

RESEARCH ARTICLE

Open Access



# Transcriptome profiling of developing testes and spermatogenesis in the Mongolian horse

Bei Li<sup>1,2,3</sup>, Xiaolong He<sup>4</sup>, Yiping Zhao<sup>1,2,3</sup>, Dongyi Bai<sup>1,2,3</sup>, Ming Du<sup>1,2,3</sup>, Lianjie Song<sup>1,2,3</sup>, Zhuang Liu<sup>1,2,3</sup>, Zhenchen Yin<sup>1,2,3</sup> and Dugarjaviin Manglai<sup>1,2,3\*</sup>

## Abstract

**Background:** Horse testis development and spermatogenesis are complex physiological processes.

**Methods:** To study these processes, three immature and three mature testes were collected from the Mongolian horse, and six libraries were established using high-throughput RNA sequencing technology (RNA-Seq) to screen for genes related to testis development and spermatogenesis.

**Results:** A total of 16,237 upregulated genes and 8,641 downregulated genes were detected in the testis of the Mongolian horse. These genes play important roles in different developmental stages of spermatogenesis and testicular development. Five genes with alternative splicing events that may influence spermatogenesis and development of the testis were detected. GO (Gene ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analyses were performed for functional annotation of the differentially expressed genes. Pathways related to “spermatogenesis,” male gamete generation,” “spermatid development” and “oocyte meiosis” were significantly involved in different stages of testis development and spermatogenesis.

**Conclusion:** Genes, pathways and alternative splicing events were identified with inferred functions in the process of spermatogenesis in the Mongolian horse. The identification of these differentially expressed genetic signatures improves our understanding of horse testis development and spermatogenesis.

**Keywords:** Mongolian horse, Developing testis, Spermatogenesis, RNA-Seq

## Background

The mammalian testis is an important male reproductive organ that produces sperm and androgens. Spermatogenesis is a complex developmental process with a unique mechanism of cell division—meiosis—as a major defining event. Sperm development consists of three stages: (1) proliferation and differentiation of spermatogonia, (2)

meiosis of spermatocytes, and (3) sperm maturation. Spermatogonial stem cells differentiate into spermatogonia and, with replication of their DNA content, give rise to primary spermatocytes. Secondary spermatocytes and spermatids are derived from the primary spermatocytes by DNA replication and tetraploid meiosis. Haploid sperm cells undergo at least four distinct morphological changes into sperm cells, namely chromatin condensation, acrosome formation, flagellar formation, and cytoplasmic reduction. In addition to spermatogenic cells, spermatogenesis also involves multiple somatic cells in testicular tissues, such as Sertoli and Leydig cells. In mammalian testes, spermatogenesis occurs within seminiferous

\* Correspondence: [dmanglai@163.com](mailto:dmanglai@163.com)

<sup>1</sup>College of Animal Science, Equine Research Center, Inner Mongolia Agricultural University, Hohhot 010018, China

<sup>2</sup>Inner Mongolia Key Laboratory of Equine Genetics, Breeding and Reproduction, Equine Research Center, Inner Mongolia Agricultural University, Hohhot 010018, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

tubules, where germ cells associate with Sertoli cells, which provide the appropriate microenvironment for spermatogenesis. Related genes in these cells play important roles in specific stages of germ cell development, and the transcription and translation of these genes vary in different stages of spermatogenesis [1, 2].

RNA sequencing technology (RNA-Seq) technology, which provides a global view of gene expression profiles, has been developed for mapping and quantifying transcriptomes [3, 4]. This methodology has several advantages over other transcriptome technologies, such as higher resolution and sensitivity, a large dynamic range of gene expression levels, and the ability to identify novel transcribed regions and splicing isoforms of known genes [5]. Ramsköld [6] assessed different mammalian tissues by RNA-Seq and found that the vast majority of genes are specifically expressed in the testis. Furthermore, Djureinovic [7] categorized 20,050 putative human genes based on RNA-Seq expression patterns. These genes were specifically expressed in normal human testicular tissues compared to 26 other normal human tissue types in seven individuals. The analysis showed that the testis has the highest number of tissue-specific genes by far. Anand [8] identified differentially expressed genes (DEGs) in the human testis relative to other tissues by microarray analysis, and a total of 2,868 upregulated and 2,011 downregulated transcripts were identified. The testis thus seems to be highly metabolically active relative to other normal tissues as indicated by functional annotation. Currently, most of the studies on the relationship between the testis development and spermatogenesis have focused on humans or mice. However, little is known regarding the spermatogenesis process in the horse.

The Mongolian horse is one of the most famous native breeds in China, and in the world. It is a dual-purpose breed that can adapt to harsh climate and extensive breeding conditions. Normal development of the testis and spermatogenesis is extremely important to ensure high level semen production and protect population continuity, and genes that are important for different stages of testis development and spermatogenesis are likely to have roles in maintaining fertility. Unfortunately, the transcriptional expression of genes during testicular development and spermatogenesis in the horse is not well understood, thereby limiting the understanding of the mechanisms of deformation and maturation in spermatogenesis. This study aimed to study the transcriptome profiles of immature and mature male Mongolian horse testes utilizing high-throughput RNA-Seq technology and bioinformatics analysis. For

this purpose, a testis transcript database was obtained, and DEGs and regulatory pathways were analyzed. This study provides new insights into gene expression and the regulatory mechanisms of horse testis development and spermatogenesis.

## Results

### mRNA expression profiles in Mongolian horse testis development

To screen for genes related to testis development and spermiogenesis, a total of six libraries—three from immature colts (BSM1, BSM2, and BSM3) and three from mature stallions (ASM1, ASM2, and ASM3)—were constructed and sequenced on the Illumina HiSeq platform. An overview of the information for these libraries is listed in Table S1. The libraries were found to contain 58,277,598 (BSM1), 59,921,568 (BSM2), 53,371,478 (BSM3), 59,523,578 (ASM1), 61,727,216 (ASM2), and 53,956,064 (ASM3) raw reads. To ensure the quality of subsequent analysis, the adapter reads, the N base > 10% and low-quality reads were removed; the results were 57,482,858 (BSM1), 58,972,988 (BSM2), 52,475,320 (BSM3), 58,411,056 (ASM1), 60,775,868 (ASM2), and 53,135,736 (ASM3) clean reads. The error rate and GC content of each library were calculated to control for the quality of libraries (Table S1). These results verify that the six libraries were of high quality.

