

乳腺癌复发相关基因的文献计量学和生物信息学分析

Bibliometrics Analysis of Breast Cancer Recurrence Related Genes

JIANG Zhao-jing, CEN Dong-zhi, YANG Fei, et al.

姜兆静¹, 岑东芝¹, 杨 菲², 张积仁¹

(1.南方医科大学附属珠江医院肿瘤中心, 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] 总结对乳腺癌复发具有重要价值的相关基因, 以便在基因检测基础上更精确有效地指导乳腺癌的分子预测, 及时发现复发病例。[方法] 以 Embase、Medline/Pubmed、BIOSIS Preview 数据库为文献来源, 采用文献计量分析法对近 10 年乳腺癌复发相关基因文献进行统计, 对其出版年、作者、国家、研究机构、发表期刊进行计量学分析, 并对涉及的基因及其产物进行功能分类及通路分析。[结果] 共有 378 篇文献入组, 涉及 650 个乳腺癌复发相关基因。生物信息学分析提示关键基因共 87 个, 瓶颈基因共 44 个, 同时为关键基因和瓶颈基因的有 TP53、ESR1、CDH1、MYC、EGFR、TNF、CCND1、PCNA、SRC、NFKB1、Bcl-2、IGF1、PPARG、TOP2A、BIRC5、CDK1、VEGFA、BRCA1、FOS、AR、IL6、CCNA2、PTGS2、CDKN1A、CD44、HIF1A、STAT1、APP、ERBB2、FOXO1、PARP1、RRM2、MMP9、RB1、IL8、MKI67、FEN1、KIAA0101。[结论] 乳腺癌复发相关的关键基因和瓶颈基因应重视, 在此基础上深入研究, 可使这些基因从临床角度更好地指导乳腺癌的复发监测, 从而改善乳腺癌患者的预后。

关键词: 乳腺癌; 复发; 基因; 计量学分析; 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)02-0127-05

乳腺癌是一种发病率和死亡率逐年升高的恶性肿瘤, 是严重威胁女性身心健康的恶性肿瘤之一, 而乳腺癌的复发又是影响其治疗疗效及预后的主要因素之一。目前有关乳腺癌复发方面的研究越来越受到重视, 宏观上, 相关学者分别从诊断、治疗等多方面进行研究阐释; 微观上, 复发相关的分子生物学及相关基因的研究逐渐成为热点。当前迫切需要寻找一个敏感、能早期预测乳腺癌复发的分子生物学标志物及抑制其复发的基因治疗靶点, 为临床对乳腺癌的复发预测提供一定的参考依据。

1 材料与方法

1.1 文献检索

本研究从 Embase 数据库、Medline/Pubmed 数据库及 BIOSIS Preview 数据库 3 个数据库中进行乳腺癌复发相关基因文献检索。为提高文献的查准率, Embase 数据库进行 EMtree 扩展检索, Pubmed/Med-

line 数据库进行 MeSH 检索, 此两个数据库均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*”。复发的英文检索词包括: ① MeSH 词“recurrence”, ② EMtree 词检索使用“relapse”、“recurrence risk”、“cancer recurrence”、“cancer relapse”及“tumor recurrence”, ③ 自由词检索使用“recurrence*”、“relapse*”。基因的英文检索词为“gene”。

1.2 文献纳入与排除

纳入全部与乳腺癌复发相关基因及基因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 排除与复发相关的非基因研究; ④ 纳入临床研究文献, 排除针对基因的结构、机理等研究; ⑤

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

排除细胞株体外研究及动物体内实验;⑥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;⑦排除其他与研究目的不符的文献。时间限定为2001~2010年,物种限定为人类,语种不限。纳入与排除文献的过程,由2名研究人员独立进行,最终结果互相比对,不一致的地方互相讨论,或向临床医生(副主任医师)、实验研究人员(研究员)请教。

1.3 文献计量学分析

对最终纳入乳腺癌复发相关基因文献分别从出版年、作者、国家、研究机构、发表期刊等角度进行计量学分析。

1.4 乳腺癌复发相关基因生物信息学分析

用在线STRING对乳腺癌复发相关基因表达产物构建蛋白质相互作用网络,然后在线软件GATHER(<http://gather.genome.duke.edu/>)根据Gene Ontology(GO,即基因本体论)分类的标准对这些基因及其产物进行功能分类,将STRING软件构建的复发相关基因相互作用网络图导入Cytoscape软件可视化,用其插件CentiScape计算网络及各个节点的拓扑特性。

