

doi: 10.3978/j.issn.2095-6959.2021.01.003

View this article at: <http://dx.doi.org/10.3978/j.issn.2095-6959.2021.01.003>

基于基因芯片的肾母细胞瘤生物信息学分析

瞿根义*, 王佳威*, 徐勇, 阳光, 聂海波, 黄文琳, 汤乘

(中南大学湘雅医学院附属株洲医院泌尿外科, 湖南 株洲 412007)

[摘要] 目的: 整合并利用生物信息学分析肾母细胞瘤基因表达谱芯片, 挖掘肾母细胞瘤发生的关键基因。方法: 利用GEO2R在线分析工具对GEO数据库肾母细胞瘤基因芯片数据GSE2712进行差异表达基因筛选, 使用R软件对差异表达基因进行火山图绘制, 进一步结合DAVID和STRING在线生物信息学工具对差异表达基因进行调控网络分析并构建蛋白质-蛋白质相互作用(protein-protein interaction, PPI)网络, 使用Cytoscape软件进行Hub基因筛选。结果: 共筛选出肾母细胞瘤差异表达基因955个, 其中表达上调基因157个, 表达下调基因798个。对差异表达基因进行GO富集分析和KEGG通路富集分析, 利用在线生物信息学工具构建PPI网络, 使用Cytoscape软件获取PPI网络中前10位Hub基因, 分别是FN1, ALB, VEGFA, KDR, IGF1, PECAM1, FLT1, TEK, FGF1和ANGPT1。结论: 应用生物信息学能有效分析基因芯片数据, 获取肾母细胞瘤发生的关键基因, 为肾母细胞瘤的治疗提供新的思路。

[关键词] 肾母细胞瘤; 生物信息学; 差异表达基因; 关键基因

Bioinformatics analysis of gene chip related to Wilms tumor

QU Genyi*, WANG Jiawei*, XU Yong, YANG Guang, NIE Haibo, HUANG Wenlin, TANG Cheng

(Department of Urology, Zhuzhou Hospital Affiliated to Xiangya School of Medicine, Central South University, Zhuzhou Hunan 412007, China)

Abstract **Objective:** To integrate and use bioinformatics to analyze the gene expression profiling chip of nephroblastoma, and to dig out the key genes of nephroblastoma. **Methods:** The GEO2R online analysis tool was used to screen the differentially expressed genes in the GEO database for renal cell tumor gene chip data GSE2712. The R software was used to map the differentially expressed genes. The volcano map was used to further analyze the regulatory networks and construct the protein-protein interaction (PPI) networks in combination with DAVID and STRING online bioinformatics tools. The Cytoscape software was used to screen the HUB genes. **Results:** A total of 955 differentially expressed genes were screened out, including 157 up-regulated genes and 798 down-regulated genes.

* 为共同第一作者

收稿日期 (Date of reception): 2020-01-05

通信作者 (Corresponding author): 徐勇, Email: tigerhnl@126.com

基金项目 (Foundation item): 湖南省株洲市科技计划项目 (2019-001)。This work was supported by the Science and Technology Guiding Program of Zhuzhou City, Hunan Province, China (2019-001).

GO enrichment analysis and KEGG pathway enrichment analysis of differentially expressed genes were performed using online bioinformatics tools to construct a PPI network. Cytoscape software was used to obtain the top 10 Hub genes in the PPI network, including *FN1*, *ALB*, *VEGFA*, *KDR*, *IGF1*, *PECAM1*, *FLT1*, *TEK*, *FGF1* and *ANGPT1*. **Conclusion:** The application of bioinformatics can effectively analyze the gene chip data, obtain the key genes of nephroblastoma, and provide new ideas for the treatment of nephroblastoma.

