# cDNA Cloning and Expression Analysis of Mouse Gene Encoding The Protein Ercc6l Which Is a Novel Member of SNF2 Family*

CHEN Xiang-Gui[1,2], LI Yong[1)**], ZANG Ming-Xi[1)], PEI Xin-Rong[1)], XU Ya-Jun[1)], GAO Li-Fang[1)]

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

[2)] *Department of Bioengineering, Xihua University, Chengdu* 610039, China)

**Abstract** SNF2 family includes proteins essential to genome replication, repair, and expression. The cDNA cloning and initial characterization of a novel mouse member of this protein family was reported, and the new member was designated as *Ercc6l* (excision repair cross-complementing rodent repair deficiency, complementation group 6-like) by MGNC. The cDNA of *Ercc6l* was 4 002 bp and firstly cloned in silico and then identified by RT-PCR in mouse embryo heart. This gene was composed of 2 extrons and mapped to mouse chromosome X. The longest ORF encoded a putative protein of 1240 amino acids. Eight highly conserved sequence regions of SNF2 protein were present in the deduced protein. Alignment against members of SNF2 family initially placed Ercc6l into ERCC6 subfamily of SNF2 family. EGFP-tagged Ercc6l protein was localized in cytoplasm of HeLa, NIH3T3 and B16 cells. BLAST search in EST database retrieved 69 homologous ESTs of *Ercc6l*, most of which were from mouse embryonic and tumorous tissues. RT-PCR performed on a panel of different mouse tissues demonstrated that *Ercc6l* strongly expressed in mouse embryo but significantly downregulated in neonate and adult normal tissue. These results suggest that Ercc6l is a novel member of SNF2 family and may be crucial for development of embryo and tumor.

**Key words** *Ercc6l*, gene cloning, SNF2 protein, expression pattern, subcellular localization

DEAD/H box-containing proteins are found in a wide variety of species ranging from bacteria to higher animals, constituting a DEAD/H box protein superfamily. This protein superfamily is essential for RNA processing, ribosome assembly, translation initiation and involved in spermatogenesis, embryogenesis, and cell growth and division[1~4]. DEAD/H box proteins contain a signature protein sequence motif, in which an DEAD/H box (a conserved Asp-Glu-Ala-Asp/His motif) is found in tandem with an A-type ATP-binding site (AXXXXGKT)[5,6]. Based upon sequence conservation, DEAD/H box-containing proteins can be subdivided into several families. One such family is the SNF2 family, the members of which share similarity to the yeast transcriptional activator protein SNF2. All members of this family have the ability to hydrolyze NTPs, and many have been shown to function in transcriptional regulation, DNA recombination and repair[5]. Here we characterize a novel member of SNF2 family, designated as *Ercc6l* (excision repair cross-complementing rodent repair deficiency, complementation group 6-like) by Mouse Genomic Nomenclature Committe (MGNC). The GenBank Accesion Number for *Ercc6l* is AY172688. This gene was firstly identified in silico and then verified by RT-PCR. The expression pattern analysis suggests that Ercc6l might implicate in embryo development and tumor genesis.

## 1 Materials and methods

### 1.1 Preparation and reverse trancription of RNA

Total RNA was prepared from balb/c mouse various organ of different developmental stages using TRIzol (Gibco BRL, Grandisland, NY) according to the manufacturer's recommendations. Proper quantity of total RNA was reverse transcribed to first strand cDNA using oligo-dT (12 ~ 18) and M-MLV reverse transcriptase (Promega). Balb/c mice used in this study were provided by Animal Center of Peking University.

### 1.2 cDNA cloning and sequencing

The cDNA cloning of *Ercc6l* was the result of our initial intention to identify novel genes related to embryo (especial embryonic heart) development by bioinformatic approach. The entire 4 002 bp *Ercc6l* cDNA (GenBank Accession number: AY172688) was firstly cloned in silico by assembling of overlapping murine expressed sequence tags (ESTs). These ESTs were retrived by BLAST algorithms. Repeated database searches were initiated with an EST (BB476905), which was from mouse 13-day embryo heart. Selection

* This work was supported by The Special Funds for Major State Basic Research of China (G1999055904), The National Nature Science Foundation of China (30030120) and the "985" Project of Peking University.

