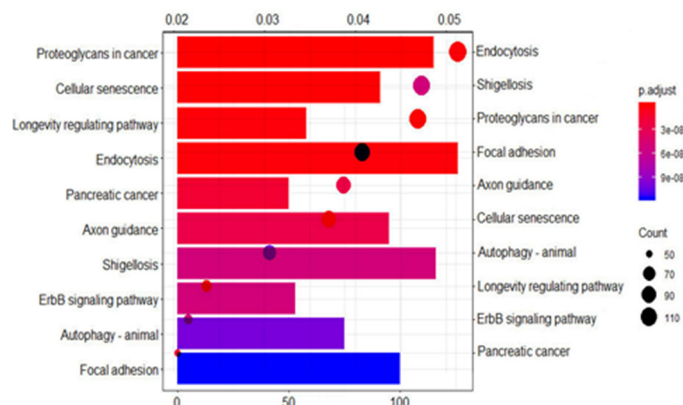**Fig15.**   Molecular function column and dot plot for GO enrichment analysis.

Fig16. KEGG pathway enrichment column, dot plot.

4 Discussion

Nephroblastoma is the most common malignant solid tumor in children, and its pathogenesis remains unclear[3]. According to existing studies, abnormal expression of oncogenes, deletion or inactivation of tumor suppressor genes, gene mutation and abnormal behavior of DNA in children with nephroblastoma can cause tumorigenesis [14,15]. At present, there are relatively few studies on the specific mechanism of lncRNA in nephroblastoma at home and abroad. These genes are not directly involved in the protein coding process, and generally affect the protein coding by regulating the expression level of genes at multiple levels. Based on transcriptome data and clinical data of childhood nephroblastoma in TARGET database, five lncRNAs with the most significant gene expression values were obtained by Cox proportional hazard regression model analysis. Although the five prognosis-related lncRNAs have been less studied in the previous academic literature, the experiments proved that these lncRNAs may participate in the tumorigenesis process of nephroblastoma.

Zhang et al. found that AL118505.1, as a protective factor, was positively correlated with the survival of glioma patients during the process of glioma transcriptome gene research[16]. Florence Jobard et al. found that AL118505 in the chromosome 20 gene overlap group Q8WUT4. C20orf75 affected some of the presumed leucine-rich protein synthesis abnormalities [17]. Chen Ting-hsu et al. found that AC055764.1 can induce activation and actively regulate immune cell transport through a transcriptome study of how seminal plasma (SP) affects human primary endometrial epithelial cells (eEC) and stromal fibroblasts (eSF)[18].

5 Conclusions

Based on the transcriptome data and clinical data of nephroblastoma in TARGET database, we firstly screened out several lncRNAs that were significantly associated with the survival of nephroblastoma patients, then through robust screening and multivariate detection, we finally determined the five lncRNAs with the most significant expression values(AC004233.3, AC055764.1, SCAT8, LINC02623, AL118505.1), and based on this, a Cox proportional hazard regression model was established for the analysis of survival expectations in clinical nephroblastoma patients. After experimental verification, we can basically determine that the risk assessment calculation model based on these five prognostically relevant lncRNAs can independently carry out prognostic analysis for patients with nephroblastoma. The transcriptome data and clinical data of patients with nephroblastoma were combined and analyzed to explore the relationship between genes and cancer. Through functional enrichment analysis, the relationship between lncRNA and the survival of patients with nephroblastoma was more intuitive, so that we had a more intuitive understanding of the protein interaction and gene-related pathways of lncRNA in nephroblastoma, and can have a more comprehensive understanding of its potential biological role. The final results are supported by sufficient data and statistical analysis results. The data obtained from the study can further understand the molecular mechanisms related to nephroblastoma, and the established model can play an auxiliary role in the clinical diagnosis of doctors, thereby improving the diagnosis, treatment and prognosis of patients with nephroblastoma.

Acknowledgments

References

1. FNM Surname (2018). Article Title. Journal Title, 10(3), 1–10.

2. F.N.M Surname, Article Title, https://www.acm.org/publications/proceedings-template.

3. F.N.M Surname and F.N.M Surname, 2018 Article Title, The title of book two (2nd. ed.). Publisher Name, City, State, Country.

4. Ian Editor (Ed.). 2018. The title of book two (2nd. ed.). University of XXX Press, City, Chapter 100. DOI: http://dx.doi.org/10.1000/0-000-00000-0.

5. Bunin GR, Meadows AT. 1993. Epidemiology and Wilms tumor: approaches and methods. MedPediatr Oncol 21(3): 169-171.

6. Verma N,Kumar A. 2016. Clinicoepidemiological profile and outcome of chil- dren with Wilms tumor in a developing country. J Pediatr Hematol Oncol 38（7）：e213－e216.

