·临床研究·

血清蛋白质谱结合人工神经网络在头颈部鳞癌诊断中的研究

曹素梅[1,2], 郭 翔[1,2], 陈福进[1,3], 杨安奎[1,3], 陈文宽[1,3], 李宁炜[1,2]

Serum Diagnosis of Head and Neck Squamous Cell Carcinoma Using Surface-enhanced Desorption Ionization Mass Spectrometry and Artificial Neural Network Analyses

CAO Su-Mei[1,2], GUO Xiang[1,2], CHEN Fu-Jin[1,3], YANG An-Kui[1,3], CHEN Wen-Kuan[1,3], LI Ning-Wei[1,2]

1. 华南肿瘤学国家重点实验室, 广东 广州 510060
2.中山大学肿瘤防治中心 鼻咽癌科, 广东 广州 510060
3. 中山大学肿瘤防治中心 头颈科, 广东 广州 510060

1 State Key Laboratory of Oncology in South China, Guangzhou, Guangdong, 510060, P. R. China
2. Department of Nasopharyngeal Carcinoma, Cancer Center, Sun Yat-sen University, Guangzhou, Guangdong, 510060, P. R. China
3. Department of Head and Neck Surgery, Cancer Center, Sun Yat-sen University, Guangzhou, Guangdong, 510060, P. R. China

通讯作者:郭 翔
Correspondence to: GUO Xiang
Tel: 86-20-87343359
Fax: 86-20-87343392
E-mail: xiangg1010@yahoo.com

基金项目:广东省社会发展领域科技项目(No. 2003-245);广东省卫生厅基金(No. 2004)

Grants: Social Development Project of Sci-Tech Research of Guangdong Province (No. 2003- 245); Research Foundation from Department of Public Health of Guangdong Province (No. 2004)

收稿日期:2006-08-10
修回日期:2007-01-15

[ABSTRACT] **BACKGROUND & OBJECTIVE**: The early diagnosis of head and neck squamous cell carcinoma (HNSCC) is the key factor that affecting the treatment result. We performed surface-enhanced desorption ionization time-of-flight mass spectrometry (SELDI-TOF-MS) using a multi-layer artificial neural network (ANN) to develop and evaluate a proteomic diagnosis approach for HNSCC. **METHODS**: Serum samples from 74 HNSCC patients and 146 healthy individuals were randomized into training set (148 samples) and test set (72 samples). At first, we detected the training set of samples using SELDI mass spectrometry and WCX2 (weak cation-exchange) chips. Using a multi-layer ANN with a back propagation algorithm, we identified a proteomic pattern that could discriminate cancer from control samples in the training set. The discovered pattern was then used to determine the accuracy of the classification system in the test set. **RESULTS**: Four top-scored peaks, at *m/z* (mass/charge) ratio of 4 469 u, 5 924 u, 8 926 u, and 16 697 u, were finally selected as the potential biomarkers for detection of HNSCC with both sensitivity and specificity of 100.0% in the training set. The classifier predicted the HNSCC with sensitivity of 85.7% (18/21) and specificity of 96.1% (49/51) in the test set. **CONCLUSION**: SELDI profiling is a useful tool to accurately identify patients with HNSCC.

KEYWORDS: Head and neck neoplasm; Squamous cell cancer; SELDI; Proteomic; Diagnosis; Artificial neural network

[摘 要] 背景与目的:头颈部鳞癌早诊率的高低是影响治疗效果的关键因素之一。本研究应用血清蛋白质谱技术结合人工神经网络建立头颈部鳞癌(head and neck squamous cell carcinoma, HNSCC)**患者的诊断模型,并评价其诊断价值。方法:对** 74 **例头颈部鳞癌患者和** 146 **例健康人的血清样本随机分为训练集**(148 **例**)**和测试集** (72 **例**)**。首先应用表面加强激光解吸电离-飞行时间质谱**(surface-enhanced desorption ionization time-of-flight mass spectrometry, SELDI-TOF-MS) **技术及** WCX2 (weak cation-exchange) **芯片检测训练集样本, 结合反向传播人工神经网络**(artificial neural network, ANN)**的方法建立诊断模型,进一步检测测试集样本并评价该模型的诊断价值。结果:在训练集样本中筛选出** 4 **个** *m/z*(**质荷比**)**峰**(4 469 u, 5 924 u, 8 926 u, 16 697 u)**区分头颈部鳞癌和对照人群的敏感度和特异度均达** 100.0%(148/148); **用该模型对测试集样本进行双盲法预测的敏感度为** 85.7% (18/21), **特异度为** 96.1%(49/51)**。结论:该模型诊断头颈部鳞癌具有较高的灵敏度和特异度,值得进一步研究。**

