

# Genome-wide occupancy reveals the localization of H1T2 (H1fnt) to repeat regions and a subset of transcriptionally active chromatin domains in rat spermatids

Shalini Vasantha

Jawaharlal Nehru Centre for Advanced Scientific Research

Utsa Bhaduri

Jawaharlal Nehru Centre for Advanced Scientific Research

Anjhana C Ravikkumar

Jawaharlal Nehru Centre for Advanced Scientific Research

Anusha Rengarajan

Jawaharlal Nehru Centre for Advanced Scientific Research

Satyanarayana Rao ( [mrsrao@jncasr.ac.in](mailto:mrsrao@jncasr.ac.in))

Jawaharlal Nehru Centre for Advanced Scientific Research <https://orcid.org/0000-0003-0201-9437>

---

## Research

**Keywords:** Spermiogenesis, linker histone, spermatid, ChIP-sequencing, histone PTMs

**Posted Date:** November 12th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-38394/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on January 6th, 2021. See the published version at <https://doi.org/10.1186/s13072-020-00376-2>.

## Abstract

**Background** H1T2/H1FNT is a germ cell specific linker histone variant expressed during spermiogenesis specifically in round and elongating spermatids. Infertile phenotype of homozygous H1T2 mutant male mice revealed the essential function of H1T2 for the DNA condensation and histone to protamine replacement in spermiogenesis. However, the mechanism by which H1T2 imparts the inherent polarity within spermatid nucleus including the additional protein partners and the genomic domains occupied by this linker histone are unknown.

**Results** Sequence analysis revealed the presence of Walker motif, SR domains and putative coiled-coil domains in the C-terminal domain of rat H1T2 protein. Genome wide occupancy analysis using highly specific antibody against the CTD of H1T2 demonstrated the binding of H1T2 to the LINE L1 repeat elements and to a significant percentage of the genic regions (promoter-TSS, exons and introns) of the rat spermatid genome. Immunoprecipitation followed by mass spectrometry analysis revealed the open chromatin architecture of H1T2 occupied chromatin encompassing the H4 acetylation and other histone PTMs characteristic of transcriptionally active chromatin. In addition, the present study has identified the interacting protein partners of H1T2 associated chromatin mainly as nucleo-skeleton components, RNA binding proteins and chaperones.

**Conclusions** Linker histone H1T2 possesses unique domain architecture which can account for the specific functions associated with chromatin remodeling events facilitating the initiation of histone to transition proteins/protamine transition in the polar apical spermatid genome. Our results directly establish the unique function of H1T2 in nuclear shaping associated with spermiogenesis by mediating the interaction between chromatin and nucleo-skeleton, positioning the epigenetically specialized chromatin domains involved in transcription coupled histone replacement initiation towards the apical pole of round/elongating spermatids.

## Background

The production of mammalian spermatozoa is a precisely controlled, complex and dynamic developmental process involving the differentiation of diploid spermatogonia to spermatocytes and finally to haploid round spermatids. The last phase of spermatogenesis namely, spermiogenesis comprises a series of morphological transformations and extreme condensation of the chromatin to form highly condensed, polarized spermatozoa. Spermiogenesis is divided into different steps based on the three developmental processes involving condensation of the nucleus followed by the formation of the flagellum and the acrosome. During the steps 1-7, the early round spermatids are characterized by a round nucleus, with the beginning of acrosome and axoneme assembly. These early round spermatids actively transcribe many of the mRNAs necessary for the whole process of spermiogenesis, while many of these mRNAs are subjected to delayed translation until the later stages of spermiogenesis and sperm function. The step 8 comprises the acrosome polarization to one side of the nucleus and the initiation of spermatid elongation followed by the assembly of the accessory structures needed for flagella function.

The sperm transcripts remain as stable as ribonucleoprotein particles (RNPs) contributing to the final sperm proteome, which can have more relevant functional roles in the zygote or early embryo (1, 2). The final stage of spermiogenesis is the spermiation, a process in which elongating spermatids undergo extensive chromatin remodeling, which have been well studied in mouse (3).

Chromatin remodeling is a unique process occurring during spermiogenesis, wherein 90% of core histones are being replaced by the testes-specific histone variants, then transition proteins (TPs), and eventually protamines (PRMs) (4, 5). The process is initiated by the incorporation of testes specific histone variants and somatic histone variants, and concurrent histone post translational modifications (6, 7). Together with histone PTMs like acetylation, ubiquitination, methylation and phosphorylation, many variants of histone H1 (H1t, H1T2 HILS1), H2A and H2B (TH2A, TH2B, ssH2B, H2A.B.bd, H2BL1, H2BL2, H2AL1-H2AL3) and H3 (H3t, H3.3) play a major role in the extensive chromatin remodeling events leading to the generation of mature sperm (8, 9). Histone replacement starts with the hyperacetylation of histone H4 (lysine residues at K5, 8, 12 and 16) at step 8 just before the elongation of spermatids, and lack of H4 hyperacetylation leads to impaired spermatogenesis in mouse (10, 11). Transition proteins incorporation is observed in steps 10 and 11, followed by protamine 1 at step 11 and protamine 2 at step 12(12). In addition to the nuclear chromatin condensation, sperm head shaping is a tightly regulated process in spermiation, which involves accurate assembly of cytoskeletal (manchette) and cytoplasmic (acrosome) structures as demonstrated by many mouse genetic studies (13, 14). Until recently it was believed that transcriptionally inert genetic material is transferred to the embryo. However, more recent studies have suggested mature sperm as carriers of epigenetic information to the next generation by retaining typical covalent histone modifications at important genomic loci and a three-dimensional architecture similar to that of somatic cells (15, 16).

Many histone variants are expressed during spermiogenesis and modulate the chromatin structure to facilitate the histone-to-protamine replacement process (17, 18). Linker histones primarily function in the formation and stabilization of the higher-order chromatin structure (19). The H1 family proteins are greatly divergent in their structure, especially in the C terminal domain which influences the chromatin binding efficiency of these linker histones (20). The C-terminal domain of germ cell specific linker histones are more divergent with the formation of additional alpha helices in comparison to other H1s (21). Spermiogenesis begins with the extensive restructuring of sperm chromatin during the meiotic prophase interval when the somatic H1s are replaced by the variant H1t. H1t is subsequently replaced by another distinct linker histone H1T2 in round spermatids which is specifically localized at the apical pole in the nucleus of round and elongating spermatids (steps 5-15 of mouse spermiogenesis) which disappears during the maturation of testicular spematozoa. H1T2 has been identified as a critical histone variant necessary for spermiogenesis as the mutant males are infertile due to delayed nuclear condensation and aberrant elongation of spermatids (22, 23). HILS1 is another linker histone variant which is specifically expressed in elongating and elongated spermatids (steps 9-15 of mouse spermiogenesis). Earlier studies on linker histone variants were mainly restricted to biochemical and biophysical investigations which showed that they generate subtle different higher order structures *in vitro*. More recently, focused studies have been initiated to understand how the sequence diversity of H1

proteins leads to distinct chromatin structures associated with different functional properties. Studies from our laboratory recently focused on the mechanism by which testes specific H1 variants contribute to diverse chromatin function by studying their genomic localization, PTMs and chromatin loci specific interacting proteins (24, 25). Linker histone H1t constituting about 50–60% of total H1 content induces local chromatin relaxation to recruit the appropriate protein machinery for establishing closed chromatin repressed structures (25). An earlier study from our laboratory had shown that HILS1, whose expression is mainly confined to the elongating spermatids is involved in the maintenance of repressed chromatin architecture by preferential binding to LINE element containing chromatin domains in rat spermatids. Unlike H1t and HILS1 whose absence can be possibly compensated by other histone variants, H1FNT (H1T2) that appears in between H1t and HILS1 has been shown to be an essential linker histone for the packaging of sperm chromatin and for the formation of the compact sperm nucleus that is required for fertilization as revealed from the infertile phenotype in homozygous H1T2/H1FNT mutant males (23). In the present study, we have addressed the potential genomic function(s) of H1T2 by analyzing the genome wide occupancy of histone H1T2 and loci specific histone PTM signature and associated chromatin proteins.

## Results

### H1FNT/H1T2 – a highly divergent germ cell specific linker histone

Similar to all metazoan linker histones, H1T2 also possess a short, flexible N-terminal tail domain (NTD), a central globular domain (GD) with a winged-helix motif (three  $\alpha$  helix) and a highly extended C-terminal tail domain (CTD). Figure 1A summarizes the comparison of H1T2 sequences from different mammalian species. A comparison of sequences of H1T2 with H1.4 (somatic) or H1t (germ cell specific) reveals that H1T2 is more diverged ( $\leq 21\%$  similarity) from other linker histones. The human orthologue of H1T2 has highly diverged from rodents, with only 37.6% homology between mouse and human H1T2 sequence, 38.4% similarity between rat and human H1T2, while sharing most similarity with the gorilla orthologue with 97.6% identity. The highly extended C-terminal domain of H1T2 consists of a Walker motif (ATP/GTP binding), SR domain with SR repeats and many phosphorylation sites (Fig 1B). At the same time, H1T2 proteins lack the 16-mer (S/TPKK) motif, which is known to have DNA condensation properties (26). In the globular domain, human H1T2 shares aminoacid identity at several positions with the rodent specific H1T2, whereas the ATP binding motif and extended C-terminus including the SR domain are not present (Fig 1B). It is interesting to note that while the general property of the C-terminal domain of linker histone is to stabilize higher order chromatin folding, functional sub domains which are present in the CTD of rodent specific H1T2 are absent in higher mammals. Comparative study of testis protein evolution in rodents identified H1T2 as one of the rapidly evolving testis-expressed proteins (27). H1T2 possesses a putative winged helix domain consisting three alpha helices in addition to a unique coiled coil domain in the C-terminal domain (Fig 1C & D). A similar domain is reported in the N-terminal domain of Drosophila linker histone Mst77f, which is responsible for the DNA aggregation property of the H1 (28). Direct alignment of Mst77f with either HILS1 or H1T2 failed to uncover any significant sequence conservation with less than 15% identity, even though HILS1 has been referred in the literature as a

putative mammalian homologue of Mst77f (28). Extensive studies on *Drosophila* MSt77f revealed dual roles in spermatogenesis, firstly involved in regional distribution of chromatin beneath the stripe of microtubules in histone-based nuclei, and secondly, it is essential for nuclear shaping through the stabilization of microtubules (9, 28). We found a high degree of similarity between Mst77f and H1T2 proteins with respect to coiled coil domains. Numerous studies have identified H1T2 as a critical testes specific histone H1 for spermiogenesis, lack of which results in reduced fertility because of delayed nuclear condensation and aberrant elongation characterized by acrosome detachment and fragmented DNA (22). With this background, we were curious to gain more insights into the genomic functions of H1T2 in the context of our recent studies on the other two testes specific histone variants, H1t and HILS1 (24, 25).

### **Characterization of antibody against CTD of rat H1T2 protein**

Since H1T2 has been shown to occupy specialized chromatin regions in the apical region of spermatids, we wanted to analyze the genomic domains bound to H1T2 within these regions by ChIP-sequencing analysis. Because of the unavailability of a ChIP grade commercial antibody against the linker histone H1T2, we raised in house antibody against the C-terminal domain of rat H1T2, using the synthetic peptide (303-EQQYVSAKEQEYVRTKEQEC-321) (Fig 1B), which has shown earlier to be a strong immunogen (22). As expected, a strong immunoreactive signal was observed with dot blot analysis using the peptide as an antigen when probed with the affinity purified antibody (Fig 2A). Western blotting analysis was performed with the tissue lysate from both liver and testes of 35-50 days old rats. The antibodies reacted strongly with the testicular lysate with a distinct band at 54Kda. In contrast, the antibody failed to recognize any of the somatic H1 subtypes in the liver lysate, confirming the specificity of the antibody (Fig 2B). There was no signal observed in the peptide competition control lane, wherein the antibody was pre-incubated with 200-fold molar excess of the immunogen. H3 antibody was used as a positive control in these experiments, wherein we observed a positive signal for both liver and testes lysate. Western blot analysis was also performed with 25 and 50 day old rat testes acid extracts, wherein a strong H1T2 signal was observed with the 50day old rat testes extract (contributed predominantly by round/elongating spermatids). 10 day old rat testes lysate acted as a negative control for the experiment, which were contributed predominantly by somatic cells and spermatogonia (Fig 2C). Taken together, these experiments confirm that the antibody raised against the C terminal domain of rat H1T2 reacts specifically with the germ cell specific linker histone H1T2, and does not react with any other linker histone variants. Additionally, H1T2 signals were not detected in the histone extracts of mature sperm collected from rat epididymis (Fig 2D). Further immunolocalization studies of H1T2 in round and elongating spermatids revealed typical localization pattern with a cap like structure in the polar region of round spermatids, agreeing with the already established localization pattern of H1T2 signal at the nuclear region beneath the acrosome (Fig 2 E) (29).

