

· 基础研究 ·

骨肉瘤潜在关键基因及相关通路的鉴定[△]

张懿明, 毛良浩, 江攀, 倪晨烈, 李建, 尹正宇, 仲新宇, 张兵, 李大鹏*

(江苏大学附属医院脊柱外科, 江苏镇江 212000)

摘要: [目的] 基于生物信息学分析骨肉瘤差异表达谱中的关键基因及相关通路。[方法] 本研究利用 GEO (gene expression omnibus, GEO) 数据库中的两个芯片数据集 (GSE11414, GSE14359) 筛选人骨肉瘤样本与人成骨细胞之间的差异表达基因, 并通过基因本体论 (Gene Ontology, GO) 和京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 富集分析挖掘其潜在的生物学功能。通过 STRING (Search Tool for the Retrieval of Interacting Genes, STRING) 数据库和 Cytoscape 软件构建蛋白互作网络, 并进一步筛选关键基因。最后利用 LOGpc (Long-term Outcome and Gene Expression Profiling database of pan-cancers, LOGpc) 数据库评估关键基因在骨肉瘤中的预后价值。[结果] GSE11414 和 GSE14359 中共有 111 个共表达差异表达基因, 包括 28 个上调和 83 个下调基因, 功能富集分析结果显示其主要富集于细胞外基质组织、整合素结合、糖胺聚糖结合、包含胶原的细胞外基质、p53 信号通路及 TGF- β 信号通路。通过蛋白互作网络分析筛选得到 10 个关键基因。无复发生存分析证实, THBS1 和 IGFBP3 的高表达与骨肉瘤患者的不良预后有关。[结论] 本研究通过生物信息学分析构建骨肉瘤差异基因的蛋白互作网络, 挖掘与骨肉瘤发病机制密切相关的枢纽基因, 可能为骨肉瘤提供新的预后标志物和治疗靶点。

关键词: 骨肉瘤, 生物信息学, 靶向治疗, 关键基因, 生物标志物

中图分类号: R738.1 **文献标志码:** A **文章编号:** 1005-8478 (2022) 07-0644-05

Identification of potential core genes and related pathways in osteosarcoma // ZHANG yi-ming, MAO Liang-hao, JIANG Pan, NI Chen-lie, LI Jian, YIN Zheng-yu, ZHONG Xin-yu, ZHANG Bing, LI Da-peng. Affiliated Hospital of Jiangsu University, Zhenjiang 212000, China

Abstract: [Objective] To analyze the core genes and functional pathways in the differential expression profile of osteosarcoma based on bioinformatics analysis. [Methods] In this study, two microarray datasets, including GSE11414 and GSE14359, from the gene expression omnibus (GEO) database were used to screen differentially expressed genes (DEGs) between human osteosarcoma and osteoblasts samples. The potential biological functions of DEGs were explored through the Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analyses. The Search Tool for the Retrieval of Interacting Genes (STRING) database and Cytoscape software were used to further construct a protein-protein interaction network and screen core genes. Finally, the prognostic value of core genes in osteosarcoma was assessed using the Long-term Outcome and Gene Expression Profiling database of pan-cancers (LOGpc) database. [Results] There were 111 DEGs in GSE11414 and GSE14359, including 28 up-regulated and 83 down-regulated DEGs. Functional enrichment analysis showed that DEGs were mainly enriched in the extracellular matrix organization, integrin binding, glycosaminoglycan binding, collagen-containing extracellular matrix, p53 signaling pathway, and TGF- β signaling pathway. Ten core genes were obtained by protein-protein interaction network analysis. Recurrence-free survival analysis confirmed that high expression of THBS1 and IGFBP3 was associated with poor prognosis in patients with osteosarcoma. [Conclusion] In this study, the protein-protein interaction network of differential genes in osteosarcoma was constructed by bioinformatics analysis. The core genes closely related to the pathogenesis of osteosarcoma were explored, which may provide new prognostic markers and therapeutic targets for osteosarcoma.

