

宫颈癌淋巴结转移相关基因的表达谱研究

朱秀兰¹, 马文丽^{1*}, 阴常欣¹, 郑文岭^{1,2}

(1. 南方医科大学 基因工程研究所, 中国广东 广州 510515 ; 2. 华南基因组中心, 中国广东 广州 510800)

摘要: 淋巴结转移是影响宫颈癌预后因素的首要因素, 但分子机理尚不清楚, 探讨宫颈癌淋巴结转移的分子机制, 筛选可能用于临床诊断和治疗的新的分子靶标尤其重要. BRB-Array Tools 软件对基因表达综合数据库 (Gene Expression Omnibus, GEO) 中有/无淋巴结转移早期宫颈癌的基因芯片表达谱数据进行统计学分析, 找出二者的差异基因, 利用 GenCLIP 软件进行文献挖掘研究这些差异表达基因的功能关系. 筛选出 71 个差异表达基因, 其中 35 个在有淋巴转移的宫颈癌中表达上调, 36 个下调. 文献挖掘发现 *FZD10*、*PDZK1IP1*、*SERPINB9*、*COMP*、*MAGEA1* 等几个新的宫颈癌淋巴结转移相关基因. 结果表明生物信息学方法能有效地分析基因芯片数据并获取生物内在信息, 为宫颈癌淋巴转移发生机制及治疗靶位开辟新思路.

关键词: 宫颈癌; 淋巴转移; 基因芯片; 差异表达

中图分类号: R737

文献标识码: A

文章编号 1007-7847(2009)04-0337-06

Study on Gene Expression Profilings of Lymph Node Metastasis in Cervical Cancer

ZHU Xiu-lan¹, MA Wen-li^{1*}, YIN Chang-xin¹, ZHENG Wen-ling^{1,2}

(1. Institute of Genetic Engineering, Southern Medical University, Guangzhou 510515, Guangdong, China;

2. Southern Genomics Research Center, Guangzhou 510800, Guangdong, China)

Abstract: Lymph node metastasis is the most important factor that can affect the prognosis of cervical cancers, but as far as the molecular mechanism is unclear yet, so to investigate the molecular pathogenesis of cervical cancer with lymph node metastasis for further identifying gene markers that may be useful in the diagnosis and treatment of the disease are significantly vital. The dataset of gene expression profilings of cervical cancer with and without lymph node metastasis was taken from GEO database and analyzed by BRB-Array tools. Then the analysis of the differentially expressed genes by the GenCLIP software was followed. 71 differentially expressed genes were identified, 35 of which were up-regulated and 36 down-regulated in cervical cancer with lymph node metastasis. Several genes (*FZD10*, *PDZK1IP1*, *SERPINB9*, *COMP*, *MAGEA1*, etc.) with potential relation to metastatic cervical cancer were identified. The results showed that bioinformatics had a high efficiency in analyzing gene chip data and revealing internal biology information. It will offer a new view for pathogenesis and treat target in cervical cancer with lymph node metastasis.

Key words: cervical cancer; lymph node metastasis; gene chip; differential expression

(*Life Science Research*, 2009, 13(4) 337~342)

收稿日期: 2008-11-14; 修回日期: 2009-05-08

基金项目: 广东省重点实验室建设基金资助项目(2004B60144)

作者简介: 朱秀兰(1980-), 女, 安徽砀山人, 硕士研究生, 医师, 从事基因芯片表达谱的研究与应用; *通讯作者: 马文丽(1964-), 女, 广东广州人, 南方医科大学基因工程研究所教授, 博士, 主要从事基因组治疗与基因分离中的 DNA 芯片技术, Tel: 020-62789098, E-mail: wenli668@gmail.com.

