

乳腺导管原位癌浸润性进展 相关基因的生物信息学研究

周军 麻怀露 丁晓飞 梁勇 陈光

【摘要】 目的 筛选介导乳腺导管原位癌(DCIS)浸润性进展的相关基因。方法 从美国国立生物技术信息中心(NCBI)公共基因芯片(GEO)数据库下载含正常乳腺(NC)、DCIS和微小浸润乳腺癌(IBC)这3种类型样本的基因芯片;使用R软件包limma分析IBC与DCIS、DCIS与NC之间的差异表达基因;使用STEM软件对上述两组差异表达基因进行趋势分析,筛选出随着NC-DCIS-IBC的趋势表达水平逐渐变化的基因;对筛选出的差异表达的基因进行KEGG信号通路富集分析。结果 通过对差异基因的趋势分析,筛选出12个随着NC-DCIS-IBC的趋势表达显著下调的基因,以及10个显著上调的基因,其中运动神经元和胰腺同源盒1(MNX1)、基质金属蛋白酶1(MMP-1)可能为介导DCIS浸润性进展的相关基因。信号通路富集分析结果显示,磷脂酰肌醇3-激酶/蛋白激酶B(PI3K/AKT)、黏附和细胞外基质/受体相互作用3个信号通路分子异常在DCIS发生及浸润性进展中均存在显著异常调控。结论 MNX1、MMP-1可能为介导DCIS浸润性进展的相关基因,PI3K/AKT信号通路可能参与介导MNX1/MMP-1促进DCIS浸润性进展。

【关键词】 乳腺癌 乳腺导管原位癌 浸润性癌 生物信息分析

Bioinformatics Analysis of invasive progression-related genes in breast ductal carcinoma in situ ZHOU Jun, MA Huailu, DING Xiaofei, et al. Department of Basic Medical Science, Medical School of Taizhou University, Taizhou 318000, China

【Abstract】 Objective To identify the key genes involved in the invasive progression of breast ductal carcinoma in situ (DCIS) by bioinformatics analysis. Methods Gene chips containing samples from normal mammary gland, breast DCIS and small invasive breast cancer (IBC) were downloaded from GEO database. The differentially expressed genes in DCIS compared to normal tissue samples, and in IBC compared to DCIS were analyzed by R language. Trend analysis of differentially expressed genes in the two groups was carried out by Short Time-series Expression Miner (STEM) software. The differentially expressed genes gradually changed with the trend of DCIS progression were screened out. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed on those differentially expressed genes. Results Through screening of differentially expressed genes, 12 down-regulated genes and 10 up-regulated genes were screened with the invasive progression of DCIS, including MNX1 and MMP-1. Those differentially expressed genes in the KEGG pathway were significantly enriched in the PI3K/AKT, cytokine-cytokine receptor interaction pathway and cell adhesion molecules. Conclusion MNX1 and MMP-1 might mediate invasive progression of breast DCIS, and PI3K/AKT signaling may be involved in MNX1/MMP-1 regulation.

【Key words】 Breast cancer DCIS IBC Bioinformatics

乳腺癌是女性最常见的恶性肿瘤之一^[1]。国家癌症中心《2018年全国最新癌症报告》显示,2014年我国女性乳腺癌新发病例约27.89万例,占女性恶性肿瘤发病的16.51%;死亡病例约6.60万例,占女性恶性肿瘤死亡的7.82%。乳腺导管原位癌(DCIS)主要发生在乳腺末梢导管小叶单位,然后进展延伸到乳腺导管小叶

系统^[2]。随着医学影像学的发展,我国DCIS的检出率日益上升,达到了7.8%~18.8%,约占乳腺癌新发病例的20%~30%^[3]。因此临床上对于DCIS发生、发展机制的研究显得极为重要。本研究从美国国立生物技术信息中心(NCBI)公共基因芯片(gene expression omnibus, GEO)数据库中下载包含乳腺癌的基因芯片,对其进行差异表达基因的筛选和趋势分析,探讨可能与DCIS浸润性进展相关的基因,以期为临床提供参考,现报道如下。

1 材料和方法

1.1 数据收集 从NCBI GEO数据库(<https://www.ncbi>.)

