

遗传性非息肉病性结直肠癌血清的蛋白质组学

魏巍, 尚秀娟, 吕保洲

魏巍, 尚秀娟, 吕保洲, 河南省新乡市中心医院肛肠科 河南省新乡市 453000

作者贡献分布: 此课题由魏巍与尚秀娟设计; 研究过程及数据分析由魏巍、尚秀娟及吕保洲操作完成; 本论文写作由魏巍完成。

通讯作者: 魏巍, 453000, 河南省新乡市金穗大道56号, 新乡市中心医院肛肠科. weew120@126.com

电话: 0373-2041093

收稿日期: 2010-12-22 修回日期: 2011-04-11

接受日期: 2011-04-21 在线出版日期: 2011-05-08

Serum proteomic analysis of hereditary nonpolyposis colorectal cancer

Wei Wei, Xiu-Juan Shang, Bao-Zhou Lv

Wei Wei, Xiu-Juan Shang, Bao-Zhou Lv, Department of Proctology, the Central Hospital of Xinxiang, Xinxiang 453000, Henan Province, China

Correspondence to: Wei Wei, Department of Proctology, the Central Hospital of Xinxiang, 56 Jinsui Avenue, Xinxiang 453000, Henan Province, China. weew120@126.com

Received: 2010-12-22 Revised: 2011-04-11

Accepted: 2011-04-21 Published online: 2011-05-08

Abstract

AIM: To identify differentially expressed serum proteins between preoperative patients with hereditary nonpolyposis colorectal cancer (HNPCC) and those with sporadic colorectal cancer to find potential markers for the diagnosis of HNPCC.

METHODS: The composition of serum proteins was detected by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) and protein chip assay in 20 preoperative HNPCC patients and 25 patients with sporadic colorectal cancer. Discriminatory serum protein profiles were analyzed with the Biomarker Wizard software and Biomarker Pattern software.

RESULTS: Two proteins (m/z 7 516.68 Da and 4 827.07 Da) were chosen to create a decision classification tree model which could discriminate the two groups of patients. In the test mode, the accuracy, sensitivity, specificity, and positive predictive value of the model for prediction of HNPCC was 86.7%, 72.3%, 95.1% and 92.5%,

respectively, while the corresponding values for blinded validation were 75.6%, 100%, 69.8% and 99.2%, respectively.

CONCLUSION: The classification tree model constructed with two differentially expressed proteins identified by SELDI-TOF-MS between HNPCC patients and patients with sporadic colorectal cancer possesses high sensitivity and specificity in the prediction of HNPCC.

Key Words: Hereditary nonpolyposis colorectal cancer; Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry; Proteomics

Wei W, Shang XJ, Lv BZ. Serum proteomic analysis of hereditary nonpolyposis colorectal cancer. *Shijie Huaren Xiaohua Zazhi* 2011; 19(13): 1417-1421

摘要

目的: 研究遗传性非息肉病性结直肠癌(HNPCC)与散发性结直肠癌患者术前血清蛋白质的表达差异, 以期发现可用于HNPCC诊断的生物学指标。

方法: 应用表面增强激光解吸电离飞行时间质谱(SELDI-TOF-MS)技术结合蛋白质芯片分别检测20例HNPCC和25例散发性结直肠癌患者术前血清蛋白质组分。将获得的蛋白质谱采用美国CiphaGen公司的Biomarker Wizard和Biomarker Pattern软件进行分析。

结果: 以质荷比为7 516.68 Da和4 827.07 Da建立的分类决策树模型能将HNPCC和散发性结直肠癌患者区分开来。该模型在测试模式下诊断准确率为86.7%、灵敏度为72.3%、总特异度95.1%、阳性预测值92.5%。盲法验证模式下诊断准确率为75.6%、灵敏度69.8%、总特异度99.2%、阳性预测值100%。

结论: 利用SELDI-TOF-MS技术建立的HNPCC诊断分类树模型具有较高的敏感度和特异度。

关键词: 遗传性非息肉病性结直肠癌; 表面增强激光解吸电离飞行时间质谱; 蛋白质组学

魏巍, 尚秀娟, 吕保洲. 遗传性非息肉病性结直肠癌血清的蛋白

■背景资料

遗传性非息肉病性结直肠癌(HNPCC)是一种常染色体显性遗传病。当前HNPCC的诊断主要依靠MMR基因突变和MSI检测。但上述手段存在检测过程耗时长、检测结果不稳定以及诊断灵敏度低等不足。因此, 探索一种新的检测手段对于HNPCC的早期诊断来讲, 显得至关重要。

