

# SH2D7 基因及其编码蛋白的生物信息学分析

杨华<sup>1,2</sup>, 杨风琴<sup>1,2</sup>, 廖国玲<sup>1,2</sup>, 刘学英<sup>1,2</sup>

1. 宁夏医科大学检验学院, 银川 750004
2. 宁夏检验医学研究所, 银川 750004

**摘要** 为了对 *SH2D7* 基因序列及其编码蛋白进行生物信息学分析, 应用 NCBI 数据库和 ExPASy 等程序, 对 *SH2D7* 基因核苷酸序列进行了分析, 并预测了其编码蛋白的理化性质及二级结构等。结果表明, *SH2D7* 基因定位于人类染色体 15q25.1, 包含了 6 个外显子和 5 个内含子。 *SH2D7* 基因编码蛋白的功能主要涉及基因表达调控、信号传导等; 分子量约为 49.8 kD, 等电点为 5.99, 是一种亲水蛋白, 主要位于细胞核中, 有 35 个磷酸化位点及 22 个抗原位点。

**关键词** 肿瘤; *SH2D7*; 生物信息学分析

**中图分类号** R73

**文献标志码** A

**doi** 10.3981/j.issn.1000-7857.2015.02.003

## Bioinformatic analysis of *SH2D7* and its encoded protein

YANG Hua<sup>1,2</sup>, YANG Fengqin<sup>1,2</sup>, LIAO Guoling<sup>1,2</sup>, LIU Xueying<sup>1,2</sup>

1. School of Laboratory Medicine, Ningxia Medical University, Yinchuan 750004, China
2. Ningxia Institute of Laboratory Medicine, Yinchuan 750004, China

**Abstract** The bioinformatic software is used to predict the nucleotide sequence of *SH2D7* and the second structure of its encoding protein. It is shown that *SH2D7* is located on the human chromosome 15q25.1, containing six exons and five introns. *SH2D7* protein is mainly involved in the regulation of the gene expression and the signal transduction. Its molecular weight is 49.8 kD and the isoelectric point is 5.99. *SH2D7* is a hydrophilic protein mainly located in the nucleus and there are 35 phosphorylation sites and 22 epitopes bit in the *SH2D7* protein sequence.

**Keywords** cancer; *SH2D7*; bioinformatic analysis

生物信息学是一门利用计算机技术研究生物系统的规律的新学科, 通过综合利用生物学、计算机技术和信息技术, 揭示大量而复杂的生物数据包含的生物学奥秘。生物信息学包括生物信息的获取、处理、存储、分配和解释。作为一个新的学科领域, 生物信息学把基因组核苷酸序列信息分析作为源头, 在获得蛋白质编码区的信息后进行蛋白质空间结构模拟和预测。随着生命科学和计算机科学的迅猛发展, 使得利用生物信息学进行科学研究变得更加快捷、简便。利用生物信息学方法对基因的功能和特性进行分析, 不但为进一步实验

提供指导, 而且可以获得关于基因功能研究的重要线索。

肿瘤通常是由一些基因的结构或表达调控的改变引起的, 利用生物信息学方法筛选肿瘤差异表达基因并对其编码蛋白的性质与功能进行预测已成为研究手段之一<sup>[1-3]</sup>。本研究利用 EST 数据库筛选得到一个肿瘤差异表达基因 SH2 domain containing 7 (*SH2D7*), 这是一种蛋白编码基因。本研究利用生物信息学技术对 *SH2D7* 基因序列进行初步分析, 推测其编码蛋白的分子生物学特征, 为进一步研究其结构与功能提供新思路。

收稿日期: 2014-07-07; 修回日期: 2014-10-07

基金项目: 宁夏自然科学基金项目 (NZ11135)

作者简介: 杨华, 副教授, 研究方向为肿瘤发生的分子机制及临床应用, 电子信箱: yanghua-126@163.com

引用格式: 杨华, 杨风琴, 廖国玲, 等. *SH2D7* 基因及其编码蛋白的生物信息学分析[J]. 科技导报, 2015, 33(2): 35-38.

