Journal of Biochemistry and Molecular Biology, Vol. 38, No. 6, November 2005, pp. 739-747

JOURNAL OF
Biochemistry and Molecular Biology

Molecular Cloning and Bioinformatic Analysis of *SPATA4* Gene

Shang-feng Liu[†,‡], Chao Ai[†], Zhong-qi Ge[†], Hai-luo Liu[‡], Bo-wen Liu[†], Shan He[†] and Zhao Wang[†,‡,*]

[†]Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, People's Republic of China
[‡]Medical School, Tsinghua University, Beijing 100084, People's Republic of China

Received 18 August 2005, Accepted 23 September 2005

Full-length cDNA sequences of four novel *SPATA4* genes in chimpanzee, cow, chicken and ascidian were identified by bioinformatic analysis using mouse or human *SPATA4* cDNA fragment as electronic probe. All these genes have 6 exons and have similar protein molecular weight and do not localize in sex chromosome. The mouse *SPATA4* sequence is identified as significantly changed in cryptorchidism, which shares no significant homology with any known protein in swissprot databases except for the homologous genes in various vertebrates. Our searching results showed that all SPATA4 proteins have a putative conserved domain DUF1042. The percentages of putative SPATA4 protein sequence identity ranging from 30% to 99%. The high similarity was also found in 1 kb promoter regions of human, mouse and rat *SPATA4* gene. The similarities of the sequences upstream of *SPATA4* promoter also have a high proportion. The results of searching SymAtlas (http://symatlas.gnf.org/SymAtlas/) showed that human *SPATA4* has a high expression in testis, especially in testis interstitial, leydig cell, seminiferous tubule and germ cell. Mouse *SPATA4* was observed exclusively in adult mouse testis and almost no signal was detected in other tissues. The pI values of the protein are negative, ranging from 9.44 to 10.15. The subcellular location of the protein is usually in the nucleus. And the signal peptide possibilities for SPATA4 are always zero. Using the SNPs data in NCBI, we found 33 SNPs in human *SPATA4* gene genomic DNA region, with the distribution of 29 SNPs in the introns. CpG island searching gives the data about CpG island, which shows that the regions of the CpG island have a high similarity with each other, though the length of the CpG island is different from each other. This research is a fundamental work in the fields of the bioinformational analysis, and also put forward a new way for the bioinformatic analysis of other genes.

Keywords: Gene structure, Promoter, *SPATA4*, Testis-specific expression

Introduction

For the full advancement of the methods of sequencing the genome of the organism, the gene bank is more and more abundant gradually. The availability of genome sequences and functional genomics data from multiple species enables us to compare the composition of biomolecular systems like biochemical pathways and protein complexes between species. From the abundant gene bank data, nowadays it is much easier to compare biomolecules from any organism, draw the evolution trees of different species, analyze the first or second structure of the protein, localize the protein in a cell, get posttranslational modification and carry topology prediction. In all, we have got a key of higher accuracy with more abundant data to access for the analyze of the biological information of all organisms. With the full advancement of the bioinformatic analysis data, we have made essential breakthrough in many fields of the biological analysis.

Testis spermatogenic cell apoptosis is a complicated process involving multiple gene interactions. It has been confirmed that the spermatogenic process is controlled by many testis spermatogenic cell apoptosis-related genes, such as *Mcl-1* (Krajewskis *et al.*, 1995), *p53* (Riou G *et al.*, 1995), *CREM* (Blendy *et al.*, 1996), *Fas* (Nagata and Golstein, 1995), *Hsp* (Dix, 1996; Sarge and Cullen, 1997), *TRAIL* (Bretz *et al.*, 1999), *c-myc* (Kodaira K *et al.*, 1996) and *TR2* (Guo *et al.*, 1999). But all of these genes mentioned above are widely

The cDNA sequences reported in this paper have been submitted to GenBank and have accession number AY040204, AF395083, AY651919, AY651920, AY653229, AY660661 and AY970819.

*To whom correspondence should be addressed.
Tel: 86-10-62772241; Fax: 86-10-62772240
E-mail: zwang@tsinghua.edu.cn

distributed. In other words, they are not testis-specific genes. From this, we can consequently think about more other genes that have not been identified but related to the testis spermatogenic cell apoptosis process. And the most important, there must be some testis-specific genes playing an important role in the apoptosis process. Cloning new apoptosis-related and testis-specific novel genes is a key to further understand the apoptosis mechanism and the biological process of germ cells, and it is of momentous significance to clarify the physiological and pathological process of spermatogenesis.

Liu *et al.* have cloned the cDNA of human (Liu *et al.*, 2004a) and mouse *SPATA*4 gene (Liu *et al.*, 2002) in 2002. Considering the homology between the different species, we use these cDNA as probes, and then screened out *SPATA*4 gene in other species such as rat (Liu *et al.*, 2004b), chimpanzee, cow and some fish such as zebrafish (Liu *et al.*, 2005) and rainbow trout (Liu *et al.*, 2005). The in-depth bioinformatic analysis on *SPATA*4 genomic sequences of various vertebrates shows that *SPATA*4 genes have similarities in these species. We compared the protein sequences of SPATA4 of these species and drew their phylogenetic tree. Other bioinformatic analysis such as subcellular localization, post-translational modification and topology prediction are also carried out for more bioinformation about *SPATA*4 gene. Because of the limits of the laboratory conditions and outlay of the task, much work is still waiting at the front. The comparison of the bioinformatics in different species provided important information to the comprehension of human *SPATA*4 gene regulation mechanism. This analysis work of *SPATA*4 gene will be the essential bricks of the scientific research building of the testis spermatogenic cell apoptosis process.

