# Construction of a protein-protein interaction network of Wilms' tumor and pathway prediction of molecular complexes

**W.J. Teng[1]\***, **C. Zhou[2]\***, **L.J. Liu[2]**, **X.J. Cao[3]**, **J. Zhuang[2]**, **G.X. Liu[3] and C.G. Sun[2]**

[1]Shanghai Municipal Hospital of Traditional Chinese Medicine,
Affiliated with Shanghai University of Traditional Chinese Medicine,
Shanghai, China
[2]Cancer Center, WeiFang Traditional Chinese Hospital, Weifang, China
[3]Clinical Institute, Shandong University of Traditional Chinese Medicine,
Shandong, China

**\***These authors contributed equally to this study.
Corresponding author: C.G. Sun
E-mail: 285758078@163.com

**ABSTRACT.** Wilms' tumor (WT), or nephroblastoma, is the most common malignant renal cancer that affects the pediatric population. Great progress has been achieved in the treatment of WT, but it cannot be cured at present. Nonetheless, a protein-protein interaction network of WT should provide some new ideas and methods. The purpose of this study was to analyze the protein-protein interaction network of WT. We screened the confirmed disease-related genes using the Online Mendelian Inheritance in Man database, created a protein-protein interaction network based on biological function in the Cytoscape software, and detected molecular complexes and relevant pathways that may be included in the network. The results showed that the protein-protein interaction network of WT contains 654 nodes, 1544 edges,

and 5 molecular complexes. Among them, complex 1 is predicted to be related to the Jak-STAT signaling pathway, regulation of hematopoiesis by cytokines, cytokine-cytokine receptor interaction, cytokine and inflammatory responses, and hematopoietic cell lineage pathways. Molecular complex 4 shows a correlation of WT with colorectal cancer and the ErbB signaling pathway. The proposed method can provide the bioinformatic foundation for further elucidation of the mechanisms of WT development.

**Key words:** Protein-protein interaction network; Wilms' tumor; Online Mendelian Inheritance in Man; Molecular complex; Pathway

## INTRODUCTION

Wilms' tumor (WT), or nephroblastoma, is the most common malignant renal cancer that affects the pediatric population (Valera et al., 2010). Great progress has been achieved in the treatment of WT. Nevertheless, the protein-protein interaction network underlying WT is not well studied.

Each disease involves complicated pathological interaction of complex genes and pathways as well as a multichannel function of a molecular network. Because individual proteins can rarely perform a given function alone, the proteins involved in the general macromolecular assemblies are formed by the interaction among proteins to perform a biological function in a biochemical process within the same cell. For such processes as copying of the genetic material, gene expression regulation, signal transduction, metabolism, cell proliferation, and apoptosis, the success of the activity depends on the interaction among proteins. Thus, research and analysis of protein interactions and of the interaction network are to be the basis of natural organization of life-related activities, processes, and functions in cells (Zhang, 2009).

Life science research shows that studies on protein interactions not only can yield system level understanding of various biological processes and reveal the development of a disease mechanism but also can help investigators to find new drug targets and facilitate novel-drug research and development (Li et al., 2010). We created protein-protein interaction networks based on biological function, detected molecular complexes, and conducted prediction of the relevant pathways that may be involved in WT. These data can form the basis for further elucidation of the mechanism of WT development.

## MATERIAL AND METHODS

### Design

A study on enrichment of genomic biological pathways.

### Data acquisition

On July 5, 2015, after inputting "(wilms tumor) AND 'Homo sapiens' [porgn:_txid9606]" respectively in the search box on the OMIM home page (http://www.ncbi.nlm.nih.gov/omim), gene information associated with WT was obtained after screening to remove duplicate genes.

## The construction of gene/protein interaction networks

We used the Cytoscape 2.8.3 software to visualize the protein interaction network. Cytoscape is a biological network and annotation data-mining tool; it can be downloaded at http://www.cytoscape.org/ for free. This tool along with the plug-in Agilent Literature Search 2.7.7 (Agilent Technologies, USA) and (Vailaya et al., 2005) helped us to construct a WT protein interaction network. It allows users to browse the data table and to filter and manipulate the related data and so on; thus, restructuring of the data could be visualized in the window of the Cytoscape immediately (Shannon et al., 2003). Visualization of nodes in a network and their attachment to each other (Connection) represent the interaction between a protein and a protein, and the size of the nodes and connections, width, shape, and color are flexible and can be customized.