DOI:10.12056/j.issn.1006-2785.2019.41.12.2019-841

基金项目:国家自然科学基金资助项目(81802657)

作者单位:318000 台州学院医学院药理学

通信作者:陈光,E-mail:gcchen@tzc.edu.cn

nlm.nih.gov/geo/)中选取包含乳腺癌样品的基因芯片^[4],包含 3 个正常乳腺(NC)样本,46 个 DCIS 样本,以及 56 个微小浸润乳腺癌(IBC)样本。使用 R 软件包 limma (Version 3.30.13, <http://www.bioconductor.org/packages/2.9/bioc/html/limma.html>)读取基因芯片数据,并对数据进行背景矫正、归一化、表达水平计算。若多个探针对

应同一个基因,取探针均值作为最后的基因表达水平。所有的样本信息均来自于美国圣路易斯华盛顿大学 Robert Lesurf 教授登记于 GEO 数据库的数据,基因芯片系列 GSE59248,基因芯片平台为 GPL13607。纳入患者的基本信息见表 1。

表 1 纳入患者的基本信息(例)

基本信息	DCIS 患者(n=46)	IBC 患者(n=56)	NC 患者(n=3)
年龄(中位数,岁)	57	57.6	55
肿瘤分期(I/II/III)	3/15/35	11/19/22	-
雌激素受体阳性/雌激素受体阴性(ER ⁺ /ER ⁻)	26/11	38/18	-
孕激素受体阳性/孕激素受体阴性(PR ⁺ /PR ⁻)	23/14	34/22	-
表皮生长因子受体 2 阳性/表皮生长因子受体 2 阴性(Her2 ⁺ /Her2 ⁻)	12/22	10/38	-
淋巴结转移阳性/阴性	0/46	13/43	-
A 型管腔癌(Lum A)	17	19	-
B 型管腔癌(Lum B)	10	17	-
基底细胞样乳腺癌	8	8	-

1.2 方法

1.2.1 差异表达基因分析 使用 R 软件包 limma 提供的经典贝叶斯方法,对样本进行差异表达基因分析^[5]。差异比较分为两组,分别是 IBC 与 DCIS、DCIS 与 NC。将 $P < 0.05$, $|\log_{2}FC| \geq 0.585$ 的基因视为差异表达基因;并对所得差异表达基因进行主成分分析。

1.2.2 差异基因表达趋势分析 针对上述筛选出的两组差异表达基因,使用 STEM 软件(version 1.3.11, <http://www.cs.cmu.edu/~jernst/stem/>)作趋势分析^[6],筛选其中表达水平随着 NC-DCIS-IBC 的趋势逐渐变化的基因,得到不同趋势的基因簇。模型配置文件的最大数目设置为 9, P 值设置为 0.05。

1.2.3 差异表达基因的信号通路富集分析 使用 R 软件包的 clusterProfiler^[7]对差异表达基因进行京都基因和基因组百科全书(KEGG)信号通路富集分析。 $P < 0.05$ 视为显著富集结果。

2 结果

2.1 IBC 与 DCIS、DCIS 与 NC 差异表达基因筛选结果 IBC 与 DCIS 比较筛选出 1 163 个差异表达基因,其中表达水平显著上调的基因 485 个,包括细胞外基质金属蛋白酶家族成员 MMP-11、MMP-13、MMP-1、MMP-3、MMP-3、MMP-2、MMP-14,磷脂酰肌醇 3-激酶/蛋白激酶 B (PI3K/AKT) 趋化因子及其受体家族成员 CX-CL9、CXCL13、CXCL11、CXCR2P1、CXCR4, PI3K/AKT/mTOR 信号通路分子 COL1A1、COL1A2、FN1、THBS2

等;表达水平显著下调的基因 678 个,包括细胞外基质金属蛋白酶家族成员 MMP-24、MMP-12, PI3K/AKT 趋化因子及其受体家族成员 CXCL6、CXCL1、CXCL17、CXCR2, PI3K/AKT/mTOR 信号通路分子 LAMA3、FGF-10 等。

DCIS 与 NC 比较筛选出 3 439 个差异表达基因,其中表达水平显著上调的基因 1 435 个,包括与细胞外基质金属蛋白酶家族成员 MMP-1、MMP-10、MMP-11, PI3K/AKT/mTOR 信号通路分子 COL6A6、GNG2、VEGFB, 细胞色素 C P450 家族成员 CYP4Z1 和 CYP4Z2P;表达水平显著下调的基因 2 004 个,包括 NMT1、ALKBH5 和 ADAMTS2 等。

差异表达基因热图见图 1(插页),显示 3 组样本的基因表达情况,可发现差异表达基因概况。主成分分析双标图见图 2,显示 DCIS 与 IBS 间有微量重叠,说明 DCIS 与 IBS 之间虽有共同之处,但是随趋势变化的基因数量较少。

