Basic Science

Identification of a Novel Testis-specific Gene in Mice and Its Potential Roles in Spermatogenesis

Aifa Tang, Zhendong Yu, Yaoting Gui, Hui Zhu, Yun Long, Zhiming Cai

Shenzhen Key Laboratory of Male Reproduction and Genetics, Shenzhen Hospital of Peking University, Shenzhen Medical Center for PKU-HKUST, Shenzhen, P.R. China

The first two authors contributed equally to this work.

> **Correspondence to:**

Zhiming Cai
Shenzhen Key Laboratory of Male Reproduction and Genetics
Peking University Shenzhen Hospital
Shenzhen Medical Center for PKU-HKUST
1120 Lianhua Rd.
518036 Shenzhen, the P.R.China
caizhiming2006@yahoo.com.cn

Aim Identification of a novel gene in mouse testis and its relation to spermatogenesis.

Methods Genes expressed during different developmental stages of the mouse testis were screened by DNA microarray. The results of chip analysis were authenticated by reverse transcription-polymerase chain reaction (RT-PCR) technique, as well as the tissue distribution of the selected genes. The characteristics of the selected genes were analyzed by bioinformatics tools.

Results A novel gene, *TSC77*, was identified and located at the mouse chromosome 2G1. The full cDNA length of *TSC77* was 2280 bp, with a 2046 bp open reading frame encoding a 681 amino acids protein with a predicted molecular weight of 77.17 kDa. The results of subcellular localization of GFP-TSC77 fusion protein indicated *TSC77* protein was located in the nucleus of Cos-7 cells. The analysis of multiple amino acid sequence alignment showed that *TSC77* protein was highly homologous with the human *CAI40813* (C2orf26 gene, 76%), and rat *XP_230651* (77%). Three putative domains including nicotine amide dinucleotide phosphate (NAD(P))-nitrite reductase (NirB) domain, uncharacterized NAD flavin adenine dinucleotide (FAD)-dependent dehydrogenases (HcaD) domain, and nicotinamide adenine dinucleotide (NADH) dehydrogenase (Ndh) domain were predicted at the protein site 168-309, 187-245, 181-216, respectively. Gene expression anlysis showed that the mouse *TSC77* is preferentially expressed in the mouse testis and its expression increased gradually from the day 9 to day 21.

Conclusion Based on its expression during mouse development, TSC77 may play an important role during mouse spermatogenesis.

Spermatogenesis is characterized by a mitotic (spermatogonia), a meiotic (spermatocytes), and a differentiated haploid (spermatids) phase. This complex process is orchestrated by the expression of thousands of genes, encoding proteins that play essential roles during specific phases of germ cell development. Investigating the mechanisms that regulate the mitotic and meiotic cell cycles in mammalian male germ cells can be useful for multiple purposes, such as: screening and characterizing key genes in sperm development; better understanding of the molecular requirements needed for spermatogenesis to occur; and development of new contraceptive targets and health care drugs (1).

Compared to the genes expressed in somatic cells, the genes in germ cells are expressed in a stage-regulated and tissue-specific manner. These stage-regulated and germ cell-specific genes include *CERM* (2), *Hlt* (3), *tesmin* (4), *Cdc2* (5), *Dmc1* (6), *Cyct* (7), *Gsk3-beta* (8), and *Ldhc* (9). Various approaches have been developed to obtain tissue-specific expressed genes including suppression subtractive hybridization (SSH), differential display reverse transcription-polymerase chain reaction (RT-PCR), and DNA microarray. Of these, DNA microarray (10) is a useful, high throughput screening method, which provides a platform to evaluate thousands of genes in parallel, allowing monitoring of changes in gene expression occurring during developmental events. Considerable progress has been made in screening germ cell-specific or testis-specific genes in mouse and human by using of DNA miroarray. Sha et al (11) compared gene expression profiles of adult and fetal human testes by a self-made cDNA chip that contained 9216 transcripts. In the chip analysis, 731 different expressing genes have been characterized. Out of 731 characterized genes, 54 were known genes, 18% of which (including *ODF2*, *CLGN*, *AKAP4*, *PGK2*, *SLP1*, *HTTA*, *LDH4*, and *LDHc*) were exclusively expressed in germ cells. In their subsequent works, Sha et al isolated and characterized a set of novel testis-specific genes, possibly spermatogenesis related including *cul-3b* (12), *NYD-SP5* (13), *BGR-like* (14), *NORPEG* (15), *DPP8* (16), *Ap2beta-NY* (17). Xin et al (18) compared gene expression profiles of normal young human testes and aged human testes by DNA miroarray, and the results of chip analysis showed that 117 different expressed genes were identified from these two samples. Of these 117 genes, *cox7a2* and *atp50* gene may be related to the decline of sexual potency in aged men. Schultz et al (1) analyzed gene expression in the mouse testis from day 1, 4, 8, 11, 14, 18, 21, 26, 29 to day 60 by the Affymetrix Mouse U74v2 chip, containing ~12 500 known mouse genes or Expressed Sequence of Tags (EST), spanning approximately 1/3 of the mouse genome. In their study, 1652 transcripts were identified, and the expression increased with or after meiosis. They estimated that >2300 genes (~4% of the mouse genome) were male germ cell-specific transcripts, >99% of which are first expressed during or after meiosis.