## Screening of molecular compounds

The plug-in Clusterviz 1.3 was used to conduct the correlation analysis of the area of the construction of biological networks (Saito et al., 2012). Molecular compounds could be screened according to the correlation integral value produced by the MCOMD algorithm in Clusterviz.

## *Identification of biological pathways*

The database for annotation, visualization, and integrated discovery (DAVID; http://david.abcc.ncifcrf.gov/home.jsp) (Huang et al., 2009) provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large lists of genes. Protein names in molecular complexes were submitted to DAVID, and biological pathways involved in WT heredity were identified.

## RESULTS

## WT-related genes in OMIM

After retrieval of data from the OMIM database, we identified 97 relevant genes: *AKR1C3, AMER1, ATF3, BIRC5, BRCA1, BRCA2, CARS, CD2AP, CDC73, CDH1, CDKN1B, CDKN1C, CITED1, CST3, CTGF, CTNNB1, DDX1, DEL11P13, DICER1, DIS3L2, DLG2, EGFR, ELP4, ERBB2, EYA1, GPC3, GPC4, H19, H19-ICR, HDAC4, HGF, HIF1A, IGF1, IGF2, IGF2-AS, IL6, KCNQ1, KCNQ1DN, KCNQ1OT1, KDR, LIN28B, LOC646517, LUZP2, MEOX2, MET, MIR370, MIR562, MTDH, MYCN, NAP1L4, NCAM1, NES, NHS, NOV, NSD1, NUP98, PAX6, PAX8, PCSK9, PDGFRB, PGR, PHLDA2, PIK3CG, POU6F2, PRTN3, PTGS2, RBM4, RET, RPL10, RRM1, S1PR2, SLC22A18, SOSTDC1, STAT1, STAT3, STIM1, TCN2, TERT, TP53, TP73, TSPAN32, TSSC1, TSSC4, USP18, VEGFA, WFDC1, WT1, WT1-AS, WT2, WT3, WT4, WTAP, WTAPP1, WTIP, YAP1,* and *ZNF224*.

## Protein interaction networks

After text mining, 97 genetically related genes showed that they formed a network

---

diagram with 654 nodes (proteins) and 1544 edges, shown in Figure 1. The diamonds represent OMIM genetic-disease-related proteins, while the circles denote the proteins obtained via the text mining.
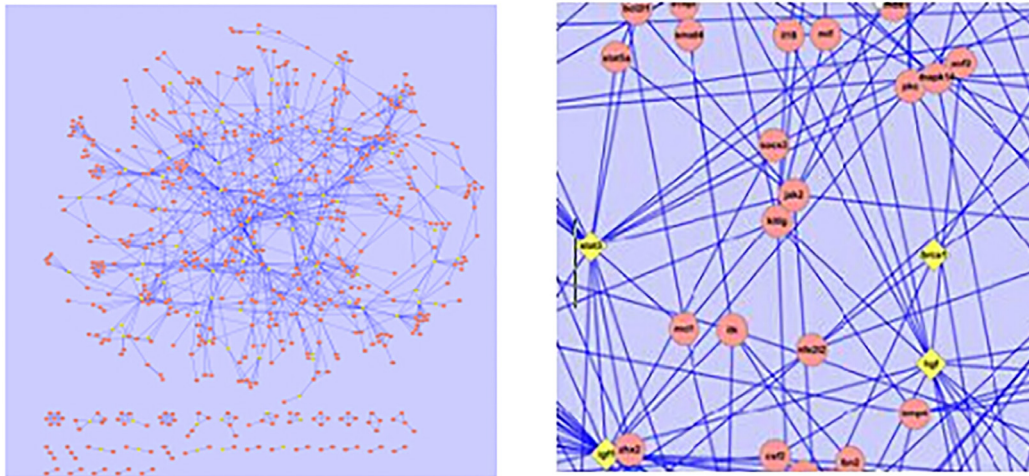


**Figure 1.** Network map of Wilms' tumor protein interactions (Overall + partial, the diamonds represent OMIM genetic disease related proteins, while the ellipses represented the proteins obtained by text mining).

