Identification of key carcinogenic genes in Wilms’ tumor

Shaohua He1†, Liu Yang2†, Zhixiang Xiao1, Kunbin Tang1 and Di Xu1*

1Department of Pediatric Surgery, Fujian Provincial Hospital, Fuzhou 350001, China
2Unimed Scientific Inc., Wuxi 214125, China

(Received 14 March 2021, accepted 21 May 2021; J-STAGE Advance published date: 1 August 2021)

This study aimed to probe carcinogenic genes and pathways associated with Wilms’ tumor (WT) onset and malignancy progression. After screening, three datasets acquired from the Gene Expression Omnibus database were analyzed. Differentially expressed genes (DEGs) were identified and GO functional enrichment, KEGG pathway enrichment and protein–protein interaction (PPI) were analyzed. The DEGs with top fold change values or top protein interaction scores were used to analyze overall survival based on the TARGET WT dataset. Together, 866 up-regulated genes in GDS1791, 585 up-regulated genes in GDS2010, and 277 down-regulated genes in GDS4802 were found, from which 46 key DEGs were selected for further analysis. In the PPI network, hub positions included COL5A1, COL4A1, ARPP21, SPARCL1, CD86, LY96 and PPP1R12B. The top DEGs (ARPP21, SYNPO, PRRC2B, PPP1R12B, EFCAB2 and LY96) were selected for survival analysis, and they consistently showed a significantly positive correlation with poor survival. Together, five key carcinogenic genes (SYNPO, PRRC2B, PPP1R12B, EFCAB2 and LY96) were highly associated with WT onset and patient survival. These risk genes, interaction networks and enrichments should improve our understanding of the complex molecular mechanisms in WT development and help clinical applications.

Key words: Wilms’ tumor, nephroblastoma, carcinogen, synaptopodin, bioinformatics analysis

INTRODUCTION

Nephroblastoma, also known as Wilms’ tumor (WT), is the most common malignant solid tumor of the urinary system in childhood. At present, the diagnosis of nephroblastoma is mainly based on the clinical manifestations of children combined with ultrasound, intravenous urography or computed tomography (Khanna et al., 2012; Arnold et al., 2014; Salvatorelli et al., 2015). Surgery and chemotherapy are the main therapeutic methods. The etiology of nephroblastoma is still unclear. Generally, genetic variation or abnormal embryonic development of the genitourinary tract may play a part in its pathogenesis. WT is also considered a prototype of differentiation failure (Emerson et al., 2004; Grieco et al., 2006; Castiglioni et al., 2013). Nevertheless, the cell origin of WT is not entirely clear. It has been proposed that renal stem/progenitor cells undergo a malignant transformation during WT development (Kimura et al., 2010; Pode-Shakked and Dekel, 2011; Tian et al., 2014; Dalpa et al., 2017). Besides common carcinogenic or tumor suppressor genes, such as TP53 and MYC, known WT-related genes including WT1, WTAP, CTNNB1 and WTX are the most strongly associated genes (Ramburan et al., 2005; Guertl et al., 2010; Tian et al., 2014). However, the exact mechanisms of WT are still to be discovered, and specific markers in WT recognition and prognosis prediction are limited. We here aimed to probe more carcinogenic genes and pathways associated with WT onset and malignancy progression. In this bioinformatics analysis, we obtained human RNA expression profiles from three datasets, identified some crucial risk genes, constructed an interaction network, and analyzed functional and pathway enrichments.
METHODS

Microarray data We used the keywords “Wilms’ tumor” or “Wilms tumor” to search gene expression data in the Gene Expression Omnibus (GEO) database. The inclusion criteria were as follows: (1) the experiment was designed to analyze RNA expression; (2) the organism was Homo sapiens; and (3) datasets were acquired through the same or similar platforms. Finally, the following three datasets were included and analyzed.

GDS1791: expression profiling of WT tissue after preoperative chemotherapy. It contains four samples: two samples of Stratagene universal reference RNAs (control) and two samples of WT RNAs. The dataset was acquired using the GPL96 platform (Affymetrix Human Genome U133A Array).

GDS2010: analysis of umbilical vein endothelial cells (HUVEC) following siRNA knockdown (KD) of Wilms’ tumor 1-associating protein (WTAP). WTAP interacts with WT1, a WT suppressor gene. The dataset contains four samples: two controls and two WTAP KD samples. It was acquired using the GPL570 platform (Affymetrix Human Genome U133A Plus 2.0 Array).

