RESEARCH NOTE

Molecular cloning and characterization of a novel human testis-specific gene by use of digital differential display

DONGSONG NIE[1,2*] and YANG XIANG[2]

[1]*Department of Chemistry and Chemical Engineering, Hunan Institute of Science and Technology, Hunan, YueYang 414000, China*
[2]*Human Reproductive and Stem Cell Engineering Institute, Central South University, 88 Xiangya Road, Changsha, Hunan 410078, China*

Introduction

Infertility is a common problem which affects 13% of couples; both female and male factors are known to contribute to it. Although several external factors have been claimed to affect male reproduction, evidence that genetic factors play important roles is accumulating (Bourrouillou *et al*. 1985; Adams *et al*. 1995). Idiopathic azoospermia or severe oligozoospermia accounts for about 20% of male infertility cases. In these cases, systematic analysis of molecular deletions of Y chromosome indicates that three nonoverlapping regions, *AZFa*, *AZFb* and *AZFc*, in band Yq11, are involved in male spermatogenesis (Vogt *et al*. 1996). A series of genes including *DAZ* and *RBM* genes have been identified in these AZF regions as candidate genes for azoospermia. They are categorized as a testis-specific type and a multiple-tissue-expression type (Ma *et al*. 1993; Reijo *et al*. 1995). However, gene-specific mutations causing the azoospermia phenotype have only been found in the *USP9Y* gene (Sun *et al*. 1999). Some autosomal genes are also associated with spermatogenesis in mouse and human, including *DAZLA*, *CREM* and *HSP70* with its human homologue *HSPA2* (Dix *et al*. 1996; Ruggiu *et al*. 1997)

The products of many genes are essential for spermatogenesis, but a small number affect spermatogenesis exclusively. Identification of these new genes and their role is of great importance in understanding the biology of spermatogenesis (Cooke *et al*. 1998). Our group has successfully cloned three new human testis apoptosis-related genes and two human testis-specific expressed genes by using suppression subtractive hybridization (SSH) (Liu *et al*. 2002, 2004; Xing *et al*. 2003). In an attempt to identify new candidate genes for spermatogenesis in testis, a new data mining tool called digital differential display was used to search for testis-specific human transcripts. Digital differential display (DDD) is an Internet-based resource for the identification of genes whose expression varies in different tissue types (http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.shtml). DDD is also a quantitative method that enables the user to determine the fold differences in expression between the libraries being compared. This resource offers exciting new avenues for exploration in the search for novel genes in health and disease; indeed it has recently been applied to the identification of transcripts associated with cancer (Robbins 1996; Schmit *et al*. 1999; Wheeler *et al*. 2000). In this study, nine testis cDNA libraries were compared against 70 other libraries, and the hits showing >10-fold differences were selected. The selected expressed sequence tags (ESTs) were assembled and validated by BLAST against dbEST and nonredundant (nr) databases. One of the ESTs was selected for further research. Reverse transcriptase polymerase chain reaction (RT-PCR) analysis revealed that it was exclusively expressed in adult testis. The results were further validated by Northern blot analysis of multiple adult tissues. Real-time RT-PCR analysis of testis of different developmental periods revealed that the EST was significantly expressed in adult testis, and moderately expressed in sperm. The full-length cDNA encompassing the entire open reading frame was cloned, and the gene was named *homo sapiens spermatogenesis-related gene 8* (*SRG8*).

Materials and methods

Data-mining of UniGene database

In this study, the DDD tool was used according to the database instructions. Nine testis-derived cDNA libraries

*For correspondence. Email: ndz1997@126.com.

Keywords. digital differential display; testis; real-time PCR; Northern blot.

(pool A) were compared against 70 other organ-specific cDNA libraries (pool B; heart, kidney, liver, spleen, muscle, etc.). The output provided a numerical value in each pool denoting the fraction of sequences within the pool that mapped to the UniGene cluster, providing a dot intensity. Fold differences were calculated by using the ratio pool A to pool B. Statistically significant hits (>10-fold differences) were selected. Of these ESTs, novel ESTs were assembled by CAP3 EST assembler to form new contigs, then the new contigs were input into a BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) search against the dbEST and nr databases.

