Animal Reproduction Science 113 (2009) 143–155

Contents lists available at ScienceDirect

Animal Reproduction Science

journal homepage: www.elsevier.com/locate/anireprosci

Identification and sequencing of remnant messenger RNAs found in domestic swine (*Sus scrofa*) fresh ejaculated spermatozoa

C.C. Yang[a], Y.S. Lin[a], C.C. Hsu[a], S.C. Wu[a,b], E.C. Lin[a], W.T.K. Cheng[a,b,*]

[a] *Department of Animal Science and Technology, National Taiwan University, Taipei 106, Taiwan, ROC*
[b] *Institute of Biotechnology, National Taiwan University, Taipei 106, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 7 March 2008
Received in revised form 25 June 2008
Accepted 1 August 2008
Available online 12 August 2008

Keywords:
Swine
Ejaculated spermatozoa
Expressed sequence tag

ABSTRACT

The existence of specific messenger RNA remnants contained within freshly ejaculated spermatozoa was described in several species. Those investigations, using high-throughput techniques to screen the population of transcripts in ejaculated spermatozoa, were limited to the probes which mostly derived from nucleic acids of testicular tissues of either human or mice. The objective of this study was to investigate mRNA remnants from ejaculated spermatozoa of the domestic swine (*Sus scrofa*), a valuable model for biomedical research. A non-redundant 5′-end complementary DNA library was generated from swine ejaculated spermatozoa. After sequence quality verification, 4562 clones remained. These clones were then clustered and assembled into 514 unique sequences including 188 contigs (36.58%) and 326 singletons (63.42%), representing those clusters containing at least two clones and those clusters without having enough similarity with other clones. These unique gene sequences were annotated in Gene Ontology (GO) hierarchy; they included biological processes (38.7%), molecular functions (39.1%) and cellular components (40.3%). Based on the analysis, a broad spectrum of messenger RNAs existed in swine ejaculated spermatozoa and was closely correlated with nucleic acid binding, structural modifications, and transcriptional regulation. All of these categories are considered to have profound effects on the male reproductive system. Therefore, our work provides initial results on potential spermatozoal gene expression for future studies regarding the tightly regulated spermiogenic processes and later fertilization events.

* Corresponding author at: Department of Animal Science and Technology, National Taiwan University, No. 50, Lane 155, Sec. 3, Keelung Road, Taipei 106, Taiwan, ROC. Tel.: +886 2 27337180; fax: +886 2 27324070.
E-mail address: wtkcheng@ntu.edu.tw (W.T.K. Cheng).

0378-4320/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.anireprosci.2008.08.012

1. Introduction

In mammals, spermiogenesis is a complex process wherein the round spermatids undergo dramatic remodeling changes, condensation of the nuclear chromatin, formation of the flagellum and development of the acrosomal cap, to form spermatozoa (Wykes et al., 1997). As a result of these processes, spermatids loose most of their cytoplasm after they are transformed into spermatozoa. Based on these observations, scientists think that mature human spermatozoa are unlikely to transcribe novel RNAs and translate protein-coding RNAs (Grunewald et al., 2005). Many of the late post-meiotically transcribed genes are most likely synthesized during the final burst of transcription prior to histone-protamine exchange in haploid spermatids, and then later actively translated in elongated spermatids (Kramer and Krawetz, 1997).

With new techniques, such as reverse transcription-polymerase chain reaction (RT-PCR), *in situ* hybridization (ISH), testicular microarray and/or serial analysis of gene expression (SAGE), the ejaculated bovine (Gilbert et al., 2007), rodent (Pessot et al., 1989) and human (Wykes et al., 1997; Miller et al., 1999; Zhao et al., 2006) spermatozoa evidently retain a complex yet specific population of transcripts. Very little is known with respect to the biological functions of these protein-coding and noncoding RNAs in ejaculated spermatozoa, because it is generally accepted that spermatozoa are dormant cells. However, recent studies suggest that these transcripts might play important roles in spermatozoal development, chromatin repackaging, genomic imprinting, early zygotic development, male-derived factors that underlie idiopathic infertility in humans, and possibly provide the key to more successful animal cloning using somatic-cell nuclear transfer (Ostermeier et al., 2004; Martins and Krawetz, 2005).

To initiate an exploration of the identity, functions and regulation of mammalian spermatozoal RNAs in spermatozoal development and the events surrounding fertilization, sequenced spermatozoal cDNAs will be an especially valuable resource. The objective of this study was to characterize the ejaculated spermatozoal RNA pool of swine, providing fundamental knowledge for developing assisted reproductive technology and study of reproductive disease in humans, by spermatozoal expressed sequence tag (EST) library construction and functional annotation. This information provides a ground work upon which to examine missing spermatozoal donated RNA as a factor in idiopathic infertility.

2. Materials and methods

2.1. Spermatozoa collection

Freshly ejaculated semen was collected from sexually mature healthy swine of the Landrace breed using the gloved hand technique. Samples were evaluated under a microscope and selected for $\geq$70% on motility and morphological appearance of spermatozoa. The ejaculated semen was then washed and diluted following the modified method described by Yang et al. (2004). Briefly, fresh semen was diluted with semen diluent (D-glucose 60 g/L, sodium bicarbonate 1.2 g/L, EDTA 3.7 g/L, trisodium citrate 3.7 g/L, streptomycin 1 g/L, penicillin 0.5 g/L, pH 6.6–6.7) (Cheng, 1985) and incubated for 30 min at 37 °C for non-motile spermatozoa and other seminal debris including somatic cells to subside and to be eliminated as much as possible at next step. The upper half of the solution, motile spermatozoa enriched fraction, then was diluted with an equal volume of phosphate buffered saline (PBS; Invitrogen, Grand Island, NY, USA) and centrifuged at 300 × *g* for 10 min at 18 °C, the process was repeated and the upper half of the supernatant fraction was combined with an equal volume PBS again but centrifuged at 1200 × *g* for 5 min at 18 °C. After washing the pellet of spermatozoa three times with PBS and centrifuging at 1200 × *g* for 5 min at 18 °C, a small aliquot of the pelleted spermatozoa was inspected by microscopy to ensure the reliability and reproducibility of this purification process as well as to examine if any adherent somatic cells or their debris remained with spermatozoa. The spermatozoa then were snap-frozen in liquid nitrogen and stored at −80 °C until used.

