Cleavage of rRNA ensures translational cessation in sperm at fertilization

G.D. Johnson¹, E. Sendler¹,², C. Lalancette¹,², R. Hauser³,⁴,⁵, M.P. Diamond², and S.A. Krawetz¹,²,*

¹Center for Molecular Medicine and Genetics, C. S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI 48201, USA ²Department of Obstetrics and Gynecology, C. S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI 48201, USA ³Vincent Memorial Obstetrics and Gynecology Service, Massachusetts General Hospital, Boston, MA, USA ⁴Department of Environmental Health, Harvard Medical School, Boston, MA, USA ⁵Department of Epidemiology, Harvard Medical School, Boston, MA 02115, USA

*Correspondence address. Tel: +1-313-577-6770; Fax: +1-313-577-8554; E-mail: steve@compbio.med.wayne.edu

Submitted on July 8, 2011; resubmitted on July 27, 2011; accepted on August 3, 2011

ABSTRACT: Intact ribosomal RNAs (rRNAs) comprise the majority of somatic transcripts, yet appear conspicuously absent in spermatozoa, perhaps reflecting cytoplasmic expulsion during spermatogenesis. To discern their fate, total RNA retained in mature spermatozoa from three fertile donors was characterized by Next Generation Sequencing. In all samples, >75% of total sequence reads aligned to rRNAs. The distribution of reads along the length of these transcripts exhibited a high degree of non-uniformity that was reiterated between donors. The coverage of sequencing reads was inversely correlated with guanine-cytosine (GC)-richness such that sequences greater than ~70% GC were virtually absent in all sperm RNA samples. To confirm the loss of sequence, the relative abundance of specific regions of the 28S transcripts in sperm was established by 7-Deaza-2′-deoxy-guanosine-5′-triphosphate RT–PCR. The inability to amplify specific regions of the 28S sequence from sperm despite the abundant representation of this transcript in the sequencing libraries demonstrates that approximately three-quarters of RNA retained in the mature male gamete are products of rRNA fragmentation. Hence, cleavage (not expulsion of the RNA component of the translational machinery) is responsible for preventing spurious translation following spermiogenesis. These results highlight the potential importance of those transcripts, including many mRNAs, which evade fragmentation and remain intact when sperm are delivered at fertilization.

Sequencing data are deposited in GEO as: GSE29160.

Key words: next generation sequencing / rRNA / sperm / spermiogenesis

Introduction

Initial investigations of sperm-retained RNAs largely focused on mRNAs, partly due to several lines of evidence suggesting they were the dominant class of transcript in the cell (Kumar et al., 1993; Chiang et al., 1994; Wykes et al., 1997; Ostermeier et al., 2002, 2004). The ability of sperm to store and protect spermatogenic messenger RNAs (mRNA) provided the rationale for the presence of any RNA despite transcriptional and translational quiescence (Kierszenbaum and Tres, 1975; Grunewald et al., 2005; reviewed in Krawetz, 2005). Subsequent studies have since attempted to identify mRNAs common to sperm from fertile men, which may be of clinical importance (Wang et al., 2004; Platts et al., 2007; Lalancette et al., 2009; Linschooten et al., 2009). In contrast, little is known about the fate of ribosomal RNAs (rRNAs) during and after spermiogenesis. These transcripts are the most abundant in all cells yet until now, assumed to be absent from the mature male gamete (reviewed in Krawetz, 2005). Instead of exhibiting defined 18S and 28S rRNA peaks, electrophoretic analysis of sperm RNAs reveals an enrichment of short-length transcripts (Ostermeier et al., 2002; Gilbert et al., 2007; Das et al., 2010). Although intact mRNAs are detected in this pool of transcripts, full-length 28S and 18S rRNAs are not (Ostermeier et al., 2005). Failure to observe rRNAs in sperm has been presumed to be due to the large reduction in cytoplasmic volume accompanying morphogenesis. Indeed, expulsion of the translational machinery during sperm maturation is supported by detection of RNA and ribosomal proteins in the residual body/cytoplasmic droplet (Breucker et al., 1985; Cappallo-Obermann et al., 2011). Together, these and other historical observations led to the assumption that spermatozoa lack rRNAs (Miller et al., 2005). However, the fate of these transcripts has recently begun to be reevaluated, suggesting the presence of rRNA components (Cappallo-Obermann et al., 2011).

