

# 基于数据挖掘对前列腺癌预后具有预测价值的差异基因筛选

郝晓明, 王玉杰

(新疆医科大学第一附属医院泌尿外科, 新疆 乌鲁木齐 830054)

**摘要:**目的 通过数据挖掘筛选对前列腺癌总生存期(OS)及无疾病生存期(DFS)有预测价值的差异基因。方法 通过挖掘 GEPIA2 数据库中与前列腺癌 OS、DFS 有关的基因以及肿瘤差异基因信息,探讨上调基因和下调基因与 OS、DFS 的关系,最后筛选出对前列腺癌 OS 及 DFS 有预测价值的差异基因。结果 共筛选出在前列腺癌差异表达的基因 3005 个,其中上调基因 681 个、下调基因 2324 个。与 OS 有关的基因 497 个,从中筛选出差异表达的基因 63 个,其中上调基因 6 个,下调基因 57 个,UBE2SP2、RP11-627G23.1、RPL7P23、TK1 的上调和前列腺癌较低的 OS 有关,而 RNA5-8S5 和 ELOVL2 高表达的患者前列腺癌 OS 高于低表达的患者。与 DFS 有关的基因 499 个,从中筛选出差异基因 184 个,其中上调基因 9 个,C2、FUCA1、C9orf152 表达上调的前列腺癌患者 DFS 较高,而 TK1、CDC20、BIRC5、TOP2A、UBE2SP2、FAM223A 高表达与较低的 DFS 有关,下调基因 175 个。UBE2SP2 和 TK1 存在正相关,且高表达与较差的 OS 和 DFS 均有关。结论 本次共筛选出 15 个与前列腺癌预后有关的上调基因,可为对前列腺癌 OS 和 DFS 有预测价值的基因靶点选择提供一定参考。

**关键词:**前列腺癌;预后;总生存期;差异基因;筛选

中图分类号:R737.25

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2024.01.003

文章编号:1006-1959(2024)01-0016-06

## Screening of Differentially Expressed Genes with Predictive Value for Prognosis of Prostate Cancer Based on Data Mining

HAO Xiao-ming, WANG Yu-jie

(Department of Urinary Surgery, the First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, Xinjiang, China)

**Abstract:** **Objective** To screen the differentially expressed genes with predictive value for overall survival (OS) and disease-free survival (DFS) of prostate cancer by data mining. **Methods** By mining the information of genes related to OS and DFS of prostate cancer and tumor differentially expressed genes in the GEPIA2 database, the relationship between up-regulated genes and down-regulated genes and OS and DFS was explored. Finally, differentially expressed genes with predictive value for OS and DFS of prostate cancer were screened. **Results** A total of 3005 differentially expressed genes in prostate cancer were screened, including 681 up-regulated genes and 2324 down-regulated genes. There were 497 genes related to OS, and 63 differentially expressed genes were screened out, including 6 up-regulated genes and 57 down-regulated genes. The up-regulation of UBE2SP2, RP11-627G23.1, RPL7P23 and TK1 was related to the lower OS of prostate cancer, while the OS of prostate cancer in patients with high expression of RNA5-8S5 and ELOVL2 was higher than that in patients with low expression. There were 499 genes related to DFS, and 184 differentially expressed genes were screened out, including 9 up-regulated genes. The DFS of prostate cancer patients with up-regulated expression of C2, FUCA1 and C9orf152 was higher, while the high expression of TK1, CDC20, BIRC5, TOP2A, UBE2SP2 and FAM223A was related to lower DFS, and 175 genes were down-regulated. There was a positive correlation between UBE2SP2 and TK1, and high expression was associated with poor OS and DFS. **Conclusion** A total of 15 up-regulated genes related to the prognosis of prostate cancer are screened, which can provide some reference for the selection of gene targets with predictive value for OS and DFS of prostate cancer.

