

具有预后价值的乳腺癌发病关键基因鉴别研究

徐久成^{1a,b}, 李成长^{1a,2}

(1.河南师范大学 a.生命科学学院细胞分化调控省部共建国家重点实验室培育基地;
b.计算机与信息工程学院河南省智慧商务与物联网工程实验室,河南 新乡 453007;
2.新乡医学院 基础医学院生理学与神经生物学教研室,河南 新乡 453003)

摘要:目的:筛选与乳腺癌发病相关的关键基因,为研究乳腺癌的诊疗提供新的潜在分子靶标.方法:使用 GEO2R 在线工具比较乳腺癌与正常乳腺组织基因表达情况,进行差异显著性分析,利用基因功能注释工具 DAVID 对差异基因进行 GO 和 KEGG 富集分析.通过 STRING 数据库构建差异基因的蛋白互作网络,基于最大团中心性(maximal clique centrality, MCC)算法鉴别关键基因,利用 Kaplan Meier 生存分析验证关键基因对患者生存时间的影响.结果:乳腺癌基因表达差异分析共产生 491 个差异基因,其中有 254 个上调,237 个下调.GO 富集分析表明差异基因显著富集于肿瘤相关的生物学过程、细胞组分和分子功能,KEGG 通路富集分析主要集中于 PI3K-Akt 信号通路、黏附斑和癌症通路.基于 MCC 算法共选取 10 个关键基因,其中的 4 个基因(ISG15, IFIT1, GBP1 和 IFI27)过表达与患者生存时间显著降低密切相关($P < 0.05$).结论:共鉴别了 4 个与乳腺癌发病密切相关的关键基因,相关研究结果可为研究乳腺癌的诊疗提供新的潜在分子靶标.

关键词:乳腺癌;关键基因;PI3K

中图分类号:TP391

文献标志码:A

尽管现代医学在癌症研究方面已取得了巨大进步,但乳腺癌仍然是全球范围内女性健康的一大主要威胁^[1].研究表明,在年龄低于 45 岁的女性中,乳腺癌是癌症相关死亡的主要病因^[2].乳腺癌的常规治疗手段主要包括:手术切除、放疗、化疗、激素疗法和分子靶向疗法^[3].对于早期发现的乳腺癌,外科切除结合化疗等各种综合治疗手段可显著提高患者的生存率^[4].但由于该病早期症状不明显,容易被忽视,导致相当一部分患者被发现时已经处于晚期,癌细胞已经向其他器官转移,手术、化疗及激素等疗法预后较差,分子靶向疗法因具有靶向性和非细胞毒性,已成为乳腺癌治疗领域的研究热点.

虽然分子靶向疗法具有细胞毒性小、副作用少和复发率低等优势,但乳腺癌发病机制复杂,不同分子亚型疾病的治疗其靶点也不一样,确定合适的靶点是该病靶向治疗的一个重要前提^[5].为探索该病治疗的合适靶点,研究人员开展了大量研究工作.离体实验表明靶向细胞毒性的小干扰 RNA(Small interfering RNA, siRNA)与人表皮生长因子受体 3(human epidermal growth factor receptor 3, HER3)可发生高度特异性结合,有效抑制 HER3 阳性乳腺癌细胞的增殖^[6].文献^[7]的研究显示蛋白激酶 R(Protein kinase R, PKR)和磷脂酰肌醇 4 激酶 2 alpha(Phosphatidylinositol 4-kinase 2-alpha, PI4K2A)基因表达对乳腺癌患者的预后具有重要的提示作用,靶向 PI4K2A / PKR 的溶酶体复合物可能是一种治疗该种癌症的有效方法.文献^[8]发现脂肪酸转运蛋白 1(Fatty acid transport protein 1, FATP1)对乳腺癌细胞与非癌细胞之间脂肪酸的转运起关键作用,抑制 FATP1 的活性可抑制肿瘤细胞获取脂肪酸的能力并导致其细胞活力降低,FATP1 因而可能是治疗乳腺癌的一个靶标.尽管已经发现多种乳腺癌治疗的潜在分子靶标,但鉴于该病的发病机制复杂,疾病亚型多,有关治疗该病的高效的分子靶标并不多见.寻找癌症潜在分子靶标的方法比较多,既有传统的实验方法,又有基于机器学习的特征选择方法^[9-14].作为一种生命科学与信息科学交叉的研究方法,生物信息

收稿日期:2019-10-17;修回日期:2020-02-28.

基金项目:国家自然科学基金(61976082;61370169;61772176;61402153)

作者简介(通信作者):徐久成(1964—),男,河南偃师人,河南师范大学教授,博士,博士生导师,研究方向为粒计算、数据挖掘、粗糙集、生物信息学等,E-mail:xjc@htu.cn.