The clean reads were aligned to the reference genome of *E. caballus*, which is the most accurate and commonly used reference genome for horses [9, 10], using TopHat v2.0.12. In total, 84.85% (BSM1), 85.05% (BSM2), 84.25% (BSM3), 84.38% (ASM1), 84.14% (ASM2), and 84.23% (ASM3) clean reads were uniquely mapped to the reference genome. In addition, more than 40% of clean reads were non-splice reads in each sample. Additionally, 27.98% (BSM1), 27.35% (BSM2), 26.92% (BSM3), 36.17% (ASM1), 34.89% (ASM2), and 34.74% (ASM3) of the clean reads were mapped to the borders of exons (also called “junction reads”) (Table S2). All RNA-Seq data have been submitted to The Gene Expression Omnibus, accession number GEO: GSE101697.

### Analysis of alternative splicing

Alternative splicing (AS) is a common occurrence in most eukaryotic cells and can result in the translation of different forms of proteins at various time points and environments, which increases the adaptability and physical fitness of the species. In our study, the AS events were classified into five kinds by rMATS (<http://rnaseq-mats.sourceforge.net/index.html>): (1) SE: Skipped exon, (2) MXE: Mutually exclusive exon, (3) A5SS: Alternative 5' splice site, (4) A3SS: Alternative

3' splice site, and (5) RI: Retained intron. The kinds and quantities of AS events were counted, and the expression of each type of AS event was analyzed. The results indicate that numerous and varied AS events exist in the spermatogenesis process (Table S3).

**Analysis of DEGs**

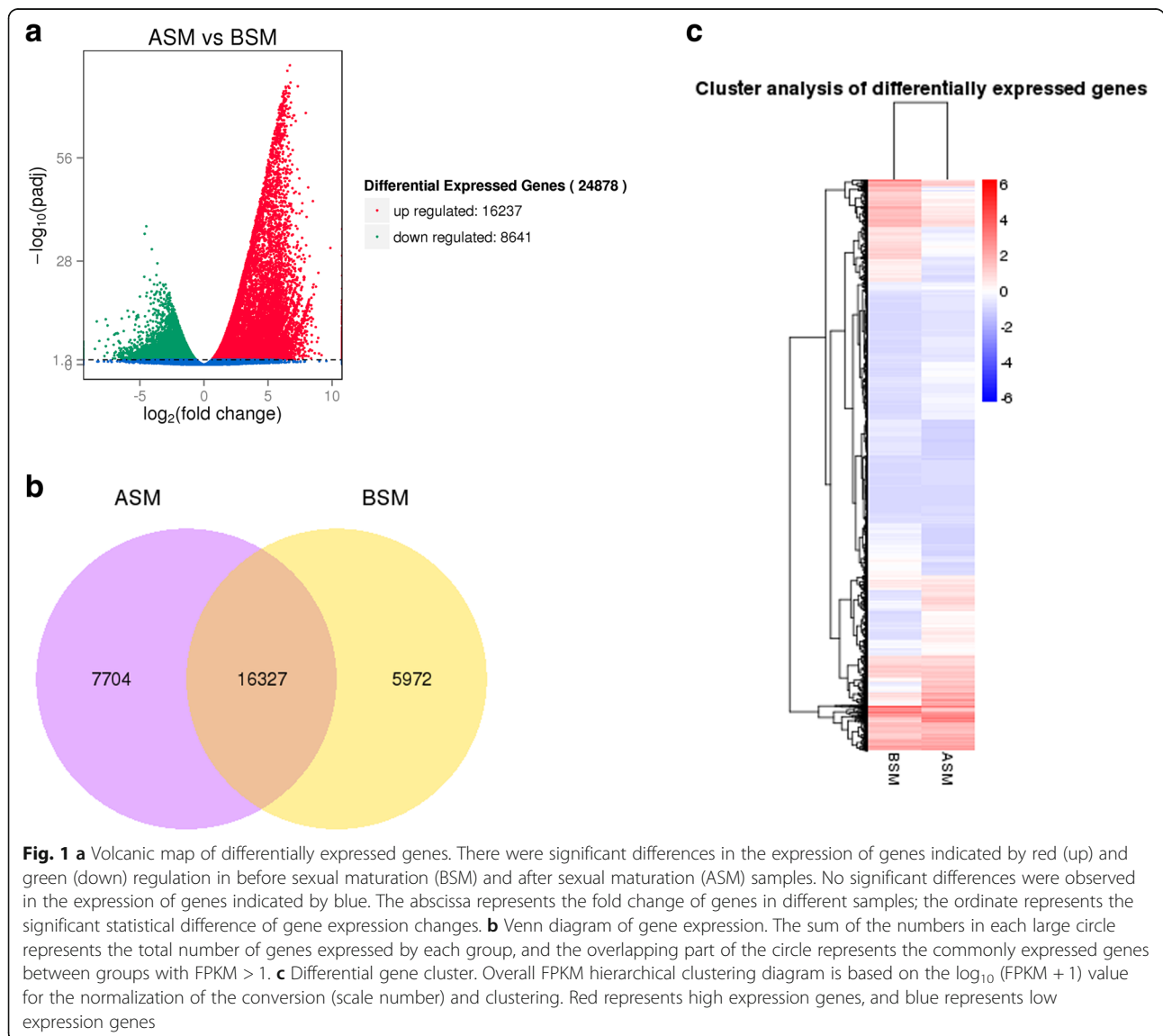
The DEGs between libraries were screened based on DESeq analysis, taking  $\text{padj} < 0.05$  as the cut-off. By considering libraries, the 16,237 up- and 8,641 down-regulated genes were identified in total that were up-regulated in ASM versus BSM (Fig. 1a). Venn diagrams were constructed to summarize the data, which show that the ASM and BSM libraries shared 16,327 genes that were expressed in both BSM and ASM (Fig. 1b). Figure 1c shows the

