

3 种人工神经网络模型预测大肠癌的初步研究

谭顺厚, 胡丽敏[△] (四川省资阳市乐至县中医医院检验科 610072)

【摘要】 目的 探讨大肠癌血清蛋白标志物、肿瘤标志物及联合多标志物人工神经网络(ANN)模型在预测大肠癌中的价值。**方法** 大肠癌与健康对照血清样本 106 例, 利用表面增强激光解吸电离飞行时间质谱(SELDI-TOF-MS)检测血清蛋白质谱并筛选大肠癌蛋白标志物, 电化学发光法检测癌胚抗原(CEA)、甲胎蛋白(AFP)、糖类抗原 72-4(CA72-4)和 CA19-9 分别建立蛋白标志物、肿瘤标志物及蛋白标志物与肿瘤标志物结合的多标志物 ANN 模型。**结果** 大肠癌患者和对照组间比较差异有统计学意义($P < 0.001$), 筛选 4 个质荷比(m/z)分别为 4 095、5 640、4 480、7 620 m/z 蛋白建立 ANN 模型, 预测大肠癌敏感度和特异度为 92.3% 和 83.3%; 肿瘤标志物模型预测大肠癌的敏感度为 73.1%, 特异度 86.7%; 联合筛选的 4 个标志蛋白和 CEA、CA72-4 建立的模型诊断大肠癌敏感度和特异度分别为 92.3% 和 96.7%。**结论** 联合蛋白标志物和肿瘤标志物建立 ANN 模型, 在预测大肠癌中显示高通量、高敏感性和高特异性的特点, 具有潜在应用价值。

【关键词】 大肠癌; 表面增强激光解吸电离飞行时间质谱; 人工神经网络; 肿瘤标志物

DOI:10.3969/j.issn.1672-9455.2011.08.022 文献标志码:A 文章编号:1672-9455(2011)08-0941-02

Application of three artificial neural network models in diagnosing colorectal carcinoma TAN Shun-hou, HU Limin[△] (Department of Clinical Laboratory, Lezhi county Hospital, Ziyang, Sichuan 646000, China)

【Abstract】 Objective To establish three artificial neural network (ANN) models based on protein markers, tumor markers, multi-markers of colorectal carcinoma and discuss the application of these in diagnosing colorectal carcinoma. **Methods** This study included 106 serum from patients with colorectal carcinoma and healthy controls. The SELDI-TOF-MS were performed to screen serum protein markers. The ANN model based on the screened protein markers was established. The levels of carcinoembryonic antigen(CEA), AFP(α -fetoprotein), CA19-9(cancer antigen) and CA72-4 in serum from patients with colorectal carcinoma were detected by electrochemiluminescence determination. The model based on the tumor markers was established. The multi-markers model was constructed by the screened protein markers and tumor markers. **Results** The result showed a significant difference between patients with colorectal carcinoma and healthy controls($P < 0.01$). Four specific protein peaks(m/z at 4 095, 5 640, 4 480, 7 620) were chosen to build an ANN model, the model showed a sensitivity of 92.3%, a specificity of 83.3% to diagnose the colorectal carcinoma. The ANN model of tumor markers had a sensitivity of 73.1% and a specificity of 86.7% to diagnose the colorectal carcinoma. The multi-markers model showed a sensitivity of 92.3%, a specificity of 96.7% to diagnose colorectal carcinoma. **Conclusion** The preliminary results suggest that the multi-markers model based on protein markers and tumor markers have the high sensibility and specificity in diagnosing colorectal carcinoma.