The aim of this study was to assess the gene expression profile of 5 time points of mouse testis development, expressed in postnatal days, using DNA microarray. The total RNA was isolated from the testes of mice 4, 9, 18, 35, 54 days and 6 months old were used for the analysis.

Materials and methods

Cell culture

Cos-7 cell line and mouse Sertoli cell lines CRL-2186 were obtained from the ATCC (American Type Culture Collection, Manassas, VA, USA) and maintained in a humidified atmosphere of 5% CO_2, 37°C and in Dullbeco's Minimal Essential Medium supplemented with 10% fetal calf serum (Life Technologies, Inc., Rockville, MD, USA), 3 mM L-glutamine, 100 µg/mL streptomycin, and 100 units/mL penicillin.

Animal care

Male and female Balb/c mice (aged 4-6 weeks) were obtained from the Laboratory Animals Center of South Medical University, Guangzhou, China and maintained in a temperature- and humidity- controlled room. All animals had free access to standard mouse food and water. Male and female mice mated naturally. The date of birth was designated as day 1. Testes were individually collected from the mice 4, 5, 9, 14, 18, 21, 35, 54, 60 days, and 6 months old (19). Other organs such as the brain, heart, lung, liver, kidney, spleen, epididymice from adult mice (6 months); ovary and uterus from adult (6 months) and newborn mice (4 days) were also collected. Testes from the day 4, 9, 18, 35, 54 of the postnatal development and 6 months old mice were collected for affymetrix chip analysis. Testes of 5, 9, 14, 18, 21, 35, 60, days and 6 months old mice, CRL-2059 cell line, and other organs (brain, heart, lung, liver, kidney, spleen, epididymis, ovary, and uterus) were collected for RT-PCR assay. Brain, heart, lung, liver, kidney, spleen, epididymis, ovary, and uterus were used to assess the assay of the distribution of *TSC77* gene. All samples were immediately frozen in RNAlater liquid (QIAGEN, Valencia, CA, USA). Animal experiments were approved by the the Animal Test Center of China.

RNA extraction and Affymetrix Genechip analysis

Total RNA was extracted from Sertoli cell lines CRL-2186 and 19 organs by using the Trizol reagent (Invitrogen, San Diego, CA, USA), according to the manufacturer's recommendations. The concentrations and the integrities of total RNA were assessed by measuring the 260:280 nm ratio and by fractionation in 1% denaturing agarose gel (formaldehyde), and the purity of total RNA was increased by using RNeasy Mini Kit (QIA-GEN, Valencia, CA, USA). Total RNA from the testes of 4, 9, 18, 35, 54 days, and 6 months old mice Balb/C mice with a 260:280 nm ratio

of 1.8 or higher was used to generate biotinylated cRNA target for the Genechip Mouse Genome 430 2.0 Array (Affymetrix, Santa Clara, CA, USA), which contained 45000 pairs of probe. The 45000 pairs of probes contained 39000 transcripts and 34000 well-characterized mouse genes. All of these procedures were carried out as described by the manufacturer. After hybridization, the array was washed, stained with streptavidin phycoerythrin using the Affymetrix GeneChip Fluidics Workstation 400, and scanned on a Hewlett-Packard gene array scanner (Hewlett-Packard Co., Palo Alto, CA, USA). After the arrays were scanned, the generated signals were determined and analyzed by MAS 5.0 software (Affymetrix, Santa Clara, CA, USA). The absolute and comparison analyses were also performed by MAS 5.0. After normalization of the data, comparison analysis compared an experimental array with a baseline array, so as to monitor changes in the expression of transcripts across the samples targeted to different arrays (1).

Semiquantitative RT-PCR

Semiquantitative RT-PCR was carried out to analyze and confirm the expression of candidate genes. RT-PCR was used to generate cDNA from various developmental stages of the testis or cells and other organs. Total RNA (1 μg) was reverse-transcribed into cDNA in a reaction primed by oligo deoxynucleotide T (dT)12–15 primer, using Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Reverse and forward oligonucleotide primers, specific to the chosen candidate genes, were designed using Primer Express 2.0 software (Applied Biosystems, Foster City, CA, USA) as described by the manufacturer. The primer sequences were as follows:

TSC77, forward primer:
TgTgTACTCgCTTTgATgACAAACTgA,
    Reverse primer:
gCTATCTCCTCCATggTAggCTTgT;

β-actin, forward primer:
AgAgggAAATCgTgCgTgAC,

    Reverse primer:
CCAAgAAggAAggCTggAAAA.

    All PCR assays were performed in triplicate.

Subcellular localization of GFP-TSC77 fusion protein

The coding sequence of mouse *TSC77* was amplified following the primers: gTgCTCgAgTA-ATgTCTgCATCTCCAgA (introducing a Xho I site in 5′), AgggAATTCgggCgCTggC-TACAgCTAA (introducing a EcoR I site in 5′), and the PCR products were double-digested with Xho I/EcoR I. The digested products were inserted into the Xho I/EcoR I site of pEG-FP-C1 to produce the fusion protein expressed vector, pEGFP-C1-TSC77, and the coding sequence of EGFP-TSC77 expressing cassette was confirmed by sequencing. The expressing plasmid, pEGFP-C1-TSC77, was transferred to Cos-7 cells by Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions. 24 hours after the transfer, the transferred cells and control pEGFP-C1-transferred cells were stained by DAPI (Sigma), according to the manufacturer's instructions, then the treated COS-7 cells were detected under the fluorescent microscopy LEICA DM4000B (Zeiss, Weyhe, Niedersachsen, Germany). PCR conditions were as follows: 94°C for 3 minutes; 30 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 40 seconds; followed by 72°C for 10 minutes.