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

Received: October 31, 2003 Accepted: December 28, 2003

of the initial EST was performed in 13-day embryo heart cDNA library (RIKEN full-length enriched, 13-day embryo heart, dbEST Library ID. 5466, Unigene Lib. 428). Because ESTs of this library is well annotated by Digital Differential Display (DDD) (http://www.ncbi.nlm.nih.gov/unigene/info_ddd.shtml), it is easy to select ESTs that might represent novel genes. In this study, the initial EST (BB476905) was selected to begin the repeated searches and extensions in EST database and finally result in cloning cDNA of *Ercc6l*.

In order to verify the correct assembly of the cDNAs and confirm the nucleotide sequences, oligonucleotide primers were designed by usage of Primer 5.0 software to PCR amplify the entire coding sequences of *Ercc6l* from first-strand cDNA prepared from 14-day mouse embryo heart RNA. The primer sequences were 1F, 5′-GCG GTT TTA ACT GGG AGT CT and 1R, 5′- ATT CAA GGG AAG AAC TGC TG. PCR experiment was carried out in a reaction buffer (50 μl) containing 4 μl template, 20 pmol primer, 100 μmol/L each of dATP, dGTP, dTTP, dCTP, 1 unit elongase mix (Invitrogen). A touch-down PCR profile was performed as following: 94℃ 3 min; 20 × (94℃ 40 s; 60 ~ 50℃ 40 s, 68℃ 4 min), 15 × (94℃ 40 s; 50℃ 40 s, 68℃ 4 min). Afer electrophoresis on 1% agarose, the 3 849 bp fragment of PCR product was purified and cloned to pGEM-T easy vetor (Promega). The inserted fragment of recombined pGEM-T easy was sequenced by the primer walking technique using ABI PRISM 377 DNA Sequencer.

### 1.3 Sequence analysis

Nucleotide and amino acid sequence queries were done with BLAST algorithms against public databases, mainly at http://www.ncbi.nlm.nih.gov. Program ORF Finder was used in translating the cDNA. General protein domain and motifs was analyzed using SMART (http://smart.embl-heidelberg.de) and PROSCAN (http://npsa-pbil.ibcp.fr/). Multiple amino acid sequence alignments were performed using the ClustalX and Bioedit software.

### 1.4 Subcellular localization

To generate a mammalian expression vector, the open reading frame of *Ercc6l* was amplified using primer 2F(5′CTGT CTC GAG GCC ATG GAG GCT TCC CAA G) and 2R(5′CGC GGT ACC TTC TCA AAT GTT CAG CTG C). The PCR product was cloned into pGEM-T easy vector. After digested by *Xho* Ⅰ and *Kpn* Ⅰ, the ORF of *Ercc6l* was subcloned into the pEGFP-C3 vector (Clontech) to obtain an in-frame fusion protein between the green fluorescent protein (EGFP) and Ercc6l. The resulting construct was called pEGFP-Ercc6l. Then, three cell lines (HeLa, NIH3T3 and B16) were transfected by pEGFP-Ercc6l and pEGFP-C3 with Lipofectin2000 (Invitrogen) according to the manufacturer's instructions. The three cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Expression of fusion protein was examined by fluorescent microscopy. Among the three cell lines, 3T3 was originally derived from mouse normal embryonic fibroblast, HeLa was a human cervical cancer cell line, and B16 was a mouse melanoma cell line. The three cell lines were easy to be transfected, and moreover, all of them were from embryonic or tumorous tissues, consistent with the expression pattern of Ercc6l. Therefore, pEGFP-Ercc6l fusion protein expressed in these cells could substantially tell the subcellular localization of Ercc6l protein.

