

联合多种高通量表达谱数据库挖掘胃癌S100基因的差异表达

刘骥, 李雪, 李纪鹏, 王为忠

■背景资料

伴随人类基因组计划的完成, 生物信息学技术、计算机技术的进步, 在线的、免费的基因数据库的日益剧增, 为研究者进行鉴定和发现疾病中关键基因或有表达差异的基因, 提供了很好的界面和挖掘工具, 有力地推动了肿瘤相关基因的研究。

刘骥, 李雪, 中国人民解放军空军杭州航空医学鉴定训练中心 浙江省杭州市 310013

李纪鹏, 王为忠, 中国人民解放军第四军医大学西京消化病医院消化三科 陕西省西安市 710032

刘骥, 主治医师, 医学博士, 主要从事胃癌的基础研究。

作者贡献分布: 此课题由刘骥与王为忠设计; 刘骥与李纪鹏负责数据挖掘与生物信息学分析; 刘骥与李雪负责统计学分析; 刘骥、李雪、李纪鹏及王为忠参与论文写作。

通讯作者: 刘骥, 主治医师, 310013, 浙江省杭州市西湖区杨公堤15号, 中国人民解放军空军杭州航空医学鉴定训练中心。
ljlonging@foxmail.com

电话: 0571-87349040

收稿日期: 2015-02-25 修回日期: 2015-03-30

接受日期: 2015-04-01 在线出版日期: 2015-05-18

In silico analysis of S100 gene expression in gastric cancer

Ji Liu, Xue Li, Ji-Peng Li, Wei-Zhong Wang

Ji Liu, Xue Li, Hangzhou Medical Identification and Training Centre of Air Force, Hangzhou 310013, Zhejiang Province, China

Ji-Peng Li, Wei-Zhong Wang, Department of Gastroenterology III, Xijing Hospital of the Fourth Military Medical University, Xi'an 710032, Shaanxi Province, China

Correspondence to: Ji Liu, Attending Physician, Hangzhou Medical Identification and Training Centre of Air Force, 15 Yanggongdi, Xihu District, Hangzhou 310013, Zhejiang Province, China. ljlonging@foxmail.com

Received: 2015-02-25 Revised: 2015-03-30

Accepted: 2015-04-01 Published online: 2015-05-18

Abstract

AIM: To explore the feasibility and strategy of in silico identification of human gastric cancer-related S100 genes.

METHODS: By combining series analysis

of gene expression, virtual Northern blot and microarray data, the expression levels of S100 family members in normal and malignant stomach tissues were systematically investigated through CGAP and GEO.

RESULTS: At least 5 S100 genes were found to be upregulated in gastric cancer by in silico analysis. Among them, four genes, including S100A2, S100A4, S100A7 and S100A10, were reported to be overexpressed in gastric cancer.

CONCLUSION: To our knowledge this is the first report of systematic evaluation of S100 gene expression in gastric cancer by in silico analysis. The results indicated that overexpression of S100 gene family members is a characteristic of gastric cancer. Reasonable use of public databases by the internet-available tools is a simple, effective approach to identify cancer-related genes, and might provide useful clues to further investigation although the results require experimental validation.

© 2015 Baishideng Publishing Group Inc. All rights reserved.

Key Words: Gastric cancer; S100; Bioinformatics; Serial analysis of gene expression; Expressed sequence tag; Microarray; Gene Expression Omnibus; Cancer Genome Anatomy Project

Liu J, Li X, Li JP, Wang WZ. In silico analysis of S100 gene expression in gastric cancer. *Shijie Huaren Xiaohua Zazhi* 2015; 23(14): 2208-2214 URL: <http://www.wjgnet.com/1009-3079/23/2208.asp> DOI: <http://dx.doi.org/10.11569/wcjd.v23.i14.2208>

■同行评议者

潘阳林, 副主任医师, 副教授, 中国人民解放军第四军医大学西京医院消化病医院消化六科

摘要

目的: 探讨利用生物信息学方法筛选胃癌组织与正常胃组织差异表达基因的可行性及具体方法, 寻找胃癌相关*S100*基因。

方法: 联合SAGE分析、虚拟电子杂交和虚拟微阵列, 挖掘癌基因组解剖计划和基因表达综合数据库资源中关于胃癌和正常胃组织差异表达*S100*基因。

结果: 5个*S100*基因在胃癌组织中上调表达, 包括*S100A2*、*S100A3*、*S100A4*、*S100A7*和*S100A10*。*S100A3*是新的胃癌过表达基因。