学方法在生命科学研究方面应用广泛并具有多方面的优势^[15-16],但其在探索乳腺癌关键基因方面的应用研究相对较少,本文利用生物信息学方法寻找乳腺发病的关键基因,为该病的分子靶向治疗提供理论依据。

1 材料与方法

1.1 数据选取

通过检索 GEO 数据库(Gene Expression Omnibus),选取乳腺癌基因表达谱数据:GSE71053,该基因表达谱数据集共包含 54 675 个探针所检测的基因表达数据,所检测的基因共计 22 185 个.该基因表达谱来源于丹麦欧登斯大学医院人体手术活检组织基因表达的检测结果,共包含 6 个乳腺肿瘤组织样本和 12 个正常乳腺组织样本。

1.2 基于 GEO2R 获取差异基因

基因芯片的设计目的是检测不同实验室条件下基因表达水平的差异,其基本原理是通过设定一定的阈值标准,比较实验组和对照组间基因表达信号的比值作为两种状态下的基因表达差异,结合统计显著性水平(P 值)获得差异表达的基因.美国国立卫生研究院创建的 NCBI 网站已经将数据预处理及差异分析的相关算法封装并整合到一个交互的在线图形用户界面差异基因分析工具 GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>)中去,该工具基于 R 语言的 Biobase, GEOquery, limma 工具包运作,可比较 GEO 数据库中不同实验条件下的两组样本基因表达情况,即使不懂相关原理,使用者也能准确对不同实验条件下数万个基因的表达水平进行差异显著性的分析。

本文利用 GEO2R 分别对上述基因表达数据集 GSE71053 中 12 个肿瘤样本和 6 个正常样本数据进行分析,所有数据分析参数采用默认设置,调整后的 P 值 $\text{adj.}P < 0.05$ 且差异基因调倍数的对数值 $|\log_2 \text{FC}| > 1$ 为差异表达基因(differentially expressed genes, DEGs)的筛选条件.利用 R 语言画出显著差异表达基因的火山图(Volcano plots)。

1.3 GO 与 KEGG 富集分析

基因本体(Gene Ontology, GO)数据库收集了对各个物种进行基因功能限定与描述的标准词汇,它是一个权威的基因功能分类描述系统,GO 富集分析常用于研究 DEGs 的生物学功能,主要包括生物过程(biological process)、细胞组成(cellular component)和分子功能(Molecular Function)3 个方面的内容。

GO 富集分析的统计学基础是超几何分布,具体来说就是对于具体某个 GO term,根据 Fisher 精确检验(Fisher exact test),计算一个 P 值:

$$P = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}},$$

其中, P 表示差异基因富集到某个 GO term 上的可信程度,当 $P < 0.05$ 或者 $P < 0.01$ 时,表示差异基因显著富集到该 GO term 上. N 表示每次研究中所有基因总数, n 代表 N 中差异基因的数量, k 指 n 中属于某个 GO term 的基因数. M 指 N 中属于特定 GO term 的基因数。

京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)是一个综合数据库,包含 16 个子数据库,其中就包含 KEGG 通路数据库,研究人员可利用该数据库分析 DEGs 显著富集的分子通路,KEGG 富集分析的原理与 GO 富集分析的原理相似.DAVID(The Database for Annotation, Visualization and Integrated Discovery)是一个强大在线 GO 和 KEGG 富集分析工具,本文利用 DAVID 对 DEGs 进行 GO 和 KEGG 通路富集分析, $P < 0.05$ 且基因 count(基因数)大于 10 作为显著富集分析的纳入标准。

1.4 构建 PPI 网络与关键基因的鉴别分析

DNA 编码区转录可形成 mRNA, mRNA 通过翻译形成蛋白,蛋白质-蛋白质之间相互作用(Protein-protein interaction, PPI)是实现其生理功能的重要途径,与多方面的生命活动密切相关,因此,研究 PPI 网络对了解 DEGs 的功能有重要启示作用.STRING(The search Tool for the Retrieval of Interacting Genes)数据库是构建 PPI 网络一个重要途径,本研究将所有 DEGs 输入到 STRING,采用默认参数构建 PPI 网络,

PPI 网络信息进一步导入到 Cytoscape 软件,利用 Cytoscape 内置的 Cytohubba 插件对网络进行拓扑分析,基于最大团中心性(maximal clique centrality, MCC)算法鉴别与乳腺癌发病相关的关键基因。

1.5 生存分析

生存分析是一种研究影响因素与生存时间和生存结局之间关系的统计分析方法,通常采用一个二维坐标图来可视化生存时间的概率分布情况,坐标轴的横轴代表时间长度,而纵轴代表某事件发生概率,描述生存时间分布的工具具有两个,分别是生存函数与风险函数,若用一个随机变量 $t(t \geq 0)$ 表示生存时间,则生存函数可定义为:

$$S(t) = Pr(T > t),$$

t 代表时间, $S(t)$ 代表生存函数, Pr 代表生存概率,随机变量 T 表示研究对象的实际生存时间,风险函数 $R(t)$ 则可定义为:

$$R(t) = 1 - S(t),$$

$R(t)$ 表示研究对象在 $T > t$ 时,感兴趣的终点事件发生的概率。

作为一种将事件结果与出现该结果所经历时间相结合进行统计分析的方法,生存分析可用于探索某个基因对患者的生存时间是否有显著影响,本课题涉及的乳腺癌关键基因生存分析主要通过 Kaplan Meier plotter(<http://kmplot.com/>)在线工具完成,该工具可在分析某个基因表达量高低对患者生存时间是否有显著影响的同时,并对分析结果画图。