■同行评议者

顾国利, 副主任医师, 中国人民解放军空军总医院普通外科

■研究前沿

HNPCC是一种常染色体显性遗传病。作为结直肠癌的一个重要亚型,其病因和临床病理类型较为特殊,同时具有肿瘤的遗传性、高发性、多发性和肿瘤发生的多器官性等特点,是目前结直肠癌和遗传性肿瘤的一个研究热点。

质组学. 世界华人消化杂志 2011; 19(13): 1417-1421
<http://www.wjgnet.com/1009-3079/19/1417.asp>

0 引言

遗传性非息肉病性结直肠癌(hereditary nonpolyposis colorectal cancer, HNPCC)是一种常染色体显性遗传病。作为结直肠癌的一个重要亚型,其病因和临床病理类型较为特殊,同时具有肿瘤的遗传性、高发性、多发性和肿瘤发生的多器官性等特点,是目前结直肠癌和遗传性肿瘤的一个研究热点^[1,2]。本实验通过对HNPCC和散发性结直肠癌患者术前血清蛋白质进行分析,探讨两者之间血清蛋白的表达差异,建立HNPCC疾病诊断模型。

1 材料和方法

1.1 材料 新乡市中心医院2005-08/2010-07住院的结直肠癌患者共计2 875例,从中选取HNPCC患者术前血清标本20例,其中男13例,女7例,年龄26-75(平均 52 ± 6)岁。入选标准参照阿姆斯特丹标准II^[3],对符合HNPCC临床诊断标准者在课题研究前先行错配修复基因突变(mismatch repair, MMR)和微卫星不稳定(microsatellite instability, MSI)检测以进一步印证HNPCC的诊断。同时随机选取同时期散发性结直肠癌患者术前血清标本25例,其中男15例,女10例,年龄30-82(平均 57 ± 8)岁。HNPCC和散发性结直肠癌病例术后均经病理学确诊。应用随机数字表把血清标本分成8个样本1个实验组。取样人群具有相似的暴露史、相同的民族。所有受试对象均知情同意并排除存在影响血清蛋白质含量的自身免疫、肝炎、过敏等相关性疾病。尿素(urea)、三羟基甲基氨基甲烷-盐酸缓冲液(Tris-HCL)、3-环乙胺-1-丙磺酸(CHAPS)、芥子酸(SPA)、乙腈(ACN)、三氟乙酸(TFA)、N-2-羟乙基哌嗪-N'-2-乙基磺酸(HEPES)均购自Sigma公司, SELDI蛋白质芯片仪(美国Ciphergen公司)、各种类型SELDI蛋白质芯片(美国BIO-RAD Laboratories公司)。

1.2 方法

1.2.1 血清标本的采集与保存: 患者于术前清晨清醒空腹状态下采集静脉血3-4 mL, 4 °C冰箱中静置2 h后离心(1 000 r/min, 30 min), 取血清按100 μ L/管分装, 置于-70 °C冰箱中保存。

1.2.2 血清标本准备: 4 °C环境下融解冰冻血清, 每个样品取10 μ L分别置于1.5 mL离心管

中, 每管加20 μ L U9缓冲液(含9 mol/L尿素, 2% CHAPS, 50 mmol/L Tris-HCl pH9.0), 4 °C环境下200 r/min振荡30 min, 使蛋白质变性。分别取120 μ L IMAC结合/洗脱缓冲液(100 mmol/L磷酸钠+500 mmol/L氯化钠, pH7.0)置于1.5 mL离心管中, 每管加入10 μ L上述变性后的血清样品后混匀, 4 °C振荡30 min。

