

曲妥珠单抗治疗乳腺癌疗效相关基因的文献计量学和生物信息学分析

Bioinformatics Analysis of Trastuzumab Related Genes in the Treatment for Breast Cancer//SONG Shu-jie, CEN Dong-zhi, LU Yan-qiang, et al.

宋述杰¹, 岑东芝¹, 鲁衍强², 王贻锴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] 了解曲妥珠单抗治疗乳腺癌疗效相关基因的研究现状及生物学信息。[方法] 检索 Embase、Medline/Pubmed 及 BIOSIS Preview 数据库, 收集曲妥珠单抗治疗乳腺癌疗效相关基因的文獻, 分析其发表年代、国家、期刊、作者及研究机构分布等; 对最终获得的 117 个基因进行生物信息学分析, 分析曲妥珠单抗相关基因及产物相互作用、分类, 并分析关键基因及瓶颈基因。[结果] 81 篇文献符合要求, 共涉及 117 个基因, 所涉及的基因及其产物处于相互联系的复杂网络之中。Bcl-2、EGFR、PIK3CA、PTEN、AKT1 和 ErbB2 同时为 Hub 基因和 Bottleneck 基因。[结论] 曲妥珠单抗治疗乳腺癌疗效相关基因的研究处于蓬勃发展时期, 将会有更多的基因得到关注。

关键词: 乳腺癌; 基因; 曲妥珠单抗; 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)02-0115-04

人类表皮生长因子受体 2 (human epidermal growth factor receptor 2, HER-2) 属于表皮生长因子受体家族, 是一个由原癌基因 *CerbB2*(*Her-2/neu*) 编码的具有受体酪氨酸激酶活性的跨膜糖蛋白。大约 20% 乳腺癌患者存在 HER-2 基因扩增, 这些患者通常更容易出现脑转移^[1]、预后差^[2]。曲妥珠单抗(商品名: 赫赛汀, herceptin) 是以 HER-2 为靶点的人源化单克隆抗体, 能特异地作用于 HER-2 过度表达的乳腺癌细胞, 接受曲妥珠单抗治疗的 HER-2 阳性转移性乳腺癌患者的有效率和总生存率明显提高, 目前已经成为治疗 HER-2 阳性乳腺癌患者的基本药物。现我们对目前曲妥珠单抗疗效相关基因的研究进行生物信息学分析。

1 材料与方法

1.1 文献检索

检索 Embase、Medline/Pubmed 及 BIOSIS Pre-

view 3 个数据库。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*”。曲妥珠单抗的英文检索词包括: ① MeSH 词“trastuzumab”, ② EMtree 词检索使用“trastuzumab”, ③ 自由词“trastuzumab”、“Herceptin”。基因的英文检索词为“gene”。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。物种限定为人类, 时间限定为 2001 年 1 月~2010 年 11 月, 语种不限。

1.2 文献纳入与排除

纳入全部与乳腺癌患者曲妥珠单抗疗效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 排除与药物代谢等相关

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

基因的研究；④排除与曲妥珠单抗疗效相关的非基因研究；⑤纳入临床研究文献，排除针对基因的结构、机制等研究；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。纳入与排除的过程，由2名研究人员独立进行，最终结果互相比较，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

对最终纳入的文献进行文献计量学分析，分析其出版年、作者、国家、研究机构、发表期刊分布。

1.4 曲妥珠单抗疗效相关基因生物信息学分析

利用在线软件 STRING (<http://string-db.org/>) 构建曲妥珠单抗疗效相关基因表达产物蛋白质相互作用网络；利用在线软件 GATHER (<http://gather.genome.duke.edu/>) 按照 Gene Ontology (GO) 分类的标准对曲妥珠单抗疗效相关基因进行功能分类，将 STRING 软件构建的曲妥珠单抗疗效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。