Keywords nephroblastoma; bioinformatics; differentially expressed genes; key genes

肾母细胞瘤(nephroblastoma)又称为Wilms瘤,是一种好发于儿童的实体肿瘤,全球范围内每1万名儿童中就有一名儿童患有此病^[1]。肾母细胞瘤是一种胚胎性肿瘤,由基质、上皮组织和间皮组织等组成,在组织学上与肾脏胚胎发生相似^[2]。尽管学者对肾母细胞瘤的了解越来越深,诊疗技术不断提升,但大多数肾母细胞瘤患者预后较差,对于Wilms瘤发病机制也仍不明确。已知部分遗传基因突变,如WT1, WTX及CTNNB1等基因突变都与肾母细胞瘤发病相关^[3]。但是其确切的分子机制仍然未知,因此进一步研究肾母细胞瘤发病机制,了解其发生的关键基因,对于指导患者的治疗及预测患者的预后具有重要的临床意义。

基因芯片技术是高效的、大规模的基因数据获取技术,可以同时研究数以万计的基因表达与疾病之间的关系,尤其是对于肿瘤的机制研究。生物信息学是将计算机技术和分子生物学相结合的技术,为基因的研究提供了明确的方向,揭示大量生物信息所含的奥秘。本研究采用生物信息学技术对肾母细胞瘤的发病相关基因芯片数据GSE2712进行整合和分析,筛选出差异基因,并进行基因富集(Gene Ontology, GO)分析和KEGG通路富集分析,制作PPI (protein-protein interaction)互作网络,旨在寻找肾母细胞瘤发病及发展的关键基因,为进一步研究肾母细胞瘤的研究机制奠定生物学基础。

1 材料与方法

1.1 材料

在GEO Datasets (<http://www.ncbi.nlm.nih.gov/geo>)数据库中以“Wilms”为关键词进行检索,获得由Cutcliffe等^[4]提交的GSE2712芯片数据。GSE2712的芯片平台是GPL96。

该数据集包含Wilms样本18个(登记号: GSM52463~GSM52480),正常肾脏样本3个(登记号: GSM52495~GSM52497)。

1.2 方法

1.2.1 获取差异基因

GEO2R(<http://www.ncbi.nlm.nih.gov/geo/geo2r/>)是一个免费的用于比较几组基因表达的在线工具^[5],利用GEO2R对GEO数据库肾母细胞瘤基因芯片数据GSE2712进行差异表达基因筛选。以 $P<0.05$ 且差异倍数1.5倍获取差异表达基因。并利用R软件绘制差异表达基因火山图。

1.2.2 差异表达基因的GO富集分析和KEGG富集分析

通过DAVID数据库(<https://david.ncifcrf.gov>)对筛选的差异具有统计学意义的基因进行GO富集分析和KEGG通路富集分析, $P<0.05$ 为统计学具有统计学意义,应用R软件及相应的clusterProfiler包进行注释及可视化。

1.2.3 差异表达基因的PPI网络分析

STRING数据库(<https://string-db.org/>)用于识别已知和预测蛋白质与蛋白质之间(PPI)的相互作用^[6-7]。使用STRING对差异表达基因进行分析并构建PPI网络,使用Cytoscape软件中的MCODE获取主要的PPI网络,以及Cytohubba插件筛选前10位Hub基因。

2 结果

2.1 肾母细胞瘤差异表达基因筛选

肾母细胞瘤基因芯片数据GSE2712包含18个Wilms样本和3个正常样本,利用GEO2R进行差异表达基因筛选,以 $P<0.05$ 且差异倍数1.5倍为显著差异获取差异表达基因。共筛选出肾母细胞瘤差异表达基因955个,其中表达上调基因157个,表达下调基因798个,并绘制火山图(图1)。

