

# The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes

LORENA PANTANO,<sup>1</sup> MERITXELL JODAR,<sup>2</sup> MADS BAK,<sup>3,4</sup> JOSEP LLUÍS BALLESÇÀ,<sup>5</sup> NIELS TOMMERUP,<sup>3,4</sup> RAFAEL OLIVA,<sup>2</sup> and TANYA VAVOURI<sup>1,6</sup>

<sup>1</sup>Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Can Ruti Campus, Badalona, Barcelona 08916, Spain

<sup>2</sup>Genetics Unit, Department of Physiological Sciences, University of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Biochemistry and Molecular Genetics Service, Hospital Clinic, 08036 Barcelona, Spain

<sup>3</sup>Center for non-coding RNA in Technology and Health (RTH), University of Copenhagen, DK-2200 Copenhagen, Denmark

<sup>4</sup>Wilhelm Johannsen Centre for Functional Genome Research, Department of Cellular and Molecular Medicine, Faculty of Health Science, University of Copenhagen, DK-2200 Copenhagen, Denmark

<sup>5</sup>Andrology Unit, Institut Clínic de Ginecologia, Obstetricia i Neonatologia, Hospital Clínic, 08036 Barcelona, Spain

<sup>6</sup>Josep Carreras Leukaemia Research Institute (IJC), ICO-Hospital GermansTrias i Pujol, Badalona, Barcelona 08916, Spain

## ABSTRACT

At the end of mammalian sperm development, sperm cells expel most of their cytoplasm and dispose of the majority of their RNA. Yet, hundreds of RNA molecules remain in mature sperm. The biological significance of the vast majority of these molecules is unclear. To better understand the processes that generate sperm small RNAs and what roles they may have, we sequenced and characterized the small RNA content of sperm samples from two human fertile individuals. We detected 182 microRNAs, some of which are highly abundant. The most abundant microRNA in sperm is *miR-1246* with predicted targets among sperm-specific genes. The most abundant class of small noncoding RNAs in sperm are PIWI-interacting RNAs (piRNAs). Surprisingly, we found that human sperm cells contain piRNAs processed from pseudogenes. Clusters of piRNAs from human testes contain pseudogenes transcribed in the antisense strand and processed into small RNAs. Several human protein-coding genes contain antisense predicted targets of pseudogene-derived piRNAs in the male germline and these piRNAs are still found in mature sperm. Our study provides the most extensive data set and annotation of human sperm small RNAs to date and is a resource for further functional studies on the roles of sperm small RNAs. In addition, we propose that some of the pseudogene-derived human piRNAs may regulate expression of their parent gene in the male germline.

**Keywords:** sperm; RNA; piRNA; pseudogene; PIWI; microRNA

## INTRODUCTION

Human mature sperm cells contain RNA (Pessot et al. 1989; Kumar et al. 1993; Rohwedder et al. 1996; Wykes et al. 1997; Martins and Krawetz 2005; Jodar et al. 2013). These RNA molecules are thought to be remnants of transcription taking place during sperm development that either have not yet reached complete degradation or have been protected from it (Ostermeier et al. 2002, 2004; Jodar et al. 2012; Sendler et al. 2013). Sperm RNAs are found in the compacted nucleus of the sperm head and in the perinuclear cytoplasmic droplet, since the rest of the cytoplasm is expelled at the end of sperm maturation (for review, see Cooper 2005; Martins and Krawetz 2005).

Two characteristic features of sperm RNA are that it consists primarily of a population of short molecules and that it contains no intact ribosomal RNA (Ostermeier et al. 2002; Johnson et al. 2011). Human sperm samples from three healthy donors were recently analyzed by high-throughput sequencing (Krawetz et al. 2011). Most of the sperm cell transcriptome was found to consist of fragments of coding and noncoding RNAs (Johnson et al. 2011; Krawetz et al. 2011; Sendler et al. 2013). A recent survey of the sperm transcriptome showed that in addition to mRNA fragments sperm also contain several small noncoding RNAs, such as microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs)

Corresponding author: [tvavouri@imppc.org](mailto:tvavouri@imppc.org)

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.046482.114>.

© 2015 Pantano et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

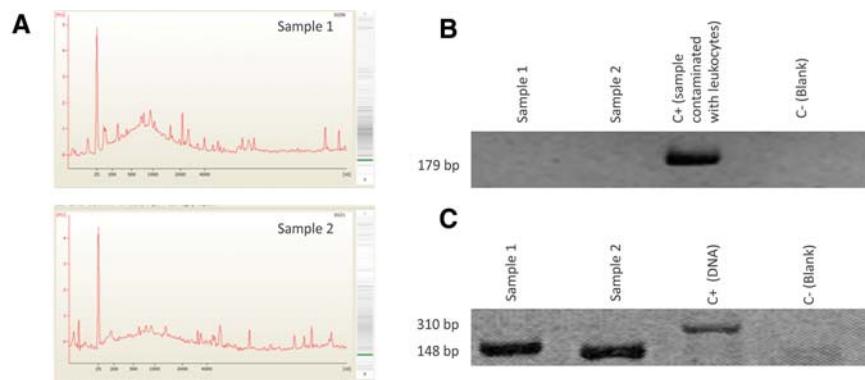
(Krawetz et al. 2011). However, a considerable portion of sperm small RNAs did not match miRNAs or piRNAs present in public databases and were therefore excluded from further analysis. We considered that understanding the potential functions and significance of sperm small RNA will benefit from a more in-depth analysis of the total small RNA content.

In this study, we have generated an extensive catalog of small RNAs present in two human sperm samples. We have analyzed the properties of these small RNAs and used them to distinguish potentially functional RNAs from degradation intermediates. We identified nearly 200 miRNAs present in sperm from two different fertile donors. The most abundant sperm miRNA is *miR-1246*, which evolved in the primate lineage and is predicted to target several sperm-specific transcripts. *miR-1246* has not previously been found in sperm. However, piRNAs are a more abundant class of regulatory RNA in sperm. By carefully analyzing the composition of human piRNA clusters using our own as well as previously published data sets we found that sperm cells contain piRNAs processed from the antisense strand of pseudogenes located inside piRNA clusters. Pseudogene-derived piRNAs are present in male germ cells in adult human testes. Parent protein-coding genes contain multiple predicted targets of pseudogene-derived piRNAs. We hypothesize that pseudogene-derived piRNAs may therefore regulate expression of their parent protein-coding genes in the male germline in the same way that typical piRNAs regulate transposons.

## RESULTS

### Extraction and sequencing of small RNAs from human sperm

We obtained sperm samples from two healthy donors with proven fertility, isolated sperm cells, and extracted their



**FIGURE 1.** Quality controls on sperm RNA samples. (A) Agilent bioanalyzer profiles of total RNA of the two samples sequenced. (B) Electrophoretic analysis of the CD45 RT-PCR products to verify the absence of leukocytes. The C+ lane (positive control) corresponds to a sperm sample contaminated with leukocytes. (C) Electrophoretic analysis of the PRM2 RT-PCR products using exon-spanning primers to check the absence of carried over DNA and the RNA integrity in sperm samples. The C+ lane (positive control) correspond to a DNA sample.