## 1 材料与方法

### 1.1 *SH2D7* 基因的获取

本研究利用CGAP网站提供的肿瘤差异表达基因筛选工具 cDNA xProfile 和 DGED, 以人类正常组织和肿瘤组织 EST 数据库为基础, 对全基因组进行筛选得到了1个肿瘤差异表达基因 *SH2D7*。

### 1.2 *SH2D7* 基因序列分析

利用NCBI网站的在线分析软件 MapViewer (<http://www.ncbi.nlm.nih.gov/mapview>) 分析 *SH2D7* 基因序列。

### 1.3 *SH2D7* 基因编码蛋白的性质、功能分析

本研究利用 ExPASy 中的 ProtParam 软件 (<http://web.expasy.org/protparam/>) 分析 *SH2D7* 基因编码蛋白分子质量、等电点 (PI)、氨基酸组成、原子组成、半衰期、稳定性、疏水性等。通过 SignalP 4.1 软件 (<http://www.cbs.dtu.dk/services/SignalP/>) 预测蛋白质是否含有信号肽。通过工具 PSORT (<http://psort.hgc.jp/form2.html>) 预测 *SH2D7* 基因编码蛋白的亚细胞定位。运用 NetPhos 2.0 软件 (<http://www.cbs.dtu.dk/services/NetPhos/>) 对 *SH2D7* 基因编码蛋白进行分析, 判断其是否含有磷酸化位点。通过巴斯德研究所 Protein Sequence Analysis 服务器中的 antigenic 软件 (<http://bioweb.pasteur.fr/seqanal/interfaces/antigenic.html>) 预测编码蛋白的抗原位点。通过 predictprotein 服务器中的 PDH 软件 (<https://www.predictprotein.org/>) 预测蛋白质的二级结构。

## 2 结果

### 2.1 基因序列分析

NCBI 网站的在线分析软件 MapViewer 分析显示, *SH2D7* 定位于人类染色体 15q25.1。该基因包含了6个外显子和5个内含子(图1), 其编码序列为1~1356。将该基因与NCBI中非冗余基因数据库进行比对, 未发现与之相似的基因序列, 证明 *SH2D7* 是一个新基因。



图1 *SH2D7* 基因序列的生物信息学分析  
Fig. 1 Bioinformatic analysis of *SH2D7* gene

### 2.2 *SH2D7* 基因编码蛋白的序列性质、功能分析

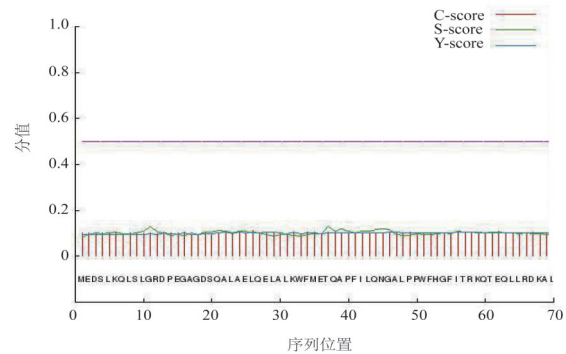
#### 2.2.1 *SH2D7* 基因编码蛋白的理化特性

用 ExPASy 中的 ProtParam 软件分析 *SH2D7* 基因编码蛋白的分子量为 49.807 kD, 等电点 (PI) 为 5.99, 总原子数为 6916, 哺乳动物中的半衰期为 30 h, 不稳定指数 55.52 (属于不稳定蛋白)。

#### 2.2.2 亲水性/疏水性、信号肽预测

用 ExPASy 中的 ProtParam 软件计算 *SH2D7* 基因编码蛋

白的疏水性平均值为-0.698 (正值表明此蛋白为疏水蛋白, 负值表明为亲水蛋白), 说明该蛋白为亲水蛋白。经 SignalP 软件对 *SH2D7* 编码蛋白分析未发现信号肽区域, 结果见图2。



# Measure	Position	Value	Cutoff	signal peptide?
max. C	26	0.111		
max. Y	26	0.107		
max. S	11	0.131		
mean S	1-25	0.102	0.450	NO
D	1-25	0.105		

Name=gi\_157820785\_ref NP\_001094874.1\_ SP='NO' D=0.105 D-cutoff=0.450 Networks=SignalP-noTM

