DOI:10.19893/j.cnki.ydyxb.2025-0030

基于GEO数据库的宫颈癌相关基因与通路的生物信息学分析

何　朵[1],马佳蕊[2],李　根[1],张　玉[1],尚亚萍[1*]

(1. 延安市人民医院病理科,陕西 延安 716000; 2. 延安大学医学院,陕西 延安 716000)

摘　要:目的　通过生物信息学方法挖掘宫颈癌发生、发展过程中的差异基因及其所涉及的信号转导通路,从而筛选可能用于宫颈癌的潜在预后生物标志物及治疗靶点。**方法**　在GEO(Gene Expression Omnibus)数据库中下载高通量测序基因表达谱的宫颈癌细胞系数据(GSE241703),并采用R语言将测序表达矩阵数据进行预处理后再分析各组细胞之间的差异基因,然后对筛选出的显著相关性基因进行基因功能注释(GO富集分析)、KEGG(Kyoto Encyclopedia of Genes and Genomes)通路富集分析以及GSEA(Gene Set Enrichment Analysis)富集分析。**结果**　HeLa细胞相较于HaCaT细胞而言(筛选条件为调整后的P值<0.05且$|\log_2FC|\geq1$),上调基因有3 558个,下调基因有5 129个;SiHa细胞相较于HaCaT细胞而言,上调基因有3 634个,下调基因有4 147个;两组细胞所得的差异基因均涉及表皮发育、细胞基质黏附等生物学过程;细胞组分均包含细胞外基质、细胞—基底连接等;分子功能均涉及肌动蛋白结合、GTP酶活性调节等;KEGG富集分析显示差异基因共同富集在Ras相关蛋白1(ras-associated protein 1,Rap1)信号通路、细胞黏附分子、丝裂原活化蛋白激酶(mitogen-activated protein kinase,MAPK)信号通路等。GSEA富集分析显示差异基因共同富集在上皮间质转化(epithelial-mesenchymal transition,EMT)信号通路、氧化磷酸化(oxidative phosphorylation)信号通路、骨髓细胞瘤病毒癌基因同源物(myelocytomatosis viral oncogene homolog,MYC)信号通路。其中HeLa细胞与SiHa细胞相较于HaCaT细胞而言,上皮细胞黏附分子(EpCAM)、细胞黏附分子1(CDH1)表达显著降低;波形蛋白(vimentin)的表达水平显著升高。**结论**　HPV感染后可能通过影响细胞黏附分子的表达,进一步诱发EMT的发生,从而促进宫颈癌的侵袭及迁移,其机制可能是通过HPV E6激活RAP1信号通路实现的。

关键词:宫颈癌;HPV E6/E7;细胞黏附;上皮间质转化;Ras相关蛋白1

中图分类号:R737.33　　**文献标识码:**A　　**文章编号:**1672-2639(2025)04-0024-08

Bioinformatics analysis of cervical cancer-related genes and pathways based on the GEO database

HE Duo[1],*MA Jiarui*[2],*LI Gen*[1],*ZHANG Yu*[1],*SHANG Yaping*[1*]

(1.Department of Pathology, Yan′an People′s Hospital, Yan′an 716000, China;
2.Yan′an Medical College of Yan′an University, Yan′an 716000, China)

Abstract: Objective To identify differentially expressed genes (DEG) and their associated signaling pathways during the progression of cervical cancer using bioinformatics approaches, and to screen for potential biomarkers and therapeutic targets for the diagnosis and treatment of cervical cancer. **Methods** High-throughput gene expression profiling data of cervical cancer cell lines (GSE241703) were obtained from the Gene Expression Omnibus (GEO) database. The expression matrix was preprocessed using R software before identifying DEG between cervical cancer cell lines and normal control cells. Subsequently, functional annotation of DEG was performed through Gene Ontology (GO) enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG)

基金项目:延安市人民医院院内培育基金普通项目(2022-PY-16)

作者简介:何　朵(1993—),女,陕西延安人,住院医师。研究方向:肿瘤病理。

* **通信作者:**尚亚萍(1994—),女,陕西吴起人,主治医师。研究方向:肿瘤病理。E-mail:3959530487@qq.com

pathway analysis, and Gene Set Enrichment Analysis (GSEA). **Results** Compared with HaCaT cells (filtering criteria: adjusted P-value<0.05 and $|\log_2$ fold change$|\geq 1$), HeLa cells exhibited 3 558 upregulated and 5 129 downregulated genes, while SiHa cells showed 3 634 upregulated and 4 147 downregulated genes. These DEG were involved in biological processes such as epidermal development and cell-matrix adhesion; cellular components included the extracellular matrix and cell-basement junctions; molecular functions were associated with actin binding and regulation of GTPase activity. KEGG pathway analysis revealed that DEG were significantly enriched in pathways including the Rap1 signaling pathway, cell adhesion molecules, and the MAPK signaling pathway. GSEA further demonstrated enrichment in pathways such as epithelial-mesenchymal transition (EMT), oxidative phosphorylation and the myelocytomatosis viral oncogene homolog (MYC) signaling pathway. Notably, compared with HaCaT cells, HeLa and SiHa cells showed significantly reduced expression of epithelial cell adhesion molecule (EpCAM) and E-cadherin (CDH1), while the expression of vimentin, a mesenchymal marker, was markedly increased. **Conclusion** HPV infection may promote EMT by modulating the expression of cell adhesion molecules, thereby enhancing the invasive and migratory potential of cervical cancer cells. This mechanism may be mediated through HPV E6-induced activation of the Rap1 signaling pathway.

