

• 实验研究 •

乳腺癌患者与健康人血清蛋白质谱的差异分析

梁燕 姜军 柴凡 王秀丽 高春芳 钟玲

【摘要】 目的 探索乳腺癌患者与健康人群的血清蛋白质谱差异,寻找能够帮助鉴别诊断乳腺癌的候选血清蛋白标志物。**方法** 收集117例乳腺癌患者和56例健康人的血清标本,随机分为训练组(74例乳腺癌和36例健康人)与测试组(43例乳腺癌和20例健康对照)。采用表面增强激光解析离子化飞行时间质谱(SELDI-TOF MS)技术检测所有血清标本的蛋白质谱。用Biomarker Wizard统计软件比较训练组乳腺癌与健康对照间的蛋白质谱差异,再用Biomarker Pattern软件筛选出一组差异蛋白构建决策分类树模型以鉴别乳腺癌病例和健康人群,最后用测试组对分类模型进行验证。**结果** 乳腺癌组和健康对照组的血清蛋白质谱存在14个差异显著的蛋白峰,以质荷比分别为3958、4288、4974、5902、8518、8930、9282和11360的8个差异蛋白构建决策树分类模型,鉴别乳腺癌与健康对照组的敏感性为82.43%(61/74),特异性为83.33%(30/36),准确性为82.73%(91/110),用测试组进行验证的敏感性为86.05%(37/43),特异性为65.00%(13/20),准确性为79.37%(50/63)。**结论** 乳腺癌与健康人群的血清蛋白质谱存在差异,SELDI-TOF MS技术筛选出的血清差异蛋白有助于乳腺癌的鉴别诊断。

【关键词】 乳腺肿瘤;蛋白质组学;表面增强激光解析离子化飞行时间质谱;血清;诊断

【中图法分类号】 R737.9 **【文献标识码】** A

Analysis of the difference of serum proteome profiles between breast cancer patients and healthy women LIANG Yan, JIANG Jun, CHAI Fan, WANG Xiu-li, GAO Chun-fang, ZHONG Ling. Breast Disease Center, Southwest Hospital, Third Military Medical University, Chongqing 400038, China

【Abstract】 Objective To investigate the difference of serum proteome between breast cancer patients and healthy women, and seek for some potential biomarkers to distinguish breast cancer patients from healthy women using a serum protein fingerprinting technique. **Methods** Total serum samples collected from 117 breast cancer patients and 56 healthy women were divided into a learn-group (74 breast cancer patients and 36 healthy women) and a test-group (43 breast cancer patients and 30 healthy women). The samples were detected by surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS). Differences of mass spectrums between breast cancer patients and healthy women from the learn-group were analyzed by Biomarker Wizard software. Some different protein peaks were picked out and a classification tree was developed for identification of breast cancer patients using Biomarker Pattern software. The classification model was validated by an independent sample from the test-group. **Results** Fourteen protein peaks were

significantly different between breast cancer and healthy women groups. Eight proteins mass/charge of 3958, 4288, 4974, 5902, 8518, 8930, 9282 and 11 360 could identify breast cancer patients from healthy women with decision tree classification method. Sensitivity, specificity and correct ratio of the method were 82.43% (61/74), 83.33% (30/36) and 82.73% (91/110), respectively. In validation test, the sensitivity, specificity and correct ratio of the method were 86.05% (37/43), 65.00% (13/20) and 79.37% (50/63), respectively. **Conclusions** There are some different proteins in serum between breast cancer patients and healthy women. Candidate protein biomarkers in serum screened by SELDI-TOF MS could facilitate the diagnosis of breast cancer.

