

DOI: 10.3724/SP.J.1005.2008.00352

小鼠睾丸特异表达基因 *TSEG-1* 的克隆及序列分析

顾朝辉¹, 童强松¹, 曾甫清¹, 刘媛¹, 王智宇¹, 郑丽端², 蔡嘉斌¹, 蒋国松¹

1. 华中科技大学同济医学院附属协和医院外科, 武汉 430022;
2. 华中科技大学同济医学院附属协和医院病理科, 武汉 430022

摘要: 从表达序列标签(expressed sequence tags, ESTs)数据库 ZooDDD 中获得小鼠正常睾丸表达的 EST, 通过 dbEST 数据库检索出与其高度同源的 EST 序列, 构建 EST 叠加群(contigs), Biolign 软件拼接, GeneScan 软件预测 contigs 对应的基因组序列中的外显子、内含子; 针对开放阅读框设计引物序列, 采用 RT-PCR 从小鼠睾丸组织中克隆新基因的 cDNA, 分析该基因在小鼠各脏器中的 mRNA 表达, 并对测序结果进行生物信息学分析。结果表明: 在小鼠 X 染色体的 1 668~2 011 kb 间克隆出一新基因 *TSEG-1*, 全长为 510 bp, 开放阅读框为 336 bp, 编码 111 氨基酸, 分子量 12.84258 kDa, 等电点 11.4000。RT-PCR 证实该基因开放阅读框正确, 在小鼠睾丸组织中特异性表达, 且与小鼠其他 cDNA 无同源性, 获得 GenBank 登录号 EU079024。功能区分析发现 *TSEG-1* 蛋白可能为一种跨膜蛋白, 跨膜区位于第 41~61 氨基酸残基。*TSEG-1* 基因与人类睾丸特异性组蛋白 2 α 变异体基因有较高同源性, 在 *TSEG-1* 基因 5'-端非编码侧翼预测发现存在 1 个启动子区域, 范围为 680 bp。 *TSEG-1* 蛋白可能有 4 个抗原性位点, 2 个特异性蛋白激酶的磷酸化位点, 其亚细胞定位可能位于线粒体。小鼠睾丸特异性基因 *TSEG-1* 的克隆为进一步研究其生物学功能和表达调控奠定了基础。

关键词: 小鼠; 睾丸; 表达序列标签; *TSEG-1*; 生物信息学

Cloning and sequence analysis of *TSEG-1*, a novel gene specifically expressed in mouse testis

GU Chao-Hui¹, TONG Qiang-Song¹, ZENG Fu-Qing¹, LIU Yuan¹, WANG Zhi-Yu¹, ZHENG Li-Duan², CAI Jia-Bin¹, JIANG Guo-Song¹

1. Department of Surgery, Union Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China;
2. Department of Pathology, Union Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

Abstract: The expressed sequence tags (ESTs) of normal mouse testis were obtained from online EST database ZooDDD. Their highly homologous EST sequences were found through the dbEST database to construct contigs, and spliced by the biomedical software Biolign. The corresponding exons and introns within genome sequences were predicted by software GeneScan. According to the open reading frame, the primers were designed. RT-PCR was applied in the cloning of novel

收稿日期: 2007-09-15; 修回日期: 2007-12-18

基金项目: 国家自然科学基金资助项目(编号: 30200284, 30600278, 30772359)和教育部新世纪优秀人才支持计划 [Supported by the National Natural Science Foundation of China (No. 30200284, 30600278, 30772359) and Program for New Century Excellent Talents in University]

作者简介: 顾朝辉(1982-), 男, 河南人, 博士生, 研究方向: 男科学、泌尿系肿瘤。Tel: 13437108099; E-mail: zgwhgch@163.com

通讯作者: 童强松 (1974-), 男, 湖北人, 副教授, 副主任医师, 医学博士, 研究方向: 小儿泌尿系发育及其调控。Tel: 027-63776478; E-mail: qs_tong@hotmail.com

gene from mouse testis and analyzing its expression pattern in various mouse tissues. The bioinformatics analysis on the sequencing results of *TSEG-1* was conducted. Results indicated that a novel gene *TSEG-1* was cloned from 1 668–2 011 kb of mouse X chromosome, with full-length sequence of 510 bp. The open reading frame (ORF) is 336 bp in length and encodes a deduced amino acid sequence of 111 residues. The molecular weight of TSEG-1 protein is 12.84258 kDa, and its pI is 11.4000. RT-PCR demonstrated the correctness of its ORF. *TSEG-1* was distinctively expressed in testis, but not in other tissues of mouse. No obvious homology with other mouse cDNA was found for *TSEG-1*. The GenBank accession number EU079024 was achieved. It was predicted that TSEG-1 is a kind of transmembrane protein, and the transmembrane domain is from 41 amino acid residue to 61 amino acid residue. Blastn analysis revealed its high homology to human testis-specific gene *H2AX*. Computational prediction of the 5'-untranslated region of *TSEG-1* gene revealed a 680 bp-length promoter region. There are four antigen binding sites and two phosphorylation sites of specific protein kinase in TSEG-1 protein, with subcellular localization in mitochondria. The cloning of mouse testis specific gene *TSEG-1* laid a foundation for subsequent research of its biological function and expression regulation.