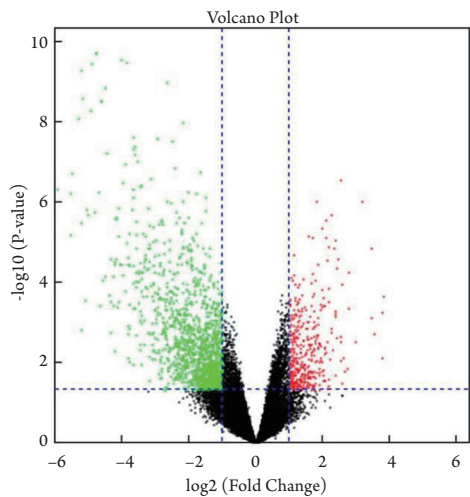


图1 两组样本之间数据的差异表达基因

Figure 1 Differentially expressed genes in the data between two samples

右上角虚线内的点表示基于 $\log FC > 1.5$ 且 $P < 0.05$ 筛选的下调基因；左上角虚线内的点表示基于 $\log FC < -1.5$ 且 $P < 0.05$ 筛选的下调基因；中间虚线外的点表示没有显著差异的基因，FoldChange表示差异倍数。

The dots within the dotted line in the upper right corner indicate down-regulated genes selected based on $\log FC > 1.5$ and $P < 0.05$; the dots within the dotted line in the upper left corner indicate down-regulated genes selected based on $\log FC < -1.5$ and $P < 0.05$; the dots outside the middle dotted line indicate genes that are not significantly different, and FoldChange indicates multiples of difference.

2.2 差异表达基因的GO富集分析和KEGG通路富集分析

通过GO富集分析和KEGG通路富集分析筛选的差异表达基因的生物学功能。在GO富集分析中包括生物学过程(biological process, BP)、细胞组成(cell composition, CC)和分子功能(molecular function, MF)。在BP中差异基因主要富集于细胞外基质组织、细胞黏附、转录的正调控和RNA聚合酶II启动子转录的正调控。在CC中差异基因主要富集于质膜的组成部分、质膜、顶质膜和细胞外区域。在MF中差异基因主要富集于整联蛋白结合、转录激活子活性和RNA聚合酶II核心启动子近端区域序列特异性结合。在KEGG通路分析中主要富集于PI3K-Akt信号通路、癌症的途径、细胞黏附分子(CAMs)和ECM-受体相互作用。主要富集结果见表1、表2和图2。

2.3 差异表达基因的PPI网络分析

通过STRING数据库对肾母细胞瘤发病的差异表达基因构建PPI网络，进一步利用Cytoscape软件中的MCODE获取主要PPI网络(图3)，利用Cytoscape软件中Cytohubba筛选PPI网络中的连接程度前10位Hub基因(图4)，分别是FN1, ALB, VEGFA, KDR, IGF1, PECAM1, FLT1, TEK, FGF1和ANGPT1。

表1 肾母细胞瘤差异表达基因GO富集分析结果

Table 1 GO enrichment analysis results of differentially expressed genes in Wilms tumor

类别	项目	GO 名称	基因数量	P
BP	GO:0030198	细胞外基质组织	36	3.68E-10
BP	GO:0007155	细胞黏附	54	2.34E-08
BP	GO:0001822	肾脏发育	20	3.04E-06
BP	GO:0045893	转录的正调控, DNA 模板	51	3.48E-05
BP	GO:0010628	基因表达的正调控	33	7.82E-05
BP	GO:0071300	对视黄酸的细胞反应	16	2.44E-04
BP	GO:0042493	对药物的反应	35	2.60E-04
BP	GO:0001525	血管生成	29	3.06E-04
BP	GO:0045944	RNA 聚合酶 II 启动子转录的正调控	75	6.61E-04
BP	GO:0030335	积极调控细胞迁移	25	0.00115861
BP	GO:0008283	细胞增殖	37	0.002576456
BP	GO:0008217	血压调节	14	0.003843518
BP	GO:0002062	软骨细胞分化	11	0.005404575
BP	GO:0030855	上皮细胞分化	14	0.009182362