【Key words】 Breast neoplasm; Proteome; Surface enhanced laser desorption/ionization time-of-flight mass spectrometry; Serum; Diagnosis

乳腺癌的诊断、预后判断和疗效检测尚缺乏有效的血清学标志物。随着蛋白质组学的迅速发展,乳腺癌的血清蛋白质组学成为了研究乳腺癌相关标志物的一个重要组成部分。为了探索乳腺癌与健康人群的血清蛋白质谱差异,本研究采用表面增强激光解析离子化飞行时间质谱(SELDI-TOF MS)技术检测了乳腺癌与健康人群的血清蛋白质谱,比较两组的差异,并筛选差异蛋白构建分类模型,以帮助鉴别乳腺癌患者与健康人群。

1 材料与方法

1.1 标本来源

乳腺癌组 117 例血清标本均来自于 2006 年 1 月至 2007 年 7 月在第三军医大学西南医院乳腺疾病中心诊断为原发性乳腺癌并行手术治疗的病例,均得到患者的知情同意。乳腺癌组年龄 29~81 岁,平均年龄为 47.2 岁,中位年龄 47 岁;均得到病理证实,其中浸润性导管癌 95 例,浸润性小叶癌 6 例,导管内癌 7 例,其他类型 9 例;临床分期 I 期 25 例,II 期 58 例,III 期 34 例;所有病例均经临床检查排除重要脏器(心、肺、肝、肾等)疾病和明确的内分泌疾病。血清标本在患者入院后未进行任何治疗时采集。健康对照组 56 例血清标本来自女性健康志愿者(均签知情同意书),年龄 23~68 岁,平均年龄为 46.1 岁,中位年龄 45 岁;乳腺专科检查无异常,既往无乳腺疾病病史。晨起空腹采集静脉血液,4℃静置 3 h 后 3000 r/min 离心 20 min,取上清液,血清标本于-80℃储存。

1.2 主要试剂和仪器

尿素(urea)、乙腈(CAN)、3-环乙胺-1-丙磺酸(CHAPS)、三氟乙酸(TFA)、芥子酸(SPA)均购自 Sigma 公司;SELDI 质谱仪及金属亲和表面蛋白芯片(IMAC30-Cu)购自 CIPHERGEN 公司。

1.3 方法

1.3.1 血清样本准备:4 °C 融解冰冻血清,各取10 μl 分别置于离心管,每管加20 μl U9 缓冲液(含9 mol/L 尿素,2% CHAPS,50 mmol/L Tris-HCl,pH9.0),4 °C 200 r/min 振荡30 min,使蛋白变性。取10 μl 变性后的样品,各加入120 μl 结合/洗脱缓冲液(100 mmol/L 磷酸钠+500 mmol/L 氯化钠,pH7.0),4 °C 振荡30 min。

1.3.2 芯片处理:芯片每孔加50 μl 100 mmol/L 硫酸铜,4 °C 200 r/min 振荡5 min,使 Cu^{2+} 结合于芯片,倒除硫酸铜溶液,去离子水冲洗后甩干。每孔中加入100 mmol/L 醋酸钠(pH4.0)50 μl ,4 °C 下200 r/min 振荡5 min后倒除醋酸钠溶液,去离子水冲洗后甩干。每孔加入150 μl 结合/洗脱缓冲液,4 °C 200 r/min 振荡5 min后除去缓冲液,重复加入一次缓冲液并振荡,每孔加入50 μl 准备好的样品,4 °C 200 r/min 振荡60 min,倾去血清/缓冲液混合物。每孔加入150 μl 结合/洗脱缓冲液,4 °C 200 r/min 振荡5 min后除去缓冲液,重复1次,风干。取0.001 mg SPA,加入CAN 15 μl ,1% TFA 15 μl ,室温下振荡5 min使SPA全部溶解成饱和溶液,离心1 min,芯片各孔分两次加入SPA溶液,每次0.15 μl 。

1.3.3 芯片检测:将芯片放入SELDI质谱仪。用加有All-in-one标准蛋白质的NP20芯片校正质谱仪,设定主要参数,激光强度为175,检测灵敏度为8,优化分子质量(M_r)范围2000~20 000,最佳聚焦中心为6000,数据采集参数范围20~80,收集点数为100次。设定读片程序,根据蛋白质飞行时间和到达探测器的瞬时电流确定蛋白质的质荷比(m/z)和相对含量。计算机绘制出质谱图,其中纵坐标为波峰强度,即相对含量,横坐标为质荷比。

