SMALL NONCODING RNA CONTENTS OF RAT EPIDIDYMAL EXTRACELLULAR VESICLES INCLUDING A PUTATIVE NOVEL SMALL RNA

Running Head: Small noncoding RNAs in epididymal EVs

Keywords: sncRNA, extracellular vesicles, epididymis, epididymosomes, miRNA, tRNA fragments, CpG islands, rat

Ross Gillette[1], Shyamal Waghwala[2], Andrea C. Gore[1†]

[1] The Center for Molecular Carcinogenesis and Toxicology, Division of Pharmacology and Toxicology, College of Pharmacy, The University of Texas at Austin, Austin, TX, 78712, USA

[2] College of Natural Sciences, The University of Texas at Austin, Austin, TX, 78712, USA

AUTHOR CONTRIBUTIONS

Conceptualization: R.G.; Methodology: R.G.; Investigation: R.G.; Formal Analysis: R.G., S.W.; Writing – Original Draft: R.G.; Writing – Review & Editing: R.G., A.C.G.; Funding Acquisition: R.G., A.C.G.

[†] Corresponding Author: andrea.gore@austin.utexas.edu

ABSTRACT

Extracellular vesicles (EVs) released from the epididymal epithelium (epididymosomes) impart functional competence on sperm as they transit the epididymis by merging with sperm and releasing a complex repertoire of molecules. The cargo of epididymosomes includes small noncoding RNAs (sncRNAs) that are modified by external factors such as stress, nutrition, and drug use. If incorporated into sperm, the EV sncRNA cargo can affect offspring and lead to heritable phenotypes. In the current study we characterized the RNA contents of EVs collected from the caput epididymis of adult male rats in order to fill a gap in knowledge in this species and to establish a sncRNA profile. Small RNAs of EVs were isolated from the caput portion of the epididymis of adult male rats, and sequenced on a NovaSeq 6000 on a SP flow cell in a single-end 50 bp configuration. The resulting reads were checked for quality, trimmed for adapter sequences, aligned to the unmasked rat genome (Rnor 6), and assigned an annotation designation. The majority of RNA reads aligned to either tRNA fragments (79.1%) or piRNA (18.1%) loci. Micro RNAs (miRNAs) accounted for a surprisingly small proportion of reads (0.18%). The third largest category of aligned reads (1.5%) was in intergenic space and not strictly associated with canonical sncRNA loci. In-depth investigation determined these latter reads (~19 nt) aligned strictly within the boundaries of known CpG islands (CpGi), which have not previously been reported to express any form of sncRNA. These newly described "CpGi sRNAs" could not consistently be accounted for by overlaying features of any other annotation type (including rRNA and piRNA). The CpGi sRNAs have characteristics of RNA fragments that can associate with the Argonaute/PIWI family of proteins and therefore could have regulatory function via RNA induced silencing or de novo DNA methylation. We propose that CpGi sRNAs constitute a new family of sncRNA that may represent an important and unreported class of regulatory RNA in gametes.

INTRODUCTION

Spermatozoa are incapable of fertilization as they exist the testes [1] and only gain motility and the capacity for fertilization as they transit the epididymis [2]. This process requires epididymosomes [3–6], a type of extracellular vesicle (EV) within the microvesicle category [7–9]. Epididymosomes are released from the epididymal epithelium via apocrine secretion [10], fuse to spermatozoa [11], aid in their maturation and functionalization [4,12,13], and mark dead or dying spermatozoa for elimination [14,15]. Epididymosomes carry a complex repertoire of ions, proteins and small non-coding RNAs (sncRNA) [4,12,16,17] that are transferred to [18,19] and presumably alter the function of the spermatozoa or impart functional consequences on the fertilized embryo [20]. Importantly, sncRNA molecules transferred to spermatozoa have the capacity to alter the epigenetic landscape of either the spermatozoa or the fertilized embryo and thereby influence subsequent generations. Furthermore, the sncRNA contents of epididymosomes are subject to modification by the hormonal milieu [21,22] and environmental challenges such as psychosocial or metabolic stress [22–27], making them a potential vector for intergenerational effects in offspring. This key role was illuminated by studies showing that early life stress caused behavioral and physiological changes in the offspring, phenotypes that were recapitulated by microinjection of a small number of microRNAs into a fertilized egg subsequently used for *in vitro* fertilization [28,29].

The sncRNA repertoire of epididymosomes is complex and diverse [30]. The most abundant sncRNA found in epididymosomes is microRNA (miRNA) followed by smaller proportions of transfer RNA (tRNA) and ribosomal RNA (rRNA) [25,30,31]. The remainder consists of piwi-interacting RNA (piRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA), as well as long non-coding RNA (lncRNA) [25]. Other species of sncRNA, such as YRNA and Vault RNA, have not yet been reported in epididymosomes but should not be ruled out for inclusion in these complex profiles. Furthermore, there are likely additional sncRNAs yet to be identified [32].

Much of the knowledge we have on the sncRNA contents of caput epididymosomes, which are crucial for the final steps of spermatozoa maturation, is from mouse models. Very little has been reported on the epididymosomes of rats, which are an important model system for endocrinology, reproduction, and epigenetic transgenerational inheritance, and in which behavioral work is considered to better translate to humans [33–35]. Here, we characterized the contents of EVs derived from the caput epididymis of the male rat and describe what we believe to be a previously undefined class of sncRNA.

MATERIALS AND METHODS

Animals and treatment

All animal experiments were conducted using humane procedures that were pre-approved by the Institutional Animal Care and Use Committee at The University of Texas at Austin and in Accordance with NIH guidelines. Three-month old male and female Sprague Dawley rats were purchased (Envigo, Indianapolis, IN) and shipped to the Animal Resource Center at the University of Texas at Austin and allowed two weeks to acclimate to the housing facility. All animals in the colony were housed in a room with consistent temperature (~22C) and light cycle (14:10 light:dark) and had *ad libitum* access to filtered tap water and a rat chow with minimal phytoestrogens (Teklad 2019: Envigo, Indianapolis, IN). After acclimation, female rats were observed for vaginal cytology indicating proestrus, and therefore receptivity. Receptive female rats were randomly paired with an experienced breeder male, receptivity was confirmed, and the pair was left together overnight. The following morning, vaginal cytology was checked for the presence of sperm and if present marked as embryonic day (E1). The pregnant dams were single-housed and provided nesting material on E18.

On E8, the dams (N = 6) were randomly split into two groups (each N = 3) and fed ¼ Nilla Wafers™ with either 3% DMSO or 1 mg/kg Aroclor 1221 (A1221) (#C-221N, Accustandard, Lot #072-202-01 - an estrogenic polychlorinated biphenol) from E8 – E18 and postnatal day (PND) 1

– 21. Treatment was not considered or analyzed in any of the data presented in this manuscript as the goal was to first characterize the composition of EV sncRNAs, with greater statistical power attained by combining the 6 litters. Treatment details are provided here for transparency. At PND 21, all pups were weaned and housed in cages of two or three. Only males were used in the present manuscript and were otherwise unmanipulated until euthanasia besides weekly handling to acclimate each rat to the experimenters in order to reduce stress. Females were used for other projects.

Sample Collection

At PND 105, 6 male rats were randomly selected (N = 1 per litter) and euthanized by rapid decapitation. The testis and epididymis were removed via a small incision in the scrotum. The epididymis was separated from the testis and segmented into three portions (caput, corpus, and cauda). The caput of the epididymis was minced into small pieces with scissors and placed in warm (37 C) M2 culture media (M7167, Millipore Sigma) with HEPES and 2% exosome depleted fetal bovine serum (A2720803, ThermoFisher). The slurry was placed on a rocker for 30 minutes to allow sperm and epididymal fluid to suspend in solution, the supernatant was removed and large tissue chunks were excluded. The resulting supernatant was centrifuged at 500 X g for 5 minutes to pellet and remove sperm. The supernatant was again removed, immediately frozen on dry ice, and stored at -80 C until use.

Extracellular Vesicle Purification and Nucleic Acid Extraction

The buffer supernatant containing extracellular vesicles from the caput of the epididymis was thawed on ice for 30 minutes prior to isolation. After it was fully thawed the buffer was mixed by pipetting and then filtered with a 0.8 um syringe filter (Millex-AA, SLAA033DD, Millipore Sigma) to remove any cells or large cellular debris. Extracellular vesicles and RNA were sequentially isolated from the media using the Qiagen exoRNeasy Midi kit (#77144, Qiagen, Germantown, MD) according to the manufacturer's protocol. Briefly, 200 ul of media from each sample was passed through a spin column that selectively binds exosomes. The column was

washed to remove contaminants and debris, and then eluted in Qiazol lysis buffer (#79306, Qiagen, Germantown, MD). Chloroform (AAJ67241AP, FischerScientific) was added, mixed by shaking, and incubated at room temperature to allow phase separation. Phase separation was aided by centrifugation at 12,000 X g for 15 minutes at 4 C and the aqueous phase was aspirated and passed through a second membrane spin column that selectively binds RNA. The membrane was washed to remove contaminants and the sample eluted in 13 ul RNase free water. M2 culture media and 2% exosome depleted fetal bovine serum was used as a negative control to determine exogenous contamination and subjected to the same extracellular vesicle and RNA isolation procedures described above. These negative controls were analyzed via particle analysis and for resulting nucleic acids to ensure there were no exogenous or contaminating exosomes or nucleic acids. None were found.

Extracellular Vesicle and RNA Quality Control

An aliquot of the extracellular vesicles from the same samples described above were isolated using the exoEasy Maxi Kit (76064, Qiagen) according to the manufacturer's protocols. The resulting EVs from two samples were analyzed using a NanoSight 300 (Malvern Panalytical) in 5 replicates to establish a size distribution of the resulting EVs and to ensure there were no contaminating particles. RNA extracted from isolated EVs was first quantified (ND-1000, ThermoScientific) and diluted for size distribution analysis and quality control using the small RNA (5067-1548, Agilent) and pico RNA kits (5067-1514, Agilent) on a BioAnalyzer 2100 (Agilent). We found there was no cellular contamination indicated by a lack of 18s and 28s ribosomal RNA and that the majority (~80%) of extracted RNA was in the 20-40 bp range, indicative of small RNA molecules.

Library Preparation, Sequencing, and Quality Control

Library preparation was performed at the University of Texas Genomic Sequencing Facility using the NEBNext Small RNA library preparation kit (E7330, New England Biolabs). Samples were prepared with 14 cycles of PCR and the final product was size selected

using a 3% gel cassette on the Blue Pippin instrument (Sage Sciences), with the parameters set to 105-165 bp. Final size selected libraries were checked for quality on the Agilent BioAnalyzer with High Sensitivity DNA analysis kit (5067-4626, Agilent) to confirm proper size selection. The Kapa Library Quantification kit for Illumina libraries (KK4602) was used to determine the loading concentrations prior to sequencing. Samples were sequenced on a NovaSeq 6000 Single Read, SR50 run, with reads counts ranging from 25 to 32 million per sample.

Analysis Pipeline

Raw RNA reads were first passed through quality control (FastQC) and checked for read quality and size distribution which showed excellent average read quality (avg. PHRED > 36) and expected read length (50 bp). The raw reads were trimmed for Illumina adaptors (QuasR, BioConductor) and again passed through quality control (FastQC) to determine per base read quality and size distribution, which again showed excellent average read quality (avg. PHRED > 36) and an expected read length distribution (~30 nt) based on electrophoresis performed before library prep. The trimmed reads were aligned (RBowtie, Bioconductor) to the unmasked rat genome (Rnor v6) with strict matching parameters to ensure reads were aligned to best-hit locations but were allowed multiple mapping locations (n = 12) due to the promiscuous nature of small RNA origins. If the alignment parameters were met but a read mapped to multiple locations, the read was randomly assigned to one of the similar best hit locations. Alignment efficiency was analyzed for cross-species contamination or PCR primer amplification artifacts and was found to be highly efficient (~85%). The aligned reads were assigned a feature annotation (Seqmonk, R, BioConductor), visualized, and analyzed for read count and length distribution. Our annotation pipeline considered reads aligned to microRNA (miRNA), piwi-interacting RNA (piRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), transfer RNA (tRNA), long non-coding RNA (lncRNA), vault RNA, and YRNA. We also obtained annotation tracks for nested and simple repeats (UCSC genome database) and microsatellites

153 (UCSC genome database). Reads were assigned an annotation designation only after considering
154 all other annotation types co-occurring in the same region.

