Published online 15 June 2015

Nucleic Acids Research, 2015, Vol. 43, No. 14 ***6847–6859***
doi: 10.1093/nar/gkv591

Chromatin and extracellular vesicle associated sperm RNAs

Graham D. Johnson[1], Paula Mackie[2], Meritxell Jodar[1,3], Sergey Moskovtsev[2,4] and Stephen A. Krawetz[1,3,*]

[1]Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI 48201, USA, [2]CReATe Fertility Centre, Toronto, ON, M5G 1N8, Canada, [3]Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI 48201, USA and [4]Department of Obstetrics and Gynaecology, University of Toronto, ON, M5G 1E2, Canada

Received November 7, 2014; Revised May 16, 2015; Accepted May 23, 2015

ABSTRACT

A diverse pool of RNAs remain encapsulated within the transcriptionally silent spermatozoon despite the dramatic reduction in cellular and nuclear volume following cytoplasm/nucleoplasm expulsion. The impact of this pronounced restructuring on the distribution of transcripts inside the sperm essentially remains unknown. To define their compartmentalization, total RNA >100 nt was extracted from sonicated (SS) mouse spermatozoa and detergent demembranated sucrose gradient fractionated (Cs/Tx) sperm heads. Sperm RNAs predominately localized toward the periphery. The corresponding distribution of transcripts and thus localization and complexity were then inferred by RNA-seq. Interestingly, the number of annotated RNAs in the CsTx sperm heads exhibiting reduced peripheral enrichment was restricted. However this included *Cabyr*, the calcium-binding tyrosine phosphorylation-regulated protein encoded transcript. It is present in murine zygotes prior to the maternal to the zygotic transition yet absent in oocytes, consistent with the delivery of internally positioned sperm-borne RNAs to the embryo. In comparison, transcripts enriched in sonicated sperm contributed to the mitochondria and exosomes along with several nuclear transcripts including the metastasis associated lung adenocarcinoma transcript 1 (*Malat1*) and several small nucleolar RNAs. Their preferential peripheral localization suggests that chromatin remodeling during spermiogenesis is not limited to nucleoproteins as part of the nucleoprotein exchange.

INTRODUCTION

Following spermatogenesis the mature male gamete remains in a transcriptionally and translationally quiescent state yet harbors a unique population of RNAs (1–4). Many of these transcripts possess a clearly defined role during spermatogenesis providing a rich source for the discovery of male fertility biomarkers (5). Still other sperm RNAs await functional assignment. Notably, these sequences in sperm may not require translation to be of functional importance within the cell or following fertilization (6,7). Regardless of their coding potential attributing function to sperm RNAs has been complicated by a poor understanding of basic principles governing these transcripts including their cellular localization. Some RNAs likely remain bound to protein chaperones as ribonucleoprotein particles that are observed following their transcription in round spermatids (8,9). Yet, others, may complex with the paternal chromatin (10,11). Nevertheless, assigning functional roles to all sperm transcripts necessitates understanding their cellular localization.

Various studies have used fluorescence *in situ* hybridization to establish the presence of specific RNAs within the context of the spermatozoon (12,13). However, localizing an internally positioned target within the mature sperm head is challenging due to the extreme compaction of this structure. The stability of proteins relative to RNAs offers greater flexibility in their immunological localization as chemical agents can be employed to increase penetrance and thereby the chance of detection (14). However, microscopy is not conducive to identifying the major sites of sperm RNA compartmentalization in a high-throughput manner.

As illustrated in Figure 1, the extra-nuclear and intranuclear compartments provide at least two sites within the limited volume of the sperm head that should be capable of harboring RNA. Polymerase chain reaction (PCR) analysis of others has shown that specific transcripts are lost when sperm nuclei are demembranated (15). This is consistent

*To whom correspondence should be addressed. Tel: +1 313 577 6770; Fax: +1 313 577 8554; Email: steve@compbio.med.wayne.edu

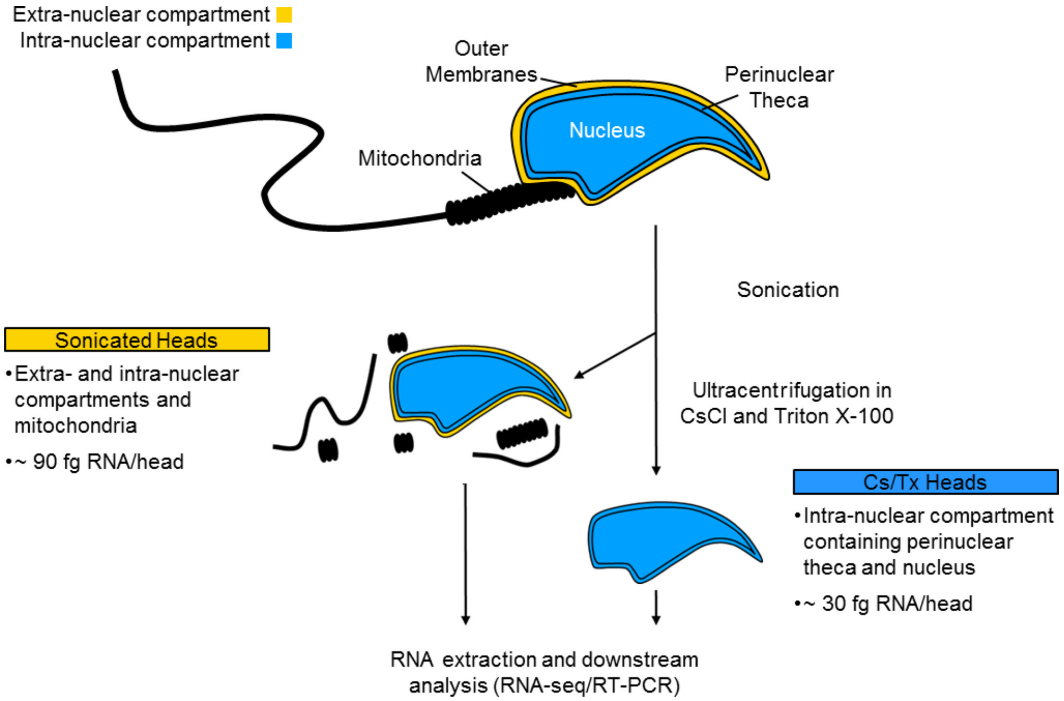

Figure 1. Potential sites of RNA localization in sperm. RNAs extracted from sonicated sperm (SS) can localize to three generalized compartments: (i) Mitochondria; (ii) extra-nuclear compartment that includes the plasma membrane, the acrosome and associated membranes; (iii) the intra-nuclear compartment which includes the nucleus, the nuclear envelope and the perinuclear theca. Fractionation and demembranation produces a population of sperm heads that remain associated with the perinuclear theca but lack the extra-nuclear compartment as well as the mitochondrial sheath and tail. RNA-seq analyses of transcripts extracted from SS or Cs/TX heads were used to identify compartment specific patterns of RNA enrichment.

with previous detergent based sperm purification methods which are expected to compromise membranes, impacting RNA-seq profiles (16).

Fractionation of nuclei and subsequent isolation and analysis by RNA-seq is a routine procedure in somatic cells (17) but has yet to be attempted in sperm. To discern the global pattern of transcript compartmentalization in mouse spermatozoa, total RNA was extracted from sonicated sperm (SS) and detergent demembranated gradient fractionated (Cs/TX) sperm heads. RNA quantification demonstrated that the majority of RNA in sperm is associated with the peripheral membranes which are lost following treatment with detergent. To identify sperm transcripts that exhibited preferential enrichment within the intra- or extra-nuclear compartments (Figure 1) RNAs extracted from SS and Cs/TX sperm heads were subjected to RNA-seq analysis. The Cs/TX heads exhibited suppressed coverage of annotated RNAs supporting their general depletion from the nucleus and perinuclear theca and localization within the outer sperm membranes. As expected and observed by RNA-seq, transcripts enriched in these samples displayed a reduced peripheral enrichment when evaluated by RT-PCR and included RNAs associated with the cytoskeleton and spermatogenesis. In comparison, within the SS samples, RNA-seq analysis highlighted a set of preferentially extra-nuclear localized RNAs, many of which are associated with extracellular vesicles. This association was supported by a cross-species comparison of the mouse and human homologs enriched in the sperm extra-nuclear compartment and exosomes recovered from semen. These results contribute to the growing evidence for the presence of exosomes on the surfaces of the male gamete (18–24). Additional classes of RNAs also appeared to be associated with the outer sperm membranes including nuclear-encoded mitochondrial transcripts and some nuclear RNAs. Their preferential peripheral localization suggests that chromatin remodeling during spermiogenesis is not limited to nucleoproteins as described in the following.

