

• 论 著 •

生物信息学在筛选影响结直肠癌预后基因中的应用*

伏小红¹,冯宦翔¹,苟伟²,杨敏¹,郭靓¹,屈春燕¹

四川省成都市第七人民医院:1. 检验科;2. 神经内科,四川成都 610000

摘要:**目的** 拟用生物信息学方法,筛选出影响结直肠癌(CRC)预后的相关基因,为寻找 CRC 早期诊断标志物提供指导。**方法** 用生物信息学手段,对正常结直肠和 CRC 组织进行对比分析,筛选出差异常表达基因(DEGs)。**结果** 共检测到 62 个 DEGs,分析发现下调的 DEGs 在肾素和胰腺分泌调节信号通路富集,相关基因为 CLCA1 和 CLCA4,而上调的 DEGs 未见明显的信号通路。对其中 6 个下调基因进行分析,发现 6 个下调基因中有 2 个对结肠腺癌预后有明显影响,通过富集分析发现,CLCA4 在肾素分泌通路中表达明显降低。**结论** CLCA1 和 CLCA4 在 CRC 中表达明显降低,有可能成为 CRC 的潜在诊断标志物和治疗靶点。

关键词:结直肠癌; 差异表达基因; 生物标记物基因

DOI:10.3969/j.issn.1673-4130.2021.10.008 **中图法分类号:**R735.3+5;R446.9

文章编号:1673-4130(2021)10-1184-05 **文献标志码:**A

Application of bioinformatics in screening genes that influence prognosis of colorectal cancer*

FU Xiaohong¹,FENG Huanxiang¹,GOU Wei²,YANG Min¹,GUO Liang¹,QU Chunyan¹

1. Department of Clinical Laboratory;2. Department of Neurology, Chengdu
7th People's Hospital, Chengdu, Sichuan 610000, China

Abstract: Objective To screen out related genes affecting the prognosis of colorectal cancer (CRC), and to provide guidance for finding early diagnostic markers for CRC by using bioinformatics methods. **Methods** Using bioinformatics methods, a comparative analysis of healthy colorectal and CRC tissues was performed to screen out differentially expressed genes (DEGs). **Results** A total of 62 DEGs were detected in this study. The analysis found that the down-regulated DEGs were enriched in the renin and pancreatic secretion regulation signaling pathways. The related genes were CLCA1 and CLCA4, while the up-regulated DEGs had no obvious signaling pathways. Analysis of the 6 down-regulated genes found that 2 of the 6 down-regulated genes had significant differences in the prognosis of adenocarcinoma of colon. Through enrichment analysis, it was found that the expression of CLCA4 in the renin secretion pathway significantly reduced. **Conclusion** The expression of CLCA1 and CLCA4 significantly reduced in CRC, and might become potential diagnostic markers and therapeutic targets for CRC.

Key words:colorectal cancer; differentially expressed genes; biomarker genes

结直肠癌(CRC)的发病率和病死率在消化系统恶性肿瘤中仅次于胃癌、食管癌和原发性肝癌^[1],近年来其发病率在我国不少地区有不断上升的趋势^[2]。有研究发现,普遍的筛查可有效提升 CRC 患者的生存率,降低其病死率^[3]。鉴于筛查的现实需要和目前筛查方式的不足,急需寻找更方便快捷的筛查方法。基因芯片和基因图谱是一种很成熟的检测技术,特别适用于筛选差异表达基因(DEGs)^[4]。运用生物信息学方法整合和分析基因芯片应用过程中积累下来的

大量 CRC 数据,并与表达谱技术相结合,非常适用于研究 CRC 可能存在的生物标记物基因。本研究结合多种数据库开展 CRC 的研究。

1 资料与方法

1.1 筛选出有差异的上调和下调基因 从基因芯片数据库(NCBI-GEO)中获取 GSE110224、GSE41328 和 GSE15960 基因在 CRC 组织及正常结直肠组织中的表达谱,均是 GPL570[G-U133_Plus_2]Affymetrix Human Genome U133 Plus 2.0 Array,包括 28 个正

* 基金项目:四川省医学科研青年创新课题(Q18058)。
作者简介:伏小红,女,主管技师,主要从事临床检验研究。
本文引用格式:伏小红,冯宦翔,苟伟,等.生物信息学在筛选影响结直肠癌预后基因中的应用[J].国际检验医学杂志,2021,42(10):1184-1188.

