

· 论 著 ·

乳腺癌高表达基因 FAM83A 的生物信息学分析*

刘 镭¹,黄 亮²,车清海³,赵恩宏²,黄 旭²,程露阳¹,肖丽君¹

(1. 承德医学院基础医学部免疫教研室,河北承德 067000;2. 承德医学院附属医院肿瘤诊治中心,河北承德 067000;3. 吉林市龙潭区妇幼保健院妇产科 132000)

摘要:**目的** 筛选乳腺癌高表达基因,用生物信息学方法预测 FAM83A 生物学功能。**方法** 使用肿瘤基因组解剖计划数据库的数字基因表达演示工具,在乳腺癌组织和乳腺良性肿瘤组织中筛选差异表达的基因。生物信息学方法预测 FAM83A 基因及其编码蛋白的序列同源性、理化性质、亚细胞定位、结构和表达谱。RT-PCR 方法测定 FAM83A 在乳腺癌、乳腺纤维瘤和正常乳腺组织中的表达。**结果** 筛选出差异表达基因 FAM83A。发现其有 3 个不同的剪接体,基因进化保守,亚细胞定位于线粒体的可能性大。预测有 PK,PKA 等磷酸化位点,MAPK 作用识别位点,SH2、SH3,RPEL 和 WH2 结合基序。二级结构主要是 α 螺旋和无规则卷曲。RT-PCR 证实 FAM83A 在乳腺癌组织中高表达,而在乳腺良性肿瘤和正常乳腺组织中均无表达。**结论** FAM83A 是一个乳腺癌高表达基因,可能作为乳腺癌分子机制研究的靶点。

关键词:乳腺肿瘤;生物信息学;SAGE 库; FAM83A

doi:10. 3969/j. issn. 1671-8348. 2013. 03. 002 文献标识码:A 文章编号:1671-8348(2013)03-0243-03

Bioinformatics analysis of the highly expressed FAM83A gene in breast cancer*

Liu Lei¹,Huang Liang²,Che Qinghai³,Zhao Enhong²,
Huang Xu²,Cheng Luyang¹,Xiao Lijun¹

(1. Department of Immunology,Basic Medical Institute,Chengde Medical College,Chengde,Hebei 067000,China;2. Cancer Diagnosis and Treatment Center,Affiliated Hospital of Chengde Medical College,Chengde,Hebei 067000,China;3. Department of Obstetrics and Gynecology, Maternal and Child Health Hospital of Longtan District,Jilin 132000,China)

Abstract:**Objective** To screen and find over-expressed gene in breast cancer,and to predict the function of FAM83A by bioinformatics.**Methods** We used digital gene expression displayer(DGED) tool in cancer genome anatomy project(CGAP) to screen the differentially expressed genes between the breast cancer and benign tumor. The sequence homology,properties,subcellular location,structure and expression profile of FAM83A were analyzed by bioinformatics software. At last,we detected the expression of FAM83A in breast cancer,breast fibroadenoma and normal breast tissue by RT-PCR.**Results** We found a differentially expressed gene FAM83A which had three transcript variants. It was evolutionary conservation. Bioinformatics analysis indicated that FAM83A protein was located likely in the mitochondria. There were PK,PKA phosphorylation sites,MAPK recognition site,SH2,SH3,RPEL and WH2 binding motif. The secondary structure mainly comprised of alpha helices and random coils. RT-PCR results confirmed that FAM83A was over-expressed in breast cancer tissues,while no expression in benign breast tumor and normal breast tissue.**Conclusion** FAM83A is over-expressed in breast cancer and maybe act as a study target for molecular mechanism.