Keywords: mouse; testis; expressed sequence tags; *TSEG-1*; bioinformatics

世界卫生组织资料显示, 西方国家有 5%~8% 的育龄夫妇可能有不育问题, 发展中国家的某些地区可高达 30%, 我国男性不育的发生率呈逐年增高趋势^[1]。近年来, 研究初步表明遗传缺陷是导致男性不育的重要原因, 而精原细胞的发育受多种基因的调控, 克隆与精原细胞发育相关的基因已成为当前的研究热点^[2]。在探索精原细胞发育分子机制的过程中, 已有不少重要基因如睾丸特异性 RNA 结合蛋白(RBM)、NYD-SP6 和雄激素受体(AR)等被相继克隆^[3]。然而已知基因间、已知基因与未知基因在精原细胞发育过程中的调控机制仍不清楚。因此, 克隆新的睾丸发育、精原细胞发育相关基因既可为阐明精原细胞的分化机制提供理论基础, 又能为在基因水平上探索精子生成障碍的发生机制提供实验依据。

如何寻找未知功能的新基因、克隆不同时空差异表达和疾病相关的新基因是人类后基因组时代亟待解决的问题。目前, 基于表达序列标签(expressed sequence tags, ESTs)的电子克隆(*in silicon* cloning)已成为发现未知基因的重要手段^[4]。本研究基于 EST 电子克隆的原理, 检索并下载睾丸组织特异性 EST 序列的同源性 EST 序列簇, 通过序列拼接、BLAST 同源性比较、GeneScan 预测等生物信息学方法, 并结合 RT-PCR 技术, 从正常小鼠睾丸中发现了一个睾丸特异性表达基因 *TSEG-1*, 并对其核酸和蛋白质的特性进行了初步分析, 旨在为进一步探讨 *TSEG-1* 基因的生物学功能奠定基础。

1 材料和方法

1.1 材料

成年昆明小鼠, 体重 25~35 g, 由华中科技大学同济医学院动物中心提供。TRI reagent RNA 抽提试剂购自美国 MRC(Molecular Research Center)公司; RevertAidTM First Strand cDNA Synthesis 和 PCR 试剂盒购自 Fermentas 公司; PCR 扩增引物由上海 Invitrogen 公司合成; DNA Marker (600 bp) 购自 Tiangen 生物技术公司; PCR 产物的回收试剂盒购自 Omega 公司。

1.2 方法

1.2.1 EST 数据库的分析

在 EST 二级数据库 ZooDDD (<http://bio301.iis.sinica.edu.tw/~ZooDDDNew/main.php>)^[5] 的 DB1, DB2 中检索小鼠睾丸组织特异性表达的 EST 簇交集 (BY076573.1、BY706589.1、BY706899.1、BY707047.1、BY707115.1、BY707245.1、BY715086.1、CB274190.1), 采用 Biolign 软件的 Assembly 程序对 EST 片段进行拼接和校对, 再将该片段在美国国家生物信息中心 (NCBI) 的小鼠 EST 和 nr 数据库中进行 BLAST 分析, 构建该基因的 EST 重叠群 (contig)。将拼接的重叠群与小鼠基因组进行 BLAST (basic local alignment search tool) 分析, 截取同源性较高基因组核酸序列, 采用 GeneScan 程序 (<http://genes.mit.edu/GENSCAN.html>) 预测基因的外显子、内含子, 确定 cDNA 序列。