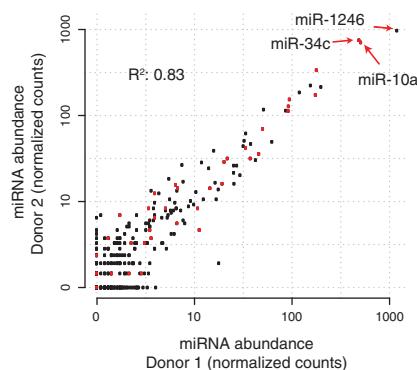
RNA (see Materials and Methods). The overall profile of extracted sperm RNA confirms the absence of intact ribosomal RNA and predominantly short-length RNA molecules, both well-known characteristics of sperm RNA (Fig. 1A). This profile also precludes RNA contamination from somatic cells. We additionally confirmed that there was no contamination from leukocytes by RT-PCR against the leukocyte-specific marker CD45 (Fig. 1B), whereas exon-spanning primers against the protamine 2 mature transcript amplified a fragment of the correct length (Fig. 1C). Molecules <50 nt were isolated from this RNA by gel-excision and were sequenced on the Genome Analyzer IIx platform. Following read quality filtering and adapter removal, we first searched for matches to miRNAs (see Materials and Methods).

### An extensive catalog of human sperm microRNAs

We first annotated the miRNA fraction of sperm small RNAs, as they are the most easily identifiable and most studied class of small RNAs. In total, we detected 314 known mature miRNAs (Supplemental Table S1) with 182 present in both samples. The correlation of abundance of these miRNAs in sperm samples from two different and unrelated donors is high ( $R^2 = 0.83$ ) (Fig. 2). Of the 182 common miRNAs, only 37 were previously reported in a recent survey (Krawetz et al. 2011) of miRNAs in human sperm. The greater coverage of sperm miRNAs in our data set compared with the previous study is most likely due to the fact that (unlike the previous study, Krawetz et al. 2011) we used an RNA isolation protocol specific for small RNAs and we sequenced to higher depth (see Materials and Methods).

Among the most abundant miRNAs in mature sperm samples is *miR-34c*. In mouse, *miR-34c* is highly expressed in the late stages of spermiogenesis and is also found in mature sperm (Bouhallier et al. 2010). Moreover, mouse *miR-34c* is transferred via sperm to the zygote where it has been reported to regulate the first cell division (Liu et al. 2012).

The second most abundant miRNA in our human sperm samples is *miR-10a*. The gene encoding this miRNA is highly conserved in animals and is found in syntenic locations inside HOX clusters in flies and human (Lund 2010). The *miR-10a* locus is packaged in nucleosomes in mature sperm (Hammoud et al. 2009). *miR-10a* is one of the few microRNAs proposed to activate rather than repress translation and in particular has been reported to enhance translation of ribosomal protein mRNAs (Orom et al. 2008). At the end of spermiogenesis and in the very early embryo, there is very little transcription and regulation happens at the post-transcriptional level.



**FIGURE 2.** Sperm miRNAs have similar abundance in small RNA samples from two different human fertile individuals. MicroRNAs in red correspond to those detected in a previous study (Krawetz et al. 2011).

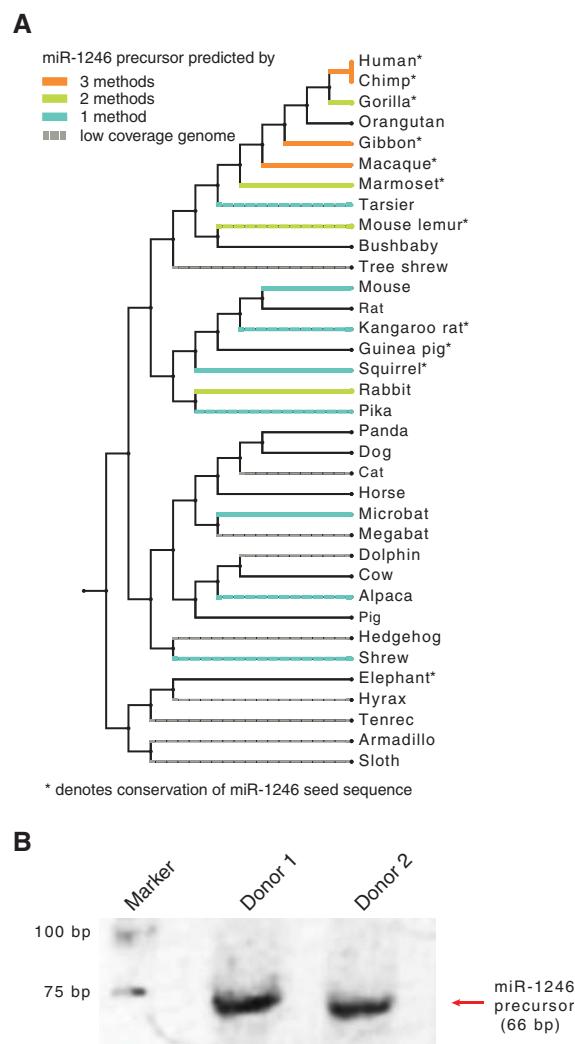
*miR-10a* may therefore be a potential novel regulator of translation in one, or both, of these developmental stages. It should be noted, however, that the quantity of microRNAs delivered to the oocyte via sperm is small and, therefore, if this microRNA has any role in the zygote, it would have to act on a small number of targets and soon after fertilization.

#### *miR-1246* is a primate-specific miRNA and the most abundant miRNA in human sperm

The most abundant miRNA in human sperm is *miR-1246* (Fig. 2). Little is currently known about the function and evolution of *miR-1246*. It was first discovered in human embryonic stem cells (Morin et al. 2008). Using miRNA gene prediction tools and comparative genomics, we found that *miR-1246* likely evolved in the primate lineage (Fig. 3A; Supplemental Table S2). Three different miRNA gene prediction methods predict the gene with a conserved seed region in human, chimp, gibbon, and macaque. The miRNA gene with a conserved seed region is also predicted in three additional primates using two of three methods. One of the three computational methods also predicts an miRNA in eight additional mammals, however, most of these predicted genes do not have the conserved seed (Supplemental Table S2). The mature *miR-1246* sequence is also found in U2 small nuclear RNAs (snRNA) and therefore fragments of U2 snRNAs could be the source of these sequences. We therefore tested whether we could detect the precursor of *miR-1246* in mature sperm. As shown in Figure 3B, we confirmed the presence of the *miR-1246* precursor in sperm by RT-PCR and sequencing. We conclude that the most abundant miRNA in sperm, *miR-1246*, evolved in primates from a previously existing sequence with miRNA-like secondary structure.

