

· 临床研究 ·

胃癌血清蛋白指纹人工神经网络诊断模型的研究*

吕连华¹, 勾宗蓉¹, 王开正^{2△}, 丁银环², 李 圃¹

(1. 四川省绵阳市人民医院检验科 621000; 2. 泸州医学院附属医院检验科, 四川泸州 646000)

摘要:目的 建立胃癌人工神经网络蛋白分子诊断模型, 寻找胃癌早期诊断的新方法。方法 采用表面增强激光解析离子化飞行时间质谱(SELDI-TOF-MS)技术及配套芯片 CM10 和 Biomarker Wizard 3.1 软件筛选胃癌差异表达蛋白, 通过人工神经网络(ANN)建立并验证胃癌的 SELDI 分子诊断模型。结果 共建立 3 个胃癌诊断模型(I、II、III), 分别为胃癌的诊断、筛查、鉴别诊断模型。选其中由 5 个差异表达蛋白(质荷比为 2502、2544、3085、8574、8740)组成的胃癌人工神经网络诊断模型 I 作为胃癌人工神经网络诊断模型, 对胃癌的诊断灵敏度为 95.0%, 特异度为 98.33%, 阳性预测值为 95.0%, 阴性预测值为 98.33%, 诊断准确度为 97.5%。结论 SELDI-TOF-MS 技术对胃癌的早期诊断具有一定的价值, 值得进一步的研究。

关键词: 胃肿瘤; 神经网络模型; 分子诊断模型

doi:10.3969/j.issn.1671-8348.2012.01.008

文献标识码: A

文章编号: 1671-8348(2012)01-0022-03

Study on serum protein finger printing coupled with ANN to establish diagnostic model of gastric cancer*

Lv Lianhua¹, Gou Zongrong¹, Wang Kaizheng^{2△}, Ding Yinhan², Li Pu¹

(1. Department of Clinical Laboratory, People's Hospital of Mianyang City, Mianyang, Sichuan 621000, China;

2. Department of Laboratory Medicine, Affiliated Hospital of Luzhou Medical College, Luzhou, Sichuan 646000, China)

Abstract: Objective To study serum protein of patients with gastric cancer and screen different expressed protein to find molecule diagnostic model and new method for early detection of gastric cancer. **Methods** The SELDI-TOF-MS, CM10 protein chip and Biomarker Wizard 3.1 software were performed to detect mass spectrogram of patients with gastric cancer and contrast persons for serum protein signature analysis to build and test ANN model. **Results** Three diagnostic models were built for diagnosis /differential diagnosis and screening of gastric cancer. Five specific protein peaks (M/Z as 2502, 2544, 3085, 8574, 8740) were chosen to develop the artificial neural network diagnostic model. The model was yielded a sensitivity of 95.0%, a specificity of 98.33%, a positive predictive value of 95.0%, a negative predictive value of 98.33%, a accuracy of 97.5%. **Conclusion** SELDI-TOF-MS is valuable to the early diagnosis of gastric cancer and need further study.

Key words: stomach neoplasms; neural networks; molecule diagnostic model

胃癌是源自胃黏膜上皮的消化道恶性肿瘤, 占消化道恶性肿瘤的首位, 死亡率居恶性肿瘤的第 2 位^[1]。胃癌患者在早期缺乏特异的临床症状及体征, 大部分胃癌患者在发现时已经是晚期, 手术切除率低, 疗效差, 术后死亡率仍保持在较高水平^[2]。据报道, 胃癌患者的 5 年生存率并不能让人满意, I a 期为 78%, I b 期为 58%, II 期胃癌为 34%, III a 期为 20%, III b 期只有 8%^[3]。目前, 还缺乏有效的筛选早期胃癌的方法, 寻找敏感性、特异性都很高的胃癌标志物仍然是胃癌早期发现的主要研究方向。本研究应用表面增强激光解析离子化飞行时间质谱(SELDI-TOF-MS)技术进行胃癌相关差异表达蛋白的筛选和人工神经网络(ANN)分子诊断模型研究。方法学中加入二维内标法, 校准和消除操作条件波动对分析结果的影响, 明显提高了实验的准确度和重复性^[4]。