155

156 **RESULTS**

157 **The size distribution of rat caput epididymosomes**

158 Particle tracking (NanoSight 300) was used to determine the size distribution of the
159 extracellular vesicles isolated from the caput of the rat epididymis using the Qiagen exoEasy Maxi
160 kit as described above. The isolated EVs from two separate samples were analyzed 5 times each.
161 The first sample had a mean EV size of 181.4 nm (SE 3), mode of 152 nm (SE 2.5), and standard
162 deviation of 40.6 nm (SE 1.1) and the second a mean of 178.7 (SE 0.5), mode of 156.1 (SE 3.8),
163 and standard deviation of 45.7 nm (SE 1.8). The size distribution and density of the extracellular
164 vesicles analyzed are shown in **Figure 1**. Approximately ~93% of all particles tracked were
165 between 50 and 250 nm in size, which are characteristic of epididymosomes [4].

166 **The small RNA contents of rat caput EVs**

167 The sncRNA contents of rat epididymosomes were dominated by reads that aligned to
168 tRNA (79.1%) or piRNA loci (18.1% - **Figure 2A**). We were surprised at the relatively low
169 abundance of miRNA in rats considering that mice models show substantial miRNA cargo in caput
170 epididymosomes [17]. All other annotation types, including miRNAs, accounted for the remaining
171 2.8% of the aligned reads. When tRNA and piRNA reads were removed from analysis and aligned
172 reads were analyzed as a proportion of the remaining reads (**Figure 2B**), the largest remaining
173 category belonged to reads that aligned to CpG islands (55.2%), which is not a known unit of
174 sncRNA expression (discussed below). The remaining reads aligned to YRNA (24.6%), lncRNA
175 (7.5%), miRNA (6.4%), rRNA (4.3%), and snoRNA (1.8%).

176 We analyzed the length of the reads aligned and assigned to each respective sncRNA
177 category to determine if our annotation pipeline performed well and to describe the subspecies of

various sncRNAs (**Table 1**). The two canonical categories with known read lengths were checked first. The average read length for miRNA (22 nt) was exactly as expected and showed a 90-percentile range of 19 – 27 nt. We set a cut-off of 10 reads per million (RPM) to determine which miRNA were loaded into caput EVs and identified 15 miRNAs, the majority of which are known to exist either in sperm or EVs in various models, and 5 uncharacterized miRNAs (**Table 2**).

We identified 81 piRNA loci that were substantially (> 10 RPM) expressed across the genome, with clusters on chromosome 1, 2, 10, 13, 14 and 17. Approximately 60% (49/81) of the expressed piRNA loci overlapped with known tRNA loci; these accounted for a significant portion of all piRNA annotated reads (82.3%). These are a known subclass of piRNAs termed tRNA-derived piRNA (td-piRNAs -- [36]) so we further analyzed their origin and determined that virtually all of these reads aligned with the 5' end of tRNA loci (98.98%). Finally, we categorized the tRNA type from which piRNAs were generated and determined that the majority were derived from either 5'-tRNA^{Gly} (53.3%) or 5'-tRNA^{Glu} (32.6%) loci. 5'-tRNA^{Lys} (8.5%) and 5'-tRNA^{Val} (2.3%) originating reads were also identified. The average read length for piRNA was also as expected (30 nt) with a 90-precentile range of 26 – 32 nt.

Small RNA fragments are prevalent in rat epididymosomes

We then analyzed other sncRNA categories where read lengths can be variable depending on how the RNA is processed. We found that reads aligned to tRNAs, which are ~70 – 100 nt [37], had an average read length of 30 nt and a 90-precentile range of 28 – 31 nt (**Table 1**). This result suggests that rat epididymosomes do not carry full length tRNAs, but rather carry either tRNA fragments (tRFs) or tRNA-halves, which are 14 – 30 nt and 30 – 40 nt in length respectively [38]. We further categorized tRNA reads by aligning them to either the 5' or 3' end of a given tRNA loci and discovered that 99.35% of all reads aligned to the 5' end (**Figure 3A**). Given their size distribution, we were able to further categorize these reads. The tRNA reads show a 28 – 30 nt size distribution indictive of tRF-5c fragments, the longest of the three known 5'-tRFs, or tRNA-halves [39]. Alignment of reads to the tRNA structure (tRNAvis) confirmed their identity as tRF-5c

fragments because reads did not align to the anti-codon stem loop, which would be characteristic of 5' tRNA-halves [38]. Because of the significant overlap of reads identified between piRNAs and tRNA loci (presented above), we analyzed the proportion of tRNA reads that could be accounted for by piRNA loci. We found that the majority tRNA reads (~81%) were unique to tRNAs and could not be accounted for by overlapping piRNA loci. Finally, we analyzed the amino-acid feature that tRNA reads were derived from and found that there was an over-representation of reads derived from Glycine-tRNAs (81.4%). Reads were also found to originate from glutamate (7.7%), histidine (6.9%), and lysine (1.7%) tRNA loci across the genome (**Figure 3B**). The size distribution for 3' tRNA reads (**Figure 3C**) was similar but more broad compared to 5' tRNA reads (**Figure 3D**).

We then looked at reads aligned to rRNA loci, which typically have a broad length range depending on their rRNA type of origin (5s: ~120 nt [40], 5.8s: ~ 150 nt [41], 18s: 1874 nt [42], and 28s: 4,802 nt [43]). In our samples, rRNAs had an average length of 22 nt and a 90-percentile range of 17 – 31 nt (**Table 1**). These results suggest that rat caput epididymosomes do not carry full length rRNA but instead carry rRNA fragments (rRFs). rRNA is expressed as a 45s pre-rRNA, which contains the transcripts for 18s, 5.8s, and 28s rRNA, organized as cassettes of tandem repeats on the short arms of chromosomes 3, 11, and 12 of the rat genome [44]. Reference genomes are poorly annotated for these repeating cassettes because the number of repeats often varies between individuals. Hence, these regions are typically masked in alignment reference genomes. Nonetheless, there are a number of 5s rRNAs, which are expressed separately from the 45s pre-rRNA, and 5.8s rRNA loci that are annotated in the rat reference genome that we analyzed. Approximately 80% of rRNA derived reads aligned to 5.8s loci (**Figure 4A**). Subcategories of rRFs have been described from human samples where short (18-19 nt), intermediate (24-25 nt), and long (32-33 nt) rRFs are expressed depending on the rRNA type (e.g. 5s vs 5.8s) [45]. Our data generally fit the categories described in human samples; 5s rRNA resulted in either short or long transcripts (**Figure 4B**) whereas 5.8s rRNA resulted in intermediate fragments (**Figure 4C**).

Finally, we observed that the Y RNA was present in caput EVs at levels ~4 times that of miRNA, which is surprising because we are not aware of any reports of Y RNA included in epididymal EVs. Full length Y RNAs are ~80 – 110 nt in length [46]. We identified 13 (of 28) Y RNA loci with substantial expression (> 10 RPKM). These reads demonstrated a sharp peak at 31 nt and a smaller peak at 22 nt with a 90-percentile range of 22 – 31 nt (**Table 1**). These results suggest that rat caput EVs include fragments of Y RNA as a part of their sncRNA repertoire.

Small RNA molecules are expressed from within CpG islands

A subset of reads could not be accounted for by known and annotated sncRNA. We identified dense clusters of expression that appeared to be expressed within the boundaries of GC-rich CpG islands that were not fully accounted for by any other overlapping features (**Figure 5A**). In total we found significant expression (>10 RPKM) within the boundaries of 50 CpG islands. We visually inspected each of these CpG islands and were able to rule out the majority of them due to overlapping features that had canonical expression patterns expected from features like rRNA and piRNA (**Figure 5B**) or tRNA (**Figure 5C**). These loci were removed from further analyses and the remaining 12 CpG islands were treated as an annotation category for sncRNAs. We found this category to be the third most abundant behind tRNA and piRNA reads (**Figure 2A** and **B**) and that the reads derived from CpG islands had a length distribution pattern that was distinct from all of the other analyzed categories (**Figure 2C**) with a median length of 19 nt and a 90-percentile range of 15 – 28 nt (**Table 1**). As expected, the GC content of reads aligned to CpG islands was substantial (~78%) compared to the average of all aligned reads (~50%).

We calculated a signal-to-noise ratio to determine if the reads we observed within CpG islands were due to random chance [45]. When compared to the number of reads per kilobase million (RPKM) over a 10 kb rolling window across the entire genome, our identified CpG islands had a 609 s/n ratio. When compared to all CpG islands across the entire genome, our identified CpG islands had a 92 s/n ratio. Put simply, the reads contained with our identified CpG islands were 609 times more abundant than a random 10kb window in the genome and 92 times more

abundant than an average CpG island, indicating that the reads we observed were very unlikely to be due to random chance.

Because CpG islands are not known units of small RNA expression, nor should they be capable of such expression, we then tried to systematically demonstrate that the observed reads were due to overlying features. We observed that CpG islands that contained more piRNA loci appeared to be associated with more reads (**Figure 6A**), but when normalized for the read depth and length of the CpG island (RPKM) the relationship was virtually non-existent ($R^2 = 0.077$, **Table 3**). We quantified the reads that aligned under known piRNA loci and found that ~48% of the total reads within the boundaries of CpG islands were also associated with piRNA loci. However, we are hesitant to classify these reads as piRNA because they do not follow the canonical length distribution associated with piRNA (~30nt) identified here and elsewhere [47]. None of the reads (0.0%) within CpG islands were expressed from known tRNA loci. In a subset of the CpG islands we observed, there was an abundance of predicted (eponine) transcriptional start sites (**Figure 6B**), but this was not always the case (e.g. **Figure 5A**).

Finally, half (6/12) of the identified CpG islands were associated with annotated 5.8s rRNA loci (**Table 3**). Functional rRNAs are expressed in tandem repeating cassettes that are processed from a 45s precursor transcript that contains the functional 5.8s, 18s, and 28s rRNA after processing. The rat genome is known to express these repeats from the short arms of chromosomes 3, 11, and 12 [44] on which none of our identified CpG islands reside. Notwithstanding, the rat genome is poorly annotated for rRNA, so to overcome this shortcoming in the available annotations we used BLAST to align the raw sequences expressed from CpG islands to the available rRNA sequences: 45s, 32s (which includes the 28s rRNA and 5' sequence between the transcription start site and 28s), 28s, 18s, and 5.8s. Approximately one third (~34.3%) of all reads derived from CpG islands aligned to some form of rRNA or precursor rRNA. Of those reads that aligned to any form of rRNA, the vast majority were derived from the 28s sequence (86.1% -- 29.6% of all CpG island reads), while the 18s (11.15% -- 3.9% of all CpG island reads), and 5.8s (0.16% -- 0.05% of all CpG island reads) accounted for a small proportion of the reads. The

283 remaining reads (2.6% -- 0.74% of all reads) aligned to the external or internal transcribed spacers
284 within the full length rRNA precursor transcript. In summary, while it appears that the CpG islands
285 we identified as expressing 19 nt small RNAs were in the vicinity of heretofore unannotated rRNA
286 or piRNA loci, the reads observed expressed within CpG islands cannot be fully accounted for by
287 either of these designations nor are their expression profiles congruent.

