Genome-wide DNA methylation profiles and small non-coding RNAs signatures study of sperm with high DNA fragmentation index

Minghua Liu
Changhai Hospital, Naval Medical University
Peiru Liu
Fudan University
Yunjian Chang
Chinese Academy of Sciences, University of Chinese Academy of Sciences
Beiying Xu
Chinese Academy of Sciences, University of Chinese Academy of Sciences
Nengzhuang Wang
Changhai Hospital, Naval Medical University
Lina Qin
Changhai Hospital, Naval Medical University
Jufen Zheng
Changhai Hospital, Naval Medical University
Yun Liu
Fudan University
Ligang Wu
Chinese Academy of Sciences, University of Chinese Academy of Sciences
Hongli Yan (hongliyan@smmu.edu.cn)
Changhai Hospital, Naval Medical University

Research Article

Keywords: DFI, DNA methylation, sncRNAs, WGBS, correlation matrices

Posted Date: February 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1336052/v1

Abstract

Background

Sperm DNA fragmentation index (DFI) is a more accurate measure of sperm quality than traditional semen analysis (sperm motility, concentration, and morphology) because DFI can better predict male fertility potential and assisted reproductive technology (ART) outcomes. However, no studies have compared the genome-wide DNA methylation profiles and small noncoding RNAs (sncRNAs) signatures of high DFI and low DFI sperm samples.

Methods

Whole-genome bisulfite sequencing (WGBS) and sncRNAs deep sequencing were performed to compare the genome-wide DNA methylation profiles and sncRNAs signatures of sperm samples of weak group(DFI > 30%) and normal group(DFI < 15%). In addition, we used the WGBS data to construct DNA methylation correlation matrices of the weak and normal group sperm.

Results

A total of 4939 diferentially methylated regions (DMRs) (3083 hypermethylated and 1856 hypomethylated) were identifed in the two group sperm samples. The hypermethylated DMRs were found to be mainly signifcantly enriched in the area of neurons and microtubule. Compared with the normal group, the global DNA methylation level of the weak group sperm showed a downward trend and the methylation correlation within the chromosomes of weak group sperm was weakened. 27 miRNAs, 151 tsRNAs and 70 rsRNAs were differently expressed between the two group sperm samples. Finally, we identified nine sncRNAs as candidate sperm quality biomarkers. In addition, the target genes of both the downregulated and upregulated miRNAs were involved in the nervous system and cell development

Conclusions

Our findings suggest that the genome-wide DNA methylation levels and sncRNAs expression of sperm were significantly correlated with sperm DFI levels. Therefore, our study provides useful biomarkers for improving the success rate of natural pregnancy and ART. The chromosomes of high DFI sperm became loose and more vulnerable to ROS attack. The increase of sperm DFI level may affect embryonic nervous system development.

Background

Infertility refers to the failure of couples to achieve clinical pregnancy for more than one year without contraception[1]. Globally, more than 15% of couples of reproductive age suffer from infertility, of which 50% are caused by male factors[2]. Currently, assessment of sperm quality in infertile men still relies on traditional semen analysis, such as sperm motility, concentration, and morphology, but these analyses cannot accurately predict male fertility potential and assisted reproductive technology (ART) outcomes[3]. In fact, about 15% of infertile men have normal sperm motility, concentration, and morphology[4], but these analyses do not reflect sperm DNA integrity[5].

Sperm DNA Fragmentation (SDF) can be caused by extrinsic factors (smoking, high temperature, chemotherapeutic drugs and environmental pollutants) and intrinsic factors (oxidative stress[OS], apoptosis failure and germ cell maturation defects)[6]. Numerous studies have shown that OS is a major factor in male infertility[7]. Reactive oxygen species (ROS) are essential for physiological processes such as apoptosis and capacitation, but excess ROS can cause

SDF[8] and affect the outcome of natural pregnancy or ART. Over the past 20 years, there has been an increasing number of studies reporting that SDF is associated with male infertility, with the main focus of SDF research being on lifestyle factors, asthenozoospermia, and varicocele[9]. In addition, there is evidence that SDF affects the health and well-being of offspring[10].

Some studies showed that changes in sperm DNA methylation in imprinted regions are associated with male infertility and abnormal semen parameters (especially oligozoospermia) [11–15]. However, no specific DNA methylation signatures that convincingly replicate male infertility or abnormal semen parameters have been found in genome-scale unbiased analyses[16]. In addition, there is no clear evidence linking sperm DNA methylation with pregnancy outcomes or offspring health[16].

A growing number of studies have shown that mammalian sperm RNA is an important source of paternal genetic information in addition to DNA. Environmental factors (mental stress, endocrine disruptors, unhealthy diet, and toxin exposure)[17–19] can alter sperm RNA signatures and affect offspring phenotype. Many types of small noncoding RNAs (sncRNAs) have been found in mammalian male germ cells, such as microRNAs (miRNAs), tRNA-derived small RNAs (tsRNAs), rRNA-derived small RNAs (rsRNAs) and Piwi-interacting RNAs (piRNAs), which play an important regulatory role in spermatogenesis[20–25].

To date, no studies have demonstrated the effect of SDF on genome-wide DNA methylation profiles and sncRNAs signatures in human sperm. Through WGBS and sncRNAs deep sequencing in two groups of sperm with DNA fragmentation index (DFI) > 30% and DFI < 15%, we revealed that the genome-wide DNA methylation levels and sncRNAs expression of sperm were significantly correlated with sperm DFI, which has important clinical significance.

Methods

Sperm samples collection and processing

From June 2020 to October 2021, we collected sperm samples of weak group (DFI>30%) and normal group (DFI<15%) at the Reproductive Medicine Center of Changhai Hospital. Due to the presence of a large number of seminal plasma and somatic cells in the semen, we treat the semen with Phosphate Buffer Saline (PBS) and Somatic Cell Lysis Buffer (SCLB) respectively, and finally obtain pure sperm.

Whole-genome bisulfte sequencing (WGBS)

DNA of sperm samples was obtained from weak group (n=6) and normal group (n=7). DNA was extracted using QIAamp DNA Mini Kit Qiagen according to the manufacturer's protocols. 1% unmethylated lambda DNA (Promega, Madison, WI, USA) was added to each sperm sample to assess bisulfite conversion efficiency. Genomic DNA (200 ng) was then fragmented to an average size of 250 bp using a Covaris M220 ultrasonicator (Covaris, Woburn, MA, USA). End repair and methylated adaptor ligation were performed using NEBNext Ultra End Repair/ dA-Tailing Module, Ligation Module and NEBNext Multiplex Oligos for Illumina (Methylated Adaptor, Index Primers Set 1; New England Biolabs, Ipswich, MA, USA). DNA fragments between 300 and 400 bp were selected for library construction using Ampure XP beads (Beckman Coulter, Brea, CA, USA). Samples were then subjected to bisulfte conversion using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA), and modified single stranded DNA fragments were amplified using Kapa HiFi U+HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) with primers (NEBNext Multiplex Oligos for Illumina). The final size selection was performed using Ampure XP beads (Beckman Coulter, USA) to enrich the library to 300 500 bp. Constructed libraries were assessed for quality using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and sequenced on the Illumina NovaSeq (Illumina, San Diego, CA, USA) using the 150 bp paired-end mode

WGBS data analysis

Sequencing reads were analyzed using Bismark v.0.22.1[26]. The reads were aligned to the human genome (hg19) as well as the lambda phage genome using Bowtie2 v.2.2.3. PCR duplication reads were then removed, methylation measurements for each CpG site were obtained, and the bisulfte conversion rates were calculated based on the spiked-in unmethylated lambda phage DNA.

Identifying DMRs in the sperm samples using the bsseq package in BSmooth [27]. Only CpGs that appeared more than twice in at least three samples from weak and normal group were analyzed. DMRs were identified using a smoothing window containing 70 CpGs or 1 kb width, whichever was larger. The presumptive DMRs must meet the following 4 criteria at the same time: (1) must located on autosomes; (2) must have a methylation difference of 10% or more; (3) must contain more than 3 CpG sites; (4) must have a t-statistic with qcutoff in the range 0.025 to 0.975.

The pathway enrichment analysis

Sperm DMRs were annotated to their nearest genes. The hypergeometric test was performed with a focus on biological processes using the Molecular Signatures Database and the R package clusterProfler v.3.6.0 enricher function with default parameters (P-value cut-of=0.05; OrgDb=org.Hs.eg.db; and p.adjust method="BH"). Pathway enrichment analysis was performed with Kyoto Encyclopedia of Genes and Genomes (KEGG) database (P-value cut-of=0.05) and R package clusterProfler v.3.6.0.

