

包含 BTB 锌指结构的新基因 Bsg2 结构特性 及其在发育阶段的组织表达特异性

杨 玮¹⁾, 岳凌生¹⁾, 姚艺桑¹⁾, 邹 星¹⁾, 慈宏亮^{1),2)}, 陈 微³⁾, 李亦平^{1),2),3)}*

¹⁾北京师范大学生命科学学院生物医学研究所, 北京 100875; ²⁾浙江赛尔生物医学研究院, 杭州 310006;

³⁾美国哈佛大学牙医学院 Forsyth 研究所, 波士顿 02115

摘要 通过消减差异筛选法寻找小鼠胚胎发育过程中在脑中特异表达的基因. 克隆得到的脑特异表达新基因 2 (brain specific gene 2, 简称 Bsg2) 长 3 691 bp, 通过生物信息学方法预测其编码一个含 713 个氨基酸的锌指蛋白. 此蛋白 N 端有一个 BTB (BR-C, ttk and bab) 结构域, C 端有 9 个连续的 C2H2 锌指结构. 该基因定位在小鼠 12 号染色体上, 包含 1 个内含子和 2 个外显子. 应用生物信息学和 RT-PCR 方法分别检验该基因在小鼠各组织中的表达. 结果表明, Bsg2 基因在小鼠胚胎及成体的各组织中普遍表达, 在脾、肾、睾丸、肠、子宫和脑的表达水平较强. 利用整体 (whole mount) 原位杂交研究其时空表达模式. 结果显示, Bsg2 在早期的小鼠胚胎和不同时期鸡胚的头部均特异表达, 在 11 d 鼠胚的肢芽里也有较强的表达. Bsg2 基因的结构和表达特征预示它编码 1 个具有 DNA 结合功能的转录调控因子, 同时揭示它在脑的发育和器官形成过程中发挥着重要作用.

关键词 生物信息学, BTB 结构域, C2H2 锌指, 原位杂交

中图分类号 Q78, Q593

Structural Characteristic and Special Expression Pattern During Development of a Novel Gene Bsg2 Containing BTB and Zinc Finger Domain

YANG Wei¹⁾, YUE Ling-sheng¹⁾, YAO Yi-sang¹⁾, ZOU Xing¹⁾, CI Hong-liang^{1),2)}, CHEN Wei³⁾, LI Yi-ping^{1),2),3)}*

¹⁾ Biomedical Research Institute, College of Life Sciences, Beijing Normal University, Beijing 100875, China; ²⁾ Zhejiang Cell Biomedical Research College, Hangzhou 310006, China; ³⁾ The Forsyth Institute, Harvard School of Dental Medicine, Boston, MA 02115, USA

Abstract The subtractive-different screening was used to clone a new gene Bsg2 expressed specifically in brain during the development of mouse embryo. The length of the gene is 3 691 bp. Bioinformatics analysis predicted that it encoded a C2H2 type zinc finger protein of 713 amino acid residues. The predicted protein contains a BTB/POZ domain in its N-terminal region and 9 typical C2H2 zinc fingers in its C-terminal region. The gene was located on mouse chromosome 12 and comprised of 1 intron and 2 exons. The mRNA levels of this gene were determined in mouse's tissues using bioinformatics and semi-quantitative RT-PCR assay. Bsg2 was expressed in various tissue of mouse embryo, especially in the spleen, liver, testis, uterus and brain. The results tested by *in situ* hybridization indicated that Bsg2 was specifically expressed in mouse brain of early stages and chicken brain of different stages, and expressed in the limbs of E11 (11 days) mouse embryo. The structure and expression pattern of Bsg2 implied that it might be a DNA binding transcription regulator. All the results indicated that Bsg2 gene might be involved in the development of brain and organic development.

Key words bioinformatics, BTB domain, C2H2 zinc finger, *in situ* hybridization

人体最复杂的器官是脑. 与其它器官相比, 脑具有独特的基因表达谱, 并且在发育的不同阶段、同一发育时期的不同部位、同一部位在不同环境下, 脑基因的表达都有所不同. 因此, 研究脑的特异基因将为了了解脑的早期发育、神经系统的构建奠定基础, 为防

收稿日期: 2003-12-19, 接受日期: 2004-02-16

浙江赛尔生物医学研究院资助项目

*联系人 Tel: 010-58806656, E-mail: YPLi @forsyth.org

Received: December 19, 2003; Accepted: February 16, 2004

Supported by Foundation of Zhejiang Cell Biomedical Research College

*Corresponding author Tel: 010-58806656, E-mail: YPLi @forsyth.org

治脑发育异常引起的疾病提供理论基础. 在人类基因组中, 脑特有基因占很大比例, 但由于多数基因的表达丰度都很低, 用常规方法研究比较困难, 所以人们对脑中特异表达的基因知之甚少. 本文利用消减差异筛选法, 建立小鼠胚胎不同时期脑部的基因表达谱, 以寻找小鼠胚胎发育过程中在脑中特异表达的基因. 消减差异筛选法曾成功应用于组织蛋白酶 K^[1]、破骨细胞特异性质子泵亚基 ATP6i^[2] 基因的克隆.