MATERIALS AND METHODS

Sonicated and Cs/Tx sperm head preparation and RNA extraction

Sperm fractions were prepared and RNAs extracted essentially as described (Figure 1;25,26,27). In brief, mature spermatozoa from transgenic line HP3.1 (28) were isolated from cauda epididymides and vas deferens harvested from individual four month old transgenic mice on ice into 50 mM Tris–HCl, pH 7.4, buffer. The cells were washed twice following filtration through an 80 micron mesh, resuspended in 0.5 ml 50 mM Tris–HCl, pH 7.4, buffer and subjected to sonication with a TekMar TM-50 sonic disruptor (TekMar, Cincinnati, OH, USA) at 70% maximum output for 2 min on ice to separate heads from tails and to lyse potential cellular contaminants. The sonicated sperm suspension was washed twice and $1 - 5 \times 10^7$ sperm per sample were diluted to a volume of 7.5 ml in 1 M sucrose buffered with 50 mM Tris–HCl, pH 7.4, containing 5 mM $MgCl_2$. A triple-step gradient was prepared by overlaying diluted samples onto cushions consisting of 2 M sucrose buffered with 50

mM Tris–HCl, pH 7.4, containing 5 mM $MgCl_2$ and 0.45 g/ml CsCl buffered with 25 mM Tris–HCl, pH 7.4, buffer containing 0.5% Triton X-100. Sperm heads were recovered by ultracentrifugation at 75 600 × rcf for 45 min at 4°C and subsequently washed twice with 25 mM Tris–HCl, pH 7.4, buffer containing 0.5% Triton X-100. Following resuspension in 0.5 ml RLT, a guanidine-thiocyanate lysis buffer (Qiagen) supplemented with 1.5% β-mercaptoethanol (Amresco), 0.2 mm stainless steel beads were added to the samples and the Cs/TX sperm heads mechanically lysed with a Disruptor Genie (Scientific Industries, Inc., Bohemia, NY, USA). After the addition of an equal volume of Qiazol (Qiagen) nucleic acids were recovered using the RNeasy system (Qiagen). Following sonication and washing sonicated sperm RNAs were equivalently extracted. Total RNAs were DNased (Turbo DNase, Ambion) and subsequently subjected to RT-PCR with intron-spanning primers to *Prm2* (Supplementary Dataset 1; 28).

Isolation of RNA from human sperm and seminal vesicles

Normozoospermic semen (WHO 2010) was collected from healthy donors with proven fertility by masturbation after 2–5 days of sexual abstinence. Following analysis excess semen was stored at −80°C. Samples were thawed at room temperature for 20 min and exosomes were isolated from the seminal fluid by differential centrifugation at 4°C as follows: 3,000 rcf for 10 min to pellet sperm fraction which was stored at −80°C, followed by 12,000 rcf for 45 min to pellet cellular debris and larger vesicles. The supernatant was centrifuged at 110,000 rcf (Beckman Coulter Optimax MLA-13 rotor) for 70 min. The resulting exosomal pellet was washed with 1 ml Phosphate buffered saline (Dulbecco's Phopshate Buffered saline), pelleted again at 110 000 rcf for 70 min and stored at −80°C until RNA extraction. The sperm pellet was thawed and processed through a 50% PureSperm (Nidacon) cushion prior to RNA extraction. Exosomal and sperm RNA-seq libraries were prepared as outlined below.

Transmission electron microscopy

Aliquots of SS and Cs/TX murine sperm were subjected to transmission electron microscopy at the Microscopy Imaging Laboratory (University of Toronto, ON, USA) using standard protocols. Briefly, cells were pelleted by centrifugation, fixed in a 4% paraformaldehyde with 1% glutaraldehyde buffered with 0.1 M sodium phosphate monobasic buffer (NaH_2PO_4), pH 7.2, at 4°C and subsequently post-fixed with 1% osmium tetroxide buffered with 0.1 M sodium phosphate dibasic buffer (Na_2HPO_4), pH 7.2, for 1 h. Dehydration of the cells was achieved with a series of increasing concentrations of ethanol to allow for infiltration of the Epon embedding resin. The resin was sectioned on an ultramicrotome (Leica) and transferred to a copper transmission electron microscopy (TEM) grid. Reynolds lead citrate and 5% uranyl acetate was used for contrast staining and micrographs were taken on a Hitachi H 7000 at 75 Kv (25uA beam current) using an AMT XR-60 digital camera.

Sequencing library preparation and analysis

Stock synthetic spike-in External RNA Controls Consortium ERCC (ERCC) RNAs (Invitrogen) were diluted 1:10 000 and pooled with 5 ng of total sperm RNAs prior to reverse transcription and amplification using the Seqplex system (Sigma; 29). Pre-amplified cDNA libraries were subjected to sequencing library construction (DNA Ultra-Low, NEB) followed by 50 cycles of paired-end sequencing on the Illumina Hi-Seq 2500 platform. Sequencing reads were aligned simultaneously to the mouse genome assembly Mm10 in addition to the ERCC FASTA sequences with Tophat2 (version 2.0.12; 30) using the following parameters: tophat2 -r 30 –mate-std-dev50 –no-coverage-search -G genes.gtf. Novel transcript structures were assembled using Cufflinks (version 2.1.1; 31) after removal of PCR-duplicates with Picard tools (http://picard.sourceforge.net). Assembled transcripts exceeding twice the average fragment length (135 bp) were combined with UCSC gene annotations (Mm10).

Preliminary analysis of aligned sequencing reads was carried out with the Samtools (version 0.1.19; 32) and Bedtools (version v2.19.1–2; 33) suites. Ranked predicted fold change and RPKM values for enrichment analyzes were calculated using uniquely aligned reads and the GFold package (34). A lower limit of detection was determined by analysis of ERCC coverage. Log normalized ERCC RPKM values were well correlated across samples ($r^2 > 0.90$; Supplementary Figure S1) demonstrating that inter-sample comparisons of RNAs exceeding 116 RPKM were reliable. To derive enrichment predictions the Cs/TX samples were independently compared to the SS sample and the intersection of transcripts with an absolute GFold value greater than zero and a RPKM value exceeding the ERCC cutoff were considered for RT-PCR validation. Differential enrichment analysis of the human ejaculate exosomes and human sperm RNA-seq samples as well as mouse embryo and oocyte analyses were performed with both with the HT-Seq and DEseq2 software packages (35–37) as well as GFold (38).

Plots and browser graphics were generated using the ggplot2 R package and the UCSC genome browser, respectively. Ontological analyzes were performed using GOrilla and associations presented as False Discovery Rate corrected *q*-values (39). A hypergeometric probability function in R was used to assess the likelihood of overlap observed between sperm datasets and nuclear encoded mitochondrial transcripts taken from the Mitocarta database (40) and human ejaculate RNA-seq datasets.

Genomic coordinates of repeat elements were retrieved from the Repeat Master track of the UCSC genome browser and analyzed with Bedtools. To determine the relative coverage of repeat families unmapped and multiply-mapped reads were aligned to the rodent RepBase v19.0 repeat sequences (41). Following alignment the number of uniquely aligned sequencing reads corresponding to each repeat family was summed and those values combined with those obtained from the initial unique alignments from the same sample. Estimates of subfamily abundance of sperm LINEs were determined by equivalently subsampling and aligning FASTQ files to canonical sequences for each of the actively

transcribed mouse LINE elements retrieved from RepBase (42). Only uniquely aligned sequencing reads were considered.

RT-PCR validation

Samples ($n = 6$) used for RT-PCR validation of the RNA-seq results were isolated as above with minor changes. Sonicated and washed sperm from a single mouse were counted on a hemocytometer and split into two equal aliquots. Subsequently sperm were either subjected to ultracentrifugation through a triple-step gradient or left on ice prior to nucleic acid extraction. SS and Cs/TX pellets were divided into two aliquots and used for separate RNA and DNA extractions. Prior to RNA extraction sperm samples were lyzed with RLT buffer supplemented with 1.5% β-mercaptoethanol and Qiazol as above and following disruption with 0.2 mm stainless steel beads stored at −80°C. Sperm pellets were digested with proteinase-K in the presence of β-mercaptoethanol and DNAs recovered by phenol–chloroform extraction with phase-lock tubes (5Prime) followed by ethanol precipitation with linear acrylamide. DNAs were quantified with Picogreen (Invitrogen) and used to equilibrate the stored RNA extractions with respect to sperm concentration per volume of lysate. For each individual animal equal volumes of the SS and Cs/TX lysate were passed through the RNeasy column such that RNA was recovered from equivalent number of match sonicated and demembranated sperm heads. Subsequently, eluted RNAs were DNased as above prior to quantification with Ribogreen (Invitrogen). First-strand cDNA was synthesized using equal volumes of total RNA with the addition of 1 μl of diluted ERCC, (1:1000), 40 units of RNase Block (Agilent), 100 ng of random primers (Invitrogen) and the Superscript III kit (Invitrogen). Diluted SS and Cs/TX cDNA products were analyzed in parallel triplicate reactions by real-time PCR using Hot-Star Taq Polymerase system (Qiagen) and 0.5 mM of each forward and reverse oligonucleotide primer (Supplemntary Datatset 1) for 50 cycles. Delta-Ct values are calculated as follows: $\Delta C_t = C_t^{Cs/TX} - C_t^{SS}$.

RESULTS AND DISCUSSION

Isolation of compartmentalized RNAs

The highly condensed mature spermatozoon possesses distinct structural features and organelles that may harbor site-specific populations of RNAs. To elucidate whether patterns of RNA retention vary between the peripheral structures contained in the extra-nuclear compartment of the gamete and those found within the intra-nuclear compartment, sperm samples were briefly sonicated and aliquots subjected to gradient fractionation in the presence of detergent. Electron microscopy confirmed the separation of sperm heads and tails following sonication as well as the absence of contaminating cell types (Figure 2). Despite the absence of an intact acrosome the sonicated sperm (SS) heads retained the inner acrosomal membrane (IAM) as well as portions of the outer acrosomal membrane (OAM; 43). Ultracentrifugation of the SS heads through a sucrose cesium chloride step gradient and detergent demembranation produced sperm heads (Cs/TX) free of tail and mitochondrial sheath remnants. In addition, the Cs/TX heads lacked all peripheral membranes while retaining the perinuclear theca. This network of cytosolic and nuclear proteins attaches the inner acrosomal membrane to the nuclear envelope and is resistant to extraction with both ionic and non-ionic detergents such as Triton X-100 used in this study (43,44).

