

中缅树鼩三个多潜能基因 cDNA 片段的克隆和序列分析

王彩云^{1, #}, 马云瀚^{2, #}, 何大健³, 杨世华^{1,*}

1. 昆明理工大学 生命科学与技术学院, 云南 昆明 650500;

2. 昆明理工大学 医学院, 云南 昆明 650500;

3. 中国科学院昆明动物研究所, 云南 昆明 650223

摘要: 该实验从中缅树鼩 (*Tupaia belangeri*) 组织中成功克隆 *Klf4*、*Sox2* 和 *c-Myc* 基因的部分序列, 长度分别为 382、612 和 485 bp, 分别编码 127、204 和 161 个氨基酸。树鼩的 *Klf4*、*Sox2* 和 *c-Myc* 序列与人相应序列的相似性分别为 89%、98% 和 89%。将树鼩 *Klf4*、*Sox2* 和 *c-Myc* 基因分别与其他物种的这 3 个基因进行系统进化分析, 发现 *Klf4* 和 *Sox2* 基因所构建的系统进化树结构较为一致, 但与 *c-Myc* 所构建的系统进化树有所不同, 表明这些基因在进化上存在差异。*Klf4*、*Sox2* 和 *c-Myc* 基因的克隆为进一步研究其功能奠定了基础。

关键词: 树鼩; *Klf4*; *Sox2*; *c-Myc*; 序列分析

中图分类号: Q75 文献标志码: A 文章编号: 0254-5853-(2013)02-0127-05

cDNA cloning and sequence analysis of pluripotency genes in tree shrews (*Tupaia belangeri*)

Cai-Yun WANG^{1, #}, Yun-Han MA^{2, #}, Da-Jian HE³, Shi-Hua YANG^{1,*}

1. Faculty of Life Science and Technology, Kunming 650500, China;

2. Faculty of Medical, Kunming University of Science and Technology, Yunnan Kunming 650500, China;

3. Kunming Institute of Zoology, the Chinese Academy of Sciences, Yunnan Kunming 650223, China

Abstract: In this paper, partial sequences of the tree shrew (*Tupaia belangeri*) *Klf4*, *Sox2*, and *c-Myc* genes were cloned and sequenced, which were 382, 612, and 485 bp in length and encoded 127, 204, and 161 amino acids, respectively. Whereas, their cDNA sequence identities with those of human were 89%, 98%, and 89%, respectively. Their phylogenetic tree results indicated different topologies and suggested individual evolutional pathways. These results can facilitate further functional studies.

Keywords: Tree shrew; *Klf4*; *Sox2*; *c-Myc*; Sequence analysis

树鼩 (tree shrews, *Tupaia belangeri*) 被认为是最接近于灵长类的小型哺乳类动物, 已被广泛应用于医学和生物学实验研究中 (Xu et al, 2013)。诱导多能干细胞 (induced pluripotent stem cells, iPSc) 是指转录因子 *Klf4*、*Sox2* 和 *c-Myc* 等促使体细胞重编程 (Takahashi & Yamanaka, 2006; Takahashi et al, 2007), 使其成为具有多潜能性的干细胞。因为 iPS 细胞能形成嵌合体动物而体现其全能性, 掀起了人类再生医学研究的热潮。*Sox2* 基因所编码的转录因子对于维持胚胎神经嵴干细胞多能性具有重要作用

用 (Laga et al, 2010)。由 *Oct4* 和 *Sox2* 基因组成的调控复合物控制基因表达对于维持早期细胞发育非常重要 (Okumura-Nakanishi et al, 2005)。*C-Myc* 是一种原癌基因, 失去该基因后细胞的重编程进度及效率均会显著降低 (Shi et al, 2008)。*Klf4* 能够连接 *Nanog* 的启动子区域并直接调控其表达, 因此, 在防止胚胎干细胞 (embryonic stem cells, ESCs) 的分化中起重要作用, 同时, 它也是维护细胞自我更新和保持多能性的重要基因 (Zhang et al, 2010)。

小鼠及人类等的体细胞均可通过转录因子

收稿日期: 2012-12-17; 接受日期: 2013-01-16

基金项目: 科技部重大科学研究计划重大科学问题导向项目 (2012CBA01300, 杨世华), 教育部“新世纪国家优秀人才”支持计划项目, 国家自然科学基金项目(31071279, 30871232), 云南省科技创新人才计划项目(2011CI009)