1.3.4 分析方法:将所有病例分为训练组(74例乳腺癌和36例健康人)与测试组(43例乳腺癌和20例健康人)。首先比较训练组乳腺癌与健康对照的血清蛋白质谱差异,然后筛选出一组差异蛋白构建分类模型鉴别乳腺癌与健康对照,最后用分类模型对测试组进行判别测试,验证模型鉴别乳腺癌的效果。

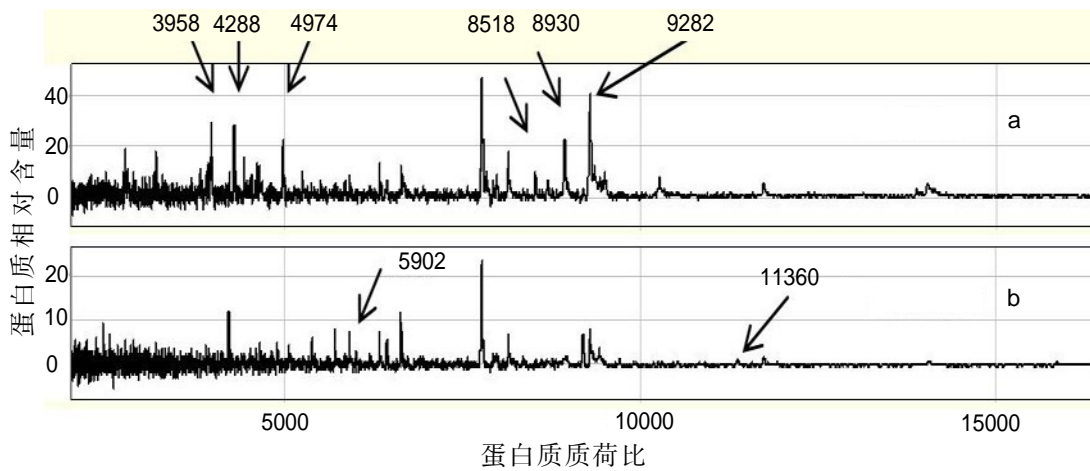
1.3.5 统计学方法:采用Biomarker Wizard统计软件对不同组相同质荷比的蛋白质相对含量进行秩和检验, $P < 0.050$ 为差异有统计学意义。用数据挖掘软件Biomarker Pattern(采用决策树方法)对数据分组及相关性进行分析。

2 结果

2.1 乳腺癌与健康对照的血清蛋白质谱差异

经SELDI-TOF-MS系统检测得到乳腺癌和健康人血清蛋白质谱图(图1)。比较训练组的乳腺癌与健康对照,两者之间共识别出33个蛋白峰,有14个蛋白峰含量的差异存在统计学意义($P < 0.050$,表1)。其中,相比健康对照,乳腺癌

组有 11 个蛋白质呈高表达, 3 个蛋白质呈低表达。



a: 乳腺癌; b: 健康对照

图 1 乳腺癌和健康对照组的血清蛋白质谱图

表 1 乳腺癌与健康对照的血清差异蛋白的相对含量

分组	质荷比(m/z)						
	11 360	5902	4212	3958	4288	8684	4974
乳腺癌	0.35±0.51	2.29±2.18	0.83±2.31	4.67±8.31	3.45±5.39	2.07±2.60	2.48±4.13
正常人	1.61±1.37	7.13±4.84	6.37±5.99	0.31±1.00	0.23±1.18	0.59±0.64	0.27±0.84
P 值	0.000	0.000	0.000	0.000	0.000	0.000	0.003