Isolation of RNAs from SS and Cs/TX heads demonstrated that the long RNA fraction (>100 nt) in sperm is largely peripherally localized (Figure 3). Considering the RNA recovered from Cs/TX sperm head fraction, 61.45% of the 87 femtograms of RNA retained in each spermatozoon must be found within the membranes removed during isolation of the nucleus and perinuclear theca (Figures 1 and 2). This supports prior work demonstrating that treatment with detergents alters the transcript profile of the sperm cell (16). Resolution of whether the RNAs recovered from the Cs/TX heads reside within the nucleus and/or the perinuclear theca remains a technical challenge. To better understand patterns of transcript localization throughout the male gamete RNA from the SS and Cs/TX sperm heads was subjected to RNA-seq analysis.

Library characteristics

Mitochondrial encoded transcripts. The efficacy of the sample fractionation established by microscopy was confirmed by the reduced sequencing coverage of the mitochondrial genome in the Cs/TX sperm heads (Figure 4A). In contrast, the SS heads exhibited a 10-fold enrichment in coverage of mitochondrial RNAs corresponding to >7% of all uniquely aligning reads. This value is within the range (3.5–36%) observed in other mouse sperm RNA-seq datasets (3,4,45) indicative of significant differences in sample and library preparation that has been resolved (16).

Ribosomal RNAs. In addition to a greater proportion of mitochondrial transcripts, the SS heads possessed greater amounts of ribosomal RNAs (rRNAs, Figure 4A). Together these transcripts and their associated pseudogenes constituted 28 and 17% of all unique sequencing reads in the SS and Cs/TX libraries, respectively, even though unlike other cell types, sperm rRNAs are fragmented (46,47). It is not obvious why the ribosomal transcripts persist at elevated levels in both sperm fractions. Within the SS samples the elevated levels of rRNA reflect the additional contribution of membrane-associated ribosomal transcripts in addition to those retained following isolation of the Cs/TX sperm head structures (Figures 1 and 2). The abundance of the rRNAs in SS relative to demembranated Cs/TX samples may reflect entrapment of cytoplasmic constituents during condensation and concurrent membrane restructuring and/or be acquired from or maintained in external structures such as membrane-bound extracellular vesicles. These structures have been observed *in vivo* in physical association with spermatozoa (18,21,23,24). Unlike cytoplasmic droplets which are osmotically sensitive structures formed by maturing spermatids as they expel their cytoplasm and often lost during routine sperm processing (48),

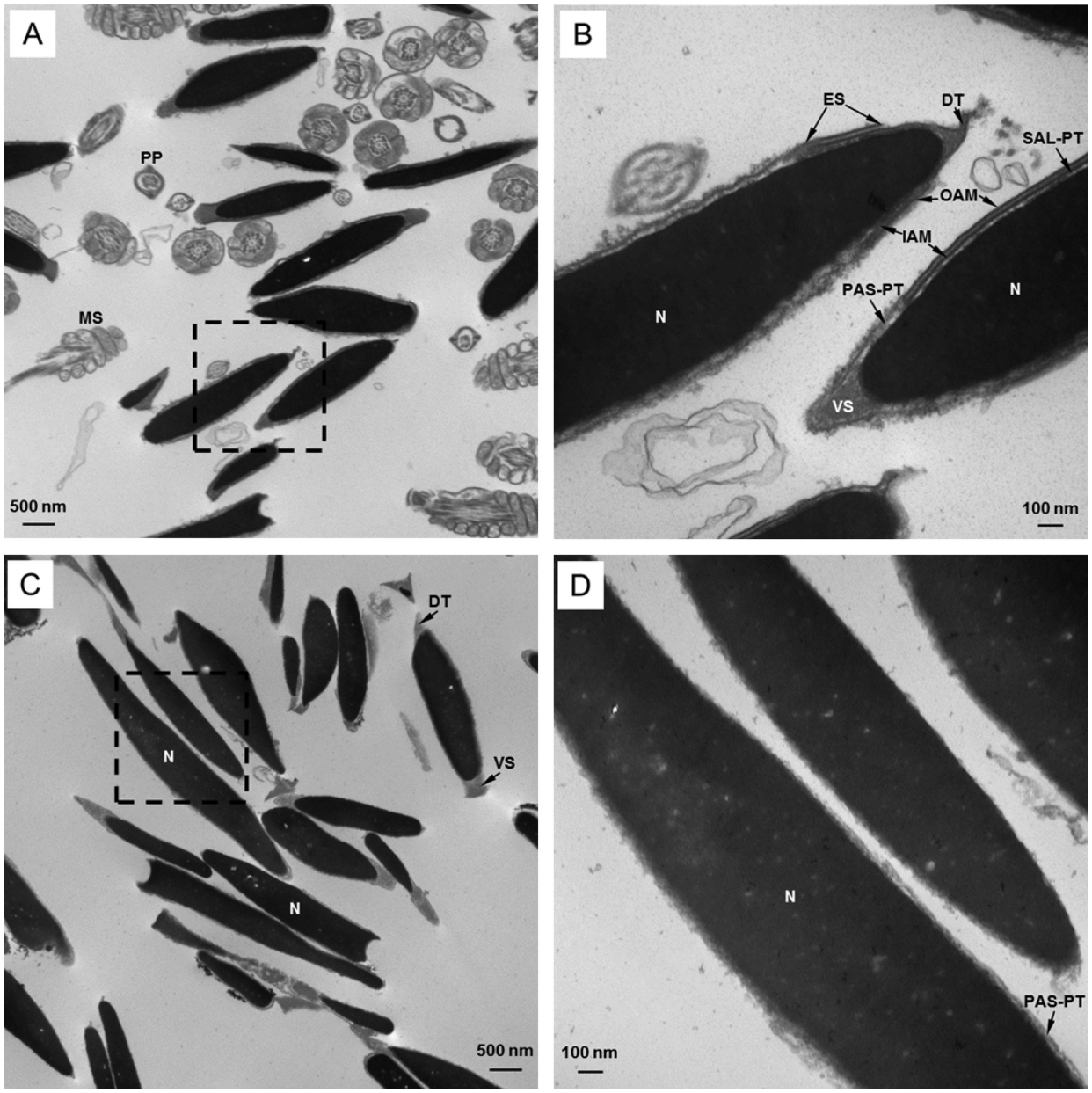

Figure 2. Cs/TX fractionated sperm heads lack peripheral membranes. (**A** and **B**) Sonication separates the tail from the sperm head while preserving the outer membrane structures. Fragments of the sperm tail (PP) and mitochondrial sheath (MS) are observed following sonication whereas potential somatic contaminants are not. (**C** and **D**) Fractionation through a sucrose–cesium chloride gradient in the presence of Triton X-100 enriches a population of sperm heads free from tail and mitochondrial remnants. The Cs/TX sperm nucleus remains surrounded by the perinuclear theca (PT) as evidenced by the dorsal tip (DT) and ventral spur (VS) structures but lacks the peripheral membranes observed in the sonicated sample. Dashed boxes in (A and C) (Magnification × 12 000; scale bar = 500 nm) are presented at higher magnification in (B and D) (Magnification × 50 000; scale bar = 100 nm). Equatorial segment (ES); inner acrosomal membrane (IAM); mitochondrial sheath (MS); nucleus (N); outer acrosomal membrane (OAM); perinuclear theca (PT) postacrosomal sheath of the PT (PAS-PT), principle piece (PP) of the sperm tail; subacrosomal layer of the PT (SAL-PT).

sperm-bound exosomes fuse with the gamete (22). Further support that the elevated levels of ribosomal transcripts in the SS cells relative to that observed in the Cs/TX heads may result from the packaging of these sequences in extracellular vesicles is the recent observation that this class of RNA is the most abundant observed in exosomes retrieved from human semen (20).

Greater than 17% of the uniquely aligned reads in the Cs/TX RNA-seq libraries corresponded to the ribosomal transcripts. Considering that rRNAs are not commonly observed in somatic nuclear RNA (17) and in the absence of evidence for their translocation into the nucleus these transcripts are presumed to be intimately associated with but likely outside of the highly condensed nucleus. In this manner these fragmented transcripts might persist within partially assembled ribosomes as has been observed in cytoplasmic droplets (46). The abundance of the rRNAs and their associated proteins would suggest that they are not

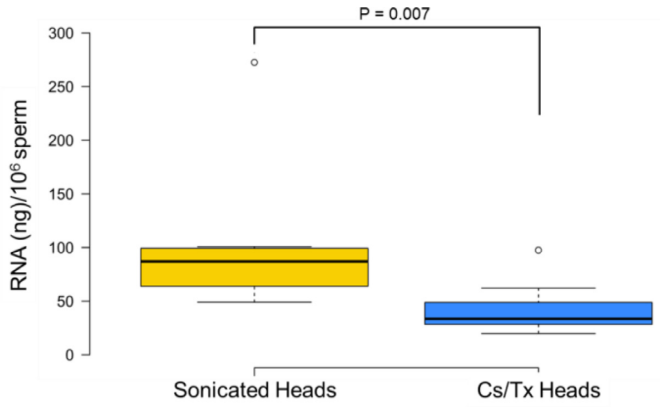

Figure 3. The majority of the RNA retained in sperm is localized to the sperm periphery. RNAs were quantified after recovery from equal numbers of sperm heads from the same animal ($n = 9$) following sonication and Cs/TX fractionation. The SS cells contained significantly more RNA than matched Cs/TX heads ($P = 0.007$, two-tailed paired t-test). After accounting for the presence of the intra-nuclear compartment within each SS sperm cell ~53.47 fg of RNA is associated within the peripheral structures of the extra-nuclear compartment. Center lines show the medians (87.03 and 33.56 fg of RNA per sperm head, SS and Cs/TX, respectively); box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots and represent matched SS and Cs/TX samples from the same animal.

