

类固醇 5- α 还原酶样基因 Srd5 α 2l2 的克隆及表达分析*

陈祥贵^{1,2)} 李 勇¹⁾** 张德礼³⁾ 成 君¹⁾ 朱文丽¹⁾ 刀京晶¹⁾

⁽¹⁾ 北京大学公共卫生学院营养学系, 北京大学分子毒理与发育分子生物学实验室, 北京 100083;

⁽²⁾ 四川工业学院生物工程系, 成都 610039;

⁽³⁾ 清华大学信息科学技术学院, 自动化系智能技术与系统国家重点实验室, 北京 100084)

摘要 采用表达序列标签 (EST) 介导的基因克隆和表达谱分析, 从小鼠心脏克隆了一个 cDNA 为 2 348 bp, 主要在心脏表达的新基因 Srd5 α 2l2 (GenBank Acc No. AF548365). 该基因由 12 个外显子组成, 3' 非翻译区富含 ATTTA 序列, 最大开放阅读框编码一个由 361 个氨基酸组成的假定蛋白, 该蛋白质 C 端含有类固醇 5- α 还原酶的保守结构域 (3-oxo-5- α -steroid 4-dehydrogenase, STEROID_DH). 生物信息学分析表明, 人 cDNA 克隆 DKFZp313D0829 (AL833108) 是 Srd5 α 2l2 的人同源基因. 经同源性检索, 支持 Srd5 α 2l2 的全部 41 条 EST 中 25 条来自小鼠心脏组织. RT-PCR 检测初步证实, 该基因主要在心脏中表达, 而在其他组织中不表达或弱表达. 综合考虑 Srd5 α 2l2 的序列特征和表达谱, Srd5 α 2l2 可能在心脏组织中发挥重要作用.

关键词 Srd5 α 2l2, 类固醇 5- α 还原酶, 新基因, 心脏特异

学科分类号 Q78

同一机体的不同细胞具有同一套基因, 细胞、组织和器官分化是基因在时空发育轴上差异表达的结果. 人类疾病的发生也常常与基因差异表达相关联. 如果某个基因在某组织和其他组织间差异表达, 就可认为该基因具有一定的组织特异性, 那么该基因的功能可能和该组织密切相关. 这类组织特异性基因不仅有助于认识特定组织的分化发育过程和生化代谢机制, 而且在疾病的分子诊断、药物设计等方面有重要的应用价值.

近年来, 生物信息学的发展为发现组织特异性基因提供了新的途径. 作为基因表达信息的重要标志物, 表达序列标签 (expressed sequence tag, EST) 在基因识别、绘制基因表达图谱、寻找新基因等研究领域获得了广泛的应用^[1]. 基于 EST 的电子克隆技术已经成为基因克隆的重要方法^[2,3]. 我们采用 EST 介导的基因克隆和表达谱分析, 成功克隆了一个在小鼠心脏相对特异表达的新基因, 其编码的假定蛋白含有类固醇 5- α 还原酶的保守结构域, 国际小鼠基因组命名委员会 (MGNC) 命名为 Srd5 α 2l2 基因 (steroid 5 α -reductase 2 like 2).

1 材料和方法

1.1 材料

TRIzol 试剂购自 GibcoBRL; 反转录试剂盒 (Reverse Transcription System)、pGEM-T easy 载体

系统、分子质量标准 Φ 174DNA/*Hae* III 购自 Promega 公司; DNA 聚合酶购自 Invitrogen 公司; 琼脂糖 DNA 纯化系统购自华美生物工程公司. Balb/C 小鼠由北京大学医学部动物部提供.