Construction of WGBS correlation matrices

We constructed WGBS correlation matrices follow the method of Kasper D. Hansen[28]. In our correlation matrix, we only included so-called open sea CpGs that are more than 4 kb away from CpG islands. We then binned each chromosome into 100 kb bins, and took the correlations of the mean methylation levels for each bin. We obtained the first eigenvector of this binned correlation matrix and gently smoothed the signal by using two iterations of a moving average with a window size of three bins. The sign of the eigenvector is chosen so that the sign of the correlation between the eigenvector and column sums of the correlation matrix is positive which ensures that positive values of the eigenvector are associated with the closed compartment.

sncRNAs library preparation and deep sequencing

Total RNA of 13 weak group sperm samples and 17 normal group sperm samples was extracted using TRIzol reagent (Takara). Approximately 50 ng RNA was used to construct the sncRNAs library according to the Illumina protocol. High-throughput RNA sequencing was performed using a Hiseq X Ten (PE150). Fastp was used to clip adaptor and filter low quality reads. Reads that do not match the adapter or reads less than 17 nt in length were discarded. Redundant sequences were collapsed as useful reads for further analysis[29]. To assess the expression levels of miRNAs, only reads with an exact match to the 5' start site of the annotated miRNA with $\leq$ 2 nt deletions at the 3' end or additional sequences derived from pri-miRNAs were counted as miRNAs. Normalize miRNA counts to total miRNAs and multiplied by 1,000,000. Expression levels and classification of tsRNAs were based on GtRNAdb, and the tsRNA counts were also normalized to total tsRNAs and multiplied by 1,000,000. rsRNAs were mapped to rRNA precursors in the order of 5S, 5.8S, 18S, 28S and 45S, and the normalized method of rsRNA was same as that of miRNA and tsRNA. The remaining 25-32 nt sequences were used to identify the piRNAs[30]. sncRNAs with more than one annotation were characterized in the following order: miRNA, tsRNA, rsRNA, snRNA, snoRNA, and piRNA. Sequences not annotated with any of the above sncRNAs were classified as other sequences.

Sources of sequences and genome assemblies

The genomic sequences of human (hg38) were downloaded from the GENECODE (https://www.gencodegenes.org/). Known sncRNA sequences were retrieved from the following databases: miRNA, miRbase[31] (version 22.0 http://microrna.sanger.ac.uk/sequences/); tsRNAs, Genomic tRNA Database (http://lowelab.ucsc.edu/GtRNAdb/Mmusc); rsRNAs and yRNA, from NCBI GenBank (http://www.ncbi.nlm.nih.gov/), snoRNAs, from Ensembl (http://www.ensembl.org/index.html).

Statistical analysis

We used the R software package for statistical analysis. The differentially expressed miRNA tsRNA and rsRNA levels were analyzed using DESeq2[32]. Only sncRNAs with P-values less than 0.05 were considered differentially expressed sncRNAs. The Student's test was used to analyze the sncRNAs ratio in the weak and normal group which is a statistical significance test, and if $P < 0.05$, the association between the miRNAs/tsRNAs/rsRNAs and the two groups could be established. Unsupervised hierarchical clustering analysis was performed using Pheatmap[33]. The PCA was performed useing the prcomp package, which was visualized by ggbiplot[34].

miRNA functional analysis

The miRNA target genes were predicted by the online computational algorithm TargetScan (http://www.targetscan.org). The target gene enrichment tests of upregulated and downregulated miRNAs were annotated based on their biological processes GO terms[35].

Results

Genome-wide DNA methylation analysis of sperm

To compare the genome-wide DNA methylation profiles of sperm samples of the weak group (DFI>30%) and normal group (DFI<15%), 6 weak group sperm samples and 7 normal group sperm samples were examined using WGBS (Table 1). After quality control and data preprocessing, all samples were found to have a bisulfite conversion rate greater than 96%.

Compared with the normal group sperm samples, the global DNA methylation level of the weak group sperm samples showed a downward trend (Figure 1A). We then performed DMRs analysis to determine epigenetic differences between the two groups. A total of 4939 diferentially methylated regions (DMRs) (3083 hypermethylated and 1856 hypomethylated) were identifed in the weak group sperm samples relative to the normal group sperm samples (Table S1) , with 2072 of them (41.95%) located in promoter regions (1 to 3000 bp upstream or downstream of transcription start site) (Figure 1B). The percentages of hypermethylated DMRs were higher than hypomethylated DMRs in all of the seven examined gene annotation groups (Figure 1C).

The top 300 DMRs included 7921 CpGs with the median length of 1282.5 bp, and most of them were located at promoter regions. We were able to separate sperm samples into two groups corresponding to DFI>30% and DFI<15% (Figure 1D) using the top 300 DMRs, thus these 300 DMRs may be potential biomarkers for sperm quality.

To characterize the functional relevance of the 3083 hypermethylated DMRs, all of the 3083 hypermethylated DMRs were associated with their nearest genes and the gene ontology(GO) and pathway enrichment analysis was performed. The 3083 hypermethylated DMRs were found to be mainly signifcantly enriched in the area of neurons, such as axonogenesis, regulation of neuron projection development, synaplic membrane, axon guidance, neuron projection guidance and distal axon(Figure 1E). These findings suggest that the increase of sperm DFI level may affect embryonic nervous system development by causing epigenetic dysregulation of genes associated with neurons. In addition, the 3083 hypermethylated DMRs were also found to be signifcantly enriched in the area of microtubule (Figure 1E) which

is an important part of the 9+2 structure of the sperm tail and is very important to ensure sperm motility. Since the sperm motility of weak group sperm was significantly lower than that of the normal group sperm (Table 1), the increase of sperm DFI level may affect sperm tail structure and sperm motility by causing epigenetic dysregulation of genes associated with microtubule.

Six representative DMRs (CD14 cluster, TENM3 cluster, MB21D2, DAPL1 cluster, GLT1D1 cluster and ZNF516 cluster) identifed by WGBS are shown in Figure 2.

Chromosome Compartments Analysis

Because the global DNA methylation level of the weak group sperm samples showed a downward trend, we speculated that there might be differences in the spatial conformation of chromosomes between the two groups of samples. We constructed the chromosome compartments of sperm from the two groups using the WGBS data, and found that the compartments of five chromosomes(13, 4, 5, 21 and Y) in the weak group sperm changed compared with the normal group, and the correlations of methylation within the five chromosomes were weakened(Figure 3) which suggested that the structure of the five chromosomes of the weak group sperm become loose. Therefore, the chromosomes of the weak group sperm is more vulnerable to ROS attacks and more prone to break. Among these five chromosomes, the chromosome compartments and the correlations of methylation of chromosome Y changed the most between the two groups(Figure 3I, 3J), which may suggests that elevated DFI levels produce more damage on sperms with chromosome Y than chromosome X.

sncRNAs deep analysis

We extracted total RNA from 13 weak group sperm samples(DFI>30%) and 17 normal group sperm samples(DFI 15%) (Table 1), and analyzed the expression of sncRNAs by deep sequencing. We found that rsRNAs, tsRNAs, yRNAs and miRNAs were abundant in sperm samples. On average, about 40.5% of the sncRNAs annotated to rsRNAs, 19.3% to tsRNAs, 10.4% to yRNAs, and 7.1% to miRNAs (Table 2, Table S2). The length distribution of these sncRNAs was similar in each sample in both groups (Figure 4). The peak of the tsRNAs, rsRNAs, and miRNAs length ranged from 17 to 40 nt, 17 to 40 nt and 20 to 23 nt, respectively. We analyzed the proportions of miRNAs, tsRNAs, rsRNAs, sn/snoRNAs, yRNAs, and piRNAs in each sample (Figure 5A) and found that the proportions of these sncRNAs were not significantly different between the two groups (Figure 5B).

A total of 632 miRNAs (average RPM > 10) were detected in sperm samples (Table S3), of which 27 miRNAs were differently expressed between the two groups (9 up-regulated, 18 down-regulated) (Figure 6A, 6B, 6C and Table S4). Figure 6D shows the difference in the expression of the top 10 miRNAs by average expression between the two groups. Furthermore, we found that a principal component analysis (PCA), which is a powerful tool for exploratory data analysis and generating predictive models, could separate the weak group sperm samples from the normal group based on these 27 differently expressed miRNAs (PC1=48.72%, PC2=19.14%) (Figure 10A). These results indicated that the 27 tsRNAs have an excellent prognostic value and can be potential biomarkers for assessing human sperm quality. Moreover, the target genes of 16 of the 27 differentially expressed miRNAs were predicted by TargetScan. An analysis of significant GO-enriched terms showed that target genes of both the downregulated and upregulated miRNAs were involved in the nervous system and cell development(Figure 7), indicating that these miRNA target genes might be important for early embryo nervous system development and other systems development.

tsRNAs were high expressed in human sperm, and a total of 3612 tsRNAs (average RPM > 10) were detected in sperm samples(Table S5), of which 151 tsRNAs were differently expressed between the two groups (76 up-regulated, 75 down-regulated) (Figure 8A, 8B, 8C and Table S6). Figure 8D shows the difference in the expression of the top 10 tsRNAs by average expression between the two groups. PCA classifier analyses showed that these 151 differently expressed

tsRNAs could also classify the samples into two groups (PC1=28.79%, PC2=23.57%), indicating that these tsRNAs have comparable predictive power and may be another type of useful biomarker for the clinical evaluation of sperm quality(Figure 10B).

We found that rsRNAs is the most highly expressed sncRNAs in human sperm. A total of 10707 rsRNAs (average RPM > 100) were detected in sperm samples(Table S7), of which 70 rsRNAs were differently expressed between the two groups(42 up-regulated, 28 down-regulated) (Figure 9A, 9B, 9C and Table S8). Figure 9D shows the difference in the expression of the top 10 rsRNAs by average expression between the two groups. PCA classifier analyses showed that these 70 differently expressed rsRNAs can be used for separating the weak and normal group sperm samples(PC1=47.06%, PC2=18.13%) (Figure 10C).

Finally, we identified nine sncRNAs as candidate sperm quality biomarkers(Table 3). PCA classifier analyses showed that these nine sncRNAs can be used for better separating the weak and normal group sperm samples(PC1=34.02%, PC2=26.28%)(Figure 10D). In the future, we also need to verify whether these nine sncRNAs can be used as sperm quality biomarkers on the basis of larger sample size. True sncRNAs biomarkers not only require the ability to distinguish the high DFI sperm from the low DFI sperm, but more importantly, can accurately predict the pregnancy outcome.