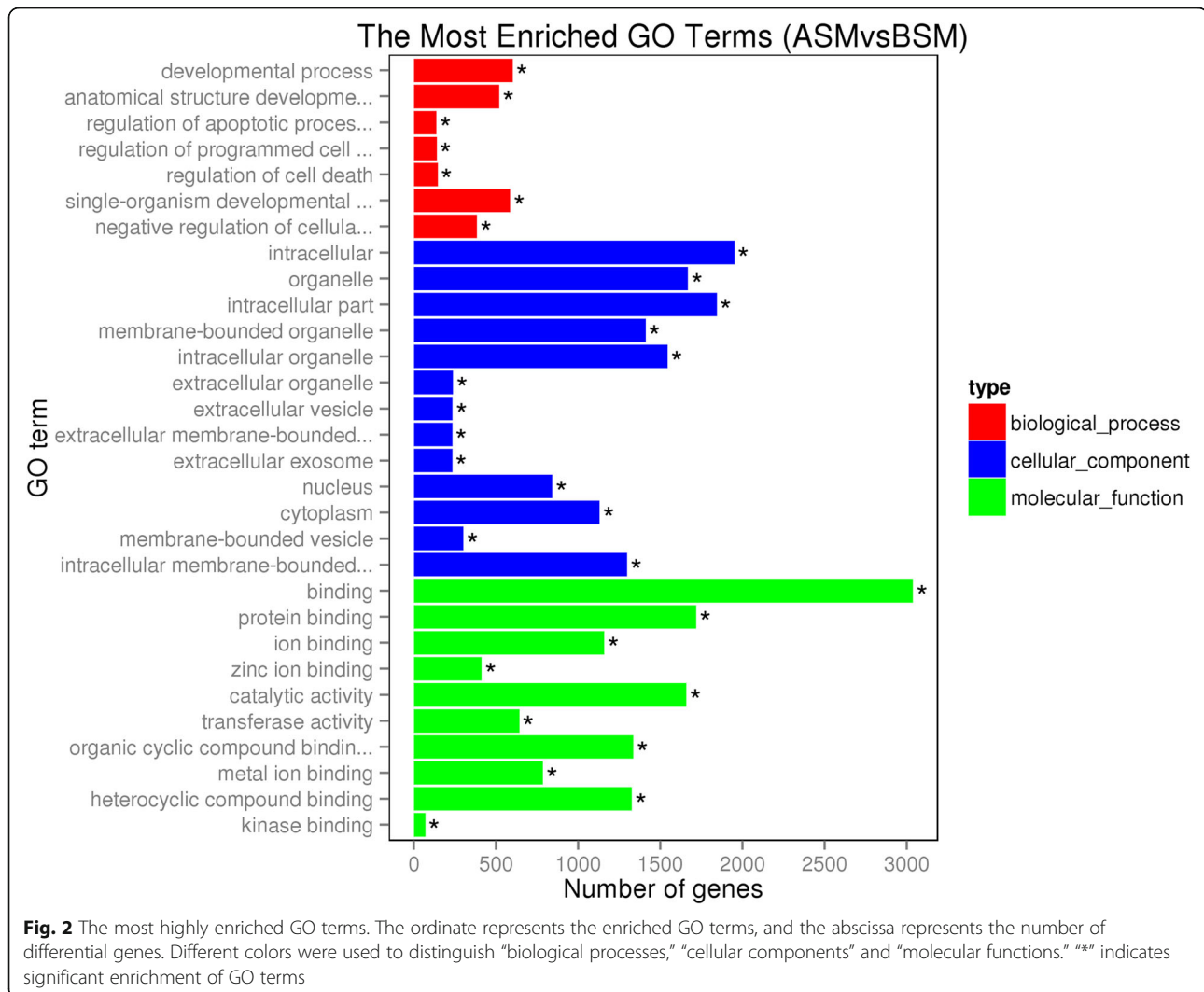
results of hierarchical clustering, with the DEGs of the six libraries divided into two clusters. These results show that the ASM and BSM libraries had overall distinct patterns of differential expression but that the replicates within each developmental stage were similar.

For further assessment of DEGs that are transcription factors. A total 922 transcription factors belong to 60 transcription factors families.

**GO enrichment analysis of DEGs**

Gene ontology (GO) enrichment analysis was used to explore the functions of the DEGs in testis development. A total of 70 GO terms related to “biological processes,” “molecular functions,” and “cellular components” were significantly enriched across the ASM vs BSM. DEG groups (Fig. 2). Approximately 32 GO terms belong to





“biological processes,” and these were mainly focused on “metabolic process” (2,454 genes), “organic substance metabolic process” (2,044 genes), and “cellular metabolic process” (1,948 genes). Approximately 20 GO terms belong to the functional category of “cellular component,” and these were mainly concentrated in “intracellular” (1,951 genes), “intracellular part” (1,844 genes), and “organelle” (1,668 genes) classification groups. Approximately 18 GO terms belong to “molecular function,” which centered on “binding” (3,037 genes), “protein binding” (1,719 genes), and “catalytic activity” (1,659 genes).

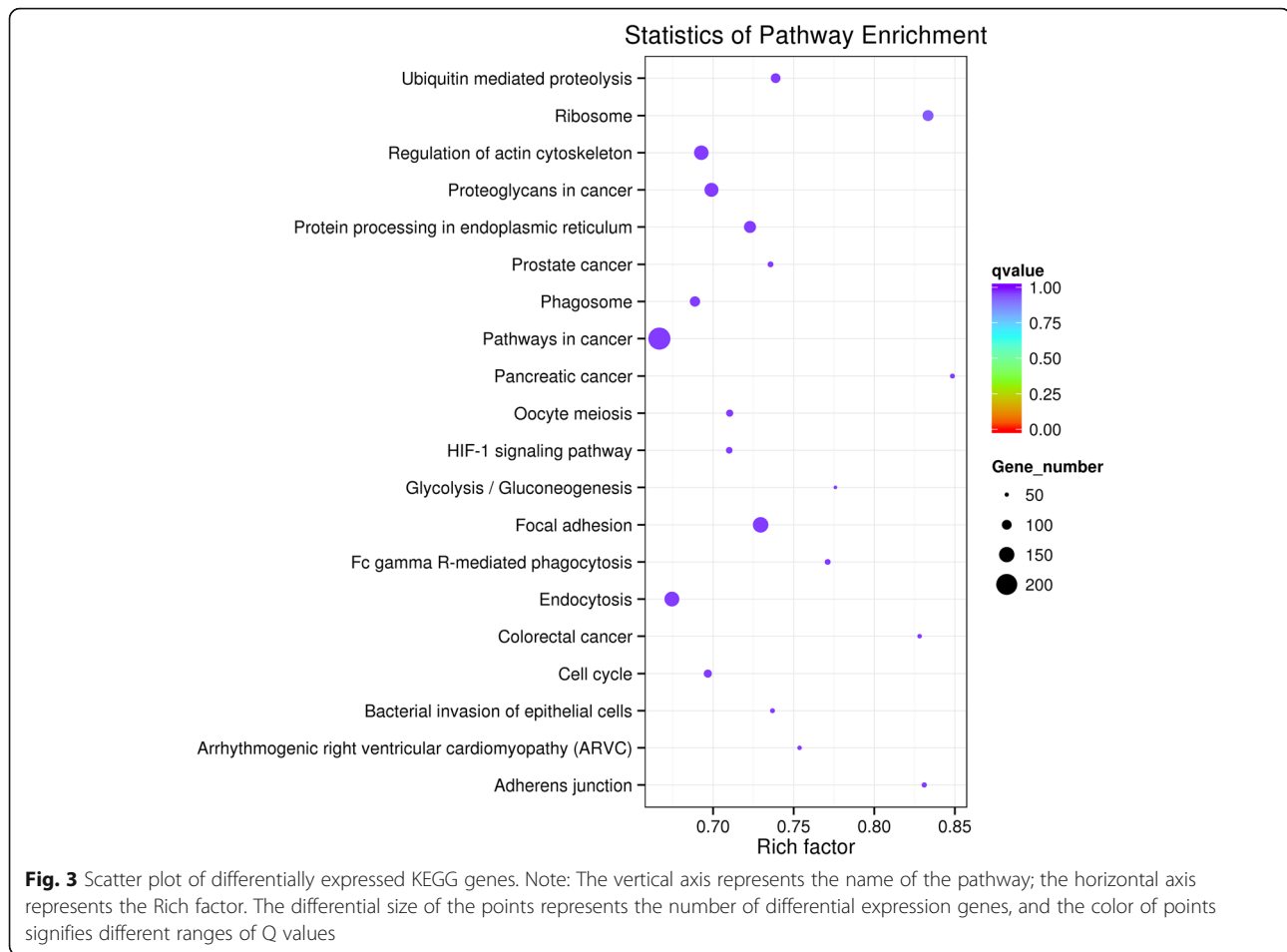
#### KEGG pathway enrichment analysis of DEGs

