

# 合浦珠母贝 iPS 细胞相关的转录因子的克隆及分析

张若云, 徐广锐, 潘 聪, 谢莉萍, 王洪钟, 张贵友, 张荣庆

(清华大学 生命科学学院, 北京 100084)

**摘要:** 转录因子 *Sox2*、*Oct4*、*c-Myc* 和 *KLF4* 在 iPS 细胞研究中具有重要作用, 将它们加入体外培养的细胞中, 可以诱导体细胞去分化成为多能干细胞。作者通过 RACE-PCR 技术从合浦珠母贝 (*Pinctada fucata*) 的外套膜组织中克隆获得了 3 个转录因子——*c-Myc*、*Sox2* 及 *KLF4* 的 cDNA。*c-Myc* 全长 1585bp, 开放阅读框的长度为 1131bp, 编码 376 个氨基酸, 蛋白结构分析表明它具有保守的 HLH 和 Myc-N 结构域。*Sox2* 全长 1908bp, 开放阅读框长度为 990bp, 编码 329 个氨基酸, 蛋白结构分析它具有保守的 HMG-box 和 Soxp 结构域。*KLF4* 全长 2268bp, 开放阅读框长 624bp, 编码 207 个氨基酸, 预测该蛋白具有两个 zf-H2C2\_2 结构域。BLAST 分析它们均与太平洋牡蛎的相关蛋白具有较高同源性, 说明其与太平洋牡蛎亲缘关系更近。RT-PCR 实验发现这 3 个基因在合浦珠母贝外套膜、足、生殖腺、内脏团、闭壳肌和鳃 6 种组织中均有表达, 表明它们是非常保守的转录因子。本研究对于合浦珠母贝细胞系建立和研究具有启发意义。

**关键词:** 诱导多功能干细胞; 转录因子; 克隆; 合浦珠母贝 (*Pinctada fucata*)

中图分类号: S823.3 文献标识码: A 文章编号: 1000-3096(2014)04-0007-08

doi: 10.11759/hyhx20130922001

胚胎干细胞(embryonic stem cells, ESCs)具有高度的分化潜能, 可以分化为机体所有类型的细胞, 在细胞治疗、组织器官移植等临床医学领域具有广泛的潜在用途<sup>[1]</sup>。诱导的多功能干细胞(induced pluripotent stem cells, iPS cells)具有与 ESC 相似的高度分化潜能, 由于它是由成体细胞诱导转化而来, 因此, 不仅规避了使用人类胚胎的伦理争议, 也可以使用病人自体细胞, 设计更加个性化的治疗方案。

自从日本学者 Yamanaka 领导的课题组先后将 4 个转录因子 *Sox2*、*Oct4*、*c-Myc* 和 *KLF4* 导入体外培养的小鼠和人的成纤维细胞, 使其去分化, 并转化为 iPS 细胞<sup>[1-2]</sup>后, iPS 细胞在全球范围内得到广泛应用。

由于这些转录因子参与维持干细胞的多能性和自我更新的能力, 故而又被称作多能性因子(Pluripotency factors)。多能性因子通常具有 3 个特点: (1)通常在胚胎发育早期表达量较高, 而在大多数已分化的组织中表达降低; (2)维持干细胞的多能性, 其表达量过高或过低均会引起胚胎干细胞分化; (3)多能性因子可协同作用<sup>[3]</sup>。

这 4 种多能因子在 iPS 细胞中起着至关重要但各不相同的作用。

*Oct4* 在干细胞、早期胚胎细胞和生殖细胞中特

异表达, 且主要通过与其他重要的转录因子, 如 *Sox2* 和 *Nanog* 等, 协同作用调节胚胎干细胞的多能性<sup>[4-5]</sup>。*c-Myc* 基因的表达与细胞的增殖分化有关, 例如, *c-Myc* 基因的表达量增强可以促进细胞生长, 导致肿瘤发生, 而在细胞分化的过程中, *c-Myc* 基因的表达量则会降低<sup>[5-6]</sup>。*Sox2* 基因在胚胎早期发育中发挥重要作用, 可以作为多能性细胞谱系的一个标志物<sup>[5]</sup>。在体外, *Sox2* 在未分化的胚胎干细胞(ES cell)、胚胎癌细胞(EC cell)中表达, 且随着细胞的分化, 其表达量降低<sup>[7-9]</sup>。*KLF4* 同样也在细胞增殖和分化过程中发挥重要作用, 可以通过作用于靶基因 *p21* 异位表达, 从而抑制细胞的生长和增殖, 也可在 *p21* 缺失的细胞中, 通过下调 *p53* 来促进细胞增殖<sup>[4]</sup>。

本实验室之前的研究证明合浦珠母贝外套膜组织在体外培养时功能与体内外套膜细胞类似, 由此为在细胞水平上研究珍珠质生物矿化提供了新方法<sup>[10]</sup>。但是合浦珠母贝外套膜细胞体外增殖和传代培养一直是体外细胞系建立与培养的难题<sup>[11-13]</sup>。此次克隆得到合

收稿日期: 2014-01-22; 修回日期: 2014-05-12

基金项目: 国家重点基础研究发展计划项目(2010CB126405)