Little is known about the targets of *miR-1246*. The only experimentally confirmed target of *miR-1246* is the Down syndrome-associated protein kinase 1A (*DYRK1A*) (Zhang et al. 2011). Examination of the statistically significant over-represented tissues where predicted *miR-1246* targets genes are expressed reveals that they include testis, in two of three

tissue gene expression data sets we used for this analysis (Supplemental Table S3). Furthermore, those genes that are predicted to be the most specific targets of *miR-1246* include several testis-specific genes: gametogenin (*GGN*), that is first expressed at the pachytene spermatocyte stage and the protein later is incorporated into the sperm tail (Jamsai et al. 2008); adenosine deaminase domain containing 1 (*ADAD1*), also known as testis-nuclear RNA binding protein (*Tennr*), a gene essential for mouse male fertility (Connolly et al. 2005); *TULP2*, a highly conserved gene of unknown function but highly expressed in testis; *CIZ1*, a gene expressed in waves during sperm cell differentiation and shown to be implicated in DNA replication and DNA break repair



**FIGURE 3.** The *miR-1246* gene evolved in the primate lineage and is a source of mature miRNAs in human sperm. (A) The species tree highlights primates in which the miRNA gene is predicted (marked in color) and those in which the miRNA seed sequence is conserved (marked with an asterisk). (B) Human sperm RNA samples contain the precursor of *miR-1246*, confirming that this is the source of mature *miR-1246* molecules in sperm.

(Greaves et al. 2012) (for the definition of target specificity see Materials and Methods). Considering the high levels of *miR-1246* in mature sperm and the sperm-enriched expression of its predicted target genes, we hypothesize that during primate evolution, *miR-1246* became a regulator of the sperm-differentiation process.

### CpG-island promoters of the most abundant miRNAs in sperm are packaged in chromatin with active histone marks

The sperm genome is highly compacted by protamines, with the few retained nucleosomes enriched at GC-rich regions of the genome, primarily promoters of housekeeping genes and developmental regulatory genes (Oliva 2006; Arpanahi et al. 2009; Hammoud et al. 2009; Oliva et al. 2009; Brykczynska et al. 2010; Vavouris and Lehner 2011; Erkek et al. 2013). Retention of nucleosomes at these promoters may influence their reprogramming in the early embryo. We tested whether miRNA abundance in sperm, and therefore expression late in sperm development, predicts the packaging of miRNA promoters in sperm. We found that 130 of 252 (~52%) miRNA promoters retain nucleosomes in sperm independently of miRNA abundance (Fig. 4A). In somatic cells, H3K4me3 is a histone modification that marks locations of active promoters (Bernstein et al. 2005). We tested whether promoters of highly abundant sperm miRNAs contain this “active” chromatin mark. Consistent with our expectation, promoters of the most abundant miRNAs are highly enriched in regions marked with H3K4me3-containing nucleosomes in mature sperm (Fisher’s exact test,  $P = 1.23 \times 10^{-3}$ , odds ratio = 3.43) (Fig. 4A).

We then stratified miRNA promoters according to their overlap with CpG-islands. Of 161 CpG-island promoters, 158 (98%) are hypomethylated and 116 (72%) retain nucleosomes in sperm (Fig. 4B, right). Stratifying CpG-island promoters according to miRNA abundance in sperm shows that

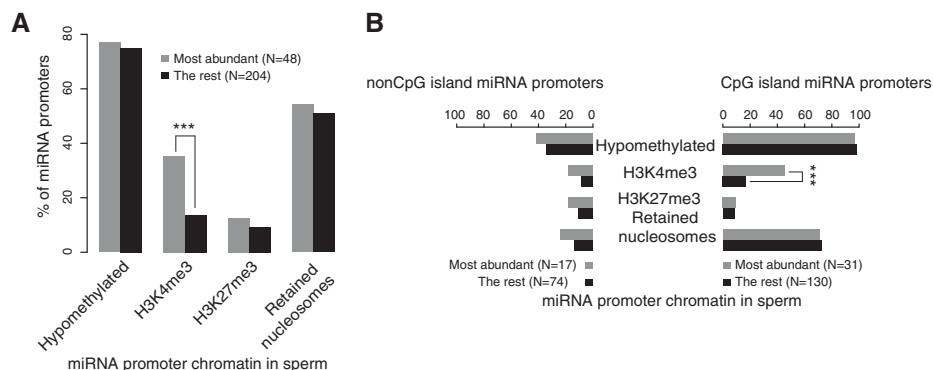
neither hypomethylation nor nucleosome retention are correlated with miRNA abundance (Fig. 4B, right). Instead, miRNA expression in late spermatogenesis predicts the likelihood that CpG-island promoters retain H3K4me3-marked nucleosomes in mature sperm (Fisher’s exact test,  $P = 1.51 \times 10^{-3}$ , odds ratio = 3.99) (Fig. 4B, right). At nonCpG-island promoters, on the other hand, none of the chromatin marks considered here are dependent on miRNA abundance in sperm (Fig. 4B, left).

We conclude that miRNA expression in sperm is not a determinant of nucleosome retention or DNA methylation levels at miRNA promoters, but rather that miRNA expression is a determinant of chromatin state in mature sperm specifically at CpG-island promoters.

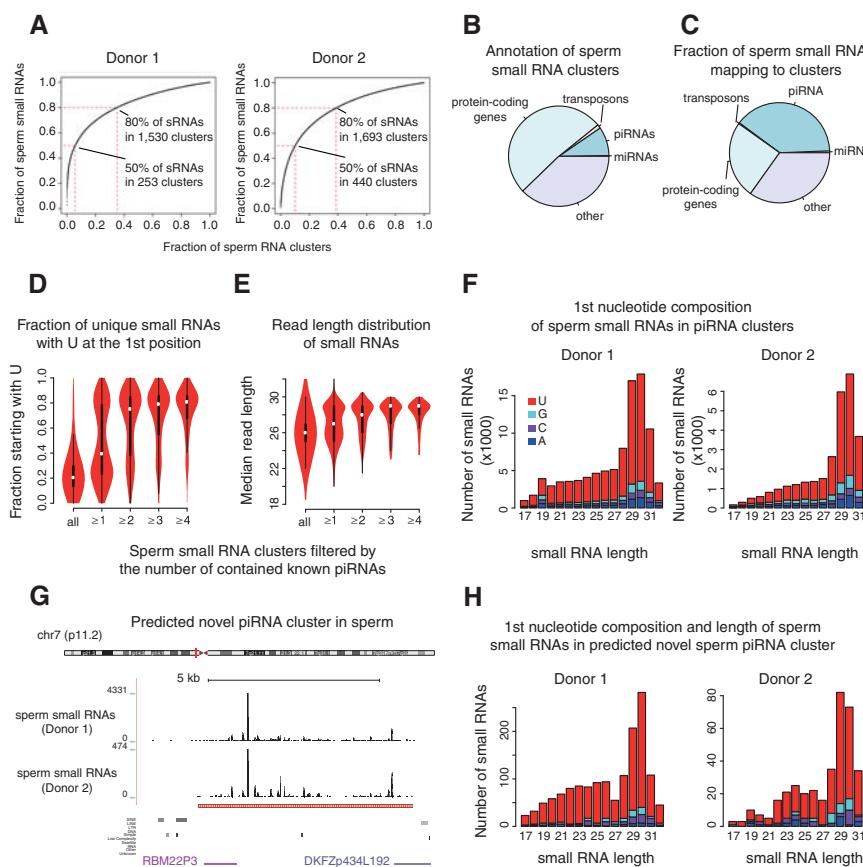
### The most abundant class of small RNAs in mature human sperm are the PIWI-interacting RNAs

For an unbiased analysis of the distribution of the rest of small RNAs along the human genome, we clustered all uniquely mapping sequences and then classified them according to the genome annotations they overlap (see Materials and Methods). We identified 7319 and 6664 clusters of small RNAs in the sample from donor 1 and donor 2, respectively, with 4376 clusters in common, covering 13.3 Mb (0.43%) of the human genome (Supplemental Table S4). Of these clusters, 6%–10% contain half of all sperm small RNAs (Fig. 5A). We focused the rest of the analysis on the 4376 small RNA clusters detected in both samples.