2 结 果

2.1 检索结果

Embase数据库检出文献655篇, Pubmed/Medline数据库检出文献264篇, BIOSIS Preview数据库检出文献986篇,共检出文献1905篇。按照纳入和排除标准,最终共有378篇文献符合要求。

2.2 文献计量学分析

纵观10年间的乳腺癌复发相关基因文献量,整体处于一个缓慢逐步上升的过程,并未出现迅速的飞跃,研究增长速度逐步提高。

共有45个国家发表了乳腺癌复发相关基因的文献,其中美国最高,发文量103篇,英国以51篇发文量位居第2,发文量前10名的国家中的另外8个国家分别为法国(38篇)、日本(31篇)、瑞典(29篇)、德国(27篇)、加拿大(22篇)、中国(19篇)、荷兰(17篇)、意大利(16篇)。

378篇论文发表在133种生物医学期刊上,平均每刊发文3.45篇,其中*Breast Cancer Research and Treatment*刊文53篇居首位,按刊文量的高低,

其他超过10篇的期刊主要有*Cancer Research*、*Clinical Cancer Research*、*Journal of Clinical Oncology*、*British Journal of Cancer*,刊文量分别为38、35、25、15篇。

共有239个研究机构进行乳腺癌复发相关基因方面的研究,其中美国的University of Texas M. D. Anderson Cancer Center和Lund University, Malmö University Hospital分别以12、10篇发文量位居前两名。其他发文量达5篇以上的机构分别为Tohoku Univ, Sch Med, Cardiff Univ, Sch Med, Thomas Jefferson University, Laboratoire d'Oncogenetique-INSERM E0017, Mount Sinai Hospital, Princess Grace Hosp, Osaka Univ, Grad Sch Med, Dept Surg Oncol, Mayo Clinic College of Medicine, Netherlands Cancer Institute, University Medical Center Nijmegen。

法国Faculte des Sciences Pharmaceutiques et Biologiques de Paris的Bieche Ivan在10年间以第一作者发表乳腺癌复发相关基因方面研究7篇,为该领域做出了巨大贡献。其他依据发文量依次为荷兰University Medical Center Nijmegen的Span Paul N、爱尔兰Dublin City University的Doolan P、日本Tohoku Univ, Sch Med的Suzuki, Takashi、阿根廷Centro de Investigaciones Inmunologicas Basicas y Aplicadas (CINIBA), Facultad de Ciencias Medicas, Universidad Nacional de La Plata的Abba Martin C、英国St George's Hospital的Al Sarakbi W、法国Institut Gustave Roussy的Apetoh Lionel、Laboratoire d'Oncogenetique-INSERM E0017的Bieche I、美国Mayo Clinic College of Medicine的Goetz Matthew P、芬兰Univ Helsinki的Joensuu H、德国University Hospital of the RWTH Aachen的Noetzel E、瑞典Lund University的Ryden L, 发文量分别为5、4、4、3、3、3、3、3、3、3、3、3篇。

2.3 乳腺癌复发相关基因及产物相互作用

应用STRING软件分析可直观说明和探讨乳腺癌复发相关基因及产物的相互作用。STRING是由EMBL、SIB和UZH共同开发管理的、由已知和预测的相互作用蛋白汇集的相互作用数据库,主要包含已有文献报道的相互作用蛋白信息以及通过共表达研究、高通量实验和基因组信息分析预测的结果,描述的相互作用包括生理上的直接相互作用和功能上的间接相互作用。378篇入组文献共涉及650个乳

