

# 乳腺癌内分泌治疗敏感相关基因的文献计量学和生物信息学分析

Analysis of Breast Cancer Endocrine Sensitivity Related Candidate Genes Based on Literature Metrological Methods and Bioinformatics

MAI Zhong-lun, CEN Dong-zhi, WANG Yi-nuo, et al.

麦仲伦<sup>1</sup>, 岑东芝<sup>1</sup>, 王贻锴<sup>2</sup>, 杨菲<sup>2</sup>, 张积仁<sup>1</sup>  
(1. 南方医科大学珠江医院肿瘤中心, 广东 广州 510282;  
2. 上海市靶向分子医学研究所, 上海 200433)

**摘要:** [目的] 了解乳腺癌内分泌治疗敏感相关基因的研究现状, 构建相关基因相互联系的网络模块。[方法] 在 Embase、Medline/Pubmed 及 BIOSIS Preview 数据库中检索 2001~2010 年发表的有关乳腺癌内分泌治疗敏感相关基因的文献进行计量学分析, 并对相关基因及其产物进行生物信息学分析。[结果] 168 篇文献纳入研究, 共涉及 267 个基因。文献计量分析结果显示近 3 年来该领域的研究目前处于快速发展阶段; 美国在该领域研究水平处于明显领先地位; *Breast Cancer Research and Treatment* 与 *Clinical Cancer Research* 为载文量最多的期刊。生物信息学分析提示关键基因和瓶颈基因具体包括 CCND1、AKT1、NFKB1、PIK3CA、TPX2、EGFR、Bcl-2、BRCA1、MYC、SRC、VEGFA、AR、ESR1、TNF、BIRC5、CDK1、TP53、Jun。[结论] 通过对乳腺癌内分泌治疗敏感相关基因的文献计量学分析及生物信息学分析, 我们可清楚了解目前该领域研究现状及基因网络模块对未来研究的预测作用, 帮助研究人员作进一步研究打下基础。

**关键词:** 乳腺癌; 内分泌治疗; 基因; 文献计量学; 生物信息学

**中图分类号:** R737.9 **文献标识码:** A **文章编号:** 1004-0242(2012)02-0123-04

内分泌治疗是乳腺癌辅助治疗的重要组成部分, 在乳腺癌治疗中的地位越来越受到人们的重视。内分泌治疗的疗效与肿瘤细胞的激素受体状态密切相关, ER 与 PR 均阳性者有效率为 60%~70%, ER 或 PR 阳性者有效率为 30%~40%, 两者均阴性者有效率小于 10%。然而在临床面临的迫切问题是内分泌治疗耐药, 即使 ER 与 PR 均阳性者中仍约有 30% 患者存在内分泌治疗原发耐药。本文对乳腺癌内分泌治疗敏感相关基因文献进行计量学分析, 并对文献所涉及的基因及其产物进行功能分类及通路分析。

## 1 材料与方法

### 1.1 文献检索

从 Embase、Medline/Pubmed 及 BIOSIS Preview 数据库中检索, 时间限定为 2001 年 1 月至 2010 年

12 月, 物种限定为人类, 语种不限。为提高文献的查准率, Embase 数据库进行 EMtree 扩展检索, 辅以自由词的题目及摘要字段的检索; Pubmed/Medline 数据库进行 MeSH 检索, 辅以自由词的题目及摘要字段的检索; BIOSIS Preview 数据库进行主题检索。乳腺癌的英文检索词包括: ① MeSH 词 “breast neoplasm”, ② EMtree 词 “breast cancer”, ③ 自由词检索使用 “breast cancer”、“breast tumor\*”、“breast neoplasm\*”、“breast carcinoma\*”、“mammary neoplasm\*”。内分泌治疗的英文检索词包括: ① MeSH 词 “aromatase inhibitors”、“estrogen antagonists”、“progesterone”、“megestrol” 及 “antineoplastic agents”、“hormonal”; ② EMtree 词检索使用 “cancer hormone therapy”; ③ 自由词检索 “aromatase inhibitors”、“estrogen antagonist\*”、“hormone therapy”、“endocrine therapy”、“progesterone” 及 “megestrol”。基因的英文检索词为 “gene”。