续表1

类别	项目	GO 名称	基因数量	P
BP	GO:0060021	上颌发展	14	0.023511386
BP	GO:0007588	排泄	10	0.027136431
BP	GO:0007507	心脏发育	22	0.042882277
BP	GO:0001649	成骨细胞分化	16	0.043775704
BP	GO:0070062	细胞外泌体	216	2.61E-21
CC	GO:0005887	质膜的组成部分	118	6.04E-12
CC	GO:0005886	质膜	250	2.26E-11
CC	GO:0016324	顶质膜	36	5.25E-06
CC	GO:0005576	细胞外区域	111	5.94E-06
CC	GO:0005578	蛋白质细胞外基质	34	8.15E-06
CC	GO:0005604	基底膜	18	1.15E-05
CC	GO:0031012	细胞外基质	34	9.41E-05
CC	GO:0016323	基底外侧质膜	25	2.21E-04
CC	GO:0005615	细胞外空间	91	5.02E-04
CC	GO:0009986	细胞表面	47	0.001048618
CC	GO:0005788	内质网腔	22	0.03063981
MF	GO:0005201	细胞外基质结构成分	17	1.36E-05
MF	GO:0005509	钙离子结合	61	1.92E-04
MF	GO:0005178	整联蛋白结合	19	3.88E-04
MF	GO:0008201	肝素结合	23	0.001021462
MF	GO:0001077	转录激活子活性, RNA 聚合酶 II 核心启动子 近端区域序列特异性结合	27	0.007466687
MF	GO:0004714	跨膜受体蛋白酪氨酸激酶活性	10	0.027496439

表2 肾母细胞瘤差异表达基因KEGG通路富集分析结果

Table 2 KEGG pathway enrichment analysis results of differentially expressed genes in Wilms tumor

类别	项目	通路名称	基因数量	P
KEGG	hsa04151	PI3K-Akt 信号通路	41	3.74E-04
KEGG	hsa04974	蛋白质消化吸收	18	0.001353958
KEGG	hsa05200	癌症的途径	43	0.001698438
KEGG	hsa04514	细胞黏附分子 (CAMs)	21	0.025091857
KEGG	hsa04512	ECM- 受体相互作用	16	0.025110252
KEGG	hsa04510	黏着斑	26	0.032915572