常结直肠组织标本和 38 个 CRC 组织标本;采用 GEO2R 在线工具对 CRC 组织标本和正常结直肠组织标本的 DEGs 数据进行分析($|\log FC|>2$ 和调整 $P<0.05$);然后再用 Venn 软件对原始数据进行分析,检测 3 个基因芯片之间的差异, $\log FC<0$ 的 DEGs 为下调基因, $\log FC>0$ 的 DEGs 为上调基因。

1.2 分析筛选出的 DEGs 的生物学特征 在 DAVID 数据库里,用基因本体论(GO)分析筛选出的 DEGs 的生物学特征,分成生物过程(BP)组、分子功能(MF)组和细胞成分(CC)组 3 组。用京都基因与基因组百科全书(KEGG)分析信号通路情况。

1.3 分析蛋白质相互作用 通过检索相互作用基因的搜索工具(String)分析蛋白-蛋白相互作用(PPI)网络,再用 MCODE 插件进一步分析。

1.4 分析筛选出的 DEGs 对生存率的影响及对 RNA 的表达 利用 UALCAN 数据库评估筛选出的基因对生存率的影响,计算出有 95% 置信区间的 logrank P 值和风险比(HR)。利用基因表达数据交互分析(GEPIA)网站对标本的 RNA 表达数据进行分析。

2 结 果

2.1 CRC 的 DEGs 分析 用于鉴别的 DEGs 共包括 38 份 CRC 组织标本和 28 份正常结直肠组织标本。通过 GEO2R 在线工具,分别从 GSE110224、GSE41328 和 GSE15960 中提取出 177、290 和 3 235 个 DEGs。用 Venn 软件对 3 个数据集中的常用数据集进行识别,共检测到 62 个常见 DEGs(21 个上调基因,41 个下调基因)。见表 1、图 1。

2.2 DEGs 的 GO 和 KEGG 通路分析 在 DAVID 数据库里采用 GO 分析筛选出的 62 个 DEGs,KEGG 通路分析表明,下调的 DEGs 在肾素分泌和胰腺分泌调节信号通路富集,主要相关的基因是 CLCA1 和 CLCA4,而上调的 DEGs 未见明显的信号通路($P<0.05$),见表 2、3。

2.3 PPI 和 MCODE 插件分析 PPI 分析 36 个 DEGs 被导入到含有 59 个节点和 60 条变的 DEGs PPI 网络复合体中,62 个 DEGs 中有 36 个被纳入到 DEGs PPI 网络,其中包括 10 个上调基因和 26 个下调基因。采用 MCODE 插件进一步分析结果显示,6 个中心节点均为下调基因,见图 2。

表 1 CRC 组织与正常结直肠组织相比得出的 3 个基因芯片都包含的 21 个上调基因和 41 个下调基因

基因表 达情况	基因名称
上调基因	CDH3,CXCL10,UBD///GABBR1,GDF15,MMP7,FOXQ1,KRT80,INHBA,AJUBA,CLDN1,EPHX4,CEMIP,CTHRC1, TESC,CLDN2,CHI3L1,CRNDE,REG1A,MSX2,KRT23,DPEP1
下调基因	HSD17B2,ZG16,UGT2B17,TRPM6,PYY,SLC26A3,PKIB,C10orf99,CLCA1,ABCA8,SLC51B,LYPD8,PCK1,ADH1C, UGT2A3,GCNT2,SST,LAMA1,CWH43,GUCA2A,DHRS9,GUCA2B,CHP2,SCARA5,CLCA4,NXPE4,SLC26A2,AN- PEP,CEACAM7,ADTRP,ISX,BEST2,ADH1B,GCG,GBA3,CLDN8,MS4A12,C2orf88,BEST4,CA1,SI