### 1.2 文献纳入与排除

纳入全部与乳腺癌内分泌敏感相关基因及基因

收稿日期: 2011-05-18; 修回日期: 2011-07-24

通讯作者: 张积仁, E-mail: zhangjiren@126.com

产物相关的临床研究。纳入与排除标准:①文献形式包括论著、会议论文及会议摘要,排除综述、个案报道、书信、书籍及社论等;②纳入文献需包括能提供足够信息的摘要,或可以获取全文资料,排除无法获取摘要的文献;③排除与药物代谢等相关基因的研究;④排除与内分泌治疗敏感性相关的非基因研究;⑤纳入临床研究文献,排除针对基因的结构、机理等研究;⑥排除细胞株体外研究及动物体内实验;⑦排除重复文献,不同数据库来源,作者、题目及出版年相同的文献按一篇统计;⑧排除其他与研究目的不符的文献。

### 1.3 文献计量学分析

对最终纳入文献的出版年、作者、国家、研究机构、发表期刊进行计量学分析。

### 1.4 内分泌治疗敏感相关基因生物信息学分析

利用在线 STRING 对乳腺癌内分泌相关基因表达产物构建蛋白质相互作用网络,在线软件 GATHER(<http://gather.genome.duke.edu/>)按照 Gene Ontology(GO,即基因本体论)分类的标准对其进行功能分类,将 STRING 软件构建的乳腺癌内分泌治疗敏感相关基因相互作用网络图导入 Cytoscape 软件可视化,利用其插件 CentiScape 计算网络及各个节点的拓扑特性。

## 2 结果

### 2.1 检索情况

Embase 数据库检出文献 360 篇, Pubmed/Medline 数据库检出文献 580 篇, BIOSIS Preview 数据库检出文献 1 200 篇, 3 个数据库共检出文献 2 140 篇。经去重及排除后, 168 篇文献符合要求。

### 2.2 文献计量学分析

近 10 年内分泌治疗敏感相关基因相关文献以 2008~2009 年间明显较多,而在 2001~2007 年间文献分布基本呈现缓慢上升趋势(图 1),这与近年来内分泌治疗扮演越来越重要的角色、内分泌治疗耐药问题亦受到越来越多的重视相关。

检索结果发现,共有 28 个国家发表了关于内分泌治疗敏感相关基因方面的研究,美国以 38 篇发文量高居榜首,其后,日本、英国、荷兰和瑞典发文量分别为 20、18、17、14 篇。总的来看,文献发文量主要集中在西欧、北美等发达地区。



168 篇论文发表在 70 种生物医学期刊上,平均每周发文 2.4 篇。其中载文量较多的期刊是 *Breast Cancer Research and Treatment* 与 *Clinical Cancer Research*, 分别为 24、23 篇,明显多于其他期刊,说明这两种期刊是内分泌治疗敏感基因研究发表文献的密集期刊。*Cancer Research* 与 *Journal of Clinical Oncology* 发文量分别为 11 篇和 9 篇。这 4 种期刊集中了约 35% 内分泌治疗敏感基因研究文献。

为避免产生重复数据,仅统计了第一作者的机构。共有 112 个研究机构进行内分泌治疗敏感相关基因方面研究,其中,荷兰的 Erasmus MC, Josephine Nefkens Institute 在 10 年间发表乳腺癌内分泌治疗敏感相关基因文献最多,达 10 篇,其次 4 个研究机构分别为 Nagoya City University Graduate School of Medical Sciences, Inst. Natl. de S./Rech. Med, Linkoping University 和 Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA, 发文量分别为 6、5、4、4 篇。