**Key words:** Prostate cancer; Prognosis; Overall survival; Differentially expressed genes; Screening

前列腺癌(prostate cancer)是泌尿生殖系统较常见的恶性肿瘤,因其早期不易诊断、异质性大、中晚期进展为激素不敏感等特点致使患者预后较差<sup>[1]</sup>。前列腺癌好发于老年男性,发病率随着年龄增

大而增高。总生存期(overall survival, OS)和无疾病生存期(disease free survival, DFS)是评估肿瘤预后的两个关键指标。诸多因素与肿瘤的预后相关,血清炎症性指标、循环肿瘤细胞、中性粒与淋巴细胞比值、mi-RNA 等均可用于前列腺癌预后的评估<sup>[2-4]</sup>。前列腺癌的发生发展有一定的遗传基础,目前已经发现了罕见但高风险的突变(如 BRCA2、HOXB13)和低风险但常见的 77 个等位基因<sup>[5]</sup>。从表观遗传学的角度来看,如果能在基因层面对前列腺癌的预后进行有效评估,精准的基因干预可能会改善前列腺癌的

作者简介:郝晓明(1986.8-),男,山西太原人,博士研究生,主治医师,主要从事泌尿系肿瘤的研究

通讯作者:王玉杰(1962.9-),男,新疆乌鲁木齐人,博士,主任医师,教授,博士生导师,主要从事泌尿系肿瘤的诊治研究

预后。肿瘤组织中表达差异的基因不一定和预后有关,而与预后有关的基因却不一定存在差异性表达,这就使得表达无差异的基因较难对肿瘤预后做出有效且精准的预测。新开发的基于比较时间序列转录组分析的 TimeMeter,能够通过建立复杂的时间和基因表达关联进而识别在时间上更先进的差异进展基因<sup>[6]</sup>。如能筛选出与预后相关的差异基因,则可能在基因层面对前列腺癌的预后进行精准预测。目前有关差异基因和前列腺癌预后的研究较少。为此,本研究通过数据挖掘筛选对前列腺癌 OS 及 DFS 有预测价值的差异基因,以期对前列腺癌预后具有预测价值的基因靶点的选择提供参考。

## 1 资料与方法

1.1 资料来源 利用 GEPIA2 数据库(该数据库整合了 TCGA 和 GTEx)检索关于前列腺癌的资料。利用“Survival Analysis”子栏“Most Differential Survival Genes”子栏“Datasets Selection”选择 PRAD(Prostate adenocarcinoma)。分别下载前列腺癌 OS 有关基因和 DFS 有关基因的数据。在“Differential Expression Analysis”子栏“Dataset”中选择 PRAD,分别下载前列腺癌 Over-expressed Genes 和 Under-expressed Genes 的数据。

1.2 数据组合分析 对下载好的数据进行整理分析。整理后分别得到前列腺癌 OS 有关基因、DFS 有关基因、上调基因、下调基因 4 组数据。然后对 4 组数据进行组合,将上调基因和 OS 有关基因、上调基因和 DFS 有关基因、下调基因和 OS 有关基因、下调基因和 DFS 有关基因、上调基因和 OS、DFS 有关基

因、下调基因和 OS、DFS 有关基因组合,组合后提取每组共同基因信息进行分析。利用 STRING、ImageGP 在线工具作图。

## 2 结果

2.1 差异基因 共筛选出差异基因 3005 个,其中上调基因 681 个、下调基因 2324 个。前列腺癌差异基因热图见图 1。

2.2 和 OS 有关的差异基因 与 OS 有关的基因 497 个,从中筛选出差异基因 63 个,其中上调基因 6 个,UBE2SP2、RP11-627G23.1、RPL7P23、TK1 这 4 个基因的上调和前列腺癌较低的 OS 有关,提示其可能是促癌基因,而 RNA5-8S5 和 ELOVL2 高表达前列腺癌的 OS 高于低表达,提示 RNA5-8S5 和 ELOVL2 对前列腺癌有抑癌作用。另外,63 个差异基因中下调基因 57 个。