288 **DISCUSSION**

289 Epididymosomes from the caput are an essential component of sperm maturation and
290 acquisition of function [5,48,49]. They are also implicated in the control of heritable non-genetic
291 phenotypes, as their sncRNA cargo is altered by stress [19,22,25], diet [23,26,50], and alcohol
292 consumption [27]. Here, we provide the first comprehensive characterization of small RNAs
293 derived from caput EVs in the rat. We show that EVs isolated from the caput epididymis match
294 the size range of epididymosomes and that their small RNA contents are dominated by tRFs and
295 piRNA and contain far fewer miRNAs than expected from other organisms (Mice - [25,30,31],
296 Humans - [51]). We also identify Y RNA fragments in caput EVs for the first time in any organism,
297 although they are expressed in EVs of other organs [52–54]. Finally, we identify a potentially
298 novel small RNA molecule that is expressed from GC-rich CpG islands that cannot be accounted
299 for by known overlying small RNA features, and have a unique size distribution that is distinct
300 from other small RNAs analyzed.

301 **The sncRNA composition of rat caput EVs differs from that in mice**

302 The small RNA contents of EVs from the mouse caput epididymis are primarily miRNAs
303 (~60% - [25,30,31]), with over 350 expressed [17]. This is in contrast to the rat in which miRNAs
304 were a small portion (0.18%) of the small RNA complement. Of these, we identified 15 miRNAs,
305 7 of which overlapped with those in mice (miR-143, let-7c, let-7i, miR-26a, miR-99a, miR-143,
306 miR-148 [17]). The most abundant miRNA we identified in rats, miR-184, did not meet the
307 threshold for abundance in mice [17]. MiR-143, which is a hallmark of EVs in mice [17] and
308 humans [30], was also expressed in rats. There were also substantial species differences in tRFs,

which are the second most abundant category (~30%) in mouse caput EVs while piRNAs are present at very low levels (<0.05% - [30,31]). Here, we show in rats that tRFs are by far the most abundant category (~79%) and piRNAs are second most abundant (~18%). Further work is needed to validate these findings and to understand why there are these dramatic species differences in the EV sncRNA cargo.

tRFs dominate the cargo of rat caput EVs

tRFs, particularly from 5'-tRNA^{Gly} loci, dominated the cargo of caput EVs in the rat. As sperm exit the testis, they contain few tRFs but their abundance gradually increases as sperm transit from the proximal caput to the cauda [19,55]. In a mouse model, paternal diet changed the abundance of 5'-tRF^{Gly} in sperm, which is implicated in the repression of endogenous retroelements in pre-implantation embryos [19]. The authors suggested that EVs could deliver these fragments, and we provide evidence that they are well poised to do so, and we confirmed the exact species they identified (5'-tRF^{Gly}). We also identify 5'-tRF^{Glu} and 5'-tRF^{His} as prevalent in rat caput EVs. The former (5'-tRF^{Glu}) has been implicated in playing a role in a high fat diet intergenerational phenotype [26] via transcriptional regulation, directing alternate splicing, and acetylation and phosphoprotein activation [27]. The latter (5'-tRF^{His}) has been implicated in responses to low protein diets [19]. Further work is needed to elucidate their functional roles in our rat model.

We found that tRFs in caput EVs are almost exclusively derived from the 5' end of tRNA loci. There are two possible explanations for the overrepresentation of 5' tRFs. First, rat caput epididymosomes may be selectively loaded with 5'-tRF fragments; these have actions similar to RNA interference as governed by miRNAs, which act through the Argonaute pathway [56] to silence endogenous retroelements in embryonic stem cells and embryos [19]. Alternatively, our observation may be due to a technical bias in the sequencing library preparation procedure [38]. Libraries are amplified with PCR during preparation, which can be prematurely aborted if an RNA molecule contains modified tRNA nucleobases that would be too short for sequencing, thereby

excluded, and not detected during analysis. In order to confirm the dominating presence of 5'-tRFs in rat caput epididymosomes, a specific analysis pipeline (streamlined platform for observing tRNA – SPOt) would need to be used to prevent observation bias [57].

piRNAs are abundant and coincide with 5' tRNA fragments

piRNAs represented a substantial proportion (18.1%) of EVs in the rat caput epididymis. This is contrary to what is found in mouse models (< 0.05% -- [30,31]) but not entirely unexpected as rat pachytene sperm are densely populated with piRNAs [58] depending on the stage of development [59,60] presumably to control the expression of transposable elements in the germline [61] by directing the catalyzation of de novo DNA methylation to suppress their expression [62]. In the rat, unlike in the mouse, piRNAs appear poised to be delivered to sperm via EVs in the caput epididymis. The reasons for this discrepancy between rats and mice could be due to a number of reasons. First, this may represent a fundamental difference in the reproductive biology between the two species. Second, it may be because the piRNAs that do exist in mouse caput EVs are typically categorized as tRNA-halves instead of piRNA. We found that a majority of reads that aligned with piRNA loci (~80%) also align to the 5' end of tRNA loci, which could explain why they are missed during analysis in the mouse. The overlap of these two features (piRNAs and tRFs) and expression from these loci has precedence. tRNA derived-piRNA (td-piRNA) have been described in the testes of marmosets where Piwi proteins were found to bind reads mapping to the 5'-tRNA^{Glu}, 5'-tRNA^{Gly}, and 5'-tRNA^{Val} loci [63]. It is intriguing that these three identified td-piRNAs account for three of the four td-piRNAs found here and we believe we are the first to identify these td-piRNA in a position poised to be delivered to sperm in the epididymis. Additionally, we append 5'-tRNA^{Lys}, the third most abundant td-piRNA in our analysis, as a potentially significant td-piRNA.

rRNA fragments are present in three distinct sizes

rRFs have long been considered RNA degradation or apoptotic by-products and have been disregarded for any functional value. Compared to miRNA, piRNA, and tRFs, little is known about

the function or categories of rRFs and they have only recently been described [45] with nearly no information existing in rats that we could find. Interestingly, rRFs appear in immunoprecipitations with the Argonaute complex in mice and humans, which suggests a role for translational regulation [64]. In humans, three distinct categories of rRFs are known with differential expression in each of the rRNA categories (e.g. 5s, 5.8s, 18s, and 28s); short (18-21 nt), intermediate (24-25 nt), and long (26-33 nt) [45]. Our ability to map rRFs to the reference genome was limited by the lack of annotations for rRNA in the rat genome, but given the annotations we did have, we identified a similar expression profile to Cherlin et al. (2020) (in human samples -- [45]) in which three distinct subtypes of rRFs were present: 18 nt & 30 nt rRFs originating from 5s loci and 22 nt rRFs from 5.8s loci. Furthermore, rRFs are sparsely reported in reproductive tissue (Humans -- [65,66], Bovine -- [67]) and we believe we are the first to report their presence in caput EV samples, which suggests a novel role for rRFs in reproduction.

19-nt small RNAs are expressed from within the boundaries of CpG Islands

There were 12 identified loci in the rat genome that corresponded with substantial expression of 19-nt small RNAs that are GC rich and expressed within the boundaries of annotated CpG islands. We postulate that this is a novel unit of expression for sncRNAs. CpG islands are generally annotated based on the characteristics of their sequence; longer than 200-bp with a GC content higher than 50% and an expected to observed ratio greater than 0.6 [68]. Many of the first CpG islands were identified at the 5' end of "housekeeping" genes [69,70] but have since been computationally and experimentally predicted across the genome, including in inter- and intragenic space [71]. Parts of CpG islands are sometimes transcribed on the 5' or 3' end of expressed genes and removed during splicing unless they extend into an exon, including sncRNAs [71,72]. This expression pattern we observed within CpG islands does not appear to have a relationship with any other annotation type; although we considered other nearby or overlying rRNA cassettes or piRNA clusters, neither accounted for all of the reads or loci, and some loci had neither feature. Nevertheless, we carefully inspected the involvement of both types of features.

For rRNAs, it is difficult to determine the precise location of cassettes because they are not well annotated in the rat genome. By using sequence similarity via BLAST alignment, we determined that ~34% of the small RNA expressed from CpG islands are found within the 45s pre-rRNA transcript, the majority of those from the 28s sequence (~86%). Because of their 19 nt length, we can say with certainty these reads are not from mature rRNAs. Furthermore, we are hesitant to term these rRNA fragments for two reasons; first, the length of the reads (19 nt) from annotated 5.8s rRNA within the identified CpG islands do not correspond to other 5.8s rRNA reads (22 nt) found elsewhere in this data set; and second, the expression profile of the 19 nt reads are bounded by the CpG island and not by the 5.8s rRNA or predicted position of 45s rRNA. It is possible that the observed CpG islands may be a part of the ~30 kb non-transcribed spacer that separates 45s repeats, but as the name implies, these portions are up or downstream of rRNA and should not be transcribed.

A second possibility is that the reads we identified arise from piRNA clusters. The function of piRNA is RNA-guided silencing mediated by Piwi proteins (Riwi in the rat) that are particularly active in the germline of the testes [73,74]. piRNAs regulate the expression of transposable elements [59,75] by inducing de novo DNA methylation [76]. Primary piRNA can be amplified by "ping-pong amplification" [77] by binding to expressed transposable elements which are cleaved to generate a secondary guide transcript that ensure the lack of transposition [59,78]. Primary piRNA are observed in developing germ cells [76] while secondary piRNAs, generated by the ping-pong cycle, are thought only to be present before pachytene stages in the basal compartment of the testes [59,76] after which only primary piRNAs should be present [76,79]. The transcripts we identified in caput EVs are well outside of the window of canonical piRNA functionality and while they appear to be associated with piRNA clusters, when read depth is corrected for by the length of a CpG island the association becomes non-existent. Furthermore, their distinct 19-nt signature differentiates them from canonical piRNAs which are typically 26-35 nt in length [80].

A final possibility we considered is that the 19-nt sRNA we observe are byproducts from the generation of secondary piRNAs during which exonuclease activity should generate fragments of a consistent size due to the portion of the transcript that is shielded by Piwi-complex itself [76]. There have been three reports we could find that have identified such byproducts [81–83]. Berninger et al. (2011) were the first to identify "19mers" expressed from both the sense and anti-sense strands of piRNA clusters solely in the testis of rats, mice, and platypuses, but did not find them in any other tissue type [81]. Oey et al. (2011) also described 19mer byproducts of piRNA with a focus on repeat elements (LINEs, SINEs, LTRs) and IAPs where ping-pong amplification is predicted [82]. They also found that 19mers that were equally present from both the sense and antisense strands, were exclusive to the testes, and while they didn't specifically analyze or mention it, their data show no GC content bias. Ichiyanagi et al. (2014) also observed 19mers immediately upstream of piRNAs but dismissed them as piRNA amplification byproducts as per Berninger et al. and Oey et al. (2011) [83]. The commonality of the piRNA amplification byproducts identified in these reports is that they are 19nt fragments expressed from a broad distribution on both strands relative to the piRNA cluster, a profile that does not fit our current observation.

Based on this evidence, we do not believe that the alternatives (rRNA fragments, piRNAs, or piRNA amplification byproducts) can account for our observation of GC-rich 19-nt small RNAs. These "CpG island small RNAs" extend outside of the boundaries of annotated rRNA and piRNA, which account for only a portion of the observations, and do not abide by their respective canonical length characteristics. It is also possible that these observations could be the 19mer byproduct of piRNA amplification observed elsewhere [81–83] but there are 3 reasons we do not think this is likely. First, we observe these reads in extracellular vesicles while the

byproducts of piRNA amplification are known only to exist in pre-pachytene spermatocytes and should be derived from both the sense and anti-sense strands, a characteristic we do not see. If this observation is due to 19mer piRNA amplification byproducts generated in the epididymis, then our data provide functional significance to their existence not previously reported. Second, the reads we observe are expressly within the boundaries of CpG islands with a GC content not reported elsewhere. The specificity of the expression loci, and the lack of high GC content in other reports make their identity as byproducts unlikely. Finally, we observe expression from multiple CpG islands not linked with known piRNA loci. If these are indeed the byproduct of piRNA amplification, our data then represent the identification of novel piRNA loci that would seem to be important for the final steps of sperm maturation. Mature sperm transiting the epididymis should be transcriptionally quiescent [84–87], and therefore transposons should not be expressed. Ascertaining a functional role of these small RNAs requires further investigation.