Discussion

Studies in non-human mammals have shown that environmental factors influence sperm DNA methylation[36–41], which is also associated with diseases such as infertility and imprinting disorders in offspring. In recent years, more and more studies have explored the human sperm DNA methylation landscape, but the impact of environmental factors on human sperm DNA methylation is still unclear, and the relationship between human sperm DNA methylation and various sperm parameters and offspring health is also not clear.

In our study, genome-wide methylation sequencing of high DFI sperm was performed for the first time. We found that relative to low DFI sperm, the overall methylation level of high DFI sperm decreased; DMRs were mainly located in the promoter regions, and the number of hypermethylated DMRs was significantly higher than that of hypomethylated DMRs. We were able to separate sperm samples into two groups corresponding to DFI > 30% and DFI < 15% using the top 300 DMRs, thus these 300 DMRs may be potential biomarkers for sperm quality. At present, there is no clear sperm quality DNA methylation biomarker, our study may provide a basis for sperm quality DNA methylation biomarker research. The hypermethylated DMRs were found to be mainly signifcantly enriched in the area of neurons and microtubule which suggest that the increase of sperm DFI level may affect embryonic nervous system development and sperm motility.

We constructed the chromosome compartments of the two groups using the WGBS data, and found that the compartments of some chromosomes in the high DFI sperm changed compared with the low DFI sperm, and the correlations of methylation within some chromosomes were weakened. Therefore, the structure of the five chromosomes of the high DFI sperm may become loose, and the chromosomes is more vulnerable to ROS and more prone to break. Therefore, patients with high DFI should be treated in time or change their bad habits to avoid the continuous increase of DFI.

Previously, sperm RNA was thought to be a simple residual product of spermatogenesis, but studies over the past few years have shown that RNA is actually dynamically regulated during spermatogenesis and is spatially compartmentalized after sperm mature[42–44]. Studies have shown that both RNA profiles and RNA modifications in mammalian sperm are sensitive to the paternal environment. Various environmental factors, such as alcohol abuse, mental stress, chemical exposure, exercise and so on, can alter the composition of sperm miRNAs, tsRNAs, and rsRNAs. There are potential interactions between different kinds of RNAs in sperm and can synergistically regulate complex

biological processes[45], such as optimizing early embryonic development[46, 47] or transmitting paternal traits to offspring[20, 48–50].

We performed deep sequencing of sncRNAs in high DFI sperm for the first time, and found that there was no significant difference in the overall expression of various sncRNAs between high DFI and low DFI sperm. However, there were many differentially expressed miRNAs, tsRNAs and rsRNAs between the two groups of sperm, and PCA classifier analyses showed that they had some discriminative effect on the two groups of samples, so they may be potential sncRNAs biomarkers of sperm quality. We screened nine sncRNAs from the differentially expressed sncRNAs as candidate sperm quality biomarkers, and PCA classifier analyses showed that these 9 sncRNAs could well distinguish the two groups of sperm samples, but we need to verify these 9 sncRNAs on the basis of more sperm samples in the future

Conclusions

In conclusion, our study showed the genome-wide DNA methylation profiles and small non-coding RNAs signatures of sperm with high DNA fragmentation index. Our study provides useful biomarkers for improving the success rate of natural pregnancy and ART. Our findings suggest that the chromosomes of high DFI sperm became loose and more vulnerable to ROS attack. The increase of sperm DFI level may affect embryonic nervous system development.

Abbreviations

DFI
DNA fragmentation index
sncRNAs
Small non-coding RNAs
WGBS
Whole-genome bisulfite sequencing
ART
Assisted reproductive technology
DMRs
Diferentially methylated regions
ROS
reactive oxygen species
SDF
Sperm DNA fragmentation
OS
oxidative stress
miRNAs
microRNAs
tsRNAs
tRNA-derived small RNAs
rsRNAs
rRNA-derived small RNAs
piRNAs
Piwi-interacting RNAs
GO
Gene ontology
PCA

Principal component analysis
PBS
Phosphate buffer saline
SCLB
Somatic Cell Lysis Buffer.

Declarations

Ethical Approval and Consent to participate

This study was approved by the ethics committee of Naval Medical University(No.chrec-2018-6) and has been performed in accordance with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of supporting data

[We intend to deposit the sequencing data in the National Genomics Data Center (NGDC) (https://ngdc.cncb.ac.cn/?lang=zh). We are awaiting database review, which may take a few days.]

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the the National Key R&D Program of China(2018YFC1005002).

Authors’ contributions

HY, ML and JZ designed the study. NW and LQ collected sperm samples. BX prepared sncRNAs library. PL and YC performed the data analysis. ML drafted the manuscript. HY, LW and YL critically review and corrected the final version of the manuscript. All authors approved the final version to be submitted.

Acknowledgments

The authors thank the staff from Reproductive Medical Center of Changhai Hospital for their assistance with the sperm samples collection. The authors also thank all participants in this study.

Authors' information

[1]Reproductive Medical Center, Changhai Hospital, Naval Medical University, Shanghai, China. [2]State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, Shanghai, China. [3]State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China

References

1. Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, et al. The international glossary on infertility and fertility care. Fertil Steril. 2017;108:39-406.
2. Choy JT, Eisenberg ML. Male infertility as a window to health. Fertil Steril. 2018;110:810-4.
3. Wang C, Swerdloff RS. Limitations of semen analysis as a test of male fertility and anticipated needs from newer tests. Fertil Steril. 2014;102:1502-7.
4. Agarwal A, Allamaneni SS. Sperm DNA damage assessment: a test whose time has come. Fertil Steril. 2005;84:850-3.
5. Guzick DS, Overstreet JW, Factor-Litvak P, Brazil CK, Nakajima ST, Coutifaris C, et al.; National Cooperative Reproductive Medicine Network. Sperm morphology, motility, and concentration in fertile and infertile men. N Engl J Med. 2001;345:1388-93.
6. Baskaran S, Cho CL, Agarwal A. Role of sperm DNA damage in male infertility assessment. In: Rizk B, Agarwal A, Sabanegh ES Jr, editors. Male infertility in reproductive medicine: diagnosis and management. Boca Raton CRC Press; 2019;205.
7. Agarwal A, Prabakaran S, Allamaneni S. What an andrologist/urologist should know about free radicals and why. Urology. 2006;67:2-8.
8. Aitken RJ, Curry BJ. Redox regulation of human sperm function: from the physiological control of sperm capacitation to the etiology of infertility and DNA damage in the germ line. Antioxid Redox Signal. 2011;14:367-81.
9. Baskaran S, Agarwal A, Panner Selvam MK, Finelli R, Robert KA, Iovine C, et al. Tracking research trends and hotspots in sperm DNA fragmentation testing for the evaluation of male infertility: a scientometric analysis. Reprod Biol Endocrinol. 2019;17:110.
10. Aitken RJ. DNA damage in human spermatozoa; important contributor to mutagenesis in the offspring. Transl Androl Urol. 2017;6(Suppl 4):S761-4.
11. Marques CJ, Carvalho F, Sousa M, Barros A. Genomic imprinting in disruptive spermatogenesis. Lancet. 2004;363:1700-1702.
12. Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, Barros A, Sousa M. Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. Mol Hum Reprod. 2008;14:67-73.
13. Dong H, Wang YX, Zou ZK, Chen LM, Shen CY, Xu SQ, Zhang J, Zhao FF, Ge SQ, Gao Q, et al. Abnormal methylation of imprinted genes and cigarette smoking: assessment of their association with the risk of male infertility. Reprod Sci. 2017;24:114-123.
14. Poplinski A, Tuttelmann F, Kanber D, Horsthemke B, Gromoll J. Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. Int J Androl. 2010;33:642-649.
15. Klaver R, Tuttelmann F, Bleiziffer A, Haaf T, Kliesch S, Gromoll J. DNA methylation in spermatozoa as a prospective marker in andrology. Andrology. 2013;1:731-740.
16. Fredrika Asenius, Amy F. Danson, Sarah J. Marzi, DNA methylation in human sperm: a systematic review, Human Reproduction Update. 2020;26:841-873.
17. Barres R & Zierath JR The role of diet and exercise in the transgenerational epigenetic landscape of T2DM. Nat. Rev. Endocrinol. 2016;12:441-451.
18. Sales VM, Ferguson-Smith AC & Patti ME Epigenetic mechanisms of transmission of metabolic disease across generations. Cell Metab.2017;25:559-571.
19. Fleming TP et al. Origins of lifetime health around the time of conception: causes and consequences. Lancet.2018;391:1842-1852.
20. Chen Q et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science.2016;351:397-400.

