**MALE REPRODUCTION: ORIGINAL ARTICLE**

# Histone Methylation Regulates Gene Expression in the Round Spermatids to Set the RNA Payloads of Sperm

Saumya Sarkar[1] · Santosh Yadav[1] · Poonam Mehta[1,2] · Gopal Gupta[1,2] · Singh Rajender[1,2]

## Abstract

Gene expression during spermatogenesis undergoes significant changes due to a demanding sequence of mitosis, meiosis, and differentiation. We investigated the contribution of H3 histone modifications to gene regulation in the round spermatids. Round spermatids were purified from rat testes using centrifugal elutriation and Percoll density-gradient centrifugation. After enzymatic chromatin shearing, immuno-precipitation using antibodies against histone marks H3k4me3 and H3K9me3 was undertaken. The immunoprecipitated DNA fragments were subjected to massive parallel sequencing. Gene expression in round spermatids and sperm was analyzed by transcriptome sequencing using next-generation sequencing methods. ChIP-seq analysis showed significant peak enrichment in H3K4me3 marks in active chromatin regions and H3K9me3 peak enrichment in repressive regions. We found 53 genes which showed overlapping peak enrichment in both H3K4me3 and H3K9me3 marks. Some of the top H3K4me3-enriched genes were involved in sperm tail formation (*Odf1, Odf3, Odf4, Oaz3, Ccdc42, Ccdc63,* and *Ccdc181*), chromatin condensation (*Dync1h1, Dynll1,* and *Kdm3a*), and sperm functions such as acrosome reaction (*Acrbp* and *Fabp9*), energy generation (*Gapdhs*), and signaling for motility (*Tssk1b, Tssk2,* and *Tssk4*). Transcriptome sequencing in round spermatids found 64% transcripts of the H3K4me3-enriched genes at high levels and of about 25% of H3K9me3-enriched genes at very low levels. Transcriptome sequencing in sperm found that more than 99% of the ChIP-seq corresponding transcripts were also present in sperm. H3K4me3 enrichment in the round spermatids correlates significantly with gene expression and H3K9me3 correlates with gene silencing that contribute to sperm differentiation and setting the RNA payloads of sperm.

**Keywords** Spermatogenesis · Histone methylation · Round spermatids · Gene transcription · ChIP-sequencing (ChIP-seq)

## Introduction

Gene expression during spermatogenesis undergoes a roller-coaster ride as this highly complex process involves cell division (mitotic), chromosome reduction (meiotic), and differentiation (post-meiotic) phases that require diverse sets of genes at work [1]. Post-meiosis, two processes must progress simultaneously: DNA compaction and cell differentiation. The process of DNA compaction would render gene expression impossible, and hence all transcripts required for the synthesis of proteins participating in the differentiation process must be synthesized at this stage. Eventually, a spurt in gene expression is seen in the round spermatids before they become transcriptionally inactive [2]. A variety of regulatory and coding RNAs generated during spermiogenesis are stored in the round spermatids, which also make their way to mature spermatozoa probably in a selective manner [1]. mRNAs generated during this phase are translationally repressed, stored, and later used for translation for the completion of spermiogenesis [2]. However, the mechanisms of gene activation in the round spermatids remain largely unknown, and the mechanisms of gene silencing remain almost entirely unknown. It is well known that histones get replaced by protamines during spermatogenesis; however, the presence of histones is still evident in mature sperm, suggesting their critical role in development and spermatogenesis [3]. It is noteworthy that epigenetic changes during this stage would define the chromatin structure, which is essential for extraordinary packaging of DNA in spermatozoa [4].

✉ Singh Rajender
rajender_singh@cdri.res.in

1 Division of Endocrinology, CSIR-Central Drug Research Institute, Lucknow, India

2 Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

Histone proteins are subjected to a number of post-translational modifications (PTMs), including methylation, acetylation, ubiquitination, and phosphorylation [5]. It is a combination of a number of these modifications, which defines the chromatin's open or closed states. The maturation of immature germ cells into mature spermatozoa is mediated by a complex interplay of histone PTMs [5]. A large number of these modifications play a crucial role in protamine transition and chromatin reorganization in the process of spermatogenesis [5, 6]. Among a number of such PTMs, modifications at Lys4, Lys9, and Lys27 have been particularly characterized for their role in chromatin condensation [6]. Histone H3 trimethylation modification on lysine 4 residue (H3K4me3) is an important histone modification that affects gene expression [7]. H3K4me3 marks near the transcriptional start sites (TSSs) evince active transcription or transcriptional readiness [8]. H3K4me3 was initially regarded as the regulator of *Hox* genes. H3K4me3 marks facilitate the open state by recruiting the chromatin remodelling factors CHD1 [9] and BPTF [10] while blocking the binding of the chromatin repressive proteins like NuRD [11] and INHAT complexes [12]. Histone H3 methylation modification on lysine 9 residue (H3K9) is the mark of condensed heterochromatin and a state of transcriptional inactiveness as di/trimethylation modification on these residues (H3K9me2/3) marks are abundantly present in the silenced genes [13]. H3K9me3 binds to the heterochromatin protein 1 (HP1) to form constitutive heterochromatin and is responsible for transcriptional repression and the maintenance of heterochromatin [14]. HP1 also recruits DNA methyltransferase 3b, providing one of the best examples of the interplay between histone methylation and DNA methylation [15].

Spermatids represent the last stage of transcriptionally active germ cells, which see significant changes in gene expression as they enter differentiation. Since transcription must be shut after this stage, these cells are ideal for studying the chromatin changes that affect gene expression and to study how these cells prepare for differentiation. In most of the studies on sperm and embryonic stem cells, H3K4me3 and Histone H3 trimethylation modification on lysine 27 residue (H3K27me3) have been studied as the activation and repressive marks, respectively [16]. In the present study, we employed chromatin immunoprecipitation sequencing (ChIP-Seq) to figure out the correlation of H3K4me3 marks with transcriptional activity and H3K9me3 marks with transcriptional silencing in the round spermatids purified from rat testes. We also undertook transcriptome sequencing in the round spermatids to find if H3K4me3- and H3K9me3-enriched regions correlate with gene expression profile in these cells. Transcripts corresponding to a large number of H3K4me3-enriched genes (~ 64%) were actually present in the round spermatids and transcripts corresponding to only 25% of H3K9me3-enriched genes were present in the round

spermatids, that too at very low levels. We also undertook sperm transcriptome sequencing and found that almost all the transcripts corresponding to open chromatin regions (suggested by H3K4me3) present in the round spermatids were also detectable in spermatozoa.

## Materials and Methods

### Animals

Male Sprague–Dawley rats of 60–70 days of age were recruited from the National Laboratory Animal Centre, CSIR-CDRI, Lucknow. All rats were kept under a 12L:12D cycle in a vivarium with proper housing with ad libitum access to water and rat chow. Four sexually mature adult animals were sacrificed by anesthetic overdose under a protocol approved by the Institutional Animal Ethics Committee of the CSIR-Central Drug Research Institute, Lucknow (IAEC/2014/49/Renew03(135/16)).

### Isolation and Purification of Round Spermatids from Rat Testes

Isolation and purification of the round spermatids were done by trypsin digestion followed by centrifugal elutriation and density gradient centrifugation described in our earlier study [17]. In brief, the testes were first decapsulated and then minced with scissors in Basal Medium Eagle (BME). Subsequently, the minced testis suspension containing Basal Medium Eagle (BME) supplemented with 0.1% trypsin (w/v), 0.1% glucose, and 17 µg/ml DNase (Sigma-Aldrich, USA) was incubated in a water bath with shaking for 15 min at 34 °C. After incubation, the enzyme reactions were abolished by adding soybean trypsin inhibitor (0.04% w/v) in the suspension, which was then filtered through a nylon mesh (36 µm), followed by sperm removal by passing it through a column of glass wool. The collected cell suspension was centrifuged at 400 g for 5 min at 4 °C to obtain the cell pellet, which was washed twice with BME. The mixed germ cell pellet was suspended in BME supplemented with DNase (2 µg/ml) and FBS (8% V/V) and kept on ice. Later, the cell suspension was elutriated with a Beckman Elutriator Rotor (JE-5) fitted with a standard chamber and mounted on a Beckman High-Speed Centrifuge (Avanti J-26S–XP, Beckman Coulter Inc, USA). A fraction was collected at 2000 rpm at flow rates of 23.0 and 40.0 ml/min containing round spermatids at around 75–80% purity. To increase the purity level, the fraction was layered over a linear gradient of 23–33% Percoll® (Sigma-Aldrich, USA), and centrifuged at 4025 g for 60 min in a swinging bucket rotor fitted on to a Sigma 3-30 K refrigerated centrifuge (SIGMA Laborzentrifugen GmbH, Germany). The major band was carefully

recovered through a puncture in the side of the tube, followed by washing and dilution with BME. Furthermore, the purity of cell preparation was visually checked under a microscope using acridine orange staining (Fig. S1).

## Chromatin Immunoprecipitation Assay

Chromatin immunoprecipitation (ChIP) assays were performed using the Simple ChIP Enzymatic Chromatin IP kit (9003, CST, USA) according to the protocol provided by the manufacturer. Round spermatids were subjected to cross-linking by adding 37% formaldehyde at a final concentration of 1% and incubating at room temperature for 10 min. The cross-linking was quenched by the addition of 125 mM glycine followed by nuclear shearing and chromatin digestion by sonication. For the input sample, 2% of enzyme digested chromatin was collected and stored at $-20$ °C. The fragmented chromatin was analyzed on agarose gel which was found to be around 150–300bp in size. The sonicated chromatin was pre-cleared by centrifugation and subjected to immunoprecipitation with mouse anti-histone H3 (tri methyl K4) antibody (2 µg/IP) (ab12209, Abcam, UK) and rabbit anti-histone H3 (tri methyl K9) antibody (2 µg/IP) (ab8898, Abcam, UK) or with rabbit anti-IgG as the negative control at 4 °C overnight with rotation. The cross-link was reversed, and desired immune complexes were recovered by incubating with magnetic protein G beads at 65 °C for 30 minutes. The DNA thus precipitated was purified by column and eluted in a buffer for further downstream processing.

## Library Preparation and High Throughput Sequencing

Immunoprecipitated DNA fragments were subjected to library preparation using TrueSeq DNA Kit (FC-121–2001, Illumina, USA). NEB Next index set 3 was used for barcoding and adapter ligation (E7710S, NEB, USA). Briefly, 5–10 ng of ChIP DNA was subjected to end repair which converts the overhangs resulted from fragmentation into blunt ends using an end repair mix. This end repair mix has 3′ to 5′ exonuclease activity which removes the 3′ overhangs and the polymerase activity fills in the 5′ overhangs. After this step, a single 'A' nucleotide is added to the 3' ends of the blunt fragments, which prevents the end-repaired fragments from ligating to each other which is called A tailing. Following this step, adapter ligation was performed where adapters having indexes were ligated to the ends of the DNA fragments, enabling them for hybridization onto a flow cell. Purification of adapter-ligated fragments followed by enrichment of library was done to amplify DNA fragments having adapter sequences. DNA sequencing was performed on the Illumina NextSeq 500 platform (Illumina Inc, USA) with a read length of 75 bp and paired end chemistry.

## ChIP-seq Data Analysis and Visualization

The ChIP-seq reads obtained from the sequencer were demultiplexed using bcl2fastq tool. Quality and adapter trimming was done using FastQC software. The processed data/reads were then aligned with rat genomic assembly rn6 by using Bowtie 1.2 (parameters -m 1 -best). Peak calling was done using MACS1.4 with $p$-value threshold of 0.001 and 10% of FDR cutoff. The bed files of peaks were subjected to motif discovery by rGADEM package and these motifs were visualized using MotIV package [18]. The resulting bed files were analyzed and visualized using R/bioconductor platform–based tool ChIPseeker [19]. Further functional enrichment analysis of the annotated peaks was carried out using R/bioconductor platform–based ReactomePA [20] and clusterProfiler [21].

## RNA Isolation and Transcriptome Sequencing

RNA was extracted using Qiagen RNeasy Micro Kit (74,004, Qiagen, Germany) according to the manufacturer's recommended protocol. The integrity and quality of the RNAs were checked by Agilent 2100 bioanalyzer (G2939BA, Agilent, USA), and the qualified RNA samples were used for sequencing. RNA was enriched with polyA tail using Dynabeads™ mRNA DIRECT™ purification kit (61,012, Thermo Fisher Sci, USA) and subjected to amplification of cDNA by SMARTer® Ultra® Low Input RNA Kit for Sequencing-v3 kit (634,848, Takara Bio Inc., USA). TruSeq RNA Library Prep Kit v2 was used for mRNA-seq library preparation (RS-122–2001, Illumina Inc, USA) and sequencing was performed on Illumina Hiseq 2500 next-generation sequencing platform (Illumina Inc, USA).

## Transcriptome Data Processing

The raw RNA-seq reads were processed for adapter sequences and quality trimmed followed by their mapping to rat genomic assembly rn6 using TopHat (v2.0.8b, http://tophat.cbcb.umd.edu/). HTSeq (http://www.huber.embl.de/users/anders/HTSeq/doc/overview.html) was used for obtaining read counts. Differentially expressed genes were analyzed using DESeq R software pack. Benjamini–Hochberg multiple testing corrections were employed to identify differentially expressed genes.