Among the genomic regions with the highest number of mapped sperm small RNAs there are many that overlap known piRNAs (Fig. 5B,C). To test this systematically, we retrieved and mapped small RNAs previously immunoprecipitated with PIWI from adult human testes (Girard et al. 2006). Indeed, we found that of the 20 most abundant clusters of sperm small RNAs, 18 contain at least one known piRNA and 13 contain more than 200. Almost two-thirds



**FIGURE 4.** CpG-island promoters of highly abundant sperm miRNAs retain nucleosomes with H3K4me3 in mature sperm. (A) Chromatin of all miRNA promoters in sperm. (B) Chromatin of nonCpG-island (left) and CpG-island miRNA promoters in sperm (right). miRNAs detected in both samples and with above median abundance were defined as those most abundant in sperm. Asterisks indicate statistically significant differences in chromatin between promoters of most abundant miRNAs and the rest. The Fisher’s exact test was used to assess statistical significance. Asterisks mark differences with  $P < 0.05$ , based on Fisher’s exact test.



**FIGURE 5.** Human sperm small RNAs contain piRNAs. (A) Most sperm small RNAs originate from a small number of clusters. The y-axis represents the fraction of all unique small RNAs in clusters detected in both samples. Panels on the left correspond to small RNAs from donor 1 and panels on the right correspond to small RNAs from donor 2. (B) Small RNA clusters were annotated according to the genomic features they overlap. The pie chart represents all clusters. (C) Here, each slice of the pie chart represents the total number of reads mapping to clusters annotated as miRNAs, piRNAs, transposons, protein-coding genes, and none of these. Note that only small RNAs uniquely mapping to the genome are considered here. (D) The violin plots show the fraction of small RNAs per cluster starting with a uracil. (E) The violin plots show the length distribution of sperm small RNA clusters. In both panels D and E, from left to right, clusters have been filtered according to the number of known testis PIWI immunoprecipitated piRNAs they overlap (using published piRNAs by Girard et al. 2006). Also, both panels show the length and first nucleotide distribution of sequences from donor 1. Sequences from donor 2 have essentially the same properties (data not shown). (F) Length distribution and first nucleotide composition of small RNAs from donor 1 and 2, that overlap three or more testis piRNAs. (G) A predicted novel human piRNA cluster located at chr7p11.2, overlapping a pseudogene and a lincRNA. (H) Small RNAs from the predicted novel piRNA cluster have the length and first nucleotide composition profile of piRNAs (i.e., enriched for uracil at the first position and predominantly 30 nt long).

(15,895/25,592) of known testis piRNAs map to the small RNA clusters detected in both sperm samples. Many of the remaining known piRNAs are scattered across the genome rather than clustered in regions (data not shown). In total, 408 sperm small RNA clusters contain one or more sequences that match piRNAs. Of these, 34 clusters contain 100 or more known piRNAs (13,509 known piRNAs map to these 34 clusters) and rank among the 100 most abundant small RNA clusters in sperm.

Mammalian piRNAs typically have uracil at the first position and are 24–31 nt long. To test whether the sperm small

RNA clusters in sperm that we have detected are piRNAs, we checked their length and first nucleotide composition. Clusters that contain matches to three or more piRNAs are highly enriched in sequences that start with a uracil and are 30 nt long (Fig. 5D,E). These are indeed the characteristics of small RNAs bound to the PIWI protein expressed in late spermatogenesis in adult testes (Girard et al. 2006). There are 176 clusters that contain matches to three or more known piRNAs and these clusters contain thousands of previously unannotated small RNAs generated from human piRNA clusters that have the length distribution and first nucleotide composition of piRNAs (Fig. 5F). Last, we have identified a potential novel human piRNA cluster on chromosome 7 that has sequences with all the hallmarks of piRNAs (Fig. 5G,H) and is therefore a candidate novel human piRNA cluster.

### Most abundant sperm piRNAs target LINE1 transposons

The best understood function of piRNAs is defense against repeat activation at the wave of genome-wide reprogramming during germline development. piRNAs loaded on PIWI proteins target transposon transcripts by extensive sequence complementarity (Reuter et al. 2011). Reuter et al. (2011) showed that complementarity in nucleotides 2–22 of the piRNA is most likely essential for targeting. To study the potential function of sperm piRNAs, we selected the 100 most abundant sperm small RNAs present in both samples and mapping to genomic clusters with at least three known piRNAs (Supplemental Table S5). To predict targets for these sequenc-

es, we searched for locations in the human genome that match nucleotides 2–22 of human sperm piRNAs. Of 136 predicted targets, 31 (23%) map to LINE1 repeats which is greater than expected based on the LINE1 repeat coverage of the human genome (expected 17%,  $P < 10^{-16}$ ). LINE1 transposons are known targets of PIWI proteins during male germline development (Aravin et al. 2007; Carmell et al. 2007; Kuramochi-Miyagawa et al. 2008; De Fazio et al. 2011; Reuter et al. 2011; Di Giacomo et al. 2013). Here we show LINE1 targeting piRNAs are retained in mature human sperm.