1.2 生物信息学方法

通过 Internet 进入美国国家生物信息中心 (NCBI) Unigene 数据库, 利用 Digital Differential Display (DDD) 系统 (http://www.ncbi.nlm.nih.gov/unigene/info_ddd.shtml), 查询 13 天小鼠胚胎心脏 cDNA 文库 (RIKEN full-length enriched, 13 days embryo heart, dbEST Library ID. 5466, NCBI UniGene Lib. 428) 的表达序列标签 (EST) 在 Unigene Clusters 中的分布, 根据 Unigene 注释信息选择其中不代表已知基因的 EST 作为查询序列, 利用 BLAST 程序搜索小鼠同源 EST 并组装成新的查询序列进行下一轮的 BLAST 搜索, 直至无新的 EST 可用于延伸查询序列. 组装的序列输入 Nr 数据库通过 BLAST 验证其是否为新 cDNA, 然后进行核酸和蛋白质序列分析. 开放阅读框 (ORF) 预测采用 ORF

* 国家重点基础研究发展规划项目 (973) (G1999055904), 国家自然科学基金资助项目 (30030120) 和北京大学“985”资助项目.

** 通讯联系人.

Tel: 010-82801177, E-mail: liyong@bjmu.edu.cn

收稿日期: 2003-04-07, 接受日期: 2003-06-09

finder 软件 (<http://www.ncbi.nlm.nih.gov/>), 蛋白质结构域分析采用 SMART 程序 (<http://smart.embl-heidelberg.de/>)^[4], 蛋白质亚细胞定位采用 PSORT 程序 (<http://psort.nibb.ac.jp/>).

1.3 总 RNA 的提取和反转录

妊娠 Balb/C 小鼠乙醚麻醉, 剖取胚胎, 剥离胚胎心脏. 15 日龄小鼠和成年小鼠则直接剖腹取心. 各发育时期心脏和取心脏后的胚体用冷 PBS 洗涤后加入 TRIzol 试剂于冰上匀浆, 然后按 TRIzol 试剂盒说明提取总 RNA, 琼脂糖凝胶电泳验证 RNA 的完整性, 经紫外检测定量后, 用 Olig T 引物按试剂盒说明进行反转录, 得到第一链 cDNA, 作为 PCR 扩增预期片段的模板.

1.4 PCR 扩增、产物克隆及测序

根据组装的重叠群, 用 Primer 5.0 设计引物扩增. 上游引物 1F 为 5' CCC AGC CTG CTA AAT AAA AC 3'; 下游引物 1R 为 5' AAT TGA ACC TAT ACA GCA AC 3'. 采用降落 PCR 循环: 94℃ 3 min; 20 × (94℃ 40 s, 60 ~ 50℃ 40 s, 68℃ 3 min); 15 × (94℃ 40 s, 50℃ 40 s, 68℃ 3 min); 68℃ 5 min. 扩增产物经琼脂糖凝胶电泳, 回收 2 000 bp 左右的条带, 将其克隆到 pGEM-T easy 载体, 然后转化 JM109. 对插入片段先用通用引物 SP6 和 T7 进行双向测序, 然后设计接力引物测出插入片段的全部序列. 引物合成和序列测定均由上海生物工程有限公司完成.

1.5 RT-PCR 检测基因表达谱

以 MTC panel (Clontech 产品) 中小鼠多种组织的 cDNA、不同发育时期心脏 cDNA 和取心脏后的胚体 cDNA 为模板进行 RT-PCR, 研究基因表达. 根据 PCR 产物测序结果, 设计 RT-PCR 引物, 上游引物 2F 为 5' TTT TCC TGG TTT CCT GTC CTA ACT 3', 下游引物 2R 为 5' CAA ACT CAT CGG GCT GAT AAT GT 3', 预期扩增片段长度 453 bp (对应图 2 的 1 074 ~ 1 526 bp). PCR 反应体积为 50 μ l, 含 5 μ l 模板 cDNA、250 μ mol/L dNTP、2.0 μ mol/L MgCl₂、0.4 μ mol/L 引物. PCR 循环: 94℃ 3 min; 38 × (94℃ 30 s, 56℃ 30 s, 72℃ 1 min); 72℃ 5 min. 每个反应管取 5 μ l 在琼脂糖凝胶电泳进行检测. 以 G3PDH 监测各组织样品中 mRNA 的表达丰度. G3PDH 的上游引物为 5' TGA AGG TCG GTG TGA ACG GA TTT GGC 3'; 下游引物为 5' CAT GTA GGC CAT GAG GTC CAC CAC 3'.