作者简介: 张若云(1987-), 女, 江苏徐州人, 硕士, 主要从事生物矿化研究, E-mail: zhangry61@126.com

浦珠母贝中的转录因子 *Sox2*、*KLF4* 及 *c-Myc* 基因,一方面为合浦珠母贝中 iPS 细胞研究奠定基石;另一方面考虑到这几个转录因子在细胞增殖和分化过程中所起的重要作用,或许也可以为合浦珠母贝建立细胞系提供支持,为生物矿化研究提供新的思路。

## 1 材料与方 法

### 1.1 材料

合浦珠母贝(*Pinctada fucata*)采自广西北海市珍珠总公司珍珠养殖场。5'/3'RACE Kit(Clontech 公司), EasyPure Quick Gel Extraction Kit(TransGen 公司), dNTPs、Reverse Transcriptase M-mLV、各种酶(TaKaRa 公司), Trizol 试剂(Invitrogen), DEPC(上海生工), 及 IPTG、X-gal(Sigma); 其他试剂为国产分析纯,各种引物由上海生工合成, DNA 测序由华大基因完成。

### 1.2 方法

#### 1.2.1 引物设计

根据本实验室构建的合浦珠母贝外套膜转录组文库中 *c-Myc*、*Sox2* 及 *KLF4* 基因的同源序列片段设计 5'RACE 和 3'RACE 引物,基因克隆所用引物序列见表 1。

表 1 *c-Myc*、*Sox2* 及 *KLF4* 的 cDNA 克隆所需引物  
Tab.1 The primers for cloning cDNA of *c-Myc*, *Sox2* and *KLF4* from *Pinctada fucata*

RACE-PCR 引物	序列 (5'-3')
myc-GSP1	AGGCAGCACAAGTAGCAGGCAGTATG
myc-NGSP1	ACGGTAACTGTCGTTTATAGCGGGAGA
myc-GSP2	CCCGCTAAAACGACAGTTACCGTGAAAG
myc-NGSP2	CCCGCTAAAACGACAGTTACCGTGAAAG
sox-GSP1	GGCACTGTATTGTTACCGCTGTTC
sox-NGSP1	GGCATCATCGGGTATCCGTTGGG
sox-GSP2	CGGGAGCCCAATCAGCCAGTT
sox-NGSP2	CGTCGGAAACAGCGGTGAACAATAC
klf-GSP1	AGCGCAAAGCGAGACTGAACCT
klf-NGSP1	TGCTGAAGGCGGTGGTAGTGGG
klf-GSP2	TGGTCCGTTTCGGCTGCTATTCC
klf-NGSP2	CCCACCTCAAAGCCCATCTACGC

在得到 *c-Myc*、*Sox2* 及 *KLF4* 基因 5'和 3'端序列的基础上设计 Full1 和 Full2 两端引物。

RT-PCR 所需引物见表 2。

#### 1.2.2 总 RNA 提取

取合浦珠母贝的外套膜组织,用液氮迅速冷冻

表 2 RT-PCR 和 Real-Time PCR 所需引物  
Tab.2 The primers for RT-PCR and Real-Time PCR

PCR 引物	序列(5'-3')
myc-F	ACCTCACACTGCCACTGACA
myc-R	ACAGCTGCAGGATCCACACA
sox-F	ACGGGAAATGGAAGCGGAGA
sox-R	GCGTTCATAGGCCGTTTCACA
klf-F	TCCCATGAACATTCCCAGCA
klf-R	ACGGCGCGTTCATCTGA

后磨碎,加适量 Trizol 试剂(50~100 mg/mL)。总 RNA 提取步骤如下:样本中加入 200  $\mu$ L 氯仿,混匀后室温放置 3 min,离心 15 min 取上清;重复上述步骤一次;上清加入 500  $\mu$ L 异丙醇,混匀静置 10 min,离心 15 min 弃上清;加入 1mL 75%乙醇,振荡混匀离心 5 min;沉淀室温晾干,加入 40  $\mu$ L DEPC 水溶解,-80 $^{\circ}$ C 保存。通过  $A_{260/280}$  检验获得的 RNA 的完整性和浓度。

#### 1.2.3 基因的克隆

5'RACE 和 3'RACE 按照 SMARTer<sup>TM</sup> RACE cDNA Amplification Kit(Clontech)试剂盒进行:首先以合浦珠母贝外套膜的总 RNA 为模板,合成 5'-RACE-Ready cDNA 和 3'-RACE-Ready cDNA。然后分别以这两个 cDNA 序列为模板,进行 5'RACE 和 3'RACE。PCR 反应程序参照试剂盒说明书。

根据得到的 5'端和 3'端的两段序列进行全长基因的 PCR 验证。PCR 反应程序:94 $^{\circ}$ C 预变性 5 min;94 $^{\circ}$ C 变性 30 s,55 $^{\circ}$ C 退火 30 s,72 $^{\circ}$ C 延伸 3 min,30 个循环;72 $^{\circ}$ C 延伸 10 min。

