RBMO

ARTICLE

Investigation of the mechanisms leading to human sperm DNA damage based on transcriptome analysis by RNA-seq techniques

BIOGRAPHY

Jin-Chun Lu, PhD, is a Chief Technician and Laboratory Director of the Reproductive Medicine Center of Zhongda Hospital Affiliated to Southeast University, China, and Chairman of the Reproductive Testing Branch of the Chinese Sexual Society. He has published more than 140 academic papers as the first or corresponding author and nine books as editor-in-chief.

Chun-Hui Zhu[1,2], Ye Wei[2], Sheng-Min Zhang[1], Fang Chen[1], Feng Li[1], Nai-Jun Dong[1], Tong-Min Xue[1], Kai-Feng Liu[1], Jin-Chun Lu[3,*], Heng-Mi Cui[2,*]

KEY MESSAGE

Differentially expressed RNAs related to sperm DNA damage and repair are identified by RNA-seq technology, which provides new insights into our understanding of sperm DNA damage and repair, and will help to discover new biomarkers related to sperm DNA damage.

ABSTRACT

Research question: What are the molecular mechanisms leading to human sperm DNA damage?

Design: Semen samples were collected and the sperm DNA fragmentation index (DFI) was assessed. Differentially expressed RNA in spermatozoa with a high (DFI ≥30%, experimental group) or normal (DFI <30%, control group) DFI were identified by RNA-sequencing (RNA-seq) technology, and Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed. Three differentially expressed RNA related to sperm DNA damage and repair, namely *PMS1*, *TP53BP1* and *TLK2*, were validated using real-time quantitative (RT-qPCR).

Results: A total of 19,970 expressed RNA were detected in the two groups. Compared with the control group, the expression levels of 189 RNA in the experimental group were significantly increased and those of 163 genes decreased. Gene Ontology enrichment analysis showed that these RNA were mainly concentrated in the ATPase-dependent transmembrane transport complex, extracellular exosome, somatic cell DNA recombination, protein binding, cytoplasm and regulation of localization. KEGG pathway analysis showed that these RNA were mainly related to the PI3K-Akt signalling pathway, endocytosis, p53 signalling pathway and cGMP-PKG signalling pathway. The RT-qPCR results showed that the expression levels of *PMS1*, *TP53BP1* and *TLK2* in the experimental group were significantly lower than in the control group (P = 0.01, 0.015 and 0.004, respectively), which was identical to the results of RNA sequencing.

Conclusions: Differentially expressed RNA related to sperm DNA damage and repair may be identified by RNA-seq technology, which provides new insights into the understanding of sperm DNA damage and repair, and will help to discover new biomarkers related to sperm DNA damage.

[1] Center for Reproductive Medicine, Northern Jiangsu People's Hospital Affiliated to Yangzhou University/Clinical Medical College, Yangzhou University, Yangzhou Jiangsu, China
[2] Institute of Epigenetics and Epigenomics, College of Animal Science and Technology, Yangzhou University, Yangzhou Jiangsu, China
[3] Center for Reproductive Medicine, Zhongda Hospital, Southeast University, Nanjing Jiangsu, China

*Corresponding authors. E-mail addresses: 406646227@qq.com (J.-C. Lu), hmcui@yzu.edu.cn (H.-M. Cui). https://doi.org/10.1016/j.rbmo.2022.08.108 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

KEYWORDS

Male infertility
mRNA
RNA-seq
Sperm DNA damage
Transcriptome

INTRODUCTION

Infertility affects 15% of couples of reproductive age, and 50% of the total number of cases are attributed to male factors (*Agarwal et al., 2015*). Abnormal expression of any functional gene in the process of spermatogenesis and maturation may affect sperm morphology, structure or function, and then induce male infertility (*Jodar et al., 2015*). Revealing the regulatory mechanisms of spermatogenesis and mining important functional genes of spermatozoa based on omics has become the focus of research in male reproductive disorders (*Zarezadeh et al., 2021*). Transcriptomes are the link between genomes and proteomes; these can reveal the spatiotemporal dynamics of gene expression, and are used to compare differences in the expression of RNA, including mRNA, long non-coding RNA (lncRNA), microRNA (miRNA) etc., in specific organs, tissues or cells of a specific species under different environmental conditions (*Dal Molin et al., 2017*).

It was previously thought that active transcription was not evident in mammalian mature spermatozoa. Mature spermatozoa contain a small amount of complex and highly fragmented RNA, which has been assumed to be functional debris reflecting events occurring during the process of spermatogenesis (*Gòdia et al., 2018*). In addition, recent research has shown that sperm RNA may play an important role in regulating spermatogenesis (*Sharma et al., 2016*), sperm function (*Rajesh et al., 2017*), fertilization (*Fang et al., 2014*; *Sendler et al., 2013*), early embryo development (*Boerke et al., 2007*) and even the offspring's phenotype (*Chen et al., 2016*). However, the analysis of sperm RNA is challenging because spermatozoa contain very small amounts of highly fragmented RNA and ejaculated semen contains some non-sperm cells such as leukocytes and spermatogenic cells, which include large amounts of non-fragmented RNA.

The methods of density gradient centrifugation (DGC) and swim-up removing non-sperm cells can be used to study sperm RNA (*Shafeeque et al., 2014*). Lian and colleagues (*Lian et al., 2021*) investigated the effect of the purification operation of DGC on porcine sperm RNA, and found that a clustering and principal component analysis showed a high degree of similarity between purified and unpurified semen, demonstrating that the removal of large amounts of somatic cells in the study of sperm transcriptome was feasible. Li and colleagues (*Li et al., 2021*) analysed bovine sperm RNA in the deformation stage of spermatids by RNA-sequencing (RNA-seq) technology, and found that up-regulated RNA in the process of sperm deformation gradually decreased, while down-regulated RNA gradually increased, indicating that although chromatin pyknosis and transcription were gradually reduced, they were not completely terminated. Gòdia and co-workers (*Gòdia et al., 2020a*, *2020b*) found that pig spermatozoa biological traits were significantly correlated with the mRNA abundances of 6128 genes by studying pig models, and that the bacterial flora of boar semen were related to sperm quality traits, suggesting that these potential pathogens and their putative antibiotic resistance genes might affect boar reproductive performance. Gòdia and collaborators (*Gòdia et al., 2019*) also studied the effects of seasonal changes (summer and winter) on the boar sperm transcriptome.

Swanson and co-workers (*Swanson et al., 2020*) found that human sperm RNA sequencing could be used to identify the presence of contaminated or potentially pathogenic bacteria by comparing sequencing data that did not map to the human genome with bacterial, viral and archaeal genomes. Based on a whole-transcriptome study of stallion spermatozoa, Das and colleagues (*Das et al., 2013*) reported that a total of 19,257 sequence tags were mapped to equine chromosomal and mitochondrial genomes, and that some mRNA and microRNA might be crucial for early embryonic development; this suggested that the abundant coding and non-coding RNA in stallion spermatozoa are not random remnants of spermatogenesis in testis but a selectively retained and functionally coherent collection of RNA.