Key words: osteosarcoma, bioinformatics analysis, targeted therapy, core gene, biomarker

骨肉瘤 (osteosarcoma, OS) 是一种常见的原发性恶性骨肿瘤, 起源于间充质组织, 在儿童和青少

DOI:10.3977/j.issn.1005-8478.2022.07.14

[△]基金项目:江苏省青年医学重点人才项目(编号:QNRC2016844);江苏省高层次卫生人才“六个一工程”拔尖人才项目(编号:LGY2019089);江苏省研究生实践创新计划项目(编号:SJCX19_0580)

作者简介:张懿明,硕士研究生,研究方向:骨科,(电话)13567650316,(电子信箱)zhangyiming0607@163.com

*通信作者:李大鹏,(电话)13921582827,(电子信箱)lidapeng706@163.com

年中发病率较高, 具有高侵袭性、易复发和远处转移的特点^[1]。通过新辅助化疗、手术、化疗及生物治疗的综合疗法, 未转移骨肉瘤患者的总生存率由不足20%逐渐提高到70%; 然而, 转移或复发患者的总体预后仍然较差, 总生存率仅为20%左右^[2]。分子生物学的发展为骨肉瘤的诊断和治疗提供了新见解, 研究发现 TP53、RB1、NOTCH1、REQL4、INK4a 的突变及 TGF- β 、c-Myc 和 IGF 通路的失调与骨肉瘤的发生发展密切相关^[3, 4], 然而具体分子机制仍有待进一步挖掘。高通量测序技术和生物信息学分析相结合, 有助于更好地了解骨肉瘤潜在的分子机制, 寻找治疗靶点, 从而延长骨肉瘤患者的生存时间^[5, 6]。有研究通过综合生物信息学分析发现 RPUSD1 基因高表达与骨肉瘤患者的不良预后相关, 并可能通过细胞代谢等相关通路发挥作用, 为进一步的体内外实验提供依据^[7]。本研究分别分析了 GSE11414 和 GSE14359 芯片表达数据集, 筛选骨肉瘤和成骨细胞之间的差异基因。通过 GO 和 KEGG 通路富集分析进一步探索两个数据集中共表达差异基因的潜在生物学功能。通过 STRING 数据库构建蛋白相互作用网络, 并使用 Cytoscape 挖掘关键基因, 最后使用 LOGpc 数据库评估关键基因的预后价值, 以期为骨肉瘤防治和研究提供新的思路。

1 材料和方法

1.1 芯片数据

两个芯片表达数据集 (GSE11414 和 GSE14359) 从美国国家生物技术信息中心 (the National Center for Biotechnology Information, NCBI) 的基因表达公共数据库 (Gene Expression Omnibus, GEO, <https://www.ncbi.nlm.nih.gov/geo/>) 获得。GSE14359 数据集由 1 个人原代成骨细胞样本、5 个普通型骨肉瘤冰冻样本以及 4 个骨肉瘤肺转移肿瘤样本组成, 每个样本均有 2 个复本, 基于 GPL96 分析平台 ([HG-U133A] Affymetrix Human Genome U133A Array)。作者选择了其中的 10 个常规骨肉瘤样本和 2 个非肿瘤原代成骨细胞样本进行进一步分析。GSE11414 数据集包含 1 个正常人成骨细胞, 1 个 U2OS 骨肉瘤细胞系, 1 个 MG63 骨肉瘤细胞系的测序结果, 每个样本有 2 个复本, 数据集基于 GPL6244 芯片分析平台 ([HuGene-1_0-st] Affymetrix human Gene 1.0 ST Array [transcript (Gene) version])。GSE11414 中的所有样本均被纳入本研究进行分析。