### **H1T2 binds to distinct genomic domains in round/elongating spermatids**

ChIP-sequencing was performed using the highly specific H1T2 antibody in rat round/elongating spermatids chromatin. Chromatin samples were collected from 35-50 day old rat testes (with a major population being represented by round and elongating spermatids). Cross linked chromatin was subjected to sonication for 45 cycles (15 sec on/15 sec off cycle), followed by immunoprecipitation using H1T2 antibody. Western blot analysis confirmed the immunoprecipitation with a positive signal at 54 KDa both in the input and IP lanes, whereas the antibody was unable to recognize any protein in the peptide competition lane (antibody was competed with ~200 fold molar excess of the immunogen). Non specific IgG also acted as a negative control for the H1T2 immunoprecipitation experiment (Fig 2F). H1T2 immunoprecipitated DNA from the soluble extract of rat round/elongating spermatids was subjected to next-generation sequencing using Illumina HiSeq X Ten system using paired-end reads of 150nt length. After preprocessing the raw reads, the clean reads were aligned against UCSC *Rattus norvegicus* genome (rn6) using Bowtie2 (version 2.2.3) (Fig 3A). To analyze the extent of H1T2 occupancy in the spermatid genome, genomic regions associated to H1T2 was determined by MACS1.4.2 software with a p value cutoff of 1e-05. A total of 11570 enriched regions were obtained for H1T2 and the chromosome wise plots were generated (Fig 3B; Additional file 1: Table S1). The number of regions bound to H1T2 is maximally associated with chromosome 1 and 2, whereas the remaining peaks were evenly distributed across other chromosomes (Fig 3C; Additional file 2: Table S2). We have also analyzed the peak length distribution and fold enrichment across different chromosomes and the data is represented in the supplementary section (Additional file 3 & 4). To analyze the distribution of H1T2 peaks in various genomic domains of rat spermatid genome, H1T2 enriched genomic regions were annotated and categorized into CpG islands, repetitive elements, exons, introns, intergenic regions, 3'UTR and 5'UTR using HOMER v4.7 (Fig 3D). H1T2 enriched regions showed a bias towards the intergenic regions (71.56%), while at the same time a significant number of peaks were in the genic regions (promoter (2.94%), 5' UTR (0.47%), 3' UTR (0.22%), exon (2.43%), intron (19.27%). Next the repeat density in the H1T2 associated intergenic regions was calculated separately for interspersed repeat types including LINEs and LTRs (long autonomous retrotransposons) and SINEs (nonautonomous short <300 bp retrotransposons that co-opt the LINE transposition machinery) (30). The peak density is much lower in SINE/LTR regions compared to LINE rich region, with around 66% of H1T2 bound intergenic regions being LINE repeats (Fig 3E; Additional file 5:Table S5). Traditional views considered spermatozoa as transcriptionally silent cells, especially LINE repeats as the successful retrotransposons which is epigenetically repressed by CpG DNA methylation in conjunction with the piRNA pathway with multiple epigenetic mechanisms (31). Earlier studies from our laboratory had analyzed the epigenetic status of the LINE repeats at different stages of spermatogenesis in the context of linker histones occupancy. In pachytene spermatocytes, H1t decorates the methylated LINE and LTR repeats in addition to heterochromatin repressive marks (25). As the sperm differentiation progresses through histone hyper acetylation and incorporation of other histone variants, H1Ls1 gets incorporated to provide a more loosened chromatin to facilitate the histone replacement, still maintaining the repressed chromatin architecture of the majorly bound LINE repeats (24). An interesting observation from this study is the association of H1T2 to a subset of promoter-TSS chromatin domains in spermatids (~3% of total ChIP-seq reads). In contrast, only 0.6% and 0.08% of the ChIP-seq reads were associated with promoter regions

by H1t and HILS1 respectively (24, 25). Detailed annotation of the H1T2 bound LINE repeats revealed the major association to a particular class of LINE repeat, LINE L1 Rn (Fig 3F; Additional file 6: Table S6) similar to our previous observation with respect to HILS1 (24). Full-length rat L1s (L1Rn) with intact open reading frames (ORFs) is considered as the functional master copies for retrotransposition. In addition, a recent study has reported strong L1Rn RNA expression and hypomethylation status in rat testes (32). By using the motif analysis program MEME-ChIP, we identified three consensus motifs to be significant in the total H1T2 peaks (Fig 3G; Additional file 7 Table S7) and the percentage of occurrence of each of the motif were found to be 55.32% (Motif A), 30.54% (Motif B) and 41.95% (Motif C).

### Functional annotation of the H1T2 associated genomic domain of spermatid genome

It was most surprising that, ChIP-sequencing analysis identified 321 gene promoters as the potential binding sites for H1T2 predominantly between (+/-) 3Kb from transcription start site in rat round/elongating spermatids. When we carried out distribution analysis of our ChIP peaks within the +/-3kb region centered around TSS of genes, the distribution pattern turned into a sharp bell shaped curve, showing a peak at the TSS position (Fig 4A). To gain insights into these regions, we carried out gene ontology analysis of the genes associated with these promoters (Fig 4B; Additional file 8: Table S8). Maximum number of genes were classified to be involved in metabolic process associated with spermatogenesis (180 genes), which have been reported to be associated with the active mark H3K4me3 in both human and mouse spermatid genome (33, 34). The most significant category was found to be belonging to cellular component organization with a p value of 2.7E-0.5. Interestingly, among these we found three important genes namely *HMGB2*, *PYGO1* and *TBPL1* which are known to be required for the spermatid nuclear differentiation. *HMGB2* and *TBPL1* are already identified as essential chromatin organizers that are necessary for proper chromocenter formation and/or maintenance, as revealed from targeted mutagenesis studies (35). In another study, loss of *HMGB2* in mice affected H1T2 localization in late round spermatids and revealed intranuclear chromatin organization is critical for correct polar localization of H1T2 (29). Additionally we also found four other genes *ASH1*, *CECR2*, *BRD2* and *BRD4* belonging to bromodomain class of proteins, which are considered as the master regulators of maintenance of intact chromocenter in spermatids (Additional file 1: Table S1) (35).

Among the promoter centric H1T2 associated genes, we observed that 23 genes belonged to Transcription factor family upon GO: Molecular function analysis (Fig 4C; Additional file 9: Table S9). The identified occupancy peaks for these 23 transcription factors are shown in Figure 4D. As mentioned above, TATA box-binding protein-like 1 (TBPL1) is one among these, which has already been identified to be important for male sterility as evidenced by the sterile phenotype in *Tbp1*<sup>-/-</sup> male mice due to a late and complete arrest of spermiogenesis at step 7 in stage VII seminiferous tubules (36). *ING2*, another candidate transcription factor whose promoter is bound to H1T2 is reported to be an essential regulator of mammalian spermatogenesis which is characterized by an increased expression level in testes and infertile knockout male mice. Furthermore a decreased *ING2* expression level is also associated with defective spermatogenesis and male infertility in humans (37). *E2F6*, a part of the mammalian polycomb complex, is another transcription factor, expressed in the later stages of spermatogenesis, which has

been identified as a master regulator through the coordinated action as both repressor and activator depending on the stage of expression. During male germ cell meiosis, *E2F6* acts as an activator of meiosis-specific genes, demonstrated from moderately impaired spermatogenesis in *E2F6* knockout mice (38). *E2F6* recruits polycomb group proteins to function as a repressor of target genes during embryo development specifically associated to developmental patterning (39, 40). It is also essential for the long-term somatic silencing of certain male-germ-cell-specific genes and dispensable for cell-cycle regulation (41). *H1T2* associates with the promoter of *JMJD1C* gene, which plays an indispensable role in mouse spermatogenesis as *Jmjd1c*-deficient males became infertile (42). It contributes to the long-term maintenance of the male germ line by promoting spermatogonial stem cell self-renewal by up-regulating *Oct4* expression (43), required for mouse embryonic stem cell (ESC) self-renewal, mechanistically with the help of pluripotency factor **KLF4** and maintains ESC identity and **somatic cell** reprogramming. Surprisingly, the promoter of *KLF4*, a pluripotency transcription factor which is highly expressed in male post meiotic germ cells is also bound to *H1T2*, which is however, dispensable for spermatogenesis. It is interesting that lack of *KLF4* alone in male germ cells does not prevent spermiogenesis and male fertility (44). Deletion of *Klf4* in male mouse germ cells has been shown to affect the post meiotic transcriptome including defined transcriptional regulators and identified as a key factor necessary for postnatal development and differentiation of the mouse testes (45). *CAMTA1*, one among the CAMTA gene family of transcription factors whose promoter is occupied by *H1T2*, is an important gene in both spermatogenesis and embryo development. *CAMTA1* gene is important for various processes of fertility namely motility, capacitation, and acrosome reaction through calmodulin mediated calcium regulation. Additionally it **is important for embryonic heart development, which is crucial for proper embryogenesis (46)**. The TSS region of *CTCF*, an architectural protein governing chromatin organization in sperm genome is also associated with linker histone *H1T2*, whose inactivation in male germ cells (*Ctcf*-cKO mice) resulted in impaired spermiogenesis and infertility (47). Interestingly, this study has also shown that *H1T2* is one among those genes which are down-regulated in *Ctcf*-cKO mice and displays the same phenotypic features of *H1T2* deleted condition namely abnormal head morphology, aberrant chromatin compaction, impaired protamine 1 incorporation (47). Being a highly conserved Zn finger protein, *CTCF* is critical for the three-dimensional organization of the genome and related functions during the development of various cell and tissue types, ranging from embryonic stem cells and gametes, to neural, muscle and cardiac cells (48)

Recent evidences accumulating in the literature are suggesting that sperm is programmed to support proper embryonic expression of genes encoding important embryonic regulators and in this context spermatids contribute epigenetic information required for proper embryonic gene expression (49). Our GO enrichment analysis also showed enrichment of *H1T2* ChIP signals at the promoters of genes of several other developmental processes related pathways (Fig 4A; Additional file 8: Table S8): GO: 0044707~single-multicellular organism process (115 genes); GO: 0048856~anatomical structure development (105 genes); GO: 0044767~single-organism developmental process (106 genes); GO: 0050793~regulation of developmental process (43 genes). Some of the important candidate genes are

*TWIST2*, *TGLF1*, *CNOT3*, *SMARCC3* and *ZFP* whose functions in early embryonic differentiation process have been well documented (50-54).

## Validation of H1T2 ChIP peaks of novel target genes by ChIP-PCR

We selected a representative set of 15 ChIP peaks for further validation of H1T2 occupancy and the *IGV visualization* of MACS *ChIP peak* and quantitative real time ChIP PCR assays of these peaks are shown in Fig 5A & B. The Primer pairs designed across the H1T2 peak regions are given in the supplementary file (Additional file 14: Table S14). IP signals were normalized relative to the signal obtained from the input chromatin and enrichment was calculated as percentage of the input. As can be seen, ChIP PCR assay clearly revealed an enrichment over IgG-negative control for all the ChIP samples analyzed (Fig5A). Among these, we have validated 7 representative transcription factor bound regions including *TBPL1*, *ING2*, *JMJD2*, *HMGB2*, *SMARCC1*, *CTCF* and *E2F6* which confirmed the enrichment of H1T2 within the transcription start site regions of these regulatory genes which are either involved in spermatogenesis or embryogenesis (36, 55-58). H1T2 ChIP signals were also observed at loci of developmentally important genes including HOX (*HOXB3*, *HOXB8*, *HOXC5*); SOX (*SOX12*, *SOX4*, *SOX14*); FOX (*FOXB1*, *FOXG1*, *FOXC1*, *FOXE1*, *FOXK2*, *FOXJ1*, *FOXM1*); PAX (*PAXX*) and TBX (*TBX1*) gene clusters, where we observe at least one *ChIP-seq peak* located within 10 kb of the TSS (Additional file S1: Table S1). Representative genomic regions of the developmentally important loci, *FOXB1*, *FOXC1* and *SOX12* were also validated by ChIP PCR using H1T2 immunoprecipitated DNA as the template which confirmed significant enrichment of H1T2 in comparison to IgG immunoprecipitated chromatin control (Fig 5B). Earlier studies have also demonstrated the nucleosomal retention and bivalent histone marks (H3K4me3/H3K27me3) to be significantly enriched at these loci of developmental importance, including HOX, SOX, FOX, TBX, PAX, CDX, and GATA family transcription factors (33, 59). All together these results suggest a prominent role of the linker histone H1T2 in organizing specific spermatid chromatin domains to potentially regulate transcription of several developmentally important embryonic genes.

We also validated the occupancy of H1T2 with some representative examples of other classes of genomic loci like LINE elements, intronic, intergenic and CpG islands by ChIP PCR (Fig 5B). ChIP PCR analysis of the two representative loci of LINE repeat regions confirmed the association of H1T2 to the LINE (intergenic) regions which was the major fraction of genomic loci possessing H1T2 ChIP peaks, with significant enrichment in IP fractions in comparison to IgG pulldown fractions. Being the most successful retrotransposon, LINE elements presents an acute challenge to germ cells, which is managed by two repressive mechanisms of piRNA pathway and accumulation of repressive histone PTMs in addition to CpG methylation (31, 60, 61). LINE-1 (L1) subclass represents some 17–20% of the total human and mouse genomes, which seen to be still accumulating in the genome, as evidenced by the presence of L1 RNA and proteins in germ cells and infrequently in differentiated tissues (62, 63). Recently it has been observed that human L1 retrotransposition events likely occur in development rather than in the germline, using L1 RNAs transcribed in the embryo and in developing germ cells, suggesting active transcription of L1 in germ cells (64, 65). Additionally, representative loci of H1T2 bound regions like intron, intergenic and

CpG were also validated using specific primers targeting the peak regions compared to a negative control locus devoid of any ChIP signal (Fig 5B).

### Identification of proteins associated with H1T2 bound genomic regions by IP/MS analysis

To gain further insights into the H1T2 occupied genomic domains, we went ahead to characterize the functional protein complexes associated with H1T2 occupied genomic loci, by subjecting the chromatin IP samples to mass spectrometry analysis as detailed in the flow diagram in Figure 6A. Analysis of the proteome data identified 425 proteins with  $\geq 4$  unique peptides, which is usually considered to be significant (Additional file 10: Table S10). We found around 34 proteins with more than 15 unique number of peptides, which includes the proteins of cytoskeleton like actin, tubulin and intermediate filament proteins (plectin, desmin, lamin, vimentin), as listed in Figure 6B. Cytoskeletal dynamics in spermatogenesis is found to be regulated by the three networks (actin, tubulin and intermediate filament proteins) which function with each other to regulate different cellular processes, such as signaling, cell adhesion, cell motility, maintenance of cell polarity and protein targeting (66-70). The most significant H1T2-chromatin associated protein was myosin (Myh9) with 43 unique peptides, which is a reported component of acroplaxome which facilitates acrosomal spreading. Dynein (DYNC1H1) with 13 unique peptides (Additional file 10: Table S10) is also associated with H1T2-chromatin which is a component of manchette microtubules and spermatid nuclear envelope, pointing towards a potential function of H1T2 in nucleo-cytoskeleton interaction during spermiogenesis (14, 71, 72). Interestingly, H1T2 bound chromatin also interacts with lamin A/C protein encoded by *LMNA* gene, whose function is essential for cytoskeletal dynamics associated with spermatogenesis and its disruption leads to male infertility (73).

Functional annotation of the mass spectrometry identified proteins using DAVID program classified the proteins into the following groups: poly (A) RNA binding (133); ATP binding (48); structural constituent of ribosome (43); cadherin binding involved in cell-cell adhesion (31); mRNA binding (23); actin filament binding (17); ubiquitin protein ligase binding (22) respectively (Fig 6C; Additional file 11: Table S11). These groups have been summarized in Figure 6C. An important category of H1T2 bound chromatin interactome was ATP binding proteins (48 numbers), suggesting a potential role of H1T2 in regulating nuclear shaping involving the energy utilizing functional processes. Notably, ATP synthase subunits (ATP5B, ATP5F1, ATP5C1, ATP5O, ATP5A1, ATP5H) were also observed in this list, whose function is essential for germ cell maturation. Interestingly, their knockdown in *Drosophila* testes resulted in male infertility and abnormal spermatogenesis (74). We also carried out functional annotation of H1T2-chromatin interactome on the basis of biological process and the most significant categories are shown in the pie diagram (Fig 6C; Additional file 12: Table S12). There were 10 groups of proteins: metabolic process (303), cellular component organization or biogenesis (185), multi-organism process (75), detoxification (10), cellular process (356), reproductive process (52), localization (147), biological adhesion (50) and developmental process (141). Further, we also observed the association of the spermatid specific linker histone H1T2-bound chromatin with a large number of proteins involved in the sperm specific functions like sperm motility, spermatid differentiation and sperm-egg recognition (Fig 6D; Additional file 13: Table S13).