2.2. Total RNA extraction and quality control

Total RNA was extracted from the washed spermatozoa using the RNeasy Lipid Tissue Mini kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. The preliminary quantity and purity of

Table 1
The primer sequences used for reverse transcription-polymerase chain reaction

Target	Accession number	Primer sequence	Location	Expected length (bp)
Beta-2-microglobulin	DT331342	5′-AACGGAAAGCCAAATTACCTG 5′-GTGATGCCGGTTAGTGGTCTG	155–175 393–413	259
CD45	CB288124	5′-AGAATACTGGCCGTCGATGG 5′-GCTGAACGCATTCACTCTCCT	364–383 581–601	238
c-kit	L07786	5′-GTTGATGACCTCGTGGAATGC 5′-CTGCTACTGCTGTCATTCCTAAGG	391–411 648–671	281
E-cadherin	BX672333	5′-GAAGCACAGAATCCCCAAGTG 5′-GGCGTGTTTGTCTTCCATTTC	476–496 611–631	156
Protamine P1	ES446457	5′-TCACCATGGCCAGATACAGAT 5′-AGTGCGGTGGTGTTGCTACT	92–112 256–275	184
Transition protein 1	AY609468	5′-TAAGAGAGGTGGCAGCAAAAGA 5′-TCTTCTCCGGCATTGACCTAA	123–144 307–327	205
Protamine P2	X16558	5′-GTCTATGGCAGGACCCACAG 5′-CATTTCCAACCAGGGGTGT	185–204 367–385	201

the extracted transcripts were measured at 260 and 280 nm with the Nanodrop® ND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Rockland, DE, USA). The transcripts were further analyzed using the Experion™ automated electrophoresis system (Bio-Rad, Hercules, CA, USA) for their integrity (Imbeaud et al., 2005). To rule out the possible contamination by leukocytes, testicular germ cells and epithelial cells, primer sets were designed for specific cell surface marker: CD45, c-kit and *E*-cadherin genes, respectively. A specific set of primers, for the beta-2-microglobulin protein (B2M) gene, located on separated exons, was used to eliminate the sample containing genomic DNA residues. For the positive control, protamine P1 (PRM1) gene, another set of primers was used. The primer sequences used are described in Table 1. Each amplifying reaction of a specific gene was performed using the SuperScript™ III One-Step RT-PCR System with Platinum® *Taq* DNA Polymerase kit (Invitrogen). The annealing temperature and thermal cycling parameters were slightly modified as follows: one cycle of 50 °C for 30 min, one cycle of 94 °C for 2 min, 5 cycles of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 40 s, 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 40 s, with a final extension of 5 min at 72 °C. Amplified products were then electrophoresed and visualized by ethidium bromide (AMRESCO, Solon, OH, USA) staining. Samples containing genomic DNAs, somatic cellular RNAs which showed PCR amplification products of gene described above or that were of low quality were discarded.

2.3. Expressed sequence tag library construction

An EST library was established using Dynabeads® Oligo$(dT)_{25}$ (Invitrogen) combined with the GeneRacer kit (Invitrogen) following the manufacturer's protocol. The magnetic beads and mRNAs were then separated by incubating the mixture at 65 °C for 5 min. Following the first-strand cDNA synthesis, the resulting cDNAs were amplified using TaKaRa EX Taq™ (TaKaRa, Kyoto, Japan) with GeneRacer™ 5′ and GeneRacer™ 3′ Primers. The amplified cDNA mixture was size-separated into 2 fractions, >700 bp and <700 bp, using electrophoresis with a molecular weight standard. Both fractions of PCR products were purified and then inserted into the pCR®-XL-TOPO® (Invitrogen), pCR®4-TOPO® (Invitrogen), and/or pCR®4Blunt-TOPO® (Invitrogen) vectors according to their sizes. The ligated cDNAs were subsequently transformed to One Shot® TOP10 Electrocomp™ *E. coli* (Invitrogen) by electroporation and the bacteria were grown at 37 °C overnight and stored at 4 °C until use.

2.4. DNA sequencing

Sequencing reactions were done using a combination of TempliPhi™ DNA Sequencing Template Amplification (GE Healthcare, Piscataway, NJ, USA) and BigDye® Terminator v3.1 Cycle

Sequencing kits (Applied Biosystems, Foster City, CA, USA) containing an M13 forward primer (5′-GTAAAACGACGGCCAG). All samples then were further purified using MultiScreen®$_{384}$-SEQ Plate (Millipore, Billerica, MA, USA) to remove free fluorescent nucleotides and then were sequenced with the 3730 DNA analyzer (Applied Biosystems, Foster City, CA, USA). Resulting sequences were identified by searching the current databases using Chromas sequence analysis software, version 2.13 (Technelysium Pty. Ltd., Helensvale, Queensland, Australia).

2.5. Sequence analysis and functional annotation

The sequence analysis was modified from Yang et al. (2007) and Xia et al. (2007). The raw chromatogram tracing files of swine ejaculated spermatozoal 5′-end EST sequences were transformed using the Phred program (University of Washington, Seattle, WA, USA) with quality-control criteria (Ewing et al., 1998). The parameter setting of base quality value was ≥20. Furthermore, high-quality sequences were obtained after removing low-complexity sequences, poly(A) tails and contaminating fragments using the Cross_Match program (University of Washington, Seattle, WA, USA) for vector-clipping (Gordon et al., 1998) and SeqClean program in The Institute for Genomic Research (TIGR; Rockville, MD, USA) software tool, respectively. Thereafter, these sequences were clustered according to sequence similarity to produce contigs and singletons utilizing TIGR Gene Indices clustering tools (TGICL; The Institute for Genomic Research, Rockville, MD, USA) software (Pertea et al., 2003). Briefly, the MegaBLAST program was first used (Zhang et al., 2000) to compare each sequence against the others and sequences with high similarity were sorted to the same clusters. In the second stage, the CAP3 sequence assembly program (Huang and Madan, 1999) was applied to generate consensus sequences for each contig via a multiple sequence alignment method. The comparison quality was set at ≥95% identity over a minimum of 55 bp. For the clustering results, contigs assembled along with singletons were analyzed by strands-alone Basic Local Alignment Search Tool (BLAST) program (Altschul et al., 1990), Blastn (NCBI, Bethesda, MO, USA), against three different public databases, the NCBI Unigene *Sus scrofa* database, the NCBI non-redundant nucleotide sequence database, and the TIGR Porcine Gene Index nucleotide sequence databases (SsGI; release 12.0), for gene identification and to identify homologous genes between species. Those contig and singleton sequences without known identity were referred to as putative novel transcripts. Subsequently, in order to classify the functions of the sequences with known identity, the genes were analyzed for functional annotation using the Gene Ontology (GO) database via Babelomics FatiGO web-based analysis tool (Al-Shahrour et al., 2005). In addition, the CLC Protein Workbench, version 3.0.2 (CLC bio, Aarhus, Denmark) software was used for protein alignment and characteristics analysis.

2.6. Northern blot analysis

Total RNA extracted from swine ejaculated spermatozoa and testes, 500 ng each, was fractionated on a 1.2% denaturing gel and transferred to a nylon Hybond™-N$^+$ membrane (GE Healthcare) using 0.05 M sodium hydrate (MERCK, Darmstadt, Germany). The PCR-generated fragments of the PRM1 and transition protein 1 (TNP1) genes were labeled with α-^{32}P using the Rediprime™ II Random Prime Labelling System (GE Healthcare). After the membrane was pre-hybridized with ULTRAHyb® solution (Applied Biosystems) at 42 °C for 2 h, the purified radio-labeled probes were added to the hybridization solution. The hybridization was continued for an additional 24 h. The hybridized membranes then were washed in a series of buffers with increasing stringency and subjected to phosphor-image analysis using a Typhoon 9200 scanner (GE Healthcare).

