

· 基础研究 ·

骨肉瘤与尤文肉瘤关键基因 CDC5L 及其生物学功能的生物信息学分析

冯娜¹, 宋文霞², 邵立平¹, 高宇航³, 赵承武⁴

(1. 吉林大学第一医院骨关节外科及运动医学护理平台, 吉林 长春 130021; 2. 吉林大学第一医院病理科, 吉林 长春 130021; 3. 吉林大学第一医院骨关节外科, 吉林 长春 130021; 4. 吉林大学第一医院运动医学科, 吉林 长春 130021)

【摘要】 目的: 利用生物信息学方法筛选骨肉瘤(osteosarcoma, OS)和尤文肉瘤(Ewing’s sarcoma, EWS)关键基因及通路预测。方法: 检索 GEO 数据库 OS(GSE16088) 及 EWS(GSE45544)数据集, 利用 GEO2R 分析差异基因, 韦恩图筛选共同参与两者发生的关键基因, 对该基因进行 GO、KEGG 通路分析及蛋白质-蛋白质相互作用 (protein-protein interaction, PPI) 网络分析, 并利用 UCSC 进行该基因的转录因子预测, GEPIA 分析转录因子与目的基因之间相关性。结果: 通过对 OS 及 EWS 数据集差异性及维恩图分析, 结果显示细胞分裂周期 5 样蛋白 (cell division cycle 5-like, CDC5L) 基因为参与 OS 及 EWS 发病的关键基因。该基因主要参与有丝分裂过程, 与 RNA 代谢、Pre-mRNA 内含子的处理、mRNA 及 Pre-mRNA 的剪接等信号通路。结论: CDC5L 作为关键基因, 在 OS 及 EWS 进展中发挥促癌基因的作用, 可能是两种原发恶性肿瘤可靠的诊疗靶点。

【关键词】 骨肉瘤; 尤文肉瘤; 生物信息学分析; CDC5L

中图分类号: R738.1

DOI: 10.12200/j.issn.1003-0034.2022.03.015

开放科学(资源服务)标识码(OSID):



Bioinformatics-based identification of key genes CDC5L and related pathways in osteosarcoma and Ewing’s sarcoma
FENG Na, SONG Wen-xia, SHAO Li-ping, GAO Yu-hang, and ZHAO Cheng-wu*. *Department of Sports Medicine, the First Hospital of Jilin University, Changchun 130021, Jilin, China

ABSTRACT Objective: Osteosarcoma(OS) and Ewing’s sarcoma (EWS) are the two most common primary malignant bone tumors in children. The aim of the study was to identify key genes in OS and EWS and investigate their potential pathways. **Methods:** Expression profiling (GSE16088 and GSE45544) were obtained from GEO DataSets. Differentially expressed genes were identified using GEO2R and key genes involved in the occurrence of both OS and EWS were selected using venn diagram. Gene ontology and pathway enrichment analyses were performed for the ensembl. Protein-protein interaction (PPI) networks were established by STRING. Further, UCSC was used to predict the transcription factors of the cell division cycle 5-like (CDC5L) gene, and GEPIA was used to analyze the correlation between the transcription factors and the CDC5L gene. **Results:** The results showed that CDC5L gene was the key gene involved in the pathogenesis of OS and EWS. The gene is mainly involved in mitosis, and is related to RNA metabolism, processing of capped intron-containing pre-mRNA, mRNA and pre-mRNA splicing. **Conclusion:** CDC5L, as a key gene, plays a role in development of OS and EWS, which may be reliable targets for diagnosis and treatment of these primary malignant tumors.

KEYWORDS Osteosarcoma; Ewing’s sarcoma; Bioinformatics; CDC5L

骨肉瘤(osteosarcoma, OS)和尤文肉瘤(Ewing’s sarcoma, EWS)是儿童和青少年中常见的原发性恶性骨肿瘤,发病人数可占到骨肿瘤的约 1/2。当前治疗方案包括手术、放疗和化疗,尽管有多模式的治疗方法,仅有不到 20%的患者可通过手术治愈,治疗失败及误诊的患者仍占有较高的比例^[1-2]。因此需要进一步探索其发病机制,发现疾病的潜在诊断和治疗

靶点。随着基因组学的重大突破,利用生物信息学方法进行大数据挖掘分析,可为疾病的发生发展提供可能的分子机制依据,为实验研究提供可行的思路。本研究应用生物信息学的方法分析基因表达谱芯片数据,筛选共同参与 OS 与 EWS 两种常见的原发骨肿瘤的基因及信号通路。为寻找新型的生物标记物,为 OS 与 EWS 发生发展提供可能分子机制及潜在的治疗靶点,并实现精准医疗的目标。