经调研统计得知,10 年间以第一作者完成乳腺癌内分泌治疗敏感相关基因方面研究 3 篇的有法国的 Bieche I、美国的 Paik S 和日本的 Zhang Z。

### 2.3 内分泌治疗敏感相关基因及产物相互作用

STRING 是蛋白质相互作用数据库,把不同物种中的蛋白质定量关联起来。168 篇符合要求的文献中涉及内分泌治疗敏感相关基因共 267 个。采用 STRING 软件对这 267 个内分泌治疗敏感相关基因表达产物构建蛋白质相互作用网络,得到关系网络图(图 2)。网络中,一个圆形节点代表一个基因,存在相互作用关系的基因用一条线将其连接起来,由图 2 可知,这些基因表达产物大部分之间存在着密切



图2 乳腺癌内分泌治疗敏感相关基因网络图

的相互作用关系,有多个基因处于网络的中心,相互之间以及与其他基因产物的关系均比较密切,处于整个通路的核心位置。整个关系网络图又可以细分为多个子网络,部分基因处在各个子网络的中心位置,起着关键连接和中介作用。

#### 2.4 乳腺癌内分泌治疗敏感相关基因及产物分类

Gene Ontology (GO)数据库是一个较大的公开的生物分类学网络资源的一部分,它包含 17 348 个注释基因,并把它们的功能分为三类:分子功能、生物学过程和细胞组分。在每一个分类中,都提供一个描述功能信息的分级结构。这样,GO 中每一个分类术语都以一种被称为定向非循环图表(DAGs)的结构组织起来。研究者可以通过 GO 分类号和各种 GO 数据库相关分析工具将分类与具体基因联系起来,从而对这个基因的功能进行描述。利用在线软件 GATHER 分析,145 个内分泌治疗敏感相关基因共涉及不同分级的分类 408 类。

#### 2.5 关键基因及瓶颈基因分析

进一步通过生物信息学方法寻找已研究的基因及其相互作用的基因中的关键基因(Hub 基因)及瓶颈基因(Bottleneck 基因)。HPRD 数据库是研究蛋白质相互作用的数据库,从 HPRD 数据库下载目前已经研究的内分泌治疗相关基因的相互作用,利用 Cytoscape 将相互作用网络可视化(图 3),利用其插件 CentiScape 计算网络及各个节点的拓扑特性。网络是由 214 个节点 3 131 条边组成,网络直径为 8,网络的平均距离为 2.73。度为节点连接的边的总数目,最大度值为 169,最小为 2,平均为 29.26±35.08,

网络中连接度高的节点对网络的稳定性起到重要作用,本研究规定度数≥均数+1SD(标准差)的蛋白所对应的基因为 Hub 基因,共有 37 个 Hub 基因,它们的功能改变可能对整个网络至关重要。Betweenness 为网络中所有最短路径中经过该节点的数量比例,最大 Betweenness 值为 1 166 281,最小为 0,平均为 378.08±720.93。相应的节点或者边在整个网络中的作用和影响力,本研究规定 Betweenness ≥ 均数+1SD 的蛋白所对应的基因为 Bottleneck 基因,共有 22 个。我们认为同时为 Hub 基因和 Bottleneck 基因的对内分泌相关基因的网络至关重要,具体包括 CCND1、AKT1、NFKB1、PIK3CA、TPX2、EGFR、Bcl-2、BRCA1、MYC、SRC、VEGFA、AR、ESR1、TNF、BIRC5、CDK1、TP53、Jun。