## Analysis of network topology attributes

In the protein interaction network diagram, each node represents a protein, whereas attachment between any two nodes represents an interaction. The network has scale-free characteristics in terms of statistical significance and the nature of the modular level (Burkard et al., 2010). In scale-free networks, an interaction between proteins and $k$ protein probability obeys the inverse power law index distribution. In this case, most of the proteins showed few interactions, while a few proteins interacted with many other proteins. We found that the connectivity of nodes in the network greater than or equal to 20 corresponds to a sharp reduction in the number of nodes (Figure 2). Therefore, we regarded the nodes whose connectivity is greater than or equal to 20 as the key nodes (hubs; Figure 2). Key nodes included *akt1*, *cdkn1c*, *il6*, *pik3ca*, *igf1*, *mapk3*, *tp53*, *egfr*, *stat3*, *vegfa*, *ervk2*, *ephb2*, *mapk14*, *yap1*, *myc*, *igf2*, *h19*, *hdac4*, *hgf*, *wt1*, *ccl2*, and *kcnq1ot1*.

## Detection of molecular complexes

According to the MCOMD algorithm analysis, there were a total of five molecular complexes (Figure 3) whose correlation integral values were higher than 3.

## Enrichment of molecular-complex pathways

We submitted the names of five protein-based molecular complexes online to obtain the relevant data shown in Table 1. We assume that the false discovery rate (FDR) value >1 is meaningful, as shown below:
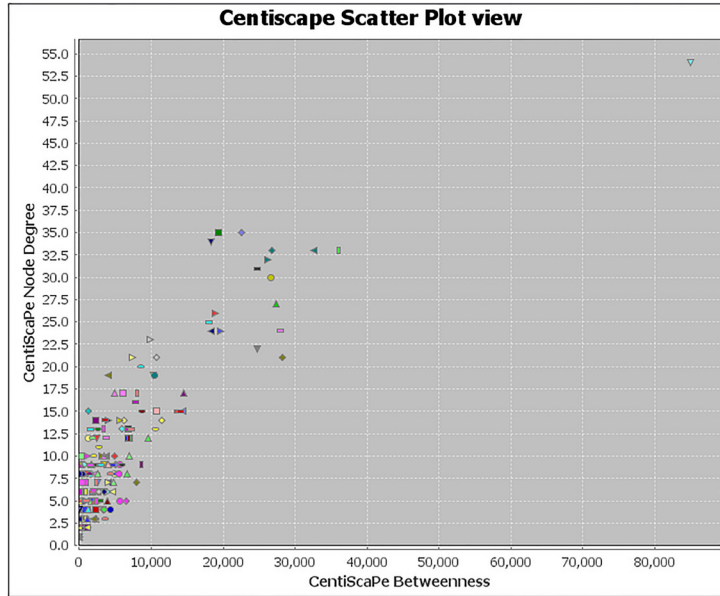
**Figure 2.** Connectivity degree of each node and betweenness comparison. (The horizontal axis represents betweenness, and the vertical axis represents the connectivity degree. The graph shows each node of the network).
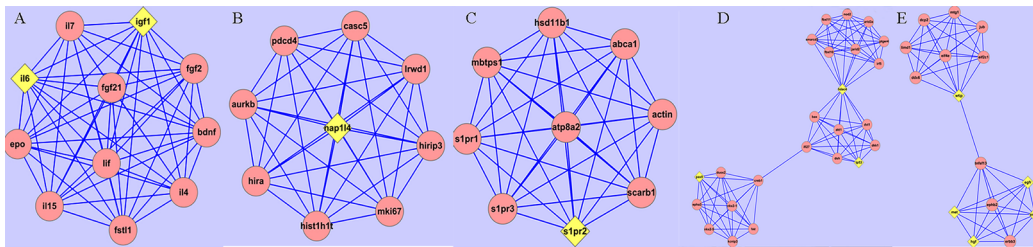


**Figure 3.** Molecular complexes obtained by MCOMD algorithm analysis. **A.** Complex 1 (relation score = 5, 11 nodes and 55 edges). **B.** Complex 2 (relation score = 4, 9 nodes and 36 edges). **C.** Complex 3 (relation score = 4, 9 nodes and 36 edges). **D.** Complex 4 (relation score = 3.875, 24 nodes and 93 edges). **E.** Complex 5 (relation score = 3.333, 15 nodes and 50 edges).