### 1.5 RT-PCR analysis of expression pattern

Reverse transcription PCR (RT-PCR) assay was performed on a panel of mouse multiple tissues cDNAs (Mouse MTC panel Ⅰ, Clontech) and the cDNA from various organs of different developmental stages using primers specific for *Ercc6l* (3F: 5′TCT CCC TTT CCA TTC TCA TCT GTG; 3R: 5′CCT CCT GTA TCT TCC CGC ACT C.) The primer pair was expected to amplify a segment of 502 bp corresponding to nt 3 185 ~ nt 3 686 of *Ercc6l* cDNA. PCR cycle was 94℃ 3 min, 36 × (94℃ 30 s; 56℃ 30 s; 70℃ 50 s); 70℃ 5 min. 5 μl of the reactions were separated on agarose gel containing ethidium bromide. Expression of housekeeping gene *G3PDH* was also assayed to check the mRNA abundance in different cDNA samples. The primer pair for *G3PDH* was: sense primer 5′ TGA AGG TCG GTG TGA ACG GAT TTG GC; antisense primer 5′ CAT GTA GGC CAT GAG GTC CAC CAC. PCR cycle for G3PDH was 94℃ 3 min; 25 × (94℃ 30 s; 68℃ 90 s); 68℃ 5 min.

## 2 Results

### 2.1 Cloning and analysis of *Ercc6l* cDNA

Repeated database searches initiated with an EST (BB476905) retrieved a large number of homologous EST sequences. Assembling of these EST produced a 4 002 bp cDNA sequence of *Ercc6l* (GenBank Acc. No AY172688). It contains an open reading frame (ORF) of 3 723 bp with a likely intiator ATG at nt 59 and stop codon at nt 3 781. The methionine residue at the beginning of the ORF is in a suitable context for the initiation of translation[7]. A polyA-signal cencus sequence ATTAAA occurs at nt 3 984. BLAST search to mouse genome mapped *Ercc6l* to mouse chromosome X. It composed of 2 extrons within 14 076 bp mouse genome. The 2 extrons are 126 bp and 3 876 bp respectively and separated by an intron of 10 074 bp.

Two turns of RT-PCR with primer 1F and 1R amplified a 3 849 bp segment from mouse 14-day embryo heart (Figure 1). Sequencing the insert fragment of recombined pGEM-T easy showed that the

3 849 bp segment was exactly the region of 4 002 bp cDNA between nt 9 to nt 3 857, containing the entire coding sequence of *Ercc6l*. This result experimentally confirmed the correct nucleotides sequence firstly cloned in silico.

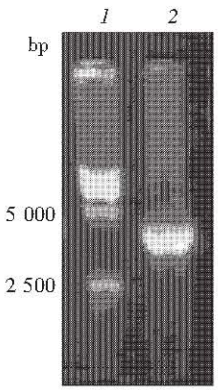


**Fig. 1 Results of PCR identification of 3 849 bp cDNA of *Ercc6l***

*1*: DNA molecular mass marker (DL15000, TaKaRa); *2*: PCR product.

## 2.2 Analysis of the predicted Ercc6l protein

The longest ORF of *Ercc6l* cDNA is capable of encoding a 1 240 amino acid (aa) polypeptide with a predicted molecular mass of 139 ku and theoretical p*I* of 5.4. Querying the general protein domain servers with the aa sequence revealed that Ercc6l protein contains a SNF2 family N-terminal domain (pfam00176, SNF2_N) and a helicase superfamily C-terminal domain (smart00490, HELICc). Proscan analysis revealed the presence of DEAD/H box, which is a conserved marker of DNA- and RNA- binding proteins classified as helicase in organism ranging from *Escherichia coli* to mouse. An A-type ATP-binding site, p-loop (AXXXXGKT) was found in tandem with DEAD/H box, indicating an ATP-dependent helicase activity. This is a common feature in DEAD/H box protein[5]. Eight highly conserved motifs, collectively considered as SNF2 domain, were present. All these features suggest that Ercc6l may be grouped to SNF2 family of DEAD/H box-containing helicase.