Key words:breast neoplasms;bioinformatics;SAGE library; FAM83A

乳腺癌是严重威胁妇女健康的恶性肿瘤,发病率已跃居国内妇女癌症发病首位^[1]。X 线片筛查受影像学特点的制约,使很多患者失去了早期发现的机会^[2]。筛查肿瘤生物标记物对乳腺癌的早期发现具有重要意义,成为目前研究的主要焦点^[3]。随着生物数据的急剧增长和计算机技术的快速提高,生物信息学这一新兴学科得到了前所未有的迅速发展,尤其是应用生物信息学方法发现新基因和进行功能研究^[4]。本文通过肿瘤基因组解剖计划(cancer genome anatomy project,CGAP)基因表达系列分析(SAGE)数据库,全面分析肿瘤组织基因表达谱,筛选新的乳腺癌高表达基因,并对其中一条功能未知的新基因 FAM83A 进行了初步的生物学分析,为进一步深入研究该基因的生物学功能及其参与乳腺癌发生发展的分子机制奠定了基础。

1 材料与方法

1.1 材料 乳腺癌细胞系 MDA-MB-231 和 MCF-7 来自承德医学院中心实验室。承德医学院附属医院收集 5 例乳腺癌手术切除标本(均经术后病理证实),相应的距癌灶边缘 2 cm 以外的癌旁组织,以及 5 例乳腺纤维瘤组织,各约 100 mg。标本装入无菌管内,迅速投入液氮罐内冷冻。

1.2 方法

1.2.1 差异基因的筛选 利用 CGAP 提供的 SAGE Genie 数据库,选取数字基因表达演示工具。选择乳腺癌组织和乳腺良性肿瘤两个池,设置 F 值为 2,即表达差异大于 2 倍以上;Q 为 0.1。差异表达的基因按差异显著性高低排序,比对出在短、长序列标签文库中都出现的差异基因。并对其中一条功能未知的高表达序列 FAM83A 进行初步的生物信息学分析。

* 基金项目:河北省人口与计划生育委员会资助项目(2009-B21);河北省卫生厅资助项目(20090579)。 作者简介:刘镭(1977~),讲师,博士,主要从事肿瘤免疫的研究。

1.2.2 序列比对 利用 BLAST 软件检索 EST 数据库、nr 数据库及 pro 数据库。将所克隆的基因核苷酸序列与 GeneBank 中已知核苷酸序列进行同源性分析。氨基酸序列相似性分析用 NCBI/blastp,多序列比对和进化树分析用 clustal W。

1.2.3 理化性质和亚细胞定位 对蛋白质分子量、等电点、疏水性等理化性质分析,用 ExPASy-protparam 工具。用 PSORT II 服务程序进行亚细胞定位预测。

1.2.4 蛋白质结构预测 ELM 预测蛋白质功能性位点(<http://elm.eu.org/>)^[5]。信号肽及剪切位点预测用 SignalIP。二级结构预测应用 SOPMA 工具。保守结构域用 NCBI/CDD 预测。

1.2.5 电子表达谱分析 以 FAM83A 为关键词,通过 UniGene 库中的 EST 表达情况分析获得 FAM83A 基因电子表达谱。

1.2.6 半定量 RT-PCR 验证性实验 TRIZOL 法提取乳腺癌细胞系、乳腺癌、癌旁、乳腺纤维瘤组织总 RNA,按逆转录试剂盒(QIAGEN)操作步骤进行逆转录。PCR 反应体系 25 μ L,其中包括 2.5 μ L cDNA 模板,2.5 μ L 10 $\times$ PCR 反应缓冲液,0.5 μ L dNTP(10 mmol/L),0.5 μ L Taq DNA 聚合酶,上下游引物各 0.5 μ L。FAM83A 上游引物为:5'-TAG AGG CAG CCA ACA AGC GTG -3';下游引物为:5'- CCT TTG GAA AGC TGG GCT TGG GAG- 3'。反应条件为 94 $^{\circ}$ C,15 s;60 $^{\circ}$ C,30 s;72 $^{\circ}$ C 1 min,30 个循环。以甘油醛-3-磷酸脱氢酶(GAPDH)作为反应内参。取 10 μ L PCR 产物,在 1%琼脂糖凝胶上进行电泳检测。

