

小鼠新基因克隆及功能预测

李其龙[#] 路瑶[#] 刘逸超 张然 马世良^{*}

(沈阳农业大学生物科学技术学院, 沈阳 110866)

摘要 该研究克隆了一个可能与小鼠核酸结合相关的新基因, 并对其表达模式和预期编码蛋白质功能进行了预测分析。该研究采用启动子捕获技术获得了具有启动子活性的基因组片段, 结合生物信息学方法预测得到该启动子活性片段下游基因。采用RT-PCR(reverse transcription PCR)技术扩增得到相应基因, 其cDNA为1 504 bp, 完整开放阅读框(open reading frame, ORF)为540 bp, 编码179个氨基酸。同源性分析表明, 该基因与褐鼠(*Rattus norvegicus*)核酸结合蛋白和黑猩猩(*Pan troglodytes*)RNA结合蛋白的基因序列分别有78%和40%的相似性。利用实时荧光定量PCR技术分析不同组织和不同生理、病理状态下该基因的表达水平。结果表明, 该基因在带瘤小鼠各组织中的表达水平均高于健康小鼠相应组织, 提示该基因可能参与了肿瘤的发生、发展。

关键词 生物信息学; 核酸结合蛋白; 基因克隆; 表达分析

Molecular Cloning and Biochemical Function Prediction of A New Gene from Mouse

Li Qilong[#], Lu Yao[#], Liu Yichao, Zhang Ran, Ma Shiliang^{*}

(College of Biological Science and Biotechnology, Shenyang Agricultural University, Shenyang 110866, China)

Abstract In this study, a new gene coding putative nucleic acid binding protein from mouse was cloned. The gene expression pattern and its predicted protein function had also been discussed. By utilizing promoter-trap technology, a DNA fragment with promoter activity was obtained from mouse genomic DNA. A putative gene located in the downstream of this DNA fragment was predicted by bioinformatics analysis. The predicted gene was 1 504 bp in length with an open reading frame of 540 bp. The predicted open reading frame was confirmed by reverse transcription PCR (RT-PCR). Homology analysis with other species showed that the sequence of this gene shares similarities with the nucleic acid binding protein from *Rattus norvegicus* and RNA binding protein from *Pan troglodytes* of 78% and 40%, respectively. The gene expression properties in diverse tissue types under different physiological status or pathological status were studied by quantitative Real-time PCR (qRT-PCR). As in all the tissue types tested from mice bearing tumor, the expression levels of this gene are much higher than that in healthy mice. These results indicate that this new gene is possibly involved in tumorigenesis.

Keywords bioinformatics; nucleic acid binding protein; gene cloning; expression analysis

收稿日期: 2016-12-01 接受日期: 2017-03-21

辽宁省十百千高端引进人才项目百人层次项目(批准号: 88030312004)和国家级大学生创新训练项目(批准号: 201610157025)资助的课题

[#]共同第一作者

*通讯作者。Tel: 024-88467163, E-mail: msl@syau.edu.cn

Received: December 1, 2016 Accepted: March 21, 2017

This work was supported by the “Financial Support for Selected Researchers Back from Abroad (2012)” from Liaoning Province (Grant No.88030312004) and the Undergraduate Innovation and Entrepreneurship Training Program (Grant No.201610157025)

[#]These authors contributed equally to this work

*Corresponding author. Tel: +86-24-88487613, E-mail: msl@syau.edu.cn

网络出版时间: 2017-04-18 14:07:09 URL: <http://kns.cnki.net/kcms/detail/31.2035.Q.20170418.1407.004.html>

随着高通量测序技术的实际应用范围越来越广,越来越多的新基因被发现和预测。由于基因组的复杂性、预测算法的局限性和测序精准度等仍存在不足^[1],目前已被发现的基因序列仅是庞大基因组序列中的一小部分。2002年,小鼠基因组计划宣布完成,基因组计划数据量为2.50 Gb,占全基因组的96%^[2]。2004年,人类基因组计划宣布完成,基因组计划数据量为2.85 Gb,占全基因组的94%,每100 000个碱基仅有一个错误,正确率达到99.999%。在人类基因组中,编码区仅占全基因组的3%,其余为非编码区^[3]。

虽然人们对非编码区的特性及作用还不清楚,但Riley等^[4]提出,在进化过程中生物不会保留如此大量且无意义的重复序列,这预示着基因组非编码区仍有一些功能未被发现。近年来对“垃圾DNA”的研究发现,在非编码区存在ncRNA、lncRNA等一系列不编码蛋白质的转录序列,可参与并调控基因表达、蛋白质翻译和加工等细胞活动^[5]。目前,基因预测是指将基因组序列通过统计算法(人工神经网络^[6]、隐马尔柯夫模型^[7])与已知的基因结构、同源序列进行比较,从而获得该基因组中可能具有转录活性片段的方法。近年来,通过基因预测方法被发现的新基因有TRPV2(transient receptor potential cation channel, subfamily V, member 2)^[8]、mSHMT(serine hydroxymethyltransferase)^[9]、CORD8(cone-rod dystrophy 8)^[10]等。虽然基因预测模型在查找传统转录序列中发挥了重大作用,但miRNA、piRNA、sRNA等不符合传统编码基因特点的转录序列被发现。这表明目前基因预测算法仍存在局限性和滞后性。虽然近年来越来越多的生物(如拟南芥^[11]、水稻^[12]、小鼠^[13])的全基因组被构建完成,但由于测序准确性不足,仍存在基因组无法被正确拼接和完全覆盖等问题。C57BL/6J小鼠和人类基因组计划虽然取得了巨大的成功,但仍存在数百个测序缺口尚未解决^[14]。由于测序缺口一般具有高度复杂的结构,所以在测序过程中聚合酶链式反应无法正确、高效地进行^[15]。这表明基因组中高度复杂的未知区域的功能还需要基因捕获技术来进一步鉴定。