分组	质荷比(m/z)						
	13 864	9076	8518	8930	9344	9282	11 715
乳腺癌	1.48±1.76	0.74±1.57	2.42±4.34	4.94±4.24	2.68±2.81	6.21±5.63	2.49±1.91
正常人	0.59±0.55	0.17±0.36	0.36±0.58	2.95±2.43	1.33±0.84	3.63±2.64	1.76±1.23
P 值	0.003	0.010	0.025	0.025	0.030	0.043	0.043

2.2 构建决策分类模型鉴别乳腺癌与健康对照

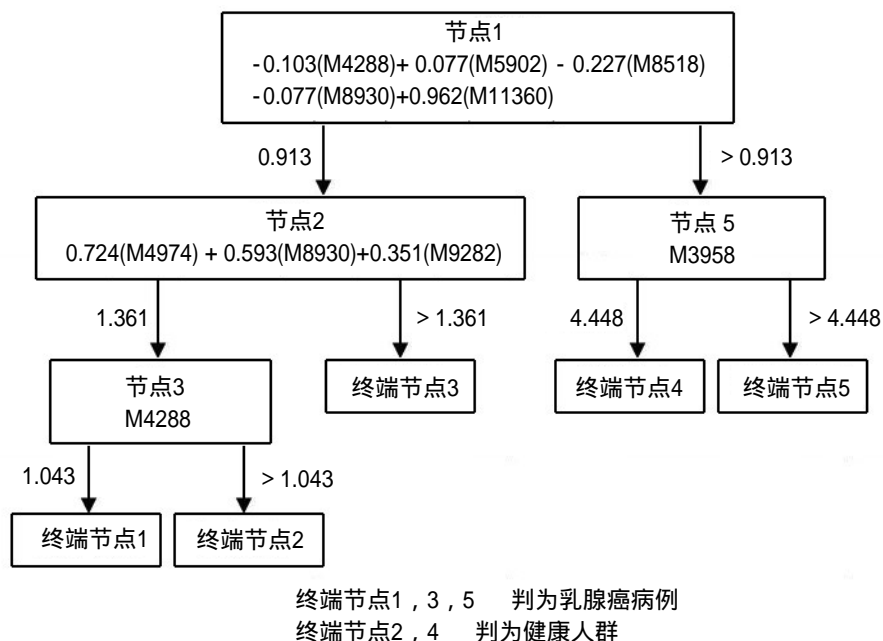
采用 Biomarker pattern 软件对差异蛋白进行分析, 通过软件计算, 筛选出以质荷比分别为 3958、4288、4974、5902、8518、8930、9282 和 11 360 的 8 个差异蛋白构建的决策树分类模型(图 2)分类效果最佳, 可以较好地将乳腺癌病例从乳腺良性疾病病例中鉴别出来, 训练模式下该鉴别模型的灵敏度为 82.43% (61/74), 特异度为 83.33% (30/36), 准确度为 82.73% (91/110)。

2.3 验证决策分类模型的鉴别效果

用建立的决策分类模型对测试组 43 例乳腺癌和 26 例健康对照的血清蛋白质谱进行盲法验证判别, 鉴别乳腺癌的敏感性为 86.05% (37/43), 特异性为 65.00% (13/20), 准确性为 79.37% (50/63)。

3 讨论

随着蛋白质组学技术的改进, 乳腺癌血清蛋白质组学研究也得到了扩展。本实验借助 SELIDI-TOF-MS 和蛋白质芯片技术, 对乳腺癌和健康对照组血



以质荷比分别为 3958、4288、4974、5902、8518、8930、9282 和 11 360 的 8 个差异蛋白构建决策树分类模型,用于鉴别乳腺癌病例与健康人群。

图 2 差异蛋白构建决策树分类模型

清标本中小分子质量蛋白质进行了高通量、高灵敏度的检测和分析。结果发现两组之间存在着 14 个差异显著的蛋白峰,其中质荷比为 3958、4288、4974、8518、8684、8930、9076、9282、9344、11 715 和 13 864 的 11 个蛋白质在乳腺癌中高表达,质荷比为 4212、5902、11 360 的蛋白质在乳腺癌中低表达。