2 结 果

2.1 乳腺癌高表达基因的筛选 最终筛选出 33 个在乳腺癌中高表达的基因,它们多数在 Gene Bank 中都被详细注释,并从中选取一条功能未知的明显高表达的基因 FAM83A。选择短标签序列搜索后结果显示,在 38 个乳腺癌短序列标签文库中,有 31 个文库具有 FAM83A 代表标签(ATGTACAGGT),共出现 1 575 次。而在 7 个良性乳腺肿瘤短序列标签文库中,有 2 个文库出现该代表标签,表达差异为 7.46 倍。同时,在 23 个乳腺癌长序列标签文库中,有 20 个文库具有 FAM83A 代表标签(ATGTACAGGTTTGTAGC),共出现 964 次。而在 5 个良性乳腺肿瘤长序列标签文库中,只有一个文库出现该代表标签,表达差异为 3.14 倍。

2.2 FAM83A 在 GeneBank 中的序列比对结果及不同剪接体组成 应用 BLAST 工具比对,发现共有 6 条序列与 FAM83A 高度同源,多为人类新 mRNA 序列。其中发现两条序列是 FAM83A 的转录变异体, GeneBank 登录号分别为 NM_207006.1 和 NM_032899.4,加上最初发现的转录变异体, FAM83A 共有 3 个不同的剪接体。经对 FAM83A 的多物种氨基酸序列相似性比较,人与长臂猿、小鼠、火鸡和爪蟾的相似性分别为 98%、87%、64%和 41%,可见氨基酸组成在脊椎动物中保守性较高。各物种序列大小接近,起始氨基酸都是蛋氨酸。基于 FAM83A 多序列比对进化树分析,人和长臂猿在进化上最为接近,在其他物种也有比较接近的进化来源。

2.3 理化性质与亚细胞定位 FAM83A(AF497803)由 367 个氨基酸组成,生物软件预测相对分子质量为 40 606,等电点为 8.97。该蛋白质的半衰期在人体网织红细胞为 30 h;不稳定指数为 51.49(>40 考虑为不稳定),分类为不稳定蛋白质;脂肪指数为 77.06,平均疏水值为-0.346,是亲水蛋白质。根据软件分析,预测该蛋白质在线粒体中表达的可能性稍大

(43.5%),其次分别是细胞核(34.8%)和细胞质(21.7%)。

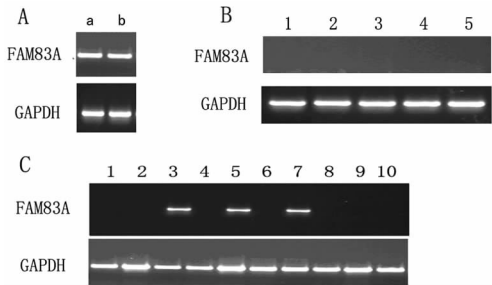
2.4 蛋白质结构 ELM 预测发现有 PK、PKA 等磷酸化位点,MAPK 作用识别位点,SH2、SH3,RPEL 和 WH2 结合基序等,未发现保守结构域存在。未预测出信号肽剪切位点。SOPMA 方法预测二级结构,发现 FAM83A 中 α 螺旋占 43.5%, β 折叠占 5.72%,延伸链和无规则卷曲结构分别占 11.99%和 48.5%。

2.5 FAM83A 基因表达 电子表达谱分析显示,FAM83A 基因在部分正常组织和多种肿瘤中均有表达,见表 1。

表 1 FAM83A 基因电子表达谱

组织	转录本/百万	组织	转录本/百万
口	1 476	头/颈癌	792
喉	414	肺癌	203
颈	103	膀胱癌	171
食道	98	前列腺癌	136
前列腺	89	胰腺癌	57
肺	74	食管肿瘤	57
膀胱	67	白血病	52
骨髓	61	乳腺癌	51

2.6 FAM83A 在乳腺癌细胞系和乳腺癌组织中的表达 RT-PCR 结果显示,FAM83A 在乳腺癌细胞系 MDA-MB-231 和 MCF-7 中均有表达(图 1A)。在 5 例乳腺癌组织中,3 例 FAM83A 高表达;而在相应的癌旁组织和乳腺纤维瘤中均不表达,见图 1B~C。



A:在乳腺癌细胞系中表达,a为MDA-MB-231,b为MCF-7;B:在5例乳腺纤维瘤中的表达;C:在5例乳腺癌及其癌旁组织中的表达;1,3,5,7,9泳道为乳腺癌组织;2,4,6,8,10泳道为相应的癌旁组织。