Eisen *et al*[8] divided SNF2 family into multiple subfamilies, each of which represented a functionally and evolutionally distinct group. Based on the method Eisen *et al*. had developed, a partial alignment was performed within SNF2 domain for Ercc6l against the representatives of subfamilies in SNF2 family (Figure 2). Less conserved regions (such as the regions flanking the SNF2 domain and the variable spacer regions) were not used, because of problem in obtaining unambiguous alignments in these regions. The result demonstrated that Ercc6l shared more similarity to ERCC6 than to others. Analyzing the sequence motifs and similarities in less conserved regions also identified its structural similarities to ERCC6. For example, we did not find bromodomain

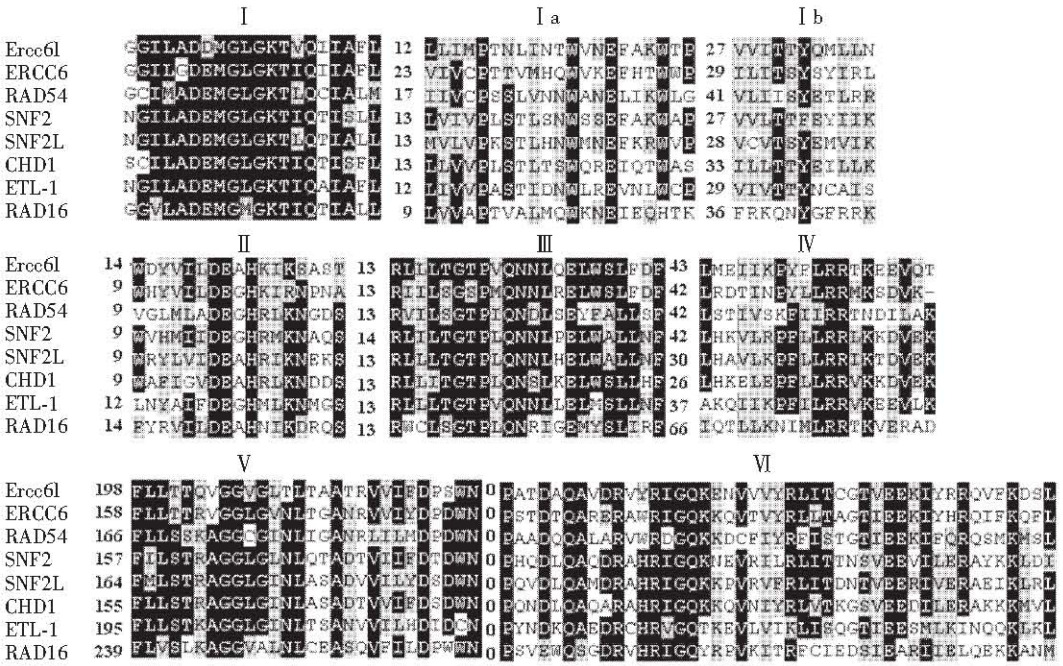


**Fig. 2 Comparison of Ercc6l with the representative members of subfamilies in SNF2 family**

Partial sequence alignment of Ercc6l and other SNF2 proteins, which are representative members of subfamilies of SNF2 family. Eight conserved regions are present, with the number of separating residues indicated between the blocked amino acids.

motif and RING finger-like motif in Ercc6l protein. According to these results and what Eisen *et al* had defined, we initially placed Ercc6l in ERCC6 subfamily, the members of which may function in transcription-coupled repair (TCR).

### 2.3 Subcellular localization of Ercc6l

The subcellular localization of fusion protein was examined by fluorescent microscopy 24 ~ 48 h after transfection. In all the three cell lines, EGFP alone was localized throughout the cells, however, the EGFP-tagged Ercc6l protein was excluded from nucleus and evenly distributed the cytoplasm of the transfected cells (Figure 3).

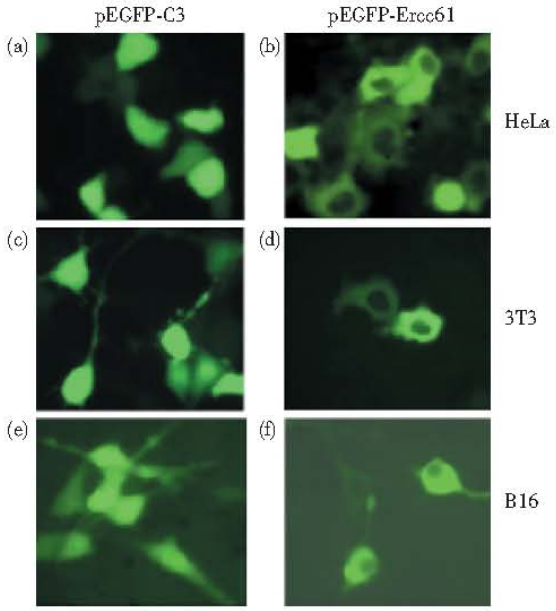


**Fig. 3 Subcellular distribution of EGFP tagged Ercc6l protein**

HeLa cells (a and b), 3T3 cells (c and d) and B16 cells (e and f) were transfected with pEGFP-C3 or pEGFP-Ercc6l respectively and analyzed by fluorescent microscopy 24 ~ 48 h after transfection.