*通信作者(Corresponding author), E-mail: yshhm@163.com

#共同第一作者(Authors contributed equally to the work)

Klf4、*Sox2*、*c-Myc* 及 *Oct4* 等诱导成为多能干细胞 (Takahashi & Yamanaka, 2006; Takahashi et al., 2007)。但树鼩多潜能转录因子的核酸序列至今未见报道, 影响了树鼩 iPSC 的研究, 因此, 本研究克隆并解析了树鼩的多潜能因子 *Klf4*、*Sox2* 和 *c-Myc* 等, 为研究树鼩 iPSC 及基因转录分析提供了重要数据。

1 材料和方法

1.1 实验材料

选取健康成体树鼩的骨髓和妊娠期~30 d 孕体的不同组织器官(肝脏、肠及大脑等)。

1.2 简并引物设计

通过 GenBank 公布的人(NM004235; NM002467; NM003106)、猕猴(NM001142793; NM001142873; NM001142940)、牛(NM001105385; NM001046074; BC133458)及猪等(EU669075; FJ882404; EU503117)的 *Klf4*、*Sox2* 和 *c-Myc* 基因序列, 用 Primier 5.0 软件设计简并引物 (Table 1)。

表 1 *Klf4*、*Sox2* 和 *c-Myc* 简并引物

Table 1 *Klf4*,*Sox2* and *c-Myc* degenerate primer

基因名称 Gene name	引物 Primer
<i>Sox2</i>	正向引物 forward primer: 5'-CGAGCGCTGCACATGAAGGA-3' 反向引物 reverse primer: 5'-GCCGCTCTGGTAGTGCTGGG-3'
<i>c-Myc</i>	正向引物 forward primer: 5'-GCTCCACCTCCAGCYTGTACCT-3' 反向引物 reverse primer: 5'-GACYCTGMCACTGTCCAATTTRG-3'
<i>Klf4</i>	正向引物 forward primer: 5'-GGCGGGAAAGGGAGAACMC-3' 反向引物 reverse primer: 5'-CCCGSATYGGATAGSTGAAG-3'

1.3 树鼩总 RNA 提取及简并 PCR 法扩增目的基因

使用 RNA prep pure tissue kit (TIANGEN 公司) 提取树鼩组织总 RNA, 提取方法按试剂盒说明书进行。

反转录反应实验按反转录试剂盒 PrimeScript™ 1st Strand cDNA Synthesis Kit (Takara公司)说明书进行。PCR 反应总体积为 25 μL: DNA 2 μL, 10× 缓冲液 2.5 μL, dNTPs 2 μL, 上、下游引物各 1 μL, easy TaqDNA 聚合酶 (5 U/μL) 0.2 μL, ddH₂O 16.3 μL。PCR 扩增程序: 95 °C 预变性 5 min, 95 °C 变性 45 s, 58~60 °C 退火 40 s, 72 °C 延伸 50 s, 共 35 个循环; 最后 72 °C 延伸 10 min。

PCR 产物通过 1% 的琼脂糖凝胶电泳分离并纯

化, 再将 *Klf4*、*Sox2* 和 *c-Myc* 纯化产物与 pMD18T 载体连接, 并转化大肠杆菌 DH5α 感受态细胞。37 °C 转化株在提供 X-gal/IPTG 的 LB 培养基中孵育 12 h。蓝白斑筛选阳性菌落, 且每个基因挑取 5 个不同的阳性菌落 37 °C 摆床培养 12 h, 挑取菌斑用于基因测序。测序由上海生工测序公司完成。

1.4 序列分析

所得序列经校对后, 进入 GenBank 数据库进行 Blast 分析, 同时通过在线软件 (<http://www.ncbi.nlm.nih.gov/orfig.cgi>) 对获得的树鼩 *Klf4*、*Sox2* 和 *c-Myc* 的部分序列进行氨基酸推导。然后从 GenBank 中下载不同物种的 *Klf4*、*Sox2* 和 *c-Myc* 基因, 利用 MEGALIGN 对序列进行排列, 并应用 Mega5.0 软件的邻接法(NJ)构建基因系统进化树。NJ 分析采用 Kimura two-parameter 距离法, 分支树可信度测试使用自展法(Bootstrap)检验, 经 1 000 次重复抽样检验得到分支树节点的支持率。系统进化树构建所用序列见图 4、图 5 和图 6。