Bioinformatics analysis

The complete coding sequence of the novel gene, *TSC77,* was derived by automated computational analysis using gene prediction method, Gnomon (*http://www.ncbi.nlm.nih.gov/genome/guide/gnomon.html*). ClustalW program (*http://www.ebi.ac.uk/clustalw/*) was used for multiple DNA sequence alignment. The amino acid sequence was analyzed by various bioinformatics tools. The hydrophobicity of the ami-

no acid sequences was analyzed by the protScale (*http://cn.expasy.org/tools/protscale.html*). SignalP (*http://www.cbs.dtu.dk/services/SignalP/*) was used for signal peptide prediction and Tmpred program (*http://ch.embnet.org/software/TMPRED_form.html*) for transmembrane domain prediction. Motives of the putative protein were predicted by the Prosite Program (*http://cn.expasy.org/prosite/*). Molecular weight and isoelectric point were also obtained from the Compute pI/Mw Program (*http://cn.expasy.org/tools/pi_tool.html*). PSORT II (*psort.nibb.ac.jp/form2.html*) was used to predict protein sorting signals and intracellular localization. Finally, we used the Conserved Domain Database at the NCBI (*http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/BLAST/cdd_search.html*) to search for any conserved important domains in the putative protein.

Results

Chip analysis

The results of the Affymetrix Genechip analysis indicated that a transcript (accession number BC004690) and its predicted complete coding sequence XM_619817, named *TSC77*, was preferentially expressed in spermatocytes. The hybridization signal intensities in the tests from Balb/C mice aged 4, 9, 18, 35, 54 days, or 6 months showed that the expression increased after 9 days of postanatal development, reaching its peak at 54 days (Figure 1A).

mRNA expression of TSC77

The RT-PCR analysis also indicated that the *TSC77* was differentially expressed in the testis. *TSC77* mRNA could not be detected at the day 5 of the postnatal development and the expression of *TSC77* increased gradually from postnatal day 9 to day 21, remaining stable from the day 21 to 6 months (Figure 1B). These results were consistent with dNA microarray analysis.
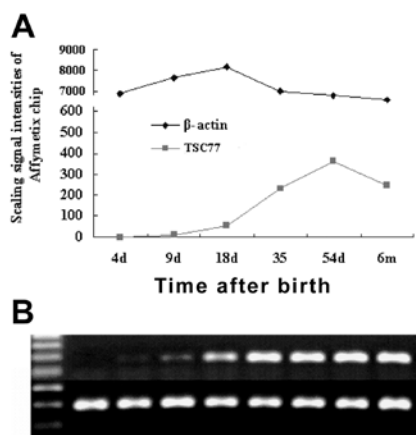
Figure 1. Time of *TSC77* expression in the testis after birth, detected by Affymetrix chip and RT-PCR. (**A**) Six time points of postnatal testis development, (d – day, m – month) were examined by Affymetrix chip analysis. (**B**) RT-PCR for *TSC77* in postnatal mouse testis. *β-Actin* was used as the internal control. M lane stands for the marker used in the detection of the products.
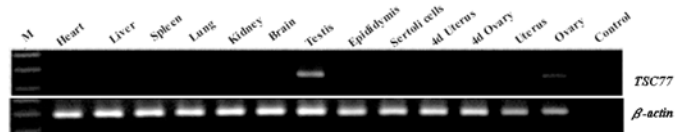

Figure 2. RT-PCR analysis of mouse *TSC*77 expression in two newborn organs (4-day uterus and 4-day ovary) and 11 adult organs. M lane stands for the detection of the RT-PCR products. *β-actin* was used as the internal control for concentration of the individual cDNA population. Control is PCR reaction mixture without template.

Among 13 different tissues examined, *TSC77* was preferentially expressed in the testis, and weakly in the adult ovary, but not in other tissues (Figure 2).

Features of cDNA and deduced protein

The full cDNA length of *TSC77* was 2280 bp, with a 2046 bp open reading frame from nt 1-2046, which encodes a 681 amino acids protein with a predicted molecular weight of 77.17 kDa, isoelectric point of 6.550, and –4.924 charge at pH 7.0 (Figure 3). A Blast search in the mouse genome database localized the *TSC77* gene to mouse chromosome 2G1. The *TSC77* gene is spliced to 11 exons and 10 introns, encompassing a 169 710 bp genomic DNA (from 38 961 384 to 39 131 093 bp) in NT_005403.16. Blast-pr showed that the predictive protein was