2.3 和 DFS 有关的差异基因 与 DFS 有关的基因 499 个,从中筛选出差异基因 184 个,其中上调基因 9 个:C2、FUCA1、C9orf152 表达上调的前列腺癌患者 DFS 较高,提示这 3 个基因抑制前列腺癌复发进展,而 TK1、CDC20、BIRC5、TOP2A、UBE2SP2、FAM223A 高表达与较低的 DFS 有关,提示其在前列腺癌的复发进展中起促进作用。下调基因 175 个。前列腺癌差异基因、OS 有关基因、DFS 有关基因的韦恩图见图 2。对 OS、DFS 有预测价值的上调基因见图 3。TK1、CDC20、BIRC5、TOP2A、UBE2SP2、FAM223AC2、FUCA1、C9orf152、RP11-627G23.1、RPL7P23、RNA5-8S5 和 ELOVL2 的蛋白互作网络见图 4。

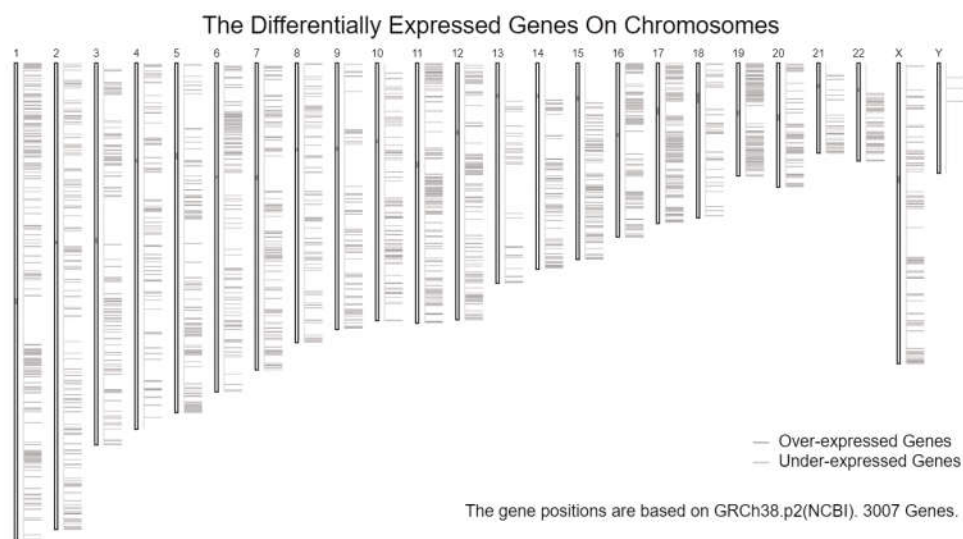


图 1 前列腺癌差异基因表达热图

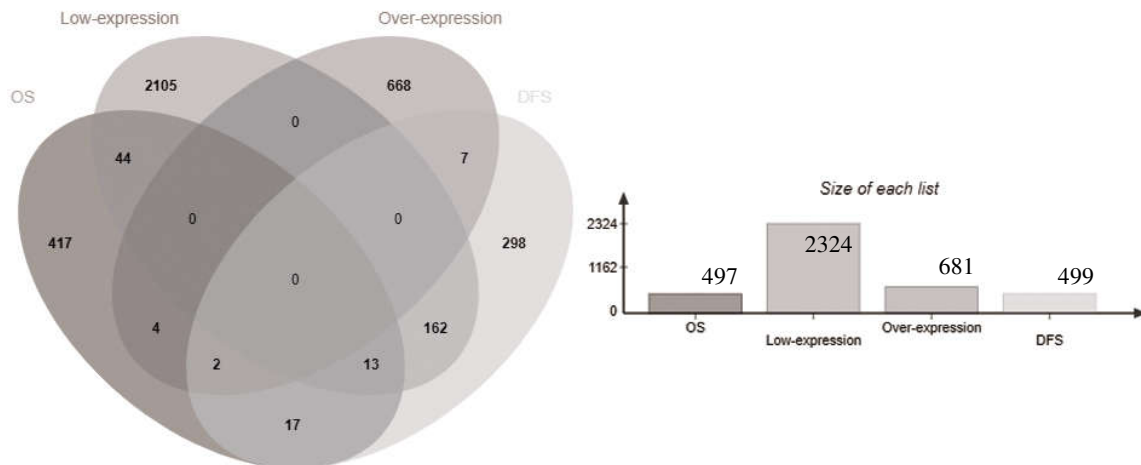


图 2 前列腺癌上调基因、下调基因、OS 有关基因、DFS 有关基因的韦恩图



图 3 15 个对 OS、DFS 有预测价值的上调基因表达情况

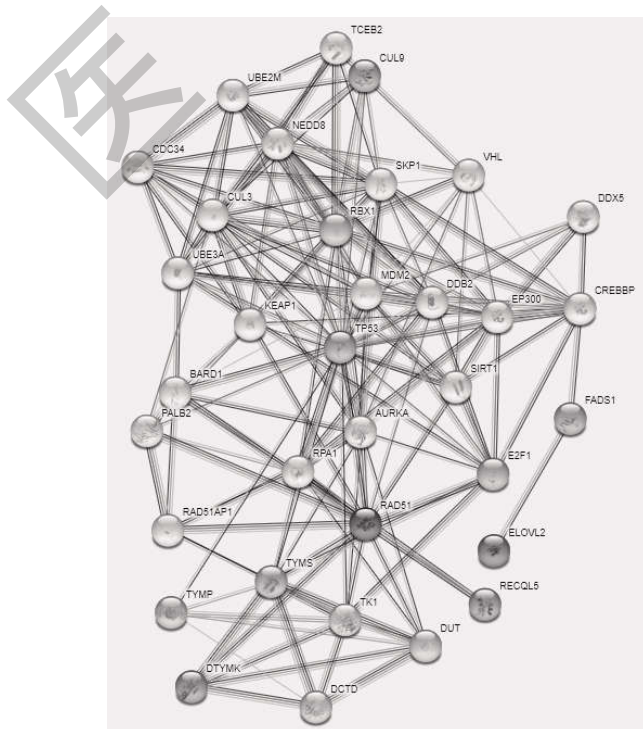
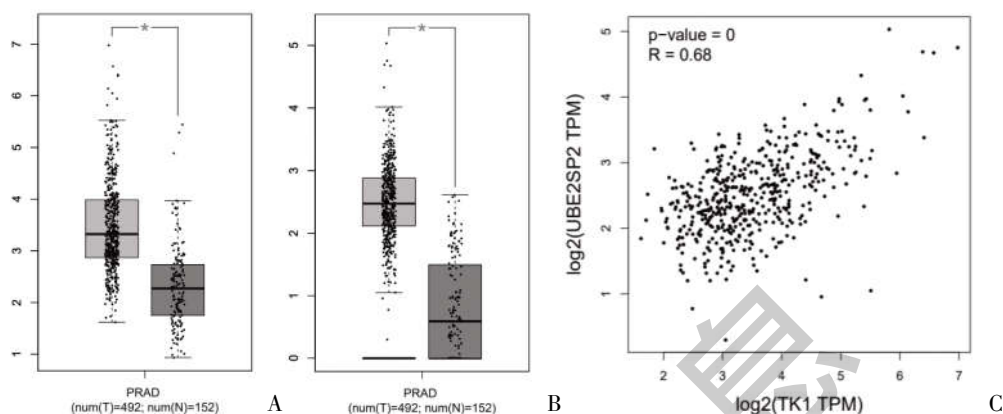