结论: 首次探讨性利用生物信息学方法系统分析胃癌*S100*基因表达。多个*S100*家族成员在胃癌中上调表达。联合多数据库挖掘肿瘤相关基因是一个可靠的方法, 给进一步生物学实验以大量的提示。

© 2015年版权归百世登出版集团有限公司所有。

关键词: 胃癌; *S100*; 生物信息学; 基因表达系列分析; 表达系列标签; 微阵列; 基因表达综合数据库; 癌症基因组解剖计划

核心提示: 本课题联合SAGE分析、虚拟电子杂交和虚拟微阵列分析, 系统分析了*S100*基因家族在胃癌组织中的表达情况, 挖掘了5个*S100*基因在3个in silico分析方法中均显示在胃癌组织中上调表达, 分别是*S100A2*、*S100A3*、*S100A4*、*S100A7*和*S100A10*。在这5个*S100*基因中, 只有*S100A3*第1次报道是胃癌相关基因, 相对正常胃组织, *S100A3*在胃癌中上调表达。

刘骥, 李雪, 李纪鹏, 王为忠. 联合多种高通量表达谱数据库挖掘胃癌*S100*基因的差异表达. 世界华人消化杂志. 2015; 23(14): 2208-2214 URL: <http://www.wjgnet.com/1009-3079/23/2208.asp> DOI: <http://dx.doi.org/10.11569/wjcd.v23.i14.2208>

0 引言

基因表达高通量研究技术及生物信息学的发展, 有力地推动了肿瘤相关基因的研究。互联网公共生物信息学数据库为研究提供了平台, 其庞大的核酸与蛋白质数据资源和成熟的数据库搜索工具为研究者提供了材料和方法。本研究利用生物信息学方法, 挖掘癌基因组解剖计划(Cancer Genome Anatomy Project, CGAP)和基因表达综合数据库(Gene Expression Omnibus, GEO)生物信息数据, 系统分析22个

*S100*家族成员在胃癌与正常胃组织的基因表达情况, 探讨利用生物信息学方法筛选胃癌与正常胃组织间的差异表达基因信息的可行性及具体方法。

1 材料和方法

1.1 材料 基因表达综合数据库(GEO, <http://www.ncbi.nlm.nih.gov/projects/geo/>): SAGE数据、微阵列数据; 癌基因组解剖计划(CGAP, <http://cgap.nci.nih.gov/>): SAGE数据、EST数据、基因发现工具(Gene Finder, <http://cgap.nci.nih.gov/Genes/GeneFinder>)、虚拟电子杂交(virtual Northern blot); 基因表达系列分析数据库(SAGE, <http://www.ncbi.nlm.nih.gov/SAGE/>); UniGene数据库(<http://www.ncbi.nlm.nih.gov/UniGene/>)。

1.2 方法

1.2.1 SAGE分析: SAGE标签的重复次数代表该转录体的表达水平, 结合生物信息学方法可以确定表达的基因种类和基因的表达丰度。对SAGE数据的分析主要包括从原始的序列中得到标签列表, 比较来胃癌组织和与正常胃组织的标签及其出现频率。截止2014-01, 收集于GEO网站的所有、可得的、关于胃癌和正常胃组织的SAGE数据用以挖掘和分析*S100*基因表达水平。获得GEO系列GSE545和GES14(分别由两个不同的实验室提供), 下载2个正常胃组织SAGE数据和8个胃癌SAGE数据(表1)。通过SAGEmap两个锚定酶(NlaIII和Sau3A)定位UniGene簇(UniGene cluster), 可以抽取、获得可靠的UniGene簇对应的20个*S100*基因标签(表2)。利用获得的*S100*基因标签, 对已下载SAGE数据进行综合分析, 从而确定各个*S100*基因在正常胃组织和胃癌的丰度值。相对正常胃组织, 在胃癌中基因差异表达大于3倍视为阳性结果。

1.2.2 虚拟电子杂交: MonochromaticSAGE/eDNAVirtualNorthern是CGAP中虚拟电子数据Northern杂交工具。通过Gene Finder, 查询*S100*基因, 获得基因信息(Gene Info), 再进行虚拟电子杂交, 可以分析该基因在EST数据库与SAGE数据库中癌组织与正常组织中的差异, 表3以*S100A2*为例列出虚拟电子杂交结果。设定*S100*基因在胃癌EST数据库与胃癌SAGE数据库中表达丰度均大于其在正常胃组织EST和