### 2.4 Expression analysis of *Ercc6l*

#### 2.4.1 Digital expression profiles of *Ercc6l*

Publicly available EST collections are a valuable source of expression data. To some extent, counting the frequency of homologous ESTs from different cDNA libraries could reveal gene expression in different tissues or developmental stages[9]. BLAST to EST database retrieved 69 homologous mouse ESTs of *Ercc6l*. These ESTs were included in various cDNA libraries such as mammary, heart, brain, spleen, kidney and so on. Further more, we carefully analyzed the database records of these 69 homologous ESTs and found that 27 ESTs were from embryo, 7 ESTs were from neonate, 19 ESTs were from tumorous cell lines or tissues (including 12 ESTs from breast cancer cell lines Jyg MC, 2 ESTs from mammary infiltrating ductal carcinoma, 3 ESTs from melanoma cell line B16, and 2 ESTs from lung spontaneous tumor), and the remaining 16 ESTs were mainly from young mouse or various kinds of prolifering cell (Figure 4).

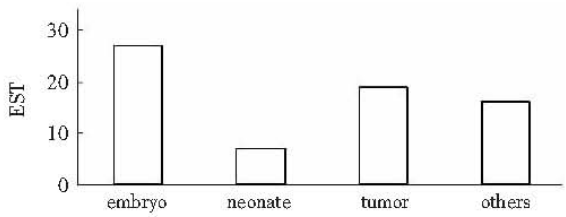


**Fig. 4 Results of counting the homologous ESTs of *Ercc6l***

#### 2.4.2 Expression profiles of *Ercc6l* analyzed by RT-PCR

RT-PCR analysis revealed that transcript of *Ercc6l* gene expressed strongly in spleen and lung, faintly in testis and not in other tissues of 8 ~ 12 weeks mouse. However, strong expression was detected in embryo especially during embryonic 11 ~ 15day (Figure 5a).

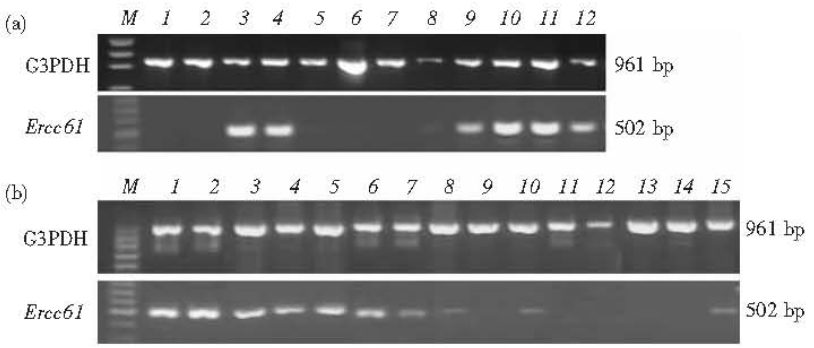


**Fig. 5 RT-PCR analysis of expression pattern of mouse *Ercc6l* trancript**

(a) RT-PCR was performed on MTC panel Ⅰ (Clonetech). Lane *1* ~ *8*: heart, brain, spleen, lung, liver, skeletal muscle, kidney and testis; Lane *9* ~ *12*: 7-day embryo, 11-day embryo, 15-day embryo and 17-day embryo. (b) RT-PCR was performed on the cDNA from five organs in different developmental stages. Lane *1* ~ *5*: heart, brain, kidney, liver, lung of 15-day embryo; Lane *6* ~ *10*: heart, brain, kidney, liver, lung of 15-day neonate; Lane *11* ~ *15*: heart, brain, kidney, liver, lung of 15-weeks adult.