2 结果与分析

2.1 重叠群获得与分析

用来自 13 天小鼠胚胎心脏的两条 EST (BB659092 和 BB477174) 作为最初的查询序列, 经过小鼠 EST 搜索拼接获得两个重叠群 (contig): contig1 和 contig2. 经对小鼠基因组进行 BLAST 搜索 (<http://genome.ucsc.edu/>), 将两 contig 分别定位于小鼠染色体 5 (-) 的 82 815 781 ~ 82 821 996 位和 82 827 715 ~ 82 892 789 位, 两者虽然互不重叠, 但相距只有约 5.7 kb. contig1 由 4 个外显子组成, contig 2 由 8 个外显子组成. 基因预测软件 SGP 和 Fgenesh 在该区域的预测基因基本上包括了这两个重叠群 (图 1). EST 组织来源分析表明, 两个 contig 的 EST 都主要来自心脏. 这些结果提示这两个重叠群可能来自同一转录本, 而且该转录本可能具有心脏特异性.

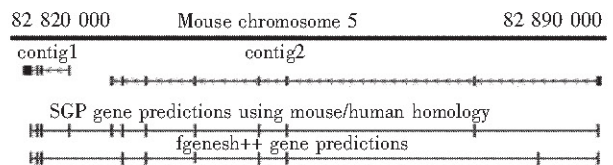


Fig. 1 Analysis of contigs

2.2 cDNA 克隆与序列测定

根据获得的两个重叠群设计的一对引物中, 上游引物位于 contig2, 下游引物位于 contig1, 经 RT-PCR, 从 14 天小鼠胚胎心脏成功扩增了一条 1 918 bp 的 cDNA. 对重组 pGEM-T easy 质粒的插入片段序列测定, 结果与重叠群 contig1 和 contig2 全程比对, 结果表明, 该 1 918 bp cDNA 覆盖了两重叠群中引物对之间的全部序列, 证实两个重叠群的确来自同一转录本. 根据 contig1 和 contig2 的序列对获得的 1 918 bp cDNA 进行延伸, 最后得到一条 2 348 bp 的 cDNA. 经同源检索确认为新的 cDNA, 其序列受小鼠基因组的完全支持, 经国际小鼠基因组命名委员会 (MGNC) 建议, 命名为小鼠 Srd5 α 2l2 (steroid 5 α -reductase 2 like 2) 基因. 将该基因的 cDNA 序列发送 GenBank, 接受号为 AF548365.

2.3 Srd5 α 2l2 序列分析

Srd5 α 2l2 基因 cDNA 全长为 2 348 bp, 其最大

的开放阅读框 (ORF) 位于 197~ 1 282 bp 编码一 361个氨基酸残基组成的假定蛋白, 该蛋白质分子质量为 41.874 ku, 等电点为 10.1, PSORT程序预测位于胞质 (52.6%) 和胞核 (26.1%), 其 210~ 361位氨基酸残基构成完整的类固醇 5- α 还原酶保守结构域 (3-oxo-5- α -steroid 4-

dehydrogenase, STEROID_ DH) (图 2). 在该最大 ORF前同一读码框内有终止密码子, 3'端有多个候选的加尾信号 AATAAA和 ATTAAA, 3'非翻译区 16次出现 ATTT序列, 包括 6处 ATTTA序列 (图 2), 这一序列特征被认为可能与 mRNA的选择性降解有关^[5 6].