■ 研究前沿

既往的研究已经证实了一些*S100*家族成员在胃癌组织中有差异表达, 包括*S100A2*、*S100A4*、*S100A7*和*S100A10*。因此, 很有必要系统的分析和验证*S100*其他家族成员在胃癌和胃正常组织中的表达情况。为“硅片实验”的数据库挖掘法筛选肿瘤相关基因提供方法学的探讨, 也为阐明*S100*和胃癌的相关性提供重要理论依据。

■ 相关报道

利用生物信息学方法, 挖掘癌基因组解剖计划 (Cancer Genome Anatomy Project, CGAP) 和基因表达综合数据库 (Gene Expression Omnibus, GEO) 的生物信息资料数据, 研究者已经发现并鉴定了多个肿瘤相关基因或新的癌基因。

表 1 系列GSE545和GES14中关于胃癌和正常胃组织的样本GSM

GSM	Delineation
GSM9103	SAGE_Hiroshima_GC_P208T
GSM8505	SAGE_Hiroshima_GC_W246T
GSM8867	SAGE_Hiroshima_GC_W226T
GSM9104	SAGE_Hiroshima_GC_P208L
GSM7800	SAGE_Hiroshima_GC_S219T
GSM2385	SAGE_gastric_cancer-G189
GSM14760	SAGE_Stomach_cancer_B_X43
GSM757	SAGE_gastric_cancer-G234
GSM784	SAGE_normal_gastric_body_epithelial
GSM14780	SAGE_Stomach_normal_B_antrum

GSM: 数据集编号.

表 2 *S100*基因对应的SAGE标签和Unigene簇号

基因	基因数据库簇号	SAGE标签
<i>S100A2</i>	516484	GATCTCTTGG
<i>S100A3</i>	433168	TCTCCCACAC
<i>S100A4</i>	81256	ATGTGTAACG
<i>S100A6</i>	275243	CCCCCTGGAT
<i>S100A7</i>	112408	GAGCAGCGCC
<i>S100A8</i>	416073	TACCTGCAGA
<i>S100A9</i>	112405	GTGGCCACGG
<i>S100A10</i>	143873	AGCAGATCAG
<i>S100A12</i>	19413	GATTTTAA
<i>S100A16</i>	515714	AGCAGGAGCA

SAGE数据库3倍为阳性结果.

1.2.3 虚拟微阵列分析: 通过GEO检索获得5个关于胃癌和胃正常组织的微阵列数据集(表4), 选择并下载包含胃癌和正常胃组织样本含量较大的3个数据集: GSE2669、GSE2701和GSE3438. 利用获得的*S100*基因标签, 对3个数据集进行分析, 从而确定各个*S100*基因在正常胃组织和胃癌的丰度值. *S100*基因在3个数据集中两个显示高表达的视为阳性结果.

统计学处理 对于连续变量, 数据均用mean±SD表示. *S100*基因在胃癌和正常胃组织表达水平的微阵列数据差异采用*T*检验. 所有数据统计分析应用SPSS11.0软件(美国芝加哥), *P*<0.05为差异有统计学意义.

2 结果

2.1 SAGE分析结果 通过分析20个*S100*基因在GEO系列GSE545和系列GES14包含的2个正常胃组织SAGE数据和8个胃癌SAGE数据的表

达水平, 发现了10个*S100*基因在胃癌的表达水平差异表达达到3倍, 相对于正常胃组织(表5). 在这10个*S100*基因中只有*S100A6*和*S100A10*在正常胃组织中也有表达, 而二者在胃癌中的表达丰值属于最高, 分别为865.9 tpm和1890.5 tpm. 其他8个*S100*基因, *S100A2*、*S100A3*、*S100A4*、*S100A7*、*S100A8*、*S100A9*、*S100A12*和*S100A16*, 平均表达丰度值的范围是2.8-637.0 tmp, 他们在胃癌组织中均无表达. *S100A11*、*S100A14*和*S1009*在胃癌和正常胃组织中的表达水平差异无统计学意义.