图2 SignalP 基于神经网络算法的蛋白质信号肽预测  
Fig. 2 SignalP-NN results for protein signal peptide prediction

#### 2.2.3 *SH2D7* 基因编码蛋白的亚细胞定位

通过 PSORT 软件进行亚细胞定位的预测, 结果显示 73.9% 的可能位于胞核, 13% 的可能位于胞浆, 13% 的可能位于线粒体。

#### 2.2.4 磷酸化位点预测

运用 NetPhos 2.0 软件分析, *SH2D7* 基因编码的蛋白质可能含有 35 个磷酸化位点, 包括 21 个丝氨酸磷酸化位点 (9, 78, 87, 117, 183, 187, 204, 209, 213, 221, 229, 275, 285, 303, 314, 336, 346, 372, 393, 407, 427 位), 7 个苏氨酸磷酸化位点 (37, 57, 161, 175, 327, 366, 433 位), 7 个酪氨酸磷酸化位点 (84, 107, 151, 243, 351, 374, 419 位), 如图3所示。

#### 2.2.5 蛋白质结构预测

通过 PredictProtein 服务器软件对 *SH2D7* 基因编码蛋白进行二级结构的推测。

根据 PDH 程序设定的标准:

$\alpha$  结构类蛋白: H>45% and E<5%

$\beta$  结构类蛋白: H<5% and E>45%

$\alpha$ - $\beta$  组合类蛋白: H>30% and E>20%

混合类蛋白: 其他

*SH2D7* 基因编码蛋白: H 4.66%, E 4.88%, L 90.47%。因此, 该蛋白属于混合类蛋白 (H 代表  $\alpha$ -螺旋, E 代表  $\beta$ -片层, L 代表其他)。 *SH2D7* 编码蛋白二级结构预测结果见图4。

#### 2.2.6 抗原位点预测

分析 *SH2D7* 基因编码蛋白发现有 22 个可能的抗原位点, 分别位于 170—181 位, 116—133 位, 295—305 位, 189—207

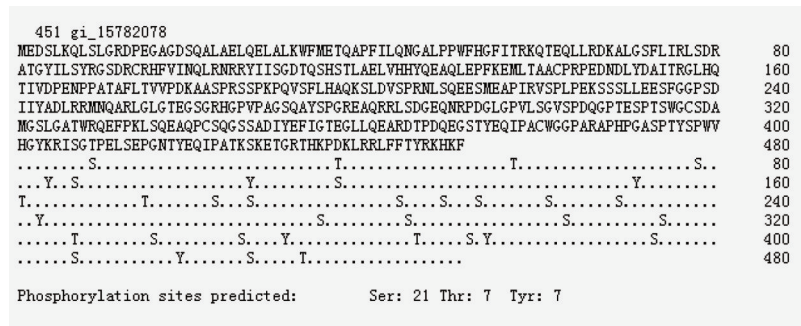
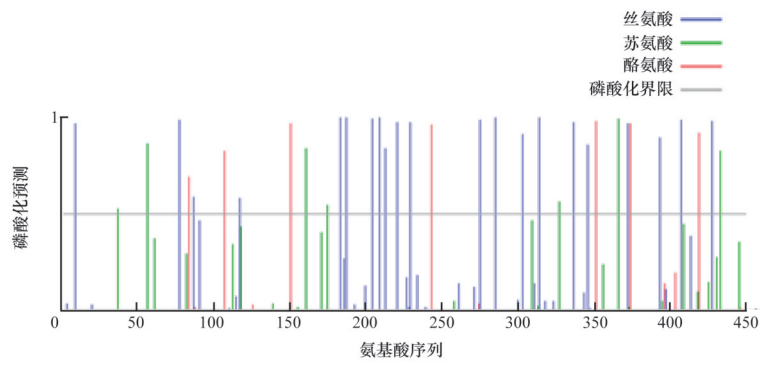