1.2.2 含完整编码区 cDNA 的克隆

采用逆转录聚合酶链反应(RT-PCR)。参照一步法 TRIzol RNA 抽提说明书提取小鼠睾丸组织总 RNA。逆转录反应体系: 模板 RNA 2 μg 、10 mmol/L dNTP 1 μL 、RNA 酶抑制剂(RNasin) 20 U、Oligo, dT₁₈ 1 μL 、AMV 逆转录酶 200 U、5 \times 逆转录酶缓冲液 5 μL 、终体积为 25 μL , 42 $^{\circ}\text{C}$ 水浴 30 min。采用 Primer Premier 5.0 软件设计 PCR 扩增引物序列: 上游引物 5'-ATGGCCAGGAAAAGGCAAAG-3', 下游引物 5'-TCAGTTGTCATCAGGTTC-3'; 同时以 GAPDH(314 bp) 作为内参照: 上游引物 5'-GCCAAGGTCATCCATGACAACCTTTGG-3'; 下游引物 5'-GCCTGCTTCA CCACCTTCTTGATGTC-3', 由上海 Invitrogen 生物技术有限公司合成。采用梯度 PCR 进行扩增: 94 $^{\circ}\text{C}$ 预变性 5 min, 94 $^{\circ}\text{C}$ 变性 1 min, 50~65 $^{\circ}\text{C}$ 复性 1 min, 72 $^{\circ}\text{C}$ 延伸 1 min, 35 个循环后 72 $^{\circ}\text{C}$ 再延伸 10 min。PCR 产物经 1.5% 琼脂糖凝胶电泳, MGIAS-1000 凝胶成像系统扫描拍照并定量。

1.2.3 DNA 序列分析

采用柱状离心法回收 PCR 产物, 参见说明书进行。回收产物在 ABI37723 自动测序仪上测序, 由大连宝生物工程有限公司完成, 与已知重叠群进行拼接, 用 ExPASy 服务器中的 Translate 和 NCBI 中的 ORF finder 进行阅读框架的识别。利用最新公布的小鼠基因组工作草图, 确定碱基序列, 进行染色体定位。

1.2.4 基因表达分析

采用断颈法快速处死成年昆明小鼠, 取大脑、心脏、肺脏、胃、肝脏、脾脏、小肠、肾脏、膀胱、前列腺、睾丸、骨骼、肌肉组织, 迅速置于液氮中, 按照上述方法抽提各组织总 RNA, RT-PCR 检测新基因在小鼠各脏器中的表达特点。

1.2.5 序列同源性和功能区的计算机分析

通过 NCBI 的 Blastn、Blastp 程序实现核酸/氨基酸序列的相似性比较。采用在线分析软件 SeqFacts (<http://bip.weizmann.ac.il/sqfbin/seqfacts>) 对 TSEG-1 蛋白质的等电点、分子量、抗原性位点进行预测, 通过 ExPASy 数据库的生物学预测软件对蛋白质的疏水区域和跨膜区域进行分析预测。联网至日本大板大学基因组信息研究中心数据库 (<http://psort.hgc.jp/>) 对蛋白质序列进行亚细胞定位分析。运用丹麦科技大学在线分析程序 NetPhos 2.0 Server 和 NetPhos 1.0 Server (<http://www.cbs.dtu.dk/services/>) 预测磷酸化位点。采用 Genomatix (<http://www.genomatix.de>) 的分析软件

Gene2promoter 进行启动子预测。

1.3 统计学方法

采用 *t* 检验, 数据用均数士标准差($\bar{X} \pm S$)表示, 运用 SPSS 10.0 统计学软件进行数据分析。

2 结果

2.1 TSEG-1 基因 cDNA 全序列的克隆及验证

以 ZooDDD 数据库中 DB1、DB2 的小鼠睾丸组织特异性表达的 EST 簇交集为模板进行杂交, 得到了 BX634402.1、CN832581.1、CN832782.1、CN835097.1、CN838816.1、CN841421.1、CJ042329.1、AA065475.1、AA144729.1 共 9 个具有高度相似性的 EST, 并将其组装成 EST 重叠群。经基因组 BLAST 分析和 GeneScan 程序预测, 得到全长为 510 bp 的基因序列, 开放阅读框为 336 bp, 由单一外显子构成。根据开放阅读框, 设计 PCR 引物, 以小鼠睾丸 cDNA 为模板进行梯度 PCR 扩增, 得到的扩增产物为 336 bp (图 1), 双向测序结果 (图 2) 与电子克隆序列完全一致, 获得 GenBank 登陆号: EU079024。

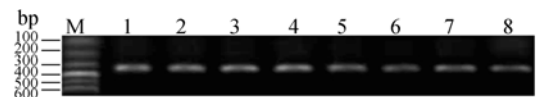


图 1 梯度 RT-PCR 从小鼠睾丸组织中扩增 TSEG-1 基因的 cDNA 序列 (336 bp)

M: DNA Marker 600 bp; 1: 49.6 ; 2: 50.9 ; 3: 52.3 ; 4: 53.8 ; 5: 55.2 ; 6: 56.6 ; 7: 58.1 ; 8: 59.4 .

Fig. 1 Cloning of TSEG-1 cDNA (336 bp) from mouse testis by gradient RT-PCR

M: DNA Marker 600 bp; 1: 49.6 ; 2: 50.9 ; 3: 52.3 ; 4: 53.8 ; 5: 55.2 ; 6: 56.6 ; 7: 58.1 ; 8: 59.4 .