图 4 有预测价值的上调差异基因蛋白互作网络图

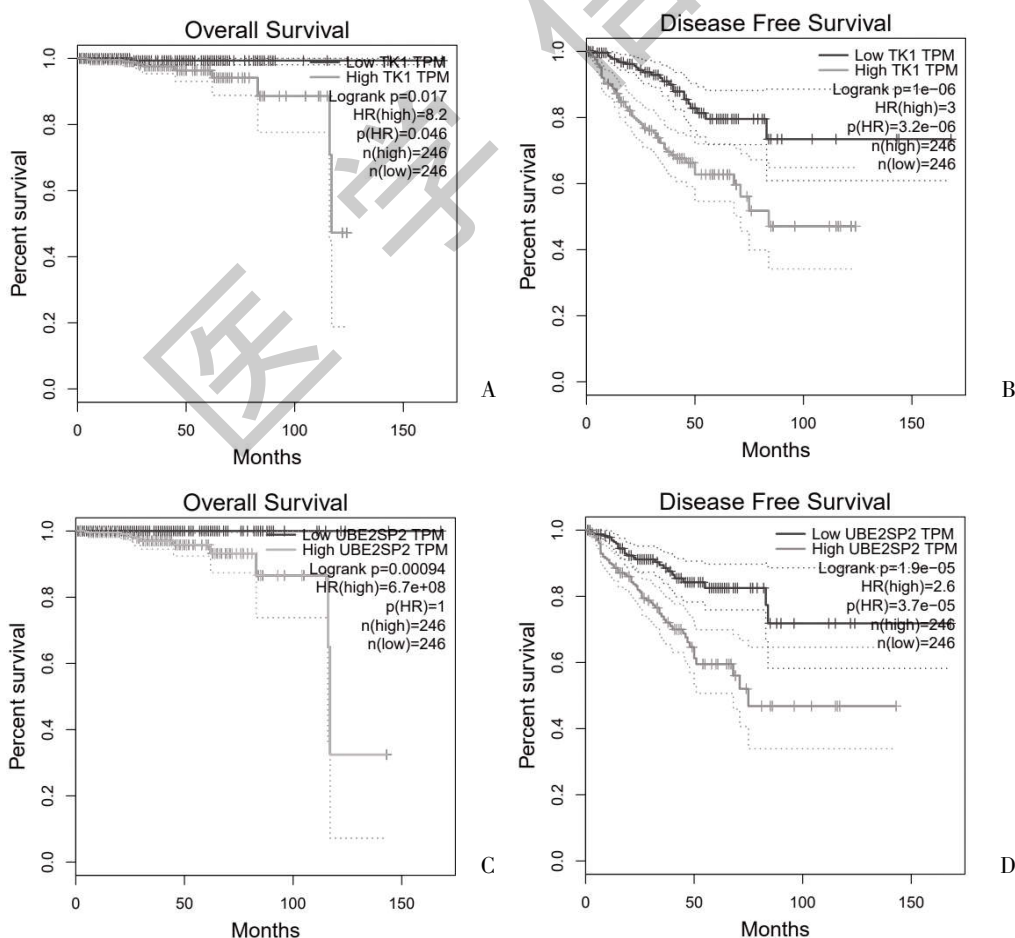
2.4 和 OS、DFS 均有关的差异基因 与 OS、DFS 均有关的基因 32 个,从中筛选出差异基因 15 个,其中上调基因 2 个,下调基因 13 个;与 OS、DFS 均有关且表达上调的 2 个基因:UBE2SP2 和 TK1。TK1 和 UBE2SP2 在前列腺癌中均高表达,且基因相关性分析显示:UBE2SP2 和 TK1 呈正相关 ( $r=0.68, P<$

0.05), 见图 5。TK1 和 UBE2SP2 与前列腺癌 OS、DFS 的生存曲线见图 6。可见高表达 TK1 的前列腺癌患者 OS 和 DFS 均低于低表达的前列腺患者, UBE2SP2 高表达的前列腺癌患者 OS 和 DFS 均低于低表达的前列腺癌患者。



注:A:TK1 在前列腺癌中差异表达;B:UBE2SP2 在前列腺癌中差异表达;C:TK1 和 UBE2SP2 基因相关性分析

图 5 和 OS、DFS 均有关的差异基因表达情况



注:A:TK1 与前列腺癌 OS 生存曲线;B:TK1 与前列腺癌 DFS 生存曲线;C:UBE2SP2 与前列腺癌 OS 生存曲线;D:UBE2SP2 与前列腺癌 DFS 生存曲线