2.2 虚拟电子杂交分析结果 通过在线虚拟电子杂交, 相对正常胃组织, 发现6个*S100*基因在胃癌中上调表达(表6). 在SAGE数据分析中阳性结果的*S100A6*、*S100A8*和*S100A16*, 通过电子虚拟杂交显示, 他们在胃癌和正常胃组织EST数据分析中则没有统计学意义. 在SAGE数据分析中在胃癌和正常胃组织中没有差异表达的*S10013*, 却在胃癌和正常胃组织EST数

表 3 *S100A2*在部分组织和对应癌组织中的EST和SAGE数据库中丰度值

人体组织	EST数据			SAGE数据		
	正常组织	癌组织	<i>P</i> 值	正常组织	癌组织	<i>P</i> 值
全部组织	58/3363544	83/2520094	0.00	10298/72145585	26873/81510018	nan
结肠	0/26958	3/153788	0.371	1/98089	57/643586	0.005
肝	0/99275	0/107967	—	0/66308	505/214987	0.000
皮肤	4/74261	6/125591	0.453	6154/26905847	25297/34522313	nan
胃	0/27005	3/66679	0.281	0/124767	63/448716	0.000

■创新亮点

联合利用多种高通量表达谱数据库资源筛选分析胃癌与正常胃组织间的*S100*基因表达情况,探讨利用生物信息学方法筛选胃癌与正常胃组织间的差异表达基因信息的可行性及具体方法,寻找胃癌发生发展相关*S100*基因。

表 4 关于胃癌和正常胃组织微阵列数据5个系列GSE

数据集	例数(<i>n</i>)		微阵列	杂交点
	正常组织	癌组织		
GSE2637	3	55	cDNA	13 K/17 K
GSE2685	8	22	Oligo-nucleotide	—7.2 K
GSE2669	10	64	cDNA	—7.4 K
GSE2701	22	90	cDNA	44 K
GSE3438	50	50	cDNA	14 K

表 5 SAGE 分析*S100*基因在胃癌和正常胃组织中的表达水平

基因	基因数据库簇号	SAGE标签	正常组织(tpm) ¹	癌组织(tpm) ¹
<i>S100A2</i>	516484	GATCTCTTGG	0.0	98.1
<i>S100A3</i>	433168	TCTCCCACAC	0.0	2.8
<i>S100A4</i>	81256	ATGTGTAACG	0.0	207.8
<i>S100A6</i>	275243	CCCCCTGGAT	245.9	865.9
<i>S100A7</i>	112408	GAGCAGCGCC	0.0	143.8
<i>S100A8</i>	416073	TACCTGCAGA	0.0	549.2
<i>S100A9</i>	112405	GTGGCCACGG	0.0	637.0
<i>S100A10</i>	143873	AGCAGATCAG	263.6	1890.5
<i>S100A12</i>	19413	GATTTTAAA	0.0	13.9
<i>S100A16</i>	515714	AGCAGGAGCA	0.0	130.6

仅列出了表达差异有意义的*S100*基因。¹tags per million.

据分析中显示了高表达。*S100A3*和*S100A12*则没有在虚拟电子杂交中显示阳性结果。

2.3 虚拟微阵列分析结果 虚拟微阵列分析用以进一步验证在胃癌中上调表达的*S100*基因。在GEO系列GSE2669、GSE2701和GSE3438分别显示有8、11、7个*S100*基因表达水平(表7)。*S100A2*和*S100A10*基因水平在3个GEO系列中均显示在胃癌组织中上调表达。*S100A3*则在GEO系列GSE2669和GSE2701中显示在胃癌组织中上调表达,而在GSE3438中无表达。*S100A4*、*S100A6*和*S100A7*仅仅在一个GEO系列中上调表达,而不表达于其他两个系列。

*S100A8*和*S100A9*在系列GSE2701中胃癌和正常胃组织之间差异表达无统计学意义。有意思的是,*S100A12*在系列GSE2701中胃癌和正常胃组织之间的差异表达则与SAGE数据分析结果相反。

2.4 综合分析结果 综合3个生物信息学方法(in silico)分析得出的关于*S100*基因在胃癌和胃正常组织中的表达情况。挖掘了5个*S100*基因在3个in silico分析方法中均显示在胃癌组织中上调表达,分别是*S100A2*、*S100A3*、*S100A4*、*S100A7*和*S100A10*。在这5个*S100*基因中,只有*S100A3*第1次报道是胃癌相关基因,相对正常

应用要点

首次探讨性利用生物信息学方法系统分析胃癌*S100*基因表达. 多个*S100*家族成员在胃癌中上调表达. EST和SAGE数据库不是没有局限性, 联合多数数据库挖掘肿瘤相关基因是一个可靠的方法. 通过这些数据库挖掘得到的差异基因都需要经过活体组织或体外细胞系模型的验证.