PCR 产物进行 1.2%琼脂糖凝胶电泳,切下目的条带,回收 PCR 产物。将回收产物与 T/A 克隆载体 pGEM-T easy vector (Promega)连接,转化大肠杆菌 Trans2-Blue Chemically Competent Cell (TransGen)菌株,通过菌落 PCR 选取阳性克隆送公司测序。

#### 1.2.4 基因的序列分析

应用 ORF Finder 查找开放阅读框;用 ProtParam (<http://web.expasy.org/protparam/>)预测编码蛋白的分子量和等电点;应用 SignalP(<http://www.cbs.dtu.dk/services/SignalP/>)预测信号肽;并且使用 Pfam(<http://pfam.sanger.ac.uk/search/sequence>)预测蛋白保守结构域。应用 blast 在线工具(<http://web.expasy.org/blast/>)比对核苷酸和氨基酸序列的同源性,使用 MEGA4 软件绘制系统进化树。

#### 1.2.5 基因的组织分布

首先采用 TaKaRa 公司的 PrimeScript<sup>TM</sup> RT

Master Mix(Perfect Real Time)试剂盒, 按照说明书条件操作, 将合浦珠母贝外套膜、足、生殖腺、内脏团、闭壳肌和鳃 6 种组织总 RNA 反转录合成 cDNA 模板。然后通过 RT-PCR 鉴定目的基因在各组织分布情况。

## 2 结果与分析

### 2.1 *c-Myc*、*Sox2* 及 *KLF4* 基因的克隆及其序列分析

通过 RACE 方法克隆获得 *c-Myc*、*Sox2* 及 *KLF4*

基因的全长 cDNA, genbank 登录号分别为 KC758857、KC758858 和 KC758859。

基因 *c-Myc*、*Sox2* 及 *KLF4* 的全长 cDNA 序列及其编码蛋白的氨基酸序列、相对分子质量、等电点等信息如表 3 及图 1~图 3 所示。

### 2.2 编码蛋白的结构预测分析

通过 SignalP、Pfam 和 SOPMA 在线分析得到以下结果。

*c-Myc* 蛋白没有信号肽, 不存在跨膜区。Pfam

表 3 *c-Myc*、*Sox2* 及 *KLF4* 的 cDNA 序列及其编码的蛋白质性质

Tab.3 Sequences and proteins of *c-Myc*, *Sox2* and *KLF4* from *Pinctada fucata*

基因	cDNA 全长	开放阅读框		编码蛋白		
		位置	长度	长度	相对分子质量	等电点
<i>c-Myc</i>	1585bp	93~1223bp	1131bp	376aa	42.8ku	5.98
<i>Sox2</i>	1908bp	372~1361bp	990bp	329aa	44.4ku	9.69
<i>KLF4</i>	2268bp	805~1428bp	624bp	207aa	24.2ku	10.11

```

1  gtgagtgttttcaaagtgaactttattcatctctttttcatcttttagcttgcctgtgcaaaatcgtaatcgtcgcacacatcaaagaaaagaag
93  atggactacgatacatttcagccttgctttatgaaaaacgagacagagctgtcacctgtaggatcatatccaaatgaagacatgtggaag
1  M D Y D T F Q P C F Y E N E T E L S P V G S Y F N E D M W K
183 aagtttgagttgcttcccaccccacccctttcccacagagagatgaccagtttgattctaccctggattggagaacctcacactgccca
31  K F E L L P T P P L S P Q R D D Q F D S T L D L E N L T L P
273 ctgacagattacattctgaatgatgatgatgcctcgcattttgagaaaaatgacttttccggaatcacctccttcgctgatttacat
61  L T D Y I L N D D D D A S L F E K M T F P E S P P S P S L H
363 tctaaattaattcaggattgtatgtggtcaggtgattctggacgttacaagacaggagacaagctgtcaacatcactggatgcaactcc
91  S K L I Q D C M W S G D S G R Y K T G D K L S T S L D V N S
453 tcagattgtgtggatcctgcagctgtcttcccataccccatgcatacactgctgctacatgagtgaacaaagctacacaacctgggt
121  S D C V D P A A V F P Y P M H S I T G Y M S E T K L H N L G
543 actgaaacaccttctgattcagaagaggagattgatgttgcactgttgaacccaacaacaagaagttccagctgtatcaggcaaaagg
151  T E T P S D S E E E I D V V T V E R Q Q Q Q V P A V S G K R
633 aaagctactcttctcatccaaagtgttaaaactggaccatcatccgaattgtcatctcccgtaaaacgacagttaccgtgaaagtga
181  K A T L L I Q S V K T G P S S E L S S P A K T T V T V K V K
723 gtagacagaccactgcagcaaacagagcagccaaggagagttgagtgctccgactgatgtacataactattgtcaaccttcaaacgtgtt
211  V D R P L Q Q T E Q P R R V E C P T D V H N Y C Q P F K R V
813 cgtcaatgccatcttcagcaagcagttcaccatcacattcaaaacgttttaaacgggaattaagtgtaccagaatttcataaacgtgta
241  R S M P S S A S S S P S H S K R F K R E L S V P E F H K R V
903 tcacagaaactacattcatcacatcgtatccgttcacactattgtagtagcaatacgtcaagtagaacatcatcagactctgaagattcc
271  S Q K L H S S H R I R S H Y C S S N T S S R T S S D S E D S
993 ccagaggcgaagaggagtcacacaaatgacttggaaactgaaacgtagaatgacctcaaataggcttcttctcctcaggatagata
301  P E G K R S Q H N D L E R K R R N D L K F S F F A L R D S I
1083 ccagacttaatgaacaggaacgtgcgcctaaagtctcattcttaaaaaggcatctgaacacatcagaaaactgggtcatgaagaacag
331  P D L M K Q E R A P K V L I L K K A S E H I R K L G H E E Q
1173 agacttgtagaagataaggaatgttggtgcaaaacacgagcagttgtaaggactttagagcatctacgtgctagagaattttaattgg
361  R L V E D K E M L V Q N T S S C * 376
1263 actcgatgatccaaactgtttactgcactgtgacaactgtaggccttgatcagtggttcaaatggcctacccaatgtgttcaaa
1353 tacattttatttctgtatcacatttggtgtgtctcattcaaatggcatttgatcactgtacatcttagtgcttgcattacatactgcc
1443 gcagcttgtctgctttatataatgaccaattttttgtgtgctgagctgttgtagcagtaactaagcagcttgacttgttctgtat
1533 ttctacatttcatttctatgcaaattttcatcagtaaatccattcaaaaaa1585
    
```