2.2 差异表达基因随 DCIS 浸润性进展的趋势分析 趋势分析结果显示随着 NC-DCIS-IBC 进程持续上调的基因共 10 个,包括三基序蛋白 49L1 (TRIM49L1)、运动神经元和胰腺同源盒 1 (MNX1)、基质金属蛋白酶 (MMP-1)、C 型凝集素家族 3 (CLEC3A)、同源盒基因 12 (HOXC12)、细胞色素 P450 (CYP4Z2P) 和未知功能基因 PP14571、LOC732272、LOC729722、LOC728606,见表 2。与之相对应的趋势下调基因有 12 个,见表 3。

2.3 差异表达基因的信号通路富集分析结果 信号

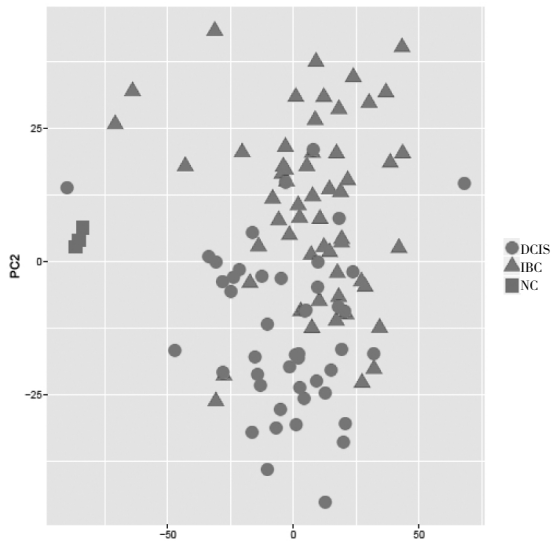


图 2 主成分分析双标图

表 2 随 DCIS 浸润性进展表达持续上调的基因

基因名称	表达差异{log ₂ [v(i)/v(0)]}	
	DCIS vs NC	IBC vs DCIS
TRIM49L1	1.21	1.05
MXN1	1.02	0.99
MMP1	0.81	1.13
CLEC3A	1.06	1.19
HOXC12	0.79	1.05
CYP4Z2P	1.24	1.08
PP14571	1	0.98
LOC732272	1.27	1.36
LOC729722	1.06	0.79
LOC728606	1.27	0.99

表 3 随 DCIS 浸润性进展表达持续下调的基因

基因名称	表达差异{log ₂ [v(i)/v(0)]}	
	DCIS vs NC	IBC vs DCIS
三合蛋白(TRDN)	-1.03	-1.22
腱调蛋白(TNMD)	-1.06	-1.19
R- 脊椎蛋白 1(RSP01)	-1	-1
碱性富含脯氨酸的泪蛋白(PROL1)	-1.08	-1.24
瘦素(LEP)	-0.89	-1.2
乳清蛋白 alpha(LALBA)	-1.98	-1.97
微丝角蛋白相关基因(KRTAP21-1)	-0.92	-1.02
热休克蛋白 B7(HSPB7)	-0.93	-1.04
酪蛋白 kappa 3(CSN3)	-1.45	-1.46
酪蛋白 kappa 2(CSN2)	-1.85	-1.77
酪蛋白 kappa 1S1(CSN1S1)	-1.31	-1.34
芳基乙胺脱乙酰酶基因(AADAC)	-0.91	-1.13