Key words: Cervical cancer; HPV E6/E7; Cell adhesion; EMT; Rap1

宫颈癌是除了乳腺癌之外第二大常见的女性生殖系统恶性肿瘤[1]。在发展中国家中，宫颈癌的发病率排在女性生殖道恶性肿瘤首位[2]。有关研究[3-4]表明，全球每年约有66.1万新发宫颈癌患者，其中约有34.8万死于宫颈癌，其发病率及死亡率均位于世界前列，且呈年轻化趋势，严重危害着女性的生命健康。

宫颈癌的发生是一个多步骤、多因素共同调控的复杂生物学过程，其中人乳头瘤病毒（human papilloma virus，HPV）的持续性感染是引发并推动其发生、发展的关键诱因，可导致宫颈上皮内瘤变（cervical intraepithelial neoplasia，CIN）等癌前病变，最终可能进展为宫颈浸润癌[5-6]。HPV是一种嗜黏膜和皮肤的小型无包膜双链环状DNA病毒，直径约为55 nm，基因组含有大约8 000个碱基对，属于乳头多瘤空泡病毒科（*Papovaviridae*）中的乳头（状）瘤病毒属，具有很强的嗜上皮性[7]。当HPV感染宿主后，可在表皮、生殖器及口腔黏膜等部位的上皮基底细胞中大量复制。当机体免疫力下降时，HPV基因组就会随机整合到宿主基因组中，此时HPV的E2基因就会被沉默，从而减弱了HPV对E6和E7基因的抑制效果，这就会致使E6和E7基因的过度表达，进而破坏细胞内部的信号转导通路，促使感染细胞无限增殖甚至发生癌变[8-9]。

但目前仍有很多问题需要解决，E2基因是如何抑制E6、E7基因的，而E6和E7基因又是如何促进宫颈癌的发生发展等问题都值得我们继续探究，所以更全面地了解HPV各个基因之间的级联效应，仍是发现、诊断、治疗及预防宫颈癌的关键所在。

本研究从GEO数据库（Gene Expression Omnibus）下载了高通量测序生成的宫颈癌细胞系基因表达谱数据（cervical cancer derived cell lines：Gene Expression Profile，GSE241703），采用R语言分析筛选不同细胞组间的差异基因，对差异基因进行GO（Gene Ontology）生物学过程富集分析、KEGG（Kyoto Encyclopedia of Genes and Genomes）通路富集分析以及GSEA（Gene Set Enrichment Analysis）富集分析，探讨不同细胞组间差异基因所影响的细胞生物学过程及分子信号通路，进而筛选出潜在的宫颈癌诊断生物标志物及治疗靶点。

1 方法

1.1 基因芯片数据的获取

本研究以GEO数据库为数据源，设定“cervical cancer”为关键词进行检索，获取GSE241703数据集，该数据集基于GPL24676（cervical cancer derived cell lines：Gene Expression Profile），旨在鉴定宫颈癌细胞系中差异表达的基因。实验所用细胞系为SiHa、HeLa和C33A宫颈癌细胞系。其中将人永生化角质形成细胞HaCaT作为对照，进行mRNA测序，每个细胞模型均进行了两次独立重复测序。

1.2 筛选差异基因

下载高通量测序基因表达谱数据（cervical cancer derived cell lines：Gene Expression Profile，

GSE241703),利用R语言对芯片原始数据进行探针注释转换,并剔除无对应基因及重复探针的数据。随后,采用GEO数据库自带的在线分析工具GEO2R(https://www.ncbi.nlm.nih.gov/geo/geo2r)筛选差异表达基因,设定筛选条件为调整后的P值<0.05且$|\log_2FC|\geq1$(其中FC表示差异倍数,即fold change)。最后,通过火山图和热图对筛选结果进行可视化展示。

1.3　差异基因的功能注释及信号通路富集分析

利用生物学信息注释数据库DAVID(The DataBase for Annotation, Visualization and Integrated Discovery)对所有差异基因进行GO富集分析、KEGG通路富集分析和GSEA富集分析。在结果筛选中,设定显著性阈值为P<0.05,并利用R语言的ggplot2包对富集分析结果进行可视化展示。