1.2 差异表达基因筛选

利用 GEO2R 在线分析工具 (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) 分别对两个芯片表达数据集 (GSE11414 和 GSE14359) 中骨肉瘤样本和正常成骨细胞样本之间的差异表达基因进行鉴定。差异基因的筛选标准为 $|\logFC| > 1$ 且 $\text{adjust } P \text{ 值} < 0.05$ 。 $\logFC > 1$ 的差异基因被认为是上调基因, $\logFC < -1$ 的差异基因被认为是下调基因。利用韦恩图在线网页工具 (bioinformatics.psb.ugent.be/webtools/Venn/) 获取两个芯片表达数据集之间的共表达差异表达基因。

1.3 GO 和 KEGG 通路富集分析

为了进一步揭示共表达差异基因的重要功能和生物学途径, 作者对其进行了 GO 功能和 KEGG 通路富集分析。GO 是一种常用的基因注释和功能基因组学分析方法, 包括生物过程 (biological process, BP)、分子功能 (molecular function, MF) 和细胞成分 (cellular component, CC)^[8]。KEGG (<http://www.genome.jp/>) 是一个广泛使用的数据库, 存储了大量关于基因组信息与生物途径、化学物质、疾病和药物的数据^[9]。利用 R 软件中的 “clusterProfiler”、“org.Hs.eg.db”、“ggplot2” 程序包对差异表达基因进行 GO 和 KEGG 通路富集分析及可视化, $P < 0.05$ 为差异具有统计学意义。

1.4 蛋白相互作用网络构建

基于筛选的共表达差异基因, 作者利用 STRING 数据库 (STRING, version 11.0, <http://string-db.org/>) 构建蛋白质相互作用 (protein-protein interaction, PPI) 网络, 并由 Cytoscape 软件可视化, 蛋白互作得分 > 0.15 为筛选标准 (满分为 1)。此外差异表达基因的 \logFC 值也被导入 Cytoscape 中以显示基因上/下调节状态。利用 Cytoscape 中的 CytoHubba 插件筛选关键基因进行下一步分析。

1.5 关键基因预后分析

LOGpc 数据库 (<http://bioinfo.henu.edu.cn>) 可提供查询分析多种癌症的预后生存数据, 其数据主要来源 GEO 和 TCGA (The Cancer Genome Atlas, TCGA) 等数据库^[10]。将由 CytoHubba 筛选得到的关键基因输入 LOGpc 数据库中进行生存分析, 选择骨肉瘤 GSE39058 数据集探究关键基因表达水平与骨肉瘤患者无复发生存期 (recurrence-free survival, RFS) 的关系, 所有参数设为默认值。

2 结果

2.1 差异表达基因筛选

以 $|\log_2FC| > 1$ 且 $\text{adjust } P < 0.05$ 为筛选标准，从 GSE11414 数据集中共得到 735 个差异基因，其中 262 个上调基因，473 个下调基因，从 GSE14359 中共筛选得到 1376 个差异基因，其中 791 个上调基因，585 个下调基因；随后，通过 Venn 分析，在两个数据集之间筛选得到 111 个共表达差异基因，其中 28 个基因显著上调（图 1a），83 个基因显著下调（图 1b）。

2.2 GO 和 KEGG 通路富集分析

为了探索共表达差异基因潜在的生物学功能，本研究使用 R 软件进行了 GO 注释和 KEGG 通路富集

分析， $P < 0.05$ 为差异有统计学意义。GO 富集分析结果显示 111 个共表达差异基因参与的生物学进程有细胞外结构组织、细胞外基质组织、细胞基质粘附等；分子功能则主要富集于整合素结合、糖胺聚糖结合、细胞外基质结构组成；细胞成分方面，差异基因主要存在于包含胶原的细胞外基质、溶酶体管腔、空泡腔（图 1c）。KEGG 通路富集分析结果显示这些差异基因主要富集于 p53 信号通路、转化生长因子 β （transforming growth factor- β , TGF- β ）信号通路及肿瘤坏死因子（tumor necrosis factor, TNF）信号通路等（图 1d）。