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed on these differentially expressed genes with a cut-off criterion of corrected  $P$ -values  $< 0.05$ . Six KEGG pathways were up-regulated, including “ubiquitin-mediated proteolysis,”

“protein processing in endoplasmic reticulum,” “RNA transport,” “metabolic pathways,” “oocyte meiosis,” and “cell cycle.” Four pathways were downregulated, including “ribosome,” “focal adhesion,” “proteoglycans in cancer,” and “ECM-receptor interaction” (Fig. 3).

#### Quantitative real time PCR (qRT-PCR) validation of DEGs

To verify the differential expression of genes in ASM vs. BSM testis, we selected eight DEGs, including *MEI1* (two different sites), *EHMT2*, *BTRC*, *CLOCK*, *IQCG*, *ANAPC5* and *SKP1*, to validate the expression patterns by qRT-PCR. These genes belong to “oocyte meiosis” KEGG pathways and GO terms “spermatogenesis,” “male gamete generation,” and “spermatid development” (Table S4). There was a relatively high correlation between the mRNA expression levels of the eight genes, as detected by qRT-PCR and RNA-Seq (correlation coefficient = 0.853). This suggests that our RNA-Seq results are reliable.



## Discussion

Currently, with the development of improved detection techniques, a growing number of sperm-specific mRNAs have been identified. RNA-Seq technology has become an efficient and inexpensive method for the discovery of novel transcripts and genes. In our study, three ASM testes and three BSM testes were used as research samples, and 24,878 genes were screened by RNA-Seq. The results led to the identification of 16,237 genes that were upregulated and 8,641 genes that were downregulated during development. Using the next-generation platform, we found that most of the highly expressed genes are differentiation-related with identifiable roles in the development of the testis and spermatogenesis. Our analysis also identified related genes that were upregulated in ASM, such as *SPATA3*, *SPATA18*, *SPATA19*, *SPATA45*, *SPATC1*, *SPEM1*, *SPERT*, *SHBG*, *TEX35*, *TEX43* and *GSG1*. Among these genes, *SPATA3*, *SPATA18*, *SPATA19*, *SPATA45* and *SPATC1*, have known roles in spermatogenesis; *TEX35* and *TEX43* are associated with testis expression; *GSG1* is related to germ cell function; *SPEM1* influences spermatid maturation; and *SHBG* might regulate sex hormone-binding

globulin. In addition, some novel genes detected are likely to be related to spermatogenesis.

Alternative splicing is a crucial mechanism for regulating gene expression and generating proteomic diversity. Recent estimates indicate that the expression of nearly 95% of human multi-exon genes involves AS [11, 12]. In metazoans, AS plays an important part in generating different protein products that function in diverse cellular processes, including cell growth, differentiation, and death. In our study, five kinds of AS events were detected, most of them involving SE. The SE events and the results of GO and KEGG analyses were comprehensively analyzed to determine the influence of AS events on the function of related genes [13]. In our study, IQ motif-containing G (*IQCG*) is regulated by 2 SE events as categorized under the GO terms “spermatogenesis,” “male gamete generation,” and “spermatid development”. The causality of the mutation was confirmed with a targeted null allele. Loss of *IQCG* disrupts spermiogenesis wherein tail formation is either incomplete or breaks apart from the sperm heads. Because IQ motif-containing genes typically regulate calmodulin, which in turn can impact the actin cytoskeleton, these findings suggest a potential role for localized calcium signaling in sperm flagellum morphogenesis,  $Ca^{2+}$  signaling is



important for regulating sperm motility, most notably in the context of flagellar function and regulation [14–19]. 11 SE and 4 MXE events of the meiotic double-stranded break formation protein 1 (*MEI1*) gene also occurred in ASM. *MEI1* was upregulated in ASM and is related to the GO terms “spermatogenesis,” “male gamete generation,” and “spermatid development.” *MEI1* is expressed specifically in the testis and the first meiosis-specific mutation identified by forward genetic approaches in mammals. Mutations in the *MEI1* gene disrupt meiosis during spermatogenesis, especially in Americans of European descent [20–22].

