

基于 Oncomine 和 GEO 数据库分析 S100P 在胰腺癌中的表达及临床意义

李丹^{1a}, 余涛², 曾智^{1b}, 吴杰^{1a}

(1. 武汉大学人民医院 a. 药学部, b. 病理科, 湖北 武汉 430060; 2. 华中科技大学同济医学院附属武汉中心医院中西医结合肿瘤科, 湖北 武汉 430014)

摘要:目的 阐明 S100P 蛋白在胰腺癌中的表达及临床意义。方法 检索 Oncomine 和 GEO 数据库中有关 S100P 的信息, 并对所获取的数据资料挖掘并进行二次分析, 对 S100P 在胰腺癌中的作用进行分析。结果 Oncomine 数据库中共收集了 438 项不同类型 S100P 的研究结果, 关于 S100P 表达差异有统计学意义的结果有 63 项, 其中 S100P 表达增高的有 38 项, 表达降低的有 25 项。胰腺癌中高表达的研究有 5 项、低表达的有 0 项。共有 11 项研究涉及 S100P 在胰腺癌组织和正常组织中的表达, 包括 387 例样本。在数据库中综合比较 11 项研究成果, 发现与对照组相比, S100P 在胰腺癌中的表达量高于正常组织 ($P < 0.05$)。通过分析 GEO 数据库中的数据集 GSE28735, 自身对照发现 S100P 在胰腺癌组织中的转录水平显著高于其癌旁组织 ($P < 0.05$)。不仅如此, 通过挖掘 GEO 数据库, 分析数据集 GSE28735 和 GSE57495, 发现高表达 S100P 的患者总体病死率较高, 低表达 S100P 的患者预后较好 ($P < 0.05$)。结论 通过深入挖掘 Oncomine 和 GEO 基因芯片数据库中肿瘤相关的基因信息, 提示 S100P mRNA 在胰腺癌组织中呈现高表达, 并与胰腺癌预后相关, 其有望成为抗胰腺癌药物的重要治疗靶点。

关键词: S100P 蛋白; 胰腺癌; Oncomine 数据库; GEO 数据库

doi: 10.3969/j.issn.1009-6469.2017.12.011

Expression of S100P and its prognostic role in pancreatic cancer: An analysis based on the data-mining of Oncomine and GEO database

LI Dan^{1a}, YU Tao², ZENG Zhi^{1b}, WU Jie^{1a}

(a. Department of Pharmacy; b. Department of Pathology, 1. Renmin Hospital of Wuhan University, Wuhan, Hubei 430060, China; 2. Oncology Department, of Integrated Traditional Chinese and Western Medicine, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, Hubei 430014, China)

基金项目: 国家自然科学基金青年项目 (81602535, 81704023)

[2] HASHIMOTO N, TOMINAGA T, MIYAMOTO S, et al. Guidelines for diagnosis and treatment of moyamoya disease (spontaneous occlusion of the circle of Willis) [J]. *Neurol Med Chir (Tokyo)*, 2012, 52(5): 245-266.

[3] 李德生, 戚鹏, 杨伟中, 等. 单侧烟雾病的临床特征[J]. *中国卒中杂志*, 2013, 8(7): 511-516.

[4] 陈光军, 李淮玉, 王国平, 等. 烟雾病的临床和影像学特点分析[J]. *安徽医药*, 2008, 12(6): 514-515.

[5] BABA T, HOUKIN K, KURODA K. Novel epidemiological features of moyamoya disease [J]. *J Neurol Neurosurg Psychiatr*, 2008, 79(8): 900-904.

[6] CEVIK B, ACU B, AKSOY D, et al. Protein S deficiency and an adult case with moyamoya syndrome that presented with primary intraventricular haemorrhage [J]. *Balkan Med J*, 2014, 31(2): 180-183.

[7] 白亮, 刘新梅, 李俊, 等. 出血型烟雾病临床诊治分析[J]. *临床急诊杂志*, 2012, 13(2): 130-131.

