

2001—2010 年乳腺癌预后基因临床研究文献的计量学分析

黄东兰, 谢菲, 岑东芝, 张积仁

Clinical Research Essays on Breast Cancer Prognostic Genes from 2001 to 2010: a Bibliometric Analysis

Huang Donglan, Xie Fei, Cen Dongzhi, Zhang Jiren

Oncology Center, Zhujiang Hospital, Southern Medical University, Guangzhou 510282, China

Corresponding Author: Zhang Jiren, E-mail: zhangjiren@126.com

Abstract: Objective To review the distribution of clinical research essays on breast cancer prognostic genes over the past decade and provide reference for further study. **Methods** We searched for essays in Embase, Pubmed and Biosis Preview database from January 2001 to December 2010. In this bibliometric analysis, all essays were managed with Endnote X3; Data were processed with Excel 2007. **Results** Eight hundred and seventy-six essays were included and the numbers of which grew slowly in the first five years but rapidly in the second one. The essays included were mainly on singular gene, specially on gene *ErbB-2* (157 essays). Though essays on multi-gene model were less than those on singular gene, they gradually increased with the year specially the essays on MammaPrint and Oncotype DX. The country, journal, institution and author with the largest number of essays published was "United States", *Breast Cancer Research and Treatment*, "Netherlands Cancer Institute" and "Bieche I", respectively. **Conclusion** Worldwide researchers pay more and more attention to breast cancer prognostic genes, and searching for its sensitive and specific genes and multi-gene profiles will be an important target for the future study.

Key words: Breast cancer; Prognostic; Genes; Bibliometrics

摘要: 目的 多角度评价近 10 年来国内外乳腺癌预后基因临床研究概况, 为进一步研究提供参考。 **方法** 以 Embase、Pubmed 及 Biosis Preview 为文献检索数据库(2001 年 1 月—2010 年 12 月), 对纳入的 876 篇文章分别对其出版年、国家、期刊、研究机构、作者及所研究基因进行计量学分析。 **结果** 近 10 年来该领域发文量逐渐增多, 且 2005 年后增长幅度明显增大; 发文量最多的国家、期刊、机构及作者分别为美国、《Breast Cancer Research and Treatment》、Netherlands Cancer Institute、Bieche I; 研究内容以非多基因模型研究为主, 其中又以研究 *ErbB-2* 基因最多(157 篇), 多基因模型研究逐年增多, 其中以 MammaPrint 及 Oncotype DX 两模型为主。 **结论** 乳腺癌预后基因临床研究日益受世界的关注, 寻找其敏感和特异的基因及多基因模型为未来的重要研究方向。

关键词: 乳腺癌; 预后; 基因; 文献计量学

中图分类号: R737.9 **文献标识码:** A **文章编号:** 1000-8578(2012)01-0091-04

0 引言

近 20 年来, 乳腺癌早期筛查的开展及综合治疗模式的应用, 乳腺癌的发病率及死亡率有所降低, 但临床上仍缺乏准确预测患者预后的有效方法。目前临床上广泛应用的乳腺癌 TNM 分期与基于 ER、PR、HER2 的分子分型在预后评价上有重要的价值, 但同一分期、同一病理类型、同一分子亚型, 甚至同一治疗方案的乳腺癌患者, 其预后仍然存在明显差

异^[1-2]。随着基因组学研究的逐渐深入, 发现基因检测技术能够提供较 TNM 分期及分子分型更为详尽的预后信息。目前已发现不少较临床病理学指标更能准确预测乳腺癌预后的基因标志物。现通过全面检索乳腺癌预后基因临床研究的相关文献, 采用计量学分析法对这些文献进行量化分析, 多角度评价近 10 年来乳腺癌预后基因相关临床研究的进展。

1 资料与方法

1.1 文献来源

以 Embase、Pubmed 及 Biosis Preview 为检索数据库。为提高文献的查准率, 对不同的数据库采取个

收稿日期: 2011-06-03; 修回日期: 2011-10-12

作者单位: 510282 广州, 南方医科大学珠江医院肿瘤中心

通信作者: 张积仁, E-mail: zhangjiren@126.com

作者简介: 黄东兰(1986-), 女, 博士在读, 主要从事乳腺癌分子预测研究

性化检索:对 Embase 进行 Emtree 扩展检索,辅以自由词的题目及摘要字段的检索;Pubmed 则为 MeSH 检索,同时辅以自由词的题目及摘要字段的检索;Biosis Preview 选择主题检索。检索时间限定为 2001 年 1 月 1 日到 2010 年 12 月 31 日,物种为人类,语种不限。“乳腺癌”的检索词包括(1)EMtree 词“breast cancer”,(2)MeSH 词“breast neoplasm”,(3)自由词检索使用“breast cancer”、“breast tumor *”、“breast neoplasm *”、“breast carcinoma *”、“mammary neoplasm *”;“预后”的检索词不论 EMtree 词、MeSH 词还是自由词检索均使用“prognosis”;“基因”的英文检索词为“gene”。