2 结 果

2.1 乳腺癌差异表达基因

利用 GEO2R 在线差异分析工具,通过对基因表达谱 GSE71053 中肿瘤样本与正常样本基因表达情况进行差异表达分析,共得到 491 个 DEGs,其中有 254 个上调 DEGs,237 个下调 DEGs,DEGs 分布情况用火山图可视化(图 1),横坐标代表显著性 P 值,纵坐标代表基因表达变化倍数,图 1 中的每个点代表一个基因,绿色的点代表表达下调的基因,红色的点代表表达上调的基因,黑色的点表示表达无显著变化的基因。

2.2 GO 与 KEGG 富集分析

利用 DAVID 对 DEGs 进行 GO 与 KEGG 富集分析, $P < 0.05$ 且基因 count > 10

作为显著富集的选取标准,并选取富集显著性排名前 10 的 GO 词汇和 KEGG 通路作图,富集分析结果显示,DEGs 的 GO 生物学过程主要富集在细胞分裂,有丝分裂核分裂,细胞外基质组织,姊妹染色单体黏附,凋亡过程的负调控,细胞增殖的正调控,细胞黏附,细胞增殖,有丝分裂细胞周期的 G2/M 转变和 GTP 酶活性的正调节(图 2(a)).细胞组分主要富集于中间体、蛋白细胞外基质,细胞外基质,细胞质,纺锤体,微管,浓缩染色体动粒,胶原三聚体,胞核和中心体(图 2(b)).分子功能方面主要富集于肌动蛋白结合,钙离子结合,转录共抑制物活性,染色质结合,同源蛋白质结合,微管结合,蛋白质结合,肝素结合,蛋白激酶结合,整联蛋白结合(图 2(c)).KEGG 通路富集分析主要集中于 PI3K-Akt 信号通路、黏附斑通路和癌症通路(图 2(d)).

2.3 PPI 网络的构建

本文利用 STRING 数据库构建蛋白质-蛋白质互作网络,互作网络中边的信息来源主要包含实验数据、基因融合、文本挖掘数据、数据库数据和基因共表达信息等,网络中的每个点代表一种蛋白,点之间存在连线代表两个蛋白之间有互作关系,采用默认参数构建蛋白质-蛋白质互作网络(图 3).利用乳腺癌 DEGs 所构建

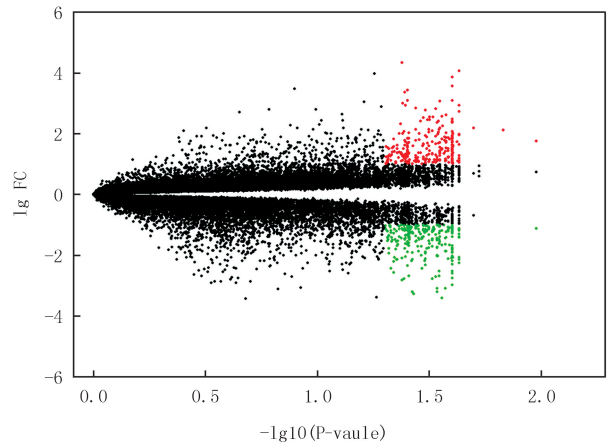
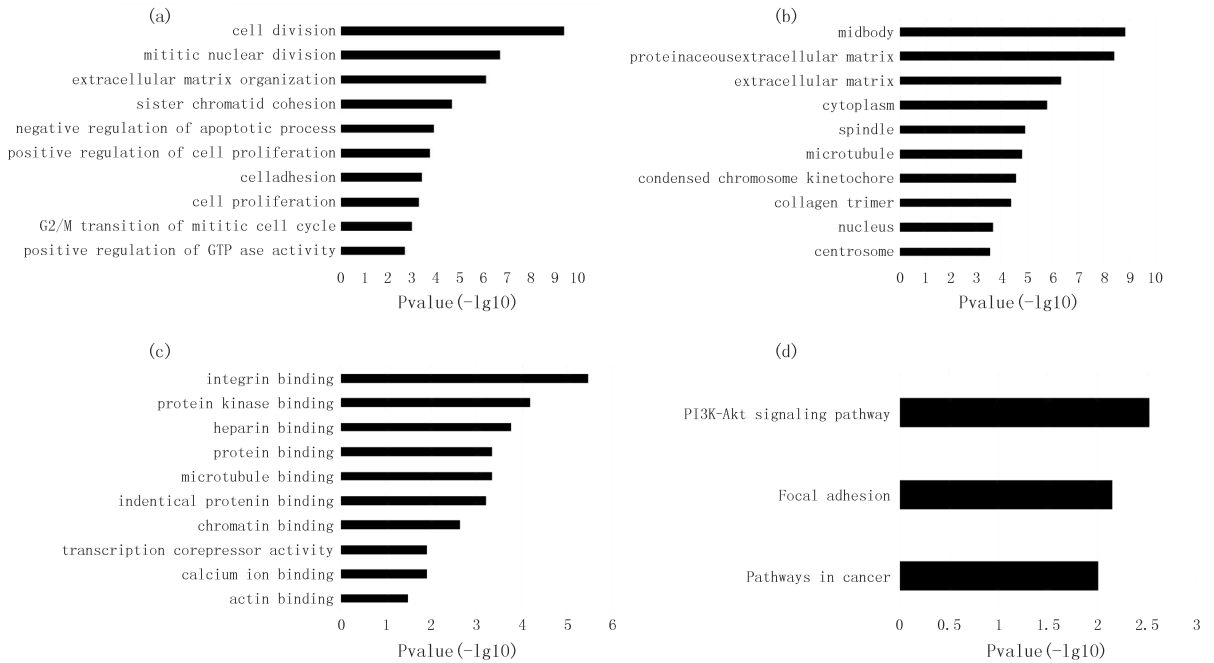