1 材料与方法

1.1 数据收集与预处理

以“Osteoarthritis normal”及“Ewing sarcoma nor-

通讯作者: 赵承武 E-mail: chengwuzhao@jlu.edu.cn

Corresponding author: ZHAO Cheng-wu E-mail: chengwuzhao@jlu.edu.cn

mal”作为检索词,在 NCBI 的基因表达汇编(GEO)数据库(<http://www.ncbi.nlm.nih.gov/geo/>)检索相应的基因芯片^[3-4]。纳入标准:(1)选取人类组织芯片。(2)表达谱的芯片,下载符合实验要求的 2 套数据集,芯片类型分别为 Affymetrix Human Genome U133A 和 Affymetrix Human Gene 1.0 ST 芯片^[5-6]。见表 1。

表 1 GEO 数据库 2 个微阵列数据库信息

Tab.1 Statistics of the two microarray databases derived from the GEO database

肿瘤类型	Dataset ID	平台	肿瘤	正常	总数
骨肉瘤	GSE16088	GPL96	14	6	20
尤文肉瘤	GSE45544	GPL6244	14	21	35

1.2 差异基因筛选

GEO2R 在线分析工具(<https://www.ncbi.nlm.nih.gov/geo/geo2r/>)检测肿瘤组与正常组织之间的差异基因,筛选标准为调整后 $P < 0.05$ 和变化幅度 ≥ 2 倍的差异基因,即 $\log_2 \text{fold change} (\log_2 \text{FC}) > 1$ 。每个数据集进行分析后,选取前 250 个差异基因,针对 1 个基因对应多个探针的情况,删除重复项,并删除无基因名称项。使用韦恩图(<http://bioinformatics.psb.ugent.be/webtools/Venn/>)选取共同参与两种肿瘤发生的差异基因。

1.3 GO 和 KEGG 分析

利用 ensembl 网站(<http://grch37.ensembl.org/index.html>)进行单基因 Gene Ontology (GO) 和 KEGG 通路分析。GO 分析可分为生物学进程(biological process, BP), 细胞组分 (cellular component, CC) 及分子功能(molecular function, MF) 3 个部分。

1.4 蛋白质相互作用网络

STRING (<https://string-db.org/>) 数据库是一

个在线搜索已知蛋白相互作用的检索工具,涵盖 5 090 种生物体,2 460 万种蛋白质,超过 30 亿种蛋白相互作用。利用该数据库构建蛋白质-蛋白质相互作用(protein-protein interaction, PPI)网络^[7-8]。

1.5 转录因子预测

使用 NCBI(<https://www.ncbi.nlm.nih.gov/>)查找细胞分裂周期 5 样蛋白 (cell division cycle 5-like, CDC5L) 基因启动子序列,在基因起始位点上游的 2 000 bp 及下游 100 bp 为潜在的启动子区域,启动子区域位置为 Chr6:44385707-44387806。利用 UCSC(<http://genome.ucsc.edu/>)进行启动子预测,筛选得分 > 600 的转录因子,采用 GEPIA(<http://gepia.cancer-pku.cn/>)网站分析转录因子与目的基因之间的相关性。

1.6 统计学处理

用 SPSS 软件进行统计分析,采用 GraphPad 8.0 软件 CDC5L 表达量作图。CDC5L 表达量呈正态分布,采用 t 检验进行统计分析。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 基因筛选

选取调整后 P 值最小的前 250 个基因,删除无基因名称及重复的差异基因后,GSE16088 及 GSE45544 基因集剩余基因数目分别为 216、214 个。利用韦恩图对两种肿瘤的数据集取交集,发现共同参与 OS 与 EWS 发生的基因为 CDC5L (图 1a)。CDC5L 基因在两种肿瘤组织中均为高表达 (图 1b, 1c)。

2.2 GO 与 KEGG 信号通路分析

CDC5L 单基因 GO 分析结果显示,主要参与的生物过程是有丝分裂的细胞周期、mRNA 剪接、DNA 修复及 mRNA 处理;主要分布在 Prp19 复合体、细胞