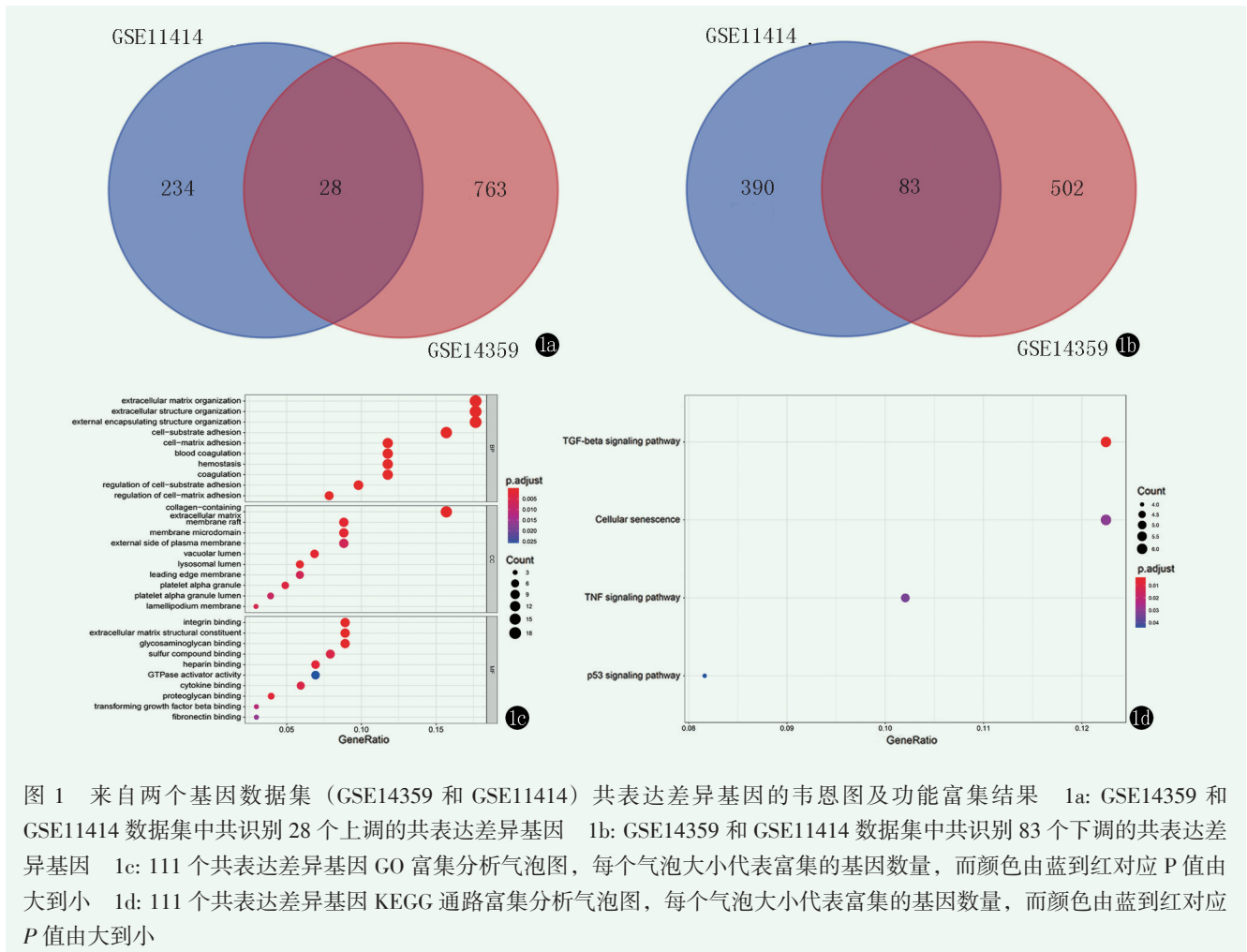


图 1 来自两个基因数据集（GSE14359 和 GSE11414）共表达差异基因的韦恩图及功能富集结果 1a: GSE14359 和 GSE11414 数据集中共识别 28 个上调的共表达差异基因 1b: GSE14359 和 GSE11414 数据集中共识别 83 个下调的共表达差异基因 1c: 111 个共表达差异基因 GO 富集分析气泡图，每个气泡大小代表富集的基因数量，而颜色由蓝到红对应 P 值由大到小 1d: 111 个共表达差异基因 KEGG 通路富集分析气泡图，每个气泡大小代表富集的基因数量，而颜色由蓝到红对应 P 值由大到小