【Key words】 colorectal carcinoma; SELDI-TOF-MS; artificial neural network; tumor marker

寻找肿瘤发生及转移相关的蛋白标志物及药物治疗靶标, 对大肠癌早期诊断及治疗具有重要意义^[1]。本文利用表面增强激光解吸电离飞行时间质谱(SELDI-TOF-MS)和电化学发光技术, 分别建立大肠癌血清蛋白标志物、肿瘤标志物、蛋白标志物和肿瘤标志物联合的人工神经网络(ANN)预测模型, 探讨 3 者在大肠癌实验诊断中的价值。

1 资料与方法

1.1 一般资料 来源于本院 2007 年 1 月至 2009 年 12 月住院患者及健康体检者血清样本 106 例。其中, 大肠癌患者 51 例(病例组), Dukes 分期: A 期 6 例, B 期 22 例, C 期 15 例, D 期 8 例。男 29 例, 女 22 例, 年龄 39~75 岁, 平均(52.1 $\pm$ 8.6)岁。健康对照组 55 例为该院同期健康体检者, 男 31 例, 女 24 例, 年龄 21~61 岁, 平均(46.4 $\pm$ 9.2)岁。所有样本均采集于清晨空腹静脉血。

1.2 主要试剂与仪器 PBS II/C 型蛋白指纹图谱仪、CM10

(弱阳离子交换表面)芯片及配套试剂(美国赛弗吉公司); 电化学发光仪(ELECSYS 2010)及癌胚抗原(CEA)、甲胎蛋白(AFP)、糖类抗原 19-9(CA19-9)和 CA72-4 配套试剂盒(美国罗氏公司)。

1.3 方法

1.3.1 采集标本 采集受试者空腹静脉血 3 mL, 分出血清后 4℃ 3 000 $\times$ g 离心 5 min。将分离的血清转入 Eppendorf 管, 再次 4℃ 3 000 $\times$ g 离心 5 min, 每管分装 50 μ L, 冻存于-80℃冰箱中备用。剩余血清用于肿瘤标志物测定。

1.3.2 蛋白标志物模型

1.3.2.1 检测蛋白质谱 病例组血清蛋白质谱检测参照姜伟^[2]等关于精神分裂症蛋白组学研究中血清蛋白质谱的检测方法。

1.3.2.2 筛选蛋白标志物 利用 Ciphergen ProteinChip 软件对获得的血清蛋白指纹图谱进行分析, 以正态因子衡量图谱间

[△] 通讯作者, E-mail: limin_hu@yahoo.com.cn.

差异的大小,当结果为 0.3~3.0 之间,则差异可接受。

1.3.2.3 建立诊断模型 ANN 技术是根据生物神经系统原理发展起来的、多个人工神经元互联组成的大规模分布式并行信息处理系统^[3]。本研究以筛选的差异蛋白丰度为输入节点,设定病例组目标输出值 1,对照组目标输出值 0,cut-off 值为 0.5。利用训练组样本建立 ANN 预测模型并反复训练,调整模型检测参数,以完全区分病例组和对照组时诊断效能最高,建立并储存模型。

1.3.2.4 盲法模型验证 采用双盲的方法将验证组中样本的蛋白标志物数据作为输入节点,利用储存的模型进行模拟仿真计算。当输出值大于 0.5 时归为病例组;当预测输出值小于或等于 0.5 时归为对照组。利用验证结果进行诊断效能评价。

1.3.3 肿瘤标志物模型

1.3.3.1 肿瘤标志物的测定 利用电化学发光仪检测血清 CEA、AFP、CA19-9、CA72-4 水平,以 $\bar{x} \pm s$ 表示。

1.3.3.2 建立诊断模型 ANN 模型节点分别设为 4、4、1,设定大肠癌目标输出值 1,对照目标输出值为 0,以 0.5 为 cut-off 值,训练次数 1 000,学习率 0.01,建立并存储模型。

1.3.3.3 盲法模型验证 采用双盲的方法,将验证组肿瘤标志物结果输入模型并模拟仿真计算。当输出值大于 0.5 时,归为病例组;当输出值小于或等于 0.5 时,归为对照组。利用验证结果进行诊断效能评价。

1.3.4 联合多标志物模型

1.3.4.1 建立诊断模型 利用筛选的蛋白标志物和肿瘤标志物数据为输入节点,分别组合并训练,以完全区分病例组和对照组时的标志物组合为最佳组合,建立联合多标志物模型。设定大肠癌目标输出值 1,对照目标输出值 0,训练次数 1 000,学习率 0.01,建立并存储 ANN 模型。