# Results

## Deep Sequencing Read Quality

Deep sequencing of the ChIP DNA fragments generated an average of 36 million raw reads per ChIP reaction with

91% of bases having a Q score of greater than 30% and 94% having greater than 20. Quality enhancement and adapter removal of raw reads fetched an average of 34.7 million good quality reads with 94% and 95% of bases possessing quality scores of 30 and 20, respectively. The maximum and minimum read length considered for alignment and peak calling was 76 bp and 50 bp, respectively, with an average read length of 75 bp. The GC content of the processed reads remained below 45% in all reads (Table S1).

## Distribution of H3K4me3 and H3K9me3 Marks in the Regulatory Regions

ChIPseeker package was used for annotation and profiling of peak calling data. After peak calling, peak locations and coverage over chromosomes were identified and visualized using the 'Covplot' function, which calculates the coverage of peak regions over chromosomes. Well-dispersed peaks were observed in the case of H3K4me3 and H3K9me3 marks, which were mutually exclusive in many instances (Fig. 1, Fig. S2a and S2b). We also compared the distribution of peaks between H3K4me3 and H3K9me3 marks. The active regions showed significant peak enrichment in H3K4me3 marks and repressive regions showed enrichment in H3K9me3 marks. Some regions also showed bivalent peak enrichment, suggesting a poised condition of certain regions/genes (Fig. 1). To map the location of

histone marks with respect to genes, heat-map and average density profiling of ChIP peaks with respect to the transcription start site (TSS) regions were generated. The read density of H3K4me3 showed enrichment in the 1.5 kb region from the TSS (Fig. 2a). On the other hand, read densities of H3K9me3 were distributed widely from the TSS (Fig. 2b). This dispersed pattern is present in the regions where no enrichment was observed with H3K4me3, suggesting discrete regions of activity and inactivity. The distribution of histone marks with respect to the TSS showed that more than 35% of total H3K4me3 peaks lie within 3 kb upstream and downstream of the TSS whereas only about 10% of H3K9me3 were seen in this region (Fig. 2c and d).

In order to get information about the genomic location of enriched peaks, we used "annotatePeak" function of the ChIPseeker package, which assigns peaks to specific genomic annotations such as TSS, exon, 5' UTR, 3' UTR, intronic or intergenic. Approximately 27% of the H3K4me3 peaks were in the promoter regions of various genes (Fig. 3a). Enrichment of H3K4me3 peaks in the promoter regions suggests overall high transcriptional activity in spermatids. About 86% of H3K9me3 marks were seen in the intergenic regions, with a very low frequency in the promoter regions (~4%) (Fig. 3b). The low frequency of H3K9me3 peaks in the promoter regions also suggested overall high transcriptional activity in spermatids. This is also shown by a high number of H3K4me3 peaks in the
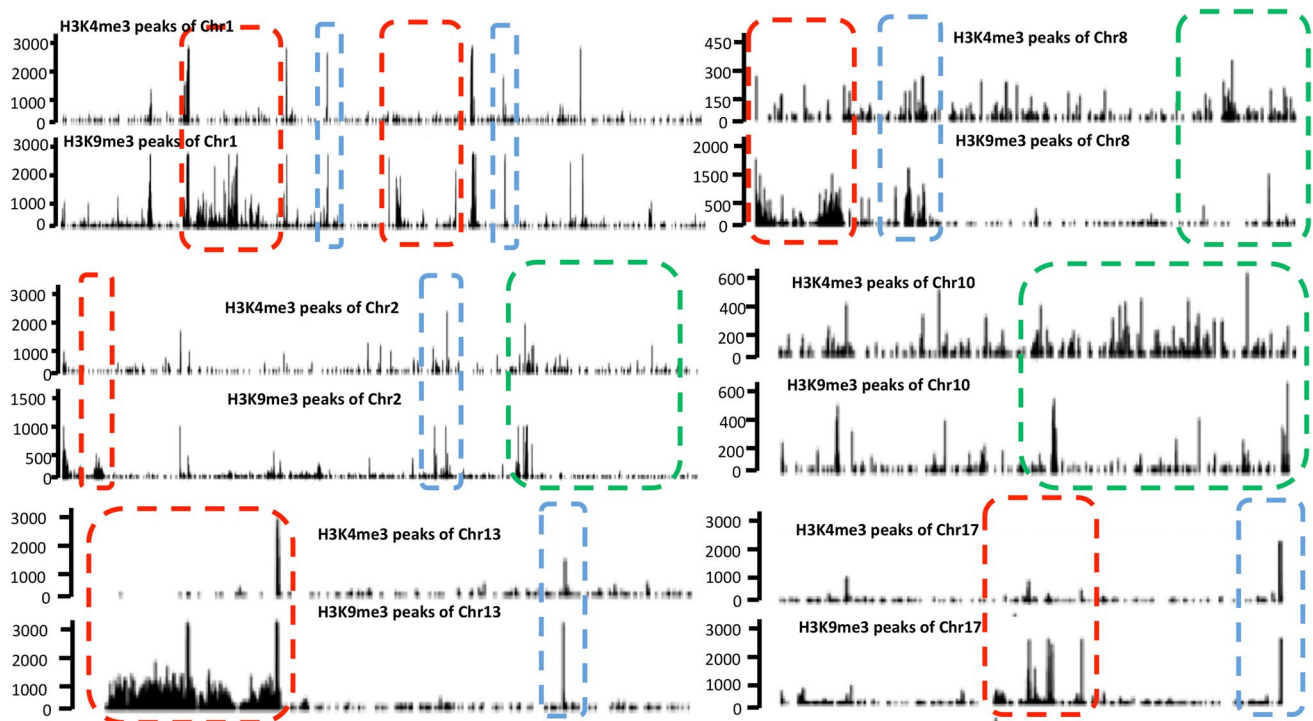


**Fig. 1** Representative peak enrichment profiles of H3K4me3 and H3K9me3 in some chromosomes showing active (green box), repressive (red box), and bivalent states (blue box)
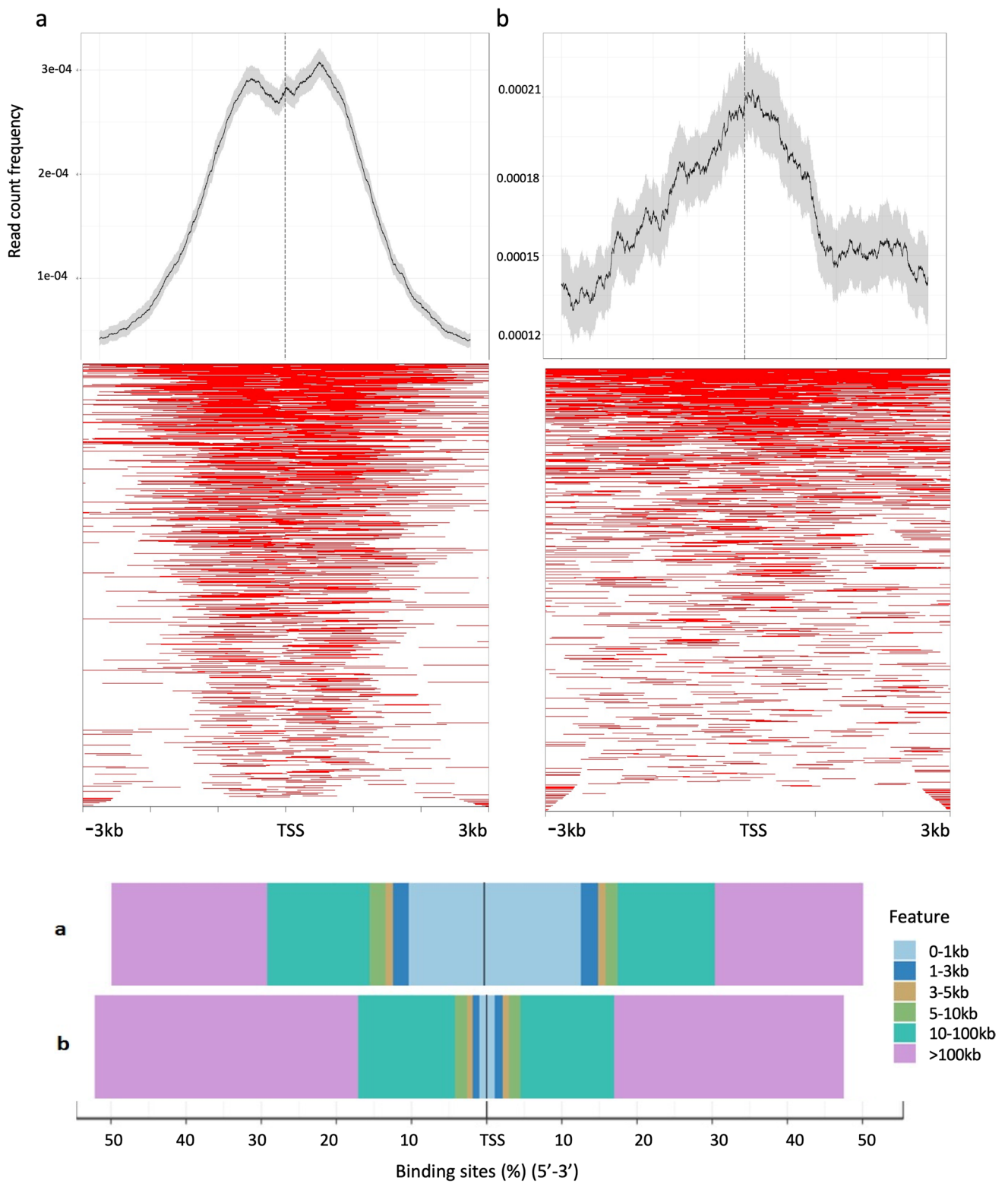
**Fig. 2** Heat map and average profile of peaks depicting their distribution in 3 Kb upstream and downstream of the TSS for H3K4me3 (**a**), H3K9me3 (**b**) (upper panel) Bar plots exhibiting the percentage distribution of peaks within the region of TSS and 100kbs for H3K4me3 (**a**) and H3K9me3 (**b**) (lower panel)
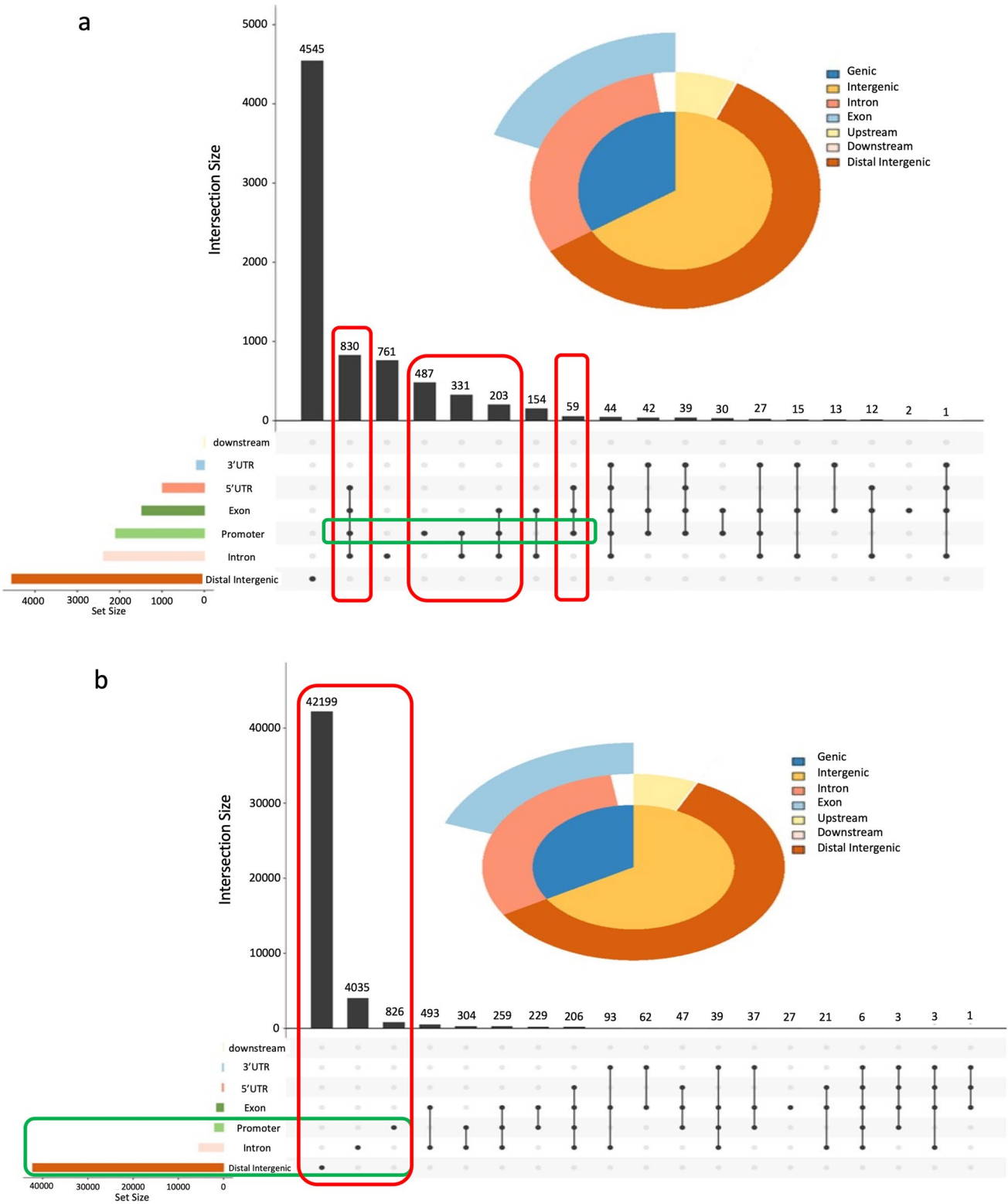
**Fig. 3** Upset plot and Venn pie plot (inset) showing the annotation overlap between different genomic regions for H3K4me3 (**a**), H3K9me3 (**b**)

## Functional Enrichment Analysis of H3K4me3 Peaks

H3K4me3 ChIP-seq identified 3958 peaks corresponding to 2807 genes. In order to identify the predominant biological pathways represented by the H3K4me3 (active) peaks, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (Fig. 4a). This analysis fetched around 40 genes involved in protein processing in the endoplasmic reticulum, 30 genes for RNA transport and 25 genes for spliceosome. Apart from these, approximately 20 genes of the cell cycle pathway were also enriched. Next, these genes were subjected to gene ontology (GO) analysis. The GO analysis highlighted the processes such as cilium development, spermatid differentiation, spermatid development, and germ cell development, which suggests gene expression relevant to germ cell differentiation (Fig. 4b). Similarly, in the cellular component domain of GO, we observed significant enrichment of sperm and flagellum-related components (Fig. 4c). In the molecular function domain, ubiquitin-related activities showed enrichment (Fig. 4d).