本文寻找到小鼠胚胎发育过程中脑特异表达基因 *Bsg2*, 此基因含有 BTB 和 C2H2 锌指结构. 锌指蛋白是指含有通过结合 Zn^{2+} 形成稳定的、可自我折叠成“指头”样结构的一类蛋白质, 是识别 DNA 最有效的结构元件之一, 并广泛参与基因表达的调控, 在细胞分化、配子形成和胚胎发育等生命过程中发挥重要作用, 且与许多重大疾病有密切关系. C2H2 型锌指蛋白含有最为典型的锌指模体结构, 其 N 末端通常伴有不同的基元, 其中包括 POZ (poxvirus and zinc finger) 结构域, 也称为 BTB 结构域. BTB/POZ 家族是类 Kuppel 锌指蛋白家族中的一种, 这个家族的第一批成员包括果蝇的 Broad-Complex、tramtrack、bric-a-brac 等, 因此被命名^[3-6]. BTB 结构域位于 5 端, 大约有 115 个氨基酸, 大多为疏水性氨基酸; 其二级结构由 4 段螺旋组成, 螺旋间由折叠连接, 所形成的三维空间结构是与其它蛋白作用的特征模体. BTB/POZ 结构域能介导同源二聚体的形成, 以及蛋白质之间的相互作用, 抑制 DNA 结合结构域与 DNA 序列的结合^[7]; 改变染色质的结构, 从而引起染色质结构的改变, 导致特定基因的开关^[8]. 目前所知的多数含 BTB/POZ 结构域的锌指蛋白均表现出转录抑制活性, 此类蛋白质对于生物体的发育、分化、染色体重组等过程的调节具有相当重要的作用, 并与癌症有密切的关系.

1 材料和方法

1.1 材料

健康小鼠购于北京大学医学部实验动物中心; M-MLV 逆转录酶购于 Invitrogen; 高保真的 PFU DNA 聚合酶购于宝生物工程大连有限公司; 地高辛标记试剂盒购于 Roche; 所有引物均由北京博润公司合成; QIAquick Gel Extraction Kit 购于 QIAGEN; Trizol 购于鼎国公司; pBluescript SK (-) 载体购于 Stratagene; 其它试剂均为国产分析纯产品.

1.2 消减差异筛选脑特异表达基因

消减差异筛选法参考文献 [2] 的方法. 收集 9.5 d 小鼠胚胎的头部, 提取 mRNA, 制备 ³²P 标记的 cDNA 探针, 并与生物素标记的 9.5 d 小鼠胚胎的躯干 mRNA 进行杂交. 除去杂交复合物后, 用 7.5 d 小鼠胚胎的 cDNA 文库进行筛选. 所得到的克隆用 ABI 377 测序仪测序.

1.3 用生物信息学方法对 *Bsg2* 基因序列特征进行分析

将所得序列在美国国家生物技术信息中心 (National Center of Biotechnology Information, NCBI) 上与已知序列进行比对, 确证所得序列为一新序列. 预测 *Bsg2* 的编码框, 并对其进行基因组定位. 将 *Bsg2* 与人、鸡的同源蛋白进行多序列比对. 用 ExPASy Proteomics^[9] 工具 (<http://us.expasy.org/tools>) 对 *Bsg2* 蛋白的结构域及其可能的功能、蛋白特点进行预测.

1.4 基因电子表达谱分析

利用 UniGene 数据库进行电子表达谱分析, 用与待分析核酸序列同源的表达序列标签 (expressed sequenced tags, EST) 序列所对应的组织来源进行推断.

利用肿瘤基因组解剖学计划 (The Cancer Genome Anatomy Project, CGAP) 的虚拟 Northern 服务器进行电子表达谱分析.

1.5 半定量 RT-PCR 表达分析

1.5.1 引物设计 用 Primer Premier 软件根据 *Bsg2* 的 cDNA 序列设计一对引物, 产物为 594 bp. 两引物序列为:

P1 (sense) : 5'-GTCCTTICCGATGICCTA-3'

P2 (antisense) : 5'-TTTGGTGTCTGCCTACC-3'

选用看家基因 *GAPDH* 作为内参照, 产物为 452 bp, 两引物序列为:

P1 (sense) : 5'-ACCACAGTCCATGCCATCAC-3'

P2 (antisense) : 5'-TCCACCACCCTGTTGCTGTA-3'

1.5.2 RT-PCR 以小鼠各组织的 RNA 为模板, 合成第一链 cDNA. 利用 *Bsg2* 及 *GAPDH* cDNA 特异的引物, 扩增 30 个循环 (退火温度分别为 55、60).

1.5.3 RT-PCR 表达分析 1.5% 琼脂糖凝胶电泳判定产物大小, 并半定量分析 (目的基因片段和 *GAPDH* 产物电泳上样量一致). 利用凝胶成像系统将图像扫描入计算机后, 用 TotalLab 软件分析电泳条带光密度值. 计算公式如下: mRNA 表达相对量 = 目的基因片段条带的光密度值 / 同一样本 *GAPDH* 条带的光密度值.

