

单细胞与孟德尔随机化鉴定前列腺癌关键基因

韩浩冉¹, 任瑞民¹

(山西医科大学第三医院, 山西白求恩医院(山西医学科学院), 太原, 030032, 同济山西医院, 泌尿外科, 太原, 030032)

摘要: 目的 利用单细胞转录组中差异表达基因结合全基因组关联研究(GWAS)数据的基因 eQTL 关联分析, 以获得因果相关的前列腺癌 (PCa) 关键基因。方法 本研究利用 GEO 数据库中的 GSE193337 数据集, 分析了 8 个前列腺组织样本的单细胞测序数据。通过 R 语言进行细胞聚类及注释并筛选各个细胞亚型标记基因。结合前列腺癌患者的 GWAS 和基因 eQTL 数据, 进行孟德尔随机化 (MR) 分析, 采用敏感性分析和异质性检验选择显著的基因结果。最后, 构建蛋白互作网络, 并进行 GO 和 KEGG 通路富集分析, 筛选相关的功能通路。结果 本研究纳入 8 例前列腺样本 (癌/正常各 4 例), 注释为 8 种细胞类型 (如 T 细胞、上皮细胞等)。MR 分析鉴定上皮细胞差异基因 IGFBP-3 (OR=0.901, P=0.017) 和 TGM3 (OR=1.042, P=0.009) 与 PCa 因果关联, 敏感性分析证实结果稳健。蛋白互作网络结合富集分析揭示关键基因参与前列腺癌、MAPK 等通路, 涉及生长调控及代谢过程。结论 本研究通过结合 scRNA-seq 和 MR 分析, 确定了 IGFBP-3 和 TGM3 是 PCa 潜在的生物标志物。

关键词: 前列腺癌; 单细胞转录组测序; 孟德尔随机化分析; 表达数量性状位点; 标记基因

中图分类号: R737.25 **文献标志码:** A

项目: 中央引导地方科技发展资金项目 (YDZJSX20231A066); 山西省研究生实践创新项目 (2024SJ170)

DOI: doi.org/10.70693/cjmsr.v1i2.1161

前列腺癌 (prostate cancer, PCa) 是男性中常见的恶性肿瘤之一, 在全球范围内发病率较高, 且随着我国人均寿命的延长, 其发病率呈上升趋势^[1-2]。对于局部晚期或转移性 PCa 患者, 雄激素剥夺治疗是主要的全身治疗手段, 但大多数患者在 2-3 年后会进展为去势抵抗性前列腺癌, 因此早期识别和精准治疗至关重要^[3]。

近年来, 单细胞 RNA 测序 (scRNA-seq) 因其在肿瘤异质性分析中的优势, 广泛应用于癌症研究^[4]。孟德尔随机化 (Mendelian Randomization, MR) 是一种基于工具变量的因果推断方法, 其利用与暴露因素密切相关的单核苷酸多态性 (Single-Nucleotide Polymorphism, SNP) 作为工具变量 (Instrumental Variable, IV), 通过整合全基因组关联研究 (Genome-Wide Association Study, GWAS) 的汇总数据, 可有效探索暴露因素与结局之间的潜在因果关系^[5]。表达数量性状位点 (expression quantitative trait locus, eQTL) 已被发现与 PCa 发生发展相关, 但基于 scRNA-seq 结合 GWAS 的基因 eQTL 分析仍较少^[6]。

前列腺癌的研究正在快速发展, 了解其流行病学特点和基因靶点机制对于提高诊断、治疗和预后具有重要意义。可根据患者的基因突变谱定制个性化的治疗方案, 如靶向特定基因突变的药物。目前前列腺癌发生发展中的相关基因还在不断探索中, 为前列腺癌的精准诊断和治疗策略开发提供新的潜在靶点。

本研究整合 PCa 的 GWAS、基因 eQTL 数据和 scRNA-seq 结果, 通过 MR 分析挖掘新的潜在诊疗靶点。

1. 资料与方法

1.1 数据采集和预处理

作者简介: 韩浩冉 (1999-), 男, 山西医科大学硕士研究生, 主要研究方向为泌尿系统肿瘤。

通讯作者: 任瑞民, 男, 副主任医师, 硕士生导师, 研究方向为泌尿系统肿瘤。

本研究的数据来自可公开访问的 GEO 数据库(www.ncbi.nlm.nih.gov/geo/)中的 GSE193337 数据集^[7], 该数据集包括来自奥地利的 8 个前列腺组织(4 个正常, 4 个癌组织), 癌组织的分期分别为 pT2c、pT2a、pT3a、pT3b。使用 Seurat 包 Read10X 函数导入并预处理 10x Genomics 格式数据。

