

乳腺癌易感性相关基因研究状况分析

Bioinformatics Analysis on Breast Cancer Predisposition Associated Gene
JIANG Mao-zhu, CEN Dong-zhi, WANG Yi-nuo, et al.

姜茂竹¹, 岑东芝¹, 王贻锴², 杨菲¹, 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] 了解乳腺癌易感性相关基因的研究现状。[方法] 对 Embase、Pubmed/Medline 和 BIOSIS Preview 数据库中 2001~2010 年期间乳腺癌易感性相关基因的研究文献进行文献计量学分析和生物信息学分析。[结果] 筛选出相关文献 1 126 篇, 近 10 年文献量呈上升趋势, 457 个研究机构进行相关研究, 涉及基因 291 个, 主要涉及 DNA 损伤修复、氧化应激反应、细胞生长、增殖和凋亡等通路。[结论] 通过文献计量学及生物信息学分析, 对乳腺癌易感性相关基因研究现状和发展趋势有了一个系统的认识。

关键词: 乳腺癌; 易感性; 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)02-0108-03

乳腺癌是妇女中常见的恶性肿瘤之一, 流行病学统计全世界每年新发 115 万例, 每年因乳腺癌死亡的女性多达 40 万例, 发达国家的发病率约为发展中国家的 3 倍^[1]。对乳腺癌流行病学的研究发现, 相对于普通人群, 一级亲属患乳腺癌的人群发病率增加 1 倍^[2,3], 同卵双生子比异卵双生子的相对患乳腺癌风险高^[4,5]。这些研究均表明, 乳腺癌的家族性侵袭更倾向于基因决定性, 而非仅是环境决定性。因此, 为了更好地掌握乳腺癌易感性相关基因的研究动态和发展趋势, 我们运用文献计量学方法和生物信息学方法, 对有关乳腺癌易感性基因的研究文献及研究所涉及的基因进行系统分析。

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 2001~2010 年期间与乳腺癌易感性相关基因研究的文献。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 tu-

mor*”、“breast neoplasm*”、“breast carcinoma*”、“mammary neoplasm*”。风险的英文检索词包括: ① MeSH 词“risk”, ② EMtree 词“cancer risk”, ③ 自由词检索使用“breast cancer risk*”、“risk* of breast cancer”。基因的英文检索词为“gene”。

1.2 文献纳入与排除

纳入与乳腺癌患病风险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、回信、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 排除与患乳腺癌风险相关的非基因研究; ④ 纳入临床研究文献, 排除针对基因的结构、机理等研究; ⑤ 物种限定为人类, 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 检索时间限定为 2001 年 1 月至 2010 年 12 月, 文献语种为英语; ⑧ 排除其他与研究目的不符的文献。

Embase 数据库检出文献 2 074 篇, Pubmed/Medline 数据库检出文献 1 796 篇, BIOSIS Preview 数据库检出文献 975 篇, 共检出文献 4 845 篇。按照纳入和排除标准, 共 1 126 篇文献符合要求。

1.3 乳腺癌易感性相关基因的挑选

对 1 126 篇文献中所研究的基因进行列表, 并

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

同时在 NCBI 中查所涉及基因的通用名、ID 号、功能研究及突变位点进行列表。

1.4 乳腺癌易感性相关基因生物信息学分析

利用在线 STRING 对乳腺癌易感性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。

2 结果

2.1 文献计量学分析

近 10 年乳腺癌易感性相关基因相关文献在 2000~2007 年间缓慢增多,2009~2010 年总体呈现明显上升趋势。见图 1。

经统计,1 126 篇论文发表在 222 种生物医学期刊上,平均每刊发文 5.07 篇,发文量前 10 位的期刊所发表的论文数占总发表论文数的 55.77%。乳腺癌易感性相关基因研究论文发表数量前 10 位的国家和地区如图 2 所示。

457 个研究机构进行乳腺癌易感性相关基因方面研究。发文量最多的是美国的范德比尔特大学医学中心,共发表 41 篇相关文献。美国南加州大学的 Lee SA 在 10 年间以第一作者发表乳腺癌易感性相关基因方面研究文献 10 篇。

2.2 乳腺癌易感性相关基因及产物关系

不考虑研究是阳性结果或阴性结果,1 126 篇文献中相关基因共 291 个。

STRING 软件对 291 个乳腺癌易感性相关基因表达产物构建蛋白质相互作用网络,并构建关系网络图(图 3)。

2.3 乳腺癌易感性相关基因的功能分类

利用在线软件 GATHER 分析发现,291 个乳腺癌易感性相关基因共涉及不同分级的分类 815 类。乳腺癌易感性相关基因主要涉及 DNA 损伤修复、氧化应激反应、细胞生长、增殖和凋亡、细胞周期、激素及物质代谢等细胞通路。