图3 SH2D7 编码蛋白磷酸化位点预测

Fig. 3 Predicted phosphorylation sites in SH2D7 encoded protein

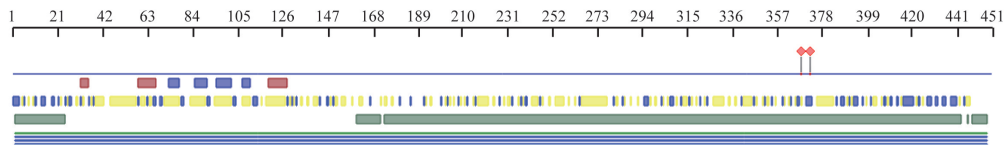


图4 PredictProtein 分析的 SH2D7 二级结构

Fig. 4 Secondary structure of SH2D7 using PredictProtein analysis

位, 93—101 位, 217—233 位, 5—11 位, 375—383 位, 70—78 位, 159—166 位, 39—47 位, 139—144 位, 20—32 位, 49—55 位, 386—404 位, 83—89 位, 240—245 位, 62—68 位, 264—275 位, 332—345 位, 439—446 位, 348—353 位。

### 3 讨论

肿瘤是危害人类健康的重要疾病,多数患者早期没有明显症状,而临床诊断又缺乏特异性的标志物,因此恶性肿瘤在中国已成为城乡居民死亡的首要原因。寻找并研究肿瘤相关基因一直是研究热点,利用 EST 数据库对全基因组进行肿瘤差异表达基因的生物信息学筛选,是一种经济、快速、有效的筛选方法<sup>[4-7]</sup>,而利用计算方法对肿瘤差异表达基因的结构与功能进行预测则是一条有效的途径。本研究通过生物信息学方法对全基因组进行筛选得到一个差异表达基因 SH2D7,并对 SH2D7 基因及其编码蛋白结构和功能进行了生物信息学分析,为研究 SH2D7 基因的功能及其在肿瘤发生发展中的作用提供了更多的理论依据。

利用 MapViewer 在线分析软件分析显示 SH2D7 定位于人类染色体 15q25.1,该基因包含了 6 个外显子和 5 个内含子,编码含有 451 个氨基酸的蛋白质。用 ExPASy 中的 ProtParam 软件在线预测 SH2D7 编码蛋白分子量为 49.8 kD,理论等电点为 5.99,碱性氨基酸残基(Arg+Lys)所占比例为 49%,酸性氨基酸残基(Asp+Glu)所占比例为 56%,哺乳动物中半衰期为 30 h,为一种不稳定的亲水性蛋白。该蛋白无信号肽,主要定位于细胞核内(73.9%)。蛋白质的磷酸化与去磷酸化是细胞内信号传递的关键步骤,在细胞代谢、生长增殖、癌变调控等方面发挥着重要作用。SH2D7 基因编码蛋白可能含有 35 个磷酸化位点。因此推测该蛋白在细胞信号传导过程中具有重要作用。

蛋白质二级结构中,α-螺旋与β-折叠的主要功能是作为蛋白质骨架起稳定作用,而无规卷曲区域决定蛋白质的功能,并且与抗原位点有关<sup>[8]</sup>。抗原位点多位于亲水性较高、溶剂可及的区域,另外表位的形成还与序列中特定氨基酸的出现、片段的可塑性及二级结构有关。蛋白质中β-转角与无规

卷曲多位于蛋白表面,结构突出,利于与抗体嵌合,形成抗原表位的可能性大。*SH2D7* 编码蛋白二级结构以无规卷曲为主,占90.47%。经生物信息学预测,*SH2D7* 编码蛋白可能存在22个抗原位点。

#### 4 结论

基因通过其编码的蛋白质发挥作用,本研究通过生物信息学分析*SH2D7* 编码蛋白属于SH2信号蛋白家族,包含一个SH2结构域,且易形成抗原位点,可能为抗肿瘤药物的设计提供思路。*SH2D7* 编码蛋白存在多个磷酸化位点,介导蛋白与蛋白相互作用,在细胞内信号传导过程中起到重要作用。其参与信号通路的异常与多种肿瘤的发生密切相关<sup>[9-12]</sup>。除了参与信号传导通路外,*SH2D7* 可以与多种蛋白结合参与转录与转录后调控及翻译后修饰。这些功能的不同取决于蛋白的性质及其配体的磷酸化状态。对*SH2D7* 核苷酸序列及其编码蛋白的结构分析为进一步研究其在肿瘤发生发展中的作用提供了参考。