**Table 1.** Enrichment of pathways related to molecular complexes.

| Complex 1 | | | | |
|---|---|---|---|---|
| Category | | P value | Genes | FDR |
| KEGG_PATHWAY | hsa04630: Jak-STAT signaling pathway | 2.82E-06 | *LIF, IL4, IL6, IL7, IL15, EPO* | 0.002420596 |
| BIOCARTA | h_stemPathway: Regulation of hematopoiesis by cytokines | 3.15E-05 | *IL4, IL6, IL7, EPO* | 0.027027912 |
| KEGG_PATHWAY | hsa04060: Cytokine-cytokine receptor interaction | 3.71E-05 | *LIF, IL4, IL6, IL7, IL15, EPO* | 0.031906535 |
| BIOCARTA | h_inflamPathway: Cytokines and Inflammatory Response | 1.76E-04 | *IL4, IL6, IL7, IL15* | 0.150820576 |
| KEGG_PATHWAY | hsa04640: Hematopoietic cell lineage | 3.65E-04 | *IL4, IL6, IL7, EPO* | 0.312776122 |
| Complex 4 | | | | |
| Category | | | | |
| KEGG_PATHWAY | hsa05210: Colorectal cancer | 4.80E-04 | *AKT1, BAX, TP53, DVL1* | 2.22E-05 |
| KEGG_PATHWAY | hsa04012: ErbB sign aling pathway | 6.70E-08 | *AKT1, NRAS, HRAS, PTK2, KRAS, JUN, CBL, MAP2K7, SRC* | 0.45050624 |
| Complex 5 | | | | |
| Category | | | | |
| EC_NUMBER | 2.7.10.1 | 5.25E-07 | *EGFR, ERBB3, ERBB2, MET, EPHB2* | 1.65E-04 |

## DISCUSSION

So far, the biological research on diseases has been focused on the study of a single protein, that is, on research into the characteristics of individual proteins and other relevant proteins able to interact with one another. In contrast, our research topic here is how to integrate these empirical data based on studies of a monomer, and how to comprehend the organism in terms of the path of the interaction and its regulation at the molecular level in a more systematic and comprehensive visual manner. We chose complex networks as a tool to study the overall characteristics of the molecular level in terms of proteins. The studies of protein interaction networks not only have theoretical significance but also hold promise for practical applications. To date, in novel-drug research and development, drugs have typically been tailored to specific targets, in order to regulate or block the operation of the network, but the targets are usually as simple as a given protein function.

For analysis of cell system network features and for identification of common bottlenecks in a metabolic pathway, the methods of current drug discovery still rely on a certain single target damage; this approach imposes limitations on the analysis (Cascante et al., 2002). So far, ~500 human proteins have been identified as drug targets, and these proteins represent the gene function and higher levels of the cell phenotype directly in a one-to-one relation (Huang, 2002).

Accurate detection of a functional module in a PPI network has substantial practical significance: this approach can help to elucidate some unknown protein's function in a process of life, to understand the nature of a physiological activity, improve disease diagnosis, and to study pathogenesis. In terms of protein function prediction, for example, if most proteins that are found in the module have the same function, a researcher can predict the protein's function without annotation: this protein should have the same function as the module. In terms of understanding the network structure and cellular function, biochemical reactions are attributed to the same functional module where proteins interact with one another in mutual coordination. Thus, detection of a functional module in the PPI network means correct understanding of this basic cellular network structure and of the function as an important tool for disease diagnosis and drug research and development. If specific protein interactions show an exception (violate the rule), this can lead to an illness, and functional-module tests can determine which proteins can work with specific proteins. Therefore, we can design drugs to prevent and block the aberrant interaction and thus improve disease treatment.

According to the 97 genes provided by OMIM in this study, we built a WT protein interaction network containing 654 nodes (proteins) and 1544 edges (interactions). Can the network describe molecular regulation of WT during tumor development? According to the existing literature, Hao et al. (1993) found that inactivation of H19, which occurs in 70% of WTs, can promote tumorigenesis via a loss of H19 RNA and/or by cis-activation of the linked and reciprocally imprinted *IGF2* gene (Steenman et al., 1994; Taniguchi et al., 1995). O'Keefe et al. (1997) studied HGF- and Met-positive WTs and demonstrated that a trend toward better prognosis may be counterintuitive when considered in the context of oncogenic signaling; however, this trend is likely to be a reflection of the specific roles of HGF and Met in renal cell differentiation and may still play an important role in diverse characteristics of Wilms' tumorigenesis. Vasei et al. (2009) found that overexpression of the *EGFR* gene in many WT cases warrants further research to determine the therapeutic benefit of EGFR inhibitors in combination with other treatments in patients with WT. The results of Chetcuti