[8] 高山, 倪俊, 黄家星, 等. 烟雾病临床特点分析[J]. *中华神经科杂志*, 2006, 39(3): 176-179.

[9] 王金, 徐小林, 王东, 等. 成人烟雾病缺血及出血性事件的危险因素[J]. *山东医药*, 2013, 53(39): 7-11.

[10] FLINT AC, ROEBKEN A, SINGH V, et al. Primary intraventricular hemorrhage: yield of diagnostic angiography and clinical outcome [J]. *Neurocrit Care*, 2008, 8(3): 330-336.

[11] TAKAHASHI JC, MIYAMOTO S. Moyamoya disease: recent progress and outlook [J]. *Neurol Med Chir (Tokyo)*, 2010, 50(9): 824-832.

[12] WANG H, WU H, HAN C, et al. Cortical microvascularization in moyamoya disease: Characteristics and the relations with surgical outcomes of encephaloduroarteriosynangiosis [J]. *CNS Neurosci Ther*, 2016, 22(4): 325-327.

[13] 张瑛, 戴启荷. 出血性烟雾病合并动脉瘤患者的治疗方法及预后分析[J]. *中国实用神经疾病杂志*, 2016, 19(22): 33-34.

[14] 李丽, 朱成铜, 郭守文. 烟雾病磁共振血管成像诊断及影像表现研究[J]. *山西医药杂志*, 2015, 44(24): 2876-2878.

[15] SUGINO T, MIKAMI T, OHTAKI S, et al. Assessment of moyamoya disease using multidetector row computed tomography [J]. *J Stroke Cerebrovasc Dis*, 2013, 22(5): 644-649.

[16] HAN C, LI ML, XU YY, et al. Adult moyamoya-atherosclerosis syndrome: Clinical and vessel wall imaging features [J]. *J Neurol Sci*, 2016, 369: 181-184.

[17] 王强, 牛玉军. CTA 与 DSA 评价烟雾病血管程度的对比研究[J]. *中国临床医学影像学杂志*, 2015, 26(2): 128-130.

(收稿日期: 2017-03-05, 修回日期: 2017-06-26)

Abstract: Objective To explore the expression of S100P and its clinical significance in pancreatic cancer. **Methods** Data about S100P were retrieved from the Oncomine and GEO database, then the role of S100P in pancreatic cancer was mined and analyzed. **Results** Totally, 438 different types of cancer studies about the expression of S100P were identified in the Oncomine database. The S100P expression was statistically significant in 63 of the studies, over-expressed in 38 (5 in pancreatic cancer) and under-expressed in the other 25 (0 in pancreatic cancer). Eleven of the studies with 387 samples covered the expression of S100P in pancreatic cancer and normal tissues. The expression of S100P was significantly higher in the cancer than in the normal tissues ($P < 0.05$). S100P mRNA expression level was also higher in pancreatic cancer tissues than in the adjacent normal tissues by analyzing the GSE28735 of GEO database ($P < 0.05$). Furthermore, GSE28735 and GSE57495 datasets of GOE showed a higher total mortality rate in the patients with higher S100P expression than in those with a lower expression ($P < 0.05$). **Conclusions** S100P is highly expressed in pancreatic cancer, and its expression is associated with the prognosis of pancreatic cancer, which may be clinically used as an important target of medication for pancreatic cancer.

Key words: S100P protein; Pancreatic cancer; Oncomine database; GEO database

胰腺癌发病约占全部恶性肿瘤的2%~3%,临床早期诊断及治疗难度大,预后极差,5年存活率<5%。由于绝大多数患者确诊时已属晚期,丧失了根治性手术的机会,而放化疗的效果不佳,约95%患者在确诊后6个月内死亡^[1]。因此,有必要在深入研究胰腺癌发生发展分子机制的同时探索有助于胰腺癌早期诊断的分子标志物。有研究显示, S100蛋白家族大部分成员与肿瘤的发生发展关系密切,其家族成员之一的S100P蛋白广泛分布在人心、脑、肺、肝脏、骨髓等正常组织,表达后定位在胞膜、胞质及胞核^[2],在细胞内外具有多种重要功能,如调节蛋白质磷酸化、酶活性、钙稳态、细胞周期、生长、存活增殖等。多项研究显示S100P的表达可能与胰腺癌的发生、发展有关^[3,4],但由于研究方法、样本量和人群的差异,其结论可靠性较低。