2.2 TSEG-1 阅读框的证实及其在小鼠多种组织中的表达

用 ExPASy 服务器中的 Translate 和 NCBI 中的 ORF finder 进行阅读框架的识别, 结果表明 RT-PCR 技术克隆了 TSEG-1 基因全长 cDNA。抽提小鼠大脑、心脏、肺脏、胃、肝脏、脾脏、小肠、肾脏、膀胱、前列腺、睾丸、骨骼、肌肉组织 RNA 进行 RT-PCR, 发现仅在睾丸组织中存在一个 336 bp 的特异性产物, 而其他组织中未见表达 (图 3), 同时看家基因 GAPDH 在各组织中均有表达。将在睾丸组织中特异表达片段测序, 测序结果与克隆结果完全一致。以 TSEG-1 基因的 cDNA 序列检索 Unigene 数据库, 获得编号为 Mn.23887 的 Unigene, 电子表达谱

```

GGAGAGCC TGGTACCAAC CACAGAGGGT CAGCTGAGCA AAAACCACCA CACAAGCATC
ATGGCCAGGA AAAGGCAAAG GCGAAGAAGA CGGAAAGTGA CCCGCTCCCA GAGAGCTGAG
M A R K R Q R R R R R K V T R S Q R A E
CTTCAATTTC CTGTTAGCCG TGTGGATCGT TTCCTACGAG AGGAAACTA TTCTAGACGC
L Q F P V S R V D R F L R E G N Y S R R
CTGAGCTCTT CTGCTCCTGT ATTCTCTCGG GGTGTGCTCG AGTACTTAAC ATCTAACATC
L S S S A P V F L A G V L E Y L T S N I
CTTGAAGTGG CTGGTGAGGT GGCCACACCC ACTGGCAGGA AGCGCATAGC TCCAGAGCAT
L E L A G E V A H T T G R K R I A P E H
GTGTGCCGTG TGGTACAGAA CAACGAACAG CTCCACCAAC TCTCAAACA AGGTGGCACA
V G R V V Q N N E Q L H Q L F K Q G G T
TCAGTGTGTTG AGCCACCAGA ACCTGATGAC AACTGA TGATGCCTGG GTCTCGAGTC
S V F E P P E P D D N
TGGAGGTTCA CCGCACTTTC AGCCTTACCC AGCTCCAATG TCCTACTAGC TGGGGGAGCT
AATCAAACAT GCTATTAAAG CATTACATC GAACCC-

```

图2 *TSEG-1* 基因的 cDNA 序列和编码的氨基酸序列

TSEG-1 基因全长 510 bp, 开放阅读框为 336 bp, 编码 111 个氨基酸, 分子量 12.84258 kDa, 等电点 11.4000。

Fig. 2 cDNA and amino acid sequences of *TSEG-1* gene

The full length of *TSEG-1* gene is 510 bp. The open reading frame is 336 bp, encoding 111 amino acid residues. The theoretical molecular weight is 12.84258 kDa, and the theoretical pI is 11.4000.

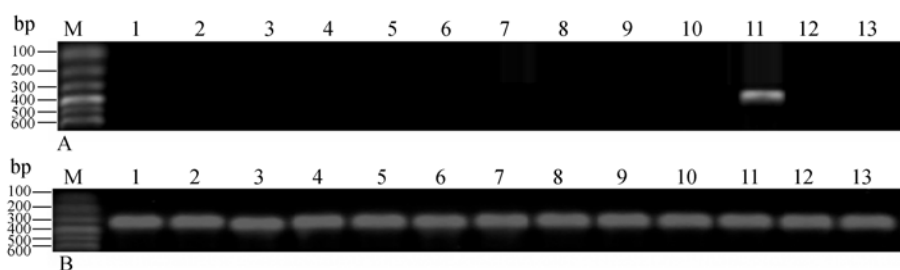


图3 RT-PCR 检测 *TSEG-1* 基因在小鼠多组织中的表达

A: *TSEG-1* 基因阅读框在小鼠不同组织中 RT-PCR 扩增产物; B: 看家基因 *GAPDH* 在小鼠不同组织中 RT-PCR 扩增产物。1: 脑; 2: 心; 3: 肺; 4: 肝脏; 5: 脾脏; 6: 胃; 7: 肠; 8: 肾脏; 9: 膀胱; 10: 前列腺; 11: 睾丸; 12: 骨骼; 13: 肌肉; M: DNA Marker 600 bp。

Fig. 3 Detection of *TSEG-1* expression in mouse tissues by RT-PCR

A: RT-PCR product of *TSEG-1* in mouse tissues; B: RT-PCR products of *GAPDH* in mouse tissues. 1: Brain; 2: Heart; 3: Lung; 4: Liver; 5: Spleen; 6: Stomach; 7: Intestine; 8: Kidney; 9: Bladder; 10: Prostate; 11: Testis; 12: Skeleton; 13: Muscle; M: DNA Marker 600 bp.