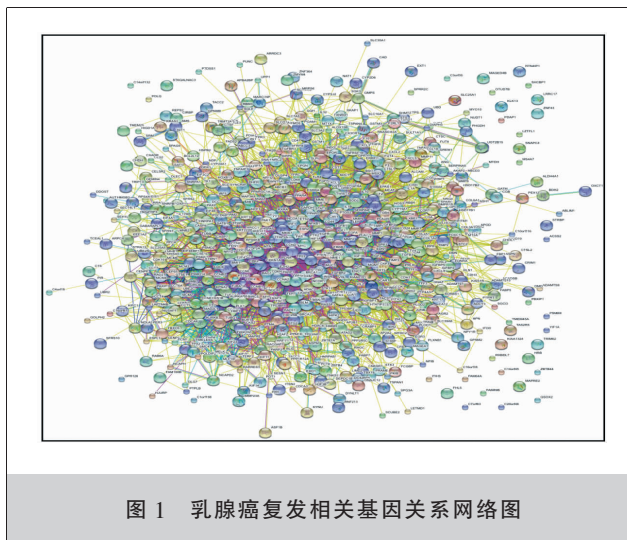


图1 乳腺癌复发相关基因关系网络图

腺癌复发相关基因。STRING 软件对这 650 个乳腺癌复发相关基因表达产物构建蛋白质相互作用网络,见图 1。这些基因表达产物大部分之间存在着密切的相互作用,部分基因处于网络的中心,相互之间以及与其他基因产物之间的关系均比较密切,处于整个通路的核心。整个关系网络图又可细分为多个子网络,部分基因处于各子网络的中心位置,起着关键连接和中介作用。

2.4 基因的功能分类

为了进一步了解目前已经研究的乳腺癌复发相关基因,我们利用在线软件 GATHER(<http://gather.genome.duke.edu/>)按照 GO 分类的标准对其进行功能分类。GO 分类法把 17 348 个注释基因产物的功能分为分子功能、生物学过程和细胞组分三类。

利用在线软件 GATHER 分析,650 个乳腺癌复发相关基因共涉及不同分级的分类 1 087 类,其中贝叶斯(Bayes)因子 ≥ 0 的分类有 129 类,贝叶斯因子 ≥ 3 的分类有 53 类,利用 GODAG 软件对 53 类 GO 分类进行可视化,了解其之间层次相互包含关系,去除过大或过细分类,按贝叶斯因子由大到小的顺序,复发相关基因主要集中于(括号内数字指所含基因数量):细胞增殖(142)、细胞周期(111)、细胞有丝分裂 M 期(32)、核分裂(33)、DNA 复制和染色体周期(29)、原浆移动 cytokinesis(22)、细胞周期检测点(12)、酶联受体蛋白信号通路(22)、凋亡(42)、DNA 复制(20)、细胞周期有丝分裂中 G₂/M 期转变(11)、应激反应(63)、G 蛋白耦联受体蛋白信号通路(18)、对非生物刺激的反应(35)、起源于 Pol II 激动

子转录(38)、DNA 代谢(44)、细胞结构和起源(60)、参与单价无机阳离子转送(1)、参与细胞有丝分裂起始控制点(4)、参与神经生理过程(15)、胶原蛋白分解代谢(6)等。

GO 分类中,乳腺癌复发相关基因涉及最多的分类是细胞增殖类,由于基因之间直接或间接的相互作用或者不同的功能分类依据之间具有一定的从属关系或相互作用,一种基因可存在于不同的功能分类中,如有关细胞周期、细胞有丝分裂、核分裂、DNA 复制等的分类依据在一定程度上亦属于细胞增殖的范畴,而细胞增殖的能力,尤其在涉及到肿瘤细胞时,往往反映其侵袭力。故此分类中的基因多与乳腺癌复发有关。同时在细胞增殖分类中的某一种基因,又可归属于前面述及的与细胞增殖分类相关的其他分类,如 CCNA2、CCNB1、CCNB2 可存在于细胞增殖、细胞周期、细胞周期有丝分裂 M 期、核分裂、原浆移动、细胞周期有丝分裂中 G₂/M 期转变等不同分类,这同时也说明了该基因所具有的不同功能。

根据基因所涉及到的分类数目的多少,现列举几个重要的基因:①BRCA1、BRCA2、ESR1:参与细胞增殖、细胞周期、细胞有丝分裂 M 期、核分裂、细胞周期检测点、凋亡、应激反应、起源于 Pol II 激动子转录、DNA 代谢、细胞结构和起源;②BRN1:细胞增殖、细胞周期、细胞有丝分裂 M 期、核分裂、DNA 复制和染色体周期;③CDK2:细胞增殖、细胞周期、细胞有丝分裂 M 期、核分裂、DNA 复制和染色体周期、原浆移动、DNA 复制、细胞周期有丝分裂中 G₂/M 期转变、参与细胞有丝分裂起始控制点;④EGFR:细胞增殖、细胞周期、酶联受体蛋白信号通路;⑤TP53:细胞增殖、细胞周期、细胞周期检测点、凋亡、应激反应、DNA 代谢、细胞结构和起源;⑥ESPL1:细胞增殖、细胞周期、细胞有丝分裂 M 期、核分裂、DNA 复制和染色体周期、原浆移动、凋亡、细胞结构和起源。