atgtctgcatctccagaggccacctgcactatacttacagatggaaagcatgcggtaaga   60
M S A S P E A T C T I L T D G K H A V R   20
gagcaagttgtgcagatacgaactcgtgcctccgatcagtttaaggctttgcctcctccgcag   120
E Q V V Q I R T R A S D Q F K A L P P Q   40
aaccccgggcttcatgctttcaggtttccattctgcataagcgctacgtcatcttccatc   180
N P G L H A F R F P F C I S A T S S S I   60
ctctctaggctgtttgcagttaggtgtgagcattatatgatctgtagcataactgtaaca   240
L S R L F A V R C E H Y M I C S I T V T   80
ggtacaggagcctctttcttcaggtgagtttcgccttgaaccacaccaacagaaaacta   300
G T G A F F L Q V S F A L N H T N R K L   100
acattagaaccgaaagtcactgtcaatgccagaattgttgtggtcggtgcgtccagtgtt   360
T L E P K V T V N A R I V V G A S S V   120
gggagtttcattcctagagacactggtgttttgctctcacctgaaatttaataatctcacc   420
G I S F L E T L V F C S H L K F N N L T   140
ctgatttcaactcatggactcccaggaaaaagacttttgcacaacgaacagaggaagttt   480
L I S T H G L P G K R L L H N E Q R K F   160
ctagcaagcgaccactgtttttaatgataaggattatgcactgatgtcactgtgctcctgg   540
L A S D H C F N D K D Y A L M S L C S W   180
gtgaacgttgtggttggtcgaatgacggccatagaccgagcagccaagcaggttgtggtt   600
V N V V V G R M T A I D R A A K Q V V V   200
tccaagagtgagatcgtgttctacgaccacctcattctctgtaccgggctgcagtaccag   660
S K S E I V F Y D H L I L C T G L Q Y Q   220
gttccgtgccccacaggggcgtgatactgagcagcacctgacgaacagagaggttttggag   720
V P C P T G A D T E Q H L T N R E V L E   240
ctcagcaagcagcggtacaccggcactgttccttccaacctgttcattctcaacgacgag   780
L S K Q R Y T G T V P S N L F I L N D E   260
gaagactgcgttaggatgctctgttggctgagaaacagctccatcctcacagaagggaag   840
E D C V R M L C W L R N S S I L T E G K   280
gtcattgtctacgggaatacactcgacgcttacacaacagtggagacactcctcaacatt   900
V I V Y G N T L D A Y T T V E T L L N I   300
ggcatgaaggggagcggcatctactttgtgcatcccccaactgaatcaaacataacctgt   960
G M K G S G I Y F V H P P T E S N I T C   320
atcaacaacaatgatgtagagagtgccgtggaagatgcactcagtgccgctggggtgaca   1020
I N N N D V E S A V E D A L S A A G V T   340
gtctttcataagcgccttgttggctcagtggaacatggccagcacccggatcctatctac   1080
V F H D A L L A Q W N H G Q H P D P I Y   360
aatgctgcttcaccactccaccaagcctatcagactggagtgttcagcgttcttcagc   1140
N A C F T T S T K P I R L E C S A F F S   380
ttctacagaagaatgtggattacgaaacatttaaagcatttaatgatgcgccttgtc   1200
F Y K K N V D Y E T F K A F N D A C L V   400
tatgatggccgacttgtgatcgatactacctccacaccaatgacatagccatcagagct   1260
Y D G R L V I D T T F H T N D I A I R A   420
gctggttccctcaccaaattctccaaccgatattactcaaatgagtggactcacagcaat   1320
A G S L T K F S N R Y Y S N E W T H S N   440
ttcagttccaaagaaattggcttccagctggctgcagctatgctcagtctctttgaccca   1380
F S S K E I G F Q L A A A M L S L F D P   460
acccttgagcctgtgactgaaccaccagctgaccttgaccgcctcatccccatgtacaag   1440
T L E P V T E P P A D L D R L I P M Y K   480
ggagccaagatccaaggagggattcttcctgggtcttatcattatctgcatatttccaaa   1500
G A K I Q G G I L P G S Y H Y L H I S K   500
cctgccatcccaactcctttgaatctacaaaggtcacagcctgattttggttcagacgta   1560
P A I P T P L N L Q R S Q P D F G S D V   520
gtgacaggtaatgcgaaaaaggggacgtacttccgactatacattaacaactataagttg   1620
V T G N A K K G T Y F R L Y I N N Y K L   540
gtggaggccatcacgtgctttcaaaggagcctttcccatcctccaactacatccgcttg   1680
V E A I T C F S K E P F P S S N Y I R L   560
tttggacaacatgagcaagtcctcaacaact**gtgtactcgctttgatgacaaactg**atc   1740
F G Q H E Q V L N N L C T R F D D K L I   580
ccagatcttttacagctacttcacagagccgtggtgcatggccctcttccacgatcgtttt   1800
P D L Y S Y F T E P W C M A L F H D R F   600
attgatctcaagacaagagttgcgccaaatcttaatctccaagcaggaggaag**acaagcc**   1860
I D L K K E L R Q I L I S K Q E E D K P   620
accatggaggagatagcctaccggttagaggaggaagaaataaacctgaatgagaagccc   1920
T M E E I A Y R L E E E E I N L N E K P   640
cagaaatacctccaaagggtttttgaggattccatctacaaaagcctgctggagaagagc   1980
Q K Y L Q R V F E D S I Y K S L L E K S   660
atcctagattaccttcattacaaccagtaccacctgcccatgtacgcatggccaggcatc   2040
I L D Y L H Y N Q Y H L P M Y A W P G I   680
attt**agc**tgctagccagcgccccacccccccgctggcatcacttccactctctgtcagccaa   2100
I *   682
gtgctgacagaacaccccaacccccatccccactattatacccttttttcctcccgtgt   2160
tttccgtgtattccagtaggtggcgctaggctgtgtggccctgctaagccagggaacttt   2220
acaaacaggcatgttcagacagaataagccactctagcgataaaaagagctcctacaaa   2280