1.6 RT-PCR 克隆 *Bsg2* 的全长编码区序列

根据 *Bsg2* 的 cDNA 序列设计一对引物,在 5 端和 3 端引物分别加入 *EcoR*、*Kpn* 酶切位点.两引物序列为:

P1 (sense):

5'-TAGAATTCGCA GAA GATGCGCAAAG CCCA-3'

P2 (antisense):

5'-TCGGTACCTTTTGGTGT TCTGCC-3'

以小鼠胚胎的总 RNA 为模板,经 M-MLV 逆转录酶合成 cDNA 第一条链.用高保真的 PFU DNA 聚合酶扩增 30 个循环(退火温度为 55 °C).纯化和回收 PCR 产物后,分别将纯化的产物和 pBluescript SK (-)载体进行 *EcoR* 和 *Kpn* 的双酶切,经纯化后,将 PCR 产物连接至载体上.转化连接产物,筛选重组子,制备质粒 DNA,经测序验证 *Bsg2* 全长编码区序列.

1.7 原位杂交

1.7.1 探针的制备 以小鼠 *Bsg2* 编码区的 cDNA 为模板,用地高辛核酸标记试剂盒制备地高辛标记的 RNA 探针.分别用不同时期的小鼠胚胎和鸡胚进行整体原位杂交.

1.7.2 原位杂交 原位杂交参照文献[10]中的方法.将固定好的胚胎经不同浓度梯度的甲醇脱水、复水后,用 6% 的 H₂O₂ 漂白胚胎,10 μg/ml 蛋白酶 K/PTW 处理胚胎 15 min,在甘氨酸终止蛋白酶 K 作用后,用 0.2% 戊二醛/4% 多聚甲醛再次固定. PTW 清洗 2 次后,用含甲酰胺等的预杂交液于 60 °C 温育 3~8 h,然后用含探针的杂交液于 60 °C 过夜.用含不同离子浓度和不同 pH 值的梯度缓冲液反复洗涤,以排除非特异性的结合,加入碱性磷酸酶偶联的抗地高辛抗体,最后以硝基四唑盐/5-溴-4-氯-3-吡啶磷酸盐(NBT-BCIP)显色.实验过程中同时作阴性对照,阴性对照组不加探针,其他处理条件相同.

2 结果

2.1 消减差异筛选脑特异表达基因

利用消减差异筛选法得到 *Bsg2* 基因.测序后,将该基因在 NCBI 上进行 BLAST (Basic Local Alignment Search Tool)分析,结果证实 *Bsg2* 是一新基因(Fig. 1).并向 NCBI 提交,序列号为 AY611626.

2.2 对 *Bsg2* 序列的生物信息学分析

2.2.1 基因组定位 *Bsg2* 基因定位在小鼠 12 号染色体的短臂上,包含 1 个内含子和 2 个外显子.该基因含有起始密码子 ATG 和终止密码子 TGA,在终止密码子后面还有 Poly A 的信号 AATAAA (Fig. 2). 外

显子的分布见 Table 1.

2.2.2 结构域分析 通过生物信息学分析,*Bsg2* 包含一个 2142 bp (482~2623) 的完整阅读框架,编码一个含 713 个氨基酸的锌指蛋白.此蛋白 N 端从第 14~121 氨基酸含有一个 BTB 结构域,C 端含有 9 个连续的 C2H2 型锌指结构域,大多数锌指结构域含有 23 个氨基酸,它们所在的氨基酸位置分别为 216~242,421~443,448~470,534~554,578~600,606~628,634~656,662~684,686~709 (Fig. 3). 这些结构域提示 *Bsg2* 基因可能是一个转录因子.

2.2.3 人、鼠、鸡基因的蛋白同源性比较 *Bsg2* 与人的 KIAA0997 基因在氨基酸水平上有 85.6% 的同源性.通过与鸡 EST 数据库^[11]进行比对 (<http://www.chick.umist.ac.uk>),发现 *Bsg2* 与鸡 EST 359288.1 的蛋白同源性达 60.8% (Fig. 4). 这说明 *Bsg2* 在不同种属里具有较高的保守性.