In different developmental stages of heart, brain, kidney, liver and lung, *Ercc6l* transcript was abundant in embryo and significantly downregulated after birth and much less in adult (Figure 5b). It is noticeable that *Ercc6l* gene expressed strongly in lung of Clonetech MTC panel I (from 8~ 12 week mouse, Figure 5a), but weakly in lung of neonate and adult mouse (their cDNAs were prepared by ourselves as described in **materials and methods**). Therefore, the inconsistent expression of *Ercc6l* gene in lung might due to the difference of cDNA preparation.

## 3 Discussion

The number of proteins assigned to SNF2 has increased rapidly, but a large number of SNF2 protein are functionally unclear. Many SNF2 family members are major components of SWI/SNF complexes that can regulate transcription by remodeling chromatin structure[10]. Genetic studies in *Drosophila*, *Arabidopsis* and mice revealed crucial roles for SNF2 homologues in proper function and development of the whole organism, for example, *Drosophila Mi-2* involved in the control of polycomb gene repression *in vivo* is required for normal germ cell development[11]. In this study, analysis of expression pattern based on EST or RT-PCR suggests that relative expression levels of Ercc6l varied with tissues and significantly down regulated with developmental stages. So we think that Ercc6l may play a role in mouse embryo development. Another remarkable characteristic is that large number of homologous EST of *Ercc6l* also present in tumorous tissue and cell lines, implying that Ercc6l may implicate in tumor. In addition, using bioinformatic analysis, we found that human cDNA FLJ90238 (GenBank Acc No: AK074719) may be partial cDNA sequence of the putative human orthologue of *Ercc6l*; most of its homologous ESTs are also from tumors(such as germ cell tumors, leiomyosarcoma, large cell carcinoma, melanoma, ovarian tumor, neuroblastoma, and adenocarcinoma), prolifering cells and embryo tissues (data not shown). So we think that *Ercc6l* and its putative human orthologue are embryo- and tumor-specific at some extent. Now the function of *Ercc6l* in embryo development and tumor genesis is under extensive investigation in our laboratory.

Using molecular phylogenetic techniques that Eisen *et al*. had developed, we placed Ercc6l into the ERCC6 subfamily of SNF2 family (in fact, this is the reason why we name it as Ercc6l). So, according to what Eisen *et al*. had pointed out, Ercc6l may function similarly to the members of this subfamily. ERCC6, also known as CSB, is the representative and most clearly characterized member of ERCC6 subfamily. ERCC6 involve in transcription coupled repair (TCR) that is essential for integrity of genome. Defective ERCC6 has been linked to human Cockayne Syndrome complementation groups B (CSB)[12]. Recent years more and more studies reported that ERCC6 might play an important role in cancers[13,14]. However, the exact linkage remains to be fully explored. In this study, we showed that Ercc6l mainly expressed in embryonic or tumorous tissues and cell lines, which was characterized by rapidly proliferating. So, we previously assumed that Ercc6l might impact on embryo development and tumor susceptibility by maintaining genome integrity and regulating cell proliferation. Based on the function prediction, Ercc6l was firstly expected to work in nucleus as most of repair/transcription proteins including some SNF2 proteins. However, the EGFP-tagged Ercc6l protein was mainly found in cytoplasm of the three transfected cell lines. Therefore, it seems to be quite reasonable that Ercc6l may enter nucleus by a time- or condition-dependent way, such cases were found in other repair/transcription proteins such as XPD[15]. However, the possibility should also be considered that EGFP tag may cause mislocalization. Hence, where and how Ercc6l protein works remains to be further experimentally investigated.