基因捕获技术可分为增强子捕获(enhancer-trap)、启动子捕获(promoter-trap)和基因捕获(gene-trap)3种方式^[16]。其中,启动子捕获是指将基因组中获得的随机DNA片段插入没有启动子的报告基因

载体中,转染细胞后,根据报告基因表达与否从而推知插入片段是否具有启动子活性^[17]。本研究基于启动子捕获技术,将小鼠乳腺癌4T1细胞基因组随机酶切并与报告基因载体连接,共获得706个含有随机DNA片段的重组载体,转染细胞后通过检测碱性磷酸酶的表达情况,筛选出15个具有启动子活性的片段。

本文将介绍第447号阳性克隆及下游假定核酸结合蛋白基因的结构及不同细胞和组织中的表达规律和其编码蛋白质的理化性质,为进一步探究该基因与乳腺癌的发生关系提供了理论参考。

1 材料与方法

1.1 实验材料

BALB/c小鼠购自长生生物科技股份有限公司;小鼠乳腺癌细胞系4T1为本实验室保存;EasyTaq[®] DNA Polymerase购于北京全式金生物技术有限公司;E.Z.N.A.[®] Gel Extraction Kit购于Omega Bio-tek公司;引物由北京鼎国昌盛生物技术有限公司合成;测序工作由金唯智生物科技有限公司完成;反转录试剂盒购于Thermo公司;QIAquick[®] PCR Purification Kit购于QIAGEN公司。

1.2 启动子捕获

提取转移性的小鼠乳腺癌细胞4T1基因组DNA,使用BamH I和Hind III进行随机酶切,将酶切片段与报告基因载体pCpGfree-basic连接,转化到大肠杆菌(GT115)感受态细胞中,提取质粒构建DNA文库。共获得706个含有随机DNA片段的重组载体。将其转染至4T1细胞中检测报告基因活性,以空载体作为阴性对照。筛选出具有启动子活性的阳性克隆进行测序,确定插入片段的大小和序列,并使用UCSC blat(University of California Santa Cruz The BLAST-Like Alignment Tool)基因组浏览器确认插入片段所在的染色体及位置。

1.3 RNA提取和cDNA合成

培养小鼠乳腺癌4T1细胞至90%汇合度,用磷酸缓冲液(PBS)冲洗3遍,使用Trizol试剂(按INVITROGENE说明书)法提取总RNA,1.0%琼脂糖凝胶电泳检测RNA完整性,NanoDrop定量及 D_{260}/D_{280} 检测纯度。按照M-MLV逆转录试剂盒说明书合成cDNA,取总RNA 3 μ L,反转录引物oligo(dT)₁₈ 1 μ L、5 \times Reaction buffer 4 μ L、RiboLock 1 μ L、10 mmol/L dNTP 2 μ L、

M-MuLV RT 1 μL , 补充ddH₂O至反应总体积20 μL 。反应程序为: 42 $^{\circ}\text{C}$ 60 min, 72 $^{\circ}\text{C}$ 5 min。

1.4 启动子序列分析

使用UCSC提取具有启动子活性片段下游3 000 bp作为启动子预测区, 使用BDGP^[18](<http://www.fruitfly.org/>)中神经网络法预测核心启动子序列, 使用Webgene^[19](<http://www.itb.cnr.it/sun/webgene/>)预测启动子相关元件并对转录因子结合位点进行分析。

1.5 转录序列预测和开放阅读框的克隆

使用FGENESH^[20](<http://www.softberry.com/berry/>)在启动子活性片段下游100 000 bp范围内预测转录序列。根据所预测开放阅读框(open reading frame, ORF)的上下游临近序列设计全长引物。其中, 在起始位点设计上游引物F₄₄₇: 5'-TTA CGG ATC CAT GCA GAA AGA TCC TAA CT-3', 在终止密码子上游设计下游引物R₄₄₇: 5'-ATT GCT CGA GTA TTA TCC TAT GAA GGG CT-3', 扩增长度为537 bp。以小鼠乳腺癌4T1细胞的cDNA为模板, 加入上述特异性引物进行PCR扩增。PCR反应程序为: 94 $^{\circ}\text{C}$ 预变性3 min; 94 $^{\circ}\text{C}$ 变性30 s, 60 $^{\circ}\text{C}$ 退火30 s, 72 $^{\circ}\text{C}$ 延伸1 min, 40个循环; 72 $^{\circ}\text{C}$ 延伸10 min。扩增完后用1.0%琼脂糖凝胶进行电泳检测, 目的片段用E.Z.N.A.[®] Gel Extraction Kit(Omega Bio-tek)回收纯化后进行测序。

1.6 编码蛋白质的功能预测

使用ProtParam(<http://web.expasy.org/protparam/>)软件分析理化性质, NCBI BLAST 2.50(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)程序分析氨基酸序列的同源性, ClustalX 2.1进行多序列比对, Mega 7最大似然法(Maximum likelihood, ML)构建系统进化树, SignalP 4.1(<http://www.cbs.dtu.dk/services/SignalP/>)软件和TMHMM(<http://www.cbs.dtu.dk/services/TMHMM/>)软件分析信号肽与跨膜区, ProtSeal软件分析疏水性, PSORT II(<http://psort.hgc.jp/>)和predictprotein(<https://www.predictprotein.org/>)软件预测亚细胞定位, Moseslab(<http://www.moseslab.csb.utoronto.ca/NLStradamus/>)预测核定位序列。

1.7 组织特异性表达分析

为了进一步研究假定核酸结合蛋白基因(putative nucleic acid binding protein)在移植肿瘤小鼠和正常小鼠各组织中的表达模式, 取小鼠乳腺癌

细胞移植重建后带瘤小鼠的肿瘤组织、肝、肺、脾组织和正常饲养未移植肿瘤小鼠的肝、肺、脾组织为材料提取总RNA, 按照M-MLV反转录试剂盒说明书合成cDNA。为了避免基因组DNA污染, 将20 μL cDNA、10 mmol/L Tris-HCl、5 mmol/L MgCl₂(pH7.4)用10 U DNase I处理2 h, 然后按照QIAquick[®] PCR Purification Kit说明书纯化, 得到15 μL cDNA。