图 1 *c-Myc* 基因全长 cDNA 序列及其编码蛋白的氨基酸序列

Fig.1 The complete cDNA sequence and deduced amino acids of *c-Myc* gene from *Pinctada fucata*

```

1   ttctctcgtattgtgtgacattttggttaagctccgccaccctcgatttcaactgtactgatcgtagttcaagccaagtcatttcag
91  cggatttgtgacaacaaaattagcatggaataatttgattggctgagtttgcgtctatttgttcactaccaaatcaatatcacatgttca
181 gtaaaagcatattagccaaaagaataagaagtacaggggtattcaagttctgcattattaccggacagtgaaggactaaaagatttata
271 tttttatgagggctgaattgggtgtgaactttattcttatctaacgacaatacactagttgacttttgggaatttattgaactttc

361 ttccacgaactatggagatcaccaacgggaaatggaagcggagatgcccaaggccccggcgagacagctctaccoccatccagtcag
1   M E Y H Q R E M E A E M P K A P G Q T A L P P S S Q

450 ggtatgcacgtcccagtgaccggaccaggagggggaatgcatcaacagcaacattcaccaggatcgcaatctaatgtaagctcaataat
31  G M H V P V T G P G G G M H Q Q Q H S P G S Q S N V S S N N

540 aataataacagtttctaatggacaagctatgaccgctgctcaaaaacaaaagaattggaccgtgtaaacggcctatgaacgctttcatg
61  N N N S S N G Q A M T A A Q K Q K E L D R V K R P M N A F M

630 gtgtggtaagaggtcagagggcggaaaatggctcaagaaaacccgaaaatgcacaattctgaaatcagtaaaagacttgggtgcagaatgg
91  V W S R G Q R R K M A Q E N P K M H N S E I S K R L G A E W

720 aaattgctatcagaagcagagaaaaagacctttcattgatgaggcaaaaagactgcgtgccattcacatgaaggagcaccagattataaa
121 K L L S E A E K R P F I D E A K R L R A I H M K E H P D Y K

810 tatagaccgagaaggaagcagaaaacccctcatgaagaaggacaaaatccctccctggaatgccacggcgcccatccaaggaggaggt
151 Y R P R R K T K T L M K K D K Y L P G M P T G A P I Q G R D

900 cccagtggtatgtaccccatgaacgggtatgtcccaacggataccggatgatgcccgatcctaaccgctaccagcacatggccaatcaa
181 P S G M Y P M N G Y M P N G Y P M M P D P N A Y Q H M A N Q

990 atgtacggctacaacattcctacacaacctatggccaccggacagatgaccaccggttcctacatgaacggaagctcaagttacacaatg
211 M Y G Y N I P T Q P M A T G Q M T T G S Y M N G S S S Y T M

1080 tccatgggaatgaatccgtattcgatgccaccgacgcaagttccaggagttaaacgggagcccaatcagcccagttccggggcccgcaa
241 S M G M N P Y S M P P T Q V P G V K R E P N Q P S S A G P Q

1170 gggggcaccggggccggccggttaaatccggagatttacgagaaatgatcagtatgtattaccggagacgggtccgatcctcatagt
271 G G Q P G R P V N P G D L R E M I S M Y L P G D G S D P H S

1260 gctcagagactcggcatgcagcagggccaggcgaagccattatcaaacacctcgtcggaaacagcgggtgaacaatacagtgcccttg
301 A Q R L A M Q Q A Q A Q A H Y Q T P S S E T A V N N T V P L

1350 actcatatgtaatttattgacaaaattatcattatctattgtttttgtactatacaacaatcatatgtgacccatgaagattttttac
298 T H M * 329

1440 aatctgctgattggaatcagcaaaattcggaaatcacataccacaccgacgcgtcatgaaaacattttttgtttttacttttccgtgctt
1530 gtatccgacagtcagttctgaaattttcgtcattgttgggtggaatattgttatagctgaattgctgaggggactcagactgaaacca
1620 ttatgcaagggcaataactccaggcatatcagcaaaacctacgagtgaaacggcggcgtgtgcatgctccagagagattctagagg
1710 gtaaaatctactctcaaaagtacgtgttgacaaatttatcaattttttgactatcatatcgaaaagaagtgccagggatcacatgttata
1800 cgttatcagttctctataagaaacgagagaattttttccatctttattttttttatttataatttcacctctattaagcataataatt
1890 ctaaatgttataaaaaaa1908
    
```