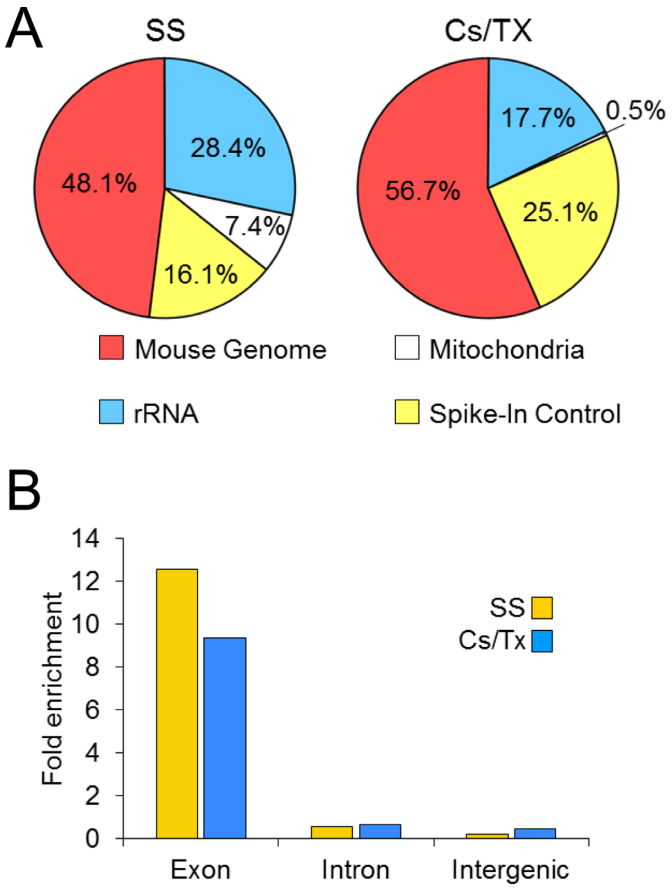

Figure 4. Sequencing coverage and enrichment in annotated regions vary between the SS and Cs/TX sperm heads. (**A**) RNAs extracted from SS or Cs/TX samples were subjected to RNA-seq. Following alignment the percentage of uniquely aligned sequencing reads attributed to specific classes of sequences was determined. The largest contribution of alignments in both the Cs/TX and SS RNA-seq libraries corresponded to the nuclear genome in a treatment dependent manner. Coverage of the rRNAs and the ERCC control RNAs in libraries also varied by preparation. The reduced sequencing coverage of the mitochondrial genome in the Cs/TX sperm heads confirmed the efficacy of the sample fractionation. (**B**) SS cells exhibited a greater enrichment of uniquely aligned sequencing reads in exonic regions (coding and non-coding) relative to that observed for the Cs/TX samples. Fold enrichment values for uniquely aligned sequencing reads within exons, introns and intragenic regions are presented for SS and Cs/TX RNA-seq libraries. Enrichment was determined by dividing the percentage of genomic bases in an element class by the percentage of unique sequencing reads aligned to that class.

completely expelled during the final stages of spermiogenesis. This may lead to the sequestration of the translational machinery within the perinuclear theca as has been observed in proteomic analysis of detergent-treated sperm nuclei (49). As these proteins are not known to contribute to this cytoskeletal nucleus-acrosomal interface they may be passively retained along with bound rRNAs through prior association. Whether this mechanism resolves other RNA associations within the sperm nucleus (15) remains to be established.

ERCC control RNAs. To determine a lower limit of detection synthetic ERCC (External RNA Controls Consortium) RNAs were added in equal amounts to all samples prior to cDNA synthesis. Coverage of the control transcripts was highly correlated across all samples ($r^2 > 0.90$; Supplementary Figure S1) while their absolute levels varied between fractionation methods. Of all uniquely aligning sequencing reads approximately, 25% Cs/TX and 16% SS sperm head reads corresponded to the ERCC transcripts in the RNA-seq libraries, respectively. The increased representation of these synthetic sequences in the Cs/TX sequencing libraries was likely a consequence of the reduced availability of biological and or accessible RNA templates following the loss of the outer sperm membranes. Supporting this observation the Cs/TX heads exhibited a 10% increase in the percentage of reads aligned to the nuclear genome with a concomitant decrease in coverage of annotated regions relative to the SS heads (Figure 4B). This is reminiscent of the observation that the use of detergents to remove somatic cell contaminants from human sperm negatively impacts transcript levels compared to samples purified by other means (16). Representation of intronic sequences were equivalently suppressed in all libraries (Figure 4B) reflecting the cessation of nascent RNA production prior to nuclear condensation.

Identification of putative SS and Cs/TX enriched transcripts

Mature spermatozoa possess a unique morphology and a comparatively small quantity of RNA. These limitations required that sonicated sperm containing nuclei surrounded by perinuclear theca be compared to the demembranated Cs/TX heads rather than independently sampling compartments as can be accomplished for somatic cells (17). By mass, ~39% of the RNA within a sonicated sperm cell can be attributed to the structures contained within a Cs/TX head (perinuclear theca and nucleus; Figures 1 and

2). Therefore transcripts enriched within the more external regions of the spermatozoon will be diluted by RNAs present within the mitochondrial sheath, the perinuclear theca and nucleus, reducing their representation when resolved by RNA-seq. Consequently, the sampling methods employed herein are expected to under- and over-report putative externally (SS) and internally (Cs/TX) localized transcripts, respectively. In addition, the limited data suggested an alternative to standard differential expression analysis pipelines was necessary to identify patterns of RNA localization within the mature spermatozoon. Accordingly, GFold was used to rank transcripts by the posterior distribution of their normalized expression values and candidate transcripts putatively enriched within either the SS or Cs/TX heads were selected for downstream RT-PCR analysis.

The GFOLD analysis identified a total of 308 transcripts that were preferentially enriched in the Cs/TX RNA-seq libraries. Together this set of RNAs included 184 annotated and 116 novel transcripts as well as all ERCC control transcripts exceeding the minimum threshold of detection (Supplementary Dataset 2). Ontological analysis of the annotated RNAs enriched in these samples identified terms related to spermatogenesis and cytoskeleton (Supplementary Dataset 3). Within this set of RNAs, the *Prm1* transcript, as well as *Erich2* and *Fam71e2* (formerly *4933404M02Rik* and *4930401F20Rik*, respectively), were previously identified within the mouse sperm nucleus by RT-PCR though their preferential localization within the gamete could not be inferred from that study (15). As expected fewer transcripts were predicted to be enriched in the SS RNA-seq libraries relative to the Cs/TX heads ($n = 152$; Supplementary Dataset 4). This set of putatively enriched sonicated sperm RNAs included many nuclear-encoded mitochondrial protein mRNAs ($n = 20$; $P < 1.9 \times 10^{-20}$; Figure 5B). These RNAs comprised 17% of the transcripts enriched in the SS RNA-seq libraries and only 2% of the Cs/TX enriched RNAs ($n = 4$; 40). This association was supported by ontological analysis of the predicted SS RNAs which revealed terms related to cytochrome-c oxidase activity and the mitochondrial membrane (Supplementary Dataset 5). Together this likely represents the evolutionarily conserved shuttling of cytosolic ribosomes bound with nuclear-encoded mitochondrial mRNAs to the surface of the mitochondria prior to condensation (50–52). These transcripts are expected to be lost along with the mitochondrial sheath and proximal structures during fractionation of the Cs/TX heads.

Additional ontological categories were also significantly associated with the RNAs enriched in the SS sample. These included terms related to exosomes ($q < 1.05 \times 10^{-9}$; Supplementary Dataset 5). The enrichment of exosomal-associated RNAs in the SS RNA-seq libraries supports prior observations that the outer sperm membranes are associated with extracellular vesicles. This set of transcripts includes guanine nucleotide binding protein (G protein), gamma 5 (*Gng5*; Figure 5A) for which both the RNA and protein products have been observed in exosomes recovered from multiple tissues in human and mouse (53–55). Interestingly, of the 42 SS enriched transcripts associated with the exosome ontological category, only three were present above background in both Cs/TX samples. This set of SS RNAs exhibited a median fold increase of 5.9 (range = 2.6 – 192.9) relative to the Cs/TX samples demonstrating that these transcripts are externally localized relative to the perinuclear theca and nucleus and therefore may be delivered to the gamete by exosomes and incorporated during epididymal maturation or within the vas deferens. Further, >45% of all mouse homologs enriched in the SS RNA-seq library were also differentially expressed in human ejaculate exosomes relative to sperm collected from the same semen samples ($P < 6.4 \times 10^{-19}$, hypergeometric probability test; Supplementary Figure S2). By comparison only 8% of the RNAs predicted to be enriched in the Cs/TX libraries were significantly enriched in the human exosome dataset ($P > 0.8$). The depletion of these transcripts following exposure to Triton-X100 reflects the loss of external sperm membranes that serve as sites of exosome attachment and likely harbor RNAs some of which may possess an origin other than sperm (19). To some extent this resolves the exosomal RNA concentration abundance question (56).