2 结果

2.1 树鼩总 RNA 的获取

树鼩总 RNA 采用琼脂糖凝胶电泳法验证质量(图 1), 所选取的两个样品的 rRNA 三条带清晰, 表明提取的 RNA 具有良好的完整性, 并适用于后续实验。

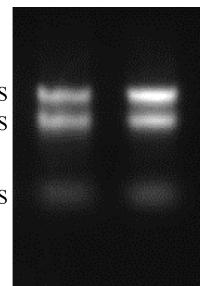


图 1 树鼩组织总 RNA

Figure 1 Total RNA extracted from tissues of tree shrews

2.2 目的片段的检测

RT-PCR 产物的琼脂糖凝胶电泳显示, 目标条带与预期目的基因大小一致(图 2)。

2.3 树鼩 *Klf4*、*Sox2* 和 *c-Myc* 基因部分片段的序列分析

获得的 *Sox2*、*c-Myc* 和 *Klf4* 部分序列长度 [GenBank 登录号分别为 *Klf4* (K1 KC480548)、*Sox2* (S1 KC480549) 和 *c-Myc* (C1 KC480547)] 与预

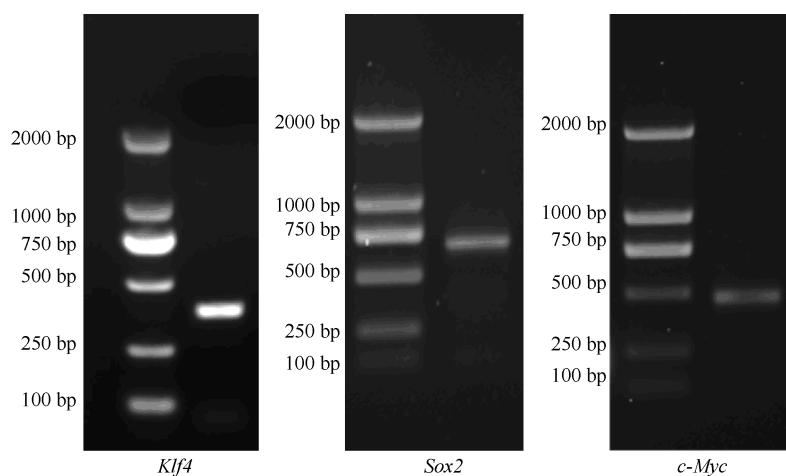


图 2 树鼩 *Klf4*、*Sox2* 和 *c-Myc* 的 RT-PCR 扩增结果
Figure 2 Amplification results of *Klf4*, *Sox2*, and *c-Myc* genes by RT-PCR

期结果一致, 分别为 612、485 和 382 bp(测序结果及推导的氨基酸序列见图 3)。树鼩 *Klf4* 部分序列与人 (*Homo sapiens*)、猕猴 (*Macaca mulatta*)、牛 (*Bos taurus*)、猪 (*Sus scrofa*) 以及鼠 (*Mus musculus*) 等相应序列的相似性分别为 89%、90%、90%、89% 和 86%, 编码 127 个氨基酸; *Sox2* 部分序列与人、猕猴、牛、猪及鼠等相应序列的相似性分别为 98%、97%、96%、97% 和 95%, 编码 204 个氨基酸, 且该氨基酸序列与猕猴的 *Sox2* 氨基酸序列仅存在 1 个氨基酸的差别, 相似性高达 99%; *c-Myc* 部分序列与人、猕猴、牛、猪及鼠等相应序列的相似性分别为 89%、89%、90%、92% 和 87%, 编码 161 个氨基酸。

2.4 树鼩 *Klf4*、*Sox2* 和 *c-Myc* 基因部分片段的进化分析

在基于 *Klf4* 部分序列以及推导氨基酸序列构建的系统树中(图 4), 各分支具有较高的置信度, 说明这些动物之间的系统进化关系具有较高可信度。由系统树可见, 人 (*Homo sapiens*)、黑猩猩 (*Pan troglodytes*) 和猕猴 (*Macaca mulatta*) 等聚为一类(灵长类); 牛 (*Bos taurus*)、羊 (*Ovis aries*) 和猪 (*Sus scrofa*) 等聚为一类(偶蹄类); 小家鼠 (*Mus musculus*) 和褐家鼠 (*Rattus norvegicus*) 等聚为一类(啮齿类)。其中树鼩以较高的支持率与灵长类聚为一支。同样, 在基于 *Sox2* 部分序列构建的系统树中(图 5), 各类动物聚类情况与 *Klf4* 类似, 树鼩同样与灵长类聚为一支, 但支持率有所降低。由于不同类群动物的 *Sox2* 基因推导的氨基酸同源性非常高, 系统分

