・论著・

乳腺癌骨转移相关的基因研究

贺功建1,杜正贵2

1. 广安市人民医院烧伤整形美容甲状腺乳腺外科(四川广安 638001)
2. 四川大学华西医院乳腺外科(成都 610041)

【摘要】目的 通过加权基因共表达网络分析 (weighted gene coexpression network analysis, WGCNA) 方法 找到与乳腺癌骨转移相关的关键基因,为寻找新的靶向治疗药物提供理论支持。方法 从基因表达综合数据库 下载 GSE2034 数据集中共 286 例乳腺癌患者的基本临床特征及肿瘤标本基因表达的信息。通过 R 软件实现基因 芯片的分析。采用 R 软件中植入的 WGCNA 包用于加权相关网络分析中的各种统计分析。采用 SPSS 软件进行 单因素 Cox 比例风险回归分析。结果 通过 WGCNA 方法选取了方差变异度最大的前 5 000 个基因做进一步的 基因富集分析,最终共构建了 15 个基因共表达模块,其中 magenta 模块与乳腺癌骨转移呈正相关关系 (*r*=0.94, *P*<0.001)。进一步分析发现,magenta 模块中与乳腺癌骨转移高度相关的 6 个关键基因为: Ral GTPase 激活蛋白 亚单位 α-1 (RALGAPA1)、B 细胞抗原受体复合物相关蛋白 α 链 (CD79A)、人免疫球蛋白 kappa 链 (IGKC)、抑制 蛋白 β2 (ARRB2)、FDCP 6 同源异构体 (DEF6)和人免疫球蛋白 lambda 链 (IGLV2)。结论 RALGAPA1、 CD79A、IGKC、ARRB2、DEF6 和 IGLV2 可能在乳腺癌骨转移中起到重要作用。

【关键词】 乳腺癌;骨转移;加权基因共表达网络分析;关键基因

Study on the gene related to bone metastasis of breast cancer

HE Gongjian¹, DU Zhenggui²

Department of Breast, Thyroid, and Burn Disease, Guang'an People's Hospital, Guang'an, Sichuan 638001, P. R. China
Department of Breast Surgery, West China Hospital of Sichuan University, Chengdu 610041, P. R. China
Corresponding author: DU Zhenggui, Email: docduzg@163.com

(Abstract) Objective To find the hub genes related to bone metastasis of breast cancer by weighted gene co-expression network analysis (WGCNA) method, and provide theoretical support for the development of new targeted therapeutic drugs. **Methods** The basic clinical features of 286 breast cancer patients and the gene expression information of tumor specimens were downloaded from the GSE2034 dataset from the Gene Expression Omnibus. R software was used to analyze the gene microarray. The WGCNA package embedded in the R software was used for various analysis in weighted correlation network analysis. Cox proportional hazard regression was performed by using SPSS software. **Results** The top one quarter genes with the greatest variance variability were selected by WGCNA, and a total of 5 000 genes were used for further enrichment analysis. Finally, 15 gene co-expression modules were constructed, and the magenta module (r=0.94, P<0.001) was significantly positively correlated with bone metastasis of breast cancer. It was further found that six hub genes highly associated with bone metastasis in the magenta module were: Ral GTPase-activating protein subunitalpha-1 (RALGAPA1), B-cell antigen receptor complex-associated protein alpha chain (CD79A), immunoglobulin kappa chain C region (IGKC), arrestin beta 2 (ARRB2), differentially expressed in FDCP 6 homolog (DEF6), and immunoglobulin lambda variable 2 (IGLV2). **Conclusion** We found that RALGAPA1, CD79A, IGKC, ARRB2. DEF6, and IGLV2 may play an important role in bone metastasis of breast cancer.