Spermatozoa contain a wide repertoire of coding and non-coding RNA, including long and short non-coding RNA (*Krawetz et al., 2011*), which may have an important role in sperm fitness. Using RNA-seq technology to study the sperm transcriptome can obtain data on sperm quality and fertility, and sperm RNA may be used as a potential biomarker of sperm quality (*Ablondi et al., 2021*; *Bissonnette et al., 2009*) and fertility (*Capra et al., 2017*). At present, the biological function of sperm RNA in the development of sperm DNA fragmentation is still unknown, and the molecular basis of selecting spermatozoa with a higher quality to improve fertility in human and animal breeding needs to be further explored.

Sperm DNA integrity is crucial for fertilization and the development of healthy offspring, and more and more studies have confirmed that sperm DNA damage is related to male infertility (*Agarwal et al., 2020*). The sperm DNA fragmentation index (DFI) can reflect the integrity of sperm DNA and is used as an important indicator to evaluate semen quality (*Malic Voncina et al., 2021*; *Yan et al., 2022*). The current study performed a bioinformatics analysis on the transcriptome data of human spermatozoa with high or normal DNA fragmentation, and screened out differentially expressed RNA. Using cluster analysis, Gene Ontology analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, the core regulatory network of sperm DNA fragmentation at the transcriptome level was obtained. The goal of this study was to obtain detailed information on RNA in spermatozoa with high DNA fragmentation, and then improve the understanding of the biological significance of sperm RNA and find potential biomarkers for evaluating sperm DNA fragmentation.

MATERIALS AND METHODS

Collection of semen samples

A total of 24 semen samples were collected from clinical patients by masturbation after 2–7 days of abstinence. Routine semen analysis was performed using a computer-aided sperm analysis system (Beijing Suijia Medical Instrument Co., China) according to the requirements of the World Health Organization guidelines (*World Health Organization, 2010*), and the basic checklist was completed (*Björndahl et al., 2016*). The remaining semen samples were used for the analysis of sperm DFI and preparation of sperm samples. These samples were divided into the experimental group and the control group according to the sperm DFI values. The values of DFI in the experimental group (n = 12) were 30% or more. (DFI

≥30%), while those in the control group (n = 12) were below 30% (DFI < 30%).

Spermatozoa were then isolated from each semen sample by the DGC method. Every four sperm samples in the experimental group and control group formed one replicate for sperm RNA analysis. Three replicate samples in the experimental group were labelled EXP1, EXP2 and EXP3, and three replicate samples in the control group were labelled CON1, CON2 and CON3. The ages of participants in the experimental and control groups were 30.42 ± 3.68 and 29.17 ± 3.97 years, respectively, and there was no significant difference between them (P = 0.432). The semen characteristics and sperm DFI values of all participants are shown in Supplementary Table 1 (https://data.mendeley.com/datasets/7ppzv7r2sb/1). This study was approved by the Northern Jiangsu People's Hospital ethics committee on 2 March 2021 (approval number: 2021ky068), and all the participants provided informed written consent.

Detection of sperm DFI

Sperm DFI was assessed using a sperm chromatin structure assay kit (Zhejiang Cellpro Biotech Co., China). First, an appropriate volume of semen was added to 0.1 ml of solution A (TNE buffer, sperm dilution) and mixed. Then, 0.2 ml of solution B (an acid solution of 0.1% Triton X-100, 0.15 mol/l NaCl and 0.08 mol/l HCl; pH 1.2) was added and mixed in. After standing for 30 s, 0.6 ml of acridine orange staining solution (6 μg/ml acridine orange, 37 mmol/l citric acid, 126 mmol/l Na_2HPO_4, 1 mmol/l Na_2EDTA, 0.15 mol/l NaCl; pH 6.0) was added and mixed in. After the spermatozoa had been staining for 3 min, the sperm DFI was detected using a fluorescence-activated cell sorter (FACSCalibur, BD Bioscience, USA). A minimum of 5000 spermatozoa were acquired, and the data were analysed using DFIView 2010 Alpha11.15 software (CellPro Biotech, China). The sperm DFI was expressed as the percentage of spermatozoa with fragmented DNA compared with the total number of spermatozoa. The variability of the replicate DFI measures was less than 5%.

As the spermatozoa used for RNA analysis were selected by DGC to remove non-sperm cells, it was wondered whether there was still a difference in sperm DFI between the experimental group and control group after DGC. To verify this, the sperm DFI of six sperm samples in each group was compared before and after DGC.

Preparation of sperm samples

Sperm samples were prepared by the DGC method according to the report of de Mateo and colleagues (*de Mateo et al., 2013*). In brief, SpermGrad lower layer (90%) and upper layer (45%) solution and SpermRinse solutions (Vitrolife, Sweden) were taken out of a refrigerator and allowed to warm to room temperature for further use. First, 1 ml of SpermGrad lower layer (90%) solution was added into a 15 ml centrifuge tube, and then 1 ml of SpermGrad upper layer (45%) solution was gently added on the surface of the SpermGrad lower layer (90%) solution. Next, normally liquefied semen was slowly added, and a clear interface between the semen and gradient solutions could be seen. After 20 min of centrifugation at 400g, the upper liquids were carefully aspirated away using a pipette, and the sperm sediments were transferred into a new centrifuge tube with the help of 3 ml of SpermRinse solution. The mixture was gently mixed, and then centrifuged for 10 min at 200g. Next, the upper liquids were carefully aspirated away using a pipette. The obtained spermatozoa were stored at –80°C and used for the extraction of RNA.

High-throughput RNA sequencing

Total sperm RNA was extracted using TRIzol reagents (batch number:15596018; Ambion, USA) according to the manufacturer's protocol, and the purity and concentration of total RNA were measured using a spectrophotometer (NanoDrop NC2000, Thermo Fisher Scientific, USA; Supplementary Table 2, https://data.mendeley.com/datasets/7ppzv7r2sb/1). When the quality of total RNA was deemed acceptable, which was when the value for the optical density 260/280 of the total RNA was between 1.8 and 2.0, the ribosomal RNA was removed from the sperm RNA using an Ribo-Zero rRNA removal kit (batch number: RZH1046; Epicentre, USA).