表 6 虚拟电子杂交分析*S100*基因在胃癌和正常胃组织中的表达水平

基因	EST标签(tpm ¹)		SAGE标签(tpm ¹)	
	正常组织	癌组织	正常组织	癌组织
<i>S100A2</i>	0.0	25.9	0.0	200.5
<i>S100A3</i>	0.0	0.0	0.0	0.0
<i>S100A4</i>	52.3	293.7	0.0	168.4
<i>S100A7</i>	0.0	0.0	0.0	304.8
<i>S100A9</i>	0.0	8.6	0.0	1339.4
<i>S100A10</i>	0.0	181.4	272.3	1483.7
<i>S100A12</i>	0.0	0.0	0.0	0.0
<i>S100A13</i>	0.0	43.2	0.0	0.0

仅列出了表达差异有意义的*S100*基因. ¹tags per million.

表 7 虚拟微阵列分析*S100*基因在胃癌和正常胃组织中的表达水平

基因	GSE3438			GSE2669			GSE2701		
	正常组织 ¹	癌组织 ¹	<i>P</i> 值	正常组织 ¹	癌组织 ¹	<i>P</i> 值	正常组织 ¹	癌组织 ¹	<i>P</i> 值
<i>S100A2</i>	-0.22	0.00	0.00	1.03	1.35	0.01	-0.95	-0.36	0.00
<i>S100A3</i>	-	-	-	0.84	1.55	0.03	-0.14	0.05	0.03
<i>S100A4</i>	-0.23	0.30	0.00	-	-	-	-	-	-
<i>S100A6</i>	-0.43	0.27	0.00	-	-	-	-	-	-
<i>S100A7</i>	-	-	-	-	-	-	-1.00	-0.35	0.00
<i>S100A8</i>	-	-	-	0.32	0.69	0.01	0.66	0.70	0.81
<i>S100A9</i>	-	-	-	1.25	3.48	0.00	0.42	0.79	0.13
<i>S100A10</i>	-0.50	0.42	0.00	0.46	1.36	0.00	0.19	1.22	0.00
<i>S100A12</i>	-	-	-	-	-	-	0.50	0.21	0.01

仅列出了表达差异有意义的*S100*基因. ¹微阵列数据; -: 数据集中不存在.

胃组织, *S100A3*在胃癌中上调表达. 下一步采用生物学实验方法在胃癌标本和胃癌细胞系中进一步验证*S100A3*的表达水平及基因功能研究.

3 讨论

大规模基因表达分析已经渐渐成为筛选肿瘤相关基因的重要工具^[1]. 两类重要的生物学实验方法是: (1)以DNA测序为基础的基因表达系列分析(SAGE)和表达系列标签(EST); (2)以斑点杂交为基础的微阵列分析. 来源于这些方法的大量生物信息数据收集和注解在多种高度组织的网络数据库内^[2,3]. CGAP和GEO是其中最重要的两个网络数据库^[4-8]. 利用生物信息学方法, 挖掘CGAP和GEO的生物信息资料数据, 研究者已经发现并鉴定了多个肿瘤相关基因或新的癌基因^[9-14].

*S100*蛋白家族是一个具有EF螺旋结构、

钙结合蛋白家族, 至少有22个家族成员^[15]. 17个*S100*家族成员基因定位均位于稳定性差、易在肿瘤发生中出现染色体重排的1号染色体长臂2区1带(1q21). 既往的研究^[16,17]已经证实了一些*S100*家族成员在胃癌组织中有差异表达, 包括*S100A2*、*S100A4*、*S100A7*和*S100A10*. 因此, 很有必要系统的分析和验证*S100*其他家族成员在胃癌和胃正常组织中的表达情况.