Conclusions

EVs from the rat caput epididymis carry a complex repertoire of small RNAs and their contents are substantially different from the mouse. The dominant features we identify are tRNA fragments and piRNAs derived from tRNA loci. MicroRNAs are poorly represented in stark contrast to mice. We also identify two types of small RNA not previously seen in caput EVs, rRNA fragments and Y RNA fragments, and described a potentially novel small RNA we have termed CpG island (CpGi) sRNAs. These data represent an exciting collection of possibilities for future researchers studying basic reproductive biology and the complexities of epigenetic transgenerational inheritance.

Conflict of Interest: The authors declare no conflict of interest.

Grant funding: Supported by and RO1 ES029464 to A.C.G. and PhRMA Foundation Postdoctoral Fellowship to R.G.

Acknowledgement: The authors recognize Mandee Bell and Lindsay M. Thompson for their assistance with animal husbandry, treatment, and sample collection. We would also like to acknowledge the excellent work (small RNA library preparation and sequencing) performed by the Genomic Sequencing and Analysis Facility at UT Austin, Center for Biomedical Research Support. RRID#: SCR_021713.

REFERENCES

1. Hermo, L.; Pelletier, R. ‐ Marc; Cyr, D.G.; Smith, C.E. Surfing the Wave, Cycle, Life History, and Genes/Proteins Expressed by Testicular Germ Cells. Part 1: Background to Spermatogenesis, Spermatogonia, and Spermatocytes. *Microsc Res Techniq* 2010, *73*, 241–278, doi:10.1002/jemt.20783.

2. Cornwall, G.A. New Insights into Epididymal Biology and Function. *Hum Reprod Update* 2009, *15*, 213–227, doi:10.1093/humupd/dmn055.

3. Sullivan, R.; Saez, F. Epididymosomes, Prostasomes, and Liposomes: Their Roles in Mammalian Male Reproductive Physiology. *Reproduction* 2001, *146*, R21–R35, doi:10.1530/rep-13-0058.

4. Sullivan, R. Epididymosomes: A Heterogeneous Population of Microvesicles with Multiple Functions in Sperm Maturation and Storage. *Asian J Androl* 2015, *0*, 0, doi:10.4103/1008-682x.155255.

5. Martin-DeLeon, P.A. Epididymosomes: Transfer of Fertility-Modulating Proteins to the Sperm Surface. *Asian J Androl* 2015, *17*, 720–725, doi:10.4103/1008-682x.155538.

6. Martin-DeLeon, P.A. Epididymal SPAM1 and Its Impact on Sperm Function. *Mol Cell Endocrinol* 2006, *250*, 114–121, doi:10.1016/j.mce.2005.12.033.

7. Fornés, M.W.; Barbieri, A.; Cavicchia, J.C. Morphological and Enzymatic Study of Membrane ‐ bound Vesicles from the Lumen of the Rat Epididymis. *Andrologia* 1995, *27*, 1–5, doi:10.1111/j.1439-0272.1995.tb02087.x.

8. Rejraji, H.; Sion, B.; Prensier, G.; Carreras, M.; Motta, C.; Frenoux, J.-M.; Vericel, E.; Grizard, G.; Vernet, P.; Drevet, J.R. Lipid Remodeling of Murine Epididymosomes and Spermatozoa During Epididymal Maturation. *Biol Reprod* 2006, *74*, 1104–1113, doi:10.1095/biolreprod.105.049304.

9. Grimalt, P.; Bertini, F.; Fornes, M.W. HIGH-AFFINITY SITES FOR Beta-D-GALACTOSIDASE ON MEMBRANE-BOUND VESICLES ISOLATED FROM RAT EPIDIDYMAL FLUID. *Arch Andrology* 2009, *44*, 85–91, doi:10.1080/014850100262245.

10. Hermo, L.; Jacks, D. Nature's Ingenuity: Bypassing the Classical Secretory Route via Apocrine Secretion. *Mol Reprod Dev* 2002, *63*, 394–410, doi:10.1002/mrd.90023.

11. Oh, J.S.; Han, C.; Cho, C. ADAM7 Is Associated with Epididymosomes and Integrated into Sperm Plasma Membrane. *Mol Cells* 2009, *28*, 441–446, doi:10.1007/s10059-009-0140-x.

12. Zhou, W.; Iuliis, G.N.D.; Dun, M.D.; Nixon, B. Characteristics of the Epididymal Luminal Environment Responsible for Sperm Maturation and Storage. *Front Endocrinol* 2018, *9*, 59, doi:10.3389/fendo.2018.00059.

13. Frenette, G.; Lessard, C.; Sullivan, R. Polyol Pathway along the Bovine Epididymis. *Mol Reprod Dev* 2004, *69*, 448–456, doi:10.1002/mrd.20170.

14. Sutovsky, P.; Moreno, R.; Ramalho-Santos, J.; Dominko, T.; Thompson, W.E.; Schatten, G. A Putative, Ubiquitin-Dependent Mechanism for the Recognition and Elimination of Defective Spermatozoa in the Mammalian Epididymis. *J Cell Sci* 2001, *114*, 1665–1675, doi:10.1242/jcs.114.9.1665.

15. D'Amours, O.; Frenette, G.; Bordeleau, L.-J.; Allard, N.; Leclerc, P.; Blondin, P.; Sullivan, R. Epididymosomes Transfer Epididymal Sperm Binding Protein 1 (ELSPBP1) to Dead Spermatozoa During Epididymal Transit in Bovine1. *Biol Reprod* 2012, *87*, Article 94, 1-11, doi:10.1095/biolreprod.112.100990.

16. Belleanne, C.; Calvo, zequiel; Caballero, J.; Sullivan, R. Epididymosomes Convey Different Repertoires of MicroRNAs Throughout the Bovine Epididymis1. *Biol Reprod* 2013, *89*, Article 30, 1-11, doi:10.1095/biolreprod.113.110486.

17. Reilly, J.N.; McLaughlin, E.A.; Stanger, S.J.; Anderson, A.L.; Hutcheon, K.; Church, K.; Mihalas, B.P.; Tyagi, S.; Holt, J.E.; Eamens, A.L.; et al. Characterisation of Mouse Epididymosomes Reveals a Complex Profile of MicroRNAs and a Potential Mechanism for Modification of the Sperm Epigenome. *Sci Rep-uk* 2016, *6*, 31794, doi:10.1038/srep31794.

18. Caballero, J.N.; Frenette, G.; Belleannée, C.; Sullivan, R. CD9-Positive Microvesicles Mediate the Transfer of Molecules to Bovine Spermatozoa during Epididymal Maturation. *Plos One* 2013, *8*, e65364, doi:10.1371/journal.pone.0065364.

19. Sharma, U.; Conine, C.C.; Shea, J.M.; Boskovic, A.; Derr, A.G.; Bing, X.Y.; Belleannee, C.; Kucukural, A.; Serra, R.W.; Sun, F.; et al. Biogenesis and Function of TRNA Fragments during Sperm Maturation and Fertilization in Mammals. *Science* 2016, *351*, 391, doi:10.1126/science.aad6780.

20. Conine, C.C.; Sun, F.; Song, L.; Rivera-Pérez, J.A.; Rando, O.J. Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice. *Dev Cell* 2018, *46*, 470-480.e3, doi:10.1016/j.devcel.2018.06.024.

21. Suryawanshi, A.R.; Khan, S.A.; Joshi, C.S.; Khole, V.V. Epididymosome‐Mediated Acquisition of MMSDH, an Androgen‐Dependent and Developmentally Regulated Epididymal Sperm Protein. *J Androl* 2012, *33*, 963–974, doi:10.2164/jandrol.111.014753.

22. Rodgers, A.B.; Morgan, C.P.; Bronson, S.L.; Revello, S.; Bale, T.L. Paternal Stress Exposure Alters Sperm MicroRNA Content and Reprograms Offspring HPA Stress Axis Regulation. *J Neurosci* 2013, *33*, 9003--9012, doi:10.1523/jneurosci.0914-13.2013.

530 23. Chen, Q.; Yan, M.; Cao, Z.; Li, X.; Zhang, Y.; Shi, J.; Feng, G.; Peng, H.; Zhang, X.; Zhang,
531 Y.; et al. Sperm TsRNAs Contribute to Intergenerational Inheritance of an Acquired Metabolic
532 Disorder. *Science* 2016, *351*, 397, doi:10.1126/science.aad7977.

533 24. Fullston, T.; Teague, E.M.C.O.; Palmer, N.O.; DeBlasio, M.J.; Mitchell, M.; Corbett, M.;
534 Print, C.G.; Owens, J.A.; Lane, M. Paternal Obesity Initiates Metabolic Disturbances in Two
535 Generations of Mice with Incomplete Penetrance to the F2 Generation and Alters the
536 Transcriptional Profile of Testis and Sperm MicroRNA Content. *Faseb J* 2013, *27*, 4226–4243,
537 doi:10.1096/fj.12-224048.

538 25. Short, A.; Fennell, K.; Perreau, V.; Fox, A.; O'Bryan, M.; Kim, J.H.; Bredy, T.; Pang, T.;
539 Hannan, A. Elevated Paternal Glucocorticoid Exposure Alters the Small Noncoding RNA Profile
540 in Sperm and Modifies Anxiety and Depressive Phenotypes in the Offspring. *Translational*
541 *Psychiatry* 2016, *6*.

542 26. Barbosa, T. de C.; Ingerslev, L.R.; Alm, P.S.; Versteyhe, S.; Massart, J.; Rasmussen, M.;
543 Donkin, I.; Sjögren, R.; Mudry, J.M.; Vetterli, L.; et al. High-Fat Diet Reprograms the
544 Epigenome of Rat Spermatozoa and Transgenerationally Affects Metabolism of the Offspring.
545 *Mol Metab* 2015, *5*, 184–197, doi:10.1016/j.molmet.2015.12.002.

546 27. Rompala, G.R.; Mounier, A.; Wolfe, C.M.; Lin, Q.; Lefterov, I.; Homanics, G.E. Heavy
547 Chronic Intermittent Ethanol Exposure Alters Small Noncoding RNAs in Mouse Sperm and
548 Epididymosomes. *Frontiers Genetics* 2018, *9*, 32–32, doi:10.3389/fgene.2018.00032.

549 28. Gapp, K.; Jawaid, A.; Sarkies, P.; Bohacek, J.; Pelczar, P.; Prados, J.; Farinelli, L.; Miska, E.;
550 Mansuy, I.M. Implication of Sperm RNAs in Transgenerational Inheritance of the Effects of
551 Early Trauma in Mice. *Nat Neurosci* 2014, *17*, 667–669, doi:10.1038/nn.3695.

552 29. Chan, J.C.; Morgan, C.P.; Leu, N.A.; Shetty, A.; Cisse, Y.M.; Nugent, B.M.; Morrison, K.E.;
553 Jašarević, E.; Huang, W.; Kanyuch, N.; et al. Reproductive Tract Extracellular Vesicles Are
554 Sufficient to Transmit Intergenerational Stress and Program Neurodevelopment. *Nat Commun*
555 2020, *11*, 1499, doi:10.1038/s41467-020-15305-w.

556 30. Trigg, N.A.; Eamens, A.L.; Nixon, B. The Contribution of Epididymosomes to the Sperm
557 Small RNA Profile. *Reproduction* 2019, *1*, R209–R223, doi:10.1530/rep-18-0480.

558 31. Hutcheon, K.; McLaughlin, E.A.; Stanger, S.J.; Bernstein, I.R.; Dun, M.D.; Eamens, A.L.;
559 Nixon, B. Analysis of the Small Non-Protein-Coding RNA Profile of Mouse Spermatozoa
560 Reveals Specific Enrichment of PiRNAs within Mature Spermatozoa. *Rna Biol* 2017, *14*, 1776–
561 1790, doi:10.1080/15476286.2017.1356569.

562 32. Wei, H.; Zhou, B.; Zhang, F.; Tu, Y.; Hu, Y.; Zhang, B.; Zhai, Q. Profiling and Identification
563 of Small RDNA-Derived RNAs and Their Potential Biological Functions. *Plos One* 2013, *8*,
564 e56842, doi:10.1371/journal.pone.0056842.