1.3.4.2 盲法模型验证 采用双盲的方法将验证组血清蛋白标志物和肿瘤标志物数据输入模型并模拟仿真计算,设定 cut-off 值 0.5。当输出值大于 0.5 时归为大肠癌;当输出值小于或等于 0.5 时归为对照组。通过盲测结果进行诊断效能评价。

1.4 统计学方法 采用 Biomarker Wizard 3.1 软件分析蛋白指纹图谱,设定有意义的蛋白质峰出现频率阈值为 10%,分别设定 5、2 两次信噪比过滤。蛋白质峰值比较时采用 $\bar{x} \pm s$ 表示,并进行 *t* 检验,以 $P < 0.01$ 为差异有统计学意义。

2 结 果

2.1 蛋白标志物模型

2.1.1 筛选蛋白标志物并建立模型 对获得的病例组和对照组血清质谱图统计后,正态系数为 2.27,正态因子为 0.3~3.0 之间,聚类分析共得到 328 个蛋白质峰。以其中 19 个蛋白比较差异有统计学意义($P < 0.01$)的蛋白作为潜在标志物优化组合,建立 ANN 预测模型。利用训练组样本反复训练,显示质荷比 4 095、5 640、4 480、7 620 *m/z* 的 4 个蛋白作为输入节点,隐含层和输出层节点分别设置为 4、1 时建立的模型,仿真计算效能最高,正确地区分开了病例组和对照组,以此建立并储存模型。

2.1.2 盲法验证结果 盲法验证结果显示 26 例大肠癌患者,有 2 例错判;而 30 例对照中有 5 例辨别错误。结果表明该模型诊断大肠癌的灵敏度为 92.3%,特异度 83.3%。

2.2 肿瘤标志物模型

2.2.1 肿瘤标志物水平 血清 CEA、AFP、CA72-4、CA19-9 检测结果见表 1。

2.2.2 盲法验证结果 利用验证组样本的肿瘤标志物结果进

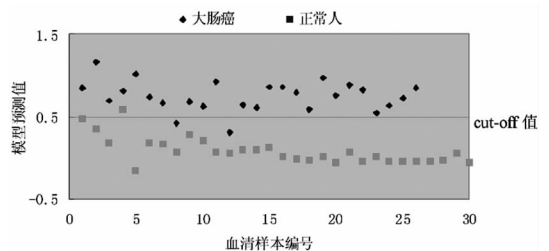
行盲法验证,结果显示,26 例大肠癌患者中有 7 例判断错误,而 30 例对照中有 4 例辨别错误。结果表明该模型诊断大肠癌灵敏度为 73.1%,特异度 86.7%。

2.3 混合标志物模型 通过优化组合并反复训练表明:利用筛选的 4 个蛋白标志物和 CEA、CA72-4 建立的模型诊断效能最高,完全区分开了病例组和对照组。利用该模型盲法检测标本,结果 26 例大肠癌中有 2 例错判,而 30 例对照中有 1 例判断错误。结果表明该模型诊断大肠癌灵敏度为 92.3%,特异度为 96.7%(图 1)。

表 1 两组血清 CEA、AFP、CA72-4 和 CA19-9 含量($\bar{x} \pm s$)

组别	<i>n</i>	CEA (ng/mL)	AFP (ng/mL)	CA72-4 (u/mL)	CA19-9 (u/mL)
对照组	55	1.7±1.0	3.2±1.5	1.9±2.0	10.5±7.8
病例组 A	6	5.9±5.3	4.1±6.5	1.6±1.4	29.7±31.7
病例组 B	22	14.5±29.1	8.1±18.9	4.8±4.8	66.3±165.7
病例组 C	15	229.6±195.4	3.87±1.47	58.9±56.8	472.4±487.1
病例组 D	8	238.4±168.1	3.67±1.98	44.1±50.1	446.9±601.8

注:病例组与对照组血清 CEA、CA72-4、CA19-9 差异均具有统计学意义($P < 0.05$)。



注:以 0.5 为 cut-off 值,26 例大肠癌(◆所示)中有 2 例误判,预测值小于 0.5,而 30 例正常人中有 1 例错判(■所示),预测值大于 0.5。