Molecular cloning of full-length cDNA

The forward primer N1: 5′-GCTCTTTGGAGCACCCACTT CAC-3′ and the reverse primer N2: 5′-CATGCTGGGCATG TATGCCAAGC-3′ were designed according to the predicted sequence. N1 and N2 were used in PCR assay with Advantage 2 DNA polymerase (Clontech) and with Marathon-Ready^{TM} cDNA of human testis (Clontech) as template. PCR amplification cycles involved initial denaturation at 95°C for 1.5 min, and 35 cycles of 94°C for 40 sec, 58°C for 30 sec, and 72°C for 40 sec, then 72°C for 5 min, and holding at 4°C. The PCR fragment was cloned into pMD18-T vectors (TaKaRa) and sequenced.

Bioinformatics analysis

The Translate program of ExPASy (http://www.expasy.org/tools/dna.html) was used to identify open reading frame (ORF). Comparison with human genome draft sequence in the GenBank database was performed to locate the new gene on the human chromosome map. The ProtParam tool (http://www.expasy.org/tools/protparam.html) was utilized to identify physicochemical parameters of the new protein sequence. TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) was used to identify transmembrane regions and protein orientation. SignalP V1.1 (http://www.cbs.dtu.dk/services/SignalP-1.1/) was used to look for signal peptide cleavage sites. PSORT WWW Server (http://psort.ims.u-tokyo.ac.jp/) was utilized to predict protein subcellular localization. SMART (http://smart.umbl-heidelberg.de/) was used to identify motifs.

RT-PCR

Total RNA from human tissues (foetal and adult lung, heart, small intestine, skeletal muscle, spleen, liver, kidney, testis, epididymis and ovary) was isolated using RNA Isolation Kit (Promega). cDNAs were synthesized according to instructions in the kit and were used as template in a PCR reaction. The PCR primer pair was S1: CTCCTGAAGAAGGGGCGTCTAAA and S2: GGAAGAGCTTCAGGGTAGGGACA. Amplification cycles were the same as above. RT-PCR product was separated in 2.0% agarose gel. *GAPDH* was amplified as control.

Real-time RT-PCR

Total RNA was extracted from six separate collections of testis and sperm using the Trizol extraction kit and treated with DNAase I. Real-time RT-PCR was performed on the iCycle QPCR system using the Quantitect SYBR Green RT-PCR kit (BioRad). Gene-specific primers were used to determine the relative expression levels of *SRG8* according to the standard curve method. Primers were designed to span introns, resulting in a single band on gels (F1: CAGGACAACTCCTGTCTCTAC, R1: CTTGAAAGGCCTCCAGCCGCA). SYBRGreen was used to detect the double-stranded DNA produced during the amplification reaction and *GAPDH* content to normalize for the input of RNA. Each reaction was performed using approximately 10 ng of total RNA for *SRG8* and *GAPDH*. RT-PCR reactions were performed in a 25 μl volume, one cycle at 95°C for 5 min, then 40 cycles at 95°C 30 sec, 60°C 40 sec and 72°C for 40 sec. A specific standard curve was established using single-use aliquots of the same stock of RNA. In all cases, reactions were performed in triplicate on the same three independent samples of testis RNA. PCR products were cloned and sequenced to confirm their identity before undertaking the study; specificity was assessed with the melting curve analysis and confirmed on a 3% agarose gel after each experiment. The relative amount of *SRG8* transcript was normalized by the amount of *GAPDH* transcript in the same cDNA.

Northern blot analysis

A cDNA probe was labelled with [α-^{32}P]dCTP in a PCR reaction. PCR amplification cycles were the same as above. The probe was purified through a Sephadex G-50 column. After prehybridization at 65°C for 1 h, the membrane with RNAs was hybridized with *SRG8* cDNA at 65°C overnight followed by washing three times with 2× SSC / 0.1% SDS at 65°C for 10 min and then twice with 0.1× SSC / 0.5% SDS at 62°C for 15 min. An autoradiograph was obtained by exposing an X-ray film at −70°C for 3 days.