2.5 关键基因及瓶颈基因分析

将 STRING 软件构建的复发相关基因相互作用网络图导入 Cytoscape 软件可视化(图 2),利用其插件 CentiScape 计算网络及各个节点的拓扑特性。网络是由 608 个节点 17 672 条边组成,直径为 7,网络的平均距离为 2.65。度为节点连接的边的总数目,最大度值为 428,最小为 2,平均为 58.13 $\pm$ 65.07,网络中连接度高的节点对网络的稳定性起到重要的作

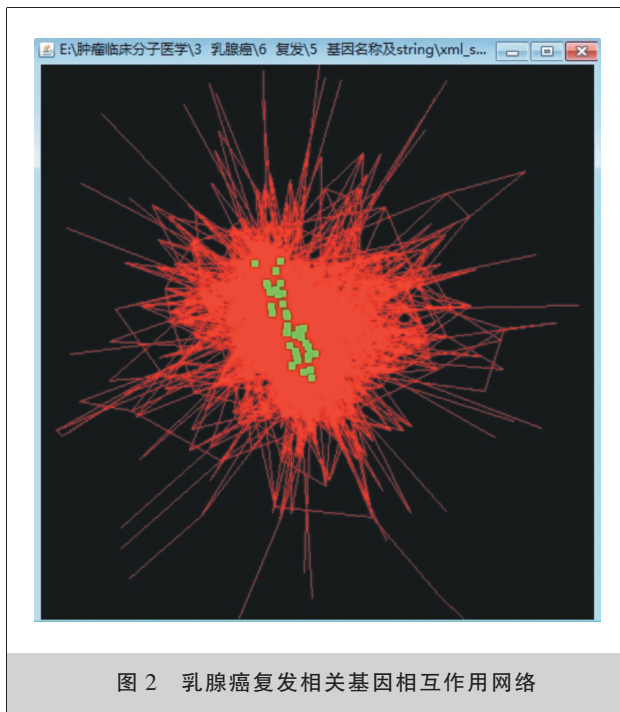


图2 乳腺癌复发相关基因相互作用网络

用。本研究规定度数 $\geq$ 均数+1SD的蛋白所对应的基因为关键基因(Hub基因),共87个,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26 288.68,最小为0,平均为 $1\,002.47 \pm 2\,047.84$ 。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大于等于均数+1SD的蛋白所对应的基因为瓶颈基因(Bottleneck基因),共44个。我们认为同时为Hub基因和Bottleneck基因的对复发相关基因的网络至关重要,有:TP53、ESR1、CDH1、MYC、EGFR、TNF、CCND1、PCNA、SRC、NFKB1、Bcl-2、IGF1、PPARG、TOP2A、BIRC5、CDK1、VEGFA、BRCA1、FOS、AR、IL6、CCNA2、PTGS2、CDKN1A、CD44、HIF1A、STAT1、APP、ERBB2、FOXO1、PARP1、RRM2、MMP9、RB1、IL8、MKI67、FEN1、KIAA0101。

3 讨论

基因芯片技术的不断完善涌现出大量乳腺癌复发相关基因,378篇乳腺癌复发相关基因研究文献共涉及复发相关基因650个。研究较多(>5次)的基因及基因组合包括HER-2、TP53、ER ESR1、BRCA1/2、

Bcl-2、MKI67、SERPINB5、TOP2A、CAV1、EGFR、GATA3、PTEN、STS、TGFB1、YWHAZ、21基因、HOXB13、IL17BR、70基因(Mammaprint)等。