Materials and Methods

Silicon cloning and computer-aided analysis of *SPATA*4 full-length cDNA in various vertebrates (Liu, 2004b; Liu, 2005; Liu BW, 2005) Based on human and mouse *SPATA*4 sequences we used Blast algorithms through the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/) to search for full-length *SPATA*4 cDNA from chimpanzee, cow, chicken and ascidian genome and submitted to the GenBank. The deduced amino acid sequence of mouse *SPATA*4 was aligned against the GenBank databases (nucleotide, EST and protein) at the National Center for Biotechnology Information and UCSC Genome Bioinformatics (http://genome.ucsc.edu/), using BLAST to search for sequence matches. Motif searches were performed with PROSITE programs (http://tw.expasy.org/). Multiple alignments and phylogenetic tree were created using the ClustalW program (Higgins *et al.*, 1994). All other bioinformatic tools used in this paper including the open reading frame (ORF) identification are at the ExPASy molecular biology WWW server of the Swiss Institute of Bioinformatics (http://www.expasy.ch/).

Protein database searching and sequence analysis NCBI's BLAST 2.0 program for protein sequences (BLAST P) was used to identify proteins exhibiting sequence similarities to human SPATA4. The sequences to be analysed were retrieved from the GenBank database using the Entrez retrieval systems (http://www.ncbi.nlm.nih.gov/Entrez). The primary analysis of the protein was carried out using ScanSitepI/Mw, PSORT and SignalP (http://tw.expasy.org/tools/#primary). The CLUSTAL W (1.74) (http://www.ch.embnet.org/software/ClustalW.html) was used for multiple sequence alignments. Other protein sequences containing the conserved amino acid motifs were retrieved from the NCBI non-

Table 1. *SPATA*4 genes in various species

Gene name	GenBank accession number	cDNA length (bp)	Exon	Isoelectric Point	gDNA length (kb)	Putative protein (aa)	Chromosome position
huaman *SPATA*4 (*Homo sapiens* spermatogenesis associated 4)	AY040204	1233	6	9.86	11.1	305	chromosome 4q34.1-34.2
chimpanzee *SPATA*4 (*Pan troglodytes* spermatogenesis associated 4)	AY660661	1169	6	9.96	10.5	305	chromosome 3_random
rat *SPATA*4 (*Rattus norvegicus* spermatogenesis associated 4)	AY653229	1086	6	9.56	9.4	323	chromosome 16
mouse *SPATA*4 (*Mus musculus* spermatogenesis associated 4)	AF395083	1088	6	9.64	9.4	295	chromosome 8B1.3
cow *SPATA*4 (*Bos taurus* spermatogenesis associated 4)	AY841904	1200	6	10.15	8.8	282	SCAFFOLD 115699
chicken *SPATA*4 (*Gallus gallus* spermatogenesis associated 4)	AY761140	860	6	9.45	6.8	250	chromosome 4
zebrafish *SPATA*4 (*Danio rerio* spermatogenesis associated 4)	AY651920	942	6	9.44	9.4	224	chromosome LG14
rainbow trout *SPATA*4 (*Oncorhynchus mykiss* spermatogenesis associated 4)	AY651919	1081	6	9.47	?	259	?
ascidian *SPATA*4 (*Ciona intestinalis* spermatogenesis associated 4)	AY970819	1214	?	9.51	?	280	?

redundant (nr) protein database. SNPs of human *SPATA*4 gene were searched in dbSNP in NCBI.

Spatio-temporal expression pattern detection SymAtlas (http://symatlas.gnf.org/SymAtlas/) was searched to determine the expression profile of human and mouse *SPATA*4 genes in multiple tissues. 44, 775 human and 36, 182 mouse transcripts including *SPATA*4 were examined in 79 different human tissues and 61 different mouse tissues using high-density oligonucleotide arrays. *SPATA*4 was searched in Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/projects/geo/) at the National Center for Biotechnology Information for spatio-temporal expression pattern.

Promoter searching and analysis of human, mouse and rat *SPATA*4 genes Comparing the cDNA sequences of *SPATA*4 with the genomic sequence revealed the structure of *SPATA*4 genes. The CpG Island Searcher (http://cpgislands.usc.edu/) was used to search for the CpG island, and genomatix (http://www.genomatix.de/) was used to the retrieval and analysis of promoters. CLUSTAL W (1.74) was used for multiple sequence alignments.

Results

Cloning full-length *SPATA*4 cDNA in various species Full-length cDNA sequences of four novel genes-*SPATA*4 in chimpanzee, cow, chicken and ascidian are identified by bioinformatic analysis using mouse *SPATA*4 or human *SPATA*4 cDNA fragment as an electronic probe (Table 1). The length of *SPATA*4 cDNA in these species ranges from 860 bp to 1233