Then strand-specific RNA libraries were prepared using a NEBNext Ultra Directional RNA Library Prep kit (batch number: E7420S; New England Biolabs, USA). Next, sperm small RNA were converted into indexed libraries for next-generation sequencing with a NEBNext Multiplex Small RNA Library Prep Set for Illumina kit (batch number: E7580L; New England Biolabs, USA). The quantities of double-stranded DNA (ds DNA) were determined using a Quant-iT PicoGreen dsDNA Assay kit (batch number: P7589; Invitrogen, USA). The quality of the final libraries was evaluated using an Agilent Bioanalyzer 2100 system and DNA High Sensitivity kit (batch number: 5067-4626; Agilent, Germany). The sequencing library was then sequenced on a NovaSeq 6000 platform (Illumina, USA).

The quality control of raw and trimmed paired-end reads was performed using FastQC v.0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimming was done with a Trimmomatic v.0.38 to remove low-quality reads and adaptors (*Bolger et al., 2014*). Trimmed reads were mapped to the human reference genome using HISAT2 v.2.0 software (https://daehwankimlab.github.io/hisat2/download/) with default parameters except for "–max seed 30" and "–k 2" (*Kim et al., 2015*). The output included reference transcripts as well as any novel transcripts that were assembled. The size of the clean dataset for each sample was 15G for mRNA and 10M for small RNA. Total RNA-Seq reads were evaluated using FastQC software for quality control. The RNA levels of genes annotated in the human genome were quantified as fragments per kilobase of transcript per million mapped reads (FPKM) by StringTie v.1.3.3b (https://ccb.jhu.edu/software/stringtie/), and annotation was performed with the –G option of Ensembl Release 104 (http://www.ensembl.org/index.html).

The abundance of differentially expressed genes between the experimental group and control group was evaluated using DESeq2. The differential expressions of mRNA were analysed based on the Blast results of the miRBase database and prediction results of MIREAP (https://help.rc.ufl.edu/doc/Mireap), and read counts were normalized using the DESeq2 R/Bioconductor package (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html). The gene set enrichment analysis of the detected genes was undertaken using the OmicsBox platform (http://www.biobam.com/omicsbox/). In addition, the canonical pathways, biological functions, gene networks and regulators of genes detected in sperm samples were identified and characterized with Ingenuity Pathway

TABLE 1 PRIMERS USED FOR REAL-TIME QUANTITATIVE PCR

Gene name	Primers sequence (5′-3′)	Amplicon size (bp)
PMS1	Forward TGACAGCAGATGACCAAAGATACAG	230
	Reverse GCGAGGTCTACATTCATAAACTTCC	
TP53BP1	Forward GGCTGTCTGATGTGGATGCTAATAC	86
	Reverse CTTGCTGGACTGTTCTGCTATGG	
TLK2	Forward TTGAGACGGCAGATTGATGAACAG	135
	Reverse GCTCTTATCTCTACACGCCATCTTC	
β-Actin	Forward TGGACTTCGAGCAAGAGATG	137
	Reverse GAAGGAAGGCTGGAAGAGTG	

Analysis (Ingenuity Systems, Qiagen, USA). If the RNA or dsDNA samples obtained in the above process needed to be stored, they were stored at –80°C.

Real-time quantitative PCR

In order to validate the analysis results of differentially expressed genes, three selected differentially expressed genes related to sperm DNA damage and repair, namely *PMS1*, *TP53BP1* and *TLK2*, were verified in sperm samples with a high DFI (n = 3) and normal DFI (n = 3) by real-time quantitative PCR (RT-qPCR). First, total RNA was extracted from each sperm sample using TRIzol reagents (Invitrogen, USA) according to the manufacturer's instructions. Then the RNA was reverse-transcribed into cDNA using a reverse transcription kit (TaKaRa, Dalian, China). Finally, an RT-qPCR kit (TaKaRa) was used to analyse the expression levels of *PMS1*, *TP53BP1* and *TLK2* relative to β-actin on a CFX Connect real-time PCR detection system (Bio-Rad Laboratories, USA). The primer sequences of *PMS1*, *TP53BP1*, *TLK2* and β-actin are shown in TABLE 1. The amplification conditions were as follows: 95°C for 30 s, and then 40 cycles of 5 s at 95°C and 35 s at 60°C.

Functional enrichment analysis

In order to explore the functions and pathways related to the differentially expressed RNA, Gene Ontology term and KEGG pathway analysis was performed using the DAVID bioinformatics tool (version 6.8, https://david.ncifcrf.gov/). The threshold was set as $P < 0.05$.

Statistical analysis

The differential expressions of sperm mRNA between the experimental and control groups were analysed using the DESeq2 R/Bioconductor package, and candidates with a value of $P < 0.05$ and a $|\log_2$ fold change$|>1$ were considered as differentially expressed genes. The differences in sperm DFI and the levels of the differentially expressed genes *PMS1*, *TP53BP1* and *TLK2* between the experimental and control groups were analysed using the Student's t-test in SPSS 22.0 statistical software (SPSS Inc., USA), with $P < 0.05$ considered statistically significant.

RESULTS

Comparison of sperm DFI between the experimental group and control group

There was a statistically significant difference in sperm DFI between the experimental group (DFI ≥30%) and the control group (DFI <30%) (40.65% versus 11.42%, $P < 0.001$; FIGURE 1A). Regardless of whether it was before DGC (35.96% versus 14.70%, $P < 0.001$) or after DGC (8.83% versus 2.34%, $P = 0.005$), the sperm DFI in the experimental group was always significantly higher than that in the control group (FIGURE 1B). The DFI results for each sperm sample are shown in Supplementary Table 1 (https://data.mendeley.com/datasets/7ppzv7r2sb/1).

Sequencing output and distribution of mRNA detected in human sperm samples

The results of high-throughput sequencing showed that there was an average of 109.0 million reads (about 16.4 Gbps) per human sperm sample after RNA extraction. Clean reads accounted for about 85.0% of the total reads, with an average of 92.4 million. About 54.8% of clean reads (50.6 million) were aligned to the human genome. A total of 19,970 transcripts were identified from human sperm samples. The majority of genes were moderately expressed in the experimental group and control group, as shown by the density distribution (FIGURE 2A) and violin plot (FIGURE 2B) of FPKM. The results for the