1.2 文献纳入与排除

纳入标准:(1)与乳腺癌患者预后有关基因及基因产物的研究及系统综述;(2)临床研究(包括肿瘤患者原代细胞产生的结果、根据临床及生物信息学模型推导出的结果);(3)文献形式为论著;(4)临床研究文献、会议论文;(5)文献需包括能提供足够信息的摘要,或可以获取全文资料;(6)当文献同时研究人类的临床研究和动物体内实验或细胞株体外研究时,纳入其中的临床研究数据。排除标准:(1)与预后相关的非基因研究;(2)综述、个案报道及通讯、回信、书籍及社论等;(3)细胞株体外研究、老鼠及其他动物体内实验;(4)针对基因的结构、机制等研究;(5)提供信息不全的会议摘要;(6)不同数据库来源,作者、题目及出版年相同的文献按 1 篇统计;(7)其他与研究目的不符的文献;(8)相同研究内容,既有会议摘要又有论著,以论著进行统计。纳入与排除的过程,两名研究人员独立进行并交叉核对,若相同文献研究员意见不一则由肿瘤临床医生(主任医师)、肿瘤专业研究员等组成的小组讨论决定。

1.3 文献计量分析

采用 Endnote X3 软件对所检索文献管理,并借助 Excel 2007 软件对最终符合纳入标准的文献分别进行出版年、国家、期刊、研究机构、作者及所研究基因进行计量分析。

2 结果

2.1 检索结果

从 Embase 中检出文献 2 104 篇,其中论著 1 438 篇,综述 415 篇,会议论文 70 篇,笔记 43 篇,会议摘要 37 篇,信件 33 篇,社论 31 篇;Pubmed 中检出文献 1 895 篇,其中综述 348 篇,其余为论著;Biosis Preview 中检出文献 1 324 篇,其中论著 978 篇,会议 321 篇,书籍目录 24 篇,信件 5 篇,书籍 1 本。三个数据库共检出文献 5 323 篇,去除重复文

献后为 3 159 篇,最后符合纳入标准的文献为 876 篇。

2.2 文献计量分析结果分析

2.2.1 文献发表年限结果分析 近 10 年乳腺癌预后基因相关文献年代分布见图 1。结果显示乳腺癌预后相关基因论文的数量总体走势为上升趋势,尤以 2005 年后增幅明显加大,由此推测,今后此领域文献仍将会保持一定的增长速度。2005 年及 2009 年发文量相对下滑,仔细研究被排除的文献中,发现这两年的文献并无漏选。故下滑原因一方面考虑为此领域研究在这两年确实无新突破;另一方面可能为预后研究所需时间较长,即使研究有增多但未能及时将研究成果在这两年发表出来。但这两年出现拐点原因是否相同以及各自具体原因尚待进一步研究。

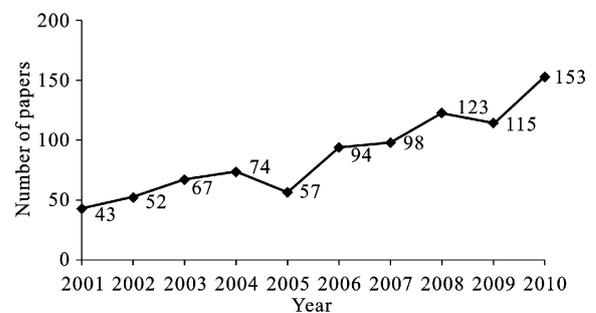


图 1 乳腺癌预后基因相关文献年份分布
Figure 1 Number of papers on breast cancer prognostic genes per year

2.2.2 文献第一作者所在国家的结果分析 统计发现进行该领域研究的国家和地区达 58 个,发文量前十名国家分布,见图 2。美国在此领域发文量最多占该领域发文总量的 17.2%,占据主导地位。并可见前十名国家中,除中国外均为发达国家,这印证了经济发达的地区,其论文的发表率也相对较高。虽然中国在该领域发文量排名第四,但中国为人口大国,其在该领域的人均发文量则难以跻身前十名,可见中国在该领域的研究仍较薄弱。