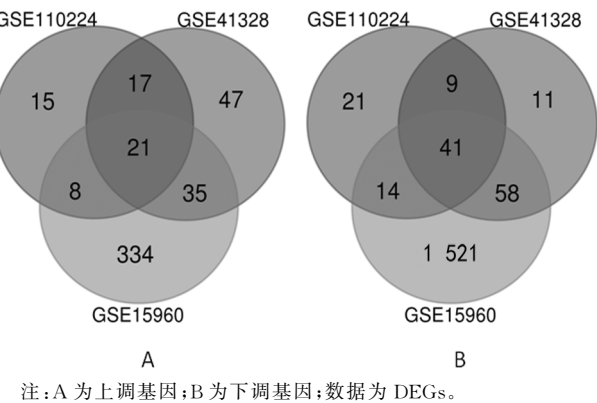


图 1 Venn 软件分析结果

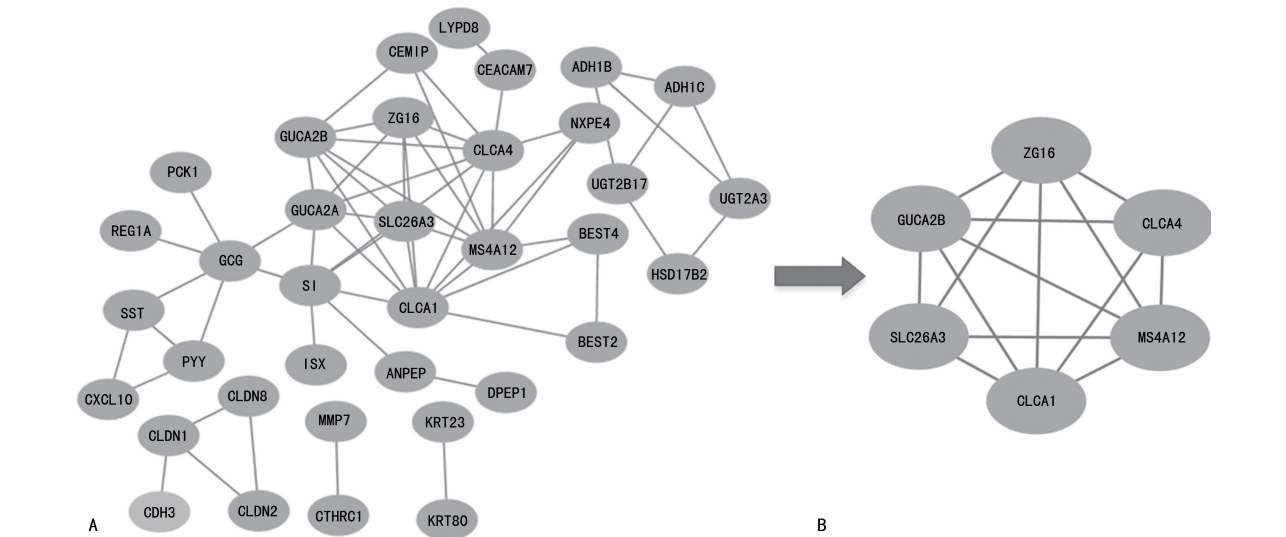
2.4 UALCAN 和 GEPIA 的核心基因分析 CRC 基因表达数据来源于数据库中的结肠腺癌(COAD)和直肠腺癌(READ)两种癌症所表达的数据。利用 UALCAN 数据库对 6 个核心基因对 CRC 的预后影

响进行分析,发现 CLCA1 和 CLCA4 基因低表达时对 COAD 的预后有明显影响($P<0.05$),对 READ 无明显影响,且剩余 4 个基因差异无统计学意义($P>0.05$)。用 GEPIA 对核心基因再进行分析发现,CRC 组织与正常结直肠组织相比,CLCA1 和 CLCA4 基因在 CRC 中表达明显降低($P<0.05$)。见图 3、4。