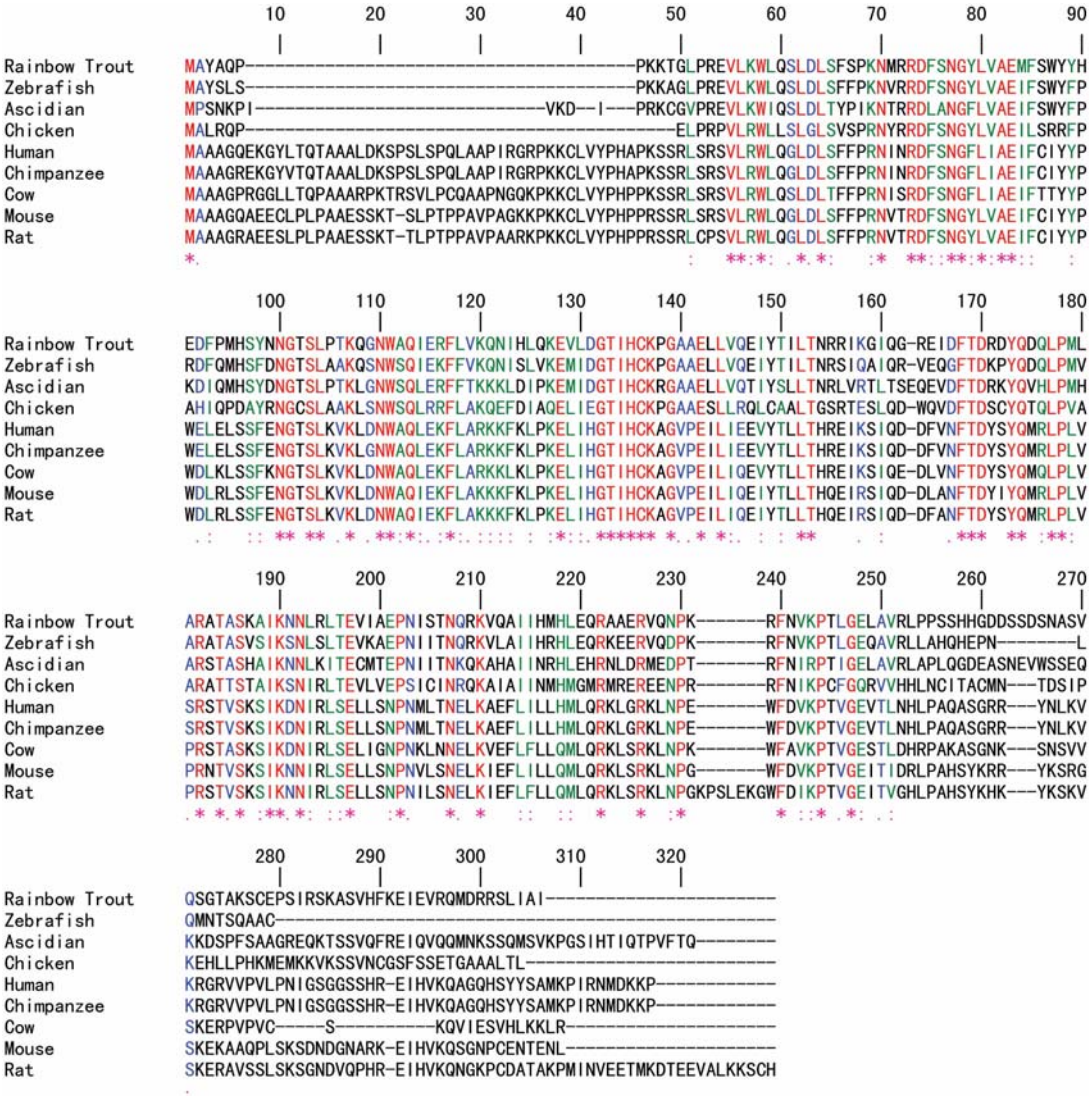

Fig. 1. Alignments of amino acid sequences for Human, Chimpanzee, Mouse, Rat, Rainbow trout, and Zebrafish SPATA4 proteins. Red means Small (small + hydrophobic (incl. aromatic-Y)) amino acid; Blue means Acidic amino acid; Magenta means Basic amino acid; Green means Hydroxyl + Amine + Basic - Q amino acid; Gray means Other amino acid; "*" means that the residues or nucleotides in that column are identical in all sequences in the alignment; ":" means that conserved substitutions have been observed; "." means that semi-conserved substitutions are observed.

(A)

	Ascidian	Chicken	Chimpanzee	Human	Cow	Mouse	Rat	Rainbow Trout	Zebrafish
Ascidian	100	46	32	32	33	33	30	56	54
Chicken		100	35	35	36	38	30	53	54
Chimpanzee			100	99	71	70	66	39	38
Human				100	72	70	66	39	38
Cow					100	70	62	43	46
Mouse						100	80	42	42
Rat							100	34	34
Rainbow Trout								100	70
Zebrafish									100

(B)

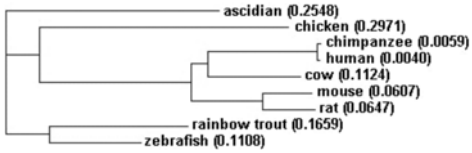

Fig. 2. Phylogenetic analysis and amino acid sequence comparison of SPATA4 protein. A: Matrix indicates the percentage identities of aligned SPATA4. B: Phylogenetic tree shows the evolutionary relationship of various vertebrates. The bootstrap tree was constructed as described in the text.

bp and the length of the putative SPATA4 protein sequences ranges from 224 aa to 323 aa. It seems that there are not any relations between the evolution degree and the length of the cDNA or putative protein from these data, and also the chromosome locations of the *SPATA*4 in different species do not have any similarities. It may be explained by the theory that evolution of autosome in different species have different ways (Fridolfsson *et al.*, 1998). Moreover, all these genes have 6 exons and have similar molecular weight and do not localize in sex chromosome though they (further study of *SPATA*4 on ascidian needs to be carried out in the future work) are gonad specific gene.

Sequence and phylogenetic analyses The sequences of SPATA4 protein from various species share high similarity with each other (Fig. 1). When the nine available amino acid sequences potentially related to one homologous family were aligned, the identity percentages of putative protein sequences ranged from 30% to 99% according to the different kinship and evolutionary degree (Fig. 2A).

The degree of SPATA4 similarity is quantified in Fig. 2B, which includes a phylogenetic tree calculated using PHYLIP software. As shown in Fig. 2B, human (*Homo sapiens*) SPATA4 protein shows the highest degree of sequence similarity with chimpanzee (*Pan troglodytes*) SPATA4 protein. Rat (*Rattus norvegicus*) and mouse (*Mus musculus*) SPATA4 are closely related when compared with zebrafish (*Danio rerio*) and rainbow trout (*Oncorhynchus mykiss*).

Protein database searching and sequence analysis When we use BLAST P to search nr database in NCBI, a 157 aa DUF1042 domain could be found in *SPATA*4 gene family (Fig. 3A). The nine available amino acid sequences were aligned as Fig. 3B. The analysis of the protein of SPATA4 from these species gives several useful information. The PI value of the protein in these species are all negative, ranging from 9.44 to 10.15. The subcellular location of the protein is usually in the nucleus, with the exception of cow's and chicken's which locate in mitochondrial. The signal anchor probabilities are all zero, and the signal peptide probabilities are also very low, but the cow and the chicken are exceptions again.