GDS4802: analysis of HEK293 embryonic kidney cells overexpressing WTX, a tumor suppressor frequently inactivated in WT. Ectopic expression of WTX in HEK293 cells activates p53 through lysine acetylation. It contains four samples: two controls and two WTX overexpression samples. This dataset was also acquired using the GPL570 platform.

Fig. 1. Differentially expressed genes (DEGs) in three datasets of human Wilms’ tumor samples. (A) GDS1791, (B) GDS2010, (C) GDS4802. (D) Venn diagram of common DEGs. In (A) and (B), some key DEGs screened in Fig. 6 are labeled.
Differentially expressed gene identification  Differentially expressed gene (DEG) analysis of each dataset was performed according to the following standards. A DEG was regarded as a gene with a fold change (FC) ≥ 2 or ≤ 0.5 and $P < 0.05$. Volcano plots were produced to represent DEGs using the ggplot2 R package, and a Venn diagram was drawn to analyze the intersection of DEGs in different datasets.

GO functional enrichment and KEGG pathway enrichment  We used the DAVID tool (https://david.ncifcrf.gov/) to calculate Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. The up-regulated DEGs in GDS1791 and GDS2010, as well as the down-regulated DEGs in GDS4802, were subjected to GO and KEGG pathway enrichment analysis. Homo sapiens was the background for enrichment analysis of all DEGs. Any term with a $P$-value < 0.05 and gene number ≥ 2 was selected as an enriched term. Fold enrichment and $P$-values acquired with the DAVID tool were used to draw bar graphs. Common differential terms (shared by two datasets) were considered as important enrichments.

Protein–protein interaction (PPI) network  Common DEGs were used to construct a PPI network using the Search Tool for the Retrieval of Interacting Genes (STRING) database (http://www.string-db.org), which provides experimental and predicted protein interaction information. The online tool of multiple proteins (organism: Homo sapiens) was used for batch analysis, and the active interaction source included all terms, including text mining, experiments, databases, co-expression,

Fig. 2. Heatmap of 46 key DEGs in Wilms’ tumor.
neighborhood, gene fusion and co-occurrence. The criterion for node connection was ≥ 0.15. Some special node connections (score ≥ 0.7) were used for survival analysis.

**Survival analysis** Among common DEGs, we screened the subclass of genes that were up-regulated in GDS1791, up-regulated in GDS2010 and down-regulated in GDS4802. Theoretically, these are possible carcinogenic (or at least risk) genes in WT. The average FC of the three datasets (up-FC in GDS1791 and GDS2010, but down-FC in GDS4802; note that 0.5 up-FC means 2 down-FC) was calculated, and the top five DEGs were used for overall survival (OS) analysis. Among nodes in the PPI network, those with the highest scores (> 0.5) were also used for OS analysis. The TARGET (Therapeutically Applicable Research to Generate Effective Treatments, focusing on pediatric tumors) database containing 657 WT samples was analyzed. For each gene, patients were divided into two groups: those with specific gene alteration (FC > 2) and those without (FC ≤ 2). OS was compared between the two groups, and a Kaplan–Meier survival curve was drawn for each screened DEG.

**RESULTS**

**DEGs in WT** In the dataset GDS1791, 866 up-regulated and 963 down-regulated genes were observed, as shown in the volcano plot in Fig. 1A. In the dataset GDS2010, there were 585 up-regulated and 611 down-regulated genes (Fig. 1B), and in the dataset GDS4802, there were 306 up-regulated and 277 down-regulated genes (Fig. 1C). The up-regulated genes in GDS2010 and GDS1791 were oncogenesis-related, while, given that the cells in GDS4802 overexpressed the WT suppressor WTX, down-regulated genes in this dataset were potentially our targets. We further analyzed the three subsets GDS1791-up, GDS2010-up and GDS4802-down, and found three intersections (Fig. 1D): 32 DEGs were shared by GDS1791-up and GDS2010-up, eight DEGs were shared by GDS1791-up and GDS4802-down, and another eight DEGs were shared by GDS2010-up and GDS4802-down. No genes were simultaneously shared by all three subsets. Together, 46 key DEGs were selected for further analysis (because the expression of two genes was not probed in GDS1791, the remaining 46 were used); these DEGs may play a role in WT onset and progression. The relative expression of the 46 key DEGs is presented as a heatmap in Fig. 2.