使用Primer 5软件根据ORF序列设计特异性实时定量引物, 并使用Oligo 7软件检测引物是否形成发夹结构和引物二聚体, 预计扩增大小为534 bp。上游引物序列为PNBP-F: 5'-AAG GAT AAT AGG TGT AGG TGA-3', 下游引物序列为PNBP-R: 5'-GTA TAA CAT CTG TCC AGG ATC-3'。以小鼠EF1(epithelial cell produced factor 1)为内参基因, 上下游引物序列分别为EF1a-F: 5'-CGG CCA CCT GAT CTA CAA AT-3'和EF1a-R: 5'-AGC ATG TTG TCA CCA TTC CA-3', 扩增片段长度为534 bp。荧光定量PCR扩增, 设定PCR程序为: 94 $^{\circ}\text{C}$ 预变性3 min; 94 $^{\circ}\text{C}$ 变性30 s, 60 $^{\circ}\text{C}$ 退火30 s, 72 $^{\circ}\text{C}$ 延伸1 min, 40个循环; 72 $^{\circ}\text{C}$ 延伸10 min。每个反应进行3次生物学重复, 采用2^{- $\Delta\Delta\text{Ct}$} 法分析实验结果。

2 结果

2.1 启动子活性片段的捕获和转录因子结合位点预测

447号阳性重组质粒的插入片段位于小鼠19号染色体反义链: 49 132 174~49 132 977, 长度为804 bp。BDGP结果显示, 在插入片段下游2 142 bp位置上含有TATA盒。转录起始点位于插入片段下游2 133 bp, 这与5'RACE实验结果相同。目前, 通过生物信息学预测结果表明, 能与假定核酸结合蛋白启动子区域结合的转录因子为AP1(transformation of anti-microbial protein 1)、AP4(transformation of anti-microbial protein 4)、GATA3(GATA binding protein 3)等。其中, GATA3可与捕获的启动子活性片段结合(图1)。GATA3可参与调控细胞因子和嗜酸性细胞中相关基因的表达^[21-22], 由此推测, 启动子活性片段可能在调控假定核酸结合蛋白(putative nucleic acid binding protein)表达中扮演着重要的角色。

2.2 转录序列预测和开放阅读框架(ORF)扩增

FGENESH结果表明, 位于活性片段下游存在2个预测基因。分别根据2个预测基因序列分别设

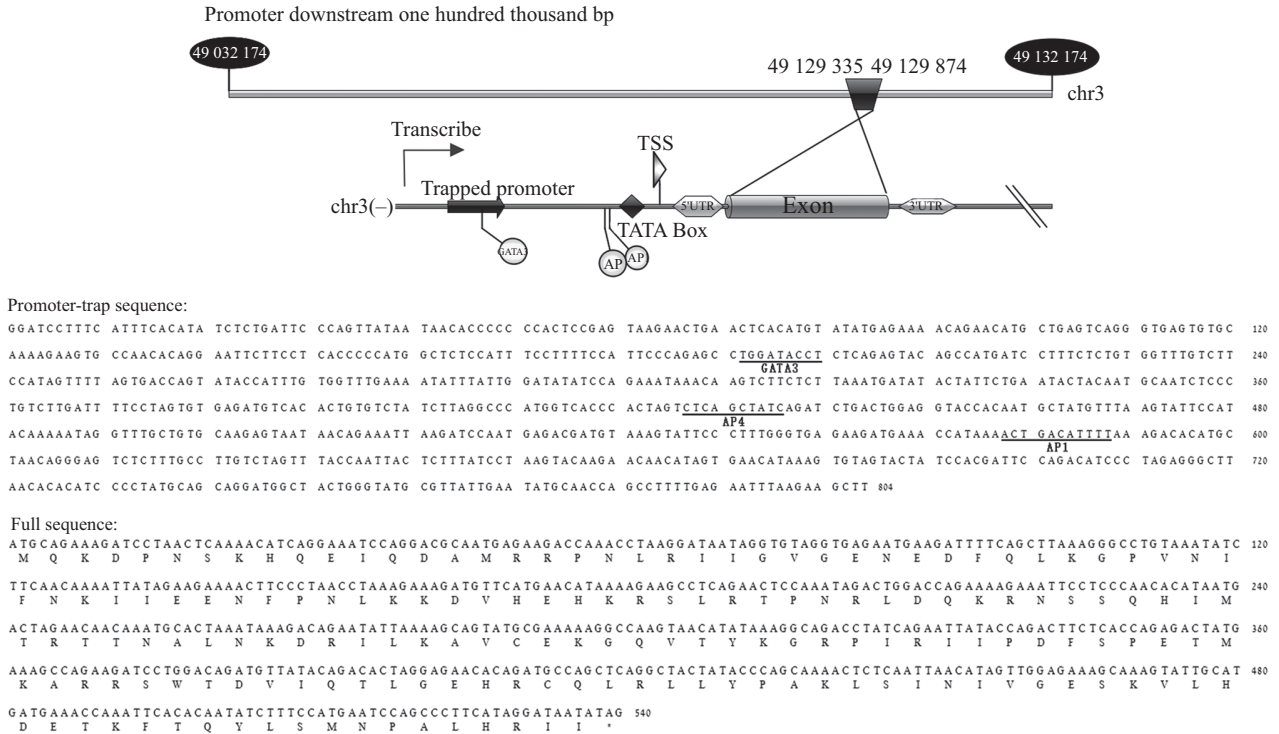


图1 假定核酸结合蛋白基因结构与序列

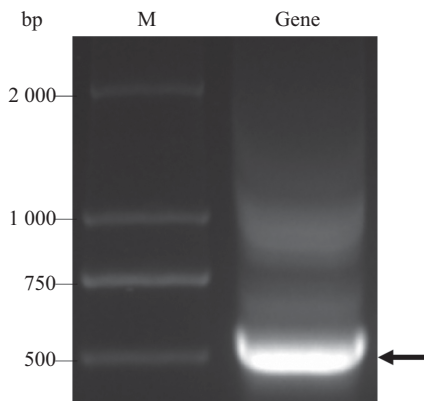
Fig.1 Structure and sequence in genome of putative nucleic acid binding protein