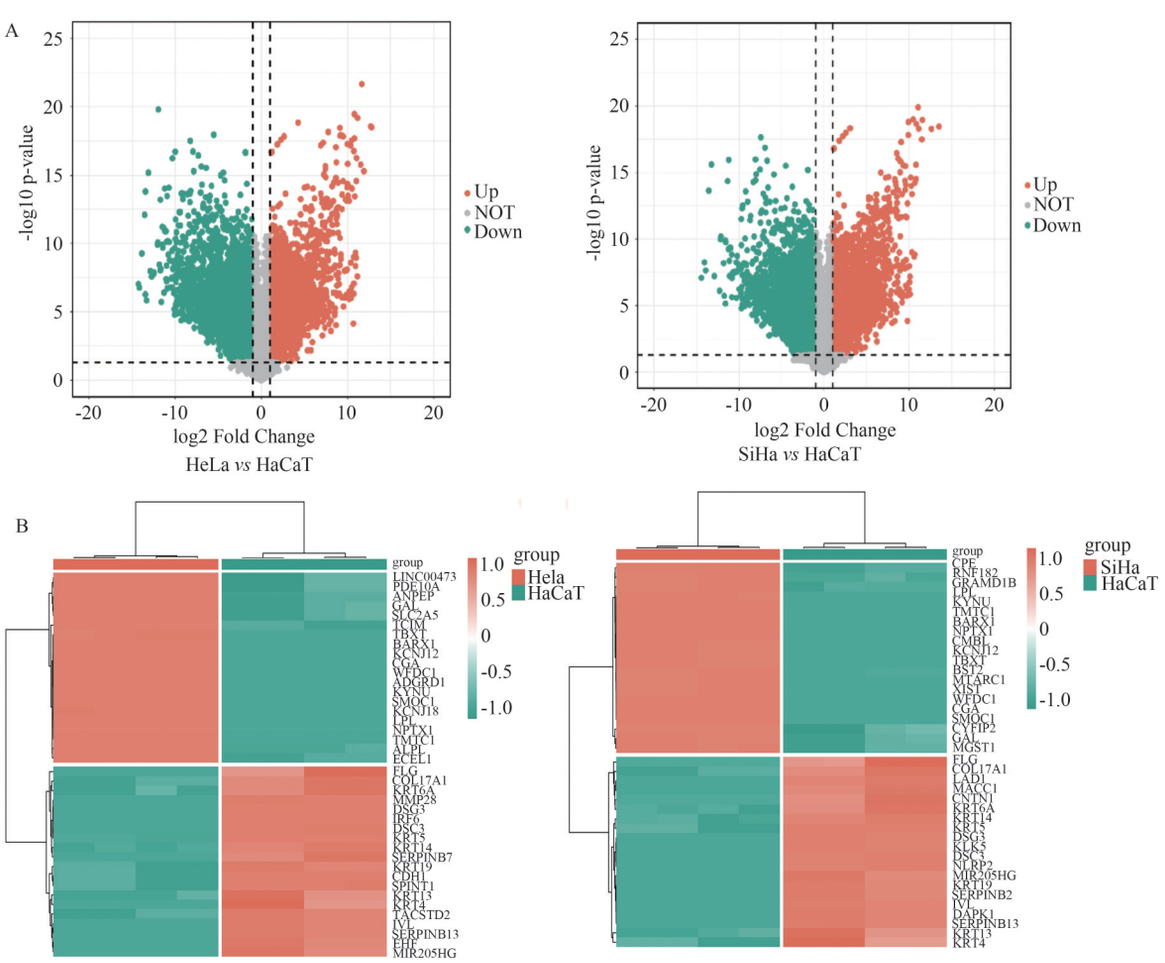
1.4　统计学方法

采用R软件4.1.3版本进行数据的统计学分析和图形绘制。采用单因素方差分析(One-way ANOVA)对多组间的计量数据进行比较,并通过Dunnett's t检验对多重比较结果进行校正。以双侧P<0.05作为判定组间差异具有统计学意义的标准。

2　结果

2.1　GEO数据库筛选差异基因

从GEO数据库中下载GSE241703数据集,使用R语言对芯片原始数据处理后获取三组细胞数据,分别为HPV18阳性的HeLa细胞、HPV16阳性的SiHa细胞、人永生化角质形成细胞HaCaT,分别将HeLa细胞与HaCaT细胞、SiHa细胞与HaCaT细胞进行比较,从而筛选HPV阳性宫颈癌细胞相较于正常细胞之间的差异基因。结果显示,HPV18阳性的HeLa细胞相较于HaCaT而言,上调的基因有3 558个,下调的基因有5 129个(图1A);HPV16阳性的SiHa细胞相较于HaCaT而言,上调的基因有3 634个,下调的基因有4 147个(图1B)。



A:差异蛋白火山图;B:差异蛋白热图。

图1　HPV阳性宫颈癌细胞系差异基因的筛选

2.2 基因生物学功能及富集分析

为了进一步明确差异基因的生物学功能以及所涉及的信号转导通路，开展GO富集分析、KEGG通路富集分析以及GSEA基因集富集分析。GO富集分析结果显示：HeLa细胞与HaCaT细胞相比、SiHa细胞与HaCaT细胞相比所得的差异基因参与的生物学过程（biological process，BP）均涉及表皮发育、细胞基质黏附等；细胞组分（cellular component，CC）均包含细胞外基质，细胞-基底连接等；分子功能（molecular function，MF）均涉及肌动蛋白结合、GTP酶活性调节等（图2）。