2.2.4 蛋白质特点及功能基序分析 *Bsg2* 分子量为 81 952.1,理论上的等电点为 5.94,亚细胞定位于细胞核.*Bsg2* 编码框的氨基酸顺序上可能有 13 个蛋白质激酶 C 磷酸化位点(123~125,131~133,135~137,203~205 等),11 个酪蛋白激酶 II 磷酸化位点(92~95,203~206,305~308,348~351 等),8 个 N-十四烷基化位点(153~158,301~306,379~384,400~405 等),5 个 N-糖基化位点(64~67,246~249,353~356,409~412,420~423),2 个 cAMP 和 cGMP 依赖性蛋白激酶磷酸化位点(194~197,303~306).从这种蛋白质的亚细胞定位和功能基序分析结果看,*Bsg2* 的生物学活性有可能接受信号途径中多种信号的调控.

2.3 *Bsg2* 基因电子表达谱分析

Unigene 分析显示,表达部位有:原肠胚、胚胎滋养层、大脑、小脑、肺、脾、肾、瘤、乳腺、胰腺、颈、睾丸等.虚拟 Northern 分析表明:*Bsg2* 在肺、脾、肾、肝、胰、胎盘、睾丸、大脑、小脑、脑下垂体、外周神经系统、头和颈、眼、胚性组织等多种组织和细胞中表达,肝癌、乳腺癌、皮肤癌组织中也有表达.

2.4 RT-PCR 检测结果

结果表明 (Fig. 5A),*Bsg2* 的特异条带大小与预计结果一致,同时证实其在小鼠各组织中均有表达.*Bsg2* mRNA 在各组织中表达的相对量依次为 0.3347 (肝),0.7552 (肠),0.8912 (脾),0.8038 (肾),0.8023 (睾丸),0.6361 (脑),0.0560 (肌肉),0.1960 (肺),0.4467 (心),0.5817 (子宫),0.0580 (小脑),0.1258 (眼).由此可见,*Bsg2* 在小鼠成体各组织中普

遍表达,在脾、肾、睾丸、肠、子宫、心和脑的表达水平最强,肝、肺、小脑、眼、肌肉里的表达相对较弱.

GAPDH 条带与预计结果相符 (Fig. 5B), 且其在各组织中的表达几乎一致.