Male infertility has also been noted upon the ablation of several genes encoding RBPs (75). Indeed, RNA binding proteins were one of the major classes of proteins associated to H1T2 bound chromatin in spermatids (Fig 6C; Additional file 11: Table S11). The major function attributed to spermatid specific RBPs is the post-transcriptional regulation of mRNAs which are highly expressed in germ cells (75). Specifically, the mRNPs functions in the storage of silenced transcripts until factors in the elongation phase of spermiogenesis trigger their translation to support final steps in spermatozoan development and fertilization (76-78). The mRNA binding proteins interacting with H1T2 bound chromatin are listed in Figure 6D. Numerous RBPs are synthesized solely in late phases of spermiogenesis, ensuring a temporal regulation of their target mRNAs (79). Embryonic Lethal Abnormal Vision (ELAV) L1 is one such candidate RBP pulled down by H1T2 antibody which is reported as a key regulator of posttranscriptional regulation in spermatocytes, whose targeted deletion results in male sterility phenotype (80). More recent studies have supported the mRNA regulatory role of Elavl1 by specifically associating to *Tnp* mRNAs to control the delayed timing of their translation (81). H1T2-chromatin IP/MS analysis also identified YbX2, another potential mRNA repressor involved in the regulation of translation whose knockout resulted in male infertility and azoospermia condition (82). Notably we also found Ddx5 in our list of H1T2-chromatin interacting proteins, which is considered to be an essential protein for both transcriptional and posttranscriptional roles in the maintenance and function of spermatogonia (83). PABP1 was the other important protein pulled down with H1T2-chromatin, whose function has been extensively characterized in germ cells and shown to have important roles in mRNA stability and translation (84). Other candidate RBPs includes PSPC1, PTBP1, PTBP2, SFPQ, RBM14, SNRPs and HNRNPs. We went ahead to further validate the association of a representative set of proteins with H1T2 bound chromatin by western blot analysis (Fig 6E). More recently, LINE repeats have been reported to harbor repressive RBPs thereby contributing to the evolution of new, lineage-specific transcripts in mammals (85). In this context, we did observe a substantial overlap between the H1T2-chromatin interacting proteins and LINE1 ORF1 interactome which function in the regulation of retrotransposons (85, 86). It is worth noting here that LINE Rn\_L1 subclass of LINE elements that we have observed to be associated with H1T2, is indeed an active element containing ORF1 and ORF2 within the repeat (32).

### **Identification of PTMs of histones associated with H1T2 occupied chromatin domain**

To investigate the epigenetic status of H1T2 associated chromatin in terms of histone PTMs, acid soluble fraction was prepared from H1T2 immunoprecipitated chromatin and subjected to mass spectrometry analysis. Histone PTMs associated to these H1T2 bound chromatin loci were identified to be acetylation marks like H2BK21ac, H2AK95ac, H4K5ac, H4K8ac, H4K12ac, H4K16ac, H3K18ac, H3K23ac and H3K27ac and methylation marks like H3K27me2 and H3K79me2 (Fig 7A). Western blot analysis confirmed a combination of acetylated H4 peptides (H4K5ac, H4K12ac, and H4K16ac) which are activation marks to be highly enriched in the immunoprecipitated chromatin, whereas no signals were observed in IgG control samples (Fig 7B). Recent evidences have shown the involvement of H1T2 in the initiation of chromatin remodeling in spermatogenesis wherein H4 acetylation acts as a key to open up the chromatin in the doughnut structure in the apical region to facilitate the histone replacement in round spermatids (7). Another study has also revealed the enrichment of BRD4, H3K9ac, and H4K5ac, H4K8ac,

H4K12ac, and H4K16ac towards TSSs of active genes in round spermatids, and their enrichment correlates with transcription levels (87). Further analysis also revealed the enrichment of other active histone PTMs like H3K4me3, H3K9ac, H3K27ac, H3K4me1 and H3K79me2, whereas the repressive modifications like H4K20me3, H3K9me3, H3K27me3 were absent in the H1T2 immunoprecipitated chromatin fraction (Fig 7B). Incorporation of H1T2 in early round spermatids can be considered as an additional player in the chromatin reorganization process in addition to H3K79 methylation accompanying H4 acetylation as has been reported previously (88, 89). Notably, we also found H1T2 occupied chromatin domains to harbor the histone variants like *H3.3* and TH2B as demonstrated by IP/western blot. These core histone variants have been shown to be involved in the formation of transitional subnucleosomal structures during histone to transition proteins/protamine transition (90-93). Taken together, the results presented so far suggests a unique property to the spermatid specific H1, H1T2 to organize the polar domain of the spermatid genome by clustering the actively transcribed or to be transcribed genomic domains and in association with extensive number of chromatin interacting partners and cytoskeleton mediated mechanical force. H1T2 may therefore contribute to the chromatin remodeling process as well as nuclear shaping event that occurs during the elongation of spermatid nucleus in rodents.

## Discussion

Linker histone H1 was thought to be a chromatin organizer protein that is primarily involved in generation and stabilization of higher order structure of eukaryotic chromatin. But more recent studies have shown that the somatic subtypes as well as its tissue specific variants have many genomic functions beyond the architectural role (94, 95). Mammalian testes express three linker histone variants, histone H1t, H1T2 and H1LS1 in a stage specific manner. As mentioned earlier, gene knockout studies have shown that H1T2 is an essential gene for spermiogenesis and these mice show an infertile phenotype (23). A more interesting finding was the distinct localization pattern of the histone in the apical region of the spermatid nucleus showing a progressive change in the localization pattern and finally disappearing from chromatin during the final stages of spermiogenesis. H1T2 mutant mice also showed an abnormal chromatin condensation pattern with a defect in histone to protamine replacement process (22, 23). With this background available in the literature, we initiated the present study to identify and characterize the genomic chromatin domains that are occupied by histone H1T2. For a meaningful analysis of the genome wide occupancy analysis of histone H1T2, it is essential that the antibody used is of high specificity and does not cross react with other somatic as well as testes specific histone H1 subtypes. All the validation experiments carried out on the authenticity and specificity of the anti histone H1T2 antibodies that we have generated show that it is highly specific for H1T2, including the ChIP experiment. Furthermore, our immunofluorescence experiments also showed the typical polar localization pattern in the apical region of spermatids as has been reported earlier in the literature (Fig 2).

The results of our ChIP-seq analysis of H1T2 have shown a predominant occupancy at intergenic regions (71.56%), in addition, a significant number of reads was observed in genic regions (promoters (2.94%) and exons (2.43%) & introns (19.27%) (Fig 3D). A more in-depth analysis of the promoter occupied

regions showed a perfect overlap with TSS region of approximately 321 genes. Most of the genes belonged to the class of spermatogenesis related functions, including many germ cell specific transcription factors like *TBPL1*, *ING2* etc and also embryonic transcription factor like *HOXC5*. A surprising finding was the occupancy of H1T2 at the TSS site of *CTCF*, a global chromatin organizer that is involved in the enhancer-promoter interaction regulating the 3D-architecture of chromatin. It may be pertinent to point here that the spermatid genome is transcriptionally active at the stages between round and elongating spermatid stages (9, 96), during which time histone H1T2 makes its appearance. The association of H1T2 with transcriptionally active spermatid chromatin was also corroborated with the association of most of the histone PTM signatures characteristic of active chromatin. In addition, histone H4K12Ac mark was also associated with these domains as revealed by IP/MS and IP/WB analysis (Fig 7). This histone mark is prominent in the post acrosomal region of the spermatids and is also preferentially distributed in the promoter region of many genes relevant for early embryonic development (97). Additionally, we also found the association of H3K79 methylation with these chromatin domains, which has shown to be a conserved feature of spermatid chromatin preceding the process of histone to protamine transition which is in agreement with the role of H1T2/H1FNT in this process (Fig 7) (88, 89). These chromatin domains were also enriched with two core histone variants H3.3 and TH2B which are known to destabilize the nucleosome particle (93, 98). Most importantly, the repressive marks like H3K9, K27 and HK20 methylation were completely absent in these domains, thus strengthening the argument that histone H1T2 is associated with open transcriptionally active chromatin domain in the spermatid nucleus.

Polar localization of H1T2 is considered as a molecular signature for the chromatin organization in spermatids since there is a delayed nuclear condensation and morphological restructuring of sperm in H1T2 disrupted mutant mice (22, 23). It is also conjectured that this linker histone variant is linked to nuclear shaping during the elongation of spermatids, which is another major morphological event during this particular stage of spermiogenesis (99). In this context, our identification of many of the proteins belonging to the class of nucleoskeleton proteins to be associated with the H1T2 occupied chromatin gains significance suggesting that these chromatin domains are tethered to the nucleoskeleton structure during the spermatid elongation process. During nuclear shaping process, component of nuclear lamina play significant role, and it is very interesting that lamin A/C, a structural component of the acroplaxome (71), is one of the very significant proteins picked up in our ChIP/MS analysis. Lamin A/C is involved in the nucleocytoskeletal force transmission and essential for acrosome biogenesis and spermatid head shaping (73). Furthermore the identification of tubulin, actin and intermediate filament proteins in the proteome of H1T2 bound chromatin (Fig 6B), suggest the possibility of H1T2 not only essential for the maintenance of polarity in spermatids, but also involved in connecting the spermatid chromatin to the cytoskeleton in the sperm head and further confirms the inevitability of polar distribution of H1T2. It is very likely that the inherent property of H1T2 to be involved in this nuclear shaping process might be contributed by the highly extended C-terminus domain containing SR repeats, ATP binding Walker motif and the putative coiled-coil domains (Fig 1B). As mentioned earlier, similar domain architecture is present in Drosophila Mst77f protein, whose function is crucial for sperm development and nuclear

shaping (9, 28). Notably, all these domains in the extended CTD are absent in human H1T2 protein, which can be correlated with the existing observation of considerable divergence of H1 CTD throughout evolution accounting for the bulk of sequence heterogeneity between H1 variants (100). It would be an interesting area of research to understand whether the sequence divergence of sperm specific proteins involved in the chromatin remodeling along with the interacting partners can also be determinants of considerable diversity in sperm morphology (101).

Recently, several epigenomic studies have shown that DNA modification status, chromatin proteins and associated histone marks, as well as sperm derived RNA are passed on to the embryo, any disruption of which can lead to male infertility, abnormal embryo development and/or trans-generational inheritance (102). Many studies have also been focused on the role of sperm derived RNAs in embryonic development (103-105), of which one study identified ~18000 mRNA are delivered to the embryo. H1T2 ChIP peaks were enriched in some of the identified genes like *AKAP4*, *FOXG1*, *CLUSTERIN* and *WNT5a* genes whose mRNAs participate in the male pro-nucleus formation and development (Additional file 1: Table S1) (104). In addition, mature spermatozoa are identified as a source of LINE-1 RNA populations which are immediately available for retro transcription early after fertilization (104-107). Majority of the H1T2 bound genomic region was identified as LINE L1 (5480 peaks; Fig 3E), possibly embedded in the active chromatin domains. It is interesting to note that full length L1 transcripts are highly expressed in germ cells, especially ORF1 expression is well reported in mouse/rat spermatids (108, 109). In addition, LINE L1 elements have been reported to be under-methylated and histone retained, specifically in the nuclear periphery where the pronuclear reverse transcription occurs in early development (107, 110-112). We found many LINE L1 specific RBPs including MATR3, ELAVL1, PTBP1 and HNRPs in our H1T2 IP/MS analysis listed proteins, which are considered as repressive RBPs which repress splicing events and 3' end processing within and around LINEs, thereby contributing to the evolution of new, lineage-specific transcripts in mammals (Fig 6D) (85). It is of considerable interest that many of the proteins that co-IP with H1T2 bound chromatin, can also be identified in the L1 ORF1 and its ribonucleoprotein interactome identified by mass spectrometry in a recent study (86). IP/WB analysis also validated several of proteins involved in splicing, mRNA export, and transcript stability like Ddx5, Ybx2, Elavl1, Pabp1 and Iif3, known to be associated with transcriptionally active chromatin domain. In addition, we also observed Srsf3 and Ruvbl1 in our proteome analysis which have been implicated recently as sperm-carried molecular factors involved in oocyte fertilization and embryo growth (Fig 6E) (113).

The process of spermatogenesis is a very complex, developmental and differentiation process involving several facets of remodeling that is influenced by the testes specific core histone and linker histone variants which appear in a definite and stage specific manner (Fig 8A). As mentioned earlier, the three mammalian testes specific linker histone variants, namely H1t, H1T2 and HILS1 appear in a sequential and stage specific manner during spermatogenesis in the testes with an overlapping presence (H1t-H1T2 and H1T2-HILS1) having unique chromatin organization property and putative specialized genomic function. We have recently delineated the genomic functions of H1t and HILS1 (24, 25), which revealed the preferential association of these two H1s to the repressed chromatin loci of spermatocytes, spermatids respectively. The results presented in this investigation have clearly shown the unique

property of H1T2, primarily occupying genomic regions which are actively transcribed during the elongating stage of spermiogenesis probably facilitating the process of transcription of many spermatogenesis related genes and also few of the genes relevant for early developmental process. This observation may look surprising since histone H1 is normally associated with repressed chromatin domains. However, there are some examples in the literature wherein multiple functions have been attributed to histone H1 not simply as chromatin architectural protein, by functioning as a part of multi-protein complexes (94, 95). The open chromatin structure of H1T2 occupied polar spermatid genome may also serve as a prerequisite to initiate the global chromatin remodeling process involving histone to transition proteins/protamine transition. Another important observation of this study is the association of the H1T2 occupied spermatid chromatin with many of the nucleo-skeleton components, which strongly suggests the possibility that this tethering event is possibly related to the nuclear shaping process which is a characteristic feature of the elongation of spermatid nucleus. We have summarized our findings in the schematic diagram depicted in Fig 8B, in the context of events that characterize the major chromatin remodeling process that occur during mammalian spermiogenesis.