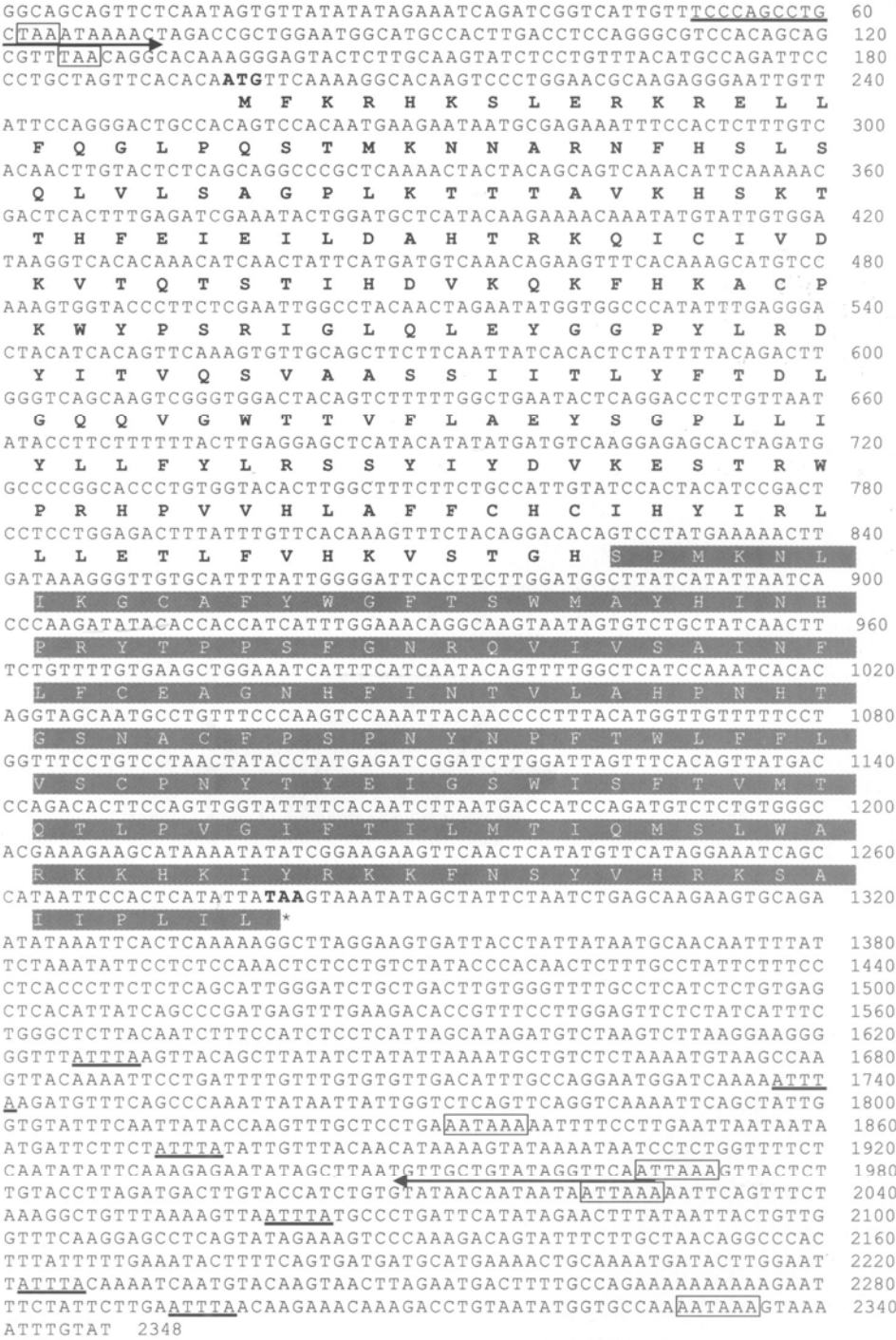


Fig. 2 Nucleotide and deduced amino acid sequence of Srd5a2L

Steroid 5- α -reductase domain is shown in white with gray background. The ATTTA sequences are underlined. The in-frame stop codon before ORF and the putative polyadenylation signal AATAAA and ATTAAA sequences are boxed. The positions of primer pair of 1 918 bp cDNA are shown by arrows.