1 GAGGAGGGCCAAAGCGACTCCGGGAGAGGGGCGACAGTGTCTGCTCCAGCGCCGCCGCCGGCCGCGCTGCGGGCAGTGCACGCTCCCGGGCC
101 GACGCAGAACTTGTGCAACCAACAGCGACCGGGTGCCTAGCCGCCAGCGGGAGTGGGTGGCGGGCGAGAGGGAGGGGAAGCTGGCGGAC
201 GCCCGCGCAACTGGTGGCAGACCCGGAGTCCGCGGTAAAGCGGGCCGGCTCAGTGGGTGCGGCAGCGCGGCTGTGCGGCAGCGGAAGTCTTTGT
301 TGCAGCGCAGTCCCTGGCAGAGCCGGAGTGTCTGGCGCAGCCCGCAGCCGCGCGCGCTCAACTGCAGCCCGAGCCCTCGCTCCGGCCCGCC
401 TTCCAGCGGCTCCGCTGCCCTCAACTCCATAAAACCTCCGACGCGCGGTCCACCGCGCCCGGTGAGTCAGCAGAAGATGGCAAAGCCAGCCACA
M A K P S H
501 GCAGCTATGCTCGCAGCAGCTCAACAACCAAGAGAGTGGGCTTCTGTGACTGCTGTATCGCTATCGACGACATCTACTTCAAGGCACAAAAGC
S S Y V L Q Q L N N Q R E W G F L C D C C I A I D D I Y F Q A H K A
601 CGTCTGGCTGCCTGATGCTTACCTTAGGATGTTTTGATGAACCATCAGCAGTACTGCACAACCTGAACCTCAGCAACATGAAGATAAGTGTCTGAG
V L A A C S S Y F R M F F M N H Q H S T A Q L N L S N M K I S A E
701 CTTTGTCTCATTITTCAGTTCATGTATTTAGGAAAGATCAGCCGCTCCTCCAGTTTCGAGCAGTTCAAAGTGGCAATGAACTACCTCCAGCTGT
C F D L I L Q F M Y L G K I M T A P S S F E Q F K V A M N Y L Q L
801 ACAATGTTCCGACGTGTGGAGAGACATACAGGACGACAGCTTCTAGTTCAAAGTGTCTGCTTCTGCTCCAGCAGGCGAGCAGCAAGATGATTTT
Y N V P D C L E D I Q D A D C S S S K C S S S A S S R Q S S K M I F
901 TGGGTGAGAATGTATGAGACACGGTGGCTAGAATGGCAATGAAGCAACAGTGGTGTGCGGAGCCAGTCAACGGTGAATACGCCGATCACAGA
G V R M Y E D T V A R N G N E A N R W C A E P S T V P H H R
1001 GAACCTGAGGAAGACTTTGCAAGTGGCCAACTCCCGGCGCTGTTGATGTGTGTAATAAAAGTTCTGTGTCCAAATATCTACTCCAAAAGAAC
E P E E E S L Q L A N F P E L F D V C K K S V S C K S V S P K E
1101 GTGTGTCGACGCTTGGACGGAGCTTACCTGTGATAGTTGTGGATTGGCTTCAGCTGTGAAAAGTTACTAGACGAACATGTGCTGACCTGTACCAA
R V S R R F G R S F R T C D S C G F G F S G E K L L D E H V L T C T N
1201 CAGACACTCATACAAAACACGACAAGAGCTTACCACCGAATAGTAGATATAGAGATGAAAAGATAGTAACATCAAAGTGAATGCTGAAAAGGAT
R H S Y Q N T T R A Y H R I V D I R D G K D S N I K A E L A E K D
1301 TCTCTAAAACATTTCTGCACAGCCGACAAATACAGAGAAGCGCAACCGCCGCTGATGATTCAGCCTGCACCACTGGAAGCAGGAAAAGCCCG
S S K T F S A Q P D K Y R E D A N Q A P D D S A S T T G S R K S T
1401 TGGAGGCTGGGATAGCTGGTGGAGAGAAAAGCAGAGCTACTGAGACAAAACGAATATCATCAAGATGGAGCCAGAGATATCCCTGCAGATGACATGAA
V E A G I A G E E K S R A T E T K R I J I K M E P E D I P A D M K
1501 AGACTCAACATTTAAAGTTACTGAGAAGACTGCAACGACTCCACCGACAATGATGAATAGAAGATGAAACCCGAGGAGCCATTTATAGATACTAT
D F N I K V T E K D C N E S T D N D E L E D E P E E P F Y R Y Y
1601 GTTGAAGAGATGTGGCATTAAAAAAGTGGTAGGAAAACCTAAAACCTCGTATGTCCATCAGCGTGGATGAGAGAGGTGGCTTAGAGAACATGAGAC
V E E D V G I K K S G R K T L K P R M S I S V D R G L E N M R
1701 CCCCATAACACAGCCCATACAAGAGGATGACAGAGAAATGATCTGTGAGCTGTGGACTCACAATAACTGAGGAGGAGCTGTGCTCTACTTACTT
P P N N T S P I Q E D A E N A S C E L C G L T I T E D E L S S H Y L
1801 GGCCAAAGCATTGAAAATATCTGTGCATGTGGTAAATGTGGACAGACTGTGCAAGGGCAGACAGCTTCCAGAACTGCTCAGAGGTGGCGAGAACCC
A K H I E N I C A C G K C G Q J L V K G R O L Q E H A O R C G E P
1901 CAGGATGACAAATGAATGGGCTAGGAAAGCTGACGAGAAAATGGACATGGAAGAGAACCCCGACGAGCAGTCTGAGATAAGAGACATGTTTGTGAAA
Q D L T M N G L G N A D E K M D M E E N P D E Q S E I R D M F V E
2001 TGCTGGATGATTCAGAGACAATCATTACCAATAAATAGTATCCAGAAAAGCAGTATTATAACACTTCTGCCTGTCTTCCGATGTCTAATGTGG
M L D L D F R D N H Y Q I N S I Q K K Q L F K H S A C P F R C P N C G
2101 CCAGCGTTTTGAAACTGAAAATTTAGTGGTGAAGATATGCTAGCTGCTAGACCAAGATATGTTCAAGGGTGCCTCATGGAAGAAAATGAAAGAGAT
Q R F E T E N L V V E H M S S C L D Q D M F K G A I M E E N E R D
2201 CACAGAAGAAAACATTTTGTAACTGTGTGGAAAAGGATTTTACAGCGCTGTCATTTGAGAGACACTACTGTGCACACTAAAGAAAAGCAGTTTG
H R R K H F C N L C G K G F Y Q R C H L R E H Y T V H T K E K Q F
2301 TTTGTGACAGATGTTGGAAGCAGTTTTTAAAGAGCGCCAGTTGCGTCTACACAATGATATGCACAAGGCATGGCCAGGTATGCTGTTCATTTGTGA
V C Q T C G K Q F L R E R Q L R L H N D M H K G M A R Y V C S I C D
2401 TCAAGAAACTCCGAAAACATGACCATGTACGCGATATGATCTGCGACCTGTCTGGTGGGAGACTATATGCCAGGTCTGCTTCCAGATATCCCAAT
Q G N F R K H D H V R H M I S H L S G G E T I C Q V C F Q I F P N
2501 AATGAACAACCTGGAGCAGCATATGGATGTCATCTGTACAGCTGTGGAATATGTGGAGCAAGTTAATTTAGGAAGGATATGAGATCGCATATAATG
N E Q L E Q H M D V H L Y T C G I C G A K F N L R K D M R S H Y N
2601 CCAAGCATTGAAAAGAACCTGATGGGTTTCTACTGTACAGATGTTTAGTTGGTAGGCAGAACCAAGCAAAAAGGCTATGAGTATTTAGAAGTGTAT
A K H L K R T *
2701 TTTATAACAAGTGGGGGAAATTTCAAGGCCCTTTTACTTTGAACGTTTTGTTTAGGATCTTAAAGTCTATTTTAGGCATTAATGTTTATCATTTTT
GTTGCTTTGAGAGCATCTTTGTCCTAGTTTTCTTAATACCATAAAAATACTTTTCTGTGAGACTGTAAAGTGGCTATATCCCTTCAATAACTATT
2801 AAAACAGTTTTTGGGTTTTTTTGGTTTTTTTAAAAAAGATAATGACTTTGCTGTTTGTATGACTTTTCATTTAAGGTTATCCTATGAAATTTT
2901 AAACCCAGATTTAGCAGCTCCTGCTTAAATTTTTAAAGAAATTTAAAAATAGTATTTAAAAATATCCTATTCATTTTTAACTTTTTTGTTTTTTTG
3001 CTTTTGGGATTTTTTGTGTGGTTTTTGGGGGTGCTTTTTGTGTCAGTTGATAGTGGGGTGGCTACTATTCCTCCCTATAAATGAAATCTATAT
3101 TCTAAAATGACAACTTATTAGGCAAGTGAAGTGCATGAATCTAACCTTAAAGTGTGACATGGGCTCAATCTGGCCTGAGATGATGGACCTAAGGA
3201 AACTCCTATAAATGAACTTTCCTATAATGATAATGTATCATTTAATAATCACATTTAAAACCTTATTAAGTGTAAATCCCAATGACTTCCCTCA
3301 CTTGAGAAATAGTGGGATAAGCCATTTGTTTTGTGATAATGAATTAACATAGTATGTTAATAAAAAGACATCTGTATTCATTAATAATATTTTA
3401 ATTTAAAAATTTCTGATTTCTAAAGTGTGTAGTTTATCAATATTTTGTTTATAAATAGACTTTAATAGCTCTTAGGAAATAGACATCTAAGAAAA
3501 ATGATTTTTATCTTTAAAATTACTTTGGAAGTGAATTTGAAAAGCTGCCATGTATATGATAAAGTAAATAAAGAAATCAATT
3601