Interrogation of sperm RNAs by Next Generation Sequencing has permitted an in-depth analysis, until now, could not be attained. This technology has recently been used to show that the male
gamete harbors many RNA species in addition to mRNAs (Johnson et al., 2011). Herein, we describe the composition of rRNAs in sperm. Sequencing of sperm total RNA suggests that though not intact, these RNAs may be present at levels approaching that observed in somatic cells. The depletion of specific subregions of the 28S transcript supports the conclusion that rRNA cleavage, not expulsion during cellular condensation, is the primary means of preventing spurious translation of mRNAs stored in mature spermatozoa and delivered at fertilization.

Materials and Methods

RNA isolation and Next Generation Sequencing

Spermatozoal RNAs were isolated from ejaculates from three fertile donors, as described previously (Goodrich et al., 2007). Briefly, following the removal of contaminating somatic and residual cells by centrifugation through a 50% percoll gradient to obtain pure populations of spermatozoa, RNA was extracted from purified sperm using the RNEasy system (Qiagen, Valencia, CA, USA). Total testis RNAs were obtained from two commercial libraries (Applied Biosystems/Ambion, Austin, TX, USA, Lot# 054P01702031A; ClonTech, Mountain View, CA, USA, Lot# 3090051). RNA size and quality were assayed using the 2100 bioanalyzer with a RNA Pico chip (Agilent Technologies, Palo Alto, CA, USA). Separate RNA-seq libraries were constructed for each sample using the Illumina mRNA-Seq kit (Illumina, San Diego, CA, USA) and HotStarTaq (Qiagen) systems, essentially as described by the manufacturer. Following random hexamer priming, second-strand synthesis was performed at 55°C for 1 h. Each 60 μl PCR contained 6 μl of the reverse transcription reaction, 0.2 mM dNTPs, 0.6 units enzyme and 0.15 mM 7-Deaza-2′-deoxy-guanosine-5′-triphosphate (GTP) (Roche, Indianapolis, IN, USA) plus 0.05 mM dGTP in all reactions. Reaction aliquots were removed every five cycles and visualized with ethidium bromide. Quenching of intercalative dyes by the nucleotide analog necessitated gel densitometric analysis of product enrichment in place of quantitative real-time PCR (Pivonkova et al., 2010). Gel images were captured on the Typhoon 9210 (Molecular Dynamics) and were well within the optical linear range of response and thus analyzed with Quantity One version 4.5.2 (BioRad).

Results

Despite robust transcription and translation during early spermiogenesis, it has been assumed that mature spermatozoa are depleted in rRNAs. This notion was supported by the transcriptionally silent state of the mature sperm, the expulsion of the majority of the cytoplasm as a droplet during maturation and the virtual complete absence of full-length 18S or 28S rRNAs following condensation (Fig. 1). As a hallmark of the mature male gamete, the lack of intact 18S and 28S rRNAs in sperm has been consistently observed in several species and is independent of extraction procedures (Ostermeier et al., 2002; Gilbert et al., 2007; Yang et al., 2009; Das et al., 2010). Whether depletion of these sequences is achieved by RNA cleavage and/or physical expulsion with bulk cytoplasm is not known. To discern the fate of these transcripts following spermiogenesis, total RNA retained in mature spermatozoa from three fertile donors was characterized by Next Generation Sequencing. The sperm RNA-seq characters were performed using the Firecrest and Bustard modules of the genome analyzer pipeline software (Illumina Pipeline software v. 1.3.0). Sequencing reads were aligned to 18S (NR_003286.2, 1869 nt length) and 28S (NR_003287.2, 5070 nt length) rRNA sequences. Alignments were performed with Novoalign (Novocraft Technologies SdnBhd, v. 2.05.43) using default parameters to align individual reads from a paired-end set separately. Alignment results were confirmed independently using the GERALD (Illumina). Only sequences uniquely aligning to the genome were considered since the addition of sequences that aligned to multiple regions had little effect. The number of sequence reads in each library was standardized as a function of the total number of reads obtained for that sample. To locally reduce non-uniformity, coverage at each position is shown as a sum of reads starting over the preceding 35 bases (Hansen et al., 2010).