图 2 Sox2 基因全长 cDNA 序列及编码氨基酸序列

Fig.2 The complete cDNA sequence and deduced amino acids of Sox2 gene from *Pinctada fucata*

软件预测该编码蛋白含有一段保守的 HLH(helix-loop-helix domain)结构域, 及 Myc-N 保守区域, 而 myc 蛋白家族是典型的 HLH 亮氨酸拉链型转录因子, 可以通过与 DNA 特异位点结合而对转录过程进行调控, 从而调控细胞的增殖和分化。据此分析, 该编码蛋白预测正确并参与转录调控<sup>[14-15]</sup>。

Sox2 蛋白未预测到信号肽, Pfam 预测含有 HMG-box(high mobility group box), 它是 Sox 蛋白的特征保守结构域, 这种高迁移率基团在需要染色质构象发生改变的 DNA 相关调控过程中发挥作用, 如复制转录和 DNA 修复等<sup>[10, 13]</sup>。编码蛋白上还含有一段与 DNA 转录调控有关的 Soxp(Sox transcription factor)结构域<sup>[16]</sup>。推测该编码蛋白参与 DNA 转录调控。

KLF4 蛋白未预测到信号肽, Pfam 预测含有两个 zf-H2C2\_2(Zinc-finger double domain)结构域, 这是 KLF 蛋白的特征结构域, 表明编码蛋白与基因的表达调控相关<sup>[17-18]</sup>。

### 2.3 c-Myc、Sox2 及 KLF4 蛋白的同源性分析

利用 BLAST 软件分析 *c-Myc*、*Sox2* 及 *KLF4* 基因编码蛋白的同源性, 并使用 MEGA4 软件绘制系统进化树(图 4)。

同源性分析结果显示: (1)*c-Myc* 基因编码蛋白与太平洋牡蛎(*Crassostrea gigas*)的 Myc 蛋白同源性最高, 约为 63%。同时, 编码蛋白与佛罗里达文昌鱼(*Branchiostoma floridae*)、青岛文昌鱼(*Branchiostoma*

```

1  taagtgttaacttctgtttattataattcactctgaaataatctgatcatagataatacctataatctaagtggaacacacgagagaa
91  ttgaagtgtttacacaattt aatcattttagatcacggcctggaagcctaatacaatgttagacactggaccatatactcagcagcgc
181  ggggatccctcactcaccggacagcatcgatagcaaaagttagctggcaggacatagaaagttggccatgatgtaccaagacaatactc
271  cttatgagaacggaatgttacaaagccctggatcagctacgtggacctaccatggtatcaacgtgcccacaaatcctaaccatc acaacaac
361  aagacgctgggtccgacctcaatcatcaacagttacagagctgtgggtgcgctgcaagcagagacacgtacgacaactctctc
451  actttgatttcactctcctcaataactatggagaatcaaaagtctgtatccagaggataaacggacataaaatcaaacagaacaatactgt
541  caccagaatcgtaccagacttccatccacatttctagacattcctgatattagtttggacaattctgtgaataatgtccaccggaga
631  tggaaatctatgaaaactgatttgaacaagtatgtaaacaggaatgcaaccttgccgttcaagttcctgtgtatccaactctttagtt

721  cattcatgggcaataaatcgccgatcatgacaatgccccgtaattgcaatatcagtcgaagttccaaccagccatccgagcagatgccg
1
1
811  tatoccatgaacattcccggcaatatttccccaccgtcctgcgcggagatgcagggtgagctcccccaagtaccgctccccctccagatg
3  Y P M N I P G N I S P P S S P E M Q G E L P K Y R A P L Q M

901  ttacaaccgagcagctcacacccccactaccaccgcttcagcaacatatacatcacagtcccccgcaacaccagcagtatctcatgccg
33  L Q P Q Q S H P H Y H R L Q Q H I H H S P P Q H Q Q Y L M P

991  catccattcttgcacaaaatgcagcctcaacatagctccaagtcaccacatcagatgataaacggccgcttctgcctcagcaactcgt
63  H P F L P K M Q P Q H M L Q V P H Q M I T P P S S P Q Q L V

1081  gacttgcttctaccacagcaatctgtcgtatcctcctgacacccccaaaaagaggggagggcgaacgtggggccgtgaagcagacaga
93  D L L L P Q Q S V D A S S V Q P K K R G R R T W G R K R Q T

1171  agccacactttagcagcatccggcctgccagaaaacctacacaaaagctcccacctcaagccatctacgcacacacacagggcgaaaa
123  S H T C T H P G C Q K T Y T K S S H L K A H L R T H T G E K

1261  ccgtacgctgcacatggaaggatgaggatggaatttgcgcctctgatgagctcaccaggcactatcgtaagcagacagagatcgt
153  P Y A C T W K G C G W K F A R S D E L T R H Y R K H T G D R

1351  ccattccaatgtcactgtgtgagcgccttctccaggagcgaccatctgtcattgcacatgaagagacacatctgaactgtccgccag
183  P F Q C H L C E R A F S R S D H L S L H M K R H I * 207

1441  gtgtcgcaccatgtctctcactcgtgagctcttatgagctcaaacactcaggagactatgttgccttatatttgactgccttcaaaaatg
1531  tttattgaatgaacatttaagcaacatttagatgtaataattttactgtaaaattgttggctaccgccagtcacaaagcaaggttcagt
1621  ctgccttggcctggagattttttggagatttggcgaagctggctggctcgttccgctgtattccggaaatttgtgatgagtt
1711  atggccataatacttgtttagccgctaccatggcaatttgaagagaaggagggttattacgctatttattatttattgttcacatcga
1801  tataaatattatgatattataatatacttgaacttgttgggtatatacacaactgtttttatgtatttgaccagtcgttcaatagttat
1891  tgccttgaataaattgtagctatgcaactgcaaaaatcgaaagtacaaat ttttagatcacaaatagagctcccaaaaatgatattta
1981  acatgagcaatttgatattttatctcactcgtacatgtctgaagtcaaatatgaaacatagaatgactggttacagatgatgaactg
2071  tacgctttacaaggtcaagtgaagctacattagaatagaatattaaatcaaatgaatttttatatttgaataaattattcttat
2161  cactgctttttcataatcgttgcgaacaattttatagaaagttatttataatcttagtattttacatagatgctactatgatttt
2251  tgttaattttttttcac2258
    
```