Construction of fusion plasmid pEGFP-C3/SRG8

To allow directional insertion of the whole ORF in the polylinker of the pEGFP-C3 vector, the forward primer P1: CGGAATTCATGGGCCCCGGCTGGGATGTC (*Eco*RI) and the reverse primer P2: ACGCGTCGAGCTTCAGGGTAGGGACACTAT (*Sal*I) were used to amplify the target gene. After PCR amplification, the *SRG8* coding sequence was subcloned in pMD18-T vectors. Subsequently, the recombinant pMD18-T-*SRG8* and the pEGFP-C3 vectors were digested with the restriction endonucleases *Eco*RI and *Sal*I (MBI). Both digestion reactions were checked by 1.7% agarose gel electrophoresis. Then, a mixture of the digested pEGFP-C3 vector and the purified fragments was added to an equal volume of phenol/chloroform (1:1) and the supernatant was precipitated with 3 volumes of 100% ethanol and washed with 70% ethanol. *E. coli* DH5α

cells were transformed with 2 μl of the ligation reaction products and spread on an LB agar plate containing 100 μg/ml kanamycin. Finally, individual clones were checked for the correct construction of the fusion plasmids by DNA sequencing.

Cell culture and transfection

HeLa cells were grown in minimal essential medium supplemented with 10% foetal bovine serum and 50 μg/ml each of penicillin and streptomycin. Cells were seeded at a density of 2×10^6 cells on 35-mm culture dishes and then transiently transfected with plasmid DNA using Lipofectamine 2000 (Invitrogen). Prior to microscopy, cells were replated after transfection onto 35-mm culture dishes and grown in medium containing 20 mM HEPES buffer, pH 7.4. The correct colonies displayed strong and even green fluorescence when excited by blue light under the fluorescence microscope. Images were captured digitally and imported into Adobe Photoshop 5.0 for formatting. Meanwhile, the transfected cells were also subjected to cell cycle analysis.

Cell cycle analysis

Cells were trypsinized, pelleted and fixed on ice for 1 h in 70% ethanol and stored at 4°C. The fixed cells were washed with PBS, then washed with PBS containing 0.1% Triton X-100, 0.1 mM EDTA. The fixed cells were treated with 100 μg/ml DNAase-free RNAase for 1 h at 37°C and then stained with 50 μg/ml propidium iodide at room temperature in the dark before analysis. Flow cytometry was performed on a Becton Dickinson FACScan flow cytometer using Lysys II software.

Results

DDD and identification of new gene

We used DDD to compare nine testis cDNA libraries against 70 cDNA libraries derived from other organs. Hits were subdivided into known proteins (115), hypothetical proteins (31) and novel ESTs (32), which included many testis-specific proteins or testis-development-related proteins such as TSP-NY, TPX1, NYD-SP20, NYD-SP26, NYD-TSP1, sperm-related proteins such as SPAG6, SPAG11, HSP40, HSP70, CREB1, TSARG1, TSARG2, some transcriptional factors, ribosomal proteins and enzymes (data not shown). We identified a new contig of ESTs (HS.326528) as testis-specific.

Molecular cloning of full-length cDNA

The PCR product of 982 bp (figure 1) was sequenced, and the result was identical with predicted sequence. BLAST analysis of the sequence against the nr database showed that it represented a new gene. This gene was named *SRG8* (GenBank accession number AY489187). The gene, whose full cDNA length is 1044 bp, containing three exons and two introns, encodes a 105-amino-acid putative protein (figure 2). There is a start codon ATG at nucleotide positions 213 to 215 and a stop codon TGA at nucleotide positions 529 to 531. The sequence GgaatgGcc, which is in accordance with Kozak's rule, was found in the start region of the ORF, and a potential polyadenylation signal (AATAAA) was found at the 3′ end. The boundaries between exons and introns are coincident with the gt-ag rule (table 1).

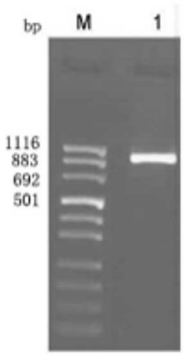


Figure 1. PCR amplification of testis-specific EST (M, DNA marker).

Bioinformatics analysis of SRG8

The gene was mapped to human chromosome 15 at 15q26.2. There is no transmembrane region or signal peptide in the predicted protein. The protein has a theoretical molecular mass of 11.7 kDa and a calculated isoelectric point of 10.09. PSORT analysis showed that there is a 52.2% possibility of localization of the protein in the nucleus. SMART analysis indicated that there are no similarities with known proteins, indicating that the protein is a novel one.

Table 1. Exon–intron junctions of *SRG8* gene.