1.2.3 芯片处理: 将芯片放入操作平台, 每孔加100 mmol/L硫酸铜50 μ L, 4 °C环境下200 r/min振荡5 min, 使 Cu^{2+} 结合于芯池中, 倒除硫酸铜溶液, 用去离子水冲洗5次后甩干。每孔中加入100 mmol/L醋酸钠(pH4.0)50 μ L, 4 °C环境下200 r/min振荡5 min后倒除醋酸钠溶液, 用去离子水冲洗5次后甩干, 以除去没有结合的 Cu^{2+} 。用去离子水冲洗5次, 吸水纸滤干操作平台。每孔加入IMAC结合/洗脱缓冲液150 μ L, 4 °C环境下200 r/min振荡5 min, 除去缓冲液, 重复1次。每孔加入稀释好的血清样品50 μ L上样, 4 °C环境下200 r/min振荡60 min后除去未结合的样品。每孔加入150 μ L IMAC结合/洗脱缓冲液, 4 °C环境下200 r/min振荡5 min, 除去缓冲液, 重复1次。吸水纸滤干操作平台, 将芯片移出操作平台, 每孔风干, 每孔加入0.5 μ L SPA饱和溶液/次, 共2次, 风干后即可上机测定。

1.2.4 芯片检测: 用加有All-in-one标准蛋白质的NP20芯片校正质谱仪, 设定仪器参数(Digitizer rate = 1 000, Mass Deflector Value = 2)。设定所有血清样品检测的蛋白质相对分子质量区间在1 500-200 000 Da, 低于1 500 Da的蛋白质或肽段将自动从波谱中被清除, 因在0-1 500 Da这一区间可存在加合物、试剂成分、SPA以及其他化学成分, 影响后续的数据分析。设定低能量激光强度读取低分子量蛋白质数据设最高分子量为50 000 Da, 最佳状态从3 000-15 000 Da, 并选择最佳激光强度和最佳检测灵敏度, 聚焦到最佳中心, 设定数据收集方法为seldi quantitation, 其数据收集参数acquisition parameters为20, delta 4, transients per to 10 ending position to 80, 设预热位置用2个点, 选择激光强度为激光最佳强度+10, 数据收集不包括预热点。设定高能量激光强度(high laser intensity)读取高分子量蛋白质数据。设最高分子量为200 000 Da, 最佳状态从10 000-50 000 Da, 设开始激光强度为低激光强度+30即为最佳激光强度, 设开始监测器灵敏度为最佳检测灵敏度, 并聚焦到最佳中心, 设定数据收集方法为seldi quantitation, 其数据收集参数

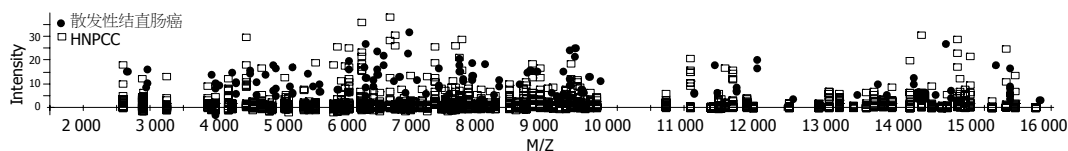


图1 HNPCC与散发性结直肠癌患者血清中相同质荷比的蛋白质相对含量比较。

■应用要点

本文利用SELDI-TOF-MS技术建立的HNPCC诊断分类树模型具有较高的敏感度和特异性。

acquisition parameters为22, delta 4, transients per to 10 ending position to 82, 设预热位置用2个点, 选择激光强度为激光最佳强度+10, 数据收集不包括预热点。

1.2.5 数据处理: 利用Ciphergen ProteinChip软件完成对蛋白质质谱信息的录入和质谱图的绘制, 其中横坐标为质荷比, 即蛋白质相对分子质量, 纵坐标为波峰强度, 即蛋白质相对含量。Bio-marker Wizard统计软件用于对各组相同M/Z条件下其血清中不同的蛋白质相对含量进行计算, 从而筛选出差异蛋白。用Biomarker Pattern数据挖掘软件通过定义一个完善的训练-验证协议把所有待检样本数据分成训练组和测试组, 在学习模式下(最优分类决策树模型建立的过程)对训练组用n倍交叉方法建立最优分类决策树模型, 以保证其正确性, 然后在测试模式(检验该模型诊断价值的过程)下用测试组对最优分类决策树模型进行验证和测试, 分别得到此模型对训练组和测试组的分组准确率、灵敏度、特异度以及阳性预测值。

统计学处理 采用Biomarker Wizard统计软件分别对HNPCC和散发性结直肠癌患者血清蛋白质的波峰强度进行t检验, 数据均以mean±SD表示, $P<0.05$ 认为差异有统计学意义。采用Biomarker Pattern软件用10倍交叉验证方法利用差异蛋白建立最优分类决策树模型。