宫颈癌是常见的妇科恶性肿瘤之一,其发病率在女性恶性肿瘤中居第 2 位,仅次于乳腺癌,全世界每年有 20 多万妇女死于该疾病,新发病例中,83%发生在发展中国家,我国每年新发病例 13.15 万,占世界宫颈癌新发病例的四分之一,居我国女性恶性肿瘤第一位,且有年轻病例逐年增加的趋势。宫颈癌筛查的普遍开展,使得早期诊断成为可能,现今宫颈癌的预后较之以往有了很大改善,尽管如此,在一组具有高危因素的宫颈癌患者中,预后并不理想。诸多影响宫颈癌预后的因素中,首要的是淋巴结转移。据文献报道,无淋巴结转移者 5 年生存率为 82.2%,有淋巴结转移者 5 年生存率为 64%~68.2%^[1],与宫颈癌的发病机制相比,对其淋巴转移的研究进展相对缓慢。分子生物学的发展,使我们可以从分子水平上揭示疾病的发生发展机制。同时,基因芯片作为一种高效、大规模地获取生物信息的技术,可对不同个体、不同组织、不同病变等情况下的 mRNA 表达情况进行检测与分析。本研究利用基因表达综合数据库(GEO)^[2]中的基因芯片数据,对有/无淋巴结转移的宫颈癌基因表达谱进行分析,以期进一步探讨宫颈癌淋巴结转移的机制,为预测和防治宫颈癌淋巴结转移提供新的分子理论基础。

1 材料和方法

1.1 材料

1) 宫颈癌表达谱数据:来源于 NCBI 的 GEO 数据库 GSE7410,由 Petra Biewenga 等提交;2) 基因芯片:该实验的芯片平台是 GPL1708,为 Agilent 公司的商业芯片 Agilent-012391 Whole Human Genome Oligo Microarray G4112A;3) BRB-Array Tools 软件包由 Richard Simon 等研发,在线 <http://linus.nci.nih.gov/BRB-ArrayTools.html> 下载此软件包;4) GenCLIP 软件由南方医科大学黄仲曦教授研发,在线 <http://www.genclip.com/> 下载此软件。

1.2 方法

1.2.1 宫颈癌表达谱数据

宫颈癌表达谱数据来源于 NCBI 的 GEO 数据库 GSE7410,由 Petra Biewenga 等提交,本研究采用有淋巴转移宫颈癌和无淋巴转移的宫颈癌数据共 40 个样本,其中有淋巴转移 19 个样本,包括 IB1 期 4 例,IB2 期 4 例,IIA 期 8 例和

3 个重复样本,无淋巴转移 21 个样本,包括 IB1 期 13 例,IB2 期 3 例,IIA 期 3 例和 2 个重复样本。所有样本均经病理证实,并于取材后立即投入液氮中冻存储存。

1.2.2 基因芯片

该实验的芯片平台 GPL1708,为 Agilent 公司的商业芯片 Agilent-012391 Whole Human Genome Oligo Microarray G4112A,其具体操作方案详见网址 <http://www.agilent.com/>。该芯片采用原位杂交寡核苷酸芯片技术,具有高分辨率、高性能的特点,可全盘观测整个基因组和 DNA 分子轮廓,提供全面的基因组信息。

1.2.3 RNA 的提取、探针标记和杂交

Trizol 一步法提取组织总 RNA,样品标记按照 Agilent 线性荧光扩增标记试剂盒说明书进行,实验组样品用 Cy5 标记,共同参照样品种用 Cy3 标记,参照样品种为所有人类肿瘤组织中总 RNA 混合而成的参照样品种 RNA 池。用 Agilent 公司的原位杂交试剂盒进行杂交。

1.2.4 图像扫描与数据采集

Agilent Technologies Scanner G2505A US-14702369 扫描仪对芯片进行扫描,然后用 Agilent Feature Extraction (version A.7.5) 软件处理图像数据,标准化信号,调校背景后输出数据,以 TXT 格式将数据上传至 GEO 数据库。

1.2.5 筛选差异表达基因

从 GEO 数据库下载上述 40 个样本数据,解压缩后放入同一个文件夹,然后上传至 BRB-Array Tools 进行在线分析。为使数据便于分析,同时获得质量高、表达信号强的基因,利用软件包中 Filter 工具的中位值方法对芯片数据进行过滤和标准化,过滤参数为: Cy3、Cy5 的荧光信号强度大于 100,芯片上点的直径大于 10,并对不符合以下标准的基因进行滤除:1) 两类样本的基因中位数值至少发生 2 倍的变化,且这种变化不少于 20%的样本数;2) 基因表达数据缺失的样本数不超过 50%;3) log-ratio 百分位数大于 75%。上传成功后,对通过过滤标准的基因进行非配对样本 *t* 检验。19 个有淋巴转移的样本为一类,21 个无淋巴转移的样本为另一类,利用 Class comparison 工具对两类样本分类对比, $P < 0.05$,找到有淋巴转移和无淋巴转移宫颈癌组织间的差异表达基因。