分析显示: 该基因仅在小鼠睾丸组织中有表达, 验证了本实验结果。

2.3 *TSEG-1* 基因结构和同源性比较

将 *TSEG-1* 基因 cDNA 序列与小鼠基因组草图 (www.ncbi.nlm.nih.gov/mouse genome blast/) 进行相似性比较, 发现该基因包含于小鼠第 2 染色体上的 2qA3 带, 分布于 17 918 165~17 918 500 kb 碱基区域内。*TSEG-1* 基因与大鼠组蛋白变异体 2 α 基因有 86% 同源性, 与人类睾丸特异性组蛋白 2 α 变异体基因有 59% 同源性, 后二者分别位于 X 染色体 25 565~25 665 kb 和 36 598~36 698 kb 之间。

2.4 *TSEG-1* 编码蛋白的生物信息学分析

Blastp 程序证实 *TSEG-1* 蛋白氨基酸序列与大鼠组蛋白变异体 2 α 蛋白 53% 同源, 与人类睾丸特异性组蛋白变异体 2 α 蛋白 50% 同源。通过 SeqFacts 程序预测小鼠 *TSEG-1* 蛋白理论等电点 pI 为 11.4000, 相对分子量 MW 为 12.84258 kDa。*TSEG-1* 蛋白有 4 个抗原性位点, 分别位于第 77~88 氨基酸之间, 第 40~70 氨基酸之间, 第 19~31 氨基酸之间, 第 91~96 氨基酸之

间(图 4)。利用 ExPASy 的 ProtScale、TMPred 软件, 分别对 *TSEG-1* 蛋白进行疏水性分析和跨膜区分析, 显示该基因表达的蛋白在第 43 位氨基酸、第 63 位氨基酸属于疏水区, 富含疏水性氨基酸(图 5)。NetPhos 2.0 Server 预测 *TSEG-1* 编码蛋白有以下基序: 5 个丝氨酸磷酸化位点(位于氨基酸残基 16、26、38、42 和 101),

- (1) Score 1.222 length 12 at residues 77~88
Sequence: APEHVCRRVQNN
77 88
- (2) Score 1.185 length 31 at residues 40~70
Sequence: RLSSAPVFLAGVLEYLTSNILELAGEVAHT
40 70
- (3) Score 1.117 length 13 at residues 19~31
Sequence: AELQFPVSRVDRF
19 31
- (4) Score 1.094 length 6 at residues 91~96
Sequence: LHQLFK
91 96

图4 SeqFacts 软件对 *TSEG-1* 蛋白的抗原性位点分析
第 19 至 31 氨基酸残基、第 40 至 70 氨基酸残基、第 77 至 88 氨基酸残基、第 91 至 96 氨基酸残基为抗原性位点。

Fig. 4 Prediction of antigenic sites of *TSEG-1* protein via SeqFacts program

The antigenic sites are located at amino acid residues 19~31, 40~70, 77~88 and 91~96, respectively.

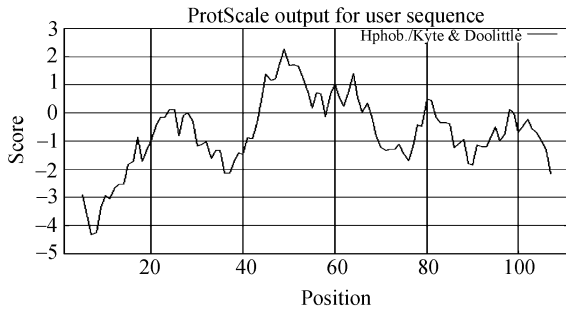


图 5 ProtScale 软件对 TSEG-1 蛋白的疏水性分析
第 43 位、第 63 位氨基酸属于疏水区, 富含疏水性氨基酸。

Fig. 5 The hydrophobic cluster analysis of TSEG-1 protein via ProtScale program

Hydrophobic domains are located at amino acid residues 43 and 63, which are rich in hydrophobic amino acids.