Figure 3. Nucleotide sequence of mouse *TSC77* cDNA and the deduced amino acid sequence. Underlining sequences show the RT-PCR primers for the determination of expression profile of mouse *TSC77* gene in mRNA level. Start codon and stop codon are marked in bold. Nucleotide and amino acid identification numbers are shown after the sequences.
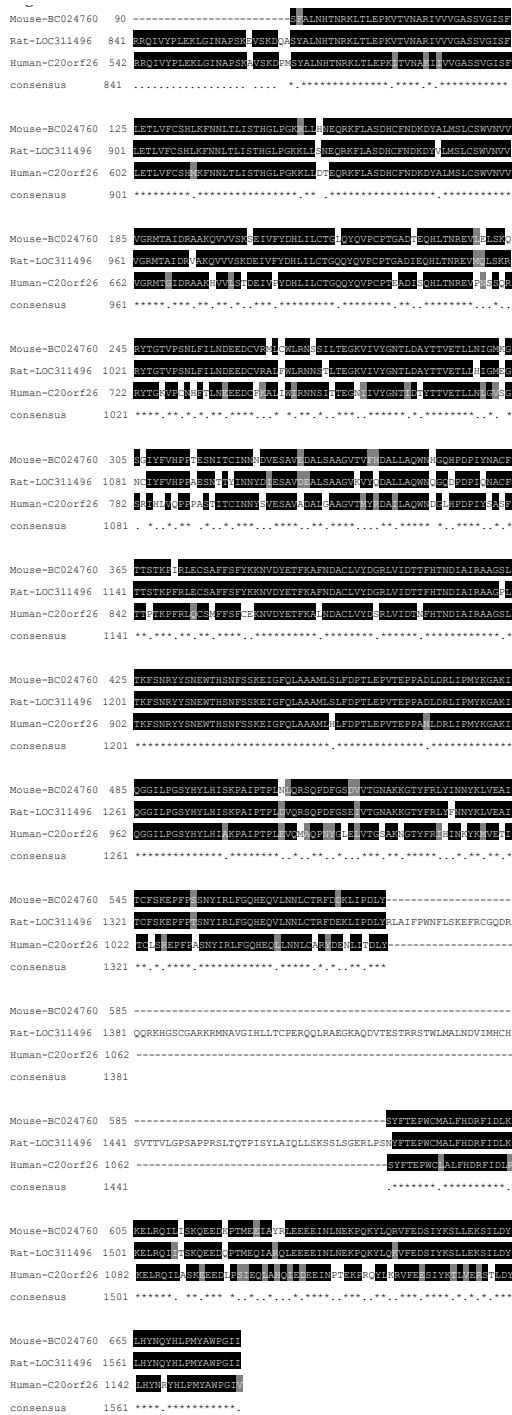
```
Mouse-BC024760    90  ------------------------SYALNHTNRKLTLEPKVTVNARIVVVGASSVGISF
Rat-LOC311496    841  RRQIVPLEKLGINAPSKRVSKDYASYALNHTNRKLTLEPKVTVNARIVVVGASSVGISF
Human-C20orf26   542  RRQIVPLEKLGINAPSKRVSKPYSYALNHTNRKLTLEPKTVNAMTVVVGASSVGISF
consensus        841  .................  .  *.*************..****.*.***********

Mouse-BC024760   125  LETLVFCSHLKFNNLTLISTHGLPGKRLLPNEQRKFLASDHCFNDKDYALMSLCSWVNVV
Rat-LOC311496    901  LETLVFCSHLKFNNLTLISTHGLPGKKLLSNEQRKFLASDHCFNDKDYWLMLSCWVNVV
Human-C20orf26   602  LETLVFCSHLKFNNLTLISTHGLPGKKLLDTLEQRKFLASDHCFNDKDYALMSLCSWVNVV
consensus        901  **********.*************..**  *.************.****...********

Mouse-BC024760   185  VGRMTAIDRAAKQVVVSKPEIVFYDHLILCTGQYQVPCPTGAPEQHLTNREVPELSKQ
Rat-LOC311496    961  VGRMTAIDRVAKQVVVSKDEIVFYDHLILCTGQYQVPCPTGADIEQHLTNREVMCLSKR
Human-C20orf26   662  VGRMTFIDRAARPVVVSPDEIVFYDHLILCTGQYQVPCPCSADISHLTNREVPLSSCR
consensus        961  *****.***.**.**.*.***.**********.*********..*..***.....*..

Mouse-BC024760   245  RYTGTVPSNLFILNDEEDCVRPLWLFNSILTEGKVIVYGNTLDAYTTVETLLNIGMPG
Rat-LOC311496   1021  RYTGTVPSNLFILNDEEDCVRALPWLRNNSPLTEGKVIVYGNTLDAYTTVETLLHGMPG
Human-C20orf26   722  RYTGAVFSNLFLANEEDGFIALSWRNNSPLTEGMIVYGNTLPYTTVETLLNIGYSS
consensus       1021  ****.**.*.*.**.****....*  *.**...***.********.*********...*  *

Mouse-BC024760   305  SHIYFVHPFPSNITCINNDVESAVRDALSAAGVTVPNDALLAQWNHEQHPDPIYNACF
Rat-LOC311496   1081  NCIYFVHPFPABSNRHINNYDESAVRALSAAGVPVPCDALLAQWNOEGQPDPNACF
Human-C20orf26   782  SHHIACFPALSITCINNYSVESAPADAVCAAGVTMPDALAQWNFASHPDPIVSAS
consensus       1081  *..*.** .*..*.*** .****.**.**..**.......  * .****** *.****.*.

Mouse-BC024760   365  TTSTKPFRLECSAFFSFYKKNVDYETFKAFNDACLVYDGRLVIDTTFHTNDIAIRAAGSL
Rat-LOC311496   1141  TTSTKPFRLECSAFFSFYKKNVDYETFKAFNDACLVYDGRLVIDTTFHTNDIAIRAAGPL