2 结果

2.1 HNPCC与散发性结直肠癌患者术前血清蛋白质表达存在差异 采用Biomarker Wizard软件对所测得20例HNPCC与25例散发性结直肠癌患者血清蛋白质进行比较, 结果显示在质荷比2 700-16 500 Da范围内检测到的26个蛋白质峰中, 有5个蛋白质荷比位点处的波峰强度有统计学意义($P<0.05$, 图1, 表1)。

2.2 差异蛋白在HNPCC和散发性结直肠癌患者血清中的表达情况 比较5个差异蛋白在血清中的相对含量, 结果显示在HNPCC患者血清中均呈显著下调。在5个差异蛋白中, 随即选取质荷比为7 516 Da和9 604 Da的两个差异蛋白在

表1 两组患者血清中5个差异蛋白的波峰强度比较

质荷比(Da)	波峰强度		P值
	散发性结直肠癌	HNPCC	
7 516.68	8.63 ± 1.41	1.45 ± 0.33	0.01278
4 827.07	10.11 ± 7.34	5.17 ± 3.31	0.02519
9 604.74	5.52 ± 4.56	2.77 ± 2.05	0.02832
6 147.44	7.02 ± 4.36	3.57 ± 1.49	0.03001
12 720.83	2.35 ± 1.50	1.89 ± 0.47	0.04948

HNPCC与散发性结直肠癌患者血清中表达的波谱图与模拟电泳图如下(图2)。

2.3 生物标志物的发现及其分类决策树模型的建立和分析 将Biomarker Wizard软件分析所得到的有差别的蛋白质峰建立数据库, 导入Biomarker Pattern统计分析软件, 选择相应条件对20例HNPCC以及25例散发性结直肠癌患者术前血清中蛋白质组成进行分析, 结果发现以质荷比为7 516.68 Da和4 827.07 Da两个差异蛋白组成的分类决策树模型(图3)可将HNPCC和散发性结直肠癌两组区分开来。用这个生物标志物数据共产生3个终节点。在节点1处共有45例结直肠癌患者, 有27例患者血清差异蛋白的相对含量满足条件 M/Z 7 516.68 Da \leq 2.218被分到节点2, 余18例被分到终节点3; 节点2处27例患者中有21例患者血清差异蛋白的相对含量满足条件 M/Z 4 827.07 Da \leq 0.282被分到终节点1, 余6例被分到终节点2。

2.4 分类决策树模型分别在软件学习和测试模式下的分类结果 在软件学习模式下对20例HNPCC和25例散发性结直肠癌进行分组, 分组准确率为86.7%(39/45), 在软件测试模式下对上述样本进行双盲检测, 其分组准确率为75.6%(34/45)。分类决策树模型在软件两种模式下对各组的错分情况见表2。

3 讨论

HNPCC是一种常染色体显性遗传病, 当前HNPCC的诊断主要依靠MMR基因突变和MSI检测^[4-6]。但上述手段存在检测过程耗时长、检测结果不稳定以及诊断灵敏度低等不足^[7-9]。因

■同行评价

本文实验数据较充足, 结论较明确, 具有一定的科学性和创新性。

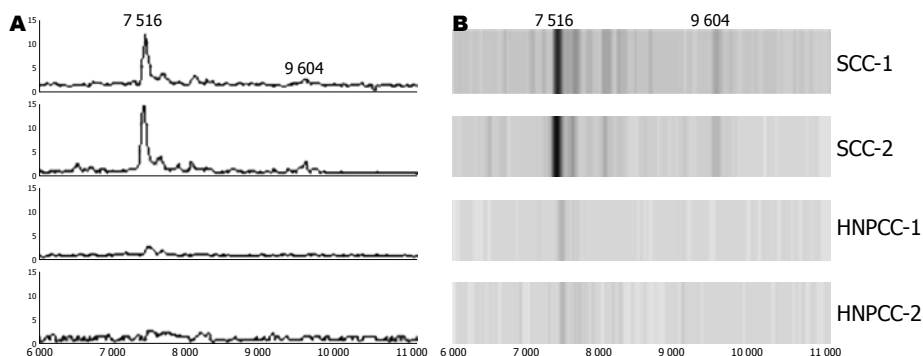


图2 质荷比为7 516 Da和9 604 Da的两个差异蛋白的波谱图及模拟电泳图。A: 波谱图; B: 模拟电泳图。

表2 分类树模型在两种模式下的错分情况 (n)