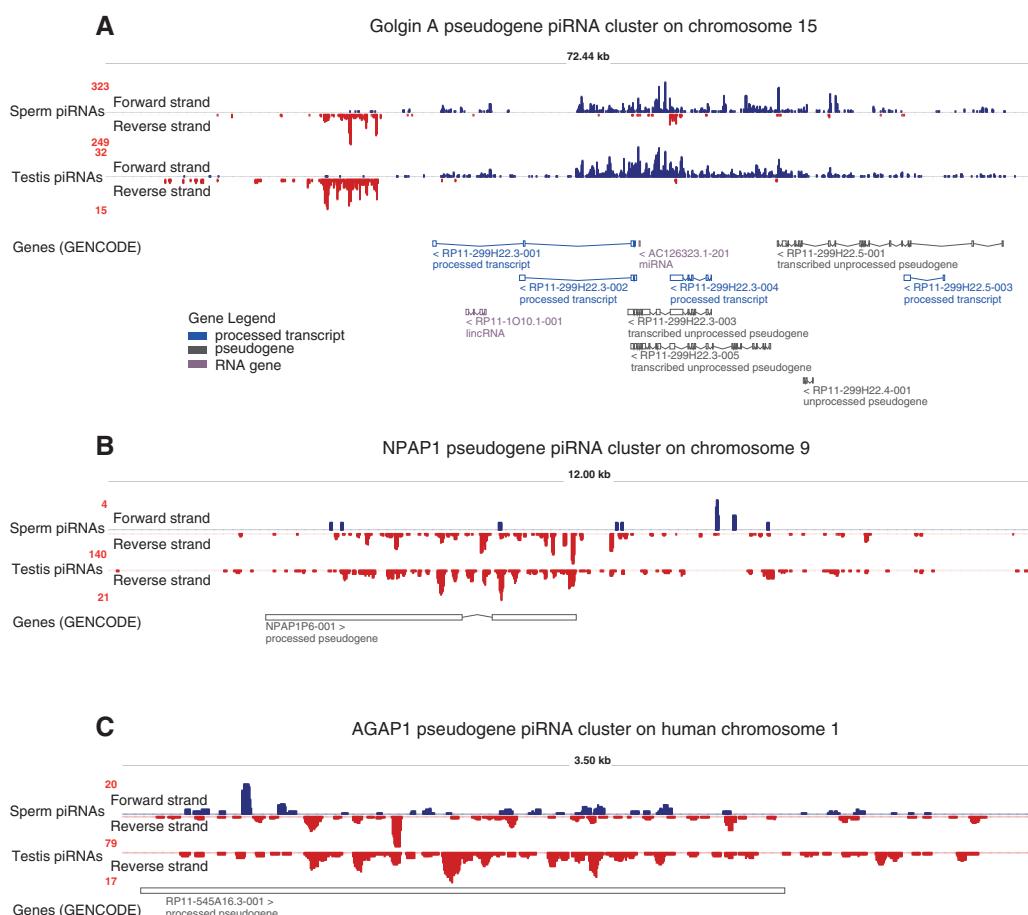
## Human piRNA clusters contain pseudogenes

Although sperm piRNA clusters are depleted of protein-coding genes, we noted that they do contain noncoding genes (Fig. 6). We found multiple cases where clusters of human piRNAs are processed from the strand that is antisense with respect to the annotated strand of long noncoding RNAs. Interestingly, these noncoding RNAs include pseudogenes of protein-coding genes. We found that 30 pseudogenes are transcribed from their antisense strand and are processed into small RNAs present in mature sperm (Fig. 6; Supplemental Table S6). One of the largest human piRNA clusters is shown in Figure 6A and contains the annotated noncoding RNA genes RP11-299.3, RP11-299.4, and RP11-229.5. According to the gold standard reference annotation of the human genome by the HAVANA team, all three are golgin A subfamily pseudogenes. Therefore, piRNA clusters do not only accumulate fragments of transposable elements (Aravin et al. 2007; Brennecke et al. 2007), but also parts of protein-coding genes. A second pseudogene-containing

piRNA cluster is shown in Figure 6B. This cluster overlaps the processed pseudogene of *NPAP1* (also known as *C15orf2*). *NPAP1* is a single-exon gene coding for a nuclear pore complex protein (Neumann et al. 2012). The function and regulation of *NPAP1* is of relevance to human disease as it lies in the imprinted critical region of the Prader–Willi syndrome (Farber et al. 2000). Also, testis dysfunction has also been associated with the Prader–Willi syndrome (e.g., Katcher et al. 1977; Siemersma et al. 2012). Multiple sperm and testis piRNAs are processed from the antisense strand with respect to the pseudogene (*NPAP1P6*). We conclude that piRNAs antisense to human protein-coding genes derive from pseudogene-containing piRNA clusters.

## Pseudogene-derived piRNAs are predicted to target protein-coding genes

We reasoned that some of the piRNAs derived from pseudogenes could potentially regulate the expression of their parent

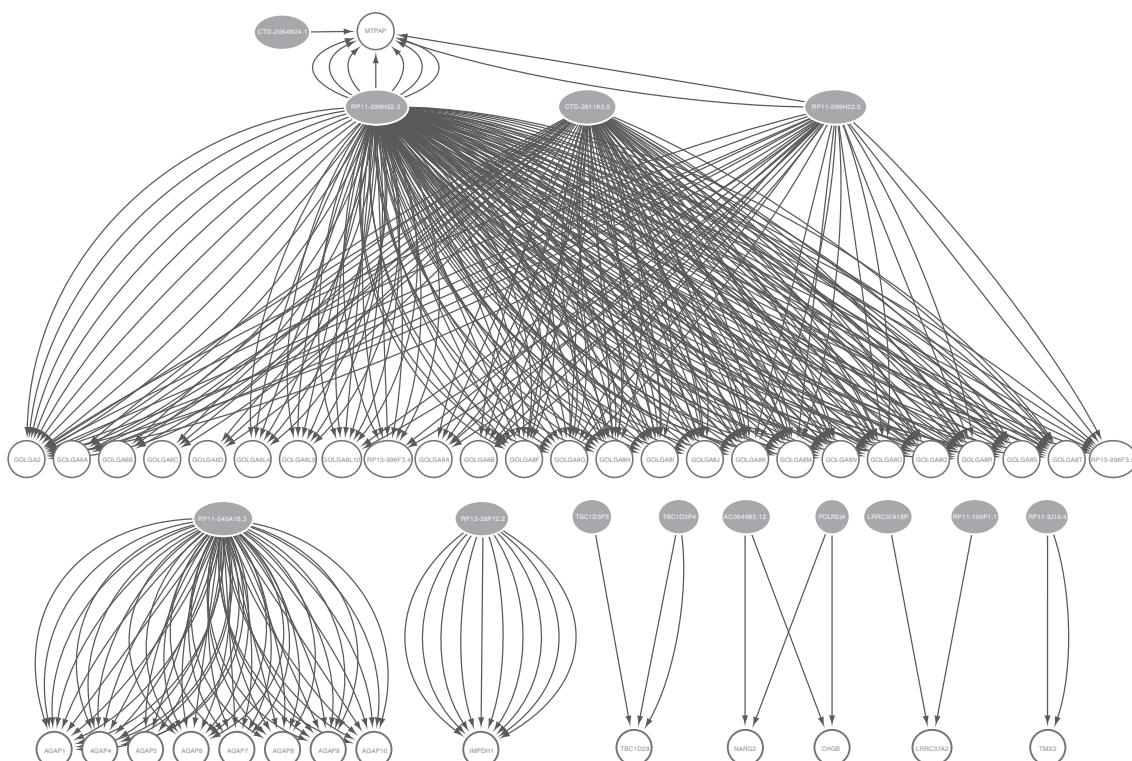


genes in the male germline in the same way that siRNAs processed from pseudogenes regulate expression of their parent genes in oocytes (Tam et al. 2008; Watanabe et al. 2008). To explore this possibility further, we retrieved all known human piRNAs from adult testes (Girard et al. 2006) and predicted targets using the 2–22 nt piRNA mammalian targeting rule by Reuter et al. (2011). We found 96 piRNAs processed from the antisense strand of 23 known human pseudogenes that target exons of 118 protein-coding genes. Being more stringent and considering only protein-coding genes targeted by at least two different piRNAs, there are 11 pseudogenes generating piRNAs that target 38 protein-coding genes (Fig. 7). Notably, there is a dense network of piRNAs generated from golgin A pseudogenes and targeting their parent protein-coding genes (Figs. 6A, 7). This dense network reflects the fact that golgin A protein-coding genes experienced a large expansion within the primate lineage and golgin A pseudogene-derived piRNAs match identical regions in many of these recently duplicated protein-coding genes. The conservation of sequence between pseudogene piRNAs and protein-coding genes may reflect evolutionary pressure to maintain piRNA targeting. Additionally, a cluster on chromosome 1 that contains an *AGAP1* pseudogene generates antisense piRNAs that target eight different *AGAP* protein-coding genes (Figs. 6C, 7). We conclude that pseu-

dogene-derived piRNAs may target homologous protein-coding genes in the human genome. These piRNAs are expressed during germline development in adult human testes and are still present in mature human sperm cells.