Gene concept networks analysis also depicted crucial spermatid and sperm-related biological linkage to genes enriched in H3K4me3 (Fig. 5). In order to identify mutually overlapping genes and gene sets in enriched GO terms, an enrichment map was created (Fig. S3). Enrichment map analysis of biological processes domain fetched a total of six clusters, among which the most notable were processes such as spermatid development, spermatid differentiation, germ cell development, sperm motility, flagellated sperm motility, and cellular processes involved in spermatogenesis. Similar analysis of cellular components produced seven clusters and the largest of them consisted of terms such as sperm part, sperm flagellum, and sperm principal piece. A thorough literature search for all these genes showed that 134 genes had an already established role in spermatogenesis (Table 1) and the remaining genes have a high plausibility of participating in spermatid differentiation (Data S1).

## Functional Enrichment Analysis of H3K9me3 Peaks

H3K9me3 ChIP-seq identified 6352 peaks corresponding to 1836 genes. KEGG pathway analysis of H3K9me3 peaks demonstrated enrichment with regard to processes other than spermiogenesis, such as retinol metabolism (31 genes), steroid hormone biosynthesis (27 genes), inflammatory mediator regulation of transient receptor potential (TRP) channels (32 genes), and chemical carcinogenesis (25 genes) (Fig. 6a). The biological process of GO analysis showed the highest number of hits in pattern specification processes (~ 110 genes) followed by synapse organization

(114 genes), regionalization (98 genes), and embryonic organ morphogenesis (88 genes) (Fig. 6b). In the cellular component domain, synaptic membrane and post-synaptic membrane exhibited the highest enrichment with 109 and 83 gene hits, respectively (Fig. 6c). Interesting enrichment of terms was seen in the molecular function domain in which DNA-binding transcription activator activity exhibited more than 100 gene hits followed by transcription factor activity with more than 75 gene hits (Fig. 6d). Gene concept networks of H3K9me3 peak-enriched genes also revealed many networks involved in processes other than spermatogenesis (Fig. S4). Enrichment map analysis of biological processes showed only two clusters, one of which was related to the embryonic development (Fig. S5). In the cellular component, the clustering of pre- and post-synaptic terms was observed (Fig. S5). Similarly, only one cluster related to the ion channel was observed in the molecular function domain (Fig. S5). Out of 1836 genes enriched in H3K9me3, only 20 genes appeared to have some role in fertility or reproduction (Table 2). All other genes (1816) were unrelated to spermatogenesis/fertility (Data S1).

## Bivalent (H3K4me3 and H3K9me3) Enriched Regions

The distribution of H3K4me3 and H3K9me3 marks are generally mutually exclusive in nature. The active regions exhibit H3K4me3 peak enrichment and the repressive regions show H3K9me3 enrichment. However, there are certain instances where both co-exist (Fig. 1), suggesting a poised state of genes. We found 53 genes which showed overlapping peak enrichment in H3K4me3 and H3K9me3 marks (Table S2). Upon functional mining of these genes, we found them to be unrelated to spermatogenesis, except a few which had roles in early and late spermatogenesis. A large number of these genes were found to be important for early embryonic development (Data S1).

## Spermatid Transcriptome Represents Most of the H3K4me3-Enriched Genes

ChIP-seq of H3K4me3 detected 3958 peaks corresponding to 2807 genes or genomic regions. We also undertook transcriptome sequencing in round spermatids, which identified about 11,000 transcripts. Upon comparing spermatid transcriptome with ChIP-seq data, we found 1800 genes to be common. Transcripts corresponding to 1800 out of 2807 (64.13%) genes suggested to be active by H3K4me3 marks were actually present in spermatids. This is a very high number as the expression of a gene is affected by a number of histone and DNA modifications. We pulled out the fragment per kilobase of transcript per million mapped reads (FPKM) values of these genes to assess their expression level in the round spermatids. We found that a large number of these

**Fig. 4** KEGG histogram for genes enriched in H3K4me3 marks showing molecular pathways (**a**), biological processes (**b**), cellular components (**c**), and molecular functions (**d**)
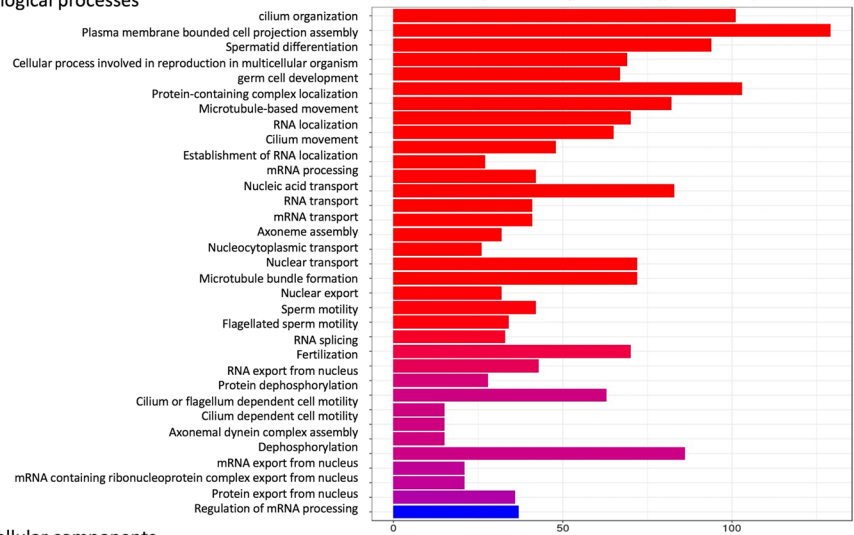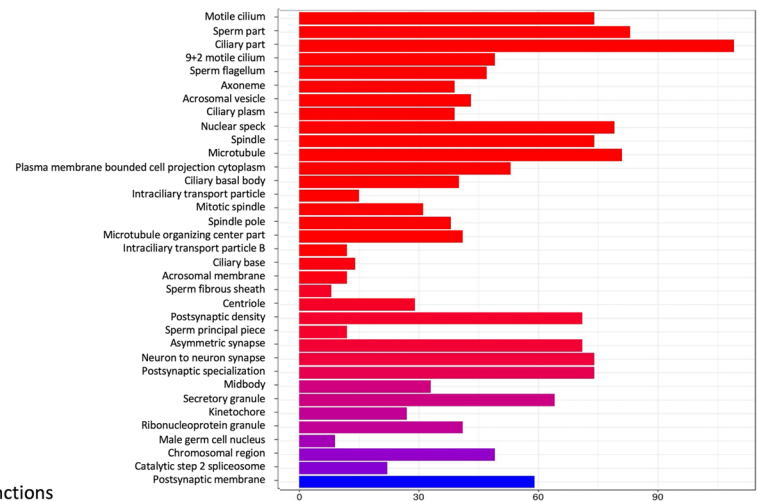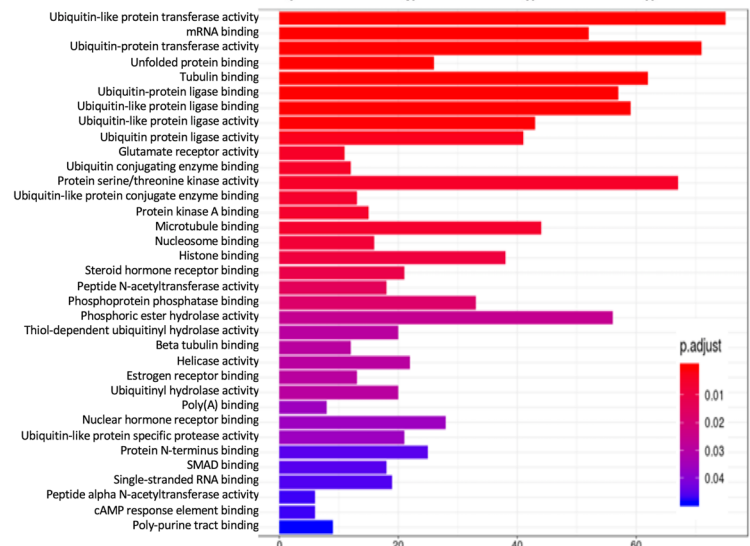
genes showed transcription at significantly high levels with the top FPKM value at 29,279 and 330 had FPKM above 100 and 902 transcripts had FPKM values above 10 (Fig. 7). This gene list consisted of top spermiogenesis candidates, such as *Acrbp, Fabp9, Catsper2, Catpser3, Catsper4, Ccdc42, Ccdc63, Ccdc181, Cftr, Dnah1, Dnah8, Dync1h1, Dynll1, Kdm3a, Gapdhs, H1fnt, Odf1, Odf3, Odf4, Oaz3, Tssk1b, Tssk2, Tssk4, Adam2, Adam26a/4, Adam3a, Tekt3, Tekt4,* and *Tekt5,* most of which encode structural and functional components of spermatozoa.

### H3K9me3-Enriched Genes Have a Scanty Presentation in Spermatid Transcriptome

ChIP-seq of H3K9me3 enriched regions detected 6352 peaks corresponding to 1836 genes or genomic regions. Upon comparing these genes with spermatid transcriptome data, we found transcripts corresponding to 413 (22.49%) genes out of 11,000 transcripts detected in the round spermatids. We also looked into the distribution of genomic features of these peaks and found that a very high number of peaks (75%) represented intronic regions of genes. We pulled out the FPKM values of these genes and found that the top FPKM value was relatively significantly low (3517 versus 29,279 in the case of H3K4me3 top transcript), with only 14 transcripts showing FPKM above 100 and 53 transcripts showing FPKM above 10 (Fig. 7). Therefore, H3K9me3 is associated with transcriptional silencing to a significant extent.

Genes with bivalent modifications were looked in the spermatid transcriptome for their expression. We found that 15 of 52 (28.8%) genes with bivalent modifications showed the actual presence of their transcripts in spermatids. Most of these genes were expressed at very low levels, with the top FPKM being 3320, which sharply dropped to 483 for the second candidate and there were only two genes having FPKM above 100 and only 5 genes having FPKM above 10 (Fig. 7).

### RNA Payloads of Sperm Are Set by Chromatin Activity in Spermatids

Since spermatozoa also carry a plethora of RNAs, we also undertook transcriptome sequencing in sperm prepared from rat epididymis. Since sperm are transcriptionally inactive, a large number of these RNAs may have origin in the round spermatids or predecessor cells [1]. We compared the list of the above genes (shared between H3K4me3-enriched regions and spermatid transcriptome) with sperm transcriptome. Interestingly, we found that sperm carried almost 100% (1788 out of 1800) transcripts corresponding to the active chromatin regions in the round spermatids as denoted by H3K4me3 marks. The top transcripts in

spermatozoa consisted of the same set of spermiogenesis-related genes as seen in spermatids, but with much lesser FPKM values (reduced 100 to 500 folds). This suggests that most of the active chromatin regions in spermatids generated their transcripts in the round spermatids, which also made their way to mature sperm.