关键词:头颈部肿瘤; 鳞状上皮癌; SELDI; 蛋白质质谱; 人工神经网络; 诊断
中图分类号:R739.63 **文献标识码:**A
文章编号:1000-467X(2007)07-0767-04

头颈部鳞癌(head and neck squamous cell carcinoma,HNSCC)的主要治疗方式是手术和放射治疗,但生存率在过去20年内并无明显的改善,重要的原因之一是早诊率没有明显的提高。所以,寻找方法简便而灵敏度、特异度高的早期诊断方法尤其重要。

近年来,随着蛋白质组学和芯片技术、生物信息学的飞速发展,为高通量、特异地筛选肿瘤标志物提供了一个平台[1,2]。本研究应用表面加强激光解吸电离-飞行时间质谱(surface-enhanced desorption ionization time-of-flight mass spectrometry,SELDI-TOF-MS)技术和WCX2(weak cation-exchange)芯片,对头颈部鳞癌患者和健康对照人群进行检测,并结合人工神经网络(artificial neural network,ANN)建立诊断模型,并评价该模型的诊断价值。

1 材料与方法

1.1 标本及来源

74例头颈部肿瘤患者均经病理确诊为鳞癌,治疗前的血清标本取自中山大学肿瘤防治中心血清库,其中喉癌10例,鼻咽癌8例,扁桃体癌10例,口底癌6例,口咽癌8例,鼻腔、上颌窦癌8例,舌癌8例,下咽癌12例,牙龈癌4例。所有患者均采用UICC分期,其中Ⅰ期8例,Ⅱ期14例,Ⅲ期36例,Ⅳ期16例。男性60例,女性14例,中位年龄52岁(44~63岁)。146例对照人群的血清标本取自健康体检者,其中男性106例,女性40例,中位年龄42岁(23~55岁)。所有血样均于清晨空腹采集,分离血清并保存于-80℃低温冰箱备用。上述所有样本按随机数字表法随机分为两组,2/3为训练集(共148例,其中HNSCC 53例,Ⅰ、Ⅱ期鳞癌有10例;对照人群95例),1/3为测试集(共72例,其中HNSCC 21例,Ⅰ、Ⅱ期癌有4例;对照人群51例)。

1.2 主要仪器和试剂

PBS-Ⅱ型表面加强激光解吸电离-飞行时间质谱仪系统(SELDI-TOF-MS)、WCX2(weak cation-exchange)蛋白芯片和能量吸收分子CHCA购于美国Ciphergen公司,Cibacron Blue 3GA(白蛋白亲和树脂)和CHAPS缓冲液购于Sigma公司。

1.3 实验方法

血清标本在冰浴中解冻后594×*g*离心5 min,取10 μl用90 μl 0.5%CHAPS(pH7.4)溶液稀释后加入到Cibacron Blue 3GA中,在4℃摇床上反应60 min,594×*g*离心5 min,取上清溶液50 μl,进一步用100 mmol/L乙酸钠(pH4.0)稀释至200 μl,加至预先用150 μl 100 mmol/L乙酸钠平衡过的WCX2芯片中,振荡器上振荡60 min。甩去血清标本,100 mmol/L乙酸钠150 μl清洗芯片3次,每次5 min,再用去离子水清洗芯片2次,每次1 min,干燥后加0.5 μl能量吸收分子CHCA到已经结合上蛋白质的WCX2芯片上,干燥后行质谱仪检测。

SELDI质谱系统用all-in-one(Ciphergen公司)校正到分子量误差<0.1%。将结合好蛋白质的WCX2蛋白芯片用质谱阅读仪分析。使用的分析参数为:最高分子量50 000 u,优化范围1 000到20 000 u,激光强度160,检测敏感度7,每个样本取91个点的平均值。所有的样本都使用相同的参数,用Ciphergen ProteinChip Software 3.1软件收集质谱信号。

1.4 生物信息学方法

应用Ciphergen ProteinChip Software 3.1软件将所有样本的质谱数据在1 000到50 000的范围内进行两次信噪比过滤,要求过滤后的质荷比(m/z)峰在10%以上的样本中同时存在,并且同一个m/z峰在不同的样本中的偏差<0.3%,并将它们的强度初始化到-1~1的范围内。