RT–PCR analysis

Total testis and sperm RNAs (20 ng) were reverse-transcribed and amplified using the SuperScriptII First-Strand Synthesis (Invitrogen, Carlsbad, CA, USA) and HotStarTaq (Qiagen) systems, essentially as described by the manufacturer. Following random hexamer priming, second-strand synthesis was performed at 55°C for 1 h. Each 60 μl PCR contained 6 μl of the reverse transcription reaction, 0.2 mM dNTPs, 0.6 units enzyme and 1 μl primer. Primer sequences are provided in Supplementary data, Table 1. Cycling conditions were 94°C for 2 min, followed by 45 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 40 s. Use of dimethylsulfoxide (DMSO) and/or betaine in all reactions was not pursued as these additives failed to rescue amplification of guanine-cytosine (GC)-rich regions. However, these sequences could be amplified in the presence of 0.15 mM 7-Deaza-2′-deoxy-guanosine-5′-triphosphate (GTP) (Roche, Indianapolis, IN, USA) plus 0.05 mM dGTP in all reactions. Reaction

![Figure 1](https://academic.oup.com/molehr/article-abstract/17/12/721/978031/722)
libraries produced an average of 58.2 million sequence reads, of which ~80% aligned to portions of the 18S and 28S rRNAs (Table I). This establishes the 18S and 28S rRNAs as the most abundant spermatozoal transcripts. However, as indicated earlier, the electrophoretic analysis of sperm total RNAs demonstrated that though highly abundant, these rRNAs are not intact (Fig. 1). The 5S and 5.8S rRNAs corresponded to <0.1% of the total number of sequence reads and were not considered further. Similarly, the mitochondrial 16S and 12S rRNAs contributed to only 1.8 and 0.8% of total sequencing reads, respectively and were not considered further. This establishes the 18S and 28S rRNAs as representing the most abundant spermatozoal transcripts. As shown in Fig. 1, these abundant rRNAs do not electro- phoretically resolve as discrete peaks. This is consonant with the recent proposal that high-quality sperm samples can be identified by a substantially lower ratio of 28S:18S rRNAs (0.11) than observed in somatic cells (Cappallo-Obermann et al., 2011). However, the total number of sequencing reads that mapped to either the 28S or 18S transcripts was proportional to their respective lengths and thus in the expected ratio of 2:1. As shown below, this apparent inconsistency reflects the fragmentation of the majority of the sperm rRNAs. This highlights the significance of those mRNAs and small non-coding RNAs that are consistently found intact.

The distribution of sequencing reads across these transcripts revealed a consistent series of peaks and valleys reiterated between donors (Fig. 2A; Supplementary data, Fig. S1). Notably, several subregions of the 28S RNA were extremely read-poor suggesting their depletion in sperm. To determine whether the absence of select rRNA sequences was a previously undescribed feature of the mature male gamete, sperm and testes poly[A+] enriched RNAs were compared by RNA-seq. As expected, poly[A+] selection diminished the representation of the 18S and 28S transcripts to <3% of that observed in the total RNAs. However, the distribution of reads across the transcripts was essentially identical to that observed in sperm prior to enrichment and strikingly concordant with that observed in the testes libraries after selection (Fig. 2B; Supplementary data, Figs S2–4, Table I).

Paradoxically, despite the presence of intact rRNAs in testis and their absence from sperm (Fig. 1), specific regions of the 18S and 28S transcripts were underrepresented. Approximately half of 28S rRNA is ≥70% GC (36 nt window and 1 nt step). Sequences exceeding this threshold were markedly reduced or absent from both the testis and sperm libraries (Fig. 2B, orange line). This trend was also observed along the 18S transcript (~10% ≥ 70% GC; Supplementary data, Fig. S4). The ~340 nt 28S RNA region bounded by nucleotide coordinates 3000–3500 underscores the relationship between GC richness and suppressed representation in the sequencing libraries. Thus, the simple comparison of 18S and 28S sequencing coverage in sperm and testis RNA-seq libraries was not sufficient to reveal potential sites of cleavage. It has been suggested that the under-representation of extremely GC-rich genomic sequences may reflect PCR amplification bias. In these cases, loss of coverage was circumvented by either altering cycling conditions or forgoing enrichment (Kozarewa et al., 2009; Aird et al., 2011; Kozarewa and Turner, 2011). However, adapting this approach to the rRNAs did not normalize coverage across the 18S and 28S rRNAs (Mamanova et al., 2010). The persistent absence of GC-rich subregions within these transcripts suggests that lack of sequencing coverage may reflect the tendency of these highly structured RNAs to intramolecular base-pair (Wellauer and Dawid, 1973; Wakeman and Maden, 1989). Stably folded rRNAs can occlude primers and reverse transcriptase, leading to their exclusion from further analysis, thus marginalizing their representation during sequencing (Zhang et al., 2001; Mortazavi et al., 2008). Although this concern is generally mitigated by fragmentation preceding cDNA synthesis, it is evident that some transcripts cannot be resolved using current sequencing protocols (Mortazavi et al., 2008).