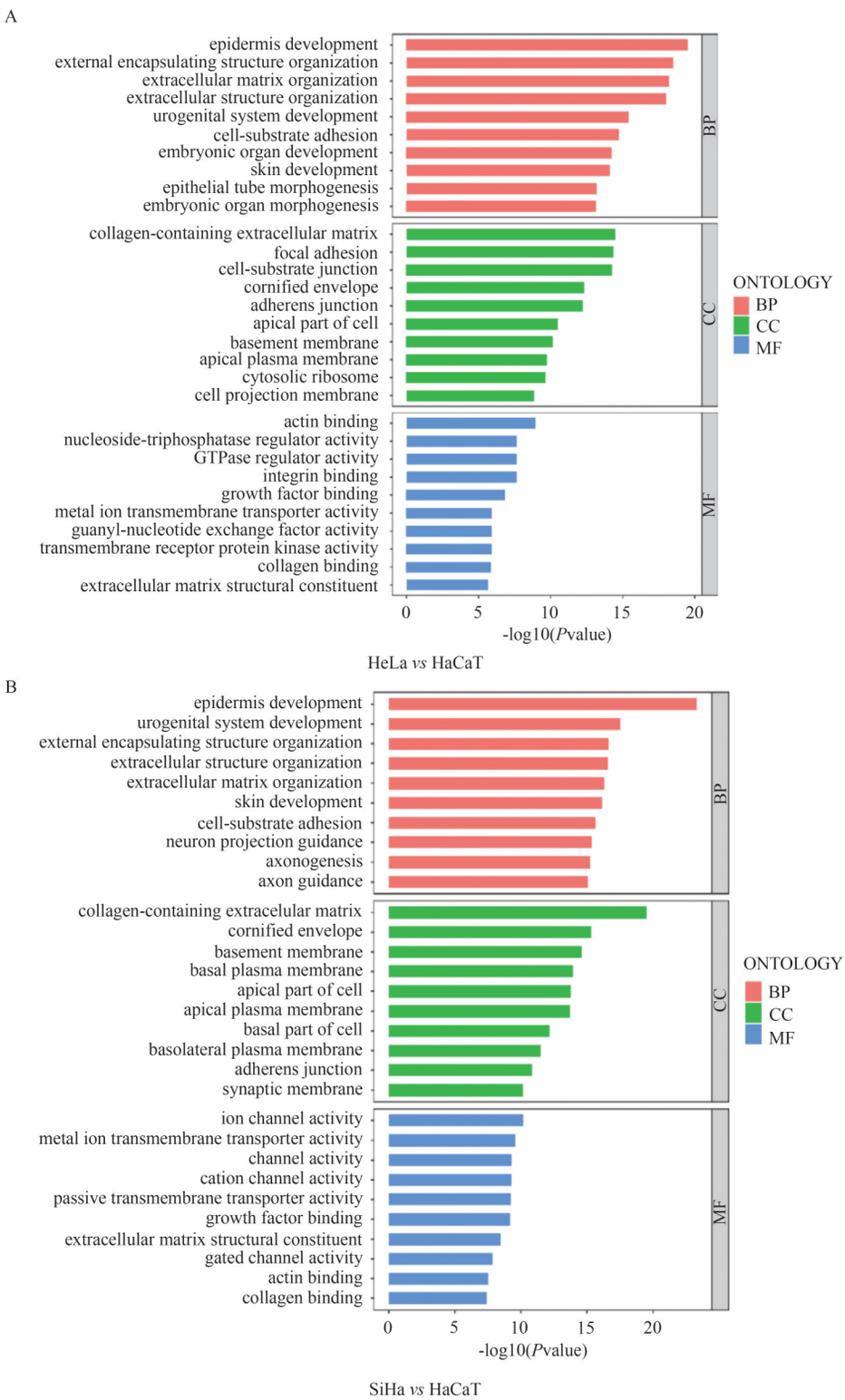


图2 HPV阳性宫颈癌细胞系差异基因的GO富集分析

KEGG 分析显示，HeLa 细胞与 HaCaT 细胞相比、所得的差异基因共涉及 39 条信号转导通路(图 3A)，SiHa 细胞与 HaCaT 细胞相比所得的差异基因共涉及 56 条信号转导通路(P<0. 05，图 3B)，上调基因共同富集在 Rap1 信号通路、细胞黏附分子、MAPK 信号通路等。