图 1 蛋白标志物和肿瘤标志物联合诊断模型对验证组样本盲法预测结果

3 讨 论

肿瘤标志物的检查是临床筛查肿瘤高危人群,评价病情进展及疗效观察的重要指标。目前应用于大肠癌的肿瘤标记物大多敏感性和特异性较差,难以用于早期诊断。从蛋白质水平研究肿瘤的发生与发展,寻找对肿瘤特定表达的标志蛋白,或许会成为疾病潜在分子诊断及治疗的新靶点。SELDI-TOF-MS 技术通过分析蛋白质水平上导致疾病的蛋白质动态变化情况实现生物标志物的发现和确证,从而辅助疾病早期筛查和鉴别诊断^[4]。

本研究利用 SELDI-TOF-MS 检测大肠癌患者和对照血清蛋白指纹图谱,通过比对分析发现 19 个表达差异的蛋白质峰。筛选 4 095、5 640、4 480、7 620 *m/z* 的 4 个蛋白作为标志物,结合 ANN 建立预测模型用于大肠癌诊断,结果显示其敏感性为 92.3%,特异性为 83.3%。4 个蛋白标志物进一步结合 CEA 和 CA72-4,建立多标志物联合预测模型,结果显示其灵敏度为 92.3%,特异性为 96.7%,显示了较高的实验诊断价值。

但本研究尚存在以下不足:样本量和病种偏少,需要扩大病例、病种检测并验证模型的可靠性,同时患者人口学特征、疾病严重程度对蛋白标志物的影响不明显。总之,采用 SELDI-TOF-MS 检测大肠癌血清特征蛋白并建立 ANN 实验诊断模型,印证了在蛋白分子水平诊断大肠癌的可行(下转第 944 页)

症状明显改善,鼻腔黏膜、下甲、中甲黏膜部分区域水肿充血,中鼻道及总鼻道有少量脓性或黏脓样分泌物;鼻窦 CT 示窦腔炎症减轻,黏膜增厚不明显,无明显积液。无效:症状无改善,鼻腔黏膜、下甲、中甲充血水肿,中鼻道及总鼻道可见脓性及黏脓性分泌物;鼻窦 CT 示窦腔积液。

1.4 统计学方法 计数资料用 χ^2 检验,数据处理由 SPSS 14.0 软件完成, $P<0.05$ 为差异有统计学意义。

2 结 果

两组治疗后病情恢复情况见表 1,可见两组好转率经检验,差异没有统计学意义($P>0.05$)。治疗组治愈率明显高于对照组,无效率明显低于对照组,经检验,差异具有统计学意义($P<0.05$)。

表 1 两组治疗后病情恢复比较[n(%)]

组别	n	治愈	好转	无效
治疗组	45	29(64.44)	14(31.11)	2(4.44)
对照组	40	12(30.0)	20(50.0)	8(20.0)
χ^2		10.062	3.148	4.936
P		0.002	0.076	0.026

3 讨 论

慢性鼻窦炎是小儿常见疾病,感染与变态反应造成的黏膜水肿是发病的主要原因,多与小儿鼻腔及鼻窦解剖、生理的特异性相关。小儿窦口较大,病原菌易经窦口侵入鼻窦。鼻腔及鼻窦腔相对狭窄,鼻窦尚未发育完全,黏膜嫩弱,淋巴管和血管丰富,感染或变态反应致黏膜充血水肿明显、分泌物较多,引起窦口阻塞,局部纤毛运动减弱^[3-4],为病原菌的繁殖提供了合适生长环境,如此形成一个恶性循环,此外患儿抵抗力弱,对外界适应能力较差,最终发生慢性鼻窦炎。