2.7. Reverse transcription-polymerase chain reaction

Total RNA extracted from ejaculated spermatozoa, 20 ng, was amplified with PRM1 and protamine P2 (PRM2) primer sets using the SuperScript™ III One-Step RT-PCR System with Platinum® *Taq* DNA Polymerase kit. The thermal cycling parameters for amplification were slightly modified: 94 °C for 15 s, 61 °C for 30 s and 72 °C for 10 s for 20, 25, 30 or 35 thermal cycles. The

amplified products were fractionated on an agarose gel and visualized by ethidium bromide staining.

3. Results

3.1. Evidence for RNA in spermatozoa

In the present study, the transcript extraction procedure estimated that each swine spermatozoon contained 0.005 pg of transcript on average. The comparative transcript profiles of normal somatic cells and spermatozoa were first visualized by electrophoresis (Fig. 1A). The RNA obtained from spermatozoa was shown to be free of 28S and 18S ribosomal RNA (rRNA) present in swine testis, but contained a homogeneous panel of RNA in various sizes under 2 kb. No RNA contaminants resulting from somatic cells in crude semen or genomic DNA residues were observed in spermartozoal RNA samples using one-step RT-PCR (Fig. 1B). Hence, it is fair to say that ejaculated swine spermatozoa do indeed contain transcriptional products which can be reverse transcribed and further amplified.

3.2. Construction and characterization of a swine spermatozoal expressed sequence tag library

The quality of the EST library was assessed prior to large-scale sequencing. Single-stranded DNA obtained from the reaction was converted into the double-stranded form and ligated into selected non-directional cloning vectors. The PCR-amplified cDNA was characterized as a distinct smear ranging from 100 to 2500 bp or longer (Fig. 2A, left) and the molecular-weight distribution was similar to the corresponding mRNA (Fig. 1A). Following the cloning reaction, colony-PCR was performed on randomly picked clones from each cloning reaction to evaluate the quality of the library. The bands located <250 bp suggested products obtained from the plasmid vector without inserts or insertion of primers. The cloning efficiency was calculated according to the result of the electrophoresis; 90% of these transformants must contain inserts $\geq$250 bp in size to be considered successful, or the cloning reaction was repeated. The size of the majority of inserts was from 300 to 800 bp with a few $\geq$1 kb (Fig. 2A, right).

Randomly picked clones, 6144 in number, were subjected to the 5′-end single-pass sequencing reaction and generating 5812 EST sequences. The low-quality bases, vector/adaptor sequences, repeat sequences and two genes within one clone were removed resulting in 4562 high-quality sequences with an average insert size of 433 bp. Clustering as well as assembling of these sequences produced 188 contigs with an average insert size of 584 bp; 326 singletons also were separated. There was only one cluster possessing two contigs. The average number of EST clones per contig was approximately 22. The 22 most abundant clusters which consisted of more than 20 EST clones were listed in Table 2. In addition, 67.5% of the full-length enriched cDNAs contained a putative translation start and stop condon. Since a large portion of these sequences have not yet been annotated, only 271 tentative unique sequences permitted mapping to Gene Ontology terms in the *Homo sapiens* field (Fig. 2B). Within a given gene ontology term at level 4 (Al-Shahrour et al., 2003), the number of clones and unique sequences were different. For the sequences that were mapped to biological process, the functions were involved mostly in metabolic and biosynthetic processes, whereas those mapped to molecular functions were mainly related to DNA, RNA and metal ion binding.

4. Discussion

4.1. Transcriptional and translational characteristics of spermatozoa

In recent years, it had been recognized that there is a population of mRNA transcripts within the mammalian mature ejaculated spermatozoon. However, the quantity of these transcripts in a spermatozoon is extremely low compared to a somatic cell, with differences sometimes reaching 600 times (Zhao et al., 2006). Use of the transcript integrity assay, double strand cDNA amplification, cloning efficiency tests or sequence analysis all indicated the swine spermatozoal RNA predominantly was composed of small-sized transcripts. This observation is in agreement with a similar study in bovine spermatozoa (Gilbert et al., 2007).