经上述处理后,将初步筛选出来的m/z峰在头颈部鳞癌患者与对照人群做t检验,按P值大小比较每一个m/z峰在二者间表达差异的显著性程度。然后根据P值大小将两组间有差异的m/z峰从小到大排列,从P值最小的单个m/z峰开始逐个增加m/z峰的数量作为输入神经元来训练ANN。将所有样本随机分组,2/3为训练集,1/3为测试集。用训练集建立神经网络,分别计算各个模型的准确性,选择输入神经元数量最少而准确性最高的m/z峰组合作为优化模型。进一步将该模型对测试集进行双盲法预测,即实验者不知所检测样本的类别,统计分析者也不知所预测样本的实际类别,并评价该模型的准确率。

2 结 果

2.1 血清蛋白质谱检测的重复性

为检验质谱检测结果的重复性,我们随机选择1名健康对照者的血清标本在同一天用3个不同批次的同类型芯片进行相同方法的检测,并在3个月后用相同方法重复检测该样本,在该样本质谱图中随机选择10条m/z峰进行比较,结果显示不同时间及不同批次芯片检测的同一样本的m/z峰最大振幅的变异系数(CV)为27.5%,如图1所示。

2.2 HNSCC诊断模型的建立和评价

通过对148例训练集标本的检测和数据分析,

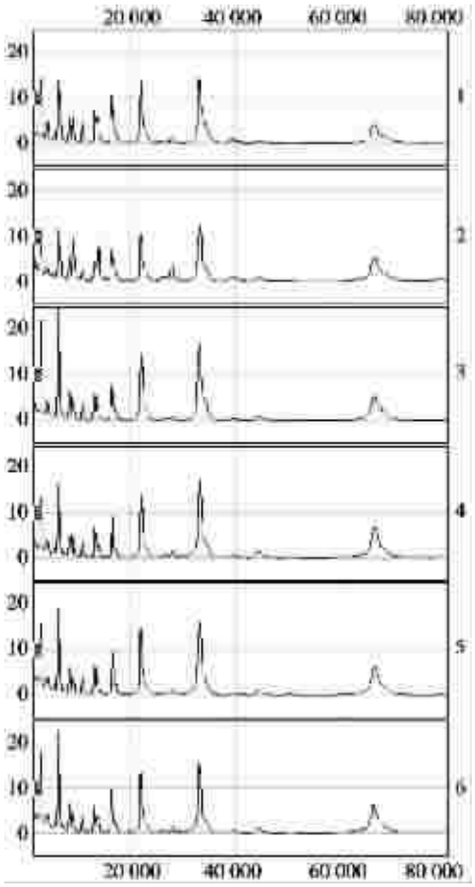

图 1 应用 SELDI-TOF 方法在不同批次的同类型芯片及不同时间检测同一血清样本所得的蛋白质谱图的比较

Figure 1 Reproducibility of SELDI-TOF mass spectra from a single serum sample processed on different batches of the same type of chip array on the same day (lanes 1-3) and on different days (lanes 4-6)

共检测到 HNSCC 患者与健康对照者间存在有差异表达的 *m/z* 峰 114 个。按 *t* 检验 *P* 值大小从小到大逐个增加 *m/z* 峰作为输入神经元训练神经网络模型,当 4 条分别为 4 469 u、5 924 u、8 926 u、16 697 u的 *m/z* 峰(表 1)作为输入神经元时,建立的诊断模型所使用的 *m/z* 峰最少而区分两组的准确性最高,灵敏度和特异度均达到 100%(148/148),见图 2。*m/z* 在 4 469 u、5924 u、8 926 u 的 3 条 *m/z* 峰在 HNSCC 患者中含量高,而 *m/z* 在 16 697 u 的质谱在对照者中含量高,其中 4 469 u 和 5 924 u 两条 *m/z* 峰如图 3 所示。进而应用这一模型对测试集的 72 例标本进行双盲法分析和预测,结果 21 例 HNSCC 患者中 18 例被正确预测;51 例健康对照者中 49 例被正确预测,灵敏度为 85.7%,特异度为 96.1%。

表 1 4 条用于构建 HNSCC 诊断模型的 *m/z* 峰在两组间的比较

Table 1 Four top-scored peaks as the potential biomarkers for detecting head and neck squamous cell carcinoma (HNSCC)