## Discussion

This study focused on scoring chromatin activity in the round spermatids, particularly with respect to the role of H3K4me3 and H3K9me3 marks and found that genes participating in the formation of structural and functional components of sperm were enriched in H3K4me3 marks and those unrelated to spermatid differentiation were enriched in H3K9me3 marks. H3K4me3-enriched genes aligned to processes such as cilium development, spermatid differentiation, spermatid development, and germ cell development, which strongly support the idea that the genes participating in sperm formation have active chromatin state in the round spermatids. Functional annotation of these genes showed their roles in sperm development, sperm flagellum, 9 + 2 motile cilium, and cilliary parts. H3K4me3 peaks also showed enrichment of functions involving ubiquitin enzymes and mRNA binding. Interestingly, a study has identified more than 30 ubiquitinating enzymes as crucial regulators of spermatogenesis, and mRNA-binding proteins are required for their repression and active translation [22]. Interestingly, *Ube2b* knockout mice have been shown to present with male infertility, characterized by abnormalities in head shape and abnormal distribution of periaxonemal structures [23]. *UBE2B* mutations have been linked with human male infertility [24, 25]. The ubiquitin system has been shown to be critical for the morphogenesis and function of sperm organelles [26]. March7 ubiquitin ligase is highly expressed in the developing spermatids and is involved in tail formation.

A thorough literature search found that 134 genes enriched in H3K4me3 have established roles in sperm differentiation, development, and sperm maturation (Table 1). In particular, *Acrbp* and *Fabp9* play roles in acrosome formation, *Catsper2, Catpser3,* and *Catsper4* participate in hyperactivated sperm motility, *Ccdc42, Ccdc63,* and *Ccdc181* participate in sperm flagella formation, *Cftr* participates in capacitation, *Dnah1* and *Dnah8* participate in cilia formation, *Dync1h1, Dynll1,* and *Kdm3a* take part in chromatin condensation and manchette formation, GAPDHS is sperm-specific glycolysis protein, *H1fnt* encodes testis-specific histone, *Odf1, Odf3, Odf4,* and *Oaz3* encode sperm tail proteins, *Tssk1b, Tssk2,* and *Tssk4* encode kinases regulating sperm motility and fertility, *Adam2, Adam26a/4,* and *Adam3a* encode proteins for sperm binding to zona
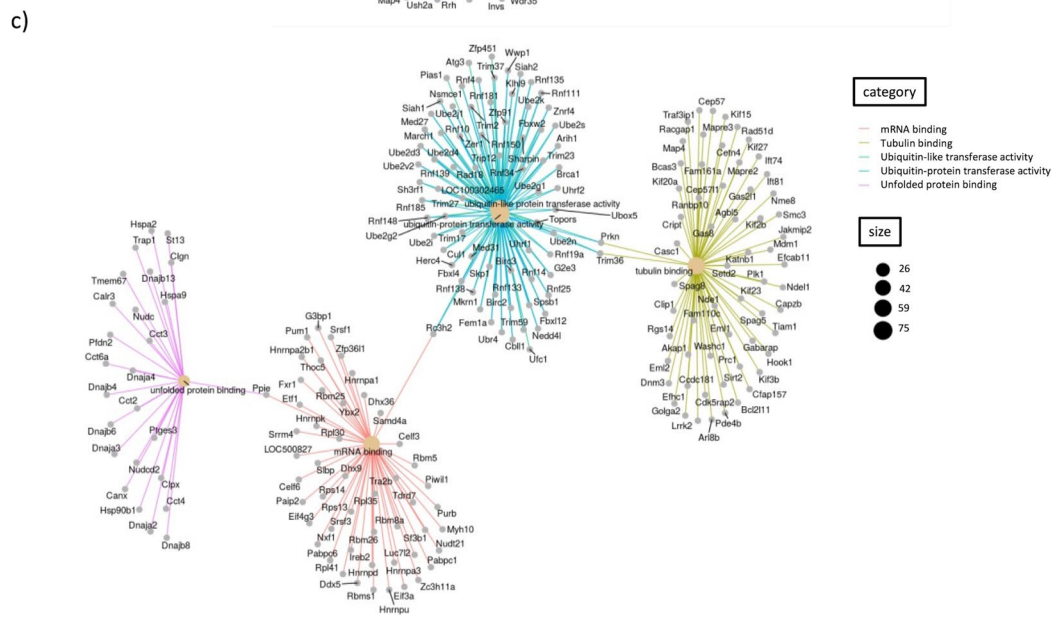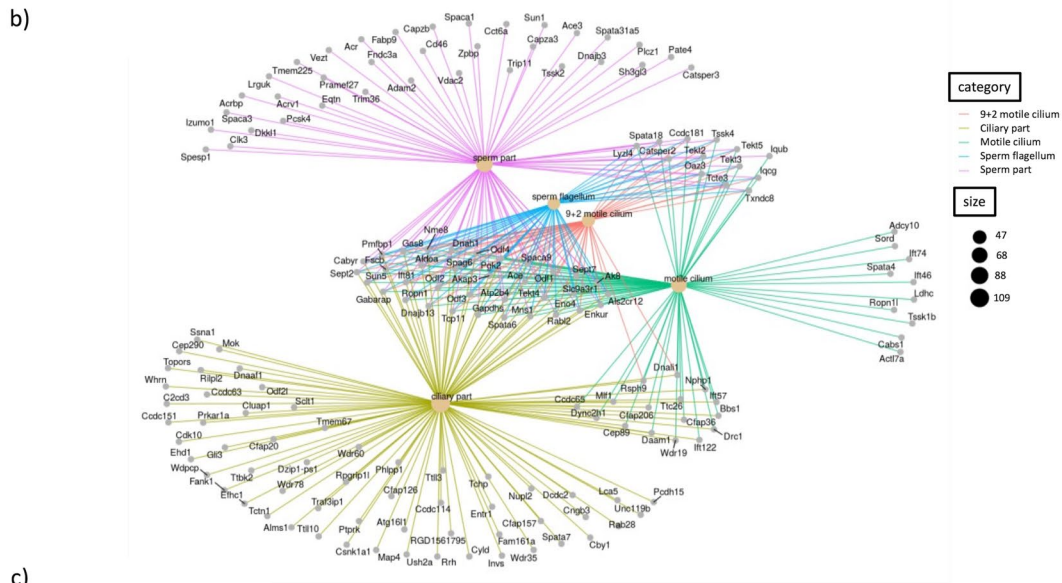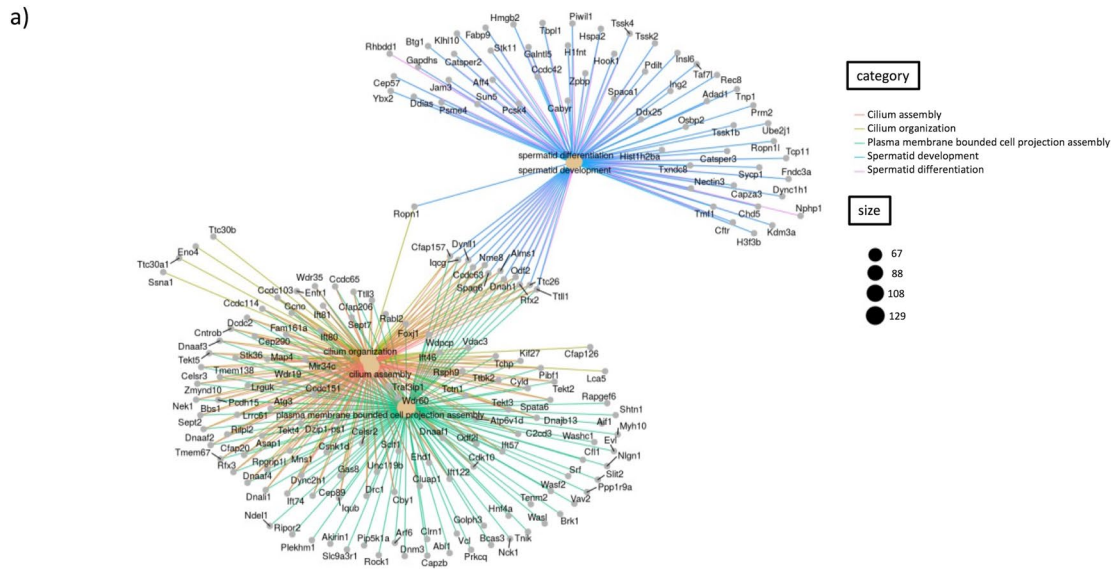
◄**Fig. 5** Gene-concept networks for genes enriched in H3K4me3 marks showing biological domains (**a**), cellular components (**b**), and molecular functions (**c**)

pellucida, and *Tekt3, Tekt4,* and *Tekt5* encode flagella formation proteins (Table 1). More than one thousand other genes which appeared in H3K4me3 ChIP profile, but do not have defined roles in spermatogenesis, are excellent candidates for investigation of their roles in the process of spermiogenesis.

Since H3K9me3 represents a repressive state of chromatin, we observed enrichment of GO terms, largely unrelated to spermatogenesis or fertility, but related to other distant processes such as neuron and synapse development. An appropriate level of H3K9me3 mediated gene inactivation of the pathways that do not participate in sperm differentiation is required for spermatogenesis to proceed normally. Accordingly, disturbances in this process have been found to affect spermatogenesis negatively. For example, a high level of H3K9me3 in germ cells has been linked to apoptosis [27] and pachytene spermatocytes lacking H3K9me3 show abnormal aggregation of chromosomes and improper synapsis [27, 28]. Various findings have shown that the removal of H3K9me3 enhances the efficiency of reprogramming by increasing the production rate and the number of pluripotent stem cells [29]. We found that 80% of the genes in the H3K9me3 list were not expressed in spermatids or showed very low expression. The H3K9me3 marks associated with the non-coding portions of the genome barricade them from transcriptional activation. This whole observation points to the fact that the presence of repressive signals during spermatogenesis is as important as that of the activation marks. The literature search identified four [30–33] and two mouse studies [34, 35] on H3K4me3 and H3K9me3, respectively, and the comparison of data identified H3K4me3 marks on 178 genes (having roles in cell differentiation) to be common across these studies while only 51 genes had common H3K9me3 marks (Fig. S6).

In comparison with a much higher percentage of genes with H3K4me3 and H3K9me3 marks, only a few genes (1%) had bivalent marks. Bivalent modifications are known to keep RNA polymerase II paused at the proximal promoter to maintain gene expression at low levels, but in the poised state. Epigenetic modifications are not only important in regulating gene expression during sperm differentiation, but also during the post-fertilization development of the embryos [36]. Accordingly, we found that most of the genes with bivalent modifications have critical roles in the post-fertilization development. The bivalent modifications may be a chromatin state to facilitate quick post-fertilization development by facilitating the process of confrontation and consolidation, which is followed by massive gene expression to support post-fertilization development. Earlier studies in sperm have shown that the co-occurrence of activation

(H3k4me3) and repressive marks (H3k27me3) represents the bivalent genes which includes many transcription factor families such as Hox, Gata, Tbx, Sox, and Pax that have crucial roles in early embryogenesis [37]. The scenario of histone modifications is beyond methylation and the real picture of chromatin state is a puzzling chimera of acetylation, ubiquitination, phosphorylation, and other modifications such as butyrylation, crotonylation, malonylation, propionylation, and succinylation, in which histones also affect each other in defining the chromatin state that condenses as spermiogenesis progresses [4].

The correlation of H3K4me3 marks with gene expression in spermatids is interesting. We found that about 64% of genes enriched in H3K4me3 marks had transcripts in spermatids. Out of 134 spermatogenically important genes identified by H3K4me3 ChIP-seq, 124 gene transcripts showed actual presence in the spermatid transcriptome. This indicates the importance of H3K4me3-based gene regulation in the round spermatids. A large number of these genes showed very high FPKM (expression) values in spermatids, with *Oaz3, Odf1, Smcp, Gsg1, Fam229b, Odf2, Irgc, Spata6, Spata18, Hmgb4, Lqcf3, Ubb, Spata20,* and *Gapx4* being some of the top candidates. Only 413 of H3K9me3 (22.49%) genes showed their transcripts in spermatids, and most of them were detected at very low levels. The top FPKM value for any H3K9me3-enriched gene was relatively significantly low (3517 versus 29,279 for top H3K4me3 transcript), with only 14 transcripts above 100 FPKM and only 53 transcripts with FPKM above 10, showing their low expression. Therefore, H3K9me3 is associated with transcriptional silencing in the round spermatids. Genes with bivalent modifications were looked in the spermatid transcriptome for their expression. We found that 15 of 53 (28.3%) genes showed the actual presence of their transcripts in spermatids. Most of these genes were expressed at very low levels, with the top FPKM being 483, which sharply dropped to 93 for the second candidate and there were only two genes having FPKM above 100 and only 5 genes having FPKM above 10.