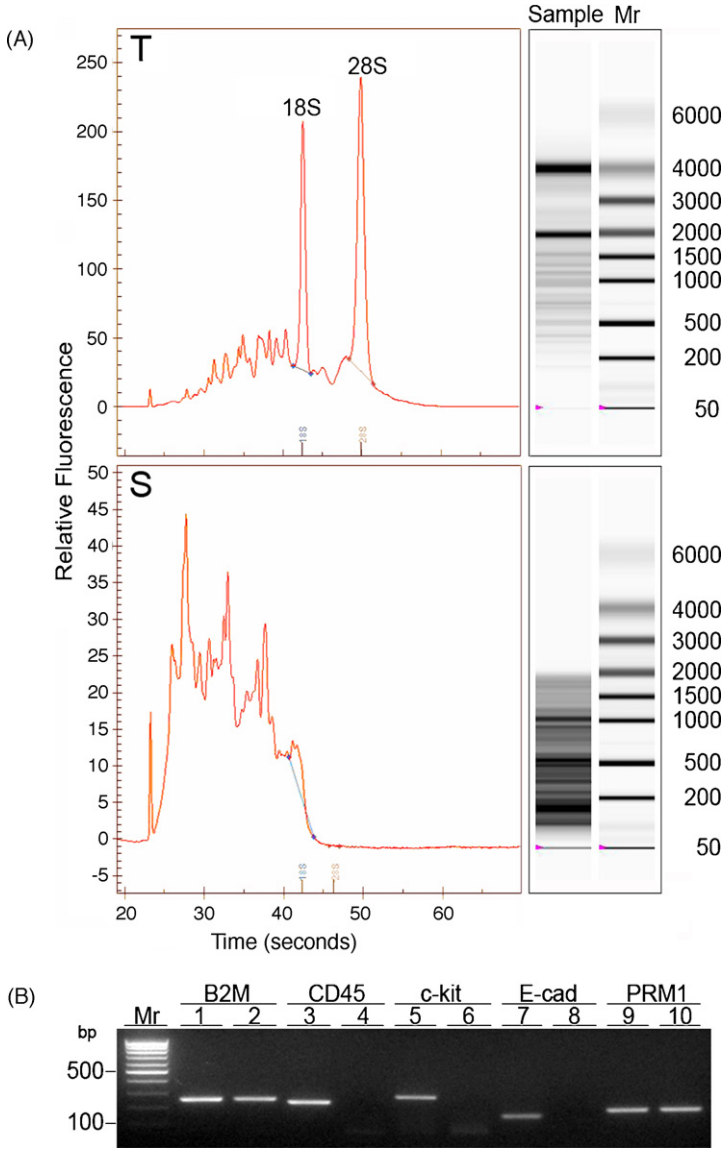

Fig. 1. Detection of domestic swine spermatozoal mRNA transcripts. (A) Total RNA isolated from testis and spermatozoa was subjected to automated electrophoresis to assess transcript integrity. Samples isolated from the testis (T) had two separated and sharp peaks, 18S and 28S rRNA, on a relatively flat baseline indicating intact RNAs. In contrast, the total RNA from spermatozoa (S) was not only free of distinct peaks of 18S and 28S rRNA, but also had highly abundant of multiple peaks at low molecular weights. Data was then transformed to create a virtual gel, exhibited at the right of the electrophoretogram; the multiple bands evident in the size distributed toward the smaller fragments in the RNA sample derived from spermatozoa. (B) To detect contamination, total RNA was extracted from freshly ejaculated semen (1, 3, 5, 7, 9) and washed spermatozoa (2, 4, 6, 8, 10) and subjected to one-step RT-PCR with gene primers, beta-2-microglobulin protein (B2M), CD45, c-kit, *E*-cadherin (*E*-cad) and protamine P1 (PRM1), for specific molecular markers. There were no products shown of unexpected length for all reactions. Mr: DNA ladder.

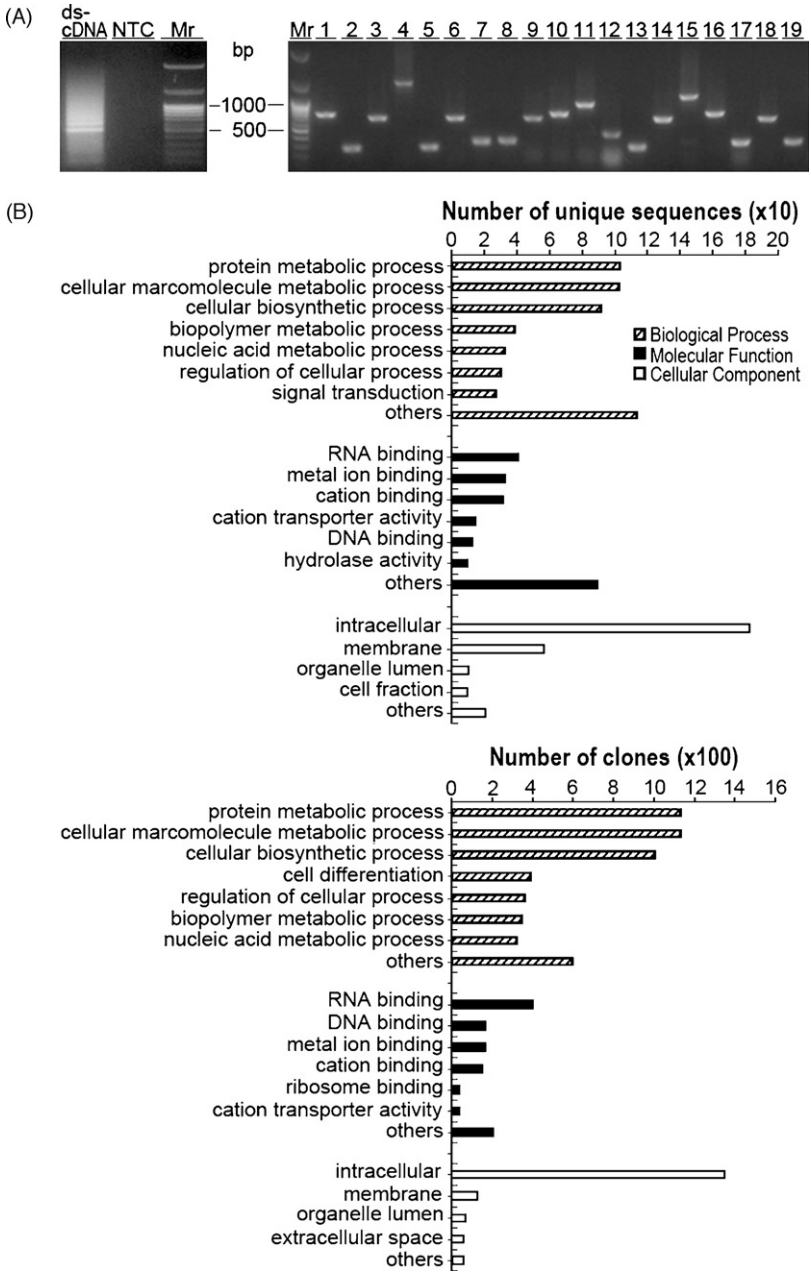

Fig. 2. Characterization of a domestic swine spermatozoal expressed sequence tag library. (A) The quality of the cDNA library was indicated by strong smear with several bright bands after amplification of a population of double-stranded cDNA fragments (left). Following gel extraction, the purified cDNA fragments were ligated into chosen vectors. The different lengths of DNA fragments were inserted (right). ds: double strand; NTC: non-template control; Mr: DNA ladder. (B) Functional annotation of swine ejaculated spermatozoal RNAs by gene ontology. The number of unique sequences or clusters of the bar chart are represented on the *X*-axis. The three categories (separated by spaces), biological processes, molecular functions and cellular components, are represented on the *Y*-axis.

Table 2
List of the BLAST search for contigs composed of the 22 most redundant expressed sequence tag clones, ≥20 clones each, in the non-normalized swine ejaculated spermatozoal cDNA library

Accession number	Species	Homologous protein	Query (bp)	*E*-value	Number of EST
NM_214253	Pig	*Sus scrofa* protamine P1	469	1.00E−146	1884
XM_583699	Cattle	*Bos taurus* CE5 protein-like	604	7.00E−53	416
X16170	Pig	*Sus scrofa* transition protein 1	420	1.00E−159	136
NM_214211	Pig	*Sus scrofa* ubiquitin/ribosomal fusion protein (UBA52)	537	0.0	101
NM_001044543	Pig	*Sus scrofa* 60S ribosomal protein L10	756	0.0	53
AF486866	Pig	*Sus scrofa* breed Landrace mitochondrion, complete genome	760	0.0	51
NM_001001633	Pig	*Sus scrofa* 40S ribosomal protein S29	323	1.00E−122	51
NM_000997	Human	*Homo sapiens* ribosomal protein L37	379	1.00E−144	42
NM_001001587	Pig	*Sus scrofa* 40S ribosomal protein S28	324	1.00E−179	36
NM_001040581	Cattle	*Sus scrofa* 40S ribosomal protein S21	360	1.00E−102	34
NM_214373	Pig	*Sus scrofa* translationally controlled tumor protein	850	0.0	33
NM_001001634	Pig	*Sus scrofa* 40S ribosomal protein S17	481	1.00E−120	30
NM_001000	Human	*Homo sapiens* ribosomal protein L39	396	1.00E−60	29
XM_001255197	Cattle	*Bos taurus* hypothetical protein LOC788003	279	1.00E−38	28
BC108179	Cattle	*Bos taurus* similar to 60S ribosomal protein L38	464	1.00E−111	28
AL137366	Human	*Homo sapiens* mRNA; cDNA DKFZp434F1626	381	1.00E−70	26
NM_001037490	Cattle	*Bos taurus* zinc finger protein 474	507	1.00E−164	26
NM_001022	Human	*Homo sapiens* ribosomal protein S19	522	0.0	24
NM_001077056	Cattle	*Bos taurus* hypothetical protein MGC133634	574	5.30E−93	22
NM_000995	Human	*Homo sapiens* ribosomal protein L34	437	1.00E−153	22
NM_001034284	Cattle	*Bos taurus* similar to chromosome 14 open reading frame 8	800	0.0	21
NM_001040577	Cattle	*Bos taurus* similar to heat shock protein, alpha-crystalline-related, B9	552	1.00E−131	20