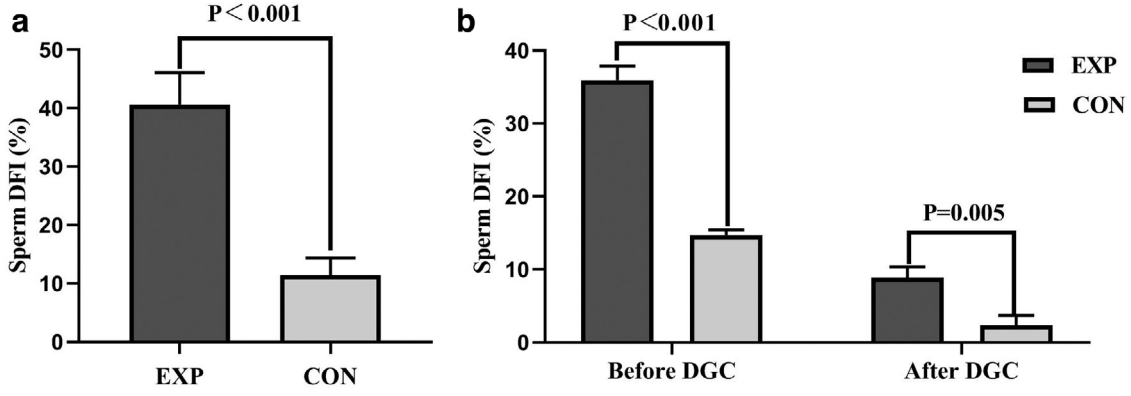

FIGURE 1 Comparison of the sperm DNA fragmentation index (DFI) between the experimental and control groups. (A) Comparison of sperm DFI between the experimental group (EXP, n = 12) and control group (CON, n = 12). Error bars represent the standard deviation. The sperm DFI in the experimental group was significantly higher than that in the control group. (B) Comparison of sperm DFI before and after density gradient centrifugation (DGC) in the experimental group (EXP, n = 6) and control group (CON, n = 6). Error bars represent the standard deviation. Regardless of whether it was before DGC or after DGC, the sperm DFI in the experimental group was always significantly higher than that in the control group ($P < 0.001$ and $P = 0.005$).

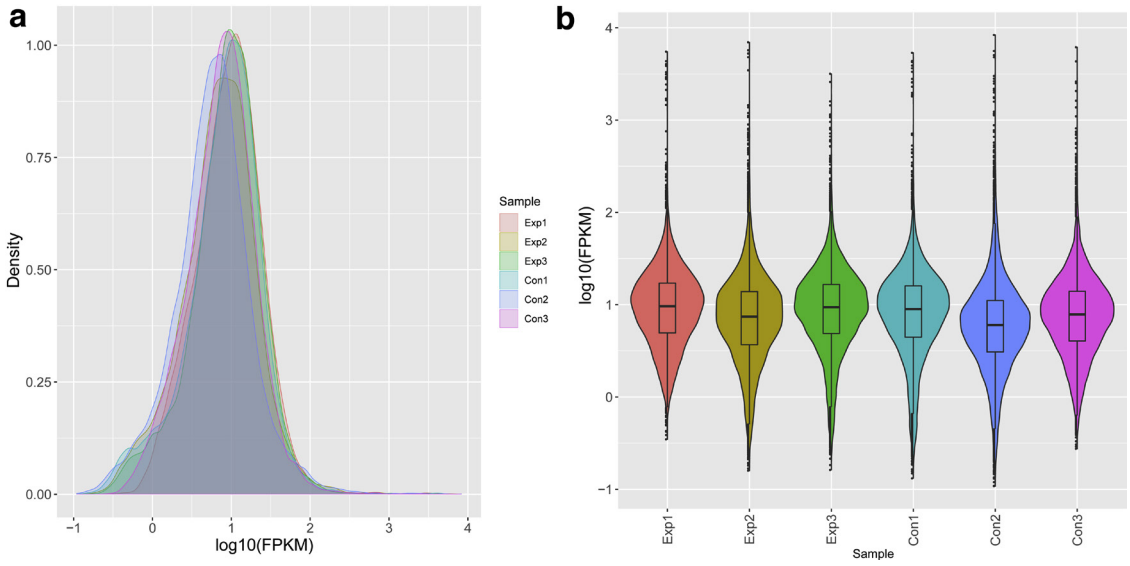

FIGURE 2 The results of RNA sequencing showed the density distribution (A) and violin plot (B) of the fragments per kilobase of transcript per million mapped reads (FPKM). Exp: experimental group, with a sperm DNA fragmentation index (DFI) ≥30%; Con: control group with a DFI <30%. The violin plot shows the mRNA levels in each sample.

RNA concentration, quality and purity of each sperm sample and mapped reads are shown in Supplementary Table 2 (https://data.mendeley.com/datasets/7ppzv7r2sb/1).

Analysis of differentially expressed genes related to sperm DNA fragmentation

The genes that were differentially expressed between the experimental group and the control group were analysed using DESeq version 1.18.0. A total of 19,970 expressed genes were detected in the two groups. Compared with the control group, the expression levels of 352 genes in the experimental group were significantly altered, of which 189 genes were significantly increased and 163 genes decreased. The cluster distances of differentially expressed genes were calculated by the Euclidean method (FIGURE 3A). The volcano plot of differentially expressed genes was drawn using the R language ggplots2 software package (https://data.mendeley.com/datasets/7ppzv7r2sb/1), and the results showed that the expressions of the normalized genes were distributed symmetrically (FIGURE 3B). According to the genomic information and the analysis results of the differentially expressed genes, the differentially expressed genes marked on the genome were distributed across all chromosomes (FIGURE 3C).

To further validate the analysis results for the differentially expressed genes, three selected differentially expressed genes related to sperm DNA damage and repair, namely *PMS1*, *TP53BP1* and *TLK2*, were verified in sperm samples from the experimental group and control group by RT-qPCR. The results of RT-qPCR showed that the expression levels of the three genes in the experimental group were significantly lower than those in the control group (P = 0.01, 0.015 and 0.004, respectively), which was identical to the results of RNA-seq (FIGURE 3D). The details for the genes that were differentially expressed between the experimental group and control group are shown in Supplementary Table 3 (https://data.mendeley.com/datasets/7ppzv7r2sb/1).

Functional annotation and identification of differentially expressed genes

Gene Ontology enrichment and KEGG pathway analysis of differentially expressed genes can help to gain insight into the biological characteristics of such genes. Gene Ontology enrichment analysis showed that genes that were differentially expressed between the experimental group and the control group were mainly concentrated in the pathways related to the ATPase-dependent transmembrane transport complex, extracellular exosome, somatic cell DNA recombination, protein binding, cytoplasm, regulation of localization, etc. (FIGURE 4A, B). KEGG pathway analysis showed that differentially expressed genes were mainly related to the PI3K-Akt (phosphatidylinositol-3-kinase-Akt) signalling pathway, endocytosis, the p53 signalling pathway, the cGMP-PKG (cyclic guanosine monophosphate-cGMP dependent protein kinase G) signalling pathway, etc. (FIGURE 4C, D), which might be involved in the mechanism of sperm DNA damage. The details of the Gene Ontology enrichment and KEGG pathway analysis of differentially expressed genes are shown in Supplementary Table 4 (https://data.mendeley.com/datasets/7ppzv7r2sb/1).

DISCUSSION

Spermatogenesis is a complex and tightly regulated process leading to the continuous production of male gametes (spermatozoa), which is one of the fundamental processes of sexual reproduction. Semen analysis remains fundamental for understanding the underlying aetiologies of male infertility. The sperm DFI has also become an important indicator of sperm functional quality, and the identification of underlying conserved molecular mechanisms can reveal the core components of the spermatogenic process, which may provide novel insight into the causes of human infertility (*Zhang et al., 2016*).