Srd5 α 2l2 cDNA 对小鼠基因组 BLAST, 结果表明: Srd5 α 2l2 定位于小鼠第 5 染色体, 由 12 个外显子组成, 完全遵守 GT/AG 规则, 基因组 DNA 序列为 77 kb, 最大 ORF 横跨全部 12 个外显子.

2.4 同源性比较

经同源检索, 在核酸水平上 Srd5 α 2l2 基因编码区与 2002 年 7 月 10 日提交的人 cDNA 克隆 DKFZp313D0829 (GenBank Accession No. AL833108) 有 85% 一致性. AL833108 cDNA 全长 3 596 bp, 经对人类基因组 BLAST 定位于人染色体 4q13.1, 由 12 个外显子组成, 其最大 ORF 也是横跨全部 12 个外

显子. 将 Srd5 α 2l2 和 AL833108 对基因组 BLAST 的结果进行比较发现, 除第 1 外显子和第 12 外显子外, Srd5 α 2l2 和 AL833108 其他对应的外显子长度完全一致, 而对应的内含子长度有一定差异 (图 3). 同时还发现 AL833108 3'非翻译区含有 11 处 ATTTA 序列, 而 ATTT 序列出现高达 43 次.

将 AL833108 按最大 ORF 翻译, 得到一个 363 个氨基酸组成的假定蛋白, 它与 Srd α 52l2 的假定蛋白有 82% 的一致性和 90% 的相似性 (图 4). 这些比较提示 AL833108 是 Srd5 α 2l2 的人同源基因.

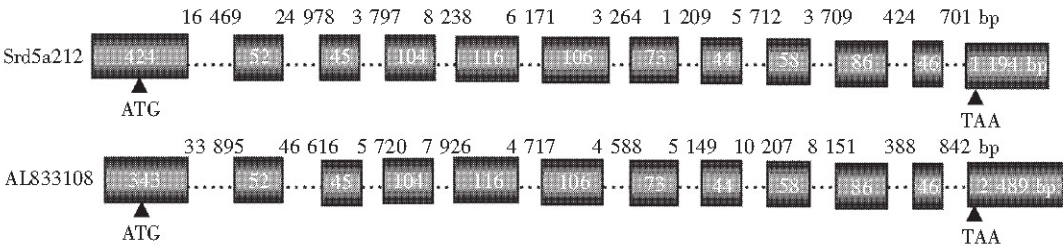


Fig. 3 Genomic structure of Srd5 α 2l2 and its putative human orthologue

The exons are figured with gray box and introns with dot line. Length of exons and introns are shown in gray box or over the dot line respectively.

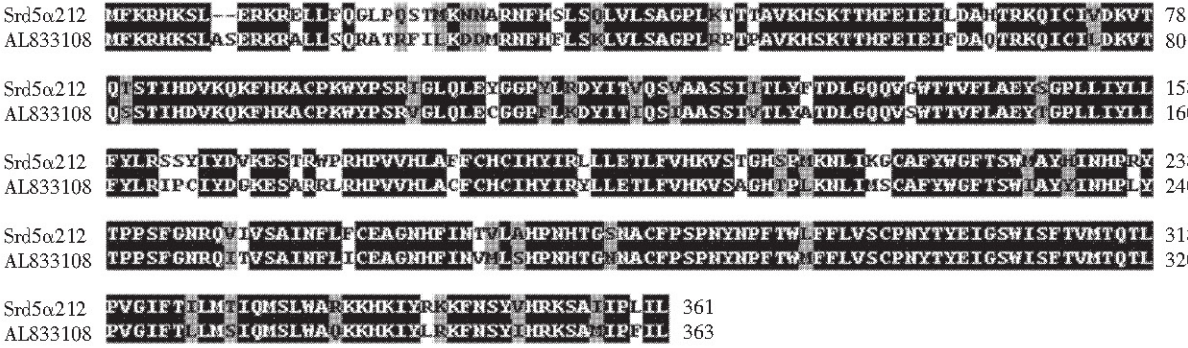


Fig. 4 Alignment of the deduced Srd5 α 2l2 protein vs its putative human orthologue