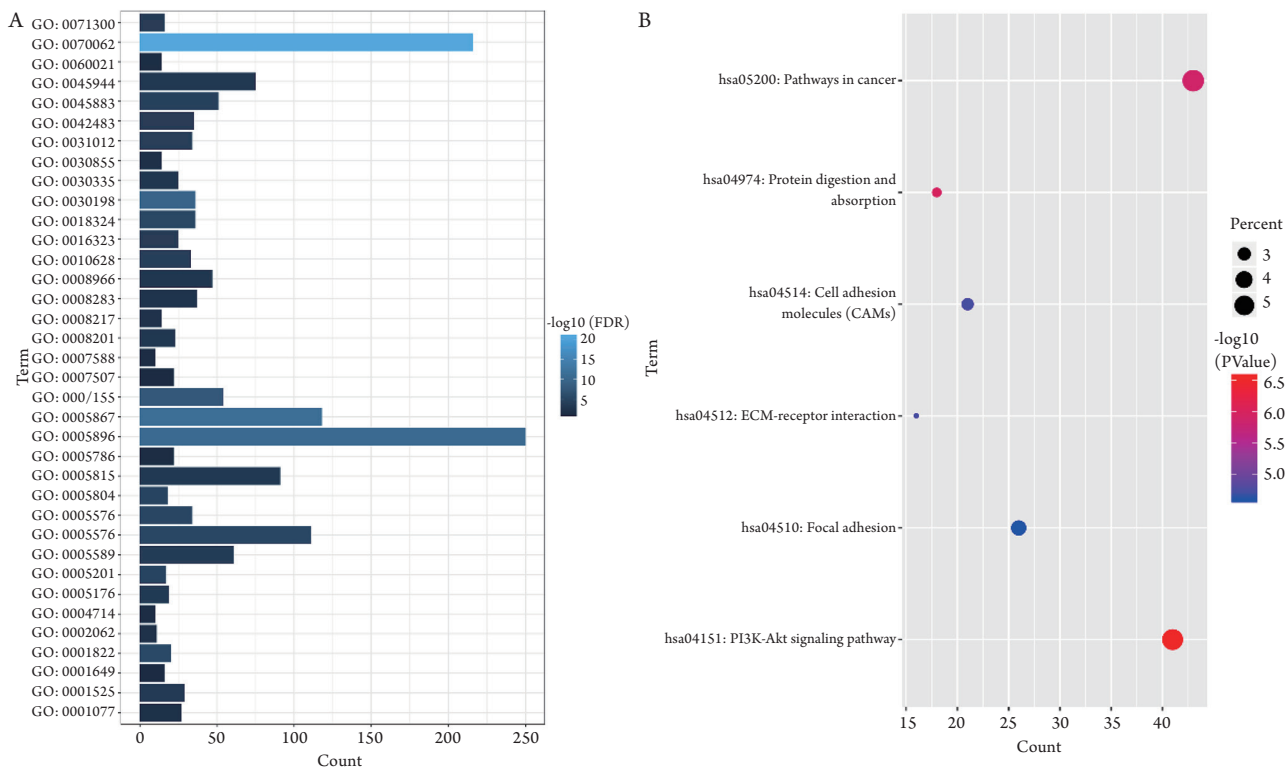


图2 差异基因富集可视化

Figure 2 Differential gene enrichment visualization

(A)差异基因GO富集可视化; (B)差异基因KEGG通路富集可视化。

(A) Visualization of GO enrichment of differential genes; (B) Visualization of KEGG pathway enrichment of differential genes.

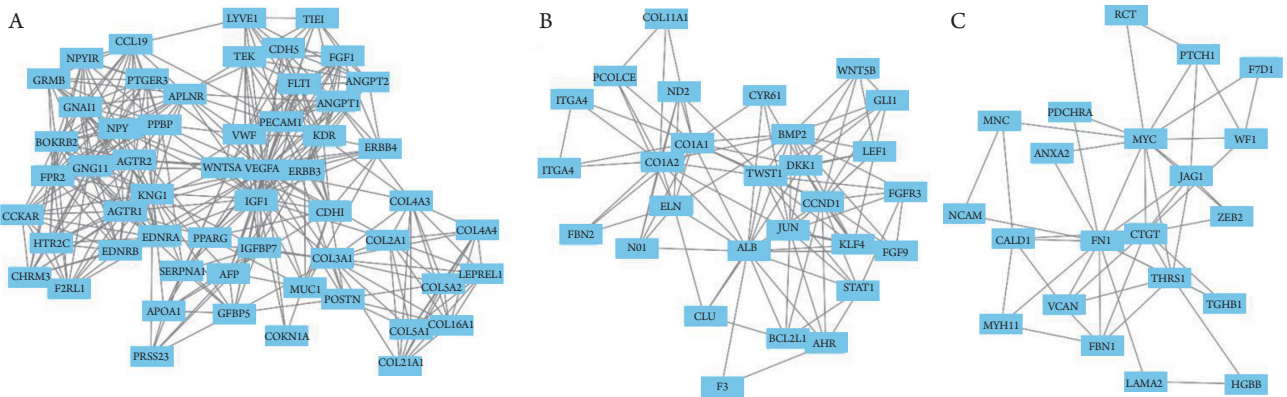


图3 差异表达基因PPI网络: PPI网络的三大模块

Figure 3 Differentially expressed genes PPI network: three modules of the PPI network

(A)模块1; (B)模块2; (C)模块3。

(A) Module 1; (B) Module 2; (C) Module 3.