通路富集分析结果显示, PI3K/AKT 和细胞黏附等信号通路分子异常在 DCIS 发生及浸润性进展中均存在显著异常调控, 见图 3。

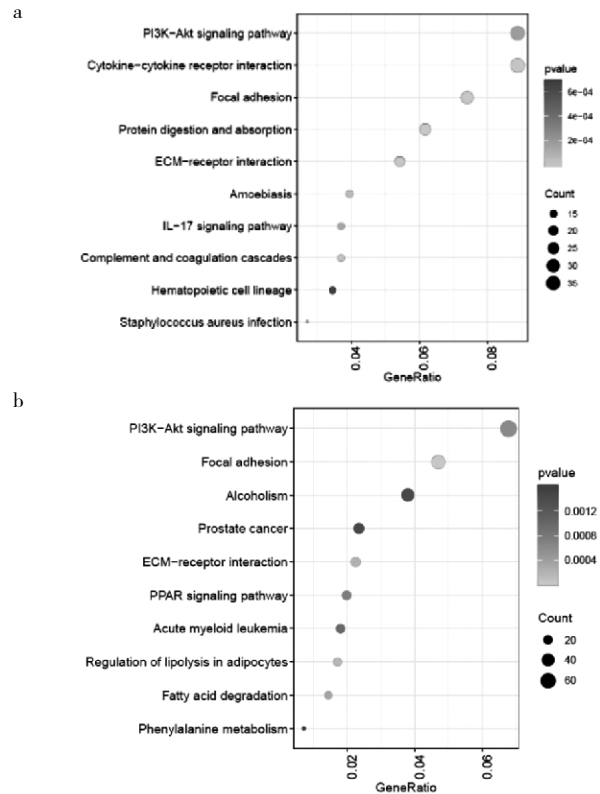


图 3 差异表达基因的信号通路富集分析结果 (a: DCIS vs NC; b: IBC vs DCIS)

3 讨论

本研究采用生物信息学研究方法整合分析了包括 NC、DCIS 与 IBC 3 类共 105 个临床样本的 GEO 数据库基因芯片 GSE59248, 发现包括 MNX1、MMP-1 在内的 10 个基因表达随着 NC-DCIS-IBC 进程呈明显的渐进性表达上调, 推测其可能在 DCIS 浸润性进展中发挥作用, 但仍需临床研究验证。乳腺上皮细胞恶性转变是一个多级过程, 包括增生、异型增生、DCIS 和 IBC 几个阶段进程。DCIS 是最常见、最易发展为 IBC 的乳腺“癌前病变”, 若不接受治疗, 约有 25% 会发展成 IBC^[8-9]。手术、放疗和辅助内分泌治疗为 DCIS 的主要治疗方法, 但存在以下问题: (1) 缺少可靠指标用以评定哪种类型 DCIS 易进展为 IBC; (2) 过度治疗。因此, 探寻 DCIS 恶化的预测性生物标记物为乳腺癌早期治疗的关键。

本研究趋势分析发现, MNX1 高表达可能为 DCIS 易进展为 IBC 的生物标志物。MNX1 基因位于人染色体 7q36.3, 全长 12 801bp, 包含 3 个外显子, 编码一种含有 403 个氨基酸的转录因子 MNX1^[10]。其蛋白结构中包含一个具有长度多态性的多聚丙氨酸重复区域、一个由外显子 2 和外显子 3 编码的同源结构域。另外, 在同源结构域上游含有一个由 82 个氨基酸组成的高度

保守区域。MNX1 蛋白具有类似螺旋-转角-螺旋结构的 3 个螺旋区域。和其他的含有同源结构域蛋白相比,最保守的部分是第 2 和第 3 螺旋,可以和 DNA 相结合。MNX1 蛋白的靶基因目前尚不清楚,和其他含有同源结构域的蛋白一样,其能够和含有 TATA 结构的 DNA 序列相互作用,并且可以通过结合位于 5' 端非编码区的 MNX1 启动子 TATA 结构来调节自身的表达^[11-12]。MNX1 基因最早因其在胰腺和淋巴组织中表达而被发现,为胰腺正常发育和脊髓运动神经元分化所必需^[13]。通过肿瘤基因组图谱(TCGA)数据库分析和组织 qRT-PCR 验证,MNX1 在乳腺癌中表达显著升高,并且 MNX1 与乳腺癌临床病理特征如 Her2 以及经期相关^[14-17]。与 MNX1 表达上调相伴,本研究差异表达基因的信号通路富集分析结果显示 PI3K/AKT 信号通路随 DCIS 浸润性进展进行性激活,提示 PI3K/AKT 信号通路可能参与 MNX1 表达调控。本研究结果也显示,MMP-1 可能为潜在的 DCIS 易进展为 IBC 的生物标志物。MMP-1 在乳腺癌和前列腺癌发生、发展中都扮演重要角色,且敲低 MMP-1 表达后,宫颈癌细胞的增殖、迁移和侵袭减弱,并且在表皮生长因子受体酪氨酸激酶抑制剂(EGFR-TKI)耐药肺腺癌的迁移和侵袭中也有起重要作用^[18]。

综上所述,本研究采用生物信息学研究方法筛选在乳腺癌中表达上调或下调显著的基因,通过趋势分析进一步筛选出了随着 NC-DCIS-IBC 趋势表达持续上调或者下调的基因,发现 MNX1、MMP-1 可能为介导 DCIS 浸润性进展的相关基因,PI3K/AKT 信号通路可能参与介导 MNX1/MMP-1 促进 DCIS 浸润性进展。但因本研究仅涉及生物信息学分析,以上潜在功能基因与乳腺癌的相关性及潜在的分子机制仍需临床及基础实验验证。