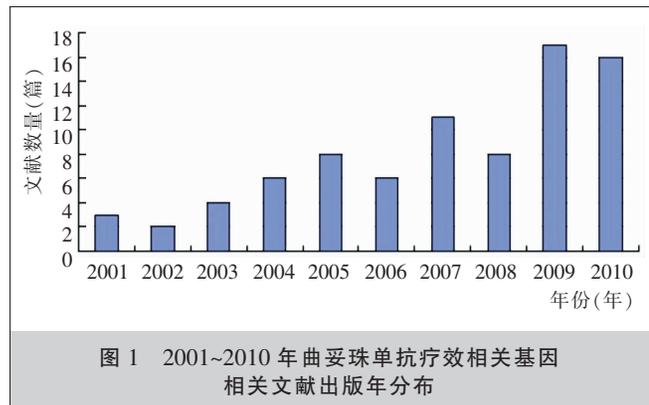
2 结果

2.1 检索结果

Embase 数据库检出文献 1 048 篇，Pubmed/Medline 数据库检出文献 403 篇，BIOSIS Preview 数据库检出文献 649 篇，3 个数据库共检出文献 2 100 篇，符合要求的文献共 81 篇。

2.2 文献计量学分析

曲妥珠单抗疗效相关基因文献在 2001~2008 年间增长缓慢，从 2009 年开始明显增加。见图 1。81 篇论文发表在 34 种生物医学期刊上，平均每刊发文 2.38 篇，发文量高于平均值的期刊为 *Journal of Clinical Oncology* (8 篇)，*Breast Cancer Research and Treatment* (6 篇)，*Cancer Research* (6 篇)，*Annals of Oncology* (5 篇)，*Clinical Cancer Research* (5 篇)，*EJC Supplements* (4 篇)，*Proceedings of the American Association for Cancer Research Annual Meeting* (4 篇)，*Anticancer Research* (3 篇)，*British Journal of Cancer* (3 篇)，*Journal of the National Cancer Institute* (3 篇) 和 *Klinicka Onkologie* (3 篇)。



共有 19 个国家发表了关于曲妥珠单抗疗效相关基因方面的研究，其中发文量超过 1 篇的共有 13 个国家，其中美国最多达 28 篇，其次为意大利 (13 篇) 和法国 (9 篇)，澳大利亚 (4 篇)、德国 (4 篇) 和日本 (3 篇) 超过 2 篇。

经统计得知，Kostler WJ 为发文最多的作者，发表相关文献 3 篇，10 年间在曲妥珠单抗疗效相关基因方面作者发文量超过 2 篇的作者还有 7 位，分别为 Fabi A, Guiu S, Gullo G, Lipton A, Morrison LE 和 Sorensen B。

发文量达 2 篇或以上的机构为 University of Texas MD Anderson Cancer Center (5 篇)，Penn State Hershey Medical Center (4 篇)，Regina Elena Cancer Institute (4 篇)，University Hospital of Vienna (3 篇)，Georges-Francois Leclerc Cancer Center (3 篇)，Jules Bordet Institute (2 篇)，Department of Clinical Biochemistry Aarhus University Hospital (2 篇)，Masarykuv Onkologicky Ustav V Brne (2 篇)，Vall d'Hebron University Hospital and Research Institute (2 篇) 和 Istituto Clinico Humanitas IRCCS (2 篇)。

2.3 曲妥珠单抗疗效相关基因及产物相互作用

STRING 是由 CPR、EMBL、SIB、KU、TUD 和 UZH 共同研发的、描述已知和预测的蛋白质相互作用的数据库，这些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，它们来源于基因组信息分析预测的结果、高通量实验、共表达研究以及已有文献报道的相互作用蛋白信息。81 篇文献中共涉及 117 个基因。采用 STRING 软件对这 117 个曲妥珠单抗疗效相关基因表达产物构建蛋白质相互作用网络 (图 2)。据图 2 可知，这些基因表达产物大部分存在着密切的相互作用，有多个基因处于网络的中心，它们相互之间以及与其他基因产物的关系均