## References


1 Suk K, Kim S, Kim Y H, *et al*. Identification of a novel human member of the DEAD box protein family. Biochim Biophy Acta, 2000, **1501** (1): 63~ 69

2 Wassarman D A, Steitz J A. Alive with DEAD proteins. Nature, 1991, **349** (6309): 463~ 464

3 Tanner N K, Linder P. DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell, 2001, **8** (2): 251 ~ 262

4 俊 江, 李龙芸, 卢光琇. DEAD/H 盒超家族新成员——人 DDX36 和小鼠 Ddx36 基因的分子克隆和特性. 生物化学与生物物理学报, 2002, **34** (5): 655~ 661
Fu J J, Li L Y, Lu G X. Acta Biochim Biophys Sin, 2002, **34** (5): 655~ 661

5 Bork P, Koonin E V. An expanding family of helicases within the "DEAD/H" superfamily. Nucleic Acids Res, 1993, **21** (3): 751~ 752

6 Luking A, Stahl U, Schmidt U. The protein family of RNA helicases. Crit Rev Biochem, Mol Biol, 1998, **33** (4): 259~ 296

7 Kozak M. Point mutationdefinea sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosome. Cell, 1986, **44** (2): 283~ 292

8 Eisen J A, Sweder K S, Hanawalt P C. Evolution of SNF2 family of proteins: subfamilies with distinct sequences and functions. Nucleic Acids Res, 1995, **23** (14): 2715~ 2723

9 钱 骏, 董 利, 张必成, 等. 表达序列标签数据库搜索鉴定小鼠 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 Biophy, 2002, **29** (2): 323~ 327

10 Peterson C L, Workman J L. Promoter targeting and chromatin

11 Kehle J, Beuchle D, Treuheit S, *et al*. dMi-2, a hunchback interacting protein that functions in polycomb repression. Science, 1998, **282**（5395）: 1897 ~ 1900
12 Troelstra C, van Gool A, De Wit J, *et al*. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell, 1992, **71**（6）: 939 ~ 953
13 Cheng L, Spitz M R, Hong W K, *et al*. Reduced expression levels of nucleotide excision repair genes in lung cancer: a case-control analysis. Carcinogenesis, 2000, **21**（8）: 1527 ~ 1530
14 Cheng L, Sturgis E M, Eicher S A, *et al*. Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. Cancer, 2002, **94**（2）: 393 ~ 397
15 Santagati F, Botta E, Stefanini M, *et al*. Different dynamics in nuclear entry of subunits of repair/transcription factor TF Ⅱ H. Nucleic Acids Res, 2001, **29**（7）: 1574 ~ 1581

# SNF2 家族新成员 Ercc6l 的 cDNA 克隆与表达分析*

陈祥贵[1,2] 李 勇[1]** 臧明玺[1] 裴新荣[1] 许雅君[1] 高丽芳[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]西华大学生物工程系，成都 610039）

摘要 SNF2 家族蛋白在基因组复制、修复与表达中具有重要作用. 报道了 SNF2 家族新成员 Ercc6l（excision repair cross-complementing rodent repair deficiency, complementation group 6-like）的 cDNA 克隆、特性与表达分析. 通过表达序列标签（EST）搜索和组装，获得了 cDNA 全长 4 002 bp 的新基因 *Ercc6l*（GenBank Acc. No AY172688），然后通过 RT-PCR 在小鼠胚胎心脏成功克隆了该基因. *Ercc6l* 在小鼠基因组中由两个外显子和一个内含子组成，定位于 X 染色体，最大开放阅读框（ORF）编码一个含 1 240 个氨基酸的假定蛋白质. 该假定蛋白质含有 SNF2 蛋白的 8 个保守基序（SNF2 结构域）. 通过与 SNF2 家族各亚家族的成员进行多重比对，初步确认 Ercc6l 属于 ERCC6 亚家族成员. 将 *Ercc6l* 编码区克隆到 pEGFP-C3 然后转染 HeLa，3T3 和 B16 细胞，融合蛋白主要定位于胞浆. BLAST 搜索检索出 69 条小鼠 EST 与 *Ercc6l* 同源，这些 EST 主要来自胚胎和肿瘤组织. 对小鼠不同发育时期的多种组织进行 RT-PCR，发现 *Ercc6l* 在胚胎期强表达，出生产后表达显著下调. 这些结果提示 *Ercc6l* 在胚胎发育和肿瘤发生中可能具有重要作用.

关键词 *Ercc6l*，基因克隆，SNF2 蛋白，表达谱，亚细胞定位
学科分类号 Q78

* 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”项目资助.
** 通讯联系人.
Tel：010-82801177，Fax：010-82801177，E-mail：liyong@bjmu.edu.cn
收稿日期：2003-10-31，接受日期：2003-12-28