565 33. Wei, Y.; Schatten, H.; Sun, Q.-Y. Environmental Epigenetic Inheritance through Gametes
566 and Implications for Human Reproduction. *Hum Reprod Update* 2015, *21*, 194–208,
567 doi:10.1093/humupd/dmu061.

568 34. Ellenbroek, B.; Youn, J. Rodent Models in Neuroscience Research: Is It a Rat Race? *Dis*
569 *Model Mech* 2016, *9*, 1079–1087, doi:10.1242/dmm.026120.

570 35. Burns, S.B.; Szyszkowicz, J.K.; Luheshi, G.N.; Lutz, P.-E.; Turecki, G. Plasticity of the
571 Epigenome during Early-Life Stress. *Semin Cell Dev Biol* 2018, *77*, 115–132,
572 doi:10.1016/j.semcdb.2017.09.033.

573 36. Honda, S.; Kawamura, T.; Loher, P.; Morichika, K.; Rigoutsos, I.; Kirino, Y. The Biogenesis
574 Pathway of TRNA-Derived PiRNAs in Bombyx Germ Cells. *Nucleic Acids Res* 2017, *45*, 9108–
575 9120, doi:10.1093/nar/gkx537.

576 37. Krahn, N.; Fischer, J.T.; Söll, D. Naturally Occurring TRNAs With Non-Canonical
577 Structures. *Front Microbiol* 2020, *11*, 596914, doi:10.3389/fmicb.2020.596914.

578 38. Shen, Y.; Yu, X.; Zhu, L.; Li, T.; Yan, Z.; Guo, J. Transfer RNA-Derived Fragments and
579 TRNA Halves: Biogenesis, Biological Functions and Their Roles in Diseases. *J Mol Med* 2018,
580 *96*, 1167–1176, doi:10.1007/s00109-018-1693-y.

581 39. Kumar, P.; Anaya, J.; Mudunuri, S.B.; Dutta, A. Meta-Analysis of TRNA Derived RNA
582 Fragments Reveals That They Are Evolutionarily Conserved and Associate with AGO Proteins
583 to Recognize Specific RNA Targets. *Bmc Biol* 2014, *12*, 78, doi:10.1186/s12915-014-0078-0.

584 40. Gray, M.W.; Burger, G.; Lang, B.F. Mitochondrial Evolution. *Science* 1999, *283*, 1476–
585 1481, doi:10.1126/science.283.5407.1476.

586 41. Walker, T.A.; Pace, N.R. 5.8S Ribosomal RNA. *Cell* 1983, *33*, 320–322, doi:10.1016/0092-
587 8674(83)90413-0.

588 42. Chan, Y.L.; Gutell, R.; Noller, H.F.; Wool, I.G. The Nucleotide Sequence of a Rat 18 S
589 Ribosomal Ribonucleic Acid Gene and a Proposal for the Secondary Structure of 18 S
590 Ribosomal Ribonucleic Acid. *J Biol Chem* 1984, *259*, 224–230, doi:10.1016/s0021-
591 9258(17)43645-3.

592 43. Hadjiolov, A.A.; Georgiev, O.I.; Nosikov, V.V.; Yavachev, L.P. Primary and Secondary
593 Structure of Rat 28 S Ribosomal RNA. *Nucleic Acids Res* 1984, *12*, 3677–3693,
594 doi:10.1093/nar/12.8.3677.

595 44. Sasaki, M.; Nishida, C.; Kodama, Y. Characterization of Silver-Stained Nucleolus Organizer
596 Regions (Ag-NORs) in 16 Inbred Strains of the Norway Rat, Rattus Norvegicus. *Cytogenet*
597 *Genome Res* 1986, *41*, 83–88, doi:10.1159/000132208.

598 45. Cherlin, T.; Magee, R.; Jing, Y.; Pliatsika, V.; Loher, P.; Rigoutsos, I. Ribosomal RNA
599 Fragmentation into Short RNAs (RRFs) Is Modulated in a Sex- and Population of Origin-
600 Specific Manner. *Bmc Biol* 2020, *18*, 38, doi:10.1186/s12915-020-0763-0.

601 46. Wolin, S.L.; Steitz, J.A. Genes for Two Small Cytoplasmic Ro RNAs Are Adjacent and
602 Appear to Be Single-Copy in the Human Genome. *Cell* 1983, *32*, 735–744, doi:10.1016/0092-
603 8674(83)90059-4.

604 47. Beyret, E.; Liu, N.; Lin, H. PiRNA Biogenesis during Adult Spermatogenesis in Mice Is
605 Independent of the Ping-Pong Mechanism. *Cell Res* 2012, *22*, 1429–1439,
606 doi:10.1038/cr.2012.120.

607 48. Sullivan, R.; Frenette, G.; Girouard, J. Epididymosomes Are Involved in the Acquisition of
608 New Sperm Proteins during Epididymal Transit. *Asian J Androl* 2007, *9*, 483–491,
609 doi:10.1111/j.1745-7262.2007.00281.x.

610 49. Caballero, J.; Frenette, G.; D'Amours, O.; Belleannée, C.; Lacroix‐Pepin, N.; Robert, C.;
611 Sullivan, R. Bovine Sperm Raft Membrane Associated Glioma Pathogenesis‐Related 1‐like
612 Protein 1 (GliPr1L1) Is Modified during the Epididymal Transit and Is Potentially Involved in
613 Sperm Binding to the Zona Pellucida. *J Cell Physiol* 2012, *227*, 3876–3886,
614 doi:10.1002/jcp.24099.

615 50. Fullston, T.; Teague, E.M.C.O.; Palmer, N.O.; DeBlasio, M.J.; Mitchell, M.; Corbett, M.;
616 Print, C.G.; Owens, J.A.; Lane, M. Paternal Obesity Initiates Metabolic Disturbances in Two
617 Generations of Mice with Incomplete Penetrance to the F2 Generation and Alters the
618 Transcriptional Profile of Testis and Sperm MicroRNA Content. *The FASEB Journal* 2013, *27*,
619 4226–4243.

620 51. Belleannée, C.; Calvo, E.; Thimon, V.; Cyr, D.G.; Légaré, C.; Garneau, L.; Sullivan, R. Role
621 of MicroRNAs in Controlling Gene Expression in Different Segments of the Human Epididymis.
622 *Plos One* 2012, *7*, e34996, doi:10.1371/journal.pone.0034996.

623 52. Kowalski, M.P.; Krude, T. Functional Roles of Non-Coding Y RNAs. *Int J Biochem Cell*
624 *Biology* 2015, *66*, 20–29, doi:10.1016/j.biocel.2015.07.003.

625 53. Driedonks, T.A.P.; Hoen, E.N.M.N.-'t Circulating Y-RNAs in Extracellular Vesicles and
626 Ribonucleoprotein Complexes; Implications for the Immune System. *Front Immunol* 2019, *9*,
627 3164, doi:10.3389/fimmu.2018.03164.

628 54. Driedonks, T.A.P.; Mol, S.; Bruin, S. de; Peters, A.-L.; Zhang, X.; Lindenbergh, M.F.S.;
629 Beuger, B.M.; Stalborch, A.-M.D. van; Spaan, T.; Jong, E.C. de; et al. Y-RNA Subtype Ratios in
630 Plasma Extracellular Vesicles Are Cell Type- Specific and Are Candidate Biomarkers for
631 Inflammatory Diseases. *J Extracell Vesicles* 2020, *9*, 1764213,
632 doi:10.1080/20013078.2020.1764213.

633 55. Stanger, S.J.; Bernstein, I.R.; Anderson, A.L.; Hutcheon, K.; Dun, M.D.; Eamens, A.L.;
634 Nixon, B. The Abundance of a Transfer RNA-Derived RNA Fragment Small RNA
635 Subpopulation Is Enriched in Cauda Spermatozoa. *Exrna* 2020, *2*, 17, doi:10.1186/s41544-020-
636 00058-x.

637 56. Gagnon, K.T.; Corey, D.R. Argonaute and the Nuclear RNAs: New Pathways for RNA-
638 Mediated Control of Gene Expression. *Nucleic Acid Ther* 2012, *22*, 3–16,
639 doi:10.1089/nat.2011.0330.

640 57. Grelet, S.; McShane, A.; Hok, E.; Tomberlin, J.; Howe, P.H.; Geslain, R. SPOt: A Novel and
641 Streamlined Microarray Platform for Observing Cellular TRNA Levels. *Plos One* 2017, *12*,
642 e0177939, doi:10.1371/journal.pone.0177939.

643 58. Stermer, A.R.; Reyes, G.; Hall, S.J.; Boekelheide, K. Small RNAs in Rat Sperm Are a
644 Predictive and Sensitive Biomarker of Exposure to the Testicular Toxicant Ethylene Glycol
645 Monomethyl Ether. *Toxicol Sci* 2019, *169*, 399–408, doi:10.1093/toxsci/kfz041.

646 59. Aravin, A.A.; Sachidanandam, R.; Girard, A.; Fejes-Toth, K.; Hannon, G.J. Developmentally
647 Regulated PiRNA Clusters Implicate MILI in Transposon Control. *Science* 2007, *316*, 744–747,
648 doi:10.1126/science.1142612.

649 60. Smorag, L.; Zheng, Y.; Nolte, J.; Zechner, U.; Engel, W.; Pantakani, D.V.K. MicroRNA
650 Signature in Various Cell Types of Mouse Spermatogenesis: Evidence for Stage‐specifically
651 Expressed MiRNA‐221, ‐203 and ‐34b‐5p Mediated Spermatogenesis Regulation. *Biol*
652 *Cell* 2012, *104*, 677–692, doi:10.1111/boc.201200014.

653 61. Newkirk, S.J.; Lee, S.; Grandi, F.C.; Gaysinskaya, V.; Rosser, J.M.; Berg, N.V.; Hogarth,
654 C.A.; Marchetto, M.C.N.; Muotri, A.R.; Griswold, M.D.; et al. Intact PiRNA Pathway Prevents
655 L1 Mobilization in Male Meiosis. *Proc National Acad Sci* 2017, *114*, E5635–E5644,
656 doi:10.1073/pnas.1701069114.

657 62. Peng, J.C.; Lin, H. Beyond Transposons: The Epigenetic and Somatic Functions of the Piwi-
658 PiRNA Mechanism. *Curr Opin Cell Biol* 2013, *25*, 190–194, doi:10.1016/j.ceb.2013.01.010.

659 63. Hirano, T.; Iwasaki, Y.W.; Lin, Z.Y.-C.; Imamura, M.; Seki, N.M.; Sasaki, E.; Saito, K.;
660 Okano, H.; Siomi, M.C.; Siomi, H. Small RNA Profiling and Characterization of PiRNA
661 Clusters in the Adult Testes of the Common Marmoset, a Model Primate. *Rna* 2014, *20*, 1223–
662 1237, doi:10.1261/rna.045310.114.

663 64. Guan, L.; Grigoriev, A. Computational Meta-Analysis of Ribosomal RNA Fragments:
664 Potential Targets and Interaction Mechanisms. *Nucleic Acids Res* 2021, *49*, gkab190-,
665 doi:10.1093/nar/gkab190.

666 65. Goodrich, R.J.; Anton, E.; Krawetz, S.A. Spermatogenesis, Methods and Protocols. *Methods*
667 *Mol Biology* 2012, 385–396, doi:10.1007/978-1-62703-038-0_33.

668 66. Johnson, G.D.; Sendler, E.; Lalancette, C.; Hauser, R.; Diamond, M.P.; Krawetz, S.A.
669 Cleavage of RRNA Ensures Translational Cessation in Sperm at Fertilization. *Mhr Basic Sci*
670 *Reproductive Medicine* 2011, *17*, 721–726, doi:10.1093/molehr/gar054.