Human-C20orf26   842  TSFTKPFRLCESAFFSFCENVDYETFKAPNDACLVYDSRLVIDEPHTNDIAIRAAGSL
consensus       1141  **.***.**.**.****.**********.*******.*****....*************.*

Mouse-BC024760   425  TKFSNRYYSNEWTHSNFSSKEIGFQLAAAMLSLFDPTLEPVTEPPADLDRLIPMYKGAKI
Rat-LOC311496   1201  TKFSNRYYSNEWTHSNFSSKEIGFQLAAAMLSLFDPTLEPVTEPPADLDRLIPMYKGAKI
Human-C20orf26   902  TKFSNRYYSNEWTHSNFSSKEIGFQLAAAMLFLFDPTLEPVTEPPAALDRLIPMYKGAKI
consensus       1201  ***********************************.**********.*.***********

Mouse-BC024760   485  QGGILPGSYHYLHISKPAIPTPINNQRSQPDFGSDVVTGNAKKGTYFRLYINNYKLVEAI
Rat-LOC311496   1261  QGGILPGSYHYLHISKPAIPTPLVQRSQPDFGSEVTGNAKKGTYFRLVPNNYKLVEAI
Human-C20orf26   962  QGGILPGSYHYLHILKPATPTPLVQYQPIGLVTGEAKGGTYFHITNSYRVVEAI
consensus       1261  **************.****.****..*....*.*..***.**.****...*..*.*.

Mouse-BC024760   545  TCFSKEPFPFSNYIRLFGQHEQVLNNLCTRFDFKLIPDLY-------------------
Rat-LOC311496   1321  TCFSKEPFPFSNYIRLFGQHEQVLNNLCTRFDEKLIPDLYRLAIFPWNFLSKEFRCGQDR
Human-C20orf26  1022  TCLSREPFPASNYIRLFGQHEQFLNNLQARFDENLFDLY-------------------
consensus       1321  **.*.****.********.*****.*.***.*.*.**.***

Mouse-BC024760   585  ------------------------------------------------------------
Rat-LOC311496   1381  QQRKHGSCGARKRMNAVGIHLLTCPERQQLRAEGKAQDVTESTRRSTWLMALNDVIMHCH
Human-C20orf26  1062  ------------------------------------------------------------
consensus       1381  

Mouse-BC024760   585  -----------------------------------------SYFTEPWCMALFHDRFIDLK
Rat-LOC311496   1441  SVTTVLGPSAPPRSLTQTPISYLAIQLLSKSSLSGERLPSNYFTEPWCMALFHDRFIDLK
Human-C20orf26  1062  -----------------------------------------SYFTEPWCMALFHDRFIDLK
consensus       1441                                           .*******.***********.

Mouse-BC024760   605  KELRQIIHSKQEELCPTMEHIAYPLEEEEINLNEKPQKYLQRVFEDSIYKSLLEKSILDY
Rat-LOC311496   1501  KELRQIIFSKQEECPTMEQIAQLEEEEINLNEKPQKYLQCVFEDSIYKSLLEKSILDY
Human-C20orf26  1082  KELRQIIASKSEELCLESREGCAIQEEEDINPTEKSRCYHKKVFEDSIYKELEDFSPLDY
consensus       1501  ******.**.**  *..*.*.....*.****.***..**..****.***.*.*.*.**.

Mouse-BC024760   665  LHYNQYHLPMYAWPGII
Rat-LOC311496   1561  LHYNQYHLPMYAWPGII
Human-C20orf26  1142  LHYNYVHLPMYAWPGIV
consensus       1561  ****.************.
```