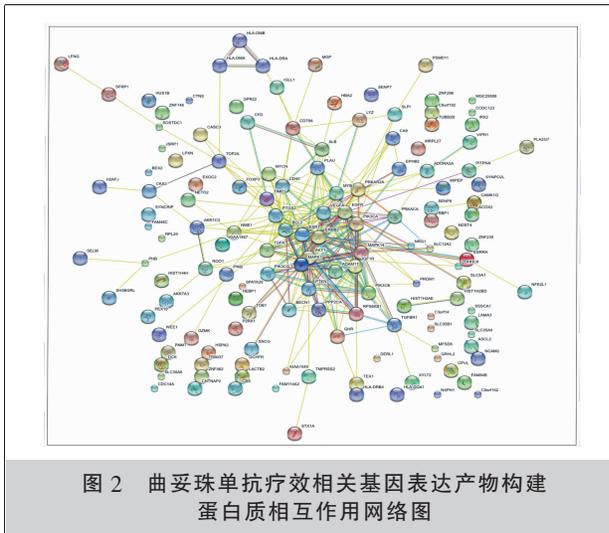


图2 曲妥珠单抗疗效相关基因表达产物构建蛋白质相互作用网络图

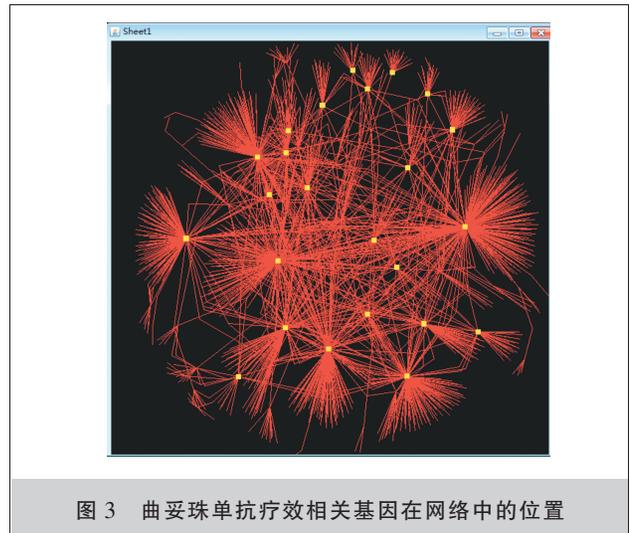


图3 曲妥珠单抗疗效相关基因在网络中的位置

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,红色方框内是目前研究最多的曲妥珠单抗相关基因,它们分别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 曲妥珠单抗疗效相关基因及产物分类

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^[3]。GO 的数据库中有 3 大独立的 Ontology 被建立起来:生物过程、分子功能及细胞组分。而这三个 Ontology 下面又可以分别独立出不同的亚层次,层层向下构成一个树型分支结构。117 个曲妥珠单抗疗效相关基因共涉及不同分级的分类 487 类,其中贝叶斯因子 ≥ 0 的分类有 53 类,贝叶斯因子 ≥ 2 的分类有 13 类,主要富集于细胞增殖、细胞周期调控、蛋白氨基酸磷酸化、一氧化氮代谢、细胞周期、磷酸化、细胞生理过程负调控、生理过程负调控、酶耦联受体蛋白信号通路、细胞分化、跨膜受体蛋白、酪氨酸激酶信号通路、磷代谢、角化细胞分化等功能分类。

2.5 关键基因及瓶颈基因分析

将 STRING 软件构建的曲妥珠单抗疗效相关基因相互作用网络图导入 Cytoscape 软件可视化(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 85 个节点 667 条边组成,直径为 7,网络的平均距离为 2.62。度为节点连接的边的总数目,最大度值为 70,最小为 2,平均为 15.69 ± 18.44 ,网络中连接度高的节点对网络的稳定性起到重要的作用,它们的功能改变可能对整个网络至关重要。本研究规定度数 \geq 均数+1SD(标准差)的蛋白

所对应的基因为关键基因(Hub 基因),共有 12 个。Betweenness 为网络中所有的最短路径中经过该节点的数量比例,最大 Betweenness 值为 1 008.47,最小为 0,平均为 144.28 ± 231.39 。相应的节点或者边在整个网络中的作用和影响力。Betweenness \geq 均数+1SD 的蛋白所对应的基因为瓶颈基因(Bottleneck 基因),共有 9 个。同时为 Hub 基因和 Bottleneck 基因对曲妥珠单抗疗效相关基因的网络至关重要,具体为 Bcl-2、EGFR、PIK3CA、PTEN、AKT1、ErbB2。