图1 乳腺癌差异表达基因的火山图

Fig.1 Volcano plot of breast cancer differentially expressed genes

PPI 网络总节点数为 344, 边数为 3 303, 平均节点度为 12.8, 局部聚类系数为 0.454, PPI 富集 P 值 $< 1.0e-16$.



(a) 代表GO生物学过程富集分析结果; (b) 代表GO细胞学组分富集分析结果; (c) 代表GO分子功能富集分析结果; (d) 代表KEGG通路富集分析结果。

图2 GO和KEGG通路富集分析结果

Fig.2 GO and KEGG pathway enrichment analysis results

2.4 关键基因的鉴别分析

将蛋白互作网络导入到 Cytoscape 软件, 利用 Cytoscape 内置的 Cytohubba 插件分析鉴别网络中的关键基因, 该插件提供了 11 种算法用于网络拓扑分析, 与其他算法比较, MCC 算法稳定性好, 结果更可靠, 已被广泛应用于关键节点的鉴别分析。在本研究中, 基于 MCC 算法共选取 10 个关键基因, 按照结点度值排序, 这 10 个的关键基因分别是: OAS1, OAS2, MX1, ISG15, IFIT1, RSAD2, IFI27, IRF9, GBP1 和 IFI44 (图 4)。

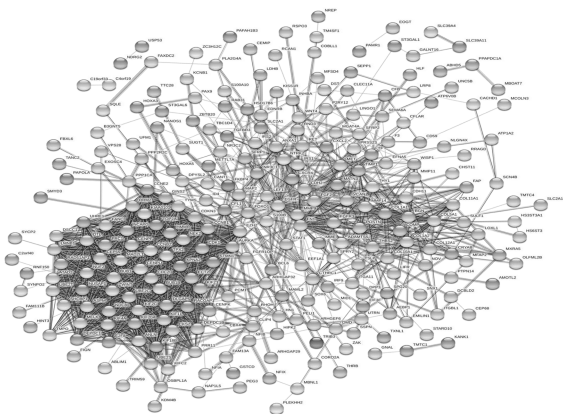


图3 基于STRING构建差异基因蛋白互作网络

Fig.3 A protein-protein interaction network constructed based on STRING database

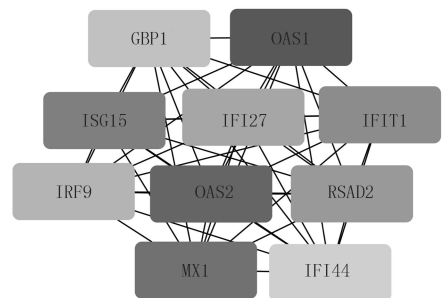


图4 基于MCC算法鉴别的关键基因及其PPI网络

Fig.4 The hub genes identified with MCC algorithm and relevant PPI network

2.5 生存分析

Kaplan Meier 主要基于荟萃分析发现和验证生存相关生物标记物, 目前利用该数据库可对 21 种肿瘤相关的 54 000 多个基因探针开展生存结局的分析。本研究利用 Kaplan Meier 数据库对所选取的排名前 10 位的关键基因进行生存分析, 结果表明 6 个关键基因过表达不能导致乳腺癌患者的生存时间显著缩短 ($P > 0.05$), 与之相对应的是, 另外 4 个基因过表达与患者生存时间显著降低密切相关 ($P < 0.05$), 这 4 个基因分