Oncomine数据库是当前世界上最大的肿瘤基因芯片数据库和整合数据提取平台,其中整合的文献及芯片数据以高质量得到学界的广泛认可。而GEO数据库是美国生物技术信息中心的基因数据库,所收录的基因芯片数据包括但不限于肿瘤。我们通过深度挖掘Oncomine数据库中收录的S100P在胰腺癌中的表达数据,并利用GEO数据库分析S100P基因与胰腺癌患者生存预后之间的关系,借此了解S100P在胰腺癌发生发展及预后中的作用,以期胰腺癌的早期诊断和治疗提供理论依据。

1 资料与方法

1.1 从 Oncomine 和 GEO 数据库中提取数据 首先用邮箱注册 Oncomine 数据库,获得使用数据库的权限,然后根据自己的需求在数据库中设定筛选和提取数据的条件。本研究中,我们在数据库检索界面输入与本研究相关的筛选条件:(1) cancer type: pancreatic cancer; (2) gene: S100P; (3) data type: mRNA 或 DNA; (4) analysis type: cancer vs normal a-

nalysiss。在 PubMed 主界面 <https://www.ncbi.nlm.nih.gov/pubmed/> 搜索栏中找到 GEO datasets, 根据自己筛选条件在对应检索栏输入“pancreatic cancer”“normal”“survival”来查找有生存分析的数据集,挖掘其中与 S100P 有关的胰腺癌患者的生存数据。

1.2 统计学方法 统计学分析处理采用 SPSS 18.0 软件,作图采用 Graphpad Prism 5.0 软件。采用 t 检验来比较正常组织与胰腺癌病例组织之间 S100P 表达的差异。采用 Kaplan-Meier 方法计算 S100P 表达与胰腺癌预后的关系。当 $P \leq 0.05$ 时认为差异有统计学意义。

2 结果

2.1 S100P 在不同类型肿瘤中的表达情况 共搜集了 Oncomine 数据库中有关 S100P 蛋白的不同类型研究结果 438 项,涉及到肿瘤病例与对照组表达的比较、不同肿瘤类型之间的比较、肿瘤不同病理类型之间的比较等。关于 S100P 表达差异有统计学意义的结果有 63 项研究,其中 S100P 表达增高的有 38 项,表达降低的有 25 项。在胰腺癌中高表达的研究有 5 项、低表达的研究有 0 项,具体数据见表 1。