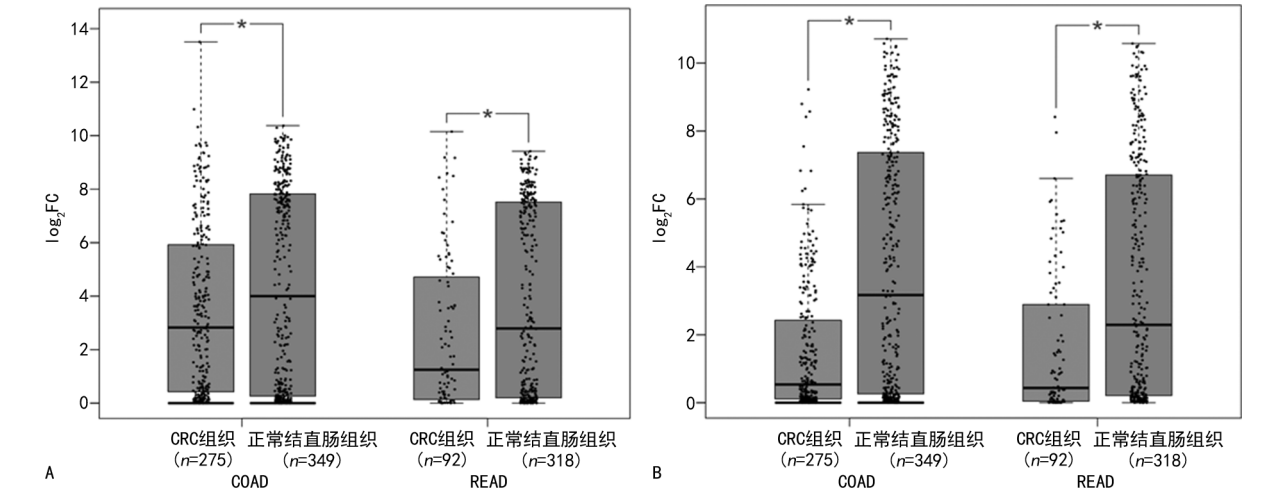
2.5 KEGG 通路富集分析 通过 DAVID 数据库重新对这 3 个筛选出的基因进行 KEGG 通路富集分析,结果显示只有 CLCA4 在肾素分泌的通路中明显表达降低($P<0.05$)。

表 2 CRC 的 DEGs 的 KEGG 通路分析

通路编号	名称	数量 (个)	百分比 (%)	P	基因
hsa04924	肾素分泌	2	66.666 67	0.009 303 678	CLCA1,CLCA4
hsa04972	胰腺分泌	2	66.666 67	0.013 519 407	CLCA1,CLCA4



注：A 为 STRING 分析 PPI 结果；B 为 MCODE 分析所得中心节点。
图 2 PPI 和 MCODE 插件分析结果



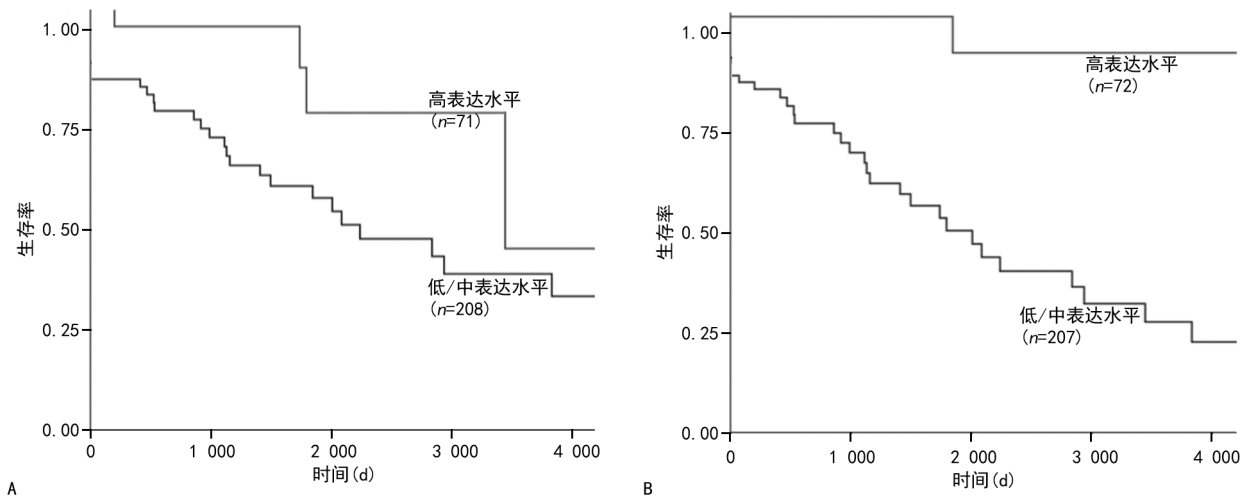
注：A 为 CLCA1 基因；B 为 CLCA4 基因；* $P < 0.05$ 。
图 3 CRC 组织较正常结肠组织表达明显下调的基因