SNP analysis Using the SNPs data in NCBI, we could found 33 SNPs in human *SPATA*4 gene genomic DNA region, with the distribution of 29 SNPs in the introns at random possibilities. There are 9 SNPs localized on intron 1, 13 SNPs on intron2, 1 on intron 3 and 5 more on intron 4. Furthermore, one untranslated SNP do exist but not marked. In contrast, the distribution of SNP on exons is more concentrated. Besides, there are 3 SNPs in exon3, and 1 SNP in exon4. Noticeably, the SNPs in different exons belong to different types, while SNPs in the same exon belong to the same class (Fig. 4).

Spatio-temporal expression pattern detection SymAtlas (http://symatlas.gnf.org/SymAtlas/) was searched to determine the expression profiles of human and mouse *SPATA*4 in multiple tissues. The results show that human *SPATA*4 has a high expression in testis, especially in testis interstitial, leydig cell, seminiferous tubule and germ cell (Fig. 5A). Mouse *SPATA*4 was observed exclusively in adult mouse testis and almost no signal was detected in other tissues (Fig. 5B).

Expression profiles of mouse *SPATA*4 transcripts was displayed in different development stages of mouse testis (Fig. 6). Mouse *SPATA*4 has a low expression before day 10 (the birth day was enacted as day 0) and is unregulated at day 14.

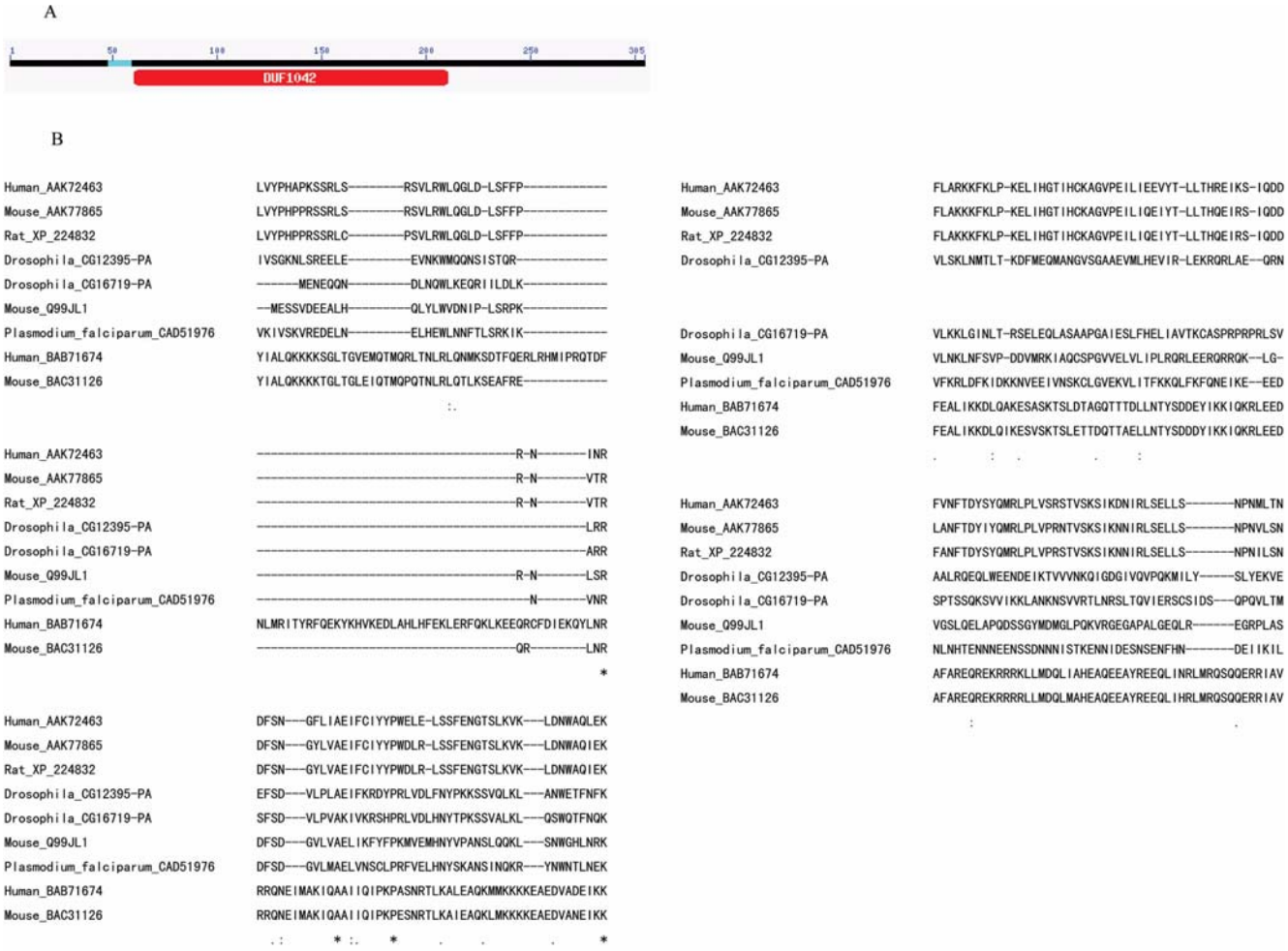

Fig. 3. Putative conserved domains of SPATA4 protein and sequence alignments of the conserved regions of proteins. (A), putative conserved domains DUF1042 in SPATA4 protein; (B), sequence alignments of the conserved regions of nine proteins.

After that, an abundant expression of mouse *SPATA4* appeared and then remained relatively stable in adult mouse testis.

CpG island analysis By carrying the CpG island searching, we get the data of CpG island from the species containing human, mouse, rat, chimpanzee, cow, chicken and zebrafish (Table 2). The data shown may give some clues to the relationship between these species. The region of the CpG island are more or less similar with each other, though the length of the CpG are different from each other. Maybe there do exist some possibilities that these species have some common transcription factors even have a similar transcription mechanism. The GC contents of these species range from 55.9% to 60.3%. Besides, we do not found CpG island in zebrafish *SPATA4* gene.