1.2 质量控制和降维

在质量控制阶段, 计算每个细胞的基因表达特征数量(nFeature_RNA)、线粒体基因表达比例(percent.mt)等指标。使用 Seurat 软件包读取 scRNA-seq 样本的基因表达谱, 并筛选异常表达的样本(nFeature_RNA > 50 and percent.mt < 15)。随后, 将得到的信息均质化归一化, 并进行主成分分析(Principal Component Analysis, PCA)。

1.3 聚类分析和单细胞类型注释

使用带有默认参数的 FindNeighbors 函数进行聚类分析, 指定 dims = 1:20, 从而使用前 20 个主成分的数据进行邻域计算。然后将 FindClusters 函数用于实际的集群, 并根据初步分析调整分辨率参数, 以优化集群粒度。我们利用 SingleR 包或细胞类型注释, 利用 SingleR 函数和从 ref_Human_all 获得的参考数据集。通过匹配我们数据集的表达谱和参考文献中的表达谱来进行注释, 为每个细胞提供预测的细胞类型。通过在 FindAllMarkers 函数中分别设置 logfc 和最小 pct 的阈值为 1.5 和 0.25, 得到各细胞亚型对应的标记基因。

1.4 孟德尔随机化数据准备

eQTL 数据来自 IEU OpenGWAS 项目(<https://gwas.mrcieu.ac.uk/>), 通过 Ensembl 网站(<https://asia.ensembl.org/>), 将 Ensembl ID 转化为 gene name。 P 值小于 5×10^{-8} 作为筛选循环蛋白强相关的单核苷酸多态性(sim-gle nucleotide polymorphism, SNP)阈值。连锁不平衡条件: $R^2 < 0.001$, kb=10000。为排除弱 IV 带来的偏倚, $F > 10$ 作为评估 IV 强度的阈值。前列腺癌患者的 GWAS 数据来自前列腺癌协会研究基因组中癌症相关变异体小组(Prostate Cancer Association Group to investigation Cancer-associated variants in the Genome, PRACTICAL)联盟, 纳入了来自欧洲 7,9148 例 PCa 病例和 61106 例对照, SNPs 20346368 个(GWAS ID=ieu-b-85)。

1.5 孟德尔随机化分析

基于 scRNA-seq 获得的不同细胞类型间的差异基因的基础上, 进一步结合 PCa 患者的 GWAS 数据和基因 eQTL 数据, 使用 ieu-b-85 为结局, 单细胞测序得出各细胞亚型的差异基因的 eQTL 数据为暴露, 进行 MR 分析。MR 分析使用 R 程序包 TwoSampleMR, 采用逆方差加权(inverse variance-weighted, IVW)法作为评估因果关系的主要分析方法, 同时使用 MR-Egger 回归法、加权中位数法、加权模式法和简单模式法作为补充。为了解释水平多效性的潜在影响, 使用了常用的 MR-Egger 方法, 该方法在截距项显著的情况下识别水平多效性的存在。采用 Cochran's Q 检验评估两样本 MR 中 snp 特异性因果效应效应量的异质性, 当 Cochran's Q 统计量的 p 值小于 0.05 时, 我们定义异质性存在。最后, 我们进行了留一法敏感性分析, 以评估单个 snp 对总体估计值的影响。