引物对其进行扩增, 但只得到1个扩增片段即假定核酸结合蛋白基因。该片段位于19号染色体反义链: 49 129 335~49 129 874, 全长540 bp, 由一个外显子构成, 共编码179个氨基酸(图1)。使用NCBI blast(2.50)与小鼠已知的mRNA数据库和EST数据库进行比对, 未发现已知的mRNA或EST序列与其完全一致, 由此推定该目的基因尚未被发现和研究。通过PCR扩增获得一条约537 bp的明亮条带(图2),

该条带测序结果与预测序列比对完全一致。

2.3 基因编码蛋白质理化特性和功能预测

ProtParam的结果显示, 假定核酸结合蛋白氨基酸残基数为179, 总原子数为2 970, 原子式为 $C_{916}H_{1501}N_{281}O_{265}S_7$, 带正电荷氨基酸(Arg+Lys)总数为32, 带负电荷氨基酸(Asp+Glu)总数为20, 相对分子质量为20 915.14 Da, 理论等电点为9.96。这表明该蛋白质是碱性蛋白质。假定核酸结合蛋白不稳定系数为48.12, 根据Guruprasad方法^[23]表明其为不稳定蛋白质。ProtSeale的结果表明, 假定核酸结合蛋白的整个氨基酸序列中存在多个疏水性肽段, 且数量明显多于亲水性肽段(图3)。该蛋白质亲水性最强位点为第150位异亮氨酸(分值为1.133), 疏水性最强位点为第6位天冬氨酸和第7位丝氨酸(分值为-3.044), 且存在较多的疏水性较强位点, 表明该蛋白质为疏水性蛋白质, 在细胞中处于非水溶性状态。SignalP 4.1的结果表明, N-端氨基酸的剪切位点分值(C-Value)和综合剪切分值(Y-Value)无明显峰值且均远低于平均值0.5(图4A), 说明该蛋白质N-端不存在剪切位点即不包含信号肽, 推测其为非分泌蛋白质。TMHMM 2.0结果显示, 该蛋白质不含跨膜结构域(图4B), 推测其不是跨膜蛋白。



M: 标志物。箭头所指为目的条带。

M: marker. The target band is indicated with arrow.

图2 小鼠假定核酸结合蛋白基因开放阅读框的克隆

Fig.2 Clone of putative nucleic acid binding protein gene open reading frame (ORF) in *Mus musculus*

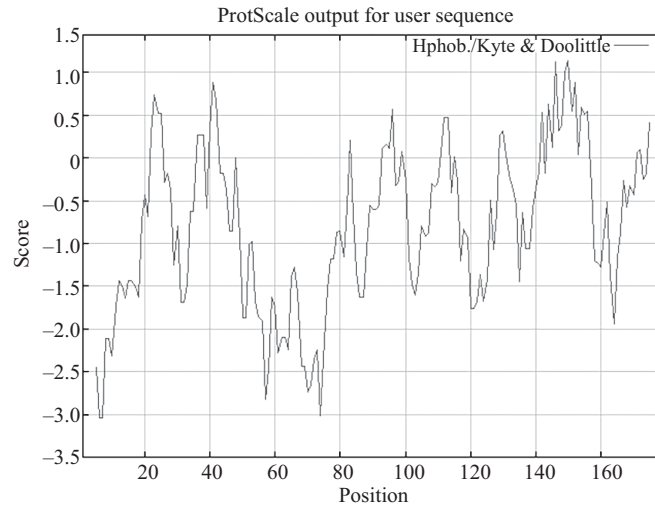


图3 小鼠假定核酸结合蛋白疏水结构分析

Fig.3 Hydrophilicity analysis of putative nucleic acid binding protein in *Mus musculus*

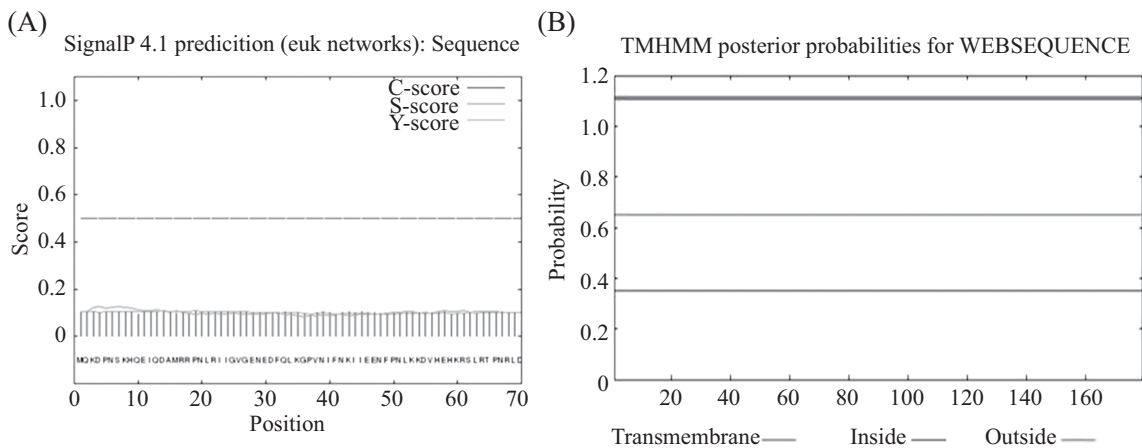


图4 小鼠假定核酸结合蛋白信号肽(A)和跨膜结构(B)预测

Fig.4 Signal peptide prediction (A) and transmembrane region prediction (B) of putative nucleic acid binding protein in *Mus musculus*