M/Z(u)	Control	HNSCC
4 469	1.95±1.00	6.70±2.94
5 924	5.53±2.60	11.18±3.03
8 926	2.82±1.55	10.05±5.08
16 697	4.56±1.49	2.54±0.81

All values are presented as mean±SD of relevant groups

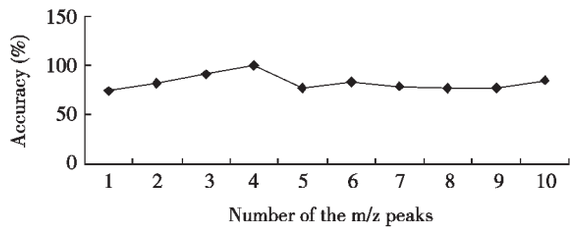

图 2 用不同数量的 *m/z* 峰组合训练 ANN 建立诊断模型区分 HNSCC 和健康对照的准确率

Figure 2 The accuracy of discriminating head and neck squamous cell carcinoma patients from control individuals using different *m/z* peak subsets to train the artificial neural network

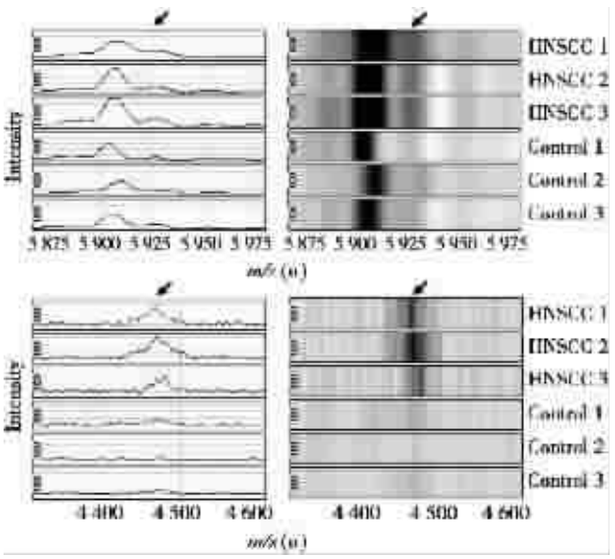

图 3 *m/z* 分别为 4 469 u 和 5 925 u 的蛋白质谱图和模拟电泳图在 3 个 HNSCC 患者和 3 个健康对照者中的比较

Figure 3 Representative surface-enhanced laser desorption/ionization mass spectra and gel view from serum of 3 HNSCC patients compared with serum from 3 controls

The arrows identify peaks with average *m/z* of 4 469 u and 5 926 u that are up-regulated in HNSCC compared with normal serum.

3 讨 论

临床研究显示,早期诊断、早期治疗头颈部鳞癌不仅能大大提高患者的生存率,而且,还决定了

能否最大限度地保存患者的发音、咀嚼及吞咽等重要的生理功能,提高生存质量。由于肿瘤的早期症状无特异性及缺乏理想的肿瘤标志物,头颈部鳞癌的早期诊断率目前仍不理想。目前,已有报道一些分子标志物在头颈部肿瘤诊断上的研究,如微卫星标记物[3]、启动子过甲基化[4,5]以及线粒体 DNA 改变[6]等,但检测这些指标存在着操作复杂、技术要求高、灵敏度及特异度不够理想等问题,因此,均未应用于临床诊断。

SELDI 系统可分析患者血清蛋白质谱变化,并且可筛选肿瘤特异性的蛋白质标记物。它是利用激光脉冲辐射使芯池中的分析物解吸形成荷电离子,根据不同 m/z 的离子在仪器场中飞行的时间长短不一,由此绘制出一张质谱图来。将它与常人或某种疾病患者的质谱图进行对照,结合先进的生物信息学方法进行分析,就能够最终发现和捕获新的疾病特异性相关蛋白及其特征。此方法具有高通量筛选、高灵敏度检测、全自动化操作、小分子量、低丰度粗样品的特点。目前,被认为是一种新的、有效的早期诊断试验方法。SELDI 质谱系统已经在肿瘤相关蛋白的识别上取得了一系列突破性进展,例如,乳腺癌[7]、卵巢癌[8]、胰腺癌[9]等肿瘤,应用该方法发现了特异性的蛋白标记物。由此可见,蛋白质芯片质谱系统虽然建立不久,但在筛选肿瘤特异性的蛋白,提高肿瘤早诊率方面显示了巨大的应用前景。