To test the proposed link between the detection of exosomal RNAs and the presence of external sperm membranes GFold was used to identify transcripts predicted to be differentially enriched in either the SS sample or in a previously published mouse sperm RNA-seq dataset prepared following detergent incubation (4). Differentially expressed human ejaculate exosome homologs were present in the set of SS enriched RNAs at levels significantly exceeding that expected by chance ($P < 1.2 \times 10^{-26}$, hypergeometric probability test). This relationship was not observed in the detergent treated sperm ($P = 0.99$). Similarly the human exosome homologs enriched in the SS sample were present at elevated levels ($P < 2.2 \times 10^{-16}$, Mann–Whitney U-test; Supplementary Figure S3) and exhibited a stronger ontological enrichment in exosome associated terms than their counterparts enriched in the detergent treated sample ($q < 3.83 \times 10^{-49}$ and $q < 3.84 \times 10^{-16}$, respectively). Analogous to the Cs/TX sperm head fractionation used in the current study, detergent treatment (4) would have been expected to produce a sperm nucleus surrounded by the perinuclear theca at the expense of external membranes. Since these external structures harbor the majority of spermatozoal RNA in addition to serving as the site of exosome attachment their loss would be expected.

Several of the SS transcripts included well known somatic nuclear RNAs such as Small Nucleolar RNAs, H/ACA Box (*Snora23*, *Snora52*, *Snora81*), Small Cajal Body-Specific RNA 13 (*Scarna13*) and the long non-coding RNA (lncRNA) metastasis associated lung adenocarcinoma transcript 1 (*Malat1*; Figure 5C). The persistence of an elevated level of *Malat1* in all sperm samples and its extra-nuclear enrichment can be ascribed to the presence of a triple helix structure at its 3′ end which stabilizes the transcript (57). Association of this nuclear RNA with more external regions of the spermatozoon suggests *Malat1*, Snora RNAs, *Scarna13* and potentially other ncRNAs are expelled along with the nucleoplasm during condensation. Bereft of a nucleous or Cajal bodies these transcripts appear as if they may not be essential to the mature male gamete or the next generation (15). However, the retention of a minor pool of nuclear-associated *Malat1* tran-

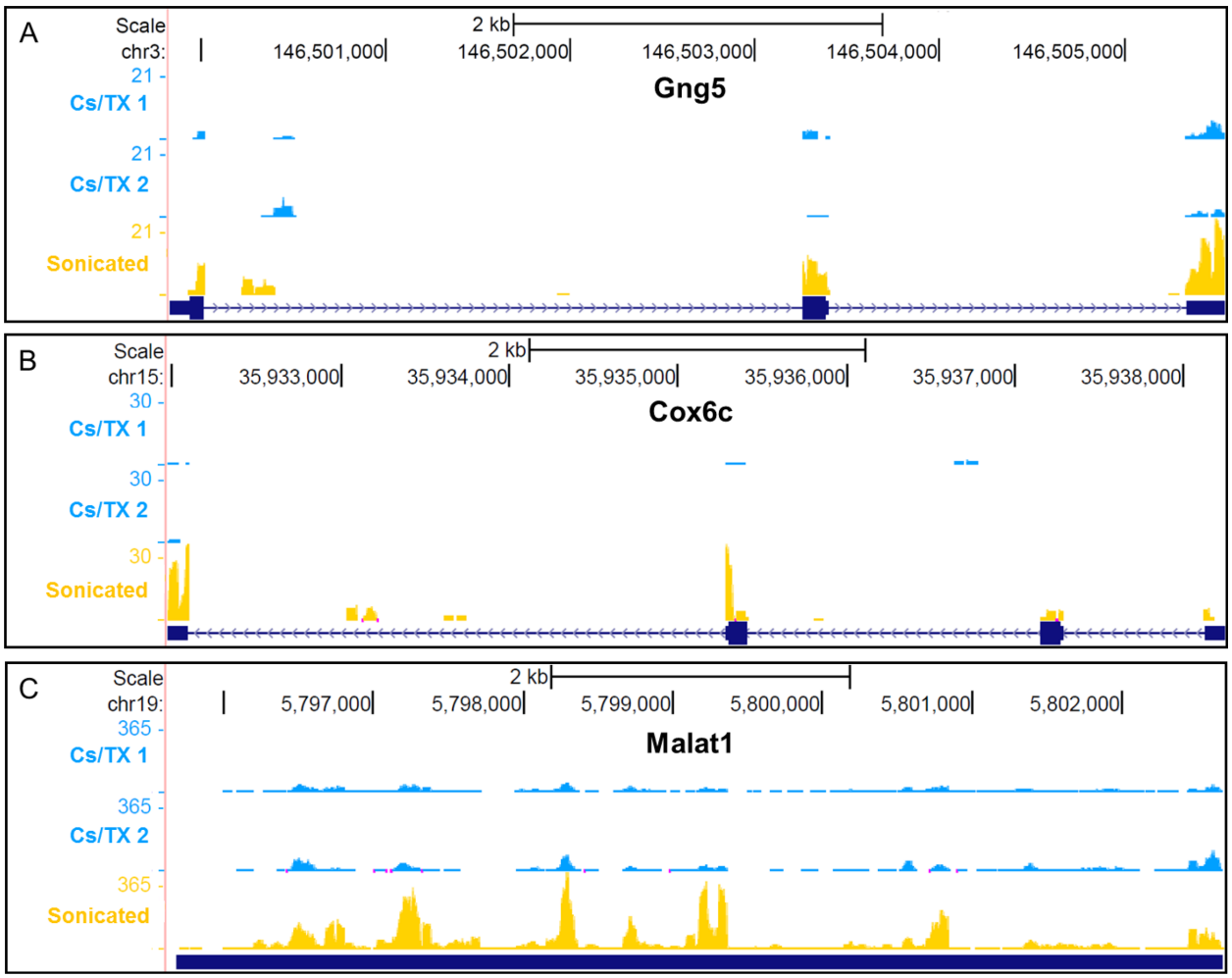

Figure 5. RNAs representative of specific cellular functions and organelles are enriched within the peripheral sperm membranes. Coverage of uniquely aligned sequencing reads are presented for three transcripts on the UCSC genome browser. Reads per million uniquely aligned reads (RPM) are presented on the Y-axis for each sample with the maximum value corresponding to that observed in the sonicated sample (SS) for each RNA panel. The direction of transcription is depicted by arrows. (**A**) *Cox6x* is nuclear-encoded mitochondrial transcript. (**B**) The *Gng5* RNA and corresponding membrane-associated G protein product are enriched in exosomes from several tissues and species. (**C**) *Malat1* is a conserved long non-coding RNA that is enriched at active loci and within paraspeckles where it recruits splicing factors.

scripts (Figure 5C) may result from the direct interaction of these lncRNAs with regions of the sperm genome which remain in a 'poised' chromatin configuration (58–61) and clarification is being actively pursued.

A series of oligonucleotide primers were designed to validate patterns of preferential transcript compartmentalization in sperm by RT-PCR. RNAs from individual mice ($n = 6$) were extracted from equal numbers of SS or Cs/TX sperm heads and reverse transcribed with random primers. The addition of ERCC RNAs to the first-strand synthesis reaction served as a loading and synthesis control (Figure 6A). Although the fold-changes estimated from the RNA-seq datasets and delta-Ct values were well correlated ($r^2 = 0.86$; Figure 6B) all transcripts queried were found to be enriched in the sonicated sperm samples suggesting that globally, RNAs are peripherally localized in the mature male gamete (Figure 2) consistent with mass distribution. This included the first and third most enriched protein-coding transcripts in the Cs/TX RNA-seq libraries, Ankyrin repeat and MYND domain containing 1 (*Ankmy1*) and ecotropic viral integration site 5-like (*Evi5l*; Supplementary Dataset 2). However, the RNAs predicted by GFold to be enriched in the Cs/TX heads exhibited significantly reduced delta-Ct values compared to those observed for the set of transcripts predicted to be enriched in the sonicated sperm ($P < 0.004$, Mann–Whitney U-test; Figure 6C). These results suggest that while the majority of RNA within the mature spermatozoon can be localized to the periphery of the cell their relative proportions vary between transcripts as some are expected to be differentially retained within the intranuclear compartment. On one hand, the relative depletion of *Malat1* and *Snora81* RNAs (Table 1) from the Cs/TX samples predicts that in the male gamete these transcripts are unlikely to contribute to nuclear organization in a manner reminiscent of that observed in somatic cells. On the other hand, some *Malat1* remains within the Cs/TX heads and may be sufficient to contribute to the packaging of the limited number of histone-associated promoters present in