[Keywords] breast cancer; bone metastasis; weighted gene co-expression network analysis; hub gene

骨是乳腺癌最常见的靶向转移器官,最常见的 骨转移部位是椎骨、骨盆、长骨和头骨的近端,出 现骨转移的乳腺癌患者往往预后不良^[1-2]。当前的 姑息性治疗手段并不能提高乳腺癌骨转移患者的 总生存率。因此,有必要深入探究乳腺癌骨转移的 机制,从而寻找新的靶向治疗药物。乳腺癌骨转移 的机制极其复杂,涉及到肿瘤细胞、成骨细胞、破



DOI: 10.7507/1007-9424.201912069 通信作者: 杜正贵, Email: docduzg@163.com

骨细胞和矿化骨基质之间的协同相互作用^[3-5]。骨 质降解、骨形成和肿瘤生长之间的相互作用受到基 因表达及信号转导通路的调控^[6-7]。随着越来越多 的乳腺癌骨转移相关基因被发现,乳腺癌骨转移的 相关机制也正逐渐被揭开。本研究拟从美国国立 生物技术信息中心(National Center for Biotechnology Information, NCBI)的基因表达综合数 据库(gene expression omnibus, GEO)下载乳腺癌 骨转移数据,并采用加权基因共表达网络分析 (weighted gene coexpression network analysis, WGCNA)方法找到乳腺癌骨转移相关的关键 (hub)基因。

1 资料与方法

1.1 乳腺癌患者肿瘤组织的基因表达数据整理

从 GEO 数据库下载 GSE2034 数据集中共 286 例乳腺癌患者的基本临床特征及肿瘤标本的基因 表达信息。基因测序平台是 GPL96, 基因芯片的探 针注释信息来自于 Affymetrix 公司下载的原始文件 [(HG-U133A) Affymetrix Human Genome U133A Array]。通过 R 软件实现基因芯片的分析。通过 affy 包将原始的 CEL 数据导入 R 软件中, 并使用 Robust Multi-Array (RMA)对各乳腺癌样本的基因 表达值进行归一化处理;然后将归一化的表达谱 用于进一步的分析。通过基因芯片的注释信息将 探针与对应的基因进行匹配。排除一个探针同时 对应一个以上基因的情况,对于一个基因同时对应 多个探针的情况应该计算所有探针表达值的平均 值。本研究选择在各数据集样本中表达超过 50% 的探针用于后续研究; 以最近邻算法 (k-nearest neighbor, KNN)补充缺失值;以limma包进行差 异表达值计算,并用贝叶斯方法进行多重检验校正。

1.2 建立基因共表达模块

首先用 R 语言中的 flashClust 工具对所有乳腺 癌样本进行聚类分析,同时检测离群值并剔除。通 过 WGCNA 算法,在构建基因模块的过程中筛选 出适当的软阈值。采用梯度法检测不同软阈值(从 1 到 20)时不同基因模块的独立性和平均连接度。 取模块独立性为 0.8 时的软阈值为最适软阈值。确 定软阈值后,使用 WGCNA 算法构建基因共表达 模块并提取出每个模块中相应基因的信息。

1.3 基因共表达模块与临床特征的关系分析

根据每个基因模块中的特征基因将连接度较 高的基因聚集在同一个模块中,进而挖掘与临床特 征相关的重要基因模块,并可容易识别出与该表型 高度相关的表达模块。在每个表达谱中,基因显著性(GS)即为基因表达谱与每种特征之间相关性的绝对值。模块成员(module membership, MM)是基因在模块内的模块隶属度。特征基因连接度(KME)值为使用主成分分析计算的特征基因的值,其将每个模块的表达模式概括为单个基因特征表达谱。

1.4 确定与骨转移相关的关键基因

将关键基因模块中每个基因的 KME 和 GS 值 分别进行排序,后将两者的排序结果通过韦恩图取 交集,将取交集后得到的基因进行单因素 Cox 比例 风险回归分析,以探索其与乳腺癌骨转移的相关 性,继而找到乳腺癌骨转移相关的关键基因。

1.5 统计学方法

采用 R 软件 (3.5.1) 中植入的 WGCNA 包用于 加权相关网络分析中的各种分析 (http://www. r-project.org/)。使用 SPSS 25.0 (SPSS for Windows) 进行单因素 Cox 比例风险回归分析。检验水准 $\alpha=0.05$ 。

2 结果

2.1 乳腺癌患者的基本临床信息

本研究纳入 GSE2034 数据集中共 286 例乳腺 癌患者的肿瘤样本。这些患者中共有 107 例发生了 乳腺癌的远处转移,其中出现骨转移的患者 69 例。在发生远处转移的患者中,有 39 例在术后 2 年内发生远处转移,63 例在术后 2 年后才发生远 处转移,另 5 例患者发生远处转移的时间不详。患 者发生远处转移的中位时间为 77.52 个月(2~171 个月)。