1 资料与方法

1.1 一般资料 选择 2007 年 9 月至 2009 年 7 月泸州医学院附属医院消化内科患者及健康者共 160 例。胃癌组 40 例, 其中男 26 例, 女 14 例, 平均年龄 52 岁。病理分期: I 期 6 例, II 期 9 例, III 期 12 例, IV 期 13 例, III、IV 期胃癌患者仅 3 例; 未见淋巴结转移, 伴脏器转移 8 例。对照组 120 例, 胃炎患者和健康体检者各 60 例。其中男 71 例, 女 49 例, 平均年龄 49 岁。

经 *t* 检验 2 组患者年龄比较差异无统计学意义($P > 0.05$)。

1.2 仪器和芯片 PBS II/C 型 SELDI-TOF-MS 系统(美国 Ciphergen 生物公司), CM10 芯片及配套试剂。

1.3 方法 利用 PBS II/C 型蛋白质指纹图谱仪检测结合在 CM10 芯片表面的血清蛋白质, 用 Biomarker Wizard 3.1 软件分析胃癌组和对照组蛋白质指纹图谱差异, 利用 ANN 建立并验证模型。芯片处理和试剂配制等具体操作按说明书要求进行。

2 结果

2.1 蛋白质峰检测结果 通过 SELDI-TOF-MS 系统对结合在 CM10 芯片上的蛋白质进行检测, 获得胃癌组、对照组的血清蛋白质指纹图。经过 Biomarker Wizard 3.1 软件分析, 作者发现以下结果。

2.1.1 胃癌组与对照组比较 建立和验证胃癌诊断模型即胃癌人工神经网络诊断模型 I, 检测到两组有显著差异的蛋白质峰 37 个(图 1), 筛选质荷比(M/Z)为 2502、2940、3085、8574、8740 的 5 个蛋白质峰建立胃癌人工神经网络诊断模型 I(表 1)。胃癌人工神经网络模型 I 双盲法验证结果(表 2)显示该诊断模型的灵敏度为 95.0%(19/20), 特异度为 98.33%(59/60), 阳性预测值为 95.0%(19/20), 阴性预测值为

* 基金项目: 国家高技术研究发展计划(863 计划)重点基金资助项目(2006AA02090407)。△ 通讯作者, Tel: 13808281350; E-mail: kaizhengw@yahoo.com。

98.33%(59/60),诊断准确度为 97.5%(78/80)。可见模型 I 对胃癌的确诊方面具有较高的应用价值。

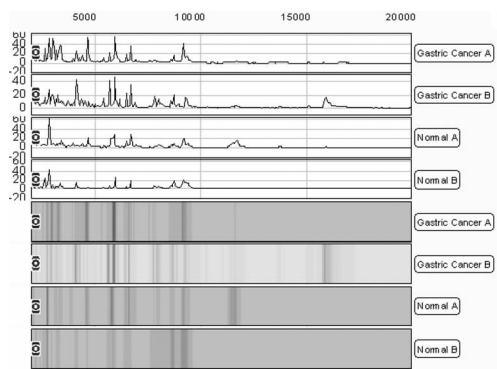


图 1 两组患者血清差异蛋白指纹图

表 1 胃癌人工神经网络诊断模型 I 的胃癌血清差异蛋白

| 质谱峰 M/Z | 胃癌组 | | | 对照组 | | | P* |
|------------|-----|-------|------|-----|------|------|----------|
| | n | Mean | SD | n | Mean | SD | |
| 2502.0 | 20 | 8.31 | 4.35 | 60 | 0.86 | 0.40 | 1.16E-07 |
| 2544.0 | 20 | 10.94 | 6.11 | 60 | 2.44 | 1.55 | 5.01E-07 |
| 3085.0 | 20 | 9.86 | 5.4 | 60 | 0.63 | 0.50 | 1.32E-08 |
| 8574.0 | 20 | 1.22 | 0.8 | 60 | 7.13 | 3.78 | 9.4E-09 |
| 8740.0 | 20 | 2.07 | 0.9 | 60 | 9.06 | 4.14 | 2.51E-08 |