Figure 4. Alignment of the amino acid sequences of the *TSC77* gene (Mouse-BC024760) and its homologous genes in rat (Rat-LOC311496) and human (Human-C20orf26) species by using software ClustlW (only C-terminal of protein sequence were shown). Protein identification numbers are shown after the species abbreviations. Dark and gray shadings indicate conserved and similar residues, respectively.

highly homologous with the human CAI40813 sequence (C2orf26 gene, 76%) and rat sequence XP_230651 (LOC311496 gene, 77%) (Figure 4). The result of subcellular localization prediction of *TSC77* protein showed that *TSC77* is located in the nucleus (87.0%). Searching for conserved function domain by Blast-p, identified 3 putative domains, including NAD(P)-nitrite reductase (NirB) domain, uncharacterized NAD (FAD)-dependent dehydrogenases (HcaD) domain, and NADH dehydrogenase (Ndh) domain, which were predicted at the protein site 168-309, 187-245, and 181-216, respectively. With TMpred search server, three transmembrane helices were found at the 46-67, 72-93, 110-131 positions in the *TSC77* gene, with the N-terminal outside the membrane sequence. The amino acid of *TSC77* was analyzed by SignalP (V 3.0) and no signal peptide was predicted in *TSC77* protein, indicating that the *TSC77* gene was a non-secretory protein.

Subcellular localization of TSC77 protein

According to the expression of fusion protein EGFP-TSC77 in Cos-7 cells (Figure 5) the protein of *TSC77* was located in the nucleus, which is consistent with the deduced results of *TSC77* in the subcellular localization predicted by PSORT II software. The control protein GFP was distributed in the entire cells of Cos-7 cell lines.

Discussion

In this study, we identified a novel gene, *TSC77,* expressed preferentially in the neonatal mouse testis. We performed Affymetrix chip analysis at six time points of postnatal mouse testis development, from the day 4 to 6 months. The results of chip analysis showed that the expression of *TSC77* increased gradually with the development of mouse testis.

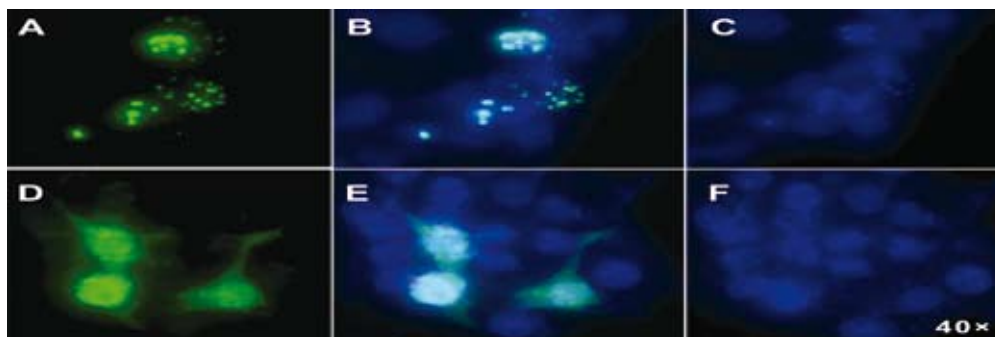RT-PCR confirmed the results of DNA microarray analysis showing that *TSC77* could

Figure 5. Localization of fusion protein, GFP-TSC77, in Cos-7 cells. (**A**) GFP-tagged TSC77 fusion protein in transfected cells under fluorescent light; (**B**) merged image of (A) and (C); (**C**) GFP-TSC77 protein-expressing cells under UV light; (**D**): GFP protein-expressing Cos-7 cells under fluorescent light; (**E**): merged image of (D) and (F); (**F**) GFP protein-expressing Cos-7 cells under UV light.

only be detected in the testis in mice older than 9 postnatal days. The expression increased gradually up to the day 21 and remained stable afterwards.

TSC77 was preferentially expressed in adult testis and weakly expressed in adult ovary, indicating that *TSC77* might be a germ cell-specific gene. In the development of a normal male mouse, the type A spermatogonia begins with mitosis taking place at the day 9 and meiosis at the day 14. According to the expression pattern with *TSC77* may play a role in the process of spermatogonia from mitosis to meiosis and may be an early marker of the meiosis during spermatogenesis in mouse.

The amino acid sequence analysis of the coding protein using the SignalP (V 3.0) predicts no signal peptide, indicating that the *TSC77* gene is a non-secretory protein, and the prediction of subcellular localization of *TSC77* protein showed that *TSC77* was located in the nucleus. The results of subcellular localization of GFP-TSC77 fusion protein in Cos-7 cells also supported the finding that *TSC77* protein was located in the nucleus. Multiple amino acid sequence alignment of human, mouse, rat, dog, and chicken homologous genes showed that *TSC77* was a highly conserved protein in mammalian species, indicating that *TSC77* may participate in the molecular events underlying mouse spermatogenesis. Three putative domains, NAD(P)-nitrite

reductase (NirB), uncharacterized NAD (FAD)-dependent dehydrogenases (HcaD), and NADH dehydrogenase (Ndh) were predicted at the protein site 168-309, 187-245, 181-216, respectively. These domains in *TSC77* protein were highly conserved among *TSC77* homologous genes in human, mouse, and rat, indicating its possible biological role during mammalian spermatogenesis.

A group of genes exclusively expressed in the testis, encoding proteins with essential roles during spermatogenesis, has been named "chauvinist" genes (20). These genes can be grouped into three categories: 1) homologous genes only in spermatogenic cells; 2) unique genes without significant similarity in nucleotide sequence to those expressed in any other cells; and 3) genes producing transcripts in both somatic and spermatogenic cells but germ cell-specific transcripts due to alternative transcription start site, transcript-splicing sites, or polyadenylation signals. A special feature of "chauvinist genes" is that their expression is developmentally regulated during meiotic and post-meiotic phases. Consistently, *TSC77* falls into the category of unique genes, expressed in developmentally regulated patterns.