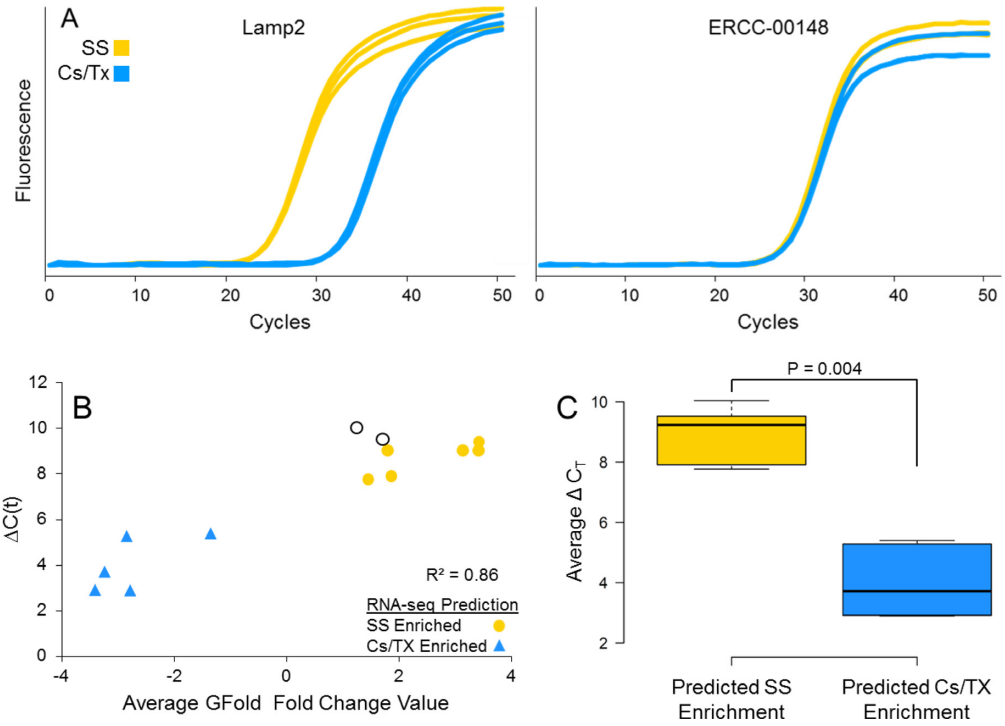

Figure 6. RT-PCR validation of peripheral sperm RNA enrichment. (**A**) To validate patterns of transcript enrichment predicted by RNA-seq RNAs from individual mice ($n = 6$) were extracted from equal numbers of sonicated or Cs/TX sperm heads and subjected to RT-PCR analysis. The transcript encoding lysosome-associated membrane glycoprotein 2 (Lamp2) is enriched in sonicated sperm indicating the preferential localization of this RNA within the outer sperm membranes. ERCC spike-in RNAs served as a loading control an exhibited no preferential enrichment. (**B**) Estimated RNA-seq fold-changes and delta-Ct values were well correlated ($r^2 = 0.86$). Transcripts exceeding the RPKM threshold (>116 RPKM; Supplementary Figure S1) and predicted by RNA-seq to be enriched in the extra- or intra-nuclear fractions are presented as solid circles or triangles, respectively. *Lamp2* and NADH dehydrogenase [ubiquinone] 1 subunit C1 (*Ndufc1*) partitioned to the extra-nuclear fraction but did not exceed the threshold (~85 RPKM in SS) and are presented as open circles. (**C**) Transcripts predicted by RNA-seq to be enriched in the SS samples ($n = 6$) exhibited significantly greater delta-C_t values ($P = 0.004$, Mann–Whitney U-test) than transcripts predicted to be enriched in the Cs/TX heads ($n = 5$). Center lines show the median delta-C_t values of the SS- and Cs/TX-predictions from RNA-seq (9.23 and 3.72, respectively); box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots.

sperm chromatin (60,61). Alternatively, other RNAs such as the repetitive RNAs may participate in packaging the paternal genome (Supplementary Tables SI, SII).

Repetitive RNAs in mouse sperm

Retained repetitive RNAs are of particular interest as they may contribute to nuclear structure as has been observed in somatic cells or may participate in genome confrontation and consolidation observed in plants and animals whereby parental genome compatibility is ensured (5,62). To better understand the role and abundance of these RNAs, the coverage of all individual repetitive elements was determined initially using uniquely aligned sequencing reads (Supplementary Tables SI, SII). This was complimented by re-aligning multiply mapped and unmapped sequencing reads to the RepBase repeat annotations. These alignments were combined with the uniquely aligned reads to determine the relative abundance of each repeat family within the sperm RNAs (Supplementary Dataset 6).

The Cs/TX libraries were enriched in repeat sequences such as satellite, simple and low complexity repeats (Supplementary Table SI). The presence of satellite RNAs may reflect opportunistic transcription as heterochromatic structures relax prior to protamine deposition. A series of com-

Table 1. Sperm RNAs are enriched in the extra-nuclear compartment

Transcript	ΔC(t)	SEM	Extra-nuclear enrichment*	GFold prediction
ERCC-00148	0.12	0.07	1.08	N/A
Novel RNA†	2.90	0.26	7.45	Intra-nuclear
Ankmy1	2.92	0.43	7.57	
Evi5l	3.72	0.39	13.19	
Tnp2	5.28	0.59	38.90	
Prm2	5.40	0.40	42.11	
Prdx1	7.77	0.68	218.02	Extra-nuclear
Rps4x	7.91	0.56	239.69	
rRNA	8.92	0.51	485.32	
Lamp2	9.03	0.55	524.37	
Ndufc1	9.04	0.55	527.92	
Cox6c	9.42	0.61	683.44	
Gng5	9.52	0.62	735.03	
Malat1	10.04	0.52	1,049.15	
Snora81	> 9.04	N/A	‡	

Paired comparisons, n = 6; standard error of the mean (SEM); $\Delta C(t) = (Ct_{Cs/TX}) - (Ct_{SS})$
*2^ΔC(t)
†Novel RNA = TCONS_00003257; Supplementary Dataset 2
‡Absence of Cs/TX amplification in some matched samples

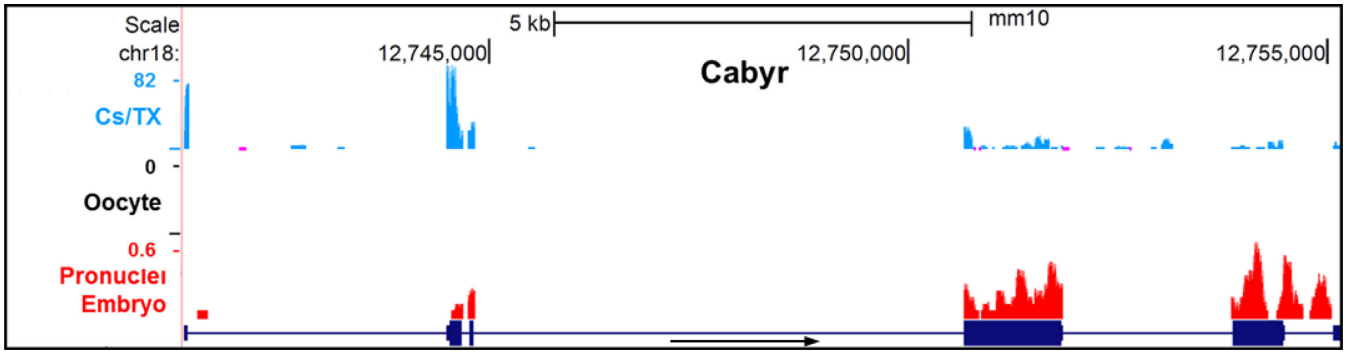

Figure 7. Detection of sperm dervied RNAs in fertilized oocytes prior to zygotic genome activation. *Cabyr* is observed in sperm (blue; Supplementary Figure S4) absent from oocytes and present in pronuclei embryos (black and red respectively; 37) following natural mating. Coverage for each sample is presented as reads per million uniquely aligned reads (RPM) on the Y-axis. The direction of transcription is depicted under the gene model.

plimentary sequences were consistently enriched in the Cs/TX sperm head RNA-seq libraries relative to sonicated sperm (Supplementary Dataset 6). The representation of these sequences within the Cs/TX samples was well above their genomic background resolving as a pattern of poly-purine or -pyrimidine RNA enrichment in these samples (Supplementary Table SII). Interestingly, GAA repeat-containing RNAs (GRC-RNAs) have been observed previously in the several somatic cell types wherein they contributed to chromatin structure by directly binding nuclear matrix proteins (63). These chromatin-associated RNAs form discrete DNase insensitive foci that are lost following RNase treatment. Confirmation that the nuclear-retained simple repeat transcripts identified in this study are poly-purine in nature would suggest that they may correspond to the population of nuclear matrix associated RNAs previously observed in sperm (11). Though assigning a similar nuclear function and even localization to the simple repeat sperm RNAs is an intriguing possibility, their lack of sequence complexity precludes RT-PCR validation and precise mapping.

Reminiscent of the results obtained for annotated and novel RNAs long interspersed nuclear elements (LINEs) exhibited greater RNA-seq coverage in the Cs/TX samples relative to sonicated sperm yet RT-PCR suggested that they were approximately eighty times more enriched in the SS samples (data not shown). Alignment to the LINE-1 (L1) repeat class on average contributed to more than 5% of all unique reads in the Cs/TX samples. Realignment to all canonical repeat sequences showed that ∼25% of all uniquely aligned reads corresponded to the L1 family in the Cs/TX samples (Supplementary Table SI). This was nearly twice the amount observed for the next most abundant family. To determine which of the L1 repeats was most actively transcribed during the final stages of spermatogenesis the RNA-seq libraries were realigned to canonical sequences for each active subfamily (42). In all samples the most abundant L1 transcripts corresponded to subfamily III of the L1MdTf lineage. The potential paternal contribution of L1 RNAs to the zygote, thought to perhaps spur autoregulated embryonic transcription of such elements, has been discussed (5,64). LINE transcripts have also recently been shown to directly contribute to somatic chromatin structure in an RNase sensitive manner (65). However, this association is not observed in the condensed somatic heterochromatin. Perhaps during the later stages of spermatogenesis the LINE transcripts are depleted either passively or actively from the sperm nucleus contributing to sperm chromatin condensation as is observed following RNase treatment in interphase cells (65,66). In the present study this would resolve as and be consistent with their observed extra-nuclear enrichment.