Juxtaposing the absence of intact 18S and 28S rRNA in the mature male gamete with their abundance in the sperm sequencing data suggested that those transcripts are fragmented during spermiogenesis. To localize potential sites of cleavage, primers were designed to five specific subregions along the 28S rRNA transcript (Fig. 2B, blue horizontal bars; Supplementary data, Table S1) and relative levels of each specific segment was compared by RT–PCR. The intact testis template served as a baseline to evaluate the relative depletion of these regions in sperm. Several products were readily detected at varying levels in both testis and sperm (Region 1, 3–733 nt; Region 3, 1301–2108 nt and Region 5, 4441–4548 nt; Fig. 2C, deaza−). As expected, sequences with a GC content exceeding 70% were not amplified in either testis or sperm (Region 2, 716–1319 nt; Region 4, 2871–3699 nt). These directly overlap the highly

---

**Table I** Percentage of Next Generation Sequencing reads of sperm and testes RNAs mapping to 18S and 28S rRNA.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Total reads</th>
<th>Reads aligned to 18S rRNA</th>
<th>% Aligned to 18S rRNA</th>
<th>Reads aligned to 28S rRNA</th>
<th>% Aligned to 28S rRNA</th>
<th>% rRNA of total RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor 1</td>
<td>52.15</td>
<td>11.49</td>
<td>22.03</td>
<td>31.37</td>
<td>60.15</td>
<td>82.2</td>
</tr>
<tr>
<td>Donor 2</td>
<td>64.45</td>
<td>14.85</td>
<td>23.1</td>
<td>41.66</td>
<td>64.75</td>
<td>87.8</td>
</tr>
<tr>
<td>Donor 3</td>
<td>58.11</td>
<td>13.60</td>
<td>23.4</td>
<td>30.98</td>
<td>53.3</td>
<td>76.7</td>
</tr>
<tr>
<td>Donor 1b</td>
<td>22.99</td>
<td>0.24</td>
<td>1.03</td>
<td>0.86</td>
<td>3.74</td>
<td>4.8</td>
</tr>
<tr>
<td>Testis CloneTech</td>
<td>20.96</td>
<td>0.21</td>
<td>0.99</td>
<td>0.67</td>
<td>3.2</td>
<td>4.17</td>
</tr>
<tr>
<td>Testis Ambionb</td>
<td>21.94</td>
<td>0.089</td>
<td>0.40</td>
<td>0.23</td>
<td>1.03</td>
<td>1.44</td>
</tr>
</tbody>
</table>

Values denote the number of high-quality sequencing reads in millions.

Poly(A+) enriched samples (shaded).
folded regions of the 28S rRNA and represent the largest gaps in sequencing coverage along this transcript (Fig. 2B; Wellauer and Dawid, 1973; Wakeman and Maden, 1989). Initial efforts to minimize RNA secondary structure, including increasing reaction temperatures, the addition of DMSO and/or betaine, did not rescue amplification of these regions (data not shown). However, partial substitution of 7-deaza-GTP for dGTP in all reactions resulted in robust amplification of all sequences from testis RNAs (Fig. 2C, deaza$^+$). Regions 1 and 3 were more abundant in testis than in sperm, while Region 5 exhibited comparable levels in both cell-types. Even after 45 cycles of amplification, Regions 2 and 4 could not be detected in sperm (Fig. 2C, deaza$^-$). In contrast, amplification of these products from the testis template proceeded to saturation and was easily detected after cycle 30. These results establish the absence of specific regions of the 28S rRNA in sperm in accordance with the view that they harbor preferential sites of cleavage. Further, the requirement of 7-deaza-GTP to resolve these sequences by RT–PCR shows that highly base-paired GC-rich regions are underrepresented and thus excluded from cDNA synthesis. This is consistent with the use of denaturants like CH$_3$HgOH to ensure the complete reverse transcription and synthesis of cDNAs from GC-rich RNAs. (Krawetz et al., 1986).