Fig. 1 cDNA sequence of the mouse Bsg2 gene

Coding sequences are shown together with the translated amino acid sequences. Initiation codon (ATG) and stop codon (TGA) are underlined. The polyadenylation signal sequences (AATAAA) are also underlined

Fig. 2 Genomic organization of Bsg2 gene

The mouse Bsg2 gene was composed of 2 exons and 1 intron distributed over approximately 18 kb. Solid bars represent exons. The locations of initiation codon ATG, stop codon TGA and polyadenylation signal (PolyA Signal) are indicated

Table 1 Exon-intron junction of the mouse *Bsg2* gene

Exon number and size (nt)	cDNA position of exon	5' splicing site	Intron number and size (kb)	3' splicing site
1 (465)	1—465	CGCCC Gtgagt	14.511	aaacagGAGGAG
2 (3226)	466—3691	TCAAT Tataaat		gcatagGTCGAG

The intron-exon boundary sequences were generally consistent with the 5' and 3' splice consensus sequences and conformed to the GT/AG rule

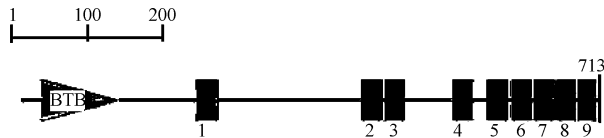


Fig. 3 The predicted domains of *Bsg2* gene

The first one represents BTB domain. The last consecutive nine bars represent nine C2H2 type zinc finger domains

Fig. 4 Alignment of mouse *Bsg2* amino acid sequence with that of human KIAA0997 and chick EST

The consensus sequences are shadowed. They share 85.6% and 60.8% similarity, respectively

Fig. 5 (A) The RT-PCR result of *Bsg2*. (B) The result of *GAPDH*

Lane 1, liver; lane 2, intestines; lane 3, spleen; lane 4, kidney; lane 5, testis; lane 6, brain; lane 7, muscle; lane 8, lung; lane 9, heart; lane 10, uterus; lane 11, cerebella; lane 12, eye.