We found that out of 1800 genes identified to be active in spermatids, the transcripts of almost 100% genes were detected in sperm, suggesting that most of the transcripts found in the round spermatids are used for encoding proteins during spermiogenesis and these transcripts also make their way to sperm. In general comparison of spermatid and sperm transcriptome, we found sharing of about 99% of transcripts between the two. The importance of the genes with active and bivalent chromatin state is also evident through clinical reports on the genes identified; for example, DPY19L2 [38] and MFSD14A [39] are associated with round head sperm, i.e., globozoospermia. Mutations in the GALNTL5 gene result in the reduction of sperm motility [40], PLCZ1 mutations hamper the oocyte activation ability and the initiation of embryonic development [41], PMFBP1 mutations associate with acephalic spermatozoa

**Table 1** H3K4me3-enriched genes (active) related to spermatid development, differentiation, spermatogenesis, and sperm structural and functional components. All references listed in the table are given in supplementary information

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Acrbp* | Acrosin-binding protein | Is a pro-acrosine protein | (1) |
| | | Functions in the formation and configuration of the acrosomal granule during early spermiogenesis | (2) |
| | | Binds to proacrosin and involved in packaging the acrosin zymogen into the acrosomal matrix | (3) |
| | | Functions in the formation and configuration of the acrosomal granule during early spermiogenesis and main function is to retain the inactive status of proacrosin in the acrosome until acrosomal exocytosis | (2) |
| *Adad1* | Adenosine deaminase domain-containing protein 1 | Plays an essential function in spermatid morphogenesis | (4) |
| *Aff4* | AF4/FMR2 family member 4 | Functions as a transcriptional regulator in testicular somatic cells and is essential for male germ cell differentiation and survival | (5) |
| *Alms1* | ALMS1, centrosome and basal body associated | Role of ALMS1 in appetite regulation, pathogenesis of the metabolic syndrome, pancreatic islet physiology, and spermatogenesis | (6) |
| | | Phenotypes of *Alms1* mutant mice include a lack of sperm flagella | (7) |
| *Btg1* | BTG anti-proliferation factor 1 | The BTG1 expression from the late prophase of the first meiotic division through spermatozoa release suggests BTG1 involvement in spermatogenesis. High levels of BTG1 mRNA at entry into terminal spermatid differentiation suggest a role consistent with that proposed for the BTG1 family of antiproliferative factors | (8) |
| *Cabyr* | Calcium-binding tyrosine phosphorylation-regulated protein | CABYR localizes to the principal piece of the human sperm flagellum in association with the fibrous sheath and is the first demonstration of a sperm protein that gains calcium-binding capacity when phosphorylated during capacitation | (9) |
| | | CABYR expresses at the late stage of spermiogenesis | (10) |
| *Capza3* | F-actin-capping protein subunit alpha-3 | Missense mutation in Capza3 is responsible for the mutant phenotype of sperm and regulation of F-actin dynamics by a spermatogenic cell-specific CAPZ heterodimer is essential for removal of the cytoplasm and maintenance of midpiece integrity during spermiation in the mouse | (11) |
| *Catsper2* | Cation channel sperm-associated protein 2 | Hyperactivated sperm motility driven by CatSper2 is required for fertilization | (12) |
| | | CatSper2 is responsible for driving hyperactivated motility, and, even with typical sperm forward velocities, fertilization is not possible in the absence of this highly active form of motility | (13) |
| *Catsper3 and Catsper4* | Cation channel sperm-associated protein 3 and 4 | Targeted disruption of murine CatSper3 cause disruption of sperm cell hyperactivated motility and male fertility but did not affect spermatogenesis or initial motility | (14) |
| *Ccdc42* | Coiled-coil domain containing 42 | Ccdc42 KO display defects in the number and location of the head–tail coupling apparatus, lack flagellated sperm, and are sterile | (15) |
| *Ccdc63* | Coiled-coil domain containing 63 | Ccdc63 removal resulted in sterile male mice due to shortened flagella and is essential for spermiogenesis | (16) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Cep57* | Centrosomal protein of 57 kDa | Study of Tsp57 expression in several purified subpopulations of spermatogenic cells confirmed maximum levels of expression in round spermatids. Consistently, Tsp57 expression was absent in testes from vitamin A-deficient mice, which do not have any round spermatids, and was reduced in RARalpha null mice, which have lowered numbers of round spermatids in their testes | (17) |
| *Cfap43* | Cilia- and flagella-associated protein 43 | Necessary to produce functional flagella in species ranging from Trypanosoma to human | (18) |
| *Cftr* | Cystic fibrosis transmembrane conductance regulator | Important for sperm capacitation | (19) |
| *Chd5* | Chromodomain-helicase-DNA-binding protein 5 | CHD5 is required for normal spermiogenesis, especially for spermatid chromatin condensation | (20) |
| *Chn2* | Chimerin 2 | Testis-specific and stage-dependent expression suggesting its role in acrosomal assembly at the late stage of spermatogenesis and cytoskeletal reorganization | (21) |
| *Ddias* | DNA damage-induced apoptosis suppressor | Loss or downregulation of Ddias leads to increased cell death and decreased number of spermatids | (22) |
| *Ddx25* | DEAD-box helicase 25 | (i) GRTH/Ddx25 is essential for completion of spermatogenesis, (ii) provides insights into intrinsic requirements for spermiogenesis, and (iii) establishes a model for studies of male infertility and contraception | (23) |
| *Dnah1* | Dynein heavy chain 1, axonemal | Cause primary cilliary dyskinesia and infertility | (24) |
| *Dpy19l2* | Probable C-mannosyltransferase DPY19L2 | DPY19L2 is associated with a complete form of globozoospermia | (25) |
| | | Patients with globozoospermia have a homozygous deletion of DPY19L2, thus indicating that DPY19L2 is necessary in men for sperm head elongation and acrosome formation | (26) |
| | | The absence of Dpy19l2 leads to the destabilization of both the nuclear dense lamina (NDL) and the junction between the acroplaxome and the nuclear envelope. Consequently, the acrosome and the manchette fail to be linked to the nucleus leading to the disruption of vesicular trafficking, failure of sperm nuclear shaping, and eventually to the elimination of the unbound acrosomal vesicle | (27) |
| *Dync1h1* | Dynein cytoplasmic 1 heavy chain 1 | Cytoplasmic dynein, possibly bound to the nuclear envelope, and manchette microtubules are involved in the protrusion of the spermatid nucleus from the cytoplasm | (28) |
| *Dynll1* | Dynein light chain LC8-type 1 | The spermiogenic stage–specific expression of DLC1 suggests a role for DLC1 in chromatin condensation, spermatid shaping, and the final release of sperm from the spermatogenic epithelium | (29) |
| *Eif4g1* | Eukaryotic translation initiation factor 4G1 | Double knockdown of eIF4G and eIF4G2 shows that these proteins act redundantly during the early stages of spermatogenesis suggesting their spatio-temporal roles of the canonical and testes-specific translation initiation factors in coordinating developmental programs during spermatogenesis | (30) |
| *Eqtn* | Equatorin, sperm acrosome associated | Eqtn $-/-$ males mated with WT females showed significant reduction in both fertility and the number of sperm attached to the zona-free oocyte | (31) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Fabp9* | Fatty acid-binding protein 9 | Protein may interact with the membrane of the acrosome and play a major role in the structural arrangement and stability of this organelle | (32) |
| *Fndc3a* | Fibronectin type-III domain-containing protein 3a | Some display degenerating pachytene spermatocytes and virtually no Sertoli cell vacuoles, while others display vacuoles without apparent elevated numbers of degenerating spermatocytes | (33) |
| | | Mutation that causes defective adhesion between spermatids and Sertoli cells within the seminiferous epithelium | (34) |
| *Galntl5* | Inactive polypeptide N-acetylgalactosaminyltransferase-like protein 5 | Genetic mutation of human GALNTL5 results in male infertility with the reduction of sperm motility and that GALNTL5 is a functional molecule essential for mammalian sperm formation | (35) |
| | | Protein expression was significantly higher in the high motility group than in the low motility group suggests that GALNTL5 is an important functional molecule during spermatogenesis | (36) |
| *Gapdhs* | Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic | GAPDHS is a sperm-specific glycolytic enzyme involved in energy production during spermatogenesis and sperm motility | (37) |
| *H1fnt* | Testis-specific H1 histone | A protein required for proper cell restructuring and DNA condensation during the elongation phase of spermiogenesis | (38) |
| *Hspa2* | Heat shock protein family A (Hsp70) member 2 | Participates in synaptonemal complex function during meiosis in male germ cells and is linked to mechanisms that inhibit apoptosis | (39) |
| *Ing2* | Inhibitor of growth protein 2 | This study establishes ING2 as a novel regulator of spermatogenesis functioning through both p53- and chromatin-mediated mechanisms, suggests that an HDAC1/ING2/H3K4me3-regulated, stage-specific coordination of chromatin modifications is essential to normal spermatogenesis, and provides an animal model to study idiopathic and iatrogenic infertility in men | (40) |
| *Insl6* | Insulin-like 6 | RNA analysis revealed a significant decrease in expression of late meiotic- and post-meiotic-specific marker genes, whereas expression of early meiotic-specific genes remains unaffected in the Insl6(-/-) testes | (41) |
| *Iqcg* | IQ domain-containing protein G | The axoneme in the Iqcg knockout sperm flagellum is disorganized and hardly any typical ("9+2") pattern of microtubule arrangement could be found in Iqcg knockout spermatids | (42) |
| | | Loss of Iqcg disrupts spermiogenesis such that tail formation either occurs incompletely or breaks apart from the sperm heads | (43) |
| *Jam3* | Junctional adhesion molecule C | Junctional adhesion molecule-C (JAM-C), is critically required for the differentiation of round spermatids into spermatozoa in mice. We found that Jam-C is essential for the polarization of round spermatids, a function that we attribute to its role in the assembly of a cell polarity complex | (44) |
| *Kdm3a* | Lysine-specific demethylase 3A | Essential for spermatogenesis, required for packaging and condensation of sperm chromatin | (45) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
| --- | --- | --- | --- |
| *Klhl10* | Kelch-like protein 10 | Klhl10 binds specifically to a unique testis-specific N-terminal Cullin-3 (TeNC) domain of Cul3(Testis) that is required for activation of effector caspase in spermatids | (46) |
| | | High-percentage chimeric and heterozygous males were completely infertile because of disrupted spermiogenesis characterized by asynchronous spermatid maturation, degeneration of late spermatids, sloughing of post-meiotic germ cells from the seminiferous epithelium, and marked reduction in the numbers of late spermatids | (47) |
| *Meioc* | Meiosis-specific coiled-coil domain-containing protein MEIOC | Meioc-deficient spermatocytes that have initiated synapsis mis-express CYCLIN A2, which is normally expressed in mitotic spermatogonia, suggesting a failure to properly transition to a meiotic cell cycle program | (48) |
| *Mfsd14a* | Hippocampus abundant transcript 1 protein | The few spermatozoa that are formed show rounded head defects similar to those found in humans with globozoospermia. Spermatogenesis proceeds normally up to the round spermatid stage, but the subsequent structural changes associated with spermiogenesis are severely disrupted with failure of acrosome formation, sperm head condensation, and mitochondrial localization to the mid-piece of the sperm | (49) |
| *Nme5* | Nucleoside diphosphate kinase homolog 5 | Gene disruption dysfunctional motile cillia | (50) |
| | | Gene plays an important role in late spermiogenesis by elevating the ability of late-stage spermatids to eliminate reactive oxygen species | (51) |
| | | Expression of Nm23-M5 plays a critical role in spermiogenesis by increasing the cellular levels of GPX-5 to eliminate reactive oxygen species | (52) |
| *Nme8* | NME/NM23 family member 8 | It is required during the final stages of sperm tail maturation in the testis and/or epididymis, where extensive disulfide bonding of fibrous sheath proteins occurs | (53) |
| *Nphp1* | Nephronophthisis 1 (juvenile) homolog (human) | Is critically required for the differentiation of early elongating spermatids into spermatozoa in mice. The roles of this gene are in the formation and maintenance of Sertoli-spermatid junctions are still under investigation | (54) |
| *Odf2* | Outer dense fiber protein 2 | Outer dense fibers (ODF) are specialized cytoskeletal elements required for the mammalian sperm tail formation, which are composed of several prominent proteins | (55) |
| *Osbp2* | Oxysterol-binding protein 2 | In ORP4-deficient testis, post-meiotic spermatids underwent extensive apoptosis, leading to a severely reduced number of spermatozoa | (56) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Pcsk4* | Proprotein convertase subtilisin/kexin type 4 | Lack of PCSK4 is associated with quantitative changes in the phosphorylation and proteolysis of sperm proteins during capacitation; therefore, alterations in signal transduction and proteolytic processing during capacitation may underlie the fertilization incompetence of PCSK4-null sperm | (57) |
| | | PCSK4-null sperm underwent capacitation at a faster rate; they were induced to acrosome react by lower concentrations of zona pellucida; and their egg-binding ability was only half that of wild-type sperm | (58) |
| | | Required during sperm capacitation and acrosome reaction | (59) |
| *Pdilt* | Protein disulfide-isomerase-like protein of the testis | Pdilt(-/-) mice were male infertile because ADAM3 could not be folded properly and transported to the sperm surface without the PDILT/CALR3 complex. Peculiarly we find that not only Pdilt(-/-), but also Adam3(-/-), spermatozoa effectively fertilize eggs when the eggs are surrounded in cumulus oophorus | (60) |
| *Piwil1* | Piwil1 | Mutations in Piwi (Miwi) knockin mice resulted in aberrant sperm showing histone retention, abnormal morphology, and severely compromised activity, which can be functionally rescued via blocking RNF8-MIWI interaction in spermatids with an RNF8-N peptide | (61) |
| | | miwi encodes a cytoplasmic protein specifically expressed in spermatocytes and spermatids. miwi(null) mice display spermatogenic arrest at the beginning of the round spermatid stage, resembling the phenotype of CREM, a master regulator of spermiogenesis | (62) |
| *Pld6* | Phospholipase D family, member 6 | The mutant mice display meiotic arrest during spermatogenesis, demethylation, and derepression of retrotransposons, and defects in primary piRNA biogenesis | (63) |
| | | Male mice deficient in either *Gykl1* or *Gk2, interactin partners* were infertile due to dysfunctional spermatozoa, which exhibited unregulated ATP production, disordered mitochondrial sheath formation, abnormal mitochondrial morphology, and defective sperm tail | (64) |
| *Psme4* | Proteasome activator complex subunit 4 | The KO sperms exhibited remarkable defects in motility, although most of them showed normal appearance in morphology | (65) |
| | | Targets the core histones for acetylation-mediated degradation by proteasomes, providing mechanisms by which acetylation regulates histone degradation, DNA repair, and spermatogenesis | (66) |
| *Ptchd3* | Patched domain-containing protein 3 | Male germ cell-specific gene and may be involved in the Hedgehog signaling to regulate sperm development and/or sperm function | (67) |
| *Rec8* | Meiotic recombination protein REC8 homolog | Impairment in the formation synaptonemal complex in spermatocyte | (68) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Rfx2* | DNA-binding protein RFX2 | The Rfx2-null round spermatids detached from the seminiferous tubules, forming large multinucleated giant cells that underwent apoptosis. In the mutants, formation of the flagellum was inhibited at its earliest stage | (69) |
| | | Rfx2-/- mice are perfectly viable but show complete male sterility, haploid cells undergo a complete arrest in spermatid development just prior to spermatid elongation | (70) |
| | | Rfx2gt/gt males are infertile because of a defect in spermatid maturation at or before the round and elongating spermatid stage | (71) |
| *Rhbdd1* | Rhomboid domain containing 1 | RHBDD1 knockdown GC-1 cells lost the ability to survive and differentiate in mouse seminiferous tubules | (72) |
| *Rimbp3* | RIMS-binding protein 3A | RIM-BP3 mutant mice display several common abnormalities, in particular with regard to the ectopic positioning of the manchette within the spermatid, a presumed cause of sperm head deformation | (73) |
| *Rnf38* | Ring finger protein 38 | Has an essential role in metaphase to anaphase transition during mouse spermatogenesis by regulating the expression of diverse families of genes | (74) |
| *Ropn1* | Rhophilin associated tail protein 1 | Defects in FS integrity, sperm motility, and PKA-dependent signaling processes, leading to male infertility | (75) |
| *Spaca1* | Sperm acrosome membrane-associated protein 1 | Disruption of Spaca1 led to the disappearance of the nuclear plate, a dense lining of the nuclear envelope facing the inner acrosomal membrane. This coincided with the failure of acrosomal expansion during spermiogenesis and resulted in the degeneration and disappearance of the acrosome in mature spermatozoa | (76) |
| *Spag6* | Sperm-associated antigen 6 | This is highly enriched in testis | (77) |
| *Spag6l* | Sperm associated antigen 6-like | This is critical for axoneme structure and function | (78) |
| *Stk11* | Serine/threonine kinase 11 | Male mice in which expression of Lkb1(S) is knocked out are sterile, with the number of mature spermatozoa in the epididymis being dramatically reduced, and those spermatozoa that are produced have heads with an abnormal morphology and are non-motile | (79) |
| *Sun5* | SUN domain-containing protein 5 | Essential role of SUN5 in anchoring sperm head to the tail and provide a promising way to treat this kind of acephalic spermatozoa-associated male infertility | (80) |
| *Sycp1* | Synaptonemal complex protein 1 | Essential for synaptonemal complex | (81) |
| *Taf7l* | TATA-box binding protein associated factor 7 like | The weight of Taf7l(-/Y) testis decreased and the amount of sperm in the epididymites was sharply reduced. Mutant epididymal sperm exhibited abnormal morphology, including folded tails. Sperm motility was significantly reduced, and Taf7l(-/Y) males were fertile with reduced litter size | (82) |