7. Luo Xin,Deng Changkai,Liu Feng,Liu Xing,Lin Tao,He Dawei,Wei Guanghui. 2019. HnRNPL promotes Wilms tumor progression by regulating the p53 and Bcl2 pathways.OncoTargets and therapy 12(12).

8. Sun Quan. 2019. LncRNA UCHL1-AS1 participates in the mechanism study of M / Z6455.5Da protein peptides in inhibiting Wilms tumor and changes in expression profile. Zhengzhou University.

9. Fu Wen. 2019. Association study of genetic variation of key genes in LIN28 / let-7 pathway and susceptibility to Wilms tumor. Southern Medical University.

10. Luo Xin, Dong Junjun, He Xingyue, Shen Lianju, Long Chunlan, Liu Feng, Liu Xing, Lin Tao, He Dawei, Wei Guanghui. 2019. MiR-155-5p: down-regulated expression in Wilms tumor and inhibits proliferation of Wilms cells , Migration and promote cell apoptosis. Journal of Southern Medical University 39 (12): 1476-1481.

11. Zhang C, Lv GQ, Cui LF, et al. 2019. MicroRNA-572 targets CDH1 to promote metastasis of Wilms' tumor. Eur Rev Med Pharmacol Sci 23(9): 3709-17.

12. Wang HL, Lou CY, Ma N. 2019. miR-140-5p alleviates the aggressive progression of wilms' tumor through directly targeting TGFBR1 gene . Cancer Manag Res 11(9): 1641-51.

13. Liu Z, He F, Ouyang S, et al. 2019.miR-140-5p could suppress tumor proliferation and progression by targeting TGFBRI/ SMAD2 / 3 and IGF-1R/AKT signaling pathways in wilms' tumor. BMC Cancer 19(1): 405-16.

14. Su HL, Wang XE, Song JJ, et al. 2019. MicroRNA-539 inhibits the progression of Wilms' tumor through downregulation of JAG1 and Notch1/3. Cancer Biomark 24(1): 125-33.

15. Zhu S, Zhang L, Zhao Z, et al. 2018. MicroRNA92a3p inhibits the cell proliferation, migration and invasion of Wilms tumor by targeting NOTCH1. Oncol Rep 40(2): 571-8.

16. Li Zhiheng, Tao Yanfang, Xu LiXiao, Zhao he, Wang Nana, Wang Jian, pan Jian. 2014. Study on the mechanism of ski-5c, a new inhibitor of SphK1, promoting the apoptosis of nephroblasts through lncrna. Proceedings of the 8th China Cancer academic conference and the 13th cross strait cancer academic conference.

17. Wu Xiaorong. 2011. Research on automatic acquisition of patient information for clinical decision support. Zhejiang University.

18. Treger Taryn Dora,Chowdhury Tanzina,Pritchard-Jones Kathy,Behjati Sam. 2019. The genetic changes of Wilms tumour. Nature reviews Nephrology Vol.15 (4), pp.240-251 DOI: 10.1038/s41581-019-0112-0.

19. Jiangwen Sun,Henry R Kranzler,Joel Gelernter,Jinbo Bi. 2020. A genome-wide association study of cocaine use disorder accounting for phenotypic heterogeneity and gene–environment interaction. Journal of Psychiatry & Neuroscience:JPN 45(1).

20. G.-H. Zhang,Q.-Y. Zhong,X.-X. Gou,E.-X. Fan,Y. Shuai,M.-N. Wu,G.-J. Yue. 2019. Seven genes for the prognostic prediction in patients with glioma. Springer International Publishing 21(10).

21. Florence Jobard, Bakar Bouadjar, Frédéric Caux, Smail Hadj-Rabia, Christina Has, Fumi Matsuda, Jean Weissenbach, Mark Lathrop, Jean-François Prud'homme, Judith Fischer. 2003. Identification of mutations in a new gene encoding a FERM family protein with a pleckstrin homology domain in Kindler syndrome. Human Molecular Genetics Volume 12, Issue 8 Pages 925–935.

22. Chen Joseph C,Johnson Brittni A,Erikson David W,Piltonen Terhi T,Barragan Fatima,Chu Simon,Kohgadai Nargis,Irwin Juan C,Greene Warner C,Giudice Linda C,Roan Nadia R. 2014. Seminal plasma induces global transcriptomic changes associated with cell migration, proliferation and viability in endometrial epithelial cells and stromal fibroblasts. Human reproduction (Oxford, England) 29(6).