**GO and KEGG enrichment** Using the above 46 key DEGs, GO functional enrichment and KEGG pathway enrichment were performed, but no significant GO functional enrichment or KEGG enrichment was found. Therefore, we input the GDS1791-up, GDS2010-up and GDS4802-down subsets. Common GO function and KEGG enrichments that appeared in two subsets are presented. GDS1791-up and GDS2010-up subsets shared several enriched biological process (BP) and cel-

![Fig. 3](image-url) Common enriched GO function terms among different subsets (GDS1791-up, GDS2010-up and GDS4802-down).
Carcinogenic genes in Wilms' tumor

The cellular component (CC) terms (Fig. 3), including regulation of the multicellular organismal process, regulation of the developmental process, lysosome, lytic vacuole, cytoplasmic vesicle membrane, vacuole, whole membrane and vesicle. In KEGG pathway enrichment, an intriguing finding was that three subsets shared a common term (Fig. 4), hypertrophic cardiomyopathy (HCM), which implied a similar cellular mechanism in WT and HCM. Additionally, the GDS1791-up and GDS2010-up subsets had the following common enrichments: rheumatoid arthritis, intestinal immune network for IgA production, oxytocin signaling pathway, focal adhesion, protein digestion and absorption, platelet activation, PI3K-Akt signaling pathway, cell adhesion molecules and cholinergic synapse. In particular, focal adhesion was found to be slightly enriched (albeit without statistical significance) when the 46 key DEGs (including the three genes COL4A1, PPP1R12B and COL5A1, fold enrichment = 6.4, P = 0.07) were used. Collectively, these functional and pathway terms may constitute the mechanisms underlying WT development.

**PPI network** Applying the STRING database, a PPI network was constructed using the 46 DEGs with a relatively low threshold (score = 0.15). In this network (Fig. 5), COL5A1, COL4A1, ARPP21, SPARCL1, CD86, LY96 and PPP1R12B were located at hub positions. The connections with the three highest scores were COL5A1–COL4A1 (0.972), CD86–CASP1 (0.762), and LY96–CASP1 (0.598).

**DEGs associated with WT patient survival** Among the 46 key DEGs, we screened the potentially most valuable ones according to the direction of their alteration. If a DEG in these 46 was up-regulated in GDS1791 and GDS2010, and down-regulated in GDS4802 (not necessarily significantly altered, given that the intersection of the three subsets was a null subset, as shown in Fig. 1D), the up-FC and down-FC values were used to calculate an average FC (in GDS4802, 0.5 up-FC meant 2 down-FC, and the down-FC value was used to assess the alteration). The top five genes (ARPP21, IGHM, SYNPO, PRRC2B and PPP1R12B) were selected for survival analysis. Besides, the five members (COL5A1, COL4A1, ARPP21, IGHM, SYNPO, PRRC2B and PPP1R12B) were selected for survival analysis.
CD86, CASP1 and LY96) contributing to the three highest connection scores in the PPI network were also used to explore their relationship with survival. In 657 patients, we compared the OS between those with specific gene alterations (average FC > 2) and those without. ARPP21 showed the highest average FC and an essential position in the PPI network, but had only a slight influence on survival (Fig. 6A). Next, high average expression of SYNPO (Fig. 6B), PRRC2B (Fig. 6C) and PPP1R12B (Fig. 6D) was associated with poor OS. IGHM was also among the DEGs with the top five average FC values, but this up-regulation was mainly due to a dramatic FC in GDS1791 and no significant correlation between IGHM and survival was found (data not shown). Additionally, EFCAB2 and LY96 ranked sixth and seventh in the average FC, and LY96 was also among the nodes with mul-

![Protein–protein interaction network of 46 key DEGs.](image-url)
Carcinogenic genes in Wilms' tumor

Multiple connections and top three interaction scores. As expected, high expression of LY96 (Fig. 6E) and EFCAB2 (Fig. 6F) was also significantly correlated with poor survival. There was no association between WT patient survival and the expression of the other four candidates (COL5A1, COL4A1, CD86 and CASP1). Together, SYNPO, PRRC2B, PPP1R12B, EFCAB2 and LY96 were highly possible onco-drivers of WT.