**Table 1**  (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Tbata* | Thymus, brain, and testes associated | Spatial in the cytosol of the early round spermatid. By the end stages of round spermatids, Spatial is concentrated at the opposite face of the acrosome near the nascent flagellum and in the manchette during the elongation process. Finally in mature sperm, Spatial persists in the principal piece of the tail. Moreover, we found that Spatial colocalizes with KIF17b, a testis-specific isoform of the brain kinesin-2 motor KIF17. This colocalization is restricted to the manchette and the principal piece of the sperm tail | (83) |
| *Tbpl1* | TATA box-binding protein-like protein 1 | TLF-deficient mice reveal defective acrosome formation in early-stage spermatids and display a fragmentation of the chromocenter | (84) |
| *Topaz1* | Testis and ovary specific PAZ domain containing 1 | Pachytene spermatocyte Topaz1–deficient become arrested before the post-meiotic round spermatid stage with extensive apoptosis. Consequently, an absence of spermatids and spermatozoa was observed in Topaz1(-/-) testis | (85) |
| *Tssk1b* | Testis-specific serine/threonine-protein kinase 1 | Phosphorylated form of this gene suggested having indispensable role in spermatogenesis. Tssk1 and 2 interact with each other in mouse and human testis | (86) |
| *Tssk2* | Testis-specific serine/threonine-protein kinase 2 | Phosphorylated form of this gene suggested having indispensable role in spermatogenesis | (86) |
| *Tssk4* | Testis-specific serine/threonine-protein kinase 4 | Male Tssk4 knockout mice exhibit an impaired sperm structure and reduced sperm motility, which affects male fertility | (87) |
| *Ttc26* | Tetratricopeptide repeat domain 26 | A mutation in Ttc26 leads to impairment in hedge-hog signalling which is a very important factor of spermatogenesis and sperm differentiation | (88) |
| *Ttll1* | Tubulin tyrosine ligase like 1 | Mice homozygous for the Ttll1 mutation exhibited male infertility along with normal biogenesis and function of sperm flagella | (89) |
| *Txndc8* | Thioredoxin domain-containing protein 8 | Overexpression of this gene is observed in morphologically abnormal human spermatozoa from infertile men | (90) |
| *Ube2j1* | Ubiquitin-conjugating enzyme E2 J1 | Male Ube2j1(-/-) mice are sterile due to a defect in late spermatogenesis. Ultrastructural analysis shows that removal of the cytoplasm is incomplete in Ube2j1(-/-) elongating spermatids, compromising the release of mature elongate spermatids into the lumen of the seminiferous tubule | (91) |
| *Ybx2* | Y box binding protein 2 | Spermatogenesis is disrupted in post-meiotic null germ cells with many misshapen and multinucleated spermatids, and no spermatozoa are detected in the epididymis | (92) |
| | | Incomplete nuclear condensation is prominent in later-stage spermatids at the time of massive spermatid loss | (93) |
| *Ace* | Angiotensin I converting enzyme | Germinal ACE is released from the testicular sperm membrane when sperm enter the epididymis hence may play role during sperm maturation | (94) |
| *Acr* | Acrosin | Extensively involved in sperm and zona pellucida reaction during fertilization | (95) |
| | | Involved in packaging of acrosomal matrix | (3) |
| *Acrv1* | Acrosomal protein SP-10 | Deficient male mice were sterile and exhibited a complete arrest in round spermatids | (96) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Adam2* | ADAM metallopeptidase domain 2 | Selective transport of specific sperm proteins including from the endoplasmic reticulum of testicular germ cells onto the cell surface | (97) |
| *Adam26a/4* | A disintegrin and metallopeptidase domain 26A (testase 3) | These proteins undergo processing during sperm maturation and knockout studies showed sperm having problem during fertilization | (98) |
| *Adam3a* | ADAM metallopeptidase domain 3A | Involved in the binding of sperm to the egg zona pellucida | (99) |
| *Ak8* | Adenylate kinase 8 | Modulates sperm flagellar movement | (100) |
| *Akap3* | A-kinase anchor protein 3 | Binds to sperm flagella and required for sperm motility | (101) |
| | | Binds to sperm flagella and required for sperm motility | (102) |
| | | Binds to sperm flagella and required for sperm motility | (103) |
| *Atp2b4* | Plasma membrane calcium-transporting ATPase 4 | Is highly enriched in the sperm tail, deletion of this protein is infertile due to severely impaired sperm motility | (104) |
| *Capzb* | Capping protein (actin filament) muscle Z-line, beta | Essential for removal of the cytoplasm and maintenance of midpiece integrity during spermiation in the mouse | (11) |
| *Ccdc181* | Coiled-coil domain-containing protein 181 | Essential for sperm motility | (105) |
| *Cct6a* | Chaperonin containing Tcp1, subunit 6a (zeta) | Required for sperm-zona pellucida interaction and for spermatozoa to gain their ability to fertilize | (106) |
| *Cd46* | CD46 antigen, complement regulatory protein | Crucial to sperm-egg interaction and fertilization | (107) |
| *Cfap69* | Cilia- and flagella-associated protein 69 | CFAP69 is necessary for flagellum assembly/stability and that in both humans and mice, biallelic truncating mutations cause autosomal-recessive multiple morphological abnormalities of the flagella and primary male infertility | (108) |
| *Dkkl1* | Dickkopf like acrosomal protein 1 | Involved in multiple processes during spermiogenesis, including the control of gene expression and the development or function of the acrosome | (109) |
| | | Involved in acrosome assembly or function | (110) |
| *Dnah8* | Dynein, axonemal, heavy chain 8 | Cause fishhook phenotype | (111) |
| *Dnajb13* | DnaJ homolog subfamily B member 13 | DNAJB13 participates in spermiogenesis and the motility of mature spermatozoa | (112) |
| | | Involved in assembly and stability of axoneme during sperm flagellum development | (113) |
| *Dnajb3* | DnaJ heat shock protein family (Hsp40) member B3 | Acrosome formation and centrosome adjustment during spermatid development | (114) |
| *Enkur* | Enkurin, TRPC channel interacting protein | Enkurin is an adaptor that functions to localize a $Ca2+$ sensitive signal transduction machinery in sperm to a $Ca2+$-permeable ion channel | (115) |
| *Eno4* | Enolase 4 | Required for normal assembly of the fibrous sheath and provides most of the enolase activity in sperm | (116) |
| *Fam170b* | Family with sequence similarity 170, member B | Anti-FAM170B antibody was remarkably found to inhibit murine in vitro fertilization, specifically blocking the acrosome reaction. We further determined that FAM170B interacts with GOPC (Golgi-associated PDZ and coiled-coil motif containing protein) during acrosome formation | (117) |
| *Fscb* | Fibrous sheath CABYR binding protein | Required for normal assembly of the fibrous sheath | (118) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| Gas8 | Growth arrest specific 8 | Gas8 production developmentally with the later stages of spermatogenesis and spatially with the sperm motility Apparatus, collectively suggest that this growth arrest-specific gene product may have a role in sperm motility | (119) |
| Glipr1l1 | GLI pathogenesis-related 1 like 1 | Role in the binding of sperm to the zona pellucida surrounding the oocyte | (120) |
| Hyal3 | Hyaluronidase-3 | Hyal3 null sperm showed delayed cumulus penetration and reduced acrosomal exocytosis | (121) |
| Ift81 | Intraflagellar transport protein 81 homolog | Helps RABL2 to delivers a specific set of effector (cargo) proteins to the sperm tail for assembly | (122) |
| Iqub | IQ and ubiquitin-like domain-containing protein | Plays a role in sperm motility | (123) |
| Izumo1 | Izumo sperm-egg fusion protein 1 | Mice without the cytoplasmic tail of IZUMO1 showed normal fertility but decreased the amount of protein | (124) |
|  |  | Has a role in sperm-egg membrane interaction | (125) |
| Ldha | Lactate dehydrogenase A | Has a role in the fibrous sheath along the length of the flagellum to provide a localized source of ATP that is essential for sperm motility | (103) |
| Lrguk | Leucine-rich repeat and guanylate kinase domain-containing protein | LRGUK-1 is required for basal body attachment to the plasma membrane, the appropriate formation of the sub-distal appendages, the extension of axoneme microtubules and for microtubule movement and organization within the manchette | (126) |
| Mns1 | Meiosis specific nuclear structural 1 | MNS1 is essential for spermiogenesis, the assembly of sperm flagella, and motile ciliary functions | (127) |
| Oaz3 | Ornithine decarboxylase antizyme 3 | Localized in the mammalian sperm tail, viz. the outer dense fibers and fibrous sheath, as well as to the connecting piece linking head and tail | (128) |
| Odf1 | Outer dense fiber protein 1 | Plays an important role in sperm tail structure and/or sperm movement | (129) |
| Odf3 | Outer dense fiber protein 3 | Plays an important role in sperm tail structure and/or sperm movement | (129) |
| Odf4 | Outer dense fiber of sperm tails 4 | Plays an important role in sperm tail structure and/or sperm movement | (129) |
| Pate4 | Prostate and testis expressed protein 4 | Functions related to fertility and reproduction | (130) |
| Pfkm | Phosphofructokinase, muscle | Functions not clear but found in principle piece region | (131) |
| Plcz1 | 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase zeta-1 | A deleterious mutation leads sperm inability to induce oocyte activation and initiate embryo development | (132) |
| Pmfbp1 | Polyamine-modulated factor 1-binding protein 1 | Important cause of infertility in men with acephalic spermatozoa syndrome | (133) |
| Racgap1 | Rac GTPase-activating protein 1 | Deletion in the germ cells prevented the formation of the intercellular bridges and induced a proliferation arrest leading to male sterility | (134) |
| Rae1 | Rae1 | Defects led to a failure to complete meiosis but allowed several aspects of spermatid differentiation to proceed, including axoneme formation and elongation | (135) |
| Sh3gl3 | SH3 domain containing GRB2 like 3, endophilin A3 | Involved in vesicular traffic in spermiogenesis by synaptojanin-clathrin complex | (136) |
| Slc2a3 | Solute carrier family 2, facilitated glucose transporter member 3 | Responsible for an energy supply to spermatocytes and spermatids | (137) |