et al. (2011) indicated that MSX1 may be associated with the invasion capacity of WT; RIN1 is a downstream effector of RAS, and Bcl2A1 functions as an antiapoptotic protein; EphB2 is an ephrin receptor and is upregulated in invasive tumors but its role needs to be confirmed in additional cases of WT. Maschietto et al. (2014a) stated that targeting IGF signaling has been tested on WT cell lines and xenograft models, which showed a significant tumor growth delay. Given IGF2 overexpression in WT, targeting IGF1R seems reasonable. IL-6 and STAT3 expression in WT may correlate with progression and predict unfavorable prognosis, highlighting a new therapeutic target for invasive or metastatic WTs (Zhang et al., 2013). Stat3 inhibits WTX expression through upregulation of microRNA-370 in WT (Cao et al., 2013). Stat3 may regulate the expression of HIF-1 and VEGF; therefore, it could be an effective target for inhibition of VEGF expression and angiogenesis in WT (Niu et al., 2007). TP53 mutational analysis improves risk stratification of WTs; this approach requires validation in an independent cohort before clinical use as a biomarker (Maschietto et al., 2014b). WT1 and IFI16 can form a complex in a positive-feedback mechanism in the case of sporadic WTs, in which IFI16 acts as a cofactor to augment the ability of WT1 to activate target genes and support cellular growth (Kim et al., 2008). According to the results of Udtha et al. (2003), in addition to the coexpression of c-MYC and WT1 in embryonic renal mesenchyme, c-MYC is a biologically relevant target of WT1 in WTs.

This knowledge stored in DAVID not only covers extensive annotation of genes in different species but also is enriched in biological information on stand-alone genes. DAVID (Dennis et al., 2003) is a biological information data platform that integrates some biological data and related analytical tools, provides a comprehensive list of genes and proteins for information about large-scale biological function annotation. It consists of a knowledge base and five tools (in a software suite) that are based on the network of functional annotation. These tools include name of gene function points, the viewer, and a genome browser of pathogens. After users upload a list of genes, the platform not only can provide a typical gene-term enrichment analysis (that can be queried) but also can provide enrichment for genes similar to the ones in the list and identify genes and proteins of conversion and other services.

The protein molecule-based biological pathways of complexes 2 and 3 do not exist. There are two possible reasons: 1) although the relevance of these molecular complexes (correlation integral) is higher, this parameter cannot prove that the complex contains a protein with similar biological functions; 2) existing studies have not yet revealed the biological pathways involved.

Molecular complex 1 is predicted to be related to the Jak-STAT signaling pathway, regulation of hematopoiesis by cytokines, cytokine-cytokine receptor interaction, cytokine and inflammatory responses, and hematopoietic cell lineage pathways. Existing literature indicates WTs are highly sensitive to cytokines: the genes of the relevant signaling pathway may help to identify a research direction for molecularly targeted therapy of WT. Molecular complex 4 shows a correlation between WT and colorectal cancer and the ErbB signaling pathway. It has been demonstrated that WT is not simply controlled by a particular gene or signaling pathway but governed by the complex process of a coordinately regulated network system that consists of a variety of signaling pathways and multiple genes. In the signaling network, there is likely some "key regulatory point" that can provide reliable directions for research on molecular pathogenesis, and this work may help to design treatments of WT.

Although we studied signaling pathways in WT here, the pathway of a specific molecular function needs further in-depth research. For the cell function module, there are

many problems that require further studies, for example, how additional types of cells function in a modular manner to perform a particular cellular function, the relation between them, and how to control and assemble these components.

## Conflicts of interest

The authors declare no conflict of interest.

## REFERENCES

Burkard TR, Rix U, Breitwieser FP, Superti-Furga G, et al. (2010). A computational approach to analyze the mechanism of action of the kinase inhibitor bafetinib. *PLOS Comput. Biol.* 6: e1001001. http://dx.doi.org/10.1371/journal. pcbi.1001001

Cao X, Liu D, Yan X, Zhang Y, et al. (2013). Stat3 inhibits WTX expression through up-regulation of microRNA-370 in Wilms tumor. *FEBS Lett.* 587: 639-644. http://dx.doi.org/10.1016/j.febslet.2013.01.012

Cascante M, Boros LG, Comin-Anduix B, de Atauri P, et al. (2002). Metabolic control analysis in drug discovery and disease. *Nat. Biotechnol.* 20: 243-249. http://dx.doi.org/10.1038/nbt0302-243

Chetcuti A, Aktas S, Mackie N, Ulger C, et al. (2011). Expression profiling reveals MSX1 and EphB2 expression correlates with the invasion capacity of Wilms tumors. *Pediatr. Blood Cancer* 57: 950-957. http://dx.doi.org/10.1002/pbc.23003

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, et al. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* 4: 3-5. http://dx.doi.org/10.1186/gb-2003-4-5-p3