## DISCUSSION

Here, we have characterized the small RNA content of human sperm. Although sperm cells have a very small total RNA content, we have shown here that they do contain a complex repertoire of small RNAs. In agreement with a previous report (Krawetz et al. 2011), we have found that although mature sperm cells contain miRNAs, these constitute only a small fraction of the total sperm small RNA content. We have detected 182 mature miRNAs in sperm RNA samples from two unrelated normozoospermic individuals. The three most abundant sperm miRNAs are *miR-1246*, *miR-34c*, and *miR-10a*. Of these, *miR-34c* is conserved in mammals and *miR-10a* is conserved in diverse animals including insects. In contrast, we show here that *miR-1246* is a recently evolved gene with a seed conserved in primates closely related to humans. There is little known about the function of *miR-1246*, as this miRNA is not conserved in model organisms and it was discovered very recently. However, here we present evidence that suggests that its predicted target genes include



many that are expressed in testes, which is consistent with a role of this small RNA in male germline development. We expect that the miRNAs we have detected in sperm regulate sperm development by either repressing or enhancing mRNA translation (Orom et al. 2008). Furthermore, although controversial, there is some evidence that sperm miRNAs are transmitted to the fertilized egg and are involved in gene regulation during pre-implantation (e.g., *miR-34c* [Liu et al. 2012] and *miR-1* [Wagner et al. 2008]). A high-quality resource of sperm miRNAs should facilitate studies on the potential impact of paternal small RNAs on early embryo development. It can also be used to identify biomarkers of male infertility.

We also analyzed the chromatin of microRNA promoters in mature human sperm and tested whether their expression during sperm development has any impact on their packaging. We found that CpG-island promoters of sperm miRNAs tend to be packaged by histones rather than protamines in mature sperm. This is consistent with previous observations that GC-rich regions of the genome, including promoter CpG-islands, retain nucleosomes in mature sperm (Vavouri and Lehner 2011; Erkek et al. 2013). We also found that retained histones at CpG-island promoters of miRNAs that are highly abundant in sperm, therefore, miRNAs highly expressed during late stages of sperm development, tend to have the “active” mark H3K4me3. In somatic cells, this mark is found at actively transcribing promoters. Although mature sperm cells are transcriptionally inactive, this mark set earlier during sperm development remains at miRNA promoters that retain nucleosomes. The presence of this active mark at some microRNA promoters in sperm may affect their activation potential in the early embryo, as previously proposed for promoters of protein-coding genes (Gardiner-Garden et al. 1998; Hammoud et al. 2009; Vavouri and Lehner 2011).

The most abundant set of regulatory small noncoding RNAs in mature sperm are piRNAs. This is a class of small RNAs bound by the PIWI proteins and found in mammals predominantly in the male germline (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006). Surprisingly, we have shown here that some of the biggest piRNA clusters contain relics of protein-coding genes that, just like retrotransposons, are transcribed and processed from the antisense strand. Many pseudogene-derived piRNAs have predicted targets in the respective parent protein-coding genes. We speculate that pseudogene-derived piRNAs may target mRNA transcripts for degradation or that they may guide the DNA methylation machinery to their parent genes for epigenetic silencing. It is worth noting that pseudogene-derived endogenous siRNAs have been found in mouse oocytes where they repress transcription from targeted loci (Tam et al. 2008; Watanabe et al. 2008). Furthermore, it has been shown that insertion of DNA targeting genes into piRNA clusters in flies and mice, leads to their processing into piRNAs and silencing of the targeted genes *in trans* (Muerdter et al. 2012; Yamamoto et al. 2013). Therefore, in

principle, the pseudogene-derived piRNAs we have identified could target and post-transcriptionally regulate expression of their parent genes in the human male germline.

In conclusion, we have found that human mature sperm cells contain a multitude of small noncoding regulatory RNAs. These include hundreds of miRNAs and tens of thousands of piRNAs. These small RNAs most likely participate in the complex program of post-transcriptional regulation of the late stages of spermatogenesis, and may also contribute to epigenetic regulation in the early embryo. Our data set and results contribute to an improved understanding of germline RNA. Questions to be answered in future studies include whether dysregulation of these small RNAs is involved in human infertility and if sperm small RNAs have a regulatory role in the pre-implantation embryo.

## MATERIALS AND METHODS

### Sperm collection, RNA extraction, and quality controls

Sperm samples were obtained after informed consent from two unrelated fertile sperm donors from the Andrology Unit of the Hospital Clinic of Barcelona. To isolate pure sperm cells, we performed a 50% Percoll gradient and washed the cells in somatic cell lysis buffer (0.1% SDS, 0.5% Triton). Total RNA including small RNAs was isolated using the Qiagen miRNeasy kit, according to the manufacturer’s recommendations. In order to achieve complete disruption of sperm we added  $\beta$ -mercaptoethanol to the QIAzol Lysis Reagent (Jodar et al. 2012). It is important to note that this commercial kit is specifically optimized to a high yield of small RNAs. The purity and concentration of RNA samples were checked spectrophotometrically at 260 and 280 nm. In order to assess the integrity of the RNA and the absence of genomic DNA, we performed RT-PCR for the protamine-2 gene with exon-spanning primers. Finally, to verify the absence of somatic cells we confirmed the lack of 18S and 28S peaks in the RNA profile using the Agilent Bioanalyzer and absence of the leukocyte-specific marker CD45 by RT-PCR (Jodar et al. 2012).

### Library preparation and sequencing

Small RNA libraries were prepared using the Small RNA Sample Prep Kit (Illumina) following the alternative v1.5 protocol with some modifications. A total of 50 ng RNA was used as input. The v1.5 sRNA 3' adapter was diluted 30 $\times$  instead of 10 $\times$ . The SRA 5' adapter was diluted 3 $\times$ . The entire ligation product was used as template in the reverse transcription reaction instead of one-third of it. The resulting library was amplified by 16 cycles of PCR. Libraries were separated on 3% MetaPhor gels, fragments of 90–110 bp (15–35 bp inserts) were excised and purified using Qiaquick Gel Extraction Kit (Qiagen). Libraries were sequenced 37 bp on a Genome Analyzer IIx (Illumina) following the manufacturer’s protocol.

### Small RNA read processing

The sequencing adapter was identified and removed from the reads using cutadapt (Martin 2011), allowing a minimum match of 5 nt,

error rate 0.125 and a minimum small RNA length of 17 nt. Reads failing to match the adapter were discarded, poly-A tails were trimmed and reads were collapsed to unique sequences, retaining the read count for each sequence.