21. Zhang Y et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. Nat Cell Biol.2018;20:535–540.
22. Peng H et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. Cell Res. 2012;22:1609-1612.
23. Sharma U et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science. 2016;351:391-396.
24. Chu C et al. A sequence of 28S rRNA-derived small RNAs is enriched in mature sperm and various somatic tissues and possibly associates with inflammation. J. Mol. Cell Biol.2017;9: 256-259.
25. Shi J et al. SPORTS1.0: a tool for annotating and profiling non-coding RNAs optimized for rRNAand tRNA-derived small RNAs. Genomics Proteomics Bioinformatics. 2018;16: 144-151.
26. Krueger F, Andrews SR. Bismark: a fexible aligner and methylation caller for Bisulfte-Seq applications. Bioinformatics. 2011;27:1571-2.
27. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfte sequencing reads to diferentially methylated regions. Genome Biol. 2012;13:R83.
28. Jean-Philippe Fortin, Kasper D. HansenReconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data, Genome Biology. 2015;16:180.
29. Gordon, A. & Hannon, G., J. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit. 2010.
30. Rosenkranz, D. & Zischler, H. proTRAC-a software for probabilistic piRNA cluster detection, visualization and analysis. BMC bioinformatics. 2012; 13:5.
31. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic acids research.2011; 39.
32. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15.
33. Kolde, R. Pheatmap: pretty heatmaps. R package version. 2012;61.
34. Vu, V. Q. ggbiplot: A ggplot2 based biplot. R package. 2011.
35. Barrell, D. et al. The GOA database in 2009-an integrated Gene Ontology Annotation resource. Nucleic Acids Res.2009;37:D396-403.
36. Dias BG, Ressier KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci. 2014;17:89-96.
37. Radford EJ, Ito M, Shi H, Corish JA, Yamazawa K, Isganaitis E, Seisenberger S, Hore TA, Reik W, Erkek S, et al. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. Science. 2014;345:785.
38. Barbosa TD, Ingerslev LR, Alm PS, Versteyhe S, Massart J, Rasmussen M, Donkin I, Sjogren R, Mudry JM, Vetterli L, et al. High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring. Mol Metab. 2016;5:184-197.
39. Youngson NA, Lecomte V, Maloney CA, Leung P, Liu J, Hesson LB, Luciani F, Krause L, Morris MJ. Obesity-induced sperm DNA methylation changes at satellite repeats are reprogrammed in rat offspring. Asian J Androl. 2016;18:930-936.
40. Sakai K, Ideta-Otsuka M, Saito H, Hiradate Y, Hara K, Igarashi K, Tanemura K. Effects of doxorubicin on sperm DNA methylation in mouse models of testicular toxicity. Biochem Biophys Res Commun. 2018;498:674-679.
41. Watkins AJ, Dias I, Tsuro H, Allen D, Emes RD, Moreton J, Wilson R, Ingram RJM, Sinclair KD. Paternal diet programs offspring health through sperm- and seminal plasma-specific pathways in mice. Proc Natl Acad Sci USA.

2018;115:10064-10069.
42. Johnson GD et al. Chromatin and extracellular vesicle associated sperm RNAs. Nucleic Acids Res. 2015;43:6847-6859.
43. Peng H et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. Cell Res. 2012;22:1609-1612.
44. Sharma U et al. Small RNAs are trafficked from the epididymis to developing mammalian sperm. Dev Cell. 2018;46:481-494.
45. Schimmel P The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nat Rev Mol Cell Biol. 2018;19:45-58.
46. Guo L et al. Sperm-carried RNAs play critical roles in mouse embryonic development. Oncotarget. 2017; 8:67394-67405.
47. Yuan S et al. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development. 2016;143:635-647.
48. Gapp K et al. Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. Mol. Psychiatry 10.1038/s41380-018-0271-6, 2018.
49. Grandjean V et al. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. Sci Rep. 2015;5:18193.
50. Gapp K et al. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci. 2014;17:667-669.

Tables

Table 1 Sample characteristics

	Samples for WGBS[a]		Samples for sncRNAs[b] sequencing	
Characteristics	**Weak(n=6)**	**Normal(n=7)**	**Weak(n=13)**	**Normal(n=17)**
Age(years)	35.8±4.6	36.1±2.4	34.7±4.4	31.4±2.9
DFI[c](%)***	46.6±12.5	6.7±2.1	36.3±4.5	8±1.8
Semen volume (mL)	2.6±0.7	2.7±0.4	2.9±1.2	2.7±0.8
Motility (%)***	21.2±9.0	63.1±10.7	29.8±3.2	62.2±5.6
Progressive motility (%)***	13.5±7.0	49.0±9.4	21±9.2	50.9±6.4
Concentration ($\times 10^6$/mL)	78.3±84.6	92.3±43.6	74±78.2	99±44.5

[a] *WGBS* whole genome bisulfte sequencing

[b] *sncRNAs* small non-coding RNAs

[c] *DFI* DNA fragmentation index

Data represent the means±SD

*** $P < 0.001$

Table 2 Percentages of different types of sncRNAs expressed in 30 human sperm samples

	Min ratio[a]	Max ratio[b]	Mean ratio[c]
miRNA	2.28%	19.39%	7.06%
tRNA	5.01%	40.36%	19.26%
rRNA	28.16%	63.66%	40.54%
sn/snoRNA	0.03%	1.50%	0.58%
piRNA	0.03%	1.46%	0.56%
yRNA	1.34%	45.79%	10.39%
other	3.60%	34.41%	21.62%

[a]*Min ratio,* [b]*Max ratio,* [c]*Mean ratio* : the minimum, maximum and average ratio of each type of sncRNAs expressed in 30 human sperm samples.

Table 3 Informations of the 9 candidate sperm quality sncRNAs biomarkers

sncRNAs[a] name	Log_2Foldchange[b]	Average mean[c]	P-value	Sequence
Gly-GCC-32-1	0.670161948	14497	0.0404	GCATGGGTGGTTCAGTGGTAGAATTCTCGCCT
Gly-GCC-31-2	0.757394615	13836	0.0073	GCATTGGTGGTTCAGTGGTAGAATTCTCGCC
Gly-GCC-33-3	1.069361698	1764	0.0013	GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTG
hsa-miR-151a-5p	-0.622466835	1414	0.0484	CCCTCGAGGAGCTCACAGTCTAGTGGG
SeC-TCA-37-4	-0.594019477	1086	0.0116	GCCCGGATGATCCTCAGTGGTCTGGGGTGCAGGCTTC
hsa-let-7b-5p	0.649849594	845	0.0178	GGGTGAGGTAGTAGGTTGTGTGGTTTCA
hsa-miR-30c-5p	-0.910409754	612	0.0198	GTGTGTAAACATCCTACACTCTCAGCTGT
Glu-CTC-40-10	-1.32067563	578	$8.44*10^{-5}$	TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCG
iMet-CAT-18-18	2.105048654	157	$2.67*10^{-7}$	AGCGGAAGCGTGCTGGGC

[a]*sncRNAs* small non-coding RNAs

[b]*$Log_2 Foldchange$* Log_2 the average expression level of sncRNAs in the weak group sperm samples / the average expression level of sncRNAs in normal group sperm samples

[c]*Average mean* average sncRNA expression in 30 sperm samples

Figures

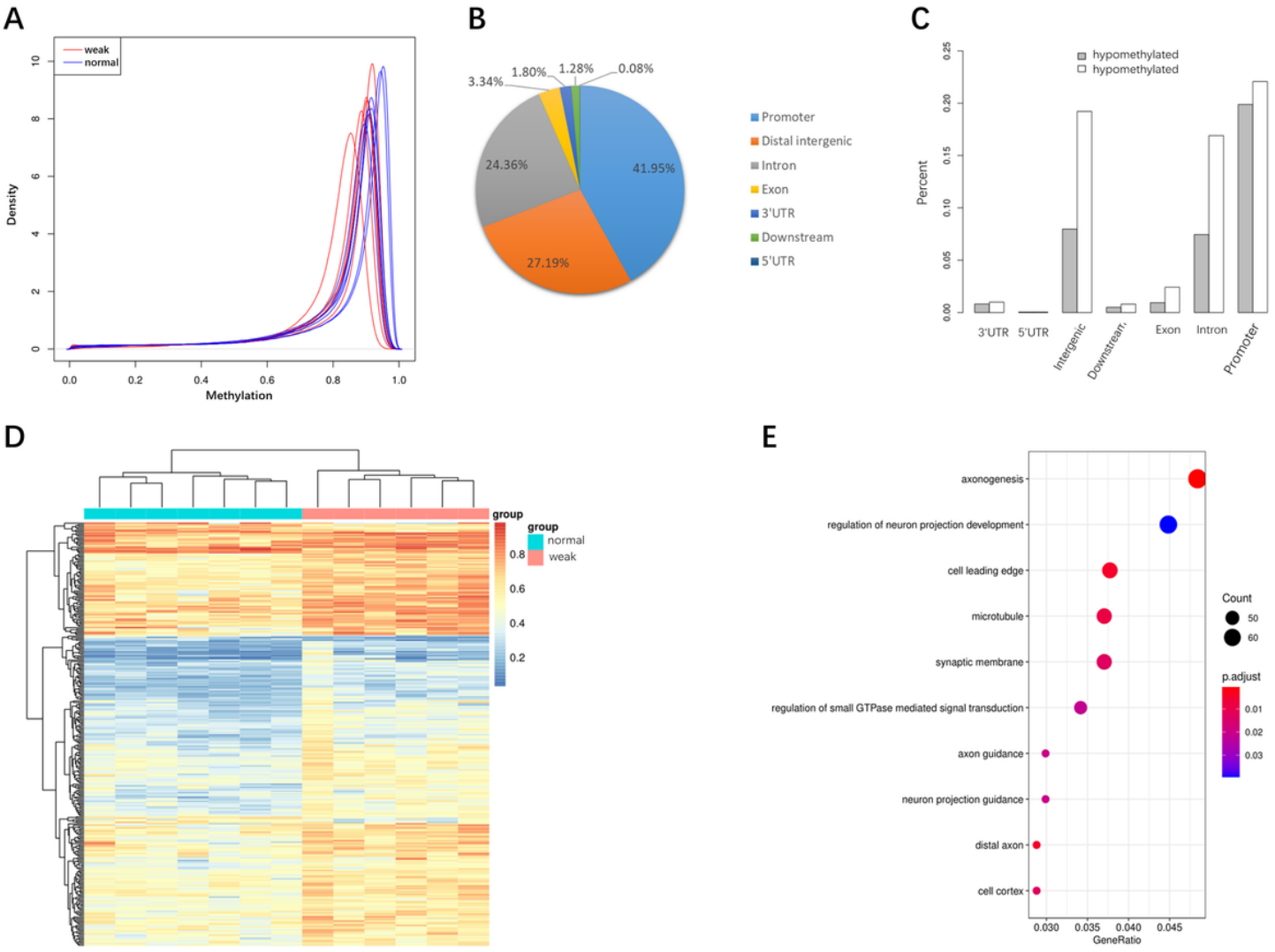

Figure 1

DNA methylation differences between the weak and normal group sperm samples.