表 3 CRC 的 DEGs 的 GO 分析

表达情况	分类	项目	数量 (个)	百分比 (%)	<i>P</i>	错误发生率 (%)
上调	GOTERM_BP_FAT	GO:0042127~细胞增殖调节	5	3.267 974	0.005 325 0	6.646 518
	GOTERM_BP_FAT	GO:0016338~不依赖钙离子的细胞黏附	2	1.307 190	0.020 945 0	23.866 400
	GOTERM_BP_FAT	GO:0016337~细胞黏附	3	1.960 784	0.027 893 0	30.539 700
	GOTERM_CC_FAT	GO:0044421~细胞外部区域部分	6	3.921 569	0.001 827 0	1.628 932
	GOTERM_CC_FAT	GO:0005615~细胞外空间	5	3.267 974	0.003 963 0	3.502 800
	GOTERM_CC_FAT	GO:0005911~细胞连接	3	1.960 784	0.015 394 0	13.003 150
	GOTERM_MF_FAT	GO:0005198~结构分子活性	5	3.267 974	0.003 808 0	3.488 039
	GOTERM_MF_FAT	GO:0008083~生长因子活性	3	1.960 784	0.012 612 0	11.140 150
	GOTERM_MF_FAT	GO:0005125~细胞因子活性	3	1.960 784	0.018 138 0	15.662 420
下调	GOTERM_BP_FAT	GO:0006811~离子转运	7	2.310 231	0.002 293 0	2.988 199
	GOTERM_BP_FAT	GO:0008610~脂类生物合成的过程	4	1.320 132	0.021 026 0	24.483 890
	GOTERM_BP_FAT	GO:0007584~对营养的反应	3	0.990 099	0.027 306 0	30.641 070
	GOTERM_BP_FAT	GO:0008272~硫酸的运输	2	0.660 066	0.027 379 0	30.708 930
	GOTERM_CC_FAT	GO:0000267~电解反应	9	2.970 297	0.003 496 0	3.663 326

续表 3CRC 的 DEGs 的 GO 分析

表达情况	分类	项目	数量 (个)	百分比 (%)	P	错误发生率 (%)
	GOTERM_CC_FAT	GO:0005903~刷毛缘	3	0.990 099	0.005 513 0	5.721 754
	GOTERM_CC_FAT	GO:0005576~胞外区域	11	3.630 363	0.016 880 0	16.592 690
	GOTERM_MF_FAT	GO:0008509~阴离子跨膜转运体活性	6	1.980 198	0.000 019 7	0.022 198
	GOTERM_MF_FAT	GO:0031404~与氯离子的结合	5	1.650 165	0.000 027 9	0.031 388
	GOTERM_MF_FAT	GO:0043168~与阴离子的结合	5	1.650 165	0.000 056 2	0.063 258
	GOTERM_MF_FAT	GO:0005254~氯离子通道活性	4	1.320 132	0.000 596 0	0.669 397



注:A 为 CLCA1 的表达水平对 COAD 患者生存率的影响;B 为 CLCA4 的表达水平对 COAD 患者生存率的影响。

图 4CLCA1 和 CLCA4 对 CRC 患者预后的影响

3 讨 论

CRC 作为全球癌症相关死亡的第 4 大原因,在 CRC 不同发展阶段诊断的患者 5 年生存率差异很大,早期诊断的患者生存率超过 70%,而晚期诊断的患者生存率不到 10%^[6]。超过一半的 CRC 患者在诊断时已有局部和远处转移,尤其是淋巴结转移严重影响患者的预后^[6]。并且有研究发现越年轻的 CRC 患者预后也越好^[7],因此如果有便捷的早期筛查方法在常规体检或者是转移发生前及早地发现 CRC,能够有效提高 CRC 的生存率。