671 67. Almiñana, C.; Tsikis, G.; Labas, V.; Uzbekov, R.; Silveira, J.C. da; Bauersachs, S.;
672 Mermillod, P. Deciphering the Oviductal Extracellular Vesicles Content across the Estrous
673 Cycle: Implications for the Gametes-Oviduct Interactions and the Environment of the Potential
674 Embryo. *Bmc Genomics* 2018, *19*, 622, doi:10.1186/s12864-018-4982-5.

675 68. Gardiner-Garden, M.; Frommer, M. CpG Islands in Vertebrate Genomes. *J Mol Biol* 1987,
676 *196*, 261–282, doi:10.1016/0022-2836(87)90689-9.

677 69. Cooper, D.N.; Gerber-Huber, S. DNA Methylation and CpG Suppression. *Cell Differ Dev*
678 1985, *17*, 199–205, doi:10.1016/0045-6039(85)90488-9.

679 70. Bird, A.P. CpG-Rich Islands and the Function of DNA Methylation. *Nature* 1986, *321*, 209–
680 213, doi:10.1038/321209a0.

681 71. Medvedeva, Y.A.; Fridman, M.V.; Oparina, N.J.; Malko, D.B.; Ermakova, E.O.;
682 Kulakovskiy, I.V.; Heinzel, A.; Makeev, V.J. Intergenic, Gene Terminal, and Intragenic CpG
683 Islands in the Human Genome. *Bmc Genomics* 2010, *11*, 48–48, doi:10.1186/1471-2164-11-48.

684 72. Hackenberg, M.; Previti, C.; Luque-Escamilla, P.L.; Carpena, P.; Martínez-Aroza, J.; Oliver,
685 J.L. CpGcluster: A Distance-Based Algorithm for CpG-Island Detection. *Bmc Bioinformatics*
686 2006, *7*, 446–446, doi:10.1186/1471-2105-7-446.

687 73. Aravin, A.; Gaidatzis, D.; Pfeffer, S.; Lagos-Quintana, M.; Landgraf, P.; Iovino, N.; Morris,
688 P.; Brownstein, M.J.; Kuramochi-Miyagawa, S.; Nakano, T.; et al. A Novel Class of Small
689 RNAs Bind to MILI Protein in Mouse Testes. *Nature* 2006, *442*, 203–207,
690 doi:10.1038/nature04916.

691 74. Girard, A.; Sachidanandam, R.; Hannon, G.J.; Carmell, M.A. A Germline-Specific Class of
692 Small RNAs Binds Mammalian Piwi Proteins. *Nature* 2006, *442*, 199–202,
693 doi:10.1038/nature04917.

694 75. Carmell, M.A.; Girard, A.; Kant, H.J.G. van de; Bourc'his, D.; Bestor, T.H.; Rooij, D.G. de;
695 Hannon, G.J. MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the
696 Mouse Male Germline. *Dev Cell* 2007, *12*, 503–514, doi:10.1016/j.devcel.2007.03.001.

697 76. Aravin, A.A.; Sachidanandam, R.; Bourc'his, D.; Schaefer, C.; Pezic, D.; Toth, K.F.; Bestor,
698 T.; Hannon, G.J. A PiRNA Pathway Primed by Individual Transposons Is Linked to De Novo
699 DNA Methylation in Mice. *Mol Cell* 2008, *31*, 785–799, doi:10.1016/j.molcel.2008.09.003.

700 77. Brennecke, J.; Aravin, A.A.; Stark, A.; Dus, M.; Kellis, M.; Sachidanandam, R.; Hannon,
701 G.J. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in
702 Drosophila. *Cell* 2007, *128*, 1089–1103, doi:10.1016/j.cell.2007.01.043.

703 78. Gunawardane, L.S.; Saito, K.; Nishida, K.M.; Miyoshi, K.; Kawamura, Y.; Nagami, T.;
704 Siomi, H.; Siomi, M.C. A Slicer-Mediated Mechanism for Repeat-Associated SiRNA 5' End
705 Formation in Drosophila. *Science* 2007, *315*, 1587–1590, doi:10.1126/science.1140494.

706 79. Betel, D.; Sheridan, R.; Marks, D.S.; Sander, C. Computational Analysis of Mouse PiRNA
707 Sequence and Biogenesis. *Plos Comput Biol* 2007, *3*, e222, doi:10.1371/journal.pcbi.0030222.

708 80. Marzec, M. Size Does Matter: PiRNA and MiRNA Targeting. *Trends Biochem Sci* 2021, *47*,
709 287–288, doi:10.1016/j.tibs.2021.12.001.

710 81. Berninger, P.; Jaskiewicz, L.; Khorshid, M.; Zavolan, M. Conserved Generation of Short
711 Products at PiRNA Loci. *Bmc Genomics* 2011, *12*, 46–46, doi:10.1186/1471-2164-12-46.

712 82. Oey, H.M.; Youngson, N.A.; Whitelaw, E. The Characterisation of PiRNA-Related 19mers
713 in the Mouse. *Bmc Genomics* 2011, *12*, 315–315, doi:10.1186/1471-2164-12-315.

714 83. Ichiyanagi, T.; Ichiyanagi, K.; Ogawa, A.; Kuramochi-Miyagawa, S.; Nakano, T.; Chuma,
715 S.; Sasaki, H.; Udono, H. HSP90α Plays an Important Role in PiRNA Biogenesis and
716 Retrotransposon Repression in Mouse. *Nucleic Acids Res* 2014, *42*, 11903–11911,
717 doi:10.1093/nar/gku881.

718 84. Kierszenbaum, A.L.; Tres, L.L. Structural and Transcriptional Features of the Mouse
719 Spermatid Genome. *J Cell Biology* 1975, *65*, 258–270, doi:10.1083/jcb.65.2.258.

720 85. Hecht, N.B. Molecular Mechanisms of Male Germ Cell Differentiation. *Bioessays* 1998, *20*,
721 555–561, doi:10.1002/(sici)1521-1878(199807)20:7<555::aid-bies6>3.0.co;2-j.

722 86. Steger, K. Transcriptional and Translational Regulation of Gene Expression in Haploid
723 Spermatids. *Anat Embryol* 1999, *199*, 471–487, doi:10.1007/s004290050245.

724 87. Fischer, B.E.; Wasbrough, E.; Meadows, L.A.; Randlet, O.; Dorus, S.; Karr, T.L.; Russell, S.
725 Conserved Properties of Drosophila and Human Spermatozoal MRNA Repertoires. *Proc Royal*
726 *Soc B Biological Sci* 2012, *279*, 2636–2644, doi:10.1098/rspb.2012.0153.

727 88. Corral-Vazquez, C.; Salas-Huetos, A.; Blanco, J.; Vidal, F.; Sarrate, Z.; Anton, E. Sperm
728 MicroRNA Pairs: New Perspectives in the Search for Male Fertility Biomarkers. *Fertil Steril*
729 2019, *112*, 831–841, doi:10.1016/j.fertnstert.2019.07.006.

730 89. Sun, J.; Zhao, Y.; He, J.; Zhou, Q.; El-Ashram, S.; Yuan, S.; Chi, S.; Qin, J.; Huang, Z.; Ye,
731 M.; et al. Small RNA Expression Patterns in Seminal Plasma Exosomes Isolated from Semen
732 Containing Spermatozoa with Cytoplasmic Droplets versus Regular Exosomes in Boar Semen.
733 *Theriogenology* 2021, *176*, 233–243, doi:10.1016/j.theriogenology.2021.09.031.

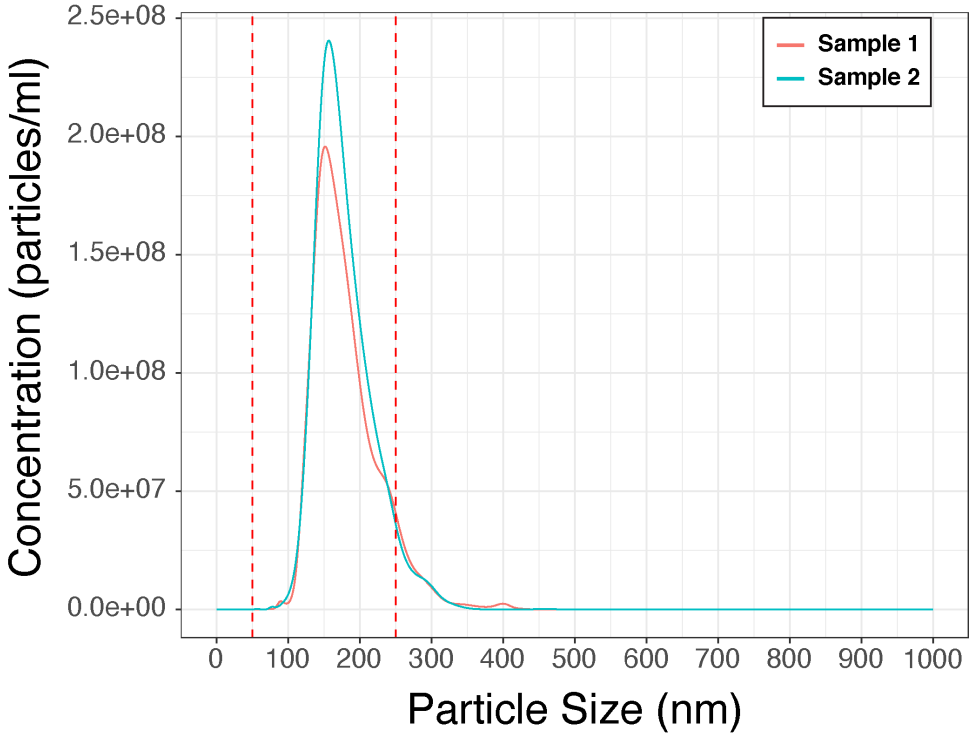
734 90. Alves, M.B.R.; Celeghini, E.C.C.; Belleannée, C. From Sperm Motility to Sperm-Borne
735 MicroRNA Signatures: New Approaches to Predict Male Fertility Potential. *Frontiers Cell Dev*
736 *Biology* 2020, *8*, 791, doi:10.3389/fcell.2020.00791.

737 91. Liu, W.-M.; Pang, R.T.K.; Chiu, P.C.N.; Wong, B.P.C.; Lao, K.; Lee, K.-F.; Yeung, W.S.B.
738 Sperm-Borne MicroRNA-34c Is Required for the First Cleavage Division in Mouse. *Proc*
739 *National Acad Sci* 2012, *109*, 490–494, doi:10.1073/pnas.1110368109.