1 个酪氨酸磷酸化位点(位于氨基酸残基 37), 1 个苏氨酸磷酸化位点(位于氨基酸残基 71), 见图 6; NetPhos 1.0 Server 预测特异性蛋白激酶的磷酸化位点: PKA 位于氨基酸残基 42 位(0.84), PKC 位于氨基酸残基 71 位

(0.84), 图 7。预测该蛋白含有一个跨膜区, 位于第 41~61 位氨基酸, 跨膜区螺旋的长度约 17 到 35 氨基酸(图 8), 经 PSORT WWW Server 分析可能位于线粒体。

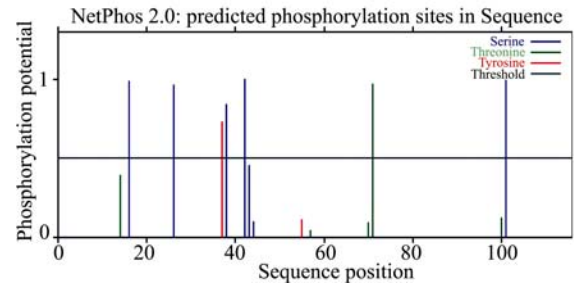


图 6 NetPhos 2.0 Server 软件对 TSEG-1 蛋白的磷酸化位点分析

5 个丝氨酸(serine) 磷酸化位点, 1 个酪氨酸(threonine)磷酸化位点, 1 个苏氨酸(tyrosine)磷酸化位点。

Fig. 6 Prediction of phosphorylation sites of TSEG-1 protein via NetPhos 2.0 Server program

The numbers of phosphorylation sites of serine, threonine and tyrosine are 5, 1 and 1, respectively.

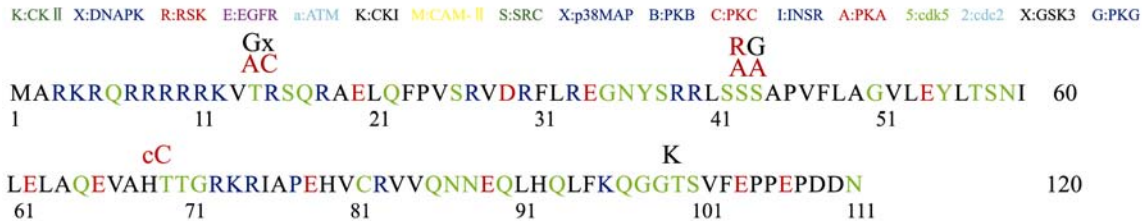


图 7 NetPhos 1.0 Server 软件对 TSEG-1 蛋白特异性蛋白激酶磷酸化位点分析

PKA 位于氨基酸残基 42 位(0.84), PKC 位于氨基酸残基 71 位(0.84)。

Fig.7 Prediction of kinase specific phosphorylation sites of TSEG-1 protein via NetPhos 1.0 Server program

The kinase specific phosphorylation sites of PKA and PKC are located at amino acid residues 42 and 71, respectively.

Sequence Length: 111
Sequence Number of predicted TMHs: 0
Sequence Exp number of AAs in TMHs: 0.66653
Sequence Exp number, first 60 AAs: 0.57111
Sequence Total prob of N-in: 0.94526
Sequence TMHMM2.0 inside 1 111

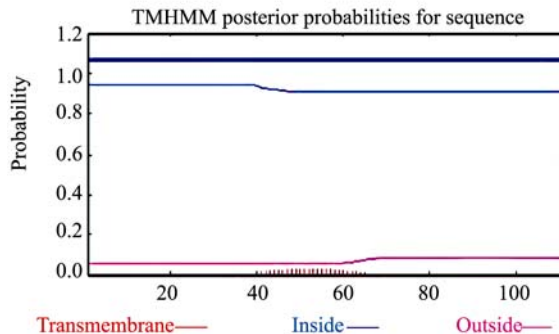
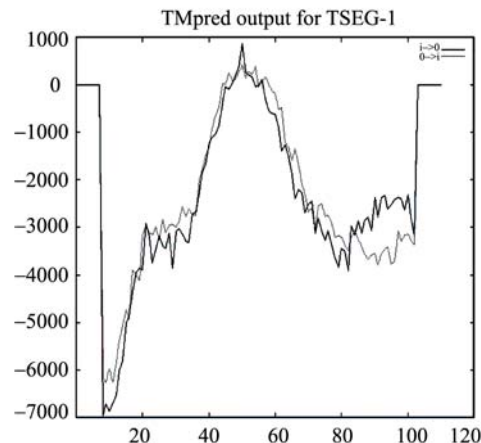


图 8 TSEG-1 蛋白的跨膜区分析

预测显示 TSEG-1 的跨膜区位于第 41 至第 61 氨基酸序列, 亚细胞定位于线粒体的可能性为 60.9%。

Fig. 8 Prediction of transmembrane domains of TSEG-1 protein

The transmembrane helices of TSEG-1 are located between 41 and 61 amino acid residue, and its subcellular localization may be in mitochondria (60.9% possibility).