Exon	Exon size (bp)	5′ splice donor	Intron size (bp)	3′ splice donor	Intron
1	296	TCCTAG**gt**a	408		1
2	109	AGGAAG**gt**g	734	c**ag**AACCTCG	2
3				c**ag**AACCTC	3

Uppercase and lowercase letters indicate exon and intron sequences, respectively. Conserved splice donor and acceptor dinucleotide are indicated in bold.

Expression pattern of SRG8

Expression analysis of *SRG8* was performed in adult human testis, heart, small intestine, skeletal muscle, spleen, liver, kidney, lung, epididymis and ovary tissues by RT-PCR. The results (figure 3) showed that *SRG8* gene was expressed only in adult testis and not expressed in other tissues. A Northern blot (figure 4) showed that *SRG8* gene had a 1.1-kb transcript and was abundantly expressed in testis, not in other tissues. Levels of *SRG8* transcript in different developmental stages of human testis were determined by real-time PCR, and considerably high level of expression was observed in adult testis, moderate expression in sperm, and low expression in foetal testis (figure 5).

Subcellular localization of EGFP–SRG8 fusion protein

To observe SRG8 fusion protein localization in mammalian cells, a pEGFP-C3/*SRG8* fusion plasmid was constructed and transiently introduced into HeLa cells by liposome transfection. Under a fluorescence microscope, the green fluorescence produced by expression of pEGFP-C3/*SRG8* was detected in nuclei of HeLa cells 24 h after transfection, while the fluorescence produced by pEGFP-C3 (no *SRG8*) was detected throughout the cells (figure 6). In consonance with the prediction by bioinformatics methods, these result suggested that the *SRG8*-encoded product is a nuclear protein.

agcgtggctctttggagcacccacttcacttctcccatgtagttcctgagtgcaggactctcctgaagaaggggcgtctaaaaggccagccca
tgcatcttgttcagaccaggctgctccatctccgctaagcccagggatccacttgtgtgtctttctgggataattgtctttagtggcagcctgggcc
tcggggaaggaagggcatgaggaatggccccggctgggatgtcaggggcccaggacaactcctgtctctaccaggaaattgccccctctttt
* M A P A G M S G A Q D N S C L Y Q E I A P S F
cagaggctgccttgtcctagaacctcgtcgcgacatttctcagaagccatgacatgtccctgcggctggaggcctttcaagggtggcccagga
Q R L P C P R T S S R H F S E A M T C P C G W R P F K G G P G G
ggcctcaagggcccggtgtggcctgcaaaggaagagaacagctgttctcacggaaggatccaaagggttcaaaggcgaagggtgccatct
L K G P V K E E N S C S H G R I Q R V Q R R R V P S A S
gcttccccactgattcagaagataaacaggagaagtgtcctgtttcacccatattgctggtcctgatttttcaatagtgtccctaccctgaagctctt
P L I Q K I N R R S V L F H P Y C W S *
ccattggtcatcatggagttagctaacagcccagggaactggtctcttttccctttccctttaccacctggaactcacctggaaagggcaatcag
ctggctctatatcagacgccaccacctggtctttgaagtttctgtcctaccagtgccacagctccctctcccacttctgaactacctccaagagag
accatcctcagcagcctagtatgttatgctagtcctggaatggtgtattctggatgtctactcaaggccaaaccccagtattgcctttcgtttcccta
ttcaaggccaaggctagattctcttcattgtctccatttctctttctccagagctcctatagggcaagacactgtttctcatatatttctggatccttgg
agcttagctcggagcttggcatacatgcccagcatgaactttctg**aataaa**ataatttcaaatgctgtatacaggaaaaaaaaaaaaaaaa

Figure 2. cDNA sequence of *SRG8* gene and predicted sequence of SRG8 polypeptide. Polyadenylation signal is underlined, stop codons are indicated by an asterisk (*).