2.5 组织表达谱

2.5.1 基于 EST 的组织表达谱分析: 小鼠 Srd5 α 2l2 基因全长 cDNA 序列受 GenBank EST 数据库中 41 条 EST 的支持, 组织来源分析表明, 其中 25 条来自心脏, 其余 EST 分别来自脑 (4 条)、肺 (2 条)、黑色素瘤 M2 细胞 (3 条)、眼球 (2 条)、B16 F10Y 细胞 (1 条) 及全胚胎 (4 条); 人同源基因 AL833108 全长受 33 条 EST 支持, 其中 17 条来自心脏, 其余 EST 分别来自肺 (4 条)、骨骼肌 (7 条)、肝 (1 条) 及混合组织样品 (4 条). 各类组织的 cDNA 文库中某个基因的 EST

数量可粗略反映该基因在不同组织中的表达丰度^[7], 因此预期 Srd5 α 2l2 及其人同源基因在心脏中表达丰度远高于其他组织, 具有一定的组织特异性.

2.5.2 组织表达谱的 RT-PCR 检测结果: RT-PCR 发现, Srd5 α 2l2 基因在小鼠心脏高表达, 在肺中等表达、在骨骼肌弱表达, 而在脑、脾、肝、肾脏、睾丸等组织未检测到表达信号; 11、15 和 17 日龄全胚胎中均检测到表达信号, 而 7 日龄胚胎未见表达. Srd5 α 2l2 表达信号的出现与心脏发生时期相一致 (图 5a). 不同发育时期的心脏中 Srd5 α 2l2 均

表达,成年鼠心脏中表达有下降趋势;去心胚体中未检测到表达信号(图5b)。这些结果提示

Srd5 α 212 主要在心脏中表达,并显示随发育调节的格局。

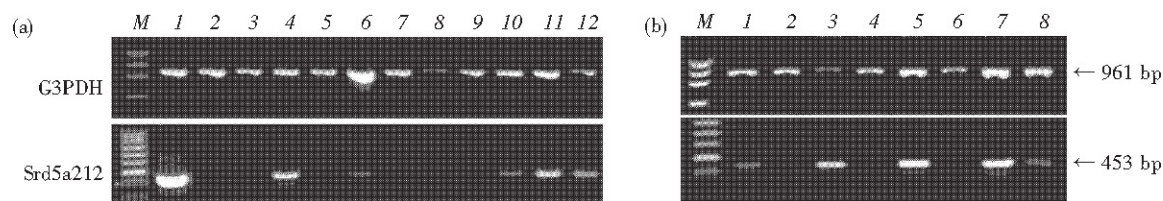


Fig. 5 RT-PCR analysis of Srd5 α 212 transcript in different mouse tissues

(a) 1: heart; 2: brain; 3: spleen; 4: lung; 5: liver; 6: skeletal muscle; 7: kidney; 8: testis; 9: 7-day embryo; 10: 11-day embryo; 11: 15-day embryo; 12: 17-day embryo. (b) 1: 11-day embryo heart; 2: 11-day heart-free embryo; 3: 15-day embryo heart; 4: 15-day heart-free embryo; 5: 17-day embryo heart; 6: 17-day heart-free embryo; 7: 15-day neonate heart; 8: 15-week adult heart. M: DNA marker (Φ 174 DNA/*Hae*III for G3PDH, 100 bp DNA ladder for Srd5 α 212)