别是:ISG15,IFIT1,GBP1 和 IFI27(图 5)。

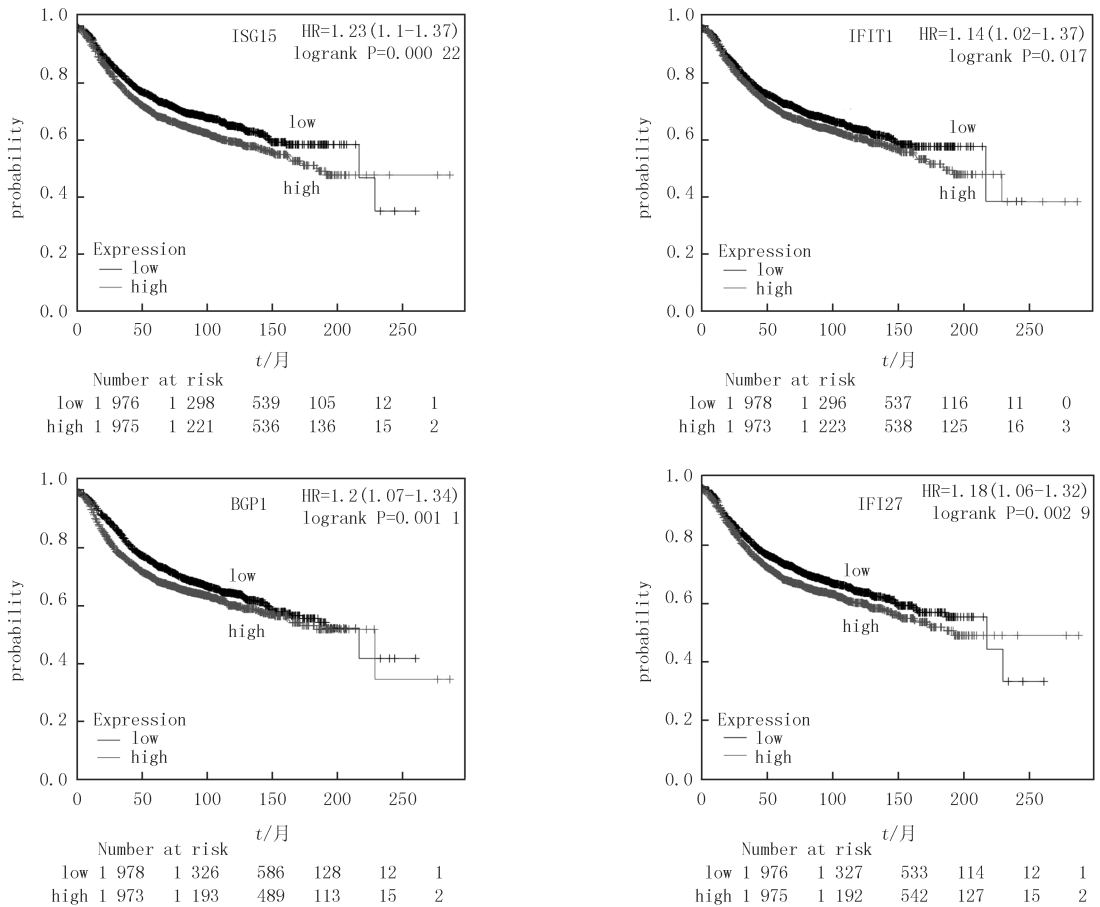


图5 与乳腺癌患者生存时间显著降低相关关键基因的生存曲线

Fig.5 Survival plots of hub genes associated with significantly reduced survival times of breast cancer patients

3 讨论

癌症是人类健康的一大威胁,临床常见癌症包括肺癌、乳腺癌、肝癌和食管癌等等^[17-20],乳腺癌是全球范围内女性癌症相关死亡的重要病因之一^[20]。近年来,在许多国家该病发病率和死亡率都呈上升的趋势,给患者、家庭及社会带来了沉重的负担。一项研究显示仅 2012 年全球就约有 170 万妇女罹患这种疾病^[21],尽管在某些类型的乳腺癌的治疗研究方面取得了重大进展,但作为一种异质性疾病,乳腺癌有几种不同的亚型,对于每种亚型的疾病,其临床和组织学特征各不相同,有关该病多方面的研究工作仍待开展。

在本文中,基于 GEO2R 在线工具对乳腺癌基因表达谱数据进行生物信息学分析,比较乳腺癌和正常乳腺组织的基因表达情况,进行差异显著性分析,共获得 491 个 DEGs,其中有 254 个上调 DEGs,237 个下调 DEGs。GO 富集分析表明这些 DEGs 差异显著富集于肿瘤相关的生物学过程、细胞组分和分子功能,表明这些差异基因所发挥的作用与这些 DGEs 的主要功能是密切相关的。KEGG 显著富集于 PI3K-Akt 信号通路、黏附斑和癌症通路。ZHOU 等人的研究表明,SPC24 基因的表达产物通过 PI3K / AKT 激酶途径引起乳腺癌的发病^[22],上述实验研究结果与本文研究结果乳腺癌最显著富集的 PI3K-Akt 信号通路相一致,本文研究结论得到了 ZHOU 等人^[22]实验证据方面的支持。