2.3 蛋白相互作用网络构建

蛋白互作得分 > 0.15 为筛选标准，通过 STRING 数据库预测了共表达差异基因之间的蛋白互作网络，并使用 Cytoscape 软件进行后续分析。整个 PPI 网络由 104 个节点和 778 条边组成，红色的节点表示上调的基因，绿色的节点表示下调的基因（图 2a）。应用 CytoHubba 插件根据 MCC 法筛选 PPI 网络中排名前 10 位的关键基因：纤连蛋白 1（fibronectin 1, FN1）、血小板凝血酶蛋白 1（thrombospondin-1, THBS1）、白细胞介

素-6（interleukin 6, IL-6）、纤溶酶原激活物抑制因子 1（serine protease inhibitor, clade E member 1, SERPINE1）、窖蛋白 1（caveolin-1, CAV1）、基质金属蛋白酶 14（matrix metalloproteinase 14, MMP14）、弹性蛋白（elastin, ELN）、类胰岛素生长因子结合蛋白 3（Insulin-like growth factor-binding protein 3, IGFBP3）、血管细胞黏附分子-1（vascular cell adhesion molecule 1, VCAM1）、细胞周期素依赖激酶抑制因子 1A（cyclin-dependent kinase inhibitor 1A, CDKN1A）（图 2b）。

2.4 关键基因预后分析

通过 LOG_{pc} 数据库对筛选得到的关键基因进行生存分析, 验证其在骨肉瘤中的预后价值。结果表

明 THBS1 ($P < 0.001$, $HR = 10.930$) 和 IGFBP3 ($P = 0.043$, $HR = 3.234$) 高表达与骨肉瘤患者的 RFS 负相关 (图 2c, 2d)。