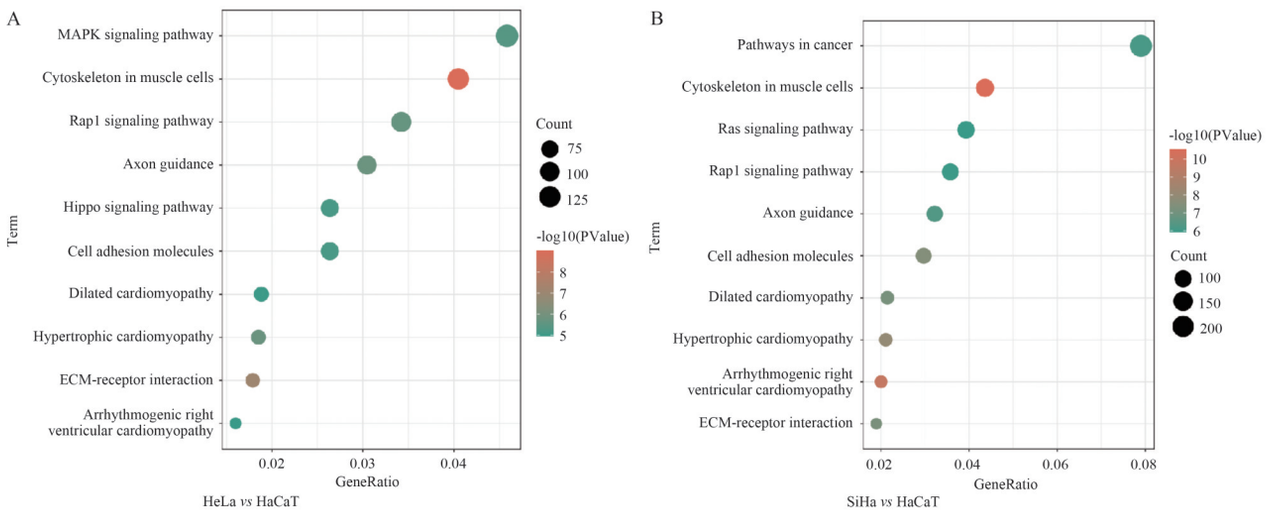


图 3　HPV 阳性宫颈癌细胞系差异基因的 KEGG 富集分析

GSEA 富集分析结果显示，HeLa 细胞与 HaCaT 细胞相比所得的差异基因主要在 10 条信号通路中富集，SiHa 细胞与 HaCaT 细胞相比所得的差异基因主要在 9 条信号通路中富集(P<0. 05)，上调基因共同富集在上皮间质转化(epithelial-mesenchymal transition，EMT)信号通路、氧化磷酸化信号通路、MYC 信号通路(图 4)。

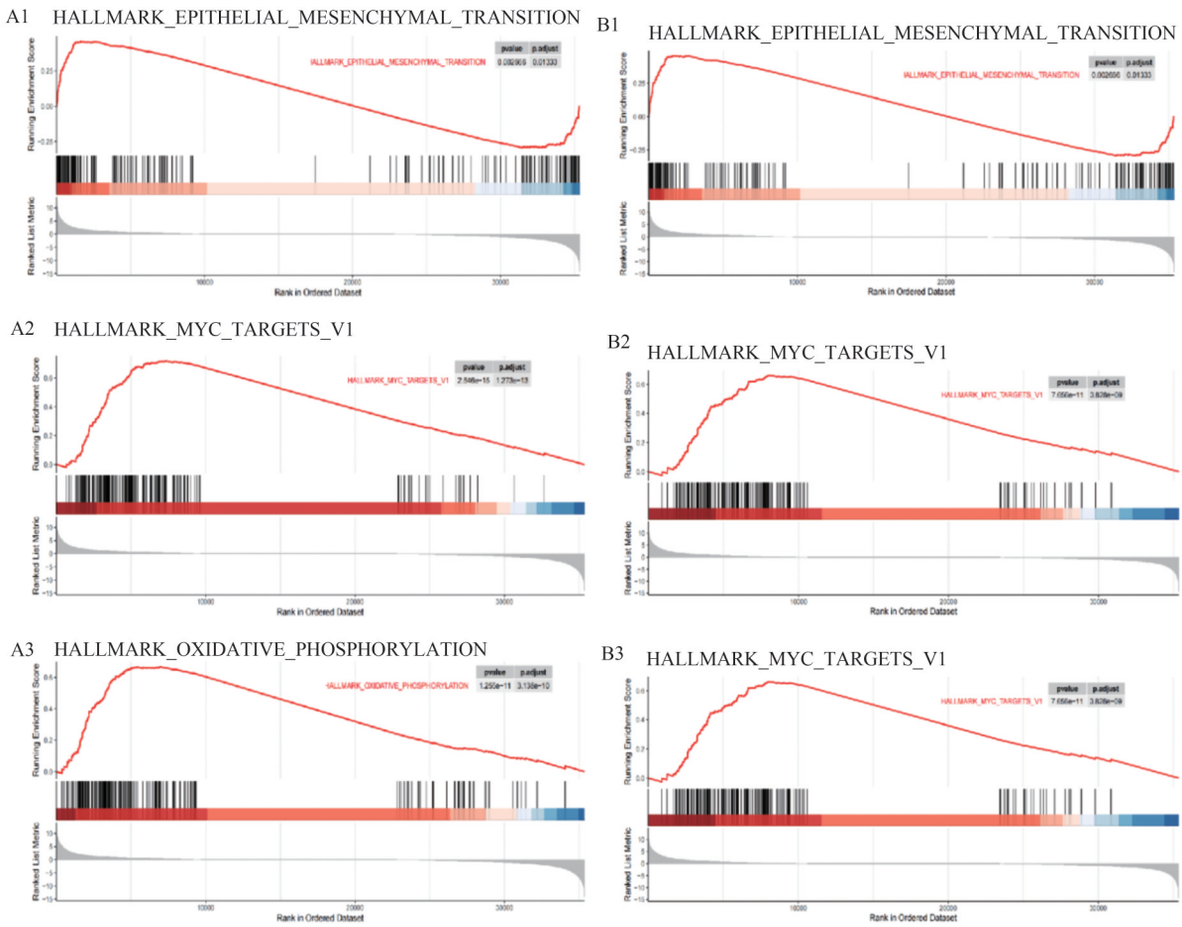
在进行本研究数据的分析过程中，我们还发现与 HaCaT 细胞相比，HeLa 细胞和 SiHa 细胞中的上皮细胞标志物：上皮细胞黏附分子(EpCAM)，细胞黏附分子 1(CDH1)的基因表达都显著下降；间充质标志物，波形蛋白(vimentin)的基因表达则显著上升。差异均具有统计学意义(图 5)。