图3 *KLF4* 基因全长 cDNA 序列及编码氨基酸序列

Fig.3 The complete cDNA sequence and deduced amino acids of *KLF4* gene from *Pinctada fucata*

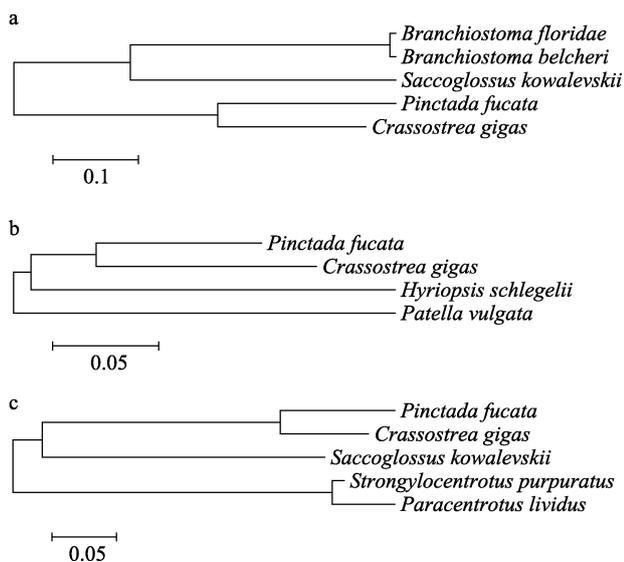


图4 *c-Myc*(a)、*Sox2*(b)和 *KLF4*(c)基因系统进化树

Fig.4 Phylogenetic tree of *c-Myc* (a), *Sox2* (b) and *KLF4* (c) gene

*belcheri*)以及半索动物长吻虫(*Saccoglossus kowalevskii*)的 Myc 蛋白也分别有约 45%、45%和 47%的相似性。(2)在分类学上,合浦珠母贝与太平洋牡蛎同为软体动物门(Mollusca)双壳纲(Bivalvia),文昌鱼与长吻虫同属脊索动物门(Chordata)。在 *c-Myc* 基因的保守性上,合浦珠母贝与太平洋牡蛎保同源性更高,而与文昌鱼和柱头虫的同源性较低。(3)*Sox2* 基因编码蛋白与太平洋牡蛎(*C. gigas*)、池蝶蚌(*Hyriopsis schlegelii*)的 *Sox2* 蛋白以及欧洲冒贝(*Patella vulgata*)的 *uSoxB* 蛋白分别有约 67%、68%和 72%的相似性。池蝶蚌也属双壳纲,冒贝属腹足纲(Gastropoda)。这 3 种生物与合浦珠母贝亲缘关系都较近,同源性比对结果也相近。(4)*KLF4* 基因编码蛋白与紫色球海胆(*Strongylocentrotus purpuratus*)的 *KLF4* 蛋白、长吻虫(*S. kowalevskii*)的类 *KLF2* 转录因子和拟正海胆(*Paracentrotus lividus*)的 *KLF2/4* 蛋白分别有