Identification of sperm-borne RNAs post-fertilization

The delivery of RNAs by human sperm to an oocyte has been shown (1,67). To ascertain whether mouse spermatozoa may potentially fulfill a similar function murine oocyte and pronuclei embryos sampled prior to zygotic genome activation were subjected to differential expression analysis (37). A total of 88 differentially enriched embryonic transcripts were identified which exhibited no coverage in the oocyte but were detected in all three sperm RNA-seq datasets in the current study. A previous mouse sperm RNA-seq study proposed that the *Wnt4* and *Foxg1* transcripts may be delivered by sperm to the oocyte (68). These RNAs were not detected in the embryonic datasets nor were they consistently present in sperm RNA-seq studies from others (Supplementary Figure S4B). Though the majority of the differentially enriched mouse embryo RNAs identified above were present at low levels in sperm (median RPKM ∼1.4) the transcript encoding the Calcium-binding tyrosine phosphorylation-regulated protein (*Cabyr*; Figure 7 and Supplementary Figure S4A) was consistently detected at elevated levels in both SS and Cs/TX samples and was observed in all available mouse sperm RNA-seq datasets (3,4,45). Interestingly, this transcript was enriched in the Cs/Tx heads by RNA-seq (Supplementary Dataset 2) suggesting a reduced peripheral enrichment similar to that observed for the set of predicted Cs/TX RNAs evaluated by RT-PCR (Figure 6 and Table). These transcripts likely persist within either the perinuclear theca or the nucleus and would therefore be expected to reach the oocyte cytoplasm following fertilization (69). It is not clear whether paternal *Cabyr* RNA would be of functional importance considering the established role of its protein product in modulating intracellular calcium levels during capacitation (70), although one could speculate a signaling function as part of

the ion flux upon fertilization. Irrespective, in human males, it is considered of diagnostic value for idiopathic male factor infertility (71). Nevertheless, detection of a testis-specific RNAs encoding a protein restricted to the post-meiotic male germline in naturally fertilized embryos prior to zygotic genome activation supports the view that sperm RNA persists throughout the normal lifecycle of the paternal gamete and are delivered to the oocyte.

The distribution of RNAs within the spermatozoon

The mature spermatozoon is host to a cadre of RNAs that are not evenly distributed throughout the limited volume of the sperm head. By mass ∼66% of the RNAs greater than ∼200 nt in length were enriched within the limited volume of the extra-nuclear compartment (plasma membrane, the acrosome and associated membranes as well as the sperm tail and mitochondrial sheath) of the sperm cell. Approximately one-third of the RNAs observed in sperm are expected to be within the nucleus/perinuclear theca. As determined by RT-PCR the strength of their preferential compartmentalization varied in accord with the RNA-seq results. Together these results provide a list of candidate extra- and intra-nuclear associated transcripts which serve as the foundation for future studies delineating their function within the mature sperm. Intriguingly, the set of RNAs exhibiting the strongest external localization included many transcripts linked to exosomes and transcripts commonly localized to specific organelles in somatic cells including the nucleus and mitochondria. Whether these transcripts are passively retained within the external sperm membranes following expulsion of the cytoplasm or are packaged within extracellular vesicles to serve a later function will be resolved when the contents of sperm-bound exosomes are fully described. The peripheral enrichment of non-coding RNAs such as *Malat1* and the LINE transcripts which are known to modulate somatic cell chromatin structure suggests that nuclear remodeling in the sperm is not limited to nucleoprotein exchange. The fraction of these RNAs that remain associated with the intra-nuclear compartment may be restricted to the few poised histone-bound accessible regions that persist following protamination. Representative of the population of transcripts is testis-specific RNA *Cabyr*. This RNA is likely delivered by sperm to the embryo and is a diagnostic element indicative of idiopathic male infertility (71).

ACCESSION NUMBER

Raw sequencing reads and processed data are deposited in the Gene Expression Omnibus repository (GEO Accession Number: GSE62874).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Charlotte B. Failing Professorship (to S.A.K.; in part). Funding for open access charge: Charlotte B. Failing Professorship (to S.A.K.; in part).

Conflict of interest statement. None declared.

REFERENCES

1. Ostermeier,G.C., Miller,D., Huntriss,J.D., Diamond,M.P. and Krawetz,S.A. (2004) Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature*, **429**, 154.
2. Sendler,E., Johnson,G.D., Mao,S., Goodrich,R.J., Diamond,M.P., Hauser,R. and Krawetz,S.A. (2013) Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res.*, **41**, 4104–4117.
3. Soumillon,M., Necsulea,A., Weier,M., Brawand,D., Zhang,X., Gu,H., Barthes,P., Kokkinaki,M., Nef,S., Gnirke,A. *et al.* (2013) Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.*, **3**, 2179–2190.
4. Hammoud,S.S., Low,D.H., Yi,C., Carrell,D.T., Guccione,E. and Cairns,B.R. (2014) Chromatin and transcription transitions of Mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell*, **15**, 239–253.
5. Jodar,M., Selvaraju,S., Sendler,E., Diamond,M.P., Krawetz,S.A. and Reproductive Medicine,N. (2013) The presence, role and clinical use of spermatozoal RNAs. *Hum. Reprod. Update*, **19**, 604–624.
6. Liu,W.M., Pang,R.T., Chiu,P.C., Wong,B.P., Lao,K., Lee,K.F. and Yeung,W.S. (2012) Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 490–494.
7. Gapp,K., Jawaid,A., Sarkies,P., Bohacek,J., Pelczar,P., Prados,J., Farinelli,L., Miska,E. and Mansuy,I.M. (2014) Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat. Neurosci.*, **17**, 667–669.
8. Kleene,K.C. (1989) Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. *Development*, **106**, 367–373.
9. Kwon,Y.K. and Hecht,N.B. (1993) Binding of a phosphoprotein to the 3′ untranslated region of the mouse protamine 2 mRNA temporally represses its translation. *Mol. Cell. Biol.*, **13**, 6547–6557.
10. Miller,D., Ostermeier,G.C. and Krawetz,S.A. (2005) The controversy, potential and roles of spermatozoal RNA. *Trends Mol. Med.*, **11**, 156–163.
11. Lalancette,C., Miller,D., Li,Y. and Krawetz,S.A. (2008) Paternal contributions: new functional insights for spermatozoal RNA. *J. Cell. Biochem.*, **104**, 1570–1579.
12. Sarrate,Z. and Anton,E. (2009) Fluorescence in situ hybridization (FISH) protocol in human sperm. *J. Vis. Exp.*, **31**, 1–2.
13. Modi,D., Shah,C., Sachdeva,G., Gadkar,S., Bhartiya,D. and Puri,C. (2005) Ontogeny and cellular localization of SRY transcripts in the human testes and its detection in spermatozoa. *Reproduction*, **130**, 603–613.
14. Johnson,G.D., Lalancette,C., Linnemann,A.K., Leduc,F., Boissonneault,G. and Krawetz,S.A. (2011) The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction*, **141**, 21–36.
15. Yan,W., Morozumi,K., Zhang,J., Ro,S., Park,C. and Yanagimachi,R. (2008) Birth of mice after intracytoplasmic injection of single purified sperm nuclei and detection of messenger RNAs and MicroRNAs in the sperm nuclei. *Biol. Reprod.*, **78**, 896–902.
16. Mao,S., Goodrich,R.J., Hauser,R., Schrader,S.M., Chen,Z. and Krawetz,S.A. (2013) Evaluation of the effectiveness of semen storage and sperm purification methods for spermatozoa transcript profiling. *Syst. Biol. Reprod. Med.*, **59**, 287–295.
17. Zaghlool,A., Ameur,A., Nyberg,L., Halvardson,J., Grabherr,M., Cavelier,L. and Feuk,L. (2013) Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues. *BMC Biotechnol.*, **13**, 99, 1–10.
18. Sullivan,R. and Saez,F. (2013) Epididymosomes, prostasomes, and liposomes: their roles in mammalian male reproductive physiology. *Reproduction*, **146**, R21–R35.
19. Cossetti,C., Lugini,L., Astrologo,L., Saggio,I., Fais,S. and Spadafora,C. (2014) Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: possible transport by exosomes. *PLoS One*, **9**, e101629.
20. Vojtech,L., Woo,S., Hughes,S., Levy,C., Ballweber,L., Sauteraud,R.P., Strobl,J., Westerberg,K., Gottardo,R., Tewari,M. *et al.* (2014) Exosomes in human semen carry a distinctive repertoire