Promoter searching and analysis of human, mouse and rat *SPATA4* genes The sequences comparison of 1 kb promoter region of human, mouse and rat *SPATA4* gene showed excellent similarity (Fig. 7). Around the transcription initiation sites of these *SPATA4* genes is a highly conserved 44 bp region with 88.6% identity between human and mouse *SPATA4* genes. The similarities of the sequences upstream of *SPATA4* promoter also have a high proportion. It may indicate that these species have some common transcription factors (data not shown). And these factors may be relevant with the specificity of the distribution in the organs.

Another phenomenon is that the TSSs (Transcription Start Sequence) of human, rat and mouse *SPATA4* gene are different from each other, while the longest is human's and the shortest is rat's. But the region of the start codon of human, mouse and rat *SPATA4* gene are at the same site.

Discussion

There will soon be over 100 available genome sequences. And there is no doubt that the number and diversity of genome sequences will continue to increase for decades with higher accuracy and integrity. To handle the volume of data, rapid and accurate automation is required. It is obvious that the tasks, such as the identification of protein coding

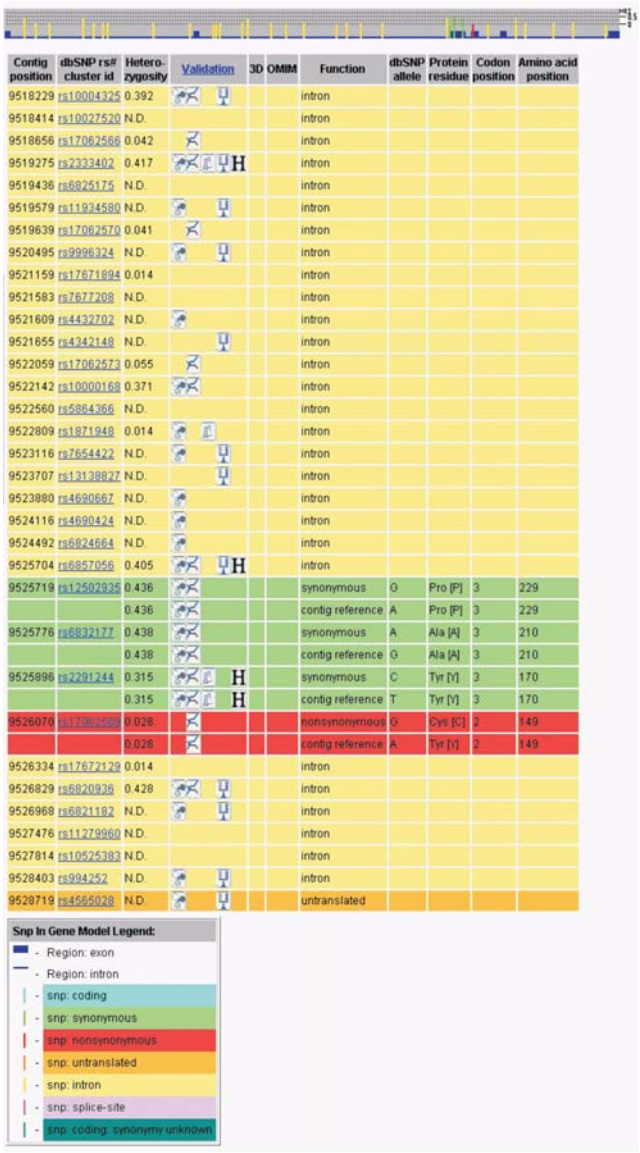

Fig. 4. Single nucleotide polymorphisms (SNPs) of human *SPATA*4. 33 SNPs were found in human *SPATA*4 gene genomic DNA region.

sequences within DNA, the extraction and translation of the sequences, the assignment of functions to all of the possibilities, are focus on the automation of the analysis on the volume of data. The biologists will be in face of the ability of getting the useful records for themselves from database (Jung and Kim, 2004).

Northern blots indicated that human *SPATA*4 gene only had a high expression in human testis, and no obvious signal was found in seven other tissues (Liu *et al.*, 2004a). Mouse *SPATA*4 gene also demonstrated the similar expression profile (unpublished). Many genes have been found to be expressed in human testis, but only a few of them were testis-specific genes (Ogi, 2001; Olesen, 2001; Imai-Senga, 2002; Weitzel, 2003). The protein of human *SPATA*4 was found mainly in the cell nucleus and could promote the growth of MCF7 cell (Liu *et al.*, 2004a). Therefore it is very different from other testis-specific genes. Bioinformational results showed that human *SPATA*4 could only be translated into 305 amino acids, so we can speculate that human SPATA4 may be a transcription factor. The Expression profiles of mouse *SPATA*4 transcripts suggests that this gene may play an important role during adolescence and maintain the spermatogenesis ability (Ogi *et al.*, 1999).

From the analysis of the protein sequences of different species, we drew an evolutionary map among these species. At the same time we made a table about the similar degree among species. It is obvious that chimpanzee is the nearest species to human being. Cow is farther than chimpanzee, while mouse, rat is farther than cow to the human. Other species such as chicken, zebrafish also accorded with traditional evolution theory. The most remarkable feature in the evolutionary tree is that ascidian, different from other vertebrates, was a sole branch from the tree. It could be suggested that ascidian was evolved from certain animals which were also the ancestor of other vertebrates from some ancient times. The ascidian evolved in a very different way compared with vertebrates. The role of the ascidian in the evolution biology and how *SPATA*4 gene functions in ascidian still need further studies.

The primary analysis of the proteins shows a visible comparison among these species. These proteins have the same value based the analyses of anchor probability, and from this data we could deduce that this SPATA4 protein is not an anchor protein, nor a secretary one. The negative value of these proteins may be a key point in their functions such as interaction with other proteins or genes. The subcellular locations of SPATA4 in many species are mostly in the nucleus, but for the cow and chicken, they are more apt to locate in the mitochondria. These locations may be a hint that SPATA4 have a high possibility to carry out its function in the nucleus.