3 讨 论

生物信息学的发展为基因研究提供了新的策略和方法。我们曾用 EST 介导的基因克隆和表达谱分析成功克隆了小鼠心脏特异性基因 *Lrrc10*^[8]。在本研究中,我们综合运用了 EST 拼接、重叠群的染色体定位分析以及 EST 介导的组织表达谱分析等生物信息学方法,得到的两个虽不重叠但位置邻近且电子表达谱相似的重叠群,并推测它们可能来自同一个在心脏特异表达的转录本,然后通过 RT-PCR 进行实验证实,成功克隆了主要在心脏表达的新基因 Srd5 α 212。这一技术路线不仅有助于发现新的组织特异性基因,而且对于克隆那些仅靠 EST 还不能拼接全编码序列的基因也有一定的参考价值。根据 Srd5 α 212 转录本在各组织中的表达丰度,可以初步确认 Srd5 α 212 具有心脏特异性,虽然还有待在不同发育时期的各种组织中进一步验证。同源比较结果表明,人 cDNA 克隆 DKFZp313D0829 (AL833108) 可能是 Srd5 α 212 的人直系同源基因,EST 来源分析提示,它与 Srd5 α 212 有极为相似的组织表达谱,也可能主要在心脏中特异表达。

蛋白质结构域分析表明, Srd5 α 212 属于类固醇 5- α 还原酶家族。GenBank 查询表明这一家族已有近 50 个成员,分布于酵母、果蝇、拟南芥、哺乳动物及人类等物种中,以类固醇 5- α 还原酶为代表,尚未见这些蛋白质在心脏特异表达的报道。类固醇 5- α 还原酶在雄性激素代谢和生物活性中起关键作用,包括两种同工酶基因: SRD5 A1 和 SRD5 A2,它们具有不同结构、功能和组织表达谱^[9~11]。类固醇 5- α 还原酶缺陷与性发育异常、

前列腺疾病有关。目前尚无证据表明该酶基因在心脏表达,因此, Srd5 α 212 是否具有类固醇 5- α 还原酶活性还有待实验确认。Srd5 α 212 及其人同源基因 cDNA 序列一个显著特征是 3' 非翻译区频繁出现 ATTTA 基序,该序列被认为是 mRNA 的降解信号,广泛存在于淋巴因子、细胞因子和原癌基因中,这些 mRNA 大多短暂表达,其 3' 非翻译区 AUUUA 序列可能促进成熟 mRNA 的降解^[5,6]。生物信息学分析还发现 Srd5 α 212 蛋白质富含潜在的磷酸化位点、糖基化位点和酰化位点,提示可能存在翻译后修饰、激活和调节机制,而磷酸化和可快速降解是信号传递控制的重要特征。根据 Srd5 α 212 的序列特征和表达谱,推测它可能是作为细胞信号通路的成员在心脏中发挥重要作用。

参 考 文 献

- 1 Wolfsberg T G, Landsman D. A comparison of expressed sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Res*, 1997, **25** (8): 1626 ~ 1632
- 2 张德礼, 丁培国, 凌伦奖, 等. 人类新基因 C17orf32 的电子克隆和编码区序列 RT-PCR 验证. *生物化学与生物物理进展*, 2002, **29** (4): 543 ~ 549
Zhang D L, Ding P G, Ling L J, *et al.* *Prog Biochem Biophys*, 2002, **29** (4): 543 ~ 549
- 3 Camargo A A, Samaia H P B, Dias-Neto E, *et al.* The contribution of 700 000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci USA*, 2001, **98** (21): 12103 ~ 12108
- 4 Schultz J, Copley R R, Doerks T, *et al.* SMART: A Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res*, 2000, **28** (1): 231 ~ 234
- 5 Shaw G, Kamen R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, 1986, **46** (5): 659 ~ 667
- 6 Raymond V, Atwater J A, Verma I M. Removal of an mRNA destabilizing element correlates with the increased oncogenicity of