析时各分支支持率很低(结果未给出), 因而未给出其氨基酸序列构建的系统树。在以 *c-Myc* 部分序列以及推导氨基酸序列构建的系统树中(图 6), 灵长类、偶蹄类和啮齿类等虽各聚为一支, 但树鼩以较高的支持率与偶蹄类动物聚为一支。

3 讨 论

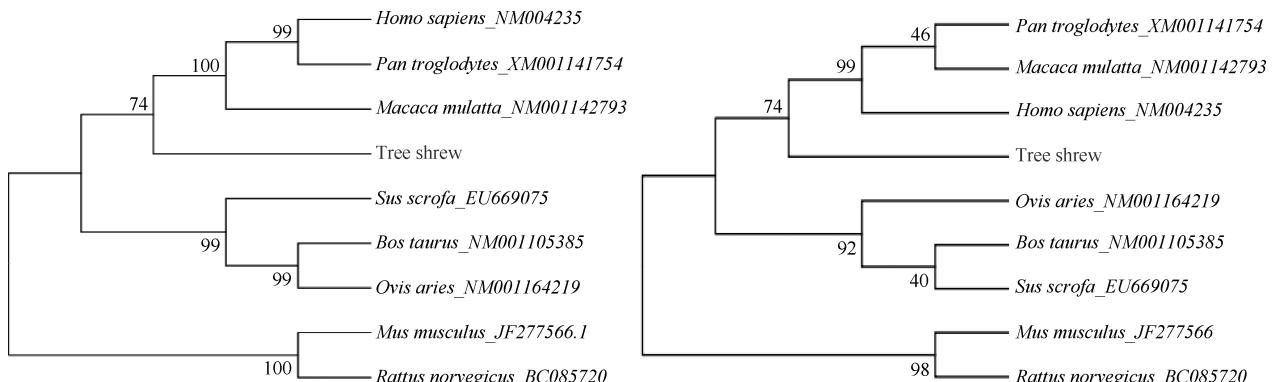
树鼩的分类学及生物学特征研究较为广泛, 已证明其许多生物学特性可以作为人类疾病模型的重要材料 (Xu et al, 2013; Ping et al, 2012; Zhang et al, 2012)。本研究结果为进一步探索树鼩的生殖工程和疾病模型提供了重要数据。