**Table 1** (continued)

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Slc9a3r1* | Na(+)/H(+) exchange regulatory cofactor NHE-RF1 | Important for sperm capacitation | (19) |
| *Spaca3* | Sperm acrosome membrane-associated protein 3 | Binding of spermatozoa to the egg plasma membrane during fertilization | (138) |
| | | Required for fertilization | (139) |
| *Spaca7* | Sperm acrosome membrane-associated protein 4 | Protein released during acrosome reaction and involved in fertilization | (140) |
| | | Accelerates cumulus dispersal and facilitates fertilization | (141) |
| *Spag8* | Sperm-associated antigen 8 | SPAG8 acts as a regulator of ACT and plays an important role in CREM-ACT-mediated gene transcription during spermatogenesis | (142) |
| *Spata18* | Spermatogenesis associated 18 | *SPATA18* in the asthenozoospermic group may suggest important function of this gene in sperm motility and fertility | (143) |
| *Spata6* | Spermatogenesis-associated protein 6 | SPATA6 is involved in myosin-based microfilament transport through interaction with myosin subunits required for normal assembly of the sperm connecting piece and tight head–tail conjunction | (144) |
| *Spesp1* | Sperm equatorial segment protein 1 | SPESP1 is necessary to produce the fully 'fusion competent' sperm | (145) |
| *Sun1* | Sad1 and UNC84 domain containing 1 | Studies suggest that these SUN1 complexes connect the differentiating spermatid nucleus to surrounding cytoskeletal structures to enable its well-directed shaping and elongation, which in turn is a critical parameter for male fertility | (146) |
| *Tcte3* | T-complex-associated testis expressed 3 | Studies suggest that it may actively take part in the development of normal gamete function in males | (147) |
| *Tekt3* | Tektin 3 | Involved in acrosome-related events, such as the acrosome reaction or sperm-egg fusion | (148) |
| *Tekt4* | Tektin 4 | Is necessary for the properly coordinated beating of the sperm flagellum and male reproductive physiology | (149) |
| *Tekt5* | Tektin 5 | Plays an important role in flagella formation during spermiogenesis as well as being implicated in sperm motility | (150) |
| *Tmem190* | Transmembrane protein 190 | TMEM190 plays an indirect role(s) in sperm-oocyte fusion | (151) |
| *Tmem225* | Transmembrane protein 225 | Plays an important role in sperm degeneration but not in spermatogenesis | (152) |
| *Trim36* | Tripartite motif containing 36 | Plays a key role in the acrosome reaction and fertilization | (153) |
| *Trip11* | Thyroid receptor-interacting protein 11 | Development of the acrosome-acroplaxome complex, the head–tail coupling apparatus (HTCA), and the spermatid tail | (154) |
| *Tsks* | Testis-specific serine kinase substrate | Predicted to play an indispensable role during spermiogenesis and infertility due to haploinsufficiency | (155) |
| *Vdac2* | Voltage-dependent anion-selective channel protein 2 | Interacts with zona pellucida glycoproteins during fertilization and maintains proper function of sperm flagellum | (156) |

syndrome [42], and non-stop mutations in the MAGEB4 gene cause X-linked azoospermia and oligospermia [43]. Overexpression of the TXNDC8 gene results in abnormal shaped human spermatozoa [44] and decreased expression of HSPA2 is seen in infertile men with ([45] or without varicocele [46]. Nonetheless, the importance of the presence of many transcripts in sperm remains to be investigated, and their roles in the post-fertilization development remain a strong proposition.
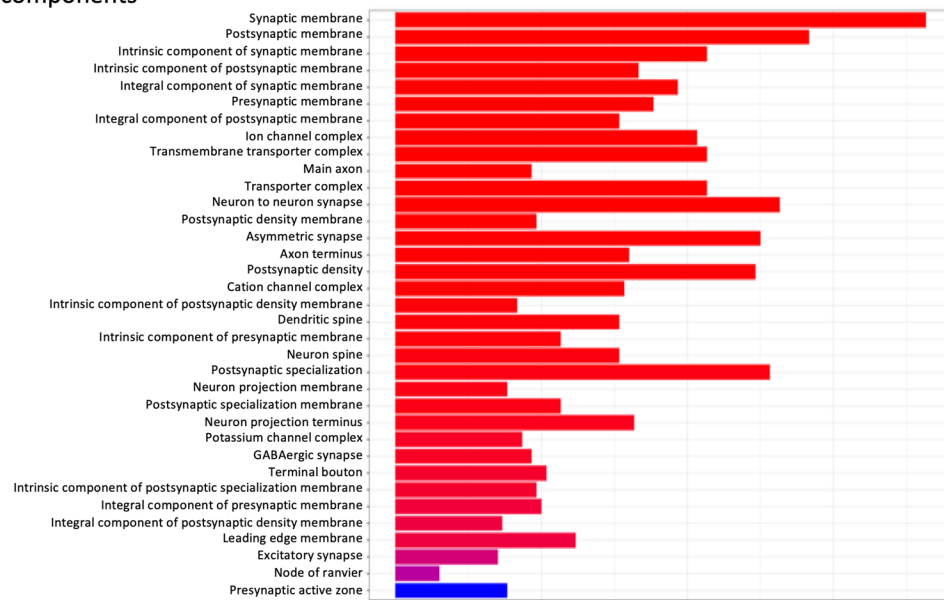
a  Molecular pathways

b  Biological processes
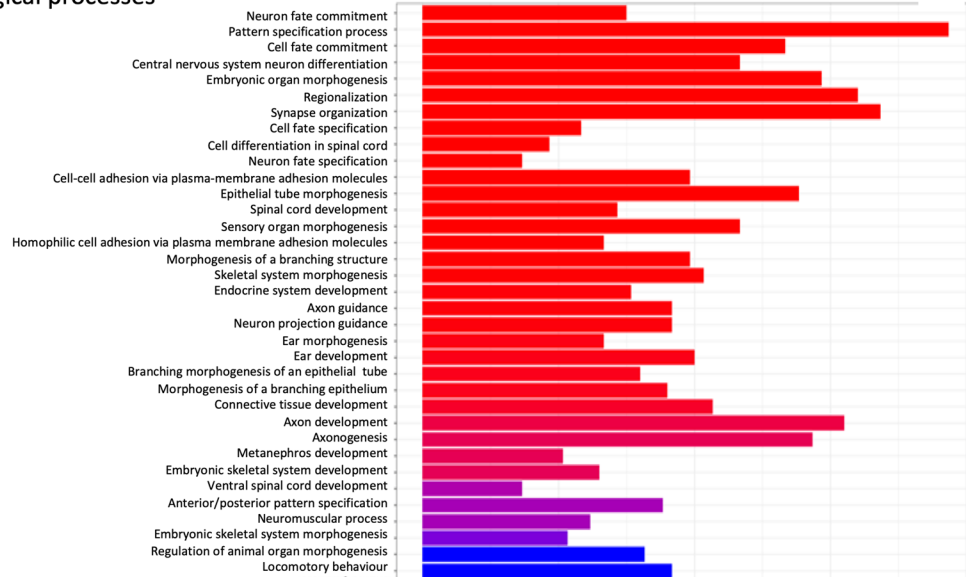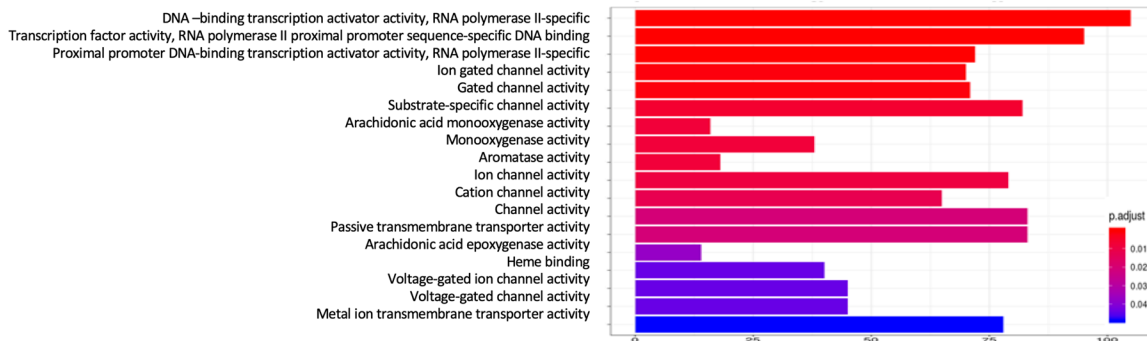
c  Cellular components

d  Molecular functions

◀**Fig. 6** KEGG histogram for genes enriched in H3K9me3 marks showing molecular pathways (**a**), biological processes (**b**), cellular components (**c**), and molecular functions (**d**)

This study has identified a strong correlation between H3K4me3 and gene expression and H3K9me3 and gene silencing. Sperm also retain a large number of histones, which do not get replaced by the protamines [16]. It would be interesting to figure out which post-translational modifications the retained histones carry. These histone marks and a heterogenous set of transcripts from the round spermatids guide further differentiation in the process of spermiogenesis. In most of the studies till date, H3K4me3 and H3K27me3 have been studied in combination and found that genes having both of these

**Table 2** H3K9me3-enriched genes (repressed) related to functions in spermatid development and spermatogenesis. All references listed in the table are given in supplementary information

| Gene symbol | Gene name | Function | Reference |
|---|---|---|---|
| *Bsph1/Bsph2* | Binder of sperm protein homolog 1 | Shares many biochemical and functional characteristics with BSP proteins secreted by seminal vesicles of ungulates, and might play a similar role in sperm functions | (157) |
| *Fndc3a* | Fibronectin type III domain containing 3A | Required during spermatid-Sertoli adhesion during mouse spermatogenesis | (34) |
| *Fshr* | Follicle-stimulating hormone receptor | Knockout of this gene results in smaller epididymis and causes low sperm count and motility | (158) |
| | | Knockout leads to delay in testicular development and compromised fertility | (159) |
| | | Genetic disruption of Fshr leads to inadequate condensation of sperm chromatin with poor quality of sperm and might contribute to reduced fertility | (160) |
| *Heph* | Hephaestus | Affects spermatid differentiation resulting in the accumulation of cysts with elongated spermatids without producing fully separated motile sperms | (161) |
| | | Necessary for proper spermatid individualization, the terminal step necessary for production of motile sperm | (162) |
| *Taz* | Tafazzin | Taz deficiency in Drosophila disrupts the final stage of spermatogenesis, spermatid individualization, and causes male sterility | (163) |
| *Adamts2* | ADAM metallopeptidase with thrombospondin type 1 motif, 2 | KO male mice exhibited marked decrease in testicular sperm and were infertile | (164) |
| *Bcl6* | BCL6, transcription repressor | Bcl6 may play a role as a stabilizer in protecting spermatocytes from apoptosis induced by stressors | (165) |
| *Ccnd2* | Cyclin D2 | Cyclin D2 have a non-cell cycle in mature Sertoli cells in the stage-dependent mechanisms regulated by AR expression and androgen action | (166) |
| *Gata4* | GATA binding protein 4 | GATA4 is a key transcriptional regulator of Sertoli cell function in adult mice | (167) |
| *Gja1* | Gap junction protein, alpha 1 | Gja1 (cx-43) expression in Sertoli cells is an absolute requirement for normal testicular development and spermatogenesis | (168) |
| *Gli1* | GLI family zinc finger 1 | GLI1 is the mediator of the Dhh signal in the testis, and that it may be a regulator of spermatogenesis | (169) |
| *Krt9* | Keratin 9 | Rat testis expresses Krt9 and that this protein is a component of the perinuclear ring of the manchette of rat spermatids | (170) |
| | | Loss results in sperm tail deformities | (171) |
| *Lrrk2* | Leucine-rich repeat kinase 2 | Plays role in embryonic germ cell development | (172) |
| *Morc1* | MORC family CW-type zinc finger 1 | Serves important biological functions in both meiotic and mitotic cells of multicellular organisms | (173) |
| | | MORC1 as an important new regulator of the epigenetic landscape of male germ cells during the period of global de novo methylation | (174) |
| | | Morc -/- males are infertile and have grossly reduced testicular mass and cause spermatogenic arrest | (175) |
| *Pou4f2* | POU class 4 homeobox 2 | Plays distinct and important functional roles in the regulation of gene expression during germ cell development | (176) |
| *Prdm9* | PR/SET domain 9 | Important for spermatogenesis and maintenance of fertility | (177) |
| *Spata16/20* | Spermatogenesis associated 16 | Plays a role in sperm maturation, fertilization, and/or embryo development, as has been shown in thioredoxin family | (178) |
| *Sstr2* | Somatostatin receptor 2 | Plays an essential role in testicular function | (179) |
| *Tp63* | Tumor protein p63 | Has an important role for maintenance of germ cell numbers, triggering or balancing the development, differentiation, and apoptosis of germ cells in the testis | (180) |
| *Wt1* | WT1 transcription factor | Promotes Sertoli cell-germ cell signaling events driving spermatogenesis | (181) |
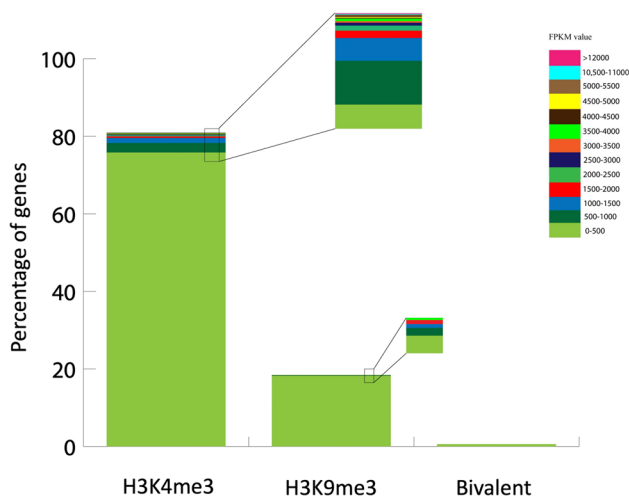
**Fig. 7** Bar diagram showing the percentage of H3K4me3, H3K9me3, and bivalent marks and the percentage of various transcripts in each category classified on the basis of expression (FPKM) values