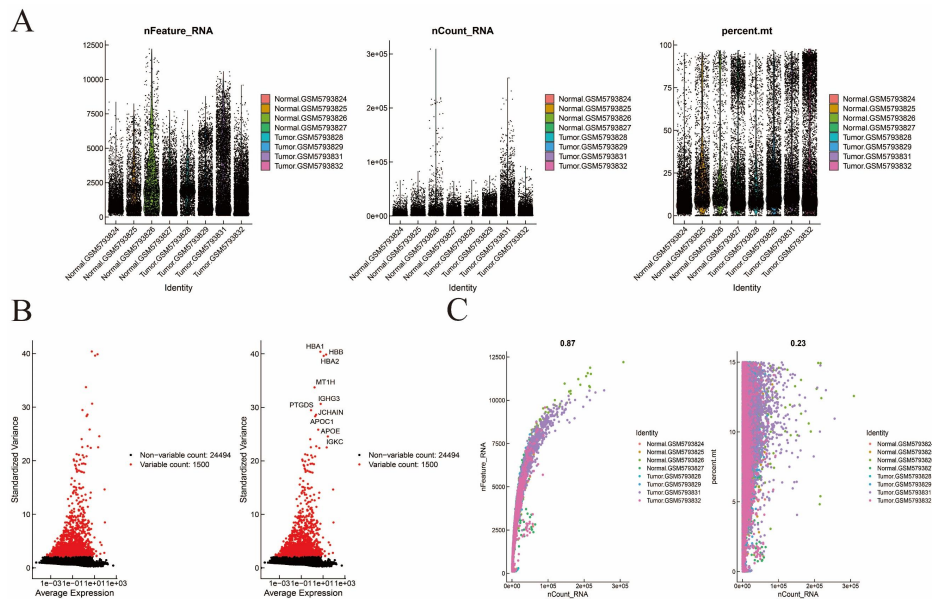
1.6 构建蛋白互作网络及富集分析

GeneMANIA(<http://www.genemania.org>) 包含基因信息、分析基因列表和具有高精度的预测算法优先排序基因以进行功能分析, 用于构建蛋白—蛋白互作网络, 分析与上文得到的关键基因的互作基因。利用上文得到的关键基因及互作基因, 基因本体论(GO)和京都基因与基因组百科全书(KEGG)通路富集 使用 R 包“clusterprofiler”进行注释和可视化, 以及差异表达基因的 GO 分析和 KEGG 通路分析, 是了解与疾病进展相关的生物学功能的常用方法。 p 值阈值设置为小于 0.05 具有统计学意义。

2. 结果

2.1 预处理单细胞测序数据表达谱的结果

本研究包括 8 例前列腺组织的 scRNA-seq, 结果其中 4 个来自正常样本组织, 4 个来自前列腺癌标本组织。为了进行后续分析, 我们只保留了 nFeature_RNA 大于 50 的细胞(图 1A)。共有 21 711 个具有特征表达水平的细胞被纳入进一步分析。筛选前 1,500 个波动最明显的基因(图 1B)。测序深度与线粒体及细胞内总序列呈正相关(图 1C)。