PSORT II程序和Predictprotein结果显示,假定核酸结合蛋白亚细胞定位于细胞核,可信度分别为94.1%和29。Moseslab结果表明,假定核酸结合蛋白的氨基酸序列中包含典型的核定位序列(nuclear localization sequence, NLS)特征序列,且包括经典的单基元型NLS,即位于第60位至第64位氨基酸序列为KRSLR。鉴于一般相对分子质量大于45 kDa的蛋白质进入细胞核时才需要自身核定位序列(nuclear localization sequence, NLS)^[24]和特异性转运系统的介导,所以上述氨基酸序列是否真正具有NLS的生物学功能还需进一步的证实。

Predictprotein结果表明,假定核酸结合蛋白分子的功能有:核苷酸结合(nucleotide binding GO:0000166)、有机环状化合物结合(organic cyclic compound binding GO:0097159)、核苷酸磷酸盐结合(nucleoside

phosphate binding GO:190126)及参与基因表达调控(regulation of gene express GO:0010468)。以上结果表明,假定核酸结合蛋白可能与核苷酸的磷酸基团结合,并具有调节基因表达的功能。

2.4 同源蛋白质比对及进化树的构建

NCBI BLAST 2.50和ClustalX 2.1的结果表明,假定核酸结合蛋白与褐鼠(*Rattus norvegicus*)核酸结合蛋白相似度为78%,其N-端氨基酸序列相对保守;与黑猩猩(*Pan troglodytes*)RNA结合蛋白(RNA-binding protein)和人类(*Homo sapiens*)p40相似度皆为40%,且C-端氨基酸序列相对保守(图5)。采用Mega 7最大似然法构建系统进化树(图6),结果表明,假定核酸结合蛋白与褐鼠的核酸结合蛋白和黑猩猩RNA结合蛋白均不聚类,说明小鼠假定核酸结合蛋白与二者亲缘关系较远,为1种新型的核酸结合蛋白。

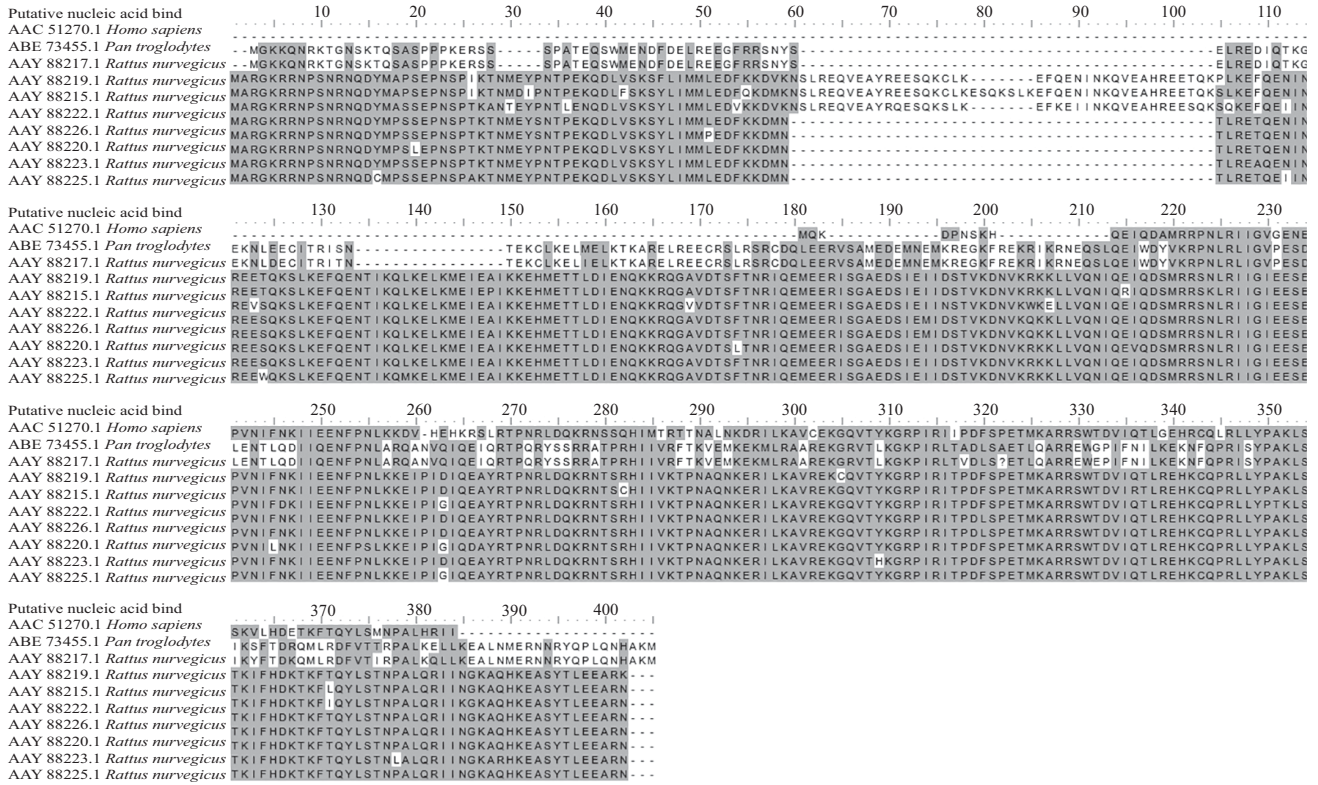


图5 小鼠假定核酸结合蛋白同源性分析

Fig.5 Homology analysis of putative nucleic acid binding protein in *Mus musculus*