人 (*Homo sapiens*) 的 *Klf4*、*Sox2* 和 *c-Myc* 全长分别为 2 949、2 520 和 2 379 bp, 本研究获得的 *Klf4*、*Sox2* 和 *c-Myc* 部分序列长度分别为 382、612 和 485 bp, 分别对应于人相应基因的 675~1 060 bp、729~1340 bp 和 1082~1566 bp。基于 *Klf4*、*Sox2* 和 *c-Myc* 部分序列的进化分析支持偶蹄动物与灵长动物的亲缘关系, 较与啮齿动物的亲缘关系更为接近, 与动物种系进化关系相一致。在基于 *Klf4* 和 *Sox2* 部分序列构建的基因树中, 树鼩与灵长类动物的亲缘关系更近, 而与偶蹄动物和啮齿动物的亲缘关系较远。但在 *c-Myc* 基因系统进化树中, 树鼩与偶蹄动物的亲缘关系更近。这说明 *Klf4*、*Sox2* 与 *c-Myc* 基因可能在进化方式或速度上存在一定的差异。另外, 由于本研究中暂未获得这几个基因的全长序列而仅克隆了较为保守的部分序列, 利用部分

```

ggcgggaaggagaagaccctgcccggcagggtgtccgaataaccgctggggaggagctctcg
A G R E K T L R P A G V P N N R W R E E L S
cacatgaagcacttccccgggtgtccctggccgccttaacgacctggcgccgacagtggccaca
H M K R L P P V L P G R P Y D L A A A T V A T
gacctggagaacggaggagccggcggttcggcgtaacaacccggcgtttacccggagagaa
D L E N G G A G A A C G G G N N P A L L P R R E
acggaggagttcaacgagctctggactttatcctctcaactcgctgtcccattccggatcg
T E E F N E L L D L D F I L S N S L S H P E S
gtccccccaccgtgtctcggtcagecctgtctcgctccggcggcggcacc
V P P T V S S S A S A S S S S S P S S S G P T
agcgtccctccacactcgatccgtatccgtcg
S A P S T C S F S Y P M R
cgagcgctgcacatgaaggagcacccggattataaccggccgggaaaaccaagacgctc
R A L H M K E H P D Y K Y R P R R K T K T L
atagaaggataagtacactctggccggcggctgtggcccccggcggcggcggcggcggc
M K K D K Y T L P G G L L A P G G N S M A S
gggtcggtggcggccggcggctggcgccggcgtgaaccaggcatggacagcta
G V G V G A G L G A G V N Q R M D S Y A H M
aacggctggagcaacggcagtcacageatgtatcgaggaccgtggctatccgcagcaccggc
N G W S N G S Y S M M Q D Q L G Y P Q H P G
cttaacgcgcacccggcggcggcggatgcagccatgcaccgtacgcgtgagcgcctgag
L N A H G A A Q M Q P M H R Y D V S A L Q Y
aactccatgagcagctcgccatgcacatgcacccgtcgcccatgcac
N S M S S S Q T Y M N G S P T Y S V S Y S Q
cagggcaccccccgcattggcgtcgccatggctggcgtggtaagtccgaggcagctccgc
Q G T P G M A L G S M G S V V K S E A S S G
ccccctgtgtttactctccactccaggccgcctgcacccgggacactccgggacatg
P P V V T S S S H S R A P C Q A G D L R D M
atagcatgtaccccccggcggcggatgtccggagccgcggcccccagcagacttcacatgtcc
I S M Y L P G A E V P E P A A P S R L H M S
cagactaccagagcgc
Q H Y Q S G
Sox2
gctccacccctcagctgtacctgcaggacactgagcggccgcctccgagtgcacccctcggt
S T S S L Y L Q D L S A A A S E C I D P S V
gtcttccctaccgcgtcaacgcggcagctcgcccaagccctgcgcctgcggactccaccgccttc
V F P Y P L N D G S S P K P C A S P D S T A F
tcccccgtctcgactccctgctctccaccaggactctccggccggccagccccggcggc
S P S S D S L L S T E S S P R A S P E P L A
ctccacgaggagacacccaccaggcagcacttgaggaagccaagggatgagaaatt
L H E E T P P T T S S D S E E D Q E D E E E I
gatgtcggttctgttagaaaagaggcggccctaccaaaaggctggatcgggtcaccctccactgga
D V V S V E K R Q P P T K R S E S G S P S T G
ggcacacagcaaaccctccacagtccactggttcaagggatggccatgtctccaccatcagcacaat
G H S K P P H S P L V L K R C H V S T H Q H N
tacccgcaccaccctccaccaggactaccaggctgccaagggccaagggtggacagtggcaga
Y A A P P S T R K D Y P A A K R A K L D S G R
gtc
V
c-Myc

```

图3 树鼩 *Klf4*、*Sox2* 和 *c-Myc* 基因部分序列及推导的氨基酸序列Figure 3 Partial nucleotide and predicted amino acid sequences of *Klf4*, *Sox2*, and *c-Myc* genes in tree shrews图4 基于 *Klf4* 基因部分序列以及推导氨基酸序列构建的系统进化树Figure 4 Phylogenetic tree of nucleotide sequences (a) and deduced amino acid sequences (b) of the *Klf4* gene

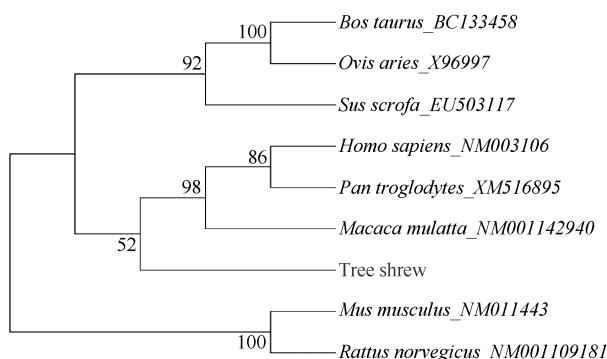


图 5 基于 *Sox2* 基因部分序列构建的系统进化树
Figure 5 Phylogenetic tree based on partial nucleotide of the *Sox2* gene