2.2 共表达基因模块的构建

通过 WGCNA 方法选取了 286 例乳腺癌患者 的肿瘤样本中方差变异度最大的前 5 000 个基因做 进一步的基因富集分析,并根据基因表达值构建基 因共表达模块(图 1a 和 1b)。

筛查合适的软阈值是进一步分析的基础,如 图 1c-1f 所示,当软阈值等于 5 时,无尺度符合指 数为 0.8,并且各基因模块的平均连接度更高,因此, 5 为最适软阈值。最终,通过 WGCNA 方法,本研 究共建立了 15 个基因共表达模块,其中每个基因 模块都用不同的颜色表示(图 1g-1i),黑色 (black)、蓝色(blue)、棕色(brown)、青色 (cyan)、绿色(green)、黄绿色(greenyellow)、灰色 (grey)、洋红色(magenta)、粉红色(pink)、紫色 (purple)、红色(red)、深粉色(salmon)、棕褐色 (tan)、天青色(turquoise)和黄色(yellow)模块中分别包含了242、826、678、38、484、102、98、185、203、167、250、47、50、843和587个基因(图1g)。

2.3 确定乳腺癌骨转移相关基因模块及关键基因

从模块-临床特征关系表中可以看出(基于 Pearson 相关性检验),在所有模块中,magenta 模 块是与乳腺癌骨转移事件呈显著正相关且相关系 数最大的模块(图 2a)。各相关基因模块与乳腺癌 骨转移关系的散点图(图 2b-2e)也显示,magenta 模块与乳腺癌骨转移呈显著正相关(r=0.94, P<0.001)。。

为了探索 magenta 模块中的关键基因,本研究 通过 KME 值对 magenta 模块中的基因进行了排 序,并取前 30 个 KME 值最大的基因;然后根据 GS 值对 magenta 模块中的基因进行排序后,同样 取排在前 30 个的基因;之后通过韦恩图将两次排 序结果取交集后,共得到 6 个基因:分别是 Ral GTPase 激活蛋白亚单位 α-1 (RALGAPA1)、B 细胞 抗原受体复合物相关蛋白 α 链 (CD79A)、人免疫 球蛋白 kappa 链(IGKC)、抑制蛋白 β2(ARRB2)、 FDCP 6 同源异构体(DEF6)和人免疫球蛋白 lambda 链(IGLV2)。通过单因素 Cox 比例风险回 归分析,本研究验证了以上 6 个基因均影响乳腺癌 骨转移患者的生存(表1)。

3 讨论

近年来,GEO 等数据库中基因微阵列数据集的数量不断增加为乳腺癌相关研究提供了极好的机会。不同于传统的差异基因表达分析,WGCNA 通过整合多个数据集,并使用无监督的层次聚类方法识别基因模块,既避免了潜在的主观决策或选择偏差,又揭示了目标表型的复杂生物学机制^[8-10]。

本研究通过 WGCNA 方法共构建了 15 个基因 共表达模块,并且找到了与乳腺癌骨转移呈显著正 相关的 magenta 基因模块。通过进一步分析,本研 究找到了 magenta 基因模块中与骨转移呈显著正相 关的 6 个关键基因,分别是 RALGAPA1、CD79A、 IGKC、ARRB2、DEF6 和 IGLV2。



图 1 示乳腺癌患者肿瘤样本的聚类分析热图、软阈值的确定、动态剪切树、交联图及基因表达网络的 3D 图

a: 286 例乳腺癌患者的肿瘤样本行聚类分析后,发现有离群值; b: 剔除离群值后,剩余样本聚类良好; c: 不同软阈值下的无尺度网络符合指数; d: 不同软阈值下的平均连接度; e: 软阈值为5时的连接度分布直方图; f: 软阈值为5时的无尺度网络拓扑检测结果, log₁₀(k)为连接度的常用对数值, log₁₀[p(k)]为概率的常用对数值; g: 通过 WGCNA 方法构建基因共表达模块, 图中的每个分支都代表一个基因,下面的每种颜色代表一个基因共表达模块; h: 共表达基因相互作用关系的交联图; i: 整个基因表达网络的 3D 图,每个基因用一个点来表示,其中点的颜色对应于该基因所属的基因模块