The two major rRNAs, essential for 80S cytoplasmic ribosome assembly within eukaryotic cells, were clearly absent in the spermatozoon in our study, as previously observed in other species (Ostermeier et al., 2002; Lambard et al., 2004; Grunewald et al., 2005; Gilbert et al., 2007). There is no clear indication of factors which cause the absence of rRNA in the mature spermatozoon. The GC-rich transcriptional domains of ribosomal DNA are generally unmethylated in normal cells and associated with active transcription of 18S, 5.8S, and 28S rRNA genes. However, if methylation occurred in the promoter region of ribosomal genes, the 5′-external transcribed spacer or the rRNA gene region, it would be associated with unexpressed rRNA (Stancheva et al., 1997; Santoro and Grummt, 2001). Adult rats have been used to demonstrate highest level of ribosomal DNA methylation in non-growing spermatozoa from the epididymal and vas deferens compared to somatic cells or to various germ cell types (Kunnath and Locker, 1982). Therefore, it is reasonable to suspect complete absence of the rRNA in ejaculated spermatozoa.

The assumption that spermatozoa are translationally silent until fertilization occurs is widely accepted. However, it was noted that within the human, mouse, bovine and rat spermatozoon, [^{35}S]methionine–[^{35}S]cysteine-labeled amino acid can be incorporated into newly synthesized polypeptide during the capacitation phase (Gur and Breitbart, 2006), indicating the stored mRNAs within the ejaculated spermatozoon can be translated prior to fertilization. In this report, scientists identified the presence of 55S mitochondrial ribosome in polysomal fractions and the mitochondrial translation machinery can direct the translation of nuclear-encoded gene in spermatozoa. It is still controversial to state that translational activity takes place in ejaculated spermatozoa of mammals, and the studies which agree with such notion all are confined to the experiments on spermatozoal mitochondria (Alcivar et al., 1989; Gur and Breitbart, 2006). Nonetheless, there were abundant and various types

of ribosomal protein mRNAs in the swine spermatozoal EST library based on the results of sequencing analysis. These ribosomal protein mRNAs within spermatozoa may reflect transcriptional and translational quiescence during spermiogenesis or may be involved in some unknown extra-ribosomal functions during fertilization. As a consequence, minimal amounts of RNA are present in mature spermatozoa and lose the apparent ability to be transcribed and/or translated; potential causes include chromatin structure condensation, cytoplasmic elimination and inefficient assembly of ribosomes.

4.2. The function of spermatozoal mRNA transcripts related to early embryo development

The existence of remnant spermatozoal mRNA transcripts within the mature ejaculated spermatozoon was confirmed, and their putative functions including functional and structural maintenance for spermatogenesis, spermiogenesis and the mature spermatozoon were observed in our analysis of gene ontology. While evidences are provided for biological functions of mostly genes in various types of male germ cells (Kramer and Krawetz, 1997; Gilbert et al., 2007), the role and the mechanism of action for mRNAs within the mature spermatozoon during fertilization and embryonic development remain to be determined. Hayashi et al. (2003) showed some of the spermatozoal mRNA transcripts did not have effect on the embryogenesis and were degraded rapidly in embryos at late 1-cell stage. However, some of these spermatozoal mRNA transcripts were found survived up to 4-cell period. Therefore, the spermatozoon delivers not only genetic materials but also spermatozoal RNAs into the oocyte during the process of fertilization. Some of the unique sequences in our EST library were identified as putative homologs of human (*H. sapiens*) genes which were expressed during embryogenesis; such as heat shock 70 kDa protein 2 (Hsp70.2), sperm-specific antigen 2 (SSFA2), Sestrin 1 (SESN1).

Hsp70.2 gene, also known as heat shock protein 2 (HspA2) gene, is a testis-specific member of the Hsp70 family and its expression is restricted mostly to meiotic and post-meiotic male germ cells in adult mice and rats (Rupik et al., 2006), but its function is unclear in spermiogenesis (Govin et al., 2006). The immature human spermatozoa which failed to express HspA2, showed cytoplasmic retention and lacked of zona pellucida binding ability (Huszar et al., 2000). Rupik et al. (2006) reported that protein of enhanced green fluorescent protein (EGFP) reporter gene driven by the promoter of rat Hsp70.2 were visualized in nervous system of transgenic mice during their embryonic development. In contrast, the Hsp70.2 gene knockout female mice underwent normal oocyte meiosis and were fertile (Dix et al., 1996), indicating that either Hsp70.2 protein was not absolutely needed for embryonic development or other HSP70 proteins could substitute for one another during embryogenesis. Hence, specific functions of the heat shock protein (HSP) gene family during embryogenesis still are not fully understood.

The spermatozoal surface antigen, sperm-specific antigen 2, was proven that involved in early cleavage of the fertilized oocyte (Ohl and Menge, 1996). Nevertheless, antibodies against SSFA2 protein were found to inhibit early cleavage of the oocyte without affecting pronuclear formation (Naz, 1992). As SSFA2 mRNAs entered the oocyte and were translated, it might provide an extranuclear signal to influence the cleavage program of the fertilized oocyte.

Sestrin 1, also known as PA26, has been initially identified as a target for the p53 tumor suppressor in adult tissues and as a potential regulator of cellular growth (Velasco-Miguel et al., 1999). Peeters et al. (2006) further demonstrated SESN1 was required for normal embryonic left–right determination. The SESN1-knockdown zebrafish embryos showed lateral disturbances both in heart and in gut providing clear evidence for the requirement of SESN1 in normal left–right asymmetric development. The SESN1 mRNA also could be detected in mouse-preimplanted embryos (Ko et al., 2000). However, it is indeterminate that these paternal mRNAs of SESN1 will be preserved and translated during embryonic development.

We do not know yet whether mRNAs retained in the spermatozoon will all acquire a function in early embryogenesis. If these spermatozoal mRNA transcripts were translated improperly within the gamete, they might likely interfere with the early embryonic development.