3　讨论

统计数据[10]显示，近些年来，全世界范围的宫颈癌的发病率与死亡率都处于日趋升高的状态，极大地危及女性的生命健康。而晚期出现复发并发生转移，则成为导致宫颈癌患者死亡的主要原因[11]，这是因为肿瘤细胞有着很强的增殖及侵袭能力[12]。而 EMT 在肿瘤细胞侵袭和转移方面发挥重要作用[13-14]。

EMT 是指上皮肿瘤细胞由于上皮细胞黏附分子如上皮细胞黏附分子(EpCAM)、钙黏附蛋白 E(E-cadherin)、α-连接蛋白(α-catenin)、β 连接蛋白(β-catenin)、细胞角蛋白(cytokeratin)等的表达降低，而丧失黏附特性。同时间充质标志物如 n-钙黏蛋白(N-cadherin)、波形蛋白(vimentin)等的表达升高，使得细胞获得运动和侵袭性的间充质特征。既往研究[15-18]证实，EMT 可通过(TGF-β、Wnt、Notch、Hedgehog 等)多条信号通路被诱导激活，并在乳腺癌[19]、肝细胞癌[20]、结肠癌[21]、卵巢癌[22]等多种恶性肿瘤的侵袭和转移过程中发挥重要作用，也因此被认为是肿瘤转移的关键驱动因素之一。

本研究通过比对 HPV 阳性宫颈癌细胞与正常细胞之间的差异基因，并对差异基因所涉及的生物学功能及信号通路进行富集分析后发现：HPV 阳性的 HeLa 细胞与 SiHa 细胞相较于 HaCaT 而言，有较多的差异基因，对差异基因进行 GO 富集分析显示，两组细胞的差异基因均涉及细胞黏附、细胞骨架以及肌动蛋白合成等生物学功能，KEGG 富集分析显示，两组差异基因均在 Rap1 信号通路、细胞黏附分子、MAPK 信号通路等富集，GSEA 富集分析显示差异基因在 EMT 信号通路、氧化磷酸化信号通路、MYC 信号通路中富集。而在本研究的数据分析中也发现，HPV 阳性的 HeLa 细胞与 SiHa 细胞相较于 HaCaT 而言，上皮细胞标志物如上皮细胞黏附分子(EpCAM)、细胞黏附分子 1(CDH1)的基因表达水平显著下降；而间充质标志物波形蛋白(vimentin)的表达水平显著升高。这些结果提示，HPV 可能通过抑制上皮细胞黏附分子(EpCAM)和细胞黏附分子 1(CDH1)的表达，进而降低由细胞黏附分子 1(CDH1)编码的钙黏蛋白 E(E-cadherin)的表达，促使上皮间质转化的发生，进一步导致宫颈癌细胞的侵袭与转移。