- proto-oncogene fos. *Oncogene Res*, 1989, **5** (1): 1~12
- 7 钱 骏, 董 利, 张必成, 等. 表达序列标签数据库搜寻鉴定小鼠 UBAP1 基因及其数字化表达分析. *生物化学与生物物理进展*, 2002, **29** (2): 323~327
Qian J, Dong L, Zhang B C, *et al.* *Prog Biochem Biophys*, 2002, **29** (2): 323~327
 - 8 陈祥贵, 李 勇, 赵如冰, 等. 心脏特异新基因 *Lrrc10* 的分子克隆与特性分析. *生物化学与生物物理进展*, 2003, **30** (4): 617~622
Chen X G, Li Y, Zhao R B, *et al.* *Prog Biochem Biophys*, 2003, **30** (4): 617~622
 - 9 Andersson S, Russell D W. Structural and biochemical properties of cloned and expressed of human and rat steroid 5-alpha-reductases. *Proc Nat Acad Sci USA*, 1990, **87** (10): 3640~3644
 - 10 Andersson S, Berman D M, Jenkins E P, *et al.* Deletion of steroid 5-alpha-reductase 2 gene in male pseudohermaphroditism. *Nature*, 1991, **354** (6349): 159~161
 - 11 Thigpen A E, Silver R I, Guileyardo J M, *et al.* Tissue distribution and ontogeny of steroid 5-alpha-reductase isozyme expression. *J Clin Invest*, 1993, **92** (2): 903~910

Molecular Cloning and Expression Analysis of *Srd5 α 2l2*: a Novel Mouse Member of Steroid 5 Alpha-reductase Family *

CHEN Xiang-Gui^{1,2)}, LI Yong^{1)*}, ZHANG De-Li³⁾, CHENG Jun¹⁾, ZHU Wen-Li¹⁾, DAO Jing-Jing¹⁾

⁽¹⁾ *Department of Nutrition & Food Hygiene, School of Public Health, Laboratory of Molecular Toxicity & Developmental Molecular Biology, Peking University, Beijing 100083, China;*

⁽²⁾ *Department of Bioengineering, Sichuan University of Science and Technology, Chengdu 610039, China;*

⁽³⁾ *State Key Laboratory of Intelligent Department of Automation, School of Information Science and Technology, Tsinghua University, Beijing 100084, China)*

Abstract A novel mouse heart-specific gene, *Srd5 α 2l2*, was cloned from mouse embryo heart by EST database searching and RT-PCR. The cDNA of *Srd5 α 2l2* was 2 348 bp and composed of 12 exons within over 77 kb in mouse genome. This gene was mapped to mouse chromosome 5 by BLAST search to mouse genome. Analysis of the 3' untranslated region of the cDNA showed the presence of ATTTA sequence, which was believed to act as degradation signal of mRNA. The longest ORF spanned all of the 12 exons and encoded a putative protein of 361 amino acids. A conserved steroid 5 alpha-reductase domain (3-oxo-5-alpha-steroid 4-dehydrogenase, STEROID_ DH) was present at the C-terminus of the deduced protein. Homologous alignment found high identity of nucleotide sequence, deduced amino acid sequence and genomic structure between *Srd5 α 2l2* and human cDNA of DKFZp313D0829 (AL833108). It may be concluded that AL833108 was the human orthologue of *Srd5 α 2l2*. BLAST search showed that cDNA of *Srd5 α 2l2* was supported by 41 ESTs from EST database, 25 out of the 41 ESTs were from mouse heart. RT-PCR performed on a panel of different mouse tissues demonstrated that *Srd5 α 2l2* strongly expressed in mouse heart but not or much more weakly expressed in other tissues. These results suggested that *Srd5 α 2l2* is a novel heart-specific member of steroid 5 alpha-reductase family. As far as it is known that no heart-specific member of this family was reported before.

Key words *Srd5 α 2l2*, steroid 5 alpha-reductase, novel gene, heart specific

* This work was supported by grants from The Special Funds for Major State Basic Research of China (1999055904) and The National Natural Sciences Foundation of China (30030120).

** Corresponding author. Tel: 86-10-82801177, E-mail: liyong@bjmu.edu.cn

Received: April 7, 2003 Accepted: June 9, 2003