图 2 示乳腺癌骨转移相关的基因模块的确定

a: 模块特征基因(MEs)与乳腺癌临床特征之间的 Pearson 相关性检验, 表格中每个单元格显示 MEs(行)和临床特征(列)之间的相关性 r 值和 P 值, 基于这种相关性, 表格根据颜色图例进行颜色编码(格子中括号内为 P 值); b-e: 乳腺癌骨转移各相关模块, 包括 megenta(b)、turquoise(c)、grey(d)和 greenyellow(e)模块的 GS 与 MM 关系的散点图, 结果 magenta 模块中 GS 和 MM 之间的相关性最强

基因	β值	标准误	OR (95% CI)	P值
RALGAPA1	1.680	0.398	5.363 (2.461, 11.690)	0.000 024
IGKC	-0.700	0.168	0.497 (0.357, 0.691)	0.000 032
IGLV2	-0.580	0.148	0.560 (0.419, 0.749)	0.000 093
CD79A	-1.364	0.324	0.256 (0.135, 0.483)	0.000 026
ARRB2	-1.397	0.344	0.247 (0.126, 0.485)	0.000 048
DEF6	-1.711	0.423	0.181 (0.079, 0.414)	0.000 053

表 1 magenta 基因模块中的关键基因表达与乳腺癌骨转移的

单因素 Cox 比例风险回归分析

RALGAPA1 基因编码了 RAL-GTP 酶激活蛋白的主要亚基管样蛋白基因1(GARNL1)。RAL由两种紧密相关的G蛋白 RALA和 RALB构成,其与RAS基因具有高度相似的序列,是小GTP 酶-RAS超家族的一个分支。RAL在细胞中以与GDP结合的未激活状态和与GTP结合的激活状态之间循环^[11]。RAL-GTP 酶的固有GTP 酶活性非常弱,因此它们主要依赖鸟嘌呤核苷酸交换因子(RALGEF)和GTP 酶活化蛋白(RALGAP)来催化GDP-GTP 交换。有研究^[12]表明,GARNL1通过与14-3-3蛋白相互作用来调节胰岛素刺激的RALA激活。RALGEF-RALA信号通路在RAS介导的肿瘤进展中起重要作用。Wu等^[13]的研究表明,RALA减少会抑制前

列腺癌细胞发生骨转移。

早期研究发现,到达骨微环境中的肿瘤细胞会分泌甲状旁腺激素相关肽或白细胞介素等因子,导致成骨细胞分化减少、破骨细胞生成增加和骨基质沉积减少^[14]。Luger等^[15]发现,B细胞受体亚单位——CD79A被肿瘤细胞分泌的细胞因子激活后,致使T细胞的增殖受到抑制、促肿瘤发生细胞因子的分泌增加,如白细胞介素6(IL-6)。IL-6过表达可促进破骨细胞的生成和骨质溶解,研究^[16]显示,IL-6的过表达与肿瘤骨转移高度相关。

IGKC和 IGLV2 基因分别编码了免疫球蛋白 kappa和 lambda 轻链蛋白。Groot Kormelink等^[17] 的研究表明,免疫球蛋白 kappa和 lambda 轻链蛋 白的表达与乳腺癌肿瘤大小、肿瘤分级、临床分期 和血管侵袭呈显著正相关。lambda 轻链蛋白的表 达与乳腺癌特异性生存更差以及无远处转移时间 更短高度相关。免疫球蛋白轻链可以以抗原特异 性方式激活肥大细胞。在包括乳腺癌、恶性黑色素 瘤等多种肿瘤的周围都发现了肥大细胞的浸润,肥 大细胞可以募集包括巨噬细胞在内的多种其他炎 性细胞,进而参与肿瘤的侵袭和转移^[18-19]。