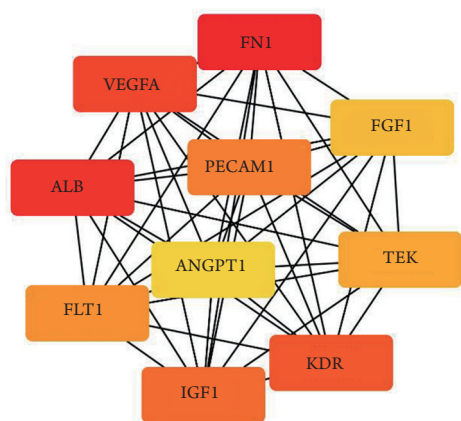


图4 PPI网络筛选的10位 hub基因

Figure 4 Ten hub genes screened by PPI networks

3 讨论

肾母细胞瘤是我国儿童最常见的恶性肿瘤之一。早前,我国肾母细胞瘤的患者预后通常不佳,上世纪30年代肾母细胞瘤患者的生存率仅有约30%^[8]。但是随着肾母细胞瘤的诊断及治疗方法在近年来发展迅速,患者的预后情况较之前得到了明显的改善。目前肾母细胞瘤的主流治疗方法包括了手术治疗、化学药物治疗、放射治疗以及肾静脉栓塞治疗等综合性治疗方法^[9];但是各种治疗方案总有其不足之处,手术治疗会发生肿瘤的局部复发,化学药物治疗可引起各种严重的不良反应,放射治疗会对患儿将来生长发育产生不良影响等。肾母细胞瘤患者的诊断和用药仍然是现在诊治肾母细胞瘤最需要解决的问题,因此关于肾母细胞瘤发生的相关关键基因的研究具有显著意义。随着基因测序技术的发展,以及第二代基因测序技术的出现,为生物信息学提供了丰富的资源^[10]。结合基因芯片大数据的生物信息学分析,在揭示肾母细胞瘤发生发展机制中发挥重要作用。

本研究采用生物信息学和基因芯片技术,首先对GEO数据库中的肾母细胞瘤基因芯片数据GSE2712进行数据挖掘,共挖掘出显著差异基因955个,其中表达上调基因157个,表达下调基因798个,并用DAVID在线工具对差异基因进行富集分析。结果显示:在BP中差异基因主要富集于细胞外基质组织、细胞黏附、转录的正调控和RNA聚合酶II启动子转录的正调控,在CC中差异基因主要富集于质膜的组成部分、质膜、顶质膜和细胞外区域,在MF中差异基因主要富集于整联蛋白结合、转录激活子活性和RNA聚合酶II核心启动子

近端区域序列特异性结合,在KEGG通路分析中主要富集于PI3K-Akt信号通路、癌症的途径、细胞黏附分子(CAMs)和ECM-受体相互作用。进一步通过STRING数据库对肾母细胞瘤发生的差异表达基因构建PPI网络,结果发现这些基因编码的蛋白质调节点主要集中在FN1, ALB, VEGFA, KDR, IGF1, PECAM1, FLT1, TEK, FGF1和ANGPT1。进一步深入挖掘这些基因的相关文献,我们发现这些基因在肿瘤中均有研究报道。