1.2.6 差异表达基因的文献挖掘

找出上述差异表达基因后,用 GenCLip 软件分析这些已知名称差异基因之间的功能关系. 先从 Medline 下载与基因有关的摘要,并提取出大于设定滤过阈值的关键词(阈值 = $t + (k/n)$, t 和 k 为主观设定的直接影响结果和噪声水平的指标,本研究采用软件默认值, n 为给定基因的相关摘要数目),再通过聚类分析等寻找基因与特定关键词之间有显著性生物学意义的关联,同时手动删除在文献中普遍存在和极罕见的关键词,并在

词列表中加入“cervix”关键词,过滤出已研究过的宫颈癌基因,进而发现新的转移相关基因^[3].

2 结果

2.1 差异表达基因

通过对芯片数据过滤和标准化,1 178 个基因通过了过滤标准. Class comparison 工具对两类样本分类对比后,共找到 71 个差异表达基因,上调的 35 个,下调的 36 个,见表 1.

表 1 有/无淋巴结转移宫颈癌差异表达基因

Table 1 Differentially expressed genes in cervical cancer with and without lymph node metastasis (with /without)

No.UniGene	Gene code	No.UniGene	Gene code	No. UniGene	Gene code
Up-regulate					
Hs.25333	<i>ILIR2</i>	Hs.512682	<i>CEACAM1</i>	Hs.95162	<i>KCNQ1</i>
Hs.592211	<i>KLHDC7B</i>	Hs.89603	<i>MUC1</i>	Hs.73793	<i>VEGF</i>
Hs.632144	<i>SAA1</i>	Hs.459072	<i>FAM108C1</i>	Hs.352018	<i>TAP1</i>
Hs.522109	<i>SLC6A14</i>	Hs.431099	<i>PDZK1IP1</i>	Hs.207631	<i>LMO7</i>
Hs.1955	<i>SAA2</i>	Hs.458425	<i>GPR109B</i>	Hs.507658	<i>ALOX5AP</i>
Hs.438292	<i>TNS4</i>	Hs.656294	<i>ZC3H12A</i>	Hs.696035	<i>RAB11FIP1</i>
Hs.270833	<i>AREG</i>	Hs.591845	<i>TUSC3</i>	Hs.654801	<i>AUTS2</i>
Hs.592069	<i>HAS3</i>	Hs.170673	<i>RDHE2</i>	Hs.631988	<i>DDR1</i>
Hs.194710	<i>GCNT3</i>	Hs.104879	<i>SERPINB9</i>	Hs.499205	<i>IRX3</i>
Hs.200412	<i>EPPK1</i>	Hs.290207	<i>ETS2</i>	Hs.89901	<i>PDE4A</i>
Hs.497636	<i>LAMB3</i>	Hs.487046	<i>SOD2</i>	Hs.443577	<i>TNFRSF21</i>
Hs.31664	<i>FZD10</i>	Hs.501574	<i>ADAM8</i>		
Down-regulate					
Hs.198760	<i>NEFH</i>	Hs.443752	<i>TAGLN</i>	Hs.696468	<i>RBMS3</i>
Hs.116724	<i>AKR1B10</i>	Hs.72879	<i>MAGEA1</i>	Hs.118727	<i>HES2</i>
Hs.592283	<i>STAG3</i>	Hs.274313	<i>IGFBP6</i>	Hs.104925	<i>ENC1</i>
Hs.521212	<i>AKR1B1</i>	Hs.472877	<i>EYA2</i>	Hs.133892	<i>TPM1</i>
Hs.116419	<i>MEI1</i>	Hs.487200	<i>SMOC2</i>	Hs.594952	<i>DES</i>
Hs.460109	<i>MYH11</i>	Hs.197922	<i>CAMK2N1</i>	Hs.472877	<i>EYA2</i>
Hs.554822	<i>UGT1A6</i>	Hs.21160	<i>ME1</i>	Hs.701977	<i>NR2F2</i>
Hs.2006	<i>GSTM3</i>	Hs.154654	<i>CYP1B1</i>	Hs.250666	<i>HES1</i>
Hs.202676	<i>SYCP2</i>	Hs.437422	<i>EPB41L1</i>	Hs.510281	<i>SLC24A3</i>
Hs.516105	<i>ACTG2</i>	Hs.8867	<i>CYR61</i>	Hs.441783	<i>AHNAK2</i>
Hs.180142	<i>CALML5</i>	Hs.1584	<i>COMP</i>	Hs.515130	<i>VANGL1</i>
Hs.160652	<i>IGF1</i>	Hs.644877	<i>TBL1XR1</i>	Hs.654465	<i>GCLC</i>