表 2 胃癌人工神经网络诊断模型 I 双盲法验证结果

| 样本分组 | 胃癌组 | 对照组 | 合计 |
|----------------------|-----|-----|----|
| ANN 预测(≥ 0.5) | 19 | 1 | 20 |
| ANN 预测(< 0.5) | 1 | 59 | 60 |
| 合计 | 20 | 60 | 80 |

2.1.2 胃癌组与对照者比较 建立胃癌筛查模型即胃癌人工神经网络诊断模型 II,检测到两组有显著差异的蛋白质峰 32 个,筛选 M/Z 为 2544、2940、3085、4124、8691 的 5 个蛋白质($P < 5.0 E-07$)建立胃癌诊断模型 II。胃癌人工神经网络模型 II 双盲法验证结果显示灵敏度为 95.0%,特异度为 96.67%,阳性预测值为 95.0%,阴性预测值为 96.67%,诊断准确度为 96.0%,可见模型 II 对健康人群筛查方面具有一定价值。

2.1.3 胃癌组与胃炎组比较 建立胃癌鉴别诊断模型即胃癌人工神经网络诊断模型 III,检测到两组有显著差异的蛋白质峰 17 个,筛选 M/Z 为 2502、4124、3085、2544、8882 的 5 个蛋白质($P < 5.0 E-07$)建立胃癌诊断模型 III。该诊断模型灵敏度为 90.0%,特异度为 100%,阳性预测值为 100%,阴性预测值 93.75%,诊断准确度为 96.0%。可见模型 III 对胃癌鉴别诊断方面具有较高价值。

2.2 胃癌人工神经网络诊断模型效能评价 如结果中所述,3 个模型侧重点和意义各有不同。选胃癌诊断模型 I (M/Z 为 2502、2940、3085、8574、8740)作为胃癌诊断模型。再将 ANN 的预测输出值应用 SPSS13.0 软件绘制胃癌诊断模型的 ROC 曲线。ROC 曲线下面积 AZ 为 0.977,标准误为 0.019, $P < 0.01$,95%可信区间是(0.939,1.015),表明诊断模型 I 诊断准确性较高。

2.3 SELDI-TOF-MS 二维内标校准法结果 内标物对血清

蛋白指纹图的影响研究。本试验采用人工合成并纯化的 ACTH 多肽分子作为内标物质,加入待测样品中进行飞行时间质谱检测,通过内标物质在质谱图中横坐标(相对分子量)和纵坐标(丰度)响应值校准,消除出于操作条件的波动而对分析结果产生的影响。结果显示加入合成 ACTH 多肽(5 $\mu\text{g/mL}$)对血清蛋白质谱图除了多 1 条 M/Z 为 2465.8 的质谱峰以外,峰的多少和强度没有明显的变化。相对分子量经内标校准后精确度从 $\pm 14.4 \text{ Da}$ 提高到 0.05 Da ($P < 2.4 \times 10^{-8}$);相对分子量测定重复性提高 1.6 倍;质谱峰强度平均变异系数降低 1.57 倍。丰度经内标校准后,其变异系数平均降低 1.57 倍。

通过二维内标法明显提高了 SELDI-TOF-MS 测定蛋白相对分子质量和丰度的准确度和重复性。同时,为保证试验的重复性和芯片间不同批次实验数据的可比性,在每张芯片上都设一个内参,即将 10 个对照样本等量混合后分装(标记为 Nmix)并在每一个芯片的 8 个点中随机选出一一点以 Nmix 做平行试验,并以不同批次试验中的 Nmix 的明显蛋白质峰的平均值校正不同批次、不同芯片间的差异。

3 讨论

3.1 目前胃癌的早期检查方法 胃癌早期起病隐匿,常因无明显症状而漏诊。诊断手段正从症状、临床体征、胃镜检查等经典手段向基因诊断、分子生物学检测手段发展,但作为后者重要组成部分之一的胃癌血清肿瘤标志物,近年来发展较缓,已发现的胃癌相关标志物主要有癌胚抗原(CEA),糖链抗原(CA),包括 CA19-9、CA125、CA50、CA72-4 等[5]。研究发现目前用于胃癌诊断的肿瘤标志物阳性都不能令人满意,最高的 CA72-4 阳性率只有 47.7%,CEA 为 25%,CA19-9 为 25%,目前,还缺乏有效的筛选早期胃癌的方法[6]。