本研究利用生物信息学方法对 3 个芯片进行了深入分析,集中研究了多个基因,与正常结直肠组织相比,CRC 组织中的 CLCA1 和 CLCA4 存在明显差异,同时发现这两个基因的低表达能够对 COAD 患者的生存率产生影响,这一发现提示 CLCA1 和 CLCA4 可能是诊断和治疗 CRC 的潜在生物标志物。但 CLCA1 和 CLCA4 的低表达通过 GEPIA 分析只对 COAD 的预后有明显影响,而在 READ 组织中的明显低表达却未对其预后有明显影响,这可能与数据库的标本收入量等有关,还需进一步做更深入的研究。

CLCA1 和 CLCA4 是钙活化氯离子通道家族中的成员,与 CLCA2 一起具有相似的一级结构。钙活化氯离子通道调控因子作为一种在氨基酸末端中具

有对称多半胱氨酸基序的蛋白质^[8],在多种癌症中都出现表达失调^[9-10]。同时,对钙活化氯离子通道蛋白的研究发现,其不仅可以调节炎症反应,还参与致癌^[11],由于其相似的一级结构,使其可能在作用效能上也存在类似情况。

有研究发现,CRC 患者的 CLCA1 血清水平较健康对照者明显降低,并且 CLCA1 血清水平和 CLCA1mRNA 表达水平与 CRC 的转移和分期呈负相关,还发现 CLCA1 可能通过下调 Wnt/ β -catenin 信号通路和上皮细胞-间质转化(EMT)发挥抑癌作用^[12]。异常的 Wnt 信号可以触发 EMT 过程,这在癌症转移中至关重要。CLCA1 表达的增加降低了 β -catenin 的表达水平,抑制了 EMT 从而抑制肿瘤的活性,这与一些文章发现的 Wnt/ β -catenin 激活和炎症性肠病是导致 CRC 的两个主要因素的结论不谋而合,这两个因素能导致肠道内肠上皮内环境稳态的破坏,如细胞增殖增加、细胞分化和凋亡减少等^[13-15]。同时,EMT 过程使上皮细胞转化为间质细胞,使细胞失去黏附和极性,从而发生迁移和侵袭^[16]。

对于 CLCA4,近期也有研究指出,CLCA4 在细胞中发挥生物学功能是通过使 PI3K/AKT 信号失活和调节 EMT 介导的,并且 CLCA4 的降低与 CRC 患者的淋巴结转移有关^[17],这与 LIU 等^[18]发现的 CLCA4 抑制肝癌细胞的增殖和侵袭的机制相似。

综上所述,应用生物信息学分析方法,本研究筛选出了 CLCA1 和 CLCA4 两个基因,对其进行更进一步的作用机制讨论,发现它们严重影响 CRC 的转移,且最终影响 CRC 患者的预后。本研究丰富了 CLCA1 和 CLCA4 在 CRC 中的认识,二者有可能可以作为 CRC 的诊断依据和治疗靶点。

参考文献

[1] KANNAN T,GAZALA K. Colorectal cancer and nutrition [J]. *Nutrients*,2019,11(1):164.

[2] 杜灵彬,李辉章,王悠清,等. 2013 年中国结直肠癌发病与死亡分析[J]. *中华肿瘤杂志*,2017,39(9):701-706.

[3] 吴春晓,顾凯,龚杨明,等. 2015 年中国结直肠癌发病和死亡情况分析[J]. *中国癌症杂志*,2020,30(4):241-245.

[4] VOGELSTEIN B,PAPADOPOULOS N,VELCULES-CU V E, et al. Cancer genome landscapes [J]. *Science*, 2013,339(6127):1546-1558.

[5] MEHRKHANI F,NASIRI S,DONBOLI K, et al. Prognostic factors in survival of colorectal cancer patients after surgery[J]. *Color Dis*,2009,11(2):157-161.

[6] QU A,YANG Y,ZHANG X, et al. Development of a pre-operative prediction nomogram for lymph node metastasis in colorectal cancer based on a novel serum miRNA signature and CT scans[J]. *EBio Medicine*,2018,37:125-133.

[7] 熊启敏,严文跃,胡静,等. 原发性年轻结直肠癌患者的病理特征及预后分析[J]. *现代肿瘤医学*,2018,26(6):888-892.