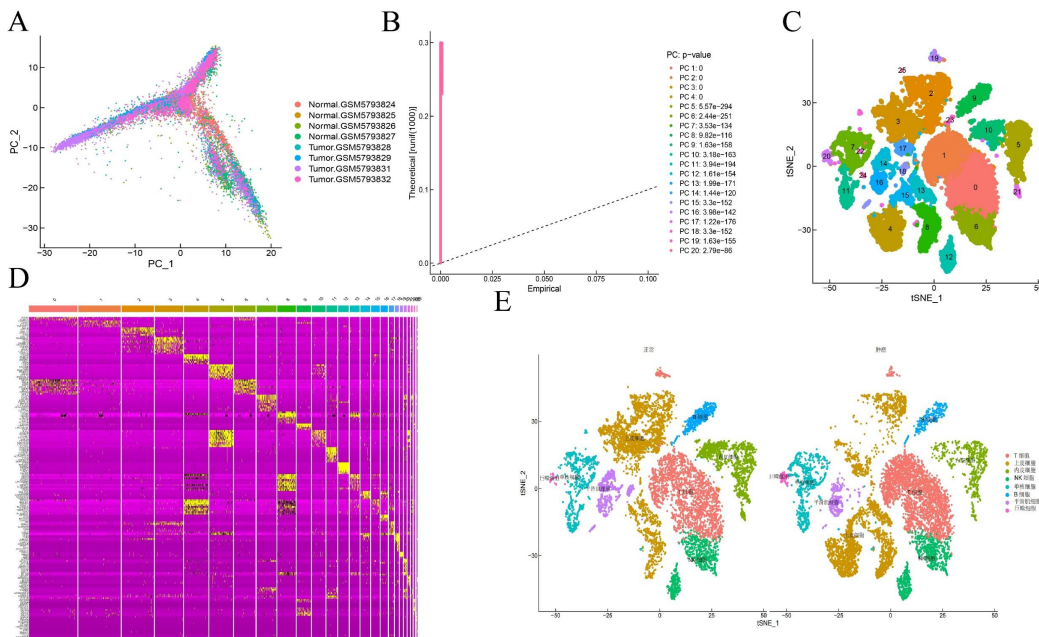


A: 样本中检测到的基因数量、测序深度、线粒体基因含量; B: 火山图显示样本中基因的波动。C: 测序深度与细胞内总序列及线粒体基因含量的关系。

图 1 单细胞测序数据预处理

2.2 单细胞样本亚型聚类分析

我们利用主成分分析(PCA)来降低差异基因的维数(图 2A)。选取差异有统计学意义的 20 个主成分进行进一步分析(图 2B), 接下来, 我们根据 t 分布随机邻居嵌入(tSNE)算法将 21 711 个 PCa 数据聚合为 26 个聚类(图 2C), 并绘制热图(图 2D)。通过“singleR”包、ref_Human_all 数据库寻找标记基因对不同的细胞簇进行注释及分组展示, 得到 8 个细胞簇, 分别是 T 细胞、上皮细胞、内皮细胞、NK 细胞、单核细胞、B 细胞、平滑肌细胞和巨噬细胞 (图 2E)。

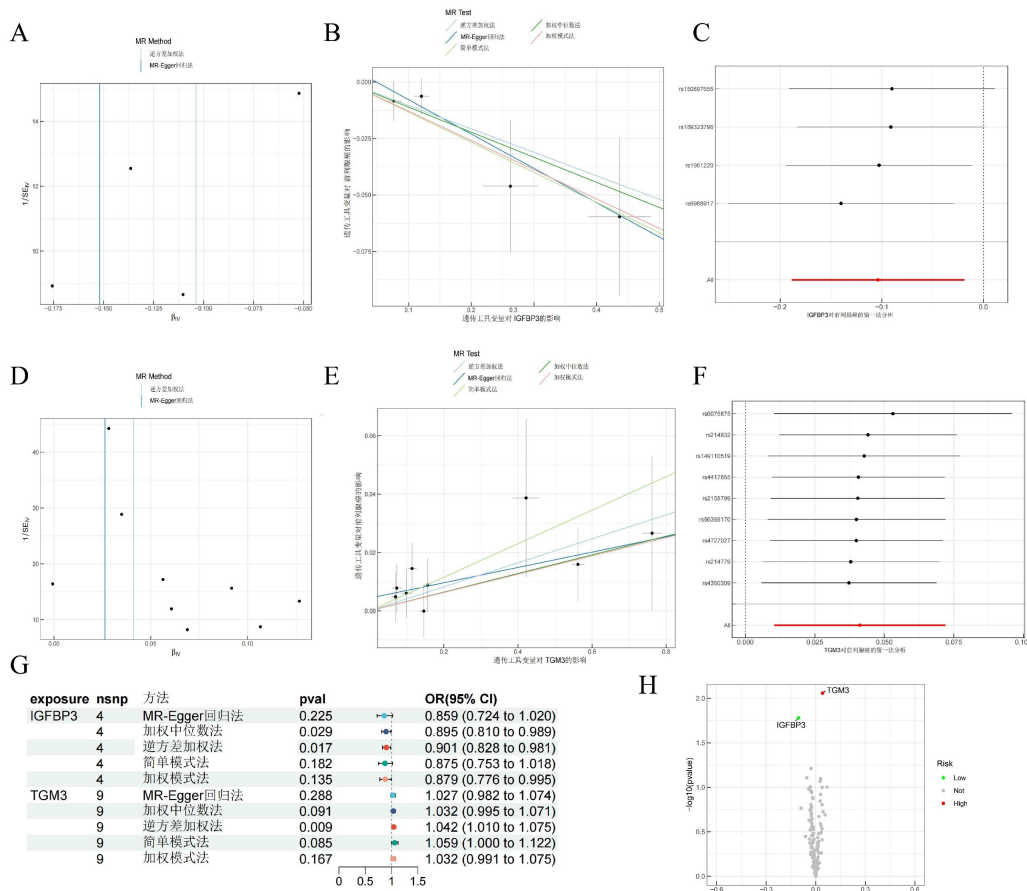


A: scRNA-seq 数据进行 PCA 初步降维。B: 识别出前 20 项主成分的 P 值; C: tSNE 算法进行细胞类型聚集; D: 各集群标记基因热图。黄色为上调基因, 紫色为下调基因。E: 分组后的细胞类型注释。