(A)Compared with the normal group sperm samples, the global DNA methylation level of the weak group showed a downward trend.

(B) DMR percentages in different genomic components (promoters were defined as 1~3000 bp upstream or downstream of transcription start site).

(C) Hypermethylated and hypomethylated DMR percentages in different genomic components.

(D) Supervised clustering with the top 300 DMRs.

(E) GO enrichment analysis with a focus on biological processes for the hypermethylated DMRs.

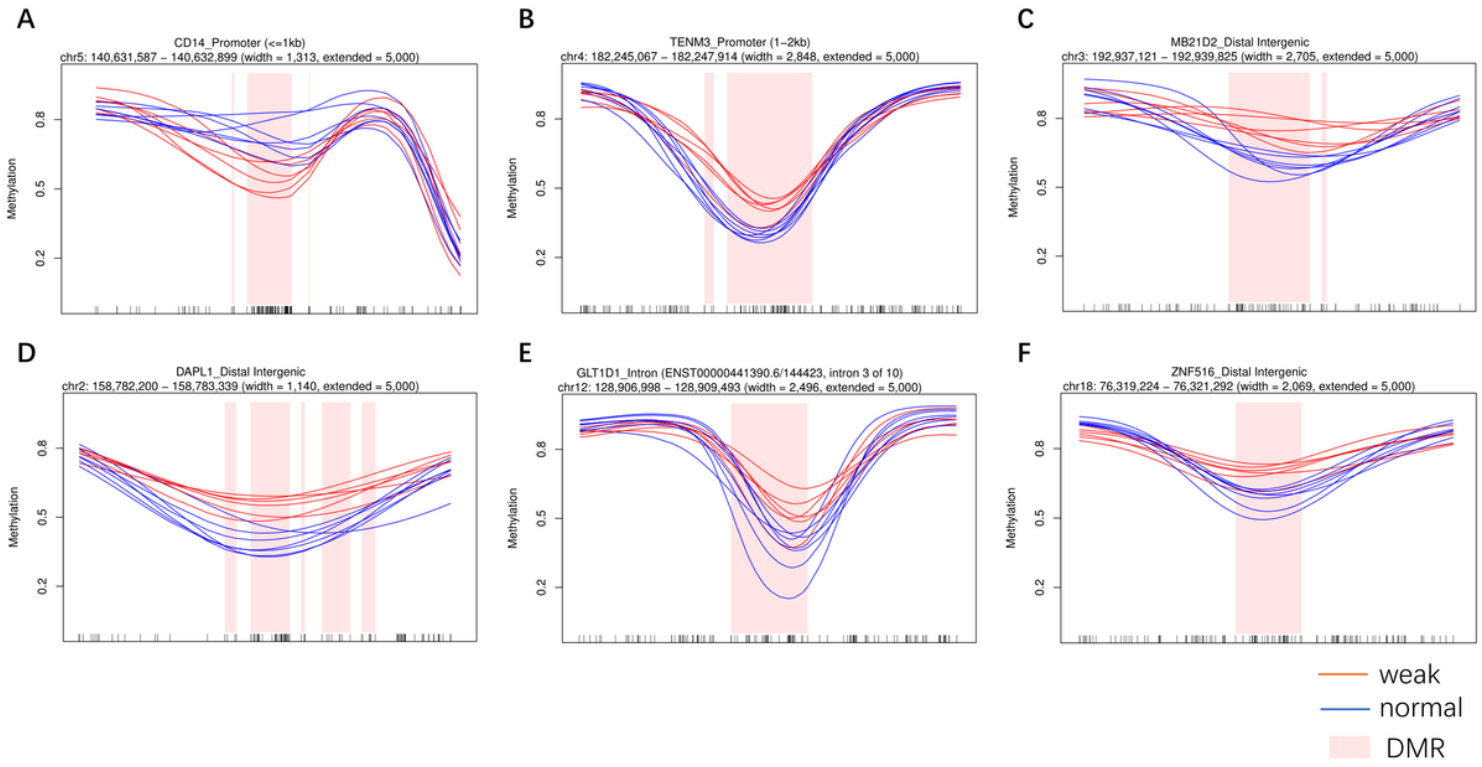

Figure 2

Representative DMRs identifed by WGBS. Smoothed methylation values were plotted with identifed DMRs indicated in pink.

Plots of the (A) CD14 cluster, (B) TENM3 cluster, (C) MB21D2, (D) DAPL1 cluster, (E) GLT1D1 cluster and (F) ZNF516 cluster.

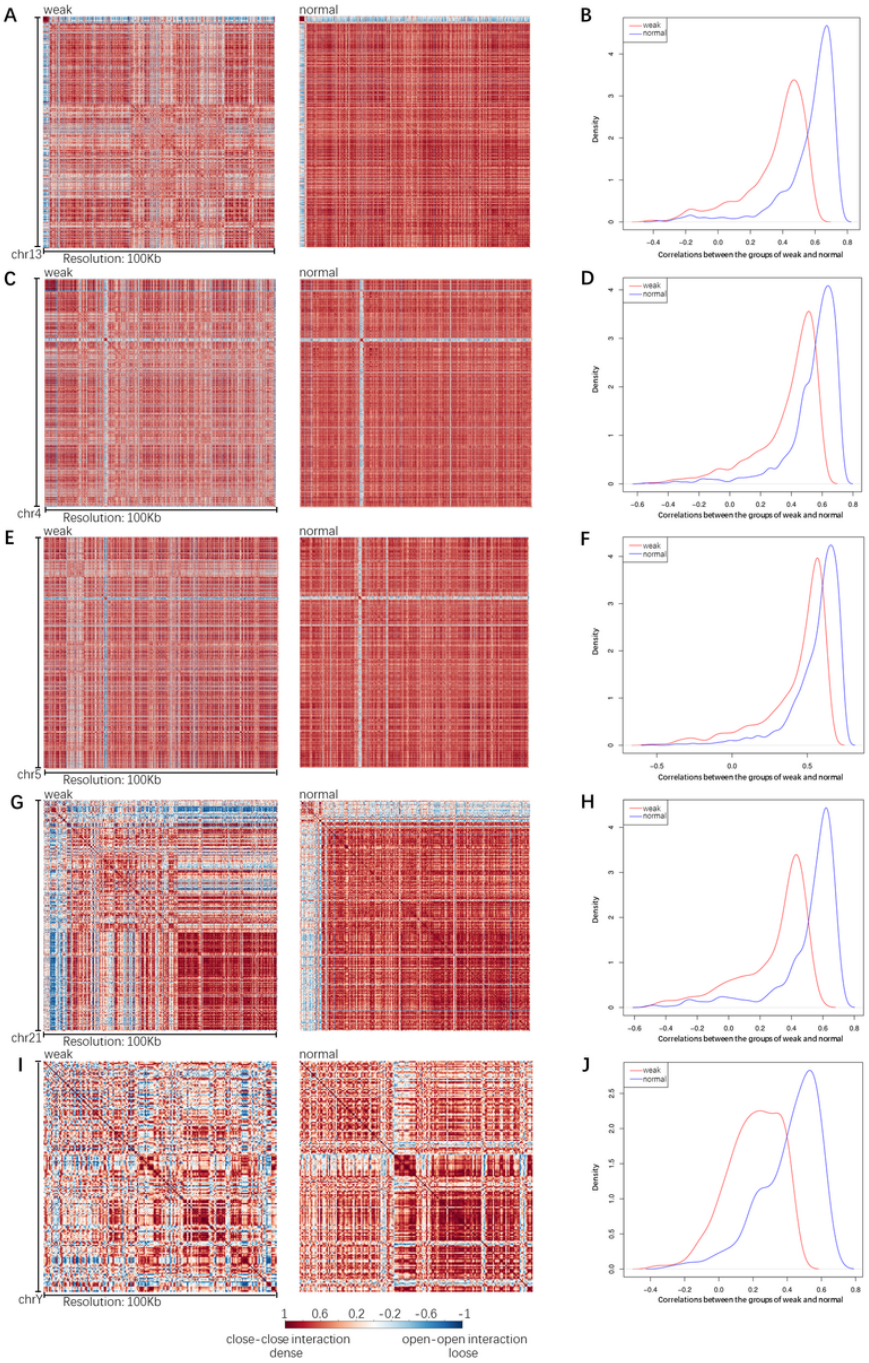

Figure 3

Chromosome compartments of the two groups of sperm samples constructed using the WGBS data.

The compartments of chromosomes 13(A), 4(C), 5(E), 21(G) and Y(I) in the weak group sperm samples changed compared with the normal group. The correlations of methylation within the chromosomes 13(B), 4(D), 5(F), 21(H) and Y(J) were weakened in the weak group sperm samples compared with the normal group.