Fig. 6. Survival analysis of key DEGs with top average FC values, or top PPI connection scores and important hub positions within the PPI network. (A) ARPP21, (B) SYNPO, (C) PRRC2B, (D) PPP1R12B, (E) LY96, (F) EFCAB2.
DISCUSSION

In this study, we used three datasets to identify key carcinogenic genes in WT. We screened 46 key candidates, among which five genes were possible risk drivers. The onset of WT may involve some important GO functional enrichments (such as regulation of the multicellular organismal process and regulation of the developmental process) and KEGG pathway enrichment (such as hypertrophic cardiomyopathy and focal adhesion). Our results provide a theoretical perspective on WT development.

Synaptopodin (SYNPO) plays a role in actin-based cell shape and motility of dendritic spines and renal podocyte foot processes. Multiple studies have pointed out that renal diseases have links to abnormal SYNPO expression (Gee et al., 2015; Li et al., 2017). SYNP0 may be regulated by the WT1 transcription factor (Kato and Mizuno, 2017). A study in 2015 identified a segregating missense mutation (R458Q) in WT1 isoform D and reported that WT1 overexpression significantly down-regulated SYNPO expression and promoted apoptosis in HEK293 cells (Hall et al., 2015). This finding is highly consistent with our conclusion. However, the oncogenic role of SYNPO needs more evidence to support it, because there are some controversial findings. For example, miR-206 was reported to be up-regulated in murine nephropathy, which causes severe podocyte injury and inhibits the expression of WT1 and SYNPO (Guo et al., 2016). Our results, for the first time, highlight that SYNPO can be a risk carcinogenic gene in WT.

PRRC2B is a novel marker that is seldom studied in the tumor field. Indeed, only one report has mentioned that fusion transcripts between O6-methylguanine-DNA methyltransferase and PRRC2B might influence the dacarbazine resistance of Hodgkin’s lymphoma cells (Kewitz et al., 2014). Another candidate, PPP1R12B, is a large regulatory subunit of myosin phosphatase. Diseases associated with PPP1R12B include hypotrichosis 1 and tumoral calcinosis. Its impact on tumorigenesis remains largely unknown, although a recent study on colorectal cancer indicated that PPP1R12B was markedly decreased in colorectal cancer tissues, enhancing suppressive effects on cancer cell proliferation and invasiveness (Ding et al., 2019). Combined with our observation, PPP1R12B may play opposite roles in colorectal cancer and WT development. EFCAB2 is also a rarely studied gene that plays a role in calcium ion binding. EFCAB2 expression is increased in meningioma (Gupta et al., 2017), and a significant variant of EFCAB2 was reported in colorectal cancer (Lu et al., 2019). LY96, also named MD2, is associated with Toll-like receptor 4 (TLR4) and confers responsiveness to lipopolysaccharide (Kumada et al., 2008; Yamazoe et al., 2008; Zimmer et al., 2008). An association between LY96 and tumorigenesis has been reported by different teams. The dominant impact was mediated by the TLR4 signaling and downstream inflammatory pathways (Hsu et al., 2011; Zhang et al., 2017; Lujan et al., 2018; Zhou et al., 2018). However, the above five risk drivers have not previously been surveyed in WT. Our analysis revealed that their expression was positively correlated with WT development and predicted poor survival. A composite application of these markers may contribute to WT diagnosis as well as prognosis prediction for WT patients.

Moreover, this is the first work reporting some new and potentially important GO function enrichments in WT, such as regulation of the multicellular organismal process and regulation of the developmental process. Furthermore, in KEGG enrichments, we observed a strong correlation between WT and hypertrophic cardiomyopathy for the first time. These novel findings provide an expanded perspective on WT onset and progression. Focal adhesion was a common enriched pathway using independent analysis of different datasets, and was found to be slightly enriched when we used 46 key DEGs, which is consistent with previous research. For example, WT1 protein and focal adhesion kinase can mediate keratinocyte growth factor signaling in breast cancer cells (Zang et al., 2008). Mutations in WT1 cause a broad spectrum of renal manifestations, which are mediated partially by alterations in the focal adhesion mechanism (Dong et al., 2015).

Nevertheless, this study has some limitations. The major limitation is that only a few samples with different gene expression levels were compared with the majority of the remaining samples in the survival analysis. The reliability of these results remains to be confirmed.

In conclusion, this study has provided an integrative analysis of gene expression profiles and interaction networks regarding WT development. We identified a list of WT markers, among which five key carcinogenic genes were highly associated with WT onset and patient survival. Further experiments are needed to validate these five risk factors. These risk genes, interaction networks and enrichments may provide a better understanding of the complex molecular mechanisms in WT development and help pave the way for clinical applications.

REFERENCES


Carcinogenic genes in Wilms’ tumor