本课题联合SAGE分析、虚拟电子杂交和虚拟微阵列分析, 系统分析了*S100*基因家族在胃癌组织中的表达情况. 有研究报道, 当序列相似性超过75%, 微阵列数据分析常常发生交叉杂交错误. 然而, *S100*家族的mRNA的序列相似性处于4%-75%, 因此我们采用的三个虚拟微阵列数据分析*S100*基因表达情况发生交叉杂交的可能性非常小. 而且, SAGE分析和虚拟电子杂交分析都是以DNA测序为基础的,

因此在挖掘基因谱方面是可靠的. 联合多种数据库和多种分析方法, 发生假阳性结果或差错的可能性就相对大大减少了. 经过综合分析, 5个*S100*基因被证实是在胃癌中上调表达, 其中4个基因以前有类似报道. 这说明了我们所采用的生物信息学方法挖掘肿瘤相关基因是可靠的、可行的.

近年来, 许多*S100*家族成员在多种恶性肿瘤中有着不同的差异表达. 虽然*S100*在肿瘤的具体机制和差异表达的功能变化仍然有待进一步研究, 但不外乎与其复杂的细胞内外作用的平衡点被打破有关. 有研究^[18]报道, *S100A2*上调表达于非小细胞肺癌、食道鳞状细胞癌、喉鳞状细胞癌、卵巢乳头癌以及胃癌. *S100A4*基因和蛋白在肺癌、食管癌、乳腺癌、前列腺癌、结肠癌、甲状腺癌、肺癌和胃癌等多种肿瘤细胞中高度表达^[19-24]. 在所有恶性肿瘤中过表达的*S100A4*都涉及肿瘤的转移以及低生存率^[25]. 在胃癌中高表达的*S100A4*与淋巴结的转移、腹膜扩散和组织病理类型密切相关, 其高表达的机制可能为*S100A4*增强子区发生超甲基化. *S100A7*在乳腺导管上皮癌的原位癌中高表达, 但在正常乳腺导管上皮及其侵袭性癌中则低表达或检测不到, 其表达水平与预后和患者生存率密切相关^[26]. *S100A7*的高表达与乳腺癌的血管发生也密切相关^[27]. 在乳腺癌的早期发展中*S100A7*通过调节免疫反应能够作为一个预警分子^[28,29]. 最近有证据显示, *S100A7*与*BRCA1*和*Jab1*的关联性对成瘤起到很大的推动作用, *BRCA1*能够在转录水平调节下游靶基因*S100A7*, *BRCA1*能够在功能上抑制其表达, 因此*BRCA1*的缺失会导致*S100A7*的上调表达^[30,31]. 有人通过微阵列技术证实了*S100A10*在鳞状非小细胞肺癌和食道鳞状细胞癌中有上调表达. 上述这4个*S100*基因, *S100A2*、*S100A4*、*S100A7*和*S100A10*都在胃癌中上调表达^[20,21].

*S100*基因在一些其他肿瘤中也有下调表达. *S100A6*在前列腺癌中下调表达可能与*S100A6*超甲基化有关. 在恶性黑色素瘤、前列腺癌、肺癌、口腔癌和乳腺癌等许多肿瘤中, *S100A2*在许多肿瘤中显示下调表达, 故被称为抑癌基因. *S100A2*与*p53*相互作用, 提高其转录活性, 影响细胞周期. 这些都说明*S100*家族在不同肿瘤的发生和发展中可能起着不同的作用.

我们后续研究将通过生物学实验方法验证和分析胃癌中差异表达的*S100*基因及其功能, 为阐明*S100*和胃癌的相关性提供重要理论依据.