740
741

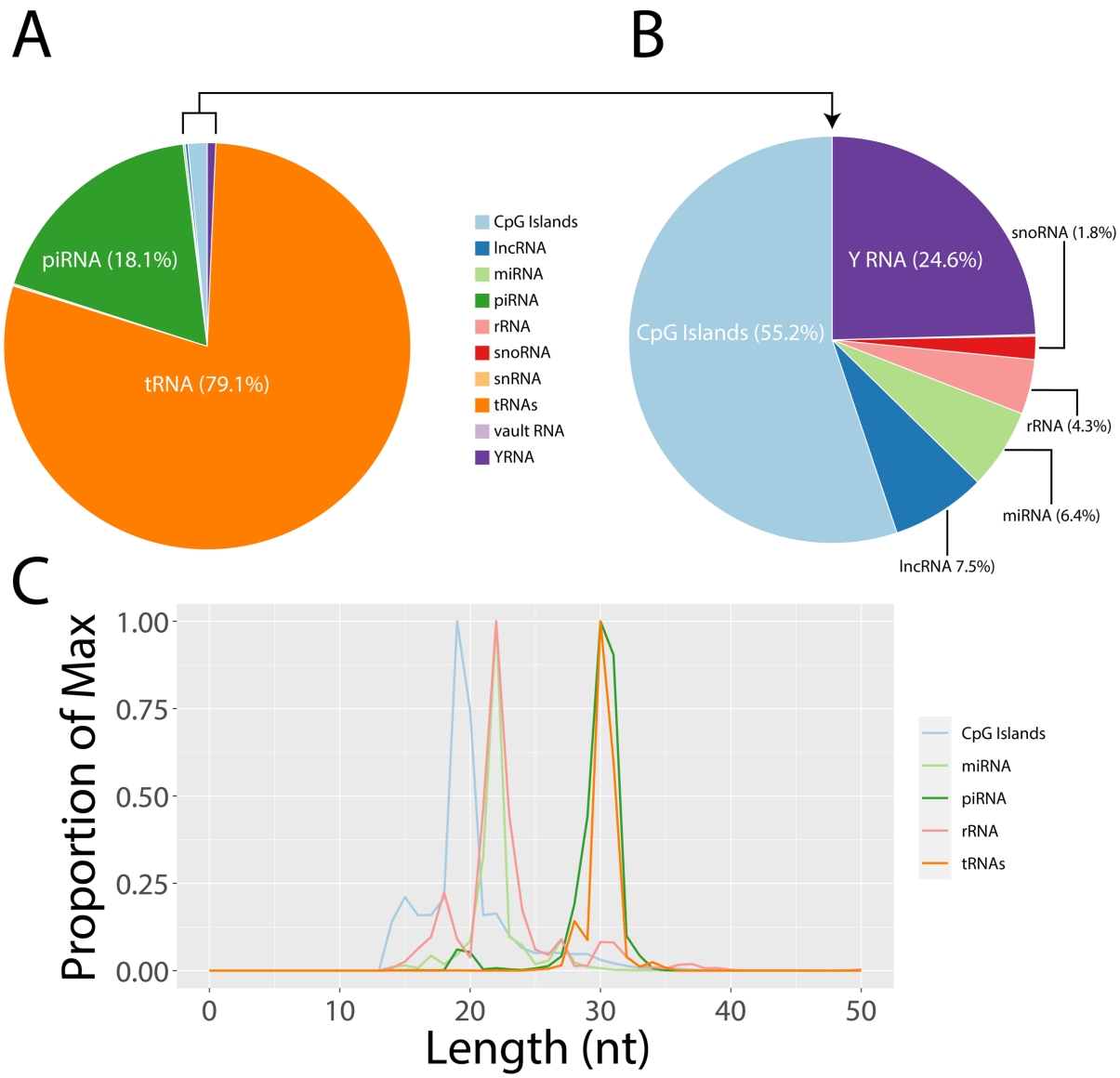
FIGURES

Figure 1.



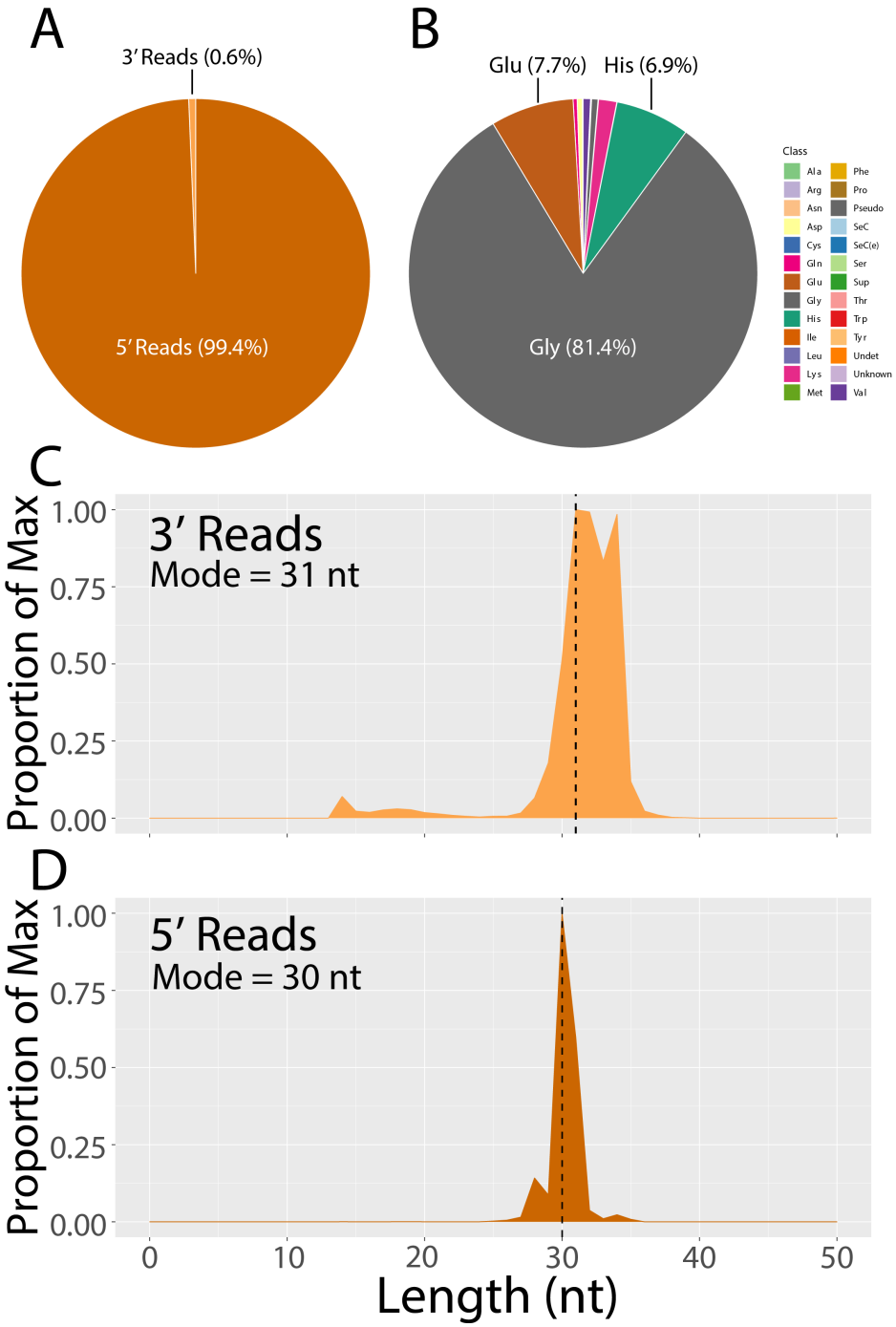
The size distribution and density of the extracellular vesicles from two separate samples are shown as analyzed by nano-particle tracking (NanoSight 300). The red dashed lines indicate the boundaries of size for canonical epididymosomes (50 – 250 nm [4]).

Figure 2.



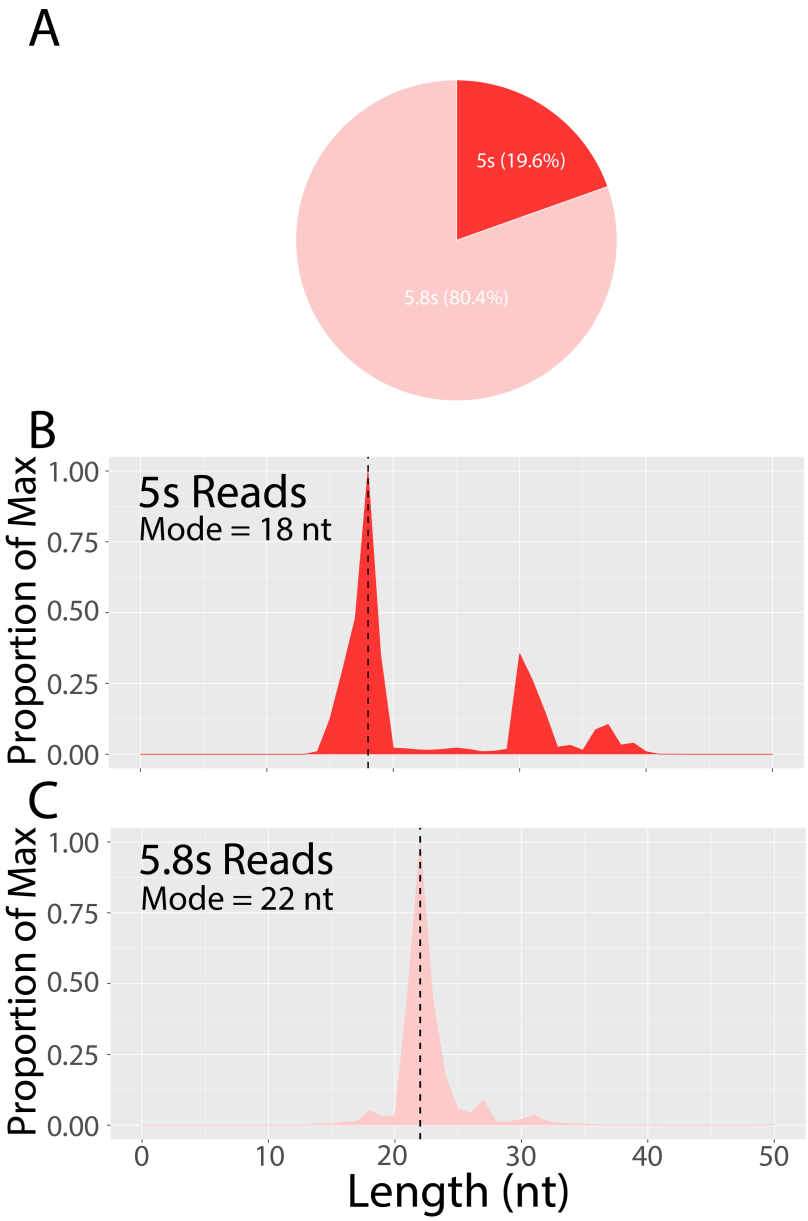
Results from sequencing small RNA derived from EVs in the rat caput are shown. **A)** The proportion of small RNA reads assigned to each of 10 annotations in our analysis; tRNA and piRNA account for the majority of small RNA in caput EVs while miRNAs are notably absent. **B)** The remaining small RNA reads are shown as a proportion of the residual 2.8% not assigned to tRNA or piRNA loci. CpG islands, which are not a known unit of small RNA expression, account for the majority of the remaining reads. **C)** The size distribution of 4 well-defined sncRNAs are shown and the one undefined category (CpG islands) demonstrates a unique size profile (19 nt). The size distribution for piRNA (29-31 nt) and miRNA (22 nt) are exactly as expected. Reads aligned to tRNA (30 - 31 nt) and rRNA (primary peak at 22 nt and small peaks at 18 nt and 30 nt) both demonstrate size specific fragmentation.