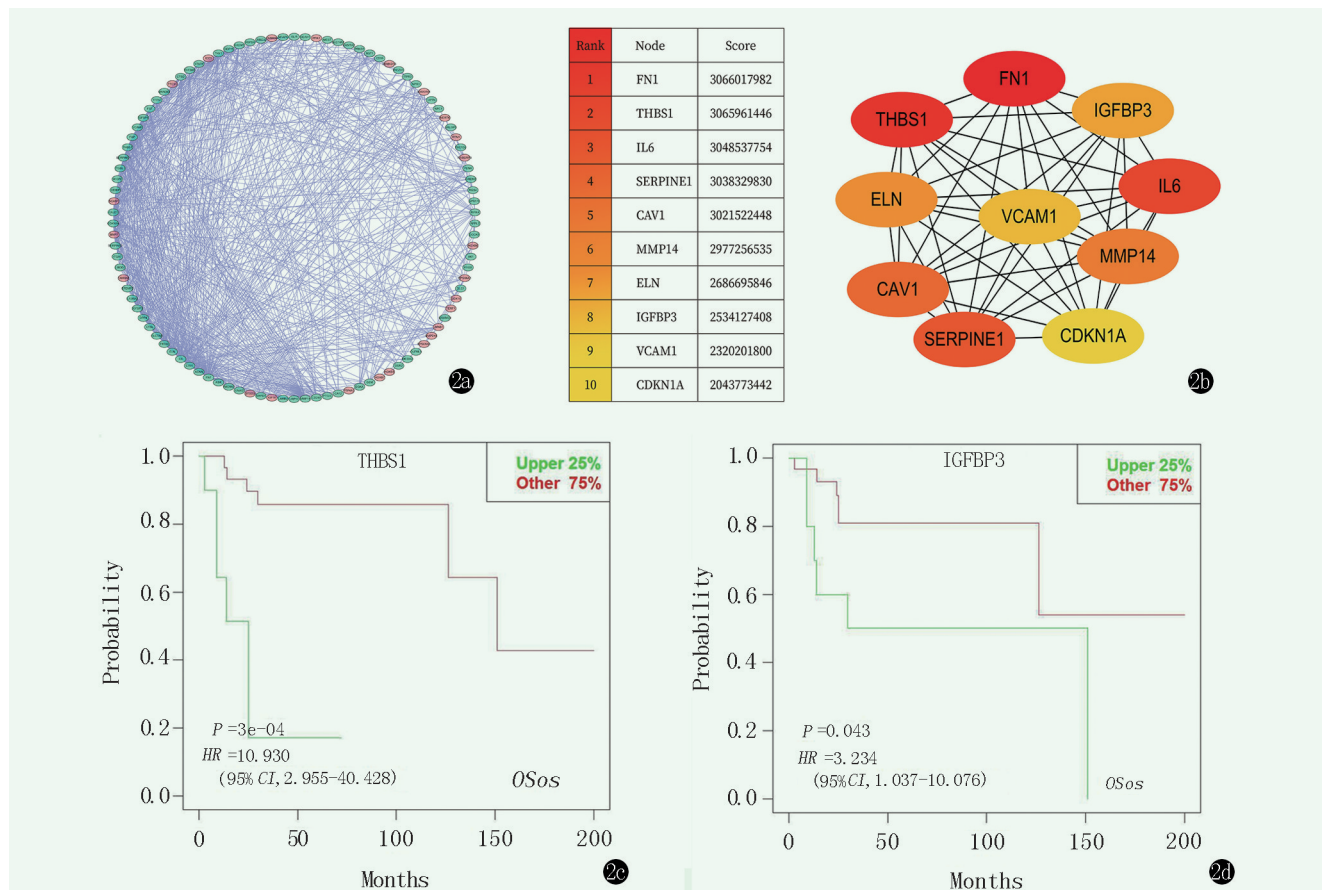


图 2 蛋白相互作用网络的构建和关键基因的生存分析 2a: PPI 网络包含 111 个差异基因, 每个节点代表一个蛋白质, 每条线代表蛋白质之间的相互作用。红色节点代表上调基因, 绿色节点代表下调基因 2b: 通过 CytoHubba 插件筛选得到的 10 个关键基因及其相互作用关系, 每条线代表蛋白质之间的相互作用 2c: 生存分析显示 THBS1 高表达与骨肉瘤患者较短的无复发生存期有关 2d: 生存分析显示 IGFBP3 高表达与骨肉瘤患者较短的无复发生存期有关

3 讨论

骨肉瘤是儿童和青少年最常见的骨肿瘤, 大多数发生在长骨的干骺端, 其特征是骨样沉积改变和高转移率^[11]。目前, 骨肉瘤患者的标准治疗策略是新辅助化疗 (术前)、手术切除 (截肢或最常见的保肢手术) 和辅助化疗 (术后)^[12]。然而, 在现有的治疗方案下, 转移性骨肉瘤患者的 10 年生存率仍低于 20%, 迫切需要新的治疗策略^[13]。近年来, 生物信息学的发展为骨肉瘤的早期诊断和治疗提供了新见解^[14, 15]。本研究通过分析 GSE11414 和 GSE14359 芯片数据, 对骨肉瘤和成骨细胞之间的差异基因进行功能富集分析, PPI 网络分析以及预后分析, 最终筛选了 10 个关键基因。