STRING 数据库是目前广泛接受的构建蛋白互作网络数据库,本文通过该数据库构建了蛋白互作网络,并将蛋白互作网络导入到 Cytoscape 软件,基于 MCC 算法获得排名前 10 位的关键基因,生存分析表明 ISG15,IFIT1,GBP1 和 IFI27 过表达与乳腺癌患者的生存时间显著下降密切相关。文献^[23]研究发现游离的 ISG15 可能通过差异调节雌激素受体- α 阳性表达和阴性表达乳腺癌细胞的致瘤特性,在乳腺肿瘤发病过

程中发挥重要作用.文献[20]的研究表明雌二醇和他莫昔芬治疗对乳腺癌的治疗效应可能是通过 ISG15 基因实现的,上述研究结果为 ISG15 作为乳腺癌治疗潜在靶点提供了更多的依据.胃癌相关研究表明,IFIT1 在胃癌中存在过表达现象^[24],但该基因在乳腺癌中过表达是否对患者生存时间有显著影响,未见相关报道.文献[25]的研究表明乳腺癌病人 STAT1 mRNA 过表达与预后不良有关,这与本文的研究结果一致.鸟苷酸结合蛋白 1(GBP1)是一种 T 淋巴细胞诱导表达的蛋白,可使乳腺癌细胞通过血脑屏障(BBB),T 淋巴细胞与乳腺癌细胞共同培养实验研究表明 GBP1 基因表达上调,可增强乳腺癌细胞通过 BBB 而发生脑转移的能力^[26],提示 GBP1 可作为一个抑制乳腺癌发生脑转移的靶点.目前,已有文献研究表明乳腺癌细胞中 IFI27 存在过表达现象^[27],但 IFI27 是否在乳腺癌发病过程中发挥重要作用,未见相关报道.

综上,本研究利用生物信息学方法鉴别乳腺癌发病相关的 4 个关键基因,这些基因过表达与患者生存时间显著减少密切相关,这一发现将为肿瘤的治疗研究提供了新的潜在靶点.

参 考 文 献

- [1] OLIVARES-URBANO M A, GRINÁN-LISÓN C, ZURITA M, et al. Matrix metalloproteases and TIMPs as prognostic biomarkers in breast cancer patients treated with radiotherapy: A pilot study[J]. *Journal of Cellular and Molecular Medicine*, 2020, 24(1): 139-148.
- [2] ANASTASIADI Z, LIANOS G D, IGNATIADOU E, et al. Breast cancer in young women: an overview[J]. *Updates in Surgery*, 2017, 69(3): 313-317.
- [3] DIABY V, TAWK R, SANOGO V, et al. A review of systematic reviews of the cost-effectiveness of hormone therapy, chemotherapy, and targeted therapy for breast cancer[J]. *Breast Cancer Research and Treatment*, 2015, 151(1): 27-40.
- [4] GUO R, SI J, XUE J, et al. Changing patterns and survival improvements of young breast cancer in China and SEER database, 1999-2017[J]. *Chinese Journal of Cancer Research*, 2019, 31(4): 653-662.
- [5] ESKANDARI E, MOTALEBZADEH J. Transcriptomics-based screening of molecular signatures associated with patients overall survival and their key regulators in subtypes of breast cancer[J]. *Cancer Genetics and Cytogenetics*, 2019, 239: 62-74.
- [6] NACHREINER I, HUSSAIN A F, WULLNER U, et al. Elimination of HER3-expressing breast cancer cells using aptamer-siRNA chimeras[J]. *Experimental and Therapeutic Medicine*, 2019, 18(4): 2401-2412.
- [7] PATAER A, OZPOLAT B, SHAO R, et al. Therapeutic targeting of the PI4K2A/PKR lysosome network is critical for misfolded protein clearance and survival in cancer cells[J]. *Oncogene*, 2020, 39(4): 801-813.
- [8] MENDES C, LOPES-COELHO F, RAMOS C, et al. Unraveling FATP1, regulated by ER- β , as a targeted breast cancer innovative therapy[J]. *Scientific Reports*, 2019, 9(1): 14107.
- [9] XU J C, MU H Y, WANG Y, et al. Feature genes selection using supervised locally linear embedding and correlation coefficient for microarray classification[J]. *Computational and Mathematical Methods in Medicine*, 2018, 2018: 1-11.
- [10] XU J C, WANG Y, MU H Y, et al. Feature genes selection based on fuzzy neighborhood conditional entropy[J]. *Journal of Intelligent & Fuzzy Systems*, 2019, 36(1): 117-126.
- [11] XU J C, WANG Y, XU K Q, et al. Feature genes selection using fuzzy rough uncertainty metric for tumor diagnosis[J]. *Computational and Mathematical Methods in Medicine*, 2019, 2019: 1-9.
- [12] 徐久成, 冯森, 穆辉宇. 基于信噪比与随机森林的肿瘤特征基因选择[J]. *河南师范大学学报(自然科学版)*, 2017, 45(2): 87-92.
XU J C, FENG S, MU H Y. Tumor feature gene selection based on signal-to-noise ratio and random forest[J]. *Journal of Henan Normal University(Natural Science Edition)*, 2017, 45(2): 87-92.
- [13] 徐久成, 李涛, 孙林, 等. 基于信噪比与邻域粗糙集的特征基因选择方法[J]. *数据采集与处理*, 2015, 30(5): 973-981.
XU J C, LI T, SUN L, et al. Feature gene selection method based on signal-to-noise ratio and neighborhood rough set[J]. *Data Acquisition and Processing*, 2015, 30(5): 973-981.
- [14] 徐久成, 穆辉宇, 冯森. 基于 PCA 和多邻域粗糙集的肿瘤特征基因选择算法[J]. *郑州大学学报(理学版)*, 2017, 49(4): 28-33.
XU J C, MU H Y, FENG S. A tumor feature gene selection algorithm based on PCA and multiple neighborhood rough sets[J]. *Journal of Zhengzhou University(Science Edition)*, 2017, 49(4): 28-33.
- [15] 俱雄, 党亚锋, 刘祥, 等. 大肠杆菌外膜蛋白 OmpC 的生物信息学分析及表位多肽疫苗的重组预测[J]. *河南农业大学学报*, 2017, 51(1): 94-100.
JU X, DANG Y F, LIU X, et al. Bioinformatics analysis of E. coli outer membrane protein OmpC and prediction of recombinant peptide epitope vaccine[J]. *Journal of Henan Agricultural University*, 2017, 51(1): 94-100.
- [16] 李文清, 李中举, 孟志鹏, 等. 牛 IRS1 基因 3' UTR 的克隆、鉴定及生物信息学分析[J]. *河南农业大学学报*, 2017, 51(1): 48-52.
LI W Q, LI Z J, MENG Z P, et al. Cloning, identification and bioinformatics analysis of bovine IRS1 gene 3' UTR[J]. *Journal of Henan Agricultural University*, 2017, 51(1): 48-52.