2.2.3 文献期刊结果分析 纳入的 876 篇论文发表在 211 种生物医学期刊上,平均每刊发文 4.15 篇。发文量排前十名的期刊分布,见图 3。显然以《Breast Cancer Research and Treatment》发文量最多,占总数的 9.68%,《Clin Cancer Res》次之。排前十名的期刊共发文量达 400 篇,已接近总数的一半,且这些期刊均为肿瘤专业杂志。可见,该领域文献刊登的期刊虽然较为分散,但又相对集中,如需快速了解该领域研究动态可首选在上述十种期刊查阅。

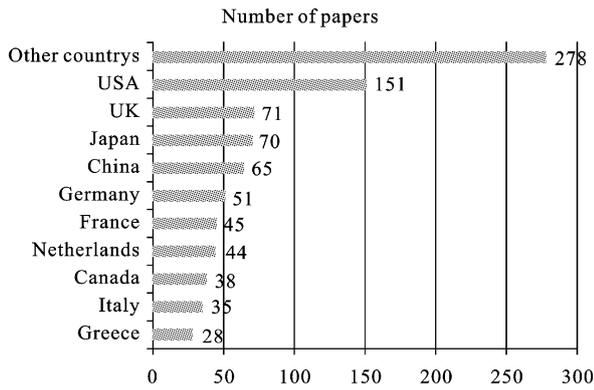


图 2 乳腺癌预后基因相关文献国家分布
Figure 2 Number of papers on breast cancer prognostic genes per country

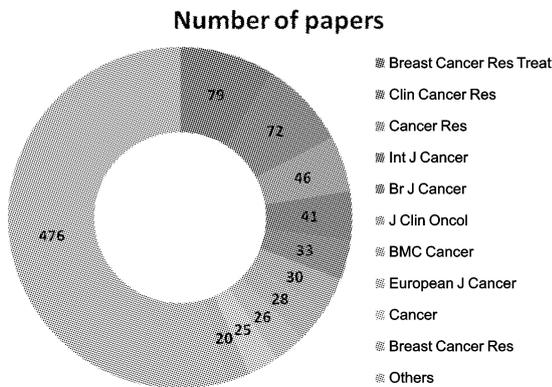


图 3 乳腺癌预后基因相关文献期刊分布
Figure 3 Number of papers on breast cancer prognostic genes per journal

2.2.4 文献的研究机构结果分析 为避免产生重复数据,仅统计通信作者的机构,共统计出 486 个研究机构进行乳腺癌预后相关基因方面研究。其中发文量 ≥ 10 篇的研究机构有 4 个:荷兰的 Netherlands Cancer Institute(19 篇),美国的 MD Anderson Cancer Center(17 篇)及 Vanderbilt University(12 篇),加拿大的 Mount Sinai Hospital(11 篇)。其中 Netherlands Cancer Institute 以研究 MammaPrint 为主,Vanderbilt University 则重视基因多态性的研究。国内相关研究机构发文量均较少,却相对集中分布,以复旦大学(8 篇)及天津医科大学(7 篇)两个机构研究比较多。上述这些机构在该领域研究较多、方法较成熟,是研究员或临床工作者快速了解该领域国内外研究的前沿动态和选择国际间的交流合作机构的重要选择机构。

2.2.5 文献的第一作者结果分析 根据洛特卡和普赖斯对科学家的生产率 and 活动规律的研究^[3],发表论文为 N 篇以上的作者为该学科的杰出科学家,即核心作者,其中 $N = 0.749 \times (\eta_{\max})^{1/2}$,式中 η_{\max} 为