4.3. Variation in spermatozoal protamine transcripts

If the residual RNAs present in spermatozoa cannot be translated, the function remains unclear. The presence of spermatozoal basic nucleoprotein mRNAs, protamines and transition proteins, prob-

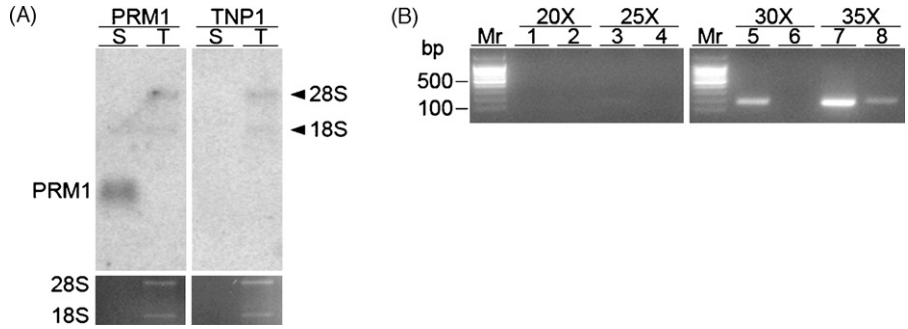

Fig. 3. The relative expression levels of remaining protamine mRNAs in domestic swine ejaculated spermatozoa. (A) Northern blot analysis for protamine P1 (PRM1) and transition protein 1 (TNP1) using total RNA, 500 ng, isolated from the swine testis (T) or ejaculated spermatozoa (S) from a sexually mature swine. The RNA was fractionated, transferred and hybridized with the ^{32}P-labeled PRM1 or TNP1 cDNA probes. The 28S and 18S rRNAs were shown as molecular weight markers. (B) Amplification of protamine P2 (PRM2) transcripts for various numbers of thermal cycles (X) using 20 ng of RNA isolated form swine ejaculated spermatozoa. The gene-specific primers, PRM1 (1, 3, 5, 7) and PRM2 (2, 4, 6, 8) were used with the one-step RT-PCR system. As the image indicated, the amplified products were observed at the 25th cycle for PRM1 and the amplicons of PRM2 were not shown until completion of the 35th cycle. Mr: DNA ladder.

ably represent the remnants of previous transcriptional products. The fluctuation of expression level or mutation in protamine genes is associated with infertility (Lee et al., 1995; Iguchi et al., 2006).

4.3.1. Protamine P1 (PRM1)

Spermatozoa of mice, humans, chimpanzees and Japanese macaques have two types of protamine, whereas the spermatozoa of bovine, swine, rats, and ovine have only PRM1 type (Lee and Cho, 1999; Yu and Takenaka, 2004). The existence of protamine mRNA in mature spermatozoa was confirmed (Lambard et al., 2004). In our study, 41% of the sequenced clones were the PRM1 gene. To confirm the high expression level of the PRM1 mRNA in swine spermatozoa, Northern blot analysis was used with only 500 ng of total RNA as starting material (Fig. 3A). It seems reasonable to assume that the protamine residue remains at a high level in spermatozoa, because each round spermatid contains 20,000 PRM1 mRNA copies on average (Braun et al., 1989) and 10–100 spermatozoa are theoretically sufficient to provide positive amplification results with semi-nested PCR (Bauer and Patzelt, 2003). Despite the presence of a large number of proamine mRNAs, microinjection of round spermatids into blastocysts produced a rapid degradation of PRM1 mRNAs and the mRNAs were undetectable at the late 2-cell stage (Hayashi et al., 2003). This result suggests that these ectopic mRNAs are not normally expressed during embryogenesis. The importance of PRM1 may be related to male infertility, but we should not rule out the possibility that these PRM1 mRNA residues in spermatozoa indirectly may influence early embryo development.

4.3.2. Protamine P2 (PRM2)

The two PRM2 EST clones in our spermatozoon library were reconfirmed (Fig. 3B), proving that the transcript for PRM2 was present in swine mature spermatozoa. Previously, the mRNA of the PRM2 gene was only found in spermatids of swine and bovine and in a very low quantity (Maier et al., 1990). The upstream elements of the PRM2 gene are needed for maximal transcription but did not explain the divergent transcription efficiencies in mouse, rat, and bovine (Kremling et al., 1992) The exon–intron structure also did not reveal any differences between species. However, mutations in the 5′ and/or 3′ control regions of the PRM2 gene in swine and bovine lead to depression of both transcription and post-transcriptional polyadenylation (Maier et al., 1990). Domenjoud et al. (1991) suggests that mammals are endowed with at least two types of protamine gene which are both transcribed but are translationally regulated in a species-specific manner. Moreover, the PRM2 mRNA can be retained in spermatozoa of some species, even after loss of all of the cytoplasm, except for a thin layer sandwiched

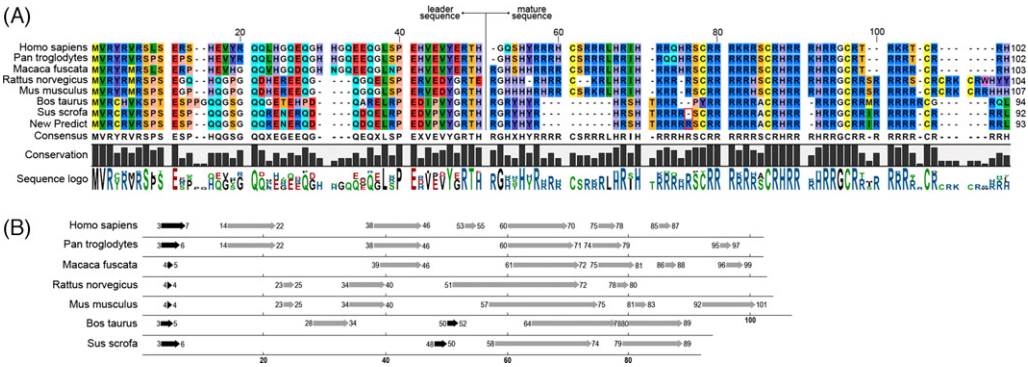

Fig. 4. Sequence analysis of seven mammalian protamine P2 (PRM2) proteins. (A) Multiple alignments of the predicted amino acid sequences; logo displayed a significant conservation among *Homo sapiens* (BC042671), *Pan troglodytes* (NM_001009084), *Macaca fuscata* (AB101300), *Rattus norvegicus* (NM_012873), *Mus musculus* (BC049612), *Bos taurus* (BC109783) and *Sus scrofa* (NM_214252). Various colors of shaded amino acid residues in all species represented the RasMol color. Red rectangle designates the differences between the published and new predicted amino acids of the swine PRM2. The parameter of consensus sequence set up limits in "majority", and ambiguous residues were represented with "X" letters. Logo color was used for distinguishing the "polarity" of amino acids. (B) Secondary structure was predicted based on the helix-/sheet-forming tendency, which indicated the patterns of residue conservation. Alpha-helix and beta-strand locus were indicated by gray and black arrows, respectively. Both analyses programs were performed using CLC Protein Workbench software, version 3.0.2.

between the nuclear and plasma membranes (Debarle et al., 1995; Lambard et al., 2004). Therefore, it is hard to prove that PRM2 mRNA is completely absent in swine spermatozoa.