Other enriched genes in this study that belong to the “spermatogenesis,” “male gamete generation,” and “spermatid development” GO terms include *SLC26A8*, *MEI1*, *EHMT2*, *DPY19L2*, *HIST1H2BA*, *CAPZA3*, *TSSK2*, *TSSK1B*, *KDM3A*, *ALMS1*, and *IQCG*. Testis anion transporter 1 (Tat1) is a protein that, in humans, is encoded by the *SLC26A8* gene. *SLC26A8* gene was localized to developing spermatocytes [23–25]. The *SLC26A8* protein specifically expressed in male germ cells and mature sperm. In the mouse, deletion of Tat1 caused male sterility as a result of lack of sperm motility, impaired sperm capacitation and structural defects of the flagella [26, 27]. It acts as a DIDS-sensitive anion exchanger that mediates chloride, sulfate, and oxalate transport. It also has critical anion exchange functions in male germ lines during meiosis, and therefore, may play a role in spermatogenesis [25]. The *DPY19L2* protein, which belongs to the *dpy-19* family, is highly expressed in testis and is required for sperm head elongation and acrosome formation during spermatogenesis. Mutations in *DPY19L2* are associated with an infertility disorder known as spermatogenic failure type 9 (SPGF9) [28–35]. Testis-specific serine kinase 2 (*TSSK2*), and Testis-specific serine kinase 1B (*TSSK1B*) are required during spermatid development. During spermatogenesis, *TSSK2* is highly expressed in the testis and is required for the transformation of a ring-shaped structure around the base of the flagellum that originates from the chromatoid body [36–41]. The above genes have well-established functions in spermatogenesis; however, the regulatory mechanisms in the horse have not been verified and thus require further investigation.

In our study, six upregulated KEGG pathways were detected. Though the “oocyte meiosis” pathway is named according to its role in female fertility, genes detected in the pathway also regulated meiosis of sperm. Meiosis plays a critical role in the formation of sperm. A total genes in the oocyte meiosis pathway were detected, with functions that influence meiosis I and II. Cell division cycle 25C (*CDC25C*), which encode threonine/tyrosine phosphatases related with in meiotic cell cycle regulation, were found to be differentially localized in rat testicular germ cells during spermatogenesis [42]. Aurora kinase a, also known as serine/threonine-protein kinase

6, is an enzyme that, in humans, is encoded by the *AURKA* gene and is a member of a family of mitotic serine/threonine kinases [43]. This gene is implicated to have an essential role during mitosis and meiosis [44], and its activity peaks during the G2 phase to M phase transition in the cell cycle [45]. The G2/mitotic-specific cyclin-B1 (*CCNB1*) is encoded by the *CCNB1* gene in humans [46]. The CCNE protein belongs to the highly conserved cyclin family, whose activity is required for cell cycle G1/S transition. This protein accumulates during the G1-S phase and is degraded as cells progress through the S-phase [47]. PLK1 is an early trigger for G2/M transition and supports the functional maturation of the centrosome in late G2/early prophase. PLK1 phosphorylates and activates components of the anaphase-promoting complex (APC). The function of APC can maintain cohesion of sister chromatids and anaphase inhibitors [48]. Meiotic recombination protein REC8 homolog is a meiosis-specific component of the cohesin complex that binds sister chromatids in preparation for the two divisions of meiosis [49, 50]. F-box only protein 43 (*FBXO43*), also called endogenous meiotic inhibitor 2 (*EMI2*), was originally identified in a yeast two-hybrid screen as a novel *Xenopus* polo-like kinase (Plx-1) [51, 52]. *FBXO43* is also essential for spermatocytes to complete meiotic divisions in mice. In male *FBXO43* knockout mice, spermatocytes cannot complete meiotic divisions, and spermatids are absent in the testes [53].

Our study focused on increasing understanding of the process of testis development and spermatogenesis in the Mongolian horse, which ultimately may lead to the development of approaches to improve fertility. Due to the lack of information available in regards to the genetics of horse spermatogenesis, we have inferred the function of certain DEGs according to available information in humans and mice. Furthermore, DEGs that have functions in development may not be directly applicable to many situations of infertility in the horse, such as azoospermia and asthenozoospermia.

## Conclusions

In our study, six mRNA libraries were constructed and sequenced from horse testes from ASM and BSM. A total of 16,237 upregulated and 8,641 downregulated genes were identified; most genes had functions within biological process categories and are known to participate in spermatogenesis and testis development. We analyzed a large number of genes related to spermatogenesis and development of the horse testis, including those that play important roles in different developmental stages. We also found that the AS events might influence the function of these genes and indirectly affect the spermatogenesis and development of the horse testis. These findings will help us to further understand how

gene expression is regulated during testis development and spermatogenesis. Future research to evaluate the influence of AS events on testis development and spermatogenesis-related genes, or studies that combine other experimental results, such as miRNA and piRNA expression patterns, will provide more fundamental information in understanding the regulatory mechanisms of equine testis development and spermatogenesis at the molecular level.

## Materials and methods

### Ethics statement

This study was approved by and conducted according to the guidelines of the Institutional Animal Ethics Committee and Animal Care Guidelines of the Inner Mongolia Agricultural University. The animals did not unnecessarily suffer at any stage of this study.

### Animals and sample collection

We received permission to geld six healthy male Mongolian horses in Xilingol League, Inner Mongolia, China. The ages of the horses were determined based on physical examination of their teeth, as well as information from the owner. Three were sexually immature colts (before sexual maturation; samples BSM1–3), between 11 and 13 months old. The other three were sexually mature stallions, between 3 and 4 years old (after sexual maturation; samples ASM1–3). We surgically collected the testes of all six horses. The excised testes were stored in cryogenic vials with RNA/DNA sample protector (TaKaRa). Small aliquots (~ 5 g) from the testes of each animal from the parenchyma, including the seminiferous tubules and Leydig cells from the middle area, were immediately frozen in liquid nitrogen for quantitative real-time polymerase chain reaction (qRT-PCR) analysis. All horses were reared by their owner after our study.

### RNA quantification and qualification

Total RNA was extracted using TRIzol™ Reagent (Invitrogen) according to the manufacturer's protocols and resuspended in nuclease-free water. Then, the samples were treated with DNase I (Invitrogen) to eliminate any contaminating genomic DNA. RNA degradation and contamination were monitored on 1% agarose gels. RNA purity was assessed using the NanoPhotometer® spectrophotometer (IMPLEN). RNA concentrations were measured using the Qubit® RNA Assay Kit with the Qubit® 2.0 Fluorometer (Life Technologies). RNA integrity was assessed using the RNA Nano 6000 Assay Kit with the Bioanalyzer 2100 system (Agilent Technologies).