除了乳腺癌细胞自身分泌的一些促进肿瘤生 长的细胞因子以外,骨微环境的其他组成部分在乳 腺癌细胞的定植中也发挥了关键的作用。如活化 的成骨细胞可以上调 TNF 相关激活诱导细胞因 子—— 核因子 κ B 受体活化因子配体 (receptor activator of nuclear κ B ligand, RANKL)的表达, 而 RANKL 与其受体核因子 κ B 受体活化因子 (RANK) 的结合是破骨细胞活化所必须的^[20]。Binder 等^[21]在 DEF6 基因缺失的炎性骨溶解小鼠模型中发现, DEF6 缺乏导致肿瘤坏死因子 α (TNF- α)诱导的破 骨细胞生成及骨吸收显著增加。DEF6 通过调节由 内源性 RANKL 或 TNF- α 诱导的干扰素- β (IFN- β) 介导的自分泌反馈路径, 来抑制关键的破骨细胞生 成因子—— 依赖性钙调神经磷酸酶 1 (NFATc1)、B 淋巴细胞诱导成熟蛋白 1 (Blimp1)和 c-Fos 的表达。

目前的一些研究^[22-24]显示,ARRB2(又名 β-arrestin-2)在肿瘤中的表达与肿瘤的侵袭转移能 力有关。如 Sun 等^[25]发现,ARRB2表达下调与肝癌 远处转移呈显著正相关(*P*=0.024),同时体外研究 的结果显示,ARRB2的过表达可以降低肝癌细胞 的侵袭和迁移能力。此外,过表达的ARRB2会导 致 E-钙黏蛋白的表达上调、波形蛋白表达下调和磷 脂酰肌醇依赖的激酶[蛋白激酶 B(Akt)]活化受 抑。近年一项研究^[26]也发现,ARRB2基因表达下调 导致趋化因子受体 2(CXCR2)基因表达上调,进而 促进非小细胞肺癌患者出现淋巴结转移。

目前乳腺癌骨转移的作用机制仍有待进一步 探索,特别是骨转移瘤微环境中免疫抑制的相关机 制,未来将会有更多免疫治疗反应的潜在作用靶点 被发现。本研究通过 WGCNA 方法分析得到与乳 腺癌骨转移密切相关的 magenta 基因模块,并确定 该基因模块中的 6 个关键基因分别是: CD79A、 RALGAPA1、IGKC、ARRB2、DEF6 和 IGLV2。以上 6 个关键基因在乳腺癌骨转移中的具体机制仍有待 进一步的研究阐释。

重要声明

利益冲突声明:本文全体作者阅读并理解了《中国普外 基础与临床杂志》的政策声明,我们没有相互竞争的利益。

作者贡献声明:本研究由贺功建和杜正贵共同设计;资 料统计、数据分析、绘图及论文写作由贺功建在杜正贵的指 导下完成。

参考文献

- Kennecke H, Yerushalmi R, Woods R, et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol, 2010, 28(20): 3271-3277.
- 2 Wang CY, Wu GY, Shen MJ, *et al.* Comparison of distribution characteristics of metastatic bone lesions between breast and prostate carcinomas. Oncol Lett, 2013, 5(1): 391-397.
- 3 Charhon SA, Chapuy MC, Delvin EE, et al. Histomorphometric

analysis of sclerotic bone metastases from prostatic carcinoma special reference to osteomalacia. Cancer, 1983, 51(5): 918-924.