The function of *TSC77* in mouse spermatogenesis remains uncertain. However, to our knowledge, this is the first time that *TSC77* was found specifically in these tissues. Further investigation of the molecular mechanisms in the dis-

tribution of *TSC77* in multiple tissues by in situ hybridization or immunohistochemical staining, its interaction with other proteins by immuno-precipitation or yeast two-hybrid system, and the down-regulation of *TSC77* by RNAi, or knock-out techniques are needed to provide adequate information on its biological functions in the mouse spermatogenesis.

Reference

1 Schultz N, Hamra FK, Garbers DL. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. Proc Natl Acad Sci U S A. 2003;100:12201-6. Medline:14526100

2 Blendy JA, Kaestner KH, Weinbauer GF, Nieschlag E, Schutz G. Severe impairment of spermatogenesis in mice lacking the CREM gene. Nature. 1996;380:162-5. Medline:8600391

3 vanWert JM, Wolfe SA, Grimes SR. Testis-specific expression of the rat histone Hlt gene in transgenic mice. Biochemistry. 1995;34:8733-43. Medline:7612613

4 Sugihara T, Wadhwa R, Kaul SC, Mitsui Y. A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation. Genomics. 1999;57:130-6. Medline:10191092

5 Yu BZ, Song YT, Yu DH, Su WH, Gasana V, Li YX, et al. Expression and immunohistochemical localization of Cdc2 and P70S6K in different stages of mouse germ cells. Cell Biochem Funct. 2006;24:113-7. Medline:16329156

6 Masson JY, Davies AA, Hajibagheri N, Van Dyck E, Benson FE, Stasiak AZ, et al. The meiosis-specific recombinase hDmc1 forms ring structures and interacts with hRad51. The meiosis-specific recombinase hDmc1 forms ring structures and interacts with hRad51. EMBO J. 1999;18:6552-60. Medline:10562567

7 Hake LE, Kuemmerle N, Hecht NB, Kozak CA. The genes encoding the somatic and testis-specific isotypes of the mouse cytochrome c genes map to paralogous regions of chromosomes 6 and 2. Genomics. 1994;20:503-5. Medline:8034327

8 Rentzsch F, Hobmayer B, Holstein TW. Glycogen synthase kinase 3 has a proapoptotic function in Hydra gametogenesis. Dev Biol. 2005;278:1-12. Medline:15649456

9 Li S, Zhou W, Doglio L, Goldberg E. Transgenic mice demonstrate a testis-specific promoter for lactate dehydrogenase, LDHC. J Biol Chem. 1998;273:31191-4. Medline:9813024

10 Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A. 1994;91:5022-6. Medline:8197176

11 Sha J, Zhou Z, Li J, Yin L, Yang H, Hu G, et al. Identification of testis development and spermatogenesis-related genes in human and mouse testes using cDNA arrays. Mol Hum Reprod. 2002;8:511-7. Medline:12029067

12 Lu L, Zhou ZM, Huang XY, Xu M, Yin LL, Wang H, et al. Identification and characterization of cul-3b, a novel hominine CUL-3 transcript variant. Asian J Androl. 2005;7:205-11. Medline:15897978

13 Yin LL, Li JM, Zhou ZM, Sha JH. Identification of a novel testis-specific gene and its potential roles in testis development/spermatogenesis. Asian J Androl. 2005;7:127-37. Medline:15897968

14 Zheng Y, Zhou ZM, Min X, Li JM, Sha JH. Identification and characterization of the BGR-like gene with a potential role in human testicular development/spermatogenesis. Zheng Y, Zhou ZM, Min X, Li JM, Sha JH. Asian J Androl. 2005;7:21-32. Medline:15685348

15 Yuan W, Zheng Y, Huo R, Lu L, Huang XY, Yin LL, et al. Expression of a novel alternative transcript of the novel retinal pigment epithelial cell gene NORPEG in human testes. Asian J Androl. 2005;7:277-88. Medline:16110356

16 Zhu H, Zhou ZM, Lu L, Xu M, Wang H, Li JM, et al. Expression of a novel dipeptidyl peptidase 8 (DPP8) transcript variant, DPP8-v3, in human testis. Asian J Androl. 2005;7:245-55. Medline:16110352

17 Zhang XD, Yin LL, Zheng Y, Lu L, Zhou ZM, Sha JH. Expression of a novel beta adaptin subunit mRNA splice variant in human testes. Asian J Androl. 2005;7:179-88. Medline:15897975

18 Xin Z, Liu W, Tian L, Yuan Y, Xin H, Fu J, et al. Studies on the testis regression related gene profile in aged male [In Chinese]. Beijing Da Xue Xue Bao. 2003;35:364-8. Medline:12947549

19 Guo R, Yu Z, Guan J, Ge Y, Ma J, Li S, et al. Stage-specific and tissue-specific expression characteristics of differentially expressed genes during mouse spermatogenesis. Mol Reprod Dev. 2004;67:264-72. Medline:14735487 20 Eddy EM.