FN1属于细胞外基质糖蛋白类别的成员,参与了细胞黏附和迁移过程,此外还参与了包括胚胎发生,血液凝固,伤口愈合以及宿主防御和转移等多种正常生理反应^[11]。该分子广泛分布在健康的黏膜,固有层,血管结构,神经和平滑肌细胞层中。研究报道在肾癌和结直肠癌中FN1出现显著的高表达,并且在肿瘤的发生发展中发挥着重要的作用^[12]。VEGFA属于血小板衍生生长因子(platelet-derived growth factor, PDGF)/血管内皮生长因子(vascular endothelial growth factor, VEGF)家族的成员,介导血管的生成,内皮细胞的生长,以及促进细胞迁移和抑制细胞凋亡^[13]。而FLT1则是VEGFA的1型受体, VEGFA与FLT1的结合会诱导受体的激活以及随后的信号转导级联反应,从而介导细胞增殖、转化、迁移与凋亡,以及血管的生成,从而在肿瘤发挥重要作用^[14]。并且研究表明FLT1在多种肿瘤中呈过度表达,对于肿瘤的发生、进展、迁移、转移以及化学耐药性都发挥着至关重要的作用^[15]。IGF1则是胰岛素样生长因子家族中的一员,其主要作用于细胞的代谢过程,包括了促进细胞的增殖,介导细胞的有丝分裂,以及抑制细胞的凋亡等。IGF1不仅仅只作用于正常的细胞代谢中,还会参与多种肿瘤的发生、发展及转移,其作用机制在于促进了肿瘤细胞的增殖与转化,并抑制肿瘤细胞的凋亡^[16]。PECAM1也参与了原发性肿瘤生长和增殖的多个过程,包括肿瘤的血管生成和肿瘤细胞外循环等^[17]。FGF家族是人体生长发育,正常生存代谢中无处不在的基因家族,从胚胎生长到器官发育,血管生成到伤口愈合,细胞有丝分裂到细胞癌变,都可见到它们的存在^[18]。多项研究^[16,19]指出FGF-1在肿瘤细胞的代谢中发挥重要作用,参与包括肿瘤细胞的增殖、发育、修复、转移、转化各个阶段。ANGPT1是一种血管生成素家族的糖蛋白,血管生成素家族主要作用是促进新生血管的发育、重塑和成熟,主要机制是与内皮细胞膜上的酪氨酸激酶受体Tie2相互结合而发挥作用^[20]。因

为内皮细胞的萌芽是血管生成的第一步骤, 所以血管生成素家族中的ANGPT1还参与了有丝分裂原活化的蛋白激酶(MAPK)和磷酸肌醇3-激酶(PI3K)等途径。肿瘤的生长可以通过灭活ANGPT1/Tie2信号转导而使肿瘤生长受到抑制^[21]。

本研究致力于分析并确定与肾母细胞瘤发病有关的基因, 共发现955个差异基因, 其中包括上调基因157个和下调基因798个。但是, 仍需要进一步的研究来阐明这些基因在肾母细胞瘤发病机制中的生物学功能, 并为肾母细胞瘤的诊断和治疗提供新的线索和方向。