of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Res.*, **42**, 7290–7304.
21. Ronquist,G., Nilsson,B.O. and Hjerten,S. (1990) Interaction between prostasomes and spermatozoa from human semen. *Arch. Androl.*, **24**, 147–157.
22. Arienti,G., Carlini,E. and Palmerini,C.A. (1997) Fusion of human sperm to prostasomes at acidic pH. *J. Membr. Biol.*, **155**, 89–94.
23. Frenette,G. and Sullivan,R. (2001) Prostasome-like particles are involved in the transfer of P25b from the bovine epididymal fluid to the sperm surface. *Mol. Reprod. Dev.*, **59**, 115–121.
24. Rejraji,H., Vernet,P. and Drevet,J.R. (2002) GPX5 is present in the mouse caput and cauda epididymidis lumen at three different locations. *Mol. Reprod. Dev.*, **63**, 96–103.
25. Ward,W.S. (2013) Isolation of sperm nuclei and nuclear matrices from the mouse, and other rodents. *Methods Mol. Biol.*, **927**, 437–444.
26. Goodrich,R., Johnson,G. and Krawetz,S.A. (2007) The preparation of human spermatozoal RNA for clinical analysis. *Arch. Androl.*, **53**, 161–167.
27. Goodrich,R.J., Anton,E. and Krawetz,S.A. (2013) Isolating mRNA and small noncoding RNAs from human sperm. *Methods Mol. Biol.*, **927**, 385–396.
28. Johnson,G.D., Platts,A.E., Lalancette,C., Goodrich,R., Heng,H.H. and Krawetz,S.A. (2011) Interrogating the transgenic genome: development of an interspecies tiling array. *Syst. Biol. Reprod. Med.*, **57**, 54–62.
29. Mao,S., Sendler,E., Goodrich,R.J., Hauser,R. and Krawetz,S.A. (2014) A comparison of sperm RNA-seq methods. *Syst. Biol. Reprod. Med.*, **60**, 308–315.
30. Kim,D., Pertea,G., Trapnell,C., Pimentel,H., Kelley,R. and Salzberg,S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36, 1–13.
31. Roberts,A., Pimentel,H., Trapnell,C. and Pachter,L. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*, **27**, 2325–2329.
32. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing,S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
33. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
34. Feng,J., Meyer,C.A., Wang,Q., Liu,J.S., Shirley Liu,X. and Zhang,Y. (2012) GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*, **28**, 2782–2788.
35. Anders,S., McCarthy,D.J., Chen,Y., Okoniewski,M., Smyth,G.K., Huber,W. and Robinson,M.D. (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Potoc.*, **8**, 1765–1786.
36. Anders,S., Pyl,P.T. and Huber,W. (2014) HTSeq— a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **2**, 1–4.
37. Xue,Z., Huang,K., Cai,C., Cai,L., Jiang,C.Y., Feng,Y., Liu,Z., Zeng,Q., Cheng,L., Sun,Y.E. *et al.* (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, **500**, 593–597.
38. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*,**15**, 550, 1–21.
39. Eden,E., Navon,R., Steinfeld,I., Lipson,D. and Yakhini,Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48, 1–7.
40. Pagliarini,D.J., Calvo,S.E., Chang,B., Sheth,S.A., Vafai,S.B., Ong,S.E., Walford,G.A., Sugiana,C., Boneh,A., Chen,W.K. *et al.* (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell*, **134**, 112–123.
41. Jurka,J., Kapitonov,V.V., Pavlicek,A., Klonowski,P., Kohany,O. and Walichiewicz,J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.*, **110**, 462–467.
42. Sookdeo,A., Hepp,C.M., McClure,M.A. and Boissinot,S. (2013) Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob. DNA*, **4**, 3, 1–15.
43. Ferrer,M., Xu,W. and Oko,R. (2012) The composition, protein genesis and significance of the inner acrosomal membrane of eutherian sperm. *Cell Tissue Res.*, **349**, 733–748.
44. Korley,R., Pouresmaeili,F. and Oko,R. (1997) Analysis of the protein composition of the mouse sperm perinuclear theca and characterization of its major protein constituent. *Biol. Reprod.*, **57**, 1426–1432.
45. Kobayashi,H., Sakurai,T., Imai,M., Takahashi,N., Fukuda,A., Yayoi,O., Sato,S., Nakabayashi,K., Hata,K., Sotomaru,Y. *et al.* (2012) Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.*, **8**, e1002440.
46. Cappallo-Obermann,H., Schulze,W., Jastrow,H., Baukloh,V. and Spiess,A.N. (2011) Highly purified spermatozoal RNA obtained by a novel method indicates an unusual 28S/18S rRNA ratio and suggests impaired ribosome assembly. *Mol. Hum. Reprod.*, **17**, 669–678.
47. Johnson,G.D., Sendler,E., Lalancette,C., Hauser,R., Diamond,M.P. and Krawetz,S.A. (2011) Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Mol. Hum. Reprod.*, **17**, 721–726.
48. Cooper,T.G., Yeung,C.H., Fetic,S., Sobhani,A. and Nieschlag,E. (2004) Cytoplasmic droplets are normal structures of human sperm but are not well preserved by routine procedures for assessing sperm morphology. *Hum. Reprod.*, **19**, 2283–2288.
49. de Mateo,S., Castillo,J., Estanyol,J.M., Ballesca,J.L. and Oliva,R. (2011) Proteomic characterization of the human sperm nucleus. *Proteomics*, **11**, 2714–2726.
50. Kellems,R.E., Allison,V.F. and Butow,R.A. (1974) Cytoplasmic type 80 S ribosomes associated with yeast mitochondria. II. Evidence for the association of cytoplasmic ribosomes with the outer mitochondrial membrane in situ. *J. Biol. Chem.*, **249**, 3297–3303.
51. Sylvestre,J., Margeot,A., Jacq,C., Dujardin,G. and Corral-Debrinski,M. (2003) The role of the 3′ untranslated region in mRNA sorting to the vicinity of mitochondria is conserved from yeast to human cells. *Mol. Biol. Cell*, **14**, 3848–3856.
52. Weis,B.L., Schleiff,E. and Zerges,W. (2013) Protein targeting to subcellular organelles via MRNA localization. *Biochim. Biophys. Acta*, **1833**, 260–273.
53. Buschow,S.I., van Balkom,B.W., Aalberts,M., Heck,A.J., Wauben,M. and Stoorvogel,W. (2010) MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis. *Immunol. Cell Biol.*, **88**, 851–856.
54. Gonzales,P.A., Pisitkun,T., Hoffert,J.D., Tchapyjnikov,D., Star,R.A., Kleta,R., Wang,N.S. and Knepper,M.A. (2009) Large-scale proteomics and phosphoproteomics of urinary exosomes. *J. Am. Soc. Nephrol.*, **20**, 363–379.
55. Valadi,H., Ekstrom,K., Bossios,A., Sjostrand,M., Lee,J.J. and Lotvall,J.O. (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.*, **9**, 654–659.
56. Chevillet,J.R., Kang,Q., Ruf,I.K., Briggs,H.A., Vojtech,L.N., Hughes,S.M., Cheng,H.H., Arroyo,J.D., Meredith,E.K., Gallichotte,E.N. *et al.* (2014) Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 14888–14893.
57. Brown,J.A., Valenstein,M.L., Yario,T.A., Tycowski,K.T. and Steitz,J.A. (2012) Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 19202–19207.
58. Arpanahi,A., Brinkworth,M., Iles,D., Krawetz,S.A., Paradowska,A., Platts,A.E., Saida,M., Steger,K., Tedder,P. and Miller,D. (2009) Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res.*, **19**, 1338–1349.
59. Hammoud,S.S., Nix,D.A., Zhang,H., Purwar,J., Carrell,D.T. and Cairns,B.R. (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature*, **460**, 473–478.
60. West,J.A., Davis,C.P., Sunwoo,H., Simon,M.D., Sadreyev,R.I., Wang,P.I., Tolstorukov,M.Y. and Kingston,R.E. (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell*, **55**, 791–802.
61. Carone,B.R., Hung,J.H., Hainer,S.J., Chou,M.T., Carone,D.M., Weng,Z., Fazzio,T.G. and Rando,O.J. (2014) High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev. Cell*, **30**, 11–22.

62. Bourc'his,D. and Voinnet,O. (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science*, **330**, 617–622.
63. Zheng,R., Shen,Z., Tripathi,V., Xuan,Z., Freier,S.M., Bennett,C.F., Prasanth,S.G. and Prasanth,K.V. (2010) Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. *J. Cell Sci.*, **123**, 3734–3744.
64. Fadloun,A., Le Gras,S., Jost,B., Ziegler-Birling,C., Takahashi,H., Gorab,E., Carninci,P. and Torres-Padilla,M.E. (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.*, **20**, 332–338.
65. Hall,L.L., Carone,D.M., Gomez,A.V., Kolpa,H.J., Byron,M., Mehta,N., Fackelmayer,F.O. and Lawrence,J.B. (2014) Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell*, **156**, 907–919.
66. Nickerson,J.A., Krochmalnic,G., Wan,K.M. and Penman,S. (1989) Chromatin architecture and nuclear RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 177–181.
67. Avendano,C., Franchi,A., Jones,E. and Oehninger,S. (2009) Pregnancy-specific {beta}-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: differential expression in fertile and infertile men and evidence of a possible functional role during early development. *Hum. Reprod.*, **24**, 270–277.
68. Fang,P., Zeng,P., Wang,Z., Liu,M., Xu,W., Dai,J., Zhao,X., Zhang,D., Liang,D., Chen,X. *et al.* (2014) Estimated diversity of messenger RNAs in each murine spermatozoa and their potential function during early zygotic development. *Biol. Reprod.*, **90**, 94, 1–11.
69. Sutovsky,P., Oko,R., Hewitson,L. and Schatten,G. (1997) The removal of the sperm perinuclear theca and its association with the bovine oocyte surface during fertilization. *Dev. Biol.*, **188**, 75–84.
70. Naaby-Hansen,S., Mandal,A., Wolkowicz,M.J., Sen,B., Westbrook,V.A., Shetty,J., Coonrod,S.A., Klotz,K.L., Kim,Y.H., Bush,L.A. *et al.* (2002) CABYR, a novel calcium-binding tyrosine phosphorylation-regulated fibrous sheath protein involved in capacitation. *Dev. Biol.*, **242**, 236–254.
71. Garrido,N., Martinez-Conejero,J.A., Jauregui,J., Horcajadas,J.A., Simon,C., Remohi,J. and Meseguer,M. (2009) Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. *Fertil. Steril.*, **91**, 1307–1310.