### Library preparation for strand-specific transcriptome sequencing

Six libraries were constructed. The three immature testes libraries were designated BSM1, BSM2, and BSM3; and the three mature testes libraries were designated ASM1, ASM2, and ASM3. Approximately 3 µg RNA per sample were used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB) according to the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-conjugated magnetic beads. Fragmentation was conducted using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV reverse transcriptase (RNaseH-). Second-strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H. In the reaction buffer, dNTPs with dTTP were replaced by dUTP. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activity. After adenylation of the 3' ends of the DNA fragments, NEBNext Adaptors with hairpin loop structure were ligated for hybridization. To select cDNA fragments of preferentially 150~200 bp in length, the libraries were purified with the AMPure XP system (Beckman Coulter). Then 3 µL of USER Enzyme (NEB) were used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min, followed by 5 min at 95 °C before PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Finally, the products were purified (AMPure XP system), and the library quality was assessed using the Agilent Bioanalyzer 2100 system.

### Clustering and sequencing

Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform and 150 bp paired-end reads were generated.

### Quality control

In order to guarantee data quality in our analysis, a Perl script was used to filter the Raw Data. The steps were as follows: The clean data were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data.

### Reads mapping to the reference genome

The reference genome and gene model annotation files were directly downloaded from the [www.ensembl.org](http://www.ensembl.org).

The index of the reference genome was built using Bowtie v2.2.3, and paired-end clean reads were aligned to the reference genome using TopHat v2.0.12 [54, 55]. TopHat was selected as the mapping tool because it can generate a database of splice junctions based on the gene model annotation file and thus provides a better mapping result than other non-splice mapping tools [56–58].

#### Quantification of gene expression

HTSeq v0.6.1 was used to count the read numbers mapped to each gene [59]. Then, the expected number of Fragments Per Kilobase of transcript sequence per Million base pairs sequenced (FPKM) of each gene was calculated based on the length of the gene and read count mapped to the gene. The FPKM simultaneously considers the effect of the sequencing depth and gene length for the read count, and is currently the most commonly used method for estimating gene expression levels [60].

#### Differential expression analysis

Differential expression analysis of two stages (three biological replicates per stage) was performed using the DESeq R package (1.18.0). DESeq provides statistical routines for determining differential expression within digital gene expression data using a model based on the negative binomial distribution. The resulting *P*-values were adjusted using the Benjamin and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted *P*-value < 0.05 found by DESeq were assigned as differentially expressed [61, 62].

#### GO and KEGG enrichment analyses of DEGs

GO enrichment analysis of DEGs was implemented by the GOseq R package, in which the gene length bias was corrected. GO terms with corrected *P*-value < 0.05 were considered significantly enriched DEGs [63]. KEGG is a database resource for understanding high-level functions and utilities of biological systems, such as the cell, the organism and the ecosystem, based on molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (<http://www.genome.jp/kegg/>) [64]. We used KOBAS software to test the statistical enrichment of DEGs in the KEGG pathways [65].

#### Differential gene transcription factor analysis

Putative transcription factors were analyzed using the animal transcription factor database AnimalTFDB 2.0 [66]. For genes included in the database, the transcription factors were directly screened by Ensembl geneID; Non-Ensembl geneID genes were screened through the horse transcription factor protein sequence database by BLASTX.

#### Novel transcript prediction and alternative splicing analysis

The Cufflinks v2.1.1 Reference Annotation Based Transcript (RABT) assembly method was used to construct and identify both known and novel transcripts from TopHat alignment results. Alternative splicing events were classified by the software rMATS (<http://rnaseq-mats.sourceforge.net/index.html>). The number of AS events in each sample was estimated separately.

#### qRT-PCR validation

A piece of testis was excised and frozen immediately in liquid nitrogen for qRT-PCR analysis. Total RNA was extracted using TRIzol™ Reagent (Invitrogen) according to the manufacturer's protocols. About 0.5 μg of total RNA was used as template to synthesize first-strand cDNA using a miScript II RT Kit (QIAGEN) according to the manufacturer's protocols. The resultant cDNA was diluted to 0.1 μg/μL for further analysis by the MX3000P Real-Time PCR System (Agilent Technologies) using SYBR® Premix Ex Taq™ II (TaKaRa). β-actin was chosen as an internal reference gene to eliminate sample-to-sample variations based on evidence of its role as a constitutive house-keeping gene for qRT-PCR and its use in normalizing changes in specific gene expressions [67]. The relative gene expression levels were calculated using the  $2^{-\Delta\Delta Ct}$  method. The DEGs between ASM and BSM testis were analyzed by t-test.

#### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12863-020-00843-5>.

**Additional file 1: Table S1.** List of data output quality.

**Additional file 2: Table S2.** Alignment of Reads and reference genome.

**Additional file 3: Table S3.** Classification and quantity statistics of AS events.

**Additional file 4: Table S4.** The results of qRT-PCR validation.

#### Acknowledgments

We would like to thank Nei Mongol BioNew Technology Co. Ltd. for providing assistance during the preparation of this manuscript.

#### Authors' contributions

BL and XH conceived the research. DM and DB collected the testicular samples of the Mongolian horse. BL, YZ, and XH performed RNA extraction, MD and LS conducted sequence preprocessing and functional annotation, ZL and ZY performed qRT-PCR validation. BL and DM wrote the paper. All authors have read and approved the final manuscript.

#### Funding

This study was supported by the Youth Fund of College of Animal Science, Inner Mongolia Agricultural University (QN201907, BZCG201901), Natural Science Foundation of Inner Mongolia (2015BS0314, 2017ZD06 and 2019ZD03), National Natural Science Foundation of China (3191101008, 31960657 and 31860636), People's Republic of China Specific Scientific and Technological Cooperation (2017YFE0108700) and Special Fund for Transformation of Scientific and Technological Achievements for Breeding of New Animals and



Plants in Inner Mongolia Agricultural University (YZGC2017001). Funding bodies had no influence on the design of the study, collection of the data, analysis, interpretation of the data or the writing of the manuscript.

#### Availability of data and materials

The accession number for the raw and processed profiling by high-throughput RNA sequencing technology reported in this paper is GEO: GSE101697.

#### Ethics approval and consent to participate

All procedures involving animals were approved and authorized by the Inner Mongolia Agricultural University. The animals used in our study were consented by their owners.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>College of Animal Science, Equine Research Center, Inner Mongolia Agricultural University, Hohhot 010018, China. <sup>2</sup>Inner Mongolia Key Laboratory of Equine Genetics, Breeding and Reproduction, Equine Research Center, Inner Mongolia Agricultural University, Hohhot 010018, China. <sup>3</sup>Scientific Observing and Experimental Station of Equine Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Equine Research Center, Inner Mongolia Agricultural University, Hohhot 010018, China. <sup>4</sup>Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot 010031, China.