GO 和 KEGG 通路富集分析结果显示差异基因主

要富集于细胞外基质组织、整合素结合、包含胶原的细胞外基质、p53 及 TGF- β 信号通路。研究表明, 细胞外基质不仅为细胞提供物理支持、调节组织发育和内环境的稳定, 还参与肿瘤细胞的增殖、侵袭、迁移、免疫逃逸^[16]。而整合素是连接细胞外基质最重要的粘附分子受体, 参与细胞粘附及信号传导, 可与受体酪氨酸激酶协同作用通过磷脂酰肌醇-3-激酶/蛋白激酶 B (phosphatidylinositol 3-kinase/protein kinase B, PI3K/AKT) 通路促进癌细胞的有丝分裂和增殖^[17]。p53 是研究最充分的肿瘤抑制因子之一, 参与调控细胞周期进展、凋亡、铁死亡、衰老、DNA 修复等多种细胞过程^[18]。而 TGF- β 在肿瘤中具有双重作用, 既可通过诱导细胞周期阻滞和凋亡来抑制肿瘤的发生; 又能刺激上皮间质转化, 作用于肿瘤微环境间接促进肿瘤的发生^[19, 20]。由此可见差异基因可能通过 p53 和 TGF- β 信号通路参与骨肉瘤的发生发

展,可以通过进一步的实验探究具体机制。

在对差异基因的PPI网络分析中,筛选得到10个关键基因,其中THBS1和IGFBP3高表达与骨肉瘤患者的RFS负相关。THBS1是血小板凝血酶蛋白家族成员之一,可通过THBS1/CD36/vasculostatin信号通路促进血管抑素的表达,抑制骨肉瘤血管生成及肿瘤生长^[21]。然而Hu等^[22]发现THBS1增加了转移相关基因的表达并通过激活FAK信号通路促进骨肉瘤的迁移、侵袭和肺转移。本研究发现THBS1高表达与骨肉瘤患者的不良预后有关。由此作者推测THBS1在骨肉瘤中可能发挥抑癌和促癌的双重作用,需要进一步实验阐明THBS1的生物学功能。而IGFBP3主要受p53调控,是循环中类胰岛素生长因子的主要载体,其在转录和翻译后水平的破坏与多种癌症的病理生理学有关^[23]。Long等^[24]研究表明circ_0000285可以通过miR-409-3p/IGFBP3轴促进骨肉瘤的增殖,侵袭并抑制其凋亡。此外IGFBP-3还可通过PI3K/AKT和活化蛋白-1信号通路促进骨肉瘤细胞迁移,同时上调VCAM1表达水平^[25]。以上研究结果提示IGFBP3可能与骨肉瘤的不良预后有关,与本研究的结果一致。

综上所述,本研究通过综合生物信息学分析为阐明骨肉瘤发生的分子机制提供了一系列关键基因和通路。但尚需进一步的体内外实验来证实这些关键基因在人骨肉瘤组织中的表达、具体功能及机制,以期为骨肉瘤靶向药物的研发和联合治疗提供新思路。