762 **Figure 3.**



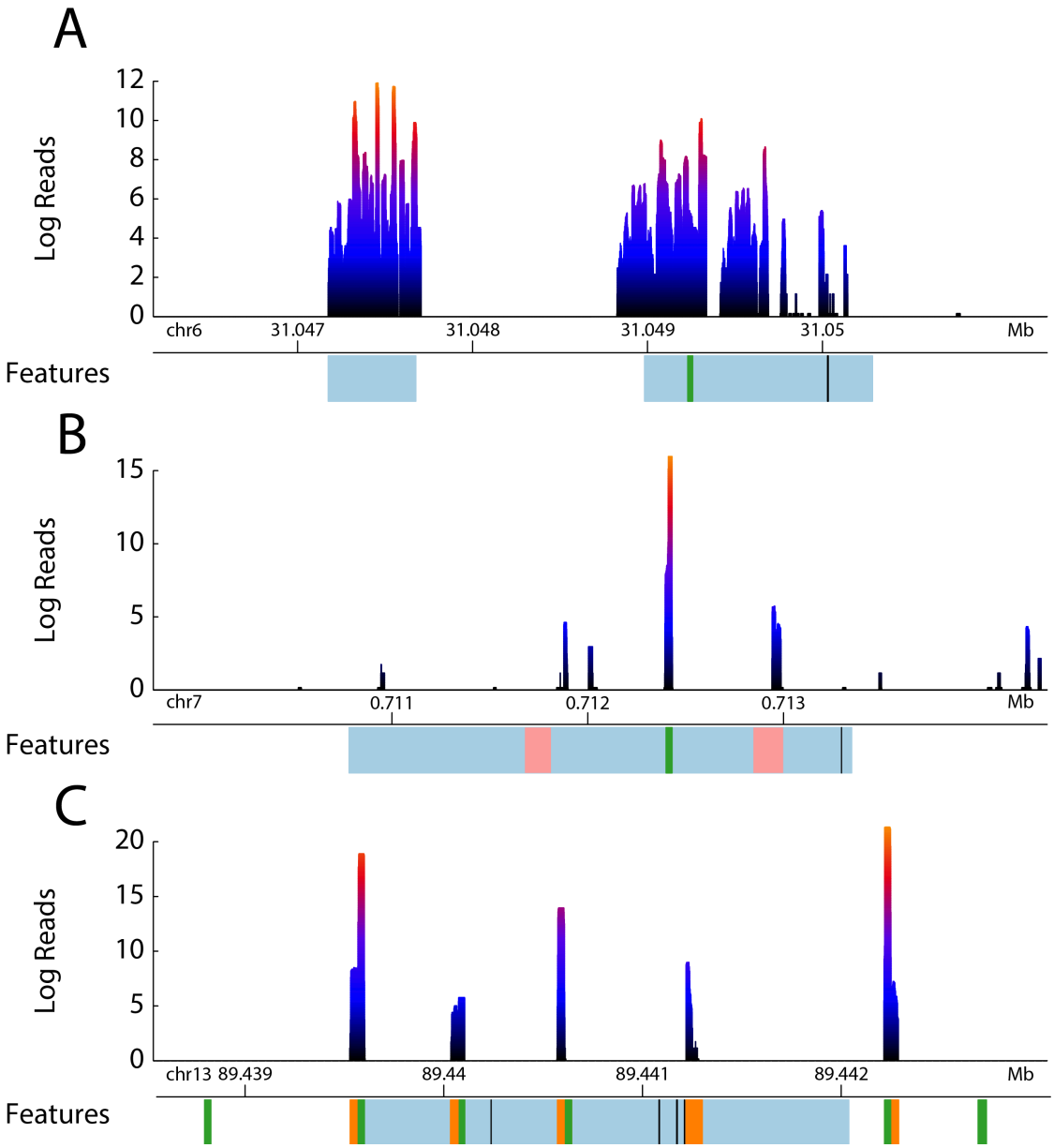
763
764 Detailed analysis of tRNA reads is shown. **A)** Reads align to the 5’ end of tRNA loci almost
765 exclusively. **B)** 5’-tRNA^{Gly} is the primary tRNA fragment found in rat caput extracellular
766 vesicles, followed by a substantially smaller proportion of 5’-tRNA^{Glu} and 5’-tRNA^{His}. **C-D)** The
767 size distribution of 3’ and 5’ tRNA reads are shown, respectively. Reads from the 3’ end of
768 tRNA are slightly longer with a broader distribution.

Figure 4.



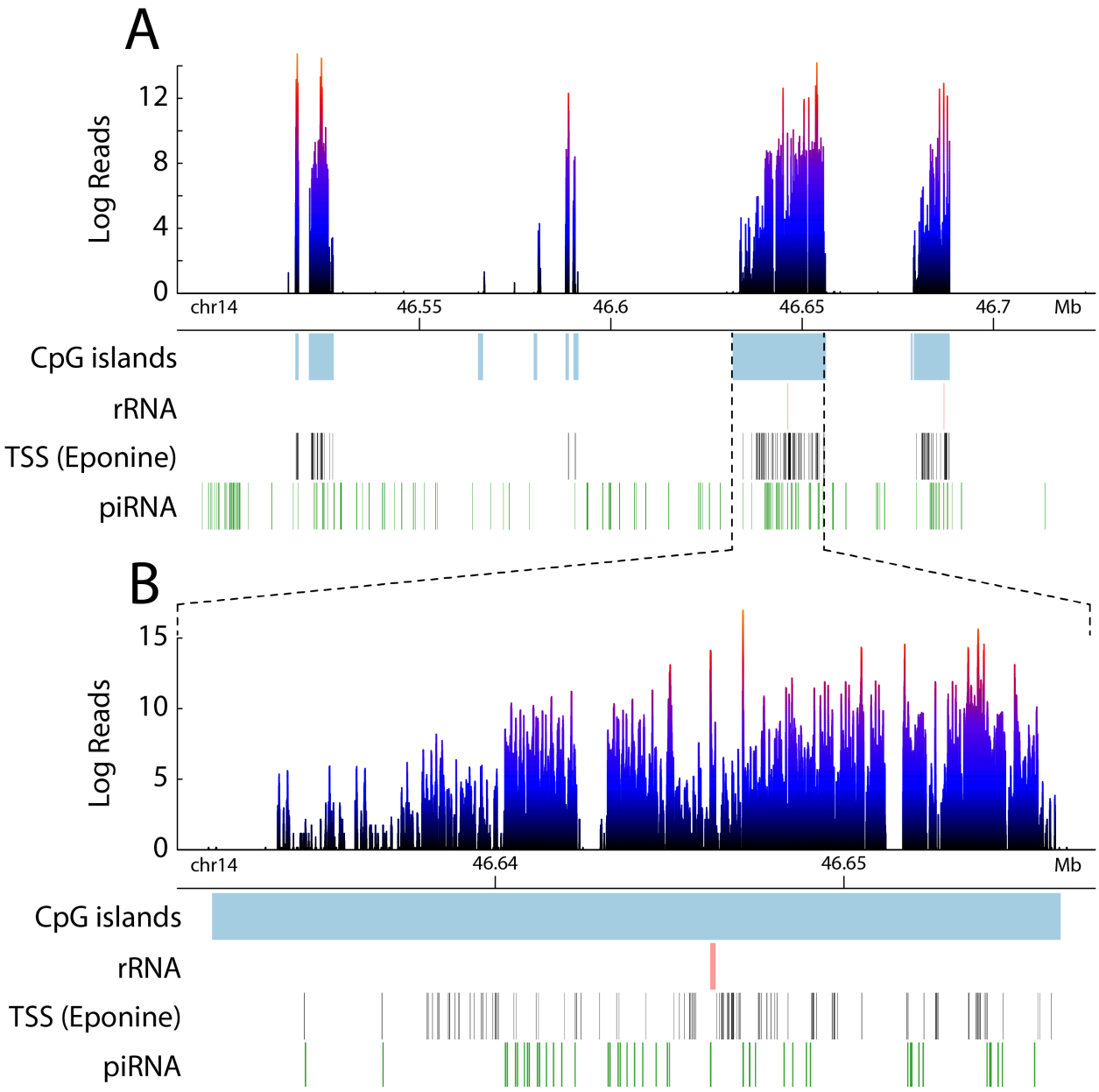
Reads that aligned to annotated rRNA loci are shown by subcategory. The rat genome is poorly annotated for rRNA, so the available 5s and 5.8s loci were used for categorization. **A)** The majority of rRNA reads aligned to 5.8s rRNA. **B-C)** The size distribution for reads aligned to 5s and 5.8s rRNA show profiles indicative of rRFs that are unique from one another. **B)** 5s rRFs show a bimodal distribution with peaks at 18 and 30 nt. **C)** 5.8s rRFs show a single peak at 22 nt.

Figure 5



The log transformed density of reads is shown across three CpG islands with different characteristics. Overlying feature annotations are shown indicated by color; Light blue = CpG island, Green = piRNA, Black = predicted transcriptional start sites, red = 5.8s rRNA, and orange = tRNA. **A)** Reads are shown expressed from within the boundaries of a CpG island on chromosome 6. The reads were not associated with the piRNA (green) within the 3' CpG island while the 5' CpG island had no other overlying features. **B)** Reads were identified within a CpG island on chromosome 7 but the majority of them aligned to a piRNA (green) and rRNA (red) feature. **C)** Reads were identified within a CpG island on chromosome 13 but they aligned exclusively to piRNA and tRNA loci and are demonstrative of their typical expression profiles. The CpG islands from **B & C** (and all other CpG islands like these) were removed from the analysis of small RNAs expressed from CpG islands. The two in **A** are representative of those that were used for further analysis.

Figure 6



The log transformed reads from a cluster of 8 CpG islands on chromosome 14 are shown with the primary overlying features indicated by color and label. Light blue = CpG islands, Red = 5.8s rRNA, Black = predicted transcriptional start sites, Green = piRNA. **A)** The boundaries of the reads shown are clearly demarcated by CpG islands and not the overlying features such as rRNA or piRNA. **B)** A detailed view of the largest CpG island in the chromosome 14 cluster. The reads are not associated with either the rRNA or piRNA features within the CpG island and are not typical of those features found elsewhere (Figure 5B & C).

TABLES

Table 1

Class	Mode	Median	Mean	SD	90% Range
CpG Islands	19	19	19.89	3.75	15 - 28
lncRNA	22	22	22.27	4.98	15 - 33
miRNA	22	22	22.04	2.34	19 - 27
piRNA	30	30	29.60	2.50	26 - 32
rRNA	22	22	22.65	4.06	17 - 31
snoRNA	27	24	23.44	5.12	15 - 32
snRNA	16	16	18.05	3.85	15 - 26
tRNAs	30	30	30.19	1.18	28 - 31
vault RNA	18	19	22.99	8.02	14 - 41
Y RNA	31	31	28.33	3.89	22 - 31

Descriptive statistics of the read length for each annotated small RNA category are shown with the mode, median, mean, standard deviation, and length range from which 90 percent of the reads are found.

Table 2

Name	Reads per Million	Found in Sperm	Found in EVs
miR-1b	16.2	--	--
miR-7a-1	19.7	[88]	[89]
miR-7a-2	19.9	[88]	[89]
miR-let-7c	10.0	[88]	[89]
miR-let-7i	11.7	[88]	[89]
miR-26a	11.7	[90]	--
miR-99a	87.8	[91]	--
miR-143	29.5	[19]	[19]
miR-148a	93.0	[88]	[89]
miR-184	493.7	[88]	[89]
AABR07005004.2	16.0		
AABR07029907.2	34.7		
AABR07057421.1	11.9		
AABR07064716.2	34.2		
AABR07064724.1	22.1		

Fifteen miRNA were identified as having substantial (> 10 RPM) expression in rat caput EVs, 10 of which are annotated and 5 that are putative. This is far fewer than expected based on mice experiments. The majority of miRNAs identified have previously been reported in either sperm or EVs (citation indicated), except for miR-1b. A miRNA considered a hallmark of caput EVs (miR-143) was also identified in our rat data set.

812 Table 3.

Chromosome	Start Location	End Location	Length (nt)	Read Count	RPKM	piRNA Loci	pRNA Protein Loci	rRNA Binding Protein Loci	rRNA Loci
Chr 1	11,903,555	11,916,838	13,283	112,070	63.16	19	1	0	1
Chr 1	11,961,297	11,975,108	13,811	278,633	151.02	27	1	0	1
Chr 5	91,124,274	91,141,016	16,742	341,585	152.72	26	1	0	1
Chr 6	30,627,068	30,633,859	6,791	177,949	196.15	13	0	0	2
Chr 6	31,047,175	31,047,676	501	10,347	154.59	0	0	0	0
Chr 6	31,048,983	31,050,286	1,303	3,282	18.85	1	0	0	0
Chr 14	46,517,594	46,518,401	807	113,831	1,055.86	1	0	0	0
Chr 14	46,521,126	46,527,599	6,473	170,612	197.30	10	0	1	0
Chr 14	46,588,206	46,589,034	828	39,955	361.21	0	0	1	0
Chr 14	46,590,294	46,591,521	1,227	2,562	15.63	1	0	1	0
Chr 14	46,631,880	46,656,213	24,333	429,235	132.04	44	0	1	1
Chr 14	46,679,234	46,688,547	9,313	160,403	128.93	17	0	0	1

813

814 The genomic coordinates (Rnor v6 – Chromosome, Start Location, End Location) of 12 CpG
815 islands found to have substantial expression (> 10 RPKM) of ~19nt small RNA transcripts are
816 shown along with the length of the CpG island, raw read count, and length-corrected read count
817 (RPKM) found within each. The number of various overlying features found overlapping each
818 CpG island is also shown.