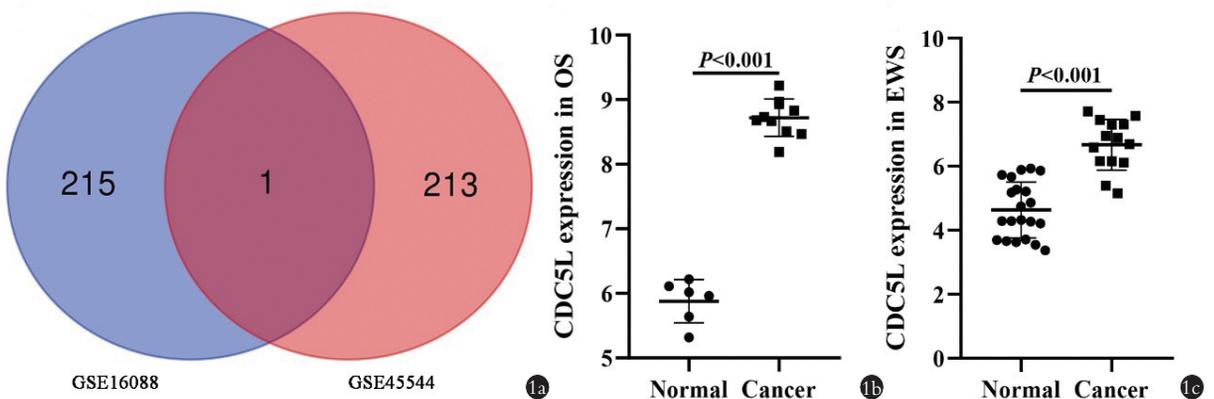


图 1 筛选差异基因及基因表达水平 1a. 共同参与 OS 与 EWS 发生基因韦恩图 1b. CDC5L 在 OS 及正常组织表达量 1c. CDC5L 在 EWS 及正常组织表达量

Fig.1 Screen differential genes and gene expression levels 1a. Venn diagram of DEGs common to all two GEO datasets 1b. CDC5L expression in OS and normal tissues 1c. CDC5L expression in EWS and normal tissues

核、细胞质及复制因子 A 蛋白复合物;分子功能主要 RNA 聚合酶 II 调控区序列特异性 DNA 结合、DNA 转录结合因子活性及 RNA 聚合酶 II 特异性、转录辅阻遏子及 DNA 结合。KEGG 通路分析结果显示 CDC5L 主要参与 RNA 代谢、Pre-mRNA 内含子的处理、mRNA 及 Pre-mRNA 的剪接等信号通路。见表 2。

2.3 单基因 PPI 分析

11 个蛋白节点组合成的蛋白互作网络,发现 10 个蛋白与 CDC5L 存在相互作用,分别为 PRPF19、BCAS2、PLRG1、EFTUD2、CWC15、SNRNP200、SF3B2、CTNNB1、SF3A1、SF3B1。见图 2。

2.4 转录因子预测

转录因子预测发现,转录因子 ZBTB33、CREM、ZNF384、BHLHE22、OLIG3、Twist2 可与基因 CDC5L 结合。由于转录因子往往上调基因的表达,因此经 GEPIA 网站预测发现 ZBTB33、ZNF384 及 Twist2 与 CDC5L 的表达呈正相关。见图 3。

3 讨论

骨肉瘤和尤文肉瘤是儿童和青少年最常见的恶性骨肿瘤,其发生与基因异常表达具有显著相关性,但具体的机制仍不清楚。近年,随着高通量技术的成熟,利用生物信息学方法对恶性肿瘤进行全基因表达谱分析,了解恶性肿瘤相关的基因及其相关的分子调控机制,为恶性肿瘤的治疗提供了新的思路^[9]。

本研究利用 GEO 数据库,发现共同参与骨肉瘤与尤文肉瘤的基因为 CDC5L,且 CDC5L 在肿瘤组织

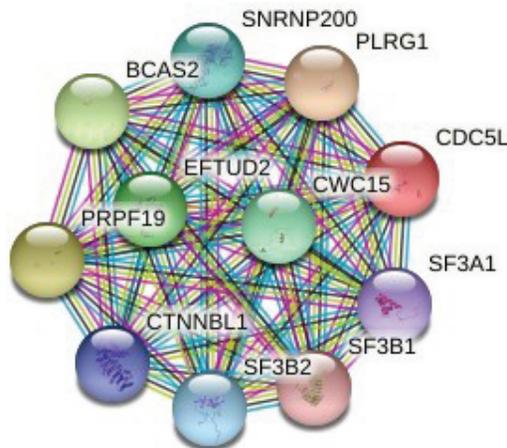


图 2 蛋白交互作用网络

Fig.2 Protein-protein interaction network constructed with the differentially expressed genes