## Conclusions

Histone H1T2 is the most divergent among the three testes specific histone H1s containing several important domains like SR domains, Walker motif, and coiled-coil domains. The three testes specific linker histone variants differ considerably in the N-terminal and C-terminal domain sequences. The C-terminus in particular is the most divergent and it will be challenging to address the role of these sequence variations among these linker histone variants in facilitating the necessary chromatin remodeling events during spermatogenesis. It would be very interesting to see the function of these individual sub-domains in many of the chromatin associated functions. Surprisingly, many of these sub-domains are absent in human H1T2, the significance of which will be an intriguing question to address. In conclusion, the present study has identified the genomic domains occupied by histone H1T2 and characterized many of the chromatin features in terms of associated proteins and the core histone modifications, which should form the basis of future investigations in delineating the molecular and cellular mechanisms of global chromatin remodeling that is a unique and characteristic feature of spermatogenesis.

## Methods

### Analysis of H1T2 sequence conservation and protein structure

Amino acid alignments were performed online using Clustal X, a graphical interface for the ClustalW multiple sequence alignment program of Jalview (<http://www.jalview.org>). Prediction of coiled coil structures was predicted using the online tool ([https://embnet.vital-it.ch/cgi-bin/COILS\\_form\\_parser](https://embnet.vital-it.ch/cgi-bin/COILS_form_parser)) (114, 115). Secondary structure of rat H1T2 sequence was predicted by Phyre2 (116).

### Antibody generation

To generate antibody against the CTD of rat H1T2 protein, rabbits were immunized with synthetic peptide, EQQYVSAKEQEYVRTKEQEC and a 12 weeks protocol was used for the generation of the antibody. The custom made antibody from the Abgenex Company (Bhubaneshwar, India) was used for the present study.

### **Extraction and purification of histones from rat tissues**

Rat testes (35-50 days old) were decapsulated and homogenized in 5 volumes of cell lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Nonidet P-40, 1 mM PMSF, 10 mM NaHSO<sub>3</sub>, 10mM sodium butyrate, protease inhibitor mixture, phosphatase inhibitor mixture). The homogenate was filtered through four layers of bandage cloth and centrifuged at 1,500 X g for 10 min at 4°C. Histones were extracted from the pellet of crude nuclei by suspending purified nuclei in 0.4N H<sub>2</sub>SO<sub>4</sub> as prescribed previously (117), homogenized and incubated on ice for 30 min. The sample was centrifuged for 10 minutes; 8,000g at 4°C to pellet the residual chromatin. The extracted proteins were precipitated at 4°C from the supernatant by making it to 30% with respect to TCA. After keeping on ice for 30 minutes, the proteins were recovered by centrifugation. The pellet of proteins so obtained was sequentially washed with cold acetone containing 0.05% HCl and ice cold acetone. The pellet recovered was dried, dissolved in water and stored in aliquot at -20°C.

### **Dot blot analysis**

Two microgram of the synthetic peptide (303-EQQYVSAKEQEYVRTKEQEC-321) was spotted on the nitrocellulose membrane, dried and subjected to western blot analysis using the anti-H1T2 antibody.

### **Western blot analysis**

The proteins resolved by SDS-PAGE were electrophoretically transferred onto nitrocellulose membranes presoaked in transfer buffer, using a semidry transfer technique. After the transfer, the membrane was blocked using the wash buffer, PBST (1x PBS with 0.05% Tween 20), containing 5% BSA or milk powder for 1 hour at room temperature. It was then washed with the same buffer lacking BSA or milk powder. A suitable dilution of the primary antibody was added in PBST supplemented with 1% BSA or milk powder and incubated for overnight on a rocking platform. The unbound antibody was removed by washing the membrane thrice for 5 minutes each using wash buffer. A suitable dilution of an appropriate secondary antibody (anti-rabbit/anti-mouse IgG) conjugated to horse raddish peroxidase (HRP) in PBST with 1% BSA or milk powder was added and incubated for 1 hour. After the antibody incubation, the membrane was washed thrice with the wash buffer for 5 minutes each. The membrane was placed in the chemiluminescence substrate solution and incubated for 2 minutes and developed using the ECL kit (Thermo Scientific). For the peptide competition assay, 50-fold molar excess of the peptide were added to the antibody solution and mixed for 3 h at 4°C before the addition to the blot.

### **Decondensation and immunofluorescence of testicular cells**

Rat testes were decapsulated and total testicular cells were decondensed by resuspending in decondensation buffer (PBS, 10mM dithiothreitol) and incubated at 4°C for 1 h in an end to end rotor (Kolthur et al., 2009). Cells obtained after centrifuging at 1,500g for 10 min at 4°C, were washed with PBS and fixed with 4% paraformaldehyde (in PBS) and permeabilized with 0.1% Triton X-100 (in PBS) for 15 minutes. Non-specific sites were blocked with 1% BSA in PBS. Smears were incubated with selected primary antibodies, anti-H1T2 and counterstained with corresponding secondary antibodies conjugated with Alexa fluor 488 or Alexa Fluor568. Nuclei were stained with DAPI (Sigma) and images were acquired and analyzed in a Zeiss confocal laser scanning (LSM 510 META, Carl Zeiss) microscope.

### ChIP-sequencing analysis of H1T2 bound chromatin and Data analysis

Chromatin immunoprecipitation (ChIP) was carried out using the published protocol (118). Briefly, the nuclei were prepared from 35-50 days postnatal rat testes and decondensed using 10 mM DTT, and were fixed using 1% formaldehyde for 10 min at room temperature, quenched with 125 mM glycine for 5 min, harvested and sonicated to generate fragments of 100 to 300 bp. For ChIP, 25 µg of sonicated chromatin was immunoprecipitated overnight with 5 µg anti-H1T2 antibodies. Immunocomplexes were captured on 50 µl protein A Dynabeads. Beads with bound immunocomplexes were washed, eluted and reverse cross-linked at 65°C overnight, and DNA were extracted by phenol–chloroform method. Input was prepared with 5% of the sonicated chromatin.

The input and ChIP DNA libraries were prepared using the NEBNext Ultra II DNA library preparation kit. Sequencing was performed with Illumina HiSeq X Ten system using paired-end reads of 150nt length. Sequenced reads containing adaptor sequences were trimmed out and filtered with Phred quality threshold as 20. Read pairs falling below 20 bp after clipping were removed. Clean reads were uniquely aligned against UCSC *Rattus norvegicus* genome (rn6) using Bowtie2 (version 2.2.3) with “no mixed”, “very sensitive” and “no discordant” parameters. All aligned files were then made free from PCR duplicates using samtools. Replicates of both the control and treated were merged respectively into a single BAM file. Thereafter, the merged BAM file of the input samples and the merged BAM file of treated samples were used for the subsequent peak calling by MACS1.4.2 with p-value cutoff of 1e-05. Read enrichments/peaks were first called for each replicate with the uniquely mapped reads by MACS1.4.2 with default parameters except the effective genome size corrected for rn6. Chromosome-wise distribution plots were generated using custom R script. H1T2 peak distribution was identified at genomic locations enriched in CpG islands, repetitive elements, exons, introns, intergenic regions, 3'UTR and 5'UTR using HOMER v4.7 (for annotation). Further classification of LINE L1 repeat elements was made using the repeat masker data of UCSC table browser. Box plots were created using the R 3.3.2 statistical package. De novo motif identification was done using 300 bp regions surrounding overlapping peak summits as target set and shuffled set of same overlap sequences as background set with MEME software.

### ChIP PCR analysis

ChIP chromatin was prepared as described above. Beads with bound immunocomplexes were washed, eluted, and reverse cross-linked at 65°C overnight, subjected for RNase (final; 40 µg/ml) and proteinase K (Final; 100 µg/ml) treatment and DNA were extracted by phenol–chloroform method. Input was prepared from 5% of the sonicated chromatin. 2µL input or ChIP sample from the recovered DNA suspension of 50µL was used for performing the qPCR analysis. Specific primers were designed against the H1T2 peaks obtained from ChIP-sequencing experiment (primers are listed in Additional file 14: Table S14). The enrichment of specific regions of the genome were validated by qPCR reaction using ChIP DNA (20 ng), 300 nM forward/reverse primers using SYBR kit from TAKARA. The Ct values of the duplicates showed minimal variability, and Ct values were used for calculating the percent of input using the following formula.

$$\text{Percent Input} = 100 \times \frac{2(\text{Adjusted input} - \text{Ct (IP)})}{\text{Ct (IP)}}$$

### MNase digestion and chromatin immunoprecipitation

MNase immunoprecipitation was done according to the previous method (Montellier et al., 2013). Briefly, the rat testes were dissected and lysed in the Lysis buffer (Tris-HCl pH 7.4 15 mM, KCl 60 mM, NaCl 15 mM, sucrose 0.34 M, EDTA 2 mM, EGTA 0.5 mM, spermidine 0.65 mM, dithiothreitol 1 mM, Triton X-100 0.03 %, glycerol 1 %, Protease Cocktail Inhibitor) followed by centrifugation for 15 min at 200 × g at 4 °C. Nuclei were suspended in wash buffer (Tris-HCl pH 7.4 15 mM, KCl 60 mM, NaCl 15 mM, sucrose 0.34 M, dithiothreitol 1 mM, Protease Cocktail Inhibitor) and directly proceeded for MNase digestion by suspending the nuclei in MNase buffer (Tris-HCl pH 7.5 10 mM, KCl 10 mM, CaCl<sub>2</sub> 2 mM). Sperm nuclei were digested with MNase in MNase buffer for 5 min 37°C. Reaction was stopped with the addition of 10 mM EDTA on ice. The soluble nucleosome fraction was prepared by centrifugation at 650×g for 10 min at 4°C, mixed with LSDB250 buffer (20% glycerol, 50-mM HEPES, 3-mM MgCl<sub>2</sub>, 250-mM KCl, protease and phosphatase inhibitor cocktail) and proceeded to immunoprecipitation protocol.

### Mass spectrometry for identification of H1T2 bound chromatin associated proteins and histone post-translational modifications

The oligonucleosomes fraction prepared by MNase digestion was incubated with either anti H1T2 antibody or pre-immune IgG for overnight at 4°C in a rotating platform. Protein a/G dynabeads were added on the next day and LSDB250 buffer was used as the wash buffer for the MNase-IP studies. Proteins bound to the beads were extracted using the elution buffer from the Pierce Co-IP kit. The eluted protein complexes were resolved on a 15% SDS gel and run for about 1cm, then the gel was subjected to Coomassie staining. The Coomassie stained wells corresponding to IP and IgG lanes were excised and subjected to downstream analysis. To determine the proteins associated to H1T2 bound chromatin, we have outsourced the samples for LC-MS/MS analysis which was carried out at the Taplin Biological Mass Spectrometry Facility, Cell Biology Department, Harvard Medical School.

Excised gel bands were cut into approximately 1 mm<sup>3</sup> pieces. As described previously (98), gel pieces were then subjected to a modified in-gel trypsin digestion procedure (119). Gel pieces were washed and dehydrated with acetonitrile for 10 min. followed by removal of acetonitrile. Pieces were then completely dried in a speed-vac. Rehydration of the gel pieces was with 50 mM ammonium bicarbonate solution containing 12.5 ng/μl modified sequencing-grade trypsin (Promega, Madison, WI) at 4°C. After 45 min, the excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate solution to just cover the gel pieces. Samples were then placed in a 37°C room overnight. Peptides were later extracted by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile and 1% formic acid. The extracts were then dried in a speed-vac (~1 hr). The samples were then stored at 4°C until analysis. On the day of analysis the samples were reconstituted in 5 - 10 μl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 2.6 μm C18 spherical silica beads into a fused silica capillary (100 μm inner diameter x ~30 cm length) with a flame-drawn tip (Peng et al., 2001). After equilibrating the column each sample was loaded via a Famos auto sampler (LC Packings, San Francisco CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted they were subjected to electrospray ionization and then entered into an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching protein databases with the acquired fragmentation pattern by the software program, Sequest (Thermo Fisher Scientific, Waltham, MA) (120). All databases include a reversed version of all the sequences and the data was filtered to between a one and two percent peptide false discovery rate. The complete list of H1T2 associated list of proteins with more than 3 unique peptides is given in the table additional file 9: Table S9.

For the mass spectrometric analysis of core histone PTMs in the immunoprecipitated chromatin complex, the final eluate was subjected to histone extraction using 0.4 n H<sub>2</sub>SO<sub>4</sub>. After precipitation of histone proteins with 30% trichloroacetic acid, the histone extracts were loaded onto a 12% SDS gel and resolved completely. 5 to 25 kDa portion that contained all core histones were excised from each gel (H1T2 IP and IgG IP lanes) and sent for LC-MS/MS analysis of the histones (PTMs).

## List Of Antibodies Used In The Present Study

<b>Antibodies</b>	<b>Host</b>	<b>Company name</b>	<b>Cat Number</b>	<b>Applications</b>
Ddx5	Rabbit	Abcam	ab126730	WB
Ruvbl1	Rabbit	Abcam	ab133513	WB
Ybx2	Rabbit	Abcam	ab154829	WB
Elavl1	Rabbit	Abcam	ab200342	WB
Srsf3	Rabbit	Abcam	ab198291	WB
Pabp1	Rabbit	Abcam	ab21060	WB
Ilf3	Rabbit	Abcam	ab92355	WB
Hspa2	Rabbit	Invitrogen	PA5-81895	WB
H3	Rabbit	Abcam	ab1791	WB
H3.3	Rabbit	Abcam	ab62642	WB
H3K4me3	Rabbit	Abcam	ab8895	WB
H3K9ac	Rabbit	Abcam	ab4441	WB
H3K27ac	Rabbit	Abcam	ab4729	WB
H3K4me1	Rabbit	Abcam	ab8895	WB
H3K79me2	Rabbit	Abcam	ab177184	WB
H4	Rabbit	Abcam	ab10158	WB
H4K5ac	Rabbit	Abcam	ab51997	WB
H4K12ac	Rabbit	Abcam	ab177793	WB
H4K16ac	Rabbit	Abcam	ab109463	WB
H3K9me3	Rabbit	Abcam	ab8898	WB
H4K20me3	Rabbit	Abcam	ab9053	WB
H3K27me3	Rabbit	Abcam	ab192985	WB

WB western blotting

## Declarations

**Ethics approval:** This work was approved by Animal Ethics Committee of JNCASR, Bangalore, India.

**Consent for publication:** Not applicable.

**Availability of data and material:** The data sets supporting the conclusions of this article are submitted to Gene Expression Omnibus repository and waiting for the Accession Number (To be updated)

**Competing interests:** The authors declare that they have no competing interests with the contents of this article.

### Funding:

M.R.S. Rao thanks Department of Science and Technology, Government of India, for SERB Distinguished Fellowship and SERB-YOS Chair Professorship, and this work was financially supported by Department of Biotechnology, Govt. of India (Grant Numbers: BT/01/COE/07/09 and DBT/INF/22/SP27679/2018). S.V acknowledges DBT for postdoctoral fellowship (Financial support from the DBT-RA Program in Biotechnology and Life Sciences is gratefully acknowledged).