SNPs which are located within coding regions could damage or alter the structures or functions of the protein if they are predicted to be in the important sites of the three dimensional structure (Salim *et al.*, 2005). It is less straightforward to deduce the functional importance of SNPs within regulatory regions. SNPs in *SPATA*4 have several classes, and these SNPs may have diverse function in the process.

Using CpG Island Searcher, we found a CpG island in 5' end of human *SPATA*4 gene (Liu *et al.*, 2004a). Following this clue we detected a 208 bp promoter sequence in this region, which should be the promoter of this testis-specific gene. We have identified this promoter by PCR assay (unpublished), and the function of this promoter would be examined in future studies. This promoter would be very important for gene therapy in which the corresponding genes pass through the blood-testis barrier and target the testis. No mutation was found by PCR-SSCP in 122 cases of azoospermia, severe oligzoospermia, and cryptorchidism (Liu *et al.*, 2004a),

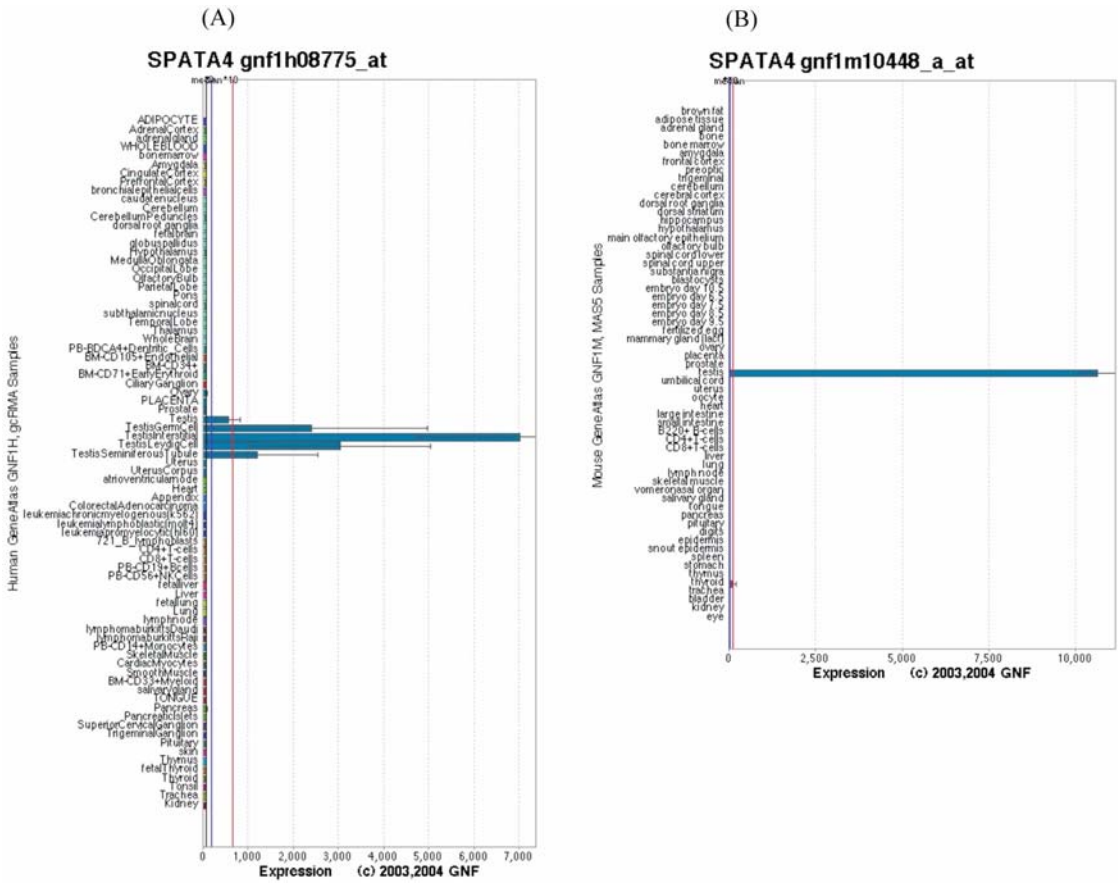

Fig. 5. Expression profiles of human and mouse *SPATA4* transcripts. Expression database on-line accession for expression profiles of human and mouse *SPATA4* (http://symatlas.gnf.org/SymAtlas/). (A), Expression profiles of human *SPATA4* gene, (B), Expression profiles of mouse *SPATA4* gene.

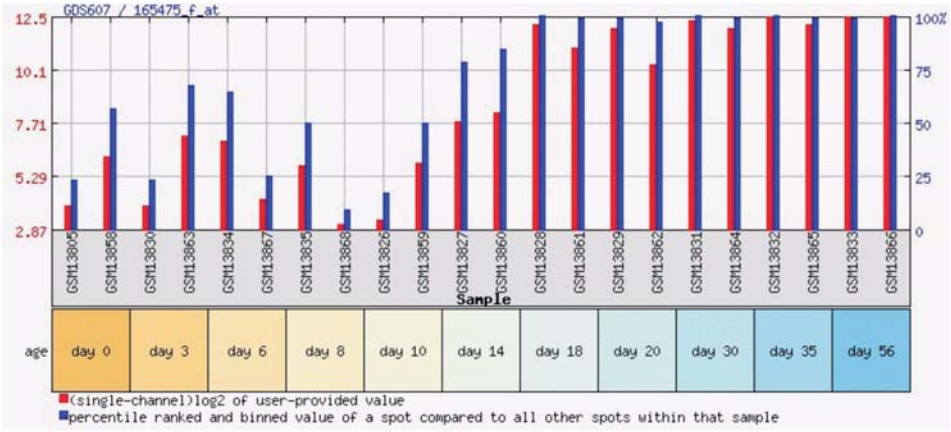

Fig. 6. Expression profiles of mouse *SPATA4* transcripts in different development stages. Mouse *SPATA4* has a low expression before day 10 (the birth day was enacted as day 0) and upregulated at day 14. After that, abundant expression of mouse *SPATA4* in testis appeared and then remained relatively stable.

suggesting that this gene might carry out its function to up-regulate or down-regulate through promoter methylation.