2.5 *TSEG-1* 基因启动子位点分析

将 *TSEG-1* 基因编码序列前 5 000 bp 基因组序列输入 Genomatix (www.genomatix.de) 的启动子

分析软件 Gene2promoter 进行检索, 发现存在 1 个可能性大于 85% 的启动子, 区域范围为 680 bp, 见图 9。

(-) strand elements (5'-3')						
Type of element	Start	End	Strand	Length	Name	Annotation/Analysis
Promoter region	17914521	17915200	-	680 bp	GXP_163933	Pormoter for GXT_21840535, GXT_13001342, GXT_13001343, GXT_13001346, GXT_13001352, GXT_13001355, GXT_13007329, GXT_13007808, GXT_13093068(1700113017Rik/GXL_136444) Analyze Promoter
						Extraction <input type="checkbox"/>

图 9 *TSEG-1* 基因启动子预测

通过启动子分析软件 Gene2promoter 进行检索, 发现存在 1 个启动子区域, 范围为 680 bp。

Fig. 9 Promoter prediction of *TSEG-1* gene

One promoter region, 680 bp in length, was predicted for *TSEG-1* gene by Gene2promoter program.

3 讨论

男性不育的发生率逐年增高越来越受到生殖医学界的重视, 研究睾丸的发育及其生精过程的发生机制是寻求男性不育病因的有效手段。然而, 影响睾丸发育的因素很多, 其生精过程受多基因的调控^[6]。由精原细胞分化为成熟的精子是一个极其复杂的过程, 多种基因在此过程中扮演重要角色。人类研究表明, 约有 2 000 个不同的基因参与男性生殖发育及其调节过程, 包括睾丸发育、生殖细胞分化、减数分裂及精子的发生各个阶段, 参与发育或调控的基因可位于性染色体, 也可定位于常染色体^[7]。因此, 克隆和研究睾丸新基因的功能及其与已知基因间的相互作用, 已成为进一步揭示男性生精过程及生精障碍分子机制的关键所在。

随着功能基因组学研究的深入, 电子克隆已成为寻找新基因的重要手段之一^[8], 即以数学模型为核心, 通过计算机和互联网, 寻找同源性EST簇, 构建包含这些EST的重叠群, 通过序列拼接克隆出新基因的全长cDNA, 从而加速对基因组中未知功能新基因的发现, 已成为功能基因组学和蛋白质组学研究的新策略^[9]。与传统电子克隆方法不同^[10], 本研究在构建EST重叠群后进行了基因组BLAST分析, 找出对应的基因组序列, 避免了EST的漏测和误测的缺点, 同时运用GeneScan程序预测基因的外显子、内含子, 能更清晰地解析基因的开放阅读框。研究结果表明, 获得的小鼠新基因*TSEG-1* 全长 510 bp, 开放阅读框为 336 bp, RT-PCR及测序结果验证了开放阅读框的准确性, 同时序列分析表明该基因具有完整基因的各项特征。此外我们发现,

TSEG-1 mRNA在小鼠睾丸组织中具有高度特异性, 这为将来进行睾丸靶向生物治疗性药物的开发提供了理论依据。通过核酸数据库的序列同源性比较, 发现小鼠*TSEG-1* 基因与大鼠组蛋白变异体 2 α 基因^[11] 86%同源, 与人类睾丸特异性组蛋白变异体 2 α 基因^[12] 59%同源, 但在对果蝇基因组进行检索后, 未发现相似性序列, 说明该基因在小鼠、大鼠、人类中非常保守, 可能在高等生物进化中起重要作用。近年来人们发现, 组蛋白变异体 2 α 蛋白在细胞周期进程中起重要作用, 被视为基因组的监视器, 它的磷酸化形式又称为 γ -组蛋白变异体 2 α , 能募集一系列DNA损伤修复蛋白和信号分子, 在DNA损伤识别、损伤信号传导及DNA修复过程中发挥关键性作用。研究显示: 组蛋白变异体 2 α 蛋白能使姐妹染色单体重组效率增加 4 倍, 组蛋白变异体 2 α 基因缺失型精母细胞在X-Y染色体配对时表现严重缺陷, 减数分裂配对障碍, 提示性染色体上的组蛋白变异体 2 α 基因在生殖发育过程中起重要作用^[13, 14]。我们推测, *TSEG-1* 可能具有类似组蛋白变异体 2 α 基因的生物学功能。利用计算机软件进行*TSEG-1* 的功能区分分析, 发现*TSEG-1* 为一种跨膜蛋白, 提示*TSEG-1* 可能在细胞信号传导中起一定的作用, 从而介导睾丸发育和精子的发生。