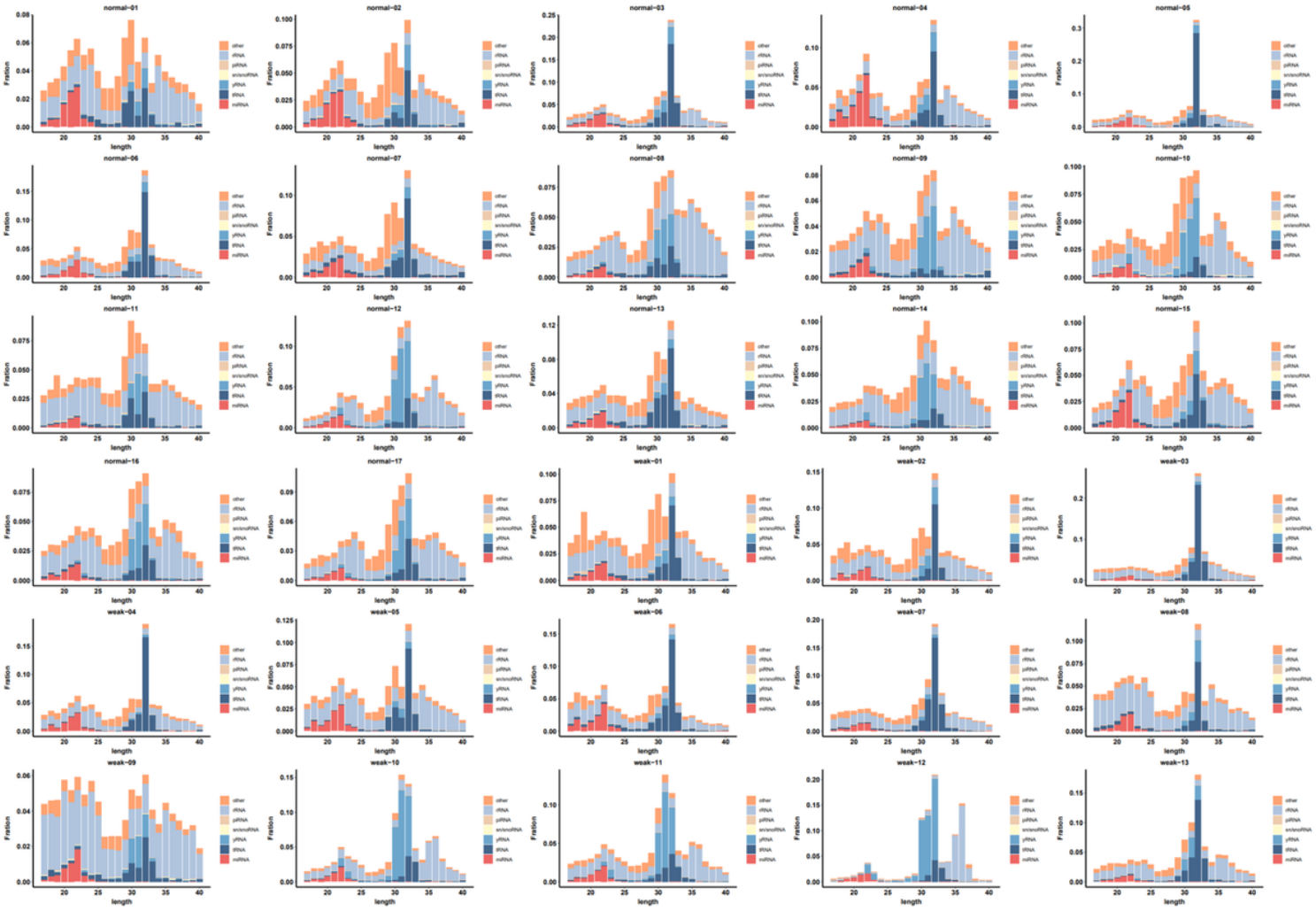

Figure 4

Length distribution of sncRNAs in 30 human sperm samples 17 normal group sperm samples and 13 weak group sperm samples

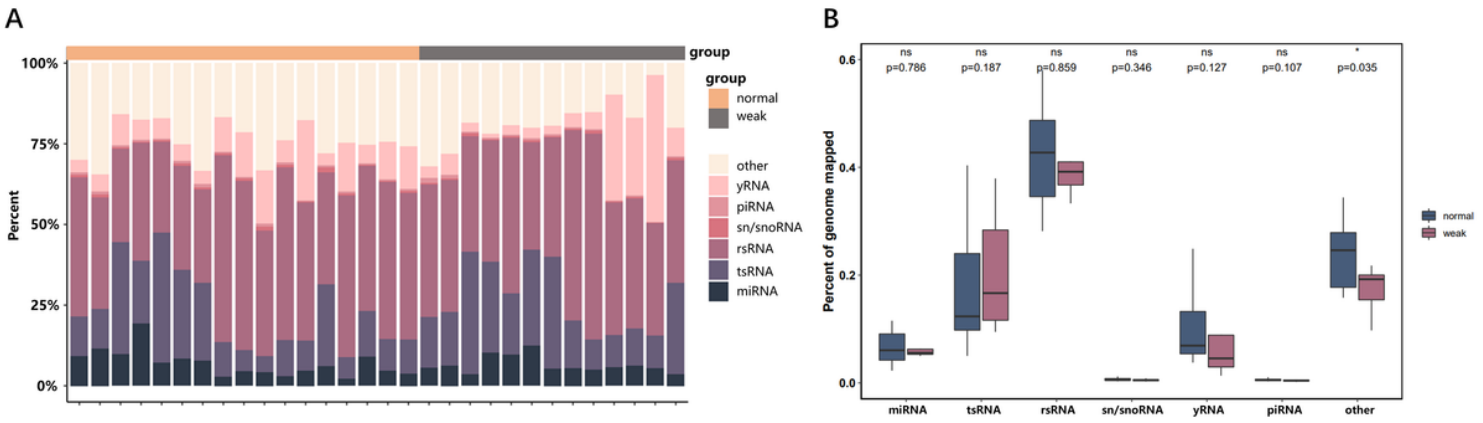

Figure 5

Comparative analysis of sncRNAs in 30 human sperm samples.

(A) Composition of sncRNA categories in 30 human sperm samples.

(B) Boxplot of the percentage of miRNA, tRNA, rRNA, sn/snoRNA, yRNA piRNA and other sncRNA in the weak and normal group sperm samples.

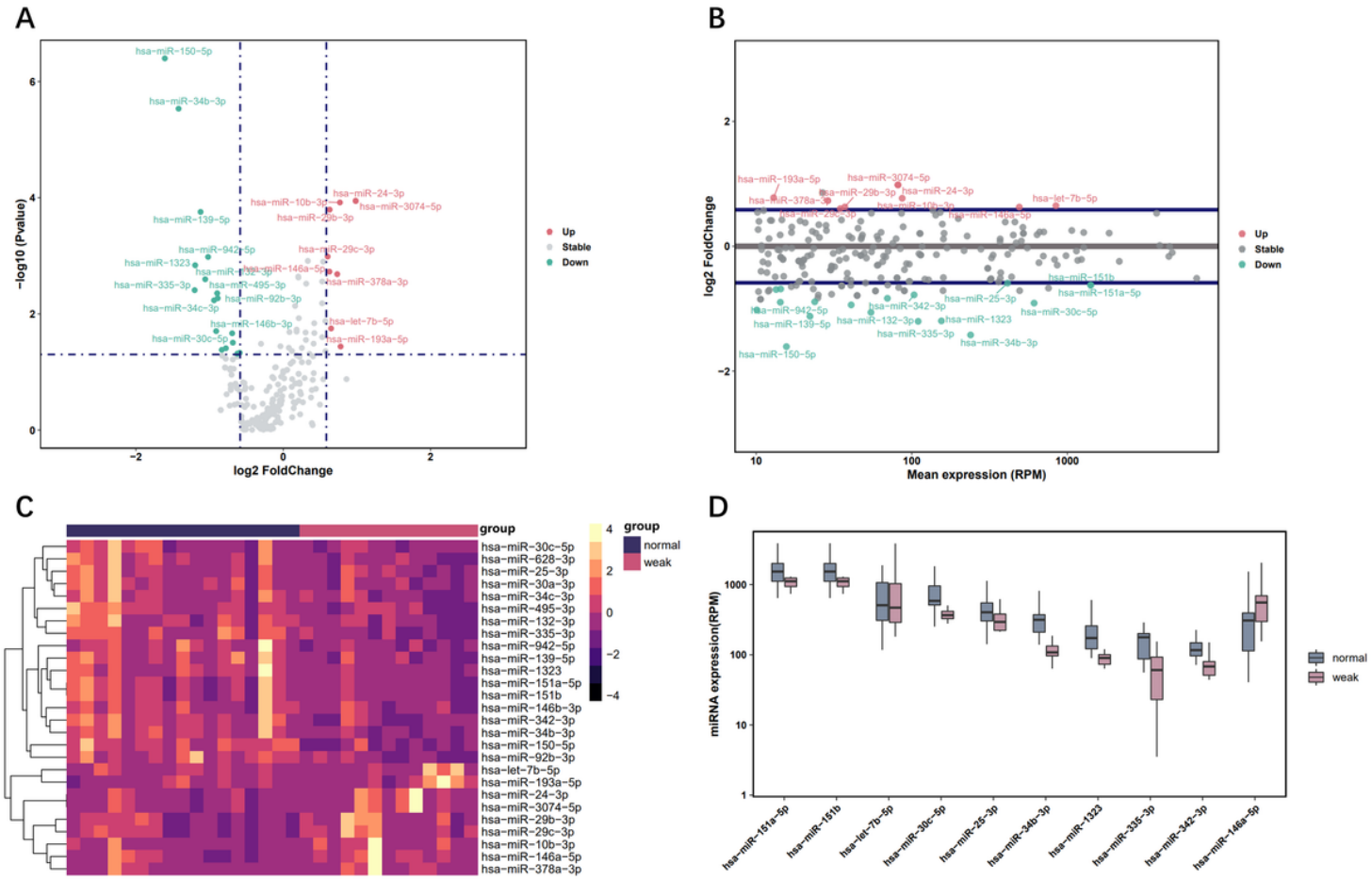

Figure 6

Differential expression miRNAs between the weak and normal group sperm samples analysis.