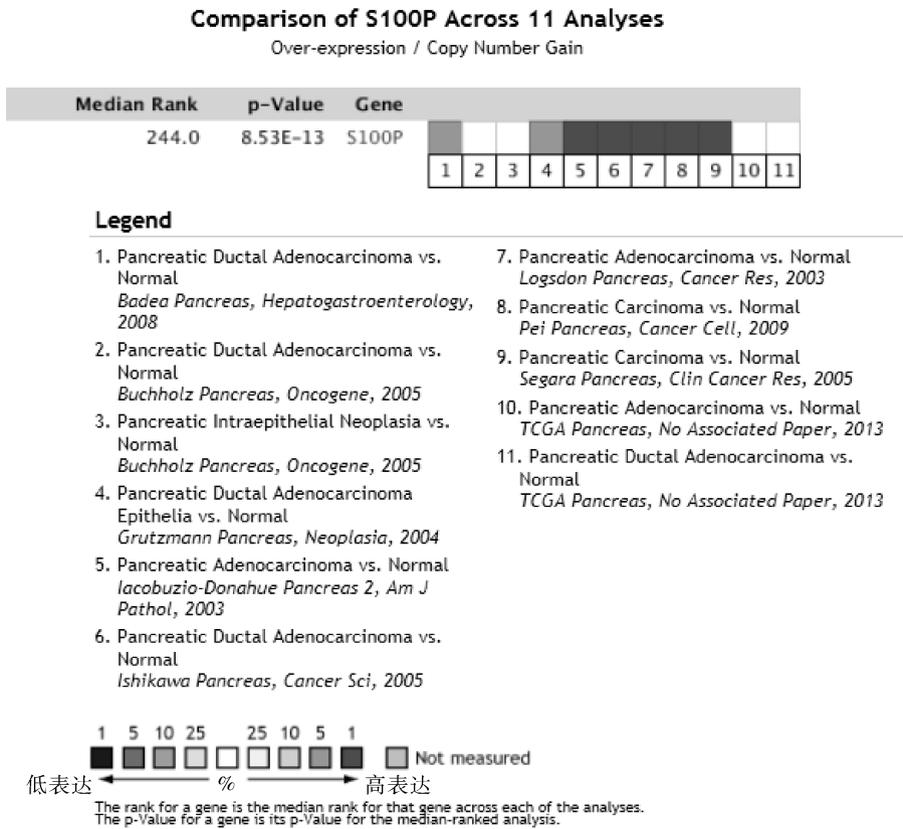
表 1 S100P 在 Oncomine 数据库中所有肿瘤相关研究中的表达/项

肿瘤类型	表达上调的研究	表达下调的研究
膀胱癌	1	0
乳腺癌	13	0
宫颈癌	1	0
结肠癌	11	1
食管癌	1	2
胃癌	0	4
头颈癌	1	1
白血病	0	9
肝癌	1	0
肺癌	3	0
淋巴瘤	0	3
黑色素瘤	0	1
卵巢癌	1	0
胰腺癌	5	0
肉瘤	0	4
合计	38	25

2.2 S100P 在胰腺癌中的表达 利用 Oncomine 数据库提取信息,我们分析发现,2003 年起共有 11 项研究涉及 S100P 在胰腺癌和正常组织中的表达,共包括 387 例样本。文章分别发表于 *American Journal of Pathology*^[5], *Cancer Research*^[6], *Cancer Cell*^[7], *Clinical Cancer Research*^[8], *Hepatogastroenterology*^[9], *Cancer Science*^[10], *Neoplasia*^[11], *Oncogene*^[12]。在 Oncomine 数据库中综合比较 11 项研究结果,发现与对照组相比,S100P 在胰腺癌中高表达($P < 0.05$),见图 1。

2.3 S100P 在不同胰腺癌研究芯片中癌组织和正

常组织之间的表达差异 Oncomine 数据库中 S100P 在不同胰腺癌研究芯片中的表达结果显示:在 Lacobuzio-Donahue^[5]、Logsdon^[6]、Pei^[7]、Segara^[8]、Badea^[9]、Ishikawa^[10]、Grutzmann^[11] 等 7 项研究中,S100P 在胰腺癌中的表达量显著高于正常组织($P < 0.05$)。表 2 为在 7 项不同研究中对应的 S100P 过表达倍数和对应的 P 值。进而,我们通过挖掘分析数据集 GSE28735 中 45 例胰腺癌患者癌与癌旁的 S100P 基因表达的丰度值,发现 S100P 在同一患者胰腺癌组织中的转录水平显著高于其癌旁组织($P < 0.001$),见图 2。



注:1~11 分别表示 11 项研究结果,蓝色代表低表达,红色代表高表达,颜色由浅变深(即从正中间%往两边各自的箭头方向)代表表达差异越大,红色越深表示 S100P 基因在该芯片中表达越高。