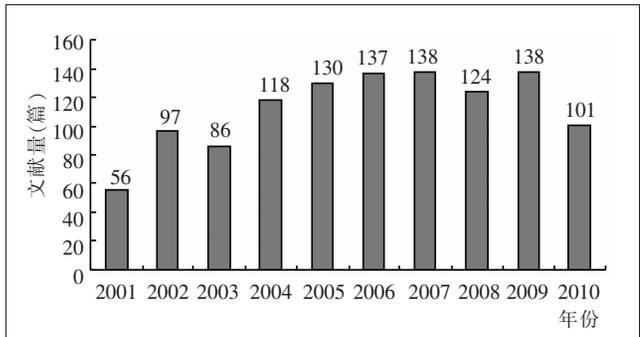


图 1 2001~2010 年三大数据库收录的乳腺癌易感性相关基因研究的论文数量

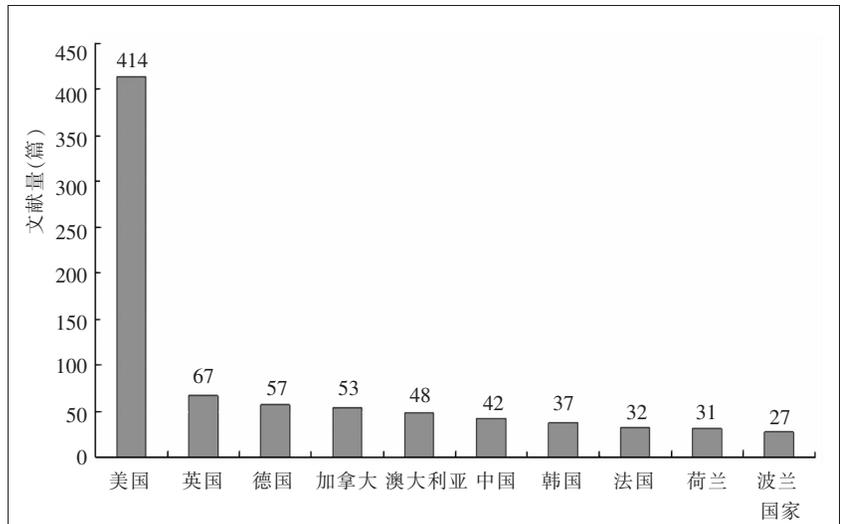


图 2 2001~2010 年乳腺癌易感性相关基因研究论文的区域分布

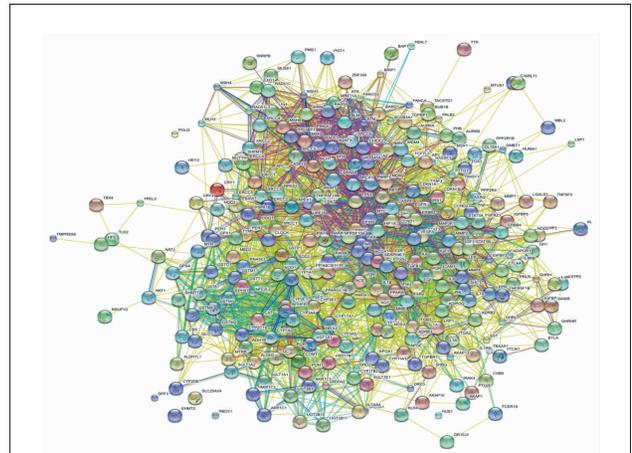


图 3 利用 STRING 软件构建乳腺癌易感性相关基因的相互作用网络图

3 讨论

本文对 2001~2010 年 Embase、Medline/Pubmed 及 BIOSIS Preview 数据库中乳腺癌易感性相关基因研究的文献进行了计量学分析,并利用生物信息学软件对所筛选的基因进行了功能分类及通路分析。对乳腺癌易感性相关基因的功能分析中,我们发现,易感基因涉及 DNA 修复、应激、细胞增殖、细胞周期、凋亡、激素代谢等多个通路。

迄今为止,已经有数种方法确定乳腺癌的易感等位基因,而且根据这些易感位点在人群中的分布频率和患癌风险大小可以分为三类:①高外显率易感等位基因:这类等位基因在人群中极为罕见,其次要等位基因频率(minor allele frequency, MAF) < 0.005 ,其患乳腺癌风险比普通人群高 10 倍以上。②较为罕见的中等外显率等位基因:其 MAF 介于 0.005~0.01 之间,相对于普通人群其患癌风险增加了 2~4 倍。③低外显率等位基因:其 MAF > 0.05 ,其患癌风险为普通人群的 1.5 倍左右^[6]。