图 2 细胞亚群与注释

2.3 标记基因的孟德尔随机化分析

为进一步确定 scRNA-seq 得出的各类细胞亚型间差异基因对 PCa 是否有因果关系。通过 MR 分析, 最终只在上皮细胞间的差异基因中筛选出了 2 个有因果关系的标记基因: IGFBP-3 (IVW-MR, $OR = 0.901$; 95% CI : 0.828 ~ 0.981; $P = 0.017$) 和 TGM3 (IVW-MR, $OR = 1.042$; 95% CI : 1.010 ~ 1.075; $P = 0.009$)。Cochran's Q 检验的 P 值与 MR-Egger 法检验多效性的 P 值均大于 0.05, 表明均无水平多效性与异质性, 具体数值可见表 1。漏斗图显示, 潜在偏倚对因果的关联影响较小(图 3A,D)。此外, 进行了留一法敏感性分析表明, 排除任何 snp 对总体误差的影响较小(图 3C,F), 这提示所选的 2 个关键基因与 PCa 具有稳健的因果关联。最后使用森林图及火山图直观地提示 TGM3 可能是 PCa 的高危因素; 而 IGFBP-3 可能是 PCa 的低危因素(图 3G-H)。



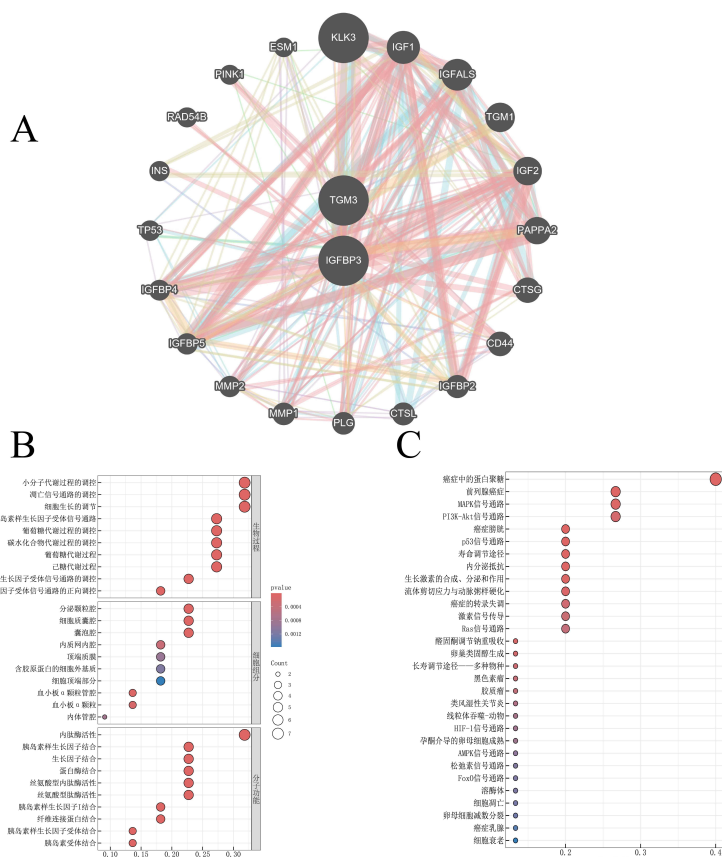
A-C: IGFBP-3 与前列腺癌之间的孟德尔随机化分析结果; D-F: TGM3 与前列腺癌之间的孟德尔随机化分析结果; G-H: 关键基因的森林和热图

图3 孟德尔随机化分析

表1 因果关系的异质性和多效性检验

基因	异质性检验		多效性检验	
	Q 值	P 值	Q 值	P 值
IGFBP-3	1.169	0.761	0.767	0.681
TGM3	3.244	0.918	2.444	0.931

2.4 蛋白互作网络构建及基因富集分析



A: 关键基因的蛋白互作网络。B-C: 关键基因和相互作用基因的富集分析。

图4 蛋白互作网络及富集分析

3. 讨论

scRNA-seq 是近年来广泛应用于生物和医学研究的一项变革技术，可以在单个细胞水平上精确分析基因表达模式。这种方法对于揭示疾病状态的细胞异质性尤其重要，特别是在复杂和异质性的癌症如 PCa 中。MR 分析是一种统计工具可用于探索暴露与结局之间的因果关系，为 PCa 的分子机制和新治疗策略的发展提供重要见解。