发文量最多的作者发表的论文数量。本研究发现法国的 Bieche I 在 10 年间发文量最大(7 篇),为该领域作出了巨大贡献,并以研究 *cga*、*ccnd1*、*serpinb5*、*erbb1*、*erbb2*、*erbb3*、*erbb4*、*mycn*、*ereg*、*shh*、*brca2*、*dnmt3b*、*ccne1* 等基因为主。由此可知 $\eta_{\max} = 7$,代入上述公式得到 $N = 2$,即发文量 ≥ 2 篇的作者为核心作者。统计发现 743 个第一作者中,核心作者共有 93 个,来自于 26 个不同国家,其中以美国(11 个)和日本(10 个)居多;美国发文量最多的为 DeMichele, A(3 篇),该作者主要研究 *il-6* 及 *tnf-alpha* 基因多态性与乳腺癌预后的关系;日本发文量最多的为 Tsutsui, S(5 篇),重点研究 *egfr*、*erbb2*、*tp53*、*pten*、*kit* 等 5 个基因表达产物与乳腺癌预后的关系;而国内发文量最多的为 Wang, Z. B(3 篇)曾研究过 *esr1*、*pgr*、*erbb1*、*erbb2*、*fhit*、*ccnd1*、*bcl2*、*bax* 等基因的表达产物,并发现乳腺浸润性导管癌患者中,*esr1*、*pgr*、*erbb1*、*erbb2*、*fhit*、*bcl-2* 等基因表达阳性与无疾病生存(DFS)显著相关,*erbb1*、*erbb2*、*fhit*、*ccnd1*、*bcl-2* 等则与总生存(OS)显著相关,而 *bax* 则与 DFS 和 OS 均无关。

2.2.6 文献所研究基因的分析结果 发现近 10 年该领域在非多基因模型研究中提及 545 个基因。近 10 年发文量最多的基因为 *erbb2*,达 157 篇。结合文献发表年限分析,可将该领域研究分为 2 个阶段,即前后 5 年各为一个阶段,发现前 5 年发文量 ≥ 10 篇基因有 *erbb2*(69 篇)、*tp53*(46 篇)、*esr1*(17 篇)、*bcl2*(16 篇)、*pgr*(13 篇)、*brca1*(13 篇)、*egfr*(11 篇)、*nme1*(10 篇)。而后 5 年发文量增多的同时,新增基因数量也随之增加,较前 5 年共增加了 344 个新基因。同时发文量 ≥ 10 篇基因除包含 *erbb2*、*tp53*、*esr1*、*bcl2*、*pgr*、*brca1*、*egfr* 外,新增有 *mki67*、*ccna1*、*abcb1*、*brca2*、*cdkn1a*、*gstp1*、*serpine1*、*pik3ca*。另外较前 5 年研究减少明显的基因为 *nme1*(减至 2 篇)。可见近 5 年来,该领域研究热点基因开始有转移的倾向,尤其是 *ccna1*、*abcb1*、*cdkn1a*、*gstp1*、*serpine1*、*pik3ca* 等基因研究增加显著。

对发文量较多的国家进行基因分析发现,其中美国研究共涉及基因数量多达 125 个,以研究 *erbb2*(29 篇)、*tp53*(9 篇)、*esr1*(8 篇)、*pgr*(10 篇)等 4 个基因为主;英国研究共有 156 个基因,基因研究分布较分散,每个基因的发文量均 ≤ 4 篇,其中发文量为 4 篇的基因为 *tp53*、*psmd9*、*myc*、*ccna1*、*bcl2*;日本研究仅有 68 个基因,同样未见明显集中趋势,以 *erbb2*(7 篇)、*tp53*(6 篇)两者居多;中国研究基因数目与日本接近达 67 个,分布较分散,仅以 *erbb2*(14

篇)及 *esr1*(8 篇)、*pgr*(7 篇)、*nme1*(6 篇)相对较多。

近 10 年来多基因模型研究逐渐增多,从 2001 年的 0 篇到 2010 的 22 篇。纳入标准的文献中共有 99 篇多基因模型研究,涉及 59 种预后预测模型,其中模型基因数目最少者为 2 基因模型,最多者为 368 基因模型。其中以 MammaPrint(23 篇)最多,而 Oncotype DX(9 篇)次之。荷兰作者 VanDe Vijver 等^[4]最早验证 MammaPrint 模型具有预测预后作用。研究 Mammaprint 模型的文献中,共有 13 篇作者研究机构在荷兰的 Netherlands Cancer Institute,除 3 篇文献的作者研究机构分别在美国、韩国、日本外,其余都在荷兰、比利时、法国及西班牙等欧洲国家。中国作为人口最多的亚洲国家尚无数据支持,可见该模型主要研究核心在欧洲,是否适合跨地域及跨种族推广有待进一步研究。关于 Oncotype DX 模型的文献中,5 篇来自于美国。此模型始于 2004 年 Paik 等^[2]提出并证实其在预测预后作用。虽已有研究^[5]对该模型进行经济效益分析并表明可显著增加成本效益比例,但 Oncotype DX 基因芯片检测费用约 3 400 美金^[6],发展中国家仍难以承担,故期待更多更好的低成本技术设备出现,并应用到多基因联合检测的研究中,从而使更多的患者受益。近年来国内研究员也逐渐认识到,建立有效而经济的乳腺癌预后多基因预测模型的必要性。但中国在此领域的研究起步较晚,最早的研究属 Jiang 等^[7]于 2006 年利用并整合 VanDe Vijver 等^[4]研究 MammaPrint 模型的数据,提出 13 基因预测模型。Feng 等^[8]通过小样本数据(60 例)证实 79 基因模型在淋巴结阳性乳腺癌患者中有预测预后价值。另有 Li 等^[9]还提出 24 基因模型可预测化疗后乳腺癌患者的预后。可见国内目前关于乳腺癌预后多基因预测模型的临床研究仍较少,样本量也较小,暂不能提供高质量的循证医学证据。