图 1 半定量 RT-PCR 检测 FAM83A 基因的表达

3 讨 论

乳腺是最常发生良、恶性肿瘤的器官,癌性增生与良性肿瘤有着本质的不同,治疗方案也完全不同^[6]。所以,在筛选乳腺癌高表达基因时,临床工作者希望寻找在癌组织中高表达,而在良性肿瘤组织中不表达或低表达的基因,对指导临床的诊断和治疗更具有意义。CGAP SAGE 数据库能够全面分析肿瘤组织基因表达谱、寻找肿瘤特异性表达新基因^[7]。在 SAGE 文库中,本研究选择了乳腺癌和良性乳腺肿瘤两个池,提交后看到良性乳腺肿瘤池的文库在短序列标签中有 7 个,长序列标签中有 5 个,主要是:乳腺纤维瘤、乳腺肌上皮瘤和良性乳腺间质瘤,并最终筛选出乳腺癌高表达基因 FAM83A。而且,半定量 RT-PCR 结果也进一步证实了生物信息学数据库的可靠性。该基因在乳腺癌细胞系和多例乳腺癌组织中均高表达。

目前,FAM83A 的研究很少,只是将该基因作为肿瘤生物

学标记物用于外周血的检测^[8-9],但关于该基因的功能少见文献报道。ELM 分析提示,该蛋白质具有 LIG_14-3-3_1,LIG_14-3-3_3 蛋白质配体,cyclin 底物识别位点和 PK、PKA 磷酸化位点等。14-3-3 蛋白质是真核生物中的一类保守蛋白质家族,它们在细胞中发挥着信号转导,细胞周期调控,凋亡和应激反应以及肿瘤恶性转化等重要的作用,是蛋白质间相互作用的媒介和调节者^[10]。Cyclin 底物识别位点在 cyclin/CDK 相互作用蛋白质中被广泛发现。该基序在 CDK 中的存在增加了(ST)Px(KR)基序的磷酸化水平。它通过 cyclin 蛋白质中的保守区域被识别,并与 p21Kip cyclin 相似的方法结合^[11]。PK 和 PKA 磷酸化位点是信号传导调控中的常见分子结构,在细胞增殖及细胞周期中有着广泛的调节作用^[12]。

以上分析发现,该蛋白质在进化上保守,从低等到高等动物都有相似蛋白质的表达,是不稳定的亲水性蛋白质,可能定位于线粒体上,有着多个功能位点和基序,提示参与信号转导和细胞周期调控。特别要提出的是,该基因存在多个转录变体,表明它在生物体发育过程中的细胞增殖和凋亡的调控方面起着重要的作用。该基因在乳腺癌中的高表达具有显著意义,很可能是通过使细胞持续增生,控制凋亡和改变细胞周期而发挥作用的。

参考文献:

[1] Yang L, Parkin DM, Ferlay J, et al. Estimates of cancer incidence in China for 2000 and projections for 2005[J]. Cancer Epidemiol Biomarkers Prev, 2005, 14(1): 243-250.

[2] 任丽, 陈瑛. 乳腺癌骨转移的实验室检测进展[J]. 国际检验医学杂志, 2010, 31(2): 137-138, 141.

[3] Liska V, Holubec L Jr, Treska V, et al. Evaluation of tumour markers as differential diagnostic tool in patients with suspicion of liver metastases from breast cancer[J]. Anticancer Res, 2011, 31(4): 1447-1451.

(上接第 242 页)

[4] Babu RP, Sekhar LN, Wright DC. Extreme lateral transcondylar approach; technical improvements and lessons learned [J]. J Neurosurg, 1994, 81(1): 49-59.

[5] Kawashma M, Tanriover N, Rhoton AL Jr, et al. Comparison of the far lateral and extreme lateral variants of the atlanto-occipital transarticular approach to anterior extradural lesions of the craniovertebral junction [J]. Neurosurgery, 2003, 53(3): 662-674.