参考文献

- [1] Brown HK, Tellez-Gabriel M, Heymann D. Cancer stem cells in osteosarcoma [J]. *Cancer Lett*, 2017, 386 (2) : 189-195.
- [2] Harrison DJ, Geller DS, Gill JD, et al. Current and future therapeutic approaches for osteosarcoma [J]. *Expert Rev Anticancer Ther*, 2018, 18 (1) : 39-50.
- [3] Yu X, Yustein JT, Xu J. Research models and mesenchymal/epithelial plasticity of osteosarcoma [J]. *Cell Biosci*, 2021, 11 (1) : 94.
- [4] Czarnecka AM, Synoradzki K, Firlej W, et al. Molecular biology of osteosarcoma [J]. *Cancers (Basel)*, 2020, 12 (8) : 2130.
- [5] Ni S, Li J, Qiu S, et al. KIF21B expression in osteosarcoma and its regulatory effect on osteosarcoma cell proliferation and apoptosis through the PI3K/AKT pathway [J]. *Front Oncol*, 2020, 10 : 606765.
- [6] Wu ZL, Deng YJ, Zhang GZ, et al. Development of a novel immune-related genes prognostic signature for osteosarcoma [J]. *Sci Rep*, 2020, 10 (1) : 18402.
- [7] 李纯璞,张立清,侯存强,等. RPUSD1基因在骨肉瘤中的表达及其预后价值分析 [J]. *中国矫形外科杂志*, 2021, 29 (2) : 149-153.
- [8] Denny P, Feuermann M, Hill DP, et al. Exploring autophagy with gene ontology [J]. *Autophagy*, 2018, 14 (3) : 419-436.
- [9] Kanehisa M, Sato Y, Furumichi M, et al. New approach for understanding genome variations in KEGG [J]. *Nucleic Acids Res*, 2019, 47 (D1) : D590-d595.
- [10] Wang Q, Xie L, Dang Y, et al. OSIMS: a web server to evaluate the prognostic value of genes in leiomyosarcoma [J]. *Front Oncol*, 2019, 9 (1) : 190.
- [11] Cascini C, Chiodoni C. The immune landscape of osteosarcoma: implications for prognosis and treatment response [J]. *Cells*, 2021, 10 (7) : 1668.
- [12] Bielack SS, Hecker-Nolting S, Blattmann C, et al. Advances in the management of osteosarcoma [J]. *F1000Res*, 2016, 5 : 2767.
- [13] Liu Y, Huang N, Liao S, et al. Current research progress in targeted anti-angiogenesis therapy for osteosarcoma [J]. *Cell Prolif*, 2021, 54 (9) : e13102.
- [14] Xiao B, Liu L, Li A, et al. Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma [J]. *Front Oncol*, 2020, 10 : 607622.
- [15] Xie B, Li Y, Zhao R, et al. Identification of key genes and miRNAs in osteosarcoma patients with chemoresistance by bioinformatics analysis [J]. *Biomed Res Int*, 2018, 2018 : 4761064.
- [16] Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer [J]. *EMBO Rep*, 2014, 15 (12) : 1243-1253.
- [17] Cooper J, Giancotti FG. Integrin signaling in cancer: mechanotransduction, stemness, epithelial plasticity, and therapeutic resistance [J]. *Cancer Cell*, 2019, 35 (3) : 347-367.
- [18] Huang J. Current developments of targeting the p53 signaling pathway for cancer treatment [J]. *Pharmacol Ther*, 2021, 220 : 107720.
- [19] Colak S, Ten Dijke P. Targeting TGF- β signaling in cancer [J]. *Trends Cancer*, 2017, 3 (1) : 56-71.
- [20] Yu Y, Feng XH. TGF- β signaling in cell fate control and cancer [J]. *Curr Opin Cell Biol*, 2019, 61 (1) : 56-63.
- [21] Jian YK, Zhu HY, Wu XL, et al. Thrombospondin 1 triggers osteosarcoma cell metastasis and tumor angiogenesis [J]. *Oncol Res*, 2019, 27 (2) : 211-218.
- [22] Hu C, Wen J, Gong L, et al. Thrombospondin-1 promotes cell migration, invasion and lung metastasis of osteosarcoma through FAK dependent pathway [J]. *Oncotarget*, 2017, 8 (44) : 75881-75892.
- [23] Cai Q, Dozmorov M, Oh Y. IGFBP-3/IGFBP-3 receptor system as an anti-tumor and anti-metastatic signaling in cancer [J]. *Cells*, 2020, 9 (5) : 1261.
- [24] Long Z, Gong F, Li Y, et al. Circ_0000285 regulates proliferation, migration, invasion and apoptosis of osteosarcoma by miR-409-3p/IGFBP3 axis [J]. *Cancer Cell Int*, 2020, 20 (1) : 481.
- [25] Chao CC, Lee WF, Yang WH, et al. IGFBP-3 stimulates human osteosarcoma cell migration by upregulating VCAM-1 expression [J]. *Life Sci*, 2021, 265 : 118758.

(收稿:2021-06-22 修回:2021-12-01)

(同行评议专家:李大森)

(本文编辑:宁桦)