3 讨论

以曲妥珠单抗为基础的治疗极大地提高了 HER-2 阳性转移性乳腺癌患者的有效率和总生存率,缩短了肿瘤进展时间^[4]。目前曲妥珠单抗治疗早期 HER-2 阳性乳腺癌患者已经成为常规,然而现在仍然无法解释即使是 HER-2 高表达患者,也会有接近一半的患者对之不敏感或在治疗中出现耐药^[5],这之中包含多种作用机制^[6]。即使是一开始对曲妥珠单抗敏感的患者,大部分也会在一年内产生耐药^[7]。因而通过分子医学的方法来筛选曲妥珠单抗敏感的患者显得尤为重要,以指导临床医生有针对性用药,从而实现个体化治疗。

运用生物信息学方法对检索文献中涉及的 117 个基因进行分析,分析曲妥珠单抗疗效相关基因及产物相互作用,这些基因表达产物大部分存在着密切的相互作用关系,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,可能是曲妥珠单抗疗效相关基因中的关键基因,对未来该领域研究有重要的参考意义,避免了人力物力的浪费,有利于资源的最优配置。通过利用在线软件 GATHER 按照 GO 分类的标准对其进行功能分类可知,曲妥珠单抗疗效相关基因主要富集于细胞增殖和细胞周期调控等功能。对关键基因及瓶颈基因进行分析,认为同时对关键基因和瓶颈基因对曲妥珠单抗疗效相关基因的网络至关重要,将是该领域未来的研究方向,有利于科研工作者的选题。

近 10 年中,前 5 年的研究主要集中在 HER-2,其他的基因的研究比较松散;近 5 年的研究主要集中在 HER-2、PTEN、PIK3CA、AKT、EGFR、IGF1R、CA153、UPA、HER-3、HER-4 和 CEA 等。因此,我们预测随着基因芯片的大规模应用,还会有更多的曲妥珠单抗疗效相关基因被发现。同时,近 5 年的研究热点基因仍然会在一段时间内继续为热点。

曲妥珠单抗在肿瘤细胞中的作用机制和耐药机制尚未完全阐明,我们需要更深入的基础和临床研究来进一步更好地选择曲妥珠单抗靶向治疗的目标人群和最佳联合治疗方案。寻找更加特异性的预测基因或者多个基因组合来预测曲妥珠单抗靶向治疗的敏感性已经成为提高曲妥珠单抗治疗有效率的关键。

参考文献:

[1] Duchnowska R, Szostakiewicz B. Correlation between

quantitative HER2 protein level and the risk of brain metastasis (BM) in patients (pts) with metastatic breast cancer (MBC) treated with trastuzumab-containing therapy [J]. *J Clin Oncol*, 2010, 28(15S): Abstr 1030.

- [2] Hudelist G, Köstler WJ, Czerwenka K, et al. Her-2/neu and EGFR tyrosine kinase activation predict the efficacy of trastuzumab-based therapy in patients with metastatic breast cancer[J]. *Int J Cancer*, 2006, 118(5): 1126-1134.
- [3] Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures [J]. *Bioinformatics*, 2006, 22(23): 2926-2933.
- [4] Svoboda M, Grell P, Simícková M, et al. A retrospective analysis of trastuzumab-based therapy in metastatic breast cancer patients at Masaryk Memorial Cancer Institute. Identification of predictive factors [J]. *Klin Onkol*, 2008, 21(6): 348-358.
- [5] Berns K, Horlings HM, Hennessy BT, et al. A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer[J]. *Cancer Cell*, 2007, 12(4): 395-402.
- [6] Végran F, Boidot R, Coudert B, et al. Gene expression profile and response to trastuzumab-docetaxel-based treatment in breast carcinoma [J]. *Br J Cancer*, 2009, 101(8): 1357-1364.
- [7] Leitzel K, Hou HY, et al. Use of pretreatment serum CA9 (carbonic anhydrase 9) to predict PFS and survival in trastuzumab-treated metastatic breast cancer[J]. *J Clin Oncol*, 2009, 27(15S): 11092.