中高表达。并对该基因进行了 GO 和 KEGG 通路分析,发现与有丝分裂及 RNA 代谢有关。并构建 CDC5L 的 PPI 网络,及预测了与 CDC5L 结合的转录因子。

3.1 CDC5L 介导肿瘤发生的分子机制

CDC5L 位于 6 号染色体,是细胞周期调节因子介导 G2/M 期转变。既往研究表明 CDC5L 在肿瘤中高表达,参与多种肿瘤的发生及发展。Zhang 等^[10]发现 CDC5L 在膀胱癌中高表达,CDC5L 可能通过调节 PI3K/AKT 信号通路促进肿瘤细胞的生长和增殖,且 CDC5L 表达与病理分级显著相关。在动物实验中,抑制 CDC5L 表达可降低人端粒酶逆转录酶表达

表 2 GO 与 KEGG 信号通路

Tab.2 Significantly enriched GO terms and KEGG pathways of differentially expressed genes

类别	条目	描述
BP term	GO:0000278	Mitotic cell cycle
BP term	GO:0000398	mRNA splicing, via spliceosome
BP term	GO:0006281	DNA repair
BP term	GO:0006397	mRNA processing
CC term	GO:0000974	Prp19 complex
CC term	GO:0005634	Nucleus
CC term	GO:0005654	Nucleoplasm
CC term	GO:0005662	DNA replication factor A complex
MF term	GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding
MF term	GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific
MF term	GO:0001222	Transcription corepressor binding
MF term	GO:0003677	DNA binding
KEGG term	R-HSA-8953854	Metabolism of RNA
KEGG term	R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA
KEGG term	R-HSA-72172	mRNA Splicing
KEGG term	R-HSA-72163	pre-mRNA splicing

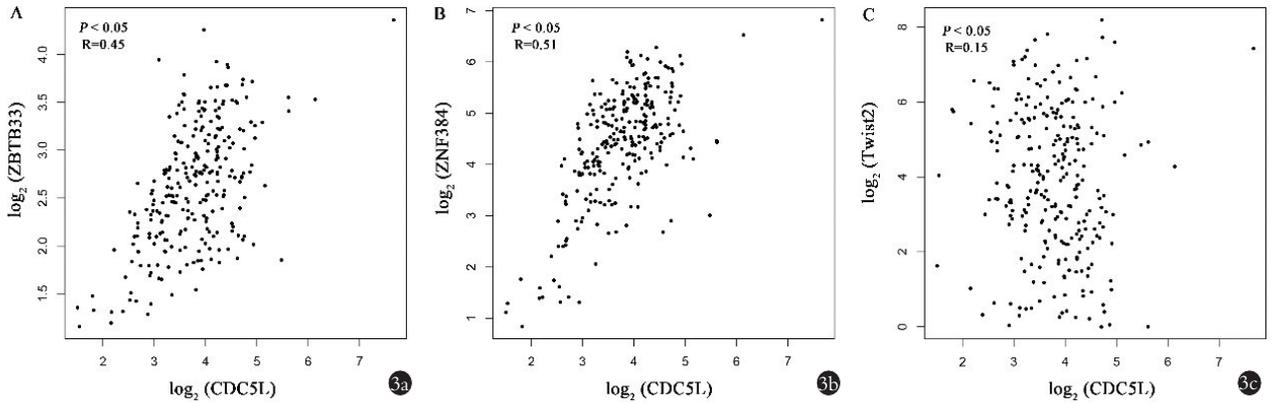


图 3 转录因子与 CDC5L 表达的相关性分析 3a. ZBTB33 与 CDC5L 表达的相关性 3b. ZNF384 与 CDC5L 表达的相关性 3c. Twist2 与 CDC5L 表达的相关性

Fig.3 Correlation analysis between the expression of CDC5L and its transcription factors 3a. The correlation between the expression of ZBTB33 and CDC5L 3b. The correlation between the expression of ZNF384 and CDC5L 3c. The correlation between the expression of Twist2 and CDC5L