A1–3：HeLa *vs* HaCaT；B1–3：SiHa *vs* HaCaT。

图4　HPV阳性宫颈癌细胞系差异基因的GSEA富集分析

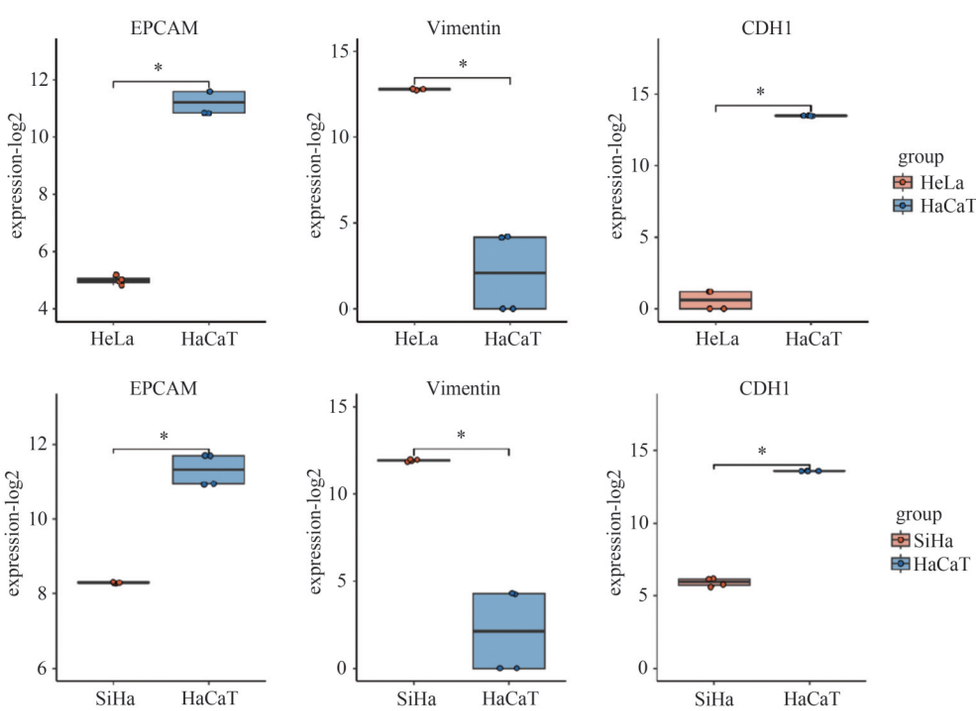


图5　HPV阳性宫颈癌细胞系差异基因上皮间质转化相关分子表达水平

秦祥川等人[23]指出HPV E6可能通过激活Rap1信号通路参与癌细胞的增殖、侵袭及迁移等过程。Rap1是一种与Ras蛋白高度同源的小GTPase蛋白，通过在GTP结合状态(即活性状态)、GDP结合状态(即非活性状态)之间切换以作为分子开关参与多种信号通路转导，并在细胞黏附与迁移等过程中发挥重要作用[24-25]。

结合上述研究结果，提示HPV感染可能通过影响宫颈癌细胞黏附分子的表达，进一步诱发EMT的发生，增强宫颈癌细胞的侵袭性和迁移能力。其潜在机制可能涉及HPV E6对Rap1信号通路的激活，但 HPV E6是否可以通过Rap1信号通路促进宫颈癌的发生发展、HPVE6通过何种信号途径激活Rap1信号通路等的过程以及Rap1信号通路与EMT之间的关系仍需我们进一步研究和发现，从而为临床预后评估及抗肿瘤治疗提供新的思路和方法。

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]SEDETA E, SUNG H, LAVERSANNE M, et al. Recent mortality patterns and time trends for the major cancers in 47 countries worldwide[J]. Cancer Epidemiology, Biomarkers & Prevention, 2023, 32(7): 894-905.

[3]JOKHADZE N, DAS A, DIZON D S. Global cancer statistics: A healthy population relies on population health [J]. CA: A Cancer Journal for Clinicians, 2024, 74(3): 224-226.

[4]CHENG L Y, ZHAO J Q, ZOU T T, et al. Cervical cancer burden and attributable risk factors across different age and regions from 1990 to 2021 and future burden prediction: Results from the global burden of disease study 2021[J]. Frontiers in Oncology, 2025, 15: 1541452.

[5]KOLIOPOULOS G, NYAGA V N, SANTESSO N, et al. Cytology versus HPV testing for cervical cancer screening in the general population [J]. The Cochrane Database of Systematic Reviews, 2017, 8(8): CD008587.

[6]HOFFMAN S R, LE T, LOCKHART A, et al. Patterns of persistent HPV infection after treatment for cervical intraepithelial neoplasia (CIN): A systematic review [J]. International Journal of Cancer, 2017, 141(1): 8-23.

[7]HARDEN M E, MUNGER K. Human papillomavirus molecular biology[J]. Mutation Research/Reviews in Mutation Research, 2017, 772: 3-12.

[8]SETAYESHI S, HASANZADEH A, YAHYAPOUR Y, et al. Evaluation of human papillomavirus type 16 viral load and genome physical status in Iranian women with cervical disease[J]. Molecular Biology Reports, 2024, 51(1): 411.

[9]BALHARA N, YADAV R, RANGA S, et al. Understanding the HPV associated cancers: A comprehensive review [J]. Molecular Biology Reports, 2024, 51(1): 743.

[10]VU M, YU J, AWOLUDE O A, et al. Cervical cancer worldwide[J]. Current Problems in Cancer, 2018, 42(5): 457-465.