2.2 上调基因的文献挖掘分析结果

从 Medline 下载与上调基因有关的摘要,并提取关键词,聚类结果显示基因按词的发生模式聚成不同的类,上调基因的聚类结果见图 1. 横轴为基因相关文献中的关键词,纵轴为基因名称,图 1 中不同区域代表该关键词在对应基因的文献中出现的频率,颜色越淡,代表该词出现的频率越高,即该基因与该关键词间的关系越密切. 图 1 显示 *SAA1*、*FZD10*、*PDZK1IP1*、*SERPINB9*

这几个基因在转移的肿瘤中高表达,但未见宫颈癌相关研究或研究极少.

2.3 下调基因的文献挖掘分析结果

下调基因的聚类结果见图 2,解释同图 1. 图 2 中显示 *COMP*、*GSTM3*、*IGFBP6*、*MAGEA1* 这几个基因在转移肿瘤中低表达,但在宫颈癌中未进行过相关研究或研究极少,因此,这些新的宫颈癌转移相关基因值得进一步研究.

大的开放性数据库,用户可提交、储存和检索多种形式的数据库,并免费使用,如基于微阵列技术实验测量得到基因表达谱数据、基因组 DNA 和蛋白质分子数据以及非阵列技术的基因表达序列分析(serial analysis of gene expression, SAGE)和质谱法原生技术得到的数据等。本研究利用生物信息学方法对该数据库中有关宫颈癌及其淋巴结转移的芯片数据进行了分析和文献挖掘,以期发现宫颈癌淋巴结转移相关的差异表达基因及其内在生物关联。

BRB-Array Tools package 是一个集合了诸多基因芯片分析工具的软件包,对双通道及单通道数据都能进行分析,且对常用的.CEL、.GPR和.TXT 格式的数据有快捷的输入途径,它能对各种基因芯片数据进行整理、注释、显示、过滤、标准化等预处理,并能对芯片数据进行散点图分析、聚类分析、分类对比分析、分类预测、二进制树形网络预测、数量性状分析、生存分析等分析。它还可以与 NCBI 的 Clone、GenBank、Unigene 等数据库实时链接。本研究利用其筛选了有无淋巴结转移宫颈癌表达谱芯片之间的差异表达基因。

GenCLIP 软件是基于文献轮廓和特定关键词之间的关系对基因功能聚类的工具,在 Medline 数据库中,每个关键词及其出现频率类似于微阵列中的基因及其表达水平,因此该软件利用基因表达谱聚类分析的方法对关键词聚类,大大提高了文献挖掘的效率,能更有效地发现新的生物信息。本研究利用该软件筛选出几个新的宫颈癌淋巴结转移相关基因。

在有淋巴结转移宫颈癌上调的基因中, SAA1 编码急性期反应蛋白,损伤、感染、炎症、肿瘤发生时此蛋白在血液中水平升高,参与细胞的黏附、迁移、增生及聚集,在很多生理和病理过程中均有重要作用。与正常宫颈组织和原位癌相比, SAA1 在浸润癌中表达上调^[5],我们的研究表明,在有淋巴结转移的宫颈癌中 SAA1 亦呈上调表达,其表达趋势和肿瘤等级呈正相关,与在结肠癌中表达情况相一致^[6]。 FZD10 属于 frizzled 基因家族,编码含 7 个跨膜区的跨膜蛋白,是 Wnt 蛋白的受体蛋白,参与 G 蛋白偶联受体蛋白信号转导途径、卷曲信号转导途径,与多种肿瘤的发生和发展有关^[7,8],研究发现其与滑膜肉瘤的关系尤其密切, Northern blotting 和免疫组化