**Discussion**

Complimenting Next Generation Sequencing analysis of sperm RNAs with RT–PCR interrogation of select regions of the 28S rRNA demonstrated that these transcripts are highly abundant though fragmented in the mature male gamete. Percoll selection prior to RNA isolation
ensured that sequencing reads were representative of RNAs from within purified sperm and not somatic contaminants nor transcripts within the residual body. The presence of intact full-length RNAs in sperm and the absence of specific subregions of the 28S transcript support the view that failure to resolve full-length rRNAs from the male gamete is not due to misdirected nuclelease activity. Rather, selective cleavage of the rRNAs likely ensures translational cessation, which clearly cannot be achieved by cytoplasmic expulsion alone.

The regions of the 28S rRNA that are preferentially cleaved in sperm have been implicated in rRNA fragmentation in other systems. Region 2 highlighted in Fig. 2 is a likely target for initial nuclease attack, since it is positioned on the surface of intact ribosomes (Han et al., 1994; Holmberg et al., 1994) and is expected to contain an exposed loop that can be preferentially cleaved prior to association with the ribosomal proteins (Leffers and Andersen, 1993; Houge et al., 1995). In comparison, Region 2, the other sequence absent from sperm, appears to be positioned internally within the ribosome (Han et al., 1994; Holmberg et al., 1994). In this case, the initial cleavage of external sequences may induce a conformational change, thereby exposing internal regions to nuclease digestion or rendering it amenable to autocatalysis (Houge et al., 1995). Regardless, the quantity of cleavage products remaining in sperm indicates that fragmentation of these transcripts does precede their complete degradation. Further, evidence of post-transcriptional polyadenylation that is proposed to target rRNA endonucleolytic cleavage products for degradation in human somatic cells was not apparent in the sequenced sperm rRNAs (Slomovic et al., 2006). Complete ablation of these transcripts should not be necessary in the male gamete as long as translational cessation is assured.

The abundance of rRNA remnants in mature sperm indicates that expulsion through the cytoplasmic droplet is not complete. Accordingly, this mechanism is not sufficient to prevent spurious protein synthesis that is only assured by fragmentation. This is of particular significance, considering that several spermatozoal mRNAs and small non-coding RNAs evade cleavage and are delivered by sperm to the oocyte. On the one hand, some, like the protamines, typify those necessary for repackageing the sperm nucleus and must remain silent until the sperm is capacitated (Avendano et al., 2009; Johnson et al., 2011) of the zygote. This would be assured by the destruction of spermatid ribosomes enabling the clearance of detrimental transcripts following fertilization. On the other hand, paternal RNAs required for early development may be protected from such pathways prior to engaging the maternal translational machinery. Perhaps, this provides a means to ensure compatibility of the gametes, as part of the consolidation confrontation pathway (Bourchis and Voïnet, 2010).

**Supplementary data**

Supplementary data are available at http://molehr.oxfordjournals.org/.

**Authors’ roles**

G.D.J. performed the experiments and C.L. created the sperm RNA sequencing library. E.S., G.D.J. and S.A.K. designed the experiments, analyzed the data and wrote the paper. M.D. and R.H. provided the samples and contributed to the discussion of the data and editing of the manuscript.

**Funding**

This work was supported, in part, by a pilot grant to SAK and RH from Harvard School of Public Health; National Institutes of Health (grant number ES017285); National Institute of Environmental Health Sciences (grant number ES009718); the Presidential Research Enhancement Program in Computational Biology and the Charlotte B. Failing Professorship (S.A.K.).

**References**


Han H, Scherpert A, Pellegrini M, Dervan PB. Mapping RNA regions in eukaryotic ribosomes that are accessible to methidiumpropyl-EDTA.Fe(II) and EDTA.Fe(II). Biochemistry 1994;33:9831–9844.


Downloaded from https://academic.oup.com/molehr/article-abstract/17/12/72/39781 by guest on 16 February 2019


Zhang YJ, Pan HY, Gao SJ. Reverse transcription slippage over the mRNA secondary structure of the LIP1 gene. Biotechniques 2001;31:1286, 8, 90, passim.