[11]GENNIGENS C, DE CUYPERE M, HERMESSE J, et al. Optimal treatment in locally advanced cervical cancer[J]. Expert Review of Anticancer Therapy, 2021, 21(6): 657-671.

[12]ADIGA D, ESWARAN S, PANDEY D, et al. Molecular landscape of recurrent cervical cancer[J]. Critical Reviews in Oncology/Hematology, 2021, 157: 103178.

[13]AOUAD P, QUINN H M, BERGER A, et al. Tumor dormancy: EMT beyond invasion and metastasis [J]. genesis, 2024, 62(1): e23552.

[14]HORTA C A, DOAN K, YANG J. Mechanotransduction pathways in regulating epithelial-mesenchymal plasticity[J]. Current Opinion in Cell Biology, 2023, 85: 102245.

[15]SCHUHWERK H, BRABLETZ T. Mutual regulation of TGFβ-induced oncogenic EMT, cell cycle progression and the DDR [J]. Seminars in Cancer Biology, 2023, 97: 86-103.

[16]CHENG Z Y, WANG H B, YANG Z W, et al. LMP2 and TAP2 impair tumor growth and metastasis by inhibiting Wnt/β-catenin signaling pathway and EMT in cervical cancer[J]. BMC Cancer, 2023, 23(1): 1128.

[17]SARKAR S, KANDASAMY T, GHOSH S S. Imatinib impedes EMT and Notch signalling by inhibiting p300 acetyltransferase in breast cancer cells [J]. Molecular Carcinogenesis, 2025, 64(2): 344-356.

[18]ANONYMOUS. FOXC1 modulates stem-like cell properties and chemoresistance through hedgehog and EMT signaling in gastric adenocarcinoma[J]. Molecular Therapy, 2024, 32(3): 852.

[19]PARFENYEV S, SINGH A, FEDOROVA O, et al. Interplay between p53 and non-coding RNAs in the regulation of EMT in breast cancer[J]. Cell Death & Disease, 2021, 12(1): 17.

[20]FYALA A S, SULTAN A. Transforming growth factor-beta and Epithelial-mesenchymal transition is considered an attractive pivotal Genetic-Biomarker for Hepatocellular carcinoma with hepatitis C virus-infection patients [J]. American Journal of Clinical Pathology, 2024, 162(Supplement_1): S126-S127.

[21]IZUTSU R, OSAKI M, SEONG H, et al. AMIGO2

enhances the invasive potential of colorectal cancer by inducing EMT[J]. Cancer Gene Therapy, 2024, 31(12): 1786-1795.

[22]DENG J L, WANG L, CHEN H M, et al. Targeting epithelial-mesenchymal transition and cancer stem cells for chemoresistant ovarian cancer[J]. Oncotarget, 2016, 7(34): 55771-55788.

[23]秦祥川,李金秋,黄晓婧,等. HPV E6通过Rap1信号通路影响宫颈癌细胞增殖、侵袭及迁移的研究[J]. 昆明医科大学学报,2024,45(9):9-16.

[24]HOPPE-SEYLER K, BOSSLER F, BRAUN J A, et al. The HPV E6/E7 oncogenes: Key factors for viral carcinogenesis and therapeutic targets[J]. Trends in Microbiology, 2018, 26(2): 158-168.

[25]SHAH S, BROCK E J, JI K, et al. Ras and Rap1: A tale of two GTPases[J]. Seminars in Cancer Biology, 2019, 54: 29-39.

[收稿日期 2025-06-17;责任编辑 刘敏丽]

(上接第23页)

[30]XU X Y, HU H M, ZENG H Z, et al. Sinisan ameliorates colonic injury induced by water immersion restraint stress by enhancing intestinal barrier function and the gut microbiota structure[J]. Pharmaceutical Biology, 2023, 61(1): 598-609.

[31]何佩宜,李思咏,黄诚宇,等. β-石竹烯的药理学研究进展[J]. 广东药科大学学报,2024,40(5):154-158.

[32]SEIF M, DEABES M, EL-ASKARY A, et al. *Ephedra sinica* mitigates hepatic oxidative stress and inflammation *via* suppressing the TLR4/MyD88/NF-κB pathway in fipronil-treated rats[J]. Environmental Science and Pollution Research International, 2021, 28(44): 62943-62958.

[33]SCANDIFFIO R, GEDDO F, COTTONE E, et al. Protective effects of (E)-β-caryophyllene (BCP) in chronic inflammation[J]. Nutrients, 2020, 12(11): 3273.

[34]VIVEIROS M M H, SILVA M G, DA COSTA J G M, et al. Anti-inflammatory effects of α-humulene and β-caryophyllene on pterygium fibroblasts[J]. International Journal of Ophthalmology, 2022, 15(12): 1903-1907.

[35]REFAAT B, EL-BOSHY M. Protective antioxidative and anti-inflammatory actions of β-caryophyllene against sulfasalazine-induced nephrotoxicity in rat[J]. Experimental Biology and Medicine, 2022, 247(8): 691-699.

[收稿日期 2025-02-04;责任编辑 刘敏丽]