The transcript sequence of the open reading frame of our swine PRM2 was a close homologue of the PRM2 in NM_214252. However, our sequence had three extra bases, CGG, at the 180th position from the start codon. A similar observation was made for other PRM2 entries in the NCBI database, e.g., CX064242, CX064047, CX063808, CX063283, and CV873251. Multiple alignment analysis of the predicted amino acid sequence (Fig. 4A) indicated these extra bases did not disarrange the original peptide sequence of swine PRM2 retaining the general characteristics of the PRM2 precursor. The consensus peptide sequence was 93 amino acids rather than 92 amino acids as stated in previous studies (Maier et al., 1990; Keime et al., 1992).

The differential expression of the PRM2 gene or synthesis of a functionally deficient product may also reflect translational regulation and subsequently result in the absence of mature PRM2. In the rat, as the swine and bovine, there is only a single type of protamine; the precursor of PRM2 could be detected in mature spermatozoa, but the mature PRM2 protein is absent (Stanker et al., 1992). Maier et al. (1990) suggests that lack of the PRM2 protein in swine spermatozoa is due to mutations or deletions within the PRM2 gene rather than quantitative differences in gene expression, and the missing motif, RLHRIH, could be essential for the presence of mature PRM2 in spermatozoa. Secondary structure predictions indicate that the beta strand-structure appears twice in bovine and swine sequences compared to other species (Fig. 4B). This extra beta strand-structure is near the missing RLHRIH motif and may lead to a change of conformation, thus, influencing the original function of PRM2. Therefore, the PRM2 variants could have modified binding sites to completely block translation or to decrease the mRNA stability (Belokopytova et al., 1993).

Finally, the zinc ion, a critical mineral for male reproductive functions and the condensation of chromosomes, is working mainly via the histidine residues of PRM2 (Bertelsmann et al., 2007). However, the numbers of histidine residues in PRM2 of swine and bovine are only about 50% compared to other species, while the number of zinc ions in the amembranous spermatozoal nucleus is similar to that of PRM2 molecules (Bench et al., 2000). Thus, the principal role of PRM2 in spermiogenesis of the swine is not quite as simple as participating in the condensation of spermatozoal chromatin and, therefore, it appears reasonable to suggest that its function should be the subject of much more intensive investigation.

5. Conclusion

Isolated RNA from ejaculated spermatozoa contained a number of stably retained transcript residues. The spermatozoal cDNA library EST sequences indicated many unique EST-encoded proteins with unknown physiological functions. The EST sequences also allowed comparisons with other mammalian species. The sequences will be useful in detecting single nucleotide polymorphisms, and other transcriptional variants associated with spermatogenesis, spermiogenesis, as well as fertilization. The non-redundant ESTs generated from this library allow construction of a cDNA microarray and an oligonucleotide array for gene expression analysis and to physically map genes for assessment of their role in regulation of spermatozoal maturation or to predict factors causing male infertility. Taken together, these resources can be applied to develop tools for functional genomic studies of the male reproductive system.

Acknowledgements

This research was conducted using funds provided by grant NSC93-2313-B-002-101 awarded by the National Science Council of Taiwan, ROC. We would like to express our sincere appreciation for this essential support.

References

Alcivar, A.A., Hake, L.E., Millette, C.F., Trasler, J.M., Hecht, N.B., 1989. Mitochondrial gene expression in male germ cells of the mouse. Dev. Biol. 135, 263–271.
Al-Shahrour, F., Herrero, J., Mateos, Á., Santoyo, J., Díaz-Uriarte, R., Dopazo, J., 2003. Using gene ontology on genome-scale studies to find significant associations of biologically relevant terms to group of genes. In: Neural Networks for Signal Processing, vol. XIII. IEEE Press, New York, pp. 43–52.
Al-Shahrour, F., Minguez, P., Vaquerizas, J.M., Conde, L., Dopazo, J., 2005. Babelomics: a suite of web-tools for functional annotation and analysis of group of genes in high-throughput experiments. Nucleic Acids Res. 33, W460–W464.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Bauer, M., Patzelt, D., 2003. Protamine mRNA as molecular marker for spermatozoa in semen stains. Int. J. Legal. Med. 117, 175–179.
Belokopytova, I.A., Kostyleva, E.I., Tomilin, A.N., Vorob'ev, V.I., 1993. Human male infertility may be due to a decrease of the protamine P2 content in sperm chromatin. Mol. Reprod. Dev. 34, 53–57.
Bench, G., Corzett, M.H., Kramer, C.E., Grant, P.G., Balhorn, R., 2000. Zinc is sufficiently abundant within mammalian sperm nuclei to bind stoichiometrically with protamine 2. Mol. Reprod. Dev. 56, 512–519.
Bertelsmann, H., Sieme, H., Behne, D., Kyriakopoulos, A., 2007. Is the distribution of selenium and zinc in the sublocations of spermatozoa regulated? Ann. N.Y. Acad. Sci. 1095, 204–208.
Braun, R.E., Behringer, R.R., Peschon, J.J., Brinster, R.L., Palmiter, R.D., 1989. Genetically haploid spermatids are phenotypically diploid. Nature 337, 373–376.
Cheng, W.T.K., 1985. In vitro fertilization in farm oocytes. Ph.D. Thesis. Cambridge University, U.K.
Debarle, M., Martinage, A., Sautiere, P., Chevaillier, P., 1995. Persistence of protamine precursors in mature sperm nuclei of the mouse. Mol. Reprod. Dev. 40, 84–90.
Dix, D.J., Allen, J.W., Collins, B.W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E.H., Eddy, E.M., 1996. Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. Proc. Natl. Acad. Sci. U.S.A. 93, 3264–3268.
Domenjoud, L., Kremling, H., Burfeind, P., Maier, W.M., Engel, W., 1991. On the expression of protamine genes in the testis of man and other mammals. Andrologia 23, 333–337.
Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using Phred I accuracy assessment. Genome Res. 8, 175–185.
Gilbert, I., Bissonnette, N., Boissonneault, G., Vallée, M., Robert, C., 2007. A molecular analysis of the population of mRNA in bovine spermatozoa. Reproduction 133, 1073–1086.
Gordon, D., Abajian, C., Green, P., 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8, 195–202.
Govin, J., Caron, C., Escoffier, E., Ferro, M., Kuhn, L., Rousseaux, S., Eddy, E.M., Garin, J., Khochbin, S., 2006. Post-meiotic shifts in HSPA2/HSP70.2 chaperone activity during mouse spermatogenesis. J. Biol. Chem. 281, 37888–37892.
Grunewald, S., Paasch, U., Glander, H.J., Anderegg, U., 2005. Mature human spermatozoa do not transcribe novel RNA. Andrologia 37, 69–71.
Gur, Y., Breitbart, H., 2006. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. Genes Dev. 20, 411–416.
Hayashi, S., Yang, J., Christenson, L., Yanagimachi, T., Hecht, N.B., 2003. Mouse preimplantation embryos developed from oocytes injected with round spermatids or spermatozoa have similar but distinct patterns of early messenger RNA expression. Biol. Reprod. 69, 1170–1176.
Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.
Huszar, G., Stone, K., Dix, D., Vigue, L., 2000. Putative creatine kinase M-isoform in human sperm is identified as the 70-kilodalton heat shock protein HspA2. Biol. Reprod. 63, 925–932.