约 96%、96%和 95%的相似性, 另外与太平洋牡蛎 *C. gigas* 的 KLF1 蛋白同源性约为 67%。(5)拟正海胆属棘皮动物门(Echinodermata), 编码 KLF4 蛋白与海胆中同一亚型蛋白同源性更高, 96%的同源性表明 *KLF4* 基因在不同物种中具有高度的保守型。与牡蛎中 KLF1 蛋白的同源性较低, 可能是由于这两个蛋白亚型自身的差异所致。(6)同源性比对结果显示合浦珠母贝中的这 3 个基因与太平洋牡蛎同源性最高, 事实上牡蛎与珍珠贝同是双壳贝类, 亲缘关系本来就相近, 这里也得到证明。(7)而 3 个基因的同源性比对数据显示, KLF4 蛋白在不同物种中的同源性最高, 表明 *KLF4* 基因在不同物种中更为保守, *Sox2* 的保守性次之, *c-Myc* 基因相对最不保守。

#### 2.4 *c-Myc*、*Sox2* 及 *KLF4* 在合浦珠母贝各组织中分布

与预期相符, *c-Myc*、*Sox2* 和 *KLF4* 是 3 个非常保守的转录因子, 它们在合浦珠母贝的外套膜、足、生殖腺、内脏团、闭壳肌和鳃各组织中都有分布。

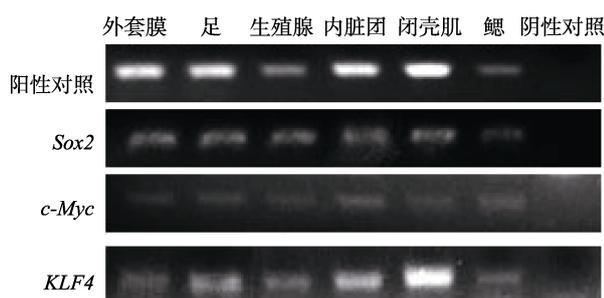


图 5 *Sox2*、*c-Myc* 和 *KLF4* 基因在合浦珠母贝各组织分布  
Fig.5 Distribution of *Sox2*, *c-Myc* and *KLF4* genes in tissues of *Pinctada fucata*

### 3 讨论与小结

自 2006 年 Yamanaka<sup>[1]</sup> 研究小组利用转录因子在体外将小鼠体细胞成功诱导成为多能干细胞之后, *Sox2*、*Oct4*、*c-Myc* 和 *KLF4* 作为多能因子被深入和广泛的研究。作者在合浦珠母贝外套膜转录组中找到 *c-Myc*、*Sox2* 及 *KLF4* 基因的片段, 并通过 RACE-PCR 技术克隆得到它们的全长序列, 不仅为以后研究合浦珠母贝中 ips 细胞奠定基础, 或许还可作为生物矿化的研究提供新的切入点。

生物信息学分析预测到 *c-Myc* 编码蛋白的一段 HLH 功能域, 以及 *Myc-N* 保守性片段, 这些都是 *c-Myc* 基因编码蛋白的典型特征, 可以通过与 DNA 特异位点结合而对转录过程进行调控, 从而起到调

控细胞增殖和分化的功能。此前已有较多文章报道过 *c-Myc* 蛋白在转录调控中所起的重要作用<sup>[14-15]</sup>。

而 *Sox2* 基因编码蛋白结构中预测到 HMG-box 和 *Soxp* 结构域, 正如前文所述, HMG-box 存在于高迁移率的基团中, 因为这种特性参与到需要改变染色质构象的一系列 DNA 相关调控过程中<sup>[13]</sup>。而 *Soxp* 结构域作为转录因子参与 DNA 转录调控<sup>[16]</sup>, 说明合浦珠母贝中 *Sox2* 编码蛋白也极有可能参与 DNA 转录调控过程中。

*KLF4* 编码蛋白中含有两个与基因表达调控有关的锌指双域。因此, 这 3 个转录因子在合浦珠母贝中同样在 DNA 水平起着调控基因表达的作用, 对它们的研究具有重要的意义。

在于其他已知基因的同源性比对中, 作者发现在与合浦珠母贝相近的物种尤其是牡蛎中都找到与这 3 种基因编码的氨基酸残基序列具有较高相似性的蛋白, 表明它们在遗传进化上具有更近的关系。而合浦珠母贝中预测的这 3 种蛋白都在 N 端序列与其他已知蛋白有较大差异, 因此根据这一点可以猜想它们功能上的差异, 可以有针对性的进一步研究。而诸如牡蛎等其他生物中相关蛋白的研究<sup>[19-21]</sup>也可为合浦珠母贝中这几个转录因子的研究提供重要参考价值。

#### 参考文献:

- [1] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors[J]. Cell, 2006, 126(4): 663-676.
- [2] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors[J]. Cell, 2007, 131 (5): 861-872.
- [3] 胡雨, 姚纪花. 斑马鱼多能性因子的研究进展[J]. 遗传, 2012, 34(9): 1097-1107.
- [4] Pan G J, Chang Z Y, Schöler H R, et al. Stem cell pluripotency and transcription factor Oct4[J]. Cell Res, 2002, 12(5-6): 321-329.
- [5] 张琦, 姚思国, 陈建泉, 等. 转录因子与诱导性多能干细胞[J]. 生命科学, 2009, 21(2): 299-302.
- [6] Adhikary S, Eilers M. Transcriptional regulation and transformation by Myc proteins[J]. Nat Rev Mol Cell Biol, 2005, 6(8): 635-45.
- [7] 陈艳玫, 姚鑫. 转录因子 Sox2 的研究进展[J]. 生命科学, 2004, 16(3): 129-134.