3.2 SELDI-TOF-MS 在临床蛋白质组学中的重要作用 近年来出现的联合蛋白质芯片和质谱方法的 SELDI-TOF-MS 技术是继基因芯片之后出现的新一代生物芯片技术,是 2002 年获得诺贝尔化学奖的“软激光解离技术”的应用与发展。它集分离、纯化、检测、分析为一体,具有灵敏、快速和高通量等特点。可定量检测蛋白质谱差异和研究蛋白质之间的相互作用[7-8]。目前,获得有意义的成果已日益凸现,尤其在肿瘤诊断上[9-10]。SELDI-TOF-MS 技术在疾病的相关蛋白筛选、肿瘤表达谱的比较、药物疗效评估和不良反应监测等方面有良好的应用前景,是近几年发展起来的一种全新的蛋白质组学研究手段[11-12]。但 SELDI-TOF-MS 技术在应用过程中也发现了一些问题或有待改进之处,主要有以下几个方面:(1)SELDI-TOF-MS 本身是一种半定量技术,顶峰高度和分子丰度之间的关系往往不呈线性关系,因此,差异蛋白的性质和血清浓度不明;(2)对不同样品设计的检测芯片,实际应用中并不能捕获所有性质相同的蛋白质;(3)不同研究者对同一疾病的最佳样品的准备也有差异,而样品的不同准备和处理常常影响结果的鉴定;(4)由于操作的高度精密性以及众多的影响因素,使不同实验室之间相同样品的检测结果常常不一致,从而使得进一步确认变得困难。为此,本实验率先建立了二维内标飞行时间质谱检测方法,明显提高了 SELDI-TOF-MS 的准确度,使得这一先进技术能够广泛应用于临床实验和蛋白质组学的研究,值得大力推广。

3.3 本项研究结果与同类文献报道的异同分析 查阅的同类文献,梁勇和刘池波[13]研究结果,其中质荷比为 8691 的结果与本研究完全相同。推测,质荷比为 8691 的蛋白质极有可能

为诊断胃癌的生物标志物。但对比其他同类研究发现结果差异很大^[14-15]。分析差异原因可能有以下几点:(1)选用的蛋白芯片类型不同,本实验中为弱阴离子芯片(WCX2),对比文献中为强阴离子芯片(SAX2,IMAC3-CU等)由于不同类型的化学表面芯片对蛋白质的选择不同,可能对捕获的蛋白质类型乃至结果产生影响;(2)数据分析软件不同,本研究为 ANN,对比文献中为决策树;(3)病例选择的区域性差异和样本类型差异;(4)方法学的不统一,本文中应用了二维内标法,消除了操作条件波动对结果产生的影响。

3.4 结论 (1)本文应用二维内标飞行时间质谱及人工神经网络共建立了 3 个胃癌诊断模型(I、II、III),交叉验证结果满意,分别对胃癌具有诊断、筛查、鉴别诊断意义。(2)M/Z 为 2544.9±4.4 的蛋白质在胃癌血清蛋白质谱中波峰强度明显升高并稳定表达,有可能是胃癌的生物标志物,需进一步验证;M/Z 为 8691 的蛋白质波峰强度明显降低且稳定表达,同类文献也见报道,说明该蛋白可能是被肿瘤细胞抑制的正常蛋白,有可能在胃癌的发病机制中起着重要作用^[8]。(3) SELDI-TOF-MS 技术在胃癌的诊断尤其是早期诊断、候选肿瘤标志物的筛选方面具有一定的价值,值得进一步的研究。(4)二维内标法,可以明显提高 SELDI-TOF-MS 技术检测的准确度,值得推广。

参考文献:

- [1] Ohgaki H, Matsukura N, Stewart BW, et al. Stomach cancer. World cancer report [M]. Lyon: IARC Press, 2003:197-208.
- [2] Doglietto GB, Pacelli F, Caprino P, et al. Independent prognostic factor in curable and far advanced gastric cancer[J]. World Surg, 2000, 24(4):459-463.
- [3] Katie PM. Gastric cancer studies focus on prolonging survival and finding molecular markers for targeted therapies [J]. Onco Log, 2004, 49(4):2-3.
- [4] 丁银环, 胡琼英, 姜伟, 等. SELDI-TOF-MS 二维内标法的建立和应用[J]. 现代检验医学杂志, 2009, 24(1):17-19.
- [5] Koike H, Ichikawa D, Ikoma H, et al. Comparison of serum aberrant methylation and conventional tumor markers in gastric cancer patients [J]. Hepatogastroenterology, 2005, 52(64):1293-1296.