目前, 该基因在睾丸发育、精子发生和受精中的确切功能仍未明确, 需进一步研究 *TSEG-1* 基因的基因组结构、时空表达特点、启动子区域和转录调控因子, 可通过基因转染、反义或 RNA 干扰等技术研究其生物学功能, 为阐明睾丸发育的内在机理奠定基础。

参考文献(References):

- [1] Turek PJ, Pera RA. Current and future genetic screening for male infertility. *Urol Clin North Am*, 2002, 29(4): 767–792. [\[DOI\]](#)
- [2] Schlecht U, Demougin P, Koch R, Hermida L, Wiederkehr C, Descombes P, Pineau C, Jegou B, Primig M. Expression profiling of mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility. *Mol Biol Cell*, 2004, 15(3): 1031–1043. [\[DOI\]](#)
- [3] Orwig KE, Shinohara T, Avarbock MR, Brinster RL. Functional analysis of stem cells in the adult rat testis. *Biol Reprod*, 2002, 66 (4): 944–949. [\[DOI\]](#)
- [4] Bouck A, Vision T. The molecular ecologist's guide to expressed sequence tags. *Mol Ecol*, 2007, 16(5): 907–924. [\[DOI\]](#)
- [5] Chen YC, Hsiao CD, Lin WD, Hu CM, Hwang PP, Ho JM. ZooDDD: a cross-species database for digital differential display analysis. *Bioinformatics*, 2006, 22(17): 2180–2182. [\[DOI\]](#)
- [6] Meyerhof W, Muller B R, Richter D. Molecular cloning of a novel putative G-protein coupled receptor expressed during rat spermiogenesis. *FEBS Lett*, 1991, 284(2): 155–160. [\[DOI\]](#)
- [7] Adams IR, McLaren A. Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis. *Development*, 2002, 129(5): 1155–1164.
- [8] Gill RW, Sanseau P. Rapid *in silico* cloning of genes using expressed sequence tags (ESTs). *Biotechnol Annu Rev*, 2000, 5: 25–44. [\[DOI\]](#)
- [9] QIAN Jun, DONG Li, ZHANG Bi-Cheng, WANG Jie-Ru, ZHOU Ming, LI Zhong-Hua, LI Wei-Fang, LI Xiao-Ling, LI Gui-Yuan. Identification and digitalized expression analysis of murine *UBAP1* gene by means of EST database searching. *Prog Biochem Biophys*, 2002, 29(2): 323–327.
- [10] ZHANG De-Li, LI Yan-Da, JI Liang. Correction of five different types of errors of model refseqs appeared in NCBI human gene database only by using two novel human genes c17orf32 and znf362. *Acta Genetica Sinica*, 2004, 31(4): 325–334.
- [11] Fernandez-Capetillo O, Chen HT, Celeste A, Ward I, Romanienko PJ, Morales JC, Naka K, Xia Z, Camerini-Otero RD, Motoyama N, Carpenter PB, Bonner WM, Chen J, Nussenzweig A. DNA damage induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nat Cell Biol*, 2002, 4(12): 993–997. [\[DOI\]](#)
- [12] Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*, 2001, 15(18): 2343–2360. [\[DOI\]](#)
- [13] Fernandez-Capetillo O, Mahadevaiah SK, Celeste A, Romanienko PJ, Camerini-Otero RD, Bonner WM, Manova K, Burgoyne P, Nussenzweig A. H2ax is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. *Dev Cell*, 2003, 4(4): 497–508. [\[DOI\]](#)
- [14] Xie A, Puget N, Shim I, Odate S, Jarzyna I, Bassing CH, Alt FW, Scully R. Control of sister chromatid recombination by H2AX. *Mol Cell*, 2004, 16(6): 1017–1025. [\[DOI\]](#)

•遗传咨询•

遗传性乳光牙本质是怎样遗传的?

问：我母亲患有乳光牙，遗传给了我。现在我到了生育年龄，怎样才能生育正常健康的宝宝？

答：遗传性乳光牙本质又称遗传性牙本质发育不全，俗称大黄牙，是一种常见的口腔疾病，儿童发病率约为 1/7000。其主要表现为牙本质的钙化异常和缺陷，牙釉质发育正常，但易从牙本质表面分离脱落，使牙本质外露而致牙冠磨损。患者牙齿萌出时，其形态大小可正常，但牙冠呈微黄色半透明，并逐渐变成乳光色或浅黄色外观。

本病一般表现为常染色体显性遗传方式，患者子女患病概率为 1/2。*DSPP* 基因的突变是导致该病的原因。由于牙齿是在出生后出现，可以利用产前基因诊断方法判断胎儿是否有致病基因及出生后是否发病。

(中国科学院遗传与发育生物学研究所 李巍)