图 1 Oncomine 数据库中关于 S100P 在胰腺癌中的表达

表 2 S100P 蛋白在不同研究中的过表达情况

第一作者	正常胰腺组织的病例数	胰腺癌组织病例数	过表达倍数	P 值	发表杂志
Lacobuzio-Donahue ^[5]	5	11	24.050	2.13×10^{-8}	Am J Pathol
Logsdon ^[6]	5	10	20.309	4.88×10^{-10}	Cancer Res
Pei ^[7]	16	36	77.931	1.61×10^{-12}	Cancer Cell
Segara ^[8]	6	11	17.727	1.23×10^{-5}	Clin Cancer Res
Badea ^[9]	39	39	13.177	8.53×10^{-13}	Hepatogastroenterology
Ishikawa ^[10]	25	24	4.247	3.00×10^{-3}	Cancer Sci
Grutzmann ^[11]	11	11	8.382	8.00×10^{-3}	Neoplasia

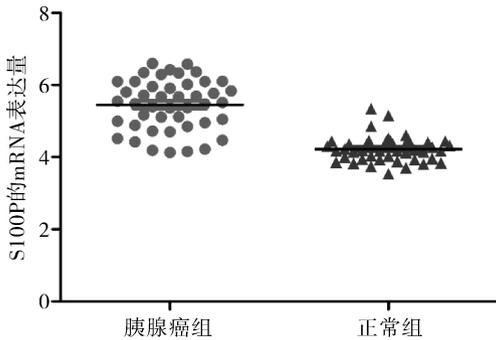


图2 S100P蛋白在胰腺癌与癌旁组织中的表达

2.4 S100P与胰腺癌预后的关系 通过分析上述数据库的信息发现, S100P在胰腺癌组织中表达增高, 为了进一步明确S100P表达与胰腺癌预后之间的关系, 我们在GEO数据库中挖掘提取S100P基因表达丰度值与胰腺癌预后相关的数据GSE57495, 利用Kaplan-Meier分析发现, S100P的表达水平与胰腺癌患者生存预后之间存在相关性, 高表达S100P的患者总体死亡率较高, 低表达S100P的患者预后较好($P < 0.05$), 见图3。

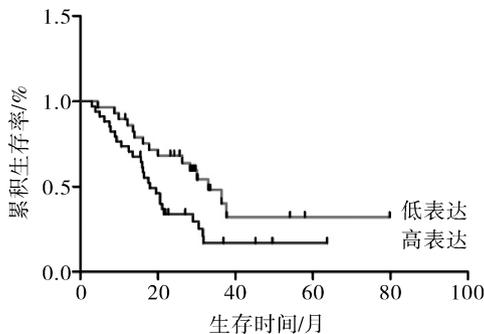


图3 数据集GSE57495中S100P表达与胰腺癌预后之间的生存曲线

3 讨论

胰腺癌近年来在我国以其高发病率, 极差的治疗效果和预后等特点, 严重危害患者健康。由于临床上缺乏胰腺癌特异性的分子靶标, 不利于胰腺癌的早期诊断、治疗方法的选择及治疗效果。这迫切要求从原发性肿瘤中寻找一些表达差异较大并有临床意义的基因来探讨肿瘤的发生、发展机制, 为临床诊断, 预后判断, 进而寻求新的基因治疗靶点提供理论依据。

S100P是S100蛋白家族中研究较少的钙结合蛋白之一, 该蛋白由95个氨基酸残基组成, 全长共510个核苷酸, 分子量为10.4 kDa, 该基因定位于4p16, 有2个外显子和1个内含子。S100P蛋白也具有两个EF手型结构, 包括氨基端和羧基端, 氨基

端含14个氨基酸, 形成环状结构, 与 Ca^{2+} 亲和性低。而羧基端由12个氨基酸组成, 与 Ca^{2+} 具有较高亲和性。S100P通过结合 Ca^{2+}/Mg^{2+} 后, 蛋白构象发生改变, 并打开暴露出疏水区域, 然后与细胞膜结合, 与相应的靶蛋白作用后发挥生物学效应^[2-3]。由于S100P在多种肿瘤组织中呈现出明显的高表达, 并与部分肿瘤的分期预后存在一定的相关性, 因此其在多种肿瘤的发生发展过程中扮演了一定的作用并对临床有积极的指导意义^[3]。