Received: 16 October 2019 Accepted: 13 March 2020

Published online: 28 April 2020

#### References

- Hecht NB. Molecular mechanisms of male germ cell differentiation. *BioEssays*. 1998;20(7):555–61.
- Eddy EM, O'Brien DA. Gene expression during mammalian meiosis. *Curr Top Dev Biol*. 1998;37:141–200.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–8.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. 2008;18(9):1509–17.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2010;10(1):57–63.
- Ramsköld D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol*. 2009;5(12):e1000598.
- Djureinovic D, Fagerberg L, Hallstrom B, Danielsson A, Lindskog C, Uhlen M, Ponten F. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol Hum Reprod*. 2014;20(6):476–88.
- Anand M, Prasad BV. The computational analysis of human testis transcriptome reveals closer ties to pluripotency. *J Hum Reprod Sci*. 2012; 5(3):266–73.
- Jun J, Cho YS, Hu H, Kim H-M, Jho S, Gadhvi P, Park KM, Lim J, Paek WK, Han K, et al. Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genomics*. 2014;15(Suppl 9):S4.
- Doan R, Cohen ND, Sawyer J, Ghaffari N, Johnson CD, Dindot SV. Whole-genome sequencing and genetic variant analysis of a quarter horse mare. *BMC Genomics*. 2012;13:78.
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008;40(12):1413–5.
- Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136(4):701–18.
- Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*. 2009; 10(11):741–54.
- Ignotz GG, Suarez SS. Calcium/calmodulin and calmodulin kinase II stimulate hyperactivation in demembrated bovine sperm. *Biol Reprod*. 2005;73(3):519–26.
- Harris TP, Schimenti KJ, Munroe RJ, Schimenti JC. IQ motif-containing G (Iqcg) is required for mouse spermiogenesis. *G3 (Bethesda)*. 2014;4(2):367–72.
- Marquez B, Ignatz G, Suarez SS. Contributions of extracellular and intracellular Ca<sup>2+</sup> to regulation of sperm motility: release of intracellular stores can hyperactivate CatSper1 and CatSper2 null sperm. *Dev Biol*. 2007; 303(1):214–21.
- Lindemann CB, Goltz JS, Kanous KS. Regulation of activation state and flagellar wave form in epididymal rat sperm: evidence for the involvement of both Ca<sup>2+</sup> and cAMP. *Cell Motil Cytoskeleton*. 1987;8(4):324–32.
- Ho H-C, Granish KA, Suarez SS. Hyperactivated motility of bull sperm is triggered at the axoneme by Ca<sup>2+</sup> and not cAMP. *Dev Biol*. 2002;250(1): 208–17.
- Darszon A, Nishigaki T, Beltran C, Treviño CL. Calcium channels in the development, maturation, and function of spermatozoa. *Physiol Rev*. 2011; 91(4):1305–55.
- Sato H, Miyamoto T, Yogev L, Namiki M, Koh E, Hayashi H, Sasaki Y, Ishikawa M, Lamb DJ, Matsumoto N, et al. Polymorphic alleles of the human MEI1 gene are associated with human azoospermia by meiotic arrest. *J Hum Genet*. 2006;51(6):533–40.
- Reinholdt LG, Schimenti JC. Mei1 is epistatic to Dmc1 during mouse meiosis. *Chromosoma*. 2005;114(2):127–34.
- Libby BJ, Reinholdt LG, Schimenti JC. Positional cloning and characterization of Mei1, a vertebrate-specific gene required for normal meiotic chromosome synapsis in mice. *Proc Natl Acad Sci*. 2003;100(26):15706–11.
- Lohi H, Kujala M, Mäkelä S, Lehtonen E, Kestilä M, Saarialho-Kere U, Markovich D, Kere J. Functional characterization of three novel tissue-specific anion exchangers SLC26A7, -A8, and -A9. *J Biol Chem*. 2002; 277(16):14246–54.
- Dirami T, Rode B, Jollivet M, Da Silva N, Escalier D, Gaitch N, Norez C, Tuffery P, Wolf J-P, Becq F, et al. Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *Am J Hum Genet*. 2013;92(5):760–6.
- Touze A, Morin L, Pineau C, Becq F, Dorseuil O, Gacon G. Tat1, a novel sulfate transporter specifically expressed in human male germ cells and potentially linked to rhogtpase signaling. *J Biol Chem*. 2001;276(23):20309–15.
- Hernández-González EO, Treviño CL, Castellano LE, de la Vega-Beltrán JL, Ocampo AY, Wertheimer E, Visconti PE, Darszon A. Involvement of cystic fibrosis transmembrane conductance regulator in mouse sperm capacitation. *J Biol Chem*. 2007;282(33):24397–406.
- Rode B, Dirami T, Bakouh N, Rizk-Rabin M, Norez C, Lhuillier P, Lores P, Jollivet M, Melin P, Zvetkova I, et al. The testis anion transporter TAT1 (SLC26A8) physically and functionally interacts with the cystic fibrosis transmembrane conductance regulator channel: a potential role during sperm capacitation. *Hum Mol Genet*. 2012;21(6):1287–98.
- Ghédér H, Ibalá-Romdhane S, Okutman O, Viot G, Saad A, Viville S. Identification of a new DPY19L2 mutation and a better definition of DPY19L2 deletion breakpoints leading to globozoospermia. *Mol Hum Reprod*. 2016;22(1):35–45.
- Coutton C, Abada F, Karaouzene T, Sanlaville D, Satre V, Lunardi J, Jouk P-S, Arnoult C, Thierry-Mieg N, Ray PF. Fine characterisation of a recombination hotspot at the DPY19L2 locus and resolution of the paradoxical excess of duplications over deletions in the general population. *PLoS Genet*. 2013; 9(3):e1003363.
- Zhu F, Gong F, Lin G, Lu G. DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations. *Mol Hum Reprod*. 2013;19(6):395–404.
- Elnati E, Kuentz P, Redin C, Jaber S, Vanden Meerschaut F, Makarian J, Kosciński I, Nasr-Esfahani MH, Demirolo A, Gurgan T, et al. Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Hum Mol Genet*. 2012;21(16):3695–702.
- Coutton C, Zouari R, Abada F, Ben Khelifa M, Mardassi G, Triki C, Escalier D, Hesters L, Mitchell V, Levy R, et al. MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. *Hum Reprod*. 2012;27(8):2549–58.
- Kosciński I, Elnati E, Fossard C, Redin C, Muller J, Velez de la Calle J, Schmitt F, Ben Khelifa M, Ray P, Kilani Z, et al. DPY19L2 deletion as a major cause of globozoospermia. *Am J Hum Genet*. 2011;88(3):344–50.
- Harbuz R, Zouari R, Pierre V, Ben Khelifa M, Kharouf M, Coutton C, Mardassi G, Abada F, Escoffier J, Nikas Y, Vialard F, Kosciński I, Triki C, Sermondade N,