进而抑制肿瘤的生长^[11]。Qiu 等^[12]发现下调 CDC5L 可以将肝癌细胞周期阻滞在 G2/M 期,抑制其增殖, CDC5L 在肿瘤细胞中高表达,还是肝癌患者不良预后的独立危险因素。既往在骨肉瘤中的研究结果显示, CDC5L 是原癌基因,在骨肉瘤中高表达^[13-14],与笔者研究结果相似。有研究发现在骨肉瘤中 6p12-p21 扩增子通过靶向调节 CDC5L 癌基因影响患者的疾病进展预后^[13]。此外,基于新辅助化疗的患者中^[15],血清外泌体中高表达的 CDC5L 与骨肉瘤的不良的化疗疗效有关^[16-17],但是 CDC5L 基因在尤文肉瘤中的作用机制尚无报道,仍需要进一步的研究。

CDC5L 主要参与 RNA 代谢、Pre-mRNA 内含子的处理、mRNA, 及 Pre-mRNA 的剪接等信号通路^[18-19]。CDC5L 是有丝分裂的关键分子,降低其表达可使有丝分裂受到抑制、纺锤体错位及纺锤体检查点持续激活^[14]。近年来,抗有丝分裂药物通过使细胞周期阻滞在 G2/M 期,逐渐成为一种有效的肿瘤治疗方法^[20-21]。因此,靶向抑制 OS 及 EWS 患者肿瘤组织中 CDC5L 的表达,有望成为该病的潜在治疗靶点,但仍需要大量的分子生物学实验进一步了解调节有丝分裂细胞死亡的分子途径,使新治疗方法的临床转化成为可能。

3.2 CDC5L 互作蛋白与肿瘤发生的分子机制

PPI 网络预测发现 PRPF19、BCAS2 (SPF27)、CWC15 (AD002)、PLRG1 (PRL1)、EFTUD2、SNRNP 200、SF3B2、CTNBL1、SF3A1、SF3B1 与 CDC5L 发生蛋白互作。PRPF19、BCAS2、CWC15、PLRG1、CTNBL1 是剪接体 PRP19-CDC5L 复合物的组成部分,PRPF19 可通过抑制 CDC5L 的表达并促进溶酶体介导的 CDC5L 降解,诱导细胞周期停滞;EFTUD2、SF3A1、SNRNP200 是 pre-mRNA 剪接所必需

的;SF3B1 和 SF3B2 形成 SF3B 复合体作为核剪切因子,抑制其表达可以降低尤文肉瘤细胞中 EWS-FLI1 基因的表达(与尤文肉瘤患者生存高度相关的基因)^[22]。因此,该预测 OS 及 EWS 催化活性剪接体核心潜在分子互作,为后续分子生物学实验提供探索方向。

此外,转录因子预测发现 ZBTB33、ZNF384 及 Twist2 与 CDC5L 结合,上调 CDC5L 的表达。其中, Twist2 为螺旋状的转录因子,有研究显示 Twist2 通过下调 E-cadherin 表达增强肿瘤细胞的侵袭性,与包括骨肉瘤在内的多种肿瘤患者的不良预后相关^[23]。

3.3 研究局限与不足

上述转录因子预测为后续分子交互作用提供线索。但本研究尚有局限性,首先,与其他肿瘤相比, OS 及 EWS 发病率低,是发生在儿童及青少年的恶性肿瘤,因此生物信息学资料有限,缺乏临床信息,无法进行诊断及预后研究。其次本研究是基于生物信息学的研究,仍需要在体内体外的分子生物学实验中进一步的验证,例如 q-RT PCR, Western Blot、集落形成分析和流式细胞术分析等分组生物学实验,以明确 CDC5L 在 OS 及 EWS 中行使功能的具体信号传导途径。

参考文献

[1] Schuetze SM. Chemotherapy in the management of osteosarcoma and Ewing's sarcoma [J]. J Natl Compr Canc Netw, 2007, 5(4): 449-455.
 [2] 李洪涛,李董董,杨晓笛,等. 尤文氏肉瘤误诊为非骨化性纤维瘤 1 例[J]. 中国骨伤, 2017, 30(1): 68-70.
 LI HT, LI DD, YANG XD, YANG XD, et al. Ewing's sarcoma misdiagnosed as nonossifying fibroma: a case report [J]. Zhongguo Gu Shang/China J Orthop Trauma, 2017, 30(1): 68-70. Chinese.
 [3] Edgar R, Domrachev M, Lash AE, et al. Gene expression omnibus;