图 6 TK1 和 UBE2SP2 与前列腺癌 OS、DFS 的生存曲线

### 3 讨论

肿瘤组织中上调或下调的差异基因在肿瘤的发生、发展中可能发挥促癌或抑癌的作用进而影响肿瘤的预后。筛选和肿瘤预后有关的差异基因能在基因层面为肿瘤的防治提供一定依据<sup>[7]</sup>。为探索前列腺癌中对 OS 及 DFS 有预测价值的差异基因。本研究筛选出差异基因 3005 个,与 OS 有关的基因 497 个,进一步筛选出和 OS 有关的差异基因 63 个,其中上调基因 6 个:UBE2SP2、RP11-627G23.1、RPL7P23、TK1、RNA5-8S5、ELOVL2。UBE2SP2、RP11-627G23.1、RPL7P23、TK1 的上调和前列腺癌较低的 OS 有关,提示这 4 个基因可能和前列腺癌较差的总体生存有关。TK1 作为 dTTP 合成的关键酶和细胞周期 S 期依赖酶,在肿瘤中高表达且与分期、淋巴转移等相关,能促进肿瘤细胞的增殖和侵袭,是影响前列腺癌预后的独立危险因素<sup>[8,9]</sup>。相似基因是通过计算出的具有同源性和蛋白质结构相似性的基因簇,其生物学功能可能相近。BIRC5、E2F1、MELK 是 TK1 的相似基因,在不同的肿瘤中高表达与不良预后有关<sup>[10,11]</sup>。这也从侧面印证了高表达 TK1 和肿瘤不良预后有关。UBE2SP2 属于假基因来源的 RNA,是一种特殊的由基因组中的假基因转录而来的非编码 RNA,在肝癌中显著高表达提示患者预后不良<sup>[12]</sup>。本研究显示,高表达 TK1 和 UBE2SP2 的前列腺癌患者 OS 更差,提示这 2 个基因可能对肿瘤发挥促癌作用,与以上报道相符。虽然关于 RP11-627G23.1、RPL7P23 在肿瘤中的研究鲜有,但本研究提示其表达上调和前列腺癌总体生存不良有关,据此推测其可能是潜在的预后不良相关基因,这值得后续进一步去探索。另外 2 个高表达基因 RNA5-8S5、ELOVL2 和前列腺癌较好的 OS 有关。ELOVL2 作为哺乳动物微粒体 ELOVL 脂肪酸酶家族的一员参与了哺乳动物各种细胞功能所需的极长链脂肪酸的延伸,不仅能抑制肿瘤活性而且还能调控 AKT、ERα 信号通路中参与耐药的基因 THEM4 从而降低乳腺癌的耐药性<sup>[13]</sup>。Hu T 等<sup>[14]</sup>发现 ELOVL2 在前列腺癌中不仅高表达而且高表达者有良好的预后,过表达 ELOVL2 可以通过上调 INPP4B 抑制 PI3K/Akt 信号通路发挥抑癌作用。而目前有关 RNA5-8S5 的报道极少,还需要更多研究去验证 RNA5-8S5 与前列腺癌预后的关系。诸多研究均表明 ELOVL2 发挥抑癌作用,本研究显示高表达的