### Identification of sperm microRNAs

Sperm miRNAs were identified using SeqBuster (Pantano et al. 2010) by comparing the small RNA sequences against miRBase (version 19) (Kozomara and Griffiths-Jones 2011). miRNAs were predicted allowing no mismatches, 3 nt trimmed from the end and no nucleotides added. miRNA expression values were scaled by dividing by the size factor estimated by DESeq2 (Anders and Huber 2010).

### Identification of microRNA targets and target gene tissue expression analysis

We retrieved the predicted *miR-1246* targets from [www.microrna.org](http://www.microrna.org) (Betel et al. 2008). As *miR-1246* itself is not deeply conserved in mammals, the targets correspond to genes that have good mirSVR scores without restricting to conserved sites (total of 2503 predicted target genes). As a background gene set for functional enrichment analysis with DAVID, we used all genes that have good mirSVR score predicted miRNA sites without conservation (total 19,778 predicted target genes). Tissue expression enrichment analysis was carried out using the functional annotation tool DAVID (release 6.7) (Huang da et al. 2009). To account for biases due to 3' UTR length, we also used the *miR-1246* target specificity score defined as the number of predicted *miR-1246* sites normalized by the total number of predicted microRNA binding sites per gene as previously described in Morin et al. (2008). With a score cutoff of 0.10 we found the following *miR-1246*-specific target genes; *TULP2*, *GGN*, *EEF1B2*, *ADAD1*, *ACOT6*, and *CIZ1*.

### Evolutionary analysis of *miR-1246*

We used Ensembl to identify the coordinates of the full *miR-1246* gene and its orthologous region in 35 other eutherian mammals. We then submitted each sequence to three programs that predict microRNA genes; MiRPara (Wu et al. 2011), Eumir (<http://miracle.igib.res.in/eumir/>) and miRPred (Brameier and Wiuf 2007). We ran all programs with default parameters and scored which program predicted that each sequence has an miRNA-like secondary structure (Supplemental Table S2). We manually searched the multiple sequence alignment for the human *miR-1246* seed (AATCCAT).

### Annotation of chromatin state of sperm miRNA promoters

We used the miRNA promoter annotation from Marson et al. (2008). We converted the coordinates of these promoters from hg17 to hg18 using the LiftOver tool at the UCSC Genome Browser (Fujita et al. 2011). We excluded miRNAs that have multiple precursors in the human genome. In the case of miRNAs with multiple annotated promoters, we used the promoter with the highest score. In the case of miRNA clusters, we defined the expression of the promoter as the maximum expression of any of the contained miRNAs. We defined the epigenetic state of miRNA promoters according to their overlap

with peaks of sperm H3K4me3, H3K27me3 retention (Hammoud et al. 2009), and sperm hypomethylated regions (HMR) (Molaro et al. 2011). Highly abundant piRNAs were considered those detected in both samples and with above median abundance value.

### Mapping, clustering, and annotation of sperm small RNAs

We mapped the remaining sequences to the human genome (version hg19) (Fujita et al. 2011) using bowtie (version 0.12.7) (Langmead et al. 2009), allowing up to one mismatch reporting only the best alignments and reporting only alignments that map to unique locations (bowtie options `-a -best -strata -v 1 -m 1`). We identified small RNAs that overlap known genomic features using the human gene annotation from Ensembl release 71 (Flicek et al. 2013), the repeat annotation from UCSC (<http://www.repeatmasker.org>; Fujita et al. 2011) and BedTools (Quinlan and Hall 2010). In addition, we retrieved known human adult testis piRNAs from Girard et al. (2006). We mapped these to the human genome using the same bowtie command described above.

We clustered together mapped small RNAs transcribed from the same strand, lying within 1 kb of each other and analyzed further clusters that contain a minimum of 10 small RNAs in each sperm sample. The two sets of sperm small RNA clusters were intersected retaining only those present in both samples and defining the coordinates of the common clusters as those of the most extremely positioned reads. We annotated clusters according to their overlap with the same annotations mentioned above. Here (with the exception of the overlap with known piRNAs), we also required that at least 50% of the cluster overlaps an annotation in order for it to be annotated as such. Clusters were annotated according to their overlap with miRNA genes first, piRNAs next, then transposons and last protein-coding genes. This was done to avoid intronic miRNAs, piRNAs and repeat clusters being classified as genic (Pei et al. 2012). Human pseudogenes were retrieved from Ensembl release 71.

### piRNA target prediction

To predict targets of piRNAs we mapped nucleotides 2–22 of small RNAs against the human genome using bowtie allowing no mismatches and reporting all hits (options `-v 0 -a`). Predicted targeted genes and repeats were considered only those in the antisense orientation with respect to the sperm small RNA hit in the genome. We predicted targets for the most highly abundant sperm small RNAs. To select these, we ranked small RNA clusters according to their read count (averaged over the two samples) and from these we identified the 100 most abundant small RNAs present in both samples. To calculate the expected percentage of targets in the genome, we calculated the proportion of the human genome that is covered by each type of annotation. We also predicted targets for all human piRNAs from adult human testes (Girard et al. 2006). The figure of the network of pseudogene-derived piRNAs targeting human protein-coding genes was generated in Cytoscape (Shannon et al. 2003).

### DATA DEPOSITION

The raw data have been submitted to the Short Read Archive (SRA) with accession number SRP029517.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

We thank Ben Lehner for comments on the manuscript. T.V. was funded by MICINN grant (BFU2011-30246), Ramon y Cajal grant (RYC-2010-07114), European Commission Framework 7 European Re-integration grant (PERG08-GA-2010-276741), and by the Institute of Predictive and Personalized Medicine of Cancer (IMPPC). N.T. and M.B. were supported by The Danish Council for Independent Research (Technology and Production Sciences) and The Lundbeck Foundation. R.O. was funded by the Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad; FEDER BFU 2009-07118 and PI13/00699), by Fundación Salud 2000 (SERONO 13-015), and by European Commission training network grant (FP7-PEOPLE-2011-ITN289880).

Received May 20, 2014; accepted February 19, 2015.

## REFERENCES

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**: 203–207.

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744–747.

Arpanah A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D. 2009. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* **19**: 1338–1349.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ III, Gingras TR, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**: 169–181.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**: D149–D153.

Bouhallier F, Allioli N, Laval F, Chalmeil F, Perrard MH, Durand P, Samarut J, Pain B, Rouault JP. 2010. Role of miR-34c microRNA in the late steps of spermatogenesis. *RNA* **16**: 720–731.

Brameier M, Wiuf C. 2007. Ab initio identification of human microRNAs based on structure motifs. *BMC Bioinformatics* **8**: 478.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**: 1089–1103.

Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schubeler D, Stadler MB, Peters AH. 2010. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* **17**: 679–687.

Carmell MA, Girard A, van de Kant HJG, Bourc'his D, Bestor TH, De Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**: 503–514.

Connolly CM, Dearth AT, Braun RE. 2005. Disruption of murine Tenr results in teratospermia and male infertility. *Dev Biol* **278**: 13–21.

Cooper TG. 2005. Cytoplasmic droplets: the good, the bad or just confusing? *Hum Reprod* **20**: 9–11.