- [17] 郭康,范文艳.突变型 P53 和肿瘤易感基因 101 在肺癌组织中的表达及意义[J].河南师范大学学报(自然科学版),2015,43(3):138-142.
GUO K,FAN W Y.Expression and Significance of Mutant P53 and Tumor Susceptible Gene 101 in Lung Cancer[J].Journal of Henan Normal University(Natural Science Edition),2015,43(3):138-142.
- [18] 王雷,王煜霞,陈丽平,等.P-gp,MRP,LRP 和 GST- π 等多药耐药蛋白在肝癌中的协同表达研究[J].河南师范大学学报(自然科学版),2015,43(4):134-137.
WANG L,WANG Y X,CHEN L P,et al.Co-expression of multidrug resistance proteins such as P-gp,MRP,LRP and GST- π in liver cancer[J].Journal of Henan Normal University(Natural Science Edition),2015,43(4):134-137.
- [19] 秦建平,陶娟,李迎迎,等.血清 miR-25、miR-223 和 miR-373 作为食管鳞癌生物标志物的评价[J].河南师范大学学报(自然科学版),2017,45(1):65-70.
QIN J P,TAO J,LI Y Y,et al.Evaluation of serum miR-25,miR-223 and miR-373 as biomarkers for esophageal squamous cell carcinoma [J].Journal of Henan Normal University(Natural Science Edition),2017,45(1):65-70.
- [20] ALSHABI A M,SHAIKH I A,VASTRAD C.Exploring the Molecular Mechanism of the Drug-Treated Breast Cancer Based on Gene Expression Microarray[J].Biomolecules,2019,9(7):282.
- [21] WOOLSTON C.Breast cancer[J].Nature,2015,527(7578):S101.
- [22] ZHOU J,PEI Y,CHEN G,et al.SPC24 Regulates breast cancer progression by PI3K/AKT signaling[J].Gene,2018,675(30):272-277.
- [23] TECALCOCRUZ A C,CRUZRAMOS E.Protein ISGylation and free ISG15 levels are increased by interferon gamma in breast cancer cells[J].Biochemical and Biophysical Research Communications,2018,499(4):973-978.
- [22] SHIOZAKI A,ARIYOSHI Y,IITAKA D,et al.Functional analysis and clinical significance of sodium iodide symporter expression in gastric cancer[J].Gastric Cancer,2019,22(3):473-485.
- [25] TYMOSZUK P,CHAROENTONG P,HACKL H,et al.High STAT1 mRNA levels but not its tyrosine phosphorylation are associated with macrophage infiltration and bad prognosis in breast cancer[J].BMC Cancer,2014,14(1):257.
- [26] MUSTAFA D,PEDROSA R,SMID M,et al.T lymphocytes facilitate brain metastasis of breast cancer by inducing Guanylate-Binding Protein 1 expression[J].Acta Neuropathologica,2018,135(4):581-599.
- [27] LI S,XIE Y,ZHANG W,et al.Interferon alpha-inducible protein 27 promotes epithelial-mesenchymal transition and induces ovarian tumorigenicity and stemness[J].Journal of Surgical Research,2015,193(1):255-264.