证实 FZD10 在几乎所有的滑膜肉瘤组织和癌细胞株中高表达。用 RNAi 技术干扰滑膜肉瘤细胞中 FZD10 的表达可明显抑制肿瘤细胞的生长^[9],此方法已用于该疾病的临床治疗。 PDZK1IP1 编码分子质量为 17 kDa 的非糖基化的膜相关蛋白,定位于细胞膜和高尔基复合体, N 端疏水区含有由 13 个氨基酸构成的 PDZ 结合域和两个跨膜区域,在卵巢癌和前列腺癌中与肿瘤的等级呈正相关^[10]。 Sugarbaker D J 等^[11]研究发现 PDZK1IP1 的点突变与恶性胸膜间皮瘤的发生有关。 SERPINB9 是一种丝氨酸蛋白酶抑制因子,可结合到丝氨酸蛋白酶并使之失活。高表达 SERPINB9 的黑色素瘤的病人提示预后不良,可能与肿瘤细胞的免疫逃逸机制有关^[12],但在急性髓细胞样白血病中, SERPINB9 表达下调^[13],具体机制有待于进一步研究。

在有淋巴转移宫颈癌下调的基因中, COMP 编码一种非胶原的细胞外基质蛋白,含有 5 个同源体糖蛋白亚基,每个亚基均有 EGF 样和钙离子结合域,分子质量 524 kD。基因本体分析表明该基因参与细胞黏附过程,其表达下调使细胞间的黏附减弱,肿瘤细胞易于脱落进而导致浸润和转移。 GSTM3 在对外源性物质、环境因素及癌化化合物的解毒过程中发挥重要作用,含有野生型纯合子 GSTM3 的个体能明显抵抗癌的发生,突变的 GSTM3 能增加对食管癌和宫颈癌的易感性^[14,15],该基因具有潜在的转移抑制作用。 IGFBP6 定位于 12q13,含有 4 个外显子,3 个内含子,可调节体外培养的癌细胞中 IGFs 的促长效能。研究发现 IGFBP6 在头颈部癌中表达上调,但当进展到转移性肿瘤时该基因的表达关闭^[16], IGFBP6 在肿瘤中的双向表达机制有待于阐明。具有双向表达的基因还有 MAGEA1, MAGEA1 的表达产物是一种肿瘤排斥抗原,可在细胞内加工产生抗原肽,并经 HLA-1 类分子递呈,被自身细胞毒性 T 淋巴细胞识别,特异性的杀伤相应肿瘤细胞。该蛋白见于多种肿瘤组织及正常的睾丸组织,而其他正常组织不表达^[17]。鉴于其存在肿瘤特异性表达及其表达产物 CTL 识别的特性,利用 MAGEA1 抗原制备肿瘤疫苗可用于多种恶性肿瘤的早期诊断和治疗。

总之,我们用新的方法对宫颈癌及其淋巴转移基因表达谱芯片进行了高效的生物信息学分析,从而大规模地获取了生物内在信息。分析结

果显示 *TNS4*、*KCNQ1*、*VEGF*、*IRX3*、*SOD2*、*TNFRSF21*等基因在肿瘤中表达上调,而 *TPM1*、*MYH11*、*IGF1*等在肿瘤中表达下调,与文献报道一致。另外通过 GenCLIP 软件对文献挖掘发现了宫颈癌上调相关基因 *SAA1*、*FZD10*、*PDZK1IP1*、*SERPINB9* 以及下调相关基因 *COMP*、*GSTM3*、*IGFBP6*、*MAGEA1*, 针对这些基因在宫颈癌中的文献报道缺乏或者研究极少,从而可作为新的宫颈癌淋巴结转移候选基因,为进一步明确宫颈癌淋巴结转移的分子生物机制提供了理论依据和治疗研究方向,为临床提供新的候选治疗靶点。同时,由于基因芯片所产生的结果很多,不可避免的出现一些假阳性结果,因而需要进一步的实验验证及更深入的研究,这将作为下一步实验计划的重点。

参考文献 (References):