本文确定了 8 种细胞亚型,并对不同细胞类型间的差异基因进行 MR 分析,确定了胰岛素样生长因子结合蛋白-3 (Insulin-like Growth Factor Binding Protein-3, IGFBP-3)和转谷氨酰胺酶 3 (Transglutaminase 3, TGM3) 的遗传变异与 PCa 的发生发展有关。IGFBP-3 是 IGF-i 的六个特征结合伴侣之一,也是在循环中发现的最丰富的 IGFBP。在多种细胞系统中, IGFBP-3 以非胰岛素样生长因子依赖的方式诱导细胞凋亡^[8]。通过抑制 mTOR 的激活, IGFBP-3 增加癌细胞对 IL-24 诱导的凋亡的敏感性,这可能在预防 PCa 方面也很重要,并且作为转移性 PCa 治疗方案的潜在辅助治疗^[9]。Chen 等^[10]的研究证实 METTL3 抑制剂通过上调 IGFBP3 抑制 AKT 通路发挥抗前列腺癌作用,且 IGFBP3 敲低可逆转其抑癌效果,表明 IGFBP3/AKT 通路是其抗前列腺癌效果的必需通路。但也有研究表明 IGFBP-3 促进了 PCa 的进展^[11]。比如在治疗方面, IGFBP-3 表达通过激活 EGFR 和 DNA- PKcs 来增强 DNA 修复能力,从而促进奥拉帕利耐药^[12]。也有研究表明 IGFBP-3/PSA 比值已被提议作为晚期 PCa 的预测标志物^[13]。总之,一方面 IGFBP-3 可以作为肿瘤的预后标志物,帮助评估肿瘤的侵袭性和预后风险,指导临床治

疗方案的选择;另一方面,IGFBP-3 作为治疗靶点的有效性仍在探索中^[8]。TGM3 是一种负责表皮形成过程的蛋白质,属于交联酶家族,已在毛发纤维、毛囊和表皮,以及其他器官,包括黏膜、脑、胃、脾、小肠、睾丸和骨骼肌中发现^[14]。有研究初步报道 TGM3 是肿瘤发生过程中的肿瘤相关抑制因子,参与细胞凋亡机制^[15]。TGM3 的下调可能参与了肿瘤细胞凋亡率的降低。一些对肿瘤细胞系的研究间接证明了 TGM3 蛋白对细胞增殖的抑制作用是由细胞凋亡引起的,TGM3 与 PCa 的关系尚未见报道^[16-17]。

本研究结合了 scRNA-seq 与 MR 进行分析 PCa 的潜在关键基因,然而其局限性主要体现在使用的数据均来自欧洲人群。因此,研究结果是否能推广至中国人群尚不明确,需要进一步的临床验证。此外,当前研究缺乏直接的体外或体内实验证实,未来仍需通过更多的实验研究来进一步验证和支持这些发现。

4. 小结

在这项研究中,通过分析 scRNA-seq 数据,确定了与 PCa 相关的各种细胞亚型的差异基因,接着对获得的基因进行基于 eQTL 和 GWAS 数据的 MR 分析,筛选到基因 IGFBP-3 和 TGM3 与 PCa 间存在因果关系,其中 TGM3 可能是 PCa 的高危因素;而 IGFBP-3 可能是 PCa 的低危因素。这两个关键基因相关潜在信息通路可以为 PCa 的发展提供额外的信息,也表明 IGFBP-3 和 TGM3 是 PCa 潜在的生物标志物。