4 参考文献

- 1 Yang S, Shin J, Park KH, Jeung HC, Rha SY, Noh SH, Yang WI, Chung HC. Molecular basis of the differences between normal and tumor tissues of gastric cancer. *Biochim Biophys Acta* 2007; 1772: 1033-1040 [PMID: 17601708 DOI: 10.1016/j.bbdis.2007.05.005]
- 2 Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995; 270: 484-487 [PMID: 7570003 DOI: 10.1126/science.270.5235.484]
- 3 Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, Altschul SF. SAGEmap: a public gene expression resource. *Genome Res* 2000; 10: 1051-1060 [PMID: 10899154]
- 4 Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, Schuler GD, Schriml LM, Tatusova TA, Wagner L, Rapp BA. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2001; 29: 11-16 [PMID: 11125038 DOI: 10.1093/nar/29.1.11]
- 5 Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R. NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* 2007; 35: D760-D765 [PMID: 17099226 DOI: 10.1093/nar/gkl887]
- 6 Ron E, Alex L. The gene expression omnibus (GEO): a gene expression and hybridization repository. *The NCBI Handbook*, 2003: 1-17
- 7 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; 30: 207-210 [PMID: 11752295]
- 8 Reverter A, McWilliam SM, Barris W, Dalrymple BP. A rapid method for computationally inferring transcriptome coverage and microarray sensitivity. *Bioinformatics* 2005; 21: 80-89 [PMID: 15308544]
- 9 Shen D, He J, Chang HR. In silico identification of breast cancer genes by combined multiple high throughput analyses. *Int J Mol Med* 2005; 15: 205-212 [PMID: 15647832 DOI: 10.3892/ijmm.15.2.205]
- 10 Murray D, Doran P, MacMathuna P, Moss AC. In silico gene expression analysis--an overview. *Mol Cancer* 2007; 6: 50 [PMID: 17683638 DOI: 10.1186/1476-4598-6-50]
- 11 Dennis JL, Vass JK, Wit EC, Keith WN, Oien KA. Identification from public data of molecular markers of adenocarcinoma characteristic of the site of origin. *Cancer Res* 2002; 62: 5999-6005 [PMID: 12414618]
- 12 Yanglin P, Lina Z, Zhiguo L, Na L, Haifeng J, Guoyun Z, Jie L, Jun W, Tao L, Li S, Taidong Q, Jianhong W, Daiming F. KCNE2, a down-regulated

■名词解释

表达谱数据库: 把高通量技术得到的组织和细胞表达谱数据加工、存储, 形成利于电子传播和使用的一类生物信息数据库. 癌基因组解剖计划 (CGAP) 和基因表达综合数据库 (GEO) 是其中最重要的两个网络数据库. 表达谱数据库从技术原理上划分为三类, 即: 表达序列标签 (EST) 文库、基因表达系列分析 (SAGE) 文库和 cDNA 微阵列数据库.

同行评价

本课题组利用多种生物信息学工具, 探讨了S100基因家族的表达变化, 发现5个基因上调, S100A3是一个新的高表达分子。这一研究有别于传统的生物学实验, 是在已有数据的基础上做出的新发现。课题设计新颖, 方法恰当, 结果丰富, 结论可靠, 对胃癌分子机制的研究有一定的借鉴价值。