Zhang and colleagues (*Zhang et al., 2019*) found that lncRNA expression was associated with sperm motility. RNA-seq technology has been successfully applied to investigations of fertilization and the relationship between the transcriptome of ejaculated spermatozoa and sperm quality or fertility (*Gòdia et al., 2020b*; *Xu et al., 2021*). It was

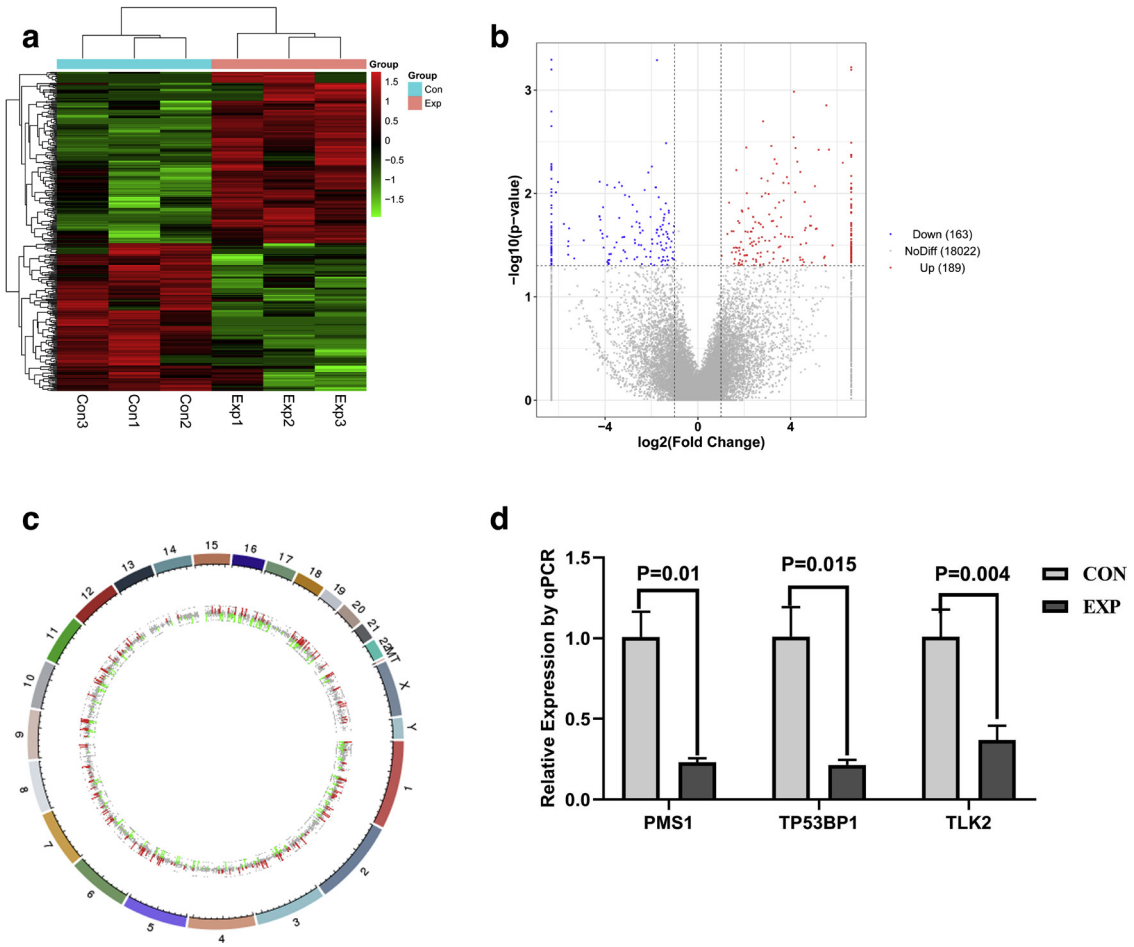

FIGURE 3 Analysis of differentially expressed genes related to sperm DNA fragmentation. (A) The heat map shows the hierarchical clustering analysis of genes detected in the control (Con1–3) and experimental (Exp1–3) groups. Horizontal lines represent genes, and each column is a sample. The red lines represent highly expressed genes, and the green lines low-expressed genes. Exp: experimental group with a sperm DNA fragmentation index (DFI) ≥30%; Con: control group with a DFI <30%. (B) The volcano plot shows the distribution of differentially expressed genes according to their *P*-values and fold changes. Candidates with a value of $P < 0.05$ and a $|\log_2$ fold change$|>1$ were considered as differentially expressed genes. Red and blue plots in the histograms of $\log_2$(fold change) represent up-regulated and down-regulated genes, respectively, and grey plots are non-differentially expressed genes. (C) The circle genome map of differentially expressed genes shows that these genes were distributed across all chromosomes. (D) The validation results of real-time quantitative PCR (qPCR) show that the expression levels of three randomly selected differentially expressed genes (*PMS1*, *TP53BP1* and *TLK2*) in the experimental group ($n = 3$) were significantly lower than those in the control group ($n = 3$), which was identical to the results of RNA sequencing. Error bars represent the standard deviation.

found that transcription in spermatozoa was associated with molecular processes such as chemoattractant-activated signal transduction, ion transport and cellular components of membranes and vesicles, and that some mRNA and microRNA might be critical for the early development of embryos. The abundant pools of coding and non-coding RNA in stallion spermatozoa were found to be not the random remnants of spermatogenesis in the testis, but selectively retained and functional consistent RNA collections (*Das et al., 2013*). RNA-seq technology was also successfully applied to the study of boar semen, which contained potential pathogens and putative antibiotic resistance genes that might affect reproductive capacity (*Gòdia et al., 2020a*; *Lian et al., 2021*). RNA-seq technology was also applied to investigating the effects of seasonal changes in the transcriptome of boar spermatozoa (*Gòdia et al., 2019*). The biological function of sperm RNA is thus constantly being rewritten. The study of differentially expressed genes at the level of transcriptome in spermatozoa with a high DFI or normal DFI by RNA-seq could provide relevant informations on the mechanisms leading to sperm DNA damage, which may be helpful for the future diagnosis and treatment of male infertility and in animal breeding.

In the current study, semen samples with a high or normal sperm DFI were first collected. In order to avoid the influence of non-sperm cells on the results of sperm RNA analysis, sperm samples were selected by DGC. Although the sperm DFI was significantly reduced after the optimization of DGC, the sperm DFI in the experimental group was still significantly higher than that in the control group. Next, differentially expressed genes at the level of the transcriptome in the optimized spermatozoa of the two groups were analysed using RNA-seq technology. By further bioinformatics analysis, quality control and comparisons of the data, a total of 352 differentially expressed genes were screened out. Compared with the control group, 189 genes were significantly increased and 163 genes were decreased in the experimental group.