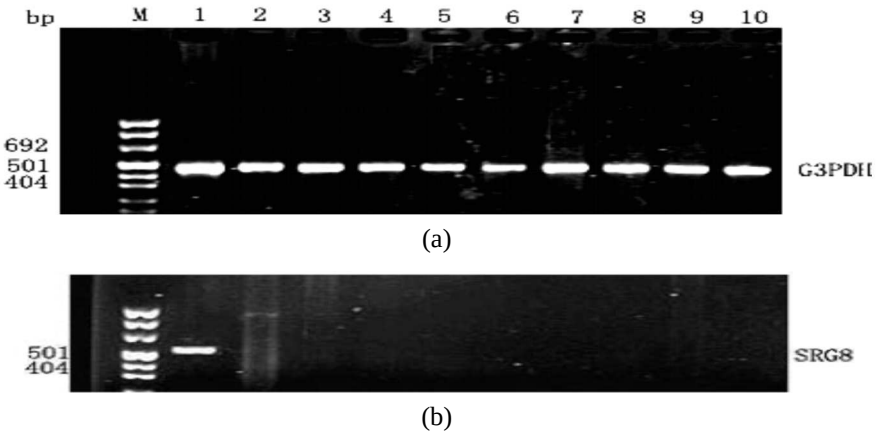


Figure 3. RT-PCR analysis of *SRG8* gene expression in various adult tissues. (a) Amplification of *GAPDH* in multiple tissues, (b) amplification of *SRG8* in adult testis (1, testis; 2, heart; 3, small intestine; 4, skeletal muscle; 5, spleen; 6, liver; 7, kidney; 8, lung; 9, epididymis; 10, ovary; M, pUC Mix8).

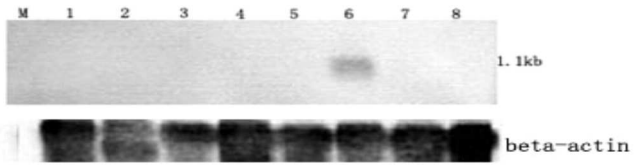


Figure 4. Northern blot analysis of *SRG8* expression. 1-8, RNA from brain, heart, liver, lung, kidney, testis, ovary, spleen, respectively. A band of about 1.1 kb is detected only in testis, while *β-actin* was detected in all tissues.

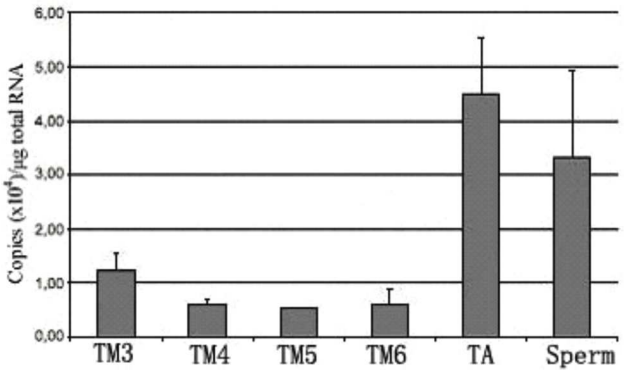


Figure 5. Quantitative analysis of *SRG8* transcript in different developmental stages of testis and sperm by real-time PCR. Standard curves for *SRG8* and *GAPDH* were generated by serial dilution of each plasmid DNA. The expression level of the *SRG8* transcript was normalized to that of the *GAPDH* transcript which was measured in the same cDNAs. Values are expressed as copy numbers of the target gene in 1 μg of total RNA. Data were obtained from triplicate experiments and are indicated as mean±s.d. (TM3–TM6, RNA from testis of 3, 4, 5, 6 months, respectively; TA, RNA from adult testis).

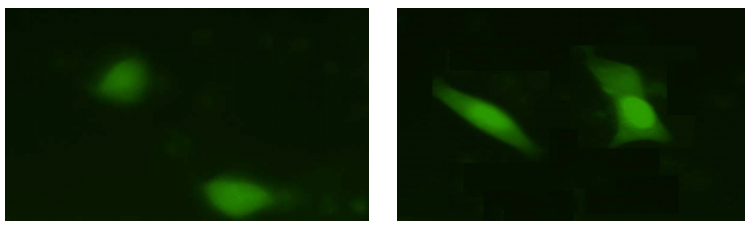
Effects of SRG8 on cell proliferation

To determine whether *SRG8* expression causes changes in cell proliferation, the effects on tumour cell growth were examined *in vitro*. The percentage of cells resident in each cell-cycle phase is indicated (figure 7). In control HeLa cells, the cell-cycle distribution is 79.4% of cells in G1 phase, 11.3% in S, and 4.3% in G2, whereas in transfected cells the distribution is 64.4% of cells in G1, 24.5% in S, and 11.1% in G2. This indicates that *SRG8* expression can accelerate traversal through the S phase and entry into G2.