序列进行进化分析, 可能不能很准确反应这些类群的真实亲缘关系。

由于目前 GenBank 数据库中尚未收录树鼩的

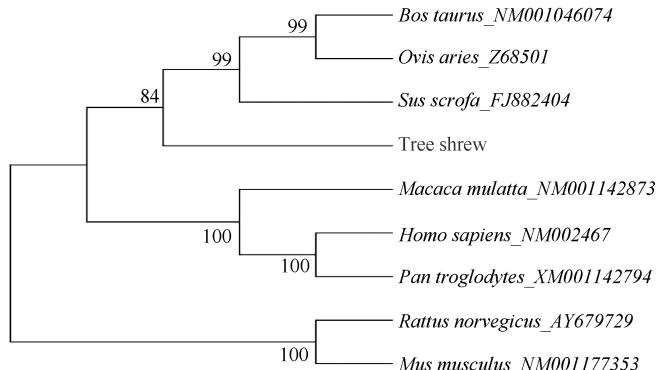


图 6 基于 *c-Myc* 基因部分序列以及推导氨基酸序列构建的系统进化树
Figure 6 Phylogenetic tree of nucleotide sequences (a) and deduced amino acid sequences (b) of the *c-Myc* gene

参考文献:

- Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM, Jaenisch R. 2007. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*, **318**(5858): 1920-1923.
- Laga AC, Lai CY, Zhan Q, Huang SJ, Velazquez EF, Yang QH, Hsu MY, Murphy GF. 2010. Expression of the embryonic stem cell transcription factor SOX2 in human skin: relevance to melanocyte and merkel cell biology. *The American Journal of Pathology*, **176**(2): 903-913.
- Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F. 2005. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *The Journal of Biological Chemistry*, **280**(7): 5307-5317.
- Ping SH, Wang CY, Tang WR, Luo Y, Yang SH. 2012. Effects of some extenders and monoamines on sperm cryopreservation in tree shrews (*Tupaia belangeri*). *Zoological Research*, **33**(1): 19-28. [平述煌, 王彩云, 唐文如, 罗瑛, 杨世华. 2012. 几种冷冻稀释液与单胺类防冻剂对中缅树鼩精子冷冻存活率的影响. 动物学研究, **33**(1): 19-28.]
- Shi Y, Desponts C, Do JT, Hahm HS, Scholer HR, Ding S. 2008. Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell*, **3**(5): 568-574.
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**(4): 663-676.
- Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**(5): 861-872.
- Xu L, Zhang Y, Liang B, Lv LB, Chen CS, Chen YB, Zhou JM, Yao YG. 2013. Tree shrews under the sp of light: emerging model of human diseases. *Zoological Research*, **34**(2): 59-69. [徐林, 张云, 梁斌, 吕龙宝, 陈策实, 陈勇彬, 周巨民, 姚永刚. 2013. 树鼩实验动物和人类疾病的树鼩模型研究概述. 动物学研究, **34**(2): 59-69.]
- Zhang PL, Andrianakos R, Yang Y, Liu CM, Lu WG. 2010. Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *Cell*, **126**(4): 663-676.
- Zhang YX, Ping SH, Yang SH. 2012. Morphological characteristics and cryodamage of Chinese tree shrew (*Tupaia belangeri chinensis*) sperm. *Zoological Research*, **33**(1): 29-36. [张远旭, 平述煌, 杨世华. 2012. 中缅树鼩精子形态特征及冷冻损伤. 动物学研究, **33**(1): 29-36.]

Klf4 和 *Sox2* 基因序列, 在一定程度上限制了树鼩的分子生物学研究。本研究采用简并 PCR 扩增, 成功克隆了树鼩 *Klf4*、*Sox2* 和 *c-Myc* 基因 cDNA 的部分序列, 为克隆这三种基因的全长提供了良好基础。

另外, 以往的研究认为 iPS 在小鼠疾病模型及胚胎干细胞研究中具有相似的功能 (Hanna et al, 2007)。因此, 我们假设可以通过导入树鼩的 3 种基因 *Sox2*、*c-Myc* 和 *Klf4* 来诱导树鼩多潜能干细胞。推测该模型的评估治疗效果较小鼠模型将更加可行。目前关于树鼩的这 3 个基因功能的研究还未见报道, 本研究首次对树鼩 *Klf4*、*Sox2* 和 *c-Myc* 三个基因 cDNA 部分片段进行了克隆和序列分析, 旨在为进一步研究诱导树鼩多潜能干细胞奠定基础。