ELOVL2 可能在前列腺癌中起抑癌作用而且与良好的整体生存有关。

本研究发现,与前列腺癌 DFS 有关的基因 499 个,从中筛选出差异基因 184 个,其中上调基因 9 个,C2、FUCA1、C9orf152 表达上调的前列腺癌患者 DFS 较高,而 TK1、CDC20、BIRC5、TOP2A、UBE2SP2、FAM223A 高表达与较低的 DFS 有关。FUCA1 是糖基水解酶 29 蛋白家族中的一员,它能通过调节糖基化发挥抑癌功能:其编码的 AFU1 够清除和乳腺癌、甲状腺癌、结直肠癌等肿瘤有关的糖蛋白末端 α-1-聚焦残基;此外,AFU1 低表达与前列腺癌病理分级高、分期高、术后残留瘤多、初治效果差、DFS 短相关<sup>[15,16]</sup>。Wu X 等<sup>[17]</sup>提出包含 C9orf152 在内的 9-mRNA 特征能更好的预测前列腺癌无生化复发生存,平均 AUC(0.81)高于 PSA、GS 或 AJCC T 分期的 AUC(0.52~0.73)。这与本研究结果一致,提示 FUCA1、C9orf152 在前列腺癌中发挥抑癌作用,二者高表达意味着更好的 DFS。作为细胞周期检查点的调节剂 CDC20 在膀胱癌和前列腺癌中高表达,而其高表达与复发和不良预后密切相关<sup>[18,19]</sup>。TOP2A 是一种必需的核酶,其高表达与肿瘤不良预后相关,上调 miR-142-5p 靶向调控 TOP2A 低表达能抑制前列腺癌细胞的增殖与侵袭能力<sup>[20]</sup>。Fu M 等<sup>[21]</sup>等建立免疫相关模型预测前列腺癌风险,BIRC5 作为预测模型中的免疫相关基因在 87.5%的前列腺癌组织中上调而且该基因的敲低可以抑制癌细胞增殖和迁移与较好的 DFS 有关。另一项研究显示非小细胞肺癌中表达上调的 BIRC5 基因通过影响免疫细胞浸润水平诱导免疫微环境失衡进而造成不良的临床预后<sup>[22]</sup>。这些研究提示 CDC20、TOP2A、BIRC5 是前列腺癌的促癌因素且其高表达与患者较差的 DFS 有关。但 UBE2SP2、FAM223A 目前在前列腺癌中的研究极少,其与前列腺癌的预后关系还需进一步明确。

本研究存在一定局限性:数据来自单一数据库,不同研究之间的异质性无法确定,国外基因信息和国人之间可能存在种族差异,下调差异基因数目较大暂未进行深入分析,可靠性有待更多高质量研究来进一步证实。希望将来能纳入更多有关国人前列腺癌差异基因和预后的数据以求更好的探索二者之间的关系。

综上所述,本研究进一步筛选出对前列腺癌 OS

可能有预测价值的 3 个差异基因:TK1、UBE2SP2、ELOVL2,TK1 和 UBE2SP2 在前列腺癌中高表达与不良 OS 有关,而 ELOVL2 的高表达与较好的 OS 有关。RP11-627G23.1、RPL7P23、RNA5-8S5 3 个差异基因在预测前列腺癌 OS 方面可能也有一定价值,因缺乏相关研究仍需进一步探索。本研究还筛选出预测前列腺癌 DFS 可能有价值的 7 个差异基因:FUCA1、C9orf152C2、UBE2SP2、TK1、CDC20、BIRC5、TOP2A,其中 FUCA1 和 C9orf152C2 在前列腺癌中发挥抑癌作用,二者高表达意味着更好的 DFS;TK1、CDC20、BIRC5、UBE2SP2、TOP2A 是前列腺癌的促癌基因,高表达和不良的 DFS 有关;C2、FAM223A 和前列腺癌的 DFS 可能有关,需进一步研究证实。最后,本研究筛选出预测前列腺癌 OS、DFS 可能均有价值的 2 个差异基因:UBE2SP2 和 TK1。TK1 和 UBE2SP2 可能均是前列腺癌的促癌基因,且两者存在正相关关联,某个基因的上调通过某种机制促进另一个基因的表达,两者协同增强促癌作用,其潜在机制值得进一步探索。