**Author contributions:** MRSR conceived the idea for the project, designed, analyzed results, and wrote the paper. SV designed, performed, analyzed results, and wrote the paper. UB analyzed the ChIP-seq data and supported the writing and figure generation. ACR and AR helped in the IF and ChIP-sequencing experiments. All authors read and approved the final version of the manuscript.

**Acknowledgements:** We thank Dr. R.G. Prakash of the Animal Facility and Suma B.S. of the Confocal Imaging Facility at JNCASR. We also thank members of the Abgenex Company (Bhubaneshwar, India) for carrying out the antibody generation project. We also thank Tomaino Ross of TAPLIN Mass spectrometry facility at Harvard for mass spectrometry-based identification of H1T2 associated proteins.

## References

1. Boerke A, Dieleman SJ, Gadella BM. A possible role for sperm RNA in early embryo development. *Theriogenology*. 2007;68 Suppl 1:S147-55.
2. Ren X, Chen X, Wang Z, Wang D. Is transcription in sperm stationary or dynamic? *The Journal of reproduction and development*. 2017;63(5):439-43.
3. O'Donnell L. Mechanisms of spermiogenesis and spermiation and how they are disturbed. *Spermatogenesis*. 2014;4(2):e979623.
4. Poccia D. Remodeling of nucleoproteins during gametogenesis, fertilization, and early development. *Int Rev Cytol*. 1986;105:1-65.
5. Balhorn R. The protamine family of sperm nuclear proteins. *Genome Biol*. 2007;8(9):227.
6. Boussouar F, Rousseaux S, Khochbin S. A new insight into male genome reprogramming by histone variants and histone code. *Cell Cycle*. 2008;7(22):3499-502.
7. De Vries M, Ramos L, Housein Z, De Boer P. Chromatin remodelling initiation during human spermiogenesis. *Biol Open*. 2012;1(5):446-57.
8. Wang T, Gao H, Li W, Liu C. Essential Role of Histone Replacement and Modifications in Male Fertility. *Front Genet*. 2019;10:962.
9. Rathke C, Baarens WM, Awe S, Renkawitz-Pohl R. Chromatin dynamics during spermiogenesis. *Biochimica et biophysica acta*. 2014;1839(3):155-68.
10. Sonnack V, Failing K, Bergmann M, Steger K. Expression of hyperacetylated histone H4 during normal and impaired human spermatogenesis. *Andrologia*. 2002;34(6):384-90.
11. Fenic I, Sonnack V, Failing K, Bergmann M, Steger K. In vivo effects of histone-deacetylase inhibitor trichostatin-A on murine spermatogenesis. *J Androl*. 2004;25(5):811-8.
12. Zhao M, Shirley CR, Hayashi S, Marcon L, Mohapatra B, Suganuma R, et al. Transition nuclear proteins are required for normal chromatin condensation and functional sperm development. *Genesis*. 2004;38(4):200-13.
13. Hermo L, Pelletier RM, Cyr DG, Smith CE. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 3: developmental changes in spermatid flagellum and

cytoplasmic droplet and interaction of sperm with the zona pellucida and egg plasma membrane. Microsc Res Tech. 2010;73(4):320-63.

14. Kierszenbaum AL, Rivkin E, Tres LL. Acroplaxome, an F-actin-keratin-containing plate, anchors the acrosome to the nucleus during shaping of the spermatid head. Mol Biol Cell. 2003;14(11):4628-40.
15. 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(7254):473-8.
16. Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausio J, Taylor J, et al. Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. Cell Rep. 2017;18(6):1366-82.
17. McCarrey JR, Geyer CB, Yoshioka H. Epigenetic regulation of testis-specific gene expression. Ann N Y Acad Sci. 2005;1061:226-42.
18. Govin J, Escoffier E, Rousseaux S, Kuhn L, Ferro M, Thevenon J, et al. Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. J Cell Biol. 2007;176(3):283-94.
19. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ, et al. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc Natl Acad Sci U S A. 1998;95(24):14173-8.
20. Hendzel MJ, Lever MA, Crawford E, Th'ng JP. The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. J Biol Chem. 2004;279(19):20028-34.
21. Kimmings S, Sassone-Corsi P. Chromatin remodelling and epigenetic features of germ cells. Nature. 2005;434(7033):583-9.
22. Martianov I, Brancorsini S, Catena R, Gansmuller A, Kotaja N, Parvinen M, et al. Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. Proc Natl Acad Sci U S A. 2005;102(8):2808-13.
23. Tanaka H, Iguchi N, Isotani A, Kitamura K, Toyama Y, Matsuoka Y, et al. HANP1/H1T2, a novel histone H1-like protein involved in nuclear formation and sperm fertility. Mol Cell Biol. 2005;25(16):7107-19.
24. Mishra LN, Shalini V, Gupta N, Ghosh K, Suthar N, Bhaduri U, et al. Spermatid-specific linker histone H1L51 is a poor condenser of DNA and chromatin and preferentially associates with LINE-1 elements. Epigenetics Chromatin. 2018;11(1):43.
25. Mahadevan IA, Kumar S, Rao MRS. Linker histone variant H1t is closely associated with repressed repeat-element chromatin domains in pachytene spermatocytes. Epigenetics Chromatin. 2020;13(1):9.
26. Pradeepa MM, Rao MR. Chromatin remodeling during mammalian spermatogenesis: role of testis specific histone variants and transition proteins. Soc Reprod Fertil Suppl. 2007;63:1-10.
27. Turner LM, Chuong EB, Hoekstra HE. Comparative analysis of testis protein evolution in rodents. Genetics. 2008;179(4):2075-89.

28. Kost N, Kaiser S, Ostwal Y, Riedel D, Stutzer A, Nikolov M, et al. Multimerization of *Drosophila* sperm protein Mst77F causes a unique condensed chromatin structure. *Nucleic Acids Res.* 2015;43(6):3033-45.
29. Catena R, Ronfani L, Sassone-Corsi P, Davidson I. Changes in intranuclear chromatin architecture induce bipolar nuclear localization of histone variant H1T2 in male haploid spermatids. *Dev Biol.* 2006;296(1):231-8.
30. Kazazian HH, Jr. Mobile elements: drivers of genome evolution. *Science.* 2004;303(5664):1626-32.
31. Di Giacomo M, Comazzetto S, Saini H, De Fazio S, Carrieri C, Morgan M, et al. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol Cell.* 2013;50(4):601-8.
32. Singh DK, Rath PC. Long interspersed nuclear elements (LINEs) show tissue-specific, mosaic genome and methylation-unrestricted, widespread expression of noncoding RNAs in somatic tissues of the rat. *RNA Biol.* 2012;9(11):1380-96.
33. Hammoud S, Liu L, Carrell DT. Protamine ratio and the level of histone retention in sperm selected from a density gradient preparation. *Andrologia.* 2009;41(2):88-94.
34. Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol.* 2010;17(6):679-87.
35. Berkovits BD, Wolgemuth DJ. The role of the double bromodomain-containing BET genes during mammalian spermatogenesis. *Curr Top Dev Biol.* 2013;102:293-326.
36. Martianov I, Fimia GM, Dierich A, Parvinen M, Sassone-Corsi P, Davidson I. Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol Cell.* 2001;7(3):509-15.
37. Saito M, Kumamoto K, Robles AI, Horikawa I, Furusato B, Okamura S, et al. Targeted disruption of *Ing2* results in defective spermatogenesis and development of soft-tissue sarcomas. *PLoS One.* 2010;5(11):e15541.
38. Kehoe SM, Oka M, Hankowski KE, Reichert N, Garcia S, McCarrey JR, et al. A conserved E2F6-binding element in murine meiosis-specific gene promoters. *Biol Reprod.* 2008;79(5):921-30.
39. Storre J, Elsasser HP, Fuchs M, Ullmann D, Livingston DM, Gaubatz S. Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2f6. *EMBO Rep.* 2002;3(7):695-700.
40. Trimarchi JM, Fairchild B, Verona R, Moberg K, Andon N, Lees JA. E2F-6, a member of the E2F family that can behave as a transcriptional repressor. *Proc Natl Acad Sci U S A.* 1998;95(6):2850-5.
41. Pohlers M, Truss M, Frede U, Scholz A, Strehle M, Kuban RJ, et al. A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr Biol.* 2005;15(11):1051-7.
42. Kuroki S, Akiyoshi M, Tokura M, Miyachi H, Nakai Y, Kimura H, et al. JMJD1C, a JmjC domain-containing protein, is required for long-term maintenance of male germ cells in mice. *Biol Reprod.* 2013;89(4):93.

43. Nakajima S, Hayashi M, Kouguchi T, Yamaguchi K, Miwa M, Yoshizaki G. Expression patterns of gdnf and gfralpha1 in rainbow trout testis. *Gene Expr Patterns*. 2014;14(2):111-20.
44. Godmann M, Kosan C, Behr R. Kruppel-like factor 4 is widely expressed in the mouse male and female reproductive tract and responds as an immediate early gene to activation of the protein kinase A in TM4 Sertoli cells. *Reproduction*. 2010;139(4):771-82.
45. Behr R, Kaestner KH. Developmental and cell type-specific expression of the zinc finger transcription factor Kruppel-like factor 4 (Klf4) in postnatal mouse testis. *Mech Dev*. 2002;115(1-2):167-9.
46. Muller-Borer B, Esch G, Aldina R, Woon W, Fox R, Bursac N, et al. Calcium dependent CAMTA1 in adult stem cell commitment to a myocardial lineage. *PLoS One*. 2012;7(6):e38454.
47. Hernandez-Hernandez A, Lilienthal I, Fukuda N, Galjart N, Hoog C. CTCF contributes in a critical way to spermatogenesis and male fertility. *Sci Rep*. 2016;6:28355.
48. Arzate-Mejia RG, Recillas-Targa F, Corces VG. Developing in 3D: the role of CTCF in cell differentiation. *Development*. 2018;145(6).
49. Teperek M, Simeone A, Gaggioli V, Miyamoto K, Allen GE, Erkek S, et al. Sperm is epigenetically programmed to regulate gene transcription in embryos. *Genome Res*. 2016;26(8):1034-46.
50. Barnes RM, Firulli AB. A twist of insight - the role of Twist-family bHLH factors in development. *Int J Dev Biol*. 2009;53(7):909-24.
51. Powers SE, Taniguchi K, Yen W, Melhuish TA, Shen J, Walsh CA, et al. Tgif1 and Tgif2 regulate Nodal signaling and are required for gastrulation. *Development*. 2010;137(2):249-59.
52. Hu Y, Yu H, Shaw G, Renfree MB, Pask AJ. Differential roles of TGIF family genes in mammalian reproduction. *BMC Dev Biol*. 2011;11:58.
53. Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, et al. Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. *Stem Cells*. 2012;30(5):910-22.
54. Schaniel C, Ang YS, Ratnakumar K, Cormier C, James T, Bernstein E, et al. Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells*. 2009;27(12):2979-91.
55. Blendy JA, Kaestner KH, Weinbauer GF, Nieschlag E, Schutz G. Severe impairment of spermatogenesis in mice lacking the CREM gene. *Nature*. 1996;380(6570):162-5.
56. Deng W, Lin H. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell*. 2002;2(6):819-30.
57. Kashiwabara S, Noguchi J, Zhuang T, Ohmura K, Honda A, Sugiura S, et al. Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP. *Science*. 2002;298(5600):1999-2002.
58. Hensen K, Braem C, Declercq J, Van Dyck F, Dewerchin M, Fiette L, et al. Targeted disruption of the murine Plag1 proto-oncogene causes growth retardation and reduced fertility. *Dev Growth Differ*. 2004;46(5):459-70.