Discussion

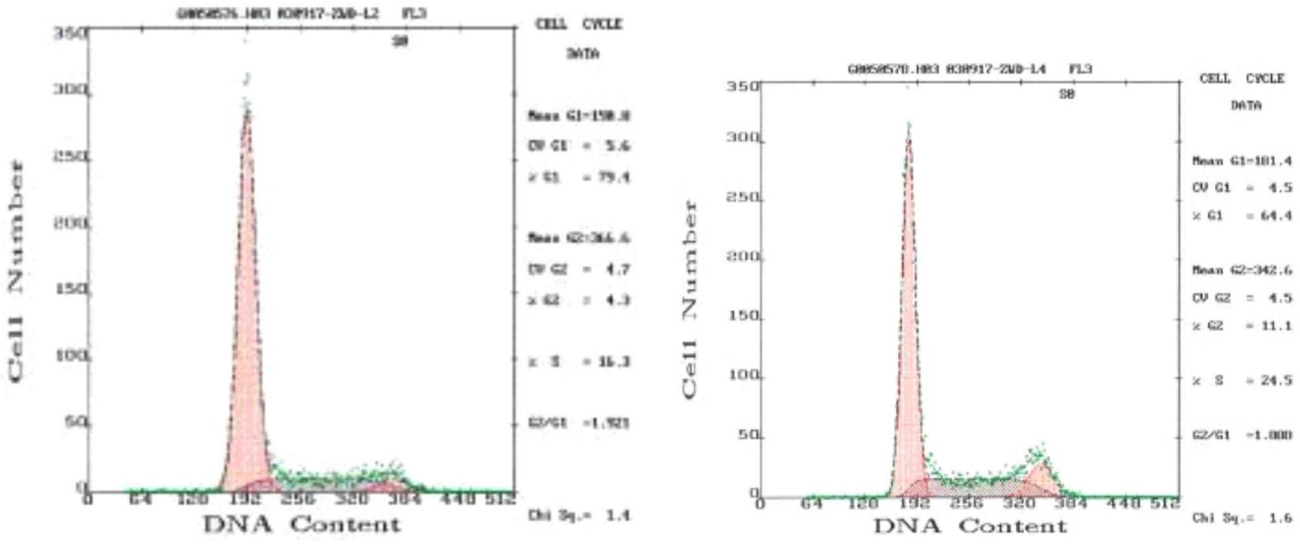
High-throughout gene expression techniques (microarrays, genechips) to identify genes are becoming available. However, the techniques are not cost-effective for average laboratories. Furthermore, these methods introduce a bias in that only a limited number, usually one-tissue-derived RNA, can be used for the initial analysis (Zhang *et al*. 1997). In this context, data-mining of databases provides a parallel approach to rapidly establish transcript-based fingerprinting. We chose the DDD protocol. A database of ESTs, both known and novel, that are differentially expressed at least 10-fold was created for nine testis cDNA libraries. An electronic Northern blot based on cDNA sources was created for each hit in DDD analysis. The vast sequence database was thus reduced to approximately 115 known proteins, 31 putative proteins and 32 novel EST contigs, which include many testis-specific proteins or testis-development-related proteins such as TSP-NY, TPX1, NYD-SP20, NYD-SP26, NYD-TSP1, sperm-related proteins such as SPAG6, SPAG11, HSP40, HSP70, CREB1, TSARG1, TSARG2, some transcriptional factors, ribosomal proteins and enzymes. These results demonstrate the utility of DDD for rapid gene discovery in human testis. Electronic expression profiling to identify testis-specific genes is likely to have false positives. However, the true lead genes can be rapidly validated by RT-PCR using appropriate cDNAs. RT-PCR validation of test genes indicated an expression profile consistent with DDD prediction.

Multiple-tissue RT-PCR and Northern blots indicated that *SRG8* gene was expressed only in human testis; no signal was found in other tissues. Real-time PCR showed that *SRG8* gene was highly expressed in adult testis, moderately expressed in sperm, and had low expression in foetal testis. Many genes could be expressed in human testis, but only a few testis-specific genes have been found, most of them structural genes (Weitzel *et al*. 2003).