- [8] Cauffman G, Van de Velde H, Liebaers I, et al. Oct-4 mRNA and protein expression during human preimplantation development[J]. *Mol Hum Reprod*, 2005, 11(3): 173-181.
- [9] Fong H, Hohenstein K A, Donovan P J. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells[J]. *Stem Cells*, 2008, 26(8): 1931-1938.
- [10] Gong N P, Ma Z J, Li Q, et al. Characterization of calcium deposition and shell matrix protein secretion in primary mantle tissue culture from the marine pearl Oyster *Pinctada fucata*[J]. *Mar Biotechnol*, 2008, 10: 457-465.
- [11] Gong N P, Li Q, Huang J, et al. Culture of outer epithelial cells from mantle tissue to study shell matrix protein secretion for biomineralization[J]. *Cell Tissue Res*, 2008, 333: 493-501.
- [12] Mount A S, Wheeler A P, Paradkar R P, et al. Hemocyte-Mediated Shell mineralization in the Eastern Oyster[J]. *Science*, 2004, 304: 297-300.
- [13] Thomas J O. HMG1 and 2: architectural DNA-binding proteins[J]. *Biochem Soc Trans*, 2001, 29 (Pt 4): 395-401.
- [14] Prochownik, E V. c-Myc: Linking transformation and genomic instability[J]. *Current Molecular Medicine*, 2008, 8(6): 446-458.
- [15] Blackwell T K, Kretzner L, Blackwood E M, et al. Sequence-specific DNA-binding by the c-Myc protein. *Science*, 1990, 250(4987): 1149-1151.
- [16] Tanaka S, Kamachi Y, Tanouchi A, et al. Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells[J]. *Mol Cell Biol*, 2004, 24: 8834-8846.
- [17] Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling[J]. *Bioinformatics*, 2006, 22(2) : 195-201.
- [18] Kiefer F, Arnold K, Künzli M, et al. The SWISS-MODEL Repository and associated resources[J]. *Nucleic Acids Research*, 2009, 37(Database issue) : 387-392.
- [19] Zhang G F, Fang X D, Guo X M, et al. The oyster genome reveals stress adaptation and complexity of shell formation[J]. *Nature*, 2012, 490(7418): 49-54.
- [20] Putnam N H, Butts T, Ferrier D E K, et al. The amphioxus genome and the evolution of the chordate karyotype[J]. *Nature*, 2008, 453 (7198): 1064-1071.
- [21] Freeman R M, Wu M, Cordonnier-Pratt M M, et al. cDNA sequences for transcription factors and signaling proteins of the hemichordate *Saccoglossus kowalevskii*: efficacy of the expressed sequence tag (EST) approach for evolutionary and developmental studies of a new organism[J]. *Biol Bull*, 2008, 214 (3): 284-302.

## Cloning and sequence analysis of iPS related transcription factors from *Pinctada fucata*

ZHANG Ruo-yun, XU Guang-rui, PAN Cong, XIE Li-ping, WANG Hong-zhong, ZHANG Gui-you, ZHANG Rong-qing

(School of Life Sciences, Tsinghua University, Beijing 100084, China)

**Received:** Jan., 22, 2014

**Key words:** ips cells; transcription factors; cloning; *Pinctada fucata*

**Abstract:** The four transcription factors named *Sox2*, *Oct4*, *c-Myc* and *KLF4* can reprogram differentiated in vitro culture cells to an embryonic-like state, in another name, ips cells. The complete gene sequence of *c-Myc*, *Sox2* and *KLF4* from *Pinctada fucata* was amplified by RACE PCR. The complete sequence of *c-Myc* gene was 1585 bp and the complete ORF length was 1131 bp which encoded a protein with 376 amino acid residues. The analysis of protein structure showed c-Myc protein had the conservative HLH and Myc-N domains, which have the DNA-binding function. *Sox2* gene has the complete sequence of 1908 bp and the complete ORF length of 990 bp which encoded a protein with 329 amino acid residues. And the Sox2 protein had the conservative HMG-box and Soxp domains, which have the DNA-regulation function. *KLF4* gene was 2268 bp long and the complete ORF length was 624 bp which encoded a protein with 207 amino acid residues. And the KLF4 protein had two zf-H2C2\_2 domains, which have the DNA-binding function. Blast analysis indicated the three deduced amino acid sequences of the c-Myc, Sox2 and KLF4 genes from *P. fucata* showed high homology with proteins from *Crassostrea gigas*. RT-PCR result shows that these three genes were expressed in all six tissues from *P. fucata* including mantle, foot, gonad, visceral mass, adductor muscle and gill, suggesting that these genes were conservative transcriptional factors. Our study inspires the work on cell culture and research in *P. fucata*.

(本文编辑: 梁德海)