GAPDH was used to control for differences in loading

2.5 Bsg2 基因全长编码区的克隆

通过 RT-PCR 获得 *Bsg2* 基因全长编码区序列的克隆, 并将它克隆至 pBluescriptII SK(-) 载体上. 扩增的目的片段与预计结果一致, 并通过测序验证. 所得到的全长编码区的序列用于原位杂交探针的制备.

2.6 原位杂交

整体 (whole mount) 原位杂交表达分析显示, *Bsg2* 在小鼠胚胎和鸡胚中均有表达. 尽管 *Bsg2* 基因普遍表达, 在小鼠胚胎和鸡胚的发育过程中该基因主要在中枢神经系统表达, 在前脑、中脑和后脑有较强的表达信号 (Fig. 6, 7), 而且在 11 d 的鼠胚, 前后肢芽里也有较强的表达 (Fig. 6D).

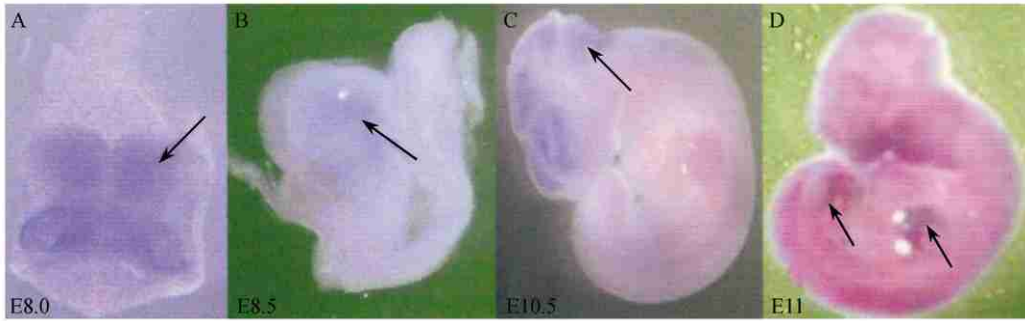


Fig. 6 Whole-mount *in situ* hybridization of mouse embryo of different stages

Bsg2 gene is expressed specifically in embryos' head and E11 embryo's limb (see arrow).

(A) E8.0 mouse embryo, (B) E8.5 mouse embryo, (C) E10.5 mouse embryo, (D) E11 mouse embryo

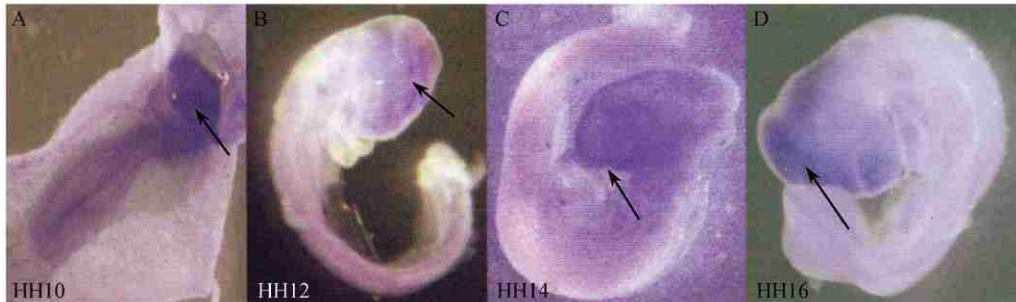


Fig. 7 Whole-mount *in situ* hybridization of chicken embryo of different stages

Bsg2 gene is expressed specifically in embryos' head (see arrow).

(A) HH10 chicken embryo, (B) HH12 chicken embryo, (C) HH14 chicken embryo, (D) HH16 chicken embryo

3 讨论

本文利用消减差异筛选法, 克隆到小鼠新基因 *Bsg2*. 通过生物信息学的方法预测该基因定位在小鼠 12 号染色体上, 它编码一个含 713 个氨基酸的锌指蛋白, 含有一个 BTB 结构域和 9 个连续的 C2H2 型锌指结构域. *Bsg2* 与人的 KIAA0997 基因在氨基酸水平上有 85.6% 的同源性, 与 Chick EST 359288.1 的蛋白同源性达 60.8%. 这说明 *Bsg2* 在不同种属里具有较高的保守性.

C2H2 型锌指结构域是一类结合核酸的重要结构域. 锌指蛋白在基因的表达调控、细胞分化、配子的形成和胚胎发育等生命过程中发挥重要作用, 其

中某些锌指蛋白还可能与特定的疾病的发生有关. BTB/POZ 结构域主要存在于锌指蛋白的 N 末端以及含有 Kelch 基序的蛋白中. BTB/POZ 结构域可以调节 BTB/POZ 蛋白的自我联合^[12], 参与同源二聚体和一些异源二聚体的形成, 以及蛋白质与蛋白质的相互作用. 目前所知的多数含 BTB/POZ 结构域的锌指蛋白均表现出转录抑制活性, 抑制 DNA 结合结构域与 DNA 序列的结合. 此类蛋白质对于生物体的发育、分化、染色体重组等过程的调节具有相当重要的作用, 并与癌症有密切的关系. *Bsg2* 基因的这些重要结构域提示 *Bsg2* 可能是一个重要的转录因子, 参与了基因的表达调控和胚胎发育等重要过程.