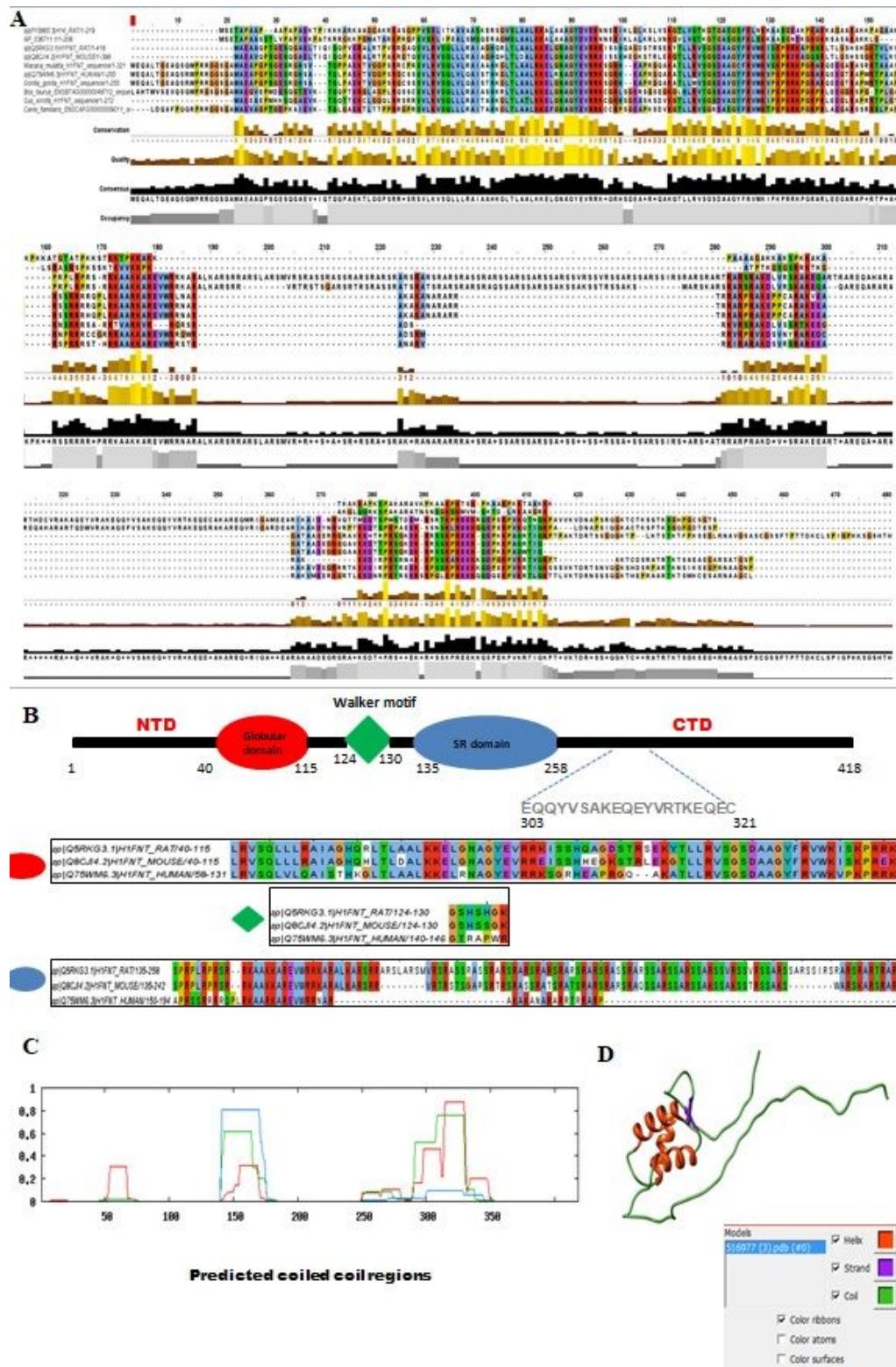
59. Gold HB, Jung YH, Corces VG. Not just heads and tails: The complexity of the sperm epigenome. *J Biol Chem.* 2018;293(36):13815-20.
60. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature.* 2004;431(7004):96-9.
61. Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, et al. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A.* 2005;102(11):4068-73.
62. Branciforte D, Martin SL. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol.* 1994;14(4):2584-92.
63. Trelogan SA, Martin SL. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl Acad Sci U S A.* 1995;92(5):1520-4.
64. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, et al. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 2009;23(11):1303-12.
65. Yin Y, Zhou L, Yuan S. Enigma of Retrotransposon Biology in Mammalian Early Embryos and Embryonic Stem Cells. *Stem Cells Int.* 2018;2018:6239245.
66. Chang L, Goldman RD. Intermediate filaments mediate cytoskeletal crosstalk. *Nat Rev Mol Cell Biol.* 2004;5(8):601-13.
67. Etienne-Manneville S. Actin and microtubules in cell motility: which one is in control? *Traffic.* 2004;5(7):470-7.
68. Ezratty EJ, Partridge MA, Gundersen GG. Microtubule-induced focal adhesion disassembly is mediated by dynamin and focal adhesion kinase. *Nat Cell Biol.* 2005;7(6):581-90.
69. Li R, Gundersen GG. Beyond polymer polarity: how the cytoskeleton builds a polarized cell. *Nat Rev Mol Cell Biol.* 2008;9(11):860-73.
70. Arnold DB. Actin and microtubule-based cytoskeletal cues direct polarized targeting of proteins in neurons. *Sci Signal.* 2009;2(83):pe49.
71. Dunleavy JEM, O'Bryan MK, Stanton PG, O'Donnell L. The cytoskeleton in spermatogenesis. *Reproduction.* 2019;157(2):R53-R72.
72. Kierszenbaum AL, Rivkin E, Tres LL. Molecular biology of sperm head shaping. *Soc Reprod Fertil Suppl.* 2007;65:33-43.
73. Shen J, Chen W, Shao B, Qi Y, Xia Z, Wang F, et al. Lamin A/C proteins in the spermatid acroplaxome are essential in mouse spermiogenesis. *Reproduction.* 2014;148(5):479-87.
74. Yu J, Chen B, Zheng B, Qiao C, Chen X, Yan Y, et al. ATP synthase is required for male fertility and germ cell maturation in *Drosophila* testes. *Mol Med Rep.* 2019;19(3):1561-70.
75. Paronetto MP, Sette C. Role of RNA-binding proteins in mammalian spermatogenesis. *Int J Androl.* 2010;33(1):2-12.

76. Kleene KC. Patterns of translational regulation in the mammalian testis. *Mol Reprod Dev.* 1996;43(2):268-81.
77. Yang J, Medvedev S, Reddi PP, Schultz RM, Hecht NB. The DNA/RNA-binding protein MSY2 marks specific transcripts for cytoplasmic storage in mouse male germ cells. *Proc Natl Acad Sci U S A.* 2005;102(5):1513-8.
78. Vemuganti SA, de Villena FP, O'Brien DA. Frequent and recent retrotransposition of orthologous genes plays a role in the evolution of sperm glycolytic enzymes. *BMC Genomics.* 2010;11:285.
79. Iguchi N, Tobias JW, Hecht NB. Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc Natl Acad Sci U S A.* 2006;103(20):7712-7.
80. Chi MN, Auriol J, Jegou B, Kontoyiannis DL, Turner JM, de Rooij DG, et al. The RNA-binding protein ELAVL1/HuR is essential for mouse spermatogenesis, acting both at meiotic and postmeiotic stages. *Mol Biol Cell.* 2011;22(16):2875-85.
81. Phillips BT, Williams JG, Atchley DT, Xu X, Li JL, Adams AL, et al. Mass spectrometric identification of candidate RNA-binding proteins associated with Transition Nuclear Protein mRNA in the mouse testis. *Sci Rep.* 2019;9(1):13618.
82. Yang J, Morales CR, Medvedev S, Schultz RM, Hecht NB. In the absence of the mouse DNA/RNA-binding protein MSY2, messenger RNA instability leads to spermatogenic arrest. *Biol Reprod.* 2007;76(1):48-54.
83. Legrand JMD, Chan AL, La HM, Rossello FJ, Anko ML, Fuller-Pace FV, et al. DDX5 plays essential transcriptional and post-transcriptional roles in the maintenance and function of spermatogonia. *Nat Commun.* 2019;10(1):2278.
84. Gorgoni B, Gray NK. The roles of cytoplasmic poly(A)-binding proteins in regulating gene expression: a developmental perspective. *Brief Funct Genomic Proteomic.* 2004;3(2):125-41.
85. Attig J, Agostini F, Gooding C, Chakrabarti AM, Singh A, Haberman N, et al. Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA Processing. *Cell.* 2018;174(5):1067-81 e17.
86. Goodier JL, Cheung LE, Kazazian HH, Jr. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* 2013;41(15):7401-19.
87. Bryant JM, Donahue G, Wang X, Meyer-Ficca M, Luense LJ, Weller AH, et al. Characterization of BRD4 during mammalian postmeiotic sperm development. *Mol Cell Biol.* 2015;35(8):1433-48.
88. Dottermusch-Heidel C, Gartner SM, Tegeder I, Rathke C, Barckmann B, Bartkuhn M, et al. H3K79 methylation: a new conserved mark that accompanies H4 hyperacetylation prior to histone-to-protamine transition in *Drosophila* and rat. *Biol Open.* 2014;3(6):444-52.
89. Dottermusch-Heidel C, Klaus ES, Gonzalez NH, Bhushan S, Meinhardt A, Bergmann M, et al. H3K79 methylation directly precedes the histone-to-protamine transition in mammalian spermatids and is sensitive to bacterial infections. *Andrology.* 2014;2(5):655-65.
90. Chen P, Zhao J, Wang Y, Wang M, Long H, Liang D, et al. H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes Dev.* 2013;27(19):2109-24.

91. Tang MC, Jacobs SA, Mattiske DM, Soh YM, Graham AN, Tran A, et al. Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. *PLoS Genet.* 2015;11(2):e1004964.
92. Meistrich ML, Bucci LR, Trostle-Weige PK, Brock WA. Histone variants in rat spermatogonia and primary spermatocytes. *Dev Biol.* 1985;112(1):230-40.
93. Montellier E, Boussouar F, Rousseaux S, Zhang K, Buchou T, Fenaille F, et al. Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev.* 2013;27(15):1680-92.
94. Fyodorov DV, Zhou BR, Skoultschi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol.* 2018;19(3):192-206.
95. Hergeth SP, Schneider R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep.* 2015;16(11):1439-53.
96. Kleene KC. Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenet Genome Res.* 2003;103(3-4):217-24.
97. Paradowska AS, Miller D, Spiess AN, Vieweg M, Cerna M, Dvorakova-Hortova K, et al. Genome wide identification of promoter binding sites for H4K12ac in human sperm and its relevance for early embryonic development. *Epigenetics.* 2012;7(9):1057-70.
98. Mahadevan IA, Pentakota S, Roy R, Bhaduri U, Satyanarayana Rao MR. TH2BS11ph histone mark is enriched in the unsynapsed axes of the XY body and predominantly associates with H3K4me3-containing genomic regions in mammalian spermatocytes. *Epigenetics Chromatin.* 2019;12(1):53.
99. Jayaramaiah Raja S, Renkawitz-Pohl R. Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus. *Mol Cell Biol.* 2005;25(14):6165-77.
100. Pan C, Fan Y. Role of H1 linker histones in mammalian development and stem cell differentiation. *Biochimica et biophysica acta.* 2016;1859(3):496-509.
101. Gu NH, Zhao WL, Wang GS, Sun F. Comparative analysis of mammalian sperm ultrastructure reveals relationships between sperm morphology, mitochondrial functions and motility. *Reproductive biology and endocrinology : RB&E.* 2019;17(1):66.
102. Champroux A, Cocquet J, Henry-Berger J, Drevet JR, Kocer A. A Decade of Exploring the Mammalian Sperm Epigenome: Paternal Epigenetic and Transgenerational Inheritance. *Front Cell Dev Biol.* 2018;6:50.
103. Kramer JA, Krawetz SA. RNA in spermatozoa: implications for the alternative haploid genome. *Mol Hum Reprod.* 1997;3(6):473-8.
104. Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature.* 2004;429(6988):154.
105. Miller D, Ostermeier GC, Krawetz SA. The controversy, potential and roles of spermatozoal RNA. *Trends Mol Med.* 2005;11(4):156-63.

106. Krawetz SA. Paternal contribution: new insights and future challenges. *Nat Rev Genet.* 2005;6(8):633-42.
107. Vitullo P, Sciamanna I, Baiocchi M, Sinibaldi-Vallebona P, Spadafora C. LINE-1 retrotransposon copies are amplified during murine early embryo development. *Mol Reprod Dev.* 2012;79(2):118-27.
108. Rosser JM, An W. L1 expression and regulation in humans and rodents. *Front Biosci (Elite Ed).* 2012;4:2203-25.
109. Lazaros L, Kitsou C, Kostoulas C, Bellou S, Hatzi E, Ladias P, et al. Retrotransposon expression and incorporation of cloned human and mouse retroelements in human spermatozoa. *Fertil Steril.* 2017;107(3):821-30.
110. Pittoggi C, Renzi L, Zaccagnini G, Cimini D, Degrassi F, Giordano R, et al. A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. *J Cell Sci.* 1999;112 ( Pt 20):3537-48.
111. Pittoggi C, Zaccagnini G, Giordano R, Magnano AR, Baccetti B, Lorenzini R, et al. Nucleosomal domains of mouse spermatozoa chromatin as potential sites for retroposition and foreign DNA integration. *Mol Reprod Dev.* 2000;56(2 Suppl):248-51.
112. Belan E. LINEs of evidence: noncanonical DNA replication as an epigenetic determinant. *Biol Direct.* 2013;8:22.
113. Cannarella R, Condorelli RA, Mongioi LM, La Vignera S, Calogero AE. Molecular Biology of Spermatogenesis: Novel Targets of Apparently Idiopathic Male Infertility. *Int J Mol Sci.* 2020;21(5).
114. Lupas A. Coiled coils: new structures and new functions. *Trends Biochem Sci.* 1996;21(10):375-82.
115. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science.* 1991;252(5009):1162-4.
116. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* 2009;4(3):363-71.
117. Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. *Nat Protoc.* 2007;2(6):1445-57.
118. Pillai S, Dasgupta P, Chellappan SP. Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications in vivo. *Methods Mol Biol.* 2009;523:323-39.
119. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem.* 1996;68(5):850-8.
120. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom.* 1994;5(11):976-89.
121. Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* 2013;3(6):2179-90.