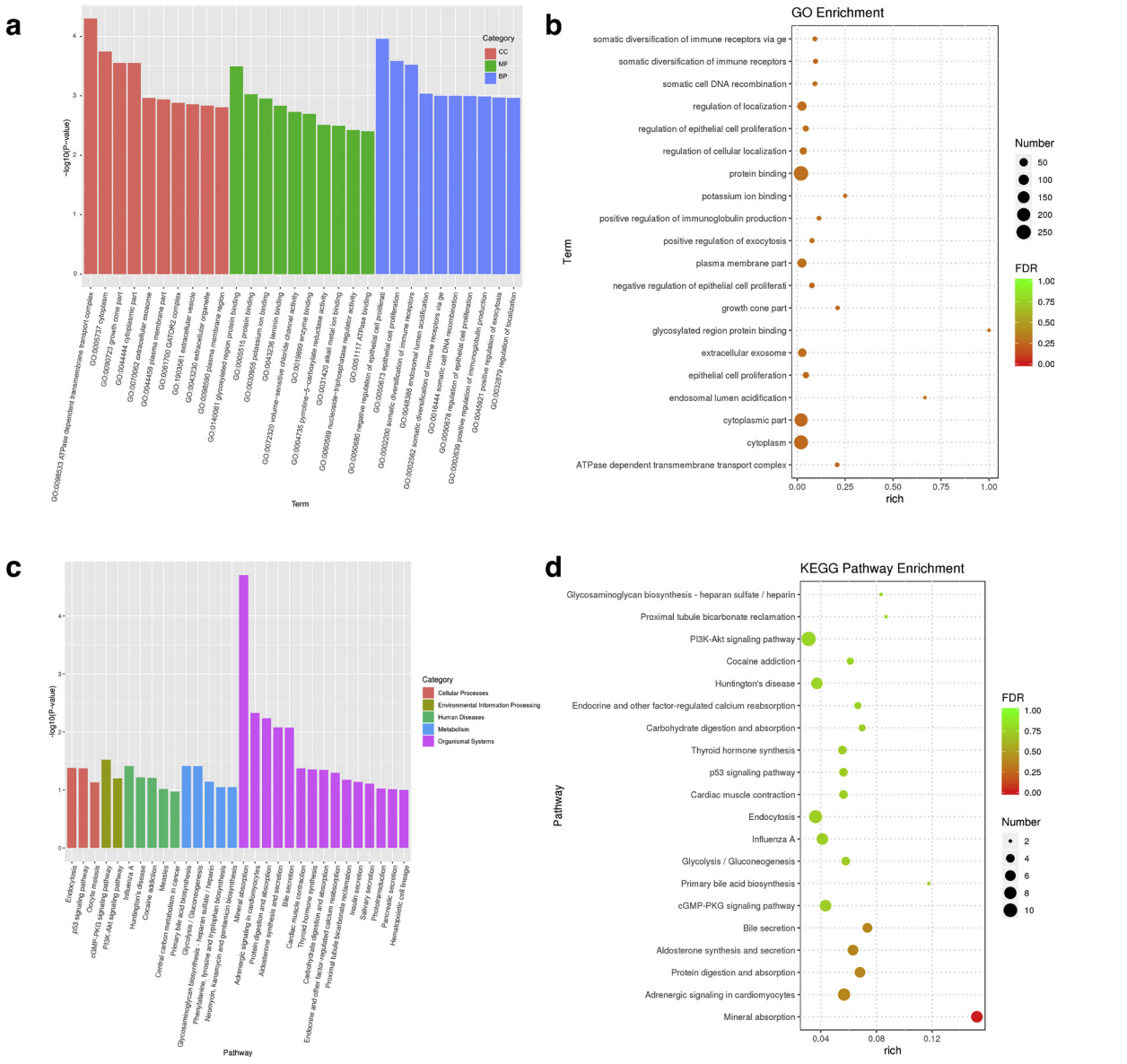

FIGURE 4 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of differentially expressed genes. (A) Histograms of the Gene Ontology analysis of differentially expressed genes. BP, biological process; CC, cellular component; MF, molecular function. (B) Bubble chart of the Gene Ontology enrichment analysis of differentially expressed genes. FDR, false discovery rate. (C) Histograms of the KEGG analysis of differentially expressed genes. (D) Bubble chart of the KEGG enrichment analysis of differentially expressed genes.

Gene Ontology analysis showed that the entries annotated as cellular components were mainly related to ATPase-dependent transmembrane transport complexes, cytoplasmic parts and extracellular exosomes, the main entries annotated to molecular functions were in glycosylation region protein binding, protein binding and potassium ion binding, and that biological processes were mainly related to items of negative regulation of epithelial cell proliferation, epithelial cell proliferation, somatic immune receptor diversification and somatic DNA recombination. Gene Ontology enrichment analysis was mainly concentrated in the areas of protein binding, cytoplasm, extracellular vesicles, cell membrane composition and somatic DNA recombination.

The KEGG analysis of differentially expressed genes showed that the main differential pathways included endocytosis, the p53 signalling pathway, the PI3K-Akt signalling pathway and the cGMP-PKG signalling pathway. KEGG enrichment analysis showed that significant pathways included aldosterone synthesis and secretion, the p53 signalling pathway and the PI3K-Akt signalling pathway. These pathways may play important roles in sperm DNA damage and repair.

In order to validate the analysis results of differentially expressed genes, the expression levels of three differentially expressed genes in the sperm samples of the experimental group and control group related to sperm DNA damage and repair, namely *PMS1*, *TP53BP1* and *TLK2*, were detected by RT-qPCR. TLKs (tousled-like kinases) belong to the nuclear serine-threonine kinase family, which are essential for genome stability, proper cell division and normal development in numerous organisms

(*Simon et al., 2022*); their activity has been linked to DNA replication, DNA repair, chromatin structure, etc. (*Segura-Bayona et al., 2017*). In humans, TLK2 may enhance ASF1 (anti-silencing function 1) histone binding, and its activity can be inhibited by the DNA damage response (*Mortuza et al., 2018*).

PMS1 (postmeiotic segregation in creased 1, mismatch repair system component) is a key element in the DNA mismatch repair (MMR) system and its role in DNA repair is authenticated by studies. MMR models have proposed that MSH (MutS homolog) proteins identify DNA polymerase and correct replication errors that may cause mutations, with the resulting MSH–MLH (MutL homolog) complex formed at a DNA lesion initiating downstream steps in repair (*Plys et al., 2012*). Mlh1–Pms1 may be relevant to signal transduction during DNA mismatch repair and to recombination, meiosis and cellular responses to DNA damage (*Drotschmann et al., 2002*). PMS1, MLH1, and MSH2 are likely to form a ternary complex during the initiation of eukaryotic DNA MMR (*Prolla et al., 1994*).