分组	n	错分	
		学习模式	测试模式
HNPCC	20	1	3
散发性结直肠癌	25	5	8

此, 探索一种新的检测手段对于HNPCC的早期诊断来讲, 显得至关重要。

SELDI-TOF-MS技术具有高通量、操作简便、样本处理简单等优点, 目前已广泛应用于卵巢癌、前列腺癌、肝癌、食管癌、结直肠癌等多种肿瘤的研究^[10-14]。我们采用SELDI-TOF-MS技术对20例HNPCC和25例散发性结直肠癌的血清标本进行了蛋白质指纹图谱的测定, 并应用相关软件进行数据分析。建立的模型分别对70例测试血清标本进行盲法验证, 得到准确率为75.6%、灵敏度69.8%、总特异度99.2%、阳性预测值100%, 明显优于当前HNPCC的检测手段(诊断准确率为40%-60%)^[7-9]。此分类决策树模型的建立, 不仅在HNPCC临床筛查及诊断方面可能发挥更大的潜力, 同时丰富了结直肠癌蛋白质诊断模板。

但由于我们的病例数量有限, 由年龄和性别带来的干扰难以通过抽样检查。因此, 我们将所有样本按年龄和性别分类后, 在以上各人群组内、组间及总样本间进行比较, 建模时将有明显差异的蛋白质剔除, 以此排除年龄和性别带来的干扰。但该模型诊断一致性强度一般, 我们考虑与以下因素有关: (1)样本量有限; (2)由于同一个M/Z可对应多个蛋白质。因此, 依据性别、年龄分组比较后筛选的蛋白质可能同时对应了性别、年龄差异蛋白及对筛查诊断有意义的蛋白, 由此损失了一些生物学信息。今后我们会不断扩大样本量, 对分类树模型进行不断的修正, 以提高诊断的准确率、灵敏度以及特异度, 并通过双盲法对该模型进行验证。同时我们

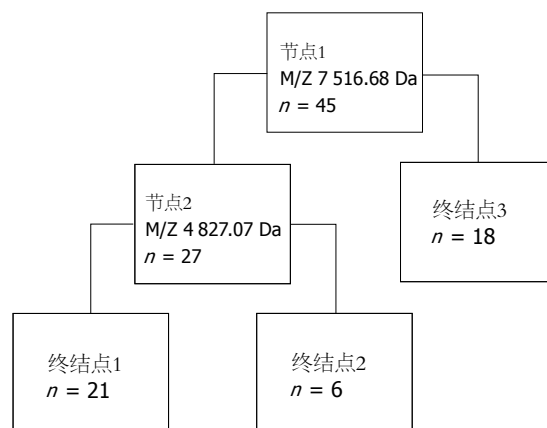


图3 分类树模型简图。Node: 节点; Terminal Node: 终结点; N: 例数。

也将对质荷比为7 516.68 Da和4 827.07 Da的这种蛋白质进行长期、动态、定量的监测, 分析术前、术后长期的变化规律, 并对其进行鉴定, 希望能对HNPCC的病因、发病机制、诊断以及治疗等研究提供理论依据。

4 参考文献

- Wang J, Luo MH, Zhang ZX, Zhang PD, Jiang XL, Ma DW, Suo RZ, Zhao LZ, Qi QH. Clinical and molecular analysis of hereditary non-polyposis colorectal cancer in Chinese colorectal cancer patients. *World J Gastroenterol* 2007; 13: 1612-1617
- 袁瑛, 曹文明, 蔡善荣, 张苏展. 中国人遗传性非息肉病性结直肠癌家系的临床表型分析. *中华肿瘤杂志* 2006; 28: 36-38
- 顾国利, 周晓武, 王石林. 遗传性非息肉病性大肠癌的研究进展. *世界华人消化杂志* 2007; 15: 3115-3121
- Papp J, Kovacs ME, Olah E. Germline MLH1 and MSH2 mutational spectrum including frequent large genomic aberrations in Hungarian hereditary non-polyposis colorectal cancer families: implications for genetic testing. *World J Gastroenterol* 2007; 13: 2727-2732
- Chialina SG, Fornes C, Landi C, de la Vega Elena CD, Nicolich MV, Dourisboure RJ, Solano A, Solis EA. Microsatellite instability analysis in hereditary non-polyposis colon cancer using the Bethesda consensus panel of microsatellite markers in the absence of proband normal tissue. *BMC Med Genet* 2006; 7: 5
- Liu SR, Zhao B, Wang ZJ, Wan YL, Huang YT.