(a) (b)

Figure 6. Flurescence microscopy of HeLa cells transfected with (a) PEGFP-C3 and (b) PEGFP-C3/ *SRG8* fusion construct (400×).



(a) (b)

Figure 7. Cell cycle analysis of (a) HeLa cells transfected with *SRG8* and (b) untransfected HeLa cells.

SRG8 is a testis-specific gene that may play an essential role in testis function. The mechanism and significance of enhanced expression of *SRG8* in different development stages of testis remains to be defined. SRG8 protein was mostly found in the cell nucleus and could accelerate cell growth, so we suppose that SRG8 may be a *trans*-acting factor.

Acknowledgement

This work was supported by Special Funds for Major State Basic Research 973 of China (No. G1999055901).

References

Adams M. D., Kerlavage A. R., Fleischmann R. F., Bult C., Lee N., Kirkness E. *et al*. 1995 Initial assessment of human gene diversity and expression patterns based upon 63 million nucleotides of cDNA sequences. *Nature* **377** (suppl.), 2–174.

Bourrouillou G., Dastugue N. and Colombies P. 1985 Chromosome studies in 952 infertile males with a sperm count below 10 million/ml. *Hum. Genet*. **71**, 366–367.

Cooke H. J., Hargreave T. and Elliott D. J. 1998 Understanding the genes involved in spermatogenesis: A progress report. *Fertil. Steril*. **69**, 989–995.

Dix D. J., Allen J. W., Collins B. W., Mori C., Nakamura N., Poorman-Allen P. *et al*. 1996 Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. *Proc. Natl. Acad. Sci. USA* **93**, 3264–3268.

Liu S. F., Li L. Y., Fu J. J., Liu G., Xing X. W. and Lu G. X. 2002 Rapid identification of human testis spermatocyte apoptosis-related gene, TSARG2, by nested PCR and draft human genome searching. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* (Shanghai) **34**, 378–382.

Liu G., Lu G. X. and Xing X. W. 2004 Molecular cloning of TSARG6 gene related to apoptosis in human spermatogenic cells. *Acta Biochim. Biophys. Sin* (Shanghai) **36**, 93–98.

Ma K., Inglis J. D., Sharkey A., Bickmore W. A., Hill R. E., Prosser E. J. *et al*. 1993 A Y chromosome gene family with RNA2 binding protein homology: Candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell* **75**, 1287–1295.

Reijo R., Lee T. Y., Salo P., Alagappan R., Brown L. G., Rosenberg M. *et al*. 1995 Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA2 binding protein gene. *Nat. Genet*. **10**, 383–393.

Robbins R. J. 1996 Bioinformatics: essential infrastructure for global biology. *J. Computat. Biol*. **3**, 465–478.

Ruggiu M., Speed R., Taggart M., Mckay S. J., Kilanowski F., Saunders P. *et al*. 1997 The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* **389**, 73–77.

Schmit A. O., Specht T. and Beckman G. 1999 Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucl. Acids Res*. **27**, 4251–4260.

Sun C., Skaletsky H., Birren B., Devon K., Tang Z., Silber S. *et al*. 1999 An azoospermic man with a *de novo* point mutation in the Y chromosomal gene *USP9Y*. *Nat. Genet*. **23**, 429–432.

Vogt P. H., Edelmann A., Kirsch S., Henegariu O., Hirschmann P., Kiesewetter F. *et al*. 1996 Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet*. **5**, 933–943.

Weitzel J. M., Shiryaeva N. B., Middendorff R., Balvers M., Radtke C. and Ivell R. 2003 Testis-specific expression of rat mitochondrial glycerol-3-phosphate dehydrogenase in haploid male germ cells. *Biol. Reprod*. **68**, 699–707.

Wheeler D. L., Chappey C., Lash A. E., Leipe D. D., Madden T. L., Schuler G. *et al*. 2000 Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res*. **28**, 10–14.

Xing X. W., Li L. Y., Fu J. J., Zhu W. B., Liu G., Liu S. F. *et al*. 2003 Cloning of cDNA of TSARG4, a human spermatogenesis related gene. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* (Shanghai) **35**, 283–288.

Zhang L., Zhou W., Velculescu V. E., Kern S. E., Hruban R. H., Vogelstein B. *et al*. 1997 Gene expression profiles in normal and cancer cells. *Science* **276**, 1268–1272.

Received 12 November 2004; in revised form 22 February 2006