- 4 Tulotta C, Lefley DV, Freeman K, *et al.* Endogenous production of IL1B by breast cancer cells drives metastasis and colonization of the bone microenvironment. Clin Cancer Res, 2019, 25(9): 2769-2782.
- 5 Nakai Y, Okamoto K, Terashima A, *et al*. Efficacy of an orally active small-molecule inhibitor of RANKL in bone metastasis. Bone Res, 2019, 7: 1.
- 6 Cao Q, Chen X, Wu X, *et al.* Inhibition of UGT8 suppresses basallike breast cancer progression by attenuating sulfatide- $\alpha V\beta 5$ axis. J Exp Med, 2018, 215(6): 1679-1692.
- 7 Zhuang X, Zhang H, Li X, *et al.* Differential effects on lung and bone metastasis of breast cancer by Wnt signalling inhibitor DKK1. Nat Cell Biol, 2017, 19(10): 1274-1285.
- 8 Bakhtiarizadeh MR, Hosseinpour B, Shahhoseini M, *et al.* Weighted gene co-expression network analysis of endometriosis and identification of functional modules associated with its main hallmarks. Front Genet, 2018, 9: 453.
- 9 Giulietti M, Occhipinti G, Principato G, et al. Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. Cell Oncol (Dordr), 2017, 40(2): 181-192.
- 10 Liang JW, Fang ZY, Huang Y, et al. Application of weighted gene co-expression network analysis to explore the key genes in Alzheimer's disease. J Alzheimers Dis, 2018, 65(4): 1353-1364.
- 11 Yan C, Theodorescu D. RAL GTPases: biology and potential as therapeutic targets in cancer. Pharmacol Rev, 2018, 70(1): 1-11.
- 12 Chen Q, Quan C, Xie B, *et al.* GARNL1, a major RalGAP α subunit in skeletal muscle, regulates insulin-stimulated RalA activation and GLUT4 trafficking *via* interaction with 14-3-3 proteins. Cell Signal, 2014, 26(8): 1636-1648.
- 13 Wu Z, Owens C, Chandra N, et al. RalBP1 is necessary for metastasis of human cancer cell lines. Neoplasia, 2010, 12(12): 1003-1012.
- 14 Taubenberger AV. *In vitro* microenvironments to study breast cancer bone colonisation. Adv Drug Deliv Rev, 2014, 79-80: 135-144.
- 15 Luger D, Yang YA, Raviv A, et al. Expression of the B-cell receptor component CD79a on immature myeloid cells contributes to their tumor promoting effects. PLoS One, 2013, 8(10): e76115.
- 16 Harmer D, Falank C, Reagan MR. Interleukin-6 interweaves the bone marrow microenvironment, bone loss, and multiple myeloma. Front Endocrinol (Lausanne), 2019, 9: 788.
- 17 Groot Kormelink T, Powe DG, Kuijpers SA, *et al.* Immunoglobulin free light chains are biomarkers of poor prognosis in basal-like breast cancer and are potential targets in tumor-associated inflammation. Oncotarget, 2014, 5(10): 3159-3167.
- 18 Khan MW, Keshavarzian A, Gounaris E, et al. PI3K/AKT signaling is essential for communication between tissue-infiltrating mast cells, macrophages, and epithelial cells in colitis-induced cancer. Clin Cancer Res, 2013, 19(9): 2342-2354.
- 19 Sloan EK, Priceman SJ, Cox BF, *et al.* The sympathetic nervous system induces a metastatic switch in primary breast cancer. Cancer Res, 2010, 70(18): 7042-7052.
- 20 Leibbrandt A, Penninger JM. RANKL/RANK as key factors for osteoclast development and bone loss in arthropathies. Adv Exp Med Biol, 2009, 649: 100-113.
- 21 Binder N, Miller C, Yoshida M, *et al.* Def6 restrains osteoclastogenesis and inflammatory bone resorption. J Immunol, 2017, 198(9): 3436-3447.
- 22 Zhang Y, He W, Zhang S. Seeking for correlative genes and signaling pathways with bone metastasis from breast cancer by integrated analysis. Front Oncol, 2019, 9: 138.
- 23 Masannat J, Purayil HT, Zhang Y, *et al.* βarrestin2 mediates renal cell carcinoma tumor growth. Sci Rep, 2018, 8(1): 4879.
- 24 Ren W, Wang T, He X, *et al.* β-arrestin2 promotes 5-FU-induced apoptosis *via* the NF-κB pathway in colorectal cancer. Oncol Rep, 2018, 39(6): 2711-2720.
- 25 Sun WY, Hu SS, Wu JJ, *et al.* Down-regulation of β-arrestin2 promotes tumour invasion and indicates poor prognosis of hepatocellular carcinoma. Sci Rep, 2016, 6: 35609.
- 26 Cong L, Qiu ZY, Zhao Y, *et al.* Loss of β-arrestin-2 and activation of CXCR2 correlate with lymph node metastasis in non-small cell lung cancer. J Cancer, 2017, 8(14): 2785-2792.

收稿日期: 2019-12-17 修回日期: 2020-03-18 本文编辑: 罗云梅