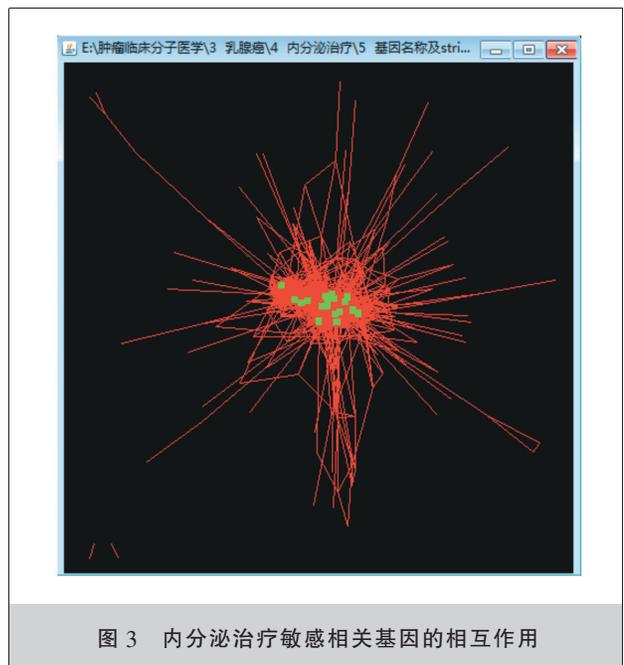


图3 内分泌治疗敏感相关基因的相互作用

## 3 讨论

### 3.1 内分泌治疗敏感相关基因的研究现状

我们通过对近 10 年内分泌治疗敏感相关基因的文献进行计量学分析显示,该领域研究正处于一个快速发展的阶段。美国是该领域相关研究最为活跃的国家。随着内分泌治疗在乳腺癌治疗中的地位日益重要,关注这一主题的文献逐渐增多,其中尤以 *Breast Cancer Research and Treatment* 与 *Clinical Cancer Research* 期刊载文量为多。荷兰的 Erasmus

MC, Josephine Nefkens Institute 是发表该主题文献最多的机构;法国的 Bieche I、美国的 Paik S 和日本的 Zhang Z 是该领域研究的高产作者,但远未形成一个高产作者群体。

### 3.2 内分泌治疗敏感研究热点基因

我们对入选的 168 篇文献中所涉及的 267 个基因进行了生物信息学分析,结果表明这些基因处于相互联系的复杂网络之中,这些基因表达产物大部分之间存在着密切的相互作用关系,该基因关系网络显示有多个基因处于网络的中心,为 HOXB13、IL17BR、ERBB2、ESR1、NCOA3、TP53、Bcl-2、AKT1、CCND1、PGR 和 CYP2D6,提示这些基因为该领域近年研究的热点基因。

他莫昔芬无反应者多存在 HOXB13 基因高表达,IL17-BR 表达则降低,HOXB12 与 IL17BR 比率高,或 HOXB13 水平高,他莫昔芬敏感性低<sup>[1]</sup>。Gruvberger-Saal 等<sup>[2]</sup>的研究发现,ER beta 在 ER(alpha)阴性患者中对内分泌治疗敏感具有重要的预测意义,与内分泌治疗较好的疗效相关联。AIB1 高表达与一些预后差的预测因子相关,例如高 HER-2、高 NHG、ER(-)、PgR(-)及存在淋巴结转移患者接受他莫昔芬治疗的 DFS 较短<sup>[3]</sup>。在原发性乳腺癌患者中,p53 蛋白质的积累亦可预测他莫昔芬的耐药性,另一方面,携带 Pro72 等位基因 ER 阳性患者更能在他莫昔芬治疗中受益,因此 Pro72 等位基因缺失的患者可考虑应用其他的治疗<sup>[4,5]</sup>。Jirstrom 等<sup>[6]</sup>和 Ahnstrom 等<sup>[7]</sup>研究发现 CCND1 扩增和 Cyclin D1 过表达可以预示 ER 阳性乳腺癌患者抗雌激素治疗无效。Cyclin D1 中度阳性在他莫昔芬治疗中明显获益,而弱阳性和强阳性存在他莫昔芬耐药<sup>[6,7]</sup>。当乳腺癌患者出现对内分泌治疗耐药时,抑制 Akt 信号途径有可能成为激素治疗耐药乳腺癌的辅助治疗策略<sup>[8]</sup>。PR-A 与 PR-B 比率高,他莫昔芬治疗耐药<sup>[9]</sup>。Bcl-2 作为评价内分泌治疗敏感性的分子标志物,与 ER 表达呈正相关,Subhawong 等<sup>[10]</sup>的研究显示,Bcl-2 表达下调可促进乳腺转移癌的生长,出现内分泌治疗耐药。CYP2D6 基因型可作为选择内分泌治疗辅助治疗方案的参考指标<sup>[11]</sup>。