- [6] Rejane M, Claudio RA, Giovanni MD, et al. Preoperative serum levels of CA72-4, CEA, CA19-9, and alpha-fetoprotein in patients with gastric cancer[J]. Rev Hosp Clin Fac Med S Paulo, 2002, 57(2):89-92.
- [7] Weinberger SR, Boschetti E, Santambien P, et al. Surface enhanced laser desorption/ionization retentate chromatography mass spectrometry (SELDI-RC-MS); a new method for rapid development of process chromatography conditions[J]. J Chromatogr B, 2002, 782(1):307-316.
- [8] Bane TK, Leblanc JF, Lee TD, et al. DNA affinity capture and protein profiling by SELDI-TOF mass spectrometry; effect of DNA methylation[J]. Nucleic Acids Res, 2002, 30(1):58-69.
- [9] Adam B, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men[J]. Cancer Res, 2002, 62(13):3609-3614.
- [10] Bandera CA, Ye B, Mok SC. New technologies for the identification of markers for early detection of ovarian cancer[J]. Curr Opin Obstet Gynecol, 2003, 15(1):51-55.
- [11] Service RF. Genetics and medicine. Recruiting genes, proteins for a revolution in diagnostics[J]. Science, 2003, 30(2):236-239.
- [12] Seibert V, Ebert MP, Buschmann T. Advances in clinical cancer proteomics; SELDI-TOF mass spectrometry and biomarker discovery[J]. Brief Funct Genomic Proteomic, 2005, 4(1):16-26.
- [13] 梁勇, 刘池波. 血清蛋白质谱模型对胃腺癌诊断的应用性研究[J]. 中国肿瘤, 15(2):127-130.
- [14] 潘春琴, 刘池波, 梁勇, 等. SELDI-TOF-MS 技术对胃炎诊断的应用[J]. 浙江医学, 2006, 28(12):970-974.
- [15] 高春芳, 李冬晖, 赵光, 等. 胃癌患者血清比较蛋白质组学研究[J]. 解放军医学杂志, 2005, 30(4):457-459.

(收稿日期:2011-08-11 修回日期:2011-09-20)

(上接第 21 页)

- [13] Kimura Y, Kimura H. Hydrogen sulfide protects neurons from oxidative stress [J]. Faseb, 2004, 18(10):1165-1170.
- [14] 王燕飞, 金红芳, 唐朝枢, 等. 气体信号分子在肺动脉高压发病中的作用[J]. 北京大学学报:医学版, 2006, 38(3):326-331.
- [15] Zhao WM, Zhang J, Lu YJ, et al. The vasorelaxant effect of H₂S as a novel endogenous gaseous KATP channel opener[J]. EMBO, 2001, 20(21):6008-6016.
- [16] Roberts WL, Moulton L, Law TC, et al. Evaluation of nine automated high-sensitivity C-reactive protein methods: implications for clinical and epidemiological applica-

tions[J]. Clin Chem, 2001, 47(3):418-425.

- [17] Dik MG, Jonker C, Hack CE, et al. Serum inflammatory proteins and cognitive decline in older persons[J]. Neurology, 2005, 64(8):1371-1374.
- [18] Oijen M, Witteman JC, Hofman A, et al. Fibrinogen is associated with an increased risk of Alzheimer disease and vascular dementia[J]. Stroke, 2005, 36(12):2637-2641.
- [19] Schmidt R, Schmidt H, Curb JD, et al. Early inflammation and dementia; a 25-year follow-up of the Honolulu-Asia Aging Study[J]. Ann Neurol, 2002, 52(2):168-174.

(收稿日期:2011-06-09 修回日期:2011-08-22)