Numerous pieces of research have revealed that TP53BP1 (tumour suppressor p53-binding protein 1) is a central transducer of DNA-damage signals to p53; it may work in the p53-mediated transcriptional activation and DNA damage-repair signalling pathways (*Ma et al., 2006*), and plays an essential role in the cellular response to DNA damage, chromosome stability and spindle bipolarity during meiotic maturation (*Jin et al., 2019*). *TP53BP1* plays an important role in both DNA repair and cell cycle control and is a key transducer of the DNA damage checkpoint signal through cooperation with damage sensors (*Wang et al., 2002*). The validation results of the current study also demonstrated that the expression levels of the three genes in spermatozoa with a high DFI were significantly lower than those with a normal DFI, and their trends in expression were identical to those of RNA sequencing, indicating that the increased sperm DFI might be related to the reduced expression of sperm DNA repair genes.

In conclusion, some differentially expressed genes and pathways related to sperm DNA damage and repair have been found, which provides new insights into the understanding of sperm DNA damage and repair and will help in the discovery of new biomarkers related to sperm DNA damage. However, this study also has some shortcomings. For example, mixed samples were used, and it was difficult to obtain RNA from spermatozoa. Because spermatogonia undergo mitosis, meiosis and sperm deformation stages during spermatogenesis, little transcription or repair takes place in the DNA of post-meiotic male germ cells, and the expression levels of RNA in terminally differentiated spermatozoa may not be enough to reflect the whole process of spermatogenesis. Future explorations into differentially expressed genes should focus on not only mature sperm, but also the entire process of spermatogenesis. The specific molecular mechanism of sperm DNA damage and repair needs to be further explored.

FUNDING

This study was funded by the Jiangsu Postgraduate Training Innovation Project (grant number KYCX17_1888) and the National Natural Science Foundation of China (grant number 81773013).

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found https://data.mendeley.com/datasets/7ppzv7r2sb/1, and in the online version, at doi:10.1016/j.rbmo.2022.08.108.

REFERENCES

Ablondi, M., Godia, M., Rodriguez-Gil, J.E., Sanchez, A., Clop, A. **Characterisation of sperm piRNAs and their correlation with semen quality traits in swine.** Anim. Genet. 2021; 52: 114–120

Agarwal, A., Majzoub, A., Baskaran, S., Panner Selvam, M.K., Cho, C.L., Henkel, R., Finelli, R., Leisegang, K., Sengupta, P., Barbarosie, C., Parekh, N., Alves, M.G., Ko, E., Arafa, M., Tadros, N., Ramasamy, R., Kavoussi, P., Ambar, R., Kuchakulla, M., Robert, K.A., Iovine, C., Durairajanayagam, D., Jindal, S., Shah, R. **Sperm DNA fragmentation: A new guideline for clinicians.** World J. Mens Health 2020; 38: 412–471

Agarwal, A., Mulgund, A., Hamada, A., Chyatte, M.R. **A unique view on male infertility around the globe.** Reprod. Biol. Endocrinol. 2015; 13: 37

Bissonnette, N., Levesque-Sergerie, J.P., Thibault, C., Boissonneault, G. **Spermatozoal transcriptome profiling for bull sperm motility: a potential tool to evaluate semen quality.** Reproduction 2009; 138: 65–80

Björndahl, L., Barratt, C.L., Mortimer, D., Jouannet, P. **How to count sperm properly': checklist for acceptability of studies based on human semen analysis.** Hum. Reprod. 2016; 31: 227–232

Boerke, A., Dieleman, S.J., Gadella, B.M. **A possible role for sperm RNA in early embryo development.** Theriogenology 2007; 68: S147–S155

Bolger, A.M., Lohse, M., Usadel, B. **Trimmomatic: a flexible trimmer for Illumina sequence data.** Bioinformatics 2014; 30: 2114–2120

Capra, E., Turri, F., Lazzari, B., Cremonesi, P., Gliozzi, T.M., Fojadelli, I., Stella, A., Pizzi, F. **Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between high- and low-motile sperm populations.** BMC Genomics 2017; 18: 14

Chen, Q., Yan, W., Duan, E.K. **Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications.** Nat. Rev. Genet. 2016; 17: 733–743

Dal Molin, A., Baruzzo, G., Di Camillo, B. **Single-cell RNA-sequencing: Assessment of differential expression analysis methods.** Front. Genet. 2017; 8: 62

Das, P.J., McCarthy, F., Vishnoi, M., Paria, N., Gresham, C., Li, G., Kachroo, P., Sudderth, A.K., Teague, S., Love, C.C., Varner, D.D., Chowdhary, B.P., Raudsepp, T. **Stallion sperm transcriptome comprises functionally coherent coding and regulatory RNAs as revealed by microarray analysis and RNA-seq.** PLoS One 2013; 8: e56535

de Mateo, S., Estanyol, J.M., Oliva, R. **Methods for the analysis of the sperm proteome.** Methods Mol. Biol. 2013; 927: 411–422

Drotschmann, K., Hall, M.C., Shcherbakova, P.V., Wang, H., Erie, D.A., Brownewell, F.R., Kool, E.T., Kunkel, T.A. **DNA binding properties of the yeast Msh2-Msh6 and Mlh1-Pms1 heterodimers.** Biol. Chem. 2002; 383: 969–975

Fang, P., Zeng, P., Wang, Z., Liu, M., Xu, W., Dai, J., Zhao, X., Zhang, D., Liang, D., Chen, X., Shi, S., Zhang, M., Wang, L., Qiao, Z., Shi, H. **Estimated diversity of messenger RNAs in each murine spermatozoa and their potential**