Another promising characteristic of human *SPATA4* promoter is temperature sensitivity. In testis, several types of heat shock proteins (HSPs) have been identified (Sarge and Cullen, 1997), but most of these proteins are widely spread in other organs. One unique function of the male reproductive system in many species was the maintenance of the testis at a temperature lower than that of the other tissues. The importance of precise thermoregulation of the testis was

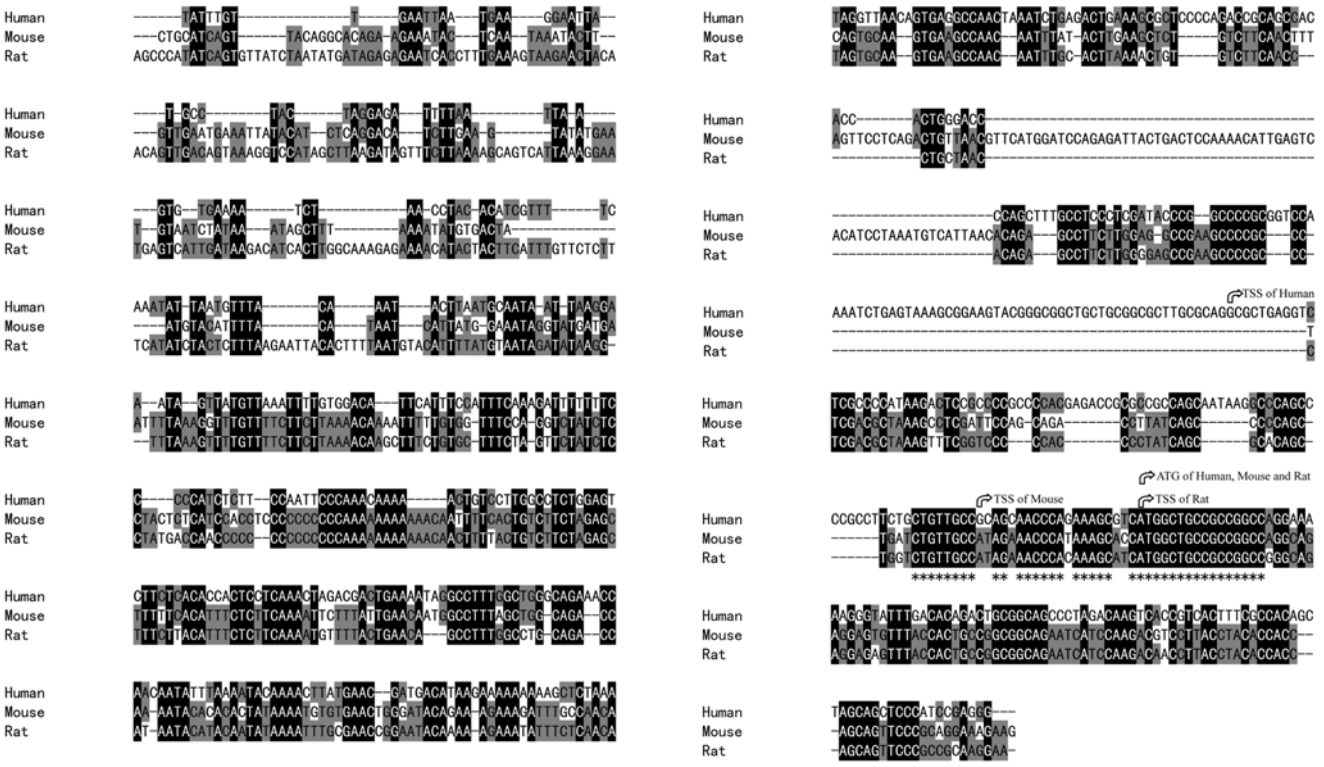

Fig. 7. Comparison of promoter sequences of human, mouse and rat *SPATA*4 genes. Promoter sequence of human, mouse and rat *SPATA*4 genes were aligned with Clustal W. Bases conserved in all these species were presented in black, while those conserved in any two species were in gray; Arrow indicated transcription start sites and translation start sites of each *SPATA*4 gene; The highly conserved 44 bp regions were marked by asterisk. Human, human *SPATA*4 gene; Mouse, mouse *SPATA*4 gene; Rat, rat *SPATA*4 gene; TSS, transcription start site; ATG, translation start site.

Table 2 The analysis of CpG island of *SPATA*4

	%GC	Length(bp)	Localization
Human	60.3	627	-295——+332
Chimpanzee	60.8	612	-280——+332
Rat	55.9	392	-182——+210
Mouse	56.1	326	-136——+190
Zebrafish		Not found	
Cow	61.0	634	-287——+347
Chicken	63.3	908	-208——+700

evidenced by the fact that even slight elevations of scrotal temperature are associated with infertility (Sarge and Cullen, 1997).

The sequence comparison of 1 kb promoter regions of human, mouse and rat *SPATA*4 gene showed excellent similarity among these promoters. Several transcription factors with common binding sites such as GATA and Sox-5 could be found in this region, suggesting that different *SPATA*4 promoters could be regulated by the same category of transcription factors (data not show). Besides, the discrepancy between the promoters of human and mouse also exists. Transient transfection assay showed that human *SPATA*4 promoter could be activated in mouse Sertoli cell line TM4, but the level is lower than that by mouse *SPATA*4 promoter (unpublished). It seemed that these *SPATA*4 genes preserved consistency during evolution, while they produced diverged mechanisms to gradually adapt to environmental requirements.