利益冲突声明: 本文所有作者均声明不存在利益冲突。

参考文献

- [1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: a cancer journal for clinicians, 2024, 74(3): 229-263.
- [2] LIU X, YU C, BI Y, et al. Trends and age-period-cohort effect on incidence and mortality of prostate cancer from 1990 to 2017 in China[J]. Public Health, 2019, 172: 70-80.
- [3] CRAWFORD E D, HEIDENREICH A, LAWRENTSCHUK N, et al. Androgen-targeted therapy in men with prostate cancer: evolving practice and future considerations[J]. Prostate Cancer and Prostatic Diseases, 2019, 22(1): 24-38.
- [4] LI H, COURTOIS E T, SENGUPTA D, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors[J]. Nature Genetics, 2017, 49(5): 708-718.
- [5] DAVEY SMITH G, HEMANI G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies[J]. Human Molecular Genetics, 2014, 23(R1): R89-98.
- [6] MORADI A, SHARMA H, SHARMA R D, et al. Identification of Candidate mRNA Isoforms for Prostate Cancer-Risk SNPs Utilizing Iso-eQTL and sQTL Methods[J]. International Journal of Molecular Sciences, 2022, 23(20): 12406.
- [7] HEIDEGGER I, FOTAKIS G, OFFERMANN A, et al. Comprehensive characterization of the prostate tumor microenvironment identifies CXCR4/CXCL12 crosstalk as a novel antiangiogenic therapeutic target in prostate cancer[J]. Molecular Cancer, 2022, 21(1): 132.
- [8] WANG Y, ZHANG H, ZHANG X, et al. The role of IGFBP-3 in tumor development and progression: enlightenment for diagnosis and treatment[J]. Medical Oncology (Northwood, London, England), 2024, 41(6): 141.
- [9] DU Y, LONG Q, SHI Y, et al. Insulin-like growth factor binding protein-3 mediates interleukin-24-induced apoptosis through inhibition of the mTOR pathway in prostate cancer[J]. Oncology Reports, 2015, 34(5): 2273-2281.
- [10] CHEN X, WANG M, WANG H, et al. METTL3 inhibitor suppresses the progression of prostate cancer via IGFBP3/AKT pathway and synergizes with PARP inhibitor[J]. Biomedicine & Pharmacotherapy, 2024, 179: 117366.
- [11] CHEN X, SHAO Y, WEI W, et al. Downregulation of LOX promotes castration-resistant prostate cancer progression via IGFBP3[J]. Journal of Cancer, 2021, 12(24): 7349-7357.
- [12] LESLIE A R, NING S, ARMSTRONG C M, et al. IGFBP3 promotes resistance to Olaparib via modulating EGFR signaling in advanced prostate cancer[J]. iScience, 2024, 27(2): 108984.

- [13] MAYO J C, HEVIA D, QUIROS-GONZALEZ I, et al. IGFBP3 and MAPK/ERK signaling mediates melatonin-induced antitumor activity in prostate cancer[J]. Journal of Pineal Research, 2017, 62(1): e12373.
- [14] ZHANG W, WU C, ZHOU K, et al. Clinical and immunological characteristics of TGM3 in pan-cancer: A potential prognostic biomarker[J]. Frontiers in Genetics, 2023, 13: 993438.
- [15] FENG Y, JI D, HUANG Y, et al. TGM3 functions as a tumor suppressor by repressing epithelial-to-mesenchymal transition and the PI3K/AKT signaling pathway in colorectal cancer[J]. Oncology Reports, 2020, 43(3): 846-876.
- [16] ES C, EV A, EA V. Transglutaminase 3: The Involvement in Epithelial Differentiation and Cancer[J]. Cells, 2020, 9(9): 1996.
- [17] GUO J, ZHAO C, ZHANG X, et al. A novel 8-gene panel for prediction of early biochemical recurrence in patients with prostate cancer after radical prostatectomy[J]. American Journal of Cancer Research, 2022, 12(7): 3318-3332.

Single-cell and Mendelian randomization were used to identify key genes in prostate cancer

Haoran Han, Ruimin Ren

Third Hospital of Shanxi Medical University, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, urinary surgery, Taiyuan, 030032, China

【Abstract】 Objective: To identify the hub genes related to prostate cancer (PCa), gene eQTL analysis combining differentially expressed genes from single-cell transcriptomics with genome-wide association study (GWAS) data. **Methods** This study analyzed single-cell sequencing data from 8 prostate tissue samples in the GSE193337 dataset. Cell clustering, annotation, and subtype marker gene identification were performed using R. Mendelian randomization, integrating GWAS and gene eQTL data, identified significant genes through sensitivity and heterogeneity testing. Finally, a protein-protein interaction network was constructed, followed by GO and KEGG pathway enrichment to identify key functional pathways. **Results** This study analyzed 8 prostate samples (4 cancer, 4 normal), annotating 8 cell types (e.g., T cells, epithelial cells). Mendelian randomization identified IGFBP-3 (OR=0.901, P=0.017) and TGM3 (OR=1.042, P=0.009) as causally linked to prostate cancer, with sensitivity analysis confirming robustness. Protein interaction and enrichment analysis highlighted key genes in prostate cancer and MAPK pathways, regulating growth and metabolism. **Conclusions** This study identified IGFBP-3 and TGM3 as potential biomarkers for prostate cancer through scRNA-seq and MR analysis.

【Key words】 Prostate cancer; single-cell RNA sequencing; Mendelian randomization analysis; expression quantitative trait loci; Marker genes