#### 参考文献(References)

- [1] Orentas R J, Nordlund J, He J, et al. Bioinformatic description of immunotherapy targets for pediatric T-cell leukemia and the impact of normal gene sets used for comparison[J]. *Frontiers in Oncology*, 2014, 4: 134.
- [2] Chu C M, Yao C T, Chang Y T, et al. Gene expression profiling of colorectal tumors and normal mucosa by microarrays meta-analysis using prediction analysis of microarray, artificial neural network, classification, and regression trees[J]. *Disease Markers*, 2014, 2014: 634123.
- [3] Leslie R, O'Donnell C J, Johnson A D. GRASP: Analysis of genotype-phenotype results from 1390 genome-wide association studies and

- corresponding open access database[J]. *Bioinformatics*, 2014, 30(12): i185-i194.
- [4] Park S M, Park S J, Kim H J, et al. A known expressed sequence tag, BM742401, is a potent lincRNA inhibiting cancer metastasis[J]. *Experimental and Molecular Medicine*, 2013, 45: e31.
- [5] Sousa J F, Torrieri R, Silva R R, et al. Novel primate-specific genes, RMEL 1, 2 and 3, with highly restricted expression in melanoma, assessed by new data mining tool[J]. *PLoS One*, 2010, 5(10): e13510.
- [6] Laversin S A, Phatak V M, Powe D G, et al. Identification of novel breast cancer-associated transcripts by UniGene database mining and gene expression analysis in normal and malignant cells[J]. *Genes Chromosomes Cancer*, 2013, 52(3): 316-329.
- [7] Auouacheria A, Navratil V, Barthelaix A, et al. Bioinformatic screening of human ESTs for differentially expressed genes in normal and tumor tissues[J]. *BMC Genomics*, 2006, 7: 94.
- [8] Barlow D J, Edwards M S, Thornton J M. Continuous and discontinuous protein antigenic determinants[J]. *Nature*, 1986, 322(6081): 747-748.
- [9] Liu W Q, Vidal M, Olszowy C, et al. Structure-activity relationships of small phosphopeptides, inhibitors of Grb2 SH2 domain, and their prodrugs[J]. *Journal of Medicinal Chemistry*, 2004, 47(5): 1223-1233.
- [10] Mandal P K, Ren Z, Chen X, et al. Structure-activity studies of phosphopeptidomimetic prodrugs targeting the Src homology 2 (SH2) domain of signal transducer and activator of transcription 3 (Stat3) [J]. *International Journal of Peptide Research and Therapeutics*, 2013, 19 (1): 3-12.
- [11] Chen K F, Tai W T, Chu P Y, et al. STAT3 mediates regorafenib-induced apoptosis in hepatocellular carcinoma[J]. *Clinical Cancer Research*, 2014, 20(22): 5768-5776.
- [12] Sanz A, Niranjani Y, Hammarén H, et al. The JH2 domain and SH2-JH2 linker regulate JAK2 activity: A detailed kinetic analysis of wild type and V617F mutant kinase domains[J]. *Biochimica et Biophysica Acta*, 2014, 1844(10): 1835-1841.

(责任编辑 王媛媛)

#### ·学术动态·



### 第11届“中国青年女科学家奖”在北京颁奖

2015年1月16日,由中华全国妇女联合会、中国科学技术协会、中国联合国教科文组织全国委员会及欧莱雅中国共同主办的第11届“中国青年女科学家奖”颁奖典礼在北京举行。

本届评审委员会由41位院士、专家组成,从111个单位和22名专家推荐、提名的183名候选人中,评选产生了10位获奖者:国家纳米科学中心研究员陈春英、中国科学院上海生命科学院研究员于翔、北京大学信息科学技术学院教授黄如、中山大学肿瘤防治中心研究员贾卫华、华南农业大学亚热带农业生物资源保护与利用国家重点实验室教授廖红、北京大学工学院力学与工程科学系研究员段慧玲、华北电力大学资源与环境研究院教授李永平、中国科学院云南天文台研究员陈雪飞、中国航天员科研训练中心研究员李英贤、西北大学地质系教授刘建妮。

详见中国科协网<http://www.cast.org.cn/n35081/n35096/n10225918/16193089.html>。