- Clinical features and mismatch repair gene mutation screening in Chinese patients with hereditary nonpolyposis colorectal carcinoma. *World J Gastroenterol* 2004; 10: 2647-2651
- 7 Church J, Simmam C. Practice parameters for the treatment of patients with dominantly inherited colorectal cancer (familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer). *Dis Colon Rectum* 2003; 46: 1001-1012
- 8 Müller A, Beckmann C, Westphal G, Bocker Edmonston T, Friedrichs N, Dietmaier W, Brasch FE, Kloor M, Poremba C, Keller G, Aust DE, Fass J, Büttner R, Becker H, Rüschoff J. Prevalence of the mismatch-repair-deficient phenotype in colonic adenomas arising in HNPCC patients: results of a 5-year follow-up study. *Int J Colorectal Dis* 2006; 21: 632-641
- 9 Sheng JQ, Chan TL, Chan YW, Huang JS, Chen JG, Zhang MZ, Guo XL, Mu H, Chan AS, Li SR, Yuen ST, Leung SY. Microsatellite instability and novel mismatch repair gene mutations in northern Chinese population with hereditary non-polyposis colorectal cancer. *Chin J Dig Dis* 2006; 7: 197-205
- 10 钦伦秀. 肿瘤的分子诊断与预测. 上海: 上海科技教育出版社, 2006: 369-402
- 11 郑树. 结直肠肿瘤基础研究与临床实践. 北京: 人民卫生出版社, 2006: 268-280
- 12 Grizzle WE, Semmes OJ, Basler J, Izbicka E, Feng Z, Kagan J, Adam BL, Troyer D, Srivastava S, Thornquist M, Zhang Z, Thompson IM. The early detection research network surface-enhanced laser desorption and ionization prostate cancer detection study: A study in biomarker validation in genitourinary oncology. *Urol Oncol* 2004; 22: 337-343
- 13 Mueller J, von Eggeling F, Driesch D, Schubert J, Melle C, Junker K. ProteinChip technology reveals distinctive protein expression profiles in the urine of bladder cancer patients. *Eur Urol* 2005; 47: 885-893; discussion 893-894
- 14 Pan YZ, Xiao XY, Zhao D, Zhang L, Ji GY, Li Y, Yang BX, He DC, Zhao XJ. Application of surface-enhanced laser desorption/ionization time-of-flight-based serum proteomic array technique for the early diagnosis of prostate cancer. *Asian J Androl* 2006; 8: 45-51

编辑 曹丽鸥 电编 何基才

ISSN 1009-3079 (print) ISSN 2219-2859 (online) CN 14-1260/R 2011年版权归世界华人消化杂志

• 消息 •

《世界华人消化杂志》参考文献要求

本刊讯 本刊采用“顺序编码制”的著录方法,即以文中出现顺序用阿拉伯数字编号排序.提倡对国内同行近年已发表的相关研究论文给予充分的反映,并在文内引用处右上角加方括号注明角码.文中如列作者姓名,则需在“Pang等”的右上角注角码号;若正文中仅引用某文献中的论述,则在该论述的句末右上角注角码号.如马连生^[1]报告……,潘伯荣等^[2-5]认为……;PCR方法敏感性高^[6-7].文献序号作正文叙述时,用与正文同号的数字并排,如本实验方法见文献[8].所参考文献必须以近2-3年SCIE, PubMed,《中国科技论文统计源期刊》和《中文核心期刊要目总览》收录的学术类期刊为准,通常应只引用与其观点或数据密切相关的国内外期刊中的最新文献,包括世界华人消化杂志(<http://www.wjgnet.com/1009-3079/index.jsp>)和World Journal of Gastroenterology(<http://www.wjgnet.com/1007-9327/index.jsp>).期刊: 序号, 作者(列出全体作者). 文题, 刊名, 年, 卷, 起页-止页, PMID编号; 书籍: 序号, 作者(列出全部), 书名, 卷次, 版次, 出版地, 出版社, 年, 起页-止页.