参考文献

- Charlton J, Pavasovic V, Pritchard-Jones K. Biomarkers to detect Wilms tumors in pediatric patients: where are we now? [J]. *Future Oncol*, 2015, 11(15): 2221-2234.
- Al-Hussain T, Ali A, Akhtar M. Wilms tumor: an update [J]. *Adv Anat Pathol*, 2014, 21(3): 166-173.
- Segers H, Kersseboom R, Alders M, et al. Frequency of WT1 and 11p15 constitutional aberrations and phenotypic correlation in childhood Wilms tumour patients [J]. *Eur J Cancer*, 2012, 48(17): 3249-3256.
- Cutcliffe C, Kersey D, Huang CC, et al. Clear cell sarcoma of the kidney: up-regulation of neural markers with activation of the sonic hedgehog and Akt pathways [J]. *Clin Cancer Res*, 2005, 11(22): 7986-7994.
- Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor [J]. *Bioinformatics*, 2007, 23(14): 1846-1847.
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets [J]. *Nucleic Acids Res*, 2019, 47(D1): D607-D613.
- Alhunaidi O, Ahmad AA, El-Nahas AR, et al. Impact of case volume per year on flexible Ureteroscopy practice: an internet based survey [J]. *BMC Urol*, 2019, 19(1): 134.
- Hamilton TE, Shamberger RC. Wilms tumor: recent advances in clinical care and biology [J]. *Semin Pediatr Surg*, 2012, 21(1): 15-20.
- 白明, 张敬悌, 王盛兴, 等. 小儿肾母细胞瘤的治疗及生存分析 [J]. *现代泌尿外科杂志*, 2018, 23(10): 765-768.
- BAI Ming, ZHANG Jingti, WANG Shengxing, et al. Treatment and survival analysis of nephroblastoma in children [J]. *Journal of Modern Urology*, 2018, 23(10): 765-768.
- 徐培杰. 生物信息学研究现状 [J]. *科技信息*, 2013, 25(10): 268-269.
- Xu Peijie. Research Status of Bioinformatics [J]. *Science & Technology Information*, 2013, 25(10): 268-269.
- Ding Y, Pan Y, Liu S, et al. Elevation of MiR-9-3p suppresses the epithelial-mesenchymal transition of nasopharyngeal carcinoma cells via down-regulating FN1, ITGB1 and ITGAV [J]. *Cancer Biol Ther*, 2017, 18(6): 414-424.
- Steffens S, Schrader AJ, Vetter G, et al. Fibronectin 1 protein expression in clear cell renal cell carcinoma [J]. *Oncol Lett*, 2012, 3(4): 787-790.
- Huang Y, Goel S, Duda DG, et al. Vascular normalization as an emerging strategy to enhance cancer immunotherapy [J]. *Cancer Res*, 2013, 73(10): 2943-2948.
- Luttun A, Tjwa M, Carmeliet P. Placental growth factor (PlGF) and its receptor Flt-1 (VEGFR-1): novel therapeutic targets for angiogenic disorders [J]. *Ann N Y Acad Sci*, 2002, 979: 80-93.
- Shibuya M. Vascular endothelial growth factor receptor-1 (VEGFR-1/Flt-1): a dual regulator for angiogenesis [J]. *Angiogenesis*, 2006, 9(4): 225-230; discussion 231.
- 苏影, 许晶晶, 李宇宁, 等. 胰岛素样生长因子1在肾癌组织中的表达研究进展 [J]. *中国医药导报*, 2019, 16(11): 27-29, 41.
- SU Ying, XU Jingjing, LI Ningning, et al. Progress in the expression of insulin-like growth factor -1 in renal cell carcinoma [J]. *China Medical Herald*, 2019, 16(11): 27-29, 41.
- Yang JF, Shi SN, Xu WH, et al. Screening, identification and validation of CCND1 and PECAM1/CD31 for predicting prognosis in renal cell carcinoma patients [J]. *Aging (Albany NY)*, 2019, 11(24): 12057-12079.
- Beenken A, Mohammadi M. The FGF family: biology, pathophysiology and therapy [J]. *Nat Rev Drug Discov*, 2009, 8(3): 235-253.
- Takeuchi Y, Daa T, Kashima K, et al. Mutations of p53 in thyroid carcinoma with an insular component [J]. *Thyroid*, 1999, 9(4): 377-381.
- Eklund L, Olsen BR. Tie receptors and their angiopoietin ligands are context-dependent regulators of vascular remodeling. *Exp Cell Res* 2006, 312(5): 630-641.
- Liang H, Ge F, Xu Y, et al. miR-153 inhibits the migration and the tube formation of endothelial cells by blocking the paracrine of angiopoietin 1 in breast cancer cells [J]. *Angiogenesis*, 2018, 21(4): 849-860.

本文引用: 瞿根义, 王佳威, 徐勇, 阳光, 聂海波, 黄文琳, 汤乘. 基于基因芯片的肾母细胞瘤生物信息学分析 [J]. *临床与病理杂志*, 2021, 41(1): 20-26. doi: 10.3978/j.issn.2095-6959.2021.01.003

Cite this article as: QU Genyi, WANG Jiawei, XU Yong, YANG Guang, NIE Haibo, HUANG Wenlin, TANG Cheng. Bioinformatics analysis of gene chip related to Wilms tumor [J]. *Journal of Clinical and Pathological Research*, 2021, 41(1): 20-26. doi: 10.3978/j.issn.2095-6959.2021.01.003