笔者将差异蛋白的质荷比和飞行质谱参数与网络数据库的参数进行了对比,发现质荷比为 4288 和 8930 的蛋白可能是间- α 胰蛋白酶抑制物重链 H4 (ITIH4) 片段和 C3a 精氨酸(C3a_{desArg})。相关报道最先见于 Li 等^[1]的研究。他们发现了可以鉴别乳腺癌和健康对照的三个血清蛋白峰:在乳腺癌中一个 M_r 4300 的蛋白减少,两个蛋白(M_r 8100 和 M_r 8900)增加。这些蛋白质峰被鉴定为间- α 胰蛋白酶抑制物重链 H4 (ITIH4, M_r 4300) 片段、C3a 精氨酸(C3a_{desArg}, M_r 8900) 以及其羟基端缩短形式的变构体(C3a_{desArg} Δ_8 , M_r 8100)^[2]。后续的独立样本研究证实只有 M_r 8100 和 M_r 8900 的 C3a 片段有所增加^[2]。但是, M_r 8100 的 C3a_{desArg} Δ_8 在第 2 次^[3] 和第 3 次^[4] 验证性试验中缺乏显著性变化,而后者甚至发现 M_r 8900 的 C3a_{desArg} 在乳腺癌中表达不升反降^[4]。此外, C3a_{desArg} 被认为与肿瘤患者生存相关,在肿瘤转移和复发时其表达会下降^[5]。随后, Song 等^[6] 发现 M_r 4300 的 ITIH4 片段是众多已知的在乳腺癌中增高的 ITIH4 片段中的一员。Villanueva 等^[7] 和 Fung 等^[8] 研究发现类似 ITIH4 的片段,在癌症中有所增加或者尚缺乏鉴别能力。在本实验中,可能相匹配的 ITIH4 片段和 C3a_{desArg} 在乳腺癌中均为高表达, M_r 8100 C3a_{desArg} Δ_8

则未观察到有特异性表达。鉴于众多研究的矛盾发现, ITIH4 片段、C3a_{desArg} 以及 C3a_{desArg} Δ ₈ 在乳腺癌诊断中的作用还不能确定。

通过检索,本实验中在乳腺癌组高表达的质荷比为 9282 的蛋白可能是载脂蛋白 A-II (Apolipoprotein A-II), Heike 等^[9]发现多西紫杉醇冲击后血清中载脂蛋白 A-II 的表达显著降低。株和组织中的高表达与转移有关^[10-13]。

目前关于肿瘤的蛋白质组学研究已得到了大量差异蛋白,但被鉴定出结构的只有极小一部分。许多使用 MALDI-TOF MS 和 SELDI-TOF MS 技术(非结构性鉴别)得到的血浆或血清蛋白质谱峰,已被用于鉴别乳腺癌患者和乳腺良性疾病和(或)健康对照人群的研究^[14-16]。但是结果缺乏验证仍然是最突出的问题。Becker 等^[17]经 SELDI-TOF MS 分析,在是、否存在 BRCA-1 突变的患者之间得到多个显著差异表达的蛋白质峰,这些蛋白质峰均未鉴定结构,也无一经过独立样本集的进一步验证。而验证才是最重要的重复性试验,可以防止系统偏倚和数据拟合过度。与上述研究相反, Belluco 等^[18]报道了由 7 个蛋白峰组成的分类模型(未进行结构鉴定),并且在距最初研究发现后的 14 个月时进行了验证性研究,得到了极好的结果。