3 展望

前后 5 年发文量 ≥ 10 篇的基因的乳腺癌预后相关特异性较差,一定程度上提示研究者在该领域的未来研究方向可着眼于寻找乳腺癌特异性预后相关基因。由于细胞增殖、分化各过程均可影响肿瘤疾病预后,且单基因研究往往忽视基因间相互协同或促进作用,故试图通过单个基因全面预测预后的目标较难实现,联合其他基因共同预测来提高预测效能不失为一种较可靠的方法。同时随着 MammaPrint 及 Oncotype DX 模型的预后预测价值逐渐得到证实并商业化,越来越多的研究朝基因模型方向发展。鉴于上述两模型尚未在中国得以大样本验

证,两模型检测费用较高,且检测标本处理要求较高,一定程度上限制了它在中国的应用,故建立系统、科学而具有中国特色的预后相关基因模型是我们未来研究的主要发展方向。而建立多基因模型,寻找有价值的建模候选基因成为关键。随着生物信息学技术的迅速发展,尤其是 hub 蛋白^[10-11]分析法的出现,使之成为现实,但要获得有价值的 hub 蛋白,首先应具备完善的临床标本基因数据库,故该研究经过文献计量分析,初步了解到哪些基因或基因模型的研究比较多、方法比较成熟、预测性能比较稳定,并能为临床标本基因检测提供参考,为进一步临床乳腺癌预后基因数据库的建立奠定基础。

(致谢:值此研究完成之际,谨向张积仁主任、郑燕芳教授、岑东芝博士、姜茂竹博士等组成的研究小组给予的指导与帮助表示感谢!)

参考文献:

- [1] Nohara T, Ryo T, Iwamoto S, et al. Expression of cell-cycle regulator p27 is correlated to the prognosis and ER expression in breast carcinoma patients [J]. *Oncology*, 2001, 60 (1): 94-100.
- [2] Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer [J]. *N Engl J Med*, 2004, 351(27): 2817-2826.
- [3] Radhakrishnan T, Kennzan R. Lotka's Law and Computer Science Literature [J]. *JASIS*, 1979, 30 (1): 51-54.
- [4] Van de Vijver MJ, He YD, Van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer [J]. *N Engl J Med*, 2002, 347(25): 1999-2009.
- [5] Lyman GH, Cosler LE, Kuderer NM, et al. Impact of a 21-gene RT-PCR assay on treatment decisions in early-stage breast cancer: an economic analysis based on prognostic and predictive validation studies [J]. *Cancer*, 2007, 109(6): 1011-1018.
- [6] Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials [J]. *J Clin Oncol*, 2008, 26(5): 721-728.
- [7] Jiang D, Zhao N. A clinical prognostic prediction of lymph node-negative breast cancer by gene expression profiles [J]. *J Cancer Res Clin Oncol*, 2006, 132(9): 579-587.
- [8] Feng Y, Sun B, Li X, et al. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients [J]. *Breast Cancer Res Treat*, 2007, 103(3): 319-329.
- [9] Li LF, Xu XJ, Zhao Y, et al. Integrated gene expression profile predicts prognosis of breast cancer patients [J]. *Breast Cancer Res Treat*, 2009, 113(2): 231-237.
- [10] Borneman AR, Leigh-Bell JA, Yu H, et al. Target hub proteins serve as master regulators of development in yeast [J]. *Genes Dev*, 2006, 20(4): 435-448.
- [11] Shen YZ, Ding YS, Gu Q, et al. Identifying the hub proteins from complicated membrane protein network systems [J]. *Med Chem*, 2010, 6(3): 165-173.

[编辑:周永红;校对:杨 卉]