represent a poised state in sperm, which may play roles in early and late embryonic development [16]. This is the first study showing that H3K4me3 and H3K9me3 bivalent modifications in germ cells/round spermatids not only correlate with gene expression, but also facilitate poised state as these genes may have roles in the post-fertilization development. The lack of replication experiments is a limitation of this study. Transcription is ultimately regulated by a number of mechanisms including histone methylation, but H3K methylation appears to be the major regulator of transcription in the round spermatids. Very likely, there are a number of other co-existent epigenetic modifications, giving rise to a multi-valent state that regulates the overall state of gene expression to guide spermatogenesis and post-fertilization development.

## Conclusions

H3K4me3 enrichment in the round spermatids correlates significantly with gene expression and H3K9me3 correlates with gene silencing. This ensures the expression of all essential genes and repression of all others. The bivalent state with these two histone marks keeps certain genes in a ready state as they may soon be required in the later stages of spermatogenesis or in the post-fertilization events. Generally, spermatids appear to be transcriptionally very active and show active and profuse expression of genes participating in sperm formation. Since the chromatin must start condensing, this may explain profuse gene expression activity seen in the round spermatids. Nevertheless, the level of transcription activity in further stages of spermatids remains to be investigated. The transcript content of the round spermatids is perfectly poised to facilitate chromatin condensation, sperm head and tail formation, sperm functional proteins

required for egg penetration, and other proteins participating in actual flagellar formation, beating for motility, and in generating the energy required for active motility. Almost all the transcripts seen in this stage of spermatids are seen in mature spermatozoa, whose functional significance would unfold many other mysterious arrangements in the complex process of spermatogenesis. Since methylation at K4 and K9 positions is not the only regulatory mechanism for gene expression, this ChIP-seq experiment may not have picked all active regions of the chromatin, which is also reflected in the number of matches between ChIP-seq peaks and the transcripts in RNA sequencing. In summary, the histone marks in the round spermatids define the battery of genes that show expression and pile up their transcripts for facilitating spermiogenesis, silence other genes that are not required during this stage, and prepare yet another set of genes (poised) for post-fertilization functions.

## References

1. Pandey A, Yadav SK, Vishvkarma R, Singh B, Maikhuri JP, Rajender S, Gupta, G. The dynamics of gene expression during and post meiosis sets the sperm agenda. Mol Reprod Dev 2019;
2. Kleene KCA. possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. Mech Dev. 2001;106:3–23.

3. Gatewood JM, Cook GR, Balhorn R, Schmid C, Bradbury EM. Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. J Biol Chem. 1990;265:20662–6.

4. Wang T, Gao H, Li W, Liu C. Essential role of histone replacement and modifications in male fertility. Front Genet. 2019;8(10):962.

5. Bao J, Bedford MT. Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. Reproduction (Cambridge, England). 2016;151(5):R55.

6. Yuen BT, Bush KM, Barrilleaux BL, Cotterman R, Knoepfler PS. Histone H3. 3 regulates dynamic chromatin states during spermatogenesis. Development. 2014;141(18):3483–94.

7. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem. 2012;81:65–95.

8. Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH, Roeder RG. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell. 2013;152:1021–36.

9. Flanagan JF, Mi LZ, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F, Khorasanizadeh S. Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature. 2005;438:1181–5.

10. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature. 2006;442:91–5.

11. Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, Reinberg D. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev. 2002;16:479–89.

12. Schneider R, Bannister AJ, Weise C, Kouzarides T. Direct binding of INHAT to H3 tails disrupted by modifications. J Biol Chem. 2004;279:23859–62.

13. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129:823–37.

14. Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol. 2003;13:1192–200.

15. Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. EMBO Rep. 2009;10(11):1235–41.

16. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. Nature. 2009;460:473–8.

17. Yadav SK, Pandey A, Kumar L, Devi A, Kushwaha B, Vishvkarma R, Maikhuri JP, Rajender S, Gupta G. The thermo-sensitive gene expression signatures of spermatogenesis. Reprod Biol Endocrinol. 2018;16:56.

18. Mercier E, Droit A, Li L, Robertson G, Zhang X, Gottardo R. An integrated pipeline for the genome-wide analysis of transcription factor binding sites from ChIP-Seq. PLoS One 2011;6:e16432

19. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31:2382–3.

20. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12:477–9.

21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

22. Hou CC, Yang WX. New insights to the ubiquitin-proteasome pathway (UPP) mechanism during spermatogenesis. Mol Biol Rep. 2013;40:3213–30.

23. Escalier D, Bai XY, Silvius D, Xu PX, Xu X. Spermatid nuclear and sperm periaxonemal anomalies in the mouse Ube2b null mutant. Mol Reprod Dev. 2003;65:298–308.

24. Suryavathi V, Khattri A, Gopal K, Rani DS, Panneerdoss S, Gupta NJ, Chakravarty B, Deenadayal M, Singh L, Thangaraj K. Novel variants in UBE2B gene and idiopathic male infertility. J Androl. 2008;29:564–71.

25. Yatsenko AN, Georgiadis AP, Murthy LJ, Lamb DJ, Matzuk MM. UBE2B mRNA alterations are associated with severe oligozoospermia in infertile men. Mol Hum Reprod. 2013;19:388–94.

26. Nakamura N. Ubiquitination regulates the morphogenesis and function of sperm organelles. Cells. 2013;2:732–50.

27. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell. 2001;107:323–37.

28. O'Carroll D, Scherthan H, Peters AH, Opravil S, Haynes AR, Laible G, Rea S, Schmid M, Lebersorger A, Jerratsch M, Sattler L, Mattei MG, Denny P, Brown SD, Schweizer D, Jenuwein T. Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. Mol Cell Biol. 2000;20:9423–33.

29. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell. 2012;151:994–1004.

30. Lesch BJ, Silber SJ, McCarrey JR, Page DC. Parallel evolution of male germline epigenetic poising and somatic development in animals. Nat Genet. 2016;48(8):888–94. Erratum in: Nat Genet. 2016;28;48(10):1296.

31. Wang X, Kang JY, Wei L, Yang X, Sun H, Yang S, Lu L, Yan M, Bai M, Chen Y, Long J, Li N, Li D, Huang J, Lei M, Shao Z, Yuan W, Zuo E, Lu K, Liu MF, Li J. PHF7 is a novel histone H2A E3 ligase prior to histone-to-protamine exchange during spermiogenesis. Development. 2019 10;146(13):dev175547. Erratum in: Development. 2020;147(8): PMID: 31189663.

32. Maezawa S, Hasegawa K, Yukawa M, Kubo N, Sakashita A, Alavattam KG, Sin HS, Kartashov AV, Sasaki H, Barski A, Namekawa SH. Polycomb protein SCML2 facilitates H3K27me3 to establish bivalent domains in the male germline. Proc Natl Acad Sci U S A. 2018;115(19):4957–62.

33. Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schübeler D, van der Vlag J, Stadler MB, Peters AH. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat Struct Mol Biol. 2013;20(7):868–75. Erratum in: Nat Struct Mol Biol. 2013 ;20(10):1236.

34. Liu Y, Zhang Y, Yin J, Gao Y, Li Y, Bai D, He W, Li X, Zhang P, Li R, Zhang L, Jia Y, Zhang Y, Lin J, Zheng Y, Wang H, Gao S, Zeng W, Liu W. Distinct H3K9me3 and DNA methylation modifications during mouse spermatogenesis. J Biol Chem. 2019;294(49):18714–25.

35. Bryant JM, Donahue G, Wang X, Meyer-Ficca M, Luense LJ, Weller AH, Bartolomei MS, Blobel GA, Meyer RG, Garcia BA, Berger SL. Characterization of BRD4 during mammalian postmeiotic sperm development. Mol Cell Biol. 2015;35(8):1433–48.

36. Teperek M, Simeone A, Gaggioli V, Miyamoto K, Allen GE, Erkek S, Kwon T, Marcotte EM, Zegerman P, Bradshaw CR, Peters AH, Gurdon JB, Jullien J. Sperm is epigenetically programmed to regulate gene transcription in embryos. Genome Res. 2016;26:1034–46.

37. Oikawa M, Simeone A, Hormanseder E, Teperek M, Gaggioli V, O'Doherty A, Falk E, Sporniak M, D'Santos C, Franklin VN,

Kishore K. Epigenetic homogeneity in histone methylation under-lies sperm programming for embryonic transcription. Nat Commun. 2020;11(1):1–6.

38. Harbuz R, Zouari R, Pierre V, Ben Khelifa M, Kharouf M, Coutton C, Merdassi G, Abada F. Escoffier J, Nikas Y, Vialard F, Koscinski I, Triki C, et al. A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. Am J Hum Genet 2011;88:351–361

39. Doran J, Walters C, Kyle V, Wooding P, Hammett-Burke R, Colledge WH. Mfsd14a (Hiat1) gene disruption causes globozoospermia and infertility in male mice. Reproduction. 2016;152:91–9.

40. Takasaki N, Tachibana K, Ogasawara S, Matsuzaki H, Hagiuda J, Ishikawa H, Mochida K, Inoue K, Ogonuki N, Ogura A, Noce T, Ito C, Toshimori K, Narimatsu H. A heterozygous mutation of GALNTL5 affects male infertility with impairment of sperm motility. Proc Natl Acad Sci U S A. 2014;111:1120–5.

41. Escoffier J, Lee HC, Yassine S, Zouari R, Martinez G, Karaouzene T, Coutton C, Kherraf ZE, Halouani L, Triki C, Nef S, Thierry-Mieg N, Savinov SN, Fissore R, Ray PF, Arnoult C. Homozygous mutation of PLCZ1 leads to defective human oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. Hum Mol Genet. 2016;25:878–91.

42. Zhu F, Liu C, Wang F, Yang X, Zhang J, Wu H, Zhang Z, He X, Zhang Z, Zhou P, Wei Z, Shang Y, Wang L, Zhang R, Ouyang YC, Sun QY, Cao Y, Li W. Mutations in PMFBP1 cause acephalic spermatozoa syndrome. Am J Hum Genet. 2018;103:188–99.

43. Okutman O, Muller J, Skory V, Garnier JM, Gaucherot A, Baert Y, Lamour V, Serdarogullari M, Gultomruk M, Ropke A, Kliesch S, Herbepin V, et al. A no-stop mutation in MAGEB4 is a possible cause of rare X-linked azoospermia and oligozoospermia in a consanguineous Turkish family. J Assist Reprod Genet. 2017;34:683–94.

44. Jimenez A, Zu W, Rawe VY, Pelto-Huikko M, Flickinger CJ, Sutovsky P, Gustafsson JA, Oko R, Miranda-Vizuete A. 2Spermatocyte/spermatid-specific thioredoxin-3, a novel Golgi apparatus-associated thioredoxin, is a specific marker of aberrant spermatogenesis. J Biol Chem. 2004;279:34971–82.

45. Esfahani MH, Abbasi H, Mirhosseini Z, Ghasemi N, Razavi S, Tavalaee M, Tanhaei S, Deemeh MR, Ghaedi K, Zamansoltani F, Rajaei F. Can altered expression of hspa2 in varicocele patients lead to abnormal spermatogenesis? International Journal of Fertility and Sterility. 2010;4(3):104–13.

46. Motiei M, Tavalaee M, Rabiei F, Hajihosseini R, Nasr-Esfahani MH. Evaluation of HSPA 2 in fertile and infertile individuals. Andrologia. 2013;45(1):66–72.