- NCBI gene expression and hybridization array data repository [J]. *Nucleic Acids Res*, 2002, 30(1):207-210.
- [4] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update [J]. *Nucleic Acids Res*, 2013, 41(Database issue):D991-995.
- [5] Specht K, Sung YS, Zhang L, et al. Distinct transcriptional signature and immunoprofile of CIC-DUX4 fusion-positive round cell tumors compared to EWSR1-rearranged Ewing sarcomas; further evidence toward distinct pathologic entities [J]. *Genes Chromosomes Cancer*, 2014, 53(7):622-633.
- [6] Agelopoulos K, Richter GHS, Schmidt E, et al. Deep sequencing in conjunction with expression and functional analyses reveals activation of FGFR1 in ewing sarcoma [J]. *Clin Cancer Res*, 2015, 21(21):4935-4946.
- [7] 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.
- [8] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible [J]. *Nucleic Acids Res*, 2017, 45(D1):D362-D368.
- [9] Chen Z, Wu H, Wang GH, et al. Identification of potential candidate genes for hypertensive nephropathy based on gene expression profile [J]. *BMC Nephrol*, 2016, 17(1):149.
- [10] Zhang Z, Mao WP, Wang LS, et al. Depletion of CDC5L inhibits bladder cancer tumorigenesis [J]. *J Cancer*, 2020, 11(2):353-363.
- [11] Li J, Zhang NN, Zhang R, et al. CDC5L promotes hTERT expression and colorectal tumor growth [J]. *Cell Physiol Biochem*, 2017, 41(6):2475-2488.
- [12] Qiu H, Zhang XB, Ni WK, et al. Expression and clinical role of Cdc5L as a novel cell cycle protein in hepatocellular carcinoma [J]. *Dig Dis Sci*, 2016, 61(3):795-805.
- [13] Lu XY, Lu YJ, Zhao YJ, et al. Cell cycle regulator gene CDC5L, a potential target for 6p12-p21 amplicon in osteosarcoma [J]. *Mol Cancer Res*, 2008, 6(6):937-946.
- [14] Mu R, Wang YB, Wu M, et al. Depletion of pre-mRNA splicing factor CDC5L inhibits mitotic progression and triggers mitotic catastrophe [J]. *Cell Death Dis*, 2014, 5(3):e1151.
- [15] 张波, 庞清江, 章海均, 等. 43 例骨肉瘤患者预后的多因素分析 [J]. *中国骨伤*, 2011, 24(12):982-986.
ZHANG B, PANG QJ, ZHANG HJ, et al. Multivariate analysis for prognostic factors among 43 patients with osteosarcoma [J]. *Zhongguo Gu Shang/China J Orthop Trauma*, 2011, 24(12):982-986. Chinese.
- [16] Martin JW, Chilton-MacNeill S, Koti M, et al. Digital expression profiling identifies RUNX2, CDC5L, MDM2, RECQL4, and CDK4 as potential predictive biomarkers for neo-adjuvant chemotherapy response in paediatric osteosarcoma [J]. *PLoS One*, 2014, 9(5):e95843.
- [17] Xu JF, Wang YP, Zhang SJ, et al. Exosomes containing differential expression of microRNA and mRNA in osteosarcoma that can predict response to chemotherapy [J]. *Oncotarget*, 2017, 8(44):75968-75978.
- [18] Ajuh P, Kuster B, Panov K, et al. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry [J]. *EMBO J*, 2000, 19(23):6569-6581.
- [19] Ajuh P, Kuster B, Panov K, et al. Identification of peptide inhibitors of pre-mRNA splicing derived from the essential interaction domains of CDC5L and PLRG1 [J]. *Nucleic Acids Res*, 2003, 31(21):6104-6116.
- [20] Asghar U, Witkiewicz AK, Turner NC, et al. The history and future of targeting cyclin-dependent kinases in cancer therapy [J]. *Nat Rev Drug Discov*, 2015, 14(2):130-146.
- [21] Schwartz GK, Shah MA. Targeting the cell cycle: a new approach to cancer therapy [J]. *J Clin Oncol*, 2005, 23(36):9408-9421.
- [22] Grohar PJ, Kim S, Rangel RGO, et al. Functional genomic screening reveals splicing of the EWS-FLI1 fusion transcript as a vulnerability in ewing sarcoma [J]. *Cell Rep*, 2016, 14(3):598-610.
- [23] Wushou A, Hou J, Zhao YJ, et al. Twist-1 up-regulation in carcinoma correlates to poor survival [J]. *Int J Mol Sci*, 2014, 15(12):21621-21630.

(收稿日期:2021-01-21 本文编辑:王玉蔓)