- [1] SAKURAGI N. Up-to-date management of lymph node metastasis and the role of tailored lymphadenectomy in cervical cancer[J]. *Int J Clin Oncol*, 2007, 12(3):165-175.
- [2] 刘华, 马文丽, 郑文岭. GEO (Gene Expression Omnibus): 高通量基因表达数据库[J]. *中国生物化学与分子生物学*. (LIU Hua, MA Wen-li, ZHENG Wen-ling. GEO (Gene Expression Omnibus): High through gene expression omnibus[J]. *Chinese Journal of Biochemistry and Molecular Biology*), 2007, 23(3) 236-244.
- [3] HUANG Z X, TIAN H Y, HU Z F, *et al.* GenCLIP: a software program for clustering gene lists by literature profiling and constructing gene co-occurrence networks related to custom keywords[J]. *BMC Bioinformatics*, 2008, 9:308.
- [4] FRUMOVITZ M, RAMIREZ P T, LEVENBACK C F. Lymphatic mapping and sentinel lymph node detection in women with cervical cancer[J]. *Gynecol Oncol*, 2008, 110(3 Suppl 2):17-20.
- [5] SONG J Y, LEE J K, LEE N W, *et al.* Microarray analysis of normal cervix, carcinoma *in situ*, and invasive cervical cancer: identification of candidate genes in pathogenesis of invasion in cervical cancer[J]. *Int J Gynecol Cancer*, 2008, 18(5):1051-1059.
- [6] GUTFELD O, PRUS D, ACKERMAN Z, *et al.* Expression of serum amyloid A, in normal, dysplastic, and neoplastic human colonic mucosa: implication for a role in colonic tumorigenesis[J]. *J Histochem Cytochem*, 2006, 54(1):63-73.
- [7] HUANG D, GAO Q, GUO L, *et al.* Isolation and identification of cancer stem-like cells in esophageal carcinoma cell lines[J]. *Stem Cells Dev*, 2009, 18(3):465-473.
- [8] BENHAI K, AKCALI K C, OZTURK M. Redundant expression of canonical Wnt ligands in human breast cancer cell lines[J]. *Oncol Rep*, 2006, 15(3):701-707.
- [9] NAGAYAMA S, FUKUKAWA C, KATAGIRI T, *et al.* Therapeutic potential of antibodies against FZD 10, a cell-surface protein, for synovial sarcomas[J]. *Oncogene*, 2005, 24(41):6201-6212.
- [10] GUIJARRO M V, LEAL J F, FOMINAYA J, *et al.* MAP17 overexpression is a common characteristic of carcinomas[J]. *Carcinogenesis*, 2007, 28(8):1646-1652.
- [11] SUGARBAKER D J, RICHARDS W G, GORDON G J, *et al.* Transcriptome sequencing of malignant pleural mesothelioma tumors[J]. *Proc Natl Acad Sci USA*, 2008, 105(9):3521-3526.
- [12] VAN HOUTD I S, OUDEJANS J J, VAN DEN EERTWEGH A J, *et al.* Expression of the apoptosis inhibitor protease inhibitor 9 predicts clinical outcome in vaccinated patients with stage and melanoma[J]. *Clin Cancer Res*, 2005, 11(17):6400-6407.
- [13] STIREWALT D L, MESHINCHI S, KOPECKY K J, *et al.* Identification of genes with abnormal expression changes in acute myeloid leukemia[J]. *Genes Chromosomes Cancer*, 2008, 47(1):8-20.
- [14] JAIN M, KUMAR S, LAL P, *et al.* Role of GSTM3 polymorphism in the risk of developing esophageal cancer[J]. *Cancer Epidemiol Biomarkers Prev*, 2007, 16(1):178-181.
- [15] SINGH H, SACHAN R, DEVI S, *et al.* Association of GSTM1, GSTT1, and GSTM3 gene polymorphisms and susceptibility to cervical cancer in a North Indian population [J]. *Am J Obstet Gynecol*, 2008, 198(3):303.e1-303.e6.
- [16] JEON G A, LEE J S, PATEL V, *et al.* Global gene expression profiles of human head and neck squamous carcinoma cell lines[J]. *Int J Cancer*, 2004, 112(2):249-258.
- [17] SZATANEK R, DRABIK G, BARAN J, *et al.* Detection of isolated tumour cells in the blood and bone marrow of patients with gastric cancer by combined sorting, isolation and determination of MAGE-1, -2 mRNA expression[J]. *Oncol Rep*, 2008, 19(4):1055-1060.