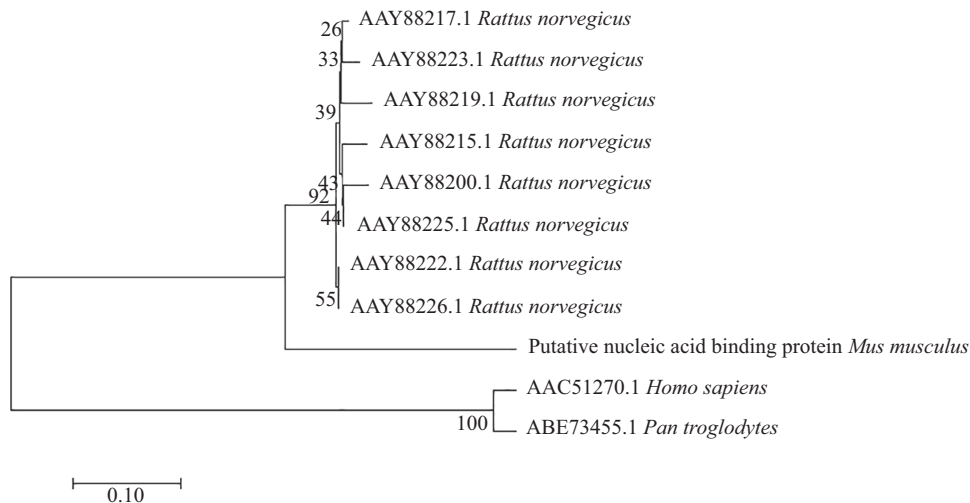


图6 小鼠假定核酸结合蛋白与其他物种的核酸结合蛋白、RNA结合蛋白氨基酸序列的进化树

Fig.6 Phylogenetic analysis of putative nucleic acid binding protein in *Mus musculus* and nucleic acid binding protein, RNA binding protein from other species

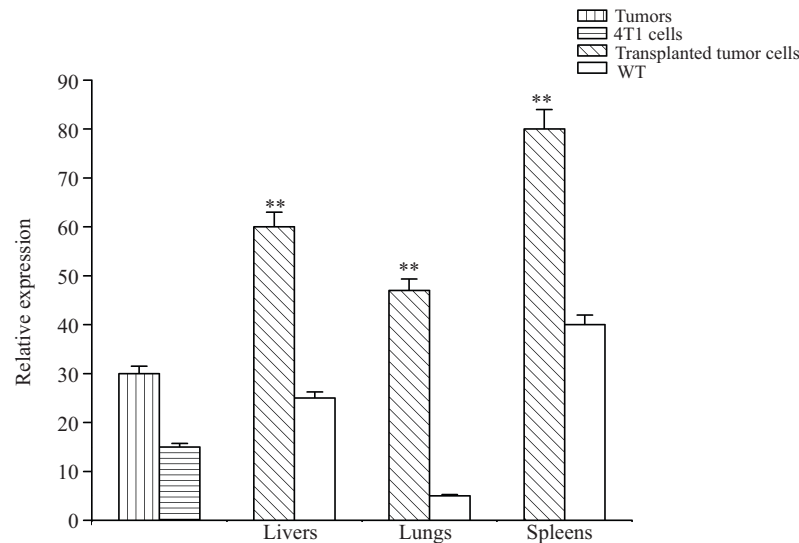
2.5 组织特异性表达

取4T1小鼠乳腺癌细胞移植重建后带瘤小鼠的肝、肺、脾、肿瘤组织和正常小鼠的肝、肺、脾组织, 以小鼠*EF1*为内参基因, 采用荧光定量PCR技术检测小鼠不同组织、不同生理状态下的假定核酸结合蛋白基因表达水平(图7)。结果显示, 在小鼠肝、肺、脾组织中假定核酸结合蛋白基因均表达, 并且该基

因在肿瘤组织中表达水平平均高于正常组织, 且移植脾中表达水平最高, 这可能与细胞增殖相关。由于该基因在正常组织和癌症组织中均有表达, 表明其表达具有组成性特点。

3 讨论

基因远距离控制是指顺式作用元件(沉默子、



该图表示假定核酸结合蛋白基因在小鼠肿瘤、4T1细胞及正常和病理情况下的肝、肺、脾组织中的表达水平,以 $EF1$ 为内参基因。每个样品3次取样,每个反应3个重复。** $P < 0.01$,与WT组比较。

The measurements represent the expression level of livers, lungs, spleens in *Mus musculus* bearing tumors and WT, and 4T1 cells. The level of each gene is relative to that of $EF1$. Three sampling each sample, three technical replicates each reaction. ** $P < 0.01$ compared with WT group.

图7 不同组织中假定核酸结合蛋白基因的表达

Fig.7 Gene expression of putative nucleic acid binding protein in different tissues of *Mus musculus*

增强子、隔离子等)参与组织和发育特定基因或肿瘤疾病的表达调控^[25]。迄今为止,已发现多种远距离调控基因,例如转录因子 $FOXL2$ (forkhead box L29)、 $FOXP3$ (forkhead box P3)等^[26]。Babraham研究所和伦敦Francis Crick研究所的研究人员运用Hi-C技术研究发现,在人类参考基因组上有近22 000个启动子存在长距离相互作用^[27]。这些远距离调控的DNA序列的微小变化会干扰基因的正常控制,可能导致发展成为一种特定的疾病。本研究捕获的启动子活性片段距离TATA盒2 142 bp,距离假定核酸结合蛋白基因2 301 bp,属于远距离启动子。但该片段具体功能还需构建报告载体进一步研究。

原核生物基因组小而简单,基因组中只含有单外显子基因且全部参与编码。而真核生物的基因组大而复杂,存在着许多不参与编码蛋白质的序列,例如重复序列、间隔序列、非编码RNA等^[28]。在真核生物(如小鼠)中,可根据外显子数目将基因类型分为2种,1种是只含有1个外显子的单外显子基因,其约占总基因数的15.8%^[29];另1种是含有1个外显子以上的多外显子基因,所占比例远远超过单外显子基因。在进化过程中,大多数单外显子基因只能编码特定蛋白质,导致生物因无法应对环境的改变而逐渐被淘汰。而在真核生物细胞中,多外显子基因可通过可变剪切的方式形成不同的转录本以应对环境