Identification of key genes in breast cancer of prognostic value

Xu Jiucheng^{1a,b},Li Chengzhang^{1a,2}

(1.a.State Key Laboratory Cultivation Base for Cell Differentiation Regulation,College of Life Science;

b.Engineering Lab of Intelligence Business & Internet of Things, Henan Province, Henan Normal University, Xinxiang 453007, China;

2.Department of Physiology and Neurobiology, School of Basic Medicine, Xinxiang Medical University, Xinxiang 453003, China)

Abstract: Objective: To identify key genes and pathways involved in the pathogenesis of breast cancer using bioinformatics methods. **Methods:** The GEO2R online tool was used to analyze the gene expression profiles of breast cancer. The differentially expressed genes (DEGs) was annotated by GO and KEGG enrichment analysis. The STRING database was used to construct a protein-protein interaction (PPI) network with the differentially expressed genes. The PPI network data was visualized using Cytoscape software. The key genes of the PPI network were identified based on MCC algorithm. The selected key genes were then subject to Kaplan Meier survival analysis to test the role of hub genes on survival time of breast cancer patients. **Results:** Differentially expressed analysis resulted in a total of 491 DEGs, of which 254 were up-regulated and 237 were down-regulated. GO enrichment analysis showed that these DEGs were significantly enriched in tumor-related biological processes, cellular components and molecular functions, and KEGG pathway enrichment analysis focused on PI3K-Akt signaling pathway, adhesion plaques and cancer pathways. Based on the MCC algorithm, a total of 10 key genes were selected and only four of them has been shown to be associated with significantly decreased survival time of the breast cancer patients ($P < 0.05$). The four key genes were ISG15, IFIT1, GBP1 and IFI27 respectively. **Conclusion:** This study results in identification four key genes contributing to the development of breast cancer. The discovery of these key genes provides new clues for studying the pathogenesis of breast cancer as well as potential molecular targets for treatment of the disease.

Keywords: breast cancer; hub genes; PI3K

[责任编辑 陈留院 赵晓华]



本期专家介绍



杨俊元, 山西大学复杂系统研究所教授, 博士, 硕士生导师, 山西省优秀青年学术带头人和山西省“131”领军人才. 曾先后在加拿大劳瑞尔大学做博士后研究(2012.1—2013.1)、在美国佛罗里达大学数学系做访问学者(2014.5—2015.4). 主要从事传染病动力学、复杂网络及博弈论研究. 主持国家自然科学基金 3 项、省部级项目 4 项, 先后在 *J Biol Math*, *Math Biosci*, *Nonlinear Anal: RWA*, *Chaos* 等杂志上发表 30 余篇论文. 在 Springer 出版专著 1 部.

徐久成, 河南师范大学二级教授, 博士, 博士生导师, 全国优秀教师. 现为教育部高等学校计算机类专业教学指导委员会委员(2018—2022), 河南省高校计算机教育研究会副理事长, 计算机科学与技术省一级重点学科带头人, 智慧商务与物联网技术河南省工程实验室主任, 计算智能与数据挖掘河南省高校工程技术研究中心主任. IEEE 国际粒计算委员会程序委员, 中国粗糙集与软计算专业委员会常务委员. 主要从事人工智能、粒计算与知识获取、生物信息学等领域的理论及应用研究. 主持国家自然科学基金项目 4 项、河南省高校新世纪优秀人才计划重大项目 1 项. 发表学术论文 130 余篇, 入选 ESI 高被引论文 2 篇. 出版专著 2 部, 主编国家规划统编教材 3 部. 获河南省高等教育省级教学成果一等奖 3 项, 作为主要参与者获国家教学成果二等奖 2 项.



杨宗献, 河南师范大学二级教授, 博士生导师, 河南省特聘教授, 河南师范大学物理学院院长, 国家级物理实验教学示范中心主任, 河南省物理学会副理事长. 留学美国、加拿大、瑞典博士后和客座教授. 主要研究领域为凝聚态物理、计算纳米科学与清洁能源材料设计. 主持国家自然科学基金项目 5 项、国际合作项目 1 项, 在 SCI 源期刊发表学术论文 160 篇. 获全国优秀科技工作者、河南省杰出人才、河南省政府津贴等奖励和荣誉.