针对验证的重要性,本组实验特别设计了独立样本来对筛选出的 8 个蛋白构成的分类模型进行盲法验证。经过验证,乳腺癌与健康对照组分类模型的鉴别准确度为 79.37%,虽然不是很高,但避免了数据过度拟合,具有较好的重复性。鉴于病例数并不大,个体差异和肿瘤异质性的影响可能较大,下一步笔者将在此基础上扩大样本,筛选更具代表性的差异蛋白构建分类模型,有望提高诊断的准确度。同时,还需后续的验证性研究,以确定结果在时间和批次上的稳定性;对于筛选出的差异蛋白,将对其开展鉴定研究,以期为寻找乳腺癌的血清标志物提供进一步帮助。

参考文献

- [1] Li J, Zhang Z, Rosenzweig J, *et al.* Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem*, 2002, 48:1296-1304.
- [2] Li J, Orlandi R, White C N, *et al.* Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clin Chem*, 2005, 51:2229-2235.
- [3] Mathelin C, Cromer A, Wendling C, *et al.* Serum biomarkers for detection of breast cancers: a prospective study. *Breast Cancer Res Treat*, 2006, 96:83-90.
- [4] Van Winden A W J, Gast M C W, Beijnen J H, *et al.* Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: a case control study. *BMC Med Genomics*, 2009, 2:4.
- [5] Goncalves A, Esterni B, Bertucci F, *et al.* Postoperative serum proteomic profiles may predict metastatic relapse in highrisk primary breast cancer patients receiving adjuvant chemotherapy. *Oncogene*, 2006, 25:981-989.

- [6] Song J, Patel M, Rosenzweig C N, *et al.* Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. *Clin Chem*, 2006, 52:1045—1053.
- [7] Villanueva J, Shaffer DR, Philip J, *et al.* Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest*, 2006, 116:271—284.
- [8] Fung E T, Yip T T, Lomas L, *et al.* Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int J Cancer*, 2005, 115:783—789.
- [9] Heike Y, Hosokawa M, Osumi S, *et al.* Identification of serum proteins related to adverse effects induced by docetaxel infusion from protein expression profiles of serum using SELDI ProteinChip system. *Anticancer Res*, 2005, 25:1197—1203.
- [10] Leong S, Christopherson R I, Baxter R C. Profiling of apoptotic changes in human breast cancer cells using SELDI-TOF mass spectrometry. *Cell Physiol Biochem*, 2007, 20:579—590.
- [11] Sanders M E, Dias E C, Xu B J, *et al.* Differentiating proteomic biomarkers in breast cancer by laser capture microdissection and MALDI MS. *J Proteome Res*, 2008, 7:1500—1507.
- [12] Ricolleau G, Charbonnel C, Lode L, *et al.* Surface enhanced laser desorption/ionization time of flight mass spectrometry protein profiling identifies ubiquitin and ferritin light chain as prognostic biomarkers in node-negative breast cancer tumors. *Proteomics*, 2006, 6:1963—1975.
- [13] Goncalves A, CharafeJauffret E, Bertucci F, *et al.* Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. *Mol Cell Proteomics*, 2008, 7:1420—1433.
- [14] Hu Y, Zhang S, Yu J, *et al.* SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast*, 2005, 14:250—255.
- [15] Laronga C, Becker S, Watson P, *et al.* SELDI-TOF serum profiling for prognostic and diagnostic classification of breast cancers. *Dis Markers*, 2003, 19:229—238.
- [16] Shin S, Cazares L, Schneider H, *et al.* Serum biomarkers to differentiate benign and malignant mammographic lesions. *J Am Coll Surg*, 2007, 204:1065—1071.
- [17] Becker S, Cazares L H, Watson P, *et al.* Surfaced-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer. *Ann Surg Oncol*, 2004, 11:907—914.
- [18] Belluco C, Petricoin EF, Mammano E, *et al.* Serumproteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage 1 breast cancer. *Ann Surg Oncol*, 2007, 14:2470—2476.

(收稿日期:2009-09-29)

(本文编辑:罗承丽)

梁燕,姜军,柴凡,等.乳腺癌患者与健康人血清蛋白质谱的差异分析[J/CD].中华乳腺病杂志:电子版,2009,3(5):535—541.