- gene identified by in silico analysis, suppressed proliferation of gastric cancer cells. *Cancer Lett* 2007; 246: 129-138 [PMID: 16677757]
- 13 Yasui W, Oue N, Ito R, Kuraoka K, Nakayama H. Search for new biomarkers of gastric cancer through serial analysis of gene expression and its clinical implications. *Cancer Sci* 2004; 95: 385-392 [PMID: 15132764 DOI: 10.1111/j.1349-7006.2004.tb03220.x]
- 14 Aouacheria A, Navratil V, Barthelaix A, Mouchiroud D, Gautier C. Bioinformatic screening of human ESTs for differentially expressed genes in normal and tumor tissues. *BMC Genomics* 2006; 7: 94 [PMID: 16640784]
- 15 Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS. Calcium-dependent and -independent interactions of the S100 protein family. *Biochem J* 2006; 396: 201-214 [PMID: 16683912]
- 16 El-Rifai W, Moskaluk CA, Abdrabbo MK, Harper J, Yoshida C, Riggins GJ, Frierson HF, Powell SM. Gastric cancers overexpress S100A calcium-binding proteins. *Cancer Res* 2002; 62: 6823-6826 [PMID: 12460893]
- 17 Yonemura Y, Endou Y, Kimura K, Fushida S, Bandou E, Taniguchi K, Kinoshita K, Ninomiya I, Sugiyama K, Heizmann CW, Schafer BW, Sasaki T. Inverse expression of S100A4 and E-cadherin is associated with metastatic potential in gastric cancer. *Clin Cancer Res* 2000; 6: 4234-4242 [PMID: 11106237]
- 18 Heizmann CW, Fritz G, Schäfer BW. S100 proteins: structure, functions and pathology. *Front Biosci* 2002; 7: d1356-d1368 [PMID: 11991838]
- 19 Emberley ED, Murphy LC, Watson PH. S100 proteins and their influence on pro-survival pathways in cancer. *Biochem Cell Biol* 2004; 82: 508-515 [PMID: 15284904]
- 20 Bartling B, Rehbein G, Schmitt WD, Hofmann HS, Silber RE, Simm A. S100A2-S100P expression profile and diagnosis of non-small cell lung carcinoma: impairment by advanced tumour stages and neoadjuvant chemotherapy. *Eur J Cancer* 2007; 43: 1935-1943 [PMID: 17689067 DOI: 10.1016/j.ejca.2007.06.010]
- 21 Dokun OY, Florl AR, Seifert HH, Wolff I, Schulz WA. Relationship of SNCG, S100A4, S100A9 and LCN2 gene expression and DNA methylation in bladder cancer. *Int J Cancer* 2008; 123: 2798-2807 [PMID: 18803290 DOI: 10.1002/ijc.23893]
- 22 Ismail NI, Kaur G, Hashim H, Hassan MS. S100A4 overexpression proves to be independent marker for breast cancer progression. *Cancer Cell Int* 2008; 8: 12 [PMID: 18771601 DOI: 10.1186/1475-2867-8-12]
- 23 Hemandas AK, Salto-Tellez M, Maricar SH, Leong AF, Leow CK. Metastasis-associated protein S100A4—a potential prognostic marker for colorectal cancer. *J Surg Oncol* 2006; 93: 498-503 [PMID: 16615153 DOI: 10.1002/jso.20460]
- 24 Cho YG, Nam SW, Kim TY, Kim YS, Kim CJ, Park JY, Lee JH, Kim HS, Lee JW, Park CH, Song YH, Lee SH, Yoo NJ, Lee JY, Park WS. Overexpression of S100A4 is closely related to the aggressiveness of gastric cancer. *APMIS* 2003; 111: 539-545 [PMID: 12887505 DOI: 10.1034/j.1600-0463.2003.1110502.x]
- 25 Matsumoto K, Irie A, Satoh T, Ishii J, Iwabuchi K, Iwamura M, Egawa S, Baba S. Expression of S100A2 and S100A4 predicts for disease progression and patient survival in bladder cancer. *Urology* 2007; 70: 602-607 [PMID: 17688917 DOI: 10.1016/j.urology.2007.04.007]
- 26 Carlsson H, Petersson S, Enerbäck C. Cluster analysis of S100 gene expression and genes correlating to psoriasin (S100A7) expression at different stages of breast cancer development. *Int J Oncol* 2005; 27: 1473-1481 [PMID: 16273201]
- 27 Salama I, Meehaimeed F, Marshall J, Jazayeri M, Thomas G, Duffy S, Jones JL. S100A7 and the basal phenotype in breast cancer. *Eur J Surg Oncol* 2006; 32(Suppl 1): S79 [DOI: 10.1016/S0748-7983(06)70697-7]
- 28 Mandal S, Curtis L, Pind M, Murphy LC, Watson PH. S100A7 (psoriasin) influences immune response genes in human breast cancer. *Exp Cell Res* 2007; 313: 3016-3025 [PMID: 17560571 DOI: 10.1016/j.yexcr.2007.03.020]
- 29 Emberley ED, Niu Y, Njue C, Kliever EV, Murphy LC, Watson PH. Psoriasin (S100A7) expression is associated with poor outcome in estrogen receptor-negative invasive breast cancer. *Clin Cancer Res* 2003; 9: 2627-2631 [PMID: 12855640]
- 30 Kennedy RD, Gorski JJ, Quinn JE, Stewart GE, James CR, Moore S, Mulligan K, Emberley ED, Lioe TF, Morrison PJ, Mullan PB, Reid G, Johnston PG, Watson PH, Harkin DP. BRCA1 and c-Myc associate to transcriptionally repress psoriasin, a DNA damage-inducible gene. *Cancer Res* 2005; 65: 10265-10272 [PMID: 16288014 DOI: 10.1158/0008-5472.CAN-05-1841]
- 31 Krop I, März A, Carlsson H, Li X, Bloushtain-Qimron N, Hu M, Gelman R, Sabel MS, Schnitt S, Ramaswamy S, Kleer CG, Enerbäck C, Polyak K. A putative role for psoriasin in breast tumor progression. *Cancer Res* 2005; 65: 11326-11334 [PMID: 16357139 DOI: 10.1158/0008-5472.CAN-05-1523]

编辑: 郭鹏 电编: 闫晋利





Published by **Baishideng Publishing Group Inc**
8226 Regency Drive, Pleasanton,
CA 94588, USA
Fax: +1-925-223-8242
Telephone: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
<http://www.wjgnet.com>



ISSN 1009-3079