改变^[30],从而被保留。本研究克隆的假定核酸结合蛋白基因仅由单一外显子组成,表明该基因可能来源于远古单核细胞或进化过程中病毒的插入^[31]。

核酸结合蛋白是一类含有DNA或RNA结合域的结合蛋白^[32]。其中,DNA结合蛋白(DNA binding protein, DBP)(含有DNA结合域)包括调节转录过程的转录因子、聚合酶、核酸酶和参与细胞核中染色体包装和转录的组蛋白等。DNA结合蛋白中一般含有锌指、螺旋-转角-螺旋和亮氨酸拉链等结构,这些结构域可促进蛋白质与核酸结合^[33]。RNA结合蛋白(RNA binding protein, RBP)是在细胞中结合双链或单链RNA并参与形成核糖核蛋白复合物(ribonucleoprotein complexes)的蛋白质。RBPs含有各种结构模体(motif),如RNA识别模体(RNA recognition motif, RRM)、dsRNA结合结构域、锌指结构等^[34]。RBPs在多种细胞过程中具有关键作用,如细胞功能、转运和定位。尤其在RNA转录后控制中扮演着重要的角色,如剪接、多聚腺苷酸化、mRNA稳定化、mRNA定位和翻译等。人类p40蛋白是由 $TP63$ (tumor protein p63)基因编码的转录因子,它含有典型DNA结合域^[35-36],表明它具有与DNA结合的作用。该基因存在多种剪切方式,能形成不同的转录本,翻译成多种蛋白质,例如p63、p51、p40等。此外,它在多种肿瘤中均表达,

例如乳腺癌^[37]、肺癌^[38]、前列腺癌^[39]等。检索THE HUMAN PROTEIN ATLAS^[40]数据库结果显示, 该蛋白存在于人类多种癌细胞系中, 例如人膀胱移行细胞乳头瘤细胞RT4、人乳腺癌细胞MCF7等。同源比对和生物信息学结果表明, 在细胞核内假定核酸结合蛋白可能参与了核酸的结合, 并调控基因表达。这与目前已知核酸结合蛋白的功能一致, 但具体功能还需进一步实验证明。

实时荧光定量PCR结果显示, 该基因在移植组织中表达水平普遍高于正常组织, 表明可能与肿瘤的转移相关, 且在肝、肺和脾中存在显著差异。这表明假定核酸结合蛋白可能与肿瘤发生有关并参与移植后肿瘤细胞的迁移和分裂。

本研究成功克隆了一个可能与小鼠核酸结合相关的基因, 并对其组织特异性表达结果进行了分析。通过生物信息学方法预测编码蛋白质功能, 为进一步阐明假定核酸结合蛋白的功能及在小鼠肿瘤发生中的作用和调控机制奠定了理论基础。

参考文献 (References)

- Blazewicz J, Glover F, Kasprzak M. DNA Sequencing—Tabu and Scatter search combined. *INFORMS* 2004; 16(3): 232-40.
- Consortium MGS, Waterston RH, Lindbladtoh K, Birney E, Rogers J, Abril JF, *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; 420(6915): 520-62.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409(6822): 860-920.
- Riley DE, Jeon JS, Krieger JN. Simple repeat evolution includes dramatic primary sequence changes that conserve folding potential. *Biochem Biophys Res Commun* 2007; 355(3): 619-25.
- Huang JF, Guo YJ, Zhao CX, Yuan SX, Wang Y, Tang GN, *et al.* Hepatitis B virus X protein (HBx)-related long noncoding RNA (lncRNA) down-regulated expression by HBx (Dreh) inhibits hepatocellular carcinoma metastasis by targeting the intermediate filament protein vimentin. *Hepatology* 2013; 57(5): 1882-92.
- Fernández-Navarro F, Hervás-Martínez C, Ruiz R, Riquelme JC. Evolutionary Generalized Radial Basis Function neural networks for improving prediction accuracy in gene classification using feature selection. *Applied Soft Computing* 2012; 12(6): 1787-800.
- 马宝山, 朱义胜. 基于隐马尔科夫模型的基因预测算法. 大连海事大学学报(Ma Baoshan, Zhu Yisheng. Gene-prediction algorithm based on hidden Markov model. *Journal of Dalian Maritime University*) 2008; 34(4): 41-4.
- Lee JY, Shim WS. Cloning of *Xenopus laevis* TRPV2 by gene prediction. *Genomics & Informatics* 2005; 3(1): 24-9.
- 刘庆坡, 冯英, 董辉. 水稻线粒体丝氨酸羟甲基转移酶基因的电子克隆. 生物信息学[Liu Qingpo, Feng Ying, Dong Hui. In silico cloning of mSHMT gene from rice (*Oryza sativa* L.). *Bioinformatics*] 2005; 3(1): 5-9.
- Khattak NA, Rauf S, Sehgal SA, Shahwar D, Ismail M. In silico gene prediction for autosomal recessive cone-rod dystrophy 8 (CORD8). *JBIC* 2013; 14(2): 104-17.
- Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, *et al.* Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* 2000; 402(6763): 761-8.
- Sasaki T, Burr B. International rice genome sequencing project: the effort to completely sequence the rice genome. *Curr Opin Plant Biol* 2000; 3(2): 138-42.
- Bradley A. Mining the mouse genome. *Nature* 2002; 420(6915): 512-4.
- Eichler EE, Clark RA, She X. An assessment of the sequence gaps: Unfinished business in a finished human genome. *Nat Rev Genet* 2004; 5(5): 345-54.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, *et al.* The sequence of the human genome. *Science* 2001; 291(5507): 1304-51.
- 李元元, 张靖溥. 基因捕获技术及其最新进展(英文). 遗传学报(Li Yuanyuan, Zhang Jingpu. Gene trapping techniques current progress. *Journal of Genetics and Genomics*) 2006; 33(3): 189-98.
- Lindsey K, Wei W, Clarke MC, Mcardle HF, Rooke LM, Topping JF. Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. *Transgenic Res* 1993; 2(1): 33.
- Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, *et al.* The BDGP gene disruption project. *Genetics* 2004; 167(2): 761-81.
- Buendia F, Cano JC. Webgene: A generative and web-based learning architecture to teach operating systems in undergraduate courses. *IEEE Xplore* 2006; 49(4): 464-73.
- 张胜利, 李东方, 张改生, 王军卫, 牛娜. 应用软件Fgenesh预测水稻基因. 中国农业科学(Zhang Shengli, Li Dongfang, Zhang Gaisheng, Wang Junwei, Niu Na. Prediction of rice gene by fgenesh. *Scientia Agricultura Sinica*) 2008; 41(6): 1567-74.
- 谢婷婷, 王芳, 赵卫东. GATA3与恶性肿瘤的相关性研究进展. 山东医药(Xie Tingting, Wang Fang, Zhao Weidong. Research progress on the correlation between GATA3 and malignant tumor. *Shandong Medical Journal*) 2008; 48(19): 149-50.
- 涂军伟, 朱以军, 朱寿田, 朱丹, 陈慧, 赵建平, 等. 大鼠肺纤维化转录因子T-bet/GATA3的表达及其与相关细胞因子的关系. 浙江医学(Tu Junwei, Zhu Yijun, Zhu Shoutian, Zhu Dan, Chen Hui, Zhao Jianping, *et al.* Expression of transcription factors T-bet/GATA3 and its association with other cytokines related to pulmonary fibrosis in rat. *Zhejiang Medicine*) 2010; 32(11): 1621-4.
- 郭宇立, 倪逸声. 从一级结构预测蛋白质稳定性: Curuprasad, Reedy和Pandit方法. 生物化学与生物物理进展(Guo Yuli, Ni Yisheng. Prediction of protein stability from primary structures: Curuprasad, reedy and pandit methods. *Prog Biochem Biophys*) 1993; 20(2): 95-9.
- 赵元茵, 王元忠, 曹念, 周度金, 李渝萍. 核定位信号及其分析策略. 中国生物化学与分子生物学报(Zhao Yuanyin, Wang Yuanyin, Cao Nian, Zhou Dujin, Li Yuping. Categories and research strategies of nuclear localization signal. *Chinese Journal*