Iguchi, N., Yang, S., Lamb, D.J., Hecht, N.B., 2006. An SNP in protamine 1: a possible genetic cause of male infertility? J. Med. Genet. 43, 382–384.
Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A., Auffray, C., 2005. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. Nucleic Acids Res. 33, e56.
Keime, S., Heitland, K., Kumm, S., Schlosser, M., Hroch, N., Holtz, W., Engel, W., 1992. Characterization of four genes encoding basic proteins of the porcine spermatid nucleus and close linkage of three of them. Biol. Chem. Hoppe Seyler 373, 261–270.
Ko, M.S., Kitchen, J.R., Wang, X., Threat, T.A., Wang, X., Hasegawa, A., Sun, T., Grahovac, M.J., Kargul, G.J., Lim, M.K., Cui, Y., Sano, Y., Tanaka, T., Liang, Y., Mason, S., Paonessa, P.D., Sauls, A.D., DePalma, G.E., Sharara, R., Rowe, L.B., Eppig, J., Morrell, C., Doi, H., 2000. Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development 127, 1737–1749.
Kramer, J.A., Krawetz, S.A., 1997. RNA in spermatozoa: implications for the alternative haploid genome. Mol. Hum. Reprod. 3, 473–478.
Kremling, H., Reinhart, N., Schlosser, M., Engel, W., 1992. The bovine protamine 2 gene: evidence for alternative splicing. Biochim. Biophys. Acta 1132, 133–139.
Kunnath, L., Locker, J., 1982. Variable methylation of the ribosomal RNA genes of the rat. Nucleic Acids Res. 10, 3877–3892.
Lambard, S., Galeraud-Denis, I., Martin, G., Levy, R., Chocat, A., Carreau, S., 2004. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. Mol. Hum. Reprod. 10, 535–541.
Lee, C.H., Cho, Y.H., 1999. Aspects of mammalian spermatogenesis: electrophoretical analysis of protamines in mammalian species. Mol. Cells 9, 556–559.
Lee, K., Haugen, H.S., Clegg, C.H., Braun, R.E., 1995. Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. Proc. Natl. Acad. Sci. U.S.A. 92, 12451–12455.
Maier, W.M., Nussbaum, G., Domenjoud, L., Klemm, U., Engel, W., 1990. The lack of protamine 2 (P2) in boar and bull spermatozoa is due to mutations within the P2 gene. Nucleic Acids Res. 18, 1249–1254.
Martins, R.P., Krawetz, S.A., 2005. RNA in human sperm. Asian J. Androl. 7, 115–120.
Miller, D., Briggs, D., Snowden, H., Hamlington, J., Rollinson, S., Lilford, R., Krawetz, S.K., 1999. A complex population of RNAs exists in human ejaculate spermatozoa: implications for understanding molecular aspects of spermiogenesis. Gene 237, 385–392.
Naz, R.K., 1992. Effects of antisperm antibodies on early cleavage of fertilized ova. Biol. Reprod. 46, 130–139.
Ohl, D.A., Menge, A.C., 1996. Assessment of sperm function and clinical aspects of impaired sperm function. Front. Biosci. 1, e96–108.
Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., Krawetz, S.A., 2002. Spermatozoal RNA profiles of normal fertile men. Lancet 360, 772–777.
Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., Krawetz, S.A., 2004. Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature 429, 154.
Peeters, H., Voz, M.L., Verschueren, K., De Cat, B., Pendeville, H., Thienpont, B., Schellens, A., Belmont, J.W., David, G., Van De Ven, W.J., Fryns, J.P., Gewillig, M., Huylebroeck, D., Peers, B., Devriendt, K., 2006. Sesn1 is a novel gene for left–right asymmetry and mediating nodal signaling. Hum. Mol. Genet. 15, 3369–3377.
Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652.
Pessot, C.A., Brito, M., Figueroa, J., Concha, I.I., Yanez, A., Burzio, L.O., 1989. Presence of RNA in the sperm nucleus. Biochem. Biophys. Res. Commun. 158, 272–278.
Rupik, W., Stawierej, A., Stolarczyk, I., Widłak, W., 2006. Promoter of the heat shock testis-specific Hsp70.2/Hst70 gene is active in nervous system during embryonic development of mice. Anat. Embryol. 211, 631–638.
Santoro, R., Grummt, I., 2001. Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. Mol. Cell 8, 719–725.
Stancheva, I., Lucchini, R., Koller, T., Sogo, J.M., 1997. Chromatin structure and methylation of rat rRNA genes studied by formaldehyde fixation and psoralen cross-linking. Nucleic Acids Res. 25, 1733–1741.
Stanker, L.H., McKeown, C., Balhorn, R., Lee, C., Mazrimas, J., Goralka, M., Wyrobek, A., 1992. Immunological evidence for a P2 protamine precursor in mature rat sperm. Mol. Reprod. Dev. 33, 481–488.
Velasco-Miguel, S., Buckbinder, L., Jean, P., Gelbert, L., Talbott, R., Laidlaw, J., Seizinger, B., Kley, N., 1999. PA26, a novel target of the p53 tumor suppressor and member of the GADD family of DNA damage and growth arrest inducible genes. Oncogene 18, 127–137.
Wykes, S.M., Visscher, D.W., Krawetz, S.A., 1997. Haploid transcripts persist in mature human spermatozoa. Mol. Hum. Reprod. 3, 15–19.
Xia, J., Radford, C., Guo, X., Magor, K.E., 2007. Immune gene discovery by expressed sequence tag analysis of spleen in the duck (*Anas platyrhynchos*). Dev. Comp. Immunol. 31, 272–285.
Yang, C.C., Chang, H.S., Lin, C.J., Hsu, C.C., Cheung, J.I., Hwu, L., Cheng, W.T.K., 2004. Cock spermatozoa serve as the gene vector for generation of transgenic chicken (*Gallus gallus*). Asian Australas. J. Anim. Sci. 17, 885–891.
Yang, K.T., Lin, C.Y., Liou, J.S., Fan, Y.H., Chiou, S.H., Huang, C.W., Wu, C.P., Lin, E.C., Chen, C.F., Lee, Y.P., Lee, W.C., Ding, S.T., Cheng, W.T.K., Huang, M.C., 2007. Differentially expressed transcripts in shell glands from low and high egg production strains of chickens using cDNA microarrays. Anim. Reprod. Sci. 101, 113–124.
Yu, S.S., Takenaka, O., 2004. Molecular cloning of protamine-2 and expression with aging in Japanese monkey (*Macaca fuscata*). Primates 45, 147–150.
Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7, 203–214.
Zhao, Y., Li, Q., Yao, C., Wang, Z., Zhou, Y., Wang, Y., Liu, L., Wang, Y., Wang, L., Qiao, Z., 2006. Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. Hum. Reprod. 21, 1583–1590.