Hao Y, Crenshaw T, Moulton T, Newcomb E, et al. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 365: 764-767. http://dx.doi.org/10.1038/365764a0

Huang S (2002). Rational drug discovery: what can we learn from regulatory networks? *Drug Discov. Today* 7 (Suppl): s163-s169. http://dx.doi.org/10.1016/S1359-6446(02)02463-7

Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57. http://dx.doi.org/10.1038/nprot.2008.211

Kim MK, Mason JM, Li CM, Berkofsky-Fessler W, et al. (2008). A pathologic link between Wilms tumor suppressor gene, WT1, and IFI16. *Neoplasia* 10: 69-78. http://dx.doi.org/10.1593/neo.07869

Li MS, Liu CJ, Li D, Liu PL, et al. (2010). Progress of literature mining for protein-protein inter-action information. *Sci. Sinica* 40: 805-819.

Maschietto M, Charlton J, Perotti D, Radice P, et al. (2014a). The IGF signalling pathway in Wilms tumours - a report from the ENCCA Renal Tumours Biology-driven drug development workshop. *Oncotarget* 5: 8014-8026. http://dx.doi.org/10.18632/oncotarget.2485

Maschietto M, Williams RD, Chagtai T, Popov SD, et al. (2014b). TP53 mutational status is a potential marker for risk stratification in Wilms tumour with diffuse anaplasia. *PLoS One* 9: e109924. http://dx.doi.org/10.1371/journal. pone.0109924

Niu ZB, Wang CL, Hou Y, Yang Y, et al. (2007). Expression of Stat3, HIF-1alpha and VEGF in Wilms' tumor. *Chin. J. Contemporary Pediatrics* 9: 461-464.

O'Keefe D, Dao D, Zhao L, Sanderson R, et al. (1997). Coding mutations in p57KIP2 are present in some cases of Beckwith-Wiedemann syndrome but are rare or absent in Wilms tumors. *Am. J. Hum. Genet.* 61: 295-303. http:// dx.doi.org/10.1086/514854

Saito R, Smoot ME, Ono K, Ruscheinski J, et al. (2012). A travel guide to Cytoscape plugins. *Nat. Methods* 9: 1069-1076. http://dx.doi.org/10.1038/nmeth.2212

Shannon P, Markiel A, Ozier O, Baliga NS, et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498-2504. http://dx.doi.org/10.1101/gr.1239303

Steenman MJC, Rainier S, Dobry CJ, Grundy P, et al. (1994). Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nat. Genet.* 7: 433-439. http://dx.doi.org/10.1038/ng0794-433

Taniguchi T, Sullivan MJ, Ogawa O and Reeve AE (1995). Epigenetic changes encompassing the IGF2/H19 locus associated with relaxation of IGF2 imprinting and silencing of H19 in Wilms tumor. *Proc. Natl. Acad. Sci. USA* 92: 2159-2163. http://dx.doi.org/10.1073/pnas.92.6.2159

Udtha M, Lee SJ, Alam R, Coombes K, et al. (2003). Upregulation of c-MYC in WT1-mutant tumors: assessment of WT1 putative transcriptional targets using cDNA microarray expression profiling of genetically defined Wilms' tumors.

*Oncogene* 22: 3821-3826. http://dx.doi.org/10.1038/sj.onc.1206597

Vailaya A, Bluvas P, Kincaid R, Kuchinsky A, et al. (2005). An architecture for biological information extraction and representation. *Bioinformatics* 21: 430-438. http://dx.doi.org/10.1093/bioinformatics/bti187

Valera ET, Brassesco MS, Scrideli CA and Tone LG (2010). Tetraploidization in Wilms tumor in an infant. *Genet. Mol. Res.* 9: 1577-1581. http://dx.doi.org/10.4238/vol9-3gmr880

Vasei M, Modjtahedi H, Ale-Booyeh O, Mosallaei A, et al. (2009). Amplification and expression of EGFR and ERBB2 in Wilms tumor. *Cancer Genet. Cytogenet.* 194: 88-95. http://dx.doi.org/10.1016/j.cancergencyto.2009.06.003

Zhang AD (2009). Protein interaction networks: computational analysis. Cambridge University Press, Cambridge.

Zhang LJ, Liu W, Gao YM, Qin YJ, et al. (2013). The expression of IL-6 and STAT3 might predict progression and unfavorable prognosis in Wilms' tumor. *Biochem. Biophys. Res. Commun.* 435: 408-413. http://dx.doi.org/10.1016/j.bbrc.2013.04.102