- of Biochemistry and Molecular Biology) 2009; 25(8): 683-9.
- 25 王玉明, 李亚. 组织和发育特异性基因的远距离转录调控. 生命的化学(Wang Yuming, Li Ya. Tissue and development of specific genes for long-term transcriptional regulation. *Chemistry of Life*) 2007; 27(2): 151-3.
- 26 Thackray VG. Fox tales: Regulation of gonadotropin gene expression by forkhead transcription factors. *Mol Cell Endocrinol* 2014; 385(1/2): 62-70.
- 27 Kind J, Pagie L, de Vries SS, Nahidiar L, Dey SS, Bienko M, *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* 2011; 163(1): 134.
- 28 林元山. 原核生物和真核生物的基因组及其基因表达调控的比较研究. 怀化学院学报(Lin Yuanshan. A comparative study on regulation of genome and gene expression of prokaryotes and eukaryotes. *Journal of Huaihua University*) 2002; 21(5): 39-44.
- 29 Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, *et al.* A physical map of the mouse genome. *Nature* 2002; 418(6899): 743-50.
- 30 Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: Diversification, exon definition and function. *Nat Rev Genet* 2010; 11(5): 345-55.
- 31 于丹丹, 汲广东. HMGB基因的结构功能与演化. 鲁东大学学报(自然科学版)[Yu Dandan, Ji Guangdong. Structural function and evolution of HMGB gene. *Journal of Ludong University (Natural Science Edition)*] 2016; 32(1): 38-43.
- 32 Lacasse EC, Lefebvre YA. Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res* 1995; 23(10): 1647-56.
- 33 王锐, 胡晓愚. DNA结合蛋白结构域的研究进展. 化学进展 (Wang Rui, Hu Xiaoyu. Progress in research on DNA-binding protein domains. *Progress in Chemistry*) 1995; 7(4): 277.
- 34 杜光伟, 周严, 袁建刚, 强伯勤. RRM RNA结合蛋白的结构与功能. 生物化学与生物物理进展(Du Guangwei, Zhou Yan, Yuan Jiangang, Qiang Boqin. Structure and function of RRM RNA-binding protein. *Biochemistry and Biophysical Advances*) 1999; 26(4): 305-07.
- 35 Strano S, Munarriz E, Rossi M, Castagnoli L, Shaul Y, Sacchi A, *et al.* Physical interaction with Yes-associated protein enhances p73 transcriptional activity. *J Biol Chem* 2001; 276(18): 15164-73.
- 36 Chi SW, Ayed A, Arrowsmith CH. Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain. *EMBO J* 1999; 18(16): 4438-45.
- 37 Papageorgis P, Ozturk S, Lambert A W, Neophytou CM, Tzatsos A, Wong CK, *et al.* Targeting IL13Ralpha2 activates STAT6-TP63 pathway to suppress breast cancer lung metastasis. *Breast Cancer Res* 2015; 17(1): 98.
- 38 Hu QY, Jin TB, Wang L, Zhang L, Geng T, Liang G, *et al.* Genetic variation in the TP63 gene is associated with lung cancer risk in the Han population. *Tumor Biol* 2014; 35(3): 1863-6.
- 39 Parsons JK, Saria EA, Nakayama M, Vessella RL, Sawyers CL, Isaacs WB, *et al.* Comprehensive mutational analysis and mRNA isoform quantification of TP63 in normal and neoplastic human prostate cells. *Prostate* 2009; 69(5): 559-69.
- 40 Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 2014; 13(2): 397.