(A) Valcano plot of the 27 differential expression miRNAs.

(B) Maplot of the 27 differential expression miRNAs.

(C) Heatmap of the 27 differentially expressed miRNAs, The branching pattern is illustrated using a dendrogram.

(D) Bar plot of the top ten differentially expressed miRNAs in average expression.

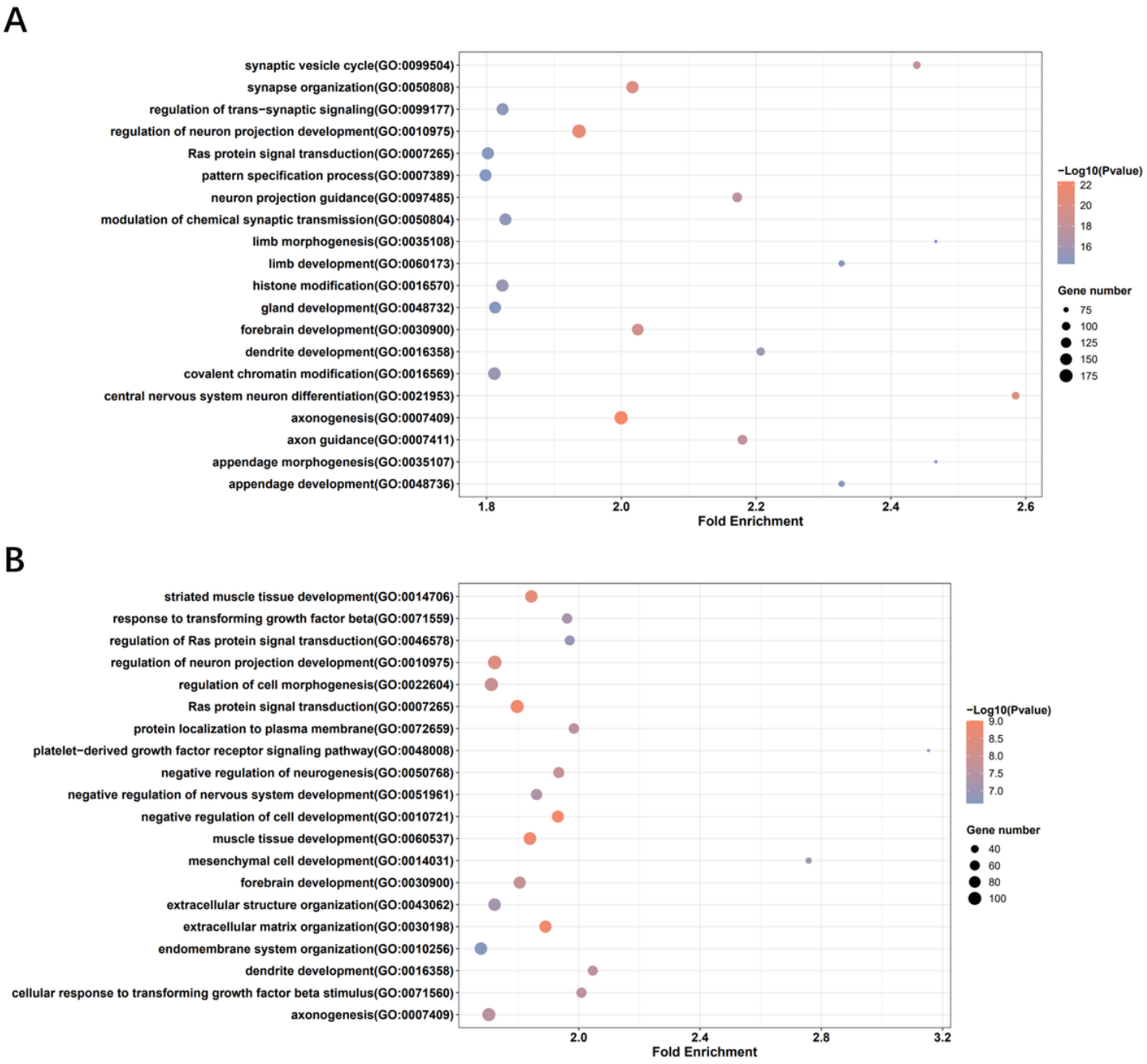

Figure 7

GO analysis of the predicted genes targeted by the differentially expressed miRNAs.

(A) Bar plot showing the GO term categories of the predicted genes targeted by the 10 downregulated miRNAs miR-150-5p, miR-139-5p, miR-132-3p, miR-30c-5p, miR-495-3p, miR-92b-3p, miR-342-3p, miR-151a-5p, miR-151b and miR-25-3p .

(B) Bar plot showing the GO term categories of the predicted genes targeted by the 6 upregulated miRNAs(miR-29c-3p, miR-146a-5p, miR-29b-3p, miR-378a-3p, miR-193a-5p and miR-24-3p). The top 20 GO term categories with the highest fold enrichment are shown.

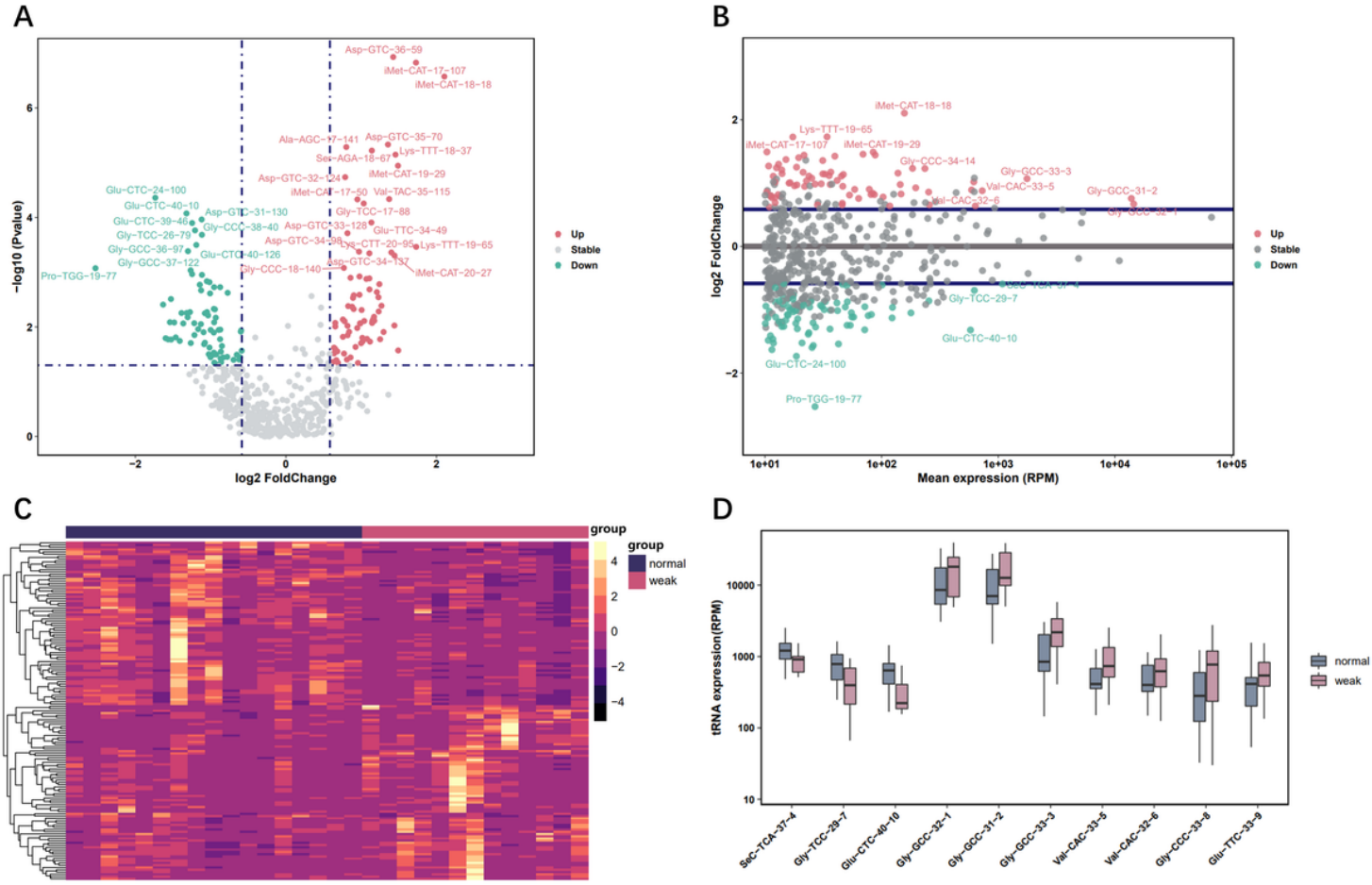

Figure 8

Differential expression tsRNAs between the weak and normal group sperm samples analysis.