在治疗上,传统的手术虽然病灶清除彻底,但创伤大,对小儿鼻窦、齿槽、面骨的发育均有影响。上颌窦穿刺冲洗术效果明显,适合于年龄较大的患儿,对于年龄偏小的患儿,配合较

差,再加上窦口肿胀狭小,滴鼻药很难进入窦内。随着内镜手术的运用,其对鼻窦炎的诊疗效果显著,但是临床实践显示,内镜手术治疗后的患儿鼻窦炎没有获得成人患者相同的疗效,考虑其原因是术后清理鼻腔困难,容易出现鼻腔及窦腔粘连。由于以上的因素,结合患儿鼻腔、鼻窦有较强的生理功能,病变后可逆性恢复能力强,故手术一般不作为治疗首选,宜首选保守治疗。负压置换术就是非手术治疗的一个重要方法^[4],鼻腔负压置换治疗是使鼻腔内产生间断性负压,从而将鼻腔和鼻窦内空气及脓性分泌物吸出,并借负压作用将药物灌入鼻窦内,使其产生抗感染、抗过敏的作用,从而缓解鼻黏膜水肿,解除阻塞,使局部组织缺氧得到改善,增加纤毛活动,改变病原菌生存环境,从而促进炎症的消散,恢复鼻腔及窦内生理^[5]。本次研究中,治疗组在给予负压置换术后,病情恢复明显,其治愈率明显高于对照组,无效率明显低于对照组,差异具有统计学意义($P<0.05$)。

综上所述,负压置换治疗对患儿慢性鼻窦炎疗效确切、痛苦小、操作简便,对设备要求不高,值得临床推广使用。

参考文献

- [1] 阎承先. 小儿耳鼻咽喉科学[M]. 天津:天津科学技术出版社,2000:418.
- [2] 闵双凤. 简易负压置换疗法治疗儿童鼻窦炎 260 例疗效观察[J]. 郑州大学学报:医学版,2007,42(4):792-794.
- [3] 周如平,周洪根,黄群. 儿童慢性鼻窦炎的诊断与治疗[J]. 江苏医药,2009,35(3):350-352.
- [4] 许庚,李源. 儿童慢性鼻窦炎手术治疗的思考与临床诊疗指引[J]. 中华耳鼻咽喉科杂志,2003,38(4):241-243.
- [5] 黄兆挑,汪吉宝. 实用耳鼻喉科学[M]. 北京:人民卫生出版社,2002:1133.

(收稿日期:2010-12-15)

(上接第 942 页)

性。同时建立肿瘤标志物智能诊断模型可利用现有血清肿瘤标志物及其特点,对大肠癌进行简便、快速、综合判断,以辅助疾病诊断和治疗评价。多标志物联合预测模型结合了蛋白标志物和肿瘤标志物,具有高通量、高灵敏度和高特异度的优点,在大肠癌诊断和疗效观察,监测肿瘤复发和转移具有潜在价值,得进一步研究与应用。

参考文献

- [1] Ward DG, Suggett N, Cheng Y, et al. Identification of serum biomarkers for colon cancer by proteomic analysis[J]. Br J Cancer, 2006, 94(12):1898-1905.
- [2] 姜伟,王开正,白克镇,等. 精神分裂症患者血清蛋白标志物的临床应用研究[J]. 中国神经精神疾病杂志,2008,34(1):27-30.
- [3] Chen YD, Zheng S, Yu JK, et al. Artificial neural net-

works analysis of surface-enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population[J]. Clin Cancer Res, 2004; 10(24):8380-8385.

- [4] 姜伟,杨永长,肖代雯,等. 尿蛋白标志物模型早期诊断糖尿病肾病的临床应用[J]. 中华检验医学杂志,2009,32(10):1101-1107.
- [5] 刘刚,柳红,刘艺萍,等. 人工神经网络在大肠癌患者术后 5 年生存期预测中的应用[J]. 中国卫生统计,2010,27(3):240-242.
- [6] 高蔚,聂绍发,施倡元. 神经网络在生存分析中的应用进展[J]. 中国卫生统计,2006,23(4):358-360.
- [7] 贺佳,张智坚,刘崎,等. BP 神经网络预测肝癌患者生存期的研究[J]. 中国卫生统计,2001,18(1):17-19.

(收稿日期:2010-12-22)