HER-2是一种原癌基因,其过表达往往提示细胞增殖。Bartlett等^[1]指出HER-2基因扩增是乳腺癌患者无复发生存期(RFS)和总生存期(OS)的独立不良预后因素。Lu等^[2]研究表明ErbB2和14-3-3zeta共同过表达的乳腺癌患者更易复发。Guiu等^[3]亦指出HER-2高扩增的乳腺癌患者表现出更短的RFS。Li等^[4]研究证实Cyclin A2和HER-2表达,尤其是两者共表达与乳腺癌高复发有关。近10年学者们从不同的层面角度阐释了HER-2与乳腺癌复发的关系,多数研究证明HER-2扩增和过表达均可在一定程度上增加患者的复发风险,是浸润性导管乳腺癌复发的独立预测因子。但对于一些乳腺癌亚型,如炎性乳腺癌,HER-2对复发的预测作用可能和患者前期接受的治疗种类有相关性。另外,联合测定与HER-2在乳腺癌中共表达的部分基因,如AP-2、TOP2A、14-3-3zeta、Cyclin A2等可更及时有效地预测乳腺癌复发。

p53是公认的抑癌基因,p53缺失可以导致多种肿瘤的发生。Ozcelik等^[5]研究证实p53发生错义突变的乳腺癌患者表现出更高的复发和死亡风险。p53是乳腺癌,特别是发生淋巴结转移的乳腺癌患者复发的独立预测因子,可预测乳腺癌早期复发、局部复发及无远处复发生存期等,联合测定VEGF、nm23H1、ErbB2等可提高其预测价值。

BRCA1和BRCA2均为抑癌基因,与遗传性(家族性)高发乳腺癌密切相关。Sharma等^[6]研究表明BRCA1超甲基化是乳腺癌复发的独立预测指标。近10年的研究显示,携带有BRCA1和BRCA2突变的乳腺癌患者,尤其是BRCA1突变的乳腺癌患者,更易发生复发,包括对侧或同侧乳腺癌复发。

目前有关乳腺癌复发相关基因研究亦存在一定的不足。根据对乳腺癌复发相关基因文献的总结,在以下几方面提出相关问题:①通过对乳腺癌复发相关基因及其产物相关作用分析所得的部分关键及瓶颈基因,如CAPN1、TUBB等在已报道的文献中研究尚少,可加强对这些基因的重视,作为进一步的研究方向;②许多复发相关基因往往是和其他基因在复发机制中相互作用,并非独立发挥作用,我们需进一步证实它们各自的作用权限;③某些基因如HER-2,

作用广泛,所涉及的机制涵盖了乳腺癌的复发、转移、对药物的敏感性等多个方面,如何更好地研究这种基因以及在临床中合理应用将是亟待解决的重要问题;④目前对乳腺癌复发相关的分子标志物研究尚局限,多数是在原位导管乳腺癌或浸润性导管乳腺癌等常见乳腺癌类型中进行研究,忽视了乳腺癌其他亚型;⑤乳腺癌复发类型多样,必要时需进行具体分析以免忽略其特定的预后意义,如早期复发、远处复发、转移复发、原发部位复发、乳腺肿瘤内复发、对侧乳腺癌复发等。

总体来讲,随着研究的深入,乳腺癌复发相关基因会越来越具体化,其相关机制及应用会更加明确,可更精确简便地在基因水平监测乳腺癌的复发,改善乳腺癌患者预后。

参考文献:

- [1] Bartlett J, Munro A, Dunn J, et al. Chromosome 17 polysomy (Ch17) as a predictor of anthracycline response: Emerging evidence from the UK NEAT adjuvant breast cancer trial [J]. Cancer Research, 2009, 69.
- [2] Lu J, Guo H, Treekitkarnmongkol W, et al. 14-3-3zeta cooperates with ErbB2 to promote ductal carcinoma in situ progression to invasive breast cancer by inducing epithelial-mesenchymal transition [J]. Cancer Cell, 2009, 16(3): 195-207.
- [3] Guiu S, Gauthier M, Coudert B, et al. Pathological complete response and survival according to the level of HER-2 amplification after trastuzumab-based neoadjuvant therapy for breast cancer[J]. Br J Cancer, 2010, 103(9): 1335-1342.
- [4] Li HP, Ji JF, Hou KY, et al. Prediction of recurrence risk in early breast cancer using human epidermal growth factor 2 and cyclin A2[J]. Chin Med J (Engl), 2010, 123(4): 431-437.
- [5] Ozcelik H, Pinnaduwa D, Bull SB, et al. Type of TP53 mutation and ERBB2 amplification affects survival in node-negative breast cancer [J]. Breast Cancer Res Treat, 2007,105(3):255-265.
- [6] Sharma G, Mirza S, Parshad R, et al. Clinical significance of promoter hypermethylation of DNA repair genes in tumor and serum DNA in invasive ductal breast carcinoma patients[J]. Life Sci, 2010, 87(3-4): 83-91.