[6] Liu JK, Ganesh R, Couldwell WT, et al. Far lateral transcondylar transtubercular approach to lesions of the ventral foramen magnum and craniovertebral junction [J]. Contemp Neurosurg, 2007, 29(10): 1-7.

[7] 蔡润, 彭翔, 屈宏彦, 等. 枕下极外侧入路手术路径显微解剖研究[J]. 华中科技大学学报: 医学版, 2008, 37(1): 101-104.

[8] 闫明, 王超, 王圣林. 正常椎动脉解剖特点及其变异概况[J]. 中国脊柱脊髓杂志, 2012, 22(2): 171-174.

[9] Guo Q, Liu J, Ni B, et al. C2 laminar screw and C1-2 transarticular screw combined with C1 laminar hooks for atlantoaxial instability with unilateral vertebral artery in-

[4] 刘娜, 郭云波, 孙显赫, 等. 比较基因组学平台的设计与实现[J]. 南方医科大学学报, 2010, 30(2): 219-223.

[5] Puntervoll P, Linding R, Gemund C, et al. ELM server A new resource for investigating short functional sites in modular eukaryotic protein[J]. Nucleic Acids Res, 2003, 31(13): 3625-3630.

[6] 陈国平, 夏立平, 翁阳, 等. 乳腺癌 90 例临床与病理分析[J]. 海南医学院学报, 2009, 15(2): 151-153.

[7] Kavak E, Unlü M, Nistér M, et al. Meta-analysis of cancer gene expression signatures reveals new cancer genes, SAGE tags and tumor associated regions of co-regulation [J]. Nucleic Acids Res, 2010, 38(20): 7008-7021.

[8] Li Y, Dong X, Yin Y, et al. BJ-TSA-9, a novel human tumor-specific gene, has potential as a biomarker of lung cancer[J]. Neoplasia, 2005, 7(12): 1073-1080.

[9] Liu L, Liao GQ, He P, et al. Detection of circulating cancer cells in lung cancer patients with a panel of marker genes[J]. Biochem Biophys Res Commun, 2008, 372(4): 756-760.

[10] Tzivion G, Avruch J. 14-3-3 proteins active cofactors in cellular regulation by serine/threonine phosphorylation [J]. Biol Chem, 2002, 277(5): 3061-3064.

[11] Lowe ED, Tews I, Cheng KY, et al. Specificity determinants of recruitment peptides bound to phospho-CDK2/Cyclin A[J]. Biochemistry, 2003, 41(52): 15625-15634.

[12] Dash PK, Moore AN, Kobori N, et al. Molecular activity underlying working memory [J]. Learn Mem, 2007, 14(8): 554-563.

(收稿日期: 2012-10-03 修回日期: 2012-12-03)

jury [J]. Arch Orthop Trauma Surg, 2011, 131(9): 1207-1210.

[10] Masashi Neo, Fujibayashi S, Miyata M, et al. Vertebral artery injury during cervical spine surgery: a survey of more than 5600 operations [J]. Spine (Phila Pa 1976), 2008, 33(7): 779-785.

[11] 王洪正, 刘窗溪, 刘健, 等. 远外侧经枕髁手术入路防止椎动脉损伤的应用解剖[J]. 中国临床解剖学杂志, 2003, 21(2): 136-139.

[12] Nanda A, Vincent DA. Far-lateral approach to intra-dural lesions of the foramen magnum without resection of the occipital condyle[J]. J Neurosurg, 2002, 96(2): 302-309.

[13] 付爱军, 朱军, 李建琨, 等. 枕下极外侧入路的显微解剖研究[J]. 华北煤炭医学院学报, 2011, 13(4): 464-466.

[14] 李聰慧, 叶建亚, 张金峰, 等. 枕下极外侧入路的显微解剖及临床应用[J]. 脑与神经疾病杂志, 2010, 18(1): 33-35.

[15] 关靖宇, 卢天舒, 佟晓杰. 枕下极外侧入路的解剖学量化研究[J]. 解剖科学进展, 2009, 15(4): 403-407.

(收稿日期: 2012-10-15 修回日期: 2012-12-27)