### 3.3 内分泌治疗敏感相关基因未来的研究方向

Hub 基因和 Bottleneck 基因对内分泌治疗敏感相关基因网络至关重要,将是该领域未来的研究方向,科技工作人员可以利用该相关敏感基因的相互

关系寻找更多更有价值的内分泌疗效预测基因,我们预测未来基因的研究热点可能为:SRC、EGFR、AR、HDAC1、YWHAZ、CDK1、Jun、BRCA1、PTK2、CAV1、NFKB1、MYC、PAK1、NCOR1、SAT1、FOS、MAPT、KDR、CAPN1、TUBB、GNAQ、KPNA2、CDKN1B、EPOR。这些基因可能成为预测内分泌治疗敏感性的生物学标志物,同时为乳腺癌的个体化综合治疗提供参考。

### 参考文献:

- [1] Jerevall PL, Brommesson S, Strand C, et al. Exploring the two-gene ratio in breast cancer-independent roles for HOXB13 and IL17BR in prediction of clinical outcome[J]. Breast Cancer Res Treat, 2008, 107(2): 225-234.
- [2] Gruvberger-Saal SK, Bendahl PO, Saal LH, et al. Estrogen receptor (beta) expression is associated with tamoxifen response in ER(alpha)-negative breast carcinoma[J]. Clin Cancer Res, 2007, 13(7): 1987-1994.
- [3] Anderson H, Bulun S, Smith I, et al. Predictors of response to aromatase inhibitors [J]. J Steroid Biochem Mol Biol, 2007, 106(1-5): 49-54.
- [4] Yamashita H, Toyama T, Nishio M, et al. p53 protein accumulation predicts resistance to endocrine therapy and decreased post-relapse survival in metastatic breast cancer [J]. Breast Cancer Res, 2006, 8(4): R48.
- [5] Wegman P, Stal O, Askmal MS, et al. p53 polymorphic variants at codon 72 and the outcome of therapy in randomized breast cancer patients [J]. Pharmacogenet Genomics, 2006, 16(5): 347-351.
- [6] Jirstrom K, Stendahl M, Ryden L, et al. Adverse effect of adjuvant tamoxifen in premenopausal breast cancer with cyclinD1 gene amplification[J]. Cancer Res, 2005, 65(17): 8009-8016.
- [7] Ahnstrom M, Nordenskjöld B, Rutqvist LE, et al. Role of cyclin D1 in ErbB2-positive breast cancer and tamoxifen resistance[J]. Breast Cancer Res Treat, 2005, 91(2): 145-151.
- [8] Tokunaga E, Kimura Y, Mashino K, et al. Activation of PI3K/Akt signaling and hormone resistance in breast cancer[J]. Breast Cancer, 2006, 13(2): 137-144.
- [9] Hopp TA, Weiss HL, Hilsenbeck SG, et al. Breast cancer patients with progesterone receptor PR-A-rich tumors have poorer disease-free survival rates [J]. Clin Cancer Res, 2004, 10(8): 2751-2760.
- [10] Subhawong AP, Molavi DW, Halushka MK, et al. Bcl-2 is infrequently upregulated in metastatic breast carcinomas [J]. Lab Invest, 2009, 89: 70A-71A.
- [11] Kiyotani K, Mushiroya T, Sasa M, et al. Impact of CYP2D6\*10 on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy [J]. Cancer Sci, 2008, 99(5): 995-999.