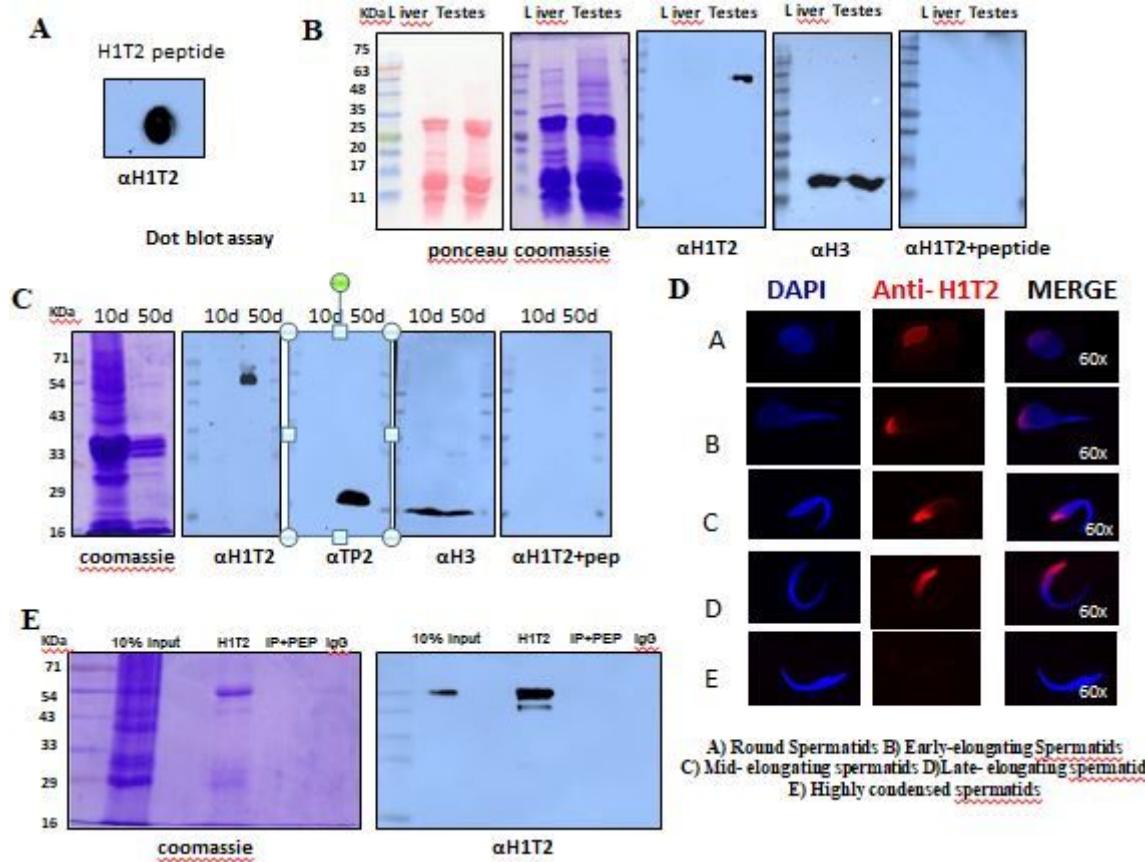
## Figures



**Figure 1**

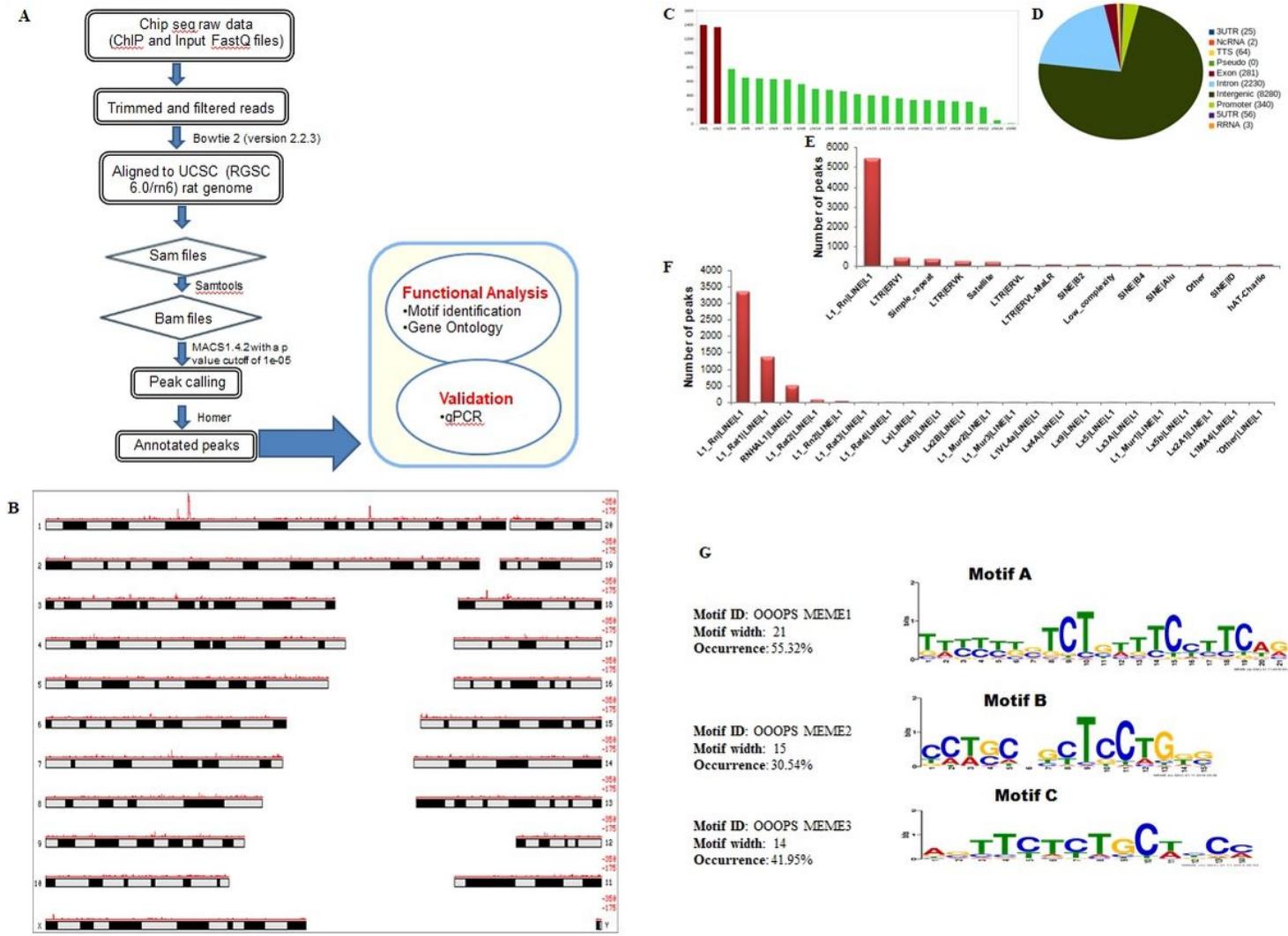
H1T2 possess highly divergent C terminal domain. A) Multiple sequence alignment of sequences of linker histone variants. Sequences from either somatic linker histone H1d or germ cell specific H1t were aligned to consensus sequence from germ cell specific H1T2 of different species. Amino acids are colored according to their similarity degree using the ClustalX option of Jalview. B) Schematic representation of domain architecture of rat H1T2 protein: Globular domain (red), ATP binding walker motif (green), serine-

arginine rich domain (blue). Peptide sequence against which the antibody is raised is highlighted in the C terminal domain. Clustal alignment of H1T2 sequence from rat, mouse and human species is also shown in the figure. Conserved globular domain, walker motif (ATP binding), and the divergent SR domain in the H1T2 sequences are represented separately. C) Coiled coil prediction analysis of rat H1T2 sequence ([https://embnet.vital-it.ch/cgi-bin/COILS\\_form\\_parser](https://embnet.vital-it.ch/cgi-bin/COILS_form_parser)). D) Secondary structure of the rat H1T2 sequence predicted using Phyre2 (PDB format).



**Figure 2**

Characterization of antibody raised against the highly divergent C terminal domain of H1T2. A) Dot blot analysis of affinity purified antibody against H1T2 with respective peptide. B) Western blot analysis of the acid extracts from rat liver (lane 1) and testes (lane 2) with antibody against H1T2. C) Western blot analysis of acid-soluble proteins from 10-days postnatal (lane 1) and 40-days postnatal (lane 2) rat testes nuclei with anti-H1T2 antibodies in the presence or absence of peptide competition with a 200-fold molar excess of the peptide used for antibody generation, as indicated. Western blot analysis with anti-TP2 and anti-H3 antibodies served as a positive control. D) Immunofluorescence analysis in spermatids using H1T12 antibody. E) Western blot analysis of the pull-down fractions with a H1T2. Input (lane 1), anti-H1T2 antibody (lane 2), anti-H1T2 antibodies + peptide (lane 3), IgG (lane 4). No signal was observed in peptide competition and pre-immune IgG lanes.



**Figure 3**

Genome-wide analysis of H1T2 in rat spermatid genome using antibody against C terminal domain. A) Schematic representation of the protocol used for ChIP-sequencing analysis B) Chromosome wise enrichment of H1T2 ChIP peaks in rat spermatid genome. C) Number of peaks distributed across different rat chromosomes (x-axis), where y-axis represents the number of peaks. Box plot representing the H1T2 associated average peak length (y-axis) across different chromosomes (x-axis) D) Pie chart distribution of H1T2 peaks for various genomic features in rat spermatids. The peaks are distributed in among intergenic, intron, promoter, exon, TTS, 5'UTR, 3'UTR regions. E) Bar diagram of number of occurrences of peaks across repeat elements. F) Bar diagram showing the number of occurrences of peaks across different classes of repeat elements. G) Motif identification by MEME. Figure represents the details of the three most significant motifs identified from the overlapping peak summits (H1T2 ChIP peaks).

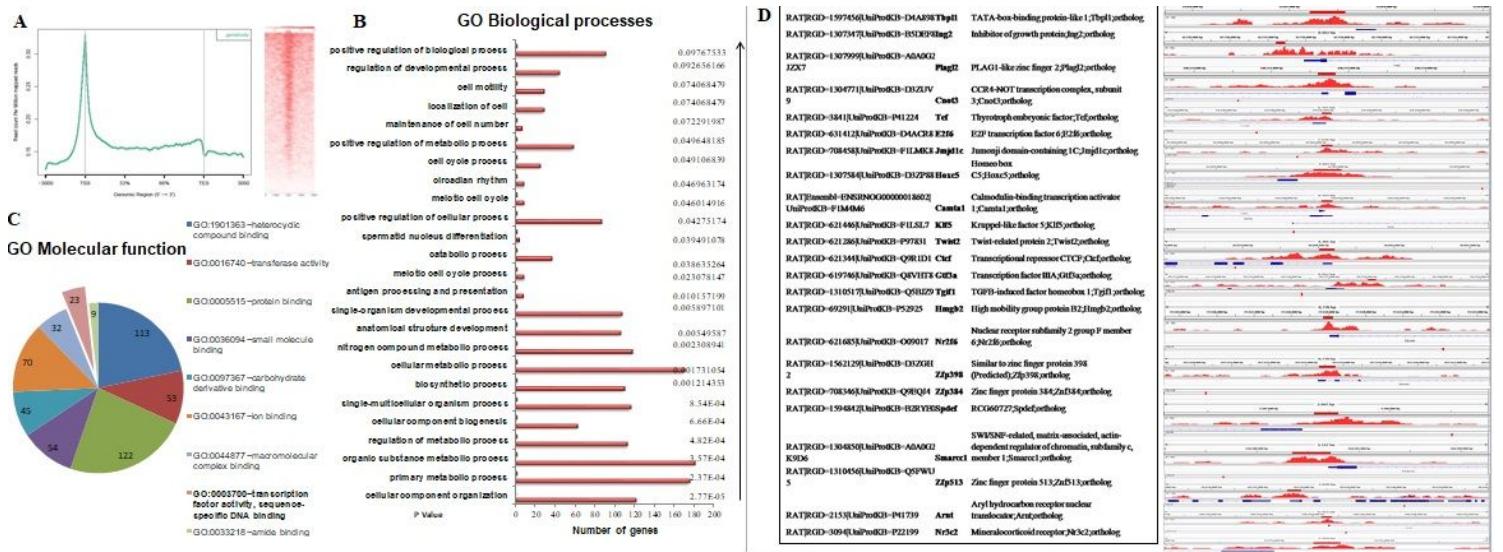
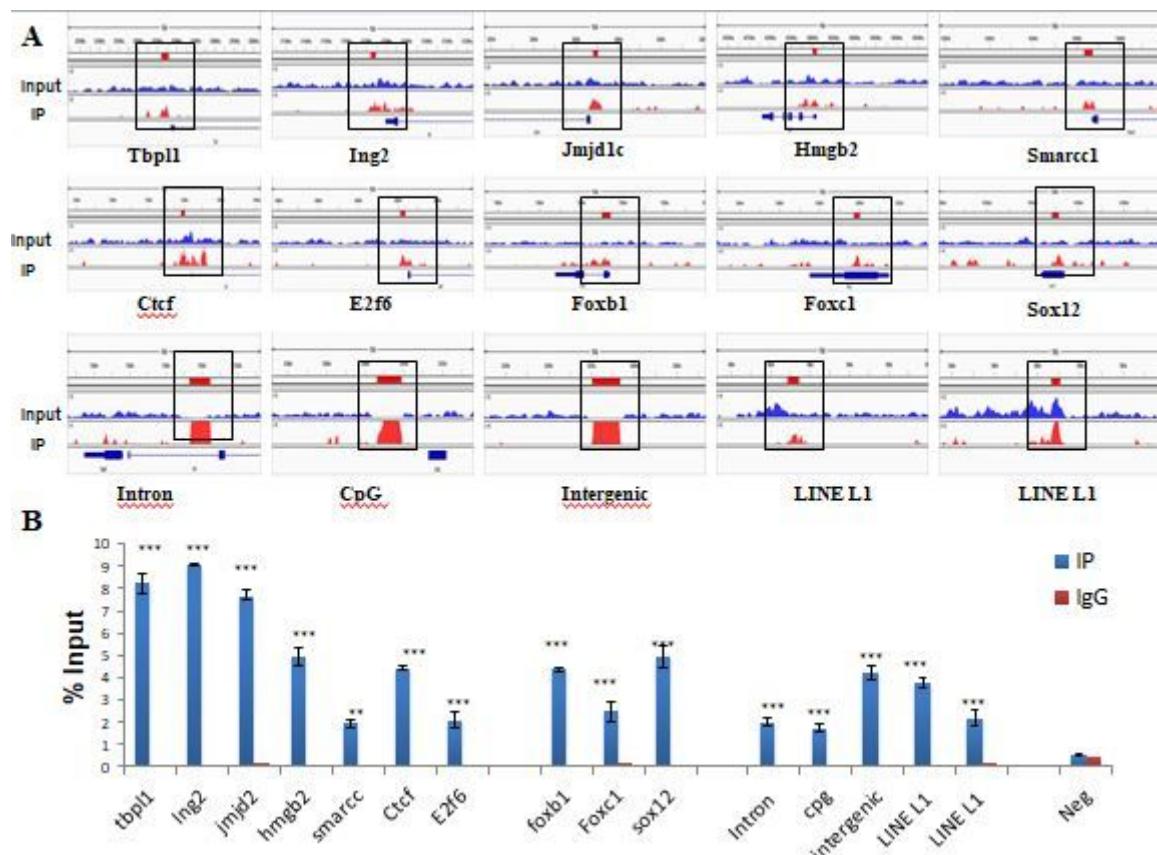