(A) Valcano plot of the 151 differential expression tsRNAs.

(B) Maplot of the 151 differential expression tsRNAs.

(C) Heatmap of the 151 differentially expressed tsRNAs. The branching pattern is illustrated using a dendrogram.

(D) Bar plot of the top ten differentially expressed tsRNAs in basic expression.

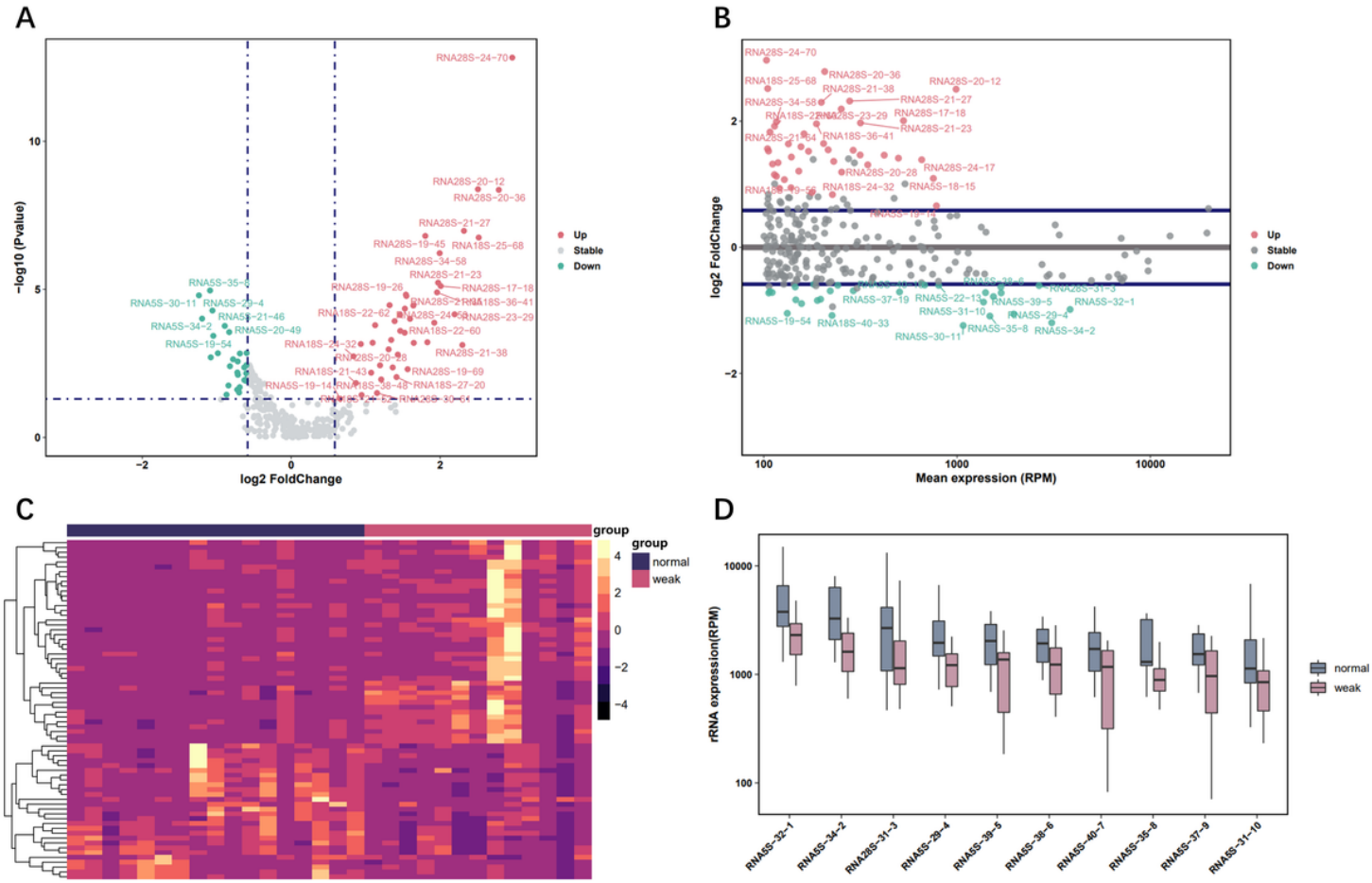

Figure 9

Differential expression rsRNAs between the weak and normal group sperm samples analysis.

(A) Valcano plot of the 70 differential expression rsRNAs.

(B) Maplot of the 70 differential expression rsRNAs.

(C) Heatmap of the 70 differentially expressed rsRNAs. The branching pattern is illustrated using a dendrogram.

(D) Bar plot of the top ten differentially expressed rsRNAs in basic expression.

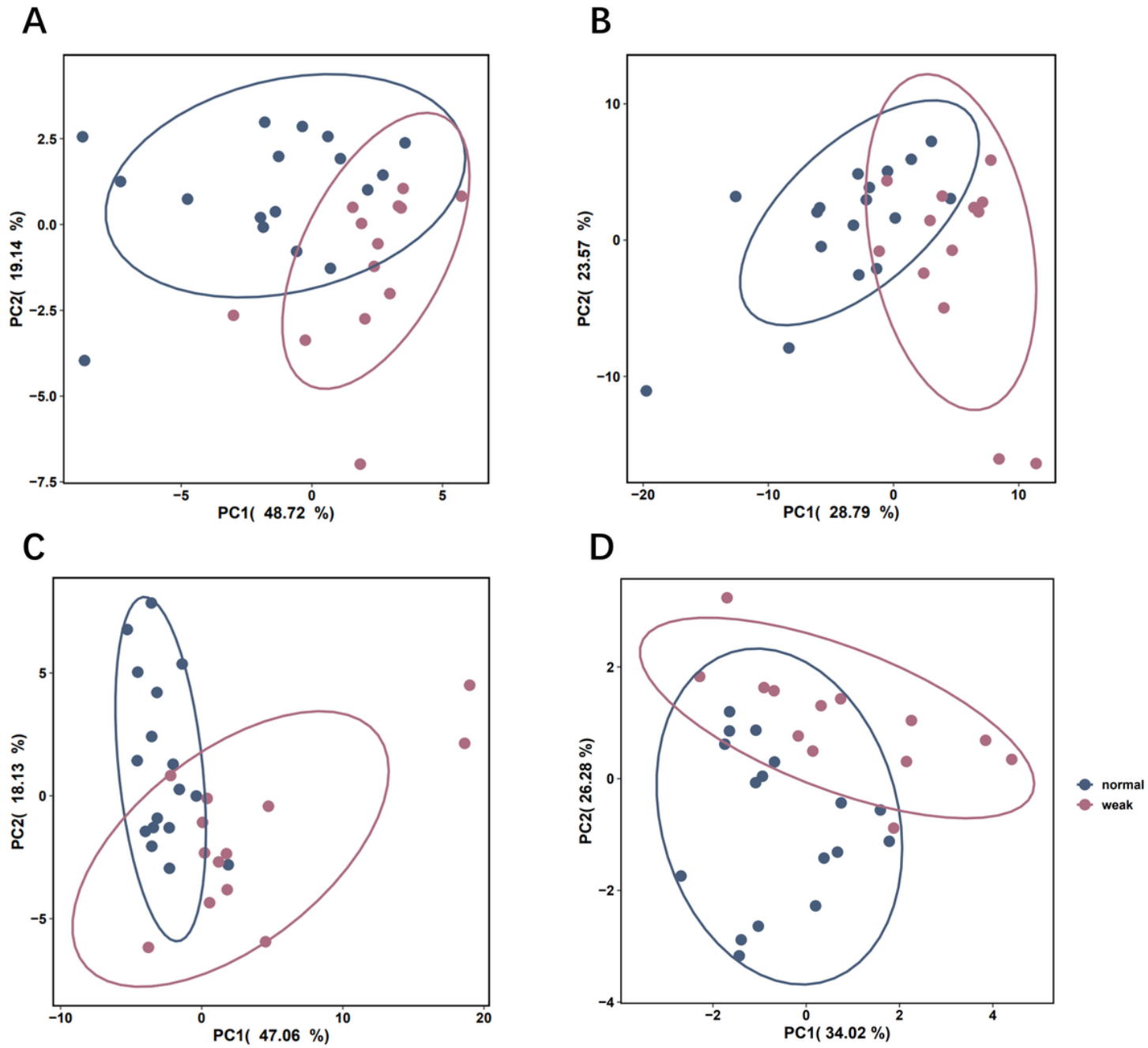

Figure 10

PCA of the sperm samples in the weak and normal groups based on differentially expressed sncRNAs.

(A) PCA of the sperm samples in the two groups based on 27 differentially expressed miRNAs. The variation values of PC1 and PC2 are 48.72% and 19.14%, respectively.

(B) PCA of the sperm samples in the two groups based on 151 differentially expressed tsRNAs. The variation values of PC1 and PC2 are 28.79% and 23.57%, respectively.

(C) PCA of the sperm samples in the two groups based on 70 differentially expressed rsRNAs. The variation values of PC1 and PC2 are 47.06% and 18.13%, respectively.

(D) PCA of the sperm samples in the two groups based on 9 candidate sperm quality sncRNAs biomarkers. The variation values of PC1 and PC2 are 34.02% and 26.28%, respectively. Points represent PCA scores of individual samples. Circles represent a general characterization of the PCA space occupied by the differentially expressed sncRNAs. Sperm samples from the two groups are shown with different colors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.xlsx