4 参考文献

- [1] Vojtek M, Marques MPM, Ferreira I, et al. Anticancer activity of palladium-based complexes against triple-negative breast cancer [J]. *Drug discovery today*, 2019, 24 (4): 1044-1058. DOI: 10.1016/j.drudis.2019.02.012.
- [2] 朱思吉,沈坤炜. 乳腺导管内癌的外科治疗进展[J]. *中华医学杂志*, 2014, 94 (48): 3865-3867. DOI: 10.3760/cma.j.issn.0376-2491.2014.48.019.
- [3] Lesurf R, Aure MR, Mork HH, et al. Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer[J]. *Cell Reports*, 2016, 16(4): 1166-1179. DOI: 10.1016/j.celrep.2016.06.051.
- [4] Law CW, Alhamdoosh M, Su S, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR[J]. *F1000 Research*, 2016 (5): 1408. doi: 10.12688/f1000research.9005.3.
- [5] Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data[J]. *BMC bioinformatics*, 2006 (7): 191. DOI: 10.1186/1471-2105-7-191.
- [6] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters [J]. *Omics: a journal of integrative biology*, 2012, 16 (5): 284-287. DOI: 10.1089/omi.2011.0118.
- [7] Groen EJ, Elshof LE, Visser LL, et al. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS) [J]. *Breast*, 2017, 31: 274-283. DOI: 10.1016/j.breast.2016.09.001.
- [8] Feinberg J, Wetstone R, Greenstein D, et al. Is DCIS Overrated? [J]. *Cancer Treat Res*, 2018, 173: 53-72. DOI: 10.1007/978-3-319-70197-4_5.
- [9] Zhang L, Wang J, Wang Y, et al. MNX1 Is Oncogenically Upregulated in African-American Prostate Cancer [J]. *Cancer Res*, 2016, 76 (21): 6290-6298. DOI: 10.1158/0008-5472.CAN-16-0087.
- [10] Leotta CG, Federico C, Brundo MV, et al. HLXB9 gene expression, and nuclear location during in vitro neuronal differentiation in the SK-N-BE neuroblastoma cell line[J]. *PLoS One*, 2014, 9 (8): e105481. DOI: 10.1371/journal.pone.0105481.
- [11] Flanagan SE, De Franco E, Lango Allen H, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man[J]. *Cell Metab*, 2014, 19 (1): 146-54. DOI: 10.1016/j.cmet.2013.11.021.
- [12] Wang Y, Wang J, Zhang L, et al. RGS12 Is a Novel Tumor-Suppressor Gene in African American Prostate Cancer That Represses AKT and MNX1 Expression[J]. *Cancer Res*, 2017, 77 (16): 4247-4257. DOI: 10.1158/0008-5472.CAN-17-0669.
- [13] Leotta CG, Federico C, Brundo MV, et al. HLXB9 gene expression, and nuclear location during in vitro neuronal differentiation in the SK-N-BE neuroblastoma cell line[J]. *PLoS One*, 2014, 9 (8): e105481. DOI: 10.1371/journal.pone.0105481. DOI: 10.1371/journal.pone.0105481.
- [14] Lv Y, Li H, Li F, et al. Long Noncoding RNA MNX1-AS1 Knock-down Inhibits Cell Proliferation and Migration in Ovarian Cancer [J]. *Cancer Biother Radiopharm*, 2017, 32 (3): 91-99. DOI: 10.1089/cbr.2017.2178.
- [15] Tian T, Wang M, Zhu Y, et al. Expression, Clinical Significance, and Functional Prediction of MNX1 in Breast Cancer[J]. *Mol Ther Nucleic Acids*, 2018, 13: 399-406. DOI: 10.1016/j.omtn.2018.09.014.
- [16] Cheng Y, Pan Y, Pan Y, et al. MNX1-AS1 is a functional oncogene that induces EMT and activates the AKT/mTOR pathway and MNX1 in breast cancer [J]. *Cancer Manag Res*, 2019, (11): 803-812. DOI: 10.2147/CMAR.S188007.
- [17] Hsiao CL, Liu LC, Shih TC, et al. The Association of Matrix Metalloproteinase-1 Promoter Polymorphisms with Breast Cancer [J]. *In vivo (Athens, Greece)*, 2018, 32 (3): 487-491. DOI: 10.21873/invi-v.11265.

(收稿日期:2019-03-20)

(本文编辑:李媚)