[8] SZABO I,LEPPLE-WIENHUES A,KABA K N, et al. Tyrosine kinase-dependent activation of a chloride channel in CD95-induced apoptosis in T lymphocytes [J]. *Proc Natl Acad Sci USA*,1998,95(11):6169-6174.

[9] WOHLRAB D,MARKWARDT F. Influence of ion channel blockers on proliferation and free intracellular Ca^{2+} concentration of human keratinocytes[J]. *Skin Pharmacol Appl Ski Physiol*,1999,12(5):257-265.

[10] KIM J A,KANG Y S,LEE Y S. Role of Ca^{2+} activated Cl^{-} channels in the mechanism of apoptosis induced by cyclosporin A in a human hepatoma cell line[J]. *Biochem Biophys Res Commun*,2003,309(2):291-297.

[11] HOU T,ZHOU L,WANG L, et al. CLCA4 inhibits bladder cancer cell proliferation, migration, and invasion by suppressing the PI3K/AKT pathway [J]. *Oncotarget*, 2017,8(54):93001-93013.

[12] LI X F,HU W X,ZHOU J J, et al. CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway [J]. *Cell Commun Signal*,2017,15(1):38-45.

[13] NELSON W J,NUSSE R. Convergence of Wnt, beta-catenin, and cadherin pathways [J]. *Science*,2004,303(5663):1483-1487.

[14] SEGDISAS S,TOMLINSON I. Colorectal cancer and genetic alterations in the Wnt pathway [J]. *Oncogene*, 2006,25(57):7531-7537.

[15] GRIVENNIKOV S I. Inflammation and colorectal cancer: colitis-associated neoplasia [J]. *Semin Immunopathol*, 2013,35(2):229-244.

[16] NIETO M A,HUANG R Y,JACKSON R A, et al. EMT:2016 [J]. *Cell*,2016,166(1):21-45.

[17] CHEN H,LIU Y,JIANG C J, et al. Calcium-activated chloride channel A4 (CLCA4) plays inhibitory roles in invasion and migration through suppressing epithelial-mesenchymal transition via PI3K/AKT signaling in colorectal cancer [J]. *Med Sci Monit*,2019,25:4176-4185.

[18] LIU Z,CHEN M,XIE L K, et al. CLCA4 inhibits cell proliferation and invasion of hepatocellular carcinoma by suppressing epithelial-mesenchymal transition via PI3K/AKT signaling [J]. *Aging (Albany NY)*,2018,10(10):2570-2584.

(收稿日期:2020-08-28 修回日期:2021-01-21)

(上接第 1183 页)

[7] PENG J,LU Y F,YU H, et al. Analysis of 2 reverse syphilis testing algorithms in diagnosis of syphilis: a Large-Cohort prospective study [J]. *Clin Infect Dis*,2018, 67(6):947-953.

[8] 庄亦晖,龚燕,许晓峰,等. 梅毒螺旋体特异性抗体检测方法在血清学诊断中的应用评估 [J]. *检验医学*,2018,33(7):633-636.

[9] SANFILIPPO A M,FREEMAN K,SCHMITZ J L. Comparison of manual and fully automated AIX1000 rapid plasma reagin assays for laboratory diagnosis of syphilis [J]. *J Clin Microbiol*,2018,56(8):e00214-e00218.

[10] GAMBIEZ C,GARET C,CHOQUET M, et al. Limited efficacy of nontreponemal test combined with treponemal test in the initial syphilis screening algorithm [J]. *J Med*

Microbiol,2019,68(2):216-220.

[11] PARK I U,FAKILE Y F,CHOW J M, et al. Performance of treponemal tests for the diagnosis of syphilis [J]. *Clin Infect Dis*,2019,68(6):913-918.

[12] 张华荣,陈正明,文海燕,等. 免疫印迹法(WB)对隐性梅毒抗体确证检测的价值探讨 [J]. *口岸卫生控制*,2013,18(2):20-23.

[13] VAN DOMMELEN L,HOEBE C J,VAN TIEL F H, et al. Confirmation of high specificity of an automated enzyme immunoassay test for serological diagnosis of syphilis: retrospective evaluation versus results after implementation [J]. *Sex Transm Dis*,2015,42(3):120-122.

(收稿日期:2020-08-24 修回日期:2020-12-29)