本研究应用 SELDI 质谱技术和 WCX2 蛋白芯片结合 ANN 技术筛选出 4 条 m/z 峰(4 469 u、5 924 u、8 926 u、16 697 u),建立了 HNSCC 的血清蛋白质谱诊断模型,并且获得了较高的灵敏度(85.7%)和特异度(96.1%)。目前,国外 Wadsworth 等[10]也进行了类似的研究,并建立了包括 33 条 m/z 峰组成的 HNSCC 患者血清蛋白质谱诊断模型,诊断的灵敏度为 83%,特异度为 90%。通过对比发现,虽然 Wadsworth 等的研究中应用的是 H4 蛋白芯片与本研究中应用 WCX2 芯片不同,但本研究诊断模型中发现的 4 469 u、5 924 u、8 926 u 的 3 条 m/z 峰在 Wadsworth 等建立的诊断模型中同样被筛选出来,说明这些蛋白质分子用于区分 HNSCC 具有较高的重复性和特异性。

SELDI 质谱技术是高灵敏度检测方法,可检测到 10^{-15} mol 的蛋白,非常微小的实验条件的改变就可能影响检测结果,因此,提高准确性和重复性是保证检测质量的重要措施。为保证研究的准确性,在本研究中分子量应用 all-in-one 校正到误差 < 0.1%。为检验本研究质谱数据的重复性,我们用不同批次的同类型芯片在不同时间应用相同方法和实验条件重复检测同一样本,结果显示随机选择的 10 条 m/z 峰最大振幅的 CV < 30%,具有较好的重复性,说明不同批次的同类型芯片及血清在不同时间的反复冻融对研究结果影响甚微。

通过这一方法获得的蛋白质谱信息量是巨大的,传统的统计方法难以完成如此大量数据的分析,因此,本研究中采用 ANN 这一生物信息学技术来参与完成。ANN 是在对人脑组织结构和运行机智的认识理解基础上模拟其机构和智能行为的一种工程系统。目前,已发展了几十种神经网络的算法。本研究中采用反向传播算法,该网络包括 4 层结构,除输入层和输出层外还包括两个分别含 100 个神经元的隐含层。在样本划分时,我们将总样本随机的划分为训练集和测试集,这样将很好的使训练和测试样本相互独立,更好的估计建立模型的准确性。

今后,我们将进一步扩大样本含量,验证本研究结果,并争取将这些特异性的蛋白质分子分离和提纯。

[参 考 文 献]

[1] Merchant M, Weinberger S R. Recent advancements in surface enhanced laser desorption/ionization-time of flight-mass spectrometry [J]. Electrophoresis, 2000,21(6):1164 -1177.

[2] Kuwata H, Yip T T, Yip C L, et al. Bactericidal domain of lactoferrin: detection, quantitation and characterization of lactoferricin in serum by SELDI affinity mass spectrometry [J]. Biochem Biophys Res Comm, 1998,245(3):764-773.

[3] Spafford M F, Koch W M, Reed A L, et al. Detection of head and neck squamous cell carcinoma among exfoliated oral mucosal cells by microsatellite analysis [J]. Clin Cancer Res, 2001,7(3):607-612.

[4] Sanchez-Cespedes M, Esteller M, Wu L, et al. Gene prometer hypermethylation in tumors and serum of head and neck cancer patients [J]. Cancer Res, 2000,60(4):892-895.

[5] Rosas SL, Koch W, da Costa Carvalho M G, et al. Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients [J]. Cancer Res, 2001,61(3):939-942.

[6] Fliss M S, Usadel H, Caballero O L, et al. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids [J]. Science, 2000,287(5460):2017-2019.

[7] Li J, Zhang Z, Rosenzweig J, et al. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer [J]. Clin Chem, 2002,48(8):1296-1304.

[8] Petricoin E F, Ardekani A M, Hitt B A, et al. Use of proteomic patterns in serum to identify ovarian cancer [J]. Lancet, 2002,359(9306):572-577.

[9] Koopmann J, Zhang Z, White N, et al. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry [J]. Clin Cancer Res, 2004,10(3):860-868.

[10] Wadsworth J T, Somers K D, Stack B C Jr, et al. Identification of patients with head and neck cancer using serum protein profiles [J]. Arch Otolaryngol Head Neck Surg, 2004,130(1):98-104.

[编辑及校对:谢汝华,杨允贵]