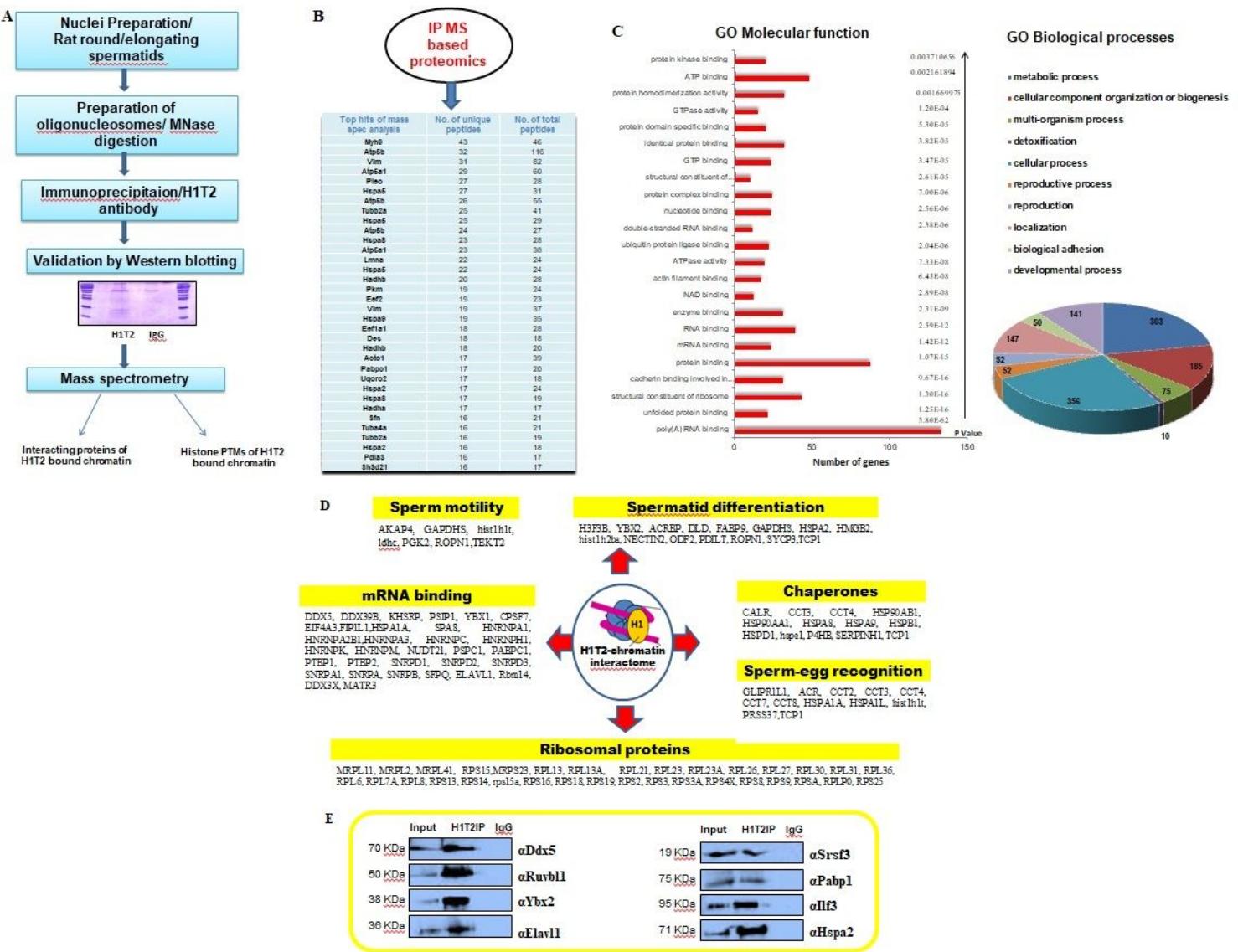
Figure 4

Functional analysis of H1T2 bound promoter-TSS regions in rat spermatids genome A) Gene ontology analysis of the 321 genes whose promoter-TSS regions are bound to H1T2. David software was used for the analysis and the results are shown in the figure. GO biological processes are represented in the figure with P values. X-axis represents the number of genes associated to the GO term B) Distribution of H1T2 peaks from +/-3kb centered around TSS of genes in the spermatids genome and heatmap showing the same observation ngs. plot showing distribution of H1T2 peaks throughout the gene C) GO analysis of 321 genes according to molecular function represented as pie diagram. The number of genes associated to the GO term is given in the pie chart. 23 transcription factors whose promoter-TSS is bound to H1T2 is highlighted in the pie diagram D) Integrative Genomics Viewer (IGV) outputs of representative genomic regions showing the occupancy of H1T2 at various transcription factors of the spermatid genome



**Figure 5**

Validation of H1T2 ChIP-sequencing data by ChIP PCR analysis of representative genes bound to the linker histone H1T2. A) Screenshots of H1T2 ChIP peaks displayed from Integrative Genomics Viewer. Shown are 15 different genomic regions associated to H1T2. The y-axis unit is reads per million (rpm). The x-axis represent the genomic positions in base pairs as follows: Chr1:23977211-23977579; Chr16:47717217-47717457; Chr20:22882800-22883032; Chr16:36080014-36080219; Chr8:118206347-118206725; Chr19:37599804-37600003; Chr6:42092121-42092366; Chr8:75994988-75995401; chr17:33949714-33950006; Chr3:147864851-147865149; Chr5:173429327-173430483; Chr1:213008609-213009881; Chr15:33478042-33479362; Chr5:1682868-1683456; ChrY:1588659-1589079. B) Bar graphs showing H1T2 ChIP-PCR results represented as % of input. The first seven regions represents the H1T2 occupancy in transcription factors and the next group of three regions represent the H1T2 bound developmental genes while the remaining regions shows the preferential binding of H1T2 to the repeat elements. A negative control region (specific regions of chr 15) was also included for ChIP PCR, where no peaks were present according to the data analysis and visualization. Results are from three independent experiments and error bars represent standard deviation. \*shows comparison with IgG control and # shows comparison with IP+pep group. \*\*\* ( $P \leq 0.001$ ), \*\*( $P \leq 0.01$ ), \*( $P \leq 0.05$ ).



**Figure 6**

Nucleosome IP/Mass spectrometry analysis for the identification of H1T2-nucleosome interacting proteins

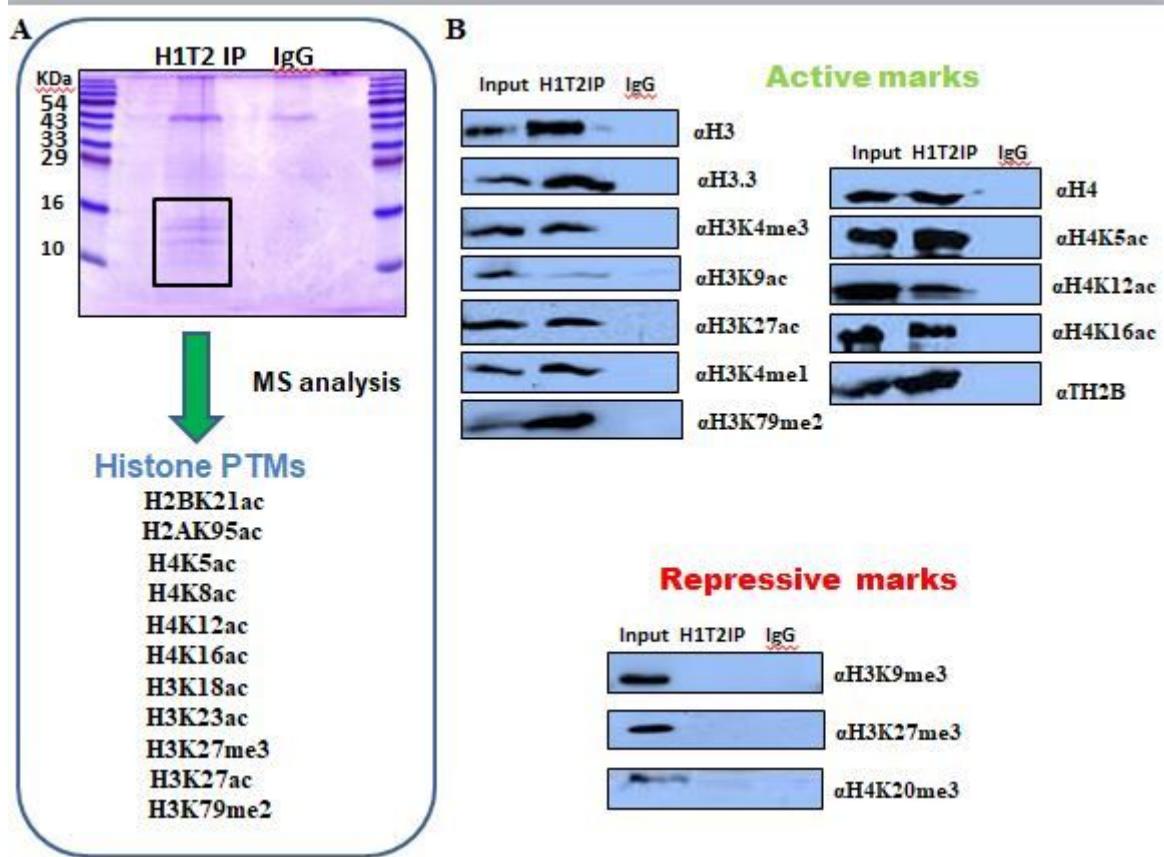
A) Schematic of the nucleosome IP and mass spectrometry protocol used

B) Schematic shows the complete list of interacting proteins of H1T2 associated chromatin with top protein hits with more than 15 unique peptides, identified by mass spectrometric analysis of the H1T2 immunoprecipitated chromatin

C) Gene ontology analysis of the proteins identified in our MS analysis. Diagrammatic representation (bar diagram) of functional classes of H1T2 interacting proteins according to molecular function (P values are given) and pie chart representation of GO biological process associated to the interacting proteins.

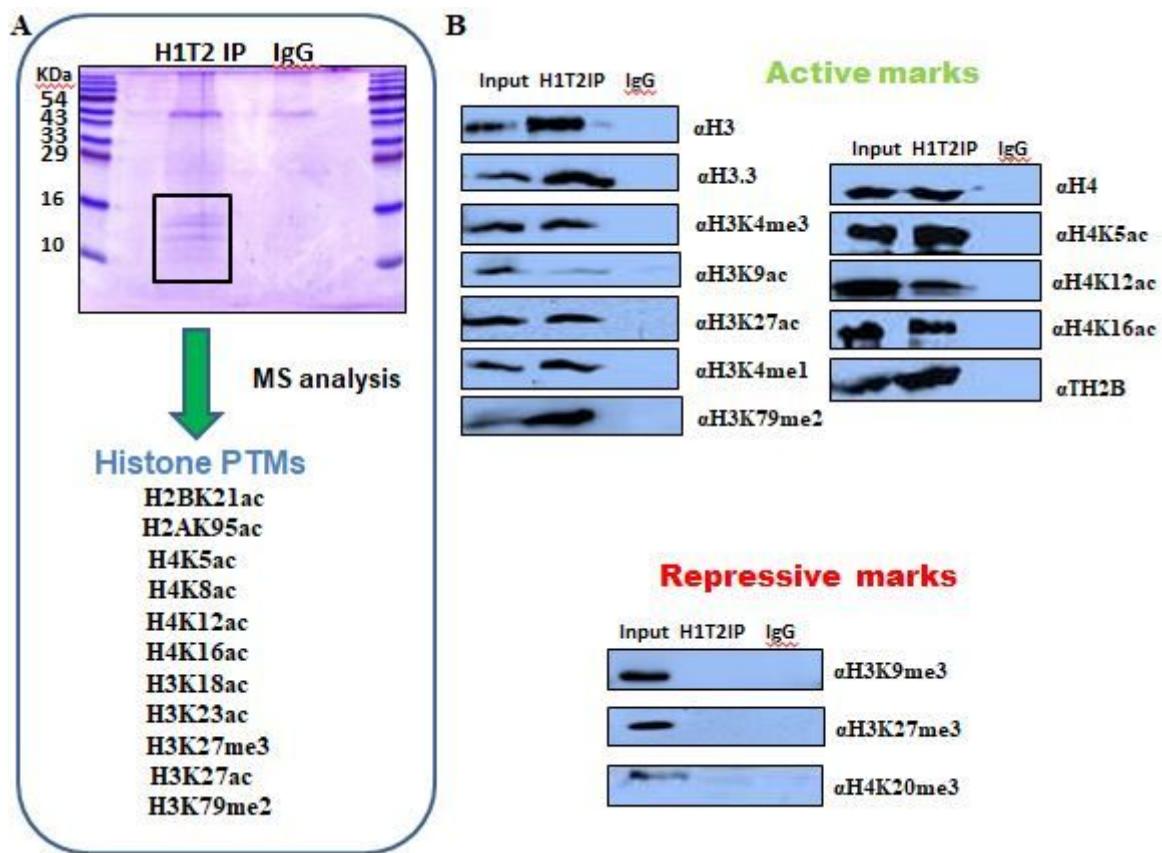
D) Different functional classes of H1T2 interacting proteins mainly related to spermatogenesis and sperm function.

E) Validation of proteins associated to H1T2-chromatin identified by mass spectrometry by IP/ WB analysis.



**Figure 7**

Identification of histone PTMs associated to H1T2 bound chromatin. A) Coomassie stained gel of the acid soluble fractions of H1T2 and IgG immunoprecipitated chromatin samples and the list of histone PTMs identified by MS analysis. B) IP/WB analysis of the histone PTMs associated to H1T2 bound chromatin. Western analysis with antibodies against active histone PTMs gave positive signal upon immunoprecipitation with H1T2 antibody, where as H1T2 bound chromatin was found to be devoid of any repressive modifications.



**Figure 8**

Schematic representation of the process of spermiogenesis and chromatin architecture of H1T2 occupied polar apical region of spermatids. A) Spermiogenesis is a lengthy post meiotic developmental process of spermatogenesis in which the round spermatids undergo several morphological and biochemical modifications to form the mature spermatozoa. Substantial remodelling of chromatin during spermiogenesis involves the replacement of standard histones with histone variants (H1t, H1T2, HILS1, H3.3A, and H3.3B) and testes-specific histones (TH2A, TH2B, and TH3) that should favor an open chromatin conformation. Hyperacetylation of H3 and H4 histones probably facilitates the substantial and continuous repackaging of DNA with transition proteins, further replaced by protamines. Leaky transcription of the spermatid genome is considered as a functional consequence of chromatin remodeling in spermatids (116). B) A schematic model showing the involvement of H1T2 in organizing the nucleoskeleton associated active chromatin architecture of the polar apical region of spermatids.

## Supplementary Files

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

- Additionalmaterial.docx
- Additionalfile1TableS1AnnotationH1T2peaks.xlsx
- Additionalfile2TableS2H1T2Peaksdistribution.xlsx

- Additionalfile3AFigureS1H1T2Peaklengthdistribution.png
- Additionalfile3BTableS3H1T2Peaklengthdistribution.xlsx
- Additionalfile4AFigureS2H1T2Foldchangedistribution.png
- Additionalfile4BTableS4H1T2Foldchangedistribution.xlsx
- Additionalfile5TableS5Repeatselementsoccurrence.xlsx
- Additionalfile6TableS6H1T2LineL1Subclassifications.xlsx
- Additionalfile7TableS7Tablesummitmotifs.xlsx
- Additionalfile8TableS8PromTSSgenesDAVIDGOBP.xlsx
- Additionalfile9TableS9PromTSSgenesDAVIDGOMF.xlsx
- Additionalfile10TableS10H1T2IPMSproteins.xlsx
- Additionalfile11TableS11H1T2IPMSDAVIDGOMF.xlsx
- Additionalfile12TableS12H1T2IPMSDAVIDGOBP.xlsx
- Additionalfile13TableS13H1T2IPMSDAVIDsperm.xlsx
- Additionalfile14TableS14ChIPPCRprimers.xlsx