function during early zygotic development. Biol. Reprod. 2014; 90: 94
Gòdia, M., Swanson, G., Krawetz, S.A. **A history of why fathers' RNA matters.** Biol. Reprod. 2018; 99: 147–159
Gòdia, M., Estill, M., Castelló, A., Balasch, S., Rodríguez-Gil, J.E., Krawetz, S.A., Sánchez, A., Clop, A. **A RNA-Seq analysis to describe the boar sperm transcriptome and its seasonal changes.** Front. Genet. 2019; 10: 299
Gòdia, M., Reverter, A., González-Prendes, R., Ramayo-Caldas, Y., Castelló, A., Rodríguez-Gil, J.E., Sánchez, A., Clop, A. **A systems biology framework integrating GWAS and RNA-seq to shed light on the molecular basis of sperm quality in swine.** Genet. Sel. Evol. 2020; 52: 72
Gòdia, M., Ramayo-Caldas, Y., Zingaretti, L.M., Darwich, L., López, S., Rodríguez-Gil, J.E., Yeste, M., Sánchez, A., Clop, A. **A pilot RNA-seq study in 40 pietrain ejaculates to characterize the porcine sperm microbiome.** Theriogenology 2020; 157: 525–533
Jin, Z.L., Suk, N., Kim, N.H. **TP53BP1 regulates chromosome alignment and spindle bipolarity in mouse oocytes.** Mol. Reprod. Dev. 2019; 86: 1126–1137
Jodar, M., Sendler, E., Moskovtsev, S.I., Librach, C.L., Goodrich, R., Swanson, S., Hauser, R., Diamond, M.P., Krawetz, S.A. **Absence of sperm RNA elements correlates with idiopathic male infertility.** Sci. Transl. Med. 2015; 7: 295re6
Kim, D., Langmead, B., Salzberg, S.L. **HISAT: a fast spliced aligner with low memory requirements.** Nat. Methods 2015; 12: 357–360
Krawetz, S.A., Kruger, A., Lalancette, C., Tagett, R., Anton, E., Draghici, S., Diamond, M.P. **A survey of small RNAs in human sperm.** Hum. Reprod. 2011; 26: 3401–3412
Li, X., Duan, C., Li, R., Wang, D. **Insights into the mechanism of bovine spermiogenesis based on comparative transcriptomic studies.** Animals (Basel) 2021; 11: 80
Lian, Y., Gòdia, M., Castello, A., Rodriguez-Gil, J.E., Balasch, S., Sanchez, A., Clop, A. **Characterization of the impact of density gradient centrifugation on the profile of the pig sperm transcriptome by RNA-Seq. Front.** Vet. Sci. 2021; 8668158
Ma, H., Hu, Z., Zhai, X., Wang, S., Wang, X., Qin, J., Chen, W., Jin, G., Liu, J., Gao, J., Wang, X., Wei, Q., Shen, H. **Joint effects of single nucleotide polymorphisms in P53BP1 and p53 on breast cancer risk in a Chinese population.** Carcinogenesis 2006; 27: 766–771
Malic Voncina, S., Stenqvist, A., Bungum, M., Schyman, T., Giwercman, A. **Sperm DNA fragmentation index and cumulative live birth rate in a cohort of 2,713 couples undergoing assisted reproduction treatment.** Fertil. Steril. 2021; 116: 1483–1490
Mortuza, G.B., Hermida, D., Pedersen, A.K., Segura-Bayona, S., López-Méndez, B., Redondo, P., Rüther, P., Pozdnyakova, I., Garrote, A.M., Muñoz, I.G., Villamor-Payà, M., Jauset, C., Olsen, J.V., Stracker, T.H., Montoya, G. **Molecular basis of Tousled-Like Kinase 2 activation.** Nat. Commun. 2018; 9: 2535
Plys, A.J., Rogacheva, M.V., Greene, E.C., Alani, E. **The unstructured linker arms of Mlh1-Pms1 are important for interactions with DNA during mismatch repair.** J. Mol. Biol. 2012; 422: 192–203
Prolla, T.A., Pang, Q., Alani, E., Kolodner, R.D., Liskay, R.M. **MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast.** Science 1994; 265: 1091–1093
Rajesh, A., Yenugu, S. **shRNA mediated ablation of prostate and testis expressed (Pate) messenger RNA results in impaired sperm function and fertility.** Andrology 2017; 5: 541–547
Segura-Bayona, S., Knobel, P.A., González-Burón, H., Youssef, S.A., Peña-Blanco, A., Coyaud, É., López-Rovira, T., Rein, K., Palenzuela, L., Colombelli, J., Forrow, S., Raught, B., Groth, A., de Bruin, A., Stracker, T.H. **Differential requirements for Tousled-like kinases 1 and 2 in mammalian development.** Cell Death Differ. 2017; 24: 1872–1885
Sendler, E., Johnson, G.D., Mao, S.H., Goodrich, R.J., Diamond, M.P., Hauser, R., Krawetz, S.A. **Stability, delivery and functions of human sperm RNAs at fertilization.** Nucleic Acids Res. 2013; 41: 4104–4117
Shafeeque, C.M., Singh, R.P., Sharma, S.K., Mohan, J., Sastry, K.V.H., Kolluri, G., Saxena, V.K., Tyagi, J.S., Kataria, J.M., Azeez, P.A. **Development of a new method for sperm RNA purification in the chicken.** Anim. Reprod. Sci. 2014; 149: 259–265
Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W., Sun, F., Song, L., Carone, B.R., Ricci, E.P., Li, X.Z., Fauquier, L., Moore, M.J., Sullivan, R., Mello, C.C., Garber, M., Rando, O.J. **Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals.** Science 2016; 351: 391–396
Simon, B., Lou, H.J., Huet-Calderwood, C., Shi, G., Boggon, T.J., Turk, B.E., Calderwood, D.A. **Tousled-like kinase 2 targets ASF1 histone chaperones through client mimicry.** Nat. Commun. 2022; 13: 749
Swanson, G.M., Moskovtsev, S., Librach, C., Pilsner, J.R., Goodrich, R., Krawetz, S.A. **What human sperm RNA-Seq tells us about the microbiome.** J. Assist. Reprod. Genet. 2020; 37: 359–368
Wang, B., Matsuoka, S., Carpenter, P.B., Elledge, S.J. **53BP1, a mediator of the DNA damage checkpoint.** Science 2002; 298: 1435–1438
World Health Organization. 2010 **WHO laboratory manual for the examination and processing of human semen.** 5th ed. World Health Organization Geneva
Xu, H., Sun, W., Pei, S., Li, W., Li, F., Yue, X. **Identification of key genes related to postnatal testicular development based on transcriptomic data of testis in Hu sheep.** Front. Genet. 2021; 12773695
Yan, B., Ye, W., Wang, J., Jia, S., Gu, X., Hu, H. **Evaluation of sperm DNA integrity by mean number of sperm DNA breaks rather than sperm DNA fragmentation index.** Clin. Chem. 2022; 68: 540–549
Zarezadeh, R., Nikanfar, S., Oghbaei, H., Rastgar Rezaei, Y., Jafari-Gharabaghlou, D., Ahmadi, Y., Nouri, M., Fattahi, A., Dittrich, R. **Omics in seminal plasma: An effective strategy for predicting sperm retrieval outcome in non-obstructive azoospermia.** Mol. Diagn. Ther. 2021; 25: 315–325
Zhang, C.W., Gao, L.Z., Xu, E.Y.J. **LncRNA, a new component of expanding RNA-protein regulatory network important for animal sperm development.** Semin. Cell Dev. Biol. 2016; 59: 110–117
Zhang, X., Zhang, P., Song, D., Xiong, S., Zhang, H., Fu, J., Gao, F., Chen, H., Zeng, X. **Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia spermdagger.** Biol. Reprod. 2019; 100: 982–993

Received 1 June 2022; received in revised form 16 August 2022; accepted 18 August 2022.