从 *Bsg2* 的电子表达谱中发现 *Bsg2* 在多种组织

(包括肝癌、乳腺癌、皮肤癌组织)中表达,说明其在多种组织中发挥作用,并很可能与癌症相关,而这需要进行进一步的验证. RT-PCR 结果亦表明 *Bsg2* 基因在小鼠胚胎及成体的各组织中普遍表达,在脾、肾、睾丸、肠、子宫和脑的表达水平较强. *Bsg2* 基因的电子表达谱与 RT-PCR 的结果基本上是一致的.

整体原位杂交表达分析显示, *Bsg2* 在小鼠胚胎和鸡胚中均有表达,验证了 *Bsg2* 在不同种属中具有较高的保守性. 尽管 *Bsg2* 基因普遍表达,在小鼠胚胎和鸡胚的发育过程中该基因主要在中枢神经系统表达,在前脑、中脑和后脑有较强的表达信号,在 11 d 鼠胚的肢芽里也有较强的表达. 由此可见, *Bsg2* 很可能在胚胎发育中与头部的发育相关. 在胚胎第 10.5 d 天时,很多器官正在形成,这暗示着 *Bsg2* 很可能在器官形成过程中发挥着重要的作用. *Bsg2* 在不同种属中显著的保守性,以及其在不同的组织器官中的表达,暗示着 *Bsg2* 在不同种属中发挥着重要的生物作用. *Bsg2* 基因的结构和表达特征预示着它编码一个具 DNA 结合功能的转录调控因子,同时揭示它在脑的发育和器官形成过程中发挥着重要作用.

参考文献 (References)

- Li Y P, Alexander M, Wucherpfennig A L, Yelick P, Chen W, Stashenko P. Cloning and complete cDNA sequence of a novel human cathepsin expressed in osteoclast. *J Bone Miner Res*, 1995, **10** (8): 1197 ~ 1202
- Li Y P, Chen W, Stashenko P. Molecular cloning and characterization of a putative novel human osteoclast-specific 116 kDa vacuolar proton pump subunit. *Biochem Biophys Res Commun*, 1996, **218** (3): 813 ~ 821
- DiBello P R, Withers D A, Bayer C A, Fristrom J W, Guild G M. The *Drosophila* Broad-Complex encodes a family of related proteins containing zinc fingers. *Genetics*, 1991, **129** (2): 385 ~ 397
- Harrison S D, Travers A A. The tramtrack gene encodes a *Drosophila* finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. *EMBO J*, 1990, **9** (1): 207 ~ 216
- Gødt D, Couderc J L, Cramton S E, Laski F A. Pattern formation in the limbs of *Drosophila*: bric a brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development*, 1993, **119** (3): 799 ~ 812
- Zollman S, Gødt D, Prive G G, Couderc J L, Laski F A. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc Natl Acad Sci USA*, 1994, **91** (22): 10717 ~ 10721
- Albagli O, Dhordain P, Deweindt C, Lecocq G, Leprince D. The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ*, 1995, **6** (9): 1193 ~ 1198
- 王重, 赵德标. BTB/POZ 蛋白质家族的结构和功能. 生命的化学 (Wang chong, Zhao De-biao. The structure and function of the family of the BTB/POZ protein. *Chem Life*), 1997, **17** (6): 10 ~ 13
- Wilkinson D G. *In situ* hybridization, In: Stern C D, Holland P W H ed. *Essential Development Biology, A Practical Approach*. Oxford: IRL Press, 1993: 257 ~ 274
- Boardman P E, Sanz-Ezquerro J, Overton I M, Burt D W, Bosch E, Fong W T, Tickle C, Brown W R A, Wilson S A, Hubbard S J. A comprehensive collection of chicken cDNAs. *Curr Biol*, 2002, **12**: 1965 ~ 1969
- 杨劲松, 郑新民, 陈诗书. NIH3T3 细胞恶变早期相关新基因的鉴定与功能分析. 中国生物化学与分子生物学报 (Yang Jin-song, Zheng Xin-min, Chen Shi-shu. Identification and function assay of two novel genes in early stages of transformed NIH3T3 cells. *Chin J Biochem Mol Biol*), 2003, **19** (2): 173 ~ 181
- Bardwell V J, Treisman R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev*, 1994, **8** (14): 1664 ~ 1677