De Fazio S, Bartonicik N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**: 259–263.

Di Giacomo M, Comazzetto S, Saini H, De Fazio S, Carrieri C, Morgan M, Vasiliauskaitė L, Benes V, Enright AJ, O'Carroll D. 2013. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol Cell* **50**: 601–608.

Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schübeler D, van der Vlag J, Stadler MB, Peters AH. 2013. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* **20**: 868–875.

Farber C, Gross S, Neesen J, Buiting K, Horsthemke B. 2000. Identification of a testis-specific gene (C15orf2) in the Prader-Willi syndrome region on chromosome 15. *Genomics* **65**: 174–183.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. *Nucleic Acids Res* **41**: D48–D55.

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. 2011. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* **39**: D876–D882.

Gardiner-Garden M, Ballesteros M, Gordon M, Tam PP. 1998. Histone- and protamine-DNA association: conservation of different patterns within the  $\beta$ -globin domain in human sperm. *Mol Cell Biol* **18**: 3350–3356.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202.

Greaves EA, Copeland NA, Coverley D, Ainscough JF. 2012. Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. *J Cell Sci* **125**: 2466–2477.

Grivna ST, Beyret E, Wang Z, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20**: 1709–1714.

Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. 2009. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**: 473–478.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.

Jamsai D, Bianco DM, Smith SJ, Merriner DJ, Ly-Huynh JD, Herlihy A, Niranjan B, Gibbs GM, O'Bryan MK. 2008. Characterization of gametogenin 1 (GGN1) and its potential role in male fertility through the interaction with the ion channel regulator, cysteine-rich secretory protein 2 (CRISP2) in the sperm tail. *Reproduction* **135**: 751–759.

Jodar M, Kalko S, Castillo J, Ballesca JL, Oliva R. 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. *Hum Reprod* **27**: 1431–1438.

Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA; Reproductive Medicine Network. 2013. The presence, role and clinical use of spermatozoal RNAs. *Hum Reprod Update* **19**: 604–624.

Johnson GD, Sendler E, Lalancette C, Hauser R, Diamond MP, Krawetz SA. 2011. Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Mol Hum Reprod* **17**: 721–726.

Katcher ML, Bargman CJ, Gilbert EF, Opitz JM. 1977. Absence of spermatogonia in the Prader-Willi syndrome. *Eur J Pediatr* **124**: 257–260.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152–D157.

Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, Draghici S, Diamond MP. 2011. A survey of small RNAs in human sperm. *Hum Reprod* **26**: 3401–3412.

Kumar G, Patel D, Naz RK. 1993. c-MYC mRNA is present in human sperm cells. *Cell Mol Biol Res* **39**: 111–117.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008.

DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908–917.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science* **313**: 363–367.

Liu WM, Pang RT, Chiu PC, Wong BP, Lao K, Lee KF, Yeung WS. 2012. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc Natl Acad Sci* **109**: 490–494.

Lund AH. 2010. miR-10 in development and cancer. *Cell Death Differ* **17**: 209–214.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**: 521–533.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10–12.

Martins RP, Krawetz SA. 2005. RNA in human sperm. *Asian J Androl* **7**: 115–120.

Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**: 1029–1041.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, et al. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* **18**: 610–621.

Muerdter F, Olovnikov I, Molaro A, Rozhkov NV, Czech B, Gordon A, Hannon GJ, Aravin AA. 2012. Production of artificial piRNAs in flies and mice. *RNA* **18**: 42–52.

Neumann LC, Markaki Y, Mladenov E, Hoffmann D, Buiting K, Horsthemke B. 2012. The imprinted *NPAP1/C15orf2* gene in the Prader–Willi syndrome region encodes a nuclear pore complex associated protein. *Hum Mol Genet* **21**: 4038–4048.

Oliva R. 2006. Protamines and male infertility. *Hum Reprod Update* **12**: 417–435.

Oliva R, de Mateo S, Estanyol JM. 2009. Sperm cell proteomics. *Proteomics* **9**: 1004–1017.

Orom UA, Nielsen FC, Lund AH. 2008. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* **30**: 460–471.

Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA. 2002. Spermatozoal RNA profiles of normal fertile men. *Lancet* **360**: 772–777.

Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. 2004. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429**: 154.

Pantano L, Estivill X, Marti E. 2010. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res* **38**: e34.

Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, Harte R, Balasubramanian S, Tanzer A, Diekhans M, et al. 2012. The GENCODE pseudogene resource. *Genome Biol* **13**: R51.

Pessot CA, Brito M, Figueiroa J, Concha II, Yañez A, Burzio LO. 1989. Presence of RNA in the sperm nucleus. *Biochem Biophys Res Commun* **158**: 272–278.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, Antony C, Sachidanandam R, Pillai RS. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**: 264–267.

Rohwedder A, Liedigk O, Schaller J, Glander HJ, Werchau H. 1996. Detection of mRNA transcripts of  $\beta$ 1 integrins in ejaculated human spermatozoa by nested reverse transcription–polymerase chain reaction. *Mol Hum Reprod* **2**: 499–505.

Sendler E, Johnson GD, Mao S, Goodrich RJ, Diamond MP, Hauser R, Krawetz SA. 2013. Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res* **41**: 4104–4117.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504.

Siemensma EP, de Lind van Wijngaarden RF, Otten BJ, de Jong FH, Hokken-Koelega AC. 2012. Testicular failure in boys with Prader–Willi syndrome: longitudinal studies of reproductive hormones. *J Clin Endocrinol Metab* **97**: E452–E459.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534–538.

Vavouris T, Lehner B. 2011. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet* **7**: e1002036.

Wagner KD, Wagner N, Ghanbarian H, Grandjean V, Gounon P, Cuzin F, Rassoulzadegan M. 2008. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev Cell* **14**: 962–969.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539–543.

Wu Y, Wei B, Liu H, Li T, Rayner S. 2011. MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. *BMC Bioinformatics* **12**: 107.

Wykes SM, Visscher DW, Krawetz SA. 1997. Haploid transcripts persist in mature human spermatozoa. *Mol Hum Reprod* **3**: 15–19.

Yamamoto Y, Watanabe T, Hoki Y, Shirane K, Li Y, Ichiiyanagi K, Kuramochi-Miyagawa S, Toyoda A, Fujiyama A, Oginuma M, et al. 2013. Targeted gene silencing in mouse germ cells by insertion of a homologous DNA into a piRNA generating locus. *Genome Res* **23**: 292–299.

Zhang Y, Liao JM, Zeng SX, Lu H. 2011. p53 downregulates Down syndrome-associated DYRK1A through miR-1246. *EMBO Rep* **12**: 811–817.



## The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes

Lorena Pantano, Meritxell Jodar, Mads Bak, et al.

RNA 2015 21: 1085-1095 originally published online April 22, 2015  
Access the most recent version at doi:[10.1261/rna.046482.114](https://doi.org/10.1261/rna.046482.114)

---

**Supplemental Material** <http://rnajournal.cshlp.org/content/suppl/2015/04/14/rna.046482.114.DC1.html>

**References** This article cites 67 articles, 31 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/21/6/1085.full.html#ref-list-1>

**Creative Commons License** This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---



---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>