临床及基础实验研究表明S100P基因水平在胰腺癌组织中的表达高于正常对照组织和胰腺炎组织: Deng等^[13]研究表明胰腺导管腺癌细胞中S100P均呈阳性表达, 而在良性细胞中多呈阴性表达, 提示S100P可作为胰腺癌良恶性肿瘤的敏感且特异的鉴别标志物。Downen等^[14]研究表明, S100P在早期胰腺癌中就有表达异常, 其中92%的胰腺导管腺癌S100P呈现高表达, 且其升高的水平与胰腺上皮癌变存在相关性, 并有助于胰腺的原位癌发展为有侵袭性的胰腺导管腺癌。Arumugam等^[15]通过体内外研究证实S100P的表达水平与胰腺癌细胞的增殖、存活、迁移和侵袭相关。

近期国际上很多学者针对S100P靶点探索了检测S100P蛋白的方法, 并尝试开发针对S100P蛋白的药物。Dakhel等^[16]开发了阻断S100P功能的anti-S100P单克隆抗体, 其除了能抑制肿瘤细胞的体外活性, 也能降低皮下移植瘤模型的肿瘤生长和肝转移, 并提出S100P特异性的单克隆抗体有望作为单药或者联合化疗药使用而开发成为治疗胰腺癌的药物。Ezzat等^[17]通过细针穿刺细胞学评价S100P和IMP3蛋白是否可以用来鉴别胰腺导管腺癌和非肿瘤胰腺组织。通过回顾性研究, 纳入72例胰腺肿块, 在内镜下引导细针穿刺细胞, 最后用组织病理学作为诊断金标准, 发现S100P作为诊断胰腺癌具有96.4%的敏感性, 93.3%特异性, 其阳性预测率和阴性预测率分别为98.2%和87.5%, 总的准确率为95.8%。Chiba等^[18]在内镜超声引导下细针穿刺标本, 建立了一种可靠的S100P蛋白含量的分析方法, 并对27例可疑的胰腺癌患者进行分析, 发现胰腺癌患者S100P蛋白水平明显高于胰腺炎患者, 为胰腺癌的早期诊断提供了一种新的定量分析方法。

尽管多数研究^[5,7,10,14-15]发现S100P在胰腺癌中高表达, 但有些研究只是细胞层面, 未涉及到人体组织样本, 同样有些研究涉及的样本量小, 易产生抽样误差等, 而导致结论的可信度不高。目前为

止一直缺乏强有力的证据。而 Oncomine 数据库包括了目前世界上最大的 TCGA 数据库,且 GEO 数据库作为全球最大的基因芯片数据库之一也存储了高通量的功能基因组数据,每年的基因表达数据量呈指数增长。因此,本次研究包含了目前为止最大的样本量,并且所有数据均来自基因芯片,研究方法一致,因此极大地增加了结论的可信度。本次研究综合两个大数据库的数据挖掘信息提出了 S100P 在胰腺癌组织中高表达,且与胰腺癌预后有关,将为胰腺癌肿瘤药物的开发提供重要的理论依据和临床指导。

参考文献

[1] SIEGEL RL, MILLER KD, JEMAL A. Cancer statistics, 2016 [J]. *CA Cancer J Clin*, 2016, 66(1):7-30.

[2] JIN G, WANG S, HU X, et al. Characterization of the tissue-specific expression of the s100P gene which encodes an EF-hand Ca²⁺-binding protein [J]. *Mol Biol Rep*, 2003, 30(4):243-248.

[3] PARKKILA S, PAN PW, WARD A, et al. The calcium-binding protein S100P in normal and malignant human tissues [J]. *BMC Clin Pathol*, 2008, 8:2.

[4] JIANG Y, LIU M, LI Z, et al. Discovery of novel candidate oncogenes in pancreatic carcinoma using high-throughput microarrays [J]. *Hepato-gastroenterology*, 2013, 60(128):1825-1832.