- Schweitzer T, Zhioua A, Zhioua F, Latrous H, Halouani L, Ouafi M, Makni M, Jouk PS, Sèle B, Hennebicq S, Satre V, Viville S, Arnoult C, Lunardi J, Ray PF. A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *Am J Hum Genet.* 2011;88(3):351–61.
35. Yassine S, Escoffier J, Martinez G, Coutton C, Karaouzène T, Zouari R, Ravanat J-L, Metzler-Guillemain C, Lee HC, Fissore R, et al. Dpy19L2-deficient globozoospermic sperm display altered genome packaging and DNA damage that compromises the initiation of embryo development. *Mol Hum Reprod.* 2015;21(2):169–85.
  36. Zhang Z, Shen X, Jones BH, Xu B, Herr JC, Strauss JF 3rd. Phosphorylation of mouse sperm axoneme central apparatus protein SPAG16L by a testis-specific kinase, TSSK2. *Biol Reprod.* 2008;79(1):75–83.
  37. Xu B, Hao Z, Jha KN, Zhang Z, Urekar C, Digilio L, Pulido S, Strauss JF, Flickinger CJ, Herr JC. Targeted deletion of Tssk1 and 2 causes male infertility due to haploinsufficiency. *Dev Biol.* 2008;319(2):211–22.
  38. Li Y, Sosnik J, Brassard L, Reese M, Spiridonov NA, Bates TC, Johnson GR, Anguita J, Visconti PE, Salicioni AM. Expression and localization of five members of the testis-specific serine kinase (Tssk) family in mouse and human sperm and testis. *Mol Hum Reprod.* 2011;17(1):42–56.
  39. Shang P, Hoogerbrugge J, Baarends W, Grootegeod JA. Evolution of testis-specific kinases TSSK1B and TSSK2 in primates. *Andrology.* 2013;1:160–8.
  40. Shang P, Baarends WM, Hoogerbrugge J, Ooms MP, van Cappellen WA, de Jong AA, Dohle GR, van Eenennaam H, Gossen JA, Grootegeod JA. Functional transformation of the chromatoid body in mouse spermatids requires testis-specific serine/threonine kinases. *J Cell Sci.* 2010;123(Pt 3):331–9.
  41. Kueng P, Nikolova Z, Djonov V, Hemphill A, Rohrbach V, Boehlen D, Zuercher G, Andres AC, Ziemiecki A. A novel family of serine/threonine kinases participating in spermiogenesis. *J Cell Biol.* 1997;139(7):1851–9.
  42. Mizoguchi S, Kim KH. Expression of cdc25 phosphatases in the germ cells of the rat testis. *Biol Reprod.* 1997;56(6):1474–81.
  43. Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, Brinkley BR, Sen S. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet.* 1998;20(2):189–93.
  44. Crane R, Gadea B, Littlepage L, Wu H, Ruderman JV, Aurora A, meiosis and mitosis. *Biol Cell.* 2004;96(3):215–29.
  45. Bischoff JR, Plowman GD. The Aurora/Ipl1p kinase family: regulators of chromosome segregation and cytokinesis. *Trends Cell Biol.* 1999;9(11):454–9.
  46. Sartor H, Ehlerf F, Grzeschik KH, Müller R, Adolph S. Assignment of two human cell cycle genes, CDC25C and CCNB1, to 5q31 and 5q12, respectively. *Genomics.* 1992;13(3):911–2.
  47. Ohtsubo M, Theodoras AM, Schumacher J, Roberts JM, Pagano M. Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol Cell Biol.* 1995;15(5):2612–24.
  48. van de Weerd BCM, Medema RH. Polo-like kinases: a team in control of the division. *Cell Cycle.* 2006;5(8):853–64.
  49. Xu H, Beasley MD, Warren WD, van der Horst GT, McKay MJ. Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Dev Cell.* 2005;8(6):949–61.
  50. Lee J, Iwai T, Yokota T, Yamashita M. Temporally and spatially selective loss of Rec8 protein from meiotic chromosomes during mammalian meiosis. *J Cell Sci.* 2003;116:2781–90.
  51. Schmidt A, Rauh NR, Nigg EA, Mayer TU. Cytostatic factor: an activity that puts the cell cycle on hold. *J Cell Sci.* 2006;119(Pt 7):1213–8.
  52. Tung JJ, Hansen DV, Ban KH, Loktev AV, Summers MK, Adler JR 3rd, Jackson PK. A role for the anaphase-promoting complex inhibitor Emi2/XErp1, a homolog of early mitotic inhibitor 1, in cytostatic factor arrest of *Xenopus* eggs. *Proc Natl Acad Sci U S A.* 2005;102(12):4318–23.
  53. Gopinathan L, Szmyd R, Low D, Diril MK, Chang HY, Coppola V, Liu K, Tessarollo L, Guccione E, van Pelt AMM7, Kaldis P. Emi2 is essential for mouse spermatogenesis. *Cell Rep.* 2017;20(3):697–708.
  54. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
  55. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9.
  56. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105–11.
  57. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36.
  58. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012;7(3):562–78.
  59. Anders S. HTSeq: analysing high-throughput sequencing data with Python. 2010.
  60. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28(5):511–5.
  61. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106.
  62. Anders S, Huber W. Differential expression of RNA-Seq data at the gene level—the DESeq package. 2012.
  63. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11(2):R14.
  64. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36(Database issue):D480–4.
  65. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics.* 2005;21(19):3787–93.
  66. Zhang H-M, Liu T, Liu C-J, Song S, Zhang X, Liu W, Jia H, Xue Y, Guo A-Y. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* 2015;43(Database issue):D76–81.
  67. Mori R, Wang Q, Danenberg K, Pinski J, Danenberg P. Both  $\beta$ -actin and GAPDH are useful reference genes for normalization of quantitative RT-PCR in human FFPE tissue samples of prostate cancer. *Prostate.* 2008;68(14):1555–60.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