Altogether, this bioinformatic research on the *SPATA*4 represents the first functional characterization of this gene that may perform important functions in the spermatogenesis as well as in testicular germ cell tumorigenesis. According to the cloning and function study of human testis spermatogenesis related genes, we have begun our trip of the understanding of the complex process of spermatogenesis, such as when testis cells apoptosis starts, how signals are transducted among those cells and what roles do they play in fertility. It appears that this work only focuses on *SPATA*4, but the significance of the work is far more important than the *SPATA*4 itself. The

methods in this work could be carried out with other genes and other biomolecules to get more information of the biosystem. It is also a more powerful mean for the further research on the bioinformation of genes and other biomolecules. This research is a fundamental work in the fields of the bioinformational analysis, and also put forward a new way for the bioinformatic analysis of other genes.

References

Blendy, J. A., Kaestner, K. H., Weinbauer, G. F., Nieschlag, E. and Schutz, G. (1996) Severe impairment of spermatogenesis in mice lacking the *CREM* gene. *Nature* **380**, 162-165.

Bretz, J. D., Rymaszewski, M., Arscott, P. L., Myc, A., Ain, K. B., Thompson, N. W. and Baker, J. R. Jr. (1999) *TRAIL* death pathway expression and induction in thyroid follicular cells. *J. Biol. Chem*. **274**, 23627-23632.

Dix, D. J., Allen, J. W., Collins, B. W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E. H. and Eddy, E. M. (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.

Fridolfsson, A. K., Cheng, H., Copeland, N. G., Jenkins, N. A., Liu, H. C., Raudsepp, T., Woodage, T., Chowdhary, B., Halverson, J. and Ellegren, H. (1998) Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc. Natl. Acad. Sci. USA* **95**, 8147-8152.

Guo, C. X., Hu, Z. Y., Zhou, Y. J., Mu, X. M. and Liu, Y. X. (1999) The expression of orphan receptor *TR2* mRNA in spermatocyte of rat and *Macaca mulatta* cryptorchidism. *Chin. Sci. Bull*. **44**, 2539-2544.

Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Imai-Senga, Y., Sun-Wada, G. H., Wada, Y. and Futai, M. (2002) A human gene, ATP6E1, encoding a testis-specific isoform of H(+)-ATPase subunit E. *Gene* **289**, 7-12.

Jung, D. H. and Kim, D. H. (2004) Characterization of isoforms and genomic organization of mouse calumenin. *Gene* **327**, 185-194.

Kodaira, K., Takahashi, R., Hirabayashi, M., Suzuki, T., Obinata, M. and Ueda, M. (1996) Overexpression of *c-myc* induces apoptosis at the prophase of meiosis of rat primary spermatocytes. *Mol. Reprod. Dev.* **45**, 403-410.

Krajewski, S., Bodrug, S., Krajewska, M., Shabaik, A., Gascoyne, R., Berean, K. and Reed, J. C. (1995) Immunohistochemical analysis of Mcl-1 protein in human tissues. Differential regulation of Mcl-1 and Bcl-2 protein production suggests a unique role for Mcl-1 in control of programmed cell death *in vivo*. *Am. J. Pathol*. **146**, 1309-1319.

Liu, B. W., Liu, S. F., He, S., Zhao, Y., Hu, H. X. and Wang, Z. (2005) Cloning and expression analysis of gonadogenesis-associated gene *SPATA4* from rainbow trout (*Oncorhynchus mykiss*). *J. Biochem. Mol. Biol*. **38**, 206-210.

Liu, S. F., Li, L. Y., Fu, J. J., Liu, G., Xing, X. W. and Lu, G. X. (2002) Molecular cloning of SRG2, a mouse testis spermatocyte apoptosis-related gene. *Acta. Biochim. Biophys. Sin*. **34**, 796-799.

Liu, S. F., Lu, G. X., Liu, G., Xing, X. W., Li, L. Y. and Wang, Z. (2004a) Cloning of a full-length cDNA of human testis-specific spermatogenic cell apoptosis inhibitor TSARG2 as a candidate oncogene. *Biochem. Biophys. Res. Commun*. **319**, 32-40.

Liu, S. F.,, He, S., Liu, B. W., Zhao, Y. and Wang, Z. (2004b) Cloning and characterization of testis-specific spermatogenesis associated gene homologous to human *SPATA4* in rat. *Biol. Pharm. Bull*. **27**, 1867-1870.

Liu, S. F., Liu, B. W., He, S., Zhao, Y. and Wang, Z. (2005) Cloning and characterization of zebrafish *SPATA4* gene and analysis of its gonad specific expression. *Biochemistry (Mosc)* **70**, 773-781.

Nagata, S. and Golstein, P. (1995) The Fas death factor. *Science* **267**, 1449-1456.

Ogi, S., Tanji, N., Iseda, T. and Yokoyama, M. (1999) Expression of heat shock proteins in developing and degenerating rat testes. *Arch. Androl.* **43**,163-171.

Ogi, T., Mimura, J., Hikida, M., Fujimoto, H., Fujii-Kuriyama, Y. and Ohmori, H. (2001) Expression of human and mouse genes encoding polkappa: testis-specific developmental regulation and AhR-dependent inducible transcription. *Genes Cells* **6**, 943-953.

Olesen, C., Larsen, N. J., Byskov, A. G., Harboe, T. L. and Tommerup, N. (2001) Human FATE is a novel X-linked gene expressed in fetal and adult testis. *Mol. Cell. Endocrinol*. **184**, 25-32.

Riou, G., Barrois, M., Prost, S., Terrier, M. J., Theodore, C. and Levine, A. J. (1995) The *p53* and *mdm-2* genes in human testicular germ-cell tumors. *Mol. Carcinog*. **12**, 124-131.

Salim, M-T. and Mohammad, A. F. (2005) Identification of functional SNPs in the 5-prime flanking sequences of human genes. *BMC Genomics* **6**, 18.

Sarge, K. D. and Cullen, K. E. (1997) Regulation of *hsp* expression during rodent spermatogenesis. *Cell. Mol. Life Sci*. **53**, 191-197.

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