[5] LACOBUZIO-DONAHUE CA, MAITRA A, OLSEN M, et al. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays [J]. *Am J Pathol*, 2003, 162(4):1151-1162.

[6] LOGSDON CD, SIMEONE DM, BINKLEY C, et al. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer [J]. *Cancer Res*, 2003, 63(10):2649-2657.

[7] PEI H, LI L, FRIDLEY BL, et al. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt [J]. *Cancer Cell*, 2009, 16(3):259-266.

[8] SEGARA D, BIANKIN AV, KENCH JG, et al. Expression of HOXB2, a

retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia [J]. *Clin Cancer Res*, 2005, 11(9):3587-3596.

[9] BADEA L, HERLEA V, DIMA SO, et al. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia [J]. *Hepatogastroenterology*, 2008, 55(88):2016-2027.

[10] ISHIKAWA M, YOSHIDA K, YAMASHITA Y, et al. Experimental trial for diagnosis of pancreatic ductal carcinoma based on gene expression profiles of pancreatic ductal cells [J]. *Cancer Sci*, 2005, 96(7):387-393.

[11] GRUTZMANN R, PILARSKY C, AMMERPOHL O, et al. Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays [J]. *Neoplasia*, 2004, 6(5):611-622.

[12] BUCHHOLZ M, BRAUN M, HEIDENBLUT A, et al. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions [J]. *Oncogene*, 2005, 24(44):6626-6636.

[13] DENG H, SHI J, WILKERSON M, et al. Usefulness of S100P in diagnosis of adenocarcinoma of pancreas on fine-needle aspiration biopsy specimens [J]. *Am J Clin Pathol*, 2008, 129(1):81-88.

[14] DOWEN SE, CRNOGORAC-JURCEVIC T, GANGESWARAN R, et al. Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer [J]. *Am J Pathol*, 2005, 166(1):81-92.

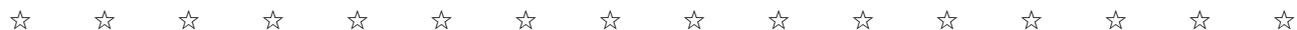
[15] ARUMUGAM T, SIMEONE D M, VAN GOLEN K, et al. S100P promotes pancreatic cancer growth, survival, and invasion [J]. *Clin Cancer Res*, 2005, 11(15):5356-5364.

[16] DAKHEL S, PADILLA L, ADAN J, et al. S100P antibody-mediated therapy as a new promising strategy for the treatment of pancreatic cancer [J]. *Oncogenesis*, 2014, 3:e92.

[17] EZZAT NE, TAHOUN NS, ISMAIL YM. The role of S100P and IMP3 in the cytologic diagnosis of pancreatic adenocarcinoma [J]. *J Egypt Natl Canc Inst*, 2016, 28(4):229-234.

[18] CHIBA M, IMAZU H, KATO M, et al. Novel quantitative analysis of the S100P protein combined with endoscopic ultrasound-guided fine needle aspiration cytology in the diagnosis of pancreatic adenocarcinoma [J]. *Oncol Rep*, 2017, 34(4):1943-1952.

(收稿日期:2017-03-23, 修回日期:2017-05-09)



《安徽医药》要求来稿中应规范统计结果的解释和表达

当 $P < 0.05$ (或 $P < 0.01$) 时,应说明对比组之间的差异有统计学意义,而不应描述为对比组之间具有显著性(或非常显著性)差别;应写明所用统计分析方法的具体名称(如:成组设计资料的 t 检验、两因素析因设计资料的方差分析、多个均数之间两两比较的 q 检验等),统计量的具体值(如 $t = 2.26, \chi^2 = 4.48, F = 8.89$ 等),应尽可能给出具体 P 值(如 $P = 0.032$);当涉及到总体参数(如总体均数、总体率等)时,在给出显著性检验结果的同时,再给出 95% 可信区间。