传统的连锁分析方法确定了 BRCA1/2 在乳腺癌发病风险中作用。我们目前尚不能完全排除存在类似 BRCA1/2 的其他乳腺癌易感基因 BRCA_x (如 BRCA3, BRCA4 等)。在对家族性乳腺癌人群(家族成员中有 ≥ 4 例早发的乳腺癌患者)的研究中发现,在排除 BRCA1/2 的作用之外,仍提示类似 BRCA1/2 的乳腺癌易感基因(BRCA_x)的存在^[7];另一方面,对非 BRCA1/2 乳腺癌家族的研究未发现 BRCA_x 与乳腺癌的显著连锁信号,这提示即使存在 BRCA_x,也不足以解释非 BRCA1/2 家族性乳腺癌的发病风险^[8]。目前所有针对非 BRCA1/2 乳腺癌家族的研究,都存在一个明显的问题:不能很好解决不同乳腺癌家族的异质性问题。以目前为止最大的一个连锁研究为例,其包含了 149 个非 BRCA1/2 乳腺癌家族,而家族成员中 ≥ 4 例在 50 岁以前被诊断为乳腺癌的家族仅有 24 个,而在 60 岁以前被诊断为乳腺癌的也只有 74 个家族^[9],这大大影响了统计效力,不足以消除不同家庭之间的异质性。基于此,BCLC(Breast Cancer Linkage Consortium)开始着手一项至少包含 250 个非 BRCA1/2 乳腺癌家族的连锁分析研究。

尽管通过病例对照研究及 GWAS 研究,发现了众多的低外显率等位基因,但这些低外显率等位基因的患乳腺癌相对风险仅为 1.3~1.5。目前尚不清楚低外显率等位基因的确切数目及其影响乳腺癌易感性机制。到目前为止,GWAS 研究虽然发现了病例对

照研究所不能发现的乳腺癌易感性基因位点,但其有限的统计效力限制了 GWAS 的应用,因而许多位点可能被遗漏。理想的研究是既达到足够的统计效力,又无需增加样本量和实验费用。将研究对象聚焦于遗传倾向明显的乳腺癌家族(或早期发病家族)将是一种可行的策略^[9]。或者将研究对象按照 BRCA1/2 阳性或阴性、组织病理分期等进行分层,有助于发现新的突变基因或位点。另外,在更多不同人群中进行 GWAS 研究也将有助于发现更多的乳腺癌易感性位点。必须指出的是,尽管目前发现的基因位点仅能解释家族性乳腺癌的一小部分(约占 5%的乳腺癌发病风险),但这种患癌风险很有可能被低估,因为 GWAS 研究中所用到的 SNPs 仅仅是某一组功能 SNPs 的标签 SNP。这就意味着 GWAS 研究很可能遗漏了许多对乳腺癌易感性有相似或更小影响的 SNPs。

随着对更多新的乳腺癌易感性基因或位点的发现及对它们功能和特性的研究,将有助于发现乳腺癌发病过程中新的致病通路,更好地预测乳腺癌发病风险。

参考文献:

- [1] Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world [J]. *J Clin Oncol*, 2006, 24(14): 2137-2150.
- [2] Campeau PM, Foulkes WD, Tischkowitz MD. Hereditary breast cancer: new genetic developments, new therapeutic avenues[J]. *Hum Genet*, 2008, 124(1): 31-42.
- [3] De Grève J, Sermijn E, De Brakeleer S, et al. Hereditary breast cancer: from bench to bedside[J]. *Curr Opin Oncol*, 2008, 20(6): 605-613.
- [4] Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland[J]. *N Engl J Med*, 2000, 343(2): 78-85.
- [5] Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer[J]. *Nat Genet*, 2000, 26(4): 411-414.
- [6] Ghossaini M, Pharoah PD. Polygenic susceptibility to breast cancer: current state-of-the-art [J]. *Future Oncol*, 2009, 5(5): 689-701.
- [7] Simon K, Geigl JB, Pristauz G. Genes beyond BRCA1 and BRCA2 for hereditary breast cancer [J]. *Wien Med Wochenschr*, 2010, 160(19-20): 478-482.
- [8] Smith P, McGuffog L, Easton DF, et al. A genome wide linkage search for breast cancer susceptibility genes [J]. *Genes Chromosomes Cancer*, 2006, 45(7): 646-655.
- [9] Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci[J]. *Nature*, 2007, 447(7148): 1087-1093.