#### 参考文献:

- [1]Ye DW,Zhu Y.Prostate cancer and prostatic diseases Best of China,2018[J].Prostate Cancer Prostatic Dis,2019,22(1):1-2.
- [2]郑燕军,汪云兴,徐秀娟.血清炎症标志物与前列腺癌预后的关系研究[J].中国卫生检验杂志,2022,32(17):2115-2118,2122.
- [3]谢张兴,高亮,王晨,等.循环肿瘤细胞在前列腺癌诊断及预后中的临床应用[J].现代泌尿外科杂志,2022,27(7):609-613.
- [4]张泉波,李玉莲,陈绮,等.基于 TCGA 数据库构建前列腺癌相关 miRNA 预后模型 [J]. 浙江医学,2021,43 (18):1971-1978,2049-2050.
- [5]Attard G,Parker C,Eeles RA,et al.Prostate cancer [J].Lancet,2016,387(10013):70-82.
- [6]Jiang P,Chamberlain CS,Vanderby R,et al.TimeMeter assesses temporal gene expression similarity and identifies differentially progressing genes[J].Nucleic Acids Res,2020,48(9):e51.
- [7]汪逊,陆雪强,侯传胜,等.前列腺癌差异基因标志物筛选有助于预后评价[J].现代泌尿外科杂志,2022,27(6):512-518.
- [8]Grasso CS,Wu YM,Robinson DR,et al.The mutational landscape of lethal castration-resistant prostate cancer [J].Nature,2012,487(7406):239-243.
- [9]McCartney A,Malorni L.Potential through simplicity: thymidine kinase-1 as a biomarker for CDK4/6 inhibitors [J].British Journal of Cancer,2020,123(2):176-177.
- [10]李雪菲,张丁丁,董丹丹,等.BIRC5 高表达在胃癌中增强细胞活力、抑制凋亡并与预后不良相关[J].中国病理生理杂志,2020,36(11):2013-2019.
- [11]张九国,袁智勇.母胚亮氨酸拉链激酶(MELK)在肺腺癌中的表达及其与预后的相关性研究[J].国际生物医学工程杂志,2021,44(3):202-206.
- [12]楼咸洋.假基因来源的 RNA 在肝癌发生和进展中的作用及其机制[D].杭州:浙江大学,2020.
- [13]Jeong D,Ham J,Kim HW,et al.ELOVL2: a novel tumor suppressor attenuating tamoxifen resistance in breast cancer [J].Am J Cancer Res,2021,11(6):2568-2589.
- [14]Hu T,Zhang H,Du Y,et al.ELOVL2 restrains cell proliferation, migration, and invasion of prostate cancer via regulation of the tumor suppressor INPP4B[J].Cell Signal,2022,96:110373.
- [15]Herrera H,Dilday T,Uber A,et al.Core-Fucosylated Tetra-Antennary N-Glycan Containing A Single N-Acetylglucosamine Branch Is Associated with Poor Survival Outcome in Breast Cancer[J].Int J Mol Sci,2019,20(10):2528.
- [16]Zhang C,Liu J,Chao F,et al.Alpha-L-Fucosidase Has Diagnostic Value in Prostate Cancer With "Gray-Zone PSA" and Inhibits Cancer Progression via Regulating Glycosylation [J].Front Oncol,2021,11:742354.
- [17]Wu X,Lv D,Eftekhari M,et al.A new risk stratification system of prostate cancer to identify high-risk biochemical recurrence patients[J].Transl Androl Urol,2020,9(6):2572-2586.
- [18]石峰,孙文功,刘增振,等.CDC20 基因的表达对前列腺癌患者临床病理特征及预后的影响[J].解放军医学杂志,2021,46(11):1092-1097.
- [19]葛新国,原小斌.细胞分裂周期蛋白 20、微小 RNA-34a-5p 在膀胱尿路上皮癌组织中的表达及与临床病理的相关性 [J].临床外科杂志,2021,29(11):1075-1078.
- [20]朱延杰,尤平洪,唐良友,等.miR-142-5p 靶向调控 TOP2A 基因对前列腺癌细胞增殖和侵袭的影响[J].山东医药,2020,60(14):13-16.
- [21]Fu M,Wang Q,Wang H,et al.Immune-Related Genes Are Prognostic Markers for Prostate Cancer Recurrence [J].Front Genet,2021,12:639642.
- [22]郑浩然,蒋爱民,史楚楚,等.BIRC5 基因在非小细胞肺癌中的表达及与免疫细胞浸润和预后的关系 [J]. 现代肿瘤医学,2022,30(4):634-640.

收稿日期:2023-03-24;修回日期:2023-04-06

编辑/成森